



ATLAS NOTE

ATLAS-CONF-2014-032

June 2, 2014



Electron efficiency measurements with the ATLAS detector using the 2012 LHC proton-proton collision data

The ATLAS Collaboration

Abstract

This note describes the algorithms used for the reconstruction and identification of prompt electrons in the central region of the ATLAS detector for the pp collisions data collected in 2012 at $\sqrt{s} = 8$ TeV. The efficiency of these algorithms is measured using $Z \rightarrow ee$, $Z \rightarrow ee\gamma$ and $J/\psi \rightarrow ee$ decays using the full dataset, corresponding to an integrated luminosity of 20.3 fb^{-1} .



1 Introduction

In the ATLAS detector [1], electrons in the central detector region are triggered by and reconstructed from energy deposits in the electromagnetic (EM) calorimeter that are matched to a track in the inner detector. Electrons are distinguished from other particles using several sets of identification criteria with different levels of background rejection and signal efficiency. These identification criteria rely on the shapes of electromagnetic showers in the calorimeter as well as on tracking and track-to-cluster matching quantities and are based either on independent cuts on these quantities or on a single cut on the output of a likelihood function taking as input these quantities.

In this document, measurements of the efficiency to reconstruct and identify prompt electrons in the central region of the ATLAS detector¹ with $|\eta| < 2.47$ in the pp collision data recorded in 2012 at a center-of-mass energy of $\sqrt{s} = 8$ TeV are presented and compared to the expectation from Monte Carlo simulation. Electrons from semi-leptonic heavy flavor decays are treated as background. The efficiency of the trigger system to select identified electrons is discussed elsewhere (see Ref. [2]).

The efficiency measurements follow the methods introduced in Ref. [3] for the 2011 ATLAS electron performance studies but are improved in several respects and adjusted for the 2012 conditions. The changes are discussed in the respective sections of the note. The measurements are based on the tag-and-probe method using the Z and the J/ψ resonances, requiring the presence of an isolated identified electron as the *tag*. Additional selection criteria are applied to obtain a clean sample of electron candidates that can be used as *probes* to measure the reconstruction or identification efficiency. The measurements described in this note span different but overlapping kinematic regions and are studied as a function of electron transverse momentum and pseudorapidity. The results are combined taking into account bin-to-bin correlations.

After briefly describing the ATLAS detector in Section 2, the algorithms to reconstruct and identify electrons are briefly summarized in Sections 3 and 4. The general methodology of tag-and-probe efficiency measurements and the decomposition of the efficiency into its different components are reviewed in Section 5. The data and Monte Carlo samples used in this work are summarized in Section 6. Sections 7 and 8 describe the identification efficiency measurements for signal electrons as well as backgrounds. Section 9 details the reconstruction efficiency measurement, which extends the identification measurement methodology, and Section 10 describes the final results for the combined identification and reconstruction efficiency measurements. Section 11 concludes the note with a summary of the final results.

2 The ATLAS detector

A complete description of the ATLAS detector is provided in Ref. [1].

The inner detector (ID) provides a precise reconstruction of tracks within $|\eta| < 2.5$. It consists of three layers of pixel detectors close to the beam-pipe, four layers of silicon microstrip detector modules with pairs of single-sided sensors glued back-to-back (SCT) providing eight hits per track at intermediate radii, and a transition radiation tracker (TRT) at the outer radii, providing about 35 hits per track (in the range $|\eta| < 2.0$). The TRT offers substantial discriminating power between electrons and charged hadrons over a wide energy range (between 0.5 and 100 GeV) via the detection of X-rays produced by transition radiation. The innermost pixel layer (also called the *b-layer*) is located just outside the beam-pipe at a

¹ ATLAS uses a right-handed coordinate system with its origin at the nominal pp interaction point at the center of the detector. The positive x -axis is defined by the direction from the interaction point to the center of the LHC ring, with the positive y -axis pointing upwards, while the beam direction defines the z -axis. The azimuthal angle ϕ is measured around the beam axis and the polar angle θ is the angle from the z -axis. The pseudorapidity is defined as $\eta = -\ln \tan(\theta/2)$. The radial distance between two objects is defined as $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$. Transverse energy is computed as $E_T = E \cdot \sin \theta$.

radius of 50 mm. It provides precise vertexing and significant rejection of photon conversions through the requirement that a track has a hit in this layer.

The main electromagnetic (EM) calorimeter is a lead-liquid argon sampling calorimeter with accordion-shaped electrodes and lead absorber plates. It is divided into a barrel section (EMB) covering $|\eta| < 1.475$ and two end-cap sections (EMEC) covering $1.375 < |\eta| < 3.2$. For $|\eta| < 2.5$, it is divided in three longitudinal layers (called strip, middle and back layers) and offers a fine segmentation in the lateral direction of the showers. At high energy, most of the EM shower energy is collected in the middle layer which has a lateral granularity of 0.025×0.025 in $\eta \times \phi$ space. The first (strip) layer consists of finer-grained strips segmented in the η -direction with a coarser granularity in ϕ . It provides excellent $\gamma - \pi^0$ discrimination and a precise estimation of the pseudorapidity of the impact point. The back layer collects the energy deposited in the tail of high energy EM showers. A thin presampler detector, covering $|\eta| < 1.8$, is used to correct for fluctuations in upstream energy losses. The transition region between the EMB and EMEC calorimeters, $1.37 < |\eta| < 1.52$ has a large amount of material in front of the first active calorimeter layer.

Hadronic calorimeters with at least three longitudinal segments surround the EM calorimeter and are used in this context to reject hadronic jets. The forward calorimeters cover the range $3.1 < |\eta| < 4.9$ and also have EM shower identification capabilities given their fine lateral granularity and longitudinal segmentation into three layers.

3 Electron reconstruction

Electron reconstruction in the *central region* of the ATLAS detector ($|\eta| < 2.47$) starts from energy deposits (clusters) in the EM calorimeter which are then associated to reconstructed tracks of charged particles in the inner detector.

3.1 Electron seed-cluster reconstruction

The $\eta - \phi$ space of the EM calorimeter system is divided into a grid of $N_\eta \times N_\phi = 200 \times 256$ towers of size $\Delta\eta^{\text{tower}} \times \Delta\phi^{\text{tower}} = 0.025 \times 0.025$, corresponding to the granularity of the EM accordion calorimeter middle layer. The energy of the cells in all longitudinal layers (the front, middle and back EM accordion calorimeter layers and for $|\eta| < 1.8$ also the presampler detector) is summed to get the tower energy. The energy of cells spanning several towers is distributed uniformly among the participating elements.

To reconstruct the EM clusters, seed clusters of longitudinal towers with total cluster transverse energy above 2.5 GeV are searched for by a *sliding-window* algorithm [4]. The window size is 3×5 towers in the $\eta - \phi$ space. A duplicate removal algorithm is applied on close-by seed clusters.

Cluster reconstruction is expected to be very efficient for true electrons. In MC simulations, the efficiency is about 95% for electrons with a transverse energy of $E_T = 7$ GeV and reaches 99% at $E_T = 15$ GeV and 99.9% at $E_T = 45$ GeV, placing a requirement only on the angular distance between the generated electron and the reconstructed electron cluster. The efficiency decreases with increasing pseudorapidity in the endcap region $|\eta| > 1.37$.

3.2 Electron-track candidate reconstruction

Track reconstruction for electrons has been improved for the 2012 data-taking period, especially for electrons which undergo significant energy loss due to bremsstrahlung in the detector, to achieve a high and uniform efficiency.

For each seed EM cluster² passing loose shower shape requirements of $R_\eta > 0.65$ and $R_{\text{had}} < 0.1$

²As in the 2011 electron reconstruction algorithm, clusters must satisfy loose requirements on the maximal fraction of energy deposited in the different layers of the EM calorimeter system: 0.9, 0.8, 0.98, 0.8 for the presampler detector, the strip,

(for the definition of these variables, see Table 1) a region-of-interest (ROI) with a cone-size of $\Delta R = 0.3$ around the seed cluster barycenter is defined. The collection of these EM cluster ROIs is retained for use in the track reconstruction.

Track reconstruction proceeds in two steps: pattern recognition and track fit. The standard pattern recognition [5] uses the pion hypothesis for energy loss at material surfaces. This has been complemented with a modified pattern recognition algorithm (based on a Kalman filter-smoother formalism [6]) which allows at most 30% energy loss at each material surface to account for possible bremsstrahlung.

If a track³ seed (consisting of three hits in different layers of the silicon detectors) with a transverse momentum larger than 1 GeV can not be successfully extended to a full track of at least seven hits using the pion hypothesis and it falls within one of the EM cluster ROIs, it is retried with the new pattern recognition using an electron hypothesis that allows for energy loss. In this way, a specific electron-oriented algorithm has been integrated into the standard track reconstruction which improves the performance for electrons and has minimal interference with the main track reconstruction.

Track candidates are then fitted either with the pion hypothesis or the electron hypothesis (according to the hypothesis used in the pattern recognition), using the ATLAS Global χ^2 Track Fitter [7]. If a track candidate fails the pion hypothesis track fit, for example, due to a high fit χ^2 (caused by large energy losses), it is refitted with the electron hypothesis.

Tracks are then considered as loosely matched to an EM cluster, if they pass either of the following two requirements:

- (i) Tracks with at least four silicon hits are extrapolated from the point of closest approach with respect to the primary vertex to the middle layer of the EM accordion calorimeter. They have to be either within 0.2 in ϕ of the EM cluster on the side the track is bending towards or within 0.05 on the opposite side. They also have to be within 0.05 in η of the EM cluster. TRT-only tracks, i.e. tracks with less than four silicon hits, are extrapolated from the last measurement point. They have to pass the same requirement for the difference in ϕ between track and cluster, $\Delta\phi$, as tracks with silicon hits but no requirement is placed on the difference in η between track and cluster, $\Delta\eta$, at this stage as their η coordinate is not measured precisely.
- (ii) The track extrapolated to the middle layer of the EM accordion calorimeter after rescaling its momentum to the measured cluster energy is either within 0.1 in ϕ to the EM cluster on the side the track is bending towards or within 0.05 on the opposite side. Furthermore, non-TRT-only tracks should be within 0.05 in η . As in (i), the track extrapolation is made from the last measurement point for TRT-only tracks and from the point of closest approach with respect to the primary vertex for tracks with silicon hits.

Criterion (ii) aims to recover tracks of typically low momentum that potentially suffered significant energy loss before reaching the calorimeter.

At this point, all electron-track candidates are defined. The track parameters of these candidates, for all but the TRT-only tracks, are precisely re-estimated using an optimized electron track fitter, the Gaussian Sum Filter (GSF) [8] algorithm, which is a non-linear generalization of the Kalman filter [6] algorithm. It yields a better estimate of the electron track parameters, especially those in the transverse plane, by accounting for the non-linear bremsstrahlung effects. TRT-only tracks and the very rare tracks (about 0.01%) that fail the GSF fit keep the parameters from the Global χ^2 Track Fit. These tracks are then used to perform the final track–cluster matching to build electron candidates and also to provide information for particle identification.

the middle and the back EM accordion calorimeter layers, respectively.

³The lower transverse momentum threshold for tracks reconstructed with the pion hypothesis is 400 MeV.

3.3 Electron-candidate reconstruction

An electron is reconstructed if at least one track is matched to the seed cluster. The efficiency of this matching and subsequent track quality cuts is measured as the reconstruction efficiency in Section 9. The track–cluster matching proceeds as described for the previous step in Section 3.2, but with the GSF refitted tracks and tighter requirements: in (i) the distance in ϕ needs to be within 0.1 (and not 0.2). Additionally, TRT-only tracks need to satisfy loose track–cluster matching criteria in η and tighter ones in ϕ : $|\Delta\eta| < 0.35$ (0.2) in the TRT barrel (endcap) and $|\Delta\phi| < 0.03$ (0.02) on the (opposite) side the track is bending towards. In this procedure more than one track can be associated with a cluster.

Although all tracks assigned to a cluster are kept for further analysis, the best matched one is chosen as the primary track which is used to determine the kinematics and charge of the electron and to calculate the electron identification decision. Thus the choice of the primary track is a critical step in the electron reconstruction chain. To favor the primary electron track and to avoid random matches between nearby tracks in case of cascades due to bremsstrahlung, tracks with at least one hit in the Pixel detector are preferred.

If more than one associated track has pixel hits, the distance between the track and the cluster is considered for each of the tracks, denoted as i and j in the following. Two angular distance variables are defined in the $\eta - \phi$ plane. ΔR is the distance between the cluster barycenter and the extrapolated track in the middle layer of the EM accordion calorimeter, while $\Delta R_{\text{rescaled}}$ is the distance between the cluster barycenter and the extrapolated track when the track momentum is rescaled to the measured cluster energy before the extrapolation to the middle layer. If $|\Delta R_{\text{rescaled},i} - \Delta R_{\text{rescaled},j}| > 0.01$, the track with the smaller $\Delta R_{\text{rescaled}}$ is chosen. If $|\Delta R_{\text{rescaled},i} - \Delta R_{\text{rescaled},j}| \leq 0.01$ and $|\Delta R_i - \Delta R_j| > 0.01$, the track with smaller ΔR is taken. For the rest of the cases, the two tracks have both similar $\Delta R_{\text{rescaled}}$ and similar ΔR , and the track with more pixel hits⁴ is chosen as primary track. A hit in the first layer of the pixel detector counts twice to prefer tracks with early hits.

All seed clusters together with their matching tracks, if there is at least one of them, are treated as electron candidates. Each of these electron clusters is then rebuilt in all four layers sequentially, starting from the middle layer, using 3×7 (5×5) cells in $\eta \times \phi$ in the barrel (endcaps) of the EM accordion calorimeter. The cluster position is adjusted in each layer to take into account the distribution of the deposited energy. These lateral cluster sizes have been optimized to take into account the different overall energy distributions in the barrel and endcap accordion calorimeters specifically for electrons⁵.

The cluster energy calibration has been improved for 2012 data [9]. In order to reduce the uncertainties on the measurement of the electron energy, the calibration procedure has been completely revisited exploiting the potential of multivariate techniques and correcting for several subtle instrumental effects. Furthermore, an in-depth investigation of the detector geometry and material distribution has resulted in an improved description by the Geant4 simulation.

After applying the electronic calibration to the calorimeter cells and building the energy clusters, a number of data pre-corrections are applied for measured time-dependent effects and non-perfectly corrected response in specific regions. Importantly, the presampler detector energy scales and the EM accordion calorimeter strip-to-middle-layer energy-scale ratios are also corrected.

The cluster energy is then determined from the energy in the three layers of the EM accordion calorimeter by applying a correction factor determined by linear regression using a multi-variate algorithm [10] trained on large samples of fully simulated single electron events. The input quantities used for electrons and photons are the total energy measured in the accordion calorimeter, the ratio of

⁴Throughout the note, when counting hits in the pixel and SCT detectors, non-operational modules that are traversed by the track are counted as hits.

⁵Unconverted (converted) photon clusters, which are used in the reconstruction efficiency measurement in Section 9, are built using 3×5 (3×7) towers in the barrel and 5×5 (5×5) towers in the endcap.

the presampler energy to the accordion energy, the shower depth⁶, the pseudorapidity of the cluster barycenter in the ATLAS coordinate system, and the η and ϕ positions of the cluster barycenter in the local coordinate system of the calorimeter. The inclusion of the cluster pseudorapidity allows to account for the passive material variations in front of the accordion. The last two variables allow to correct for the increase of lateral energy leakage for particles that hit a cell close to the edge and for the variation of the response as a function of the particle impact point with respect to the calorimeter absorbers.

As a last step, correction factors are derived in situ using a large sample of collected $Z \rightarrow ee$ events. They are applied to reconstructed electrons as final energy calibration in data events, while an energy smearing is done for simulated events.

The four-momentum of *central electrons* ($|\eta| < 2.47$) is computed using information from both the final cluster and the track best matched to the original seed cluster. The energy is given by the cluster energy. The ϕ and η directions are taken from the corresponding track parameters, except for TRT-only tracks for which the cluster ϕ and η directions are used.

4 Electron identification

Not all objects built by the electron reconstruction algorithms are signal electrons. Background objects include hadronic jets as well as background electrons from photon conversions, Dalitz decays and semi-leptonic heavy flavor hadron decays. In order to reject as much of these backgrounds as possible while keeping the efficiency for signal electrons high, electron identification in ATLAS is based on discriminating variables, which are combined into a menu of selections of varying background rejection. Both sequential cuts and multivariate analysis (MVA) techniques are employed.

In the *central region* of $|\eta| < 2.47$, variables describing the longitudinal and transverse shapes of the electromagnetic showers in the calorimeters, the properties of the tracks in the inner detector, as well as the matching between tracks and energy clusters are used to discriminate against the different background sources. These variables [11, 12, 3] are detailed in Table 1. Table 2 summarizes which variables are used for the different selections of the cut-based and likelihood (LH) identification menus.

4.1 Cut-based identification

The cut-based selections, *loose*, *medium*, *tight* and *multilepton*, are optimized in 10 bins in $|\eta|$ and 11 bins in E_T . This binning allows to take into account the variation of the electrons' characteristics due to e.g. dependence of the shower shapes on the amount of passive material traversed before entering the electromagnetic calorimeter. Shower shapes and track properties also change with the energy of the particle. The selection *tight* is a subset of *medium* and *loose*, while *medium* is a subset of *loose*. With increasing tightness, more variables are added and cuts are tightened on the variables already used in the looser selections.

Due to its simplicity, the cut-based electron identification [11, 12, 3], which is based on sequential cuts on selected variables, has been used by the ATLAS Collaboration for identifying electrons since the beginning of data-taking. In 2011, its performance (defined in terms of efficiency and background rejection) was improved by loosening cuts on existing variables and introducing additional variables, especially at the looser operating points [3]. In 2012, due to higher instantaneous luminosities provided by the LHC, the amount of overlapping collisions (pile-up) and therefore the number of particles in an event⁷ increased. Due to the higher amount of energy per event, the shower shapes, even of isolated electrons, tend to look more background-like. In order to cope with this, cuts were loosened on the

⁶The shower depth is defined as $X = \Sigma_i X_i E_i / \Sigma_i E_i$, where E_i and X_i are the cluster energy and the calorimeter thickness (in radiation lengths) in layer i , including the presampler detector layer where present.

⁷Here an "event" refers to a triggered bunch crossing with all its hard and soft pp interactions, as recorded by the detector.

Table 1: Definition of electron discriminating variables.

Type	Description	Name
Hadronic leakage	Ratio of E_T in the first layer of the hadronic calorimeter to E_T of the EM cluster (used over the range $ \eta < 0.8$ or $ \eta > 1.37$)	R_{Had1}
	Ratio of E_T in the hadronic calorimeter to E_T of the EM cluster (used over the range $0.8 < \eta < 1.37$)	R_{Had}
Back layer of EM calorimeter	Ratio of the energy in the back layer to the total energy in the EM accordion calorimeter	f_3
Middle layer of EM calorimeter	Lateral shower width, $\sqrt{(\sum E_i \eta_i^2)/(\sum E_i) - ((\sum E_i \eta_i)/(\sum E_i))^2}$, where E_i is the energy and η_i is the pseudorapidity of cell i and the sum is calculated within a window of 3×5 cells	$W_{\eta 2}$
	Ratio of the energy in 3×3 cells over the energy in 3×7 cells centered at the electron cluster position	R_ϕ
	Ratio of the energy in 3×7 cells over the energy in 7×7 cells centered at the electron cluster position	R_η
Strip layer of EM calorimeter	Shower width, $\sqrt{(\sum E_i (i - i_{\text{max}})^2)/(\sum E_i)}$, where i runs over all strips in a window of $\Delta\eta \times \Delta\phi \approx 0.0625 \times 0.2$, corresponding typically to 20 strips in η , and i_{max} is the index of the highest-energy strip	w_{stot}
	Ratio of the energy difference between the largest and second largest energy deposits in the cluster over the sum of these energies	E_{ratio}
	Ratio of the energy in the strip layer to the total energy in the EM accordion calorimeter	f_1
Track quality	Number of hits in the B-layer (discriminates against photon conversions)	n_{Blayer}
	Number of hits in the pixel detector	n_{Pixel}
	Number of total hits in the pixel and SCT detectors	n_{Si}
	Transverse impact parameter	d_0
	Significance of transverse impact parameter defined as the ratio of d_0 and its uncertainty	σ_{d_0}
	Momentum lost by the track between the perigee and the last measurement point divided by the original momentum	$\Delta p/p$
TRT	Total number of hits in the TRT	n_{TRT}
	Ratio of the number of high-threshold hits to the total number of hits in the TRT	F_{HT}
Track-cluster matching	$\Delta\eta$ between the cluster position in the strip layer and the extrapolated track	$\Delta\eta_1$
	$\Delta\phi$ between the cluster position in the middle layer and the extrapolated track	$\Delta\phi_2$
	Defined as $\Delta\phi_2$, but the track momentum is rescaled to the cluster energy before extrapolating the track to the middle layer of the calorimeter	$\Delta\phi_{\text{res}}$
	Ratio of the cluster energy to the track momentum	E/p
Conversions	Veto electron candidates matched to reconstructed photon conversions	isConv

Table 2: The variables used in the different selections of the electron identification menu.

Name	Cut-based				Likelihood		
	<i>loose</i>	<i>medium</i>	<i>tight</i>	<i>multilepton</i>	LOOSE	MEDIUM	VERY TIGHT
$R_{\text{Had}(1)}$	✓	✓	✓	✓	✓	✓	✓
f_3		✓	✓	✓	✓	✓	✓
$W_{\eta 2}$	✓	✓	✓	✓	✓	✓	✓
R_η	✓	✓	✓	✓	✓	✓	✓
R_ϕ					✓	✓	✓
w_{stot}	✓	✓	✓	✓	✓	✓	✓
E_{ratio}	✓	✓	✓	✓	✓	✓	✓
f_1					✓	✓	✓
n_{Blayer}		✓	✓	✓	✓	✓	✓
n_{Pixel}	✓	✓	✓	✓	✓	✓	✓
n_{Si}	✓	✓	✓	✓	✓	✓	✓
d_0		✓	✓			✓	✓
σ_{d_0}						✓	✓
$\Delta p/p$				✓	✓	✓	✓
n_{TRT}		✓	✓	✓			
F_{HT}		✓	✓	✓	✓	✓	✓
$\Delta\eta_1$	✓	✓	✓	✓	✓	✓	✓
$\Delta\phi_2$			✓				
$\Delta\phi_{\text{res}}$				✓	✓	✓	✓
E/p			✓				
isConv			✓				✓

most pileup-sensitive variables ($R_{\text{Had}(1)}$ and R_η) and tightened on others to keep the performance (efficiency/background rejection) roughly constant as a function of the number of reconstructed primary vertices. A cut on f_3 was added in 2012, as well. Furthermore a new operating point was added, called *multilepton*, which is optimized for the low energy electrons in the $H \rightarrow ZZ^* \rightarrow 4\ell$ analysis. For these electrons, *multilepton* has a similar efficiency to the *loose* operating point, but a better background rejection. In comparison to *loose*, cuts on the shower shapes are loosened and more variables are added, including those measuring bremsstrahlung effects.

4.2 Likelihood identification

Multivariate analysis (MVA) techniques are powerful, since they allow the combined evaluation of several properties when making a selection decision. Out of the different MVA techniques, the LH has been chosen for electron identification because of its simple construction.

The electron LH makes use of signal and background probability density functions (PDFs) of the discriminating variables. Based on these PDFs, which are treated as uncorrelated, an overall probability is calculated for the object to be signal or background. The signal and background probabilities for a given electron are combined into a discriminant $d_{\mathcal{L}}$ on which a cut is applied:

$$d_{\mathcal{L}} = \frac{\mathcal{L}_S}{\mathcal{L}_S + \mathcal{L}_B}, \quad \mathcal{L}_S(\vec{x}) = \prod_{i=1}^n P_{s,i}(x_i) \quad (1)$$

where \vec{x} is the vector of variable values and $P_{s,i}(x_i)$ is the value of the signal probability density function of the i^{th} variable evaluated at x_i . In the same way, $P_{b,i}(x_i)$ refers to the background probability function.

Signal and background PDFs used for the electron LH identification are obtained from data. As in the *multilepton* cut-based selection, variables measuring bremsstrahlung effects are included. Furthermore,

additional variables with significant discriminating power but also a large overlap between signal and background that prevents explicit cuts (like R_ϕ and f_1) can be included. The variables counting the hits on the track are not used as PDFs in the LH, but are left as simple cuts, as every electron should have a high quality track to allow for a robust 4-vector measurement.

The LOOSE, MEDIUM, and VERY TIGHT LH selections are designed to roughly match the electron efficiencies of *multilepton*, *medium* and *tight* cut-based selections, but to have better rejection of light-flavor jets and conversions⁸.

Each LH selection cuts on a LH discriminant, made with a different set of variables. The LOOSE LH features variables most useful for discrimination against light-flavor jets (in addition, a cut on n_{Blayer} is applied to reject conversions). In the MEDIUM and VERY TIGHT regimes, additional variables (d_0 , conversion bit) are added for further rejection of heavy-flavor jets and conversions. Although different variables are used for the different operating points, a sample of electrons selected using a tighter LH is a subset of the electron samples selected using the looser ones to a very good approximation.

The LH for each operating point consists of 9×6 sets of PDFs, divided into 9 $|\eta|$ bins and 6 E_T bins. This binning is similar, but coarser than the binning used for the cut-based operating points. It is chosen to balance the available data statistics with the variation of the PDF shapes in E_T and $|\eta|$.

4.3 Electron isolation

In order to further reject hadronic jets mis-identified as electrons, most analyses require electrons to pass some isolation cut in addition to the identification cuts described above. The two main alternatives are:

- Calorimeter based isolation:

The calorimetric isolation variable $E_T^{\text{cone}\Delta R}$ is defined as the sum of the transverse energy deposited in the calorimeter cells in a cone of ΔR around the electron, excluding the contribution within $\Delta\eta \times \Delta\phi = 0.125 \times 0.175$ around the electron cluster barycenter. It is corrected for energy leakage from the electron to the isolation cone and for the effect of pile-up using a correction parametrized as a function of the number of reconstructed primary vertices.

- Track based isolation:

The track isolation variable $P_T^{\text{cone}\Delta R}$ is the sum of the transverse momentum of the tracks with $p_T > 0.4$ GeV in a cone of ΔR around the electron, excluding the track of the electron itself. The tracks considered in the sum must come from the primary vertex associated to the electron track and be of good quality; i.e. they must have at least nine silicon hits, one of which should be in the innermost pixel layer.

Both types of isolation are used in the tag-and-probe measurements, mainly in order to tighten the selection criterion of the tag. Whenever isolation is applied to the probe in this work (this only happens in the J/ψ analysis described in Section 7.2), the criteria are chosen such that the effect on the efficiency is estimated to be small.

5 Efficiency measurement methodology

5.1 The tag-and-probe method

Measuring the identification and reconstruction efficiency requires a clean and unbiased sample of electrons. The method of choice is the tag-and-probe method, which, in the analyses described below, makes

⁸ An operating point, TIGHT, was also developed whose background rejection matches the background rejection of the *tight* cut-based selection. This operating point has a higher efficiency than VERY TIGHT and might be used in physics analyses in the future.

use of the characteristic signatures of the $Z \rightarrow ee$ and $J/\psi \rightarrow ee$ decays. In both cases, strict selection criteria are applied on one of the two decay electrons (called “tag”), and the second electron candidate (“probe”) is used for the efficiency measurements. Additional selection criteria on the event properties are applied to further reject background: Only events passing data-quality criteria, in particular concerning the inner detector and the calorimeters, are considered. Furthermore, at least one reconstructed primary vertex with at least three tracks should be present in the event. The tag-and-probe pairs must also pass requirements on their reconstructed invariant mass. In order to not bias the selected probe sample, each valid combination of electron pairs in the event is considered; an electron can be the tag in one pair and the probe in another.

The probe samples are contaminated by background objects (for example hadrons misidentified as electrons, electrons from semi-leptonic heavy flavor decays or from photon conversions). This contamination is estimated using either background template shapes or combined fits of background and signal analytical models to the data. The number of electrons is independently estimated at the probe level and at the level where the probe passes the tested criteria. The efficiency ϵ is defined as the fraction of probe electrons passing the tested criteria.

The efficiency to detect an electron in the ATLAS detector is not measured as a single quantity but is divided into different components, namely trigger, reconstruction and identification efficiency, as well as the efficiency to pass additional analysis criteria, like isolation. The full efficiency ϵ_{total} for a single electron can be written as:

$$\epsilon_{\text{total}} = \epsilon_{\text{reconstruction}} \times \epsilon_{\text{identification}} \times \epsilon_{\text{trigger}} \times \epsilon_{\text{additional}} \quad (2)$$

The efficiency components are defined and measured in a specific order to preserve consistency: The reconstruction efficiency, $\epsilon_{\text{reconstruction}}$, is measured with respect to clusters reconstructed in the electromagnetic calorimeter; the identification efficiency, $\epsilon_{\text{identification}}$, is determined with respect to reconstructed electrons. Trigger efficiencies are calculated for reconstructed electrons passing a given identification criterion. Therefore, for each identification selection a dedicated set of trigger efficiency measurements is performed, as documented in Ref. [2]. Additional selection criteria are often applied in analyses of collision data, for example on the isolation of electrons (introduced in Section 4.3). The efficiency of these selection criteria is dependent on the specific analysis and is therefore determined in each case separately and not covered in this note.

The determination of $\epsilon_{\text{identification}}$ and $\epsilon_{\text{reconstruction}}$ is described in Sections 7 and 9. The efficiencies are measured in data and in simulated $Z \rightarrow ee$ and $J/\psi \rightarrow ee$ samples. To compare the data values with the MC estimates, the same cuts are used to select the probe electrons. However, no background subtraction needs to be applied on the simulated samples; instead, we require the reconstructed electron track to be matched to an electron trajectory provided by the MC simulation within $\Delta R < 0.2$. In the matching procedure electrons coming from converted photons that are radiated off an electron originating from a Z or J/ψ decay are also accepted by the analyses. In the reconstruction efficiency measurement, electrons can be reconstructed as clusters without a matching track. To select these in the simulated $Z \rightarrow ee$ samples, the Z decay electrons provided by the MC simulation are matched to the reconstructed cluster within $\Delta R < 0.2$.

5.1.1 Data-to-MC correction factors

The accuracy with which the MC based detector simulation models the electron efficiency plays an important role in cross-section measurements and various searches for new physics. In order to achieve reliable physics results, the MC samples need to be corrected to reproduce the measured data efficiencies as closely as possible. This is achieved by a multiplicative correction factor defined as the ratio of the efficiency measured in data and to that in the simulation. These data-to-MC correction factors are

usually close to unity. Deviations stem from the mis-modeling of tracking properties or shower shapes in the calorimeters.

Since the electron efficiencies depend on the transverse energy and pseudorapidity, the measurements are performed in two-dimensional bins in (E_T, η) . Residual effects coming from kinematic differences of the physics processes used in the measurements are expected to cancel out in the data-to-MC efficiency ratio. This also applies to differences coming from e.g. using different physics objects such as $Z \rightarrow ee$ and $J/\psi \rightarrow ee$. Therefore, the combination of the different efficiency measurements is carried out using the data-to-MC ratios instead of the efficiencies themselves. The procedure for the combination is described in Section 7.3.

5.2 Determination of central values and uncertainties

For the evaluation of the results of the measurements and their uncertainties using a given final state ($Z \rightarrow ee$, $Z \rightarrow ee\gamma$ or $J/\psi \rightarrow ee$), the following approach has been chosen. In order to estimate the impact of the analysis choices and potential imperfections in the background modelling, different variations of the efficiency measurement are carried out, modifying for example the selection of the tag electron or the background estimation method. For the measurement of the data-to-MC correction factors, the same variations of the selection are applied consistently in data and MC. The final result (the central value) of a given efficiency measurement using one of the $Z \rightarrow ee$, $Z \rightarrow ee\gamma$ or $J/\psi \rightarrow ee$ processes is taken to be the average value of the results from all the variations (including that of the different background subtraction methods, e.g. Z_{iso} and Z_{mass} for the $Z \rightarrow ee$ final state).

The systematic uncertainty is estimated to be equal to the root mean square (RMS) of the measurements with the intention of modeling a 68% confidence interval. Therefore, if the RMS does not cover at least 68% of all the variations, an appropriate scaling of the uncertainty is applied to get the final uncertainty.

The statistical uncertainty is taken to be the average of the statistical uncertainties over all investigated variations of the measurement. The statistical uncertainty on a single variation of the measurement is calculated following the approach of Ref. [13].

6 Data and Monte Carlo samples

The results in this note are based on the pp collision data collected with the ATLAS detector in 2012. The data was produced by the LHC at a center of mass energy of $\sqrt{s} = 8$ TeV. After requiring good data quality, in particular concerning the inner detector and the electromagnetic and hadronic calorimeters, the total integrated luminosity used for the measurements is 20.3 fb^{-1} .

The measurements are compared to expectations from MC simulation. The $Z \rightarrow ee$ and $Z \rightarrow ee\gamma$ MC samples are generated by POWHEG [14, 15, 16] interfaced to PYTHIA8 [17]. The $J/\psi \rightarrow ee$ events are simulated using PYTHIA8 both for prompt ($pp \rightarrow J/\psi + X$) and for non-prompt ($b\bar{b} \rightarrow J/\psi + X$) production. All MC samples are processed through the full ATLAS detector simulation [18] based on GEANT4 [19].

The distributions of material in front of the presampler detector and the EM accordion calorimeter as a function of η are shown on the left of Fig. 1 for the detector material distribution used in Monte Carlo simulation. The contributions of the different detector elements up to the ID boundaries, including the services and thermal enclosures, are detailed on the right.

The peak in the amount of material before the EM accordion calorimeter at $|\eta| \approx 1.5$, corresponding to the transition region between the barrel and endcap EM accordion calorimeters, is due to the cryostats, the corner of the barrel electromagnetic accordion calorimeter, the inner detector services and the tile scintillator. The sudden increase of material at $|\eta| \approx 3.2$, corresponding to the separation between the

endcap calorimeters and the forward calorimeter, is mostly due to the cryostat that acts also as a support structure.

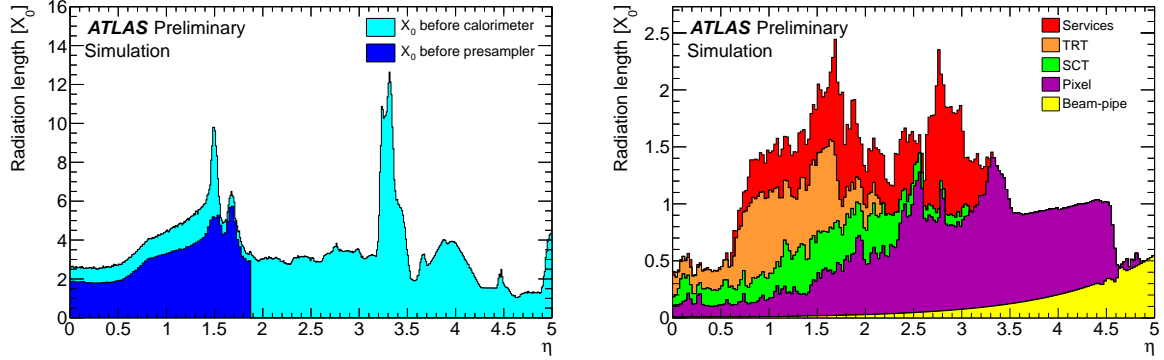


Figure 1: Amount of material, in units of radiation length X_0 , traversed by a particle as a function of η : (left) material in front of the presampler detector and the EM accordion calorimeter, and (right) material up to the ID boundaries. The contributions of the different detector elements, including the services and thermal enclosures are shown separately by filled color areas. The primary vertex position models the distribution in the data.

The simulation also includes realistic modeling (tuned to the data) of the event pile-up from the same and from previous and subsequent bunch crossings. The energy of the electron candidates in simulation is smeared to match the resolution in data and the MC events are weighted to reproduce the distributions of the primary vertex z position and the number of vertices in data, the latter being a good indicator of pile-up. Figure 2 shows the number of primary collision vertices for events with an electron probe (having $15 \text{ GeV} < E_T < 30 \text{ GeV}$ and $30 \text{ GeV} < E_T < 50 \text{ GeV}$) in the $Z \rightarrow ee$ data set used for the reconstruction efficiency measurement described in Section 9. The distribution does not depend on the probe electron transverse energy selection.

7 Identification efficiency measurement

The efficiencies of the identification criteria (cut-based *loose*, *medium*, *tight*, *multilepton* and LH LOOSE, MEDIUM, VERY TIGHT) are determined in data and in the simulated samples with respect to reconstructed electrons with associated tracks that have at least 1 hit in the pixel detector and at least 7 hits in the pixel or SCT detectors (this requirement is referred to as “track quality” below). The efficiencies are calculated as the ratio of the number of electrons passing a certain identification selection (numerator) to the number of electrons with a matching track passing the track quality requirements (denominator).

For the identification efficiencies determined in this note, three different decays of on-shell produced resonances are used: radiative decays of the Z boson, $Z \rightarrow ee\gamma$, for electrons with $10 \text{ GeV} < E_T < 15 \text{ GeV}$, $Z \rightarrow ee$ for electrons with $E_T > 15 \text{ GeV}$ and $J/\psi \rightarrow ee$ for electrons with $7 \text{ GeV} < E_T < 20 \text{ GeV}$. For all of these cases, the distributions of the probes passing the *tight* identification selection are depicted in Fig. 3 as a function of η (left) and E_T (right). This gives a good indication of the statistics available for each of the measurements in the respective η and E_T bin. The E_T spectrum of probes from $J/\psi \rightarrow ee$ is discontinuous, as the sample is selected by a number of triggers with different E_T thresholds as discussed in Section 7.2.

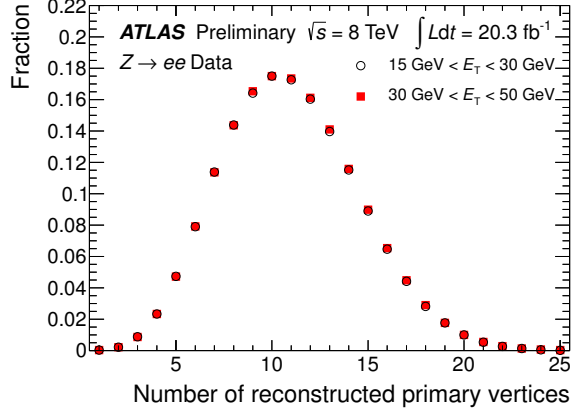


Figure 2: Number of reconstructed primary collision vertices for events with an electromagnetic cluster probe with $15 \text{ GeV} < E_T < 30 \text{ GeV}$ (open circles) and $30 \text{ GeV} < E_T < 50 \text{ GeV}$ (filled squares) in the $Z \rightarrow ee$ dataset used for the reconstruction efficiency measurement described in Section 9.

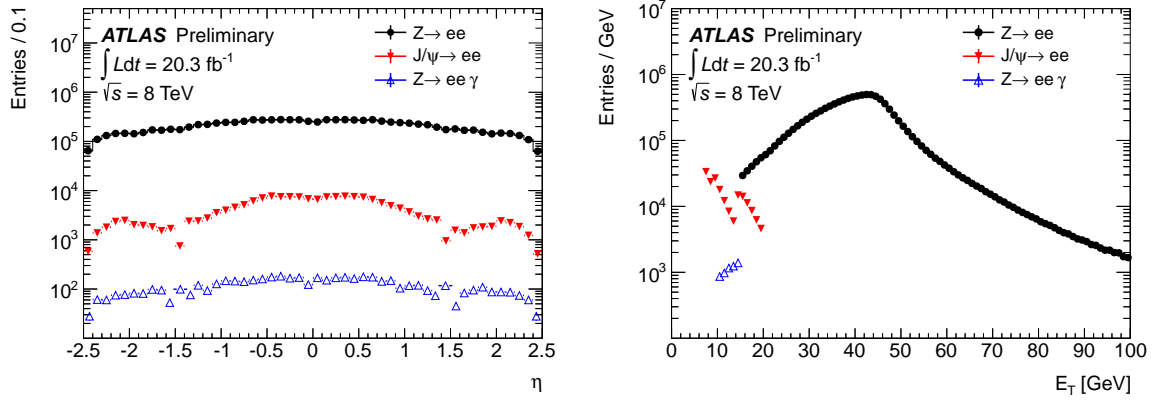


Figure 3: Probe electrons passing the *tight* identification criterion in the $Z \rightarrow ee$ (full circles), $Z \rightarrow ee\gamma$ (empty triangles) and the J/ψ (full triangles) samples as a function of η and E_T .

7.1 Tag-and-probe with $Z \rightarrow ee$ events

7.1.1 Introduction

$Z \rightarrow ee(\gamma)$ decays are used to measure the identification efficiency of electrons with $E_T > 10 \text{ GeV}$. Tag-and-probe using $Z \rightarrow ee$ decays provides a clean sample of electrons, especially when the probe electrons have $E_T > 25 \text{ GeV}$. For lower transverse energies, background subtraction becomes important. Two different distributions are used to discriminate signal electrons from background: the invariant mass of the tag-and-probe pair is used in the Z_{mass} method, and, as a systematic variation, the isolation distribution of the probe is used in the Z_{iso} method.

For probe electrons with E_T between 10 and 15 GeV, the invariant mass is computed from three objects: the tag, the probe and a photon. This explicitly selects probe electrons from $Z \rightarrow ee\gamma$ decays that have lost their energy due to final state radiation (FSR), a topology which, at low E_T , has a reduced background compared to $Z \rightarrow ee$ decays.

7.1.2 Event selection

Events are selected using an OR between two single electron triggers, one with an E_T threshold of 24 GeV requiring medium identification and one with an E_T threshold of 60 GeV and loose identification cuts⁹.

Events are required to have at least two reconstructed electron candidates in the central region of the detector, $|\eta| < 2.47$, with opposite charges¹⁰. One of the two electron candidates, the tag, is required to have a transverse energy $E_T > 25$ GeV, be matched to a trigger object with $\Delta R < 0.15$ and be outside the transition region between barrel and endcap of the electromagnetic calorimeter, $1.37 < |\eta| < 1.52$. Furthermore, it has to pass the *tight* identification cuts (*medium* for $Z \rightarrow ee\gamma$). The probes must have $E_T > 10$ GeV and pass the track quality criteria. The invariant mass of the tag-probe (tag-probe-photon for $Z \rightarrow ee\gamma$) system is required to be within 15 GeV of the Z mass. About 15.5 million electron probes are selected for further analysis.

For the $Z \rightarrow ee\gamma$ method, in addition to the tag and the probe, a photon is selected passing *tight* photon identification cuts [20] and fulfilling $E_T(\text{probe}) + E_T(\text{photon}) > 30$ GeV. Cuts are applied on the radial distance between the photon and the electron candidates to avoid double counting of objects: $\Delta R(\text{tag-photon}) > 0.4$ and $\Delta R(\text{probe-photon}) > 0.2$. The reason for the asymmetry between tag and probe cuts is an isolation cut with a cone of 0.4 which is applied to the tag as one of the systematic variations. Furthermore, FSR photons from the probe tend to be closer to the probe than the tag. Further selections are applied on the tag-probe and tag-photon invariant mass to select events with FSR: $m(\text{tag+photon}) < 80$ GeV, $m(\text{tag+probe}) < 90$ GeV. For a probe electron transverse energy of $10 \text{ GeV} < E_T < 15 \text{ GeV}$, about 13000 probes are selected integrated over the full $|\eta| < 2.47$ range.

7.1.3 Background estimation and systematic variations for the Z_{mass} method

The invariant mass of the tag and probe pair (and the photon in the case of $Z \rightarrow ee\gamma$) is used as the discriminating variable between signal electrons and background.

In order to form background templates, probes are chosen that are reconstructed as electrons with an associated track that passes track quality criteria. In addition, identification and isolation cuts are inverted to minimize the contribution of signal electrons. A study has been performed testing the shape biases of possible background templates due to cut inversion and contamination from signal electrons, and the least biased templates have been chosen. The remaining signal electron contamination in the background templates is estimated using simulated events.

The normalization of the background template is determined by a side-band method: For the denominator, the templates are normalized to the invariant mass distribution above the Z peak ($120 \text{ GeV} < m_{ee} < 250 \text{ GeV}$ for $Z \rightarrow ee$ and $100 \text{ GeV} < m_{ee\gamma} < 250 \text{ GeV}$ for $Z \rightarrow ee\gamma$). Care is taken to remove the contribution of signal electrons in the tails of the distribution of all probes before normalizing the background template to them. For the numerator, the same templates are used as in the denominator, but they are normalized to the same-sign invariant mass distribution (all numerator cuts are applied on the probe). The normalization is done in the same ranges as in the denominator. The same-sign distribution is used as reference because it has less signal contamination than the opposite-sign distribution, an effect that is more important in the numerator. Figure 4 shows the $Z \rightarrow ee$ tag and probe invariant mass distribution in one example bin for both the numerator and the denominator, including the normalized background

⁹The electron identification selection in the trigger is looser than or equivalent to the corresponding analysis cuts.

¹⁰The charge-identification efficiency is not measured in this note. It depends on the passive detector material distribution before the EM calorimeter and thus decreases with increasing pseudorapidity. It was measured in 2011 ATLAS data [3] to be above 99% for reconstructed electrons everywhere in barrel region, decreasing to 93% at the boundaries of the tracking acceptance. It increases with tightening identification requirements. The agreement between data and simulation was found to be good except at the outermost edge of the acceptance where the simulation predicts a higher misidentification probability. The material description in the simulation has been improved recently in that region.

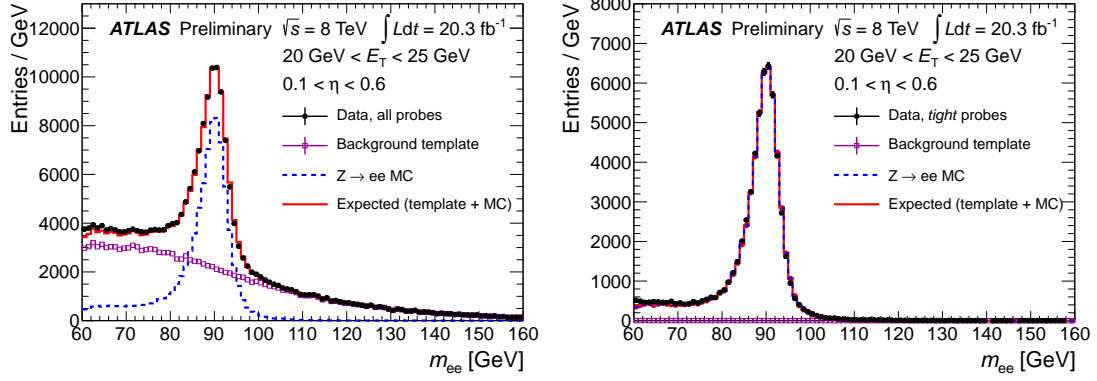


Figure 4: Illustration of the background estimation using the Z_{mass} method in the $20 \text{ GeV} < E_T < 25 \text{ GeV}$, $0.1 < \eta < 0.6$ bin, at reconstruction+track-quality level (left) and for probes passing the cut-based *tight* identification (right). The background template is normalized in the range $120 \text{ GeV} < m_{ee} < 250 \text{ GeV}$. The tag electron passes cut-based *medium* and isolation cuts. The signal MC is scaled to match the total estimated signal in the Z -mass window.

template and the MC $Z \rightarrow ee$ prediction. Figure 5 shows the same for the $Z \rightarrow eey$ invariant mass distribution.

In order to assess systematic uncertainties, the efficiency measurements from the following analysis variations are considered: The mass window is changed from 15 GeV to 10 and 20 GeV around the Z mass, the tag requirement is varied by applying a cut on the calorimetric isolation variable $E_T^{\text{cone}0.4} < 5 \text{ GeV}$ and, in the $Z \rightarrow ee$ case, by loosening the identification requirement to *medium*. Furthermore, for $E_T < 30 \text{ GeV}$, two normalization regions, below and above the Z peak are used. The normalization range below the peak is $60 \text{ GeV} < m_{ee} < 70 \text{ GeV}$. For $E_T > 30 \text{ GeV}$, the number of events in the low mass region is too small for a reliable normalization, so instead two different background template selections are considered. All possible combinations of these variations are produced and taken into account as described in Section 5.2.

7.1.4 Background estimation and systematic variations for the Z_{iso} method

The calorimeter isolation distribution $E_T^{\text{cone}0.3}$ of the probe electrons is used as a discriminating variable.

The background templates are formed as subsets of all probes used in the denominator of the identification efficiency calculation. The electrons for the background template are required to be reconstructed as electrons with a matching track that passes track quality criteria, however they are required to fail some of the identification cuts, namely the requirements of w_{stot} and the F_{HT} . A study has been performed on possible background templates and the bias due to cut inversion and contamination from signal electrons. The least biased templates have been chosen. As illustrated in Fig. 6, the background templates are normalized to the isolation distribution of the probes using the background dominated upper end of the isolation distribution.

To assess the systematic uncertainty of the efficiency, the parameters of the measurement procedure are varied. The threshold for the sideband subtraction is chosen between $E_T^{\text{cone}0.3} > 10 \text{ GeV}$ and $> 15 \text{ GeV}$. As in the Z_{mass} case, the mass window is changed from 15 GeV to 10 and 20 GeV around the Z mass, the tag requirement is varied by applying a cut on the calorimetric isolation variable $E_T^{\text{cone}0.4} < 5 \text{ GeV}$. In addition, different identification cuts are inverted to form two alternative templates and an alternative isolation variable with a larger isolation cone size $E_T^{\text{cone}0.4}$ is used as discriminant. As in the Z_{mass} case, all possible combinations of these variations are considered.

For Z_{mass} and Z_{iso} method together, there are in total 90 systematic variations, which are treated as

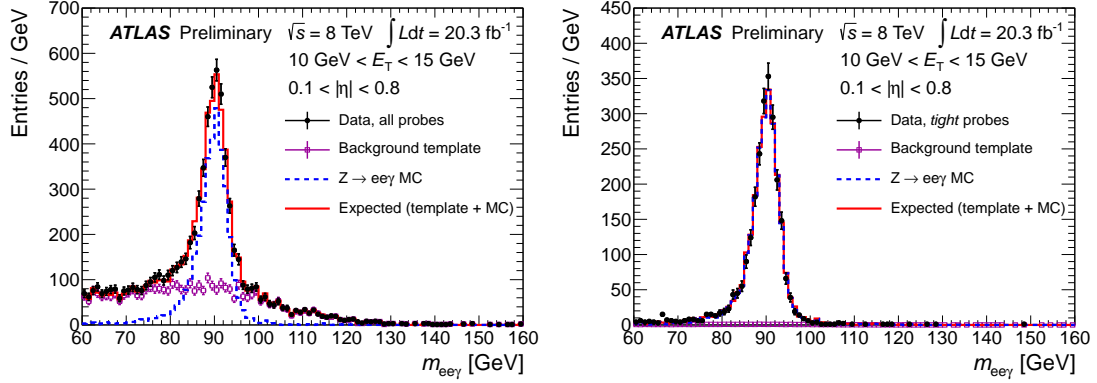


Figure 5: Illustration of the background estimation using the $Z \rightarrow ee\gamma$ method in the $10 \text{ GeV} < E_T < 15 \text{ GeV}$, $0.1 < |\eta| < 0.8$ bin, at reconstruction+track-quality level (left) and for probes passing the cut-based *tight* identification (right). The background template is normalized in the range $100 \text{ GeV} < m_{ee} < 250 \text{ GeV}$. The tag electron passes cut-based *medium* and isolation cuts. The signal MC is scaled to match the total estimated signal in the Z -mass window.

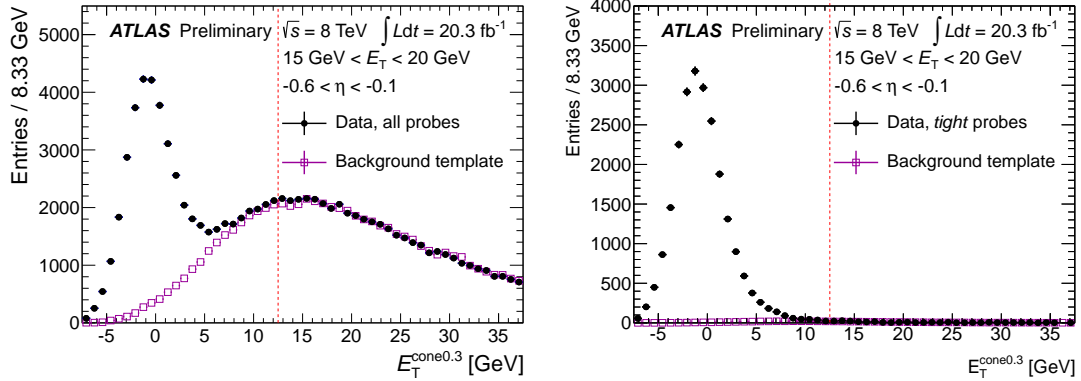


Figure 6: Illustration of the background estimation using the Z_{iso} method in the $15 \text{ GeV} < E_T < 20 \text{ GeV}$, $-0.6 < \eta < -0.1$ bin, at reconstruction+track-quality level (left) and after applying the *tight* cuts (right). The tag electrons are selected using the cut-based *tight* identification and a moderate Z mass window of 15 GeV is applied. The threshold chosen for the sideband subtraction is $E_T^{\text{cone}0.3} = 12.5 \text{ GeV}$.

variations of the same measurement. This allows to cover the systematic uncertainty due to the background estimation method, as the two methodologies use completely different approaches.

7.2 Tag-and-probe with $J/\psi \rightarrow ee$ events

7.2.1 Introduction

$J/\psi \rightarrow ee$ events are used to measure the electron identification efficiency for $7 \text{ GeV} < E_T < 20 \text{ GeV}$. At such low energies, the probe sample suffers from a significant background fraction, which can be estimated using the reconstructed di-electron invariant mass (m_{ee}) of the selected tag-and-probe pairs. Furthermore, the J/ψ sample is composed of two contributions. In *prompt production*, the J/ψ meson is produced directly in the proton-proton collision via strong interaction or from the decays of directly produced heavier charmonium states. The electrons from the decay of prompt J/ψ particles are expected to be isolated and therefore to have identification efficiencies close to those of isolated electrons from other physics processes of interest in the same transverse energy range, such as Higgs-boson decays. In *non-prompt production*, the J/ψ meson comes from b -hadron decays and its decay electrons are expected to be less isolated.

Experimentally, the two production modes can be distinguished by measuring the displacement of the $J/\psi \rightarrow ee$ vertex with respect to the primary vertex. Due to the long lifetime of b -hadrons, electron-pairs from non-prompt J/ψ production will have a measurably displaced vertex, while prompt decays occur at the primary vertex. To reduce the dependence on the J/ψ transverse momentum, the variable used in this analysis to discriminate prompt and non-prompt production, called pseudo-proper time [21], is defined as

$$\tau = \frac{L_{xy} \cdot m_{\text{PDG}}^{J/\psi}}{p_T^{J/\psi}}. \quad (3)$$

Here, L_{xy} measures the displacement of the J/ψ vertex w.r.t. the primary vertex in the transverse plane, while $m_{\text{PDG}}^{J/\psi}$ and $p_T^{J/\psi}$ are the mass [22] and the reconstructed transverse momentum of the J/ψ particle.

Two methods have been developed to measure the electron efficiency using $J/\psi \rightarrow ee$ decays. The *short- τ method*, already used in Ref. [12, 3], considers only events with short pseudo-proper time, selecting a subsample dominated by prompt J/ψ production. The remaining non-prompt contamination is estimated using MC simulation and the ATLAS measurement of the non-prompt fraction in $J/\psi \rightarrow \mu\mu$ events [23]. The *τ -fit method*, used in Ref. [3], utilizes the full τ -range and extracts the non-prompt fraction by fitting the pseudo-proper time distribution both before and after applying the identification cuts.

7.2.2 Event selection

Events are selected by five dedicated $J/\psi \rightarrow ee$ triggers. These require tight trigger electron identification¹¹ and an electron E_T above a threshold for one of the two trigger objects, while only requiring an electromagnetic cluster above a certain E_T threshold for the other.

Events with at least two electron candidates with a transverse energy $E_T > 5 \text{ GeV}$ and within the pseudorapidity range $|\eta| < 2.47$ are considered. The tag electron candidate must be matched to a tight trigger electron object within $\Delta R < 0.005$ and satisfy the cut-based *tight* identification selection. To further clean the tag electron sample an isolation criterion is applied in most of the analysis variations. The other electron, the probe, needs to pass the track quality criteria. It is also required to match an electromagnetic trigger object of the $J/\psi \rightarrow ee$ triggers within $\Delta R < 0.005$ and have a transverse energy

¹¹The tight electron identification selection applied in the J/ψ trigger is looser than the corresponding analysis cuts. In particular, no selection is applied on $\Delta\phi_2$, E/p and isConv .

that is at least 1 GeV higher than the corresponding trigger threshold. To ensure that the measured efficiency corresponds to well-isolated electron objects an isolation cut is applied on the probe as well. The isolation criterion has less than 1% effect on the identification efficiency in simulated events. It is further required that the tag and the probe candidates are separated by $\Delta R_{\text{tag-probe}} > 0.2$ to prevent one object from affecting the identification of the other. The pseudo-proper time of the reconstructed J/ψ candidate is restricted to $-1 \text{ ps} < \tau < 3 \text{ ps}$ in the τ -fit method and typically to $-1 \text{ ps} < \tau < 0.2 \text{ ps}$ in the short- τ method. At this stage no requirement is made on the charge of the electrons and all possible tag and probe pairs are considered. About 700000 electron probes are selected for $E_T = 7 - 20 \text{ GeV}$, of which about 190000 pass the *tight* selection, within the wider range of $-1 \text{ ps} < \tau < 3 \text{ ps}$ and integrated over $|\eta| < 2.47$.

7.2.3 Background estimation and systematic variations

The invariant mass of the tag-and-probe pair is used to discriminate signal electrons against background. The most important contribution to the background, even after requiring the tag-and-probe pair to have opposite-sign (OS) charges, comes from random combinations of two particles which do not originate from a resonance decay. This can be evaluated – assuming charge symmetry – using the mass spectrum of same-sign (SS) charge pairs. The remaining background is small and can be described using an analytical model. For this, the invariant mass distribution of the two electrons is fitted with the sum of three contributions: J/ψ , $\psi(2S)$ and background, typically in the range of $1.8 \text{ GeV} < m_{ee} < 4.6 \text{ GeV}$. To model the J/ψ component, a Crystal-Ball [24, 25] (or alternatively to better describe the tail a Crystal-Ball + Gaussian) function is used. The $\psi(2S)$ is modeled with the same shape except for an offset corresponding to the mass difference between the J/ψ and $\psi(2S)$ states. Finally the residual background is modeled by a Chebychev polynomial (or alternatively by an exponential function) with the parameters determined from a combined signal + background fit to the data. The background estimated using SS events is either added to the residual background in the binned fit (see Fig. 7 for the short- τ method) or subtracted explicitly before performing the unbinned fit (see Fig. 8 for the τ -fit method). The number of J/ψ candidates is counted within a mass window of $2.8 \text{ GeV} < m_{ee} < 3.3 \text{ GeV}$.

In the τ -fit method, the prompt component is then extracted by an unbinned fit of the pseudo-proper time distribution in the range $-1 \text{ ps} < \tau < 3 \text{ ps}$, after correcting the contribution for the estimated background by subtracting the τ distribution in the mass sidebands $2.3 \text{ GeV} < m_{ee} < 2.5 \text{ GeV}$ and $4.0 \text{ GeV} < m_{ee} < 4.2 \text{ GeV}$ normalized to the estimated background within the signal mass window as given by the m_{ee} fit. The non-prompt component is modeled by an exponential decay function convoluted with the sum of two Gaussians, while the shape of the prompt component is described by the sum of the same Gaussians describing the detector resolution, as shown in Fig. 9.

In the short- τ method, strict requirements on τ are made, requiring it to be below 0.2 or 0.4 ps. The resulting non-prompt contamination is below $\sim 20\%$, decreasing with decreasing probe E_T . The measured efficiency is compared to the MC prediction after mixing the simulated prompt and non-prompt $J/\psi \rightarrow ee$ samples according to the ATLAS measurement of the non-prompt J/ψ fraction in the di-muon final state at $\sqrt{s} = 7 \text{ TeV}$ [23].

Systematic uncertainties arise predominantly from the background estimation and the probe definition. They are estimated by varying the tag-and-probe selection (such as the isolation and the τ requirements), the fit parameters (background and signal shapes, fit window and sideband definitions) and the size of the mass window (changed by $\pm 40\%$) for signal counting after the mass fit. In total 186 systematic variations have been considered in each (E_T, η) bin, using the two methods as variations of the same measurement, to determine the efficiency and its uncertainty.

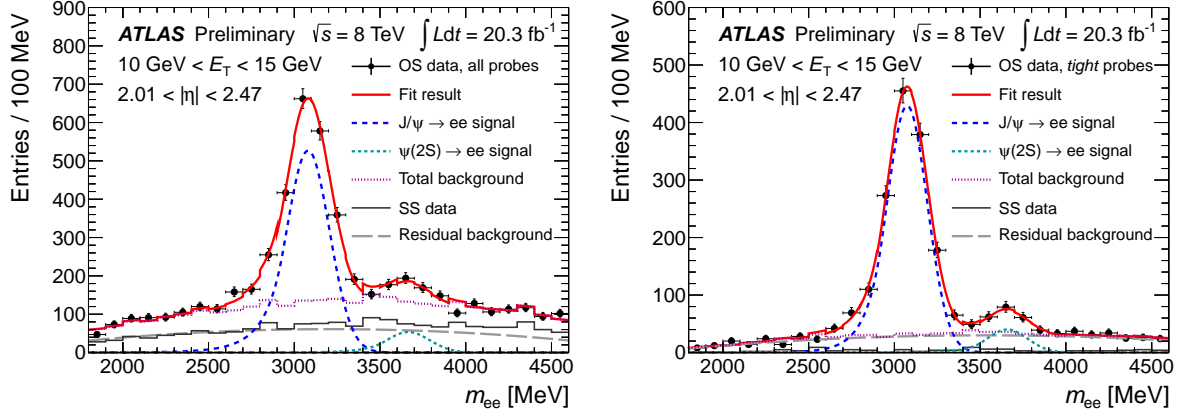


Figure 7: The figure demonstrates the background subtraction as carried out in the short- τ method. Shown is the di-electron invariant mass fit for all electron probes having a good-track quality (left) and for probes passing the cut-based *tight* identification (right) for $10 \text{ GeV} < E_T < 15 \text{ GeV}$ and $2.01 < |\eta| < 2.47$. Dots with error bars represent the opposite-sign (OS) pairs for data, the fitted J/ψ signal is shown by the dashed blue and the $\psi(2S)$ by the dashed light blue lines (both modeled by a Crystal-Ball function). A background fit is carried out using the sum of the same-sign (SS) distribution (solid gray) from data and a Chebychev polynomial of 2nd order describing the residual background (dashed gray). The sum of the two background contributions is depicted as a purple dotted line. A track isolation cut of $P_T^{\text{cone}0.2}/E_T < 0.15$ is applied on the probe. The pseudo-proper time is required to be $-1 \text{ ps} < \tau < 0.2 \text{ ps}$.

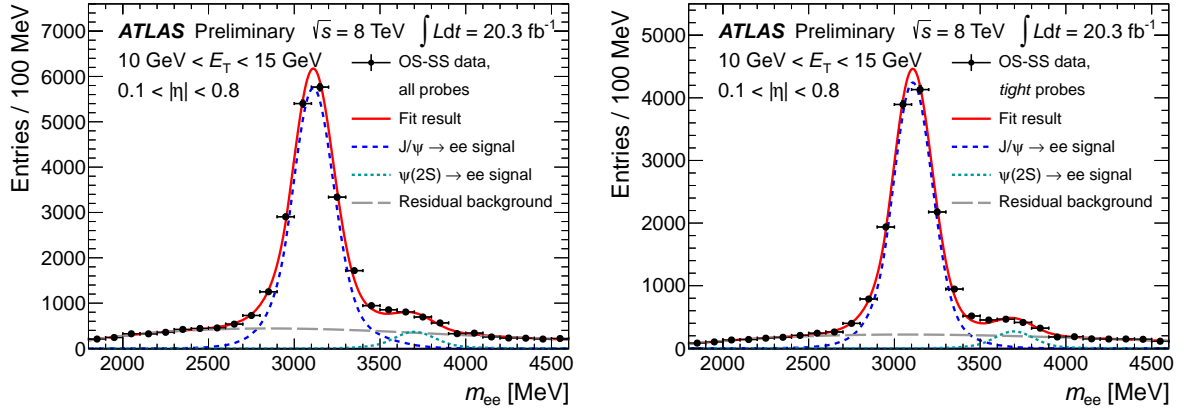


Figure 8: Illustration of the background determination for the J/ψ analysis, in the τ -fit method. The di-electron invariant mass fit for all probes passing track-quality requirements (left) and for probes passing the cut-based *tight* identification (right) for $10 \text{ GeV} < E_T < 15 \text{ GeV}$ and $0.1 < |\eta| < 0.8$ is shown. Dots with error bars represent the OS minus SS data, the fitted J/ψ signal is shown by the dashed blue and the $\psi(2S)$ by the dashed light blue lines (both modeled by Crystal-Ball + Gaussian function). The residual background (Chebychev polynomial of 2nd order) is shown by the dashed gray line. A track isolation cut of $P_T^{\text{cone}0.2}/E_T < 0.15$ is applied on both the tag and the probe. The pseudo-proper time is required to be $-1 \text{ ps} < \tau < 3 \text{ ps}$.

7.3 Combination

To calculate the final results on the identification efficiency, the data-to-MC correction factors are combined. This has the advantage that efficiency differences due to kinematic differences between the pro-

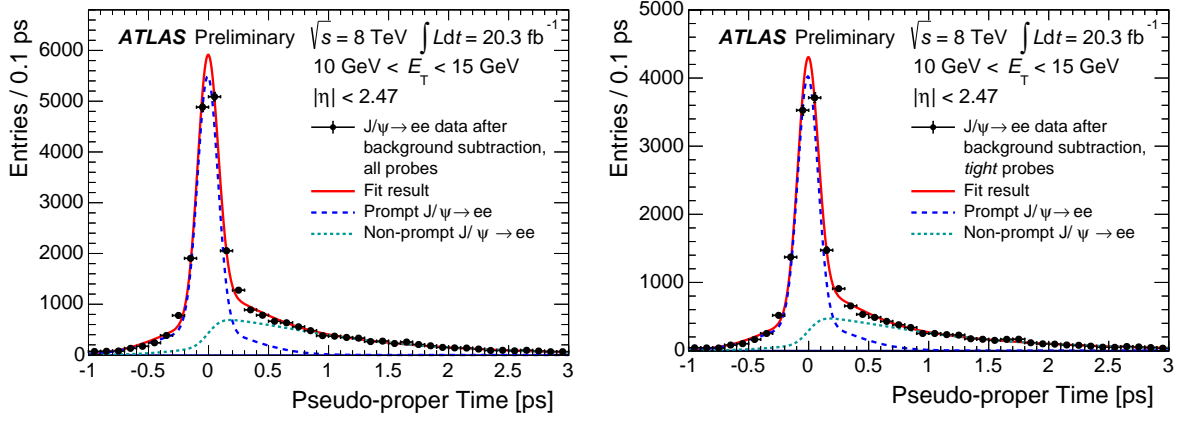


Figure 9: Pseudo-proper time fit for all probes passing reconstruction+track-quality criteria (left) and for probes passing the *tight* identification criteria (right) for $10 \text{ GeV} < E_T < 15 \text{ GeV}$, integrated over $|\eta| < 2.47$. Dots with error bars represent the OS minus SS data with the residual background subtracted using the reconstructed di-electron mass distribution sidebands. The signal prompt component is shown by the dashed blue line (sum of two Gaussians) and the signal non-prompt component is shown by the light blue dashed line (exponential decay function convoluted with the sum of two Gaussians). A calorimetric isolation cut of $E_T^{\text{cone}0.2}/E_T < 0.2$ is applied on the probe in this case.

cesses are expected to cancel out. The following measurements are used in the different E_T bins:

- 7-10 GeV: $J/\psi \rightarrow ee$,
- 10-15 GeV: $J/\psi \rightarrow ee$ and $Z \rightarrow e e \gamma$,
- 15-20 GeV: $J/\psi \rightarrow ee$ and $Z \rightarrow ee$,
- 20 GeV and above: $Z \rightarrow ee$.

Only the two E_T bins 10-15 GeV and 15-20 GeV allow a combination of independent measurements, which is done using a code originally developed for the HERA experiment [26] and used in Ref. [3]. It performs a χ^2 fit over all bins, adjusting the input values taking into account correlations of the systematic uncertainties in η and E_T bins.

Both the χ^2 (ranging from 3.4 to 12.3 for 12 degrees of freedom, depending on the identification selection) and the pulls of the combination indicate good agreement for the measurements in the 10-15 GeV and 15-20 GeV bins.

7.4 Results

The combined data efficiencies are derived by applying the combined data-to-MC efficiency ratios to the MC efficiency prediction from simulated $Z \rightarrow ee$ decays. Similarly, when comparing the results of different measurements, the measured data-to-MC efficiency ratios are used to correct the $Z \rightarrow ee$ MC sample.

The measured efficiencies for the various identification criteria are presented as functions of the electron η , E_T and the number of reconstructed primary collision vertices in the event. The latter is a measure of the amount of activity due to overlapping collisions which affects the reconstructed electron object, for example, by making the calorimeter shower shapes more background like due to close-by particles. The efficiency dependence in bins of primary vertices is only measured for electrons with

$E_T > 15$ GeV using $Z \rightarrow ee$ events with the Z_{mass} method, as the $J/\psi \rightarrow ee$ sample size is not large enough.

Figure 10 shows a comparison between efficiencies computed for $Z \rightarrow ee$ decays in the two E_T bins in which different measurements overlap. The agreement between the methods is good.

Figures 11 and 12 show overlays of the combined data efficiencies for the different identification criteria in all considered (E_T, η) bins. The efficiency integrated over E_T or η , as well as the dependence on the number of primary vertices is shown in Fig. 13. These distributions assume the E_T, η distribution of electrons from $Z \rightarrow ee$ decays.

With tighter cuts on more variables, the overall identification efficiency decreases, while the dependence on E_T and η increases, as expected. The efficiency of the *multilepton* cuts show less variation with the number of primary vertices than the *loose* cuts, as they rely less on the pile-up sensitive variables R_η and R_{Had} . Overall, the 2012 update of the cut-based menu (see Section 4.1) was quite successful, and the remaining pile-up dependence is small (variation below 4% from 1-30 vertices). The improvement of the 2012 menu regarding the pile-up robustness of the cuts is demonstrated in Fig. 14, where the efficiencies for the *loose*, *medium* and *tight* cuts as a function of the number of reconstructed primary vertices are compared for 2011 and 2012.

The LOOSE LH is tuned to match the efficiencies of the *multilepton* cuts, while the (MEDIUM) VERY TIGHT LH is tuned to match those of the (*medium*) *tight* cuts. The efficiency figures show that this tuning was successful in almost all bins. While the efficiencies match, the background rejection of the LH selections is better. The background efficiencies have been estimated both in data and in MC simulation, as described in Section 8: for hadrons and conversions, they are reduced by a factor of about two when comparing the cut-based identification to the corresponding LH operating points.

The efficiencies as a function of E_T and η , for example in Fig. 13, show some well understood features. As expected, the identification efficiencies in general rise as a function of E_T because electrons with higher E_T are better separated from the background in many of the discriminating variables. For the lowest (7-10 GeV) as well as for the highest (above 80 GeV) E_T bin, a significant and somewhat discontinuous increase in the identification efficiency can be observed. This is explained by the fact that at very low and very high E_T some cuts are relaxed. For the high E_T bin the E/p cut is dropped, because the measurement of the electron's track momentum becomes unreliable due to its poor resolution for high p_T and can therefore not safely be used to distinguish electrons from backgrounds. It was checked that the data-to-MC correction factor measured for electrons above 80 GeV is applicable to electrons even at energies of $E_T > 400$ GeV using the Z_{iso} method. Within the large statistical uncertainties, data-to-MC correction factors binned in E_T for the high- E_T region were found to agree with the combined data-to-MC correction factor above 80 GeV that is presented in this note. The lowest E_T bin (7-10 GeV) was tuned separately from the other bins, choosing the signal efficiency to be a bit higher which leads to higher background contamination.

The shape of the identification efficiencies as a function of η is predominantly due to features of the detector design and the selection optimization procedure that is typically based on the signal-to-background ratio. A small gap between the calorimeter wheels and in the TRT around $|\eta| \approx 0$ explains the slight drop in efficiency. Another, larger drop in efficiency can be observed between $1.37 < |\eta| < 1.52$, where the transition region between the barrel and endcap calorimeters is situated. At high $|\eta|$ the efficiencies are lower due to the larger amount of material in front of the endcap calorimeters.

In general, all of these broad features of the detector described above are well modeled in the simulation, which shows the same general behavior as a function of η . There are however some small differences between data and MC efficiencies. Comparisons between the data and MC efficiencies are shown in Figs. 15 to 17 for the cut-based *multilepton*, *loose*, *medium* and *tight* cuts and in Fig. 18 to 20 for LOOSE, MEDIUM and VERY TIGHT LH. These figures depict in their lower panels the data-to-MC correction factor introduced in section 5.1.1, which are applied to the MC simulation in physics analyses. As can

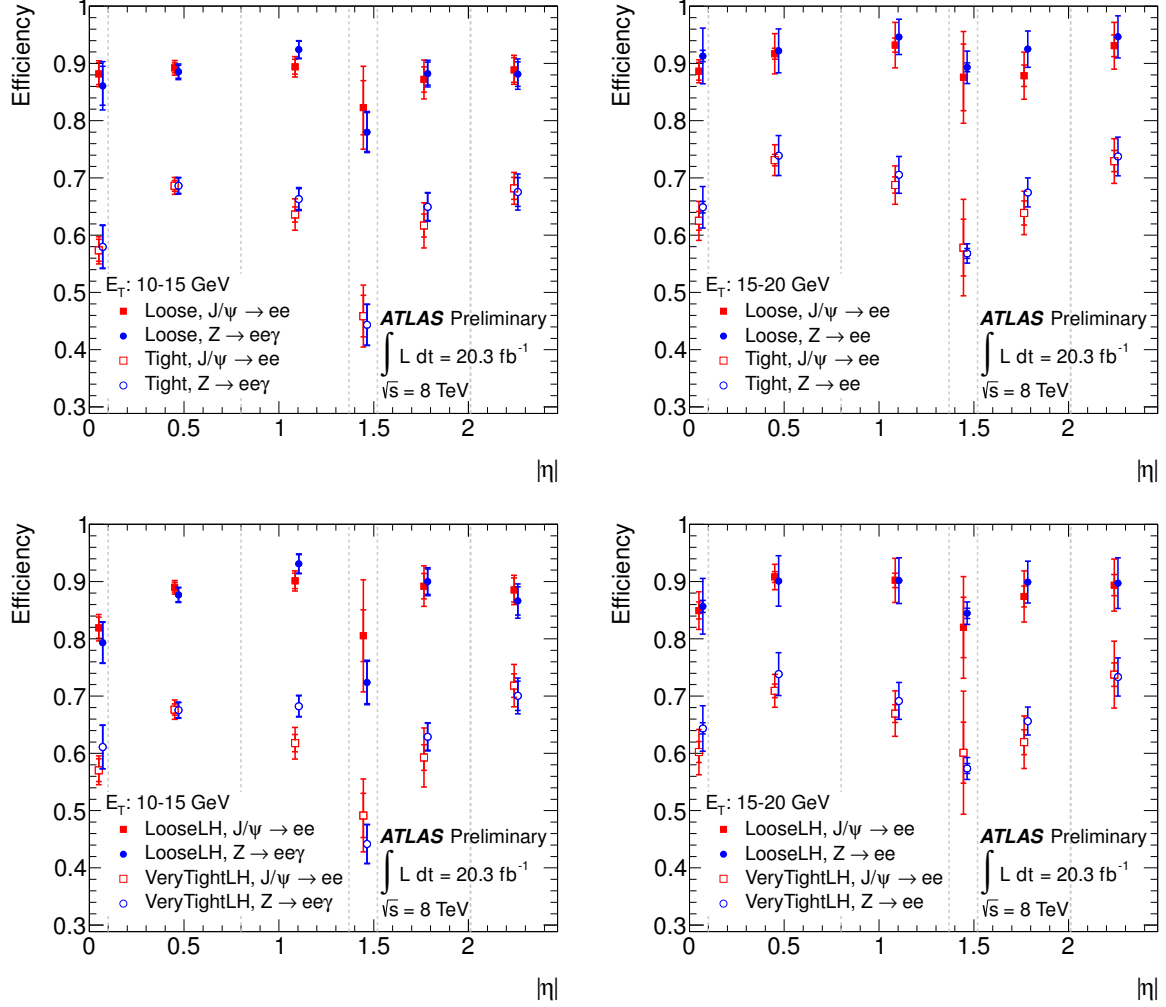


Figure 10: Measured identification efficiency as a function of $|\eta|$ for $E_T = 10 - 15$ GeV (left) and $E_T = 15 - 20$ GeV (right) for the cut-based *loose* and *tight* selections (top) and for LOOSE and VERY TIGHT LH (bottom). The data efficiency is derived by applying the measured data-to-MC efficiency ratios, determined with either the J/ψ or the Z methods, on the MC prediction from $Z \rightarrow ee$ decays. The uncertainties are statistical (inner error bars) and statistical+systematic (outer error bars). The dashed lines indicate the bins in which the efficiencies are calculated.

be seen, these correction factors are very close to one, with cut-based selections showing a better data - MC agreement than LH ones. Only for low E_T or high values of η , corrections reaching 10% have to be applied for the more stringent selection criteria. The uncertainties on the data-to-MC correction factors range from 0.5% to 10%, with the highest uncertainties found in the transition region of the calorimeter, $1.37 < |\eta| < 1.52$. A large contribution to the uncertainties is of statistical nature and is considered uncorrelated between bins when propagating the uncertainties to the final results of analyses.

Figures 21 and 22 show the identification efficiencies when integrated over E_T or η , and as a function of the number reconstructed primary vertices. As discussed in detail in Ref. [12], the discrepancy between identification efficiencies in data and MC can be tracked down to differences in the distribution of the variables used in the identification, particularly the shower shape variables and the TRT high threshold hit ratio F_{HT} , the latter being defined only for $|\eta| < 2$. The distribution of the lateral shower shapes are not well modelled by the GEANT4-based simulation of the detector: in comparison to MC predictions, most shower shapes in data are wider and shifted more towards the background distributions. These effects lead to higher efficiencies in MC. F_{HT} , on the other hand, is underestimated in the simulation for $|\eta| > 1$, leading to higher efficiencies in data than in the simulation. These two effects cancel each other, as can be seen in Fig. 21, where the data and MC efficiency values of the *tight* cuts are quite close to each other for $1 < |\eta| < 2$.

Figures 21 and 22 show that the data has a more significant dependence on pile-up than predicted by simulation. For the *multilepton* and *loose* cuts, the data-to-MC ratio is almost constant as a function of the number of primary vertices, while it decreases for the *medium* and *tight* cuts as well as the likelihood operating points by about 2% from 1 to 30 primary vertices.

This effect is primarily caused by the mis-modeling in MC of the $R_{had(1)}$, w_{stot} and F_{HT} variables. The F_{HT} variable is sensitive to the pile-up conditions due to higher occupancies in events with many vertices, which can lead to hit overlaps in the TRT straws increasing the chance of a high threshold hit. The effect is not expected to be particularly well modeled by the simulation, independent of the modeling of the pile-up itself. Both the $R_{Had(1)}$ and w_{stot} variables are known to be not well modeled by GEANT4. Additional energy deposits from pile-up particles are thus not expected to be well modeled for these variables either, leading to differences as a function of pile-up between data and MC. The pile-up profile of the collision data analyses which use the results of these efficiency measurements is very close to the pile-up profile of the efficiency measurements presented here. The data-to-MC correction factors will therefore adjust the MC efficiencies in the collision data analyses for the residual pile-up dependence.

In general, the mis-modeling of the distributions affects cut-based and LH selections differently. For cut-based selections, a mis-modeling in MC is reflected in the efficiency only if it occurs around the cut value. In the case of the LH, a mis-modeling anywhere in the distribution can affect the efficiency. The harder the cut on the discriminant of the LH, the larger the effect of the differences between data and MC distributions on the data-to-MC correction factors, as can be seen in Figs. 18-20 and 22.

8 Identification efficiency for background processes

The three main categories of electron background (in descending order of abundance after electron reconstruction) are light-flavor hadrons, background electrons from conversions and Dalitz decays, and non-isolated electrons from heavy flavor decays. The background efficiencies of the different identification selections have been studied using both MC simulation and data.

8.1 Background efficiency from Monte Carlo simulation

The efficiencies of the different identification selections for backgrounds have been studied using MC simulation of all relevant $2 \rightarrow 2$ QCD processes filtered at particle-level to mimic a level-1 electromag-

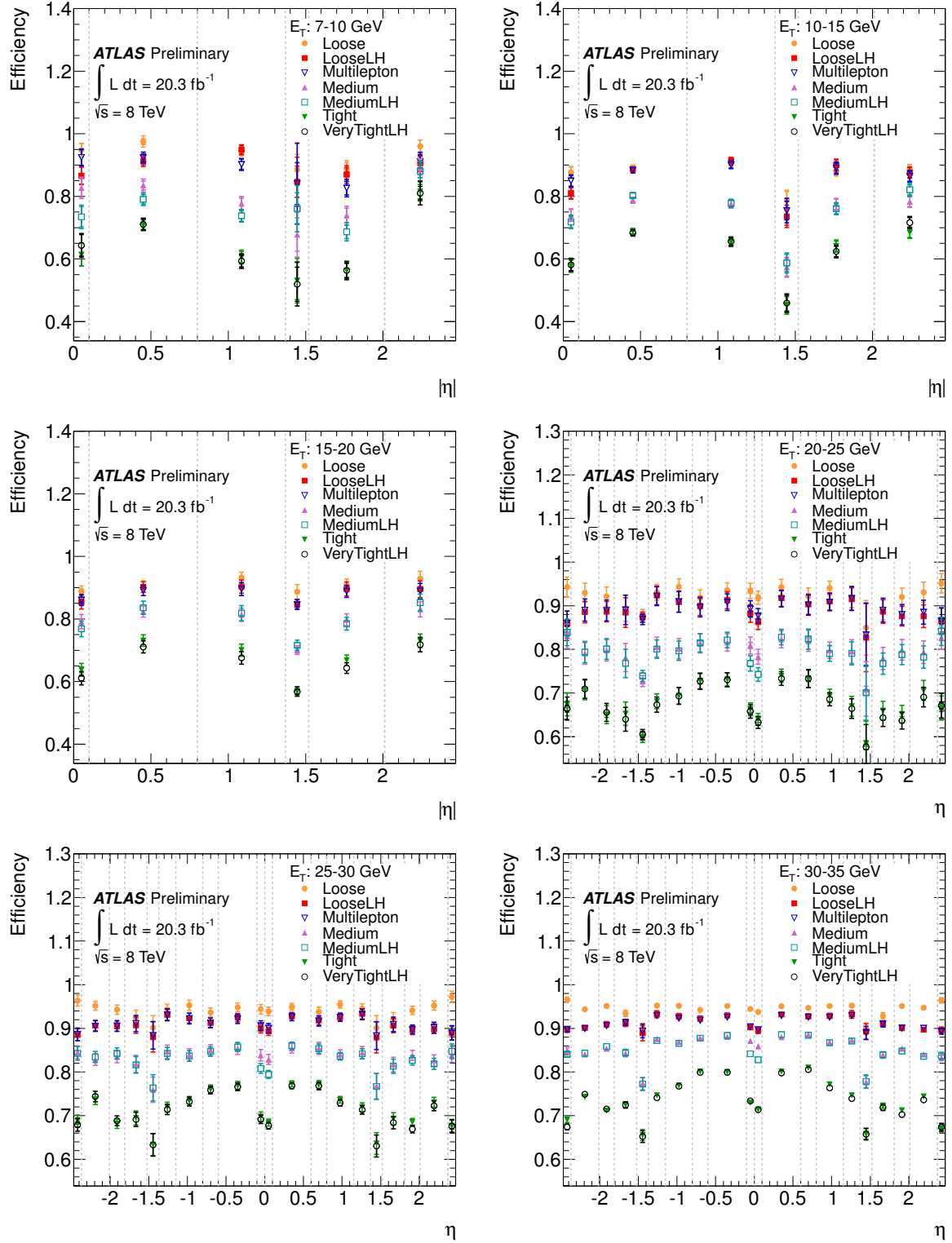


Figure 11: Measured identification efficiency for the various cut-based and LH selections as a function of η for different E_T bins ($E_T < 35$ GeV) for electrons. The data efficiency is derived from the measured data-to-MC efficiency ratios and the MC prediction for $Z \rightarrow ee$ decays. The uncertainties are statistical (inner error bars) and statistical+systematic (outer error bars). The dashed lines indicate the bins in which the efficiencies are calculated.

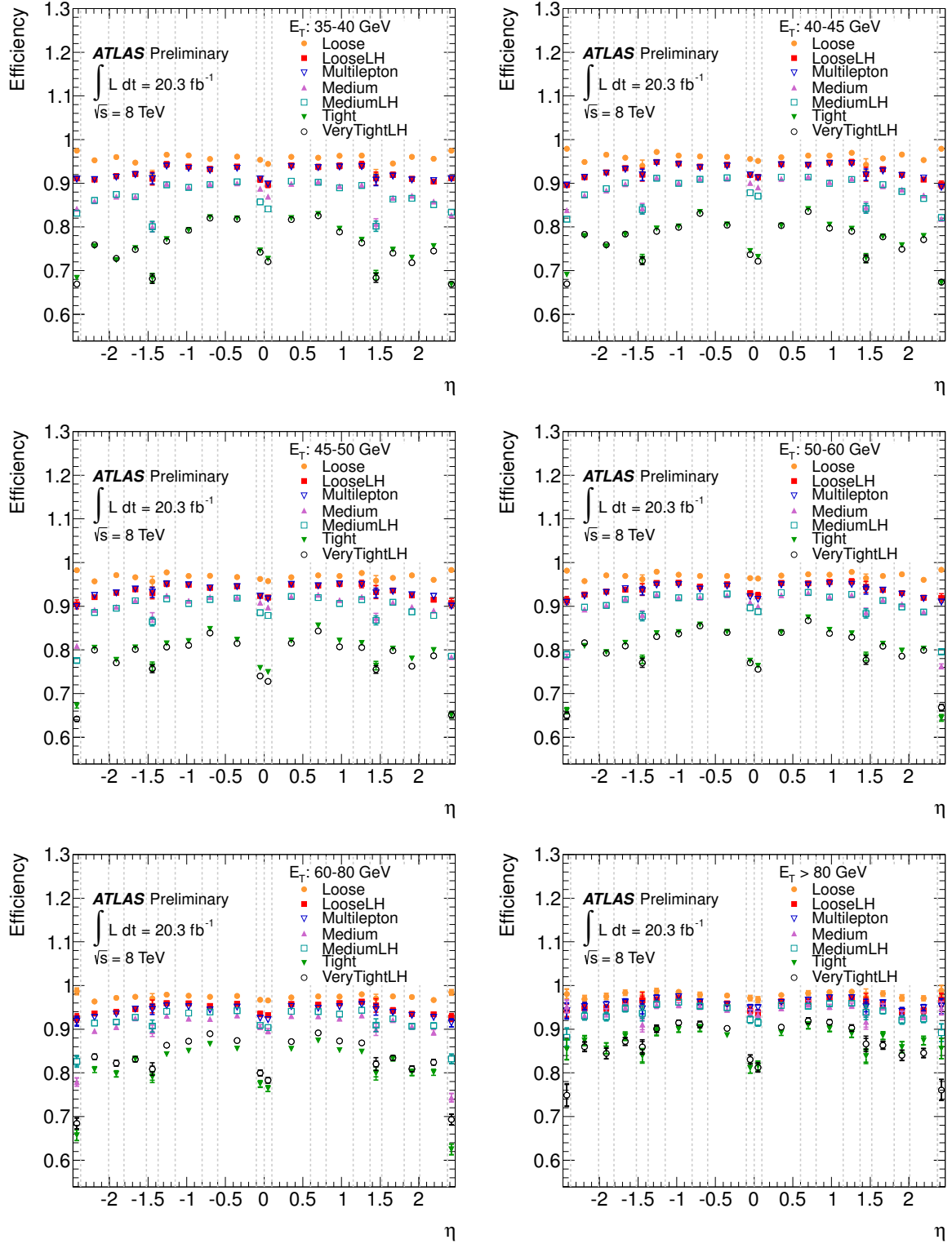


Figure 12: Measured identification efficiency for the various cut-based and LH selections as a function of η for different E_T bins ($E_T > 35$ GeV) for electrons. The data efficiency is derived from the measured data-to-MC efficiency ratios and the MC prediction from $Z \rightarrow ee$ decays. The uncertainties are statistical (inner error bars) and statistical+systematic (outer error bars). The dashed lines indicate the bins in which the efficiencies are calculated.

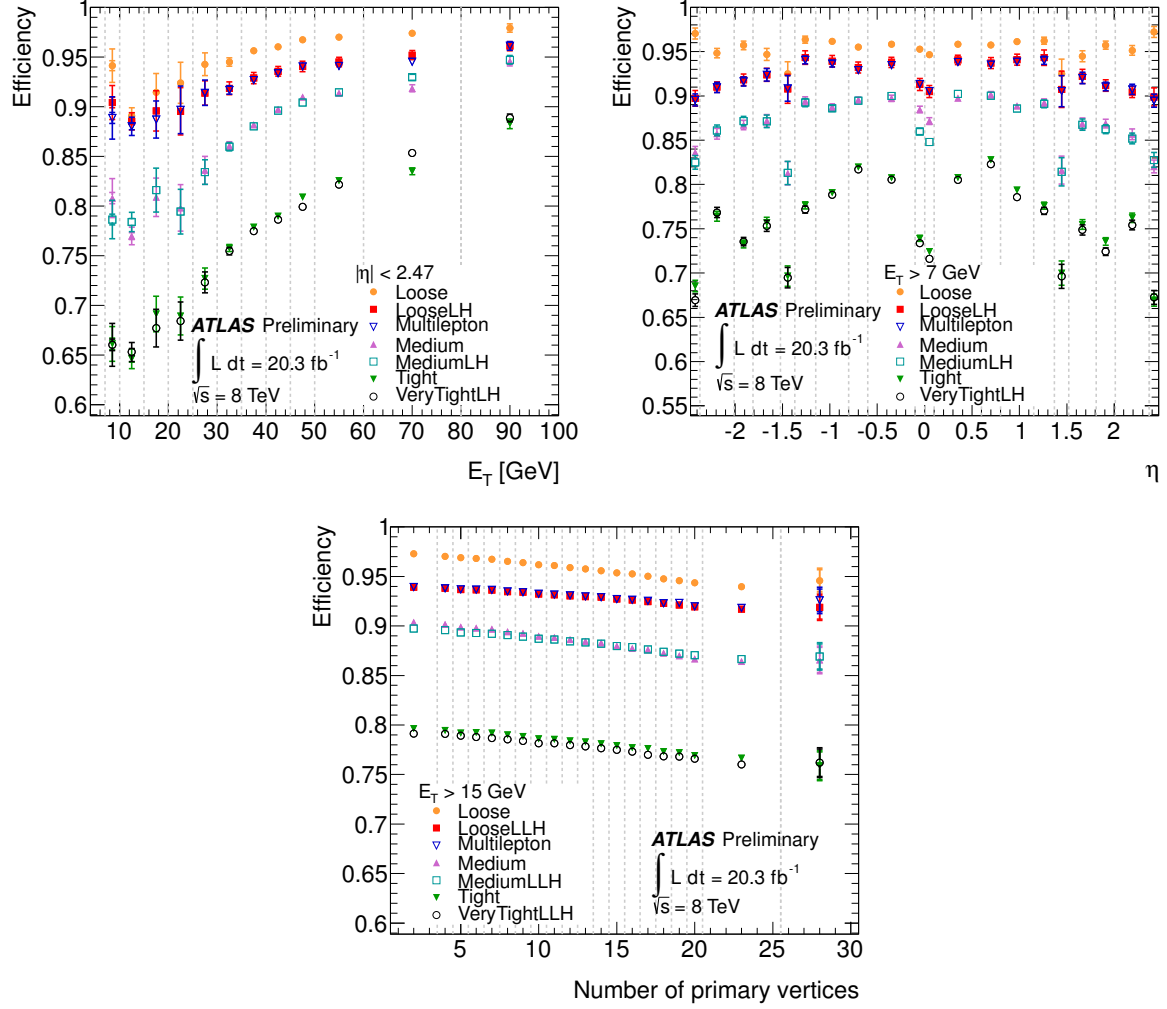


Figure 13: Measured identification efficiency for the various cut-based and LH selections as a function of E_T (top left), η (top right) and the number of reconstructed primary vertices (bottom). The data efficiency is derived from the measured data-to-MC efficiency ratios and the MC prediction from $Z \rightarrow ee$ decays. The uncertainties are statistical (inner error bars) and statistical+systematic (outer error bars). The last bin in E_T and number of primary vertices includes the overflow. The dashed lines indicate the bins in which the efficiencies are calculated.

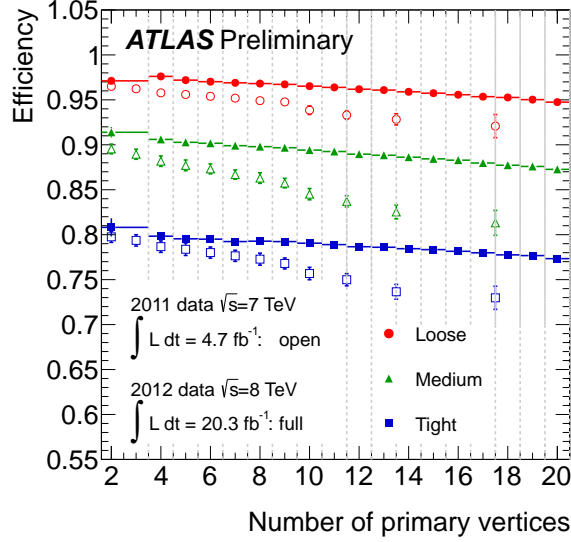


Figure 14: Identification efficiency in data for the various cut-based selections measured with 2011 and 2012 data as a function of the number of reconstructed primary vertices.

netic trigger requirement. The sample is enriched in electron backgrounds, with electrons from W and Z decays excluded at particle level using the MC truth record. Furthermore, the sample is required to pass a set of electron and photon triggers without identification criteria, to allow for better comparison with data-driven measurements. The estimated background efficiency and the composition of the background are shown in Table 3 for reconstructed electrons passing track quality requirements with transverse energies between 20 and 50 GeV. The quoted uncertainties are statistical only. The composition of this background-enriched reconstructed electron sample is categorized according to simulation information: non-isolated electrons from heavy-flavor decays, background electrons from conversions and Dalitz decays, and hadrons. No explicit isolation cut is applied.

The background after applying the looser cut-based selections generally consists of hadrons and background electrons in similar fractions, with a small contribution of electrons from heavy-flavor decays. As the cut-based selections get tighter, heavy-flavor decays begin to dominate the remaining background, followed by background electrons. In contrast, the LOOSE LH selection retains significantly less hadronic background than its cut-based counterpart; instead, non-isolated and background electrons dominate in this regime. After the VERY TIGHT LH selection, hadrons become highly suppressed and the sample is dominated by non-isolated electrons. To suppress these further, in many analyses of collision data isolation and tighter impact parameter cuts are added on top of the electron identification selection.

To estimate absolute background efficiencies, it is necessary to determine the efficiency of background objects to pass the denominator requirement of the relative efficiencies listed in Table 3. An unfiltered MC sample consisting of minimum bias, single- and double-diffractive physics is used. The numerator consists of reconstructed electrons passing the trigger and track quality requirements with transverse electron energy $E_T > 20$ GeV. The denominator is defined as the numerator plus any object reconstructed as a hadronic jet using the anti- k_t jet reconstruction algorithm [27], with a cone size of $\Delta R = 0.4$, and transverse jet energy $E_{T,\text{jet}} > 20$ GeV. While hadronic jets are the dominant background at this level, heavy flavor electrons and photons are mostly reconstructed as jets as well, so they are not considered separately. Jets overlapping with reconstructed electrons within a ΔR of 0.4 are removed to prevent double-counting. Reconstructed objects matched to simulated electrons from W and Z decays are also removed from the calculation. Using this methodology, we find that $8.89\% \pm 0.16\%$ (stat.) of

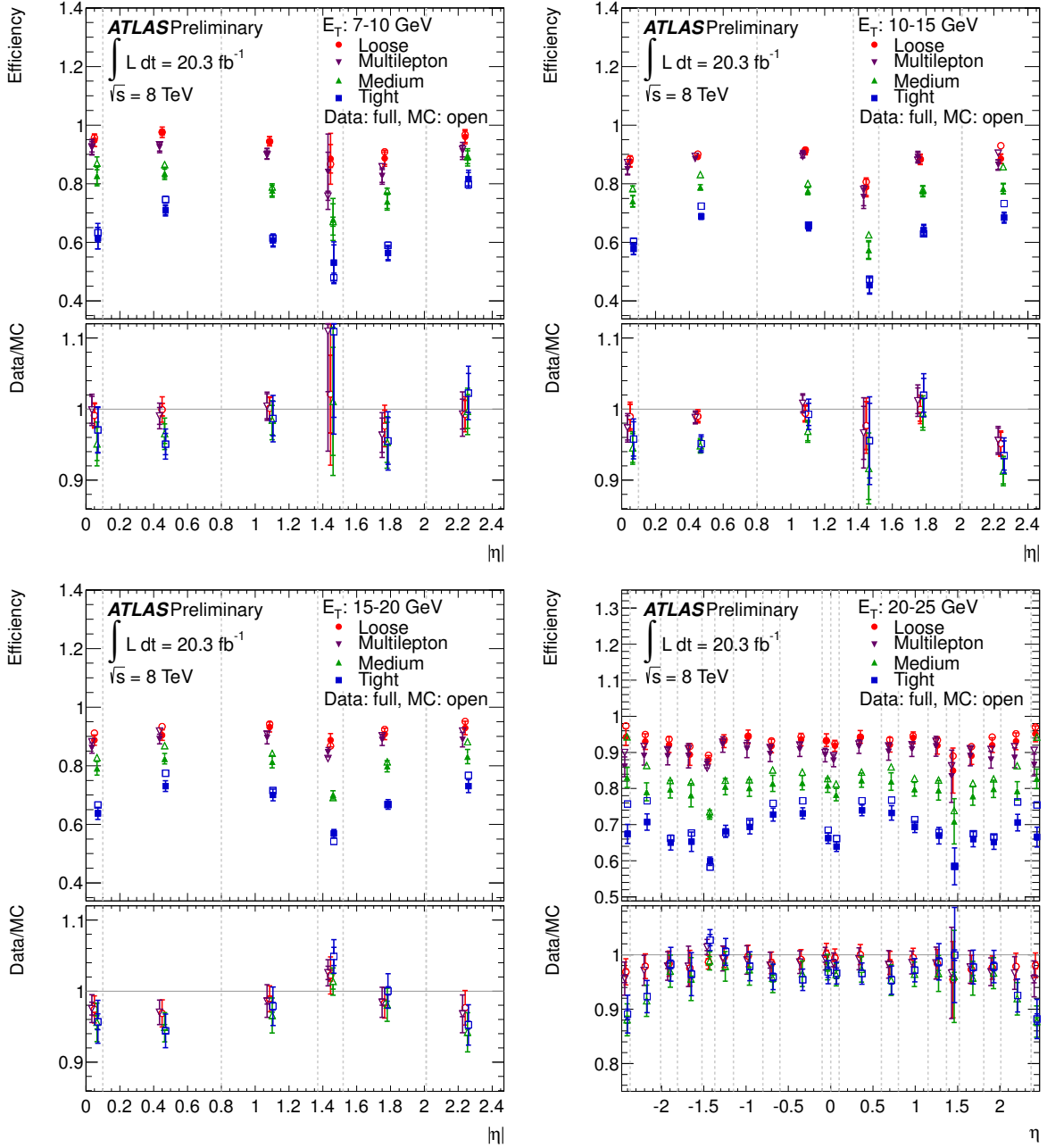


Figure 15: Identification efficiency in data as a function of η for different E_T bins ($E_T < 25$ GeV) for the cut-based *loose*, *multilepton*, *medium* and *tight* selections, compared to the MC expectation for electrons from $Z \rightarrow ee$ decay. The data efficiency is derived from the measured data-to-MC efficiency ratios and the MC prediction for electrons from $Z \rightarrow ee$ decays. The uncertainties are statistical (inner error bars) and statistical+systematic (outer error bars). The dashed lines indicate the bins in which the efficiencies are calculated.

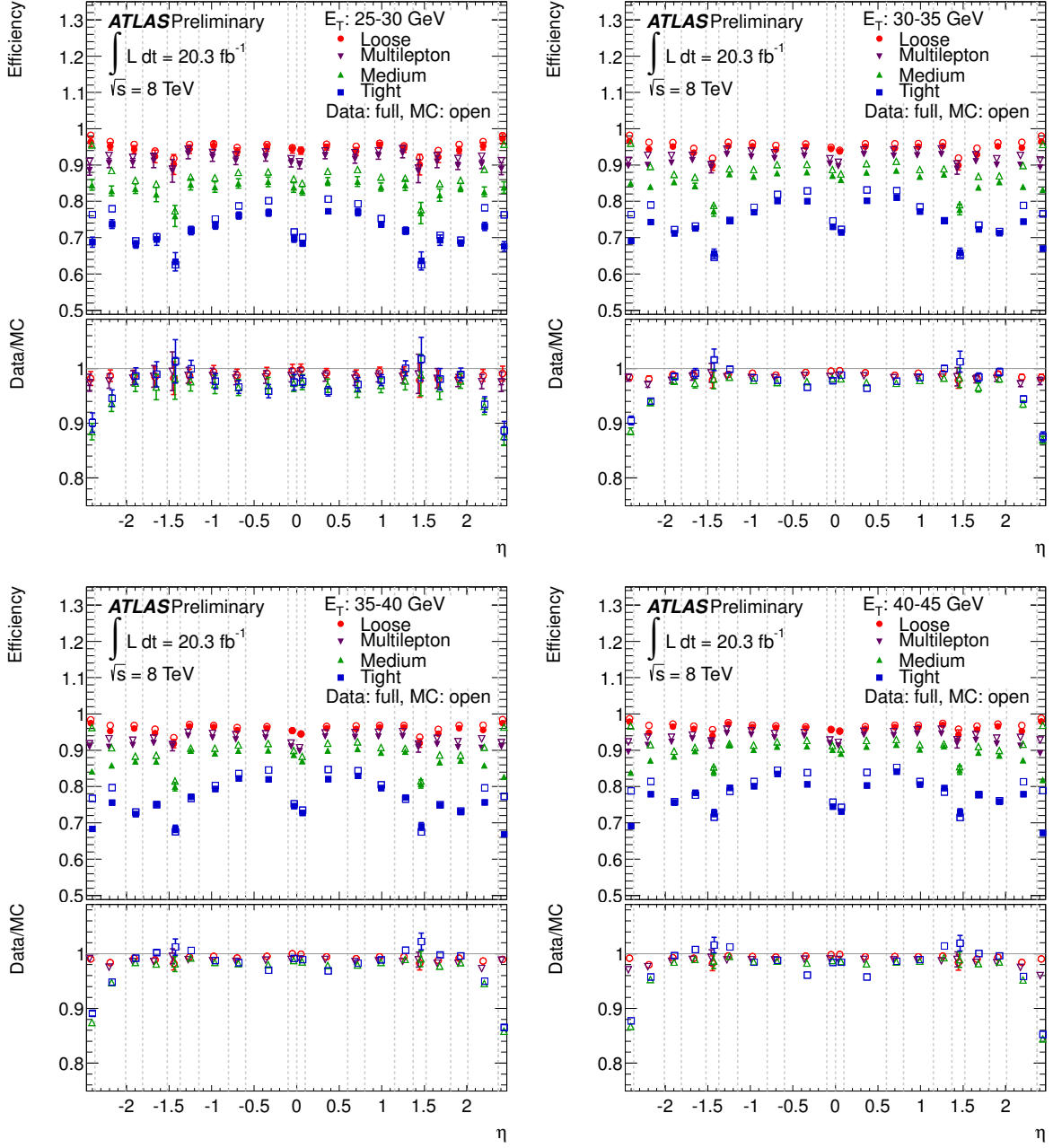


Figure 16: Identification efficiency in data as a function of η for different E_T bins ($25 \text{ GeV} < E_T < 45 \text{ GeV}$) for the cut-based *loose*, *multilepton*, *medium* and *tight* selections, compared to the MC expectation for electrons from $Z \rightarrow ee$ decay. The data efficiency is derived from the measured data-to-MC efficiency ratios and the MC prediction for electrons from $Z \rightarrow ee$ decays. The uncertainties are statistical (inner error bars) and statistical+systematic (outer error bars). The dashed lines indicate the bins in which the efficiencies are calculated.

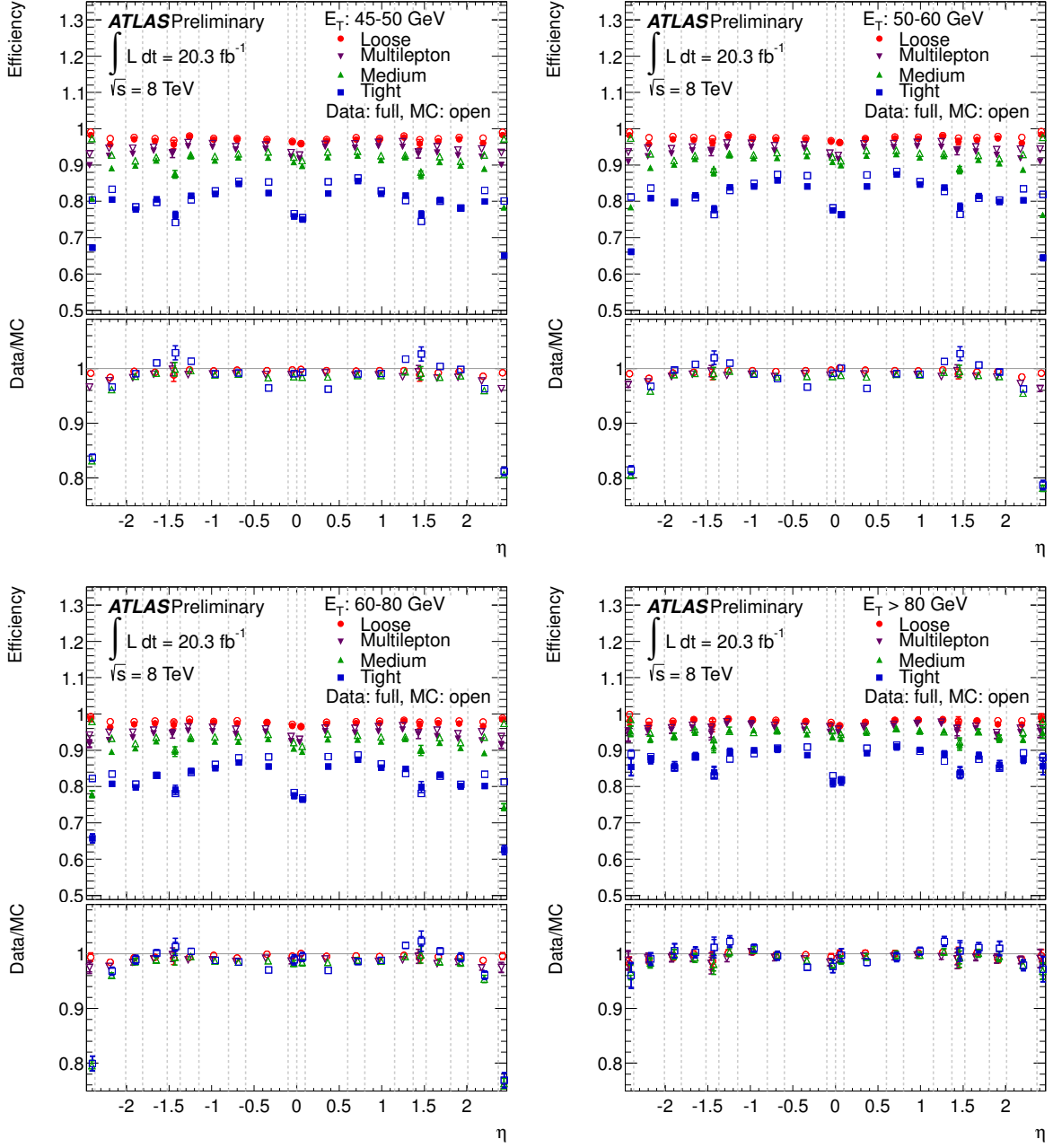


Figure 17: Identification efficiency in data as a function of η for different E_T bins ($E_T > 45$ GeV) for the cut-based *loose*, *multilepton*, *medium* and *tight* selections, compared to the MC expectation for electrons from $Z \rightarrow ee$ decay. The data efficiency is derived from the measured data-to-MC efficiency ratios and the MC prediction for electrons from $Z \rightarrow ee$ decays. The uncertainties are statistical (inner error bars) and statistical+systematic (outer error bars). The dashed lines indicate the bins in which the efficiencies are calculated.

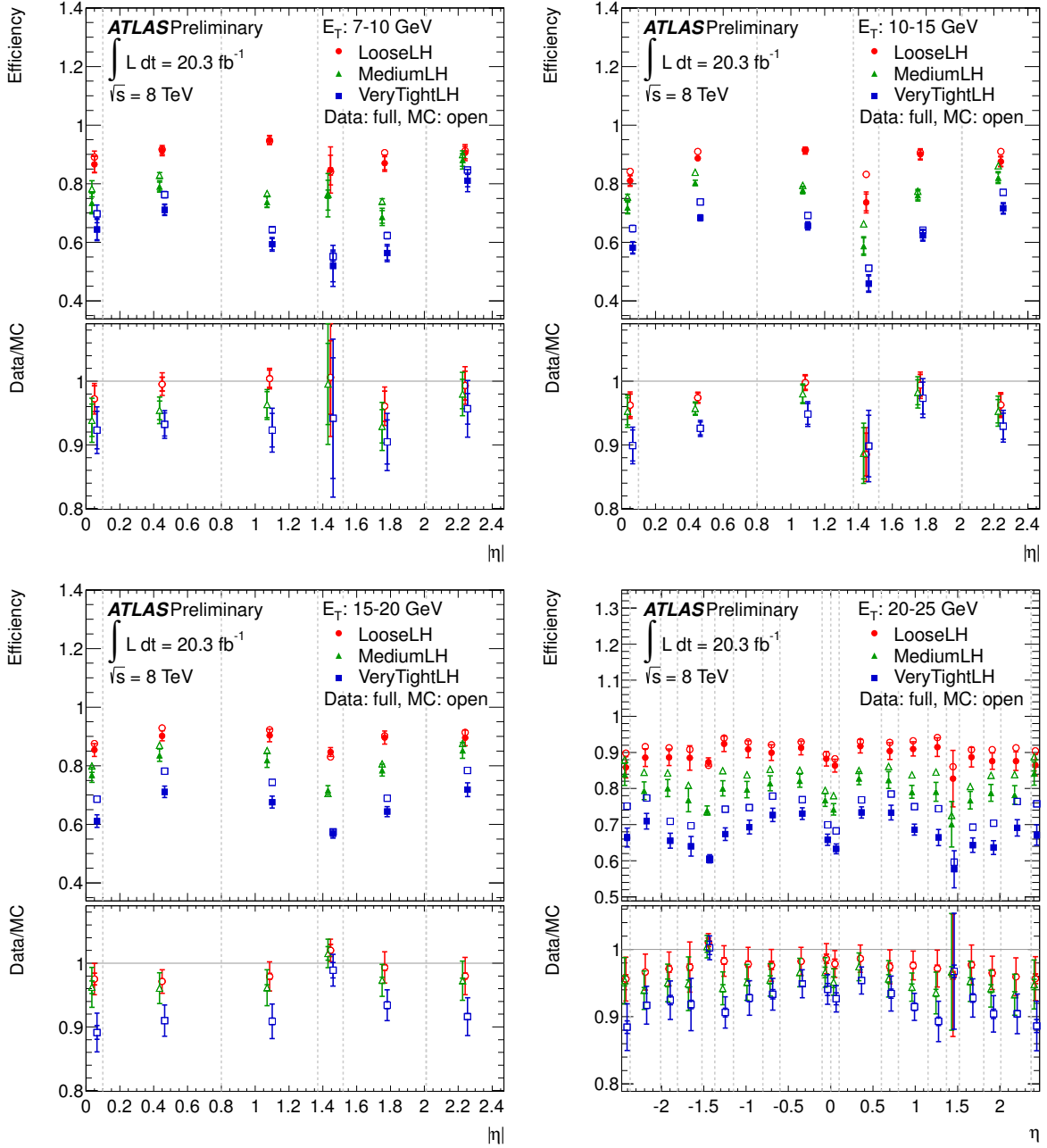


Figure 18: Identification efficiency in data as a function of η for different E_T bins ($E_T < 25$ GeV) for LOOSE LH, MEDIUM LH and VERY TIGHT LH, compared to MC expectation for electrons from $Z \rightarrow ee$ decay. The data efficiency is derived from the measured data-to-MC efficiency ratios and the MC prediction for electrons from $Z \rightarrow ee$ decays. The uncertainties are statistical (inner error bars) and statistical+systematic (outer error bars). The dashed lines indicate the bins in which the efficiencies are calculated.

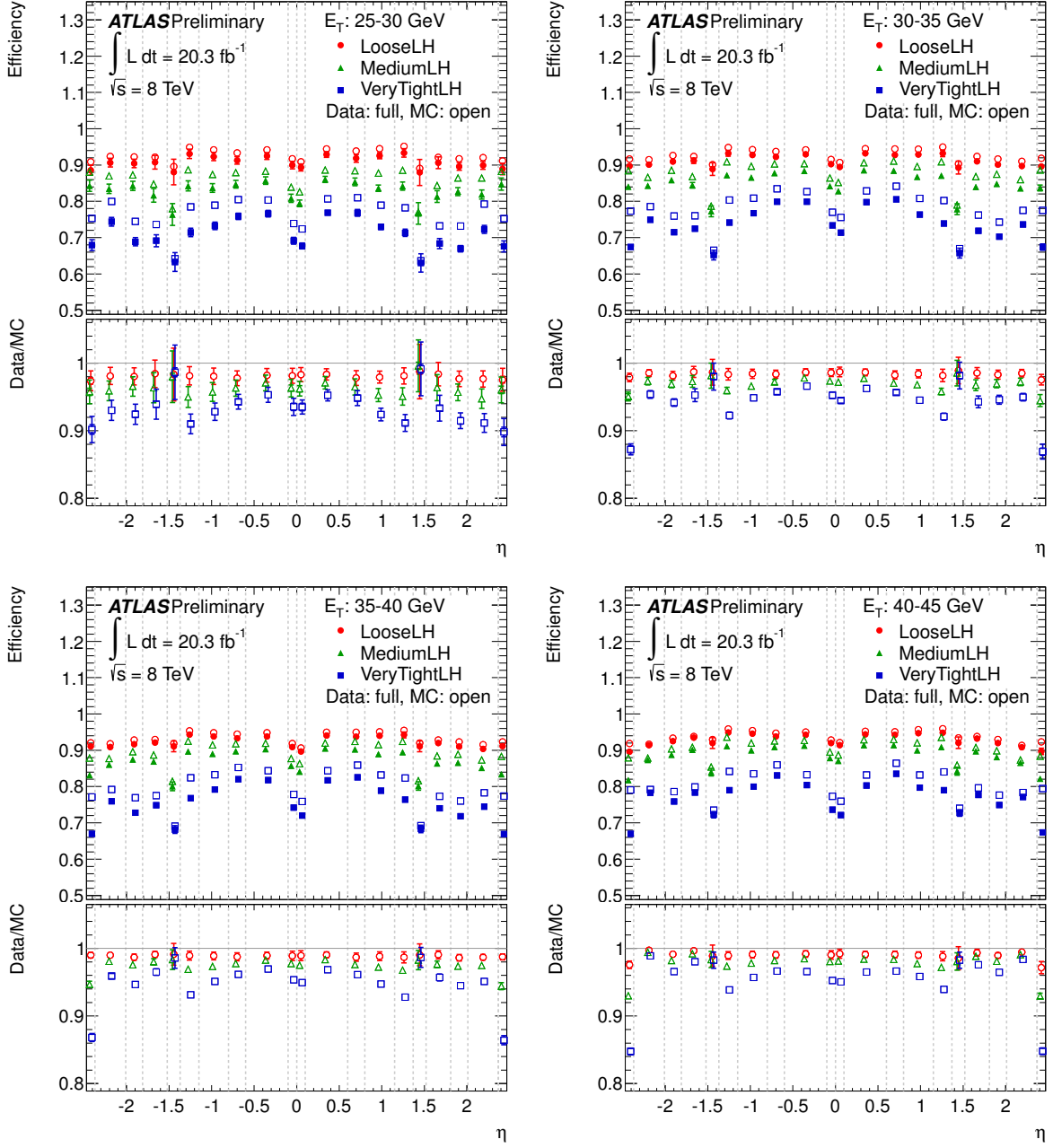


Figure 19: Identification efficiency in data as a function of η for different E_T bins ($25 \text{ GeV} < E_T < 45 \text{ GeV}$) for LOOSE LH, MEDIUM LH and VERY TIGHT LH, compared to MC expectation for electrons from $Z \rightarrow ee$ decay. The data efficiency is derived from the measured data-to-MC efficiency ratios and the MC prediction for electrons from $Z \rightarrow ee$ decays. The uncertainties are statistical (inner error bars) and statistical+systematic (outer error bars). The dashed lines indicate the bins in which the efficiencies are calculated.

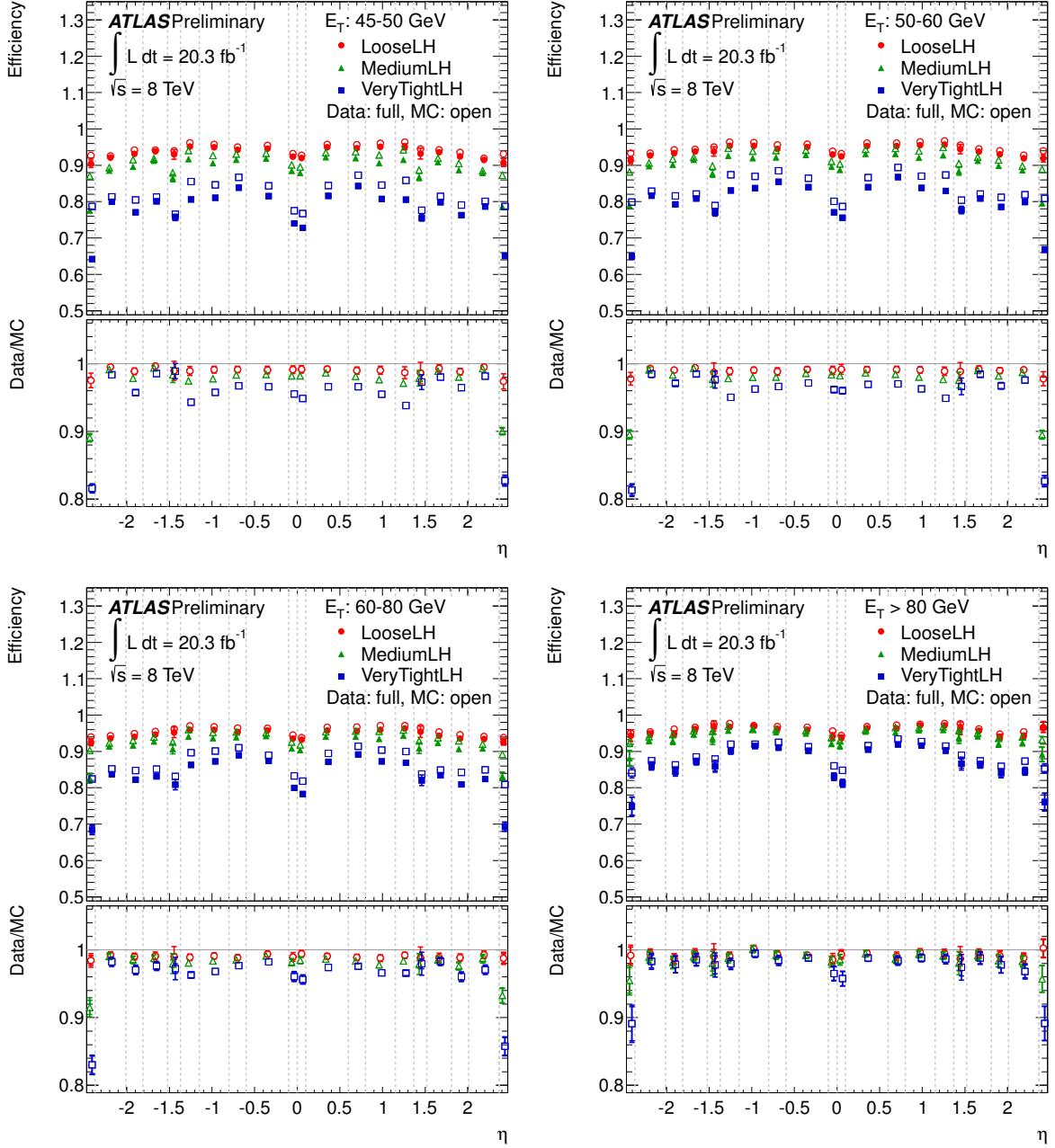
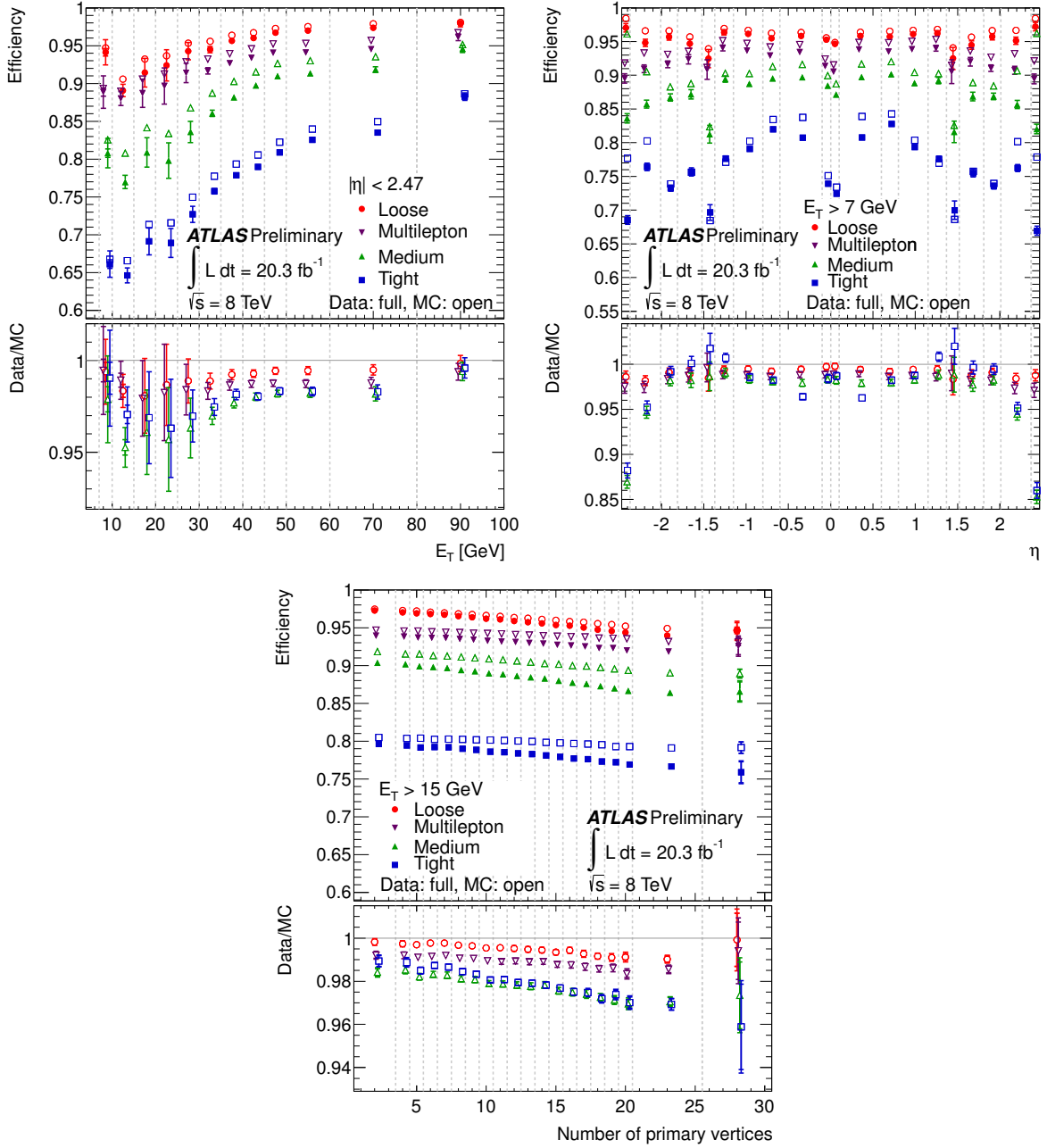


Figure 20: Identification efficiency in data as a function of η for different E_T bins ($E_T > 45$ GeV) for LOOSE LH, MEDIUM LH and VERY TIGHT LH, compared to MC expectation for electrons from $Z \rightarrow ee$ decay. The data efficiency is derived from the measured data-to-MC efficiency ratios and the MC prediction for electrons from $Z \rightarrow ee$ decays. The uncertainties are statistical (inner error bars) and statistical+systematic (outer error bars). The dashed lines indicate the bins in which the efficiencies are calculated.



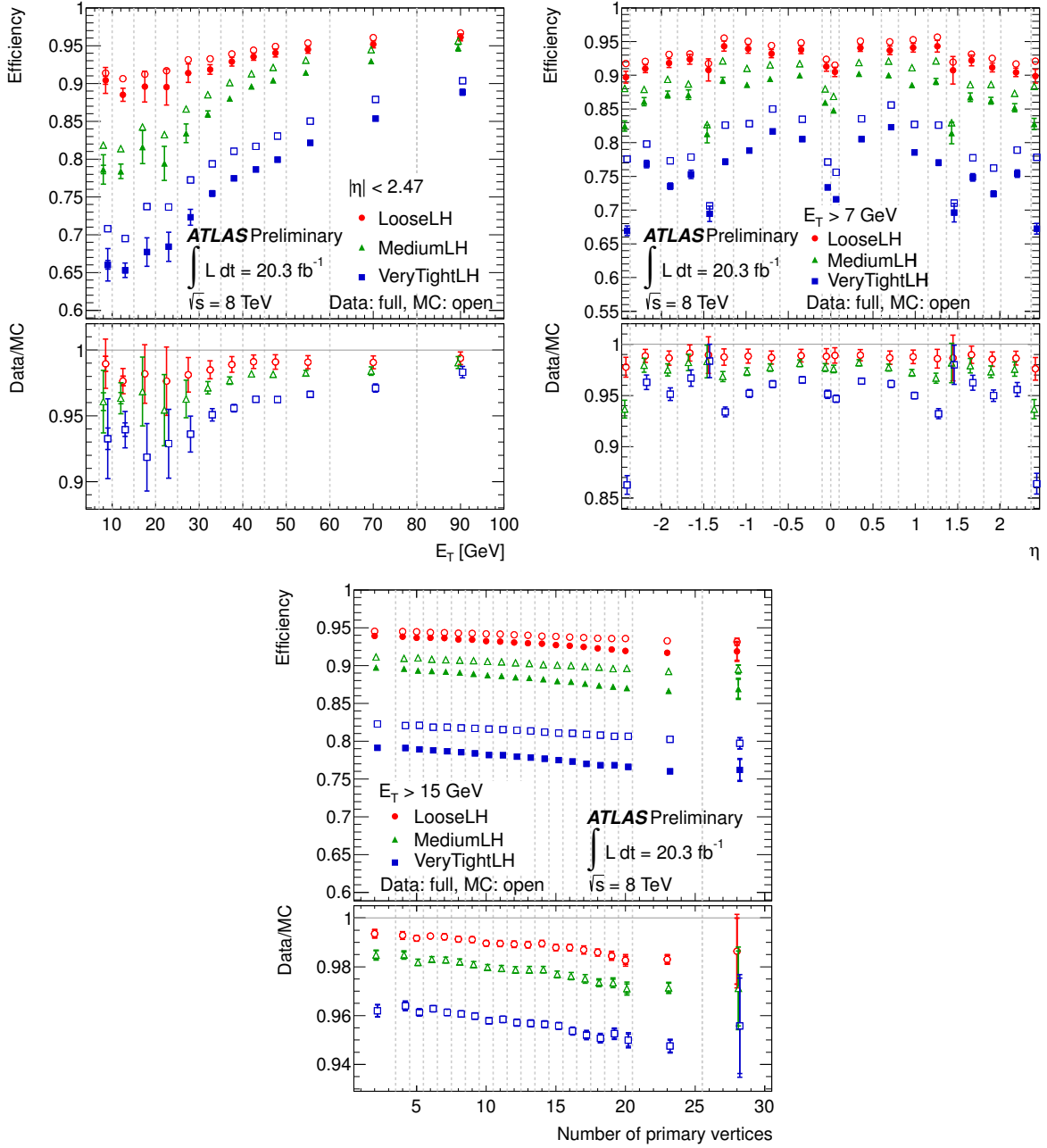


Figure 22: Identification efficiency in data as a function of E_T (top left), η (top right) and the number of reconstructed primary vertices (bottom) for LOOSE LH, MEDIUM LH and VERY TIGHT LH selections, compared to MC expectation for electrons from $Z \rightarrow ee$ decay. The lower panel shows the data-to-MC efficiency ratios. The data efficiency is derived from the measured data-to-MC efficiency ratios and the MC prediction for electrons from $Z \rightarrow ee$ decays. The uncertainties are statistical (inner error bars) and statistical+systematic (outer error bars). The dashed lines indicate the bins in which the efficiencies are calculated.

20 < E _T < 50 GeV								
Selection	Data efficiency Z → ee signal	MC efficiency Background (prompt e excluded)	Background composition (%)			MC efficiency (%) for background categories		
			non-iso e	bkg e	hadron	non-iso e	bkg e	hadron
Track Quality	100.0	100.00	1.1	16.1	82.8	100.0	100.00	100.000
<i>loose</i> cuts	95.7 ± 0.2	4.76 ± 0.04	7.4	48.4	44.2	32.5±0.8	14.3±0.2	2.54±0.03
<i>multilepton</i> cuts	92.9 ± 0.2	1.64 ± 0.02	22.5	34.5	43.0	34.2±0.8	3.51±0.08	0.85±0.02
<i>medium</i> cuts	88.1 ± 0.2	1.11 ± 0.02	25.8	50.5	23.7	26.5±0.8	3.46±0.08	0.32±0.01
<i>tight</i> cuts	77.5 ± 0.2	0.46 ± 0.01	54.5	29.9	15.6	23.0±0.7	0.85±0.04	0.086±0.006
LOOSE LH	92.8 ± 0.2	0.94 ± 0.02	40.2	42.0	17.9	34.8±0.8	2.44±0.07	0.20±0.01
MEDIUM LH	87.8 ± 0.3	0.51 ± 0.01	48.8	40.6	10.7	23.1±0.7	1.29±0.05	0.066±0.005
VERY TIGHT LH	77.0 ± 0.3	0.29 ± 0.01	63.7	28.9	7.4	16.9±0.7	0.51±0.03	0.026±0.003

Table 3: Background efficiency of different identification selections taken from a MC simulation containing all relevant $2 \rightarrow 2$ QCD processes. The reconstructed electrons are required to have transverse energies between 20 and 50 GeV and electrons from W and Z decays are removed at particle level. Furthermore, the sample is required to pass a set of electron and photon triggers without identification criteria, to allow for better comparison with data-driven measurements. The composition of the sample is categorized according to MC simulation information: non-isolated electrons from heavy-flavor decays, background electrons from photon conversions and Dalitz decays, and hadrons. The background efficiency for each category is also quoted. The efficiency is always quoted with respect to reconstructed electrons passing the track quality requirement. For completeness, the isolated electron efficiency for $Z \rightarrow ee$ decays, measured from data, is also given. The uncertainties are statistical only.

the simulated jets are reconstructed as electrons with $E_T > 20$ GeV and pass trigger and track quality requirements. The efficiencies in Table 3 can be multiplied by this number to obtain absolute background efficiencies for jets with $E_T > 20$ GeV.

8.2 Background efficiency ratios measured from collision data

Studying the electron backgrounds in MC can give an approximate estimate of the background efficiency. However, the MC description has several limitations: mis-identification efficiencies depend on the tail regions of many discriminating variable distributions, which are typically more susceptible to mis-modeling than the core of the distribution. Furthermore, a small deviation in shape can lead to a large data-to-MC efficiency correction factor due the low fraction of candidates in the tails. A data-driven estimate of the background efficiency is therefore essential. In this section, the ratio of background efficiencies between cut-based and LH menus is determined using data.

An inclusive background sample is selected by a set of electron and photon triggers with different E_T thresholds and no identification requirement. To prevent contamination from isolated electrons from W and Z decays, the electron candidate (matched to the trigger object) is rejected if it forms a pair with an invariant mass of 40-140 GeV with an electron candidate passing the *medium* requirement. Likewise, the electron candidate is also rejected if there is significant missing transverse energy in the event $E_T^{\text{miss}} > 25$ GeV, or if the transverse mass calculated with the missing momentum is compatible with W production, $M_T > 40$ GeV. These kinematic cuts are furthermore applied on simulated $Z \rightarrow ee$ and $W \rightarrow e\nu$ samples; the surviving events are scaled to the 2012 integrated luminosity and subtracted from the data yields before the background efficiency calculation.

The background sample is dominated by light-flavor hadrons, followed by photon conversions and a small fraction of heavy-flavor decays. The ratio of the background efficiency for a LH and the closest-efficiency cut-based selection is shown in Fig. 23. It can be seen that the LH selections let through only about 40-60% of the background compared to the cut-based selections, while it is shown in Section 7.4

that they retain approximately the same signal electron efficiency. These results can not be directly compared to those derived from MC simulation and given in Table 3, as the composition of the samples might differ. Nonetheless, the data-driven and MC-based estimations show the same trend when comparing the background rejection of cut-based and LH selections.

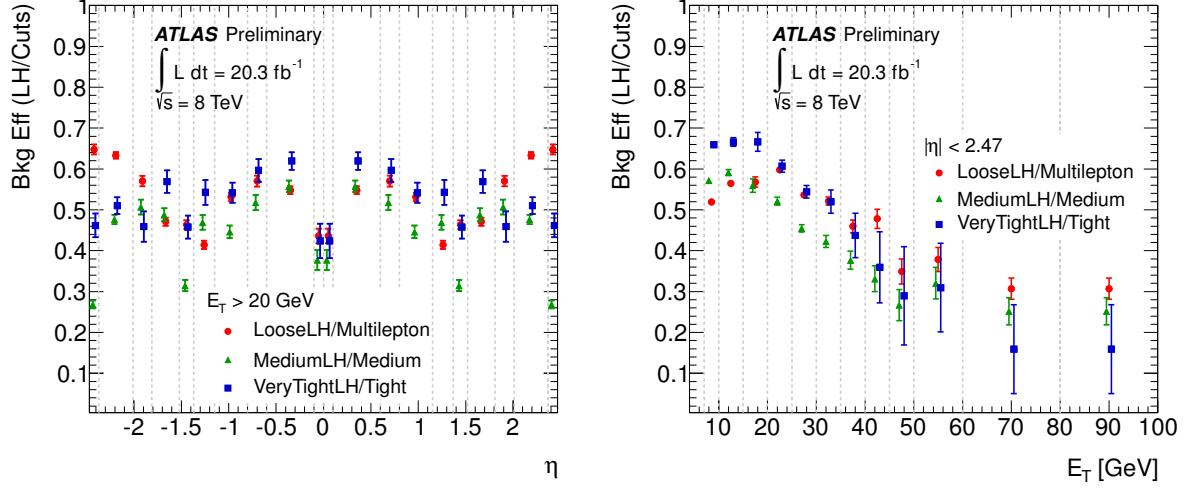


Figure 23: Ratio of background efficiencies for a LH and the closest-efficiency cut-based selections as a function of η (left) and E_T (right), as obtained using an inclusive background sample (see text). The uncertainties are statistical as well as systematic: a systematic uncertainty of 21% is assigned to the subtraction of signal events using the simulation; this uncertainty is dominated by the mis-modeling of the missing transverse energy.

9 Reconstruction efficiency measurement

9.1 Tag-and-probe with $Z \rightarrow ee$ events

9.1.1 Introduction

Electrons are reconstructed from EM clusters that are matched to tracks in the inner detector, as described in Section 3. The tracks are required to pass the track quality criteria, that is to have at least 1 hit in the pixel detector and in total at least 7 hits in the pixel or SCT detectors. The measurement of the efficiency to detect an energy cluster in the EM calorimeter using the sliding window algorithm is very challenging in data and not performed here. In MC simulation, it is found to be above 99% for $E_T > 15$ GeV as discussed in Section 3. EM clusters form the starting point of the reconstruction efficiency measurement.

The reconstruction efficiency is defined as the ratio of the number of electrons reconstructed as a cluster matched to a track passing the track quality criteria (numerator) to the number of clusters with or without a matching track (denominator). This reconstruction efficiency is measured using a tag-and-probe analysis which is very similar to the Z_{mass} method introduced in Section 7. In comparison to the measurement of the identification efficiency, the probe definition is relaxed to include all EM clusters. The background estimation is adapted to include the contribution of EM clusters with no associated track. The measurement is only performed for probe electrons with $E_T > 15$ GeV, as the background contamination of the sample becomes too high at lower E_T .

9.1.2 Event selection

The general event selection as well as the criteria on the tag electron are identical to the ones used in the Z_{mass} method, described in Section 7.1.2. Each event is required to have at least one tag electron and one probe object, which in this case is an EM cluster. In order to veto EM clusters of converted photons, no cluster within $\Delta R < 0.4$ of a reconstructed electron is considered. No requirement on the charge of the tag and the probe is applied, since there is no charge associated to EM clusters unless they are matched to a track.

9.1.3 Background estimation and systematic variations

The background estimation for the numerator of the reconstruction efficiency (electrons passing the reconstruction cuts) follows that of the Z_{mass} method described in Section 7.1.3. However, for the denominator (all reconstructed EM clusters) an additional contribution from photon candidates must be determined separately. The total background at the denominator level is the sum of two contributions: background to electrons reconstructed as a cluster with and without an associated track. The background estimation for these two contributions is explained below.

Background estimate for electrons reconstructed as clusters with no associated track Electrons reconstructed as EM clusters but not matched to any track are interpreted as photons. In order to estimate the photon background, which, unlike the signal electrons, has a smoothly falling invariant mass shape, a 3rd order polynomial is fitted to the invariant mass distribution of the selected electron-photon pairs (corresponding to the tag and the probe). The fit is carried out using the two side band regions above and below the Z mass peak, as illustrated in Fig. 24. Residual signal electron contamination in the background dominated side-band regions is subtracted using MC simulation before the fit. Systematic uncertainties on the MC scaling and MC description of the inefficiency to associate an electron with a track are 10 – 20 % and are not shown in Fig. 24.

Background estimate for electrons reconstructed as clusters with an associated track The method to estimate the background to EM clusters with an associated track is almost the same as for the identification efficiency measurement, described in Section 7.1.3: A background template is selected in data by inverting identification selection criteria on the probes and normalized to the data in a control region of the invariant mass distribution of the tag-and-probe pair. The backgrounds in the signal region are determined separately for clusters with tracks passing or failing the track quality selection criteria. Therefore, the track quality selection criteria are required to be passed (failed) in the background template selection for the invariant mass distribution of EM clusters passing (failing) the electron reconstruction procedure.

Figure 25 shows the invariant mass distributions of the EM cluster probes (composed of clusters with and without a track match at the probe level) for two selected bins both at the probe level and after the probe passed the reconstruction criteria. The estimation of the two background components is also depicted. As demonstrated by the figure, the measured data agrees well with the expectation, and the background subtraction procedure performs well.

The systematic uncertainty is estimated as for the identification efficiency. In addition to the variations listed in Section 7.1.3, the sidebands for the polynomial fit used for the estimation of the background to electrons without associated track are varied among these choices: $[70, 80 \text{ GeV}] \cup [100, 110 \text{ GeV}]$, $[60, 80 \text{ GeV}] \cup [100, 120 \text{ GeV}]$, $[50, 80 \text{ GeV}] \cup [100, 130 \text{ GeV}]$ and $[55, 70 \text{ GeV}] \cup [110, 125 \text{ GeV}]$.

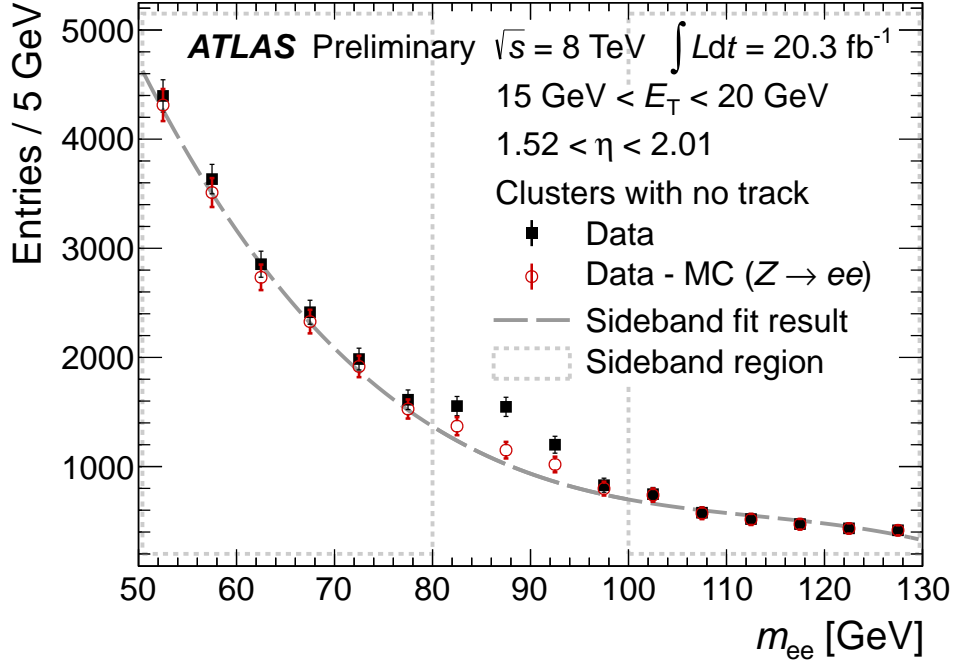


Figure 24: Estimation of the background to the selected EM clusters with no associated track for $15 \text{ GeV} < E_T < 20 \text{ GeV}$ and $1.52 < \eta < 2.01$. A polynomial fit (shown by a dashed dark gray line) is carried out in the side band region (indicated by dashed light grey boxes) on the invariant mass distribution of data events from which genuine electrons have been subtracted using MC simulation (the data are shown by filled squares before the MC subtraction and by open circles afterwards). In the signal region, here defined as the events with an invariant mass of 80 to 100 GeV, the fit result is used to obtain a data-driven estimate, which is compared to the data minus MC prediction. Only statistical uncertainties are shown for the data-MC prediction; the systematic uncertainty on the MC scaling and MC description of the inefficiency to associate an electron with a track is 10 – 20 %. This explains the small deviation in the signal region between the data-MC prediction and the polynomial fit to the sidebands. The MC prediction enters only in the subtraction of the very small residual signal in the sideband regions used to perform the polynomial fit. The resulting uncertainty on the measured reconstruction efficiency is negligible.

9.2 Results

The reconstruction efficiency, like the identification efficiency, is measured differentially in (E_T, η) bins. The efficiency to reconstruct an electron associated to a track of good quality varies from 95% to 99% between endcap and barrel region for low E_T electrons ($E_T < 20 \text{ GeV}$). For very high E_T electrons ($E_T > 80 \text{ GeV}$) the efficiency is $\sim 99\%$ over the whole η range. The results are shown in Figs. 26 and 27. The measured efficiency agrees well with the MC prediction. The data-to-MC correction factors are at most 1-2% different from unity and in most of the measurement they are within only a few permille of one. The total uncertainty is $< 0.5\%$ for electrons with E_T between 25 and 80 GeV. It is larger at lower E_T , varying between 0.5% and 2.0%. The statistical and systematic uncertainties are of the same order. The good data - MC agreement observed for $E_T > 15 \text{ GeV}$ gives confidence in the MC description of the detector response, which is relied on for electrons with $E_T < 15 \text{ GeV}$. In this low E_T region, the data-to-MC correction factor is assumed to be 1.0 with an uncertainty of 2% in the barrel and 5% in the endcap region.

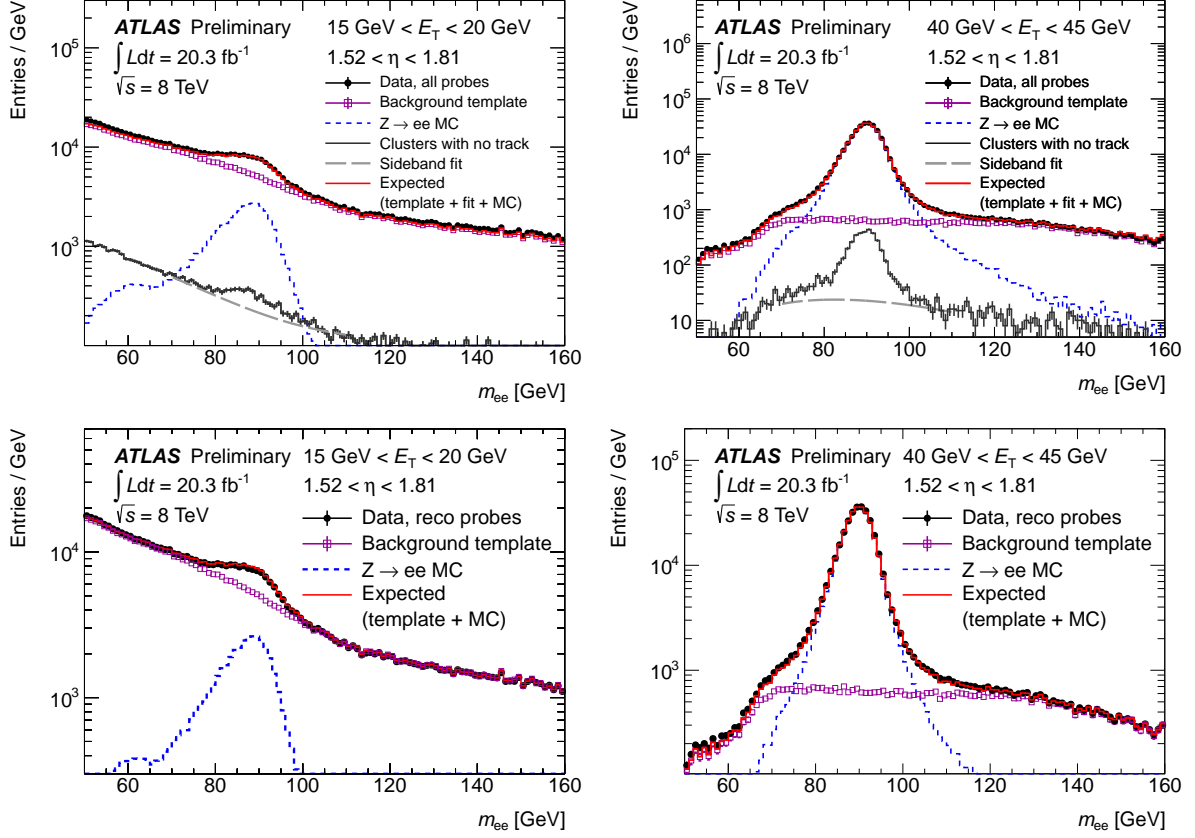


Figure 25: Invariant mass distributions of the tag-and-probe pairs for probes with $1.52 < \eta < 1.81$ and $15 \leq E_T \leq 20$ GeV (left) or $40 \leq E_T \leq 45$ GeV (right), before (top) and after (bottom) requiring the reconstruction criteria. The data (black dots with error bars) at the *all probes* level is composed of two components: clusters with no matching track (dark gray histogram with error bars) and clusters with a matching track. The background is evaluated separately for these two components. A 3rd order polynomial (gray dashed line spanning from 70 to 110 GeV) depicts the estimated photon background from a fit performed in the sideband regions as explained in Section 9.1.3 and shown in Figure 24. A background template normalized in this case to the high mass tail (magenta markers) is used to estimate the background with a matching track. This background template is obtained by requiring some of the identification criteria to be failed. Additionally the track quality selection criteria are required to be passed or failed depending on whether the background to the electrons passing or failing the reconstruction criteria is determined (cf. Section 9.1.3). For illustration only, the MC signal prediction (blue dashed line) is also displayed. The sum of the normalized background template and the MC signal prediction (red line, shown for comparison but not used in the measurement) agrees well with the data points.

The efficiencies are also measured as a function of the number of primary vertices in order to investigate the dependence of the electron reconstruction on pile-up. Figure 28 shows that the reconstruction efficiency for electrons with $E_T > 15$ GeV does not change with the number of primary vertices.

As described in Section 3, for the 2012 data, a new track reconstruction algorithm has been introduced in order to improve the reconstruction of electrons that have undergone significant bremsstrahlung. Figure 29 compares the reconstruction efficiencies measured in the 2011 and 2012 data. The new track fitting algorithm improves the overall electron reconstruction efficiency by $\sim 5\%$. Most of this improvement is in the low E_T range, where the electron reconstruction efficiency increases by more than $\sim 7\%$.

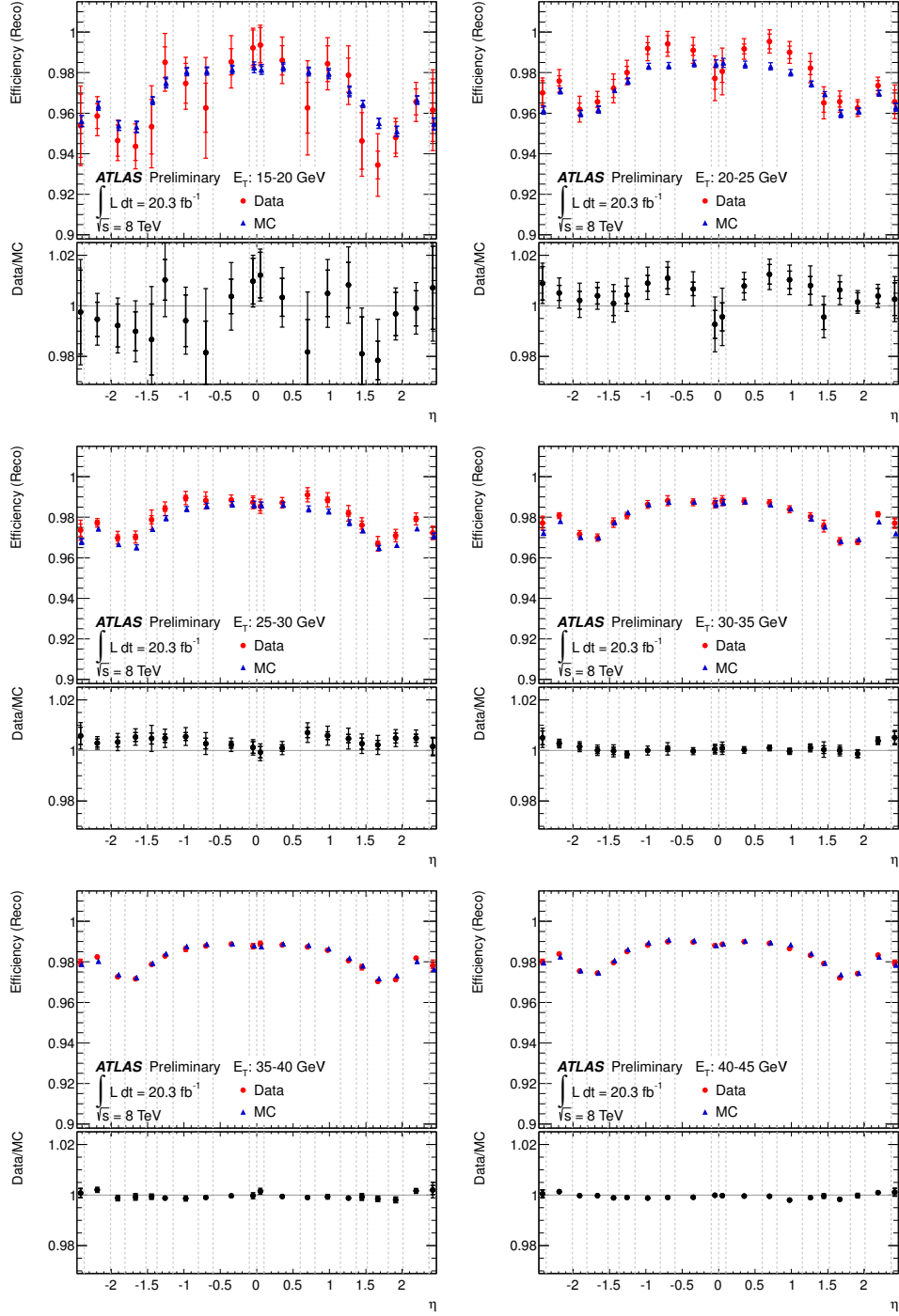


Figure 26: Measured reconstruction efficiency (circles) as a function of η for different E_T ranges for $E_T < 45$ GeV compared to the MC prediction (triangles). The lower panel shows the data-to-MC efficiency ratios. The uncertainties are statistical (inner error bars) and statistical+systematic (outer error bars).

This constitutes a significant gain for important measurements such as the measurement of the Higgs boson in the channel $H \rightarrow 4\ell$.

The gain in efficiency from the new track reconstruction algorithm flattens the distribution of the

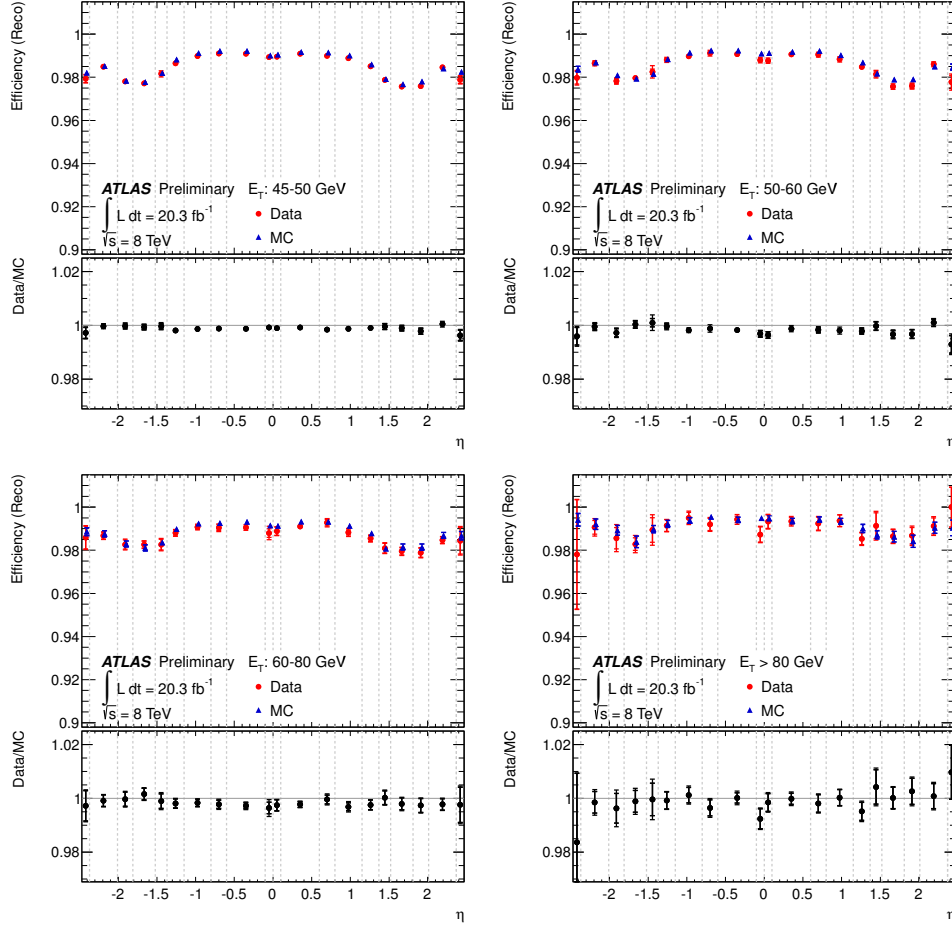


Figure 27: Measured reconstruction efficiency (circles) as a function of η for different E_T ranges for $E_T > 45$ GeV compared to the MC prediction (triangles). The lower panel shows the data-to-MC efficiency ratios. The uncertainties are statistical (inner error bars) and statistical+systematic (outer error bars).

reconstruction efficiency in η . For the 2011 data, a large drop in efficiency was observed for the end-cap regions, where more bremsstrahlung occurs due to a higher amount of material. For the 2012 data, this drop has become much smaller. Furthermore, the 2012 results are more precise than the final 2011 results, partly because of the increase in data statistics, but also due to improvements in the background subtraction method.

10 Combined reconstruction and identification efficiencies

Figure 30 shows the combined efficiencies to reconstruct and identify electrons with respect to reconstructed energy clusters in the electromagnetic calorimeter for all identification operating points. The efficiencies are shown as a function of E_T and η . As described in Section 7.4, the measured data-to-MC correction factors are applied to a simulated $Z \rightarrow ee$ sample. The resulting efficiencies correspond to the measured data efficiencies and can be compared to the efficiencies of simulated electrons in $Z \rightarrow ee$ events as done in Fig. 31 and 32. For electrons below $E_T < 15$ GeV, the reconstruction efficiency cannot be measured and is taken instead from the MC simulation.

The combined efficiency to reconstruct and identify an electron from $Z \rightarrow ee$ with E_T around 25 GeV

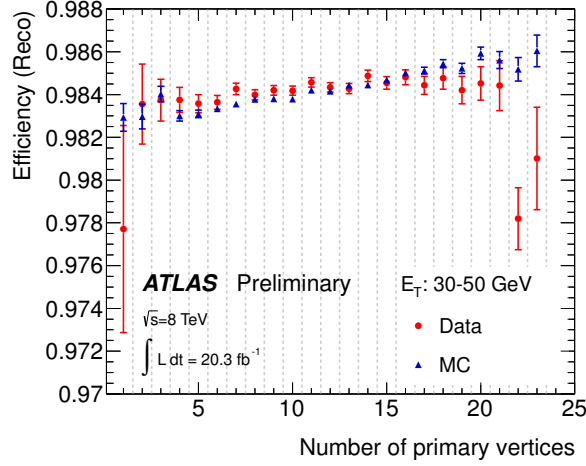


Figure 28: Measured reconstruction efficiency (red circles) as a function of the number of reconstructed primary vertices for $30 \text{ GeV} < E_T < 50 \text{ GeV}$ and integrated over η , compared to the MC prediction (blue triangles). The uncertainties are statistical+systematic. The dashed lines indicate the bins in which the efficiencies are calculated.

is about 92% for the *loose* cuts and around 68% for the *tight* cuts as well as the *VERY TIGHT* LH. It is lower (higher) at lower (higher) E_T , with a more pronounced turn-on as well as a greater η dependence for the tighter selections. Since the reconstruction efficiency is flat, the shapes are predominately determined by the variation of the identification efficiency (see Section 9 and 7).

11 Conclusion

Using the full 2012 data set, 20.3 fb^{-1} of 8 TeV pp collision data, the reconstruction and identification efficiencies of central electrons in the ATLAS detector are determined with the tag-and-probe method. Reconstruction efficiencies are measured for electrons from $Z \rightarrow ee$ decays. The identification efficiency measurements from J/ψ and Z decays are combined using data-to-MC efficiency ratios, increasing the precision of the results.

In 2012, a new track reconstruction algorithm and improved track-cluster matching were introduced, dedicated to recover efficiency losses due to electrons undergoing bremsstrahlung. As a result, the overall electron reconstruction efficiency is increased by roughly 5%. Averaged over η , it is about 97% for electrons with $E_T = 15 \text{ GeV}$ and reaches about 99% at $E_T = 50 \text{ GeV}$. For electrons with $E_T > 15 \text{ GeV}$, the efficiency varies from 99% at low η to 95% at high η . Below 15 GeV, the reconstruction efficiency is not measured due to the overwhelming background contamination of the sample. The uncertainty on the reconstruction efficiency is below 0.5% for $E_T > 25 \text{ GeV}$, and between around 0.5 – 1.5% at lower transverse energy.

The electron identification was improved by loosening the selection criteria on the shower shapes in the electromagnetic calorimeter that are most affected by the increased instantaneous luminosities provided by the LHC in 2012. To compensate for the loss in rejection power, new selection criteria were introduced and cuts on less pileup-sensitive variables were tightened. Additionally, new identification selections were developed: the cut-based *multilepton* selection, optimized for low energy electrons, as well as an identification based on the likelihood approach. Using the likelihood identification selections, the background rejection is significantly improved while maintaining the same signal efficiency as that of the cut-based selections. The identification efficiency depends strongly on E_T and, for the tighter

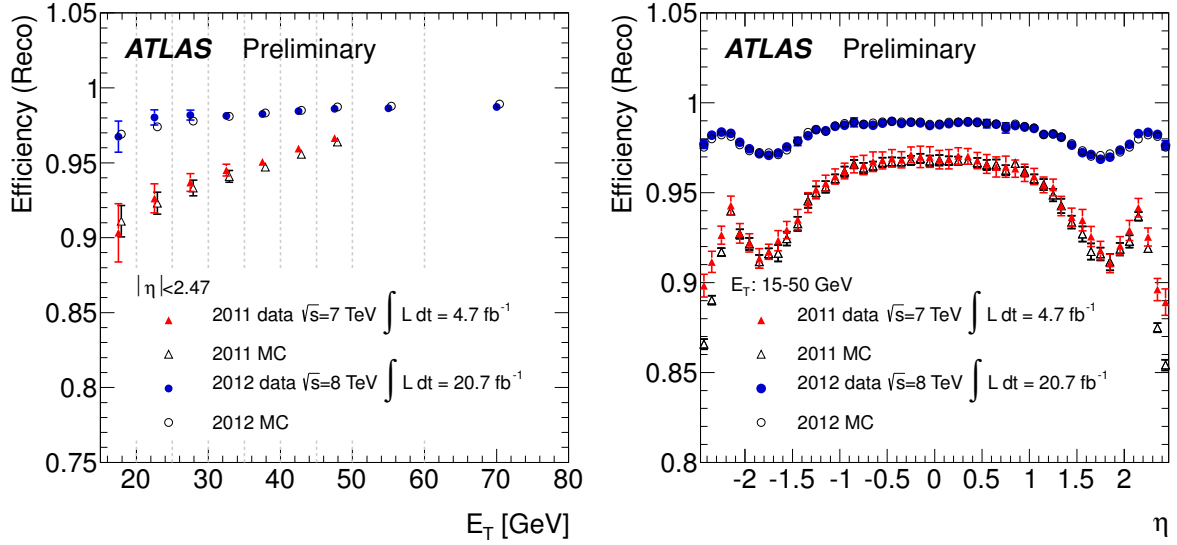


Figure 29: Measured reconstruction efficiencies as a function of E_T integrated over the full pseudorapidity range (left) and as a function of η for $15 \text{ GeV} < E_T < 50 \text{ GeV}$ (right) for the 2011 (triangles) and the 2012 (circles) datasets. For illustration purposes a finer η binning is used. The dashed lines in the left plot indicate the bins in which the efficiencies are calculated.

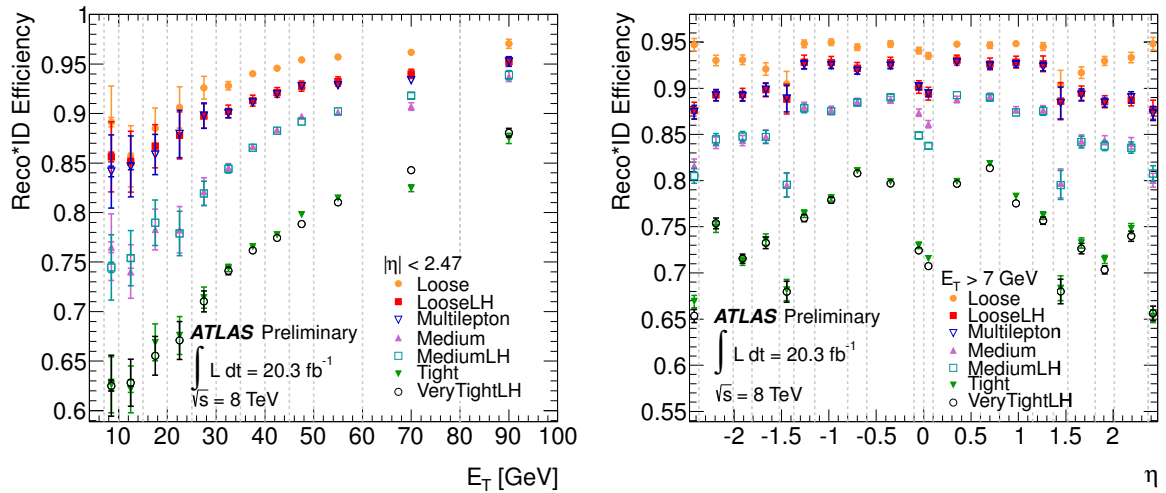


Figure 30: Measured combined reconstruction and identification efficiency for the various cut-based and likelihood selections as a function of E_T (left) and η (right) for electrons. The data efficiency is derived from the measured data-to-MC efficiency ratios and the MC prediction from $Z \rightarrow ee$ decays. The uncertainties are statistical (inner error bars) and statistical+systematic (outer error bars). The last E_T bin includes the overflow.

criteria, on η . Calculated with respect to reconstructed electrons passing quality criteria on their tracks, it averages between 96% (cut-based *loose*) and 78% (VERY TIGHT LH) for electrons with $E_T > 15 \text{ GeV}$. The measured pileup dependence is below 4% for 1–30 reconstructed primary collision vertices per bunch crossing for all sets of selection criteria. Some differences between the behavior in data and MC are observed, but understood. The total uncertainties on the identification efficiency measurements are

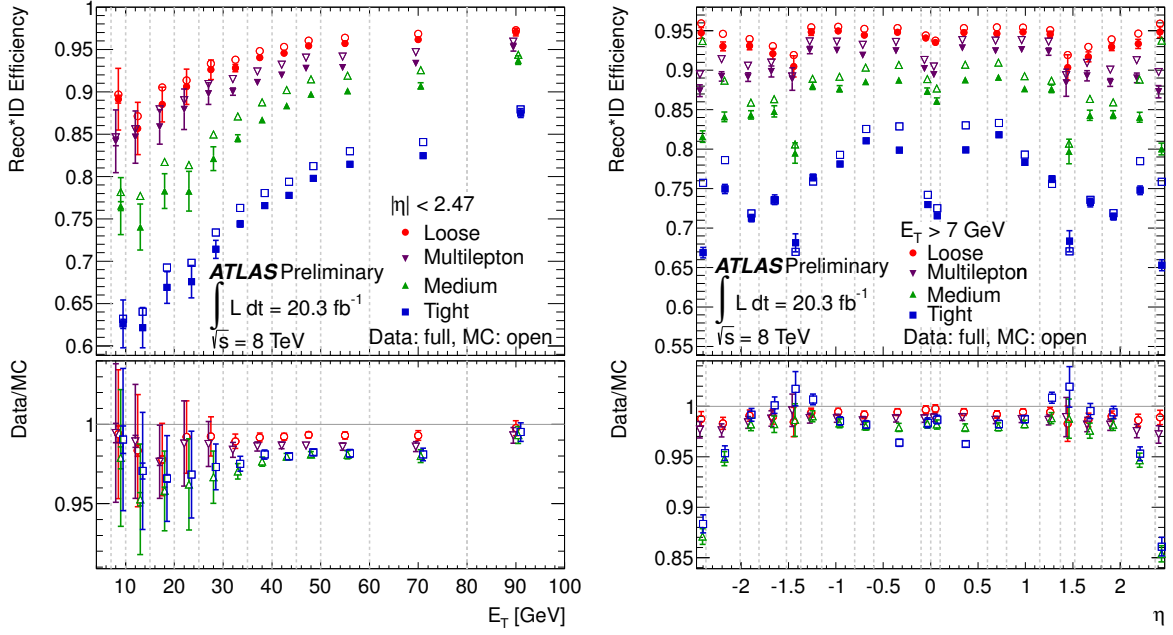


Figure 31: Measured combined reconstruction and identification efficiency as a function of E_T (left) and η (right) for the cut-based *loose*, *multilepton*, *medium* and *tight* selections, compared to MC expectation for electrons from $Z \rightarrow ee$ decay. The lower panel shows the data-to-MC efficiency ratios. The data efficiency is derived from the measured data-to-MC efficiency ratios and the MC prediction for electrons from $Z \rightarrow ee$ decays. The uncertainties are statistical (inner error bars) and statistical+systematic (outer error bars).

around 5-6% (1-2%) for electrons below (above) $E_T = 25$ GeV.

The measured data-to-MC efficiency ratios are applied as correction or scale factors in analyses, such as the measurement of the properties of the newly discovered Higgs boson. They are close to 1 with deviations larger than a couple of percent from unity occurring only for low E_T or high η regions.

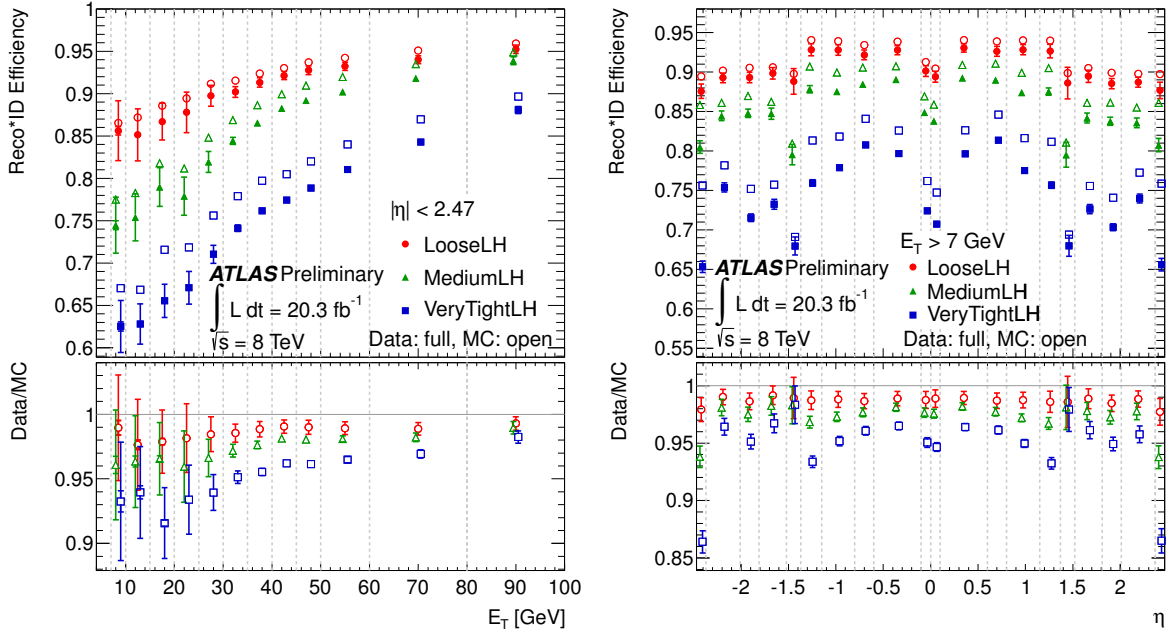


Figure 32: Measured combined reconstruction and identification efficiency as a function of E_T (left) and η (right) for the LOOSE LH, MEDIUM LH and VERY TIGHT LH selections, compared to MC expectation for electrons from $Z \rightarrow ee$ decay. The lower panel shows the data-to-MC efficiency ratios. The data efficiency is derived from the measured data-to-MC efficiency ratios and the MC prediction for electrons from $Z \rightarrow ee$ decays. The uncertainties are statistical (inner error bars) and statistical+systematic (outer error bars).

References

- [1] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, **JINST** **3** (2008) S08003.
- [2] ATLAS Collaboration, *Evolution and performance of the ATLAS electron-photon trigger in 2010-2012*, to be published.
- [3] ATLAS Collaboration, *Electron reconstruction and identification efficiency measurements with the ATLAS detector using the 2011 LHC proton-proton collision data*, **arXiv:1404.2240 [hep-ex]**, submitted to Eur. Phys. J. C.
- [4] W. Lampl et al., *Calorimeter Clustering Algorithms: Description and Performance*, **ATL-LARG-PUB-2008-002** (<https://cdsweb.cern.ch/record/1099735>), 2008.
- [5] T. Cornelissen et al., *Concepts, design and implementation of the ATLAS new tracking (NEWT)*, **ATL-SOFT-PUB-2007-007** (<https://cds.cern.ch/record/1020106>), 2007.
- [6] R. Frühwirth, *Application of Kalman filtering to track and vertex fitting*, **Nucl. Instrum. Meth.** **A262** (1987) 444–450.
- [7] T. Cornelissen et al., *The global χ^2 track fitter in ATLAS*, **J. Phys. Conf. Ser.** **119** (2008) 032013.
- [8] ATLAS Collaboration, *Improved electron reconstruction in ATLAS using the Gaussian Sum Filter-based model for bremsstrahlung*, **ATLAS-CONF-2012-047** (<https://cdsweb.cern.ch/record/1449796>), 2012.
- [9] ATLAS Collaboration, *Electron and photon energy calibration with the ATLAS detector using LHC Run 1 data*, to be submitted to Eur. Phys. J. C.
- [10] A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne and H. Voss, *TMVA: Toolkit for Multivariate Data Analysis*, PoS ACAT **040** (2007) , **arXiv:physics/0703039 [physics]**. <http://tmva.sourceforge.net/>.
- [11] ATLAS Collaboration, *Expected electron performance in the ATLAS experiment*, **ATL-PHYS-PUB-2011-006** (<http://cdsweb.cern.ch/record/1345327>), 2011.
- [12] ATLAS Collaboration, *Electron performance measurements with the ATLAS detector using the 2010 LHC proton-proton collision data*, Eur. Phys. J. **C72** (2012) 1909, **arXiv:1110.3174 [hep-ex]**.
- [13] C. Blocker, *Uncertainties on Efficiencies*, **CDF/MEMO/STATISTICS/PUBLIC/7168** (http://www-cdf.fnal.gov/physics/statistics/notes/cdf7168_eff_uncertainties.ps), 2004.
- [14] P. Nason, *A New method for combining NLO QCD with shower Monte Carlo algorithms*, **JHEP** **0411** (2004) 040, **arXiv:hep-ph/0409146 [hep-ph]**.
- [15] S. Frixione, P. Nason, and C. Oleari, *Matching NLO QCD computations with Parton Shower simulations: the POWHEG method*, **JHEP** **0711** (2007) 070, **arXiv:0709.2092 [hep-ph]**.
- [16] S. Alioli, P. Nason, C. Oleari, and E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, **JHEP** **1006** (2010) 043, **arXiv:1002.2581 [hep-ph]**.

- [17] T. Sjostrand, S. Mrenna, and P. Z. Skands, *A brief introduction to PYTHIA 8.1*, **Comput. Phys. Commun.** **178** (2008) 852–867, [arXiv:0710.3820 \[hep-ph\]](#).
- [18] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, **Eur. Phys. J.** **C70** (2010) 823–874, [arXiv:1005.4568 \[physics.ins-det\]](#).
- [19] GEANT4 Collaboration, S. Agostinelli et al., *GEANT4: A Simulation toolkit*, **Nucl. Instrum. Meth.** **A506** (2003) 250–303.
- [20] ATLAS Collaboration, *Measurements of the photon identification efficiency with the ATLAS detector using 4.9 fb^{-1} of pp collision data collected in 2011*, **ATLAS-CONF-2012-123** (<https://cds.cern.ch/record/1473426>), 2012.
- [21] CDF Collaboration, D. Acosta et al., *Measurement of the J/ψ meson and b -hadron production cross sections in $p\bar{p}$ collisions at $\sqrt{s} = 1960\text{ GeV}$* , **Phys. Rev.** **D71** (2005) 032001, [arXiv:hep-ex/0412071 \[hep-ex\]](#).
- [22] Particle Data Group, *Review of particle physics*, **J. Phys. G** **37** (2010) 075021.
- [23] ATLAS Collaboration, *Measurement of the differential cross-sections of inclusive, prompt and non-prompt J/ψ production in proton-proton collisions at $\sqrt{s} = 7\text{ TeV}$* , **Nucl. Phys.** **B850** (2011) 387–444, [arXiv:1104.3038 \[hep-ex\]](#).
- [24] M. Oreglia, *A Study of the Reactions $\psi' \rightarrow \gamma\gamma\psi$* , **SLAC-0236, UMI-81-08973** (1980) .
- [25] J. Gaiser, *Charmonium Spectroscopy From Radiative Decays of the J/ψ and ψ'* , **SLAC-0255, UMI-83-14449-MC** (1982) .
- [26] H1 Collaboration, *Measurement of the Inclusive ep Scattering Cross Section at Low Q^2 and x at HERA*, **Eur. Phys. J.** **C63** (2009) 625–678, [arXiv:0904.0929 \[hep-ex\]](#).
- [27] M. Cacciari, G. P. Salam, and G. Soyez, *The Anti- k_t jet clustering algorithm*, **JHEP** **0804** (2008) 063, [arXiv:0802.1189 \[hep-ph\]](#).

Appendix: Auxiliary material

Figures 4, 5 and 6 in Section 7 demonstrate the background subtraction techniques employed by the Z_{mass} , the $Z \rightarrow ee\gamma$ and the Z_{iso} methods respectively. For probes passing the *tight* identification, the backgrounds are highly suppressed and not visible well in linear scale. Therefore, the same figures are collected here with a logarithmic scale.

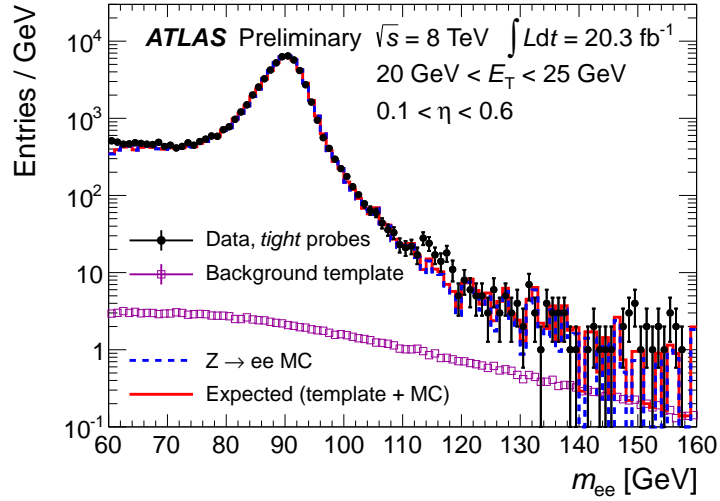


Figure 33: Illustration of the background estimation using the Z_{mass} method in the $20 \text{ GeV} < E_T < 25 \text{ GeV}$, $0.1 < \eta < 0.6$ bin for probes passing the cut-based *tight* identification. The background template is normalized in the range $120 < m_{ee} < 250 \text{ GeV}$. The tag passes cut-based *medium* and isolation cuts. The MC is scaled to match the total estimated signal in the Z-mass window.

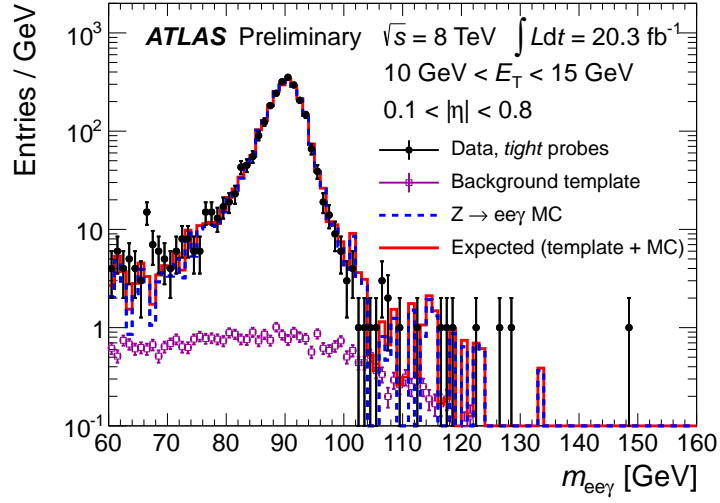


Figure 34: Illustration of the background estimation using the $Z \rightarrow ee\gamma$ method in the $10 \text{ GeV} < E_T < 15 \text{ GeV}$, $0.1 < |\eta| < 0.8$ bin for probes passing the cut-based *tight* identification. The background template is normalized in the range $100 \text{ GeV} < m_{ee} < 250 \text{ GeV}$. The tag passes cut-based *medium* and isolation cuts. The MC is scaled to match the total estimated signal in the Z -mass window.

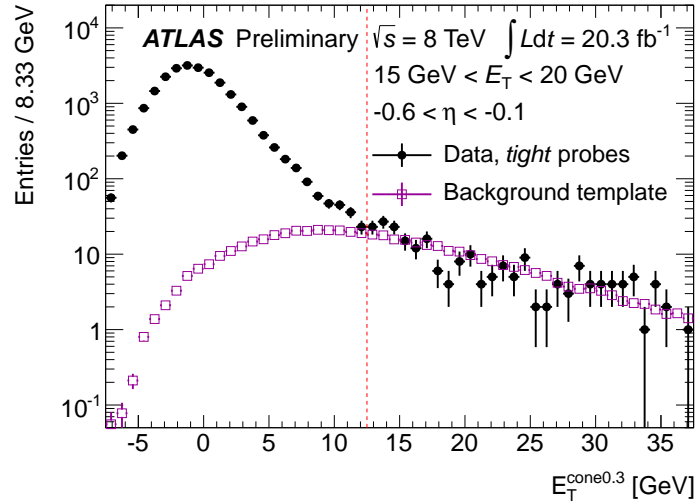


Figure 35: Illustration of the background estimation using the Z_{iso} method in the $15 \text{ GeV} < E_T < 20 \text{ GeV}$, $-0.6 < \eta < -0.1$ bin for probes passing the cut-based *tight* identification. The tag passes cut-based *tight* and the invariant mass of tag and probe must be within a Z mass window of 15 GeV. The threshold chosen for the sideband subtraction is $E_T^{\text{cone0.3}} = 12.5 \text{ GeV}$.