Development of a multivariate algorithm for the classification of B mesons at the LHCb experiment

Nico Guth

Bachelor talk, 20.07.2022

Arbeitsgruppe Albrecht
Fakultät Physik

Goal of my thesis:

Develop an algorithm that distinguishes between B_d^0 and B_s^0 mesons based on tracks associated with the signal B meson without tracks of the signal decay. (in pp-collisions at the LHCb detector)

Goal of my thesis:

Develop an algorithm that distinguishes between B_d^0 and B_s^0 mesons based on tracks associated with the signal B meson without tracks of the signal decay. (in pp-collisions at the LHCb detector)

Structure of this talk:

- Motivation
- B meson production in pp-collisions
- The LHCb detector
- Development of a B meson classifier
  - Identification of same side tracks using a BDT
  - Classification of the B meson using a DeepSet
  - Testing on real LHCb data
- Conclusion and Outlook

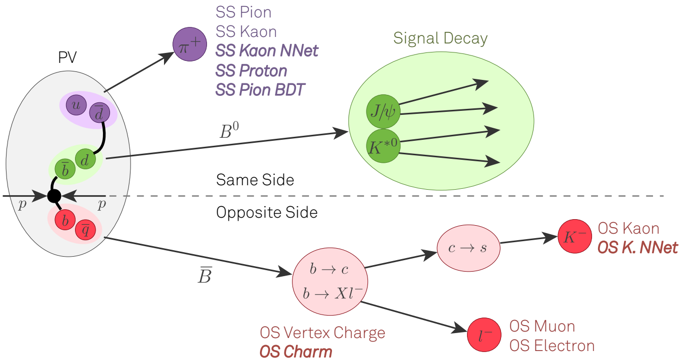
Motivation

- support background reduction where B_d^0 $(\bar{b}d)$ or B_s^0 $(\bar{b}s)$ is unwanted
  - *partial backgrounds* with missing information in the signal decay
  - backgrounds with similar signal kinematics
  - e.g. $B_s^0 \to D_s^+ K^-$ with B_d^0 backgrounds in the signal region

Motivation

- support background reduction where B_d^0 $(\bar{b}d)$ or B_s^0 $(\bar{b}s)$ is unwanted
  - *partial backgrounds* with missing information in the signal decay
  - backgrounds with similar signal kinematics
  - e.g. $B_s^0 \to D_s^+ K^-$ with B_d^0 backgrounds in the signal region
- excluding the signal decay
  $\to$ independence of the signal decay channel
- associated event contains enough information (in principle)
  - mass difference of B_d^0 and B_s^0 (87 MeV)
  - different fragmentation processes

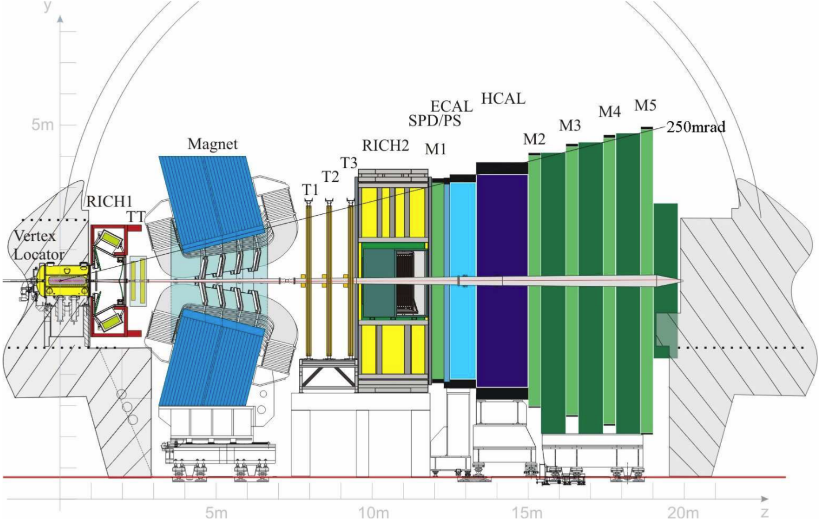
B meson production in pp-collisions



https://twiki.cern.ch/twiki/bin/view/LHCb/FlavourTaggingConferencePlots

- pp-collisions produce many particles
- gluon-fusion may lead to a $b\bar{b}$-pair
- hadronisation → B meson and fragmentation particles
- Lorentz boosted signal B → distinguish secondary from primary vertex
- for B_d^0 vs B_s^0 only same side (SS) relevant
- here: exclude the signal decay

The LHCb detector



https://iopscience.iop.org/article/10.1088/1748-0221/3/08/S08005

Development of a *B* meson classifier

Strategy:

- same side track identification using a BDT
- *B* meson classification using a DeepSet
- test on real LHCb data

Development of a *B* meson classifier

Strategy:

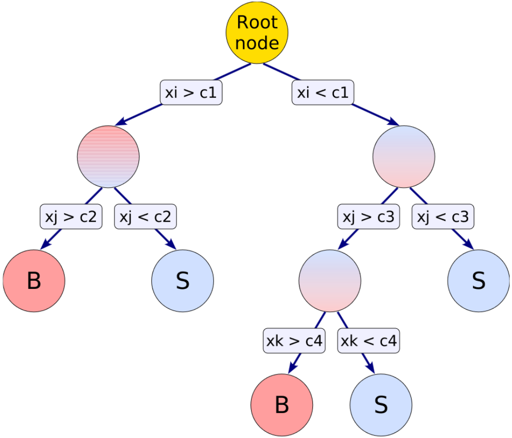
- same side track identification using a BDT
- *B* meson classification using a DeepSet
- test on real LHCb data

Training dataset:

- training with LHCb simulation
- combined dataset:
  - $B_d^0 \to J/\psi K^*$
  - $B_s^0 \to D_s^+ \pi^-$
- found differences by year and simulation version
  $\to$ chose 2016 and same simulation version
- dataset contains 0.4 million events and 18 million tracks

Boosted Decision Tree (BDT)

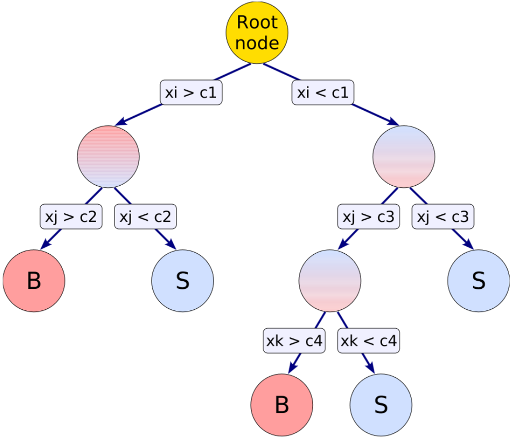
Simple Decision Tree:



https://arxiv.org/abs/physics/0703039

Boosted Decision Tree (BDT)

Simple Decision Tree:



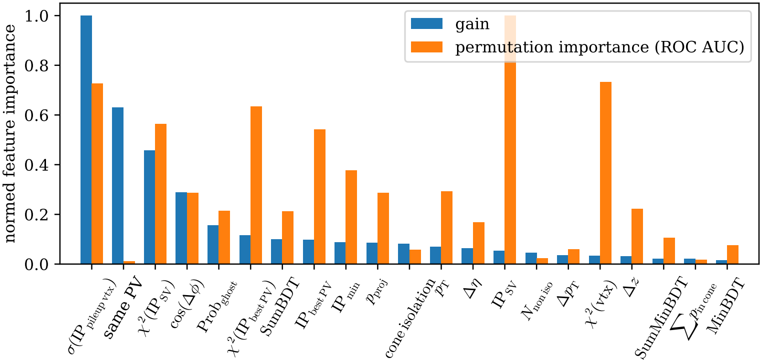
https://arxiv.org/abs/physics/0703039

Boosted Decision Tree:

- ensemble of multiple small Decision Trees
- weighted sum transformed with logistic function
  → estimated class probabilities
- iterative training through gradient boosting
  → minimum of a loss function

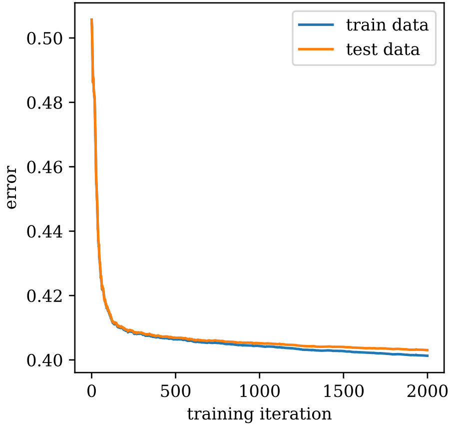
SS track identification: Feature Selection

track features	
p_T	IP_SV
p_proj	$\chi^2(\mathrm{IP}_\mathrm{SV})$
Δp_T	$\sigma(\mathrm{IP}_\mathrm{pileup\ vtx})$
Δz	$\mathrm{IP}_\mathrm{best\ PV}$
$\Delta\eta$	$\chi^2(\mathrm{IP}_\mathrm{best\ PV})$
$\cos(\Delta\phi)$	IP_min
$\mathrm{Prob}_\mathrm{ghost}$	same PV
$\chi^2(\mathrm{vtx})$	cone isolation
SumBDT	$N_\mathrm{non\ iso}$
MinBDT	$\sum p_\mathrm{in\ cone}$
SumMinBDT	



SS track identification: BDT training and results

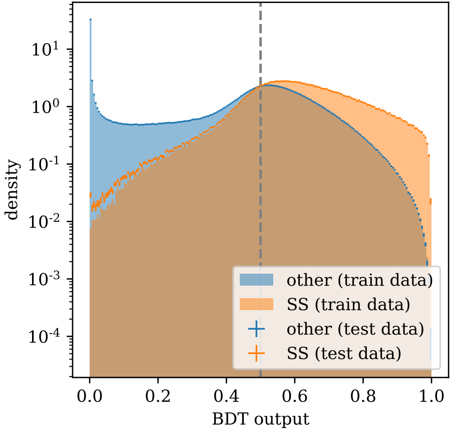
Error rate during training



- 60% training data, 40% test data
- 2000 decision trees with maximum tree depth of 4
- loss: logistic regression for binary classification
- output: $\mathrm{Prob}_{SS} \in [0, 1]$

SS track identification: BDT training and results

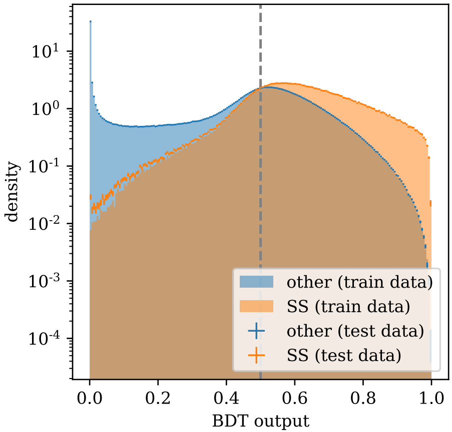
Distribution of Prob_{SS}



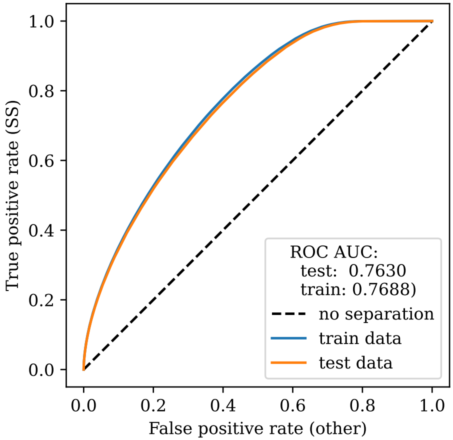
- 60% training data, 40% test data
- 2000 decision trees with maximum tree depth of 4
- loss: logistic regression for binary classification
- output: $\mathrm{Prob}_{SS} \in [0, 1]$

SS track identification: BDT training and results

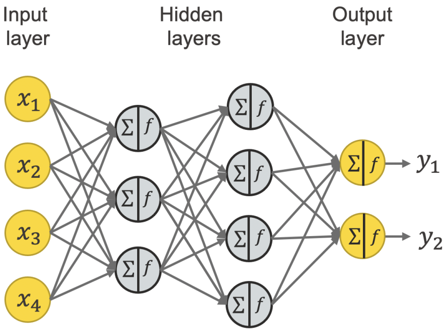
Distribution of Prob_{SS}



ROC curve of the BDT predictions



Neural Network (NN)

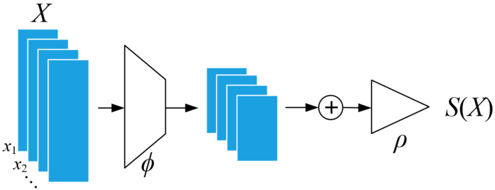


https://www.knime.com/blog/a-friendly-introduction-to-deep-neural-networks

- non-linear transformation $\vec{x} \to \vec{y}$
- multiple steps called layers of activation
  $\to \vec{a}^{(n)} = f^{(n)}\left(W^{(n)} \cdot \vec{a}^{(n-1)} + \vec{b}^{n}\right)$
- activation functions used here:
  - $f_{\text{ReLU}}(z) = \max(0, z)$
  - $f_{\text{Sigmoid}}(z) = \frac{1}{1+e^{-z}}$
- iterative training through backpropagation (gradient descent)

DeepSet

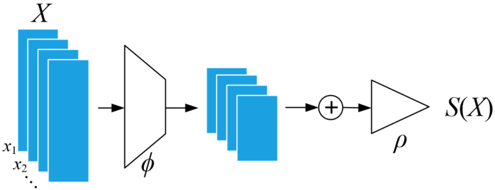
- extension of NNs to allow inputs of sets of vectors
  → variable input length
  → permutation invariant
- $f(X) = \rho\left(\sum_{x_i \in X} \phi(x_i)\right)$



https://arxiv.org/abs/1703.06114

DeepSet

- extension of NNs to allow inputs of sets of vectors
  → variable input length
  → permutation invariant
- $f(X) = \rho\left(\sum_{x_i \in X} \phi(x_i)\right)$



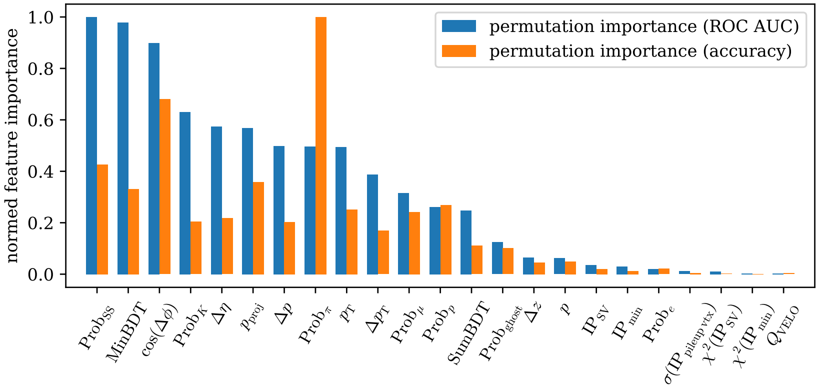
https://arxiv.org/abs/1703.06114

DeepSet for B meson classification:

- one set X per event
- one vector x_i per track
- ϕ-network layer sizes: 23, 64, 128, 64
- ρ-network layer sizes: 64, 128, 64, 1
- f_{ReLU} for hidden layers
- f_{Sigmoid} for the output layer
- output: $\text{Prob}_{B_s} \in [0, 1]$

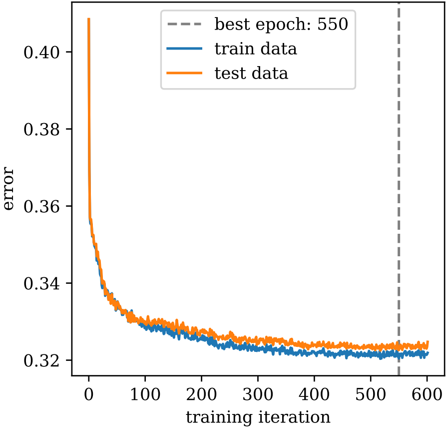
B meson classification: Feature Selection

track features	
p	**$\mathrm{Prob}_{\mathrm{SS}}$**
p_{T}	Prob_{e}
p_{proj}	$\mathrm{Prob}_{\mathrm{ghost}}$
Δp	Prob_{K}
Δp_{T}	Prob_{μ}
Δz	Prob_{p}
$\cos(\Delta\phi)$	Prob_{π}
$\Delta\eta$	$\sigma(\mathrm{IP}_{\mathrm{pileup\ vtx}})$
$\mathrm{IP}_{\mathrm{SV}}$	Q_{VELO}
$\chi^2(\mathrm{IP}_{\mathrm{SV}})$	SumBDT
$\mathrm{IP}_{\mathrm{min}}$	MinBDT
$\chi^2(\mathrm{IP}_{\mathrm{min}})$	



B meson classification: DeepSet training and results

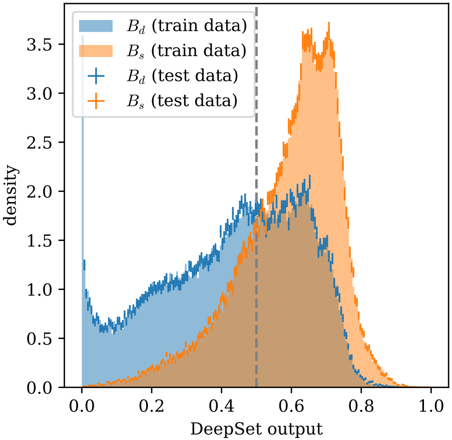
Error rate during training



- 60% training data, 40% test data (standard scaled)
- regularisation:
  - early stopping after 50 iterations
  - Dropout of 50%
- loss: binary cross entropy
- optimizer: Adam
- output: $\mathrm{Prob}_{B_s} \in [0, 1]$

B meson classification: DeepSet training and results

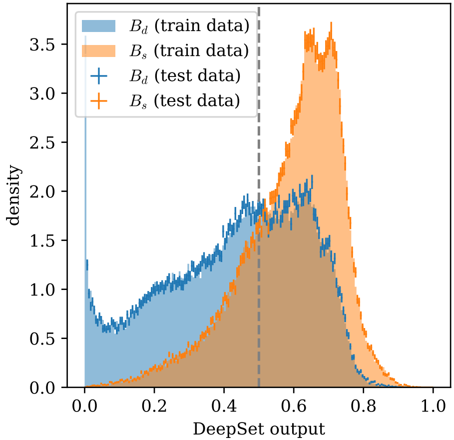
Distribution of Prob_{B_s}



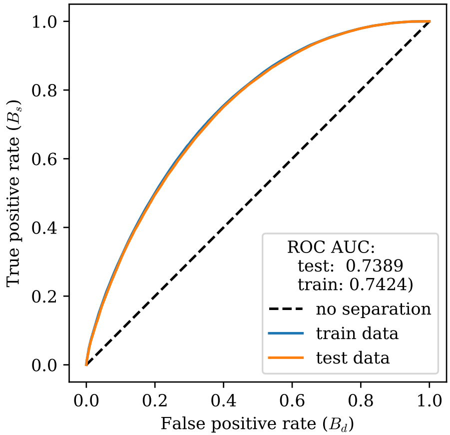
- 60% training data, 40% test data (standard scaled)
- regularisation:
  - early stopping after 50 iterations
  - Dropout of 50%
- loss: binary cross entropy
- optimizer: Adam
- output: $\mathrm{Prob}_{B_s} \in [0, 1]$

B meson classification: DeepSet training and results

Distribution of Prob$_{B_s}$



ROC curve of the DeepSet predictions



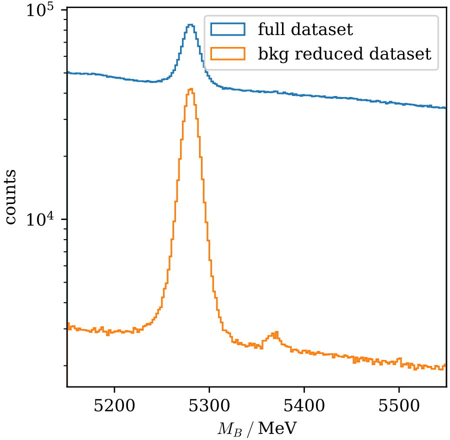
Testing on LHCb data: Overview

- run 2 LHCb data selected for B_d^0 or $B_s^0 \to J/\psi\, K_S^0$
- based on an ongoing analysis

Testing on LHCb data: Overview

- run 2 LHCb data selected for B_d^0 or $B_s^0 \to J/\psi\, K_S^0$
- based on an ongoing analysis
- visible B_s^0 peak after background reduction:
  - trained BDT with 13 features on $B_d^0 \to J/\psi K_S^0$ simulation as signal and upper mass sideband ($\geq 5450\,\mathrm{MeV}$) as combinatorial background
  - manual cuts for Λ^0 and K^* background that got misidentified as K_S^0

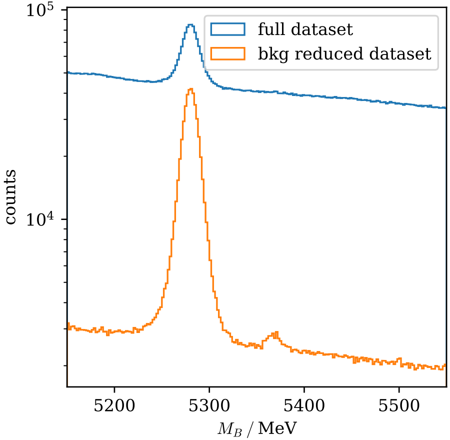
Signal B mass after background reduction
(peaks at $M(B_d) = 5280\,\mathrm{MeV}$ and $M(B_s) = 5367\,\mathrm{MeV}$)



Testing on LHCb data: Overview

- run 2 LHCb data selected for B_d^0 or $B_s^0 \rightarrow J/\psi\, K_S^0$
- based on an ongoing analysis
- visible B_s^0 peak after background reduction
- testing strategy:
  - apply the developed algorithm $\rightarrow$ Prob_{B_s} for every event
  - estimate counts of B_d^0 and B_s^0 events by fitting the mass distribution and integrating the B_d^0 and B_s^0 components
  - scan through the Prob_{B_s} distribution

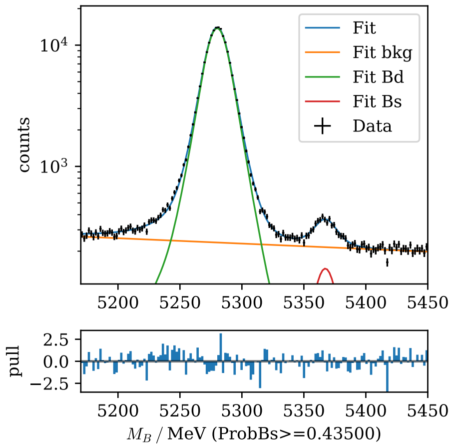
Signal B mass after background reduction (peaks at $M(B_d)$ = 5280 MeV and $M(B_s)$ = 5367 MeV)



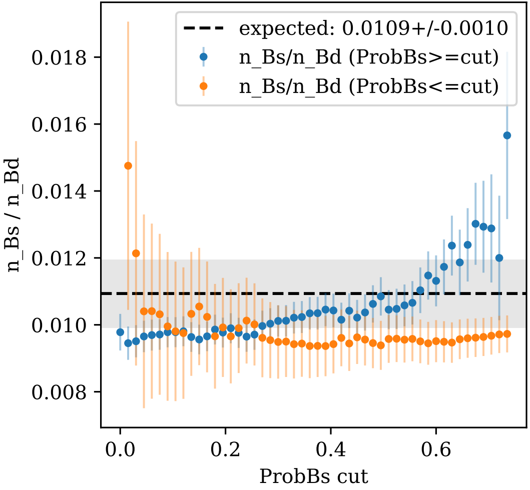
Testing on LHCb data: Overview

- run 2 LHCb data selected for B_d^0 or $B_s^0 \to J/\psi\, K_S^0$
- based on an ongoing analysis
- visible B_s^0 peak after background reduction
- testing strategy:
  - apply the developed algorithm $\to \mathrm{Prob}_{B_s}$ for every event
  - estimate counts of B_d^0 and B_s^0 events by fitting the mass distribution and integrating the B_d^0 and B_s^0 components
  - scan through the Prob_{B_s} distribution

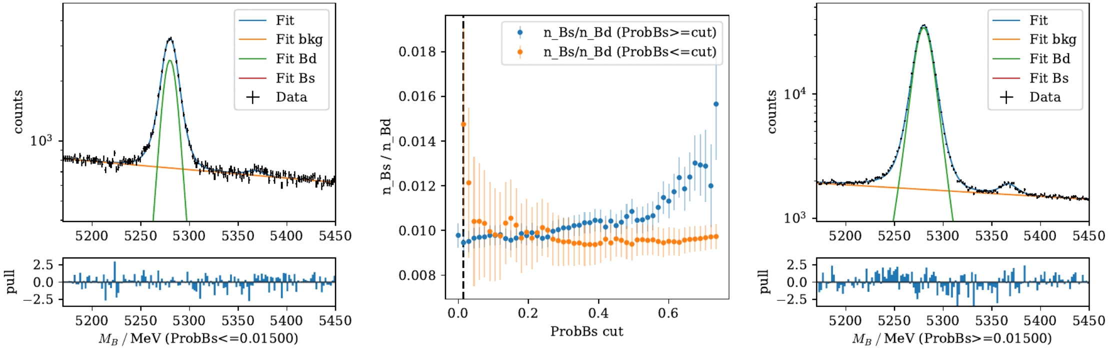
Example fit of the mass distribution



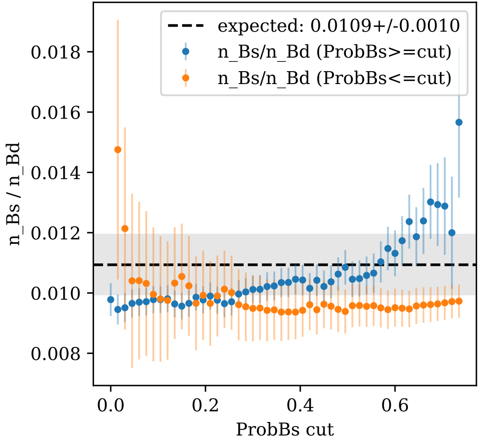
Testing on LHCb data: Results (ratio n_{B_s}/n_{B_d} by Prob_{B_s} cut value)



Testing on LHCb data: Animation of n_{B_s}/n_{B_d} and the corresponding fits



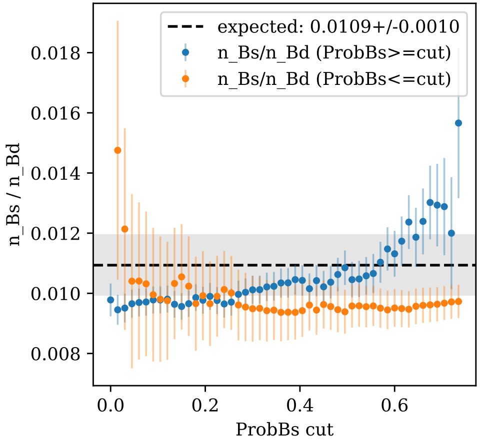
Testing on LHCb data: Results (ratio n_{B_s}/n_{B_d} by Prob$_{B_s}$ cut value)



- without separation: constant ratio n_{B_s}/n_{B_d}
- expected value (with perfect selection efficiencies):

$$\frac{\mathrm{BR}(B_s \to J/\psi\, K_S^0)}{\mathrm{BR}(B_d \to J/\psi\, K_S^0)} \cdot f_s/f_d(13\,\mathrm{TeV}) = 0.0109 \pm 0.0010$$

Testing on LHCb data: Results (ratio n_{B_s}/n_{B_d} by Prob_{B_s} cut value)

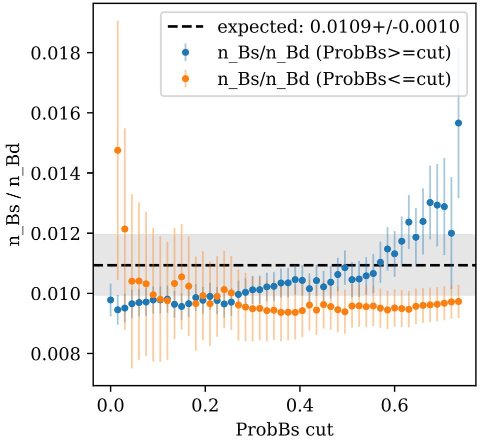


- without separation: constant ratio n_{B_s}/n_{B_d}
- expected value (with perfect selection efficiencies):

$$\frac{\mathrm{BR}(B_s \to J/\psi\, K_S^0)}{\mathrm{BR}(B_d \to J/\psi\, K_S^0)} \cdot f_s/f_d(13\,\mathrm{TeV}) = 0.0109 \pm 0.0010$$

- $\mathrm{Prob}_{B_s} \leq x$: mostly constant, no clear B_s^0 peak for low x

Testing on LHCb data: Results (ratio n_{B_s}/n_{B_d} by Prob_{B_s} cut value)

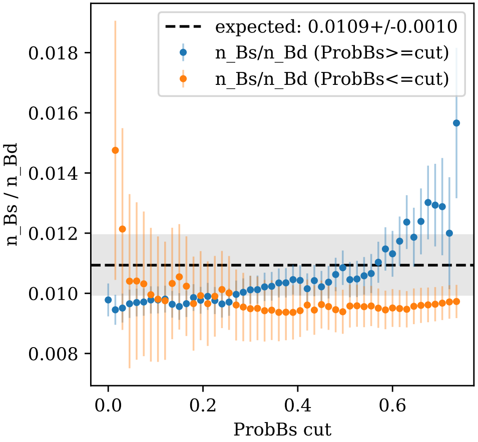


- without separation: constant ratio n_{B_s}/n_{B_d}
- expected value (with perfect selection efficiencies):

$$\frac{\mathrm{BR}(B_s \to J/\psi\, K_S^0)}{\mathrm{BR}(B_d \to J/\psi\, K_S^0)} \cdot f_s/f_d(13\,\mathrm{TeV}) = 0.0109 \pm 0.0010$$

- $\text{Prob}_{B_s} \leq x$: mostly constant, no clear B_s^0 peak for low x
- $\text{Prob}_{B_s} \geq x$: starts constant, then increases

Testing on LHCb data: Results (ratio n_{B_s}/n_{B_d} by Prob_{B_s} cut value)



- without separation: constant ratio n_{B_s}/n_{B_d}
- expected value (with perfect selection efficiencies):

$$\frac{\mathrm{BR}(B_s \to J/\psi K_S^0)}{\mathrm{BR}(B_d \to J/\psi K_S^0)} \cdot f_s/f_d(13\,\mathrm{TeV}) = 0.0109 \pm 0.0010$$

- $\mathrm{Prob}_{B_s} \leq x$: mostly constant, no clear B_s^0 peak for low x
- $\mathrm{Prob}_{B_s} \geq x$: starts constant, then increases
- clearly achieved some separation between B_d^0 and B_s^0

Conclusion and outlook

Results:

- on simulation:
  - BDT can identify SS tracks (ROC AUC: 0.76) and helps the DeepSet (feature importances)
  - DeepSet achieves a clear separation of B_d^0 and B_s^0 events (ROC AUC: 0.74)
- on LHCb data: prove of concept shown
- reasons for incomplete performance portability unknown:
  - selection differences in training dataset? (combination of $B_d^0 \to J/\psi K^*$ and $B_s^0 \to D_s^+ \pi^-$)
  - mismodeled simulation features?

Conclusion and outlook

Results:

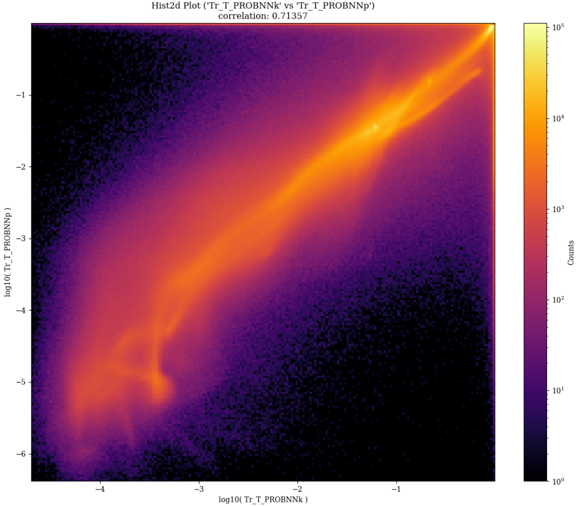
- on simulation:
  - BDT can identify SS tracks (ROC AUC: 0.76) and helps the DeepSet (feature importances)
  - DeepSet achieves a clear separation of B_d^0 and B_s^0 events (ROC AUC: 0.74)
- on LHCb data: prove of concept shown
- reasons for incomplete performance portability unknown:
  - selection differences in training dataset? (combination of $B_d^0 \to J/\psi K^*$ and $B_s^0 \to D_s^+ \pi^-$)
  - mismodeled simulation features?

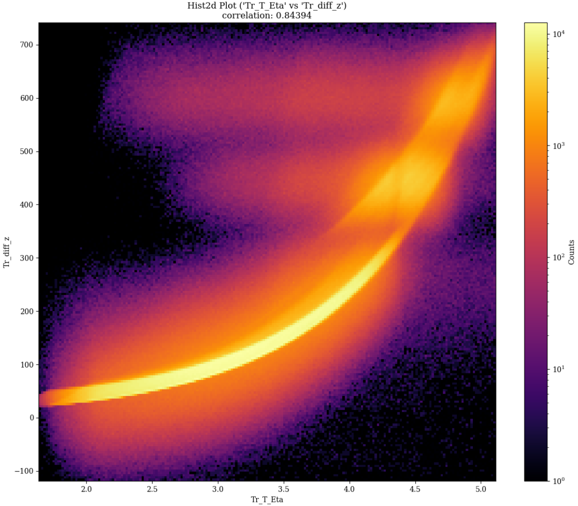
Outlook and suggestions:

- feature validation: compare simulation and data
- ensure that kinematic differences originate only from the mass difference:
  - training dataset with the same final-state particles for both B mesons
  - reweighting the training data to equalize kinematics
- possible extension to include other b hadrons ($B^\pm$, $B_c^\pm$, Λ_b^0, ...)

Thank you for your attention!

Here is some art I found in the data (2D histograms):

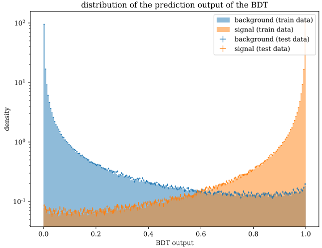


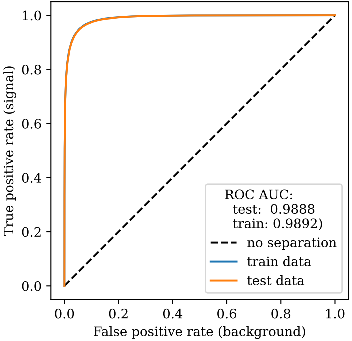


Backup

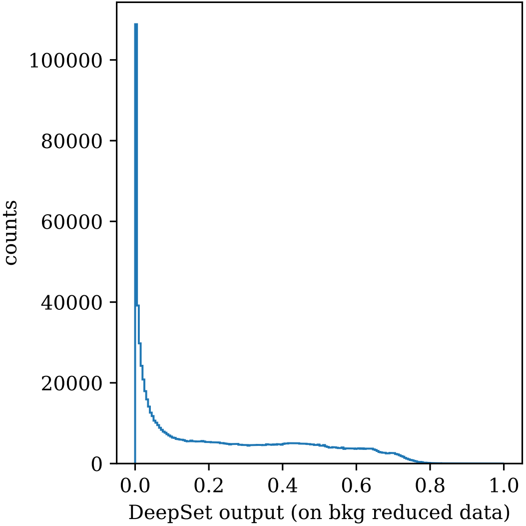
Background BDT

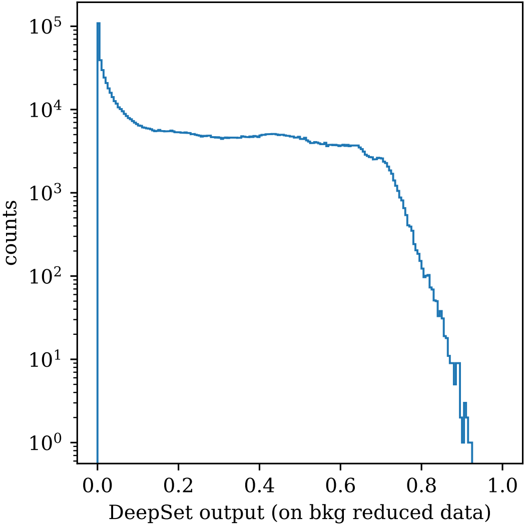
signal features	
$\mathrm{IP}(B^0)$	$p_\mathrm{T}(\pi^+)$
$\mathrm{IP}(J/\psi)$	$p_\mathrm{T}(\pi^-)$
$\mathrm{IP}(K_S^0)$	$p_\mathrm{T}(K_S^0)$
$\mathrm{IP}(\mu^+)$	$\eta(B^0)$
$\mathrm{IP}(\mu^-)$	$\eta(K_S^0)$
$\mathrm{FD}(K_S^0)$	$p_z(K_S^0)$
$\chi^2(\mathrm{fit})$	



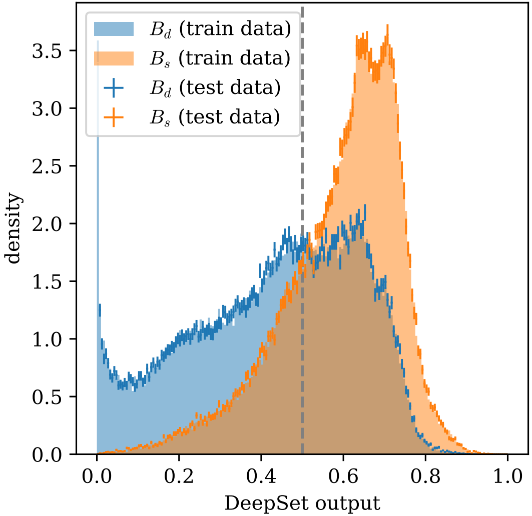


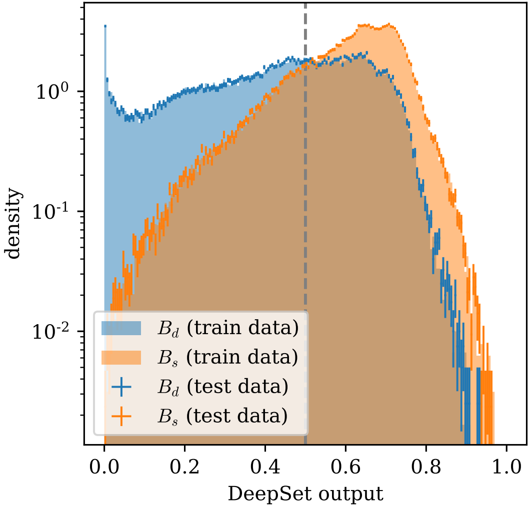
Test on LHCb data: DeepSet output



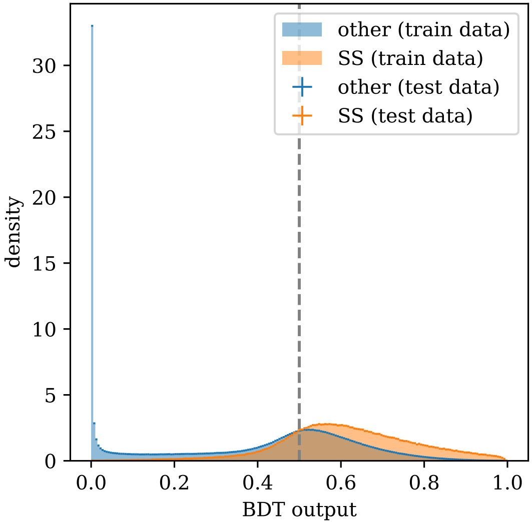


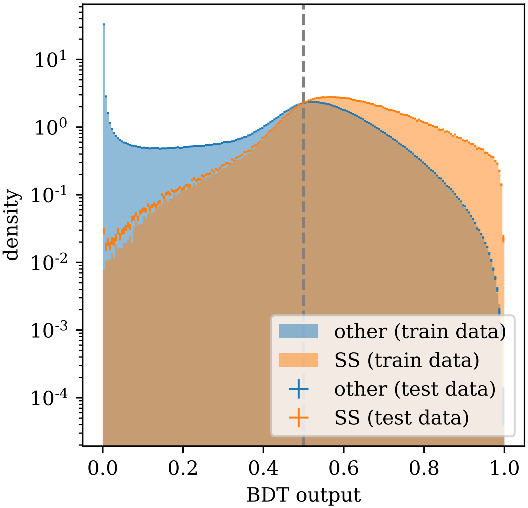
B meson classification: DeepSet output



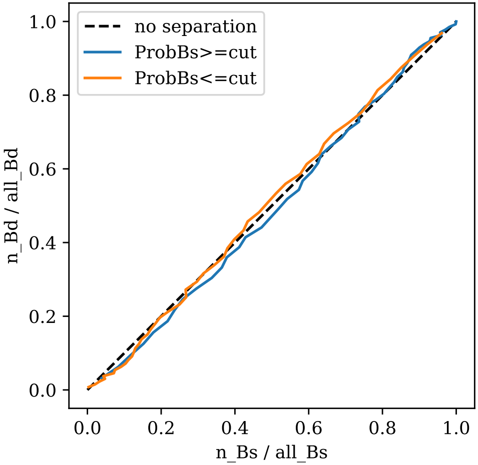


SS track identification: BDT output





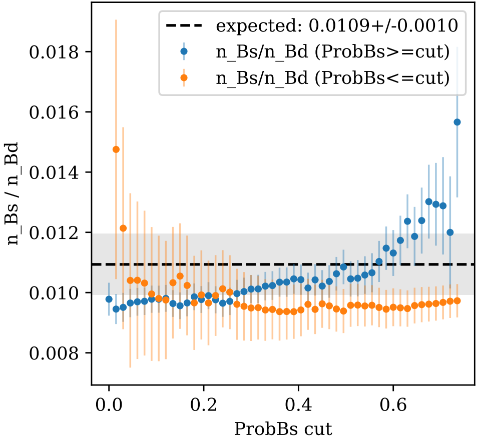
Testing on LHCb data: Results (efficiencies, similar to a ROC curve)



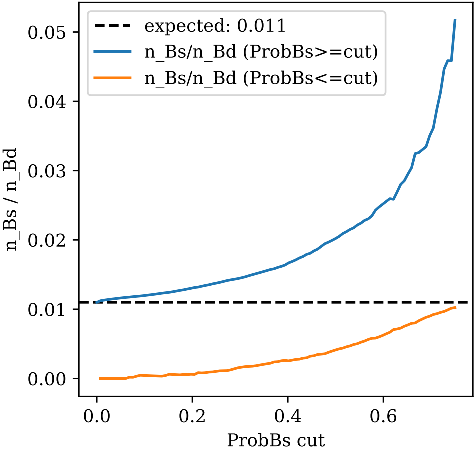
- calculated efficiencies $\varepsilon_B = n_B(x)/n_B(\text{no cut})$
- plot ε_{B_d} against ε_{B_s}
- should be similar to a ROC curve
- separation not really visible

Testing on LHCb data: Results (ratio n_{B_s}/n_{B_d} by Prob_{B_s} cut value)

Achieved separation on data

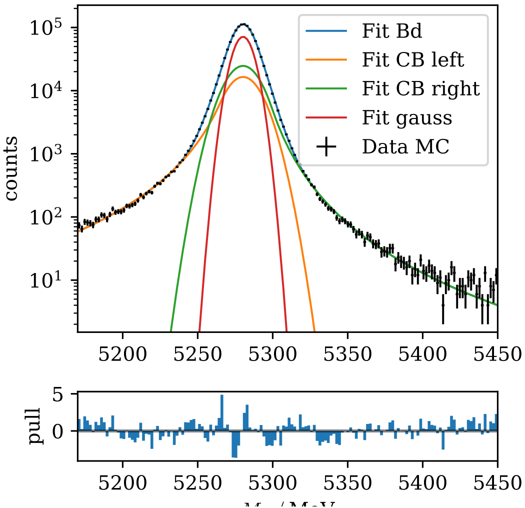


Achieved separation on simulation

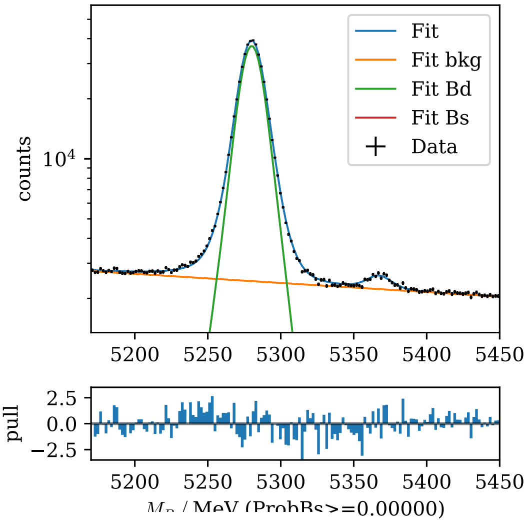


Testing on LHCb data: Fits

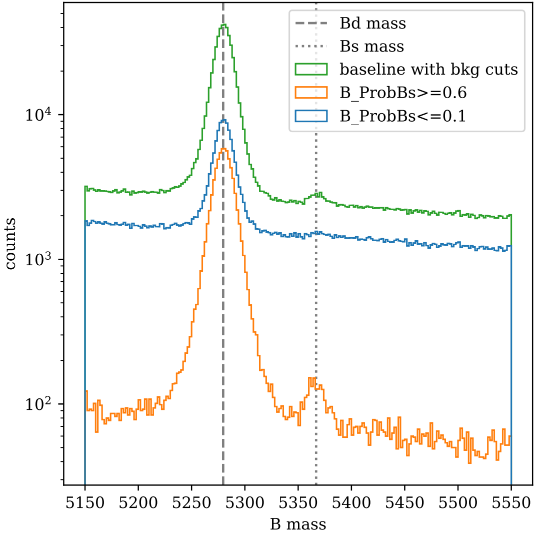
Fit of B_d^0 mode on simulation



Fit without ProbBs selection



Data cut comparison:



Fit functions

$$F(M_B) = N_{\text{bkg}} \cdot F_{\text{bkg}}(M_B) + N_{B_d} \cdot F_{B_d}(M_B) + N_{B_s} \cdot F_{B_s}(M_B)$$

$$F_{\text{bkg}}(M_B) = \exp(-\lambda \cdot M_B)\,.$$

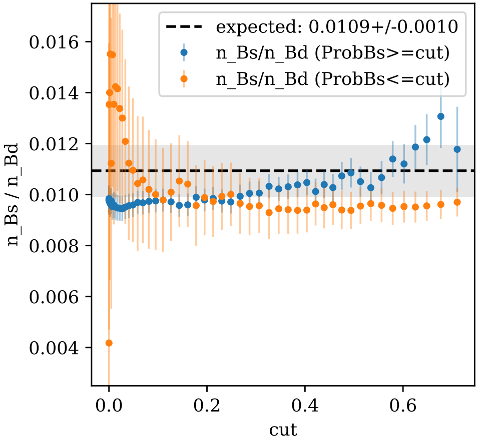
$$\begin{aligned} F_B(M_B) = & f_1 \cdot f_2 \cdot F_{\text{CB}}\left(\frac{M_B - \mu}{\sigma_1}, \beta_1, m_1\right) \\ & + (1 - f_1) \cdot f_2 \cdot F_{\text{CB}}\left(-\frac{M_B - \mu}{\sigma_2}, \beta_2, m_2\right) \\ & + (1 - f_1) \cdot (1 - f_2) \cdot F_{\text{gauss}}\left(M_B, \mu, \sigma_3\right)\,, \end{aligned}$$

$$F_{\text{CB}}(x, \beta, m) = \begin{cases} N \cdot \exp(-\frac{x^2}{2}) & \text{for } x > -\beta \\ N \cdot \left(\frac{m}{|\beta|}\right)^m \cdot \exp\left(-\frac{\beta^2}{2}\right) \cdot \left(\frac{m}{|b|} - |b| - x\right)^{-m} & \text{for } x \leq -\beta \end{cases}$$

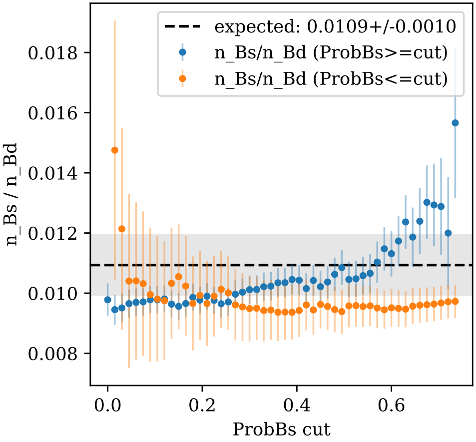
$$F_{\text{gauss}}(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma}} \cdot \exp\left(-\frac{1}{2}\left(\frac{x - \mu}{\sigma}\right)^2\right)$$

Ratio plot in the thesis and the newest plot (different cut values and slightly different results due to fit instabilities at the edges)

In the thesis

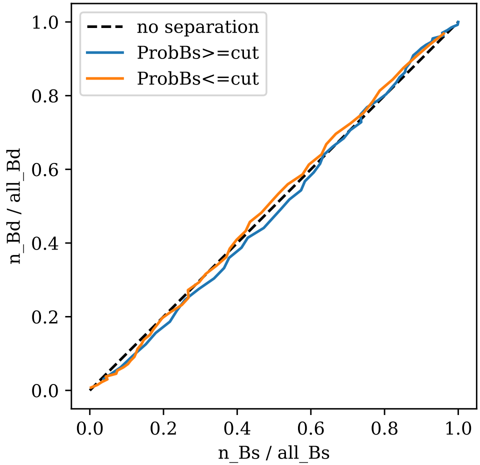


In the presentation

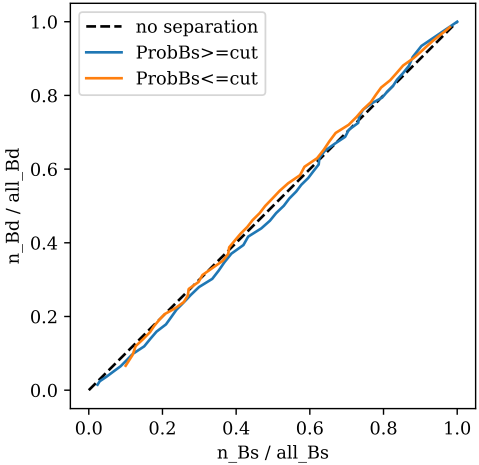


ROC plot in the thesis and the newest plot

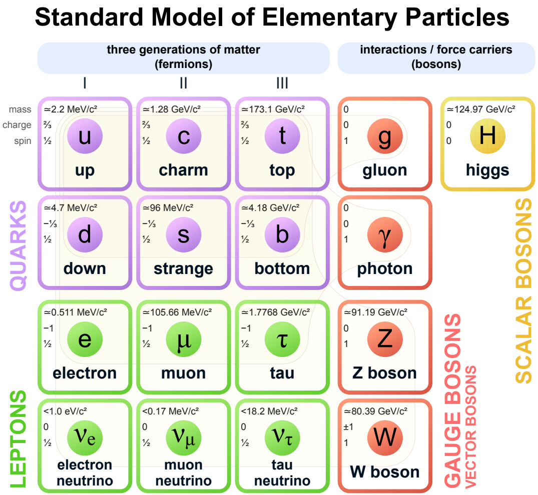
In the thesis



In the presentation

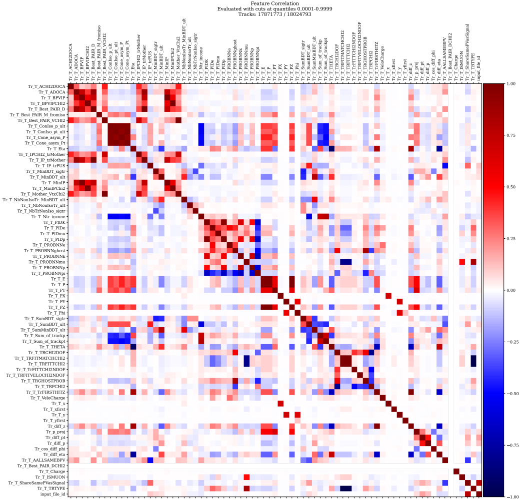


The Standard Model of particle physics

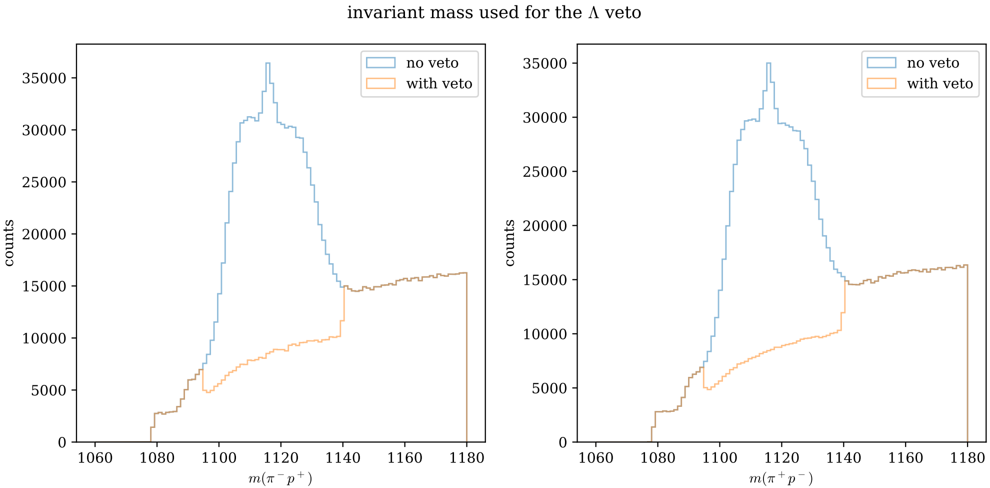


https://en.wikipedia.org/wiki/Standard_Model

Correlation Matrix



Lambda Veto


invariant mass used for the Λ veto
no veto
with veto
counts
35000
30000
25000
20000
15000
10000
5000
0
1060
1080
1100
1120
1140
1160
1180
$m(\pi^- p^+)$
$m(\pi^+ p^-)$