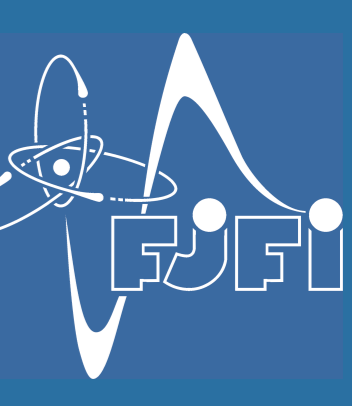




Divergence Techniques in High Energy Physics

Petr Bouř, Václav Kuř, Jiří Franc

Department of Mathematics, Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University in Prague
petr.bour@jfifi.cvut.cz



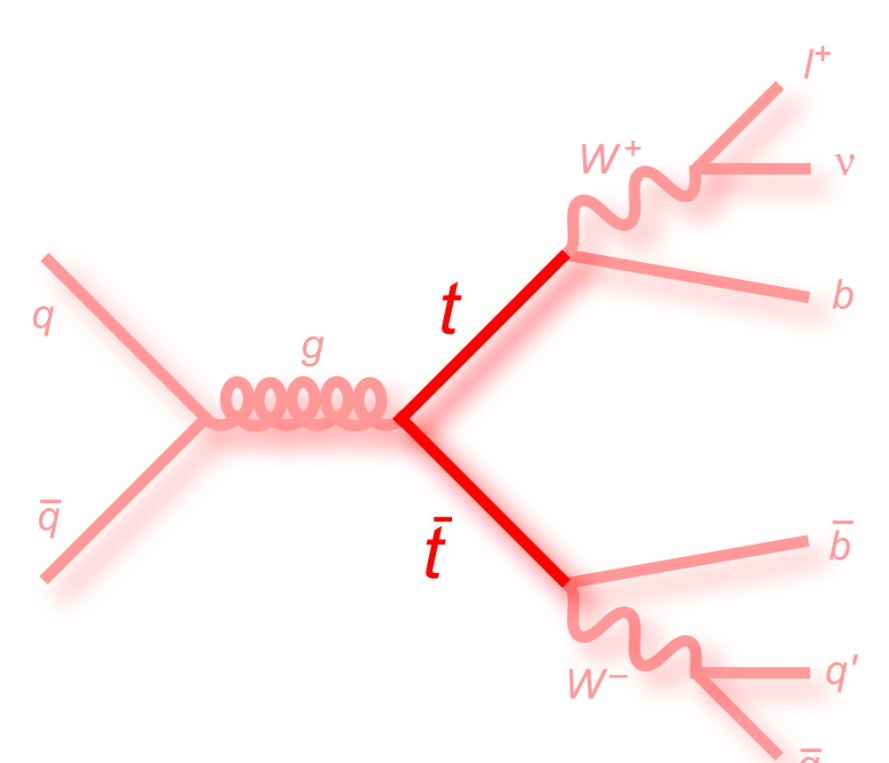
Summary

Binary decision trees are a widely used tool for supervised classification of high-dimensional data, for example among particle physicists. We present our proposal of the supervised binary divergence decision tree with nested separation method based on kernel density estimation. A key insight we provide is the clustering driven only by a few selected physical variables. The proper selection consists of the variables achieving the maximal divergence measure between two different subclasses of data. Further we apply our method to Monte Carlo data set from the particle accelerator Tevatron at the $D\bar{O}$ experiment in Fermilab. We also introduce the modification of statistical tests applicable to weighted data sets in order to test homogeneity of the Monte Carlo simulation and real data.

I. INTRODUCTION

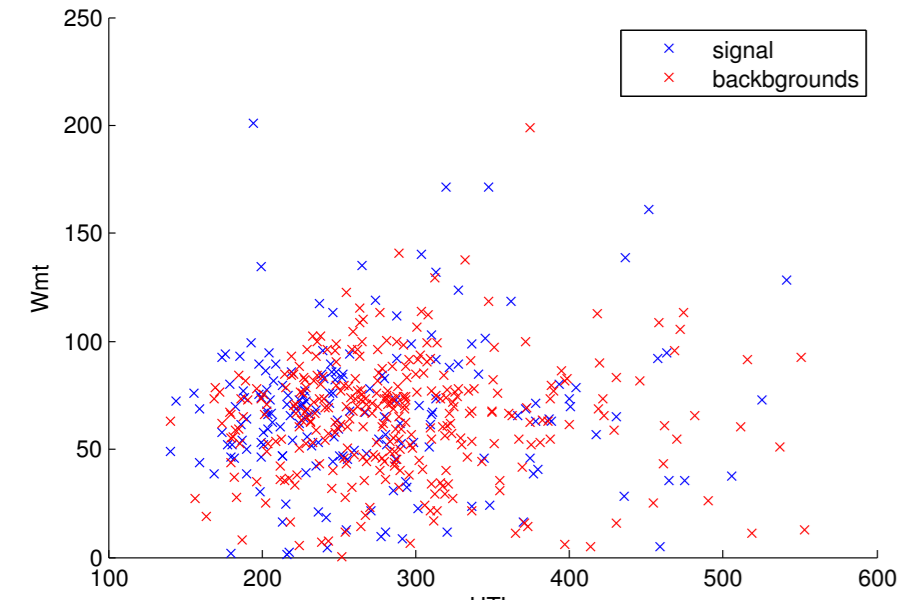
We are interested in top-antitop quark pair production at the Tevatron synchrotron within $D\bar{O}$ Experiment in Fermilab. Our goal is to develop and test our new statistical techniques. We focus mainly on the homogeneity testing of Monte Carlo simulation (MC) vs measured data (DATA) and signal discrimination from massively contaminated data sets.

$D\bar{O}$ Experiment and top pair production



- ▶ lepton+jets channel
- ▶ center-of-mass energy of 1.96 TeV
- ▶ 6 analysis channels:
 - ▶ Electron: 2 Jets, 3 Jets, 4+ Jets
 - ▶ Muon: 2 Jets, 3 Jets, 4+ Jets
 - ▶ weighting applied

Figure: Feynman diagram of the top-antitop quark pair decay.



	2 Jets	3 Jets	4+ Jets
Electron	1.13	12.24	39.09
Muon	0.88	10.98	39.34

Figure: Example of 2D data sample from MC channel Muon 4+ Jets. There are no clusters visible.

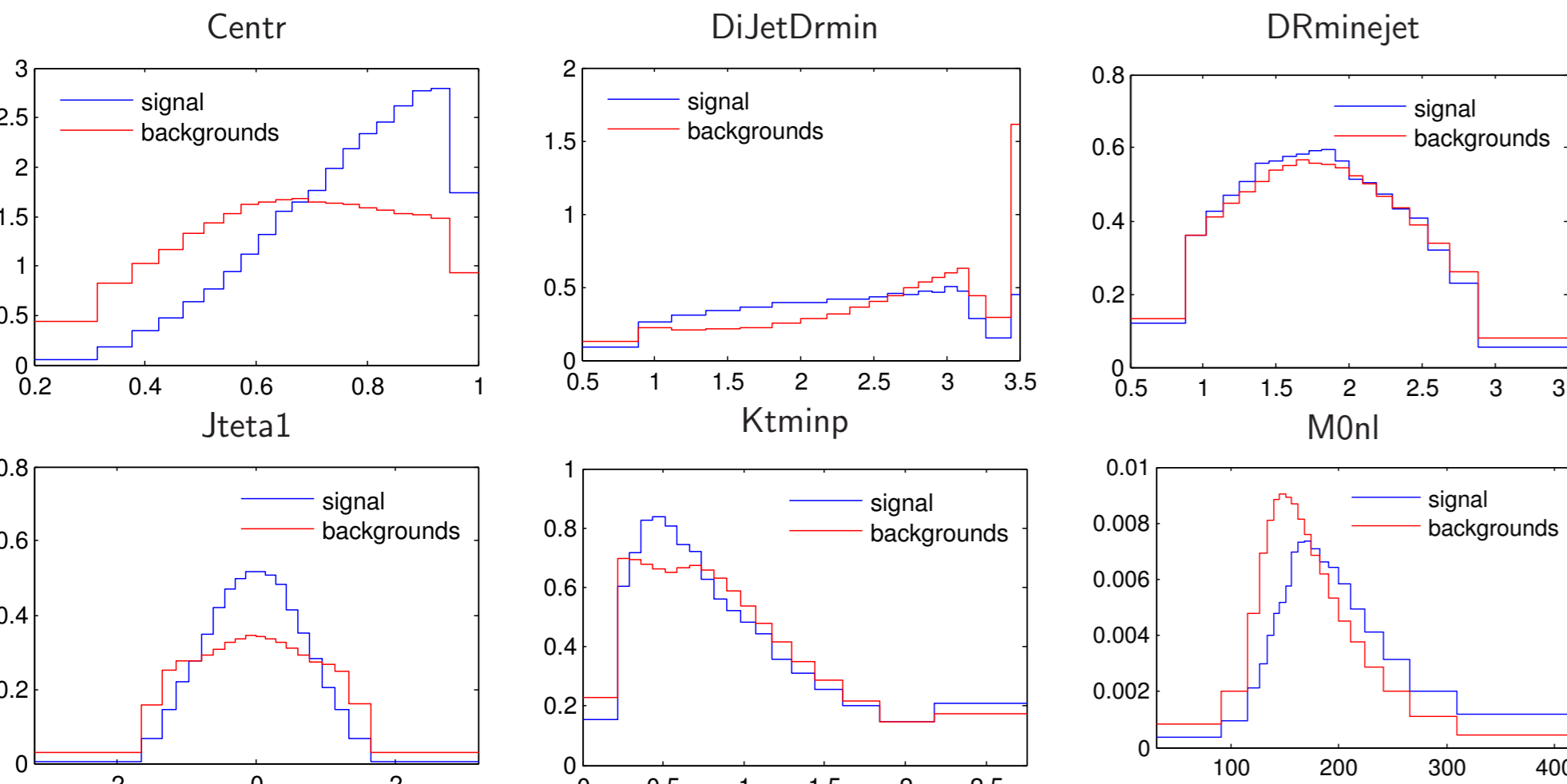


Figure: Quantile histograms of signal vs background distribution in MC channel Electron 2 Jets. The univariate distributions of signal/backgrounds are very similar.

II. RESULTS

- ▶ divergence analysis of signal separation potential

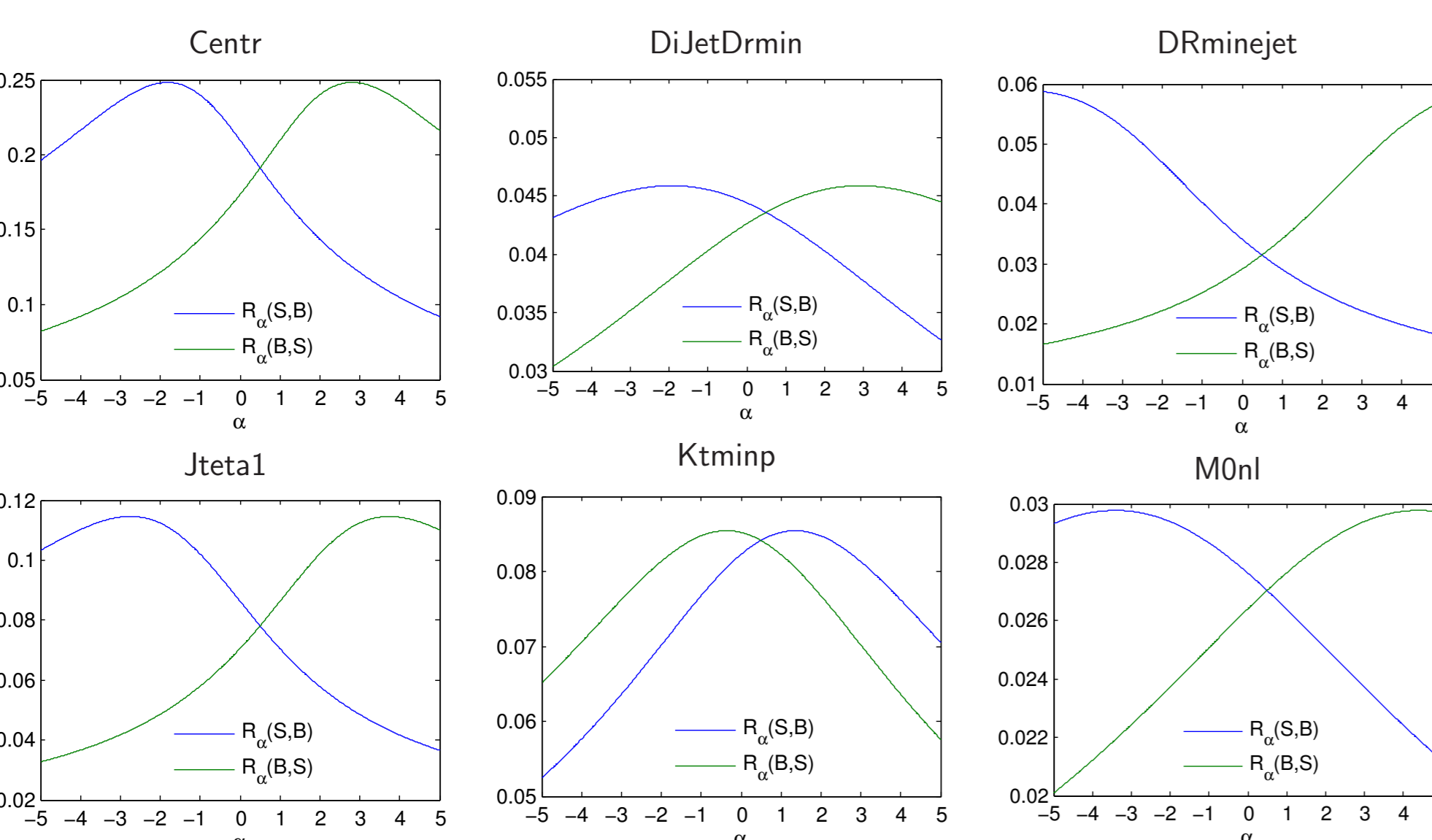


Figure: Dependence of the normalized Rényi distance for signal and background distributions on the parameter α . The higher value of the Rényi divergence corresponds to the higher signal separability.

- ▶ using only the best 12 variables yields 80 % AUC
- ▶ rejection of variable Mva_max with high signal discrimination potential (85 % AUC using only Mva_max)

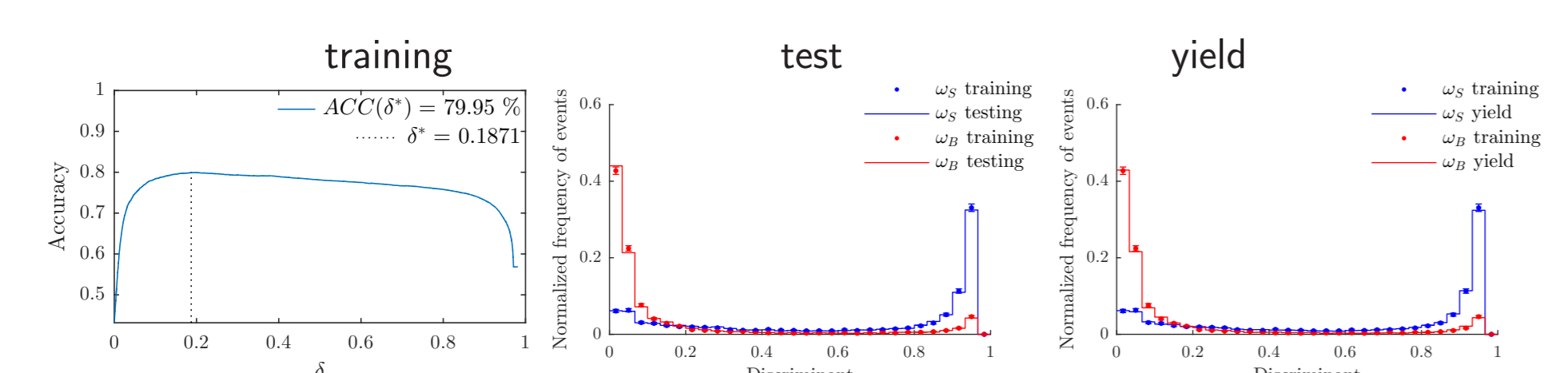


Figure: Control plots for classification using only the variable Mva_max.

Final signal separation results

- ▶ SDDT method with nested KDE utilized for final $D\bar{O}$ signal discrimination
- ▶ good results despite significant restriction regarding variable selection from homogeneity testing phase

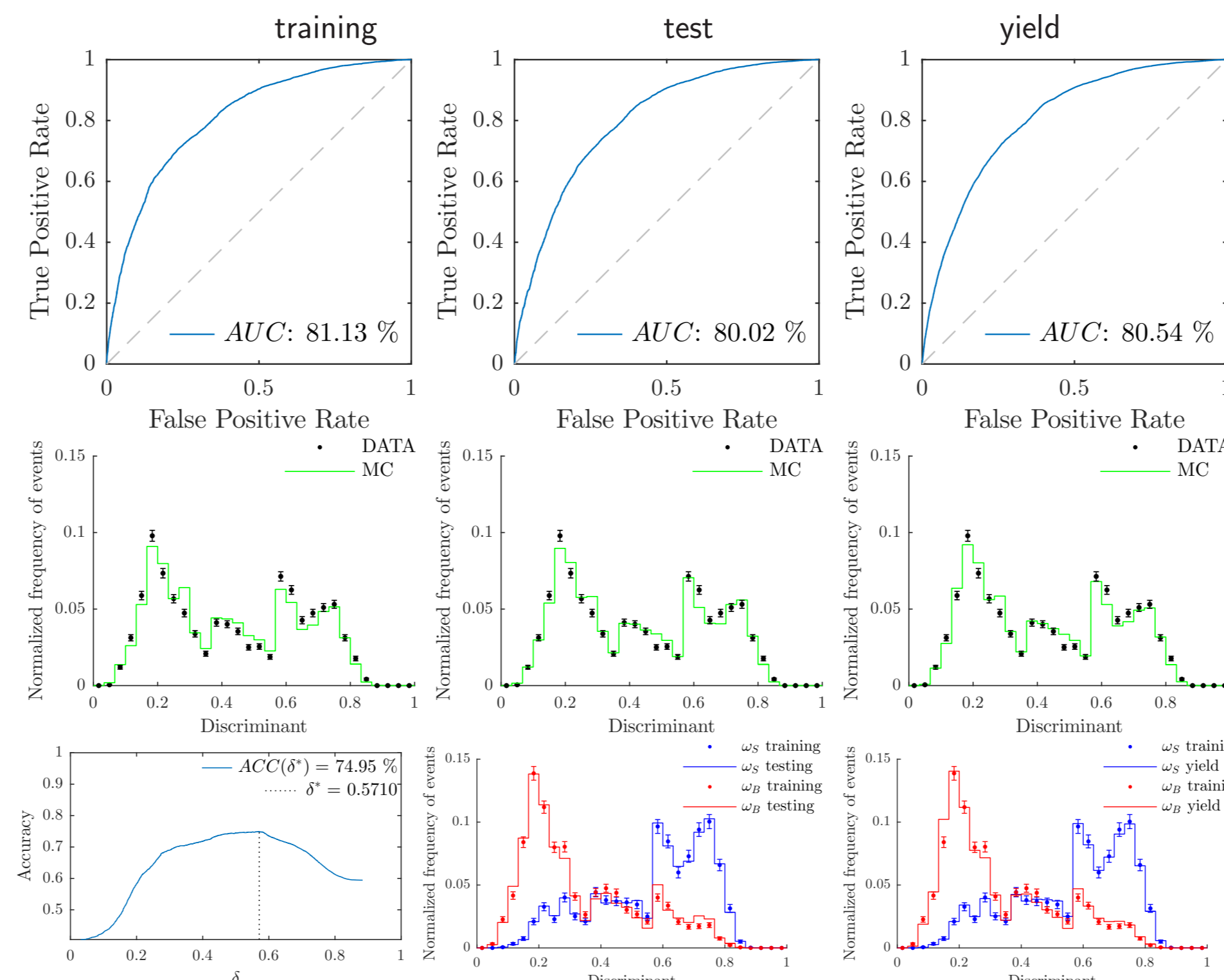


Figure: SDDT-KDE final classification of MC channel Muon 4+ Jets using only 41 eligible variables, excluding Mva_max.

III. APPROACH

- ▶ separation potential analysis via the normalized Rényi distance of order α , $\alpha \in \mathbb{R}$, $\alpha \neq 0$, $\alpha \neq 1$, for distributions P, Q with densities p, q :

$$R_\alpha(P, Q) = \frac{1}{\alpha(\alpha-1)} \int p^\alpha q^{1-\alpha} d\mu$$

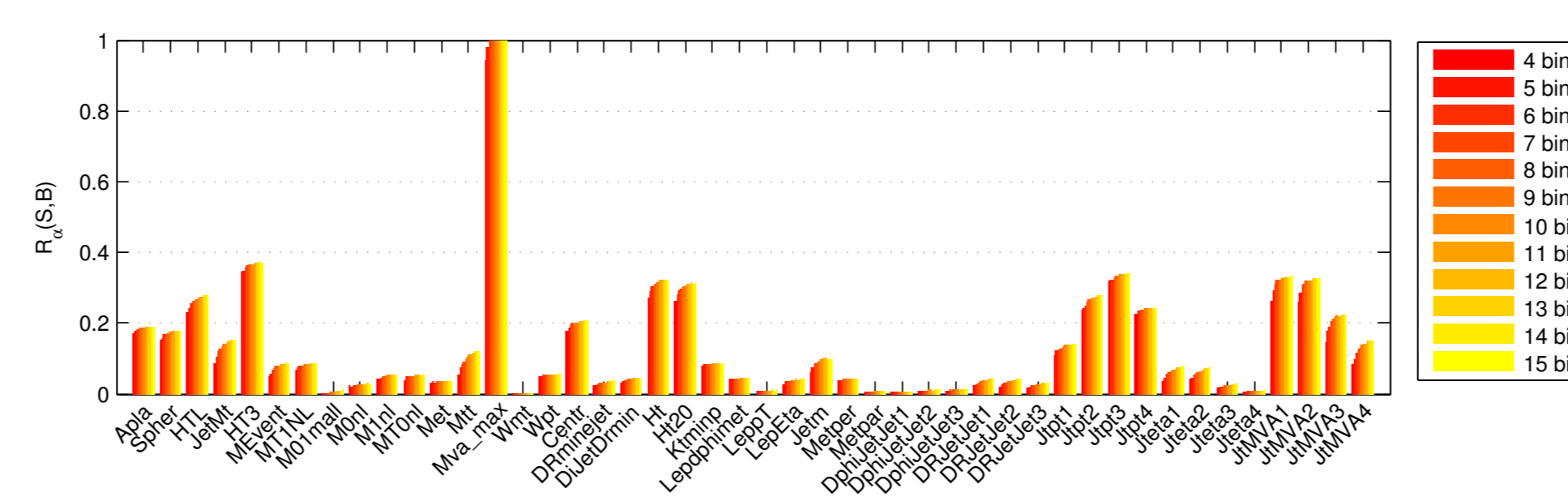


Figure: Normalized Rényi distance of order $\alpha = \frac{1}{2}$ in MC channel Muon 4+ Jets. Note the divergence is not dependent upon the number of bins.

Modified homogeneity testing

- ▶ rigorous statistical approach applicable to weighted data
- ▶ Two sample Kolmogorov-Smirnov test:

$$p\text{-value} = 2 \sum_{k=1}^{\infty} (-1)^{k-1} e^{-2k^2 \lambda_0^2}$$

where $\lambda_0 = \sqrt{\frac{W_1 W_2}{W_1 + W_2} D_{W_1, W_2}}$ and $D_{W_1, W_2} = \sup_{x \in \mathbb{R}} |F_{n_1}^{W_1}(x) - G_{n_2}^{W_2}(x)|$, where $F_{n_1}^{W_1}, G_{n_2}^{W_2}$ denotes the weighted empirical distribution function of signal or backgrounds, resp.

- ▶ Two sample Anderson-Darling test statistic:

$$A_{n_1 n_2}^2 = \frac{n_1 n_2}{N} \int_{-\infty}^{+\infty} \frac{[F_n(x) - G_n(x)]^2}{H_n(x)[1 - H_n(x)]} dH_n(x)$$

where H_n denotes the EDF of the pooled sample, i.e., $H_n(x) = [n_1 F_{n_1}(x) + n_2 G_{n_2}(x)]/N$ and $N = n_1 + n_2$ (weighted modification analogous to the Kolmogorov-Smirnov test)

- ▶ Divergence tests statistic:

$$H_\phi(\hat{\theta}) = \frac{2N}{\phi''(1)} D_\phi(\hat{p}, p^*(\hat{\theta})) = \frac{2N}{\phi''(1)} \sum_{j=1}^2 \sum_{i=1}^k \frac{n_i N_j}{N N} \phi\left(\frac{N_{ij} N}{n_i N_j}\right)$$

where ϕ is a certain function from the non-negative valued convex family

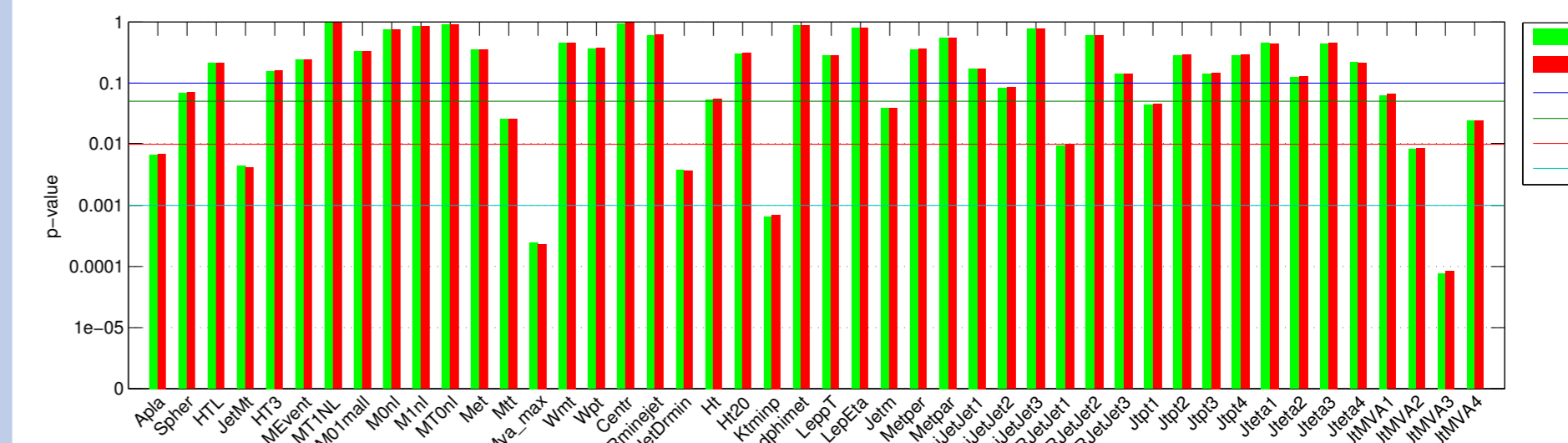


Figure: Numerical verification of modified weighted Kolmogorov-Smirnov test: p-value for all variables in weighted MC (modified test) channel Electron 4+ Jets and modified ensemble MC_{trans} without weighting (standard test).

Adaptive kernel density estimation (AKDE)

- ▶ whitening/sphering modification of standard KDE:

$$\hat{f}(t) = \frac{1}{nh^d \sqrt{\det(\hat{\Sigma})}} \sum_{j=1}^n K\left(\frac{t - X_j}{h}\right)$$

where $\hat{\Sigma}$ is an estimate of the data covariance matrix

- ▶ adaptive bandwidth with respect to the local density:

$$\hat{f}_\beta(t) = \frac{1}{n} \sum_{j=1}^n \frac{1}{(h\lambda_j)^d} K\left(\frac{t - X_j}{h\lambda_j}\right)$$

for $0 \leq \beta \leq 1$, where $\lambda_j = \left(\frac{f_\beta(X_j)}{g_n}\right)^{-\beta}$, $g_n = \sqrt[n]{\prod_{j=1}^n f_\beta(X_j)}$ and $\hat{f}_\beta(X_j) > 0$ for all $j \in \hat{n}$ is a pilot estimate

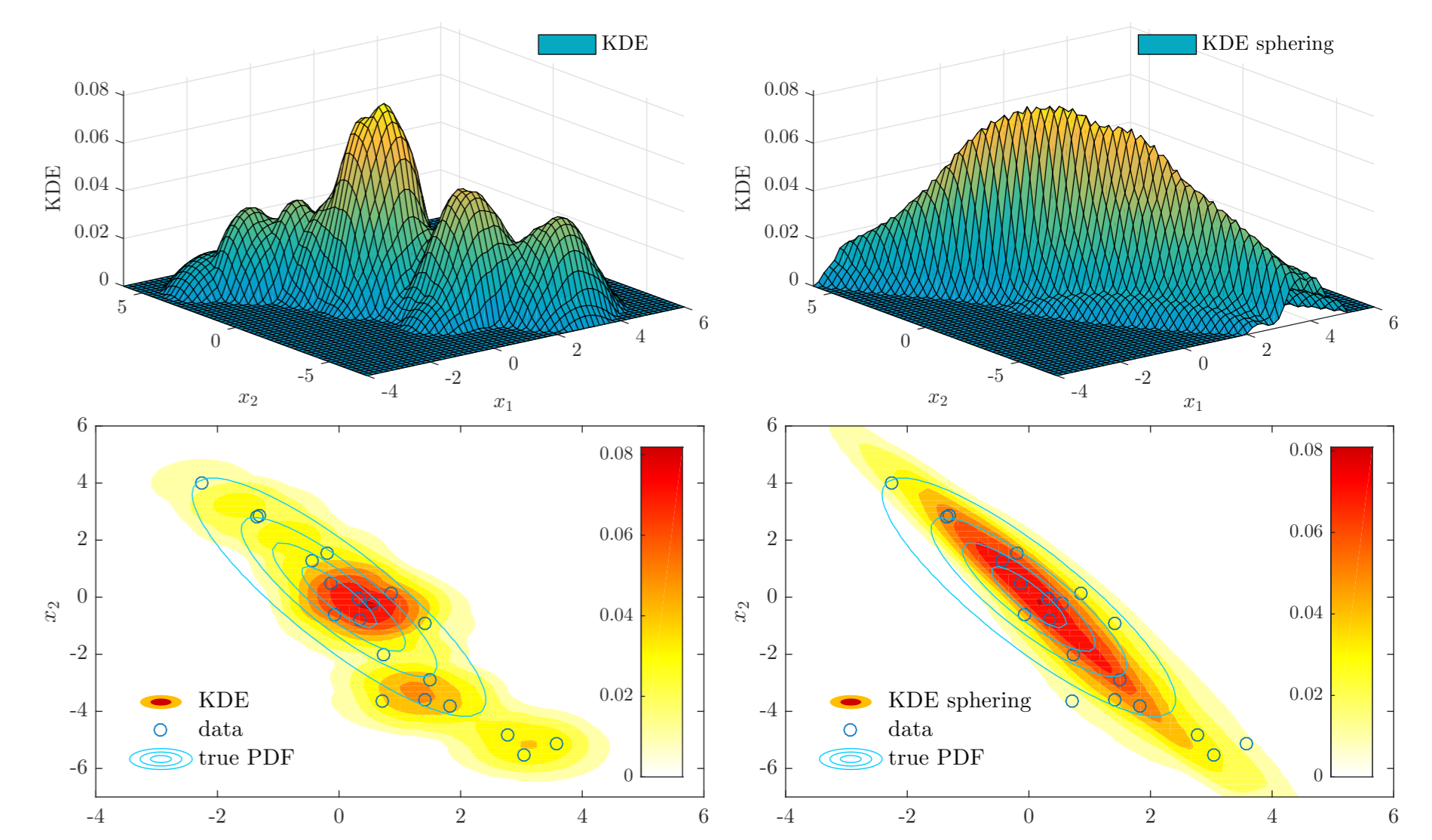


Figure: Density estimation examples using standard KDE and KDE with sphering transform.

Supervised divergence decision tree (SDDT)

- ▶ universal classification algorithm based upon certain localization of separation among the nodes of the tree
- ▶ locally selects the j^* -th combination of variables such that the ϕ -divergence $D_\phi^{(j^*)}(S, B)$ is maximal for the current node: $j^* = \arg \max_{j \in \{1, \dots, (q)\}}$ $D_\phi^{(j)}(S, B)$
- ▶ employs a nested separation method (AKDE/SVM/GLM) using only the selected variables in the current node

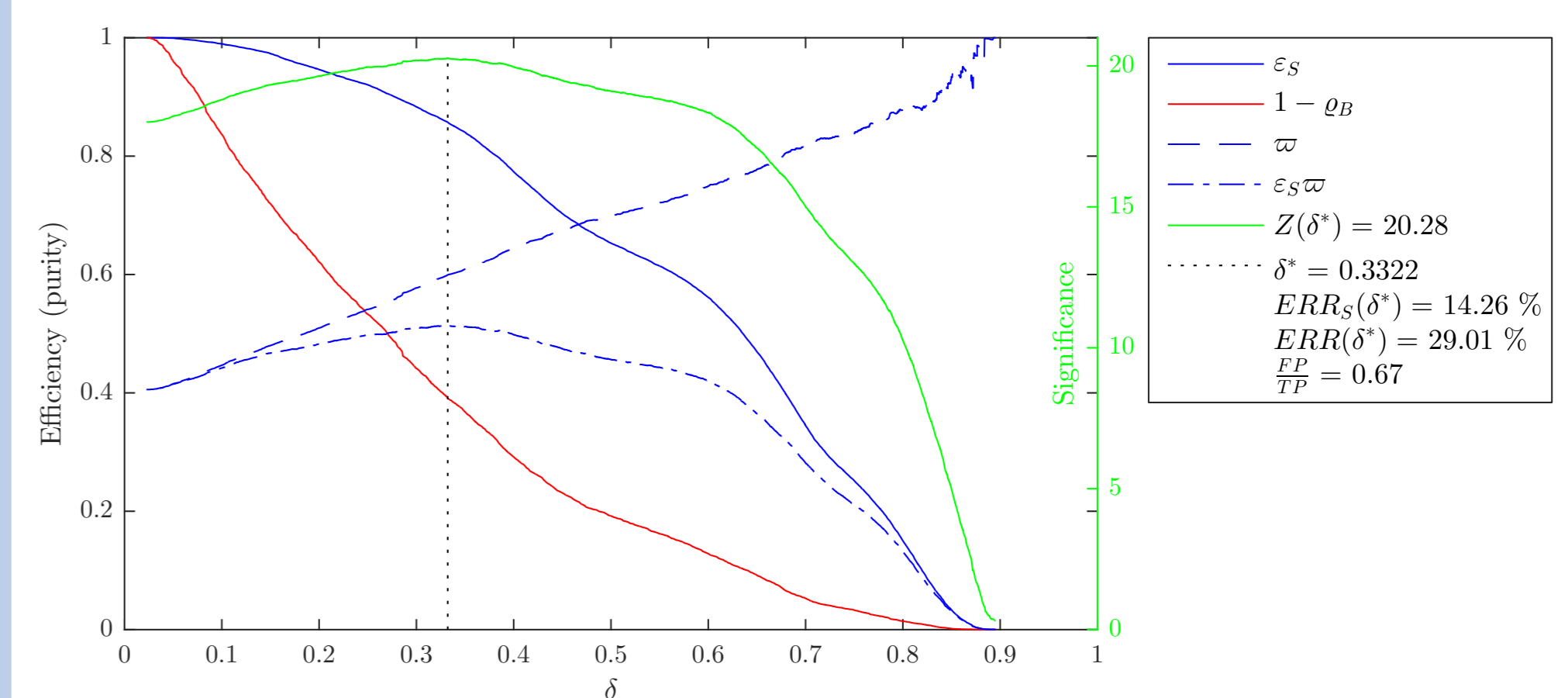


Figure: Figures of merit and optimal discriminant cut for SDDT-AKDE classification of MC channel Muon 4+ Jets (signal efficiency ϵ_S , background rejection ϵ_B , signal purity ω_S , significance Z).

IV. DISCUSSION

- ▶ AKDE smooths and polarize the final discriminant distribution in SDDT classification

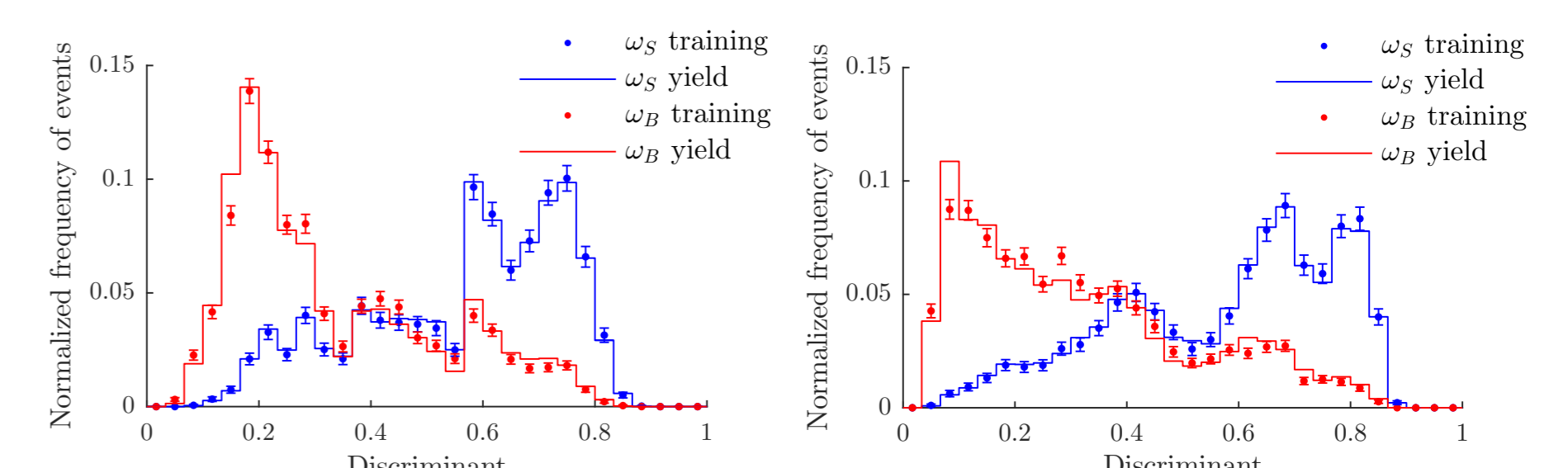


Figure: Control plots for final SDDT classification with KDE (left) and AKDE (right). There is a better correspondence between training and test/yield discriminant distribution for SDDT-AKDE (p-value of 0.145 whereas 0.063 for SDDT-KDE).

- ▶ SDDT-AKDE improves overall classification up to 82 % AUC

Acknowledgements

This work was supported by the grants LM2015068 (MYES), LG15047 (MYES), GA16-09848S (GACR) and SGS15/214/OHK4/3T/14 (CTU).