Electron Likelihood Efficiency in p17

Lei Wang
University of Maryland, College Park, USA

Jonathan Hays
Northwestern University, Evanston, Illinois 60208, USA

Christian Schwanenberger
Universitat Bonn, Physikalisches Institut, Bonn, Germany
University of Manchester, Manchester, United Kingdom

Jovan Mitrevski
Columbia University, New York, New York 10027, USA

September 12, 2006

DØ Note 5114

Abstract

We looked at three new electron Likelihood discriminants in p17 for both p14 Pass 2 data and p17.03.03 reprocessed data. The Likelihood efficiency is measured for both $Z/\gamma* \rightarrow e^+e^-$ signal events and EM+jet QCD background events. The three new Likelihood discriminants show similar performances as the old Likelihood discriminant used in p14.

Contents

1 Introduction

Electron Likelihood is used to efficiently select high p_T, isolated electrons by combining several preselected electron identification variables. To construct an Electron Likelihood, first, distributions of each variable are obtained for signal and background samples. Then the distributions of each variable x_i can be used to assign a probability for an EM object to be signal or background: $P_{sig}(x_i)$ and $P_{bkg}(x_i)$. Assuming the variables are uncorrelated, these probabilities can be multiplied to get the overall probability for the EM object

$$P(\mathbf{x}) = \prod_i P(x_i) \quad , \tag{1}$$

where $\mathbf{x}$ is the vector of the variables for the electron. The electron Likelihood is then defined as:

$$L(\mathbf{x}) = \frac{P_{sig}(\mathbf{x})}{P_{sig}(\mathbf{x}) + P_{bkg}(\mathbf{x})} \quad . \tag{2}$$

When $L(\mathbf{x})$ is approaching 1, the EM object is more signal-like and when it is approaching 0, it is more background-like.

In p14[1], seven variables are selected to construct the electron Likelihood LHood. They are:

- f_{em}
- H-matrix(7)
- E_T/p_T
- $Prob(\chi^2_{spatial})$
- Distance of Closest Approach(DCA), which measures the shortest distance of the selected track to the line parallel to the $z-$axis which passes through the primary vertex position.
- Number of tracks in a ΔR=0.05 cone, around and including the candidate track
- Total track p_T in a ΔR=0.4 cone around, but excluding the candidate track

The study described in [2] shows that there is a significant correlation between f_{em} and H-matrix(7). So three new Likelihood discriminants with good discriminating power and small correlation variables are constructed. Five variables were selected:

- N_{trks} in a ΔR=0.05 cone
- $\sum p_T^{trk}$ between ΔR=0.05 and ΔR=0.4
- $\chi^2_{spatial}$
- f_{iso}
- E_T/p_T

Using the first three variables the 3-variable Likelihood LHood3 was constructed. By adding f_{iso} or E_T/p_T to the three variables the 4-variable Likelihoods LHood4Iso and LHood4EOP, respectively, are defined.

In [2] p14 Pass 2 data was used to look at the new Likelihood discriminant efficiencies. That analysis used `top_analyze`, while in this analysis we use `em_cert` which utilizes d0reco and d0correct. First, it was checked for p14 Pass 2 data that the results derived with the two different analysis tools give consistent results. Then p17 reprocessed data was analyzed.

2 Event Selection

For the p14 Pass 2 data, events are reconstructed using p14.03.00, p14.03.01, p14.03.02, p14.05.00, p14.05.02 and p14.06.00 versions of the DØ code and a d0correct v8.3 code. For p17 data, events are reconstructed using p17.03.03 and a d0correct v9 code. We analyzed the electron Likelihood efficiency for both signal and background. For the $Z/\gamma* \rightarrow e^+e^-$ signal we use the Common Sample Group's 2EM skims.

For p14 Pass 2 data, the SAM dataset definitions are:

- CSskim-2EM-PASS2-p14.03.00
- CSskim-2EM-PASS2-p14.03.01
- CSskim-2EM-PASS2-p14.03.02
- CSskim-2EM-PASS2-p14.05.00

- CSskim-2EM-PASS2-p14.05.02
- CSskim-2EM-PASS2-p14.06.00
- CSskim-2EM-PASS2-p14.06.01

For p17 data, the SAM dataset definition is CSskim-2EMhighpt-PASS1-p170303raw.

The event selection criteria are:

There are two EM objects passing:

- $p_T > 15$ GeV
- f_{em} >0.9, $f_{iso} < 0.15$, H-matrix(7)<50
- Loose spatial track match

We also require that:

- At least one of the EM objects passes single EM trigger
- The invariant mass of the two EM objects is between [80,100] GeV

As to background we use a sample with one EM object and one jet which are back-to-back (EM+jet). We use the Common Sample Group's EM1TRK skims.

For p14 Pass 2 data, the SAM dataset definitions are:

- CSskim-EM1TRK-PASS2-p14.03.00
- CSskim-EM1TRK-PASS2-p14.03.01
- CSskim-EM1TRK-PASS2-p14.03.02
- CSskim-EM1TRK-PASS2-p14.05.00
- CSskim-EM1TRK-PASS2-p14.05.02
- CSskim-EM1TRK-PASS2-p14.06.00
- CSskim-EM1TRK-PASS2-p14.06.01

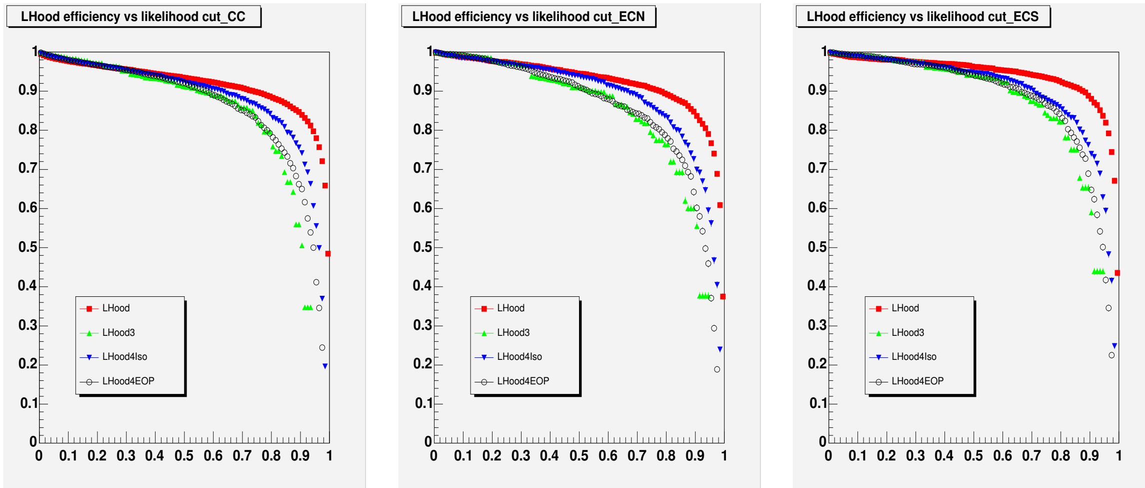
For p17 data, the SAM dataset definition is CSskim-EMinclusive-PASS1-p170303raw and we set the skim flag as "SKIM_EM1TRK" to only select the EM1TRK skim.

The event selection is as following:

- Exactly one EM object satisfies the single EM cuts above
- Event passes EMQCD trigger
- $|\Delta\phi(EM, MET)| > 1.5$
- $p_T^{jet} > 15$ GeV to be a good jet
- 2≤number of good jets≤4
- p_T(leading jet)>25 GeV, η_{det}(leading jet)<2.5
- $\Delta\phi$(leading jet,MET)<1.5
- Missing E_T <15 GeV to remove W's

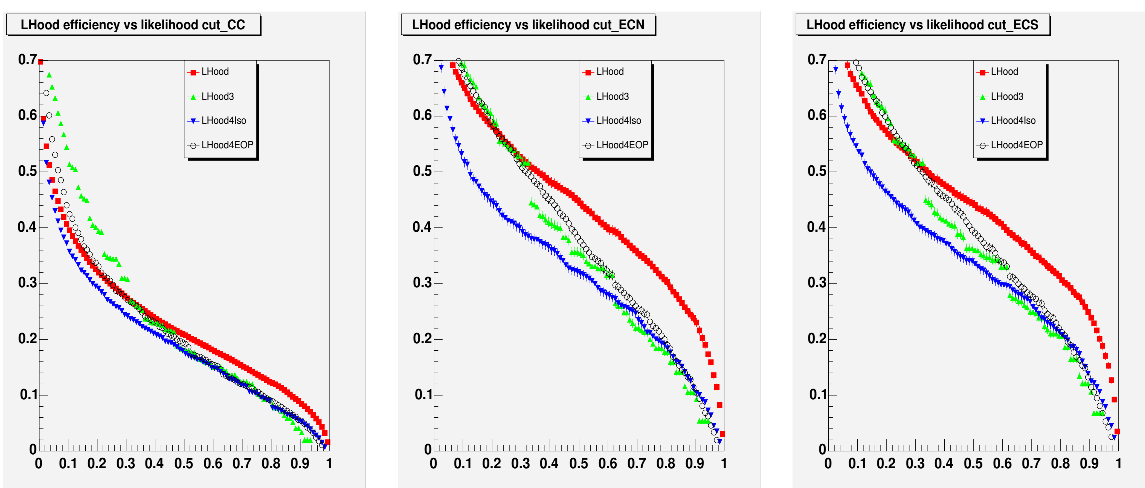
3 Electron Likelihood Efficiency for p14 Pass 2 Data

The Likelihood efficiency is measured using a tag and probe method for both signal and background. Details can be found in [3, 4, 5]. The program package used for this analysis was `em_cert` [6, 7]. All results shown here are consistent with those of [2]. Figure 1 shows the Likelihood efficiencies as functions of Likelihood cuts in different calorimeter regions for $Z/\gamma* \rightarrow e^+e^-$ signal. Here we compared the three new Likelihood discriminants LHood3, LHood4EOP and LHood4Iso with the old Likelihood discriminant LHood. Figure 2 shows the same plots for the QCD background.



(a) CC region (b) ECN region (c) ECS region

Figure 1: Likelihood efficiency as function of Likelihood cut for $Z/\gamma* \rightarrow e^+e^-$ signal in p14 data.

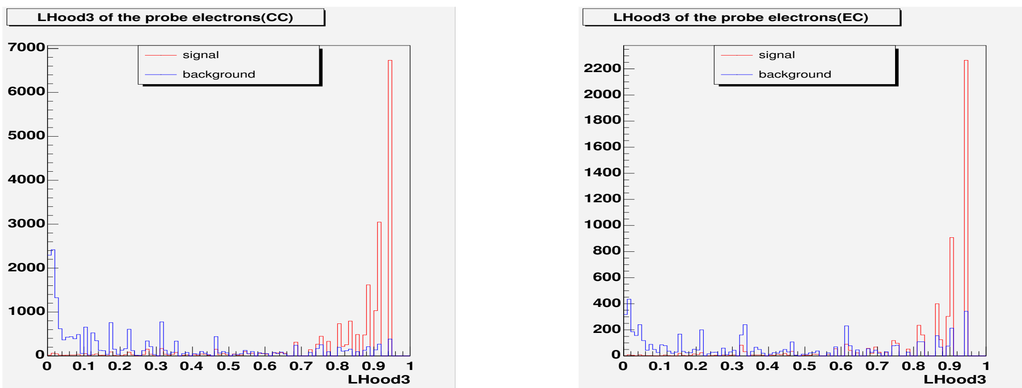


(a) CC region (b) ECN region (c) ECS region

Figure 2: Likelihood efficiency as function of Likelihood cut for EM+jet background in p14 data.

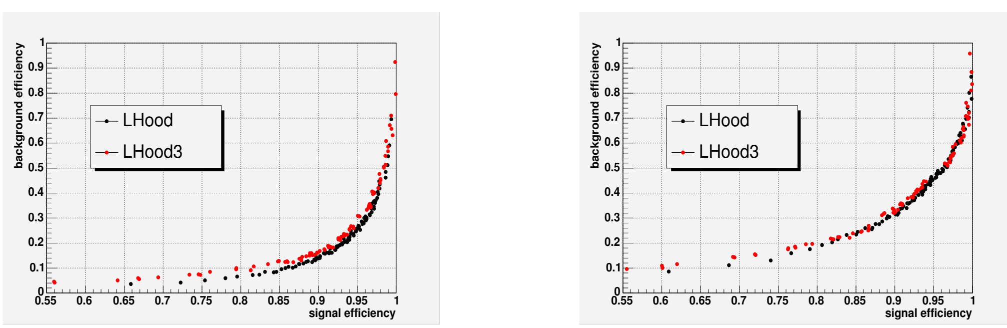
3.1 3-variable Likelihood LHood3

Figure 3 shows the 3-variable Likelihood distributions for both $Z/\gamma* \rightarrow e^+e^-$ signal and EM+jet background. We can see that background and signal have very different Likelihood distributions so that they can easily be separated. The separation power can be derived from Figure 4, where the background efficiency vs. the signal efficiency is plotted for both LHood3 and LHood. We can see that the 3-variable Likelihood performs similar to the 7-variable Likelihood. However, the performance is worse in CC where e.g. for a signal efficiency between 90-95% the fake rate is higher by 3-5%. But in the EC region, the performance of the 3-variable Likelihood almost matches with that of the 7-variable Likelihood.



(a) CC region (b) EC region

Figure 3: 3-variable Likelihood distribution for signal(red) and background(blue) in p14 data.

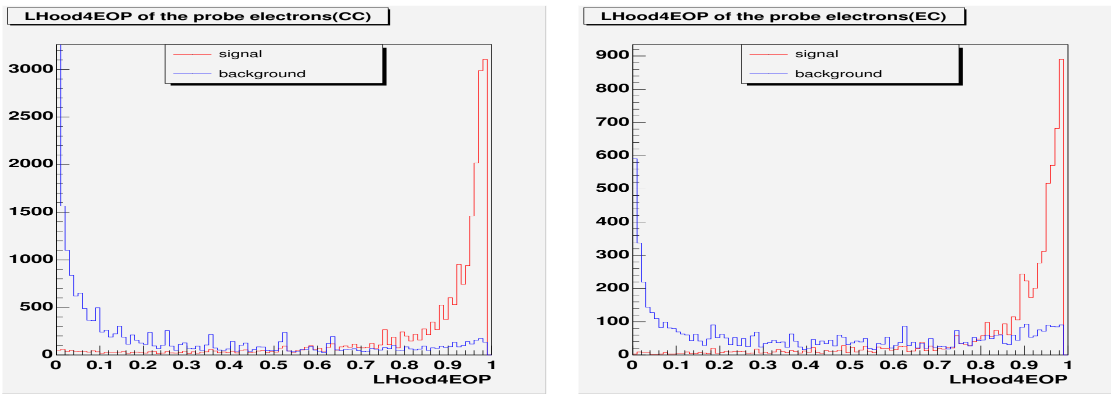


(a) CC region (b) EC region

Figure 4: Background efficiency vs. signal efficiency for LHood3 compared to LHood in CC and EC region in p14 data.

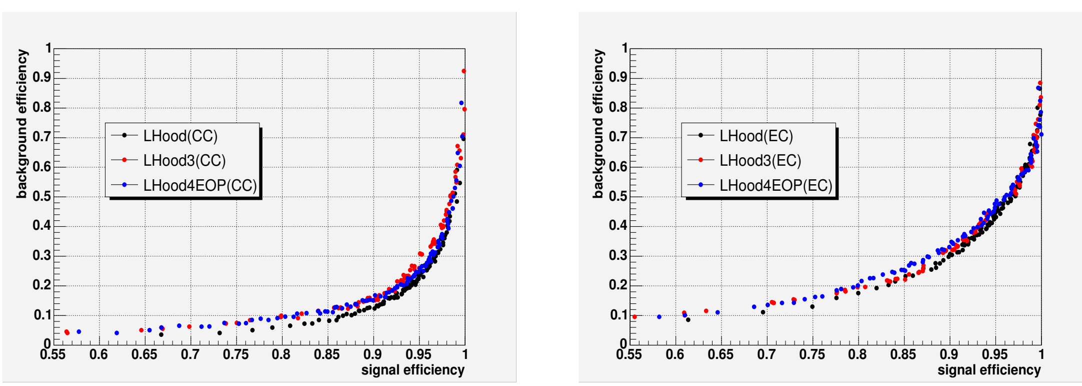
3.2 4-variable Likelihood

With the variable E_T/p_T added, the 4-variable Likelihood LHood4EOP works better than LHood3. Shown in Figure 5, is the Likelihood distributions for both signal and background. In Figure 6, the background efficiencies vs. the signal efficiencies are compared for 3,4 and 7-variable Likelihood discriminants. We can see that in the CC region, LHood4EOP performs better than LHood3. Above a signal efficiency of 90% it almost matches with that of LHood. In the EC region, however, its performance is a little bit worse than that of LHood3 and LHood. E.g. for a signal efficiency between 90-95% the fake rate is higher by 3-5%.



(a) CC region (b) EC region

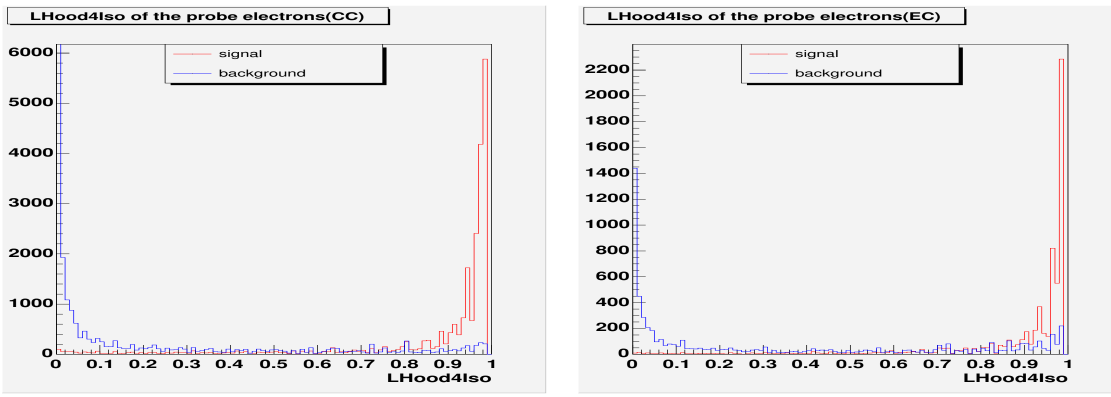
Figure 5: LHood4EOP distribution for signal (red) and background (blue) in p14 data.



(a) CC region (b) EC region

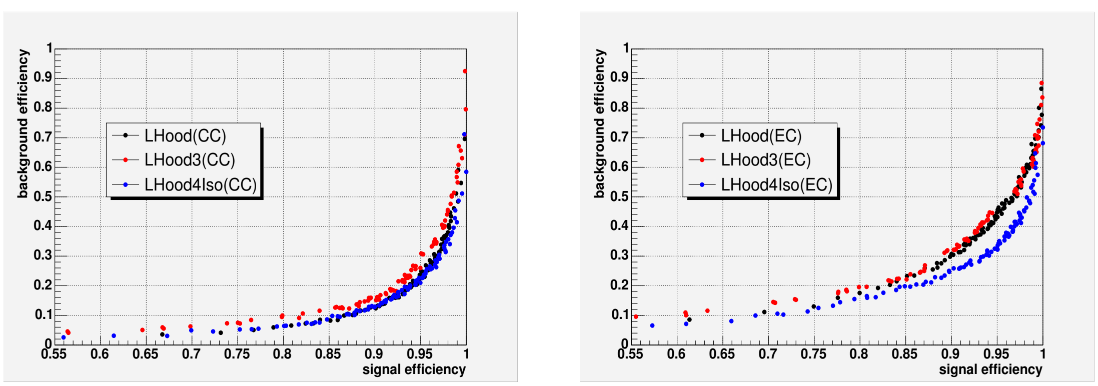
Figure 6: Background efficiency vs. signal efficiency for LHood4EOP compared to LHood3 and LHood in CC and EC region in p14 data.

With the variable f_{iso} added, the 4-variable Likelihood LHood4Iso is defined. In Figure 7 the Likelihood distributions for both signal and background are shown. In Figure 8, the background efficiency vs. the signal efficienciy is plotted for LHood3, LHood4Iso and LHood. In both the CC and the EC region, LHood4Iso performs better than LHood3. In the CC, LHood4Iso matches with LHood and performs even better above a signal efficiency of 95%. In the EC, its performance is much better than that of LHood. E.g. for a signal efficiency between 90-95% the fake rate is lower by 5-10%.



(a) CC region (b) EC region

Figure 7: LHood4Iso distribution for signal (red) and background (blue) in p14 data.

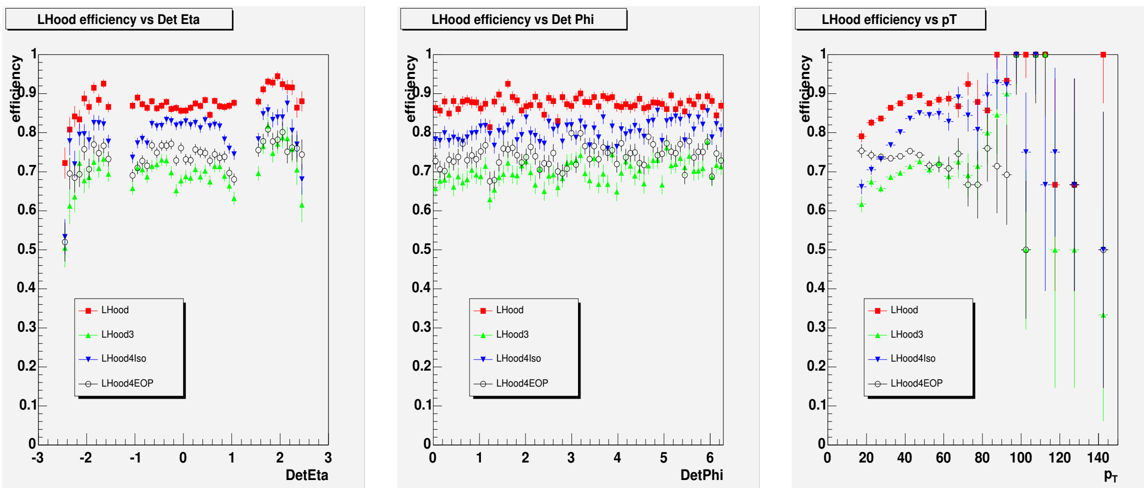


(a) CC region (b) EC region

Figure 8: Background efficiency vs. signal efficiency for LHood4Iso compared to LHood3 and LHood in CC and EC region in p14 data.

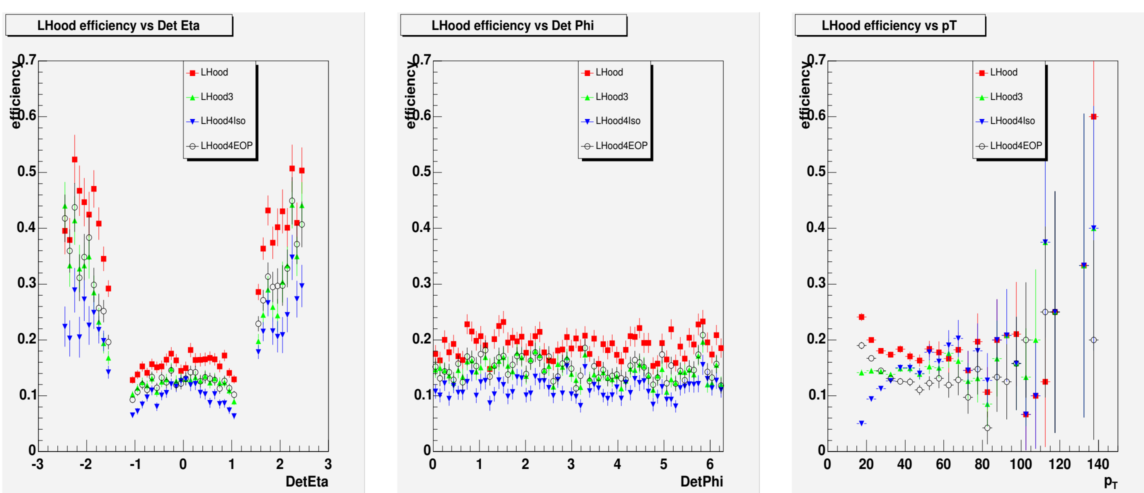
3.3 Kinematic Dependence of the Likelihood Efficiency

We also examined the kinematic(p_T, η, ϕ) dependence of the Likelihood discriminants efficiencies for the signal is shown in Figure 9. Here we set Likelihood cuts to be larger than 0.85. Figure 10 shows the same dependences for the background. The Likelihood cuts are selected so that the corresponding signal efficiency is 85%.



(a) η_{det} dependence (b) ϕ_{det} dependence (c) p_T dependence

Figure 9: Kinematic dependence of the Likelihoods efficiencies for signal in p14 data.

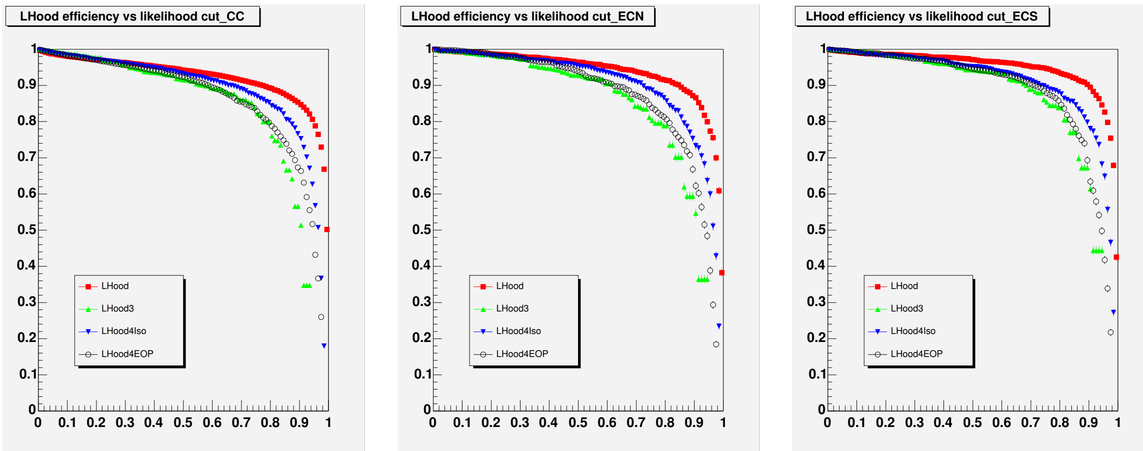


(a) η_{det} dependence (b) ϕ_{det} dependence (c) p_T dependence

Figure 10: Kinematic dependence of the Likelihoods efficiencies for background in p14 data.

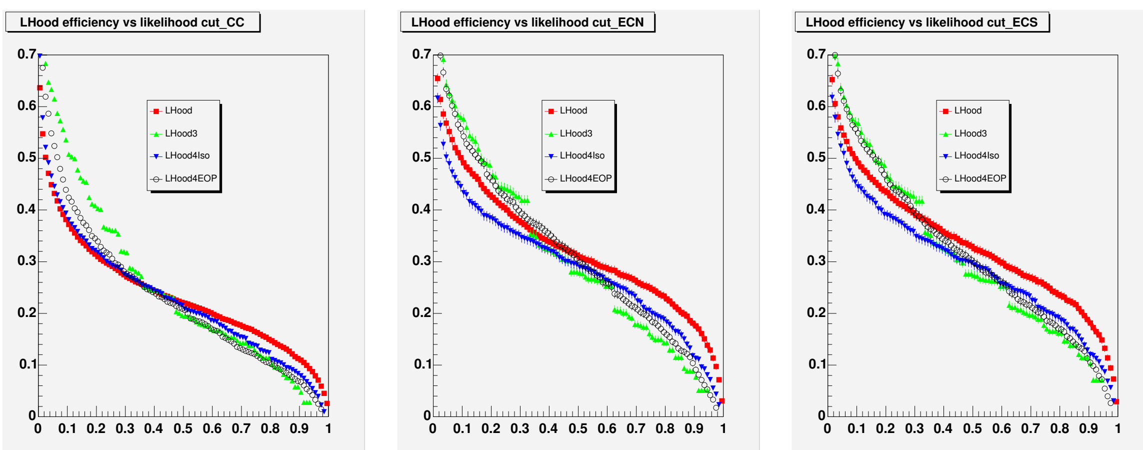
4 Electron Likelihood Efficiency for p17 Data

Figures 11 and 12 show the four different Likelihood efficiencies as function of the Likelihood cut for signal and background, respectively, in different regions of the calorimeter.



(a) CC region (b) ECN region (c) ECS region

Figure 11: Likelihood efficiency as function of Likelihood cut for $Z/\gamma* \rightarrow e^+e^-$ signal in p17 data.

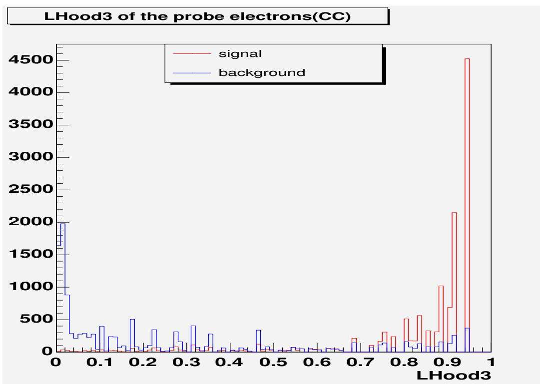


(a) CC region (b) ECN region (c) ECS region

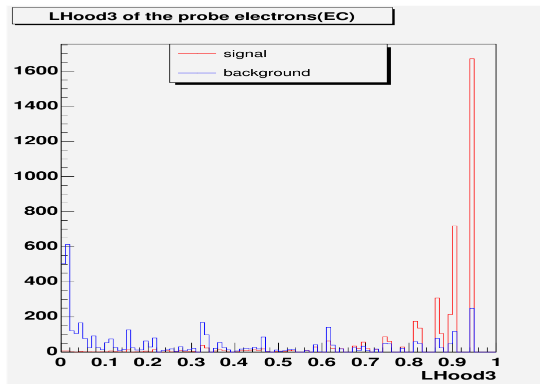
Figure 12: Likelihood efficiency as function of Likelihood cut for EM+jet background in p17 data.

4.1 3-variable Likelihood LHood3

Figure 13 shows the Likelihood distributions for both $Z/\gamma* \rightarrow e^+e^-$ signal and EM+jet background. The separation power can be shown in Figure 14, where the background efficiency vs. the signal efficiency is plotted for both LHood3 and LHood. We can see that the 3-variable Likelihood performs worse in the CC with the fake rate being higher by 5% for a signal efficiency between 90-95%. In the EC region, its performance is worse with the fake rate being higher by 5-10% for a signal efficiency between 90-97%. This is different from the results for p14 data where LHood3 almost matches LHood.

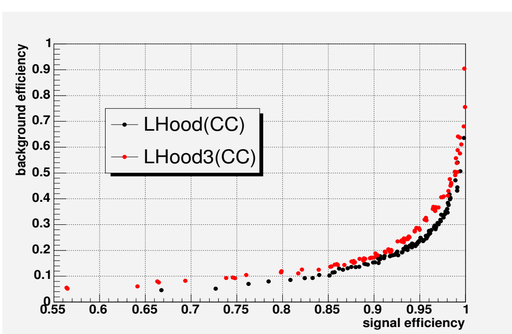


(a) CC region

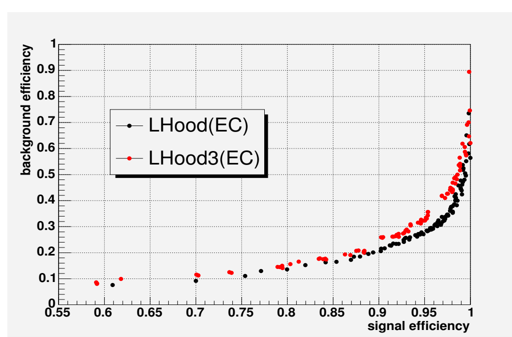


(b) EC region

Figure 13: 3-variable Likelihood distribution for signal (red) and background (blue) in p17 data.



(a) CC region

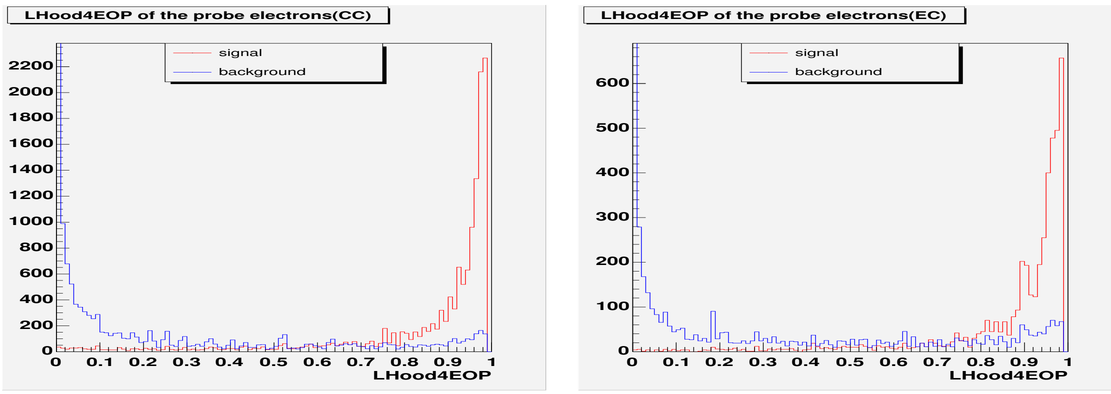


(b) EC region

Figure 14: Background efficiency vs. signal efficiency for LHood3 compared to LHood in CC and EC region in p17 data.

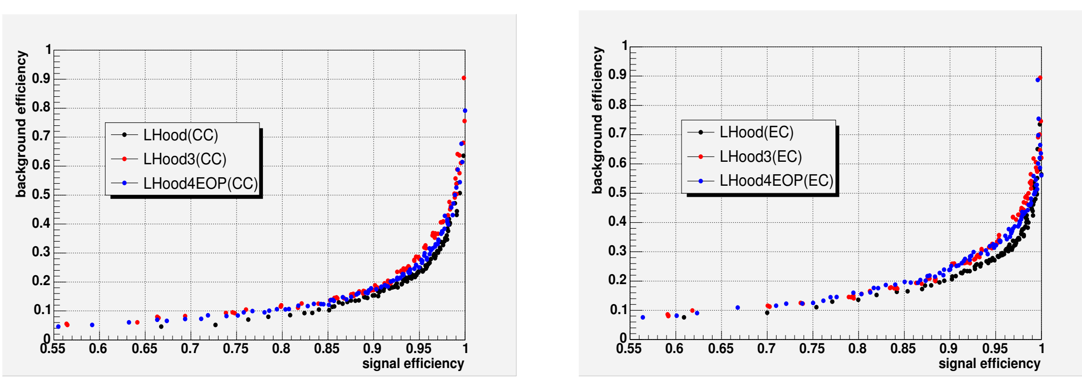
4.2 4-variable Likelihood

With the variable E_T/p_T added, the 4-variable Likelihood LHood4EOP performs better than LHood3 for p17 data as was the case for p14 data. Shown in Figure 15, is the Likelihood distributions for both signal and background. In Figure 16, the background efficiency vs. the signal efficiency is compared for the 3- ,4- and 7-variable Likelihood discriminants. In the CC region, for a signal efficiency of 90% LHood4EOP is better than LHood3 and almost reaches the performance of LHood. In the EC region, it is similar to LHood3 but significantly worse than LHood.



(a) CC region　　　(b) EC region

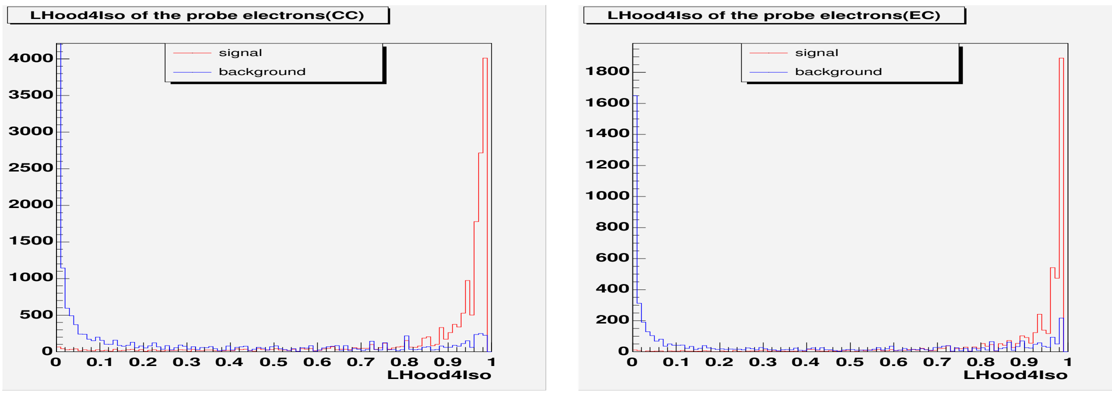
Figure 15: Lhood4EOP distribution for signal (red) and background (blue) in p17 data.



(a) CC region　　　(b) EC region

Figure 16: Background efficiency vs. signal efficiency for LHood4EOP compared to LHood3 and LHood in CC and EC region in p17 data.

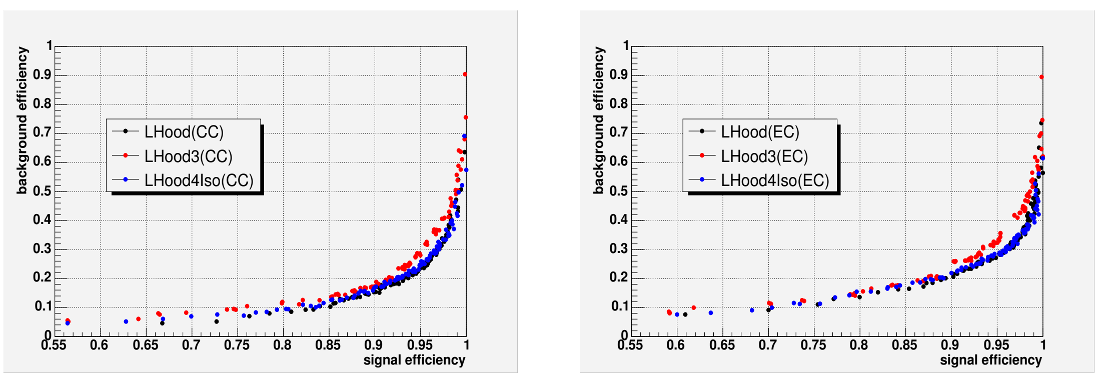
With the variable f_{iso} added, the 4-variable Likelihood LHood4Iso works better than LHood3. In Figure 17 the Likelihood distribution for both signal and background is shown. In Figure 18, the background efficiency vs. the signal efficiency is plotted for LHood3, LHood4Iso and LHood. In both the CC and the EC region, LHood4Iso delivers results similar to LHood. This is different from the case for p14 data.



(a) CC region

(b) EC region

Figure 17: LHood4Iso distribution for signal (red) and background (blue) in p17 data.



(a) CC region

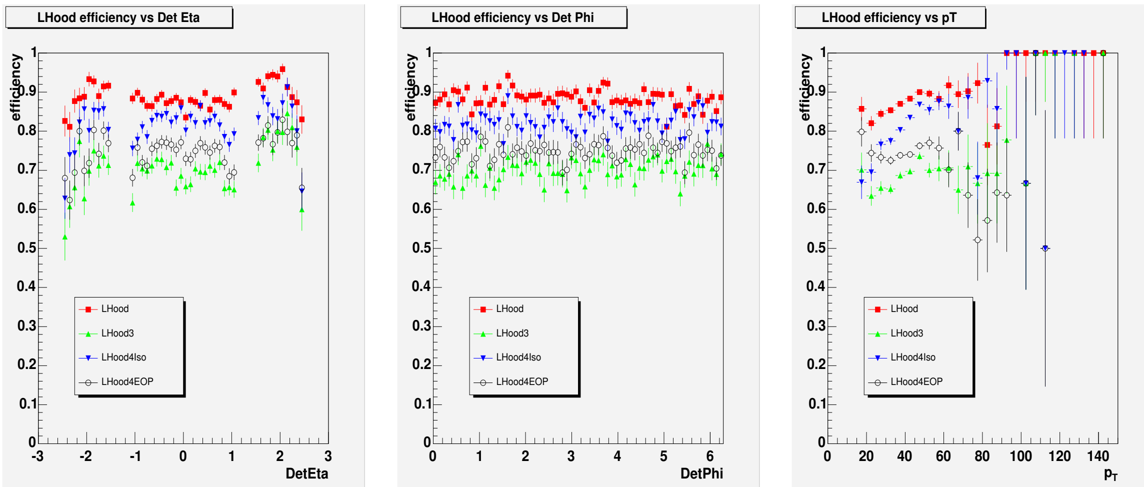
(b) EC region

Figure 18: Background efficiency vs. signal efficiency for LHood4Iso compared to LHood3 and LHood in CC and EC region in p17 data.

4.3 Kinematic Dependence of the Likelihood Efficiency

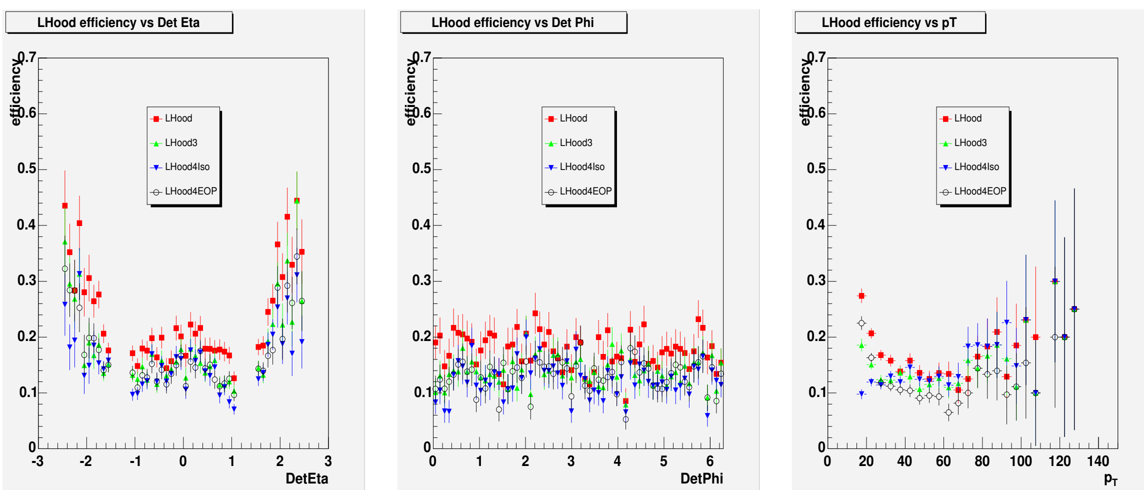
We examined the kinematic(p_T, η, ϕ) dependence of the Likelihood discriminants efficiencies for signal, as is shown in Figure 19. Here we require the

Likelihood to be larger than 0.85. Figure 20 shows the same dependences for the background. The Likelihood cuts are selected such that the corresponding signal efficiency is 85%.



(a) η_{det} dependence (b) ϕ_{det} dependence (c) p_T dependence

Figure 19: Kinematic dependence of the Likelihoods efficiencies for signal in p17 data.



(a) η_{det} dependence (b) ϕ_{det} dependence (c) p_T dependence

Figure 20: Kinematic dependence of the Likelihoods efficiencies for background in p17 data.

5 Conclusions

We measured the efficiencies for three new electron Likelihood discriminants. They perform pretty well comparing to the old Likelihood in both the CC and the EC region of the calorimeter. We also examined the kinematic dependences for the efficiencies. We have checked that this analysis is consistent with earlier results by comparing the results for p14 Pass2 data with [1]. We find a good agreement between these different analyses. We also looked at p17 reprocessed data. Here the new Likelihood discriminants show a similar performance as in p14 data. Furthermore we found that the 4-variable Likelihood LHood4Iso shows a similar performance as the 7-variable Likelihood currently used by default for physics analyses.

However, the Likelihood has to be retrained for p17 data. Then the comparison of the performance of the different Likelihood discriminants has to be redone. This will be the subject of a future study.

References

[1] J. Kozminski, R. Kehoe, H. Weerts, S.-J. Park, A. Quadt, J. Gardner, S. Jabeen , "Electron Likelihood in p14", DØ Note 4449, April 5, 2004.

[2] A. Kumar, J. Kozminski, R. Kehoe, J. Hays, J. Stark, "Electron Likelihood Study", DØ Note 4769, March 30, 2005.

[3] J. Hays, J. Mitrevski, C. Schwanenberger, T. Toole, "Electron Certification in p14", DØ Note 4783.

[4] J. Hays, J. Mitrevski, C. Schwanenberger, T. Toole, "Single Electron Efficiencies in p17 Data and Monte-Carlo", DØ Note 5025.

[5] J. Hays, J. Mitrevski, C. Schwanenberger, T. Toole, "Single Electron Efficiencies in p17 Data and Monte-Carlo Using p18.05.00 d0correct", DØ Note 5105.

[6] J. Hays, J. Mitrevski, C. Schwanenberger, "The Program Package em_cert: Version v00-01-35", DØ Note 4782.

[7] J. Hays, J. Mitrevski, C. Schwanenberger, "The Program Package em_cert: Version p18-br-20", DØ Note 5070.