

## Electron and Photon Identification with p20 data

Oleksiy Atramentov<sup>1</sup>, Dmitry Bandurin<sup>2</sup>, Xuebing Bu<sup>3</sup>, Betty Calpas<sup>4</sup>, Edgar Carrera<sup>1</sup>,  
Daniel Duggan<sup>1</sup>, Alexey Ferapontov<sup>2</sup>, Maiko Takahashi<sup>5</sup>, Tonya Uzbyakova<sup>6</sup>, Hang Yin<sup>3</sup>

<sup>1</sup>*Florida State University, FL, USA*

<sup>2</sup>*Kansas State University, KS, USA*

<sup>3</sup>*University of Science and Technology of China, Hefei*

<sup>4</sup>*Centre de Physique des Particules, Marseille, France*

<sup>5</sup>*University of Manchester,*

*Manchester, United Kingdom*

<sup>6</sup>*Moscow State University, Moscow, Russia*

(Dated: 16th September 2008)

### Abstract

This note describes Electron and Photon certification for the data recorded with the DØ detector during Run IIb from August 2006 to February 2008 and reconstructed using p20 version of the DØ software. It also contains description of the new identification variables introduced during the data analysis.

## Contents

<b>1. Introduction.</b>	3
1.1. Data Samples.	3
1.2. Monte Carlo Samples.	4
<b>2. Object Identification.</b>	4
2.1. Choice of selection variables.	5
2.2. Photon Identification.	17
2.2.1. Definitions.	17
2.2.2. Efficiencies.	17
2.2.3. Scale factors and systematics.	20
2.2.4. Anti-track matching cut.	20
2.3. Electron Identification.	25
2.3.1. Definitions.	25
2.3.2. Efficiencies.	25
2.3.3. Scale factors and systematics.	26
<b>3. Conclusion.</b>	40
<b>4. Acknowledgments.</b>	40
<b>References</b>	41
<b>5. Appendix.</b>	42
5.1. Dependence of photon selection efficiencies on $L_{inst}$ , $R(e, jet)$ and $njets$ .	42
5.2. Dependence of electron selection efficiencies on $L_{inst}$ , $R(e, jet)$ and $njets$ .	44

## 1. INTRODUCTION.

This note describes the work done to measure electron and photon identification efficiencies as well as data/Monte Carlo scale factors for the p20 data and Monte Carlo samples. Here we follow identification techniques applied earlier during p17 electron [2] and photon [1] certifications.

Main challenge in the analyzed Run IIb data is caused by increased Tevatron instantaneous luminosity by a factor of two, on the average, as compared with Run IIa data sample. This fact resulted in degrading electron/photon efficiencies for similar definitions. Another challenge is caused by the new physics goals, primarily directed on the searches (and setting limits) for Higgs boson and new particles.

It necessitates some re-optimization of EM ID tools existed earlier in the p17 era in order to introduce ID variables and definitions less sensitive to the effects related with the luminosity increase and which, at the same time, would allow to preserve high signal efficiencies and low fake rates. For this reason some EM ID variables have been re-defined (like calorimeter isolation, spatial track matching, electron likelihood), some have been introduced for the first time (like electron and photon neural networks). Some new electron ID variables have been borrowed from the previous p17 photon ID certification [1].

Another problem that we tried to solve is low tracking efficiency in the forward detector region that degraded in Run IIb even more due a higher fraction of fake hits and track segments. Trying to keep electron efficiencies in this region as high as possible five new trackless electron definitions have been introduced that differ from each other by signal and fake efficiencies. A possible solution in near future is direct hit analysis using hits-on-the-road (likelihood based) method like it was done for the central region [13] with a possible account of forward per-shower (FPS) clusters in the shower layers.

A new part of p20 electron ID is certification of the inter-cryostat (ICR) detector region. It allows to increase electron acceptance in various searches by 20-25%. The ICR certification is beyond the scope of current note and can be found here [3].

### 1.1. Data Samples.

For both electron and photon ID we use  $Z \rightarrow ee$  events found in 2EMhighpt skim provided by Common Sample Group. Following skim have been used:

Pass2 (preshutdown 2007) data

CSG\_CAF\_2EMhighpt\_PASS2\_p21.10.00

Pass4 (postshutdown 2007)

CSG\_CAF\_2EMhighpt\_PASS4\_p21.10.00\_p20.12.00

CSG\_CAF\_2EMhighpt\_PASS4\_p21.10.00\_p20.12.01

CSG\_CAF\_2EMhighpt\_PASS4\_p21.10.00\_p20.12.02

Data in this skim require at least two EM objects with  $|ID|=10$  or  $11$  and having  $p_T > 7$  GeV. In analysis of signal events we have also required at least one of calorimeter single or double EM trigger to be fired.

For the fakes study, we have used these QCD skims:

CSG\_CAF\_QCD\_PASS4\_p21.10.00\_p20.12.00

CSG\_CAF\_QCD\_PASS4\_p21.10.00\_p20.12.01

CSG\_CAF\_QCD\_PASS4\_p21.10.00\_p20.12.02

Data in QCD skim require to have fired at least one of single or di-jet triggers with  $p_T$  threshold from 8 GeV and higher.

All the used data are in the CAF format and obtained with the new track matching constants [5] and electron likelihood provided by EM ID group based on analysis of p20 data.

### 1.2. Monte Carlo Samples.

For electron ID we have used  $Z \rightarrow ee$  events simulated with Pythia Monte Carlo event generator [6] in the invariant mass range of 60 – 130 GeV using these datasets:

CSG\_CAF\_MCv4-55052\_p21.11.00

CSG\_CAF\_MCv4-55052\_p21.11.00

For photon IDs we have used photon+jet and di-photon samples from following datasets:

CSG\_CAF\_MCv4-66353\_p21.11.00

CSG\_CAF\_MCv4-66354\_p21.11.00

CSG\_CAF\_MCv4-66355\_p21.11.00

CSG\_CAF\_MCv4-66356\_p21.11.00

CSG\_CAF\_MCv4-66357\_p21.11.00

for photon+jet events (which correspond to simulations with  $\hat{p}_\perp^{min}$  from 10 to 320 GeV) and

CSG\_CAF\_MCv4-80052\_p21.11.00

CSG\_CAF\_MCv4-80053\_p21.11.00

for di-photon events.

## 2. OBJECT IDENTIFICATION.

To derive photon and electron efficiencies in p20 data we used MC and data samples listed in sections 1.1 and 1.2.

Any event we consider should contain at a primary vertex with  $|z| < 60$  cm with at least three matched tracks. *Tag and probe method* has been used for analysis of  $Z \rightarrow ee$  events described, for example, in [1, 2].

For probe electron we have used following basic requirements:

1. EM cluster with  $p_T > 10$  GeV
2.  $|\text{ID}| = 10$  or  $11$
3.  $\text{Iso} < 0.15$
4.  $\text{EMfrac} > 0.90$
5.  $\text{IsoHC4} < 6$  GeV
6.  $\text{SigPhi} < 20$ .

The meaning and notations for the variables are described in section 2.1. Probe electron can be located in either CC ( $|\eta| < 1.1$ ) or EC ( $1.5 < |\eta| < 3.2$ ) region. Definition of the probe coincides with those used earlier in [1, 2] and differs by two new last lines. Those cuts are almost 100% efficient (see section 2.1) and have been introduced for additional background suppression. Efficiencies of all the cuts under study are calculated with respect to those definitions which are also called preselection criteria. The preselection efficiencies for p20 data are measured separately [7] and not described in this note. They are pretty close to those found earlier in p17 electron certification [2].

For the tag electron we require

1. EM cluster with  $p_T > 25$  GeV
2.  $|\text{ID}| = 10$  or  $11$
3.  $\text{Iso} < 0.07$
4.  $\text{EMfrac} > 0.97$
5.  $\text{HMx7} < 20$
6.  $\text{IsoHC4} < 2$  GeV
7.  $\text{Lhood8} > 0.8$
8.  $\text{ANN-7 output} > 0.7$ .

Both the tag and the probe electron are required to give invariant mass  $80 < M_{inv}(e, e) < 100$  GeV and to be located in the  $\eta$  fiducial region.

Also, fake sample in data should

- contain at least one probe EM object,
- which is separated from any *good* jet by  $dR(\eta, phi) > 0.7$
- have at least one *good* jet with the *jetcorr* corrected  $p_T > 15$  GeV that have  $\Delta\phi(e, jet) > 2.8$
- have missing  $E_T < 10$  GeV
- have invariant mass of the probe EM object with any other EM object (if it exists)  $m_{inv} < 60$  or  $m_{inv} < 130$  GeV.

## 2.1. Choice of selection variables.

Electrons and photons in most of final states of the hard interactions are produced as single well-isolated objects. For this reason, electron/photon ID variables and criteria should reflect this nature for all sub-detectors.

Most of the criteria have been used in the previous electron/photon IDs [1, 2] and some of them are new ones or obtained by re-optimization/re-training using p20 data [3-5, 8, 10].

For electron and photon ID we use following variables:

- Calorimeter isolation *Iso*. It is obtained from a regular calorimeter isolation after subtraction of the energy coming to the EM cluster isolation cone ( $R = 0.4$ ) due to additional minimum bias interactions and parametrized versus instantaneous luminosity [10]. Figure 1 shows electron efficiencies in data and MC events for the relative isolation cut  $\text{Iso} < 0.07$  after the energy corrections in CC and EC regions. The

plots also show the fake rates <sup>1</sup>. We see that MC and data efficiencies are in a very good agreement and show stable behavior versus instantaneous luminosity.

The new isolation is available by calling method *Isolation()* in both, Electron and Photon selectors.

- Fraction of the EM cluster energy deposited in the electromagnetic calorimeter *EMfr*. It is available in the standard EMreco via method *emfrac()*.
- Track isolation of the EM cluster which is calculated in Electron and Photon selectors in method *trk\_iso()*. It returns total track  $p_T$  (for tracks with  $p_T^{track} > 0.5$  GeV) in the hollow cone  $0.05 < R < 0.4$  around the EM cluster and denoted by *IsoHC4*. Similar variable *IsoHC7* is calculated for a cone of  $0.05 < R < 0.7$  <sup>2</sup>.
- HMx7 and HMx8 that characterize lateral and longitudinal shower shapes of EM cluster, which are available via *HMx7()* and *HMx8()* methods.
- Energy (squared) weighted CPS strips width *cps\_rms* (and *cps\_sq\_rms*), that was studied in detail in this note [11] and used in p17 photon ID [1]. It is available now through *E\_RMS\_CPS()* (and *E2\_RMS\_CPS()*) methods of *TMBEMCluster*.
- Shower widths of the EM cluster at 3rd layer of the EM calorimeter in  $r-\phi$  and  $r-z$  planes, respectively, *SigPhi* and *SigZ* (called by *flrS1(3)* and *flrS2(3)* methods).
- Probability of spatial track matching, implemented in *track\_match\_spatialchi2prob()* method. This probability has been recalculated for p20 data by using new values of  $\sigma_\phi$  and  $\sigma_\eta$  and average shifts between the track extrapolated to EM3 layer of EM cluster and gravity center of EM cluster at this layer. Those values are calculated separately positive and negative signs of the product  $q_{track} \cdot B_z$  (i.e. for different track curvatures). Due this splitting we can reject more jet fakes without loosing a signal efficiency. New dependences on  $p_T$  and  $\eta$  of EM object has been found and implemented as well. Figures 2 and 3 show efficiencies to pass cut *track\_match\_spatialchi2prob()*  $> 0.001$  for  $Z \rightarrow ee$  and di-jet events in data. One can see that additional factor of about 1.25 – 1.5 per electron candidate is achieved with the new track matching constants. More details can be found in this D0 Note [5].
- “Hits on the road” discriminant [13] *HoR*.
- Electron likelihood trained using p20 data [4]. Here we use the 8-variable likelihood *Lhood8()*.
- Artificial neural networks, trained using p20 data for electron and photon ID [8]. Here we use electron ANN with 7 input variables for CC and ANN with 3 input variables for EC regions. We have also introduced photon ANN with 6 input variables (same as used in ANN-7 for electron ID). Certification of those neural nets is described in [8] and [9]. Figs. 4 and 5 show efficiency of signal (electrons in MC and data) and background events versus cuts on the ANN output for ANN-7 in the CC and ANN-3 in the EC regions. All the neural networks are available in both Electron and Photon selectors.

Efficiency versus cuts on those variables in CC and EC regions are presented in Figs.6–10. These plots tell us which cuts would be optimal for a given variable and explain our choices made for electron and photon ID definitions in sections 2.2 and 2.3. The only exclusions are *SigPhi* and *SigZ* variables, choice of the cuts for which in EC region is motivated by their  $|\eta_{det}|$  dependence, found during photon ID certification using p17 data [1] <sup>3</sup> Those dependencies turned out to be still useful for p20 data. We used following parameterizations for these variables in the EC region

$$SigPhi < 7.3 * \eta_{det} * \eta_{det} - 35.9 * |\eta_{det}| + 45.7 \quad (1)$$

<sup>1</sup> All the efficiencies are calculated with respect to the preselection cuts.

<sup>2</sup> Not used for ID due to a smaller signal efficiency. Nut it was used for JES purposes to select tight photon candidates [12].

<sup>3</sup> Due to decreasing *SigPhi* and *SigZ* and narrowing their distributions with growing  $|\eta_{det}|$  we have to apply tighter cuts for bigger  $|\eta_{det}|$ .

$$SigZ < 7.5 * \eta_{det} * \eta_{det} - 36.0 * |\eta_{det}| + 44.8 \quad (2)$$

While for CC region a straight cut should be applied for *SigPhi* (and no useful cut for *SigZ*).

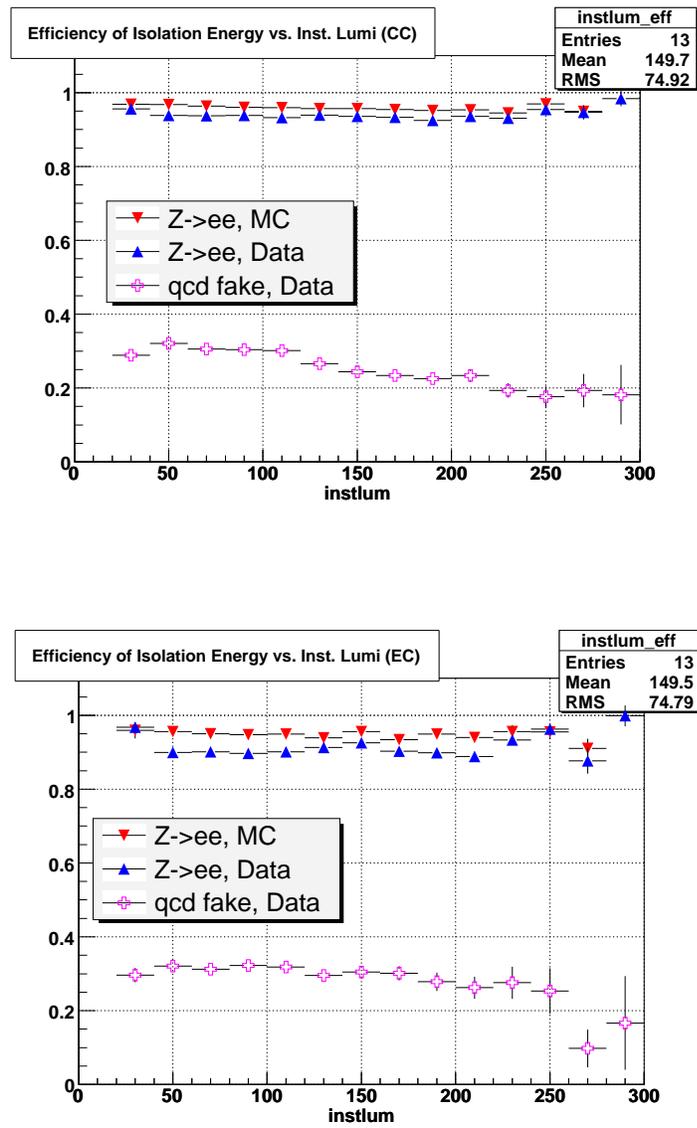


Figure 1: Electron efficiencies to pass the cut  $\text{Iso} < 0.07$  in data and MC in CC (top) and EC (bottom) regions.

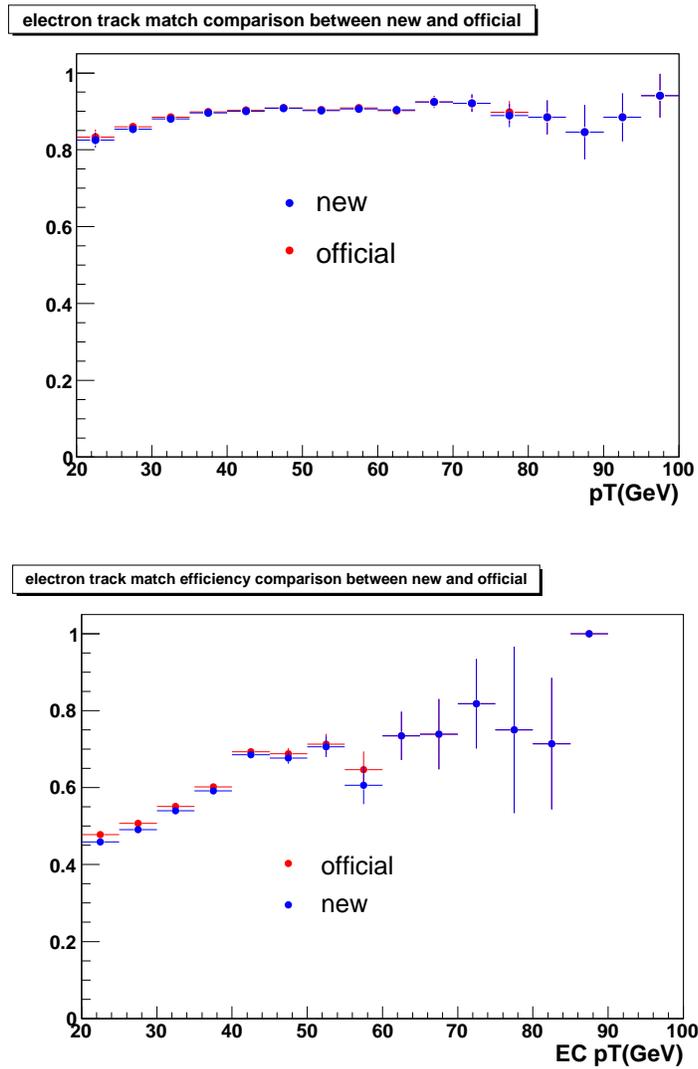


Figure 2: Efficiencies to pass cut  $track\_match\_spatialchi2prob() > 0.001$  for  $Z \rightarrow ee$  events in data in CC (top) and EC (bottom) regions.

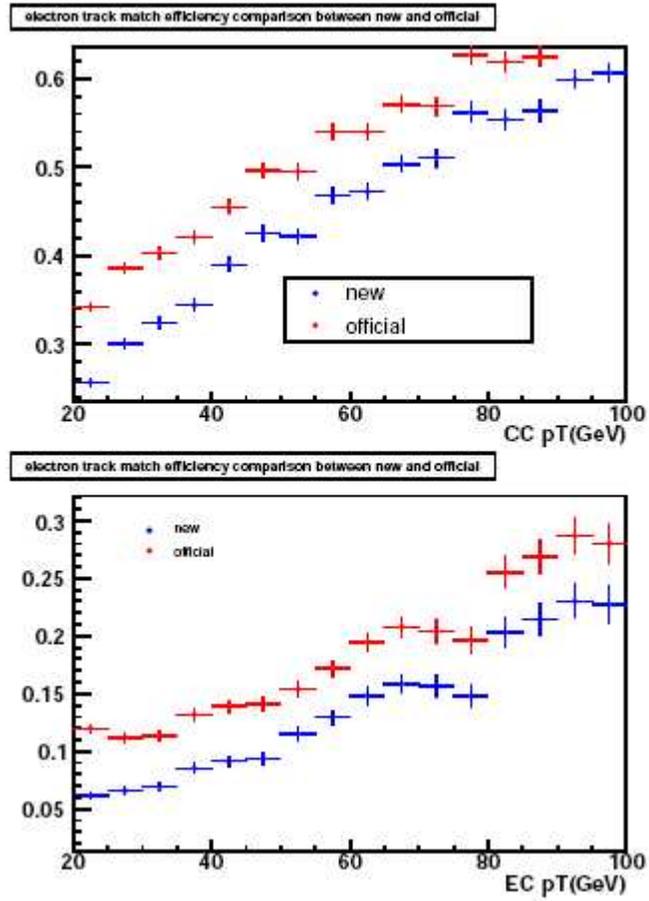


Figure 3: Same as in Fig. 2 but for QCD jet events in data.

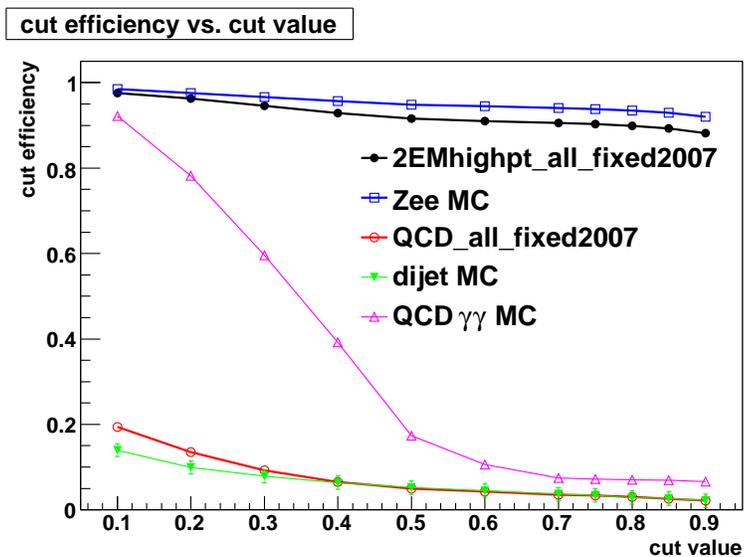


Figure 4: Efficiency vs. cut value for ANN-7 for signal and background events in the CC region.

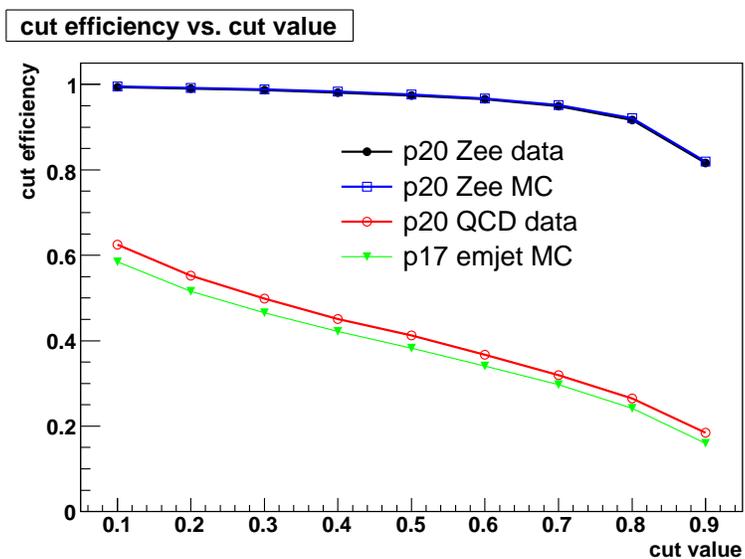


Figure 5: Efficiency vs. cut value for ANN-3 for signal and background events in the EC region.

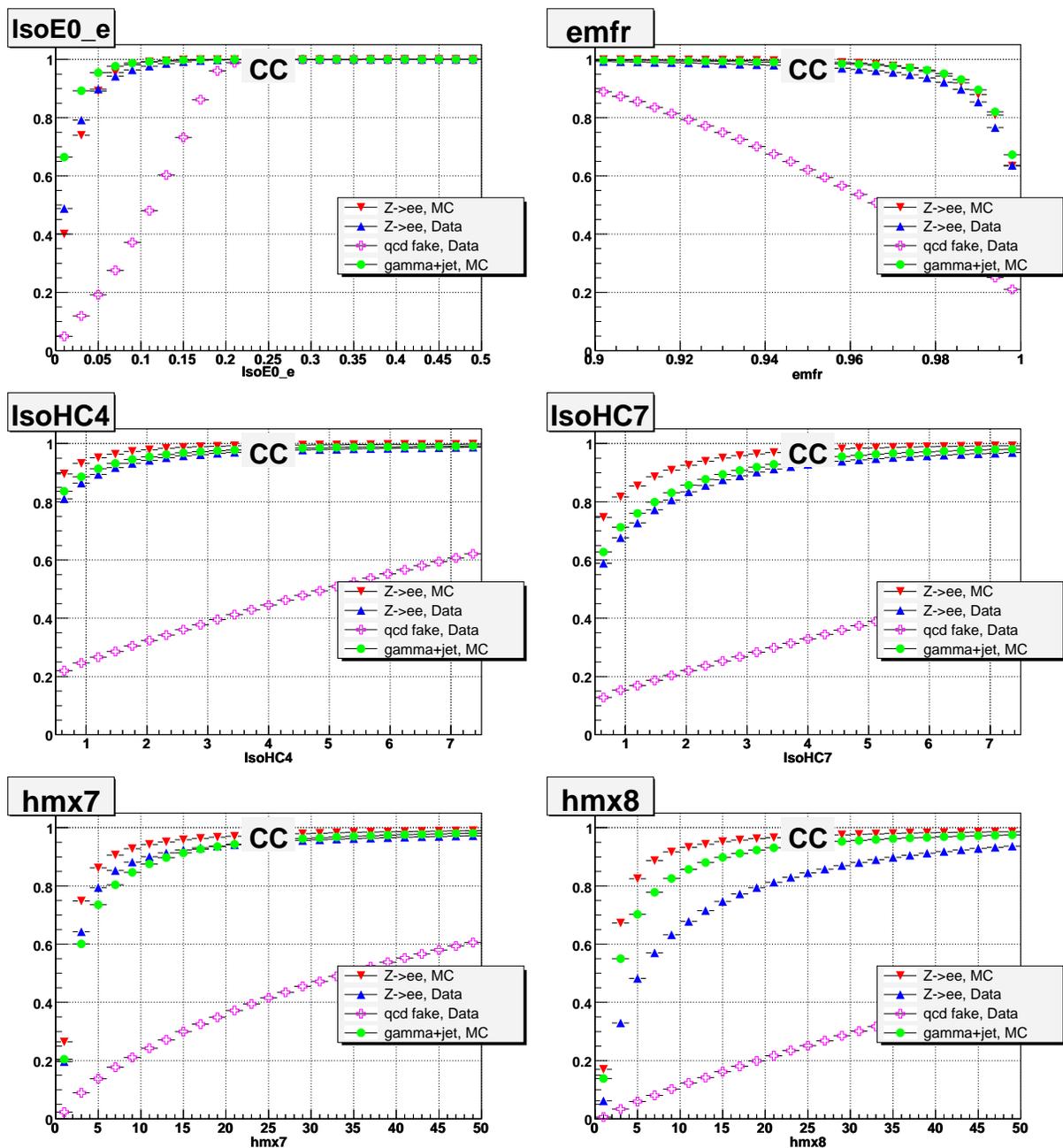


Figure 6: Efficiency vs. cut on fractional isolation ( $IsoE/E$ ), EM fraction ( $emfr$ ), track isolation ( $IsoHC4$  and  $IsoHC7$ ), and H-matrix ( $hmx7$  and  $hmx8$ ) in the CC region.

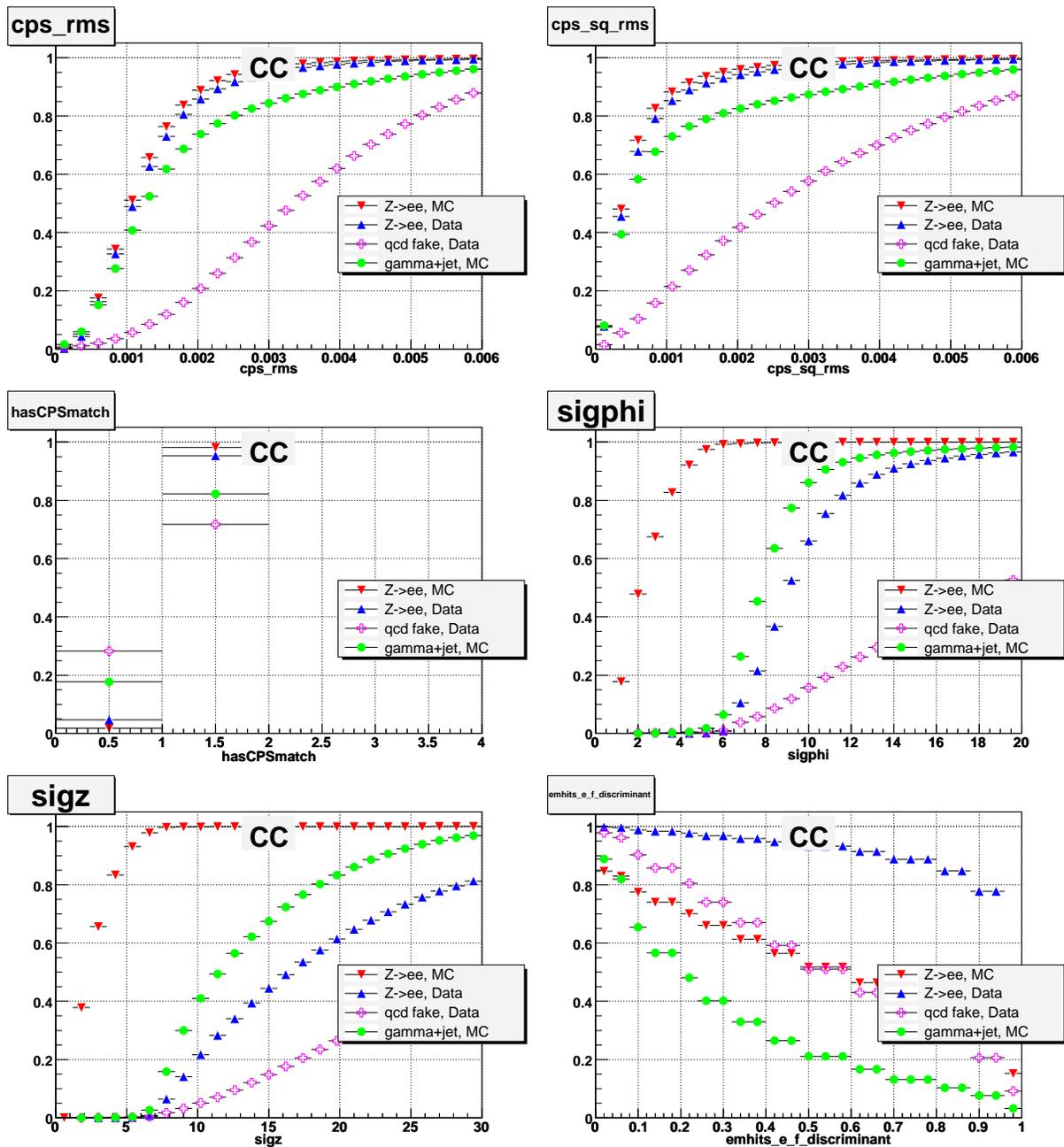


Figure 7: Efficiency *vs.* cut on energy weighted and energy squared weighted RMS of the re-mapped CPS cluster ( $cps_{rms}$  and  $cps_{sq_{rms}}$ ), existence of the CPS match ( $hasCPSmatch$ ), squared EM cluster width in  $r \times \phi$  and  $r \times z$  space in the third layer of the EM calorimeter ( $SigPhi$  and  $SigZ$ ), and the EM/fakes discriminant ( $emhits\_e\_f\_discriminant$ ) in the CC region.

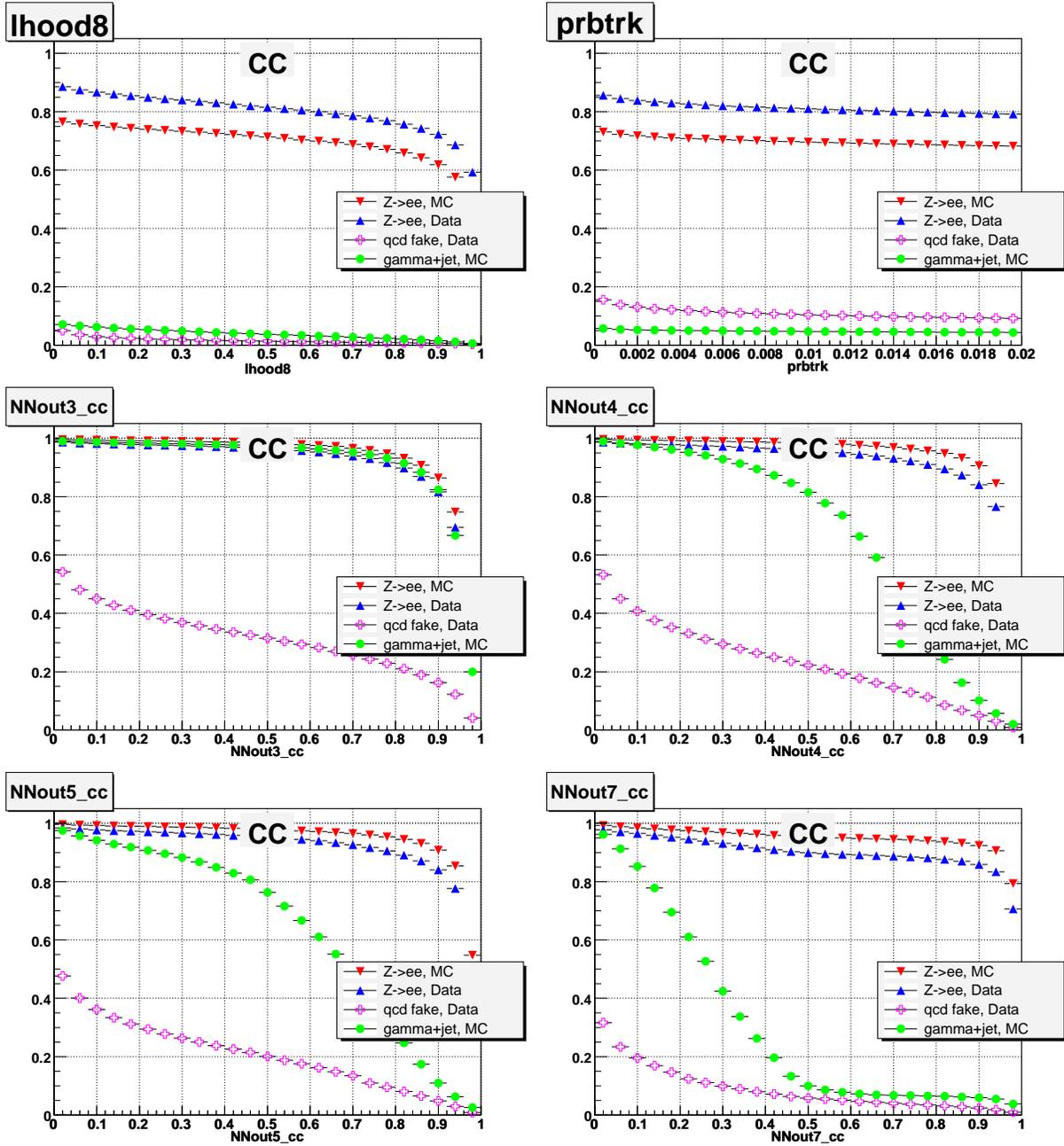


Figure 8: Efficiency *vs.* cut on EM likelihood ( $LHood8$ ), spatial match  $\chi^2$  probability of a track ( $prbtrk$ ), 3-, 4-, 5- and 7-parametric Neural Net variables ( $NNout3_{cc}$ ,  $NNout4_{cc}$ ,  $NNout5_{cc}$  and  $NNout7_{cc}$ ) in the CC region,

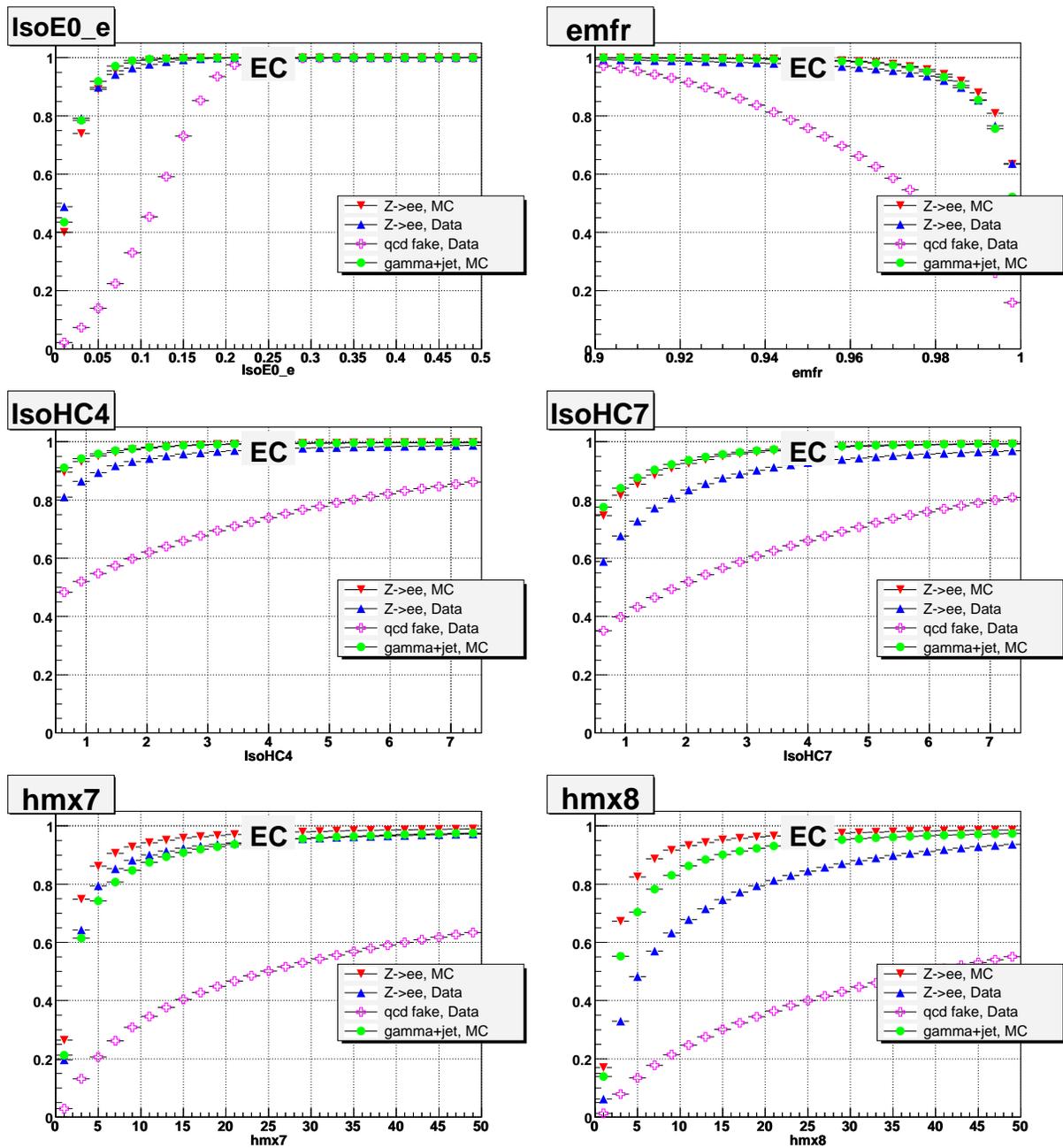


Figure 9: Efficiency *vs.* cut on fractional isolation ( $IsoE/E$ ), EM fraction ( $EMfr$ ), track isolation ( $IsoHC4$  and  $IsoHC7$ ), and H-matrix ( $HMx7$  and  $HMx8$ ) in the EC region.

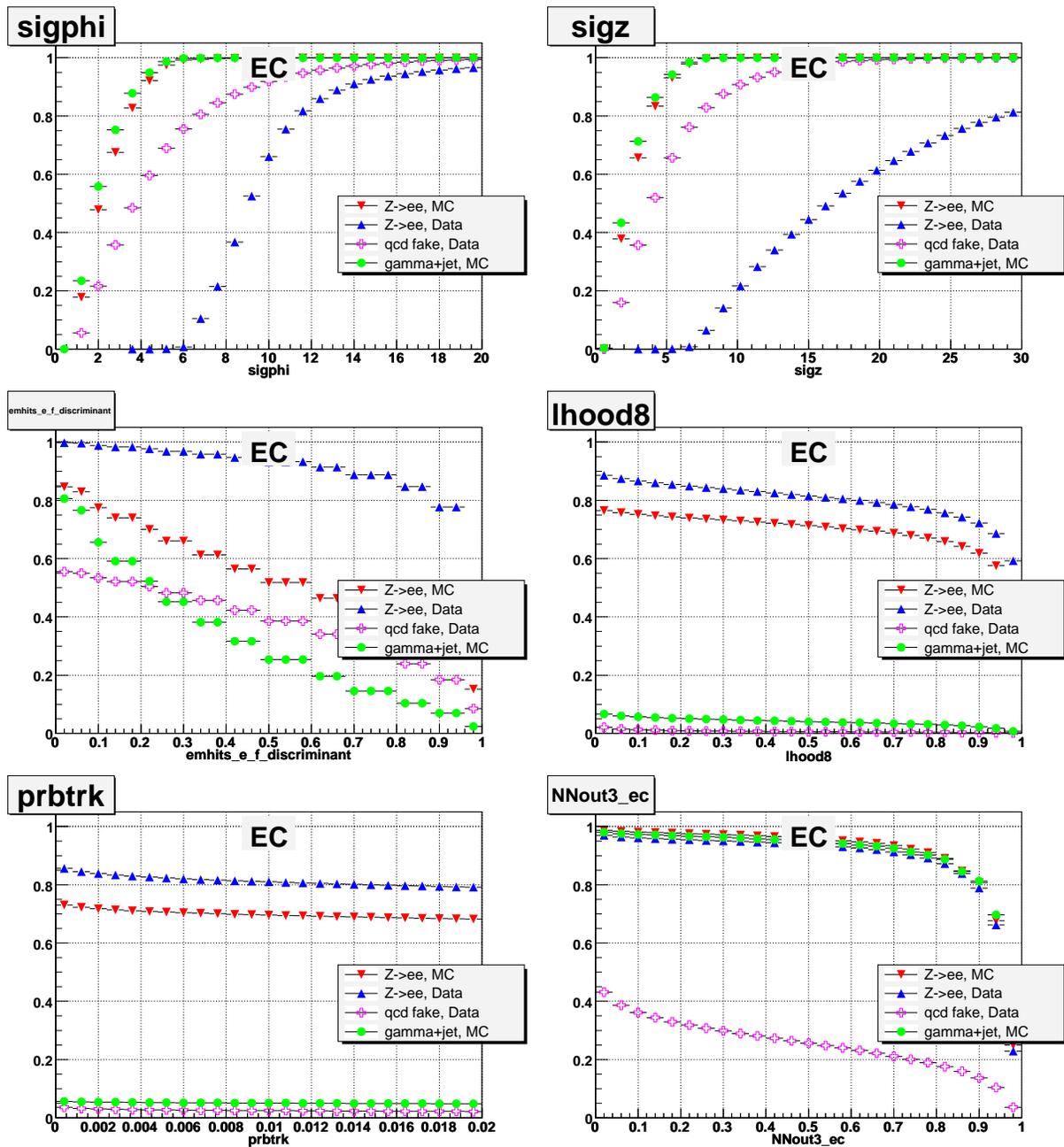


Figure 10: Efficiency *vs.* cut on squared EM cluster width in  $r \times \phi$  and  $r \times z$  space in the third layer of the EM calorimeter (*sigphi* and *sigz*), EM/fakes discriminant (*emhits\_e\_f\_discriminant*), EM likelihood (*lhoo8*), spatial match  $\chi^2$  probability of a track (*prbtrk*), and 3-parametric Neural Net in the EC region.

## 2.2. Photon Identification.

### 2.2.1. Definitions.

To detect photons with p20 data the two sets of base (core) definitions in the CC and EC regions are suggested. In the both, CC and EC regions we call them Core1 and Core2. The sets of selection criteria for them are comprised in Table I.

Table I: Photon definitions for CC and EC regions.

Variables	Core1(CC)	Core2(CC)	Core1(EC)	Core2(EC)
Iso <	0.10	0.07	0.10	0.07
EMfr >	0.95	0.95	0.95	0.95
IsoHC4 <	2.5	2.0	2.0	1.5
Hmx8 <	—	—	—	10
<i>SigPhi</i> <	18	16	eq.(1)	eq.(1)
<i>SigZ</i> <	—	—	eq.(2)	eq.(2)

The choice of the variables and selection cuts has been optimized using “ $\gamma$ +jet” MC and  $Z \rightarrow ee$  events in MC, data and QCD jet data. In the suggested definitions we tried to keep high selection efficiencies with small corrections on data/MC difference and low background.

Basically, these p20 definitions are very close to those chosen for p17 Photon ID [1] with some variations of the cuts on the used variables (plus added Hmx8 for Core2 in EC).

Efficiencies for those definitions are presented in section 2.2.2.

### 2.2.2. Efficiencies.

Figures 11 and 12 show efficiencies for for Core1 and Core2 definitions for electrons in data and MC, QCD fakes and MC photons as functions of  $p_T$ ,  $\eta_{det}$  and  $\phi_{det}$  in the CC and EC regions, respectively. Note that the QCD fake rates being preselected with cuts eq.(8) should contain real direct photons which tend to increase efficiency of the mixture. As we see, the electron efficiencies between Core1 and Core2 definitions in CC differ by 3% (7%) at  $p_T=40$  (25) GeV while fake rates differ by approximately factor of two. Analogous difference in EC region is 7% (10%) at  $p_T=40$  (25) GeV with fake rates differing by approximately factor of 1.5.

Figures 28 and 29 of Appendix 5.1 also show efficiencies versus instantaneous luminosity, distance to a closest jet ( $dR$ ) and the number of jets in the event for the both definitions. The luminosity dependence in CC is flat, while there is about 5% drop in efficiency between very low and very high luminosities for Core2 definition in EC. All the definitions (maybe excepting Core1 in CC) show the  $dR$  dependence resulting in about 4-7% drop of efficiencies from  $dR = 3.5 - 4.0$  to  $dR = 0.7$  in both data and MC. And all the definitions show strong dependence on the number of jets. We may also see that these dependencies are very well modeled in MC  $Z \rightarrow ee$  events.

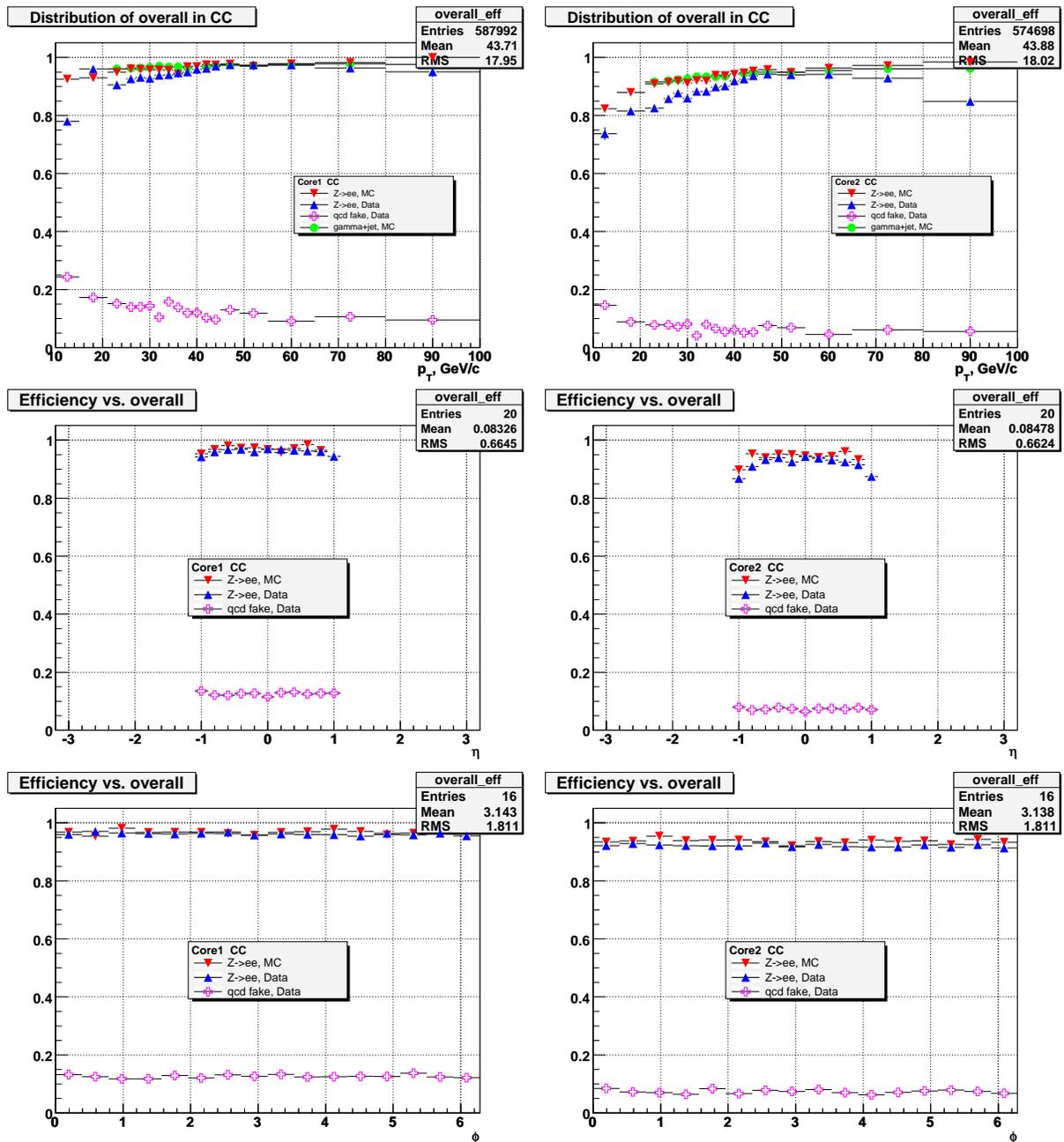


Figure 11: Efficiency for Core1 (left column) and Core2 (right column) definitions for electrons in data and MC, QCD fakes and MC photons as functions of  $p_T$ ,  $\eta_{det}$  and  $\phi_{det}$  in the CC region.

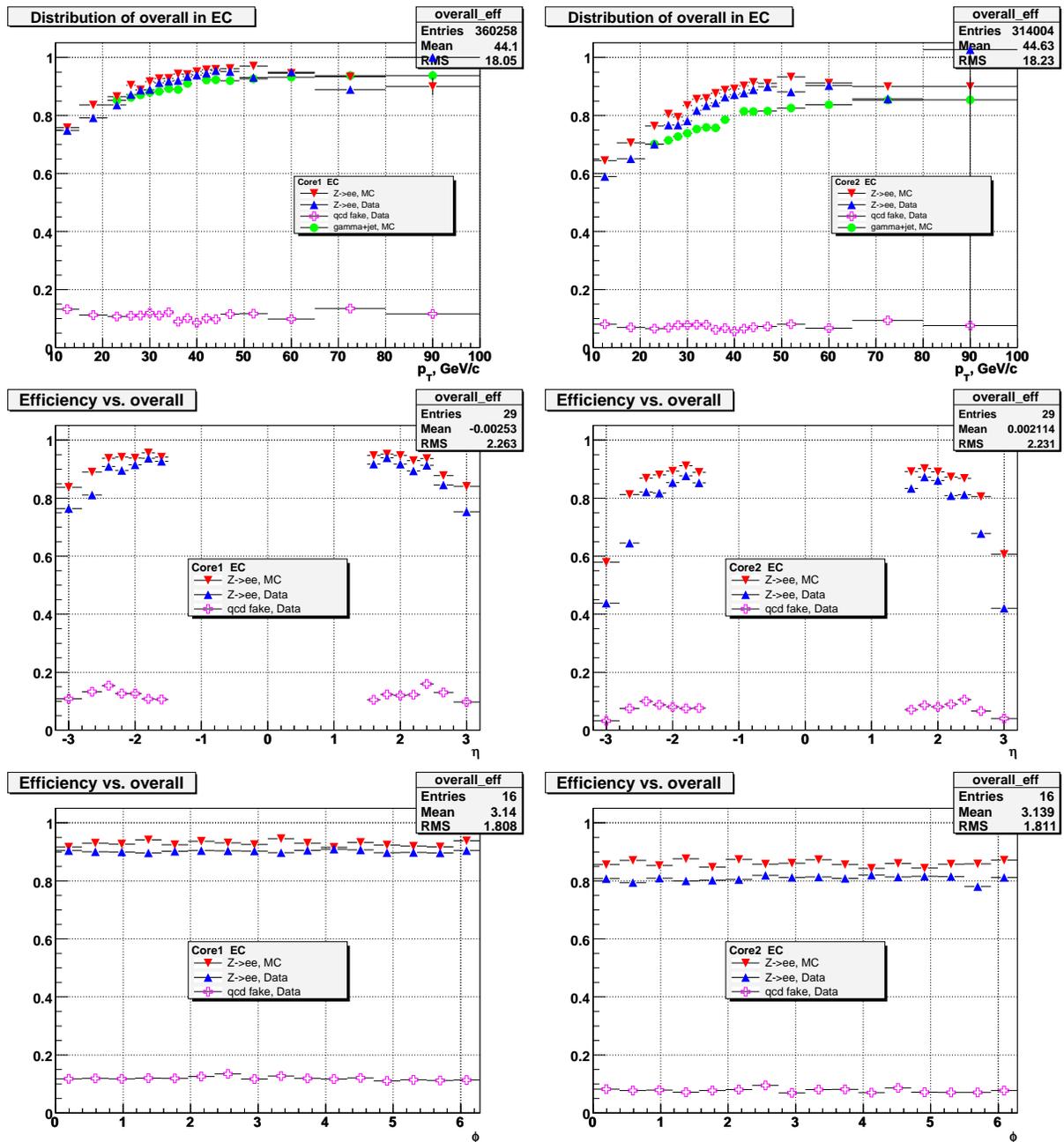


Figure 12: Efficiency for Core1 (left column) and Core2 (right column) definitions for electrons in data and MC, QCD fakes and MC photons as functions of  $p_T$ ,  $\eta_{det}$  and  $\phi_{det}$  in the EC region.

Table II: Fit parameters for electron efficiencies in data and MC for Core1 and Core2 definitions.

	Core1 CC	Core2 CC	Core1 EC	Core2 EC
$p_0^{data}$	4.50811	9.41755	32.0539	52.9853
$p_1^{data}$	-1.24265	-1.24425	-1.65435	-1.57038
$p_0^{MC}$	1.10466	2.47171	20.9924	39.9682
$p_1^{MC}$	-0.945545	-0.997348	-1.59897	-1.56245

### 2.2.3. Scale factors and systematics.

Here we present data/MC correction factors (also called ‘‘scale factors’’) found as a function of  $p_T$  for each of the definition separately in CC and EC. To get them we are following method used in [1]. Specifically, for this aim first we are fitting efficiencies versus  $p_T$  for data and MC electrons and then divide the fits to get the scale factor (SF) for a given definition in a given region.

Figures 13 and 14 show efficiency fits for Core1 and Core2 photon definitions (left column), and data/MC ratios (right column) for electrons as functions of  $p_T$  in the CC region.

To fit the  $p_T$  efficiencies following functional forms have been used:

$$\epsilon = \frac{1}{1 + p_0 \cdot p_T^{p_1}}. \quad (3)$$

Fit parameters for function (3) for the electron efficiencies in data and MC for all the definitions are summarized in Table II. They can be used in the user codes to get correction factor for a chosen definition using this ratio:

$$\text{SF} = \frac{\epsilon^{data}}{\epsilon^{MC}} = \frac{1 + p_0^{MC} \cdot p_T^{p_1^{MC}}}{1 + p_0^{data} \cdot p_T^{p_1^{data}}}. \quad (4)$$

Main uncertainties for this SF come from the errors on the fits for electron in data and MC shown in Figures 13 and 14. We also add in quadrature uncertainties coming from the background subtraction procedure which we took 1% for MC and 1.5% for data. The values of the final uncertainties for SFs found using eq.(II) are shown in Table III. We see that the uncertainties drop fast from about 8 – 10% at  $p_T = 10$  GeV to about 4% at  $p_T = 20$  GeV and become almost stable (2.3 – 2.5%) for  $p_T \geq 50$  GeV.

### 2.2.4. Anti-track matching cut.

For photon ID users also apply requirement of absence of any spatially matched track to the photon EM cluster. This requirement is usually expressed as

$$\text{track\_match\_spatialchi2prob}() < 0.001 \quad (5)$$

(section 2.1). Efficiency for this cut for MC photons is shown in Figure 15 for CC and EC regions.

As we see, the efficiency for this cut drops by 2 – 3.5% at small  $p_T^\gamma$  and practically stable at  $p_T^\gamma > 35$  GeV with  $\epsilon_{atr}^{mc} = 0.94 \pm 0.01$ .

To make a confirmation for this number in data, we have used photons from  $Z\gamma \rightarrow ee\gamma$  and  $Z\gamma \rightarrow \mu\mu\gamma$  event samples selected by tight cuts described in [8]. We found following efficiencies:

Table III: Relative uncertainties on the scale factors for the four photon definitions.

$\langle p_T \rangle$ , GeV/c	Core1 (CC)	Core2 (CC)	Core1 (EC)	Core2 (EC)
	relative uncertainty d(SF)/SF			
10	0.087	0.076	0.100	0.101
13	0.061	0.056	0.068	0.074
18	0.035	0.036	0.037	0.044
23	0.028	0.029	0.029	0.034
26	0.026	0.028	0.027	0.031
28	0.025	0.027	0.027	0.030
30	0.025	0.027	0.026	0.029
34	0.025	0.026	0.026	0.028
40	0.025	0.026	0.025	0.027
52	0.025	0.025	0.025	0.026
60	0.025	0.025	0.024	0.026
73	0.025	0.025	0.024	0.025
90	0.025	0.025	0.024	0.025
100	0.024	0.025	0.023	0.024

$p_T^\gamma > 10$  ( $\langle p_T^\gamma \rangle = 18.7$ ) GeV:  $\epsilon_{atrak}^{data} = 0.951 \pm 0.014$ ,

$p_T^\gamma > 15$  ( $\langle p_T^\gamma \rangle = 22.9$ ) GeV:  $\epsilon_{atrak}^{data} = 0.938 \pm 0.019$ ,

$p_T^\gamma > 25$  ( $\langle p_T^\gamma \rangle = 30.5$ ) GeV:  $\epsilon_{atrak}^{data} = 0.923 \pm 0.043$ .

The efficiencies in data and MC are same within statistical uncertainties but the average efficiency numbers of  $\epsilon_{trkmatch}^{data}$  and  $\epsilon_{trkmatch}^{mc}$  differ for same  $p_T^\gamma$  bins by 1 – 2.5%. The small difference can be caused by admixture of remaining QCD jets [14] that have less efficiency [15]. To be conservative, we suggest to take 2% as a systematic uncertainty for the antitrack matching cut (5).

Also,  $\epsilon_{trkmatch}^{data}$  obtained for  $p_T^\gamma > 10$  GeV in p17 and p20 data using  $Z\gamma \rightarrow \mu\mu\gamma$  sample do not differ within uncertainties with  $0.938 \pm 0.018$  in p17 and  $0.935 \pm 0.015$  in p20 data.

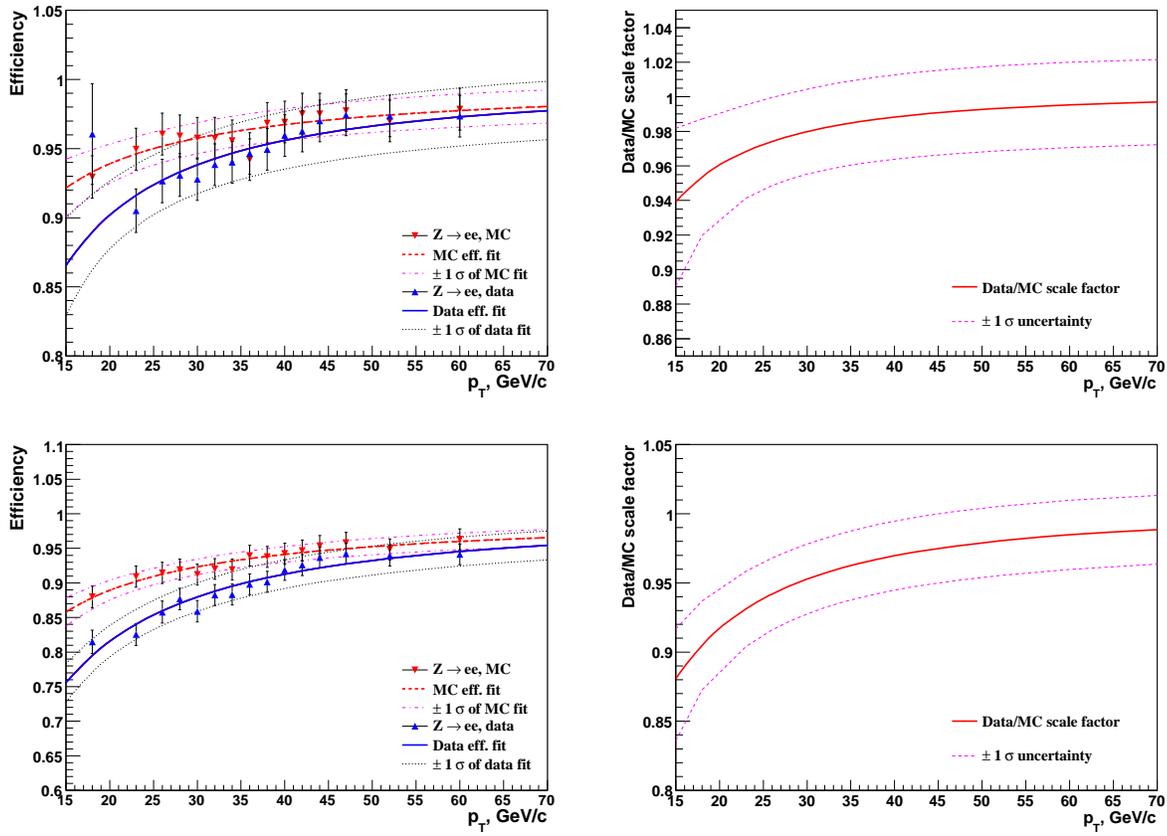


Figure 13: Efficiency fits for Core1 and Core2 photon definitions, and data/MC ratios for electrons as functions of  $p_T$  in the CC region.

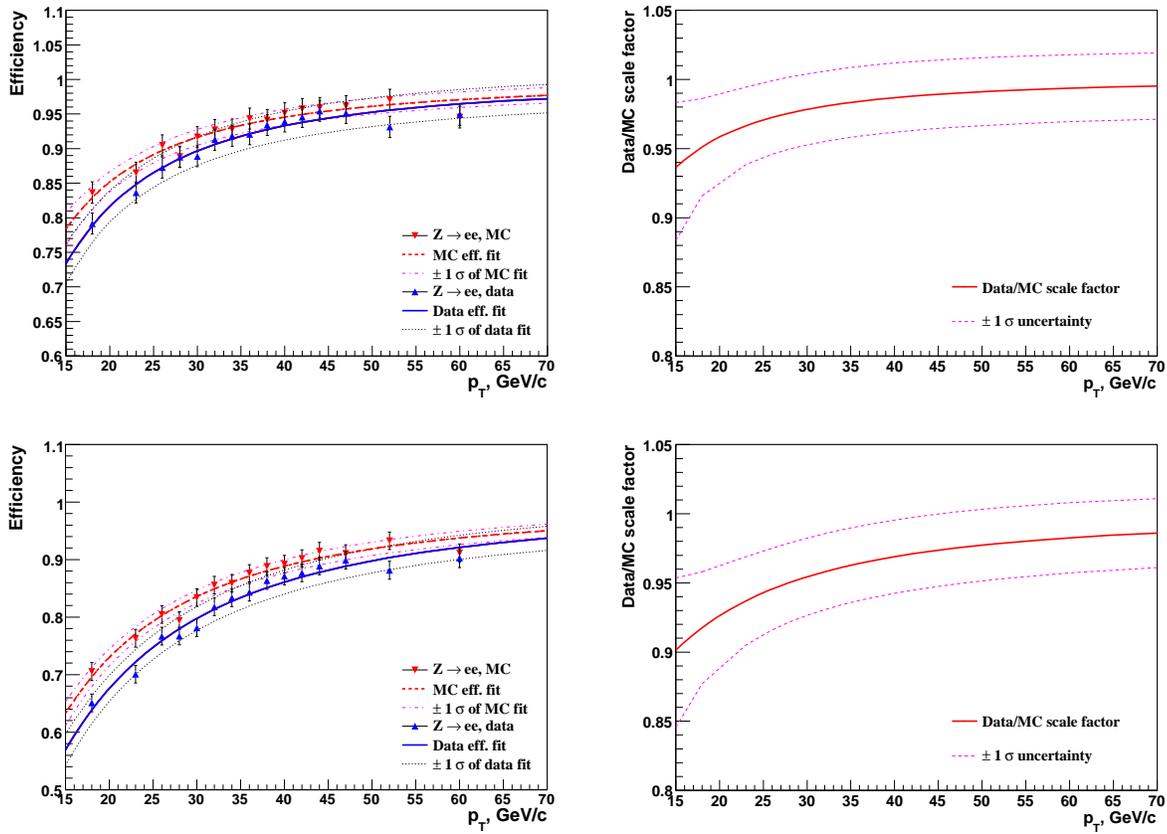


Figure 14: Efficiency fits for Core1 and Core2 photon definitions, and data/MC ratios for electrons as functions of  $p_T$  in the EC region.

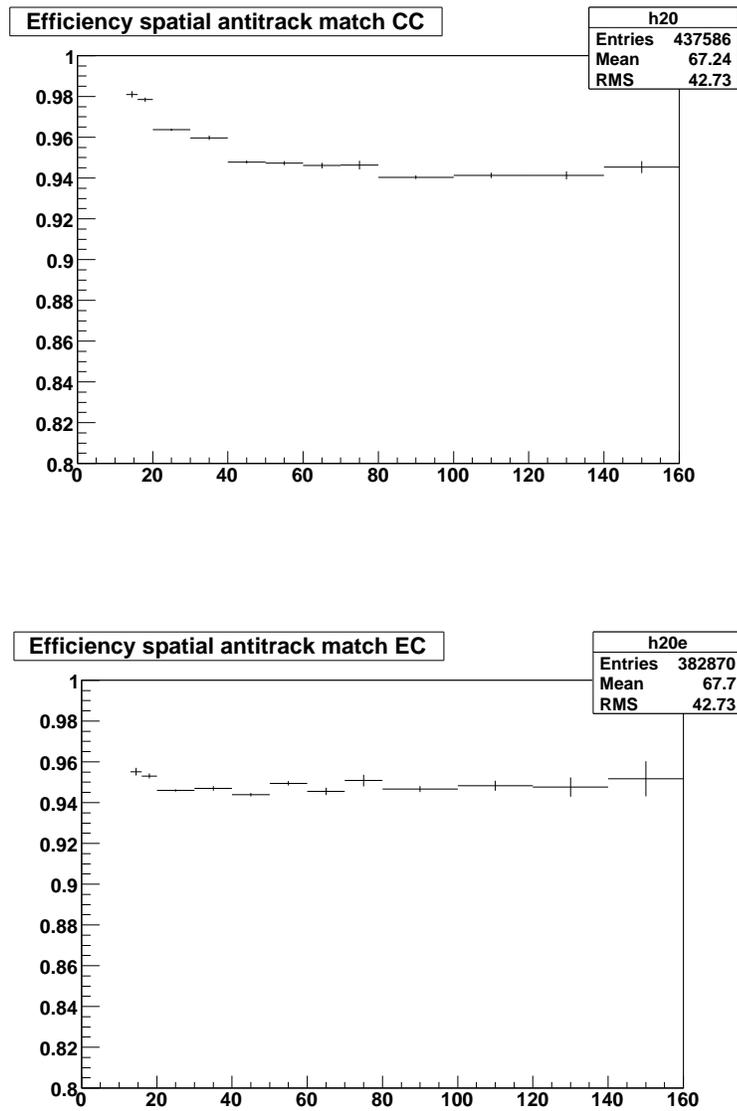


Figure 15: Efficiencies to pass cut  $track_{match\_spatialchi2prob} < 0.001$  for MC photons in CC (top) and EC (bottom) regions.

### 2.3. Electron Identification.

#### 2.3.1. Definitions.

Using selection variables described in section 2.1., we have formulated a few new electron definitions for CC, EC and ICR regions. Electron ID for ICR region is not covered in this note and discussed separately in [3]. All the electron definitions for CC and EC regions are summarized in Tables IV and V.

In most cases the definitions are built by tightening cuts on the variables or adding new variables with corresponding criteria for them. Very loose definition “VLoose” contains set of four basic cuts for both CC and EC regions:  $Iso < 0.10$ ,  $EMfr > 0.96$ ,  $IsoHC4 < 3.5$  GeV and  $HMx7(8) < 35$ . “Loose” differs from “VLoose” definition in CC by adding requirements on track matching, ORed with hits-on-the-road cut for EM clusters that do not have matched track, adding cut on the electron neural net output (ANN-7) and tightening the track isolation  $IsoHC4$ . But there are exclusions when cuts on some variables are replaced by cuts on some other variables, for example, when we go from “MLoose2” to “Medium” (or “Tight”) the track matching requirement is replaced by the electron likelihood cut  $LHood8 > 0.2$  for “Medium” (or  $> 0.8$  for “Tight”). As one can see from Table V, five of six definitions in EC are trackless and differ by a tightness of calorimeter based cuts and track isolations. Just “Tight” definition requires a spatially matched track with probability  $> 0.001$ .

The definitions built in such a way to keep have difference in electron efficiencies by 4 – 7% between any two adjacent definitions and keeping fake rate as low as possible for such a change in signal efficiency. Exclusions are differences between trackless and track definitions in EC (“Medium” and “Tight”), where total efficiency dropped much more (15 – 30%) due to a low tracking efficiency in EC. That is the main reason why, in a difference with the CC region, we used five trackless electron definitions in EC, trying to keep the efficiency as high as possible.

Table IV: Electron definitions for CC region.

Variables	VLoose	Loose	MLoose1	MLoose2	Medium	Tight
Iso <	0.10	0.10	0.10	0.07	0.07	0.07
EMfr >	0.95	0.95	0.97	0.97	0.97	0.97
IsoHC4 <	3.5	3.0	2.5	2.5	2.5	2.5
HMx7 <	35	35	25	25	25	25
NNout7_cc >	—	0.2	0.2	0.6	0.6	0.6
LHood8 >	—	—	—	—	0.2	0.8
TrkM >	—	0.001	0.001	0.001	0.001	0.001
or HoR >	—	0.4	0.4	0.5	0.5	0.5

#### 2.3.2. Efficiencies.

Figures 16–21 show efficiencies for for six electron definitions, shown in Tables IV and V, for electrons in data and MC. as functions of  $p_T$ ,  $\eta_{det}$  and  $\phi_{det}$  in the CC and EC regions. Similarly to the photon case, all the  $p_T$  dependences have been fitted using the functional form (3) and contain also the  $1\sigma$  error bands caused by the statistical uncertainties and the background subtraction procedure. One can see that difference between MC and data efficiencies for the first five definitions in CC (“Loose” – “Medium”) never exceeds 10%, being smallest for “VLoose” and “Loose” definitions of just 1-3%. It becomes a bit higher

Table V: Electron definitions for EC region.

Variables	VLoose	Loose	MLoose1*	MLoose2*	Medium*	Tight*
Iso <	0.10	0.10	0.10	0.07	0.05	0.07
EMfr >	0.95	0.95	0.95	0.95	0.97	0.97
IsoHC4 <	3.5	2.0	2.0	2.0	1.0	2.0
HMx8 <	35	20	15	15	10	15
NNout3_ec >	—	0.4	0.4	0.4	0.4	0.4
TrkM >	—	—	—	—	—	0.0

\* Electron definitions MLoose1, MLoose2, Medium and Tight additionally also have cuts on the shower shapes  $SigPhi$  and  $SigZ$  [eq.(1), and eq.(2)], which are not shown in this table.

for “Tight” criterion and reaches 15% for small  $p_T$ ’s. The same is true for the five trackless definitions in EC. Just for “Tight” cuts the noticeable data/MC corrections are required.

Figures 30 – 35 of Appendix 5.2 also show efficiencies versus instantaneous luminosity  $L$ , distance to a closest jet ( $dR(e, jet)$ ) and the number of jets for all the definitions in CC and EC. All the CC/EC definitions, excepting “VLoose” and “Loose”, show a luminosity dependence, the stronger the tighter cuts are used: the difference between lowest and highest lumi is just about 4 – 6% for “MLoose1” and about 20% for the definitions containing electron likelihood cuts, “Medium” and “Tight” in CC and “Tight” in EC. But the luminosity dependencies observed in data is very well reproduced in MC modeling. For most of the definitions we also observe dependence on  $dR(e, jet)$  and very strong dependence on  $n_{jets}$ . A good thing is that all the dependencies are very well reproduced in MC samples.

### 2.3.3. Scale factors and systematics.

In this section we present data/MC scale factors vs.  $p_T, \eta_{det}$  and  $\phi_{det}$ . As we mentioned in the previous subsection, all the  $p_T$  efficiencies we have been fitted using functional form (3). Fit parameters for this function for the electron efficiencies in data and MC for all the definitions are summarized in Tables VI for CC region and VII for EC region. The ratio of the efficiencies in data to that in MC gives correction factor for a chosen definition (see eq.(4)).

Table VI: Fit parameters for electron  $p_T$  efficiencies in data and MC for the six definitions in CC.

	VLoose	Loose	MLoose1	MLoose2	Medium	Tight
$p_0^{data}$	2.55901	0.433514	2.57156	1.78059	2.3748	2.6072
$p_1^{data}$	-1.24807	-0.470056	-0.83386	-0.58809	0.65164	-0.562246
$p_0^{MC}$	0.648609	0.320142	1.29136	0.978957	1.62978	1.30179
$p_1^{MC}$	-0.996734	-0.525554	-0.774526	-0.621237	-0.715462	-0.554375

The obtained correction factors are shown in Figs. 22–27. The uncertainties for the correction factors are calculated from those for data and MC efficiencies added in quadrature. Similar to the photon ID efficiencies we also add in quadrature uncertainties coming from the background subtraction procedure which we took 1% for MC and 1.5% for data. The values of the final uncertainties for SFs are shown in Tables VIII and IX. From a comparison of  $p_T$  bins in Table VIII we see that the uncertainties are especially large at small  $p_T$  (6 – 7% for 10-12 GeV), drop to 3 – 4% at 18-20 GeV and become frozen at

Table VII: Fit parameters for electron  $p_T$  efficiencies in data and MC for the six definitions in EC.

	VLoose	Loose	MLoose1	MLoose2	Medium	Tight
$p_0^{data}$	19.9804	6.82224	17.0333	47.2372	44.289	55.4933
$p_1^{data}$	-1.7997	-1.26424	-1.40745	-1.58234	-1.33177	-1.13561
$p_0^{MC}$	14.0309	3.98723	19.9417	65.2758	68.1618	32.1626
$p_1^{MC}$	-1.84555	-1.18972	-1.54188	-1.79109	-1.5686	-1.21516

$p_T > 28$  GeV at about 2.5 – 3%. About the same behavior for uncertainties in the EC region (Table IX), where uncertainties vary from 8 – 11% at  $p_T \simeq 10 - 15$  GeV to  $\sim 2.5\%$  at  $p_T \geq 30$  GeV for “VLoose” – “Medium” definitions.

We have studied other possible sources of systematics uncertainty as well. We have looked at efficiencies in data and MC vs. the distance to a closest jet  $dR(jet, e)$ . They are shown in Figs. 32 and 32 for CC and EC definitions. One can see that efficiencies drop by about 5% from  $dR(jet, e) = 3.5 - 4$  to  $dR = 0.7$  for “MLoose1” – “Tight” definitions in both MC and data and almost does not change for “VLoose” and “Loose” definitions.

Another effect we studied is a dependence on the total number of jets ( $njets$ ) in the event. Here we see a clear linear dependence: the larger  $njets$  the smaller efficiency. Since this effect is very well reproduced in MC, the only uncertainty may come from different average number of jets in MC and data. From analysis of  $Z \rightarrow ee$  events we found, for example, that those numbers are pretty close to each with a largest difference of about 0.20 jets [17]. According to the parametrization of the  $njets$  dependencies, it may give an additional just  $\lesssim 0.5\%$  uncertainty <sup>4</sup> on the SFs presented in Tables VIII, IX.

Another systematics may be caused by a description of the material distribution in the MC simulations. For this aim, we have considered samples of single electrons generated by JES group [16] with various amount of radiation lengths added in the solenoid region ( $+0.17 X_0$ ,  $+0.36 X_0$ ) in CC region and using the W mass group GEANT tunings <sup>5</sup>. Using those samples we found that typical variations of the electron ID efficiencies w.r.t. the default MC samples is about 0.3 – 1.2% and the efficiencies are consistent withing statistical uncertainties.

<sup>4</sup> For example, for Loose definition in EC we can parametrize the  $njets$  dependence as  $0.924 - 0.0272 \cdot njets$  with  $\langle njets \rangle$  in data and MC equal to 1.06 and 0.86. It gives approximately  $0.0272\% \cdot 0.2 = 0.54\%$  difference in the data/MC efficiencies.

<sup>5</sup> According to estimates of the W mass group about  $0.28 X_0$  more are probably required to better describe the longitudinal behavior of the electron shower in the EM calorimeter sections. See also [16].

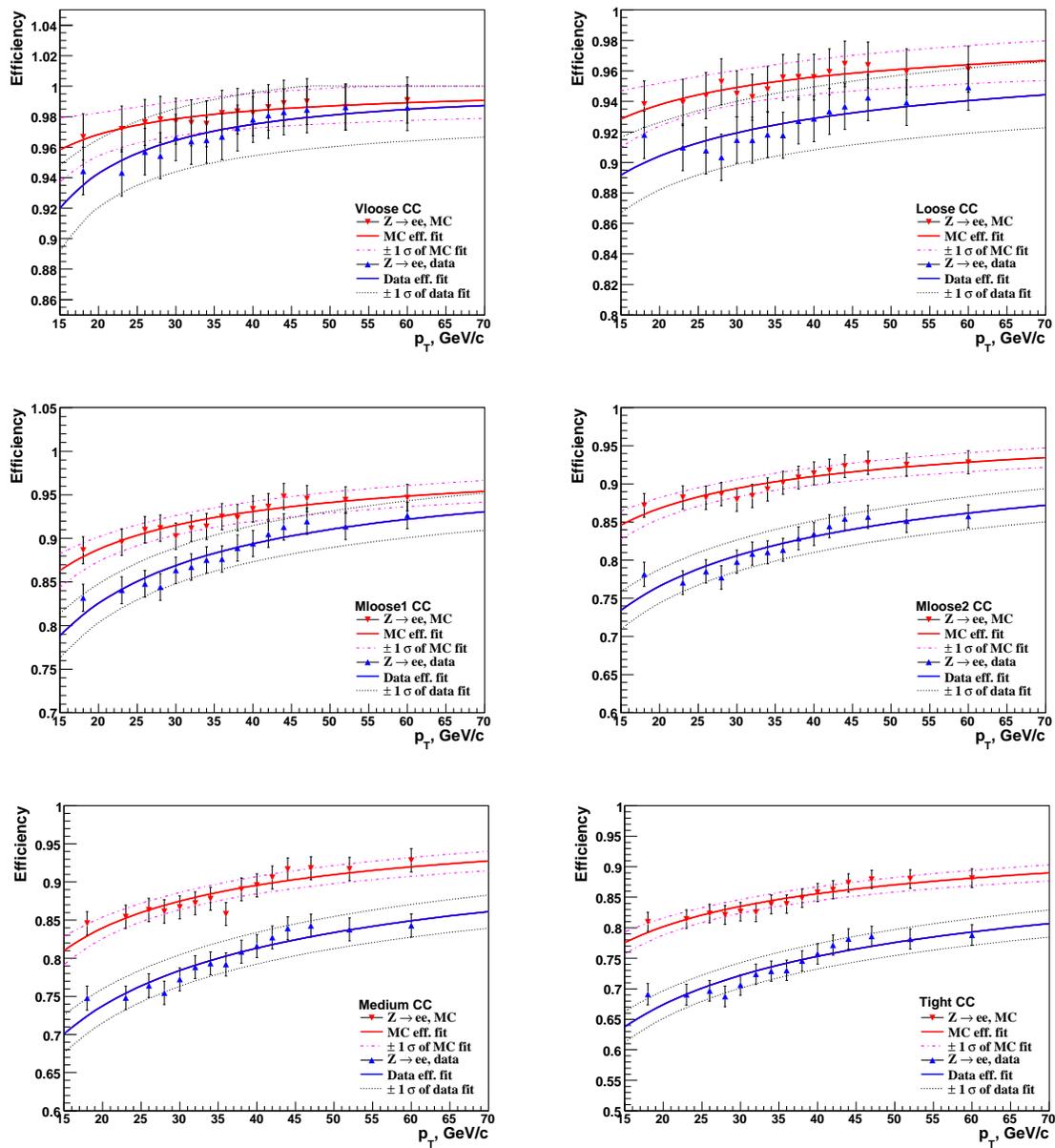


Figure 16: Efficiency in data and MC vs. electron  $p_T$  for all six definition in CC.

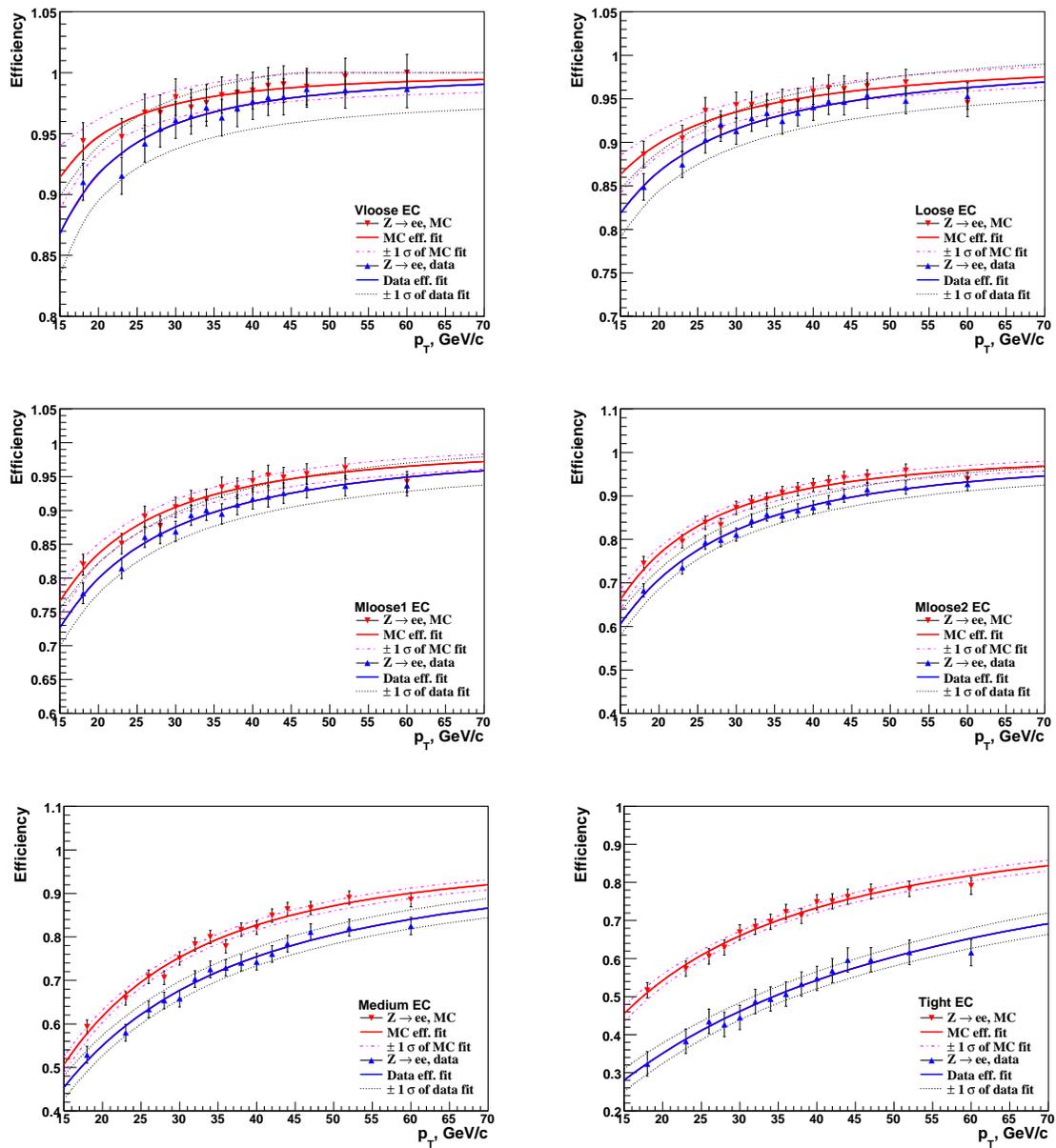


Figure 17: Efficiency in data and MC vs. electron  $p_T$  for all six definition in EC.

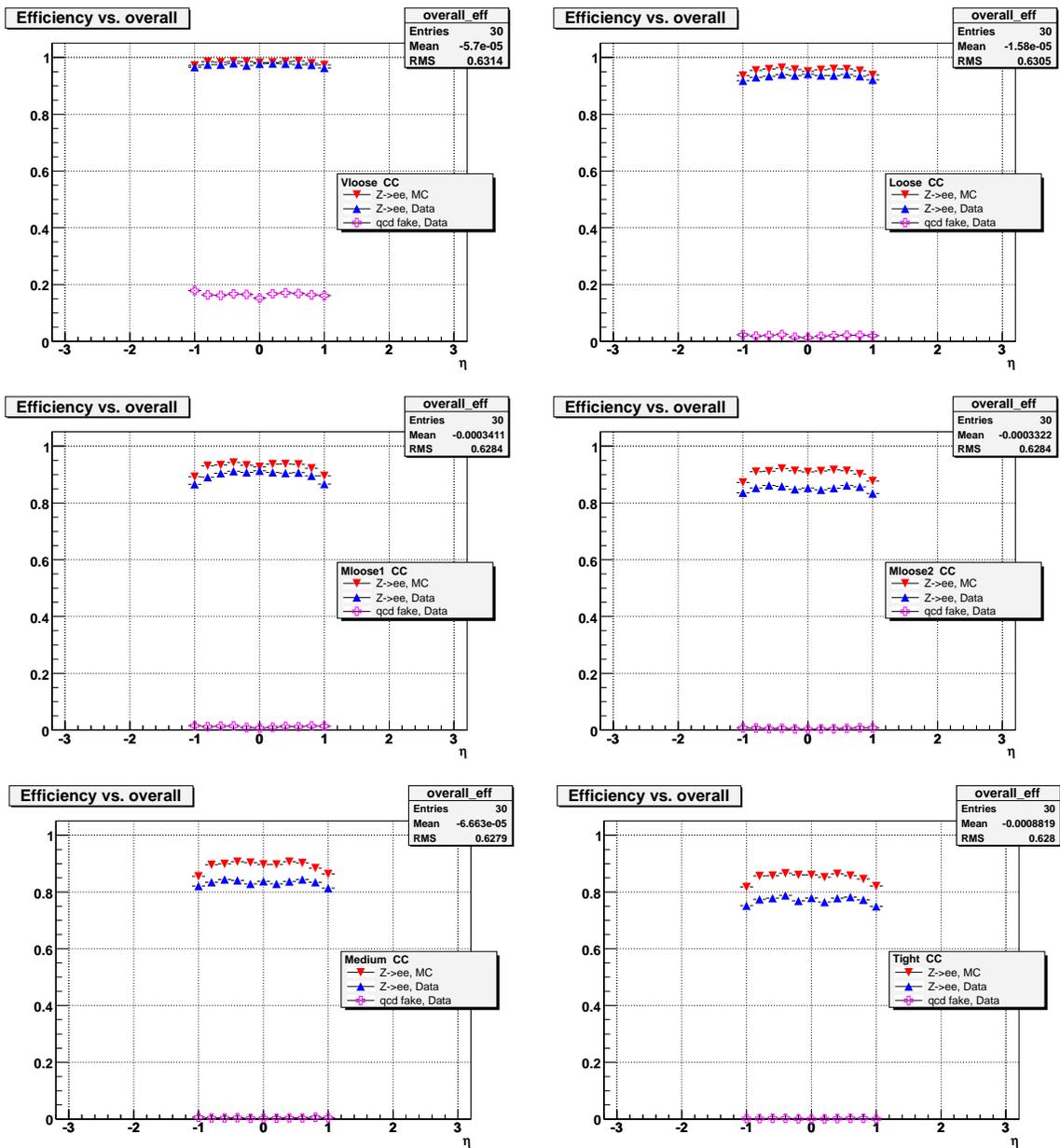


Figure 18: Efficiency in data and MC vs. electron  $\eta_{det}$  for all six definition in CC.

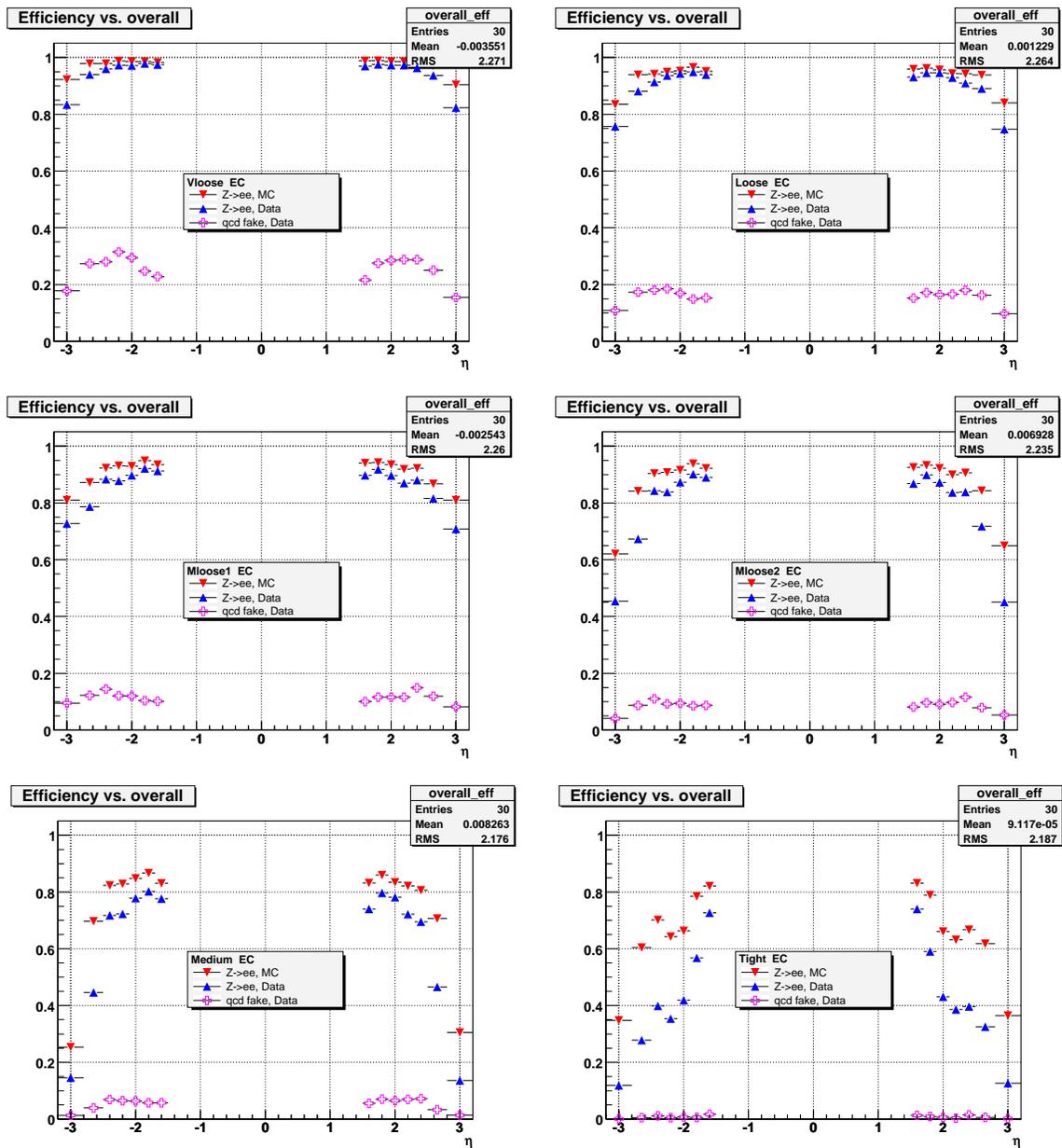


Figure 19: Efficiency in data and MC vs. electron  $\eta_{det}$  for all six definition in EC.

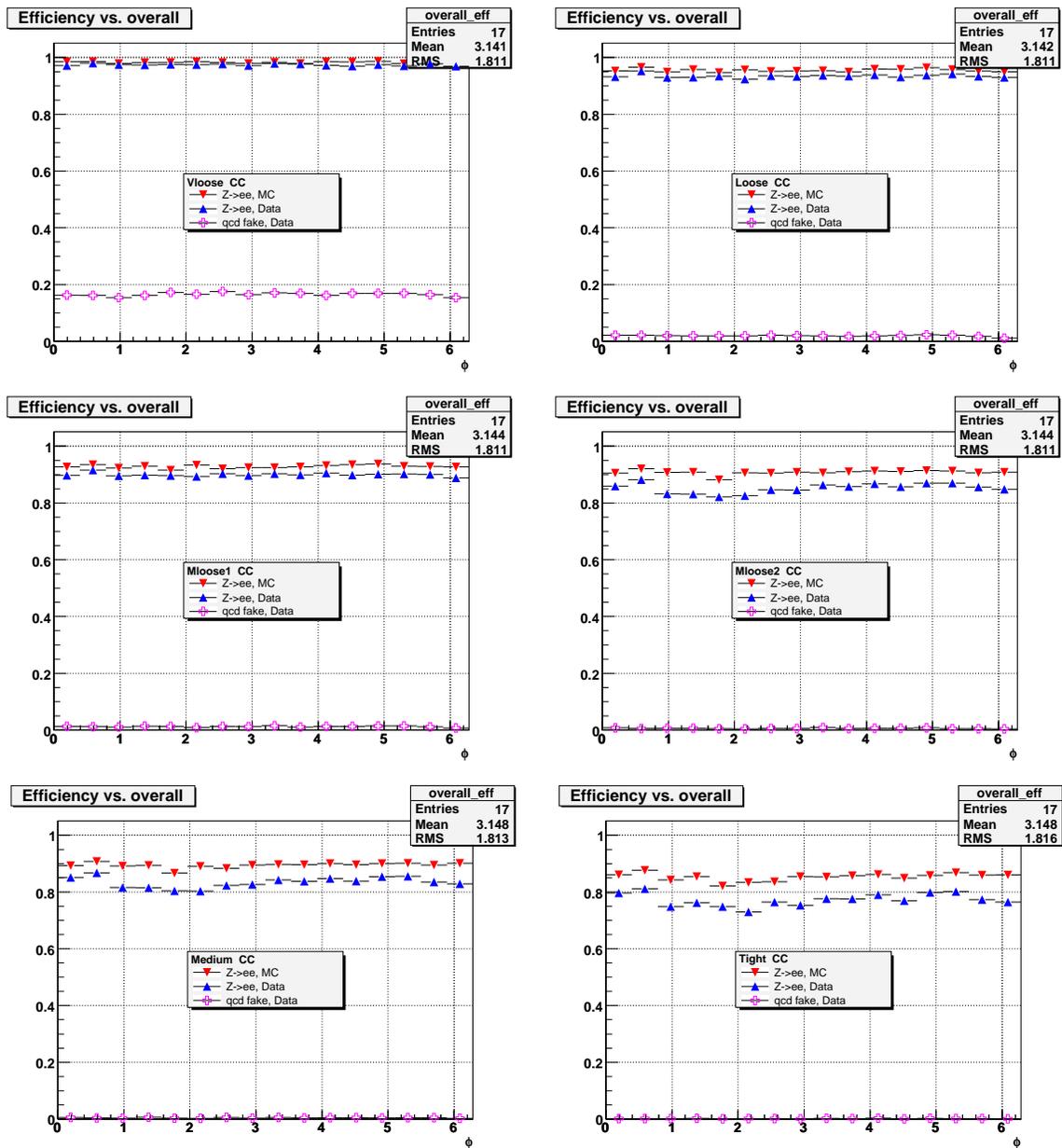


Figure 20: Efficiency in data and MC vs. electron  $\phi_{det}$  for all six definition in CC.

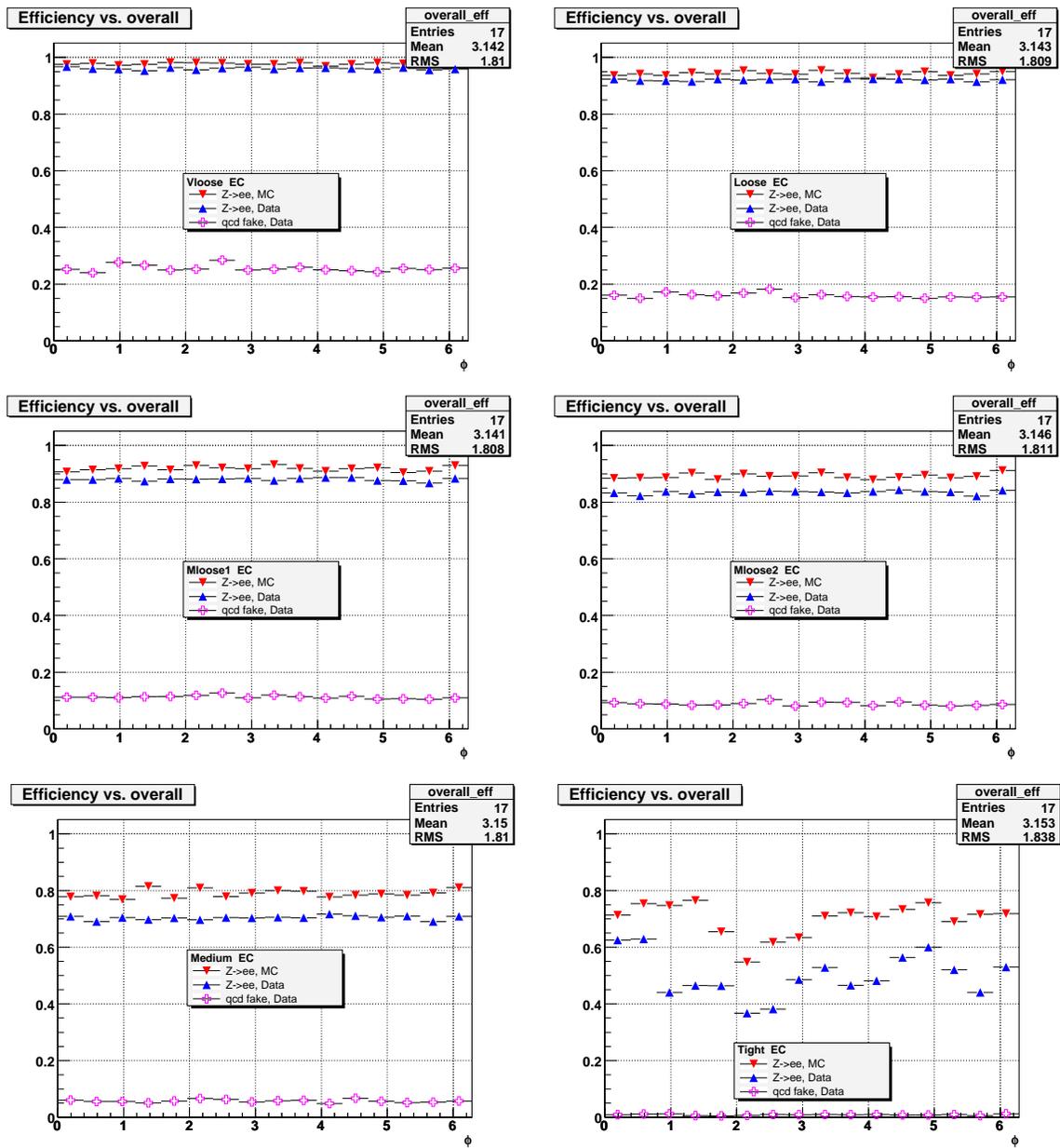


Figure 21: Efficiency in data and MC vs. electron  $\phi_{det}$  for all six definition in EC.

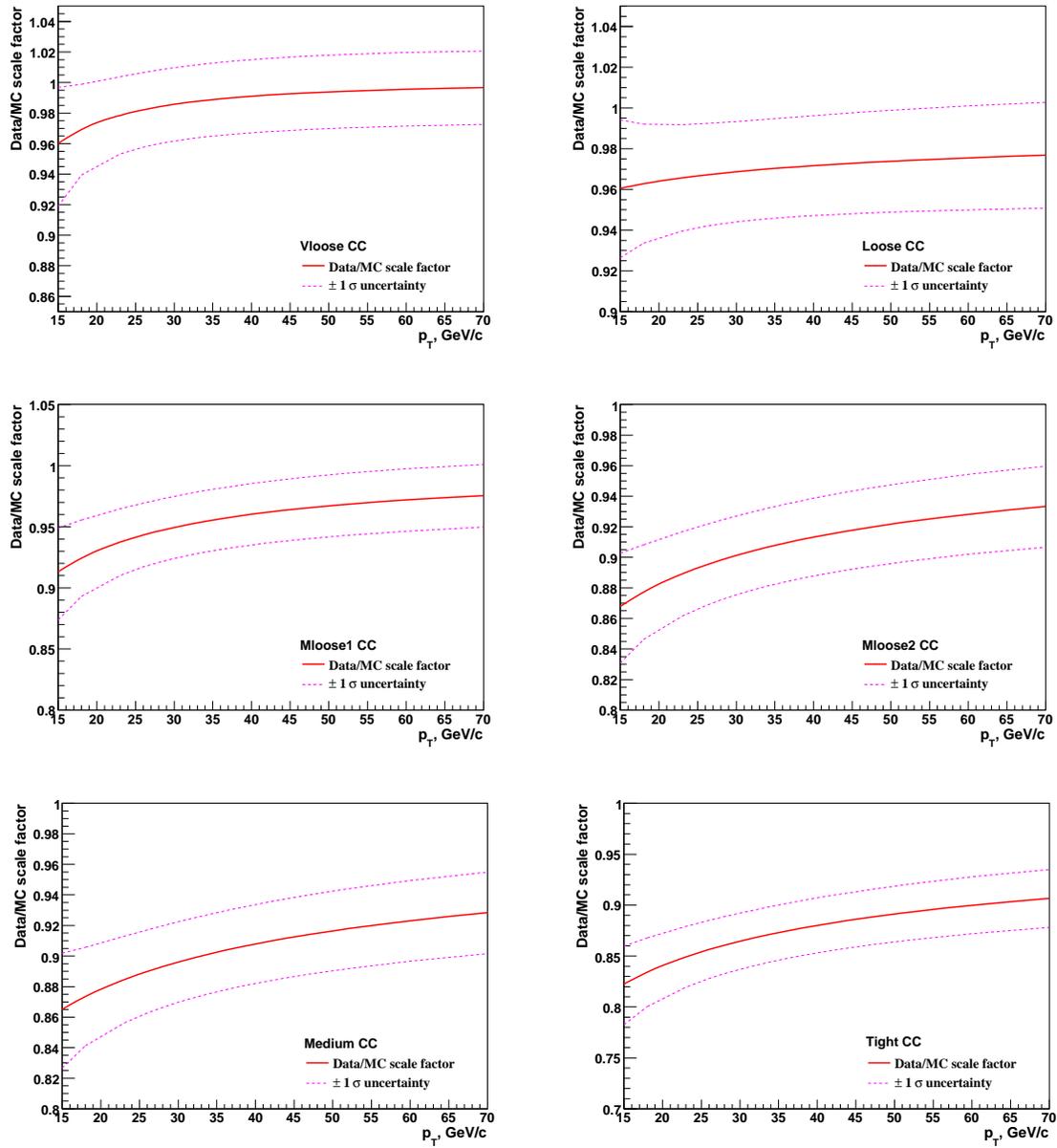


Figure 22: Data/MC efficiency ratio vs. electron  $p_T$  for all six definition in CC region. The shown error band is total uncertainty for the ratio.

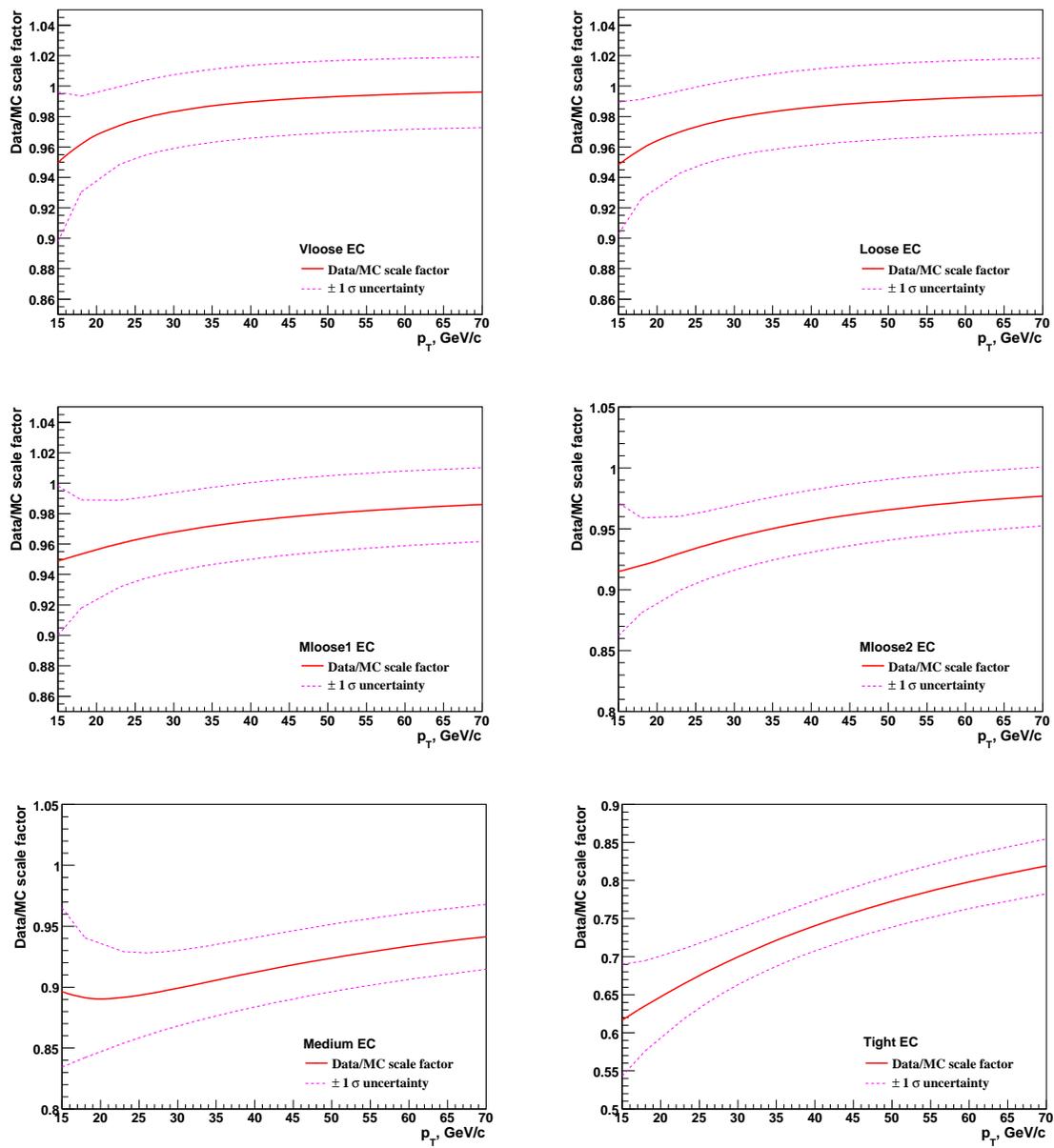


Figure 23: Same as in Fig. 22 but for the EC region.

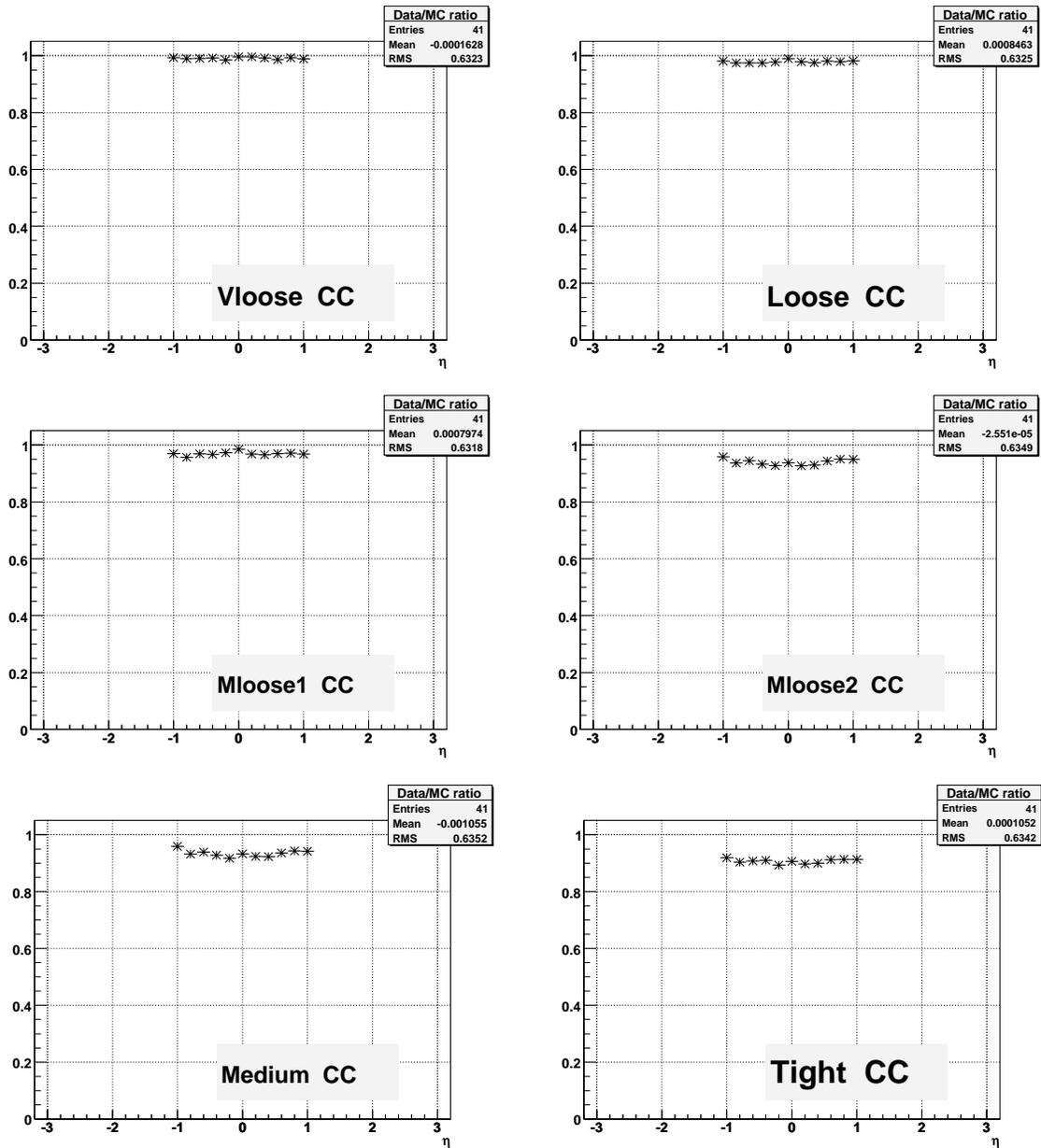


Figure 24: Data/MC efficiency vs. electron  $\eta_{det}$  for all six definition in CC.

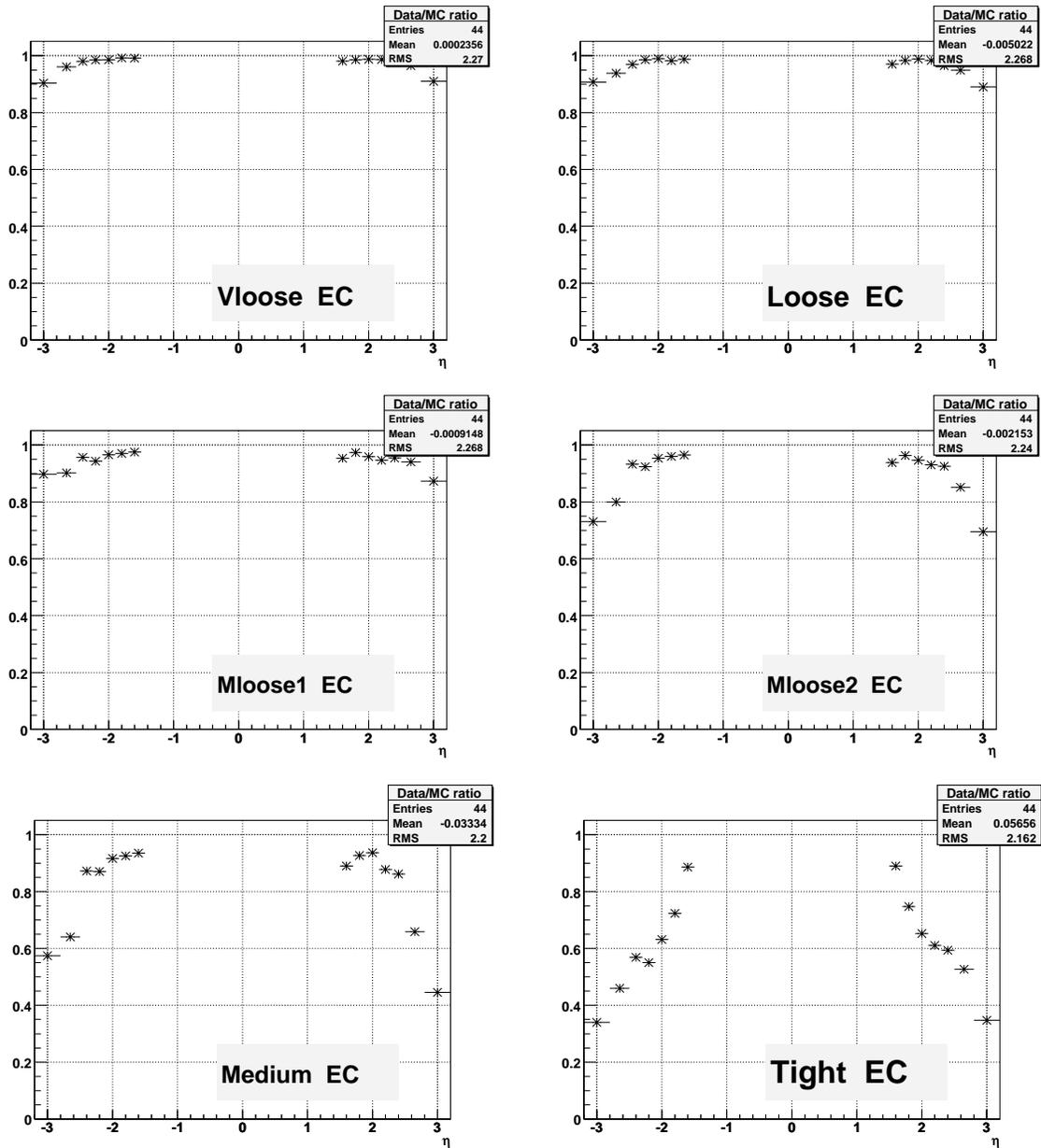


Figure 25: Data/MC efficiency vs. electron  $\eta_{det}$  for all six definition in EC.

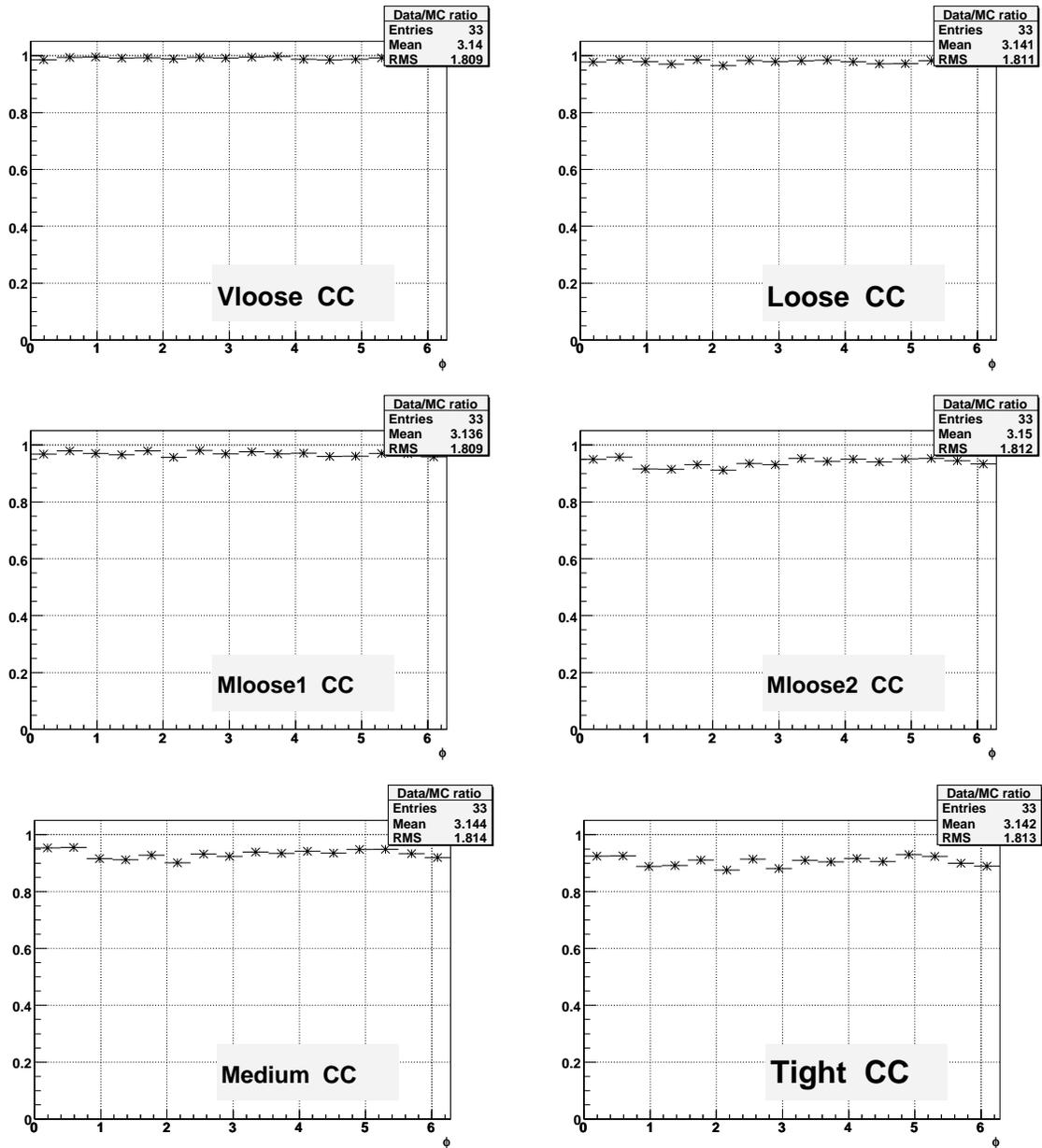


Figure 26: Data/MC efficiency vs. electron  $\phi_{det}$  for all six definition in CC.

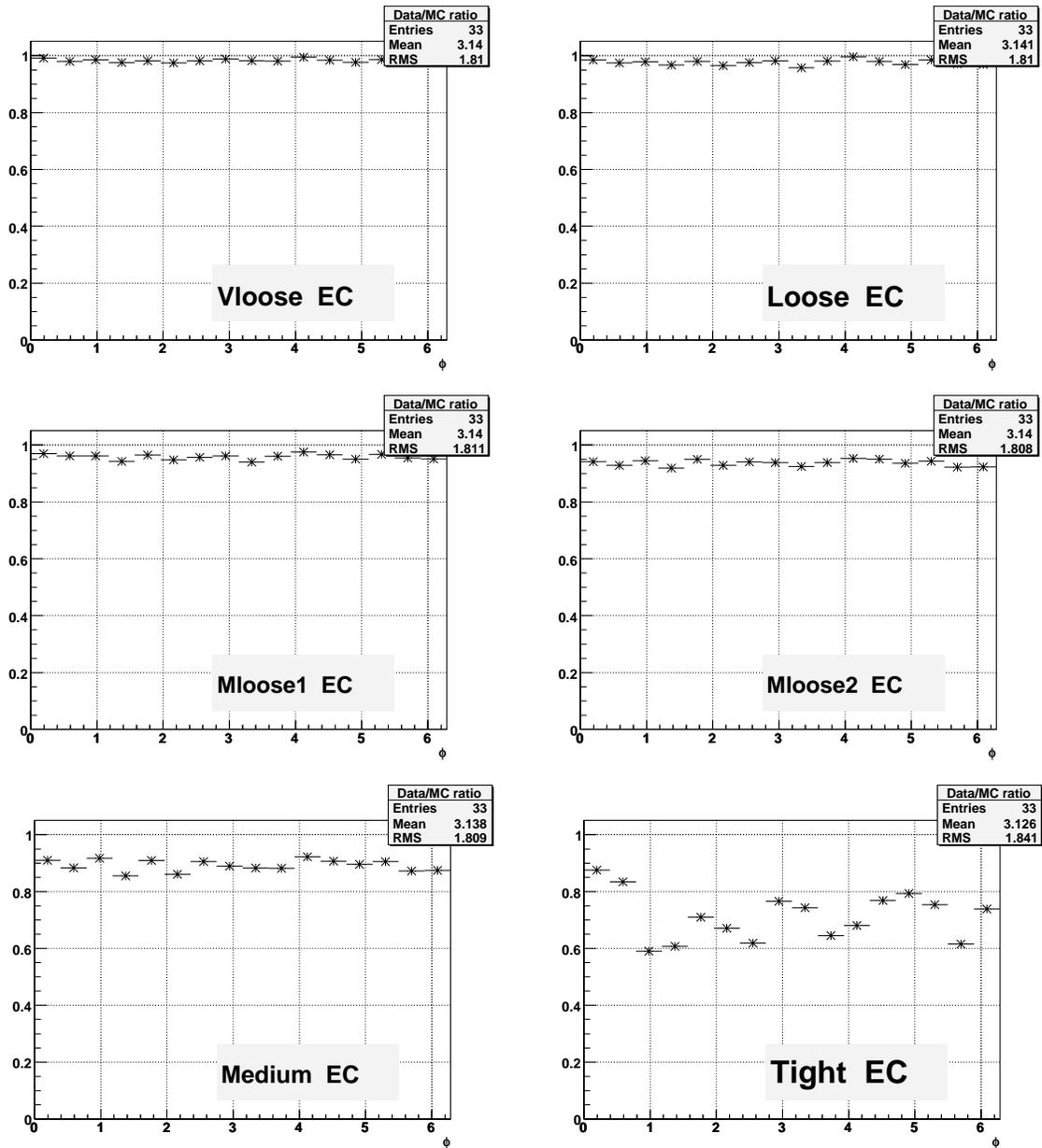


Figure 27: Data/MC efficiency vs. electron  $\phi_{det}$  for all six definition in EC.

Table VIII: Relative uncertainties on the scale factors for the six electron definitions in CC.

$\langle p_T \rangle$ , GeV/ $c$	VLoose	Loose	MLoose1	MLoose2	Medium	Tight
	relative uncertainty d(SF)/SF					
10	0.070	0.048	0.060	0.056	0.060	0.063
13	0.049	0.039	0.047	0.047	0.049	0.053
18	0.031	0.030	0.034	0.035	0.037	0.040
23	0.026	0.027	0.029	0.031	0.032	0.035
26	0.025	0.026	0.028	0.030	0.031	0.033
28	0.025	0.026	0.027	0.029	0.030	0.032
30	0.024	0.025	0.027	0.029	0.029	0.032
40	0.024	0.025	0.026	0.028	0.028	0.031
60	0.024	0.026	0.026	0.028	0.029	0.031
70	0.024	0.027	0.026	0.029	0.029	0.031

Table IX: Relative uncertainties on the scale factors for the six electron definitions in EC.

$\langle p_T \rangle$ , GeV/ $c$	VLoose	Loose	MLoose1	MLoose2	Medium	Tight
	relative uncertainty d(SF)/SF					
10	0.110	0.079	0.092	0.104	0.114	0.159
13	0.068	0.056	0.064	0.074	0.087	0.122
18	0.033	0.034	0.037	0.042	0.055	0.083
23	0.026	0.028	0.030	0.033	0.042	0.069
26	0.025	0.026	0.028	0.030	0.038	0.060
28	0.025	0.026	0.027	0.029	0.036	0.056
30	0.025	0.026	0.027	0.028	0.035	0.052
40	0.024	0.025	0.026	0.027	0.032	0.045
60	0.023	0.025	0.025	0.025	0.030	0.044
70	0.023	0.024	0.025	0.025	0.029	0.044

### 3. CONCLUSION.

We have presented the certification results for photon and electron definitions for Run IIb data and MC, reconstructed within p20 reco version. All the presented electron definitions are new ones, not used in earlier certifications. Along with regular efficiencies for all definitions versus  $p_T, \eta_{det}$  and  $\phi_{det}$ , we also present data/MC correction factors with total uncertainties. We also demonstrate dependencies of the efficiencies on instantaneous luminosity, distance to a closest hadron jet and the number of jets in the event. The provided large variety of definitions should allow to choose a proper one for any particular physics analysis.

### 4. ACKNOWLEDGMENTS.

We thank Common Sample Group, especially Herbert Greenlee and Slava Shary, for providing data and MC samples as well as permanent useful discussions. We are very also grateful to V+jets and Higgs

groups, especially to Taka Yasuda, Björn Penning for permanent cross-checks and feedbacks. We are also very thankful to Aurelio Juste, Stephan Soldner-Rembold and Erich Varnes for support and many useful discussions.

- 
- [1] Oleksiy Atramentov, Andrew Askew, Dmitri Bandurin, Dan Duggan, Alexey Ferapontov, Yuri Gershtein, Yurii Maravin, Greg Pawloski, “Photon Identification in P17 Data”, D0 Note 4976.
  - [2] J. Hays, J. Mitrevski, C. Schwanenberger, T. Toole, “Single Electron Efficiencies in p17 Data and Monte-Carlo Using p18.05.00 d0correct”, D0 Note 5105.
  - [3] James Kraus *et al*, “p20 ICR Electron Identification”, D0 Note 5691.
  - [4] Masato Aoki *et al*, “Electron Likelihood for p20 data”, D0 Note 5675.
  - [5] Dmitry Bandurin, Oleksiy Atramentov, Yanwen Liu, Hang Yin, “p20 electron track match parameters study”, D0 Note 5635.
  - [6] T. Sjostrand, *Comp.Phys.Comm.* **82** (1995)74.
  - [7] DØ note in preparation.
  - [8] Xuebing Bu, Yanwen Liu, Dmitry Bandurin, Oleksiy Atramentov, “Artificial neural network using central preshower detector information for electron and photon selection”, D0 Note 5650.
  - [9] Xuebing Bu, Yanwen Liu, “Artificial neural network for Run IIb electron and photon identification”, D0 Note 5545.
  - [10] Dmitry Bandurin, Daniel Duggan, “”, D0 Note xxxx.
  - [11] A. Askew, D. Duggan, “CPS Variables for Photon Identification”, D0 Note 4949.
  - [12] “Jet Energy Scale Determination at DØ Run II”, DØ Note 5382.
  - [13] Oleksiy Atramentov and Yurii Maravin, “Utilizing CFT and SMT hits count for photon and electron reconstruction”, D0 Note 4444.
  - [14] A. Ferapontov, Y. Maravin, “Study of  $Z(\gamma) \rightarrow \ell\ell\gamma$  events in DØRunII p17 Data”, DØ Note 5156. D0 Collaboration, V.M. Abazov *et al.*, *Phys. Lett. B* **653**, 378 (2007).
  - [15] D. Bandurin, G. Golovanov, D. Korablev and N. Skachkov, D0 Note 5368, “Measurement of Triple Differential Photon Plus Jet Cross Sections in  $p\bar{p}$  Collisions At 1.96 TeV In D0”. Submitted to *Phys.Lett. B*.
  - [16] “Jedett Energy Scale Determination at DØ Run II (final p17 version)”, DØ Note 5382.
  - [17] See EMID meeting on May 15, 2008; slide 6 of A. Ferapontov and D. Bandurin’s report.

## 5. APPENDIX.

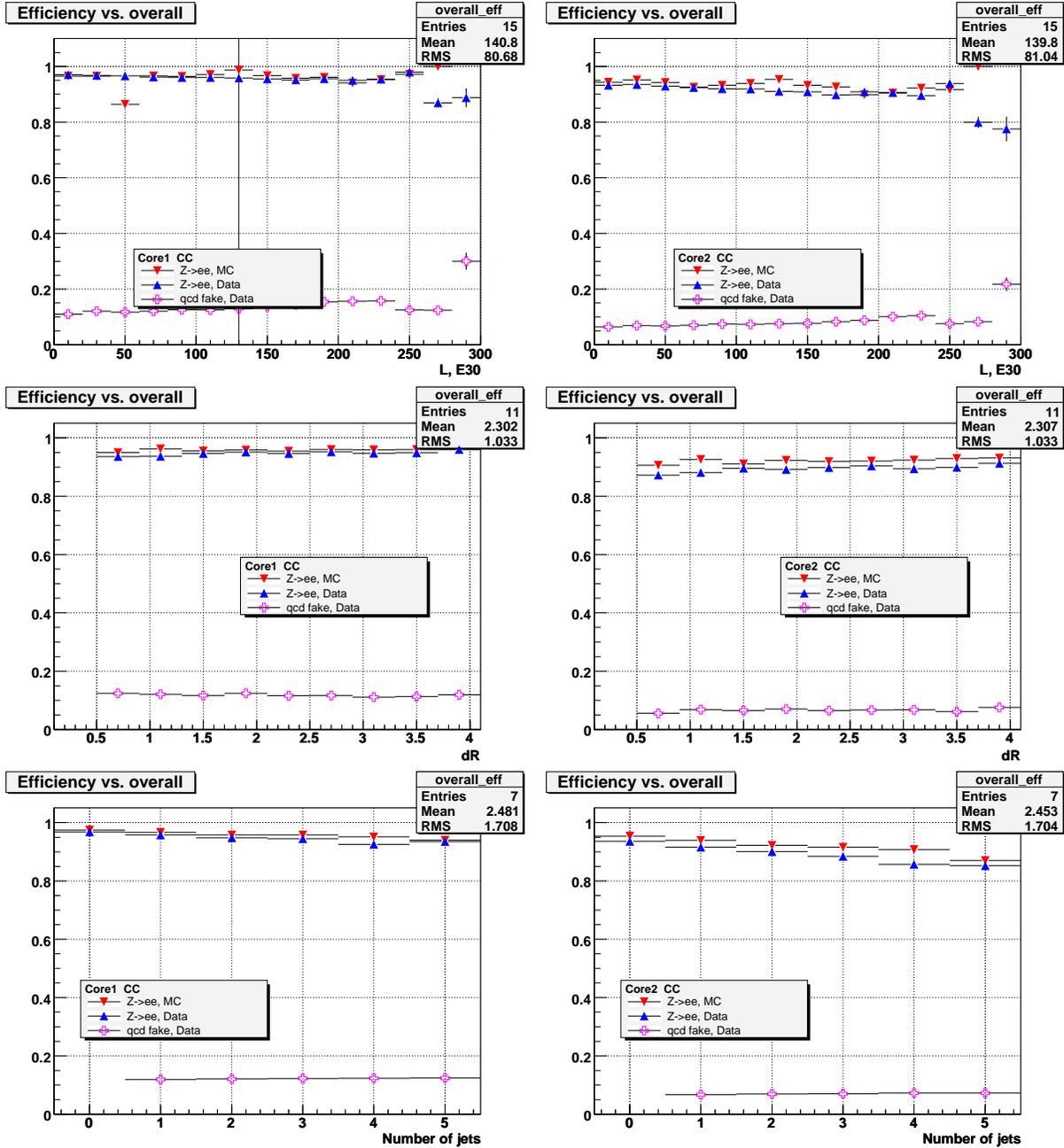
5.1. Dependence of photon selection efficiencies on  $L_{inst}$ ,  $R(e, jet)$  and  $n_{jets}$ .

Figure 28: Efficiency for photon definitions Core1 (left column) and Core2 (right column) definitions for electrons in data and MC, QCD fakes and MC photons as functions of instantaneous luminosity  $L$ ,  $dR(e, jet)$  and the number of jets  $n_{jets}$  in the CC region.

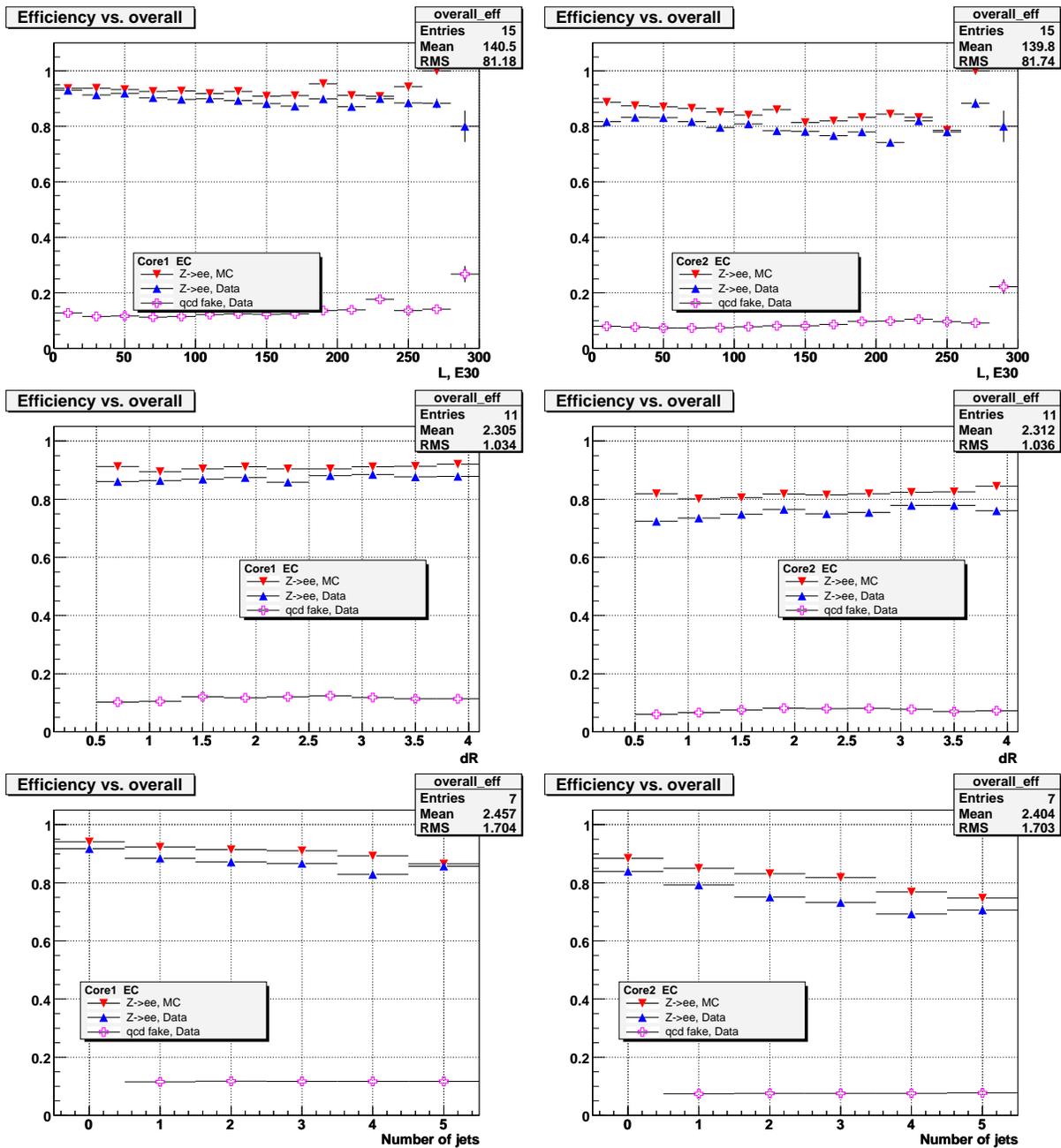


Figure 29: Efficiency for photon definitions Core1 (left column) and Core2 (right column) definitions for electrons in data and MC, QCD fakes and MC photons as functions of instantaneous luminosity  $L$ ,  $dR(e, jet)$  and the number of jets  $n_{jets}$  in the EC region.

## 5.2. Dependence of electron selection efficiencies on $L_{inst}$ , $R(e, jet)$ and $n_{jets}$ .

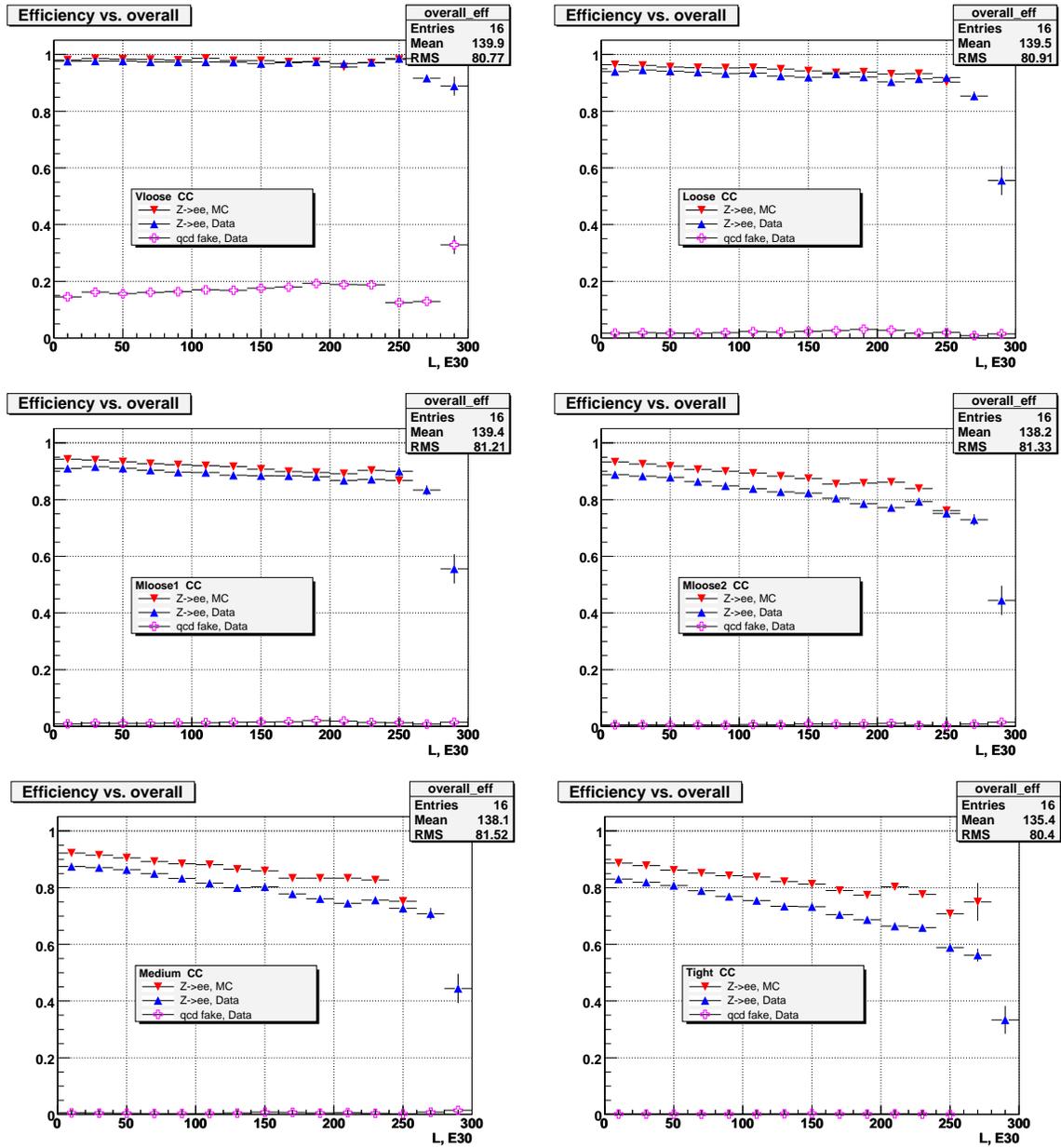


Figure 30: Efficiency in data and MC vs. instantaneous luminosity  $L$  for all six electron definition in CC.

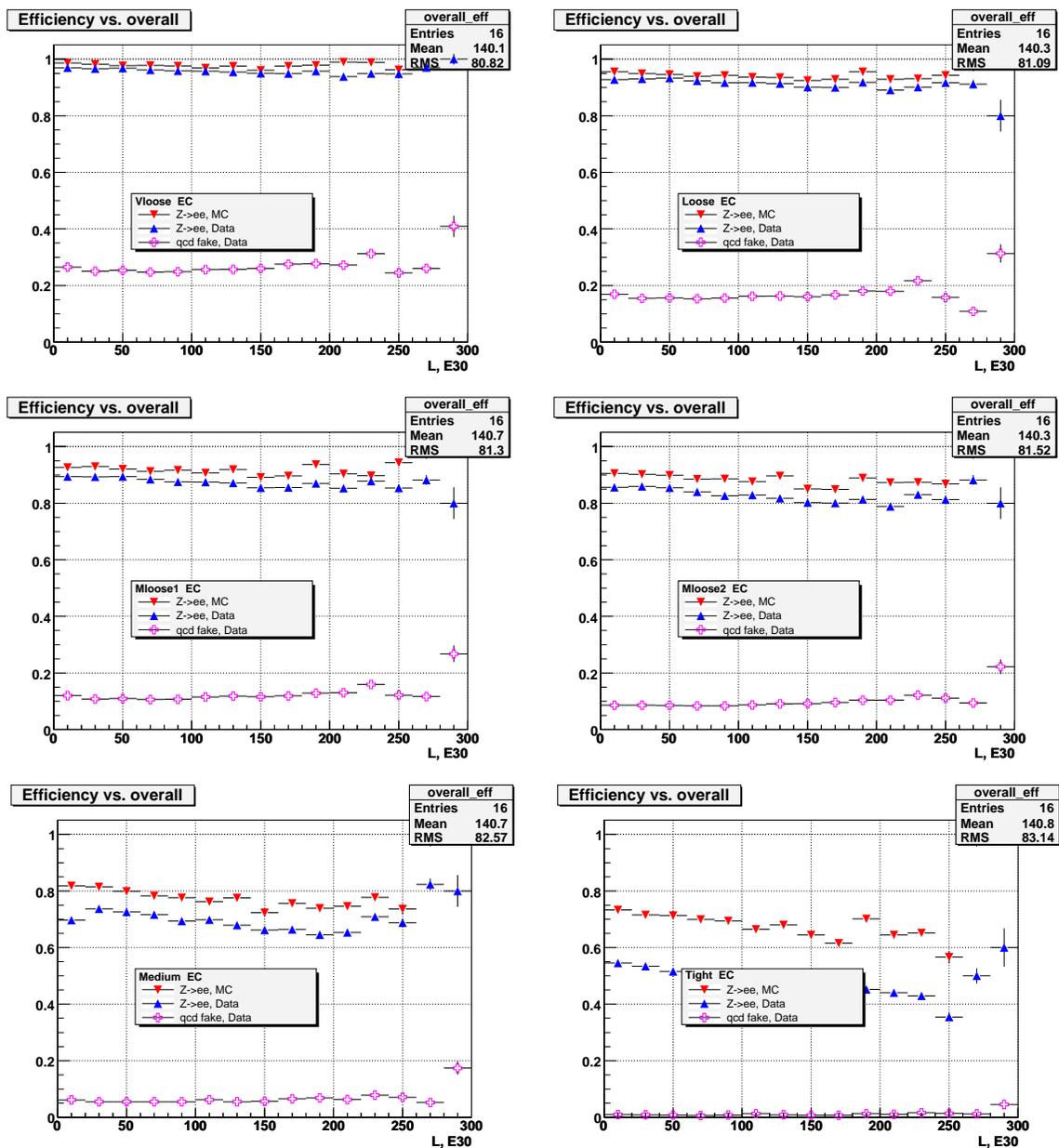


Figure 31: Efficiency in data and MC vs. instantaneous luminosity  $L$  for all six electron definition in EC.

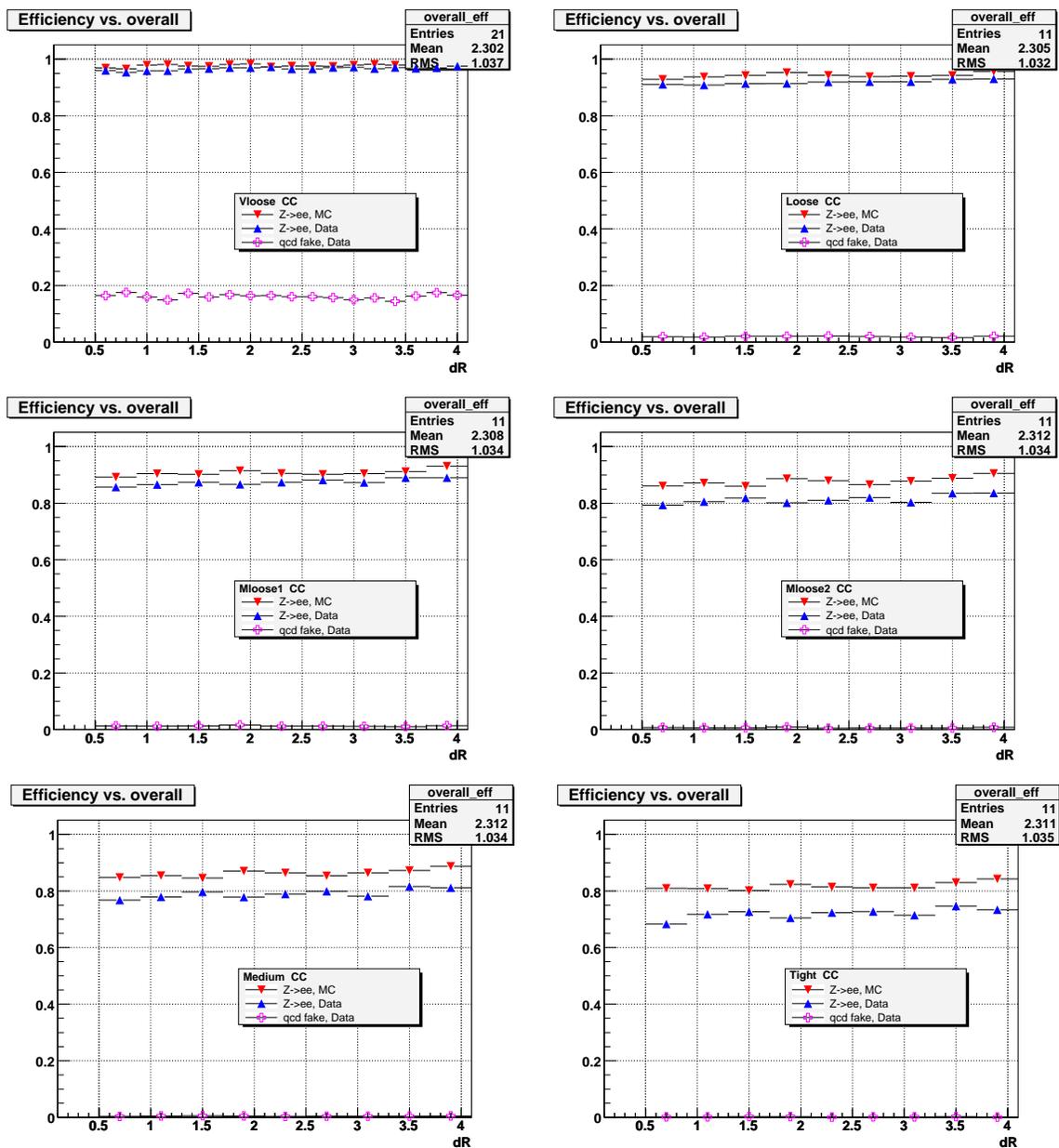


Figure 32: Efficiency in data and MC vs. distance to a closest jet  $R(e, jet)$  for all six electron definition in CC.

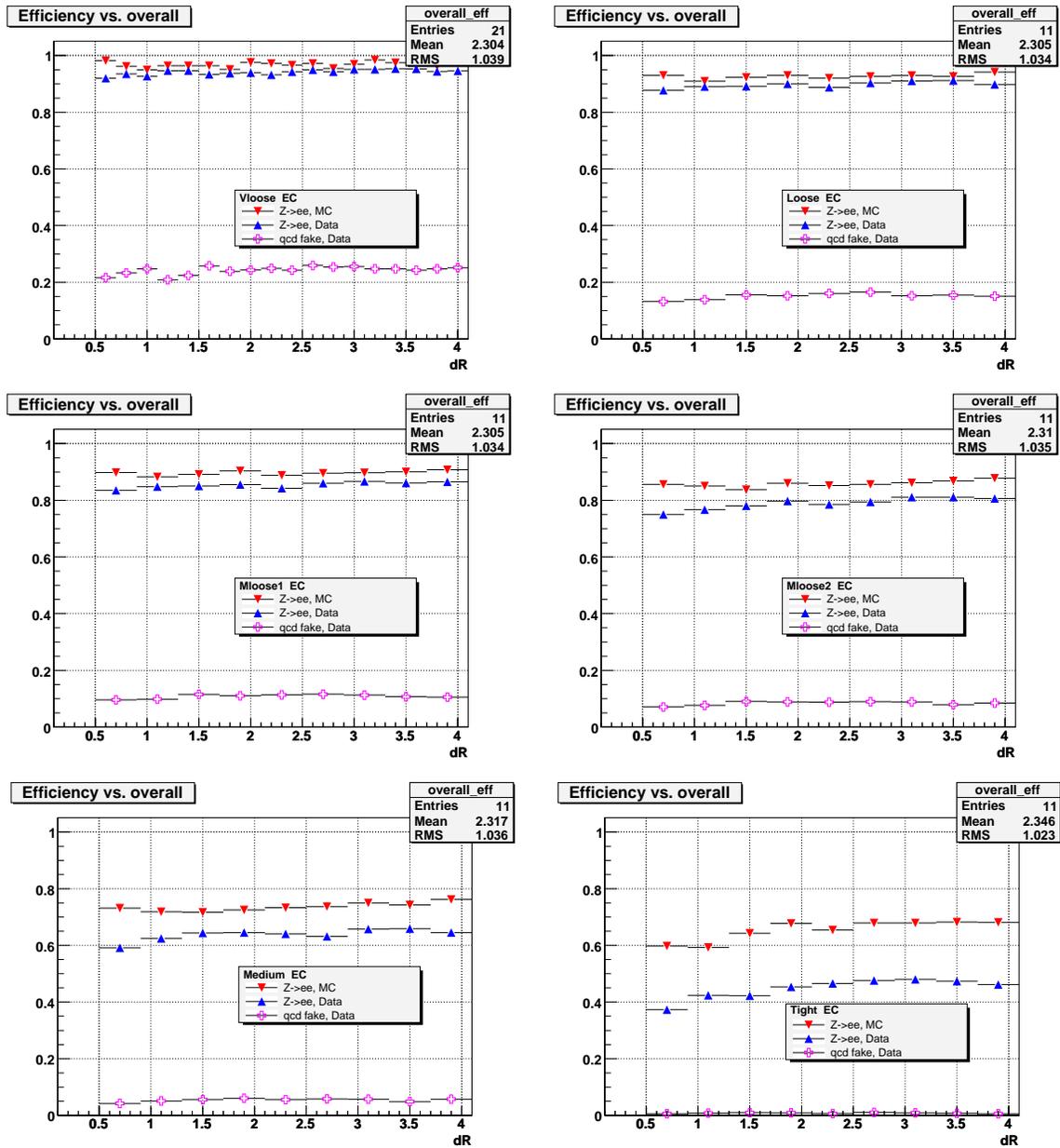


Figure 33: Efficiency in data and MC vs. distance to a closest jet  $R(e, jet)$  for all six electron definition in EC.

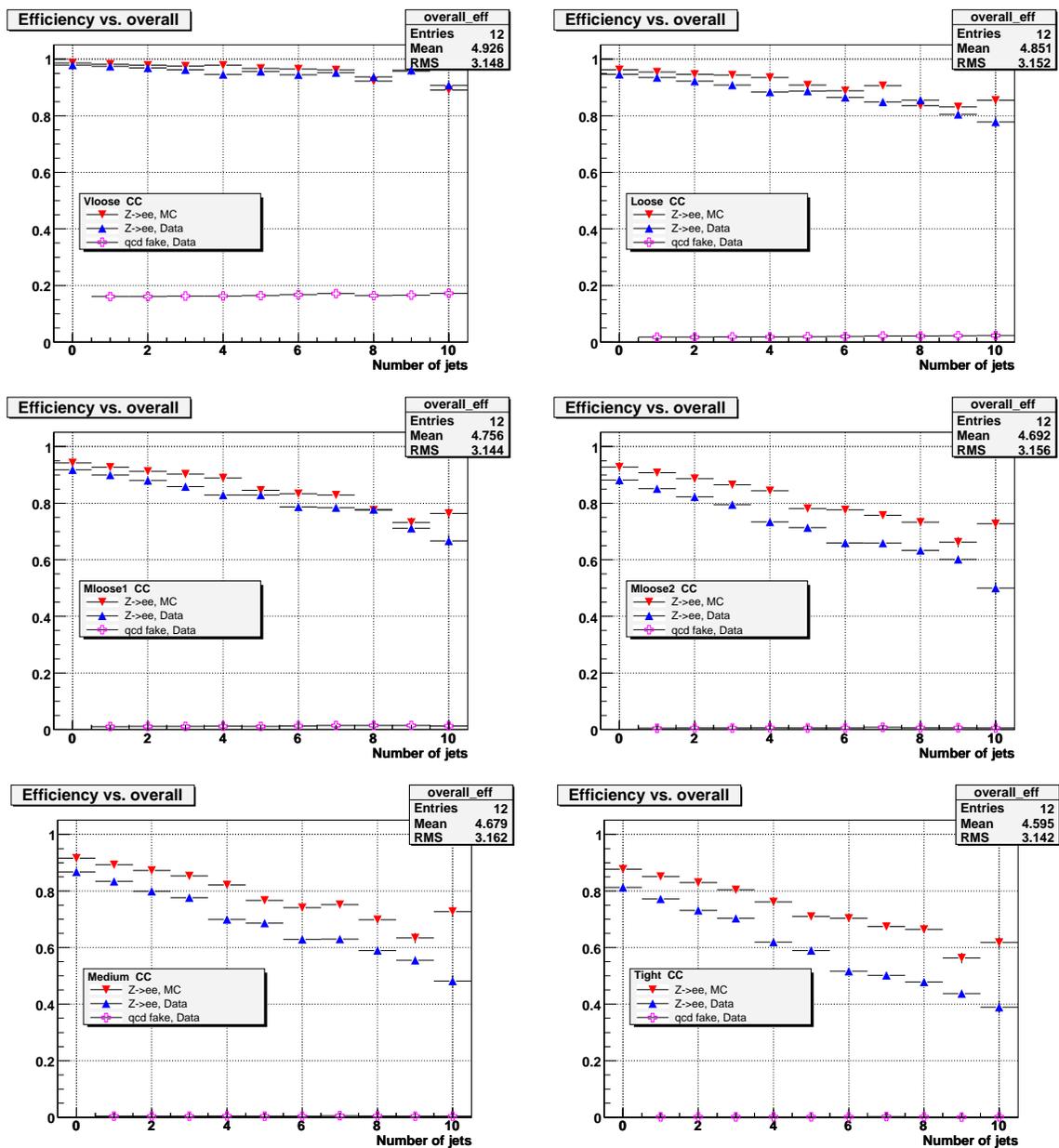


Figure 34: Efficiency in data and MC vs. the number of jets in event  $n_{jets}$  for all six electron definition in CC.

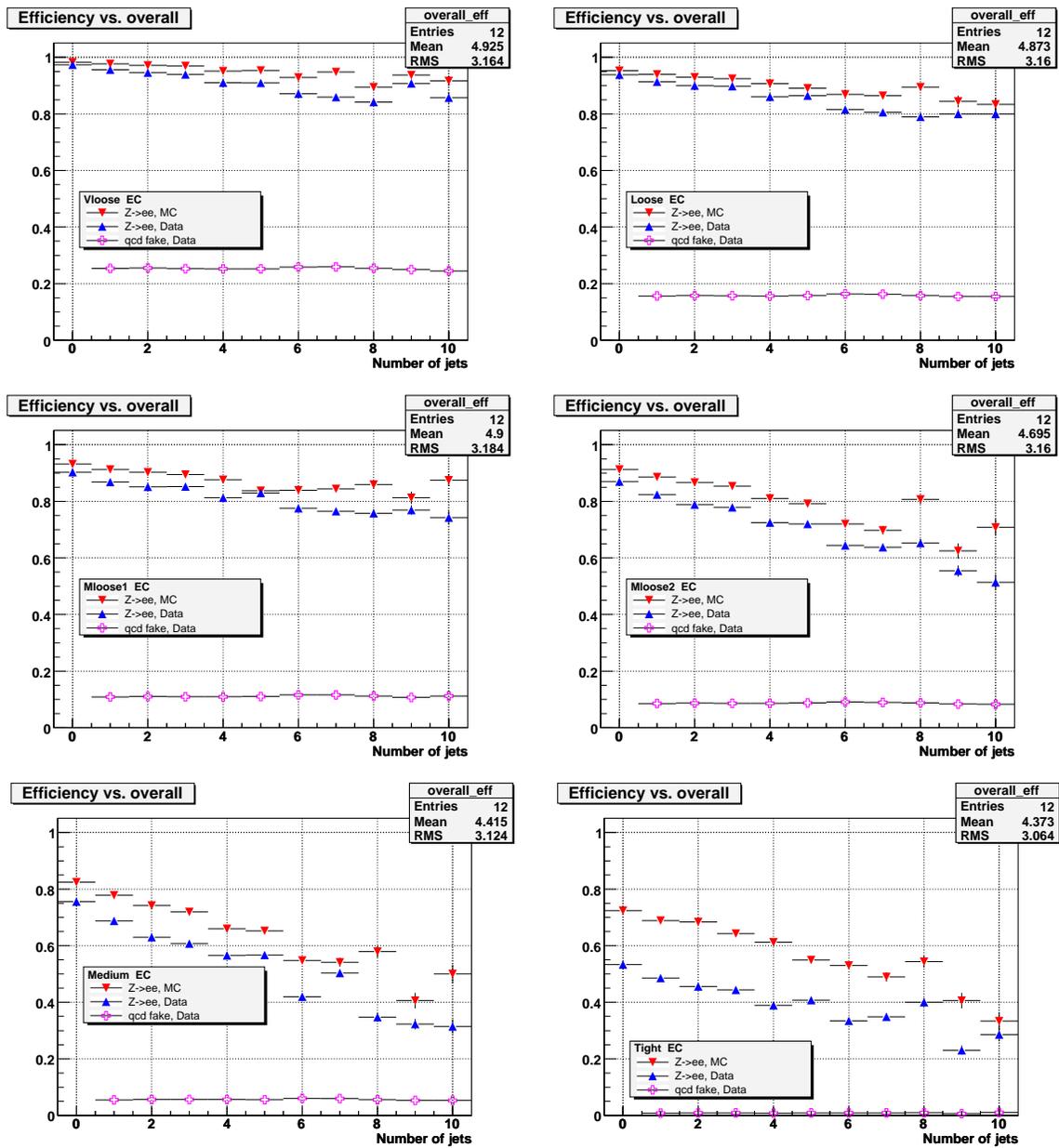


Figure 35: Efficiency in data and MC vs. the number of jets in event  $n_{jets}$  for all six electron definition in EC.