



ATLAS NOTE

ATLAS-CONF-2016-024

1st June 2016



Electron efficiency measurements with the ATLAS detector using the 2015 LHC proton-proton collision data

The ATLAS Collaboration

Abstract

This note summarises the electron efficiency measurements using the 2015 LHC pp collision data at 13 TeV centre-of-mass energy. The data, collected by the ATLAS detector, correspond to an integrated luminosity of 3.2 fb^{-1} . The studies are focused on electron reconstruction, identification, trigger and isolation algorithms. The combination of these algorithms leads to a large variety of operating points for the electron measurement, with diverse signal efficiency versus background rejection performance. Samples of electron candidates with transverse energies above 7 GeV, in the central region of the detector defined by the pseudorapidity range $|\eta| < 2.47$ are selected using the tag-and-probe method for $Z \rightarrow ee$ and $J/\psi \rightarrow ee$ processes. The measurements performed on data are compared to the measurements done on Monte Carlo simulations in the same phase space. The deviations are expressed as data-to-MC ratios, which are provided for a variety of operating points, and used in the physics analyses in order to correct the Monte Carlo simulations.



1. Introduction

In the ATLAS detector [1], electrons and positrons, collectively referred to as electrons, in the central region¹ give rise to tracks in the inner detector and energy deposits in the electromagnetic calorimeter. The calorimeter signals are used in the first level trigger system (L1), and are combined with tracks to reconstruct electron candidates used for the high level trigger (HLT) decision algorithms.

The electron candidates are then further selected against background – such as hadrons and background (non-prompt) electrons originating predominantly from photon conversions and heavy flavour hadron decays – 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. Additionally, electrons can be required to be isolated from other activity in the calorimeter or inner detector to further distinguish them from background objects.

The accuracy of the Monte Carlo (MC) detector simulation to model the electron measurement efficiency plays a crucial role for cross section measurements and searches for new physics. In order to achieve reliable results, the MC samples are corrected to reproduce the efficiencies measured with data. The efficiency measurements of the various steps described above (trigger, reconstruction, identification and isolation) 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. The data measurements are compared to the simulation to obtain corrections (scale-factors) as a function of the electron’s transverse energy E_T and its pseudorapidity η . Previous efficiency measurements were performed for Run-1 data at 7 TeV [2] and 8 TeV [3] centre-of-mass energy.

This note contains the description of the electron reconstruction, identification, isolation and trigger algorithms, as well as the measurements of the corresponding experimental efficiencies for the data collected at a centre-of-mass energy of 13 TeV in 2015 and with 25ns bunch spacing. During this period, the ATLAS detector recorded a data sample corresponding to an integrated luminosity of 3.2 fb^{-1} with all relevant detector components fully functional. The methodology used to extract the various efficiency measurements follows, with small modifications, the one used during Run-1 and described in Ref. [3].

A brief description of the ATLAS detector is presented in Sect. 2. The description of various algorithms together with the optimisations performed in order to adapt these algorithms to the Run-2 running conditions are described in Sects. 3–6. The general methodology of the efficiency measurements and the data and Monte Carlo samples used in this work are presented in Sect. 7. Sect. 8 describes the identification efficiency measurement, while Sect. 9, Sect. 10 and Sect. 11 present the reconstruction, isolation and trigger efficiency measurements, respectively. Sect. 12 contains the conclusions.

¹ ATLAS uses a right-handed coordinate system with its origin at the nominal pp interaction point at the centre of the detector. The positive x -axis is defined by the direction from the interaction point to the centre of the LHC ring, with the positive y -axis pointing upwards, while the nominal beam direction defines the z -axis. The transverse momenta p_T and transverse energies E_T (used thereafter for calorimetric measurements) are measured in the $x - y$ plane. 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 $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$. This note describes measurements performed in the central region, defined by $|\eta| < 2.47$.

2. The ATLAS Detector

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

The inner detector provides a precise reconstruction of tracks within $|\eta| < 2.5$. It consists of four layers of pixel detectors close to the beam-pipe, four layers of silicon microstrip detector modules (SCT) with pairs of single-sided sensors glued back-to-back, and a transition radiation tracker (TRT) at the outer radii (in the range $|\eta| < 2.0$). A track from a charged particle traversing the barrel detector typically has 12 silicon measurement points (hits), of which four are pixel and eight SCT hits, and approximately 35 TRT straw hits.

The innermost pixel layer, the insertable B-layer (IBL) [4], was added between Run-1 and Run-2 of the LHC. It is composed of 14 lightweight staves arranged in a cylindrical geometry, each made of 12 silicon planar sensors in its central region and four 3D sensors at each end. The IBL pixel dimensions are $50 \times 250 \mu\text{m}^2$ in the ϕ and z directions (compared with $50 \times 400 \mu\text{m}^2$ for other pixel layers). The smaller radius and the reduced pixel size result in improvements of both the transverse and longitudinal impact parameter resolutions. Moreover, new detector services for the pixel detector have been implemented which significantly reduce the material at the boundaries of the active tracking volume.

The TRT offers substantial discrimination between electrons and charged hadrons over a wide energy range through the detection of transition radiation photons. During Run-1, several leaks developed in the TRT exhaust system, leading to a large loss of highly expensive xenon gas. In order to reduce the operation costs in 2015, a few parts of the TRT (internal layer of the Barrel TRT and one endcap wheel on each side of the detector) were running with argon gas mixture.

The central 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 endcap sections (EMEC) covering $1.375 < |\eta| < 3.2$. For $|\eta| < 2.5$, it is divided into three longitudinal 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 pre-sampler 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 ($|\eta| < 2.47$) proceeds in several steps:

- **Seed-cluster reconstruction:** A sliding window with a size of 3×5 in units of 0.025×0.025 , corresponding to the granularity of the EM calorimeter middle layer, in $\eta \times \phi$ space is used to search for electron cluster "seeds" as longitudinal towers² with total cluster transverse energy above 2.5 GeV. The clusters are then formed around the seeds using a clustering algorithm [5] that allows for duplicates to be removed. The cluster kinematics are reconstructed using an extended window depending on the cluster position in the calorimeter. The efficiency of this cluster search ranges from 95% at $E_T = 7$ GeV to more than 99% above $E_T = 15$ GeV.
- **Track reconstruction:** Track reconstruction proceeds in two steps: pattern recognition and track fit. The standard ATLAS pattern recognition uses the pion hypothesis for energy loss due to interactions with the detector material. This has been complemented with a modified pattern recognition algorithm which allows up to 30% energy loss at each intersection of the track with the detector material to account for possible bremsstrahlung. If a track 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 region of interest³, a second attempt is performed with the new pattern recognition using an electron hypothesis that allows for larger energy loss. Track candidates are then fit 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 [6]. If a track candidate fails the pion hypothesis track fit (for example, due to large energy losses), it is refit with the electron hypothesis. In this way, a specific electron-oriented algorithm has been integrated into the standard track reconstruction. It improves the performance for electrons and has minimal interference with the main track reconstruction.
- **Electron specific track fit:** The obtained tracks are loosely matched to EM clusters using the distance in η and ϕ between the position of the track, after extrapolation, in the calorimeter middle layer and the cluster barycentre. The matching conditions account for energy-loss due to bremsstrahlung and the number of precision hits in the silicon detector. Tracks that have significant number of precision hits (≥ 4) and are loosely associated to electron clusters are refit using an optimised Gaussian Sum Filter (GSF) [7], which takes into account the non-linear bremsstrahlung effects.
- **Electron candidate reconstruction:** The matching of the track candidate to the cluster seed completes the electron reconstruction procedure. A similar matching as the one described above is repeated for the refit track with stricter conditions.

If several tracks fulfil the matching condition, one track is chosen as "primary" track. The choice is based on an algorithm using the cluster-track distance R calculated using different momentum hypotheses, the number of pixel hits and the presence of a hit in the first silicon layer [3]. Electron candidates without any associated precision hit tracks are removed and considered to be photons. The efficiency of this association and subsequent track quality cuts is measured as the "reconstruction efficiency" in Sect. 9. The electron cluster is then re-formed using 3×7 (5×5) longitudinal towers

² The $\eta \times \phi$ space of the EM calorimeter is divided into a grid of $N_\eta \times N_\phi = 200 \times 256$ elements of size $\Delta\eta^{\text{tower}} \times \Delta\phi^{\text{tower}} = 0.025 \times 0.025$, called towers. Inside each of these elements, the energy of cells in all longitudinal layers (including the front, middle, and back calorimeter layers as well as, for $|\eta| < 1.8$, the presampler) is summed into the tower energy, with the energy of cells spanning several towers distributed uniformly among the participating towers.

³ For each seed EM cluster passing loose shower shape requirements of $R_\eta > 0.65$ and $R_{\text{had}} < 0.1$ (for the definition of these variables, see Table 1) a region of interest with a cone-size of $\Delta R = 0.3$ around the seed cluster barycenter is defined.

of cells in the barrel (endcaps) of the EM calorimeter. The energy of the clusters is calibrated to the original electron energy using multivariate techniques [8] based on simulated MC samples.

The four-momentum of the electrons is computed using information from both the final calibrated energy cluster and the best track matched to the original seed cluster. The energy is given by the final calibrated cluster, while the ϕ and η directions are taken from the corresponding track parameters with respect to the beam-line.

For Run-2 analyses, the electron measurements are performed by requiring the track associated with the electron to be compatible with the primary interaction vertex of the hard collision, in order to reduce the background from conversions and secondary particles. The track parameters are calculated in a reference frame where the z-axis is taken along the measured beam-line position. The following conditions are applied together with all the identification operating points considered in this note: $d_0/\sigma_{d_0} < 5$ and $\Delta z_0 \sin \theta < 0.5$ mm, where the impact parameter d_0 is the distance of closest approach of the track to the measured beam-line, z_0 is the distance along the beam-line between the point where d_0 is measured and the beam-spot position, and θ is the polar angle of the track. To assess the compatibility with the primary vertex of the hard collision the Δz_0 between the track and the primary vertex is employed. This vertex is selected from the reconstructed primary vertices (compatible with the beam-line) as the one with the highest sum of transverse momenta of the associated tracks. σ_{d_0} represents the estimated uncertainty of the d_0 parameter, and θ is the polar angle of the track. The efficiency of these requirements in data and MC is estimated together with the efficiency of the various identification operating points.

4. Electron Identification

To determine whether the reconstructed electron candidates are signal-like objects or background-like objects such as hadronic jets or converted photons, algorithms for electron identification (ID) are applied. The ID algorithms use quantities related to the electron cluster and track measurements including calorimeter shower shapes, information from the transition radiation tracker, track-cluster matching related quantities, track properties, and variables measuring bremsstrahlung effects for distinguishing signal from background. These quantities are summarised in Table 1.

For Run-2, several changes to the input variables used for electron ID have been introduced. Taking advantage of the IBL, the number of hits in this innermost pixel layer is used for discriminating between electrons and converted photons. This criterion was also used in Run-1, but with what is now the second-to-innermost pixel layer.

Moreover, the change in the TRT gas led to modifications in the detector response and prompted the introduction of a new discriminating variable in the electron identification algorithms. In Run-1, only the fraction of high-threshold hits was used from the TRT as a signature of transition radiation to distinguish electrons from hadrons. In Run-2, a likelihood method based on the TRT high-threshold hits is introduced to compensate for the lower transition radiation absorption probability of the argon. The TRT likelihood method uses the high-threshold probability of each TRT hit to construct a discriminant variable, referred to here as eProbabilityHT. The probability for each TRT hit to exceed the high level threshold depends on the straw gas type, the Lorentz factor γ calculated from the track p_T under a particle type hypothesis, and the geometry: detector partition, straw layer, track-to-wire distance and the hit coordinates (z for the barrel and radius for the endcaps).

Table 1: Definitions 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. This variable is only used below 100 GeV because it is known to be inefficient at high energies.	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 conditions	Number of hits in the innermost pixel 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 with respect to the beam-line	d_0
	Significance of transverse impact parameter defined as the ratio of d_0 and its uncertainty	d_0/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	Likelihood probability based on transition radiation in the TRT	eProbabilityHT
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 track extrapolated from the perigee	$\Delta\phi_2$
	Defined as $\Delta\phi_2$, but the track momentum is rescaled to the cluster energy before extrapolating the track from the perigee to the middle layer of the calorimeter	$\Delta\phi_{\text{res}}$
	Ratio of the cluster energy to the track momentum	E/p

The re-optimisation of the ID algorithms for Run-2 is based on MC simulation samples. Electron candidates from MC simulations of $Z \rightarrow ee$ and dijet events are used, in addition to $J/\psi \rightarrow ee$ and minimum bias events at low E_T . However, the distributions of several input variables in data tend to be wider or shifted more towards the background distributions than in the MC, due to inaccuracies in the detector description and the modelling of the shower shapes in GEANT [9]. To account for this, corrections derived from data in the form of simple linear shifts or width adjustments to the MC distributions were applied to the electron candidates' distributions used in the optimisation, to ensure that the performance of the operating points derived in MC would correspond to the one observed in the data. This procedure is only used for the described optimisation, while MC samples used for physics analyses are corrected using event weights to account for mismodelling of tracking properties or shower shapes, and to reproduce the efficiencies measured with data.

The baseline ID algorithm used for Run-2 data analyses is the likelihood-based (LH) method. It is a multivariate analysis (MVA) technique that simultaneously evaluates several properties of the electron candidates when making a selection decision. The LH method uses the signal and background probability density functions (PDFs) of the discriminating variables⁴. Based on these PDFs, an overall probability is calculated for the object to be signal or background. The signal and background probabilities for a given electron are then combined into a discriminant $d_{\mathcal{L}}$ on which a requirement is applied:

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

where \vec{x} is the vector of discriminating 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. This allows for better background rejection for a given signal efficiency than a "cut-based" algorithm that would use selection criteria sequentially on each variable. In addition to the variables used as input to the LH discriminant, simple selection criteria are used for the variables counting the number of hits on the track.

Three levels of identification operating points are typically provided for electron ID. These are referred to, in order of increasing background rejection, as *Loose*, *Medium*, and *Tight*. If not otherwise specified, the operating points will refer to the LH identification algorithm. The *Loose*, *Medium* and *Tight* operating points are defined such that the samples selected by them are subsets of one another. Each operating point uses the same variables to define the LH discriminant, but the selection on this discriminant is different for each operating point. Thus, electrons selected by *Medium* are all selected by *Loose*, and *Tight* electrons are all selected by *Medium*.

The distributions of electron shower shapes depend on the amount of material the electrons pass through, and therefore vary with the pseudorapidity of the electron candidates. In addition, significant changes to the shower shapes and track properties are expected with increasing energy. The ID operating points were consequently optimised in several bins in $|\eta|$ and E_T . The performance of the LH identification algorithm is illustrated in Fig. 1. Depending on the operating point, the signal (background) efficiencies for electron candidates with $E_T = 25$ GeV are in the range from 78 to 90% (0.3 to 0.8%) and increase (decrease) with E_T .

⁴ All variables presented in Table 1 are used as PDFs except E/p , w_{stot} and $\Delta\phi_2$, and the variables counting the number of hits on the track. Simple selection criteria are always applied to the latter, while the first three are only used as rectangular cuts in either the *Tight* operating point at high E_T or the cut-based ID, as described below.

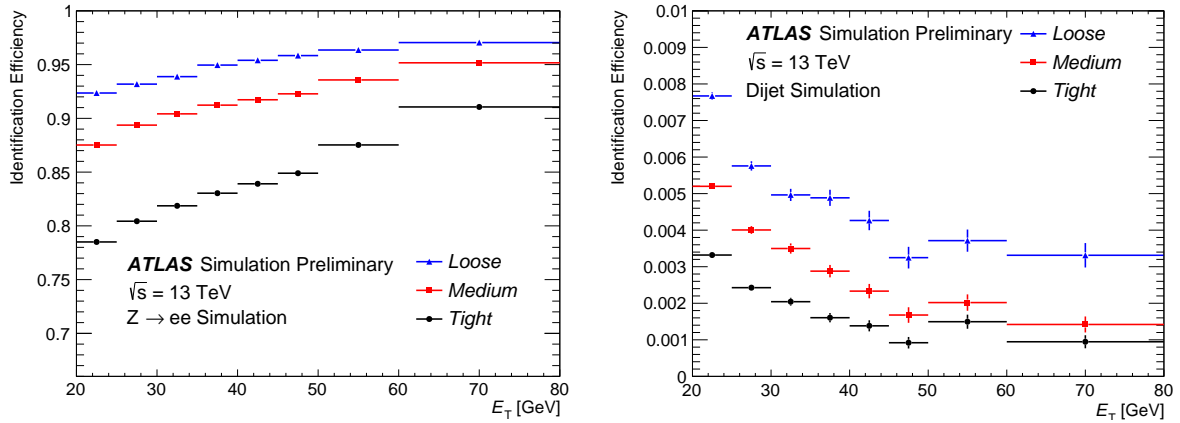


Figure 1: The efficiency to identify electrons from $Z \rightarrow ee$ decays (left) and the efficiency to identify hadrons as electrons (background rejection, right) estimated using simulated dijet samples. The efficiencies are obtained using Monte Carlo simulations, and are measured with respect to reconstructed electrons. The candidates are matched to true electron candidates for $Z \rightarrow ee$ events. For background rejection studies the electrons matched to true electron candidates are not included in the analysis. Note that the last bin used for the optimisation of the ID is 45-50 GeV, which is why the signal efficiency increases slightly more in the 50 GeV bin than in others, and the background efficiency increases in this bin as well.

The electron identification performance may be influenced by the parasitic collisions taking place in the same beam crossing (in-time pileup) or a consecutive bunch crossing (out-of-time pileup) as the hard pp collision producing the electron candidate. The number of reconstructed primary vertices is indicative of the level of pileup in each event, with the average number of primary vertices (eight per event) corresponding to an average pileup of 13.7. Since some shower shape distributions depend on the number of pileup collisions per bunch crossing, the cut on the LH discriminant value is loosened as a function of the number of primary vertices. This is done to ensure that the LH identification remains efficient at high pileup, without drastically increasing the amount of background accepted by the LH selection. The optimisation included simulations with a number of pileup collisions of up to 40, covering the range of the pileup observed in 2015.

At high E_T , some of the calorimeter variable distributions are different from the typical distributions obtained with $Z \rightarrow ee$ and used to construct the LH PDFs. Higher energy electrons tend to deposit relatively smaller fractions of their energy in the early layers of the EM calorimeter, and more in the later layers of the EM calorimeter or even in the hadronic calorimeter. *Loose* and *Medium* were deemed to be loose enough to be robust against these E_T -dependent changes. However, the tighter requirement used in *Tight* would lead to inefficiencies at high E_T , if not handled properly. Thus, for electron candidates with E_T above 125 GeV, *Tight* uses the same discriminant selection as *Medium* but adds rectangular cuts on w_{stot} and E/p , which were found to be particularly effective at discriminating signal from background at high E_T .

In addition to the multivariate approach used in the LH method described so far, a cut-based method using a set of rectangular cuts on the electron ID discriminating variables was used in Run-1. This method encompasses a similar set of operating points. The cut-based *Loose* operating point relies primarily on information from the hadronic calorimeter and the first two layers of the EM calorimeter for distinguishing signal from background. The cut-based *Medium* operating point adds information from the TRT, the transverse impact parameter, and the third layer of the EM calorimeter, in addition to tighter cuts on the

variables from the cut-based *Loose* ID. Finally, the cut-based *Tight* operating point adds track-cluster matching variables such as E/p and $\Delta\phi_2$, and uses tighter cuts than the cut-based *Medium* ID for the remaining variables. The cut-based algorithms were optimised for Run-2 and used as for cross-checks during the 2015 data taking. These selection criteria are used in the analyses presented in this note for the background selections used in the tag-and-probe measurements. The cut-based algorithms are not used in physics analyses in Run-2 and therefore their efficiencies are not presented in this note.

5. Electron Isolation

In addition to the identification criteria described above, most analyses require electrons to fulfil isolation requirements, to further discriminate between signal and background. The isolation variables quantify the energy of the particles produced around the electron candidate and allow to disentangle prompt electrons (from heavy resonance decays, such as $W \rightarrow e\nu$, $Z \rightarrow ee$) from other, non-isolated electron candidates such as electrons originating from converted photons produced in hadron decays, electrons from heavy flavour hadron decays, and light hadrons mis-identified as electrons. Two discriminating variables have been designed for that purpose:

- a *calorimetric* isolation energy, $E_T^{\text{cone0.2}}$, defined as the sum of transverse energies of topological clusters [5, 10], calibrated at the electromagnetic scale, within a cone of $\Delta R = 0.2$ around the candidate electron cluster. Only clusters with reconstructed positive energy are considered in the sum. The E_T contained in a rectangular cluster of size $\Delta\eta \times \Delta\phi = 0.125 \times 0.175$ centred around the electron cluster barycentre is subtracted. An (E_T, η) dependent correction is then applied to account for the electron energy leakage outside this cluster. The contribution from pileup and the underlying event activity is corrected for on an event-by-event basis using the technique outlined in Ref. [11]⁵;
- a *track* isolation, $p_T^{\text{varcone0.2}}$, defined as the sum of transverse momenta of all tracks, satisfying quality requirements, within a cone of $\Delta R = \min(0.2, 10 \text{ GeV}/E_T)$ around the candidate electron track and originating from the reconstructed primary vertex of the hard collision, excluding the electron associated tracks, *i.e.* the electron track and additional tracks from converted bremsstrahlung photons. The track quality requirements are :
 - $E_T > 1 \text{ GeV}$;
 - $n_{\text{Si}} \geq 7, n_{\text{Si}}^{\text{hole}} \leq 2, n_{\text{Pixel}}^{\text{hole}} \leq 1, n_{\text{mod}}^{\text{sh}} \leq 1$, where $n_{\text{Si}}^{\text{hole}}$ and $n_{\text{Pixel}}^{\text{hole}}$ are the number of missing hits in the silicon (pixel and SCT) and pixel detector respectively, and $n_{\text{mod}}^{\text{sh}}$ is the number of hits in the silicon detector assigned to more than one track;
 - $|\Delta z_0 \sin \theta| < 3 \text{ mm}$.

The distributions of these two discriminating variables are illustrated in Fig. 2 for electrons of E_T greater than 27 GeV satisfying the *Tight* requirements, selected in events consistent with originating from the $Z \rightarrow ee$ process. The negative tail of $E_T^{\text{cone0.2}}$ originates from the correction for pileup and the underlying event activity. A slight discrepancy is observed in the region at large $E_T^{\text{cone0.2}}$ and $p_T^{\text{varcone0.2}}$ values, where the background dominates (note that no background subtraction is applied here).

⁵ To account for the variation of the ambient energy density as a function of the pseudorapidity, it is estimated in two wide pseudorapidity bins : $|\eta| < 1.5$ and $1.5 < |\eta| < 3.0$.

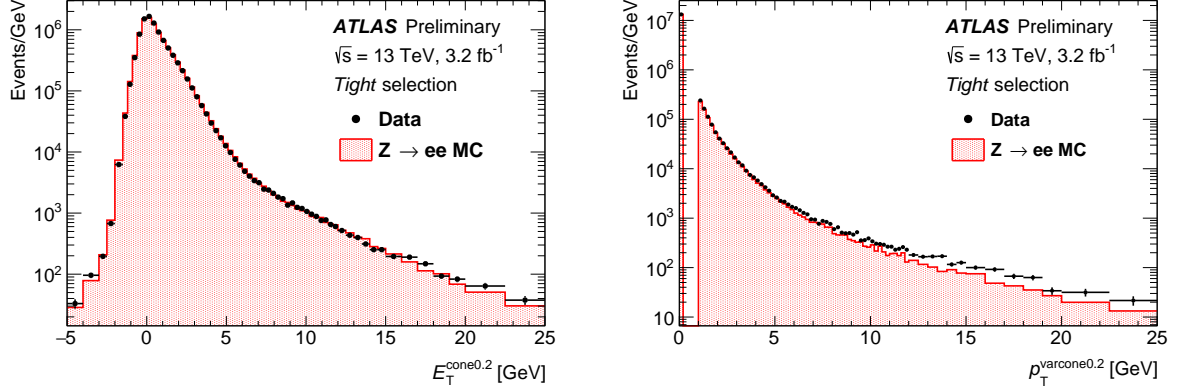


Figure 2: Distributions of $E_T^{\text{cone0.2}}$ (left) and $p_T^{\text{varcone0.2}}$ (right) for electrons from $Z \rightarrow ee$, in the data (dots with error bars) and the simulation (full histogram, normalised to data).

A variety of selection requirements on the quantities $E_T^{\text{cone0.2}}/E_T$ and $p_T^{\text{varcone0.2}}/E_T$ have been defined to select isolated electron candidates. The resulting operating points are divided into two classes:

- efficiency targeted operating points: varying requirements are used in order to obtain a given isolation efficiency ε_{iso} , which can be either constant or as a function of E_T . Typical ε_{iso} values are 90(99)% for $E_T = 25(60)$ GeV, estimated for electrons from simulated $Z \rightarrow ee$ events. For transverse energies between 7 and 15 GeV, simulated electrons from J/ψ decays have been used to determine the upper cuts on $E_T^{\text{cone0.2}}$ and $p_T^{\text{varcone0.2}}$, and only events satisfying $\Delta\phi > 0.3$ have been considered, where $\Delta\phi$ is the azimuthal separation between the two electrons.
- fixed requirement operating points: in this case the upper thresholds on the isolation variables are constant. These operating points were optimised by maximising the expected sensitivities of $H \rightarrow 4\ell$ and multilepton supersymmetry searches.

Table 2 shows the definition of the various operating points used for electron isolation. Fixed requirement operating points are typically preferred in physics analyses using low E_T electrons and requiring high background rejection, while in the high E_T range looser operating points are adopted in order to maintain high signal efficiency.

6. Electron Trigger

The ATLAS online data processing reconstructs and identifies electron candidates both at the L1 trigger and at the HLT. At L1, the electron triggers use the signals recorded in the electromagnetic (EM) and hadronic calorimeters within regions of 4×4 trigger "towers" (corresponding to $\Delta\eta \times \Delta\phi \approx 0.4 \times 0.4$) to calculate the energy in the inner region (core) and the surrounding (isolation) region. The E_T thresholds used in the trigger decision can be set differently for different η regions that are delimited by the η values of $N \times 0.1$ ($N = 1, 2, \dots$). This allows a coarse tuning of thresholds taking into account different energy responses in different regions. A veto on the hadronic leakage can also be applied at L1 by requiring that the amount of energy measured in the hadronic calorimeter behind the core of the EM cluster relative to the EM cluster energy is less than a certain value. An isolation cut on the transverse energy in an annulus

Operating point	Efficiency		
	calorimeter isolation	track isolation	total efficiency
LooseTrackOnly	-	99%	99%
Loose	99%	99%	~ 98%
Tight	96%	99%	~ 95%
Gradient	$0.1143\% \times E_T + 92.14\%$	$0.1143\% \times E_T + 92.14\%$	90/99% at 25/60 GeV
GradientLoose	$0.057\% \times E_T + 95.57\%$	$0.057\% \times E_T + 95.57\%$	95/99% at 25/60 GeV

Operating point	Cut value	
	calorimeter isolation	track isolation
FixedCutLoose	0.20	0.15
FixedCutTightTrackOnly	-	0.06
FixedCutTight	0.06	0.06

Table 2: Electron isolation operating point definitions. The upper table illustrates the efficiency targeted operating points, and the numbers expressed in percents represent the target efficiencies used in the operating point optimisation procedure. For the Gradient and GradientLoose operating points, E_T is in GeV. The fixed requirement operating points are shown in the lower table. The calorimeter and track isolation refer to the selection based on $E_T^{\text{cone}0.2}/E_T$ and $p_T^{\text{varcone}0.2}/E_T$, respectively.

of calorimeter towers around the EM candidate relative to the EM cluster transverse energy can also be used at L1. The isolation and hadronic leakage veto requirements are not used for electron candidates with a transverse energy of greater than 50 GeV.

At the HLT, electron candidates are reconstructed and selected in several steps to reject potential background candidates early, thereby reducing the event rate to a level where more precise offline-like algorithms, using information from calorimeter and tracking, can be applied in the allowed latency range. Fast EM calorimeter algorithms build clusters from the calorimeter cells within the region of interest (RoI) ($\Delta\eta \times \Delta\phi = 0.4 \times 0.4$) identified at the L1 step, using the second layer of the EM calorimeter to find the cell with the largest deposited transverse energy in the RoI. Electrons are identified by applying requirements on the energy deposit in the hadronic calorimeter R_{had} , on the R_η shower shape variable, on E_{ratio} as well as on the transverse energy of the cluster. Tracks reconstructed using a simplified (fast) tracking algorithm, with a minimum p_T of 1 GeV, are associated to clusters within $\Delta\eta < 0.2$. The second step of the HLT relies on precise offline-like algorithms, based on candidates selected in the first step. EM calorimeter clusters are built following the same techniques described in Sect. 3. Additional requirements on the shower shapes are applied to reduce the rate further before precision tracking. Electrons are reconstructed with clusters matched to precision tracks extrapolated to the second layer of the EM calorimeter within $|\Delta\eta| < 0.05$ and $|\Delta\phi| < 0.05$. The ID of electrons at the HLT is performed employing the same discriminating variables as the offline ID described in Sect. 4.

During Run-1, the electron trigger used a cut-based ID at the HLT. However, inefficiencies can arise by applying a LH identification selection offline with a cut-based selection in the trigger, so the electron LH identification was adapted to also work at the HLT for Run-2.

The likelihood-based identification for the online environment is similar to the offline identification. The discriminant is optimised with the online reconstructed shower shapes and track variables. The composition of the likelihood is the same as offline with the exception of momentum loss due to bremsstrahlung, $\Delta p/p$, which is not accounted for in the online environment. The cut on the LH identification discriminant

is adjusted as a function of the average number of interactions per crossing in order to account for pileup effects.

The efficiency $\varepsilon_{\text{trigger}}$ for a given trigger is defined with respect to a given offline identification algorithm and isolation operating point. It is defined as the fraction of events having been selected by the trigger in a sample of events with reconstructed electrons identified by the offline algorithm.

7. Efficiency Measurement Methodology

The tag-and-probe method has been used in all of the analyses described below. The method employs events containing well-known resonance decays to electrons, namely $Z \rightarrow ee$ and $J/\psi \rightarrow ee$. A strict selection on one of the electron candidates (called "tag") together with the requirements on the di-electron invariant mass, and in case of J/ψ , on the lifetime information, allows for a loose pre-identification of the other electron candidate ("probe"). In order not to bias the selected probe sample, each valid combination of electron tag-probe pairs in the event is considered, such that an electron can be the tag in one pair and the probe in another. The probe is used for the measurement of the reconstruction, identification, isolation and trigger efficiencies, after accounting for the residual background contamination. This method delivers sufficient data events to cover a range in E_T from 7 to 200 GeV and $|\eta| < 2.47$, with a good granularity. The low E_T range (typically from 7 to 20 GeV) is covered by the $J/\psi \rightarrow ee$ decays, while $Z \rightarrow ee$ events are used for measurements above 15 GeV.

The efficiency to find and select an electron in the ATLAS detector is not measured as a single quantity but is divided into different components, namely reconstruction, identification, isolation, and trigger efficiencies. The total efficiency $\varepsilon_{\text{total}}$ for a single electron can be written as:

$$\varepsilon_{\text{total}} = \varepsilon_{\text{reconstruction}} \times \varepsilon_{\text{identification}} \times \varepsilon_{\text{isolation}} \times \varepsilon_{\text{trigger}} \quad (2)$$

with the various efficiency components measured with respect to the previous step.

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 simulated samples need to be corrected to reproduce the measured data efficiencies as closely as possible. For this reason, the efficiencies are estimated both in data and in simulation. For the efficiencies measured in simulated samples, the same cuts are used to select the probe electrons as in data. However, no background subtraction needs to be applied on the simulated samples; instead, the reconstructed electron track is required to have hits in the inner detector which originate from the true electron during simulation. The ratio between data and MC efficiencies is used as a multiplicative correction factor for MC. These data-to-MC correction factors are usually rather close to unity. Deviations stem from the mismodelling of tracking properties or shower shapes in the calorimeters.

Since the electron efficiencies depend on E_T and η , the measurements are performed in two-dimensional bins in (E_T, η) , as specified in Tables 3 and 4. Residual effects coming from 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 Sect. 8.3.

be referred to as “track quality” hereafter). 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 efficiency measurements described in this note, two different decays of on-shell produced resonances are used: $Z \rightarrow ee$ for electrons with $E_T > 15$ GeV and $J/\psi \rightarrow ee$ for electrons with $7 \text{ GeV} < E_T < 20$ GeV. In the overlapping 15-20 GeV bin, results from $Z \rightarrow ee$ and $J/\psi \rightarrow ee$ are combined.

8.1. Tag-and-Probe with $Z \rightarrow ee$ Events

The tag-and-probe method is used with $Z \rightarrow ee$ decays to obtain a sample of unbiased electron candidates with $E_T > 15$ GeV. These electron candidates can then be used to determine the electron identification efficiency for the various operating points. As there is significant background contamination, particularly for $E_T < 25$ GeV, the background needs to be statistically subtracted. To distinguish between signal electrons and background, two methods are used: the Z_{mass} method uses the invariant mass of the tag-probe pair as the discriminating variable for background subtraction, while the Z_{iso} method uses the calorimetric isolation distribution of the probe electron. These methods are treated as systematic variations on a single measurement.

8.1.1. Event Selection

Events are selected using a single electron trigger with an E_T threshold of 24 GeV and *Medium* identification requirements. These events are then required to have at least two reconstructed electron candidates in the central region of the detector, $|\eta| < 2.47$, with opposite charges. One of the two electrons, the tag, must have $E_T > 25$ GeV, be matched to a trigger electron object within $\Delta R < 0.07$, and be outside the transition region between the barrel and the endcap of the electromagnetic calorimeter, $1.37 < |\eta| < 1.52$. The tag must satisfy an identification requirement, nominally *Tight*. The probes, meanwhile, are required to have $E_T > 15$ GeV and pass the track quality criteria. The discriminating variables for the probe are displayed in Appendix A (Figs. 18 and 19).

8.1.2. Z_{mass} Method

In the Z_{mass} method, the identification efficiency is measured using the invariant mass spectrum of the tag-probe pairs in each E_T and η bin of the probe electron. The signal is obtained from tag-probe pairs with an invariant mass m_{ee} within 15 GeV of the Z boson nominal mass. The background is subtracted using templates which are constructed using probes that fail identification and isolation criteria, in order to model the background distribution with minimal contribution from signal electrons. Any residual signal contamination is estimated using simulated events and removed from the templates. The templates are then normalised to the data using the sidebands of the invariant mass distribution.

Different challenges exist for determining the numerator and the denominator for the ID efficiency in the Z_{mass} method. For the denominator, there is a significant amount of background which must be properly estimated, and this is the primary source of systematic uncertainty in the measurement. For the numerator, meanwhile, the background contribution is typically several orders of magnitude smaller than the signal. However, as the majority of objects which pass the numerator selection are real electrons, there is significant signal contamination in the template normalisation regions at numerator level.

Thus, the calculation of the numerator and denominator is performed in different ways, to ensure that the challenges of each are properly addressed. Note that the same template distributions are used for both the numerator and the denominator, but the normalisation procedure differs between the two. Nominally, the normalisation region is the high invariant mass region $120 \text{ GeV} < m_{ee} < 250 \text{ GeV}$.

As the invariant mass distribution of same-sign tag-probe pairs has less signal contamination than that of the opposite-sign tag-probe pairs, the numerator normalisation is performed using the same-sign sample. At denominator level, however, the opposite-sign distribution must be used for the normalisation procedure as there is significant opposite-sign background present—and this is particularly important because some background processes such as W +jets predominantly contribute to the opposite-sign invariant mass distribution. Signal contamination at denominator level is accounted for, and estimated from the number of *Tight* events in the normalisation region, divided by the *Tight* efficiency. This contamination is then subtracted from the normalization region.

Examples of the invariant mass distribution for the numerator and denominator for one bin in (E_T, η) are shown in Fig. 3.

To assess the systematic uncertainties, several variations are considered. Since the invariant mass distribution is used for the background subtraction, the m_{ee} region of interest for determining the numerator and denominator is varied between mass windows of 10, 15, and 20 GeV on each side of the Z mass. The tag identification criterion is varied by adding a calorimeter isolation requirement in addition to the nominal *Tight*, or by loosening the tag requirement to *Medium* with calorimeter isolation. Finally, the background template is varied. For $E_T < 30 \text{ GeV}$, the background contamination is large, so the template normalisation is tested by using as a variation the invariant mass region $60 \text{ GeV} < m_{ee} < 70 \text{ GeV}$ instead of the nominal $120 \text{ GeV} < m_{ee} < 250 \text{ GeV}$ range. For $E_T > 30 \text{ GeV}$, the signal contamination in the templates is a larger concern as the background is much smaller. Moreover, there are far fewer events in this low mass region due to the kinematics, so only the $120 \text{ GeV} < m_{ee} < 250 \text{ GeV}$ normalisation is used. Instead, for $E_T > 30 \text{ GeV}$, a different template selection criterion is used as a variation. All possible combinations of these variations are used to determine the systematic uncertainties, as described in Sect. 8.3. The template choice is found to be the most important systematic effect at low E_T , while the variation on the invariant mass window is dominant at higher E_T .

8.1.3. Z_{iso} Method

The Z_{iso} method is another method of background subtraction used to obtain the electron identification efficiency and scale-factors for electron energies of $15 \text{ GeV} < E_T < 200 \text{ GeV}$ and $|\eta| < 2.47$ using the $Z \rightarrow ee$ decay. As a discriminating variable between signal and background, the calorimetric isolation $E_T^{\text{cone}0.3}$ of the probe is used. The calorimetric isolation $E_T^{\text{cone}0.3}$ follows the definition described in Sect. 5 with a cone of $\Delta R = 0.3$ around the electron. Therefore, signal electrons should be accumulated around low values of the isolation variable whereas background should be found mainly at high values of $E_T^{\text{cone}0.3}$, allowing for the definition of a signal and background dominated region.

To estimate the background, a model is constructed representing the shape of the isolation distribution for background objects. For the construction of a background template, reconstructed probe electrons passing track quality criteria but failing the cut-based *Loose* identification or, alternatively, selected shower shape cuts, are selected. Furthermore, these probes are required to have the same charge as the tag electron. Real electrons passing the described background selection are estimated by applying the background selection to a $Z \rightarrow ee$ MC sample and are then subtracted from the background template. The $Z \rightarrow ee$ MC is

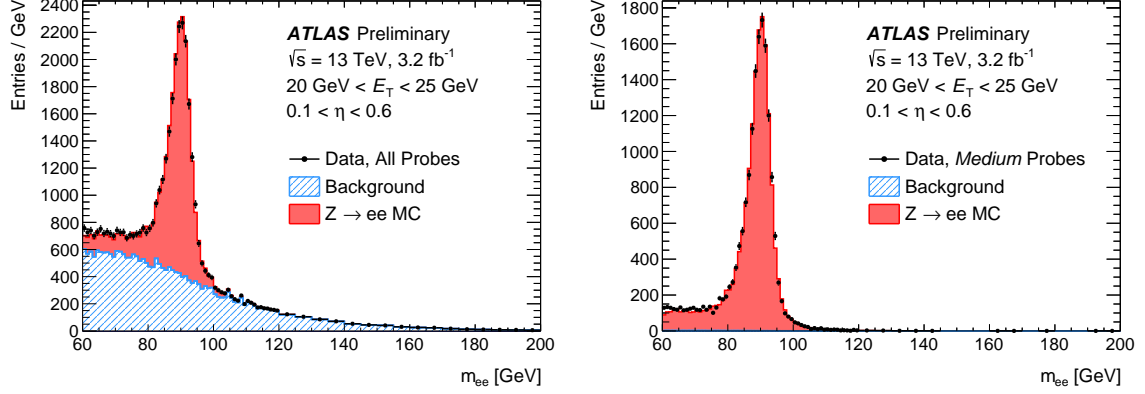


Figure 3: Illustration of the background estimation using the Z_{mass} method in the $20 \text{ GeV} < E_T < 25 \text{ GeV}$, $0.10 < \eta < 0.60$ bin, at reconstruction level (left) and for probes passing the *Medium* identification (right). Efficiencies are determined by taking the ratio of background-subtracted probes passing the likelihood identification over background-subtracted probes at reconstruction level. The simulated $Z \rightarrow ee$ sample is shown for illustration and is scaled to match the background-subtracted data in the m_{ee} window.

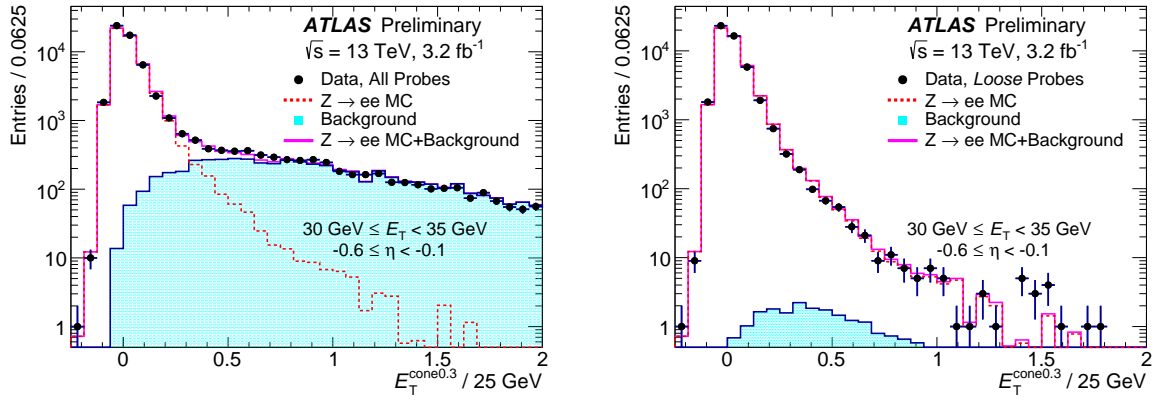


Figure 4: Examples for denominator (left, reconstruction level) and numerator (right, *Loose* identification) probe isolation distribution. The black dots represent data, the red dashed lines the $Z \rightarrow ee$ MC distribution. The light blue shaded area indicates the background model scaled to the tail of the data probe isolation distribution taking into account the signal contamination. The purple line represents the sum of the $Z \rightarrow ee$ MC and the background distribution which should describe the data points.

scaled to data by reconstructing the Z boson mass peak using oppositely charged tag-probe pairs passing a tight selection to reduce the background. The obtained background template is scaled to data in the background dominated tail of the probe isolation distribution, defined by $E_T^{\text{cone}0.3}/25 \text{ GeV} > 0.5$. An example of the probe isolation distribution and the obtained background template for one bin in (E_T, η) is shown in Fig. 4.

To evaluate the systematic uncertainties, several variations are considered. These include a modification of the tag-probe pair invariant mass window cut to 10 GeV or 20 GeV instead of 15 GeV around the Z boson mass, as done in the Z_{mass} method. The tag selection is varied by adding a requirement on the calorimeter isolation ($E_T^{\text{cone}0.3}/25 \text{ GeV} < 0.3$) to the required *Tight* selection. In addition, different identification cuts are inverted to form two alternative background templates, and the definition of the background scaling

region is varied from $E_T^{\text{cone}0.3}/25 \text{ GeV} > 0.5$ to $E_T^{\text{cone}0.3}/25 \text{ GeV} > 0.4$ or 0.6 . A different choice of the isolation variable with a larger cone ($E_T^{\text{cone}0.4}$) around the probe electron allows for an evaluation of the uncertainty due to the choice of the discriminating variable. All possible combinations of the variations are considered. The main systematic effect originates from the background template variations.

8.2. Tag-and-Probe with $J/\psi \rightarrow ee$ Events

$J/\psi \rightarrow ee$ events are used to measure the identification efficiency of electrons with $7 \text{ GeV} < E_T < 20 \text{ GeV}$. At such low energies, the probe sample suffers from a significant background contamination, which can be estimated using the reconstructed di-electron invariant mass (m_{ee}) of the selected tag-probe pairs. Furthermore, the J/ψ sample is composed of two contributions with significantly different electron identification efficiencies. The first contribution comes from the *prompt production* of the J/ψ mesons, which are produced directly in proton-proton collisions or in radiative decays of directly produced heavier charmonium states. The second contribution arises from the *non-prompt production* of the J/ψ mesons, where the J/ψ come from b -hadron decays. The electrons from the decay of prompt J/ψ particles are expected to be more isolated, and therefore to have identification efficiencies closer to those of isolated electrons from other physics processes of interest in the same transverse momentum range, compared to those from non-prompt production.

The two production modes can be distinguished by exploiting the long lifetime of b -hadrons, and the consequently displaced decay vertex of non-prompt J/ψ . This displacement can be estimated using the following variable, called pseudo-proper time:

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

where L_{xy} is the distance between the J/ψ vertex and the primary vertex in the transverse plane, and $m_{\text{PDG}}^{J/\psi}$ and $p_T^{J/\psi}$ are the mass and the reconstructed transverse momentum of the J/ψ meson, respectively.

Both methods of efficiency measurement described in Ref. [3] have been used for these measurements. The key difference between the two methods is in the different approach for separating between prompt and non-prompt components using the pseudo-proper time distribution. The first method is called J/ψ τ -cut and applies an explicit cut on the τ distribution to enrich the selected sample with prompt J/ψ . The residual fraction of the non-prompt component is determined from MC simulation and the ATLAS measurement of the non-prompt fraction in $J/\psi \rightarrow \mu\mu$ events [20]. The J/ψ τ -fit method utilises the full sample and extracts the non-prompt fraction by fitting the pseudo-proper time distribution both before and after applying the identification cuts.

8.2.1. Event Selection

The baseline event selection is the same for both the J/ψ τ -cut and J/ψ τ -fit methods. Events are required to have at least two electrons with $E_T > 5 \text{ GeV}$ and pseudorapidity $|\eta| < 2.47$. At least one of five dedicated J/ψ two-object triggers must fire in the event. Each of these triggers requires a likelihood *Tight* trigger electron identification and E_T above certain threshold for one trigger object, while only demanding the electromagnetic cluster E_T to be higher than some other threshold for the second object. Tag electron candidates must match a *Tight* trigger electron object within $\Delta R < 0.07$ and satisfy the offline *Tight*

identification selection. The probe electron candidates are required to satisfy track quality criteria and to be matched to an electromagnetic trigger object within $\Delta R < 0.07$ and have E_T at least 1 GeV higher than the corresponding trigger threshold. It is further required that the tag and the probe candidates are separated by $\Delta R_{\text{tag-probe}} > 0.15$ to prevent one object from affecting the identification of the other. On top of this baseline selection, isolation criteria are applied both on the tag and the probe for most of the variations.

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 τ -cut method. At this stage no requirement is made on the charge of the electrons and all possible tag-probe pairs are considered.

8.2.2. Background Estimation

The invariant mass of the tag-probe pair is used to discriminate signal electrons against background. 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 $1.8 \text{ GeV} < m_{ee} < 4.6 \text{ GeV}$. The most important contribution to the background comes from random combinations of two particles which do not originate from a resonance decay. This can be estimated – assuming charge symmetry – using the mass spectrum of same-sign (SS) pairs. The mass distribution of the SS events can be modelled with a Chebychev polynomial and included as an extra component in the opposite-sign (OS) fit (see Fig. 5 for the τ -cut method). Alternatively, the SS events can be subtracted explicitly from the OS events in order to remove the estimated combinatoric background component from both the OS mass and lifetime distribution. This is the procedure followed in the τ -fit method. The remaining OS background is modelled by a Chebychev polynomial. To model the signal J/ψ component, a Crystal-Ball function [21, 22] is used. The $\psi(2S)$ is modelled with the same shape except for an offset corresponding to the mass difference between the J/ψ and $\psi(2S)$ states. The ratio of J/ψ over $\psi(2S)$ is constrained around the measured value [23]. The number of J/ψ candidates is finally counted within a mass window of $2.8 \text{ GeV} < m_{ee} < 3.3 \text{ GeV}$.

In order to reduce the non-prompt J/ψ component, in the τ -cut method strict requirements are applied on τ , constraining it 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 $\sqrt{s} = 7 \text{ TeV}$ [20].

In the τ -fit method, the prompt component is extracted by fitting the pseudo-proper time distribution in the range $-1 < \tau < 3 \text{ ps}$, after subtracting the contribution from the estimated OS background using the τ distribution in the mass sidebands $2.3 \text{ GeV} < m_{ee} < 2.5 \text{ GeV}$ and $4.0 \text{ GeV} < m_{ee} < 4.2 \text{ GeV}$. The non-prompt component is modelled by an exponential decay function convolved 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. 6.

Systematic uncertainties are estimated by varying the tag-and-probe selection (such as the isolation), the fitting procedure (background and signal shapes, fit window, and sideband definitions) and the mass window for signal counting after the mass fit. The main systematic effects are found to be related to the isolation cut on the probe, the choice of the background function, as well as the choice of the Crystal-Ball tail parameters (either fixed to values obtained from simulation or allowed to vary).

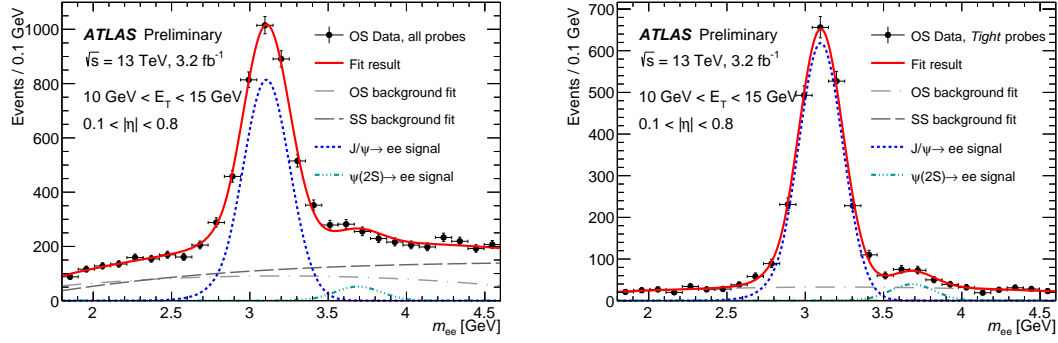


Figure 5: The figure demonstrates the procedure of background subtraction in the τ -cut method. Shown is the dielectron invariant mass fit for all electron probes at reconstruction level (left) and for probes passing the *Tight* identification (right) for $10 \text{ GeV} < E_T < 15 \text{ GeV}$ and $0.1 < |\eta| < 0.8$. 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 modelled by a Crystal-Ball function). A background fit is carried out using the sum of the same-sign (SS) distribution fit (dashed grey) from data and a Chebychev polynomial of 2nd order describing the residual background (dashed grey). The pseudo-proper time is required to be $-1 \text{ ps} < \tau < 0.2 \text{ ps}$.

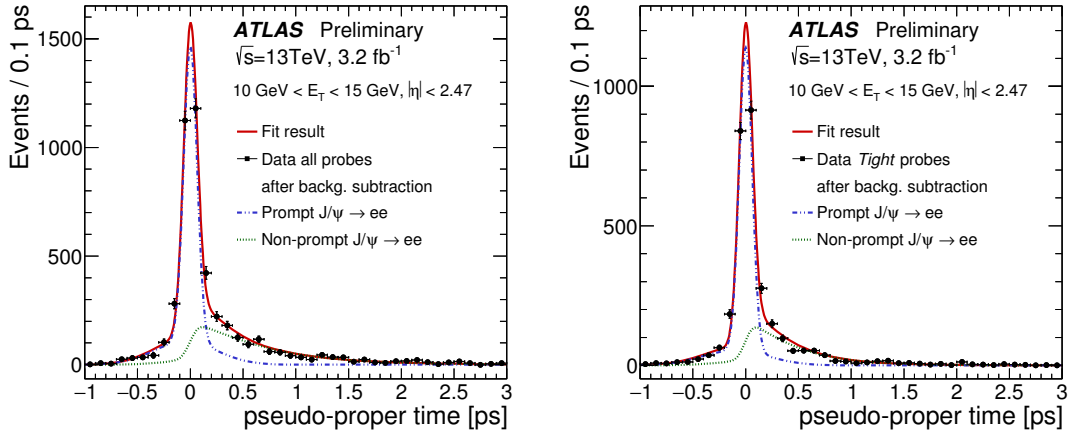


Figure 6: Pseudo-proper time fit for all probes passing reconstruction and track quality criteria (left) and for probes passing the *Tight* identification criteria (right) for $10 \text{ GeV} < E_T < 15 \text{ GeV}$, integrated over $|\eta| < 2.47$. Dots with error bars represent the OS minus SS data with the residual background subtracted using the reconstructed dielectron 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 dashed green line (exponential decay function convolved with the sum of two Gaussians). A track isolation cut of $p_T^{\text{varcone0.2}}/E_T < 0.15$ is applied on both the tag and the probe.

8.3. Combined Results

The measurement of identification efficiencies is performed with two methods for the evaluation of the background in $Z \rightarrow ee$ events, namely the Z_{mass} and Z_{iso} , and two methods for the evaluation of the fraction of prompt $J/\psi \rightarrow ee$ decays, namely the τ -cut and τ -fit methods. The Z_{iso} method is performed using a subset of the data of the Z_{mass} method, the τ -cut method uses a subset of the data of the τ -fit method. The individual measurements are therefore statistically correlated. Moreover, the results are also systematically correlated since part of the systematic variations are shared between methods, e.g. both the

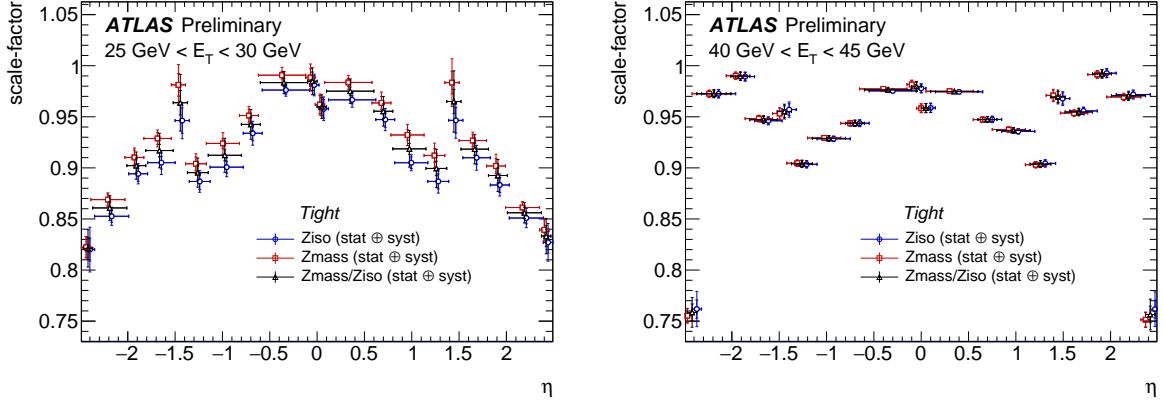


Figure 7: Data-to-MC ratio measured in $Z \rightarrow ee$ events with the Z_{mass} and Z_{iso} methods and the combined result. The data-to-MC ratios are shown as a function of pseudorapidity η for probe transverse energies in the ranges $25 \text{ GeV} < E_T < 30 \text{ GeV}$ (left) and $40 \text{ GeV} < E_T < 45 \text{ GeV}$ (right). Shown are numbers for the *Tight* identification criterion. The inner error bars show the statistical uncertainty, the outer error bars show the combined statistical and systematic uncertainty.

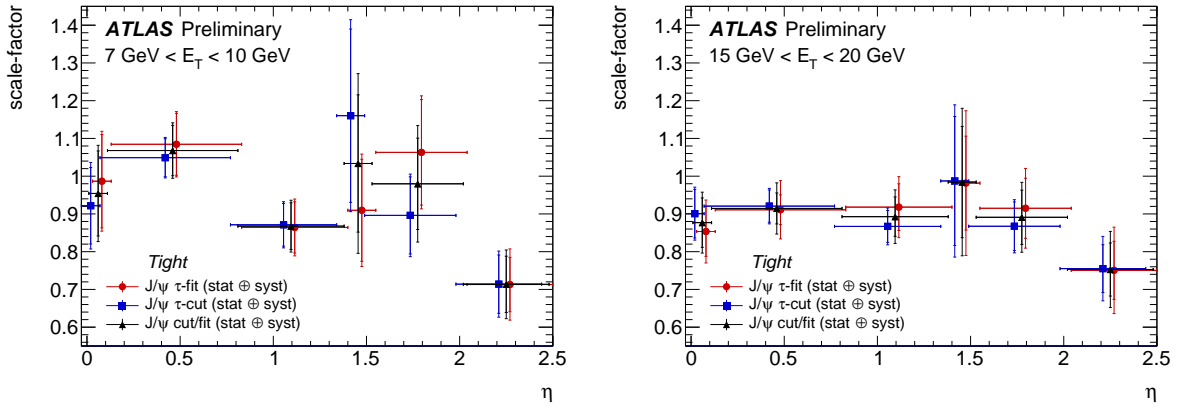


Figure 8: Data-to-MC ratio measured in $J/\psi \rightarrow ee$ events with the τ -fit and τ -cut methods and the combined result. The data-to-MC ratios are shown as a function of pseudorapidity η for probe transverse energy in the ranges $7 \text{ GeV} < E_T < 10 \text{ GeV}$ (left) and $15 \text{ GeV} < E_T < 20 \text{ GeV}$ (right). Shown are numbers for the *Tight* identification criterion. The inner error bars show the statistical uncertainty, the outer error bars show the combined statistical and systematic uncertainty.

Z_{mass} and Z_{iso} methods use a variation of the invariant mass window to select a Z candidate event.

Therefore, when combining the individual results from the Z_{mass} and the Z_{iso} methods, and from the τ -fit and the τ -cut methods, they are considered as systematic variations of one another.

The central value of the data-to-MC ratios is calculated from all systematic variations used in the individual methods, i.e. the combined results from $Z \rightarrow ee$ decays is calculated from all Z_{mass} and Z_{iso} variations, the combined results from $J/\psi \rightarrow ee$ decays is calculated from all τ -cut and τ -fit variations. If the individual methods have different numbers of variations, the contributions to the average are weighted accordingly. Fig. 7 and Fig. 8 show a comparison of the combined results and the ones from the individual Z tag-and-probe and J/ψ tag-and-probe measurements, respectively, as a function of pseudorapidity η .

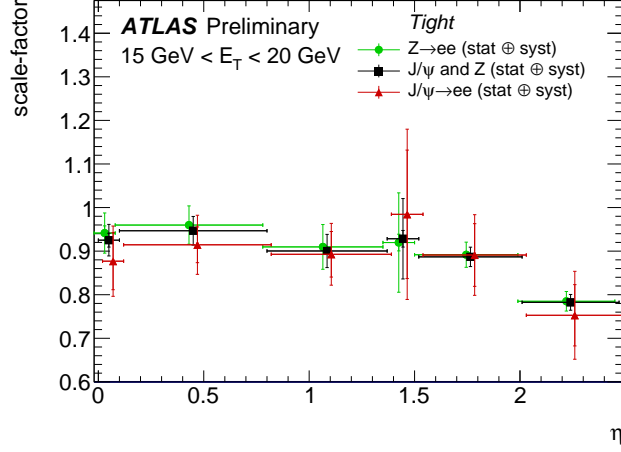


Figure 9: Data-to-MC ratio measured in $J/\psi \rightarrow ee$ and $Z \rightarrow ee$ events and the combined result for the *Tight* identification criterion. The ratios are shown as a function of pseudorapidity η for probe transverse energy of $15 \text{ GeV} < E_T < 20 \text{ GeV}$. The inner error bars show the statistical uncertainty, the outer error bars show the combined statistical and systematic uncertainty.

The results are presented for the *Tight* identification criterion in two selected E_T bins.

The methods employed for $J/\psi \rightarrow ee$ decays are found to be in good agreement. For $Z \rightarrow ee$ decays the uncertainty of the combined result in some cases does not cover the envelope of the individual results. This is most visible for probes with a transverse energy in the range $15 \text{ GeV} < E_T < 20 \text{ GeV}$. To ensure that the systematic uncertainty accounts for differences between results from the two methods the uncertainty is inflated such that the RMS covers at least 68% of all the variations, similar to the prescription used in Ref. [24]. It was found that scaling the uncertainties of the measurements by 10% covers observed differences between the methods.

The measurements in $Z \rightarrow ee$ and $J/\psi \rightarrow ee$ decays overlap for $15 \text{ GeV} < E_T < 20 \text{ GeV}$. A χ^2 -fit is used to combine the measurements, where the systematic uncertainties are included as nuisance parameters, constrained by a penalty term as explained in Ref. [25]. A comparison of results from $J/\psi \rightarrow ee$ and $Z \rightarrow ee$ decays is shown in Fig. 9 for $15 \text{ GeV} < E_T < 20 \text{ GeV}$. The precision of the combined result is improved compared to the individual results.

The $Z \rightarrow ee$ tag-and-probe efficiency cannot be precisely measured for $E_T > 200 \text{ GeV}$ due to a lack of statistics. The data-driven shower shape corrections used in the LH optimisation and described in Sect. 4, and other corrections⁶ can be applied to high E_T electrons from MC events prior to the efficiency measurement. Using the ratio of efficiencies for the nominal (uncorrected) MC to that of the corrected MC, an additional systematic uncertainty is conservatively assigned using the largest deviation from 1. For *Tight*, shown in Fig. 10, this corresponds to an additional uncertainty of 2.5% for electrons with $E_T > 125 \text{ GeV}$, while *Medium* and *Loose* are assigned an additional systematic uncertainty of 2%.

The efficiencies measured in data and in MC simulation are shown in Fig. 11 as a function of the number of measured primary vertices. The slow dependence of the efficiencies on the number of primary vertices as

⁶ In addition to differences in electron shower shapes between data and simulation, there are known distortions to the calorimeter affecting the $\Delta\eta_1$ variable in data, which were accounted for in this study by injecting distortions of a similar magnitude to the MC.

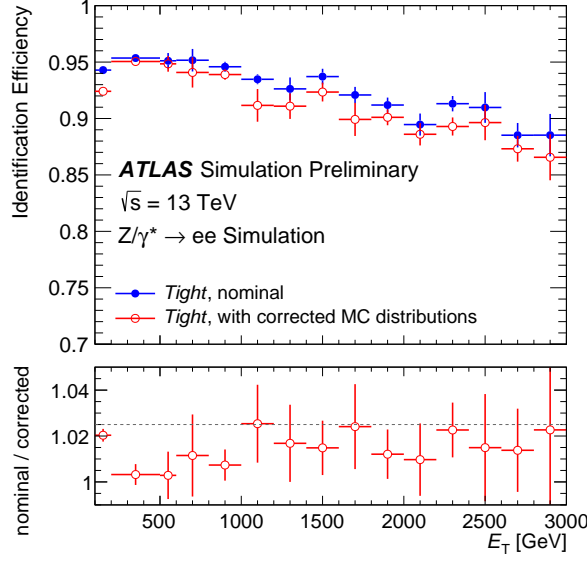


Figure 10: Identification efficiencies for high E_T electrons from MC Drell-Yan ($Z/\gamma^* \rightarrow ee$) events, using the nominal, uncorrected MC, and MC which has had corrections to the shower shapes and the $\Delta\eta_1$ variable applied to it. The ratio was used to determine the additional systematic uncertainty assigned for high E_T electrons.

well as the flatness of the data to Monte Carlo ratio is the result of the optimisation discussed in Sect. 4.

9. Reconstruction Efficiency Measurement

The reconstruction efficiency is defined as the ratio of the number of electrons reconstructed as described in Sect. 3 with a matching track passing the track quality requirements, to the total number of EM clusters from electrons with or without a matching track. The efficiency to detect an energy cluster with the sliding window algorithm in the EM calorimeter is found to be greater than 99% for $E_T > 15$ GeV [3]. Therefore, EM clusters are the starting point of the reconstruction efficiency measurement.

A tag-and-probe method similar to the Z_{mass} method, presented in Sect. 8, is employed for estimating the reconstruction efficiency. The tag passes strict requirements whereas the probe definition is relaxed to include all EM clusters. Events are required to have at least one tag electron and one EM cluster, serving as probe as described in Sect. 8. No requirement is made on the charge of the probe, unlike for the measurement of the identification efficiency. The measurement is only performed for probe electrons with $E_T > 15$ GeV since the background contamination becomes too large at lower E_T .

The reconstruction efficiency $\varepsilon_{\text{reco}}$ is defined as:

$$\varepsilon_{\text{reco}} = \frac{N^{\text{Quality-Track}} - B^{\text{Quality-Track}}}{(N^{\text{Quality-Track}} - B^{\text{Quality-Track}}) + (N^{\text{No-Quality}} - B^{\text{No-Quality}}) + (N^{\text{No-Track}} - B^{\text{No-Track}})}, \quad (4)$$

where N (B) represents the number of reconstructed (background) probes, “Quality-Track” represents probes associated with a good quality track having at least 7 precision hits and 1 pixel hit, “No-Quality” refers to probes with a track with silicon hits but no pixel or less than 7 precision hits, and “No-Track”

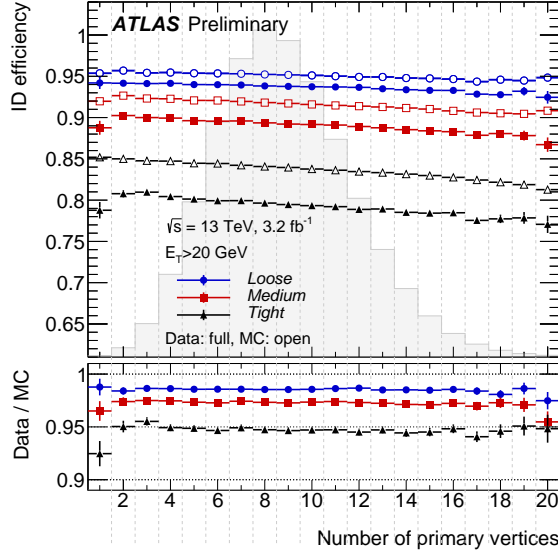


Figure 11: Efficiencies for the different likelihood operating points obtained using the Z tag-and-probe method, as a function of the number of reconstructed primary vertices, which is related to the number of collisions in each event (pileup). The total statistical and systematic uncertainties are displayed. The distribution of the number of reconstructed primary vertices in the selected data events is overlaid in grey.

refers to probes with no associated track. Splitting the denominator into 3 terms allows for an optimised background determination for each case.

The $B^{\text{Quality-Track}}$ and $B^{\text{No-Quality}}$ in Eq. 4 are estimated in a similar way to the one used for the identification efficiency measurement described in Sect. 8.1.2. The $B^{\text{No-Track}}$ is modelled by a third order polynomial. The fit is performed in an invariant mass region excluding the Z-peak, resulting in sidebands of low ($m_{ee} < 80$ GeV) and high masses ($m_{ee} > 100$ GeV). The residual signal contamination in the sideband regions is subtracted using MC.

The systematic uncertainties are estimated similarly to the identification measurements. In addition to similar variations as the ones described in Sect. 8.1.2, four sideband variations for the fit of the *No-Track* EM clusters are performed: $[70, 80]$ GeV \cup $[100, 110]$ GeV, $[60, 80]$ GeV \cup $[100, 120]$ GeV, $[50, 80]$ GeV \cup $[100, 130]$ GeV and $[55, 70]$ GeV \cup $[110, 125]$ GeV. The MC prediction enters only in the subtraction of the small residual signal in the sidebands used for the fit. The uncertainty on the MC description of the inefficiency to associate an electron cluster with a track is 20%- 30% [3]. The resulting uncertainty on the measured reconstruction efficiency is negligible.

Fig. 12 shows the invariant mass distribution of the tag-probe pairs, computed with the cluster energies and the direction from the associated tracks (except for probes with no such track, for which the cluster position is used instead). Two selected bins at the cluster level (denominator) are shown; the background components are also depicted.

Fig. 13 shows the efficiency to reconstruct an electron with a good track quality in data and MC as a function of E_T and η . The efficiency to reconstruct an electron associated with a good quality track varies from 97% to 99% between the endcap and barrel regions for electrons with $E_T > 15$ GeV. Lower efficiencies (95%) are obtained in the calorimeter transition region. For very energetic electrons ($E_T > 80$ GeV) the efficiency is $\sim 99\%$ over the whole η range. The good agreement between the reconstruction

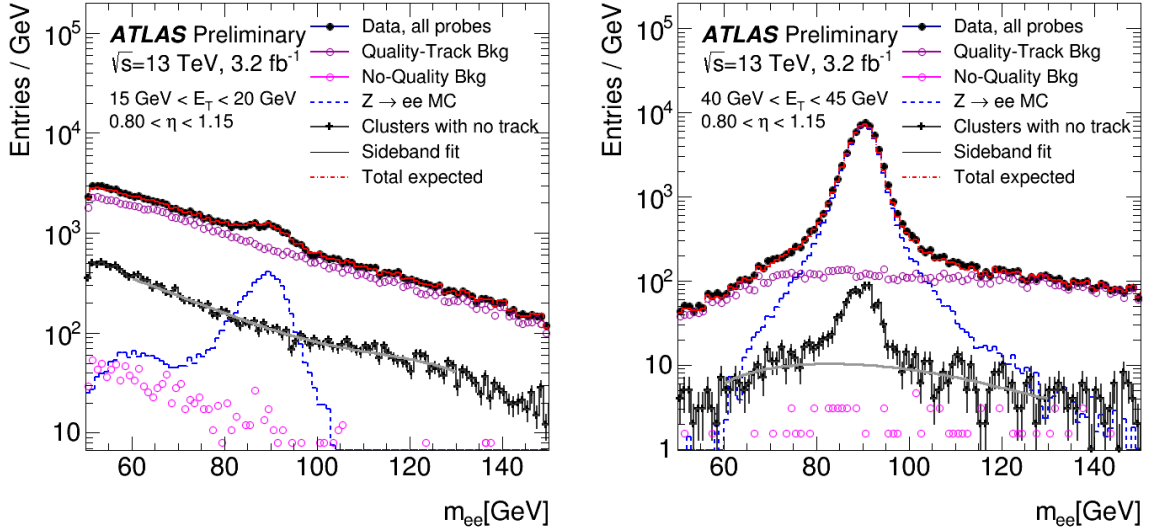


Figure 12: Invariant mass distributions of the tag-probe pairs for probes with $0.80 < \eta < 1.15$ and $15 \text{ GeV} < E_T < 20 \text{ GeV}$ (left) or $40 < E_T < 45 \text{ GeV}$ (right), before requiring the reconstruction criteria. The data (black dots) at the *all probes* level is composed of three components: clusters with no matching track (dark grey histogram with error bars) and clusters with a matching track either satisfying or not satisfying track quality criteria. The background is evaluated separately for these three components. A 3rd order polynomial (grey dashed line spanning from 60 to 130 GeV) depicts the estimated background for clusters with no associated track from a fit performed in the sideband regions. Background templates normalised in this case to the high mass tail (purple and magenta markers) are used to estimate the background with a matching track (with/without track quality). These background templates are obtained by requiring some identification criteria to be failed. The sum of the normalised templates and the MC signal prediction (red line, only shown for comparison) agrees well with the data points.

results obtained in data and in MC gives confidence in the MC description of the detector response 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.

The reconstruction efficiency is comparable to the efficiency observed in data taken in 2012 at 8 TeV, except for the calorimeter transition region ($1.37 < |\eta| < 1.52$) where the efficiency degrades by 2%. The drop in efficiency is caused by inefficiencies in the matching of the track to the calorimeter section (barrel/endcap).

The combined reconstruction and identification efficiencies are shown in Fig. 14. In order to be able to compare the different efficiency measurements, 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. The uncertainties associated to the measurements, also shown in Fig. 15, are obtained with pseudo-experiments, treating the statistical uncertainties from the different (E_T, η) bins as uncorrelated.

10. Isolation Efficiency Measurement

The isolation efficiency measurement uses the tag-and-probe method with $Z \rightarrow ee$ decays as described in Sect. 8. Instead of using $E_T > 15 \text{ GeV}$, the probe electron candidates are obtained with a lower

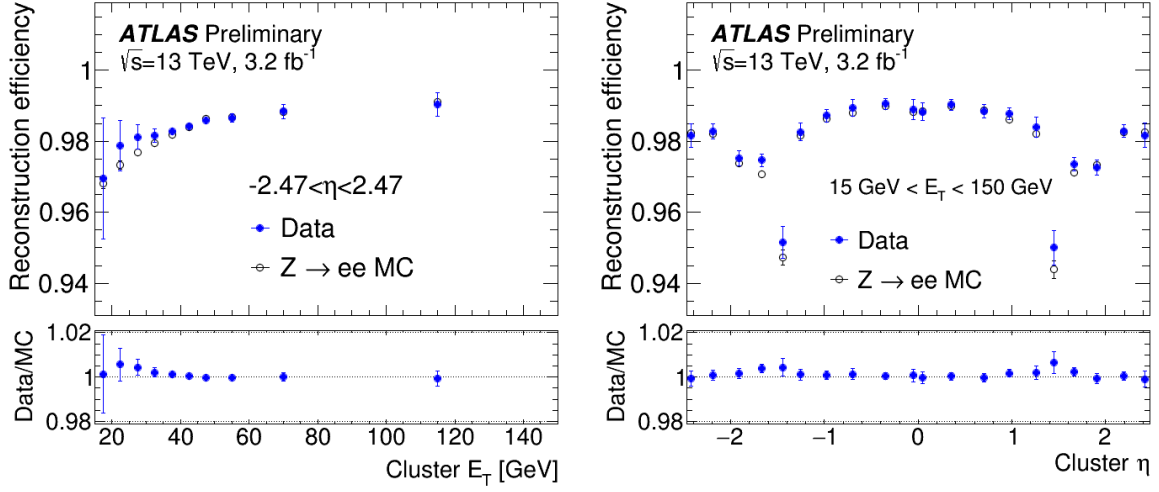


Figure 13: Measured reconstruction efficiencies as a function of E_T integrated over the full pseudorapidity range (left) and as a function of η for $15 \text{ GeV} < E_T < 150 \text{ GeV}$ (right) for the 2015 dataset. The shown uncertainties are statistical plus systematic.

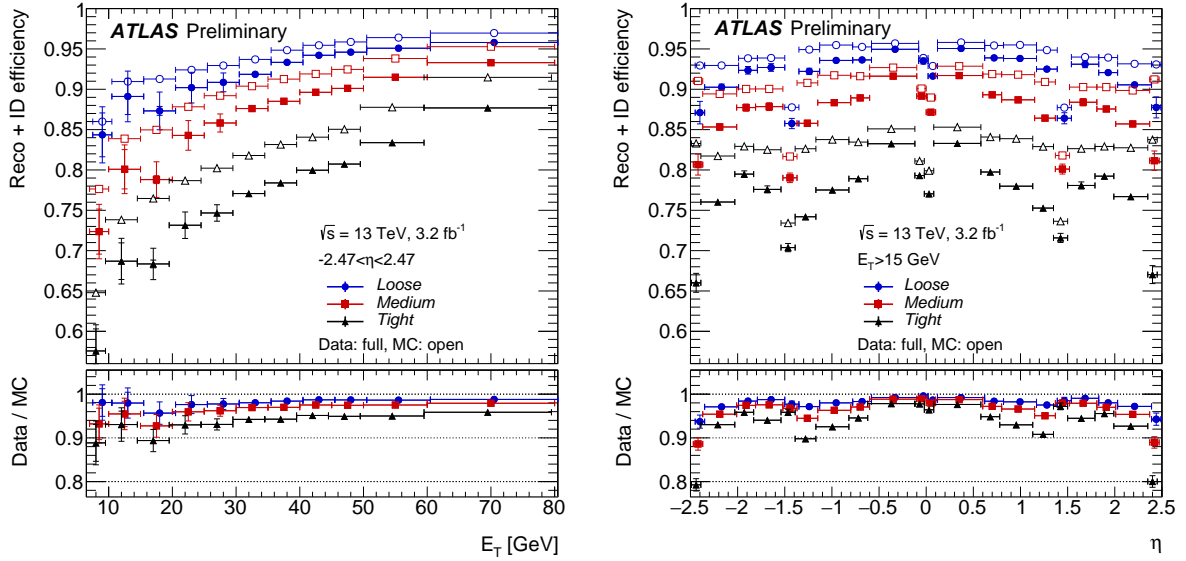


Figure 14: Combined electron reconstruction and identification efficiencies in $Z \rightarrow ee$ events as a function of the transverse energy E_T , integrated over the full pseudorapidity range (left), and as a function of pseudorapidity η , integrated over the full E_T range (right). The data efficiencies are obtained from the data-to-MC efficiency ratios measured using J/ψ and Z tag-and-probe, multiplied by the MC prediction for electrons from $Z \rightarrow ee$ decays. The uncertainties are obtained with pseudo-experiments, treating the statistical uncertainties from the different (E_T , η) bins as uncorrelated. Two sets of uncertainties are shown: the inner error bars show the statistical uncertainty, the outer error bars show the combined statistical and systematic uncertainty.

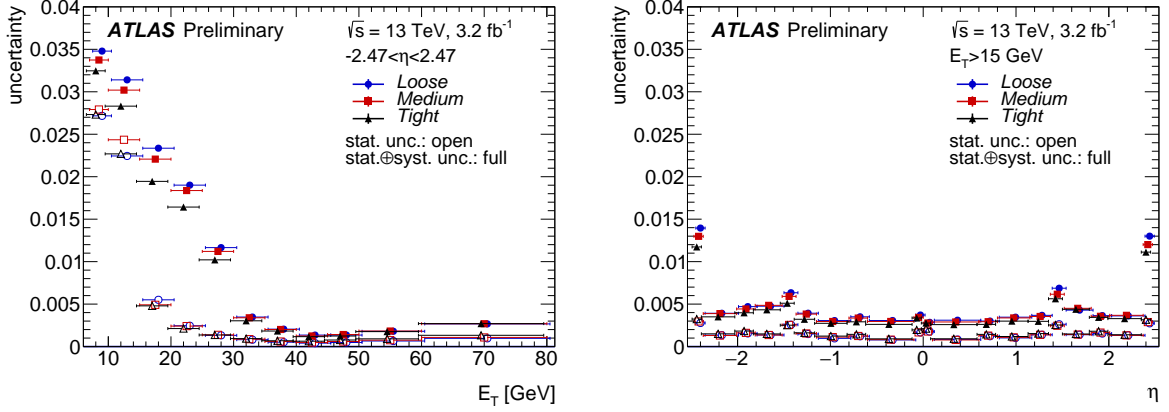


Figure 15: Absolute uncertainties on the combined electron reconstruction and identification efficiencies in $Z \rightarrow ee$ events as a function of the transverse energy E_T , integrated over the full pseudorapidity range (left), and as a function of pseudorapidity η , integrated over the full transverse energy range. The data efficiencies are obtained from the data-to-MC efficiency ratios measured using J/ψ and Z tag-and-probe, multiplied by the MC prediction for electrons from $Z \rightarrow ee$ decays. The uncertainties are obtained with pseudo-experiments, treating the statistical uncertainties from the different (E_T, η) bins as uncorrelated. Two sets of uncertainties are shown: the statistical and the combined statistical and systematic uncertainty.

threshold, $E_T > 7$ GeV. After subtracting the background contamination using the Z_{mass} method, the isolation efficiency is calculated by taking the ratio between the number of probe electrons that pass the identification and isolation criteria and the number of probe electrons passing the identification requirement only. Efficiencies are computed for all the isolation operating points listed in Table 2, with respect to the three likelihood identification operating points, *Loose*, *Medium* and *Tight*.

Since the efficiencies depend on the transverse energy and pseudorapidity, the measurements are performed in two-dimensional bins in (E_T, η) . Fig. 16 shows the electron isolation efficiency and data-to-MC ratios as a function of the transverse energy E_T and pseudorapidity η , respectively. Good agreement between data and MC is observed for $E_T > 20$ GeV, while larger discrepancies are observed for $E_T < 20$ GeV. Good data-to-MC agreement is also observed as a function of the electron η with slightly larger discrepancies at the level of 1% in the regions $|\eta| \approx 1.5$. The same systematic uncertainties presented for the Z_{mass} method are also used in this case, and the most relevant source is found to come from the variation of the background template, particularly at low E_T .

11. Trigger Efficiency Measurement

The trigger efficiencies are determined in $Z \rightarrow ee$ events using the same tag-and-probe method as described in Sect. 8. Background subtraction is performed similarly to the Z_{mass} method. The trigger efficiency is defined as the ratio of the number of probe electrons that match to the requested HLT electron within $\Delta R < 0.07$ to the total number of probe electrons. In addition to the tag-and-probe event selection, all offline probe electrons are required to pass an isolation and identification criterion. The efficiencies for each trigger operating point are determined for each combination of the three identification operating points *Loose*, *Medium*, *Tight* and for all isolation operating points as listed in Sect. 5.

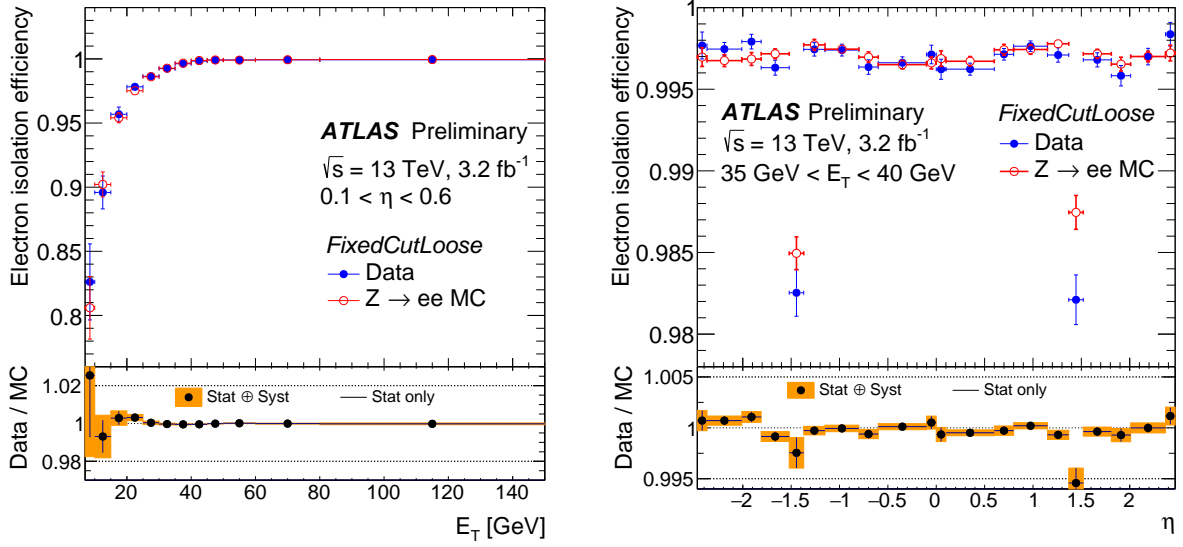


Figure 16: Efficiency of the FixedCutLoose isolation requirement ($E_T^{\text{cone0.2}}/E_T < 0.2$ and $p_T^{\text{varcone0.2}}/E_T < 0.15$) for electrons from $Z \rightarrow ee$ as a function of the transverse energy E_T , for $0.1 < \eta < 0.6$ (left) and a function of pseudorapidity η , for $35 \text{ GeV} < E_T < 40 \text{ GeV}$ (right). The electrons are required to fulfill *Tight* identification.

Fig. 17 shows the efficiencies measured in data and $Z \rightarrow ee$ MC for the main single electron trigger with respect to the *Medium* offline operating point and no isolation applied. In the turn-on region a higher efficiency is measured for MC than for data. Good agreement can be seen for electrons with transverse energy far above the threshold. The large discrepancy seen very close to the threshold originates from different thresholds in the L1 seeds of data and simulation used for the comparison. In the endcap region ($|\eta| > 1.52$) the efficiency measured in data generally is higher than the one measured in MC.

The efficiencies and scale-factors are measured as a function of E_T and η . For η , the same binning as for the ID and isolation efficiency measurements is used. The E_T binning is modified to have one bin starting at 1 GeV above the HLT threshold value for each supported trigger operating point.

12. Conclusions

The 2015 LHC pp collision data collected by ATLAS at 13 TeV centre-of-mass energy, corresponding to an integrated luminosity of 3.2 fb^{-1} , has been used to measure the efficiencies for the electron identification and reconstruction as well as the efficiencies for trigger and isolation selection. The multiplication of these efficiencies yields the total efficiency to measure electrons in the ATLAS data analyses.

The algorithms used to measure the electrons have been adapted to the new experimental conditions. In particular, the detector modifications, including the insertion of the new innermost silicon detector (IBL) and the change in the TRT gas have been taken into account. The identification algorithms based on a multivariate likelihood discriminator have been optimised such that only a mild dependence on the pileup conditions remains. This dependence is well reproduced by the Monte Carlo simulation.

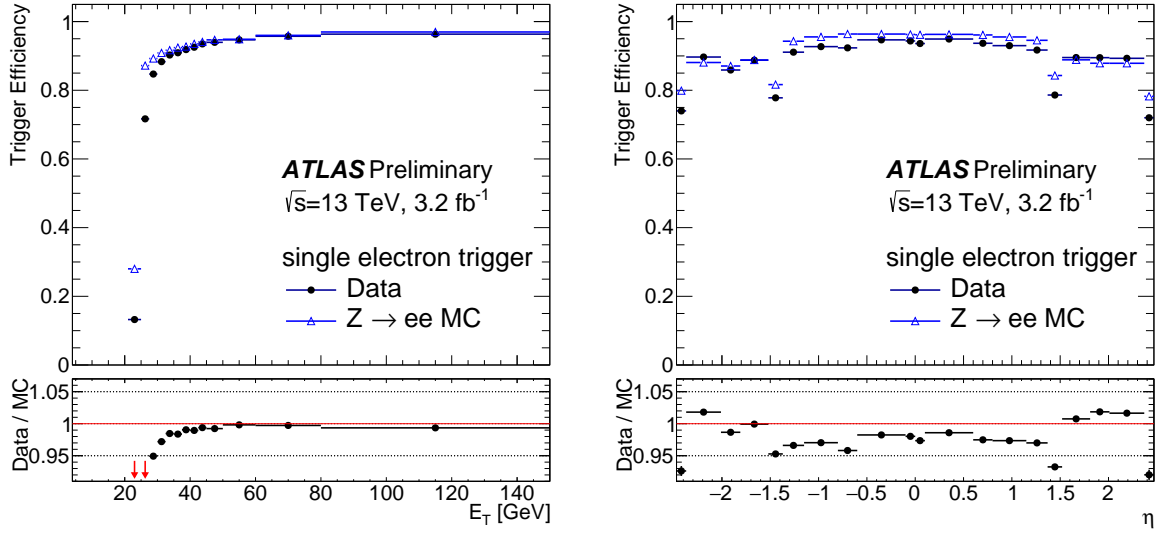


Figure 17: Trigger efficiency for the main single electron trigger measured in $Z \rightarrow ee$ events with the tag-and-probe method as a function of transverse energy E_T (left) and pseudorapidity η (right). The offline electron is required to pass the *Medium* identification. The efficiencies for each E_T bin are integrated over η and the efficiencies in each η bin are shown for electrons with $E_T > 25$ GeV. The bottom pad shows the data-to-MC ratio.

The various efficiency measurements are based on the tag-and-probe method using Z and J/ψ events, and are similar to the ones used in Run-1 [3]. The differences between data and MC simulation have been incorporated into data-to-MC ratios (scale-factors), obtained by combining the Z and J/ψ measurements, as a function of the transverse energy and the pseudorapidity of the electrons. The precision of the scale-factors reaches a few percent at low E_T and is below 1% at high E_T . The scale-factors obtained by the analyses presented in this note are used in all ATLAS analyses based on the data sample collected in 2015.

References

- [1] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, [*JINST* **3** \(2008\) S08003](#).
- [2] ATLAS Collaboration, *Electron reconstruction and identification efficiency measurements with the ATLAS detector using the 2011 LHC proton-proton collision data*, [*Eur. Phys. J.* **C74** \(2014\) 2941](#), arXiv: [1404.2240 \[hep-ex\]](#).
- [3] ATLAS Collaboration, *Electron efficiency measurements with the ATLAS detector using the 2012 LHC proton-proton collision data*, ATLAS-CONF-2014-032, 2014, URL: <http://cdsweb.cern.ch/record/1706245>.
- [4] ATLAS Collaboration, *ATLAS Insertable B-Layer Technical Design Report*, CERN-LHCC-2010-013, 2010, URL: <https://cds.cern.ch/record/1291633>.
- [5] W. Lampl et al., *Calorimeter Clustering Algorithms: Description and Performance*, ATL-LARG-PUB-2008-002, 2008.
- [6] T. Cornelissen et al., *The global χ^2 track fitter in ATLAS*, [*J. Phys. Conf. Ser.* **119** \(2008\) 032013](#).
- [7] ATLAS Collaboration, *Improved electron reconstruction in ATLAS using the Gaussian Sum Filter-based model for bremsstrahlung*, ATLAS-CONF-2012-047, 2012, URL: <http://cdsweb.cern.ch/record/1449796>.
- [8] ATLAS Collaboration, *Electron and photon energy calibration with the ATLAS detector using LHC Run 1 data*, [*Eur. Phys. J. C* **74** \(2014\) 3071](#), arXiv: [1407.5063 \[hep-ex\]](#).
- [9] S. Agostinelli et al., *GEANT4: A Simulation toolkit*, [*Nucl. Instrum. Meth.* **A506** \(2003\) 250](#).
- [10] ATLAS Collaboration, *Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1*, (2016), arXiv: [1603.02934 \[hep-ex\]](#).
- [11] M. Cacciari and G. P. Salam, *Pileup subtraction using jet areas*, [*Phys. Lett.* **B659** \(2008\) 119](#), arXiv: [0707.1378 \[hep-ph\]](#).
- [12] C. Blocker, *Uncertainties on Efficiencies*, [CDF/MEMO/STATISTICS/PUBLIC/7168](#), 2004.
- [13] P. Nason, *A New method for combining NLO QCD with shower Monte Carlo algorithms*, [*JHEP* **0411** \(2004\) 040](#), arXiv: [hep-ph/0409146 \[hep-ph\]](#).
- [14] 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 \[hep-ph\]](#).
- [15] S. Alioli et al., *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, [*JHEP* **1006** \(2010\) 043](#), arXiv: [1002.2581 \[hep-ph\]](#).
- [16] T. Sjostrand, S. Mrenna and P. Z. Skands, *PYTHIA 6.4 physics and manual*, [*JHEP* **05** \(2006\) 026](#), arXiv: [hep-ph/0603175](#).
- [17] T. Sjostrand, S. Mrenna and P. Z. Skands, *A brief introduction to PYTHIA 8.1*, [*Comput. Phys. Commun.* **178** \(2008\) 852](#), arXiv: [0710.3820 \[hep-ph\]](#).
- [18] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, [*Eur. Phys. J.* **C70** \(2010\) 823](#), arXiv: [1005.4568 \[physics.ins-det\]](#).

- [19] ATLAS Collaboration, *Summary of ATLAS Pythia 8 tunes*, ATL-PHYS-PUB-2012-003, 2012, URL: <https://cds.cern.ch/record/1474107>.
- [20] ATLAS Collaboration, *Measurement of the differential cross-sections of inclusive, prompt and non-prompt J/ψ production in pp collisions at $\sqrt{s} = 7$ TeV*, Nuclear Physics B **850** (2011) 387, arXiv: [1104.3038 \[hep-ex\]](#).
- [21] M. Oreglia, *A Study of the Reactions $\psi' \rightarrow \gamma\gamma\psi$* , SLAC-0236, UMI-81-08973 (1980).
- [22] J. Gaiser, *Charmonium Spectroscopy From Radiative Decays of the J/ψ and ψ'* , SLAC-0255, UMI-83-14449-MC (1982).
- [23] CMS Collaboration, *J/ψ and ψ_{2S} production in pp collisions at $\sqrt{s} = 7$ TeV*, JHEP **1202** (2012) 011, arXiv: [1111.1557 \[hep-ex\]](#).
- [24] K. A. Olive et al., *Review of Particle Physics*, Chin. Phys. **C38** (2014) 090001.
- [25] F.D. Aaron et al., H1 Collaboration, *Measurement of the Inclusive ep Scattering Cross Section at Low Q^2 and x at HERA*, Eur. Phys. J. **C63** (2009) 625, arXiv: [0904.0929 \[hep-ex\]](#).

A. Electron Variable Distributions

The distributions of the variables used as input to the electron LH, as measured using the tag-and-probe method for probes from $Z \rightarrow ee$ events with $E_T > 15$ GeV are shown in Fig. 18 and Fig. 19. The *Loose* ID was applied on these probes, to reduce contamination from background for the data distributions.

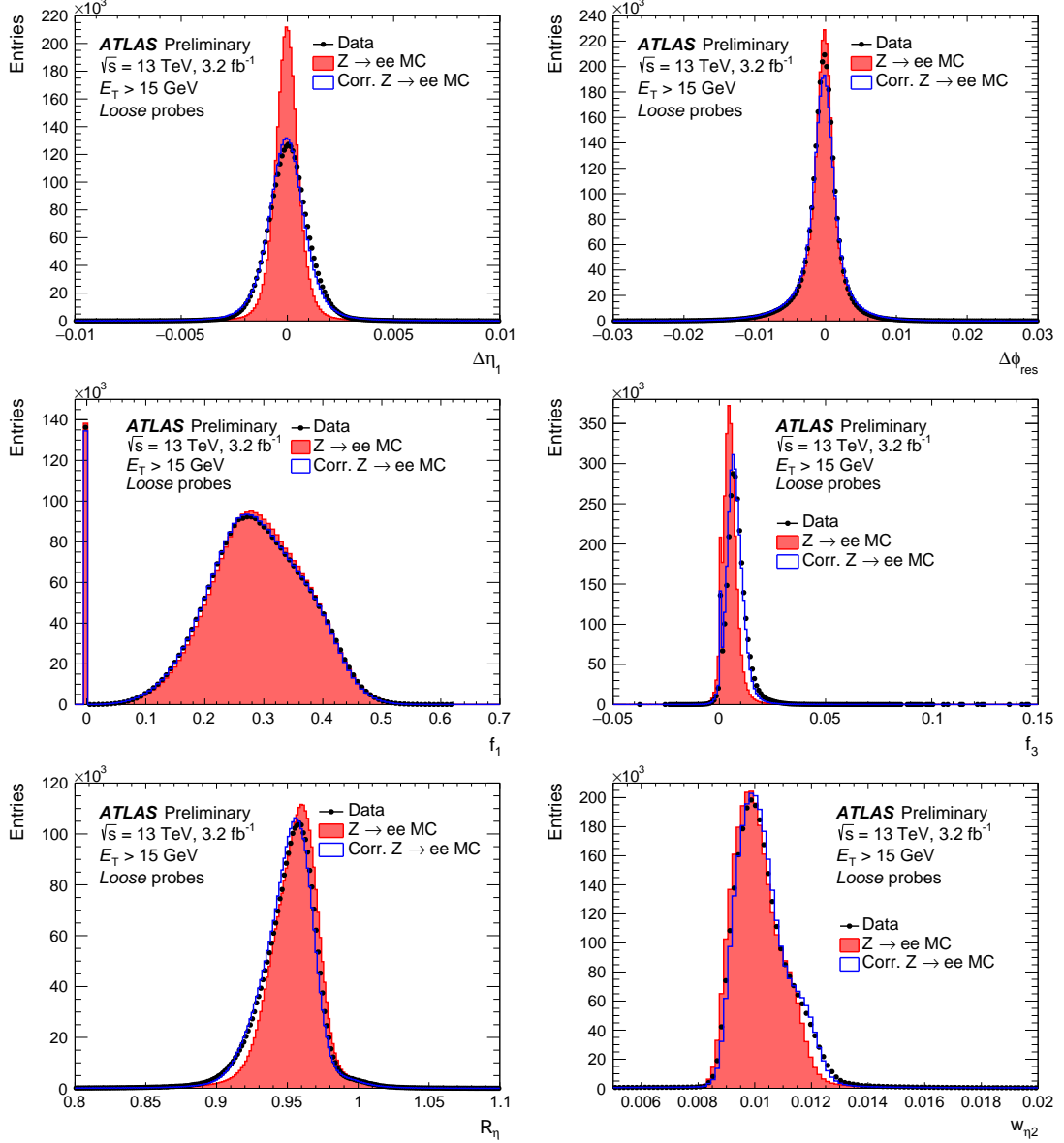


Figure 18: Distributions of the variables used in the electron ID, measured using $Z \rightarrow ee$ tag-and-probe. To reduce the effect of background, these are shown only for *Loose* probes. Corrected MC (denoted as “Corr. MC”) refers to the MC distributions with shifts or widths corrections applied, as was done for the PDFs used by the electron LH for these particular variables.

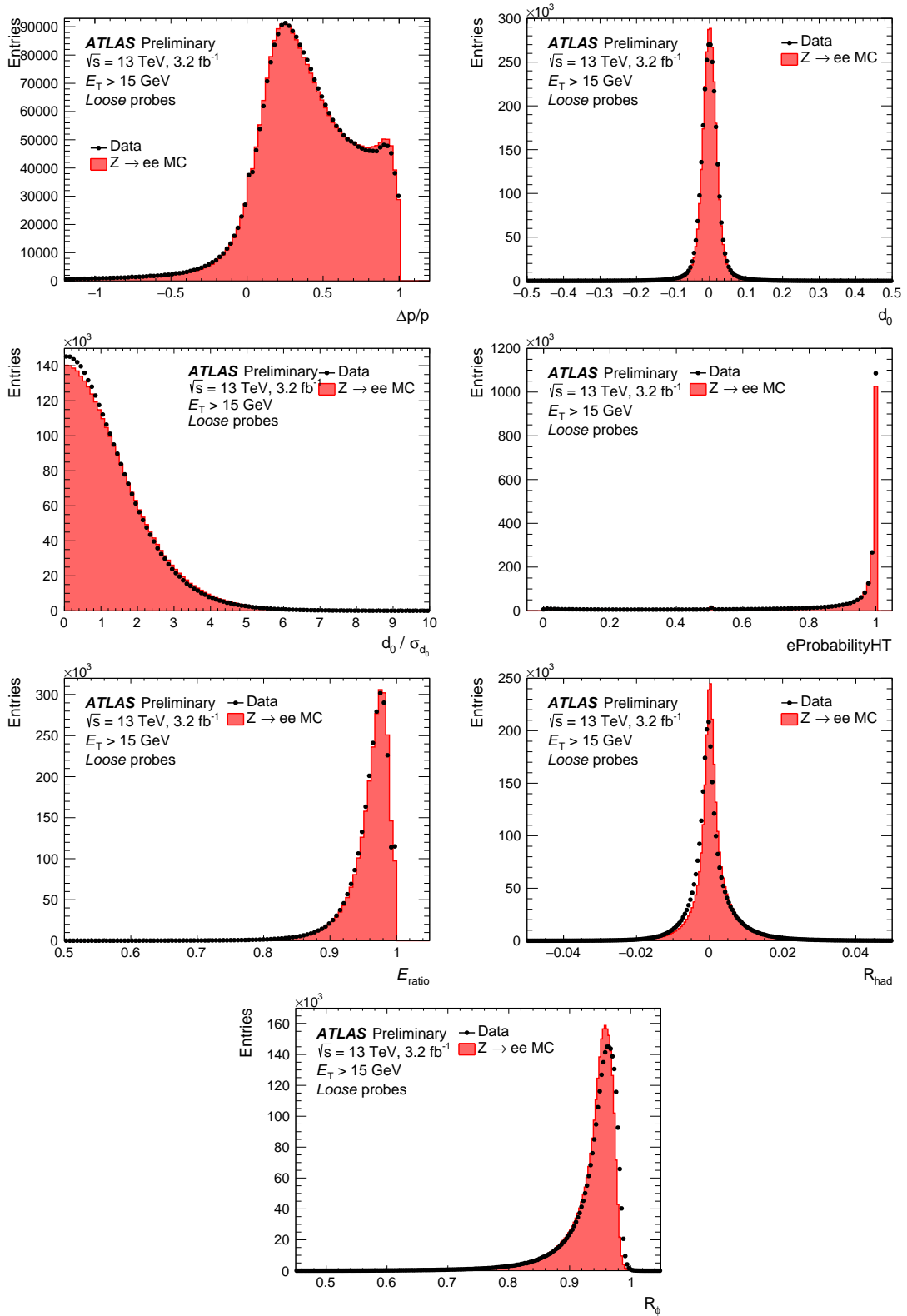


Figure 19: Distributions of the variables used in the electron ID, measured using $Z \rightarrow ee$ tag-and-probe. To reduce the effect of background, these are shown only for *Loose* probes. Here, the distributions are shown for the variables which did not have shifts or widths corrections applied during the optimisation of the electron LH PDFs. Note that $eProbabilityHT$ is the only variable currently being used in the electron ID that was not used in Run-1.

B. Electron Reconstruction and Identification Sketch

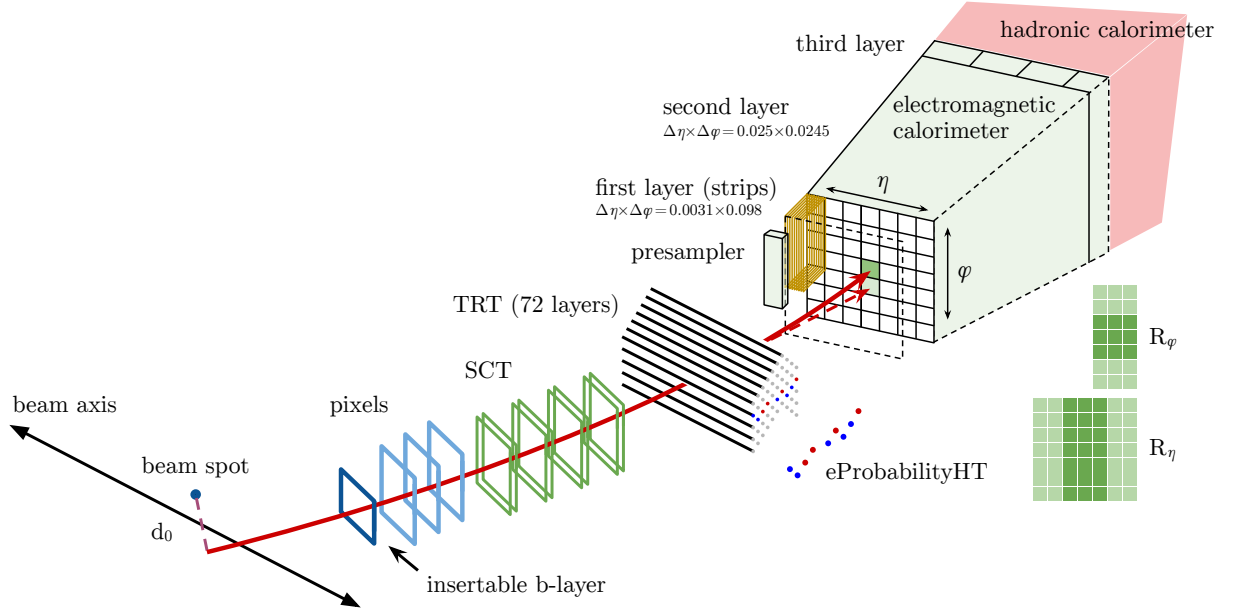


Figure 20: Schematic view of the electron reconstruction and identification (see Table 1).