Preshower Update

Brad Wray
The University of Texas at Austin
February 18, 2009
Topics

• What to do with EmcCands having no DIRC info

• MLP with different inputs
  – Different combination of input DIRC variables

• MLP trained over energy and theta bins

• Loose Ends
  – Dip in efficiency vs theta plots
New MLP Input Variables

Inputs to Current MLP

- nCrystals
- absZer20
- absZer42
- secondMomentTP
- lateral moment
- s1s9
- s9s25
- fitFailed
- nHits

Inputs to New MLP

- nCrystals
- absZer20
- absZer42
- secondMomentTP
- lateral moment
- s1s9
- s9s25
- thetaCError -f(θ) variable

Removed

- fitFailed

Added

- ndf (=nBkgPhotons+nHits)
- chiSqr
- rBkgSig (=nBkgPhotons/ndf)
- thetaC
Quality of DIRC Information

- DIRC info accessed through EmcCand via a pointer to an EmcPreshowerInfo object

- EmcPreshowerInfo object stores DIRC info which has been retrieved from DrcAbsThetaMeas object
Quality of DIRC Information

• Three possible scenarios
  – 1) EmcPreshowerInfo pointer is null (7%/6%)
    • Generally because “faked” track passes between DIRC bar boxes
  – 2) EmcPreshowerInfo pointer exists, but value of all DIRC ring fit variables is zero (pointer to DrcAbsThetaMeas object is null) (13%/33%)
    • Hard to track down, but appears to be due to some condition not being met in the DIRC software (no DIRC hits could be associated with our fake track)
  – 3) EmcPreshowerInfo pointer exists and DIRC variables are non-zero (80%/60%)
    • Best case scenario for a preshower

Percentage of true preshowers/non-preshowlers in Bbbar sample
Quality of DIRC Information

• Should we include candidates with no DIRC information?
  – Current MLP has purity of 46% and efficiency of 2% for cases 1 & 2.
  – Using the current MLP, including candidates from 1 & 2 increases the overall efficiency by 0.4% and decreases the overall purity by 0.5%

• When no DIRC information exists classify candidate as a preshower since we can only identify about 2% of them
“New” DIRC Variable

- ThetaC Error (rad) vs Theta (rad)
- Can we exploit differences in the shapes of the distributions?
  - Cut not effective (VERY low efficiency)
  - Find function \( f(\theta) \) which is different for preshowers and non-preshowers and use this as input to MLP
- Try to separate populations first

2-3 populations present in histograms

48 theta bins

Good DIRC Info
“New” DIRC Variable

- Isolate cands in these peaks

- Histogram of ndf for cands in big peak and little peak over all theta bins

- Few events lie in big peak past ndf = 4 for preshowers → Use ndf ≥ 4 as means to separate populations
“New” DIRC Variable

- Populations seem to be correlated with $ndf$:

<table>
<thead>
<tr>
<th></th>
<th>$ndf &lt; 4$</th>
<th>$ndf \geq 4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preshowers</td>
<td>20%</td>
<td>60%</td>
</tr>
<tr>
<td>Non-preshowers</td>
<td>40%</td>
<td>20%</td>
</tr>
</tbody>
</table>
“New” DIRC Variable

Preshowers: Theta Bin = 10

- ndf < 4

- ndf ≥ 4

Preshowers: Theta Bin = 30

- ndf < 4

- ndf ≥ 4
“New” DIRC Variable

• Construct $f(\theta)$ by fitting peak values (central value of most populated bin) of thetaC Error vs theta for preshowers with fifth order polynomial for $\text{ndf} \geq 4$

• New variable is thetaCError – $f(\theta)$

• Try as input to MLP with the following DIRC vars: thetaC, ndf, chiSqr, rBkgSig ($= \frac{n\text{BkgPhotons}}{\text{ndf}}$)
Results Using New Input Variables

Cuts chosen to give same efficiency in order to make fair comparison of purities.
Results Using New Input Variables

• ~3% improvement in purity using new of input variables

• Should we keep cands with ndf<4
  – Look very non-preshower like
  – Under current MLP these cands have efficiency of 6% and purity of 41%
  – Contribute 1.2 % to overall efficiency in current scheme and decrease the purity by 1.7%
Binned Training

Is the MLP cut properly optimized in these regions?
Binned Training (Theta)

Trained & tested in 3 theta bins
\[ \theta_{\text{min}} \leq \theta \leq 0.75 \text{ rad}, \ 1.0 \text{ rad} \leq \theta \leq 1.4 \text{ rad}, \ 1.7 \text{ rad} \leq \theta \leq \theta_{\text{max}} \]

Applied correction in 3 different theta bins
\[ \theta_{\text{min}} \leq \theta < 0.8 \text{ rad}, \ 0.8 \text{ rad} \leq \theta < 1.6 \text{ rad}, \ 1.6 \text{ rad} \leq \theta \leq \theta_{\text{max}} \]

\[ \text{Eff} = 42.2\%/42.3\% \]
\[ \text{Pur} = 55.8\%/55.7\% \]

Cuts chosen to give same efficiency in order to make fair comparison of purities.

Binned Training (Energy)

Trained & tested in 2 energy bins
100 MeV ≤ E ≤ 300 MeV, 600 MeV < E

Applied correction in 3 different theta bins
50 MeV ≤ E ≤ 450 MeV, 450 MeV < E

unbinned/binned

Eff = 42.2%/44.6%

Pur = 55.8%/52.8%

Cuts chosen to give same efficiency in order to make fair comparison of purities.
Binned Training (Energy & Theta)

Cuts chosen to give same efficiency in order to make fair comparison of purities.
## Energy & Theta Binned Results

### Efficiency (unbinned/binned)

<table>
<thead>
<tr>
<th></th>
<th>Theta Bin 1</th>
<th>Theta Bin 2</th>
<th>Theta Bin 3</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>E Bin 1</td>
<td>41.6%/41.8%</td>
<td>33.8%/33.6%</td>
<td>50.1%/50.0%</td>
<td>39.6%/39.6%</td>
</tr>
<tr>
<td>E Bin 2</td>
<td>57.7%/57.3%</td>
<td>50.5%/50.2%</td>
<td>63.3%/63.6%</td>
<td>54.7%/54.4%</td>
</tr>
</tbody>
</table>

### Purity (unbinned/binned)

<table>
<thead>
<tr>
<th></th>
<th>Theta Bin 1</th>
<th>Theta Bin 2</th>
<th>Theta Bin 3</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>E Bin 1</td>
<td>49.7%/49.6%</td>
<td>56.0%/54.8%</td>
<td>59.0%/59.4%</td>
<td>54.7%/54.2%</td>
</tr>
<tr>
<td>E Bin 2</td>
<td>50.8%/53.4%</td>
<td>58.5%/61.7%</td>
<td>60.9%/60.5%</td>
<td>55.1%/57.6%</td>
</tr>
</tbody>
</table>

No real improvement in any binning scenario
Dip in Efficiency vs Theta

<nHits> vs theta has similar shape to efficiency vs theta
Dip in Efficiency vs Theta

Avg number of reconstructed DIRC photons in dimuon events

- In communication with Dave Aston and Nicolas Arnaud. Dave seemed to think dip in $<n_{\text{Hits}}>$ in BBbar is to be expected and that the flatness of $<n_{\text{Hits}}>$ in MMG is odd. Also thought drop off of $<n_{\text{Hits}}>$ above $\cos \theta = 0.7$ ($\theta < 0.8$ rad) in BBbar to be odd.

- Suggested making the $<n_{\text{Hits}}>$ plots for data and trying to isolate the neutral clusters more (make sure there are no other clusters near by, charged or neutral)

Conclusions

• Candidates with no DIRC info are classified as non-preshowers

• Should MLP with different variables be used since a small improvement is possible?
  – What should be done with ndf<4?

• Binning in theta, energy or energy & theta
  – Only questionable improvement and increasing complexity for implementation and user.

• Will produce a note describing studies and documenting instructions for use of the preshower correction