Searching for Supersymmetric Higgs Bosons

Daniel C E Bunting & Amalia Madden

June 21, 2015

Abstract

1 Introduction

2 Theory

The standard model represents our current understanding of fundamental particles and their interactions. Its development stemmed from the discovery in the 1970s of a symmetry between the weak and electromagnetic forces which allowed them to be unified as a single force. The equations of this unified theory correctly described the electroweak force and its associated bosons, but with a serious issue that all emerged without mass. Whilst true for a photon, we know that the W and Z bosons have significant mass[1].

To solve this problem, a mechanism was proposed by theorists Robert Brout, Franois Englert and Peter Higgs, whereby particles are given mass through interaction with the Higgs field. Mass of a particle is dependent on the strength of its interaction with the field for example photons do not interact with it at all. The particle associated with the fundamental Higgs field is the Higgs boson[1].

Observation of a new particle with a mass of 126 Gev was announced on 4 July 2012 by the ATLAS and CMS experiments at CERN's Large Hadron Collider. This particle is thought to be consistent with the Higgs Boson, and lead to the award of the Nobel prize in 2013 to Franois Englert and Peter Higgs.

Quantum field theories such as the standard model comprise two basic types of field, fermions that lead to matter particles and bosons that lead to force-carrier. The spin-statistics theorem asserts that these two types are differentiated in that bosons have integer spins and fermions half-odd-integer spins. The creation of a symmetry that relates fermions and bosons forms the basis supersymmetry, whose major proponent is the existence of a new set of partner particles to those in the standard model. The superpartners differ by a spin of a half, so that bosons will be partnered with fermions and vice-versa[2].

The existence of these new particles could solve a major problem of fixing the mass of the Higgs Boson. The observed mass of the Higgs boson is puzzlingly light when interactions with the particles of the standard model should make it very heavy. The addition of the extra particles in supersymmetry could cancel out contributions to the Higgs mass from their Standard-Model partners.

The minimal supersymmetric standard model (MSSM) investigated here is the supersymmetry model with the minimum number of new particle states and interactions that are consistent with phenomenology[3]. Whereas in the Standard Model the Higgs field is introduced as a weak

isospin doublet, in the MSSM the presence of two of these doublets leads to five physical Higgs bosons, three neutral (collectively denoted as ϕ): h, H, and A ; and two charged: H^- and H^+ [4].

To describe the MSSM Higgs model, two free parameters are conventionally chosen - $\tan\beta$, the ratio of the vacuum expectation values of the two Higgs doublets, and the mass of A, one of the MSSM Higgs. Although $\tan\beta$ is a free parameter in MSSM, a large value of $\tan\beta \approx 35$ would explain the ratio of the top to bottom quark mass, and also provide good reason for the observed density of dark matter.

The value of $\tan\beta$ provides an enhancement factor for the couplings of the Higgs to fermions in the SM compared to the MSSM. The couplings of the Higgs in the MSSM are proportional to those in the SM and depend also on the type of quark and Higgs. For large $\tan\beta$, two of the Higgs Bosons A and either h or H have approximately the same mass. Whilst their coupling to down-type quarks is enhanced by $\tan\beta$ compared to the SM, their coupling to up-type quarks is reduced. Due to this enhancement, the most probable mode of decay for the three neutral Higgs bosons is $\phi \to b\bar{b}$ with a branching ratio near 90% (the next most probable being $\phi \to \tau\bar{\tau}$). Due to large multijet backgrounds a direct search for $\phi \to b\bar{b}$ is difficult. Instead, the case $\phi b \to b\bar{b}b$ where ϕ is produced in association with one b quark is used which.

3 Method

3.1 Event Simulation

The Monte Carlo simulation of the events used in the training and evaluation of the neural networks is beyond the scope of this project, so will only be briefly summarised here. Signal $\phi b \to b\bar{b}b$ events were simulated using the PYTHIA event generator [5] to leading order, with a mass hypothesis of $M_H = 110\text{GeV/c}^2$ and then corrected to next-to-leading order with MCFM [6]. The background multijet events were generated using ALPGEN [7]. The simulated events were then passed through a model of the D∅ detector [8]. Finally various preprocessing cuts were applied to the data excluding unlikely signal events, giving a final total of 53,960 events.

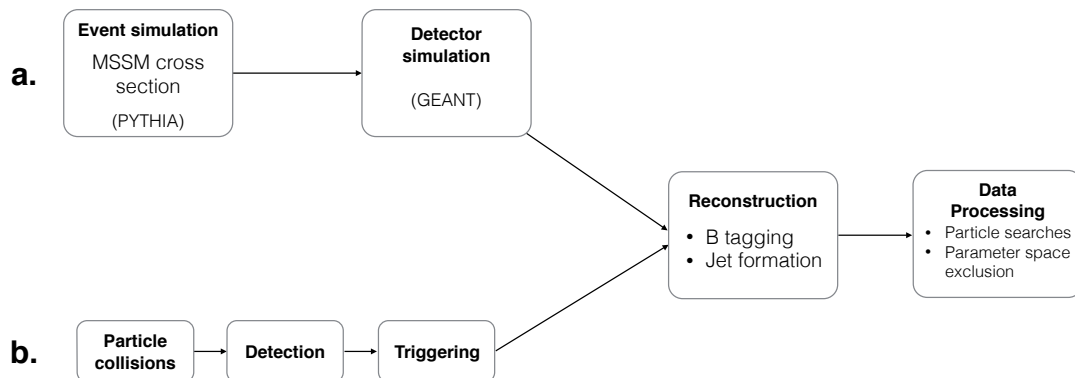
Figure 1: Steps in the information processing pipeline for a. simulated and b. measured data.

3.2 Feature Selection

In any scenario where statistical methods are used to interpret data selecting features in order to maximise the discriminatory or explanatory power of the model plays a critical role. From the b-tagged jets, those with the highest and second highest transverse momentum P_t were identified as being the putative product of the decaying Higgs boson, with all the following variable differences are considered to be taken between these leading jets. Following [4] seven feature variables were extracted from the raw simulation data, and used as inputs to the classification neural network.

Pseudorapidity difference $\Delta\eta$

Te pseudorapidity of a jet is defined as

$$\eta = -\ln\left[\tan\left(\frac{\theta}{2}\right)\right] \tag{1}$$

where θ is the angle between the jet and the beam axis. The pseudorapidity is preferred to θ as the difference between two pseudorapidity is Lorentz invariant.

Momentum balance $P_{balance}$

The momentum balance of the leading b jet pair is defined

$$P_{balance} = \frac{|p_1 - p_2|}{|p_1 + p_2|} \tag{2}$$

Sphericity S

The sphericity is defined as

$$S = \frac{3}{2}(\lambda_2 + \lambda_3) \tag{3}$$

where λ_2 and λ_3 are the second and third largest eigenvalues of the sphericity tensor $\hat{\mathbf{S}}^{\alpha\beta}$

$$\hat{\mathbf{S}}^{\alpha\beta} = \frac{\sum_i p_\alpha^i p_\beta^i}{\sum_i |p^i|^2} \tag{4}$$

3.2.1 Invariant Mass

$$M_H = \sqrt{\sum_{Dijet} E^2 - p^2} \tag{5}$$

As can be seen in Table 1 the invariant dijet mass is a strong predictor of whether an event is Higgs decay and including it in the ANN significantly increased its discriminatory power. However since the Higgs mass is not predicted by the Standard Model it must be hypothesised during the simulation stage, so in order to avoid training the ANN to depend too heavily on a particular value of M_H we tested models both with and without M_H

Additionally we define the **azimuthal angle difference** $\Delta\phi$, the **combined pseudorapidity** η_H, the **difference between the leading jet and the Higgs** $\eta_H - \eta_1$.

In order to evaluate the the feature selection the correlation matrices between the features for the signal and background events were calculated and shown in Figure 2. The correlations between each feature and the output variable were also calculated and are summarised in Table 1. The features selected for inclusion in a model should be both highly correlated with the
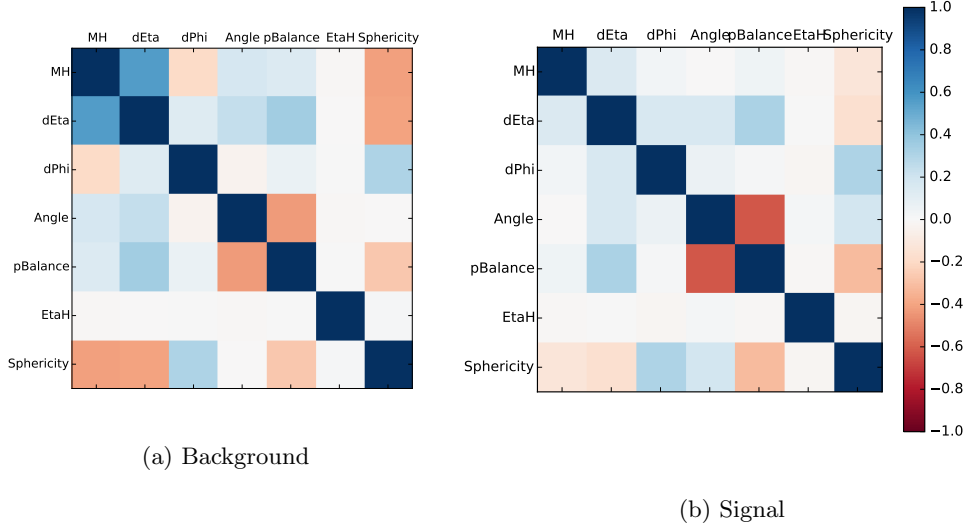
(a) Background

(b) Signal

Figure 2: Matrices of the Pearson product-moment correlation coefficients between each of the features

Feature	Correlation
$\Delta\eta$	0.329
M_H	0.204
$\Delta\Phi$	0.176
Sphericity	0.150
Momentum Balance	0.134
$\eta_H - \eta_1$	0.0343
η_H	0.00359

Table 1: The correlation between each of the features and the output variable

output and weakly correlated with each other. A high correlation between two features suggests redundancy, which should be avoided as unnecessarily complex models take longer and require more data (which may not be available) to train. η_H is a potential candidate for removal since it is very weakly correlated with the output.

3.3 Artificial Neural Networks

Artificial neural networks (ANNs) are a machine learning approach to classification/regression problems. They consist of stacked layers of neurons, each implementing a nonlinear activation function, interlinked with linear synapses. We utilise a class of feedforward ANNs called Multi-layer Perceptrons (MLPs), where activation function $g(x)$ is taken as a sigmoid and the output node a single logistic regression node

As a supervised machine learning problem, ANNs are trained by minimising the negative log likelihood (NLL) function $L(\mathbf{W})$ w.r.t. the parameters,

(a) Invariant mass

(b) $\eta_H - \eta_1$

(c) $\Delta\eta$

(d) $\Delta\phi$

(e) η_H
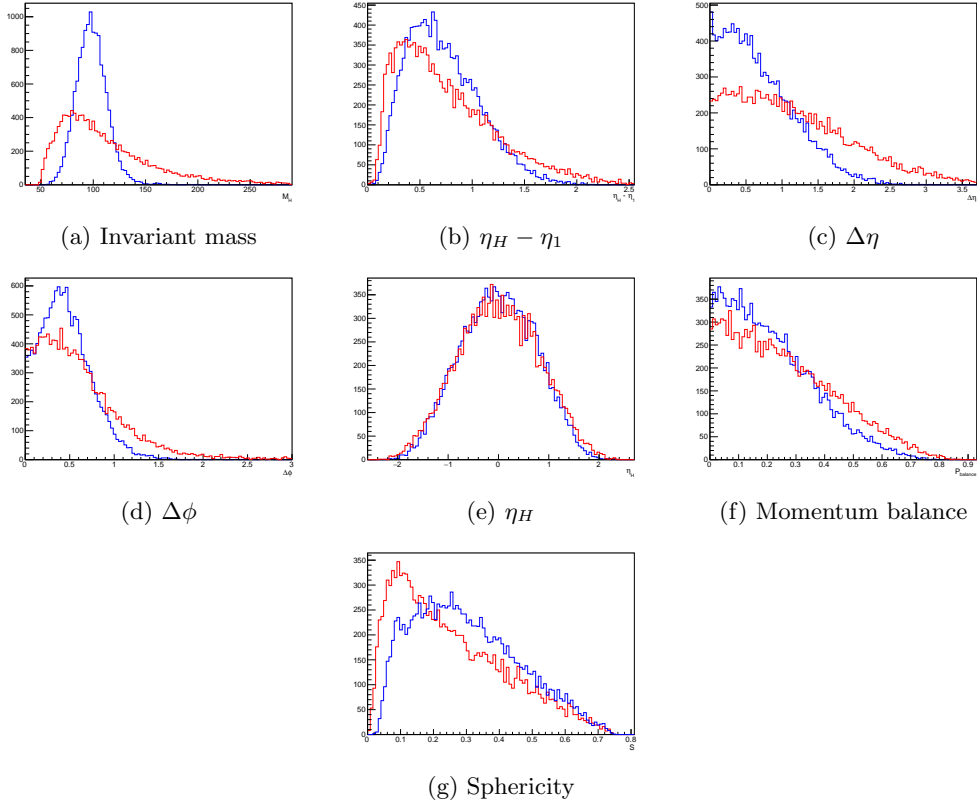
(f) Momentum balance

(g) Sphericity

Figure 3: Distributions of the input features. All y scales are events. The signal is in blue and the background in red.
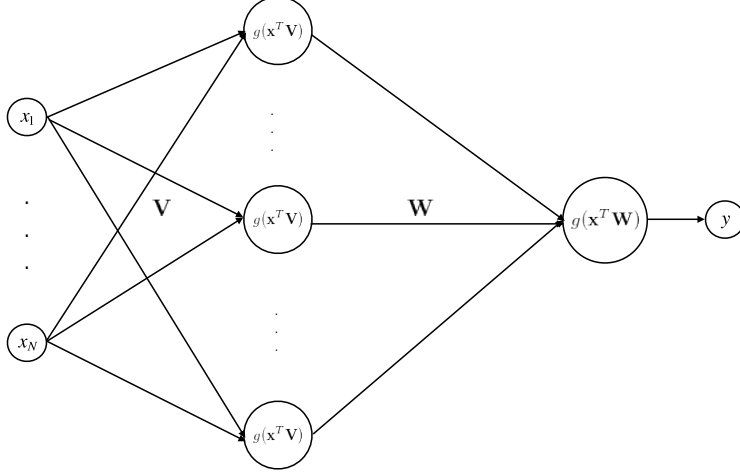
Figure 4: An example MLP with one hidden layer, N input variables $\mathbf{x}$ and single output $0 < y < 1$. The transfer function $g(\mathbf{x^T V}) = \frac{1}{1+e^{-\mathbf{x^T V}}}$ is sigmoid and the weight matrices are $\mathbf{V}$ and $\mathbf{W}$

$$L(\mathbf{W}) = -\log\left[P(\mathbf{W}|\mathcal{D})\right] \tag{6}$$

$$= -\sum_{i=1}^{N_{train}} \left[y^i \log \hat{y}^i + (1-y^i)\log(1-\hat{y}^i)\right] \tag{7}$$

where $\mathbf{W}$ denotes the weight matrices collectively and $\mathcal{D}$ the data. Due to the nonlinearity of the activation function the NLL is in general a non-convex function of $\mathbf{W}$, so it is minimised using random restarts of stochastic gradient descent. A strength of ANNs is that the gradient of the NLL can be efficiently computed by the backpropagation algorithm [9], which allows that gradient at the nth hidden layer to be calculated in terms of the error at the $(n + 1)$th layer, so computation can start at the output layer and work backwards. Once the gradient has been calculated the weights are optimised and the output estimates $\hat{y}$ updated by forward propagating the inputs through the network. A single backpropagation, optimisation, forward propagation cycle is called an epoch. In order to avoid overfitting, the input data is split and 50% reserved as a "test set" so that after each epoch the network can be evaluated on the unseen test set and the misclassification error $e = \frac{1}{N_{test}} \sum_{i=1}^{N_{test}} \mathbf{1}\{y^i \neq \hat{y}^i\}$ calculated. Training is stopped when e stops declining significantly, indicating that the network has learned all it can and further training is likely to decrease performance. We used the Cern ROOT [10] implementation of an MLP, `TMultiLayerPerceptron` due to its tight integration with the rest of the ROOT data analysis frame, despite it lacking some desirable features.

3.3.1 Model Selection

The performance of a neural network is highly dependent on both the number of nodes and the network topology. In theory by the Universal Approximation Theorem, an ANN with sufficiently

many nodes arranged in single hidden layer is able to approximate an arbitrary decision boundary. However in practice deep neural networks can perform better, so a variety of sizes and topologies were investigated and evaluated.

A commonly used criterion for measuring signal/background separation performance is the significance, $\frac{S(c)}{\sqrt{B(c)}}$ where $S(c)$ and $B(c)$ are the weighted number of signal and background events respectively, scoring above the signal acceptance cutoff c. Typically, as a function of $c \in (0,1)$, $\frac{S(c)}{\sqrt{B(c)}}$ has a maximum which can be used as a comparison.

Another evaluation parameter common in machine learning applications is the AUC, the area under the ROC (receiver operator characteristic) curve. The ROC curve is the parametric plot of the true positive rate (tpr) against the false positive rate (fpr), parameterised by the cutoff.

(a) $\frac{S}{\sqrt{B}}$

(b) ROC curve of the classifier

(c) Distribution of model scores for signal (red) and background (blue) events
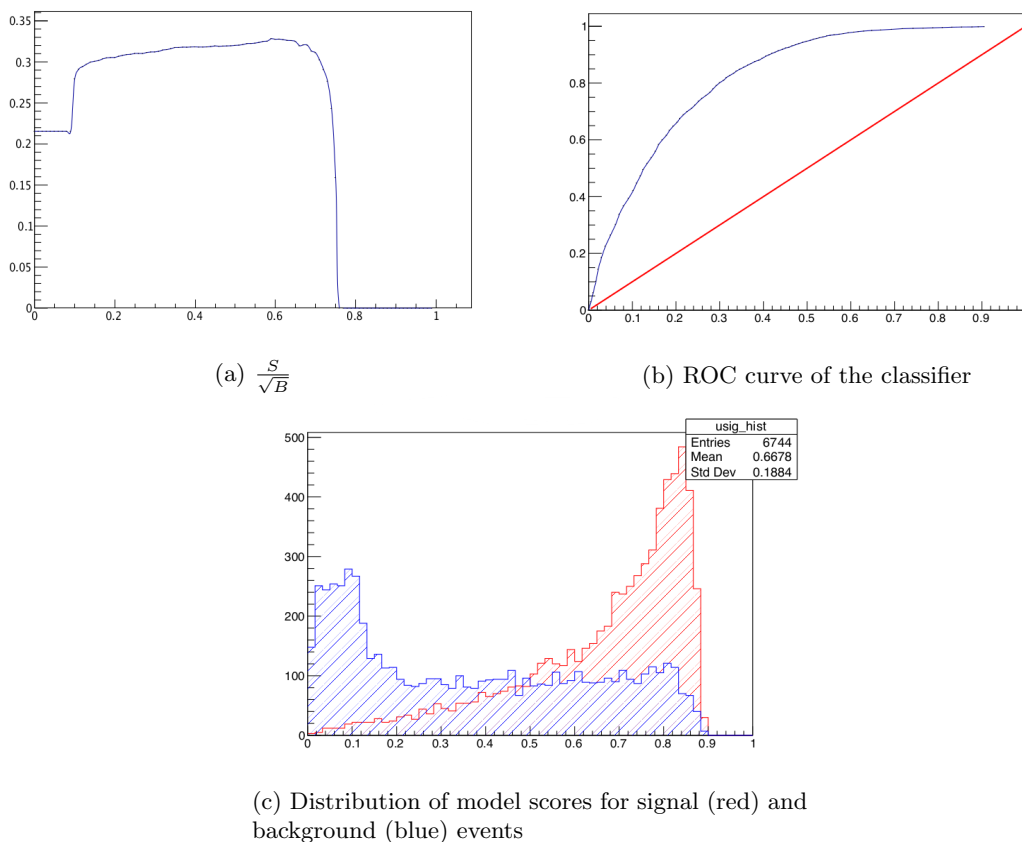
Figure 5: Diagnostic plots for a train ANN with two layers of seven and six hidden nodes

In order to evaluate the reliance of the network on the (hypothesised) invariant dijet mass, the network was trained and evaluated both with and without M_H over single layer configurations with between 1 and 50 hidden nodes. As expected Figure 6 shows that removing M_H as a feature had a significant negative effect on the accuracy of the network.

We also considered the effect of a second hidden layer on the discriminatory power of the ANN
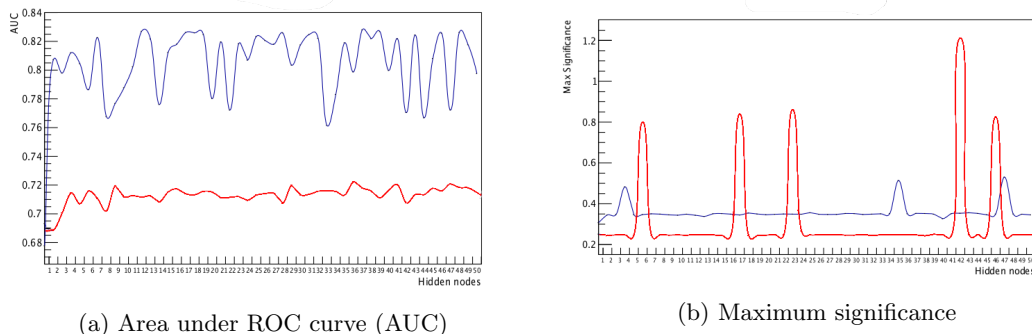
7

(a) Area under ROC curve (AUC)

(b) Maximum significance

Figure 6: The ANN evaluated with (blue) and without (red) the invariant dijet mass M_H as a feature

4 Conclusion

In this project artificial neural networks were used to identify Higgs boson decay events in the $b\bar{b}b$ channel for the purpose of identifying the excess of $\phi \to b\bar{b}$ decays predicted by minimal supersymmetric standard model compared to the standard model. A set of features describing the collision events measured at the D∅ were identified and used to train ANNs with a variety of topologies, with the goal of identifying an optimal configuration by cross validation. However this proved difficult due to the time required to train the networks and the apparent lack of a trend in the summary statistics (AUC and $\frac{s}{\sqrt{b}}$) we used to compare networks. A potential alternative approach would to start with large network and use a regularising Bayesian prior on the weights [9], or a separate algorithm such as Optimal Brain Surgery[11] to shrink the network to an optimal size.

Author Contributions

DB wrote section 3. AM wrote sections 1 and 2. We wrote section 4 together.

Source Code

Throughout the project we used GitHub for collaborating and version control and our entire source code tree can be found at `https://github.com/dnlbunting/tevatron-higgs/`.

References

[1] CERN. The origins of the brout-englert-higgs mechanism. http://home.web.cern.ch/topics/higgs-boson/origins-brout-englert-higgs-mechanism.

[2] King's College London Neil Lambert. Supersymmetry (cm439z/cmms40) course notes. http://www.mth.kcl.ac.uk/ lambert/SUSY.pdf.

[3] Howard Baer and Xerxes Tata. *Weak scale supersymmetry: From superfields to scattering events.* Cambridge University Press, 2006.

[4] The DØ Collaboration. Search for neutral higgs bosons in the multi-b-jet topology in of collisions at. *Physics Letters B*, 698(2):97 – 104, 2011.

[5] Torbjörn Sjöstrand, Stephen Mrenna, and Peter Skands. Pythia 6.4 physics and manual. *Journal of High Energy Physics*, 2006(05):026, 2006.

[6] John M Campbell and RK Ellis. Mcfm for the tevatron and the lhc. *Nuclear Physics B-Proceedings Supplements*, 205:10–15, 2010.

[7] Michelangelo L Mangano, Fulvio Piccinini, Antonio D Polosa, Mauro Moretti, and Roberto Pittau. Alpgen, a generator for hard multiparton processes in hadronic collisions. *Journal of High Energy Physics*, 2003(07):001, 2003.

[8] Rene Brun, Federico Carminati, Simone Giani, et al. Geant detector description and simulation tool. *CERN program library long writeup W*, 5013:1993, 1993.

[9] Kevin P Murphy. *Machine learning: a probabilistic perspective.* MIT press, 2012.

[10] Rene Brun and Fons Rademakers. Root - an object oriented data analysis framework. In *AIHENP'96 Workshop, Lausane*, volume 389, pages 81–86, 1996.

[11] Babak Hassibi, David G Stork, and Gregory J Wolff. Optimal brain surgeon and general network pruning. In *Neural Networks, 1993., IEEE International Conference on*, pages 293–299. IEEE, 1993.