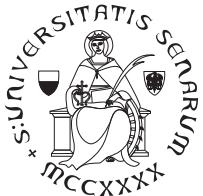


UNIVERSITA' di SIENA

DIPARTIMENTO DI SCIENZE FISICHE, DELLA TERRA E DELL' AMBIENTE
SEZIONE DI FISICA



UNIVERSITÀ
DI SIENA
1240

**Search for new resonances in p-p
collisions using fully leptonic W^+W^-
decays with the CMS detector**

Tesi di Dottorato di Ricerca in Fisica Sperimentale
In partial fulfillment of the requirements fo the Ph.D.
Thesis in Experimental Physics

Candidato:
Dr. Lorenzo Russo

Supervisor:
Prof. Vitaliano Ciulli

Tutor:
Prof.ssa Maria Agnese Ciocci

dedica a...

Abstract

This thesis presents a search for a possible heavy Higgs boson, X , decaying into a pair of W bosons, in the mass range from 200 GeV to 3 TeV. The analysis is based on proton-proton collisions recorded by the CMS experiment at the CERN LHC in 2016, corresponding to an integrated luminosity of 35.9 fb^{-1} at $\sqrt{s} = 13 \text{ TeV}$. The W boson pair decays are reconstructed in the $2\ell 2\nu$ and $\ell\nu q\bar{q}$ final states. Both gluon fusion and electroweak production of the scalar resonance are considered. Dedicated event categorizations, based on the kinematic properties of the final states, are employed for an optimal signal-to-background separation. Combined upper limits at the 95% confidence level on the product of the cross section and branching fraction excludes a heavy Higgs boson with Standard Model-like couplings and decays in the range of mass investigated.

Contents

Introduction	ix
1 The Standard Model, the Higgs Boson and New Scalar Particles	1
1.1 Phenomenology of the Standard Model	1
1.2 The Higgs Boson	5
1.3 New Scalar Particles	12
1.4 The WW decay channel for high mass Higgs-like particles	18
2 The CMS experiment at LHC	25
2.1 The Large Hadron Collider	25
2.2 The Compact Muon Solenoid experiment	26
2.3 Data samples and future plans	35
3 Monte Carlo Generators	39
3.1 General Overview	39
3.2 Hard process	41
3.3 Parton shower	43
3.4 Multiple Interaction	48
3.5 Hadronization	48
3.6 Hadronic Decays and Electromagnetic Radiation.	51
3.7 Jets	51
3.8 Main Monte Carlo generators	51
3.9 Monte Carlo samples in High Mass Analysis	53
4 Event Reconstruction	59
4.1 The Particle Flow	59
4.2 Tracking	60
4.3 Calorimeter clusters	62
4.4 Link algorithm	63
4.5 Lepton reconstruction and identification	63
4.6 Jet reconstruction and identification	67
4.7 b-jet identification	69
4.8 The Missing Transverse Energy	72
4.9 Fake Lepton Background Estimation	74

5 Search for high mass resonances	77
5.1 Overview of the fully leptonic analysis	77
5.2 Main Background processes	78
5.3 Lepton Efficiencies from Tag and Probe Method	80
5.4 Data sample and Trigger used	81
5.5 Study of the Interference effects	81
5.6 Discriminating variable	83
5.7 Opposite Flavor final state	87
5.8 Same Flavor final state	100
5.9 Systematic uncertainties	110
6 Semileptonic final state	115
6.1 Overview of the semileptonic analysis	115
6.2 Jets for Boosted and Resolved selection	116
6.3 Event categorisation	117
6.4 Backgrounds	120
6.5 Systematic uncertainties	120
7 Results and Interpretation	123
7.1 Statistical interpretation	123
7.2 Signal interpretation: EW singlet and 2HDM	125
7.3 Fully and same leptonic results	127
7.4 Combination among the fully and the semi leptonic analysis	129
Conclusion	137
A Boost-invariant variables	139
B Parton Distribution Function	141

Introduction

The discovery of the Higgs boson by the LHC experiments ATLAS and CMS in 2012 has been a major breakthrough in modern experimental particle physics. The discovered particle is compatible with the Standard Model (SM) Higgs mechanism predictions: the only unknown parameter, the boson’s mass, has been measured to be close to 125 GeV. Nevertheless, in order to determine whether the SM Higgs sector is complete, precise measurements of the Higgs boson coupling strengths, CP structure and transverse momentum are required. A complementary and important strategy is the search for additional heavy scalars, that would prove the presence of beyond-the-SM (BSM) physics in the form of a non-minimal Higgs sector. The existence of such a sibling Higgs boson, denoted X , is motivated in many BSM scenarios, so the search for additional scalar resonances in the full mass range accessible at the LHC remains one of the main objectives of the experimental community. The search for a high mass Higgs boson has been performed using Run-I and early Run-II data in many different decay channels and upper limits on its cross section have been determined as a function of the X mass. With the full 2016 data collected by CMS experiment at $\sqrt{s} = 13$ TeV, approximately 36 fb^{-1} , it is now possible to set very tight upper limits on high mass Higgs boson cross section. One of the most sensitive decay channels, for masses above 200 GeV, is to a pair of W bosons. The fully leptonic, $2\ell 2\nu$, and the semi leptonic, $\ell\nu q\bar{q}$ final state (with $\ell = e$ or μ), are considered in this analysis. The fully leptonic channel has a clear signature due to the presence of the two isolated leptons and moderate missing-transverse-energy (MET) that provides indirect evidence of the neutrinos presence. For the semi leptonic channel the leptonically decaying boson is reconstructed as a single isolated lepton and MET. If the hadronic decay products are resolved then the boson decay products may be reconstructed using two quark-jets. The search is performed in a wide range of masses from 200 GeV up to 3 TeV. The events are classified in different categories, optimised for the gluon-gluon fusion and for the vector-boson fusion (VBF) production mechanisms. The signal is interpreted in terms of the electroweak (EW) singlet model, including a detailed simulation of the interference between X signal, SM Higgs boson and WW backgrounds. In addition, since Two-Higgs-doublet models (2HDM) are a well motivated extension of the SM, a final interpretation of the results in term of 2HDM models is also performed.

The general picture of the SM, the relative open question and the Higgs boson sector are described in the first chapter, Cap. 1. The Electroweak Singlet model and Two Higgs Doublet Models, used in this analysis, are also introduced in the same chapter. In Cap. 2, the Compact Muon Solenoid (CMS) experiment at the CERN

Large Hadron Collider (LHC) is described, focusing the attention on its components, useful to reconstruct the hunted final objects (leptons and hadrons). Due to the complexity of hadron-hadron collision at LHC, Monte Carlo generators are used to simulate the realistic result of such collisions. The various simulations steps and the review of the main Monte Carlo generators are presented in Cap. 3. The reconstructed objects and the identification algorithms used to obtain the high level objects, targeting in the analysis, are illustrated in Cap. 4. In the Cap. 5, starting from the reconstructed objects, the analysis strategy for fully leptonic final state is described, focusing the attention on the signal modellization and signal identification. Following the semi leptonic final state analysis is described, Cap. 6. The results for the the fully leptonic and same leptonic analysis including their combination are shown in Cap. 7. The analysis is under review of the CMS collaboration and it is targeting to journal publication in the first months of 2019.

This work is partially funded by MIUR Italy, under contract 2012Z23ERZ of PRIN 2012, “H-TEAM: Trigger, Elettronica Avanzata e Metodi innovativi per misure di precisione nel settore dell’ Higgs ad LHC”.

Chapter 1

The Standard Model, the Higgs Boson and New Scalar Particles

In this chapter is introduced the theoretical bases for the research performed in this analysis. The Standard Model, successfully describing the elementary particles and three of the four fundamental interactions, is brief overview in the beginning, including the Higgs sectors and the experimental results. However there are still open question (dark matter, dark energy, matter-antimatter asymmetry) in the SM. Two Beyond the SM models, that could resolve the pending issues, are introduced: the Electroweak Singlet and the Two Higgs Doublet Model. These models predict the existence of high mass Higgs-like particle(s). A completed description of such models is performed in the last section of the chapter, including the experimental results. I have contributed personally in exclusion limits on the cross section for high mass particle using 2015 data, Fig. 1.16.

1.1 Phenomenology of the Standard Model

The Standard Model (SM) of particle physics [1] is a description of the nature which best explain the fundamental structure of matter and the fundamental forces which govern all known phenomena. The SM gives a quantitative description of three of the four interactions in nature: electromagnetism, weak interactions and strong nuclear force. Developed in the early 1970s by Glashow [2], Weinberg [3] and Salam [4], it has successfully explained almost all experimental results and precisely predicted a wide variety of phenomena. It is a renormalizable quantum field theory, compatible with special relativity.

General Picture

The main constituents of the SM are shown in Fig. 1.1. These are the particle composing the ordinary matter and responsible of the forces. The particles involved are characterized by the spin, the mass, and the quantum numbers determining their interactions. The quarks are subject to all the three forces and, in particular, are the only fermions to possess a “colour” charge, which is responsible of the strong nuclear force, as described by Quantum Chromo Dynamics (QCD). Because of the QCD

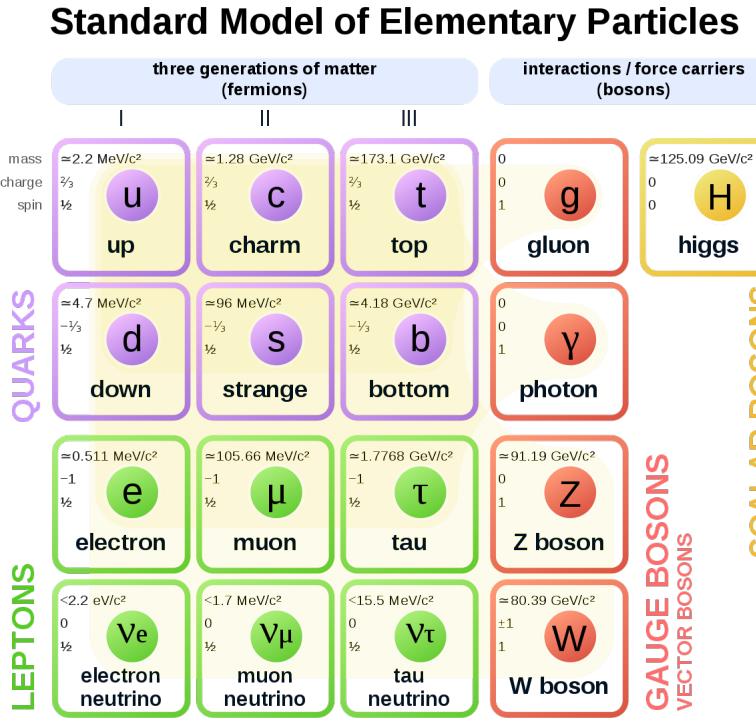


Figure 1.1. Main constituents of the Standard Model.

colour confinement properties, quarks do not exist as free states but can be experimentally observed only as bound states. The proton and neutron, are composed by three quarks (called baryons). The particle composed by a quark-antiquark are called “meson”. Quark flavour is conserved in electromagnetic and strong interactions but not in weak ones, as quark mass eigenstates do not correspond to the weak interaction eigenstates. Their mixing is described by the Cabibbo–Kobayashi–Maskawa (CKM) matrix. The leptons have no colour charge and are subject only to the electromagnetic and weak forces. The charged leptons of the three families are respectively denoted as the electron (e), muon (μ) and tau lepton (τ). The only stable lepton is the electron. To each lepton corresponds a neutrino. The mass of neutrino is unknown but their flavour oscillations prove a non-zero mass [5]. The gluons, the W^\pm bosons, the Z boson and the photon (γ) are boson that compose the SM gauge sector. The gluons are the mediators of the strong interactions. They are massless, electrically neutral and carry color quantum number and they can interact with themselves. The W^\pm and Z bosons are the mediator of the weak interactions. Their mass is $\sim 81 \text{ GeV}$ and $\sim 90 \text{ GeV}$ respectively. These particles are unstable and decay in other particles. Finally the photon is massless, chargeless, non self-interacting and mediates the electromagnetic interactions.

To summarize, the SM Lagrangian may be written as the sum of three parts:

$$\mathcal{L}_{SM} = \mathcal{L}_{QCD} + \mathcal{L}_{EWK} + \mathcal{L}_H , \quad (1.1)$$

where \mathcal{L}_{QCD} is the quantum chromodynamics Lagrangian that describes the interactions of quarks and gluons, the \mathcal{L}_{EWK} is the electroweak Lagrangian that describes the interactions of the fermions with the Z and W^\pm bosons. The \mathcal{L}_H is the Higgs part of the Lagrangian.

Quantum Chromodynamics

The Quantum Chromodynamics (QCD) describes the interactions of quarks and gluons, mediated by the strong force through the colour charge. The QCD Lagrangian has the local gauge invariance under the $SU(3)_C$ group and it is given by [6],

$$\begin{aligned} \mathcal{L}_{QCD} = & \sum_q [\bar{\psi}_{q,a}(i\gamma^\mu \partial_\mu \delta_{ab} - g_s \gamma^\mu t_{ab}^C A_\mu^C - m_q \delta_{ab}) \psi_{q,b} \\ & - \frac{1}{4} F_{\mu\nu}^A F^{A\mu\nu}] , \end{aligned} \quad (1.2)$$

where repeated indices are summed over. The γ^μ are the Dirac γ -matrices. The $\psi_{q,a}$ are quark-field spinors for a quark of flavor q and mass m_q , with a color-index a that runs from $a = 1$ to $N_c = 3$, i.e. quarks come in three “colors”. Quarks are said to be in the fundamental representation of the $SU(3)$ color group. The A_μ^C correspond to the gluon fields, with C running through the eight kinds of gluon. The t_{ab}^C correspond to eight 3×3 matrices and are the generators of the $SU(3)_C$ group. The quantity g_s is the QCD coupling constant. Finally, the field tensor $F_{\mu\nu}^A$ is given by,

$$F_{\mu\nu}^A = \partial_\mu A_\nu^A - \partial_\nu A_\mu^A - g_s f_{ABC} A_\mu^B A_\nu^C , \quad (1.3)$$

$$[t^A, t^B] = i f_{ABC} t^C , \quad (1.4)$$

where the f_{ABC} are the structure constants of the $SU(3)_C$ group. Neither quarks or gluons are observed as free particles. Hadrons are color-singlet (i.e. color-neutral) combinations of quarks, anti-quarks, and gluons. The fundamental parameters of QCD are the coupling g_s (or $\alpha_s = \frac{g_s^2}{4\pi}$) and the quark masses m_q . If the quark masses are fixed, there is only one free parameter in the QCD Lagrangian, that is α_s . This constant is not a physical observable but rather a quantity defined in the context of perturbation theory, which enters in the prediction for experimental observables.

Electroweak Interaction

The electroweak interactions are based on the gauge group $SU(2)_L \otimes U(1)_Y$. The $SU(2)_L$ group refers to the weak isospin charge (I), and $U(1)_Y$ to the weak hypercharge (Y). Left-handed (L) fermions are paired in $I = 1/2$ isospin doublets, whereas right-handed (R) fermions are $I = 0$ singlets. The presence of these local gauge symmetries introduces four vector bosons: three for the $SU(2)$ group, the W_i fields ($i = 1, 2, 3$), and one for $U(1)$, the B field. This gives rise to a quantum field theory,

invariant under local gauge symmetries, whose Lagrangian is expressed as:

$$\mathcal{L}_{EW} = \sum_f \bar{\psi} i\gamma^\mu \mathcal{D}_\mu \psi - \frac{1}{4} F_{\mu\nu} F^{\mu\nu} - \frac{1}{4} \vec{E}_{\mu\nu} \cdot \vec{E}^{\mu\nu}, \quad (1.5)$$

where the sum is extended over all the fermions f and where covariant derivatives which preserve the local gauge invariance have the following form:

$$\begin{aligned} \mathcal{D}_\mu &= \partial_\mu + ig\vec{W}_\mu \cdot \frac{\vec{\tau}}{2} + i\frac{g'}{2}YB_\mu, \\ F_{\mu\nu} &= \partial_\mu B_\nu - \partial_\nu B_\mu, \\ E_{\mu\nu}^\alpha &= \partial_\mu W_\nu^\alpha \partial_\nu W_\mu^\alpha - g\epsilon^{\alpha\beta\gamma}W_\mu^\beta W_\nu^\gamma, \end{aligned} \quad (1.6)$$

where $\vec{\tau}$ indicates the three Pauli matrices, g and g' are the coupling constants which correspond respectively to $SU(2)_L$ and $U(1)_Y$. The physical fields are obtained as linear combinations of these fields:

$$\begin{aligned} A_\mu &= \sin\theta_W W_\mu^3 + \cos\theta_W B_\mu, \\ Z_\mu &= \cos\theta_W W_\mu^3 - \sin\theta_W B_\mu, \\ W_\mu^\pm &= \frac{W_\mu^1 \mp W_\mu^2}{\sqrt{2}}. \end{aligned} \quad (1.7)$$

The above equations represent two neutral particles (the photon, described by the A_μ field, and the Z boson) and two charged particles (the W^+ and W^- bosons). We have further introduced the angle θ_W which is known as the weak mixing angle or Weinberg angle. Up to here, the theory is necessarily incomplete: all particles it describes are massless, contradicting experimental evidence. The Lagrangian symmetries, on the other hand, seem to forbid the introduction of mass terms without spoiling its gauge invariance. Higgs' proposal solves this problem by spontaneously breaking the Lagrangian symmetry, Sec. 1.2.

Experimental evidence

The experimental study of the Standard Model has made a quantum leap in the last 30 years. First of all the theory predicted the existence of W^\pm and Z bosons in the mass range from 60 to 93 GeV [7]. In 1976 Rubbia, Cline and McIntyre proposed the transformation of an existing high-energy proton accelerator into a proton–antiproton collider as a quick and relatively cheap way to achieve collisions above threshold for W and Z production. This proposal was adopted at CERN Super Proton Synchrotron (SPS) collider and the first proton–antiproton collisions were collected in 1981. In the following years the W and Z bosons have been observed by UA1 and UA2 experiments with a mass of 80 GeV and 91 GeV respectively, Fig. 1.2. With the electron-positron colliders at a center-of-mass energy equal to Z mass, precise measurements of the fundamental parameters of electroweak theory could be made. In 1989, two e^+e^- colliders operation started: the Stanford Linear Collider (SLC) at SLAC, and the circular Large Electron Positron collider (LEP) at CERN. Precision electroweak tests covering the measurements at the Z pole have been conducted by SLC and LEP experiments. W boson properties were also

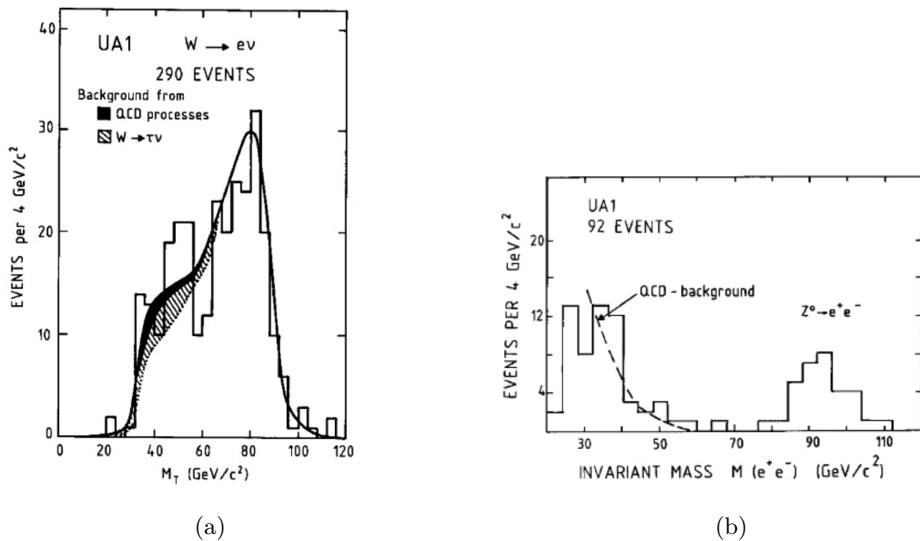


Figure 1.2. (a) Transverse mass distribution for all $W \rightarrow e\nu$ events recorded by UA1 between 1982 and 1985. (b) Invariant mass distribution of all e^+e^- pairs recorded by UA1 between 1982 and 1985

measured at the LEP collider, which reach 209 GeV center-of-mass energy, well above the threshold for W^+W^- pair production, Fig. 1.4. The results of measurements confirmed the SM prediction. The next missing SM piece was top quark. It was a necessary component of the SM of electroweak interactions, but there was no consistent theoretical guidance as to what its mass should be. The only way to observe a top quark with such a high mass was at the collider with the highest-energy, the Tevatron antiproton-proton collider at Fermilab. The existence of the top quark was firmly established in 1995 with simultaneous announcements by both the CDF [8] and the DO [9] experiments with results that demonstrated a mass of around 174 GeV, Fig.1.5. After the top quark discovery, the only missing part to the SM was the Higgs boson particle that has been observed at ATLAS and CMS experiment at LHC proton-proton collider at CERN in 2012 (see Sec.1.2).

1.2 The Higgs Boson

A Lagrangian is said to have a symmetry when it is invariant under a group of transformations. However the fact that the weak gauge bosons have a mass different from zero indicates that the gauge symmetry is broken. Also the fermion masses can not be included without violating gauge symmetry in the \mathcal{L}_{QCD} . The mass terms can be introduced with the Spontaneous Symmetry Breaking Mechanism, adding \mathcal{L}_H , that gives mass to the weak bosons and fermions and leaves the photon massless. This mechanism has been proposed in 1964 independently by Higgs [10] and Brout and Englert [11]. With the spontaneous symmetry breaking mechanism, a new particle which couples to the massive fermions and to the boson emerges. This particle is called Higgs boson and its mass is a free parameter of the theory. In 2012, 48 years after this hypothesis was formulated, the Higgs boson has been

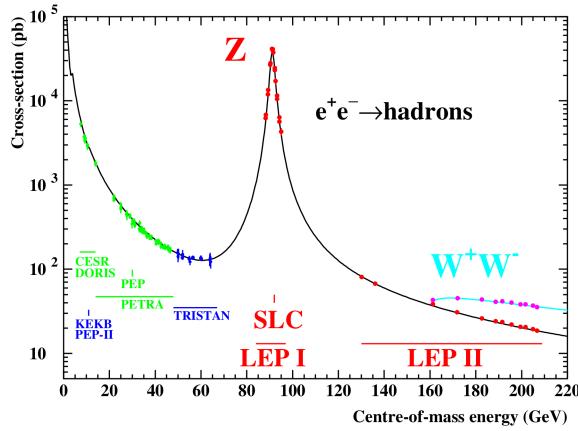


Figure 1.3. The cross-section for the production of hadrons in e^+e^- annihilations. The measurements are shown as dots with error bars. The solid line shows the prediction of the SM



Figure 1.4. The measured W-pair production cross section compared to the SM and alternative theories not including trilinear gauge couplings.



Figure 1.5. Top quark mass measurements.

observed by ATLAS [12] and CMS [13] experiments at LHC.

The Brout–Englert–Higgs mechanism

The symmetry of SM Lagrangian, \mathcal{L}_{SM} , is $SU(2)_L \otimes U(1)_Y \otimes SU(3)_C$, where L , Y and C refers to isospin, hypercharge and color quantum numbers, respectively. To break the symmetry, a scalar field Φ (Higgs field) is introduced. The field is an isospin doublet ($SU(2)_L$):

$$\Phi = \begin{pmatrix} \Phi^+ \\ \Phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \Phi_1 + i\Phi_2 \\ \Phi_3 + i\Phi_4 \end{pmatrix}$$

where Φ_j with $j = 1, 2, 3, 4$ are real fields used to manifest the complexity of Φ^+ and Φ^0 . The simplest Lagrangian of a self-interacting scalar field is,

$$\mathcal{L}_H = (D^\mu \Phi^\dagger)(D_\mu \Phi) - \mu^2 \Phi^\dagger \Phi - \lambda (\Phi^\dagger \Phi)^2, \quad (1.8)$$

where λ needs to be positive for the potential to be bounded from below and μ^2 is a mass term for the Φ field. The ground state (vacuum) of the theory is defined as the state where the energy density is at a minimum. If the μ parameter is chosen so that $\mu^2 < 0$, the symmetry of the potential may be broken, and the minimum value is,

$$v \equiv \sqrt{\frac{-\mu^2}{\lambda}} = \Phi^\dagger \Phi. \quad (1.9)$$

The choice for the sign of the parameters μ^2 and λ gives to the potential $V(\Phi) = \mu^2 \Phi^\dagger \Phi + \lambda (\Phi^\dagger \Phi)^2$ the “mexican hat” shape, as illustrated in Fig. 1.6. In the

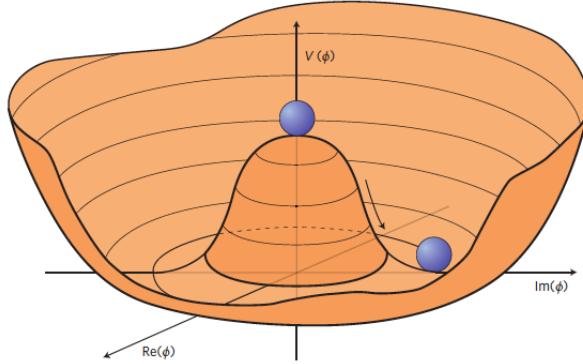


Figure 1.6. Higgs fields potential with two degree of freedom.

perturbation theory, the Φ fields is expanded around the minimum, that is chosen among the set of states which satisfy Eq. 1.9. All this states break the rotational symmetry of the Lagrangian. The Φ field is expressed as,

$$\Phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}. \quad (1.10)$$

The $h(x)$ field gives a particle with mass equal to,

$$m_H = \sqrt{2}\mu = v\sqrt{2\lambda}, \quad (1.11)$$

that is a free parameter of the theory. However some theoretical constrains could be imposed. The value of v parameter in Eq. 1.11 could be determinate by the Fermi constant, G_F , as,

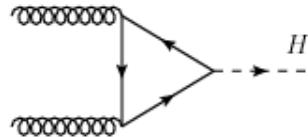
$$v = \frac{2m_W}{g} = (\sqrt{2}G_F)^{-1/2}, \quad (1.12)$$

with the actual G_F measurements obtained with the muon lifetime [14], a value of ~ 250 GeV is obtained for v . However the model is not predictive on the value of the λ parameter. Nevertheless, additional theoretical arguments place approximate upper and lower bounds on m_H [6]. The lower bound on the Higgs mass is given by the vacuum state stability, that leads to requiring λ to be positive at all energies. The upper edge is imposed by the Planck scale. So a Higgs boson in a range of $130 < m_H < 180$ GeV is consistent with the theory.

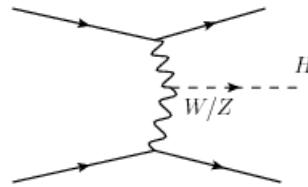
The Higgs boson at LHC

The Higgs boson has been searched in several experiments (Fig. 1.7) located at different colliders (LEP, SLC, Tevatron) without a clear evidence of such particle. Therefore, the Large Hadron Collider, a proton-proton collider located at CERN, has been designed with the primary goal of discovering or excluding the Higgs boson. In a proton-proton collision the Higgs boson can be produced in different ways: via gluon-gluon fusion, the vector boson fusion (VBF), the vector-boson associated production and the top-quark associated production. Below are the details of the different mechanisms:

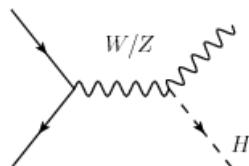
- *Gluon-gluon fusion* ($gg \rightarrow H$): this is the main Higgs boson production mode. Here a couple of gluons interact via a heavy quark loop and give rise to a Higgs boson.



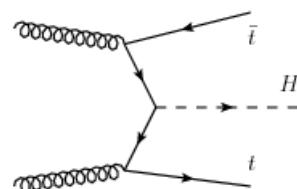
- *Vector Boson Fusion* (VBF) ($qq \rightarrow qqH$): each incoming quarks emits a virtual W or Z boson that interact generating a Higgs boson. The quarks after emitting the vector bosons proceed in the forward direction and represent the peculiar signature of this production mode.



- *Vector boson associated production* (or *Higgsstrahlung*): the Higgs boson is emitted from a W^\pm or a Z boson which has been produced by a pair quark-antiquark.



- *Top-quark associated production*: a pair of top quarks, originated from the splitting of two incoming gluons, interacts to yields to a Higgs boson.



The Higgs boson cross section of the different production modes, depend on the center of mass energy of the collider and by the supposed mass of the particle. The



Figure 1.7. An example of simulated data modeled for the CMS particle detector on the Large Hadron Collider (LHC) at CERN.

cross section increases with the energy of the colliding particles (\sqrt{s}) and decreases with the mass of the boson, m_H . The second behaviour is shown in Fig. 1.11. The Higgs boson is an unstable particle and decays in a variety of different final states, Fig 1.12. For a Higgs boson of mass $m_H \sim 125$ GeV (that is the mass measured by ATLAS and CMS) the channel with the largest branching ratio is in $b\bar{b}$ quark, followed by the W^+W^- , $\tau^+\tau^-$ and ZZ channels.

Experimental Results

The discovery or exclusion of the SM Higgs boson was one of the primary scientific goals of LHC. On the 4th of July 2012, as the observation of a new boson by the ATLAS [12] and the CMS [13] collaborations has been announced, as an excess of events near 125 GeV was reported by both experiments, Fig. 1.8. Latest results from both the experiments confirm that the properties of the discovered particle are consistent with the hypothesis of being the Higgs boson predicted by the Standard Model. The measurements of Higgs boson couplings via exclusive production modes and decay channels, of its spin-parity, and of its differential production cross section, need to be thoroughly investigated to verify that they correspond precisely to the SM predictions. It is being done using the new data, collected from 2015 at LHC with increased center-mass-energy of 13 TeV. Precision measurements of the Higgs boson's properties can help understanding why the Higgs boson mass is near the electroweak scale. In particular, it is important to measure the Higgs boson couplings. In the k -framework [15] coupling modifiers are introduced in order to test for deviations in the couplings of the Higgs boson to other particles. Indeed, the couplings to up and down type fermions can be nonuniversal. Additionally, in these models, it is possible



Figure 1.8. (a) Distribution of the four-lepton invariant mass for the $ZZ \rightarrow 4\ell$ analysis of CMS. (b) The observed (solid) p_0 local as a function of m_H in the low mass range in ATLAS detector.

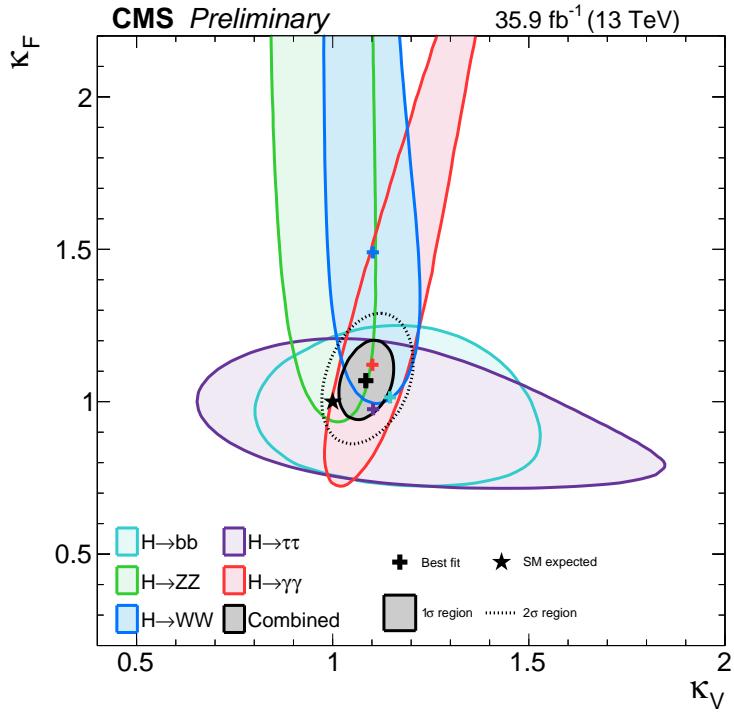


Figure 1.9. The CL regions in the k_F vs k_V parameter space for the model assuming a common scaling of all the vector boson and fermion couplings.

that the Higgs boson will couple differently to fermions (k_F) and vector bosons (k_V), Fig. 1.9. This results so far shows a good agreement among the experimental data and the theory.

1.3 New Scalar Particles

Open Questions

The Standard Model is presently the best description of the subatomic nature, but this theory does not provide a complete picture of the world. Sure enough there are still open questions: the gravity is not described in the model, the explanation of the dark matter is not clear and the neutrino mixing and mass are not well understood. The gravity is one of the four fundamental interactions but it is not described in the SM. It is so different from the three other forces. The purpose to establish a common theory that describe all forces is so difficult. The gravity is described in the Einstein's General Relativity (GR) theory. To combine the SM with the GR it is necessary a quantum theory with a new field associated to gravity as mediator: a spin 2 particle called graviton. Right now there are no experimental evidence of the existence of this kind of particle. An other deficit of the SM regards the dark matter. In fact, from astronomical observations, only the 5% of the matter and energy content of our universe is formed by the ordinary matter (hadrons and leptons), the other 95% is composed by dark matter ($\sim 25\%$) and dark energy ($\sim 70\%$). The SM does not offer good candidates or explanations for the dark matter

and dark energy problems. Concerning the neutrinos, in the SM they were assumed massless. However flavour oscillation implies that they must have non zero mass differences. It is not clear if the small neutrino masses can arise from the same electroweak symmetry breaking mechanism that is in act for the other SM particles.

These open questions represent deficiencies of the Standard Model. The presence of a hidden sector, defined here to mean extra states that have no SM gauge charge but are charged under some other exotic gauge symmetry, does not necessarily solve any of the problems above. However, in order to identify whether the SM Higgs sector is complete, the searches of additional heavy scalars are performed. They would prove the presence of beyond-the-SM (BSM) physics in the form of a non-minimal Higgs sector [16]. The existence of sibling Higgs boson, denoted X , is motivated in many BSM scenarios, so the research in the full mass range accessible at colliders remains one of the main objectives of the experimental community. This program needs to be continued within the full mass range that is accessible to current and future experiments.

Higgs Singlet Extension

The simplest extension of the SM Higgs sector consist in adding to the complex $SU(2)_L$ doublet Φ a real scalar S which is a singlet under all SM gauge groups. The most general gauge-invariant and renormalisable scalar Lagrangian is,

$$\mathcal{L}_s = (D_\mu \Phi)^\dagger D_\mu \Phi + \partial^\mu S \partial_\mu S - V(\Phi, S), \quad (1.13)$$

where $V(\Phi, S)$ is the scalar potential,

$$V(\Phi, S) = -m^2 \Phi^\dagger \Phi - \mu^2 S^2 + \lambda_1 (\Phi^\dagger \Phi)^2 + \lambda_2 S^4 + \lambda_3 \Phi^\dagger \Phi S^2. \quad (1.14)$$

Here, Z_2 ($S \rightarrow -S$) symmetry is imposed which forbids additional terms in the potential. To determine the condition for $V(\Phi, S)$ to be bounded from below, it is necessary to study its behaviour for large field values. The two vacuum expectation values (VEVs) are defined as,

$$\langle \Phi \rangle \equiv \begin{pmatrix} 0 \\ \frac{v}{\sqrt{2}} \end{pmatrix}, \quad \langle S \rangle \equiv \frac{x}{\sqrt{2}}, \quad (1.15)$$

with v and x real and non-negative. With this definition of the VEVs, the extrema of V are determined using the following set of equations:

$$\frac{\partial V}{\partial v}(v, x) = v \cdot (-m^2 + \lambda_1 v^2 + \frac{\lambda_3}{2} x^2) = 0, \quad (1.16)$$

$$\frac{\partial V}{\partial x}(v, x) = x \cdot (-\mu^2 + \lambda_2 x^2 + \frac{\lambda_3}{2} v^2) = 0. \quad (1.17)$$

The physically interesting solutions have $v, x > 0$:

$$v^2 = \frac{\lambda_2 m^2 - \frac{\lambda_3}{2} \mu^2}{\lambda_1 \lambda_2 - \frac{\lambda_3^2}{4}}, \quad (1.18)$$

$$x^2 = \frac{\lambda_1 \mu^2 - \frac{\lambda_3}{2} m^2}{\lambda_1 \lambda_2 - \frac{\lambda_3^2}{4}}. \quad (1.19)$$

For the determination of the extrema we evaluate the Hessian matrix:

$$\mathcal{H}(v, x) \equiv \begin{pmatrix} \frac{\partial^2 V}{\partial v^2} & \frac{\partial^2 V}{\partial v \partial x} \\ \frac{\partial^2 V}{\partial v \partial x} & \frac{\partial^2 V}{\partial x^2} \end{pmatrix} = \begin{pmatrix} 2\lambda_1 v^2 & \lambda_3 vx \\ \lambda_3 vx & 2\lambda_2 x^2 \end{pmatrix}. \quad (1.20)$$

So, the scalar potential $V(\Phi, S)$ is bounded from below if the following conditions are fulfilled,

$$4\lambda_1 \lambda_2 - \lambda_3^2 > 0, \quad (1.21)$$

$$\lambda_1, \lambda_2 > 0, \quad (1.22)$$

where if the first condition is fulfilled, the extremum is a local minimum. The second condition , guarantees that the potential is bounded from below for large field values. The Higgs fields, Φ and S , have non-zero vacuum expectation, denoted by v and x , respectively. Following the unitary-gauge prescription, the Higgs fields is given by,

$$\mathcal{H} \equiv \begin{pmatrix} 0 \\ \frac{h+v}{\sqrt{2}} \end{pmatrix}, \quad S \equiv \frac{h' + x}{\sqrt{2}}. \quad (1.23)$$

Expansion around the minimum leads to the squared mass matrix

$$\mathcal{M}^2 = \begin{pmatrix} 2\lambda_1^2 v^2 & \lambda_3 vx \\ \lambda_3 vx & 2\lambda_1^2 x^2 \end{pmatrix}, \quad (1.24)$$

with the mass eigenvalues

$$m_h^2 = \lambda_1 v^2 + \lambda_2 x^2 - \sqrt{(\lambda_1 v^2 - \lambda_2 x^2)^2 + \lambda_3 (xv)^2}, \quad (1.25)$$

$$m_H^2 = \lambda_1 v^2 + \lambda_2 x^2 + \sqrt{(\lambda_1 v^2 - \lambda_2 x^2)^2 + \lambda_3 (xv)^2}, \quad (1.26)$$

where h and H are the scalar fields of definite masses m_h and m_H respectively, with $m_h^2 < m_H^2$. The gauge and mass eigenstates are related via the mixing matrix

$$\begin{pmatrix} h \\ H \end{pmatrix} = \begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix} \begin{pmatrix} \tilde{h} \\ h' \end{pmatrix}, \quad (1.27)$$

where the mixing angle $-\frac{\pi}{2} \leq \alpha \leq \frac{\pi}{2}$ is given by,

$$\sin 2\alpha = \frac{\lambda_3 xv}{\sqrt{(\lambda_1 v^2 - \lambda_2 x^2)^2 + \lambda_3 (xv)^2}}, \quad (1.28)$$

$$\cos 2\alpha = \frac{\lambda_2 x^2 - \lambda_1 v^2}{\sqrt{(\lambda_1 v^2 - \lambda_2 x^2)^2 + \lambda_3(xv)^2}}. \quad (1.29)$$

By the mixing matrix it is clear that the light (heavy) Higgs couplings to SM particles are now suppressed by $\cos \alpha$ ($\sin \alpha$). The heavy Higgs is a new version of the SM Higgs with rescaled couplings to the matter contents and to the gauge fields of the SM. In fact, the only novel channel with respect to the light Higgs case is $H \rightarrow hh$. The corresponding partial decay width, Γ , is given by [17],

$$\Gamma_{H \rightarrow hh} = \frac{|\mu'|^2}{8\pi m_H} \sqrt{1 - \frac{4m_h^2}{m_H^2}}, \quad (1.30)$$

where the coupling strength μ' is,

$$\mu' = -\frac{\sin 2\alpha}{2vx} (\sin \alpha v + \cos \alpha x) (m_h^2 + \frac{m_H^2}{2}). \quad (1.31)$$

In collider phenomenology, is important:

- the suppression of the production cross section of the two Higgs states induced by the mixing
- the suppression of the Higgs decay modes to SM particles,

For the high mass scenario, i.e. the case where the heavy Higgs boson is identified with the discovered Higgs state at ~ 125 GeV, $|\sin \alpha| = 1$ corresponds to the complete decoupling of the second Higgs boson and therefore the SM-like scenario.

Minimal Supersymmetric Standard Model

With the a Higgs boson discovery, possible extensions of the Standard Model has been considered. One of the simplest ways to extend the scalar sector of the SM is to add one more complex doublet to the model. The resulting two Higgs doublet models (2HDMs) [18] can provide additional CP-violation coming from the scalar sector and can easily originate dark matter candidates. Also more, models with additional field content can have a 2HDM scalar sector, like the the Minimal Supersymmetric Standard Model. The 2HDMs have a richer Higgs particle spectrum with one charged and three neutral scalars. All neutral scalars could in principle be the scalar discovered at the LHC.

Two versions of the 2HDM, one CP-conserving and the other explicitly CP-violating, exist. From the Yukawa Lagrangian arise scalar scalar exchange flavour changing neutral currents (FCNCs) that are constrained by experiments results. A simple way to take in account the FCNCs constrains is to impose a discrete Z_2 symmetry¹ [19] on the scalar doublets ($\Phi_1 \rightarrow \Phi_1$ and $\Phi_2 \rightarrow -\Phi_2$) and a corresponding symmetry to

¹A Z_n group describes a symmetry of a plane figure invariant after a rotation of $2\pi/n$ degrees.

	Φ_1	Φ_2	u_R	d_R	ℓ_R	Q_L, L_L
Type-I	+	-	-	-	-	+
Type-II	+	-	-	+	+	+
Type-X	+	-	-	-	+	+
Type-Y	+	-	-	+	-	+

Figure 1.10. Variation in charge assignments of the Z_2 symmetry.

the quark field.

These arguments lead four Yukawa models: types-I, types-II, types-Y (flipped) and types-X (Lepton Specific) [20, 21]. For example, the Higgs sector of the minimal super-symmetric standard model (MSSM) is the 2HDM with a supersymmetric relation among the parameters of the Higgs sector, whose Yukawa interaction is of type-II, in which only a Higgs doublet couples to up-type quarks and the other couples to down-type quarks and charged leptons. On the other hand, a model that tries to explain neutrino masses and dark matter corresponds to a type-X Yukawa interaction. In this model, the Higgs sector is the two Higgs doublets with extra scalar singlets, in which only a Higgs doublet couples to quarks, and the other couples to leptons. Therefore it is important to experimentally determine the type of the Yukawa interaction in order to select the true model from the different 2HDMs. The most general Yukawa interaction under the Z_2 symmetry can be written as,

$$\mathcal{L}_{Yukawa}^{2HDM} = -\bar{Q}_L Y_u \tilde{\Phi}_u u_R - \bar{Q}_L Y_d \Phi_d d_R - \bar{L}_L Y_l \Phi_l l_R + h.c., \quad (1.32)$$

where Φ_f , with $f = u, d, l$, is either Φ_1 or Φ_2 . There are four independent Z_2 charge assignments on quarks and charged leptons, Fig 1.10. In the type-I, all quarks and charged leptons obtain their masses from the VEV of Φ_2 . In the type-II, masses of up-type quarks are generated by the VEV of Φ_2 , while those of down-type quarks and charged leptons are acquired by that of Φ_1 . The Higgs sector of the MSSM is a special 2HDM whose Yukawa interaction is of type-II. The type-X Yukawa interaction (all quarks couple to Φ_2 while charged leptons couple to Φ_1) is used in the model described in [21]. The remaining one is referred to as the type-Y. The most general Higgs potential under the broken Z_2 symmetry is given by,

$$\begin{aligned} V^{2HDM} = & m_1^2 \Phi_1^\dagger \Phi_1 + m_2^2 \Phi_2^\dagger \Phi_2 - m_3^2 (\Phi_1^\dagger \Phi_1 + \Phi_2^\dagger \Phi_2) + \frac{\lambda_1}{2} (\Phi_1^\dagger \Phi_1)^2 \\ & + \frac{\lambda_2}{2} (\Phi_2^\dagger \Phi_2)^2 + \lambda_3 (\Phi_1^\dagger \Phi_1 + \Phi_2^\dagger \Phi_2) + \lambda_4 (\Phi_1^\dagger \Phi_2 + \Phi_2^\dagger \Phi_1) \\ & + \lambda_5 [(\Phi_1^\dagger \Phi_2)^2 + (\Phi_2^\dagger \Phi_1)^2], \end{aligned} \quad (1.33)$$

where the parameters m_3^2 and λ_5 are complex, in general but in the following , assuming CP invariance, they are real. The Higgs doublet fields can be parametrized as,

$$\Phi_i = \begin{pmatrix} \omega_i^+ \\ \frac{1}{\sqrt{2}}(v_i + h_i + iz_i) \end{pmatrix}, \quad (1.34)$$

and the mass eigenstates are defined by,

$$\begin{pmatrix} h_1 \\ h_2 \end{pmatrix} = R(\alpha) \begin{pmatrix} H \\ h \end{pmatrix},$$

$$\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = R(\beta) \begin{pmatrix} z \\ A \end{pmatrix}, \quad (1.35)$$

$$\begin{pmatrix} \omega_1^+ \\ \omega_2^+ \end{pmatrix} = R(\beta) \begin{pmatrix} \omega^+ \\ H^+ \end{pmatrix},$$

where the rotation matrix is given by,

$$R(\theta) = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}. \quad (1.36)$$

There are five physical Higgs bosons, i.e., two CP-even states h and H , one CP-odd state A , and a pair of charged states H^\pm , while z and ω^\pm are Nambu-Goldstone bosons that are eaten as the longitudinal components of the massive gauge bosons. The eight parameters, m_i^2 ($i = 1, 3$) and λ_j ($j = 1, 5$) in the Higgs sector are replaced by eight physical parameters:

- the VEV $v = \sqrt{v_1^2 + v_2^2}$,
- the mixing angles α and β ($\tan \beta = v_1/V_2$) with $\sin(\beta - \alpha) = 1$,
- the boson masses m_h, m_H, m_A, m_H^\pm ,
- the soft breaking mass parameter, $M = m_3/\sqrt{\sin \beta \cos \beta}$.

For the successful electroweak symmetry breaking, a combination of quartic coupling constants should satisfy the condition of vacuum stability. In addition one has also to take into account bounds from perturbative unitarity to restrict parameters in the Higgs potential. The top and bottom Yukawa coupling constants cannot be taken to large. The requirement $|Y_{t,b} < \pi|$ at the tree level can provide a milder constraint $0.4 < \tan \beta < 91$.

The wide range of possibilities for Higgs boson mass spectrum hierarchies and branching ratios in 2HDMs yields a diversity of production and decay channels that are relevant for multi-lepton signatures at the LHC. Multi-lepton final states become especially important when the decay of one Higgs scalar to a pair of Higgs scalars or a Higgs scalar and a vector boson is possible.



Figure 1.11. Higgs production cross sections at the LHC for various production mechanisms as a function of the Higgs mass. QCD corrections are included except for Higgs bremsstrahlung [22].

1.4 The WW decay channel for high mass Higgs-like particles

Let's now concentrate on the $X \rightarrow WW$ channel, that is the final state used in the following for the high mass searches. The two dominant production mechanisms of high mass SM-like Higgs boson are the gluon-gluon fusion and Vector Boson Fusion, Fig. 1.11. The first one is the main mechanism for mass values below 1 TeV, above the VBF production mechanism becomes more and more important as m_X increases. In the search for high mass Higgs boson, in many models, the WW final state, along with ZZ , is the dominant decay channel of X for masses above $2m_Z$ threshold. This fact is evident in Fig. 1.12, where the WW branching ratio (in green) dominates in the high mass region. More detailed results on the decays $H \rightarrow WW$ and $H \rightarrow ZZ$ with the subsequent decay chain are presented in Fig. 1.13 and Fig. 1.14. However even if the yield for the decay channels started by the $X \rightarrow WW$ decay is higher, the presence of neutrinos in the final state does not allow to have a complete reconstruction of the decay. This fact makes this channel very challenging. As we have already noticed, gluon-gluon fusion cross section is between one and two orders of magnitude larger than that of VBF for a wide range of Higgs masses. Nevertheless, the VBF becomes competitive when the mass approaches to 1 TeV. Furthermore, in case of gluon-gluon fusion production mechanism, in addition to the two leptons and two neutrinos, there are rarely one or more jets coming from initial state radiation, Sec. 3.3. The VBF production, providing two more jets (the VBF jets coming from the hadronization quarks from production) to the final state,



Figure 1.12. Higgs branching ratios and their uncertainties for the full mass range [23].

benefits from a highly reduced background with respect to the gluon-gluon fusion production mode, such that even if the VBF production mechanism has a branching ratio smaller than the gluon-gluon fusion, a higher signal-to-noise ratio is expected.

$X \rightarrow WW$ searches at colliders

The search of high mass particle with WW final state has been widely performed at experiments at hadron colliders to search for new particles beyond the SM. The resonant WW production has been studied at both the Fermilab Tevatron Collider, and the CERN Large Hadron Collider, with the progressively increasing collision energy and integrated luminosity. Each machine in its time has therefore probed the highest masses of resonances accessible. A review of the different searches performed by D0, CDF, ATLAS and CMS, their techniques, data, results, and limits on new particles decaying to WW are described. After the discovery of the Higgs boson at LHC, the two experiments, ATLAS and CMS, have been focused on high mass searches using the data collected at $\sqrt{s} = 7, 8$ (Run-I) and 13 TeV (Run-II). A search for a heavy Higgs boson in the $H \rightarrow WW$ and $H \rightarrow ZZ$ decay channels has been performed by the CMS experiment based on proton-proton collision data samples corresponding to an integrated luminosity of up to 5.1 fb^{-1} at $\sqrt{s} = 7$ TeV and up to 19.7 fb^{-1} at $\sqrt{s} = 8$ [24]. This analysis is performed in a mass range $145 < m_H < 1000$ GeV and searches for a heavy Higgs boson in the EW singlet extension of the SM. The peculiarity of this analysis is that the full Run-I statistics is considered and the mass range reach 1 TeV for the first time in CMS. In the case of a high Higgs boson decaying into a pair of W bosons, the fully leptonic ($X \rightarrow WW \rightarrow 2\ell 2\nu$) and semileptonic ($X \rightarrow WW \rightarrow \ell\nu qq$) final states are considered in this analysis. Decays containing four charged leptons ($X \rightarrow WW \rightarrow 2\ell 2\ell'$), two charged leptons



Figure 1.13. Higgs branching ratios for the different $H \rightarrow 4\ell$ and $H \rightarrow 2\ell 2\nu$ final states and their uncertainties for the full mass range [23].

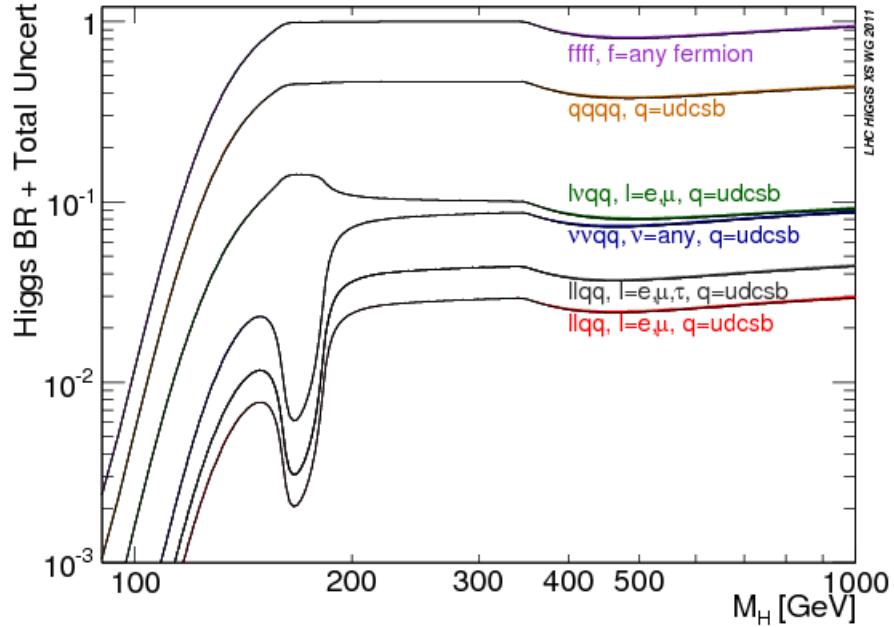


Figure 1.14. Higgs branching ratios for different final states and their uncertainties for the full mass range [23].

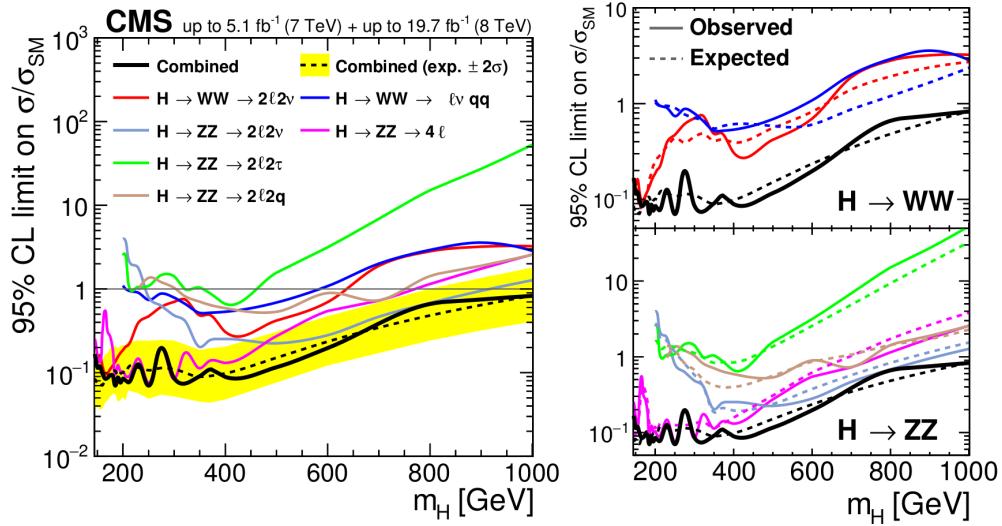


Figure 1.15. Upper limits at the 95% CL for each of the contributing final states and their combination. [24].

and two quarks ($X \rightarrow ZZ \rightarrow \ell\nu qq$) and two charged leptons and two neutrinos ($X \rightarrow ZZ \rightarrow 2\ell 2\nu$) are considered. No significant excess over the expected SM background has been observed and exclusion limits have been set. The combined results obtained for a heavy Higgs boson with SM-like couplings for all the different contributing final states are displayed in Fig. 1.15. On the left, the observed 95% CL limit is shown for each final state. On the right, the expected and observed limits are displayed for each of the individual channels as well as the combined result. Using the 13 TeV proton-proton collision data produced at the LHC in 2015, corresponding to an integrated luminosity of 2.3 fb^{-1} , CMS has performed the search for a final state with different flavour leptons ($X \rightarrow WW \rightarrow 2\ell 2\nu$) in the mass range $200 < m_H < 1000 \text{ GeV}$ [25]. The search has been carried out in the 0-jets, 1-jet and VBF categories in order to increase the signal sensitivity to different production mechanisms and maximize the exclusion limits. No significant excess with respect to the SM background expectation has been observed, and the exclusion limits on the cross section times branching ratio to $WW \rightarrow 2\ell 2\nu$ ave been reported, Fig. 1.16. The ATLAS collaboration published the results of a search for a heavy neutral scalar decaying to two W bosons using the datasets collected in 2015 and early 2016 at a centre-of-mass energy $\sqrt{s} = 13 \text{ TeV}$ corresponding to an integrated luminosity of 13.2 fb^{-1} in the mass range between 300 GeV and 3 TeV [26]. In this analysis, categories with one- and at least two-jets are optimised for a vector boson fusion-like signal and the remaining category is quasi-inclusive for a gluon-gluon fusion-like signal. The search sensitivity depends on the assumed Higgs boson width. Two different hypotheses are tested: a narrow width approximation, where the width of the heavy Higgs boson is smaller than the experimental resolution, and a large width assumption, where widths of 5%, 10%, and 15% of the heavy Higgs boson mass are considered. Upper limits are set on the production cross section times the $H \rightarrow WW$ branching ratio in two scenarios: a high-mass Higgs boson with a

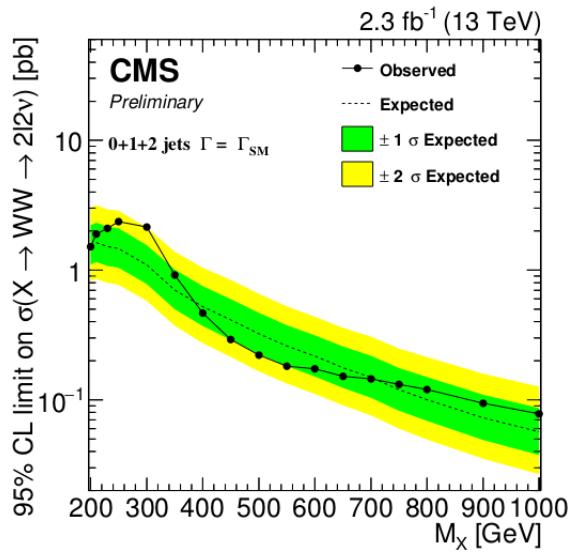


Figure 1.16. Expected and observed exclusion limits at 95% CL on the sum of gluon-gluon fusion and VBF cross sections times branching fraction for the combination of the three jet categories as a function of the resonance mass. The black dotted line corresponds to the central value while the yellow and green bands represent the $\pm 1\sigma$ and $\pm 2\sigma$ uncertainties respectively.

narrow width, and with intermediate widths (of 5, 10, 15% of the heavy Higgs boson mass), Fig. 1.17. Values above 4.3 pb (1.4 pb) at $m_H = 300$ GeV (400 GeV) and above 0.051 pb (0.071 pb) at 3 TeV are excluded at 95% CL by the gluon-gluon fusion quasi-inclusive NWA (LWA 15%) analysis. For the VBF NWA case, the upper exclusion limit ranges between 1.1 pb at $m_H = 300$ GeV to 0.03 pb at 3 TeV.

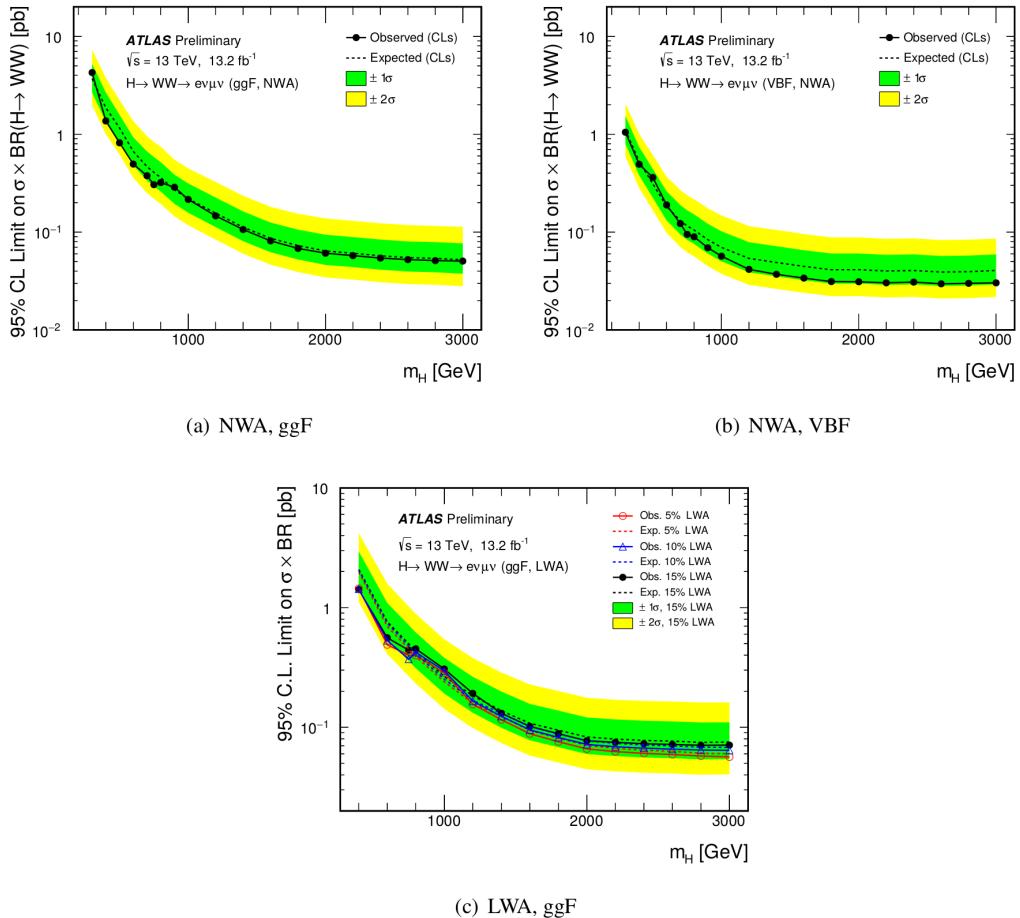


Figure 1.17. 95% CL upper limits on the Higgs production cross section times branching ratio in the analysis, for signals with narrow-width (gluon-gluon fusion or VBF) in the top row and the 5%, 10% and 15% width lineshapes (gluon-gluon fussion only) in the bottom. The green and yellow bands show the $\pm 1\sigma$ and $\pm 2\sigma$ uncertainties on the expected limit.

Chapter 2

The CMS experiment at LHC

The data used in this analysis has been collected by the Compact Muon detector. It is one of the main experiments located at the Large Hadron Collider at CERN. This chapter provides a general description of the accelerator system and of the detector involved in producing and recording the data. In particular I have been involved in the CMS data taking and in data validation operation since 2016, Fig. 2.10.

2.1 The Large Hadron Collider

The Large Hadron Collider (LHC) [27] at CERN, which started operations in 2008, is the largest and most powerful hadron collider ever built. Installed in the underground tunnel which housed the Large Electron Positron Collider (LEP), in operation until 2000, the LHC accelerator has approximately the shape of a circle with a length of about 27 km and is located underground at a depth varying between 50 m to 175 m, straddling the Franco-Swiss border near Geneva. It is designed to collide two 7 TeV counter-circulating beams of protons resulting in a center-of-mass energy of 14 TeV, or two beams of heavy ions, in particular lead nuclei at an energy of 2.76 TeV/nucleon in the center-of-mass frame. The transition from a leptonic collider to a hadronic collider entailed the following advantages: first, it has been possible to build a machine that having the same size of the previous one (and therefore accommodated in the same tunnel, substantially reducing the cost and time of construction), could reach a higher energy in the center-of-mass frame. This is due to the much lower amount of energy loss through synchrotron radiation emitted by the accelerated particles, that is proportional to the fourth power of the ratio E/m between their energy and their mass. Secondly, the composite structure of protons compared to the elementary structure of electrons allows LHC to be able to access simultaneously a wider energy spectrum, despite the production of many low energies particles in a complex environment. This feature is particularly important for a machine dedicated to the discovery of “new” physics.

A schematic description of the accelerator complex installed at CERN is shown in Fig. 2.1. The acceleration is performed in several stages. The protons source is a Duoplasmatron: the protons are obtained by removing electrons from a source of hydrogen gas and then sent to the LINAC2, a 36 m long linear accelerator which generates a pulsed beam with an energy of 50 MeV using Radio Frequency

Quadrupoles (RFQ) and focusing quadrupole magnets. The beam is subsequently sent to the Proton Synchrotron Booster (PSB), a circular accelerator consisting of four superimposed synchrotron rings with a circumference of about 160 m, which increases the proton energy up to 1.4 GeV. Then, protons are injected into the Proton Synchrotron (PS), a single synchrotron ring with a circumference of about 600 m where the energy is increased to 25 GeV. The sequential combination of these two synchrotrons also allows to create a series of protons bunches interspersed by 25 ns (i.e. at the frequency of 40 MHz) as required for the final correct operation of LHC. The final proton injection stage is the Super Proton Synchrotron (SPS), a synchrotron with a circumference of approximately 7 km where protons reach an energy value of 450 GeV. Subsequently, protons are extracted and injected into the LHC ring via two transmission lines, to generate two beams running in opposite directions in two parallel pipes and which are accelerated up to the energy of interest.

The beams collide at four interaction points where the four main experiments are located: ALICE, ATLAS, CMS and LHCb. Two small experiments, TOTEM and LHCf, which focus on forward particles, have also been installed near, respectively, CMS and ATLAS. The 7 TeV per-beam-energy limit on the LHC is not determined by the electric field generated by the radiofrequency cavity but by the magnetic field necessary to maintain the protons in orbit, about 8.3 T, given the current technology for the superconducting magnets.

One of the most important parameter of an accelerator is the *instantaneous luminosity* \mathcal{L} . For a given process having a cross section σ , the instantaneous luminosity \mathcal{L} is defined by the relation

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

where N is the rate of occurrence of the process. Then the *integrated luminosity* L is defined as the integral of the instantaneous luminosity over time, i.e.

$$L = \int \mathcal{L} dt. \quad (2.2)$$

2.2 The Compact Muon Solenoid experiment

The Compact Muon Solenoid (CMS) is a general purpose detector optimized for the analysis of the proton-proton interactions with the expected energy and luminosity of the LHC. CMS is able to accurately identify muons, electrons, photons and hadrons. It has been designed to investigate a wide range of physics, with the search for the Higgs boson as one of the main highlights. Search for new physics is also an important goal of the experiment, as well as top physics and, of course, Standard Model precision measurements. The collaboration that run the CMS experiment is one of the largest international scientific community in history, involving more than 4000 people (particle physicists, engineers, technicians, students and support staff) from about 180 universities and research institutes in more than 40 countries. The experimental apparatus is placed in a cavern 100 m underground in the area called Point 5 (an old LEP access point) near the village of Cessy, in France.

The coordinate system used in CMS is a right-handed Cartesian system, having the origin in the nominal beam collision point inside the detector. The x-axis points



Figure 2.1. Schematic layout of the accelerator complex installed at CERN.



Figure 2.2. An exploded view of the CMS detector with all subdetectors.

radially towards the center of the LHC ring, the y -axis is directed upwards along the vertical and the z -axis is oriented along the direction of the beams, along the counter-clockwise direction of the ring if seen from above. The cylindrical symmetry of CMS design and the invariant description of proton-proton physics suggest to define also a coordinate system based on pseudo-angular coordinates, given by the triplet (r, ϕ, η) where r is the distance from the z -axis, ϕ is the azimuthal angle measured on the $x - y$ plane starting from the x -axis and η is the pseudorapidity (please refer to Appendix A for more details).

The CMS detector, shown in Fig. 2.2, is 21.6 m long, has a diameter of 15 m and a weight of about 12,500 tons. The constructive element that characterizes the experiment is a solenoidal superconducting magnet which produces an internal constant magnetic field of 3.8 T along the direction of the beams. The CMS detector is designed as a dodecagonal base prism. The central part of the prism, named barrel, contains several layers of detectors with quasi-cylindrical symmetry, coaxial with respect to the direction of the beams. A set of detector disks, called endcaps, close the detector at its ends, to ensure hermeticity.

From the inner region to the outer one, the various components of CMS are:

- Silicon Tracker: it is placed in the region $r < 1.2$ m and $|\eta| < 2.5$. It consists of a silicon pixel vertex detector and a surrounding silicon microstrip detector, with a total active area of about 215 m^2 . It is used to reconstruct charged particle tracks and vertices;
- Electromagnetic Calorimeter (ECAL): it is placed in the region $1.2\text{ m} < r <$

1.8 m and $|\eta| < 3$. It consists of scintillating crystals of lead tungstate and it is used to measure the direction and energy of photons and electrons;

- Hadron Calorimeter (HCAL): it is placed in the region $1.8 \text{ m} < r < 2.9 \text{ m}$ and $|\eta| < 5$. It consists of brass layers alternated with plastic scintillators and it is used to measure the direction and the energy released by the hadrons produced in the interactions;
- Superconducting Solenoidal Magnet: it is placed in the region $2.9 \text{ m} < r < 3.8 \text{ m}$ and $|\eta| < 1.5$. It generates an internal uniform magnetic field of 3.8 T along the direction of the beams, necessary to deflect the charged particles in order to allow a measurement of their momentum through the curvature observed in the tracking system. The magnetic field is closed with an iron yoke 21.6 m long with a diameter of 14 m, where a residual magnetic field of 1.8 T is present, opposite in direction with respect to the 3.8 T field;
- Muon System: it is placed in the region $4 \text{ m} < r < 7.4 \text{ m}$ and $|\eta| < 2.4$. It consists of Drift Tubes (DT) in the barrel region and Cathode Strip Chambers (CSC) in the endcaps. A complementary system of Resistive Plate Chambers (RPC) is used both in the barrel and in the endcaps. This composite tracking system for muons is used to reconstruct tracks released by muons that pass through it. The muons chambers are housed inside the iron structure of the return yoke that encloses the magnetic field.

The Tracker

The silicon tracker is the detector closest to the beams collision point. Its goal is the high resolution reconstruction of the trajectories of charged particles originating from the collision region and the identification of the position of secondary vertices produced by particles with a short mean life time (in particular hadrons containing the b quark, that decay after few hundreds of μm). The events produced in the proton- proton collisions can be very complex and track reconstruction is a challenging pattern recognition problem. Indeed, at the nominal instantaneous luminosity of operation, an average of about 20 pile-up events overlapping to the event of interest are expected, leading to about 1000 tracks to be reconstructed per event.

In order to ease the pattern recognition, two requirements are fundamental: a low occupancy detector and a large redundancy of the measured points (hits) per track. The first requirement is achieved building a detector with high granularity. The redundancy of the hits is instead achieved having several detecting layers, and is necessary to reduce the ambiguity on the assignment of the hits to a given track. Nevertheless, the amount of tracker material has to be as low as possible, in order to avoid jeopardizing the measurement of the particle trajectory. An excessive amount of material would indeed deteriorate the measurement, mainly because of multiple scattering and energy loss for charged particles, bremsstrahlung for electrons and nuclear interactions for hadrons. The outer detectors such as ECAL are influenced by the material as well, for example because of the increased probability for a photon to convert to an electron-positron pair in the tracker material.

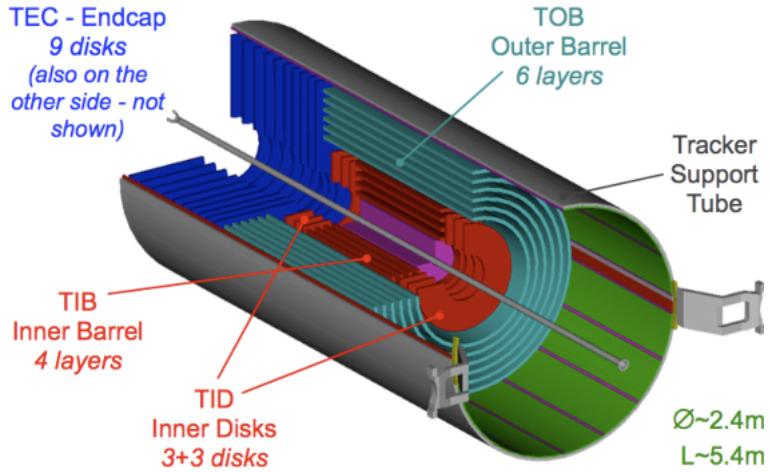


Figure 2.3. Three-dimensional schematic view of the CMS silicon tracker.

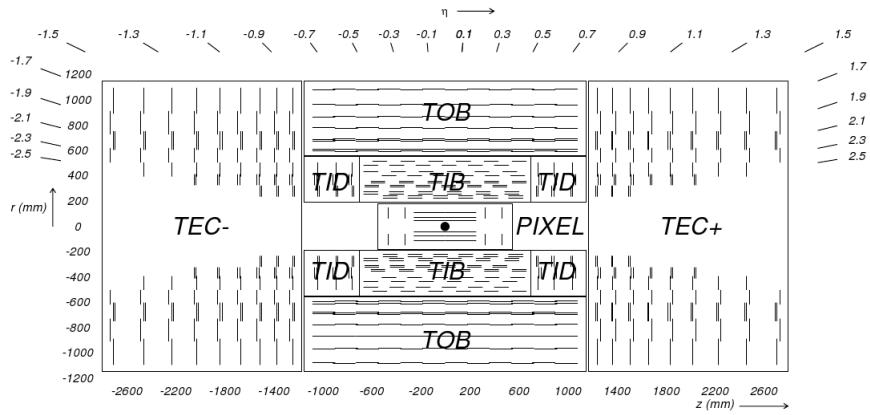


Figure 2.4. Pictorial view of a tracker slice in the r-z plane. Pixel modules are shown in red, single-sided strip modules are depicted as black thin lines and strip stereo modules are shown as blue thick lines.

For these reasons, the tracker layers are limited in number and thickness. The tracker comprises a large silicon strip detector with a small silicon pixel detector inside it. In the central η region, the pixel tracker consists of three co-axial barrel layers at radii between 4.4 cm and 10.2 cm and the strip tracker consists of ten co-axial barrel layers extending outwards to a radius of 110 cm. Both subdetectors are completed by endcaps on either side of the barrel, each consisting of two disks in the pixel tracker, and three small plus nine large disks in the strip tracker. The endcaps extend the acceptance of the tracker up to $|\eta| < 2.5$. A three-dimensional schematic view of the tracker is shown in Fig. 2.3, while in Fig. 2.4 a pictorial representation of a slice of the tracker is displayed, showing the various layers of the subdetectors. The whole tracker has a cylindrical shape with a length of 5.8 m and a diameter of 2.5 m, with the axis aligned to the beams direction. The number of hits per track is 12-14, allowing high reconstruction efficiency and low rate of fake tracks.

The Pixel Vertex Detector The pixel vertex detector is mainly used in CMS as a starting point for the reconstruction of tracks and is essential for the reconstruction of the primary vertex (PV) and any possible secondary vertices. It is placed in the region closest to the collision point, where the particle flux is maximum. It covers the region $|\eta| < 2.5$ and is composed of a central part (barrel) and by two forward parts (endcaps).

The barrel consists of three concentric cylindrical sectors 53 cm long, located at an average distance r of 4.4 cm, 7.3 cm and 10.2 cm. Each half-cylinder is made up of ladders and half ladders that serve as support and cooling structure for the pixel modules, with each ladder containing 8 modules. In total, the barrel is composed of 768 modules.

Each endcap is composed of two disks placed at a distance of 34.5 cm and 46.5 cm from the nominal beams impact point. They cover a radius r in a range between 6 cm and 15 cm in such a way that each track within the detector acceptance passes through at least two layers. Each disk is divided into 24 segments, on each of which 7 modules of different sizes are mounted, for a total of 672 modules on all the endcaps.

A module is composed by a silicon sensor with a thickness of 250 μm and the corresponding readout chips. In order to optimize the reconstruction of track and vertices near the interaction point, the sensor is segmented into rectangular pixels with a size of $150 \times 100 \mu\text{m}^2$, with the $100 \mu\text{m}$ side oriented along the $r\phi$ direction in the barrel region and along the r direction in the endcap region. The hit position resolution is about 10-15 μm in the barrel and about 15 μm in the endcaps.

The Silicon Microstrip Detector In the region of the detector that is more than 20 cm far from the beam, the flux of charged particles is sufficiently limited to allow the use of a silicon microstrip detector. Overall, this detector (Silicon Strip Tracker, SST) consists of 15400 elementary units or modules, composed by one or two sensors stuck onto a support of carbon fiber together with the readout electronics. In some of the layers and in the innermost rings, special double-sided modules are able to provide accurate three-dimensional position measurement of the charged particle hits by having the two silicon sensors rotated in order to have strips forming an angle of 100 mrad. This “stereo” combination, although of lower resolution, is preferable compared to a pixel segmentation because it requires a lower number of readout channels.

The silicon microstrip tracker is 5.4 m long, extending up to a distance of 1.1 m from the axis of the beams. It consists of a barrel and two endcaps and it is divided into four distinct parts, TIB and TOB, and TID and TEC.

The Electromagnetic Calorimeter (ECAL)

The main function of an electromagnetic calorimeter is to identify electrons and photons measuring accurately their direction and energy. The electromagnetic calorimeter (Fig. 2.5) of CMS (ECAL, Electromagnetic CALorimeter) is a homogeneous calorimeter with cylindrical geometry, whose active elements are scintillating crystals of lead tungstate (PbWO_4) with a truncated pyramidal shape. It consists of an ECAL Barrel (EB) with 61200 crystals and two ECAL Endcaps (EE) containing 7324 crystals each. Crystals are grouped into 5×5 matrices called towers.



Figure 2.5. Schematic representation of the electromagnetic calorimeter ECAL.

The barrel has an inner radius of 129 cm, a length of 630 cm and it extends in the region $|\eta| < 1.479$. Crystals in the ECAL barrel have the following dimensions: $22 \times 22 \text{ mm}^2$ at the front face, $26 \times 26 \text{ mm}^2$ at the rear face, and a length of 23 cm, corresponding to $25.8 X_0$. Each submodule, consisting in a 5×2 crystals arrays mounted on a glass fiber structure, forms the elementary unit of EB. To avoid that cracks might align with particle trajectories, the crystal axes are tilted by 3 degrees with respect to the direction from the interaction point, both in the η and ϕ direction.

Each endcap covers the region $1.479 < |\eta| < 3$ and is formed by two semicircular halves of aluminum called dees. Crystals in endcaps have a length of 22 cm and frontal area equal to $28.6 \times 28.6 \text{ mm}^2$. They are arranged in supercrystals with 5×5 elementary unity. Unlike the crystals in the barrel, arranged in a $\eta - \phi$ geometry, the endcap crystals are arranged according to a xy geometry.

Two preshower detectors are placed in front of the endcaps in order to disentangle the showers produced by a primary γ from those produced by a primary π_0 . This detector, which covers the region $1.653 < |\eta| < 2.6$, is a sampling calorimeter and it consists of two disks of lead converters that start the electromagnetic shower of the incident photon/electron, alternating with two layers of silicon microstrip detectors in which a measurement of the released energy and the identification of the shower profile are performed. The strips are arranged orthogonally in the two planes, according to a xy configuration.

The choice of the PbWO_4 crystals as scintillating material for ECAL is due to several reasons. First, the high-density, the short radiation length and the reduced Molière radius ($R_M = 2.2 \text{ cm}$) allow to build a compact and high granularity calorimeter. Furthermore, the 15 ns scintillation decay time allows to collect about 80% of the emitted light during the 25 ns that exist between two consecutive bunch crossings in the LHC. Finally, the PbWO_4 crystals have a good intrinsic radiation hardness and therefore they can operate for years in the hostile LHC environment, with a modest deterioration in performance. The main disadvantage of these crystals is the low light yield which makes an internal amplification for the photodetectors necessary. This is achieved through the use of silicon avalanche photodiodes in barrel and single stage photomultipliers in the endcaps, both resistant to the radiation and

to the strong magnetic field of CMS.

The energy resolution of a homogeneous calorimeter is usually expressed by the sum in quadrature of three terms, according to the formula,

$$\frac{\sigma_e}{E} = \frac{a}{\sqrt{E}} \oplus \frac{b}{E} \oplus c, \quad (2.3)$$

The stochastic term a is dominant at low energies: it includes the contribution of statistical fluctuations in the number of photoelectrons generated and collected. The noise term b includes contributions from the electronic noise, both due to the photodetector and to the preamplifier, and from pileup events. The constant term c , dominant at high energies, takes into account several contributions: the stability of the operating conditions (in particular of temperature and voltage), the presence of dead material in front of the crystals and the rear leakage of the electromagnetic shower, the longitudinal non uniformity of the crystal light yield, the intercalibration errors and the radiation damage of the crystals. The ECAL barrel energy resolution for electrons in beam tests has been measured to: $a = 2.8\% \text{ GeV}^{-1/2}$, $b = 12\% \text{ GeV}$, $c \approx 0.3\%$, where the energy E is measured in GeV.

The Hadronic Calorimeter (HCAL)

The hadronic calorimeter HCAL (Hadronic CALorimeter) complements the electromagnetic calorimeter to build up a complete calorimetric system for the jet energy and direction measurement. Furthermore, thanks to its hermeticity, it can provide a measurement of the properties of non-interacting particles, such as neutrinos, by measuring the missing energy deposited in the transverse plane, E_T^{Miss} or MET.

The CMS hadronic calorimeter is a hermetic sampling calorimeter that covers the region $|\eta| < 5$. As shown in Fig. 2.6, it is divided into four subdetectors: HB (Barrel Hadronic Calorimeter), located in the barrel region inside the magnet, extending up to pseudorapidities $|\eta| \sim 1.4$; HE (Endcap Hadronic Calorimeter), situated in the endcaps region inside the magnet, extends in the pseudorapidity region $1.3 < |\eta| < 3$, partially overlapping the HB coverage; HO (Outer Hadronic Calorimeter, also called tail-catcher, placed along the inner wall of the magnetic field return yoke, just outside of the magnet; HF (Forward Hadronic Calorimeter), a sampling calorimeter made of quartz fibers sandwiched between iron absorbers, consisting of two units placed in the very forward region ($3 < |\eta| < 5$) outside the magnetic coil. At the passage of charged particles, Cherenkov light is emitted in the quartz fibers and this light is detected by radiation resistant photomultipliers.

In order to maximize particle containment for a precise missing transverse energy measurement, the amount of absorber material was maximized, reducing therefore the amount of the active material. Since HCAL is mostly placed inside the magnetic coil, a non-magnetic material like brass was chosen as absorber. HB and HE are therefore made with brass absorber layers interleaved with plastic scintillators coupled to wavelength shifting optical fibers which transmit the light to the HPD (Hybrid Photodiodes) photodetectors.



Figure 2.6. A schematic rz view of a quadrant of the CMS hadronic calorimeter HCAL.

The Solenoidal Magnet

The CMS magnet (shown in Fig. 2.7) houses the tracker, the electromagnetic and the hadronic calorimeters and is the biggest superconducting solenoid ever built in the world. The solenoid achieves a magnetic field of 3.8 T in the free bore of 6 m in diameter and 12.5 m in length. The energy stored in the magnet is about 2.6 GJ at full current. The superconductor is made of four Niobium-Titanium layers. In case of a quench, when the magnet loses its superconducting property, the energy is dumped to resistors within 200 ms. The magnet return yoke of the barrel is composed by three sections along the z -axis; each one is split into 4 layers (hosting the muon chambers in the gaps). Most of the iron volume is saturated or nearly saturated, and the field in the yoke is about half (1.8 T) the field in the central volume.

The Muon chamber

The CMS Muon System is dedicated to the identification of muons and measurement of their transverse momentum, p_T , in combination with the tracker. Furthermore, it provides a time measurement of the bunch-crossing and also works as trigger for events involving muons. Momentum measurement, in the muon system, is determined by the muon bending angle at the exit of the 3.8 T coil, considering the interaction point as the origin of the muon. Up to p_T values of 200 GeV, the resolution of the muon system is dominated by multiple scattering and the best resolution is rather given by the silicon tracker, as shown in Fig 2.8. The system is placed outside the magnetic coil, embedded in the return yoke, to fully exploit the 1.8 T return flux. It consists of three independent subsystems, as shown in Fig. 2.9: drift tubes (DT), cathode strip chambers (CSC) and resistive plate chambers (RPC). The DT and the CSC provide an excellent spatial resolution for the measurement of charged particle momentum; the RPC are used for trigger issues because of the very good timing. The active parts of the muon system are hosted into stations which

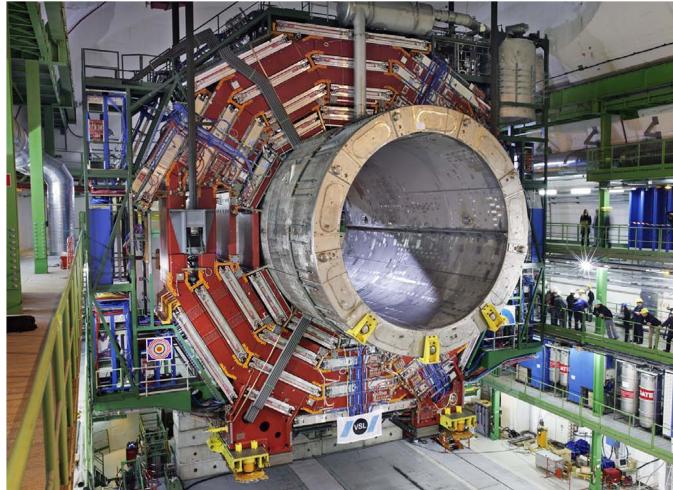


Figure 2.7. Arrival of the magnet in the tunnel on February 28, 2007.

are interleaved by the iron layers of the return yoke of the magnet.

Trigger and Data Acquisition

LHC can produce interactions at 40 MHz frequency, but only a small fraction of these events can be written on disk. On one hand the speed at which data can be written to mass storage is limited, on the other hand the vast majority of events produced is not interesting, because it involves low transverse momentum interactions (minimum bias events). Thus, a trigger system is needed to select interesting events at the highest possible rate. The maximum rate of events written on disk is about 800 Hz. CMS has chosen a two-level trigger system, consisting of a Level-1 Trigger (L1) and a High Level Trigger (HLT). Level-1 trigger runs on dedicated processors, and accesses coarse level granularity information from calorimetry and muon system. A L1 Trigger decision has to be taken for each bunch crossing within $3.2 \mu\text{s}$. Its task is to reduce the data flow from 40 MHz to about 100 kHz. The High Level Trigger is responsible for reducing the L1 output rate down to a maximum rate of the order of 1 kHz. The HLT code runs on a farm of commercial processors and can access the full granularity information of all the subdetectors.

2.3 Data samples and future plans

The first proton beam circulated in the LHC on September 2008, after more than a decade of construction and installation. An incident occurred in two magnets, causing the release of helium into the tunnel and mechanical damage. After that, in March 2010 started the Run-I, a fruitful data taking era that lasted until 2012. It was decided not to operate the LHC at its design parameters, and proton-proton collisions took place at a centre-of-mass energy of 7 TeV and 8 TeV. The amount of recorded data (in CMS) in this period is reported in Fig. 2.10. At the end of 2012, LHC operations halted for two years due to the first long shutdown (LS1). In 2015, with centre-of-mass energy of 13 TeV, the proton-proton collision restarted (Run-II) and,

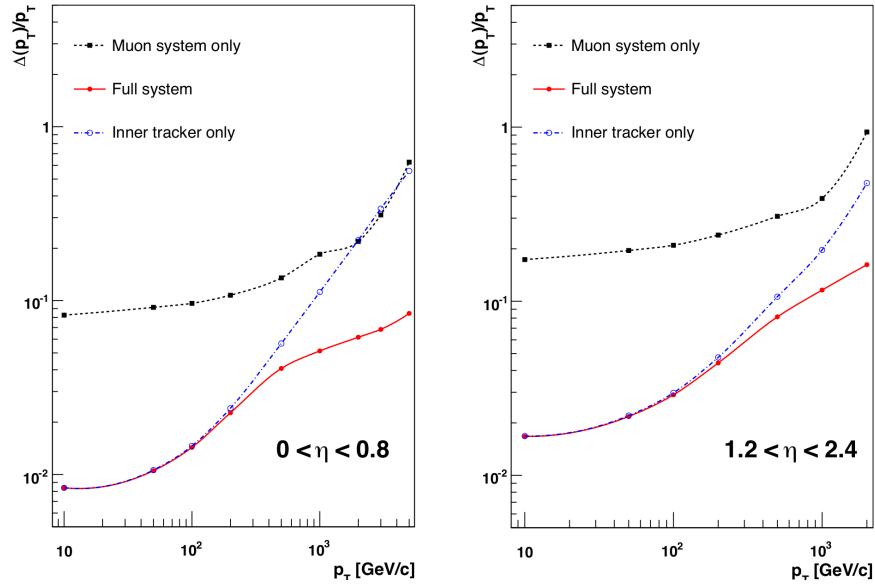


Figure 2.8. Muon transverse momentum resolutions for the tracking system. On the left the barrel zone. On the right the endcap.



Figure 2.9. Schematic overview of the muon chambers.



Figure 2.10. Run-I and Run-II integrated luminosity.

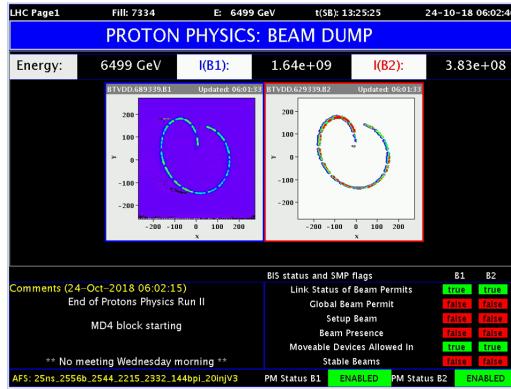


Figure 2.11. Last proton-proton beam dump at the end of Run-II.

in the 2016, LHC was ready to deliver a large dataset to the experiments. The data collected in 2016 are used in the high mass analysis that is the subject of this thesis. Overall, the data correspond to 35.9 fb^{-1} of data validated for the physics analyses. The 2016 LHC operations can be grouped into several time-periods, labelled with a letter from A to H, corresponding to slightly different conditions of data taking. In 2017 and 2018 operations continued and the total integrated luminosity for Run-II is now around $\sim 150 \text{ fb}^{-1}$. In October 2018 the protons collisions have been stopped, Fig. 2.11, and all operation (ions collisions after the proton stop) will halt in 2019 for a second long shutdown of ~ 2.5 year (LS2) for the machine and experiments upgrade. The Run-III will start in 2021 with an energy of 14 TeV. After that, as shown in Fig. 2.12, the third long shutdown (LS3) starting in 2024, will see a substantial upgrade of the LHC and of both Atlas and CMS experiments. The high-luminosity LHC (HL-LHC) starting in ~ 2026 will represent an unprecedented way to study very rare phenomena at the LHC. The machine is expected to deliver, during a decade of operations, an integrated luminosity of $\sim 3000 \text{ fb}^{-1}$.



Figure 2.12. Schedule of LHC and HL-LHC operations.

Chapter 3

Monte Carlo Generators

As we have seen in the collisions between high energy protons hundreds of particles are generally present in the final state. Given the complexity of the events it is necessary to use Monte Carlo generators, i.e. programs that allow to simulate the realistic result of the collisions assuming a certain model for the processes involved. In this chapter are described the various stages beyond the Monte Carlo events simulation. In particular I have participated personally to the Monte Carlo production and validation for the CMS collaboration.

3.1 General Overview

The use of Monte Carlo generators is necessary because it is impossible to predict what happens event-by-event: in fact, in quantum mechanics we can only calculate the probability of having a certain result. The simulation of an event is carried out in successive steps [28, 29], as schematized in Fig. 3.1, thus subdividing the problem into several parts of lower complexity.

The various steps are summarized here:

- Hard process: the incident protons are composed by partons (quarks and gluons) and the hard process consists in a collision between two partons, coming from different hadrons. The matrix element of the process is evaluated perturbatively and often only the lowest perturbative order, called leading order (LO), is calculated.
- Parton shower: the incoming or outgoing partons participating in the hard process can emit gluons. In fact, in analogy with the electromagnetic interaction, a particle with an accelerated color charge can radiate for the bremsstrahlung. The gluons in turn, can produce quark-antiquark pairs thus generating the parton showers. The emission of additional partons takes place mainly in the collinear space respect to the initial parton and progressively with less energies. In the final state there will be a set of partons, called jet, located in the collinear with respect to the initial parton. This probabilistic process can be simulated as a Markov process and is implemented in the parton shower algorithms we will discuss later.



Figure 3.1. Schematic representation of an event generated within an event generator. The partons coming from the protons indicative participate in both the hard process and multiple interactions. Subsequently there is the hadronization.

- Multiple interactions: in a single collision, it may happen to have more pairs of partons interacting. In this case it is said that there are multiple interactions in addition to the hard process.
- Hadronization: in the evolution of the event the partons are generated with gradually lower relative momenta. For momentum values of 1 about GeV the confinement forces prevail. At these energy scales, the perturbation theory fails in the description, so we resort to non perturbative models which describe the formation of real hadrons. This hadronization process preserves the jet structure which can therefore be observed experimentally.
- Decaying of unstable particles: many of the particles produced in the primary process are unstable and they decay unless they interact before with the detector.

In the Monte Carlo simulation all these steps are considered sequentially: the result of each step is the starting point of the next. At the end, in a single event, there are hundreds of particles each of which has a dozen degrees of freedom (mass, flavor, impulse, average life, spin, vertex production, etc.), so there is a high number of parameters that came into play and must be simulated for each event. The final aim is to provide a realistic description of what happens in high-energy collisions, in order to compare the Monte Carlo model with the experimental data. Schematically, the cross section of the final state is given by,

$$\sigma_{final\ state} = \sigma_{hard\ process} \cdot \mathcal{P}_{tot,\ hard\ process \rightarrow final\ state}, \quad (3.1)$$

integrated over the total phase space and summed over all possible final states (e.g. the production of two or more jets). This is the measurable quantity associated with the hard process.

3.2 Hard process

In many processes of interest to LHC, high momenta come into play, to produce high mass particles or energetic jet. The simulation of these events is the main goals of the Monte Carlo generators. The cross section for a scattering $ab \rightarrow n$ process is given [29] by,

$$\begin{aligned} \sigma &= \sum_{a,b} \int_0^1 dx_a dx_b \int f_a^{h_1}(x_a, \mu_F) f_b^{h_2}(x_b, \mu_F) d\hat{\sigma}_{ab \rightarrow n}(\mu_F, \mu_R) \\ &= \sum_{a,b} \int_0^1 dx_a dx_b \int d\Phi_n f_a^{h_1}(x_a, \mu_F) f_b^{h_2}(x_b, \mu_F) \\ &\times \frac{1}{2\hat{s}} |\mathcal{M}_{ab \rightarrow n}(\Phi_n, \mu_F, \mu_R)|^2, \end{aligned} \quad (3.2)$$

where

- $f_a^h(x, \mu)$ are the parton density functions (PDF) that depend on the x fraction of the parton a 's energy (Bjorken variable) respect to the hadron h_i ($i = 1, 2$), and on the μ_F factorization scale, App B.

- $\hat{\sigma}_{ab \rightarrow n}$ is the partonic cross section of the process $ab \rightarrow n$. The total differential cross section is given by the product of the corresponding square matrix element, $|\mathcal{M}_{ab \rightarrow n}|^2$, with the incident particle flow $1/(2\hat{s}) = 1/(2x_a x_b s)$, where \sqrt{s} is the energy of the system's center of mass.
- The matrix element $|\mathcal{M}_{ab \rightarrow n}(\Phi_n, \mu_F, \mu_R)|^2$ can be written as the sum on all Feynman diagrams,

$$\mathcal{M}_{ab \rightarrow n} = \sum_i \mathcal{F}_{ab \rightarrow n}^{(i)}. \quad (3.3)$$

- $d\Phi_n$ it is the phase space differential for n particles in the final state.

The phase space will not be all physical space but will contain cuts for two reasons: the first is that the cuts will reflect the geometry and acceptance of the detector; the second because it is necessary put a cut on the transverse impulse of the particles produced in the process to avoid divergences in the calculation of the cross section ¹. In general, the calculation of the matrix element would require the calculation of all the Feynman diagrams which grow in a factorial way (Fig. 3.2) with the number of particles in the final state. Usually the Monte Carlo events generators can compute

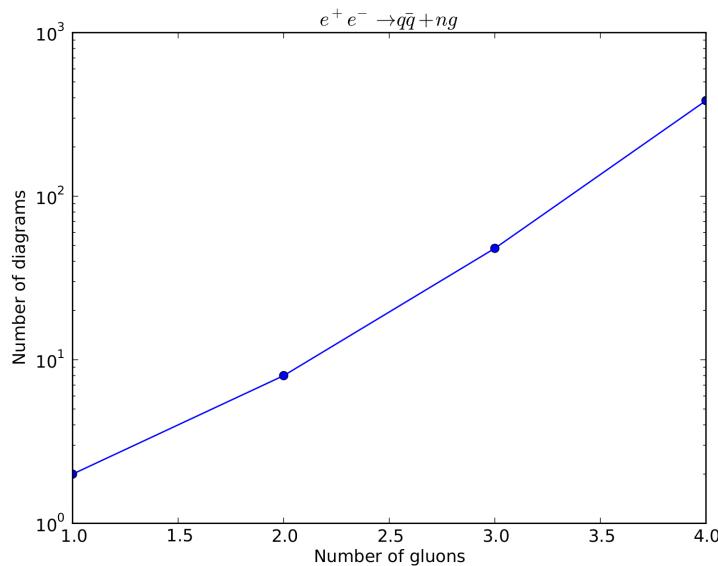


Figure 3.2. Trends in the number of Feynman diagrams as the number n of gluons increases in the process $e^+e^- \rightarrow q\bar{q} + ng$.

the matrix element at the leading order for the Standard Model $2 \rightarrow 1$, $2 \rightarrow 2$ and $2 \rightarrow 3$ [30] processes.

However, if we stop at the first perturbative order, we would have only a rough description of the process: in fact, subsequent orders involve important corrections both to the shape of the distributions and to the total cross section. LO is useful for a first study but it is important to evaluate next-to-leading-order (NLO) ².

¹You can imagine having a singularity similar to that in classical Coulomb scattering.

²For some particularly important processes, for example $gg \rightarrow H$, the next-next-to-leading-order (NNLO) calculations are even available.

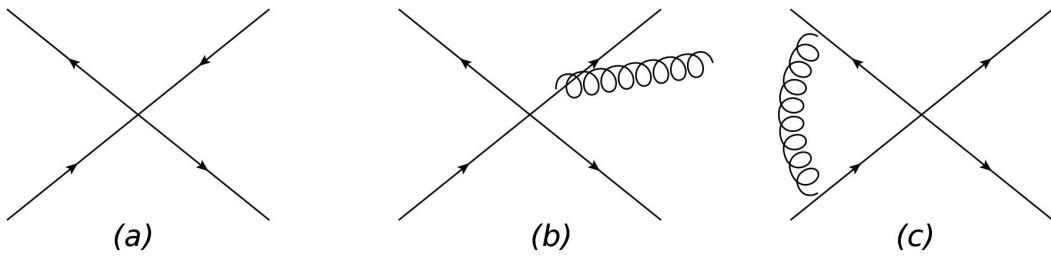


Figure 3.3. Examples of Feynman diagrams (a) Born, (b) real, (c) virtual.

The cross section calculated at the NLO is composed of three parts: the LO part or Born, by the real and by the virtual part of the emission corrections (Fig. 3.3),

$$d\sigma^{NLO} = d\tilde{\Phi}_n [\mathcal{B}(\tilde{\Phi}_n) + \alpha_s \mathcal{V}(\tilde{\Phi}_n)] + d\tilde{\Phi}_{n+1} \alpha_s \mathcal{R}(\tilde{\Phi}_{n+1}), \quad (3.4)$$

where \mathcal{B} , \mathcal{R} and \mathcal{V} denote the Born, the real and the virtual part respectively. The integral must be made on the n or $n + 1$ final state particles and on the Bjorken variables related to the incident partons. Consider, in the Born approximation, the process $2 \rightarrow 2$. If you want to go to the next order, NLO, you have to keep the element with an additional parton in the final state, the $2 \rightarrow 3$ process, and virtual correction with a loop in the $2 \rightarrow 2$ process. It should be noted that the cross-section for processes of the type $2 \rightarrow 3$ is divergent when the energy of one of the partons tends to zero (soft divergence) or when two partons are collinear (collinear divergence).

3.3 Parton shower

In a collision between partons a charge of color is accelerated, so there will be bremsstrahlung emission. When studying a process of the type $2 \rightarrow n$, where n represents the number of partons in the final state, the LO matrix elements (called tree-level) will have divergences in the collinear and soft case. In particular, the processes that suffer from this type of divergence are $q \rightarrow qg$, $\bar{q} \rightarrow \text{bar}qg$, $g \rightarrow gg$: the first are similar processes to $e \rightarrow e\gamma$ in QED, while the third is due to the fact that QCD is not an Abelian theory. The process $g \rightarrow q\bar{q}$ does not have this type of divergence. The divergences of the tree-level matrix element can be removed by introducing the virtual corrections into the calculation, but they will be in the next order; these calculations are therefore particularly complex and they are only possible for a limited number of processes. The parton shower [?] algorithms offer an alternative and simple way to eliminate the collinear and soft divergences through:

- an iterative structure that combines the three states suffering from divergences in a single multi-partonic state,
- the introduction of the form factor of Sudakov.

The incoming or outgoing partons, which are far (temporally) from hard process, are called on-shell. Indeed the module of their four-momenta is equal to the mass at rest. However, the closer one gets to the interaction, the more partons can be

off-shell, i.e. the module of the their four-momenta does not correspond to the mass at rest due to the uncertainty principle ($\Delta E \Delta t \sim \hbar$). For this reason they are able to emit other partons and the energy of emitted partons is higher if they are closer to the scattering. If the emission occurs before the scattering, it is called initial state radiation (ISR), while after the interaction it is called final state radiation (FSR). Each parton is characterized by a “virtuality scale” Q^2 that corresponds roughly to a shower temporal scale. It is important to stress that different definitions are available for Q^2 ; however regardless of the chosen convention, the Q^2 scale increases as it approaches the hard process, in the ISR, and decreases away, in the FSR. If we take the FSR, the evolution starts at a Q_{max}^2 scale that is related to the hard process and it ends when a limit scale is reached, Q_0 , which will be on the order of 1 GeV. The most common choice used is to set $Q^2 = p^2 = E^2 - |\vec{p}|^2$. With this convention in a process of type $a \rightarrow bc$, in FSR case, $Q^2 > 0$, Q is time-like, decreases until the limit scale Q_0 is reached. The ISR case is complicated. In this case, if c is an emitted parton that will not participated in the hard interaction, a and b are off-shell. They are space-like and in order to guaranteed the increase order of Q^2 , i.e. $Q_b^2 > Q_a^2$, it is better to define $Q_i^2 = -m_i^2$. In contrast c is time-like Therefore its shower will evolve like that of the FSR.

Final State Radiation

In the parton shower approach, the final state radiation is modeled through a series of divisional processes of the type $a \rightarrow bc$. This is evident from the process $q\bar{q}g$, Fig. 3.3 (b), where the first order matrix element corrections correspond to the emission of a gluon. The evolution of the shower is described by two parameters: the fraction of energy carried by one of the two outgoing partons, $z = E_b/E_a$, and the order variable t . As we said, a possible choice for t is the virtuality Q_a^2 of the incoming parton. In the collinear limit the probability of division $d\mathcal{P}_{a \rightarrow bc}$, in z and $t = \ln(Q^2/\Lambda^2)$ is:

$$d\mathcal{P}_{a \rightarrow bc} = \frac{\alpha_{abc}}{2\pi} P_{a \rightarrow bc} dt dz, \quad (3.5)$$

where $dt = \frac{dQ^2}{Q^2}$, α_{abc} it is the coupling constant that regulates the division process and $P_{a \rightarrow bc}$ is the kernel splitting; these are universal functions and are valid in the collinear limit:

$$\begin{aligned} P_{q \rightarrow qg} &= \frac{4}{3} \frac{1+z^2}{1-z}, \\ P_{g \rightarrow gg} &= 3 \frac{(1-z(1-z))^2}{z(1-z)}, \\ P_{g \rightarrow q\bar{q}} &= \frac{n_f}{2} (z^2 + (1-z)^2), \end{aligned} \quad (3.6)$$

where n_f is the quarks flavour number. However the probability evaluated is larger than the unity. Indeed it suffers from the same divergences of the matrix element at the LO. The expression 3.5 is evaluated in the collinear approximation. In particular, there are two types of divergences: collinear, due to the dependency of type $1/Q^2$, and soft which corresponds to the limit $z = 1$.

To solve this problem, in the parton shower approach, the probability of dividing t

and $t + dt$ is evaluated; this is obtained with the integration of Eq 3.5 over z in the intervals $[z_{min}(t), z_{max}(t)]$:

$$d\mathcal{P}_a = \left(\sum_{bc} \int_{z_{min}(t')}^{z_{max}(t')} \frac{\alpha_{abc}}{2\pi} P_{a \rightarrow bc} dz \right) dt. \quad (3.7)$$

As in other physical situations ³ the probability of something happening at t is given by the probability that this happens between t and $t + dt$, multiplied by the probability that this has not already occurred between the initial instant t_0 and t . In this case then the division probability at t is:

$$d\mathcal{P}_a^{\text{FSR}}(t) = d\mathcal{P}_a \cdot \exp \left(- \sum_{bc} \int_{t_0}^t dt' \int_{z_{min}(t')}^{z_{max}(t')} \frac{\alpha_{abc}}{2\pi} P_{a \rightarrow bc}(z) dz \right), \quad (3.8)$$

where t_0 is the shower starting scale. The exponential term is called Sudakov factor and it represents the probability of non-division. If you want to interpret it in terms of Feynman diagrams, this represents the virtual corrections of LO matrix element. This total process can be combined to have more emissions at different steps: this will result in a partons shower which will be ordered in decreasing Q . Finally, the description given by parton shower is correct if you have collinear jet and it fail in configurations where there are well separated partons.

Initial State Radiation

The initial state radiation evolution is much more complicated than the final state. Indeed the quark and the gluons are emitted and absorbed continuously, inside the incoming proton. The initial stare radiation is already present during the hard scattering. The ISR simulation could start from the on-shell parton before the interaction and after could evolve to higher and higher Q^2 scales until the hard process. However, this approach is very inefficient because the interesting process is particularly rare and it has the same probability to happen as in nature. In the event generators a different approach is used: first the hard process is produced and then we try to rebuild back what may have happened. This procedure is called backward evolution, Fig. 3.4. It is necessary to evaluate the probability for the process of type $a \rightarrow bc$, that a parton b has been produced by the parton a . For this reason the partonic density function is introduced. This evolves according to the DGLAP [?] equation,

$$\frac{df_b(x, t)}{dt} = \sum_{ac} \int_x^1 \frac{dx'}{x'} f_a(x', t) \frac{\alpha_{abc}}{2\pi} P_{a \rightarrow bc} \left(\frac{x}{x'} \right), \quad (3.9)$$

where $f_{a,b}(x, t)$ are the parton PDFs a, b , that which have x fraction of the incident and scale proton momenta $t = \ln(Q^2/\Lambda^2)$, instead $P_{a \rightarrow bc}$ is the kernel splitting function.

In the backward evolution the probability that the parton b has been generated from

³For example radioactive decay.



Figure 3.4. Evolution of the initial state. The bold line corresponds to the part that will undergo the hard process (represented by a cross). Thin lines represent the partons that can not recombine, while the dashed lines are fluctuations that may or may not recombine.

a in the interval between t and $t - dt$ is given by:

$$d\mathcal{P}_b(t) = \frac{df_b(x, t)}{f_b(x, t)} = |dt| \sum_{ac} \int \frac{dx'}{x'} \frac{df_a(x', t)}{f_b(x, t)} \frac{\alpha_{abc}}{2\pi} P_{a \rightarrow bc} \left(\frac{x}{x'} \right), \quad (3.10)$$

while the probability of non-division between the scale t_{max} and $t < t_{max}$ is:

$$S_b(x, t, t_{max}) = \exp \left(- \int_t^{t_{max}} dt' \sum_{ac} \int \frac{dx'}{x'} \frac{df_a(x', t')}{f_b(x, t')} \frac{\alpha_{abc}}{2\pi} P_{a \rightarrow bc} \left(\frac{x}{x'} \right) \right), \quad (3.11)$$

Finally the probability of combining b in a in the range between t and $(t - dt)$ from is:

$$\begin{aligned} d\mathcal{P}_b^{\text{ISR}}(t) &= - \frac{dS_b(x, t, t_{max})}{dt} dt \\ &= \sum_{ac} \int \frac{dx'}{x'} \frac{df_a(x', t)}{f_b(x, t)} \frac{\alpha_{abc}}{2\pi} P_{a \rightarrow bc} \left(\frac{x}{x'} \right) \cdot S_b(x, t, t_{max}) dt \end{aligned} \quad (3.12)$$

In this case the Sudakov form factor is different respect to FSR as it contains the PDFs. This means that the parton shower results do not depend only on the algorithm but also on the PDFs used.

Resummation

When calculating an observable predicted by QCD in a perturbative way, the expansion in powers of α_S contains terms of the type $\alpha_S^n L^k$ ($k < 2n$), where $L = \ln(q_{cut}/s)$, being q_{cut} the cut on resolvable emission. When we consider small values of q_{cut} the logarithm of the perturbative expansion becomes large and the perturbative series diverges. The main perturbative order of the expansion is n only if the successive terms of the series are negligible, however this is not guaranteed if there are high value of L . It is therefore necessary to consider the terms that have a high value of the logarithm. The study of these terms is called resummation and is done by putting the terms together in the perturbation series according to their degree of divergence: $\alpha_S^n L^{2n}$ are the leading log (LL) terms, $\alpha_S^n L^{2n-1}$ are the next-to-leading log (NLL) terms, and so on. At the end all α_S orders terms are added. For many processes calculations are available at the NLL. The parton shower approximate the effects of resumming at the NNLL.

Merging among ME and PS

The two different approaches for the matrix element calculation and for the parton shower have advantages and disadvantages. Regarding the ME we have:

- the LO matrix element calculations can be performed exactly in the cases where there are many jet (of the order of six) in the final state,
- a good description of separate partons is performed,
- the perturbative calculations are correct,
- however, the cross section diverges in the collinear and soft case, so an exhaustive description of the internal structure of jet is not possible.

On the other hand the PS:

- it is a universal approach that produces a realistic configuration of the partons,
- the divergences, in the collinear limit, are treated with the introduction of the form factor of Sudakov. So we have an appropriate description of the jet evolution
- however, the method fails when describing separate partons, since the collinear approximation in this case can not be valid.

Clearly the two methods are complementary and their merging is desirable. There are different approaches that combine ME with PS. The main difficulty is to cover the total phase space without overlaps or holes: we want to describe a process in which there are n well separated partons in the final state, using the LO matrix element but also including the large logarithms resummation (LL, NLL) which is typical of the PS. A schematic description of the combination for four jet is given in Fig. 3.5. On the horizontal axis are the α_S coupling orders, while on the vertical axis the logarithm. The PS includes the LL ($m = 2n$) and the NLL ($m = 2n - 1$) green circles (e.g. in the case of $n = 2$, $m = 4, 3$ the two circles in green marked as “4”). The circles that describe the event with four jet, when combining the ME and PS, are green, the blue and three red circles marked with the “4”. The difficulty arises because the ME describes exactly all the circles marked with the “4”: so if we simply sum up the two approaches we would double counts these green circles. The most used approaches to merge the ME and PS are:

- parton shower reweight: the basic idea is to start from the process to the lowest order and then re-evaluate the output of the PS as if it had been produced by the ME. This approach does not change the cross section, which remains at the lowest order, but improves the population of the phase space [31, 32].
- CKKW prescription: the phase space is divided into two zones using k_{perp} which is a measure of the cut Q_0^2 : the region in which the jet is produced is filled with the ME, that of evolution with the PS [33, 34].
- The MLM prescription, which is also very widespread, is based on the same principle, but is implemented in a different way.

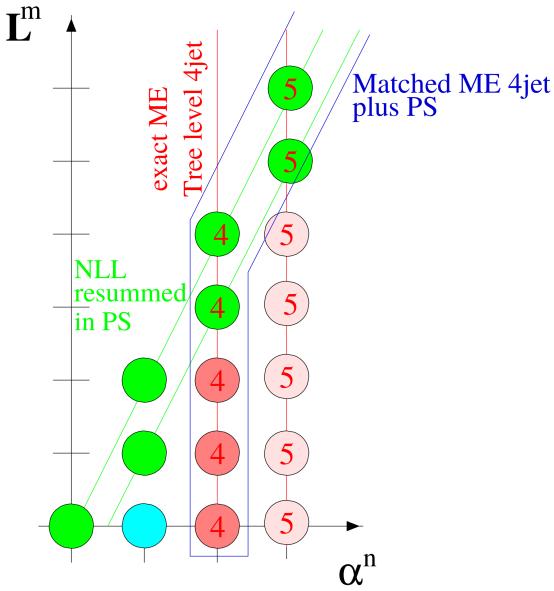


Figure 3.5. Merging among ME and PS.

3.4 Multiple Interaction

Incident protons participating in the interaction are composed by large number of partons (quark and gluons) that can interact independently with each other in addition to the hard process. The total cross section for the QCD process $2 \rightarrow 2$ is dominated by the t -channel, so the cross section diverges as dp_{perp}^2/p_{perp}^4 for $p_\perp \rightarrow 0$ [?]. So when simulating a real event, in addition to the hard event, characterized by having large transverse transverse momentum, we must also take into account the additional collisions at small p_\perp . If these occur independently then a Poisson distribution is expected, $P_n = \langle n \rangle^n \exp(-\langle n \rangle)/n!$. However, conservation of energy and momenta means that interactions are not effectively independent, thus suppressing the possibility, for $p_{perp} \rightarrow 0$, of having a high number of interactions. It should also be noted that in order to eliminate the divergence it is necessary to introduce a cut-off value of the transverse pulse, below which no collisions are generated.

3.5 Hadronization

In this context, the process of hadronization is a particular model, used in event generators, which describes the transition from the final partonic state to the final hadron state, which is an experimental observable. It is important to underline that this transition is treated in a phenomenological way and not by a rigorous approach. The two most important classes for tuning are the string model and the cluster model. The difference is that the former transforms the partonic systems directly into hadrons, while the second takes an intermediate step where it groups the objects to a scale of ~ 1 GeV.

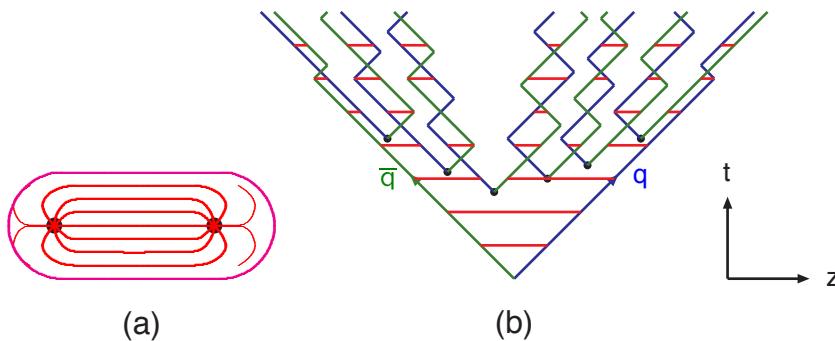


Figure 3.6. (a) The flow tube between a quark and an antiquark moving away. (b) Motion and breaking of a system string.

String Model

The Lund model is the most complete string model: we know from QCD that there is a linear confinement force between the partons that increases with distance. Consider, as an example, the final state in which there are two quarks, $q\bar{q}$. As the partons move away the color flow tube is stretched between q and \bar{q} , Fig. 3.6 (a). The transverse dimensions of the tube are the typical dimension for the hadrons, therefore about 1 fm. If the tube is assumed to be uniform, the potential increases linearly, $V(r) = \kappa r$, with $\kappa \approx 1$ GeV/fm, string constant. At short distances it would be necessary to introduce an additional Coulomb term, $\sim \frac{\alpha_s}{r}$, however in the Lund model this term is negligible. As the quark and antiquark move away from the interaction vertex, the potential energy accumulated in the string increases until it breaks, giving rise to a pair $q'\bar{q}'$. So the system is divided into two new color singlets $q\bar{q}'$ and $q'\bar{q}$. These two systems will move away repeating the process below. The evolution of the system in space-time is represented in Fig. 3.6 (b). At the end of the process a series of $q_i\bar{q}_i$ pairs are presented, each of which will form a hadron. For now, only the case $q\bar{q}$ has been considered. However, if more partons come from the interaction, the string model becomes more complicated. For an event in which there is an additional gluon, $q\bar{g}g$, the string is stretched between q and g and between g and \bar{q} , Fig. 3.7.

Cluster Model

This hadronization model is based on the pre-confining property of the parton shower: the distribution of the invariant mass of a single pair of opposite-colored partons is the same at any Q^2 scale. The distribution increases rapidly at low value , Fig. 3.8 (a).

In the model, the gluons from parton shower are represented by pairs of color-anticolor lines connected to the vertex. Each color line, near the cutoff, is connected to another colorless line present at the same scale. At this point the contiguous color/anticolor lines are interpreted, in the non-perturbative limit, as quark-antiquark pairs which give rise to mesons, which are observable objects in the final state. This mechanism is represented in Fig. 3.8 (b).

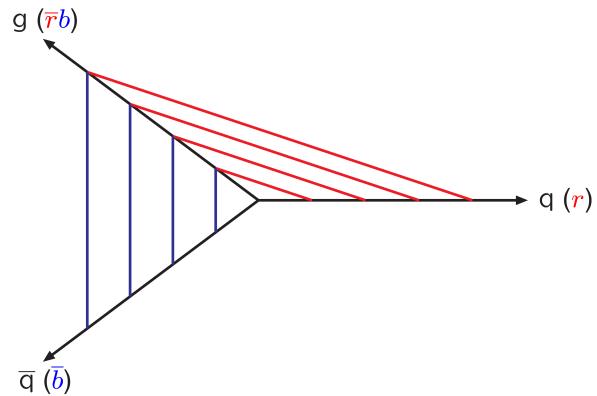


Figure 3.7. Motion of the string in the case $q\bar{q}g$.



Figure 3.8. (a) Invariant mass distribution for singlets. (b) Parton shower structure in the cluster model.

3.6 Hadronic Decays and Electromagnetic Radiation.

In the hadronization step, unstable hadrons which decay into other particles can be produced. So the final state is the result of the convolution between the hadronization and the decay. The information necessary for the decay of unstable particles is generally taken from the “Particle Data Book” (PDG) [?] which provides the properties (e.g. average life) of a large number of particles. In general, in an event generator, it is necessary to choose which hadrons to include in the simulation and then select the possible decay channels. In addition to hadronic decays, it is also necessary to simulate the emission of electromagnetic radiation. The most common approach adopted is to use algorithms similar to those used to simulate the emission of QCD in parton shower.

3.7 Jets

At the end of this process, after the hadronization and the decaying of unstable particles it is still possible to estimate the four-momenta of the partons generated in the hard process as the direction and energy of the jets that are reconstructed from the final state particles [?]. The jets are reconstructed by an algorithm that calculates the distance, d_{ij} , between two objects (particles or pseudo-jet) defined as,

$$d_{ij} = \min(k_{ti}^{2p}, k_{tj}^{2p}) \frac{\Delta_{ij}^2}{R^2}, \quad (3.13)$$

where $\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$ and k_{ti} , y_i and ϕ_i are the transverse momentum, the rapidity and the azimuthal angle of i respectively. The constant R is the radial parameter. The distance between i and the beam is also introduced, $d_{iB} = k_{ti}^{2p}$. The algorithms proceed by calculating the minor distance d_{ij} between all the pairs of particles i, j . The four-momenta of the two particles with the smaller distance are added. The d_{iB} is evaluated for every i and if it is less than the distance d_{ij} with all other particles j , i , than is considered a jet and it is removed from list of objects present in the event. Finally the distances are recalculated and this whole procedure is repeated until there are no more objects to be added. For a value of $p = -1$ the algorithm is called anti- k_t [35]. This is what has been used in this work, Sec. 4.6.

3.8 Main Monte Carlo generators

For a proton-proton collision at the LHC, different Monte Carlo generators are available. Each of these has different methods for combining the ME with the PS. A short introduction of the main generators is given below.

Madgraph_aMC@NLO

The MADGRAPH [30] approach is very ambitious, in fact the purpose of this generator is to calculate the cross section at the NLO including automatically both real and virtual contributions in the calculation. The hard process is produced by the ME while the soft emissions are added by the PS. The first step is to compute the ME

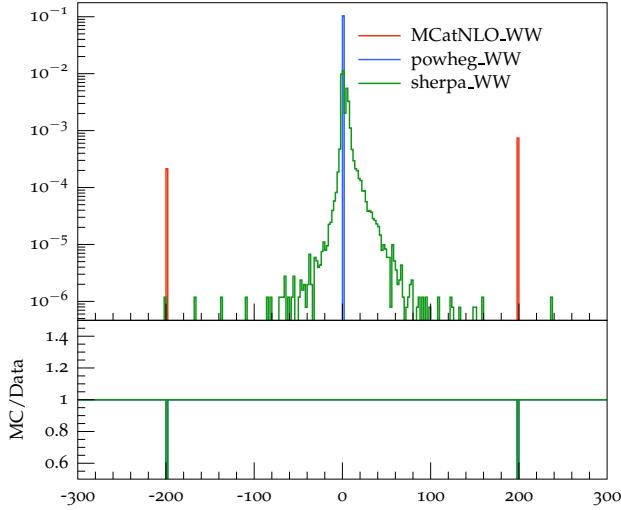


Figure 3.9. Weight distribution, for different Monte Carlo generators, with cross section normalization of 1 fb^{-1} .

NLO corrections for a process involving n partons. It results from $n + 1$ partons due to the real corrections and from n partons due to the virtual corrections. As next step, it is evaluated how the parton shower populates the phase space for $n + 1$ parton, excluding the Sudakov form factor. To get the state with $n + 1$ partons, MADGRAPH subtracts the PS from the ME. The PS, without the Sudakov, and the ME are in agreement in the soft and collinear limits, i.e. the singularities are deleted thus obtaining a finite value for the cross section for n and $n + 1$ partons. A technical problem arises in the collinear limit. Here, there is no certainty that the ME overhangs the PS everywhere. This problem is solved by introducing a fraction of negative-weight events, Fig. 3.9. Finally, the parton shower is applied, which includes the Sudakov factor and thus allows a finite and correct result to be obtained at the NLL.

POWHEG

The idea behind POWHEG [36] is to generate the hardest radiation first, and then pass the event to the parton shower generator. In parton shower generators, the production, ordered in a transverse pulse, of the harshest radiation is always the first; so POWHEG simply replaces this with the NLO emission. In POWHEG events are produced with a positive and constant weight (Fig. 3.9).

PYTHIA8

PYTHIA 8 [37] is a generator that can calculate the ME for processes with two particles or partons in the final state, but above all it generates the parton shower and the subsequent synchronization. The parton shower is ordered in a transverse impulse, p_T , and the first issue is corrected with reweight method. For hadronization is used the Lund model.

SHERPA

SHERPA [38] is a Monte Carlo generator that like PYTHIA8 provides a complete description of hadronic collisions, from the calculation of the matrix element, up to the stable particles. The parton shower includes both QCD and QED emissions, i.e. photons. It can calculate the ME for the main processes (eg $gg \rightarrow H$) at the NLO and combine the ME with the PS. The code is written completely in C ++ language.

3.9 Monte Carlo samples in High Mass Analysis

Several Monte Carlo generators were used in the searching of a high mass particle to simulated the signal and the backgrounds. All processes are simulated using the NNPDF3.0 [39, 40] parton distribution functions (PDF) for NLO ME generators, while the LO version of the same PDF is used for LO ME generators. All the event generators are interfaced to PYTHIA 8.1 [41] for the showering of partons and hadronization, as well as for including a simulation of the underlying event (UE) and multiple interactions (MPI) based on the CUET8PM1 tune [42]. For all processes, the detector response is simulated using a detailed description of the CMS detector, based on the GEANT4 package [43]. The simulated samples are generated with distributions for the number of pileup interactions that are meant to roughly cover, though not exactly match, the conditions expected for the different data-taking periods. In order to factorize these effects, the number of true pileup interactions from the simulation truth (as stored in the Monte Carlo) is reweighted to match the data. The re-weighting is propagated automatically to both the in-time pile up and the out-of-time one. The pileup histogram for reweighting is calculated using the *pileupCalc* tool as described in [44]. Different calculations are used to obtain the cross sections for the all processes at 13 TeV. All simulated sample are summarized in Tab. 3.1.

Signal

In order to perform the resonance search in a large part of the mass spectrum, several signal samples for the gluon-gluon fusion and the vector boson fusion mechanisms have been generated corresponding to different Higgs boson masses in the range between 200 GeV to 3 TeV. All signal samples have been simulated with POWHEG v2 [45, 46, 47], designed to describe the full NLO properties of these processes. In particular, for Higgs produced via gluon fusion [48], and vector-boson-fusion (VBF) [49], the decay of the Higgs boson into two W boson and subsequently into leptons was done using JHUGen [50]. The signals which correspond to a Higgs boson mass of 125 GeV have been simulated accordingly and are treated as backgrounds in the following analysis, including the associated production with a vector boson (W^+H , W^-H , ZH) [51], and gluon fusion produced ZH ($ggZH$). For associated production processes the Higgs boson decay was done via PYTHIA 8.1 [41]. For Higgs signals, the cross sections used are the ones reported by the LHC Higgs Cross Section Working Group [52], computed at NNLO and NNLL QCD and NLO EW for gluon fusion, and at NNLO QCD and NLO EW for the rest of the production

Process	MC generator used	$\sigma \times \text{BR}$ [pb]
Signal gluon-gluon fusion	Powheg V2+JHUGen	Various
Signal VBF	Powheg V2+JHUGen	Various
$t\bar{t} \rightarrow WW$ $b\bar{b} \rightarrow 2l2\nu b\bar{b}$	Powheg V2	87.31
$q\bar{q} \rightarrow WW \rightarrow 2l2\nu$	Powheg V2	12.178
$q\bar{q} \rightarrow WW \rightarrow l\nu qq$	Powheg V2	0.59
$gg \rightarrow WW \rightarrow 2l2\nu$	MCFM 13TeV	0.5905
Single top	Powheg+pythia8	71.70
Drell-Yan	amcatnloFXFX+pythia8 Madgraph+pythia: Inclusive Madgraph+pythia: HT 100-200 Madgraph+pythia: HT 200-400 Madgraph+pythia: HT 400-600 Madgraph+pythia: HT 600-Inf.	1867 6025.26 147.4 40.99 5.678 2.198
Multibosons	$WZ \rightarrow 2\ell 2q$: amcatnloFXFX+pythia8 $ZZ \rightarrow 2\ell 2q$: amcatnloFXFX+pythia8 WW : amcatnlo+pythia8 WZZ : amcatnlo+pythia8	5.5950 3.2210 0.1651 0.05565
$W+jets$	Inclusive: amcatnloFXFX+pythia8 $W p_T$ 0-50: amcatnloFXFX+pythia8 $W p_T$ 50-100: amcatnloFXFX+pythia8 $W p_T$ 100-200: amcatnloFXFX+pythia8 $W p_T$ 250-400: amcatnloFXFX+pythia8 $W p_T$ 400-600: amcatnloFXFX+pythia8 $W p_T$ 600-Inf: amcatnloFXFX+pythia8	61526.7 57280.0 3258.0 676.3 23.94 3.031 0.4524

Table 3.1. Simulated samples for the signal, $t\bar{t}$, WW , DY, Multiboson and $W+jet$ production.

modes. The branching fractions are the ones reported in [15].

The WW sample

The WW production, irreducible background for the analysis, was simulated in different ways. POWHEG v2 [53] was used for $q\bar{q}$ induced WW in different decays. The cross section used for WW processes produced via $q\bar{q}$ was computed at NNLO. The WW , produced via gluon-gluon fusion, was generated, with and without Higgs diagrams, using MCFM v7.0 [54]. The cross section for normalizing WW , produced via $q\bar{q}$, was computed at next-to-next-to-leading order (NNLO). The leading-order (LO) cross section for $ggWW$ is obtained directly from MCFM. For gluon fusion, the difference between LO and NLO cross sections is significantly big. A scale factor of 1.4 is theoretically calculated [55] and applied to the $gg \rightarrow WW$ background.

In the analysis two different WW Monte Carlo samples are merged: the “ $WW \rightarrow 2l2\nu$ ” at NLO and the “ WW plus 2 quark” at LO. The second sample, “ WW plus 2 quark” at LO, contains final state with two quarks or a gluon-quark system: only the final state with two quarks interferes with the signal. To avoid double count between the two sample a cut on di-jet mass at generator level, mjj_{GenLev} , is applied. In particular the sample “ $WW \rightarrow 2l2\nu$ ” at NLO is used for $mjj_{GenLev} < 100$ GeV and the “ WW plus 2 quark” at LO is adopted for $mjj_{GenLev} > 100$ GeV. The distribution for the reco di-jet mass is shown in Fig.3.10. In particular the red distribution correspond the “ $WW \rightarrow 2l2\nu$ NLO” sample with a cut of $mjj_{GenLev} < 100$, the blue distribution

to “WW plus 2 quark” with $m_{jjGenLev} > 100$ GeV. The sum of the red and blue distributions is shown in black. There is a good agreement between the black distribution and the “ $WW \rightarrow 2l2\nu$ NLO” without any $m_{jjGenLev}$ distribution in green.

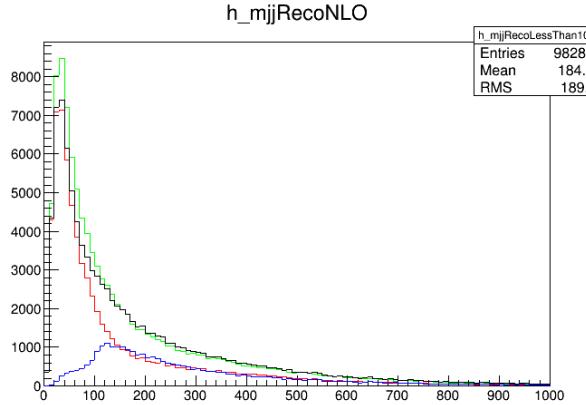


Figure 3.10. Distribution for m_{jj} at RECO level for the merged WW sample.

The Top sample

In order to control the top quark background processes, the analysis is performed in jet bins as described in Cap. 5. The jet binning enhances the importance of logarithms of the jet p_T , spoiling the convergence of fixed-order calculations of the $q\bar{q} \rightarrow WW$ process and requiring the use of dedicated resummation techniques for an accurate prediction of differential distributions [56, 57]. Since the p_T of the jets produced in association with the WW system is strongly correlated with its transverse momentum, p_T^{WW} , the simulated $q\bar{q} \rightarrow WW$ events are reweighted to reproduce the p_T^{WW} distribution from the p_T -resummed calculation. A $t\bar{t}$ dilepton sample was also generated using POWHEG v2. The cross sections of the different single top processes are estimated by the LHC Top Working group [58] at NLO. The $t\bar{t}$ cross section is also provided by the LHC Top Working group [59], and it is computed at NNLO, with NNLL soft gluon resummation.

The DY sample

For the Drell-Yan backgrounds we use two different sets of samples. For the opposite flavor analysis (Sec 5.7), selecting events with an electron and a muon, a dedicated sample in which only the $Z/\gamma^* \rightarrow \tau\tau \rightarrow e\mu\nu\nu$ decay is simulated. For the same flavor analysis (Sec. 5.8), in which pairs of electrons or muons are selected, a soup of different H_T binned DY samples is used. A detailed study about this soup is given below. Drell-Yan production of Z/γ^* is generated using MADGRAPH [60] and the cross section is scaled using a LO to NNLO k-factor equal to 1.23. Given the lack of MC statistics in the LO inclusive DY sample the H_T -binned samples are used. This helps increasing the MC statistics especially in the VBF category of the same flavor analysis, which is characterized by large values of H_T . The LO

inclusive sample is used for events with $H_T < 100$ GeV and it has been merged to the other samples selecting events with H_T below 100 GeV using the parton level information. The cross sections of those samples have been scaled applying the LO to NNLO k-factor. In Fig. 3.11 the H_T distribution of the sample after the merging is reported, showing a smooth transition between different H_T samples. To further check the correct behaviour of the H_T binned samples we compared

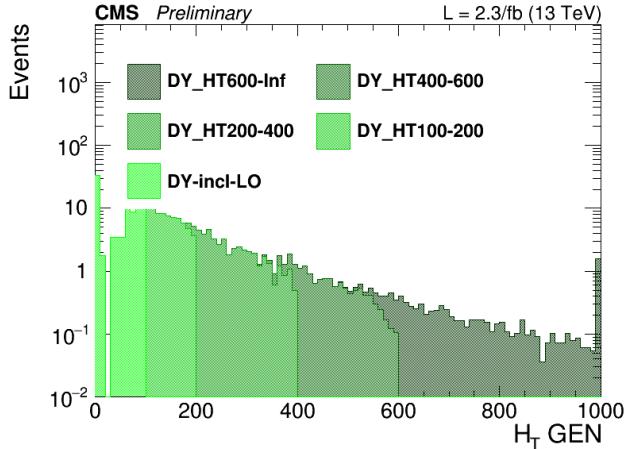


Figure 3.11. H_T distribution for the merged DY sample.

them to the inclusive LO sample, selecting only the events with a generator level H_T above 100 GeV. The comparison is done in a control region with two same flavor leptons with $p_T > 20$ GeV and $m_{\ell\ell} > 50$ GeV, showing very good agreement between the two samples. The distributions of some variables are shown in Fig. 3.12 To check the differences between the LO inclusive sample and the NLO sample simulated with MC@NLO, the two samples have been compared in a same flavor control region and some variables of interest are shown in Fig. 3.13. The control region is defined requiring two same flavor leptons with $p_T > 20$ GeV and with $m_{\ell\ell} > 50$ GeV. The simulated samples are generated with distributions for the number of pileup interactions that are meant to roughly cover, though not exactly match, the conditions expected for the different data-taking periods. In order to factorize these effects, the number of true pileup interactions from the simulation truth is reweighted to match the data. The re-weighting is propagated automatically to both the in- time pile up and the out-of-time one. In Fig. 4.1, the effect of this reweighting on a sample enriched in Drell-Yan events is shown. In order to select this sample, events with two electrons with $p_T > 25$ GeV for the leading one and $p_T > 13$ GeV for the trailing one, are selected only if $|m_{\ell\ell} - m_Z| < 10$ GeV.

Other processes

Other multiboson processes, such as WZ, ZZ, and VVV ($V=W/Z$), are generated with amC@NLO and normalized to the cross section obtained at NLO in generation. The cross sections for the remaining processes were directly obtained from the generator itself.

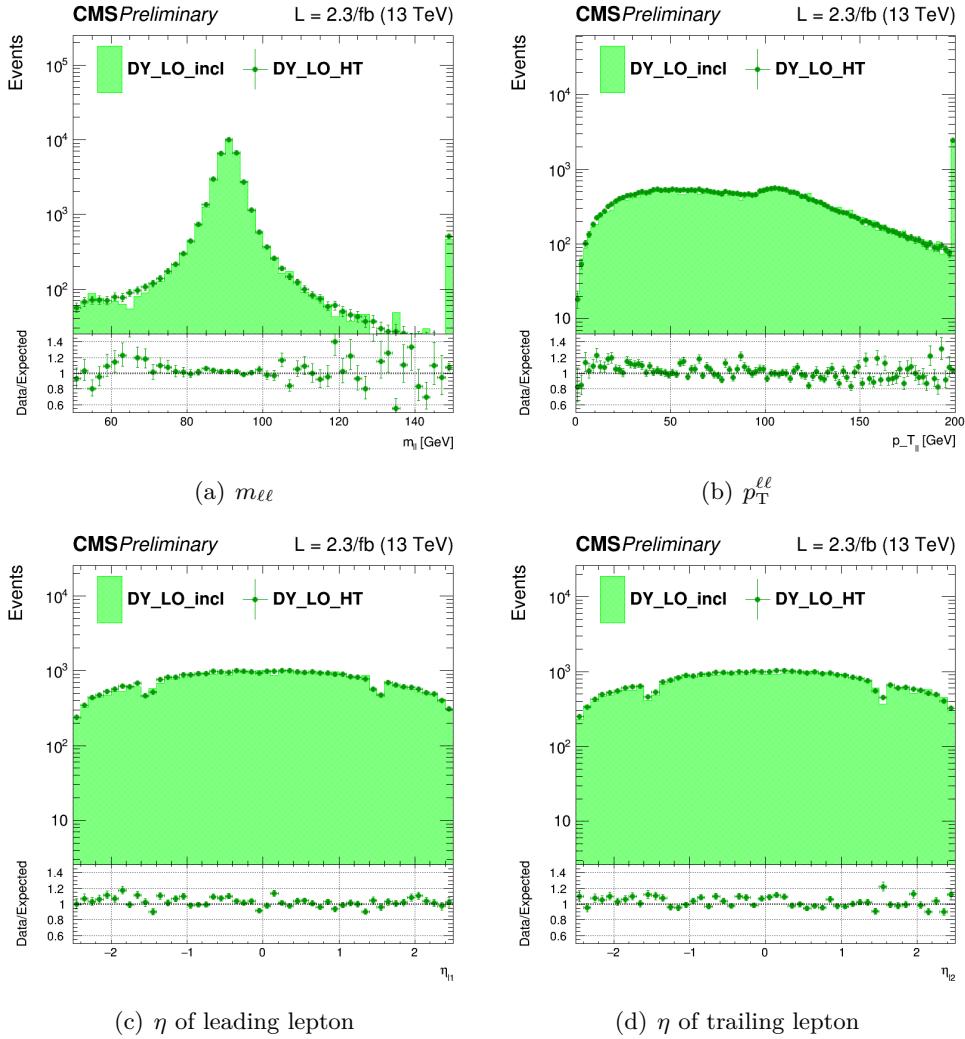


Figure 3.12. Comparison between the inclusive LO DY sample and the H_T binned samples.

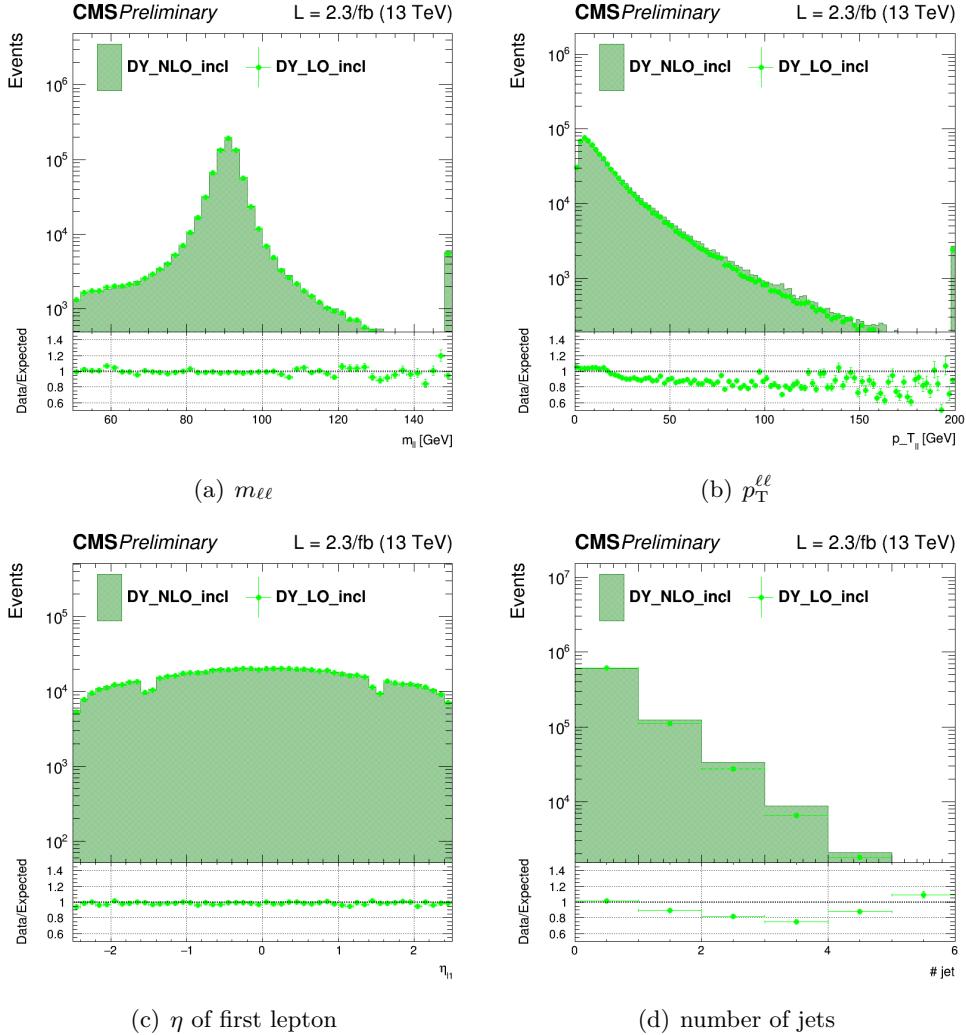


Figure 3.13. Comparison between the LO and NLO DY samples.

Chapter 4

Event Reconstruction

In the search of high mass resonances in the W^+W^- channel, leptons (specifically, electrons or muons and neutrinos) and jets are searched for to select signal events. The reconstruction of such objects follow standard algorithms and procedure which have been developed by the experiment. In this chapter the reconstruction of the interested objects are described. I have contributed in the evaluation of the best working point for Standard Model Higgs boson analysis with 2015 data [61] and marginally for the same analysis using 2016 data [62], Tab. 4.5.

4.1 The Particle Flow

The algorithm for the event reconstruction is called Particle Flow (PF) [63]. Its goal is to reconstruct all stable particles that emerge from a hadron-hadron collision, such as muons, electrons, photons, charged hadrons, and neutral hadrons. The PF technique uses all the informations coming from the CMS sub-detectors and combines them to obtain: type of particle, direction, momentum, and energy. The reconstruction of charged particles is done from the hits in the silicon tracker that is immersed in a uniform magnetic field of 3.8 T. Transverse momentum is precisely measured down to about 150 MeV. The photon and electron reconstruction is performed using the high resolution and high granularity of ECAL together with the excellent tracking system. This combination provides a very good energy resolution for such particles. Muons are reconstructed in the muon chambers together with the tracking system. The reconstruction of charged-particles in the tracker, energy clusters in the calorimeters and muons in the muon chambers are the first steps of the PF. The informations are then connected to each others, making blocks of elements which are topologically compatible. Starting by blocks, the candidate particles (PF Candidates) are fully reconstructed and identified as:

- Muons: the combination of a track in the tracker and a track in the muon system gives rise to a PF muon. The corresponding track is removed from the block after the identification.
- Electrons: a charged-particle track is linked and one or more ECAL clusters (if presents)

- Charged hadrons: PF charged hadrons are obtained from the remaining tracks. Tracks are linked to ECAL and HCAL clusters, and the deposited energy is determined taking into account the momentum of the track.
- Photons and Neutral hadrons: PF photons are given by ECAL clusters not compatible with charged-tracks; PF neutral hadrons instead are given by unaccounted HCAL deposits.

After that, once the list of PF Candidates is defined, the PF jets are clustered. At the LHC, at each bunch crossing there are of the order of 20 minimum bias proton-proton interactions, which pollute any interesting hard events with many soft particle. The kinematic measurements for jets will be adversely affected by pileup (PU), with resolution and absolute energy measurements suffering significantly. Therefore effects of PU needs to be taken into account and corrected for. At the end the PF MET is evaluated as the opposite of the transverse momentum-vector sum over all reconstructed PF Candidates.

In the following, each step of the reconstruction is described in details.

4.2 Tracking

Hit reconstruction in the pixel and strip detector

The first step of the reconstruction process is referred to as local reconstruction [64]. It consists of the clustering of zero-suppressed signals above specified thresholds in pixel and strip channels into hits, and then estimating the cluster positions and their uncertainties defined in a local orthogonal coordinate system (u, v) in the plane of each sensor (pixel and strip). The hit reconstruction in the pixel detector and in the strip detector is described below:

- The pixel detector: in the data acquisition system of the pixel detector, zero-suppression is performed in the readout chips of the sensors, with adjustable thresholds for each pixel. This pixel readout threshold is set to a single-pixel threshold corresponding to an equivalent charge of 3200 electrons. Offline, pixel clusters are formed from adjacent pixels, including both side-by-side and corner-by-corner adjacent cells. Each cluster must have a minimum charge equivalent to 4000 electrons.
- The strip detector: the clusters are seeded by any channel passing zero-suppression that has a charge at least a factor of three greater than the corresponding channel noise. Neighbouring strips are added to each seed, if their strip charge is more than twice the strip noise. A cluster is kept if its total charge is a factor five larger than the cluster noise. The position of the hit corresponding to each cluster is determined from the charge-weighted average of its strip positions, corrected for the Lorentz drift.

Track reconstruction

Different algorithms are used in CMS for track reconstruction, [64, 65]. All methods use the reconstructed positions (hits) of the passage of charged particles inside

the CMS silicon detectors to determine the trajectories of the charged tracks and therefore measure their directions and momenta. The Combinatorial Track Finder (CTF) is the main standard algorithm and proceeds in three steps:

- Seeding: pairs of hits, that are compatible with a charged particle, originated from the interaction region and with p_T above a lower threshold, are considered as possible candidates of charged tracks. Pixel hits provide the best track seeding, given their three-dimensional position information and lower occupancy.
- Finding: is based on a standard Kalman Filter pattern recognition approach. The track trajectory is extrapolated to the neighboring tracker layers, starting from the seeded parameters. Compatible hits are assigned to the track. The Kalman Filter, a succession of alternating prediction and filtering steps, updates the track parameters at each steps with new hits. The updated tracks are assigned a quality and only the best ones are kept for further propagation.
- Fitting: the final estimate of the parameters of each track helix is completed in the last steps applying again the Kalman Filter for the trajectory fitting. Each trajectory is refitted using a least-squares fit. In the central region, $|\eta| < 1$, the p_T resolution is better than 1% for tracks with $p_T < 10\%$. At higher p_T resolution worsens approximately as $\Delta(1/p_T) \sim 0.2 \text{ TeV}^{-1}$.

The reconstruction of single muons with the CTF algorithm is almost fully efficient over the whole acceptance range. Electrons, being charged particles, can be reconstructed through the standard track reconstruction. However, as electrons lose energy primarily through bremsstrahlung, rather than ionization, large energy losses are common. The energy loss distribution is highly non-Gaussian, and therefore the standard Kalman filter is not appropriate. Electron candidates are reconstructed using the information from the tracker but also from the ECAL, Sec 4.5. For charged hadrons, finally, there are inefficiencies in the reconstruction due to their nuclear interactions in the tracker material.

Primary-vertex reconstruction

The goal of primary-vertex (PV) reconstruction [66] is to measure the location, and the associated uncertainty, of all proton-proton interaction vertices in each event, including the “signal” vertex and any vertices from pileup collisions, using the available reconstructed tracks. It consists of three steps:

- selection of the tracks;
- clustering of the tracks that appear to originate from the same interaction vertex;
- fitting for the position of each vertex using its associated tracks.

Track selection involves choosing tracks consistent with being produced promptly in the primary interaction region, by imposing requirements on the maximum value of significance of the transverse impact parameter, the number of strip and pixel hits

associated with a track and the normalized χ^2 from the fit to the trajectory [64]. The selected tracks are then clustered on the basis of their z-coordinates at their point of closest approach to the centre of the beam spot using a *deterministic annealing* (DA) algorithm [67]. The DA process finds not only positions and assignments of tracks to vertices but also the number of vertices. In Fig. 4.1 is shown the distribution of the number of vertices in a Drell-Yan sample. This distribution is related to the number of PU events. Therefore MC events are reweighted according to the ratio between data and simulation in the number of primary vertices in order to correct for the different pileup.

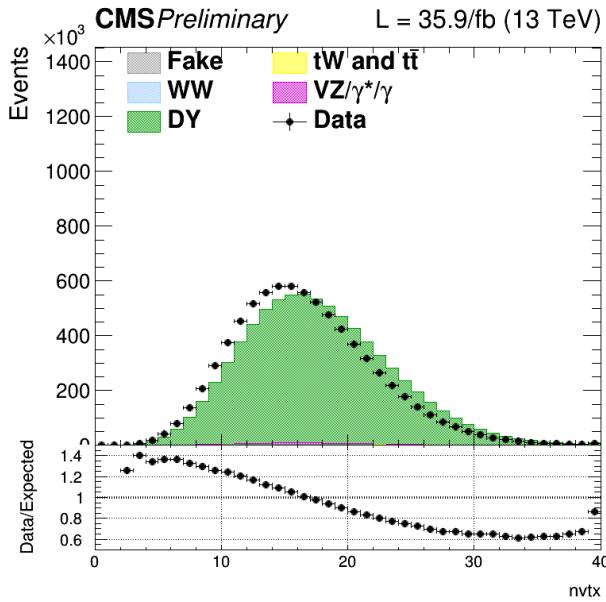


Figure 4.1. Distributions of the number of vertices in a Drell-Yan enriched sample ($Z \rightarrow ee$) in data

4.3 Calorimeter clusters

The purpose of the clustering algorithm in the calorimeters are [68]:

- detect and measure the energy and direction of stable neutral particles such as photons and neutral hadrons;
- separate these neutral particles from charged hadron energy deposits;
- reconstruct and identify electrons and all accompanying bremsstrahlung photons;
- help the energy measurement of charged hadrons for which the track parameters were not determined accurately, which is the case for low-quality and very high p_T tracks.

For the PF event reconstruction, a specific clustering algorithm was developed. The clustering is performed separately in each subdetector: ECAL (barrel and endcaps)

and HCAL (barrel and endcaps) In the algorithm, first, *cluster seeds* are identified as cells with an energy larger than a given seed threshold, and larger than the energy of the neighbouring cells. Second, *topological clusters* are grown from the seeds by aggregating cells with at least a corner in common with a cell already in the cluster.

4.4 Link algorithm

A given particle is, in general, expected to give rise to several PF elements in the various CMS subdetectors. The reconstruction of a particle therefore proceeds first with a link algorithm that connects the PF elements from different subdetectors. The link between a track in the central tracker and a calorimeter cluster is established as follows: the track is first extrapolated from its last measured hit in the tracker to the ECAL at a depth corresponding to the expected maximum of a typical longitudinal electron shower profile, and to the HCAL at a depth corresponding to one interaction length. The track is linked to a cluster if its extrapolated position is within the cluster area, defined by the union of the areas in the HCAL and the ECAL in (ϕ, η) plane. After, Calorimeter cluster-to-cluster links are sought between HCAL clusters and ECAL clusters. Charged-particle tracks may also be linked together through a common secondary vertex, for nuclear-interaction reconstruction. Finally, a link between a track in the central tracker and information in the muon detector is established to form global and tracker muons, Sec 4.5.

4.5 Lepton reconstruction and identification

Electron reconstruction and identification

The combination of tracker and ECAL information is used for electrons reconstruction. The electrons, emerging from a collision, when interact with the silicon tracker, radiate bremsstrahlung photons that reaches the ECAL with a significant spread in the azimuthal direction ϕ . These energy depositories measured in ECAL are the starting point of the electrons reconstruction algorithm. They are associated in clusters and in superclusters (clusters of clusters) that take into account the spread in the ϕ direction of the bremsstrahlung energy, Fig. 4.2. When superclusters are identified, the reconstruction algorithm tries to match them to track seeds. The track seeds are identified as pairs or triplets of hits in the inner tracker layers. The electrons trajectories are reconstructed using a dedicated modeling that take in account energy loss in the tracker layers via bremsstrahlung radiation: non-Gaussian contributions to the event-by-event fluctuations of the calorimetry and tracking measurements are introduced due to the bremsstrahlung radiation. A preselection is applied to solve ambiguous cases where several tracks are reconstructed. It is based on matching between the GSF track and the supercluster in η and ϕ . To achieve a good resolution the electron supercluster must be inside the ECAL acceptance volume, that meaning $\eta < 2.5$, and outside the ECAL barrel-endcap overlap region, $1.4442 < \eta < 1.566$. Several identification variables are used to achieve a good discrimination. These variables are:

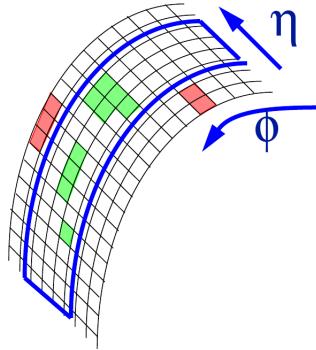


Figure 4.2. EM cluster spread in η and ϕ .

- $\Delta\eta_{trk,SC}$ and $\Delta\phi_{trk,SC}$, that measure the spacial matching between the track and the supercluster.
- $\sigma_{in,in}$ that measures the width of the ECAL supercluster in the η direction. It is the calorimeter shower shape.
- H/E: is the ratio among the energy deposit in the HCAL tower and the energy of the seed supercluster.
- $|1/E - 1/p|$ the difference of $1/E$ measured in ECAL and $1/p$ measured in the tracker.
- Number of missing hits
- d_{xy} and d_z , impact parameters with respect to the reconstructed primary vertex (PV).
- $\gamma \rightarrow e^+e^-$ veto based on missing hits in the inner layers of the tracker.

Different selections on the above variables define different working points (WP). The cuts are also different for electrons in the ECAL barrel or in the endcap. The WPs are:

- Tight WP: this corresponds to an average 70% selection efficiency for electrons with $p_T > 20$ GeV. This working point is used where backgrounds are very large. The $X \rightarrow WW$ high mass analysis has large backgrounds like W+jets, where the second lepton is a “fake”, i.e. a jet identified as an isolated lepton. So, in this analysis we use Tight WP for electrons.
- Medium WP: the average efficiency is about 80% for electrons with $p_T > 20$ GeV. This is also a good starting point for measurements of W and Z cross-sections.
- Loose WP: this working point is used only for very clean final states. The average efficiency is about 90%.
- Veto WP: generally, this is not used for signal selection. However, it is found to be useful for extra lepton veto counting of electrons. The average efficiency is about 95%.

Observable	Barrel (EB) cut	Endcap (EE) cut
$ \Delta\eta_{trk,SC} $	0.00308	0.00605
$ \Delta\phi_{trk,SC} $	0.0816	0.0394
$\sigma_{in,in}$	0.011	0.031
H/E	0.060	0.065
$ 1/E - 1/p $	0.013	0.013
Number of missing	1	1
$ d_{xy} $	0.05	1
$ d_z $	1	1
conversion veto	true	true
$ISO_{EA \text{ corrected}}^{Rel}$	0.0588	0.0571

Table 4.1. Electron identification criteria for the Tight working point and isolation requirements.

Selected electrons are also required to pass the isolation criteria that include a pile-up mitigation correction based on the electron effective catchment area. The isolation variable is computed for each electron as,

$$ISO_{EA \text{ corrected}}^{Rel} = [\sum_{ChH} (p_T) + \max(0, \sum_{Ph} (p_T)) + \sum_{NH} (p_T) - \rho EA]/p_T^{electron}, \quad (4.1)$$

where ChH is an index which runs on the charged hadrons, Ph on the photons, NH on the neutral hadrons, ρ is the pile up energy density and A is an effective area. The sum is done in a isolated cone of $\Delta R < 0.4$ around the electron direction. The identification and isolation criteria used for the $X \rightarrow WW$ analysis are summarised in Tab. 4.1.

Muon reconstruction and identification

Muons produced in the interaction point can pass through all the detector, with a negligible energy loss, and give a signal in the muon chambers. They are thus seen both in the silicon tracker and in the external muon chambers. The reconstruction starts from the measurements of DT, CSC and RPC sub-detectors. This reconstructed track in the muon spectrometer is called Stand-alone Muon. In parallel, muon tracks are also reconstructed in the inner silicon tracker, as described before for generic charged particles. The tracker track is then combined with the Stand-alone Muon track in order to construct a global track, which defines a Global Muon. A global fit is performed for each pair of tracks reconstructed in the inner tracker and in the muon system. If more than one track matching the stand-alone track is found, then the one giving the best χ^2 in the global fit is chosen. A complementary approach consists in considering all tracker tracks with $p_T > 0.5$ GeV as potential muon candidates. These tracks are extrapolated to the muon system taking into account the magnetic field. If at least one muon segment (a short track stub made of DT or CSC hits) matches the extrapolated tracks, the corresponding tracker track is identified as a Tracker Muon. Quality requirements are applied to ensure a good quality of the reconstruction:

Observable	Value Cut
Is global muon	true
Is PF muon	true
Tracker layers with measurements	> 5
Number of valid pixel hits	> 0
Number of valid muon hits	> 0
Number of matched muon stations	> 1
χ^2 / ndof	< 10
d_{xy} (PV)	< 0.2 cm
d_z (PV)	< 0.5 cm

Table 4.2. Muon identification and isolation requirements.

- the tracker track has to be reconstructed from at least 5 tracker layers with hits;
- at least one hit must be present in the pixel detector;
- at least one hit must be present in the muon detector;
- at least one muon chamber hit should be included in the Global Muon track fit;
- the normalized χ^2 of the Global Muon track fit should be less than 10;
- the muon track reconstructed in the tracker must have a distance to the primary vertex smaller than 2 mm in the transverse plane and smaller than 5 mm in the longitudinal direction.

The requirements are summarized in Tab.4.2

For the analysis goals, the muons are expected to be isolated, as they are generated by W boson decay. Indeed, prompt muons are expected to be isolated in the event, contrary to non-prompt muons that are generally produced within jets and characterized by many nearby particles. Muons coming from W are therefore requested to pass an isolation criterion, which includes a PU mitigation, called “ $\Delta\beta$ correction”, in order to ensure its independence and robustness on the number of simultaneous interactions. The isolation variable used is,

$$ISO_{\Delta\beta}^{Rel} = [\sum_{ChH} (p_T) + \max(0, \sum_{Ph} (p_T) - 0.5 \times \sum_{ChHPU} (p_T))] / p_T^{electron}, \quad (4.2)$$

where ChH is the charged hadrons, Ph is photons and $ChHPU$ is charged hadrons not coming from the primary vertex. The sum is performed in a cone of 0.4 units in ΔR around the muon.

A bias in the muon p_T is introduced by the imperfect knowledge of the magnetic field and the effect of the material distribution. In addition the p_T measurement is sensitive to the alignment of the tracker and muon chambers. To estimate the muon p_T scale and resolution events with two muons from a J/Ψ or a Z resonance decay are used.

4.6 Jet reconstruction and identification

Jets are the experimental signature of quarks and gluons produced in the hadron collision. They result from the parton hadronization and they play a major role in a hadronic collider where final state with jets have very large cross-section. Using the particle-flow algorithm described in Sec. 4.1, the jets are reconstructed by clustering of the four-momentum vectors of PF candidates. The momentum and spatial resolutions of these jets reconstructed with the PF technique are greatly improved respect to jets obtained using the calorimeters information only, because the tracking detector allows a better resolution of p_T for the charged particles. The four-vectors of input particles are combined with a sequential and iterative jet clustering algorithm, that should have ideally the following features:

- infrared safety: no infrared singularity appears in perturbative calculations for the results of this algorithm;
- collinear safety: same as before for collinear singularities;
- invariance under boosts: the algorithm should find the same solutions independently by boosts in the longitudinal direction.
- order independence: the algorithm should find the same jets at parton, particle, and detector level;
- straightforward implementation: the algorithm should be straightforward to implement in perturbative calculations.

The jet algorithm should also follow these experimental criteria:

- detector independence: the performance of the algorithm should be independent from the detector;
- minimization of resolution smearing and angle biases: the algorithm should not amplify the inevitable effects of resolution smearing and angle biases;
- stability with luminosity: jet finding should not be strongly affected by pileup;
- efficient use of computing resources: minimum of computer time consumption;
- maximal reconstruction efficiency: the algorithm should efficiently identify all jets;
- ease of calibration and use: algorithm should not obstruct a reliable calibration and be straightforward to implement.

Jet definition is not unique, being the parton not a well-defined object, from the experimental point of view, so several approaches for jet clustering are available. Two main algorithms have been developed: the “conical recombination” where jets are defined from the dominant directions of energy flow; the “sequential recombination” that works by defining a distance between pairs of particles, then performing subsequent recombinations of pairs of closest particles, and stopping when all resulting objects are too far apart. The standard algorithms adopted by CMS are the SISCone

in the conical recombination class, and the k_t , Cambridge-Aachen (CA) and anti- k_T algorithms for the sequential, Sec 3.7.

Jet Energy Calibration

The purpose of the jet energy calibration is to relate, on average, the energy measured for the detector jet to the energy of the corresponding true particle jet. A true particle jet results from the clustering (with the same clustering algorithm applied to detector jets, Sec. 4.6) of all stable particles originating from the fragmenting parton, as well as of the particles from the underlying event (UE) activity.

The jet energy calibration is related, on average, on the energy measured in the detector to the true energy of the corresponding final state particle jet or parton jet. From the clustering of all stable particles, a true particle jet results. The correction is applied as a multiplicative factor C to each component of the raw jet four-momentum vector, p_μ^{raw} , as:

$$p_\mu^{corr} = C \cdot p_\mu^{raw}, \quad (4.3)$$

where correction factor C is composed of the offset correction C_{offset} , the MC calibration factor C_{MC} , and the residual calibrations C_{rel} and C_{abs} for the relative and absolute energy scales, respectively. The various components are applied in sequence:

$$C = C_{offset}(p_\mu^{raw}) \cdot C_{MC}(p'_T, \eta) \cdot C_{rel}(\eta) \cdot C_{abs}(p''_T), \quad (4.4)$$

where p'_T is the transverse momentum of the jet after applying the offset correction and p''_T is the transverse momentum of the jet after all previous corrections. The details of each component are:

- C_{offset} : a energy offset is produced by the pileup of multiple proton-proton collisions and by electronic noise. The goal of the offset correction is to subtract, on average, the unwanted energy from the jet. For the offset correction estimation, the Jet Area Method has been used [69].
- C_{MC} : is based on the simulation and corrects the energy of the reconstructed jets such that it is equal on average to the energy of the generated jets (GenJets). The GenJets reconstruction algorithm is identical to the one applied to the data. The response variable, $R = p_T^{reco}/p_T^{gen}$, in each bin of the GenJet transverse momentum p_T^{gen} , is recorded as jet p_T^{reco} . The average correction in each bin is defined as,

$$C_{MC}(p_T^{reco}) = \frac{1}{\langle R \rangle}, \quad (4.5)$$

- C_{rel} : the goal of the relative jet energy scale correction is to make the jet response flat versus η . The size of this residual correction is of the order of 2-3% in the central η region, while it goes up to about 10% in the forward region.

- C_{abs} : the goal of the absolute jet energy scale correction is to make the jet response flat versus p_T . Once a jet has been corrected for η dependence, it is corrected back to particle level.

Each type of correction has uncertainties arising from many different sources. These sources are categorized as: physics modeling in MC, MC modeling of true detector response and potential biases in the methodologies used to estimate the corrections. In CMS more than 16 such sources of uncertainties have been identified. Several are related and can be combined into groups that are relative to the absolute scale, relative scale, extrapolation in p_T , pileup, jet flavor and time stability.

Jets in $X \rightarrow WW$ analysis

The clustering algorithm used for jet reconstruction in the $X \rightarrow WW$ analysis is the anti-kt algorithm with a distance parameter equal to 0.4. Pileup mitigation is done with the Charge Hadron Subtraction algorithm [70] which removes charged particles coming from pileup vertices before clustering. To reject jets originating from calorimeter or readout electronics noise the loose working point of the PF Jet Identification is applied. This analysis selects jets with

$$p_T^{Jet} > 30 \text{ GeV}, |\eta| < 5.0. \quad (4.6)$$

Kinematic distributions of the jets after applying corrections are shown in Fig. 4.3 for a Drell-Yan enriched sample. As can be seen from the plot there is a good agreement between simulation and data.

4.7 b-jet identification

To identify the b jets, the peculiarities that characterize hadrons containing the quark b are used. These hadrons have a relatively large mass, around 5 GeV, and a long lifetime, of about 1.5 ps. Given that they can have an impulse of several tens of GeV, the distance that they travel in the detector it is of the order of:

$$\begin{aligned} \delta &\approx \langle t \rangle v \approx \gamma \tau \beta c \\ &\approx 10 \cdot 1.5 \cdot 10^{-12} \text{s} \cdot 1 \cdot 3 \cdot 10^8 \text{ m/s} \\ &\approx 5 \text{ mm}, \end{aligned} \quad (4.7)$$

a distance that can be measured since the accuracy of the position of the secondary vertices is $\sim 100 \mu\text{m}$. The CMS collaboration has developed different algorithms to identify the jets coming from quarks b (b-tag algorithms) [71]. Each of these algorithms is characterized by the efficiency of signal identification (i.e. to select jets from b quark) versus the probability of rejecting the background (jets that are not from b quark), which in general depend on the transverse impulse, p_T , and on the pseudorapidity of the jet. To build an observable (or discriminator) which can separate jets originated by b from those originating from light quark reconstructed objects such as tracks, vertices and leptons are used. Some simple algorithms use a single observable in input, while others, more complexes, combine several variables. Each algorithm produces, in output, a value of the discriminator for each jet of

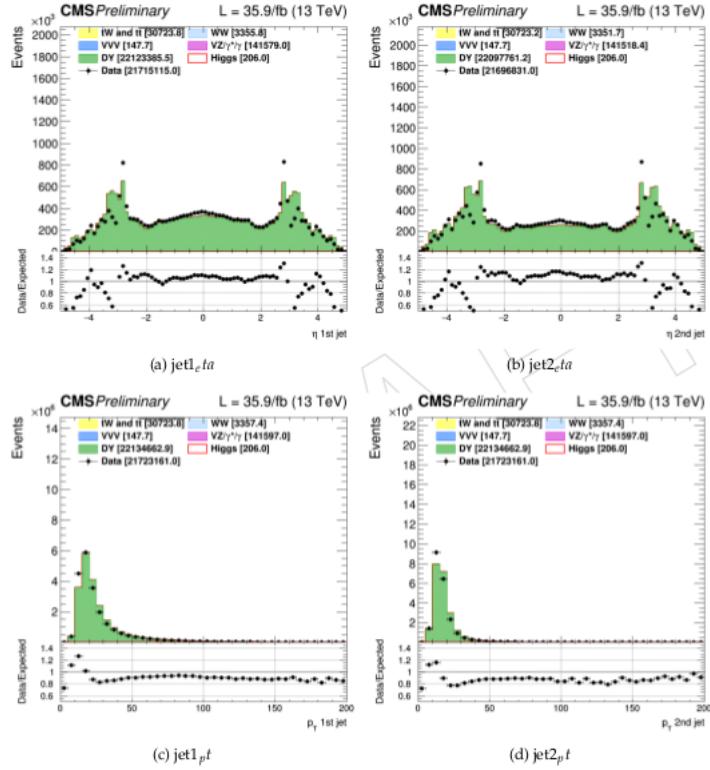


Figure 4.3. The jet kinematic distributions

the event. The first step in identifying of b-jets is the reconstruction of all the jets in the the event, using, as mentioned, the anti- k_T algorithm. However b-jet identification algorithms require a sample with well reconstructed and high purity tracks. Therefore further cuts to the tracks in the jet are applied. First of all, to reduce the number of tracks reconstructed incorrectly, transverse momentum must be greater than 1 GeV and at least two hits must be present in the silicon pixel detector. Then, at least eight hits per each track must be present and the fit must have $\chi^2/\text{d.o.f.} < 5$, where d.o.f. it is the number of degrees of freedom. A selection is also applied on the impact parameter (see below): this is used to increase the fraction of well reconstructed tracks and to reduce the contamination due to long life particles like neutral kaons. The transverse distance d_{xy} and longitudinal d_z between the track and the primary vertex are required to be smaller than 0.2 cm and 17 cm respectively. In Fig. 4.4, a schematic representation of the parameters used in the identification of the b jets is given.

Impact Parameter algorithm

The impact parameter (IP) of a track respect to the primary vertex, can be used to distinguish the hadrons b decay respect to background tracks (prompt). The IP is calculated in three dimensions thanks to the excellent resolution of the pixel detector. The impact parameter has the same sign of the scalar product between the direction of the jet and the vector between the primary vertex and the nearest point of the

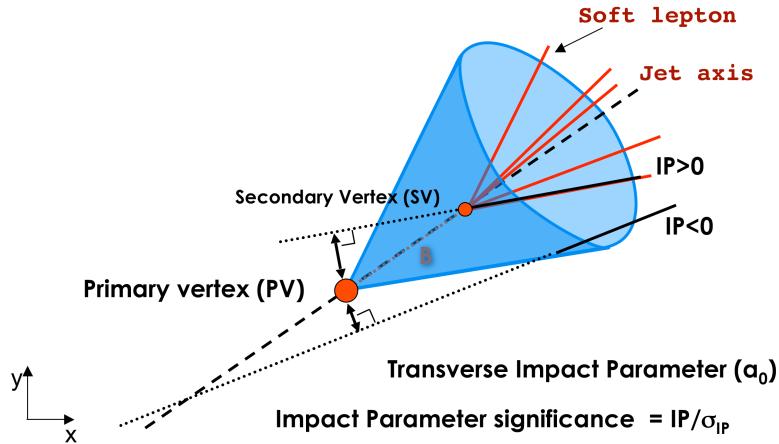


Figure 4.4. Cluster spread in η and ϕ .

track. The tracks due to the decay of a hadron with a long lifetime have a positive IP, while for tracks originated from the primary vertex the IP is equally negative or positive. The variable used as observable is the significance of the impact parameter, S_{IP} , defined as the ratio between IP and its estimated uncertainty. The Track Counting (TC) algorithm orders according to their IP significance the tracks of the jet, from the highest to lowest. There are two versions of this algorithm that differ for which value of the IP significance is used as the discriminator: Track Counting High Efficiency (TCHE) uses the S_{IP} of the second track, while the Track Counting High Purity (TCHP) uses the third. A natural extension of the TC algorithms is the combination of the several tracks impact points. This is implemented in the Jet Probability (JP) and Jet B Probability (JBP) algorithms. The former uses an estimator for the likelihood that all the tracks in the jet come from the primary vertex of the interaction. This estimator, P_{jet} , is defined as,

$$P_{jet} = \Pi \cdot \sum_{i=0}^{N-1} \frac{(-\ln \Pi)^i}{i!} \quad \text{con} \quad \Pi = \prod_{i=0}^N \max(P_i, 0.005), \quad (4.8)$$

where N is the number of tracks considered and P the probability that the track is originated in the primary vertex. Instead the JBP algorithm gives more weight to the tracks (up to a maximum of four) that have a high value of the IP significance.

Secondary Vertex identification

The presence of a secondary vertex in the event and related variables can be used to discriminate jets coming from a quark b respect to the others. The main variables associated with the secondary vertex are the distance, the flight direction with respect to the primary vertex, the invariant mass and the energy of the tracks associated to secondary vertex. To identify a secondary vertex it is required that:

- less than 65% of its tracks are associated with the primary vertex and the significance of the radial distance between the two vertices is beyond 3σ ;
- the flight distance of each candidate is in a cone of $\Delta R < 0.5$ around the jet direction;

- it has not a radial distance of more than 2.5 cm from the primary vertex and a mass compatible with that of K_0 or exceeding 6.5 GeV (this is needed to reduce contamination from long-lived mesons and interactions with the detector material).

The Simple Secondary Vertex (SSV) algorithm uses as a discriminating variable the significance of the flight distance, given by the ratio of the flight distance to its uncertainty. Similar to the TC algorithm, there are two versions of SSV: High Efficiency (SSVHE) which uses vertices to which there are at least two associated tracks, and High Purity (SSVHP) that requires at least three associated tracks. There are more complicated algorithms that consider the secondary vertices together with the information on the lifetime. This makes possible to construct a discriminant even when there is no secondary vertex, increasing efficiency with respect to SSV; two of these algorithms are the Combined Secondary Vertex (CSV) and the Combined Secondary Vertex Version 2 + Inclusive Vertex Finder (CSVV2IVF). The difference between the two is that the CSV algorithm uses information from the jets only to reconstruct the vertices while the CSVV2IVF algorithm uses all the tracks present in the event.

b-tag performance in $X \rightarrow WW$ analysis

In order to assess which tagger and working point are performing better, we have calculated the signal significance for different taggers and working points in events with 0 or 1 jet with $p_T > 30$ GeV, which are the most sensitive to the dominant gluon fusion production mode. In events with 0 jets with $p_T > 30$ GeV, the b-tag algorithm is applied to jets with $20 < p_T < 30$ GeV, if any. The significance has been computed running the full analysis for the Standard Model Higgs at 125 GeV [62], but including only the systematic uncertainties associated to the b-tagging [71]. The results are shown in Tab. 4.5. The results show that the usage of CMVA (loose WP) or Deep CSV (medium WP) leads to a comparable signal significance in the combined 0+1 jet category. The CMVA tagger with loose WP has been found to be the best option for the analysis, given the good performance and the nice agreement that has been observed between data and MC, Fig. 4.6 and Fig. 4.7.

4.8 The Missing Transverse Energy

The longitudinal momentum (along the beam axis) in the collision is not known, so the measurement of the total missing energy is impossible. However the initial transverse momentum, carried by the incoming partons, is zero, so in the final state of the collision, for the conservation of the momentum components, the sum of momenta of all the particles must be zero. If a missing transverse momentum, \vec{p}_T^{miss} , in the transverse plane is present, it is the evidence of invisible particles, such as neutrinos or other weakly interacting particles predicted by some BSM models. The \vec{p}_T^{miss} is defined as,

$$\vec{p}_T^{\text{miss}} = - \sum_{\text{PF } \text{Obj}s} \vec{p}_T^{\text{PF } \text{Obj}}, \quad (4.9)$$

Tagger (WP)	Significance		
	0 jet	1 jet	0+1 jet
CMVA (loose)	7.31	4.86	8.76
CMVA (medium)	7.39	4.52	8.66
CMVA (tight)	7.35	4.16	8.44
CSV (loose)	7.12	4.64	8.47
CSV (medium)	7.37	4.47	8.62
CSV (tight)	7.36	4.15	8.45
Deep CSV (loose)	7.19	4.80	8.61
Deep CSV (medium)	7.38	4.79	8.80
Deep CSV (tight)	7.39	4.25	8.52

Figure 4.5. Signal significance for different taggers and working points in the 0 and 1 jet categories. The significance for the combination of the 0 and 1 jet categories is shown as well. Only the systematic uncertainties associated to the b tagging scale factors are taken into account for the significance evaluation.

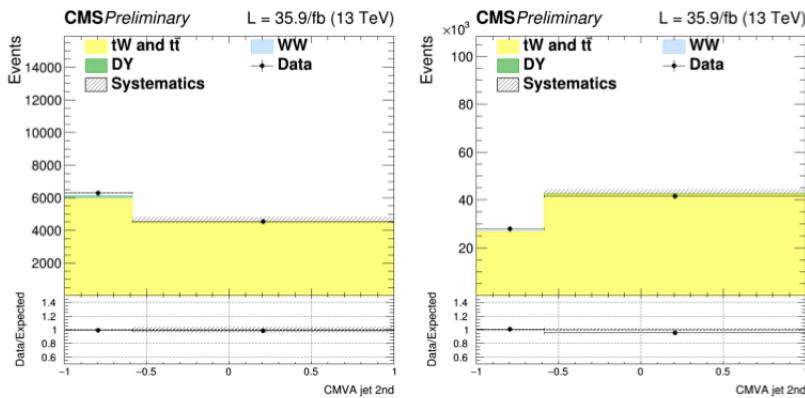


Figure 4.6. CMVA discriminator for jets above 30 GeV (a) and between 20 and 30 GeV (b) in a top enriched control region, after applying the scale factors. The top background normalization is scaled to match data. The systematics band comprises only the uncertainties related to the b-tagging.

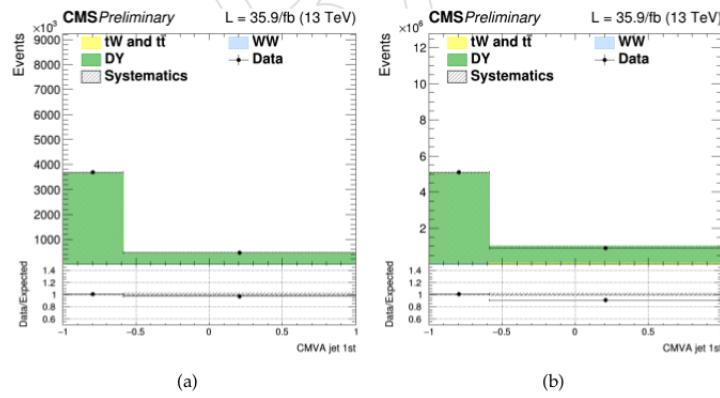


Figure 4.7. CMVA discriminator for jets between 20 and 30 GeV (a) and above 30 GeV (b) in the Z enriched control region. The normalization of the DY background is scaled to match data. The systematics band comprises only the uncertainties related to the b-tagging.

where the sum extends over all the PF objects. The module of \vec{p}_T^{miss} is usually referred to as missing transverse energy, $E_T^{miss} = |\vec{p}_T^{miss}|$. Inefficiencies of the tracking algorithm, minimal thresholds in the calorimeter energy estimation, and nonlinearities of the energy response of the calorimeters for hadronic particles can introduce a bias in the \vec{p}_T^{miss} . A correction is applied by propagating the jet energy corrections to the \vec{p}_T^{miss} formula:

$$\vec{p}_T^{miss\,corr} = \vec{p}_T^{miss} - \sum_{Jets} (\vec{p}_T^{JEC} - \vec{p}_T), \quad (4.10)$$

where the superscript JEC refers to corrected jets. To estimate the E_T^{miss} systematic uncertainty, the following sources have been taken into account:

- jet p_T ;
- jet resolution;
- muon p_T ;
- electron p_T ;
- unclustered energy.

The E_T^{miss} systematic uncertainty has been computed varying each of these uncertainties and adding in quadrature the difference with respect to the nominal value. For the ϕ uncertainty, the largest ϕ variation has been used. The distributions of various E_T^{miss} variables, in a Drell-Yan enriched sample, are shown in Fig. 4.8

4.9 Fake Lepton Background Estimation

Lepton fake rates are measured as a function of the lepton p_T and η , in a single lepton triggered sample. The test of the method and systematic errors are described below. In the analysis the primary source of background from misidentification is

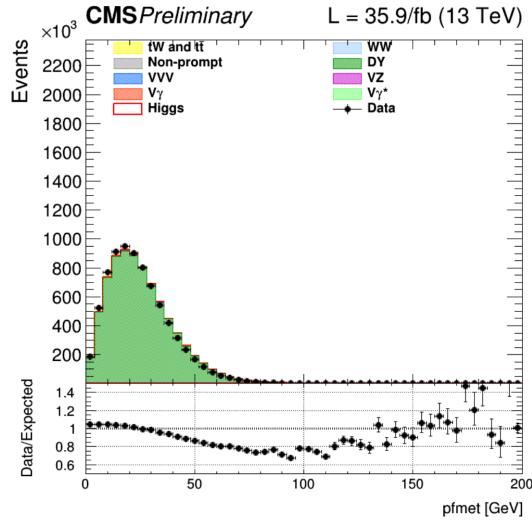


Figure 4.8. The E_T^{miss} distribution.

W+Jets. QCD multi-jet and hadronic top backgrounds are also present at much smaller level. Events in which W bosons are produced in association with jets give rise to background to WW events when a jet is misidentified as a lepton. These events contain a real lepton and real missing energy from the W decay. With the jet misidentified as a lepton, the W+Jets events have two identified leptons, missing energy, and no other significant event characteristics. As a result, the W+Jets events cannot be easily suppressed by event selection. This background is particularly important at low p_T . The estimation of the fake lepton contribution is based on the “fakeable object” data-driven method and provides a measurement of the yield and the kinematic distributions of fake background. It is a general technique, applicable to any physics analysis in which particle level selection criteria are used to suppress background. The method can be used with any number of final state particles and is independent of the event selection. The fundamental idea of the fakeable object method is simple: select a control sample of events enriched in the background being estimated, and then use an extrapolation factor to relate these events to the background in the signal region. The method is data-driven provided the control sample is selected in data, and the extrapolation factor is measured with data. For background arising from particle misidentification, the extrapolation is done in particle identification space (p_T and η) of the lepton. The control sample is defined using a looser particle selection criteria that are chosen such that the rate of misidentification is increased. The extrapolation factor relates background misidentified with this criteria, to background misidentified as passing the full particle selection of the signal region.

Chapter 5

Search for high mass resonances

In this chapter the $X \rightarrow W^+W^- \rightarrow 2\ell 2\nu$ analysis using 2016 data is reported. To increase the sensitivity of this search, the signal must be selected in the most efficient way by reducing the presence of the backgrounds with a similar signature: the selection criteria are described in detail. I have participated at all stages of the analysis selection (signal simulation, categorization, background estimation, etc.) and I have been responsible for the whole analysis.

5.1 Overview of the fully leptonic analysis

The analysis strategy for the high mass search in the $X \rightarrow W^+W^- \rightarrow 2\ell 2\nu$ final state must take into account the production modes of the new scalar, the main backgrounds, and the interference among the different processes.

The main production mode for a Higgs-like particle over the all mass spectrum is the gluon-gluon fusion process. However the ratio of the VBF cross-section to the gluon-gluon fusion cross-section increases with m_X (see Fig. 1.11), making the VBF production mechanism more and more important.

Among the SM processes that have the same final state of the signal or a similar one the most important are non-resonant WW production, top production, Drell-Yan. The WW production is an irreducible background and needs to be determined together with the signal in the fit. To estimate the other background processes, control regions are defined on data and compared to simulation.

The events are first divided according the flavour in the final state:

- opposite-flavour final state, $e^\pm\mu^\mp$,
- same-flavour final state, e^+e^- and $\mu^+\mu^-$.

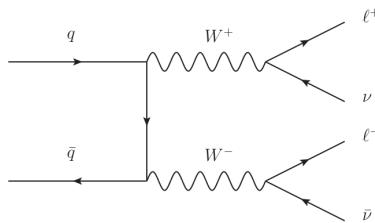
In the opposite-flavour final state four different jets categories are defined: the 0-jet, the 1-jet, the 2-jet and the VBF, Sec. 5.7. In the same-flavour final state only the VBF category is considered. Indeed, only the VBF selection cuts are sufficiently tight to reduce the overwhelming Z plus jets background to a manageable level, Sec. 5.8. The jets categories improve the sensitivity of the analysis, because each category has different contributions from signal production modes and backgrounds.

The signal is interpreted in terms of the EWK singlet and MSSM models as described in Sec 1.3. The Higgs boson width and lineshape is reweighted at generator level according to the parameters defined in the model. The interference effects between the signal produced via gluon-gluon fusion, the WW background also from gluon-gluon fusion, and SM Higgs boson, are expected to change the shape of the signal distribution and have been fully taken into account. A similar treatment is also applied for the interference between high mass signal produced via VBF, the WW plus two quarks background (emerging from the same initial state) and the SM Higgs generated with VBF production mechanism. In general, the interference becomes more and more important as the mass of X increase and it is studied in detail in Sec 7.2. Finally, the interference between the $W^+W^- \rightarrow 2\ell 2\nu$ and $ZZ \rightarrow 2\ell 2\nu$ is negligible due to the different phase space characteristic of these processes.

5.2 Main Background processes

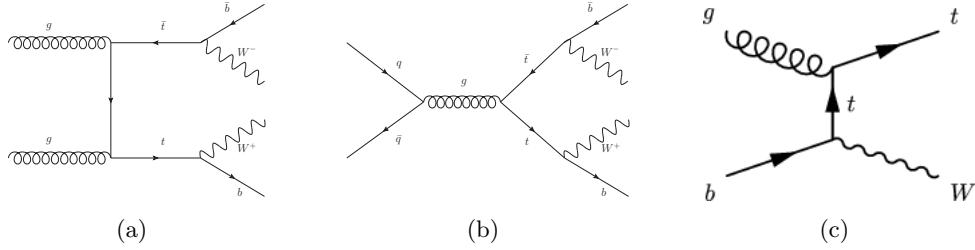
Inside the SM there are several processes that have the same or a similar final state of the signal. The most important background processes contributing to this final state are non resonant $q\bar{q} \rightarrow W^+W^-$, the top production ($t\bar{t}$ and single-top) and the Drell-Yan process. Other backgrounds are due to $W+$ jets events, where a jet is misidentified as a lepton, and multibosons events. All these processes have been simulated with Monte Carlo generators and the simulation details have been discussed in Sec 3.9. A detailed description of these background processes is given below:

- *Non-resonant WW ($q\bar{q} \rightarrow W^+W^-$):* this background is characterized by a final state identical to the signal, however the lepton kinematics for signal and $q\bar{q} \rightarrow W^+W^-$ processes is rather different. For the signal process, the W bosons originate from a spin-0 particle decay and their spins must therefore be antiparallel, implying that the charged leptons produced in their decays appear preferentially in the same hemisphere [72]. In contrast, there is no preferential spin direction in the background case. For this reason the azimuthal angle difference between the two leptons is on average smaller for signal than for background, resulting in a smaller dilepton invariant mass in the former case. The most relevant Feynman diagram of the process is shown below.

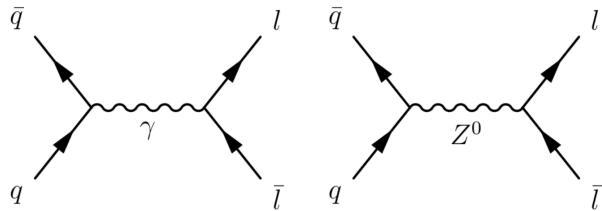


- *Top ($t\bar{t}$ and single-top):* the $t\bar{t}$ events give a signal-like signature if both decays $t \rightarrow Wb$ are followed by $W \rightarrow \ell\nu$. In such a case, in fact, there are two leptons and missing transverse energy, plus two jets (from the hadronization of the b

quark) in the final state. This process is especially important when the signal is produced via VBS or when the signal is associated with jets coming from initial or final state radiation. The single-top production is characterized by the presence of a W boson and a top quark, so after the top decay, again by two W 's, but only one b-jet. Following, some examples of Feynman diagrams for the top background: the $t\bar{t}$ in (a) and (b) diagrams, the single-top in (c).



- *Drell-Yan*: the Drell-Yan process is defined as the annihilation of a quark-antiquark pair into a lepton-antilepton pair. This process is described at leading order by the two shown Feynman diagrams. This kind of background is particularly important for the same flavour final state of the signal



- *W+jet*: this background is characterized by a W boson, decaying in $\ell\nu$, produced in association with a jet. If a fake lepton arise from the misidentified jet, these events have the same final state of the signal (two leptons and missing-transverse-energy).
- *Other*: other background processes involve multibosons production, such as $WZ/W\gamma^*$, ZZ^* and $Z\gamma$ with γ conversion.

The main background processes, the WW production and the top production, are estimated using data. Instrumental backgrounds arising from non-prompt leptons in $W+jets$ production and mis-measurement of E_T^{miss} in Drell-Yan events are also estimated from data. The contribution from $W\gamma^*$ is estimated partly from data. The contribution of other sub-dominant backgrounds is obtained directly from simulated samples. The different data-driven background estimations are explained in the following sections.

5.3 Lepton Efficiencies from Tag and Probe Method

One of the well established data-driven approach for measuring the particle efficiencies is the so called Tag and Probe method. The Tag and Probe method uses a known mass resonance (e.g. J/Ψ , Z) to select particles of the desired type, and probe the efficiency of a particular selection criterion on these particles. In general the “tag” is an object that passes a set of very tight selection criteria designed to isolate the required particle type. Tags are often referred to as a “golden” electrons or muons and the fake rate for passing tag selection criteria should be very small. A generic set of the desired particle type (i.e. with potentially very loose selection criteria) known as “probes” is selected by pairing these objects with tags such that the invariant mass of the combination is consistent with the mass of the resonance. Combinatorial backgrounds may be eliminated through any of a variety of background subtraction methods such as fitting, or sideband subtraction. The definition of the probe objects depend on the specifics of the selection criterion being examined. The simple expression to get the efficiency is given below:

$$\epsilon = \frac{N_{Pass}^{Probes}}{N_{Pass}^{Probes} + N_{Fail}^{Probes}} \quad (5.1)$$

Electrons

The Tag and Probe is used here to get the identification and isolation efficiency of electrons. In this case, the Tag is a well identified and isolated electron which also passes an electron trigger. Once the Tag electron is selected then another object that pass the kinematic electron selection, Tab. 4.1, is searched for. The invariant mass of the Tag and the Probe electron pair is reconstructed and must be in window around the Z boson mass. After that, to compute the efficiency, the Probe electron is required to pass the identification tight working point. This procedure is performed for the data and the MC samples. Once data and MC efficiencies have been calculated, the electron scale factors are estimated as the ratio among the data and MC efficiencies. These scale factors are calculated as a function of p_T and η and used to correct the difference in efficiencies between data and MC in the analysis.

The electron efficiency is about 95%, on the trigger plateau.

Muons

The muon identification and isolation efficiency is also studied and compared to the prediction of MC in order to understand if a correction is needed. The muon efficiency is obtained as

$$\epsilon_\mu = \epsilon_{TRK} \times \epsilon_{Tight} \times \epsilon_{ISO\ Tight}, \quad (5.2)$$

where ϵ_{TRK} is the tracker muon efficiency. The ϵ_{Tight} is muon efficiency under the assumption that the muon passes the kinematic selections summarized in Tab. 4.2. The $\epsilon_{ISO\ Tight}$ is the efficiency of the isolation criteria under the assumption that the muon passes the same selection. The ϵ_{Tight} and $\epsilon_{ISO\ Tight}$ are determined using

Trigger name	HLT Threshold
SingleElectron	$p_T > 27 \text{ GeV}$
SingleMuon	$p_T > 22 \text{ GeV}$
MuonElectron	$p_T > 17 \text{ GeV}$ and 8 GeV
DoubleMuon	$p_T > 17 \text{ GeV}$ and 8 GeV
DoubleElectron	$p_T > 23 \text{ GeV}$ and 12 GeV

Table 5.1. Transverse momentum thresholds applied in the lepton triggers at the HLT level. Double set of thresholds indicates the thresholds for each leg of the double lepton triggers.

the Tag and Probe method. The Tag muon is obtained by applying the kinematic selections and the isolated muon trigger with $p_T^\mu > 20 \text{ GeV}$. The Probe muon is requested to pass the kinematic selection and the isolation criteria. The efficiency value on the plateau is around 93%-99%.

5.4 Data sample and Trigger used

Data recorded in proton proton collisions at 13 TeV during all 2016 was used in the analysis, with a total integrated luminosity of 35.9 fb^{-1} . All runs are taken with a 25 ns LHC filling scheme and recorded in seven different periods, namely Run2016C, Run2016D, Run2016E, Run2016F, Run2016G, and Run2016H.

Events are required to fire one of the unprescaled single-electron, single-muon or muon-electron triggers. Due to the rather high LHC instantaneous luminosity the single-lepton triggers must have high HLT p_T thresholds, otherwise the rate of these triggers would be too large to be sustained. The double-lepton triggers allow the p_T thresholds to be lowered while keeping a sustainable trigger rate. The triggers used in this analysis are summarized in Table 5.1.

The trigger efficiencies are calculated for leptons that pass the identification and isolation criteria. In Fig. 5.1 the trigger efficiency for a gluon-gluon fusion signal with mass 300 GeV is shown for electrons (left) and muons (right). An average trigger efficiency greater than 99% is found, as shown in Figure 5.2. The triggers used in the analysis are summarized in Table 5.1.

5.5 Study of the Interference effects

When a resonance X , with a non negligible width is considered, it is important to take into account also the interference effects both with the WW background, in case of same initial and final state, and with the Higgs boson off-shell tail. The interference is taken into account both when the new signal X is produced via gluon-gluon fusion and when it is produced via vector-boson-fusion: the effects are shown in 5.3 and 5.4 for the two different production mechanisms. The contribution of the interference of the high mass resonance X with the WW background and with the Higgs boson have opposite sign and partially cancel out. This cancellation effect

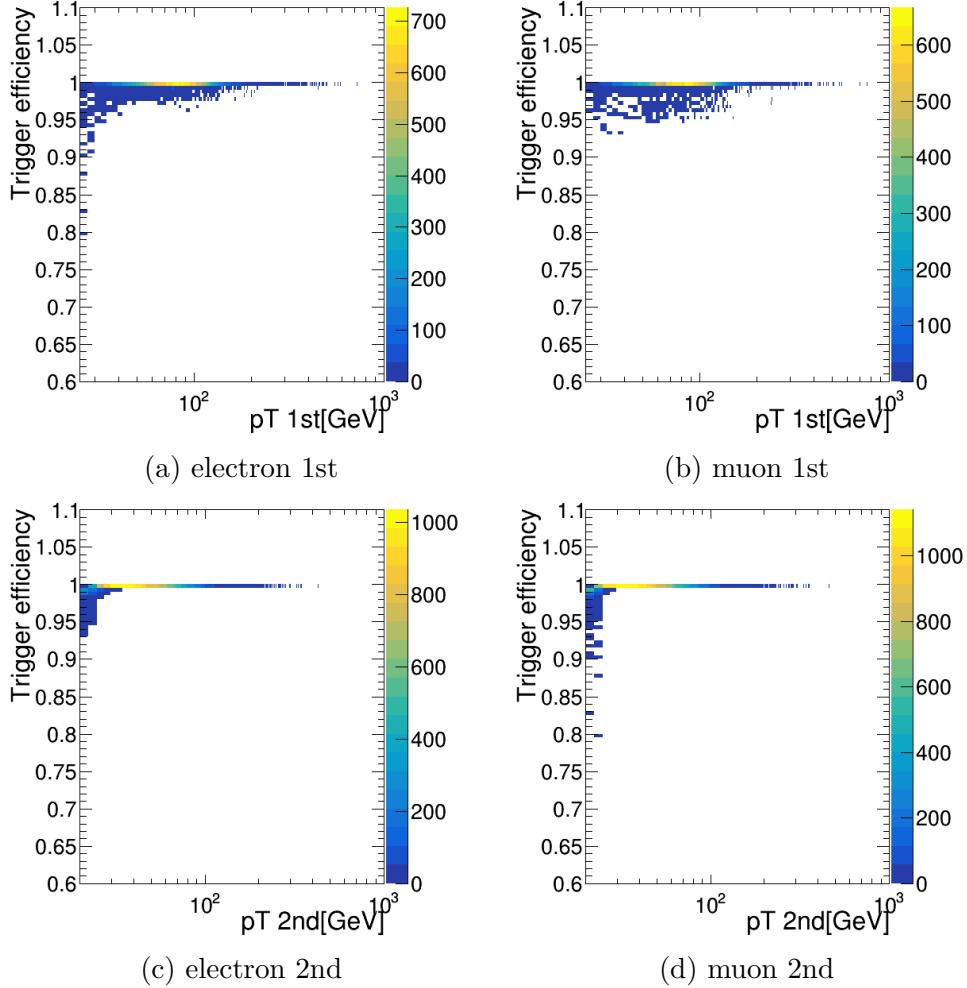


Figure 5.1. Trigger efficiency per event as a function of the lepton p_T for electrons (a) where leading lepton is an electron, and (c) where trailing lepton is an electron, and muons (b) where leading lepton is a muon, and (d) where trailing lepton is a muon, for a gluon fusion 300 GeV MC sample. In this plots the other lepton not shown is integrated. An average trigger efficiency greater than 99% is found.

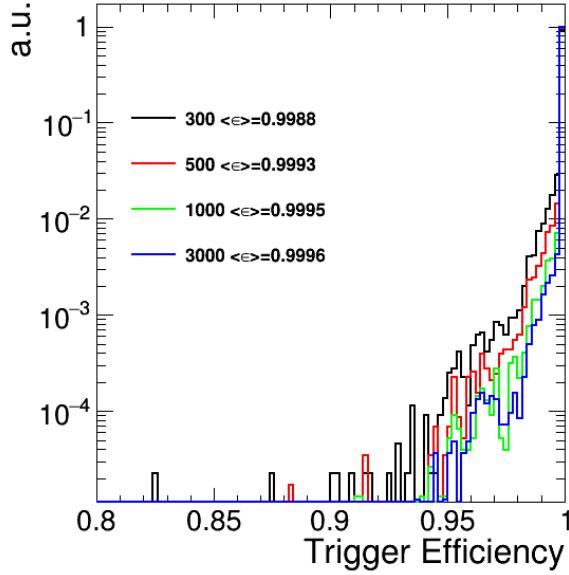


Figure 5.2. Trigger efficiency distribution for four MC samples corresponding to masses of 300, 500, 1000 and 3000 GeV. An average trigger efficiency greater than 99% is found.

is different for different values of the resonance mass. The interference contribution is in general non negligible and it is included in the fit, Sec 7.1.

5.6 Discriminating variable

This analysis is a shape analysis, meaning that after applying selection cuts, the events are not simply counted, but rather an histogram of a discriminating variable is filled from the data and fitted with the sum of signal and background templates, and finally the signal yield is extracted. In principle, the variable with the best discriminating power would be the invariant mass of the four lepton system, however, due to the presence of the neutrinos, it can not be measured. Usually, in the Higgs boson to $WW \rightarrow 2\ell 2\nu$, the variables used in the analysis are:

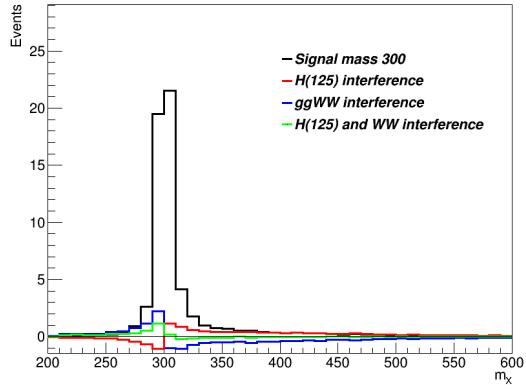
- the transverse mass, m_T^H , defined as,

$$m_T^H = \sqrt{2p_T^{\ell\ell}E_T^{\text{miss}}(1 - \cos\Delta\phi(\ell\ell, \vec{p}_T^{\text{miss}}))} \quad (5.3)$$

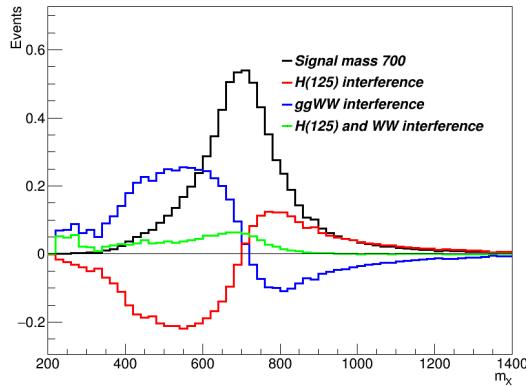
where $\Delta\phi(\ell\ell, \vec{p}_T^{\text{miss}})$ is the azimuthal angle between the dilepton momentum and \vec{p}_T^{miss} ;

- the di-lepton mass, $m_{\ell\ell}$.

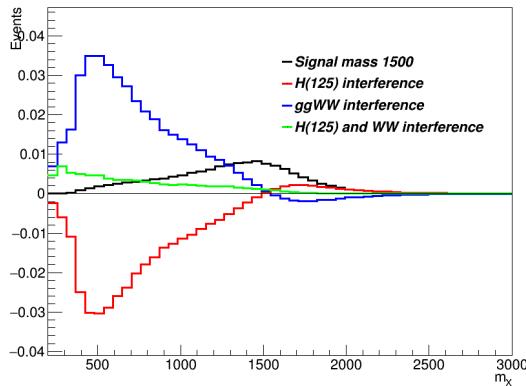
However, both m_T^H and $m_{\ell\ell}$, are not sensitive to different signal mass hypothesis. For this reason a new variable, the visible transverse mass, m_T^I , has been used. This variable, studied for this specific analysis, is defined as the invariant mass of the



(a) Mass 300 GeV.

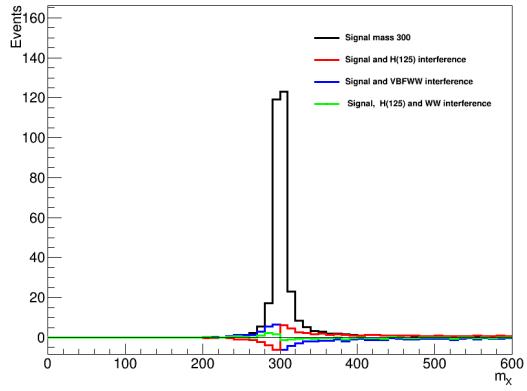


(b) Mass 700 GeV.

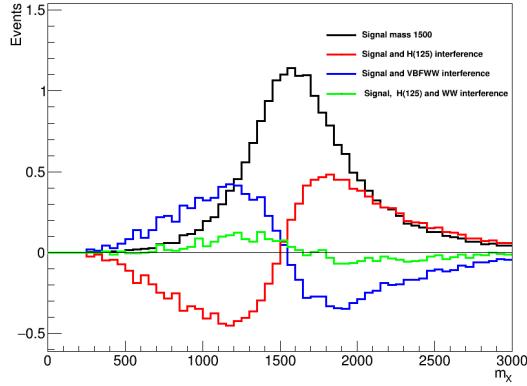


(c) Mass 1500 GeV.

Figure 5.3. Distribution of for the X mass resonance, produced via gluon-gluon fusion for different masses. In black the high mass signal. In red the interference between the high mass signal and the Higgs boson. In blue the interference between the high mass signal and the background. In green the total interference i.e. high mass signal, Higgs bison and background.



(a) Mass 300 GeV



(b) Mass 1500 GeV

Figure 5.4. Distribution of for the X mass resonance, produced via vector-boson-fusion fusion for different masses. In black the high mass signal. In red the interference between the high mass signal and the Higgs boson. In blue the interference between the high mass signal and the background. In green the total interference i.e. high mass signal, Higgs bison and background.

four momentum resulting from the sum of the two leptons four-momenta and the missing four-momentum:

$$m_T^I = \sqrt{(p_{\ell\ell} + E_T^{\text{miss}})^2 - (\vec{p}_{\ell\ell} + \vec{p}_T^{\text{miss}})^2}. \quad (5.4)$$

The distribution of all the variables defined above can be compared in Fig. 5.5, where it is visible the better power of m_T^I in discriminating different mass hypotheses with respect to m_T^H and $m_{\ell\ell}$. The usage of m_T^I also provides a good discriminating power between signal and background.

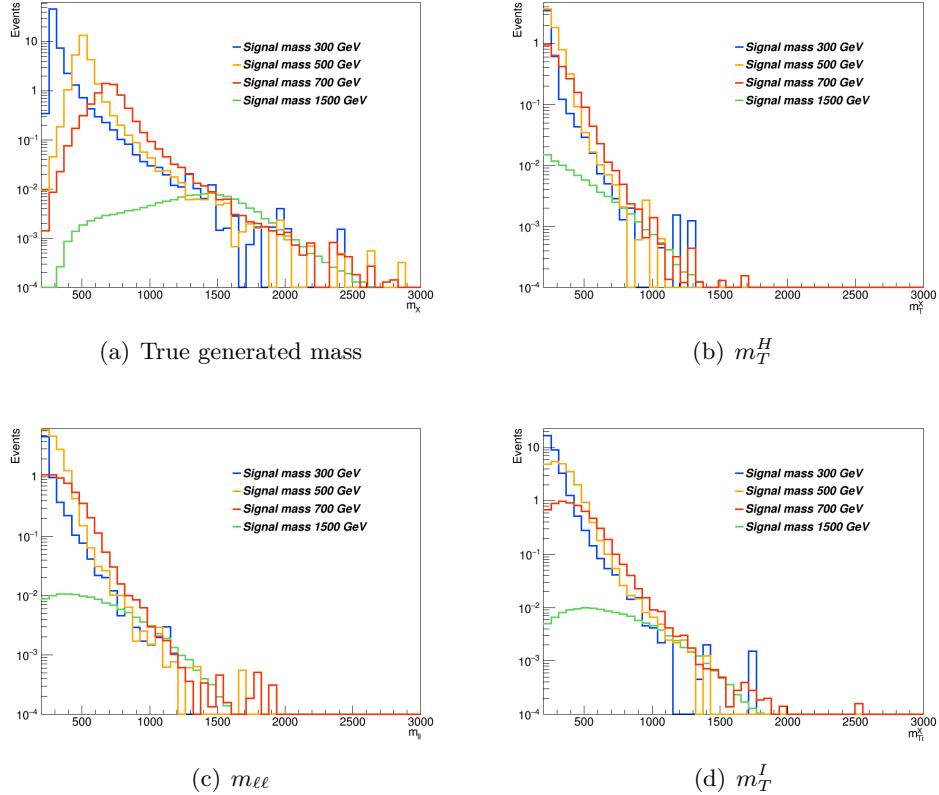


Figure 5.5. Distributions of the generated mass (no possible reconstruction), m_T^H , $m_{\ell\ell}$ and m_T^I variables for different X mass hypothesis. It is clear that the most discriminating variable is m_T^I .

5.7 Opposite Flavor final state

General overview

The events are subdivided in signal and control regions. The signal events are expected to fill mainly to first one. Instead the second one are expected to be fill up by backgrounds events and they are used to normalized the backgrounds shape.

To avoid a possible bias of the experimenter in the process of developing the analysis strategy and optimizing the selection requirements, firstly a blind analysis has been performed by defining a background-only sideband control region and only looking at data events falling in such a region. Precisely, the control region is defined according to selection criteria that exclude the signal events. Data in the signal region, i.e. were not examined until the event selection criteria were settled, only relying on Monte Carlo simulations for what concerns signal. In this way, the analysis strategy has been decided looking at the simulated events. Only when data distributions in signal-depleted regions confirmed the robustness of the analysis, data in signal region were inspected. For these reasons, the agreement between data and Monte Carlo was checked on a sideband region.

Signal region

The events are requested to pass single or double lepton triggers, and exactly one electron and one muon are requested to be reconstructed in the event. One of the two leptons is requested to have a p_T greater than 25 GeV, the other is requested to have p_T greater than 20 GeV and both leptons are requested to be well identified and isolated, to reject non-prompt leptons and leptons coming from QCD sources. To suppress background processes with three or more leptons in the final state, such as ZZ, WZ, Z γ , W γ or triboson production, no additional identified and isolated lepton with $p_T > 10$ GeV should be reconstructed. The low dilepton invariant mass region dominated by QCD production of leptons is not considered in the analysis and $m_{\ell\ell}$ is requested to be higher than 50 GeV to reduce the SM Higgs boson ($m_H=125$ GeV) contamination. A moderate cut on the missing-transverse-energy is applied ($\text{MET} > 20$ GeV) due to the presence of neutrinos in the final state. Since a high mass signal is searched a cut on $m_T^I > 100$ GeV is applied. A cut on the transverse momentum ($p_T^{\ell\ell} > 30$ GeV) and on the $m_T^H > 60$ GeV are applied against $DY \rightarrow \tau\tau$ background. Finally, against the top background, all jets above 20 GeV are requested not to be identified as b-jets according to the cMVAv2 tagger, loose WP. The full selection, defined as the “WW opposite-flavour selection”, is summarized:

- Two isolated leptons with different charge and flavor ($\mu^\pm e^\mp$);
- p_T of the leading lepton > 25 GeV;
- p_T of the trailing lepton > 20 GeV;
- Third lepton veto: veto events if a third lepton with $p_T > 10$ GeV;
- $m_{\ell\ell} > 50$ GeV, to reduce H(125) contamination;

- MET > 20 GeV;
- $m_T^I > 100$ GeV;
- $p_T^{\ell\ell} > 30$ GeV;
- $m_T^H > 60$ GeV;
- no b-tagged (cMVAv2 loose WP) jets with $p_T > 20$ GeV;

Events passing the “WW opposite-flavour selection” are categorized according to the jet multiplicity, counting jets above 30 GeV, to enhance the sensitivity, especially against the top background.

- **0 jet**, no jets are required in the event;
- **1 jet**, exactly 1 jet is required in the event;
- **2 jet**, exactly 2 jets are required in the event and in addition the condition $\Delta\eta_{jj} < 3.5$ **or** $m_{jj} < 500$ GeV;
- **VBF**, exactly 2 jets are required in the event and in addition the condition $\Delta\eta_{jj} > 3.5$ **and** $m_{jj} > 500$ GeV;

where the 2 jet and VBF regions are mutually exclusive by construction. To extract high mass boson signals in these four categories, the strategy is followed: the m_T^I distribution is fitted as the sum of signal and background templates. Different binnings have been chosen for the m_T^I distributions in the different categories. The binning was chosen to have at least 10 top Monte Carlo events in each bin of the template. The chosen bins are:

- **0/1/2 jet**, [100,150,200,250,300,350,400,450,500,550,600,650,700,750,800,900,1000,2000]
- **VBF**, [100,150,200,250,300,350,400,500,700,1000,2000]

where the first number represents the lower edge of the first bin while the other numbers represent the upper edges. The last bin is an overflow bin. The m_T^I distributions for the signal regions are presented in the four categories in Figs. 5.6.

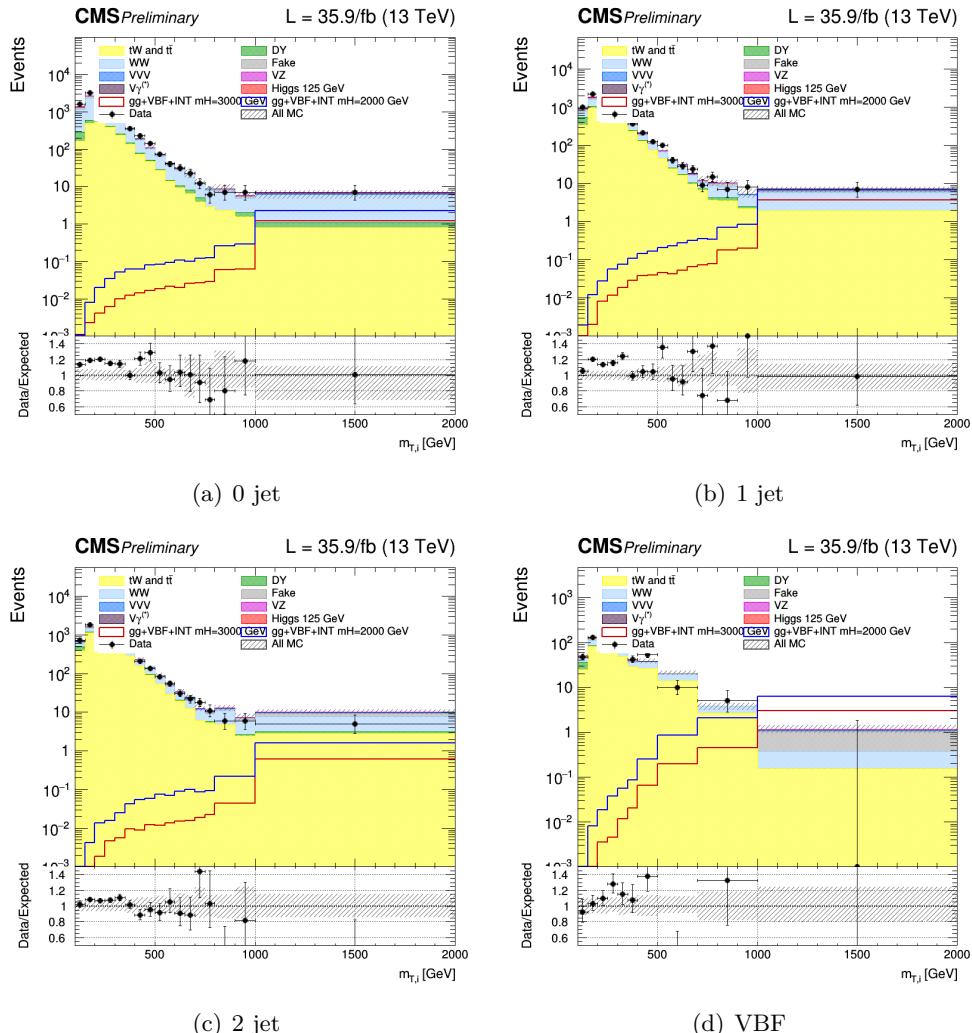


Figure 5.6. Unblinding distributions m_T^I in the signal region for 0, 1, 2 and VBF categories. The signal hypothesis corresponding to m_X of 2 and 3 TeV.

Drell-Yan $\tau\tau$ control region

To normalize the Drell-Yan $\tau\tau$ background to the data, control regions have been defined, as close as possible to the signal region, but enriched in $Z \rightarrow \tau^+\tau^-$. In particular, the “WW OF selection” is used with inverted m_T^H cut, i.e. $m_T^H < 60$. In addition a cut on the invariant mass of the two leptons $50 \text{ GeV} < m_{\ell\ell} < 80 \text{ GeV}$ is requested to exclude possible contribution from non-prompt leptons (low limit) and from $t\bar{t}$ (high limit). For each signal category, a corresponding Drell-Yan $\tau\tau$ control regions is defined. We thus have 4 total Drell-Yan $\tau\tau$ control regions, for 0 jets, 1 jets, 2 jets and VBF. The control plots for several variables in a Drell-Yan enriched phase space for the four jets categories are shown in Figs. 5.7, 5.8, 5.9, 5.10. In general there is a good agreement between data and MC.

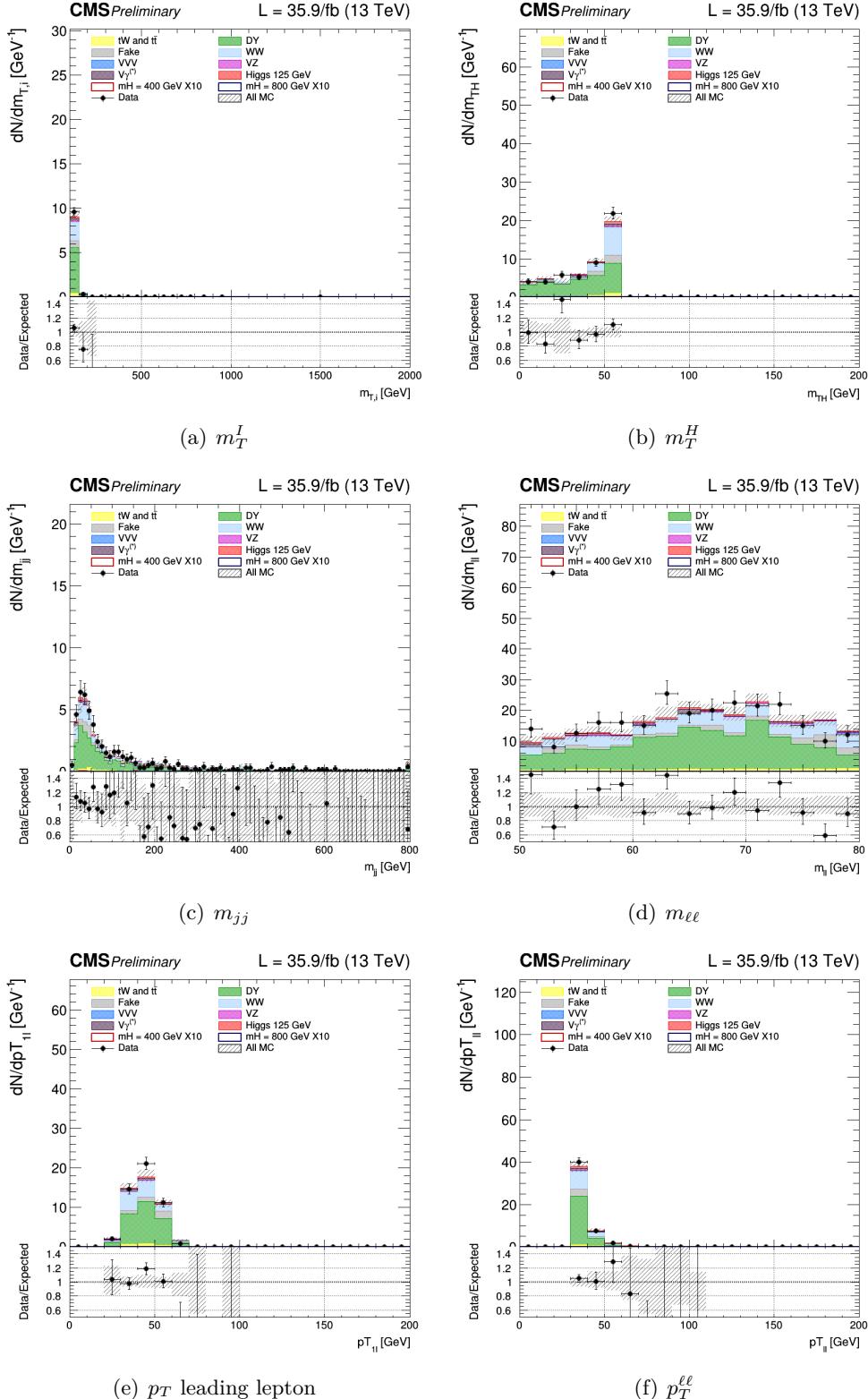


Figure 5.7. Control plots for several variables in a Drell-Yan enriched phase space for events with 0 jet.

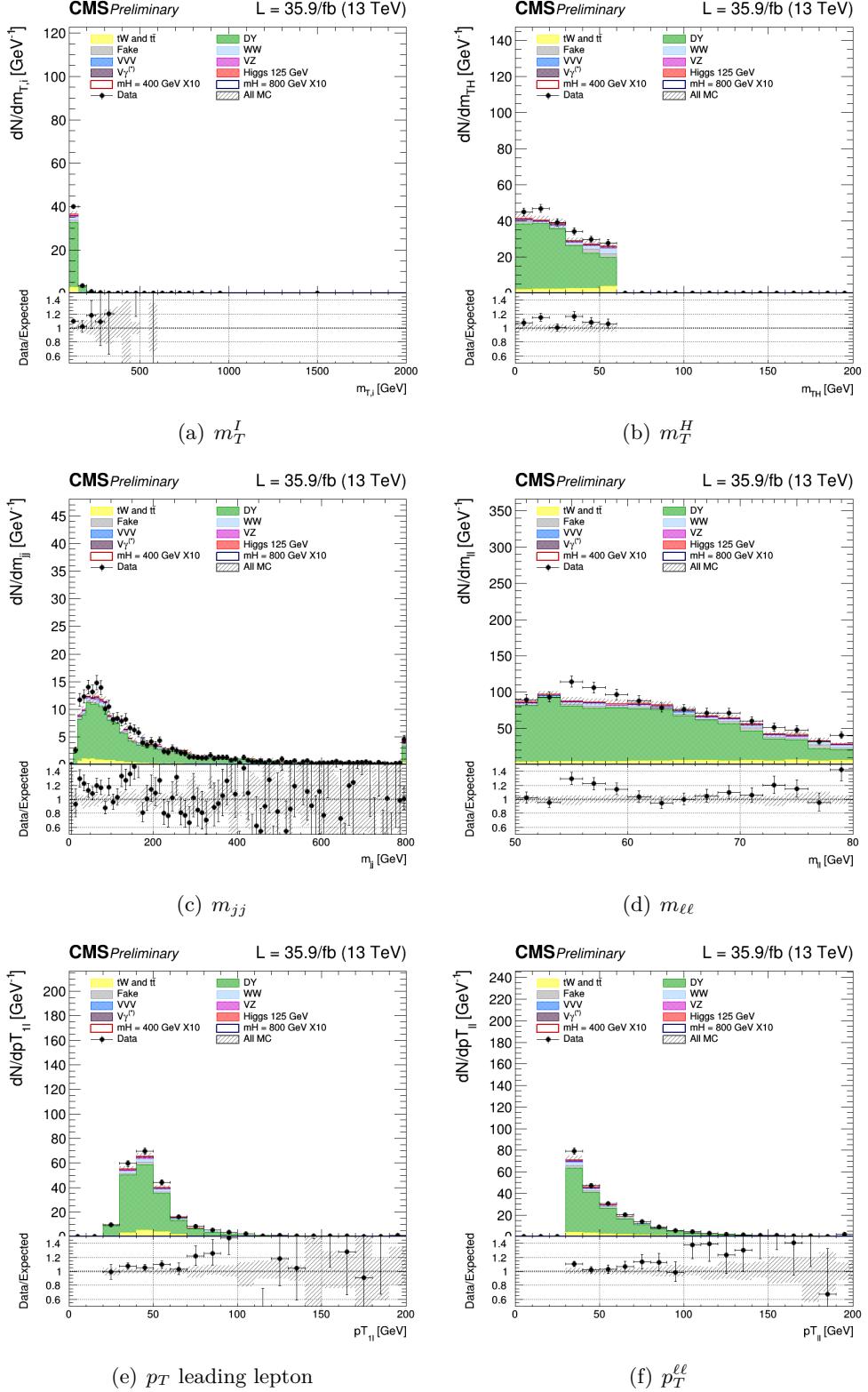


Figure 5.8. Control plots for several variables in a Drell-Yan enriched phase space for events with 1 jet.

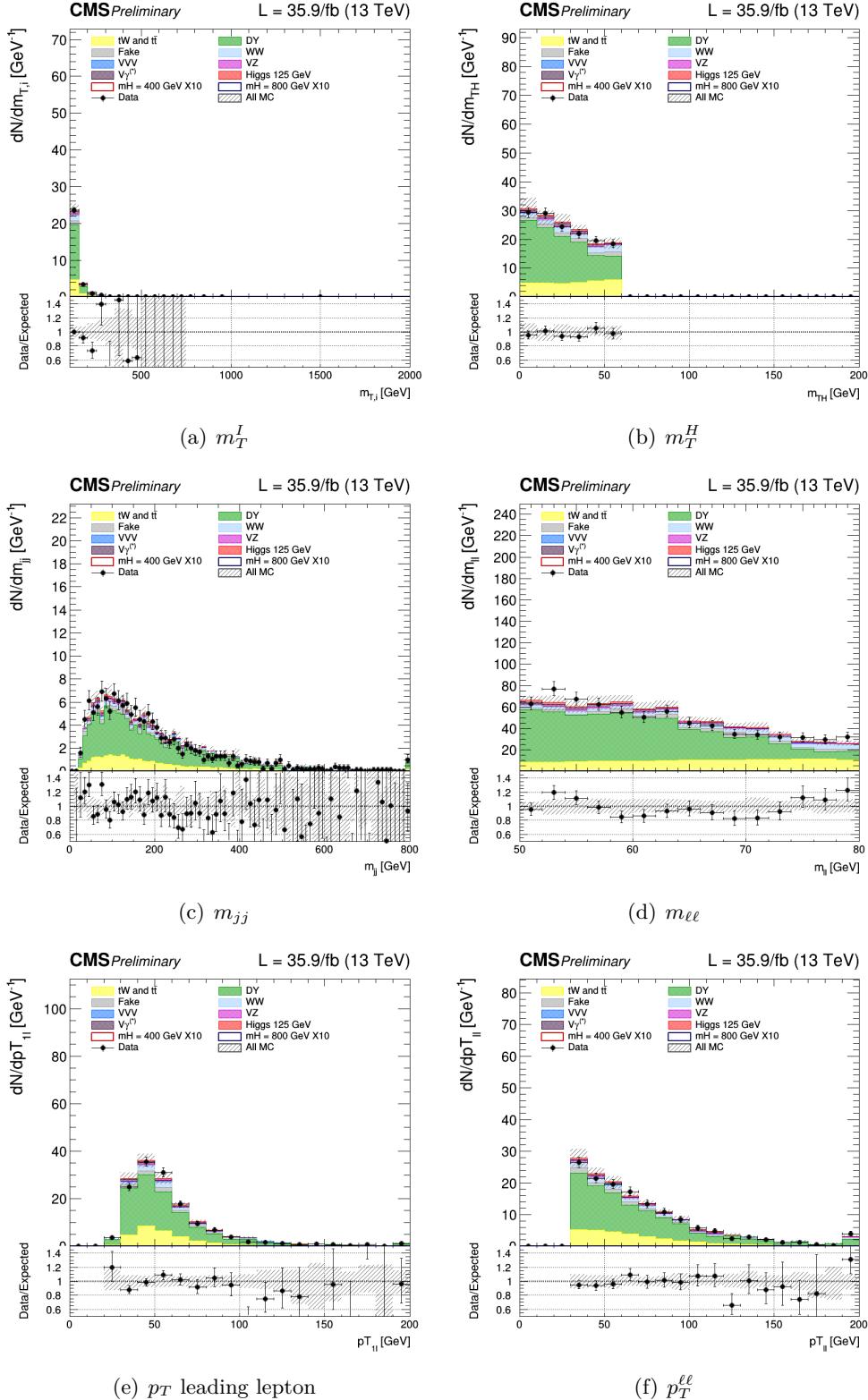


Figure 5.9. Control plots for several variables in a Drell-Yan enriched phase space for events with 2 jet.

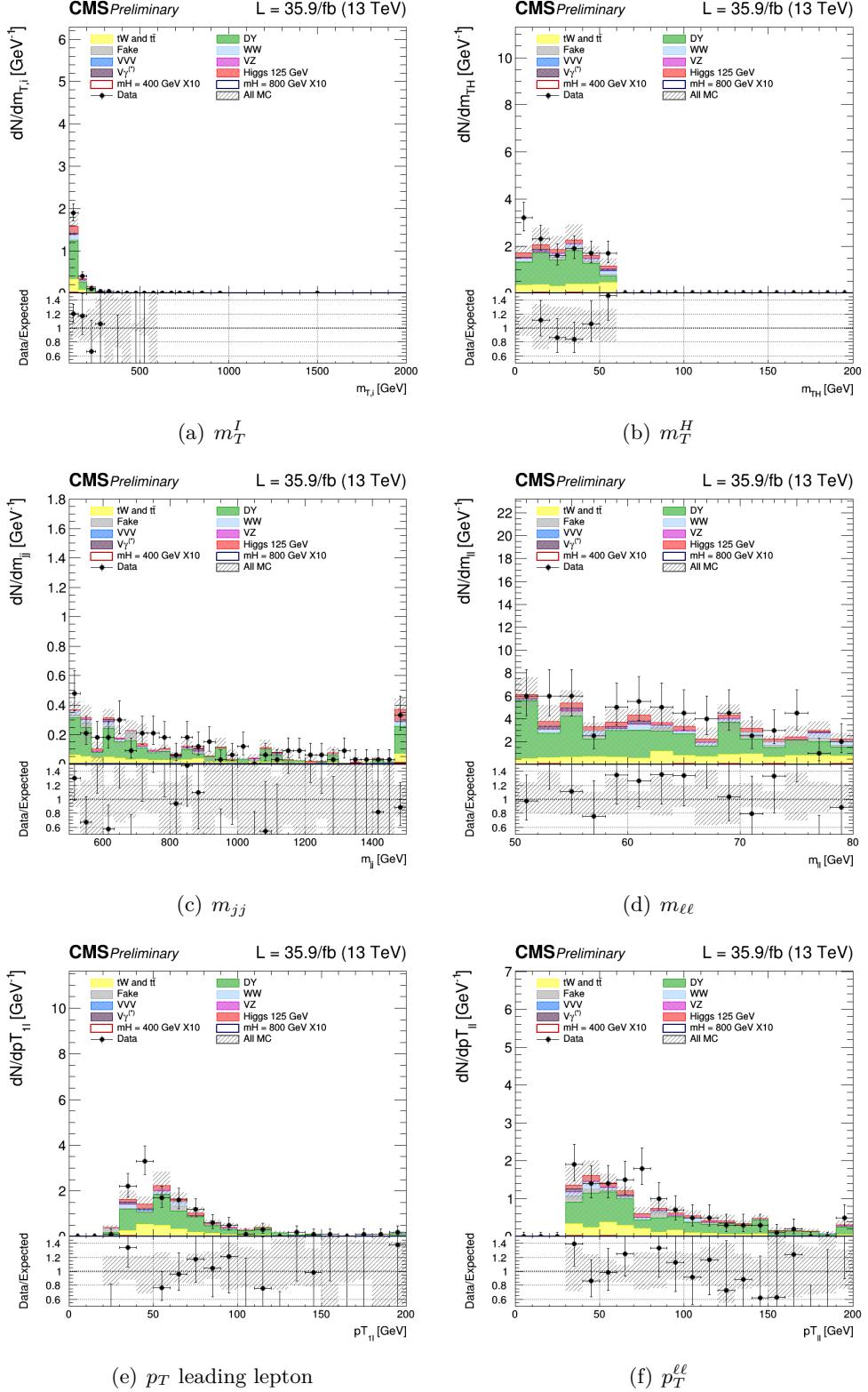


Figure 5.10. Control plots for several variables in a Drell-Yan enriched phase space for events for VBF.

Top control region

Similarly to the Drell-Yan $\tau\tau$ case, control regions are defined for the Top background, and they are used to normalize the top background to data. The “WW OF selection” is used with inversion of the veto on b-jets. In particular the following conditions are imposed to select a top enriched control region for each of the 4 signal regions:

- **0 jet**, at least one b-tagged jet with $20 < p_T < 30$ GeV is required;
- **1 jet**, exactly one b-tagged jet with p_T above 30 GeV is required;
- **2 jet**, exactly 2 jets with at least one of them b-tagged and in addition the condition $\Delta\eta_{jj} < 3.5$ **or** $m_{jj} < 500$ GeV;
- **VBF**, exactly 2 jets with at least one of them b-tagged and in addition the condition $\Delta\eta_{jj} > 3.5$ **and** $m_{jj} > 500$ GeV.

A jet is considered b-tagged if its cMVAv2 score is above the threshold defining the loose working point. The control plots for several variables in a top enriched phase space for events are shown in the Fig. Figs. 5.11, 5.12, 5.13, 5.14. The last bin in the distribution is the overflow.

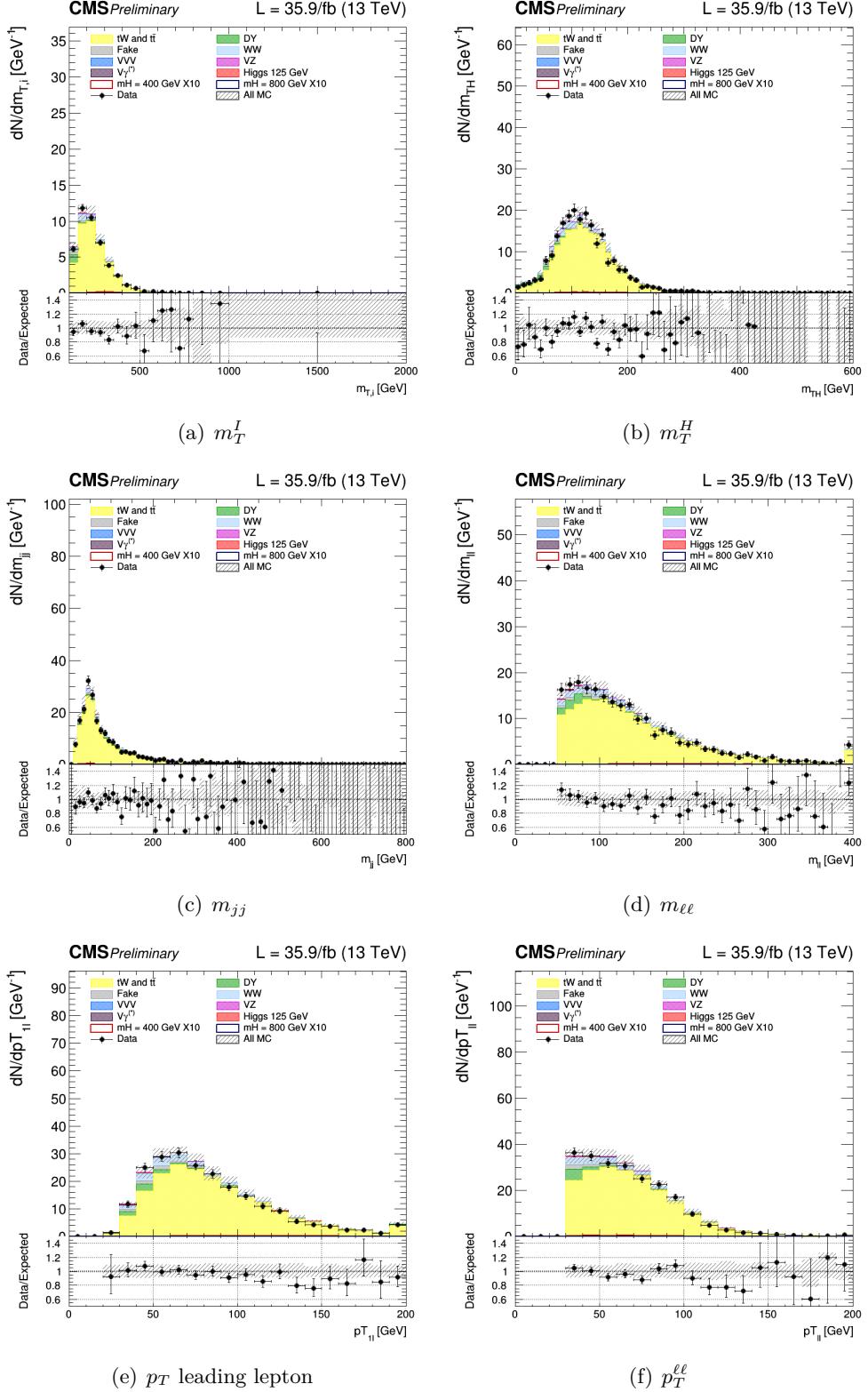


Figure 5.11. Control plots for several variables in the Top enriched phase space for events with 0 jet.

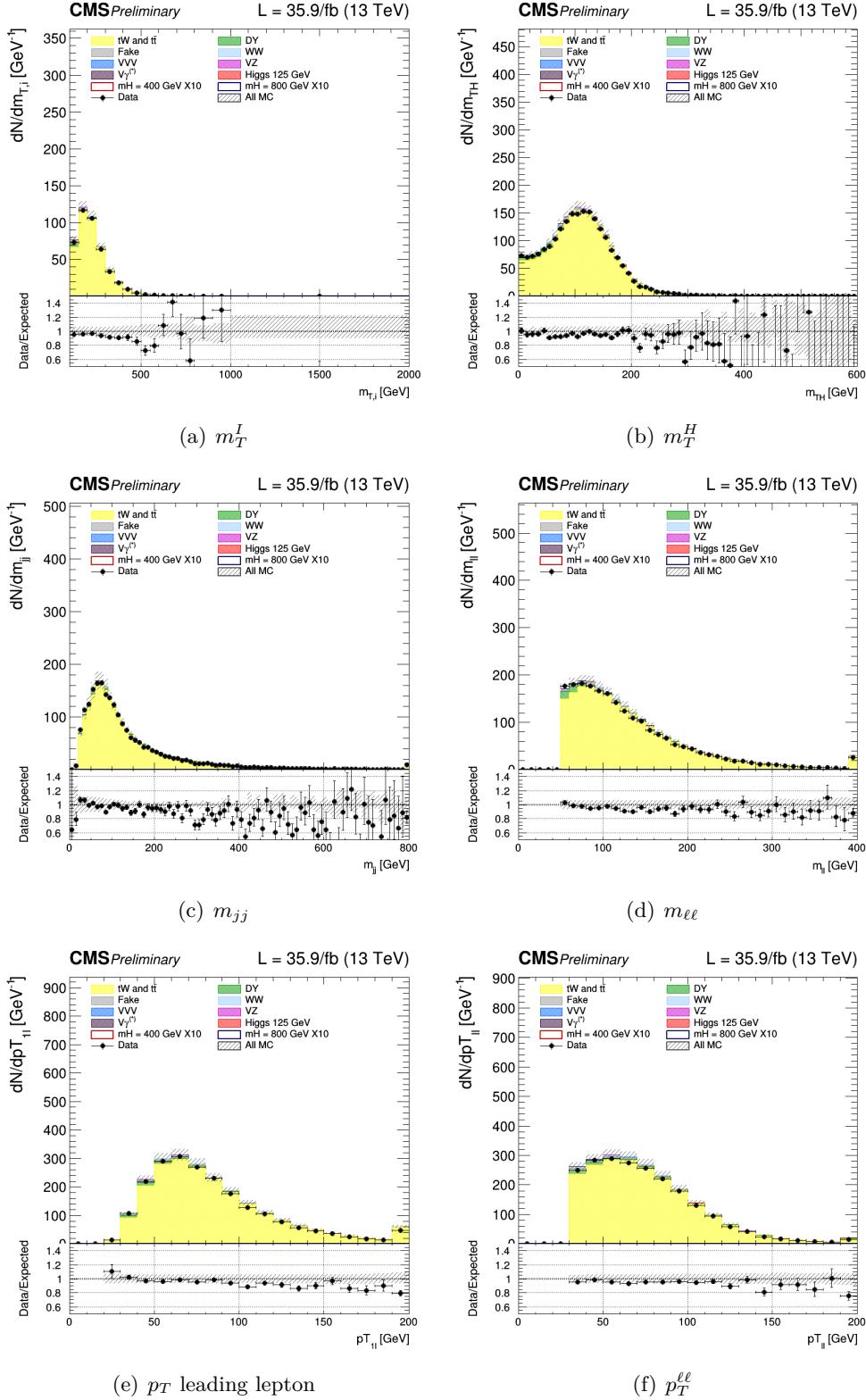


Figure 5.12. Control plots for several variables in the Top enriched phase space for events with 1 jet.

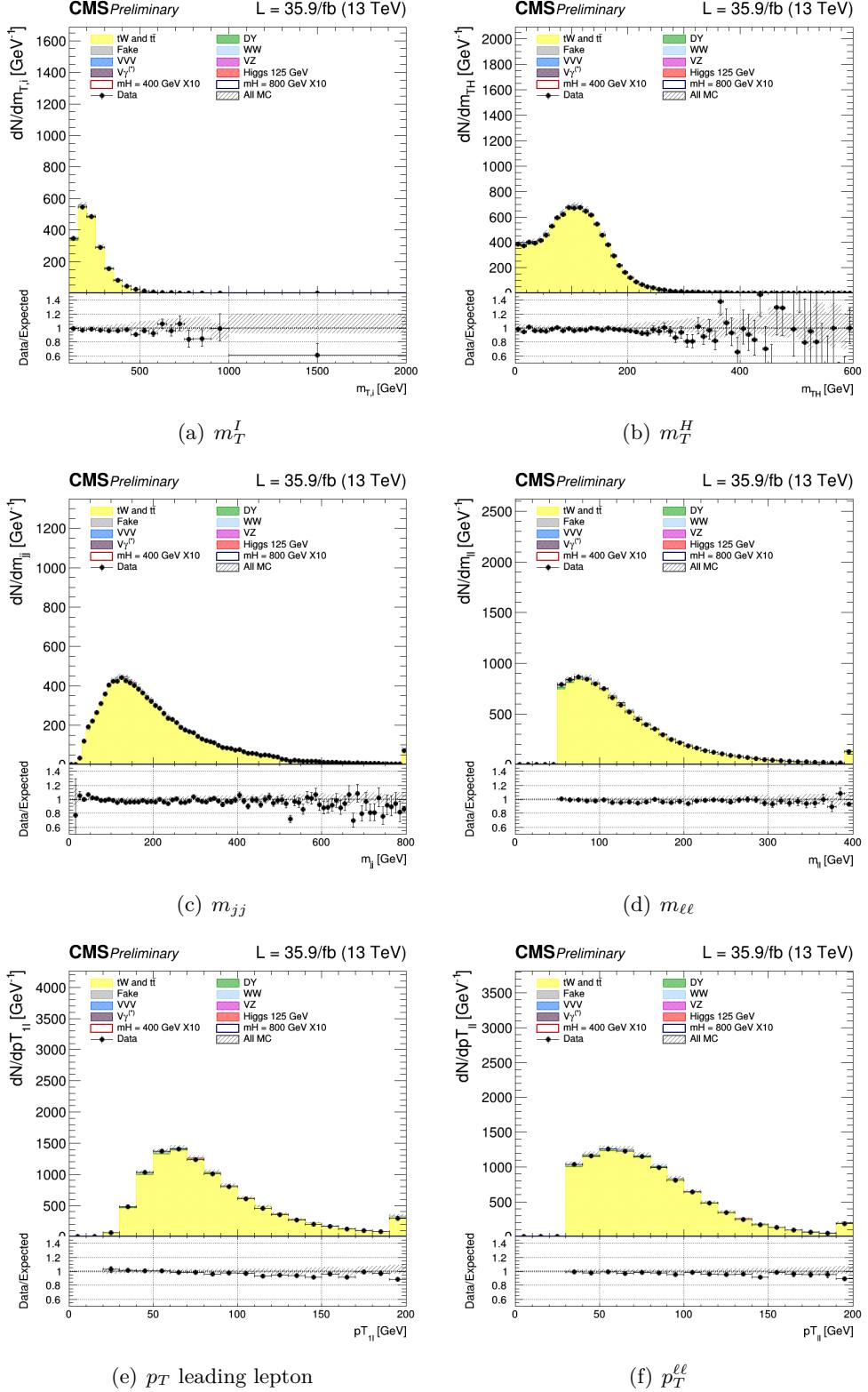


Figure 5.13. Control plots for several variables in the Top enriched phase space for events with 2 jet.

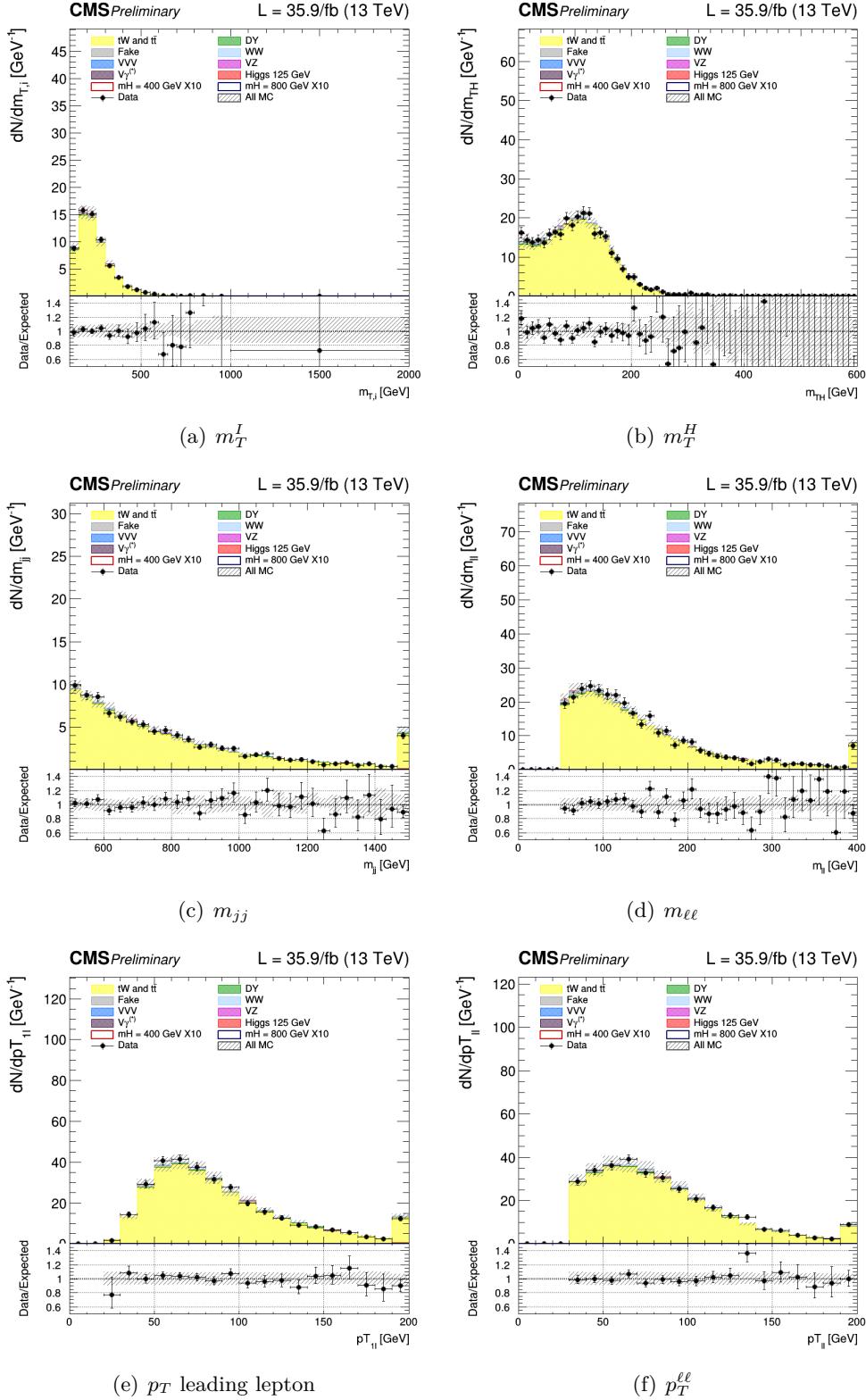


Figure 5.14. Control plots for several variables in the Top enriched phase space for events in VBF region.

5.8 Same Flavor final state

The analysis if the same-flavour final state $W^+W^- \rightarrow \mu^\pm\mu^\mp 2\nu$ and $W^+W^- \rightarrow e^\pm e^\mp 2\nu$ is described.

Signal region

Events are requested to pass single or double lepton triggers and all the physics objects definitions are the same as in the OF analysis. The final state consists of two well identified electrons or two muons with $p_T > 20$ GeV, opposite charge, and large missing transverse energy from the undetected neutrinos.

In addition to the backgrounds described for the OF final state, the background from $DY \rightarrow \mu^+\mu^-$ and $DY \rightarrow e^+e^-$ is very large in this final state. Indeed, due to this very large background, the SF analysis only targets the VBF topology, where the DY background is suppressed by the tight jet requirements. In addition, an invariant mass of the two leptons larger than 120 GeV is requested. The full selection, defined as the “WW same flavour selection”, is :

- Two isolated leptons with same flavor and opposite charge ($\mu^\pm\mu^\mp$ and $e^\pm e^\mp$);
- p_T of the leading and trailing lepton > 20 GeV;
- Third lepton veto: veto events if a third lepton with $p_T > 10$ GeV;
- $m_{\ell\ell} > 120$ GeV
- $p_T^{\ell\ell} > 30$ GeV;
- MET > 50 GeV;
- $m_T^I > 100$ GeV;
- At least 2 jets non b-tagged (according to cMVAv2 loose WP) with $p_T > 30$ GeV.
- $\Delta\eta_{jj} > 3.5$;
- $m_{jj} > 500$ GeV;;

Similarly to the opposite-flavour analysis, the signal is extracted from a template fit of the m_T^I distribution. The m_T^I distributions has the following binning:

- **VBF**, [100,150,200,250,300,350,400,450,500,600,700,1000];

where the first number represents the lower edge of the first bin while the other numbers represent the upper edges. The last bin is an overflow bin. The binning has been chosen in order to have at least 10 expected Top-backgrounds event and at least 10 expected Drell-Yan events in each bin of the template.

The distributions for the signal region of m_T^I variable is shown in Fig. 5.6

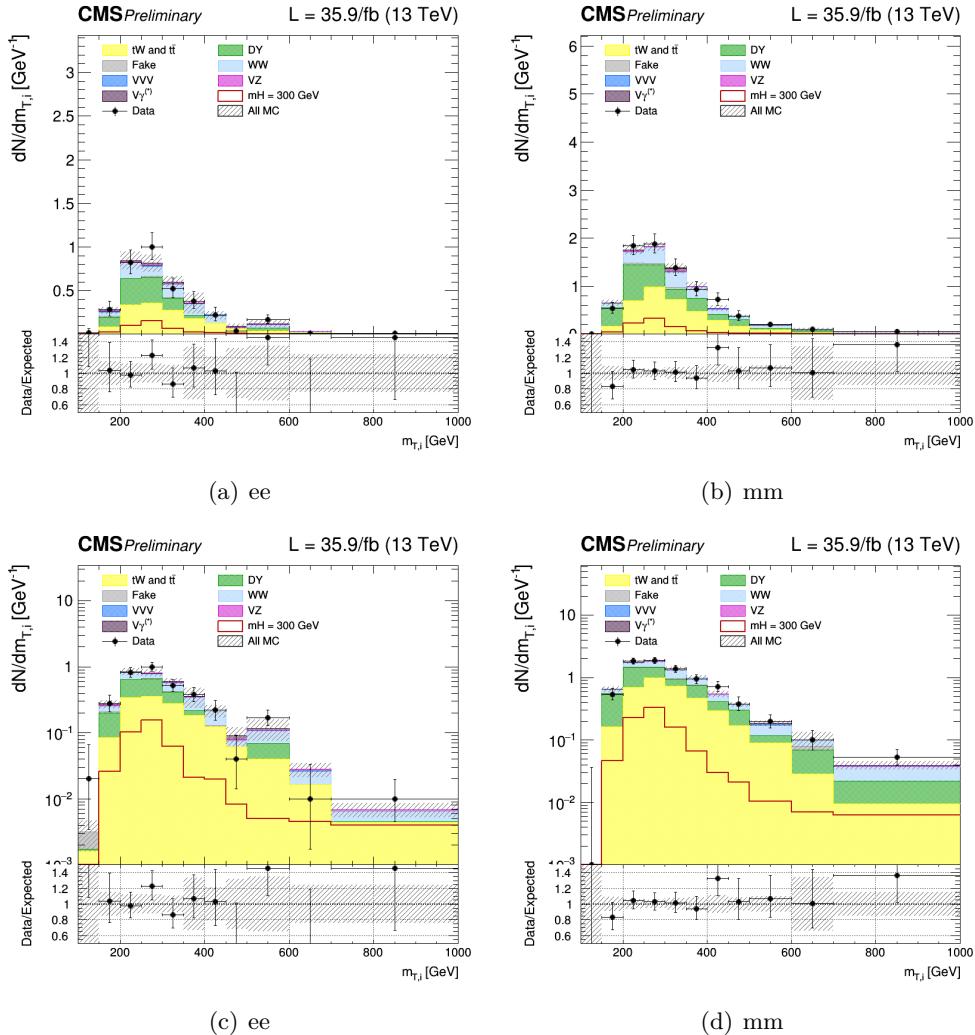


Figure 5.15. Unblinding distributions m_T^I in the signal region for ee and $\mu\mu$ categories in linear and log scale. The signal hypothesis corresponding to m_X of 300 GeV.

Drell-Yan control region

The main background for the SF analysis is the DY. A control region has been defined, as close as possible to the signal one to be used for the normalization of the DY background, separately for electrons and muons. The control region is defined by the “WW same flavour selection”, except for the $m_{\ell\ell}$ requirement which is changed to $70 \text{ GeV} < m_{\ell\ell} < 120 \text{ GeV}$ to include the Z boson. The missing transverse energy distribution in the data shows discrepancies respect to Monte Carlo simulation in ee and $\mu\mu$ Drell-Yan control regions. A correction is applied reweighting all the simulated samples with a weight per event which depends on the MET value. The weight is evaluated as the ratio between data, one subtracted all backgrounds except the DY, and the Drell-Yan itself, in each bins of the distribution, separately for ee and $\mu\mu$ categories. The weight is assumed to be linear as function of the MET value.

This kind of reweighting allows to correct for shape differences between data and MC, , Fig. 5.16. The control plots for several variables in a Drell-Yan enriched phase space for the ee and $\mu\mu$ are shown in Figs. 5.18 for the dielectron case and Figs. 5.19 for the dimuon case. In general there is a good agreement between data and MC.

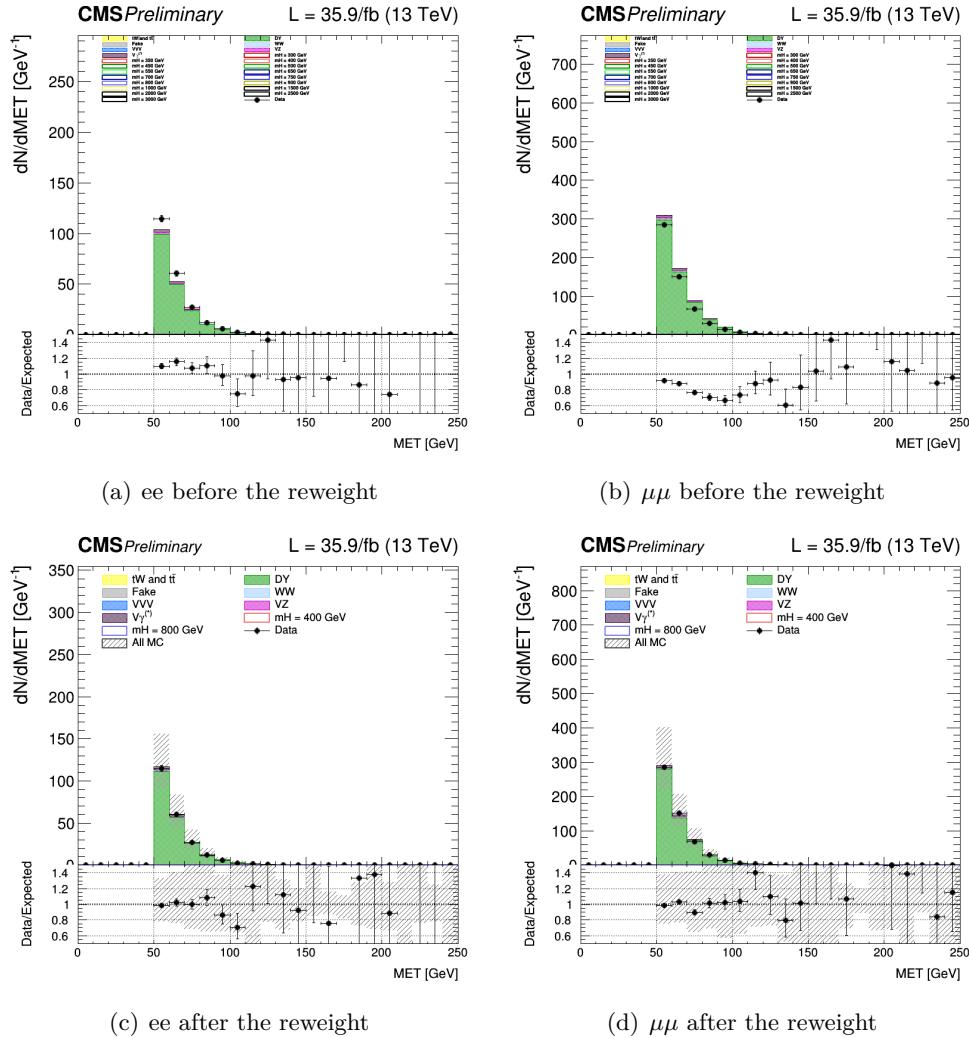


Figure 5.16. MET control plots for Drell-Yan fot ee categories in *a* and for $\mu\mu$ in *b* before the reweight. In *c* and *d* the same distribution after the correction.

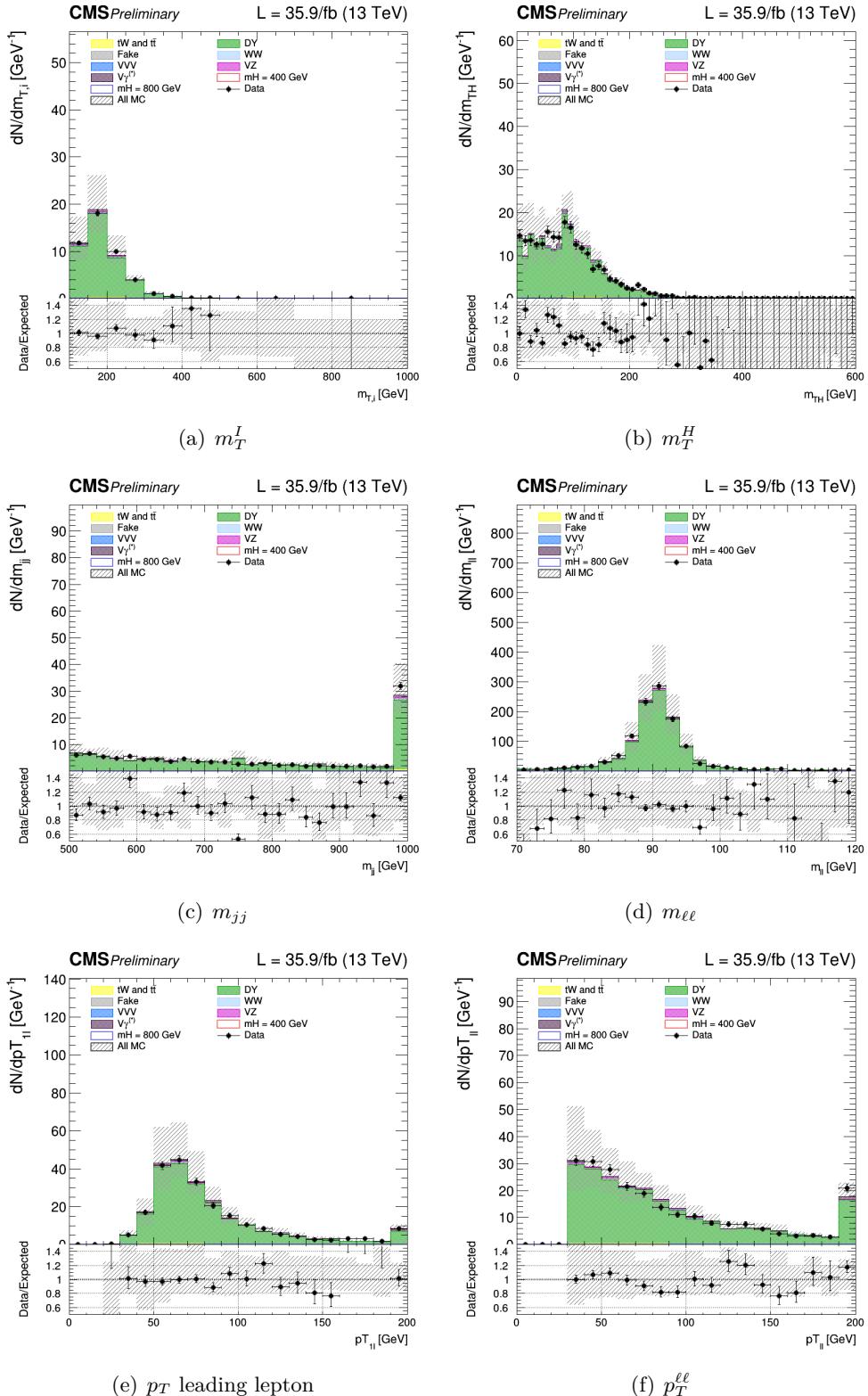


Figure 5.17. Control plots for several variables in a Drell-Yan enriched phase space for ee.

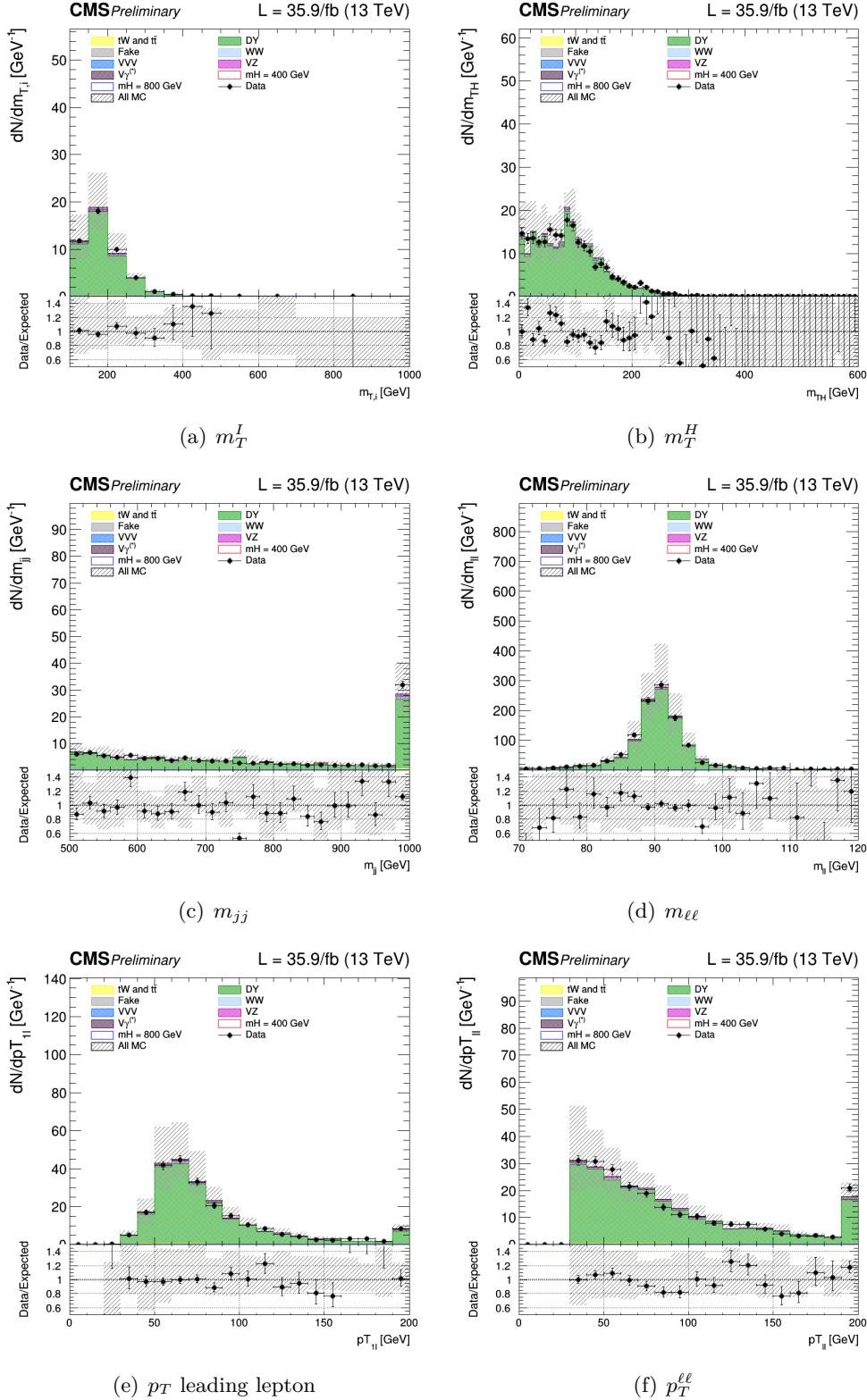


Figure 5.18. Control plots for several variables in a Drell-Yan enriched phase space for ee.

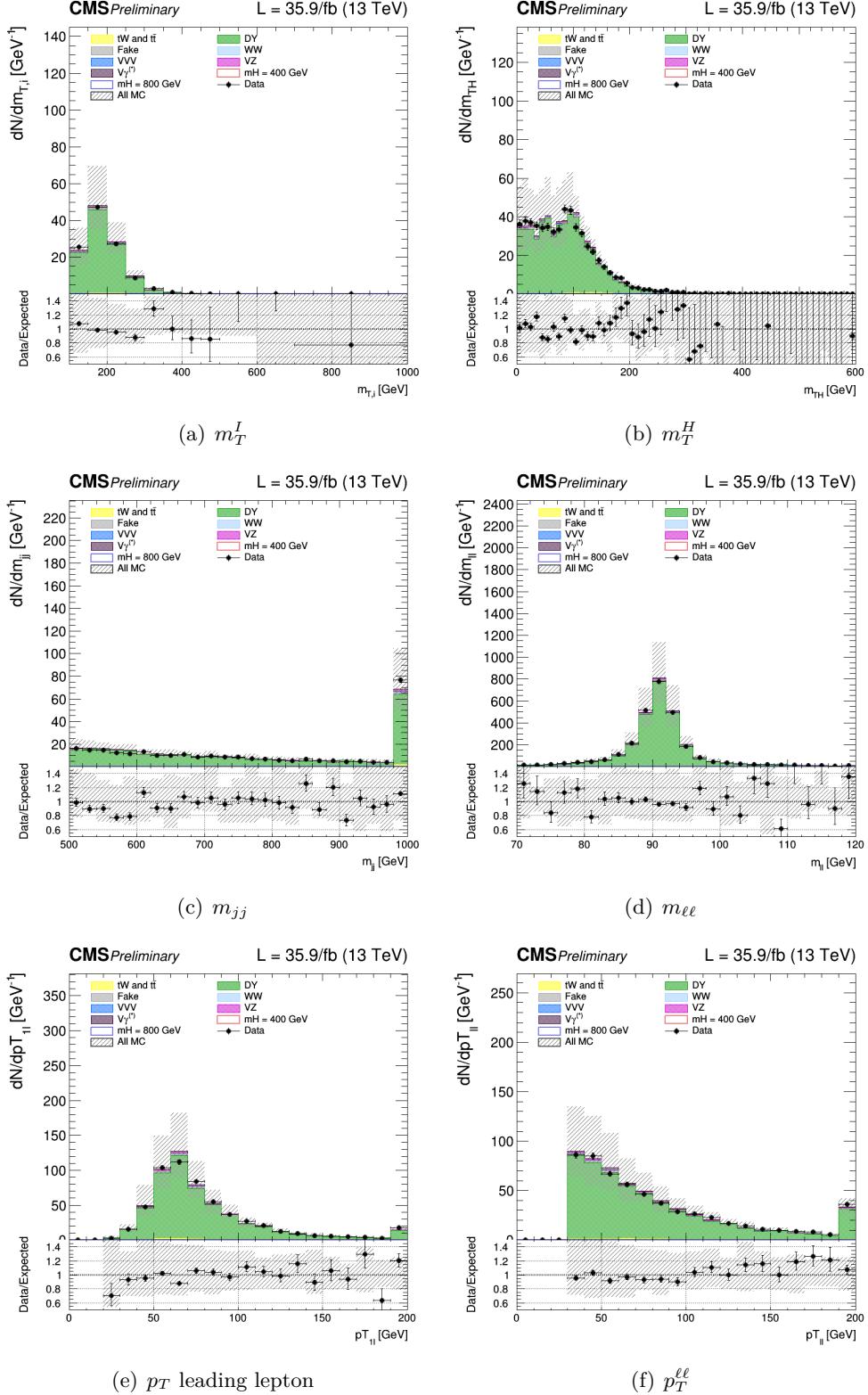


Figure 5.19. Control plots for several variables in a Drell-Yan enriched phase space for $\mu\mu$.

Top control region

A top-enriched control region is defined to normalize the top backgrounds, separately for electrons and muons. The “WW SF selection” is required with the inversion of the b-tagging requirement, i.e. the two jets are both requested to be b-tagged according to cMVAv2 loose WP. The control plots for several variables in a top enriched phase space for events are shown in the Figs. 5.20 for the dielectron case and 5.21 for the dimuon case. Good agreement is observed between data and MC.

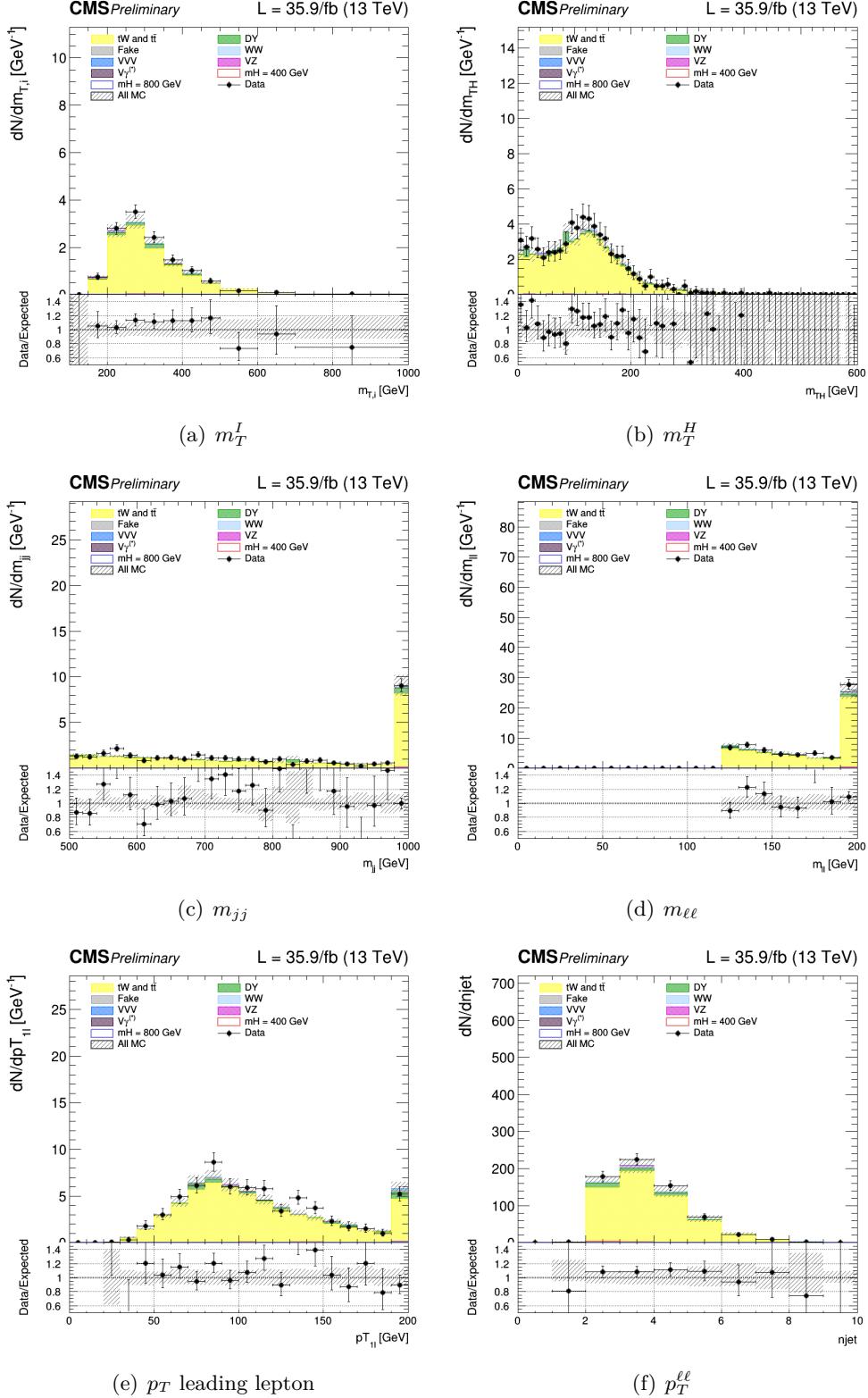


Figure 5.20. Control plots for several variables in a Top enriched phase space for ee.

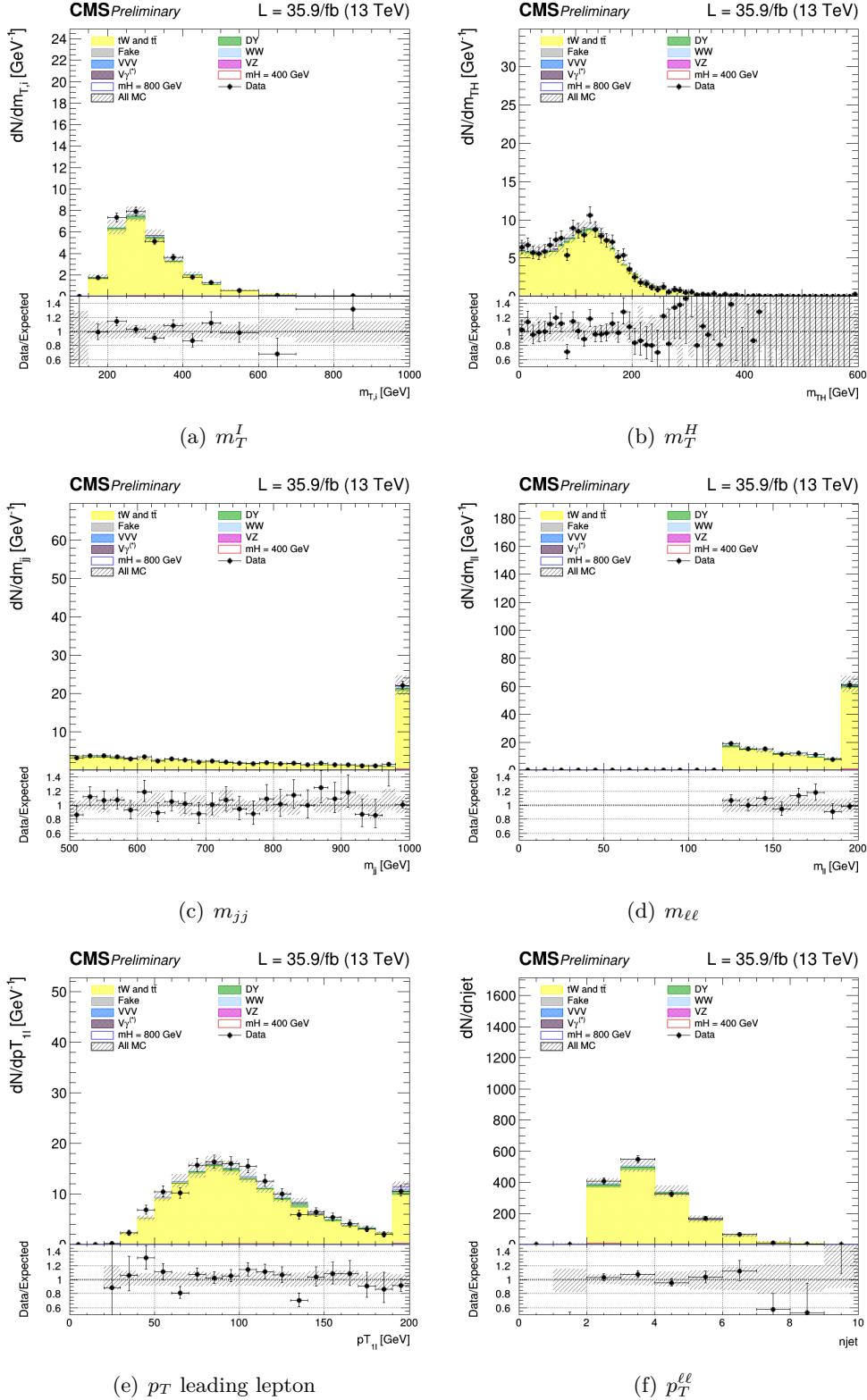


Figure 5.21. Control plots for several variables in a Top enriched phase space for $\mu\mu$.

5.9 Systematic uncertainties

Systematic uncertainties are introduced as nuisance parameters in the fit and can affect the normalization and the shape of the different contributions.

Statistical uncertainties from MC simulated events are taken into account. Systematic uncertainties are represented by individual nuisance parameters with log-normal or shape-based distributions. The uncertainties affect the overall normalization of the signal and backgrounds as well as the shape of the predictions across the distribution of the observables. Correlations between systematic uncertainties in different categories and final states are taken into account. Systematic uncertainties play an especially important role in this analysis where no strong mass peak is expected due to the presence of undetected neutrinos in the final state. Below we describe in detail sources and quantities of systematics in this analysis and their effects on the signal and background processes. A list of the most important background uncertainties is given below.

Background normalization uncertainties

One of the most important sources of systematic uncertainty is the normalization of the backgrounds that are estimated on data control samples whenever is possible. The signal extraction is performed subtracting the estimated backgrounds to the event counts in data. The amount of uncertainty depends on the considered background:

- jet-induced background: normalization and kinematic shapes are derived from a data control region and both normalization and shape systematic uncertainties are considered. A conservative 30% uncertainty on the fake rate is assumed correlated across the different analysis regions. The contribution to the uncertainty in the signal region due to the limited electron statistics in the background enriched control regions is about 10%, while the contribution due to the limited muon statistics 3%.
- WW background: The normalization of the WW background is performed independently in each jet multiplicity via the rateParam feature of combine. A WW electroweak (VBS) sample is used in addition to the standard WW sample in the phase spaces with at least two jets, where its contribution becomes non negligible. The uncertainty in the cross section for this process is evaluated using the variations of the renormalization and factorization QCD scales, as well as the PDF variations, and amounts to 11%.
- $t\bar{t}$ and tW backgrounds: Top events are estimated with b-tagging in data control regions. The two top background enriched control regions are defined as additional categories in the fit while the kinematic shapes are taken from the simulation corrected for the b-tagging discriminant scale factors. The top normalization is correlated between the top control region and the Higgs signal categories separately in the different jet multiplicities, and these normalizations are left unconstrained using the rateParam feature of combine.
- Drell-Yan background: The Drell-Yan background enters the different flavor analysis via the leptonic decays of the τ leptons from $Z\gamma^* \rightarrow \tau\tau$. In the

different flavor analyses the normalization of these background is controlled via the rateParam feature of combine and with a dedicated control region in each jet bin category.

- $W\gamma^*$ background: The kinematic shape of this background is predicted by simulation, normalized to its data-driven estimate, and constrained within the respective uncertainty, which is 25%.
- WZ : The kinematic shapes of this backgrounds are predicted by simulation and normalized to their theoretical predictions in the different and same flavour analysis.
- $Z\gamma^*$: The kinematic shapes of this backgrounds are predicted by simulation and normalized to their theoretical predictions in the different and same flavor analysis.
- ZZ: The kinematic shapes of this backgrounds are predicted by simulation and normalized to their theoretical predictions in the different and same flavor analysis.

Experimental uncertainties

Effects from experimental uncertainties are studied by applying a scaling and/or smearing of certain variables of the physics objects, followed by a subsequent recalculation of all the correlated variables. This is done for MC simulation, to account for possible systematic mismeasurements of the data. All experimental sources except luminosity are treated both as normalization and shape uncertainties. For background with a data-driven normalization estimation, the shape uncertainty is considered only. The following experimental systematic sources have been taken into account.

- Luminosity: The uncertainty determined by the CMS luminosity monitoring is 2.3% for 13 TeV data.
- Lepton trigger systematics: Lepton trigger systematics are of the order of less than 1%. These uncertainties are computed by varying the tag selection as well as the Z window in the tag and probe method used to compute the corresponding scale factors.
- Lepton reconstruction and identification efficiency: The lepton reconstruction and identification efficiencies are measured with the tag and probe method in data. To correct for the difference in the lepton identification efficiencies between data and MC, data/MC scale factors dependent on p_T and η are applied to the MC. The resulting uncertainty in the signal region is 1% for electrons and 2% for muon.
- Muon momentum and electron energy scale: Uncertainties on both the scale and resolution individually amount to 0.6-1% for electrons and 0.2% for muons.

- MET miss modelling: The MET miss measurement is affected by the possible mismeasurement of individual particles addressed above, as well as the additional contributions from the pile-up interactions. The effect of the missing transverse momentum resolution on the event selection is studied by propagating each component of the MET uncertainty to the absolute value and direction of MET.
- Jet energy scale (JES) uncertainties: We estimate this uncertainty applying the official jet uncertainties on the JES and compute the variation of the selection efficiency. JES uncertainty affects the rates in the signal region at the level of 10%.
- b-jet misidentification modelling: The uncertainties on the selection of non-b jets is taken into account by looking at the b-jet misidentification efficiency. The uncertainties on these scale factors need to be taken into account, and are of the order of a few percent.

Theoretical uncertainties

- PDF and higher-order corrections (renormalization and factorization scale): PDF uncertainties and the missing knowledge on higher-order corrections, evaluated by means of scale variation, directly affect the cross section, as well as the acceptance of a simulated process. The uncertainties that arise from using different PDF sets were obtained by reweighting events with different PDF sets.
- Underlying event and parton shower modelling: The underlying event (UE) and parton shower (PS) modelling uncertainties are estimated by comparing samples interfaced with different parton showers (Pythia vs Herwig) and UE tunes
- Single top tW and tt ratio: The ratio between the single top and top pair cross section is varied by the uncertainty on the ratio between their cross sections, calculated considering scale variations,
- A QCD and PDF scales for the signal samples at different masses. The uncertainties are taken from the Yellow Report 3 and the same values are used both for the large width hypothesis and for different values of C' . The effect of QCD and PDF scale uncertainties on the analysis selection has also been taken into account.
- The categorization of events based on jet multiplicity introduces additional uncertainties related to higher order corrections. These uncertainties are associated to the ggH production mode and are evaluated independently following the recipe described in [73] and are 5.6% for the 0-jet and 13% for the 1-jet and 20% for the 2-jet and VBF categories.

The top background shape is estimated from simulation and corrected using a data driven b-tagging scale factor. The normalization is measured in a top quark

enriched control region obtained inverting the b-veto requirement of the signal region. Three control regions are defined, one for each jet bin category. A nuisance parameter is added to take into account the effect of the parton shower uncertainty on the top background.

The DY background shape is also estimated from simulation and analogously to the Top background, the DY normalization is measured with a data driven technique in three control regions enriched in DY events.

A dedicated nuisances for MET reweighting in DY control region is introduced in SF analysis. It is evaluated separately for ee and $\mu\mu$ categories. The uncertainty is quoted as maximum and minimum best-fit lines of the linear fit.

Chapter 6

Semileptonic final state

In this chapter the search in the semileptonic final state is presented. In the considered mass range the hadronically decaying boson may be sufficiently boosted that its decay products are contained in a single jet. Thus merged jet reconstruction and substructure techniques are used, but the case where the hadronic decay products are resolved is also considered. I have contributed in particular to the synchronisation between this analysis and the fully leptonic (in terms of signal modelization, uncertainties, backgrounds estimation) among the fully- and the semi-leptonic analysis.

6.1 Overview of the semileptonic analysis

The $X \rightarrow WW \rightarrow \ell\nu q\bar{q}$ final state is one of the dominant decay channels of a SM-like Higgs boson for masses above 200 GeV, as shown Fig. 1.14 in green. In the mass regime of interest the hadronically decaying boson may be sufficiently boosted that its decay products are contained in a single merged jet. Jet substructure techniques are used to identify merged jets with two well defined subjets and to determine the merged jet mass, helping to discriminate vector boson decays from QCD jets coming from quarks and gluons. If the hadronic decay products are resolved then the boson decay may be reconstructed as two quark-jets (a dijet). In this analysis it is first attempted to reconstruct boson candidates using merged jets, if no boosted candidates are found then dijet reconstruction is attempted. The leptonically decaying boson is reconstructed as a single isolated lepton and missing transverse energy corresponding to the neutrino. An estimate of the neutrino longitudinal momentum is derived by imposing the constraint of the W mass on the invariant mass of the $\ell\nu$ system. The dominant background processes are from $W+jets$ and top production, with a smaller contribution from diboson, $Z+jets$ and QCD multijet events. Unlike for the signal, the mass distribution of the background events is not resonant, providing a useful handle to isolate signal events. Data from signal-free control regions are used to normalize and tune the $W+jets$ and top MC background prediction reducing the dependence on the simulation. The kinematic information of the final state particles is fully exploited with the use of a matrix element based kinematic discriminant to enhance the sensitivity to signal. Identifying additional jets in the event also leads to increased sensitivity to signal produced through vector-boson fusion. Effects due to the interference between the

signal, the SM Higgs boson and the WW continuum background are also considered.

6.2 Jets for Boosted and Resolved selection

Hadronic jets are reconstructed with the anti- k_T clustering algorithm using two different radius parameters, $R = 0.8$ (AK8) and $R = 0.4$ (AK4). The AK8 algorithm is adopted for reconstructing the hadronic W decay in a single merged jet when the decay products are highly collimated, while the AK4 algorithm is used to reconstruct the hadronic W decay when the decay products are resolved. Charged Particle Flow constituents not associated to the primary vertex are not used in the jet clustering procedure. To avoid double counting of the same object reconstructed in different collections, jets are required to be separated from the selected isolated leptons by $\Delta R(lj) > 0.8$ for AK8 jets by $\Delta R(lj) > 0.4$ for AK4 jets. For the selection of W candidates, two AK4 jets both with $p_T > 30$ GeV or an AK8 jet with $p_T > 200$ are required. The final hadronic W candidate requirements for the merged and resolved analyses will be discussed in detail in the following. To reduce the top background, a veto on the events that contain an additional jet with $p_T > 20$ GeV which is b-tagged is applied.

For AK8 jets the softdrop algorithm [74] is implemented to suppress the effects of pileup and underlying event radiation. This technique declusters the jet and recursively removes soft wide-angle radiation, pushing QCD jets towards lower values while maintaining the jet mass for V-jets around the boson-mass. Each AK8 jet has two associated subjets corresponding to the protojets obtained by undoing the last iteration of the soft drop jet declustering procedure.

This analysis also makes use of the n-subjettiness variable [75] to quantify the compatibility of AK8 jets with a substructure hypothesis of n-subjets. It is based on the distribution of the jet constituents with respect to the subjet axes. The ratio of 2-subjettiness and 1-subjettiness, τ_2/τ_1 , allows the dipole structure of hadronic W decays to be distinguished from the monopole structure of QCD jets. In order to mitigate the effect of pileup on the jet substructure observables used in this analysis (softdrop mass and n-subjettiness), the pileup per particle identification (PUPPI) [76] method is implemented.

Boosted selection

For the boosted merged jet analysis an AK8 jet with corrected PUPPI softdrop mass $m_J > 40$ GeV and, to suppress the background from QCD multijet events, $MET > 40$ GeV are required. The $W \rightarrow \ell\nu$ and $W \rightarrow q\bar{q}$ decay candidates satisfying the boosted analysis requirements are combined into $\ell\nu J$ resonance candidates. To suppress the dominant $W + \text{jets}$ background, candidates are selected in the PUPPI softdrop mass region $65 < m_J < 105$, that is the signal region (SR). Outside of this signal region, i.e. for $m_J < 65$ or $m_J > 105$, the candidates are used for background determination (sideband region or SB). A top-enriched control region (Top CR) is also defined by reversing the b-veto and requiring at least one AK4 b-tagged jet in the event. For the final selection a cut on the PUPPI n-subjettiness ratio $\tau_2/\tau_1 < 0.4$ is used to identify boosted hadronic boson candidates. The entire selection procedure

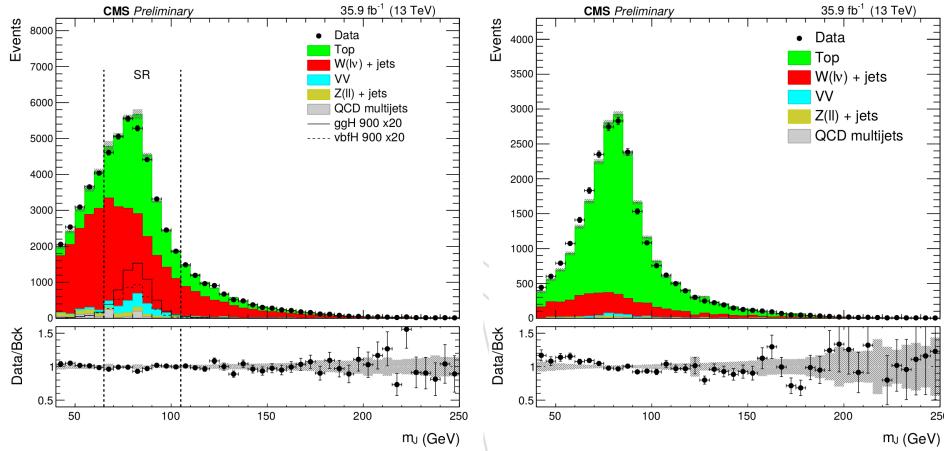


Figure 6.1. Hadronic W candidate PUPPI softdrop mass for events passing the boosted selection in the SR and SB (left), and the Top CR (right).

described above is referred to as the “boosted selection”. The m_J distribution in SR, SB and Top CR is shown in Fig. 6.1.

Resolved selection

For events which do not pass the boosted jet selection, it is attempted to reconstruct a resolved hadronic W decay using two AK4 jets. A kinematic fit is performed to the dijet system using the W mass constraint and, in events with more than two jets, the dijet pair with the smallest χ^2 is chosen. To suppress the background from QCD multijet events we require $MET > 30$ GeV and $W_{m_T} > 50$ GeV, where W_{m_T} is the transverse mass of the leptonic W candidate. The $W \rightarrow \ell\nu$ and $W \rightarrow q\bar{q}$ decay candidates satisfying these requirements are combined into $\ell\nu jj$ resonance candidates. To suppress background $\ell\nu jj$ candidates are selected in the dijet mass signal region (SR) $65 < m_{jj} < 105$ GeV. Sideband (SB) and top-enriched control regions (Top CR) used for background estimation are defined as for the “boosted selection”. The entire selection procedure described above is referred to as the “resolved selection”. The m_{jj} distribution for events passing the resolved selection in SR, SB and Top CR is reported in Fig. 6.2.

6.3 Event categorisation

In this analysis the events are categorised by the flavour of the lepton (electron and muon) and type of hadronic W candidate (boosted and resolved). To increase the signal sensitivity further we also divide events into categories based on the tagging of VBF production and gluon-gluon fusion for the high mass X production:

- **The VBF category:** for the identification of a VBF high mass candidate event, it is required two additional AK4 jets with $p_T > 30$ GeV and $\eta < 4.7$. If there are more than two additional jets then the pair with the highest invariant mass is chosen. An event enters the VBF category if the dijet pair has a

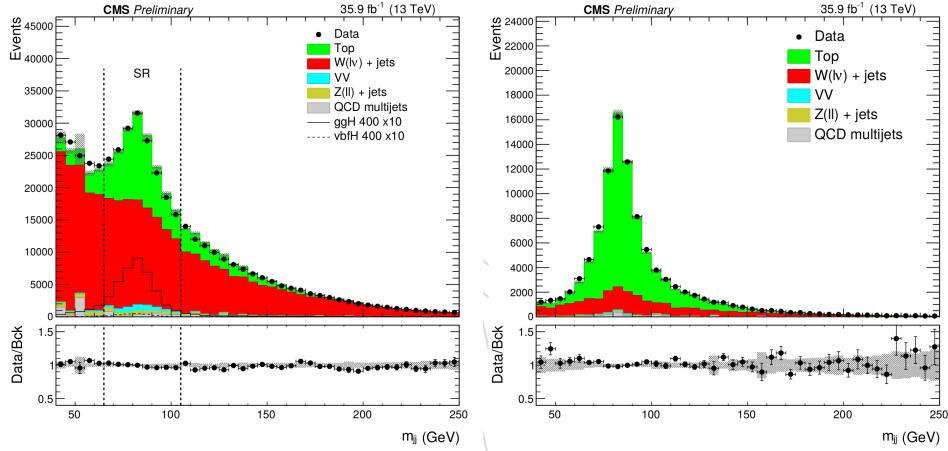


Figure 6.2. Hadronic W dijet mass in data and simulation for events passing the resolved selection in the SR and SB (left), and the Top CR (right).

separation in η , $\Delta\eta_{jj}$, greater than 3.5 and an invariant mass, m_{jj} greater than 500 GeV.

- **The gluon-gluon fusion category:** those events which are not tagged as VBF selection are considered for the gluon-gluon fusion category. The tagging of gluon-gluon fusion is achieved using a kinematic discriminant based on the angular distributions of the X candidate decays products. A kinematic discriminant is constructed as :

$$KD = \left(1 + \frac{c \cdot P_{Bkgr}}{P_{Sig}}\right)^{-1}, \quad (6.1)$$

where c is a parameter that is tuned to adjust the relative normalisations of probabilities. P_{Bkgr} and P_{Sig} are the probability for an event to come from either background or signal respectively calculated using JHUG EN and MCFM. The discriminant is continuously distributed between 0 and 1, with signal being closer to 1 and background closer to 0. In this analysis a cut at 0.5 is used to tag gluon-gluon fusion.

- **The untagged category:** events failing the kinematic discriminant requirement gluon-gluon fusion category and the VBF category enter the untagged category.

Since the decay products of the Higgs-like resonance can be fully reconstructed, it is possible to use its reconstructed mass to discriminate signal events from background events. In fact the mass will peak around the true resonance mass, m_X , for the signal, while for background processes it will have a broader distribution.

The $m_{\ell\nu J}$ and $m_{\ell\nu jj}$ mass distributions in the singnal region, for the three categories (untagged, gluon-gluon fusion and VBF) are depicted in Fig. 6.3. The mass distributions are compared to the background predictions after fitting to the data in the SR with the SB and Top CR used to constrain the background normalisations.

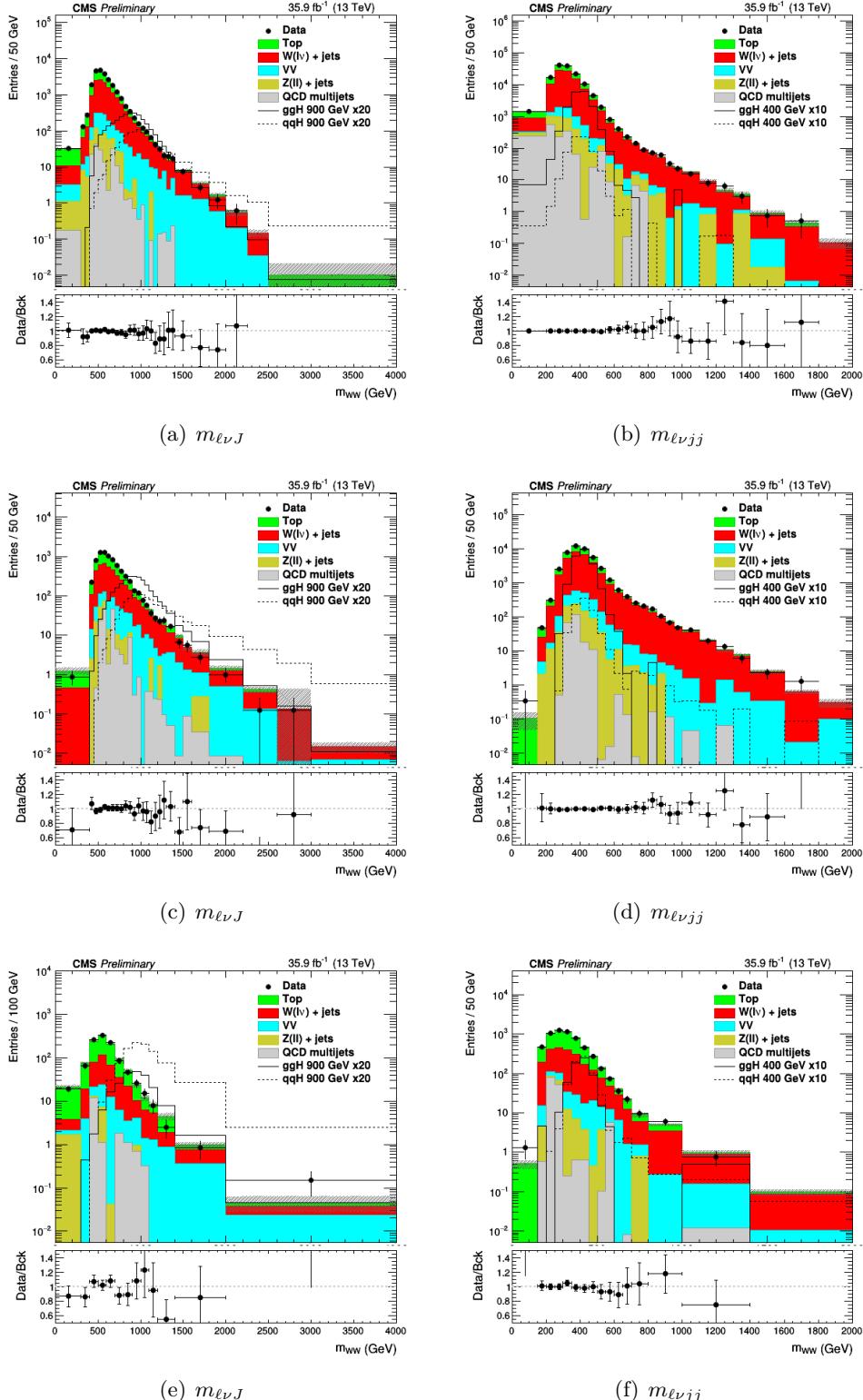


Figure 6.3. The $m_{\ell\nu J}$ distribution in boosted selection on the left, and $m_{\ell\nu jj}$ distribution in resolved selection on the right for data and for simulation for events passing the SR selections in the untagged (top), gluon-gluon fusion (middle) and VBF categories (bottom). The $W+\text{jets}$ and top background normalisations are taken from a fit to the data in the SR, SB and Top CR.

6.4 Backgrounds

The main backgrounds in this analysis are from W +jets and top production. Subdominant backgrounds considered are SM diboson production, Z +jets and QCD multijet production. Background samples are modeled by MC simulation that has been properly reweighted to account for known discrepancies between data and simulated events. The QCD multijet background is greatly suppressed in this analysis and is estimated with PYTHIA. Most events in the selected sample come from W +jets and top production. A good level of agreement between data and simulation is demonstrated by the SR, SB and Top CR distributions. For the final analysis, the W +jets and top X candidate mass distributions are extracted from the simulation. However in each category an alternative normalization to that predicted by the MC is allowed. The normalization of each background is parametrized by a multiplicative factor that is allowed to float free in the fits to the data and is constrained by the observed yields in the SR, SB and Top CR.

6.5 Systematic uncertainties

In this section the systematic uncertainties affecting the boosted and resolved categories are presented:

- Luminosity uncertainty: the uncertainty on LHC luminosity is 2.5%.
- Subdominant background normalization: a 100% uncertainty on the normalization of the diboson, Z +jets and QCD multijets backgrounds is assigned.
- Lepton trigger, identification and isolation: Lepton trigger, reconstruction, identification and isolation scale factors have been computed using tag-and-probe technique for both muons and electrons. The scale factors are applied in this analysis and the related uncertainties estimated by shifting up and down the scale factors by their uncertainty. Additionally, for muons, flat systematic uncertainties are assigned for the trigger (0.5%), identification (1%) and isolation (0.5%). For electrons with $p_T > 80$ GeV a flat systematic uncertainty of 1% is also assigned to the reconstruction. The total uncertainty for muons is 1.5%, while for electrons it is 2.9%.
- Lepton energy scale: Uncertainties related to the lepton energy scale are small. The effect on normalisation is found to be around 0.5%.
- Jet energy scale and resolution: uncertainties related to the jet energy scale are calculated by changing the jet energies by ± 1 sigma error of the corresponding jet energy corrections for AK8 and AK4 jets. Uncertainties related to the jet energy resolution are accessed by applying the recommended resolution scale factors, which typically have a value of 1.1, and comparing the result with the central analysis. The variations in jet energy scale and resolution are also propagated to the MET measurement.
- Heavy quark flavor tagging: the uncertainties related to the b-tagging efficiency of b-jets and to the mistag probability of light jets has been estimated by

independently varying the corresponding data-to-Monte Carlo scale factors by one sigma, as a function of p_T and η of the jets.

- Boosted V tagging efficiency: the uncertainty related to the boosted V tagging efficiency for signal, top and diboson background events is taken into account as a data-to-Monte Carlo scale factor uncertainty of 6% for the selection cuts.
- Boosted V mass scale and resolution: the PUPPI softdrop mass scale (JMS) and resolution (JMR) have been measured using boosted W bosons from top decays. For signal and backgrounds containing hadronic W decays, uncertainties related to the JMS and JMR have been estimated by varying the mass and resolution calibrations within their measured uncertainties. These uncertainties are found to have a negligible effect on the $m_{\ell\nu J}$ and $m_{\ell\nu jj}$ shapes and so are treated as normalisation systematics only.

Chapter 7

Results and Interpretation

In the research of high mass Higgs boson, processes that have been predicted but not yet seen are searched . Given that no excesses over the SM expectation are seen in the mass spectra, the upper limits on the cross sections are computed. In this chapter the 95% CL exclusion limits on the production cross-section are evaluated. These exclusion limits are also interpreted in term of 2HDM. I have performed the limits evaluation for the fully-leptonic analysis and I have participated in the limit combination among fully- and same-leptonic analysis.

7.1 Statistical interpretation

The Bayesian and the classical frequentist [77], with a number of modifications, are two statistical approaches commonly used in high energy physics for characterising the absence of a signal. Both methods allow one to quantify the level of incompatibility of data with a signal hypothesis, which is expressed as a confidence level (C.L.) [78]. For excluding a signal the C.L. 95% is a common choice. The C.L. probabilistic interpretation is used when stating the non-existence of a signal is not straightforward and the subject of a vast body of literature as in the high mass analysis. The procedure used to establish the upper limits calculation is based on frequentist test using a likelihood ratio as a test statistic. In addition to the parameter of interest such as the cross section of the signal, the signal and the background models contain a nuisances parameters whose values are not taken in account as known *a priori* but rather must be fitted from the data [79]. In the following the frequentist approach is described. The expected high mass signal event yields will be generically denoted as s and the backgrounds as b .

The frequentist approach is built to discriminate signal from background events. The most powerful statistic test, in accordance to the Neyman-Pearson lemma [77], is the likelihoods ratio $\lambda(\mu)$,

$$\lambda(\mu) = \frac{\mathcal{L}(\text{data}|\mu s + b)}{\mathcal{L}(\text{data}|b)} \quad (7.1)$$

where, \mathcal{L} is the likelihood function from the product of Poisson probabilities and μ is the strength of the signal process (the case $\mu = 0$ corresponds to background only hypothesis, $\mu = 1$ the nominal signal hypothesis). One can see that $0 \leq \lambda(\mu) \leq 1$,

μ near 1 is a evidence of good agreement among data and the hypothesized μ value. It is convenient, for numerical reason, to use the test statistic q_μ defined as,

$$q_\mu = -2 \ln \lambda(\mu) \quad (7.2)$$

where high value of q_μ correspond to more likely incompatibility between data and μ , i.e. background only hypothesis.

Using the statistic test q_μ , is possible to quantify the level of disagreement between the data and the hypothesis, p -value, defined as,

$$p_\mu = \int_{q_\mu, obs}^{\infty} f(q_\mu | \mu) dq_\mu \quad (7.3)$$

where $q_{\mu, obs}$ is the value of statistic test q_μ observed from the data and $f(q_\mu | \mu)$ is the pdf of q_μ under the assumption of the signal strength μ .

The systematic uncertainties on signal $s(\theta)$ and background $b(\theta)$ rates are introduced in test statistic. The test statistic then would take the following form:

$$q_\mu = \frac{\mathcal{L}(data | \mu, \hat{\theta}_\mu)}{\mathcal{L}(data | 0, \hat{\theta}_0)}, \quad (7.4)$$

where $\hat{\theta}_\mu$ and $\hat{\theta}_0$ are maximum likelihood estimators for the signal+background hypothesis (with the signal strength factor μ) and for the background-only hypothesis ($\mu = 0$). The profile likelihood test statistic is introduced to prevent negative signal as,

$$\tilde{q}_\mu = \frac{\mathcal{L}(data | \mu, \hat{\theta}_\mu)}{\mathcal{L}(data | \hat{\mu}, \hat{\theta})}, \quad 0 \leq \hat{\mu} \leq \mu, \quad (7.5)$$

where $\hat{\mu}$ and $\hat{\theta}$ gives the global maximum of the likelihood. The constrain $0 \leq \hat{\mu}$ is due to a positive signal rate, while the $\hat{\mu} \leq \mu$ is imposed by hand in order to guarantee a one-sided confidence interval.

At this point is useful to evaluate the observed statistic test \tilde{q}_μ^{obs} and the nuisance parameters $\hat{\theta}_0^{obs}$, $\hat{\theta}_\mu^{obs}$ that escribing the experimentally observed data for the background-only and signal+background hypotheses, respectively. With this in mind, the pdf of the test statistic in constructed by generating toy MC pseudo-data for both the background-only and signal+background hypotheses, $f(\tilde{q}_\mu | \mu, \hat{\theta}_\mu^{obs})$ and $f(\tilde{q}_\mu | \mu, \hat{\theta}_0^{obs})$. The corresponding p -value for the signal+background and background-only hypotheses, p_μ and p_b are given by:

$$p_\mu = P(\tilde{q}_\mu \geq \tilde{q}_\mu^{obs} | signal + background) = \int_{q_\mu, obs}^{\infty} f(\tilde{q}_\mu | \mu, \hat{\theta}_\mu^{obs}) d\tilde{q}_\mu \quad (7.6)$$

$$1 - p_b = P(\tilde{q}_\mu \geq \tilde{q}_\mu^{obs} | background - only) = \int_{q_0, obs}^{\infty} f(\tilde{q}_\mu | 0, \hat{\theta}_0^{obs}) d\tilde{q}_\mu. \quad (7.7)$$

The $CL_s(\mu)$ is given by the ratio,

$$CL_s(\mu) = \frac{p_\mu}{1 - p_b} \quad (7.8)$$

To quote the 95% of confidence level upper limits on μ , μ is adjusted until it reaches $CL_S=0.05$. For the background-only hypothesis, the expected median upper-limit and $\pm 1\sigma$ and $\pm 2\sigma$ bands are generated with a large set of background-only pseudo-data. The CL_S is evaluated for each of them. Then, one can build a cumulative probability distribution of results by starting integration from the side corresponding to low event yield. The point at which the cumulative probability distribution crosses the quantile of 50% is the median expected value. The $\pm 1\sigma$ (68%) band is defined by the crossings of the 16% and 84% quantiles. Crossings at 2.5% and 97.5% define the $\pm 2\sigma$ (95%) band.

In the high mass analysis, the interference contribution is not negligible, as described in 5.5, and it is included as part of the signal. In particular during the fit the interference term is scaled by $\sqrt{\mu}$. However, to prevent possible negative probability distribution function of the interference, during the fit the signal yield is computed as,

$$Yield = \sqrt{\mu} \times (S + B + I) + (\mu - \sqrt{\mu}) \times (S) + (1 - \sqrt{\mu}) \times (B) \quad (7.9)$$

where S is the signal, B the background and I the interference.

7.2 Signal interpretation: EW singlet and 2HDM

The signal is interpreted in terms of the electroweak singlet model and in 2HDM model. The model theorist details are described in Sec. 1.3.

Electroweak singlet model

The EW singlet represents a scalar mixing among the high mass particle and the Higgs boson. This model relies on two parameters: the scale factor of the couplings of the high mass resonance with respect to the SM, C' , and the branching fraction of the electroweak singlet to non-SM decays modes, BR_{new} . The electroweak singlet signal strength, μ' and the modified width, Γ' , are related with the parameters in the model by the following equations:

$$\mu' = C'^2 \cdot (1 - BR_{\text{new}}) \quad (7.10)$$

$$\Gamma' = \Gamma_{\text{SM}} \cdot \frac{C'^2}{1 - BR_{\text{new}}} \quad (7.11)$$

The high mass signal samples for different mass hypothesis have been reweighted according to this model. To evaluate the sensitivity with respect the C' parameter, different templates (m_T^I , $m_{\ell\ell}$ and m_T) are studied for high mass boson of 700 GeV, Fig. 7.1. The value of $BR_{\text{new}} = 0$ is considered in all cases. The signal shape is not very sensitive to different C' values, as evident from the distribution, so in the following only the $C' = 1$ hypothesis has been investigated.

2HDM and MSSM models

The 2HDM is a well motivated extension of the SM. It contains two Higgs doublets, from which a total of five Higgs bosons are predicted: two CP-even bosons h and H ,

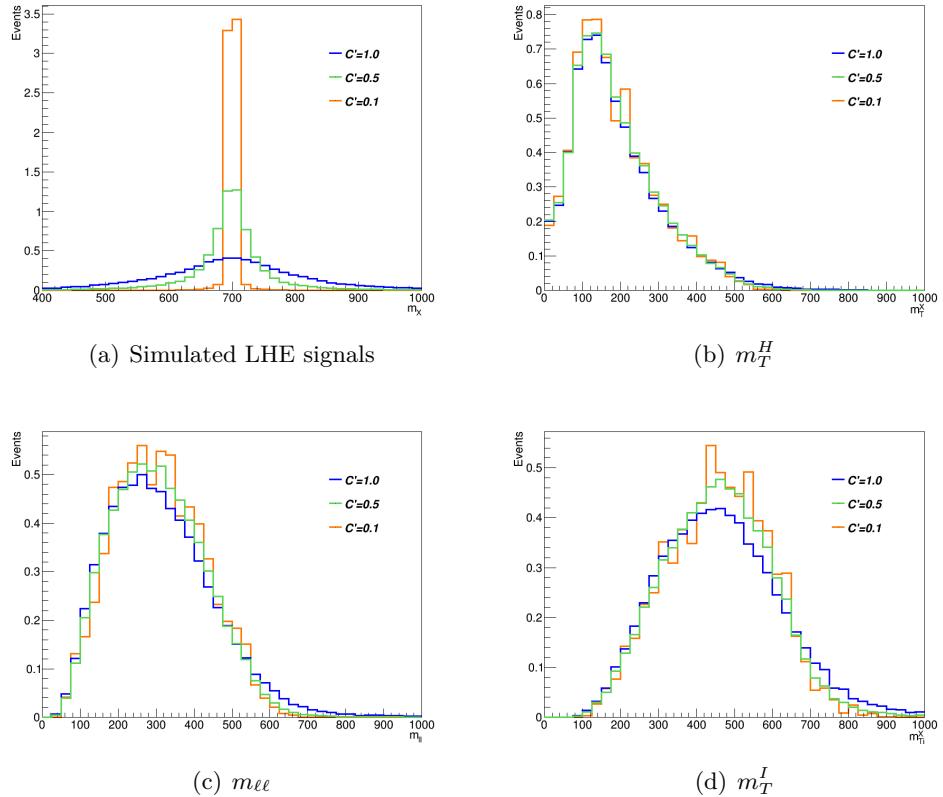


Figure 7.1. Distributions of the MC generator level for signals (no possible reconstruction), for m_T^H , for $m_{\ell\ell}$ and for m_T^I variables using different values of C' (1,0.5,0.1).

a CP-odd boson A and two charged bosons H^\pm . In most theories, h exhibits the features of the SM Higgs boson, while H is a CP-even Higgs boson at a higher mass. The 2HDM comprises many free parameters. Two of these are of particular interest:

- $\tan \beta$: The ratio $\frac{v_u}{v_d}$ of the vacuum expectation values of the two Higgs doublets.
- α : The mixing angle of the two scalar Higgs bosons h and H .

The quantity $\cos(\beta - \alpha)$ is also of interest, as the coupling of the heavy scalar Higgs boson H to two vector bosons is proportional to this factor. In the decoupling limit, which occurs at $\cos(\beta - \alpha) = 0$, all couplings become SM-like. A 2HDM of type-2 is considered in this study. Here up-type quarks couple to one doublet, while down-type quarks and leptons couple to the other doublet.

The MSSM is a type-2 2HDM. On tree level only two parameters are left free. By convention, these parameters are chosen to be $\tan \beta$ and m_A , the mass of the pseudoscalar Higgs boson. The exclusion limits can be set in a two-dimensional plane as a function of these two parameters. Due to higher order diagrams additional free parameters occur. Benchmark scenarios are then used in order to constrain these parameters. Here two MSSM scenarios are used: the m_h^{mod+} scenario and the hMSSM scenario [80]. The exclusion limits obtained for the MSSM scenarios are displayed in the m_A vs $\tan \beta$ plane. A fine grid is chosen in this plane, and for each point of this grid a maximum likelihood fit is performed after the m_A and/or $\tan \beta$ dependent values of the model, such as cross sections and masses of the Higgs bosons are calculated. These fits are done using the asymptotic method. Performing a maximum likelihood fit in this manner is equivalent to a hypothesis test, where the signal hypothesis is tested against the SM and background hypothesis. The signal hypothesis for a combination of m_A and $\tan \beta$ is excluded at 95% confidence level. In the two-dimensional plane this limit is determined from interpolation between the points of the grid. The limits in the more general 2HDM are obtained in the same way, although a different parameter is chosen in place of m_A .

7.3 Fully and same leptonic results

Electroweak singlet sector

No evidence for an excess of events with respect to the SM predictions is observed and thus, exclusion limits in the product of the cross section production times the BR of the decay to two W bosons is evaluated for the EW singlet model. The final binned fit is performed using the m_T^I histogram for all signals and the number of events for the backgrounds for every mass point from 200 GeV up to 3 TeV and the 95% CL upper exclusion limits are calculated. The SM fraction among gluon gluon fusion and VBF mechanism production is considered.

In the fully leptonic analysis, a combination among the opposite flavour and same flavour analysis is performed. The expected and observed limit is shown in Fig. 7.2. This analysis excludes the existence of a resonance in the mass range between 200 GeV and ~ 2.5 TeV: the predicted high mass cross section (the red line in Fig. 7.2

is above the observed data in this masses range. This limit evaluated represent a considerable improvement respect to the high mass search done with 2015 data and the expected and observed limits are compatible with the ATLAS results for a similar analysis, Fig. 1.17.

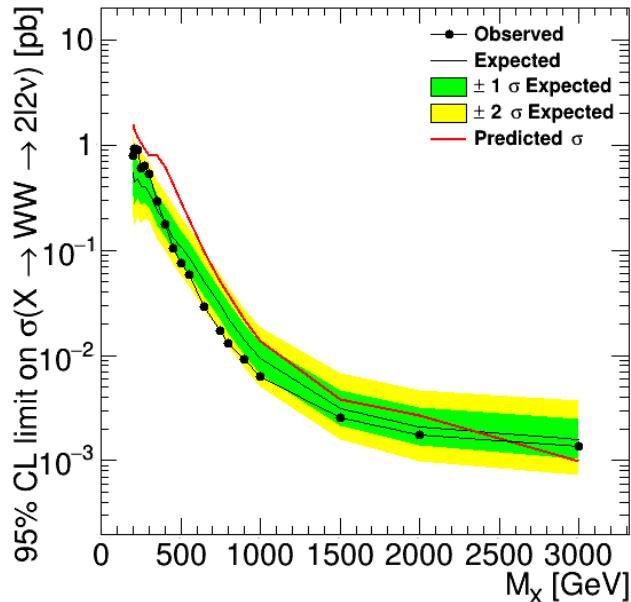


Figure 7.2. Fully leptonic exclusion limits at 95% CL on the cross section (gluon gluon fusion and VBF) times branching ration in $WW \rightarrow 2\ell 2\nu$ as a function of the mass. The red line represent the predicted cross-section for EW high mass bosons.

The same procedure is adopted to evaluated cross section limits for the same leptonic final state, Fig 7.3. The fully and same leptonic analysis are complementary: the former puts a powerful limit at low mass (< 1 TeV), the other instead at very high mass. To perform a right comparison among the two limits, Fig. 1.17 and Fig 7.3, the fully leptonic limits should be multiplied by a factor of ~ 10 to take in account the BR in $2\ell 2\nu$.

MSSM sector

The exclusion limits are also extended to the MSSM model interpretation in terms of m_A . The exclusion limits for the m_h^{mod+} scenario and for the hMSSM scenario are reported for the fully leptonic final state in Fig. 7.4. The dashed line marks the expected limit, while the azure area shows excluded limit obtained with the data. The bands in gray and dark gray surrounding the limit indicate the $\pm 1, 2\sigma$ contours, respectively. For both scenarios the region at low values of m_A (approximately < 150 GeV) and low $\tan \beta$ value (approximately < 20) are excluded. These results complement well with the exclusion limit given by the MSSM $H \rightarrow \tau\tau$ analysis, where the sensitivity is lower for low m_A masses and low $\tan \beta$ values [81].

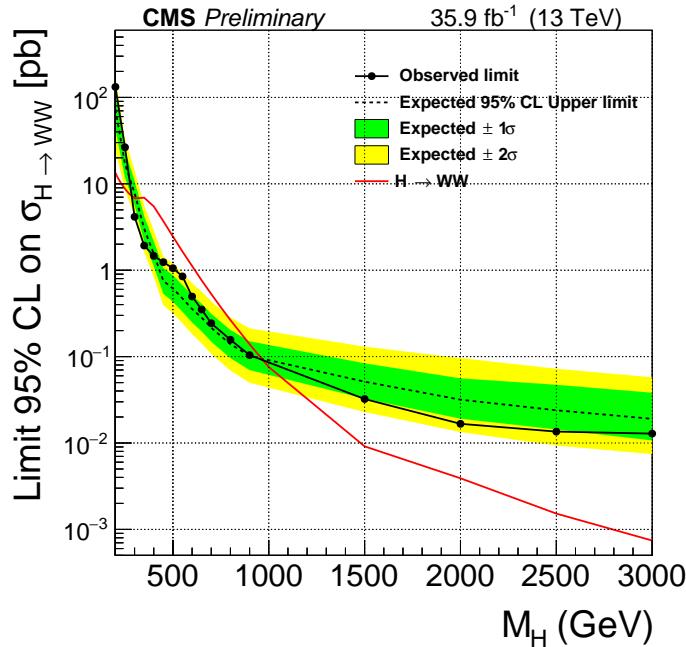


Figure 7.3. Semi leptonic exclusion limits at 95% CL on the production cross section (gluon gluon fusion and VBF) times branching ration in WW as a function of the mass. The red line represent the predicted cross-section for EW high mass bosons.

7.4 Combination among the fully and the semi leptonic analysis

Using the EW model interpretation, the combination of the fully leptonic and semi leptonic analysis has been performed for cross section limits, Fig. 7.5. The limit result represents a improvements for the high mass searches. The value of upper limit are comparable with the results obtained in CMS for $X \rightarrow ZZ$ with final state (4ℓ , $2\ell 2\nu$ and $2\ell 2q$) [82].

The production of the X resonance has been considered to be either in gluon gluon fusion and in VBF with the assumption of the SM fraction. However a scan of the fraction of VBF mechanism production (f_{VBF}) is possible to evaluate various upper limits in different scenarios. The $\sigma_X \times f_{VBF}$ represents the VBF cross section fraction respect to the total cross section. The gluon gluon fusion cross section is so given by $\sigma_X \times (1 - f_{VBF})$. Three different values of f_{VBF} are studied: f_{VBF} free to floating, $f_{VBF} = 0$ and $f_{VBF} = 1$. The upper limits for three cases are shown in Fig. 7.6 for the combination of the fully and same leptonic analysis.

Also the MSSM limits in function of m_A are computed for the combination among the two analysis, Fig. 7.7. The exclusion limits for a 2HDM model are also evaluated. In particular, the 2HDM limits for both the type-1 and the type-2 models are displayed in a $\cos(\beta - \alpha)$ - $\tan \beta$ plane, Fig. 7.8, in which the masses of neutral heavy Higgs are $m_H = m_A = 200, 300, 500$ GeV and the convention $\sin(\beta - \alpha) > 0$ is used.

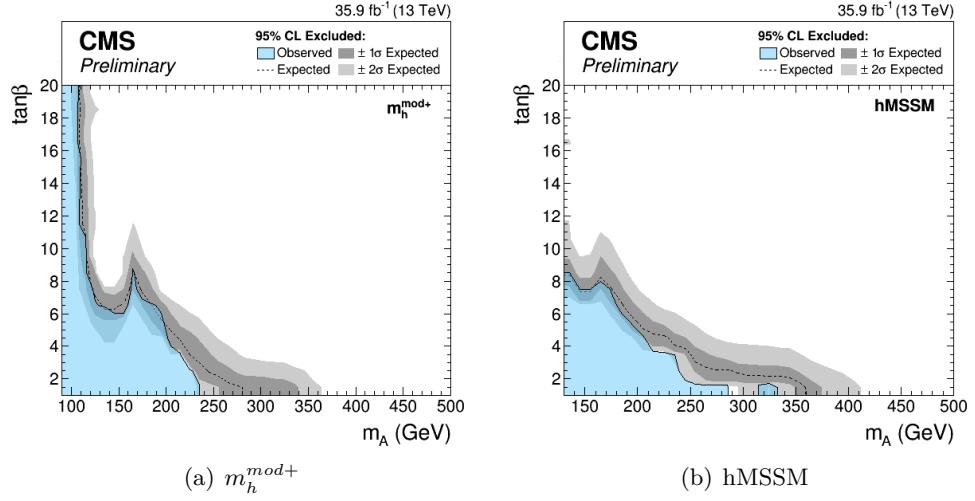


Figure 7.4. (a) 95% CL exclusion limits for the MSSM m_h^{mod+} scenario (b) for the hMSSM scenario.

Instead the limit in the m_H - $\tan\beta$ plane are shown in Fig. 7.9. Here it is again assumed that $m_H = m_A$ and $\sin(\beta - \alpha) > 0$, but here the relationship between β and α is $\cos(\beta - \alpha) = 0.1$.

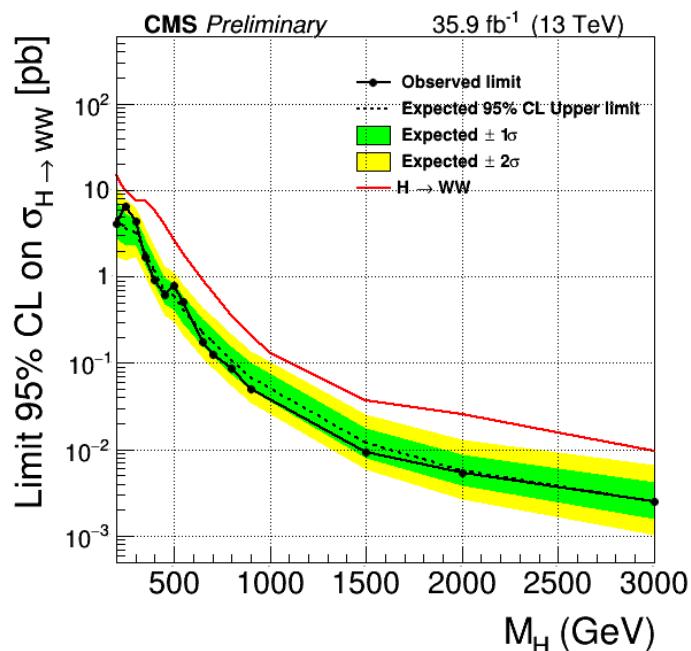


Figure 7.5. Expected and observed exclusion limits at 95% CL on the sum of ggH and VBF cross sections times branching fraction for the combination of all the analysis categories as a function of the resonance mass. The black dotted line corresponds to the central expected value while the yellow and green bands represent the $\pm 1\sigma$ and $\pm 2\sigma$ uncertainties respectively.

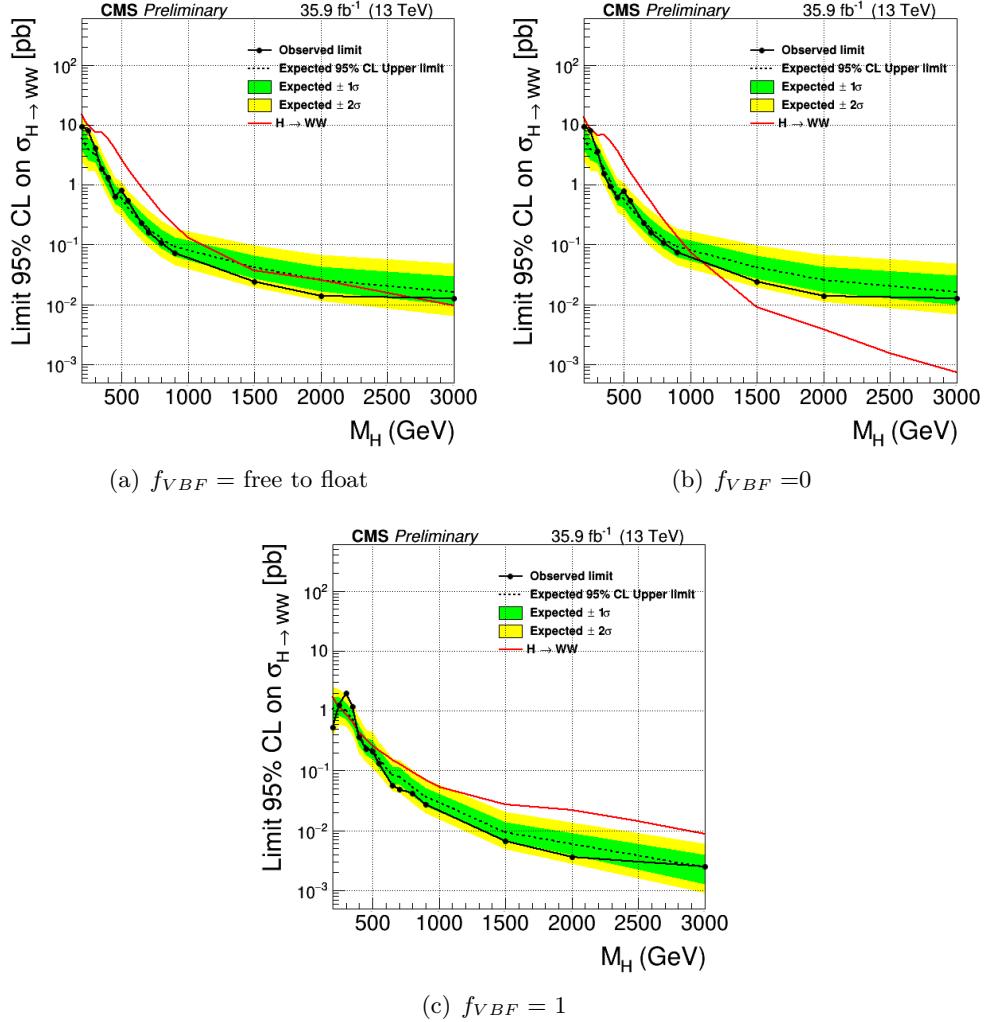


Figure 7.6. 95 CL exclusion limits, on the production gluon-gluon fusion and VBF cross section times branching for different f_{VBF} fraction. The red line represent the predicted cross-section for EW high mass bosons in the free floating case (a), only the predicted vector boson cross-section in f_{VBF} case (b) and only the predicted gluon-gluon fusion cross-section in f_{VBF} case (c).

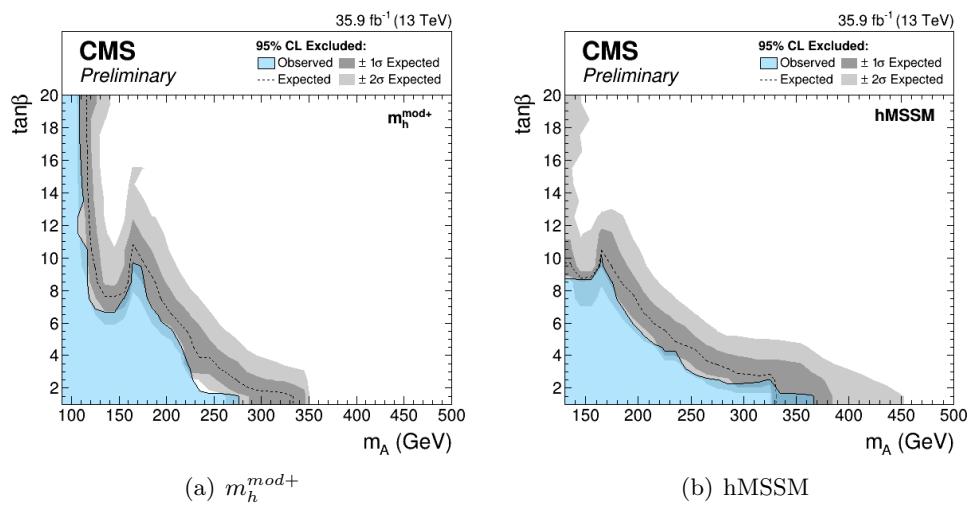


Figure 7.7. (a) 95% CL exclusion limits for the MSSM m_h^{mod+} scenario (b) for the hMSSM scenario.

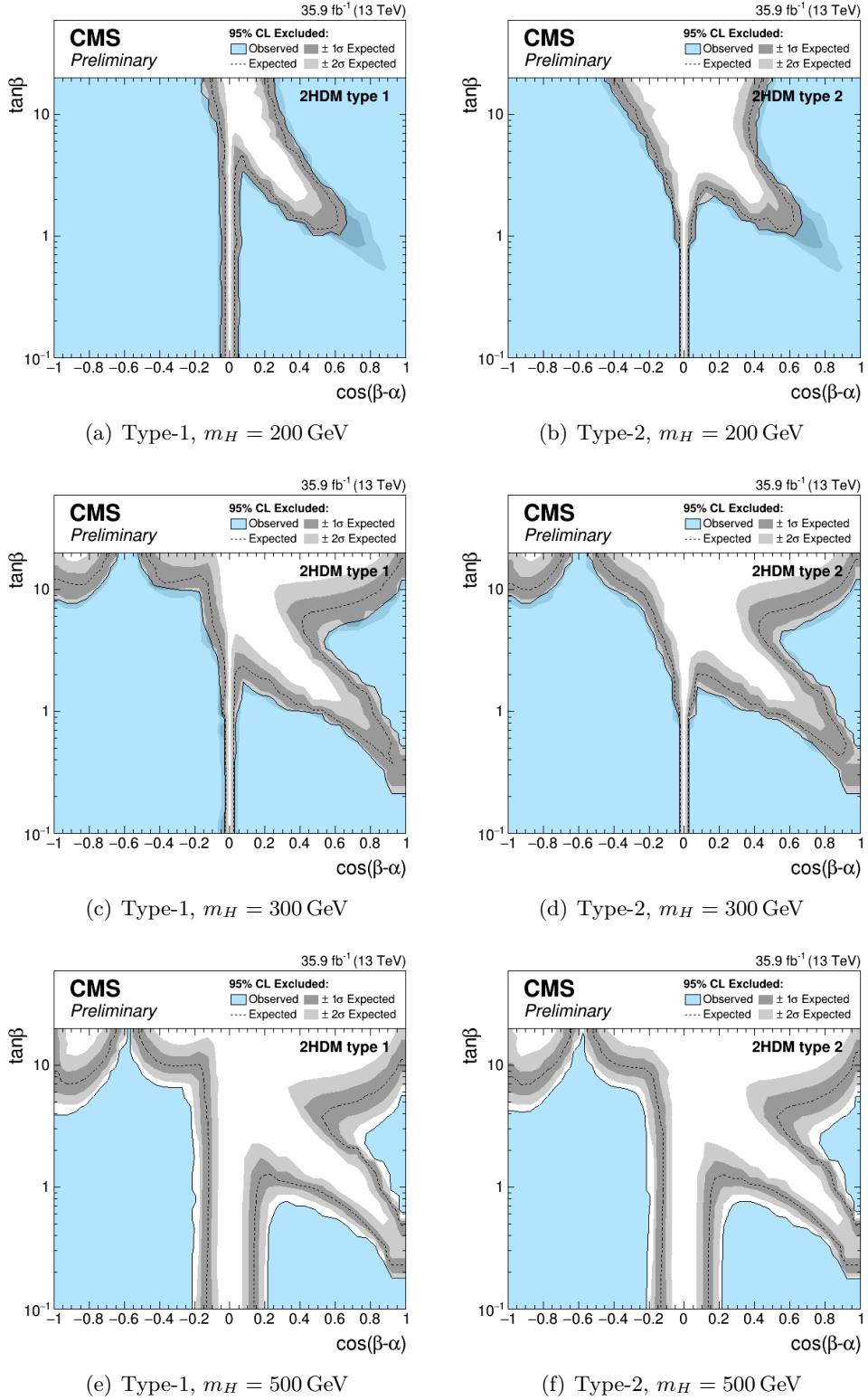


Figure 7.8. 95% CL exclusion limits on a 2HDM with $\cos(\beta - \alpha)$ on the x-axis. Limits are shown for a type-1 and type-2 2HDM for different masses $m_H = 200, 300, 500$ GeV.

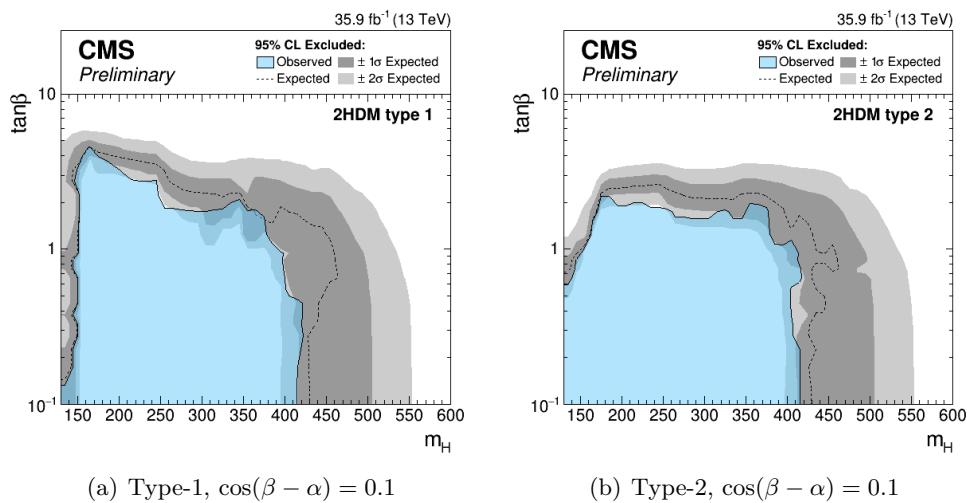


Figure 7.9. 95% CL exclusion limits on a 2HDM with m_H on the x-axis. Limits are shown for a type-1 and type-2 2HDM for $\cos(\beta - \alpha) = 0.1$.

Conclusion

A search for a Standard Model-like Higgs boson in $X \rightarrow W^+W^-$ with fully leptonic and same leptonic final state, produced via gluon-gluon fusion and VBF mechanism, has been performed in the mass range between 200 and 3000 GeV. Data samples collected by the CMS experiment during 2016 at $\sqrt{s} = 13$ TeV for a total luminosity of 35.9 fb^{-1} , have been used. This channel has a high branching ratio but the reconstruction of the final state is not possible due to the presence of neutrino(s). The main sources of background in this channel are represented by WW, Drell-Yan, Top, W+jet and multiboson. The interference between the high mass signal, the WW background and the SM Higgs boson has been studied for the gluon-gluon fusion and the VBF mechanism production. This contribution is not negligible, especially for high mass resonances, so it is included as part of the signal. Different kinematic cuts are applied, in order to tag the fully or semi leptonic final states and to roughly reduce background. To increase to sensitivity to the possible scalar resonance, events are categorized in according to the final states, the number of jets and the jet kinematics. The final statistical analysis is a shape analysis performed on the “improved transverse mass”, a dedicate variable introduced for this specific analysis.

Two models are investigated: the Electroweak singlet and the Two Higgs Doublet Model. Since no signal excess was observed in the distribution of the reconstructed invariant mass of the X candidate, exclusion limits have been set at 95% CL on the production cross section of a scalar neutral boson as function of its mass, in the range 200-3000 GeV for the Electroweak singlet model. The results were also interpreted in the context of the MSSM and 2HDM type-I and type-II scenarios.

Appendix A

Boost-invariant variables

Since the center-of-mass of the parton-parton scattering is normally boosted in the beam direction with respect to that of the two incoming hadrons, it is therefore useful to classify the final state in terms of variables that are invariant under Lorentz transformations in that direction. For this purpose, the variables rapidity y , transverse momentum p_T , and azimuth angle ϕ are introduced. In term of these variables, the four-momentum of a particle of mass m may be written as,

$$p^\mu = (E, p_x, p_y, p_z) \quad (\text{A.1})$$

where where p_x , p_y and p_z are the Cartesian coordinates of the momentum \vec{p} . The rapidity y is defined by the relation,

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

and is not invariant for relativistic transformations but since it transforms according to the law $y' = y + \beta$, (where β is the relative velocity between two frames), the rapidity differences Δy are boost invariant. In the approximation of ultra-relativistic particles rapidity can be approximated by the pseudorapidity η , defined as:

$$\eta = -\ln\left(\tan\frac{\theta}{2}\right) \sim y \quad (\text{A.3})$$

where θ is the angle between the particle and the beam direction, and is therefore directly measurable in the detector.

Appendix B

Parton Distribution Function

In the hadron colliders, the cross-section of $pp \rightarrow F + X$, where F is the final state, is given by,

$$d\sigma_{pp \rightarrow F} = \int_0^1 dx_1 \int_0^1 dx_2 f_1(x_1, \mu_F^2) f_2(x_2, \mu_F^2) d\hat{\sigma}_{pp \rightarrow F}(\mu_F^2), \quad (\text{B.1})$$

where $\hat{\sigma}$ is the partonic cross section, x_1 and x_2 are the transverse momentum fraction respect to the incoming parton, $f_{1,2}(x)$ are the parton distribution function (PDF) giving the probability that a parton has a x momentum fraction of the incoming proton, μ_F is the factorization scale. The PDF are evaluated via experiments, and their evolution is evaluated with perturbative method, following the Altarelli-Parisi equations (DGLAP) [83].

Bibliography

- [1] F. Halzen and Alan D. Martin. *QUARKS AND LEPTONS: AN INTRODUCTORY COURSE IN MODERN PARTICLE PHYSICS*. 1984.
- [2] Sheldon L. Glashow. Partial-symmetries of weak interactions. *Nuclear Physics*, 22(4):579 – 588, 1961.
- [3] Steven Weinberg. A model of leptons. *Phys. Rev. Lett.*, 19:1264–1266, Nov 1967.
- [4] Abdus Salam. Weak and Electromagnetic Interactions. *Conf. Proc.*, C680519:367–377, 1968.
- [5] A. B. Balantekin and W. C. Haxton. Neutrino Oscillations. *Prog. Part. Nucl. Phys.*, 71:150–161, 2013.
- [6] Particle Data Group, J. Beringer, et al. Review of Particle Physics. *Phys. Rev. D*, 86:010001, 2012.
- [7] Luigi Di Lella and Carlo Rubbia. *The Discovery of the W and Z Particles*, pages 137–163.
- [8] et alt. Abe. Evidence for top quark production in $p^- p$ collisions at $\sqrt{s} = 1.8$ tev. *Phys. Rev. Lett.*, 73:225–231, Jul 1994.
- [9] S. Abachi et al. Observation of the top quark. *Phys. Rev. Lett.*, 74:2632–2637, 1995.
- [10] P.W. Higgs. Broken symmetries, massless particles and gauge fields. *Physics Letters*, 12(2):132 – 133, 1964.
- [11] F. Englert and R. Brout. Broken symmetry and the mass of gauge vector mesons. *Phys. Rev. Lett.*, 13:321–323, Aug 1964.
- [12] 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.*, B716:1–29, 2012.
- [13] Serguei Chatrchyan et al. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phys. Lett.*, B716:30–61, 2012.

- [14] Timo van Ritbergen and Robin G. Stuart. On the precise determination of the Fermi coupling constant from the muon lifetime. *Nucl. Phys.*, B564:343–390, 2000.
- [15] J R Andersen et al. Handbook of LHC Higgs Cross Sections: 3. Higgs Properties. 2013.
- [16] Tania Robens and Tim Stefaniak. Status of the Higgs Singlet Extension of the Standard Model after LHC Run 1. *Eur. Phys. J.*, C75:104, 2015.
- [17] Robert M. Schabinger and James D. Wells. A Minimal spontaneously broken hidden sector and its impact on Higgs boson physics at the large hadron collider. *Phys. Rev.*, D72:093007, 2005.
- [18] A. Barroso, P. M. Ferreira, Rui Santos, Marc Sher, and Joao P. Silva. 2HDM at the LHC - the story so far. 2013.
- [19] Sheldon L. Glashow and Steven Weinberg. Natural conservation laws for neutral currents. *Phys. Rev. D*, 15:1958–1965, Apr 1977.
- [20] V. Barger, J. L. Hewett, and R. J. N. Phillips. New constraints on the charged higgs sector in two-higgs-doublet models. *Phys. Rev. D*, 41:3421–3441, Jun 1990.
- [21] Mayumi Aoki, Shinya Kanemura, Koji Tsumura, and Kei Yagyu. Models of yukawa interaction in the two higgs doublet model, and their collider phenomenology. *Phys. Rev. D*, 80:015017, Jul 2009.
- [22] Abdelhak Djouadi. Higgs physics at future colliders: Recent theoretical developments. *Pramana*, 62(2):191–206, Feb 2004.
- [23] A. Denner, S. Heinemeyer, I. Puljak, D. Rebuzzi, and M. Spira. Standard Model Higgs-Boson Branching Ratios with Uncertainties. *Eur. Phys. J.*, C71:1753, 2011.
- [24] Vardan Khachatryan et al. Search for a Higgs boson in the mass range from 145 to 1000 GeV decaying to a pair of W or Z bosons. *JHEP*, 10:144, 2015.
- [25] Search for high mass Higgs to WW with fully leptonic decays using 2015 data. Technical Report CMS-PAS-HIG-16-023, CERN, Geneva, 2016.
- [26] Search for a high-mass Higgs boson decaying to a pair of W bosons in pp collisions at $\sqrt{s}=13$ TeV with the ATLAS detector. Technical Report ATLAS-CONF-2016-074, CERN, Geneva, Aug 2016.
- [27] Thomas Sven Pettersson and P Lefèvre. The Large Hadron Collider: conceptual design. Technical Report CERN-AC-95-05-LHC, Oct 1995.
- [28] T. Sjostrand. Monte Carlo generators. 2006.
- [29] A. Buckley et al. General-purpose event generators for LHC physics. *Phys. Rept.*, 504:145–233, 2011.

- [30] T. Stelzer and W. F. Long. Automatic generation of tree level helicity amplitudes. *Comput. Phys. Commun.*, 81:357–371, 1994.
- [31] M. H. Seymour. Matrix element corrections to parton shower algorithms, *Comp. Phys. Commun.* 90 (1995) 95–101, arXiv:hep-ph/9410414.
- [32] G. Miu and T. Sjostrand. W production in an improved parton shower approach,” *Phys. Lett. B*449 (1999) 313–320, arXiv:hep-ph/9812455.
- [33] R. Kuhn S. Catani, F. Krauss and B. R. Webber. QCD matrix elements + parton showers,” *JHEP* 11 (2001) 063, arXiv:hep-ph/0109231.
- [34] F. Krauss. Matrix elements and parton showers in hadronic interactions,” *JHEP* 08 (2002) 015, arXiv:hep-ph/0205283.
- [35] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti- k_t jet clustering algorithm. *JHEP*, 04:063, 2008.
- [36] C. Oleari. The POWHEG-BOX. *Nucl. Phys. Proc. Suppl.*, 205-206:36–41, 2010.
- [37] T. Sjostrand, S. Mrenna, and P. Skands. PYTHIA 6.4 physics and manual. *JHEP*, 05:026, 2006.
- [38] T. Gleisberg et al. SHERPA 1.alpha, a proof-of-concept version. *JHEP*, 02:056, 2004.
- [39] Richard D. Ball, Valerio Bertone, Stefano Carrazza, Luigi Del Debbio, Stefano Forte, Alberto Guffanti, Nathan P. Hartland, and Juan Rojo. Parton distributions with QED corrections. *Nucl. Phys.*, B877:290–320, 2013.
- [40] Richard D. Ball, Valerio Bertone, Francesco Cerutti, Luigi Del Debbio, Stefano Forte, Alberto Guffanti, Jose I. Latorre, Juan Rojo, and Maria Ubiali. Unbiased global determination of parton distributions and their uncertainties at NNLO and at LO. *Nucl. Phys.*, B855:153–221, 2012.
- [41] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. A Brief Introduction to PYTHIA 8.1. *Comput. Phys. Commun.*, 178:852–867, 2008.
- [42] Vardan Khachatryan et al. Event generator tunes obtained from underlying event and multiparton scattering measurements. 2015.
- [43] S. Agostinelli et al. GEANT4—a simulation toolkit. *Nucl. Instrum. Meth. A*, 506:250, 2003.
- [44] 2015 pileup json files. https://twiki.cern.ch/twiki/bin/view/CMS/PileupJSONFileforData#2015_Pileup_JSON_Files.
- [45] Paolo Nason. A New method for combining NLO QCD with shower Monte Carlo algorithms. *JHEP*, 11:040, 2004.
- [46] Stefano Frixione, Paolo Nason, and Carlo Oleari. Matching NLO QCD computations with Parton Shower simulations: the POWHEG method. *JHEP*, 11:070, 2007.

- [47] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re. A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX. *JHEP*, 06:043, 2010.
- [48] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re. NLO Higgs boson production via gluon fusion matched with shower in POWHEG. *JHEP*, 04:002, 2009.
- [49] Paolo Nason and Carlo Oleari. NLO Higgs boson production via vector-boson fusion matched with shower in POWHEG. *JHEP*, 02:037, 2010.
- [50] A. V. Gritsan *et. al.* S. Bolognesi, Y. Gao. Jhugen. <http://www.pha.jhu.edu/spin/>.
- [51] Gionata Luisoni, Paolo Nason, Carlo Oleari, and Francesco Tramontano. $HW^\pm/HZ + 0$ and 1 jet at NLO with the POWHEG BOX interfaced to GoSam and their merging within MiNLO. *JHEP*, 10:083, 2013.
- [52] Sm higgs production cross sections at $\sqrt{s} = 13\text{-}14$ tev. <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageAt1314TeV>.
- [53] Tom Melia, Paolo Nason, Raoul Rontsch, and Giulia Zanderighi. W+W-, WZ and ZZ production in the POWHEG BOX. *JHEP*, 11:078, 2011.
- [54] John M. Campbell, R. Keith Ellis, and Ciaran Williams. Bounding the Higgs width at the LHC: Complementary results from $H \rightarrow WW$. *Phys. Rev.*, D89(5):053011, 2014.
- [55] Morad Aaboud et al. Measurements of gluon-gluon fusion and vector-boson fusion Higgs boson production cross-sections in the $H \rightarrow WW^* \rightarrow e\nu\mu\nu$ decay channel in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. 2018.
- [56] Patrick Meade, Harikrishnan Ramani, and Mao Zeng. Transverse momentum resummation effects in W^+W^- measurements. *Phys. Rev.*, D90(11):114006, 2014.
- [57] Prerit Jaiswal and Takemichi Okui. Explanation of the WW excess at the LHC by jet-veto resummation. *Phys. Rev.*, D90(7):073009, 2014.
- [58] Nlo single-top channel cross sections. <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/SingleTopRefXsec>.
- [59] Nnlo+nnll top-quark-pair cross sections. <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/TtbarNNLO>.
- [60] 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*, 07:079, 2014.
- [61] Higgs to WW measurements with 15.2 fb^{-1} of 13 TeV proton-proton collisions. Technical Report CMS-PAS-HIG-16-021, CERN, Geneva, 2017.

- [62] Measurements of properties of the Higgs boson decaying to a W boson pair in pp collisions at $\sqrt{s} = 13$ TeV. Technical Report CMS-PAS-HIG-16-042, CERN, Geneva, 2018.
- [63] Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET. Technical Report CMS-PAS-PFT-09-001, CERN, Geneva, Apr 2009.
- [64] Serguei Chatrchyan et al. Description and performance of track and primary-vertex reconstruction with the CMS tracker. *JINST*, 9(10):P10009, 2014.
- [65] W. Adam, B. Mangano, T. Speer, and T. Todorov. Track reconstruction in the CMS tracker. 2005.
- [66] Thomas Speer, Kirill Prokofiev, R Frühwirth, Wolfgang Waltenberger, and Pascal Vanlaer. Vertex Fitting in the CMS Tracker. Technical Report CMS-NOTE-2006-032, CERN, Geneva, Feb 2006.
- [67] K. Rose. Deterministic annealing for clustering, compression, classification, regression, and related optimization problems. *Proceedings of the IEEE*.
- [68] A. M. Sirunyan et al. Particle-flow reconstruction and global event description with the CMS detector. *JINST*, 12(10):P10003, 2017.
- [69] Matteo Cacciari and Gavin P. Salam. Pileup subtraction using jet areas. *Phys. Lett.*, B659:119–126, 2008.
- [70] Pileup Removal Algorithms. Technical Report CMS-PAS-JME-14-001, CERN, Geneva, 2014.
- [71] Serguei Chatrchyan et al. Identification of b-quark jets with the CMS experiment. *JINST*, 8:P04013, 2013.
- [72] John Ellis and Dae Sung Hwang. Does the ‘Higgs’ have Spin Zero? *JHEP*, 09:071, 2012.
- [73] Radja Boughezal, Xiaohui Liu, Frank Petriello, Frank J. Tackmann, and Jonathan R. Walsh. Combining Resummed Higgs Predictions Across Jet Bins. *Phys. Rev.*, D89(7):074044, 2014.
- [74] Andrew J. Larkoski, Simone Marzani, Gregory Soyez, and Jesse Thaler. Soft drop. *Journal of High Energy Physics*, 2014(5):146, May 2014.
- [75] Jesse Thaler and Ken Van Tilburg. Maximizing Boosted Top Identification by Minimizing N-subjettiness. *JHEP*, 02:093, 2012.
- [76] Daniele Bertolini, Philip Harris, Matthew Low, and Nhan Tran. Pileup per particle identification. *Journal of High Energy Physics*, 2014(10):59, Oct 2014.
- [77] G. Cowan. Statistical data analysis. 1998.
- [78] Procedure for the LHC Higgs boson search combination in Summer 2011. (CMS-NOTE-2011-005. ATL-PHYS-PUB-2011-11), Aug 2011.

- [79] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells. Asymptotic formulae for likelihood-based tests of new physics. *Eur. Phys. J.*, C71:1554, 2011. [Erratum: Eur. Phys. J.C73,2501(2013)].
- [80] Stefania Gori, Ian-Woo Kim, Nausheen R. Shah, and Kathryn M. Zurek. Closing the Wedge: Search Strategies for Extended Higgs Sectors with Heavy Flavor Final States. *Phys. Rev. D*, 93(WSU-HEP-1508. WSU-HEP-1508):075038. 30 p, Feb 2016. Comments: 30 pages, 10 figures.
- [81] CMS Collaboration. Search for additional neutral MSSM Higgs bosons in the di-tau final state in pp collisions at $\sqrt{s} = 13$ TeV. 2017.
- [82] Albert M Sirunyan et al. Search for a new scalar resonance decaying to a pair of Z bosons in proton-proton collisions at $\sqrt{s} = 13$ TeV. *JHEP*, 06:127, 2018.
- [83] G. Altarelli and G. Parisi. Asymptotic Freedom in Parton Language. *Nucl. Phys.*, B126:298, 1977.