Determination of electron identification and isolation efficiency in early Run2 data at CMS experiment.

Alena Kolchanova[1,2]

[1] Institute for Theoretical and Experimental Physics, Russia

[2] National Research Nuclear University MEPhI, Russia

Supervisors:
Alexis Kalogeropoulos
Alexei Raspereza

Abstract

This document reports on measurement of the efficiency of the electron identification and isolation at the CMS experiment. The study is performed on proton-proton collision data collected with the CMS detector at center-of-mass energy of 13 TeV. The analyzed data set corresponds to an integrated luminosity of 41 pb^{-1}. The electron identification and isolation efficiency is determined using tag-and-probe method applied to the sample of the Z boson decays into electron-positron pairs. The related scale factors, accounting for differences in electron identification efficiency in data and simulated samples, are derived and applied to simulated samples. We have found that these corrections improve overall agreement between data and simulation.

Contents

1. Introduction

The Large Hadron Collider at CERN is designed to explore mechanism of electroweak symmetry breaking, study electroweak and strong interactions and search for new physics in proton-proton collisions at highest energy frontier attained so far. In years 2011 – 2012, the LHC has delivered more than 25 fb^{-1} of data at center-of-mass energies 7 and 8 TeV. This data has been thoroughly analyzed by two groups of physicists, operating two multipurpose detectors, ATLAS and CMS. The studied performed by ATLAS and CMS collaborations culminated in the discovery of a new scalar boson, whose properties are consistent with expectation for the Standard Model Higgs boson. In July 2015 the LHC has been brought into operation at center-of-mass energy of 13 TeV and rich physics program at this new energy frontier has started.

Many analyses at the LHC deal with final states characterized by the presence of isolated high momentum electrons. Examples are given by measurements of the Z or W boson production followed by the $Z \rightarrow ee$ ($W \rightarrow e\nu$) decays. Another example is the study of the Higgs boson decay into tau leptons in final states, where one or both tau leptons decay into electron and two neutrinos. The precise knowledge of the electron identification efficiency is important prerequisite for such analyses. In this report we present study aiming at the measurement of electron identification and isolation efficiency and determination of the related scale factors which has to be applied in order to correct simulation for possible mis-modeling of electron identification and isolation. The analysis is done on first proton-proton collision data collected with the CMS detector at the center-of-mass energy of 13 TeV.

2. Standard Model of elementary particles

The Standard Model is the quantum field theory, which classifies elementary particles and describes fundamental interactions between them. The matter constituents are represented by fermions - spin ½ particles. Fundamental fermions are classified into two groups: leptons and quarks. The family of leptons includes

electron (e), muon (μ) and tau (τ), and corresponding neutrinos, . Quarks exist in six flavors known under the names “up”, “down”, “strange”, “charm”, “bottom” and “top”. These twelve fermions and their anti-particles make up the matter we observe in the universe. The forces through which these fermions interact are mediated by spin-1 particles, so-called gauge bosons. The Standard Model describes three types of interactions.

- Strong interactions are mediated by eight massless gluons (g).
- Weak interactions are mediated by massive $W^{\pm}$ and Z bosons.
- Electromagnetic interactions are mediated by massless photon (γ).

The Standard Model is based on a principle of local gauge invariance. To generate masses of fermions and weak bosons in a gauge invariant way, the Higgs field is introduced in theory. All fermions and weak bosons acquire mass through interaction with the ground state of this field which has non-zero vacuum expectation value. Excitation of this ground state manifests itself as one more physical state in the model – spin-0 particle, the Higgs boson. This particle has been discovered by the ATLAS and CMS collaborations in decays to pairs of gauge bosons, $H \rightarrow W^{+}W^{-}$, ZZ, γγ [1,2]. Later on, a more than 3σ evidence of the Higgs boson decays into tau leptons, $H \rightarrow \tau\tau$, has been also established [3,4].

3. CMS experiment

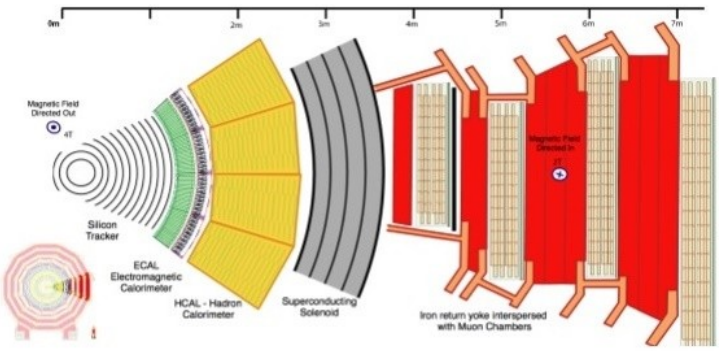


Figure 1. Transverse Slice of the Compact Muon Solenoid (CMS) Detector.

The Compact Muon Solenoid (CMS) is a general-purpose detector at the LHC [5]. It is designed to investigate a broad range of physics, including physics

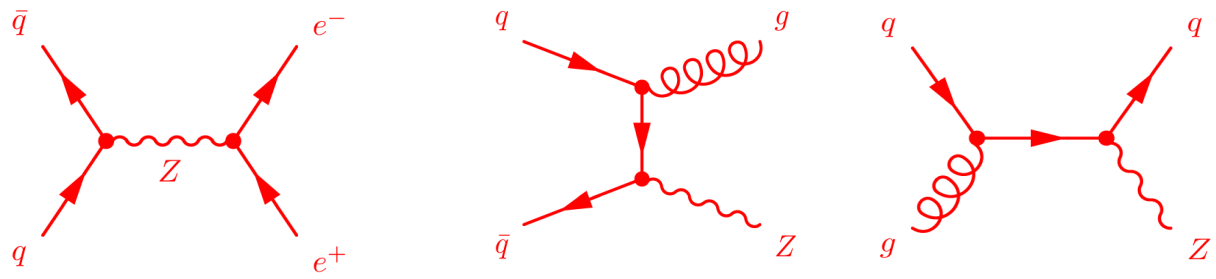
beyond the SM. The detector is 21 m long, 15 m wide and 15 m high. In Fig. 1 a transverse slice of the detector is shown. The detector has an onion-like structure. The main feature of the detector is a huge solenoid magnet, that generates field of 3.8 T. Within the volume of the magnetic field essential
parts for detecting particles of different types are embedded:

- Silicon pixel detector, allowing to measure the vertex position and paths of particles with high precision.
- Strip tracker is used to reconstruct tracks of charged particles and measure their momenta.
- Electromagnetic calorimeter made of lead tungstate, measures energy of electrons and photons.
- Hadron calorimeter to detect and measure energy of hadrons. It uses layers of absorber and scintillator material that produces a rapid light pulse when a particle passes through.
- Muon detectors are placed outside solenoid and consist of multiple layers of resistive plate chambers, cathode strip chambers and drift tubes, providing reconstruction of muon tracks in outer part of the whole detector.

The origin of the coordinate system is placed in the geometrical center of the CMS detector. The x-axis points to the center of the LHC ring. The y-axis goes vertically upwards. The z-axis is aligned with the direction of the proton beam. The polar angle θ is measured with respect to the positive z-axis, and the azimuthal angle φ is measured in the x-y plane. The pseudo-rapidity of reconstructed particles is defined as $\eta = -\log(\tan(\theta/2))$.

4. Dataset and studied processes

The analyzed data sample corresponds to an integrated luminosity of 41 pb^{-1} of proton-proton collisions collected by CMS at center-of-mass energy 13 TeV. The study focuses of Drell-Yan production of electron-positron pairs. The leading and next-to-leading order Feynman diagrams are presented in Fig. 2.



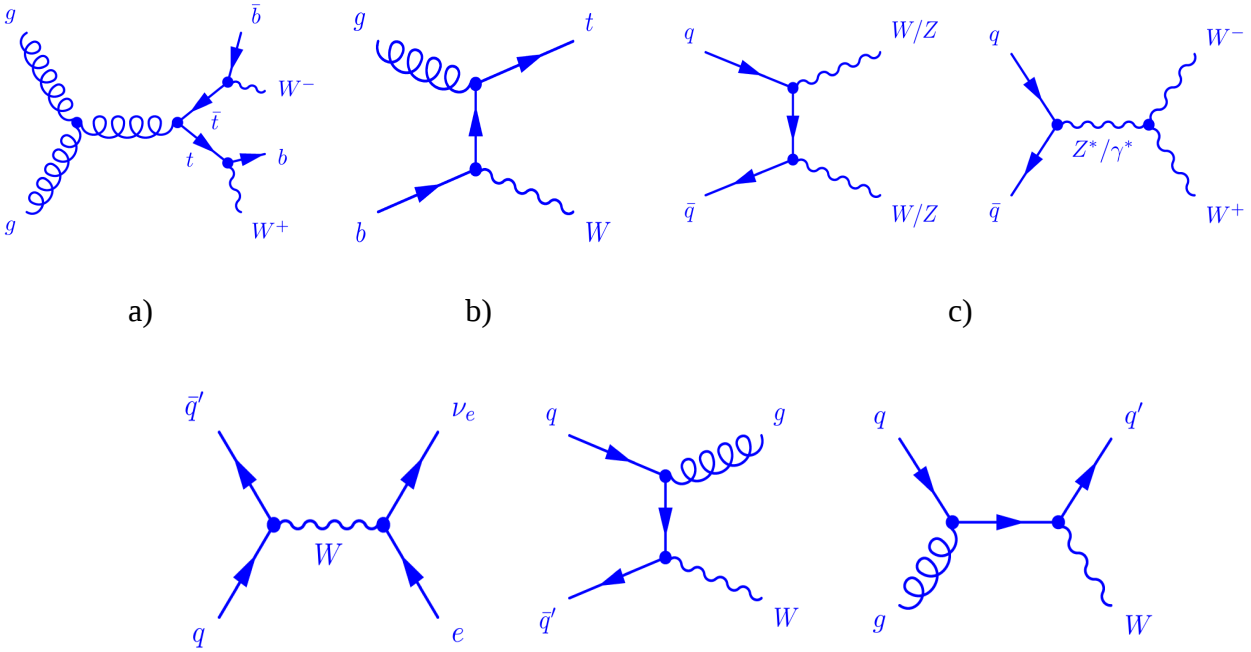
a) b)

Figure 2. Feynman diagrams for the studied processes. a) Leading order diagram for Drell-Yan e+e- production. b) Next-to-leading order diagrams describing production of the Z boson in association with jets.

In our study we considered the following background processes.

- Top-quark pair production;
- Single top production in association with W boson;
- W boson production in association with jets;
- Pair production of weak bosons: WW, WZ, ZZ.

Feynman diagrams for background processes are presented in Fig. 3.



a) b) c)

d)

Figure 3. Feynman diagrams for background processes: a) top pair production; b) top+W production; c) Diboson production; d) W+jets production.

For all studied processes, a dedicated simulated Monte Carlo (MC) samples have been produced by the generator group of the CMS Collaboration. Details of these MC samples are given in Table 1.

Physics Process	Theoretical cross section	Number of generated events	Monte Carlo generator
$Z \rightarrow ll$ ($10 < m(ll) < 50$ GeV)	1.8 nb	30535559	aMC@NLO
$Z \rightarrow ll$ ($m(ll) > 50$ GeV)	6 nb	28825132	aMC@NLO
W+Jets, $W \rightarrow lv$	62 nb	24151270	aMC@NLO
top-quark pair production	0.83 nb	42730273	aMC@NLO
top+W^+ , anti-top+W^-	71 pb	1995000	POWHEG
WW	63 pb	994416	PYTHIA8
WZ	23 pb	991232	PYTHIA8
ZZ	10 pb	996168	PYTHIA8

Table 1. Simulated MC samples used in the analysis.

5. Analysis

5.1 Electron identification and isolation at CMS

Several strategies are used in CMS to identify prompt isolated electrons (signal), and to separate them from background sources, mainly originating from photon conversions, jets misidentified as electrons, or electrons from semileptonic decays of b and c quarks. Simple and robust algorithms have been developed [6] to apply sequential selections on a set of discriminants. More complex algorithms combine variables in an MVA analysis to achieve better discrimination. In addition, dedicated selections are used for highly energetic electrons.

Variables that provide discriminating power are grouped into three main categories:

- Observables that compare measurements obtained from the ECAL and the tracker (track–cluster matching, including both geometrical as well as SC energy–track momentum matching).
- Purely calorimetric observables used to separate genuine electrons (signal electrons or electrons from photon conversions) from misidentified electrons (e.g., jets with large electromagnetic components), based on the transverse shape of electromagnetic showers in the ECAL and exploiting the fact that electromagnetic showers are narrower than hadronic showers. Also utilized are the energy fractions deposited in the HCAL (expected to be small, as electromagnetic showers are essentially fully contained in the ECAL), as well the energy deposited in the preshower in the endcaps.
- Tracking observables employed to improve the separation between electrons and charged hadrons, exploiting the information obtained from the Gaussian-Sum-Filtered(GSF)-fitted track, and the difference between the information from the Kalman-Filter and GSF-fitted tracks.

We follow recommendation of the E/Gamma Physics Object Group of the CMS Collaboration who proposed a set of optimized (p_T,η)-dependent cuts on the MVA discriminant. These criteria ensure selection of electrons, having transverse momentum greater than 20 GeV, with efficiency close to 80% at the fake rate of less than 1%.

Electrons originating from decays of W and Z boson as well as those emerging in the H → decays are expected to be isolated. In CMS analyses, the electron isolation is defined as

$$\mathrm{Iso^e} = \frac{\sum_{\Delta R<0.3}(\mathrm{p_T^{h+}} + \mathrm{p_T^{\gamma}} + \mathrm{p_T^{h0}} - 1.5 \cdot \mathrm{p_T^{pu}})}{\mathrm{p_T^e}}, \qquad [1]$$

The numerator in Eq. [1] contains p_T -sum of all charged particles, photons and neutral hadrons within the isolation cone $\Delta R<0.3$. The sum is corrected for pileup effects by subtracting contribution from charged particles, originating from pileup vertices. This contribution is multiplied by a factor 1.5, which accounts for the fact that in QCD processes the total energy of neutral hadrons and the total energy of

charged hadrons are produced on average with ratio 1:2. The denominator contains electron p_T. To isolate electrons in the analysis, an upper cut is placed on variable Iso^e.

5.2 Selection of dielectron events

Events with electron-positron pairs, are selected online with the single electron trigger, implementing p_T threshold of 27 GeV and requiring $|\eta|<2.1$. We require p_T > 29 (10) GeV for leading (trailing) electron. Pseudo-rapidities are requested to be $|\eta| < 2.1$ (2.5) for leading (trailing) electron. Leading electron must match trigger object, associated with the single electron trigger. As in our analysis we expect prompt electrons originating from the Z boson decays, the cuts are applied on the transverse and longitudinal impact parameters: $|d_{XY}| < 0.45$ mm, $|d_Z| < 2$mm.

In order to exclude electrons, originating from heavy-quark decays, we require electrons to be isolated by imposing cut $Iso^e < 0.15$, where Iso^e is defined according to Eq. [1]. Electrons are requested to be separated by > 0.5.

5.3 Measurement of electron identification and isolation efficiency

The Tag-and-Probe (T&P) method [7] allows to determine from data (or simulation) efficiencies and other physical quantities representing the fraction of events passing (or failing) a given selection. It is based on the determination of the total number of events and that of those passing the cuts. The method is used frequently in detector studies due to the quality of its results, as in good conditions the extracted value can be affected only by a very small bias compared to the "real" value being estimated. It is easier to explain it using a concrete example: the selection efficiency of a series of identification cuts for real electrons in experimental data.

Measuring the efficiency of a set of identification cuts for real electrons presents a challenge derived from conflicting needs. Indeed, it is desirable to

tighten the pre-selection of the electrons so that the measurement is performed on a very pure sample.

Unfortunately, tightening the electron pre-selection biases the set of electrons over which the efficiency is being estimated, to the point that if high purity is achieved the measurement will likely be excessively optimistic. This happens because genuine electrons which were difficult to identify had been filtered out by the tight pre-selection.

One possible solution to this dilemma is using as the dataset on which the measurement is performed pairs of electrons emerging from the decay of well-known massive particles, such as J/Ψ and Z boson. The line shape of the distribution in the invariant mass of the e^+e^- pair is known from theory, and the line shape of the non-resonant background can be modeled by a properly chosen function.

In the case Z resonance, the shape of its peak is modeled by double gaussian distribution. The shape of the background is modeled by exponential function.

The tight selection is imposed on one (and only one) of the electrons in the pair, this electron is called the tag. This tight selection is chosen as to have an acceptable $Z \rightarrow ee$ signal to background ratio in the studied dataset. We require tag electron to pass electron identification criteria as proposed by the E/Gamma Physics Object Group (see Section 5.1) and to match trigger object, associated with the single muon trigger. The other electron, the probe, is subject to a set of selection and identification criteria that are designed to keep the bias of the measurement small. In our case, the sample of probes consists of all GSF-fitted electrons, matching the reconstructed cluster of energy deposits in the electromagnetic calorimeter. Then the invariant mass distributions of the tag and probe electron pairs are constructed and the $Z \rightarrow ee$ signal yields are extracted by means of the fit both for the subsample, in which the probe electrons pass the selection criteria under study (“passing probes”), and for the subsample, in which the probe electrons fail the same criteria (“failing probes”). These measurements are done differentially in (p_T,η) bins of the probe electrons. Figure 4 shows few

representative examples of the fits of the tag-and-probe electron invariant mass distributions.

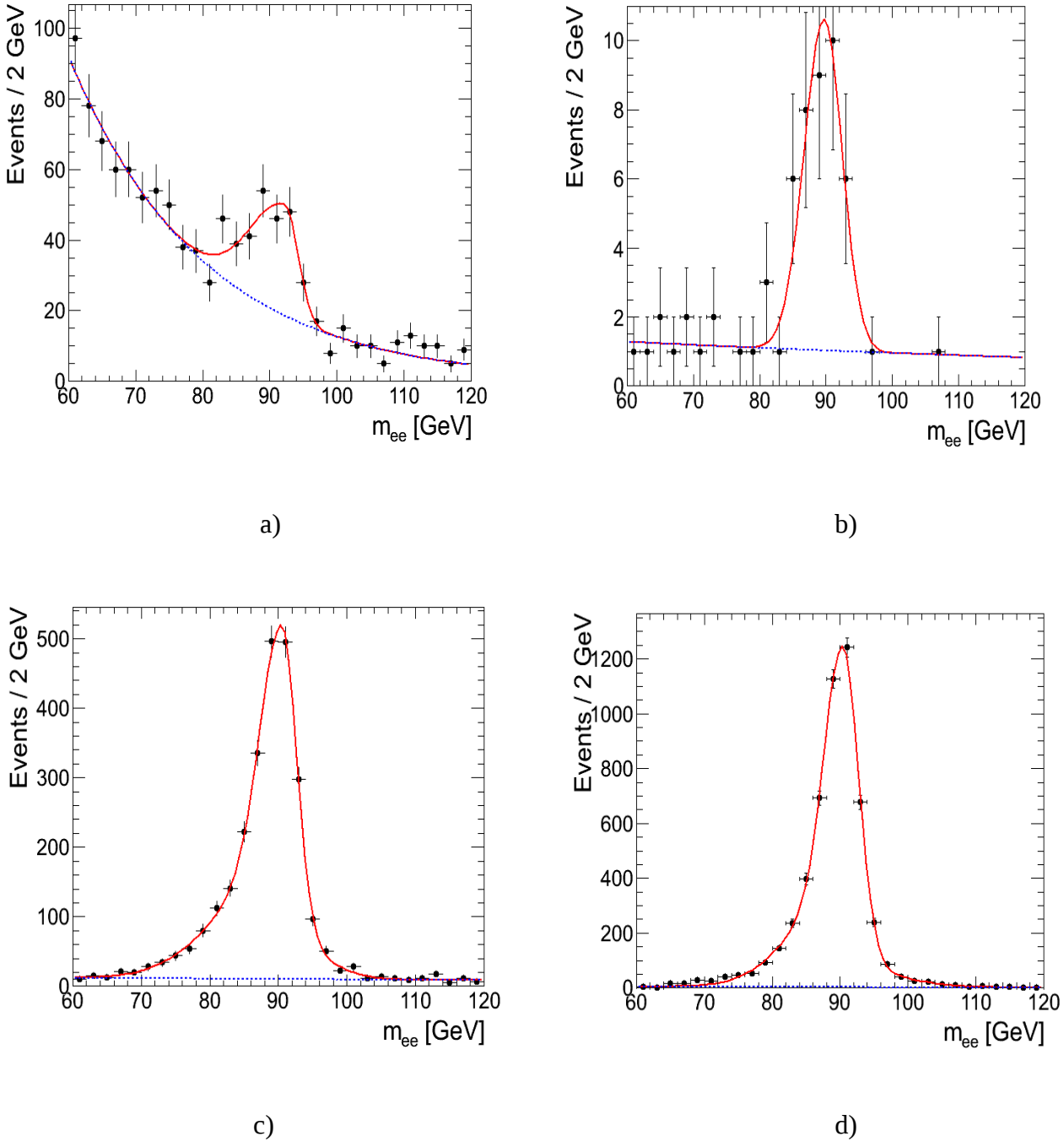


Figure 4. Example of fits of the tag-and-probe electron pair invariant mass spectra. a) Sample of failing probes for electrons with $10 < p_T < 15$ GeV in the barrel region ($|\eta| < 1.48$). b) Sample of passing probes for electrons with $10 < p_T < 15$ GeV in the barrel region. c) Sample of failing probes for electrons with $30 < p_T < 40$ GeV in the endcap region ($1.5 < |\eta| < 2.5$). d) Sample of passing probes for electrons with $30 < p_T < 40$ GeV in the endcap region.

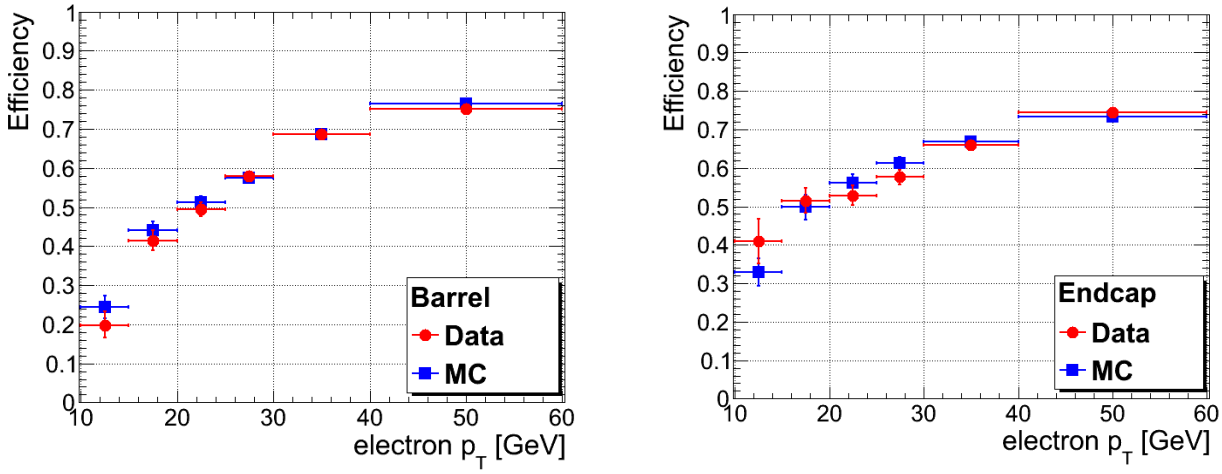
The $Z \rightarrow ee$ signal yields in the tag-and-probe electron mass distributions are extracted according to the formula

$$N_{Z\rightarrow ee} = \frac{\int S(m)dm}{\Delta_{bin}}, \qquad [2]$$

where the numerator contains the integral of the fitted signal function (double gaussian), and the denominator is the bin width in the fitted histograms. The efficiency is computed as

$$\epsilon = \frac{N_{pass}}{N_{pass} + N_{fail}}, \qquad [3]$$

with N_{pass} being the $Z \rightarrow ee$ signal yield computed in the sample of passing probes, and N_{fail} is the signal yield computed in the sample of passing probes. The efficiency measured in data and MC is presented in Fig. 5 as a function of the probe electron p_T for two pseudo-rapidity regions, barrel ($|\eta| < 1.48$) and endcap ($1.48 < |\eta| < 2.5$).



a) b)

Figure 5. Efficiency of the electron identification and isolation as a function of electron p_T in the barrel region (a) and endcap (b). Efficiency measured in data (circles) is compared to the efficiency determined in the simulated $Z \rightarrow ee$ MC sample (squares).

5.4 Validation of the electron efficiency scale factors in MC simulation

To correct for slight differences in the electron identification and isolation efficiencies between data and simulation, each MC event is assigned weight

$$w_{\text{eff.}} = \frac{\epsilon_{\text{data}}\big(p_T(\text{lead}), \eta(\text{lead})\big)}{\epsilon_{\text{MC}}\big(p_T(\text{lead}), \eta(\text{lead})\big)} \times \frac{\epsilon_{\text{data}}\big(p_T(\text{trail}), \eta(\text{trail})\big)}{\epsilon_{\text{MC}}\big(p_T(\text{trail}), \eta(\text{trail})\big)}, \quad [4]$$

where

- $\varepsilon_{data}(p_T,\eta)$: (p_T,η)-dependent efficiency measured in data;
- $\varepsilon_{MC}(p_T,\eta)$: (p_T,η)-dependent efficiency determined in MC simulation;
- $(p_T(\text{lead}),\eta(\text{lead}))$: leading electron p_T and η;
- $(p_T(\text{trail}),\eta(\text{trail}))$: trailing electron p_T and η.

The MC corrections are also known under the name “electron efficiency scale factors”. To estimate effect of applying corrections in simulation, we compare data with the MC predictions before and after applying corrections in several key kinematic distributions:

- leading electron p_T;
- leading electron η;
- trailing electron p_T;
- trailing electron η;
- dielectron invariant mass, m_{ee}.

Figure 6 shows representative example of such a comparison in the distribution of the trailing electron p_T. To quantify effect of the MC corrections, we perform χ^2-test for the key kinematic distributions. The χ^2-value, quantifying consistency between data and simulation, is computed taking into account Poisson errors in each bin of the distribution:

$$\chi^2 = \sum_{\text{bins}} \frac{\big(N_{\text{data}}(i) - N_{\text{MC}}(i)\big)^2}{N_{\text{MC}}(i)}, \quad [5]$$

where

- $N_{data}(i)$ is the content of bin “i” in the data distribution;
- $N_{MC}(i)$ is the content of bin “i” in the simulated distribution;

- sum runs over all bins in the tested distribution.

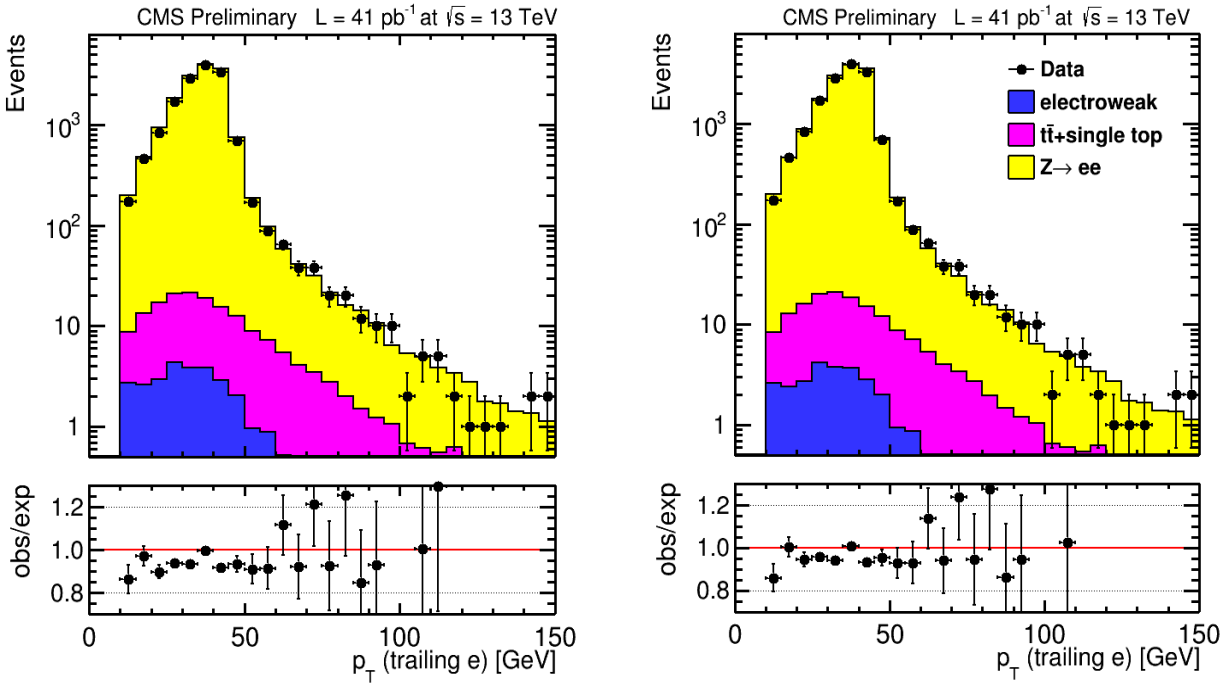


Figure 6. Distribution of the trailing electron p_T before (a) and after (b) applying electron efficiency scale factors to MC samples. The MC corrections improve slightly agreement between data and simulation.

The results of χ^2-tests are summarized in Table 2. In all tested distributions, the χ^2-value gets lower after applying MC corrections. This indicates that electron efficiency scale factors improve agreement between data and simulation.

Distribution	χ^2-value w/o corrections	χ^2-value with corrections
p_T of the leading electron	78	56
p_T of the trailing electron	82	57
η of the leading electron	94	80
η of the trailing electron	88	64
dielectron invariant mass	192	167

Table 2. The χ^2-values, quantifying consistency between data and MC predictions, for key kinematic distributions before and after applying electron efficiency scale factors applied.

The key kinematic distributions obtained after applying electron efficiency scale factors are presented in Figures 7-9. Good agreement between data and simulation is found in all kinematic distributions.

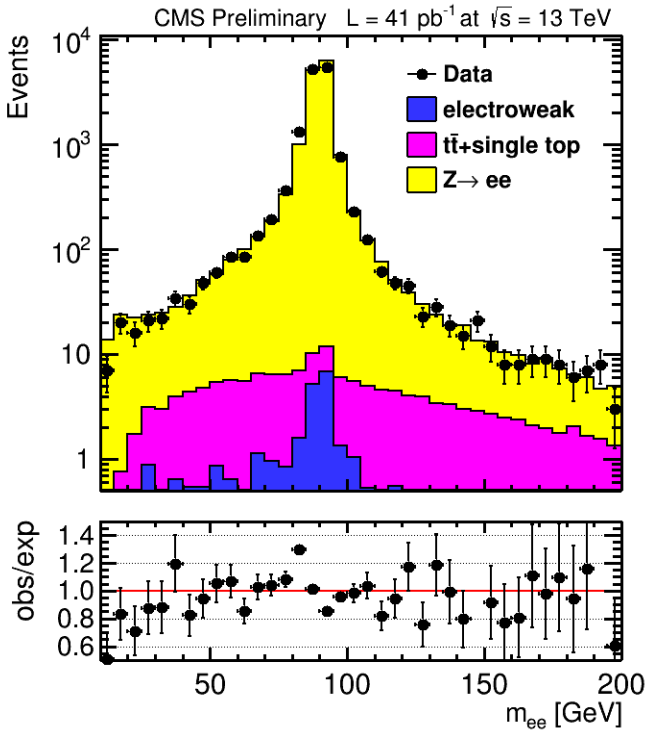
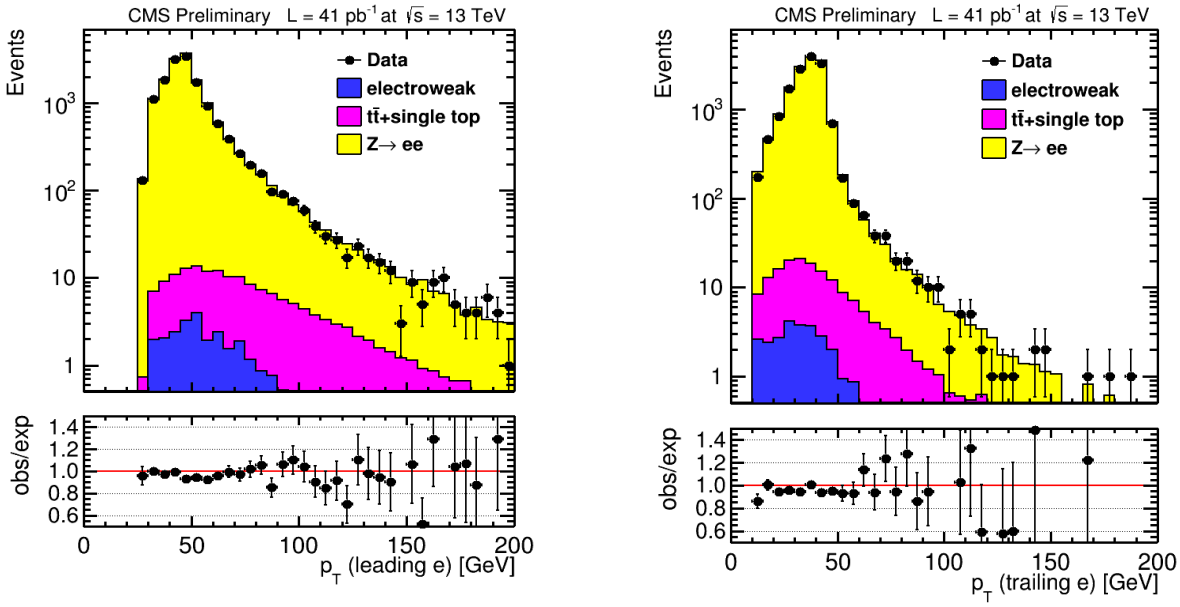
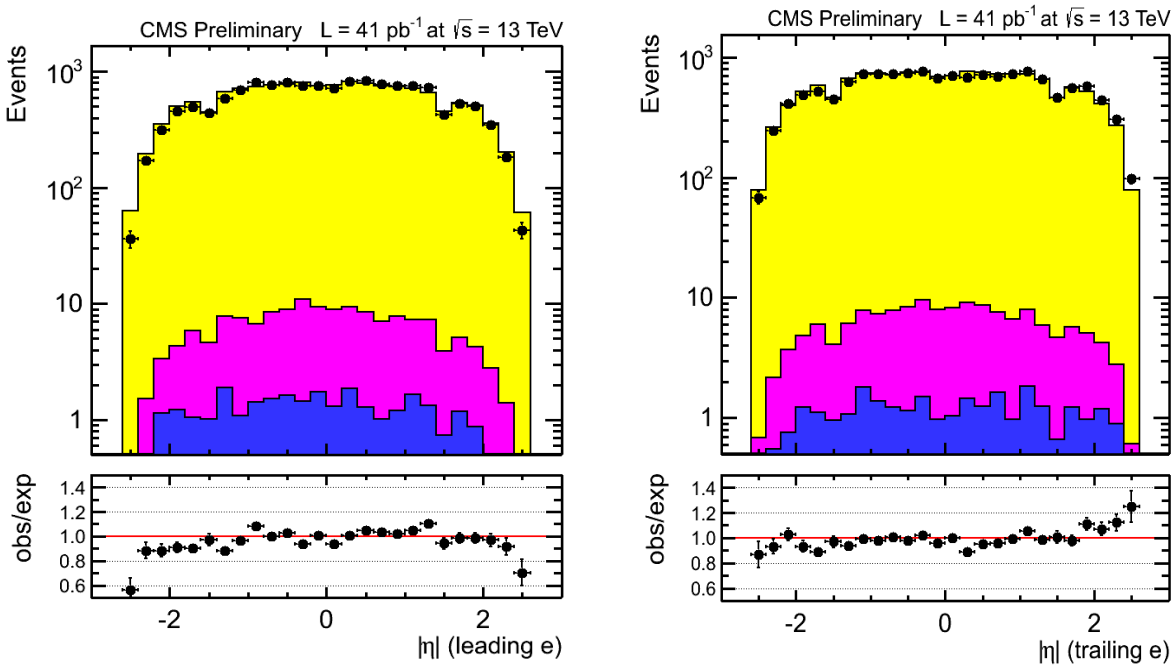


Figure 7. Distribution of the dielectron invariant mass after applying MC corrections.



a) b)

Figure 8. Distributions of the leading electron (a) and trailing electron (b) p_T after applying MC corrections.



a) b)

Figure 9. Distributions of the leading electron (a) and trailing electron (b) η after applying MC corrections.

6. Summary

The dielectron channel is studied in early Run2 proton-proton collision data collected with the CMS experiment at center-of-mass energy of 13 TeV. The analyzed data set corresponds to an integrated luminosity of 41 pb^{-1}. The electron identification and isolation efficiency is determined in data and simulated samples differentially in bins of electron p_T, and η, using tag-and-probe technique applied to the sample of $Z \rightarrow ee$ events. The related scale factors are derived and applied to the MC samples. The corrections are found to improve agreement between data and MC predictions in key kinematic distributions of electrons.

References

[1] The ATLAS Collaboration, “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”, Phys. Lett. B **716** (2012) 1.

[2] The CMS Collaboration, “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”, Phys. Lett. B **716** (2012) 30.

[3] The CMS Collaboration, “Evidence for the 125 GeV Higgs boson decaying to a pair of leptons”, JHEP **05** (2014) 104.

[4] The ATLAS Collaboration, “Evidence for the Higgs-boson Yukawa coupling to tau leptons with the ATLAS detector”, JHEP **03** (2015) 088.

[5] The CMS Collaboration, “The CMS experiment at the CERN LHC”, JINST 3 (2008) S08004.

[6] The CMS Collaboration, “Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at 8 TeV”, arXiv:1502.02701.

[7] Luigi Calligaris, "Search for a Heavy Neutral Higgs Boson decaying into Tau Pairs with the CMS Experiment at the LHC", PhD thesis, DESY 2015.