Fermi National Accelerator Laboratory

FERMILAB-Conf-94/261-E

DØ

Search for the Top Quark using Multivariate Analysis Techniques

Pushpalatha C. Bhat
The DØ Collaboration

Fermi National Accelerator Laboratory
P.O. Box 500, Batavia, Illinois 60510

August 1994

Proceedings of the *Eighth Meeting of the Division of Particles and Fields of the American Physical Society (DPF'94)*, Albuquerque, NM, August 2-6, 1994.

Operated by Universities Research Association Inc. under Contract No. DE-AC02-76CHO3000 with the United States Department of Energy

Disclaimer

SEARCH FOR THE TOP QUARK USING MULTIVARIATE ANALYSIS TECHNIQUES

PUSHPALATHA C. BHAT *
Fermi National Accelerator Laboratory,
Batavia, IL 60510, U.S.A.

ABSTRACT

The DØ collaboration is developing top search strategies using multivariate analysis techniques. We report here on applications of the H-matrix method to the $e\mu$ channel and neural networks to the e+jets channel.

1. Introduction

Top quark events are being searched for in the di-lepton, lepton+jets and all-jets channels at the Collider detectors at Fermilab, in $p\bar{p}$ collisions at $\sqrt(s)$ =1.8 TeV. The DØ collaboration has been applying multivariate techniques such as the Covariance matrix (H-matrix) method, Probability Density Estimation (PDE) method and Neural Networks, in the search for the top quark. In this paper, we present a brief discussion of these techniques and report on some aspects of the on-going analyses.

2. Multivariate Techniques

Multivariate classifiers provide a discriminating boundary between the signal and background in multidimensional space that can yield discrimination close to the theoretical maximum (Bayes' limit[1]). If $P(s|x)$ ($P(b|x)$) is the probability that a given event with feature vector x is a signal (background) event, then the optimal way to partition the feature space is to cut on the ratio of these probabilities. This ratio is the Bayes discriminant function,

$$R(x) = \frac{P(s|x)}{P(b|x)} = \frac{P(x|s)P(s)}{P(x|b)P(b)}. \quad (1)$$

$P(s|x), P(b|x)$ are also known as Bayesian probabilities. The quantities $P(x|s), P(x|b)$ are the likelihood functions for signal and background, respectively (hereafter referred to as f(x)). The ratio of the prior probabilities $\frac{P(s)}{P(b)}$ is the ratio of signal and background cross-sections. Different multivariate classifiers approximate the likelihood functions with different functional forms and attempt to arrive at the Bayes discriminant. The three classifiers being used at DØ are briefly described below.

*Representing the DØ Collaboration

To be published in the Proceedings of the Eighth meeting of the APS Division of Particles & Fields, 1994.

2.1. H-Matrix Method

This is the familiar covariance matrix method which is also known as the Gaussian Classifier. The likelihood function is taken to be gaussian,

$$f(x) = A.exp\{-\frac{1}{2}\sum_{i,j}(x_i - \bar{x}_i)^T M^{-1}(x_j - \bar{x}_j)\} = A.exp(-\chi^2) \quad (2)$$

where $H = M^{-1}$ is the covariance matrix. We use Fisher's formulation of the discriminant, which is $F = \frac{1}{2}(\chi_b^2 - \chi_s^2)$, where χ_b^2 and χ_s^2 are the χ^2 terms of a sample calculated using background and signal H-matrices respectively. The Bayes discriminant in terms of the Fisher variable F can be shown to be $R(x) = exp(F)$, when P(s)=P(b).

2.2. Probability Density Estimation (PDE) Method

The likelihood function is approximated as,

$$f(x) = \frac{1}{N_{events}h}\sum_{i=1}^{N_{events}}\prod_{j=1}^{d} K(\frac{x_i - x_{ij}}{h_j}) \quad (3)$$

where K is a kernel which we take to be a multivariate gaussian centered at each data point x_{ij} with variance h_j^2 (for the jth variable). The Bayes discriminant is $R(x) = \frac{f_s(x)}{f_b(x)}$.

2.3. Neural Networks

It has been shown[2] that neural networks do not calculate the likelihood function for each class separately, but arrive at the Bayesian probability for the signal directly. The discriminant in this case is the output of the network

$$O(x) = g(\sum_j w_{kj} g(\sum_i w_{ji} x_i)) = P(s|x) \quad (4)$$

(assuming a three layer feed-forward neural network). The x_i's are the input variables, g represents a non-linear function ($e.g.$, $\frac{1}{(1+e^{-2x})}$), w_{kj} and w_{ji} are the weights that are adjusted during the "learning" process. Descriptions of the neural network approach and details of the training algorithms are available in many articles and books. The Bayes discriminant in terms of the network output will be $R(x) = \frac{O(x)}{(1-O(x))}$.

3. H-matrix Analysis of $e\mu$ data

From the conventional analysis,[3] DØ has one top candidate event in the $e\mu$ channel. The dominant backgrounds are $Z \to \tau\tau$, WW and instrumental fake events. We have applied the H-matrix method to enhance the signal to background particularly w.r.t. $Z \to \tau\tau$. We have built H-matrices using the variables E_T^e, P_T^μ, E_T^{jet1}, E_T^{jet2}, $\not{E}_T^{cal}$($\not{E}_T$ in the calorimeter), $H_T(\Sigma E_T$ of jets), $M_{e\mu}$, $\Delta\phi_{e\mu}$ for $t\bar{t}$, $Z \to \tau\tau$ Monte Carlo (MC) and data events, after applying loose electron identification (ID) criteria and requiring $P_T^e > 11$ GeV and $P_T^\mu > 11$ GeV. We use data to represent one set of backgrounds ($b\bar{b} \to e\mu$ and fakes -bkg1) and $Z \to \tau\tau$ as the other background (bkg2). These H-matrices are then applied to data, $t\bar{t}$ (top mass 140, 160 and 180 GeV), $Z \to \tau\tau$ and WW samples and χ^2 values calculated. The lego plots of χ^2_{Top} vs χ^2_{bkg1} are shown in

Fig.1. It can be seen that the data and $Z \to \tau\tau$ events have small χ^2 values with both signal and background H-matrices whereas top events have small χ^2_{Top} but large χ^2_{bkg1}. We define two Fisher discriminants $F_1 = \frac{1}{2}(\chi^2_{bkg1} - \chi^2_{Top})$ and $F_2 = \frac{1}{2}(\chi^2_{bkg2} - \chi^2_{Top})$. In Fig. 2, are shown F_1 for various event samples. We search for a combination of F_1 and F_2 to keep the same efficiency as in the conventional analysis for 140 GeV top events and to maximize background rejection. By applying $F_1 > 15$ and $F_2 > 3$ we have 16%, 22% and 25% efficiency for top events with top mass of 140, 160 and 180 GeV respectively. The signal to background ratio (S/B) is about 18 w.r.t. to $Z \to \tau\tau$ and 10 w.r.t. WW events for 180 GeV top mass. For lower masses the S/B is higher.

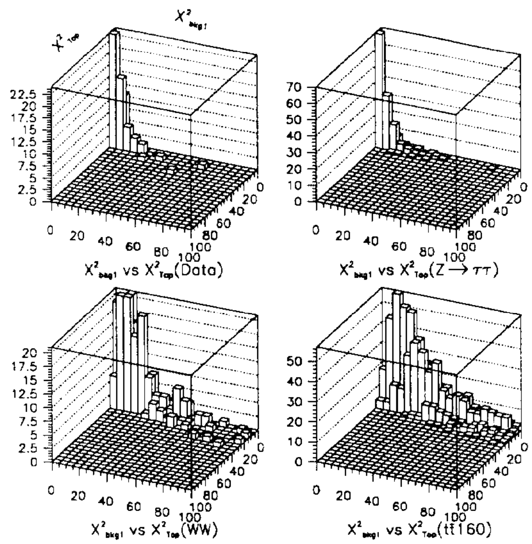

Fig. 1. χ^2_{Top} vs χ^2_{bkg1} for DØ data, $Z \to \tau\tau$, WW and $t\bar{t}160$ samples

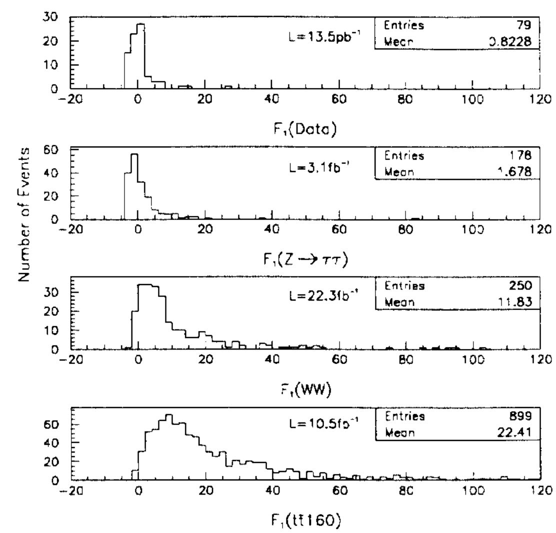

Fig. 2. Fisher variable distributions

4. Neural Networks Analysis of $e +$ jets data

Details of our conventional analyses in the $e +$ jets channel can be found elsewhere in these proceedings.[4,5] Here, we have analysed e + 4jets events using neural networks after applying the following kinematic cuts: $E_T^e \geq 20$ GeV (+tight electron ID), $\not{E}_T \geq 20$ GeV and $E_T(jet4) \geq 15$ GeV. We use two separate neural networks to handle the dominant backgrounds to this channel *viz.*, $W +$ jets and QCD events where one of the jets gives a false electron ID (QCD fake).

Figure 3 shows a schematic of using networks in parallel. Their outputs can be fed into another network to make higher level decisions. We train the first network with $t\bar{t}160$ and $W +$ jets samples and the second network with $t\bar{t}160$ and QCD fake data sample (obtained from multi-jet triggers at DØ). We have carried out analyses with different sets of input variables. In a 2-variable analysis we use network with 2 input nodes, 3 hidden nodes (one hidden layer) and 1 output node. We use the H_T (sum of E_T of jets with $\eta_{jet} \leq 1.7$) and the aplanarity[4] (A) of the event as input variables. In Fig. 4, we show the distributions of the output for $t\bar{t}160$ and $W +$ jets from network 1.

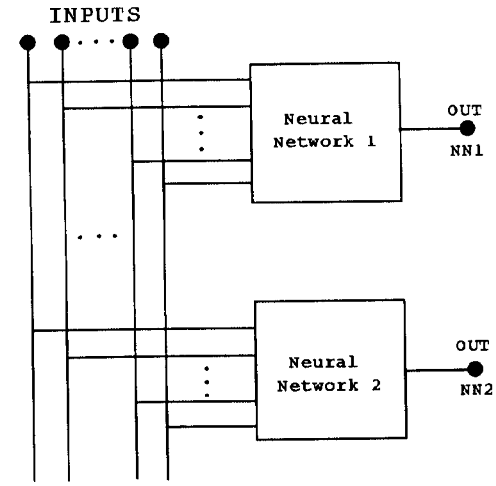

Fig. 3. Processing with many Neural Networks

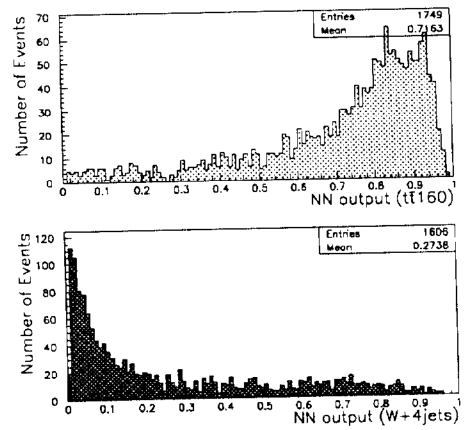

Fig. 4. Distributions of output from network-1 trained on $t\bar{t}160$ and $W+$ jets events

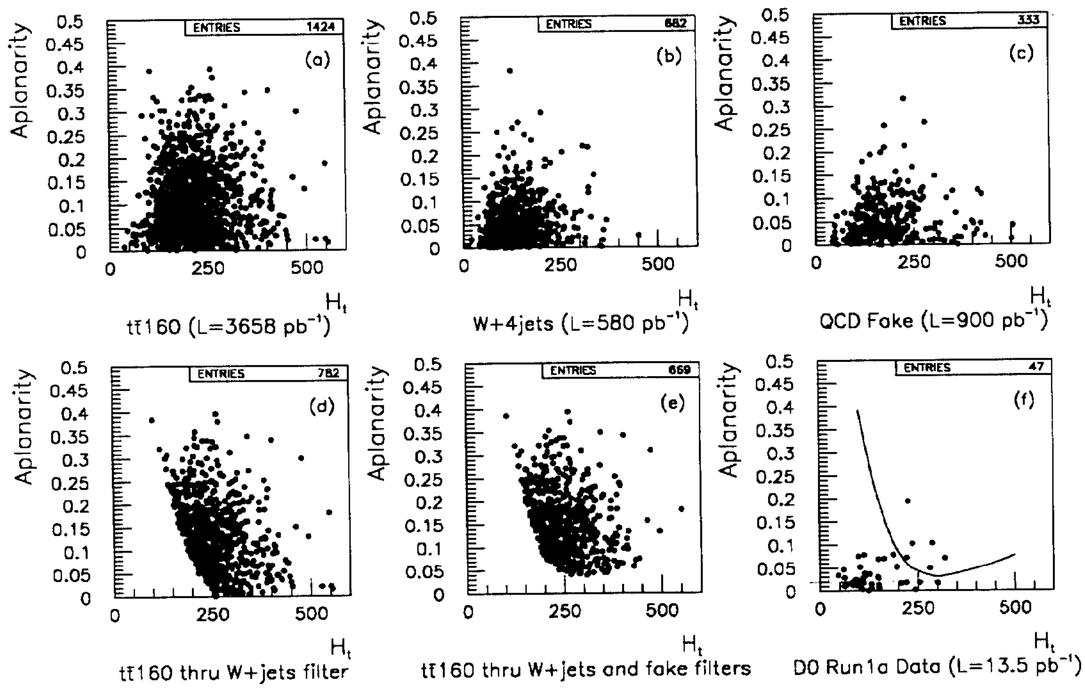

Fig. 5. H_t vs Aplanarity plots from the 2-variable analysis for (a)$t\bar{t}160$, (b)W+4jets (VECBOS MC), (c)QCD fakes, (d)$t\bar{t}160$ after cut on network 1 (NN1>.8), (e)$t\bar{t}160$ after cuts on networks 1 & 2 (NN1>.8, NN2>.6) and (f)D0 data with contour from the cuts NN1>.8,NN2>.6.

From the plots shown in Fig. 5, it can be seen that a combination of neural networks can provide good rejection to individual backgrounds in different parts of the multi-dimensional phase space. The DØ data with the decision boundary generated by the neural networks (NN1> 0.8,NN2>.6) is shown in Fig. 5(f). Figure 6 shows the compound probability (signal probability surface) for the two networks together as a function of H_T and A. When the analysis is extended to many dimensions, it is still possible to examine the distributions of the variables and their correlations before and after the network selection to understand the decision boundary. Figure 7 shows such distributions for a 6 variable analysis done using H_T, A, E_T^e, $\not{E}_T$, $E_T(jet4)$, $\theta(e)$ (6 inputs, 6 hidden nodes and 1 output node used).

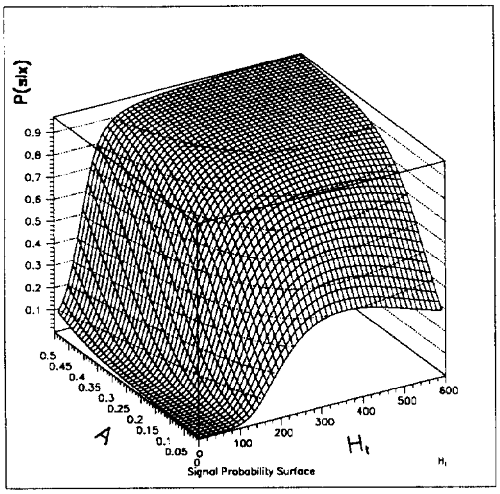

Fig. 6. Signal probability surface for the two networks together as a function of H_T and A

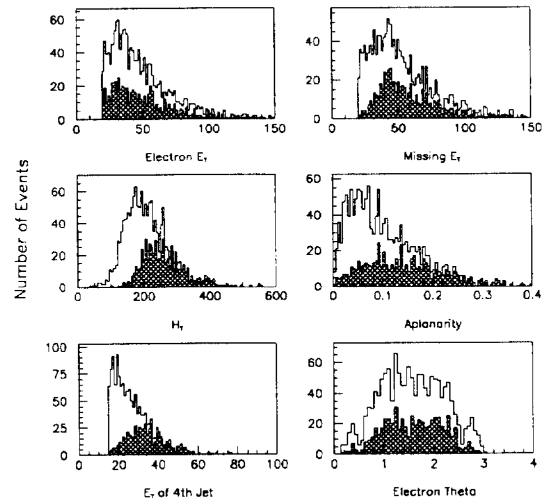

Fig. 7. Distributions of variables in 6-variable analysis before (open histograms) and after(hatched histograms) neural network cuts

5. Summary

A discussion of the multivariate techniques being used in the top search at DØ and some preliminary results have been presented. For example, the H-matrix applied to the $e\mu$ channel yields a S/B=18 for top(180 GeV) to $Z \rightarrow \tau\tau$. This is a significant improvement over conventional analyses.[3] Preliminary results from the neural networks analysis of the lepton+jets channels show promise for better background rejection and higher efficiency than conventional analysis techniques.

This work is supported in part by the U.S. Department of Energy.

References

1. R.O. Duda and P.E. Hart, *Pattern Classification and Scene Analysis* (Wiley, New York, 1973).
2. D.W. Ruck *et al.*, "The multilayer perceptron as an approximation to a Bayes optimal discriminant function", IEEE Trans. Neural Networks 1 (4), 296-298(1990).
3. K. Genser, DØ Collaboration, these proceedings.
4. D. Chakraborty, DØ Collaboration, these proceedings.
5. W. Cobau, DØ Collaboration, these proceedings.