UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Fisica e Astronomia "Galileo Galilei"

Tesi di Laurea Magistrale in Fisica

Search for Anomalous Production of Higgs Boson Pairs with the CMS Detector

Relatore:
Dott. Tommaso Dorigo
Correlatore:
Dott.ssa Mia Tosi

Laureando: Carlo Alberto Gottardo
Matricola: 1080461

Anno Accademico 2014/2015

Abstract

This thesis consists of a study of the double Higgs production as the only means to probe the Higgs self-coupling λ. The final state chosen, featuring four b-jets, has a tiny cross section (9.96 fb at 8 TeV) and is overwhelmed by irreducible background. This study takes advantage of statistical and multivariate methods to characterize the signal and extract a limit on the signal strength on the data collected by the Compact Muon Solenoid (CMS) experiment at the LHC. Beyond Standard Model scenarios, accounting for anomalies in the the Higgs trilinear coupling λ and the Yukawa coupling with top quarks within the known constraints, are explored too. A limit has been extracted in all the scenarios considered, producing, in the Standard Model case, a result compatible with the one obtained by the ATLAS collaboration.

CONTENTS

CHAPTER 1

INTRODUCTION

The discovery of a 125 GeV scalar boson recognized as the Standard Model (SM) Higgs at the Large Hadron Collider (LHC) three years ago paved the way for a detailed study of the mentioned particle. The boson properties are currently being investigated by the scientific community both to confirm its supposed identity, and to verify new physics scenarios which may produce visible effects in the Higgs sector [1]. After the already determined mass, the most distinctive feature of the boson is the self-coupling λ [2] which can be directly measured only through the study of double Higgs production. According to the Standard Model the non-resonant di-Higgs production from proton-proton collisions has unfortunately a very low cross section of about 10 fb [3] at 8 TeV of center-of-mass energy (34 fb at 13 TeV [4]) which is well below the sensitivity of the current acquired data. However a limit on the cross section is sufficient to impose an experimental constraint on the λ coupling that translates into a test for several Beyond Standard Model (BSM) physics theories and for the SM itself.
To understand the motivation of this research the present work begins with a theoretical introduction about the Higgs couplings and its role in the SM, continues with a phenomenological section about BSM possibilities and then turns to the actual analysis performed on 8 TeV data. This analysis of the double Higgs production is performed on the particular decay channel in which each boson decays into a pair of b quarks, with a final state featuring four hadronic jets. Our final goal consists in the determination of an upper limit on the process cross section but also in tracing a path for the analysis on the upcoming data from LHC at 13 TeV.

1.1 THE STANDARD MODEL

1.1.1 The electroweak symmetry breakdown

In the next pages we will briefly go through the theory to show, without much justification, how the Higgs couples to itself and to other particles. To do so we will review the electroweak symmetry breakdown avoiding confusing heuristic explanations. A complete exposition can be found in [5].
Before the breakdown the electroweak theory is invariant under the action $SU(2)_L \times U(1)_Y$ gauge group, where the L reminds that the interaction is chiral and the Y that the group charge is the hypercharge (and not the electric one). The gauge field associated to $SU(2)_L$ is the Lorentz

vector W^μ, which can be written in terms of the algebra generator τ^a according to eq. (1.1).

$$W^\mu = W_a^\mu \tau^a = \frac{1}{2}\begin{pmatrix} W_3^\mu & W_1^\mu - iW_2^\mu \\ W_1^\mu + iW_2^\mu & -W_3^\mu \end{pmatrix} \tag{1.1}$$

The Lorentz-vector field associated to $U(1)_Y$ is called B^μ. Calling the field strengths $W^{\mu\nu}$ and $B^{\mu\nu}$, the kinetic Lagrangian is

$$\mathcal{L} = -\frac{1}{2g^2} Tr[W^{\mu\nu} W_{\mu\nu}] - \frac{1}{4g'^2} B_{\mu\nu} B^{\mu\nu} \tag{1.2}$$

Usually the fields W and B are redefined to gW and $g'B$ to absorb the couplings at the denominator. The electric charge, i.e. the generator of the electromagnetic U(1) symmetry, is defined by

$$Q = \tau^3 + Y \tag{1.3}$$

Making a global transformation of W and B fields under Q we discover that B and W_3 have no charge while $W_+ = \frac{W_1 - iW_2}{\sqrt{2}}$ has charge +1 and $W_- = \frac{W_1 + iW_2}{\sqrt{2}}$ has charge -1.
Even if the charge spectrum might sound familiar all the fields introduced are massless and mass terms in the Lagrangian are not gauge invariant. However we know that the weak interaction, being a short range one, must have massive mediators. We will now show how the introduction of the Higgs field grants the weak mediators mass. The Higgs field is a SU(2) doublet of complex scalar fields,

$$\Phi = \begin{pmatrix} \varphi_u \\ \varphi_d \end{pmatrix}. \tag{1.4}$$

Assigning to it hypercharge Y=1/2 the upper component φ_u has charge +1 while φ_d is neutral. The Higgs doublet is subjected to the most general $SU(2) \times U(1)$ invariant renormalizable potential which is

$$V[\Phi] = -\mu^2 \Phi^\dagger \Phi + \lambda(\Phi^\dagger \Phi) = \lambda \left(\Phi^\dagger \Phi - \frac{v^2}{2} \right)^2 \qquad v = \frac{\mu}{\sqrt{\lambda}} \tag{1.5}$$

The minima of the potential lie in the points

$$< \Phi >= \begin{pmatrix} 0 \\ v/\sqrt{2} \end{pmatrix} = \begin{pmatrix} 0 \\ \frac{\mu}{\sqrt{2}v^2}. \end{pmatrix} \tag{1.6}$$

up to an SU(2) transformation and electric charge redefinition. Since the field has to be quantized in the vicinity of a minimum, it has to be redefined in order to have $< \Phi >= 0$:

$$\Phi(x) = e^{i\Pi_a(x)\tau^a} \begin{pmatrix} 0 \\ \frac{v+h(x)}{\sqrt{2}} \end{pmatrix} \tag{1.7}$$

We can get rid of the first exponential by making an opposite gauge transformation. The Φ gauge invariant Lagrangian is given by:

$$\mathcal{L}_\mathcal{H} = (D_\mu \Phi)^\dagger (D^\mu \Phi) - V[\Phi] \tag{1.8}$$

where $D_\mu\Phi$ is the covariant derivative applied to the field, explicitly

$$D_\mu\Phi = \partial_\mu\Phi - igW_\mu\Phi - ig'Y(\Phi)B_\mu\Phi \tag{1.9}$$

$$= \begin{pmatrix} 0 \\ \frac{1}{\sqrt{2}}\partial_\mu h \end{pmatrix} - \frac{i}{\sqrt{2}}(v+h)\begin{pmatrix} \sqrt{2}gW_\mu^+ \\ g'B_\mu - gW_\mu^3 \end{pmatrix}. \tag{1.10}$$

The kinetic term of the Higgs Lagrangian becomes

$$\mathcal{L}_h^{kin} = \frac{1}{2}\partial_\mu h\partial^\mu h + \frac{1}{2}(v+h)^2(2g^2|W_\mu^+|^2 + (g'B_\mu - gW_\mu^3)^2). \tag{1.11}$$

Expanding the square $(v+h)^2$ the following mass terms come up:

$$g^2v^2|W^+|^2 \qquad \rightarrow m_W = (1/2)gv \tag{1.12}$$

$$\frac{1}{2}v^2(g'B_\mu - gW_\mu^3)^2 \qquad \rightarrow \text{mixed mass term} \tag{1.13}$$

The last term hands mass to a combination of B and W^3 while we expect a massive Z boson and a massless photon. In other words B and W^3 are not mass eigenstates. So we change base defining two new fields A^μ and Z^μ that will correspond to the electromagnetic vector potential and the Z boson field, respectively:

$$\begin{pmatrix} Z_\mu \\ A_\mu \end{pmatrix} = \begin{pmatrix} cos\theta_W & -sin\theta_W \\ sin\theta_W & cos\theta_W \end{pmatrix}\begin{pmatrix} W_\mu^3 \\ B_\mu \end{pmatrix} \tag{1.14}$$

where θ_W is called Weinberg angle. Plugging the new fields in the mixed mass term we obtain a massless A and a Z with $m_Z = \frac{m_W}{cos\theta_W}$.
The Higgs mass can be read from the potential expansion

$$V[\Phi] = \lambda\left(\frac{(v+h)^2}{2} + \frac{v^2}{2}\right) = \lambda v^2h^2 + \lambda vh^3 + \frac{\lambda}{4}h^4 \tag{1.15}$$

and yields $m_h = \sqrt{2\lambda}v$.
The presented theory has so far a spectrum made of four particles, a massive scalar boson (the Higgs), two oppositely charged weak bosons W of equal mass and a neutral weak boson, the Z, with mass related to the W. Any discussion above the Weinberg angle will be neglected for the sake of brevity.

1.1.2 The quark masses

In the SM the quarks are the constituents of hadrons such as the proton or the neutron. Quarks have both electric and color charge where the latter is the one associated to the Strong Interaction. They are divided into three families which are up-down, charm-strange, top-bottom. Mathematically a single quark is a fermion described by a Dirac spinor. From now on we will neglect anything regarding QCD such as asymptotic freedom and quark confinement in order to concentrate on the relation between quarks and the Higgs boson.
The first quark family is made up by a left-handed $SU(2)_L$ doublet and two right-handed $SU(2)_L$ singlets as follows:

$$Q_L = \begin{pmatrix} u_L \\ d_L \end{pmatrix} \in 2_{Y=1/6}SU(2), \qquad u_R \in 1_{Y=2/3}SU(2), \qquad d_R \in 1_{Y=-1/3}SU(2) \tag{1.16}$$

where we have written the hypercharge as a subscript to the SU(2) representation dimension. We remind that a SU(2) representation has dimension 2j+1 where j is the Casimir.
Now the easiest way one could think to confer mass to the quarks could be a term like

$$\bar{u}u = \bar{u}_L u_R + \bar{u_R} u_L \tag{1.17}$$

but such a term would not be $SU(2) \times U(1)$ invariant as it contains objects transforming to different representations. Actually the quarks acquire a mass by interacting with the Higgs field. For the sake of brevity we will now assume the existence of just one quark family The most general gauge invariant Higgs-quark Lagrangian term is given by a Yukawa interaction

$$\mathcal{L}_{Yukawa} = -Y_u \bar{Q}_L \Phi^c u_R - Y_d \bar{Q}_L \Phi d_R + h.c. \tag{1.18}$$

where Φ^c is the charge conjugated Higgs field $\Phi^c_\alpha = i(\sigma_2)_{\alpha\beta}\Phi^{*\beta}$ which is still a doublet and in practice, after the electroweak symmetry breakdown, it equals Φ but has the components swapped.
After the symmetry breaking the same Lagrangian will yield an interaction and a mass term

$$\mathcal{L} = -\frac{v+h}{\sqrt{2}} \left(\bar{u}_L Y_u u_R + \bar{d}_L Y_d d_R \right) \tag{1.19}$$

The quark mass is $m_{u/d} = Y_{u/d} v/\sqrt{2}$ while the interaction coupling constant is given by $-iY_{u/d}/\sqrt{2}$, which, plugging the expression for the quark mass becomes $-im_{u/d}/\sqrt{2}$, hence the saying "The Higgs couples with quarks proportionally to their masses". The extension to three families is mathematically trivial but has important physics outcomes as it leads to mismatching between mass eigenstates and interaction ones. Processes leading to change of quark flavor will be allowed by charged weak interaction weighted by the Cabibbo, Kobayashi, Maskawa matrix elements.

1.1.3 Lepton masses

For the sake of completeness we show also how charged leptons become massive leaving neutrino massless. A right-handed neutrino would be neutral under every fundamental interaction (except gravity) and the prediction of a unobservable particle has discomforted even philosophers. However, leptons come in three leptonic flavour families: electronic, muonic, and tauonic but here we assume the existence of just one. Each family is made of a SU(2) doublet and one singlet. In the doublet the upper component is the neutrino while the lower is the charged lepton:

$$L_L = \begin{pmatrix} \nu_L \\ \ell_L \end{pmatrix} \in 2_{Y=-1/2} \qquad \ell_R \in 1_{Y=-1}. \tag{1.20}$$

Leptons acquire mass form a Yukawa interaction with the Higgs field

$$\mathcal{L} = -Y_L \bar{L} \Phi \ell_R + h.c.. \tag{1.21}$$

After the symmetry breaking the charged lepton acquires a mass equal to $m_\ell = \frac{v}{\sqrt{2}} Y_L$.
In this case the introduction of the other two families poses no problem: the Y_L becomes a 3×3 matrix that can be diagonalized redefining the fields. Each charged lepton gets a different mass. The inclusion of a right-handed neutrino will allow the neutrino to be massive but also causes a mismatch between mass and interaction neutrino eigenstates allowing neutrino flavor oscillations.

1.1.4 Higgs boson couplings

Focusing on the Higgs Lagrangian we saw that its coupling are described by three parameters: μ, λ and the vacuum expectation value (v.e.v) v.
The boson mass is related only to μ according to $m_h = \sqrt{2}\mu$ so, after the 2012 Higgs mass determination μ is known.
The v.e.v., that parametrizes the W and Z masses as well as the Fermi constant G_F, is related to both λ and μ according to eq. (1.5) and it has been indirectly estimated at $v \sim 246GeV$. As a result λ is expected to be around 0.12 according to the SM, however no direct measurement has been performed yet. Since λ appears only in the trilinear and quadriliear Higgs vertices, we need to study the double Higgs production to probe it.

1.2 DOUBLE HIGGS PRODUCTION

At an hadronic collider double Higgs production can be realized, according to the Standard Model, through gluon-gluon fusion (ggF). The process, depicted in fig. 1.1, is the main production channel with a cross section $\sigma = 9.96fb \pm 9\%(scale) \pm 2\%(pdf)$ at $\sqrt{s} = 8TeV$. As anticipated the cross section is very low, especially considering the integrated luminosity of $20fb^{-1}$ collected by the CMS experiment.

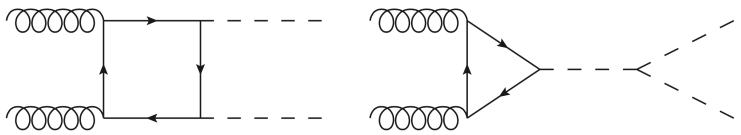


Figure 1.1: Feynman (interfering) diagrams of ggF Higgs pair production. The quarks in the fermionic loops are assumed to be top due to their much stronger coupling with respect to other flavors.

Given the already low cross-section we will not examine the other production channels[(1)] concentrating, instead, on the possible final states.
Each of the bosons can decay, in decreasing branching ratio [6], into $b\bar{b}$ ($\mathcal{B}r = 60\%$), WW ($\mathcal{B}r = 23\%$), gg ($\mathcal{B}r = 8\%$), $\tau\tau$ ($\mathcal{B}r = 7\%$), ZZ ($\mathcal{B}r = 3\%$), $\gamma\gamma$ ($\mathcal{B}r = 0.1\%$).
Each channel has its features, for example the $\gamma\gamma$ allowed the discovery of the Higgs in 2012 thanks to its clear signature despite of the low branching ratio. The $b\bar{b}$ channel benefits from the large branching ratio but also suffers from a large QCD background. Nevertheless the advantage of a large branching fraction makes it competitive with the other more distinctive final states. Other suitable choice would be $\gamma\gamma b\bar{b}$, which takes advantage from the two photon with 125 GeV invariant mass and $WWb\bar{b}$.

1.3 HIGGS BEYOND THE STANDARD MODEL

In the introduction we referred to SM extensions regarding our process. A simple Effective Theory involving the Higgs pairs production by gluon-gluon fusion that assumes negligible

(1) these, for the records, are Higgstrahlung, Vector Boson Fusion (VBF), top associated production (ttHH)

Higgs couplings to light fermions and the absence of any other light state in addition to the SM particles is given by eq. (1.22) [7].

$$\mathcal{L}_h = \frac{1}{2}\partial_\mu h \partial^\mu h - \frac{1}{2}m_h^2 h^2 - \kappa_\lambda \lambda_{SM} v h^3 - \frac{m_t}{v}(v + \kappa_t h + \frac{c_2}{v} h h)(\bar{t_L} t_R + h.c.)$$
$$+ \frac{1}{4}\frac{\alpha_s}{3\pi v}(c_g h - \frac{c_{2g}}{2v} h h) G^{\mu\nu} G_{\mu\nu}$$

where κ_λ and κ_t are multiplicative deviation factors regarding the Higgs trilinear coupling and the Yukawa interaction with the top quark respectively, while the c_2, c_g and c_{2g}, not present in the SM, account for Higgs contact interactions with gluons and top quarks.
To give a picture of our current knowledge let us be aware that the c_g and κ_t parameters are already constrained by single Higgs measurement [8][9], while the remaining ones are experimentally completely unconstrained. To be precise [10] k_t has to lay between 0.5 and 2.5 at 95% C.L., c_g can be at least of order $O(10^{-1})$ [11] [12] [13]. The other ranges are found one by one considering one parameter free, the others fixed and comparing the corresponding cross section to the known limits.

It has been calculated that some combination of the parameter values (within the designated ranges) causes an enhancement of the cross section which is only limited by the experimental result [14] [15] by ATLAS, $\sigma^{SM}_{hh\to 4b} < 202 fb$, which is 62.4 times greater than the SM value.
However, due to the interference among the contributing diagrams (fig. 1.2), different combinations of the parameter values besides the cross section also drastically change the kinematics of the process, making the known experimental limit no longer applicable and requiring different optimized analyses. For this purpose a clustering technique has been developed [7] to reduce the large set of possibilities to a handful of representative ones on which analyses can be more effectively performed .

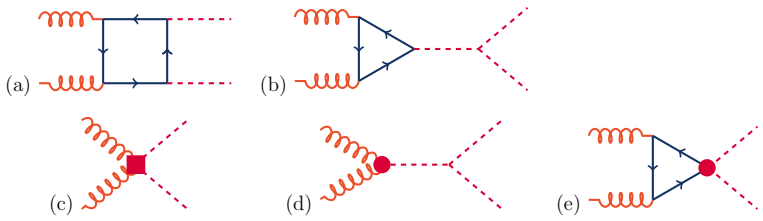


Figure 1.2: ggF Higgs pairs production Feynman (interfering) diagrams. The quarks into the fermionic loops are assumed to be top thanks to stronger coupling with respect to other flavors.

CHAPTER 2

THE COMPACT MUON SOLENOID

2.1 THE DETECTOR

The Compact Muon Solenoid (CMS) is a 21.6 meters long and 15 meters wide multi-purpose hermetic detector placed in the Large Hadron Collider (LHC) facility at CERN. The LHC has been designed to collide bunches of protons at a center-of-mass energy up to $\sqrt{s} = 14TeV$ and heavy ions up to $\sqrt{s} = 5.5TeV$ per nucleon pair. The bunches traveling clockwise and counterclockwise in two different pipes cross each other inside the detector every 50 ns, yielding about 20 inelastic collisions per crossing. To take advantage of the collider capabilities CMS had been equipped with a high-resolution silicon tracker to guarantee a high track reconstruction efficiency despite the pile-up,a high granularity electromagnetic calorimeter and a muon detection system. The exceptional feature of CMS is the superconducting solenoid that encloses both the tracker and the calorimeters and provides an homogeneous and strongly bending 3.8 T magnetic field. The muon detectors, which provide a tracker-independent measurement of the muons momentum, are instead embedded into the iron return yoke used both as absorber and a mean to confine the magnetic field. A brief description of each subdetector will follow, with particular attention to the tracking system which plays the leading role in b-tagging (i.e. the ability of distinguish between jets originating from bottom quarks among the other possible flavors). A detailed description can be found in [16].

2.1.1 CMS Coordinate System

A Cartesian coordinate system is defined in the CMS detector as follows. The x-axis points to the center of the LHC ring, the y-axis vertically upwards and the z-axis is directed along the beam towards the Jura mountains; the origin is located in the interaction point. The azimuthal angle ϕ is measured in the xy plane from the x axis and the radial coordinate in this plane is denoted as r. The polar angle θ is defined in the rz plane but usually is expressed in terms of the pseudorapidity $\eta = ln(tan(\theta/2))$. The pseudorapidity is 0 for a particle moving perpendicular to the beam direction while approaches $\pm\infty$ for a particle moving parallel (anti-parallel) to the z-axis. The pseudo-rapidity for a massless particle is equal to the rapidity $y = \frac{1}{2} ln\left(\frac{E+p_z}{E-p_z}\right)$ which is invariant under boost along the z. The momentum component transverse to the beam direction, denoted p_T, is computed from the x-and y-components, and similarly the transverse energy is defined as $E_T = Esin\theta$. Both $\Delta\eta$ and $\Delta\phi$ between two particles are independent of Lorentz boosts, therefore the distance between two particles can be measured with a third Lorentz-invariant variable, called ΔR and defined as $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$.

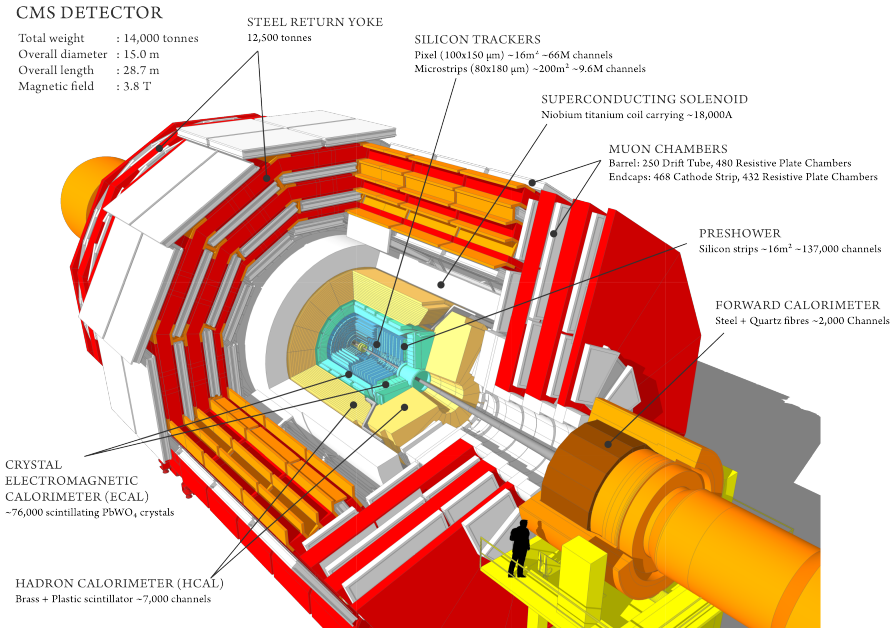


Figure 2.1: Overview of the detector

2.1.2 The tracker

The tracker relies completely on radiation-hard silicon hybrid pixels and strips. The pixel modules are arranged in three cylindrical layers at radii 4.4, 7.3 and 10.2 cm from the interaction point plus two disks at each end. The strips are employed in the rest of the tracker subsystems, namely the 4 layers of Inner Barrel (TIB) and the three Disks (TID) per side, the six layers of the Outer Barrel (TOB) and finally the nine layers per side of End Caps (TEC). The tracker altogether covers angles up to $|\eta| = 2.5$ with a resolution of about $15\mu m$ both in $r - \Phi$ and z which allows an efficient jet b-tagging. Considering only the barrel, the strips of the TIB provide a resolution on the $r - \phi$ plane ranging from 23 to 35 μm while the ones in the TOB from 35 to 53 μm.

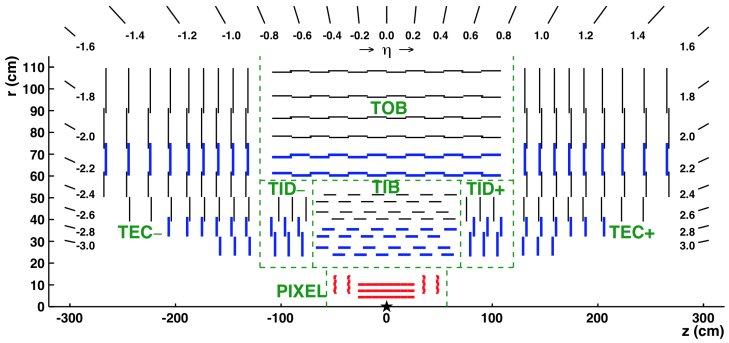


Figure 2.2: Tracker system schematic.

2.1.3 The ECAL

The Electromagnetic calorimeter is designed to contain the electromagnetic showers caused mainly by photons and electrons and is placed just ouside the tracker at r = 1.29m. It covers a region up to $|\eta| < 2.5$ and is made up of 74848 lead-tungstate ($PbWO_4$) crystals grouped together into 5×5 towers for triggering purposes. This dense material ($8.3 g/cm^3$) has been chosen because it has a short radiation length X_0 of 0.89 cm, a small Moliére radius R_M of 2.2 cm, a fast response (80% of light emitted within 25 ns) and clearly it is transparent to its relaxation light (440nm). In the history of scintillators, lead-tungstate has emerged only in the last 20 years due to its low light yield of about of 100 photons/MeV [17] at $18 \circ C$ (the nominal operating temperature of the ECAL [18]). The high shower containment capability of these crystals made possible to include the ECAL inside the magnetic coil. The light coming from each crystal is then converted into an electrical impulse by an avalanche photodiode (APD) if in the barrel, or a vacuum phototriode (VPT) if in the endcap. A pre-shower device made of two disks of lead absorber at $2X_0$ and $3X_0$, and of two planes of silicon strip detectors completes the ECAL in front of the endcaps. It allows the rejection of photon pairs from π_0 decays and improves the estimation of the photon direction, enhancing the two-photon invariant mass resolution.
The ECAL energy resolution can be parametrized by three different contributions:

$$\frac{\sigma_E}{E} = \frac{a}{\sqrt{E}} \oplus \frac{b}{E} \oplus c \tag{2.1}$$

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

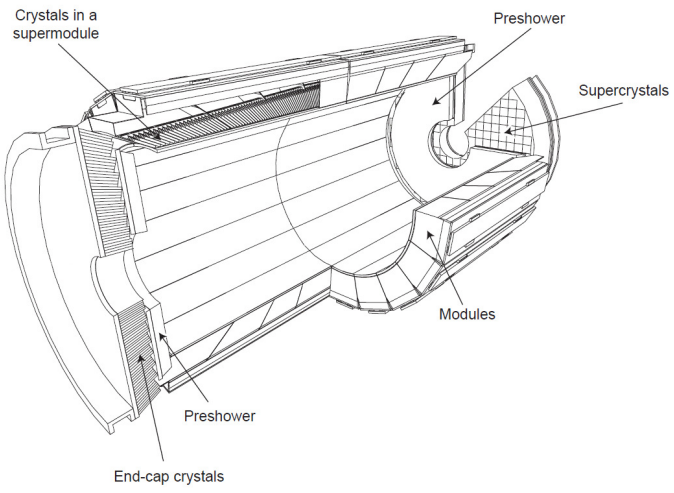


Figure 2.3: Three-dimensiona CAD of ECAL.

2.1.4 The HCAL

The hadronic calorimeter [18] is responsible for the measurement of hadron jets energy and, being hermetic, also of the reconstruction of the apparent missing energy due to neutrinos. In order to keep it between the outer extent of the ECAL and the inner extent of the magnetic

coil but assure the containment of the hadronic shower of jets with energy in the multi-TeV range, the plastic scintillators are separated by brass absorbing plates. The thickness of the absorber layers is between 60 mm in the barrel and 80 mm in the end-caps making the HCAL barrel extend to 5.46 interaction lengths at $\eta \approx 0$ to 10.8 at $\eta \approx 1.3$. The light produced by the active medium is brought by a wavelength shifter fibers to hybrid pixelated photodiodes, able to operate in a high magnetic field environment. To cover the biggest possible portion of the solid angle the barrel and endcaps are complemented by a hadron forward calorimeter, which is placed outside the magnet return yokes, at 11 m from the interaction point, with a total coverage of $3 < |\eta| < 5.3$. Moreover, an outer "tail catcher" hadronic calorimeter is placed in the first muon absorber layer in order to enhance the containment of high-energy jets in the central region of the detector.

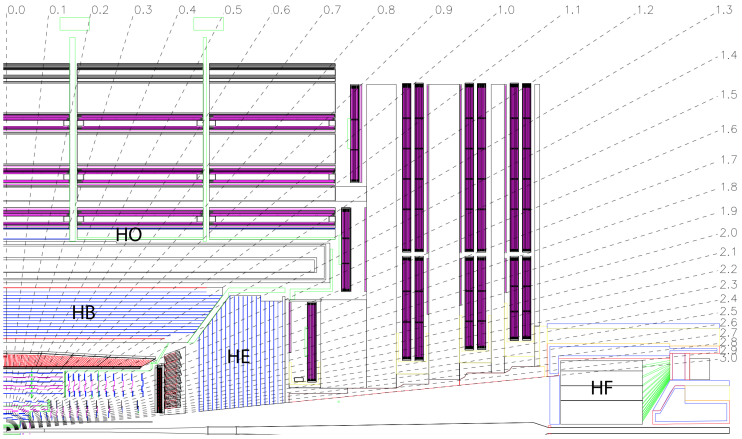


Figure 2.4: HCAL schematic.

2.1.5 The muon chambers

Since a large number of physics process present a clear muon signature, special care has been taken in the design and construction of the muon system. The muon [19] detectors are hosted in the return yoke outside the solenoid. The yoke, necessary to contain the magnetic field, provides a magnetic field pointing in the opposite direction with respect to the one in the solenoid. This allows a muon momentum measurement indipendent from the tracker one. There are three types of subdetectors serving different regions.

Drift Tubes (DT): in the barrel ($|\eta| < 1.2$) four layers of Drift Tube chambers are placed in the barrel region arranged in 5 wheels along the z-axis, each one divided into 12 azimuthal sectors. Each DT chamber, on average $2 \times 2.5m$ in size, consists of 12 aluminum layers, divided in three groups of four, each with up to 60 tubes: the middle group measures the coordinate along the direction parallel to the beam and the two outside groups measure the perpendicular coordinate. The muon position in each DT is reconstructed by measuring the drift time of the ionization electrons, and converting it into a distance from the wire. Each one of the 250 DT chambers has a resolution of $\approx 100\mu m$ in $r - \phi$ and 1 mrad in ϕ.
Cathode Strip Chambers (CSC): in the two endcaps ($0.8 < |\eta| < 2.4$), where the flux of hadron punch-through and radiation is much higher and the magnetic field is strongly varying, 540 Cathode Strip Chambers 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 in r, and 75 to 150 μm in the $r - \phi$ coordinate.
Resistive Plate Chambers (RPC): In both the barrel and the endcaps, a system of 912 Resistive Plate Chambers is installed, ensuring redundancy to the measurement. RPCs provide a rougher spatial resolution than DTs and CSCs, but the fast response with a good time resolution (1 ns) is used for triggering purposes.

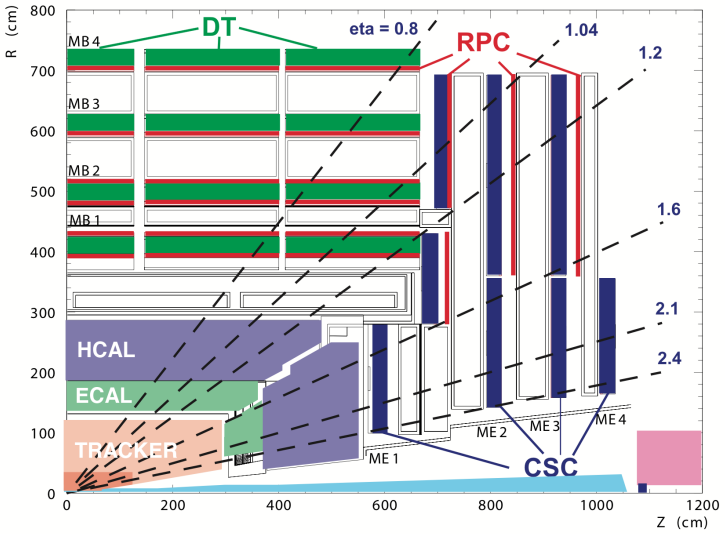


Figure 2.5: Muon system schematic.

2.2 TRIGGER

Even neglecting the pile-up, the pp collision rate delivered by LHC is of $40MHz$. Since an event requires several hundreds of kilobytes the limited bandwidth of electronics represents a bottleneck. To overcome it an online tigger system selects events reducing substantially the storage rate but maintaining a high efficiency on the potentially interesting events [20]. Different selection criteria can be used by the trigger. Each criterion, or better, logic is identified by the trigger path which is a string containing the name of the object, some thresholds, and how many objects are required.
The CMS trigger is implemented at two levels.
Level 1 trigger (L1): it lowers the rate from $40MHz$ to $\approx 100kHz$ in less than $3\mu s$. The L1 trigger, powered by custom programmable processors, exploits only calorimetric measurements and muon system information to take a decision. In detail the trigger response is based on the so-called "trigger primitive", that is the presence and the number of objects like electrons, photons, muons, jets and t-jets, and missing energy with a transverse energy or momentum above a given threshold.
High level trigger (HLT): relying on 7552 Intel Xeon cores (as of May 2012), the HLT reduces the event rate from $\approx 100kHz$ to about 300 Hz before data storage. The first step, called L2, consists in the identification and measurements of particle candidates and global variables using only the information coming from calorimeters and muon system. In the next step, denoted as

L2.5, the tracker data is used to reconstruct tracks and primary vertices. The last step, known as L3, runs the same algorithms employed by the offline reconstruction. Since at trigger level computing time is more critical than reconstruction accuracy, the algorithms are modified in order to be faster, even with a slightly lower precision. In order to meet the timing requirements given by the L1 input rate, events can be discarded before being fully reconstructed, as soon as the available information is enough to take the decision, or reconstruction can be limited only to a restricted region of the detector, identified by the L1 trigger object.

2.3 Particle Identification

In order to identify the particles produced in a pp collision, by means of the tracks and energy deposits recorded, an algorithm that combines all the information coming from the sub-detectors is needed. In the CMS experiment the main algorithm developed and employed to accomplish this task is called Particle Flow (PF) [21] [22]. In a nutshell the algorithm works as follows: the tracks reconstructed from the tracker are extrapolated through the calorimeters, if they pass close to one or several energy deposits, also called clusters, the deposits are associated to the track. To this set of track and cluster(s) is associated a charged hadron and is not considered anymore by the algorithm. The muons are identified beforehand so that their track does not give rise to a charged hadron candidate.
The reconstruction of electrons needs particular care as their bremsstrahlung photons generate several clusters in the ECAL, each of which has to be associated to the electron avoiding a double counting.
Once all the tracks have been matched, the remaining clusters are associated to photons if in the ECAL or to neutral hadrons if in the HCAL. Then the nature of the particles can be assessed, and the information of the sub-detectors combined to determine optimally their four-momentum.
The resulting list of particles, namely charged hadrons, photons, neutral hadrons, electrons and muons, is then used to reconstruct the jets, the missing transverse energy, to reconstruct and identify the taus from their decays products and to measure the isolation of the particles.

2.4 Jets

2.4.1 Phenomenology

When two protons collide at $\sqrt{s} = 8TeV$ deep inelastic scattering may take place. In the deep inelastic scattering the proton is regarded as a set of quarks each carrying a fraction x of the proton longitudinal momentum. Since the strong interaction binding the quarks in the proton reference frame is non-perturbative, the probability that a quark of flavor q contributes with a fraction x to the proton momentum is described by a *parton distribution function* $p(x, q)$ based on phenomenological models. The partonic collision, on the other hand, is governed by perturbative QCD processes. Just as accelerated electric charges emit photons, the colored particles arising in the final state irradiate gluons. Gluons, being colored, emit further QCD radiation creating the so-called parton shower. As the shower develops the energy of its components decreases and, around 1 GeV, the QCD interaction regime turns into a non-perturbative phase, which cannot be calculated exactly. At this stage hadronization processes recombine the resulting partons into observable color singlets hadrons.
The dynamics of the hadronization process is not yet fully understood, hence its simulation

relies on models that are tuned to fit the data [23] such as the *Cluster Model*, *Color String Model* and *UCLA Model*.

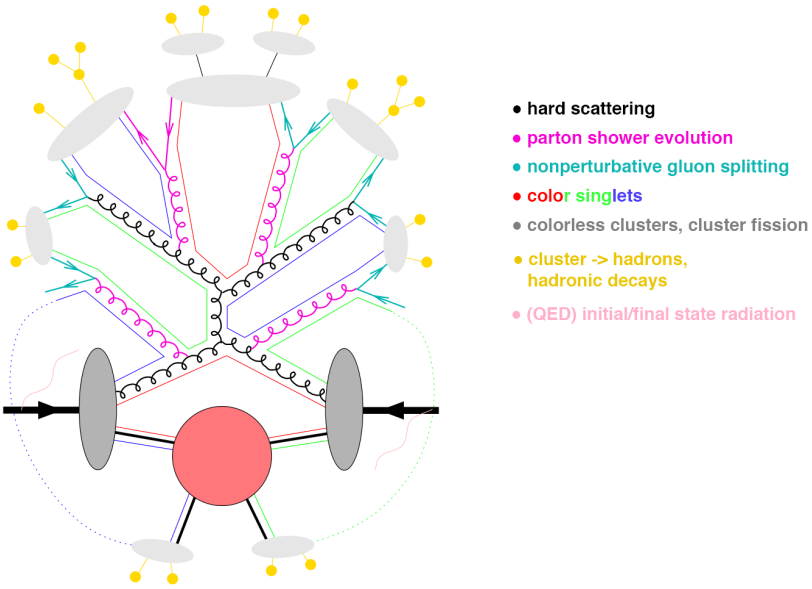


Figure 2.6: Representation of showering and hadronization after a deep inelastic scattering between two protons (in grey). The showering originates from a gluon-gluon scattering and terminates with colorless clusters. The hadronization transforms the clusters into hadrons (in yellow). On the bottom of the image the pink circle represents an underlying event that generally consists in a soft scattering among the rest of the partons and the production of jets with low η and p_T.
Taken from Dieter Zeppenfeld's PiTP 2005 lectures.

2.4.2 Jet reconstruction algorithms

From the experimental point of view a parton from the final state of the inelastic scattering appears as a bunch of collimated hadrons called *jet*. In practice to map the observed hadrons onto a set of jets a precise jet definition algorithm is needed. Two main classes of algorithm exist: cone and clustering based.
A generic cone algorithm [24] starts with some seed particle i, sums the momenta of all particles j within a cone of opening-angle R, typically defined in terms of pseudorapidity and azimuthal angle. Then it takes the direction of this sum as a new seed and repeats until the cone is stable, and calls the contents of the resulting stable cone a jet if its transverse momentum is above some threshold $p_{T,min}$. The parameters R and $p_{T,min}$ should be chosen according to the needs of a given analysis.
A clustering algorithm, instead, iteratively merges pairs of particle candidates in order of increasing relative transverse momentum into jets, until a stopping requirement is achieved, typically when the "distance" between adjacent jets is greater than some value.

Given two candidates i and j this "distance" may e.g. be defined as:

$$d_{ij} = min(p_{Ti}^{2\kappa}, p_{Tj}^{2\kappa}) \frac{\Delta R_{ij}^2}{R^2} \tag{2.2}$$

where ΔE is the distance in the $r - \phi$ plane, R and κ are arbitrary parameters. One of the most widely clustering algorithm used in CMS, called *anti-kT* [25] [26], is based on d_{ij} with κ set to -1. With this choice, the distance d_{ij} between a soft and a hard particle is dominated by the hard-particle p_T. Instead, two soft particles with a similar separation ΔR_{ij} would have a larger distance d_{ij}. As a consequence, soft particles will tend to cluster with hard ones before clustering among themselves. If a hard particle has no hard neighbors within a distance 2R, then it will simply accumulate all the soft particles within a circle of radius R, resulting in a conical jet.

2.4.3 Jet reconstruction information

The jet algorithms may be used with one or two recombination schemes for adding constituents. In the *enery scheme*, constituents are simply added as four-vectors and this produces massive jets. In the E_T*scheme*, massless jets are produced by equating the jet transverse momentum to the $\sum E_T$ of the constituents and then fixing the direction of the jet in one of two ways. In all cases the jet E_T is equal to p_T. In CMS four different jets types are considered depending on the information provided to the reconstructing algorithm.

Calorimeter jet (CALOjets): jets are reconstructed using energy deposits in the electromagnetic (ECAL) and hadronic (HCAL) calorimeter cells, combined into calorimeter towers. A calorimeter tower consists of one or more HCAL cells and the geometrically corresponding ECAL crystals. The association between HCAL cells and ECAL crystals is more complex in the end-cap regions of the electromagnetic calorimeter. In order to suppress the contribution from calorimeter readout electronics noise, thresholds are applied on energies of individual cells when building towers for event pile-up calorimeter towers with transverse energy of $E_{towers} < 0.3GeV$ are not used in jet reconstruction. CALOjets based selection is often employed for triggering purpose thanks to the calorimeter fast response.

Jet-Plus-track (JPT): this method exploits the excellent performance of the CMS tracking detectors to improve the p_T response and resolution of calorimeter jets. Calorimeter jets are reconstructed first as described above, then charged particle tracks are associated with each jet based on spatial separation in $\eta - \phi$ between the jet axis and the track momentum measured at the interaction vertex. The associated tracks are projected onto the surface of the calorimeter and classified as in-cone tracks if they point to within the jet cone around the jet axis on the calorimeter surface. If the 3.8 T magnetic field of CMS has instead bent the track out of the jet cone, it is classified as a out-of-cone track. The momenta of both in-cone and out-cone tracks are then added to the energy of the associated calorimeter jet. For in-cone tracks the expected average energy deposition in the calorimeters is subtracted based on the momentum of the track. The direction of the axis of the original calorimeter jet is also corrected by the algorithm

Particle Flow (PFJet): the information used to feed the reconstruction algorithm is collected according to the Particle Flow technique described in section 2.3. The jet momentum and spatial resolutions are expected to be improved with respect to calorimeter jets as the use of the

tracking detectors and of the excellent granularity of the ECAL allows to resolve and precisely measure charged hadrons and photons inside jets, which constitute 90 % of the jet energy.

Track Jets: jets are reconstructed from tracks of charged particles measured in the central tracker. Only well-measured tracks, based on their association with the primary vertex and their quality, are used by the algorithm. The method is completely independent from the calorimetric measurements, allowing for crosschecks.

2.5 B-TAGGING

Among all the jets it is possible to distinguish the ones originated by a bottom quark. The small cb and ub CKM matrix elements make the proper decay length of the b-quark about $c\tau \sim 450\mu m$, which corresponds to several millimeters in the laboratory frame for jets of the energy typical of Higgs decay. From the experimental point of view this means that part of the particles of b-jets originate from a *secondary vertex* which is displaced from the primary one.
The compatibility of a track to the primary vertex is evaluated through the *impact parameter (IP)*. The IP, as depicted in fig. 2.7 is the distance between the primary vertex and the line tangent to the track in the point corresponding to minimum distance to jet axis. A sign is assigned to the IP according to the scalar product between the IP and jet axis unit vectors. So that a positive (negative) IP corresponds to a downstream (upstream) decay with respect to the jet.

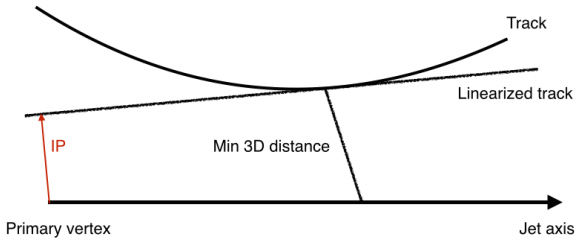


Figure 2.7: Impact parameter definition in 3D

Only tracks fulfilling the following criteria are used for b-tagging:

- angular distance between track and jet axis $\Delta R < 0.3$;
- number of pixel hits ≥ 2 and number of tracker hits (including pixel) ≥ 8;
- distance smaller than 0.2 cm (17 cm) in the transverse plane (along the beam axis) between the track and the primary vertex at the point of closest approach of the trajectory to the PV in the transverse plane;
- transverse momentum $p_T > 1GeV$;
- normalized $\chi^2 < 5$;

- distance to jet axis < 0.07 cm, defined as the spatial distance between the trajectory and the jet axis at their point of closest approach, where the jet axis is reconstructed with respect to the primary vertex;
- decay length < 5 cm, defined as the spatial distance between the primary vertex and the point of closest approach between the track trajectory and the jet axis.

Different b-tagging algorithms have been developed in CMS:

Simple Secondary Vertex (SSV) This algorithm uses the significance of the flight distance (the ratio of the flight distance to its estimated uncertainty) as the discriminating variable but its efficiency is limited by the secondary vertex reconstruction to about 65 %. [27]

Jet Probability (JP) It entails computing the compatibility of a set of tracks with the hypothesis of having originated from the primary vertex. Tracks with negative impact parameter are used to extract a resolution function, which is used to calibrate the impact parameter significance distribution. The (signed) probability is flat between -1 and 1 for tracks coming from the primary vertex, and positive and concentrated near 0 for tracks with large impact parameter significance.

Soft Lepton It tags b-jets by searching for electrons or muons from the semi-leptonic B hadron decay, which typically has a large momentum with respect to the jet axis and a large impact parameter. The b tag discriminator is the output of a neural net trained on four characteristic variables, the p_T^{rel} (the lepton p_T relative to the jet direction), the 3D impact parameter significance of the lepton track, the ratio between the lepton p_T and the jet energy, and the angular separation between the lepton and the jet axis. Although the efficiency of these lepton-based algorithms is limited by the intrinsic $B \to \ell\nu + X$ branching ratio, the information can be integrated in the more performing combined algorithms.

Inclusive Vertex Finder (IVF) The IVF technique[28] reconstructs secondary vertices independently of jet reconstruction and is particularly suited for B-hadrons decays at small angles which would lead to overlapping jets, or completely merged jets. First the tracks characterized by high three-dimensional impact parameter (IP) significance (IP normalized on its uncertainty) are selected and labeled *seeding tracks*.The *seeding tracks* are clustered with their surrounding tracks according to a compatibility requirement evaluated in terms of separation distance in three dimensions, separation distance significance, and angular separation. The clustered tracks are then fitted to a common vertex with an outlier-resistant fitter [29] [30]. The vertices sharing more than 70% of the tracks compatible within the uncertainties are merged. As a final step, all tracks are assigned to either the primary or the secondary vertices on the basis of the significance of the track to vertex distance.
Each secondary vertex is associated to the decay of a b-quark. The efficiency can be improved requiring conditions on the secondary vertex.

Combined Secondary Vertex (CSV) It has been designed to improve the efficiency with respect to the others methods and is now widely used in CMS. The b-tagging efficiency of CSV is reported in fig. 2.8. The CSV algorithm [31] is able to b-tag a jet even if no secondary vertex has been reconstructed in it, using all the variables listed below:

- the number of tracks in the jet;
- the number of tracks associated to the secondary vertex (if present);
- the secondary vertex mass (if present);
- the 2D flight distance significance (σ_{IP}/IP) of the secondary vertex (if present);
- the ratio of the energy carried by tracks at the secondary vertex with respect to all tracks in the jet;
- the pseudorapidities η of the tracks at the vertex with respect to the jet axis;
- he 2D IP significance of the first track ordered by decreasing IP significance that raises the secondary vertex mass above the charm threshold;
- the 3D IP significances for each track in the jet.

Two likelihood ratios are built from these variables, to discriminate between b and c jets and between b and light-parton jets. The two likelihood ratios are then combined into a unique scalar CSV disciminator ranging from 0 (no-btag) to 1 (good b-tag). For practical reasons three CSV working points have been defined: *loose cut* (CSV = 0.244), *medium cut* (CSV=0.679), *tight cut* (CSV = 0.898).

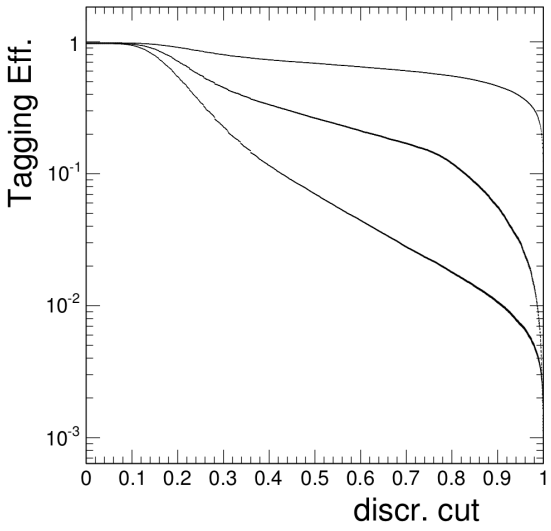


Figure 2.8: Efficiencies to tag a jet as b-jet versus the cut on the discriminator output for b-jets (top curve), c-jets (middle curve) and udsg-jets (bottom curve) [31].

CHAPTER 3

$hh \rightarrow 4b$ 8 TEV ANALYSIS

3.1 INTRODUCTION AND ANALYSIS STRATEGY

The analysis aims to impose a limit on the cross section of the double Higgs production in the final state of largest branching ratio, in which each boson decays into a bottom and an anti-bottom quark.
The final state we are looking at features two b-dijets (that is two pairs of b-jets) almost back-to back in ϕ and with invariant mass compatible with the Higgs mass. The reconstruction is complicated by the matching the right jets into correct pairs that has to cope with the irreducible QCD multijet background.
The strategy adopted is the following: all the kinematical variables will be tested in terms of discrimination power and the most relevant ones will feed a multivariate algorithm. The response of this algorithm, together with the jets b-tagging variables, will be used to define a control and a signal region used to extract a limit on the signal strength from a counting experiment. The signal strength will be also calculated by means of a bi-dimensional fit on the algorithm response and the dijet mass.
The analysis will take advantage of a multivariate b-tag algorithm, namely CMVA, that combines the output discriminants of several different b-tagging estimators with a neural network. In detail the combined algorithms are CSV, JP, and soft lepton taggers, described in [32], and IVF, described in [28]. In [32] and [33] the CMVA is proved to be more efficient than the CSV as shown by fig. 3.1. The CMVA response, just like the CVS one, is a scalar ranging between 0 (not b) and 1 (most likely b); the working points are the same of the CSV.

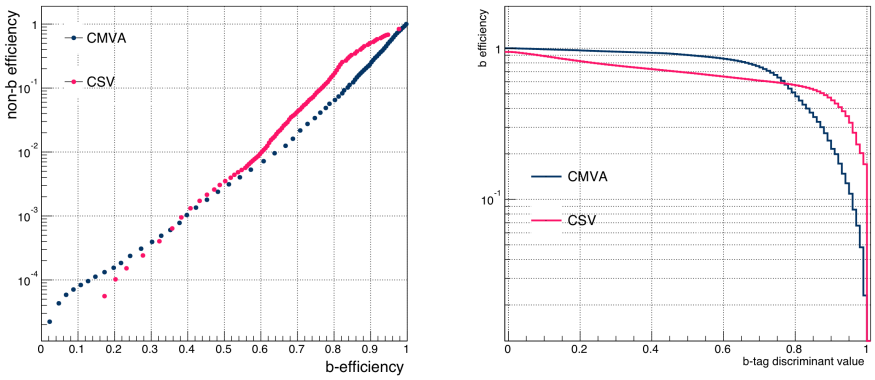


Figure 3.1: Misidentification probability as a function of b-tagging efficiency (left), identification efficiency as a function of the CSV and CMVA discriminant values (right), evaluated in a simulated $t\bar{t}$ sample for central jets with $p_T > 30GeV$ [33].

3.2 Signal, Background and Data samples

Signal The signal features were modeled to create a 300,000 $hh \to b\bar{b}b\bar{b}$ events private Monte Carlo sample. The sample was generated by MadGraph5 _ aMC NLO [34] using the `CT10` [35] parton distribution functions. Pythia 6 simulated parton showering and hadronization while GEANT 4 [36] the interaction of the stable particles with the detector. Finally the events were recontructed and processed with the same algorithms, including the trigger ones, used for real data.
Given the cross section of $9.96 \pm 0.92 fb$ [37] and the branching ratio of $58\% \pm 3.2$ the sample has an equivalent luminosity of $88.99 \pm 10 pb^{-1}$.

Background The background, being almost entirely produced by irreducible QCD processes, is difficult to simulate reliably and with high enough luminosity. As a consequence the background features have been modeled on a small fraction [(1)] of data events. The choice of a data-driven analysis was, in this case, unavoidable but partly justified by the tiny signal expectation, considered the low cross section and the limited statistic available for the search.

Data The data used for the analyses had been collected by CMS during the LHC "Run I" in 2012. Specifically only the datasets on which the `HLT-DiPFJet80-DiPFJet30-BTagCSVd07d05` trigger path is present were considered. This trigger, suitable for our study, had been implemented on May 9^{th} 2012 making the collected statistic equal to $17.9 fb^{-1}$ out of the total $19.7 fb^{-1}$ acquired by CMS in Run I.
The samples employed for the analysis are listed in table 3.1.

[(1)] the amount will be specified later on.

dataset	luminosity (fb^{-1})
/BJetPlusX/Run2012B-13Jul2012-v1	4.412
/BJetPlusX/Run2012C-24Aug2012-v2	0.474
/BJetPlusX/Run2012C-PromptReco-v2	6.330
/BJetPlusX/Run2012D-PromptReco-v1	6.712
total	17.928

Table 3.1: Datasets employed for the analysis.

Trigger path The final state characterized by 4 b-jets is particularly challenging to trigger on in the crowd of QCD jets. To keep the low thresholds for jet p_T appropriate for our analysis and an acceptable rate, a trigger exploiting the b-tag information at HLT level has been used. The HLT-DiPFJet80-DiPFJet30-BTagCSVd07d05 seeds on events that contain two Level 1 (L1) jets above a threshold, then, it selects events with four HLT anti-KT5 jets above four p_T thresholds, and requires a minimum HLT b-tagging value for two of the four E_T leading jets at Level 3 (L3) with full regional HLT tracking information. The trigger employs an online version of the Combined Secondary Vertex algorithm described in [31].

- L1 DoubleJetC56 or L1 DoubleJetC64
- Reconstruct anti-kT 0.5 L1FastJetCorrected CaloJets

  2 Jets with $|\eta| < 2.6$ and $p_T > 75$ GeV

  4 Jets with $|\eta| < 2.6$ and $p_T > 25$ GeV
- Fast Primary Vertex Reconstruction

  $|z| < 25$ cm, r < 2 cm
- Online-Combined Secondary Vertex (CSV) computation

  1 CaloJet with $p_T > 20$ GeV must have CSV > 0.7

  2 CaloJets with $p_T > 20$ GeV must have CSV > 0.5
- PF Reconstruction Sequence

  2 PFJets with $|\eta| < 2.6$ and $p_T > 80$ GeV

  4 PFJets with $|\eta| < 2.6$ and$p_T > 30$ GeV

3.3 SIGNAL PRE-SELECTION

Events from the Monte Carlo sample contain up to 13 jets with an average of 6. Each jet is supplied with a CMVA value but, without looking at the MC truth, there is no obvious way to identify the jets from the Higgs bosons and match them in correct pairs. Several algorithms can be designed and implemented, but all have to cope with the b-tagging efficiency (70% around the CMVA medium cut working point and 40% around the tight one) and with the reserving of some variables for further steps of the analysis. For example if the final result, say the signal strength, is going to be extracted from a shape fit involving a set of a variables, make use of such variables in an event pre-selection will introduce a bias.

We anticipate that, in our case, an upper limit on the signal strength, which is precisely our final goal, will be obtained from

1. a shape fit on the dijet mass and the response of a multivariate tool;
2. a counting experiment based on the number of b-tags and, again, the response of a multivariate tool.

This means that we cannot operate directly on the number of b-tags or the dijet masses in this phase, turning down two essential signal features.
Incidentally let us observe that in our signal topology not all the four h-jets (i.e. jets coming from the two bosons) have to be necessarily b-tagged: if one or more has $|\eta| > 2.5$ it will not (poorly) be b-tagged due to the lack of tracker data. The tracker, in fact, does not cover such low angles.
Nevertheless an efficient selection, hereby described, was conceived. In each event that passes the trigger requirements, only jets with $p_T > 20$ GeV are retained. Then the first three jets in b-tag ranking are considered, provided that their CMVA is above the medium cut. Each of the remaining jets, one a time, is provisionally considered the fourth jet and the four are then matched in pairs of dijets. The matching yielding the least invariant mass difference between the dijets is recorded. Finally all the pairings recorded are scrutinized and the one characterized, again, by least invariant mass between dijets, is retained. In other words we require the dijets mass to be nearby equal without imposing it to be the closest to 125 GeV.
In fig. 3.2 we see the dijets masses.

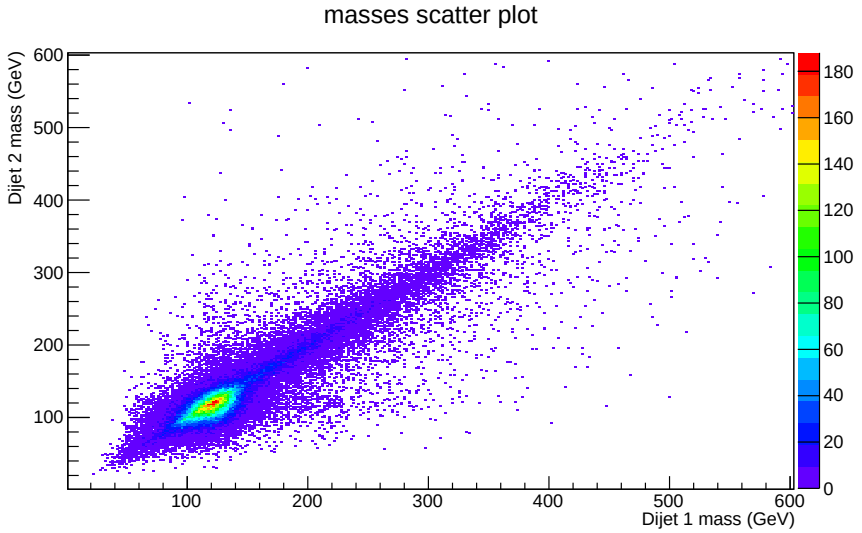


Figure 3.2: $b\bar{b}$ dijets mass scatter plot for the signal MC sample. Dijet 1 is the one with the highest p_T between the two.

The spot around 125 GeV for each dijet tells us that many jet pairs are matched properly, but at the same time the diagonal spray of events tells us that our criterion often finds a match that does not correspond to the true Higgs decay in jet pairs.
Only the events in which the four b-jets are correctly matched and associated to the relative bosons are helpful in modeling variables that we will use to distinguish the signal from the underlying QCD. Some of these variables, in fact, will describe the kinematics of the two

	Signal (MC)	Background (DATA)
total events	298165	27691653
after trigger	119259 (40%)	10724673 (39%)
after jets pairing	68454 (23%)	433621 (2%)
inside mass region	33198 (11%)	80668 (0.3%)

Table 3.2: Datasets employed for the analysis.

bosons reconstructed from the jet pairs and, obviously, only a perfect jets-to-Higgs matching prospects a successful outcome. We regard the events properly reconstructed as the ones inside a square of 50 GeV side centered at 125 GeV for each dijet [(2)]. The efficiency of this selection procedure is shown in table 3.2; it has been applied to the signal Monte Carlo and to a data sample, namely `/BJetPlusX/Run2012B-13Jul2012-v1`, regarded as background.

Among all the possible discriminating variables we choose the ones listed in table 3.3. Looking forward to applying a multivariate analysis all of them are one-entry-per-event variables.

Single h-jet:	p_T; η; *CMVA*; min and average of the *three highest CMVAs*.
Other jets:	*centrality*; max, min, average of p_T, η, *cmva*.
Single h-dijets:	p_T; $\Delta\phi$; $\Delta\eta$; ΔR; τ.
Two h-dijet system:	*hh invariant mass*; $cos(\theta_{CS})$; $cos(\theta^*)$; $\Delta\phi$; $\Delta\eta$; ΔR; *centrality*; *absolute value of* p_T *vector sum*.
Global:	*MET*.

Table 3.3: Full list of kinematic variables.

The Collins Soper angle (θ_{CS}) is described in [38] while the twist τ and centrality are defined as:

$$centrality = \sum_{jets} \frac{p_T}{E} \qquad \tau = tan^{-1}\left(\frac{\Delta\phi}{\Delta\eta}\right) \tag{3.1}$$

The *centrality* is an indicator of how much hard is the scattering, its value lies between 0 and 1. The *twist* is a longitudinal boost-invariant version of the rotation of the $h\,b\bar{b}$ plane with respect to the beam-h plane, it is 0 when the jets are separated in η, and $\pi/2$ when separated in ϕ.

3.4 THE KINEMATICAL STUDY

A study of the kinematics was performed: every variable from table 3.3 was tested for discrimination power. The aim of the step is to catch the peculiar features of the signal, as a consequence only the events in the dijets mass square defined in the previous section were examined. Three statistical methods were applied to rank the variables by signal-background discrimination potential: the Kolmogorov-Smirnov, the Anderson-Darling and a Likelihood based one.

Likelihood based test (L) Let s_i and b_i be the i-th bin content of signal and background sample respectively. Now let us build a toy sample that mixes the two samples according to a bin-by-

[(2)] 50 GeV is a rough round up of the mass resolution, estimated by the width of the peak, which is about 30 GeV.

bin Poisson distribution. Explicitly the toy sample i-th bin content n_i is picked from

$$\frac{s_i + b_i}{n!} e^{-(s_i+b_i)} \tag{3.2}$$

Then a bin-by-bin fit of the toy sample is performed. The pdf used is a Poisson distribution with mean μ_i given by

$$\mu_i = \frac{f_s \cdot s_i + (1 - f_s) \cdot b_i}{N} \tag{3.3}$$

where f_s is the free parameter representing the fraction of signal and N is the new sample integral. Maximizing the log-likelihood

$$logL = \sum_{i=1}^{N_{bins}} log\left(\frac{\mu^{n_i}}{n_i!} e^{-\mu_i}\right) \tag{3.4}$$

we get f_s and its error which is a clue of how different are the distributions. The smaller the error, the more different the distributions.

Kolmogorov-Smirnov test (KS) This test measures the probability that two samples belong to the same parent population according to the maximum distance between their cumulative distribution functions (CDF) [39]. If $F(x)$ and $G(x)$ are CDFs the statistic is given by

$$D_{KS} = max|F(x_i) - G(x_i)| \tag{3.5}$$

The KS test is very popular as it is distribution independent, simple to implement and its critical values are well known. However it is more sensitive to differences in the central part of the *pdf*s as, by definition, cumulative distributions approach 0 and 1 at the ends. In addition if there are repeated deviations in the pdf the *cdf*s cross each other multiple times and the maximum distance between them is not a good estimator. For the reasons above this test has been considered mainly as a check.

Anderson-Darling test (AD) A better test that makes use of the area between two samples cumulative distributions, weighting it near the tails is the Anderson-Darling. The statistic is given by [40]

$$A_{nm} = \frac{nm}{N} \int \frac{(F_n(x) - G_m(x))^2}{H_N(x)(1 - H_{N](x))}} dH_N(x) \tag{3.6}$$

where m and n are the samples number of entries and F and G the cdfs; H is the combined sample cdf defined as

$$H_N = (nF_n(x) + mG_m(x))/N \qquad N = m + n \tag{3.7}$$

The AD test is also distribution-independent and its critical values are tabulated.

L-rank	L	KS	KS-rank	AD	AD-rank	label
1	0.0176	0.377	4	9086	3	Dijet 2 p_T
2	0.0178	0.384	1	9274	1	Dijet 2 ΔR
3	0.0185	0.381	3	9068	4	Dijet 1 ΔR
4	0.0189	0.382	2	9205	2	Dijet 1 p_T
5	0.0213	0.252	13	4281	10	4^{th} jet CMVA
6	0.0223	0.312	5	7142	5	average CMVA of first 3 selected jets
7	0.0232	0.289	7	6070	6	3^{th} jet CMVA
8	0.0266	0.293	6	4360	9	Dijet 1 $\Delta\phi$
9	0.0276	0.262	9	4987	7	minimum CMVA of first 3 selected jets
10	0.0284	0.289	8	4229	11	Dijet 2 $\Delta\phi$
11	0.0284	0.221	14	3683	14	3^{th} jet p_T
12	0.0287	0.262	10	4987	8	2^{nd} jet CMVA
13	0.0297	0.261	11	3841	13	hh invariant mass
14	0.0309	0.261	12	4049	12	$cos(\theta^*)$
15	0.0316	0.129	25	1140	25	ΔR between dijets
16	0.0350	0.169	21	1951	22	4^{th} jet p_T
17	0.0351	0.194	18	2683	17	centrality (selected jets)
18	0.0352	0.212	15	2796	16	$\Delta\eta$ between dijets
19	0.0352	0.158	23	2243	20	2^{nd} jet p_T
20	0.0372	0.141	24	1791	23	1^{st} jet p_T
21	0.0374	0.205	16	2451	18	Dijet 2 $\Delta\eta$
22	0.0379	0.198	17	3414	15	1^{st} jet CMVA
23	0.0397	0.184	20	2383	19	$\Delta\phi$ between dijets
24	0.0410	0.189	19	2224	21	DiJet 1 $\Delta\eta$
25	0.0469	0.159	22	1631	24	$cos(\theta_{CS})$
26	0.0555	0.089	29	364	29	max CMVA among discared jets
27	0.0575	0.090	28	435	27	Dijet 1 invariant mass
28	0.0575	0.060	31	320	30	3^{rd} jet η
29	0.0586	0.121	26	764	26	Dijet 2 invariant mass
30	0.0645	0.092	27	388	28	average p_T of discarded jets
31	0.0674	0.052	34	146	37	max p_T among discared jets
32	0.0695	0.052	33	229	33	4^{th} jet η
33	0.0724	0.048	37	146	38	Dijets modulus of vector p_T sum
34	0.0729	0.064	30	275	32	average CMVA of discarded jets
35	0.0740	0.032	42	61	41	max $\|\eta\|$ amond discarded jets
36	0.0758	0.049	36	168	35	min CMVA among discarded jets
37	0.0771	0.028	43	52	43	min η among discarded jets
38	0.0823	0.057	32	200	34	Centrality (discarded jets)
39	0.0852	0.040	38	156	36	2^{nd} jet η
40	0.0957	0.033	40	104	39	1^{st} jet η
41	0.1048	0.040	39	65	40	min p_T among discarded jets
42	0.1266	0.051	35	299	31	Number of jets
43	0.1342	0.020	46	10	46	average η among discarded jets
44	0.1572	0.032	41	39	44	Dijet 1 τ
45	0.1878	0.027	44	53	42	MET
46	0.1935	0.024	45	21	45	Dijet 2 τ

Table 3.4: Variables ranking.

Variables ranking results The variables mentioned above are listed in table 3.4 by least f_s error ie. by greatest S/B difference according to the likelihood based method, as a reference. The ranking claimed by the KS and AD tests are also presented and are in good agreement with the L test. A correlation plot of the results for each pair of tests is presented in figure 3.3, as a proof of the consistency.

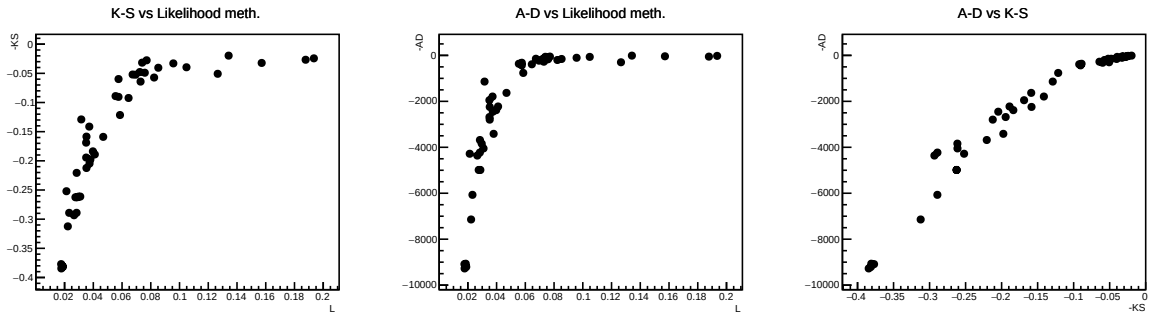
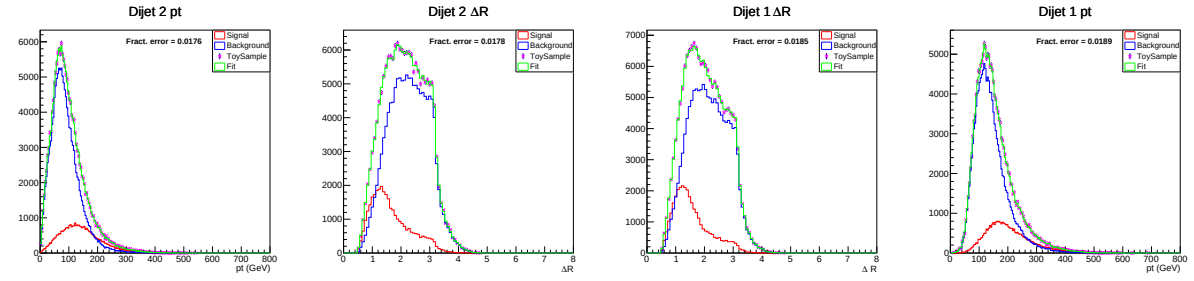


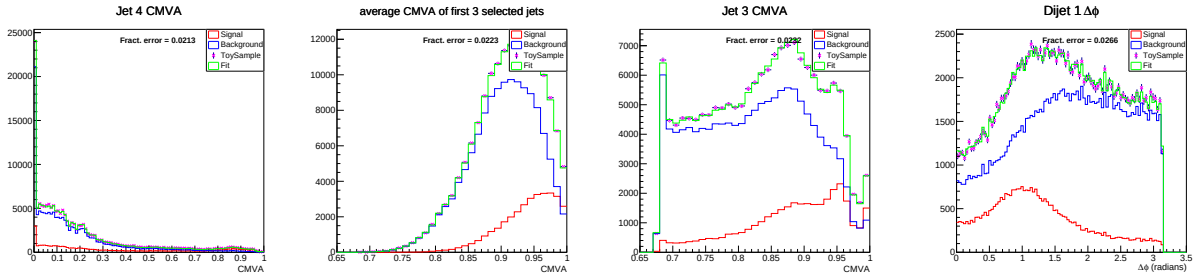
Figure 3.3: Correlation among different test results

The twenty-four distributions ranked as the ones with the highest discrimination potential by the likelihood method are reported in figure 3.4, 3.5, 3.6, 3.7 and 3.8 while the eight with the lowest presented in figure 3.9 and 3.10. Note that the histograms are scaled on purpose to have 20% signal and 80% background just to fix the fraction of signal that the likelihood based test fits. Any other fraction of signal would have been equally acceptable since the L-test is weakly dependent on such parameter choice [(3)].

[(3)]The test has been repeated with different signal fraction (1%, 5%, 10%, 40%, 50%) and small discrepancies were found in the exact rank. The first seven position are always taken by the same variables while the next ranks are mixed within a maximum range of 3 ranks.

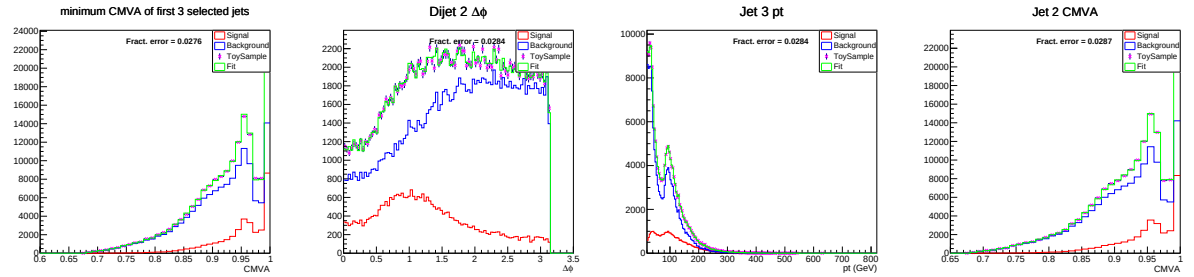


(a) 1^{st} (b) 2^{nd} (c) 3^{rd} (d) 4^{th}



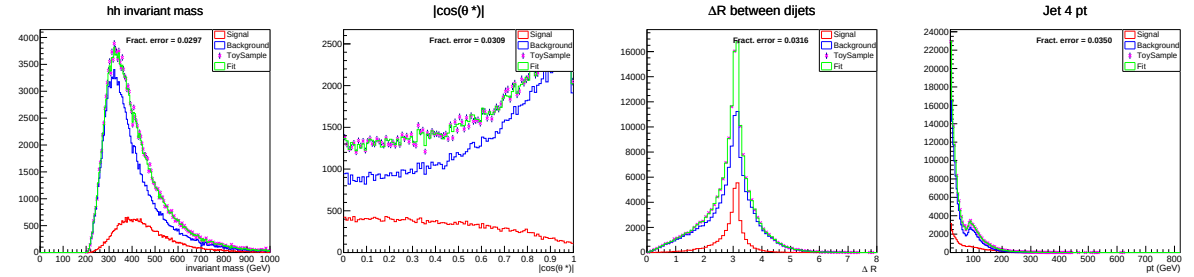
(e) 5^{th} (f) 6^{th} (g) 7^{th} (h) 8^{th}

Figure 3.4: 1^{st} to 8^{th} ranked variables distributions.



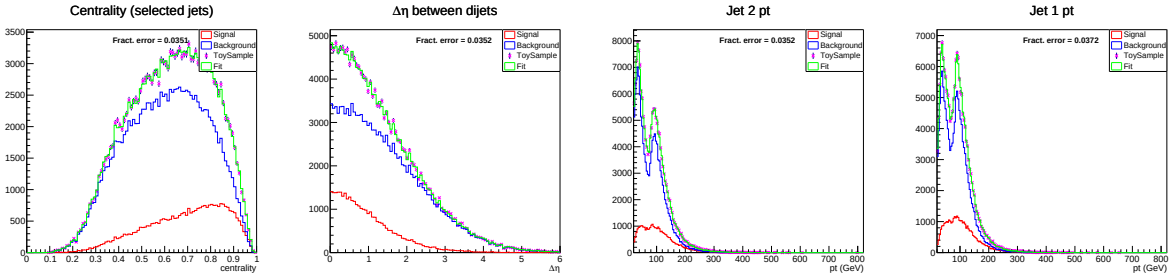
(a) 9^{th} (b) 10^{th} (c) 11^{th} (d) 12^{th}

Figure 3.5: 9^{th} to 12^{th} ranked variables distributions.



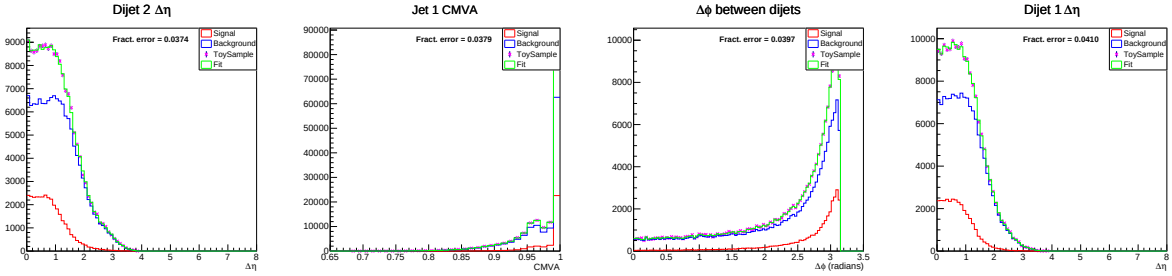
(a) 13^{th} (b) 14^{th} (c) 15^{th} (d) 16^{th}

Figure 3.6: 13^{th} to 16^{th} ranked variables distributions.



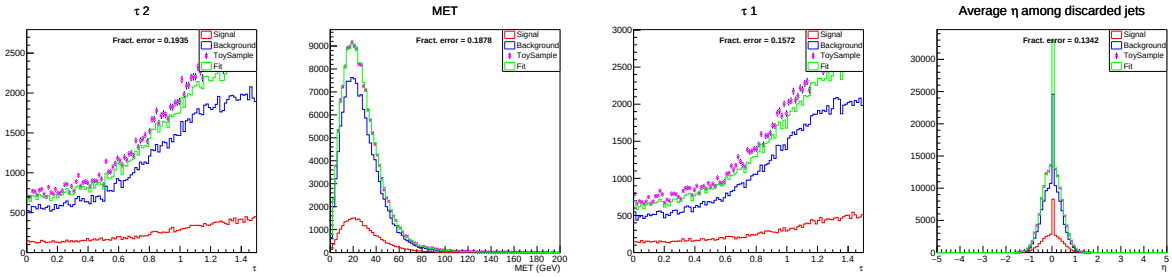
(a) 17^{th} (b) 18^{th} (c) 19^{th} (d) 20^{th}

Figure 3.7: 17^{th} to 20^{th} ranked variables distributions.



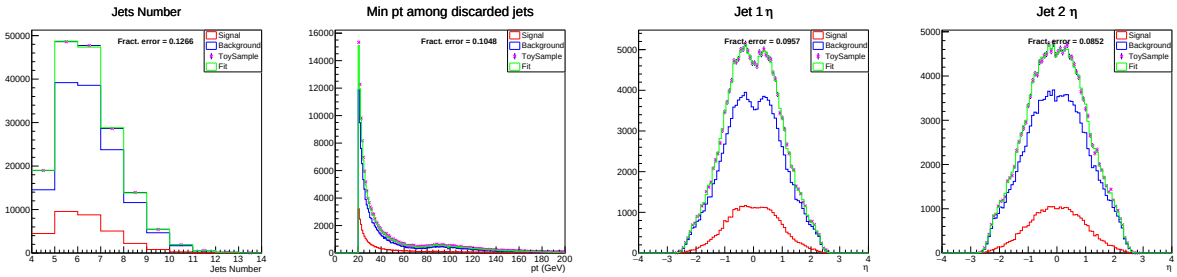
(a) 21^{st} (b) 22^{nd} (c) 23^{trd} (d) 24^{th}

Figure 3.8: 21^{st} to 24^{th} ranked variables distributions.



(a) 46^{th} (b) 45^{th} (c) 44^{th} (d) 43^{rd}

Figure 3.9: 46^{th} to 43^{rd} ranked variables distributions.



(a) 42^{nd} (b) 41^{st} (c) 40^{th} (d) 39^{th}

Figure 3.10: 42^{st} to 39^{th} ranked variables distributions.

3.5 MULTIVARIATE ANALYSIS

3.5.1 Introduction

The advantage of a multivariate analysis lies in the possibility of exploiting interesting structures in the n-dimensional space of the variables that may not be visible projecting the whole space onto one or two dimensions. If applied to binary classification, e.g. signal (S) vs background (B), these techniques generally rely on a *training* phase in which they learn from pure and known samples how to distinguish between the categories. Once the tool is trained is can be *applied* on unknown data to measure, for example, the amount of signal events. In the midst of *training* and *application* an optional *testing* phase can be included. The testing is an application on a limited part of the training samples performed mainly to check for over-training. Over-training happens when the algorithm separates the two categories perfectly in the training set, but its error on the test set is much larger because it learned on fine features not corresponding to real features of the samples, but arising as statistical fluctuations.
Our multivariate analysis will take advantage of the TMVA [41] package embedded in ROOT [42] that provides several methods among which the ones we will use: the Likelihood Ratio (LR) and the Boosted Decision Tree (BDT).
In general it is not granted that a multivariate analysis (MVA) performs better than a cut-based one, hence usually they are both carried out. The improvement from the LR result of a MVA method betrays how much use the MVA makes of the correlation among the input variables.
In this study, instead of a cut based analysis, we chose to take as reference a likelihood ratio test because of its well-known asymptotic properties.
In fact, according to the Neyman-Pearson lemma the likelihood ratio is the most *powerful* test statistic [(4)] for two-sample tests. However the lemma holds only if the variables characterizing the problem are independent i.e. that our multi-dimensional distributions should be factorized into-one dimensional independent ones [43] and clearly it is not our case, so the likelihood ratio test serves as a benchmark.

3.5.2 Projective Likelihood

The likelihood method [44] consists of building a model out of probability density functions (PDF) that reproduces the input variables for signal and background. For a given event, the likelihood for being of signal type is obtained by multiplying the signal probability densities of all input variables, which are assumed to be independent, and normalizing this by the sum of the signal and background likelihoods. As we already mentioned this would be the best method if the variables were independent. In formulas the likelihood ratio for the event i is given by

$$y_{\mathcal{L}} = \frac{\mathcal{L}_S(i)}{\mathcal{L}_S(i) + \mathcal{L}_B(i)} \tag{3.8}$$

where the likelihood is obtained by

$$\mathcal{L}_{S(B)}(i) = \prod_{k=1}^{n_{var}} p_{S(B),k}(x_k(i)) \tag{3.9}$$

[(4)]The power of a test statistic is the probability of rejecting the null hypothesis when the alternative is true.

where the product runs on the variables and $p_{S(B),k}(x_k(i))$ is the signal (background) PDF, normalized to 1, for the k-th input variable x_k. Since the parametric form of the PDFs is generally unknown, the PDF shapes are empirically approximated from the training data by parametric functions.
The response of the method is a scalar ranging from 0 (background) to 1 (signal).

3.5.3 Boosted Decision Tree

A Decision Tree is an algorithm that, through a sequence of binary splits of the data, is able to categorize events into different classes which, in our scope, are just two: signal and background. The value of these splits, operated on the variables characterizing each event, are determined in the training phase where the decision tree is exposed to pure signal and background samples.

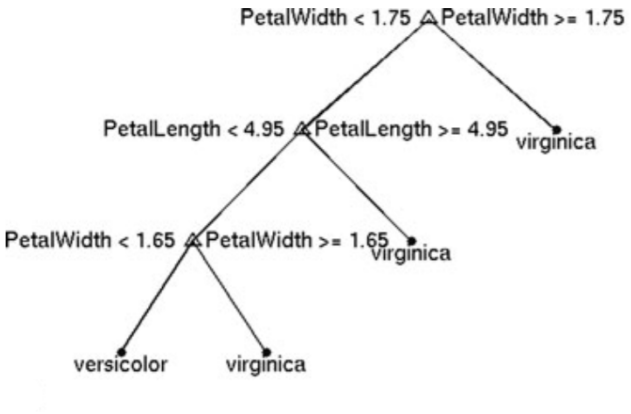


Figure 3.11: Example of a decision tree with four leaves applied to iris flower categorization [45].

To understand how an optimal split can be determined let us concentrate on the training phase: to begin with all the N_0 events are in one node t_0, each event has a label $y = S$ or B that indicates the event class. A split on the variables defines two child nodes: t_L with N_L events and t_R with N_R.
To each node is associated a probability $P(t) = N_t/N; t \in \{0, L, R\}$ and a posterior probability $P(S(B)|t) = N_{S(B)}/N_t$, where $N_{S(B)}$ is the number of events in the node of class signal (background) [45].
Since a tree node predicts into the class with the largest posterior, the training error is $\epsilon(t) = min_{y\in\{S,B\}}P(y|t)$. If a tree node contains observations of one class only, it is called pure and its training error is zero. We would like a tree with pure nodes because it would confidently separate the classes, still, we do not want nodes with low probabilities $P(t)$ because a leafy tree would likely overfit the data. Growing a tree with big pure nodes may not be possible because the class distributions overlap. To choose the best split, we need a measure of impurity. Let us define the node impurity as a function of class probabilities

$$i(t) = \phi(P(A|t), P(B,t)) = \phi(p,q) \tag{3.10}$$

where an optimal choice [45], for ϕ is given by the quadratic function called *Gini diversity index*.

$$\phi(p,q) = 1 - p^2 - q^2 \tag{3.11}$$

A good decision split should minimize the impurity. Above, we have defined impurity for one node but a binary split produces two so we minimize the average impurity for the two children. The two nodes can differ in size, and for averaging it would be natural to weight their impurities by the node probabilities. The weighted node impurity is

$$I(t) = P(t)i(t) \tag{3.12}$$

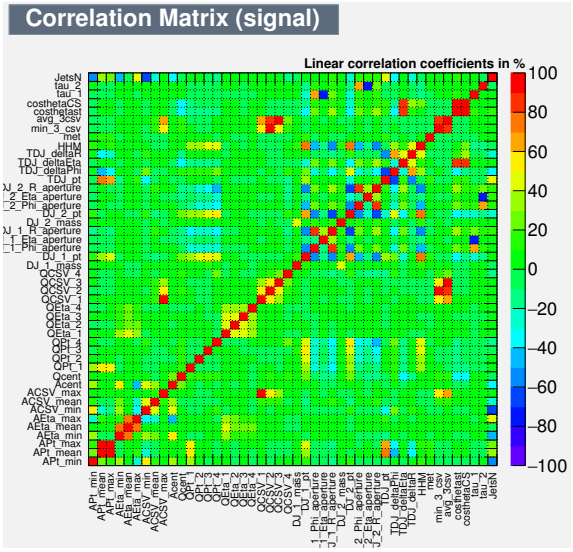
and the *impurity gain* after the splitting is

$$\Delta I = I(t_0) - I(t_L) - I(t_R) \tag{3.13}$$

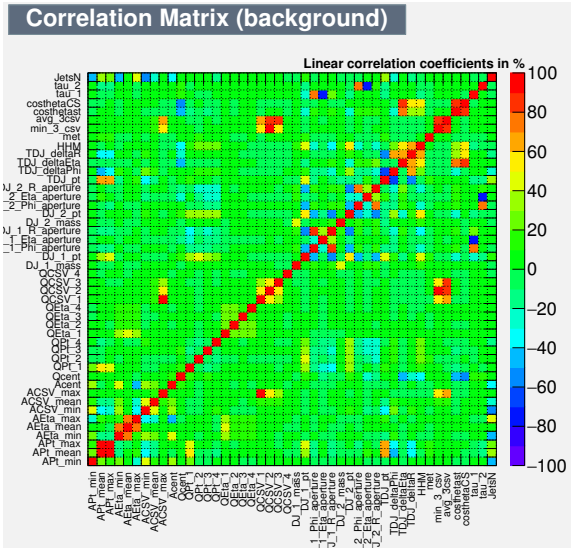
where $I(t_L) + I(t_R)$ accounts as an average.
The best splitting rule is the one that maximizes the impurity gain ΔI over all possible splits for all variables. If there are no splits with positive gain, the node cannot be split and becomes a leaf, or terminal node.
Often the growth of a tree is stopped before the positive-gain-splits are over. Usually the stop condition involves the tree depth (i.e. the maximum number of consecutive splittings) or the nodes purity.
A decision tree, as described here, is powerful but still unstable because of its dependency on the statistical fluctuations of the training sample. To overcome this shortcoming the technique of *boosting* has been developed. The training events that are misclassified by a decision tree have their weights increased (boosted) and a new tree is trained. This procedure is repeated several times obtaining a so-called "forest" of trees. The final classification is based on a majority vote of the classifications done by each tree. Different algorithms have been designed for this task, one of the most popular being AdaBoost [46].

3.5.4 TMVA preliminary study

Correlations The whole TMVA training was performed on a background sample of 25,000 events from dataset `/BJetPlusX/Run2012B-13Jul2012-v1` and a signal one of 25,000 events from the MC sample dropping the condition on the dijet mass.
However to proceed with the training the variable ranking is not sufficient to choose which variables are relevant as it does not account for correlations. In fig. 3.12 we see that many variables are highly correlated and train on all of them does not add much information. Clearly the minimum and maximum among the jets CMVA is correlated to the single jet values by definition, and the same holds for ΔR, $\Delta\phi$ and $\Delta\eta$. Less obvious is the correlation among $cos\theta^*$, $cos\theta_{CS}$ and $\Delta\eta$ between dijets.



(a) Signal correlation matrix



(b) Background correlation matrix

Figure 3.12: variables correlations calculated by TMVA

Training and Application events region From now on we will concentrate on the BDT method. The choice of dropping the condition on dijets mass to be inside the *"mass window"* of 100-150 GeV has been taken after careful understanding of the advantages and disadvantages. In an initial trial three possibilities regarding the region of BDT training and application were tested:

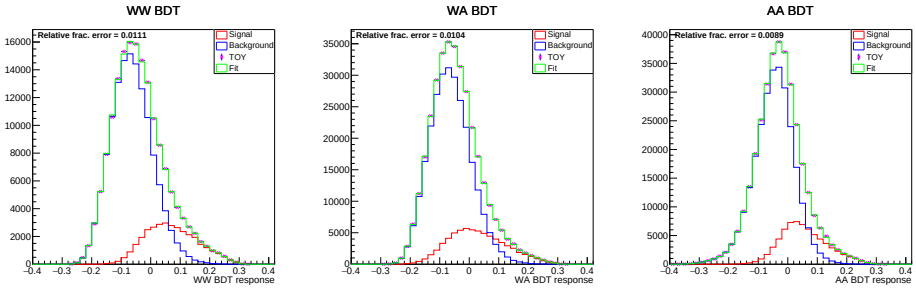
- **WW** training and application on the events inside the mass window
- **WA** training on the events inside the mass window and application on all the space
- **AA** training and application on all the events (mass condition dropped)

Ten variables, chosen accounting for the ranking and the correlations, were used to feed the BDT: $cos\theta^*$, dijet 1 and 2 p_T, dijet 1 and 2 ΔR, centrality, CMVA of 2^{nd}, 3^{rd} and 4^{th} jet, hh system invariant mass. For each of the three options the BDT training phase produces one response distribution for the signal and one for the background. These differences between these two distributions were measured by means of the likelihood-based method used in the variables ranking. Briefly the two histograms have been combined in a known fraction f_s that constitutes the fit parameter. The smaller is the error on f_s the more distinguishable are the distributions. The result of this procedure, shown [5] in fig. 3.13, suggest to proceed according to the third option, i.e. discard the dijet mass condition.

3.5.5 TMVA training and application

Our analysis strategy envisages the extraction of the signal strength from a counting experiment and from a bi-dimensional fit on the BDT response and dijet mass. We anticipate that the counting experiment will rely on control and signal regions defined according to jet b-tagging.

[5] The native TMVA BDT response has range [-1,1], here no care was taken to set to the customary [0,1] but it is not relevant to the point.



(a) WW option, f_s error = 1.11 %

(b) WA option, f_s error = 1.04 %

(c) AA option, f_s error = 0.89 %

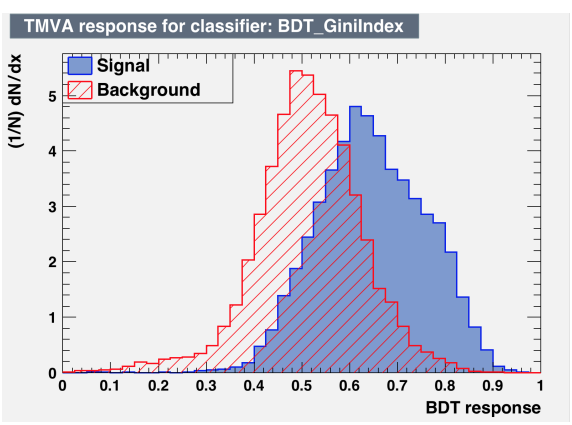
Figure 3.13: Different BDT performances depending on events selection.

These intents forced us to remove the CMVA-related variables and the dijets masses from the BDT input. Table 3.5 lists the selected nine input variables.

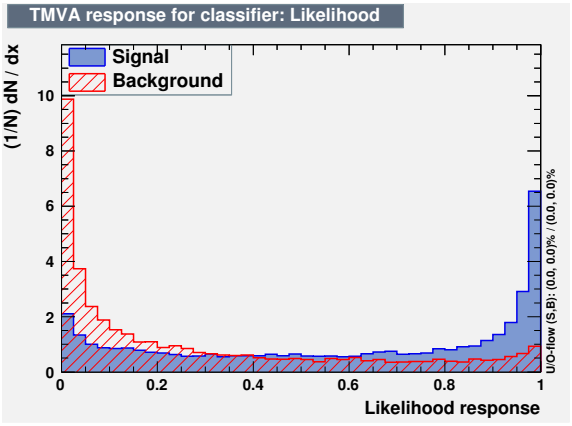
dijet 1 and 2 p_T
dijet 1 and 2 ΔR
centrality
3^{rd} jetp_T
hh system invariant mass
ΔR between dijets
$cos\theta^*$

Table 3.5: List of TMVA input variables

To reduce the correlations between the p_T and the ΔR of each dijet a principal component decomposition was applied. The principal component decomposition, or analysis, (PCA) is [44] a linear transformation that rotates a sample of data points such that the maximum variability is visible. In the PCA-transformed coordinate system, the largest variance by any projection of the data comes to lie on the first coordinate, denoted as"first principal component", the second largest variance on the second coordinate and so on. This practice could be employed to reduce the dimensionality of the problem by ignoring high orders components, but that is not our case.
The BDT and ProjectiveLikelihood methods have been again trained on 25,000 background events from the dataset `/BJetPlusX/Run2012B-13Jul2012-v1` and other 25,000 from the signal MC sample. Fig. 3.14 presents the outcome of the training phase performed on 8000 events from each sample. Finally the trained tool has been *applied* to all the datasets.



(a) BDT response



(b) Projective Likelihood response

Figure 3.14: Response of the two multivariate methods on testing sample

In fig. 3.15 the performance of the methods is shown by means of the ROC curve. As predicted the Likelihood ratio is not the best test statistic.

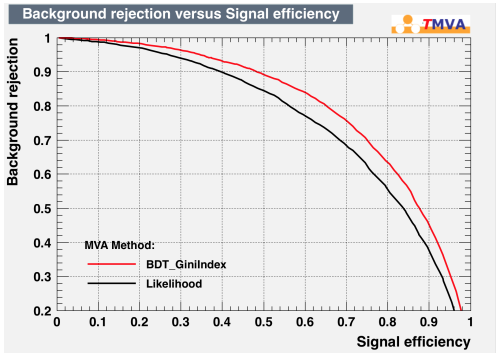


Figure 3.15: BDT and ProfileLikelihood ROC curve

3.5.6 BDT response cut

A cut value for the BDT response has been chosen maximizing a Figure of Merit (FOM) i.e. a function that quantifies the efficiency in signal selection. Several FOMs are present in literature [47], but we considered just the two in eq. (3.14), namely the Bityukov-Krasnikov S_{12} [48] and the likelihood-ratio-based $S_{\mathcal{L}}$ proposed in [49].

$$S_{12} = 2(\sqrt{s+b} - \sqrt{b}) \qquad S_{\mathcal{L}} = \sqrt{2ln(\mathcal{L}_{\mathcal{S}+\mathcal{B}}/\mathcal{L}_{\mathcal{B}})} \tag{3.14}$$

In S_{12} s and b are the number of signal and background events present above the cut i.e. in the signal region. In $S_{\mathcal{L}}$ $\mathcal{L}_{\mathcal{S}+\mathcal{B}}$ is the maximum likelihood value obtained in the full signal-plus-background fit, and $\mathcal{L}_{\mathcal{B}}$ is the maximum likelihood from the background fit only. The explicit expression is given is eq. (3.15). Of course the signal sample has been re-scaled to match the background luminosity.

$$\mathcal{L}_b = exp(-b)\frac{b^{b+s}}{(b+s)!} \qquad \mathcal{L}_{s+b} = exp(-s-b)\frac{(b+s)^{b+s}}{(b+s)!} \tag{3.15}$$

Both the FOMs have been computed and the outcome is shown in fig. 3.16. The optimal cut lies around 0.55. The surprising similarity of the plots is a proof of consistency.

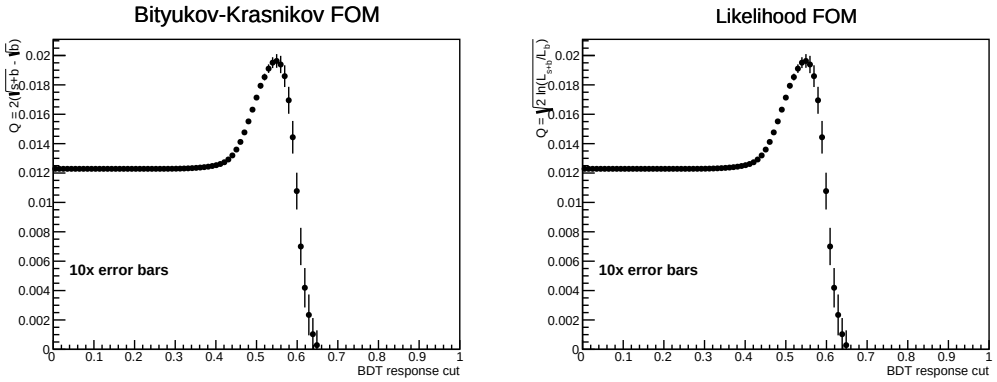
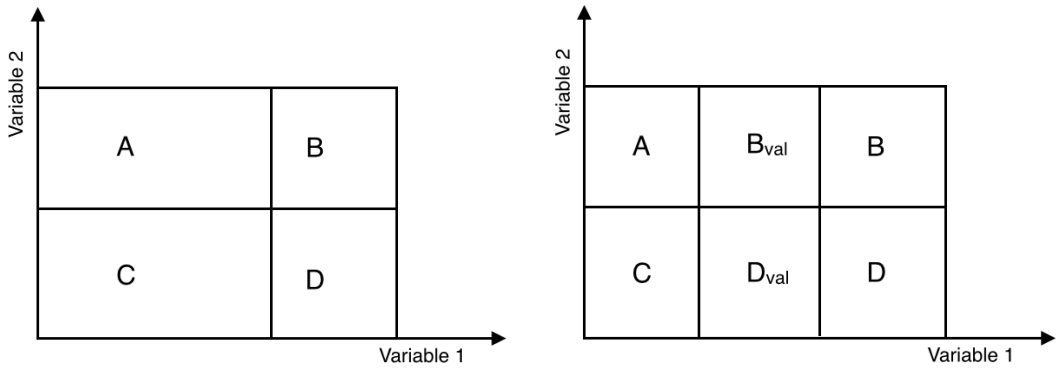


Figure 3.16: Figures of merit as functions of the BDT response cut.

3.6 ABCD COUNTING EXPERIMENT

The ABCD method is a technique that allows to estimate the number of background events expected in a given signal region. To explain it let us consider a general case in which some data in characterized by two independent variables, *var*1 and *var*2. Two cuts on these variables define four regions called A, B, C and D as depicted in fig. 3.17(a). The cuts are set in such a



(a) ABCD regions.

(b) ABCD plus validation regions.

Figure 3.17: Schematic representation of the ABCD method.

way that our signal populates prevalently the D region hence D takes the name of *signal region*. If the other regions, called collectively *control regions*, are sufficiently evenly populated it is possible to extract the number of expected background events $D^{bkg.exp.}$ present in D through the simple eq. (3.16).

$$N_D^{bkg.exp.} = \frac{N_C \cdot N_B}{N_A} \tag{3.16}$$

In general the regions do not need to be consecutive: an exclusion or a validation region can be inserted. With reference to fig. 3.17(b) we can exploit the B_{val} and D_{val} regions to estimate

a systematic uncertainty on $N_D^{bkg.exp.}$. Applying eq. (3.16) to A, B_{val} , C and D_{val} we obtain $N_{D_{val}}^{bkg.exp.}$. As the D_{val} is still in the control region, we can legitimately compare the background prediction with the actual value and extrapolate an estimation of the systematic uncertainty on $N_D^{bkg.exp.}$. First we define the source Δ of this systematic error and then we compute its statistical error σ_Δ by simple propagation. The relative systematic uncertainty on $N_D^{bkg.exp.}$ is then predicted according to eq. (3.17). The practice of subtracting to a source of systematic uncertainty its statistical uncertainty is motivated in [50].

$$\sigma^{syst.}_{D_{bkg.exp.}} = \sqrt{\Delta^2 - \sigma_\Delta^2} \qquad \Delta = \frac{N_{D_{val}} - N_{D_{val}}^{bkg.exp.}}{N_{D_{val}}^{bkg.exp.}} \tag{3.17}$$

3.6.1 Custom ABCD counting with b-tagging matrix

An extension of the ABCD method was developed for the present analysis. Nine regions were defined. To understand how they were delineated let us aid ourselves with fig. 3.18. On the x-axis, where the BDT response lies, two cuts are marked: one is the validation region cut and the other is the FOM-cut obtained by maximizing the figure of merit (see section 3.5.6). The exact value validation cut will be discussed later. On the y-axis lies the number of b-tags or rather the number of jets with CMVA greater than the medium cut, respectively two, three and four.

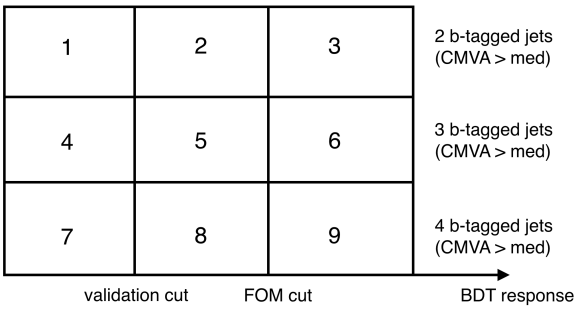


Figure 3.18: Schematic of the custom ABCD method developed.

The events in each region are then parameterized on the p_T, $|\eta|$ and number of constituent tracks of the third jet (the single jets associated to the two Higgs bosons are still sorted by decreasing CMVA). In other words each region is actually a 3-dimensional histogram and eq. (3.16) evolves (simplifying the notation) into eq. (3.18). The bins, 10 per variable, are not equally sized in order to avoid fine binning into low populated regions.

$$D_{bkg.exp.} = \sum_{i=1}^{p_T} \sum_{j=1}^{\eta} \sum_{k=1}^{trks} \frac{C_{ijk} \cdot B_{ijk}}{A_{ijk}} \tag{3.18}$$

The reason of this parametrization lies in the hidden dependency between the b-tagging and the BDT response. We already mentioned that the variables used in the ABCD method should be independent, however our BDT, even if not trained on the CMVA values, is instructed to recognize our four b-jets signal. As a consequence its response, even if not directly, is correlated

to the b-tagging. To account for this correlation we find some kinematical variables on which the b-tag probability C/A depends and then parametrize eq. (3.16) on them. To found that the number of b-tags is dependent on the third jet p_T, $|\eta|$ and number of constituent tracks. The ratios of events with ≥ 3 to $= 2$ b-tagged jets was plotted for each of the three variable and is shown in fig. 3.19.

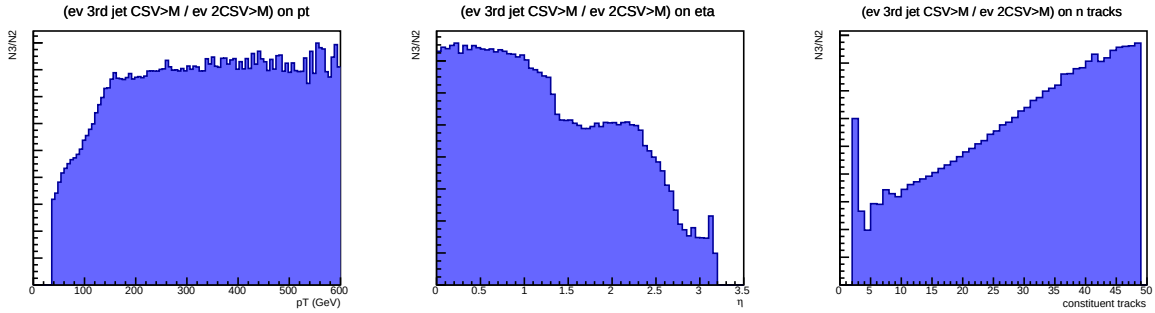


Figure 3.19: Ratio of events with ≥ 3 b-tagged jets on events with just 2 in p_T, η, constituent tracks in `/BJetPlusX/Run2012B-13Jul2012-v1` dataset.

The validation cut was fixed minimizing the mismatch Δ selecting regions A = 1, B = 3, C = 4, D = 6 which mimics our final extrapolation in regions with one less b-tagged jet. A wide range of validation cuts (0.38-0.52 BDT response where 0.55 is the FOM cut) was probed, fig. 3.20 reports the different values of Δ. A minimum is found at 0.40.

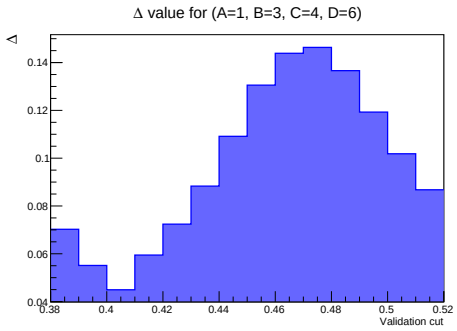


Figure 3.20: Relative mismatch between actual and background events predicted in region D = 6.

Our signal region is the number 9, that corresponds to 4 b-tagged jets and high BDT response. The expected background in the signal region is estimated assigning A = 4, B = 6, C = 7, D = 9. The relative systematic uncertainty is evaluated as explained in section 3.6 using A = 1, B = 3, C = 4, D = 6.
At the end we have D observed events with an expectation of $D_{bkg.exp}$ background events, and $N_{sig.exp} = \sigma_{hh} \cdot \mathcal{B}r \cdot \mathcal{L} \cdot \epsilon$ signal events. The efficiency ϵ is the ratio of the number of events of the Monte Carlo that fit in the D region to the total.

Source of systematic uncertainty	Impact on signal (%)	Impact on background (%)
luminosity	2.6	2.6
cross section	11.0	-
b-tagging scale factor	12.7	-
trigger scale factor	10	-
uncertainty on bkg. events	-	6.4

Table 3.6: Systematic uncertainties

3.6.2 Systematic uncertainties

The list of systematic uncertainties affecting the signal and background templates is summarized in table 3.6. Being the background modeled from data it is immune to many of the systematics.

- **luminosity:** An uncertainty of 2.5% has been measured for the integrated luminosity of $17.93 fb^{-1}$ of 8 TeV data used in this analysis. It affect also the expected signal because of the luminosity rescaling.

- **cross section:** The cross section $\sigma_{hh\to 4b} = 3.35 \pm 0.37 fb$. The uncertainty affects the precdicted number of events.

- **b-tagging:** the b-tagging scale factors for the CMVA algorithm are evaluated in [33].

- **trigger scale factor:** the uncertainty comes from the modeling of the trigger response in Monte Carlo simulations. The value used here is a round up of the one used in [33].

- **uncertainty on bkg. events** it is the uncertainty of the number of background events expected in the signal region. The uncertainty comes mainly from the propagation of eq. (3.16).

3.6.3 Limit extraction

The limit on the cross-section is extracted according to the prescriptions of the LHC Higgs Combination Group [51] using the Combine tool developed by the CMS Collaboration. The method used, called modified frequentist construction (CL_s), is determined by the choice of the test statistic and the treatment of the nuisance parameters.
The test statistic is based on a profile likelihood ratio [52] the evaluates the compatibility of data with the background-only and signal+background hypotheses, where the signal is allowed to be scaled by a factor μ, the signal strength modifier.
For every μ under test the observed value of the test statistic is calculated. Then the values of the nuisance parameters best describing the observed data for the background-only and signal+background hypotheses are respectively found. These two values of the nuisances are used to construct the *pdf*s of the test statistic under the two hypotheses through Monte Carlo

pseudo-data. Having constructed the two *pdf*s, two p-values are associated with the actual observation: p_μ for the signal+backgound hypothesis, p_b for the background-only.
The $CL_s(\mu)$ is defined as the ratio of these two probabilities (eq. (3.19))

$$CL_s(\mu) = \frac{p_\mu}{1 - p_b} \tag{3.19}$$

To quote the 95% Confidence Level upper limit on the signal strength, μ is adjusted until the condition $CL_s = 0.05$ is reached.

3.6.4 Results

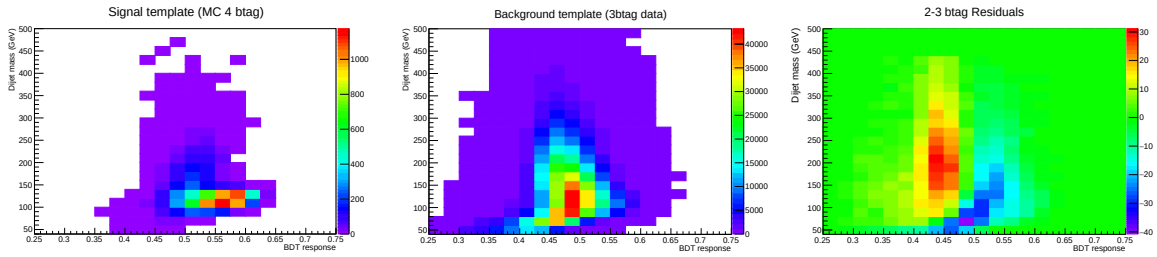
The final result is presented in section 3.6.4. The limit is expressed as a signal strength i.e. in units of Standard Model predicted cross section.

Events in signal region:	1711
Background expected:	$1929 \pm 88(\text{stat.}) \pm 86(\text{syst.})$
Signal Strength	$r < 135.5$

Table 3.7: Results from the counting experiment.

3.7 SHAPE FIT

Another way to extract a limit on the signal strength is to exploit a shape fit to test the data against a signal and a background hypothesis. The variables considered for this fit are the BDT response and the mean of the two dijets masses. The signal template is provided by the Monte Carlo events with 4 b-tagged (CMVA > medium cut) jets while the background template by all the data events with 3 b-tagged jets. The data to test are all the events with 4 b-tags.



(a) Signal template. (b) Background template. (c) 2-3 btag residuals (data).

Figure 3.21: Signal and background templates, test on background template.

Initially the background template was tested for reliability: if 3 b-tag events are regarded as background, 2 b-tag events should be background *a fortiori* and the two sets should have a similar shape. Fig. 3.21(c), which is a residual plot between the two shapes, shows some discrepancies, and that is fair since 3-btag events, being more signal-like, hold a higher BDT response. For our fit purpose this discrepancy represent an acceptable source of systematic uncertainty which has be taken into account. Calling the mass vs BDT histograms h_{ij}^{btags}, where

Source of systematic uncertainty	Impact on signal (%)	Impact on background (%)
luminosity	2.6	2.6
cross section	0.11	-
b-tagging scale factor	12.7	-
trigger scale factor	10	-
normalization on 4btag events	-	50
2-3 btag systematics	-	shape

Table 3.8: Systematic uncertainties.

btag stands for the number of b-tagged jets and *ij* are the bin indices, the systematic uncertainty is estimated for each bin according to eq. (3.20) and eq. (3.21).

$$\sigma_{ij}^{syst.} = \sqrt{\Delta_{ij}^2 - \sigma_{\Delta,ij}^2} \tag{3.20}$$

$$\Delta_{ij} = h_{ij}^{3b} - h_{ij}^{2b} \cdot \sum_{mn} \frac{h_{mn}^{3b}}{h_{mn}^{2b}} \tag{3.21}$$

The signal shape was rescaled by a factor $f' = \mathcal{L}_b/\mathcal{L}_s$ to match the data luminosity, while the background shape in the 4-btag region was extrapolated by a simple rescaling of a factor $f = \int 4btag \; / \int 3btag$ where the integrals stands for the number of events with four and three b-tags respectively.

3.7.1 Systematic uncertainties

The list of systematic uncertainties affecting the signal and background templates is summarized in table 3.8. Being the background modeled from data it is immune to many of the systematics.

- **luminosity:** An uncertainty of 2.5% has been measured for the integrated luminosity of $17.93 fb^{-1}$ of 8 TeV data used in this analysis. It affects also the expected signal because of the luminosity rescaling.
- **cross section:** The cross section $\sigma_{hh\to 4b} = 3.35 \pm 0.37 fb$. The uncertainty affects the predicted number of events.
- **b-tagging:** the b-tagging scale factors for the CMVA algorithm are evaluated in [33].
- **trigger scale factor:** the uncertainty comes from the modeling of the trigger response in Monte Carlo simulations. The value used here is a round up of the one used in [33].
- **normalization on 4btag events** it derives from the uncertainty on the f' factor of section 3.7.
- **2-3** this systematic error refers to eq. (3.20).

3.8 RESULTS

In order to make the Combine algorithm fit the data on the two templates, all the two-dimensional histograms were cut in slices and the sliced put head to tail into some new one-dimensional histograms. The maximum likelihood fit, accounting all the systematics, yielded the following result:

$$\text{Best fit } r : 61.4^{+31.5}_{-26.8} \text{ (68\% CL)}$$
$$\text{Observed Limits } r < 117.3$$

The result is compatible with the limit $r < 56.1$ found by the ATLAS collaboration [53] in the same final channel. The fit is shown in fig. 3.22. The data points in the first 120 bins, corresponding to low BDT values, are not properly fitted but their inclusion or exclusion does not alter sensitively the result, mainly because there is no signal at such BDT values.

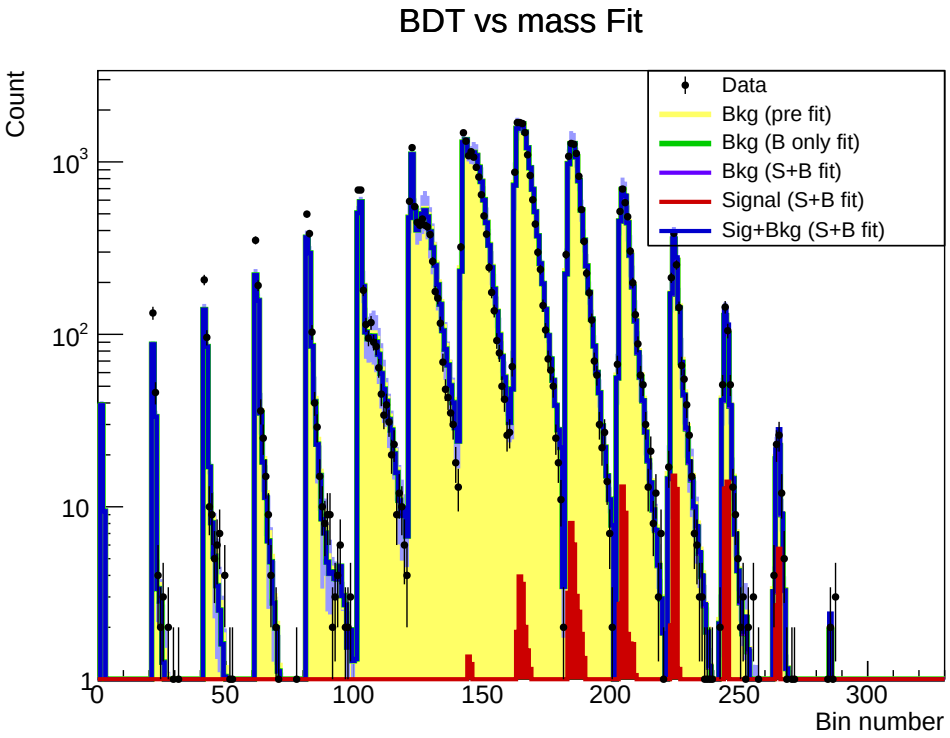


Figure 3.22: Fit result.

CHAPTER 4

BSM ANALYSIS

In this chapter we want to probe the anomalous production of Higgs boson pairs with the aim of setting a limit on the signal strength on each BSM scenario considered. We will rely on the effective theory presented in section 1.3 that introduces two parameters accounting for deviations of the Higgs tri-linear coupling (κ_λ), of the Higgs-top quark Yukawa interaction (κ_t) respectively, and three couplings not genuinely predicted by the Standard Model.
As we already pointed out the a wide variety of the kinematics arising from different values of the parameters motivated the creation of a clustering technique [7] capable of identifying parameter space regions sharing the same kinematics .
We will take full advantage of the clustering technique as it will allow us to probe a large parameter space optimizing the analysis for a handful of kinematical scenarios. From each scenario a signal strength will be extracted replicating the analysis performed in chapter 3.

4.1 PARAMETER SPACE SAMPLING

We chose to probe a two-dimensional parameter space defined by κ_λ and κ_t setting the non-SM-genuine couplings to zero. The space probed extends from 0.5 to 2.5 in κ_t and from -20 to 20 in κ_λ reflecting the known constraints. A evenly-spaced grid was defined and 54 *gen-level* samples of 20k events were generated on its nodes. The clustering algorithm identified 9 samples, out of the 54, as *benchmarks* i.e. as representatives of a parameter space region that yields a homogeneous kinematic.
The 9 regions of the parameter space are depicted in section 4.1 while fig. 4.2(b) is a proof of the kinematic variability at the *gen-level*. The benchmarks were re-generated in order to have a 300k events using the same tools of the SM sample. Let us notice that the SM point defined by $\kappa_\lambda = \kappa_t = 1$ lies in the region represented by the benchmark number 6 that corresponds to $\kappa_\lambda = 1$, $\kappa_t = 1.6$.

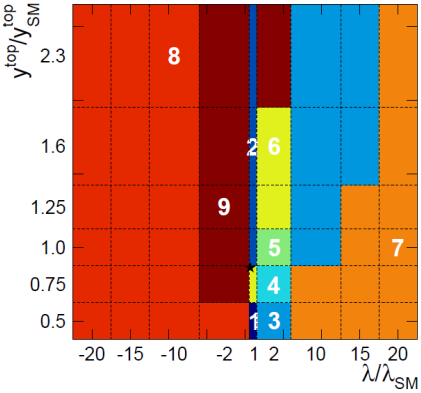
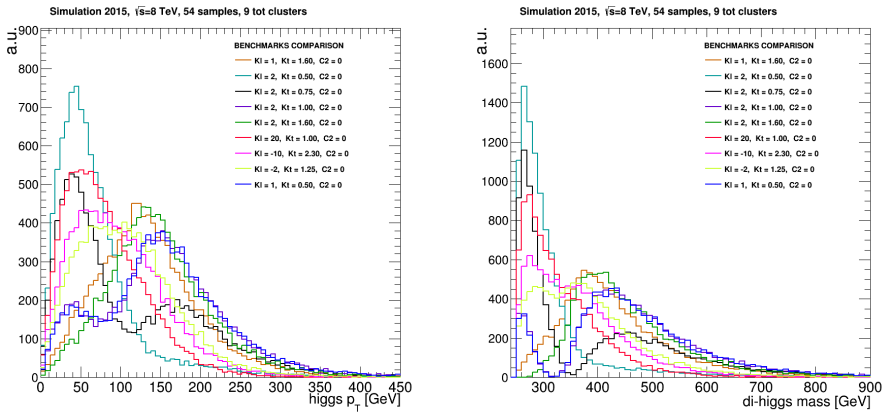


Figure 4.1: Coverage of the parameter space by benchmarks. The star pinpoints the SM values.



(a) Transverse momentum of one boson at the *gen-level*.

(b) Higgs pair invariant mass at the *gen-level*.

Figure 4.2: P_T and hh invariant mass at the gen-level as a proof of kinematic variability. Here$L$ and y are aliases for $k_\lambda$and k_t respectively.

4.2 Cross sections

The cross section of the benchmarks was calculated according to the parametrization used in [7] that, considering only κ_{lambda} and κ_t becomes

$$\sigma = \sigma_{SM} \times (A_1\kappa_t^4 + A_3\kappa_t^2\kappa_\lambda^2 + A_7\kappa_t^3\kappa_\lambda) \tag{4.1}$$

where the coefficients, at 8 TeV, are $A_1 = 2.19 \pm 0.03$, $A_2 = 0.324 \pm 0.006$, $A_7 = -1.52 \pm 0.03$. The cross section of the 9 benchmarks are listed in table 4.1.

Number	Benchmark label	κ_λ	κ_t	cross section σ (fb)	σ/σ_{SM}
1	L1y05	1	0.5	0.286 ± 0.049	0.029
2	L1y16	1	1.6	89.4 ± 8.5	9.0
3	L2y05	2	0.5	0.83 ± 0.12	0.08
4	L2y075	2	0.75	1.45 ± 0.31	0.16
5	L2y1	2	1	4.59 ± 0.78	0.46
6	L2y16	2	1.6	52.5 ± 5.7	5.27
8	Lm10y23	-10	2.3	4150 ± 386	416.2
9	*Lm2y125	-2	1.25	132 ± 12	13.3

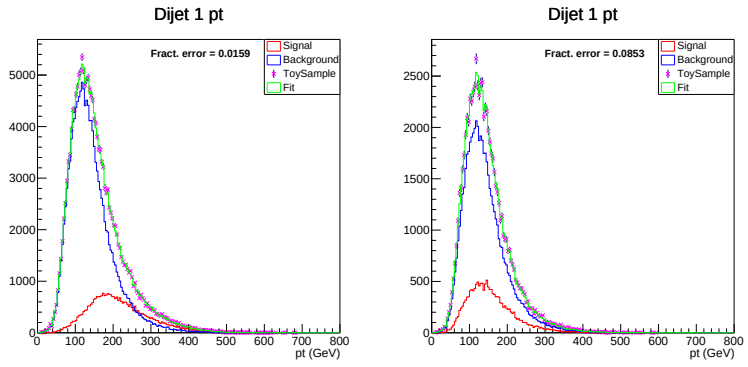
Table 4.1: Cross sections of the selected samples - the *benchmarks*.
The last sample was not analyzed further due to MC generation failure

4.3 Analysis adaptations

The repetition of the whole analysis on each benchmark required some adjustments that encompass the choice of the variables on which train the BDT, the calculation of the optimal cut on the BDT response, the definition of the *validation cut* for the ABCD method and of course the modification of the factors directly dependent on the cross section. To report these changes and the peculiarities of the outcomes a subsection is hereby dedicated to each step of the analysis. Unfortunately the analysis was performed only on 8 out of 9 benchmark due to a problem with the generation of the MC sample corresponding to the last cluster, Lm2y125.

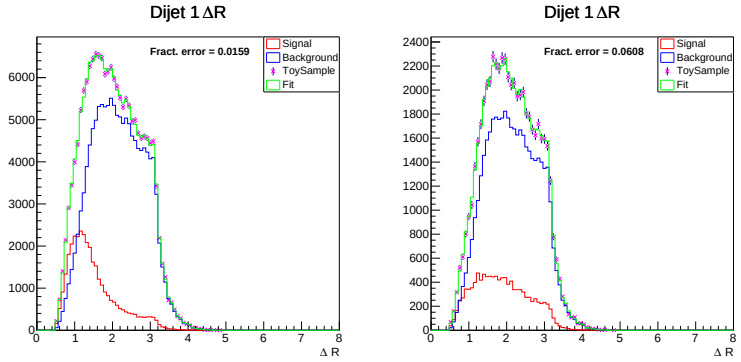
4.3.1 Variables ranking

After the pre-selection (see section 3.3) the benchmarks underwent variables ranking with no need of adjustments. The outcome showed that, as expected, not all benchmarks share the same variable ranking: in some cases (L2y05, L20y1, Lm10y13) the two leading variables of the SM case that are the dijet p_T and ΔR slip down the ranks, often (L2y05, L20y1) making room for the angular variables $cos\theta^*$ and $cos\theta_{CS}$. Figure fig. 4.3 and fig. 4.4 show an example of shape-shifting of the SM leading variables.



(a) L2y1 (b) L20y1

Figure 4.3: Transverse momentum of one dijet comparison in a significant example.



(a) L2y1 (b) L2y05

Figure 4.4: ΔR between dijets comparison in a significant example.

4.3.2 TMVA training

Surveying the ranking outcomes and on the correlations calculated by TMVA, the changes listed in table 4.2 were applied on the TMVA training with respect to the SM analysis.

Label	variable added	variable removed	notes
L1y05	1^{st}, 2^{nd}, 3^{rd} jet p_T	-	-
L1y16	-	-	-
L20y1	$cos\theta_{CS}$	hh inv. mass	$cos\theta_{CS}$ and dijets ΔR decorrelation
L2y05	$cos\theta_{CS}$, 2^{nd} jet p_T	3^{rd} jet p_T	$cos\theta_{CS}$ and ΔR between dijets decorrelation.
L2y075	-	-	-
L2y1	$cos\theta_{CS}$	-	$cos\theta_{CS}$ and dijets ΔR decorrelation.
L2y16	-	-	-
Lm10y23	$cos\theta_{CS}$	-	$cos\theta_{CS}$ and dijets ΔR decorrelation.

Table 4.2: Adjustments on the TMVA training input.

For each scenario considered a BDT was trained and applied on the benchmarks as well as to the entire CMS dataset. Then the cuts on the BDT response were determined using the same figures of merit of section 3.5.6 accounting for the different equivalent luminosities of the MC samples.

4.3.3 ABCD counting experiment

Once the validation cut was set according to the same criterion applied in section 3.6 the ABCD method yielded the results in table 4.3. To extract the limits, presented in table 5.1 together with the shape fit outcome, only two nuisance parameters had to be updated: the uncertainty on the signal cross section and the uncertainty on the predicted background events in signal region.

$\kappa_\lambda = 1, \kappa_t = 0.5$	
Events in D region:	1355
Background events expected:	1818.19 $\pm$ 198.369 (stat.) $\pm$ 90.2656 (syst.)
$\kappa_\lambda = 1, \kappa_t = 1.6$	
Events in D region:	2803
Background events expected:	3263.78 $\pm$149.101 (stat.) $\pm$ 189.853 (syst.)
$\kappa_\lambda = 20, \kappa_t = 1$	
Events in D region:	12243
Background events expected:	17374 $\pm$ 1292.83 (stat.) $\pm$ 2868.74 (syst.)
$\kappa_\lambda = 2, \kappa_t = 0.5$	
Events in D region:	11761
Background events expected:	18156.8 $\pm$ 3100.95 (stat.) $\pm$ 516.139 (syst.)
$\kappa_\lambda = 2, \kappa_t = 0.75$	
Events in D region:	1041
Background events expected:	1180.08 $\pm$ 97.3422 (stat.) $\pm$ 85.7917 (syst.)
$\kappa_\lambda = 2, \kappa_t = 1$	
Events in D region:	1089
Background events expected:	1254.76 $\pm$ 88.9982 (stat.) $\pm$ 22.5465 (syst.)
$\kappa_\lambda = 2, \kappa_t = 1.6$	
Events in D region:	1801
Background events expected:	2182.98 $\pm$ 206.725 (stat.) $\pm$ 46.5053 (syst.)
$\kappa_\lambda = -10, \kappa_t = 1.3$	
Events in D region:	7862
Background events expected:	9297.73 $\pm$ 449.361 (stat.) $\pm$ 1796.43 (syst.)

Table 4.3: Expected background in signal region.

4.3.4 Shape fit

Again we constructed the dijet-mass vs BDT-response 2D histogram for the datasets events within the signal region (i.e. with 4 b-tagged jets) and we tested it against the appropriate signal template (MC events with 4 b-tags) and background template (dataset events with 3 b-tags). Each benchmark scenario required different templates: even if the events in the background template are actually the same, the BDT-response associated to each events depends on the training that the BDT underwent. Still the procedure was completed automatically except for the the recalculation of the equivalent luminosity of each MC sample.
The nuisance parameters were left untouched with respect to the SM fit except for the cross section uncertainty. The signal and background template shapes for each scenario are reported in appendix A and the fits in appendix B. The final results are presented in table 5.1.

CHAPTER 5

FINAL RESULTS

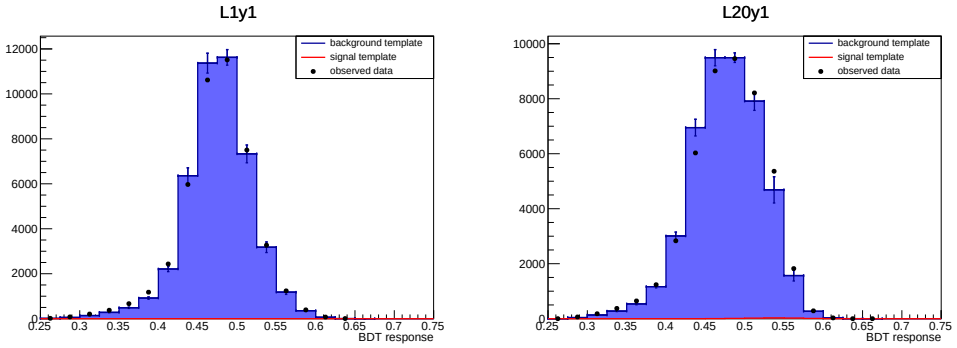
		SHAPE	SHAPE	ABCD
	σ (fb)	best fit σ (fb)	limit σ (fb)	limit σ (fb)
SM	9.96	612^{+314}_{-267}	1168	1350
† $\kappa_\lambda = 1, \kappa_t = 0.5$	0.286	167^{+208}_{-167}	567	1340
$\kappa_\lambda = 1, \kappa_t = 1.6$	89.4	1067^{+445}_{-372}	1857	2079
* $\kappa_\lambda = 20, \kappa_t = 1$	1012	8238^{+2145}_{-1741}	12053	851294
* $\kappa_\lambda = 2, \kappa_t = 0.5$	0.83	14152^{+3760}_{-2971}	20877	6204671
$\kappa_\lambda = 2, \kappa_t = 0.75$	1.45	737^{+278}_{-326}	1504	2067
$\kappa_\lambda = 2, \kappa_t = 1$	4.59	400^{+266}_{-223}	889	1076
$\kappa_\lambda = 2, \kappa_t = 1.6$	52.5	641^{+326}_{-273}	1218	1761
* $\kappa_\lambda = -10, \kappa_t = 2.3$	4150	4441^{+1245}_{-996}	6640	21197

Table 5.1: Summary of the results. The rows marked by an asterisk * or a dagger † present an unsatisfactory or unstable result respectively. A further discussion can be found in the next section.

5.1 RESULTS REVIEW

Both the ABCD method and the shape analysis have some weaknesses. The ABCD method suffers from large systematics the cause of which lies in the non-even population of the nine regions that can be figured looking at the BDT response of fig. 3.14. As a consequence the particular choice of the validation cut, for which there is not a definitive criterion, influences strongly the prediction and its error. This reason motivated the development of another strategy. The shape analysis, based again on the BDT output and the number of b-tags, follows a

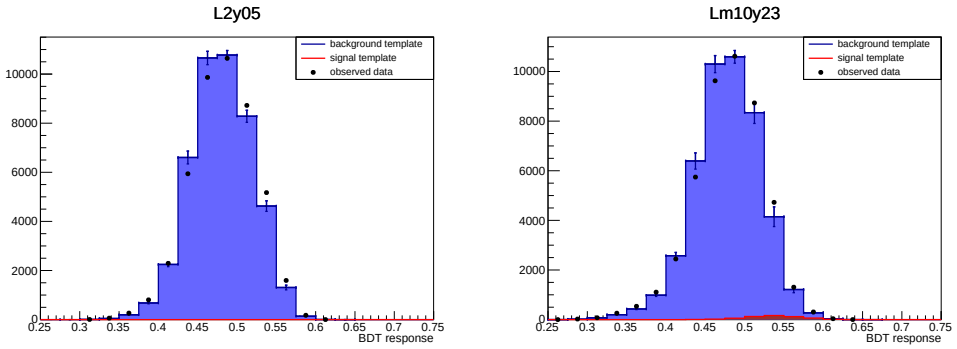
different approach and provides a valid alternative to the ABCD. The shape analyis, however, leans on a rough extrapolation of the background shape in the signal region. The limits are expected to be homogeneous within the uncertainties even if the signal templates have different cross sections, as their extraction depends mainly on the difference between the observed data and the background template. Yet, for some parameter space points, marked in table 5.1, abnormal (* mark) or unstable († mark) limits were found. A revision of the systematics did not spot any errors in the procedure but revealed that, in some cases, the observed data shape exhibits deviations from the background template larger that the background uncertainties. As a consequence high limits are obtained when the fit algorithm boosts the signal template to compensate these deviations. In fig. 5.1 and fig. 5.2 the bi-dimensional shapes corresponding to the unsatisfactory results have been projected on the BDT axis showing how the observed data is distributed with respect to the background template. We conclude that background template extrapolation is not always adequate in spite of the sound systematics estimation.
A more detailed study of the template shapes would be needed to obtain more credible results; however these are beyond the scope of the present pilot study and cannot fit within this thesis work.



(a) $\kappa_\lambda = 1, \kappa_t = 1$ SM, as reference

(b) * $\kappa_\lambda = 20, \kappa_t = 1$

Figure 5.1: Projection of the data and templates shapes on the BDT axis for the unsatisfactory results.



(a) $\kappa_\lambda = 2, \kappa_t = 0.5$

(b) $\kappa_\lambda = 10, \kappa_t = 2.3$

Figure 5.2: Projection of the data and templates shapes on the BDT axis for the unsatisfactory results.

CHAPTER 6

CONCLUSIONS

The subject of this thesis was the study of Higgs pair production, which is the only means to probe the Higgs self-coupling λ. The final state chosen, featuring four b-jets, has a tiny cross section and is overwhelmed by irreducible background. The study took advantage of statistical and multivariate methods to characterize the signal and extract both a limit and a best fit value on the signal strength. Thanks to [7] also Beyond Standard Model scenarios, accounting for anomalies in the the Higgs trilinear coupling λ and the Yukawa coupling with top quarks, have been explored.
The data processed amounts to the full dataset ($17.9 fb^{-1}$) collected by the Compact Muon Solenoid at LHC during Run I with a suitable trigger. A limit and a best fit signal strength have been extracted successfully.
As expected this analysis did not produce a radical advance in the knowledge of the topic, but sharpened the tools necessary for a state-of-the-art analysis on the upcoming 13 TeV data. This work, in fact, points out several critical points of the analysis:

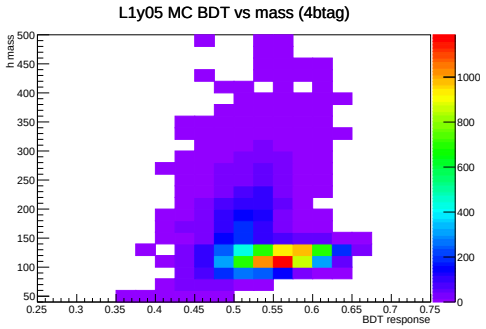
- A study on the signal pre-selection, that is the matching of the jets to the two boson, should be carried on looking at the Monte Carlo truth. The selection of the jets coming from the double Higgs decay could be based on a finer criterion than the minimization the dijets mass difference, for example it could be based on a multivariate discriminant trained on correct and purposely wrong jet matching extracted from the MC truth.

- The shape fit method could be improved extrapolating the background template shape in the signal region in a more sophisticated way, instead of a simple rescaling, for example by including a matrix reweighing based on the b-tag probability, as done in the ABCD method.

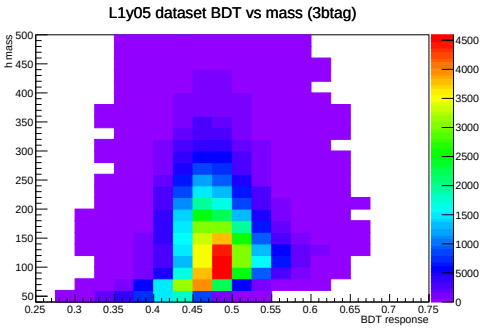
- The multivariate algorithm and the variable ranking are considered reliable and sound.
- The analysis would benefit from a background simulation and a dedicated trigger.

APPENDIX A

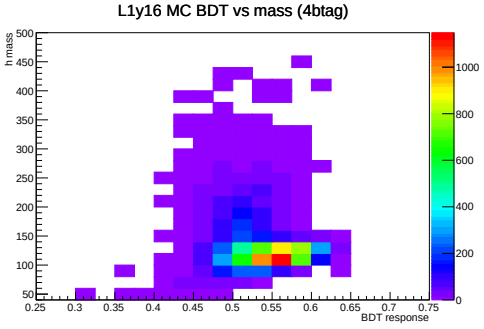
SHAPES OF THE BSM SCENARIOS



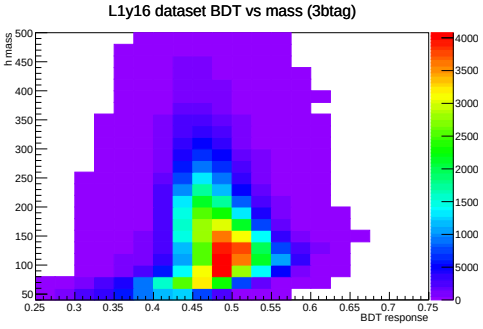
(a) $\kappa_\lambda = 1, \kappa_t = 0.5$ signal template



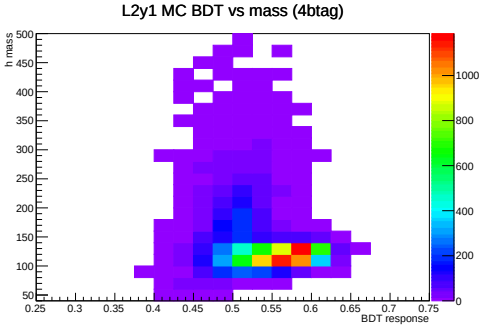
(b) $\kappa_\lambda = 1, \kappa_t = 0.5$ background template



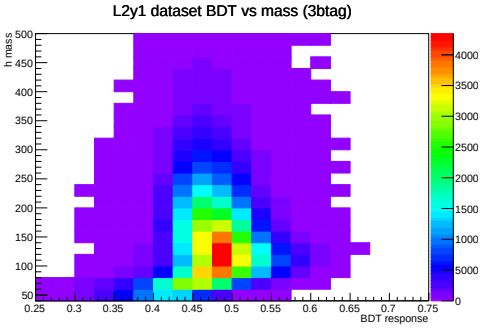
(c) $\kappa_\lambda = 1, \kappa_t = 1.6$ signal template



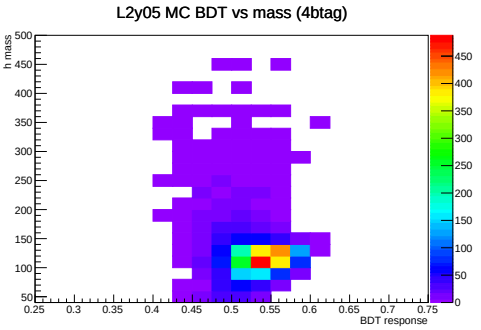
(d) $\kappa_\lambda = 1, \kappa_t = 1.6$ background template



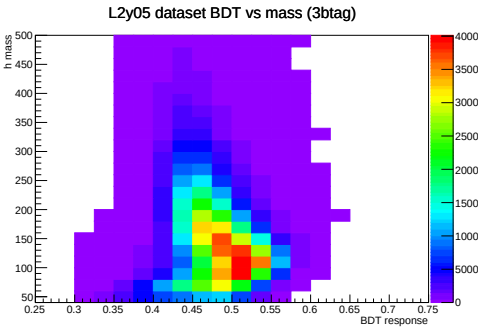
(e) $\kappa_\lambda = 2, \kappa_t = 1$ signal template



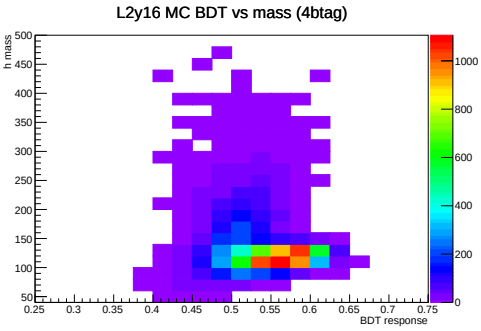
(f) $\kappa_\lambda = 2, \kappa_t = 1$ background template



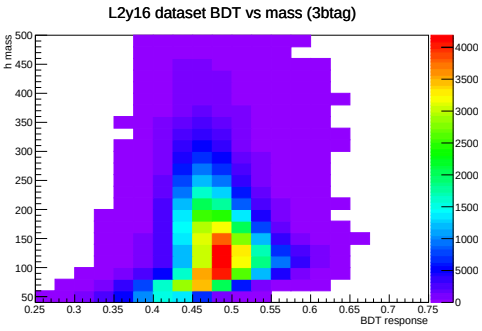
(g) $\kappa_\lambda = 2, \kappa_t = 0.5$ signal template



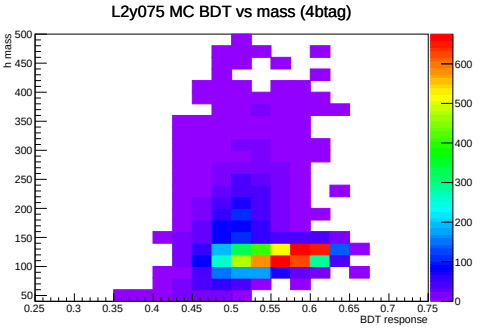
(h) $\kappa_\lambda = 2, \kappa_t = 0.5$ background template



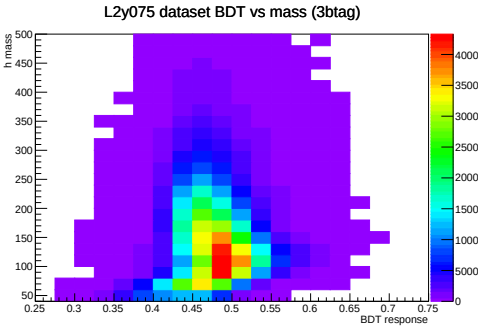
(i) $\kappa_\lambda = 2, \kappa_t = 1.6$ signal template



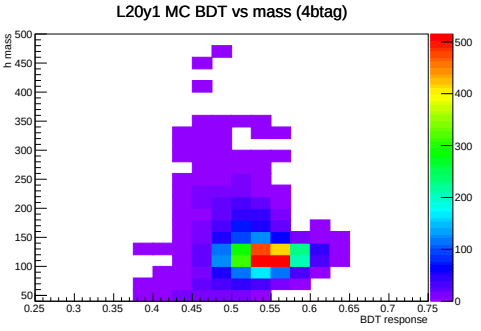
(j) $\kappa_\lambda = 2, \kappa_t = 1.6$ background template



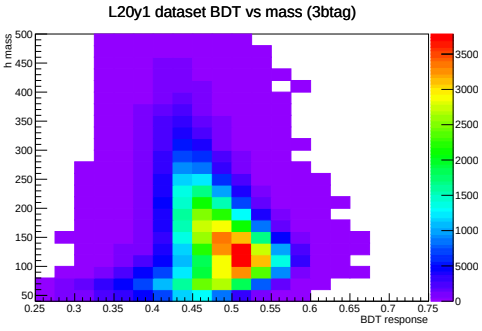
(k) $\kappa_\lambda = 2, \kappa_t = 0.75$ signal template



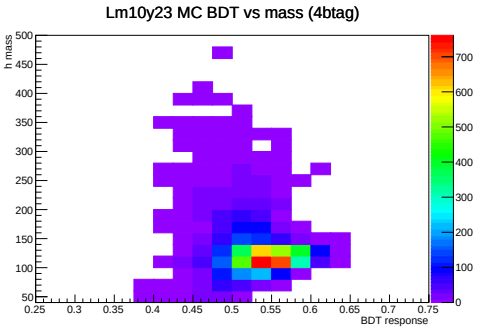
(l) $\kappa_\lambda = 2, \kappa_t = 0.75$ background template



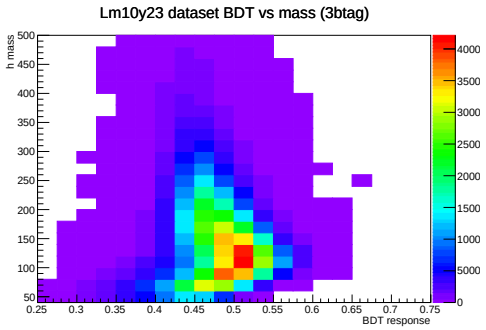
(m) $\kappa_\lambda = 20, \kappa_t = 1$ signal template



(n) $[\kappa_\lambda = 20, \kappa_t = 1$ background template



(o) $\kappa_\lambda = -10, \kappa_t = 1.3$ signal template

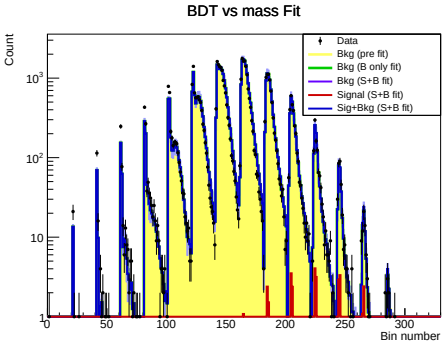


(p) $\kappa_\lambda = -10, \kappa_t = 1.3$ background template

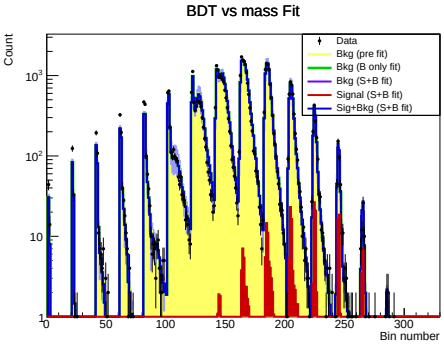
Figure A.0: Dijet mass vs BDT response shape.

APPENDIX B

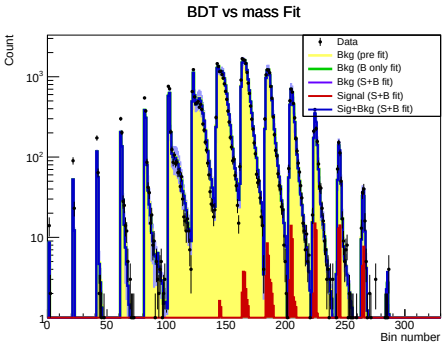
FIT SHAPE RESULTS



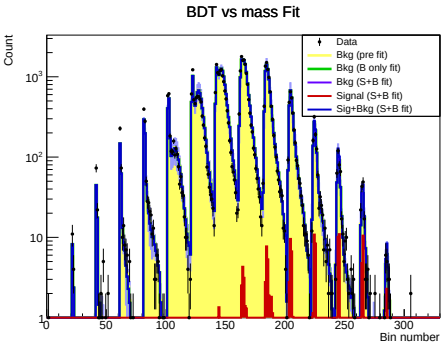
(q) $\kappa_\lambda = 1, \kappa_t = 0.5$



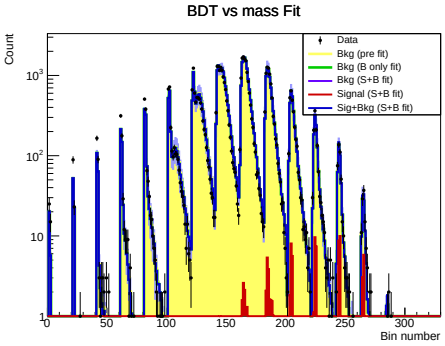
(r) $\kappa_\lambda = 1, \kappa_t = 1.6$



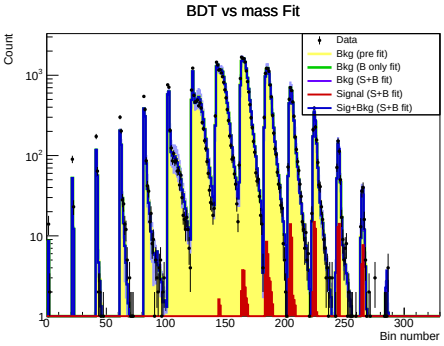
(s) $\kappa_\lambda = 2, \kappa_t = 0.5$



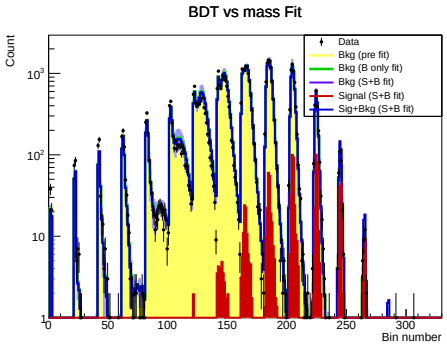
(t) $\kappa_\lambda = 2, \kappa_t = 0.75$



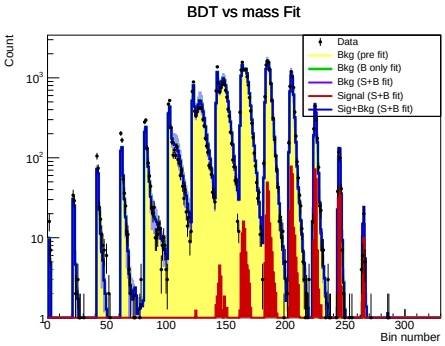
(u) $\kappa_\lambda = 2, \kappa_t = 1$



(v) $\kappa_\lambda = 2, \kappa_t = 1.6$



(w) $\kappa_\lambda = 20, \kappa_t = 1$



(x) $\kappa_\lambda = -10, \kappa_t = 2.3$

Figure B.0: Fit results

BIBLIOGRAPHY

[1] S. Dawson, A. Ismail, and Ian Low. "What's in the loop? The anatomy of double Higgs production". In: *Phys. Rev. D* 91 (11 June 2015), p. 115008. DOI: 10.1103/PhysRevD.91.115008. URL: http://link.aps.org/doi/10.1103/PhysRevD.91.115008 (cit. on p. 1).

[2] Tilman Plehn and Michael Rauch. "Quartic Higgs coupling at hadron colliders". In: *Phys. Rev. D* 72 (5 2005) (cit. on p. 1).

[3] D. de Florian and J. Mazzitelli. "Higgs Boson Pair Production at Next-to-Next-to-Leading Order in QCD". In: *Phys. Rev. Lett. 11* 111 (2013), pp. 161–168. DOI: 10.1103/PhysRevLett.111.201801. arXiv: arXiv:1309.6594 (cit. on p. 1).

[4] D. de Florian and J. Mazzitelli. "Next-to-Next-to-Leading Order QCD Corrections to Higgs Boson Pair Production". In: *PoS* LL2014 (2014), p. 029. DOI: 10.1103/PhysRevLett.111.201801. arXiv: 1405.4704 (cit. on p. 1).

[5] Michael E. Peskin and Daniel V. Schroeder. *An Introduction To Quantum Field Theory*. Westview Press; First Edition edition, 1995 (cit. on p. 1).

[6] LHC Higgs Cross Section Working Group. *Handbook of LHC Higgs Cross Sections: 2. Differential Distributions*. Tech. rep. arXiv:1201.3084. CERN-2012-002. Geneva, 2012. URL: http://cds.cern.ch/record/1416519 (cit. on p. 5).

[7] Martino Dall'Osso et al. "Higgs Pair Production: Choosing Benchmarks With Cluster Analysis". In: (2015). arXiv: 1507.02245 [hep-ph] (cit. on pp. 6, 41, 42, 48).

[8] Vardan Khachatryan et al. "Precise determination of the mass of the Higgs boson and tests of compatibility of its couplings with the standard model predictions using proton collisions at 7 and 8 TeV". In: (2014). arXiv: 1412.8662 [hep-ex] (cit. on p. 6).

[9] ATLAS Collaboration. *Measurements of the Higgs boson production and decay rates and coupling strengths using pp collision data at $\sqrt{s}$ = 7 and 8 TeV in the ATLAS experiment*. Tech. rep. ATLAS-CONF-2015-007. Geneva: CERN, Mar. 2015. URL: http://cds.cern.ch/record/2002212 (cit. on p. 6).

[10] P. Artoisenet et al. "A framework for Higgs characterisation". In: *JHEP* 1311 (2013), p. 043. DOI: 10.1007/JHEP11(2013)043. arXiv: 1306.6464 [hep-ph] (cit. on p. 6).

[11] Beranger Dumont, Sylvain Fichet, and Gero von Gersdorff. "A Bayesian view of the Higgs sector with higher dimensional operators". In: *Journal of High Energy Physics* 2013.7, 65 (2013). DOI: 10.1007/JHEP07(2013)065. URL: http://dx.doi.org/10.1007/JHEP07(2013)065 (cit. on p. 6).

[12] J. Elias-Miro' et al. "Higgs windows to new physics through $d = 6$ operators: constraints and one-loop anomalous dimensions". In: *Journal of High Energy Physics* 2013.11, 66 (2013). DOI: 10.1007/JHEP11(2013)066. URL: http://dx.doi.org/10.1007/JHEP11(2013)066 (cit. on p. 6).

[13] Rick S. Gupta, Alex Pomarol, and Francesco Riva. "BSM Primary Effects". In: (2014). arXiv: 1405.0181 [hep-ph] (cit. on p. 6).

[14] G. Aad et al. "Search For Higgs Boson Pair Production in the $\gamma\gamma b\bar{b}$ Final State using pp Collision Data at $\sqrt{s} = 8$ TeV from the ATLAS Detector". In: *Phys.Rev.Lett.* 114.8 (2015), p. 081802. DOI: 10.1103/PhysRevLett.114.081802. arXiv: 1406.5053 [hep-ex] (cit. on p. 6).

[15] Georges Aad et al. "Search for Higgs boson pair production in the $b\bar{b}b\bar{b}$ final state from pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector". In: (2015). arXiv: 1506.00285 [hep-ex] (cit. on p. 6).

[16] S. Chatrchyan et al. *The CMS experiment at the CERN LHC*. Tech. rep. 08. 2008, S08004. URL: http://stacks.iop.org/1748-0221/3/i=08/a=S08004 (cit. on p. 7).

[17] CMS Collaboration. *CMS-ECAL Technical Design Report*. CMS Technical Design Report CMS-ECAL-TDR-4. 1997 (cit. on p. 9).

[18] The CMS Collaboration. "The CMS Experiment at the CERN LHC". In: *JINST* 3 (S08004 2008) (cit. on p. 9).

[19] CMS Collaboration. *The CMS muon project: Technical Design Report*. Technical Design Report. 1997. URL: https://cms-physics.web.cern.ch/cms-physics/public/PFT-09-001-pas.pdf (cit. on p. 10).

[20] Sergio Cittolin, Attila Rácz, and Paris Sphicas. *CMS The TriDAS Project: Technical Design Report, Volume 2: Data Acquisition and High-Level Trigger. CMS trigger and data-acquisition project*. Technical Design Report CMS. Geneva: CERN, 2002. URL: https://cds.cern.ch/record/578006 (cit. on p. 11).

[21] Florian Beaudette. "The CMS Particle Flow Algorithm". In: *Proceedings, International Conference on Calorimetry for the High Energy Frontier (CHEF 2013)*. 2014, pp. 295–304. arXiv: 1401.8155 [hep-ex]. URL: https://inspirehep.net/record/1279774/files/arXiv:1401.8155.pdf (cit. on p. 12).

[22] CMS Collaboration. *ParticleFlow Event Reconstruction in CMS and Performance for Jets, Taus, and E_T^{miss}*. CMS Physics Analysis Summary CMS-PAS-PFT-09-001. 2009. URL: https://cms-physics.web.cern.ch/cms-physics/public/PFT-09-001-pas.pdf (cit. on p. 12).

[23] B. R. Webber. "Fragmentation and hadronization". In: *Int. J. Mod. Phys.* A15S1 (2000). [,577(1999)], pp. 577–606. DOI: 10.1142/S0217751X00005334. arXiv: hep-ph/9912292 [hep-ph] (cit. on p. 13).

[24] K.A. Olive et al. "Review of Particle Physics". In: *Chin.Phys.* C38 (2014), p. 090001. DOI: 10.1088/1674-1137/38/9/090001 (cit. on p. 13).

[25] M. Cacciari, G. Salam, and G. Soyez. "Dispelling the myth for the jet-finder". In: *Physics Letter B* 641 (1 2006), pp. 57–61. DOI: http://dx.doi.org/10.1016/j.physletb.2006.08.037. (cit. on p. 14).

[26] M. Cacciari and G. Salam. "The anti-kt clustering algorithm". In: *Journal of High Energy Physics* 4 (63 2008) (cit. on p. 14).

[27] The CMS collaboration. "Identification of b-quark jets with the CMS experiment". In: *JINST* 8 (P04013 2013). DOI: 10.1088/1748-0221/8/04/P04013 (cit. on p. 16).

[28] V. Khachatryan and al. "Measurement of $B\bar{B}$ angular correlations based on secondary vertex reconstruction at $\sqrt{s} = 7$ TeV". In: *Journal of High Energy Physics* 2011.3, 136 (2011). DOI: 10.1007/JHEP03(2011)136. URL: http://dx.doi.org/10.1007/JHEP03(2011)136 (cit. on pp. 16, 18).

[29] Wolfgang Waltenberger. *Adaptive Vertex Reconstruction*. Tech. rep. CMS-NOTE-2008-033. Geneva: CERN, July 2008. URL: http://cds.cern.ch/record/1166320 (cit. on p. 16).

[30] R Frühwirth, Wolfgang Waltenberger, and Pascal Vanlaer. *Adaptive Vertex Fitting*. Tech. rep. CMS-NOTE-2007-008. Geneva: CERN, Mar. 2007. URL: http://cds.cern.ch/record/1027031 (cit. on p. 16).

[31] C. Weiser. *A Combined Secondary Vertex Based B-Tagging Algorithm in CMS*. CMS Note 2006/014. 2006 (cit. on pp. 16, 17, 20).

[32] CMS Collaboration. "Search for di-Higgs resonances decaying to 4 b-jets in pp collisions at 8 TeVs". In: (2013) (cit. on p. 18).

[33] Suvik Das and Jacobo Konigsberg et al. *Search for di-Higgs resonances decaying to 4 b-jets in pp collisions at 8 TeV*. CMS Note 2013/227 (cit. on pp. 18, 19, 37, 39).

[34] Johan Alwall et al. "MadGraph 5: going beyond". In: *Journal of High Energy Physics* 2011.6, 128 (2011). DOI: 10.1007/JHEP06(2011)128. URL: http://dx.doi.org/10.1007/JHEP06(2011)128 (cit. on p. 19).

[35] Jun Gao et al. "CT10 next-to-next-to-leading order global analysis of QCD". In: *Phys. Rev. D* 89 (3 Feb. 2014), p. 033009. DOI: 10.1103/PhysRevD.89.033009. URL: http://link.aps.org/doi/10.1103/PhysRevD.89.033009 (cit. on p. 19).

[36] S. Agostinelli et al. "GEANT4: A Simulation toolkit". In: *Nucl. Instrum. Meth.* A506 (2003), pp. 250–303. DOI: 10.1016/S0168-9002(03)01368-8 (cit. on p. 19).

[37] Daniel de Florian and Javier Mazzitelli. "Higgs Boson Pair Production at Next-to-Next-to-Leading Order in QCD". In: *Phys. Rev. Lett.* 111 (20 Nov. 2013), p. 201801. DOI: 10.1103/PhysRevLett.111.201801. URL: http://link.aps.org/doi/10.1103/PhysRevLett.111.201801 (cit. on p. 19).

[38] John C. Collins and Davison E. Soper. "Angular distribution of dileptons in high-energy collisions". In: *Physical Review D* 16 (1977), pp. 2219–2225 (cit. on p. 22).

[39] R. J. Barlow. *A Guide to the Use of Statistical Methods in the Physical Sciences*. John Wiley & Sons Ltd., 1989 (cit. on p. 23).

[40] A. N. Pettitt. "A two sample Anderson-Darling rank statistic". In: *Biometrika* 63 (1976), pp. 161–168 (cit. on p. 23).

[41] Andreas Hoecker et al. "TMVA: Toolkit for Multivariate Data Analysis". In: *PoS* ACAT (2007), p. 040. arXiv: physics/0703039 (cit. on p. 28).

[42] Rene Brun and Fons Rademakers. "ROOT – An object oriented data analysis framework". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 389.1–2 (1997), pp. 81–86. DOI: http://dx.doi.org/10.1016/S0168-9002(97)00048-X (cit. on p. 28).

[43] Luca Lista. *Statistical Methods for Data Analysis in Particle Physics*. Springer, 2016 (cit. on p. 28).

[44] A. Hoecker et al. *TMVA - Toolkit for Multivariate Data Analysis*. PoS ACAT:040,2007. 2007. eprint: arXiv:physics/0703039 (cit. on pp. 28, 32).

[45] Ilya Narsky and Frank C. Porter. *Statistical Analysis Techniques in Particle Physics*. Wiley-VCH, 2013. ISBN: 978-3-527-41086-6 (cit. on pp. 29, 30).

[46] Yoav Freund and Robert E Schapire. "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting". In: *Journal of Computer and System Sciences* 55.1 (1997), pp. 119–139. ISSN: 0022-0000. DOI: http://dx.doi.org/10.1006/jcss.1997.1504. URL: http://www.sciencedirect.com/science/article/pii/S002200009791504X (cit. on p. 30).

[47] Valeria Bartsch and Gunter Quast. *Expected Signal Observability at Future Experiments*. Tech. rep. CMS-NOTE-2005-004. Geneva: CERN, Feb. 2005. URL: http://cds.cern.ch/record/824351 (cit. on p. 33).

[48] S. I. Bityukov and N. V. Krasnikov. "New Physics Discovery Potential in Future Experiments". In: *Mod. Phys. Rev. Lett. A* 13 (40 1998), pp. 3225–3249 (cit. on p. 33).

[49] Robert Cousins, Jason Mumford, and Vyacheslav Valuev. "Detection of Z' Gauge Bosons in the Dimuon Decay Mode in CMS". In: (2005) (cit. on p. 33).

[50] Joel Heinrich and Louis Lyons. "Systematic Errors". In: *Annu. Rev. Nucl. Part. Sci.* 57.1 (2007), pp. 145–169. DOI: 10.1146/annurev.nucl.57.090506.123052. URL: http://dx.doi.org/10.1146/annurev.nucl.57.090506.123052 (cit. on p. 35).

[51] "Combine Tool". In: *CMS twiki* (). URL: https://twiki.cern.ch/twiki/bin/viewauth/CMS/SWGuideHiggsAnalysisCombinedLimit (cit. on p. 37).

[52] Glen Cowan et al. "Asymptotic formulae for likelihood-based tests of new physics". In: *Eur. Phys. J.* C71 (2011), p. 1554. arXiv: 1007.1727 (cit. on p. 37).

[53] Georges Aad et al. "Search for Higgs boson pair production in the $b\bar{b}b\bar{b}$ final state from pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector". In: (2015). arXiv: 1506.00285 [hep-ex] (cit. on p. 40).

[54] T. Plehn U. Baur and D. Rainwater. "Probing the Higgs self-coupling at hadron colliders using rare decays". In: *Phys. Rev. D* 9 (5 2004).