Chapter 5

ESTIMATING THE RATES OF ELECTRON CHARGE MIS-IDENTIFICATION

Many physics analyses involve charged leptons in their final states, where leptons typically refer to electrons or muons. Not only are the kinematic quantities associated with these particles measured, their charges have to be determined as well, using the curvatures of the tracks which result from the inner detector magnetic field. As will be discussed below, the measured charges are not always correct, causing what is called charge mis-identification.

Charge mis-identification is important for analyses that involve same-sign leptons [1] in the final state, such as measurements of the same-sign WW scattering [48], Higgs production in association with a $t\bar{t}$ pair ($t\bar{t}H$), or SUSY search with two same-sign leptons [51]. In general, electron charge mis-identification rates occur on the order of O(1%), whereas Standard Model processes that provide opposite-sign dileptons, dominantly $Z \to e^+e^-$, occur approximately 10^3 times more commonly than genuine Standard Model sources of same-sign leptons (dominantly WZ production). Accordingly, opposite-sign sources of dileptons that suffer from charge mis-identification can constitute a significant background in these searches, and must therefore be estimated as precisely as possible.

This chapter describes a method for estimating the rate of charge mis-identification using a likelihood function. Section 5.1 discusses briefly how electron charge mis-identification might arise at ATLAS. Section 5.2 discusses the likelihood method, including the Poisson likelihood used as well as how it is applied to $Z \to e^+e^-$ events to measure the charge mis-identification rates. Finally, Section 5.3 provides some conclusions.

The data used was collected with the ATLAS detector in 2012, at 8 TeV center-of-mass energy and corresponds to an integrated luminosity of 20.3 fb^{-1}.

[1] Muon charge mis-identification is negligible at ATLAS [49]. Compared to electrons, muons are much less likely to undergo bremsstrahlung and pair-production in the detector. Moreover, muon tracks are measured in the inner detector as well as in the muon spectrometer, prodiving a larger lever arm for curvature measurements.

For all practical purpose, muon charge mis-identification...—> more careful because the charge flip rate of a 1 TeV muon is not negligible, but essentially no one looks at these, too rare

Chapter 5

ESTIMATING THE RATES OF ELECTRON CHARGE MIS-IDENTIFICATION

Many physics analyses involve charged leptons in their final states, where leptons typically refer to electrons or muons. Not only are the kinematic quantities associated with these particles measured, their charges have to be determined as well, using the curvatures of the tracks which result from the inner detector magnetic field. As will be discussed below, the measured charges are not always correct, causing what is called charge mis-identification.

Charge mis-identification is important for analyses that involve same-sign leptons [1] in the final state, such as measurements of the same-sign WW scattering [48], Higgs production in association with a $t\bar{t}$ pair ($t\bar{t}H$), or SUSY search with two same-sign leptons [51]. In general, electron charge mis-identification rates occur on the order of O(1%), whereas Standard Model processes that provide opposite-sign dileptons, dominantly $Z \to e^+e^-$, occur approximately 10^3 times more commonly than genuine Standard Model sources of same-sign leptons (dominantly WZ production). Accordingly, opposite-sign sources of dileptons that suffer from charge mis-identification can constitute a significant background in these searches, and must therefore be estimated as precisely as possible.

This chapter describes a method for estimating the rate of charge mis-identification using a likelihood function. Section 5.1 discusses briefly how electron charge mis-identification might arise at ATLAS. Section 5.2 discusses the likelihood method, including the Poisson likelihood used as well as how it is applied to $Z \to e^+e^-$ events to measure the charge mis-identification rates. Finally, Section 5.3 provides some conclusions.

The data used was collected with the ATLAS detector in 2012, at 8 TeV center-of-mass energy and corresponds to an integrated luminosity of 20.3 fb^{-1}.

[1] For all practical purposes, muon charge mis-identification is negligible at ATLAS [49]. Compared to electrons, muons are much less likely to undergo bremsstrahlung and pair-production in the detector. Moreover, muon tracks are measured in the inner detector as well as in the muon spectrometer, prodiving a larger lever arm for curvature measurements.

We assume that

Spellchecker please?

5.1 Electron Charge Mis-identification

At ATLAS, the sign of the charge of an electron is determined from the curvature of its track in the inner detector (Section 3.2.2.1). Charge mis-identification occurs mainly because of two reasons:

- The electron may radiate photons as it passes through the detector and interacts with the detector materials. These radiated photons may in turn convert to electron-positron pairs. A charge mis-identification occurs when the electron candidate is matched to the wrong track. This is the dominant source of charge mis-identification.
- The reconstructed track associated with the electron has a small curvature, which may happen at very high momentum or at large pseudorapidity, the latter case because of the limit of the lever arm of the tracker. Indeed, for $|\eta| \geq 2.0$, the track is oriented in the endcap region of the ATLAS detector and will not reach the full available lever arm of ~ 1.2 m transverse to the beam of the inner detector.

5.2 The Likelihood Method

We assume there is a probability associated to charge mis-identification and seek to determine this rate in a sample of electrons. At ATLAS, $Z \to e^+e^-$ events are used for this purpose because they are a dominant source of opposite-sign electrons as compared to other Standard Model sources. A very clean and high-statistics sample of electrons may be obtained by selecting two isolated electrons around the invariant Z mass peak. Due to charge mis-identification, not only are opposite-sign pair of electrons observed, same-sign pairs will be encountered as well, from which the charge mis-identification rates could be determined. More specically, the mis-identification rates to be extracted are parameters of a Poisson likelihood function that will be discussed below.

The rates obtained will be applied to an opposite-sign control sample in data, or to correct the MC simulation, to estimate the electron charge mis-identification background in a same-sign lepton analysis.

5.2.1 The $Z \to e^+e^-$ Sample

At ATLAS, electron charge mis-identification rates are extracted from $Z \to e^+e^-$ events using a likelihood function (Section 5.2.2). These events, which are also called tag-and-probe $Z \to e^+e^-$ events, are selected by applying the following selections.

Event selections

- A logical OR between two single-electron triggers, one with $E_T > 24$ GeV plus Medium identification, and one with $E_T > 60$ GeV plus Loose identification.

5.1 Electron Charge Mis-identification

At ATLAS, the sign of the charge of an electron is determined from the curvature of its track in the inner detector (Section 3.2.2.1). Charge mis-identification occurs mainly because of two reasons:

- The electron may radiate photons as it passes through the detector and interacts with the detector materials. These radiated photons may in turn convert to electron-positron pairs. A charge mis-identification occurs when the electron candidate is matched to the wrong track. This is the dominant source of charge mis-identification.
- The reconstructed track associated with the electron has a small curvature, which may happen at very high momentum or at large pseudorapidity, the latter case because of the limit of the lever arm of the tracker. Indeed, for $|\eta| \geq 2.0$, the track is oriented in the endcap region of the ATLAS detector and will not reach the full available lever arm of ~ 1.2 m transverse to the beam of the inner detector.

5.2 The Likelihood Method

We assume that there is a probability associated to charge mis-identification and seek to determine this rate in a sample of electrons. At ATLAS, $Z \to e^+e^-$ events are used for this purpose because they are a dominant source of opposite-sign electrons as compared to other Standard Model sources. A very clean and high-statistics sample of electrons may be obtained by selecting two isolated electrons around the invariant Z mass peak. Due to charge mis-identification, not only are opposite-sign pair of electrons observed, same-sign pairs will be encountered as well, from which the charge mis-identification rates could be determined. More specifically, the mis-identification rates to be extracted are parameters of a Poisson likelihood function that will be discussed below.

The rates obtained will be applied to an opposite-sign control sample in data, or to correct the MC simulation, to estimate the electron charge mis-identification background in a same-sign lepton analysis.

5.2.1 The $Z \to e^+e^-$ Sample

At ATLAS, electron charge mis-identification rates are extracted from $Z \to e^+e^-$ events using a likelihood function (Section 5.2.2). These events are required to undergo the following preliminary selections. Further selections that will be applied before the rates are extracted are discussed in Section 5.2.3.

Preliminary event selections

- A logical OR between two single-electron triggers, one with $E_T > 24$ GeV plus Medium identification, and one with $E_T > 60$ GeV plus Loose identification.

I added a note on the previous page too

IMPORTANT: I didn't pick that up until now but I thought that the two electrons were treated equally. It HAS to be the case as if not the charge flip rate would not be the same for the tag and the probe

preced by "truth-level"

- At least two reconstructed electron candidates with $|\eta| < 2.47$. One electron, called the tag candidate, is required to pass the Tight identification requirement; it must also have $E_T > 25$ GeV and $1.37 < |\eta| < 1.52$, and must be associated to a triggered electron within $\Delta R < 0.15$. The other electron, called the probe candidate, is required to have $E_T > 10$ GeV; moreover, the tracks associated with the electron must have at least one hit in the pixel detector and at least seven hits in the pixel and SCT detectors. The invariant mass of the tag-probe pair must be within ± 15 GeV of the Z mass.

Figure 5.1 [38] shows the invariant mass distribution m_{ee} in data and simulation for E_T between 25 GeV and 50 GeV and $0.0 < \eta < 0.8$ (left) or $2.0 < \eta < 2.47$ (right). Due to charge mis-identification same-sign electron pairs also exist in addition to opposite-sign pairs, indicating a charge mis-identification rate of $\sim 10^{-3}$ in the central region and $\sim 10\%$ in the high η region. The higher rates in the latter is expected because of the larger amount of material and the limited lever arm in the forward region. In both cases, same-sign pairs show a broader peak that is also slightly shifted towards lower values, consistent with the fact that the radiation that causes charge mis-identification also results in energy loss.

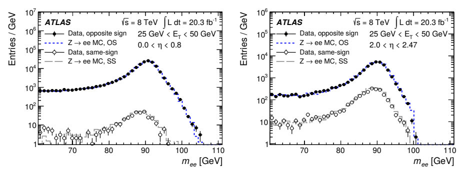

Figure 5.1: Distribution of the invariant mass m_{ee} for E_T between 25 and 50 GeV and $|\eta|$ between 0.0 and 0.8 [38]. Due to charge mis-identification same-sign pairs as well as opposite-sign pairs are observed.

The next section discusses the Poisson likelihood function that is used to fit the data.

5.2.2 The Poisson Likelihood

In a truth-level e^+e^- pair, which will also be called a truth-level opposite-sign pair, if the charge of any one of the electrons is mis-identified, then a same-sign pair will be observed instead[2]. Assuming a probability p that a truth-level opposite-pair will be identified as a same-sign pair, then in considering n truth-level pairs e^+e^-, the probability that exactly n_{ss} same-sign pairs will be counted follows the binomial distribution

[2] In order to distinguish between truth-level electron pairs and identified ones, we will always write truth-level to indicate the former. Thus if a pair is not preceded by truth-level, it is tacitly understood to be an identified pair.

- At least two reconstructed electron candidates with $|\eta| < 2.47$. The invariant mass of the electron pair must be within ± 15 GeV of the Z mass.

Figure 5.1 [38] shows the invariant mass distribution m_{ee} in data and simulation for E_T between 25 GeV and 50 GeV and $0.0 < \eta < 0.8$ (left) or $2.0 < \eta < 2.47$ (right). Due to charge mis-identification same-sign electron pairs also exist in addition to opposite-sign pairs, indicating a charge mis-identification rate of $\sim 10^{-3}$ in the central region and $\sim 10\%$ in the high η region. The higher rates in the latter is expected because of the larger amount of material and the limited lever arm in the forward region. In both cases, same-sign pairs show a broader peak that is also slightly shifted towards lower values, consistent with the fact that the radiation that causes charge mis-identification also results in energy loss.

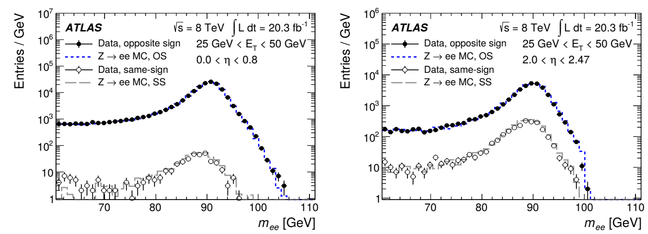

Figure 5.1: Distribution of the invariant mass m_{ee} for E_T between 25 and 50 GeV and $|\eta|$ between 0.0 and 0.8 [38]. Due to charge mis-identification same-sign pairs as well as opposite-sign pairs are observed.

The next section discusses the Poisson likelihood function that is used to fit the data.

5.2.2 The Poisson Likelihood

In a truth-level e^+e^- pair, which will also be called a truth-level opposite-sign pair, if the charge of any one of the electrons is mis-identified, then a same-sign pair will be observed instead[2]. Assuming a probability p that a truth-level opposite-pair will be identified as a same-sign pair, then in considering n truth-level pairs e^+e^-, the probability that exactly n_{ss} same-sign pairs will be counted follows the binomial distribution

$$P(n_{ss}) = \binom{n}{n_{ss}} p^{n_{ss}} (1-p)^{n-n_{ss}}.$$

The charge mis-identification probability p is typically small while the sample of n pairs of electrons considered is typically very large, and therefore the Poisson

[2] In order to distinguish between truth-level electron pairs and identified ones, we will always write truth-level to indicate the former. Thus if a pair is not preceded by "truth-level", it is tacitly understood to be an identified pair.

$$P(n_{ss}) = \binom{n}{n_{ss}} p^{n_{ss}}(1-p)^{n-n_{ss}}.$$

The charge mis-identification probability p is typically small while the sample of n pairs of electrons considered is typically very large, and therefore the Poisson distribution may be used instead. Thus, let

$$m_{ss} = np \tag{5.1}$$

denote the expected number of same-sign pairs, then the Poisson distribution

$$P(n_{ss}) = \frac{m_{ss}^{n_{ss}} e^{-m_{ss}}}{n_{ss}!} \tag{5.2}$$

gives the probability of counting n_{ss} same-sign pairs, given the expected number of same-sign pairs m_{ss}. This will be used as a likelihood function, to be maximized to extract the charge mis-identification rates, as will be explained further below.

The probability p that a truth-level opposite-pair will be identified as a same-sign pair may be written directly in term of the probability of charge mis-identification associated to an individual electron. If ϵ denotes the latter probability, then because a same-sign pair will be observed precisely when only one of the electrons has its charge mis-identified, we may write

$$p = (1-\epsilon)\epsilon + \epsilon(1-\epsilon). \tag{5.3}$$

The Poisson likelihood of Equation 5.2 may now be written to depend explicitly on ϵ:

$$P(n_{ss}|\epsilon) = \frac{m_{ss}^{n_{ss}} e^{-m_{ss}}}{n_{ss}!}, \quad \text{where} \quad m_{ss} = np = n(1-\epsilon)\epsilon + \epsilon(1-\epsilon). \tag{5.4}$$

The maximization of this function gives the mis-identification rates ϵ's.

On the other hand, because charge mis-identification rates are expected to show strong dependencies on p_T and η of the electrons (Section 5.1), they are often measured in bins of these two quantities. In such a situation the electrons in a pair generally belong to different bins and that needs to be taken into account in the likelihood function. Thus, the electrons are assigned charge mis-identification probabilities ϵ_i and ϵ_j, where the indices i and j indicate the bins, and we write

- The probability

  $$p_{ij} = (1-\epsilon_i)\epsilon_j + \epsilon_i(1-\epsilon_j) \tag{5.5}$$

  in place of the probability p in Equation 5.3. This is the probability an opposite-sign pair may be seen as a same-sign pair in the bin pair (i, j)

- The number of electron pairs considered, n_{ij}, in the bin pair (i, j)

48

distribution may be used instead. Thus, let

$$m_{ss} = np \tag{5.1}$$

denote the expected number of same-sign pairs, then the Poisson distribution

$$P(n_{ss}) = \frac{m_{ss}^{n_{ss}} e^{-m_{ss}}}{n_{ss}!} \tag{5.2}$$

gives the probability of counting n_{ss} same-sign pairs, given the expected number of same-sign pairs m_{ss}. This will be used as a likelihood function, to be maximized to extract the charge mis-identification rates, as will be explained further below.

The probability p that a truth-level opposite-pair will be identified as a same-sign pair may be written directly in term of the probability of charge mis-identification associated to an individual electron. If ϵ denotes the latter probability, then because a same-sign pair will be observed precisely when only one of the electrons has its charge mis-identified, we may write

$$p = (1-\epsilon)\epsilon + \epsilon(1-\epsilon). \tag{5.3}$$

The Poisson likelihood of Equation 5.2 may now be written to depend explicitly on ϵ:

$$P(n_{ss}|\epsilon) = \frac{m_{ss}^{n_{ss}} e^{-m_{ss}}}{n_{ss}!}, \quad \text{where} \quad m_{ss} = np = n(1-\epsilon)\epsilon + \epsilon(1-\epsilon). \tag{5.4}$$

The maximization of this function gives the mis-identification rates ϵ's.

On the other hand, because charge mis-identification rates are expected to show strong dependencies on p_T and η of the electrons (Section 5.1), they are often measured in bins of these two quantities. In such a situation the electrons in a pair generally belong to different bins and that needs to be taken into account in the likelihood function. Thus, the electrons are assigned charge mis-identification probabilities ϵ_i and ϵ_j, where the indices i and j indicate the bins, and we write

- The probability

  $$p_{ij} = (1-\epsilon_i)\epsilon_j + \epsilon_i(1-\epsilon_j) \tag{5.5}$$

  in place of the probability p in Equation 5.3. This is the probability an opposite-sign pair may be seen as a same-sign pair in the bin pair (i, j)

- The number of electron pairs considered, n_{ij}, in the bin pair (i, j)

- The expected number of same-sign pairs

  $$m_{ss,ij} = n_{ij} p_{ij} \tag{5.6}$$

  in place of the expected number of same-sign pairs in Equation 5.1

- The expected number of same-sign pairs

$$m_{ss,ij} = n_{ij} p_{ij} \tag{5.6}$$

  in place of the expected number of same-sign pairs in Equation 5.1

- The Poisson likelihood

$$P(n_{ss,ij}|\epsilon_i, \epsilon_j) = \frac{m_{ss,ij}^{n_{ss,ij}} e^{-m_{ss,ij}}}{n_{ss,ij}!}, \quad \text{where} \quad m_{ss,ij} = n_{ij} p_{ij} = (1-\epsilon_i)\epsilon_j + \epsilon_i(1-\epsilon_j). \tag{5.7}$$

  in place of the Poisson likelihood in Equation 5.4. This will also be denoted simply as L_{ij}

These equations are valid whether the rates are extracted in only p_T bins, only η bins, or both, because in the latter case the grid of two-dimensional bins may be treated as a long one-dimensional sequence of bins. On the other hand, all the possible bin pairs (i, j) need to be used and therefore, assuming statistically-independent rates, the rates ϵ_i to be extracted come from the maximization of the likelihood function

$$L = \prod_{i,j} L_{ij},$$

the data being n_{ij}, the numbers of electrons observed in the bin pair (i, j), and $n_{ss,ij}$, the number of same-sign electron pairs observed in the bin pair (i, j).

Background subtractions Backgrounds to $Z \to e^+e^-$ events consistly mostly of events involving top quarks, diboson events, and W+jets events. They are assumed to be flat in the invariant Z mass peak selection and are subtracted by a method called the sideband method. To this end, we will denote the invariant mass interval around the Z mass peak by (m_l, m_h), where $m_l = 15$ GeV is the low mass point and $m_h = 15$ GeV the high mass point. Then an interval of 15 GeV is selected to the left of m_l and to the right of m_h, i.e. $m_l = m_h = 15$ GeV and there are now two side intervals $(m_l - w_l, m_l)$ and $(m_h, m_h + w_h)$ in addition to the original interval (m_l, m_h). The side intervals are assumed to be dominated by background events and are used to compute the backgrounds in the (m_l, m_h) interval, i.e. to subtract background contamination in n_{ij} and $n_{ss,ij}$, quantities that need to be counted in the (m_l, m_h) interval. We will write $b(n_{ij})$ for the background contamination in n_{ij}, and $b(n_{ss,ij})$ for the background contamination in $n_{ss,ij}$; they will be computed as weighted quantities:

$$b(n_{ij}) = \frac{w_l \times n_{ij}^l + w_h \times n_{ij}^h}{w_l + w_h}, \qquad b(n_{ss,ij}) = \frac{w_l \times n_{ss,ij}^l + w_h \times n_{ss,ij}^h}{w_l + w_h}.$$

The terms n_{ij} and $n_{ss,ij}$ and the background terms $b(n_{ij})$ and $b(n_{ss,ij})$ are put into the Poisson likelihood (Equation 5.7):

49

- The Poisson likelihood

$$P(n_{ss,ij}|\epsilon_i, \epsilon_j) = \frac{m_{ss,ij}^{n_{ss,ij}} e^{-m_{ss,ij}}}{n_{ss,ij}!}, \quad \text{where} \quad m_{ss,ij} = n_{ij} p_{ij} = (1-\epsilon_i)\epsilon_j + \epsilon_i(1-\epsilon_j), \tag{5.7}$$

  in place of the Poisson likelihood in Equation 5.4. This will also be denoted simply as L_{ij}

These equations are valid whether the rates are extracted in only p_T bins, only η bins, or both, because in the latter case the grid of two-dimensional bins may be treated as a long one-dimensional sequence of bins. On the other hand, all the possible bin pairs (i, j) need to be used and therefore, assuming statistically-independent rates, the rates ϵ_i to be extracted come from the maximization of the likelihood function

$$L = \prod_{i,j} L_{ij},$$

the data being n_{ij}, the numbers of electrons observed in the bin pair (i, j), and $n_{ss,ij}$, the number of same-sign electron pairs observed in the bin pair (i, j).

Background subtractions Backgrounds to $Z \to e^+e^-$ events consistly mostly of events involving top quarks, diboson events, and W+jets events. They are assumed to be flat in the invariant Z mass peak selection and are subtracted by a method called the sideband method. To this end, we will denote the invariant mass interval around the Z mass peak by (m_l, m_h), where $m_l = 15$ GeV is the low mass point and $m_h = 15$ GeV the high mass point. Then an interval of 15 GeV is selected to the left of m_l and to the right of m_h, i.e. $m_l = m_h = 15$ GeV and there are now two side intervals $(m_l - w_l, m_l)$ and $(m_h, m_h + w_h)$ in addition to the original interval (m_l, m_h). The side intervals are assumed to be dominated by background events and are used to compute the backgrounds in the (m_l, m_h) interval, i.e. to subtract background contamination in n_{ij} and $n_{ss,ij}$, quantities that need to be counted in the (m_l, m_h) interval. We will write $b(n_{ij})$ for the background contamination in n_{ij}, and $b(n_{ss,ij})$ for the background contamination in $n_{ss,ij}$; they will be computed as weighted quantities:

$$b(n_{ij}) = \frac{w_l \times n_{ij}^l + w_h \times n_{ij}^h}{w_l + w_h}, \qquad b(n_{ss,ij}) = \frac{w_l \times n_{ss,ij}^l + w_h \times n_{ss,ij}^h}{w_l + w_h}.$$

The terms n_{ij} and $n_{ss,ij}$ and the background terms $b(n_{ij})$ and $b(n_{ss,ij})$ are put into the Poisson likelihood (Equation 5.7):

$$P(n_{ss,ij}|\epsilon_i, \epsilon_j) = \frac{m_{ss,ij}^{n_{ss,ij}} e^{-m_{ss,ij}}}{n_{ss,ij}!}$$

in which the background terms make a contribution to the expected number of same-sign $m_{ss,ij}$ in the likelihood, modifying it from $m_{ss,ij} = n_{ij} p_{ij}$ (Equation 5.6) to

$$m_{ss,ij} = (n_{ij} - b(n_{ij})) \times p_{ij} + b(n_{ss,ij}).$$

No indentation, also below, check everywhere

Punctuate with a "." And the one above with ",". Check everywhere

$$P(n_{ss,ij}|\epsilon_i,\epsilon_j) = \frac{m_{ss,ij}^{n_{ss,ij}} e^{-m_{ss,ij}}}{n_{ss,ij}!}$$

in which the background terms make a contribution to the expected number of same-sign $m_{ss,ij}$ in the likelihood, modifying it from $m_{ss,ij} = n_{ij}p_{ij}$ (Equation 5.6) to

$$m_{ss,ij} = (n_{ij} - b(n_{ij})) \times p_{ij} + b(n_{ss,ij})$$

The first quantity on the right in the equation above is the same-sign contribution from signal events where the background events have to be subtracted, and the second quantity is the contribution from background events.

5.2.3 Charge Mis-identification Rates and Uncertainties

The rates are obtained upon the maximization of the likelihood function discussed in the previous section. The statistical uncertainties associated with the estimated rates depend on the statistics of the data, and are given by the statistical tool that maximizes the Poisson likelihood.

The following sources of systematic uncertainties are evaluated:

- Systematic uncertainty that comes from background subtraction, which is evaluated by determining the rates with and without background subtraction. The inclusion of this uncertainty ensures a conservative figure of systematic uncertainty in the charge mis-identification rates; it has a small impact because the background is small.
- The invariant mass interval (m_l, m_h) may be varied, from 15 GeV around the Z mass to 10 and 20 GeV additionally. This provides an estimation of the impact of the choice of mass window on the measure rates.
- The invariant mass widths w_l and w_h may be varied, taking values 20, 25, or 30 GeV. Thus, the uncertainty on the rates due to the choice of a mass width is taken into account.

The actual rates are estimated for the following three sets of requirements:

- Medium: Medium identification requirements
- Tight + isolation: Tight identification requirements plus track isolation cut $p_T^{\text{cone }0.2}/E_T < 0.14$.
- Tight + isolation + impact parameter: Tight identification plus $E_T^{\text{cone }0.3}/E_T < 0.14$ and $p_T^{\text{cone }0.2}/E_T < 0.07$ and additionally $|z_0| \times \sin\theta < 0.5$ mm and $|d_0|/\sigma_{d_0} < 5.0$

Figure 5.2 [38] shows the estimated rates in data and simulation, for electron E_T between 25 and 50 GeV as a function of η, the variable upon which they depend the most. The dashed lines indicate the bins in which the rates are calculated. The total

50

The first quantity on the right in the equation above is the same-sign contribution from signal events where the background events have to be subtracted, and the second quantity is the contribution from background events.

5.2.3 Charge Mis-identification Rates and Uncertainties

The rates are obtained upon the maximization of the likelihood function discussed in the previous section. The statistical uncertainties associated with the estimated rates depend on the statistics of the data, and are given by the statistical tool that maximizes the Poisson likelihood.

The following sources of systematic uncertainties are evaluated:

- Systematic uncertainty that comes from background subtraction, which is evaluated by determining the rates with and without background subtraction. The inclusion of this uncertainty ensures a conservative figure of systematic uncertainty in the charge mis-identification rates; it has a small impact because the background is small.
- The invariant mass interval (m_l, m_h) may be varied, from 15 GeV around the Z mass to 10 and 20 GeV additionally. This provides an estimation of the impact of the choice of mass window on the measure rates.
- The invariant mass widths w_l and w_h may be varied, taking values 20, 25, or 30 GeV. Thus, the uncertainty on the rates due to the choice of a mass width is taken into account.

The actual rates are estimated for the following three sets of requirements:

- Medium: Medium identification requirements
- Tight + isolation: Tight identification requirements plus track isolation cut $p_T^{\text{cone }0.2}/E_T < 0.14$.
- Tight + isolation + impact parameter: Tight identification plus $E_T^{\text{cone }0.3}/E_T < 0.14$ and $p_T^{\text{cone }0.2}/E_T < 0.07$ and additionally $|z_0| \times \sin\theta < 0.5$ mm and $|d_0|/\sigma_{d_0} < 5.0$

Figure 5.2 [38] shows the estimated rates in data and simulation, for electron E_T between 25 and 50 GeV as a function of η, the variable upon which they depend the most. The dashed lines indicate the bins in which the rates are calculated. The total uncertainty, which is computed as the sum in quadrature of statistical and systematic uncertainties, is also showed. Charge mis-identification rates vary from below 1% in the central region to $\sim 10\%$ in high η region, reflecting the correlation of the rates with bremsstrahlung, and thus a dependency on the amount of the material traversed. On the other hand, tighter selection criteria, in particular requirements uncertainty, which is computed as the sum in quadrature of statistical and systematic uncertainties, is also showed. Charge mis-identification rates vary from below 1% in the central region to $\sim 10\%$ in high η region, reflecting the correlation of the rates with bremsstrahlung, and thus a dependency on the amount of the material traversed. On the other hand, tighter selection criteria, in particular requirements on the isolation or track parameters, may decrease the charge misidentification probability by a factor of up to four, depending on the additional selection requirements[3]. Moreover, as is seen, simulation over-estimates the rates as compared to the data by 5-20% depending on η and electron requirements.

Charge mis-identification rates are known to show a positive correlation with p_T as well (Figure 5.3).

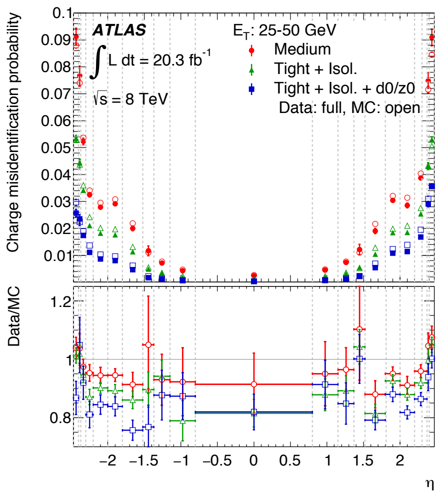

Figure 5.2: Charge mis-identification probabilities in η bins, E_T between 25 GeV and 50 GeV [38]. Three different sets of selection requirements (Medium, Tight + Isolation, and Tight + Isolation + impact parameter) are shown, along with simulation expectations. Displayed in the lower panel is the data-to-simulation ratios. The uncertainties are the total uncertainties from the sum in quadrature of statistical and systematic uncertainties. The dashed lines indicate the bins in which the rates are calculated.

[3]The energy in the cone around an electron could indicate the amount of energy deposited by bremsstrahlung, and large values of the track impact parameters could mean that the track matched to the electron is not a prompt track from the primary vertex but from a secondary interaction or bremsstrahlung and a subsequent conversion [38].

51

on the isolation or track parameters, may decrease the charge misidentification probability by a factor of up to four, depending on the additional selection requirements[3]. Moreover, as is seen, simulation over-estimates the rates as compared to the data by 5-20% depending on η and electron requirements.

Charge mis-identification rates are known to show a positive correlation with p_T as well (Figure 5.3).

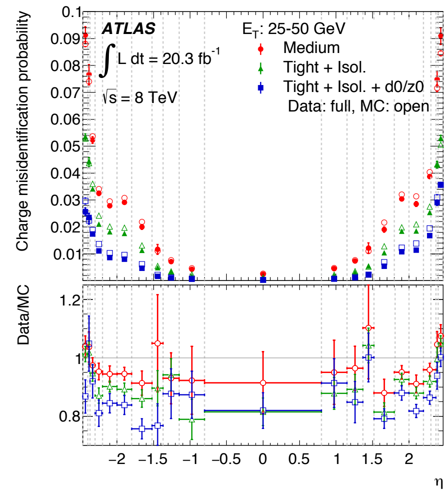

Figure 5.2: Charge mis-identification probabilities in η bins, E_T between 25 GeV and 50 GeV [38]. Three different sets of selection requirements (Medium, Tight + Isolation, and Tight + Isolation + impact parameter) are shown, along with simulation expectations. Displayed in the lower panel is the data-to-simulation ratios. The uncertainties are the total uncertainties from the sum in quadrature of statistical and systematic uncertainties. The dashed lines indicate the bins in which the rates are calculated.

5.2.4 Estimating Charge Mis-identification Background from the Charge Mis-identification Rates

In this section we give an example of how the charge mis-identification rates may be used to estimate the charge mis-identification background in analysis with a same-

[3]The energy in the cone around an electron could indicate the amount of energy deposited by bremsstrahlung, and large values of the track impact parameters could mean that the track matched to the electron is not a prompt track from the primary vertex but from a secondary interaction or bremsstrahlung and a subsequent conversion [38].

Add a sentence: This technique is used in the SUSY same-sign leptons search. With a reference at the end.

Again

Integrate to main text

5.2.4 Estimating Charge Mis-identification Background from the Charge Mis-identification Rates

In this section we give an example of how the charge mis-identification rates may be used to estimate the charge mis-identification background in analysis with a same-sign lepton pair signature. Thus, given $n_{\mathrm{ss,ij}}$ of same-sign electron pairs that has been selected in the bin pairs (i, j) (Section 5.2), we want to determine the charge mis-identification contribution to it.

To begin, the number of same-sign electron pairs $n_{\mathrm{ss,ij}}$ that has been selected under a set of selection requirements is to be distinguished from the number of truth-level same-sign electron pairs. The latter is what would be counted in the bin pairs (i, j) if there were no charge mis-identification. In the following we will write it by $\bar{n}_{\mathrm{ss,ij}}$.

A charge mis-identification contribution occurs whenever there is a truth-level opposite-sign pair of electrons in which one of the electron has its charge mis-identified. The probability for this to happen is, according to Equation 5.5,

$$p_{ij} = (1 - \epsilon_i)\epsilon_j + \epsilon_i(1 - \epsilon_j,)$$

where ϵ_i and ϵ_j are the charge mis-identification rates in the bins. Now, in the same bin pair (i, j) the number of opposite-sign pairs obtained from the same selection requirements may be counted as well, we will write it as $n_{\mathrm{os,ij}}$. Moreover, as for the same-sign case, this has to be distinguished from the number of truth-level opposite-sign pairs, which will be denoted $\bar{n}_{\mathrm{os,ij}}$. The number of interest is $\bar{n}_{\mathrm{os,ij}}$, because given the mis-identification rate p_{ij}, the charge mis-identification contribution to $n_{\mathrm{ss,ij}}$ is simply $\bar{n}_{\mathrm{os,ij}} \times p_{ij}$.

The only quantities known are $n_{\mathrm{ss,ij}}$, $n_{\mathrm{os,ij}}$, and the mis-identification rates ϵ_i and ϵ_j, while $\bar{n}_{\mathrm{ss,ij}}$ and $\bar{n}_{\mathrm{os,ij}}$ are unknown. However, the following relation holds

$$n_{\mathrm{os,ij}} = \bar{n}_{\mathrm{os,ij}} - \bar{n}_{\mathrm{os,ij}} \times p_{ij} + \bar{n}_{\mathrm{ss,ij}} \times p_{ij},$$

which reflects the fact that the number of opposite-sign lepton pairs counted in the bin pair (i, j) is the corresponding truth-level number minus the portion that is identified as same-sign plus the contribution from truth-level same-sign pairs. This may be re-written as

$$n_{\mathrm{os,ij}} = \bar{n}_{\mathrm{os,ij}} \times (1 - p_{ij}) + \bar{n}_{\mathrm{ss,ij}} \times p_{ij}.$$

Similarly we have the following relation

$$n_{\mathrm{ss,ij}} = \bar{n}_{\mathrm{ss,ij}} \times (1 - p_{ij}) + \bar{n}_{\mathrm{os,ij}} \times p_{ij}$$

Thus there are two equations in two unknowns and as a result $\bar{n}_{\mathrm{os,ij}}$ and $\bar{n}_{\mathrm{ss,ij}}$ may be solved.

At ATLAS, charge mis-identification rates are also provided to different analyses as scale factors[4], to be applied to charge mis-identification rates in simulations to match the data. If charge mis-identification rates on data are provided directly instead of the scale factors we can avoid the need for the use of all systematic uncertainties that are associated with the use of simulation samples.

[4] These are the ratios of charge mis-identification rates in data over those in simulation.

sign lepton pair signature. This technique is used in the SUSY same-sign leptons search [51]. Thus, given $n_{\mathrm{ss,ij}}$ of same-sign electron pairs that has been selected in the bin pairs (i, j) (Section 5.2), we want to determine the charge mis-identification contribution to it.

To begin, the number of same-sign electron pairs $n_{\mathrm{ss,ij}}$ that has been selected under a set of selection requirements is to be distinguished from the number of truth-level same-sign electron pairs. The latter is what would be counted in the bin pairs (i, j) if there were no charge mis-identification. In the following we will write it by $\bar{n}_{\mathrm{ss,ij}}$.

A charge mis-identification contribution occurs whenever there is a truth-level opposite-sign pair of electrons in which one of the electron has its charge mis-identified. The probability for this to happen is, according to Equation 5.5,

$$p_{ij} = (1 - \epsilon_i)\epsilon_j + \epsilon_i(1 - \epsilon_j,),$$

where ϵ_i and ϵ_j are the charge mis-identification rates in the bins. Now, in the same bin pair (i, j) the number of opposite-sign pairs obtained from the same selection requirements may be counted as well, we will write it as $n_{\mathrm{os,ij}}$. Moreover, as for the same-sign case, this has to be distinguished from the number of truth-level opposite-sign pairs, which will be denoted $\bar{n}_{\mathrm{os,ij}}$. The number of interest is $\bar{n}_{\mathrm{os,ij}}$, because given the mis-identification rate p_{ij}, the charge mis-identification contribution to $n_{\mathrm{ss,ij}}$ is simply $\bar{n}_{\mathrm{os,ij}} \times p_{ij}$.

The only quantities known are $n_{\mathrm{ss,ij}}$, $n_{\mathrm{os,ij}}$, and the mis-identification rates ϵ_i and ϵ_j, while $\bar{n}_{\mathrm{ss,ij}}$ and $\bar{n}_{\mathrm{os,ij}}$ are unknown. However, the following relation holds

fixed also

$$n_{\mathrm{os,ij}} = \bar{n}_{\mathrm{os,ij}} - \bar{n}_{\mathrm{os,ij}} \times p_{ij} + \bar{n}_{\mathrm{ss,ij}} \times p_{ij},$$

which reflects the fact that the number of opposite-sign lepton pairs counted in the bin pair (i, j) is the corresponding truth-level number minus the portion that is identified as same-sign plus the contribution from truth-level same-sign pairs. This may be re-written as

$$n_{\mathrm{os,ij}} = \bar{n}_{\mathrm{os,ij}} \times (1 - p_{ij}) + \bar{n}_{\mathrm{ss,ij}} \times p_{ij}.$$

Similarly we have the following relation

$$n_{\mathrm{ss,ij}} = \bar{n}_{\mathrm{ss,ij}} \times (1 - p_{ij}) + \bar{n}_{\mathrm{os,ij}} \times p_{ij}$$

fixed also

Thus there are two equations in two unknowns and as a result $\bar{n}_{\mathrm{os,ij}}$ and $\bar{n}_{\mathrm{ss,ij}}$ may be solved.

At ATLAS, charge mis-identification rates are also provided to different analyses as scale factors (the ratios of charge mis-identification rates in data over those in simulation), to be applied to charge mis-identification rates in simulations to match the data. If charge mis-identification rates on data are provided directly instead of the scale factors we can avoid the need for the use of all systematic uncertainties that are associated with the use of simulation samples.

5.3 Conclusions

This chapter describes the electron charge mis-identification problem at ATLAS and how the charge mis-identification rates are extracted by fitting a Poisson likelihood function using the $Z \to e^+e^-$ data sample, collected at 8 TeV LHC center-of-mass energy in 2012 with the ATLAS detector and corresponding to an integrated luminosity of 20.3 fb^{-1}. Three sets of charge mis-identification rates are measured and provided to ATLAS analyses, corresponding to three different sets of selection requirements (Medium, Tight + Isolation, and Tight + Isolation + impact parameter). The rates show a variation from less than 1% to nearly 10% depending on η and p_T. It is also observed from the measurements that, in general, simulation underestimates the charge mis-identification rates as compared to those in the data.

In Run 2, in addition to measuring the charge mis-identification rates, a separate effort was started by the physics team at Université de Montréal, aiming at reducing charge mis-identification. The technique relies on the output of a boosted decision tree using a simulated sample of single electrons. Figure 5.3 shows the impact of applying the BDT requirement on charge mis-identification rates; it has been demonstrated to reduce charge mis-identification rates by about a factor of 10 while maintaining a 97% efficiency on signal electrons. More details may be found in Ref. [38].

53

5.3 Conclusions

This chapter describes the electron charge mis-identification problem at ATLAS and how the charge mis-identification rates are extracted by fitting a Poisson likelihood function using the $Z \to e^+e^-$ data sample, collected at 8 TeV LHC center-of-mass energy in 2012 with the ATLAS detector and corresponding to an integrated luminosity of 20.3 fb^{-1}. Three sets of charge mis-identification rates are measured and provided to ATLAS analyses, corresponding to three different sets of selection requirements (Medium, Tight + Isolation, and Tight + Isolation + impact parameter). The rates show a variation from less than 1% to nearly 10% depending on η and p_T. It is also observed from the measurements that, in general, simulation underestimates the charge mis-identification rates as compared to those in the data.

In Run 2, in addition to measuring the charge mis-identification rates, a separate effort was started by the physics team at Université de Montréal, aiming at reducing charge mis-identification. The technique relies on the output of a boosted decision tree using a simulated sample of single electrons. Figure 5.3 shows the impact of applying the BDT requirement on charge mis-identification rates; it has been demonstrated to reduce charge mis-identification rates by about a factor of 10 while maintaining a 97% efficiency on signal electrons. More details may be found in Ref. [38].

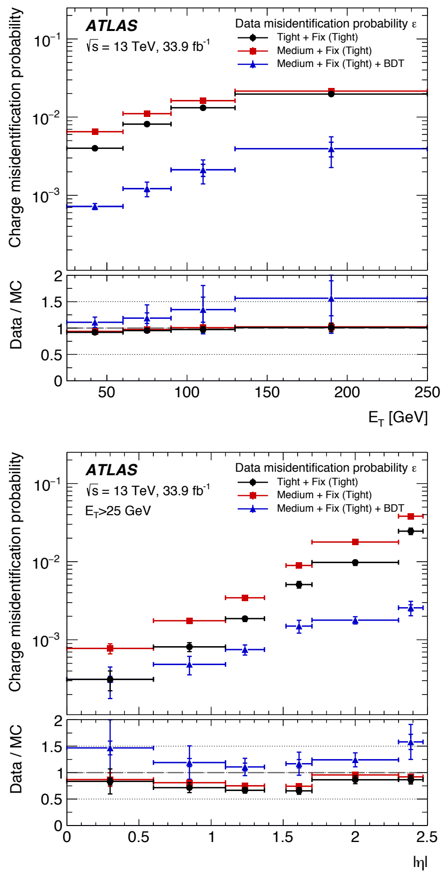

Figure 5.3: Charge mis-identification probabilities in 2016 data and $Z \to e^+e^-$ events as a function of E_T (top) and $|\eta|$ (bottom) that shows also the impact of applying the BDT requirement (in blue) to suppress charge mis-identification.

54

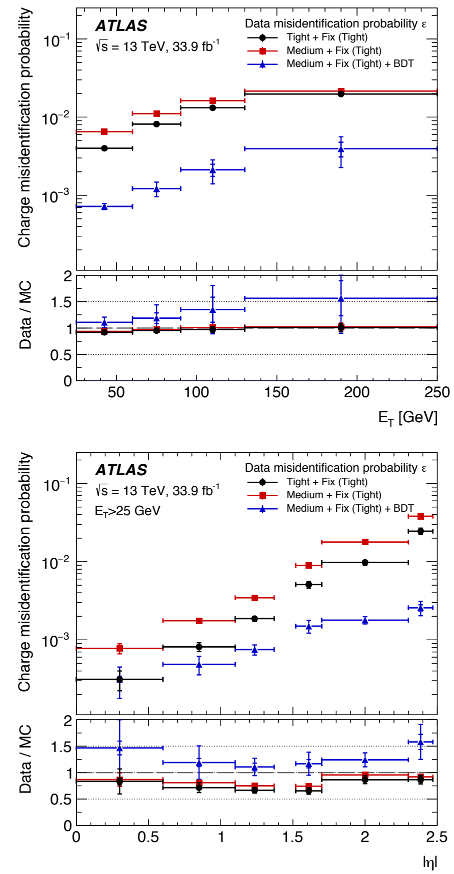

Figure 5.3: Charge mis-identification probabilities in 2016 data and $Z \to e^+e^-$ events as a function of E_T (top) and $|\eta|$ (bottom) that shows also the impact of applying the BDT requirement (in blue) to suppress charge mis-identification.