A multivariate selection of charmless B^+ decays to three mesons in LHCb detector

J. Helder Lopes
IF-UFRJ

for the UFRJ-CBPF LHCb BnoC Working Group

July, 31th 2018

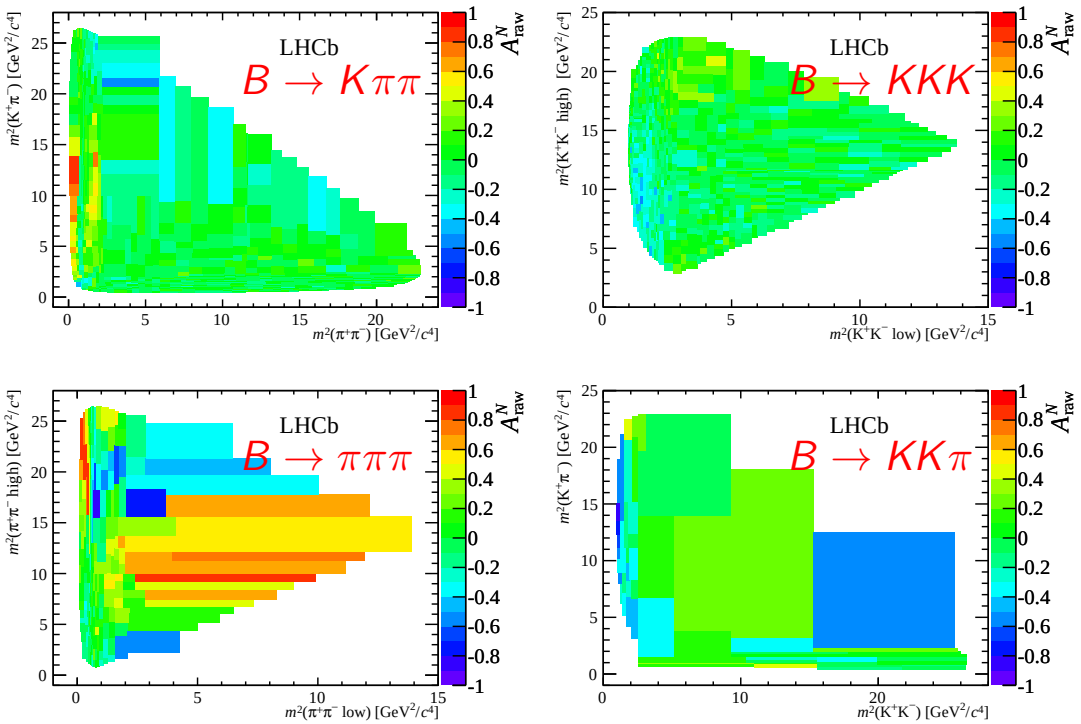
Overview

- Motivation for the analysis
- Motivation for the MVA
- MVA in Run I
- MVA in Run II
  - Update from Run I
  - New reduced set of variables
    - Motivation
    - New MVA optimization
    - Comparison with old one
    - MVA optimal cuts and invariant mass distributions
- Conclusions

Motivation

From Run I analysis *(Phys.Rev.D90(2014), no.11 112004)*

- Strong evidence for integrated CP asymmetry in the four channels
  - $A_{CP}(B^{\pm} \to K^{\pm}\pi^{+}\pi^{-}) = +0.025 \pm 0.004 \pm 0.004 \pm 0.007$, Significance: 2.8 σ
  - $A_{CP}(B^{\pm} \to K^{\pm}K^{+}K^{-}) = -0.036 \pm 0.004 \pm 0.002 \pm 0.007$, Significance: 4.3 σ
  - $A_{CP}(B^{\pm} \to \pi^{\pm}\pi^{+}\pi^{-}) = +0.058 \pm 0.008 \pm 0.009 \pm 0.007$, Significance: 4.2 σ
  - $A_{CP}(B^{\pm} \to \pi^{\pm}K^{+}K^{-}) = -0.123 \pm 0.017 \pm 0.012 \pm 0.007$, Significance: 5.6 σ
- Large asymmetries in regions of the phase space
  - Particularly large in the $\pi\pi \Leftrightarrow KK$ re-scattering region



Hot color: Excess of B^- Cold color: Excess of B^+

- Run I: 2011+2012
- Run II: 2015+2016 $\sim 2\times$Run I

Motivation for MVA

- Traditionally: Cuts on each variable of a set
  - You know what you are doing
  - An event is rejected based in just one variable
  - Hard for many variables
  - Computationally expensive to automate
  - Lower perfomance

- Multivariate methods (for classification): Combine the variables into a single one
  - Many ways to combine. Find the best for your case
  - Sometimes hard to understand/follow how they are combined
  - An event "rejected" by one variable may be recovered in the combination with others
  - Smooth phase space borders
  - In general, better performance

We still use loose cuts in trigger, reconstruction and pre-selections. But MVA is already there ...

Motivation for MVA

Some methods:

- Projective Likelihood Estimator
- Fisher discriminant
- **Artificial Neural Networks**
- **Boosted Decision Trees - BDT**
- Support Vector Machine
- ...

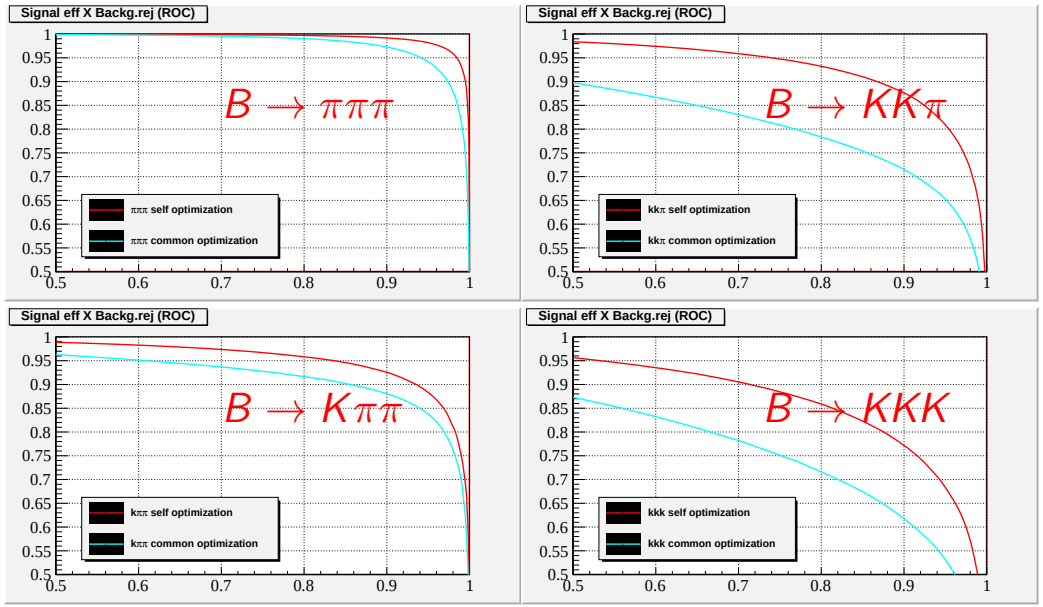
Some packages

- **TMVA: Toolkit for Multivariate Data Analysis**
  (http://tmva.sourceforge.net)
- NeuralBayes
  (https://twiki.cern.ch/twiki/bin/view/Main/NeuroBayes)
- scikit-learn: Machine Learning in Python
  (http://scikit-learn.org/stable)
- ...

MVA in Run I analysis (*Phys.Rev.D90(2014), no.11 112004*)

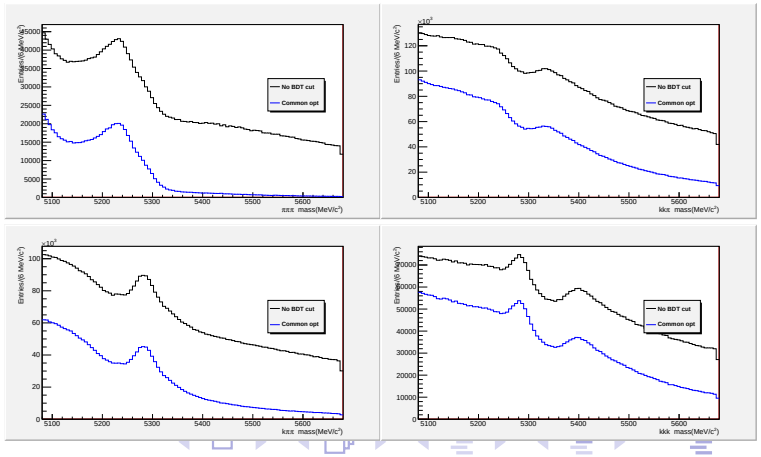
Note: Cuts used in Phys.Rev.Lett.112, 011801 (2014), Phys.Rev.Lett.111, 101801 (2013)

- Training:
  Signal: MC simulation
  Background: Data in a right side invariant mass window
- Input: 23 discriminating variables
- Data pre-processing: PCA (Principal component analysis)
- MVA method: BDT
- One single optimization and cut for all four channels
  Training with joint MC from all channels and background from $B \to \pi\pi\pi$



No PID selection

MVA cut: $BDTPCA > 0$



MVA in Run II analysis

- Data from 2015 and 2016. May include 2017
- Pre-selection:
  - Trigger, loose cuts (*stripping*), loose PID
- Update from Run I
  - Two more variables (25 in total)
  - Extended background region: $(5400 < B\_m < 6300)\ MeV/c^2$
  - Joint 2015+2016 data

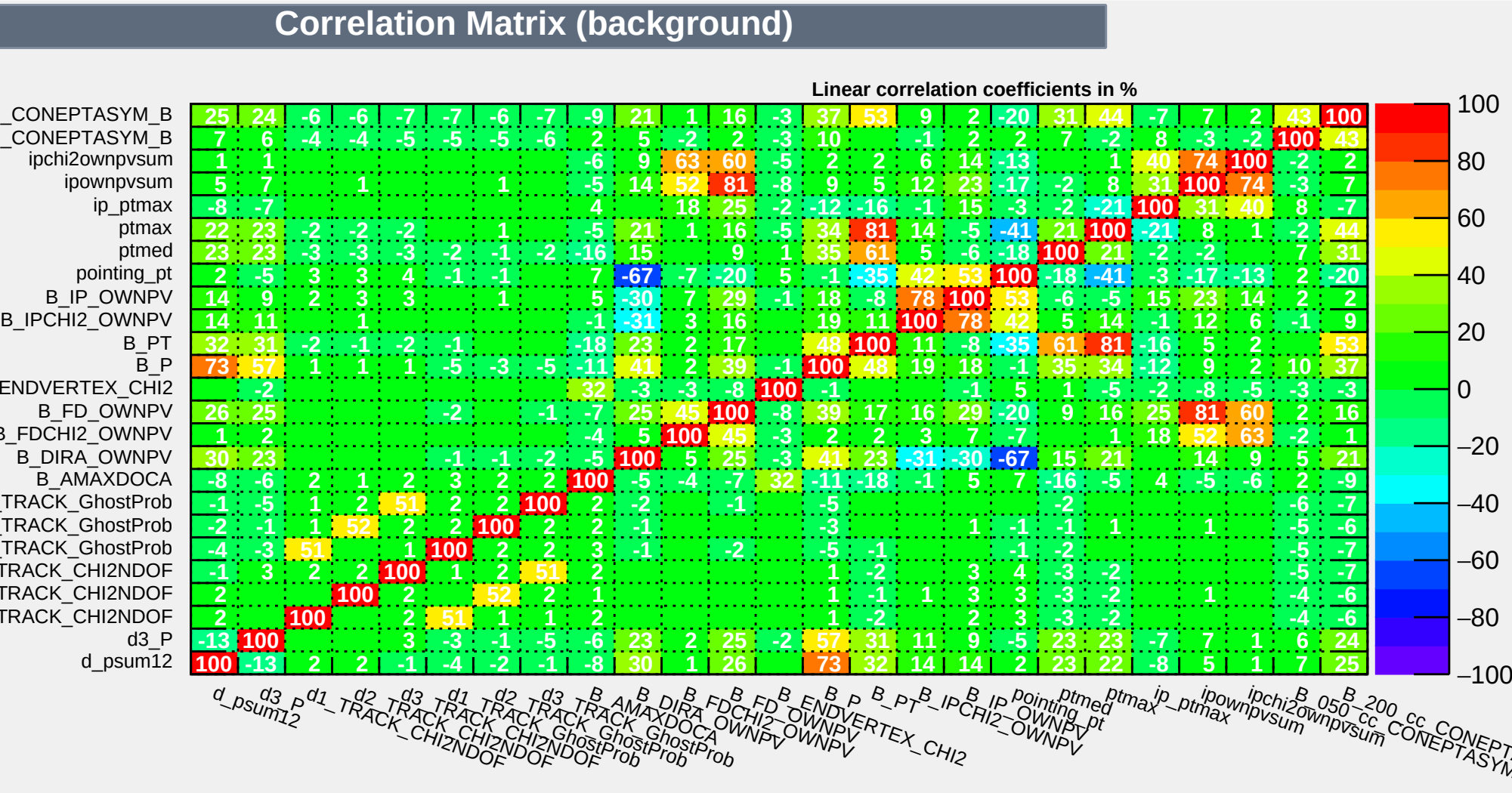
  ➠ Method: BDT with PCA
- New MVA optimization: Reduced set of input variables
  - Too many variables
  - Many are highly correlated
  - Some don't show good MC-data agreement

  ➠ Set of just 10 discriminating variables

  ➠ Method: BDT, no need for PCA

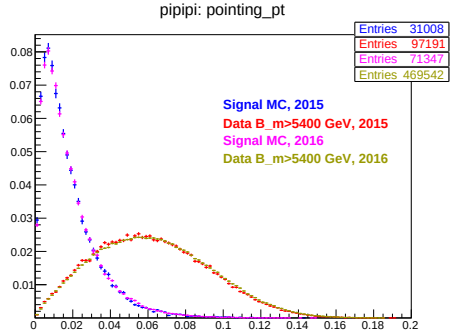
Showing some sample plots next slides

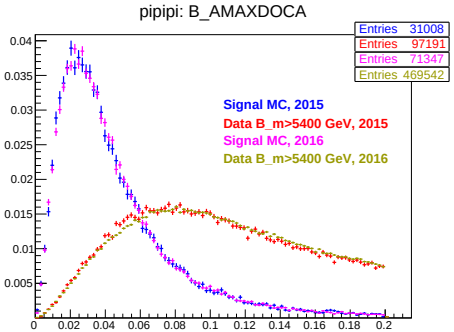
Correlations for $B \to K\pi\pi$ background 25 variables

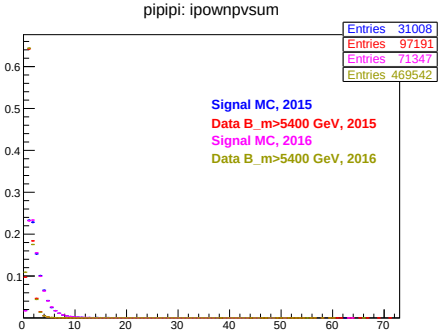


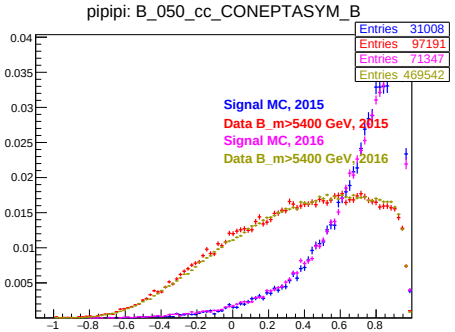
Similar correlations hold for the signal MC and for other channels

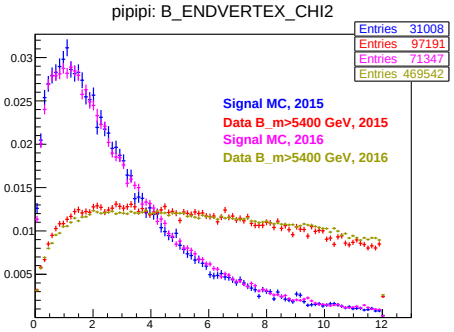
New input variables for $B \to \pi\pi\pi$

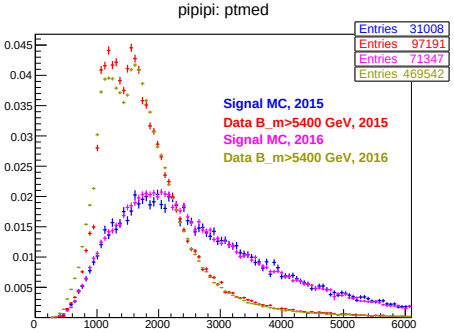


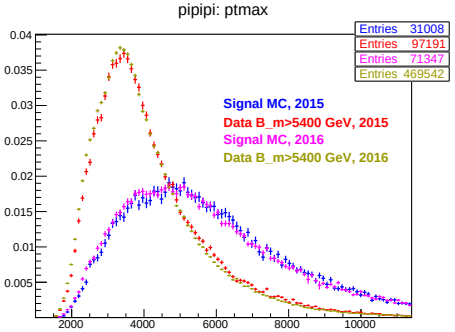


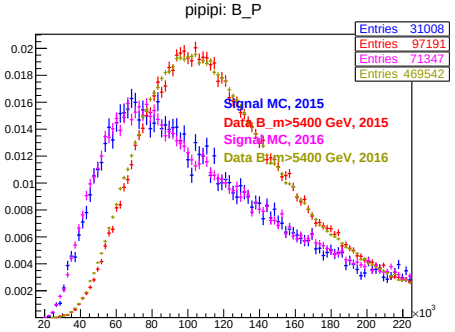


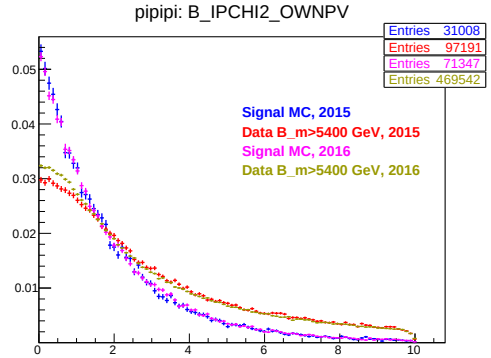


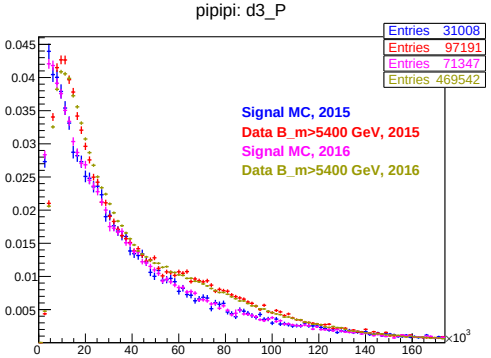




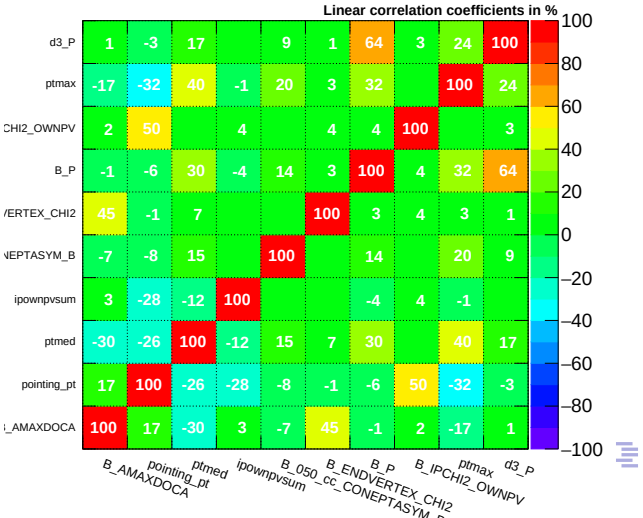




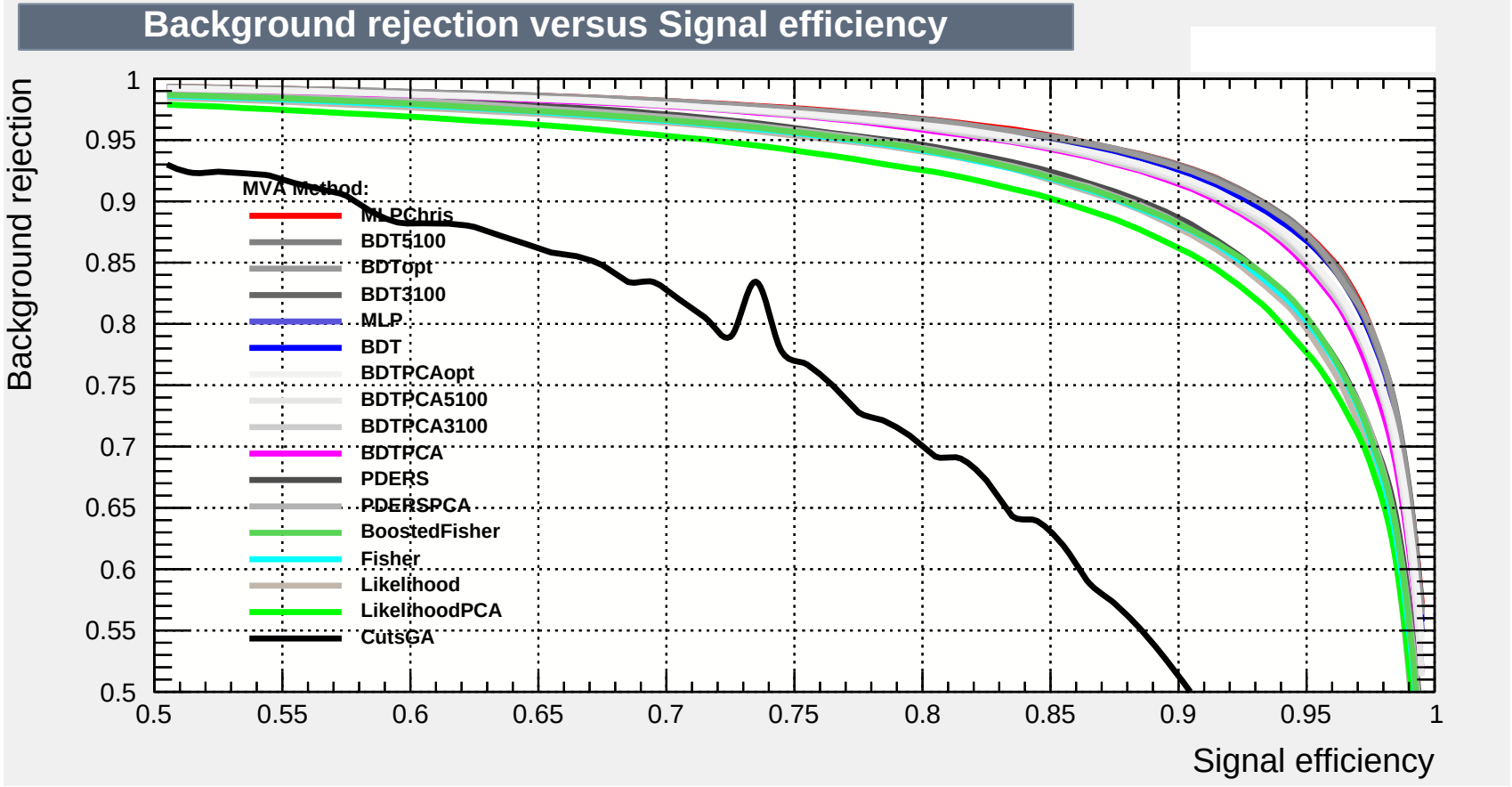




Correlation Matrix (signal) $B \to KK\pi$

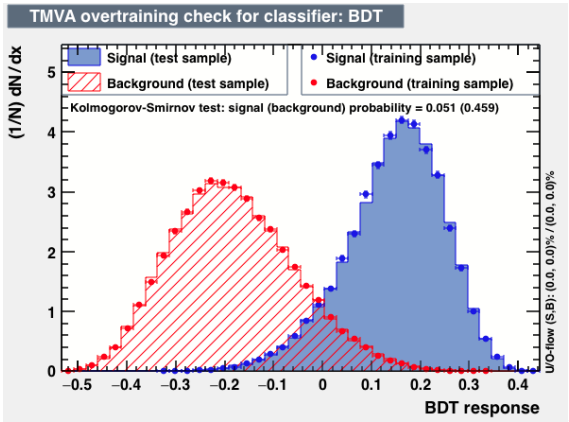


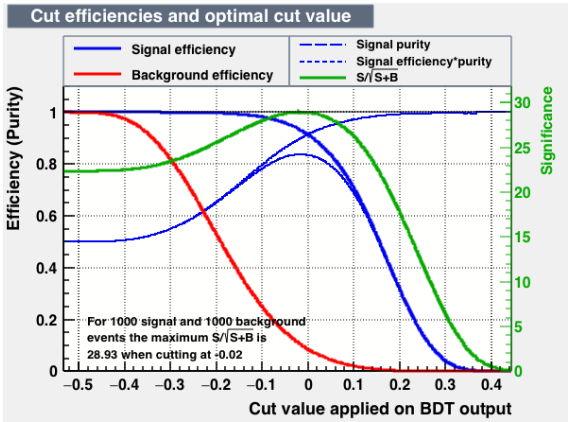
Choice of MVA method for $B \to K\pi\pi$



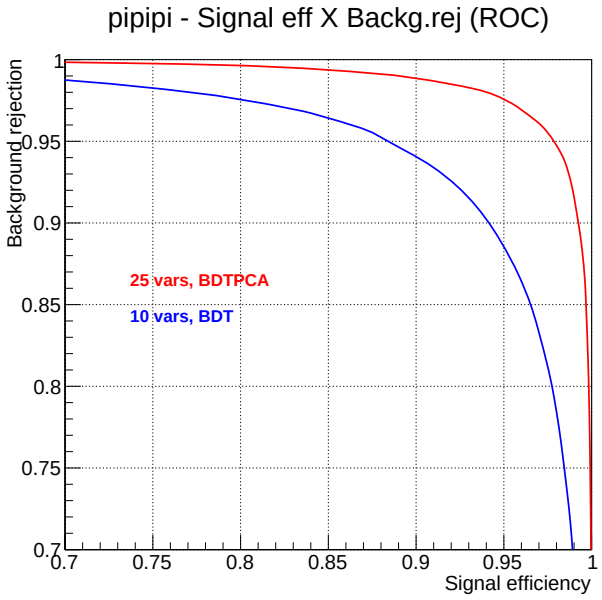
Signal(%) for $B = 10\%$
Overtraining checks

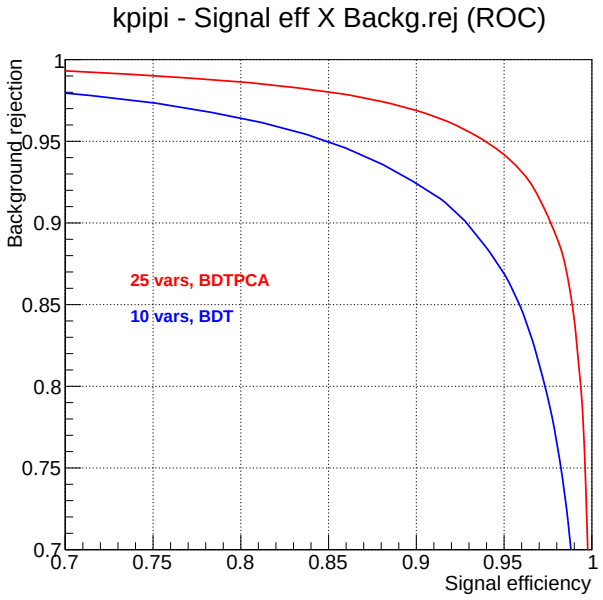
	test	train
MLPChris	93.2	93.5
BDT	92.7	92.8
BDTPCA	91.6	91.8

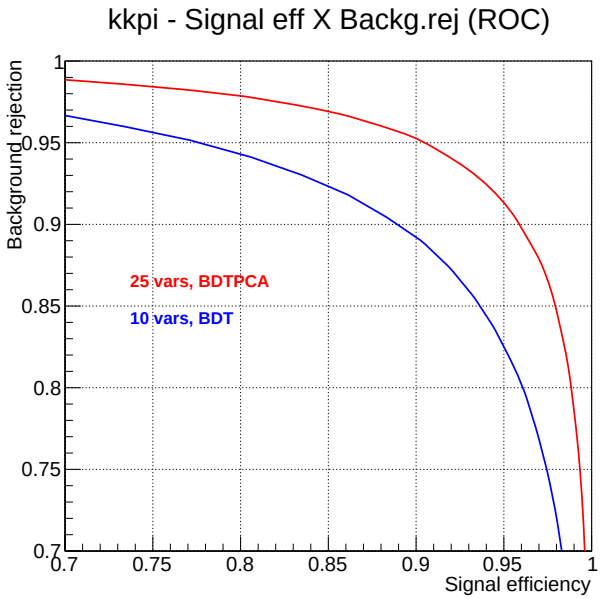


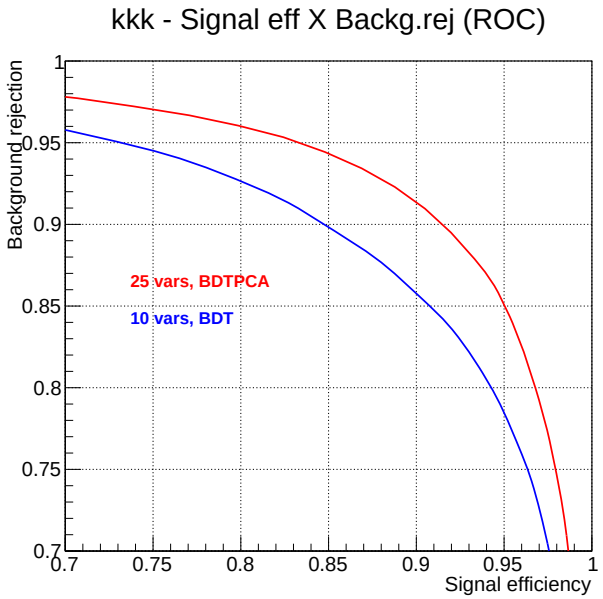


Performances with 25 and 10 variables (loose PID requirement)









Performances of optimizations with 25 and 10 variables

- From training/testing samples: Set with 10 variables gives performances somewhat inferior compared to the set with 25 variables
- However, when MVA cut is optimized combined, with tight PID requirements, we can easily recover the lost efficiency with a small penalty in significance

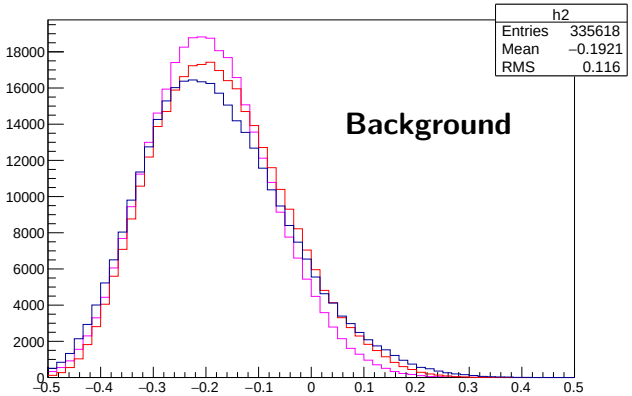
Strategy for MVA cut optimization:

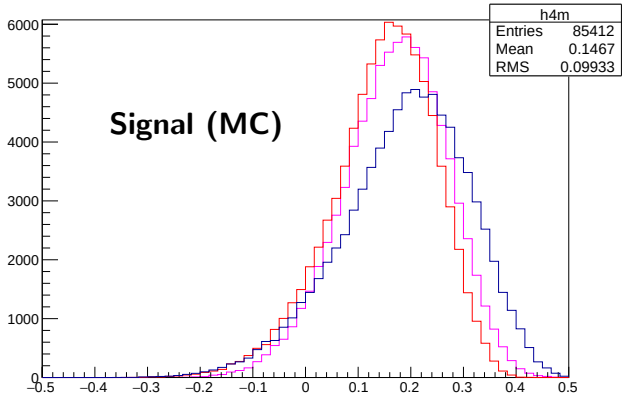
- S is taken from MC sample: S_{MC}
- $(S + B)$ is the number of events in the signal region $(abs(B\_m - 5284) < 40)$ in the data spectrum

$$Significance = \frac{S_{MC}}{\sqrt{(S+B)_{Data}}}$$

- Choose MVA cut as the value **close** to the maximum significance that gives good signal efficiency

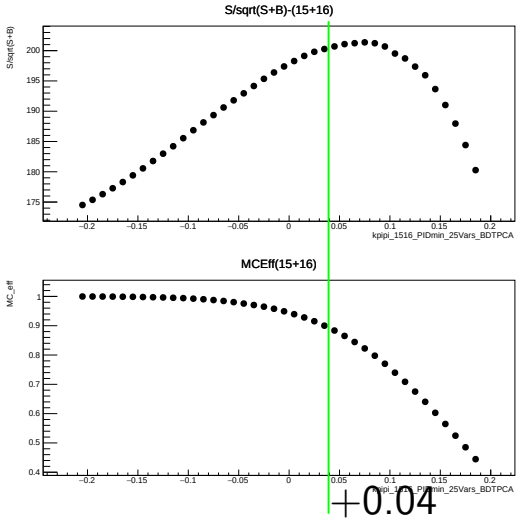
MVA cut optimization - $B \to K\pi\pi$ (With tight PID requirements)



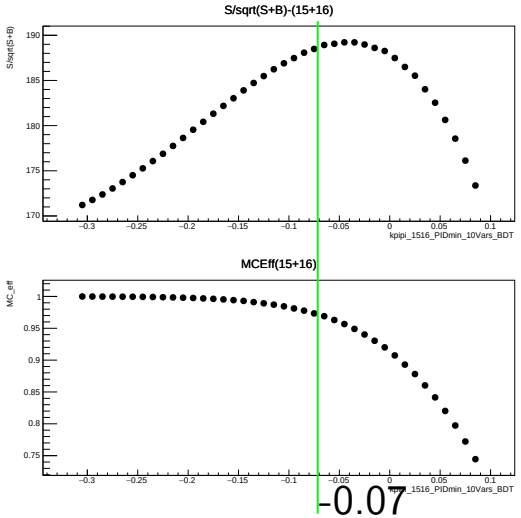


Self, 25 vars (BDTPCA)
Self, 10 vars (BDT)
Common*, 10 vars (BDT)

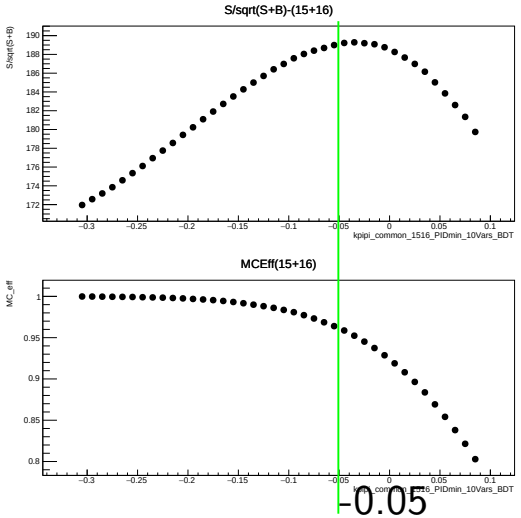
Self, 25 vars



Self, 10 vars



Common, 10 vars

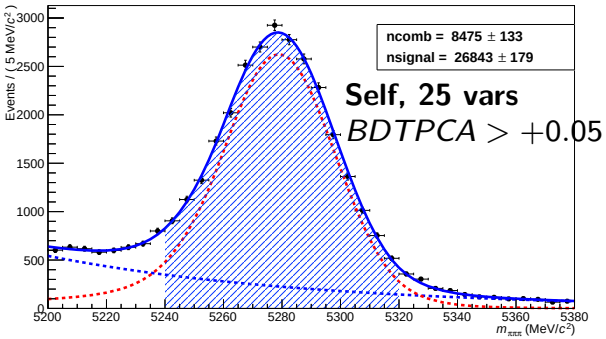


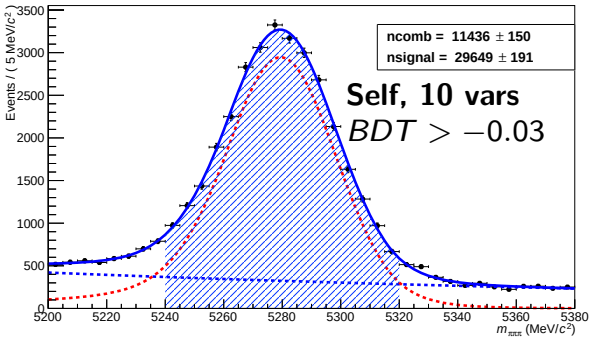
Cut is a compromise between high significance and high efficiency

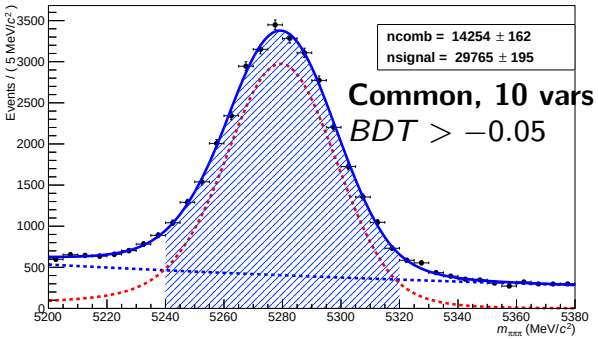
*Common: Training with joint MC from all channels and background from pipipi
Same MVA cut for all channels

MVA selected events invariant mass fits

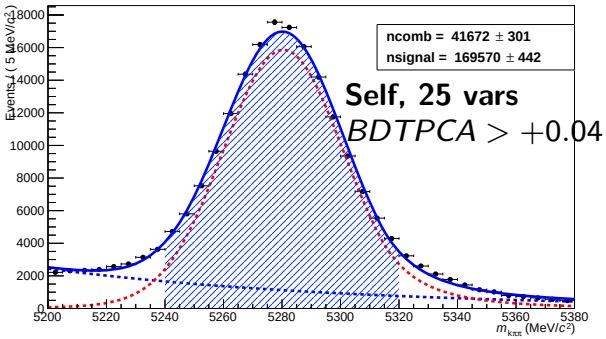
$B \to \pi\pi\pi$

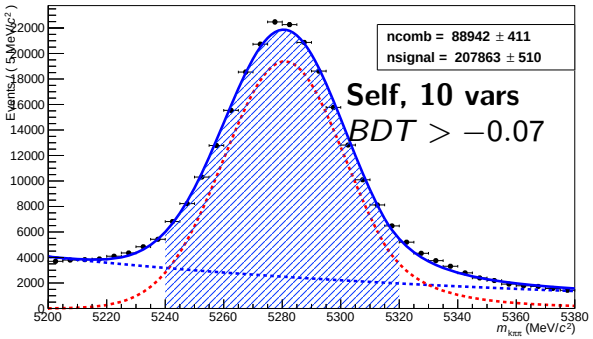


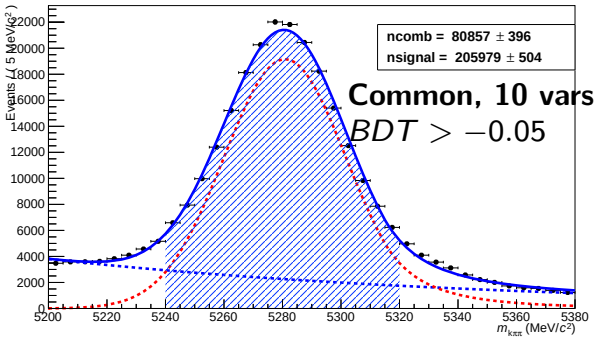




$B \to K\pi\pi$

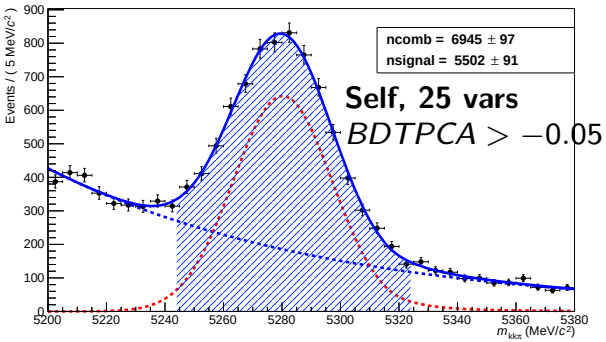


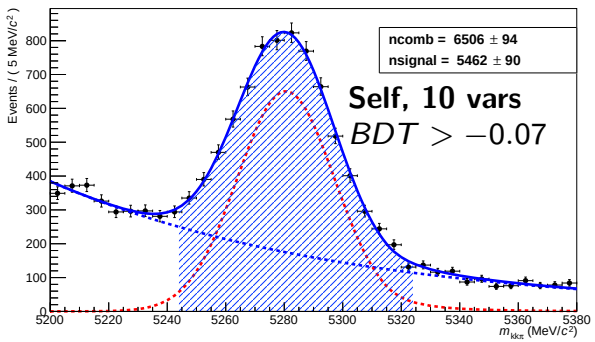


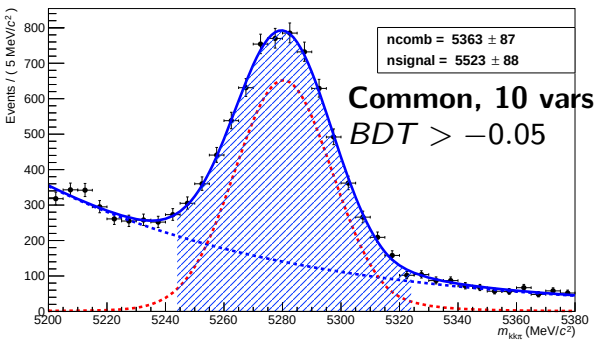


MVA selected events invariant mass fits

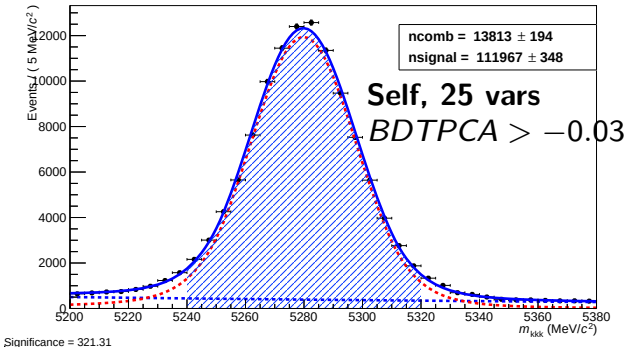
$B \to KK\pi$

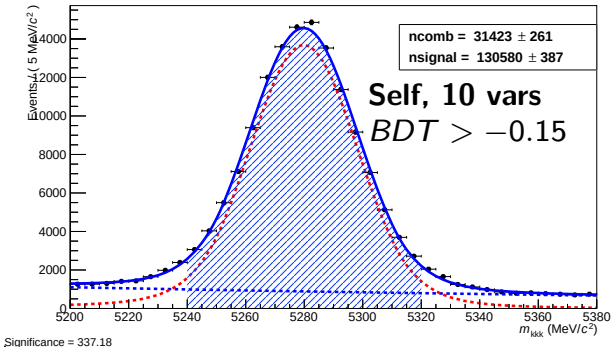


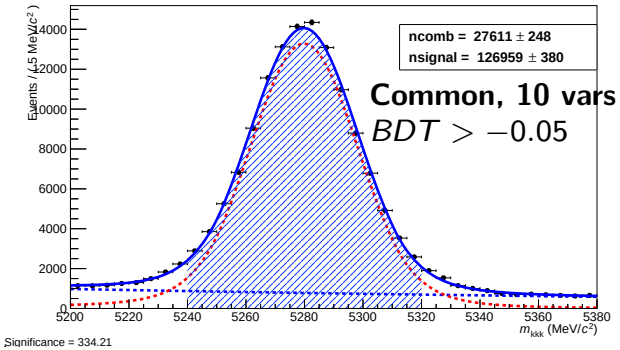




$B \to KKK$







Conclusions

Run II analysis of $B \to hhh$: MVA with 10 input variables

- Cleaner, simpler analysis
- Less correlations
- Better MC-data agreement

- MVA method: BDT *(Was 25 vars with BDT+PCA)*
- Final performances similar to the ones with 25 vars
  Both for the self or for the common optimizations
- Background levels are acceptable and well behaved