

Study of dimuon channel with the first Run2 data at CMS

Miriam Olmi, Università degli Studi di Firenze, Italy

Supervisor Elisabetta Gallo, DESY and University of Hamburg, Germany

Abstract

This report describes the analysis of the first Run2 data at the LHC, collected with the CMS detector and corresponding to an integrated luminosity of 42 pb^{-1} . The events are selected requiring two high transverse momentum muons and the $Z \rightarrow \mu\mu$ decay is used as "standard candle" process and compared to Drell-Yann Monte Carlo samples. Backgrounds to this process are diboson production, the $W + jets$ and $t\bar{t} + jets$ processes and QCD processes, here estimated with a data-driven method. The MC samples were reweighted for pile-up and an efficiency correction was applied to the muons, derived with a "tag and probe" method. After all these corrections, a very good agreement between data and MC is obtained for all the distributions except for the missing transverse energy as not all the corrections have been yet implemented in this variable. This work has allowed to commission the muon objects to be used in the Higgs analyses of the DESY CMS group.

Contents

1	Introduction	3
1.1	The CMS experiment at LHC	4
2	Data Sample	5
3	Monte Carlo Sample	5
4	Data selection	7
4.1	Preselection	8
4.2	$Z \rightarrow \mu\mu$ selection	9
5	Control plots	9
6	Efficiency	13
6.1	Tag and probe method	13
6.2	Efficiency calculation	13
7	Conclusions	18

1 Introduction

The Large Hadron Collider (LHC) is the world's largest and most powerful particle accelerator. It first started on the 10th of September 2008, and remains the latest addition to CERN's accelerator complex. The LHC consists of a 27-kilometre ring of superconducting magnets with a number of accelerating structures to boost the energy of the particles along the way. During the LHC first run, protons collided at energies of up to 8 teraelectronvolts (TeV) in bunches that were spaced apart by 50 ns. After a 2-year technical stop that prepared the machine for running at almost double the energy of the LHC first run, on the 3rd June 2015 the second run started with a center-of-mass energy of 13 TeV and an initial bunch spacing of 50 ns.

At the LHC and in pp collisions in general a wide range of physics processes is accessible. The high centre-of-mass energy and high luminosity allows to explore a completely new regime. At LHC protons travel at nearly the speed of light and this energy allows them to overcome their mutual electromagnetic repulsion so that quarks and gluons of different protons can interact via the strong nuclear force. When quarks of different protons become close, they start to feel each other through gluons and the new formed gluons go under tension when the colliding protons go away from each other. At some point the tension becomes so strong that new particles production is energetically favored. The high energy allowed the discovery of the Higgs boson and opens a new regime for searches of new physics.

Six experiments are installed on the LHC: two large and multipurpose experiments, ATLAS (A Toroidal LHC ApparatuS) and CMS (Compact Muon Solenoid); two smaller and customized experiments, LHCb for b-physics and the study of CP violation and ALICE (A Large Ion Collider Experiment) for heavy ion physics and the study of the quark gluon plasma; and two more experiments designed to study collisions in which protons or other new particles emerge at small angles with respect to the beam-pipe, LHCf (Large Hadron Collider forward) situated along the LHC beam-line in the ATLAS forward region and TOTEM (TOTAl Elastic and diffractive cross section Measurement) placed in the CMS forward region.

The current report describes the analysis of the $Z \rightarrow \mu\mu$ decay channel with the CMS experiment. This analysis was performed with data collected during the first months of the second run at the LHC corresponding to a total integrated luminosity of 42 pb⁻¹ and a bunch separation of 50 ns. The analysis of this dimuon channel is part of the preliminary study of the $Z \rightarrow \tau\tau \rightarrow \mu\mu$ process. The measurement of $Z \rightarrow \mu\mu$ events is important as it is one of the most commonly used process in order to calibrate the detector. Moreover these events constitute one of the major source of irreducible background to the search of the Higgs decay into τ leptons. The $H \rightarrow \tau\tau \rightarrow \mu\mu$ process has as signature two high transverse momentum (p_T) muons and missing transverse energy (MET). It is therefore clear that establishing a signal in the $Z \rightarrow \tau\tau \rightarrow \mu\mu$ channel is a necessary step before attempting to search for the Higgs signal in the same final state.

Another result included in this report is the muon identification and isolation selection

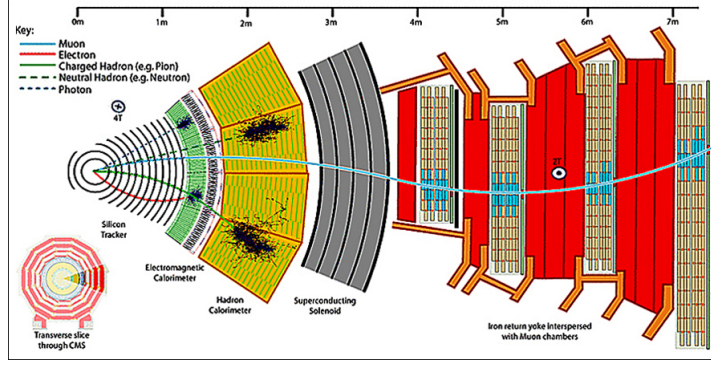


Figure 1: This illustration is a slice of the CMS detector from the collision point to the perimeter as shown in the lower left hand corner. The various colored curved lines represent paths that a particle might take, for example, the continuous blue line represents a muon.

efficiency. This efficiency, combined with the electron one, is essential in order to completely characterize the Z leptonic decay channel.

1.1 The CMS experiment at LHC

The Compact Muon Solenoid (CMS) has a broad physics program ranging from studying the Standard Model (including the Higgs boson) to searching for extra dimensions and particles that could make up dark matter. The CMS detector is 21.6 metres long, 15 metres in diameter, and weighs about 14,000 tonnes. An image of the CMS detector is shown in figure 1.

The CMS detector, as the name itself gives away, is built around a huge solenoid magnet that takes the form of a cylindrical coil of superconducting cable that generates a field of 3.8 Tesla. The innermost layer is a silicon-based tracker. Surrounding there is a scintillating crystal electromagnetic calorimeter, which is itself surrounded by a sampling calorimeter for hadrons. Both the tracker and the calorimetry are compact enough to fit inside the solenoid. Outside the magnet there are the large muon detectors, which are inside the return yoke of the magnet. The strong magnetic field combined with the high-precision tracking system allows precise momentum measurement even for very high energy particles as electrons, muons, and other products of the collisions [3].

2 Data Sample

As previously mentioned, the analysis presented in the following pages was performed with data from the early LHC Run2 with a luminosity of 42 pb^{-1} and a bunch separation of 50 ns. In particular the data sample used in this analysis is the “SingleMuon” sample with the following run range: 251244 – 251883.

This data sample has been collected with the High Level Trigger (HLT) “IsoMu24_eta2p1_v”, which is based on the request of at least one muon with a p_T higher than 24 GeV and within the $|\eta| = 2.1$ acceptance of the muon detector.

3 Monte Carlo Sample

As presented in the introduction, the process that we want to study is the $Z \rightarrow \mu\mu$ decay. As this is a Drell-Yan process [1], in order to simulate our signal we used a Drell-Yan Monte Carlo (MC) sample. In figure 2 the diagram of a generic Drell-Yan process is shown.

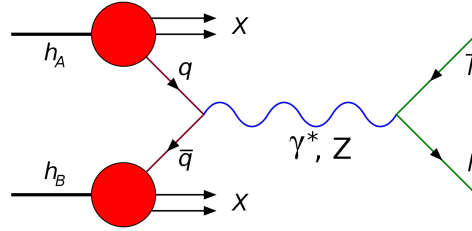


Figure 2: Drell-Yan process. The two red balls represent the interacting hadrons. Two quarks, one from each hadron, represented by the purple arrows interact and a virtual photon or a Z boson, represented by the blue wavy line, is created. This boson then produces a pair lepton-antilepton, represented by the green arrows.

Focusing on the criteria used to collect the data sample, we can see that the possible background consists of all the processes that have at least one of two high p_T muons and may be accompanied by jets and MET. Because of this we have to consider the following background MC samples:

- Diboson production: WW, WZ, ZZ
- $W + jets$
- $t\bar{t} + jets$

In table 1 all the MC samples with the respective theoretical cross section are shown.

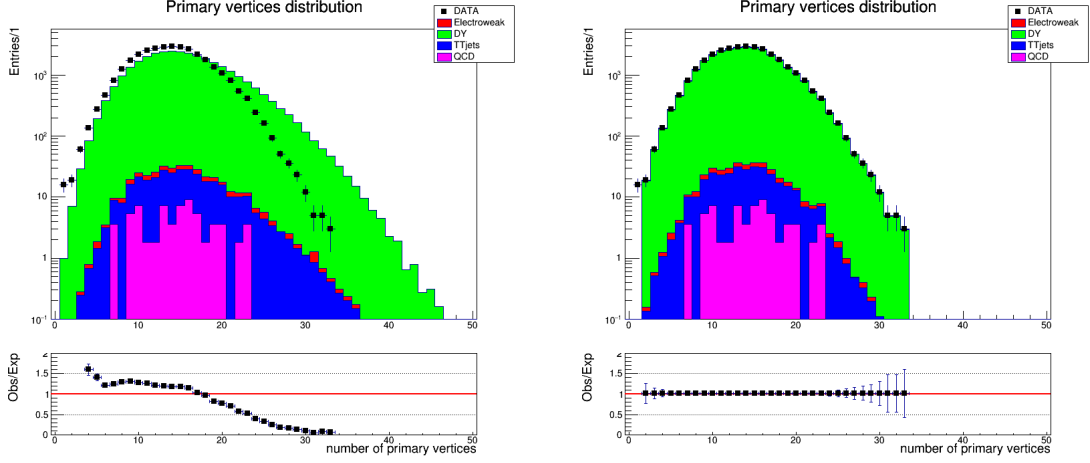
Table 1: MC sample and their relative theoretical cross section.

MC sample	cross section (pb)
Drell-Yan Mass=10-50 GeV	18610
Drell-Yan Mass < 50 GeV	6025.2
WW	63.21
WZ	22.82
ZZ	10.32
$W + jets$	61526.7
$t\bar{t} + jets$	831.76

Another possible background is the QCD contribution. However in this case we could not use MC samples because of statistical and simulation problems. Therefore we decided to obtain a QCD sample with a data-driven method, requiring events with two leptons with same charge and rescaling them by a factor of 1.8 as obtained in Run1 data [3].

Figure 3a shows the distribution of the number of primary vertices both for the data and the MC samples used in this analysis. In data the primary vertices are reconstructed using the so-called Deterministic Annealing (DA) clustering of tracks and the requirement on the reconstructed vertices are listed in the following section. From all the vertices that pass all these requirements, the vertex with the maximum value of the square of the sum of the tracks p_T associated with the vertex is chosen as the hard interaction vertex. The initial guess in the MC generator on the number of primary vertices is not so precise, so one of the correction needed in this analysis is the so called “pile-up reweighting”¹. This correction consists of weighting all the MC distributions so that the MC primary vertices distribution matches with the one in the data. In figure 3b the primary vertices distribution for both the data and the MC sample is shown after the “pile-up” correction.

¹There are a lot of weights that can be used to fill all the MC histograms. In this work we used only the pile-up weight and the genweight, which is the weight used by the generator to simulate the MC sample, to normalize the MC samples to the data luminosity. The MC histogram shown in figure 3a is filled only with the genweight. The MC histogram in figure 3b is filled with the following weight: $weight = (genweight) * (pileup\ weight)$.



(a) Primary vertices distribution before the pile-up reweighting. (b) Primary vertices distribution after the pile-up reweighting.

Figure 3: Primary vertices distribution before and after the pile-up reweighting. The full points represent the selected data and the error bars represent the statistical errors. The full histograms represent the stacked MC samples weighted to the data luminosity. In the images, the lower plots show the ratio data/MC.

4 Data selection

In order to select only the events of interest, we applied several cuts based on the following quantities:

- transverse momentum $p_T = \sqrt{p_x^2 + p_y^2}$;
- pseudorapidity $\eta = -\log(\tan(\theta/2))$, where θ is the polar angle with respect to the beam line;
- distance d_{xy} between the primary vertex and the muon in the transverse plane;
- distance dz between the primary vertex and the muon along the beam line;
- relative isolation of the muon iso ;
- *isMedium*, a flag that indicates a “good” muon;
- $nDoF$, the number of degrees of freedom of the fit used to reconstruct the primary vertex;
- the spread Z_{Vertex} of the colliding particles distribution along the beam line;
- the spread D_{Vertex} of the colliding particles distribution in the transverse plane;
- the distance ΔR between two tracks in the detector given by the formula $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$.

4.1 Preselection

This first selection is applied to select good muons:

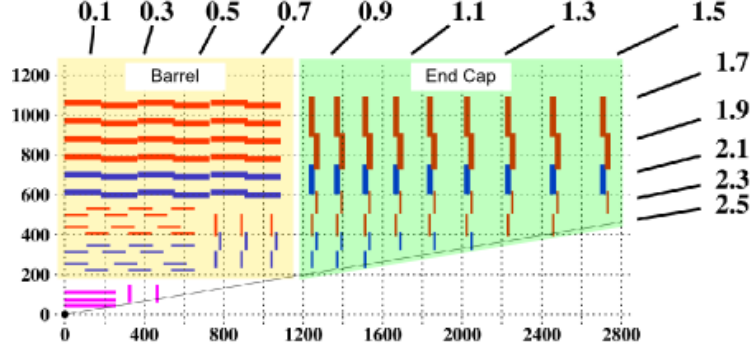


Figure 4: The CMS tracker consists of 15148 detector modules arranged in a so-called barrel and endcap configuration. A sketch of the detector layout is shown in the figure to the left. The red and blue lines indicate the position and orientation of the detector modules in the r-z plane where r is the distance to the beam axes and z is the distance to the IP along the beam axes.

- $p_T > 10 \text{ GeV}$: This cut is necessary to eliminate most of the QCD background and is related to the trigger used in the data;
- $|\eta| < 2.4$: This cut is needed to take into account the acceptance of the tracker (see figure 4);
- $|dxy| < 0.045 \text{ cm}$: This cut requires muons close to the primary vertex in the transverse plane;
- $|dz| < 0.2 \text{ cm}$: This cut requires muons close to the primary vertex along the beam line;
- $iso < 0.3$: This cut requires the muons to be isolated;
- $isMedium = true$: This selection requires good identified muons;
- $nDoF > 4 \wedge |Z_{Vertex}| < 25 \wedge |D_{Vertex}| < 2$: These are the “*quality cuts*” on the primary vertex reconstruction.

Note that we selected only events close to the primary vertex because we know that the Z boson has a half-life of about $3 \times 10^{-25} \text{ s}$ and so it is produced very close to the primary vertex.

4.2 $Z \rightarrow \mu\mu$ selection

This second selection is applied only to those muons that passed the first selection. Through this more specific selection we can consider now only the pair of muons that fired the trigger and that satisfied some additional topological features. The following cuts are applied:

- $|\Delta R_{\mu-trigger}| < 0.5$: This selection is to require a matching between the muon and the trigger object that fired the “IsoMuonTrigger”;
- *opposite charge = true* : This selection is to select only pairs of opposite charge muons as the Z boson is electrically neutral;
- $p_T > 25 \text{ GeV}$: This selection is applied to both the muons and it is needed to take into account the trigger used to produce the data sample (see section 2);
- $|\eta| < 2.1$: This selection is needed because the muon trigger is up to $|\eta| = 2.1$;
- $|\Delta R_{\mu_1-\mu_2}| > 0.5$: This selection is needed to require the two muons to be almost back to back or at least well separated².

In table 2 we can see how all these selections affect the number of events in each sample.

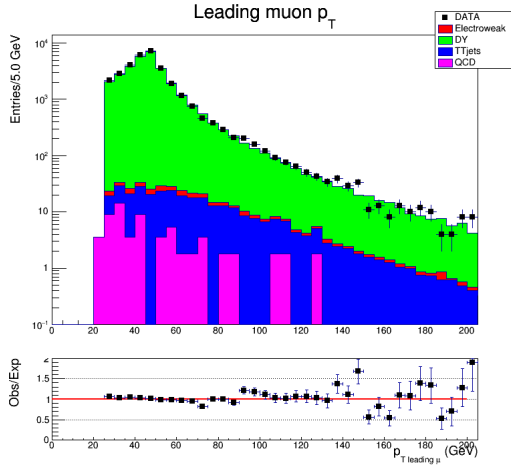
Table 2: Number of muon pairs in each sample before and after the selection. Note that the following numbers are weighted to the data luminosity.

type	Sample	Initial muon pairs	Selected muon pairs
data	SingleMuon	49162	32146
signal	Drell-Yan 10-50	1585	908
	Drell-Yan > 50	34774	30169
background	WW	24	13
	WZ	22	16
	ZZ	18	14
	$W + jets$	3774	14
	$t\bar{t} + jets$	1035	246

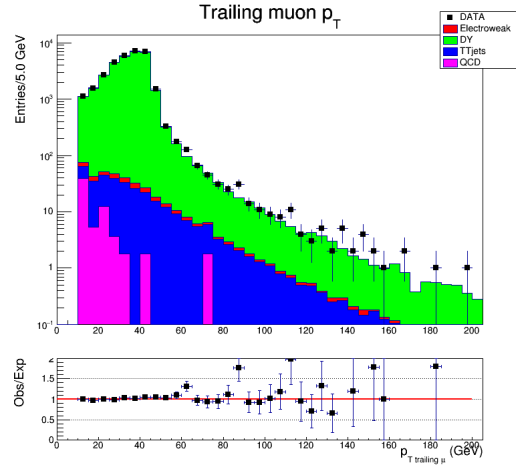
5 Control plots

After preselection and selection control plots for data and MC comparison are obtained. All MC samples are reweighted to the data luminosity and pileup.

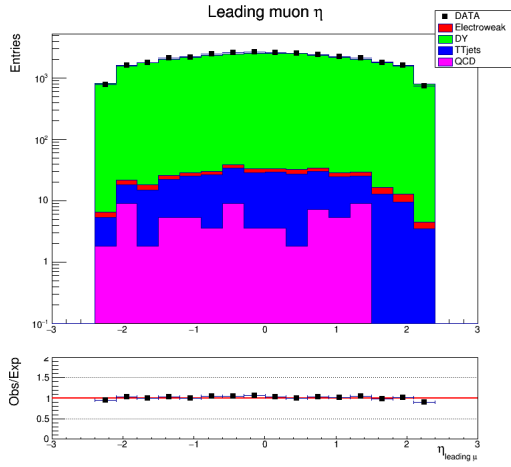
²This because the Z boson is produced with a very low momentum. If it would be produced at rest the two muons should be exactly back-to-back.



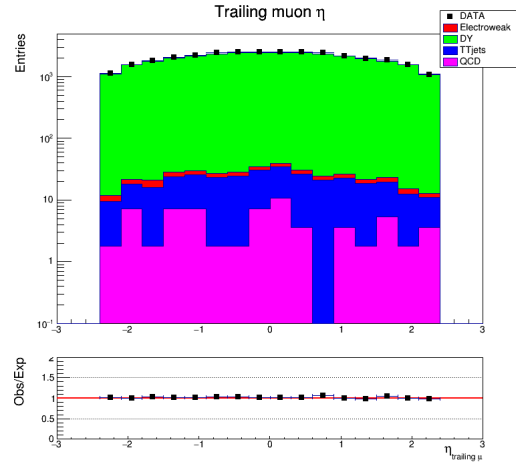
(a) P_T leading muon distribution



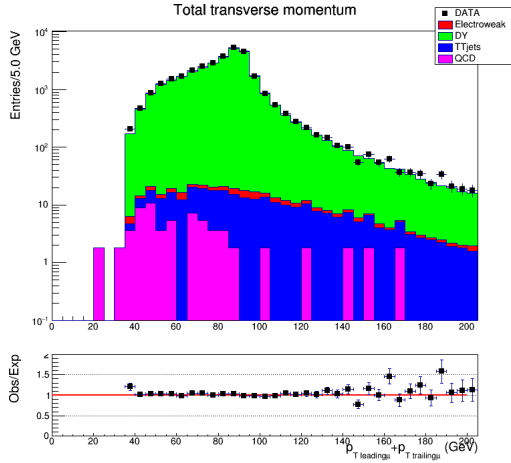
(b) P_T trailing muon distribution



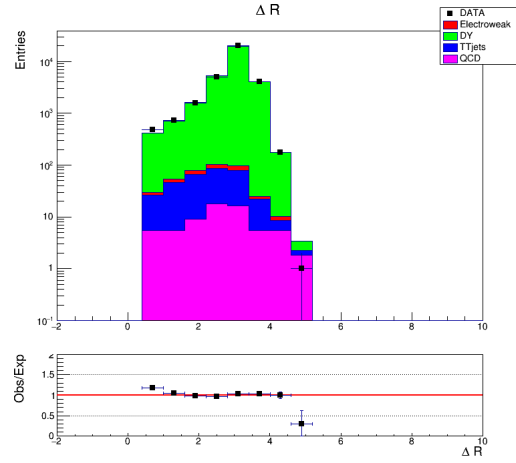
(c) η leading muon distribution



(d) η trailing muon distribution



(e) Sum of the leading and trailing muon p_T



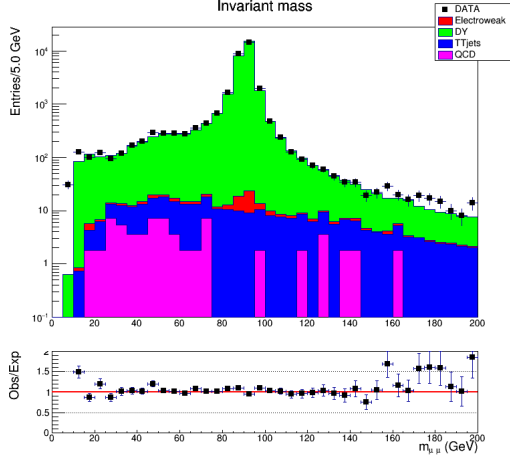
(f) ΔR distribution

Figure 5: Distributions of the leading and the trailing muons. The full points represent the selected data and the error bars represent the statistical errors. The full histograms represent the stacked MC samples weighted to the data luminosity. For all the images, the ratio plot data/MC is also shown.

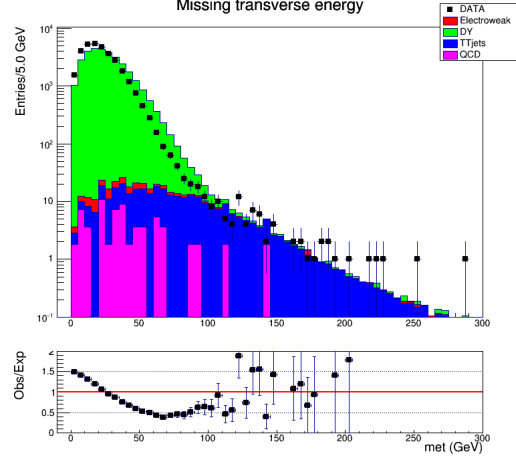
In figure 5 the main distributions for the leading and the trailing muons are shown. The p_T distribution for the leading and the trailing muons are shown in figure 5a and 5b, respectively, and we can see that both peak around 45 GeV. This is because the Z boson is produced by pp collisions almost at rest and it decays into two particles of same mass. In figure 5c and 5d the distribution of the pseudorapidity is shown for the leading and the trailing muon, respectively. Noticing that the plots are on a logarithmic scale, these distributions show that there is no preferential direction in which the muons are emitted. Looking at the distribution of the sum of the leading and trailing muon p_T , which is shown in figure 5e, we can see that it has a peak at ~ 90 GeV. This is in perfect agreement with the value of the Z boson mass which is (91.1876 ± 0.0021) GeV, the mass of the muon being completely negligible ($mass_\mu = (105.6583715 \pm 0.0000035)$ MeV)[2]. The ΔR distribution showed in figure 5f shows that most of the selected muons are emitted back-to-back as we were expecting because of the total momentum conservation. In fact this distribution has a peak at ~ 3 .

In figure 6 we can see some other important distributions. The invariant mass distribution is shown in figure 6a and we can clearly see the Z resonance at a mass value around 90 GeV with a very good agreement between data and MC as shown also by the ratio plots. In figure 6b the missing transverse energy distribution is shown. We can see that the agreement between data and MC is not so good and this is due to the poor simulation of the MET corrections and resolution. The number of jets distribution is shown in figure 6c and the distribution of the sum of the p_T of the jets (H_T) is shown in figure 6d. The jets are defined with ParticleFlow objects and must satisfy the requirement $p_T > 30$ GeV and $|\eta| < 2.4$. Both histograms are obtained with jets that passed the jets identification selection³. From the number of jets distribution we can see that most of the events have no jets as we were expecting and this is the main reason of the peak at zero in the H_T distribution.

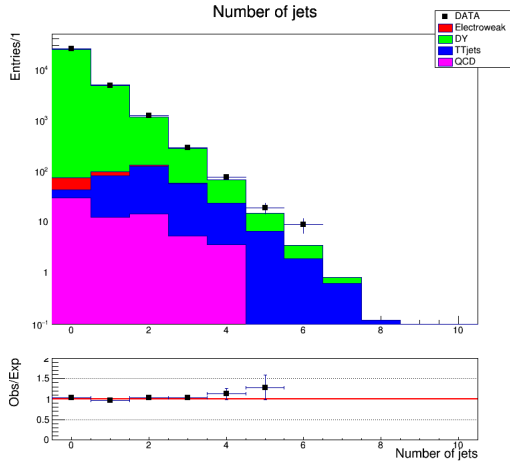
³The jets identification consists of some requirements about the energy in both the electromagnetic and hadronic calorimeter and the numbers of neutral and charged tracks.



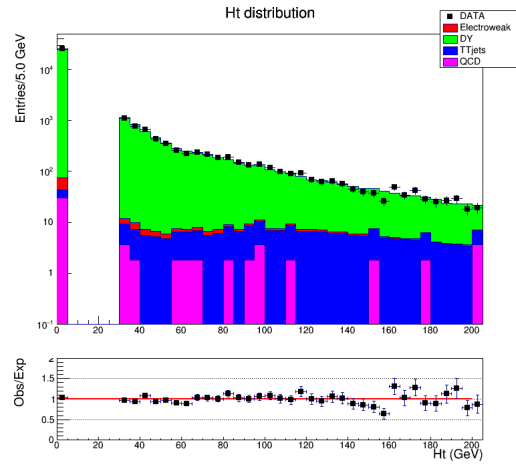
(a) Invariant mass distribution



(b) Missing transverse energy distribution



(c) Number of jets distribution



(d) H_T distribution

Figure 6: Distributions of the invariant mass, missing transverse energy, number of jets and sum of the jets p_T . The full points represent the selected data and the error bars represent the statistical errors. The full histograms represent the stacked MC samples weighted to the data luminosity. For all the images, the ratio plot data/MC is also shown.

6 Efficiency

Another important correction that we have to consider is the efficiency of our selection. One of the most common method used to calculate the efficiency is the “Tag and probe method” which allows for the determination of efficiencies from real data.

6.1 Tag and probe method

The first step is to take one of the muons that passed all the selections and this muon will be the “tag-muon”. Notice that this muon is matched with the trigger object and has passed all the isolation, identification and impact parameters⁴ cuts.

The second step is to check among all the other muons before the selections, the so called “probe-muons”, how many muons has passed or not the selection we are working on. In both cases, two sample are collected: the passing-probe-sample and the failing-probe-sample. These two probe-samples are used to obtain the efficiency ϵ with the following equation:

$$\epsilon = \frac{\text{passing probes}}{\text{total probes}} = \frac{\text{passing probes}}{\text{failing probes} + \text{passing probes}}. \quad (1)$$

6.2 Efficiency calculation

In this analysis the total selection efficiency is calculated for different values of p_T and $|\eta|$ both for the data and the Drell-Yan MC samples⁵.

As shown in figure 7 the invariant mass distribution was fitted with each probe-sample and then fitted using a double Gaussian function for the peak and an exponential function for the background estimation. Dividing the integral of these fitted functions by the width of the bins, the number of failing and passing probes can be obtained both for data and MC.

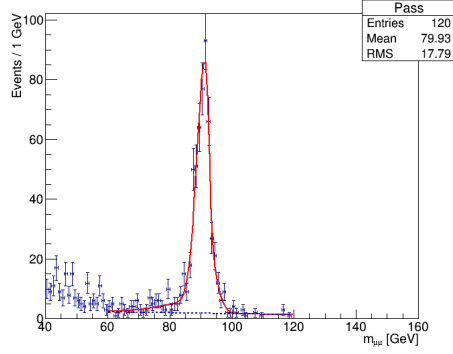
Finally through equation 1 the efficiencies can be obtained both for data and MC and their trend as a function of p_T is shown in figure 8 for the barrel and the endcap.

Considering that the resulting efficiency is a function of p_T and $|\eta|$, each muon pair has to be reweighted using the correct efficiency value as described in the following formula:

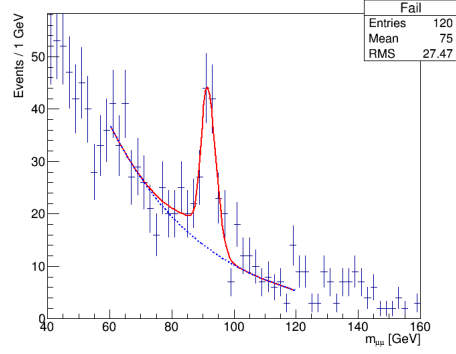
$$\begin{aligned} weight_{ID} &= weight_{ID}(\text{leading } \mu) \cdot weight_{ID}(\text{trailing } \mu) = \\ &= \frac{\epsilon^{DATA}(p_{T\text{Leading } \mu}, \eta_{\text{Leading } \mu})}{\epsilon^{MC}(p_{T\text{Leading } \mu}, \eta_{\text{Leading } \mu})} \cdot \frac{\epsilon^{DATA}(p_{T\text{Trailing } \mu}, \eta_{\text{Trailing } \mu})}{\epsilon^{MC}(p_{T\text{Trailing } \mu}, \eta_{\text{Trailing } \mu})} \end{aligned} \quad (2)$$

⁴The so called “Impact Parameter” (IP) are the transverse momentum, the pseudorapidity and the distances between the impact point and the primary vertex both in longitudinal and transverse plane.

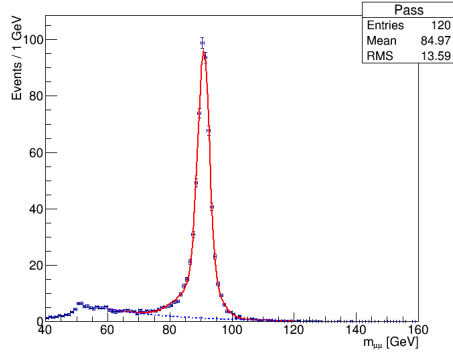
⁵The efficiency for all the other MC samples is not calculated because they do not have resonances and therefore their efficiencies should be negligible.



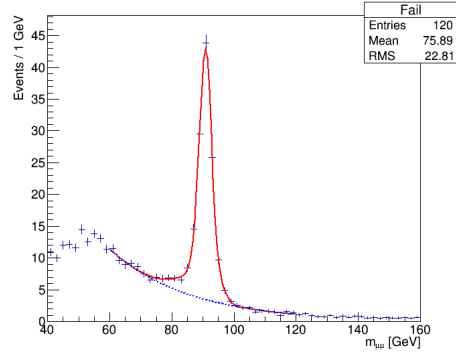
(a) Example of passing-probe histogram filled with data samples



(b) Example of failing-probe histogram filled with data samples



(c) Example of passing-probe histogram filled with MC samples



(d) Example of failing-probe histogram filled with MC samples

Figure 7: Example of passing and failing probes histogram both for data and MC. The points are the entries of the histogram. The full red colored line is the double Gaussian function used to fit the signal. The dashed blue line is the exponential function used to fit the background. This example is for the p_T bin in the range $[10 - 20] \text{ GeV}$ and for $|\eta|$ in the endcap.

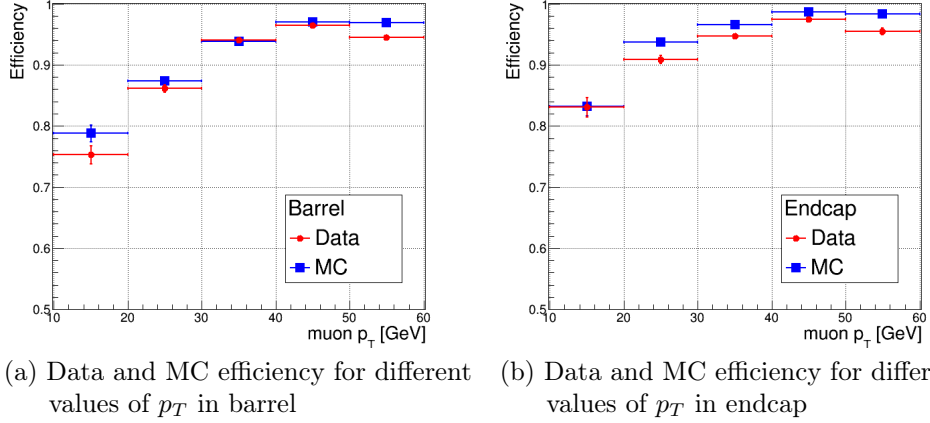


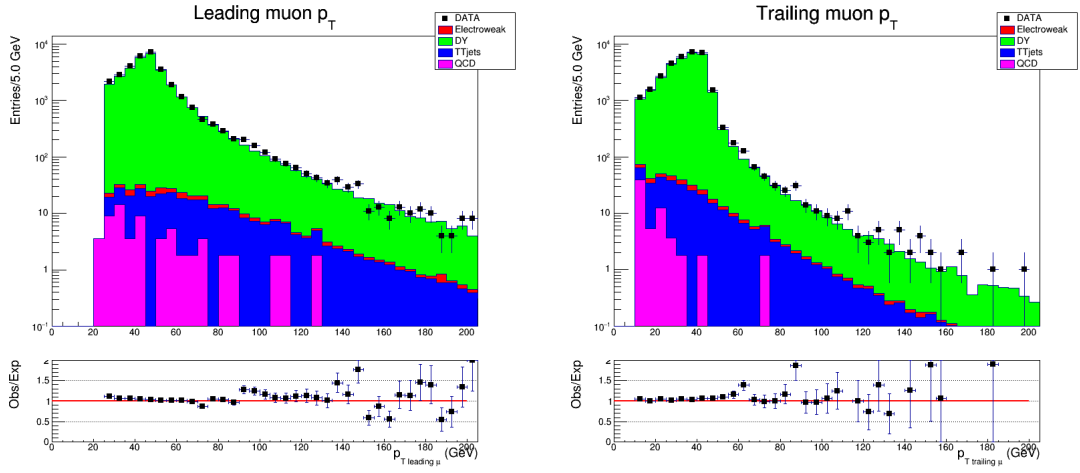
Figure 8: Distributions of the efficiency as a function of the p_T for two different intervals of $|\eta|$. The barrel corresponds to the interval $|\eta| \in [0, 1.48]$ and the endcap corresponds to the interval $|\eta| \in [1.48, 2.4]$.

Table 3: Efficiency scale factors $\frac{\epsilon^{DATA}}{\epsilon^{MC}}$ for different values of p_T and η .

	Barrel $[0 < \eta < 1.48]$	Endcap $[1.48 < \eta < 2.4]$
$p_T \in [10, 20] \text{ GeV}$	0.96	0.99
$p_T \in [20, 30] \text{ GeV}$	0.98	0.97
$p_T \in [30, 40] \text{ GeV}$	1.01	0.98
$p_T \in [40, 50] \text{ GeV}$	0.99	0.99
$p_T > 50 \text{ GeV}$	0.97	0.97

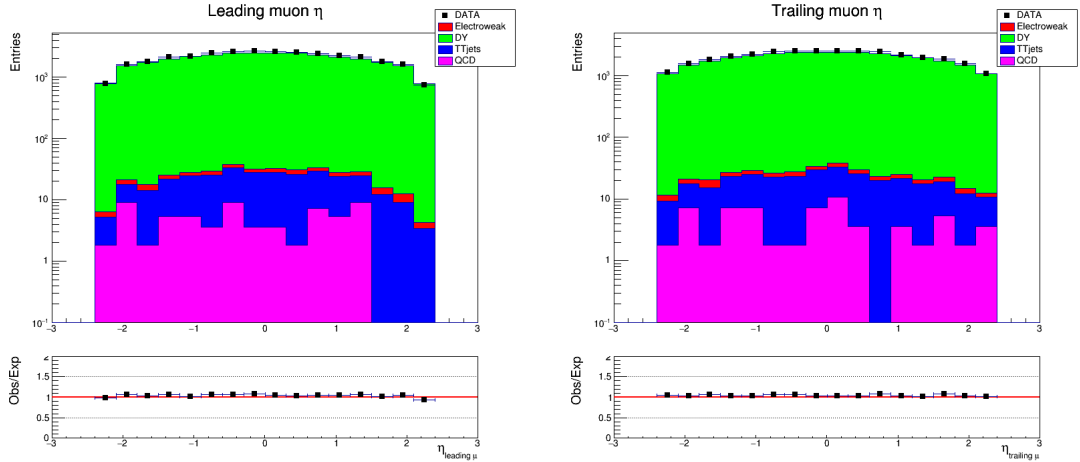
The efficiency scale factors used for the reweighting are listed in table 3. From this table the following mean values are obtained for barrel and endcap respectively: $0.98^{+0.02}_{-0.03}$, $0.98^{+0.02}_{-0.01}$, where the errors are the maximum deviations from the mean value. In figure 9 the main distributions after correcting for the scale factors for the leading and the trailing muons are shown.

In figure 10 we can see some other important distributions.



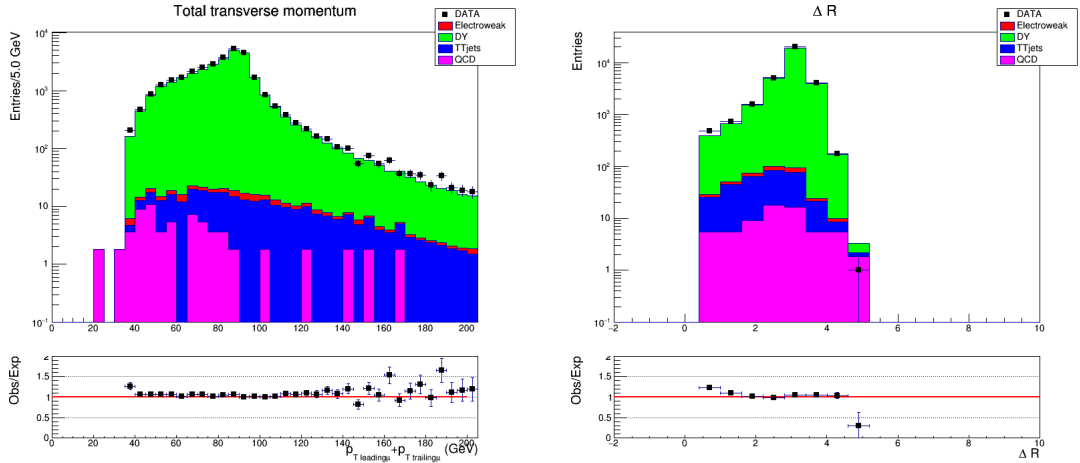
(a) P_T leading muon distribution after efficiency reweighting

(b) P_T trailing muon distribution after efficiency reweighting



(c) η leading muon distribution after efficiency reweighting

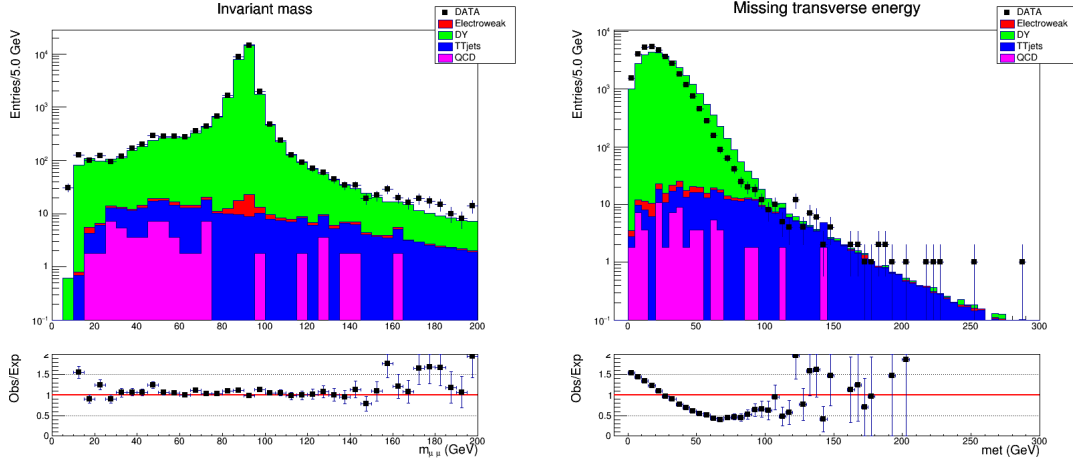
(d) η trailing muon distribution after efficiency reweighting



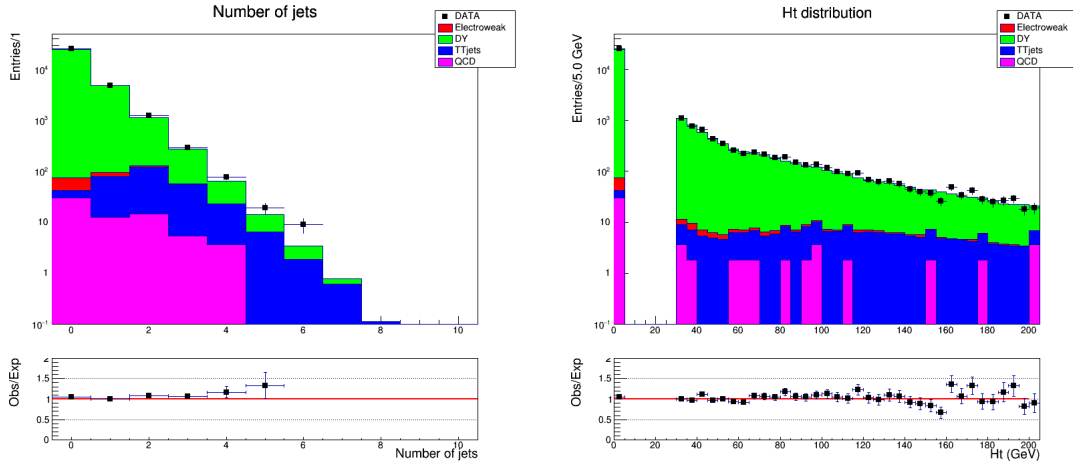
(e) Sum of the leading and trailing muon p_T after efficiency reweighting

(f) ΔR distribution after efficiency reweighting

Figure 9: Distributions of the leading and the trailing muons after the efficiency reweighting. The full points represent the selected data and the error bars represent the statistical errors. The full histograms represent the stacked MC samples weighted to the data luminosity. For all the images, the ratio plot data/MC is also shown.



(a) Invariant mass distribution after efficiency reweighting (b) Missing transverse energy (MET) distribution after efficiency reweighting



(c) Number of jets distribution after efficiency reweighting (d) H_T distribution after efficiency reweighting

Figure 10: Distributions of the invariant mass, missing transverse energy, number of jets and sum of the jets p_T after the efficiency reweighting. The full points represent the selected data and the error bars represent the statistical errors. The full histograms represent the stacked MC samples weighted to the data luminosity. For all the images, the ratio plot data/MC is also shown.

7 Conclusions

This analysis was performed with data collected during the early second run data at the LHC with a total integrated luminosity of 42 pb^{-1} and a bunch separation of 50 ns. In this analysis the "standard candle" process is the $Z \rightarrow \mu\mu$ decay where the selected data were compared to a Drell-Yan MC sample. Possible backgrounds to this process are the diboson production, the $W + jets$ and $t\bar{t} + jets$ processes and QCD processes here estimated with a data-driven method.

After the selection cuts, the pile-up reweighting is applied to the MC samples. With the "Tag and Probe" method the total efficiency for the identification and isolation of the two selected muons is obtained for different value of p_T and η . Scaling factors for the MC efficiency were obtained and applied to the distributions.

After all these corrections, we can see that the agreement between data and MC is very good for all the distributions. The only exception is the missing transverse energy distribution as not all the corrections have been implemented yet in this variable. Also for the efficiency the agreement between data and MC is very good both for the barrel and the endcap.

This study allows to commission the muon objects to be used by the Higgs in tau tau group in the future analysis of the $H \rightarrow \tau\tau$ process.

References

- [1] *Drell, S.D.; Yan, T.-M.* (1970). "Massive Lepton-Pair Production in Hadron-Hadron Collisions at High Energies". *Physical Review Letters* 25 (5): 316-320. Bibcode:1970PhRvL..25..316D. doi:10.1103/PhysRevLett.25.316.
- [2] *K.A. Olive et al.* Particle Data Group, *Chin. Phys. C*, 38, 090001 (2014).
- [3] *Luigi Calligaris* "Search for the Heavy Neutral Higgs Boson decaying into Tau Pairs with the CMS Experiment at the LHC". Dissertation, Hamburg 2015.