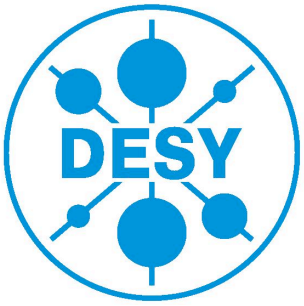

Multivariate optimization of the Higgs to gamma gamma analysis in the vector boson fusion production channel

Ana Rosario Cueto Gómez, Universidad Autónoma de Madrid

September 4, 2013

Abstract

Since the observation of the Higgs boson by ATLAS and CMS collaborations a big effort has been undertaken to study its properties. Their precise measurements are essential to verify the role of the Higgs boson in the electroweak symmetry breaking mechanism and mass generation. In this project the VBF production channel of the Higgs to di-photon decay is analyzed. We use PARADIGM decision-making framework to assess the selected parameter space. We find that interference between the baseline η_{j1}, η_{j2} and $\Delta\eta_{jj}$ variables reduces the performance of the classification process between signal and background and that proposed third jet kinematic variables have negligible effect in the results.

Contents

1 Introduction

The Standard Model explains how the basic building blocks of matter interact, governed by four fundamental forces. In the 1970's physicists realized that the weak and electromagnetic forces are unified in a single force: the electroweak force with four force carrying particles, the photon and the W and Z bosons. A problem remains, all this particles emerge without mass.

According to the Brout-Englert-Higgs mechanism the W and Z bosons acquire their mass by interacting with the "Higgs field" which permates the Universe. The more a particle interacts with this field, the heavier it is. As a quantized scalar field, the Higgs field has a massive, spinless and positive parity particle associated to it: the Higgs Boson.

The Higgs boson couples to all particles (including itself) proportional to their masses and tends to decay to the heaviest states allowed by the phase space. At low masses there is a loop-induced decay to two photons. To extract a Higgs signal from the backgrounds we measure the four momenta of the Higgs decay products and reconstruct the invariant mass of them. The signal peaks around m_H, while backgrounds are expected to be flat. The LHC detectors are designed to precisely measure the photon momentum and energy. The resolution on $m_{\gamma\gamma}$ is a factor of 10 better than any other decay channel (except for muons). Moreover, the photons do not decay further, and we can use all photon events.

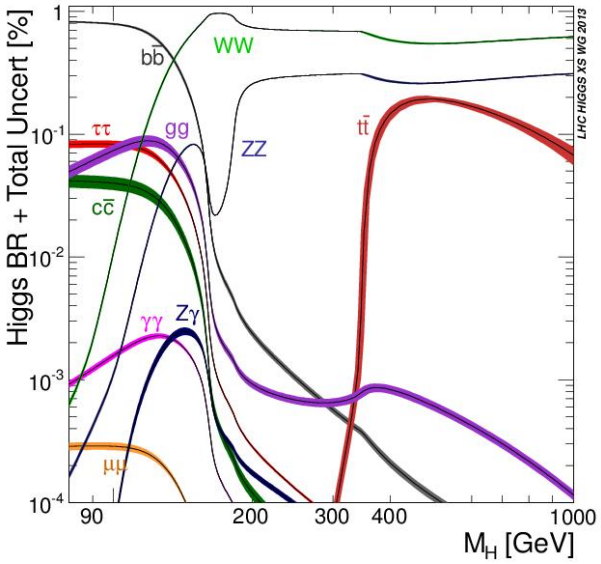

Figure 1: Standard Model Higgs boson decay branching ratios and total width

2 Vector Boson Fusion production channel

In the Vector Boson Fusion (VBF) production two longitudinally polarized bosons merge to produce a Higgs particle. It is one of the dominant production channels together with gluon-gluon fusion (ggF). However, ggF has a branching ratio 10 times bigger in the low mass region.

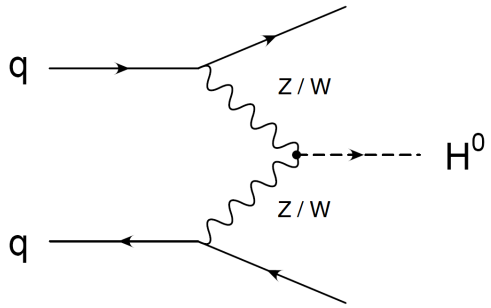

Figure 2: Vector Boson Fusion in the t-channel

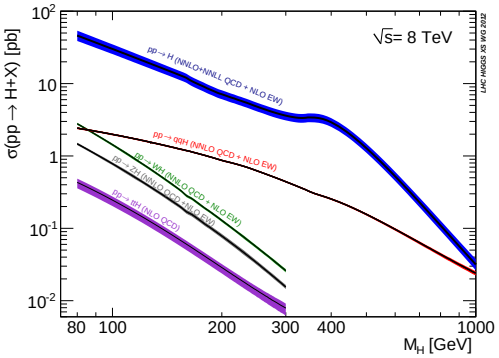

Figure 3: Theoretical branching ratios for the Higgs production channel at a energy of the center of mass of 8GeV

The VBF production channel is characterized by a particular topology that helps to suppress the background. Because of collinear enhancement (cross section diverges like $\log(pT)$), both jets tend to be forward in the detector. Additional jet activity is limited to collinear jets from the initial state and final-state quarks. The background $t\bar{t}$+jets can be suppressed with a veto to additional hard jets in the central region. The Higgs and its decay products are expected in the central pseudorapidity gap of the detector. The table in Fig. 4 shows the number of selected di-photon events in the data (N_D) and expected number of SM Higgs signal events (N_S) $m_H = 126.5GeV$ from the $H \to \gamma\gamma$ analysis, for each category in the mass range $100 - 160GeV$ at $\sqrt{s} = 8TeV$ [1].

$\sqrt{s}$	8 TeV						
Category	N_D	N_S	$gg \to H$ [%]	VBF [%]	WH [%]	ZH [%]	ttH [%]
Unconv. central, low p_{Tt}	10900	51.8	93.7	4.0	1.4	0.8	0.2
Unconv. central, high p_{Tt}	553	7.9	79.3	12.6	4.1	2.5	1.4
Unconv. rest, low p_{Tt}	41236	107.9	93.2	4.0	1.6	1.0	0.1
Unconv. rest, high p_{Tt}	2558	16.0	78.1	13.3	4.7	2.8	1.1
Conv. central, low p_{Tt}	7109	33.1	93.6	4.0	1.3	0.9	0.2
Conv. central, high p_{Tt}	363	5.1	78.9	12.6	4.3	2.7	1.5
Conv. rest, low p_{Tt}	38156	97.8	93.2	4.1	1.6	1.0	0.1
Conv. rest, high p_{Tt}	2360	14.4	77.7	13.0	5.2	3.0	1.1
Conv. transition	14864	40.1	90.7	5.5	2.2	1.3	0.2
Loose high-mass two-jet	276	5.3	45.0	54.1	0.5	0.3	0.1
Tight high-mass two-jet	136	8.1	23.8	76.0	0.1	0.1	0.0
Low-mass two-jet	210	3.3	48.1	3.0	29.7	17.2	1.9
$E_{\mathrm{T}}^{\mathrm{miss}}$ significance	49	1.3	4.1	0.5	35.7	47.6	12.1
One-lepton	123	2.9	2.2	0.6	63.2	15.4	18.6
All categories (inclusive)	118893	395.0	88.0	7.3	2.7	1.5	0.5

Figure 4: Statistics in the $H \to \gamma\gamma$ decay channel

As Fig. 4 shows, only the 7.3% of all the data events correspond to VBF, but in categories involving two high-mass jets becomes the dominant process. Due to the lack of statistics, it is essential to use multivariate techniques in this channel.

3 Baseline Analysis

In 2012 a multivariate analysis of this channel was carried out with eight discriminating variables ref[1]. The variables are required to have little or no correlation with the mass of the di-photon system ($m_{\gamma\gamma}$). The modeling of the background is based on Monte Carlo samples where the $\gamma-\gamma$, $\gamma+jet$, $jet+jet$ and the Drell-Yan (produced with SHERPA) are combined according to the fractions determined by data driven-measurements. It also includes jets with inverted isolation.

Variables selected for the analysis are listed below and shown in Figures 5-12.

- Dijet mass m_{jj}
- $pTt_{\gamma\gamma}$
- Pseudorapidity of the leading jet η_{j1}
- Pseudorapidity of the subleading jet $\eta_j 2$
- Pseudodrapidity separation between the leading and subleading jet $\Delta\eta_{jj}$
- Zeppenfelder variable $Zepp = \eta_{jj} - \frac{\eta_{j1}+\eta_{j2}}{2}$
- $\Delta\phi$ between Higgs and dijet system $\Delta\phi_{\gamma\gamma,jj}$
- $\Delta R_{min}(\gamma, jet)$

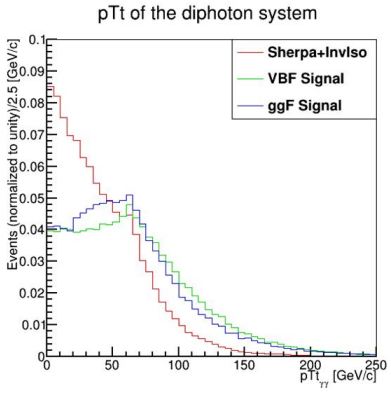

Figure 5: $pTt_{\gamma\gamma}$ normalized to unity

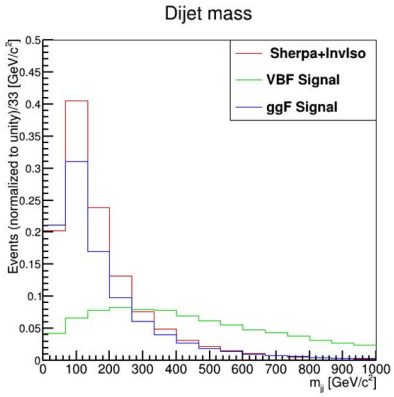

Figure 6: m_{jj} normalized to unity

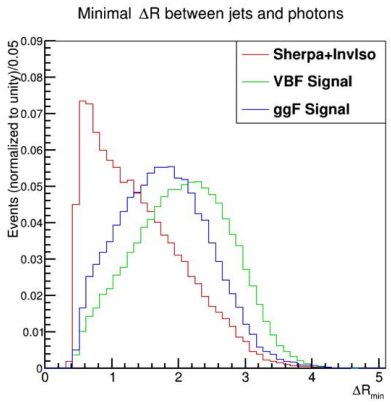

Figure 7: ΔR_{min} normalized to unity

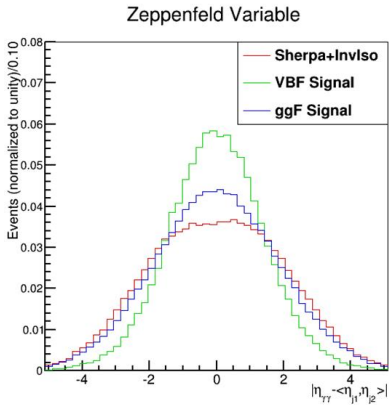

Figure 8: $Zepp$ normalized to unity

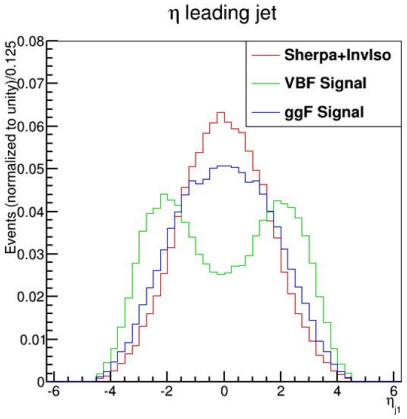

Figure 9: η_{j1} normalized to unity

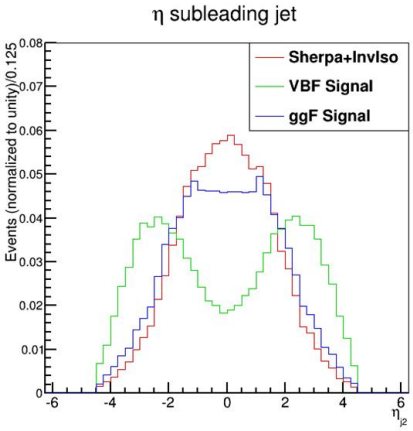

Figure 10: η_{j2} normalized to unity

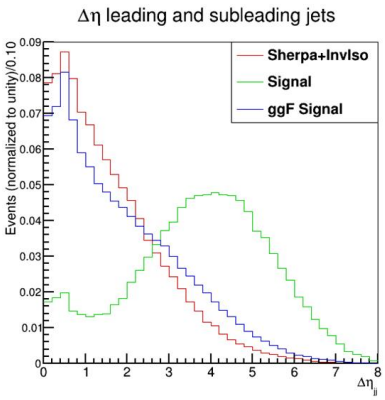

Figure 11: $\Delta\eta$ normalized to unity

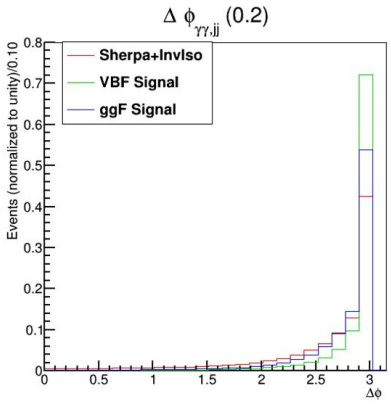

Figure 12: $\Delta\phi_{\gamma\gamma,jj}$ normalized to unity

As can be seen in Figures 5-12 the ggF production mode is similar to the background, therefore separating the signal from the QCD background and inverted isolation requirements also works to separate VBF from ggF background. Signal and background are normalized to unity so that the shapes can be easily compared.

3.1 New Variables

With the purpose of improving the classification between signal and background additional variables related with the kinematics of a third jet were proposed. They are listed below:

- pT_{j3}
- η_{j3}
- ϕ_{j3}
- $\Delta\phi_{j1j3}$
- $\Delta\phi_{j2j3}$

The following two variables were also included: $pT_{jj,\gamma\gamma}$ and $\Delta\phi_{jj}$. The plots of all the variables are shown in Figures 13-19

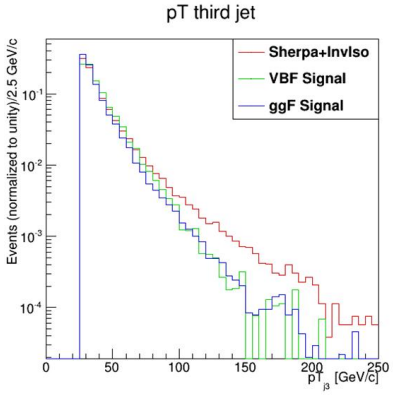

Figure 13: pT_{j3} normalized to unity

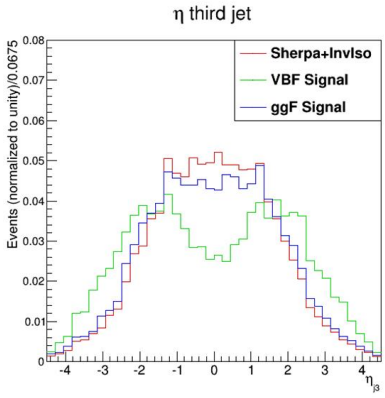

Figure 14: η_{j3} normalized to unity

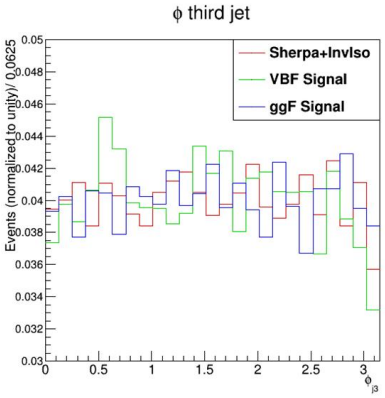

Figure 15: ϕ_{j3} normalized to unity

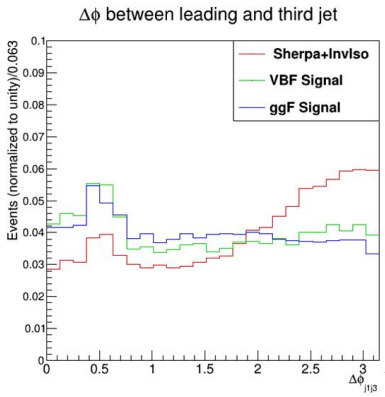

Figure 16: $\Delta\phi_{j1j3}$ normalized to unity

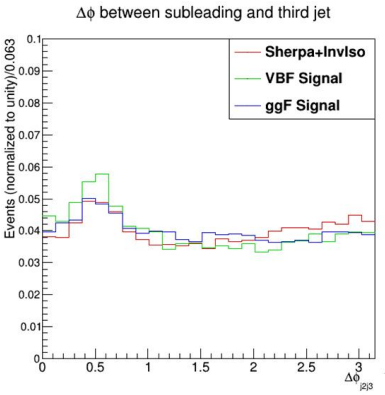

Figure 17: $\Delta\phi_{j2j3}$ normalized to unity

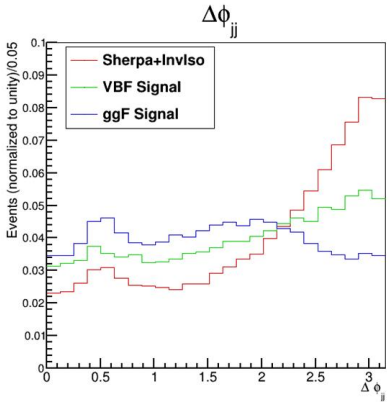

Figure 18: $\Delta\phi_{jj}$ normalized to unity

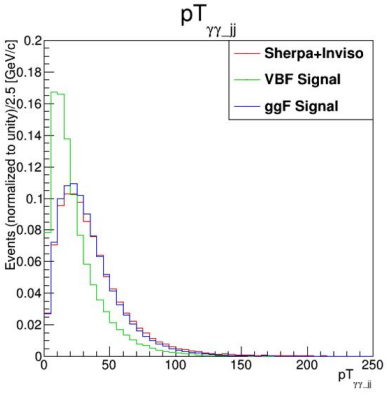

Figure 19: $pT_{\gamma\gamma,jj}$ normalized to unity

4 PARADIGM

We apply PARADIGM decision-making framework [2] oriented to High-Energy Phisics to this analysis. One of the main features of PARADIGM is Gleyzer-Prosper variable importance, described in Section 4.3.1, and Global Loss Function described in Section 4.3.2. It is classifier independent and can help in classifier selection. We use Gleyzer-Prosper variable importance to evaluate the parameter space selection are investigate further the possibility to reduce the parameter space using the Global Loss Function.

4.1 Classifiers

Classification aims to properly categorize an instance (event) to a predefined set of classes based on its attributes. In our analysis we train two similar types of classifiers: decision trees (DT) and boosted decision trees (BDT).

4.1.1 Decision Trees

A decision tree is one of the most popular classifiers. Each attribute measures some important feature of the class. The inputs are a collection of attributes whose class is known. The output is the classification decision to be made: signal or background. In order to classify an event, we start at the root of the tree. Each attribute is evaluated based on how well it separates signal from background in the training sample. The best attribute is selected and used at the root node of the tree to create two new decision branches. The process is repeated for each daughter branch to make further branches. The process continues until a leaf is encountered, at which time the object is asserted to belong to a particular class.

We use c4.5 [3] to construct decision trees used in the analysis and a sample decision tree is shown in Figure 20

4.1.2 Boosted decision trees

Boosting is a well-known algorithm for enhancing the performance of decision trees [5]. The construction of boosted decision trees consists of several iterations. In each iteration, greater weight is assigned to misclassified events of the previous round. The result of the process is an ensemble of weighted decision trees.
In our analysis boosted decision trees are generated using TMVA [4] with the following characteristics

```
NTrees = 400 :  nEventsMin = 100 :  BoostType = Grad :  Shrinkage = 0:05 :
   UseBaggedGrad :  GradBaggingFraction = 0:5 :  nCuts = 30 :  MaxDepth = 3
```

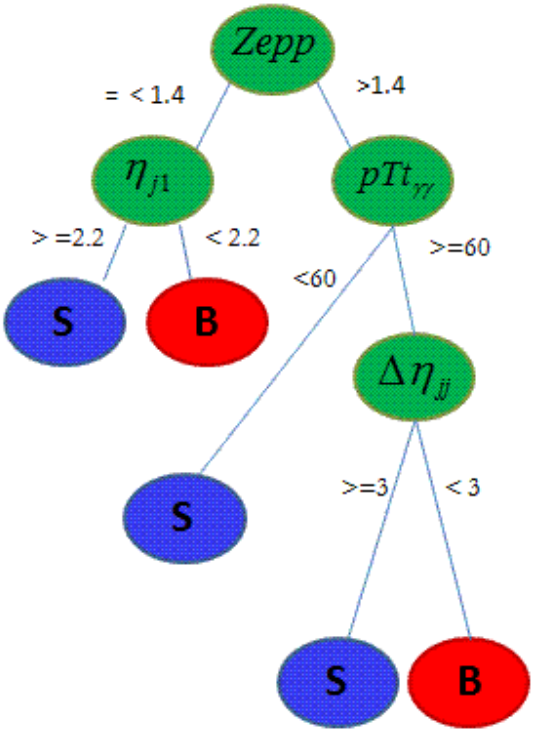

Figure 20: Example decision tree. The decision nodes are colored in green and the class names are blue for the signal and red for the background

These are the same options that were used in the original VBF analysis [1], for a reduced number of trees to minimize computation time without sacrificing more than 2/1000 in the performance measure of the classifier. Also, following [1], cuts $|\Delta\eta_{jj}| > 2$ and $|Zepp| < 5$ were applied in order to focus on the region of the phase space where most of the signal events are present.

4.2 The ROC curve

After a classifier is built, it is desirable to evaluate its accuracy. The data are split into two parts: the training and the testing set. We use the testing set to evaluate the classifier performance.

The Receiver Operating Characteristic (ROC) curve shows the background rejection rate versus signal efficiency. The area under the ROC curve is calculated

as

$$AUC = \frac{1}{2}\left(\sum_{k=1}^{n}(\epsilon_{s_k} - \epsilon_{s_{s-1}})(\epsilon_{b_k} + \epsilon_{b_{k-1}})\right) \tag{1}$$

and is a commonly-used measure of the performance of the classifier. The ROC curve for the eight baseline variables is shown in Figure 4.2. When an instance is classified without any previous knowledge in a binary classification, this area is 0.5. The maximum AUC value is 1 which corresponds to a perfect classifier.

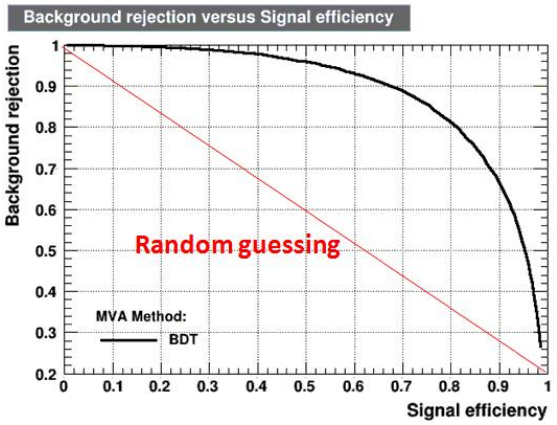

Figure 21: ROC curve for BDT with 8 baseline variables in TMVA

4.3 A decision making-framework

4.3.1 The Gleyzer-Prosper Variable Importance

Gleyzer-Prosper variable importance (GPI) is useful to understand which is the relevance of a particular variable to the classification task.

Given an initially-chosen variable set $\{V\} = \{X_1, ..., X_N\}$, the GPI of variable X_i is defined in the following way:

$$GPI(X_i) \equiv \sum_{S \subseteq V; X_i \in S} F(S) \cdot W_{X_i}(S) \tag{2}$$

where $\{S\} = \{X_i, ..., X_j\}$ are subsets that include variable X_i, $F(S)$ is the performance measure of the classifier built with $\{S\}$, and W_{X_i} is defined as

$$W_{X_i} \equiv 1 - \frac{F(S - X_i)}{F(S)} \tag{3}$$

It accounts for individual variable's share of the classifier performance $F(S)$ if the variable is removed from the classifier. GPI are normalized by

$$N \equiv \sum_{X_i} F(S) \cdot W_{X_i}(S) \tag{4}$$

Often in classification one encounters variable interactions.

4.3.2 Global Loss Function

Global Loss Function is an information measure that takes into account variable interactions. It is specific to variable reduction and measures the predictive power loss relative to an upper bound of performance of classifiers that remain:

$$GF(S') \equiv 1 - \frac{1 - \sum_{S \subseteq V - S'} F(S)}{2^{|V-S'|}} \tag{5}$$

where $S' \subset V$ is the subset considered for reduction. The absolute scale limit is given by:

$$\sum_{S \subseteq \{V-S'\}} F(S)_{max} = 2^{|V-S'|} \tag{6}$$

5 Results

5.1 ROC Integrals

For a full set of variables

$$\{pTt_{\gamma\gamma}, m_{jj}, Zepp, \Delta R_{min}, \Delta\phi_{jj,\gamma\gamma},$$
$$y_{j1}, y_{j2}, \Delta y_{jj}, pT_{j3}, \eta_{j3},$$
$$\phi_{j3}, \Delta\phi_{j1,j3}, \Delta\phi_{j2,j3}, \Delta\phi_{jj}, pT_{\gamma\gamma,jj}\}$$

the ROC results for DT and BDT are shown for four different cardinalities in Figure 22

The maximum ROC values are around 0.9 and there is still potential for improvement by selecting better-discriminating variables. As Figure 22 shows, boosting affects more the subsets with higher cardinality, creating a gap in the distribution.

5.2 Gleyzer-Prosper Variable Importance

Gleyzer-Prosper variable importance results for decision trees and boosted decision trees are shown in Figures 23 and 24. Variables with a positive value of GPI are shown in green while those with a negative value are shown in red. In the x-axis, the colors differentiate between the baseline variables (black), the third jet variables (red) and others (blue).

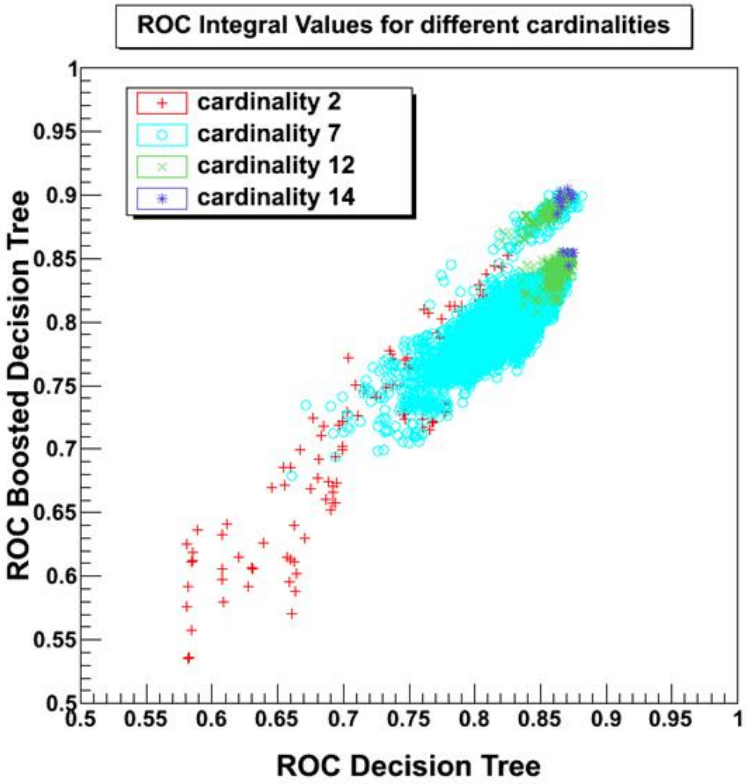

Figure 22: ROC integral for BDT and DT for cardinalities 2,7,12,14

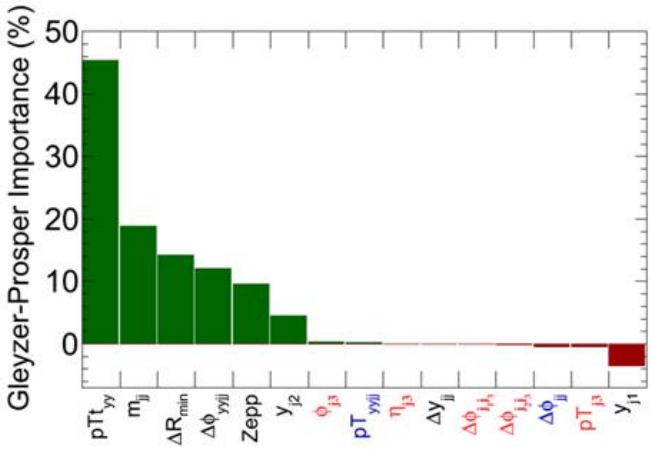

Figure 23: Gleyzer-Prosper importance for decision trees

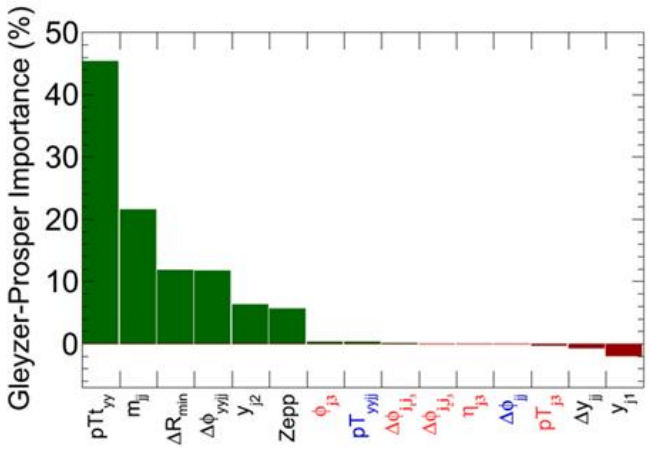

Figure 24: Gleyzer-Prosper importance for boosted decision trees

As can be seen in figures 23 and 24, variables associated with the third jet do not significantly contribute to the analysis. Moreover, one of the variables in the baseline analysis (η_{j1}) shows negative importance, implying a presence of significant interactions between it and other variables of the set. This is not surprising since η_{j2} and $\Delta\eta_{jj}$ are both present as well.

5.3 Global Loss Function

Figure 25 the global loss function is shown for decision trees. The x-axis shows an index associated with each of the subsets S', while the y-axis shows the amount of information loss of the classifier if the subset is removed.

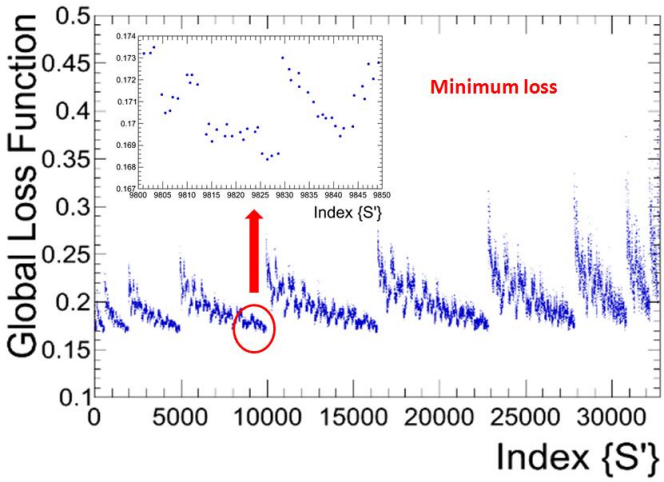

Figure 25: Global Loss function for decision trees

6 Evaluation and conclusions

We have analyzed the parameter space for the Vector Boson Fusion production channel of the Higgs to di-photons decay channel.

We see that the proposed parameter space is not as optimal as it can be. The maximum values of the area under the ROC curve are around 0.9. There is strong potential for improvement from choosing additional variables that improve classification performance and by removal of existing variables with strong negative interactions.

Three of the originally chosen variables η_{j1}, η_{j2} and $\Delta\eta_{jj}$ are not independent and show interference in classification causing classifier performance loss when all are present, as shown by the negative GPI of η_{j1}. One of these three variables should be removed, most likely η_{j1}.

We observe that boosting has a greater effect on sets with higher cardinality, as shown by the ROC gap in Figure 22.

Finally the addition of third jet variables does not contribute to an improvement in classification performance as shown by Figures 23 and 24. Their GPI values are very close to zero, meaning they are not meaningful for the analysis.

7 Future work

We are currently studying additional variables such as ϕ^* and other angular variables that can be used to further separate the VBF signal from the background and improve classifier performance of this production channel.

After this variables are added, we will use the global loss function, available in PARADIGM, to select the most optimal variable set to be used.

References

[1] Measurementes of the properties of the Higg-like boson in the two photon decay channel with the ATLAS detector using 25 fb^{-1} proton-proton collision data. *ATLAS collaboration*

[2] PARADIGM, a Decision-Making Framework for Variable Selection and Reduction in High Energy Physics, Proceedings of the XII. International Workshop on Advanced Computing and Analysis Techniques in Physics. Research, Erice, Italy, November 3-7, 2008 *Sergei V. Gleyzer, Harrison B. Prosper*

[3] c4.5: Programs for machine learning *J. Ross Quinlan (University of Sydney)*

[4] TMVA - Toolkit for Multivariate Data Analysis *A. Hoecker, P. Speckmayer, J.Stelzer, J.Therhaag, E. von Toerne, H. Voss, M. Backes, T.Carli, O.Cohen, A. Christov, D. Dannheim, K. Danielowski, S. Henrot-Versille, M. Jachowski, K.Kraszewski, A. Krasznahorkay Jr., M. Kruk, Y. Mahalalel, R. Ospanov, X. Prudent, A. Robert, D. Schouten, F. TEgenfeldt, A. Voigt, K. Voss, M. Wolter, A. Zemla*

[5] A Decision-Theoretic Generalization of on-Line Learning and an Application to Boosting *Yoav Freund, Robert E. Shapire*