UNIVERSITÀ DI PISA

FACOLTÀ DI SCIENZE MATEMATICHE, FISICHE E NATURALI

Dipartimento di Fisica

SEARCH FOR NEW EXOTIC RESONANCES IN SEMILEPTONIC ZH FINAL STATE AT CMS

Master thesis

Andrea Mauri

UNIVERSITÀ DI PISA

FACOLTÀ DI SCIENZE MATEMATICHE, FISICHE E NATURALI

Corso di Laurea Magistrale in FISICA

SEARCH FOR NEW EXOTIC RESONANCES IN SEMILEPTONIC ZH FINAL STATE AT CMS

Master thesis of:

Andrea Mauri

Supervisor:

Prof. Luigi Rolandi .....

Internal supervisor:

Prof. Giovanni Batignani .....

Candidate:

Andrea Mauri .....

3 February 2014

Academic year 2012-2013

Contents

Chapter 1

Introduction

In the summer of 2012, ATLAS and CMS [1] [2] have announced the discovery of a new boson with mass around 125 GeV. This boson is compatible with a standard model Higgs boson, as far as the theoretical predicted cross section and decay branching ratio are concerned.

This discovery has completed the Standard Model (SM) as known so far [3], and it opens a new age of research in particle physics. The current challenge is to find signals of new physics Beyond the Standard Model (BSM) since, according to the scientific community, SM cannot explain all phenomena observed in the universe (dark matter). Moreover, there is no natural explanation for the difference between the electroweak scale $O(100$ GeV) and the Plank scale $O(10^{19}$ GeV): the Higgs mass correction requires unnatural fine-tuning. In the last 30 years, many natural theories have been formulated in order to solve the naturalness problem, and all of them share the prediction of visible experimental effects at the TeV energy scale (1-10 TeV). Therefore, in 1994, the construction of the Large Hadron Collider (LHC) at CERN has been approved. It is a 27 kilometres circular accelerator that will be able to collide proton bunches at $\sqrt{s} = 14$ TeV, exploring for the first time the TeV scale mass region.

Four different experiments (ATLAS, CMS, ALICE and LHCb) are placed in the four collision points of the ring, and in 2009 there was the first proton collision at low energy in all the detectors. Then the LHC ran at $\sqrt{s} = 7 \div 8$ TeV from March 2010 to February 2013 with few short interruptions for the heavy ions collisions. LHC is now in a shutdown phase to prepare the collider for a higher energy and luminosity.

LHC has already collected a large amount of data and the discovery of the Higgs boson has made possible to study final states with the Higgs boson itself. The interest of these studies lies in the expectation that these events might be sensitive to new physics. In this thesis I present a data analysis performed at the CMS experiment. The intent of this work is to search for new ZH resonances at high mass, as predicted by the partial compositeness

model [4].

In the first chapter the theoretical model and predictions are briefly presented, while the second chapter contains a description of the apparatus, with special attention to the performance involved in this search. In the last section the complete analysis is shown. It is divided into two parts: chapter 3, that contains a study of high mass events peculiarities and discusses the most adequate tools to be used in the analysis, and chapter 4, that contains the data analysis itself. Chapter 5 presents a comparison with other searches and future perspective.

1.1 The naturalness problem

The SM has been very successful in describing the high-energy physics phenomena observed so far. However, the light mass of the Higgs boson leaves open the questions about the naturalness of the SM and the presence of new physics.

The electroweak energy scale is

$$v \approx 246\, GeV \tag{1.1}$$

where $v/\sqrt{2}$ is the vacuum expectation value ('vev') of the scalar Higgs field, $\langle 0|\phi_0|0\rangle = v/\sqrt{2}$, and sets all the masses in the theory

$$M_W = \frac{gv}{2} \tag{1.2}$$

$$M_Z = v\frac{\sqrt{g^2+g'^2}}{2} \tag{1.3}$$

$$M_H = v\sqrt{\frac{\lambda}{2}} \tag{1.4}$$

where g and g' are the SU(2) and U(1) gauge coupling constants of the theory, and λ is the strenght of the Higgs self-interaction in the scalar potential

$$V = -\mu^2\phi^\dagger\phi + \frac{\lambda}{4}(\phi^\dagger\phi)^2 \tag{1.5}$$

where $\lambda > 0$ and $\mu^2 > 0$.

The negative sign of the coefficient of the first term in the potential is essential for the spontaneous symmetry-breaking of the SM. With this assignement the minimum of the potential in equation 1.5 is given by

$$\phi = \mu\sqrt{\frac{2}{\lambda}} = v/\sqrt{2} \tag{1.6}$$

where $\mu \equiv \sqrt{\mu^2}$. Note that if we fix a positive sign to the $-\mu^2$ coefficient the equilibrium point of the potential would have been simply $v = 0$.

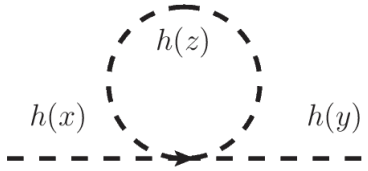

Figure 1.1. One-loop self interaction graph in ϕ^4 theory.

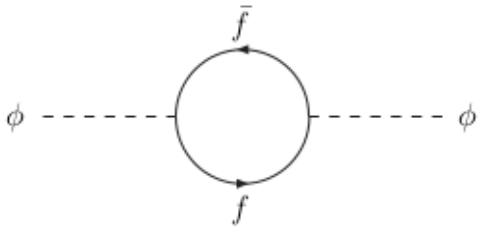

Figure 1.2. A fermion anti-fermion diagram of the contribution to the self energy of the Higgs boson in the Standard Model.

In a quantum field theory, divergencies may appear when integrating the virtual momenta to infinity. In renormalizable theories, instead, physical quantities cannot contain divergencies that cannot be absorbed by the bare lagrangian parameters. If we compute the contribution of the self-energy of a scalar field, building a Feynman diagram with a four legs vertex as in figure 1.1, we obtain [5]

$$\frac{\lambda}{2}\frac{1}{(2\pi)^4}\int d^4k \frac{1}{k^2 - M_H^2} \tag{1.7}$$

where k is the momentum of the virtual particle in the loop that should be integrated to infinity. It is evident that this quantity is quadratically divergent and, introducing a cut-off Λ in order to keep the integral finite, it scales as

$$\sim \lambda\Lambda^2\phi^\dagger\phi. \tag{1.8}$$

If one includes the rest of the SM particles in the Higgs mass correction, things get worse. For example considering a fermionic loop (figure 1.2), as in [6], leads to

$$g_f^2\frac{N_c}{(2\pi)^4}\int d^4k \frac{k^2 + m_f^2}{(k^2 - m_f^2)^2} \tag{1.9}$$

where g_f is the coupling between the Higgs boson and the fermions and N_c the number of color in the theory. This correction gives again a quadratic divergence. Bosonic loops produce also divergent contributions, but with opposite sign. However in the SM they do not cancel.

In conclusion, one gets from all the 1 loop diagrams a positive correction as in equation 1.8 [6] to the 'bare' $-\mu^2\phi^\dagger\phi$ in the potential that affects the physical value of the theory as

$$\mu^2_{phys} = \mu^2 - \lambda\Lambda^2 \tag{1.10}$$

Re-minimizing the potential V, one obtains again expression 1.6, but replacing μ with μ_{phys}. Then, v is fixed phenomenologically by equation 1.1 and the solution to the minimum of the potential provides a relation between the two unknown parameters μ_{phys} and λ:

$$\mu_{phys} \approx \sqrt{\lambda}\,123 \text{ GeV}. \tag{1.11}$$

If we want to be able to treat the the Higgs coupling λ perturbatively and to have a small Higgs mass as observed, μ_{phys} can not be greater than a few hundred GeV.

On the other hand the cut-off Λ represents the energy scale where new physics starts to appear and the existing SM theory must be modified. This certainly happens at the Planck scale

$$M_P = (G_N)^{-1/2} \simeq 1.2 \times 10^{19} \text{ GeV} \tag{1.12}$$

where the strengh of gravity becomes of the same order of the other forces.

If no new physics lays between these two energy scales, we should consider $\Lambda \sim M_P \sim 10^{19}$ GeV and, in order to obtain a final $\mu_{phys} \sim 100$ GeV, the lagrangian parameter μ should be extremely fine-tuned [5].

On the contrary, a fine-tuning of the Higgs mass of the order of 1-10% requires the existence of new physics at the TeV scale.

1.2 Beyond the Standard Model (BSM)

Several models have been formulated in order to solve the naturalness problem, but none of their predictions has been observed in the experiments. In this section we briefly present the most popular theories and how they extend the SM.

Supersymmetry (SUSY) aims for the extension of the SM where each SM particle has a supersimmetric partner in the SUSY sector. If one introduces bosonic partners for each SM fermion and vice versa fermionic partners for each SM boson contributing to the one loop correction to the Higgs mass, one finds out that the SUSY corrections have opposite sign as the SM particles and lead to the cancellation of the quadratic divergence [7].

Extra-dimensions (ED) predicts the presence of further dimensions beyond the 3 + 1 experienced in our everyday life. The space-time would then consist of $4 + N$ dimensions, where N is the number of additional ED. In this scenario the boson carrier of the gravitational force, the *graviton*, can propagate in the extra-dimensions reducing the strenght of the gravitational force. In this case the cut-off Λ is given by the typical size of the extra-dimensions.

Possible observations that could verify this model consist of detecting the decay of the graviton, whose branching ratios better couple to the heaviest SM particles. The main searches look for pairs of SM bosons (WW, ZZ, HH) or $t\bar{t}$ in the final state, predicted respectively with the proportion 2:1:1:9 [8] [9].

Little Higgs This model predicts the existence of heavy gauge bosons Z_H and W_H similar to the SM bosons Z and W but much heavier. The value of the Higgs mass is protected by an adequate combination of discrete symmetries. The decay of these new particles includes a large combination of final states, ZH, WH in addition to ZZ, WW, HH, $f\bar{f}$, and three body decays are not negligible [10] [11].

Hidden Valley models admit BSM heavy particles of any spin decaying to gg, $\gamma\gamma$, ZZ, WW, γZ but not ZH [12].

1.2.1 Partial compositeness

Let us now focus on the model and theoretical predictions that motivate this search in the ZH channel. We follow the minimal scenario described by R. Contino et al. in [4], this is a schematic interpretation of general extra-dimensions models and of other several theories that predict strong dynamics at the TeV scale. The resulting set-up goes under the name of *partial compositeness*[1].

This model predicts a weakly coupled elementary particles sector, described by $\mathcal{L}_{elementary}$, and a second strongly interacting sector, made of bound composite states, including the Higgs doublet, described by $\mathcal{L}_{composite}$. The elementary sector coupling constants are roughly $g_{el} \sim 1$, while the coupling force between the different composite states, g_*, is stronger, $1 < g_* \ll 4\pi$.

These two sectors couple to each other through a mixing mechanism given by $\mathcal{L}_{mixing}$, which primarily results in mass-mixing. The total lagrangian can be written as

[1]This truncation, while losing some of the explanatory power and precision of the full higher-dimensional warped theory, greatly simplifies phenomenological considerations and computations.

$$\mathcal{L} = \mathcal{L}_{elementary} + \mathcal{L}_{composite} + \mathcal{L}_{mixing}, \tag{1.13}$$

and mass eigenstate are superposition of elementary and composite states. The lighest ones are identified as the SM particles,

$$|SM_n\rangle = \cos\varphi_n|\text{elementary}_n\rangle + \sin\varphi_n|\text{composite}_n\rangle, \tag{1.14}$$

while its orthogonal combination

$$|\text{heavy}_n\rangle = -\sin\varphi_n|\text{elementary}_n\rangle + \cos\varphi_n|\text{composite}_n\rangle \tag{1.15}$$

is interpreted as state of the TeV-scale new physics. The mixing angles denoted with φ_n parametrize the degree of "compositeness".

Considering the combination of the states and the coupling of the two sectors, the order of magnitude of the interaction of every mass eigenstate with each other can be found, obtaining

$$\begin{aligned}
g_{\text{SM}_1\text{SM}_2\text{SM}_3} &\sim g_{el}\cos\varphi_1\cos\varphi_2\cos\varphi_3 + g_*\sin\varphi_1\sin\varphi_2\sin\varphi_3 \\
g_{\text{SM}_1\text{SM}_2\text{heavy}_3} &\sim -g_{el}\cos\varphi_1\cos\varphi_2\sin\varphi_3 + g_*\sin\varphi_1\sin\varphi_2\cos\varphi_3 \\
g_{\text{SM}_1\text{heavy}_2\text{heavy}_3} &\sim g_{el}\cos\varphi_1\sin\varphi_2\sin\varphi_3 + g_*\sin\varphi_1\cos\varphi_2\cos\varphi_3 \\
g_{\text{heavy}_1\text{heavy}_2\text{heavy}_3} &\sim -g_{el}\sin\varphi_1\sin\varphi_2\sin\varphi_3 + g_*\cos\varphi_1\cos\varphi_2\cos\varphi_3.
\end{aligned}$$

In this model the SM Higgs is a full composite state with no elementary component, that is $\sin\varphi_{Higgs} = 1$ and $\cos\varphi_{Higgs} = 0$. Consequently the other SM particles couple to it with strenght $\sim \sin\varphi_1\sin\varphi_2$ acquiring mass as result. The heavier SM particles correspond to larger mixing angles, and they are more strongly coupled to the new TeV-scale physics.

This correlation also explains why no visible effects of new physics have ever been observed, even if low-energy precision tests of the SM are studied to be sensitive to heavy physics up to ~ 10 TeV: the maximum experimental sensitivity is given by the lighest SM particles, and in this framework they have really small coupling to the composite sector.

Heavy gauge boson production and decays

The elementary sector contains all the SM fields with the exception of the Higgs field, while the composite sector can be seen as an excited copy of the SM fermions and gauge bosons. One can denote the elementary gauge fields, where the gauge group is the usual $[SU(3)_c \otimes SU(2)_L \otimes U(1)_Y]^{el}$, with

$$A_\mu \equiv \{G_\mu, W_\mu, B_\mu\}, \tag{1.16}$$

and the fermionic doublets and singlets with

$$\psi_L = \{q_{Li} = (u_{Li}, d_{Li}), (l_{Li} = \nu_{Li}, e_{Li})\} \tag{1.17}$$

and

$$\tilde{\psi}_R \equiv \{u_{Ri}, d_{Ri}, \nu_{Ri}, e_{Ri}\}. \tag{1.18}$$

On the other hand, calling ρ_μ the gauge bosons of the composite sector, one obtains [4]

$$\begin{aligned}\mathcal{L}_{composite} &= -\frac{1}{4}\rho_{\mu\nu}^2 + \frac{M_*^2}{2}\rho_\mu^2 + |D_\mu H|^2 - V(H) \\ &\quad +\bar{\chi}(i\not{D} - m)\chi + \bar{\tilde{\chi}}(i\not{D} - m)\tilde{\chi} \\ &\quad -\bar{\chi}(Y_{*u}\tilde{H}\tilde{\chi}^u + Y_{*d}H\tilde{\chi}^d) + \text{h.c.}\end{aligned} \tag{1.19}$$

where χ is the SU(2) composite doublet and $\tilde{\chi}$ the singlet. Then one divides the heavy vector excitations into those that mix with the elementary gauge bosons, ρ_μ^*, and those that do not mix, $\tilde{\rho}_\mu$:

$$\rho_\mu = \{\rho_\mu^*, \tilde{\rho}_\mu\}. \tag{1.20}$$

In analogy with the elementary sector we rename ρ_μ^* with:

$$\rho_\mu^* = \{G_\mu^*, W_\mu^*, B_\mu^*\} \tag{1.21}$$

Partial compositeness is realized by adding a set of mass-mixing term to our model [4],

$$\mathcal{L}_{mixing} = -M_*^2\frac{G_{el}}{g_*}A_\mu\rho_\mu^* + \frac{M_*^2}{2}\left(\frac{g_{el}}{g_*}A_\mu\right)^2 + (\bar{\psi}_L\Delta\chi_R + \bar{\tilde{\psi}}_L\tilde{\Delta}\tilde{\chi}_L + \text{h.c.}), \tag{1.22}$$

where Δ and $\tilde{\Delta}$ are the mixing mass parameters. One can now write more completely the combination between the elementary and composite sectors of equations 1.14 and 1.15. Diagonalizing $\mathcal{L}_{mixing}$ one obtains

$$\begin{aligned}
\begin{pmatrix} A_\mu \\ \rho_\mu^* \end{pmatrix} &\rightarrow \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}\begin{pmatrix} A_\mu \\ \rho_\mu^* \end{pmatrix}, & \tan\theta &= \frac{g_{el}}{g_*}, \\
\begin{pmatrix} \psi_L \\ \chi_L \end{pmatrix} &\rightarrow \begin{pmatrix} \cos\varphi_{\psi_L} & -\sin\varphi_{\psi_L} \\ \sin\varphi_{\psi_L} & \cos\varphi_{\psi_L} \end{pmatrix}\begin{pmatrix} \psi_L \\ \chi_L \end{pmatrix}, & \tan\varphi_{\psi_L} &= \frac{\Delta}{m}, \\
\begin{pmatrix} \tilde{\psi}_R \\ \tilde{\chi}_R \end{pmatrix} &\rightarrow \begin{pmatrix} \cos\varphi_{\tilde{\psi}_R} & -\sin\varphi_{\tilde{\psi}_R} \\ \sin\varphi_{\tilde{\psi}_R} & \cos\varphi_{\tilde{\psi}_R} \end{pmatrix}\begin{pmatrix} \psi_L \\ \chi_L \end{pmatrix}, & \tan\varphi_{\tilde{\psi}_R} &= \frac{\tilde{\Delta}}{\tilde{m}}.
\end{aligned} \tag{1.23}$$

For semplicity we renominate the SM fields as the elementary $\{A_\mu, \psi_\mu, \tilde{\psi}_\mu\}$, and the new physics mass eigenstates as the composite fields $\{\rho_\mu^*, \chi_\mu, \tilde{\chi_\mu}\}$. The mixing angles are parameters of every SM multiplet,

$$\begin{aligned}\theta &\equiv \theta_1, \theta_2, \theta_3 \\ \varphi_{\psi_L} &\equiv \varphi_{q_{Li}}, \varphi_{l_{Li}} \\ \varphi_{\tilde{\psi}_R} &\equiv \varphi_{\tilde{u}_{Ri}}, \varphi_{\tilde{d}_{Ri}}, \varphi_{\tilde{\nu}_{Ri}}, \varphi_{\tilde{e}_{Ri}}\end{aligned} \tag{1.24}$$

where the indices for θ refer to the $SU(3) \otimes SU(2) \otimes U(1)$ factors of the SM gauge group, while the indices i are generational.

Implementing the traformation of equation 1.23 in the total lagrangian we get the effective couplings of the theory. In the case of a heavy gauge boson, ρ^*, interacting with SM fermions we find [4]

$$g(\sin^2\varphi\cot\theta - \cos^2\varphi\tan\theta). \tag{1.25}$$

where the SM gauge coupling g has the form

$$g = \frac{g_{el}g_*}{\sqrt{g_{el}^2 + g_*^2}} \simeq g_{el}, \qquad \text{for } g_{el} \ll g_*. \tag{1.26}$$

Therefore heavy boson production can be obtained in colliders through Drell-Yan scattering

$$q\bar{q} \to \rho^*, \tag{1.27}$$

while electroweak heavy excitations $W^{*\pm}, W^{*3}, B^*$ can also be produced through vector boson fusion. Despite the large coupling to longitudinal W's and Z's, this process is subdominant for large masses.

Once produced, the heavy vectors will mainly decay to pairs of third-generation SM quarks, $t\bar{t}$, $b\bar{b}$ or $t\bar{b}$, and pairs of longitudinally SM vector bosons, $Z_L H$, , $W_L^+ W_L^-$ or $W_L^\pm Z_L$. When kinematically allowed, decays to a SM top or bottom plus one excited top or bottom quark ($T\bar{t}, T\bar{b}$, ect.), or even to any heavy fermions, are also important.

We report for example the branching ratios of the neutral W^{*3} in the approximation of high masses [4]. The following expressions are presented as function of the free parameters of the theory:

$$\begin{aligned}
\Gamma(W^{*3} \to q\bar{q}) = 3\Gamma(W^{*3} \to l\bar{l}) &= \frac{g_2^2 M_*}{32\pi}\tan^2\theta_2, \\
\Gamma(W^{*3} \to t\bar{t}) = \Gamma(W^{*3} \to b\bar{b}) &= \frac{g_2^2 M_*}{32\pi}(\sin^2\varphi_{t_L}\cot\theta_2 - \cos^2\tan\varphi_{t_L})^2 \\
\Gamma(W^{*3} \to Zh) = \Gamma(W^{*3} \to W^+W^-) &= \frac{g_2^2 M_*}{192\pi}\cot^2\theta_2, \\
\Gamma(W^{*3} \to T\bar{t}) = 3\Gamma(W^{*3} \to B\bar{b}) &= \frac{g_2^2 M_*}{32\pi}\frac{\sin^2\varphi_{t_L}\,\cos^2\varphi_{t_L}}{\sin^2\theta_2\,\cos^2\theta_2} \\
&\quad \times \left(1 - \frac{1}{2}\frac{m_*^2}{M_*^2} - \frac{1}{2}\frac{m_*^4}{M_*^4}\right)\left(1 - \frac{m_*^2}{M_*^2}\right).
\end{aligned} \tag{1.28}$$

Some of these branching ratios are plotted in figure 1.3 for a particular choice of the theoretical parameters, in particular the coupling g_* is fixed to $g_* = 4$ [13]. The asymptotic limit corresponds to the high mass approximation of equation 1.28.

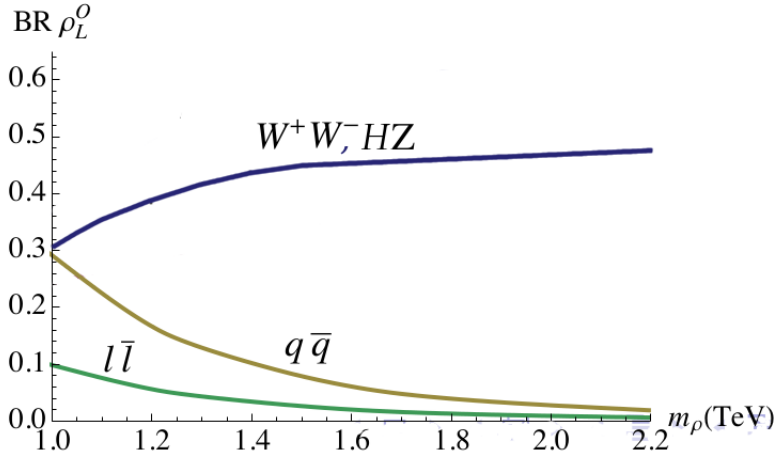

Figure 1.3. Branching ratio of a neutral spin 1 heavy resonance as function of its mass.

Ideally, one would like to measure several of these channels in order to get all parameters and test the model. Events with only two SM quarks in the final states are almost impossible to isolate from the QCD background, the only possible signal of partial compositeness in this channel would be a large violation of flavour universality.

Decays of the heavy W^{*3} resonance in lepton pairs can be separated from the background. This will be briefly discussed in section 5.3. However this branching ratio is sizeable only for relatively small masses, below 1.5 TeV.

Decays into a pair of bosons (W^+W^- and ZH) can also be separated from the QCD background and have large branching ratios[2].

In this thesis we perform a data analysis on the search of new resonances in the ZH channel, with the Z boson decaying in two leptons and the Higgs decaying hadronically ($q\bar{q}$ or gg). We take partial compositeness as the theoretical framework of reference but we follow a sufficiently generic approach so that our analysis can be also applied to other new physics scenarios. Figure 1.4 shows the Feynman diagram of the process where X is the generic resonance to investigate.

[2]Decays in pairs of bosons with identical quantum number are forbidden because of the antisymmetric coupling [14].

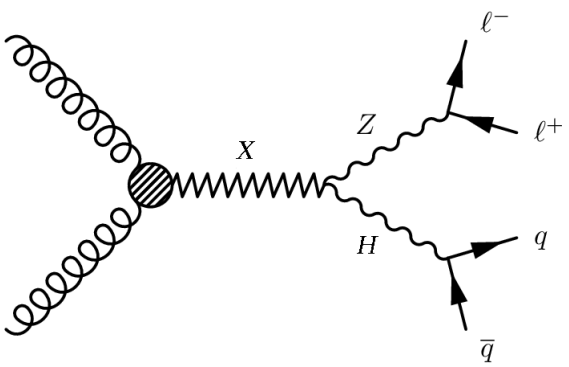

Figure 1.4. Feynman diagram of the process studied in this analysis.

Chapter 2

The Large Hadron Collider (LHC)

The Large Hadron Collider is a proton-proton collider operating at CERN in the 27 km ring tunnel since 2009. Other accelerators, that had been originally built at CERN for previous experiments, work as an injection chain for the LHC. The proton beam starts from LINAC, a small linear accelerator, where its energy firstly reaches 50 MeV. It then passes through a booster, then in the PS where it is accelerated up to 25 GeV, and then in the SPS up to 450 GeV. The beam is finally injected in the LHC ring, where it is accelerated up to 3.5 TeV and to 4 TeV, as regards the 2010-2011 and 2012 datasets respectively. From 2015, instead, it is should reach an energy close to the maximum design value of 7 TeV per beam. The LHC ring and the intermediate steps of the acceleration are sketched in figure 2.1.

The high luminosity of LHC is obtained with a high frequency bunch crossing and a high density of protons per bunch. Each proton beam at full intensity consists of 2808 bunches, and each bunch contains 1.15×10^{11} protons. The transverse dimension of the beam, when it is fully squeezed at the collision points, is around 16 μm. Figure 2.2 shows the total collected luminosity in the years 2011-2012.

One critical aspect of accelerating protons to such high energies is the magnetic field. It is obtained employing a total number of about 9300 superconducting magnets along the ring, operating at a temperature of 1.9 K. The accelerating power is actually given by radiofrequecy cavities working at 400.79 MHz. The energy gain per rivolution is 485 keV, in spite of the 7 keV loss per turn due to synchrotron radiation.

2.1 CMS

The Compact Muon Solenoid (CMS) experiment has been designed to cope the very high rate of interactions expected to take place at the high LHC

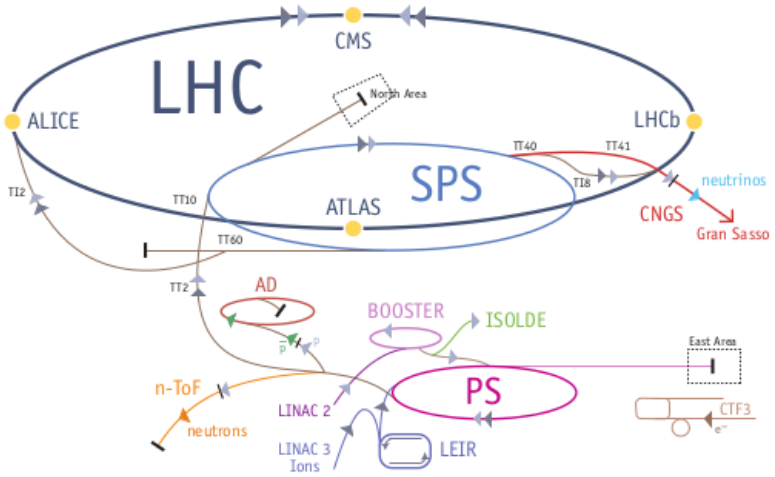

Figure 2.1. Skecth of the LHC accelerator with experiments.

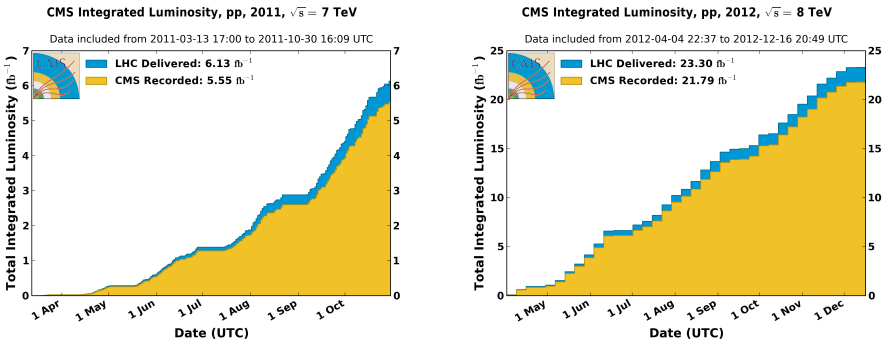

Figure 2.2. Integrated luminosity produced by LHC and recorded by the CMS experiment in the year 2011 (left) and 2012 (right).

luminosity. It has the typical structure of experiments at hadron colliders: it has a central region (*barrel*) enclosed by two disks (*endcaps*) as can be seen in figure 2.3.

LHC proton bunches collide every 25 ns: the detector has so short a time to get ready to analyse a new event[1]. This on-line event selection task is entrusted to the trigger that selects events with interesting features and records them on a disk.

CMS features a powerful superconducting coil, generating a solenoidal magnetic field of ~ 3.8 T in a large volume, which hosts different subdetectors. The magnetic field lines close through steel jokes in the outer region and the distinct subdetectors are designed in order to obtain the highest possible resolution and the largest acceptance for every kind of particles.

[1] 40 MHz is the design frequency. Untill now LHC has been operated at 20 MHz.

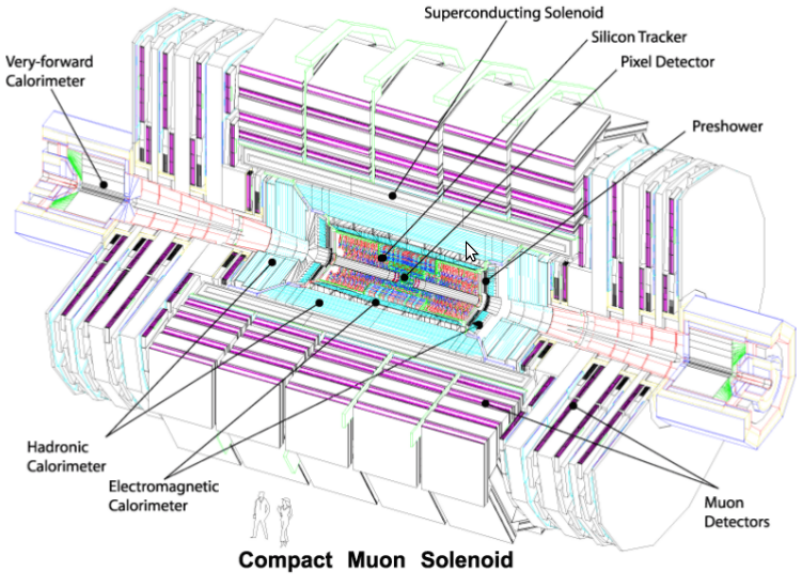

Figure 2.3. Schematic view of the CMS apparatus.

Figure 2.3 shows a schematic view of the CMS detector and figure 2.4 sketches the possible interactions of the particles crossing the detector. The CMS coordinate system is oriented such that the x-axis points south to the center of the LHC ring, the y-axis points vertically upward and the z-axis is in the direction of the beam to the west. The azimuthal angle φ is measured from the x-axis in the xy plane and the radial coordinate in this plane is denoted by r. The polar angle θ is defined in the rz plane and the pseudorapidity is $\eta = -\ln\tan(\theta/2)$. The momentum component transverse to the beam direction, denoted by p_T, is computed from the x- and y-components, while the transverse energy is defined as $E_T = E\sin\theta$.

The different constituents of the CMS detector can be characterized as follows.

2.1.1 Tracker

The tracking system is the innermost subdetector of CMS. It is placed inside the coil of the 3.8 Tesla superconductive solenoid and it is designed to reconstruct the vertices of the interaction and to measure the momentum of charged particles. This system has to work at the full LHC luminosity with very high density of particles and it is composed of three different parts:

- The pixel tracker detector. It is composed of approximately 66 million pixel cells, with size $100 \times 150\ \mu m^2$, that allow a fine 3D vertex

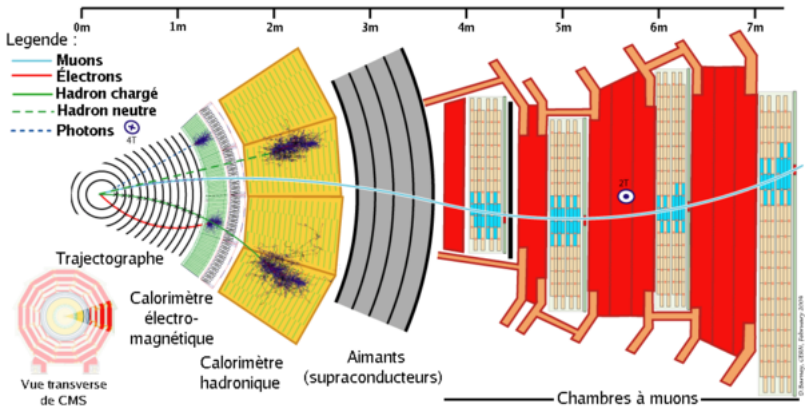

Figure 2.4. Transverse section of the CMS experiment and schematic possible interaction of particles produced in collisions.

reconstruction. Both $r\varphi$ and z coordinates are important, therefore the cells are nearly square-shaped. The pixel detector covers both the barrel and the endcaps with a pseudorapidity range $|\eta| < 2.5$.

- Strip detector. The barrel region is divided into two parts, the Tracker Inner Barrel (TIB) and the Tracker Outer Barrel (TOB). The former is composed of four layers of silicon sensors with a thickness of 320 μm and of strip pitches varying from 80 to 120 μm. The TOB is made of six layers. In this subdetector thicker silicon sensors (500 μm) are employed, while the strip pitch varies from 120 to 180 μm. The end cap region ($|\eta| > 1.6$) is covered by the Tracker Inner Disks (TID) and the Tracker End Cap (TEC).

A track at CMS has on average three hits in the pixel detector and about ten hits in the strip detector.

Thanks to the bending power of the magnetic field, the tracker provides a measurerement of the transverse momentum of charged particles by the relation

$$p_T[\text{GeV}] = 0.3 \cdot B[\text{T}] \cdot R[\text{m}] \tag{2.1}$$

where R is the radius of the helicoidal trajectory. The resolution obtained by the tracker is dominated by two components: a term proportional to the transverse momentum due to the uncertainty on the curvature measurement, and a second term independent of the momentum due to the multiple scattering in the material. In order to keep the latter source of uncertainty as small as possible, it is important to minimize the thickness of the material crossed by the detected particle. In figure 2.5 the number of radiation lengths of the tracker is reported as function of η.

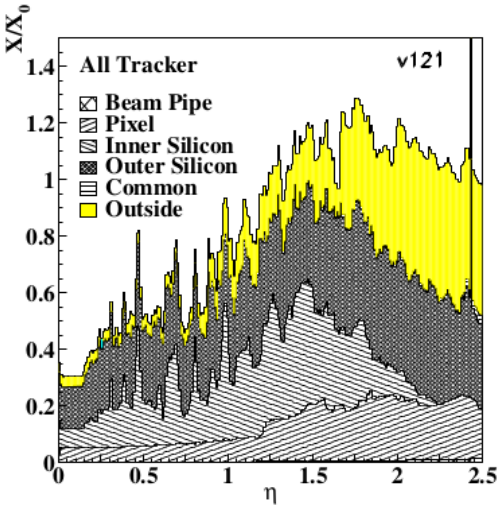

Figure 2.5. Material radiation lengths of the barrel as function of $|\eta|$ in CMS.

2.1.2 Electromagnetic calorimeter

The Electromagnetic Calorimeter (ECAL) measures the energy of photons, electrons and positrons. It it is placed immediately outside the tracker, but still inside the solenoid. ECAL is made of 74848 lead-tungstate ($PbWO_4$) crystals: this material is characterized by a high density (8.28 g/cm^3), which gives the crystals a very compact form and makes them particularly suitable to be placed inside the magnetic coil.

This material has also a fast temporal response ($\sim$ 10 ns) and its radiation length (X_0) of 0.89 cm and Moliere radius of 2.2 cm give ECAL the possibility to fully contain the expansion of the electromagnetic shower. The barrel crystals have a front face area of 2.2 $\times$ 2.2 cm^2, a length of 23 cm (25.8 X_0) and they are positioned at r = 1.29 m. The crystals in the end caps have a 2.47 $\times$ 2.47 cm^2 front face, a 22 cm length and they are positioned at z = 3.17 m. For trigger purpose the ECAL crystals are grouped together into 5 $\times$ 5 trigger towers.

A pre-shower device is placed in front of the endcaps. It is made of two disks of lead absorber at $2X_0$ and $3X_0$, and of two planes of silicon strip detectors. It allows the rejection of photon pairs from π_0 decays and improves the estimation of the direction of photons, to enhance the measurement of the two-photon invariant mass.

The energy resolution of the ECAL is given by three different contributions:

$$\frac{\sigma_E}{E} = \frac{a}{\sqrt{E}} \oplus \frac{b}{E} \oplus c \tag{2.2}$$

where the first term is statistical in nature and contains fluctuation in showering and in the amplification through photodiodes (a=1.8 %), the second one considers electronic noise and pile-up (b=4 %) and the last term is mainly due to the calibration (c=0.5 %).

2.1.3 Hadronic calorimeter

The hadronic calorimeter (HCAL) is placed just outside ECAL and inside the magnet coil. It plays an important role in the reconstruction of jets and missing energy. The design is strongly influenced by this aim, hence an important requirement is the high hermeticity (the ability to capture every particle emerging from the collisions). This means that the detector must cover the biggest possible portion of the solid angle.

For this reason, the barrel and endcaps are complemented by a hadron forward calorimeter, which is placed outside the magnet return yokes, with a total coverage of $3 < |\eta| < 5.31$ at 11 m from the interaction point. Moreover, an outer hadronic calorimeter is placed in the first muon absorber layer in order to enhance the containement of high energy jets in the central region of the detector.

HCAL is a sampling calorimeter, whose active elements are plastic scintillators interleaved with brass absorber plates and read out by wavelength-shifting fibres. Brass has been chosen as absorber material for its short interaction length and because it is non-magnetic.

The thickness of the absorber layers is between 60 mm in the barrel and 80 mm in the endcaps. The barrel ranges between 5.46 interaction lengths at $\eta = 0$ to 10.82 at $\eta = 1.3$, while the endcaps have an average of 11 interaction lengths.

The photodetection readout is based on multi-channel hybrid photodiodes, able to operate in a high magnetic field, that give an amplified response, proportional to the original signal, for a large range of particle energies.

The HCAL energy resolution (E in GeV) is

$$\frac{\sigma}{E} \simeq \frac{a}{\sqrt{E}} \oplus 5\% \tag{2.3}$$

where $a \simeq 65\%$ in the barrel, $a \simeq 85\%$ in the endcaps and $a \simeq 100\%$ in the hadron forward calorimeter.

2.1.4 Muon chambers

The efficient detection of muons is of primary importance (to such an extent that it named the experiment), as they represent a clear signature for a large number of processes. The muon system fulfills three purposes: muon identification, momentum measurement, and triggering. It is hosted

in the return yoke, thanks to which the detectors are shielded from charged particles other than muons. Indeed the muon detector is separated from the interaction point by 16 interaction lengths and the only particles that are able to pass through so much material are muons. Before entering the calorimeters and the iron joke, the muon momentum is also measured in the inner tracking system.

The measurement of the momentum of muons, using the muon system only, is performed through the determination of the muon bending angle at the exit of the coil, taking the beam line (known with a precision of $\sim 20\,\mu m$) as the origin of muons. It measures the momentum with a resolution of about 10 % up to momenta of 1 TeV. The minimum energy required to cross the muon system is about 5 GeV.

To identify and measure muons, three types of subdetectors are used. In the barrel ($|\eta| < 1.2$), where the muon rate and the residual magnetic field are low, four layers of drift tube chambers (DT) are used. The chamber segmentation follows that of the iron yoke, that consists of 5 wheels along the z axis, each one divided into 12 azimuthal sectors. Each chamber has a resolution of $\sim 100\,\mu m$ in $r\varphi$ and 1 mrad in φ. In the two endcaps ($0.8 < |\eta| < 2.4$), where the muon rate and the residual magnetic field are higher, cathode strip chambers (CSC) are used. In each of the endcaps, the chambers are arranged in 4 disks perpendicular to the beam, and in concentric rings (3 rings in the innermost station, 2 in the others). Each chamber has a spatial resolution of about 200 μm (100 μm for the first station of chambers) and an angular resolution of about 10 mrad. In both the barrel and the endcaps, a system of resistive plate chambers (RPC) is installed to ensure redundancy to the measurement. RPCs provide a rougher spatial resolution than DTs and CSCs, and a fast response with a good time resolution which is useful for triggering.

2.1.5 Trigger

At design luminosity LHC produces a huge number of interactions (10^9 events/second), while the storage system is able to save events only with a rate of about 1000 Hz. On the other hand, only a minimal part of the interactions produce phenomena that turn out interesting to study. The task of the trigger system is to reduce the storage rate while keeping a high efficiency on the potentially interesting events.

Triggering procedure in CMS is implemented in two levels.

- Level-1 trigger: very fast response (3.2 μs), it has to reduce the rate from 20 MHz (with 30 events per bunch crossing) to 10^5 Hz. The L-1 trigger involves the calorimetric measurements and the muon system without looking at the tracker. The trigger decision is based on the so called "trigger primitive", that is the presence of objects like electrons,

photons, muons, and jets with a E_T or a p_T above a given threshold. L-1 trigger is able to identify within an event the presence of the following objects: muons, isolated electrons or photons, jets, τ-jets, missing energy and multiple objects (2 muons, 4 jets, etc.).

- High level trigger: HLT uses all the information from the detector, including tracker, pre-shower and maximal resolution of the calorimeters. The reconstruction algorithms are the same of the off-line analysis. However triggering procedure doesn't need maximal precision and therefore these algorithms are modified in order to be faster, even if with a lower resolution. Furthermore, reconstruction information is required only for a restricted region of the detector, identified by the L-1 trigger.

2.1.6 Muons

The muon reconstruction algorithm at CMS takes advantage of the redundancy of detection methods. The first step is track reconstruction, done independently in the tracker ("tracker track") and in the muon system ("standalone-muon track"). Then, one tries to match the two types of tracks.

Two different approaches can be used here:

- *Outside-in*: starting from a standalone-muon track, a matching tracker-track is found; then, the fit of the track is repeated using the hits both in the tracker and in the muon system. The resulting object is called a *global muon*, and its resolution is improved, at high p_T , with respect to the tracker-only fit;

- *Inside-out*: each tracker track is extrapolated to the muon system region; this is a delicate step, because it is necessary to take into account the energy loss and the uncertainty due to the multiple scattering in a large amount of material (calorimetry and coil). If a muon segment (i.e. a subset of a real track in the muon system) is found to match the extrapolation, this object is called a *tracker muon*. This algorithm is useful for low-p_T muons, that are not expected to fully penetrate the muon system, and therefore do not generate a full standalone-muon track, but only a few hits.

If no match is found when extrapolating outside-in, the standalone-muon track is stored as a *standalone muon*. This happens only for less than 1% of the muons produced in a collision [15].

2.1.7 Electrons

Electrons have a less distinguishable signature than muons in the detector: since the electrons are stopped inside the ECAL, they are more affected by

jet induced background. The relevant subdetectors for electron reconstruction are the inner silicon tracker and the electromagnetic crystal calorimeter. Both of them have a very good energy resolution and high granularity.

Similarly to the case of muons, there are two different approaches to build up a reconstruction algorithm for electrons [16]:

- *Tracker seeding:* the track is fitted starting from the tracker hits, and afterwards the calorimeter information is added; this method gives the best results for low-pT electrons inside jets.
- *ECAL-driven seeding:* first of all, one reconstructs an ECAL Super-Cluster, that is a group of clusters of energy deposits in the calorimeter. Then the information about the characteristic width in η and φ (where the latter depends also on the bending of the track due to the magnetic field) of the electron cluster is taken into account. Afterwards, the supercluster position is used to make a match to tracker seeds, and finally a global fit is performed, using an appropriate modelling of energy loss in the tracker material (not negligible for electrons). This method is the most adequate for electrons in a higher p_T range.

2.1.8 Jets

Since gluons and quarks cannot exist in free state, they fragment into hadrons. The result of this fragmentation is a jet of particles depositing energy in the detectors.

CMS exploits a complex and powerful algorithm called particle flow [17] to measure at best all particles included in the jet exploiting the high granularity of the detector.

There is a wide range of algorithms for jet reconstruction which takes into account also biases coming from detectors effects (jet energy calibration). Cone-based algorithms define a metric $\Delta R = \sqrt{(\eta_1 - \eta_2)^2 + (\phi_1 - \phi_2)^2}$ based on the pseudorapidity and the azimuthal angle, while k_T based algorithms use a E_T metric. The former uses any input object to look for the most stable cone, without any E_T threshold. To find the stable cones, the four-momenta of the input objects into a cone are added together and the direction of the cone is compared to the summed four-momentum of the objects enclosed in this cone. A cone is considered stable when the cone axis is aligned with the direction of the sum of the three momenta of the particles in the cone.

On the other hand, the k_T based algorithm uses all the informations available from all detectors and subdetectors to measure the energy and the momentum of each particle which leaves a signal. There are two important requirements for the stability of jet algorithms:

- *Collinear safe*: collinear splitting shouldn't change jets
- *Infrared safe*: soft emissions shouldn't change jets

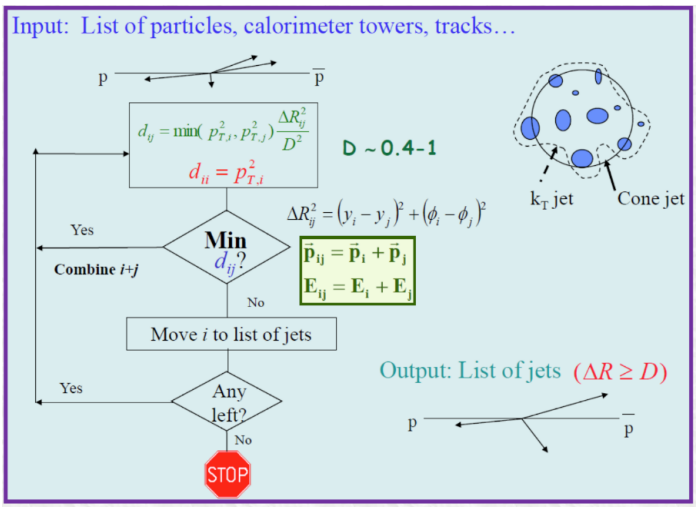

Figure 2.6. Sketch of the k_T algorithm [17].

and the k_T algorithm respects both.

k_T algorithm

k_T algorithms are extensively used in the CMS experiment for the reconstruction of jets. It is implemented as follows: for each input object i and for each pair of input objects i, j two quantities (distances) are calculated:

$$d_{ii} = p_{T,i}^{2p} \tag{2.4}$$

$$d_{ij} = \min(p_{T,i}^{2p}, p_{T,j}^{2p})\frac{\Delta R_{ij}^2}{D^2} \tag{2.5}$$

where D is a parameter related to the tipical size of the jet and p can be chosen as follows:

- $p = 1$: regular k_T jet algorithm;
- $p = 0$: Cambridge/Aachen jet algorithm;
- $p = -1$: Anti-k_T jet algorithm;

The algorithm checks the smallest value among d_{ii} and d_{ij}; if it is d_{ii} the first object is removed from the list of candidates and inserted into the list of jets, while if it is d_{ij} the two candidate jets are merged. The procedure is repeated until all jets are found. Figure 2.6 sketches the k_T algorithm and figure 2.7 shows the resolution of the reconstructed jet [17] [18].

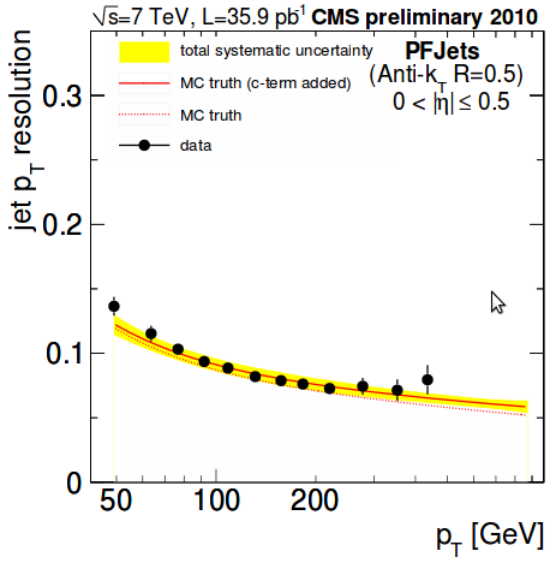

Figure 2.7. Particle flow jet resolution σ_{p_T}/p_T [18].

Chapter 3

Physics tools in the analysis

As introduced in section 1.2.1, this analysis searches for signal of partial compositeness in the $X \to ZH$ channel. This should be one of the two favourite decays to test the model. In particular, the Z boson is selected leptonically (with electron or muon final state) while the Higgs is chosen to decay fully hadronically ($q\bar{q}$ or gg). Combining the two branching ratios, one gets only BR($ZH \to llq\bar{q}(gg)$) = 4.65%.

Despite the small final branching ratio, this channel is found to be a reasonable compromise between a strong signature and an acceptable statistics. The two leptons are easily identified by the detector and limit the presence of the background, while the hadronic Higgs decay collects the largest possible fraction of Higgs events but it needs a more elaborated off-line reconstruction.

This chapter contains a study at generator level of the signal and background simulated events. This is a fundamental step of the analysis because it checks the accuracy of the MC simulation and allows us to study in detail the physical process under consideration. Considerations on the particular topology of the events at high mass are also presented, as well as the tools needed to select the jet of the Higgs decay.

3.1 Monte Carlo Samples and Data Sets

3.1.1 Signal MC

As signal hypotesis, the heavy resonance (X) is tested using a wide set of masses from 800 GeV to 2000 GeV, one masspoint every 100 GeV (Table 3.1). The signal is simulated by MADGRAPH5 [19] as a narrow spin 1 resonance and is forced to decay in the $X \to ZH \to llqq$ channel. Showering and hadronization are performed with PYTHIA [20] version 6.426.

Sample	Number of events
Zprime_ZH_llqq_M800	10710
Zprime_ZH_llqq_M900	10209
Zprime_ZH_llqq_M1000	19997
Zprime_ZH_llqq_M1100	9370
Zprime_ZH_llqq_M1200	10710
Zprime_ZH_llqq_M1300	9369
Zprime_ZH_llqq_M1400	10497
Zprime_ZH_llqq_M1500	19999
Zprime_ZH_llqq_M1600	8950
Zprime_ZH_llqq_M1700	9369
Zprime_ZH_llqq_M1800	10708
Zprime_ZH_llqq_M1900	10498
Zprime_ZH_llqq_M2000	19999

Table 3.1. Signal samples used in the analysis.

3.1.2 Background MC

Since we are looking for new resonances decaying in semileptonic final state, the background of this analysis is originated by all SM events with two leptons and at least one jet as final state. By far, the dominant contribution is the production of Z boson with jets. This Z+jets background is produced by MADGRAPH and divided into two samples depending on the Z p_T, higher than 100 GeV or between 70 and 100 GeV. The contribution of events with Z p_T less than 70 GeV is found to be negligible due to further cut on the objects p_T in the selection requirements.

Other sources of background considered are $t\bar{t}$ production, generated by POWHEG [21], and di-boson production (WW, WZ and ZZ) generated by PYTHIA6. SM direct non resonant ZH production trough Higgs-strahlung mechanism is totally irrelevant due to its much smaller cross section of $\sim$ 0.4 pb.
Samples and related statistics are reported in Table 3.2.

3.1.3 Data Samples

In this analysis we use 19.8 fb^{-1} of data collected in 2012, some of these datasets were reprocessed with other parked data samples and take the 'parked' epithet in the name[1].

All datasets are collected with a double muon or a double electron trigger,

[1]A parked dataset is a sample of data collected with a lower trigger threshold. This lower threshold corresponds to higher rate, too high to be reconstructed and stored. Therefore these datasets are parked waiting for a later reconstruction.

Sample	Number of events	σ[pb]
DYJetsPt100	12511326	39
DYJetsPt70To100	11764538	63
$t\bar{t}$	6540800	225
WW	10000431	57
WZ	9955839	33
ZZ	9799908	8.0

Table 3.2. Background samples used in the analysis.

as explained in detail in chapter 4. The trigger algorithm employed for the electron samples doesn't use any information from the tracker but only the energy deposite in the ECAL. This expedient is implemented in order to avoid any possible inefficiencies due to the presence of two tracks very close to each other when the Z has high p_T and its decay products are very collimated. Such a trigger is contained in the Photon/DoublePhotonHighPt dataset.
All samples are listed in Table 3.3.

Sample	L[pb]$^{-1}$
Photon_Run2012A_22Jan2013	889
DoublePhotonHighPt_Run2012B_22Jan2013	4429
DoublePhotonHighPt_Run2012C_22Jan2013	7152
DoublePhotonHighPt_Run2012D_22Jan2013	7318
DoubleMu_Run2012A_22Jan2013	889
DoubleMuParked_Run2012B_22Jan2013	4426
DoubleMuParked_Run2012C_22Jan2013	7114
DoubleMuParked_Run2012D_22Jan2013	7318

Table 3.3. Data samples used in the analysis.

3.2 Event topology

This analysis is performed in a high mass region (TeV scale). Figure 3.2 shows the rapidity distribution of the resonance generated by simulation, defined as

$$y = \ln \frac{E + p_z}{E - p_z}. \tag{3.1}$$

The rapidity can be expressed also in term of the fractions of the momentum of the proton carried by the colliding partons, x_1 and x_2. The relationship

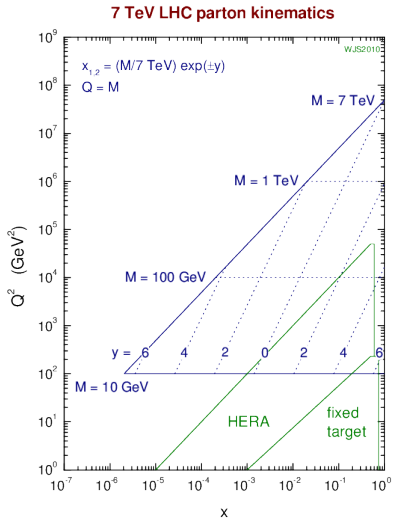

Figure 3.1. Parton kinematics at $\sqrt{s}$=7 TeV.

is

$$y = \frac{1}{2} \ln \frac{x_1}{x_2} \tag{3.2}$$

and is strictly related to the energy available in the center of mass of the partons collision

$$s' = M_X^2 = x_1 x_2 s, \tag{3.3}$$

where $\sqrt{s}$ is the energy in the laboratory frame of reference.

Figure 3.1 shows the available range for the x_1 and x_2 quantities and the consequent rapidity value. From this constraint the minimun possible fraction of the proton momentum carried by the partons is $x_{min} = M^2/s$. For lower masses the x_1 and x_2 range available is larger and the collision can be more asymmetric.

Finally figure 3.2 presents the rapidity distribution of the generated heavy resonance, it can be easily seen that the higher the signal mass is, the more central the resonance is produced.

Madgraph generates the resonance produced in the collision with $p_T = 0$. In the next step of the simulation, during the hadronization, Pythia adds the QCD ISR (initial state radiation) and consequently a resonance p_T different from 0. The p_T distribution of the heavy resonance after the Pythia simulation is shown in figure 3.3. The typical p_T is small compared to the mass of the resonance, in fact two thirds of the events have p_T smaller than 50 GeV.

The $X \to ZH$ process is a two body decay and, in the heavy resonance frame of reference, the energy of its decay products Z and H are univocally defined. In figure 3.4 the p_T distributions of Z and H at generator level are reported. These distributions have a Jacobian peak close to $M_X/2$, that

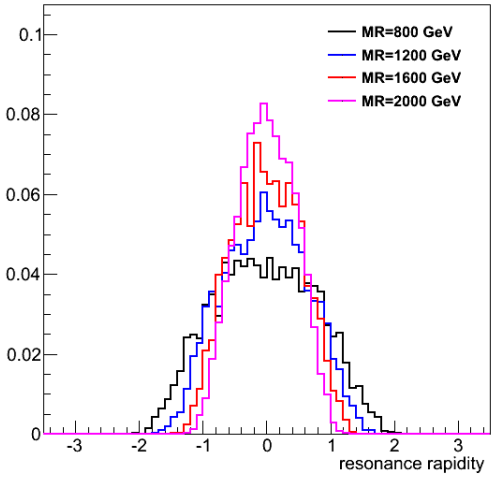

Figure 3.2. Spin 1 resonance rapidity distribution at generator level for different masses.

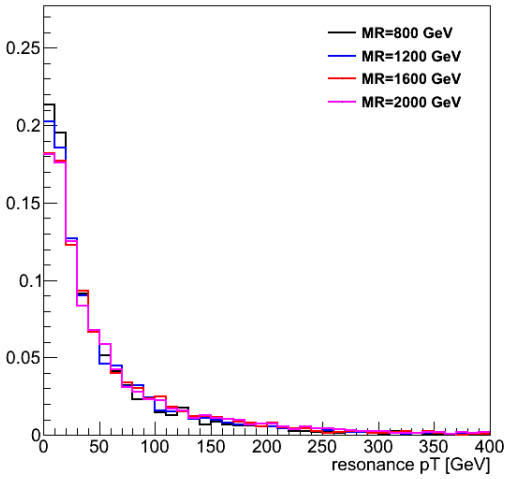

Figure 3.3. Spin 1 resonance p_T distribution at generator level for different masses.

corresponds to the value in p_T of a two body decay with massless products. In our case, the mass of the produced particles, Z and Higgs bosons, is about 91 and 125 GeV respectively [22] and their mass can be considered negligible compared to the mass of the decayed resonance (TeV scale).

The tail on the right of the peak is due to non-zero p_T of the original resonance, while the smooth shape at lower p_T is due to events with a sizable

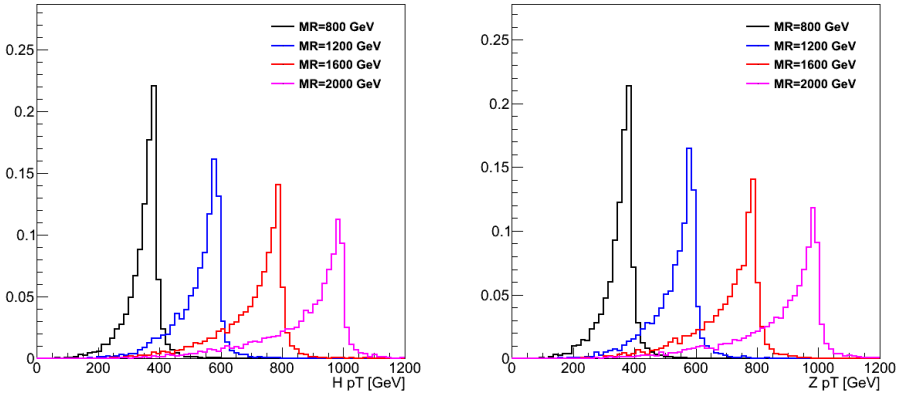

Figure 3.4. p_T distribution of H (left) and Z (right) at generator level for different heavy resonance masses.

p_z component.

Figure 3.5 shows the ΔR distribution at generator level of the Higgs and Z decay products respectively, where for the Higgs decay we simply consider the direction of the partons. It is defined as

$$\Delta R = \sqrt{d\phi^2 + d\eta^2}. \tag{3.4}$$

Both are peaked at

$$\Delta R = 2\frac{m_{H(Z)}}{p_{T,H(Z)}} \simeq 4\frac{m_{H(Z)}}{M_X}, \tag{3.5}$$

that corresponds to the configuration in which the final particles are emitted perpendicularly to the direction of motion of the Higgs or Z; this configuration is preferred by the phase space (Jacobian peak). For masses larger than 1000 GeV ΔR is often smaller than 0.5, that is the tipical size of the jet cone (section 2.1.8).

This consideration leads us to a particular event topology, where the jets produced from the quarks or gluons hadronization merge in a single fat jet. The process is sketched in figure 3.6.

3.3 H-tagging

It is now clear how the SM Z+jets production, where the Z decays leptonically, represents the main background in the analysis. In fact, these events have the same topology but the jet is generated by different processes: jets from background events are produced by one single parton, while jets from the signal samples are generated by a pair of quarks or gluons. So it is impor-

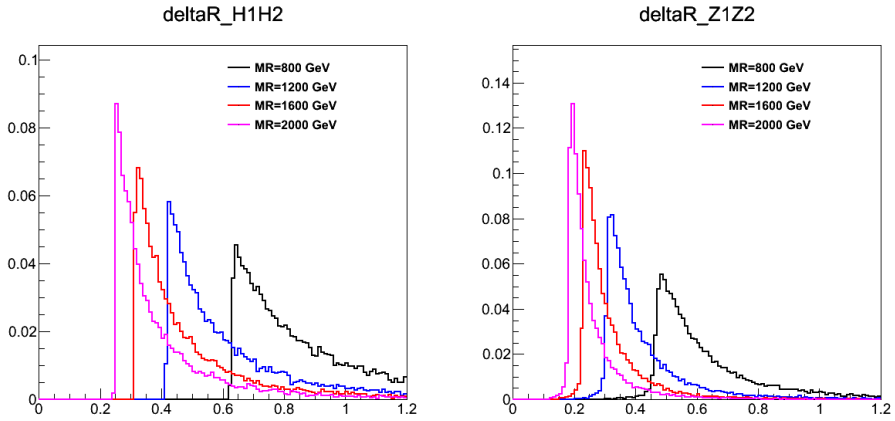

Figure 3.5. Delta R between the two quarks of the Higgs decay (left) and the two leptons from the Z (right) at generator level for different heavy resonance masses. The spatial separation of the Higgs products is simply computed taken into account the generated partons directions.

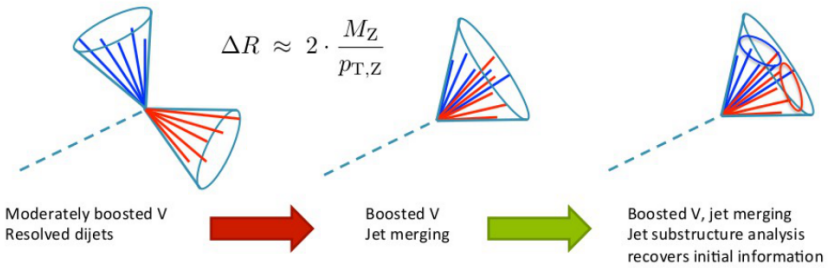

Figure 3.6. Highly boosted jet merging.

tant to distinguish as much as possible jets produced by QCD interactions from merged jets produced in the Higgs decay.

The variables most often used in analyses involving jets are usually the jet direction and the transverse momentum (p_T), but they cannot distinguish signal from background. However, being the jets composite objects, their mass and internal structure contain additional information. The jet mass is defined as the invariant mass of all the objects contained inside the jet: the pion mass is associated to charged hadronic tracks, while the reconstructed photons are considered massless.

The identification of jets produced by Higgs decay is based on two ideas:

- The jet mass: jets produced by the decay of a massive particle should have the invariant mass around the nominal mass of the original particle. Oppositely, jets originated by QCD radiation are produced by the emission of quarks or gluons and typically have smaller invariant

mass.

- The identification of jet substructures: looking inside the stucture of jets can help the discrimination of the original seed of the jet. Indeed, H-jets are produced by two partons merged into a single fat jet.

3.3.1 Jet grooming algorithms

The jet mass is the main observable in distinguishing a H-jet from a QCD jet. Jet grooming consists in the suppression of uncorrelated UE/PU (underlying event and pile-up) radiation from the target jet and improves the discrimination pushing the jet mass for QCD jets towards lower values while mantaining the jet mass for V(H)-jets around the boson-mass. [23] [24]

Tipically three different grooming algorithms are considered, trimming [25], filtering [26] or pruning [27] [28]. Studies of these different grooming methods in CMS are presented in [29].

Trimming algorithm Trimming is a technique that ignores regions within a jet that falls below a minimum p_T threshold. Trimming reclusters the jet's constituents with a radius R_{sub} and then accepts only the subjets that have $p_{T,sub} > f_{cut}$, where f_{cut} is tipically taken proportional to H_T, the scalar sum of the p_T of all jet reconstructed in the event.

Filtering algorithm This procedure provides a hierarchical structure for the clustering like the K_T algorithm, but in angles rather than in relative transverse momenta. It creates a series of n new subjets $s_1, s_2 ... s_n$ ordered in descending p_T. The final jet is redefined as the sum of the four-momenta of the three highest p_T subjet: $\sum_i^{\min(n,3)} s_i$.

Pruning algorithm The idea is to take a jet of interest and then to recluster it using a vetoed sequential clustering algorithm. [23]
Clustering proceeds as explained in section 2.1.8, but it is vetoed if the particles are too far away in ΔR

$$\Delta R_{ij} > D_{cut} = \alpha \frac{m_J}{p_{T_J}} \tag{3.6}$$

and the energy sharing is too asymmetric

$$z_{ij} = \frac{\min(p_{T_i}, p_{T_j})}{p_{T_{i+j}}} < z_{cut}, \tag{3.7}$$

where z_{cut} and α are parameters of the algorithm. If both these conditions are satisfied the softer of the two particles is not considered.

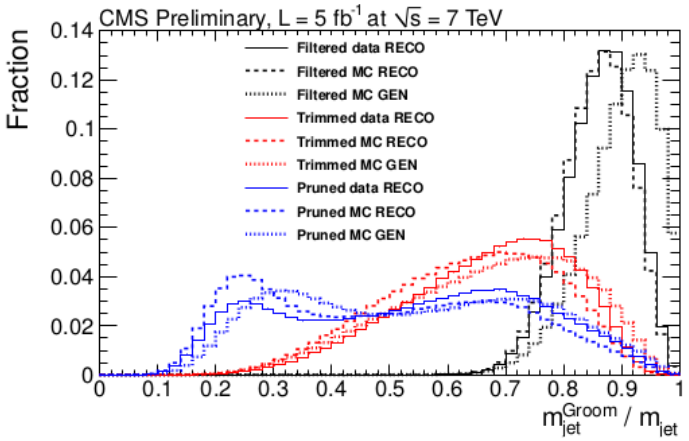

Figure 3.7. Comparison of the jet mass in generic QCD events from the groomed jets divided by the jet mass of matched ungroomed jets for the three grooming techniques, for both data and the PYTHIA 6 Monte Carlo. [30] Events are collected with a single jet trigger.

In figure 3.7 the effect on the jet mass for the different grooming algorithms is shown. This study has been performed at $\sqrt{s} = 7$ TeV with 5 fb^{-1} of data [30]. Events are triggered by a single calorimeter jet.

In general, the filtering algorithm is the least aggressive grooming technique, with groomed jet masses close to the original case. The trimming algorithm is moderately aggressive and produces a much wider final mass distribution. Pruning is the most aggressive tecnique and a bimodal distribution begins to appear: in cases where the pruned jet mass is small, jets usually have most of their energy configured in core components with little gluon radiation, which leads to narrow jets. Instead, when the pruned jet mass is large, the jets are split more symmetrically.

In this analysis we use the pruned jet mass because of its capability to improve the jet mass resolution and backgroun rejection (section 3.3.2).

3.3.2 Pruned jet mass

We can now proceed and study the effect of the pruning algorithm on the MC signal and evaluate the discrimination compared to the background. Figure 3.8 shows the jet mass and the pruned jet mass distributions for signal and MC background. The two plots are obtained with a full simulation of the detector and for a resonance mass of 1000 GeV. The signal histogram is scaled arbitrarily in order to be visible compared to SM background and is fitted with a gaussian distribution. As expected, the jet mass of the dominant Z+jets background decreases exponentially at high masses. On the other hand, SM di-boson productions (ZZ, ZW, WW) presents a peak around the vector boson mass but their cross section is so small that their

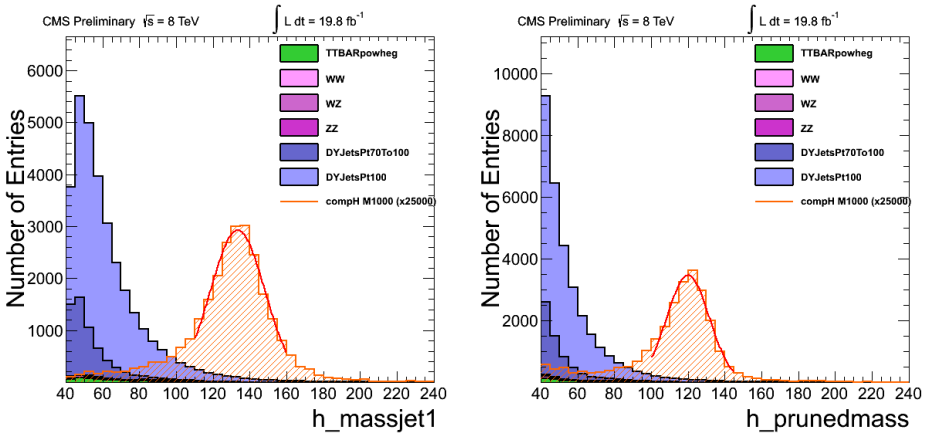

Figure 3.8. Mass (left) and pruned mass (right) of the jet for MC signal and background.

	Jet mass	**Jet pruned mass**
Sig. mean ($\bar{m}$) [GeV]	133.3	119.9
Sig. sigma (σ) [GeV]	14.72	11.58
Num. of MC background events in $[\bar{m} \pm 3\sigma]$	3684	1975
Num. of MC signal events (1 fb $\times$ 2500) in $[\bar{m} \pm 3\sigma]$	23692	21799

Table 3.4. Mean value and standard deviation of the jet mass distribution before and after the pruning algorithm. The number of expected events in computed in a windows of three sigma around the mean value of the signal.

contribution cannot be seen in a plot with linear scale.

In order to evaluate the effect of the pruning algorithm, one can compare the signal over background ratio around the signal region. Table 3.4 shows the fit parameters of the original and pruned jet mass and the total amount of signal and background yields in a window of three sigma around the signal peak. The application of the pruned algorithm reduces the number of background events in the signal window almost to one half, while the 92 % of the signal is kept. Moreover, the pruning procedure slightly improves the resolution of the jet mass.

We can now investigate in detail how the Higgs boson is reconstructed by observing the merged jet. It could be interesting to consider the possible Higgs decay channels separately, since quarks and gluons hadronize in different ways due to different color charge. Figure 3.9(a,b,c) presents the

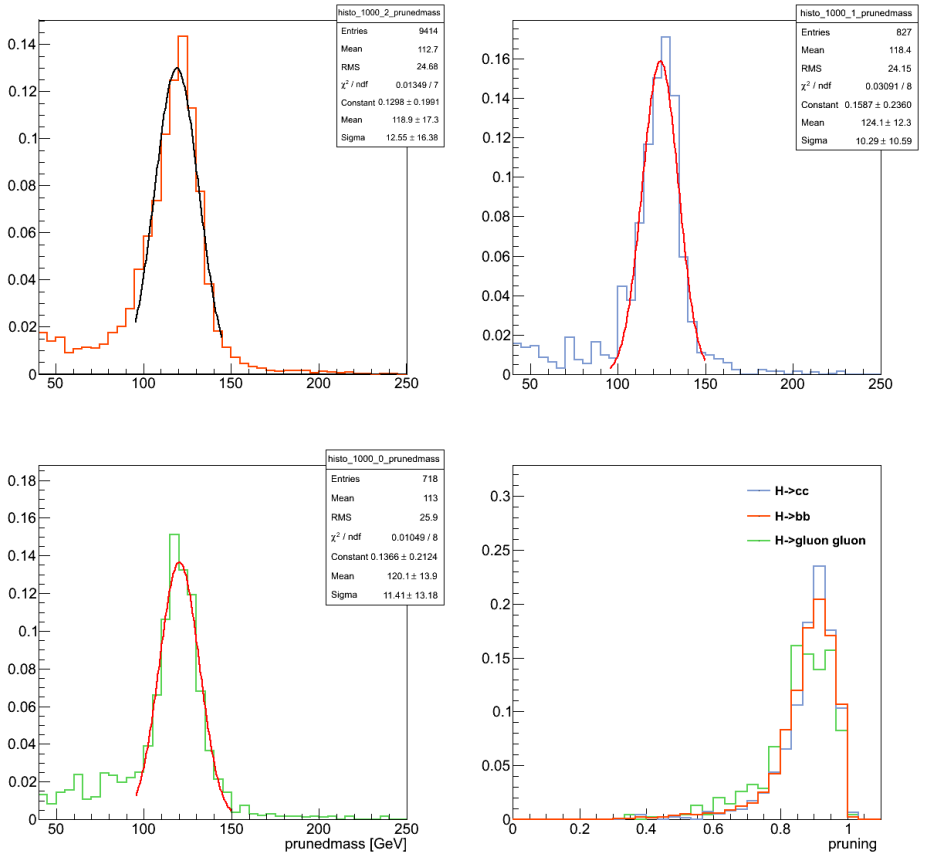

Figure 3.9. Pruned mass distribution fitted separately for the three Higgs decay channel considered: $b\bar{b}$ (top left); $c\bar{c}$ (top right); gg (bottom left). The bottom right plots shows the ratio of the pruned mass over the original jet mass. All the samples are simulated for a resonance mass of $M = 1000$ GeV.

reconstructed pruned jet mass for different signal decay modes. A gaussian fit is performed in the central core of the distribution. Figure 3.9(d) shows the fraction of the final pruned mass respect to the original jet mass. A small difference between the two $H \to q\bar{q}$ and the $H \to gg$ samples can be observed: amoung the three, the gluon decay case loses the highest fraction of the original jet mass. The output fit parameters are summarized in table 3.5.

No significant differences are observed. The peak of the $H \to c\bar{c}$ sample is slighly shifted at higher mass value and the width of the $H \to b\bar{b}$ channel is slightly larger. Figure 3.10 compares the three distributions and shows the total shape of the pruned jet mass obtained summing up all the signal channels.

The resolution of the reconstructed jet mass for different signal mass

[GeV]	$b\bar{b}$	$c\bar{c}$	gg
Signal MC mean	118.9	124.1	120.1
Signal MC sigma	12.55	10.29	11.41

Table 3.5. Mean value and standard deviation of the pruned jet mass distribution. The fit is performed separately for the three signal samples.

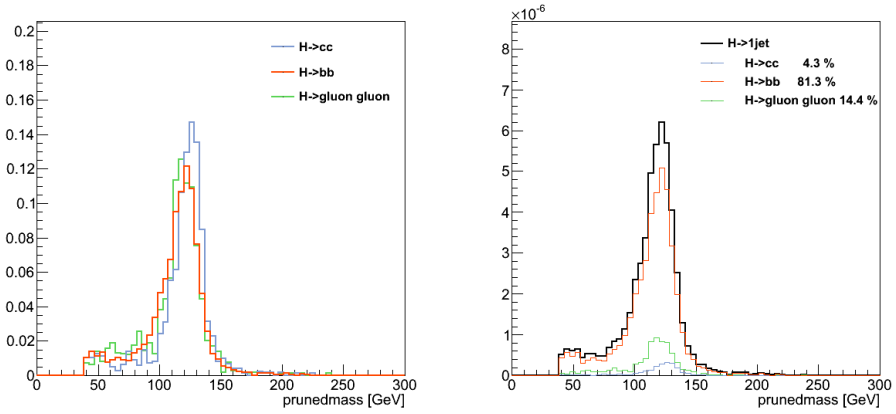

Figure 3.10. Normalized (left) total (right) jet pruned mass distribution in the different Higgs decay channels for a signal sample of $M = 1000$ GeV.

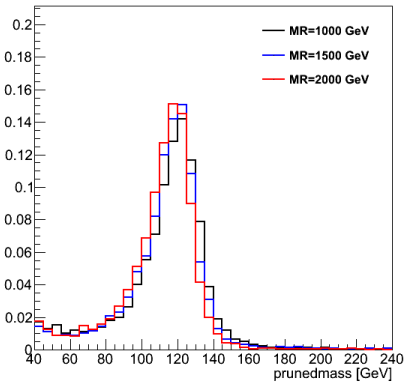

Figure 3.11. Pruned mass jet distribution for different signal mass hypotheses.

hypotheses is studied in figure 3.11. Both the mean value and the resolution of the pruned jet mass are stable as function of the resonance mass.

Figure 3.12 shows the correlation between the pruned mass of the Higgs boson candidate and of the dominant background (DY + jets, with $Z/\gamma p_T >$ 100 GeV) and their transverse momenta. The p_T threshold at 40 GeV is

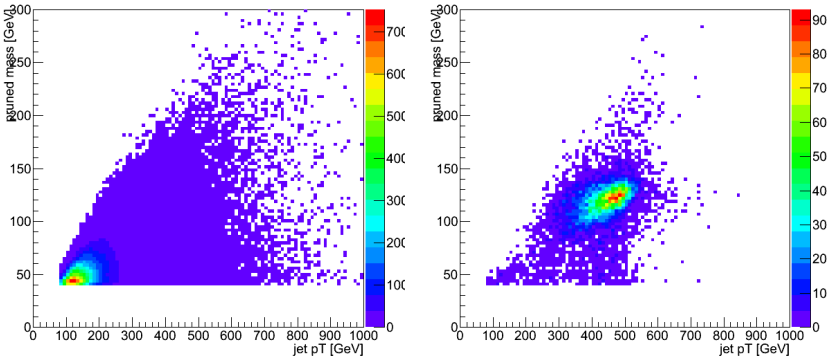

Figure 3.12. Correlation between jet pruned mass and p_T for a DY sample with $Z/\gamma p_T > 100$ GeV (left) and a heavy resonance of $M = 1000$ GeV (right).

given by the off-line selection (section 4.5).

3.3.3 N-subjettiness

In order to further discriminate signal from background, it useful to investigate the inner structure of the jet. Studying the distribution of the jet constituents with respect to the jet axis allows us to test the hypothesis of the existence of multiple substructures, that could be evidence of jets originated by more than one parton. This procedure proceeds as follows: the costituents of the jet are clustered again with the usual algorithm, however the procedure is stopped when one obtains N subjets. Then, a new variable, the N-subjettiness, is introduced. It is defined as

$$\tau_N = \frac{1}{d_0} \sum_k p_{T,k} \min((\Delta R_{1,k})^\beta, (\Delta R_{2,k})^\beta ... (\Delta R_{N,k})^\beta), \tag{3.8}$$

where β is an arbitrary parameter, the index k runs over the jet constituents and the distances $\Delta R_{n,k}$ are calculated with respect to the axis of the $n-th$ subjet. The normalization factor d_0 is calulated as $d_0 = \sum_k p_{T,k} R_0^\beta$, setting R_0 to the radius of the original jet.

The N-subjettiness is always included in the interval from 0 to 1 and represents the compatibility of the jet structure with an N-subjet hypothesis: small values correspond to high compatibility. Indeed, τ_N weights the transverse momentum of the jet costituents by their angular distance to the closest subjet.

In this analysis the N-subjettiness is calculated from the ungroomed jet with the parameter $\beta = 1$. Let's now write explicitly the subjettiness related to the one and two subjet hypothesis,

$$\tau_1 = \frac{1}{d_0} \sum_k p_{T,k} \Delta R_k \tag{3.9}$$

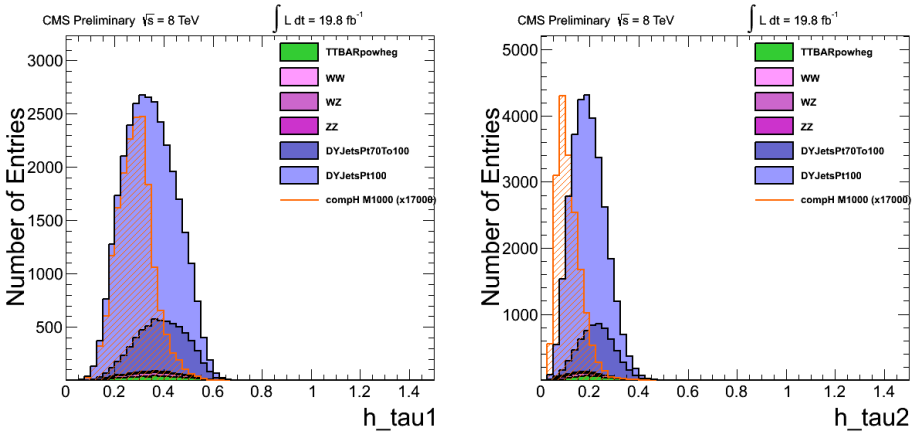

Figure 3.13. τ_1 and τ_2 distributions for SM background and MC signal. The signal has been scaled arbitrarily in order to compare the two shapes.

and

$$\tau_2 = \frac{1}{d_0} \sum_k p_{T,k} \min((\Delta R_{1,k}), (\Delta R_{2,k})). \tag{3.10}$$

Figure 3.13 shows the τ_1 and τ_2 variables for the SM background and the MC signal.

In principle, these two quantities should allow us to distinguish the dipole-like nature of the showering of the Higgs decay from the classic monopole structure of QCD jets. In particular, the variable that best discriminates between H-jets and QCD jets is the ratio of 2-subjettiness and 1-subjettiness,

$$\tau_{21} = \frac{\tau_2}{\tau_1}. \tag{3.11}$$

Figure 3.14 shows the discriminating power of the ratio τ_2/τ_1 for all sources of SM background compared to the the signal, Higgs di-subjet decays. The different shapes and the percent contribution to the total τ_{21} distribution of the three considered channels are presented in figure 3.15. In this case the different behaviour during the hadronization of quarks and gluons causes non-trivial effects, in fact we see that an Higgs jet originated by a pair of gluons is more background-like than an Higgs jet produced by quarks. This is due to the higher number of particles produced during the hadronization, typical of gluons. These additional tracks fill the space between the two partons making less evident the jet substructure.

In figure 3.16 the discriminant τ_{21} variable is presented for different resonance mass. A small shift towards a more background-like shape is observed for high mass signal.

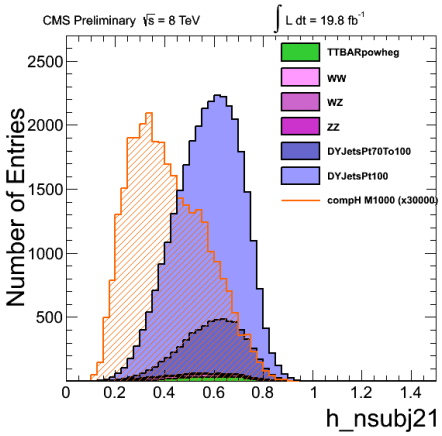

Figure 3.14. τ_{21} distribution for signal and background.

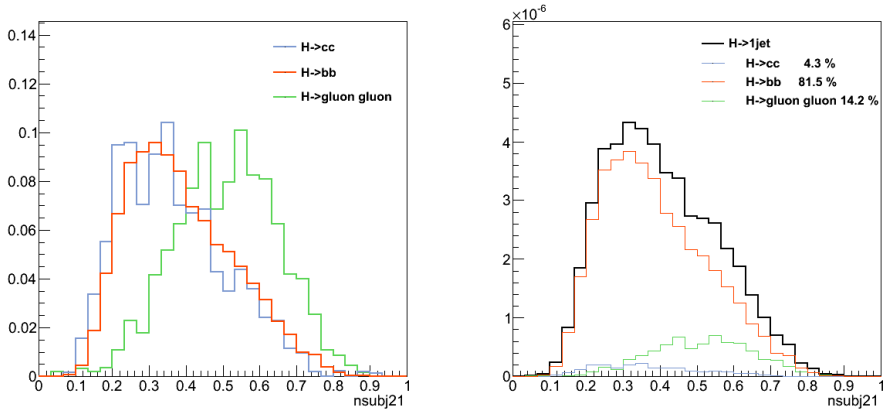

Figure 3.15. Normalized (left) total (right) τ_{21} distribution in the different Higgs decay channels for a signal sample of $M = 1000$ GeV.

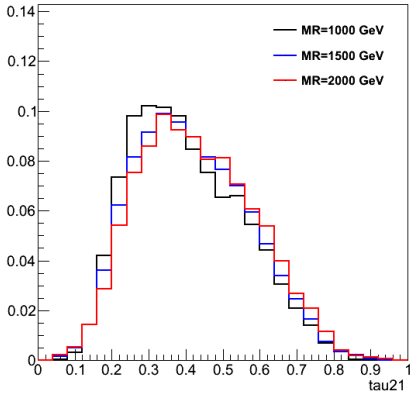

Figure 3.16. Normalized τ_{21} distribution for different mass samples.

Chapter 4

Data analysis

In this chapter the whole analysis is presented. Firstly, the selection criteria are discussed, then all relevant data and MC distributions are shown. Finally, the SM background prediction and the results of the analysis are discussed in details.

4.1 Trigger

Since we search for a final state with two same flavour leptons and at least one jet, we perform this analysis on the DoubleMu and Photon/DoublePhoton-HighPt datasets. The first dataset is triggered by two muons, the second one is triggered by by two electrons. Each of these datasets contains at least one un-prescaled trigger with looser requirements than our offline selections. The lowest un-prescaled trigger is used, and this threshold changes as instantaneous luminosity rises. These triggers are:

- `HLT_Mu22_TkMu8*`

for the DoubleMu dataset, and

- `HLT_DoubleEle33_*`

for the Photon/DoublePhotonHighPt dataset.

The muon trigger has a double threshold, p_T higher than 22 GeV for the leading muon and higher than 8 GeV for the sub-leading muon. Differently, the electron trigger requires an unique (and higher) threshold of 33 GeV. The off-line selection (section 4.3) is such that the trigger efficiency in the acceptance is very close to 100 %.

4.2 Pile up

At the tipical luminosity provided by the LHC, it is common to reconstruct more than one vertex per event. The main event vertex is defined as the

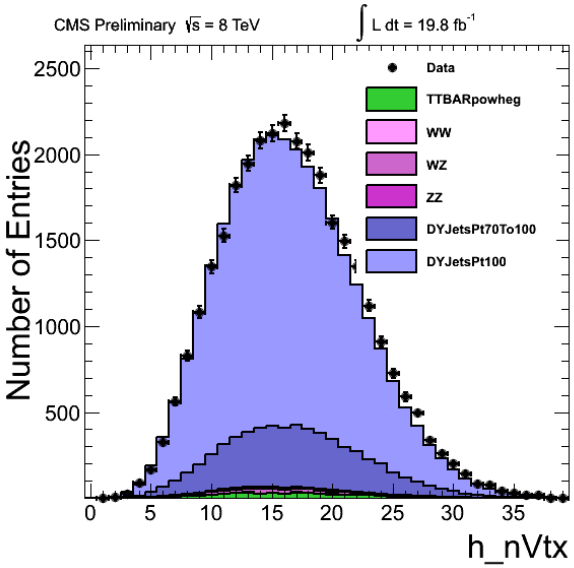

Figure 4.1. Number of reconstructed vertices in data and MC for events at the pre-selection stage (see section 4.6).

one with the highest sum of the p_T^2 of the associated tracks. The presence of additional interactions, known as pile-up (PU), is expected to affect this analysis in two ways:

- tracks and calorimetric towers from PU energy deposits may add energy to the jets from the high p_T collision and bias their energies and angles;
- additional low p_T jets fully composed of PU energy are added to the high p_T collision;

The total number of vertices reconstructed in the data collected for this analysis is shown in figure 4.1.

4.3 Lepton requirements

4.3.1 Muon Selection

In the analysis, we use both tracker muons and global muons [31]. We select muon candidates with $p_T > 20$ GeV and $|\eta| < 2.4$ and at least one of the two muons must have a transverse momentum higher than 40 GeV. The kinematic cuts are the lowest possible compatible with trigger efficiencies flat in p_T . Furthermore, the muons must pass one of these two off-line selections [32].
The high-p_T muon selection:

- muon identified as a `GlobalMuon`
- number of muon hits larger than zero
- number of matched muon stations larger than one
- number of pixel hits larger than zero
- number of tracker layer with hits larger than eight
- transverse impact parameter d_{xy} with respect to the primary vertex smaller than 0.2 cm
- longitudinal impact parameter `dZ` with respect to the primary vertex smaller than 0.5 cm
- relative error on the track transverse momentum $\sigma_{p_T}/p_T < 0.3$

or the tracker-based muons selection:

- muon identified as a `TrackerMuon`
- number of matched muon stations larger than one
- number of pixel hits larger than zero
- number of tracker layer with hits larger than eight
- transverse impact parameter d_{xy} with respect to the primary vertex smaller than 0.2 cm
- longitudinal impact parameter `dZ` with respect to the primary vertex smaller than 0.5 cm
- relative error on the track transverse momentum $\sigma_{p_T}/p_T < 0.3$

An additional variable, useful for the lepton identification, is the isolation. It is defined as the scalar sum of the p_T of the reconstructed objects within a cone in the (η-ϕ) space around the lepton track. The typical size of the cone is ΔR=0.3. Obviously, the transverse momentum of the lepton itself is not included in the sum. A relative isolation definition, obtained dividing the simple isolation by the lepton p_T ($I_{rel} = Iso/p_T^{lep}$), is more frequently used.

An isolation requirement helps in the identification of leptons produced directly in the high p_T collision, which are expected to be isolated, and rejects leptons originated inside jets.

In this analysis, a different isolation criteria is used. The two muons originated from decays of high-p_T Z are close to each other due to the boost of the boson (sec. 3.2) and consequently the presence of each muon could spoil

the isolation of the other muon in the pair. In order to solve this problem we use a track-based isolation relative quantity, explicitely removing from the momentum flow any other muon passing our muon selection. Moreover, a tracker-based isolation is well motivated also by two additional aspects: it is more independent of pile up (pile-up tracks tipically do not match the primary vertex) and does not include possible muon radiation.

Finally, the modified requirement is

$$I^{mod}_{trkrel} < 0.1. \tag{4.1}$$

4.3.2 Electron Selection

Electrons are selected with $p_T > 40$ GeV and $|\eta| < 2.4$. Other identification requirements are:

- transversal supercluster energy larger than 35 GeV
- supercluster pseudorapidity (η_{SC}) smaller than 1.442 (for barrel electrons) or in the range 1.56-2.5 (for endcap electrons)
- have either $E_{2\times5}/E_{5\times5}$ larger than 0.94 or $E_{1\times5}/E_{5\times5}$ larger than 0.83
- ratio of hadronic energy (HCAL deposit) to electromagnetic energy (ECAL deposit) smaller than 0.05
- number of inner layer lost hits smaller than 2
- have d_{xy} smaller than 0.02 (0.05) cm for barrel electrons (endcap electrons)

where $E_{i\times j}$ is the energy contained in a $i \times j$ block around the seed crystal (defined as the highest deposit of energy of the cluster). The idea is sketched in figure 4.2 and it is used to study the shape of the shower in the ECAL [33]. The asimmetry in the (η-ϕ) space is due to the presence of bremsstrahlung radiation. Electrons can emit photons that continue straight and impact the ECAL at the same η but at shifted ϕ due to the curvature of the electron track in the transverse plane.

Again, as for muons, we need a redefinition of the isolation requirement because of the small ΔR between electrons. In this case there are three variables that have to be changed:

- the track isolation variable I_{trk} is defined as the scalar p_T sum of the tracks within a $\Delta R = 0.3$ cone around the electron, excluding both an inner core of dimensions 0.03×0.3 in (η, ϕ) around the electron in question and additional cones of dimensions 0.03×0.3 in (η, ϕ) around any other electron passing the criteria given above.

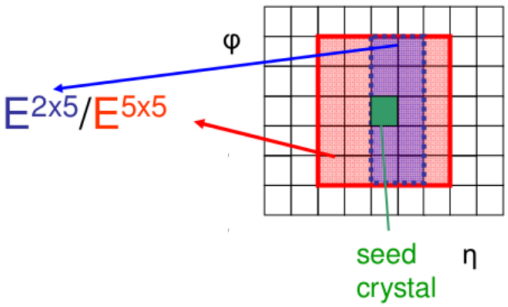

Figure 4.2. Example of the $E_{2\times5}/E_{5\times5}$ computation in the ECAL crystals [33].

- the electromagnetic calorimeter isolation I_{ECAL} is defined as the scalar sum of E_T of the crystals in a $\Delta R = 0.3$ cone around the electron, excluding both an inner area of full-width 3 crystals around the electron in question and a 4-crystals ΔR cone around any other electron. The dimension of the ECAL crystals corresponds roughly at $\Delta R \sim$ 0.01-0.02 (see section 2.1.2).

- the hadronic calorimeter isolation variable I_{HCAL1} is defined as the scalar sum of E_T of the HCAL caloTowers with centres in a $\Delta R = 0.3$ cone around the electron, excluding those lying within $\Delta R = 0.15$ of the electron itself and of any other electron [34].

The final modified tracker isolation requirement is

$$I_{trk}^{mod} < 5 \text{ GeV}, \tag{4.2}$$

while, for the ECAL and HCAL isolations, a threshold varying with the electron transverse energy is chosen (see table 4.1).

4.4 Jet requirements

Jets are clustered from the list of Particle Flow (PF) candidates that are reconstructed in the event [17]. Charged hadrons originating from vertices other than the primary vertex are not used in the jet clustering procedure. In this analysis the CA8 (Cambridge-Aachen) algorithm with a cone radius of $R = 0.8$ is used for the identification of jets and jet candidates are selected with $p_T > 30$ GeV and $|\eta| < 2.4$.

Furthermore jets are required to pass the following loose identification criteria:

- muon energy fraction smaller than 0.99

- photon energy fraction smaller than 0.99
- charged electromagnetic energy fraction smaller than 0.99
- neutral hadron energy fraction smaller than 0.99
- charged hadron energy fraction larger than 0
- number of constituent particles larger than 1.

4.5 Reconstruction of the bosons

Z candidates are formed from oppositely charged same flavour lepton pairs, with invariant mass in the range [70,110] GeV. Since this search focuses on high mass resonances, we can apply a preselection threshold for the Z $p_T > 80$ GeV without any loss of efficiency.

Accordingly to the single jet topology (section 3.2), we select H candidates choosing jets with pruned mass larger than 40 GeV and $p_T > 80$ GeV.

No selections on τ_{21} are applied at the pre-selection level.

4.6 Pre-selection level: MC signal and background

In this section all the control plots at the pre-selection level are presented. Table 4.1 reports a summary of the pre-selection requirement described in section 4.3, 4.4 and 4.5. Figures 4.3 and 4.4 present the identification variables for the muon and electron selections described in sections 4.3.1 and 4.3.2 respectively. The data and MC comparison generally presents a fair agreement, any discrepancies are then corrected with the tag and probe technique [35].

Figure 4.5 shows kinematic distributions separatly for the electron and muon channels, the electron η distribution presents a gap due to the veto in the barrel/endcaps transition region of the selection (section 4.3.2).

Figures 4.6 and 4.7 show the modified isolation variables. All but the calorimeter based isolation have a large peak at zero and their distributions are cut at the value given by the selection requirements of sections 4.3.2 and 4.3.1.

The ΔR (defined in eq. 3.5) between leptons and between lepton and jet for the two combined channels are reported in figure 4.8. The different signal shape of the lepton ΔR compared to the MC background reflect the different behaviour of the Z p_T signal and background distributions and the high p_T expected for the resonance products (fig. 4.11 (left)).

At the pre-selection level all the possible $X \to ZH \to ll1jet$ candidates are considered. The number of jets and the number $X \to ZH$ candidates in the event are reported in figure 4.9.

Selection	Value	Comments
Trigger		
	HLT_Mu22_TkMu8	DoubleMu dataset
	HLT_DoubleEle33_	DoublePhoton dataset
Lepton selections		
Leading lepton p_T	$p_T > 40$ GeV	Same for electrons and muons
Subleading lepton p_T	$p_T > 40$ GeV	For electrons
Subleading lepton p_T	$p_T > 20$ GeV	For muons
Muon η	$\lvert\eta\rvert < 2.4$	
Electron η	$\lvert\eta\rvert < 2.5$	
Electron fiducial	$\lvert\eta\rvert$ out of [1.4442, 1.566]	Avoid the ECAL gap
Muon ID	High p_T	
	Tracker based	
Muon Isol. I^{mod}_{trkrel}	< 0.1	
Electron ID	HEEP modified	
Ele. Isol.		
I^{mod}_{trk}	< 5 GeV	
$I^{mod}_{ECAL} + I^{mod}_{HCAL1}$	< 2 GeV $+0.03\,E_T$	EB electrons
	< 2.5 GeV	EE ele. with $E_T < 50$ GeV
	< 2.5 GeV $+0.03\,E_T$	EE ele. with $E_T > 50$ GeV
jet selections		
Jet ID	Loose working point	
Jet p_T	$p_T > 30$ GeV	
Jet η	$\lvert\eta\rvert < 2.4$	
Boson selections		
m_{LL}	$70 < m_{LL} < 110$ GeV	
m_J	$m_J > 40$ GeV	
Z p_T	$p_T > 80$ GeV	
H p_T	$p_T > 80$ GeV	

Table 4.1. Pre-selection requirements used in the analysis.

Kinematic distributions of the Higgs and Z bosons are plotted in figure 4.10 separately for the $Z \to ee$ and $Z \to \mu\mu$ channels and in figure 4.11 for the combined Z and for the Higgs boson. The 80 GeV threshold of the p_T distributions is given by the pre-selection cut. Figure 4.12 shows the resolution in the reconstruction of the Z boson in the two leptonic channels.

Since the requirement on the sub-leading lepton p_T is different for the electron and muon channels, as a consequence of the different trigger threshold, the SM background is more populated (roughly twice) in the muon channel than in the electron channel. On the contrary, the predicted signal yield is similar in the two channels because it has little sensitivity to the low threshold applied in the preselection. The sensitivity is low because the signal has high p_T bosons in the final state.

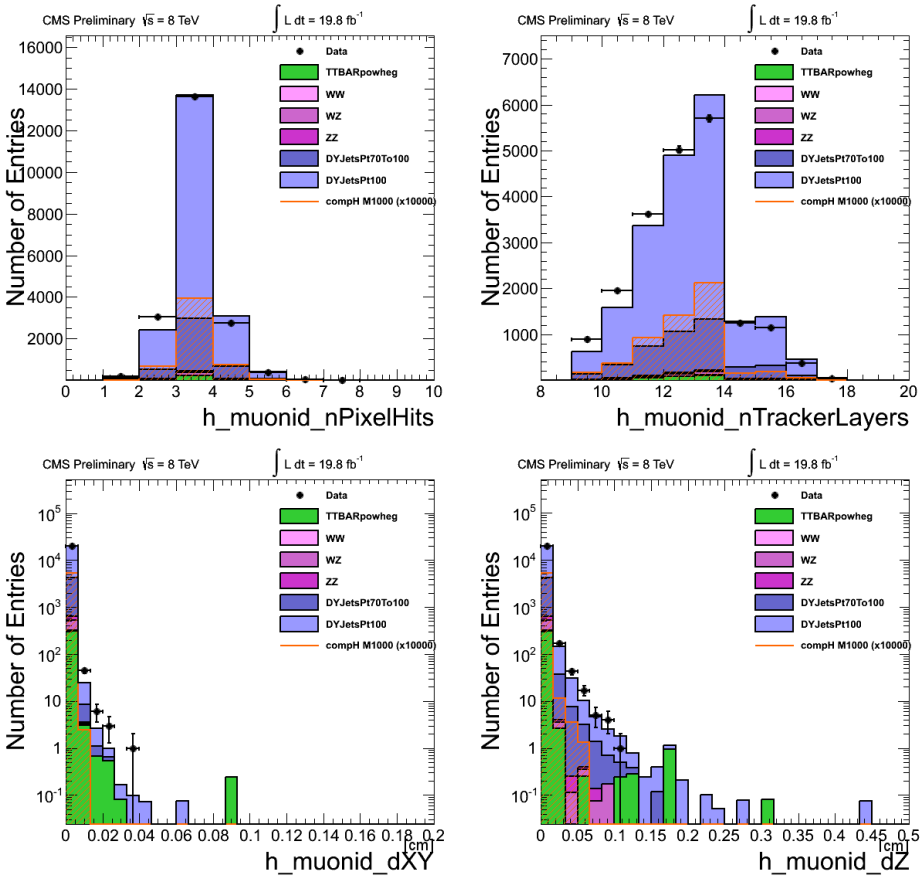

Figure 4.3. Muon identification and selection variables. Top left: number of pixel hits; top right: number of tracker layers with hits; bottom left: transverse impact parameter with respect to the primary vertex; bottom right: longitudinal impact parameter with respect to the primary vertex.

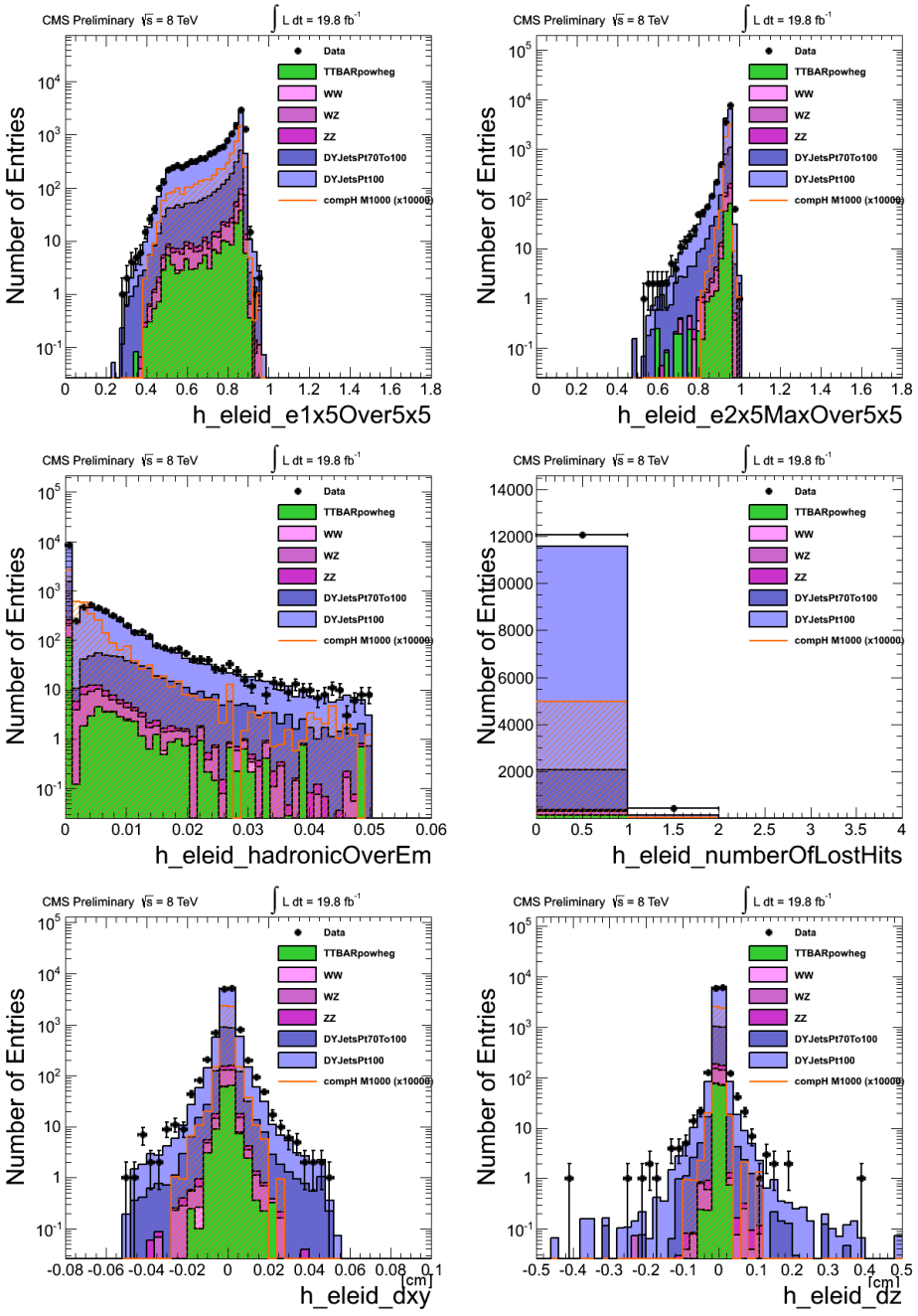

Figure 4.4. Electron identification and selection variables. Top left: $E_{1\times5}/E_{5\times5}$ ratio of the ECAL energy deposits; top right: $E_{2\times5}/E_{5\times5}$ ratio of the ECAL energy deposits; middle left: ratio of the hadronic to electromagnetic energy deposits; middle right: number of inner layer lost hits; bottom left: transverse impact parameter with respect to the primary vertex; bottom right: longitudinal impact parameter with respect to the primary vertex.

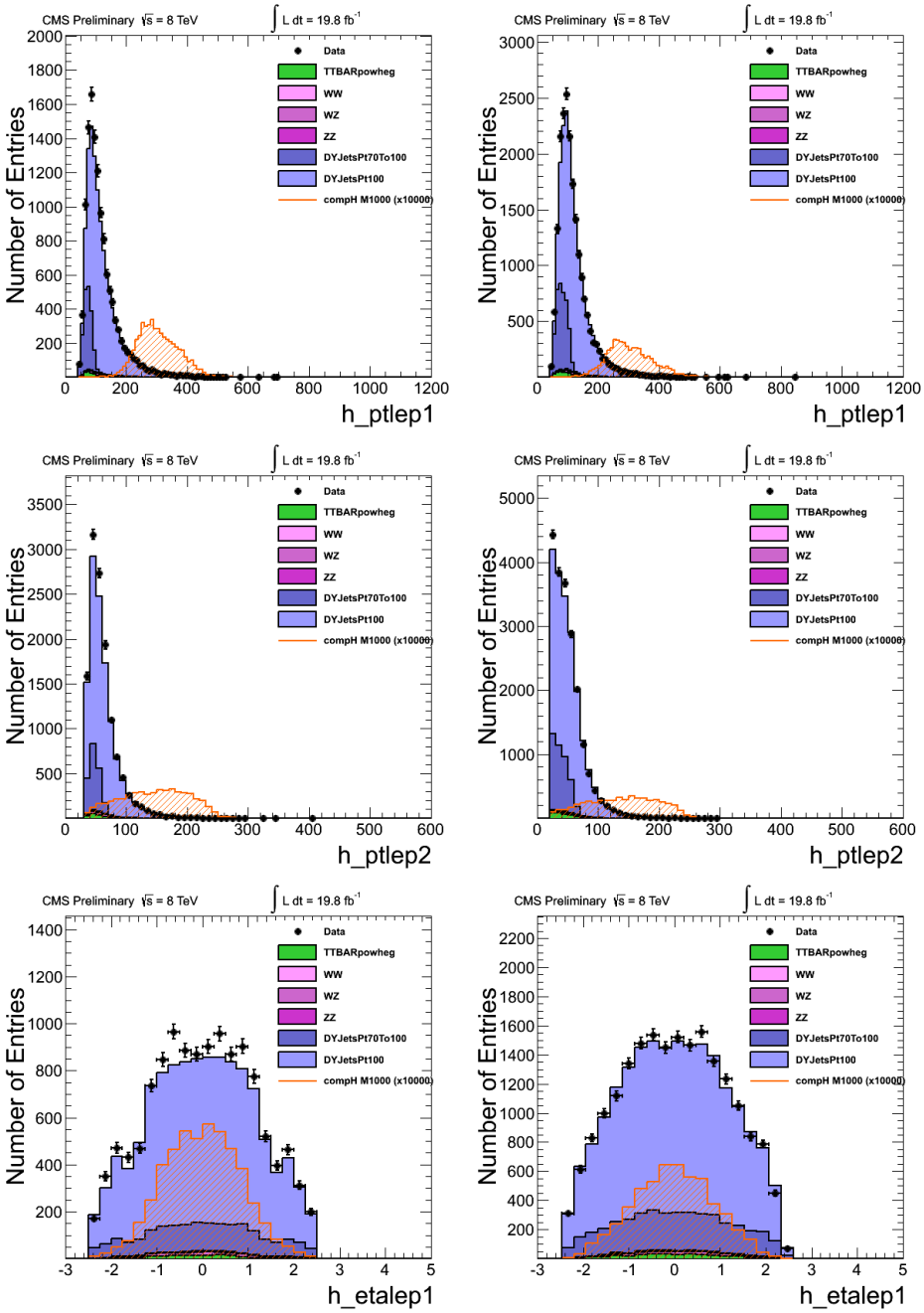

Figure 4.5. Kinematical distribution for electrons (left) and muons (right). Top: p_T of the leading lepton; middle: p_T of the subleading lepton; bottom: η of the leading lepton.

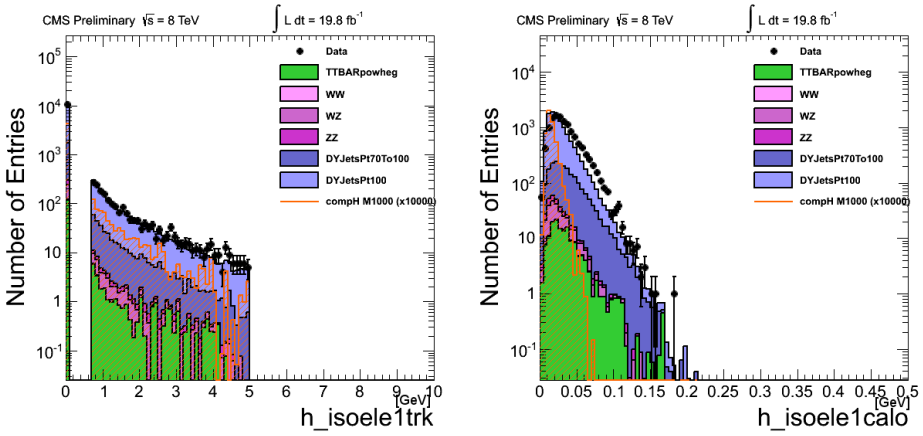

Figure 4.6. Modified isolation variables for the leading electron after the pre-selection requirements; tracker based (left) and calorimeter based isolation (right).

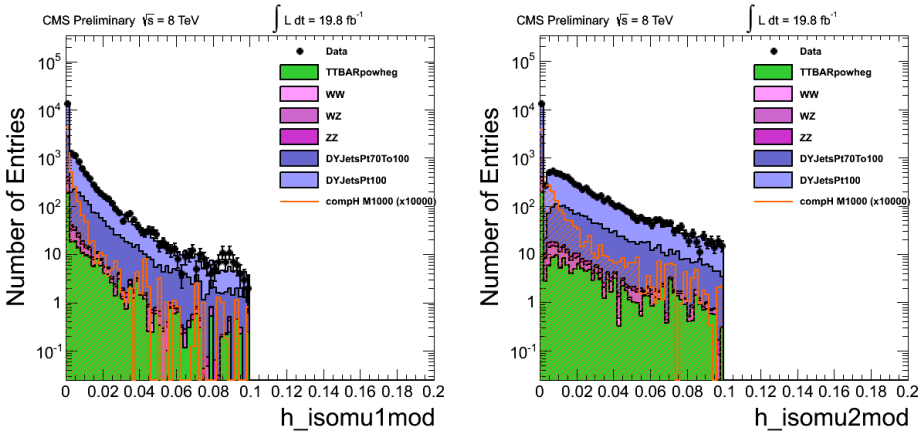

Figure 4.7. Modified isolation for the leading (left) and subleading (right) muon after the pre-selection requirements.

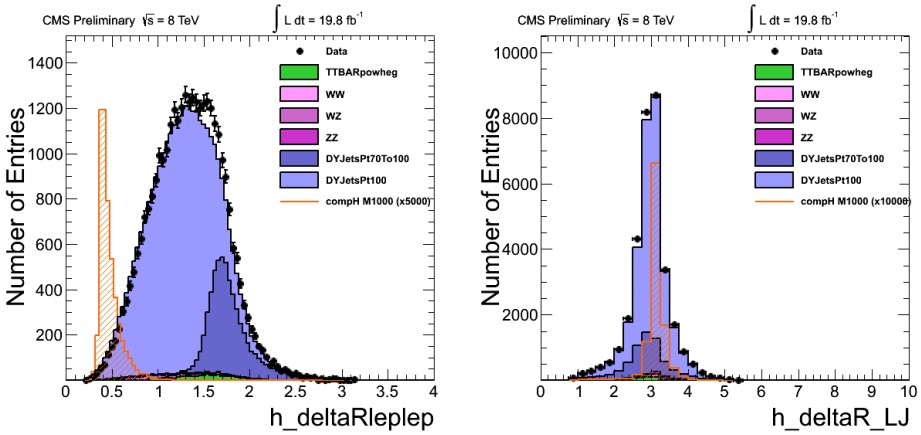

Figure 4.8. Delta R distribution between the two leptons (left) and between the leading lepton and the jet (right) for the combined channel ($lep = e, \mu$).

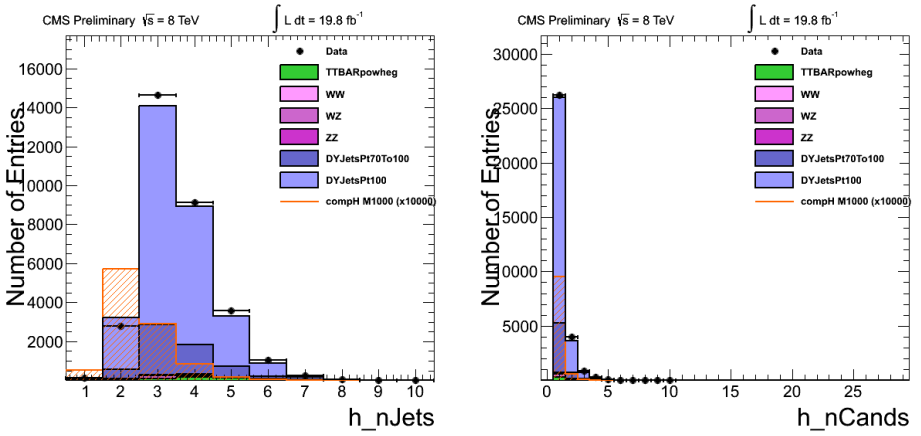

Figure 4.9. Number of jets and number of $X \to ZH$ candidates in the event for the combined channel ($lep = e, \mu$).

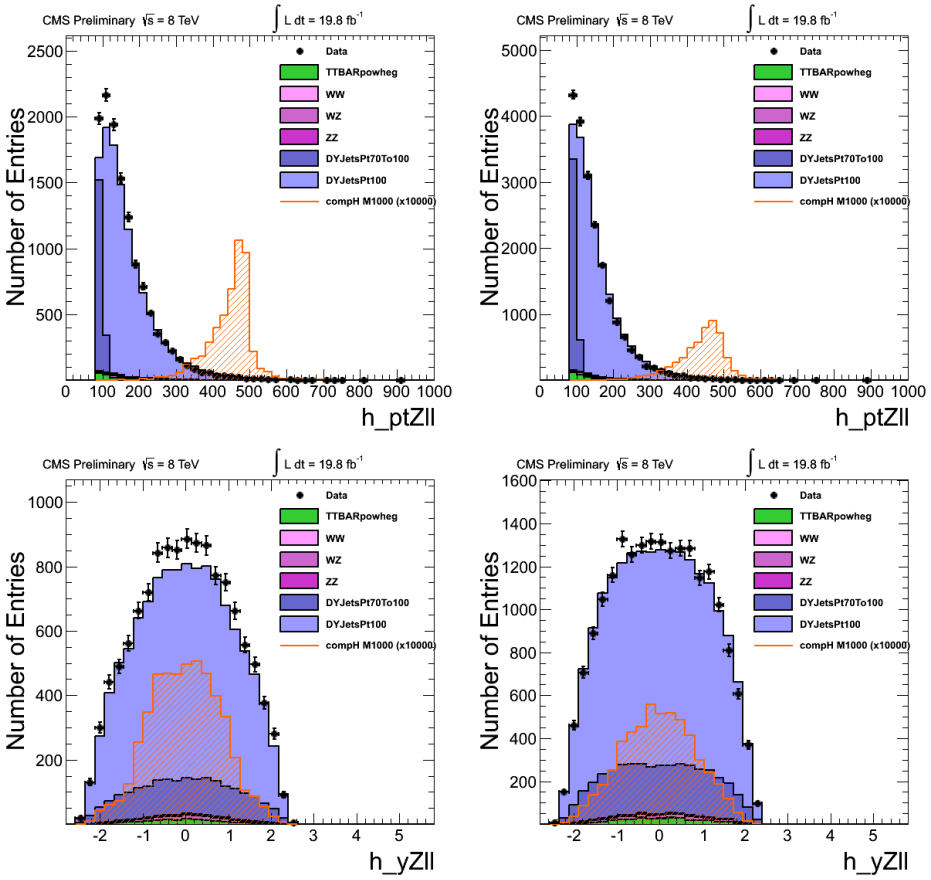

Figure 4.10. Kinematic distributions at the pre-selection level for the Z boson. p_T (top) and η (bottom) separately for electron (left) and muon channel (right).

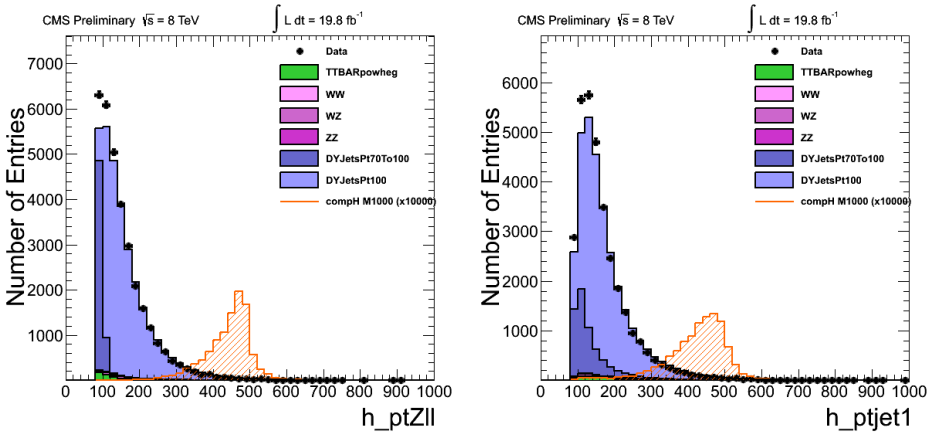

Figure 4.11. Combined p_T distribution for the Z (left) and Higgs (right) bosons at the pre-selection level.

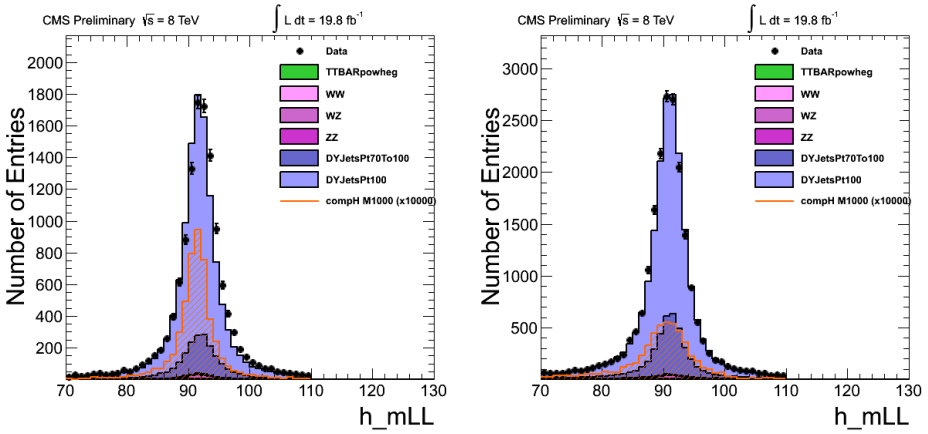

Figure 4.12. Recostructed Z mass in the electron (left) and muon (right) channel.

4.7 Final selection

At this step of the analysis we can still have more than one candidate per event, due to the presence of multiple leptons or jets (figure 4.9). We select our best candidate by choosing the one with the lowest value of the variable

$$D = (m_{ll} - M_Z)^2 + (M_J - M_H)^2. \tag{4.3}$$

4.7.1 Signal region

As already discussed in section 3.3.2, the most discriminating tool to separate signal from the dominant background is the requirement on the pruned mass of the jet (fig. 3.8 (right)). In this analysis the pruned mass of the jet is required to be in the range [110, 140] GeV in order to pass the final selection. The range is chosen in order to contain as much signal as possible without overlapping the signal region of this analysis with other searches of new resonances, in particular with the semileptonic $BulkG \to ZZ \to llq\bar{q}$ search [32]. In the latter, the signal events are the same with the exception of the presence of a second Z boson instead of the Higgs, and the expected backgrounds are identical. Since in the $BulkG \to ZZ$ search the signal region is set around the Z mass in the range [70, 110] GeV, we do not extend further the lower edge of our region.

Figure 4.13 shows the signal region superimposed on the pruned mass distribution. The gaussian fit on the peak of the distribution has as output parameters a mean value around 120 GeV and σ around 12 GeV. The difference of the peak mass respect to the real value of the Higgs mass is due to the pruning algorithm applied to the jet, that reduces its reconstructed mass.

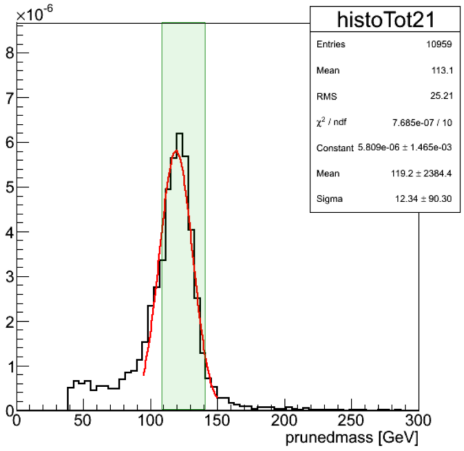

Figure 4.13. Jet pruned mass distribution for a MC signal of 1000 GeV whose peak is fitted with a gaussian core. The signal region is painted in green.

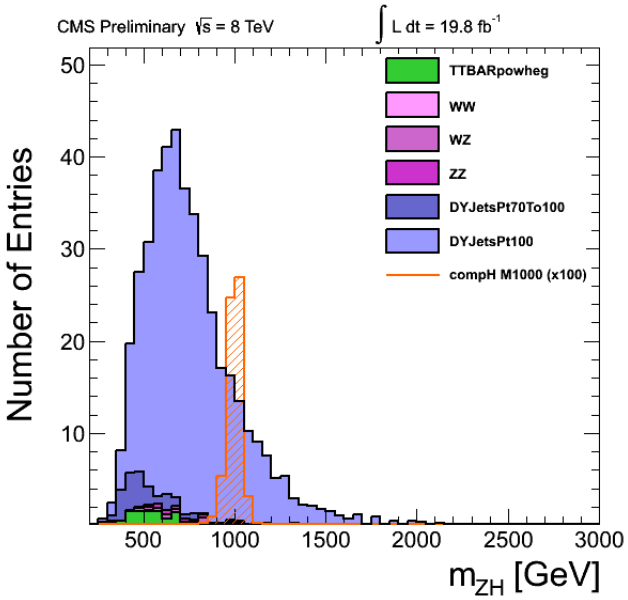

Figure 4.14. Invariant mass distribution after the cut on the pruned jet mass (H-tagging). The MC background simulation is compared with a signal hypotesis of 1000 GeV for the combined channel ($lep = e, \mu$).

Finally, figure 4.14 shows the invariant mass of the remaining MC background after the pruned jet cut compared with a 1000 GeV signal hypotesis.

4.7.2 τ_{21} cut optimization

After the selection on the pruned mass, the discriminating power of the ratio τ_{21}, described in section 3.3.3, is reduced since the mass cut and τ_{21} cut are correlated. Figure 4.15 shows the τ_{21} distributions for MC signal and background before and after the requirement on the pruned mass of the jet.

In this section we want to study the performances and the optimization procedure of the selection on this variable.

Punzi figure of merit

In searches for new phenomena it is important to define the sensitivity of the experiment, in order to maximize the potentiality of the analysis.

In a simple counting experiment, the whole problem is reduced to one question: whether the number of observed events n is compatible with the background B, predicted by the standard theory H_0 (and fluctuating following the Poisson statistics), or it is more compatible with a larger mean value $B + S_m$ of the Poisson distribution that includes the new signal S_m, tipically function of the unknown free parameters (m) of the new theory H_m.

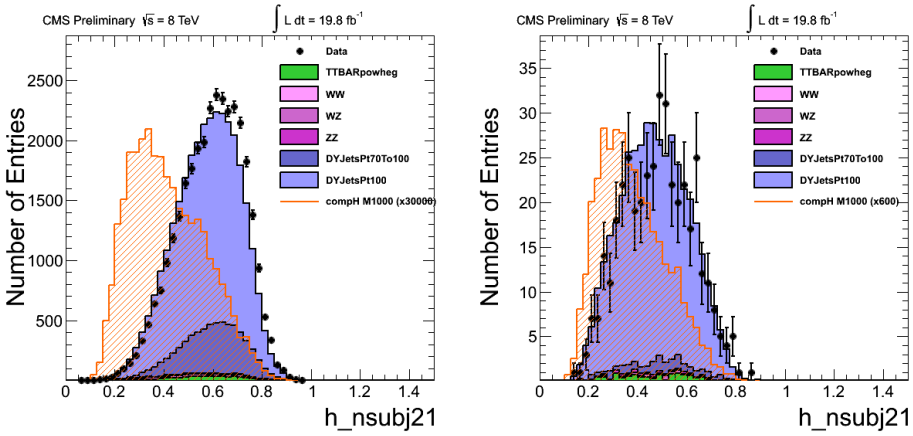

Figure 4.15. τ_{21} distribution for MC signal and background before (left) and after (right) the selection in the signal region [110-140] GeV.

The Poisson statistics leads to a probability of observing a discrete number of events n given by

$$p(n) = e^{-B}\frac{B^n}{n!} \tag{4.4}$$

in the case of an only background scenario, and

$$p(n) = e^{-(B+S_m)}\frac{(B+S_m)^n}{n!} \tag{4.5}$$

in presence of signal.

In this case the critical region, defined as the set of values of the observed variable (number of events) that rejects the H_0 hypothesis with a given probability $(1-\alpha)$, takes the form

$$S_m > S_{min}, \tag{4.6}$$

and depends on the significance level of the test (α). Figure 4.16 shows the value of $n_{min} = B + S_{min}$ as a function of B, for given values of α [36].

The parameter space that satisfies equation 4.6 is called the *sensitivity region* of the experiment and corresponds to the set of values that can be confirmed by a new discovery or excluded with a certain confidence level (CL) if no evidence is found for a deviation from the standard theory.

Therefore, the final aim is to find the largest sensitivity region changing the different experimental settings. This optimization procedure (done on simulated data) can be performed with the help of an analytic parametrization.

Two common sensitivity variables are tipically used:

1. $\frac{S_m}{\sqrt{B}}$

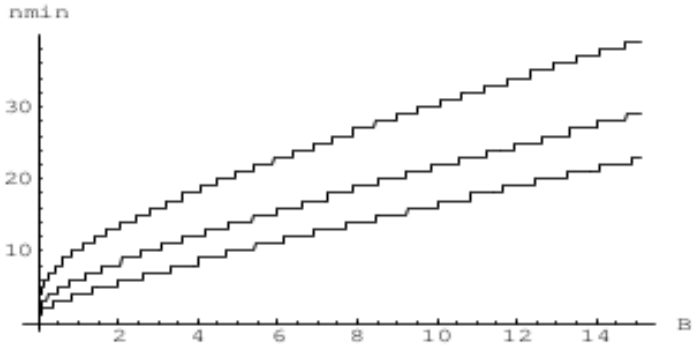

Figure 4.16. Minimum number of observed events needed to claim discovery with 95%, 3σ, 5σ significance, vs expected background [36].

2. $\frac{S_m}{\sqrt{B+S_m}}$.

but both present problems. The former diverges for small values of background overstimating the sensitivity at low background, the latter, being not linear in S_m, explicitely requires the knowledge of the cross section of the signal.

As a consequence we adopt as figure of merit the quantity introduced in [36] called Punzi significance

$$P = \frac{\varepsilon_S}{1 + \sqrt{B}} \tag{4.7}$$

where ε_S is the signal selection efficiency and B is the remaining background.

Optimization procedure

For each mass point we want to establish which is the best value of the τ_{21} ratio to discriminate signal from background. The procedure is implemented as follows:

- set a window of $\pm 15\%$ around the signal resonance mass;
- plot the expected τ_{21} variable for signal and background, for the events that passed all the other selection requirements;
- integrate the expected τ_{21} distributions of signal and background up to a threshold τ_{21}^{max}. The values obtained are an estimation of the signal selection efficiency and the amount of background;
- calculate the figures of merit.

This procedure is repeated for values of τ_{21}^{max} ranging form 0.05 to 0.95 in steps of 0.05. In figures 4.17, 4.18 and 4.19, the results of the optimization procedure for signal of 1000, 1500 and 2000 GeV are reported. The

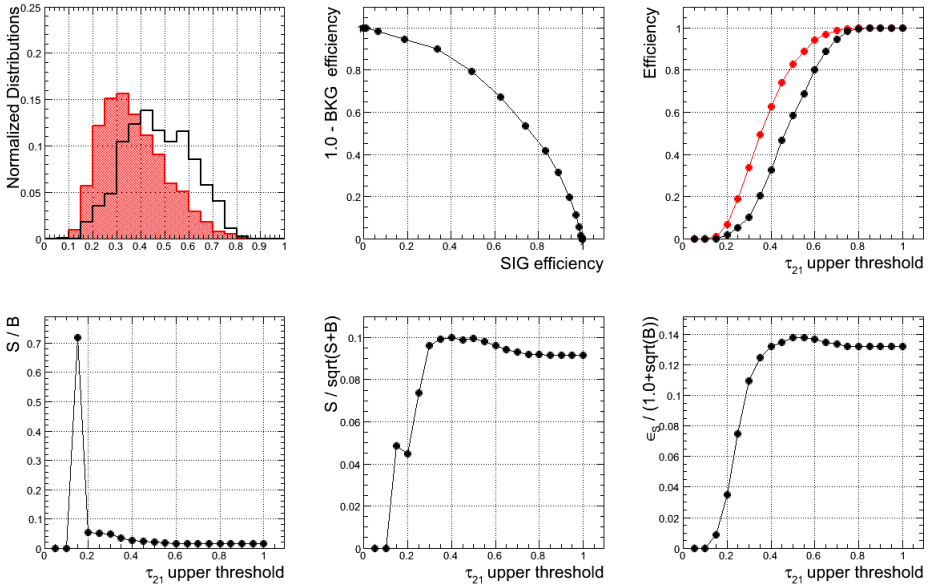

Figure 4.17. Optimization procedure for a signal sample of 1000 GeV. Top left: distributions of τ_{21} for signal (red) and background (black). Top middle: signal efficiency $\times$ background rejection (ROC curve). Top right: signal (red) and background (black) efficiencys as function of τ_{21} cut. Bottom left: signal S over background B as function of τ_{21} cut. Bottom middle: $S/\sqrt{S+B}$ as function of τ_{21} cut (the signal cross section has been arbitrary fixed at 0.1 fb). Bottom right: Punzi significance (as defined in eq. 4.7) as function of τ_{21} cut.

large peak of the bottom left plot of figure 4.17 shows the instability of a simple $S/\sqrt{B}$ discriminating ratio when the background tends to zero. The top left plot of figure 4.19, instead, highlights a lack of statistics for the background MC samples at high invariant mass. It can be deduced from the huge fluctuations of the background distribution.

For each mass the optimized τ_{21} threshold is given by the maximum of the Punzi figure of merit (bottom right plot). The trend of the best cut as a function of the signal mass is shown in figure 4.20.

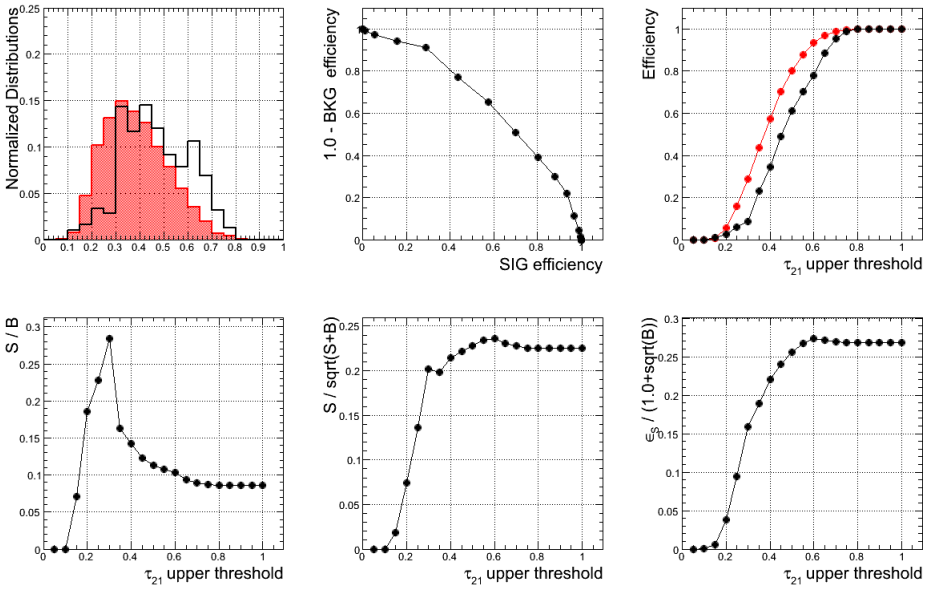

Figure 4.18. Optimization procedure for a signal sample of 1500 GeV. Top left: distributions of τ_{21} for signal (red) and background (black). Top middle: signal efficiency $\times$ background rejection (ROC curve). Top right: signal (red) and background (black) efficiencys as function of τ_{21} cut. Bottom left: signal S over background B as function of τ_{21} cut. Bottom middle: $S/\sqrt{S+B}$ as function of τ_{21} cut (the signal cross section has been arbitrary fixed at 0.1 fb). Bottom right: Punzi significance (as defined in eq. 4.7) as function of τ_{21} cut.

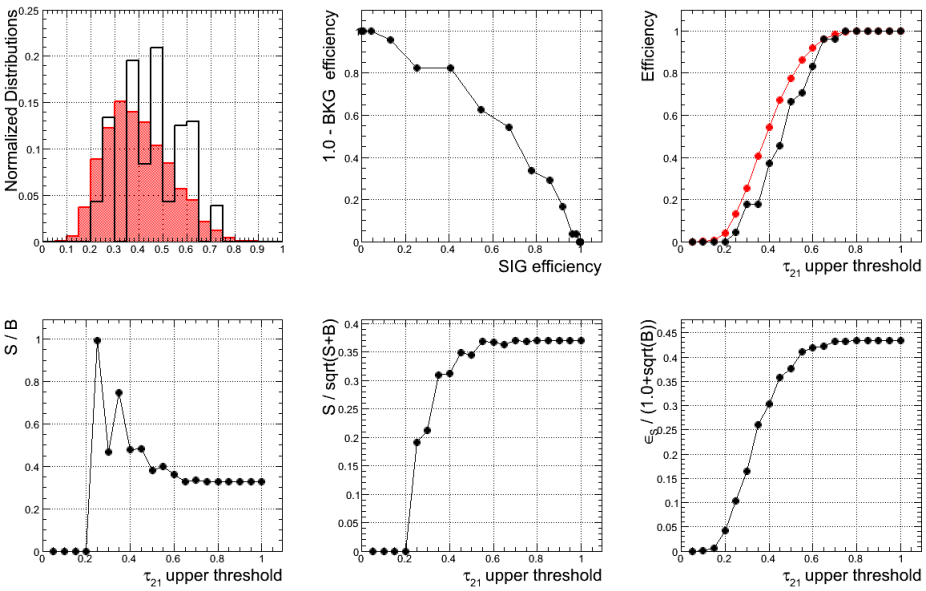

Figure 4.19. Optimization procedure for a signal sample of 2000 GeV. Top left: distributions of τ_{21} for signal (red) and background (black). Top middle: signal efficiency $\times$ background rejection (ROC curve). Top right: signal (red) and background (black) efficiencys as function of τ_{21} cut. Bottom left: signal S over background B as function of τ_{21} cut. Bottom middle: $S/\sqrt{S+B}$ as function of τ_{21} cut (the signal cross section has been arbitrary fixed at 0.1 fb). Bottom right: Punzi significance (as defined in eq. 4.7) as function of τ_{21} cut.

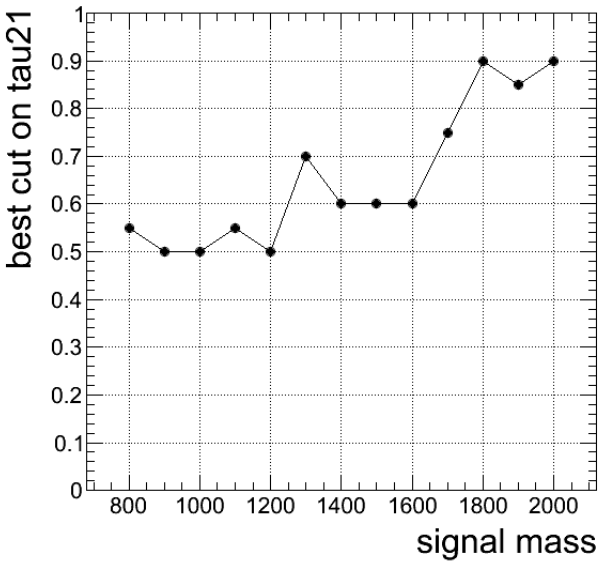

Figure 4.20. Best τ_{21} cut as a function of the mass sample as described in the optimization procedure.

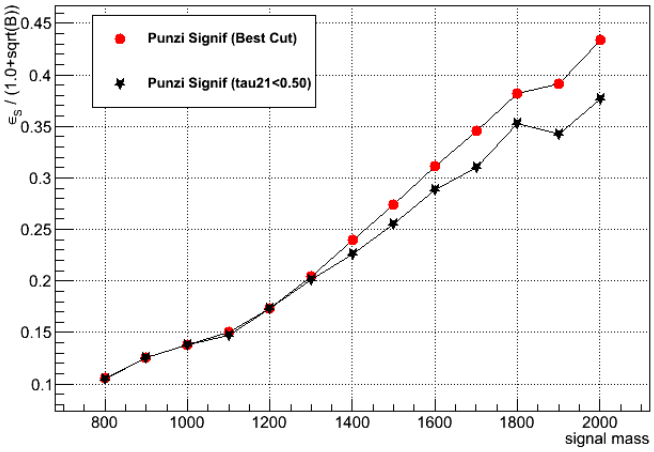

Figure 4.21. Values assumed by the Punzi figure of merit for the best τ_{21} cut (red), given by the optimization procedure, and the fixed cut at $\tau_{21} = 0.5$ applied in the analysis (black) as a function of the mass of the signal.

4.7.3 τ_{21} categorization

One can see from figure 4.20 that there is no optimal choice for the τ_{21} requirement in the mass range considered. It is evident that for high masses the optimal cut on τ_{21} should be relaxed so much as to be no longer needed above 1800 GeV, where the small amount of background makes any restriction on the signal itself useless. This fact can be observed from the trend of the Punzi significance in figure 4.19, that never reaches a well defined maximum but is flat up to the high edge of the τ_{21} range investigated.

The situation is different at low masses (below 1300 GeV), where the maximum of the sensitivity discriminant is around 0.5. A third intermediate case is rapresented by the signal samples with mass in the middle of the range explored, where the value of the discriminant variable, after having reached a maximum, stays on a high plateau only slightly smaller than the value at the maximum: this means that the selection on the τ_{21} variable is not really relevant.

Figure 4.21 plots the different values of the Punzi discriminant for the best τ_{21} cut and for a $\tau_{21} = 0.5$ cut. Figure 4.22 shows the loss of efficiency due to the introduction of a threshold on the τ_{21} ratio at high masses.

On the basis of these observations we have decided to adopt a cut of τ_{21} of 0.5 and to use in a different category also the events above the cut. We define two regions for the analysis:

- High-purity category (HP): events in the signal region with $\tau_{21} < 0.5$.
- Low-purity category (LP): events in the signal region with $0.5 < \tau_{21} < 0.75$.

The advantage of this strategy is the recovery of the efficiency at high masses and the search will be done combining the two categories.

Figure 4.23 shows the residual MC background after the final selection divided respectively in the two categories. One can note how the high purity category has a further improvement in the signal/background discrimination respect to the selection on the jet mass alone (fig. 4.14).

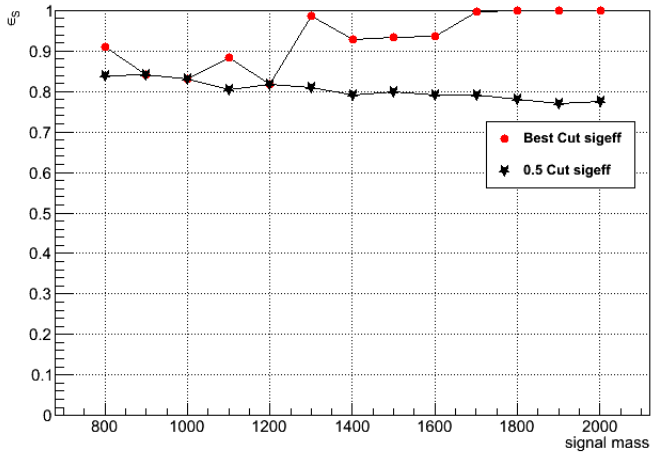

Figure 4.22. Efficiency of the selection on the τ_{21} ratio calculated for the best τ_{21} cut (red), and the fixed cut at $\tau_{21} = 0.5$ applied in the analysis (black) as a function of the mass of the signal.

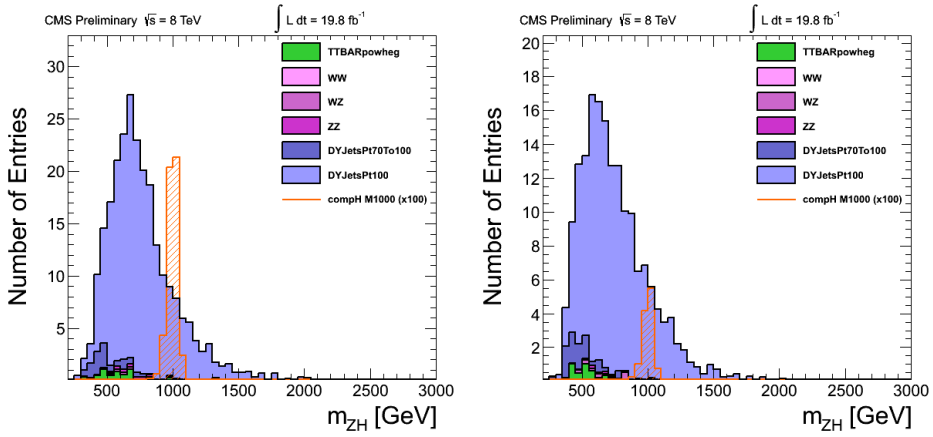

Figure 4.23. Invariant mass distribution after the cut on the pruned jet mass (H-tagging) of the MC background simulation separately for the high purity (left) and low purity (right) category. A rescaled signal hypotesis of 1000 GeV is added as a comparison.

4.7.4 Data/Montecarlo comparison in the signal region

Now that the selection has been optimized we can look at the data in the signal region. A comparison between data and MC of the invariant ZH mass distribution in the signal region (defined in section 4.7.1) is shown in figure 4.24.

Inspecting the two distributions, one can see that data do not present any significant deviation from the MC expectation, so we decide to put limits on the production cross section times the branching ratio for the $pp \to X \to ZH$ process.

In order to put these limits we need two more ingredients: the efficiency of our selection and the evaluation of the background in the signal region. These will be computed also using data in the sideband region, defined as the events where the H invariant mass lies in the [50, 110] GeV interval. An accurate choise of the sideband region will be presented in section 4.8.1.

4.7.5 Signal efficiency

The signal reconstruction and selection efficiencies are computed separately for the four categories (HP and LP, both divided in the electron and muon channels) defining the ratio

$$\varepsilon_{SIG} = \frac{Nr.\,of\,events\,passing\,the\,full\,selection}{Total\,nr.\,of\,events\,generated}, \tag{4.8}$$

where the total number of events generated includes only leptonic (ee, $\mu\mu$) Z decays and hadronic ($q\bar{q}$, gg) Higgs decays. The total efficiency of the $pp \to X \to ZH \to llq\bar{q}(gg)$ process in this analisys is obtained by summing up the efficiencies of equation 4.8 computed in each of the four categories. In order to compute the efficiency for the $pp \to X \to ZH$ process we scale by the SM branching ratios for the Z and H decays: $BR(Z \to l^+l^-(e,\,\mu)) = 6.7\%$ and $BR(H \to q\bar{q}(gg)) = 69.2\%$. The restriction to the $H \to q\bar{q}(gg)$ channel can lead to a systematic error in the evaluation of the efficiency since other fully hadronic Higgs decays as $H \to WW \to 4q$ and $H \to ZZ \to 4q$ can pass our selection. However, this effect is expected to be small because of the strong background-like behaviour of the other fully hadronic Higgs decays, these will be discussed in detail in section 5.2.1.

The efficiency defined in equation 4.8 is computed using the MC simulation, and the resulting values are compared with the data based tag and probe technique [35]. Then, in the case that the two efficiencies are different, a scale factor and systematic uncertainties are introduced.

An other important point is to verify that the efficiency of the pruning and of the τ_{21} cut are correctly reproduced by the MC simulation. Figure 4.25 shows the data/MC comparison in the sideband region for the τ_{21} variable and one can see that the agreement is not perfect: a pure MC efficiency computation is not totally trustworthy in this case.

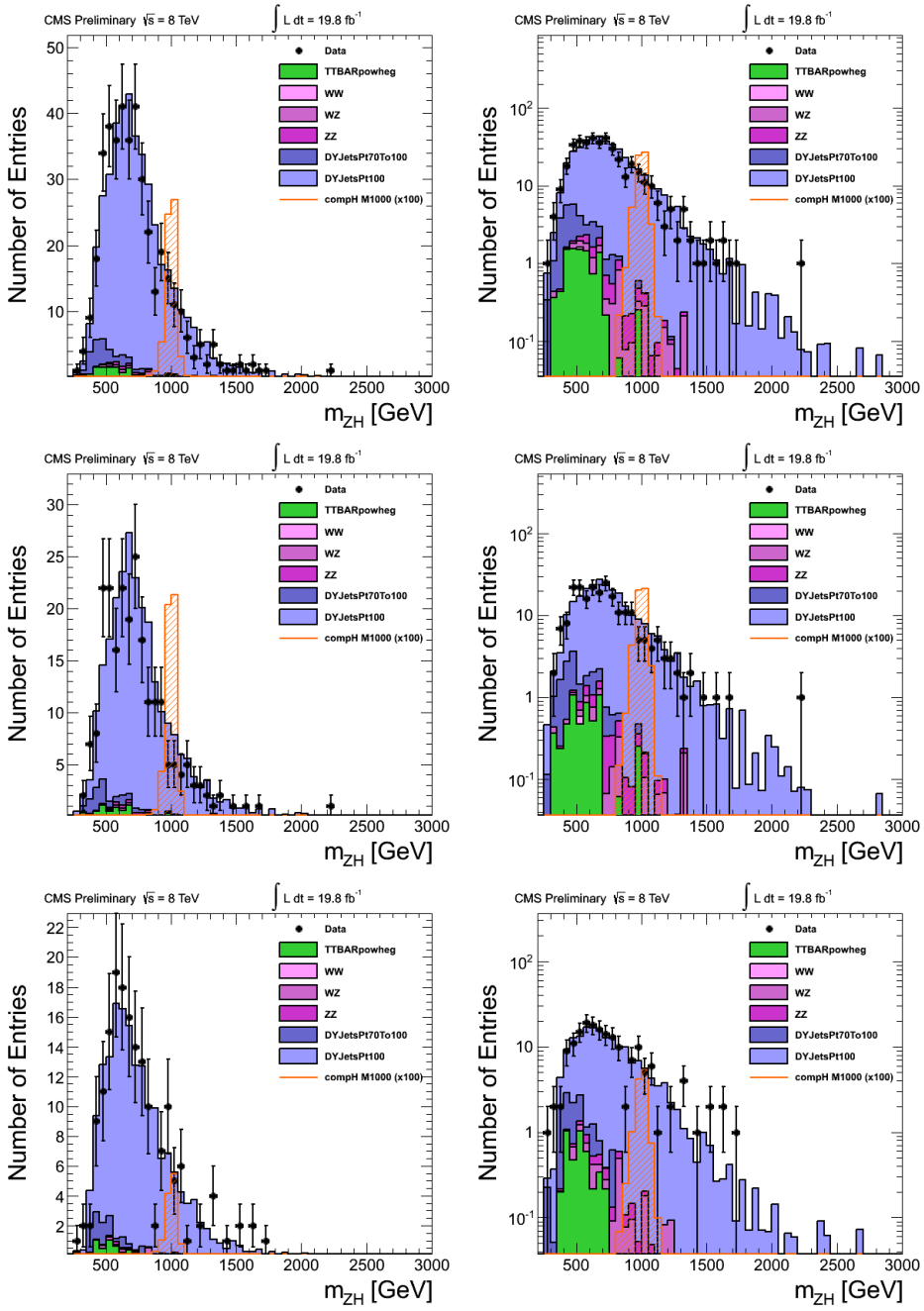

Figure 4.24. Invariant mass distribution of the MC background simulation compared to the observed data in the signal region. The histograms are presented with no category distinction (top), in the HP category (middle) and in the LP category (bottom) plotted in linear scale (left) and logarithmic scale (right). A rescaled signal hypotesis of 1000 GeV is added as a compariso; the MC distributions are the same as in figures 4.14 and 4.23.

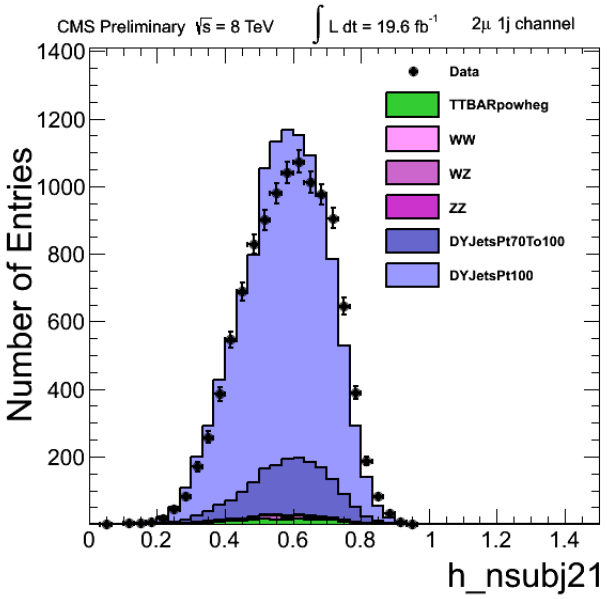

Figure 4.25. τ_{21} distribution for data and MC background in the sideband region.

Consequently, in this study it has been used a data driven approach already tested in these other analyses [32], [37] and [38]. It is based on a $t\bar{t}$ control sample that provides a boosted W boson sample on which the same selection and τ_{21} categorization of the main analysis are applied. The data/MC scale factors obtained with this boosted W sample, reported in table 4.2, are applied also in analyses with boosted Z [32] and in this case with a boosted Higgs.

Cat.	Efficiency scale factor
HP	0.93 ± 0.08
HL	1.10 ± 0.30

Table 4.2. τ_{21} selection efficiency scale factor.

Finally the τ_{21} selection MC efficiency is corrected with this scale factor accordingly in the high purity or low purity category.

Table 4.3 summarizes the total efficiency for the four channels. This efficiency is interpolated to all mass points considered in the analysis (figure 4.26). The function used for the fit is a third degree polynomial. We choose this analytical approach in order to be as much independent as possible from the mass of the samples and to be able to cover analytically the complete range between the highest and the lowest mass signal sample. It's obvious that the more samples we use and the more accurate the predicted efficiency

M_X [GeV]	Signal efficiency				
	electron channel		muon channel		Total Sum
	HP	LP	HP	LP	
800	0.085	0.016	0.093	0.024	0.219
900	0.102	0.023	0.120	0.025	0.271
1000	0.118	0.030	0.132	0.030	0.310
1100	0.125	0.037	0.126	0.037	0.325
1200	0.135	0.037	0.134	0.033	0.340
1300	0.127	0.039	0.137	0.035	0.340
1400	0.126	0.035	0.126	0.043	0.331
1500	0.130	0.035	0.128	0.039	0.333
1600	0.119	0.038	0.120	0.035	0.313
1700	0.117	0.039	0.116	0.032	0.305
1800	0.120	0.039	0.107	0.038	0.303
1900	0.116	0.038	0.105	0.036	0.296
2000	0.113	0.040	0.105	0.033	0.290

Table 4.3. Total signal efficiency as defined in equation 4.8 devided in the four different categories and summed up to their combination.

is.

Note that, expecially in the high purity category, we see a small decrease of the efficiency at high masses; this loss is due to the case in which the Z and Higgs bosons are highly boosted and consequently the products of the Higgs (quarks or gluons) are emitted with small ΔR (fig. 3.5). In this configuration the pruning algorithm presented in section 3.3.1 starts to be suboptimal, removing an excessive fraction of the jet mass. This effect is enhanced by the fact that the low edge of the signal region is quite high compared to the resolution (see discussion in section 4.7.1).

4.8 Background extrapolation

The computation of the expected background is the most difficult part of the analysis. Since, as we already said, the final aim of this analysis is to compare the predicted SM background with the observed data, it is important to elaborate a trustworthy strategy for the background estimation. Indeed, despite the good description of the event kinematics provided by the MC simulation, it is more advisable to minimize the dependence on the MC and develop a data driven strategy.

We have already defined in section 4.7.1 our signal region, we need now a sideband region to be used as a pure background control region. Two possibilities have been taken into account

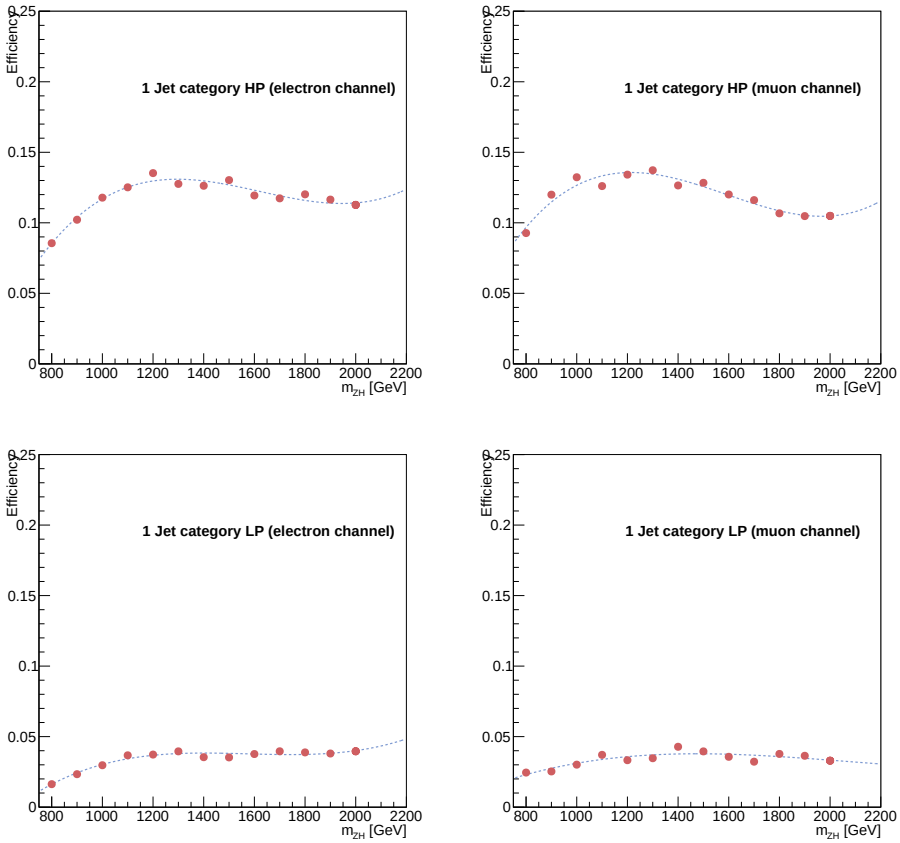

Figure 4.26. Signal efficiency divided in the four categories: high purity (top row), low purity (bottom row) and electron channel (left) and muon channel (right) as function of mass fitted with a third degree polynomial .

- pruned jet mass $\in$ [50, 70] GeV
- pruned jet mass $\in$ [50, 110] GeV.

Both of them are coherently defined out of the signal region and a study on the different performances is presented in the next section.

In order to estimate the final background, we consider the m_{ZH} MC mass spectrum in the signal and sideband region. A ratio $\alpha(m_{ZH})$ of the two is created. This α factor allows a prediction of the mass spectrum in the signal region starting from the measured distribution in the sideband. Under the assumption that this estrapolation from the sideband to the signal region works in the same way both for data and MC, we can estimate the final background distribution by multipling the m_{ZH} mass spectrum observed in

the sideband by this α ratio, obtaining

$$N_{bkg}(m_{ZH}) = N_{sb}^{data}(m_{ZH}) \times \frac{N_{bkg}^{MC}(m_{ZH})}{N_{sb}^{MC}(m_{ZH})} = N_{sb}^{data}(m_{ZH}) \times \alpha(m_{ZH}) \quad (4.9)$$

The same procedure is applied also in [32] and its trustworthiness is guaranteed by the good agreement between data and MC in the whole range of the pruned mass distribution (fig. 4.27)[1].

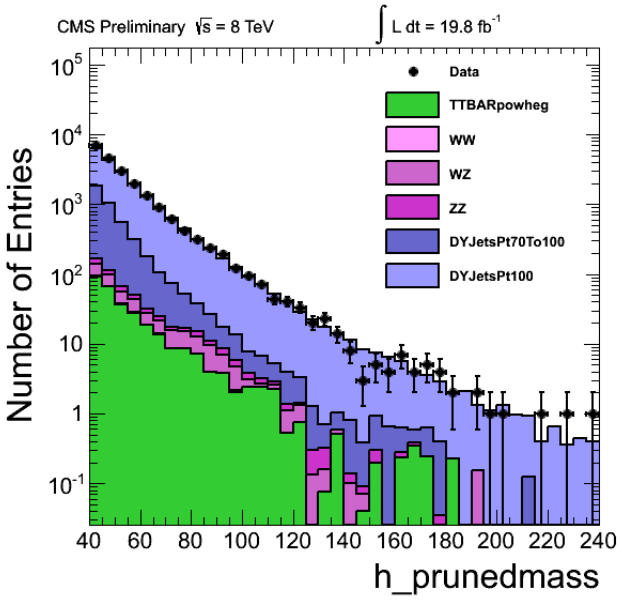

Figure 4.27. Pruned mass distribution in the all m_J pre-selection range.

4.8.1 Sidebands region

We have already mentioned the importance of a well known control region where we can check the correct behaviour of the MC (background) simulation compared to the observed data. Indeed, such a control region should contain a pure background sample and it is tipically defined as the sidebands of the signal region in the distribution of the main discriminating variables. In our case, we don't consider a right sideband of the pruned mass distribution, higher than 140 GeV, because of the poor statistics and the excessive contribution of $t\bar{t}$ events.

At this point we have to select wisely an adequate left sideband region. We have already stressed that this work follows in the ZZ analysis's footsteps,

[1] An additional confirmation is given in [32]: a closure test is performed and the higher mass sideband is extracted by the lower one; in our case we can not proceed this test because of the low statistics for jet masses above 140 GeV (figure 4.27).

the latter has its own signal and sidebands region and it has published exclusion limits with no evidence of new resonances in [32]. The two possible selection regions of this analysis are skecthed in table 4.4. From now on, we will refer to the two cases as:

1. thin sideband: [50, 70] GeV
2. large sideband: [50, 110] GeV

The former reproduces the same sideband than the ZZ analysis, skipping the whole intermediate range corresponding to the signal region of the ZZ search. The weakness of this choice is the lack of statistics at high masses, due to the small range and low value of the sideband considered. In fact, although the background is exponentially distributed in term of the invariant ZH mass (fig. 4.14), the jet mass and the final invariant mass are strongly correlated and the extension of the sideband up to 110 GeV largely helps the increasing of the population of the high invariant mass region (above 1800 GeV).

This second approach is still coherent and we are confident that no bias is added by the enlarged range, firstly because the theoretical model at the base of this search (presented in section 1.2.1) does not allow the ZZ decay of the predicted resonance, and then because the results, published internally in CMS, don't show any evidence of signal of new physics in the ZZ channel.

Pruned jet mass	$X \to ZH$	
[GeV]	thin sideband	large sideband
Signal region	[110,140]	
Sidebands region	[50, 70]	[50, 110]

Table 4.4. Signal and sidebands range selection related to the jet pruned mass distribution. In the first column is added the selection adopted in the $BulkG \to ZZ$ analysis [32].

4.8.2 $\alpha(m_{ZH})$ ratio

We divide the spectrum in 14 not uniform width bins (table 4.5), accordingly to the decreasing statistics in the high mass tail.

The MC background distribution in the signal region is used to explore the range where the invariant mass is well described by an exponential function.

From figure 4.28 we see that the exponential trend only starts from around 850 GeV (marked in green in the plot). Accordingly, the extrapolation procedure of the background starts from the fifth bin of table 4.5.

We can now proceed with the computation of the $\alpha(m_{ZH})$ ratio separately for the HP and LP categories. In figures 4.29, 4.30, 4.31 and 4.32 are

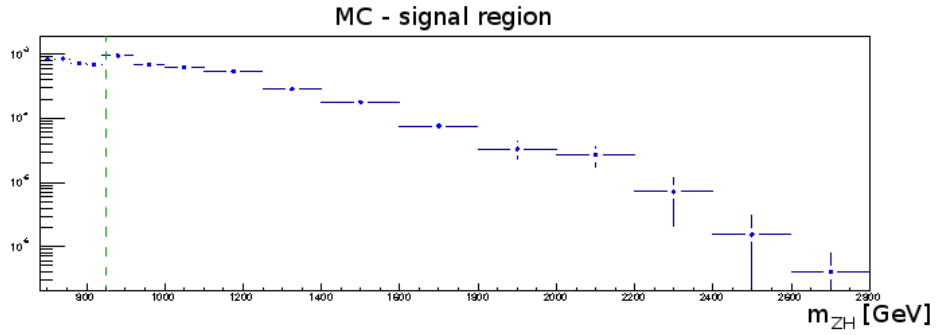

Figure 4.28. Distribution of the invariant mass in the signal region for the background MC simulation.

Bin	GeV
1	[680, 720]
2	[720, 760]
3	[760, 800]
4	[800, 840]
5	[840, 920]
6	[920, 1000]
7	[1000, 1100]
8	[1100, 1250]
9	[1250, 1400]
10	[1400, 1600]
11	[1600, 1800]
12	[1800, 2000]
13	[2000, 2200]
14	[2200, 2400]

Table 4.5. Binning of the X invariant mass range.

reported the MC histograms and the resulting α ratio as a function of m_{ZH} in the four cases: thin sideband/large sideband and high purity/low purity. In all of them, the top histogram contains the MC background prediction, it is fitted either with a simple exponential or a leveled exponential defined in Eq. 4.10. The middle row shows the MC sideband region fitted with a simple exponential, and, finally, the bottom plot shows the ratio of the two histograms. A smoothing algorithm is applied to the α ratio, it is needed expecially at high masses, for the thin sideband case in the high purity configuration. Indeed, in this case, the lack of statistics leads to large statistical fluctuations. The smoothing makes the α factor more continuos: when the content of a bin is too far from the values of its neighbours, it is fixed to the

their mean.

Finally, one notes a higher stability given by the [50, 110] GeV sideband sample. Expecially in the high purity category, where we expect less background events, the α ratio shows less fluctuations and the related error bars are smaller.

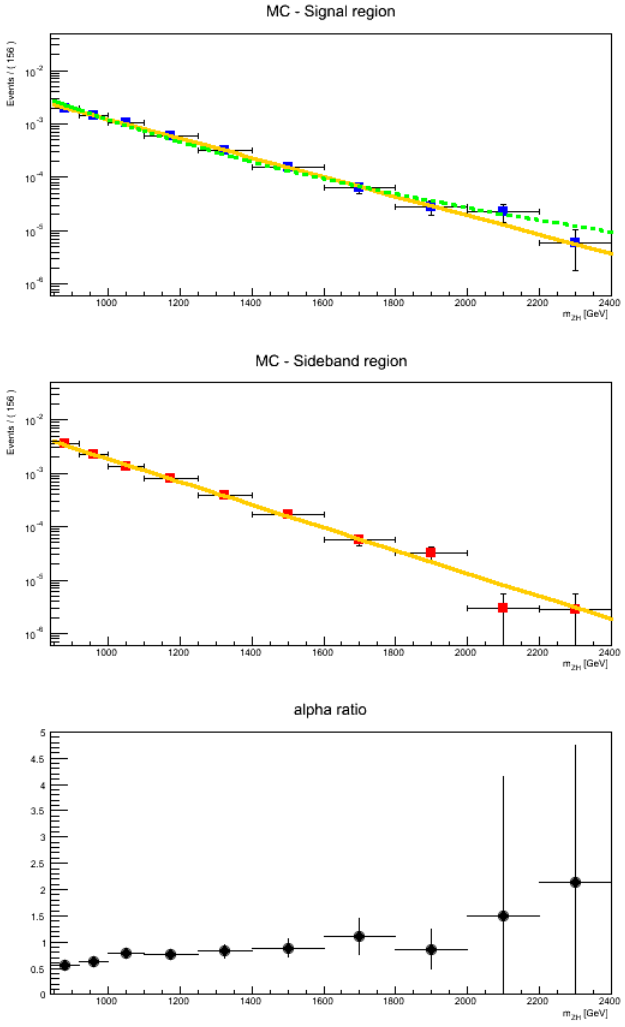

Figure 4.29. $\alpha(m_{ZH})$ ratio computation procedure using thin sideband in the high purity category. Top row: MC background simulation in the signal region fitted either with a simple exponential (yellow) or with a leveled exponential (green). Middle row: MC background simulation in the sideband region, an exponential fit is shown in yellow. Bottom row: the α ratio $MC^{\text{sigreg}}/MC^{\text{sb}}$ as function of m_{ZH}.

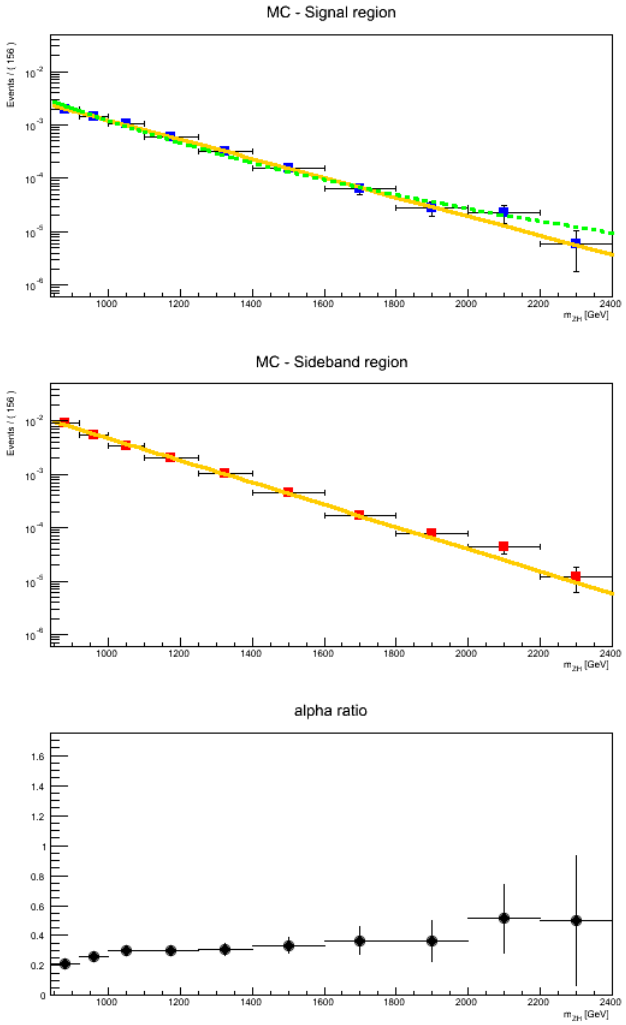

Figure 4.30. $\alpha(m_{ZH})$ ratio computation procedure using large sideband in the high purity category. Top row: MC background simulation in the signal region fitted either with a simple exponential (yellow) or with a leveled exponential (green). Middle row: MC background simulation in the sideband region, an exponential fit is shown in yellow. Bottom row: the α ratio $MC^{\text{sigreg}}/MC^{\text{sb}}$ as function of m_{ZH}.

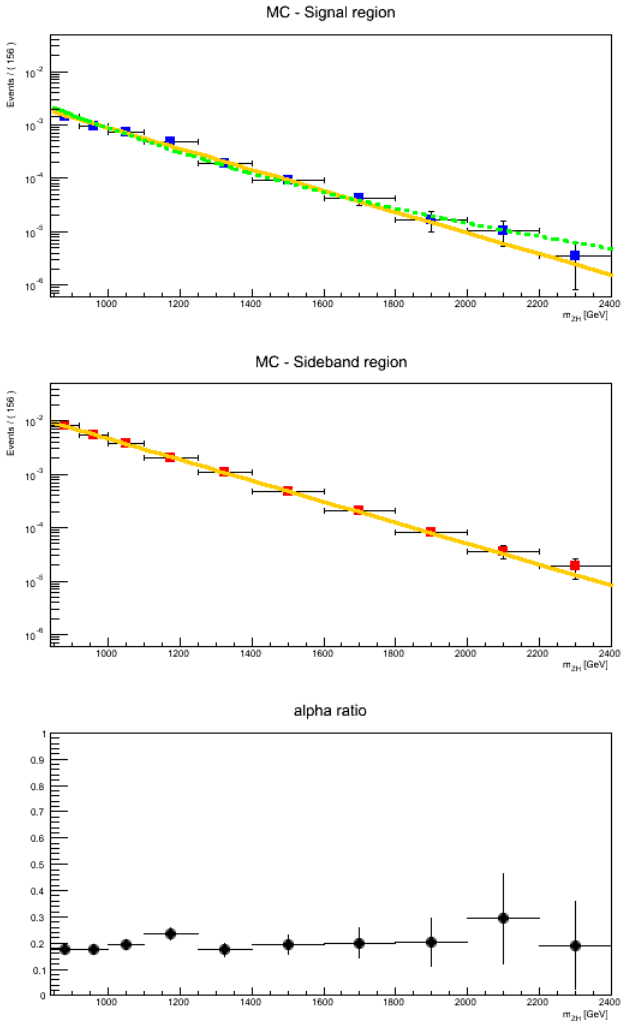

Figure 4.31. $\alpha(m_{ZH})$ ratio computation procedure using thin sideband in the low purity category. Top row: MC background simulation in the signal region fitted either with a simple exponential (yellow) or with a leveled exponential (green). Middle row: MC background simulation in the sideband region, an exponential fit is shown in yellow. Bottom row: the α ratio $MC^{\text{sigreg}}/MC^{\text{sb}}$ as function of m_{ZH}.

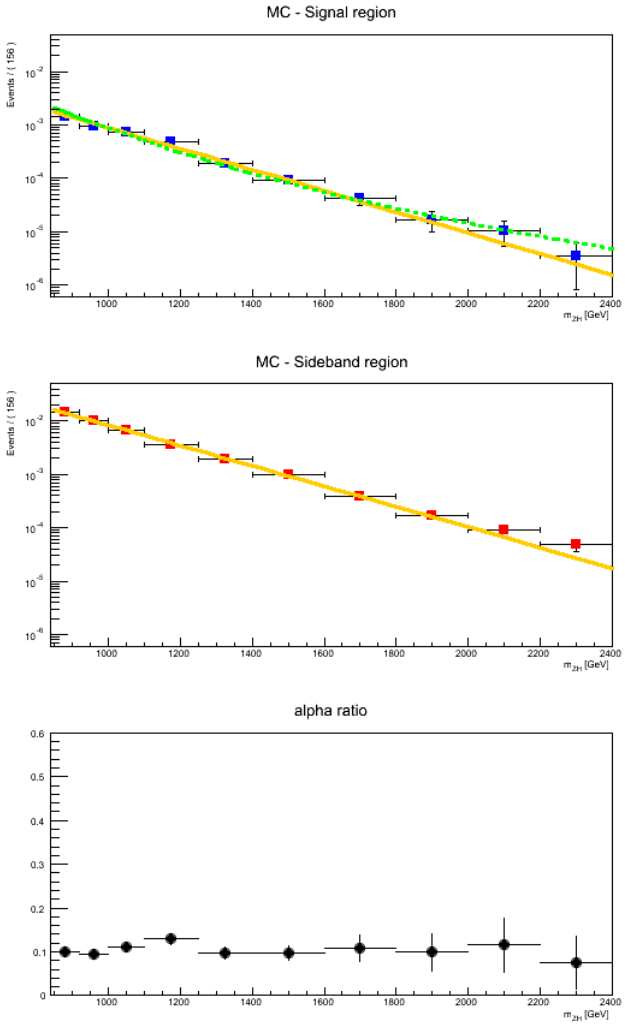

Figure 4.32. $\alpha(m_{ZH})$ ratio computation procedure using large sideband in the low purity category. Top row: MC background simulation in the signal region fitted either with a simple exponential (yellow) or with a leveled exponential (green). Middle row: MC background simulation in the sideband region, an exponential fit is shown in yellow. Bottom row: the α ratio $MC^{\text{sigreg}}/MC^{\text{sb}}$ as function of m_{ZH}.

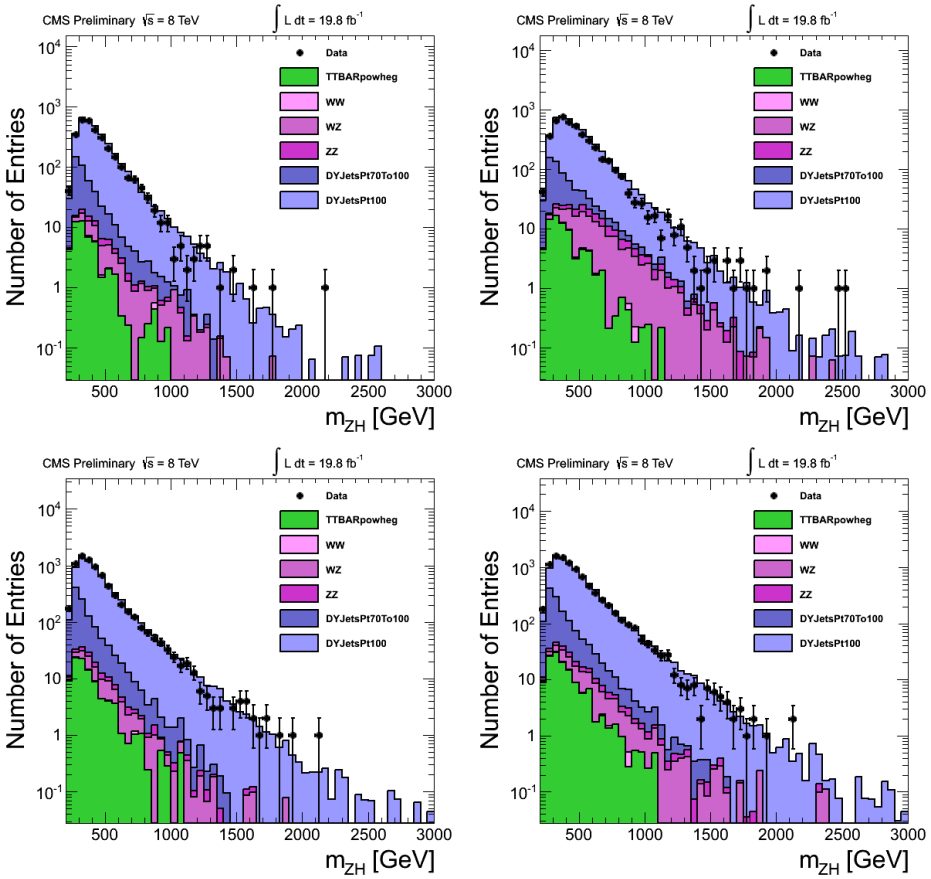

Figure 4.33. Data and MC m_{ZH} observed distribution in the sideband region: thin sideband (left), large sideband (right) and high purity (top) and low purity category (bottom).

4.8.3 Fit to background

Once we have the α ratio for the two categories, following the strategy of section 4.8, we look to the data in the sideband region. Figure 4.33 shows the data/MC agreement of the m_{ZH} distribution in the sideband region in the four considered cases. Again, with the enlarged sideband range, an important improvement in the statistics of the high mass bins is observed.

Therefore, we obtain the final data-driven background estimation as in equation 4.9. Its shape is then fitted using either a simple exponential or a leveled exponential function defined as:

$$f(m_{ZH}) = \exp\left[\frac{-(m_{ZH} - m_0)}{\sigma + p_0(m_{ZH} - m_0) + p_1(m_{ZH} - m_0)^2}\right]. \qquad (4.10)$$

From this definition it follows that for the values of the parameters $p_0 =$

$p_1 = 0$ one recovers the simple exponential.

For semplicity criteria, we limit our attempts to functions with at most two free parameters: either p_0 or p_1 is fixed to zero. The diagonalization procedure of the covariance matrix, implemented in order to obtain the errors of the fit, becomes computationally harder and less stable with the increasing number of fit parameters.

Figures 4.34 and 4.35 show the background distribution fitted with the different functions, separately for the high purity and low purity category. The fit is performed in the range [850, 2400] GeV and the m_0 parameter is set to 750 GeV.

One can immediately see that the extension of the sideband up to 110 GeV considerably improves the shape of the background histogram at high masses, expecially because the [50, 70] GeV data may not fill every bin of the mass range. Then, one can note that in the large sideband case, the leveled exponential fit is so close to the simple exponential function that there are no distinctions between the first and the second degree polynomial denominators. On the other hand, in the thin sideband computation, at high masses one can see the typical higher tail of the second degree polynomial leveled function.

Finally, we conclude that there is no risk to work with the enlarged sideband, indeed, it is more convenient from a statistical point of view. About the choise between the simple and the leveled exponential fit function, a study presented in [32] concluded that it is more trustworthy to fit the background shape with a leveled exponential. Otherwise, if the background really follows a leveled distribution, fitting with a simple exponential would introduce some biases in the signal.
It is also clear that there is no need to add other parameters more than the second degree polynomial of equation 4.10.

Table 4.6 summarizes the final values of the fit parameters that are used for the limit setting procedure. One notices that the values of the p_0 parameter are compatible with zero in the errors of the fit. This is the consequence of the fact that the simple exponential describes well the data, as already discussed above.

	HP	LP
σ [GeV]	$2.8265\,10^2 \pm 3.72\,10^2$	$1.9525\,10^2 \pm 2.98\,10^1$
p_0	$1.5376\,10^{-3} \pm 1.47\,10^{-1}$	$1.5513\,10^{-2} \pm 3.18\,10^{-2}$

Table 4.6. Final values of the background fit parameters in the large sideband case and with a leveled exponential with $p_1 = 0$. The associated errors contain only the uncertainty given by the statistics of the data in the sideband region.

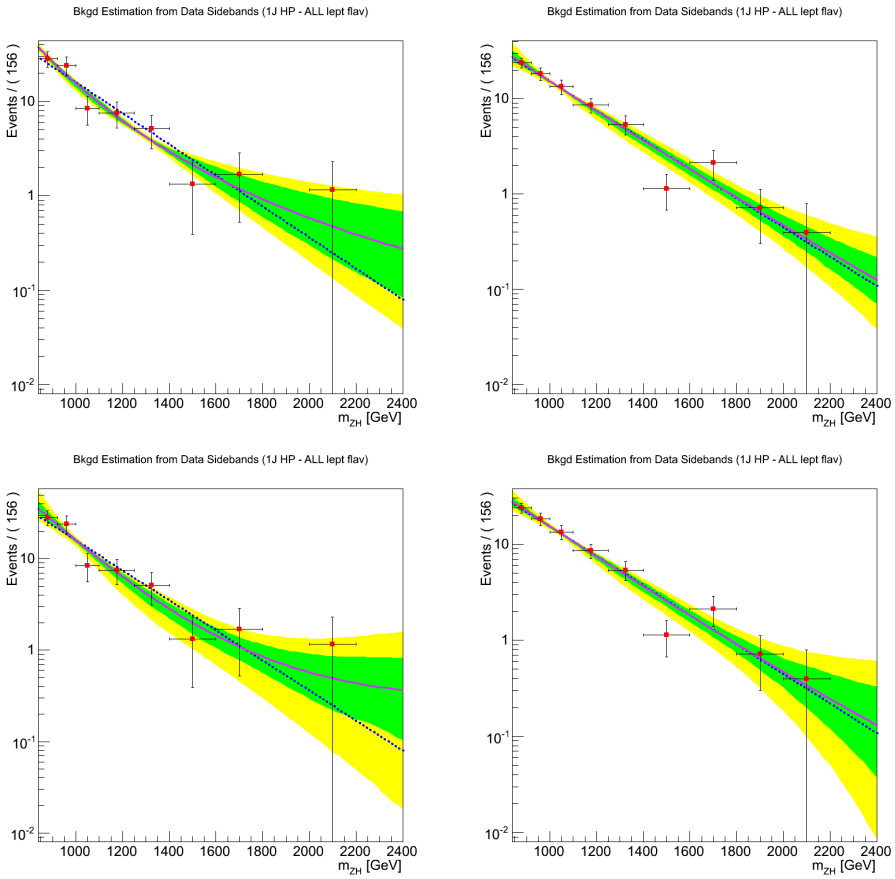

Figure 4.34. Final background prediction in the high purity category based on thin sideband (left) or large sideband (right). The purple line is the fit with a leveled exponential of first degree [$p_1 = 0$] (top) and of second degree [$p_0 = 0$] (bottom). The green and yellow area represent respectively 1σ and 2σ of the statistical uncertainty of the fit. The fit with the simple exponential [$p_0 = 0$ and $p_1 = 0$] is also plotted with a dashed blue line.

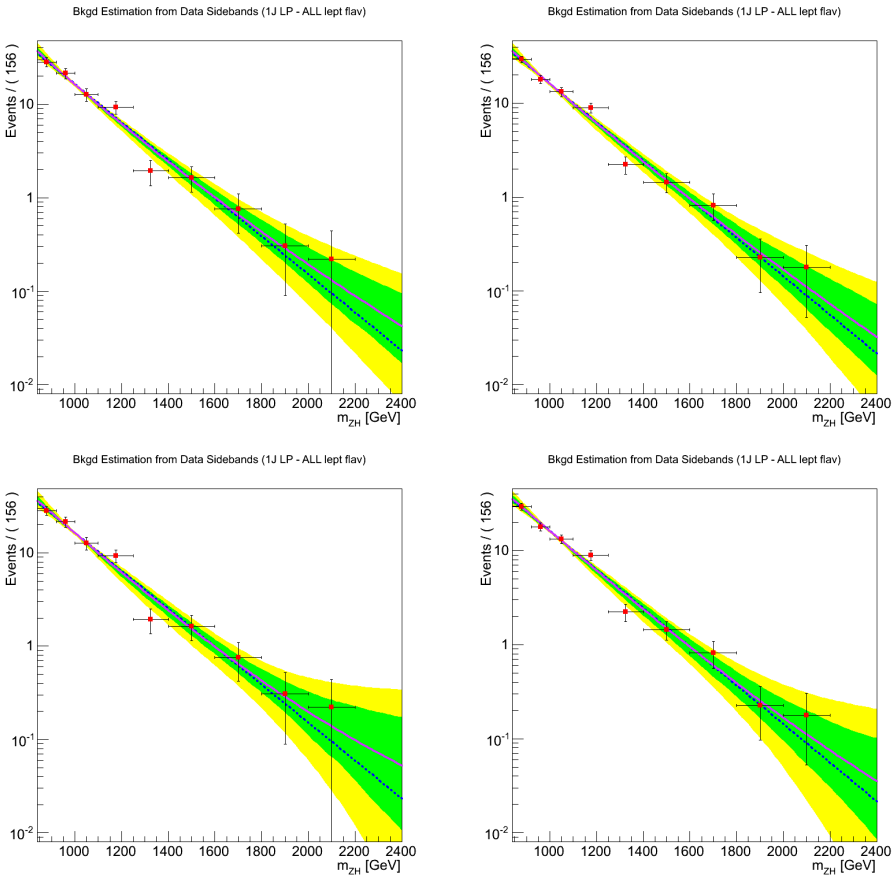

Figure 4.35. Final background prediction in the low purity category based on thin sideband (left) or large sideband (right). The purple line is the fit with a leveled exponential of first degree [$p_1 = 0$] (top) and of second degree [$p_0 = 0$] (bottom). The green and yellow area represent respectively 1σ and 2σ of the statistical uncertainty of the fit. The fit with the simple exponential [$p_0 = 0$ and $p_1 = 0$] is also plotted with a dashed blue line.

The uncertainties that affect the predicted background are statistical in nature and come from the limited size of the data and MC samples. The error due to the fluctuations of the data in the sideband region is reported in the graph as errorbars and included in the fitting procedure. The uncertainty associated to the α factor, coming from the finite size of the MC samples, is calculated in a different way: a large number (500) of alternative sets of extrapolation factors (α_{ZH}) is randomly generated from their nominal value and accondingly to the statistical fluctuations of the MC samples (figures 4.30 and 4.32). Then, the whole background estimation procedure is applied for each of these pseudo-experiments; they are multiplied by the data spectrum in the sideband (fixed) and the result is fitted with a leveled exponential. The standard deviation of the fit parameter distributions is associated to the systematic uncertainty of the fit.

Figure 4.36 shows the α ratio distribution of the 500 pseudo experiments generated, the central value and the statistics uncertainty, on which is based the random generation, are superimposed in red. Figure 4.37 shows the distribution of their resulting fit parametres. Note that this systematic error is of the same order of magnitude of the statistical uncertainty of the fit (marked on the the same plots).

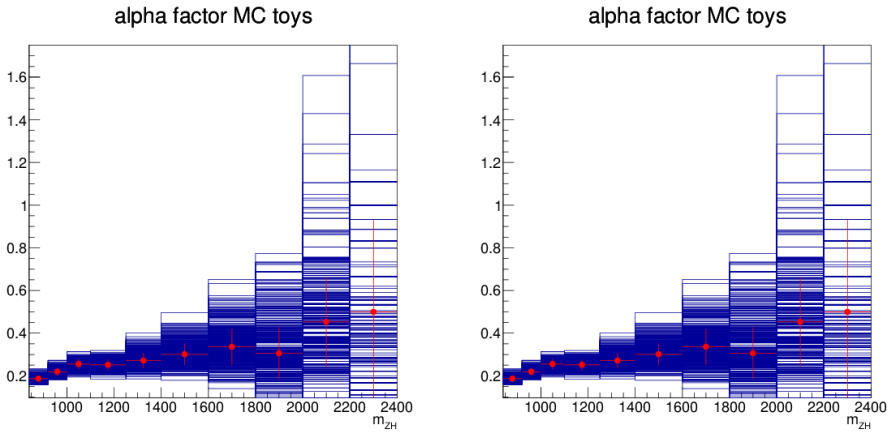

Figure 4.36. Distribution of the α factor of the 500 pseudo-experiments generated from the MC samples.

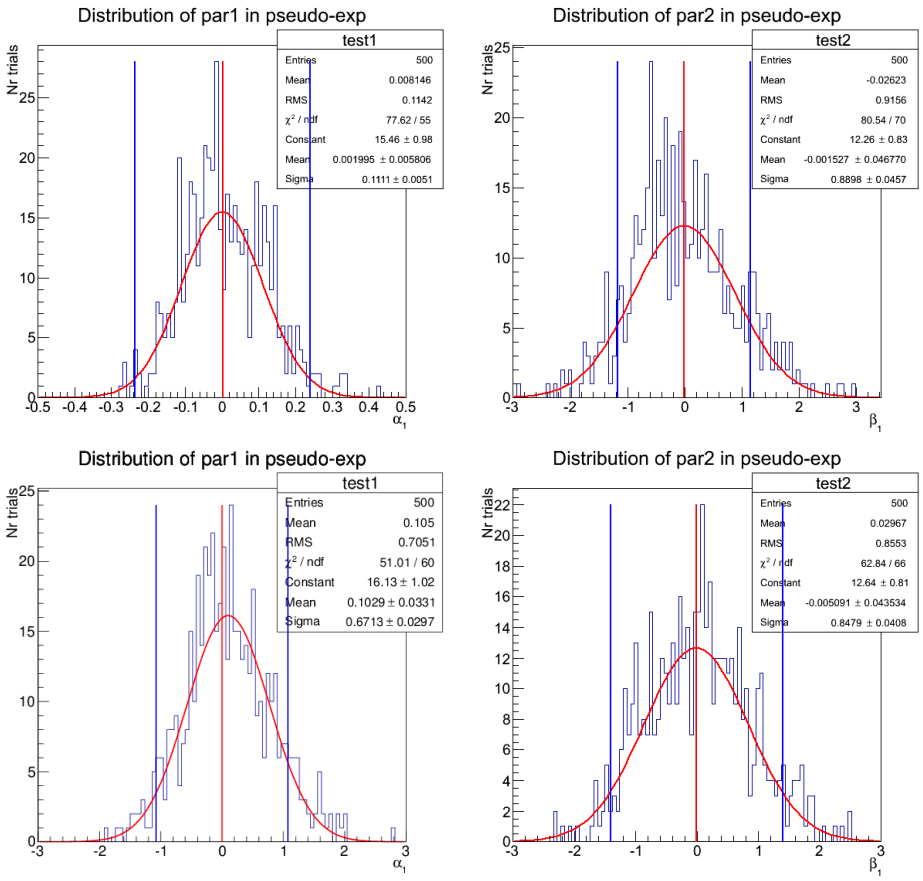

Figure 4.37. Distribution of the difference of the output fit parameters in the 500 generated pseudo-experiments compared to the nominal values (red line) and to the statistical uncertainty on the nominal value (blue lines) calculated from the covariance matrix of the fit done assuming that the error on the background extrapolation is due only to the statistics of the data sideband. Top row: high purity category; bottom row: low purity category. Left: σ; right: p_0.

4.9 Signal shape

The natural width of the resonance, generated by the samples of table 3.1, is sufficiently small to be neglected when compared to the detector resolution. The reconstructed signal distribution varies depending on the resonance mass and the lepton flavour.

Figures 4.38, 4.39, 4.40, 4.41 and 4.42 show all the 13 mass points divided in the electron and muon channels. The shapes are fitted with a double Crystal-Ball. The double Crystal-Ball function is composed by a gaussian core with powerlaw tails on both sides and it is found to describe fairly the CMS detector resolution. It is defined as

$$dCB(x;\alpha_1,n_1,\alpha_2,n_2,\bar{x},\sigma) = N\cdot\begin{cases} A_1\cdot(B_1-\frac{x-\bar{x}}{\sigma})^{-n_1}, & \text{for } \frac{x-\bar{x}}{\sigma} > -\alpha_1 \\ \exp\left(-\frac{(x-\bar{x})^2}{2\sigma^2}\right), & \text{for } \alpha_1 < \frac{x-\bar{x}}{\sigma} < \alpha_2 \\ A_2\cdot(B_2-\frac{x-\bar{x}}{\sigma})^{-n_2}, & \text{for } \frac{x-\bar{x}}{\sigma} > \alpha_2 \end{cases} \tag{4.11}$$

where α_1, n_1, α_2, n_2, $\bar{x}$ and σ are free parameters and A_1, B_1, A_2 and B_2 are functions of the previous parameters to make the total function continuous.

Table 4.7 reports the output parameters of the gaussian core of the fit.

Mass [GeV]	electron channel		muon channel	
	CB mean [GeV]	CB sigma [GeV]	CB mean [GeV]	CB sigma [GeV]
800	803	23	800	28
900	900	26	898	31
1000	1002	30	995	38
1100	1103	31	1097	42
1200	1205	34	1196	40
1300	1303	33	1292	44
1400	1404	35	1391	54
1500	1502	39	1490	62
1600	1600	43	1596	50
1700	1698	43	1694	63
1800	1797	46	1800	60
1900	1901	47	1895	79
2000	2002	47	1986	82

Table 4.7. Output fit parameters of the signal shape resolution separately for electron and muon channels.

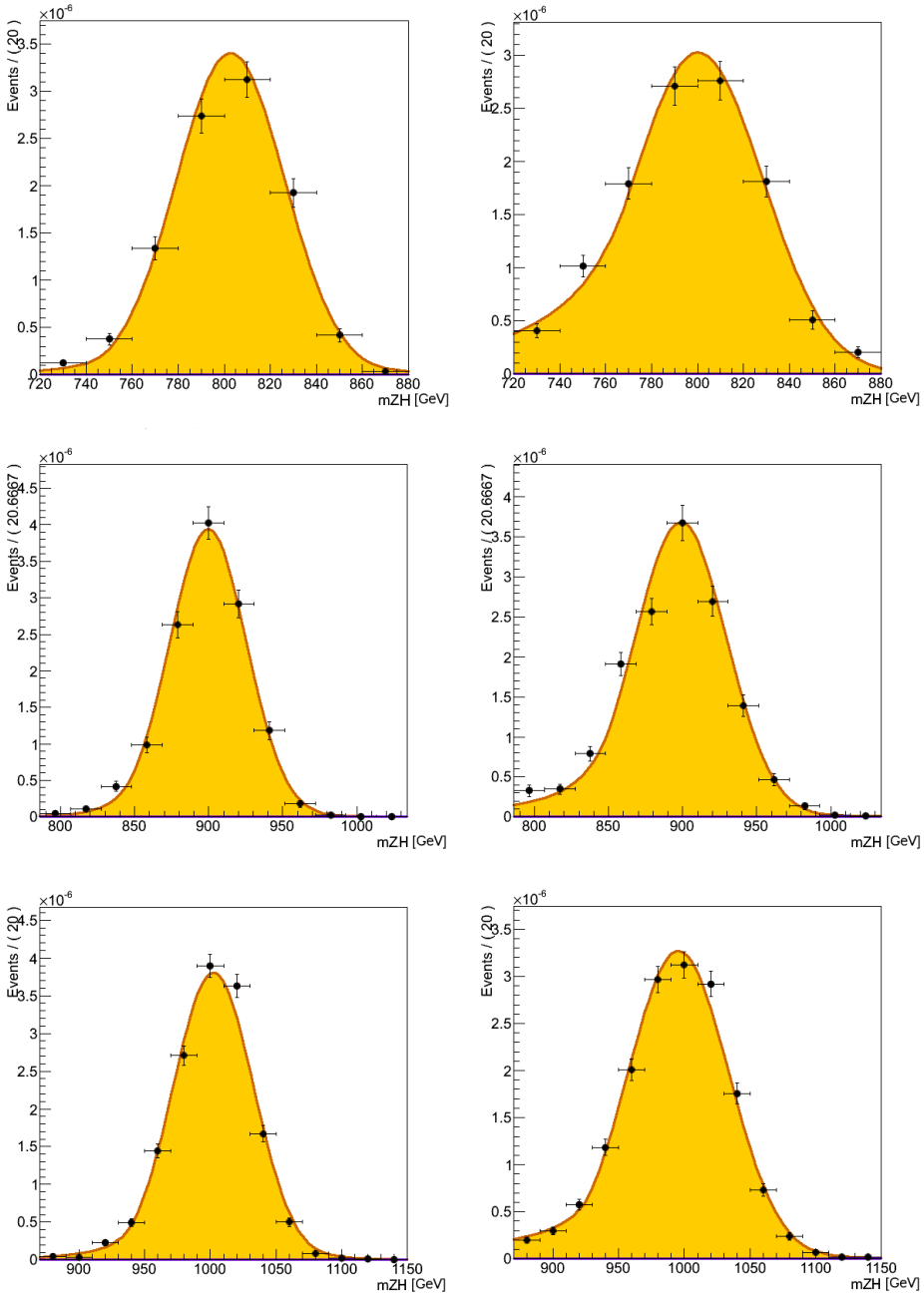

Figure 4.38. Mass distribution of the signal samples in the electron (left) and muon (right) channels. From top to bottom: resonance mass of 800, 900 and 1000 GeV.

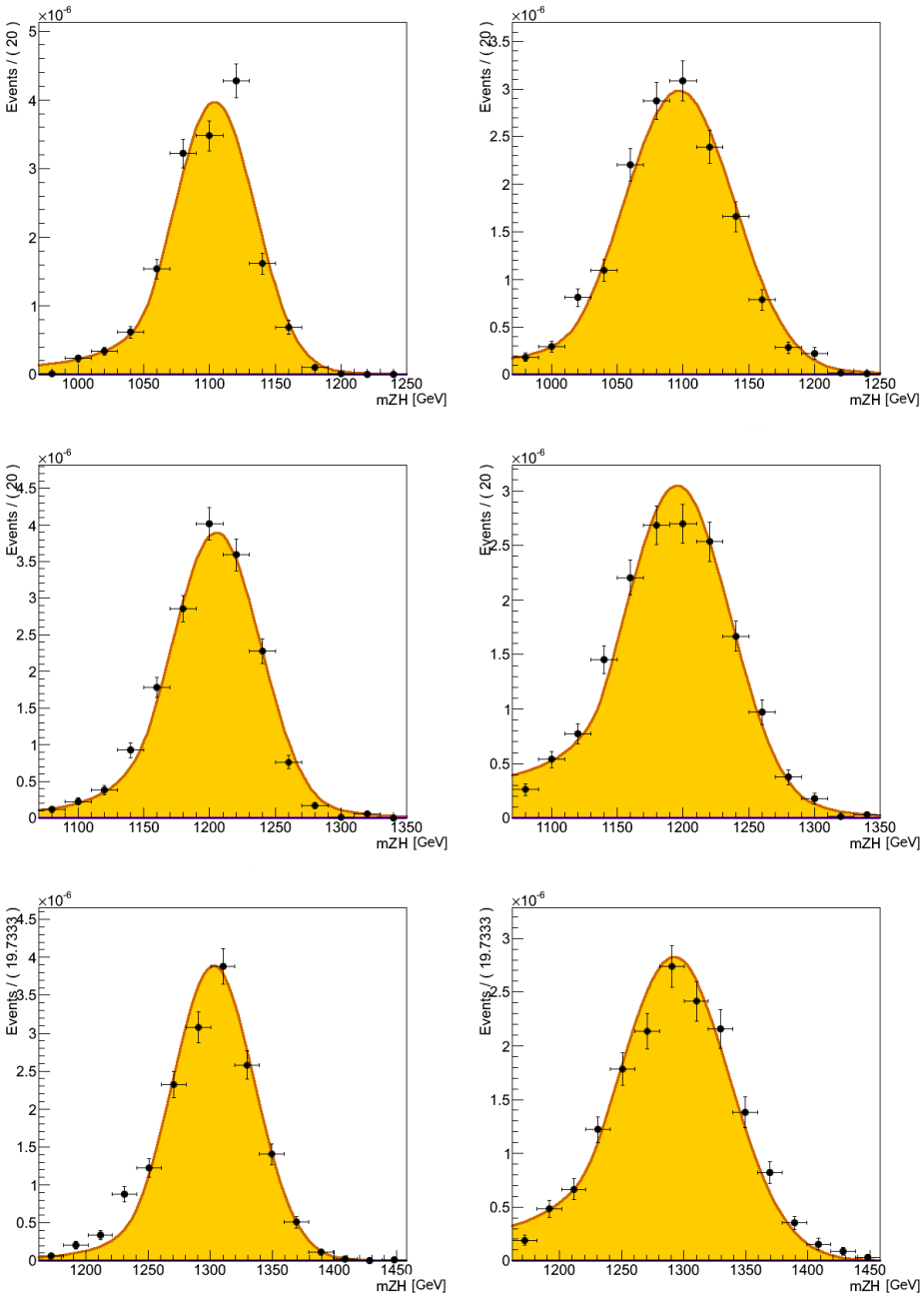

Figure 4.39. Mass distribution of the signal samples in the electron (left) and muon (right) channels. From top to bottom: resonance mass of 1100, 1200 and 1300 GeV.

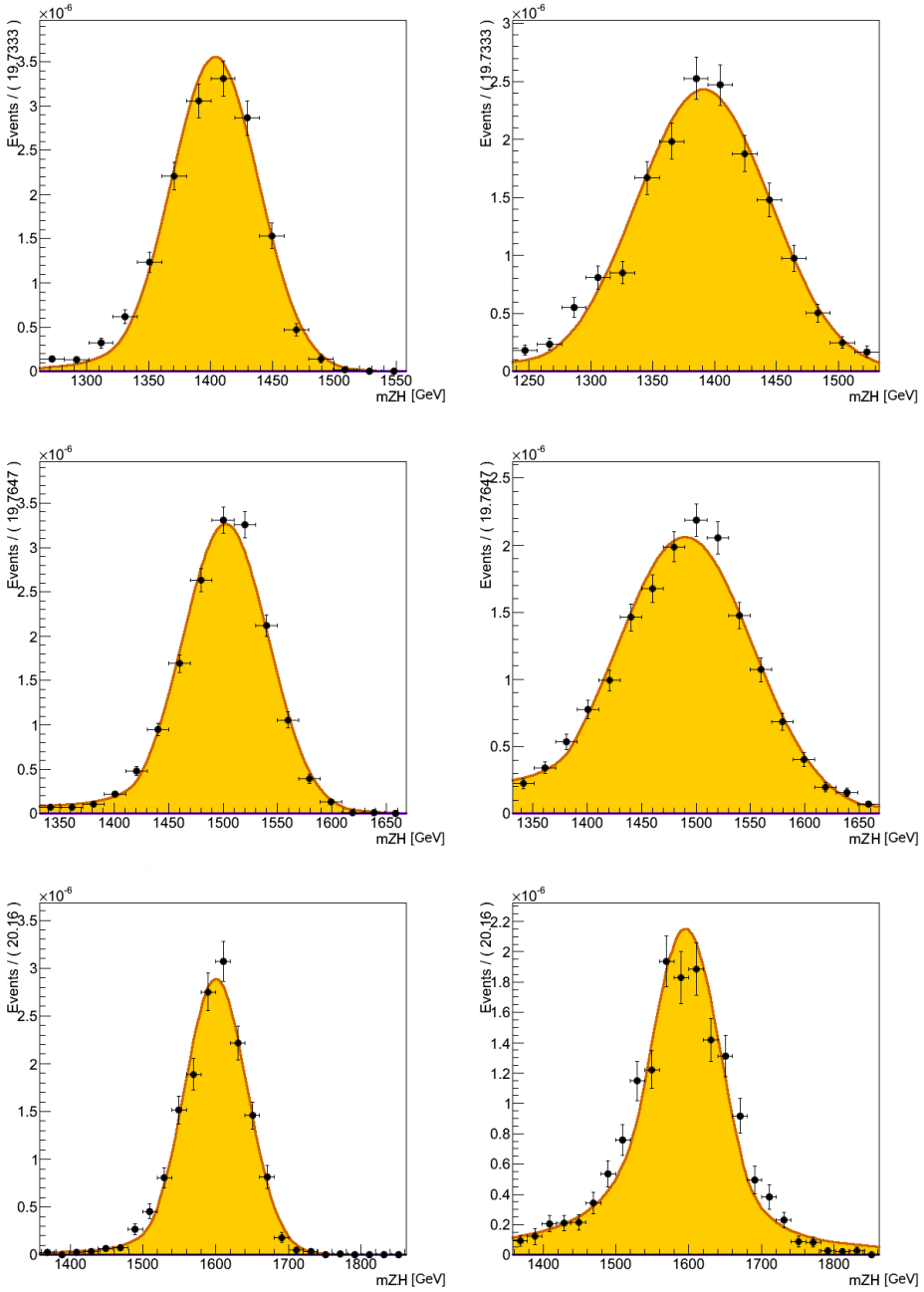

Figure 4.40. Mass distribution of the signal samples in the electron (left) and muon (right) channels. From top to bottom: resonance mass of 1400, 1500 and 1600 GeV.

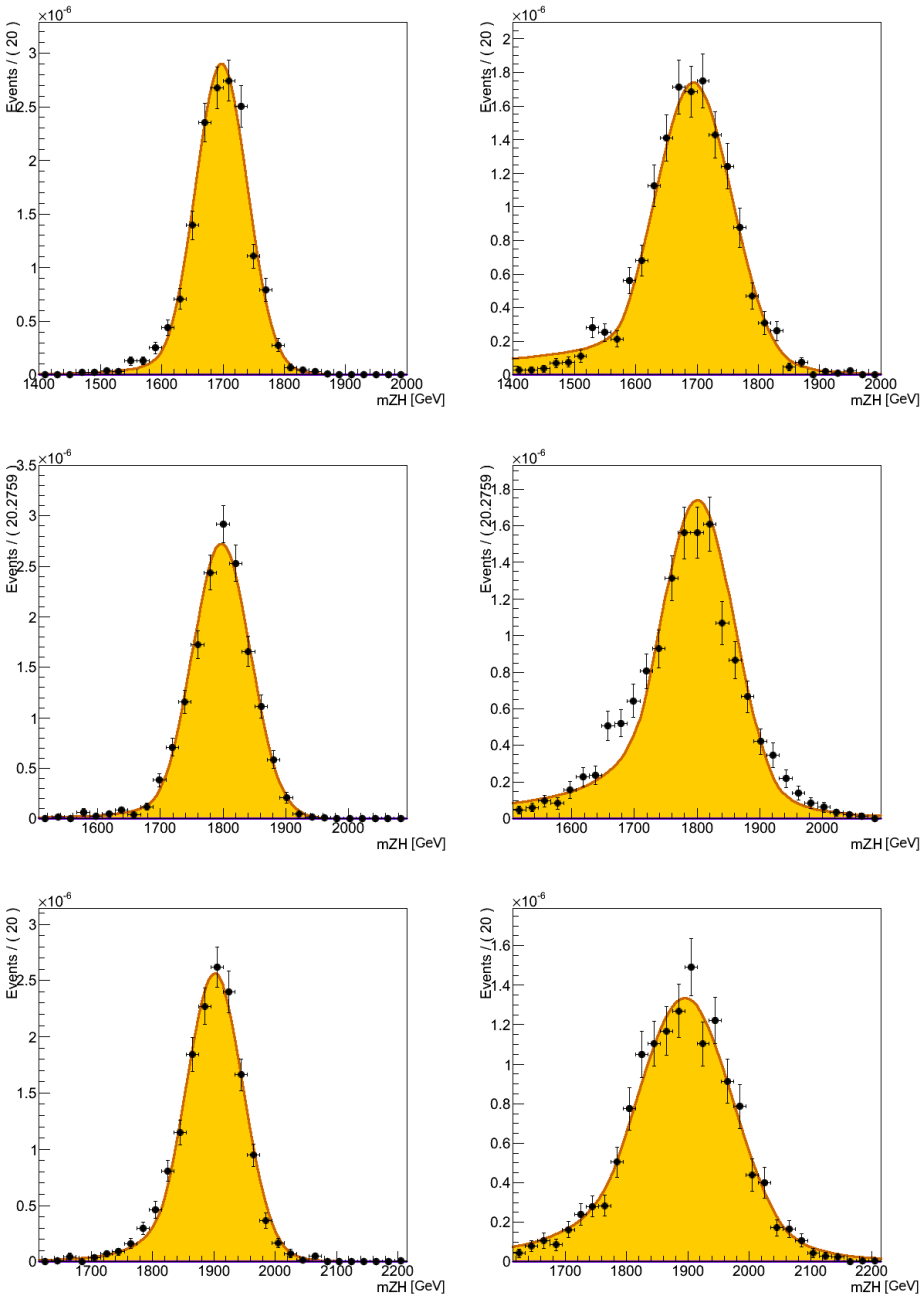

Figure 4.41. Mass distribution of the signal samples in the electron (left) and muon (right) channels. From top to bottom: resonance mass of 1700, 1800 and 1900 GeV.

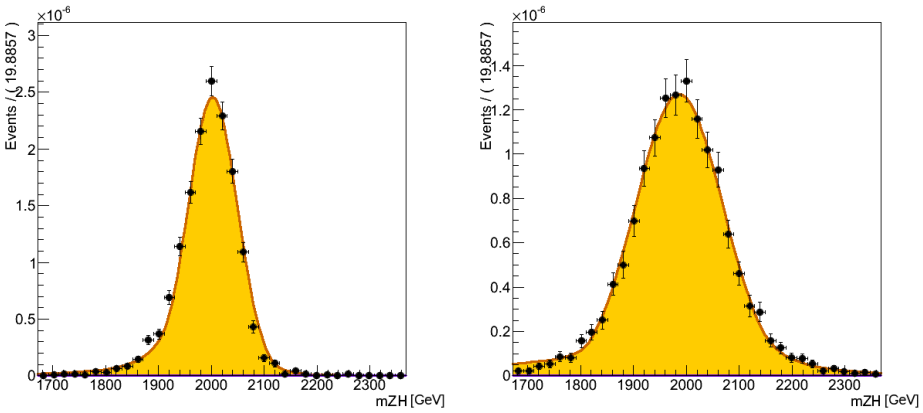

Figure 4.42. Mass distribution of the signal samples in the electron (left) and muon (right) channels for a resonance 2000 GeV.

4.10 Systematic uncertainties on the signal yield

The main systematic uncertainties on signal yield are summarized in table 4.8. The lepton and trigger efficiency uncertainties are estimated with the tag-and-probe technique, while effects of jet reconstruction are evaluated by changing the jet energy and resolution within calibration uncertainties. Variations on the lepton and jet energy scales and resolutions affect the reconstruction of the heavy resonance in three possible ways: changes on the final reconstruction efficiency and changes on the peak and on the width of the signal shape in the m_{ZH} distribution [32].

The dominant systematic error is given by the uncertainty associated to the H-tagging scale factor introduced in section 4.7.5. This scale factor is computed on a sample of boosted W bosons, the large uncertainty is

Source	HP	LP
Muon trigger & ID	5%	
Muon scale & resolution	2%	
Electron trigger & ID	$\sim$ 3%	
Electron scale & resolution	$<$ 0.1%	
Jet energy scale	1%	
H-tagging	$\sim$9%	$\sim$27%
Proton PDFs	0.4%	
Luminosity	4.4%	

Table 4.8. Summary of the main systematic uncertainties on signal [32].

statistical in nature and is mostly driven by the available number of events in the $t\bar{t}$ control region.

4.11 Results

We are now at the final step of the analysis, we have the background SM prediction and the signal shape hypothesis. Table 4.9 shows the expected yields from background and from signal in the high purity category at different masses. The background number of events is calculated integrating the leveled exponential function, obtained by the fit in figure 4.34, in a mass window of $\pm 15\%$ around the nominal value. The range is chosen in order to totally contain the signal, whose resolution is found to be around 3-4% (table 4.7). Figure 4.43 shows the m_{ZH} invariant mass distribution of the selected events after all the event selections for data and MC background in the four different categories (HP/LP, electron/muon); the red line represents the data driven background estimation presented in section 4.8. The muon and the electron samples are treated separately in the fit since they have different mass resolution. The four categories are summed up in figure 4.44, where the difference divided by the uncertainty $((data - MC_{bkg})/err)$ is also shown.

Mass [GeV]	**N. bkg expected** (HP)	**N. sig expected** (HP) [$\sigma = 1$ fb]
800	45.2	3.5
900	39.3	4.4
1000	30.9	4.9
1100	24.2	4.9
1200	18.7	5.3
1300	14.4	5.2
1400	11.1	5.0
1500	8.5	5.1
1600	6.4	4.7
1700	4.9	4.6
1800	3.7	4.5
1900	2.8	4.3
2000	2.1	4.3

Table 4.9. Expected number of background events in a mass window of $\pm 15\%$ of the nominal mass value compared to the signal production with a cross section of reference of 1 fb^{-1}. These values concern to the final $pp \to ZH \to ll1jet$ selection and assume: $\sigma(pp \to X \to ZH) = 1$ fb, BR$(ZH \to llqq(gg)) = 4.65\%$ (see section 4.7.5) and the HP selection efficiency.

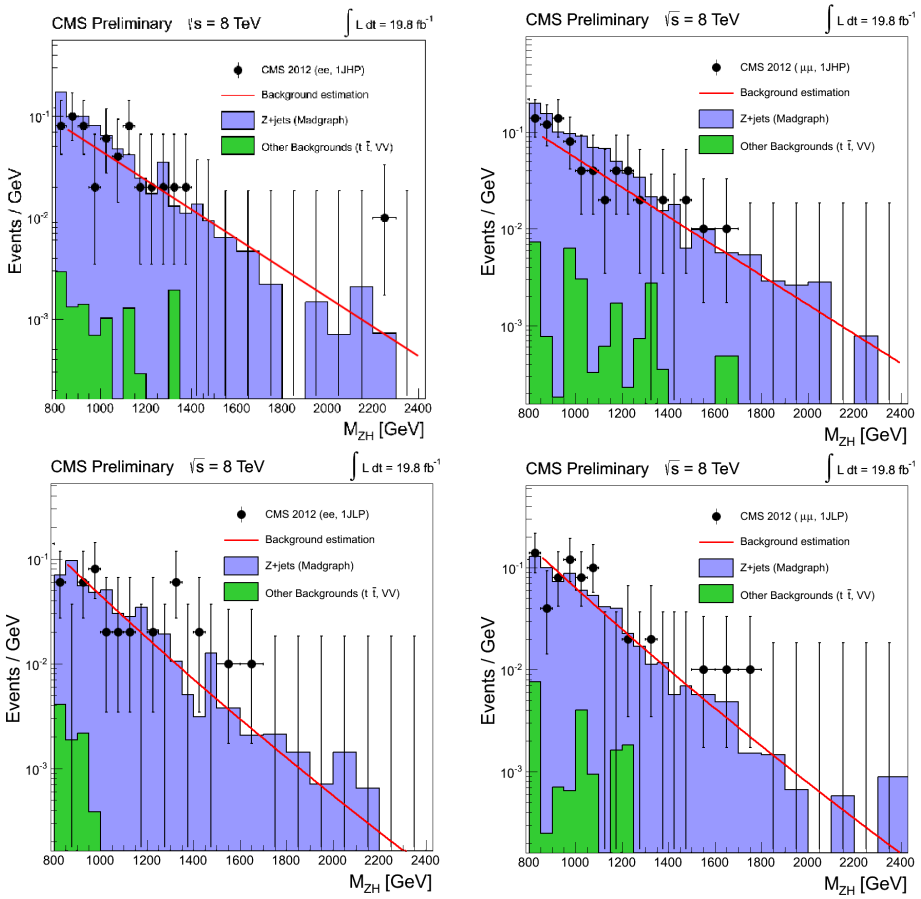

Figure 4.43. m_{ZH} invariant mass distribution after the final selection in the electron (left) and muon (right) channel and for the high purity (top) and low purity (bottom) category. Data with error bars are superimposed on the MC background simulation, the red line is the data-driven background estimation of section 4.8. The bin width is of 50 GeV for masses below 1500 GeV and of 100 GeV for higher masses. These plots have already been shown for the combined lepton channel in fig. 4.24.

No excess respect to the SM prediction is observed in the investigated range.

4.11.1 Extraction of the upper limit

As already discussed in section 4.7.4, this analysis is only able to set an upper limit on the studied process.

We tested 22 different mass hypotheses in the range $m_{ZH} \in [950, 2000]$ GeV at regular steps of 50 GeV. We can not extend the mass range lower than 950 GeV because we have to make sure that the whole signal, whose resolution is about 3-4 % (table 4.7), is included in the confident region of

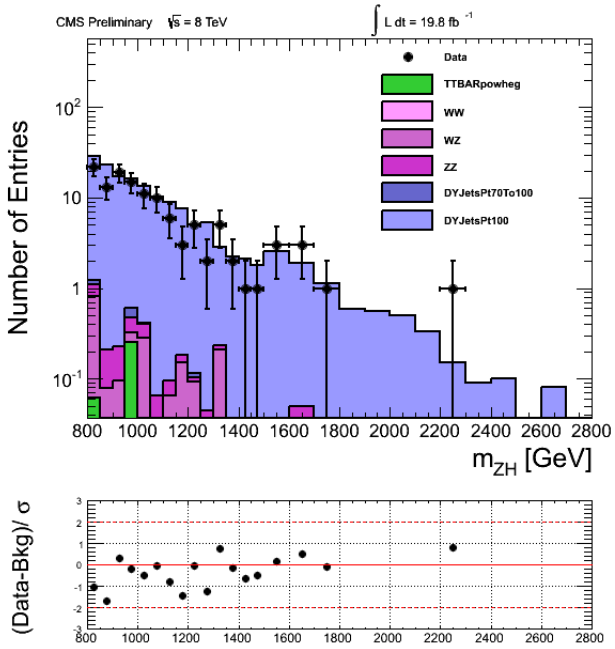

Figure 4.44. Invariant m_{ZH} mass distribution (top) after the final selection in the combined leptonic channel for data and MC and difference of the two divided by the measured error (bottom). The bin width is of 50 GeV for masses below 1500 GeV and of 100 GeV for higher masses.

the fit [850, 2400] GeV. For signal mass points where there is no generated MC sample, we estimated the signal shape and efficiency interpolating the corresponding quantities from the neighbouring mass points.

We follow the CL_S prescription described in [39] and [40] to set the final 95% exclusion limit on the cross section of the process $pp \to X \to ZH$[2] where the different categories have been combined as in [41]. The result is presented in figure 4.45. Figure 4.46 shows the same limit separately for the two categories, one can appreciate how the sensitivity of the high purity category is the dominant one while the low purity contributes only marginally.

[2]In the upper limit computation the brancing ratio of the Z and Higgs boson decays are not included in order to set a more general limit for searches of ZH resonances.

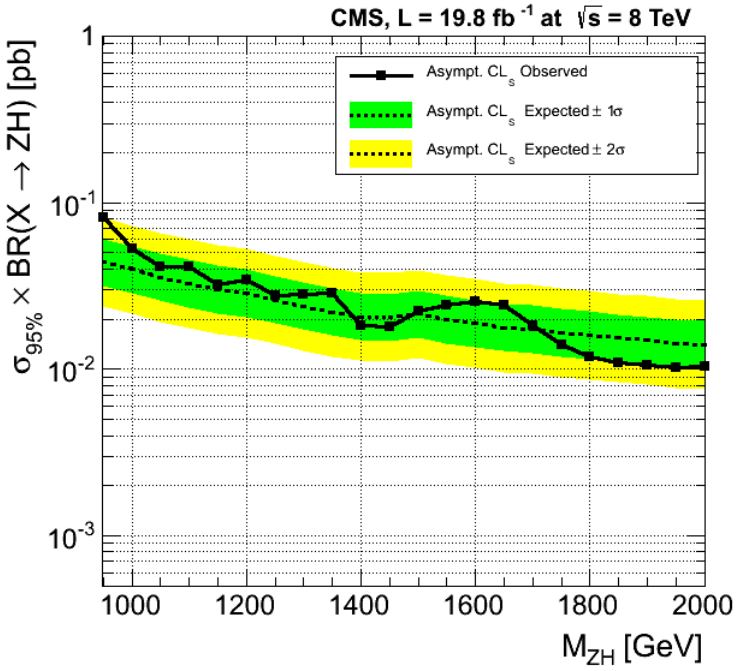

Figure 4.45. Observed and expected 95 % CL upper limit on the product of the new spin 1 resonance cross section and the branching fraction of $X \to ZH$ using 19.8 fb^{-1} of data. The 1 σ and 2 σ ranges of expectation are also shown with green and yellow bands.

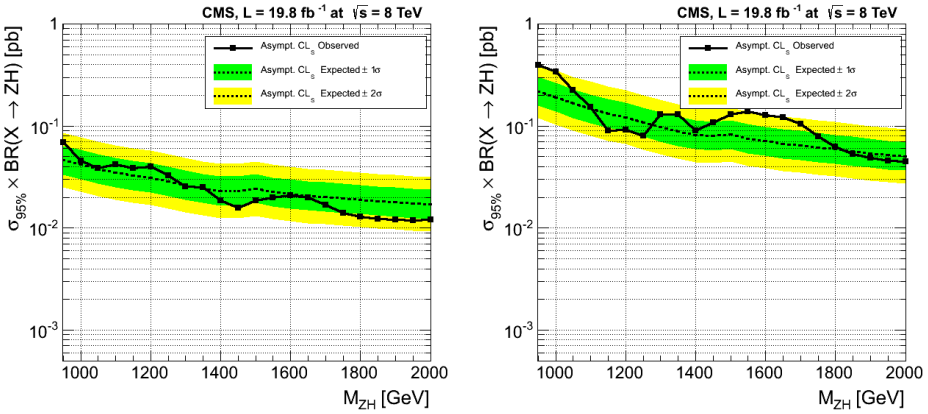

Figure 4.46. Observed and expected 95 % CL upper limit on the product of the new spin 1 resonance cross section and the branching fraction of $X \to ZH$ using 19.8 fb^{-1} of data. The 1 σ and 2 σ ranges of expectation are also shown with green and yellow bands. The left-hand panel presents the limit using only the high purity category, the right-hand panel uses only the low purity category.

Chapter 5

Conclusions

5.1 Results and conclusion

This analysis puts an upper limit at 95 % confidence level on the cross section of $pp \to ZH \to 2l2q$ at $\sqrt{s} = 8$ TeV. The analysis is performed with a categorization based on the study of a kinematic quantity, τ_{21}, sensitive to the substructure inside the jet. This categorization is implemented in order to enhanche the sensitivity of the search. A data driven technique is used to estimate the expected background in the signal region.

No excess above the expected SM background was found. The result is interpreted as an exclusion limit on the production cross section times the branching ratio in the ZH channel as a function of the resonance mass. Upper limits at 95 % confidence level are set in the range from 80 fb to 10 fb for resonance masses between 950 GeV and 2000 GeV, respectively.

Unfortunately the theoretical cross sections predicted by the composite Higgs model, taken as specific benchmark for the signal, are not available yet for a comparison with the experimental limit. We are waiting for the results of the theoretical group of R. Contino[1], that is currently computing quantitatively the production cross section for suitable values of the model parameters.

5.2 Performances and future improvement

In experimental physics, it is important to evaluate the performance of the experiment and how to improve a given measurement. These evaluations are crucial in order to plan future projects and, not last, to get funded by the scientific community.

In this analysis the result is mostly limited by the small amount of data, and only higher luminosity can improve this limitation. Moreover, the first source of systematics comes from the uncertainty on the H-tagging scale

[1]Roberto Contino, CERN & EPFL Lausanne

factor, which is computed studying the $t\bar{t}$ sample and it is again limited by the statistics. In conclusion, with higher luminosity we will be able to set lower limits and to reduce both the statistical error and the main systematic uncertainty. This will be discussed in section 5.2.3.

5.2.1 Other fully hadronic Higgs channels

In order to improve the results of this analysis it is also possible to follow a wider strategy and to consider other Higgs decay channels different from the $H \to q\bar{q}(gg)$ used in this analysis. Table 5.1 summarizes the Higgs boson branching ratios predicted by the standard model.

Channel	BR (%)
$H \to b\bar{b}$	57.7
$H \to WW\,(\to 4q)$	21.5 (9.82)
$H \to gg$	8.57
$H \to \tau\tau$	6.2
$H \to c\bar{c}$	2.9
$H \to ZZ\,(\to 4q)$	2.6 (1.27)
$H \to \gamma\gamma$	0.23
$H \to Z\gamma$	0.16
$H \to \mu\mu$	0.02

Table 5.1. Standard model Higgs branching ratios.

The present analysis can be extended including another fully hadronic channel, where the Higgs decays to a pair of vector bosons (WW or ZZ) and the vector bosons decay hadronically. Since the branching fractions of the W and Z to hadrons are large (67.6 % and 69.9 %), and the decay of the Higgs to WW and ZZ are also quite frequent (21.5 and 2.6 %), this channel can contribute effectively to the search.

Therefore, this ZH semileptonic search can easily include two additional channels that present the same final state, two leptons from the decay of the Z boson and the fully hadronic Higgs products that merge into a sigle jet.

However, there is an important difference between the basic approach and this extended strategy, due to the different substructure of the jet. In section 3.3.3 we have seen how the τ_{21} variable, inspired by the di-pole structure of the jet, plays a central role in the signal/background discrimination and in the categorization procedure. In the $H \to WW$ and $H \to ZZ$ cases we expect a quad-pole substructure of the jet and the analysis has to be modified to reflect this difference.

Figure 5.1 shows that the selection on τ_{21} is no longer incisive in the case of the $H \to VV$ channels: as expected, they have background-like shapes since the τ_{21} variable is built in order to distinguish a dipole jet

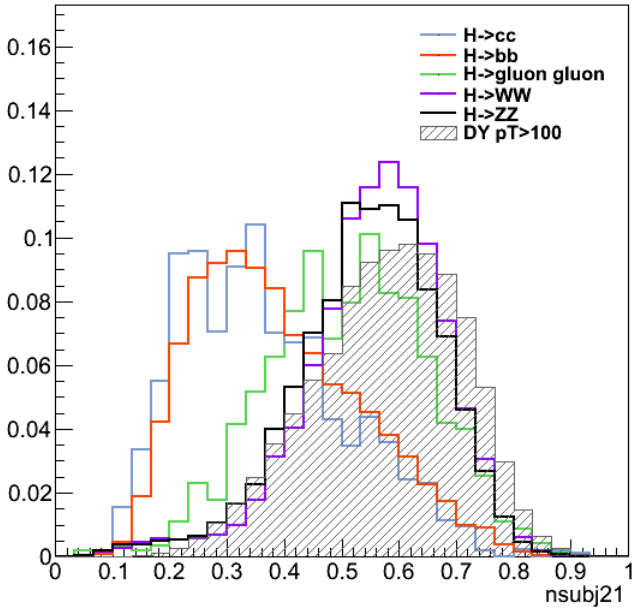

Figure 5.1. Normalized τ_{21} distribution of the five sub-samples with fully hadronic Higgs decay (M=1000 GeV) and of the background.

structure (eq. 3.11). In figure 5.2 one can see the relative contribution of each channel, with all the percentages reported in the legend. They sum up to unity and include all the Higgs fully hadronic channels. Their values can differ from those (rescaled) of table 5.1, because the plots report the status at the preselection level and the selection cuts can act differently on the various sub-samples.

The dominant $H \to b\bar{b}$ is reduced to two thirds of the total events and the new $H \to VV$ channels cover ~ 20 % of the potential signal. Since the total τ_{21} shape is very different (and much more background-like) from the distribution of the basic analysis (figure 3.14 and 3.15), the optimization study of section 4.7.2 has to be redone and new strategies for the categorization are needed.

The simplest idea is to look at other combinations of N-subjettiness involving τ_3, τ_4 and their possible ratios τ_{31}, τ_{32}, τ_{41}, ect. We expect variables containing τ_4 to be the best discriminant of the $H \to VV$ channels, even if distinguishing a quad-pole core in a jet is an harder task than the usual di-pole. This is one of the first cases where an analysis aims at looking so boldly inside a jet structure.

Unfortunately, the available MC samples don't provide the τ_4 variable of the jet, definitely needed to perform the analysis at its best. In this section

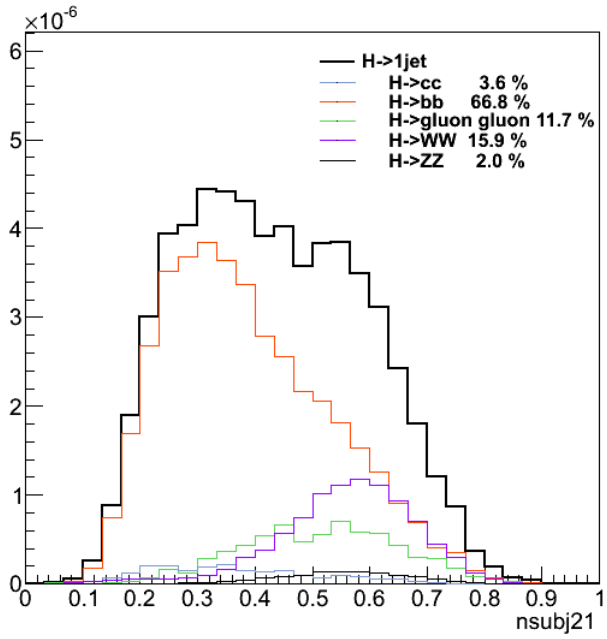

Figure 5.2. τ_{21} distribution of the five sub-samples with fully hadronic Higgs decay for a heavy resonance of 1000 GeV. Each contribute is plotted separately (colored thin line) and summed up with the relative weights in the total signal shape (thick black line).

we limit our effort to an overview of the possible strategies involving τ_3.

Figure 5.3 shows the distributions of the different subjettiness hypoteses and their possible ratios for the five fully hadronic Higgs decays sub-samples compared to the background. As discussed in section 3.3.3, it's their ratio that has the strongest discriminating power limited to the $H \to q\bar{q}$ channels, while it is $\tau_{32} = \tau_3/\tau_2$ that is found to be the best discriminant variable in the case of $H \to WW$ and $H \to ZZ$ decays. The background discrimination provided by τ_{32} is not as performant as τ_{21} in the basic analysis. One can note again that the $H \to gg$ sample presents extremely background-like shapes.

The existence of two exclusive discriminating variables suggests a 2D categorization in the plane (τ_{21}-τ_{32}). Figure 5.4 shows the correlation of these two variables for all the considered samples and background.

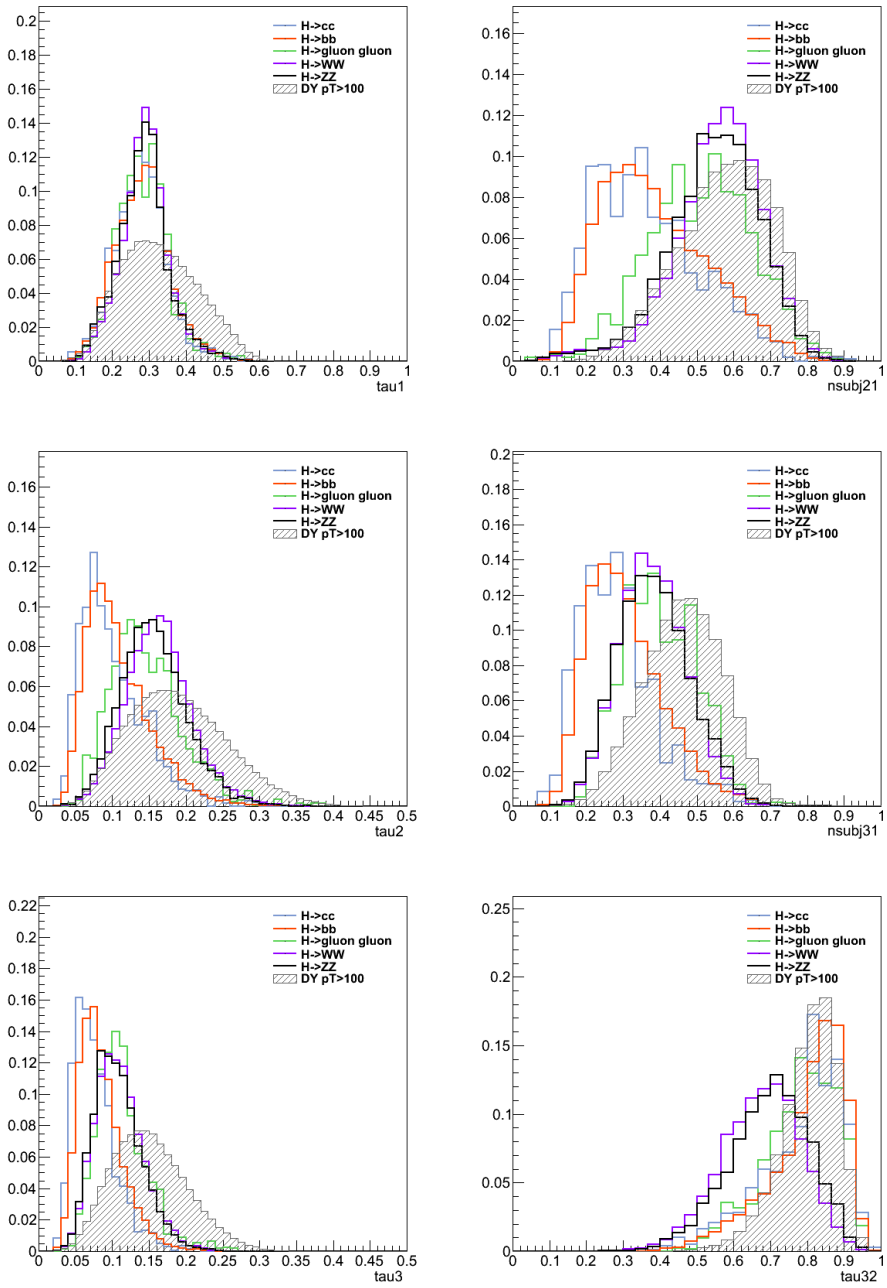

Figure 5.3. Normalized N-subjettiness distribution in the five sub-samples with fully hadronic Higgs decay for a heavy resonance of 1000 GeV and for background. From top to bottom, left: τ_1, τ_2, τ_3; right: τ_{21}, τ_{31}, τ_{32}.

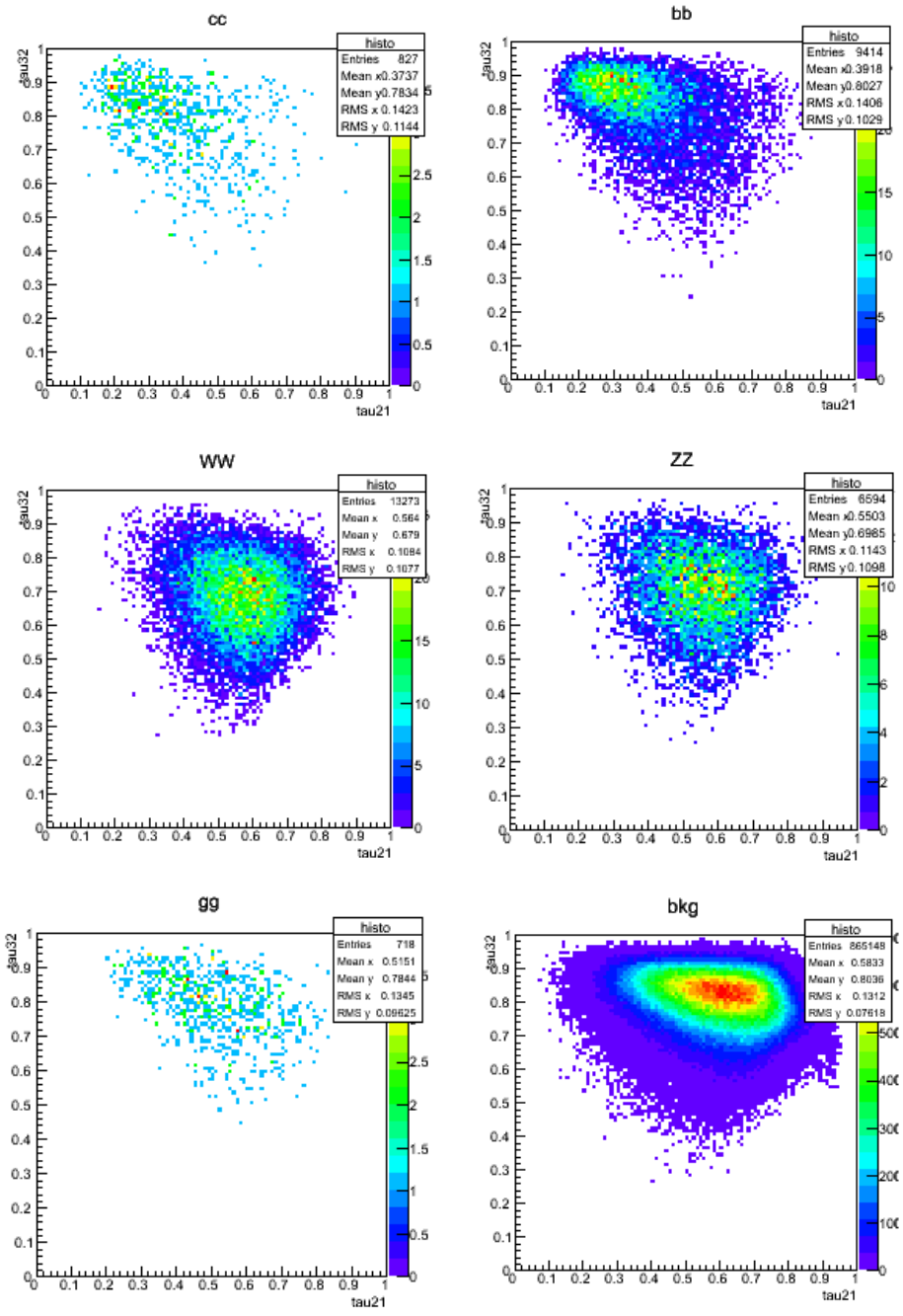

Figure 5.4. 2D correlation of the τ_{21} and τ_{32} variables for all the hadronic Higgs decay modes and for the background. Top left: $H \to cc$; top right: $H \to bb$; middle left: $H \to WW \to 4q$; middle right: $H \to ZZ \to 4q$; bottom left: $H \to gg$; bottom right: Z+jets background.

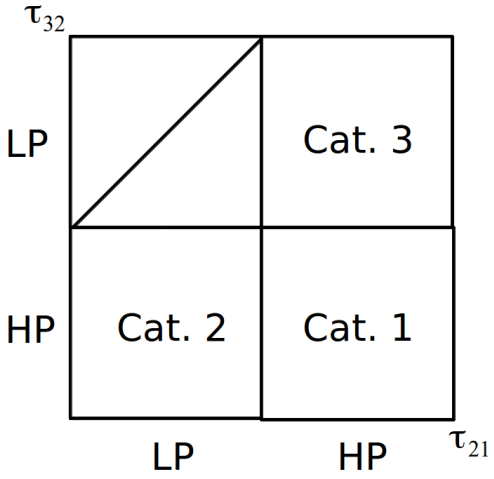

Figure 5.5. Possible 2D categorization in case of in analysis including $H \to VV$ fully hadronic samples.

The hypotetical 2D categorization would divide this plane in four quarters as sketched in figure 5.5.

5.2.2 Subjet b-tagging

Since the $b\bar{b}$ channel represents at least the two thirds of the total hadronic Higgs decays, the introduction of a b-tagging requirement applied on the two subjet can greatly improve the signal background discrimination. In [42] the application of the b-tagging algorithm was studied both on the fat jet and on the single jet substructures. Results are extremely comforting and they are reported in figure 5.6.

5.2.3 $\sqrt{s} = 14$ TeV upgrade at LHC

We want now to discuss the expectation of this analysis with the upcoming upgrade of the proton-proton energy collision at roughly 14 TeV. Since the heavy resonance is mainly produced through Drell-Yan scattering (see section 1.2.1), one can study the parton luminosities in order to define what is the gain in running at $\sqrt{s} \sim 14$ TeV. From figure 5.7 one can deduce a factor between 2.5 and 7 for invariant mass of 1 or 2 TeV [43]. This means that at $\sqrt{s} = 14$ TeV one can set the same 95% CL upper limit for the 2 TeV mass point with only about 3 fb^{-1}.

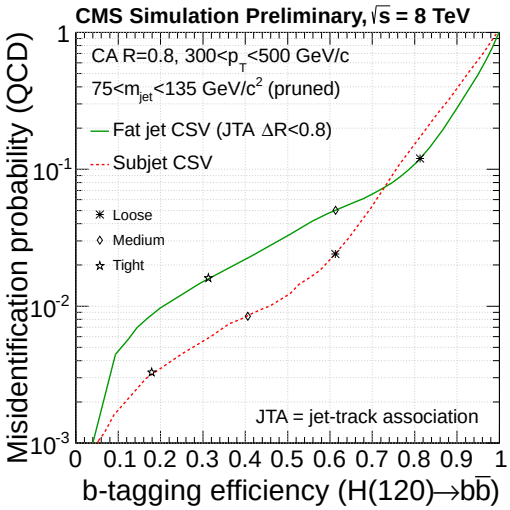

Figure 5.6. Background misidentification probability and b-tagging efficiency for a boosted $H \to b\bar{b}$. Two strategies are tested: (green line) the b-tagging algorithm is applied to the CA8 fat jet; (red line) at least one of the identified subjets is required to pass the b-tagging [42].

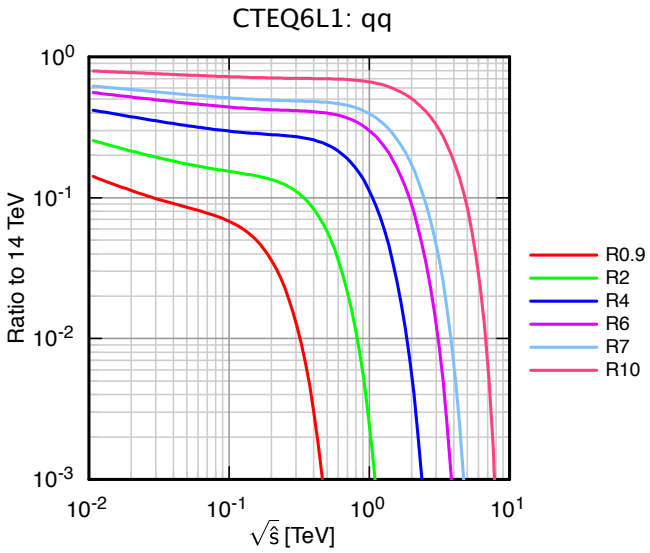

Figure 5.7. Ratio of parton luminosities for different pp collision energies compared to $\sqrt{s} = 14$ TeV [43].

5.3 Other searches

Simultaneously to this analysis in the ZH channel, the CMS EXO-VV group is performing other searches for the heavy spin 1 resonance X in other decay channels as W^+W^- and WH. These are still at a preliminary state and

haven't set any limit yet.

The strongest exclusion limit is currently set by the LHC search performed at 8 TeV with ∼20 fb^{-1} in the dilepton final state. This is possible because the resolution in the dileptonic final state in very good and the branching ratio of X decaying to a lepton pair is non negligible for moderate masses (see figure 1.3). The 95% CL upper limit is plotted in figure 5.8. It is expressed in term of the ratio

$$R_\sigma = \frac{\sigma(pp \to X) \times BR(X \to ll)}{\sigma(pp \to Z) \times BR(Z \to ll)}, \tag{5.1}$$

and $\sigma(pp \to Z) \times BR(Z \to ll)$ at $\sqrt{s} = 8$ TeV is measured to be 1.12 nb [44].

In order to compare such a result with our limit of figure 4.45, we have to evaluate the ratio of the two branching fractions $BR(X \to Zh)$ and $BR(X \to ll)$ of equation 1.28 and set an exclusion limit for the $\sigma(pp \to X)$ production.

One can find with qualitative substitutions,

$$\frac{BR(X \to ZH)}{BR(X \to ll)} = \frac{1}{192}\frac{3 \cdot 32}{2}\frac{\cot^2\theta_2}{\tan^2\theta_2} = \frac{1}{4}\left(\frac{g_{*2}}{g_2}\right)^4 = O(10^2), \tag{5.2}$$

where l include both the electron and muon channel and the mixing parameter $\tan\theta$ has been replaced with g_{el}/g_* as in equation 1.23. The strong composite coupling has been assumed moderately large, roughly $2 < g_{*2} < 4$.

In the high mass region (∼ 2 TeV), the performance of this analysis is comparable to the di-lepton search. In fact, evaluating qualitatively the corresponding limit for the di-lepton direct decay, one obtains

$$\sigma(pp \to X) \cdot BR(X \to Zh) = R_\sigma \cdot (1.12 \text{ nb }) \cdot \frac{BR(X \to Zh)}{BR(X \to ll)} = O(10^{-2}\text{pb}). \tag{5.3}$$

At lower masses (∼1 TeV), when the decay in pairs of SM bosons starts to decrease to advantage of the SM fermionic channels, the di-lepton limit is more stringent.

Other possible approaches to test the partial compositeness model consist of indirect measurements of the Higgs mass and of its coupling to the SM sector. This kind of analyses are particularly appropriate for linear lepton colliders, like the $\sqrt{s} = 500$ GeV ILC [45] and the $\sqrt{s}$ =1-5 TeV TLEP [46], thanks to their high precise measurements.

We can imagine a scenario where no clear indications of new particles emerged at the LHC, precise measurements of the single Higgs production can prove Higgs compositeness. For instance, a heavy particle with mass m_ρ and coupling to the Higgs g_ρ modifies the low-energy Higgs coupling by a relative amount of order $(g_\rho v/m_\rho)^2$. Consequently, in the absence of new

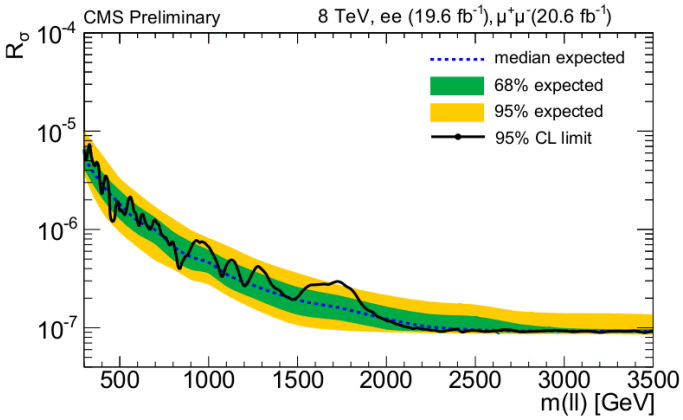

Figure 5.8. Upper limit on the ratio between the cross section times branching ratio for production of a resonance X decaying to lepton pairs and the Z cross section times branching ratio to lepton pairs. [49].

states below a certain scale M, an hypothetical observed deviation of order δ_h in single Higgs production would imply a lower bound on the coupling [47]

$$g_\rho > \sqrt{\delta_h}\frac{M}{v}. \tag{5.4}$$

The observation of the rare processes of double and triple Higgs production through vector boson fusion, $VV \to hh$ and $VV \to hhh$, would be even more sensible to SM deviations [47].

Finally, studies on the performances of direct and indirect searches [47] conclude that direct production of resonances even at a high-energy hadron collider, like the LHC with $\sqrt{s} = 33$ TeV [48], may not be competitive compared to precise measurements at linear colliders.

Glossary

Multiple scattering is a stochastic process where a charged particle is affected by many scattering interaction within a given material. The final bending angle is approximately gaussianly distributed with $\sqrt{\langle\theta^2\rangle} \propto \frac{1}{p}\sqrt{\frac{L}{X_0}}$ with respect to the initial direction of motion.

Radiation length X_0 is the mean distance (usually measured in $g\,cm^{-2}$) to reduce the energy of a high-energy electron by the factor $1/e$. It is also the appropriate scale length for describing high-energy electromagnetic cascades.

Moliere radius is a characteristic constant of a material giving the scale of the transverse dimension of the fully contained electromagnetic showers initiated by an incident high energy electron or photon. By definition, it is the radius of a cylinder containing on average 90% of the shower's energy deposition. It is related to the radiation length X_0 by the following approximate relation: $R_M = 0.0265X_0(Z + 1.2)$, where Z is the atomic number.

Interaction length λ_I is the mean path length required to reduce the numbers of relativistic hadronic particles by the factor 1/e, as they pass through matter. Differently from the radiation length, the undergoing processes that cause the energy loss are inelastic nuclear interactions.

Initial state radiation refers to a radiative emission that is not to due particle annihilation but it is emitted by the partons before the collision.

QCD (quantum cromodynamics) is a quantum field theory of the strong interaction (color force), a fundamental force describing the interactions between quarks and gluons which make up hadrons. QCD is a non abelian gauge theory of the SU(3) gauge group obtained by taking the color charge to define a local symmetry.

Fat jet A larger radius of 0.8 is chosen for the jet reconstruction algorithm, instead of the typical value of 0.5.

Tag and probe It is a data-driven technique used to compute the efficiency of a given selection based on pairs of resonant leptons: the first lepton (*tag*) is selected with a tight selection, the second one with extremely loose requirements so as to be biased by the studied selection as little as possible. The tag & probe efficiency is given by the fraction of the resonant events (collected with a selection on the inveriant mass of the system) that passed the given requirement.

Bibliography

[1] CMS collaboration. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Physics Letters B*, 716(1):30 – 61, 2012.

[2] ATLAS collaboration. Observation of a new particle in the search for the standard model higgs boson with the ATLAS detector at the LHC. *Physics Letters B*, 716(1):1 – 29, 2012.

[3] Precision Electroweak Measurements and Constraints on the Standard Model. *arXiv*, 1012.2367, 2010.

[4] R. Contino, T. Kramer, M. Son, and R. Sundrum. Warped/composite phenomenology simplified. *Journal of High Energy Physics*, 2007(05):074, 2007.

[5] I. Aitchison. *Supersimmetry in Particle Physics, An elementary introduction.* Cambridge University Press, 2007.

[6] D. Santopietro. The gauge hierarchy problem.
. http://lifshitz.ucdavis.edu/santopietro/hierarchyproblem.pdf.

[7] M. Drees. An Introduction to supersymmetry. *hep-ph/9611409*, 1996.

[8] L. Fitzpatrick, J. Kaplan, L. Randall, and L.T. Wang. Searching for the kaluza-klein graviton in bulk rs models. *Journal of High Energy Physics*, 2007(09):013, 2007.

[9] O. Antipin, D. Atwood, and A. Soni. Search for RS gravitons via W(L)W(L) decays. *Phys.Lett.*, B666:155–161, 2008.

[10] G. Azuelos, K. Benslama, D. Costanzo, G. Couture, J.E. Garcia, et al. Exploring little Higgs models with ATLAS at the LHC. *Eur.Phys.J.*, C39S2:13–24, 2005.

[11] J.E. Garcia, M. Lechowski, E. Ros, and D. Rousseau. Search for the decays $Z_H \to Zh$ and $W_H \to Wh$ in the Little Higgs model assuming m(h) = 120 GeV. ATL-PHYS-2004-001, 2003.

[12] J.E. Juknevich, D. Melnikov, and M.J. Strassler. A pure-glue hidden valley I: States and decays. *Journal of High Energy Physics*, 2009(07):055, 2009.

[13] A. Kaminska. Spin-1 resonances as a signature of composite higgs at the LHC, SUSY 2013, Trieste, 29.08.13.

[14] M. Pierini. Private conversation, 2013.

[15] S. Chatrchyan et al. Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV. *JINST*, 7:P10002, 2012.

[16] CMS Collaboration. Electron reconstruction and identification at $\sqrt{s} =$ 7 TeV. CMS-PAS-EGM-10-004, 2010.

[17] CMS Collaboration. Commissioning of the Particle-Flow reconstruction in Minimum-Bias and Jet Events from pp Collisions at 7 TeV. CMS-PAS-PFT-10-002, 2010.

[18] CMS Collaboration. Jet Energy Resolution in CMS at sqrt(s)=7 TeV. CMS-PAS-JME-10-014, 2011.

[19] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer, and T. Stelzer. MadGraph 5 : Going Beyond. *JHEP*, 1106:128, 2011.

[20] T. Sjostrand, S. Mrenna, and P. Skands. PYTHIA 6.4 Physics and Manual. *JHEP*, 0605:026, 2006.

[21] S. Frixione, P. Nason, and G. Ridolfi. The POWHEG-hvq manual version 1.0. 2007.

[22] J. Beringer et al. (Particle Data Group). The Review of Particle Physics, 2012.

[23] P. Tran et al. Measurement of jet mass in V+Jets events. CMS-AN-12-137, 2012.

[24] A. Bonato, R. Gerosa, et al. Identifying hadronically decaying W bosons merged into a single jet. CMS-AN-13-072, 2013.

[25] J. Thaler D. Krohn and L.T. Wang. Jet Trimming. *JHEP*, 100:242001, 2010.

[26] J. M. Butterworth et al. Jet substructure as a new Higgs search channel at the LHC. *PhysRevLett*, 100:242001, 2008.

[27] S. Ellis, C. Vermilion, and J. Walsh. Techniques for improved heavy particle searches with jet substructure. *Phys.Rev.*, D80:051501, 2009.

[28] S. Ellis, C. Vermilion, and J. Walsh. Recombination algorithms and jet substructure: Pruning as a tool for heavy particle searches. *Phys.Rev.*, D81:094023, 2010.

[29] S. Chatrchyan et al. Studies of jet mass in di-jet and W/Z + jet events. *JHEP*, 1305:090, 2013.

[30] The CMS collaboration. Jet mass in di-jet and vector boson plus jet events. *arXiv*, CMS PAS SMP-12-019, 2012.

[31] S. Chatrchyan et al. Performance of CMS muon reconstruction in *pp* collision events at $\sqrt{s} = 7$ TeV. *JINST*, 7:P10002, 2012.

[32] S. Ellis, C. Vermilion, and J. Walsh. Search for a BSM resonance decaying to Z vector bosons in the semileptonic final state. CMS AN-13-040, 2013.

[33] S. Harper and C. Shepherd-Themistocleous. Improving Sigma Eta Eta Showershape Studies: E^{2x5}/E^{5x5}, 2007.

[34] CMS Collaboration. Modified isolation cuts for the selection of high p_T electrons from the decay of boosted Z bosons. CMS-AN-2012/168, 2012.

[35] N. Adam, J. Berryhill, V. Halyo, A. Hunt, and K. Mishra. Tag and Probe Tutorial CMSSW_3_1_2.
. https://twiki.cern.ch/twiki/bin/view/CMS/TagAndProbe, 2009.

[36] G. Punzi. Sensitivity of searches for new signals and its optimization. *arXiv:physics*, 0308063v2:8–11, 2003.

[37] CMS Collaboration. Search for a BSM resonance decaying to W vector bosons in the semileptonic final state. CMS-AN-2013/045, 2013.

[38] CMS Collaboration. Search for Exotic resonances decaying to WW in the semi-leptonic channel. CMS-AN-2013/139, 2013.

[39] G. Cowan, K. Cranmer, E. Gross, and O. Vitells. Asymptotic formulae for likelihood-based tests of new physics. *Eur.Phys.J.*, C71:1554, 2011.

[40] A. Read. Presentation of search results: The CL(s) technique. *J.Phys.*, G28:2693–2704, 2002.

[41] S. Chatrchyan et al. Combined results of searches for the standard model Higgs boson in *pp* collisions at $\sqrt{s} = 7$ TeV. *Phys.Lett.*, B710:26–48, 2012.

[42] CMS Collaboration. Performance of b tagging at sqrt(s)=8 TeV in multijet, ttbar and boosted topology events. CMS-PAS-BTV-13-001, 2013.

[43] C. Quigg. LHC Physics Potential versus Energy. *arXiv*, 0908.3660, 2009.

[44] CMS Collaboration. Inclusive W/Z cross section at 8 TeV. CMS-PAS-SMP-12-011, 2012.

[45] H. Baer, T. Barklow, K. Fujii, Y. Gao, A. Hoang, et al. The International Linear Collider Technical Design Report - Volume 2: Physics. *arXiv*, 1306.6352, 2013.

[46] M. Bicer, H. Duran Yildiz, I. Yildiz, G. Coignet, M. Delmastro, et al. First Look at the Physics Case of TLEP. *arXiv*, 1308.6176, 2013.

[47] R. Contino, C. Grojean, D. Pappadopulo, R. Rattazzi, and A. Thamm. Strong Higgs Interactions at a Linear Collider. *arXiv*, 1309.7038, 2013.

[48] A. De Roeck, J. Ellis, C. Grojean, S. Heinemeyer, K. Jakobs, et al. From the LHC to Future Colliders. *Eur.Phys.J.*, C66:525–583, 2010.

[49] CMS Collaboration. Search for Resonances in the Dilepton Mass Distribution in pp Collisions at sqrt(s) = 8 TeV. CMS-PAS-EXO-12-061, 2013.