1. Numu   CC/ NC Separation in the ND

2. Slicing & Reco Efficiencies

N. Saoulidou, Fermilab, NC phone meeting 11-11-04

Outline

Slicing & Reco Efficiencies in the ND

- Slicing efficiency for CC & NC events. Overall numbers and also as a function of
  - Visible energy in the detector
  - True Neutrino energy
  - Y
  - Muon momentum (CC) & Hadronic shower energy (NC)

- Event reconstruction efficiency for CC & NC events. Overall numbers and also as a function of
  - Visible energy in the detector
  - True Neutrino energy
  - Y
  - Muon momentum (CC) & Hadronic shower energy (NC)

NC / CC Separation

- Method (ANN)

- Sample

- Updated Results

- A few event displays of misclassifications...CC categorized as NC-LIKE

- Summary – On going work

Visible energy in the Detector

- So far I was extracting slicing and reconstruction efficiencies using the MC truth information to decide as to whether a CC or NC event "existed" or not. But I wanted to visually scan ALL the actual MC (not reco with MC or slicing with some MC info) events.

- I modified the "Cheezy Display" (Roberts package) in order to be able to visually scan through the digits of each MC event in the detector (U Z & VZ). After scanning through a few events I realized (maybe late…?) that there is a fare number of them (especially NC events) that have ~ "NOTHING" visible in the detector.

- After doing that, I also modified the NtpMCTruth code in order to be able to record for every MC event the total visible PH (in ADC counts) per view and the total number of detector digits per view in the standard NtpMC ntuple.

- Then having that information I recalculated the slicing and event reconstruction efficiencies for CC and NC events and as a function of various quantities.
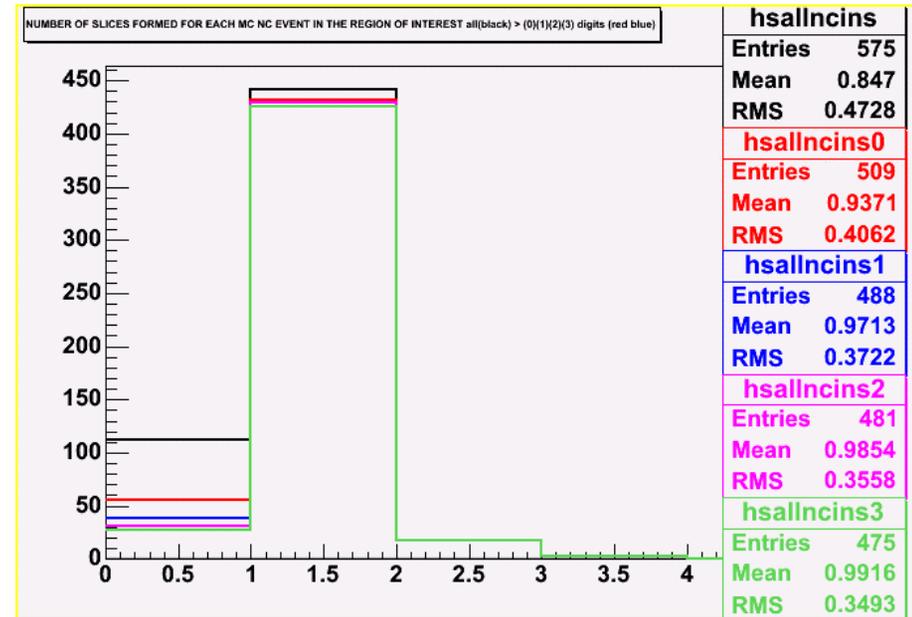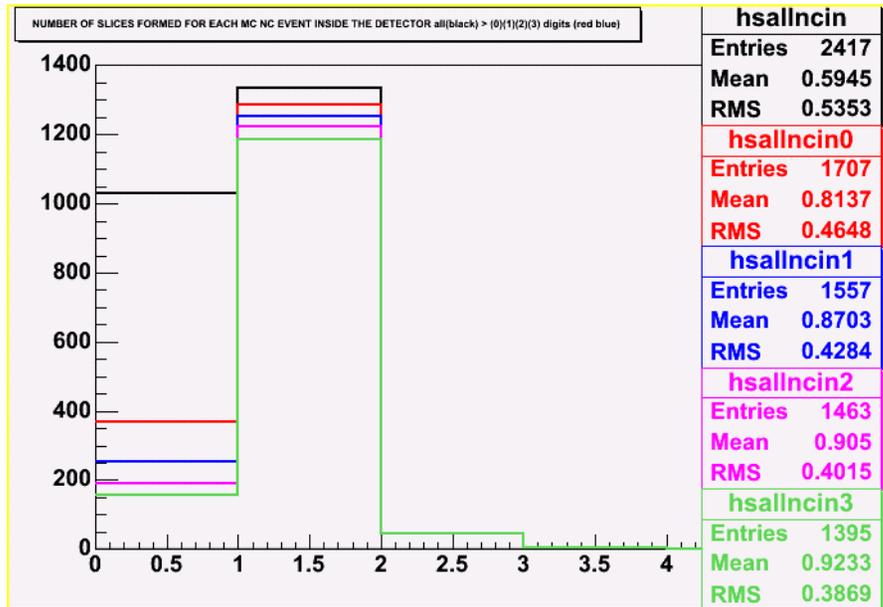
Various Cuts & definitions

- The left plots that I am showing next are for MC events occurring anywhere **inside the calorimeter**, and the right for MC events **1m around the beam center and 6 < start-plane < 110.**

- The various colors represent the following:
  - **NO CUTS on PH (or equivalently number of digits)**

  - **# Digits in each view > 0**
  - **# Digits in each view > 1**
  - **# Digits in each view > 2**
  - **# Digits in each view > 3**

- The major difference between the first region (all calorimeter) and the second (1 m around the beam) is the partially and fully covered planes (the second region "sees" only fully covered planes, while the first sees both , fully covered and 1 /5 covered)

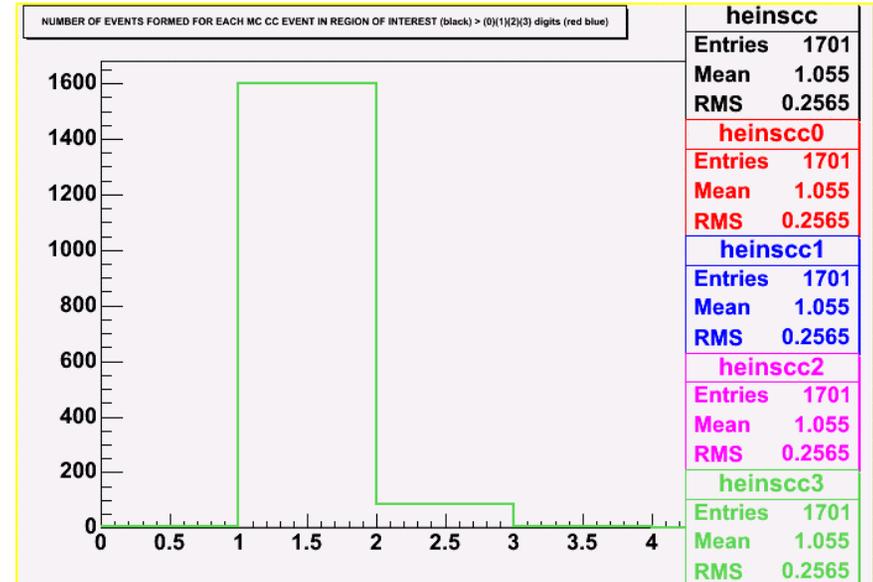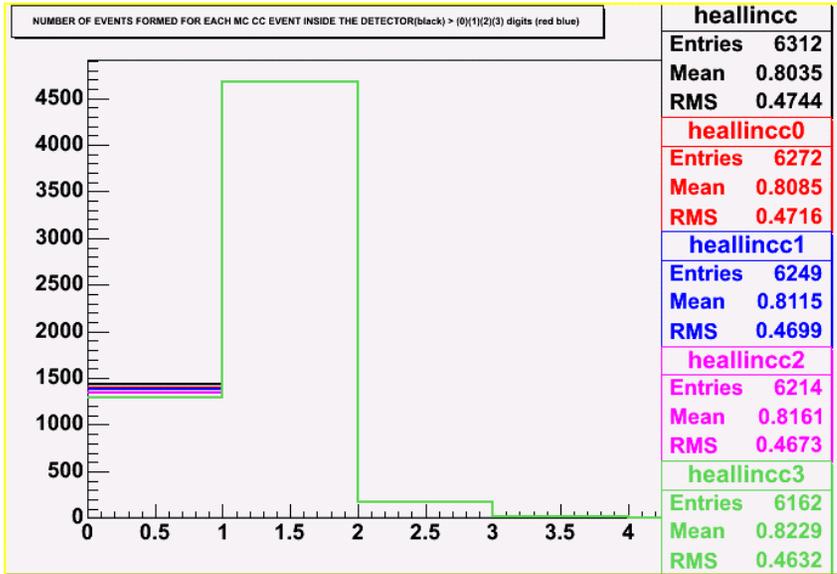# Slicing efficiency for CC events: Total Numbers
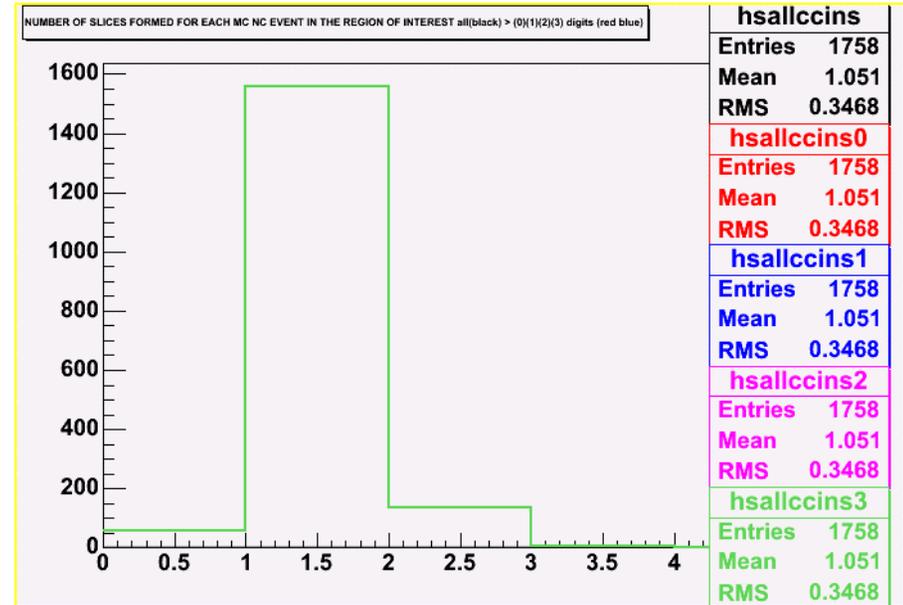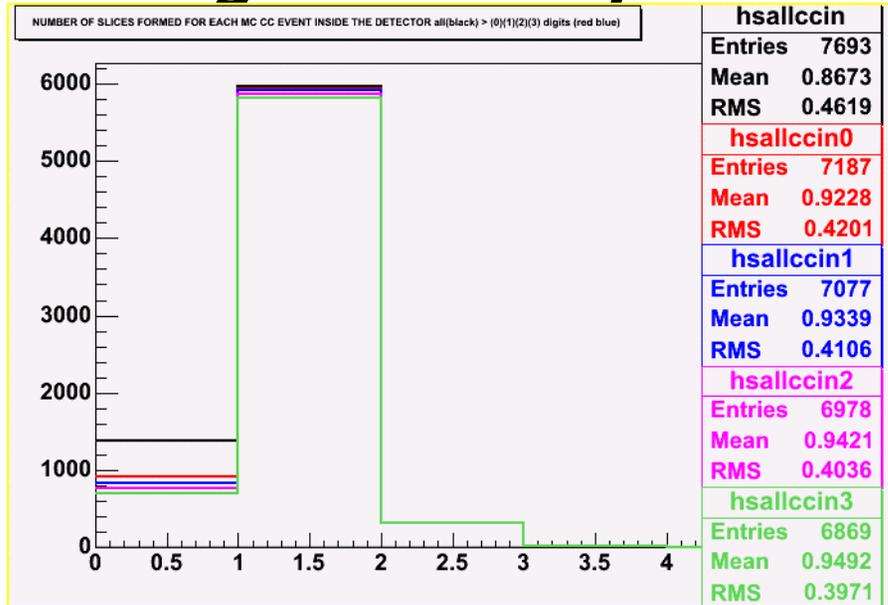
- These plots show the Number of slices one gets per MC event :
  - 0 means that this particular MC event does not have a slice pointing back to it. Slices point back to MC events based on their majority of charge weighted hits. This category might have small MC events totally eaten up by bigger slices or MC events split up in small pieces in different slices or small MC events not satisfying the slicing criteria. The slicing criteria are pretty loose : More than 2 strips and strip PH > 2 PEs.
  - 1 means that the MC event has a slice pointing to it.
  - 2 means that the MC events was split in two, 3 in three e.t.c.
- Initially without any cut for the "larger" calorimeter region 18 % of the CC MC events don't have a slice…
- Applying the first basic cut, "At least 1 digit per view", ~half of the MC events with no slice disappear, and the percentage of CC events with no slice decrease to 7 % of the total.

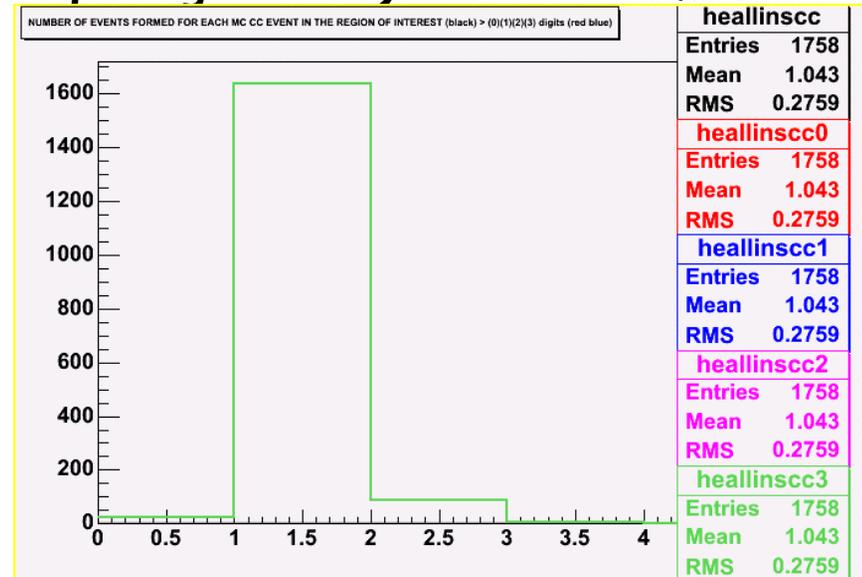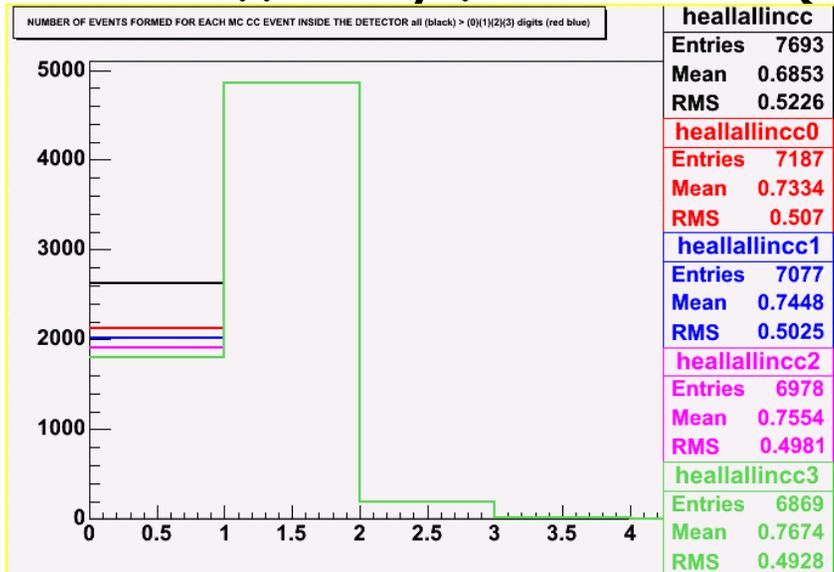# Slicing efficiency for NC events: Total Numbers

hsallncin	
Entries	2417
Mean	0.5945
RMS	0.5353

hsallncin0	
Entries	1707
Mean	0.8137
RMS	0.4648

hsallncin1	
Entries	1557
Mean	0.8703
RMS	0.4284

hsallncin2	
Entries	1463
Mean	0.905
RMS	0.4015

hsallncin3	
Entries	1395
Mean	0.9233
RMS	0.3869

hsallncins	
Entries	575
Mean	0.847
RMS	0.4728

hsallncins0	
Entries	509
Mean	0.9371
RMS	0.4062

hsallncins1	
Entries	488
Mean	0.9713
RMS	0.3722

hsallncins2	
Entries	481
Mean	0.9854
RMS	0.3558

hsallncins3	
Entries	475
Mean	0.9916
RMS	0.3493

- Initially without any cut for the "larger" calorimeter region 43 % of the NC MC events don't have a slice...

- Applying the first basic cut, "At least 1 digit per view", 60% of the MC events with no slice disappear, and the percentage of NC events with no slice decrease to 22 % of the total.

- In the "fully" covered area the percentage of NC events with no slice is smaller : 19% and applying the minimal cut it further decreases to 10 % going down to 6 % with the 3 (4 & 5) digit per view requirement.

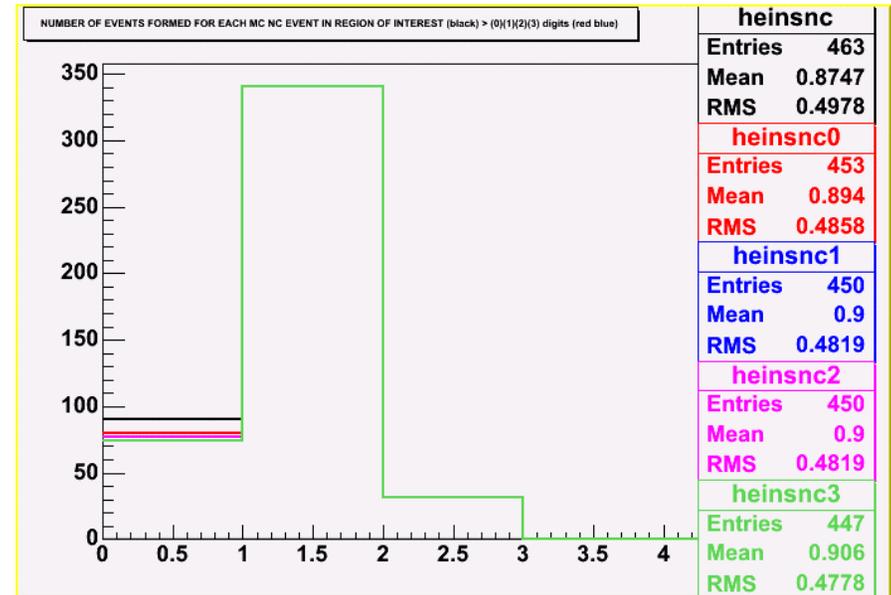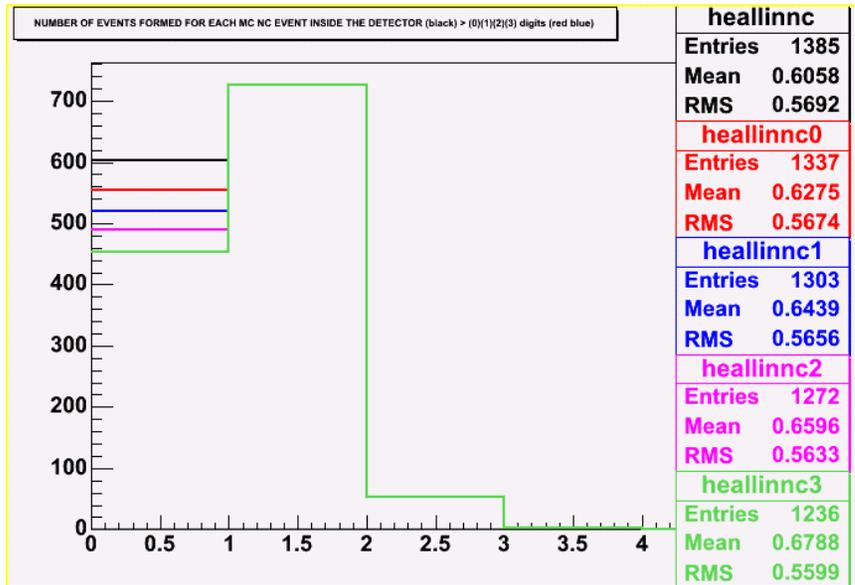# Reco efficiency for CC events (requiring at least 1 slice) : Total Numbers

Left plot — **heallincc** (NUMBER OF EVENTS FORMED FOR EACH MC CC EVENT INSIDE THE DETECTOR(black) > (0)(1)(2)(3) digits (red blue)):

	Entries	Mean	RMS
heallincc	6312	0.8035	0.4744
heallincc0	6272	0.8085	0.4716
heallincc1	6249	0.8115	0.4699
heallincc2	6214	0.8161	0.4673
heallincc3	6162	0.8229	0.4632

Right plot — **heinscc** (NUMBER OF EVENTS FORMED FOR EACH MC CC EVENT IN REGION OF INTEREST (black) > (0)(1)(2)(3) digits (red blue)):

	Entries	Mean	RMS
heinscc	1701	1.055	0.2565
heinscc0	1701	1.055	0.2565
heinscc1	1701	1.055	0.2565
heinscc2	1701	1.055	0.2565
heinscc3	1701	1.055	0.2565

- In order to study reconstruction efficiencies we can :
  - Require a slice for this particular MC event and check if it is reconstructed or not.
  - Not require a slice. This way we can examine how many MC events are eaten up by bigger ones but are still reconstructed.
- These plots show the number of reconstructed events that point back to a particular MC event (the looping here is done over all MC events that had a slice)
- In the "larger" calorimeter region the reco efficiency **(I am not examining here, and in the slices also, the quality of the reconstruction i.e completeness, purity e.t.c)** for CC events is ~ 78 % and does not change significantly after imposing the "digit" criteria .
- In the "fully" covered area the reconstruction efficiency for CC events that had a slice is 100 %.

Reco efficiency for CC events (not requiring 1 slice) : Total Numbers

| heallallincc	
Entries	7693
Mean	0.6853
RMS	0.5226

heallallincc0	
Entries	7187
Mean	0.7334
RMS	0.507

heallallincc1	
Entries	7077
Mean	0.7448
RMS	0.5025

heallallincc2	
Entries	6978
Mean	0.7554
RMS	0.4981

heallallincc3	
Entries	6869
Mean	0.7674
RMS	0.4928

heallinscc	
Entries	1758
Mean	1.043
RMS	0.2759

heallinscc0	
Entries	1758
Mean	1.043
RMS	0.2759

heallinscc1	
Entries	1758
Mean	1.043
RMS	0.2759

heallinscc2	
Entries	1758
Mean	1.043
RMS	0.2759

heallinscc3	
Entries	1758
Mean	1.043
RMS	0.2759

NUMBER OF EVENTS FORMED FOR EACH MC CC EVENT INSIDE THE DETECTOR all (black) > (0)(1)(2)(3) digits (red blue)

NUMBER OF EVENTS FORMED FOR EACH MC CC EVENT IN THE REGION OF INTEREST (black) > (0)(1)(2)(3) digits (red blue)

- These plots show the number of reconstructed CC events that point back to a particular MC event (the looping here is done over all MC events regardless of whether they had a slice or not).

- In the "larger" calorimeter region the reco efficiency for CC events is initially (no digit cut) ~ 66 % and increase to ~ 71% after imposing the "digit" criteria. What is interesting here (and I need to examine by visual scan) is that the number of MC CC events that are reconstructed (the 1 bin) is higher (500 events more) than the previous plot were I was requiring at least one slice. That means that MC events that don't point back to a slice (possibly merged with other) do get reconstructed.

- In the "fully" covered area the reconstruction efficiency for CC events is 99 % and does not change with the "digit" cuts. Also the number of MC CC that are reconstructed but did not have a slice (previous slide) is nearly zero here.

Reco efficiency for NC events (requiring at least 1 slice) : Total Numbers

heallinnc	
Entries	1385
Mean	0.6058
RMS	0.5692
heallinnc0	
Entries	1337
Mean	0.6275
RMS	0.5674
heallinnc1	
Entries	1303
Mean	0.6439
RMS	0.5656
heallinnc2	
Entries	1272
Mean	0.6596
RMS	0.5633
heallinnc3	
Entries	1236
Mean	0.6788
RMS	0.5599

NUMBER OF EVENTS FORMED FOR EACH MC NC EVENT INSIDE THE DETECTOR (black) > (0)(1)(2)(3) digits (red blue)

heinsnc	
Entries	463
Mean	0.8747
RMS	0.4978
heinsnc0	
Entries	453
Mean	0.894
RMS	0.4858
heinsnc1	
Entries	450
Mean	0.9
RMS	0.4819
heinsnc2	
Entries	450
Mean	0.9
RMS	0.4819
heinsnc3	
Entries	447
Mean	0.906
RMS	0.4778

NUMBER OF EVENTS FORMED FOR EACH MC NC EVENT IN REGION OF INTEREST (black) > (0)(1)(2)(3) digits (red blue)

- In the "larger" calorimeter region the reco efficiency **(I am not examining here, and in the slices also, the quality of the reconstruction i.e completeness, purity e.t.c)** for NC events is initially (with no digit cut) ~ 57 % increases to ~ 60% after imposing the digit cuts .

- In the "fully" covered area the reconstruction efficiency for NC events that had a slice is initially ~ 81 % and increases slightly to 82% after imposing the digit criteria.
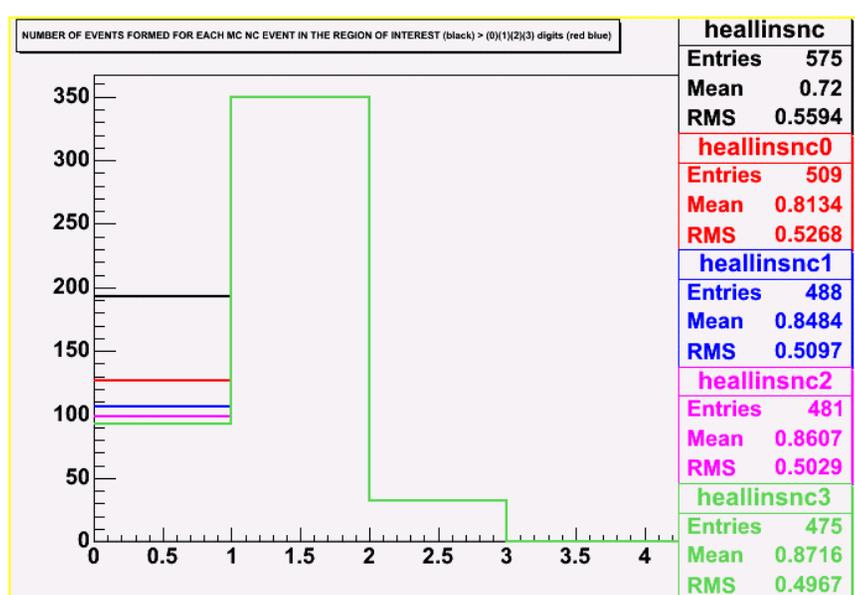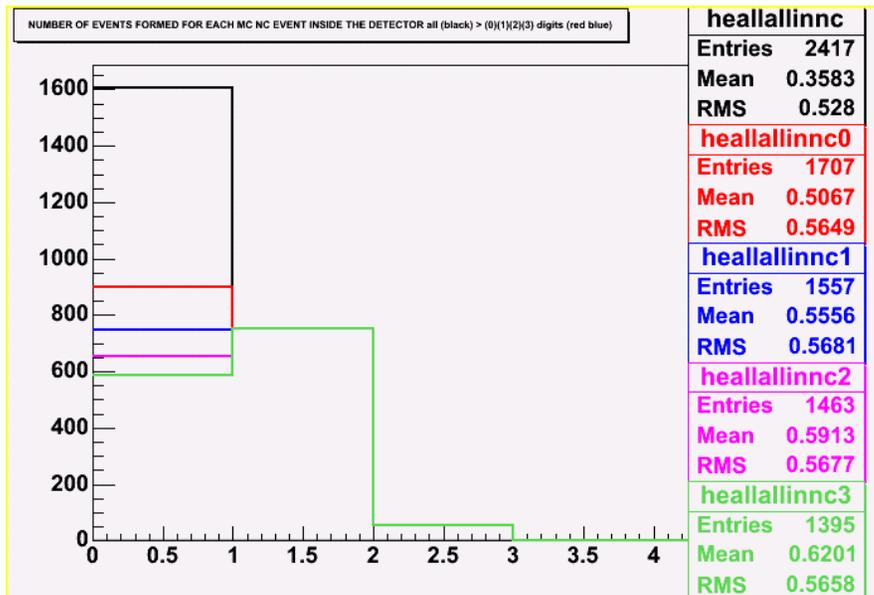
- In the "larger" calorimeter region the reco efficiency for NC events is initially (no digit cut) ~ 34 % and increase to ~ 47 % after imposing the "digit" criteria. Again what is interesting here is that the number of MC NC events that are reconstructed (the 1 bin) is higher (100 events more) than the previous plot were I was requiring at least one slice. That means that MC events that don't point back to a slice (possibly merged with other) do get reconstructed.

- In the "fully" covered area the reconstruction efficiency for NC events initially is 67 % and increases to significantly to 75 -80 % with the "digit" cuts. Also the number of MC CC that are reconstructed but did not have a slice (previous slide) is very small zero.

Efficiencies (slice & reco) as a function of other quantities

- After examining the total number for the slicing and reco efficiencies I wanted also to see them as a function of :

    - Visible energy
    - True neutrino energy
    - Y
    - Muon momentum and shower true energy

- I did that only for events in the region of interest (1 m around the beam) and requiring:

    - At last 1 digit in each view for the slices
    - At least 2 digits in each view for the events

- I also tested what kind of MC events are the ones that have 0 digits in either view (regardless of whether they are sliced or reconstructed) and that is what I am going to show next. This served as a "sanity" check since we expect these events to have certain characteristics (especially Y distribution, muon momentum, shower true energy e.t.c). Again I am going to show the plots only for the region of 1 m around the beam (the rest are also available if someone wants to look at them).

Etrue,Y, Muon momentum & Evis of CC events (all MC events and remaining after digit cuts)
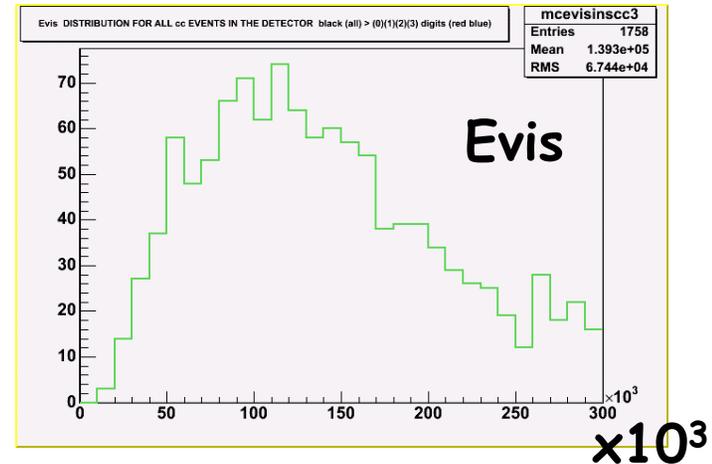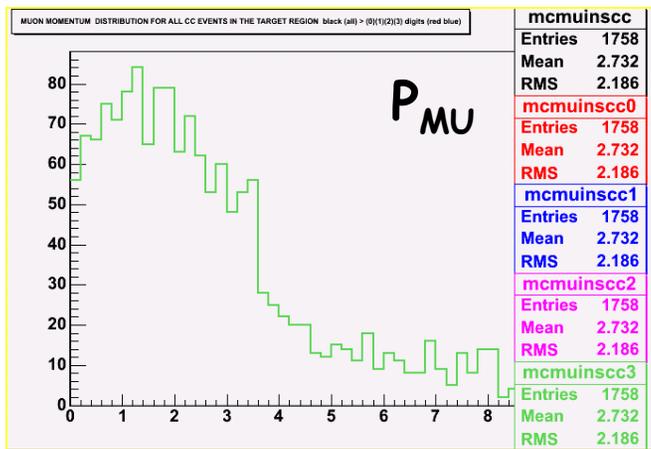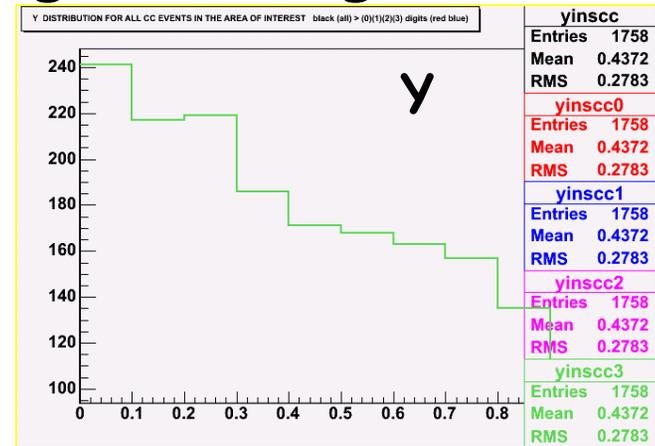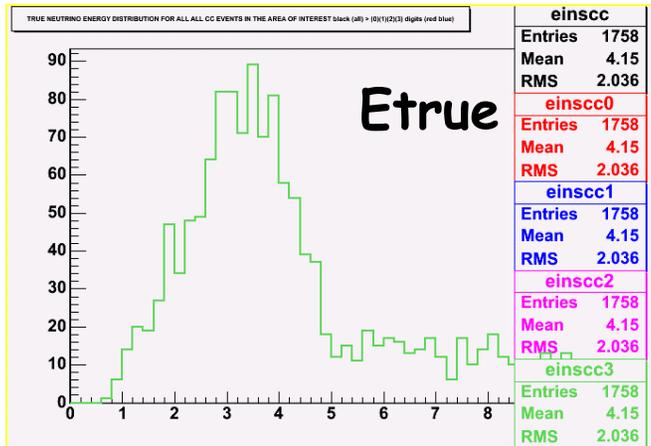
- All MC CC events in the "fully" covered calorimeter region have at least 5 digits in each view.

Etrue, Y, Eshw & Evis of NC events (all MC events and remaining after digit cuts)
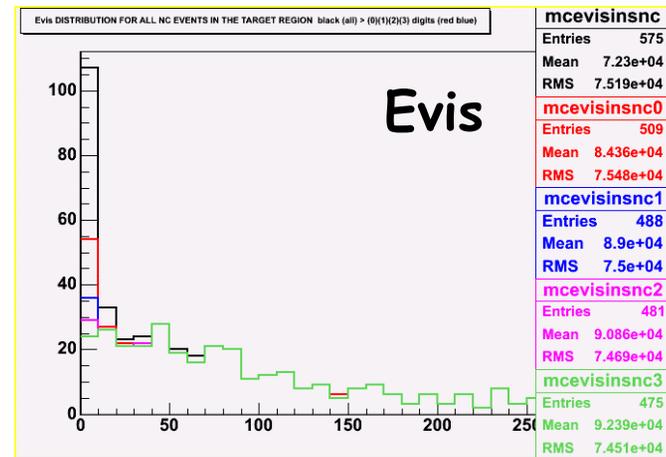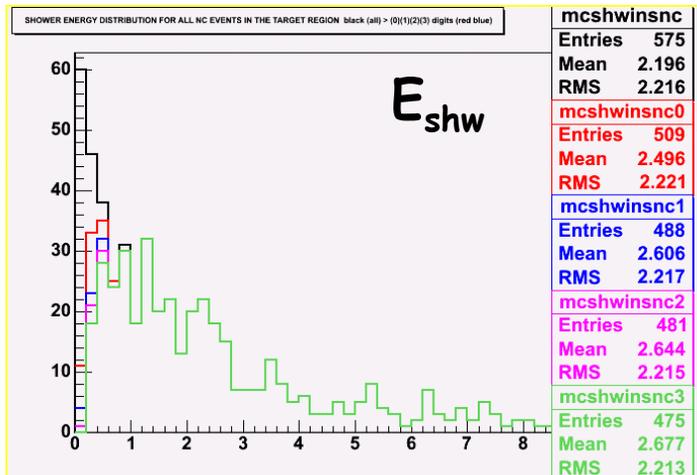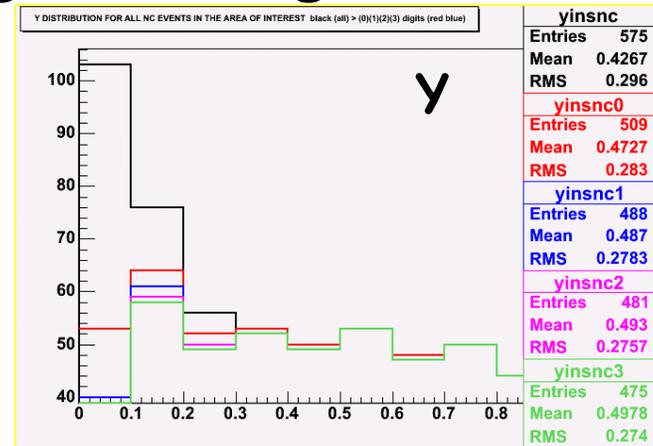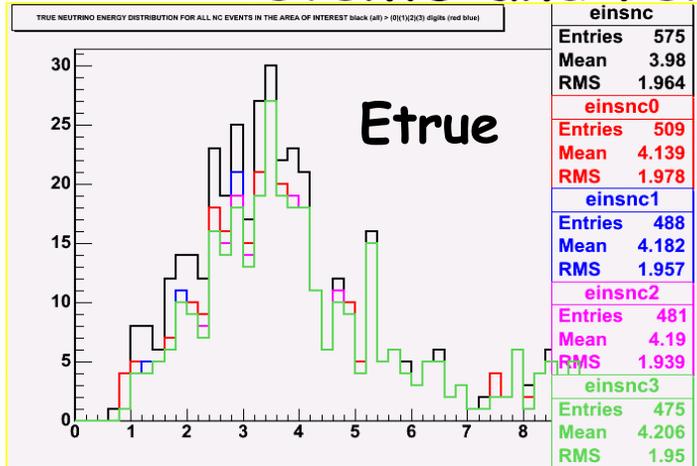
×10³

- Low Y (low hadronic shower energy) NC events are the ones that are cut after applying the digit cut (as expected).

Etrue,Y, Muon momentum & Evis for all MC CC events and the ones that are sliced.
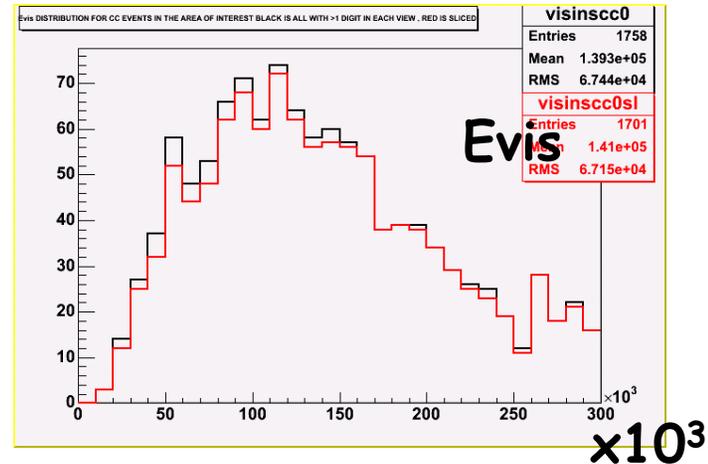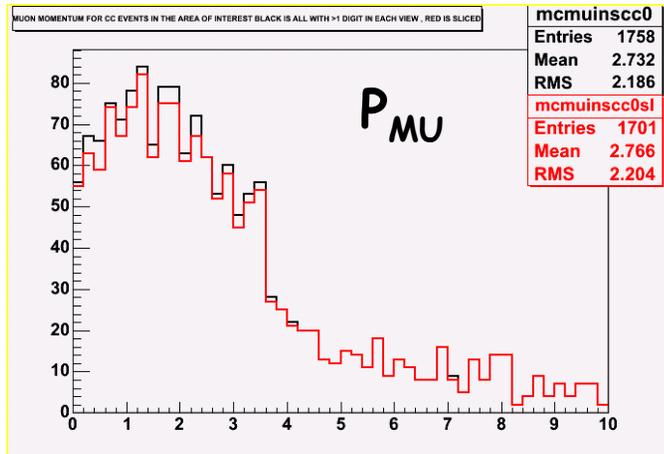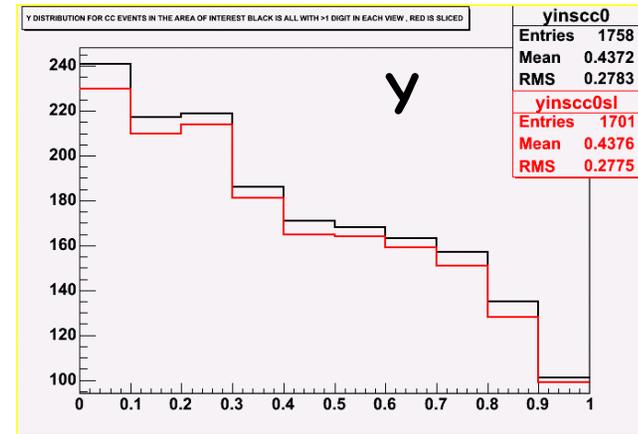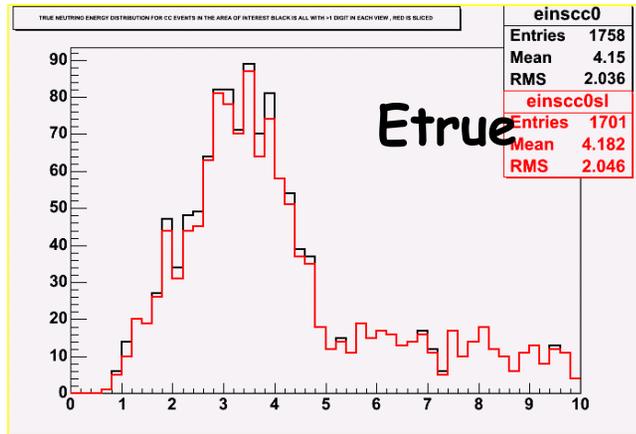
• Nearly all MC CC events in the "fully" covered calorimeter region (with more than 1 digit in each view) are sliced. There are very few missed in low y that are most probably slicing failures.

Etrue, Y, Muon momentum & Evis for all MC NC events and the ones that are sliced.
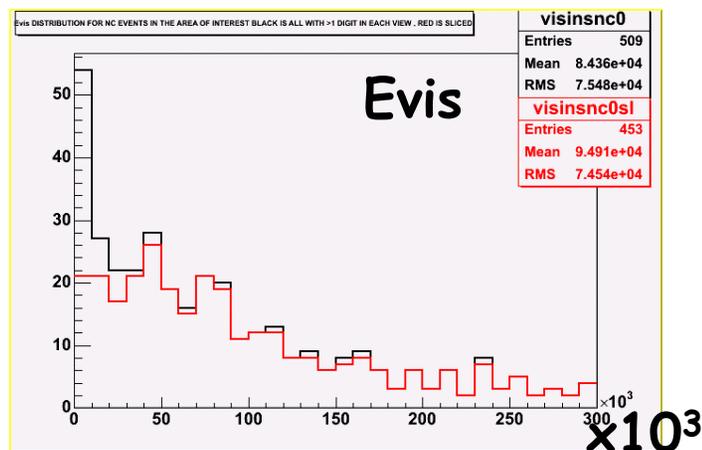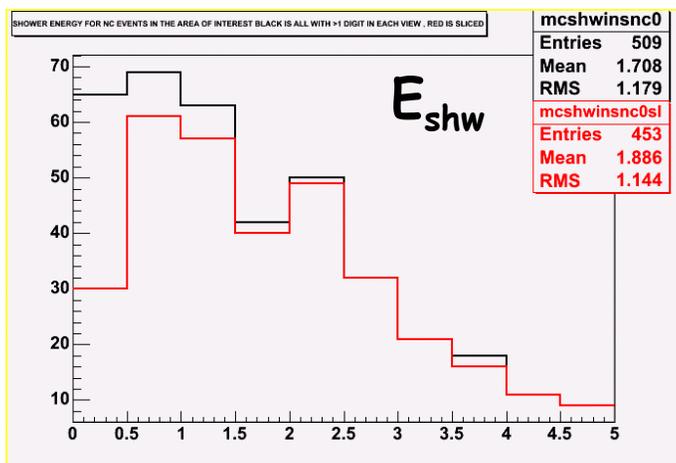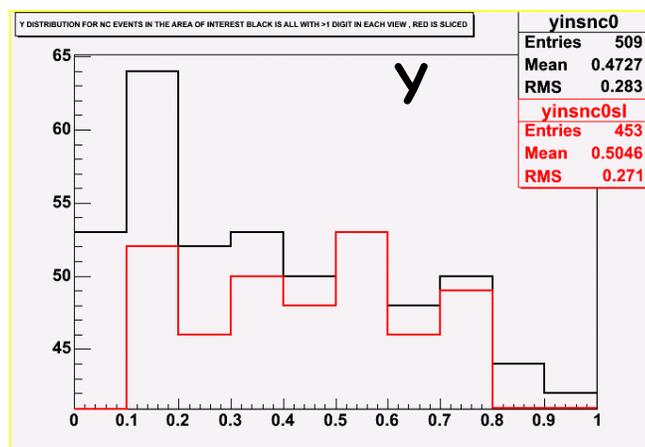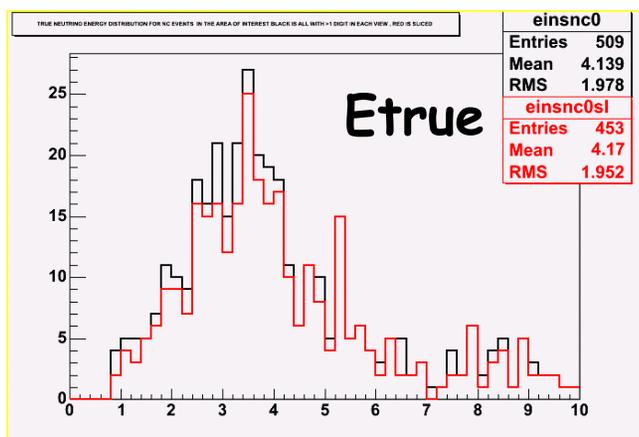
- Nearly 90 % of MC NC events in the "fully" covered calorimeter region (that have at least 1 digit in each view) are sliced. Most of the events missed are in low y as expected.

Etrue,Y, Muon momentum & Evis for all MC CC events and the ones that are reconstructed.
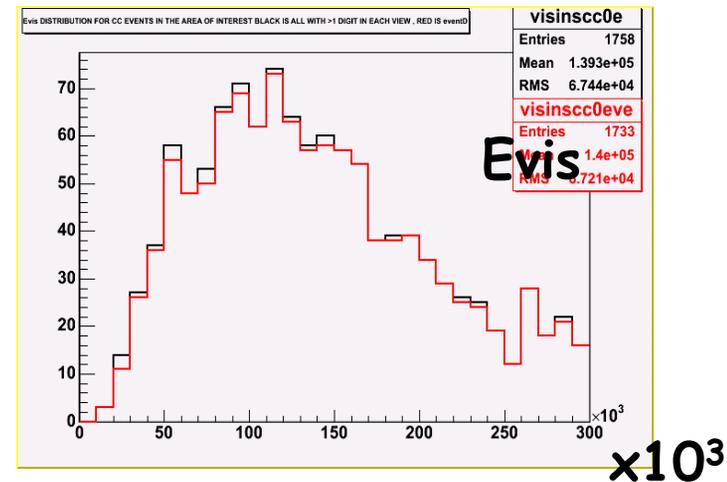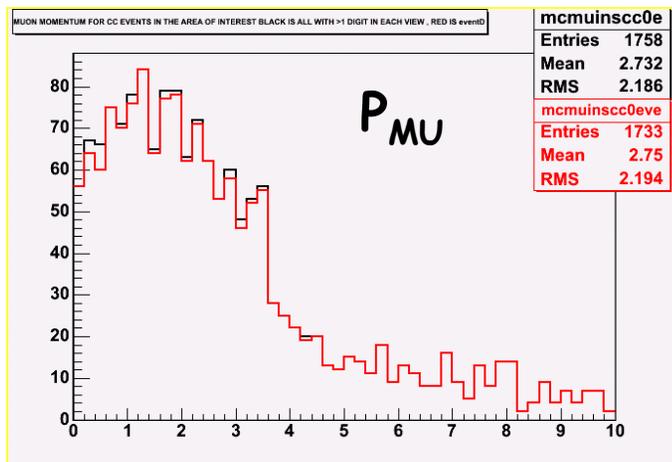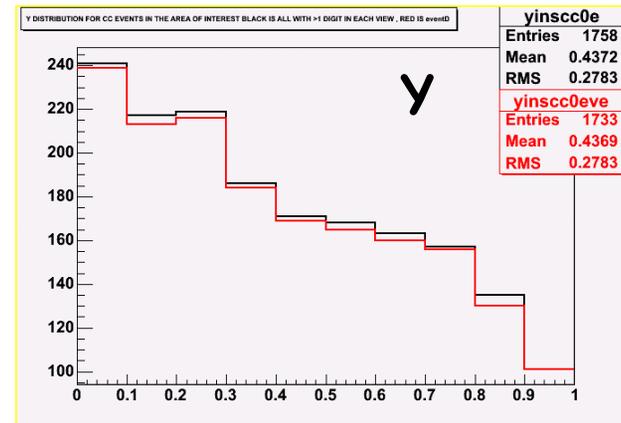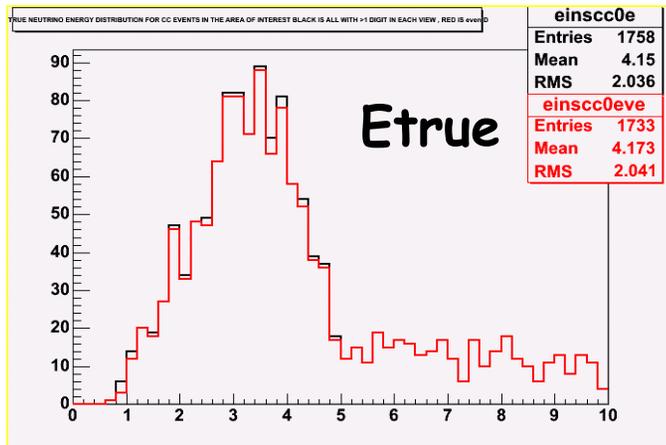
• Nearly all MC CC events in the "fully" covered calorimeter region (with at least 2 digits in each view) are reconstructed (one way or another).

Etrue, Y, Eshw & Evis for all MC NC events and the ones that are reconstructed.
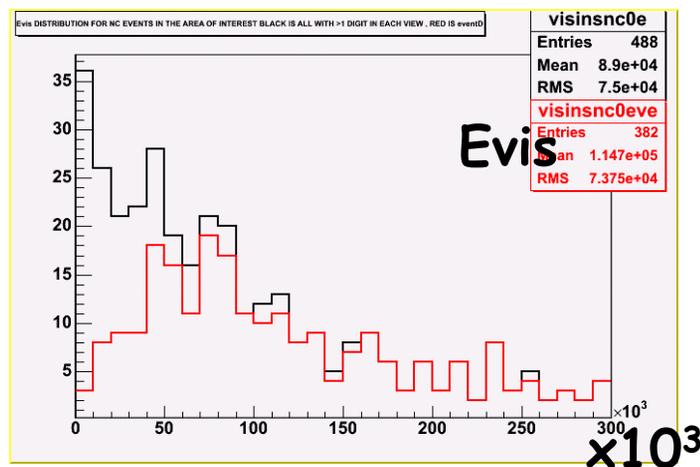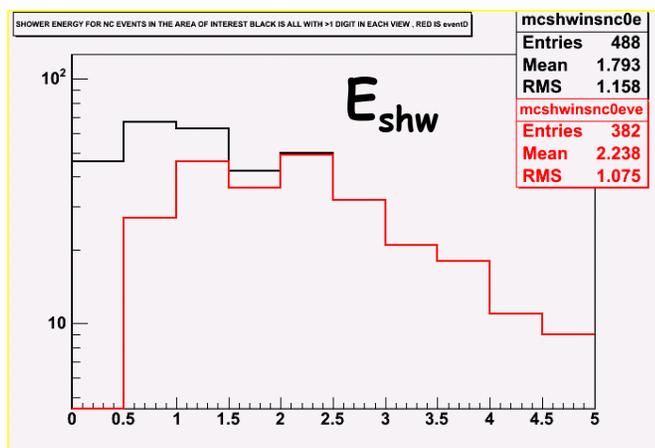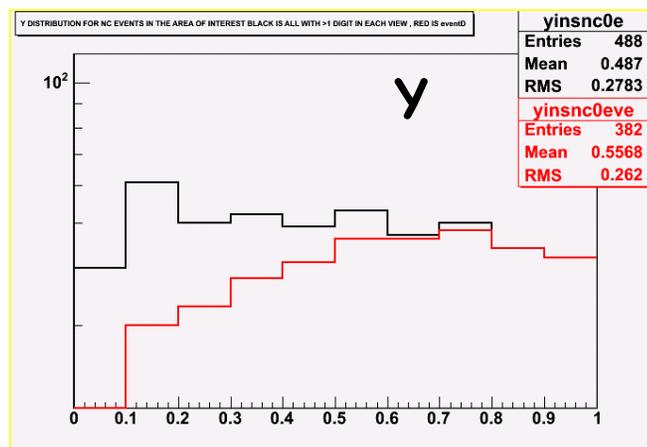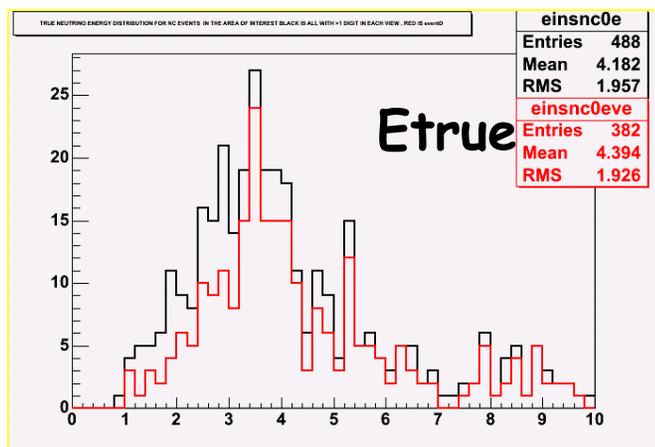
• 78% of the MC NC events in the "fully" covered calorimeter region (with at least 2 digits in each view) are reconstructed. The 20 % that are missed are low y events with very small visible energy (below 30×10^3 ADC counts). Maybe a fraction can be recoverable (I will scan them and have more next time)

Slicing & Reco Efficiencies : Summary

- After imposing a simple "digit" cut (at least 1 digit in each view) the slicing efficiencies in the central (1m around the beam) region of the detector are (modulo a small double-counting fraction):

  - 100 % for CC Events

  - 90 % for NC Events

- After imposing a simple "digit" cut (at least 2 digits in each view) the reconstruction efficiencies in the central (1m around the beam) region of the detector are (modulo a small double-counting fraction):

  - 99 % for CC Events

  - 78 % for NC Events

- The addition of the total PH & the total number of digits in each view in the NtpMC ntuple were quite useful for this simple analysis. Also the addition of the true digit information for overlay events (for single   event is what Robert had written already) in the Cheezy Display are is also quite useful if one wants to visually scan the actual MC event in the detector.(I don't know fo any display offering that option but maybe I am wrong…)

ANN for NC/CC Separation : Method, Events used, & additions from the discussion in the previous NC meeting.

- Reconstructed, using development release, 9 MDC ND overlaid files writing out NtpSR, NtpMC, NtpTH (Jims latest truth variables are ABSOLUTELLY essential for any kind of analysis that needs to know CC and NC events in overlaid files)

- The statistics are still quite poor but with the R1.12 MDC production things will get better…

- This time I only studied events in the central region (1m around the beam center) since these are the ones we will be looking for physics analysis. The geometry of the detector in the partially covered region introduces strong biases of all sorts CC vs NC, Near vs Far e.t.c.

- Used a plane cut to select an extremely pure sample of CC events (52 % of CC events).

- Constructed an ANN (Usinf the MLPfit package) to classify the remaining CC and NC events with total length < 40 planes.

- Replaced any variable that had PH in PEs with PH in SigCorr (after various people suggesting that in the previous meeting) and used trklike number of planes instead of trk (after Mayly's suggestion).

- Visually scanned misclassified CC events as NC-LIKE (I have a few examples)

ANN architecture & Input variables

ANN Architecture
7 inputs
2 hidden layers with 4 & 3 neurons respectively
1 output

Input Variables :

Pulse height per plane
Pulse height per strip
Number of tracks
Number of planes
Number of strips
Total PH
Shower PH per digit

I have examined a large number of variables (including the ones I presented in the previous meeting.)
I am using the ones that seem to have the greatest separation power this new set of events (last time I was examining events anywhere in the calorimeter and not only in the central part).

Numu CC – NC variables
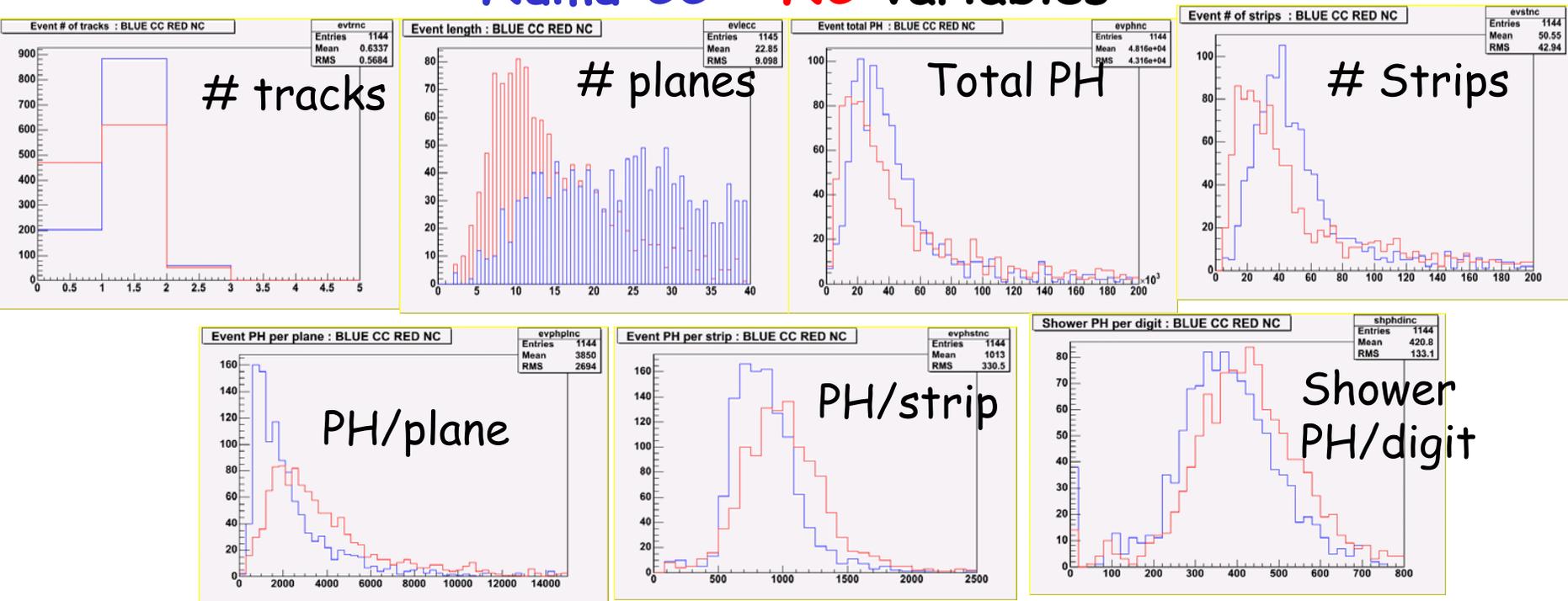
- These distributions are normalized to have ~ the same number of entries.

- The ratio of CC events that survive the 40 plane cut to NC events is ~ 2.3

- The actual MC ratio (with events with > 1 digits in each view) is ~ 1.8

- I expect the results to be slightly better than the previous time (mostly due to the smaller a priori ratios).

Sanity test with ANN

ANN distributions for signal (numu CC) and background (NC) events
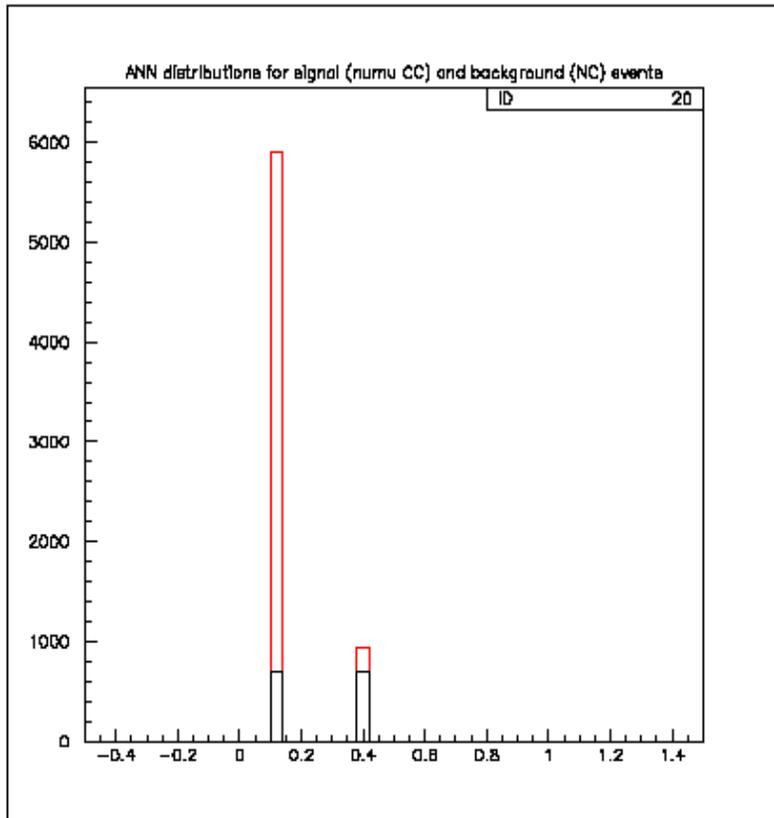
- Use just number of tracks to see in the worst possible case what I would get.

- What the ANN says is that in this case CC events (red) with zero tracks would be classified as NC and NC events (black) with one track as CC which is what we expect.

- The number of CC and NC events that you see are in the proportions expected.

- If the other variables don't add any significant information for the event classification then we expect to see ~ the same picture.

Final ANN Results

A priori probabilities 1:2.3   **A priori probabilities 1:4.5**

NEW RESULT   PREVIOUS RESULT

ANN distributions for signal (numu CC) and background (NC) events
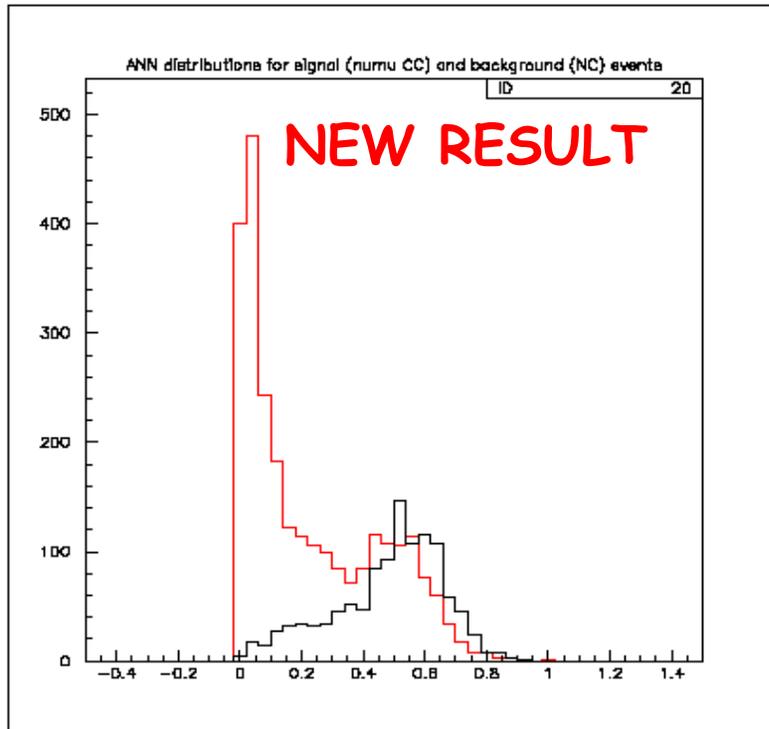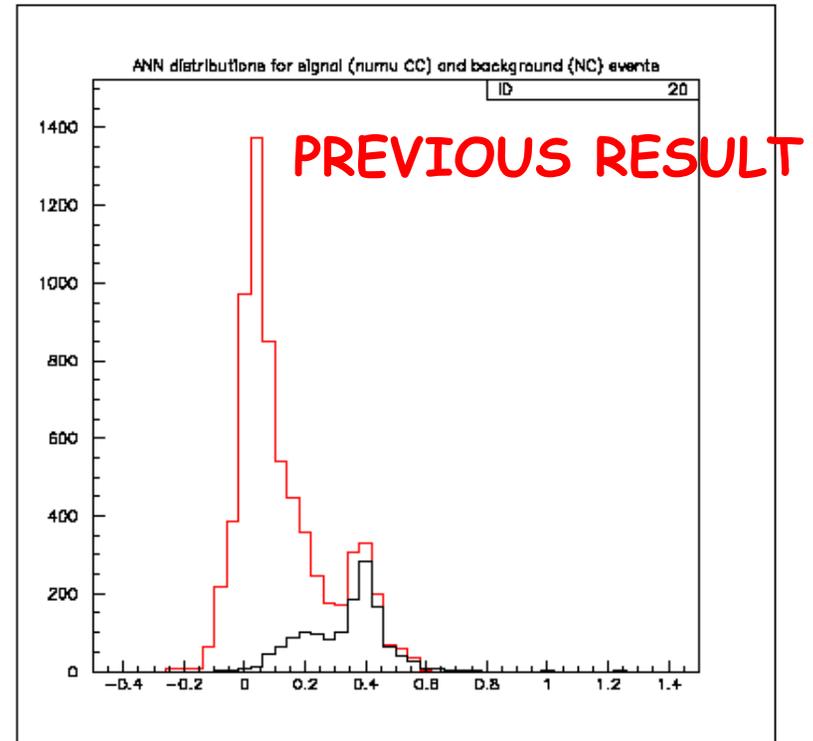
Event Probability

- **The ANN performs as expected :** Higher purity **for** CC selection **and** lower for NC **selection but the new results are certainly improved compared to the ones I showed in the previous meeting**

Final ANN Results con't
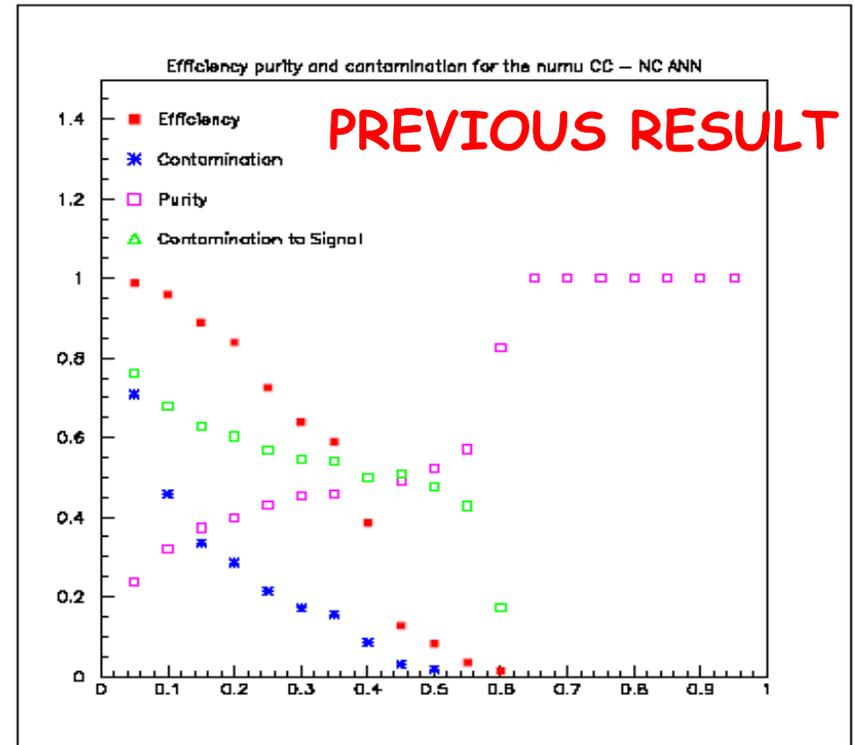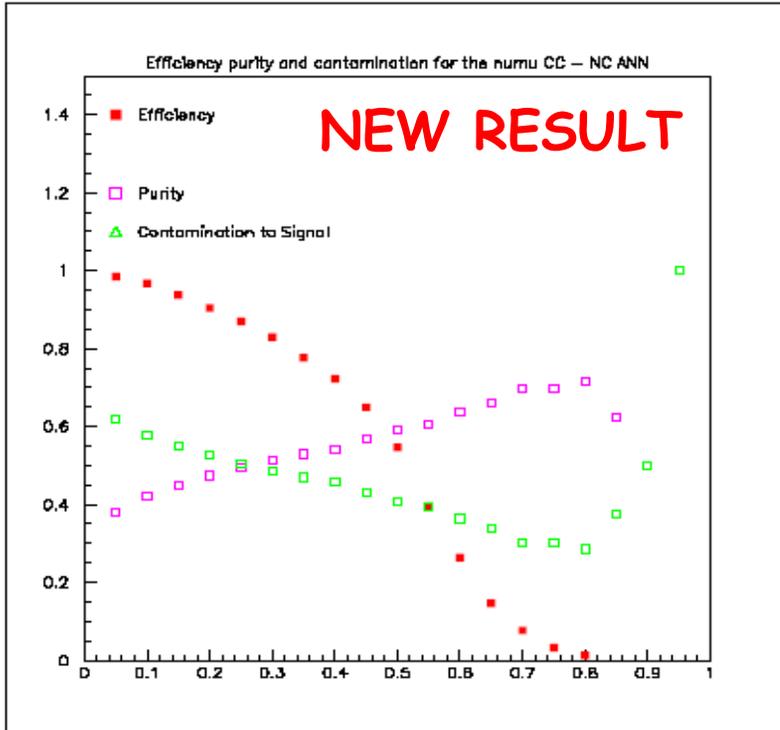
Efficiency (red) and purity (magenta) as a function of cut in the ANN output function for the signal (NC events)

- **If we set the cut @ 0.4 (i.e) we have an efficiency of 70% and a purity of 55 %…**

Final ANN Results con't

NEW RESULT

PREVIOUS RESULT
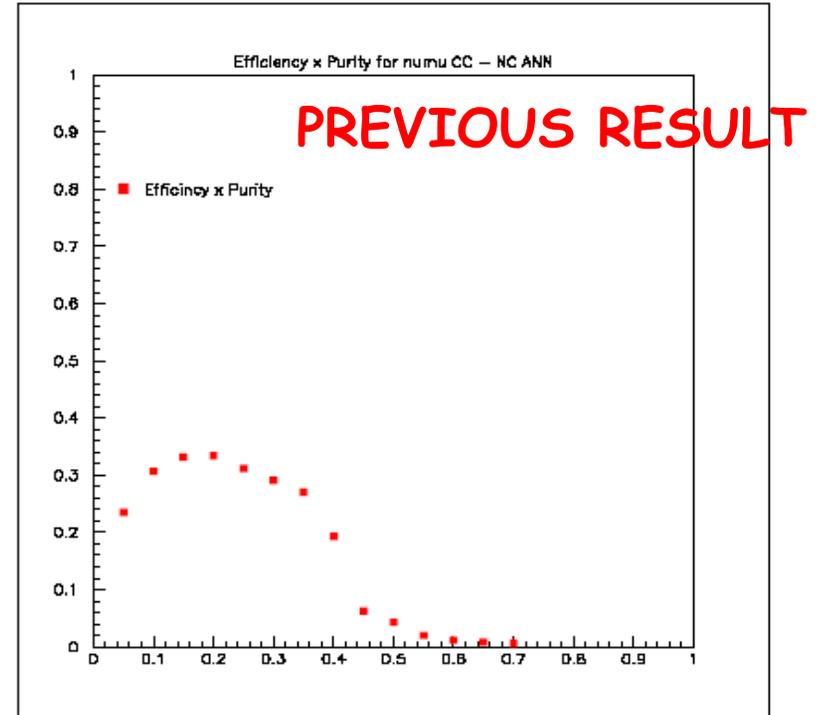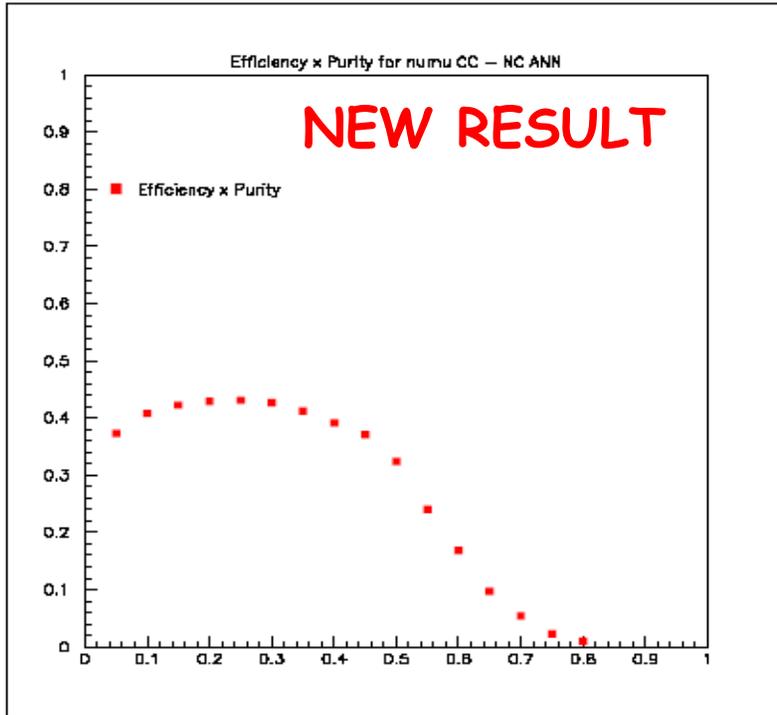
Efficiency x Purity

- **The new results are clearly improved.**

# Characteristics of correctly and wrongly classified CC events
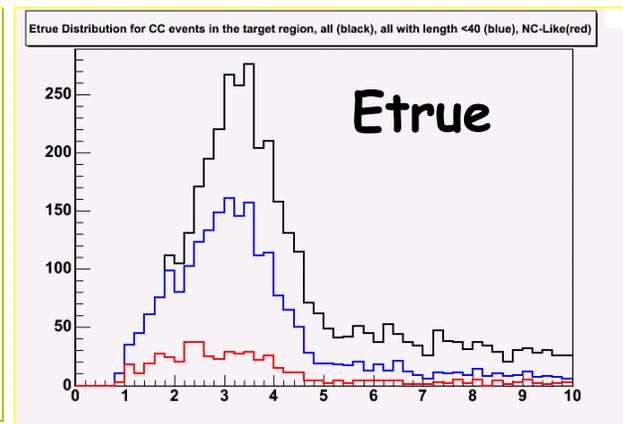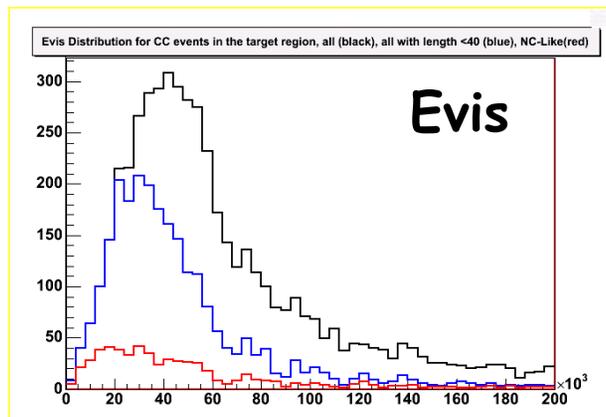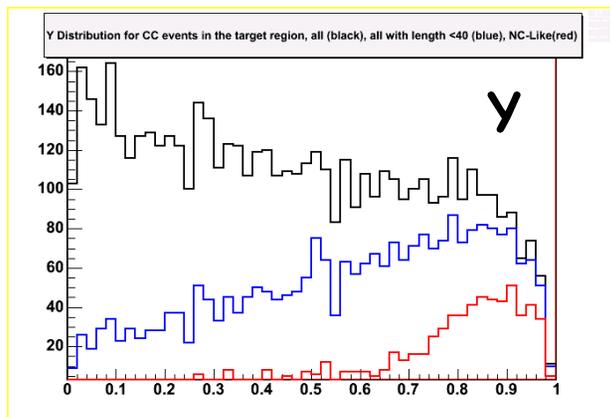
- **Black are all CC events in the central calorimeter region**
- **Blue are the ones that have total length < 40 planes**
- **Red are the ones that are mis-classified as NC-Like (P>0.4)**

- Most of the misclassified CC events are high Y events where the muon is not very energetic.

- These events will either have zero or quite sort tracks.

Characteristics of correctly and wrongly classified NC events
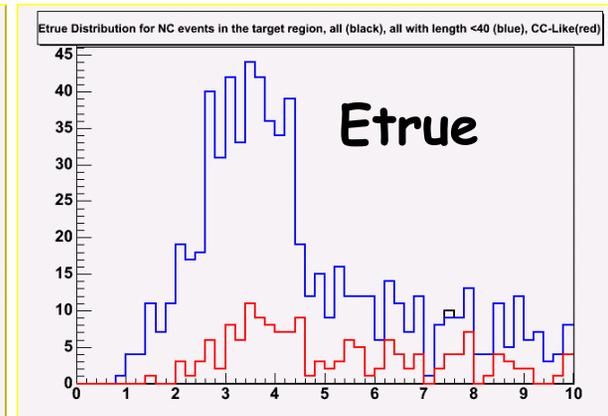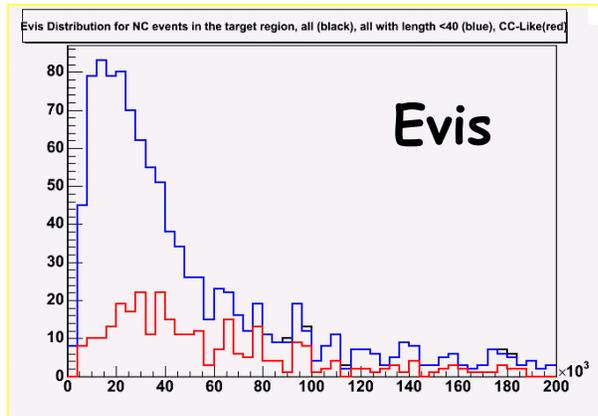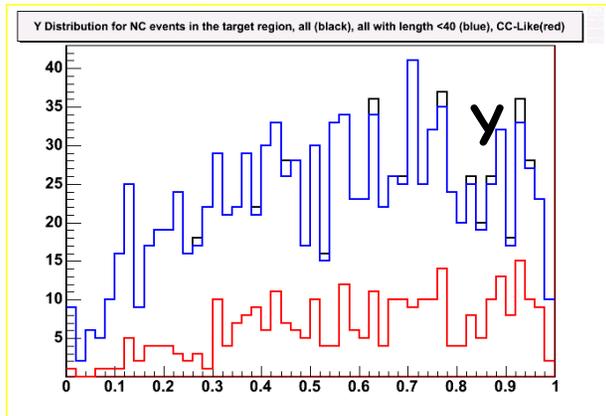
- **Black are all NC events in the central calorimeter region**
- **Blue are the ones that have total length < 40 planes**
- **Red are the ones that are mis-classified as CC-Like (P<0.4)**

- The misclassified NC events seem to be evenly distributed with a slight preference to higher Y events. Most probably regardless of whether these events have a track or not are very similar with the high Y CC events
- As a start I scanned CC events that are classified as NC-Like (20 of them as they appeared in one of the input files) and I intend to do the same thing with NC classified as CC-Like.

Slicing Characteristics of wrongly classified NC & CC events
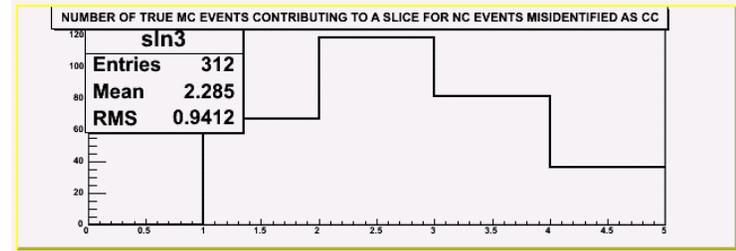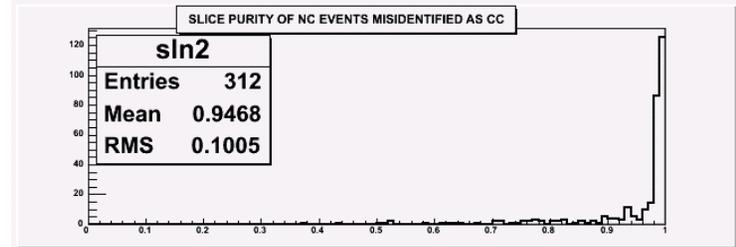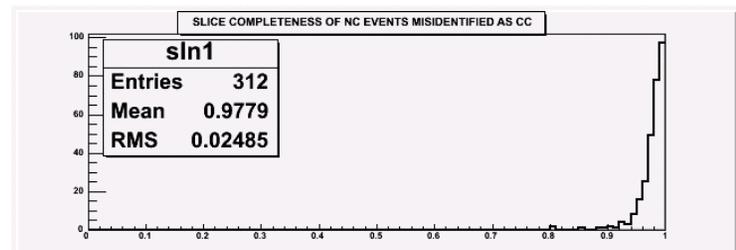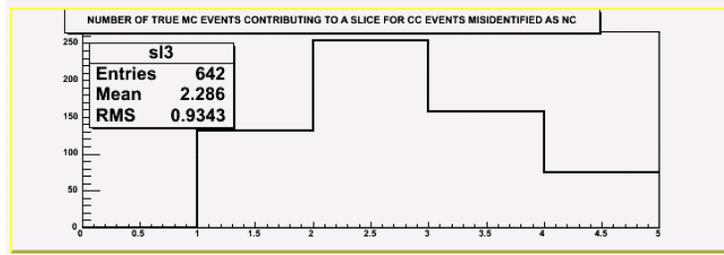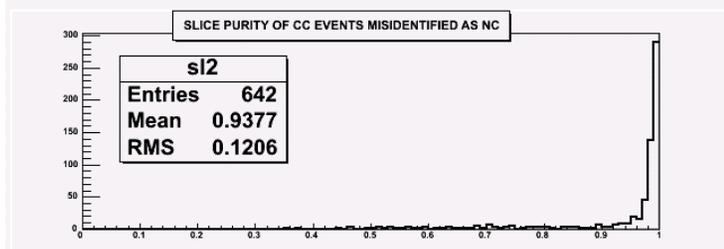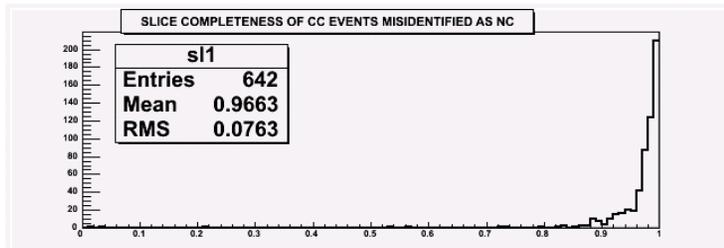
- Before I started scanning I wanted to see if there is anything obviously wrong with the slice characteristics for the misclassified events…
- I don't see any particular preference of the "wrong" classifications in lower completeness or purity slices

CC event classified as NC-like 1
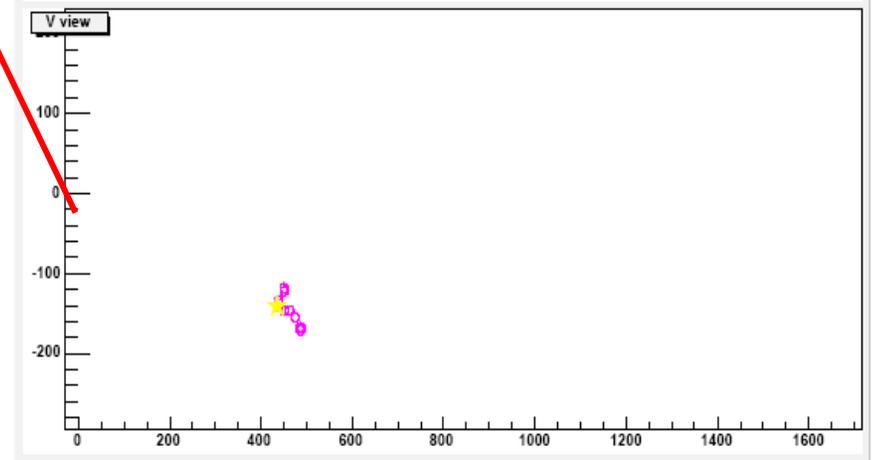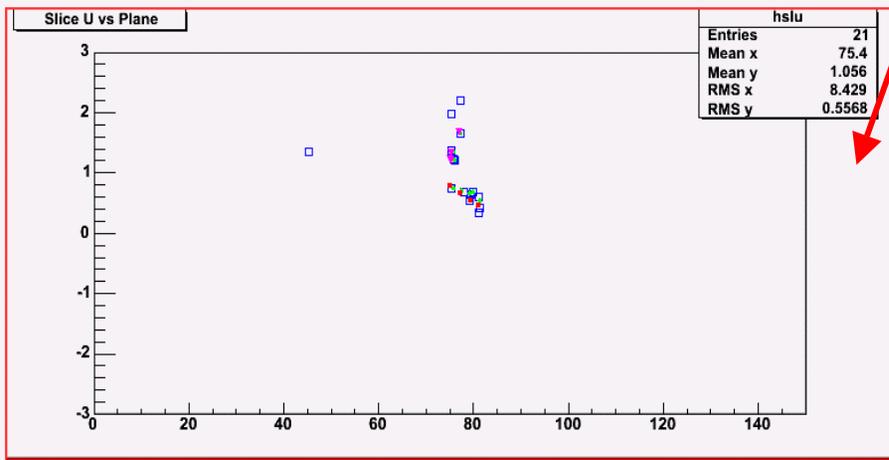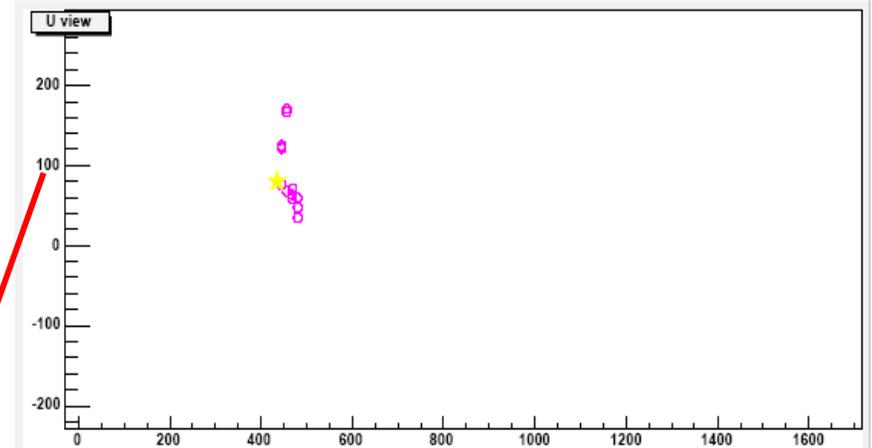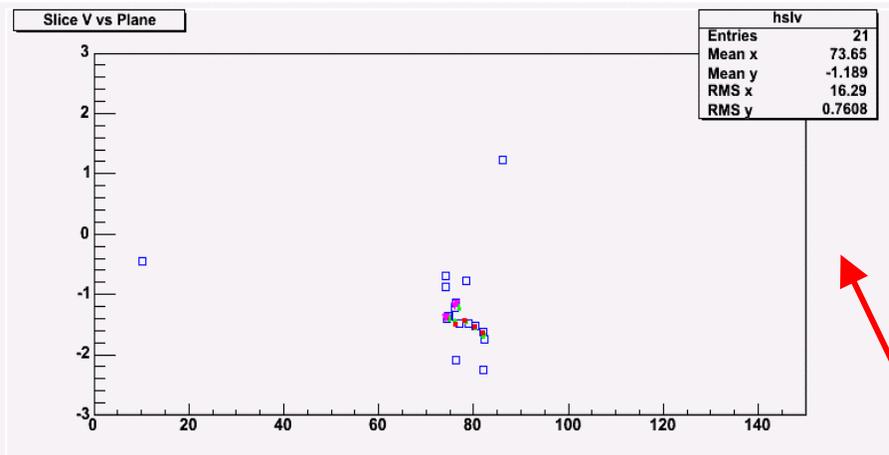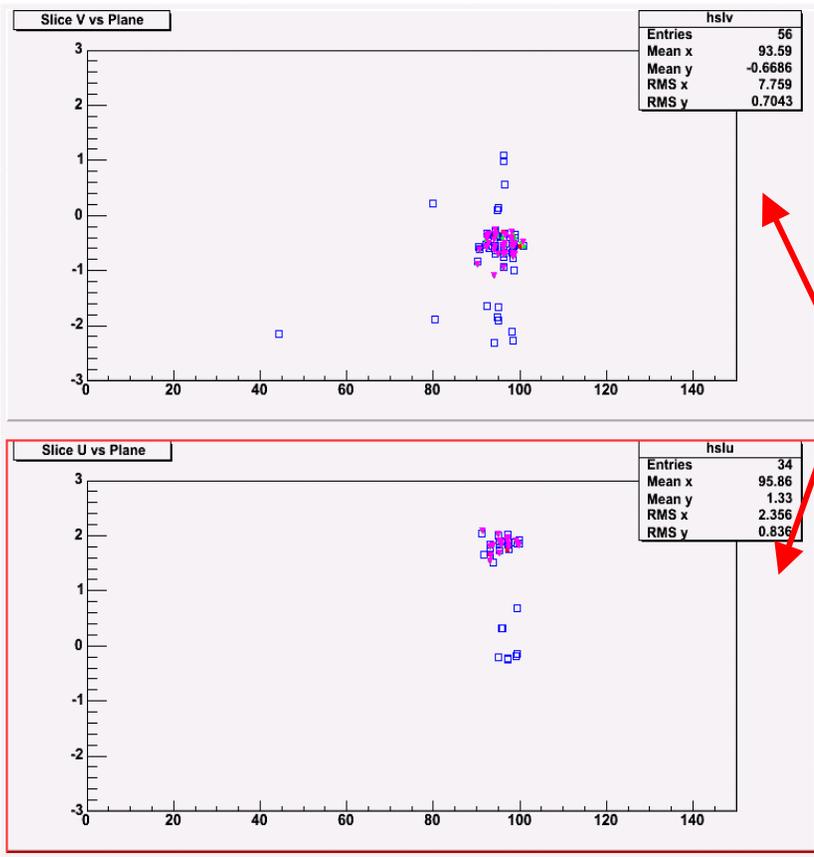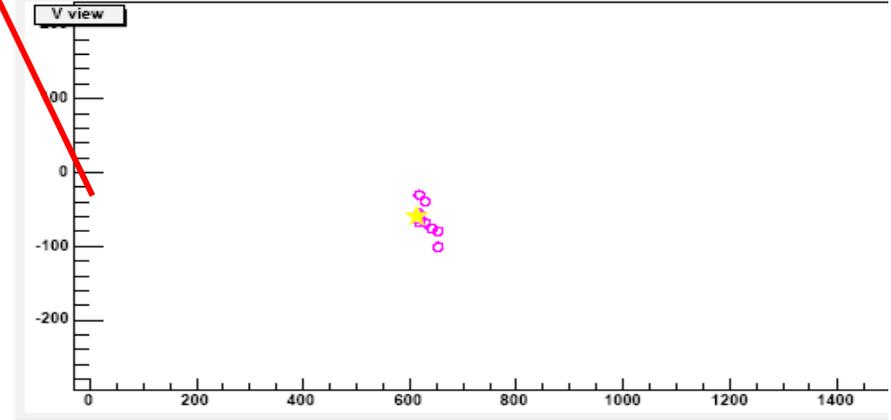

Blue squares : slice hits
Red triangles : track hits
Magenta triangles : shower hits

TRUTH DISPLAY
Etrue =1.7 Emu =0.7   Y=0.56

Sort track resembling NC events with a pion i.e
(I need to scan NC as well …more in next meeting)
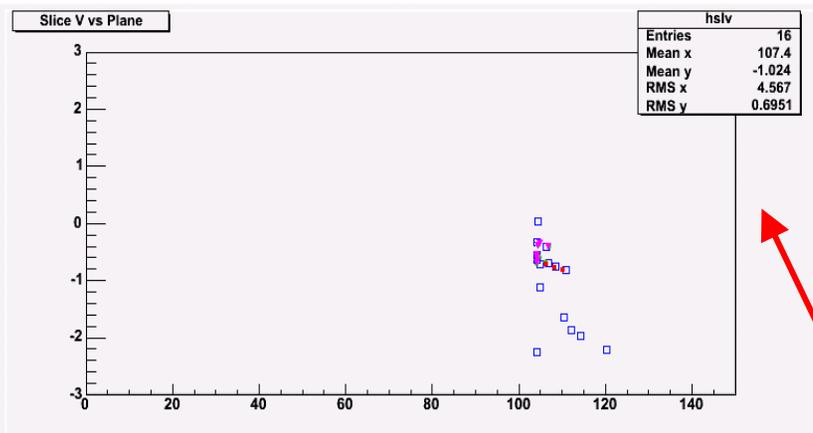
CC event classified as NC-like 2
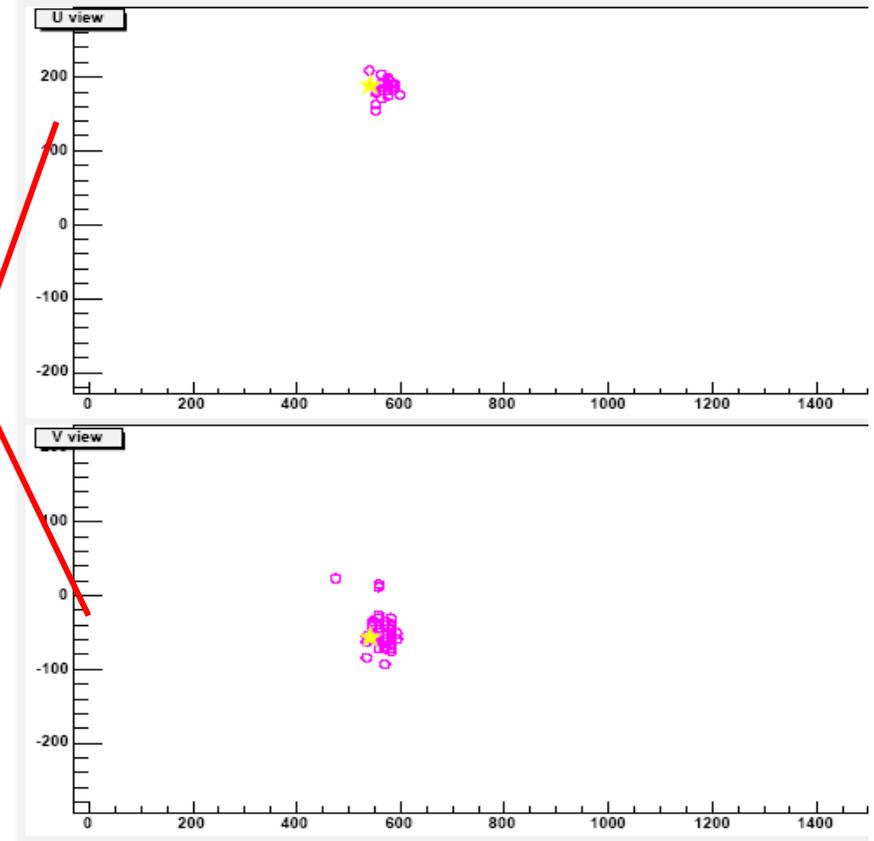
Blue squares   : slice hits
Red triangles   : track hits
Magenta triangles : shower hits

TRUTH DISPLAY
Etrue =4.5 Emu =0.7   Y=0.8

No obvious & reco track, resembling NC

CC event classified as NC-like 3
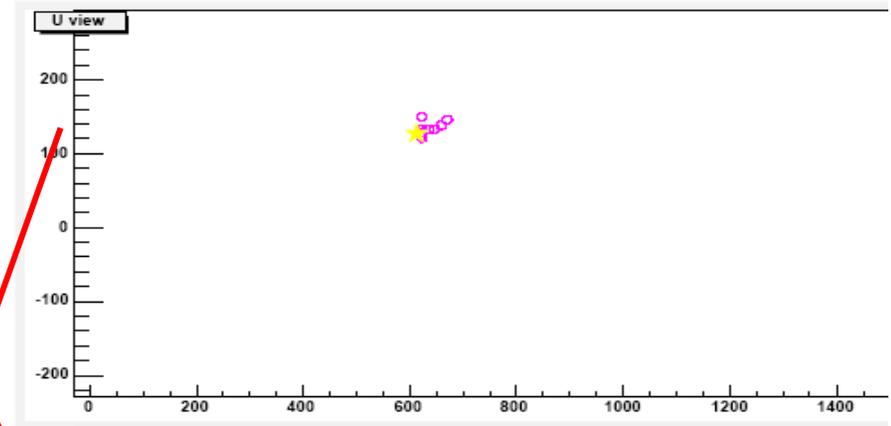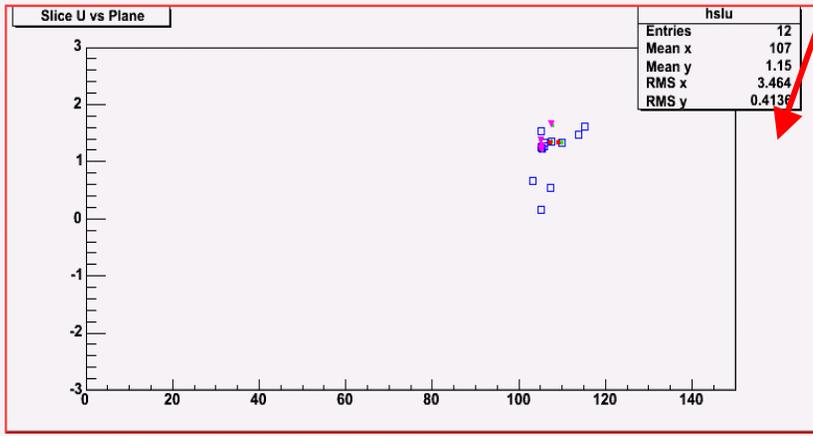
Blue squares   : slice hits
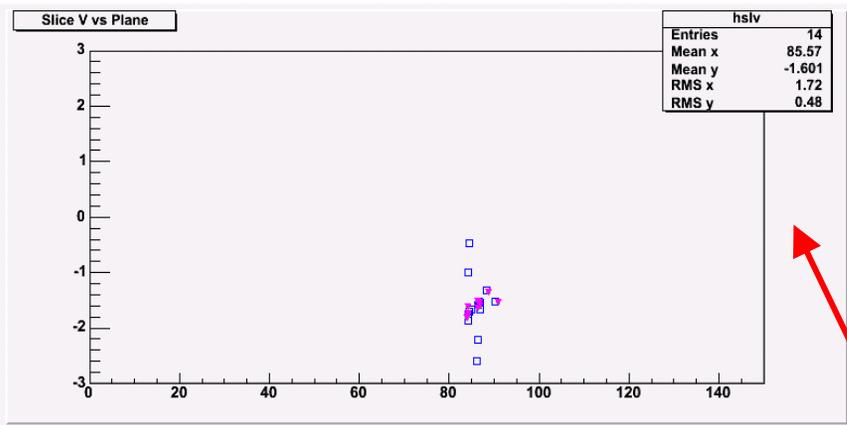**Red triangles**   : track hits
Magenta triangles : shower hits

Sort track, resembling NC

TRUTH DISPLAY
Etrue =1.3 Emu =0.4   Y=0.8

CC event classified as NC-like 4
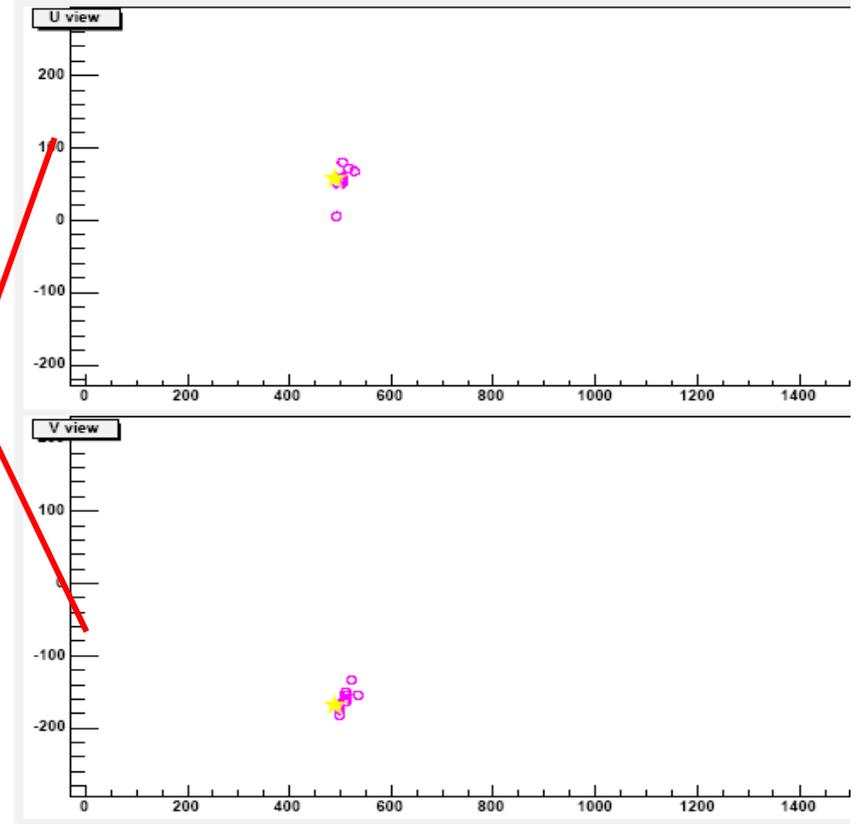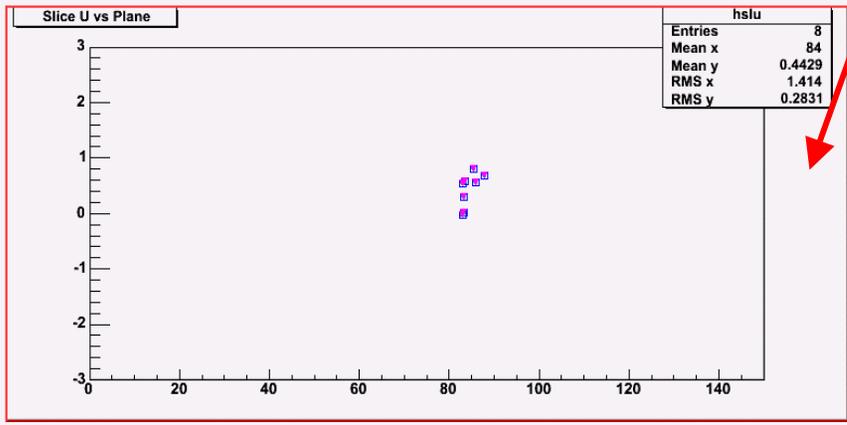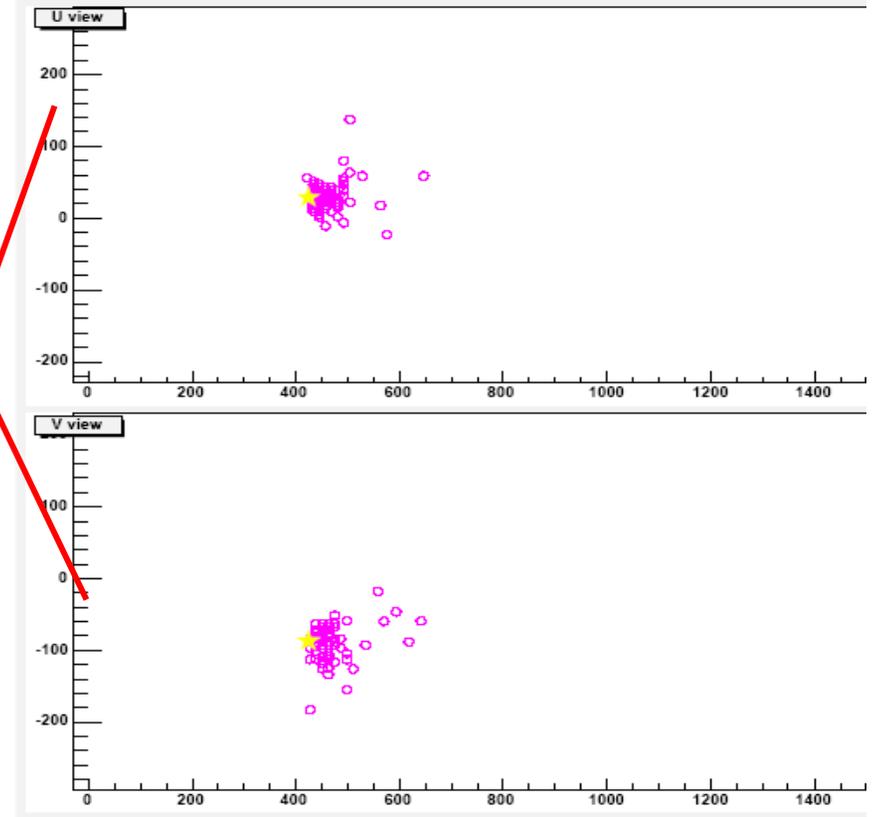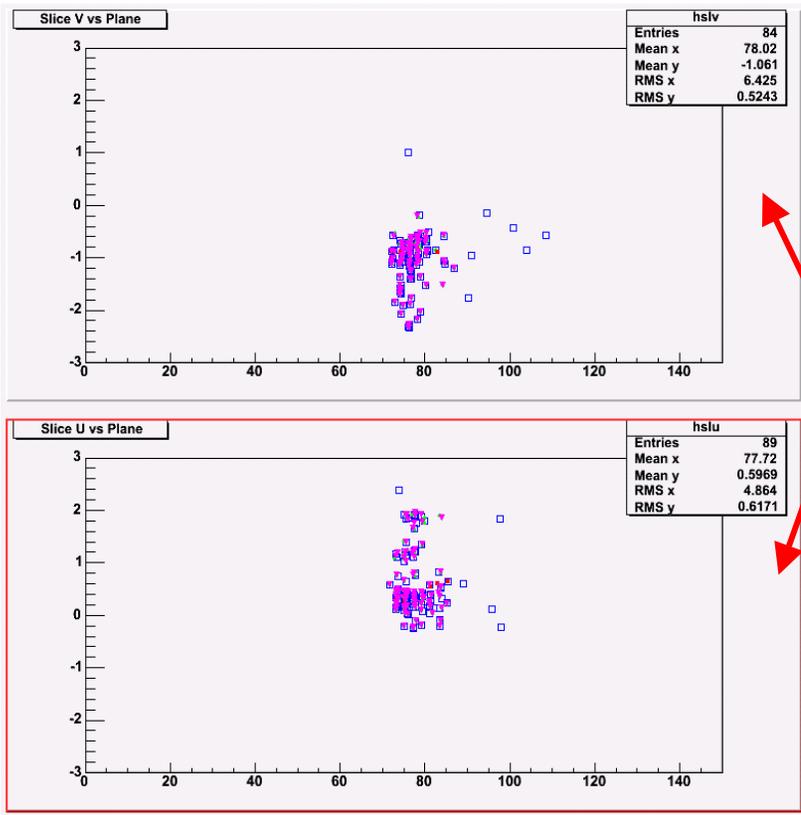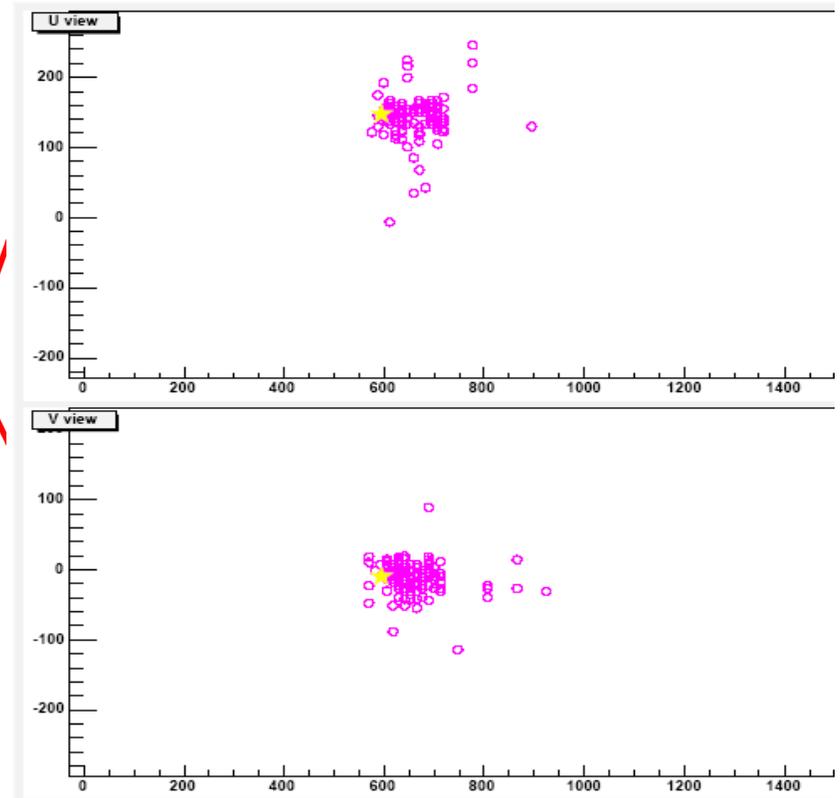

Blue squares   : **slice hits**
Red triangles   : **track hits**
Magenta triangles : shower hits

TRUTH DISPLAY
Etrue =2.1 Emu =0.3   Y=0.9

No obvious and reco rack, resembling NC

CC event classified as NC-like 5

Blue squares   : **slice hits**
Red triangles   : **track hits**
Magenta triangles : **shower hits**

TRUTH DISPLAY
Etrue =11.4 Emu =0.6   Y=0.95

No obvious and reco rack in the big shower , resembling NC

CC event classified as NC-like 6

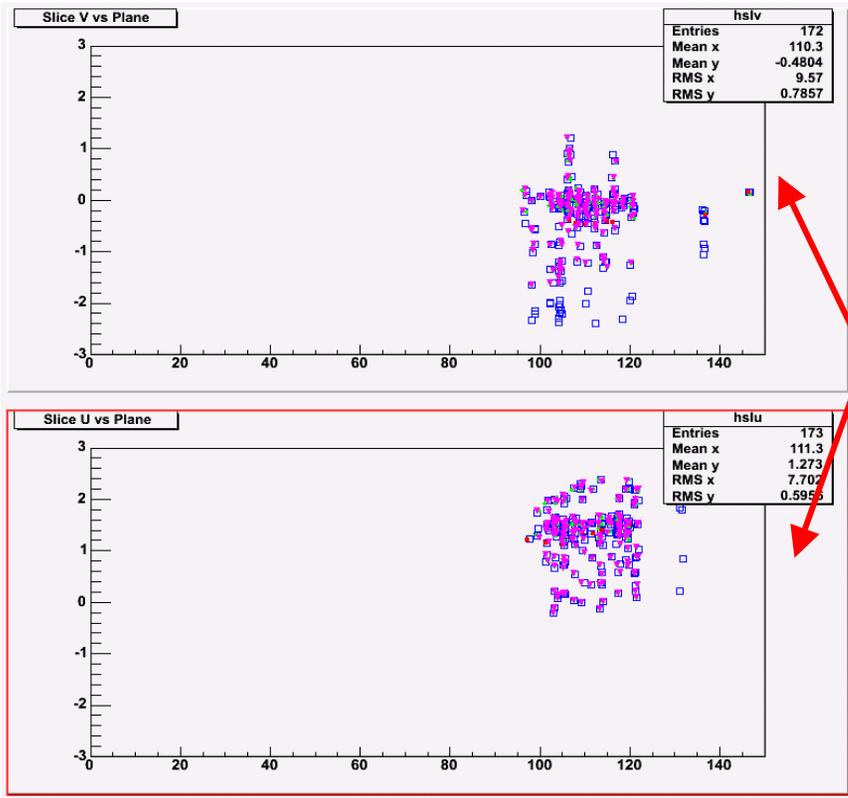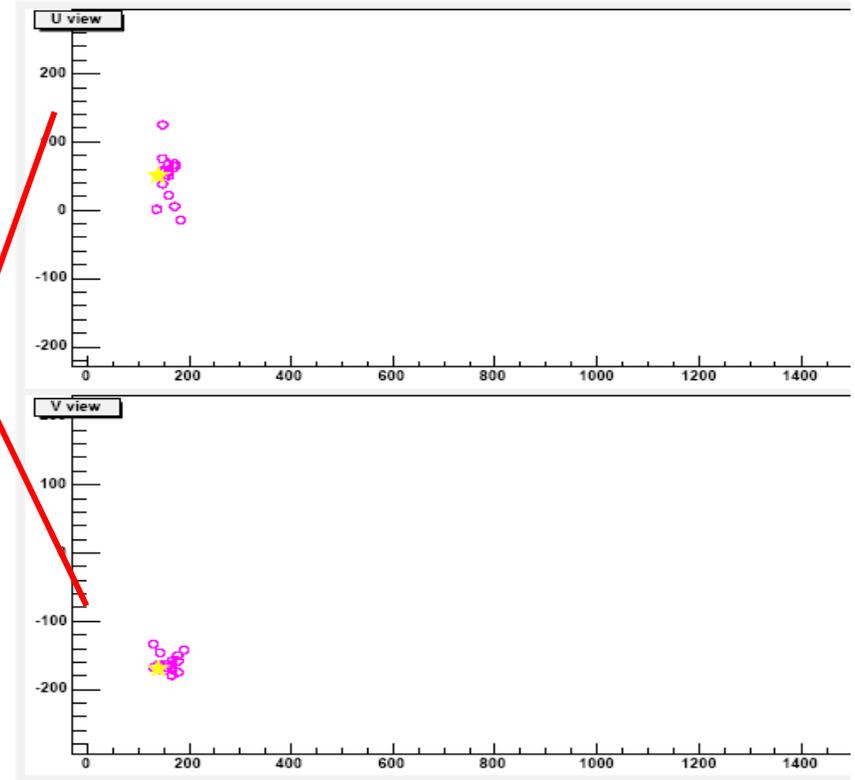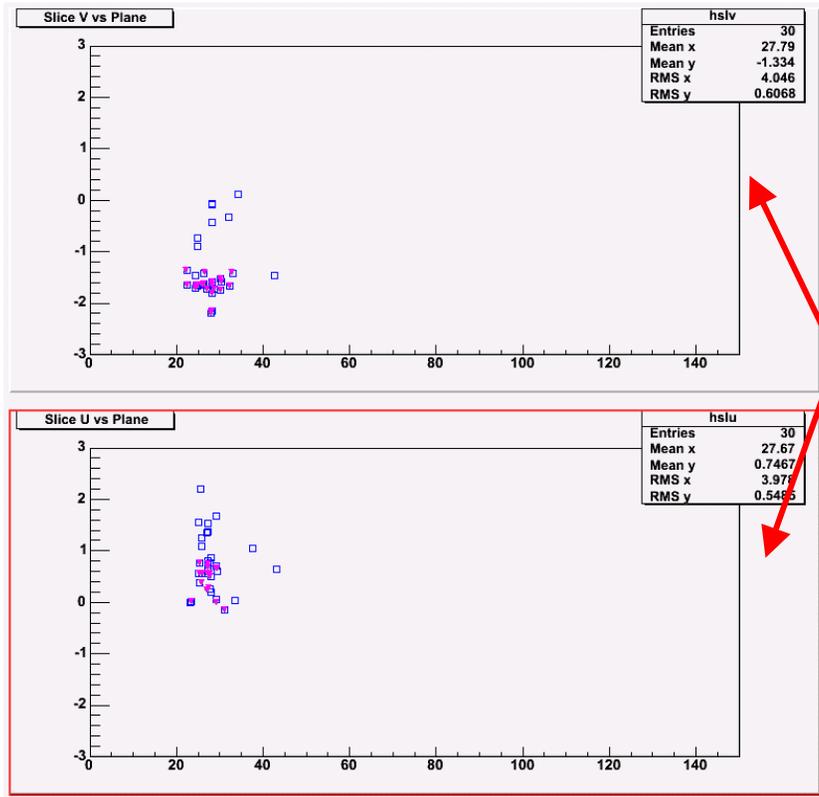**Blue squares**   : slice hits
Red triangles   : track hits
Magenta triangles : shower hits

TRUTH DISPLAY

Etrue =21   Emu =0.95   Y=0.95

No reco track, big shower , resembling NC
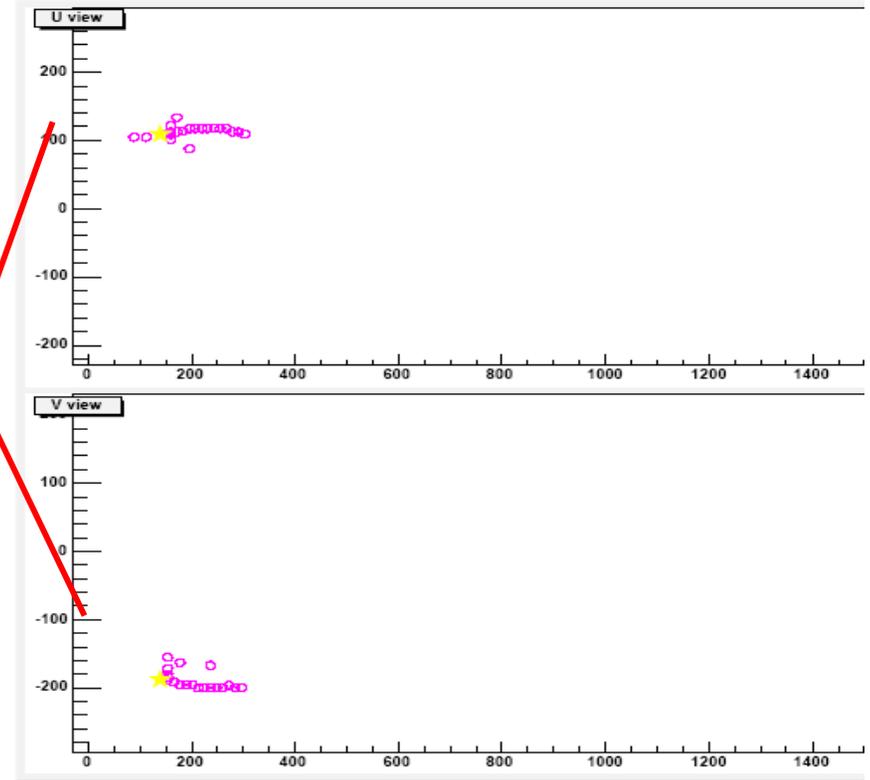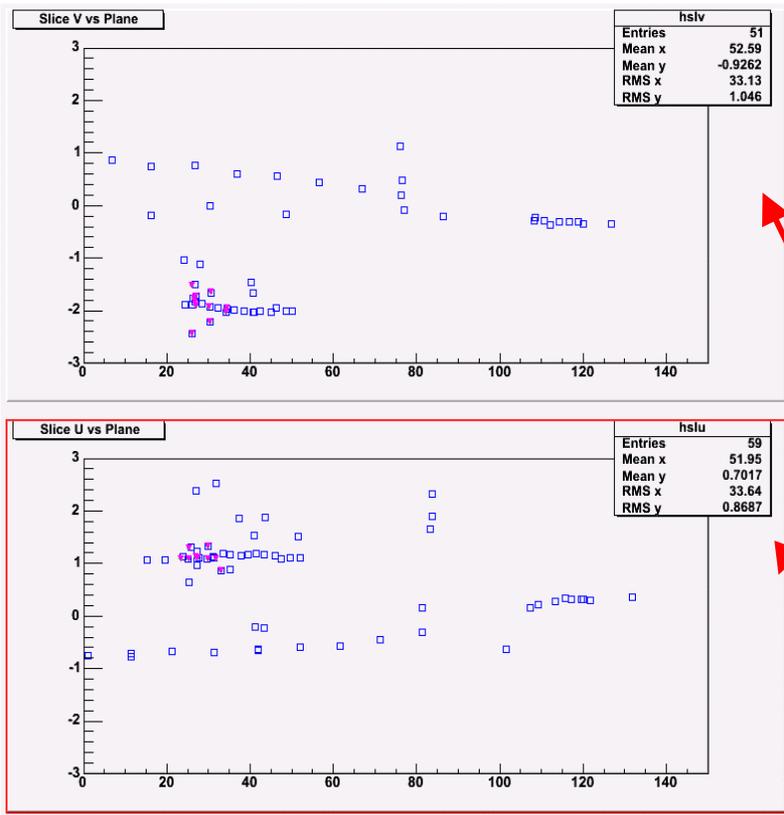
Blue squares : slice hits
Red triangles : track hits
Magenta triangles : shower hits

TRUTH DISPLAY
Etrue =2.3   Emu =0.6   Y=0.7

No reco track, maybe a small reconstructable track (that most probably would not help it to change category) resembling NC

Blue squares : slice hits
Red triangles : track hits
Magenta triangles : shower hits

TRUTH DISPLAY
Etrue =2.1   Emu =1.1   Y=0.5

No reco track but there should be one clear reconstructed track for this event (the other event of this slice which is a penetrating muon is reconstructed properly). Recoverable….

Summary / On going work

- The slicing and reco efficiency study that I did shows that both CC and NC slicing and reco efficiencies are quite high.

- The addition of visible energy and number of digits for all MC events (sliced/reco or not) in the NtpMC ntuple was very useful for me in order to understand clearly where all these NC events were going…

- This simple display I am using with the truth info (digits) in the detector is very helpful in order not to make guesstimates as to whether a slice/events/track is truly complete or not.

- The ANN for the NC/CC separation is performing better than before but still the NC purity is relatively low.

- The CC scanning so far revealed no major problems with the slicing or the reco. 1-2 /20 events has a clear muon that is missed (5%-10%) so there is a potential for recovering these events.

- I want to visually scan more events and in particular NC events misclassified as CC to search for possible pathologies and also new better variables (more clever).

- More results hopefully in the next meeting…