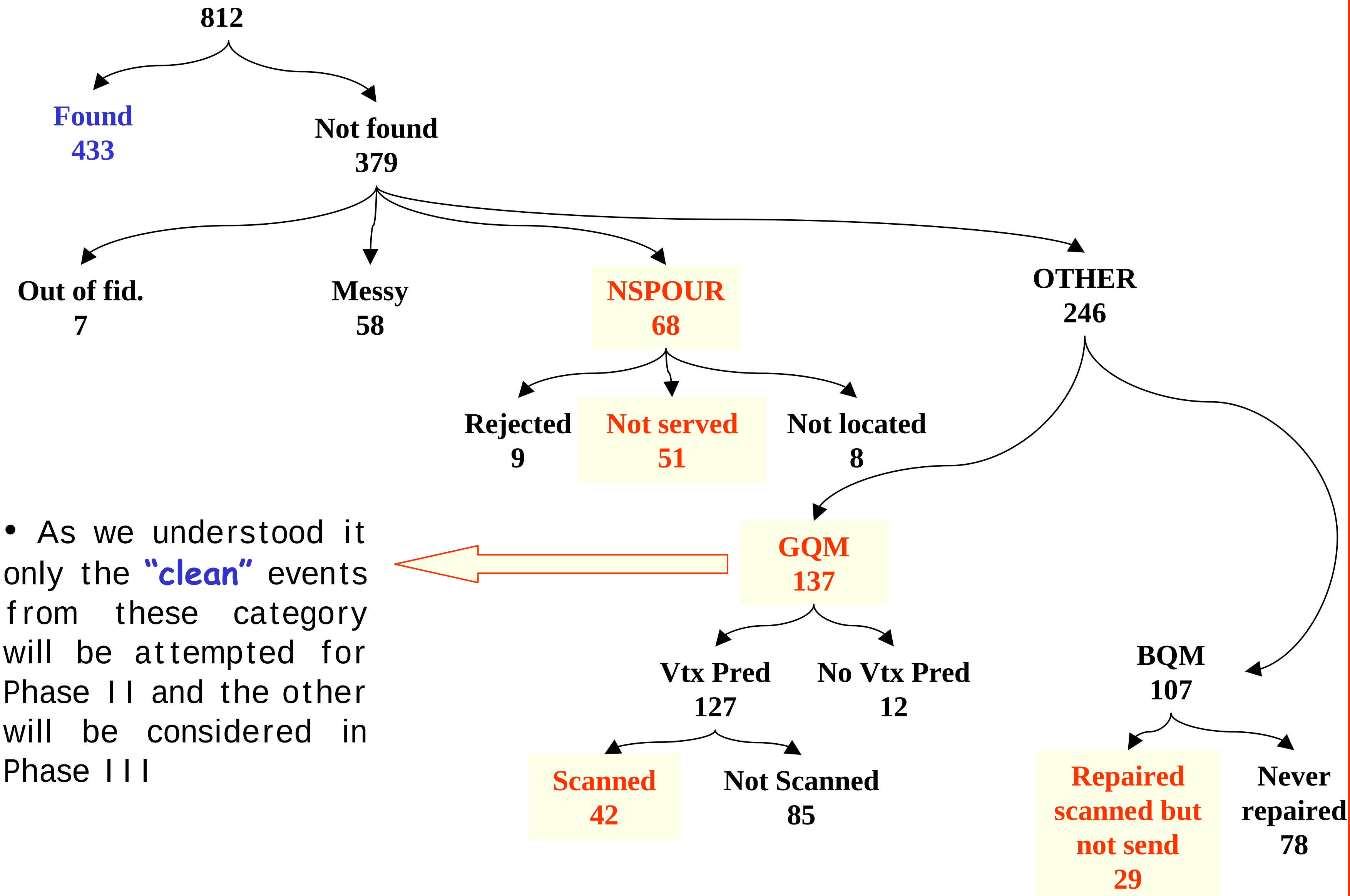
Progress Report

Outline

- Event Status Summary from the Nagoya meeting
  - Clean events with good quality m-files
- ANN on ν_τ CC - hadron scattering
  - Bayesian & ANN analysis revisited
  - ANN results on all Phase I and II kinks
  - Background estimation
- ANN on ν_τ CC - charm one prong kink decays
  - ANN results on ν_τ CC events
  - Background estimation
- Decay Search

Event Status from the Nagoya meeting



- As we understood it only the **"clean"** events from these category will be attempted for Phase II and the other will be considered in Phase III

"Clean" events from the GQM category

- As we understood from the Nagoya meeting for the **"clean"** events from the good quality m-files category, **event location** can be **attempted** so as to be included in the **"Phase II" data sample**.

- **"Clean" events** are the ones where there are **no large showers** developed in the **SFT** and therefore **vertex predictions** have **higher possibility** of being the **correct** ones.

- We have gone through all **139 events** with good quality m files and ended up with **21** that can be considered as **"clean"** out of which :

  - **Four are already located or are 2ry interactions**
  - **For another six we processed all available m-files and send emulsion vertex candidates to Nonaka**
  - **For another six we send new vertex predictions**
  - **The remaining five have only one large angle spectrometer track and therefore CS - NETSCANBACK could be attempted for the ones that seem "interesting"**

- We have send this information to Nonaka (and everyone else) and we are waiting for reply.

ANN Probability (review)

ANN analysis : Minimization of an Error (Cost) Function

$$\mathbf{E_N} = \frac{1}{\mathbf{N}}\sum_{\mathbf{N}}(\mathbf{f}(\mathbf{x_i},\mathbf{w}) - \mathbf{t_i})^2, \; \mathrm{w = weights, f(x,w) = ANN\ output, x = feature\ vector}$$

$$\mathrm{t = desired\ ANN\ output\ (1\ Signal\ \&\ 0\ background)}$$

$$\mathrm{E_N} = \frac{\mathbf{N_S}}{\mathbf{N}}\frac{1}{\mathbf{N_S}}\sum_{\mathbf{S}}(\mathbf{f}-1)^2 + \frac{\mathbf{N_B}}{\mathbf{N}}\frac{1}{\mathbf{N_B}}\sum_{\mathbf{B}}(\mathbf{f}-0)^2$$

$$\lim_{\mathbf{N},\mathbf{N_S},\mathbf{N_B}\to\infty}\mathbf{E_N} = \lim_{\mathbf{N},\mathbf{N_S},\mathbf{N_B}\to\infty}\left(\frac{\mathbf{N_S}}{\mathbf{N}}\frac{1}{\mathbf{N_S}}\sum_{\mathbf{S}}(\mathbf{f}-1)^2 + \frac{\mathbf{N_B}}{\mathbf{N}}\frac{1}{\mathbf{N_B}}\sum_{\mathbf{B}}(\mathbf{f}-0)^2\right)$$

$$\mathrm{but}\quad \lim_{\mathbf{N},\mathbf{N_S}\to\infty}\frac{\mathbf{N_S}}{\mathbf{N}} = \mathbf{P(S)}\ \&\ \lim_{\mathbf{N},\mathbf{N_B}\to\infty}\frac{\mathbf{N_B}}{\mathbf{N}} = \mathbf{P(B)}$$

$$\mathrm{and}\quad \lim_{\mathbf{N_S}\to\infty}\frac{1}{\mathbf{N_S}}(\mathbf{f}-\mathbf{s})^2 = \int(\mathbf{f}-\mathbf{s})^2\mathbf{P}(\mathbf{x}/\mathbf{S})\mathbf{dx}\ \ldots..$$

$$\ldots\ldots \mathrm{f = P(S/x)}$$

The ANN output is the Bayes a posteriori probability & in the proof no special assumption has been made on the a priori P(S) and P(B) probabilities (absolute normalization)…..TRUE BUT THEIR VALUES DO MATTER ………(They should be what nature gave us)

ANN probability (review)

- **Bayesian a posteriori probability :**

$$\mathbf{P(S/x)} = \frac{\mathbf{P(x/S)*P(S)}}{\mathbf{(P(S)*P(x/S)+P(B)*P(x/B)}}$$

$\mathbf{P(S)} = \mathbf{a}$ priori signal probabilty $\quad$ P(x/S) = Signal probabilty density function

$\mathbf{P(B)} = \mathbf{a}$ priori background probabilty $\quad$ P(x/B) = Background probabilty density function

- ANN output : P(S/x)
- ANN training examples : P(x/S) & P(x/B)
- ANN number of Signal Training Examples P(S)
- ANN number of Background Training Examples P(B)

The MLP (ann) analysis and the Maximum Likelihood Method (Bayes Classifier) are equivalent.
(c_{11} c_{22} = cost for making the correct decision &
c_{12} c_{21} = cost for making the wrong decision)

$$\Lambda(\mathbf{x}) = \frac{\mathbf{P(x/S)}}{\mathbf{P(x/B)}} \ \& \ \xi = \frac{\mathbf{P(B)(c_{12} - c_{11})}}{\mathbf{P(S)(c_{21} - c_{22})}}$$

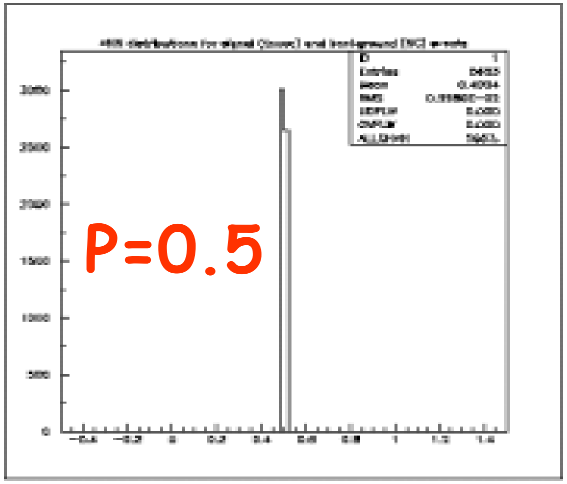
$$\textbf{if} \ \ \mathbf{c}_{11} = \mathbf{c}_{22} = 0 \ \& \ \mathbf{c}_{12} = \mathbf{c}_{21} \Rightarrow$$

$$\Lambda(\mathbf{x}) > \xi \Leftrightarrow \frac{\mathbf{P(x/S)}}{\mathbf{P(x/B)}} > \frac{\mathbf{P(B)}}{\mathbf{P(S)}} \Leftrightarrow \mathbf{P(x/S)*P(S)} > \mathbf{P(x/B)*P(B)} \Leftrightarrow$$

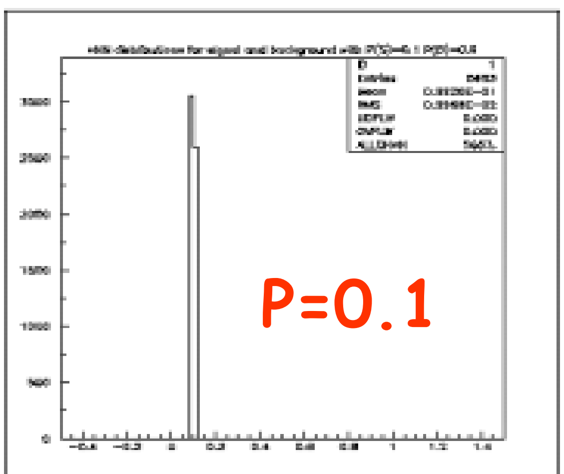
$$\Leftrightarrow \frac{\mathbf{P(x/S)*P(S)}}{\mathbf{P(x)}} > \frac{\mathbf{P(x/B)*P(B)}}{\mathbf{P(x)}} \Leftrightarrow \mathbf{P(S/x)} > \mathbf{P(B/x)} \Leftrightarrow$$

$$\Leftrightarrow \mathbf{P(S/x)} > (1 - \mathbf{P(S/x)}) \Leftrightarrow \mathbf{P(S/x)} > 0.5$$

ANN Probability cont.

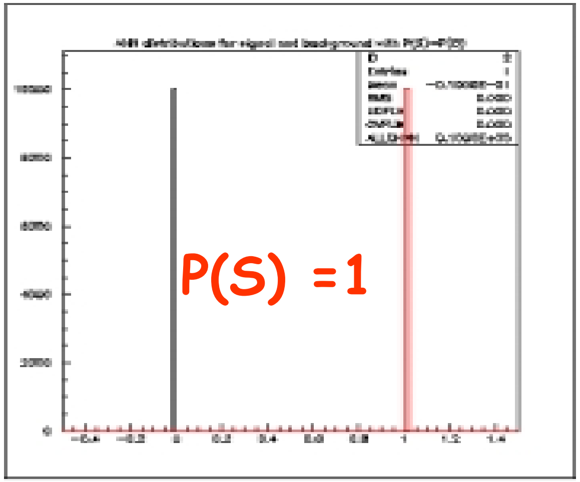


ANN output

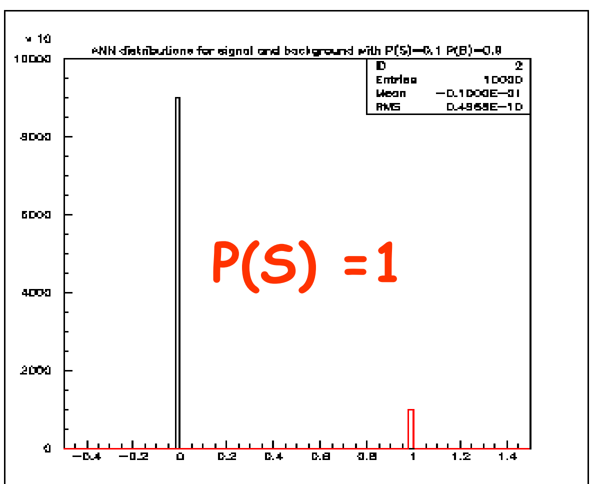


- **Worse hypothetical case :** One variable characterizing the populations, which is identical for S and B.
- P(S)=0.1 & P(B)=0.9
- If we train with equal numbers for signal and background the ANN will wrongly compute P(S)=0.5.
- If we train with the correct ratio for signal and background the ANN will correctly compute P(S)=0.1 which is exactly what Bayes a posteriori probability would give also.

ANN Probability cont.



ANN output



- **Best hypothetical case :** One variable characterizing the populations, which is completely separated (different) for S and B.
- P(S)=0.1 & P(B)=0.9
- If we train with equal numbers for signal and background the ANN will compute P(S)=1.
- If we train with the correct ratio for signal and background the ANN will again compute P(S)=1.
- In this case it does not matter if we use the correct a priori probabilities or not.

ANN Probability (final...)

- The MLP output approximates the Bayesian a posteriori probability and the a priori class probabilities P(S) and P(B) should be considered correctly.

- The more similar the characteristics of the populations are, the more important the a priori probabilities are, in calculation of the final a posteriori probability by the MLP.

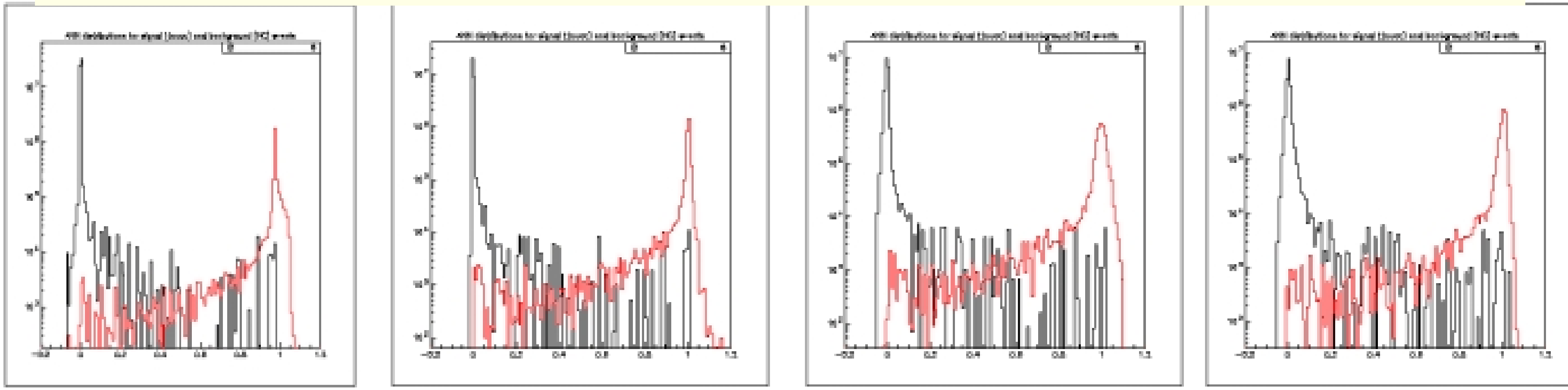
- In addition the more close to the boundary surface (between the two populations) an event is , the more sensitive it's a posteriori probability is to changes in the a priori probabilities.

ANN ν_τ CC - hadron scattering

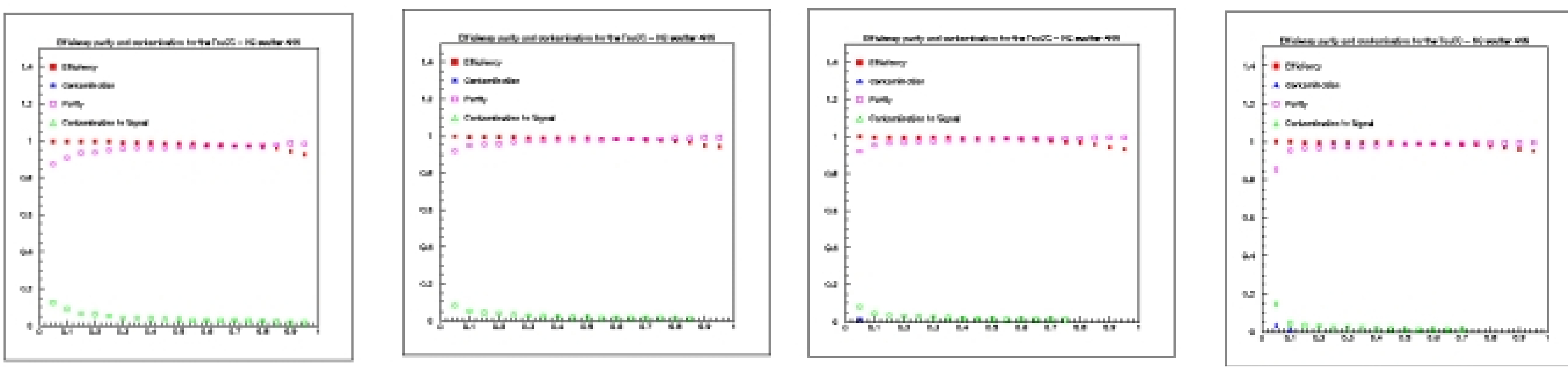
- Since we do not know the precise expected ratio for the **hadron scattering** events with respect to the ν_τ CC interactions we have re-constructed the ANN with the following ratios (which are close to the actual) :
  - **(ν_τ CC / hadron scattering) :** 0.5/10 - 1.0/10 - 1.5/10 - 2.0/10
- We have smeared the daughter momentum by 10%, 30% , 50%, 70% , and since the actual resolution is of the order of 30% we present the results for this case (4 ANNs corresponding to the four different ratios) .
- We have processed all phase I and phase II kinks with no leptons identified from the primary vertex.

ANN ν_τ CC - hadron scattering cont.

Output ANN function (in log scale) (momentum smeared by 30%)



Efficiency, Purity and contamination



- The results are very similar with ones obtained with P(S)=P(B) since the characteristics of the two populations are very different
- But the individual event probabilities for the events that their variables have values close to the "boundary surface" between the two populations, change.

MC Distributions of ν_τ CC & hadron scattering events

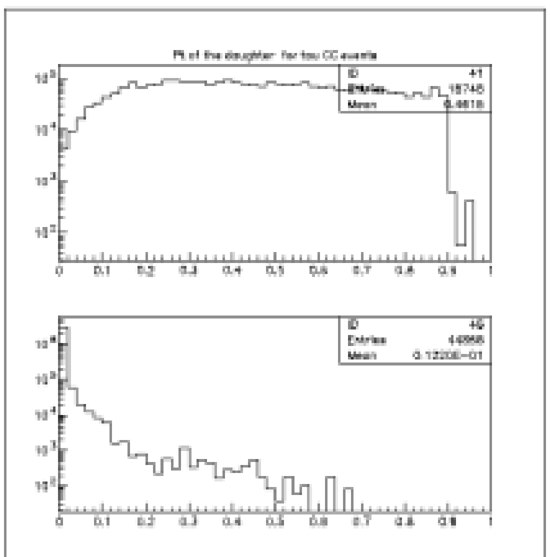
Daughter P_T

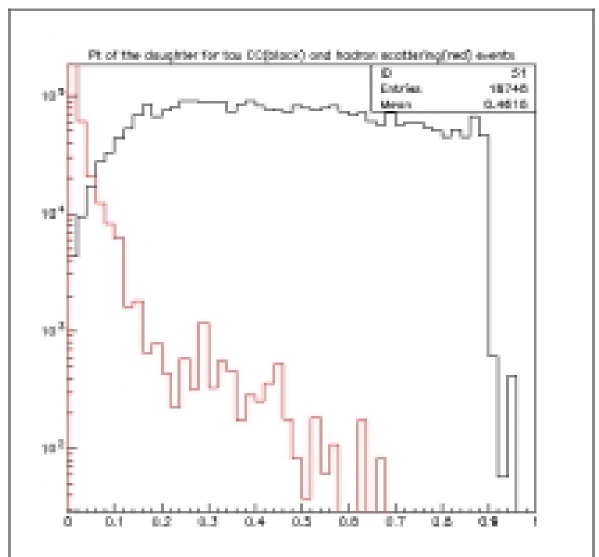
ν_τ CC (top)

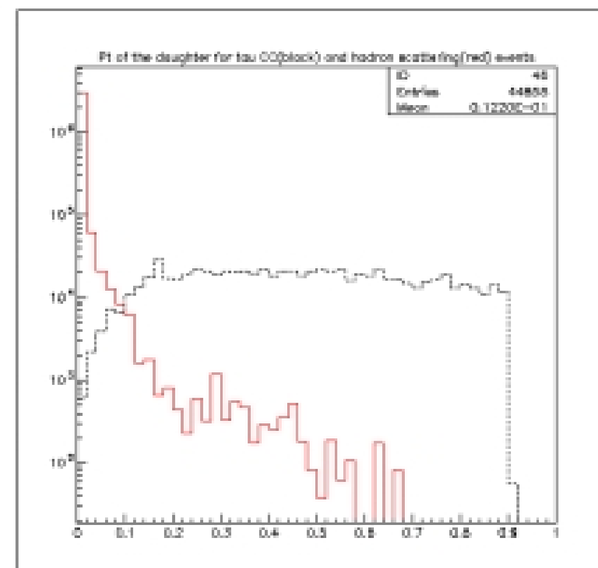
Hadron Scattering (bottom)

S/B = 1

S/B ~ 1/10







MC Distributions of ν_τ CC & hadron scattering events

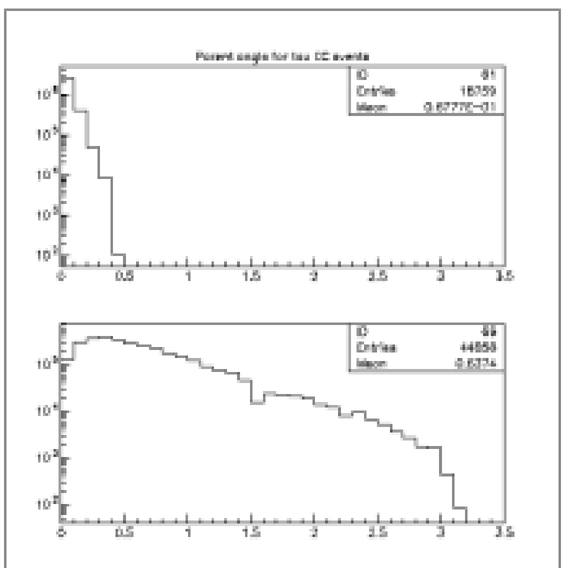
Parent Angle

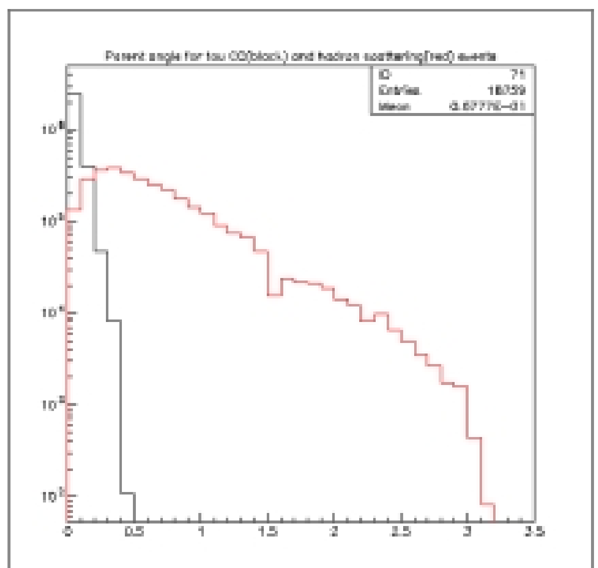
ν_τ *CC* (top)

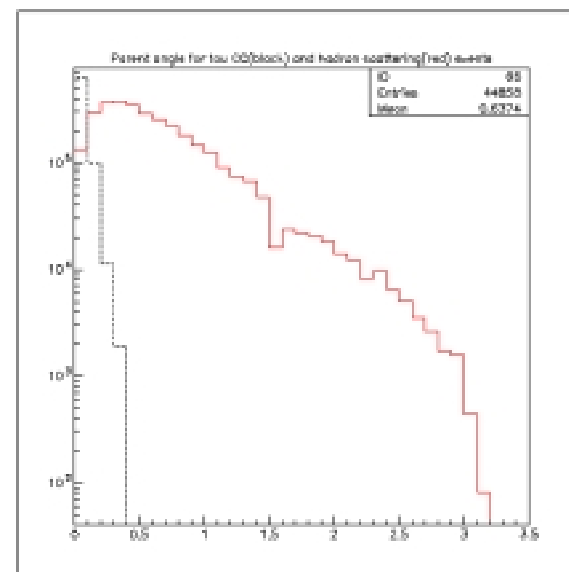
Hadron Scattering (bottom)

S/B = 1

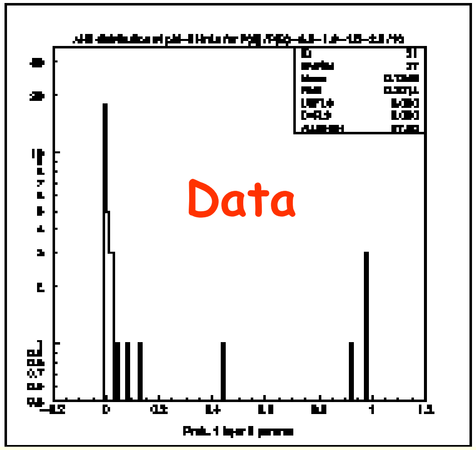
S/B ~ 1/10



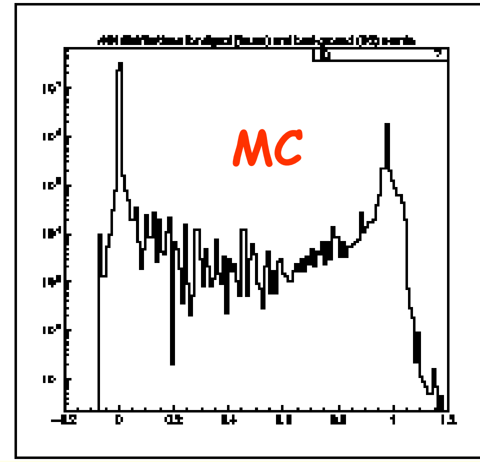




ANN ν_τ CC - hadron scattering results on Phase I & II kinks



Δp/p=30%



<u>EVENTS THAT EXCEED THE 0.5 CUT IN THE ANN OUTPUT FUNCTION (P(S) > P(B))</u>

RUN	EVENT	P_d	θ_d	P_T	L_d	θ_p	$\Delta\varphi$	Probabilities for 0.5 -1.0 1.5 2.0 /10				
3263	25102	1.900	0.1300	0.247	1890.1	0.1772	0.176	0.084	0.136	0.144	0.407	0.336
3024	30175	2.900	0.0936	0.271	4504.8	0.0279	1.027	0.929	0.971	0.956	0.988	0.984
3039	1910	4.600	0.0895	0.412	276.5	0.0653	2.684	0.979	1.000	1.000	1.000	1.000
3333	17665	21.400	0.0130	0.278	564.6	0.0154	2.806	0.976	1.000	1.000	0.999	0.988
3193	1361	20.000	0.0187	0.374	1863.6	0.0838	2.341	0.977	1.000	1.000	1.000	1.000

ANN ν_τ CC - hadron scattering results on Phase I & II kinks (background estimation)

EVENTS THAT EXCEED THE 0.5 CUT IN THE ANN OUTPUT FUNCTION (P(S) > P(B))

RUN	EVENT	P_d	θ_d	P_T	L_d	θ_p	$\Delta\varphi$	Probabilities for 0.5	-1.0	1.5	2.0	/10
3263	25102	1.900	0.1300	0.247	1890.1	0.1772	0.176	0.084	0.136	0.144	0.407	0.336
3024	30175	2.900	0.0936	0.271	4504.8	0.0279	1.027	0.929	0.971	0.956	0.988	0.984
3039	1910	4.600	0.0895	0.412	276.5	0.0653	2.684	0.979	1.000	1.000	1.000	1.000
3333	17665	21.400	0.0130	0.278	564.6	0.0154	2.806	0.976	1.000	1.000	0.999	0.988
3193	1361	20.000	0.0187	0.374	1863.6	0.0838	2.341	0.977	1.000	1.000	1.000	1.000

Considering as "Signal" events (ν_τ CC) the ones with probabilities P > 0.5 we can compute the background to these events by adding 1-P. Therefore :

	0.5/10	1.0/10	1.5/10	2.0/10
Bkg =	0.139	0.029	0.013	0.028

ANN ν_τ CC - charm one prong kink decays (were lepton is missed)

- Since we roughly know the ratio for the **charm one prong kink decays where the lepton is missed** events with respect to the **ν_τ CC** interactions (from Phase I background analysis) we have re-constructed the ANN with the following ratio
  - **ν_τ CC / Charm one prong kink decays with missed lepton : 4.1/0.21**
- We have used the actual daughter momentum and also the one smeared by 30% (since this variables is almost identical for both populations, it does not make much difference)
- We have processed phase I and phase II **ν_τ CC -like (3 tau neutrino events from phase I and one additional (under study) from phase II)** events with no leptons identified from the primary vertex in order to mainly estimate the background.

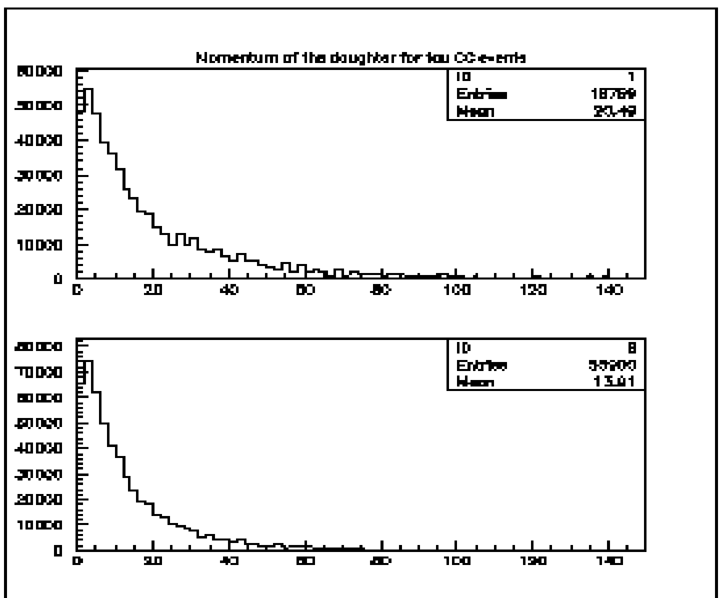
MC Distributions of ν_τ CC & Charm events

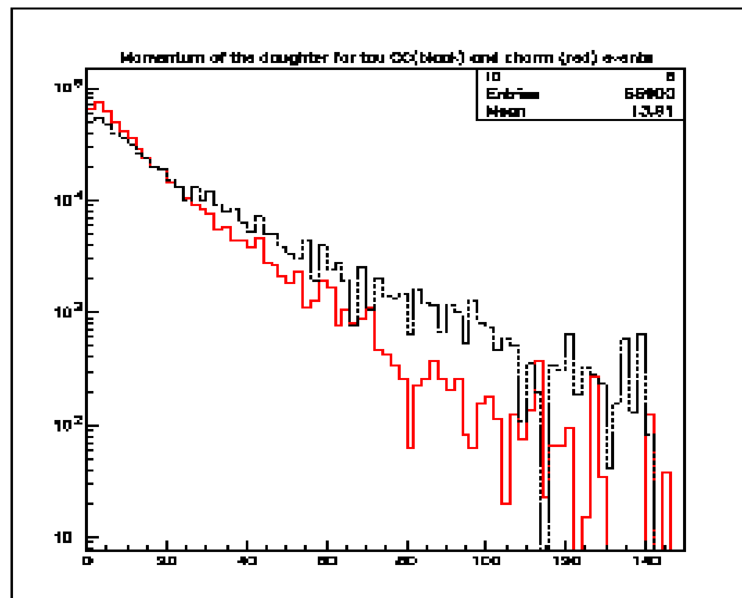
Daughter Momentum

ν_τ *CC* (top)

Hadron Scattering
(bottom)

Equal Numbers



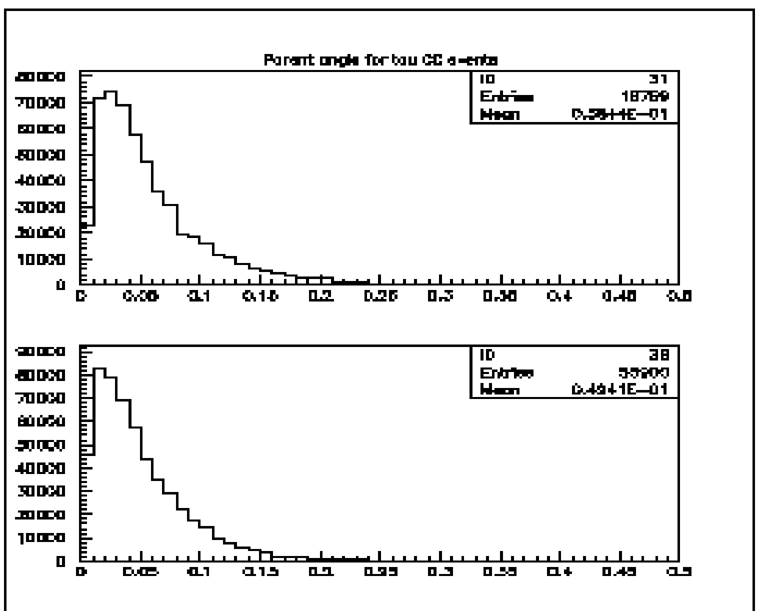


MC Distributions of ν_τ CC & Charm events

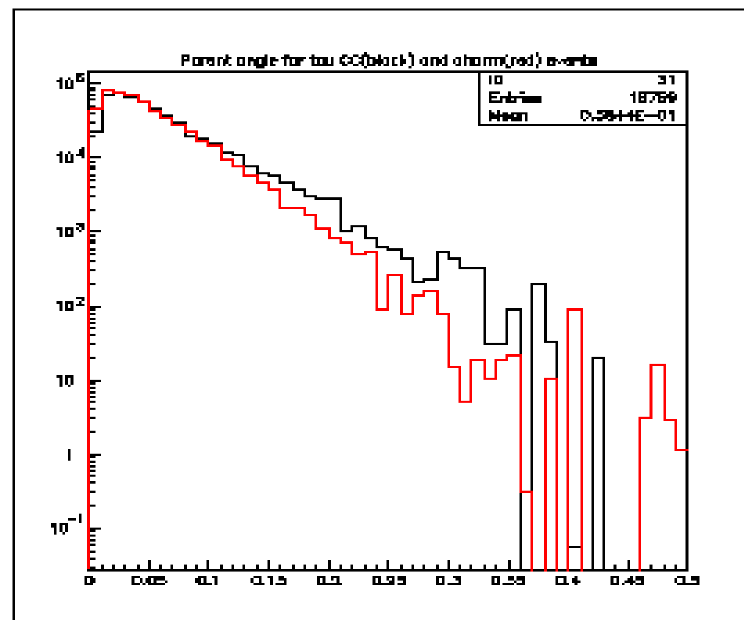
Parent angle

ν_τ *CC* (top)

Hadron Scattering (bottom)



Equal Numbers

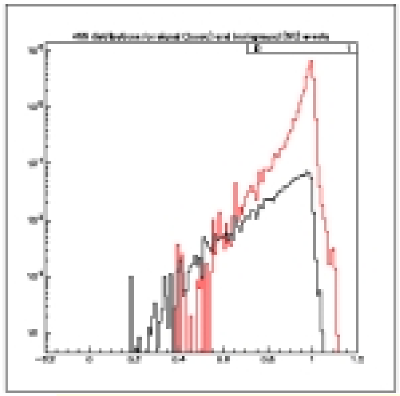


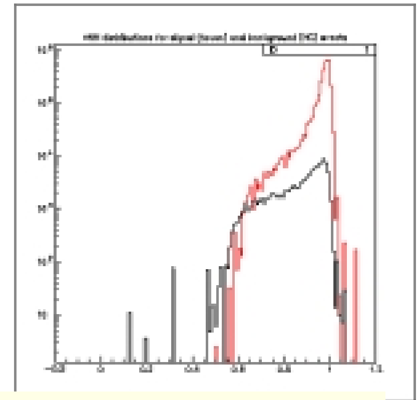
dou & G

ANN ν_τ CC - charm one prong kink decay cont.

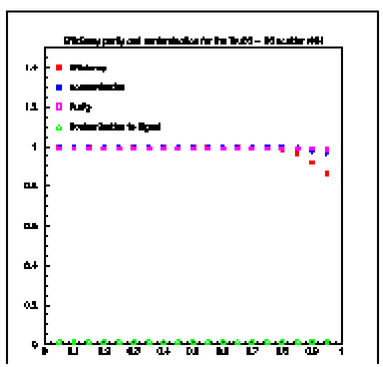
Output ANN function (in log scale)

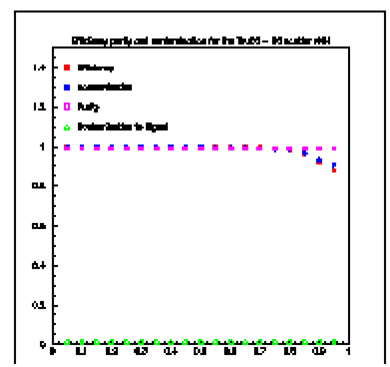
(actual momentum & momentum smeared by 30%)





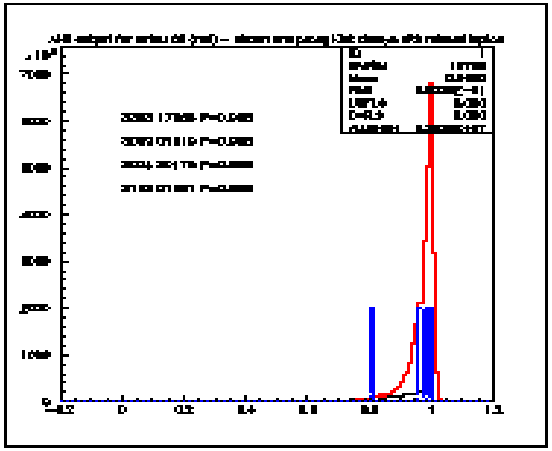
Efficiency, Purity and contamination

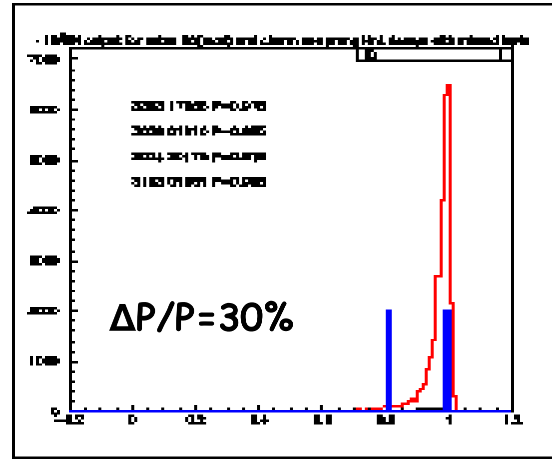




- The classification is very poor (as expected), since all variables characterizing these two populations are almost identical.
- However the event probabilities obtained from this ANN analysis can be used to compute the background from this second source (charm one prong kink decays were the lepton from the primary is missed)

ANN ν_τ CC - charm one prong kink decay background estimation





ν_τ CC (red) charm (black) 4 ν_τ CC -like events (3 ν_τ CC events from phase I and an additional under study from phase II, blue)

RUN	EVENT	P_d	θ_d	P_T	L_d	θ_p	$\Delta\varphi$	Prob. for P	ΔP/P=30%
3024	30175	2.900	0.0936	0.271	4504.8	0.0279	1.027	0.805	0.815
3039	1910	4.600	0.0895	0.412	276.5	0.0653	2.684	0.995	0.985
3333	17665	21.400	0.0130	0.278	564.6	0.0154	2.806	0.985	0.975
3193	1361	20.000	0.0187	0.374	1863.6	0.0838	2.341	0.965	0.985

We compute the background to these events by adding 1-P. Therefore :

	no daugther momentum smearing	30% daughter momentum smearing
Bkg =	0.25	0.24

Decay Search

- The decay search code tested on both MC and experimental data has a quite large efficiency for detecting kinks (~ 86 % on MC and > 95% no Data)
- So far we have only processed part of Phase II data were there was only one kink not detected by Nagoya.
- For this kink the reply we got from Kodama is that it is rejected because it does not satisfy certain code criteria(?).
- We plan to process the remaining phase II events as a cross-check to the results from Nagoya.

Conclusions - on Going work

- The 139 events with good quality m-files have been visually scanned in order to separate the "clean" ones.
- We processed 21 such "clean" events and we are waiting for Nonakas response.
- The ANN analysis on signal from background separation has been corrected and is used for both classification and background estimation.
- This analysis yielded one quite interesting event from phase II 3193_01361 for which we are waiting for Nonakas reply...
- We need to know with greater accuracy the S/B ratio (a priori probabilities) for both sources of background in order to examine their effect on both classification and background estimation.
- We are going to perform decay search analysis on the remaining phase II events as a crosscheck with Nagoya results