

# $\tau$ identification with neural networks for pass2 data

The DØ  $\tau$ -id group

*M. Arov, C. Biscarat, D. Chakraborty, F. Charles, C. Galea, A. Gay, Y. Hu, Y. Gerstein  
A-C. Le Bihan, S. Nelson, C. Nöding, M. Owen, A. Patwa, Y. Pogorelov, S. Protopopescu*

8th April 2005

## Abstract

This note describes the neural network (NN) algorithms developed to identify tau leptons in pass 2 data. We compare the performance of two NN algorithms, which use slightly different input variables on various Monte Carlo and selected data samples.

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Reconstruction algorithm</b>	<b>2</b>
<b>3</b>	<b>Datasets</b>	<b>5</b>
<b>4</b>	<b>Discriminating Variables</b>	<b>5</b>
<b>5</b>	<b>Neural network to reduce the electron contamination</b>	<b>12</b>
<b>6</b>	<b>Neural Net</b>	<b>14</b>
<b>7</b>	<b>Performances of the two NN's</b>	<b>16</b>
<b>8</b>	<b>Performance on <math>\mu\tau</math> final state</b>	<b>21</b>
<b>9</b>	<b>Appendix</b>	<b>32</b>
9.1	<i>NN calculation</i> . . . . .	32
9.2	<i>Calculating NN variables</i> . . . . .	33

# 1 Introduction

Good identification of  $\tau$  leptons is essential for many measurements and searches in the DØ physics program. As already shown in DØ notes 4210 and 4453, it is possible to develop a universal ID algorithm for tau leptons for many physics processes. The efficiency of the algorithm is different for different cases but the algorithm itself remain close to optimum. This ID algorithm is based on neural network (NN) techniques.

A first set of networks was developed and described in DØ note 4210. Later, some improvement was observed training the networks with more variables. This algorithm is described in DØ note 4622.

Since then modifications have been made to the reconstruction program (d0reco) that affect the performance of NN trained with events reconstructed with the earlier versions of d0reco. We refer to pass 2 data as the data reconstructed with p14 (p14.05 or higher) versions of d0reco and reprocessed with the pass 2 version of the TMBfixer. The main difference with pass 1 data (affecting  $\tau$ 's) is the use of T42 algorithm for clustering calorimeter information. Thus, the training of both sets of neural networks had to be repeated with the new data. The results of the training and comparison between MC and data are described in this note.

## 2 Reconstruction algorithm

This section summarises the reconstruction algorithm. More details can be found in DØ notes 4210 and 4453.

The reconstruction program finds  $\tau$  using two algorithms, one seeded using calorimeter energy clusters and the other using high pT track(s). A reconstructed  $\tau$  consists of the following elements:

- **Calorimeter Cluster**, found by simple cone algorithm, cone size  $R = 0.3$ , isolation cone size  $R_{iso} = 0.5$ ,
- **EM Sub-clusters**, i.e. individual  $\pi^0$ 's from  $\tau$  decays; these subclusters are found by a nearest neighbour algorithm in the EM3 layer of the calorimeter. If such clusters are found, then EM cells in other layers and preshower hits are attached to them,
- **Tracks**, which are likely to have been produced in the  $\tau$  decay.

In the reconstruction and further analysis we distinguish three final states, which we refer to as  $\tau$ -types:

1.  $\tau \rightarrow \pi^- \nu_\tau$  - one track with calorimeter cluster and no associated EM sub-cluster,
2.  $\tau \rightarrow \rho^- \nu_\tau \rightarrow \pi^0 \pi^- \nu_\tau$ <sup>1</sup> - one track with calorimeter cluster and at least one associated EM sub-cluster,
3.  $\tau \rightarrow h^- h^+ h^- (\pi^0) \nu_\tau$  - more than one track, with calorimeter cluster and with or without associated EM sub-clusters.

The first two types belong to the "1-prong" category while type 3 is "3-prong".

No attempt is made at the reconstruction stage to separate electrons from  $\tau$ s. Indeed, electrons will be reconstructed with very high efficiency as  $\tau$ -type 2 taus, except in the ICD region ( $1.1 < |\eta_d| < 1.4$ ) where they will pass as  $\tau$ -type 1.

Figures 1 and 2 show the ratio of jets to  $\tau$  for each  $\tau$ -type as function of  $E_T$  and  $\eta$ . A sample of 150,000 events satisfying a single  $\mu$  trigger and only one reconstructed  $\mu$  ( $p_T > 7$  GeV) and

---

<sup>1</sup>There can be more than one  $\pi^0$  in the final state

at least one 0.5 cone jet ( $E_T > 8$  GeV) was used for those plots (only jets with  $\Delta R > 0.5$  respect to the  $\mu$  are counted).

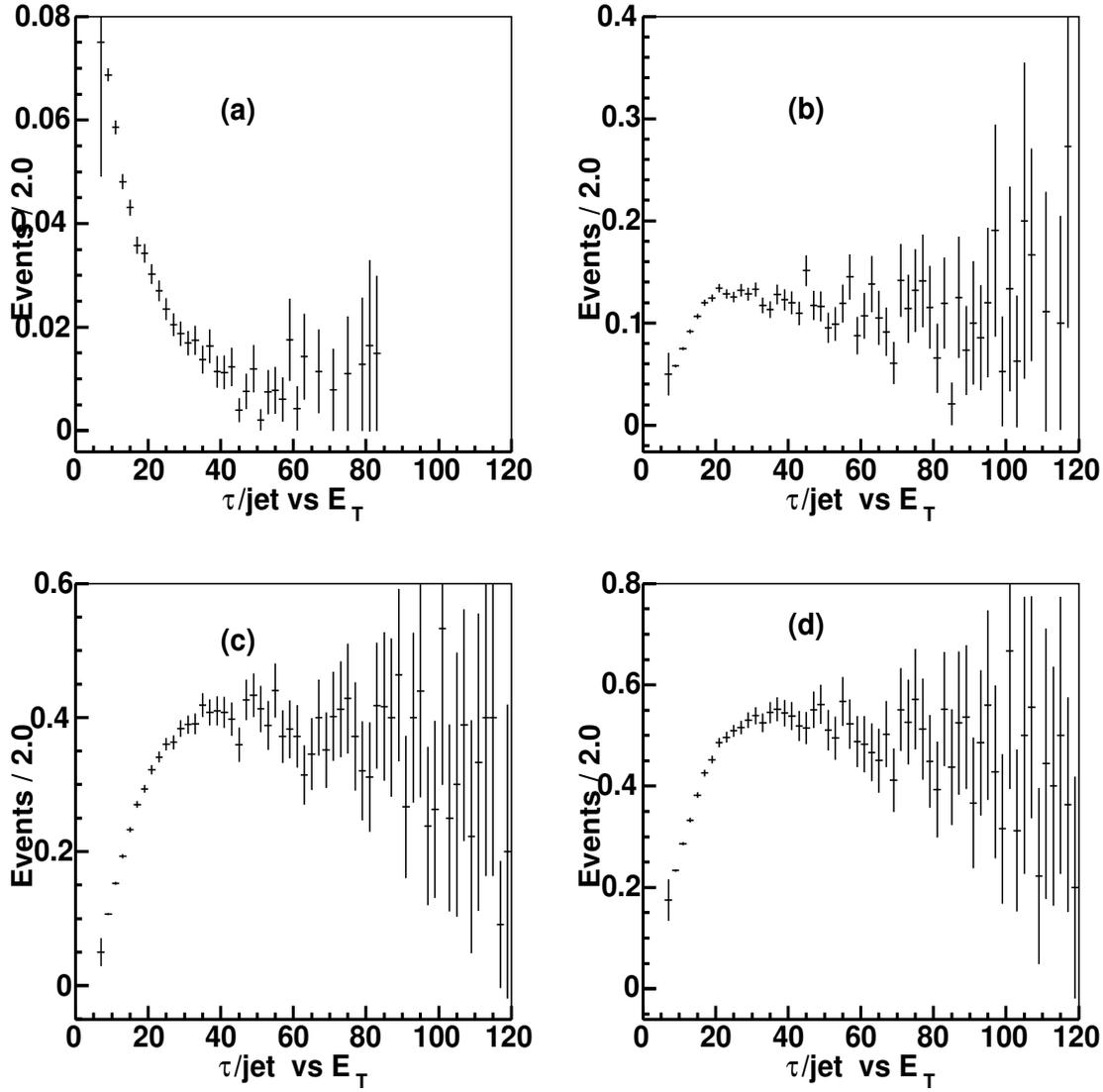


Figure 1: Efficiencies as function of  $E_T$  for jets reconstructed as  $\tau$  candidates: (a)  $\tau$ -type 1, (b)  $\tau$ -type 2, (c)  $\tau$ -type 3, (d) sum over  $\tau$ -types.

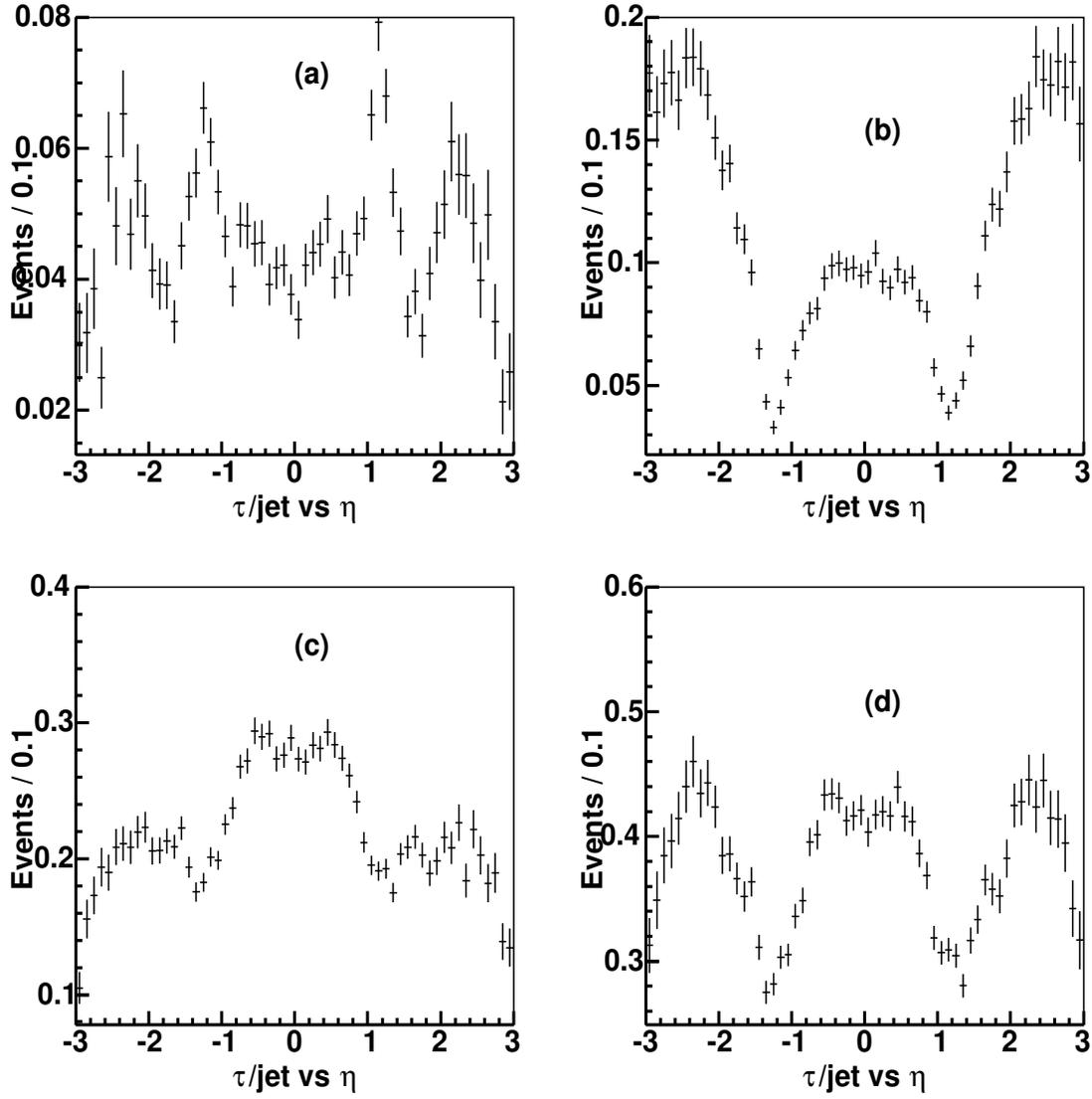


Figure 2: Efficiencies as function of  $\eta$  for jets reconstructed as  $\tau$  candidates: (a)  $\tau$ -type 1, (b)  $\tau$ -type 2, (c)  $\tau$ -type 3, (d) sum over  $\tau$ -types.

### 3 Datasets

For the training and the testing of the neural networks, we used  $\tau$  candidates defined by the algorithm described in the previous section.

The signal sample consists of 100 000 single  $\tau$  events generated with pythia without constraints on the decay of the  $\tau$  and reconstructed using reco version p14.05.02. Out of this sample we made 3 sub-samples of hadronically decaying  $\tau$ s, each with only  $\tau$ s of the same type. Each of these sub-samples is in fact doubled: in the first case we restrict the decay of the tau to hadronic modes only (using MC information), in the second case we also allow  $\tau \rightarrow e$  decay. The training of the neural networks will be repeated for these two cases.

The background was modeled from the NP 1mutrk skim (p14.06.01). Events with tight muons within a jet ( $\Delta(R)_{\tau,\mu} < 0.5$ ) and a  $\tau$  candidate with same sign as the muon have been selected to form a data QCD sample. A further requirement  $\Delta(\phi)_{\tau,\mu} > 0.7$  was applied. This sample has been also split into 3 sub-samples, one for each  $\tau$ -type.

### 4 Discriminating Variables

In this section are described the sets of discriminating variables used for the two NN trainings. We will refer to each NN as  $NN_I$  and  $NN_{II}$  respectively.

- Variables used for both NN's:

1.  $profile = (E_{T_1} + E_{T_2})/E_T^\tau$ , where  $E_{T_1}$  and  $E_{T_2}$  are the  $E_T$  of the two most energetic calorimeter towers. Used for all  $\tau$ -types and both NN's.
2.  $caliso = (E_T^\tau - E_T^{core})/E_T^{core}$ . A calorimeter isolation parameter used for all  $\tau$ -types and both NN's.
3.  $trkiso = \Sigma p_T^{trk} / \Sigma p_T^{\tau trk}$ , where  $p_T^{trk}$  ( $p_T^{\tau trk}$ ) is the  $p_T$  of a track within a  $R < 0.5$  cone not associated (associated) with the  $\tau$  candidate. A track isolation parameter used for all  $\tau$ -types.
4.  $em12isof = (E^{EM_1} + E^{EM_2})/E^\tau$  in a  $R < 0.5$  cone, where  $E^{EM_1}$  and  $E^{EM_2}$  are the energies deposited in the first two layers of the EM calorimeter. A parameter used for  $\tau$ -type 1 to reject jets with one energetic charged track and soft  $\pi^0$  mesons.
5.  $\delta\alpha = \sqrt{(\Delta\phi/\sin\theta)^2 + (\Delta\eta)^2}/3.1416$ , where the differences are between  $\Sigma\tau$ -tracks and  $\Sigma EM$ -clusters. In the small angle approximation the observed  $\tau$  mass is given by  $e_{12} \cdot E_T^\tau \cdot \delta\alpha$ . Used for  $\tau$ -types 2 and 3.

- Variables used for  $NN_I$  only:

6.  $e_{12} = \sqrt{\Sigma p_T^{\tau trk} \cdot E_T^{EM}}/E_T^\tau$ , where  $E_T^{EM}$  is the transverse energy deposited in the EM layers of the calorimeter. Used for  $\tau$ -types 2 and 3.
7.  $p_T^{\tau trk1}/E_T^\tau$ , where  $p_T^{\tau trk1}$  is  $p_T$  of the highest  $p_T$  track associated with the  $\tau$ . Used for  $\tau$ -type 1.
8.  $p_T^{\tau trk1}/(E_T^\tau \cdot caliso)$ . A parameter used for  $\tau$ -type 2 that measures the correlation between track and energy deposition in isolation annulus.

- Variables used for  $NN_{II}$  only:

6.  $rms_\tau = \sqrt{\Sigma_{i=1}^n [(\Delta\phi_i)^2 + (\Delta\eta_i)^2] E_{T_i}/E_T^\tau}$ . A measure of  $\tau$ -cluster width. Used for all  $\tau$ -types.
7.  $fhf$  is the fraction of  $E_T^\tau$  fine hadronic section of the calorimeter. Used for  $\tau$ -types 1 and 2.

8.  $E_T^{em}/E_T^\tau$ , where  $E_T^{em}$  is the transverse energy of the EM subclusters. Use for  $\tau$ -types 2 and 3.
9.  $prf3$  is the transverse energy of the leading EM subcluster divided by the transverse energy in the layer 3 of the calorimeter in a  $R < 0.5$  cone.
10.  $E_T^\tau/(E_T^\tau + \Sigma p_T^{\tau, trk})$  used for  $\tau$ -types 2 and 3.

The figures 3 to 8 show the input variables distributions for the single tau MC sample and the background.

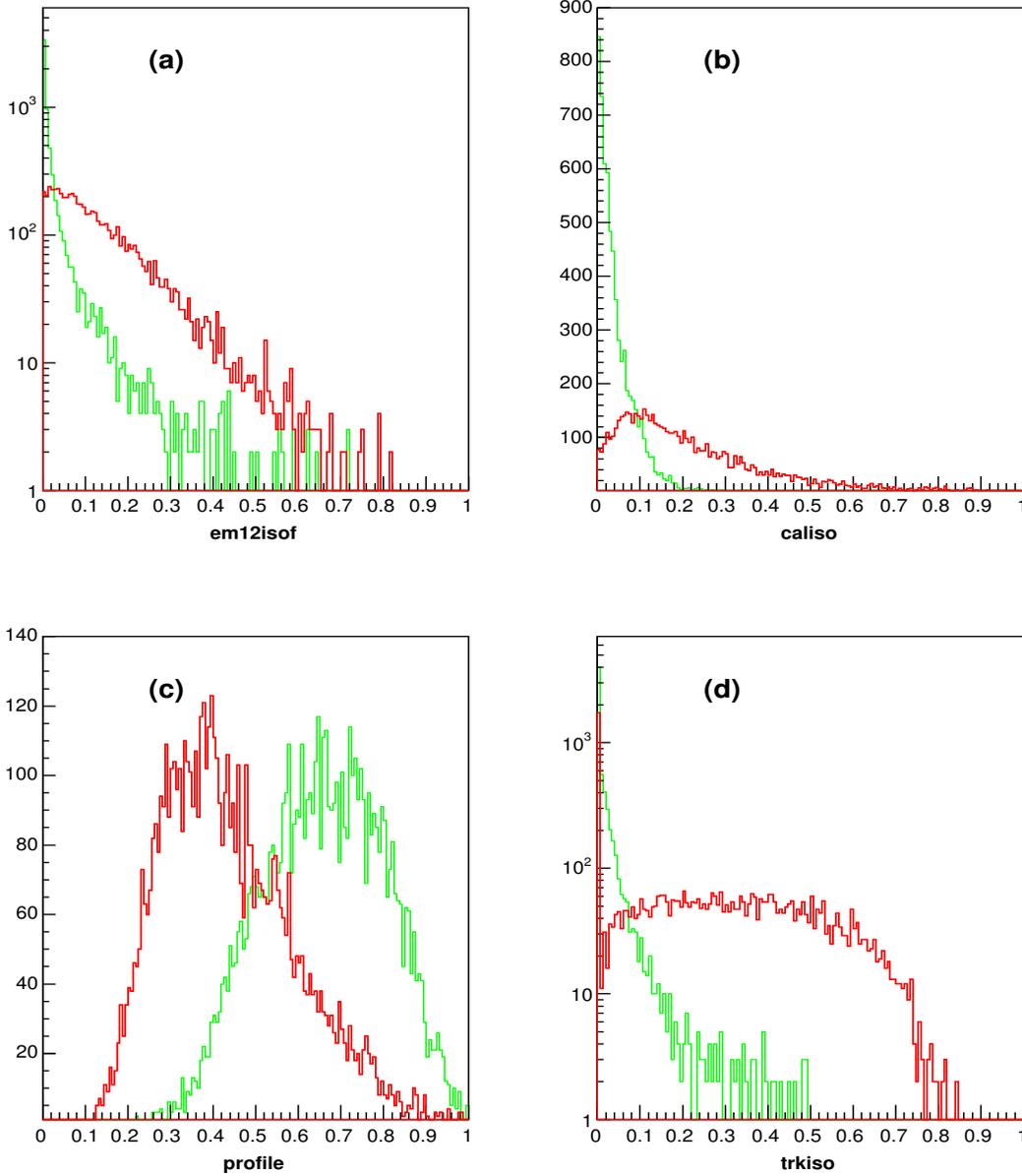


Figure 3: NN input variables for  $\tau$ -type 1. The signal (background) is represented by the green (red) line: (a) *em12isof*, (b) *caliso*, (c) *profile*, (d) *trkiso*

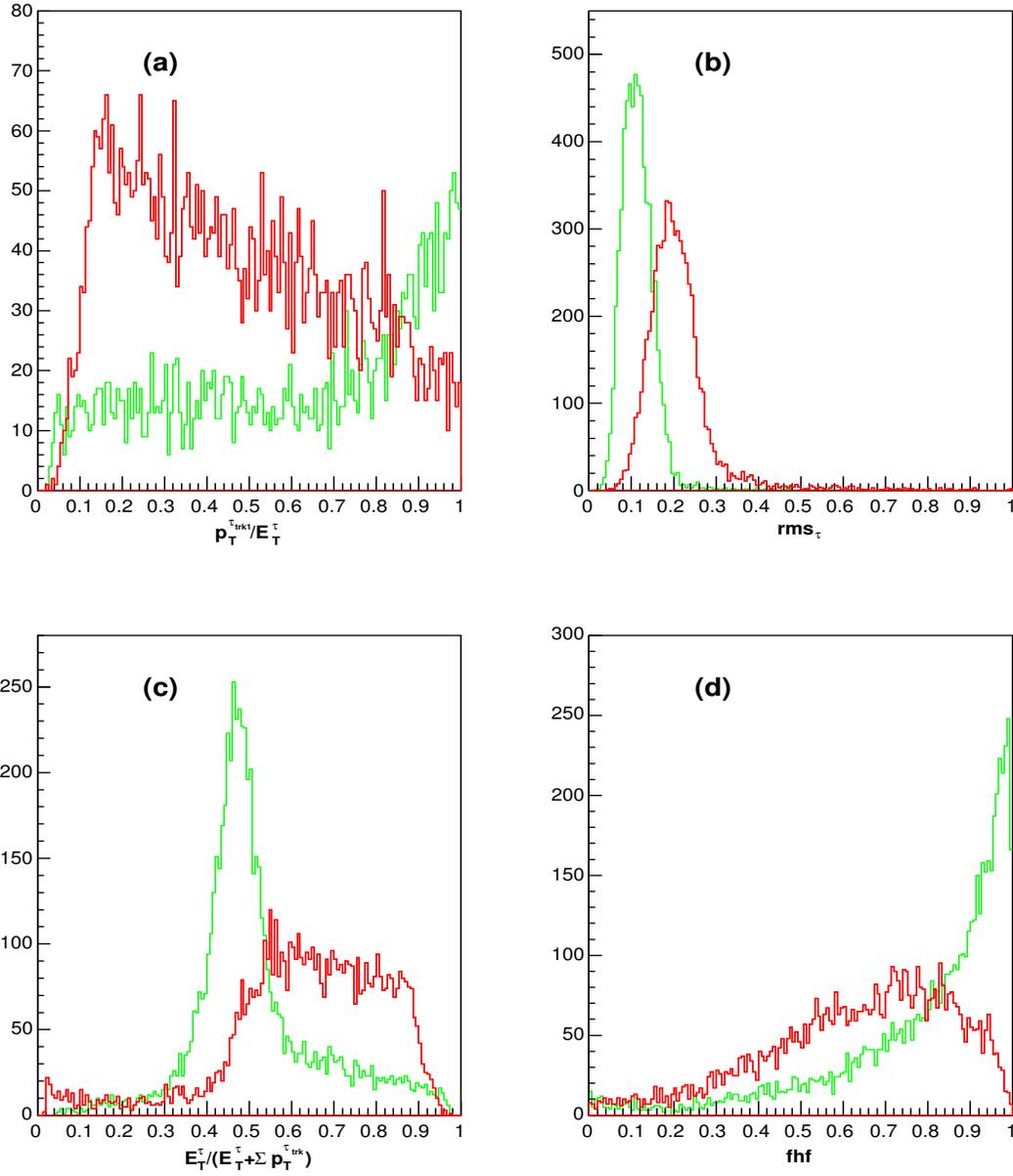


Figure 4: Additional NN input variables for  $\tau$ -type 1 . The signal (background) is represented by the green (red) line: (a)  $p_T^{trk1}/E_T^\tau$ , (b)  $rms_\tau$ , (c)  $E_T^\tau/(E_T^\tau + \Sigma p_T^{trk})$ , (d)  $fhf$

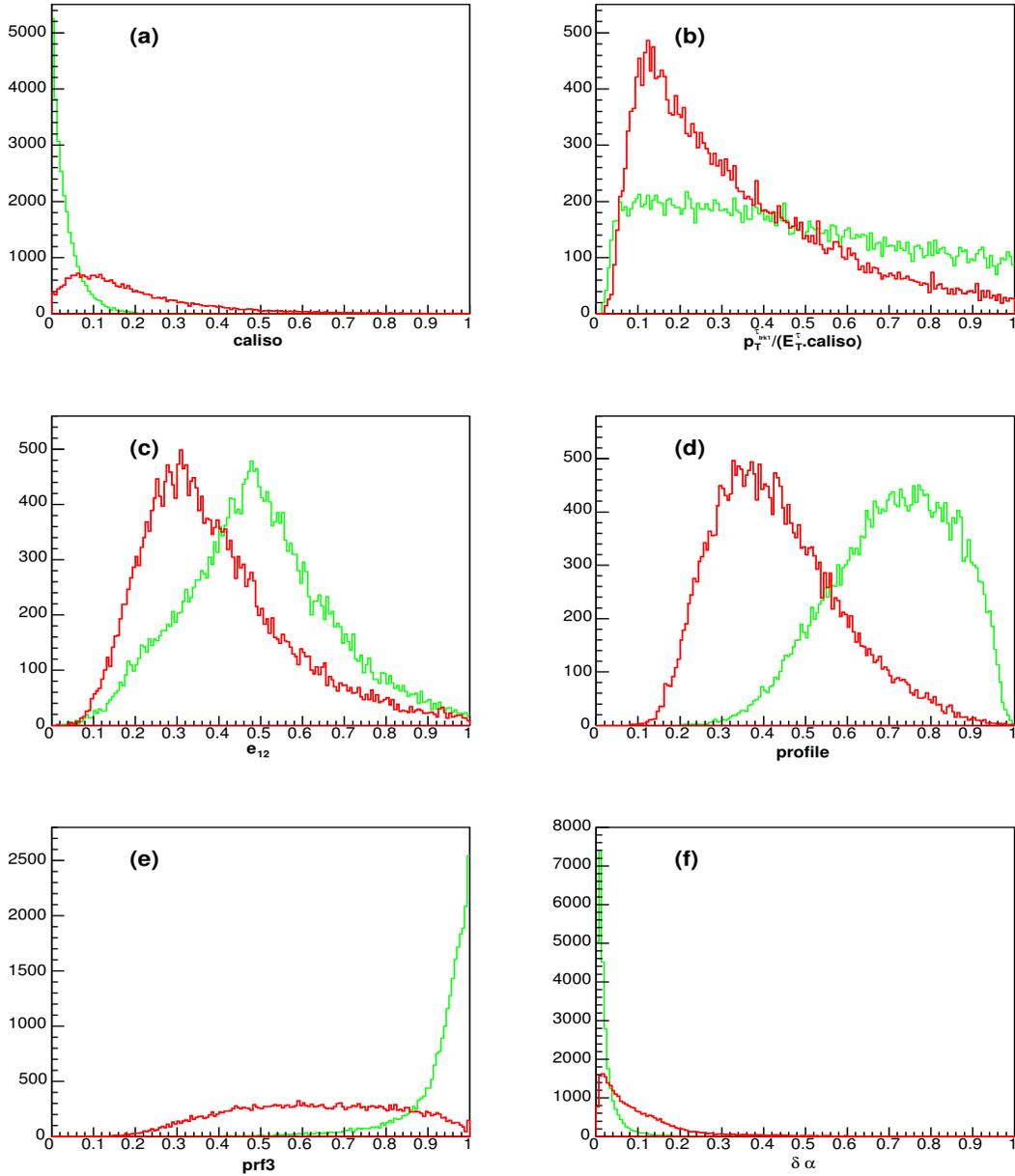


Figure 5: NN input variables for  $\tau$ -type 2 . The signal (background) is represented by the green (red) line: (a)  $caliso$ , (b)  $p_T^{trk1}/(E_T^{caliso})$ , (c)  $e_{12}$ , (d)  $profile$ , (e)  $prf3$ , (f)  $\delta\alpha$

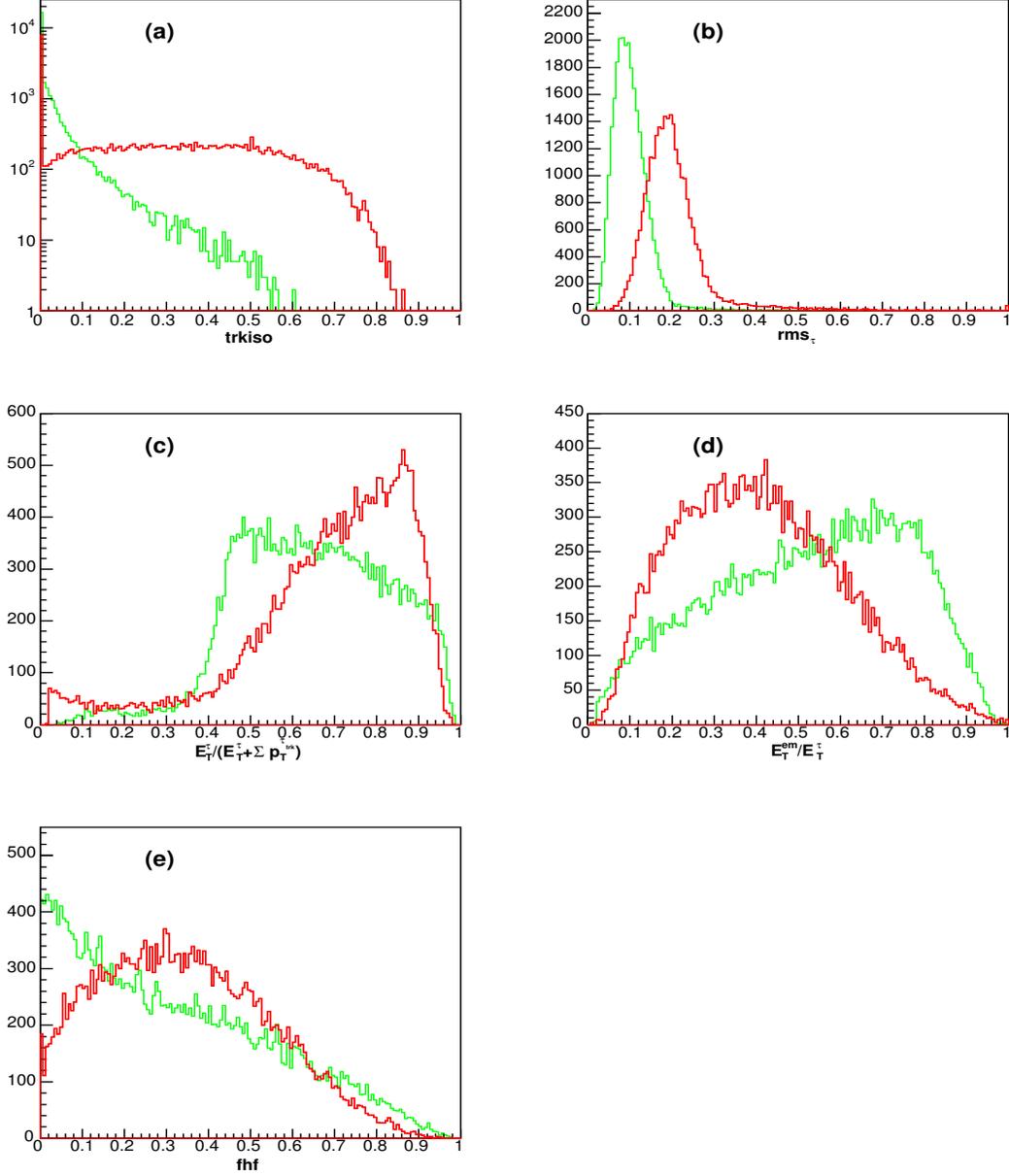


Figure 6: Additional NN variables for  $\tau$ -type 2 . The signal (background) is represented by the green (red) line: (a)  $trkiso$ , (b)  $rms_\tau$ , (c)  $E_T^\tau / (E_T^\tau + \Sigma p_T^{\tau_{trk}})$ , (d)  $E_T^{em} / E_T^\tau$ , (e)  $fhf$

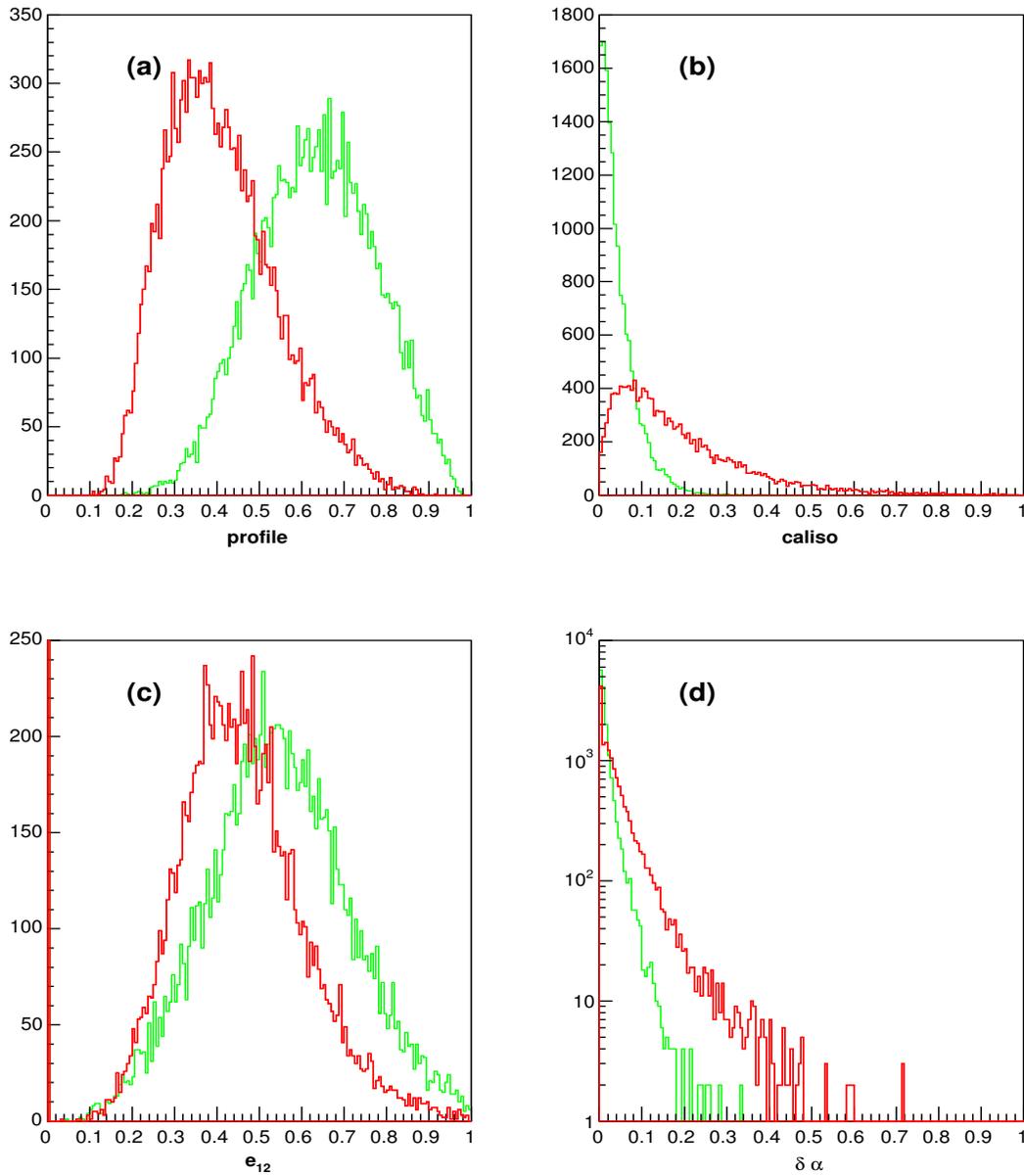


Figure 7: NN input variables for  $\tau$ -type 3 . The signal (background) is represented by the green (red) line: (a) profile, (b) caliso, (c)  $e_{12}$  (d)  $\delta\alpha$

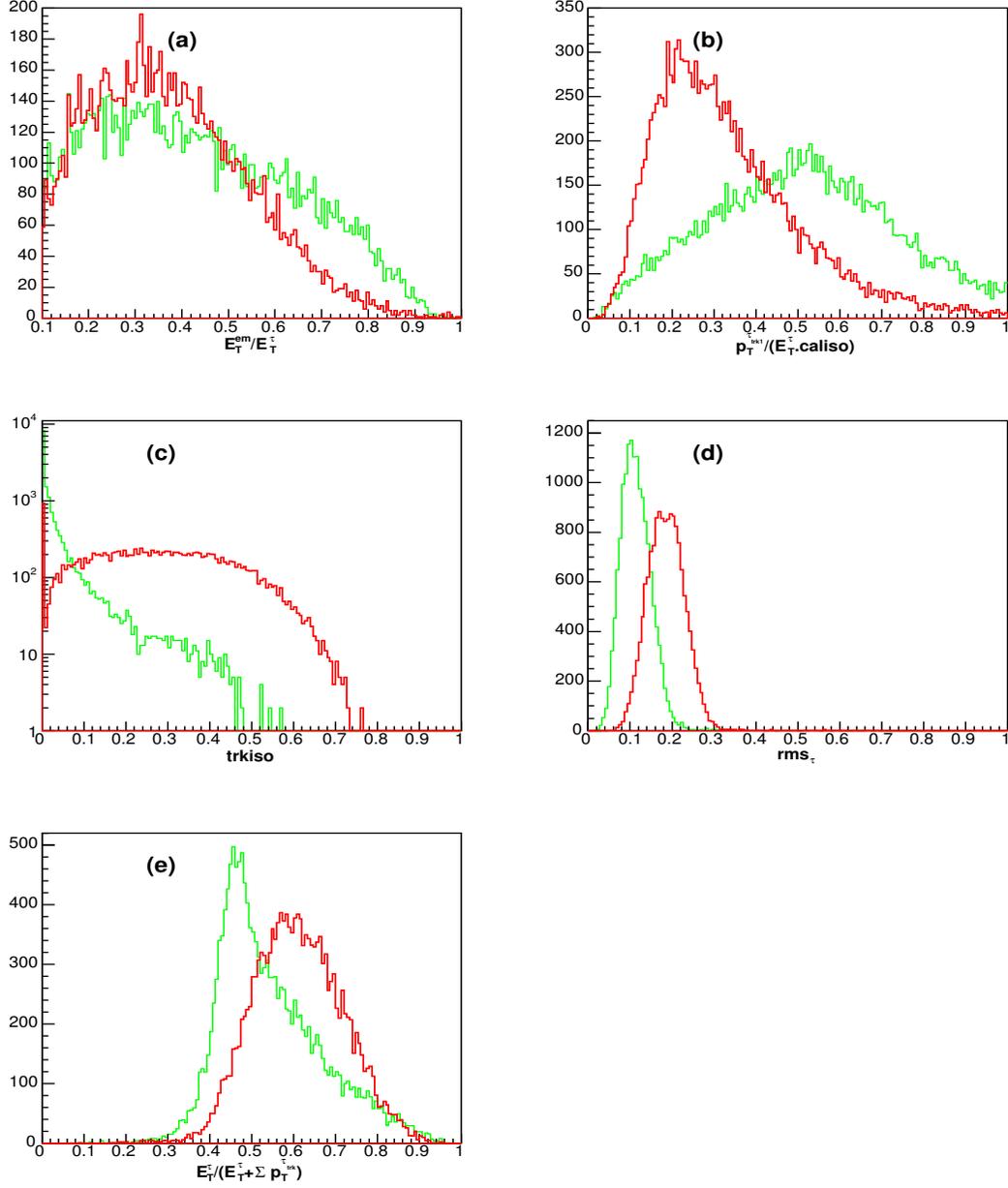


Figure 8: Additional NN input variables for  $\tau$ -type 3. The signal (background) is represented by the green (red) line: (a)  $E_T^{em}/E_T^\tau$ , (b)  $p_T^{\tau_{trk1}}/(E_T^\tau \cdot caliso)$ , (c)  $trkiso$ , (d)  $rms_\tau$ , (e)  $E_T^\tau/(E_T^\tau + \Sigma p_T^{\tau_{trk}})$

## 5 Neural network to reduce the electron contamination

In addition to the neural network developed to disentangle QCD jets from hadronic taus, a neural network dedicated to separate electrons from type 2 taus was developed ( $NN_e$ ). Indeed, most of the electrons are recognized as type 2 taus.

For the signal, the training sample is the same as the one used to train the previous networks. The background sample is made out of Monte Carlo  $Z \rightarrow ee$  events generated with PYTHIA and reconstructed with D0reco version p14.05.01 and keeping only electrons identified as  $\tau$ -type 2.

The variables used for the training of  $NN_e$  are:

1. *profile*
5.  $\delta\alpha$
6.  $rms_\tau$
7.  $E_T^{em}/E_T^\tau$
9.  $E_T^\tau/(E_T^\tau + \Sigma p_T^{\tau_{trk}})$

Figure 9 shows the  $NN_e$  performance. These variables are defined in previous section and are shown on the figure 10. Figure 11 compares  $NN$  for data and MC. It shows that the  $NN$  obtained with data is very close to that obtained with MC events for  $NN_e$  trained with MC events.

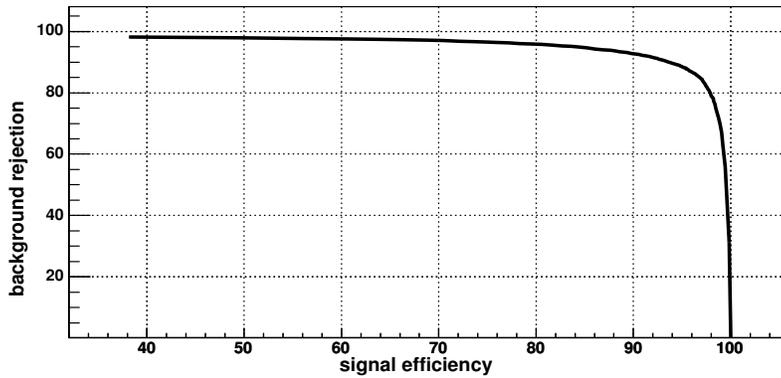


Figure 9: *Background rejection vs signal efficiency for the NN dedicated to separate electrons from type 2 taus.*

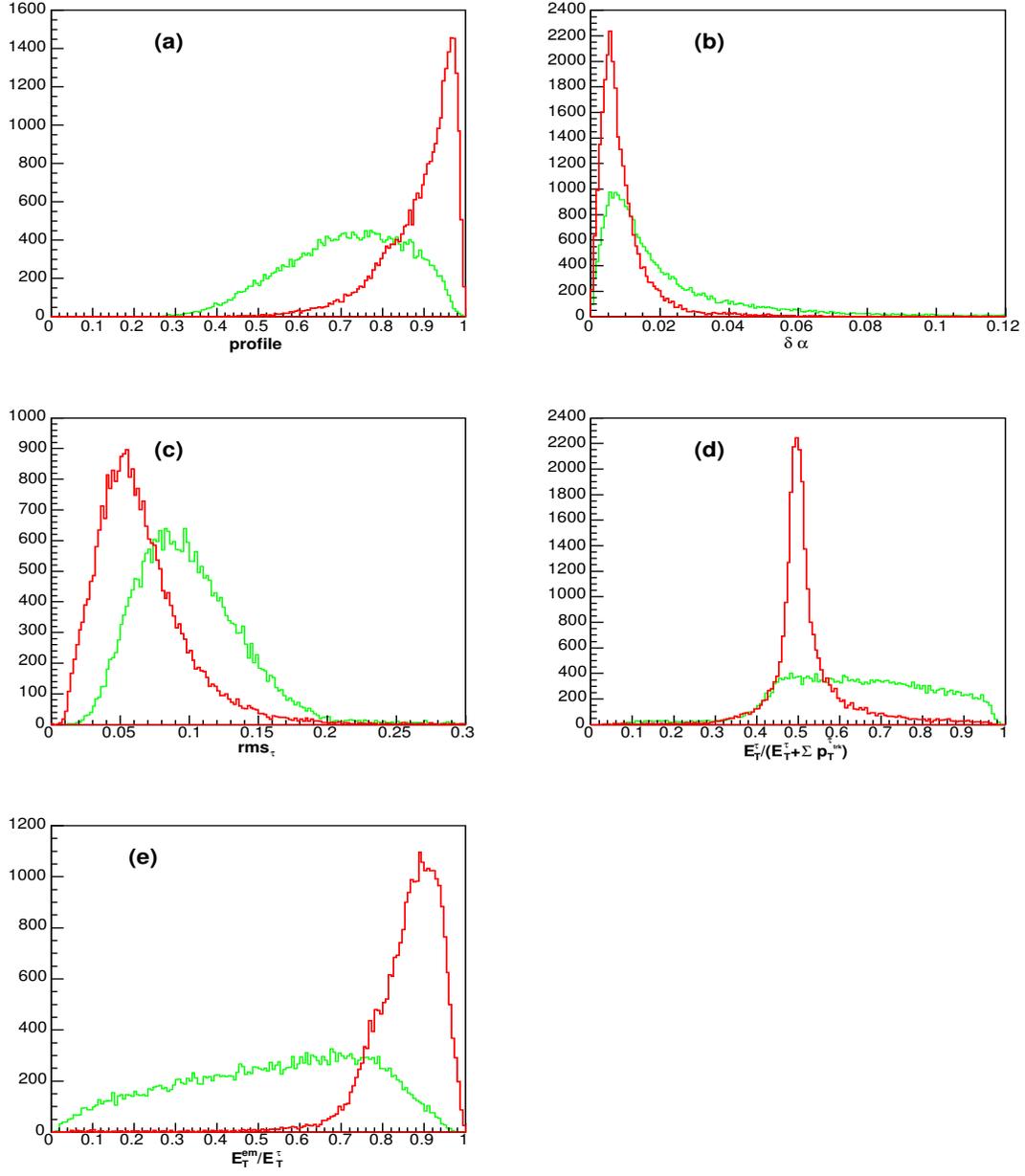


Figure 10:  $NN_e$  input variables. The signal (background) is represented by the green (red) line: (a) profile, (b)  $\delta\alpha$ , (c)  $rms_\tau$ , (d)  $E_T^\tau / (E_T^\tau + \Sigma p_T^{trk})$ , (e)  $E_T^{em} / E_T^\tau$

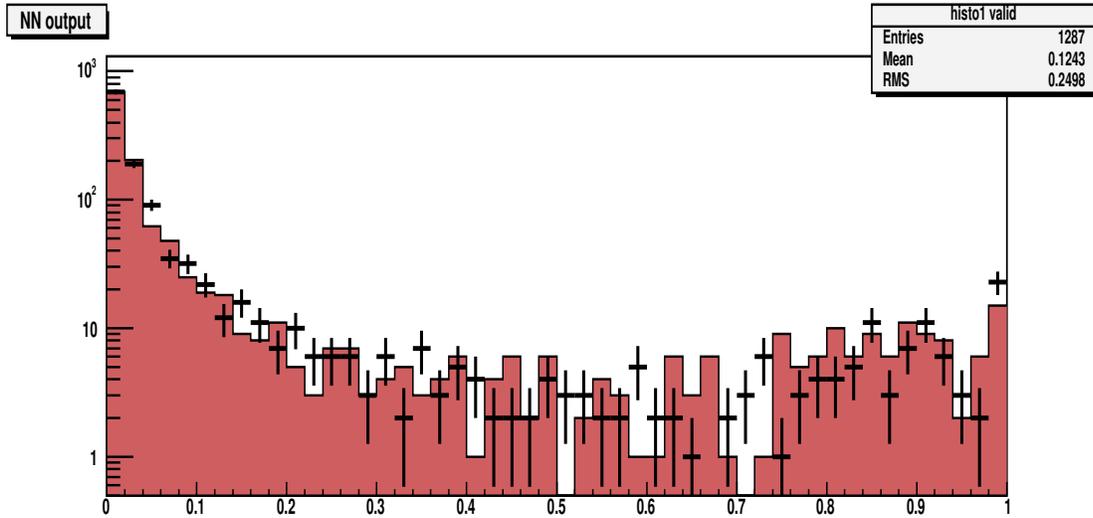
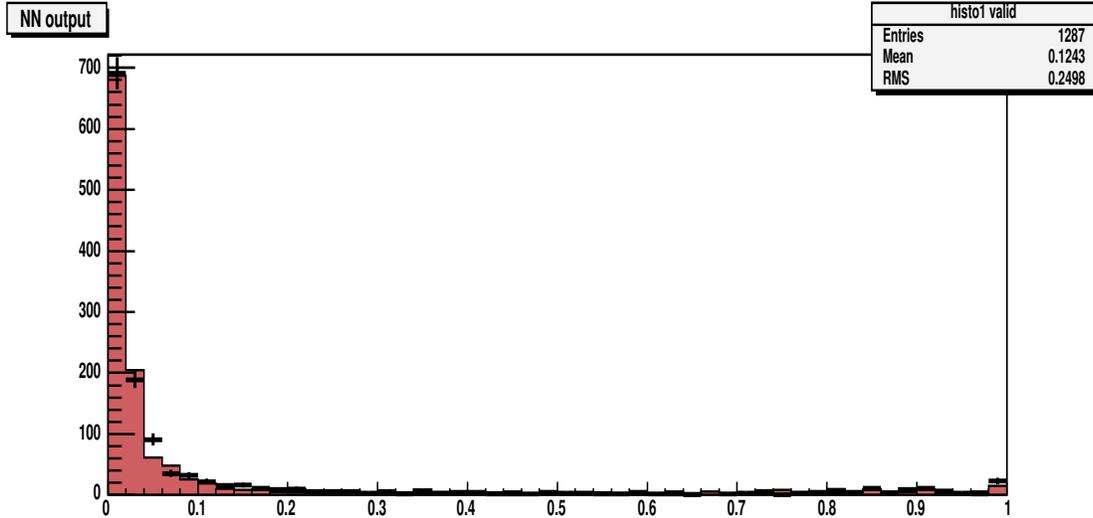


Figure 11: Comparison of NN from  $NN_e$  between  $Z^- \rightarrow ee$  MC (points) and  $Z^- \rightarrow ee$  data (histogram)

## 6 Neural Net

Three separate NN were built, one for each  $\tau$ -type. The input samples for each NN uses only the sub-samples selected with the  $\tau$ -type one wants to identify.

We have chosen the neural network package from the ROOT example applications <sup>2</sup> which uses a vanilla back-propagation method especially suited for particle physics classification tasks. It also has a clean and easy interface and offers remarkable flexibility in choices of network parameters. So far we used only the simplest configuration of a network, i.e., one which has a single input layer consisting of several nodes (one for each measured variable), a single hidden layer consisting of several nodes, and a single output. There is no connection between any two nodes of a given layer, nor is there any direct connection between the input nodes and the output.

<sup>2</sup>Jean-Pierre Ernenwein <http://e.home.cern.ch/e/ernen/www/NN/index.html>

Each hidden node  $h_j$  represents a neuron that performs a linear combination of input signals,

$$h_j = \sum_{i=1}^{n_i} W_{ij} x_i$$

where  $x_i$  is  $i$ -th input and generates an output in a normalized sigmoid form

$$s(h_j + B_j) = \frac{1}{1 + \exp(-(h_j + B_j))}$$

(the bias  $B_j$  is added in order to improve the response of the sigmoid function)

The output is simply a linear combination of the hidden node outputs, on which a sigmoid function is also applied. For  $n_i$  inputs and  $n_h$  hidden nodes, there are all together  $n_h(n_i + 1)$  free parameters for the weights and  $n_h + 1$  free parameters for the biases. Given a set of signal and a set of background events for training, the program determines the weights and biases by iterative function minimization for optimal signal selection. The weights and biases are corrected in proportion to the error (desired output value - computed value) they have generated and to the derivative of the sigmoid function  $s(s-1)$ . This leads to a more important correction if the output of a given neuron doesn't peak near to 0 or 1. According to the convention, the NN output ( $NN$ ) is chosen to be 1 (0) for the desired signal (background). Each neural network has been trained with a reasonably high number of epochs (500 - 1000). The events used have been selected with a visible calorimetric  $p_T$  above 10 GeV and a physics eta within -3 and 3.

## 7 Performances of the two NN's

Figures 12 and 13 show the background rejection vs signal efficiency and the background efficiency vs signal efficiency for  $NN_I$  and  $NN_{II}$  for similar background and signal samples as used in the training. Figure 14 show the  $NN$  distribution for the two NN's. Figures 15 to 18 show the signal and background efficiency vs  $p_t$  and  $\eta$ . On these samples  $NN_{II}$  significantly outperforms  $NN_I$ .

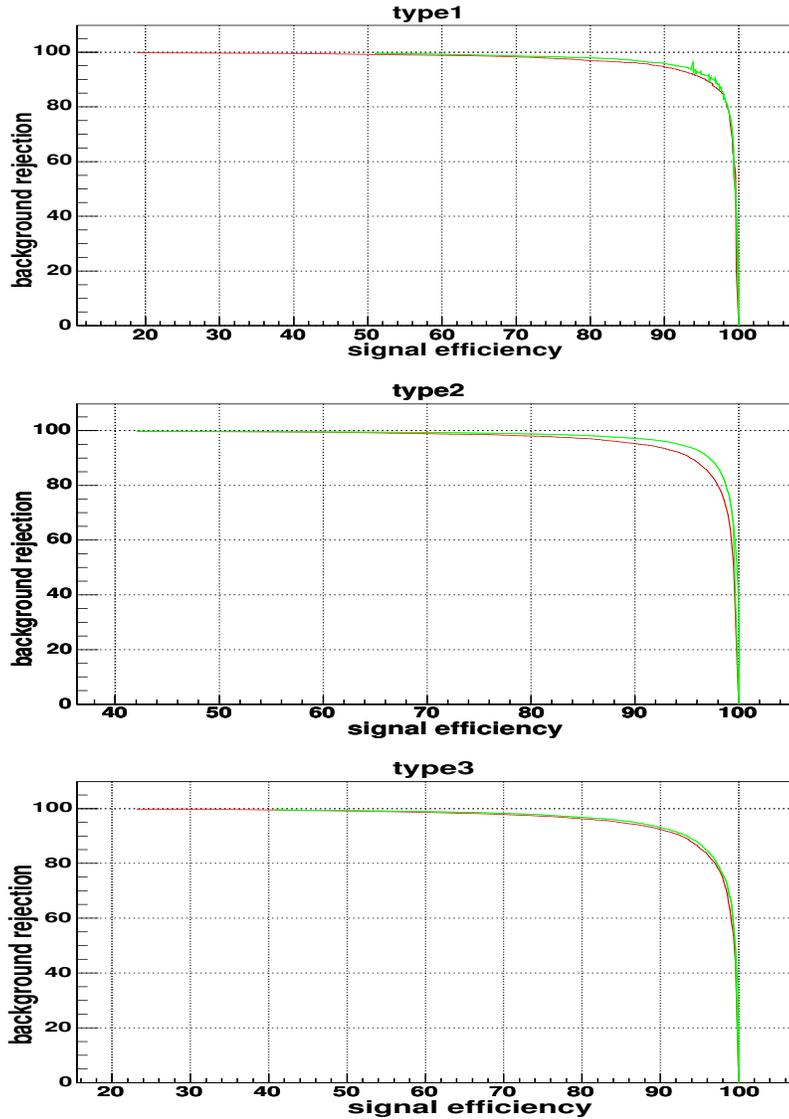


Figure 12: *Background rejection vs signal efficiency for  $NN_I$  (red) and  $NN_{II}$  (green).*

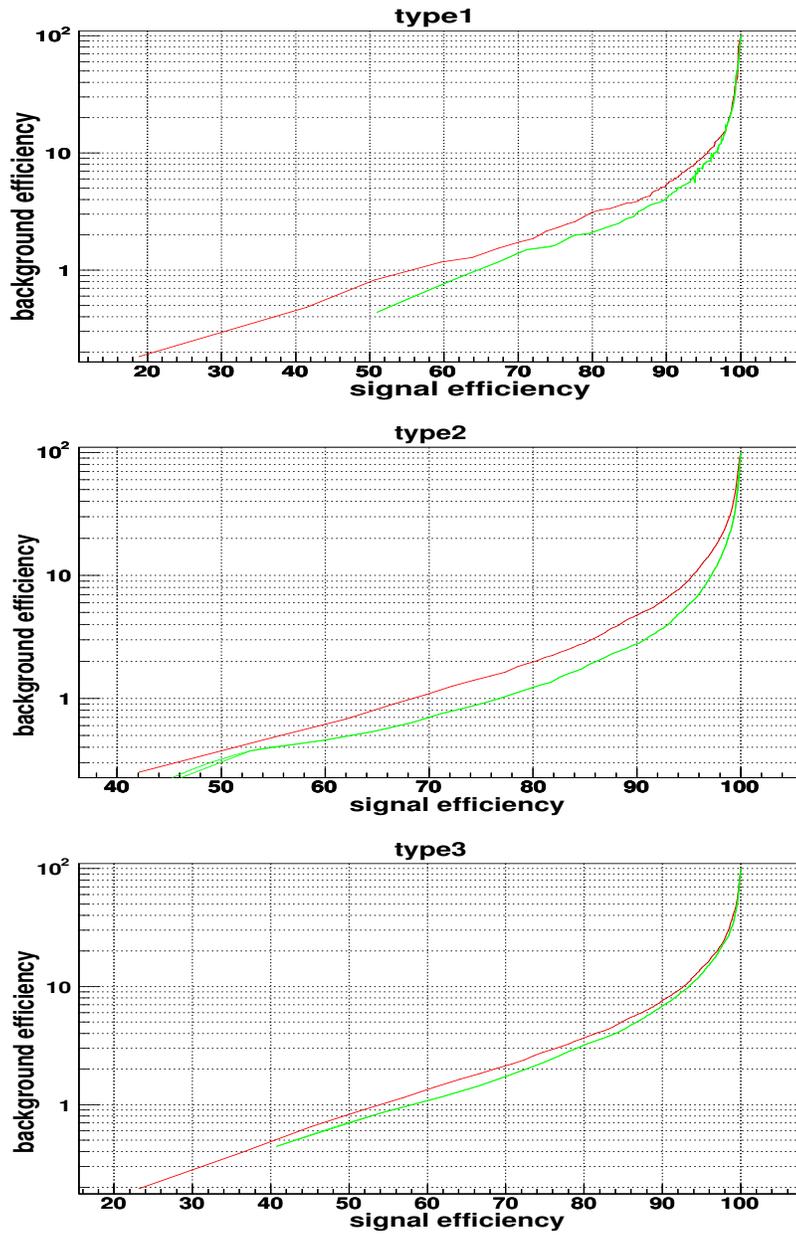


Figure 13: *Background efficiency vs signal efficiency for  $NN_I$  (red) and for  $NN_{II}$  (green).*

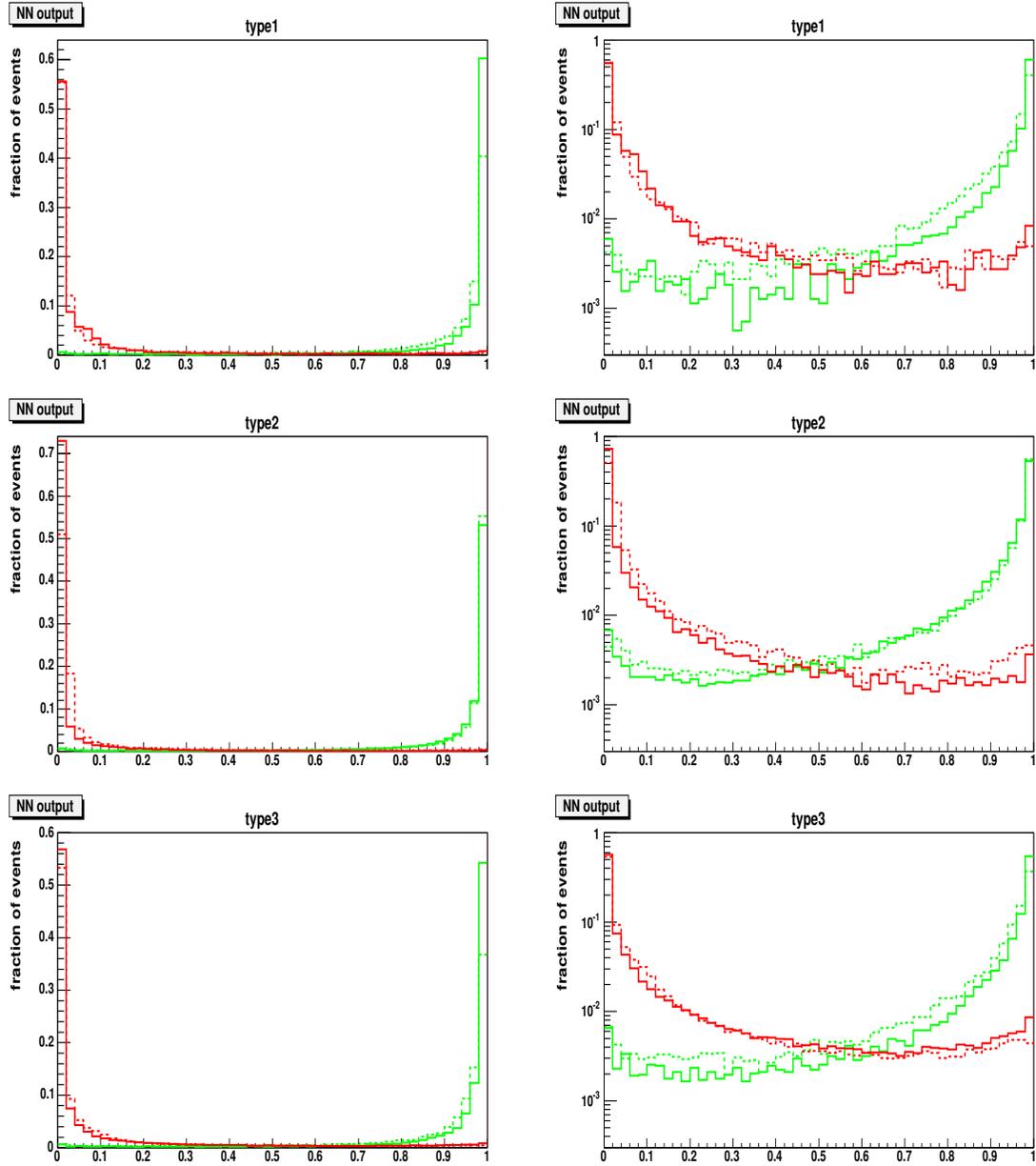


Figure 14:  $NN$  distribution for signal (green) and background (red) for  $NN_I$  (dashed line) and for  $NN_{II}$  (solid line). Left: linear scale; right: logarithmic scale.

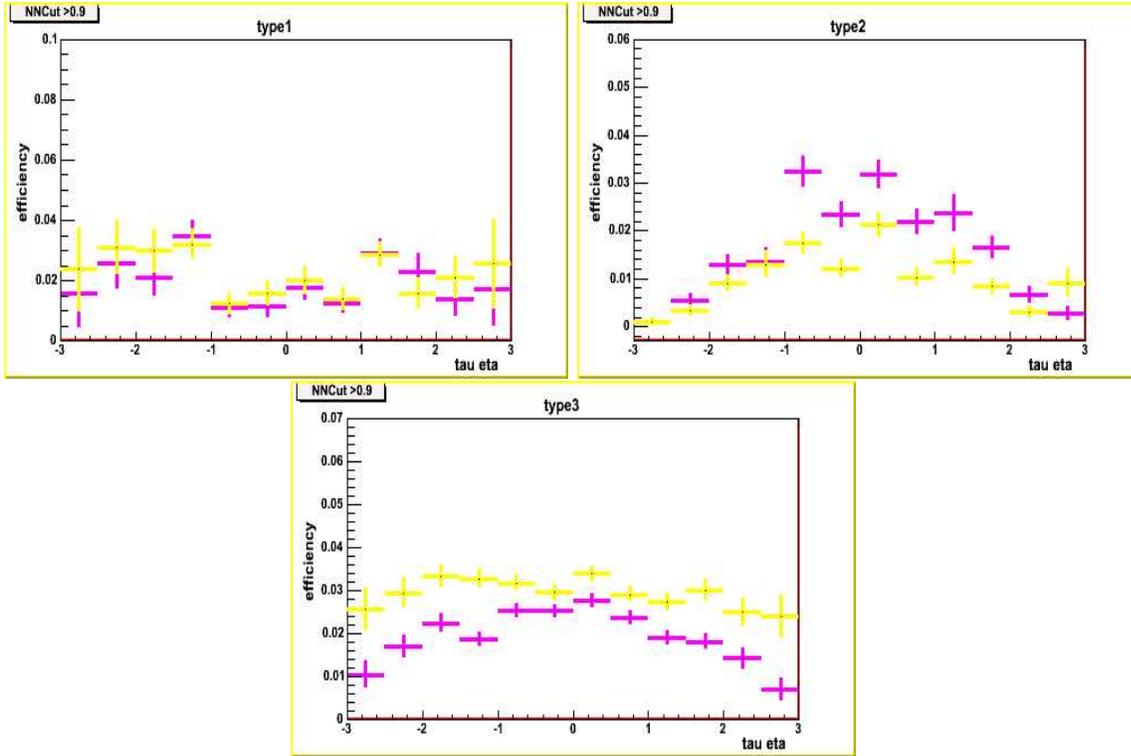


Figure 15: Background efficiencies as function of  $\eta$  for  $NN > 0.9$ ,  $NN_{II}$  (yellow)  $NN_I$  (magenta)

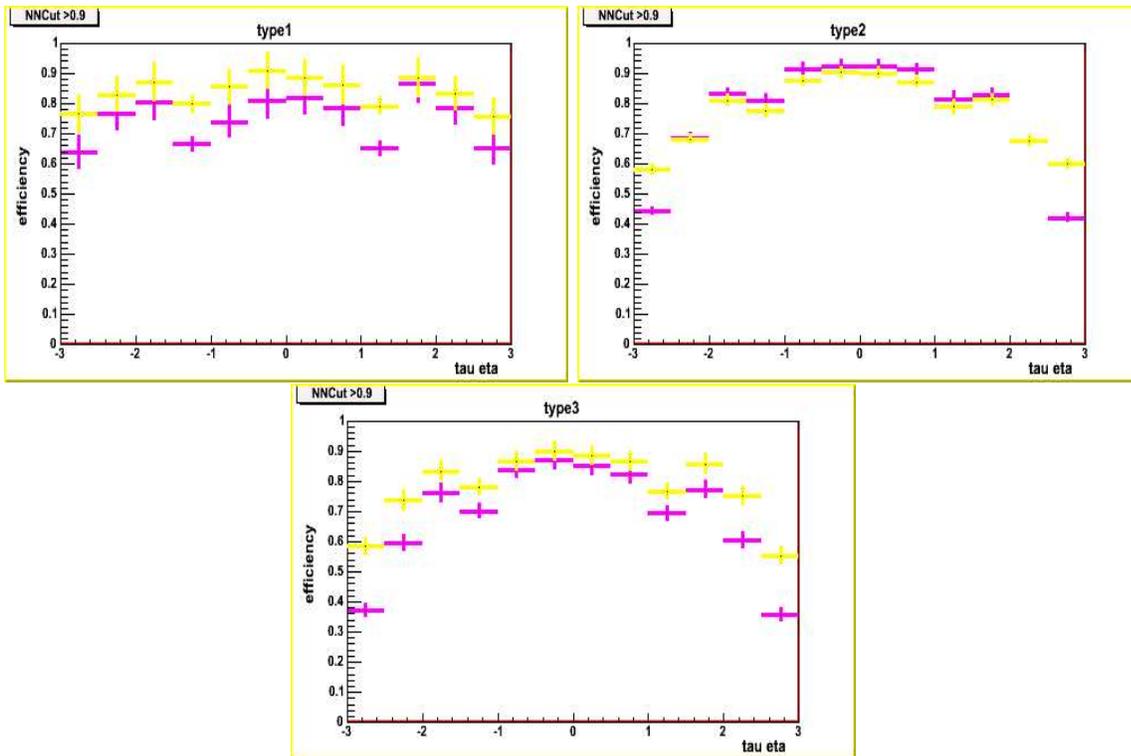


Figure 16: Single  $\tau$  efficiencies as function of  $\eta$  for  $NN > 0.9$ ,  $NN_{II}$  (yellow)  $NN_I$  (magenta)

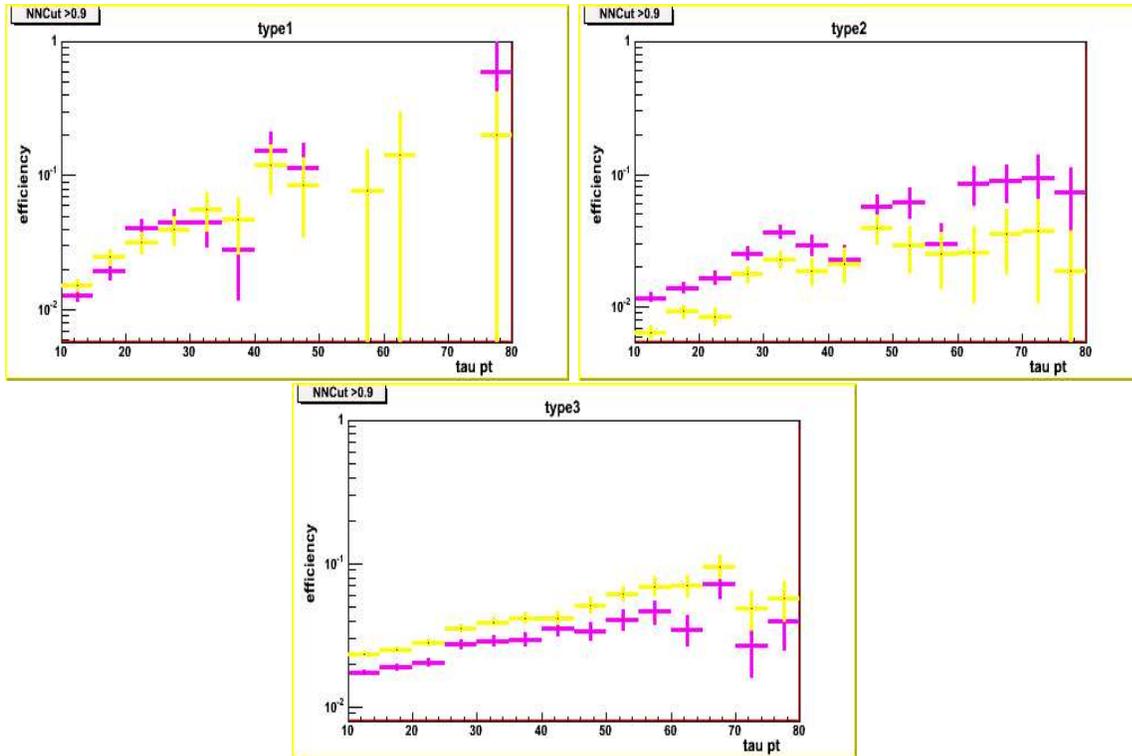


Figure 17: Background efficiencies as function of  $pt$  for  $NN > 0.9$ , new (yellow) old(magenta)

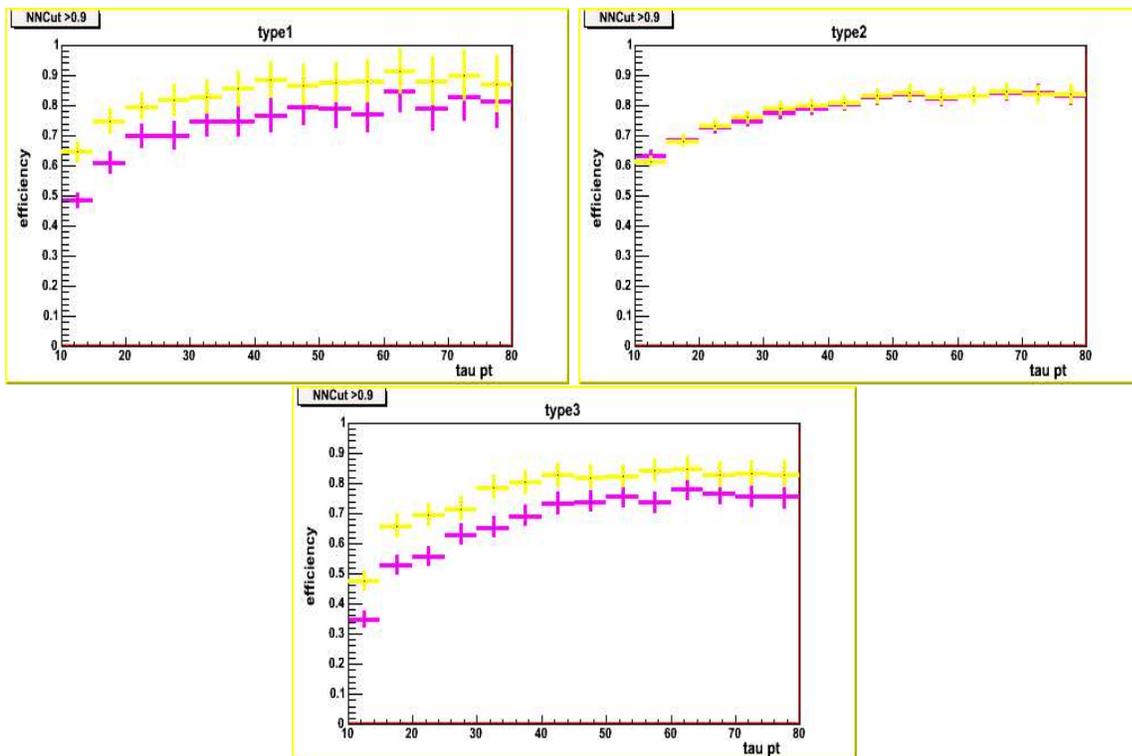


Figure 18: Single  $\tau$  efficiencies as function of  $pt$  for  $NN > 0.9$ ,  $NN_{II}$  (yellow)  $NN_I$ (magenta)

## 8 Performance on $\mu\tau$ final state

In this section we study the performance of the pass 2 NN's on selected events with  $\mu$  and  $\tau$  candidates. The pass2 data includes all good events in run range 162000-196580. The luminosity of the sample is on the order of  $300 \text{ pb}^{-1}$ . Root tmb tree files were made from the 1MUloose skim with the requirements:

- $E_T^\tau > 10(5) \text{ GeV}$ ,  $\Sigma p_T^{\text{trk}} > 7(5) \text{ GeV}$ , (*tau-type 2*).
- semi-isolated  $\mu$ ,  $p_T > 12 \text{ GeV}$ 
  - cal isolation= $[E(R < 0.4) - E(R < 0.1)] < 6.0 \text{ GeV}$
  - trk isolation= $\Sigma p_T < 6.0 \text{ GeV}$

Those files can be found in

```
/rooms/outhouse/projects/pass2_1MUloose_MuTauFilter_p1406  
/work/olemiss-clued0/arov/TAUMERGE/
```

From these events two subsets were made: (1) isolated  $\mu$  sample (191,934 events)

- only one loose  $\mu$
- cal isolation= $[E(R < 0.4) - E(R < 0.1)] < 4.0 \text{ GeV}$
- cal energy along  $\mu = E(R < 0.1) < 4.0 \text{ GeV}$
- trk isolation= $\Sigma p_T < 2.5 \text{ GeV}$

files stored in

```
/work/patchogue-clued0/serban/mutau
```

(2) anti-isolated  $\mu$  sample (291725 events):

- only one loose  $\mu$
- cal isolation= $[E(R < 0.4) - E(R < 0.1)] > 4.0 \text{ GeV}$
- cal energy along  $\mu = E(R < 0.1) > 4.0 \text{ GeV}$
- trk isolation= $\Sigma p_T > 2.5 \text{ GeV}$

files stored in

```
/work/patchogue-clued0/serban/nomuiso
```

In addition there is a sample of 403,000 MC  $Z/\gamma^* \rightarrow \tau\tau$  (mass 60-130 GeV) events in

```
/rooms/bordello/TMBTree_pythia_gam-z-tautau_recop14.05.xx_p14.fixtmb2.03  
/kinmass_60-130
```

Made file with the same requirements as the isolated  $\mu$  sample (20,207 events):

/work/patchogue-clued0/serban/test\_pass2/zmutau\_sel.root

In same directory, files are available with events satisfying the selection cuts used in the  $Z \rightarrow \tau\tau$  cross section measurement (except for the different  $\mu$  isolation criteria).

```

bos*.root    OS with non-isolated muons
bss*.root    SS      ''
os*.root     OS with isolated muons
ss*.root     SS      ''

```

No trigger requirements were imposed. The main effect of asking for events firing the triggers used in the  $Z \rightarrow \tau\tau$  measurement is to reduce the number of events by 30%.

The dominant contributor to isolated  $\mu$  with real  $\tau$ 's is Drell-Yan  $\tau$  pair production ( $Z/\gamma^* \rightarrow \tau^+\tau^-$ ). There are significant backgrounds from fake  $\tau$ 's, mainly from three channels: (1) QCD jets (mainly  $b\bar{b}$ ) with one jet faking a  $\tau$  and the other giving an isolated  $\mu$ , (2)  $W \rightarrow \mu\nu$ +jet with one jet faking a  $\tau$ , and (3)  $Z \rightarrow \mu^+\mu^-$  with one  $\mu$  faking a  $\tau$ . Since (1) is by far the largest when the requirement is made that  $\mu$  and  $\tau$  be back-to-back ( $|\phi^\mu - \phi^\tau| > 2.5$ ) we will concentrate on comparing the behaviour of the NN with old and new set of variables on the MC  $Z/\gamma^* \rightarrow \tau\tau$  sample, the non-isolated  $\mu$  sample and compare that to the isolated  $\mu$  sample. Samples are split into OS ( $\mu$  and  $\tau$  opposite sign charge) and SS ( $\mu$  and  $\tau$  same sign charge)

Figure 19 shows the  $NN$  distributions of each  $\tau$ -type for the MC  $Z/\gamma^* \rightarrow \tau\tau$  OS sample (green points for  $NN_{II}$ , red for  $NN_I$ ).  $NN_{II}$  clearly gives more events for  $NN$  close to 1 for  $\tau$ -types 1 and 2 but slightly less for  $\tau$ -type 2. The number of events for each  $\tau$ -type are given in table 1, before and after an  $NN > 0.9$  cut. Figure 20 show the corresponding  $NN$  distributions for the SS sample of non-isolated  $\mu$ 's. This sample is completely dominated by QCD background. The number of events are given in table 2. For these events the  $NN$  distributions for OS and SS are very similar, with a small excess of OS over SS (see figs 21 and 22).

Table 1: Number of MC  $Z/\gamma^* \rightarrow \tau\tau$  events

	$\tau$ -type 1	$\tau$ -type 2	$\tau$ -type 3	sum
		$NN > 0.0$		
# OS events	1951	10208	4486	16645
	$NN_{II}$ ( $NN_I$ ) $NN > 0.9$			
# OS events	503 (342)	7029 (7316)	3247 (2732)	11704 (11177)
$\epsilon_{OS}^Z$	73.2% (57.9%)	68.9% (71.7%)	72.4% (60.9%)	70.3% (67.1%)

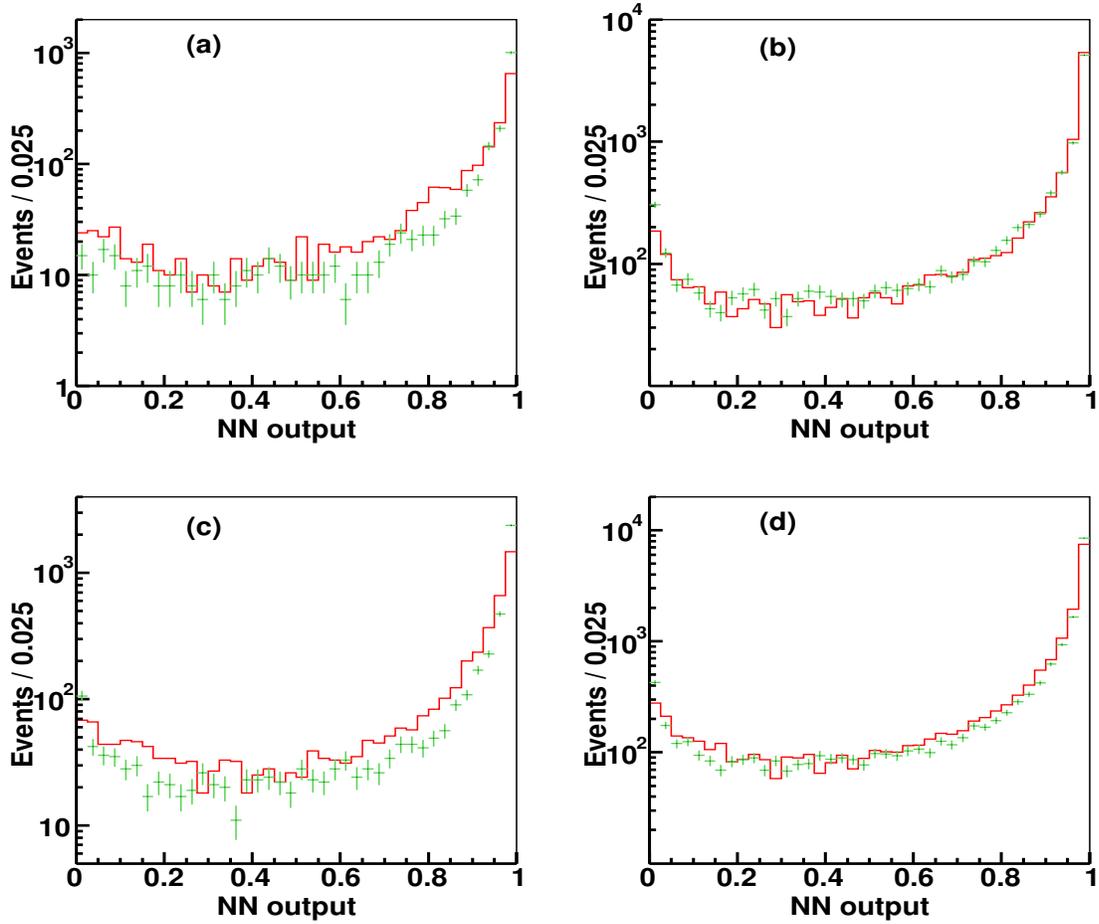


Figure 19:  $NN$  distributions of  $\tau$  candidates OS  $Z/\gamma^* \rightarrow \tau\tau$  Monte Carlo events. Green points for  $NN_{II}$ , red histogram  $NN_I$ . (a)  $\tau$ -type 1, (b)  $\tau$ -type 2, (c)  $\tau$ -type 3, (d) sum over  $\tau$ -types

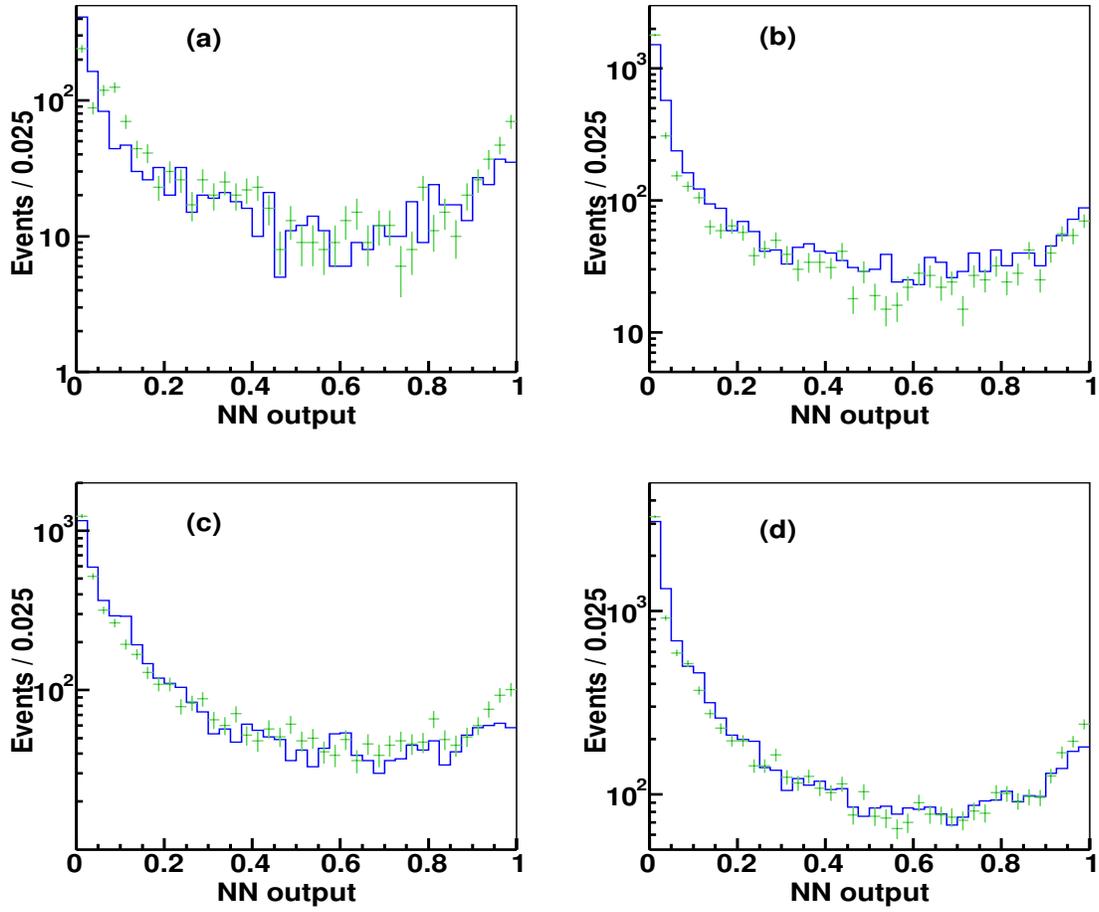


Figure 20:  $NN$  distributions of  $\tau$  candidates in anti-isolated  $\mu$  events. Green points for  $NN_{II}$ , blue histogram for  $NN_I$ . (a)  $\tau$ -type 1, (b)  $\tau$ -type 2, (c)  $\tau$ -type 3, (d) sum over  $\tau$ -types.

The  $NN$  distributions and number of events for the isolated  $\mu$  sample are given in figures 24 and 25, and table 3. Figure 20 There is no obvious gain from using new vs old  $NN$  in the pass  $2\mu\tau$  data. At the fixed value of  $NN > 0.9$  cut whenever signal acceptance is higher for one so is the background acceptance.

Table 2: Number of anti-isolated  $\mu$  events

	$\tau$ -type 1	$\tau$ -type 2	$\tau$ -type 3	sum
	$NN > 0.0$			
# OS events	1436	3912	5033	10381
# SS events	1365	3729	4834	9928
	$NN_{II} (NN_I) NN > 0.9$			
# OS events	163 (127)	186 (260)	369 (268)	718 (655)
$\epsilon_{OS}^n$	11.3% (8.8%)	4.7% (6.6%)	7.3% (5.3%)	6.9% (6.3%)
# SS events	180 (123)	220 (259)	330 (238)	730 (620)
$\epsilon_{SS}^n$	13.2% (9.0%)	5.9% (6.9%)	6.8% (4.9%)	7.3% (6.2%)

The jets from QCD background are mostly b jets while those from the W background are mostly light quark jets. An interesting question is whether the efficiency of a  $NN$  cut is different for the different types of jets. To greatly enrich the proportion of W+jet events in the sample we can cut on transverse mass ( $M_T$ ) and require  $p_T^\mu > 25$  GeV. Figure 26 shows the  $M_T$  distributions for: (i) anti-isolated  $\mu$  events, (ii) 2 OS  $\mu$ 's with one isolated  $\mu$  and a  $\mu\tau$  overlap, (iii) MC  $Z \rightarrow \tau^+\tau^-$ , and (iv) isolated  $\mu$  with standard cuts except  $|\phi_\mu - \phi_{\tau\text{au}}|$  for OS and SS separately. The  $\cancel{E}_T$  for calculating  $M_T$  is obtained by subtracting from the standard calorimeter  $\cancel{E}_T$  the  $\mu$  momenta. The  $M_T$  distributions show clearly the Jacobian peak expected from  $W$ 's for both SS and OS. By selecting events with  $50 < MT < 90$  GeV we get a  $W$  dominated sample. The estimated contribution from  $Z \rightarrow \tau^+\tau^- + Z \rightarrow \mu^+\mu^-$  to the OS sample in that region is of order 0.5% (before a  $NN$  cut). Table 4 gives the number of OS and SS events with  $50 < MT < 90$  GeV,  $p_T^\mu > 25$  GeV and  $|\phi_\mu - \phi_\tau| < 2.5$  cut. One can immediately see that SS has 30% fewer events than OS. After  $NN > 0.9$  cut (we use  $NN_{II}$  for this study), and after correcting for the expected contributions from  $Z$  events, roughly the same excess remains. The efficiencies for both OS and SS are similar and comparable to that of QCD events. We can conclude that within our somewhat limited statistics there is no significant difference in the  $NN$  efficiencies for jets from  $W$ 's and b jets. The only significant difference is that the charge correlations between  $\mu$  from b jets and  $\tau$  candidates is or order 5% while it is quite large (about 30%) in  $W$  events.

Table 3: Number of isolated  $\mu$  events

	$\tau$ -type 1	$\tau$ -type 2 $NN > 0.0$	$\tau$ -type 3	sum
# OS events	2977	11695	14151	28817
# SS events	2274	10271	12070	24617
	$NN_{II} (NN_I) NN > 0.9$			
# OS events	588 (448)	1782 (1950)	1537 (1089)	3907 (3487)
$\epsilon_{OS}^i$	19.7% (15.0%)	15.2% (16.7%)	10.9% (7.7%)	13.6% (12.1%)
# SS events	243 (172)	565 (681)	908 (570)	1716 (1429)
$\epsilon_{SS}^i$	10.7% (7.6%)	5.5% (6.6%)	7.5% (4.7%)	7.0% (5.8%)

Table 4: Number of isolated  $\mu$  events,  $p_T^\mu > 25$  GeV,  $50 < M_T < 90$  GeV,  $|\phi_\mu - \phi_\tau| < 2.5$

	$\tau$ -type 1	$\tau$ -type 2	$\tau$ -type 3	sum
$NN > 0.0$				
# OS events	457	1807	2868	5132
Estimated from Z	8	20	4	32
# SS events	321	1350	2188	3859
new $NN > 0.9$				
# OS events	49	174	241	464
Estimated from Z	5	10	2	17
$\epsilon_{OS}^W$	$9.6 \pm 1.6\%$	$9.1 \pm 0.8\%$	$8.3 \pm 0.6\%$	$8.7 \pm 0.4\%$
# SS events	37	93	138	268
$\epsilon_{SS}^W$	$11.5 \pm 1.8\%$	$6.9 \pm 0.7\%$	$6.3 \pm 0.7\%$	$6.9 \pm 0.6\%$

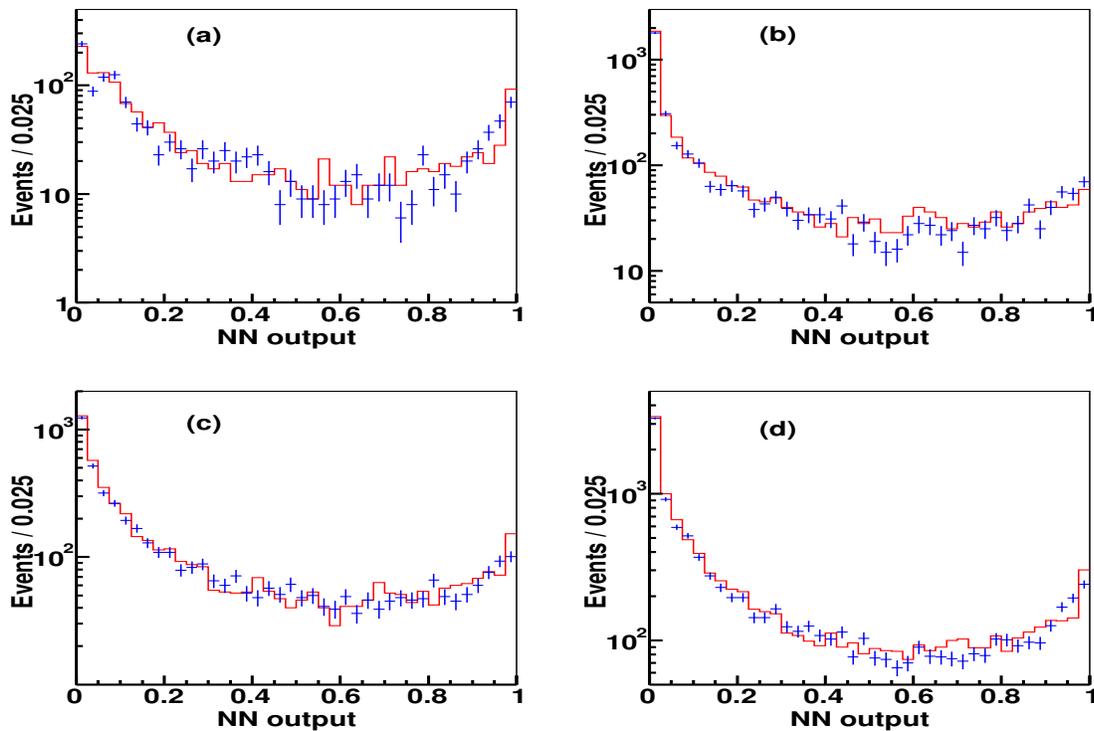


Figure 21:  $NN_{II}$  distributions for  $\tau$  candidates in OS (red) SS (blue) anti-isolated  $\mu$  events. (a)  $\tau$ -type 1, (b)  $\tau$ -type 2, (c)  $\tau$ -type 3, (d) sum over  $\tau$ -types.

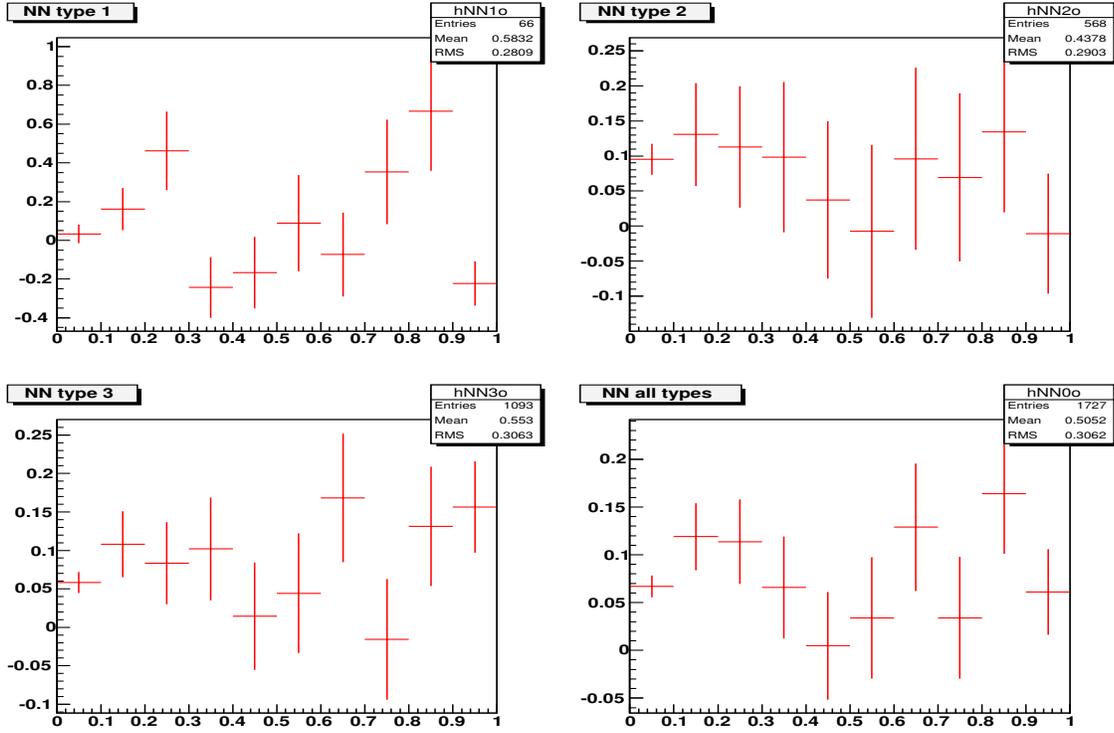


Figure 22: Ratio OS-SS/SS as function of  $NN_{II}$  for anti-isolated  $\mu$ 's. (a)  $\tau$ -type 1, (b)  $\tau$ -type 2, (c)  $\tau$ -type 3, (d) sum over  $\tau$ -types

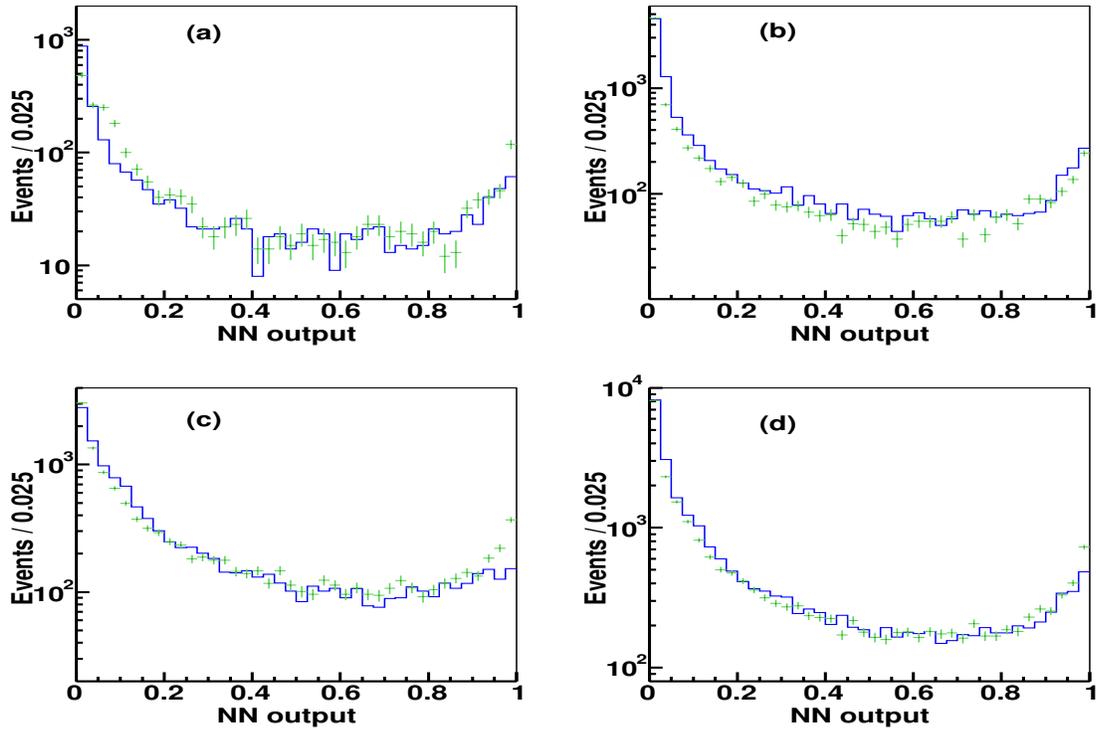


Figure 23:  $NN$  distributions of  $\tau$  candidates in SS isolated  $\mu$  events. Green points for  $NN_{II}$ , blue histogram for  $NN_I$ . (a)  $\tau$ -type 1, (b)  $\tau$ -type 2, (c)  $\tau$ -type 3, (d) sum over  $\tau$ -types.

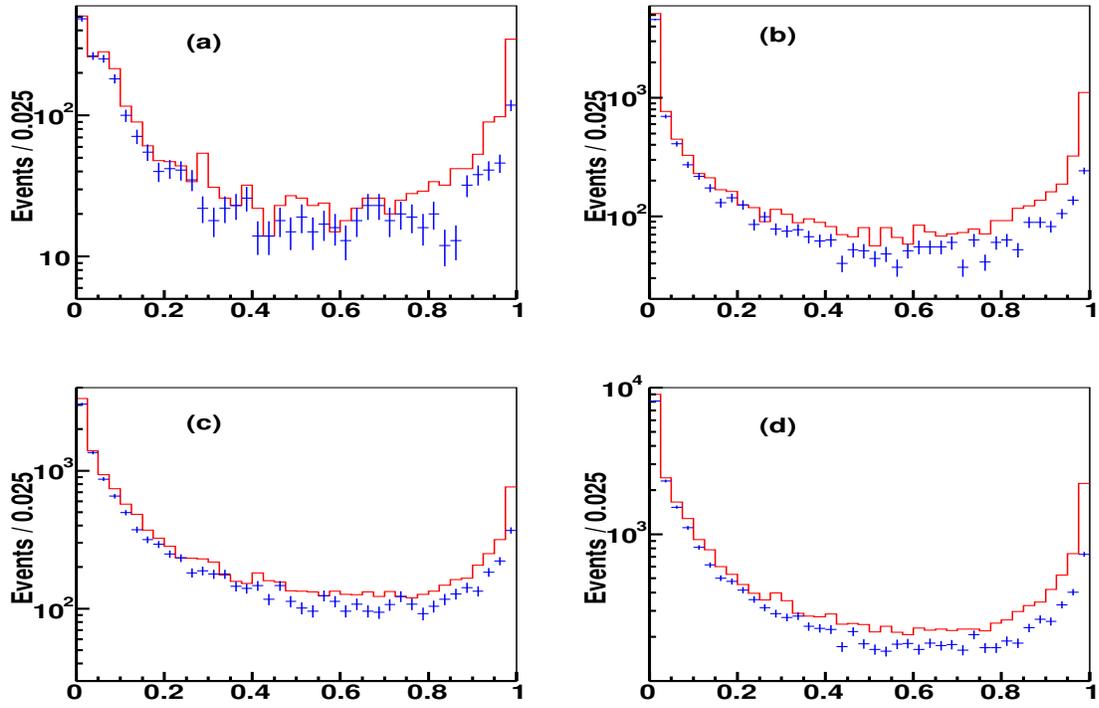


Figure 24:  $NN_{II}$  distributions of  $\tau$  candidates in OS(red) SS(blue) isolated  $\mu$  events. (a)  $\tau$ -type 1, (b)  $\tau$ -type 2, (c)  $\tau$ -type 3, (d) sum over  $\tau$ -types.

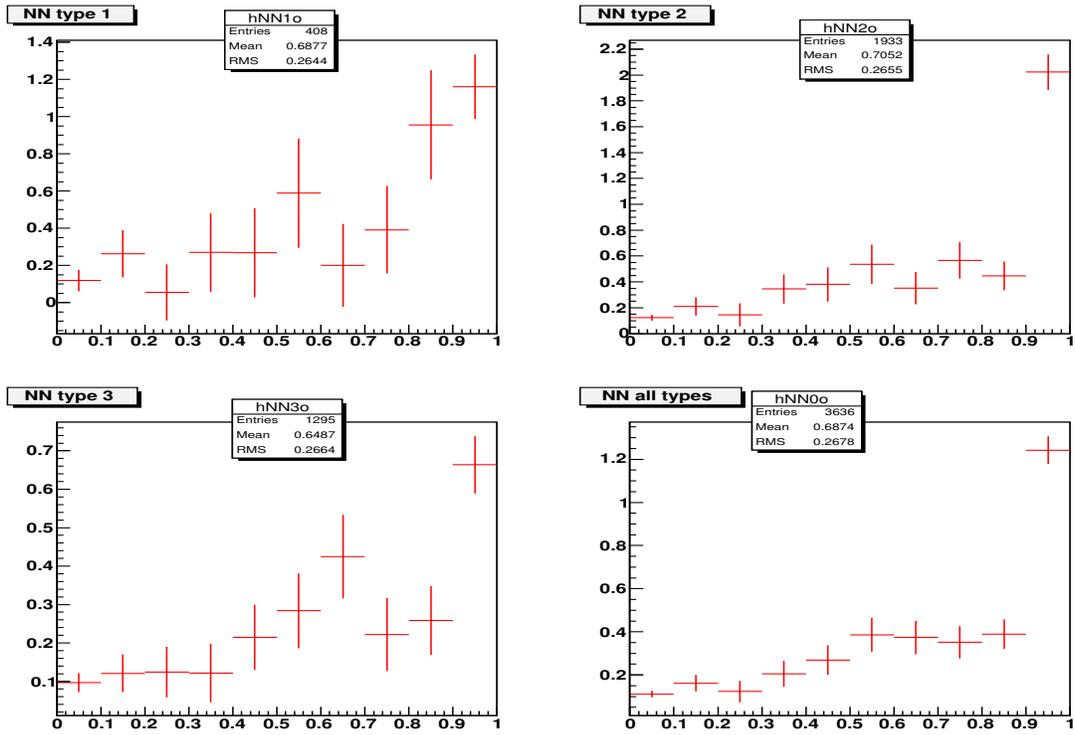


Figure 25: Ratio OS-SS/SS as function of  $NN_{II}$  for isolated  $\mu$ 's

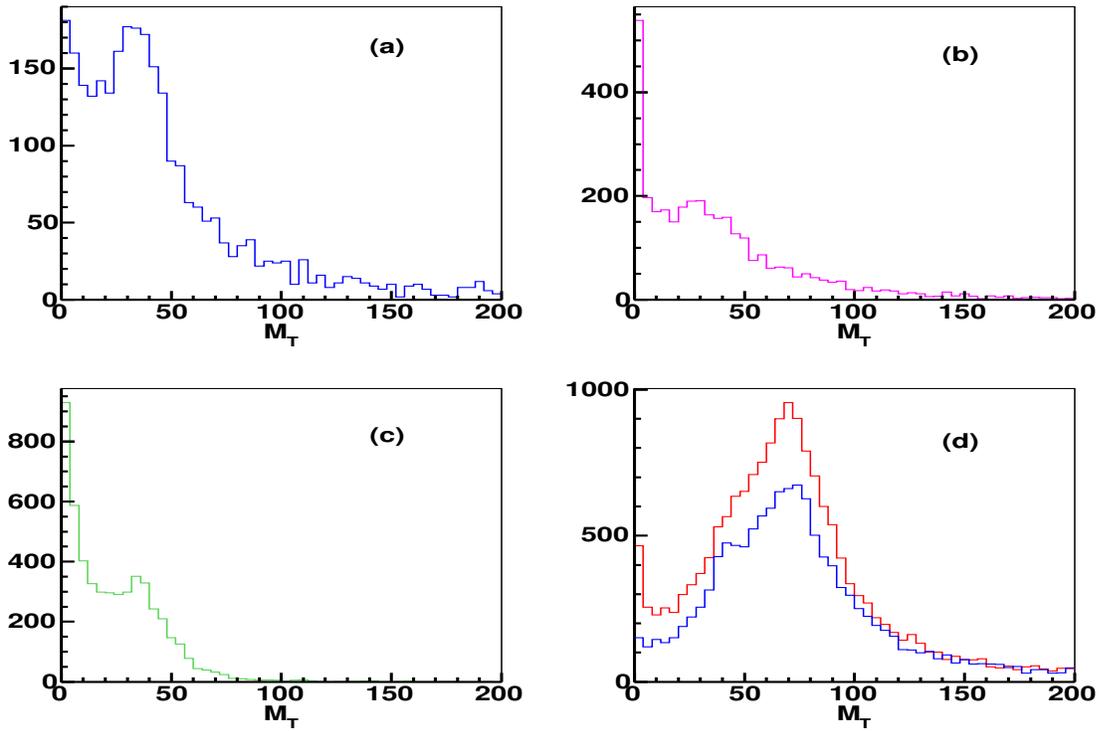


Figure 26: Transverse mass distributions for: (a) anti-isolated  $\mu$ , (b) two OS  $\mu$ 's with one isolated  $\mu$  and a  $\mu\tau$  overlap, (c) MC  $Z \rightarrow \tau^+\tau^-$ , (d) OS (red) SS (blue) isolated  $\mu$  events. Events require  $p_T^\mu > 25$  GeV, no  $NN$  cut, no  $|\phi_\mu - \phi_{\tau au}|$  cut.

Another instrumental background to consider comes from  $\mu$ 's misidentified as  $\tau$ 's. We can study that background using events with  $\mu^+\mu^-$  pairs with one of the  $\mu$ 's reconstructed as a  $\tau$  candidate. This background contributes only to  $\tau$ -type 1 and  $\tau$ -type 2. The background can be reduced by cutting on the variable  $\mathcal{R}_{trk}^\tau = (E^\tau - E_{CH}^{trk})/p_T^{trk}$ , where  $E_{CH}^{trk}$  is the energy deposited in a window of  $5 \times 5$  towers (each tower of size  $\phi \times \eta=0.1 \times 0.1$ ) around the  $\tau$ -track in the coarse hadronic (CH) section of calorimeter. This variable turns out to be less effective in pass 2 data than in pass 1. The reason is that the T42 algorithm tends to remove energy from the coarse hadronic section of the calorimeter, where on average a  $\mu$  will deposit most of the energy it loses. Figure 27 (a) show the distribution for that variable in pass 1 and pass 2 data. A cut of  $\mathcal{R}_{trk}^\tau > 0.7$  removes only 50% of this background in pass 2 as opposed to 70% in pass 1. Note that the  $\tau$ 's from misidentified  $\mu$  tend to have lower  $E_T^\tau$  in pass 2 than in pass 1 data, see figure 27 (b). The  $NN$  distributions ( $NN_{II}$ ) for these events are given in figure 28. Table 5 gives the number of events before and after  $NN > 0.9$  cut. There are more events of  $\tau$ -type 2 than  $\tau$ -type 1. Figure 27 (c) gives the mass calculated using the  $\mu$  and the  $\tau$  track. The distribution peaks around 60 GeV, although there are also more events than in pass 1 near the Z mass. Most of  $\tau$ -type 2 events are muons with a hard brehmstrahlung. We can recalculate the mass adding to the  $\tau$  track the energy in the EM calorimeter. The resulting distribution in 27 (d) shows a clear Z peak. One can remove close to 50% of this remaining background by requiring  $80 < M_{\mu, \tau} < 100$  GeV.

Table 5: Number of  $\mu^+\mu^-$  events

	$\tau$ -type 1	$\tau$ -type 2	sum
$NN > 0.0$	812	1947	2759
$NN > 0.9$	407	605	1012
$NN > 0.9$ & $80 < M(\mu, \tau) < 100$ GeV	258	296	554

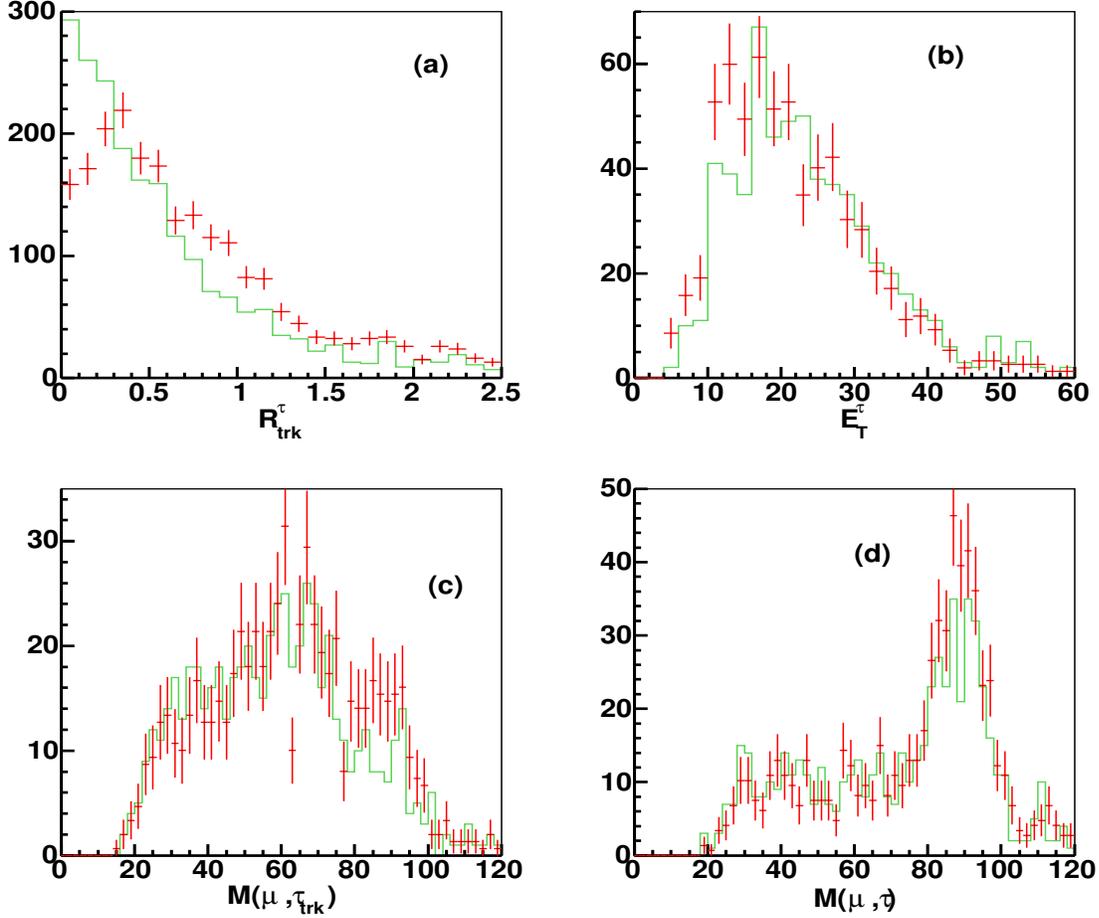


Figure 27: Distributions from events with misidentified  $\mu$ 's for pass 2 (red points) and pass 1 (green histogram) (a)  $\mathcal{R}_{trk}^\tau$  with no  $NN$  cut, (b)  $E_T^\tau$  after  $NN > 0.9$  cut, (c)  $M(\mu, \tau_{trk})$  after  $NN > 0.9$  cut, (d)  $M(\mu, \tau)$  after  $NN > 0.9$  cut.

## 9 Appendix

A set of utility functions for  $\tau$  root tree objects are provided in the `tau_cand cvs` module. Version `p16-br-03` is the most up-to-date version at this time. All general utility functions are in `tau_utils.C (.h)`.  $NN$  data files for all versions are also available (`*.dat`). Instructions on how to use the utilities are in `tau_cand/README.txt`. and `tau_utils.h`

### 9.1 $NN$ calculation

To calculate  $NN$  one needs to create an instance of `TauNNOutput` for each tau type supplying the wished for data file as argument. Below is an example of how to calculate  $NN$  for the new  $NN$  trained including electrons:

- Before event loop:

```
TauNNOutput NN1('NN_type1_mcm_pass2_with_e.dat');
TauNNOutput NN2('NN_type2_mcm_pass2_with_e.dat');
TauNNOutput NN3('NN_type3_mcm_pass2_with_e.dat');
```

- In event loop (prepare input variables and get output):

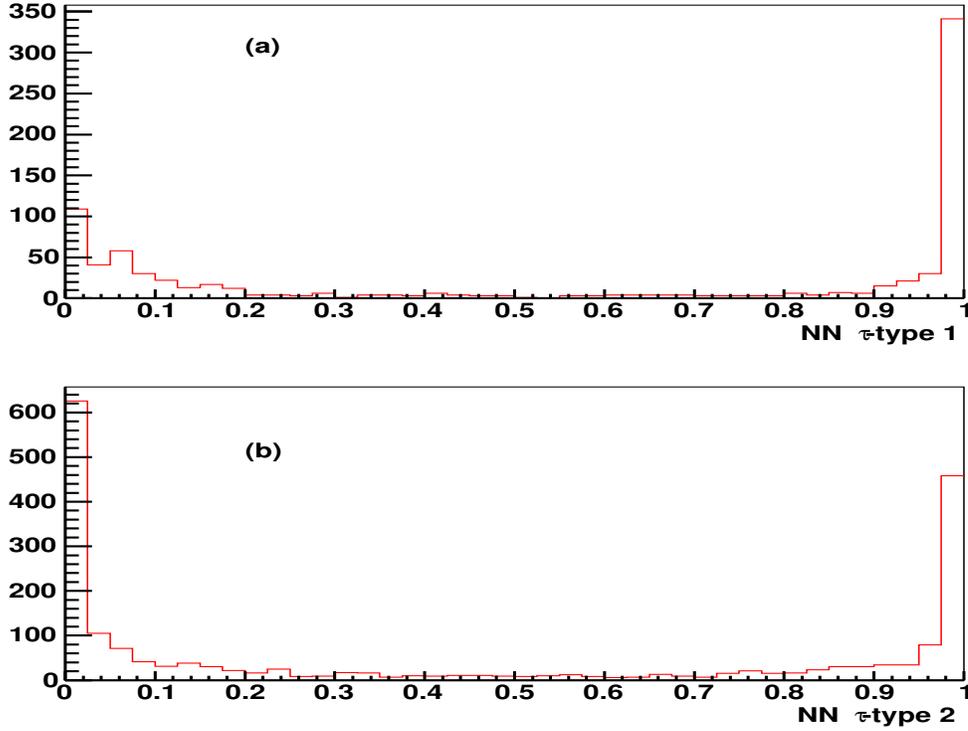


Figure 28: NN distributions from events with misidentified  $\mu$ 's for pass 2 events.

```
float nn;
if(tau->type()==i){
  NNi.setInputTypei(tau);
  nn=NNi.output();
}
```

where  $\tau$  is a TMBTaus pointer and  $i=1,2,3$ .

## 9.2 Calculating NN variables

In this section we give explicit instructions on how to calculate the NN input variables defined in section III using TMBTaus methods. Given pointer `const TMBTaus* tau` each input variable is given by

- Variables used for both NN's:
  1.  $profile = \tau \rightarrow profile()$ ;
  2.  $caliso = \tau \rightarrow iso()$ ;
  3.  $trkiso = \tau \rightarrow ettr()$   
 $/( \tau \rightarrow ettr() + \tau \rightarrow ett1() + \tau \rightarrow ett2() + \tau \rightarrow ett3() )$ ;
  4.  $em12isof = \tau \rightarrow EM12isof()$ ;
  5.  $\delta\alpha = \tau \rightarrow dalpha() / 3.1416$ ;
- Variables used for NN<sub>I</sub> only:
  6.  $e_{12} = \tau \rightarrow ett1() / \tau \rightarrow pT()$ ;
  7.  $p_T^{\tau_{trk1}} / E_T^\tau = \tau \rightarrow ett1() / \tau \rightarrow pT()$ ;

$$8. p_T^{\tau_{trk1}} / (E_T^\tau \cdot caliso) = \text{tau} \rightarrow \text{ett1}() / \text{tau} \rightarrow \text{Et\_iso}();$$

- Variables used for NN<sub>II</sub> only:

$$7. rms_\tau = \text{tau} \rightarrow \text{rms}();$$

$$8. fhf = 1 - (\text{tau} \rightarrow \text{emf}() + \text{tau} \rightarrow \text{chf}() / \text{tau} \rightarrow \text{pT}());$$

$$9. E_T^{em} / E_T^{tau} = \text{tau} \rightarrow \text{empt}() / \text{tau} \rightarrow \text{pT}();$$

$$10. prf3 = \text{tau} \rightarrow \text{emcl\_et1}() / \text{tau} \rightarrow \text{EM3\_Et\_iso}();$$

$$11. E_T^\tau / (E_T^\tau + \Sigma p_T^{\tau_{trk}}) = \text{tau} \rightarrow \text{pT}() / (\text{tau} \rightarrow \text{pT}() + \text{tau} \rightarrow \text{ett1}() + \text{tau} \rightarrow \text{ett2}() + \text{tau} \rightarrow \text{ett3}());$$