Identification of $\tau \to 3-prong$ Decays Using BackPropagation Neural Network

G.M.Chen H.J.Yang

L3 Beijing Group

Institute of High Energy Physics

Beijing 918(2),100039

P.R.China

Abstract

A serial of Neural Networks are constructed to reject γ conversion and to identify $\tau^{\pm} \to 3h^{\pm}n\pi^{0}\nu(n=0,1,\geq 2)$ decay channels using 94' M.C. $e^{+}e^{-} \to \tau^{+}\tau^{-}$ events. Purity of the three channels is 87.6%, 71.5% and 50.6% respectively. It should be noted that none $\tau\tau$ backgrounds are not taken into account in the note.

(L3 Note)

1. Introduction

In this note, we describe our attempts to reject γ conversion and to identify $\tau^{\pm} \rightarrow 3h^{\pm}n\pi^{0}\nu(n = 0, 1, \geq 2)$ decay channels using a serial of BackPropagation Neural Networks, where $h^{\pm}$ is either a $\pi^{\pm}$ or $K^{\pm}$ meson. Such efforts are motivated to measure branching ratios of $\tau^{\pm} \rightarrow 3h^{\pm}n\pi^{0}\nu(n = 0, 1, \geq 2)$ decays, the mass of $a_1(M_{a_1})$ and its width Γ_{a_1} etc..

132K M.C. $e^{+}e^{-} \rightarrow \tau^{+}\tau^{-}$ events generated by KORALZ 4.0, simulated by GEANT 3.5, are used to do all analysis. Real detector simulation is based on $94'$ database, and AFLW3P program developed by Kunin is used to reconstruct all M.C events.

2. Rejecting γ conversion events

M.C. studies indicate that the 3-prong sample has a considerable contamination from 1-prong decays, background from 5-prong decays is negligible, see $Table1(a)$,where only $N_{trk} = 3$ is required. One sees that 14.4% of the sample come from 1-prong decay due to γ conversion.

Source (τ decays)	Purity(%)
$\tau \rightarrow$ 3-prong	84.68
$\tau \rightarrow$ 1-prong with $K_s^0 \rightarrow \pi^+\pi^-$ vertex	0.90
$\tau \rightarrow$ 1-prong with $\pi^0 \rightarrow e^+e^-\gamma$ decays	13.58
$\tau \rightarrow$ 1-prong with radiative $\gamma \rightarrow$ e^+e^- conversion	0.80
$\tau \rightarrow$ 5-prong with 2 tracks missing	0.04
Total	100

$Table1(a)$: Purities of τ decay channels where $N_{trk} = 3$ is required.

As to identify γ conversion , the most useful information would be dE/dx and invariant mass of the two tracks. Unfortunately, there is no good dE/dx information of tracks in L3 detector. Tracks from γ conversion are e^+e^-, their momentums are rather low, and their invariant masses are very small,half of them is less than 60MeV. If we use simple cuts to identify γ conversion, the rejection ratio is rather poor. So we decided to use BackPropagation Neural Network technique to identify them.

A BackPropagation Neural Network [1] is constructed with one input layer, one hidden layer and one output layer. The output layer has only one node with output value from 0 (background) to 1 (signal). The input layer contains 12 neurons, each of them corresponds to a measured variable of a τ jet. The following variables are chosen as input of Neural Network:

1) N_{ng}: The number of neutral-gamma particle in the cone.

2) N_{nk}: The number of neutral-K0 particle in the cone.

3) N_{bmp}: The number of bumps in the cone.

4) E_{bgo}: The energy deposited in the BGO.

5) E_{09}/E_{25}: The ratio of energy sum of 9 crystals over 25 crystals in the most energetic bump, it shows the transversal distribution of the shower in the BGO.

6) $E_{hcal}/(E_{hcal}+E_{bgo})$: E_{hcal} is the energy deposited in HCAL. The variable show the longitudinal development of the shower.

7-8) $P_1,\ P_1+P_2$: P_1 is the momentum of track 1 in the TEC.

9-11) $M_3,\ M_2+M_3\ or\ M_1+M_2+M_3$: M_3 is the smallest invariant masses of the third track with other tracks.

12) $(E_{bgo}-E_1)/(P_2+P_3)$: E_1 is the energy of track 1.

Fig.1 show the all twelve variables. All these variables are normalized to be within the range [0,1] .

The network is trained by 5000 $\tau \rightarrow$ 3-prong events as signal and $\tau \rightarrow \pi ee$ as background where e^+e^- come from γ conversion. The “temperature” is set to be 1, the “momentum” α chosen to be 0.07 and the learning strength parameter η decreases from 0.03 to 0.001 during the training process. The initial values of weight W and threshold C are set randomly between [-0.1, 0.1]. Overtraining may bias for training sample but not for test sample. After attempts and comparison, one million loops is decided to train the network.

The trained network is tested by another M.C sample called test sample, which is statistically independent from the training sample. Output (y) of network is shown in Fig.2(a). The selection efficiency is defined as the number of selected signal events divided by the total number of signal when N_{trk}=3. The purity is defined as the number of selected signal (bg) events divided by the total number of selected candidates with $y > Y_{cut}$. The selection efficiency and purity as the function of $Y_{cut}(y > Y_{cut})$ are shown in Fig.2(b). Fig.2(c,d) represents the selection efficiency as the function of momentum of tracks. The purity of selected events are listed in $Table1(b)$ with $Y_{cut} = 0.3$. Here the signal lose 6%, about 80% of bg is rejected, it is much more effective than simple-cut method. After γ conversion rejection, the background from τ sources is reduced to 3.64% .

Source (τ decays)	Purity(%)
$\tau \to$ 3-prong	96.36
$\tau \to$ 1-prong with $K_s^0 \to \pi^+\pi^-$ vertex	1.06
$\tau \to$ 1-prong with $\pi^0 \to e^+e^-\gamma$ decays	2.37
$\tau \to$ 1-prong with radiative $\gamma \to\ e^+e^-$ conversion	0.17
$\tau \to$ 5-prong with 2 tracks missing	0.04
Total	100

$Table1(b)$: Purities of τ decay channels after rejecting γ conversion events, where $N_{trk} = 3$ and $Y_{cut} = 0.3$.

3. Identification of $\tau^\pm \to 3h^\pm n\pi^0\nu$ (n=0,1,$\geq$ 2) decays

3.1 The number of neutrals

$Fig.3(a)$ shows the number of neutrals distribution of different $\tau \to 3 - prong$ decay channels. For $\tau^\pm \to 3h^\pm\nu$, neutrals should be 0 in theory, but $Fig.3(a)$ shows that it can be 0, 1, 2, 3, or more. Obviously most of them are junk which mainly from hadron shower split-off or shower fluctuation. Neutrals can be divided into three types roughly according to their coincidence level of energy and angle to π^0 and γ in generator level, they are π^0 like, γ like and junk like. When the cut $E_n > 0.5GeV$ and $|cos\theta| > 0.95$ are used , about 77% of junk like neutrals are rejected, at the same time 14% of π^0 and γ like neutrals are lost. $Fig.3(b)$ shows Neutrals distribution after rejecting junk like neutrals with the above two cuts.

3.2. Neural Network with N=0 (Number of Neutrals)

A neural network named n0p0 (where,n0 means number of neutrals is 0 , while p0 means 3h without π^0 is regarded as signal, others are background in the training process, the output of the neural network is yn0p0, the name regulation is the same throughout the note) is used to identify $\tau \to 3h\nu$ and $\tau \to 3h\pi^0\nu$, by taking $\tau \to 3h\nu$ as the signal and all others as background. The following variables are choesn as input of n0p0 network(see Fig.4):

1) $M_{3\pi}$: Invariant mass of 3 π.

2-3) $M^1_{2\pi}, M^2_{2\pi}$: Invariant mass of $\pi^+\pi^-$.

4) E_{bgo}: Energy in the BGO.

5) $E_{bgo}/(E_{hcal} + E_{bgo})$.

6-8) E_{bgo}/P: P means P_1, P_1+P_2 or P_1+P_2 +P_3, P_1 is the momentum of track 1.

The training condition of n0p0 is quite similar to the above. We obtain the network output(yn0p0) distribution as shown in Fig.9(a) .

3.3. Neural Network with N=1

Two neural networks named n1p0 and n1p2 are constructed when N=1. Totally 19 variables are selected as input of the two networks(see Fig.5):

1-8): The first 8 variables are the same as those of n0p0.

9-11)E_{tot}/P: E_{tot} means total energy in the cone.

12)$E_{tot} - (E_1 + E_2 + E_3)$: E_1 is the energy of particle 1.

13)E_{n1}: Energy of neutral 1.

14-16)$M_{h_1h_2n}$, $M_{h_1h_3n}$ or $M_{h_2h_3n}$: Invariant masses of two hadrons and neutral.

17)$M_{h_1h_2h_3n}$: Invariant mass of 3 hadrons and the neutral.

18)E_{n1}/E_{bgo}: E_{n1} is the energy of the neutral.

19)$E_{n1}/(P_1+P_2+P_3)$: E/P distribution of the neutral.

The ouput distributions of n1p0(yn1p0) and n1p2(yn1p2) are shown in Fig.9(b,c).

3.4. Neural Network with N=2

Similar to N=1, there are two networks(n2p0 and n2p2). The network n2p0 has 17 variables, while n2p2 only has 13 variables. The following show variables of n2p0:

1) $M_{3\pi}$: Invariant mass of 3 π.

2-3) $M_{2\pi}$, $M_{2\pi}^2$: Invariant mass of $\pi^+\pi^-$.

4) E_{bgo}: Energy in BGO.

5) $E_{bgo}/(E_{hcal}+E_{bgo})$.

6-8) E_{bgo}/P: P means P_1, P_1+P_2 or P_1+P_2 +P_3, P_1 is the momentum of track 1.

9)M_{n12}: Invariant mass of two neutrals.

10-11)$E_{n1}, E_{n1}+E_{n2}$: Energy of neutrals.

12-17) variables are similar to (14-19) when N=1.

Fig.6 shows distribution of these variables.

Variables of n2p2 network are listed below(see Fig.7):

1) E_{bgo}: Energy in BGO.

2-4) E_{bgo}/P: P means P_1, P_1+P_2 or P_1+P_2 +P_3, P_1 is the momentum of track 1.

5-7)E_{tot}/P: E_{tot} means total energy in the cone.

8)$E_{tot}-(E_1+E_2+E_3)$: E_1 is the energy of particle 1.

9)M_{n12}: Invariant mass of two neutrals.

10-11)$E_{n1}, E_{n1}+E_{n2}$: Energy of neutrals.

12)$E_{n1}/(P_1+P_2+P_3)$: the ratio of neutral energy to energy in BGO.

13)$(E_{n1}+E_{n2})/(P_1+P_2+P_3)$: E/P distribution of neutrals.

Each network is trained half million loops. Fig.9(d,c) show the output distributions of the above two networks.

3.5. Neural Network with N≥ 3

Two networks named n3p0 and n3p2 are constructed when $N \geq 3$ .Total 17 variables are chosen as networks input(see Fig.8):

1) $M_{3\pi}$: Invariant mass of 3 hadrons.

2) E_{bgo}: Energy in BGO.

3-5) E_{bgo}/P: P means P_1, P_1+P_2 or P_1+P_2 +P_3, P_1 is momentum of track 1.

6-8)E_{tot}/P: E_{tot} means total energy in one cone.

9-11)M_{n12}, M_{n13} and M_{n23}: Invariant masses of two neutrals.

12-14)E_n: E_n means $E_{n1}, E_{n1} + E_{n2}$ or $E_{n1} + E_{n2} + E_{n3}$, energy of the neutrals.

15-17)$E_n/(P_1 + P_2 + P_3)$: E/P distribution of neutrals.

The two networks are trained 0.1 million loops , The output distributions are shown in Fig.9(f,g).

3.6. Summary

To identify $\tau^{\pm} \rightarrow 3h^{\pm}n\pi^0\nu$ (n=0,1,≥ 2), we use totally 7 networks according to the number of neutrals. The selection of network outputs are listed in $Table2(a)$, where networks can be used to select signal or to reject background, for example, in order to select $\tau \rightarrow 3h\pi^0\nu$ decay , we use $yn2p0 < 0.5$ to reject $\tau \rightarrow 3h\nu$ and $yn2p2 < 0.6$ to reject $\tau \rightarrow 3h \geq 2\pi^0\nu$, so the rest is $\tau \rightarrow 3h\pi^0\nu$. The relative selection efficiency and purity of different channels are shown in $Table2(b)$, where branching ratios is assumed by KORALZ . Having the above selection, we obtain invariant masses spectrum of $\tau^{\pm} \rightarrow 3h^{\pm}n\pi^0\nu$ (n=0,1,≥ 2), see Fig.10 .

Decay modes	N=0	N=1	N=2	N≥3
$\tau \rightarrow\ 3h\nu$	$yn0p0 > 0.2$	$yn1p0 > 0.7$	$yn2p0 > 0.5$	
$\tau \rightarrow\ 3h\pi^0\nu$		$yn1p0 < 0.3$ $yn1p2 < 0.95$	$yn2p0 < 0.5$ $yn2p2 < 0.6$	$yn3p0 < 0.4$ $yn3p2 < 0.6$
$\tau \rightarrow\ 3h \geq 2\pi^0\nu$			$yn2p2 > 0.85$	$yn3p2 > 0.85$

$Table2(a)$: The cuts of network outputs for selecting $\tau^\pm \rightarrow 3h^\pm n\pi^0\nu$(n=0,1,$\geq$ 2) decay channels.

τ decay modes	$3h\nu$		$3h\pi^0\nu$		$3h \geq 2\pi^0\nu$	
	efficiency(%)	purity(%)	efficiency(%)	purity(%)	efficiency(%)	purity(%)
$\tau \rightarrow$ 3hν	**83.2**	**87.6**	2.7	11.8	0.1	3.1
$\tau \rightarrow$ 3h$\pi^0\nu$	22.4	10.3	**37.8**	**71.5**	2.8	35.8
$\tau \rightarrow$ 3h$\geq 2\pi^0\nu$	5.2	0.4	26.3	9.2	**21.8**	**50.6**
$\tau \rightarrow$ 1-prong	5.4	1.6	6.2	7.4	1.3	10.2
$\tau \rightarrow$ 5-prong	33.3	0.02	16.7	0.05	16.7	0.3

$Table2(b)$: The relative selection efficiency and purity of different τ decay modes when identify $\tau^\pm \rightarrow 3h^\pm n\pi^0\nu$(n=0,1,$\geq$ 2)

4. Outlook

In this note we only use $\tau\tau$ M.C. events to do the analysis, a lot of things need to be done, all other backgrounds should be analysed in detail, real data in 1992-1995 should be reconstructed by AFLW3P and M.C. should be compared with real data. After then, B.R. of $\tau^\pm \rightarrow 3h^\pm n\pi^0\nu(n = 0, 1, \geq 2)$, M_{a1} and Γ_{a1} analysis can be carried out.

References

[1] For an introduction, see P.Wasserman, "Neural computing: Theory and Practice", Van Nostrand Reinhold, N.Y.(1989).

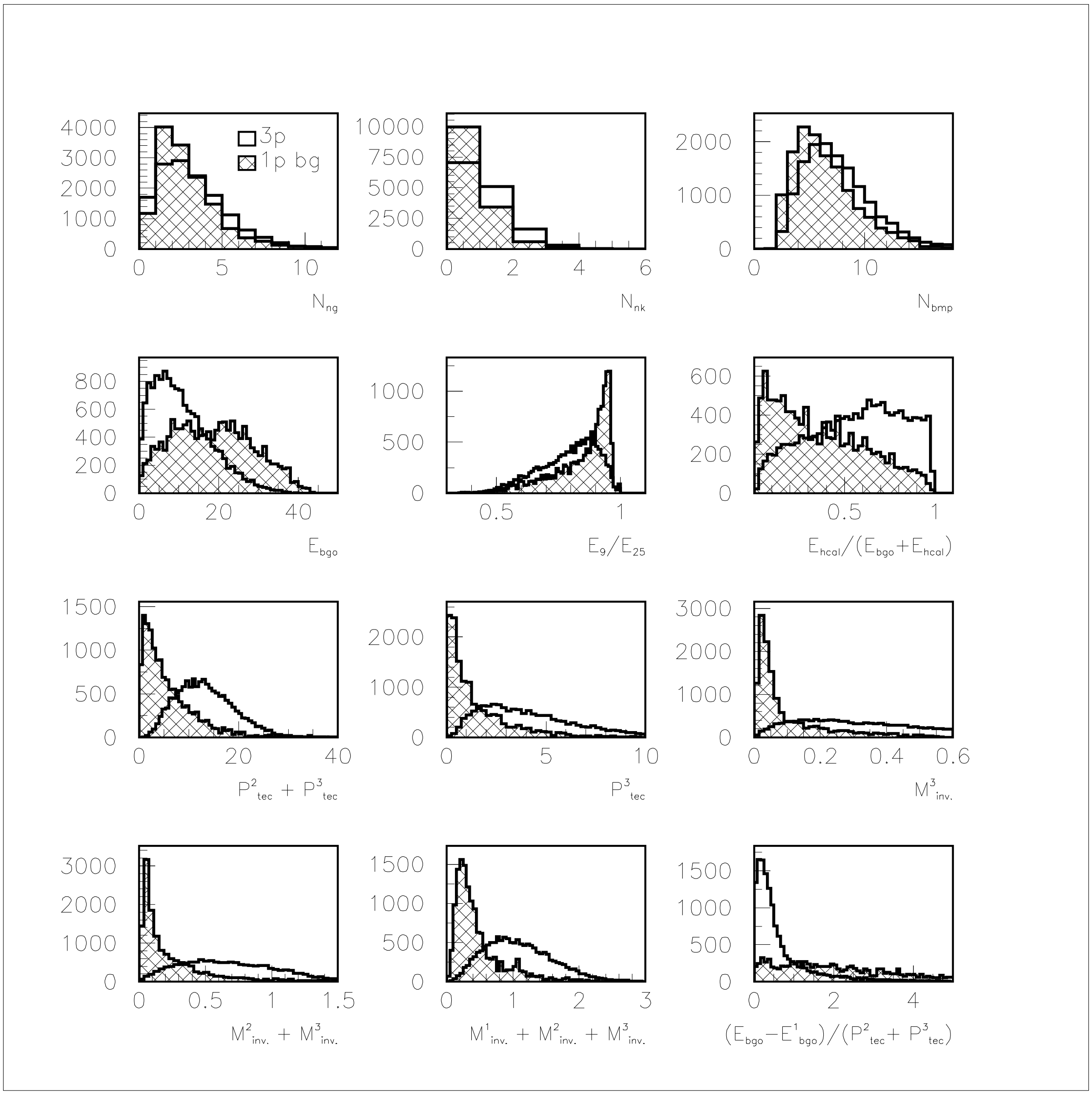

Figure 1: Input variables of network for rejecting γ conversion

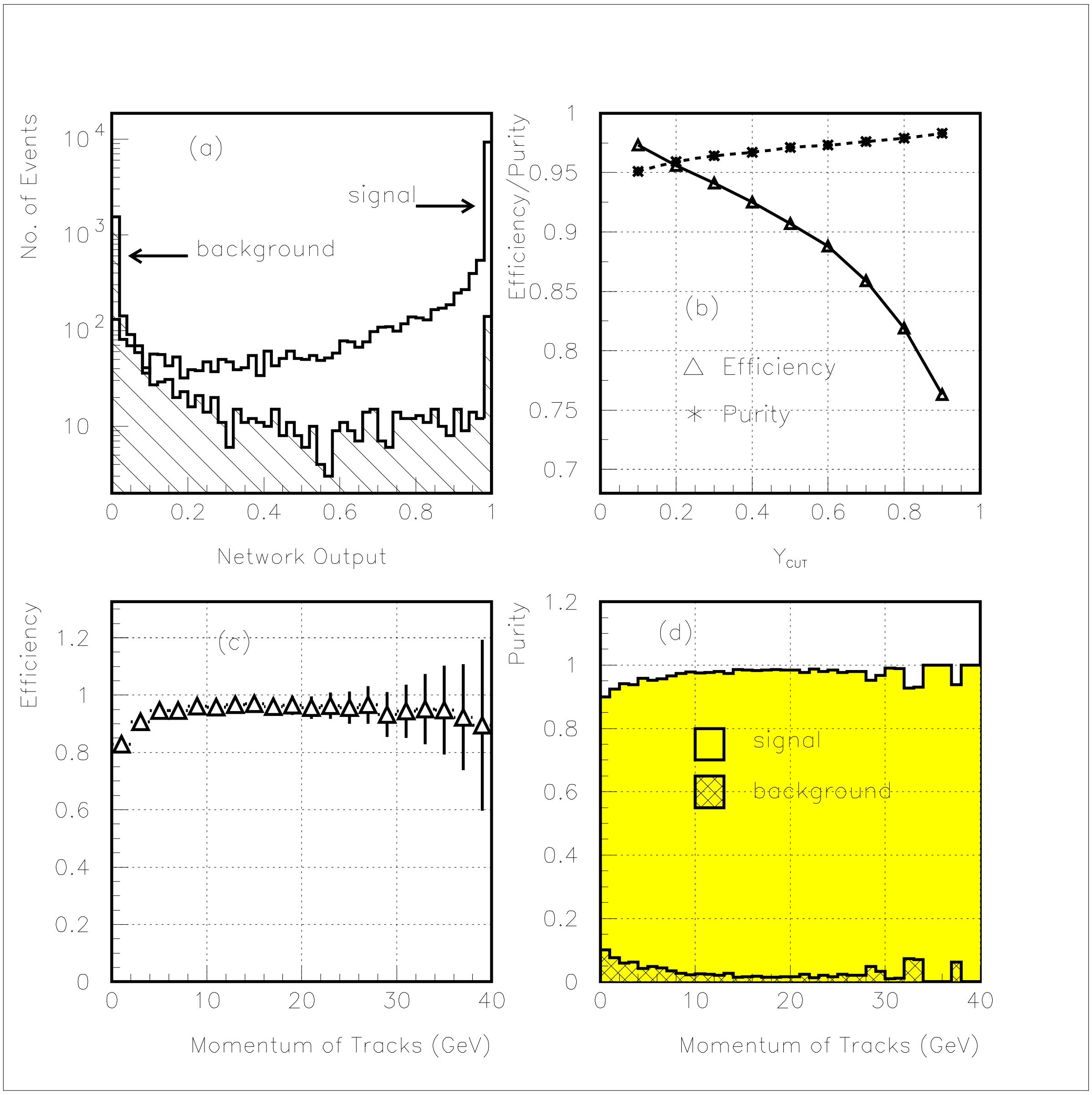
(a)
No. of Events
signal
background
Network Output
(b)
Efficiency/Purity
Efficiency
Purity
Y_{CUT}
(c)
Efficiency
Momentum of Tracks (GeV)
(d)
Purity
signal
background
Momentum of Tracks (GeV)

Figure 2:

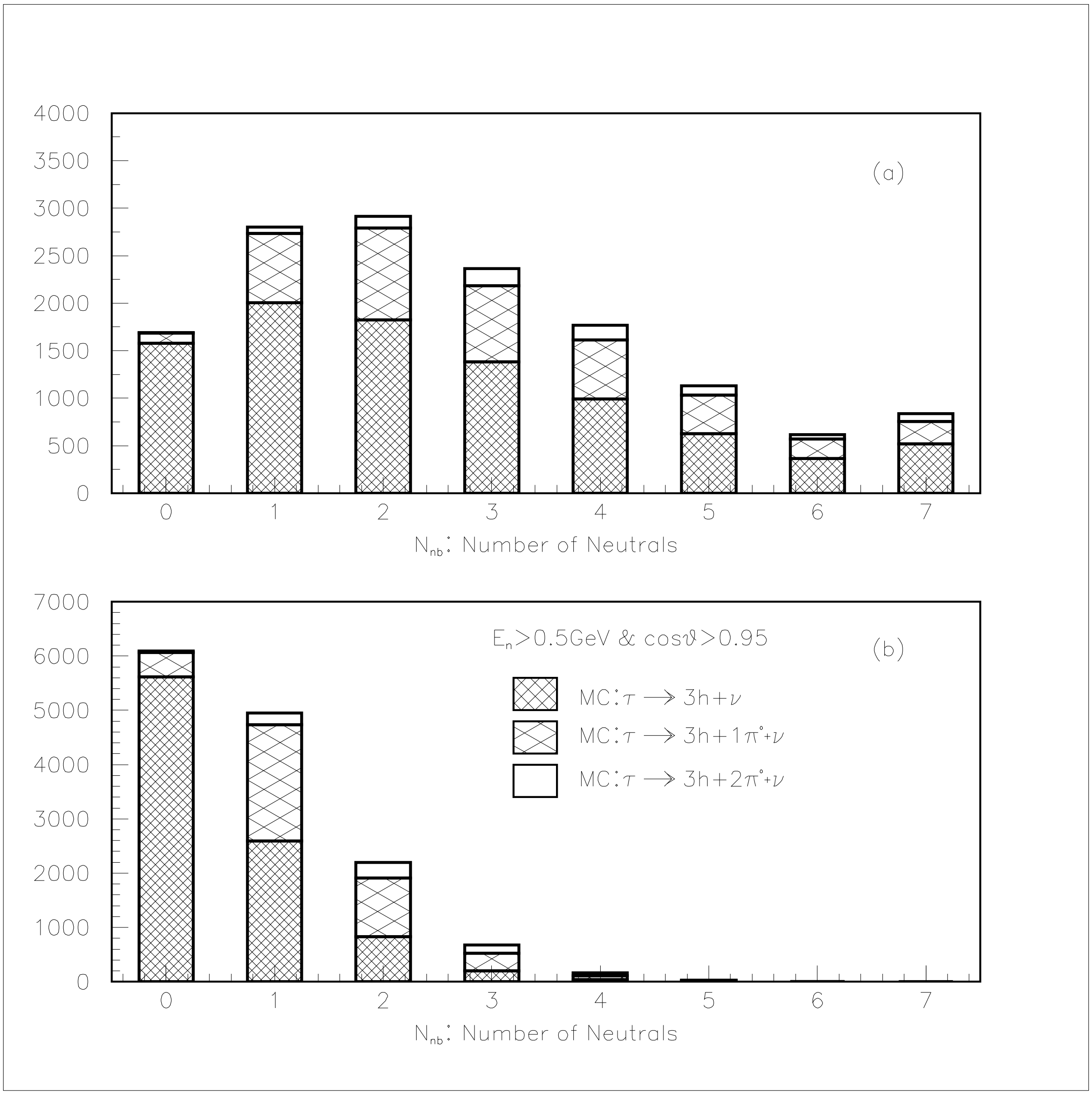

Figure 3: Number of neutrals

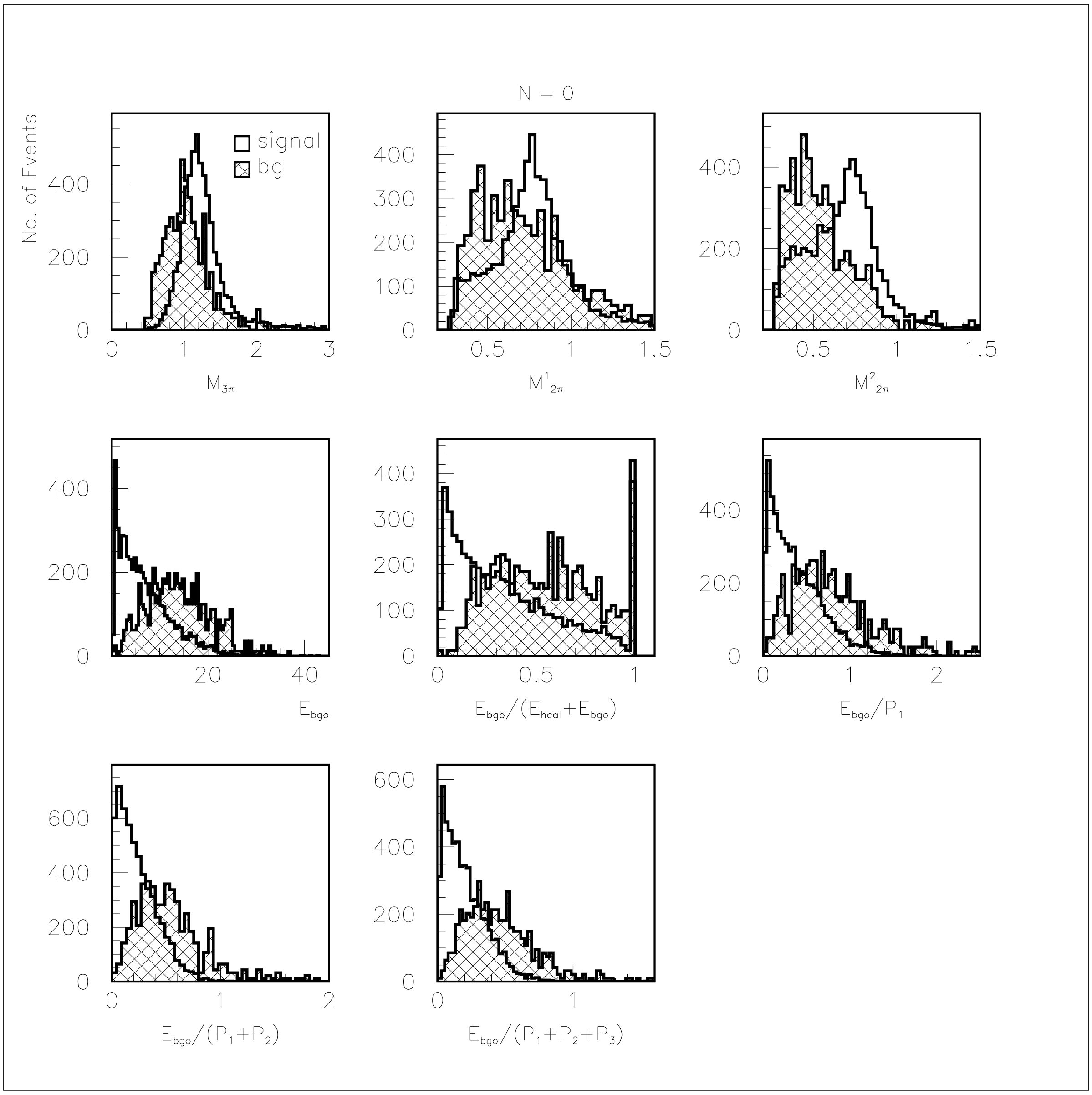

Figure 4: Input variables of network with N = 0

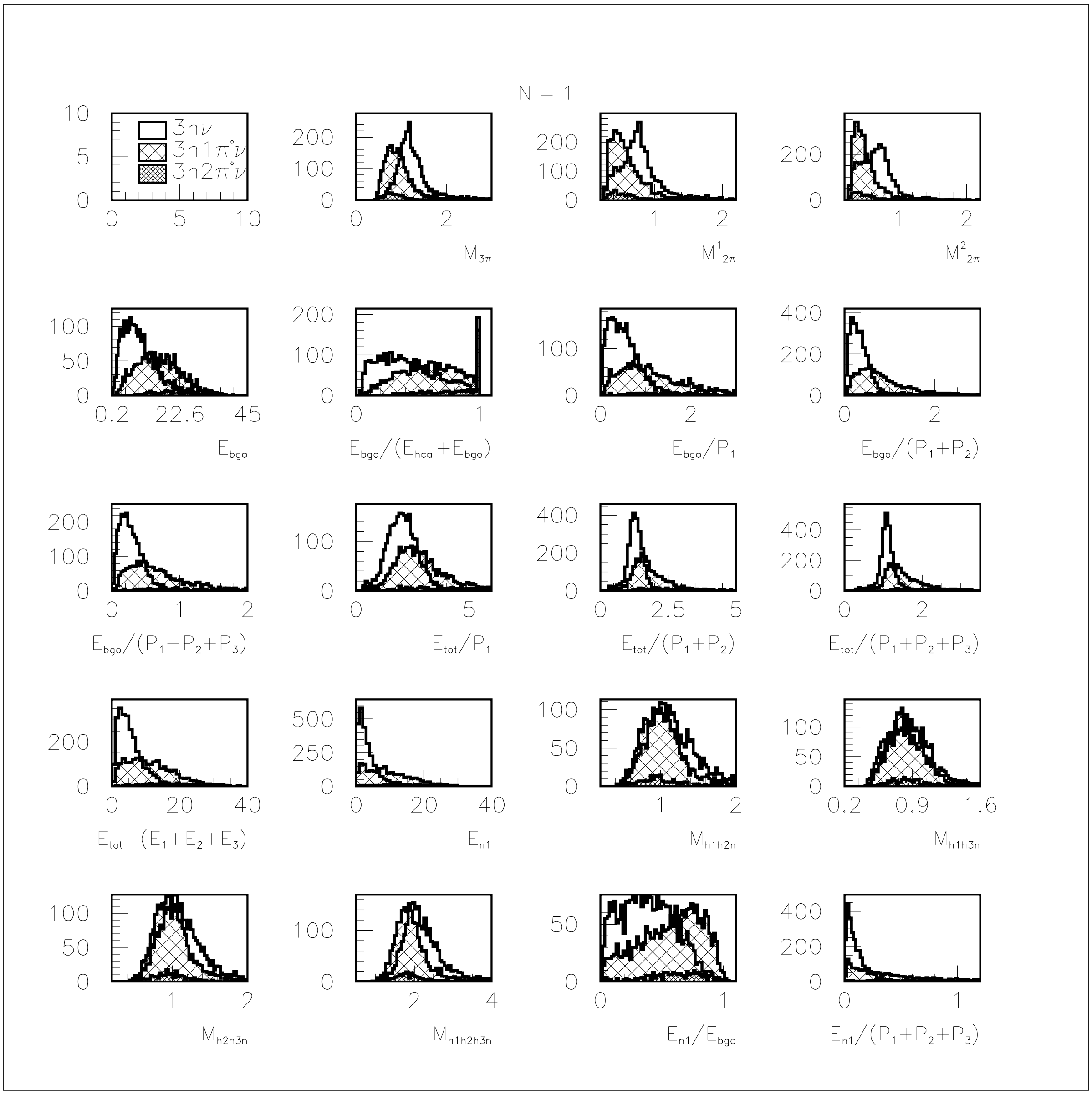

Figure 5: Input variables of networks with N = 1

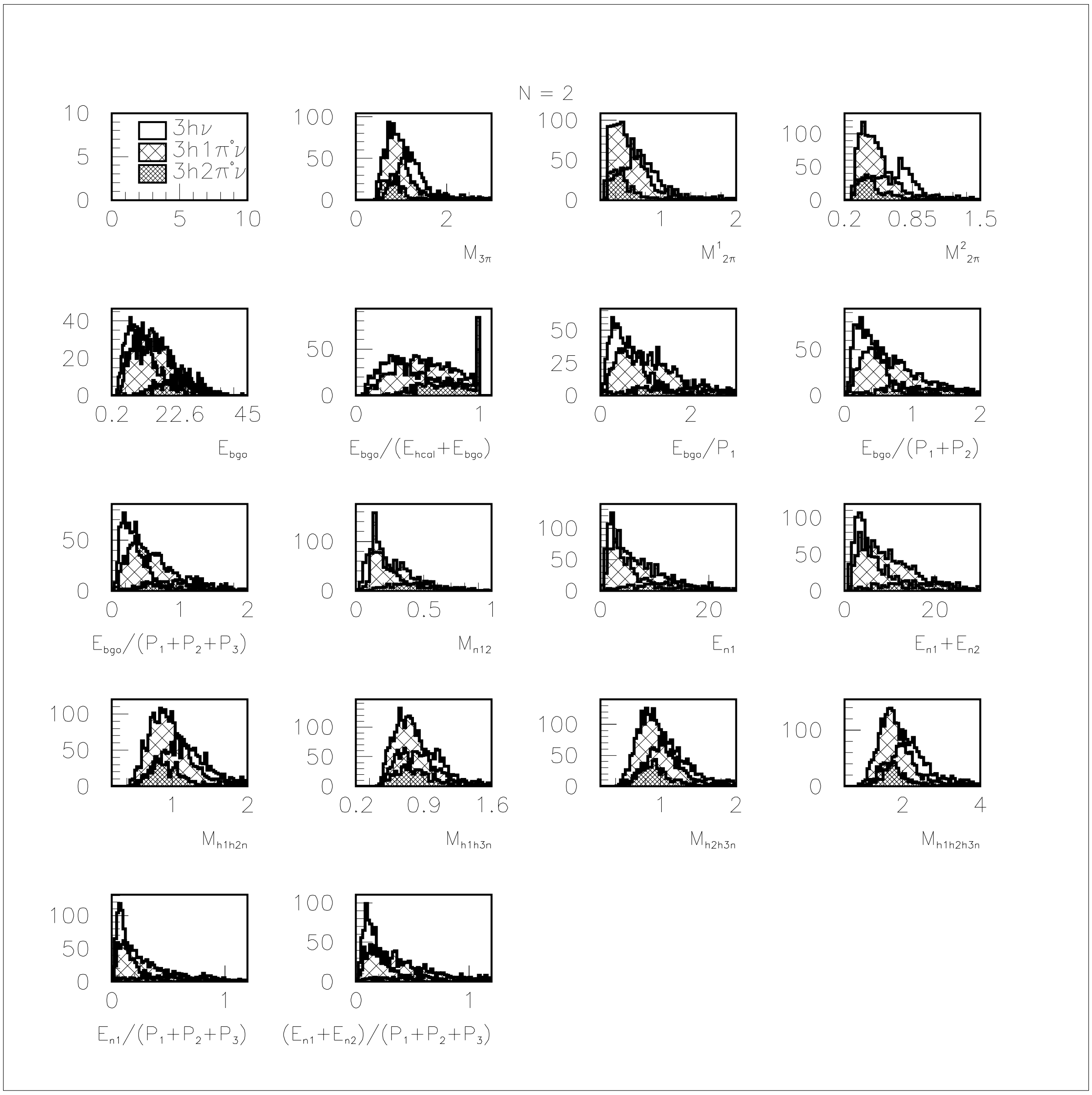

Figure 6: Input variables of network n2p0 with N = 2

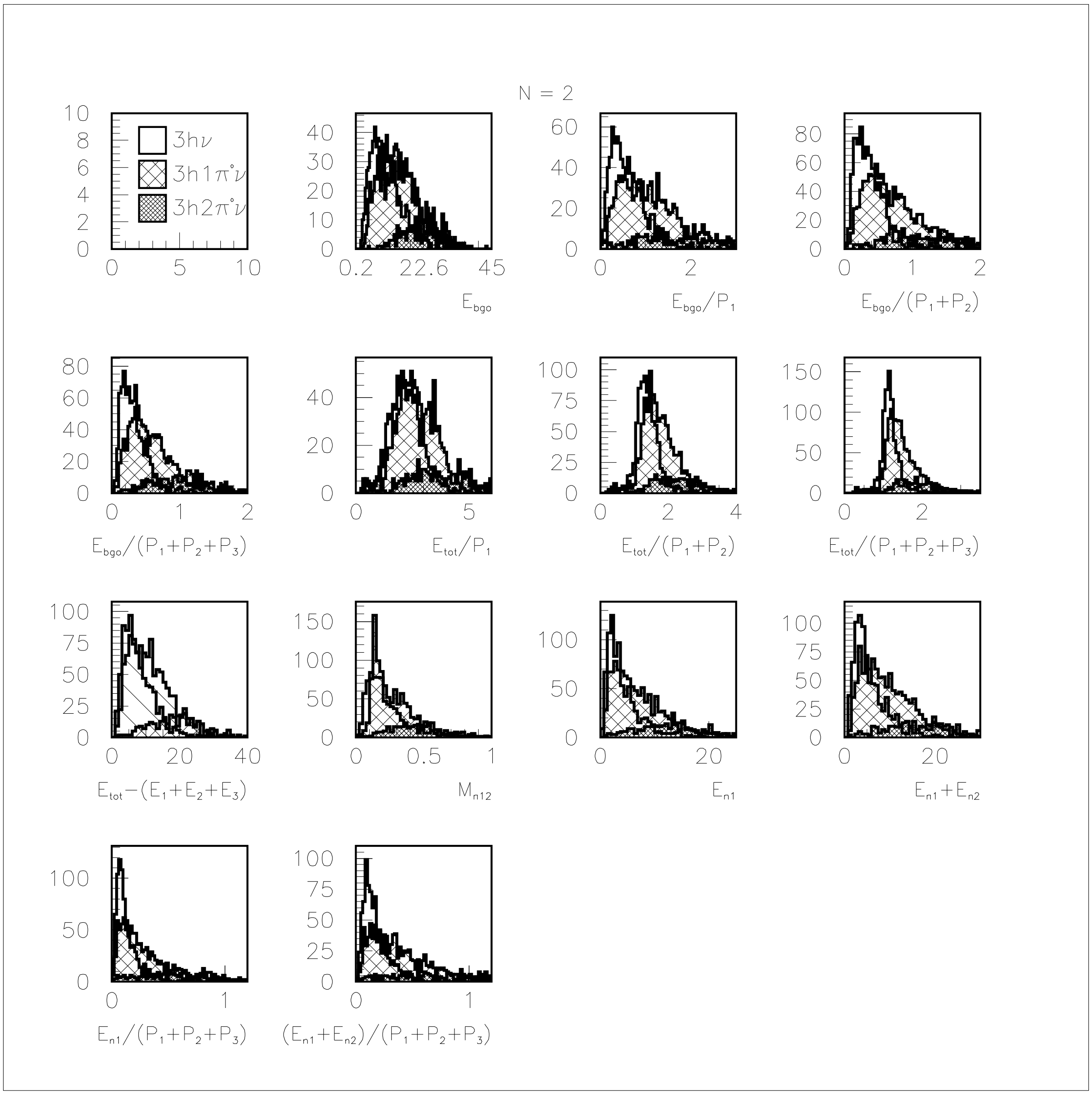

Figure 7: Input variables of network n2p2 with N = 2

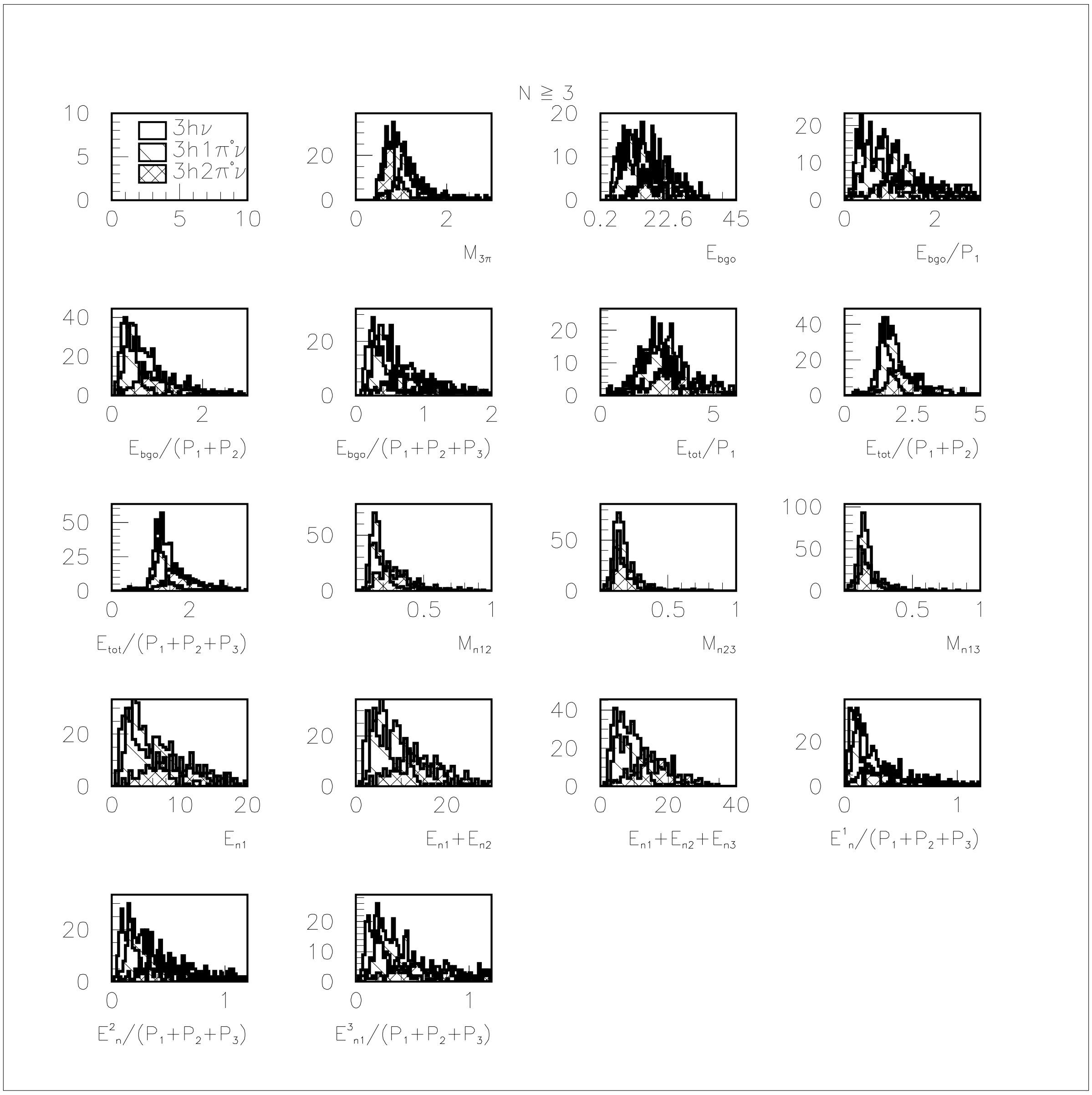

Figure 8: Input variables of networks with N≥ 3

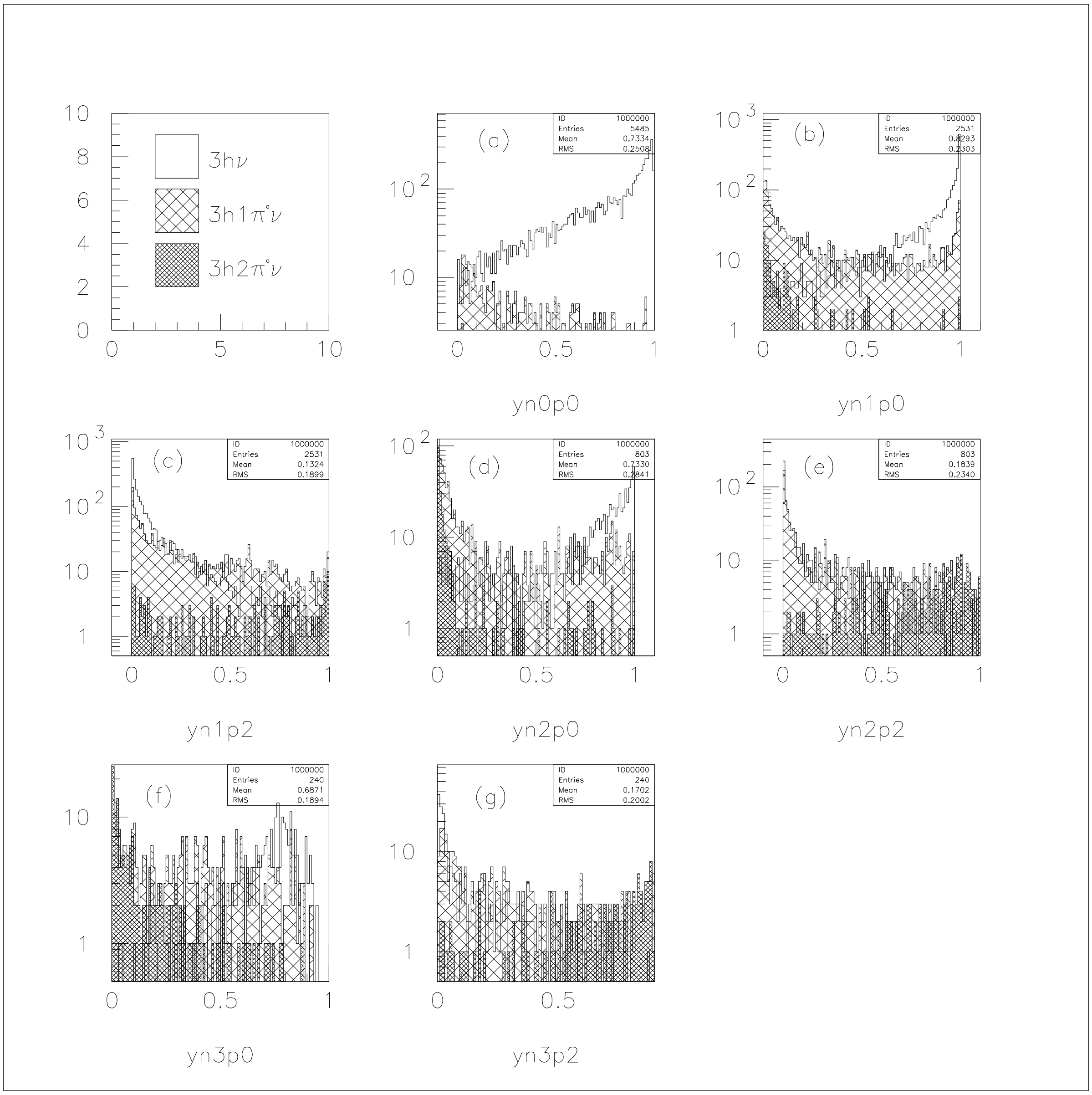

Figure 9: Output distributions of Neural Network

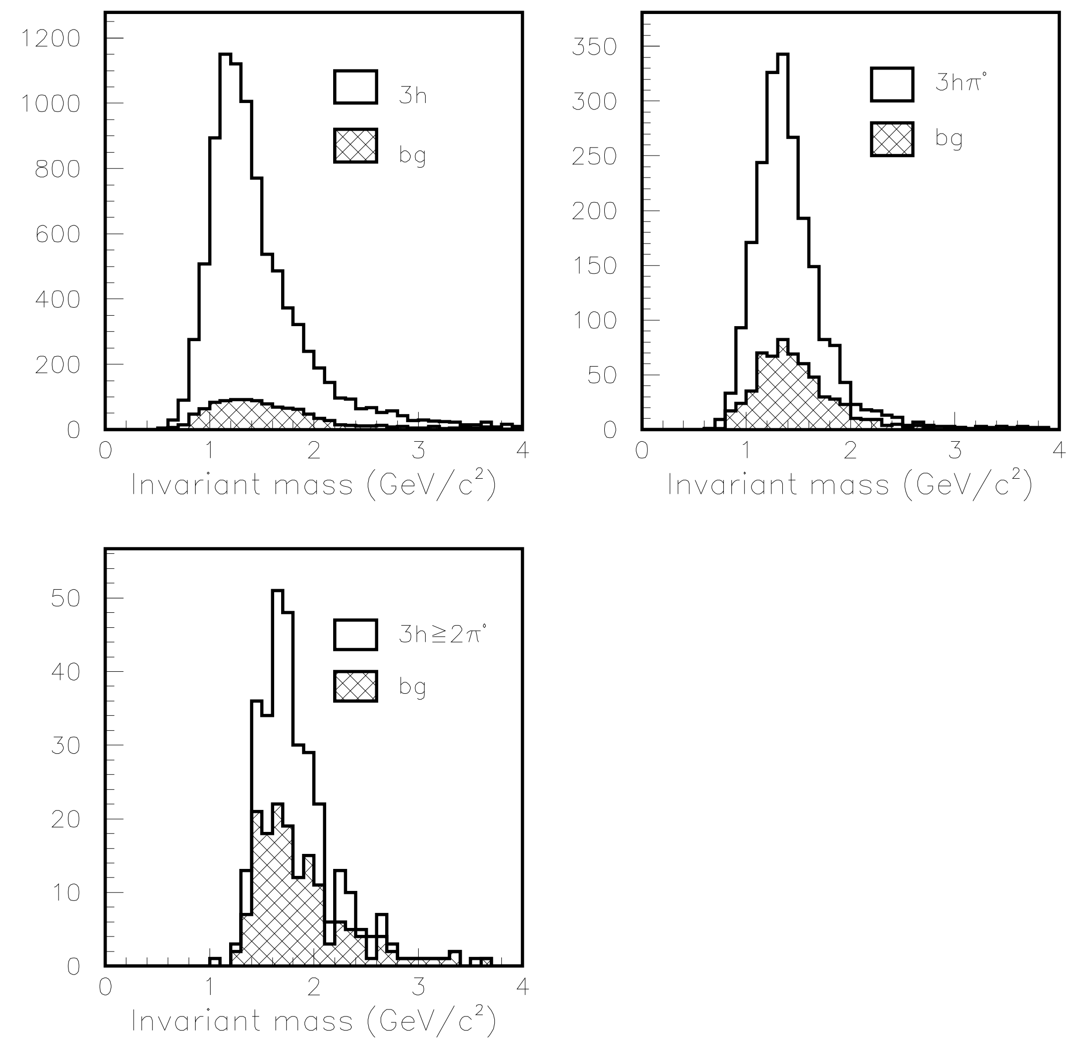

Figure 10: Invariant masses distribution of different $\tau \rightarrow$ 3-prong decay channels