Analysis of Decay Processes Separation

Jiřina, Marcel
2008

Institute of Computer Science
Academy of Sciences of the Czech Republic

Analysis of Decay Processes Separation

Marcel Jiřina and František Hakl

Technical Report No. V-1035

November 2008

Abstract
This report summarizes results obtained on simulated data from ATLAS detector for Higgs boson search.

Keywords:
multivariate data, Higgs boson, data classification, neural network, genetic optimization.

Analysis of Decay Processes Separation

Marcel Jirina and František Hakl

Institute of Computer Science AS CR, v.v.i., Pod Vodárenskou věži 2, 182 07 Prague 8 – Libeň, Czech Republic, marcel@cs.cas.cz, hakl@cs.cas.cz

Contents

I. Introduction

This report summarizes results in the following areas of research:

a. Analysis of possibility to use genetic neural networks with categorial variables and systems with estimation of correlation dimension for improvement of reconstruction algorithms of decay scheme.

b. Analysis of separation of decay events with Higgs boson and gluon based on data from computer simulations.

c. Analysis (enhancement) of statistical differences of events with Higgs boson and background events by the use of neural networks and correlation dimension.

In this report we neither describe mathematical background of methods applied nor particular software tools. The report presents results in form of tables, ROC curves and separation curves, only.
We used simulated data from FZU as well as older data sets dealing with the same problem to get more general results.
We found that genetically optimized neural networks as well as unconventional approaches based on the theory of effective dimension, especially correlation dimension of data, can significantly enhance statistical differences between events with Higgs boson and background events.

II. DATA SET I. ("JURANEK DATA- POMERON")

These data correspond to Higgs boson production process as a signal and very similar process with gluon as a background.

The process is an exclusive Higgs production, the simplest case where no other protons arise. In the data set there are for quadruples of variables, namely E, p_x, p_y, p_z, i.e. energy and three coordinates of momentum. The quadruples correspond to proton 1, proton 2, bjet1, and bjet2 thus forming 16 numbers for each event.

A.Elementary analysis

We studied first correlation of individual variables with the class mark and found very little correlation as shown in Table AA.

Variable	1	lrn: 0.0180	tst: -0.0497
	2	-0.0286	-0.0499
	3	0.0221	0.0480
	4	0.0180	-0.0497
	5	-0.0188	-0.0423
	6	-0.0141	-0.0792
	7	-0.0232	0.0403
	8	0.0188	0.0423
	9	-0.0653	0.0593
	10	0.0027	0.0100
	11	0.0125	0.0262
	12	-0.0125	0.0042
	13	0.0364	0.0196
	14	0.0038	0.0038
	15	-0.0210	-0.0342
	16	-0.0204	-0.0208

Table AA. Correlation coefficients between individual variables and class mark.

It is seen that the largest value of the correlation coefficient is 0.0593 that means no correlation at all if any noise is assumed. This points to the fact that these data cannot be separated with good efficiency. Nevertheless we show that some minor possibility exists.

B. Signal and background separation

The data set at hand has total 16 variables and we have used all these variables. We have found that elimination even of a variable, which is clearly, in fact, a duplicate of some another variable leads to worse results.

For separation we used correlation dimension based classifier. For technical details see e.g. [7], [8], [9]. Programs are written in c++ and can be run under Unix/Linux as well as under Windows environment.

C. Results obtained

In Fig. 1 the ROC curve is shown. As the desired course is to approach to the left upper corner as much as possible, this nearly diagonal curve signalizes bad separation of the signal from the background. Even though this bad impression one can find that for threshold on classifiers output set to 0.310333 there is a

minimal classification error 0.098022, i.e. less than ten per cent. Unfortunately it holds for signal acceptance (efficiency, sensitivity) 0.99933 and background acceptance (background error, purity) 0.97493 that are in practice unacceptable values. At the same time, the ROC curve crosses the curve for the quality

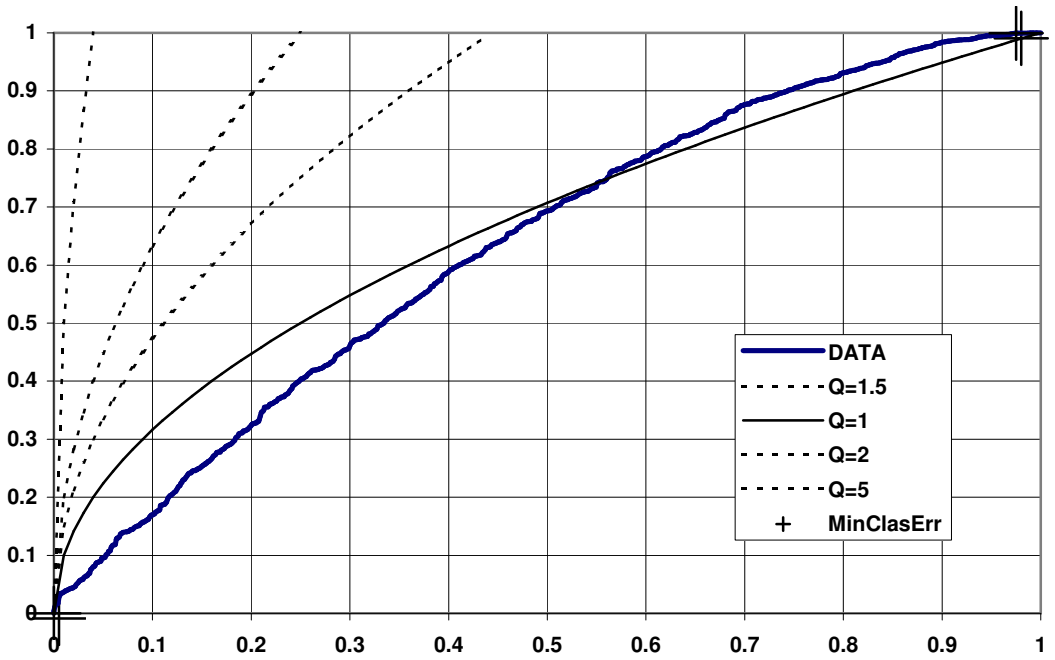

Fig. 1. The ROC curve obtained. In the figure the double cross at the upper right part denotes state of minimal classification error.

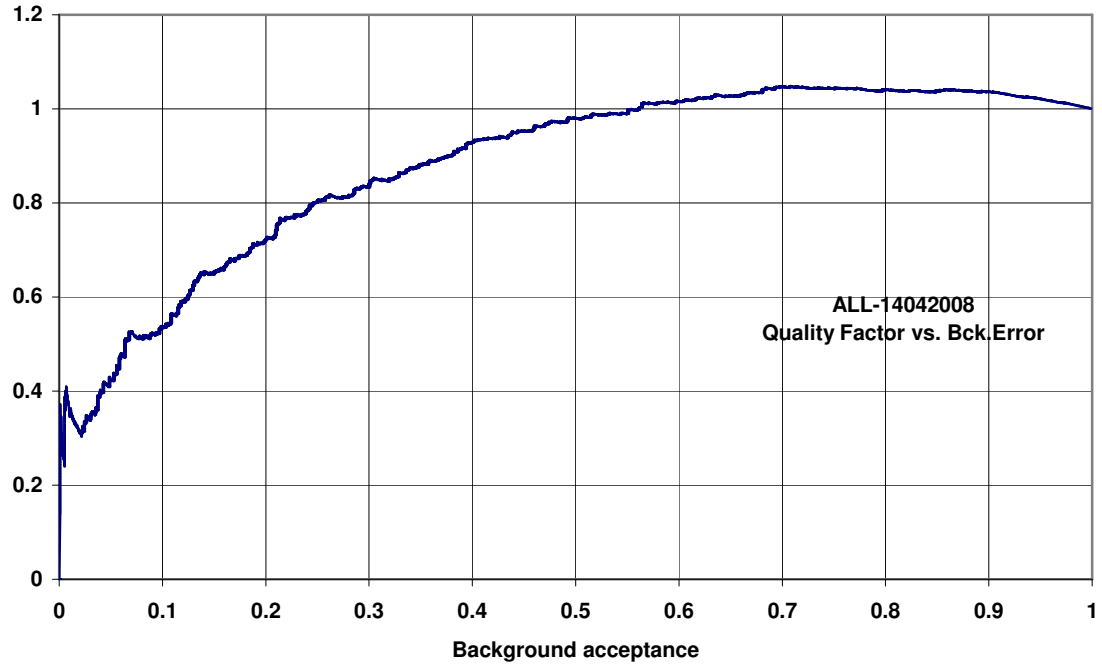

Fig. 2. The quality factor [Q = S/sqrt(B)] as the function of background acceptance. For values of Q larger than 1 there is no loss of statistical significance.

factor (see the thin smooth line in Fig.1 or also horizontal line at level 1 in Fig. 2), i.e. ratio S/sqrt(B) equal to one at point nearly in the middle of the picture. This means that we can suppress the background to 56.09 % and at the same time keep 74.89 % of the signal without loss of statistical significance.

In Fig. 2 the dependence of the quality factor Q = S/sqrt(B) on the background acceptance is shown. There is the same horizontal axis as in Fig. 1 and thus these figures correspond to each other. The value of Q crosses value 1 at the point for background acceptance (background error, purity) 0.5609, which corresponds to the threshold on the output variable of the classifier 0.4644. Thus setting the threshold for the output variable of this classifier to 0.4644 we get acceptances 56.09 % and 74.89 % mentioned above.

III. DATA SET II ("ERW DATA")

Identification of hadronic τ decays will be the key to the possible Higgs boson discovery in the wide range of the MSSM parameter space [1]. The $H / H / A \to \tau\tau$ and $H^{\pm} \to \tau\nu$ are promising channels in the mass range spanning from roughly 100 GeV to 800 GeV. The sensitivity increases with large $\tan\beta$ and decreases with rising mass of the Higgs boson. The $H \to \tau\tau$ decays will give access to the Standard Model and light Minimal Supersymmetric Standard Model Higgs boson observability around $m_{\mathrm{H}} = 120$ GeV, with Higgs boson produced by vector-boson fusion [2]. Data was kindly provided by Dr. Elsbieta Richter-Was [1].

A.Identification of variables

For the classification procedure calorimetric observables as described in details in [4] are used. Separately we optimize identification procedure for single-prong (τ_{1p}) and three-prong (τ_{3p}) candidates. The τ_{1p} is seeded by the leading hadronic track at vertex (track η and φ at the vertex). The τ_{3p} is seeded by the bary-center of three nearby tracks. The calorimetric observables are calculated from energy deposition in cells within a distance from a seed of $\Delta R = 0.2$.
The following calorimetric and tracking variables are used to build discriminating observables:

- Track transverse momenta of a leading track p_T^{track} (or scalar sum of tracks transverse momenta in case of τ_{3p} candidates)
- Electromagnetic radius of the τ-candidate, R_{em}
- Number of strips N_{strips}^{τ}, strips with energy deposition above a certain threshold
- The width of energy deposition in strips, W_{strips}^{τ}
- The fraction of the transverse energy deposited, $fracET_{R12}$, in the $0.1 < \Delta R < 0.2$ radius with respect to the total energy in the cone $\Delta R = 0.2$. Cells belonging to all layers of the calorimeter are used.
- The ratio of energy deposited in the hadronic calorimeter $E_T^{chrgHAD}$ and track transverse momenta, $\frac{E_T^{chrgHAD}}{p_T^{track}}$ (or sum of transverse momenta in case of τ_{3p} candidates)
- The ratio of energy deposited in calorimeters in a a ring $0.2 < \Delta R < 0.4$, with respect to the total energy deposited in a cone $\Delta R < 0.4$, $E_T^{chrgHAD} / E_T^{calo}$ and $E_T^{chrgHAD} / E_T^{calo}$.

The variables above are used either directly or to build up in total 6 discriminating variables: N^{τ}_{strips}, W^{τ}_{strips}, $fracET_{R12}$, R_{em}, $\frac{E_T^{chrgHAD}}{p_T^{track}}$, $\frac{E_T^{otherEM}+E_T^{otherHAD}}{E_T^{calo}}$. Classifiers use them separately, without any assumptions on the possible correlations.

B.Standard cuts method

The cuts-based approach uses a sequence of properly tuned cuts for individual variables. The cuts used here for reference selection are as follows:

- $N^{\tau}_{strips} < 15$;
- $W^{\tau}_{strips} < 0.004$;
- $fracET_{R12} < 0.4$ for τ_{1P} (< 0.6 for τ_{3P});
- $R_{em} < 0.08$;
- $\frac{E_T^{chrgHAD}}{p_T^{track}} < 1.0$;
- $\frac{E_T^{otherEM}+E_T^{otherHAD}}{E_T^{calo}} < 0.15$ for τ_{1P} (< 0.25 for τ_{3P}).

It is obvious that the order of cuts in the sequence has no impact on the final acceptance.

C.Results obtained

Standard cuts method was compared with different classifiers. We have used

- Off-the-shelf product Statistica 7 by Statsoft Co. mostly used for analysis of general statistical data often from area of economics. Due to relative complexity most users use default settings and it appears that it is usually the best setting. We used default settings as well; the only exception is switching between classifier regime and approximator regime.
- NNSU with genetic optimization [3], [4]
- GMDH neural network
- Correlation dimension–based method; in fact the variant which uses relation of correlation dimension to harmonic numbers [9].

Our results for ERW data are presented in Fig. 3.

For Statistica 7 package we obtained two sets of curves, one set for classifier regime, the other for approximator regime. Each set consists of four curves corresponding to four different neural networks generated by Statistica package for the task given. Only four best results are shown from total ten networks generated in each regime.

NNSU and GMDH were used as approximators and gave nearly the same ROC curve shown in bold black in Fig. 3. It is seen that generally give nearly the same result as Statistica package being slightly better in some part of the ROC curve.

The correlation dimension based method in this case surprisingly overcomes other methods used here. We must warn before too much optimism, as this need not be a rule.

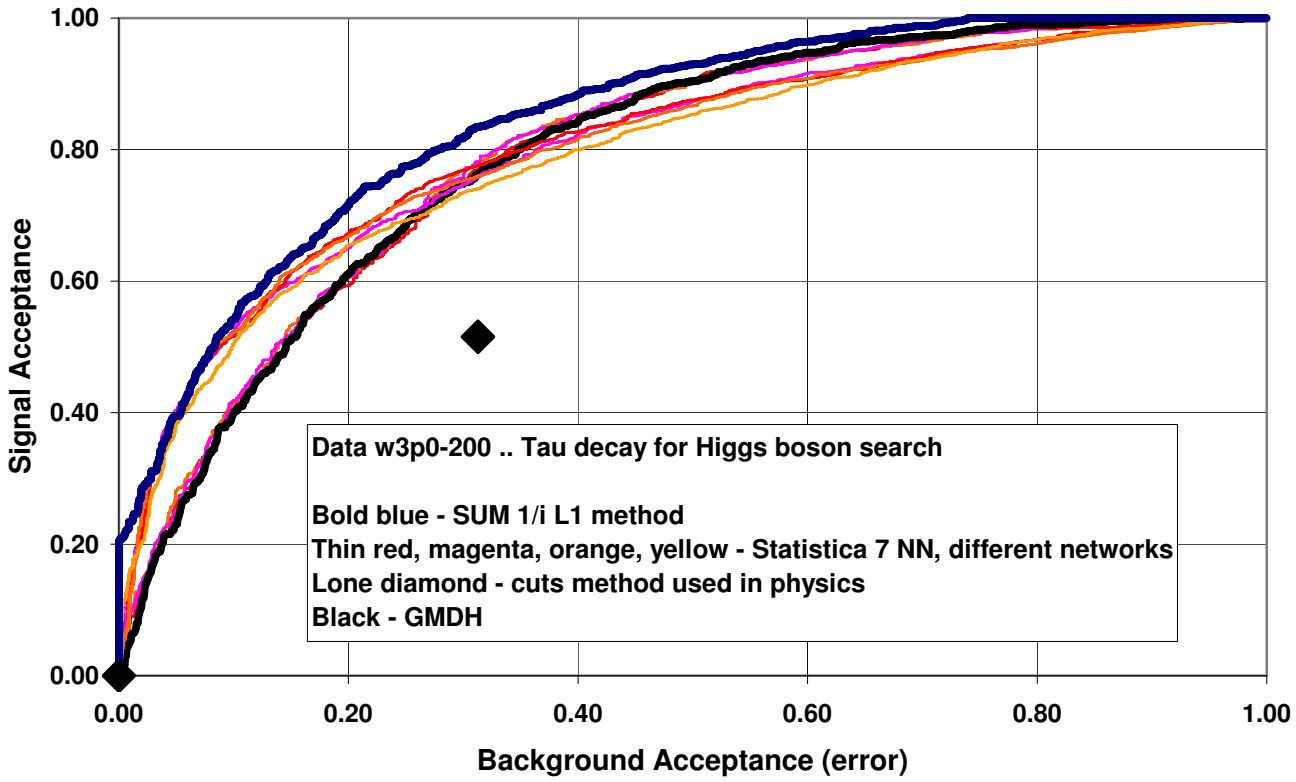

Fig. 3. Roc curves for different signal and background separators.

IV. Data set III (ERW data B)

This part describes an application of a neural network approach to SM (standard model) and MSSM (minimal supersymmetry standard model) Higgs search in the associated production $t\bar{t}H$ with $H \to b\bar{b}$. This decay channel is considered as a discovery channel for Higgs scenarios for Higgs boson masses in the range 80 - 130 GeV. Neural network model with a special type of data flow is used to separate $t\bar{t}jj$ background from $H \to b\bar{b}$ events.

A. Data description

Data is produced during proton-proton collision with energy 14 TeV (in centroid mass system). There is a certain probability that Higgs bosons are produced in this collision (see Fig. 4, a)). In the case that mass of this Higgs boson is $m \leq 200\ GeV/c^2$ decay $H \to b\bar{b}$ is dominating. The main background of process above is process without production of Higgs boson but with the same final state (see Fig. 4, b)). Instead of Higgs boson in this case a gluon is radiated and produce a $b\bar{b}$ pair.

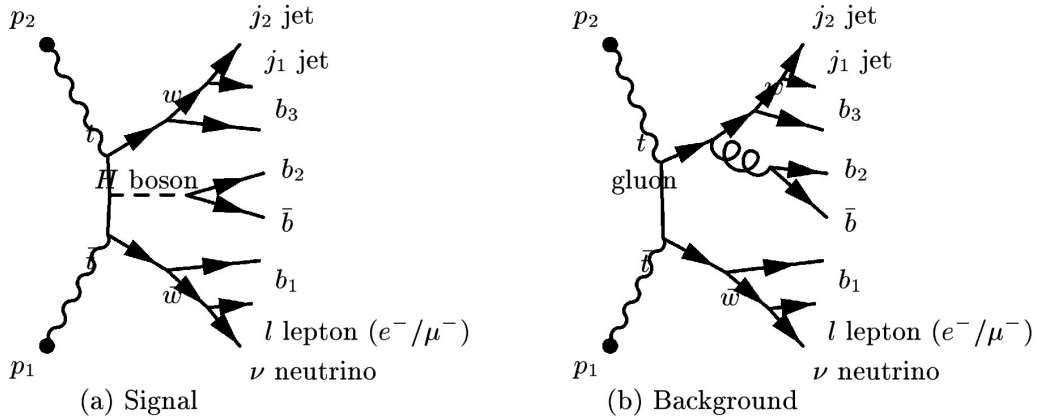

Fig. 4 Higgs boson and gluon decay trees.

Thus products of this type of collision are two jets from one *w* decay, four *b*-jets and one lepton (electron or muon) plus missing energy from unobserved neutrino. Each visible particle (2 jets, 4 *b*-jets and lepton) is described by three values P_T - transversal momentum [GeV/c^2], η - pseudorapidity and φ - direction angles [rad], corresponding to particle vectors with negligible mass. Neutrino is described by two missing energy values $E_{\text{miss x}}$, $E_{\text{miss y}}$.

Fundamental variable, which can be used in Higgs boson search, is an effective mass M_{bb} of two *b*-'s, which can arise either form Higgs boson decay or from gluon decay after $p\overline{p}$ collision. There is lot of events with gluon decay (background) and much less events with Higgs decay (signal). Each of these two classes of events have different statistics of effective mass M_{bb}. Statistics corresponding to Higgs boson decay is theoretically of Gaussian distribution with mean 120 GeV/c^2 and $\sqrt{\sigma}$ = 15 GeV/c^2, whereas statistics corresponding to gluon decay is much broader. Difference between those two statistics can be exploited to decide if Higgs boson decay is present in the data or not.

In addition, other physical reasons reject all events, in which at least one of following conditions has been satisfied [5], [6]:

- At least one jets has $P_{\text{T}} < 15\text{GeV}$.
- At least one jet has pseudorapidity out of the range (–2.5, 2.5).
- Lepton is electron and $P_T^{lep} < 20$ GeV.
- Lepton is muon and $P_T^{lep} < 6$ GeV.

All events used passed these restrictions.

Data was kindly provided by Dr. Elsbieta Richter-Was.

B. Analysis

Data really measured do not provide information about presence of Higgs decay in the event, hence for real data we have available total distribution of M_{bb} (see Fig. 5, upper curve) only. For data simulated, we can plot two histograms of M_{bb}, one for background only and the second one for plain signal (see Fig. 5, two bottom curves).

Application of neural networks covers the case when we know distribution of separated signal and separated background (e.g. bottom curves in the Fig. 5) because neural networks should provide information if a given event is signal or background (up to some misclassification, of course). So the main idea how to exploit neural network to confirm Higgs decay presence is based on filtering of events in such a

way that percentage of signal will be increased after filtering and at the same time significance $\frac{S_a}{\sqrt{S_a + B_a}}$ will stay on the same level.

C. Results obtained

Some experiments were performed with signal and background data described above. We use raw data as they were produced by package PYTHIA. It is evident that neural nets with switching units are able to partly separate signal with Higgs decay and background without Higgs decay. As we can seen on the plot of final outputs from neural networks, there is possible to choose an interval in which signal prevails over background, in case of discussed experiment such interval should be (0.45; 0.75). We call such interval "best signal window". On the other hand we can take interval, in which the signal is suppressed and background will be of dominant importance. In the case discussed now, as this interval should serve (0.20; 0.45). We call such interval consistently "best background window".

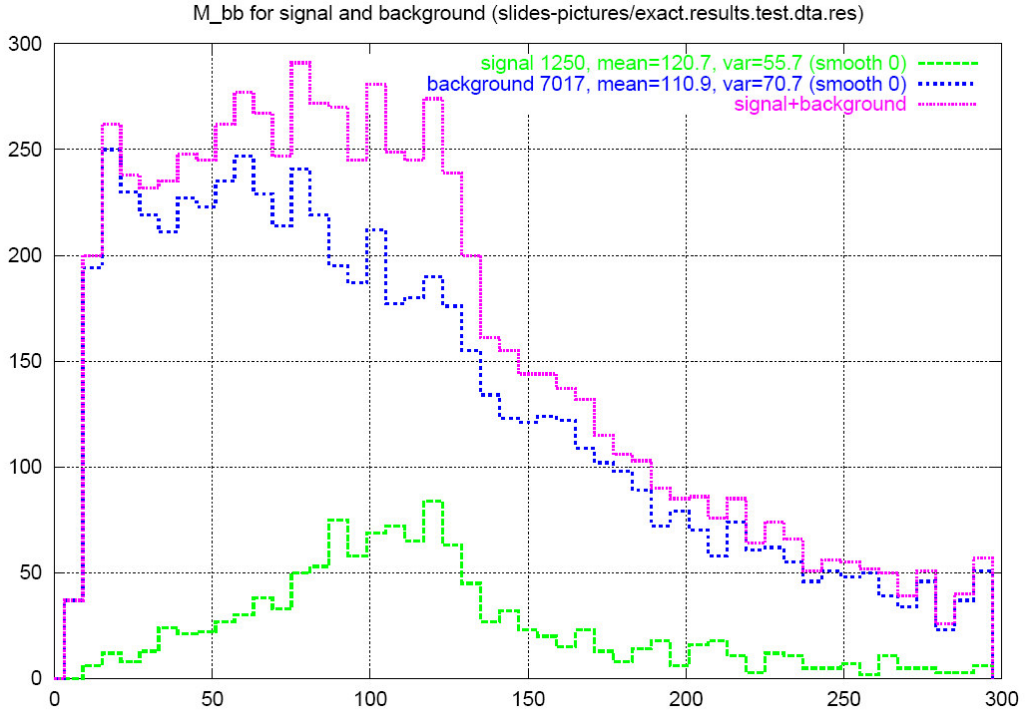

Fig. 5: Histogram of Mbb for signal and background (first number in the legend is number of events accepted by rejecting algorithm, means is average value of Mbb, var is mean square error and smooth is a smoothing factor used to plot the histogram).

If Higgs decay is present than we can assume that plot of M_{bb} over all events that are mapped by neural network into the best signal window will differ from the next one, based on events mapped into the best background window. Really, for our simulated data these plots differ, see figures 6a) and 6b). We can see visual differences between these two plots. Of course these plots should be different from the similar plot in Fig. 5. Hence our experiments convinced us that chosen approach to separation of Higgs decay seem to be applicable and promise useful detection methods.

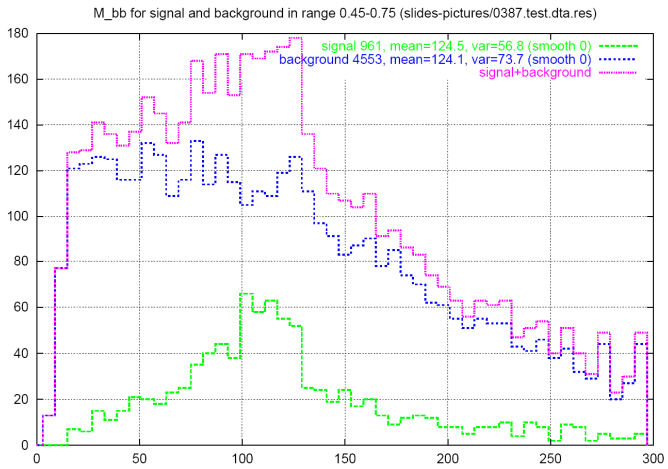

Fig. 6a. Signal and background distributions in the signal window.

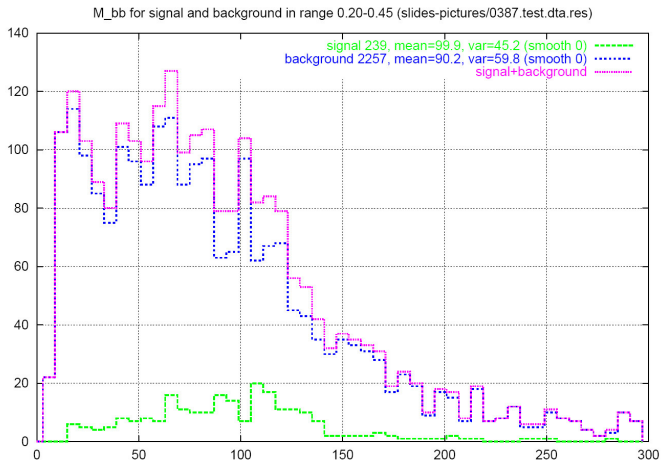

Fig. 6b. Signal and background distributions in the background window.

V. Conclusions

Our results show that NN and CD approaches are applicable to the problem of Higgs boson detection. These filters can be used to emphasize difference of M_{bb} distribution for events accepted by filter (with better *signal/background* rate) and M_{bb} distribution for original events (with original *signal/background* rate) under condition that there is no loss of significance. This improvement of the shape of M_{bb} distribution can be used as a criterion of existence of Higgs boson decay in considered discovery channel. Finally we point out that developed separation methods based on neural networks with switching units or on CD based classifier are universal separation methods, which can be used for various particle recognition problems.

Acknowledgement

This work was supported in part by the Institute of Physics of the Academy of Sciences of the Czech Republic under contract to ISC AS CR and in part by the Ministry of Education of the Czech Republic under the project Center of Applied Cybernetics No. 1M0567.

References

[1] E. Richter-Was, H. Przyseniak, F. Tarrande: Exploring hadronic tau identification with DC1 data samples: a track based approach. ATLAS Physics Communication ATL-COM-PHYS-2004-057, 22 Sept. 2004, rev. 24.10.2004.

[2] F. Hakl, M. Jirina, E. Richter-Was: Hadronic tau's identification using artificial neural network. ATLAS Physics Communication ATL-COM-PHYS-2005-044, rev. 26 Aug. 2005.

[3] 1-F. Hakl and M. Jirina.: Design of a neural net for level-two triggering in atlas nuclear experiments. Neural Networks World, vol. 6 (6): pp. 951-973, 1996.

[4] 2-P. Bitzan, J. _Smejkalova, and M. Kucera: Neural networks with switching units. Neural Network World, vol. 4, pp. 515-526, 1995.

[5] 3-M. Sapinski: The ttbb background to the Higgs searches: Pythia versus CompHEP rates. Tech. rep. no. ATL-PHYS-2000-020, CERN Geneve, 2000.

[6] 4-T. Sjöstrand: PYTHIA 5.7 and JETSET 7.4 Physics and Manual. tech. report CERN - TH . 7112/93, Dec. 1993.

[7] M. Jirina, M. Jirina, jr.: Simple and Effective Probability Density Estimation and Classification. CD-DVD-ROM proceedings of SICE-ICASE International Joint Conference 2006, Oct.18-21, 2006 in Bexco, Busan, Korea, paper No. SA01-3.pdf, 2pp.

[8] M. Jirina, M. Jirina, jr.: Probability Density Estimation by Decomposition of Correlation Integral. Proc of the 2008 International Conference on Artificial Intelligence and Pattern Recognition (AIPR-08) Orlando, Florida, USA, July 7-10, 2008, B. Prasad, P.Sinha (Eds.), pp. 113-119 (paper No. AIPR165).
[9] M. Jirina, M. Jirina, jr.: Device for stating of value of control variable (in Czech). Patent pending at Industrial Property Office in Prague, Czech Rep. under No. PV 2008-245; Z 7576 .