

Application of Neural Networks to b -quark jet detection in $Z \rightarrow b\bar{b}$

Stephen Poprocki

REU 2005

Department of Physics, The College of Wooster, Wooster, Ohio 44691

Advisors: Gustaaf Brooijmans, Andy Haas

Nevis Laboratories, Irvington, NY 10533

August 5, 2005

Abstract

A multilayer perceptron (MLP) artificial neural network (NN) was trained with Monte Carlo data to detect b -jets. A variety of NNs were tested to maximize performance. The best NN was run on data with different reconstruction options. It was found that a simple MLP NN with 6 variables and 6 hidden neurons performed better than using only a decay length significance cut for detecting b -jets. The NN was run on data from the $D\bar{O}$ experiment to measure the mass of the Z boson from $Z \rightarrow b\bar{b}$ decays. Simple background subtraction techniques were used to yield a shifted Z peak at 73.3 ± 3.4 GeV.

1 Introduction

The Standard Model (SM) has largely been successful in explaining almost all observed particle phenomena to date [1, 2]. However, the mechanism for electro-weak symmetry breaking is not yet understood. Many theories predict at least one Higgs field which gives mass to the W^\pm and Z weak bosons. The theories also predict one or more Higgs bosons. The most promising channels for Higgs boson detection at the Tevatron are $p\bar{p} \rightarrow WH \rightarrow l\nu b\bar{b}$, $p\bar{p} \rightarrow ZH \rightarrow l^+l^-b\bar{b}$, and $p\bar{p} \rightarrow ZH \rightarrow \nu\bar{\nu}b\bar{b}$ [3]. Due to the extremely small chance of detecting these events, improved b -jet detection is useful for possible Higgs discoveries. In this paper we focus on the $Z \rightarrow b\bar{b}$ channel of the DØ experiment. Since the mass of the Z boson is well known, measuring it from $Z \rightarrow b\bar{b}$ can provide a means of calibrating b -quark calorimeters.

2 Purity & Efficiency

Only taggable MC jets were used to measure b -tagging efficiency and purity. That is, each jet must be matched with 2 tracks within $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} < 1.0$. Further, we only use MC jets with at least 1 secondary vertex and assume 0 vertex jets are non- b -jets. To correct for these 0 vertex jets, we use the following formulas:

$$\text{efficiency} = \frac{\text{correctly tagged } b\text{-jets}_{\text{NV}>0}}{b\text{-jets}_{\text{NV}>0} + b\text{-jets}_{\text{NV}=0}},$$
$$\text{purity} = \frac{\text{incorrectly tagged non-}b\text{-jets}_{\text{NV}>0}}{\text{non-}b\text{-jets}_{\text{NV}>0} + \text{non-}b\text{-jets}_{\text{NV}=0}}.$$

Purity versus efficiency plots can be obtained by varying a cut on the NN output.

3 Neural Networks

The NNs were implemented in ROOT 4.04/02 with the TMultiLayerPerceptron class. A MLP is a simple feed-forward network. Figure 1 shows a simple model of a MLP NN.

Each neuron is connected to all neurons on the layer below it. We utilize only one output neuron, namely, the probability of the jet having a b -quark. The output of a neuron is a linear combination of the previous neurons and their weights. The sigmoid function is used to normalize the output between 0 and 1. Multiple hidden layers are possible but typically only complicate matters.

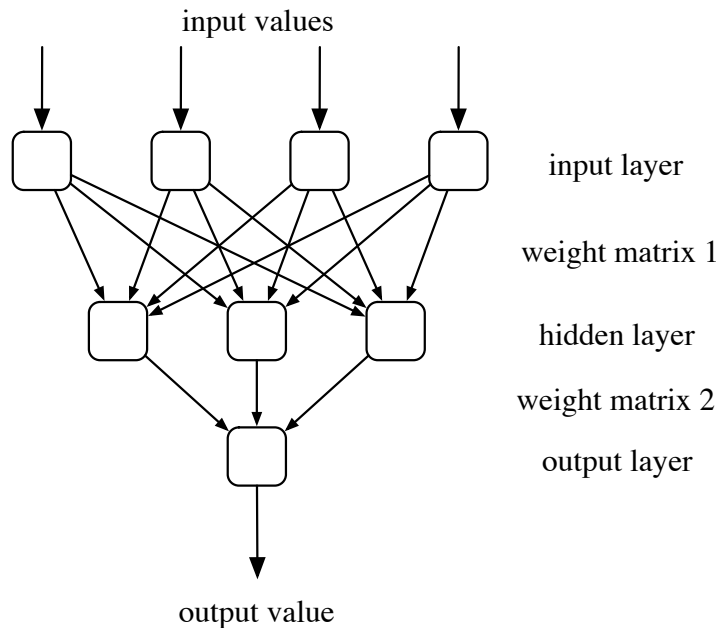


Figure 1: Simple multilayer perceptron neural network.

3.1 Training

Training is the process of minimizing the error of the NN output. Out of a total of 263 767 jets, half were used for training and the other half for testing. The `TMultiLayerPerceptron` class implements six different training methods:

1. Stochastic minimization
2. Steepest descent with fixed step size (batch learning)
3. Steepest descent algorithm
4. Conjugate gradients with the Polak-Ribiere updating formula
5. Conjugate gradients with the Fletcher-Reeves updating formula
6. Broyden, Fletcher, Goldfarb, Shanno (BFGS) method

The training methods adjust the weights based on the error in the output. One weight update is called an epoch, cycle, or iteration. In principle, too few epochs results in an under-trained NN and too many epochs results in an over-trained NN. An over-trained NN is not desirable because it has learned features that are specific to the training data.

Figure 2 shows the errors as a function of the epoch for each training method. Figure 3 shows the signal and background outputs of the NN for the different training methods. Figure 4 shows the purity versus efficiency for the different training methods. The stochastic minimization method was chosen for its simplicity, speed, and slightly superior performance.

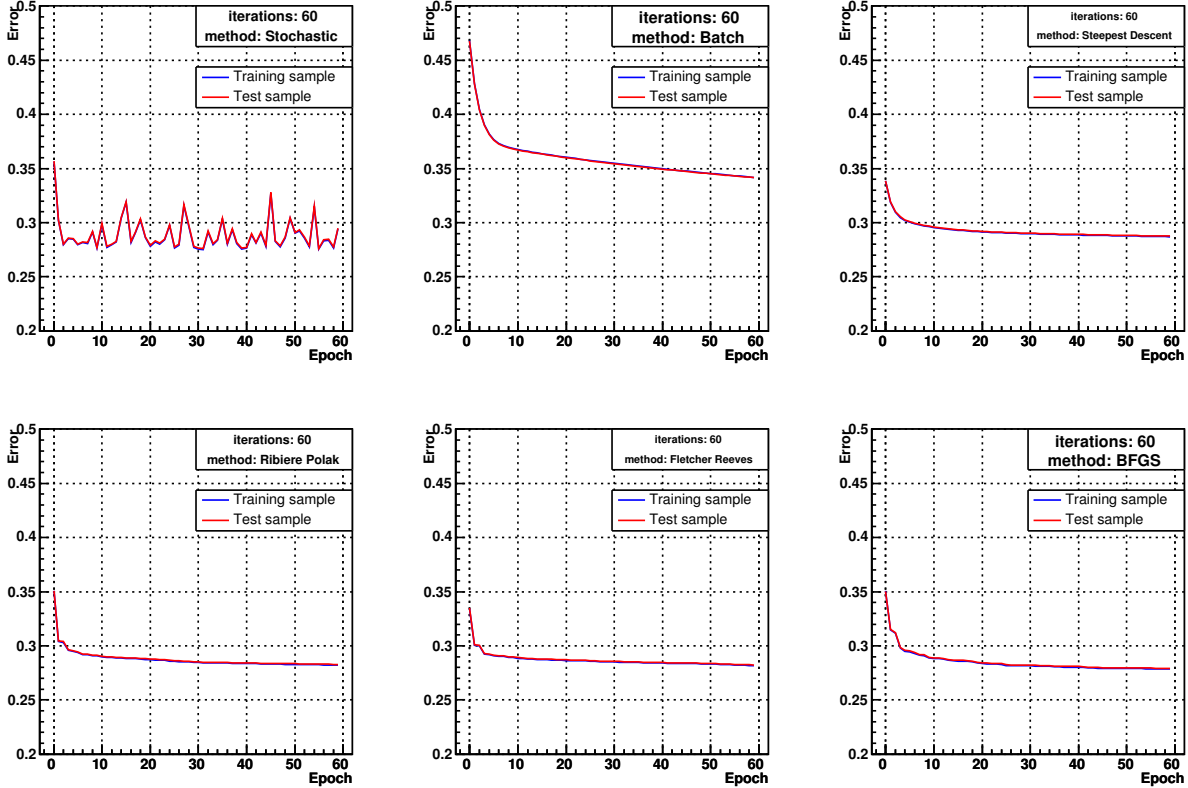


Figure 2: Training error for various training methods as a function of number of epochs.

3.2 Variable Sets

Initially, the NNs were trained with many of combinations of variables. Then two variable sets were chosen for comparison:

“simple”:

- 3D decay length significance (`sdls3d`)
- ΔR (`dr`)
- `mass`
- 2D decay length (`d12d`)
- χ^2 (`chi2`)
- multiplicity (`mult`)

“fancy”: All of “simple” plus:

- number of vertices (`nv`)
- tertiary vertex:
 - 3D decay length significance (`sdls3d2`)

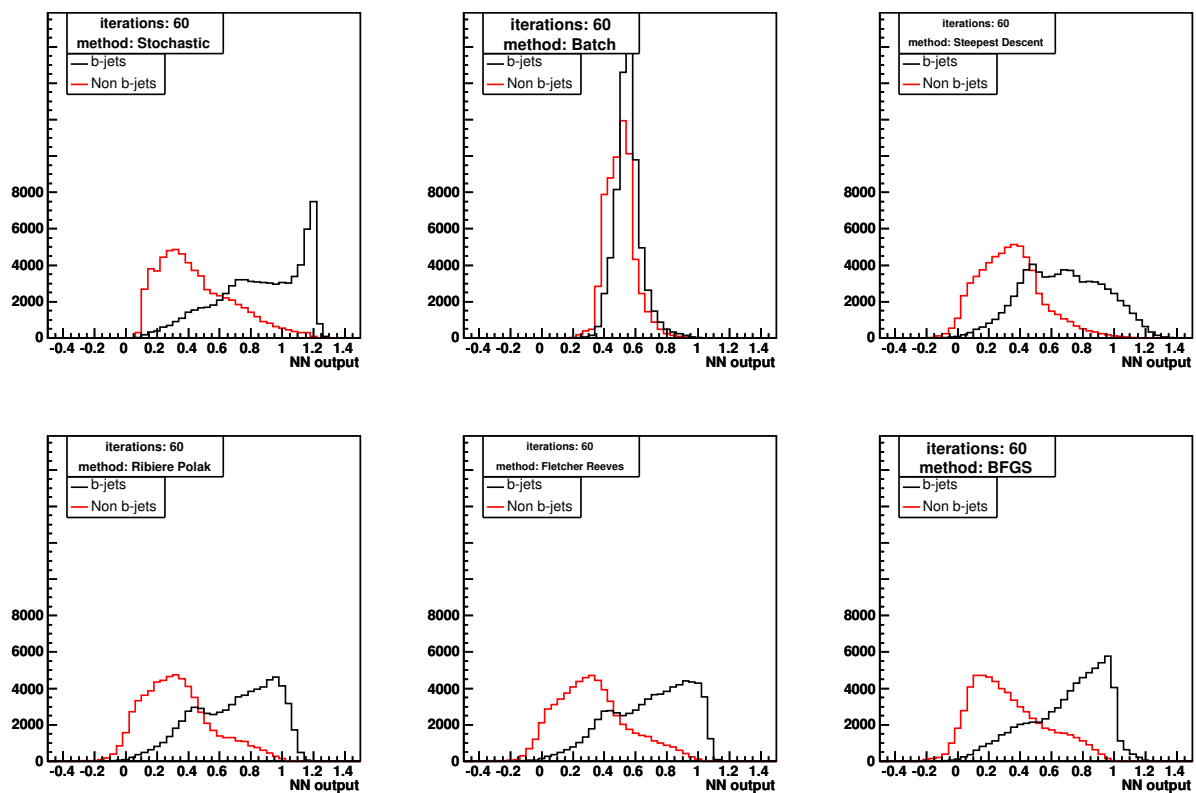


Figure 3: Signal and background outputs of the NN for the different training methods.

- ΔR (`dr2`)
- mass (`mass2`)
- χ^2 (`chi22`)
- multiplicity (`mult2`)

The purpose is to see if adding the tertiary vertex variables yields an improvement. Figures 5 and 6 show no improved performance with the tertiary vertex variables.

3.3 Hidden Neurons

The NNs were trained with varying number of hidden neurons. The result was that six hidden neurons for “simple” and 12 hidden neurons for “fancy” were best. That is, the same number of hidden neurons as variables.

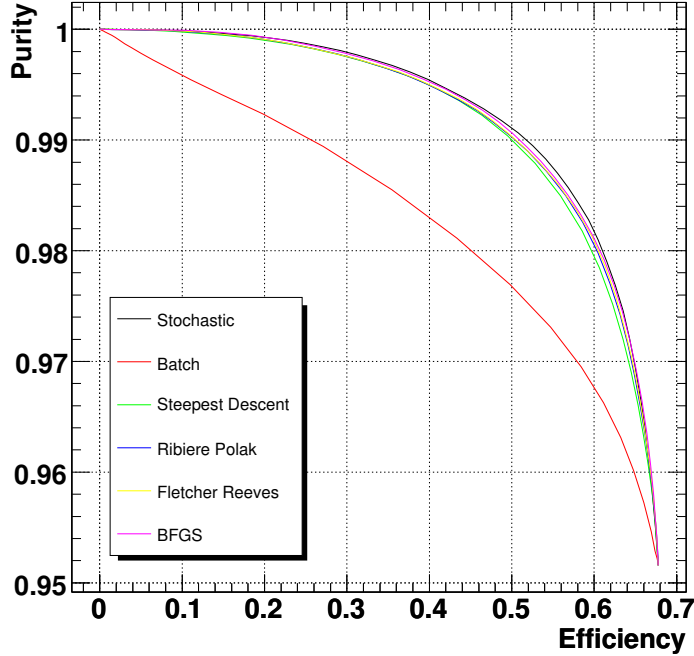


Figure 4: Purity versus efficiency for the different training methods.

3.4 Reconstruction Options

MC data from a total of five different reconstruction options were compared:

- All of Andy Haas' options for reconstruction (`cab3`).
- Default secondary vertexing options such as 2D track impact parameter significance (`cab_default_sv`).
- Default trackjet options (`cab_default-tj`).
- Default primary vertex options (`cab_noadapt`).
- No merging of overlapping trackjets (`cab_no-tj-merge`).

The result is that all but `cab_default_sv` had the same performance, and `cab_default_sv` had poor performance. Figure 7 compares their performances.

3.5 Neural Network Results

For b -jet detection, NNs provide an improvement over cutting on 3D decay length significance (`sdls3d`) alone. This is because the NN is able to learn the significances of the minor variables. Figure 8 compares the NN with the 3D decay length significance cut.

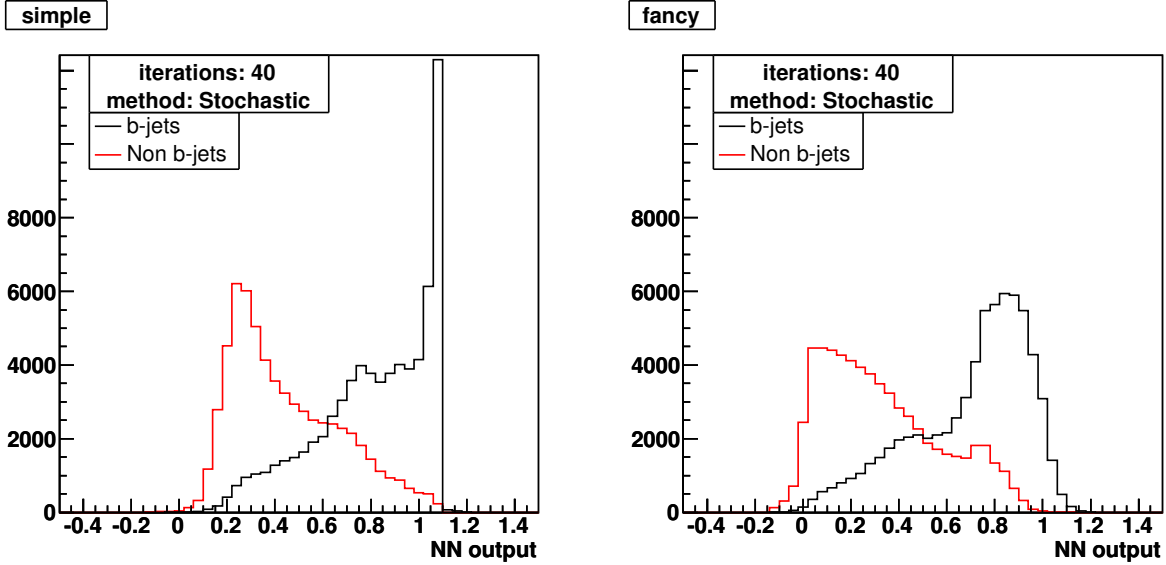


Figure 5: Signal and background outputs of the NN for the different variable sets.

4 $Z \rightarrow b\bar{b}$ Analysis

Using the neural network technique as previously described, the invariant mass of the Z boson was obtained from $Z \rightarrow b\bar{b}$ events from approximately 100 pb^{-1} of Run II data with the `cab3` reconstruction options. The NN output cut for tagging b -jets was set to yield an efficiency of 50% and purity of 99% in the MC data.

4.1 Background Subtraction

For this analysis, an understanding of the background in the double b -tagged data is essential. The background is essentially composed of heavy-flavor dijet production and mistagged gluon/light-quark jet production, and these events cannot be accurately simulated with current techniques [4]. If the background shape can be accurately derived, it can simply be subtracted from the data. The method used here is similar to [4] and makes use of a tag-rate function (TRF) to estimate the amount of double b -tagged background. Figure 9 shows the TRF derived from the single-tagged data. The TRF is then used to estimate the double-tagged background (Fig. 10). This estimated background is then subtracted from the double-tagged data to yield Figure 11.

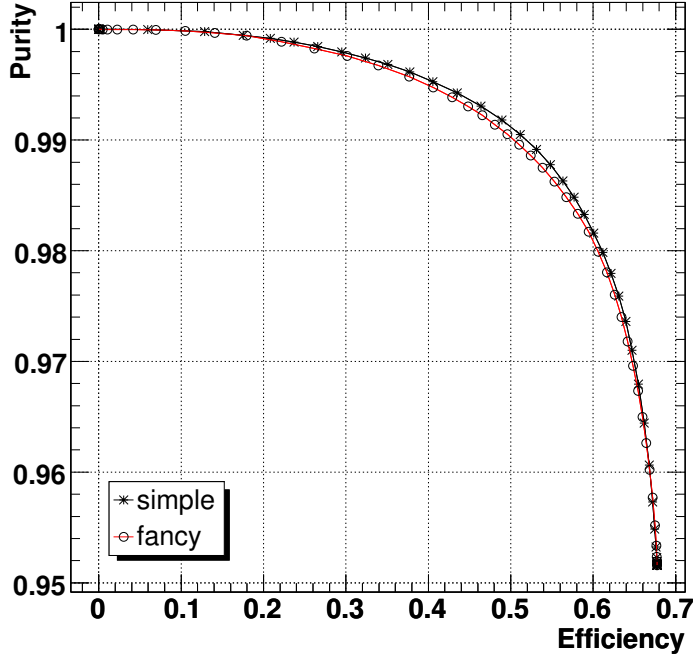


Figure 6: Purity versus efficiency for the different variable sets.

4.2 $0 \rightarrow 1$ Correction

There is one important correction we have made. The double-tagged data has more heavy-flavor jets than the single-tagged data which the TRF is applied to. This difference in heavy-flavor jets is called the $0 \rightarrow 1$ shift. The shift is measured using untagged data compared to the single-tagged data. This $0 \rightarrow 1$ correction is then subtracted from the expected double-tagged background. Figure 12 compares the uncorrected and $0 \rightarrow 1$ corrected Z peaks.

5 Conclusion

We have shown that NNs yield better b -jet detection performance than a decay length significance cut alone. It was also found that tertiary vertex NN variables don't yield an improvement, but the MC reconstruction options can make a difference. We note that there is much room for improvement for the background subtraction technique. Additional background correction techniques can be utilized as in [4]. Unfortunately, due to the short time-span of this REU, these correction techniques were not studied.

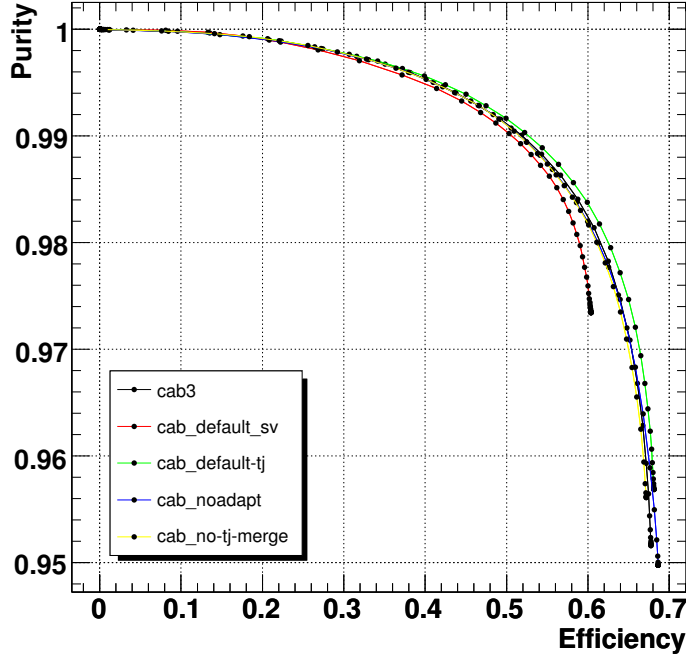


Figure 7: Purity versus efficiency for the different reconstruction options.

References

- [1] J. Erler and P. Langacker, in *Proceedings of the 5th International WEIN Symposium: A Conference on Physics Beyond the Standard Model*, Santa Fe, NM, 1998; e-print [hep-ph/9809352](#).
- [2] G. Altarelli, CERN report, CERN-TH/97-278; e-print [hep-ph/9710434](#).
- [3] P. C. Bhat, R. Gilmartin, H. B. Prosper, *A Strategy for Discovering a Low-Mass Higgs Boson at the Tevatron*, e-print [hep-ph/0001152](#).
- [4] A. Jenkins, P. Jonsson, G. Davies, A. Haas, *Observation of the $Z \rightarrow b\bar{b}$ Decay at $D\bar{O}$* , $D\bar{O}$ note in preparation.

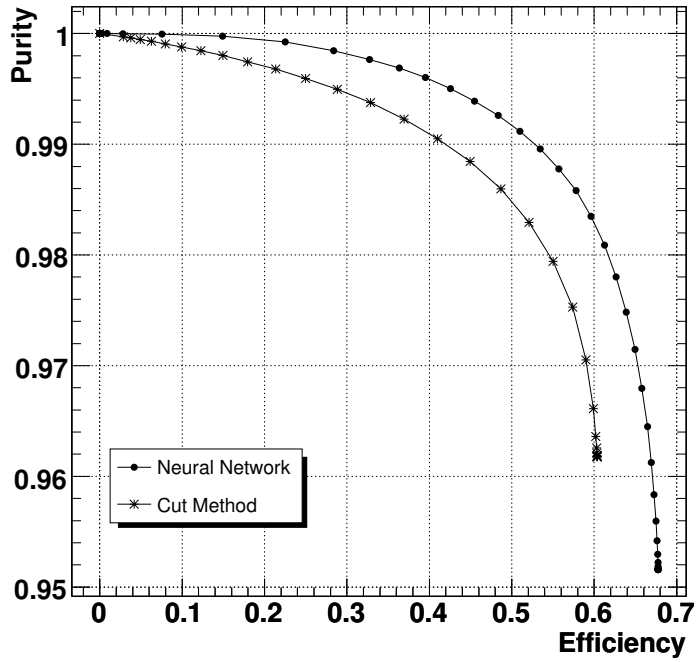


Figure 8: Purity versus efficiency for the neural network and the decay length significance cut method.

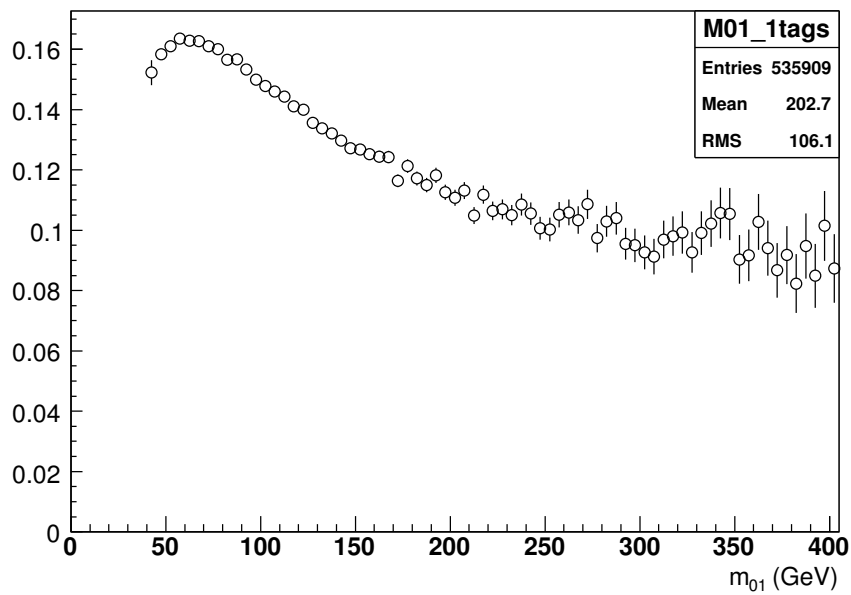


Figure 9: The TRF derived from the single b -tagged data used to estimate the double-tagged background.

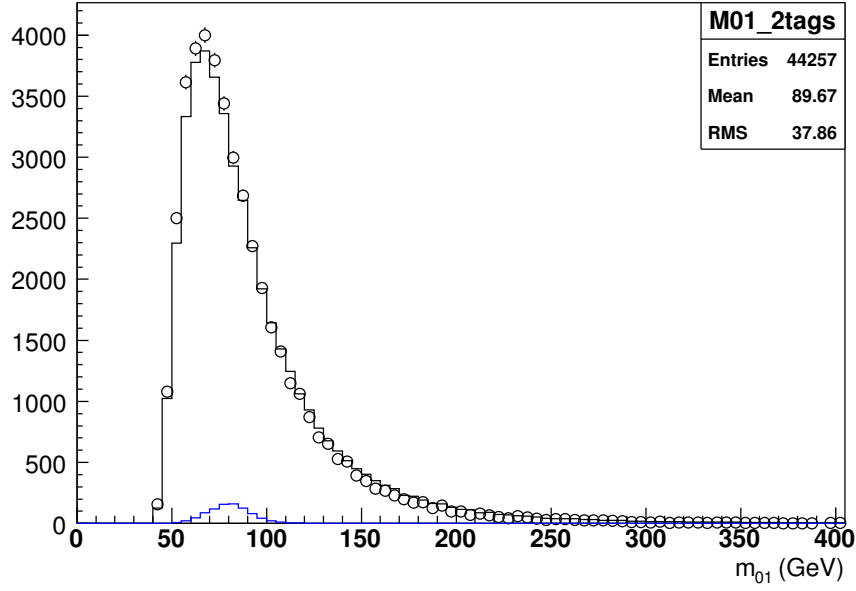


Figure 10: Comparison between double b -tagged data (points) and the expected background before any background corrections. The distribution from the MC is shown for comparison.

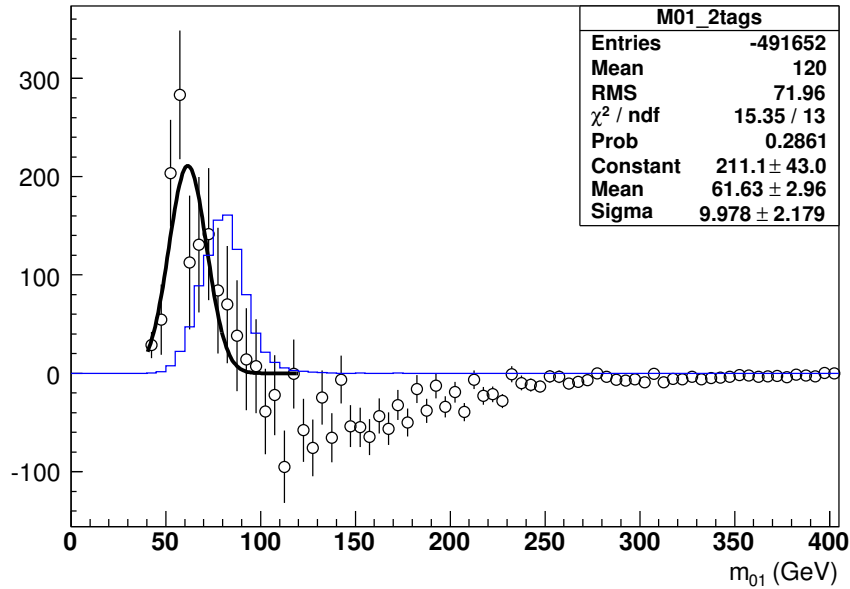


Figure 11: The $Z \rightarrow b\bar{b}$ peak derived from the data after no corrections. The distribution shape from the MC is shown for comparison.

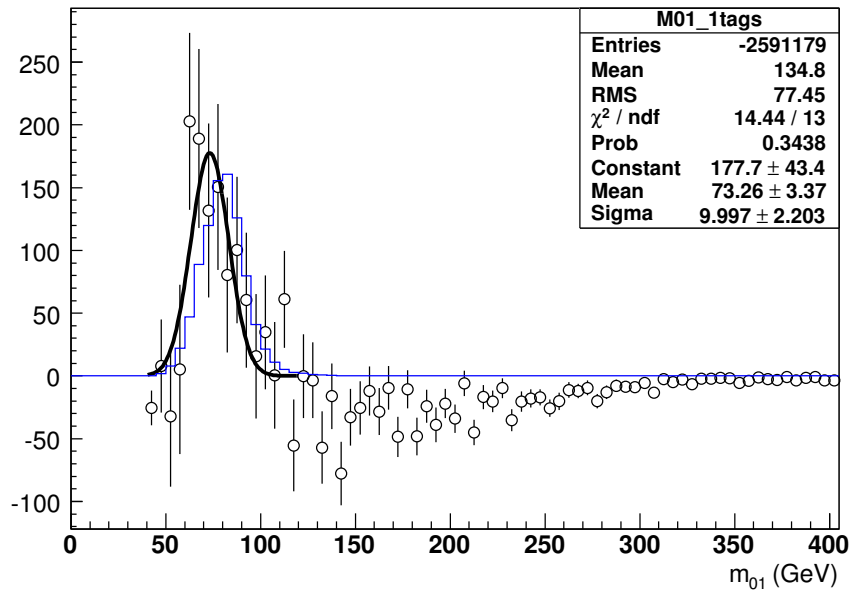


Figure 12: The $Z \rightarrow b\bar{b}$ peak derived from the data after the $0 \rightarrow 1$ correction. The distribution shape from the MC is shown for comparison.