Particle identification at DØ using Neural Networks

Dhiman Chakraborty

Supriya Jain

Vishnu Zutshi

ACAT2000

16-20 Oct, 2000, Fermilab

Introduction

- Experimental High Energy Physics = effective particle identification.
- Most interesting processes are buried in enormous backgrounds. At the crudest level, $S/B < 10^{-6}$ is quite common.
- Need "smart" discriminants to identify particles that feature in signal, but not in background.
- Electrons and tau leptons (heavier, short-lived, cousins of electrons) appear in the final states of many interesting processes in High Energy Physics.

Signal characteristics

- An electron is a stable charged particle whose
  - momentum is measured by plotting its trajectory in a magnetic field (central tracker).
  - energy is measured by stopping it using an instrumented absorber (calorimeter) where it deposits all of its energy in a short and narrow volume.
- A tau lepton is an unstable charged particle that decays before reaching any detector component.
  - The decay products consist of an odd number of charged particles that can be tracked.
  - One or more neutral daughters (neutrinos) leave no visible trace, but the energy of the rest is deposited in the calorimeter in a combined volume somewhat larger than that for an electron.

Background characteristics

- Jets of hadrons are the main background to electrons and taus. They tend to have
  - higher charged particle multiplicity,
  - deeper and wider volume of energy deposition in the calorimeter.
- However, the rate is so high that there are enough “outliers” that mimic the signal characteristics well enough to defeat a simple selection based on “cut”s on a few directly measured quantities.
- Multivariate tools that can exploit characteristic correlations among many variables to discriminate signal from background are needed.

Multivariate selection methods

- H-matrix

$$\chi^2 = \sum_{i,j} X_i H_{ij} X_j \tag{1}$$

where

$$H_{ij} = (\sigma^{-1})_{ij} \tag{2}$$

and

$$\sigma_{ij} = \langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle \tag{3}$$

- The MLPfit neural network

$$y = k + \sum_j p_j s(h_j) \tag{4}$$

with

$$s(x) = (1 + e^{-x})^{-1} \tag{5}$$

and

$$h_j = c_j + \sum_i w_{i,j} x_i \tag{6}$$

- An H-matrix is a simple linear transformation which is optimal for variables that follow multivariate Gaussian distribution.
- However, the observables available to us are asymmetrically distributed and bear non-linear correlations. Under such conditions, Neural Networks offer better separation between signal and background.
- MLPfit is a simple, flexible feed-forward neural network with an easy user-interface. (see `http://home.cern.ch/~schwind/MLPfit.html`)

Selection criteria for electron id:

- Narrow calorimeter cluster with large EM fraction and matching track.
- $|\eta| < 1.0$
- 10 GeV $< E <$ 100 GeV

Samples used for training:

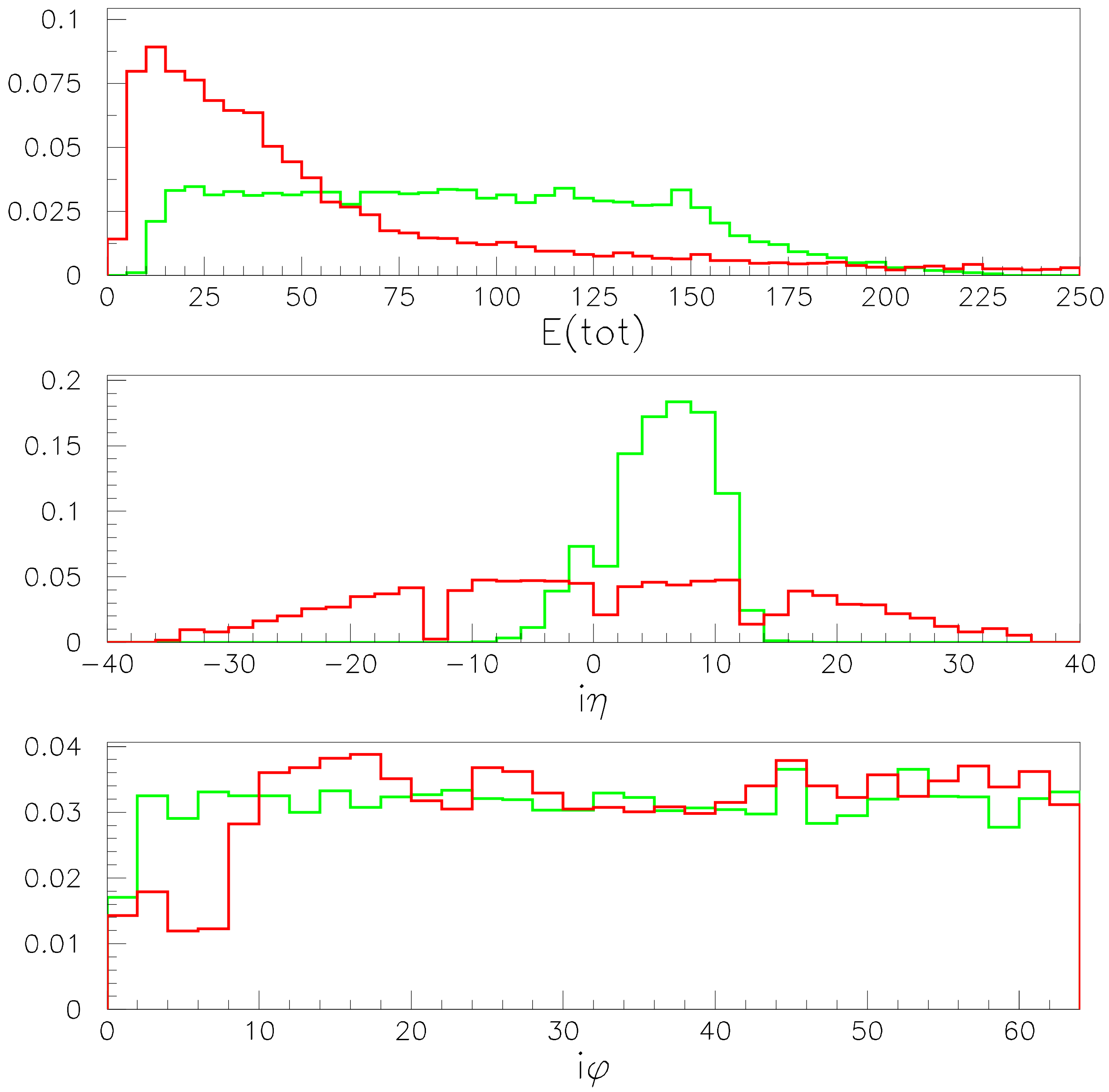
- Signal: 2600 single electrons, uniform in 10 GeV $< E_T <$ 150 GeV.
- Background: 1600 QCD $2 \to 2$ events (clumps with different E_T thresholds), π^0 fraction > 0.65 required.

Variables (x_i) used by MLPfit: (7 inputs, 1 hidden layer with 16 nodes,1 output):

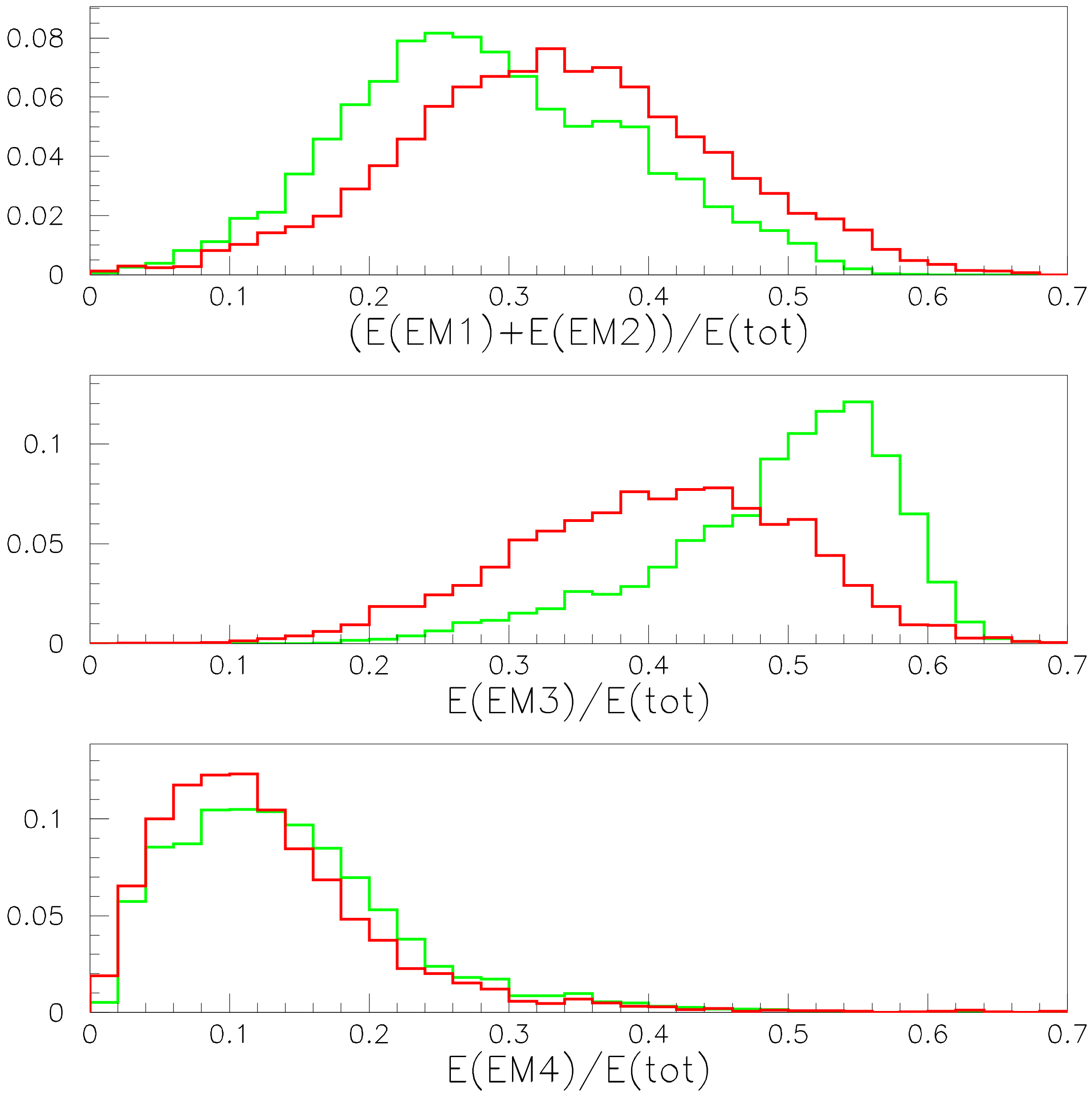
1. $E_{\mathrm{EM1+EM2}}/E_{\mathrm{total}}$
2. $E_{\mathrm{EM3}}/E_{\mathrm{total}}$
3. $E_{\mathrm{EM4}}/E_{\mathrm{total}}$
4. $E_{\mathrm{CPS}}/E_{\mathrm{total}}$
5. $E_{\mathrm{total}}/100$
6. σ_ϕ (cal)
7. σ_η (cal)

Trained over 400 epochs

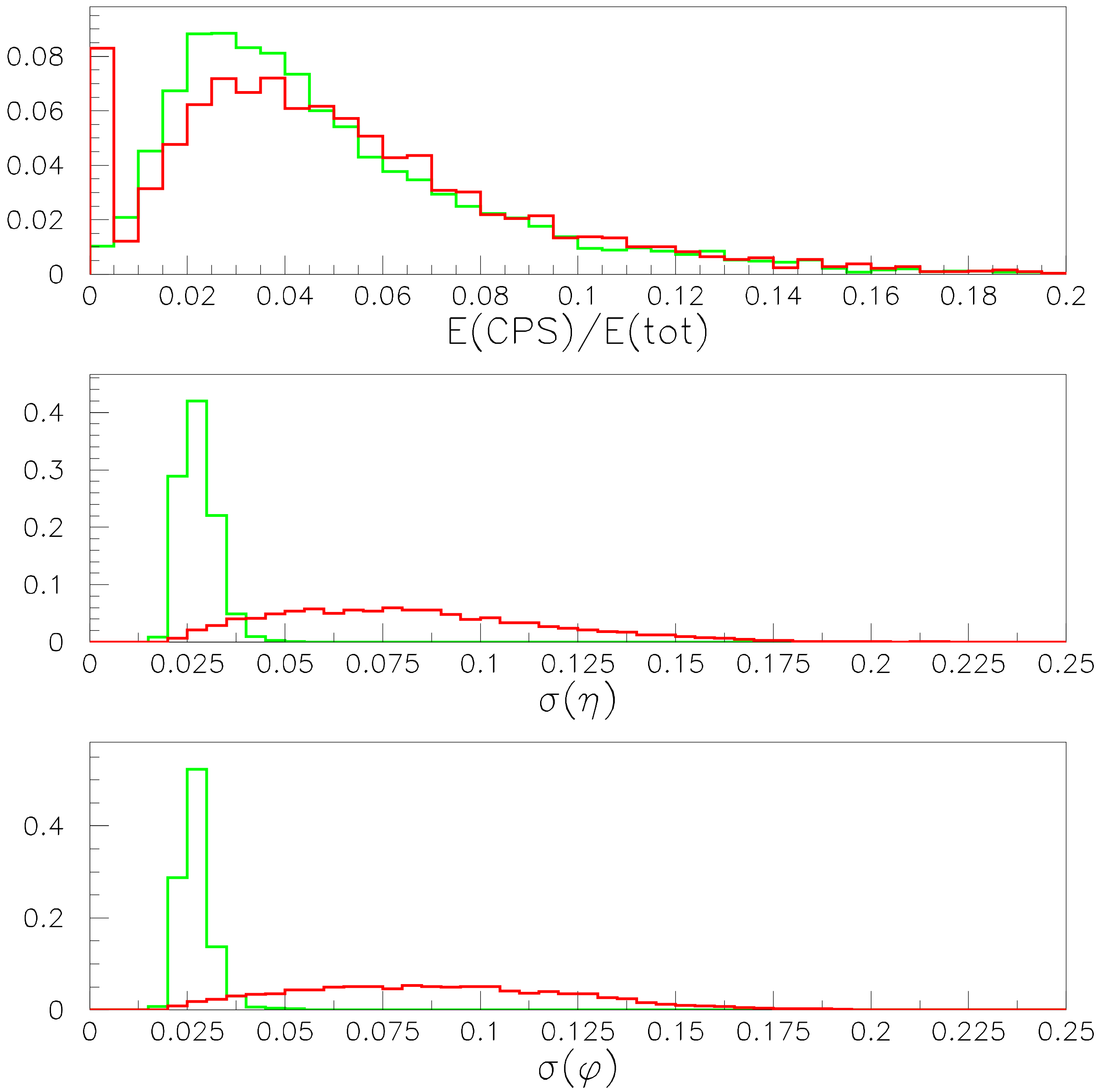
E, η, ϕ


0.1
0.075
0.05
0.025
0
0 25 50 75 100 125 150 175 200 225 250
E(tot)
0.2
0.15
0.1
0.05
0
−40 −30 −20 −10 0 10 20 30 40
iη
0.04
0.03
0.02
0.01
0
0 10 20 30 40 50 60
iφ

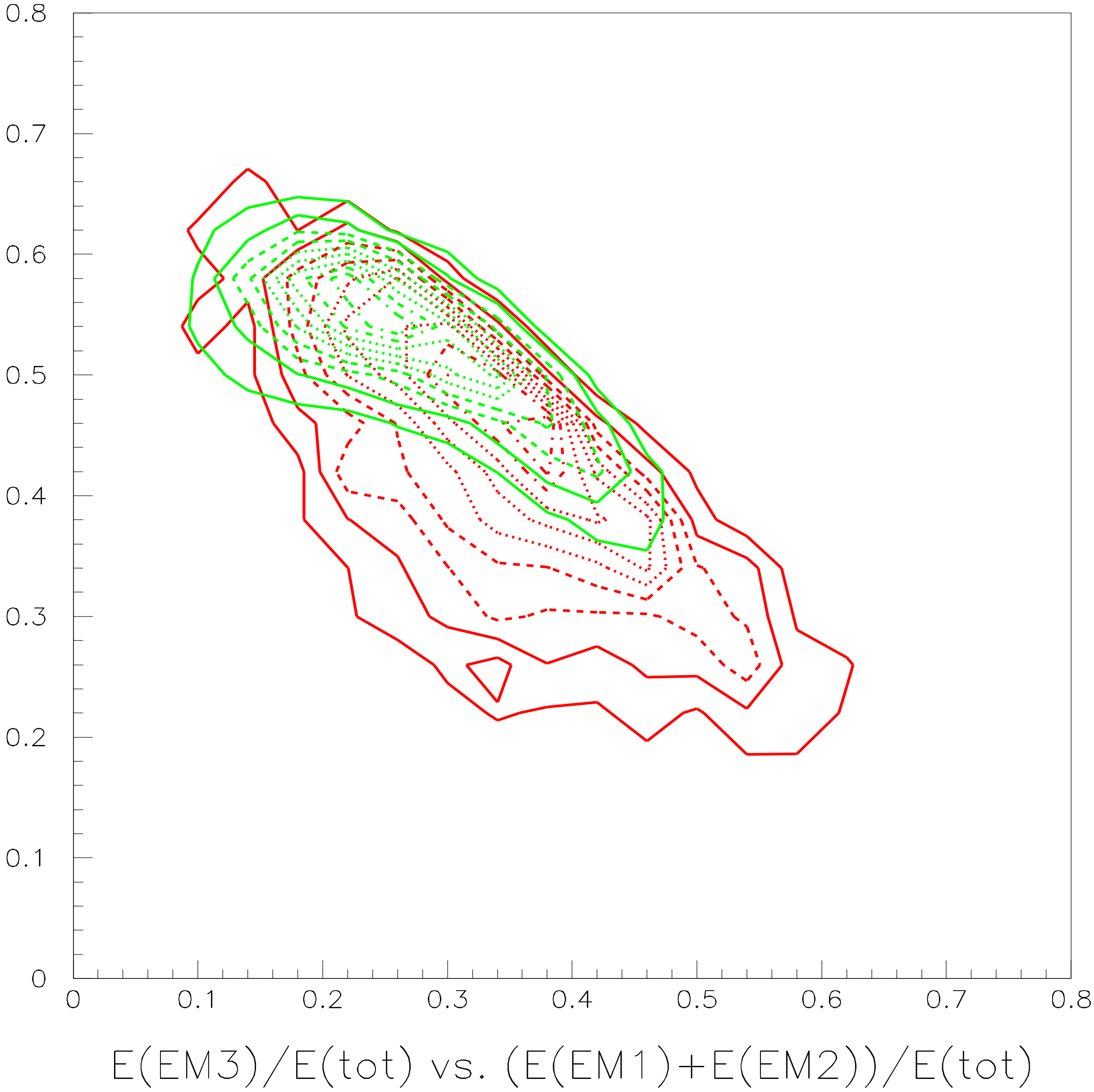
EM fractions



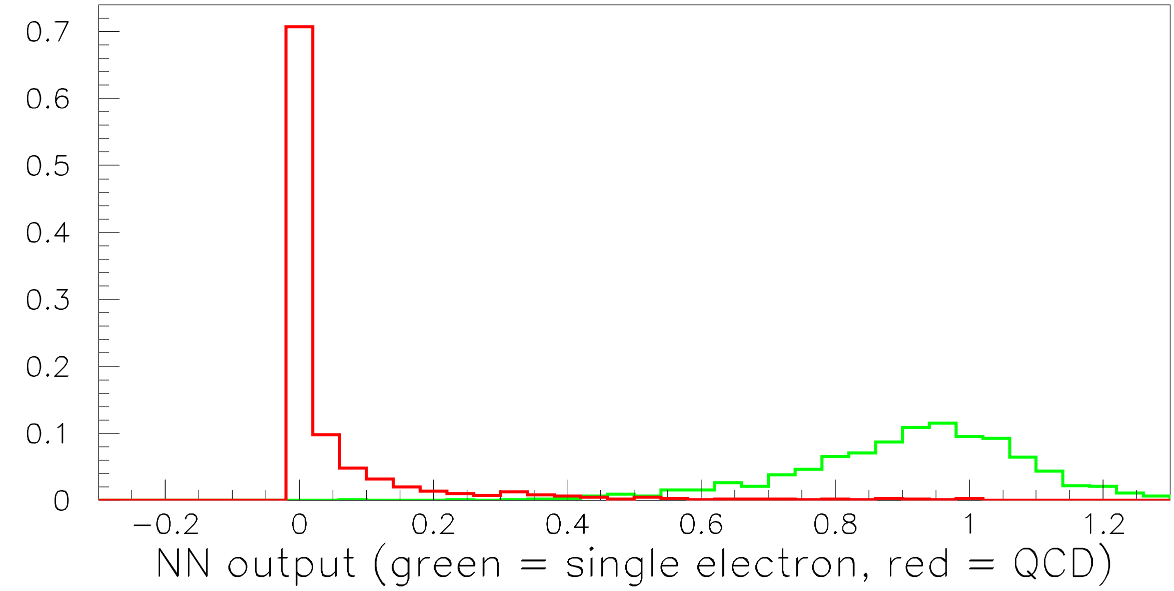
CPS fraction, σ_η, σ_ϕ

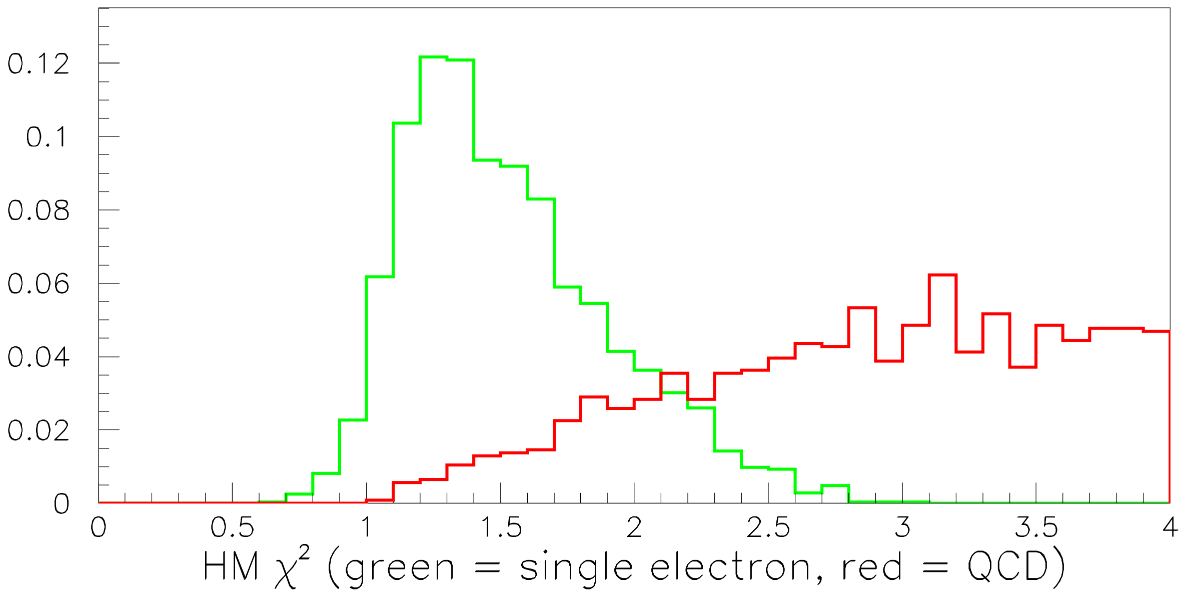


Correlation between input variables

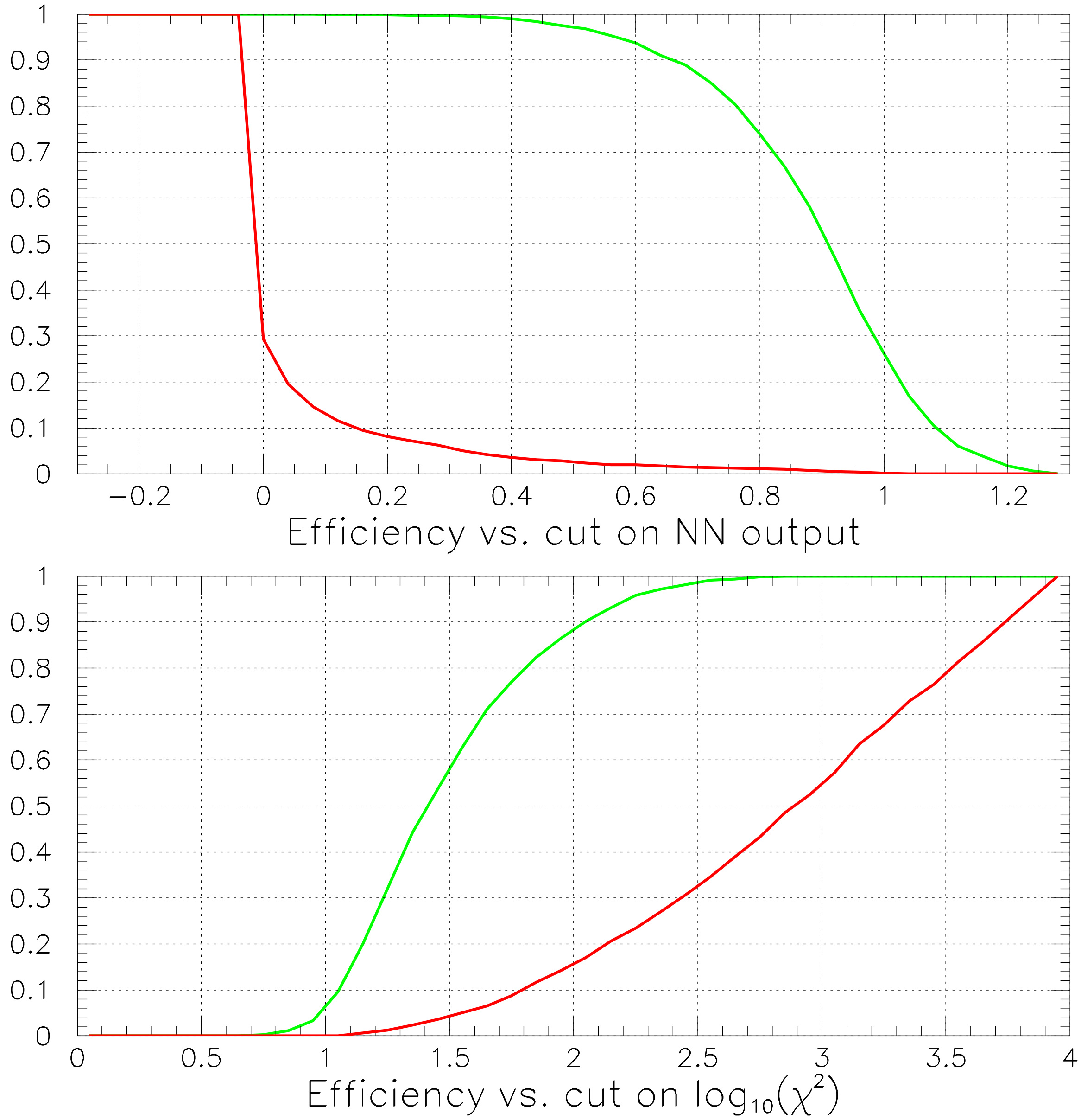


NN output and H-matrix χ^2: S and B

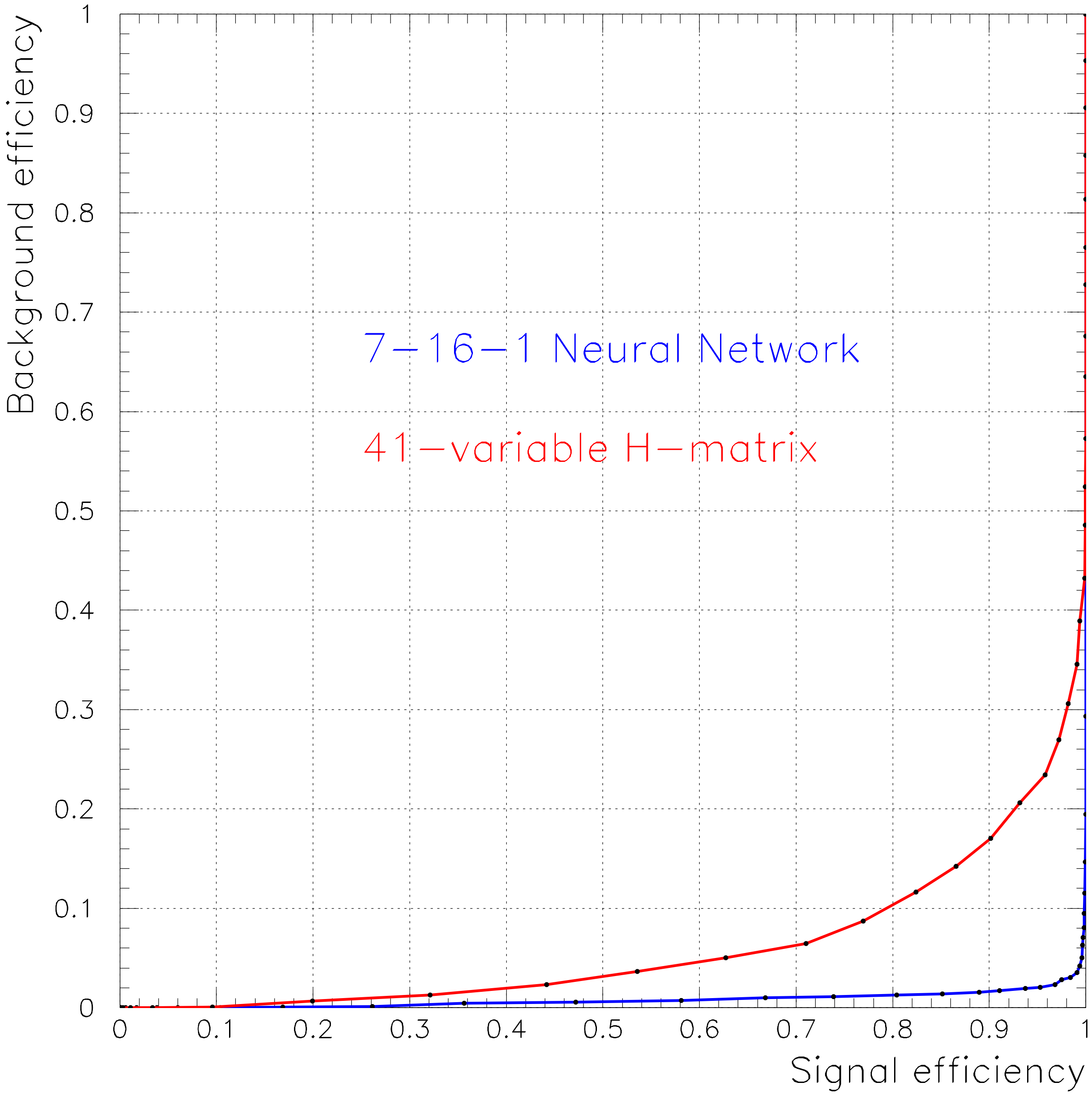




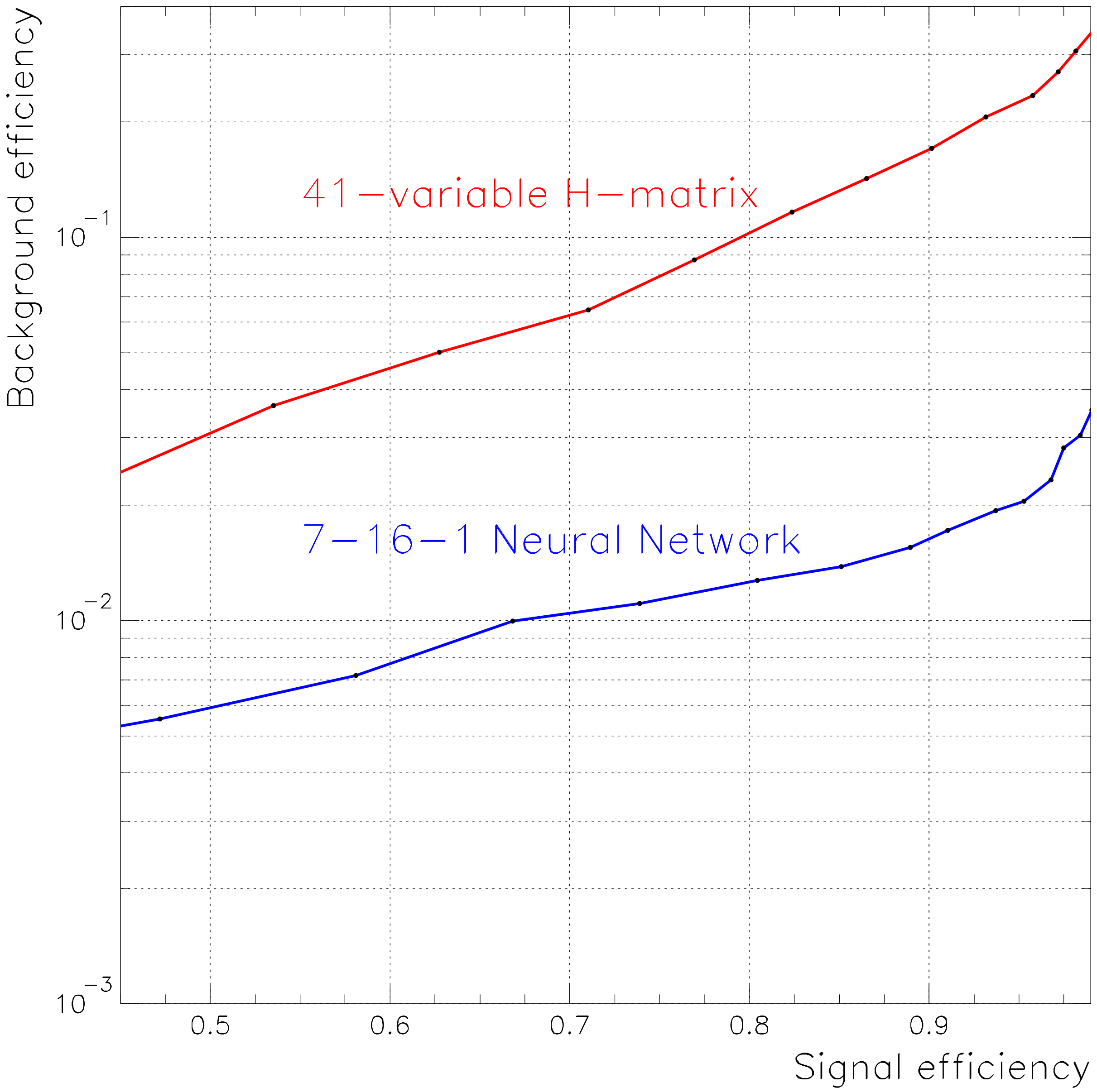
Efficiencies vs cuts on NN output and H-matrix χ^2



H-matrix and NN efficiencies B vs. S



H-matrix and NN efficiencies B vs. S



Selection criteria for tau id:

- Calorimeter cluster ($\Delta R = 0.7$)
- $|\eta| < 1.0$
- $E_T > 10$ GeV

Samples used for training:

- Signal: $Z \to \tau\tau$ + 0 jet + 1.1 (avg) minbias, Pythia, 2000 events (1157 candidates used by NN)
- Background: QCD ($p_T > 20$ GeV) + 1.1 (avg) minbias, Pythia,

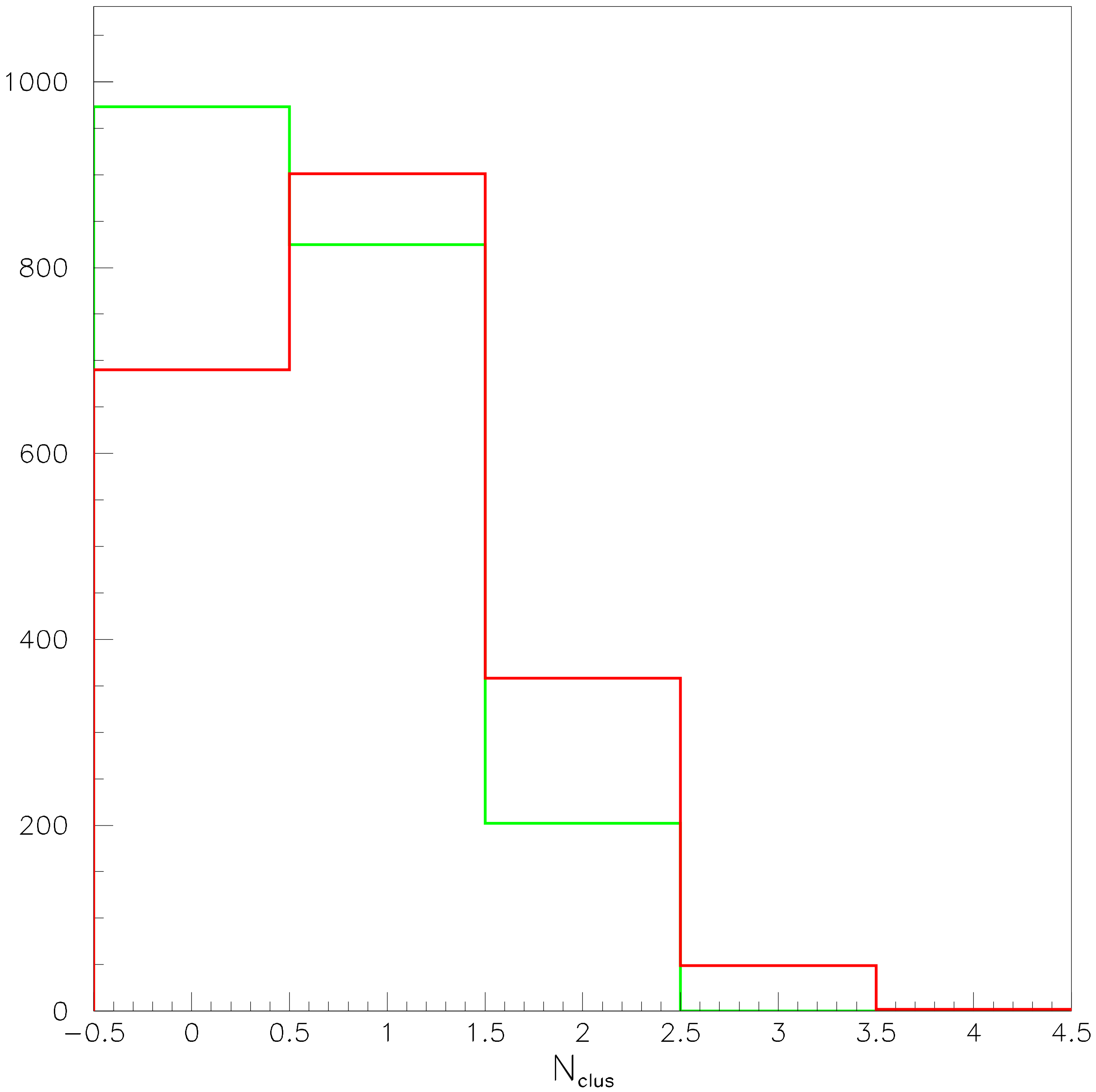
Variables (x_i) used by MLPfit for τs: (9 inputs, 1 hidden layer with 20 nodes,1 output)

1. $E_{\mathrm{EM3}}/E_{\mathrm{total}}$
2. $E_{\mathrm{EM4}}/E_{\mathrm{total}}$
3. $E_{\mathrm{FH}}/E_{\mathrm{total}}$
4. RMS (cal)
5. $E_{\mathrm{hot2}}/E_{\mathrm{total}}$
6. $\log_{10} E_{\mathrm{total}} - 1$
7. $\log_{10}(N_{0.4} + 1)$
8. $(N_{0.4} - N_{0.2})/(N_{0.4} + 1)$
9. RMS (trk)

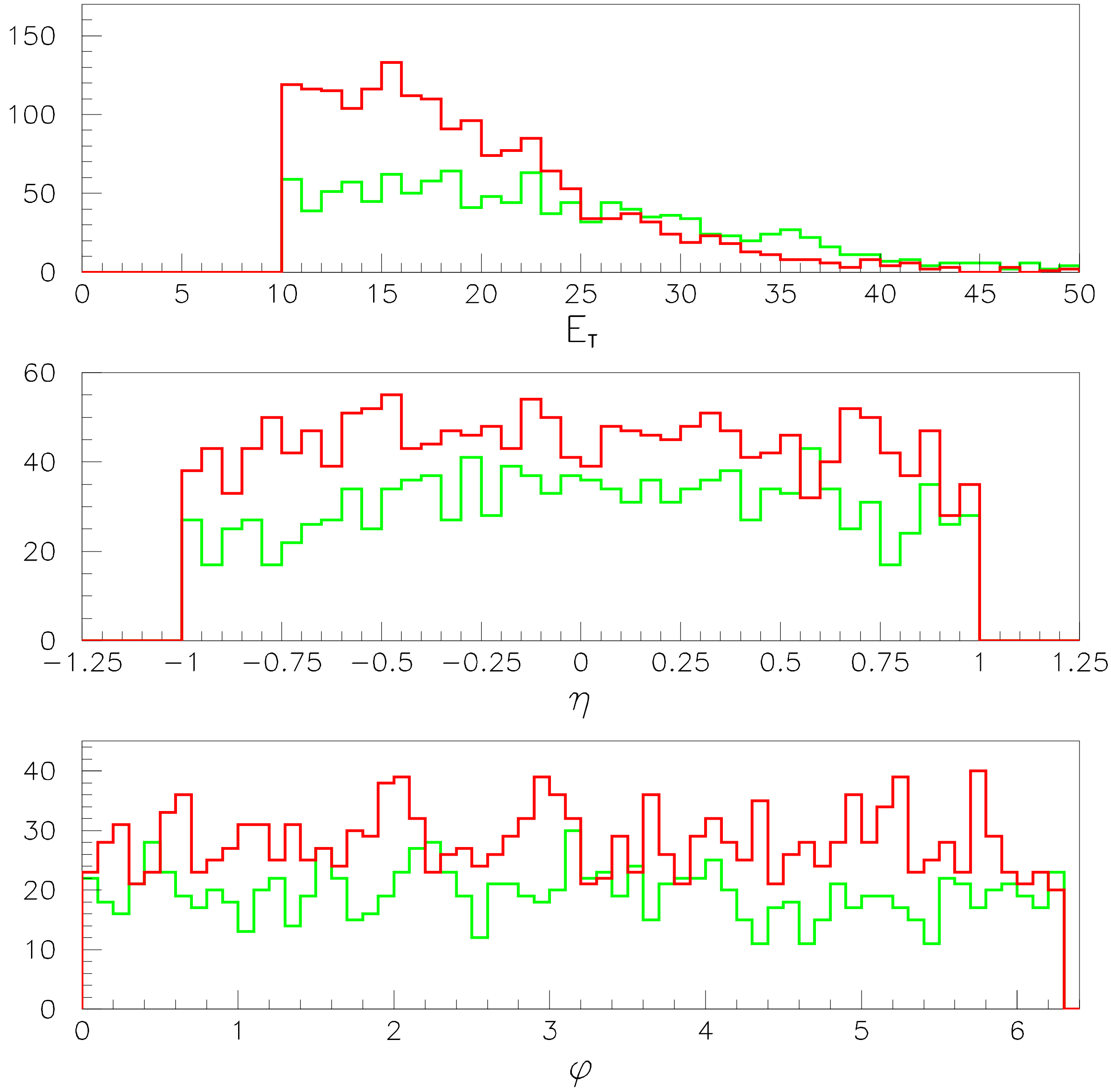
Additional variables used by the H-matrix (all of the above with minor differences, except RMS (cal), plus):

1. $E_{\mathrm{EM1}}/E_{\mathrm{total}}$
2. $E_{\mathrm{EM2}}/E_{\mathrm{total}}$
3. $E_{3\times 3}/E_{\mathrm{total}}$
4. $E_{5\times 5}/E_{\mathrm{total}}$
5. $p_{\mathrm{total}}/E_{\mathrm{total}}$
6. $z_v/\sigma(z_v)$

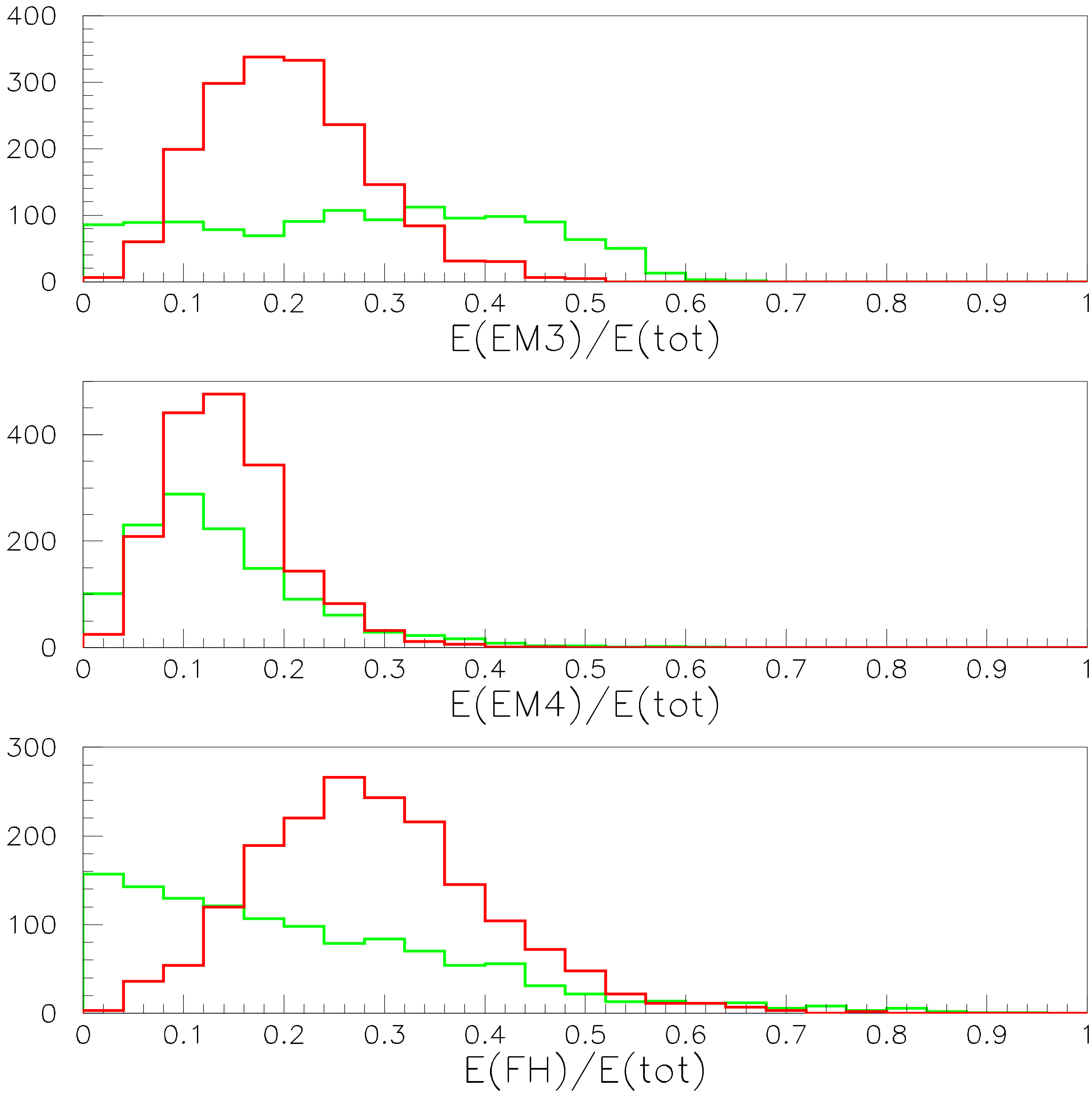
Number of tau candidate clusters in S and B



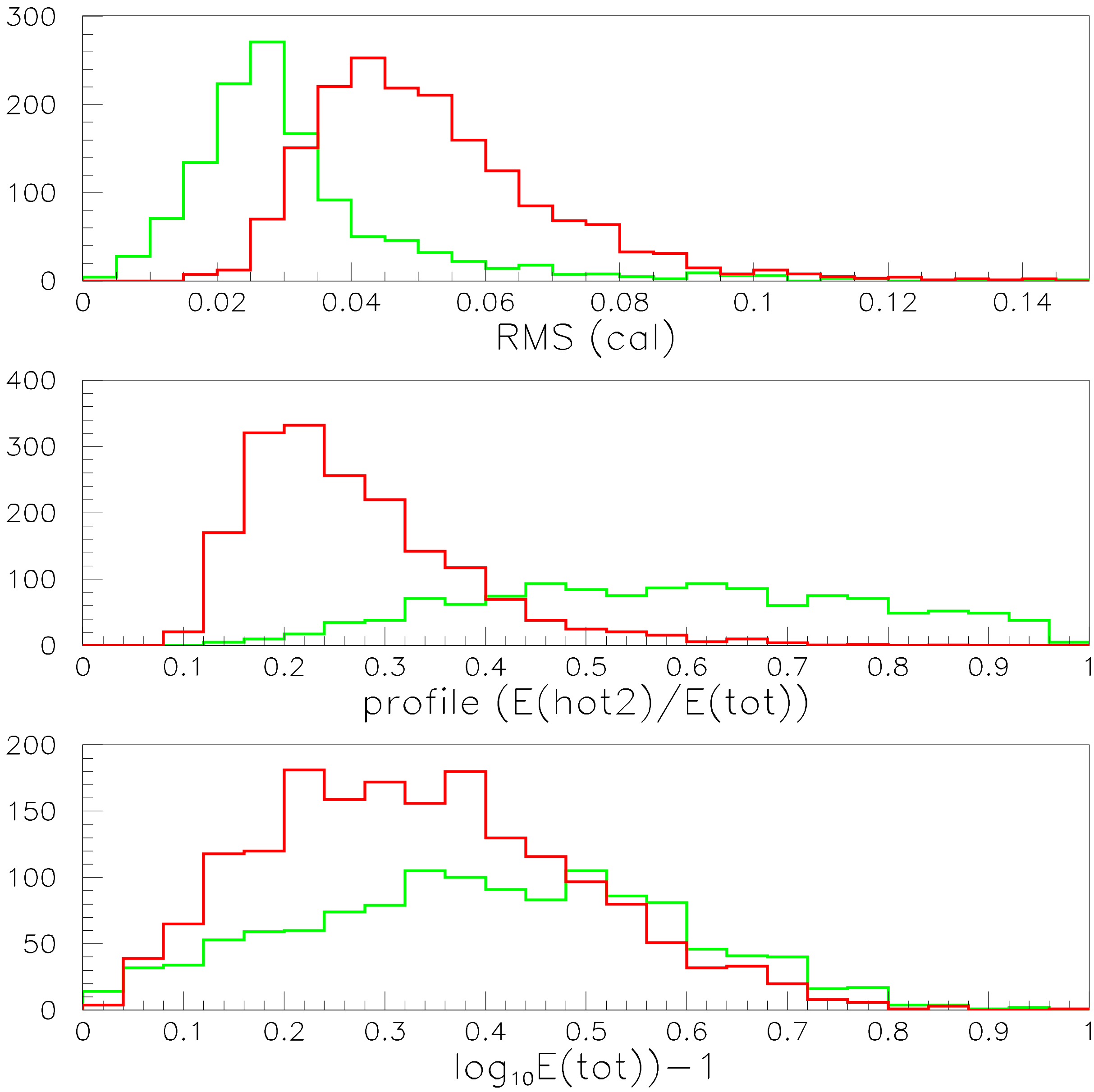
E_T, η, ϕ


150
100
50
0
0 5 10 15 20 25 30 35 40 45 50
E_T
60
40
20
0
−1.25 −1 −0.75 −0.5 −0.25 0 0.25 0.5 0.75 1 1.25
η
40
30
20
10
0
0 1 2 3 4 5 6
φ

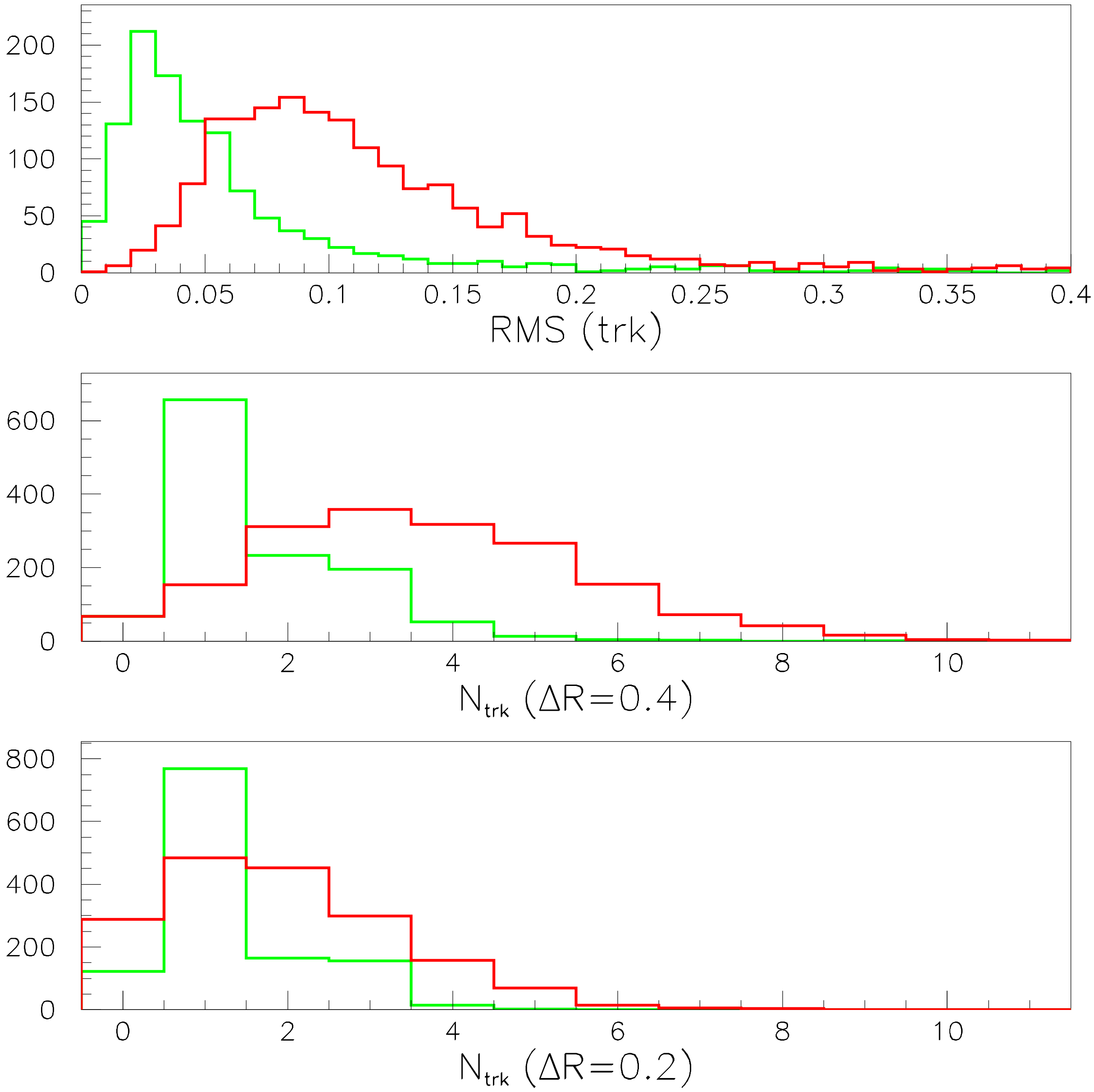
Calorimeter cluster longitudinal variables



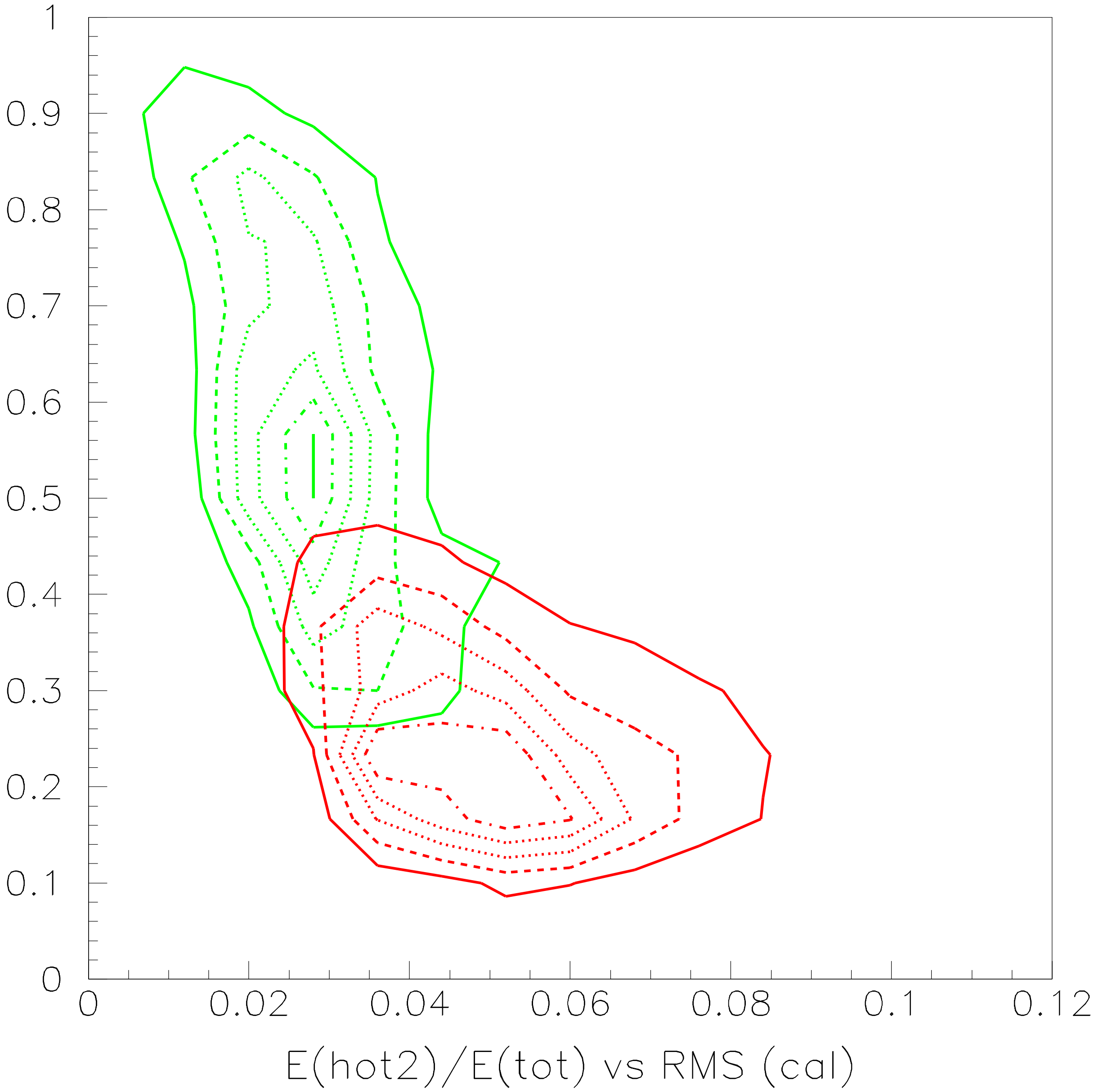
Calorimeter cluster lateral variables and total energy



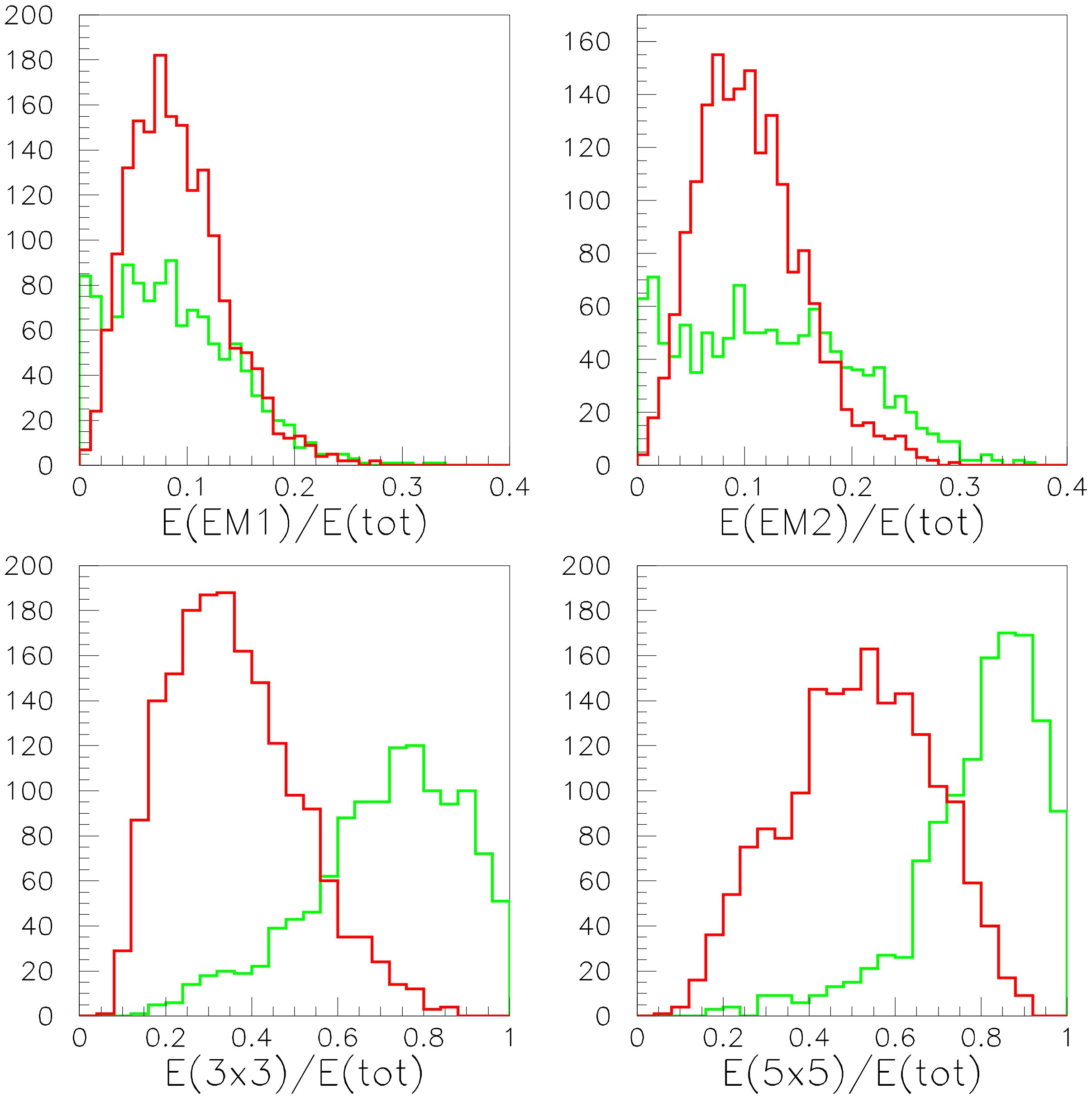
Tracking variables



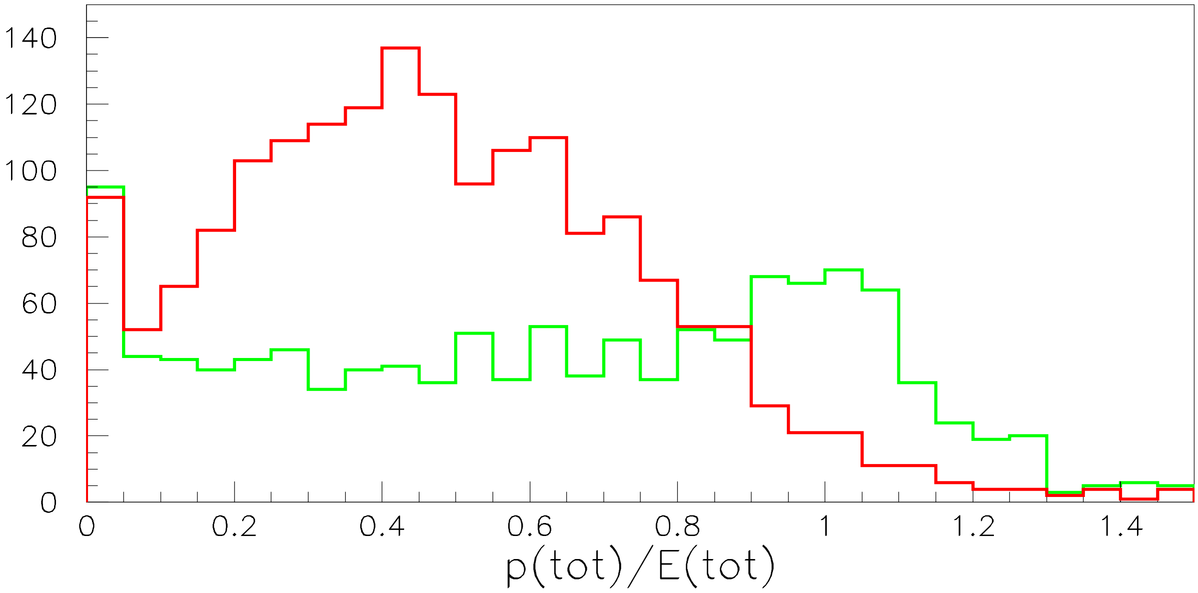
Correlation between input variables

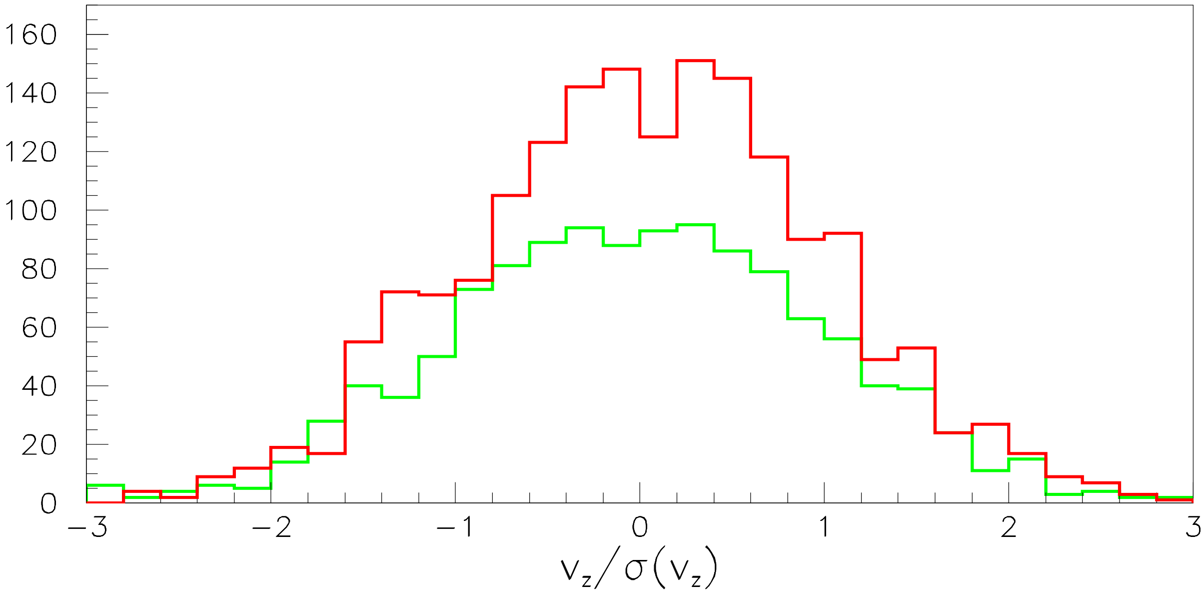


Additional variables used in H-matrix

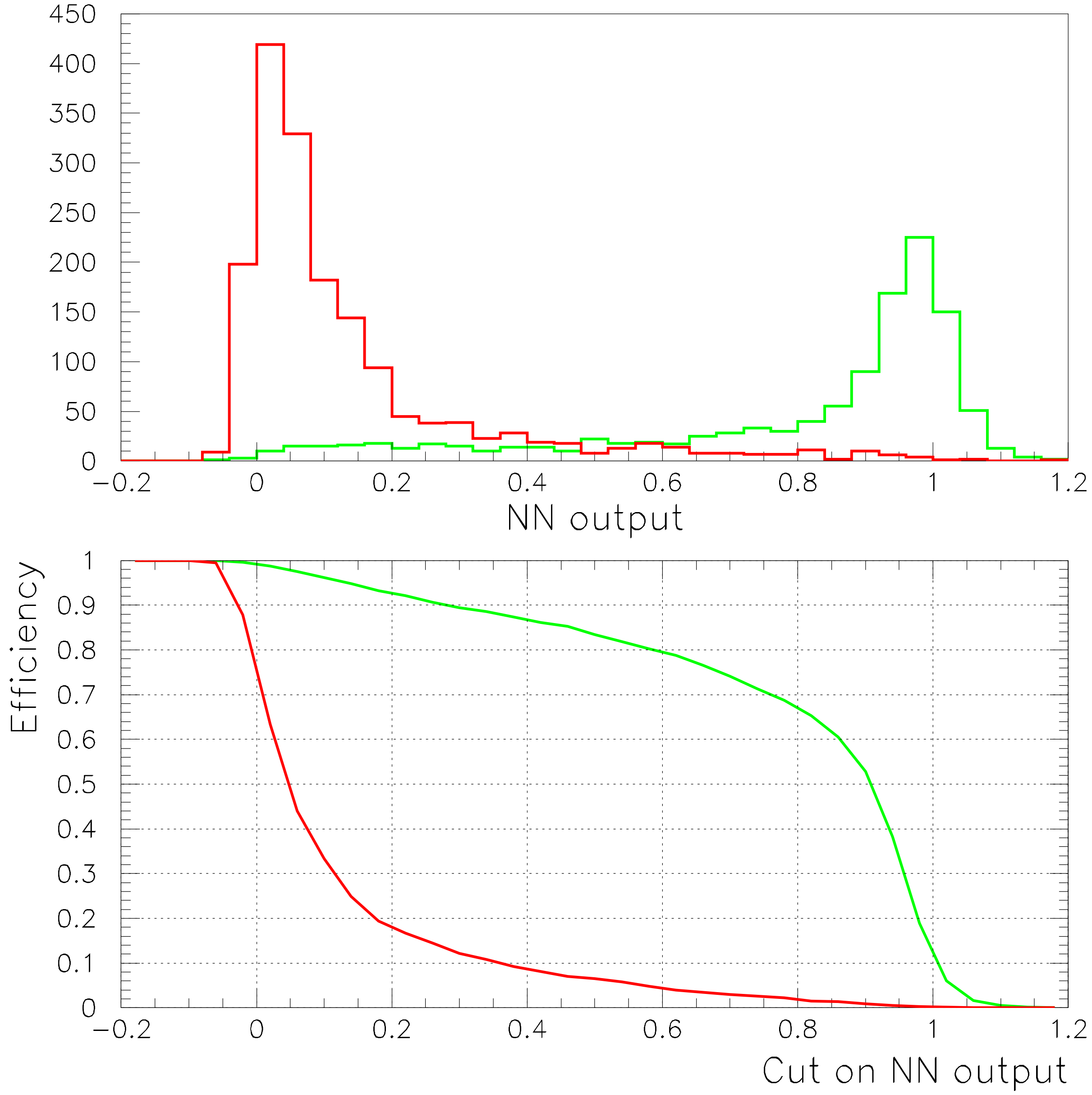


Additional variables used in H-matrix (contd)

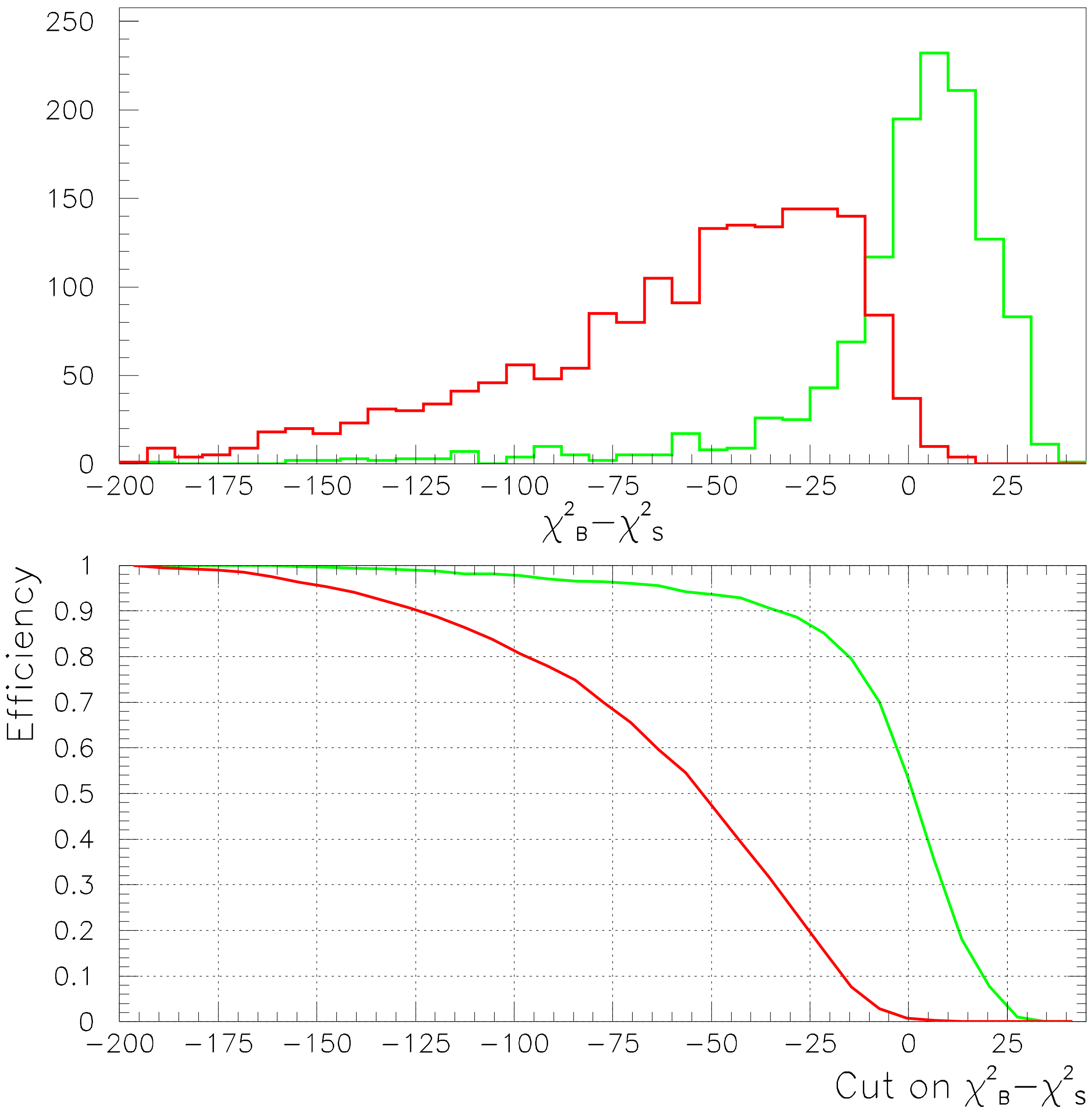




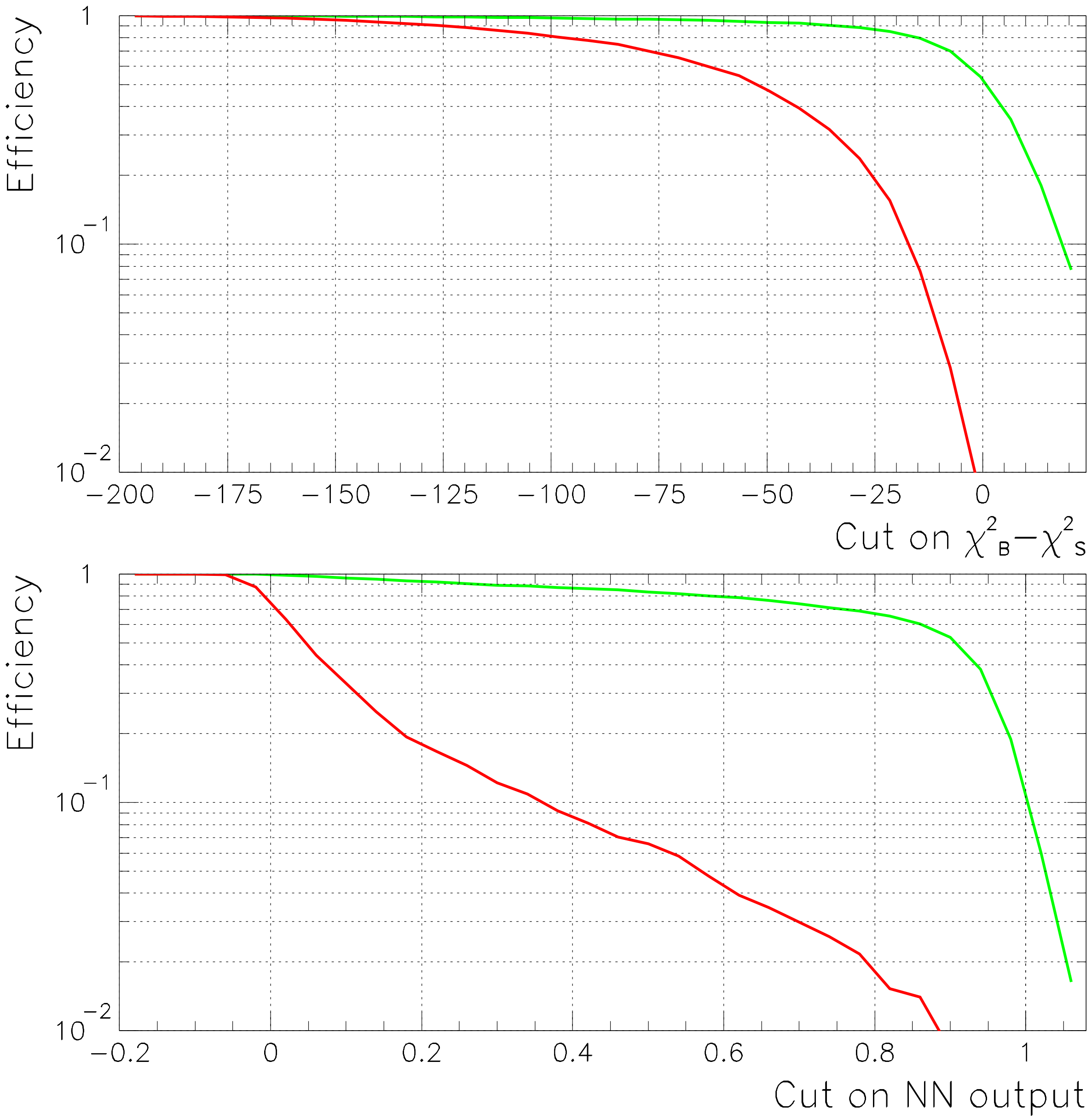
NN output and efficiencies: S and B



H-matrix Fisher discriminant and efficiencies: S and B



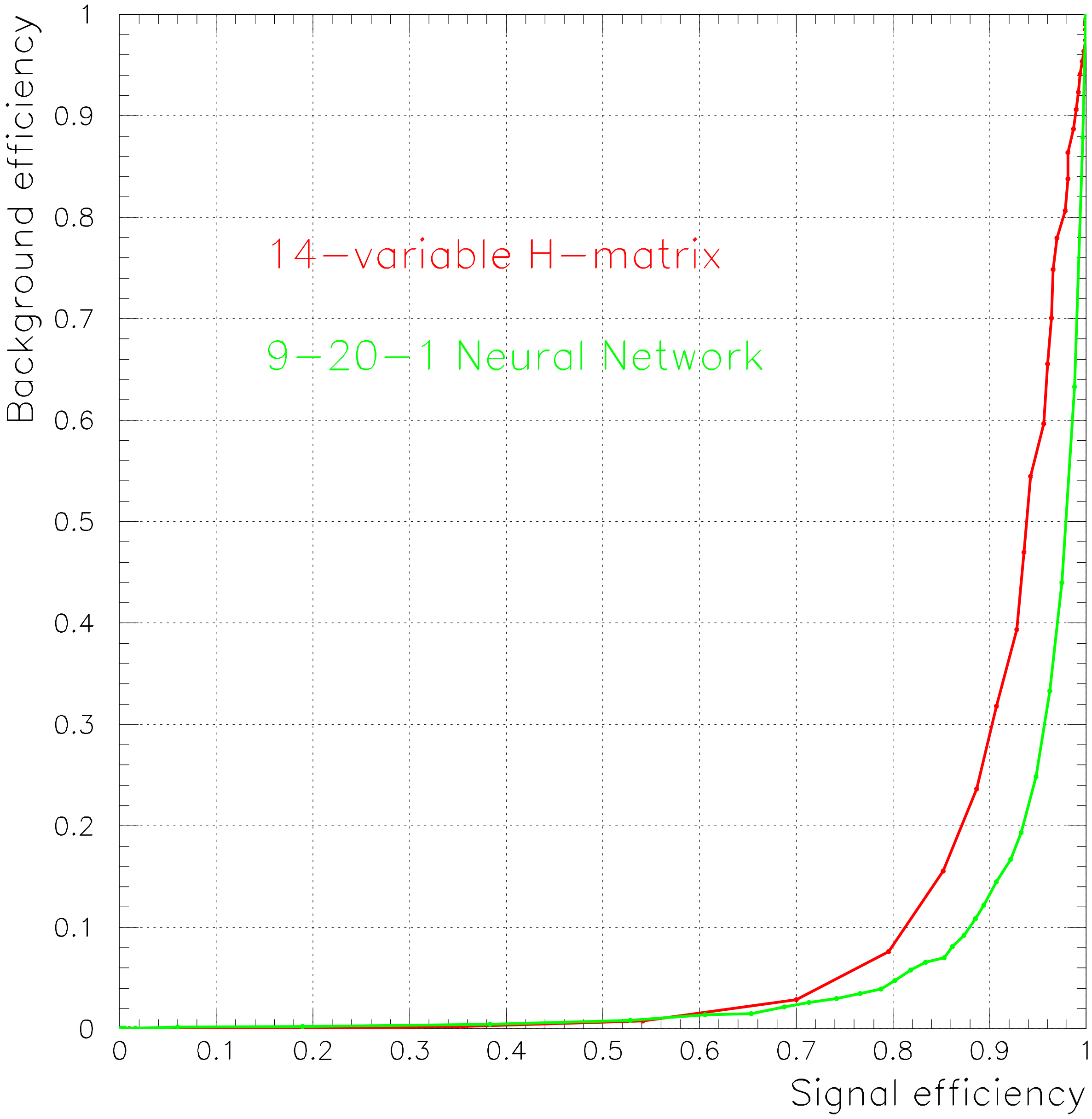
H-matrix and NN efficiencies: S and B



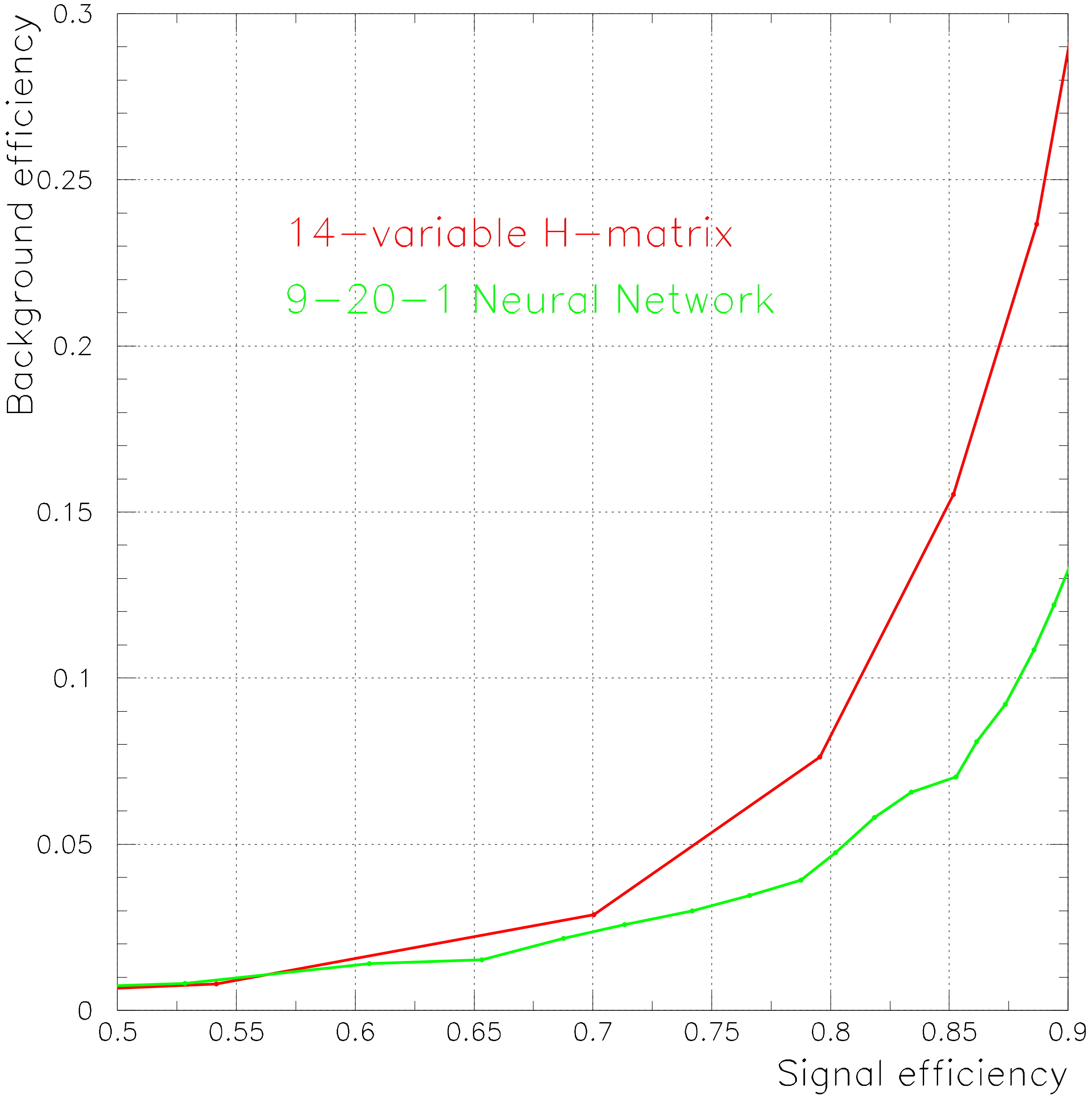
Note: Two differences between HM and NN have compensating effects:

- H-matrix is trained only on clusters selected by TauReco, but tested on all 0.7 cone clusters. It is therefore not as efficient in finding real τs that fail the TauReco preselection.
- On the other hand, the H-matrix is trained on a much larger sample which includes the testing sample. This results in a superior efficiency for the testing sample.

H-matrix and NN efficiencies: B vs S



H-matrix and NN efficiencies: B vs S



Summary

- The Neural Network looks very promising: factor of 2 (in tau id) to 10 (in electron id) improvement over H-matrices can be achieved with fewer variables and little optimization.
- Using the NN as the sole offline discriminator will enhance S/B especially when there are other handles in an event.
- Initial studies indicate that significant further improvements can be achieved by using additional variables available to us.