

UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

SEARCHING FOR SUPERSYMMETRIC PARTICLES AT THE LARGE HADRON
COLLIDER USING THE ATLAS DETECTOR

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

OTHMANE RIFKI
Norman, Oklahoma
2016

SEARCHING FOR SUPERSYMMETRIC PARTICLES AT THE LARGE HADRON
COLLIDER USING THE ATLAS DETECTOR

A DISSERTATION APPROVED FOR THE
HOMER L. DODGE DEPARTMENT OF PHYSICS AND
ASTRONOMY

BY

Dr. Brad Abbott, Chair

Dr. Peter Parker, Outside Member

Dr. Mike Strauss

Dr. Chung Kao

Dr. Eric Abraham

To my loving parents and sister.

Acknowledgements

I would like to acknowledge...

Table of Contents

List of Tables	vii
List of Figures	x
Abstract	xi
1 Theoretical Background	2
2 The LHC and the ATLAS Experiment	3
3 The Region of Interest Builder	4
4 Object Reconstruction	5
5 Data-driven techniques to estimate fake lepton backgrounds	6
5.1 The problem of fakes	6
5.2 Common processes for faking leptons	7
5.3 Monte Carlo Template Method	9
5.3.1 Motivation	9
5.3.2 Description of the method	10
5.3.3 Correction factors	12
5.3.4 Control regions	13
5.4 Matrix Method	16
5.4.1 Events with one object	17
5.4.2 Dynamic matrix method	18
5.4.3 Propagation of uncertainties	20
5.5 Real lepton efficiency	21
5.5.1 The Z tag-and-probe method	21
5.5.2 ϵ Background subtraction	24
5.5.3 Real lepton efficiencies	30
5.5.4 ϵ Cut efficiencies	30
5.5.5 ϵ Real lepton efficiencies	32
5.5.6 ϵ Tag-and-probe method and truth matching comparisons	34
5.5.7 ϵ Data-to-MC comparisons	36
5.5.8 ϵ Real lepton efficiency versus pileup	36
5.5.9 Sources of systematic uncertainties	38
5.5.10 ϵ Measurement systematics	38
5.5.11 ϵ Trigger bias	43
5.5.12 ϵ Extrapolation to signal regions	44
5.5.13 ϵ Final uncertainties	52
5.6 Fake lepton rate	54
6 Search for new physics in events with two same sign leptons or three leptons	57

List of Tables

5.1	The FNP and charge flip multipliers obtained after minimizing the likelihood function using Pythia. The uncertainty in the multipliers takes into account the limited statistics of simulated events.	18
5.2	The FNP and charge flip multipliers obtained after minimizing the likelihood function using Sherpa. The uncertainty in the multipliers takes into account the limited statistics of simulated events.	19
5.3	The list of single lepton and dilepton triggers used for the real lepton efficiency measurements. The dilepton triggers are used for studying the systematic uncertainties causing by the trigger.	22
5.4	The definition of the background templates used to estimate the background contamination associated with the Z tag-and-probe method. The candidate template is used to estimate the background contamination and the variation 1 and 2 templates, which have looser and tighter requirements, are used to assess the systematic caused by the background contamination.	26
5.5	The estimated background contamination using the background template. The p_T and $ \eta $ binnings correspond to the one used for the final measurements.	29
5.6	The real electron efficiencies in percentage before and after applying the background subtraction. The efficiency in three different $ \eta $ regions for $10 < p_T < 15\text{GeV}$ and $15 < p_T < 20\text{GeV}$ are shown.	29
5.7	The measurement systematic uncertainties in percentage on the real electron efficiencies. The background subtraction is applied on the first two p_T bins on the electron case so there are three different sources of the measurement systematic uncertainties, varying templates, varying fitting ranges, and varying $m_{\ell\ell}$ windows denoted by t, f, and m, respectively. Only the variations of $m_{\ell\ell}$ windows are applied on the electron case when $p_T > 20\text{GeV}$	41
5.8	The measurement systematic uncertainties in percentage on the real muon efficiencies. Because the background subtraction is not applied on the muon case, only the variations of $m_{\ell\ell}$ windows are applied on the muon case for all p_T regions.	42
5.9	The systematic uncertainties in percentage due to the different trigger strategies for the real electron efficiencies. The uncertainties of each trigger strategy are computed with respect to the one applied single lepton trigger and required tag trigger matched. And the total uncertainties are the quadratic sum of the uncertainties of each trigger strategy.	47
5.10	The systematic uncertainties in percentage due to the different trigger strategies for the real muon efficiencies. The uncertainties of each trigger strategy are computed with respect to the one applied single lepton trigger and required tag trigger matched. And the total uncertainties are the quadratic sum of the uncertainties of each trigger strategy.	48
5.11	Systematic uncertainties on the measured real lepton efficiency due to extrapolation to busy environments using	53

5.12	The final uncertainties of the real electron efficiencies. The final uncertainties are calculated using the quadratic sum of the statistical uncertainties and the systematic uncertainties. The systematic uncertainties contain the measurement uncertainties, the trigger uncertainties, and the uncertainties come from the difference between truth matching and tag-and-probe methods. The trigger uncertainties are included for $p_T > 20\text{GeV}$ only. The uncertainties in the busy environment do not incorporate in the final uncertainties calculation because it is a function of p_T and ΔR .	55
5.13	The final uncertainties of the real muon efficiencies. The final uncertainties are calculated using the quadratic sum of the statistical uncertainties and the systematic uncertainties. The systematic uncertainties contain the measurement uncertainties, the trigger uncertainties, and the uncertainties come from the difference between truth matching and tag-and-probe methods. The uncertainties in the busy environment do not incorporate in the final uncertainties calculation because it is a function of p_T and ΔR .	56

List of Figures

5.1	Lepton classification.	11
5.2	Pre-fit distributions for ee channel (left), for $e\mu$ channel (middle), and for $\mu\mu$ channel (right) from CR0b that were used in the fit to extract the FNP lepton and charge flip multipliers. The generator used in these plots is Powheg. The hashed band represents the sum of systematic uncertainties on the predictions.	14
5.3	Post-fit distributions for ee channel (left), for $e\mu$ channel (middle), and for $\mu\mu$ channel (right) from CR0b that were used in the fit to extract the FNP lepton and charge flip multipliers. The generator used in these plots is Powheg. The hashed band represents the sum of systematic uncertainties on the predictions.	15
5.4	Pre-fit distributions for ee channel (left), for $e\mu$ channel (middle), and for $\mu\mu$ channel (right) from CR1b that were used in the fit to extract the FNP lepton and charge flip multipliers. The generator used in these plots is Powheg. The hashed band represents the sum of systematic uncertainties on the predictions.	16
5.5	Post-fit distributions for ee channel (left), for $e\mu$ channel (middle), and for $\mu\mu$ channel (right) from CR1b that were used in the fit to extract the FNP lepton and charge flip multipliers. The generator used in these plots is Powheg. The hashed band represents the sum of systematic uncertainties on the predictions.	17
5.6	The distributions of the tag-and-probe pair invariant mass ($m_{\ell\ell}$) computed using $Z+jets$ MC and 2015 + 2016 data. The blue color represents the Z truth matched events, the red color stands for the Z tag-and-probe events, and the black dots are data. The top row shows the m_{ee} computed with probe electrons in the $10 < p_T < 15\text{GeV}$ (left) and $15 < p_T < 20\text{GeV}$ (right) intervals. The bottom row shows the $m_{\mu\mu}$ cases. The MC distributions are scaled to the data using a Gaussian fit of the Z mass peak ($85 < m_{\ell\ell} < 95\text{GeV}$).	23
5.7	The m_{ee} distributions of the three different background templates. The m_{ee} distributions in the left and right plots are computed using probe electrons with $10 < p_T < 15\text{GeV}$ and $15 < p_T < 20\text{GeV}$, respectively. Because the variation 1 template has looser calorimeter and track isolation requirements, we see a peak in the Z mass region. Compared with the variation 1 template, the candidate and the variation 2 templates have tighter cuts so no peak can be found in the Z mass region.	25

5.8	Plots illustrate the background subtraction procedure. The top row represents the $10 < p_T < 15\text{GeV}$ region and the the bottom row shows the $15 < p_T < 20\text{GeV}$ region. The three columns, from left to right, are the $0 < \eta < 0.8$, $0.8 < \eta < 1.37$, and $1.52 < \eta < 2.0$ regions, respectively. The m_{ee} distributions for data before the background subtraction (full black dots) and after the background subtraction (blue squares) are shown along with the corresponding $Z \rightarrow ee$ MC distributions (open black circles). The MC distribution is normalized to the data after the background subtraction using a Gaussian fit of the Z peak region ($85 < m_{ee} < 95\text{GeV}$). The lower panels are the ratio between data after the background subtraction and the MC. The corresponding background templates (red triangles) and their respective fitting results (green lines) are also shown.	28
5.9	Efficiencies of the signal electron (left) and muon (right) definition cuts as a function of p_T . The leptons are selected from data samples using Z tag-and-probe method. The black points correspond to the total real lepton efficiency. The red squares represent the loose to medium likelihood cut efficiency. The green triangles stand for the calorimeter isolation cut efficiency and the blue triangles are the distribution of track isolation cut efficiency. The longitudinal and tranverse impact parameters cut efficiencies are presented by magenta diamonds and cyan crosses, respectively.	31
5.10	The real lepton efficiencies as a function of p_T and $ \eta $ measured using the Z tag-and-probe method. The left plot corresponds to the real electron efficiencies in three $ \eta $ regions and the right plot stands for the real muon efficiencies in four $ \eta $ regions. The $ \eta $ binning used for the real electron efficiencies measurement corresponds to the geometry of the electromagnetic calorimeter and removing the creak region. A homogeneous $ \eta $ binning has been chosen for the muon case.	33
5.11	The real lepton efficiencies computed using Z tag-and-probe method and truth matching. The top row is the electron case and the bottom row is the muon case. The three columns from the left to the right are the real lepton efficiencies as a function p_T , $ \eta $, and $\Delta R(\ell, jet)$, respectively. The real lepton efficiencies computed using Z tag-and-probe are denoted by the red dots and the one calculated using truth matching are indicated by the blue triangles. The lower pads show the ratio with respect to the Z tag-and-probe method.	35
5.12	The real lepton efficiencies as a function of p_T , $ \eta $ and $\Delta R(\ell, jet)$ measured on data and MC using the Z tag-and-probe method. The plots on the top row correspond to the real electron efficiencies and the plots on the bottown row correspond to the real muon efficiencies. The 2015 + 2016 data are denoted by the black dots, the $Z \rightarrow \ell\ell$ MC are shown using red squares and are reweighted to the data pile-up. The uncertainties shown in the plots are corresponding to the statistical uncertainties only. Some differences between data and $Z \rightarrow \ell\ell$ MC can be seen in the low $\Delta R(\ell, jet)$ region. The binning used for the real electron efficiencies as a function of η corresponds to the geometry of the electromagnetic calorimeter.	37

5.13	The real lepton efficiencies as a function of the average interactions per crossing $\langle \mu \rangle$. The data is shown as black dots and $Z \rightarrow \ell\ell$ tag-and-probe and truth matching are in red squares and blue triangles, respectively. The real lepton efficiencies calculated using $t\bar{t}$ and $\tilde{g} \rightarrow t\bar{t}\tilde{\chi}_1^0$ MC are also shown in magenta diamonds and yellow crosses. And they show the real lepton efficiencies with different event topologies. The $ \eta < 2$ requirement has been applied on the $t\bar{t}$ and $\tilde{g} \rightarrow t\bar{t}\tilde{\chi}_1^0$ MC samples for the electron case.	39
5.14	The electron and muon real efficiencies as a function of p_T . The black dots stands for data, the red squares and blue triangles indicate $Z \rightarrow \ell\ell$ tag-and-probe and truth matching, respectively. The $t\bar{t}$ and $\tilde{g} \rightarrow t\bar{t}\tilde{\chi}_1^0$ are represented by magenta diamonds and yellow crosses, respectively. All of the real lepton efficiencies computed by MC processes are agree with the data one when $p_T > 40\text{GeV}$. The differences in the $p_T < 40\text{GeV}$ region come from the different event topologies.	40
5.15	The real electron efficiencies as a function of p_T in 3 different $ \eta $ regions. Four different trigger strategies are applied. The nominal values of the real electron efficiencies are measured using the single lepton trigger with tag trigger matched. The differences between the nominal values and the values measured using other strategies are assigned as the systematic uncertainties.	45
5.16	The real muon efficiencies as a function of p_T in 4 different $ \eta $ regions. Four different trigger strategies are applied. The nominal values of the real muon efficiencies are measured using the single lepton trigger with tag trigger matched. The differences between the nominal values and the values measured using other strategies are assigned as the systematic uncertainties.	46
5.17	The kinematic distributions of the candidate lepton from the processes considered for the systematic uncertainty study. The top row shows the p_T distributions for the candidate electrons on the left hand side and muons on the right hand side. The bottom row shows the $ \eta $ distributions. The $\tilde{g} \rightarrow t\bar{t}\tilde{\chi}_1^0$ process is more boosted and centralized than the $Z \rightarrow \ell\ell$ and $t\bar{t}$ processes.	50
5.18	The $\Delta R(\ell, jet)$ and the N_{jets} distributions of the candidate leptons extracted from $Z \rightarrow \ell\ell$, $t\bar{t}$, and $\tilde{g} \rightarrow t\bar{t}\tilde{\chi}_1^0$ processes. Most of the statistics of $\tilde{g} \rightarrow t\bar{t}\tilde{\chi}_1^0$ are located in $\Delta R(\ell, jet) < 1$ region and higher N_{jets} region. In the contrat, the statistics of $Z \rightarrow \ell\ell$ processes are populated at higher $\Delta R(\ell, jet) < 1$ region and lower N_{jets} region.	51

Abstract

Well now this is my abstract...

Introduction

Chapter 1

Theoretical Background

Chapter 2

The LHC and the ATLAS Experiment

Chapter 3

The Region of Interest Builder

Chapter 4

Object Reconstruction

Chapter 5

Data-driven techniques to estimate fake lepton backgrounds

5.1 The problem of fakes

The reconstructed objects (leptons, photons, b jets, etc.) in a collision event are used to perform a wide range of SM measurements or searches for evidence of BSM physics. The assumption is that these objects are ‘real’ representing the desired particles in the final state used in the analysis. In practice, the reconstructed objects might not be always ‘real’. In fact, they may be something completely different that was mistakenly reconstructed as the desired object, called ‘fake’. For the purpose of the analysis presented in this thesis, the focus is on ‘fake’ leptons. To illustrate the problem, a hadronic jet may deposit more energy in the electromagnetic calorimeter than the hadronic calorimeter, or that it leaves a narrow deposit of energy leading the reconstruction algorithms to mistake this jet for an electron. From the analysis point of view, the ‘fake’ electron will pass all the selection criteria and will be indistinguishable from a ‘real’ electron. It is important for the analysis that requires a reconstructed electron to model the fake electron background to get a sound result. This example was given with electrons, but can be generalized to muons as well. In short, any analysis that uses leptons in the final state must account for the ‘fake’ lepton background. This background can be more or less important depending on the detector, the analysis selection,

and the number of leptons required. To estimate this background it is important to first understand what type of processes lead to fake leptons.

5.2 Common processes for faking leptons

The reconstruction of ‘fake’ leptons can be an instrumental effect related to the inability to identify the object based on its measured properties by the detector. In this case, the reconstructed lepton is not a real lepton and the production process will be different for electrons and muons.

The reconstruction of electrons relies on the observation of well aligned particle hits in the layers of the ID that are consistent with an energy deposition in the EM calorimeter. Photons can mimick this signature since they deposit energy in the EM calorimeter that happens to be aligned with a charged track. A jet for example containing charged and neutral pions can lead to such scenarios. It is possible for the jet to have one charged pion leaving a track similar to that of an electron. The decay of π^0 mesons to photons in this jet can deposit energy in the EM calorimeter leading to the required signature. Another mechanism that can lead to fake electrons is the emission of photons via Brehmstrahlung from high energy muons. The muon track can be mistaken for that of an electron and the photons interact with the EM calorimeter leading to a signature similar to that of electrons. An additional process is that of photon conversions into a e^+e^- .

The reconstruction of muons relies on the observation of tracks from the ID matched to tracks from the muon spectrometer. It is possible for charged hadrons

with long lifetime to traverse the calorimeter layers and leave hits in the muon spectrometer. These hits may coincide with other hits from the ID due to the random activity in the event. As a result, a muon can get reconstructed. Another instance may occur when pions or kaons decay in-flight to muons in the muon spectrometer and happen to align with the primary vertex.

The leptons that are used in the physics analyses must be coming from the hard scatter, generally referred to as prompt leptons. There is another case where the reconstructed lepton is a real lepton but is not a lepton coming from the hard interaction, referred to as non-prompt leptons. Non-prompt leptons can be produced from heavy flavor meson decays with a low energy activity around the lepton which allows it to pass isolation requirements. A good example of this type of process is the semi-leptonic decay of top quark pair which contribute to final states with two leptons.

For the rest of the thesis, the fake leptons will be referred to as fake/non-prompt (FNP) leptons. There are several methods used to perform the estimation of FNP lepton backgrounds. A method that the author developed will be described next along with a standard method for estimating this type of backgrounds. The benefit of having two methods for estimating the FNP lepton background is to have enough confidence in the final estimate. The two methods use different assumptions which naturally leads to a more robust estimation of this difficult background. Moreover, the final estimate of the FNP lepton background is taken as a statistical combination of the estimates from the two methods leading to a reduction of the systematic uncertainties on the estimate.

5.3 Monte Carlo Template Method

5.3.1 Motivation

The processes leading to FNP leptons depend on the selection applied in the analysis. For instance, a selection with same-sign leptons will have contributions from top quark pair production ($t\bar{t}$) or the associated production of a vector boson and jets (W +jets or Z +jets). These processes cannot give two leptons of the same electric charge unless there is a charge mis-measurement (mainly affecting electrons) or that a FNP lepton was produced. It is possible to generate the processes that can contribute to a FNP lepton, such as $t\bar{t}$ or V +jets, with Monte Carlo event generators processed through Geant4 detector simulation of the ATLAS detector. This approach will yield an estimate however it might not be reliable. For instance, the detector simulation itself might not reproduce the true behavior of the interaction of the physics objects with the detector, particularly when looking at rare processes such as the production of FNP leptons. The second limitation is in the generation of enough MC events to probe the region of the phase space targetted by the analysis which affects the statistical uncertainties in the estimates. The latter concern is addressed by ensuring that the simulations for the major backgrounds ($t\bar{t}$ and V +jets) have much higher event count than the corresponding number of events observed in the data sample. In fact, these backgrounds have a large number of simulated events because they are important for many analyses (including SM measurements and BSM searches). The rest of the section will concentrate on addressing the former limitation.

low statistics of the tri-lepton samples and the variety of sources of fake leptons. It is difficult to rely on data-driven (including ME-based) methods that rely on inverting the isolation requirement given we deal with multi-jet final states. Inversion of the isolation requirement produces a sample with much more energetic jets than in the sample with the un-inverted isolation.

” picked variables that offer good separation between processed with prompt and “fake” leptons. It is nice that the corrections are close to one but they do not have to be. For example, their deviations from unity depend on the decay modes of heavy flavor hadrons that are implemented in simulations.

Is the showering of the samples used to measure the correction of the fake rate and the that of the samples it is applied to the same? if not, is there a potential systematic difference that has to be taken into account?

We used Herwig for showering of all the samples that require corrections to the fake rates.

5.3.2 Description of the method

The MC template method relies on the correct modelling of FNP leptons kinematics in MC simulation to extrapolate background predictions from control regions to the signal regions. The method assumes that the kinematic shapes for each source of FNP lepton is correctly modeled in the simulations, and the normalization for each source is extracted in a combined fit to data control regions. The number of normalization factors depend on the number of identified origins of the FNP

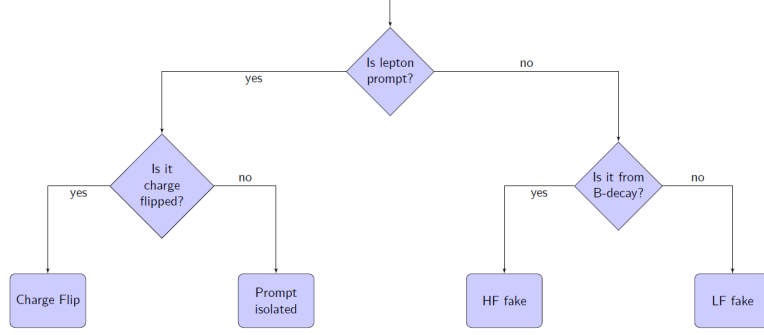


Figure 5.1: Lepton classification.

lepton in the signal regions and the control regions are designed to constrain these factors in regions enriched with FNP leptons from the same origin.

To illustrate the approach, we describe the application of the method in SS/3L analysis later described in this thesis. The processes of interest that may lead to a FNP lepton or a charge flip are $t\bar{t}$ and V +jets. FNP leptons are classified using an algorithm that navigates the generator particle record to determine where the FNP lepton is originating from. The lepton is classified as either an electron or a muon that is prompt from decays of on-shell W and Z bosons, non-prompt from a heavy flavor b decay (HF), or fake from mis-identification of a light flavor jet or a photon (LF). In the case of an electron, we further classify the prompt electrons to prompt electrons with the correct charge or with a charge mis-measurement, commonly named charge flip. In total, five categories referred to as MC templates are constructed following the classification illustrated in figure 5.1.

5.3.3 Correction factors

The FNP estimate relies on kinematic extrapolation using processes expected to contribute via FNP leptons from control regions with low jet multiplicity and E_T^{miss} , to the signal regions that require high jet multiplicity and E_T^{miss} . The control regions are chosen to separate FNP leptons from HF origins and FNP leptons from LF origins. For instance, a control sample characterized by the presence of a b -jet will be enriched in processes with one FNP lepton that is coming from a HF decay, while a sample characterized by the absence of a b -jet will have one FNP lepton from LF decay. The presence of one FNP lepton in the control sample allows the correction of the production rate of these FNP leptons by performing a fit to data.

For example, if a $Z \rightarrow \mu\mu + \text{LF jet}$ event is reconstructed as a $\mu^+\mu^-e^+$ event, then the electron is fake. Therefore, a correction of $\text{LF jet} \rightarrow e$ ($\text{Fr}(\text{LF} \rightarrow e)$) is applied to the rate of $\mu\mu e$ events. The correction $\text{Fr}(\text{LF} \rightarrow e)$ is constrained by a fit to data in control regions dominated by $\text{LF jet} \rightarrow e$ type fakes. Similarly, three other corrections are defined as $\text{LF jet} \rightarrow \mu$ ($\text{Fr}(\text{LF} \rightarrow \mu)$), $\text{HF jet} \rightarrow e$ ($\text{Fr}(\text{HF} \rightarrow e)$), $\text{HF jet} \rightarrow \mu$ ($\text{Fr}(\text{HF} \rightarrow \mu)$). An additional correction is applied to correct the charge flip rate predicted by simulation. For example, a $Z \rightarrow e^+e^-$ event is reconstructed as e^+e^+ or e^-e^- . The simulation takes into account the charge flip rate but it might be off. The charge flip ($\text{Cf}(e)$) correction derived from a data fit is expected to recover this mis-modeling. The charge flip rate only concern electrons as the muon charge flip rate is negligible.

A likelihood fit is defined as the product of the Poisson probabilities describing the observed events in the binned distributions from the expected number of events rescaled by the five multipliers which are left free to float in the fit. These multipliers are applied to the MC predictions in the signal regions to obtain an estimation of the charge flip and FNP backgrounds.

5.3.4 Control regions

The corrections depend on the simulated sample, the reconstructed final state, and the flavor of the leptons. As a result, care must be taken when designing the control regions used to perform the fit of the FNP leptons and electron charge flip templates. For instance, each template needs to be constrained in a selection that is representative of the processes leading to FNP leptons and charge flip electrons present in the kinematic region targetted by the search for BSM physics.

In the SS/3L analysis discussed in this thesis the control regions are defined with at least two same-sign leptons, $E_T^{\text{miss}} > 40$ GeV, two or more jets. This preselection ensures that the FNP leptons are not from fakes originating from QCD like event topologies. They are further split in regions with or without b -jets to constrain the HF and LF leptons respectively. In addition, they are also split with different flavours of the same-sign lepton pair ee , $e\mu$, and $\mu\mu$, giving a total of six control regions. Any event entering the signal region is vetoed. The ee channel will constrain the charge flip correction factor, fake leptons from LF decays in the selection without b -jets, and non-prompt decay from HF in the selection with b -jets. The $\mu\mu$ channel will constrain the muon fake rates in the

LF and HF decays for the selection without or with b -jets, respectively. The $e\mu$ channel will constrain both the electron and muon fakes for events containing both lepton flavors.

The six distributions are chosen for variables that provide the best separation between processes with prompt leptons and processes with FNP leptons and charge flip and are shown before and after the fit in Figures 5.2-5.4 and Figures 5.3-5.5, respectively.

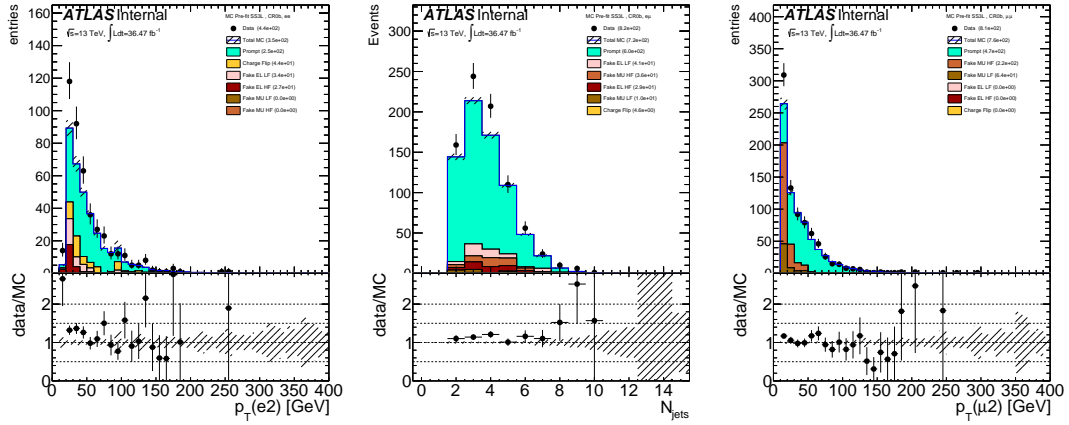


Figure 5.2: Pre-fit distributions for ee channel (left), for $e\mu$ channel (middle), and for $\mu\mu$ channel (right) from CR0b that were used in the fit to extract the FNP lepton and charge flip multipliers. The generator used in these plots is Powheg. The hashed band represents the sum of systematic uncertainties on the predictions.

The minimization of the negative log likelihood using the MINUIT package leads to the multipliers shown in Tables 5.1 and 5.2. The tables represent the multipliers obtained from the fit upon using two different parton showers, POWHEG-BOX and SHERPA for the processes that lead to FNP leptons and charge flips. The

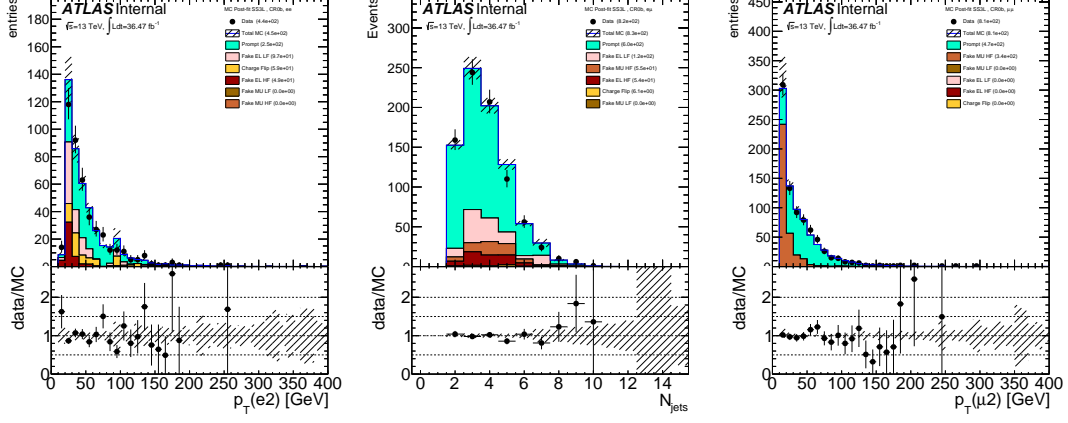


Figure 5.3: Post-fit distributions for ee channel (left), for $e\mu$ channel (middle), and for $\mu\mu$ channel (right) from CR0b that were used in the fit to extract the FNP lepton and charge flip multipliers. The generator used in these plots is Powheg. The hashed band represents the sum of systematic uncertainties on the predictions.

systematic uncertainty is obtained by varying the generator from POWHEG-Box to SHERPA and evaluating the impact on the expected background from FNP and charge flip leptons. It is found to be the dominant contribution to the systematic uncertainty of the method (up to 80%). The uncertainties in the multipliers themselves correspond to how much the parameter needs to be varied for a one standard deviation change in the likelihood function. This uncertainty takes into account the limited number of simulated events and is included as a systematic uncertainty on the expected number of background events.

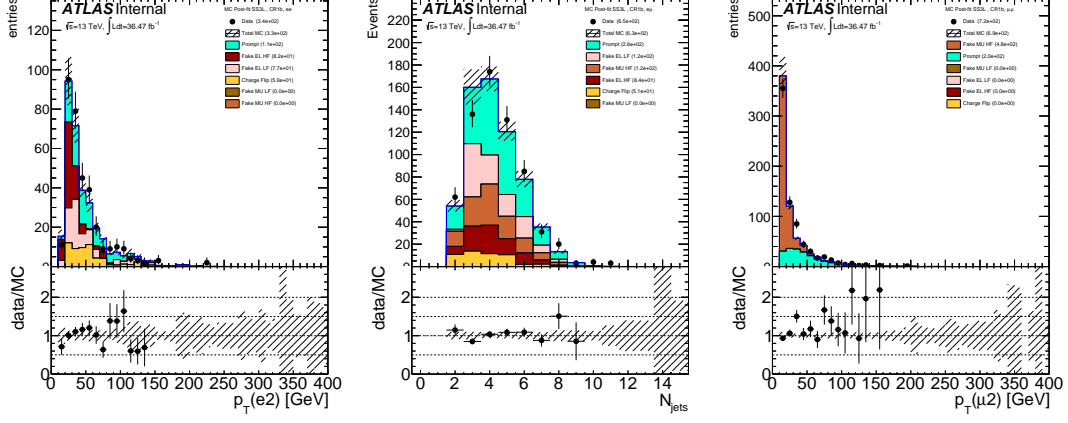


Figure 5.5: Post-fit distributions for ee channel (left), for $e\mu$ channel (middle), and for $\mu\mu$ channel (right) from CR1b that were used in the fit to extract the FNP lepton and charge flip multipliers. The generator used in these plots is Powheg. The hashed band represents the sum of systematic uncertainties on the predictions.

with events containing one object. Then a generalized treatment that can handle events with an arbitrary number of leptons in the final states will be discussed.

5.4.1 Events with one object

Given the probabilities ε/ζ for a real/FNP candidate lepton to satisfy in addition the signal lepton criteria, one can relate the number of events with one candidate lepton passing/failing signal requirements ($n_{\text{pass}}/n_{\text{fail}}$) to the number of events with one real/FNP signal leptons ($n_{\text{real}}/n_{\text{FNP}}$):

$$\begin{pmatrix} n_{\text{pass}} \\ n_{\text{fail}} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ \frac{1-\varepsilon}{\varepsilon} & \frac{1-\zeta}{\zeta} \end{pmatrix} \begin{pmatrix} n_{\text{real}} \\ n_{\text{FNP}} \end{pmatrix}; \quad (5.1)$$

Table 5.1: The FNP and charge flip multipliers obtained after minimizing the likelihood function using Pythia. The uncertainty in the multipliers takes into account the limited statistics of simulated events.

Category	Multiplier	Uncertainty
chFlip	1.49	0.58
HF EL	2.80	0.98
LF EL	2.89	0.88
HF MU	1.59	0.31
LF MU	1.00	1.34

allowing to determine the unknown number of events n_{FNP} from the observed n_{pass} and n_{fail} .

5.4.2 Dynamic matrix method

This readily generalizes to events with more than one lepton, with a formalism described in details. An implementation of this formalism (known also as the dynamic or generalized matrix method) can handle an arbitrary number of leptons in the event. The method should be applied event-by-event, effectively resulting into a weight being assigned to each event. The predicted yield of events with FNP leptons is simply the sum of weights.

The application of the matrix method to multilepton final states comes with two specificities. Firstly, contributions of events with charge-flip electrons would

Table 5.2: The FNP and charge flip multipliers obtained after minimizing the likelihood function using Sherpa. The uncertainty in the multipliers takes into account the limited statistics of simulated events.

Category	Multiplier	Uncertainty
chFlip	1.34	0.58
HF EL	2.40	0.85
LF EL	1.83	1.04
HF MU	1.17	0.16
LF MU	2.40	0.81

bias a straightforward matrix method estimate (in particular for a final state formed by two leptons with same electric charge). This happens because the candidate-to-signal efficiency for such electrons is typically lower than for real electrons having a correctly-assigned charge. One therefore needs to subtract from n_{pass} and n_{fail} the estimated contributions from charge-flip. This can be performed by including as well events with pairs of opposite-sign candidate leptons in the matrix method estimate, but assigning them an extra weight corresponding to the charge-flip weight. Thanks (again) to the linearity of the matrix method with respect to n_{pass} and n_{fail} , this weight-based procedure is completely equivalent (but more practical) to the aforementioned subtraction.

Secondly, the analytic expression of the matrix method event weight depends on the lepton multiplicity of the final state. This concerns events with three or

more candidate leptons: one such event takes part both in the evaluation of the FNP lepton background for a selection with two signal leptons or a selection with three signal leptons, but with different weights². Therefore, for a given event used as input to the matrix method, one should consider all possible leptons combinations, each with its own weight and its own set of kinematic variables. For example, a $e^+e^-\mu^+$ event is comptabilized in the background estimate both as an $e^+\mu^+$ event (with a weight w_1) and as an $e^+e^-\mu^+$ event (with a weight $w_2 \neq w_1$).

5.4.3 Propagation of uncertainties

The two parameters (ε and ζ respectively) can be measured in data, and depend on the flavor and kinematics of the involved leptons. Systematic uncertainties resulting from the measurement of these two parameters, and their extrapolation to the signal regions, can be propagated to uncertainties on the event weight through standard first-order approximations. The different sources of uncertainties should be tracked separately so that correlations of uncertainties across different events can be accounted for correctly. The resulting set of uncertainties on the cumulated event weights can be then added in quadrature to form the systematic uncertainty on the predicted FNP lepton background yield. The corresponding statistical uncertainty can be taken as the RMS of the event weights.

²This can appear for inclusive selections: for example an event with two signal leptons may or not contain additional candidate leptons, in a transparent way

5.5 Real lepton efficiency

The real lepton efficiency parameter should be measured in a region dominated by real leptons, usually with the same topology as the real leptons dominating the signal regions. It can be a control region dominated by $Z \rightarrow \ell\ell$ or “ $W \rightarrow \ell\nu$ ” or $t\bar{t}$, etc. events. This section is dedicated to the real lepton efficiency measurement using the Tag and Probe method. The real efficiency is obtained by taking the ratio between the number of probe leptons passing the signal requirement (n_{signal}) and the number of probe candidate leptons ($n_{\text{candidate}}$). The estimated number of the background events ($N_{\text{candidate}}^{bkg}$) associated with the low p_T leptons (as low as 20 GeV) are considered and estimated using a background template method. Equation 5.3 shows the real lepton efficiency calculation.

$$\epsilon = \frac{N_{\text{signal}}}{N_{\text{candidate}} - N_{\text{candidate}}^{bkg}} \quad (5.2)$$

5.5.1 The Z tag-and-probe method

The leptons used for the efficiency measurements are extracted from data using the Z tag-and-probe method. Events are required to have at least two candidate leptons. One of the two leptons, the tag lepton, is required to have a transverse momentum $p_T > 25\text{GeV}$ and pass the signal lepton requirements specified in Table ???. In the case of the $Z \rightarrow ee$ events, an additional $|\eta| < 2$ requirement is used for the tag and probe selection. The probe lepton, which is used for the real efficiency measurements, is required to pass the candidate lepton requirements and have

to carry opposite charge and same flavor with respect to the corresponding tag lepton. This requirement ensures the selection of real leptons to allow the real lepton efficiency measurement. The tag and the probe leptons are required to match the single lepton triggers or dilepton triggers listed in Table 5.3.

Trigger	lepton	2015	2016
Single lepton trigger	electron	e24_lhmedium_iloose_L1EM20VH	e26_lhtight_nod0_ivarloose
	muon	mu20_iloose_L1MU15	mu26_ivarmedium
Dilepton trigger	electron	2e12_lhloose_L12EM10VH	2e17_lhvloose_nod0
	muon	mu18_mu8noL1	mu22_mu8noL1

Table 5.3: The list of single lepton and dilepton triggers used for the real lepton efficiency measurements. The dilepton triggers are used for studying the systematic uncertainties causing by the trigger.

All possible combinations of the tag-and-probe pairs are considered to avoid any bias and to increase the statistics. The invariant mass of the tag-and-probe pair system should satisfy $80 < m_{\ell\ell} < 100\text{GeV}$. The data-to-MC comparison of the tag-and-probe pair invariant mass distributions are shown in Figure 5.6, and indicates the need of subtracting the background, especially when the probe electron has $p_T < 20\text{GeV}$.

The background contamination associated with these low p_T electrons is estimated using a background template method similar to the one used by the e/γ performance group for their efficiency measurements [?]. Because the real

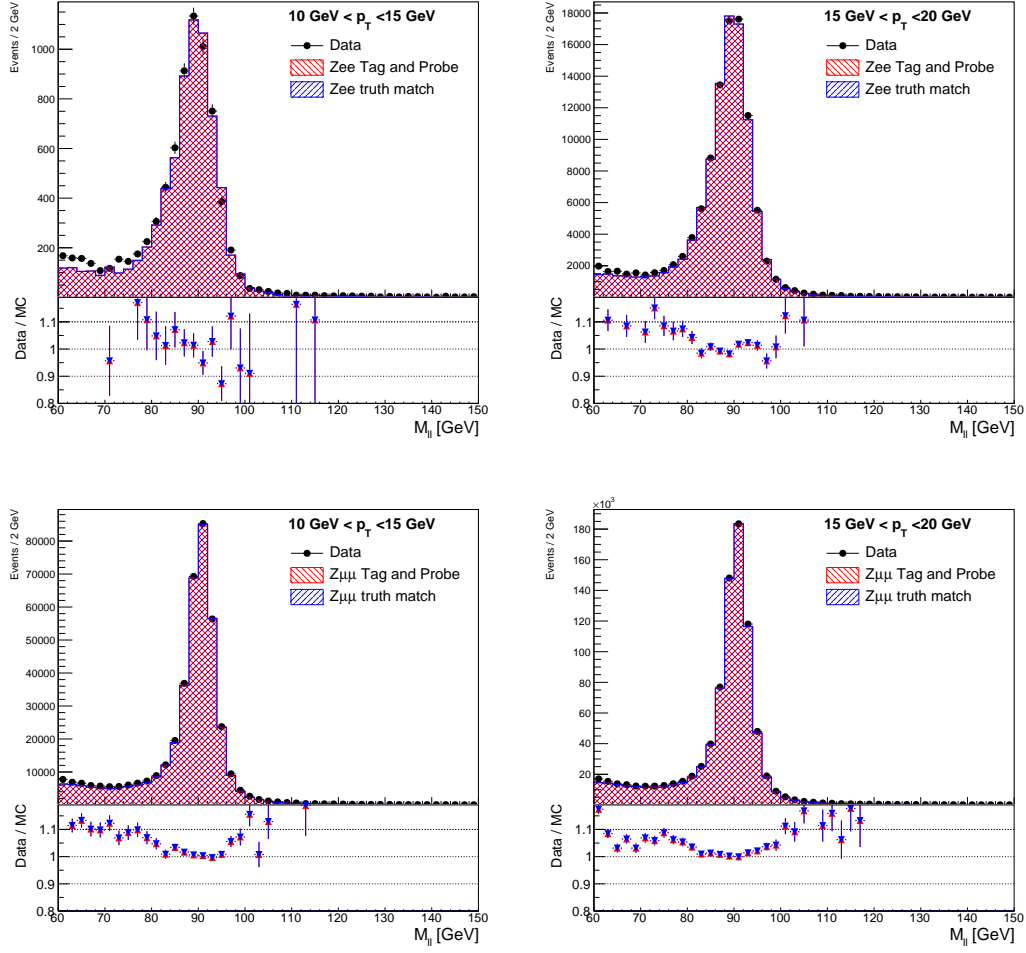


Figure 5.6: The distributions of the tag-and-probe pair invariant mass ($m_{\ell\ell}$) computed using $Z + jets$ MC and 2015 + 2016 data. The blue color represents the Z truth matched events, the red color stands for the Z tag-and-probe events, and the black dots are data. The top row shows the m_{ee} computed with probe electrons in the $10 < p_T < 15\text{GeV}$ (left) and $15 < p_T < 20\text{GeV}$ (right) intervals. The bottom row shows the $m_{\mu\mu}$ cases. The MC distributions are scaled to the data using a Gaussian fit of the Z mass peak ($85 < m_{\ell\ell} < 95\text{GeV}$).

lepton efficiency is obtained by computing the ratio between the number of probe leptons passing the signal requirements (N_{signal}) and the number of probe leptons passing the candidate requirements ($N_{\text{candidate}}$), the estimated background contamination ($N_{\text{candidate}}^{bkg}$) in the candidate probe leptons needs to be subtracted. As the background contamination is found to be negligible for signal leptons, no background subtraction is performed at the numerator. Equation 5.3 shows the real lepton efficiency calculation.

$$\epsilon = \frac{N_{\text{signal}}}{N_{\text{candidate}} - N_{\text{candidate}}^{bkg}} \quad (5.3)$$

5.5.2 Background subtraction

The background contamination has been evaluated on data using a background template method. A sample enriched in background is obtained by inverting the calorimeter isolation and track isolation cuts and requesting the electron object to fail the medium LH identification. In order to asses a systematic on the background template definition, three variations of background template definitions are considered. Table 5.4 summarizes the background template definitions and the Figure 5.7 shows the background template m_{ee} distributions computed with probe electrons in $10 < p_T < 15\text{GeV}$ and $15 < p_T < 20\text{GeV}$, respectively. The invariant mass distribution of the template events (m_{ee}^{template}) is then used to estimate the amount of background in the measurement region ($80 < m_{\ell\ell} < 100\text{GeV}$).

As the template cuts remove background objects, the background template has to be normalized to provide the correct background estimation. The $120 < m_{ee} <$

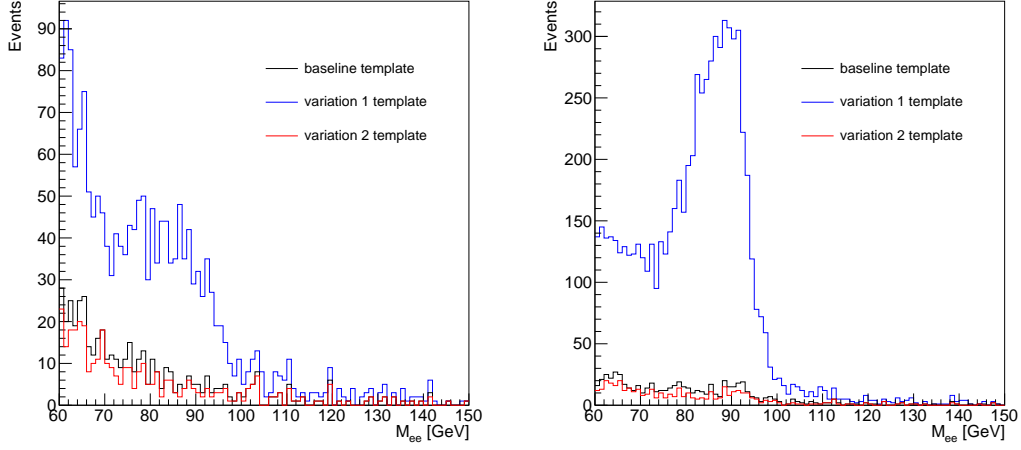


Figure 5.7: The m_{ee} distributions of the three different background templates. The m_{ee} distributions in the left and right plots are computed using probe electrons with $10 < p_T < 15$ GeV and $15 < p_T < 20$ GeV, respectively. Because the variation 1 template has looser calorimeter and track isolation requirements, we see a peak in the Z mass region. Compared with the variation 1 template, the candidate and the variation 2 templates have tighter cuts so no peak can be found in the Z mass region.

cut	variation 1 template	candidate template	variation 2 template
Identification	-	fail medium LH	fail medium LH
Calorimeter isolation	$E_T^{topocone20}/p_T > 6\%$	$E_T^{topocone20}/p_T > 15\%$	$E_T^{topocone20}/p_T > 20\%$
Track isolation	$p_T^{varcone20}/p_T > 6\%$	$E_T^{topocone20}/p_T > 8\%$	$E_T^{topocone20}/p_T > 15\%$

Table 5.4: The definition of the background templates used to estimate the background contamination associated with the Z tag-and-probe method. The candidate template is used to estimate the background contamination and the variation 1 and 2 templates, which have looser and tighter requirements, are used to assess the systematic caused by the background contamination.

150GeV region is used for this normalization as a smaller prompt electron contribution is expected in this region. The background in this tail region is estimated by integrating the candidate m_{ee} distribution in the tail region after subtracting the prompt electron contamination which is estimated by integrating the m_{ee} distribution in the tail region using the $Z \rightarrow ee$ MC simulation. Equation 5.4 shows the estimation of the number of background events in the tail region using the candidate electrons.

$$N_{bkg}^{\text{tail}} = N_{\text{candidate}}^{\text{tail}} - N_{\text{MC, prompt}}^{\text{tail}} \quad (5.4)$$

The candidate electron selection criteria already provides a relatively pure sample of prompt electrons. Therefore, the background template suffers from low

statistics in the tail region. To avoid any bias in the normalization factor due to statistical fluctuations, the template is fitted using an exponential function. The fitting range is $60 < m_{ee}^{\text{template}} < 120\text{GeV}$. In order to minimize the prompt lepton contamination arising from $Z \rightarrow ee$ events, the $80 < m_{ee}^{\text{template}} < 100\text{GeV}$ region is excluded. Because of the low statistics in the tail, the fit is mostly driven by the $60 < m_{ee}^{\text{template}} < 80\text{GeV}$. After applying the fit, the template in the tail region, $N_{\text{template}}^{\text{tail}}$, is normalized to the background in the tail, N_{bkg}^{tail} , to get the correct estimated number of background. Figure 5.8 shows the candidate m_{ee} distributions before and after applying the background subtraction using the background template. The data after applying the background subtraction is compared to the MC simulations and the background template distribution and their corresponding fit are also shown. The simulated m_{ee} distribution of $Z \rightarrow ee$ MC are normalized to the data after background subtraction using a Gaussian fit in $85 < m_{ee} < 95\text{GeV}$ region. The top and the bottom rows correspond to the $10 < p_T < 15\text{GeV}$ and $15 < p_T < 20\text{GeV}$, respectively. The left, middle, and right columns correspond to the $0 < |\eta| < 0.8$, $0.8 < |\eta| < 1.37$, and $1.52 < |\eta| < 2.0$ regions, respectively. After applying the background subtraction, the data agree with the MC simulation within the statistical uncertainties.

Then, the background contamination in the Z mass region ($80 < m_{ee} < 100\text{GeV}$) is calculated using:

$$N_{bkg}^{80 < m_{ee} < 100\text{GeV}} = \int_{80}^{100} N_{\text{template}} dm_{ee} \cdot \frac{N_{bkg}^{\text{tail}}}{N_{\text{template}}^{\text{tail}}} \quad (5.5)$$

The background estimations are summarized in Table 5.5. The largest improve-

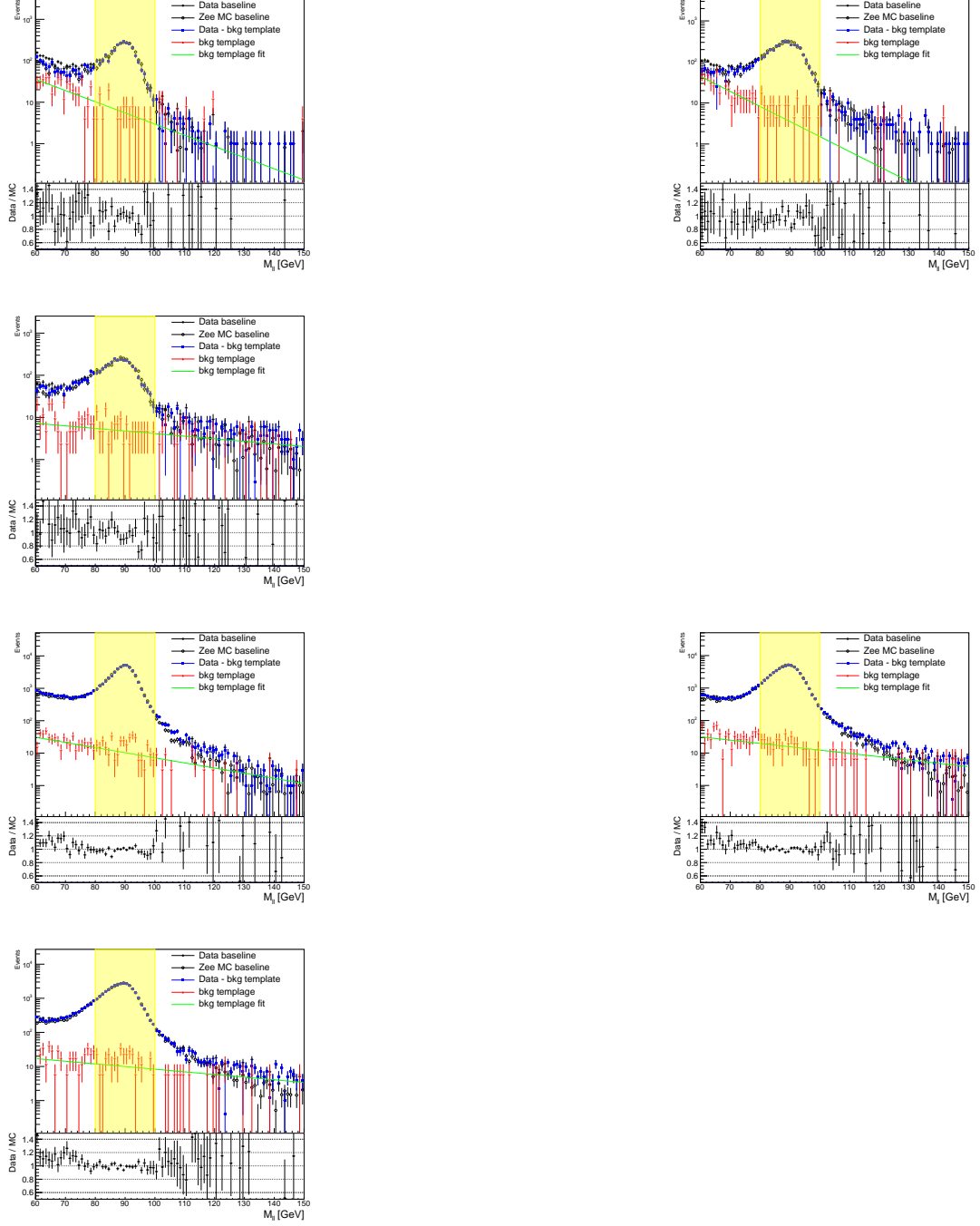


Figure 5.8: Plots illustrate the background subtraction procedure.

The top row represents the $10 < p_T < 15\text{GeV}$ region and the the bottom row shows the $15 < p_T < 20\text{GeV}$ region. The three columns, from left to right, are the $0 < |\eta| < 0.8$, $0.8 < |\eta| < 1.37$, and $1.52 < |\eta| < 2.0$ regions, respectively. The m_{ee} distributions for data before the background subtraction (full black dots) and after the background subtraction (blue squares) are shown along with the

ments are observed in the $10 < p_T < 15\text{GeV}$ range which a sizeable background contamination is subtracted. The background contamination in $15 < p_T < 20\text{GeV}$ is relatively small and indicates the Z tag-and-probe method provides high purity sample of prompt leptons. The real electron efficiencies before and after applying the background subtraction are shown in Table 5.6.

	$0 < \eta < 0.8$	$0.8 < \eta < 1.37$	$1.52 < \eta < 2.0$
$10 < p_T < 15\text{GeV}$	4.04%	2.10%	3.17%
$15 < p_T < 20\text{GeV}$	0.44%	0.58%	0.76%

Table 5.5: The estimated background contamination using the background template. The p_T and $|\eta|$ binnings correspond to the one used for the final measurements.

	background subtraction	$0 < \eta < 0.8$	$0.8 < \eta < 1.37$	$1.52 < \eta < 2.0$
$10 < p_T < 15\text{GeV}$	before	57.4 ± 0.9	66.6 ± 0.8	53.2 ± 0.9
	after	59.9 ± 1.9	68.0 ± 1.8	55.0 ± 1.7
$15 < p_T < 20\text{GeV}$	before	64.5 ± 0.2	69.4 ± 0.2	62.0 ± 0.3
	after	64.8 ± 0.5	69.8 ± 0.5	62.5 ± 0.6

Table 5.6: The real electron efficiencies in percentage before and after applying the background subtraction. The efficiency in three different $|\eta|$ regions for $10 < p_T < 15\text{GeV}$ and $15 < p_T < 20\text{GeV}$ are shown.

5.5.3 Real lepton efficiencies

5.5.4 μ Cut efficiencies

The real lepton efficiency is defined as the ratio between the number of prompt leptons passing the signal lepton definitions and the number of prompt leptons passing the candidate lepton definitions (as defined in [?]). Thus, it measures the efficiency of the signal lepton definition cuts with respect to the candidate ones. The efficiencies associated to each signal cut with respect to candidate definitions are shown in Figure 5.9.

The left plot in Figure 5.9 shows that the prompt electron efficiency increases from $\sim 62\%$ in the $10 < p_T < 15\text{GeV}$ region to $\sim 98\%$ when $p_T > 80\text{GeV}$. The dominant contribution to the electron efficiency losses is the calorimeter isolation. The loose to medium likelihood (LH) cut efficiency increases from $\sim 92\%$ to $\sim 96\%$ in the $10 < p_T < 30\text{GeV}$ region then reaches a plateau for $30 < p_T < 50\text{GeV}$ and increases again to $\sim 98\%$ when $p_T > 60\text{GeV}$. The calorimeter isolation cut efficiency increases from $\sim 69\%$ in the $10 < p_T < 15\text{GeV}$ region to $\sim 98\%$ when $p_T > 70\text{GeV}$. The efficiency associated to the track isolation cut is $\sim 89\%$ for $10 < p_T < 15\text{GeV}$ and increases up to $\sim 100\%$ when $p_T > 60\text{GeV}$. The contribution of the longitudinal impact parameter cut to the real efficiency is $\sim 98\%$ for $10 < p_T < 15\text{GeV}$ and increases up to $\sim 100\%$ when $p_T > 15\text{GeV}$.

As the same muon identification is used for the candidate and the signal definitions, the muon efficiencies are much higher than the electron ones (Figure 5.9, right hand side). The associated efficiencies computed using $Z \rightarrow \mu\mu$ events

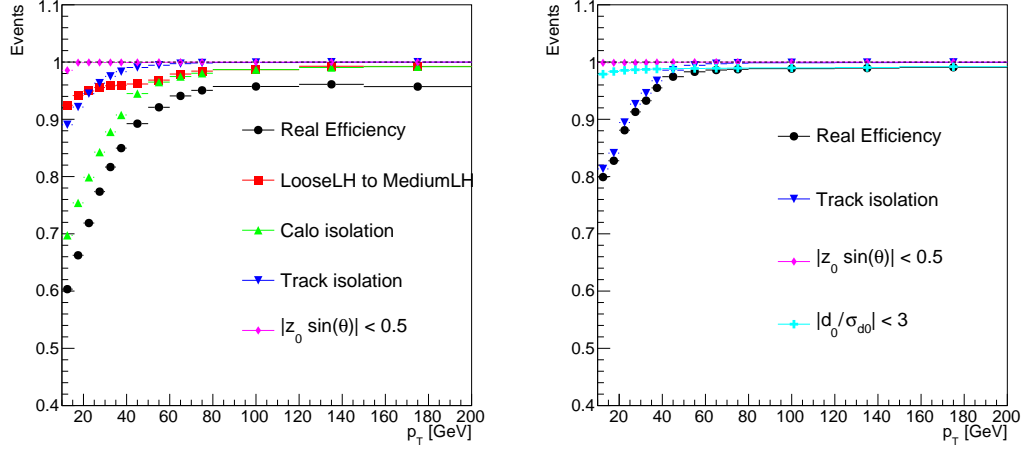


Figure 5.9: Efficiencies of the signal electron (left) and muon (right) definition cuts as a function of p_T . The leptons are selected from data samples using Z tag-and-probe method. The black points correspond to the total real lepton efficiency. The red squares represent the loose to medium likelihood cut efficiency. The green triangles stand for the calorimeter isolation cut efficiency and the blue triangles are the distribution of track isolation cut efficiency. The longitudinal and tranverse impact parameters cut efficiencies are presented by magenta diamonds and cyan crosses, respectively.

increase from $\sim 80\%$ for $10 < p_T < 15\text{GeV}$ to $\sim 98\%$ when $p_T > 50\text{GeV}$. The dominant contribution to the muon efficiency is the track isolation cut. The associated efficiency increases from $\sim 82\%$ to 98% when $p_T > 50\text{GeV}$. Another difference with the electron case is that a transverse impact parameter cut is used in the muon signal definitions while this cut is already applied at the candidate level in the electron case. The contributions of the longitudinal and transverse impact parameter cuts on muon efficiency are very small because the cut efficiency of transverse impact parameter is $\sim 99\%$ and the longitudinal one is 100% .

5.5.5 μ Real lepton efficiencies

Figure 5.10 shows the real lepton efficiencies as a function of p_T and $|\eta|$ used by the matrix method. The three $|\eta|$ binnings used for the electron case are driven by the electromagnetic calorimeter geometry. The background subtraction has been applied on the $10 < p_T < 15\text{GeV}$ and $15 < p_T < 20\text{GeV}$ for the electron case. Because the crack region, $1.37 < |\eta| < 1.52$, is not considered in the analysis, this region is removed from the real electron efficiency study. The uncertainties are the quadratic sum of the statistical uncertainties and the measurement systematic uncertainties. The electron efficiencies in $1.52 < |\eta| < 2.01$ are lower than the other two $|\eta|$ regions. This is expected as the electron identification is better in the central region of the calorimeter.

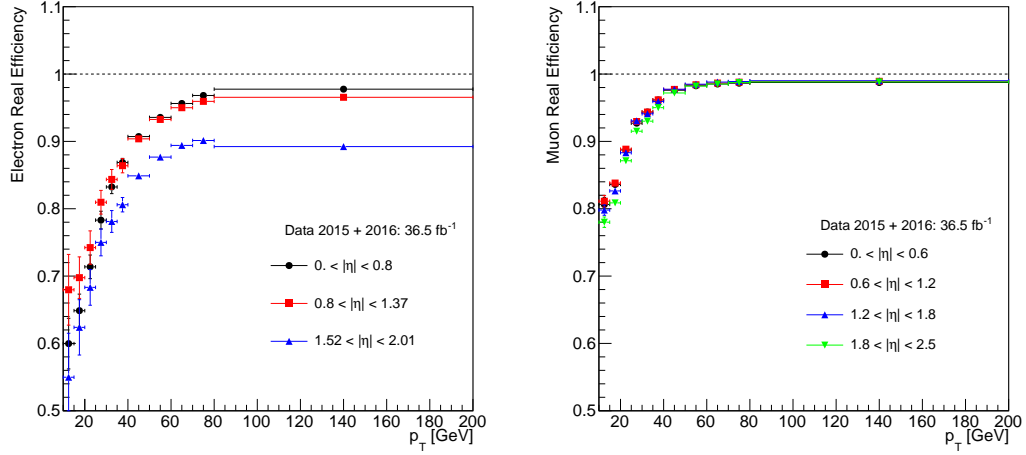


Figure 5.10: The real lepton efficiencies as a function of p_T and $|\eta|$ measured using the Z tag-and-probe method. The left plot corresponds to the real electron efficiencies in three $|\eta|$ regions and the right plot stands for the real muon efficiencies in four $|\eta|$ regions. The $|\eta|$ binning used for the real electron efficiencies measurement corresponds to the geometry of the electromagnetic calorimeter and removing the creak region. A homogeneous $|\eta|$ binning has been chosen for the muon case.

5.5.6 Z Tag-and-probe method and truth matching comparisons

Although the Z tag-and-probe method could select probe leptons accurately, we would like to verify the accuracy of the Z tag-and-probe method by comparing it with the truth matched information using the $Z \rightarrow \ell\ell$ MC samples. Figure 5.11 shows the real lepton efficiencies computed using Z tag-and-probe method and truth matching as a function of p_T , $|\eta|$, and $\Delta R(\ell, jet)$, respectively. The uncertainties shown in the plots correspond to the statistical uncertainties only. There are as large as 7% differences in the low p_T region for the real electron efficiencies as function of p_T computed by these two different methods. The differences decrease when p_T increases and no differences can be seen when $p_T > 50\text{GeV}$. The largest differences is about 3% in the real electron efficiencies as a function of $|\eta|$. Larger differences between these two methods can be seen in the real electron efficiencies as a function of $\Delta R(e, jet)$ when the $\Delta R(e, jet) < 0.4$ because the overlap removal requirement has been applied on the candidate electrons. The overlap removal removes the electrons so the $\Delta R(e, jet) < 0.4$ region lacks statistics. The efficiency differences between Z tag-and-probe method and truth matching are less than 1% for the muon case except the real muon efficiencies as a function of $\Delta R(\mu, jet)$ which has larger differences between these two methods when $\Delta R(\mu, jet) < 0.4$ also because of the overlap removal requirement. The comparisons showing small differences between methods indicate the probe leptons selected by the Z tag-and-probe method are reliable and the differences may be considered as the systematic uncertainties.

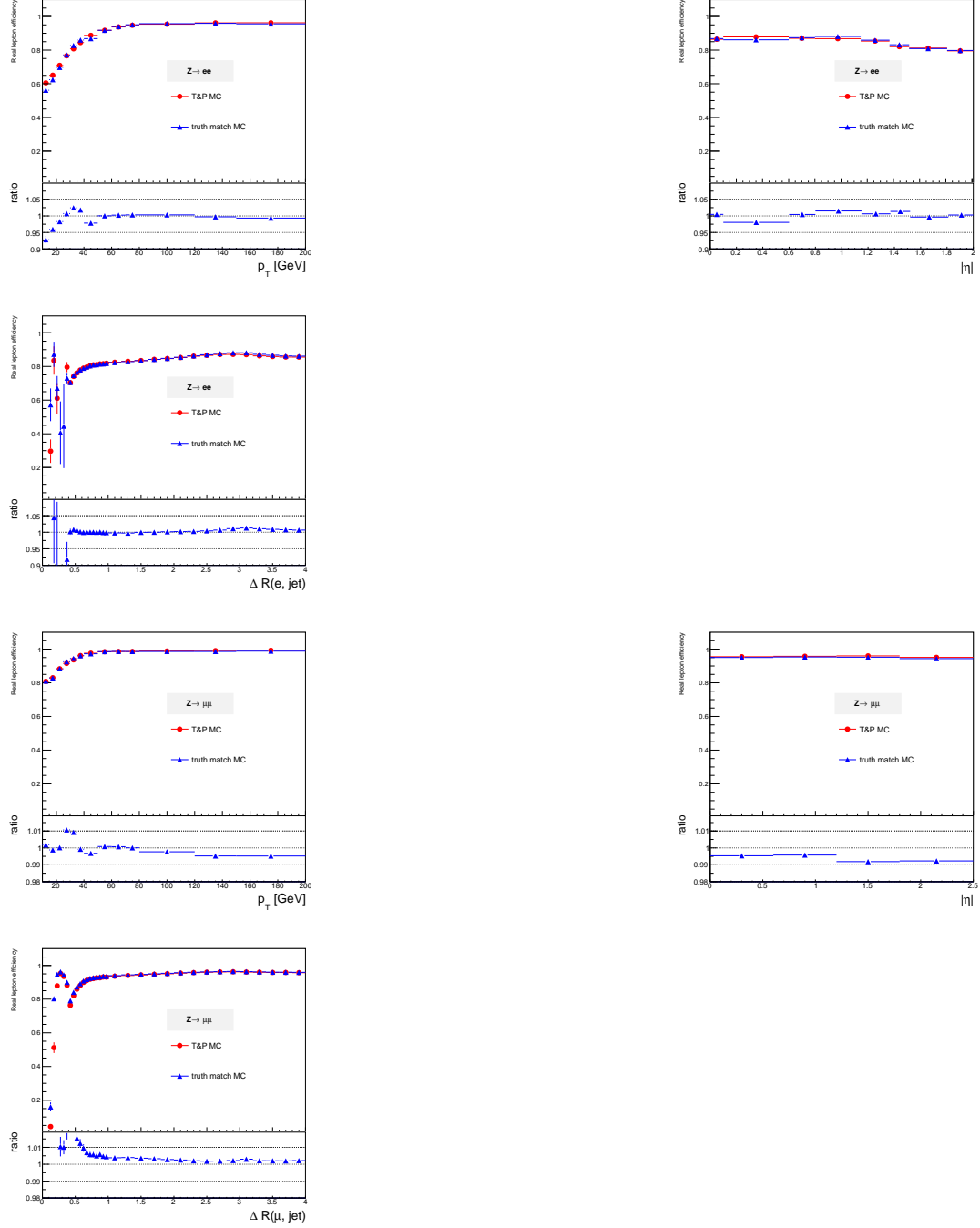


Figure 5.11: The real lepton efficiencies computed using Z tag-and-probe method and truth matching. The top row is the electron case and the bottom row is the muon case. The three columns from the left to the right are the real lepton efficiencies as a function p_T , $|\eta|$, and $\Delta R(\ell, jet)$, respectively. The real lepton efficiencies computed using Z tag-and-probe are denoted by the red dots and the one calculated using truth matching are indicated by the blue triangles. The lower

5.5.7 Data-to-MC comparisons

The real lepton efficiencies computed using the Z tag-and-probe method in data are compared to those using the simulated $Z \rightarrow \ell\ell$ MC processes. All 2015 and 2016 data are considered, and corresponds to an integrated luminosity of 36.5 fb^{-1} after good run list requirements. All the MC lepton scale factors provided by the CP group are applied and the simulation is reweighted to the pile-up observed in the data. Figure 5.12 shows the electron and the muon real efficiencies as a function of p_T , $|\eta|$ and $\Delta R(\ell, jet)$ measured on data and with simulated $Z \rightarrow ee$ and $Z \rightarrow \mu\mu$ MC processes, respectively. The associated uncertainties correspond to the statistical uncertainties only. A reasonable data to MC agreement is observed except the low $\Delta R(\ell, jet)$ region because of lacking statistics. The real electron efficiencies as a function of $|\eta|$ computed using the $Z \rightarrow ee$ MC are slightly lower than the one computed using data. The differences come from the efficiencies drop of $Z \rightarrow ee$ MC in the $40 < p_T < 50 \text{ GeV}$.

5.5.8 Real lepton efficiency versus pileup

The relations between the real lepton efficiencies and the pileup are also studied. The efficiencies calculated using 2015 + 2016 data are compared with those obtained using Z tag-and-probe method and truth matching MC samples. The $t\bar{t}$ and $\tilde{g} \rightarrow t\bar{t}\tilde{\chi}_1^0$ MC samples are also considered because we are interested in the behavior of the efficiencies with different event topologies. Figure 5.13 shows the electron and muon real efficiencies as a function of the average interactions per crossing $\langle \mu \rangle$. The real electron efficiencies are about 92% to 93% at low $\langle \mu \rangle$

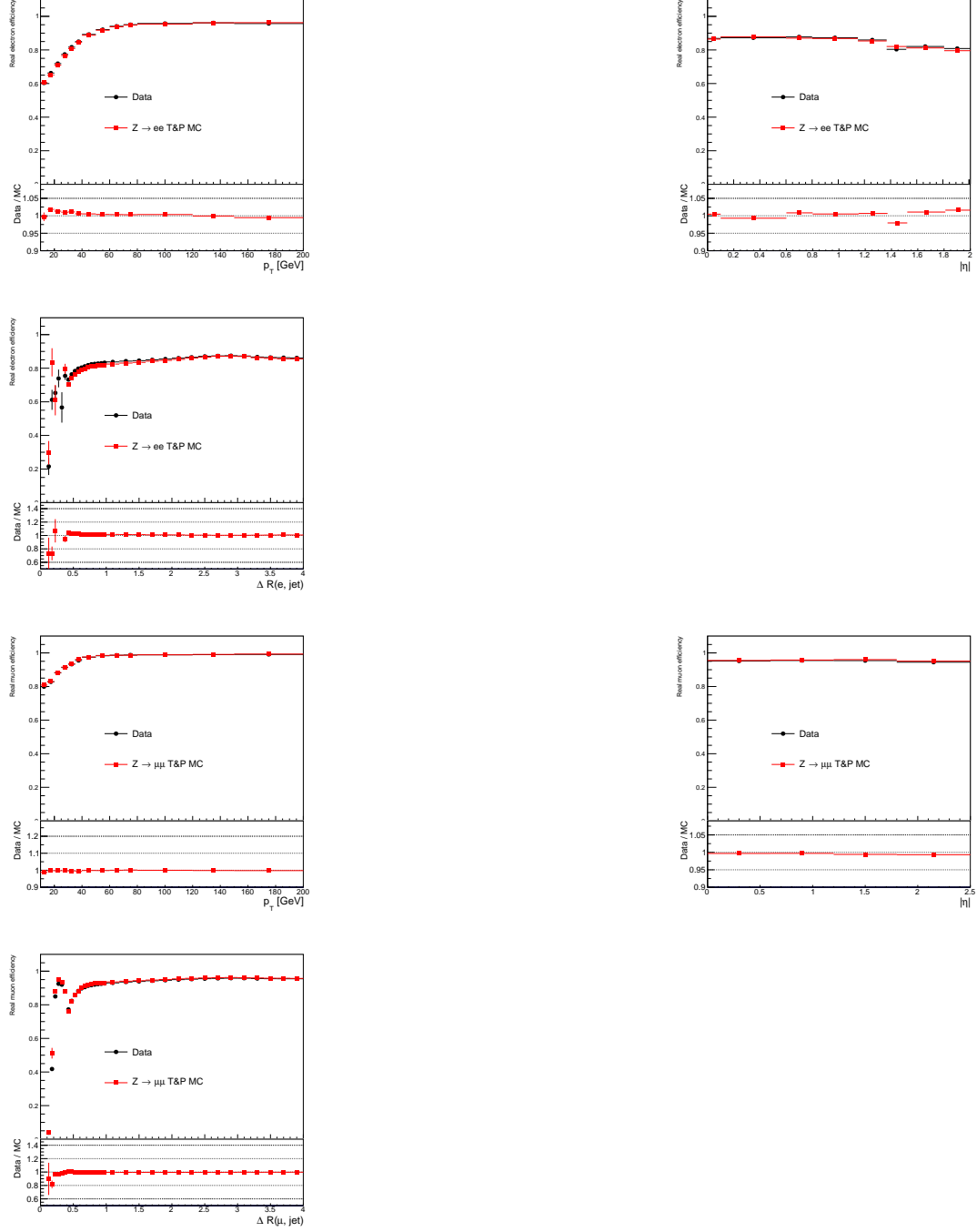


Figure 5.12: The real lepton efficiencies as a function of p_T , $|\eta|$ and $\Delta R(\ell, \text{jet})$ measured on data and MC using the Z tag-and-probe method. The plots on the top row correspond to the real electron efficiencies and the plots on the bottom row correspond to the real muon efficiencies. The 2015 + 2016 data are denoted by the black dots, the $Z \rightarrow \ell\ell$ MC are shown using red squares and are reweighted to the data pile-up. The uncertainties shown in the plots are corresponding

and decrease when $\langle \mu \rangle$ increase. Comparing with the real lepton efficiencies calculated by data samples, the efficiencies computed using $\tilde{g} \rightarrow t\bar{t}\tilde{\chi}_1^0$ MC sample have lower efficiencies in both electron and muon cases. But the real lepton efficiencies computed by the $t\bar{t}$ have similar efficiencies with the data one in the electron case and have lower efficiencies in the muon case. This is because the $t\bar{t}$ MC has lower real muon efficiencies when $p_T < 40\text{GeV}$ with respect to data. If we require a $p_T > 40\text{GeV}$ requirement on the probe leptons, then the real muon efficiencies for $t\bar{t}$ MC agree with data. The electron and muon real efficiencies as a function of p_T compared with data and $Z \rightarrow \ell\ell$ are shown in Figure 5.14. From which we notice the real lepton efficiencies of $t\bar{t}$ process is lower than the data one when $p_T < 40\text{GeV}$ and the $\tilde{g} \rightarrow t\bar{t}\tilde{\chi}_1^0$ process has lower real lepton efficiencies in both electron and muon cases.

5.5.9 Sources of systematic uncertainties

5.5.10 Measurement systematics

The systematic uncertainties associated with the Z tag-and-probe method have been studied by varying the background template definitions, the template fitting ranges, and the $m_{\ell\ell}$ windows used for the real lepton efficiency measurements. The variations of the background template definitions are presented in Table 5.4 and the additional template fitting ranges are $[60 - 70] \cup [100 - 120]$ GeV and $[65 - 75] \cup [100 - 120]$ GeV. The two other measurement $m_{\ell\ell}$ windows considered are $75 < m_{ee} < 105\text{GeV}$ and $85 < m_{ee} < 95\text{GeV}$. In total 27 variations of the measurement methods are considered in $p_T < 20\text{GeV}$ and 3 variations when

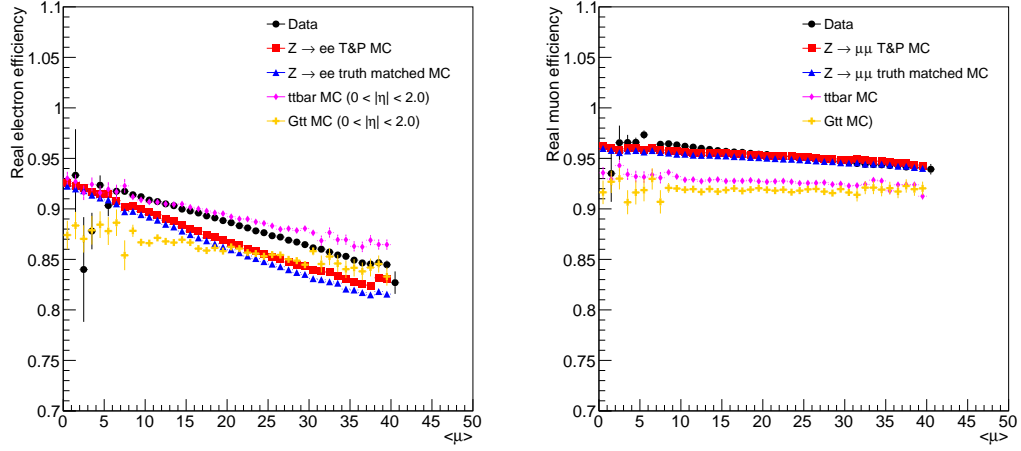


Figure 5.13: The real lepton efficiencies as a function of the average interactions per crossing $\langle \mu \rangle$. The data is shown as black dots and $Z \rightarrow \ell\ell$ tag-and-probe and truth matching are in red squares and blue triangles, respectively. The real lepton efficiencies calculated using $t\bar{t}$ and $\tilde{g} \rightarrow t\bar{t}\tilde{\chi}_1^0$ MC are also shown in magenta diamonds and yellow crosses. And they show the real lepton efficiencies with different event topologies. The $|\eta| < 2$ requirement has been applied on the $t\bar{t}$ and $\tilde{g} \rightarrow t\bar{t}\tilde{\chi}_1^0$ MC samples for the electron case.

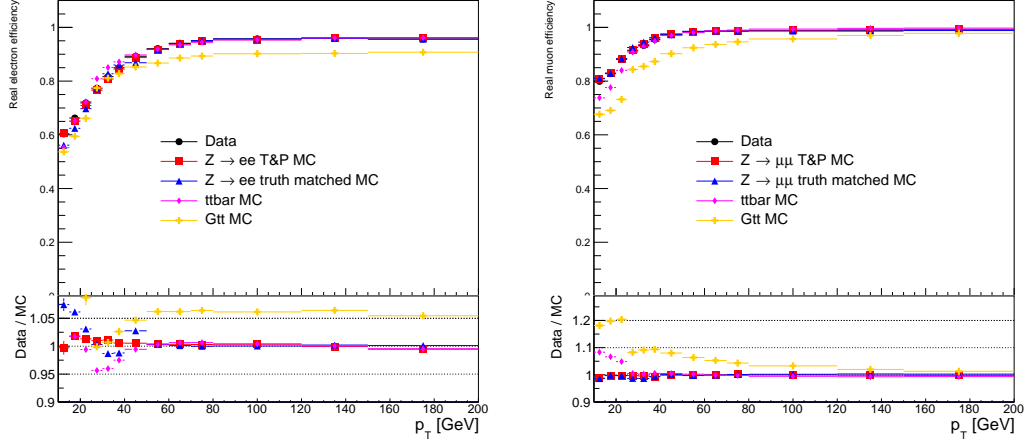


Figure 5.14: The electron and muon real efficiencies as a function of p_T . The black dots stands for data, the red squares and blue triangles indicate $Z \rightarrow \ell\ell$ tag-and-probe and truth matching, respectively. The $t\bar{t}$ and $\tilde{g} \rightarrow t\bar{t}\tilde{\chi}_1^0$ are represented by magenta diamonds and yellow crosses, respectively. All of the real lepton efficiencies computed by MC processes are agree with the data one when $p_T > 40\text{GeV}$. The differences in the $p_T < 40\text{GeV}$ region come from the different event topologies.

$p_T > 20\text{GeV}$ for the electron case. Because the background subtraction is applied on the electron case only, no background templates and template fitting ranges are used in the muon case. Therefore, the systematic uncertainties of real muon efficiency are studied by varying the 3 different $m_{\ell\ell}$ windows only. Table 5.7 and Table 5.8 show the measurement uncertainties for the electron and muon cases, respectively.

Electrons (measurement)			
$ \eta $	[0, 0.8]	[0.8, 1.37]	[1.52, 2.0]
$10 < p_T < 15\text{GeV}$	2.32%(t) / 2.85%(f) / 5.06%(m)	4.84%(t) / 1.99%(f) / 5.66%(m)	5.90%(t) / 0.28%(f) / 10.31%(m)
$15 < p_T < 20\text{GeV}$	1.39%(t) / 0.00%(f) / 3.55%(m)	2.01%(t) / 0.01%(f) / 3.94%(m)	2.05%(t) / 0.15%(f) / 6.19%(m)
$20 < p_T < 25\text{GeV}$	2.46%	3.34%	3.88%
$25 < p_T < 30\text{GeV}$	1.69%	2.17%	2.66%
$30 < p_T < 35\text{GeV}$	1.19%	1.75%	2.07%
$35 < p_T < 40\text{GeV}$	0.70%	1.23%	1.32%
$40 < p_T < 50\text{GeV}$	0.20%	0.30%	0.42%
$50 < p_T < 60\text{GeV}$	0.15%	0.17%	0.20%
$60 < p_T < 70\text{GeV}$	0.13%	0.14%	0.21%
$70 < p_T < 80\text{GeV}$	0.10%	0.17%	0.18%
$80 < p_T < 120\text{GeV}$	0.12%	0.11%	0.19%
$120 < p_T < 150\text{GeV}$	0.10%	0.16%	0.06%
$150 < p_T < 200\text{GeV}$	0.11%	0.03%	0.19%

Table 5.7: The measurement systematic uncertainties in percentage on the real electron efficiencies. The background subtraction is applied on the first two p_T bins on the electron case so there are three different sources of the measurement systematic uncertainties, varying templates, varying fitting ranges, and varying $m_{\ell\ell}$ windows denoted by t, f, and m, respectively. Only the variations of $m_{\ell\ell}$ windows are applied on the electron case when $p_T > 20\text{GeV}$.

Muon (measurement)				
$ \eta $	[0, 0.6]	[0.6, 1.2]	[1.2, 1.8]	[1.8, 2.5]
$10 < p_T < 15\text{GeV}$	1.29%	1.06%	0.96%	0.98%
$15 < p_T < 20\text{GeV}$	0.44%	0.38%	0.56%	0.64%
$20 < p_T < 25\text{GeV}$	0.19%	0.22%	0.38%	0.56%
$25 < p_T < 30\text{GeV}$	0.09%	0.12%	0.22%	0.36%
$30 < p_T < 35\text{GeV}$	0.06%	0.11%	0.23%	0.32%
$35 < p_T < 40\text{GeV}$	0.05%	0.07%	0.13%	0.26%
$40 < p_T < 50\text{GeV}$	0.04%	0.04%	0.05%	0.07%
$50 < p_T < 60\text{GeV}$	0.06%	0.06%	0.09%	0.07%
$60 < p_T < 70\text{GeV}$	0.06%	0.07%	0.08%	0.08%
$70 < p_T < 80\text{GeV}$	0.06%	0.08%	0.11%	0.05%
$80 < p_T < 120\text{GeV}$	0.07%	0.07%	0.12%	0.07%
$120 < p_T < 150\text{GeV}$	0.06%	0.06%	0.15%	0.05%
$150 < p_T < 200\text{GeV}$	0.09%	0.11%	0.15%	0.06%

Table 5.8: The measurement systematic uncertainties in percentage on the real muon efficiencies. Because the background subtraction is not applied on the muon case, only the variations of $m_{\ell\ell}$ windows are applied on the muon case for all p_T regions.

The largest contribution to the systematic uncertainties arises from the $m_{\ell\ell}$ window variations. This result is expected as electrons extracted from the $m_{\ell\ell}$ tail region are affected by bremsstrahlung effects. Thus, the efficiency computed with electrons extracted with a large $m_{\ell\ell}$ window will be lower than the one computed using a tighter $m_{\ell\ell}$ window. In the $10 < p_T < 15\text{GeV}$ region, the order of magnitude of the $m_{\ell\ell}$ window variation is 9% whereas the background subtraction one is 5%. This result shows the robustness of the background subtraction method.

5.5.11 Trigger bias

Besides the measurement systematic uncertainties, the different trigger strategies for the analysis are also considered as sources of systematic uncertainties. The leptons entering in the signal regions are required to fire one of the di-lepton triggers. For example, if the event fires the E_T^{miss} trigger or the considered lepton is the third leading lepton, then no trigger matching should be applied for the efficiency computation. On the other hand, if the event fires the di-lepton trigger and the considered lepton is the leading lepton or the sub-leading lepton, then a trigger matching should be applied before the real lepton efficiency measurement. The nominal value of the real lepton efficiencies are measured with events triggered by the single lepton triggers listed in Table 5.3 and the tag lepton must match to the corresponding single lepton trigger. The tag trigger matching is required in order to provide unbiased probe leptons for the real lepton efficiency measurements. The systematics uncertainties of the real lepton efficiencies are assigned as the differences between the nominal values and the values measured with different

trigger. Moreover, as a $p_T > 20\text{GeV}$ requirement is applied on the two leading leptons, the leptons with $p_T < 20\text{GeV}$ will never be trigger matched to the dilepton trigger. Therefore, no systematics are assigned in the $10 < p_T < 20\text{GeV}$ range. Figure 5.15 shows the real electron efficiencies computed with the different trigger strategies as a function of p_T in different $|\eta|$ regions. Because the crack region, $1.37 < |\eta| < 1.52$, is not considered in the analysis, this region is removed from the real electron efficiency study. And Figure 5.16 shows the real muon efficiencies computed with the different trigger strategies as a function of p_T in different $|\eta|$ regions. These plots indicate that the trigger strategy does not affect the real muon efficiency measurement. Table 5.9 and Table 5.10 show the systematic uncertainties in percentage due to the different trigger strategies.

5.5.12 \tilde{g} Extrapolation to signal regions

The real lepton efficiencies are measured with a sample enriched in $Z \rightarrow \ell\ell$ events characterized by well isolated leptons. The different processes entering in the signal region are accompanied by many (b -)jets and with a different event topology. Thus, the leptons presented in the final state are not necessary well isolated. As tighter isolation cuts are used for the signal lepton definitions, their associated real efficiencies can be smaller. A SUSY benchmark model $\tilde{g} \rightarrow t\bar{t}\tilde{\chi}_1^0$ is used and the boosted event topologies are selected by applying the requirement $\Delta m = m_{\tilde{g}} - m_{\tilde{\chi}_1^0} > 1\text{TeV}$ on the model. These differences in the real lepton efficiencies between the $Z \rightarrow \ell\ell$ processes and the $\tilde{g} \rightarrow t\bar{t}\tilde{\chi}_1^0$ process are assigned as system uncertainties. As the topology of one of the main irreducible backgrounds

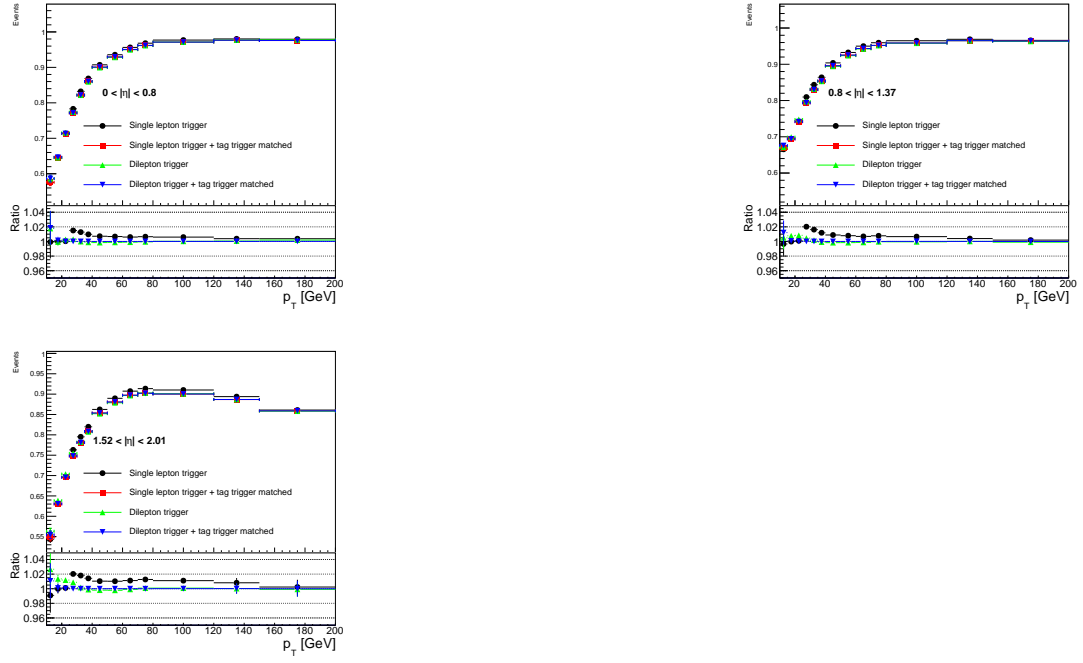


Figure 5.15: The real electron efficiencies as a function of p_T in 3 different $|\eta|$ regions. Four different trigger strategies are applied. The nominal values of the real electron efficiencies are measured using the single lepton trigger with tag trigger matched. The differences between the nominal values and the values measured using other strategies are assigned as the systematic uncertainties.

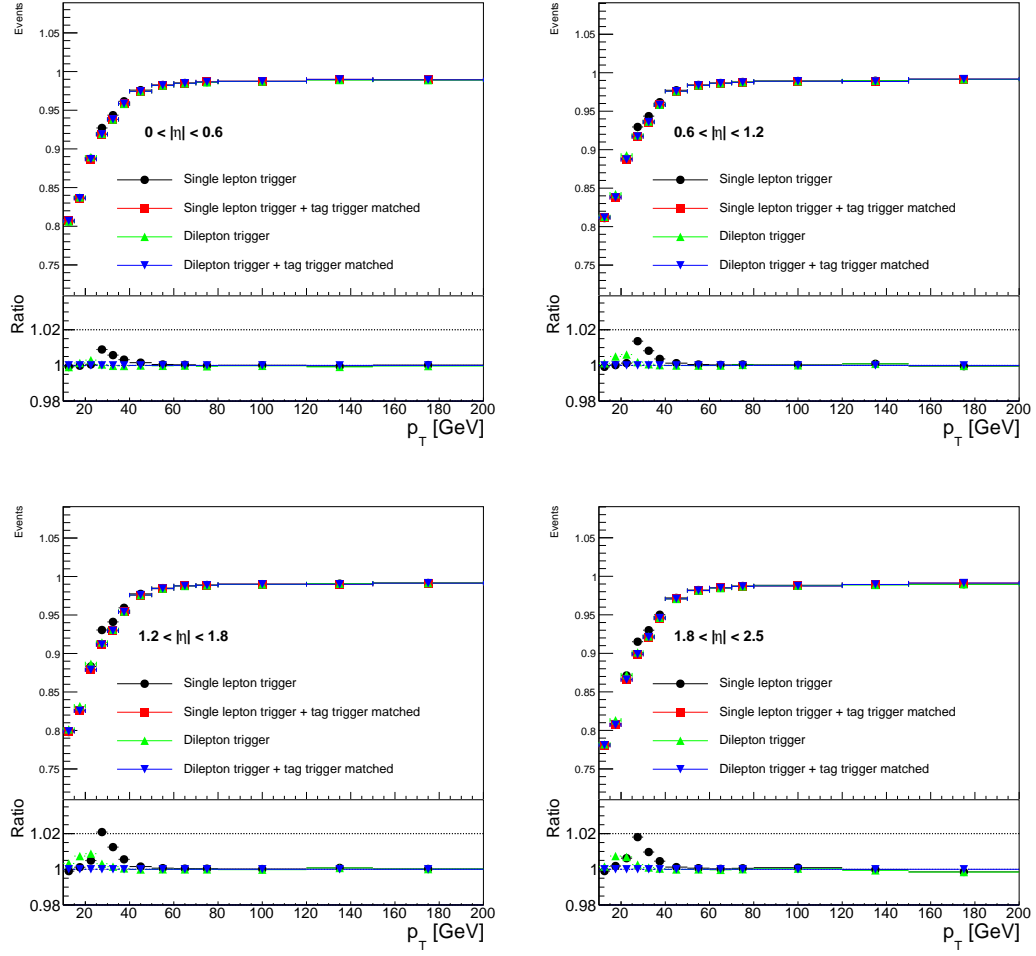


Figure 5.16: The real muon efficiencies as a function of p_T in 4 different $|\eta|$ regions. Four different trigger strategies are applied. The nominal values of the real muon efficiencies are measured using the single lepton trigger with tag trigger matched. The differences between the nominal values and the values measured using other strategies are assigned as the systematic uncertainties.

Electrons (trigger)			
$ \eta $	[0, 0.8]	[0.8, 1.37]	[1.52, 2.0]
$10 < p_T < 15\text{GeV}$	2.46%	1.32%	3.02%
$15 < p_T < 20\text{GeV}$	0.16%	0.78%	1.33%
$20 < p_T < 25\text{GeV}$	0.29%	0.84%	1.18%
$25 < p_T < 30\text{GeV}$	1.53%	2.07%	2.20%
$30 < p_T < 35\text{GeV}$	1.28%	1.63%	1.81%
$35 < p_T < 40\text{GeV}$	0.98%	1.19%	1.42%
$40 < p_T < 50\text{GeV}$	0.73%	0.90%	1.05%
$50 < p_T < 60\text{GeV}$	0.68%	0.81%	1.05%
$60 < p_T < 70\text{GeV}$	0.61%	0.70%	1.13%
$70 < p_T < 80\text{GeV}$	0.65%	0.77%	1.27%
$80 < p_T < 120\text{GeV}$	0.60%	0.66%	1.11%
$120 < p_T < 120\text{GeV}$	0.38%	0.40%	0.79%
$150 < p_T < 200\text{GeV}$	0.43%	0.22%	0.25%

Table 5.9: The systematic uncertainties in percentage due to the different trigger strategies for the real electron efficiencies. The uncertainties of each trigger strategy are computed with respect to the one applied single lepton trigger and required tag trigger matched. And the total uncertainties are the quadratic sum of the uncertainties of each trigger strategy.

Muons (trigger)				
$ \eta $	[0, 0.6]	[0.6, 1.2]	[1.2, 1.8]	[1.8, 2.5]
$10 < p_T < 15\text{GeV}$	0.11%	0.15%	0.34%	0.19%
$15 < p_T < 20\text{GeV}$	0.14%	0.50%	0.75%	0.77%
$20 < p_T < 25\text{GeV}$	0.30%	0.63%	1.01%	0.93%
$25 < p_T < 30\text{GeV}$	0.90%	1.38%	2.12%	1.83%
$30 < p_T < 35\text{GeV}$	0.58%	0.84%	1.27%	0.99%
$35 < p_T < 40\text{GeV}$	0.33%	0.37%	0.57%	0.46%
$40 < p_T < 50\text{GeV}$	0.16%	0.13%	0.16%	0.13%
$50 < p_T < 60\text{GeV}$	0.06%	0.06%	0.07%	0.07%
$60 < p_T < 70\text{GeV}$	0.04%	0.04%	0.04%	0.04%
$70 < p_T < 80\text{GeV}$	0.05%	0.06%	0.04%	0.06%
$80 < p_T < 120\text{GeV}$	0.03%	0.03%	0.03%	0.11%
$120 < p_T < 150\text{GeV}$	0.05%	0.13%	0.10%	0.08%
$150 < p_T < 200\text{GeV}$	0.04%	0.03%	0.02%	0.19%

Table 5.10: The systematic uncertainties in percentage due to the different trigger strategies for the real muon efficiencies. The uncertainties of each trigger strategy are computed with respect to the one applied single lepton trigger and required tag trigger matched. And the total uncertainties are the quadratic sum of the uncertainties of each trigger strategy.

$t\bar{t}V$ is close to the $t\bar{t}$ one, the efficiencies measured with the leptons from $t\bar{t}$ are also considered. The efficiency comparisons are made for each p_T bin considering different $\Delta R(\ell, jets)$ ranges.

The kinematic distributions of the candidate leptons from the $Z \rightarrow \ell\bar{\ell}$, $t\bar{t}$, and $\tilde{g} \rightarrow t\bar{t}\tilde{\chi}_1^0$ are shown in Figure 5.17. The top row shows the p_T distributions for the candidate electrons on the left hand side and muons on the right hand side. The bottom row shows the $|\eta|$ distributions. These plots show that the leptons from SUSY process are more boosted and more central than the ones from Z and $t\bar{t}$ processes. The $\Delta R(\ell, jet)$ and the N_{jets} distributions of the candidate leptons extracted from $Z \rightarrow \ell\bar{\ell}$, $t\bar{t}$, and $\tilde{g} \rightarrow t\bar{t}\tilde{\chi}_1^0$ processes are shown in Figure 5.18. The left hand side is the electron case and the right hand side is the muon case. The $\Delta R(\ell, jet)$ distribution associated to the SUSY signals peak at 0.5 and most of the statistics are contained in $\Delta R(\ell, jet) < 1$ region. In comparison, the leptons from the $Z \rightarrow \ell\bar{\ell}$ processes are not accompanied with a signal jet and the $\Delta R(\ell, jet)$ distribution peak about $\Delta R(\ell, jet) = 3$. The jet multiplicity of the $Z \rightarrow \ell\bar{\ell}$ peak at 4 jets for the electron case, 3 jets for the muon case, but for the SUSY signal peaks at 9 jets. These plots confirm that the leptons produced in the SUSY signal are accompanied with many more jets and are therefore less isolated than the $Z \rightarrow \ell\bar{\ell}$ processes. This extreme topology enables us to assess a conservative SUSY signal extrapolation systematic uncertainty that should cover all SUSY signal processes considered by the analysis.

The real lepton efficiencies as a function of p_T using $Z \rightarrow \ell\bar{\ell}$, $t\bar{t}$, and $\tilde{g} \rightarrow t\bar{t}\tilde{\chi}_1^0$ are shown in Figure 5.14. The lower panel shows the ratio with respect to the

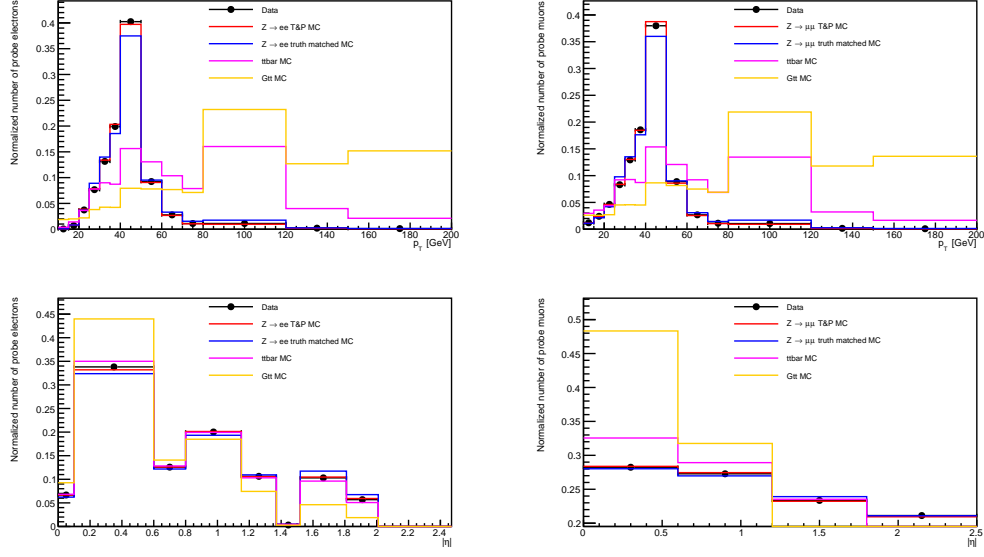


Figure 5.17: The kinematic distributions of the candidate lepton from the processes considered for the systematic uncertainty study. The top row shows the p_T distributions for the candidate electrons on the left hand side and muons on the right hand side. The bottom row shows the $|\eta|$ distributions. The $\tilde{g} \rightarrow t\bar{t}\tilde{\chi}_1^0$ process is more boosted and centralized than the $Z \rightarrow \ell\ell$ and $t\bar{t}$ processes.

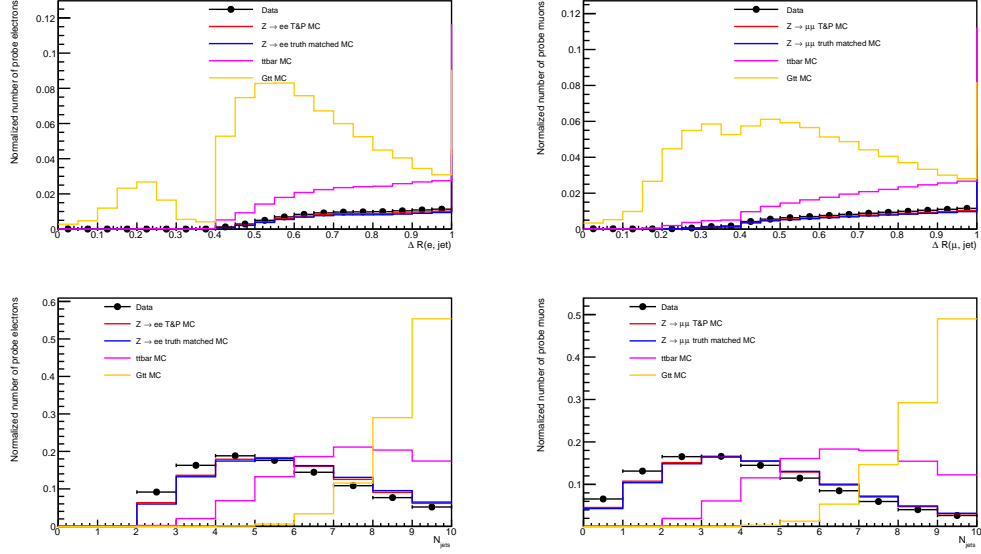


Figure 5.18: The $\Delta R(\ell, jet)$ and the N_{jets} distributions of the candidate leptons extracted from $Z \rightarrow \ell\ell$, $t\bar{t}$, and $\tilde{g} \rightarrow t\bar{t}\tilde{\chi}_1^0$ processes. Most of the statistics of $\tilde{g} \rightarrow t\bar{t}\tilde{\chi}_1^0$ are located in $\Delta R(\ell, jet) < 1$ region and higher N_{jets} region. In the contrast, the statistics of $Z \rightarrow \ell\ell$ processes are populated at higher $\Delta R(\ell, jet) < 1$ region and lower N_{jets} region.

data. In the left hand side plot shows the real electron efficiencies as a function of p_T and the right hand side shows the real muon efficiencies as a function of p_T . The left hand side plot shows that the real electron efficiencies are p_T dependent in the $p_T < 50\text{GeV}$ region and become stable when $p_T > 50\text{GeV}$. The efficiencies computed using $\tilde{g} \rightarrow t\bar{t}\tilde{\chi}_1^0$ are about 8% lower than the efficiencies calculated using $Z \rightarrow ee$. The observed differences in the low p_T region are mostly due to the calorimeter isolation and the track isolation cuts. And the track isolation and d_0/σ_{d_0} cuts are the main reasons cause the efficiency differences in the muon real efficiencies.

The average efficiencies of $Z \rightarrow \ell\ell$ are calculated and the relative efficiency differences are computed with respect to the average efficiencies. The relative efficiency differences as a function of $\Delta R(\ell, jet)$ are considered for each measured p_T bins and are shown in Table 5.11.

5.5.13 Final uncertainties

The final uncertainties of the real lepton efficiency study are the quadratic sum of the statistical uncertainties and the systematic uncertainties. The systematic uncertainties under the consideration include the measurement uncertainties, the trigger uncertainties, the uncertainties come from the differences between truth matching and tag-and-probe methods, and the uncertainties in the busy environment. Because the uncertainties in the busy environment are a function of p_T and ΔR and the other systematic uncertainties are a function of p_T and $|\eta|$, we keep the uncertainties in busy environment in Table 5.11 and do not

electrons (busy environments)								
$\Delta R(e, jet)$	[0, 0.1]	[0.1, 0.15]	[0.15, 0.2]	[0.2, 0.3]	[0.3, 0.35]	[0.35, 0.4]	[0.4, 0.6]	[0.6, 4]
10 GeV < p_T < 20 GeV	-	-	-	-	-	-	25.31%	6.5%
20 GeV < p_T < 30 GeV	-	-	-	-	-	73.37%	10.21%	0.37%
30 GeV < p_T < 40 GeV	-	-	-	97.71%	48.22%	15.54%	7.29%	0.58%
40 GeV < p_T < 50 GeV	-	-	-	52.81%	22.80%	16.73%	7.68%	1.10%
50 GeV < p_T < 60 GeV	-	-	-	29.96%	21.49%	20.23%	6.99%	2.78%
60 GeV < p_T < 80 GeV	-	-	55.89%	24.31%	17.40%	24.77%	6.20%	2.87%
80 GeV < p_T < 150 GeV	-	57.52%	30.24%	16.45%	12.73%	20.92%	4.44%	2.73%
150 GeV < p_T < 200 GeV	88.54%	40.16%	19.34%	8.45%	14.66%	16.57%	2.57%	1.90%
muons (busy environments)								
$\Delta R(\mu, jet)$	[0, 0.1]	[0.1, 0.15]	[0.15, 0.2]	[0.2, 0.3]	[0.3, 0.35]	[0.35, 0.4]	[0.4, 0.6]	[0.6, 4]
10 GeV < p_T < 20 GeV	-	-	-	-	-	-	33.59%	5.18%
20 GeV < p_T < 30 GeV	-	-	-	-	-	82.34%	22.27%	3.39%
30 GeV < p_T < 40 GeV	-	-	-	98.54%	56.36%	31.89%	14.22%	2.24%
40 GeV < p_T < 50 GeV	-	-	-	53.10%	21.33%	13.90%	6.81%	1.45%
50 GeV < p_T < 60 GeV	-	-	-	24.98%	13.72%	9.62%	3.83%	0.79%
60 GeV < p_T < 80 GeV	-	-	44.41%	13.75%	6.14%	4.76%	2.04%	0.15%
80 GeV < p_T < 150 GeV	-	29.94%	7.14%	3.16%	1.30%	1.04%	0.07%	0.57%
150 GeV < p_T < 200 GeV	82.26%	4.14%	1.02%	0.17%	0.29%	0.62%	1.02%	1.13%

Table 5.11: Systematic uncertainties on the measured real lepton

efficiency due to extrapolation to busy environments using

combine it with other uncertainties. The Table 5.12 and Table 5.13 show the final uncertainties.

5.6 Fake lepton rate

The FNP leptons efficiency parameter should be measured in a CR dominated by same FNP lepton sources as the signal regions. The definition of this control region is analysis dependent, and dedicated truth composition studies should be performed to understand the different sources of FNP leptons in the signal and FNP leptons enriched CR regions.

Electrons (final uncertainties)			
$ \eta $	[0, 0.8]	[0.8, 1.37]	[1.52, 2.0]
$10 < p_T < 15\text{GeV}$	0.047	0.063	0.089
$15 < p_T < 20\text{GeV}$	0.027	0.042	0.062
$20 < p_T < 25\text{GeV}$	0.018	0.031	0.041
$25 < p_T < 30\text{GeV}$	0.029	0.024	0.027
$30 < p_T < 35\text{GeV}$	0.023	0.021	0.023
$35 < p_T < 40\text{GeV}$	0.014	0.018	0.018
$40 < p_T < 50\text{GeV}$	0.007	0.010	0.010
$50 < p_T < 60\text{GeV}$	0.008	0.010	0.010
$60 < p_T < 70\text{GeV}$	0.007	0.010	0.010
$70 < p_T < 80\text{GeV}$	0.008	0.011	0.012
$80 < p_T < 120\text{GeV}$	0.010	0.010	0.011
$120 < p_T < 120\text{GeV}$	0.005	0.005	0.011
$150 < p_T < 200\text{GeV}$	0.005	0.003	0.020

Table 5.12: The final uncertainties of the real electron efficiencies.

The final uncertainties are calculated using the quadratic sum of the statistical uncertainties and the systematic uncertainties. The systematic uncertainties contain the measurement uncertainties, the trigger uncertainties, and the uncertainties come from the difference between truth matching and tag-and-probe methods. The trigger uncertainties are included for $p_T > 20\text{GeV}$ only. The uncertainties in the busy environment do not incorporate in the final uncertainties calculation because it is a function of p_T and ΔR .

Muons (final uncertainties)				
$ \eta $	[0, 0.6]	[0.6, 1.2]	[1.2, 1.8]	[1.8, 2.5]
$10 < p_T < 15\text{GeV}$	0.014	0.010	0.008	0.011
$15 < p_T < 20\text{GeV}$	0.005	0.006	0.008	0.011
$20 < p_T < 25\text{GeV}$	0.003	0.006	0.010	0.010
$25 < p_T < 30\text{GeV}$	0.011	0.015	0.022	0.019
$30 < p_T < 35\text{GeV}$	0.007	0.009	0.014	0.011
$35 < p_T < 40\text{GeV}$	0.004	0.004	0.006	0.006
$40 < p_T < 50\text{GeV}$	0.002	0.001	0.002	0.001
$50 < p_T < 60\text{GeV}$	0.001	0.001	0.001	0.001
$60 < p_T < 70\text{GeV}$	0.001	0.001	0.001	0.002
$70 < p_T < 80\text{GeV}$	0.002	0.001	0.001	0.002
$80 < p_T < 120\text{GeV}$	0.004	0.002	0.002	0.002
$120 < p_T < 150\text{GeV}$	0.006	0.005	0.005	0.005
$150 < p_T < 200\text{GeV}$	0.005	0.005	0.005	0.006

Table 5.13: The final uncertainties of the real muon efficiencies. The final uncertainties are calculated using the quadratic sum of the statistical uncertainties and the systematic uncertainties. The systematic uncertainties contain the measurement uncertainties, the trigger uncertainties, and the uncertainties come from the difference between truth matching and tag-and-probe methods. The uncertainties in the busy environment do not incorporate in the final uncertainties calculation because it is a function of p_T and ΔR .

Chapter 6

Search for new physics in events with two same sign leptons or three leptons

Conclusion