
CMS Internal Note

The content of this note is intended for CMS internal use and distribution only

October 1, 2010

Fake Lepton Background Estimate for the $Z \rightarrow e^+e^-$ Cross Section Measurement Using the Fake Rate Method

G. Bauer^{a)}, J. Bendavid^{a)}, E. Butz^{a)}, M. Chan^{a)}, V. Dutta^{a)}, P. Everaerts^{a)}, G. Gómez-Ceballos^{a)},
M. Goncharov^{a)}, K. Hahn^{a)}, P. Harris^{a)}, M. Klute^{a)}, I. Kravchenko^{c)}, S. Nahn^{a)}, C. Paus^{a)}, D. Ralph^{a)},
M. Rudolph^{a)}, K. Sumorok^{a)}, K. Sung^{a)}, S. Tkaczyk^{b)}, S. Xie^{a)}, M. Yang^{a)}, M. Zanetti^{a)}

Abstract

A data driven estimate of the fake lepton backgrounds for the $Z \rightarrow e^+e^-$ cross section measurement is presented for the first 2.88pb^{-1} of CMS data. We obtain an estimate of 4.5 ± 5.1 (statistical) ± 2.2 (systematic) events, representing a sub-percent level uncertainty on the cross section measurement. A comprehensive set of systematic uncertainties are accounted for.

^{a)} Laboratory for Nuclear Science, Massachusetts Institute of Technology, Cambridge, USA

^{b)} Fermilab National Accelerator Laboratory, Batavia, USA

^{c)} University of Nebraska, Lincoln, USA

1 Introduction

One important aspect of the first $Z \rightarrow e^+e^-$ cross section measurement [1] that was missing was a robust, data driven, estimate of the fake lepton backgrounds. The Monte Carlo simulation is known to be unreliable in predicting backgrounds containing fake electrons. The two main reasons for this deficiency in the Monte Carlo simulation is that a fake isolated electron is the result of a jet fragmentation occurring in a very extreme region of phase space where the fragmentation and hadronization models are not well constrained, and that the exact detector material is not precisely known. As a result one typically makes use of data driven techniques to estimate these backgrounds. We use one such method where we extrapolate from the yield of loose electrons to the yield of tight electrons using an orthogonal calibration sample dominated by fake electrons, where the tight electrons refer to those passing the electron selection that is used in the final event selection. The precise definition of the loose electron can be tuned in order to minimize systematic uncertainties,

The fake electron dominated calibration sample is defined by selecting events passing the HLT_Jet30U trigger. From here on, we will refer to the VBTF 95% working point electron selection as the tight electron selection. The fake rate is measured by counting the number of electron candidates passing the tight electron selection and dividing by the number of electron candidates passing the looser denominator selection. It is therefore a function of the particular definition that one uses for the loose candidates. We will use two different denominator selections for two different purposes, as explained in greater detail in Section 2.

In order to remove any bias introduced by the jet trigger requirements, we veto any candidates that lie within a $0.3 \Delta R$ cone to the trigger jet with the highest measured transverse momentum. To account for dependencies on the kinematic quantities of the candidates, the fake rates are measured in bins of transverse momentum and pseudorapidity.

2 Fake Lepton Background Estimation Method

For the $Z \rightarrow e^+e^-$ analysis, there are three background processes involving fake electrons: QCD multijet events, where both electrons are jets faking electrons, γ +Jets events where one electron is a photon faking an electron and the other electron is a jet faking an electron, and W+Jets events where one electron is a real electron and the other one is a jet faking an electron. One feature of the $Z \rightarrow e^+e^-$ analysis, which generates some complications for estimating fake lepton backgrounds, is the fact that the signal to background ratio is particularly large, even after imposing various anti-signal selection requirements. As a result, one must pay particular attention to signal contaminating the background sample. We use two different types of anti-signal selection and combine them to derive a total fake lepton background estimate.

In the first case, referred to as the “Loose+Loose” extrapolation, we define the denominator to be any reconstructed GSF electrons falling inside the electron acceptance region. We select events containing two such denominator electrons, which fail the tight electron selection, whose mass lies inside of the Z mass window, between $60 \text{ GeV}/c^2$ and $120 \text{ GeV}/c^2$. The requirement of two anti-selected electrons is sufficient to suppress signal contamination to a negligible degree. These events are subsequently weighted by the product of the two individual fake rates, defined to be the efficiency that a denominator object, originating from a jet, passes the tight electron selection requirements, in order to construct a background estimate. This extrapolation makes the implicit assumption that all such denominator object pairs originate from jets, and essentially ignores the W+Jets and γ +Jets background. As a result, this estimate accounts for the QCD background only.

For the second case, denoted as the “Tight+Jet” extrapolation, we define the denominator to be a reconstructed jet falling inside the electron acceptance region. The anti-Kt jet algorithm is used through the analysis. We select events containing one electron passing the full electron selection criteria and one such denominator jet, whose mass lies inside of the Z mass window, between $60 \text{ GeV}/c^2$ and $120 \text{ GeV}/c^2$. The events are weighted by the fake rate for a jet to be matched to a reconstructed GSF electron passing the full electron selection. For the QCD component of the background, both selected electrons are fake electrons. Since the Tight+Jet selection does not specify which of the two fake electrons is required to pass the tight selection, our selection includes both cases. This results in a combinatorial double counting of the QCD background.

We correct for this overlap by subtracting the Loose+Loose prediction from the Tight+Jet prediction.

3 Electron Fake Rate Measurements

The fake rate measurements are performed in data events triggered by the presence of a 30 GeV Jet (HLT_Jet30U trigger). The objects matching the trigger jet with the largest transverse momentum is ignored in the fake rate computation, in order to reduce bias from the jet trigger requirements. Details of the procedure can be found in reference[2].

Figure 1 shows one dimensional projections of the GSF electron fake rate as a function of the transverse momentum and pseudorapidity of the GSF electron. Figure 2 show various slices of the two dimensional fake rate in bins of the transverse momentum. The measurement in the Jet30 data is in close agreement with the prediction from the Monte Carlo simulation.

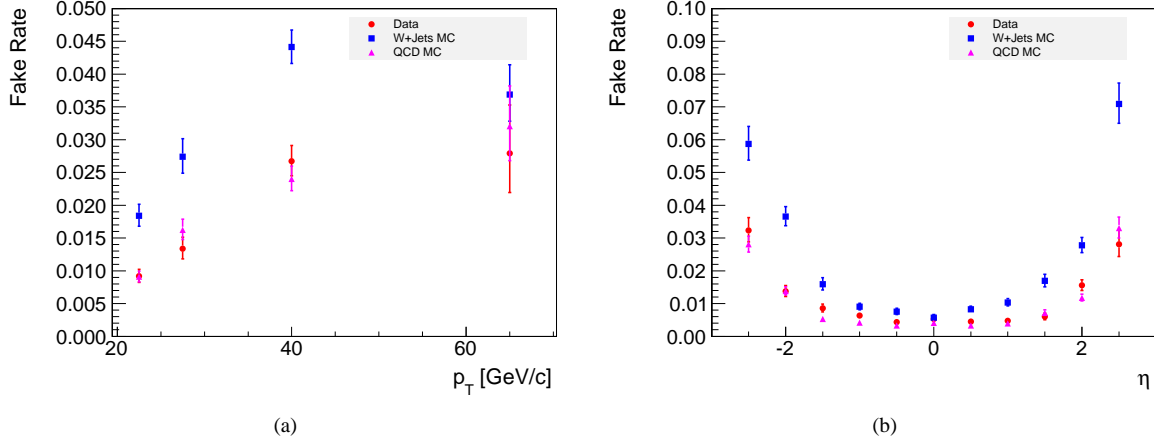


Figure 1: The fake rate for a GSF electron to pass the VBTF95 electron selection measured in the Jet30 data sample are compared to the measurements performed in the corresponding QCD Monte Carlo sample. They are shown as a function of the p_T (a) and η (b) of the GSF electron.

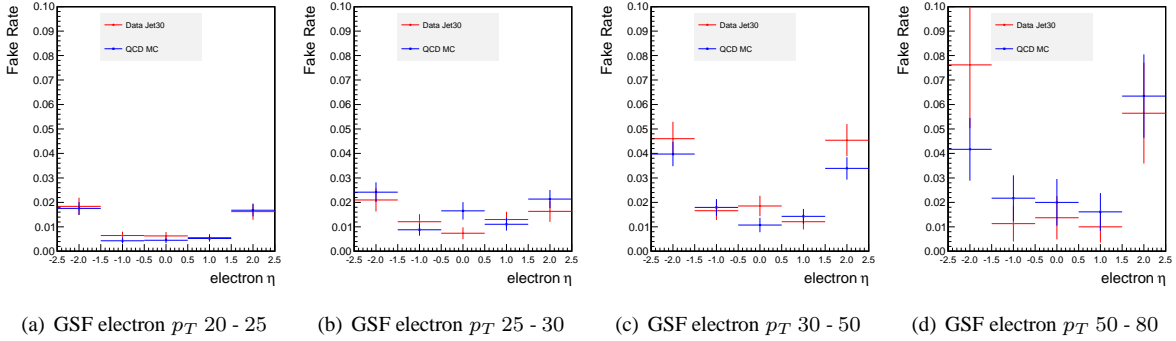


Figure 2: The fake rate for a GSF electron to pass the VBTF95 electron selection measured in the Jet30 data sample are compared to the measurements performed in the corresponding QCD Monte Carlo sample. They are shown as a function of η in various bins of transverse momentum.

One dimensional projections of the fake rate for jets to pass the VBTF95 electron selection are shown in Figure 3 as a function of the transverse momentum and pseudorapidity of the jet. Figure 4 show various slices of the two dimensional fake rate in bins of the transverse momentum. The measurement in the Jet30 data is in close agreement with the prediction from the QCD Monte Carlo simulation, while there is a clear difference with respect to the W+Jets Monte Carlo. This difference is attributed to different fake electron composition and different quark-gluon content, and accounted for in the systematic uncertainties.

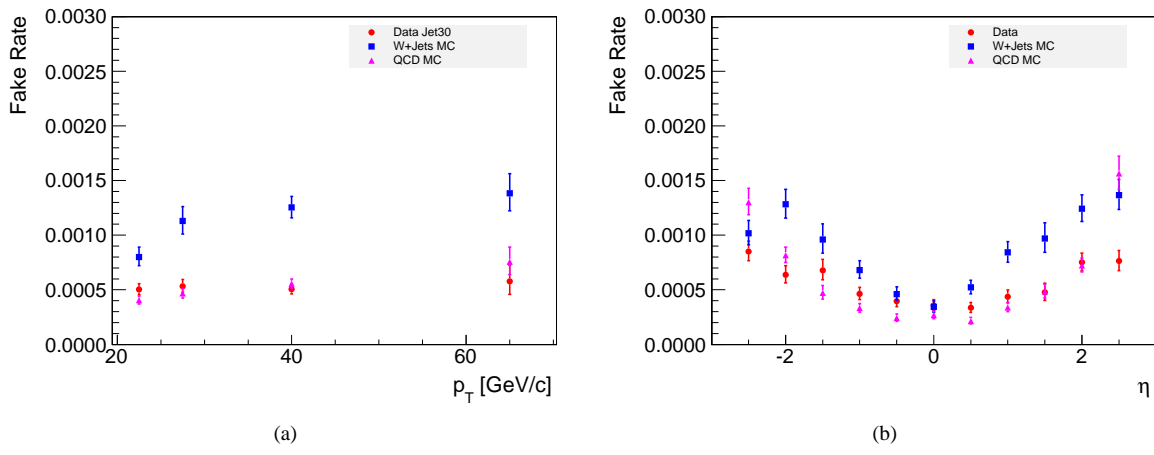


Figure 3: The fake rate for a jet to pass the VBTF95 electron selection measured in the Jet30 data sample are compared to the measurements performed in the corresponding QCD Monte Carlo sample. They are shown as a function of the p_T (a) and η (b) of the GSF electron.

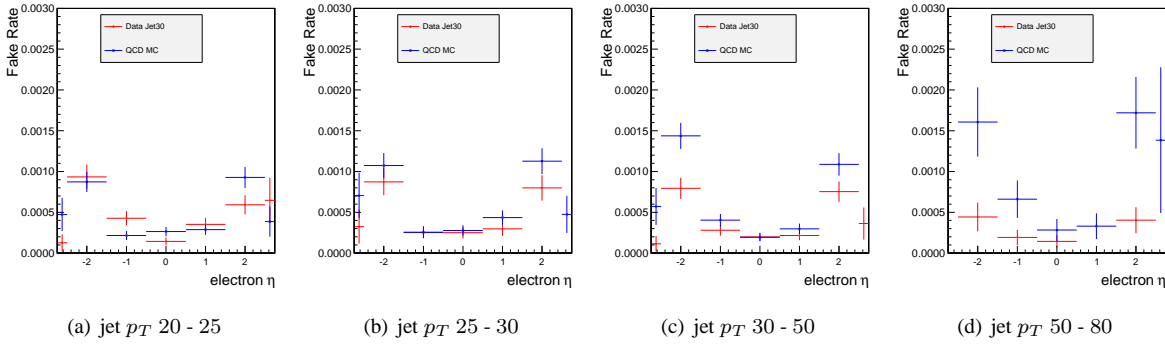


Figure 4: The fake rate for a jet to pass the VBTF95 electron selection measured in the Jet30 data sample are compared to the measurements performed in the corresponding QCD Monte Carlo sample. They are shown as a function of η in various bins of transverse momentum.

4 Performance on Monte Carlo Simulation

We first perform a check of the method using the Monte Carlo simulation, where we apply fake rates computed in simulated QCD events on simulated QCD, W+Jets, and Photon+Jets events and compare the estimate with the prediction of the Monte Carlo simulation. For the QCD simulation prediction, a different Monte Carlo sample has been used where certain generator level cuts have been introduced in order to enhance the number of simulated events with fake electrons. Figures 5, 6, and 7 show the comparison of the estimation using the fake rate method and the simulation prediction as a function of the dielectron mass, for the QCD, and W+Jets and γ +Jets background respectively. One observes agreement to within the total uncertainties of the method for the QCD background. The W+Jets background is underestimated due to the fact that the fake rate in W+Jets events is higher than the fake rate in QCD events due to different fake electron composition seen in Figure 3. This difference is included in the final systematic uncertainty estimate. Figure 8 shows the analogous comparison for the sum of all fake lepton backgrounds. The resulting predictions for the number of background events in the 60 GeV to 120 GeV mass window is shown in Table 1 for each of the relevant background processes separately.

The effect of signal contamination on these predictions are estimated by performing the extrapolation procedure on signal Monte Carlo. Figure 9 shows the amount of signal contamination in the Loose+Loose and Tight+Jet predictions compared to the total fake lepton background. The contamination is on the order of 1% for the Tight+Jet prediction and 0.01% for the Loose+Loose prediction. This yields a negligible effect on the final background estimate.

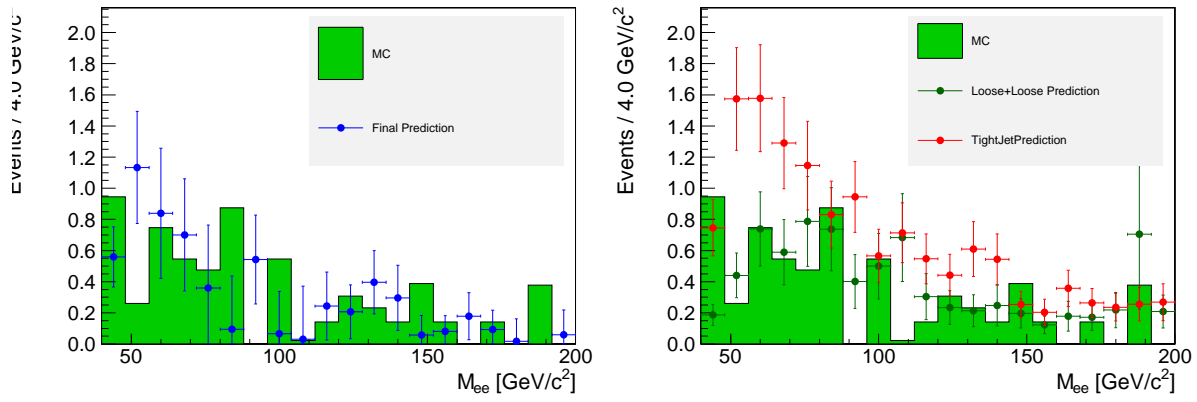


Figure 5: The QCD background prediction using the Tight+Jet and Loose+Loose extrapolation performed on Monte Carlo simulation is compared with the direct prediction from the simulation in (a). The Tight+Jet and Loose+Loose extrapolations are compared to the prediction from the simulation in (b).

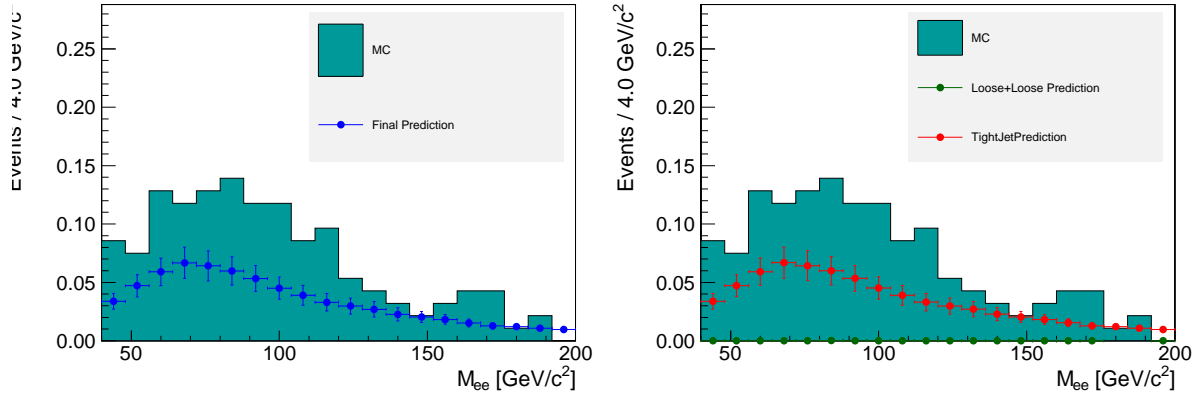


Figure 6: The QCD background prediction using the Tight+Jet and Loose+Loose extrapolation performed on Monte Carlo simulation is compared with the direct prediction from the simulation in (a). The Tight+Jet and Loose+Loose extrapolations are compared to the prediction from the simulation in (b).

Background Process	Tight+Jet Prediction	Loose+Loose Prediction	Final Prediction	MC Simulation Prediction
QCD	8.1	5.0	3.1	3.7
W+Jets	0.5	0.001	0.5	1.0
Photon+Jets	1.0	0.03	1.0	1.4
Total	9.5	5.0	4.5	6.0

Table 1: The predictions obtained from the fake rate method performed on the Monte Carlo are compared with the direct Monte Carlo simulation prediction for each of the relevant background processes separately.

5 Data Results

The data driven estimate of the fake lepton background is compared to the prediction from the Monte Carlo simulation in Figure 10. The Tight+Jet and Loose+Loose predictions are also separately compared to the background predicted by the Monte Carlo simulation. The background estimate inside the Z mass window is summarized in Table 2, along with the uncertainties.

6 Systematic Uncertainties

The dominant systematic uncertainties are those associated with the fake rate measurement itself. There may be differences between the calibration sample and the dielectron sample, in both the composition of the exact fake

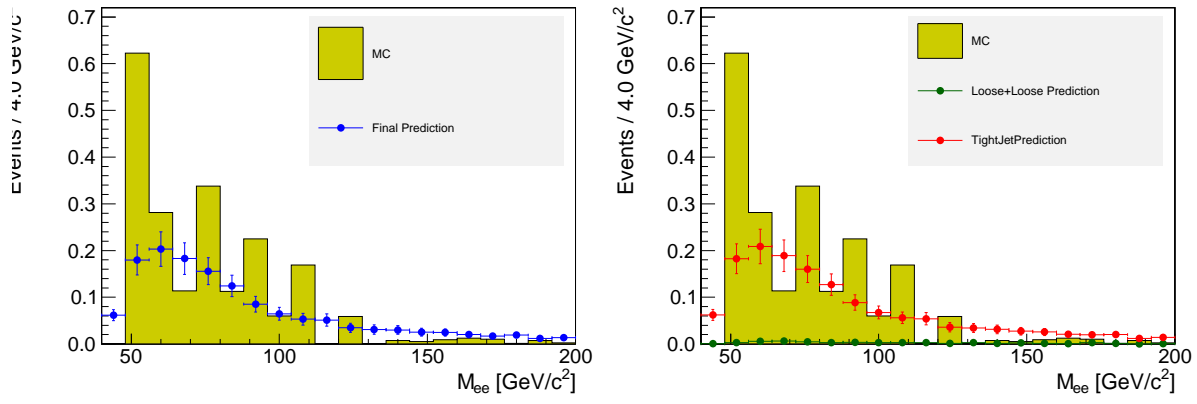


Figure 7: The QCD background prediction using the Tight+Jet and Loose+Loose extrapolation performed on Monte Carlo simulation is compared with the direct prediction from the simulation in (a). The Tight+Jet and Loose+Loose extrapolations are compared to the prediction from the simulation in (b).

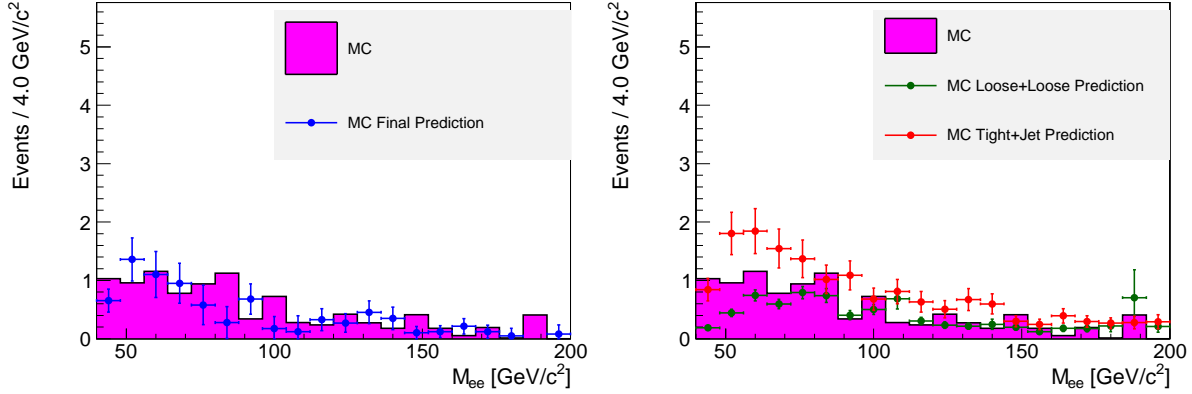


Figure 8: As a consistency check, the total fake lepton background prediction derived from the fake rate method performed on Monte Carlo simulation is compared with the direct prediction from the simulation in (a), while the Tight+Jet and Loose+Loose extrapolations are separately compared to the simulation prediction in (b).

Data Tight+Jet Events	15820
Tight+Jet Prediction	7.9 ± 4.7 (stat) ± 2.2 (W+Jets systematics)
Data Loose+Loose Events	13906
Loose+Loose Prediction	3.5 ± 2.0 (stat)
Final Prediction	4.5 ± 5.1 (stat) ± 2.2 (W+Jets systematics)
MC Simulation Prediction	6.0 ± 0.9

Table 2: The data driven fake lepton background estimates are summarized and compared to the result from the MC simulation. Only statistical uncertainties are given and dominated by the statistical uncertainty of the binned fake rate measurements.

processes, each with different fake rates, and in the amount of jet activity, which affects the isolation fake rate.

To account for the difference in jet activity, we take the difference between the Jet30 data measurement and the Jet15 data measurement as a systematic uncertainty. However, the comparison, shown in Figures 11 and 12, yields a negligible difference.

To account for the composition effect, we take the difference between the fake rate measurement from the Jet30 data measurement and the W+Jet fake rate from Monte Carlo as a systematic uncertainty for the fraction of the background that originate from W+Jets and γ +Jets events. This fraction is calculated from the Tight+Jet and Loose+Loose predictions as in equation 1 and found to be roughly 0.2 in data.

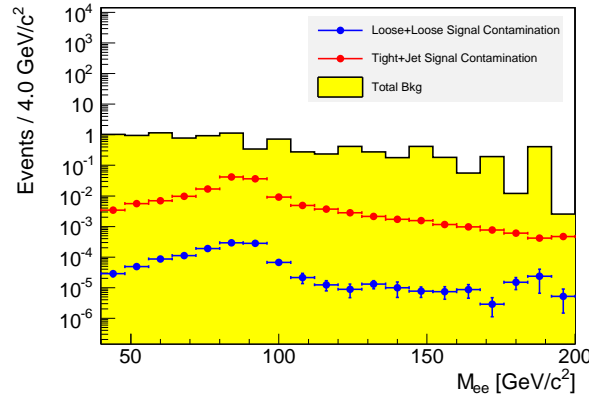


Figure 9: The signal contamination from $Z \rightarrow e^+e^-$ events in the Tight+Jet and Loose+Loose background prediction calculated from Monte Carlo simulation is compared with the total background prediction.

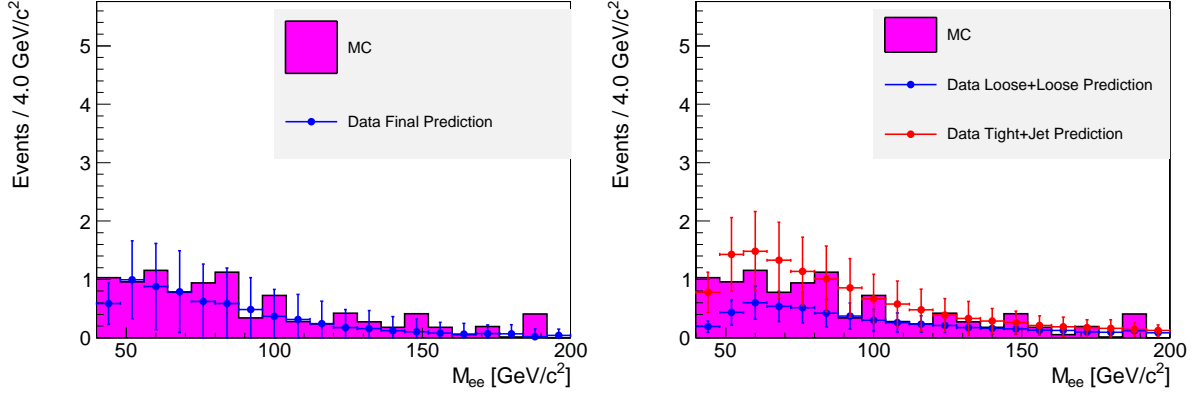


Figure 10: The data driven fake lepton background prediction is compared with the prediction from the Monte Carlo simulation in (a), while the Tight+Jet and Loose+Loose extrapolations from data are separately compared with the prediction from the Monte Carlo simulation in (b).

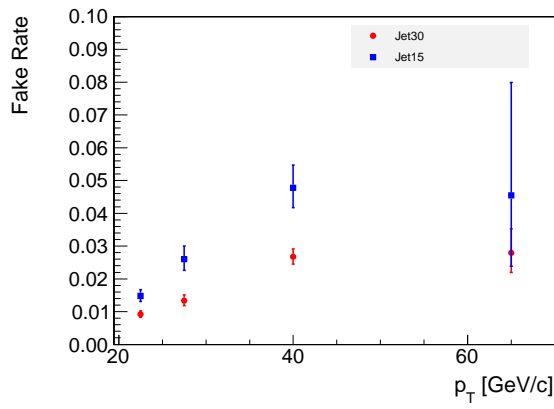
$$\text{Fraction of non - QCD Background} = 1 - \frac{N_{\text{Loose+Loose}}}{N_{\text{Tight+Jet}} - N_{\text{Loose+Loose}}} \quad (1)$$

This difference between the W+Jets fake rate and the fake rate measured in the Jet30 data is illustrated in Figure 3 for the jet fake rate. This systematic uncertainty is only included for the Tight+Jet extrapolation because the Loose+Loose prediction does not account for any W+Jets background by construction. When propagated to the Tight+Jet prediction, we obtain a systematic uncertainty of 2.2 events.

Finally, an additional systematic uncertainty due to jet angular resolution is considered. Since an electron inside a jet may point significantly away from the central jet axis, it is possible for a jet whose direction is outside of the electron acceptance region in η to give a fake electron whose direction points inside the acceptance region. To estimate the magnitude of this effect, we compute the Tight+Jet prediction for jets with $|\eta|$ between 2.5 and 2.75. The resulting number of events for the dataset under consideration is 0.3, which is roughly 7% of the total background estimate. This represents a negligible part of the total systematic uncertainty.

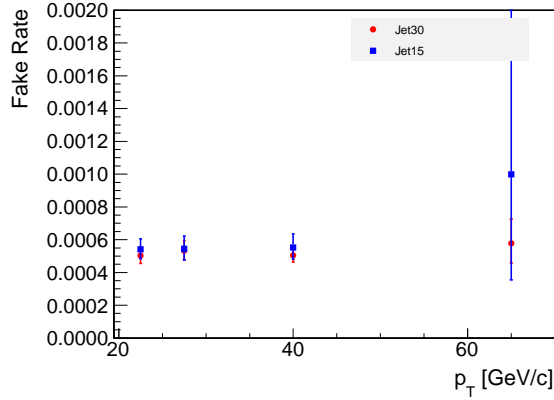
7 Conclusion

We have combined two different data driven techniques to estimate the total fake lepton background for the $Z \rightarrow e^+e^-$ analysis in 2.88pb^{-1} . We estimate a total fake lepton background of 4.5 ± 5.1 (statistical) ± 2.2 (systematic) events. A comprehensive set of systematic effects are taken into account and propagated to the final result.



(a)

Figure 11: A comparison is made between the fake rate for GSF electrons to pass VBTF95 electron selection (ϵ_{fake}) measured in the Jet30 data sample and the Jet15 data sample.



(a)

Figure 12: A comparison is made between the fake rate for jets to be matched to a GSF electron passing the VBTF95 electron selection (ϵ_{fake}) measured in the Jet30 data sample and the Jet15 data sample.

The current total uncertainty on the fake lepton background determination using this method represents a roughly 0.6% uncertainty on the $Z \rightarrow e^+e^-$ cross section measurement, dominated by the statistical uncertainty from the measurement of the fake rate in jet data, which will decrease with an increasing data sample. This measurement serves as a supplemental documentation for the main W and Z inclusive cross section measurement described here[3].

References

- [1] CMS AN-2010/116, “Measurements of Inclusive W and Z Cross Sections in pp Collisions at $\sqrt{s} = 7$ TeV”, The Vector Boson Task Force of the CMS Collaboration *et al.*
- [2] CMS AN-2009/120, “Study of Data-Driven Methods For Estimation of Fake Lepton Backgrounds”, G. Bauer *et al.*
- [3] CMS AN-2010/264, “Updated Measurements of Inclusive W and Z Cross Sections at $\sqrt{s} = 7$ TeV”, J. Alcaraz Maestre, *et al.*