

# Bayesian non parametric modelling of Higgs pair production

Bruno Scarpa<sup>1,a</sup> and Tommaso Dorigo<sup>2</sup>

<sup>1</sup>*Department of Statistical Sciences, University of Padova*

*Via Cesare Battisti 241 - 35121 Padova, Italy*

<sup>2</sup>*INFN, Padova*

**Abstract.** Statistical classification models are commonly used to separate a signal from a background. In this talk we face the problem of isolating the signal of Higgs pair production using the decay channel in which each boson decays into a pair of  $b$ -quarks. Typically in this context non parametric methods are used, such as Random Forests or different types of boosting tools. We remain in the same non-parametric framework, but we propose to face the problem following a Bayesian approach. A Dirichlet process is used as prior for the random effects in a logit model which is fitted by leveraging the Polya-Gamma data augmentation. Refinements of the model include the insertion in the simple model of P-splines to relate explanatory variables with the response and the use of Bayesian trees (BART) to describe the atoms in the Dirichlet process.

## 1 Introduction

After the 2012 discovery of the Higgs boson by the LHC experiments, a lot remains to be understood about the phenomenology of the newly found particle. In this context, a particularly interesting question is the value of the Higgs boson self-coupling constant, which can be determined by studying the production of Higgs boson pairs in proton-proton collisions. Here we are interested in analysing data collected by the Run 1 of the CMS experiment at the LHC (Large Hadron Collider) in 2012, and our goal is to fit a statistical model to isolate the signal of Higgs boson pairs decays in the final state characterised by 4 jets of  $b$ -quark, that is the particular decay channel in which each boson decays into a pair of  $b$ -quarks, with a final state featuring four hadronic jets.

Although this decay channel is the most frequent, since it is verified in about 33% of cases, it suffer of a quite large background noise that make difficult to isolate the signal (LHC Higgs Cross Section Working Group, 2012). Data collected from Run 1 of CMS correspond to a luminosity of 20 inverse femtobarns of proton-proton collisions collected at a center-of-mass energy of 8 TeV; this dataset is not sufficient to observe the presence of signal over the background events. In order to identify a good model we considered a Monte Carlo simulation to create events representing the signal (Johan Alwall et al., 2011 e Jun Gao et al., 2014).

Once cleaned the data set, we first applied the main statistical learning classification tools, such as random forests, boosting decision trees and gradient boosting, both by considering all the possible feature variables defined from the kinematical event characteristics and including only a limited

---

<sup>a</sup>e-mail: [scarpa@stat.unipd.it](mailto:scarpa@stat.unipd.it)

number of variables chosen by physicists. As a second approach we applied some Bayesian non parametric models, such as the non parametric logistic Bayesian model, Bayesian  $P$ -splines and BART (Bayesian additive regression trees).

As pre-analysis we applied on both background and simulated data the  $b$ -tagging multivariate algorithm, known as CMVA (Das et al., 2013).

## 2 Preprocessing

In this paper we consider the decay channel of each boson in a pair of  $b$ -quarks, with a final state of 4 hadronic jets. According to the Standard Model the non-resonant di-Higgs production from proton-proton collisions has unfortunately a very low cross section of about 10 fb at 8 TeV of center-of-mass energy (de Florian and Mazzitelli, 2013) which is well below the sensitivity of the current acquired data. In fact only one collision over 10000 billions will generate a pair of Higgs. Therefore, we need to isolate the decay signal of our interest. We consider the data collected by CMS during the LHC Run I in 2012 including 1259973 observations. In addition to these data, we consider a Monte Carlo simulation of 300000 signal events (Alwall et al., 2011 and Gao et al., 2014). Therefore, the considered dataset is the union of two types of data: the sampled form CMS, representing the background, and the simulated ones, representing the signal.

The first step of analysis consist in the identification of the observations corresponding to the hadronization of  $b$ -quark. A characteristic of the decay allows to distinguish jets corresponding to  $b$ -quarks, said  $b$ -jets, and possible other jets generated during the collision. In fact, the lifetime of hadrons containing  $b$ -quarks, even if very short, allow them to run a short distance before decay; from the experimental point of view this means that part of the particles of  $b$ -jets originate from a secondary vertex which is displaced from the primary one. This property has a key role in the identification of Higgs boson, and a number of  $b$ -tagging algorithms have been proposed to identify the presence of a secondary vertex (Weiser, 2006).

A multivariate CMVA  $b$ -tagging algorithm (Das et al., 2013) has been applied both to signal and background data. The algorithm implements neural networks to combine outputs from other  $b$ -tagging algorithms, by giving, for each jet, a scalar ranging between 0 (not  $b$ ) and 1 (most likely  $b$ ). We selected the events corresponding to 3 or more  $b$ -tag, that is those for which CMVA was higher than a threshold fixed to 0.679. We therefore joined jets corresponding to pairs  $bb$ , or  $dijets$  corresponding to each of the Higgs bosons. As result of these pre-analysis, the dataset has been reduced to 60454 simulated and 433621 background events.

In order to predict signal we considered for each event a number of available variables. All kinematic variables representing the physical phenomenon are included, in particular related to the characteristics of the jets generated by the collision, such as the transverse momentum, pseudorapidity, mass, and all physical quantities connected to them. In Table 1 all the available variables are presented; we considered a total of 46 kinematic variables divided in 4 groups: variables related to the selected jets by the  $b$  tagging algorithm, variables related to the non selected jets, corresponding to other interactions and decays, variables related to the pair production of the 4 jets in 2 pair of dijets, and general variables.

Figure 1 shows the pattern of correlations between the variables. It is evident that some pairs of variables are very correlated to each other, so that some selections will be appropriate in order to implement multivariate prediction models.

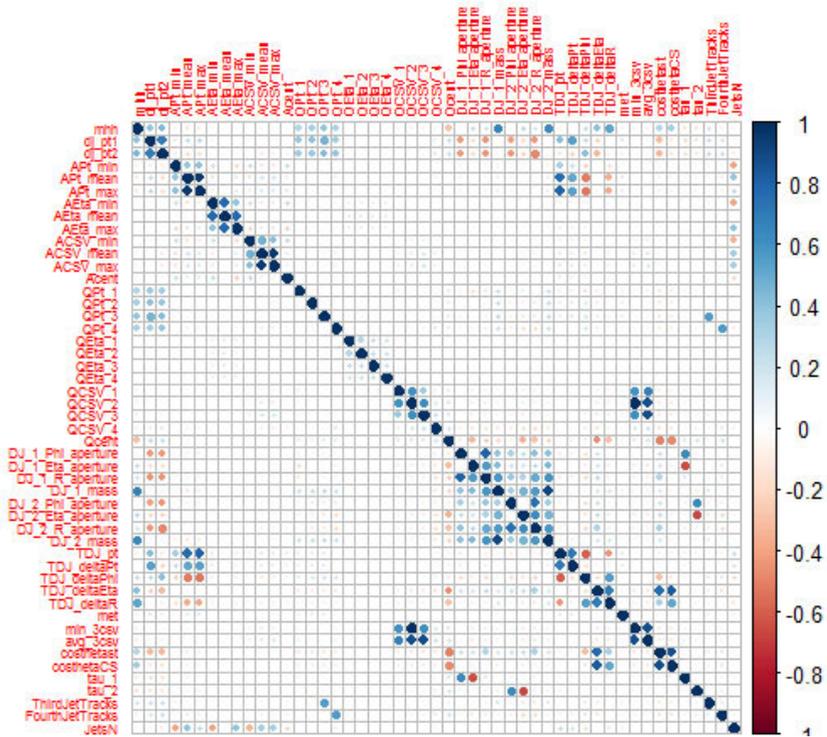
**Table 1.** Variables considered as predictors of signal and background

Name	Variable
Variables related to selected jets	
$QP_1$	Transverse momentum - first jet
$QP_2$	Transverse momentum - second jet
$QP_3$	Transverse momentum - third jet
$QP_4$	Transverse momentum - fourth jet
$QEta1$	Pseudorapidity - first jet
$QEta2$	Pseudorapidity - second jet
$QEta3$	Pseudorapidity - third jet
$QEta4$	Pseudorapidity - fourth jet
$QCMVA1$	CMVA - first jet
$QCMVA2$	CMVA - second jet
$QCMVA3$	CMVA - third jet
$QCMVA4$	CMVA - fourth jet
$QCent$	Centrality of 4 jets
Variables related to non-selected jets	
$AP_{i,min}$	minimum $p_i$ among non-selected jets
$AP_{i,mean}$	$p_i$ mean among non-selected jets
$AP_{i,max}$	maximum $p_i$ among non-selected jets
$AEtamin$	minimum $\eta$ among non-selected jets
$AEtamean$	mean $\eta$ among non-selected jets
$AEtamax$	maximum $\eta$ among non-selected jets
$ACMVAmin$	minimum CMVA among non-selected jets
$ACMVAmean$	mean CMVA among non-selected jets
$ACMVAmax$	maximum CMVA among non-selected jets
$ACent$	centrality non-selected jets
Variables related to pair of jets, corresponding to each Higgs	
$DJ1mass$	mass for the first dijet
$DJ1P_i$	$p_i$ of the first dijet
$DJ1Phi$	$\Delta\Phi$ of the first dijet
$DJ1Eta$	$\Delta\eta$ of the first dijet
$DJ1R$	$\Delta R$ of the first dijet
$\tau_1$	twist of the first dijet
$DJ2mass$	mass of the second dijet
$DJ2P_i$	$p_i$ of the second dijet
$DJ2Phi$	$\Delta\Phi$ of the second dijet
$DJ2Eta$	$\Delta\eta$ of the second dijet
$DJ2R$	$\Delta R$ of the second dijet
$\tau_2$	twist of the second dijet
Other variables	
$TDJP_i$	vectorial sum of $P_i$ of two dijets
$TDJ\Delta\Phi$	$\Delta\Phi$ between the two dijets
$TDJ\Delta\eta$	$\Delta\eta$ between the two dijets
$TDJ\Delta R$	$\Delta R$ between the two dijets
$HHM$	invariant mass of the two dijets
$MET$	Missing transverse energy
$min3cmva$	minimum CMVA of the first 3 jets
$avg3cmva$	mean CMVA of the first 3 jets
$cos\theta^*$	cosine of $\theta$ in the c.d.m. reference system of the two H
$cos\theta_{CS}$	cosine of $\theta$ in the Collins Saper reference system
$JetsN$	number of jets in the event

### 3 Prediction models

In order to introduce statistical models to predict the signal behaviour, we decided to follow two different approaches. In order to favour interpretation and contain error propagation, a small number of 9 variables has been selected (see Table 2), by choosing the ones typically used by physicists in this type of prediction analysis, which are the most relevant from a physical point of view.

A second approach, finalised to identify the best classification and separation between signal and background, consists in using all the available variables. However, in this case, we decided to eliminate from the predictors the variables related to dijets masses and  $b$ -tagging values, since those are



**Figure 1.** Correlation plot of the available variables

**Table 2.** List of the 9 variables selected

$TDJ\Delta R$	$HMM$	$QP_13$
$cos\theta^*$	$QCent$	$DJ1P_1$
$DJ1R$	$DJ2R$	$DJ2P_1$

typically used by physicists in following steps to identify cross sections, remaining with 38 variables used to prediction. Finally a third set predictors adds to these 38 variables a new predictive variable obtained as the sum of the four transverse momenta for the selected jets, resulting in an other set of models with 39 predictors.

Both these approaches have been applied considering the following estimation strategy. In order to balance bias and variance of the predictions, and given the large quantity of available data, we decided to divide the dataset in three subset. A first training set of 50000 balanced events has been used to fit models, a second test set of 16000 units has been used to test the results and select the best model in each class of model considered. A final validation set of other 16000 units has been used to validate the results and identify the best model overall.

### 3.1 Typical statistical learning models

As a benchmark we first fit a number of typical statistical learning models (see for example, Azzalini and Scarpa, 2012) to classify signal or background.

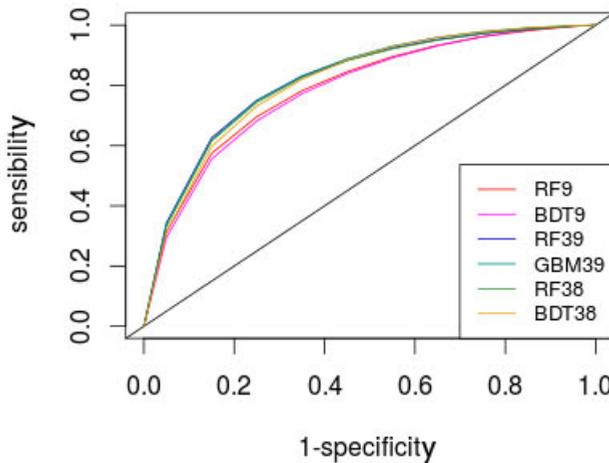
Linear and logistic models, multiple adaptive regression splines (MARS), generalized additive models (GAM), classification trees (CART), Neural networks, projection pursuit regression (PPR) have been implemented and results compared.

Combination of classifiers have also been implemented and used for prediction. Boosting models have been implemented with different specific algorithms. The ones reported are the Boosting decision tree (Drucker, 1997), typically used by physicists and Gradient boosting (Friedman, 2002). Finally, random forest, a quite popular refinement of bagged trees (e.g. Hastie, Tibshirani and Friedman, 2009), have been fitted to data.

These models over performed the simpler ones, so in Table 3 we only show results for ensemble methods. Results from random forest by using all the 39 variables is the best model both considering test and validation set, showing the lower misclassification error and the highest area under the ROC curve (AUC).

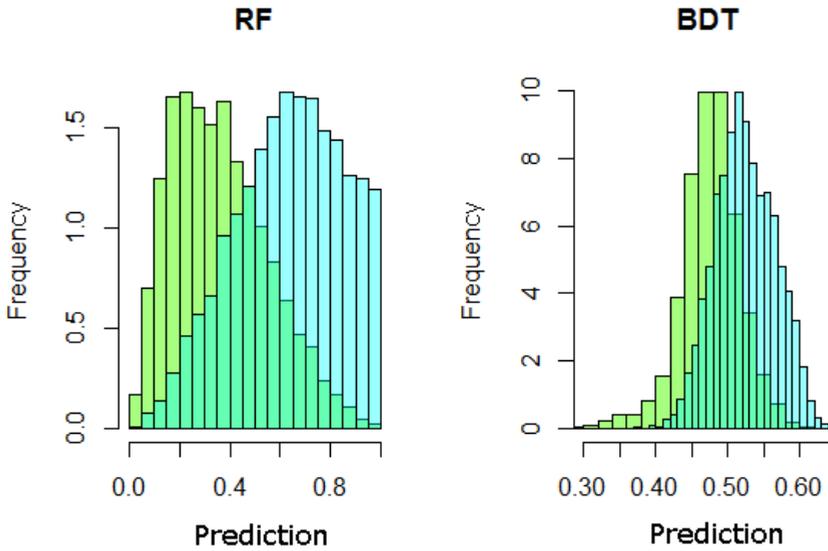
**Table 3.** Ensemble methods - Misclassification error and AUC on the test and validation set.

Model	misclassification error		AUC	
	test set	validation set	test set	validation set
BDT9	0.2821	0.2695	0.7969	0.7934
RF9	0.2756	0.2685	0.7982	0.8004
GBM9	0.2851	0.2858	0.7888	0.7911
BDT38	0.2596	0.2637	0.8204	0.8227
RF38	0.2540	0.2521	0.8224	0.8263
BDT39	0.2340	0.2464	0.8349	0.8400
RF39	0.2320	0.2424	0.8369	0.8424
GBM39	0.2431	0.2477	0.8278	0.8328



**Figure 2.** ROC curve - ensemble methods

Figure 2 shows ROC curves for the best selected models. Figure 3 compare the frequencies of the predictions for the background and the signal units, showing that the difference between frequency



**Figure 3.** Comparison RF and BDT- green: background, blue: signal

**Table 4.** List of interactions between variables selected via random forest

$DJ1P_i$	$DJ1R$	$QCent$	$QP_i3$
$QP_i3$	$TDJ\Delta R$	$QP_i3$	$mhh$
$TDJ\Delta R$	$mhh$	$TDJ\Delta R$	$DJ1R$
$DJ1R$	$DJ2P_i$	$DJ1R$	$DJ2R$
$QCent$	$DJ1R$	$QP_i3$	$DJ2R$
$QCent$	$mhh$	$mhh$	$DJ2R$
$mhh$	$DJ1R$	$QP_i3$	$DJ1R$

curves are more evident for the Random forest results than for the boosted decision tree, where the distributions are closer each other.

It is well known that random forests overemphasizes the role of interactions, so that we decided to use results provided by this model to identify which may be the most interactive kinematic variables to be considered in predicting signal from background in the following considered models. Maximal subtree analysis (Ishwaran et al., 2010, 2011) has been proposed for identify interactions. Smaller minimum depth of a variable with respect to the maximal subtree for another variable indicates that interaction between these two variable are important in predicting response. Thus, we choose to select couples of variables with maximal subtree smaller than a threshold (in our case we selected a threshold of 0.045).

### 3.2 Bayesian non parametric mixed effect model

Models fitted in section 3.1, thanks to the non parametric approach inherited, are very flexible and misclassification errors are quite low, considering the large variability shown in the background. In this section we propose a new model with the goal to keep the flexibility of a non parametric approach but also including some structure related with the specificity of the problem analysed.

Let  $y_i$  be the binary variable encoding signal or background for in event  $i$ . A classical generalized mixed model formulation assumes

$$\begin{aligned} y_i | \pi_i &\sim \text{Bern}(\pi_i) \\ \text{logit}(\pi_i) &= \eta_i \\ \eta_i &= \mu_i + f(\mathbf{x}_i) \end{aligned} \quad (1)$$

where  $\mathbf{x}_i$  is the  $p \times 1$  vector including all the explanatory variables for each event  $i$ ,  $p$  is the number of available explanatory variables,  $\mu_i$  is a random effect for each event, and  $f(\cdot)$  is a generic function to be estimated from data. In the simplest model specification we consider  $f(\mathbf{x}_i) = \mathbf{x}_i^T \beta$  where  $\beta$  is a vector of parameters:

$$\eta_i = \mu_i + \mathbf{x}_i^T \beta$$

In order to provide a flexible formulation and adaptively modelling the determinants of signal we take a Bayesian approach by specifying prior distributions for the parameters  $(\mu_i, \beta)$ . This approach allows for uncertainty in the distribution of the random intercepts, while avoiding over-shrinking and favouring clustering effects. The simplest prior choice for  $\beta$  is

$$\beta \sim \mathcal{N}_p(\mathbf{b}, \mathbf{B})$$

where  $\mathbf{b}_{p \times 1}$  is a prior mean vector and  $\mathbf{B}_{p \times p}$  is a covariance matrix, while for the random intercepts, we suppose a Dirichlet Process

$$\begin{aligned} \mu_i &\sim P \\ P &\sim DP(\alpha P_0), \quad \alpha > 0 \end{aligned}$$

where  $DP(\alpha P_0)$  indicates the Dirichlet Process with base measure  $P_0$  and precision parameter  $\alpha$ .

The Dirichlet process  $DP(\alpha P_0)$  represents a fully flexible prior with support on the set of distributions on the real line, allowing  $P$  to be unknown with  $P_0$  indicating the best guess for such distribution and  $\alpha$  quantifying the confidence in such guess. In our case, we define  $P_0$  as a normal distribution  $\mathcal{N}(0, \sigma^2)$  where  $\sigma^{-2} \sim \text{Gamma}(a, b)$  (i.e. prior for  $\sigma$  is inverse Gamma) and  $\alpha \sim \text{Gamma}(a_\alpha, b_\alpha)$  to favor learning of cluster effects in the data.

In order to estimate the model, we need for an efficient algorithm for posterior computation. We exploit the well known stick-breaking representation of the Dirichlet Process (Sethuraman, 1994). This representation is the following:

let

$$\begin{aligned} V_h &\stackrel{iid}{\sim} \text{Beta}(1, \alpha) & \theta_h &\stackrel{iid}{\sim} G_0 \\ \pi_h &\sim V_h \prod_{l < h} (1 - V_l) & G &= \sum_{h=1}^{\infty} \pi_h \delta_{\theta_h} \end{aligned}$$

therefore  $G \sim DP(\alpha, G_0)$ , where  $\delta_\theta$  indicates a mass point concentrated in  $\theta$ .

From this representation of the DP, it is clear that a realization of the this process is discrete in nature, thus it favours ties among random intercepts: events in the same cluster have equal random intercept values. Denoting with  $S_i$  the grouping variable, the stick-breaking representation shows clustering effects among events, providing  $\mu_i = \theta_{S_i}$ , with the number of clusters stochastically increasing with  $\alpha$ . This clustering property is particularly useful in our signal detection, favouring events with common kinematic features to share the same effect. Conditionally on the grouping indicator  $S_i$ , the Gaussian base measure  $P_0$  is conjugate, favoring the implementation of an efficient Gibbs sampling algorithm.

Therefore, for posterior computation we exploit blocked Gibbs sampler algorithm by Ishwaran and James (2001) a recently proposed data-augmentation scheme based on Pólya-Gamma (PG) distribution (Polson et al. 2013).

We fitted to the data the simple Dirichlet model by obtaining a misclassification error on the test set of 0.49 much higher than the 0.32 we obtained with the random forest. Also the AUC of this model is much lower, being 0.511 compared with 0.74 of the previous results. Clearly the model proposed shows some drawbacks, mainly related with the choice to use a parametric function to specify  $f(\mathbf{x}_i)$ , the relation between predictors and response variable. This choice does not admit nonlinearities in the relationship and interactions among variables are not allowed in the function.

We therefore propose a model which keep parametric the random intercept, but allows for nonlinearities in the relationship, by estimating the function  $f(\cdot)$  through an additive model by using  $P$ -splines (Lang and Brezger, 2004). Model (1) is now specified as follow

$$\begin{aligned}
 y_i | \pi_i &\sim \text{Bern}(\pi_i) \\
 \text{logit}(\pi_i) &\sim \eta_i \\
 \eta_i &\sim f(\mathbf{x}_i) \\
 f(\mathbf{x}_i) &= f_1(x_{i1}) + \dots + f_p(x_{ip}) \\
 f_j(x_j) &= \sum_{r=1}^{M_j} \beta_{jr} B_{jr}(x_j), \quad j = 1, \dots, p
 \end{aligned} \tag{2}$$

where  $B_{jr}$ ,  $r = 1, \dots, M_j$  are base function for  $P$ -splines and  $\beta_j = (\beta_{j1}, \dots, \beta_{jM_j})$  is a parameter vector and  $M_j$  is the number of basis needed for the  $j$ th variable.

In addition to the original variables, we also created new variables defined by the interactions identified as most influent by the Random forest and included also these new variables in the additive form in model (2) Include interactions identified with Random Forests, as new variables.

Misclassification error on the test set for this model was 0.34 and AUC 0.71, comparable with the random forest results.

A third non parametric model we fitted is a Bayesian version of the random forest. The formal specification of BART (Bayesian additive regression tree) proposed by Chipman, George and McCulloch (2010) modify model (1) as follows

$$f(\mathbf{x}_i) = \sum_{j=1}^m g(\mathbf{x}_i; T_j, M_j)$$

where  $g(\mathbf{x}_i, T_j, M_j)$  denotes the predicting function assigning a value to  $\mathbf{x}_i$  given the Bayesian tree  $T_j$  and parameters  $M_j$ , so that a BART is a sum of trees. Variables to be selected at each node, cut points and depth of the trees are parameters and need for a prior distribution.

An efficient algorithm is available to fit a BART, which is a tailored version of Bayesian backfitting MCMC (Hastie and Tibshirani, 2000) that iteratively constructs and fits successive residuals

A tree automatically includes interactions so we don't need to include among variables the interactions identified in section 3.1.

Misclassification error on the test set, for BART is 0.35 and AUC is 0.71, almost equal to the  $P$ -splines ones and comparable with random forest.

### 3.3 Combination of Bayesian non parametric models

A final set of models has been considered by combining the Bayesian non parametric model described in section 3.2.

**Table 5.** Final comparisons

Model	Misclassification Error	False Positives	False Negatives	AUC
RF	0.3200	0.2678	0.3678	0.7393
GB	0.3300	0.2971	0.3602	0.7234
BDT	0.3351	0.3461	0.3389	0.7198
DP	0.4920	0.4121	0.5663	0.5118
P-splines	0.3400	0.3975	0.2874	0.7102
BART	0.3480	0.3083	0.3846	0.7078
BART as atoms of DP	0.4620	0.4346	0.4867	0.5429
DP+P-splines	0.3300	0.3347	0.3295	0.7259
BART as atoms of DP + P-splines	0.3240	0.2970	0.3455	0.7321
BART+P-splines	0.3140	0.3138	0.3141	0.7417

First we propose the mixed effect model 1 including both a DP with Gaussian atoms for the random intercept  $\mu_i$  and Bayesian  $P$ -splines for the function  $f(\mathbf{x}_i)$ .

A second combination model include the relation between predictors and classes in be DP itself by allowing its atoms to depend on covariates through a BART model.

Finally, a last model has been fitted by considering a sum of a BART model and a Bayesian  $P$ -splines additive model:

$$f(\mathbf{x}_i) = f_1(x_{i1}) + \dots + f_p(x_{ip}) + \sum_{j=1}^m g(x; T_j, M_j).$$

## 4 Results

Table 5 summarizes all the results obtained by the different models. We only selected the best random forest, gradient boosting and boosted decision tree, to be compared with the other proposed models.

Combination of models, as expected, produce the best results both in terms of misclassification error and of AUC. The final model joining BART and  $P$ -splines seems to be the one showing the lower error and the higher AUC, even if other combination model, both Bayesian and classical, show similar results. In table 5 we also include the proportions of false negative and false positive which show the same behaviour than the other indices.

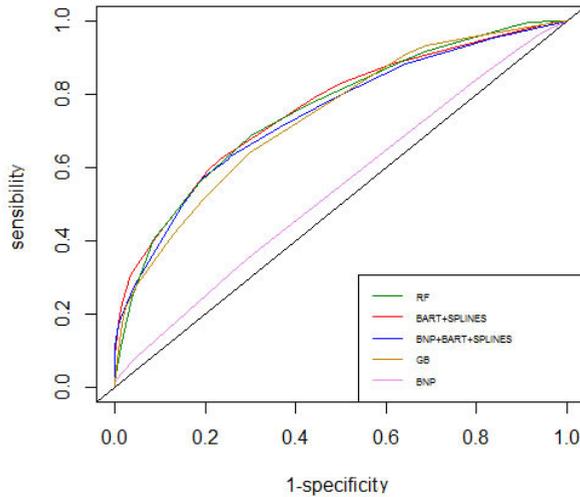
Figure 4 shows the ROC curves for the BNP models compared each other and with the random forest model. For most of the curve the model joining BART and  $P$ -splines outperform all other models. Only in a small section of the plot, for low values of specificity random forest is preferable, and for another small portion (values around 0.7) gradient boosting has the performance.

## 5 Discussion

The choice of the most appropriate model for predicting signal of Higgs pair production can be driven by knowledge of its strength. However, it is always better to fit different models and compare them on different data, in order to choose the best classifier.

Combination of models may work better than single models if each model has a different strength compared to the others: committees can help when different learners have complementary strengths for a given task.

Choice of 9 most predictive variables has been done by physicists by looking to marginal correlation and meaning of the variables.



**Figure 4.** ROC curves

Most of models with 39 variables have some procedure of automatic selection of complexity and of variables, compromising bias and variance. We used a test set to train and select complexity level of each model.

## Acknowledgments

This research was partially supported by the AMVA4NewPhysics Marie Skłodowska-Curie ITN Network.

## References

- Alwall J., Herquet M, Maltoni F., Mattelaer O., Stelzer T., (2011), Madgraph 5: going beyond, *Journal of High Energy Physics*, 128, 033009.
- Azzalini A., Scarpa B. (2012), *Data Analysis and Data Mining*, Oxford University Press.
- Chipman H.A., George E.I., McCulloch R.E., (2010), BART: Bayesian Additive Regressive Trees. *The Annals of Applied Statistics*, 4(1), 266D298.
- Das S., Konigsberg J., Vernieri C., Rizzi A., (2013). Search for di-Higgs resonances decaying to 4  $b$ -jets in pp collisions at 8 TeV. *CMS Analysis Note* 2013/227.
- de Florian D., Mazzitelli J. (2013), Higgs Boson Pair Production at Next-to-Next-to-Leading Order in QCD. *Phys. Rev. Lett.* 111 111 , pp. 161D168. DOI: 10.1103/PhysRevLett. 111.201801.
- Friedman J. H., (2002), Stochastic gradient boosting, *Computational Statistics & Data Analysis* 38, 367D378.
- Gao J., Guzzi M., Huston J., Lai H.-L., Li Z., Nadolsky P., Pumplin J., Stump D., Yuan C.-P., (2014), Ct10 next-to-next-to-leading order global analysis of qcd, *Physical Review*, 89.

Hastie, T., Tibshirani, R., Friedman J., (2000), Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors), *The Annals of Statistics* 28, 337–407.

Hastie T., Tibshirani R., Friedman J., (2009) *The Elements of Statistical Learning*, Springer Series in Statistics.

Ishwaran H., James L. F. (2001), Gibbs Sampling Methods for Stick-Breaking Priors. *Journal of the American Statistical Association*, 96(453), 161–173.

Ishwaran J., Kogalur U.B., GORODESKI E.Z., MINN A.J., LAUER M.S., (2010), High-dimensional variable selection for survival data, *Journal of the American Statistical Association*, 105, 205–217.

Ishwaran J., Kogalur U.B., Chen X., Minn A.J., (2011), Random survival forests for high-dimensional data, *Statistical Analysis and Data Mining*, 4, 115-132.

Lang S., Brezger A., (2004) Bayesian P-splines *Journal of computational and graphical statistics* 13 (1), 183-212.

LHC Higgs Cross Section Working Group, (2012), *Handbook of LHC Higgs Cross Sections: 2. Differential Distributions*. Tech. rep. arXiv:1201.3084. CERN-2012-002. Geneva, 2012. URL:<http://cds.cern.ch/record/1416519>

Polson N.G., Scott J.G., Windle J., (2013) Bayesian inference for logistic models using PólyaGamma latent variables. *Journal of the American Statistical Association*, 108(504):1339–1349,

Weiser C., (2006), A Combined Secondary Vertex Based B-Tagging Algorithm in CMS. *CMS Note* 2006/014.