Machine learning techniques for heavy-flavour baryon production measurements at the LHC

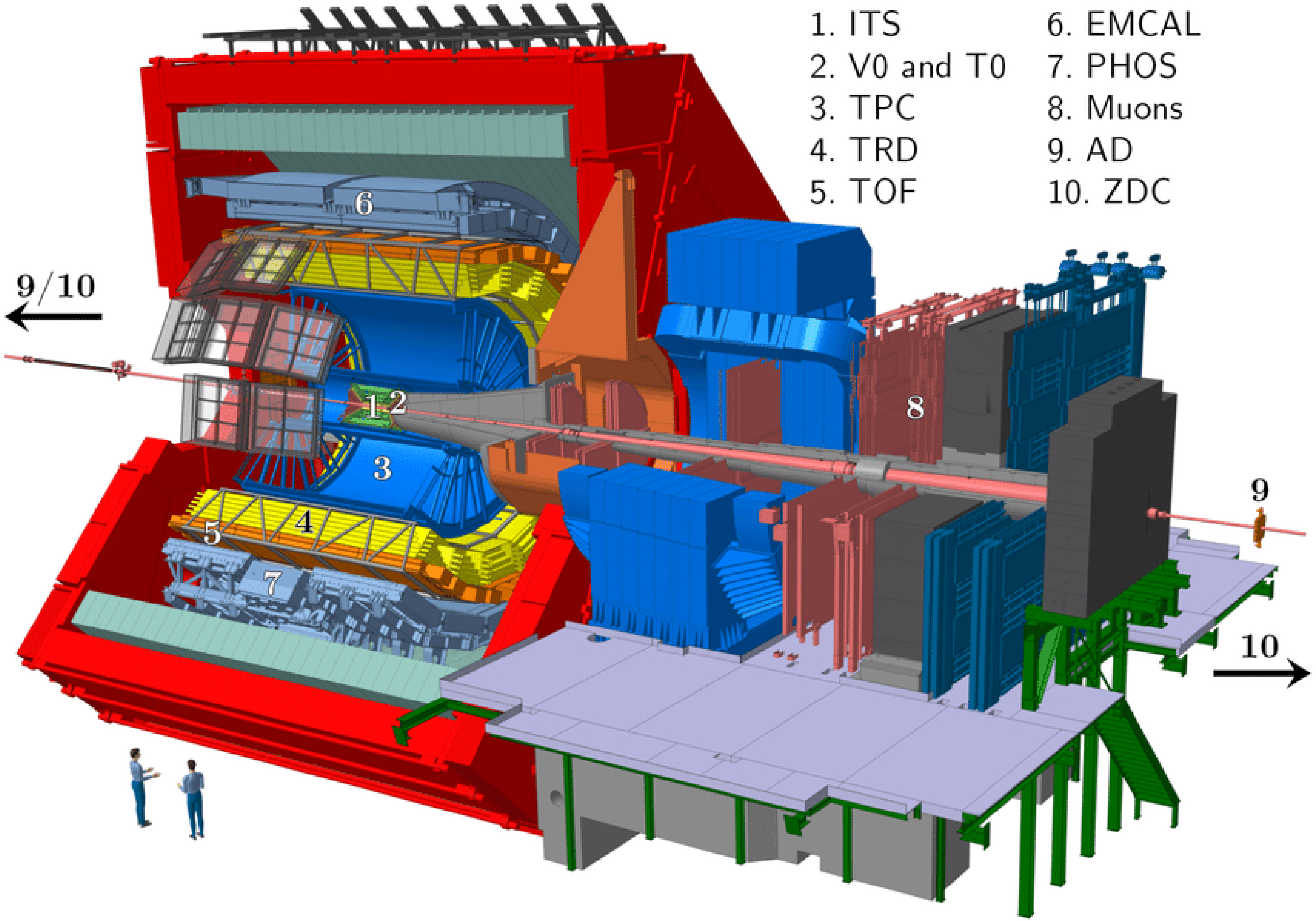
THESIS SUPERVISOR:
PROF. ANDREA ALICI

PRESENTED BY:
MARCO CRUCIANI

The ALICE detector at LHC

Heavy-ion collisions to study:

- the properties of strongly interacting matter
- quark-gluon plasma (QGP)

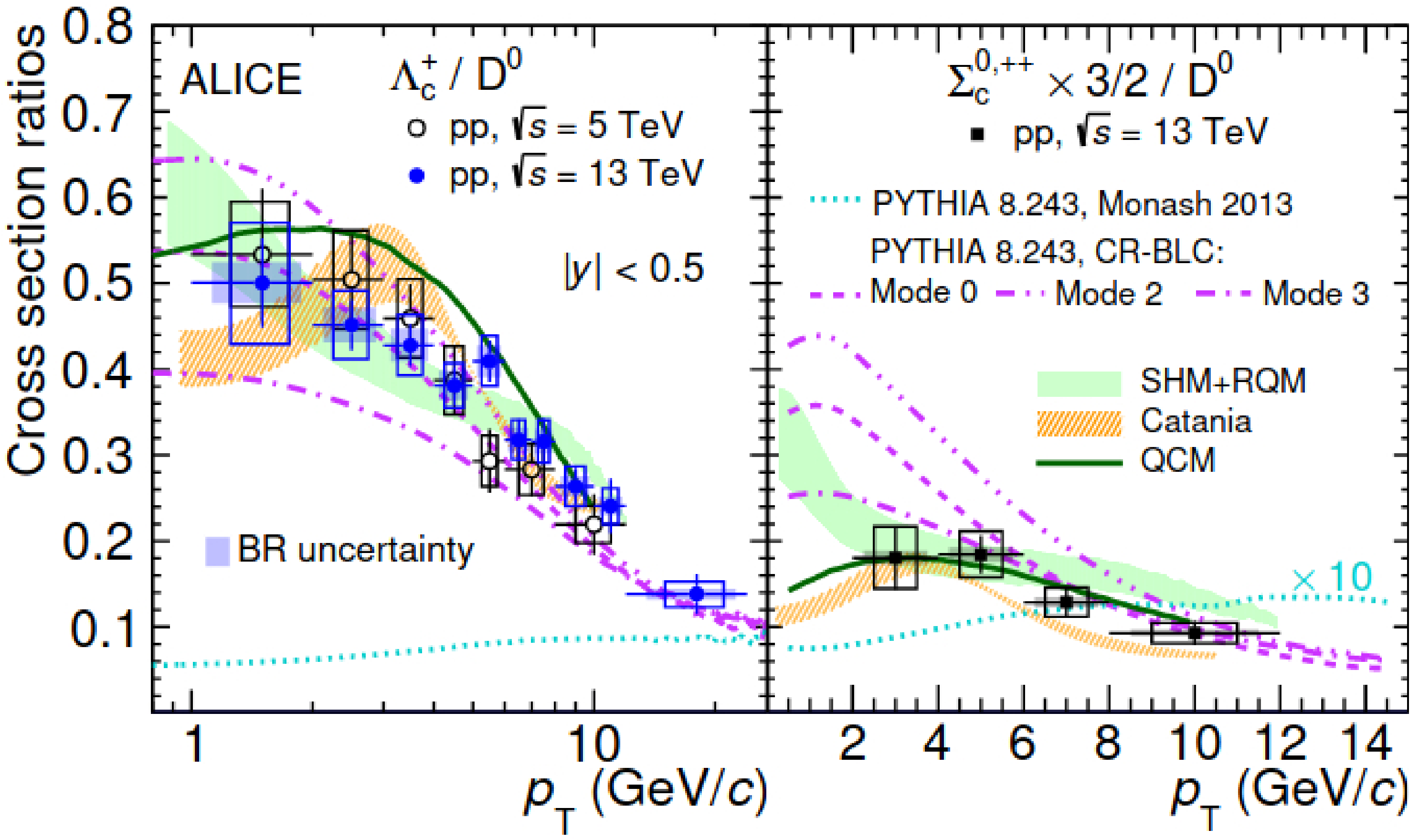


Different hadronization models

- Cross section ratios of Λ_c^+/D^0 and Σ_c/D^0 for pp collisions at 5 TeV and 13 TeV
- PYTHIA Monash does not reproduce data

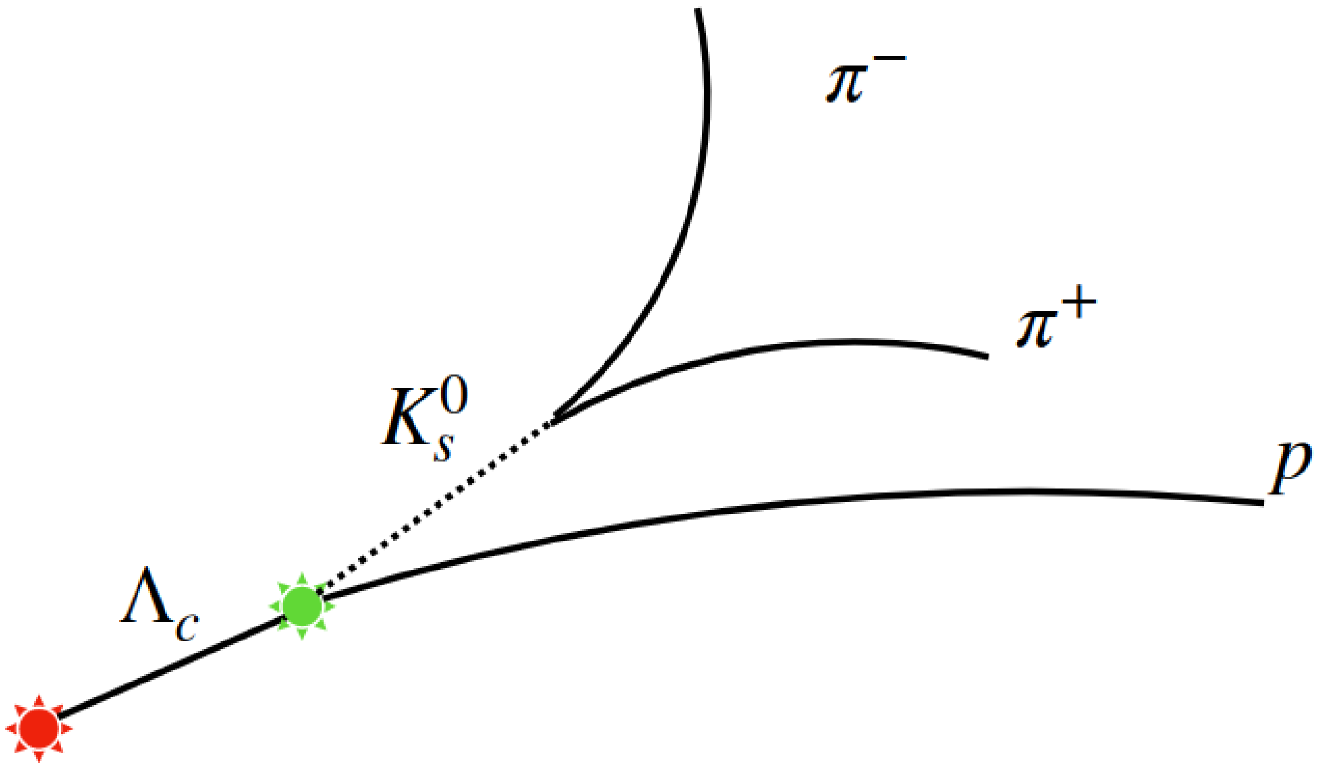
Factorization theorem

$$\frac{d\sigma_{pp}^{h}}{dyd^2p_T} = K\sum_{abcd}\int dx_a dx_b f_a(x_a,Q^2) f_b(x_b,Q^2)\frac{d\sigma}{d\hat{t}}(ab\rightarrow cd)\frac{D_{h/c}^0}{\pi z_c}$$



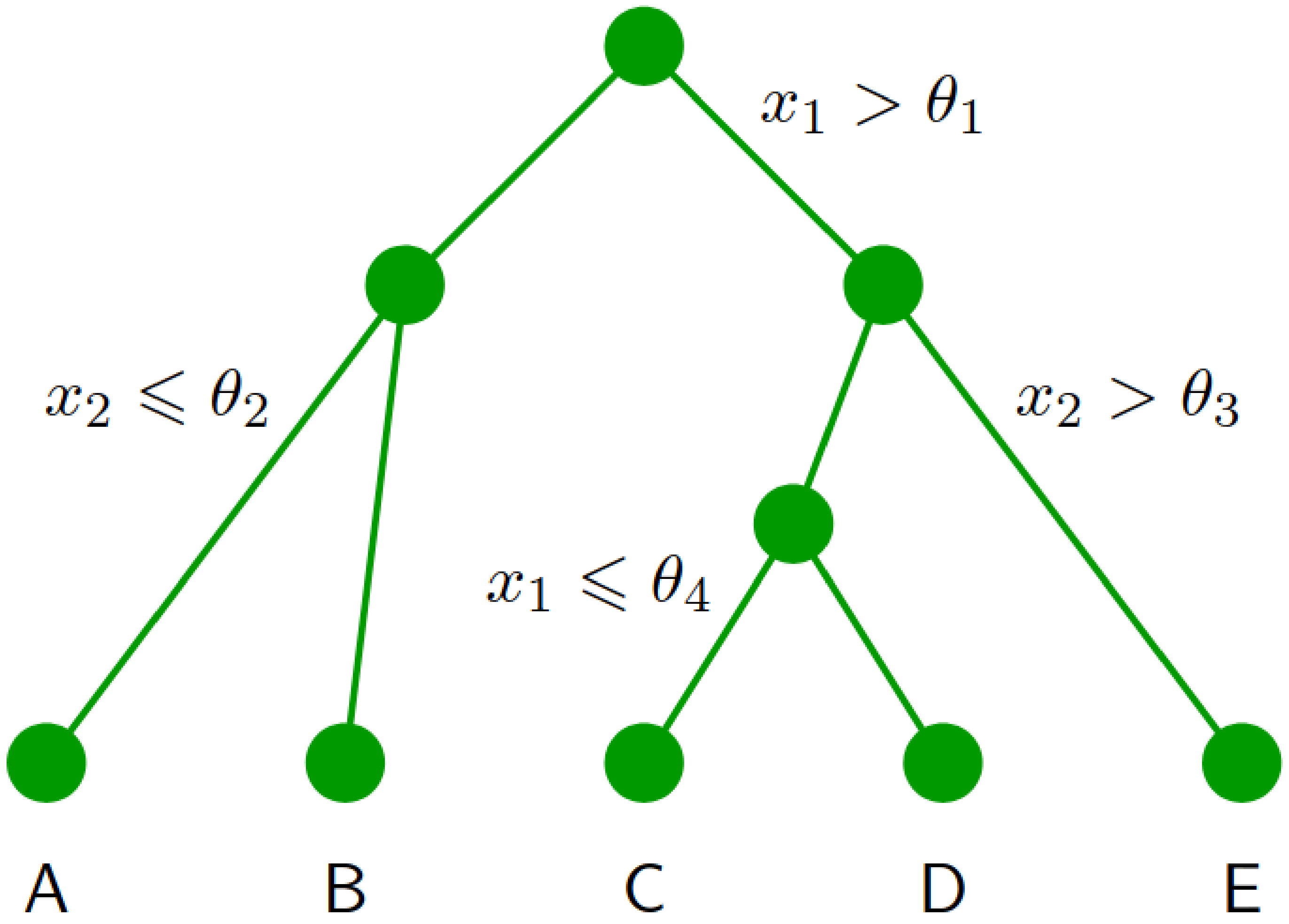
The decay channel under consideration

- $\Lambda_c \to p\, K_s^0$
- BR = (1.59 ± 0.08)%
- resolution for vertex reconstruction is $\sim 100\ \mu m$ while Λ_c decays in $\sim 60\ \mu m$



Decision trees

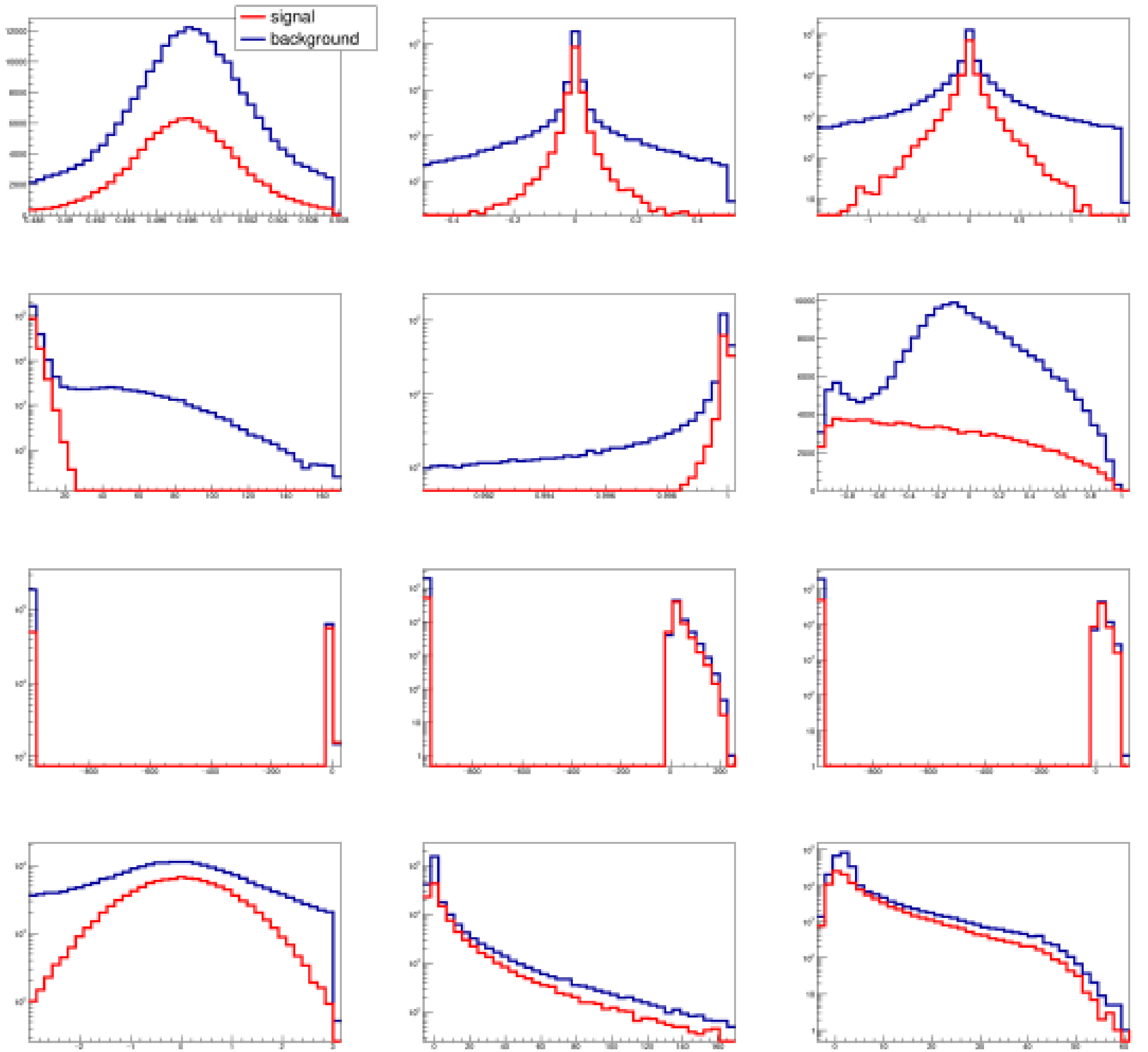
- Machine learning method available in the ROOT library TMVA
- Simple classification based on a set of binary questions
- The boosting process forms a new classifier by combining the information from several decision trees



Input variables

massK0s	nSigmaTOFpr
tImpParBach	nSigmaTOFpi
tImpParV0	nSigmaTOFka
ctK0s	nSigmaTPCpr
cosPaK0S	nSigmaTPCpi
CosThetaStar	nSigmaTPCka

Distribution of input variables in the range $p_T \in [2,4]\ GeV/c$

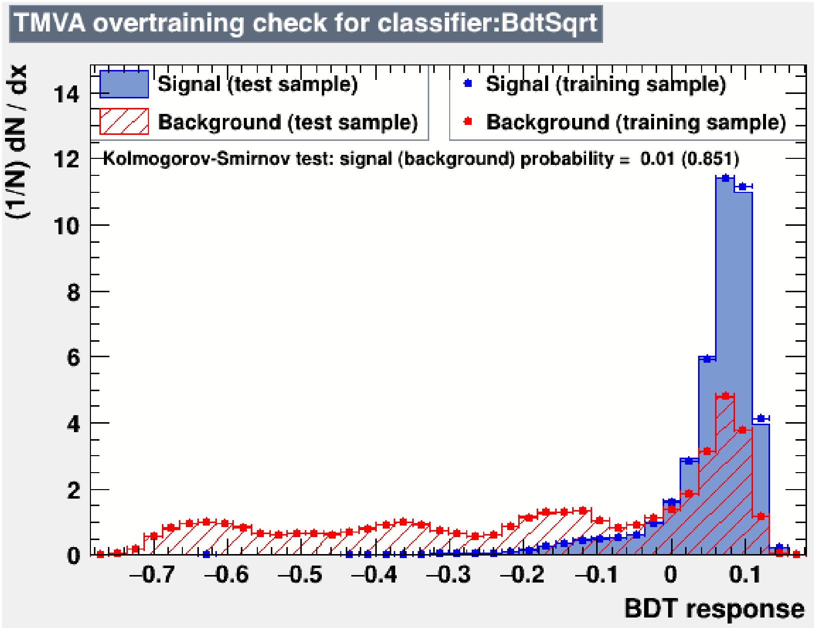


The three methods

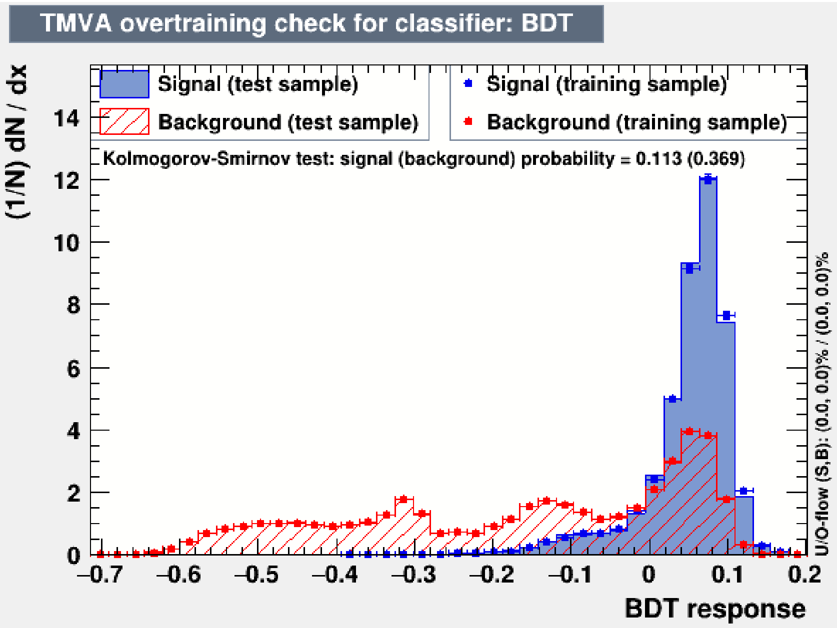
- Boosted Decision Trees (BDT)
- BDT Category (BdtCat), we separate into two sets one with TOF variables available and the other with missing TOF variables
- BDT root sum square (BdtSqrt), where identification variables from TOF and TPC are used together.

$$if\ TOF > -999 \Rightarrow n_\sigma(p) = \sqrt{n^2_{\sigma,TOF}(p) + n^2_{\sigma,TPC}(p)}$$

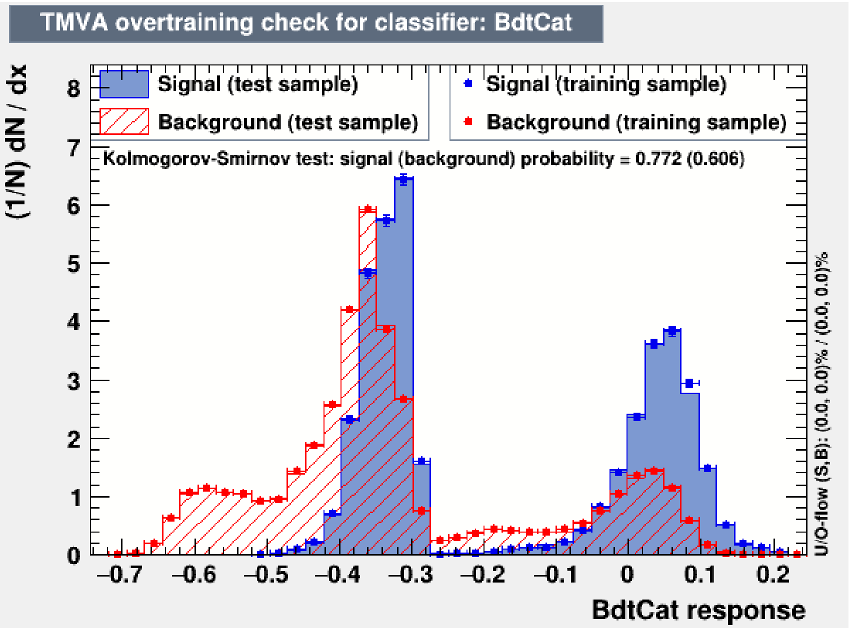
$$otherwise \Rightarrow n_\sigma(p) = n_{\sigma,TPC}(p)$$



BdtSqrt



BDT



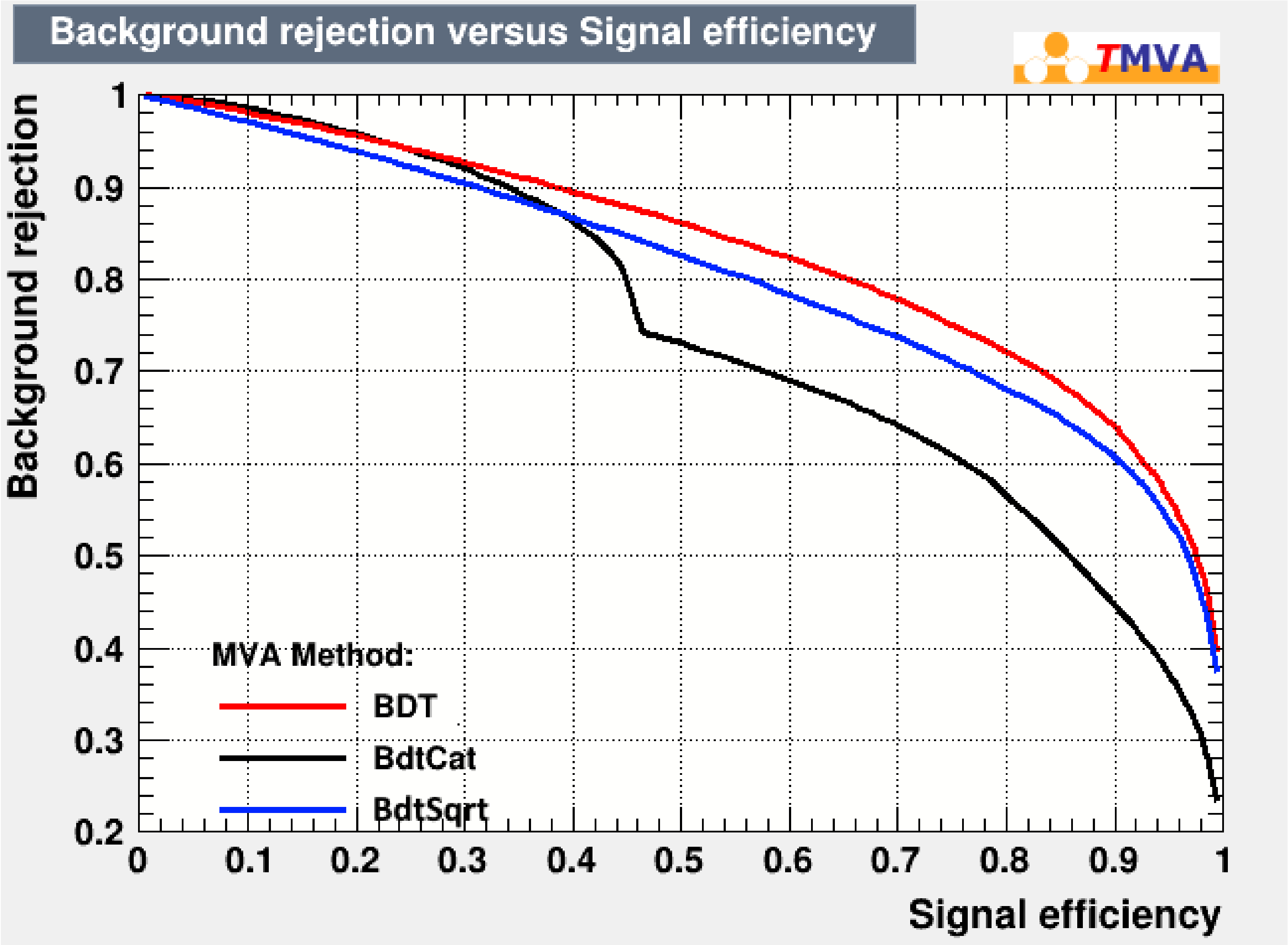
BdtCat

BDT response for the three different methods

$(p_T \in [0,1]\ GeV/c)$

ROC curves

p_T	BDT	BdtSqrt	BdtCat
[0,1]	0,831	0,804	0,744
[1,2]	0,841	0,819	0,802
[2,4]	0,841	0,783	0,795
[4,6]	0,816	0,743	0,714
[6,8]	0,845	0,781	0,746
[8,12]	0,868	0,808	0,737

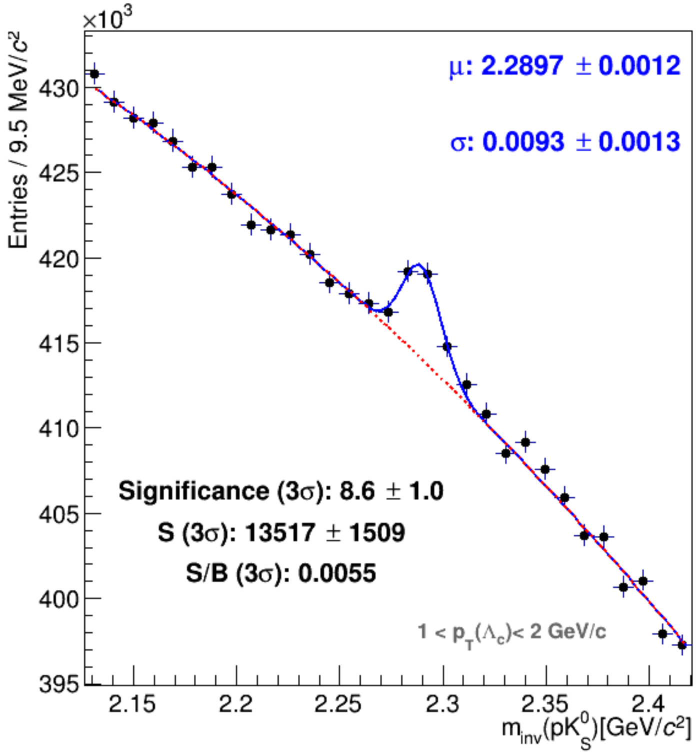


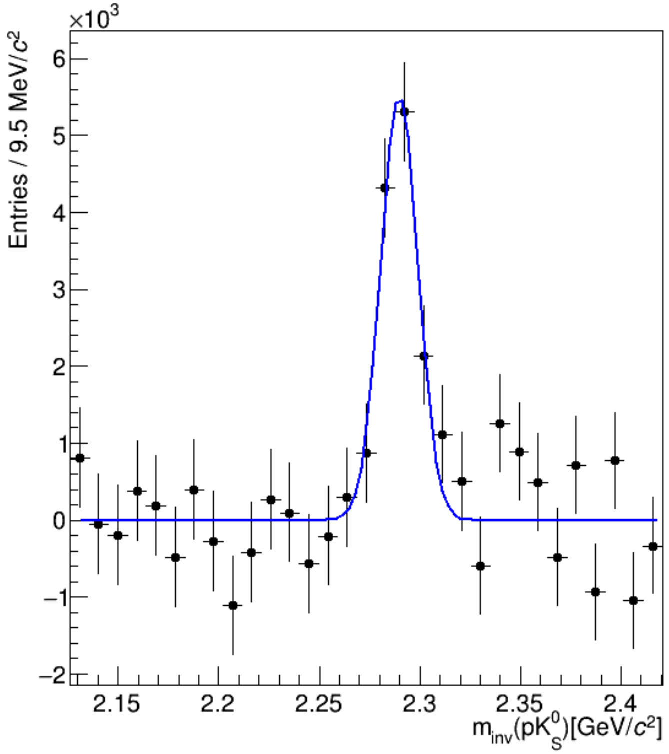
$p_T \in [0,1]\ GeV/c$

Reconstructed Λ_c

- Signal fitted with gaussian
- Background fitted with a 2^{nd} or 3^{rd} degree polynomial

p_T	Signale (3σ)	Signific. (3σ)
[0,1]	4748 ± 1193	3.9 ± 1.0
[1,2]	13517 ± 1509	8.6 ± 1.0
[2,4]	18591 ± 1264	13.8 ± 0.9
[4,6]	7413 ± 523	13.1 ± 0.9
[6,8]	2034 ± 183	11.0 ± 1.0
[8,12]	770 ± 91	7.9 ± 0.9





$p_T \in [1,2]\ GeV/c$

Conclusions

- The out-of-the-box BDT method gives the best performance in spite of missing data
- We have shown a method that can estimate the production of Λ_c^+ baryons, and thus can allow estimation of the ratio of production baryon/meson Λ_c^+/D^0
- Improvement in these techniques and in experimental precision will allow to determine which model best describes hadronization mechanisms