News

in contact with Ryo about differences in LCFIPlus performance

optimization of input variables

- leave-one out
  - a lot of possbilities

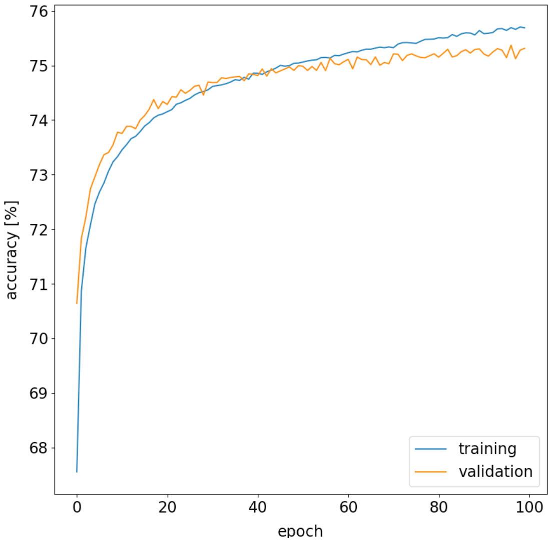
➡ taylor expansion of NN output w.r.t. input features
  - gives idea of relevance of the input feature

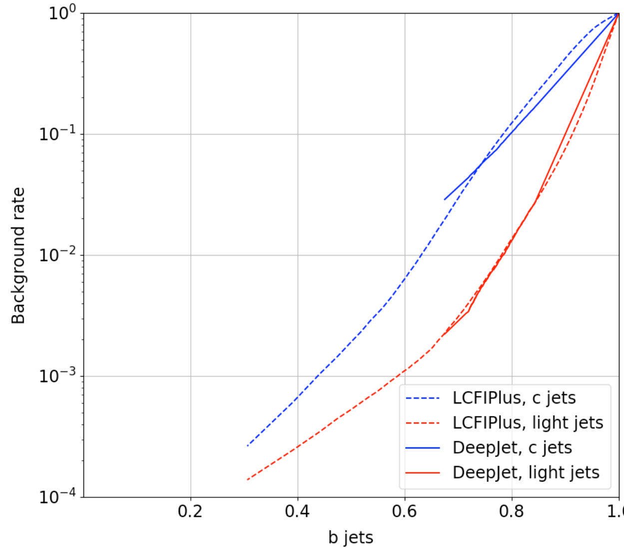
➡ 1D shape comparison plots

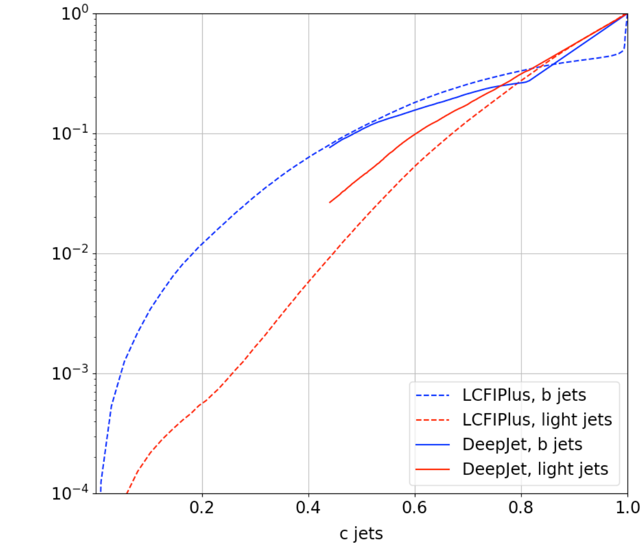
➡ 2D plots to study correlations

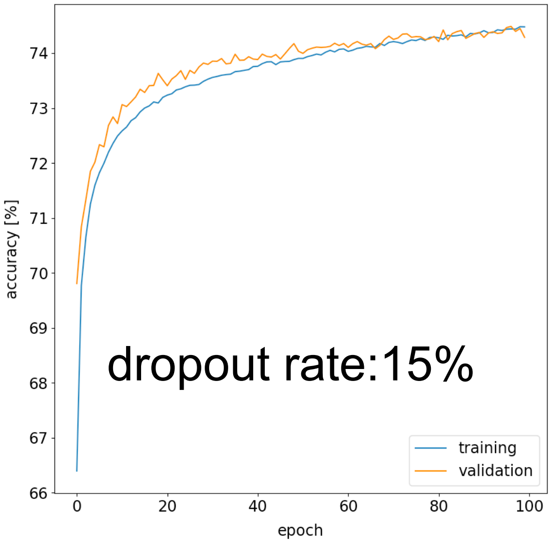
- PCA
  - best results when using 7 jet variables, 20 variables for charged constituents, 6 variables for neutral, 9 for secondary vertices (7, 23 , 6, 16)

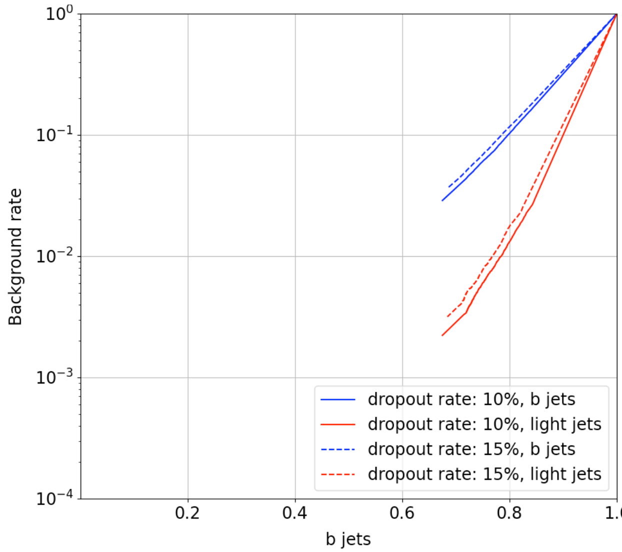
Baseline

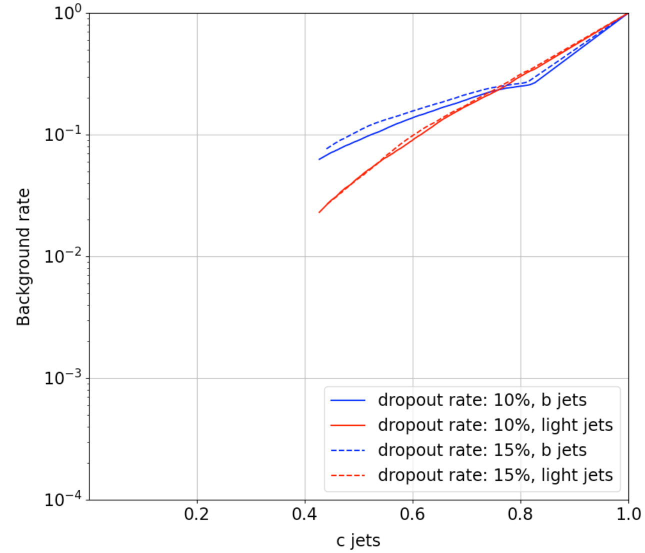

accuracy [%]
epoch
training
validation


Background rate
b jets
LCFIPlus, c jets
LCFIPlus, light jets
DeepJet, c jets
DeepJet, light jets


c jets
LCFIPlus, b jets
LCFIPlus, light jets
DeepJet, b jets
DeepJet, light jets


accuracy [%]
epoch
dropout rate:15%
training
validation


Background rate
b jets
dropout rate: 10%, b jets
dropout rate: 10%, light jets
dropout rate: 15%, b jets
dropout rate: 15%, light jets


Background rate
c jets
dropout rate: 10%, b jets
dropout rate: 10%, light jets
dropout rate: 15%, b jets
dropout rate: 15%, light jets

Relevance propagation: idea

S.Wunsch, R. Friese, R.Wolf, and G. Quast, Computing and Software for Big Science 2 (Sep, 2018) 5, doi:10.1007/s41781-018-0012-, arXiv:1803.08782

- relate output space of NN to input space

➡ identify characteristics of input space that have large influence on output for a given task

- **decompose NN function into Taylor expansion in each element of the input space**
- Taylor coefficients contain information about the sensitivity of the NN response to the inputs
- dependence on phase space: mean of absolute values of Taylor coefficients evaluated for all elements of test sample

set of input variables, evaluated for element k of test sample

$$\langle t_i \rangle \equiv \frac{1}{N} \sum_{k=1}^{N} \left| t_i\left(\{x_j\}\big|_k \right) \right| \qquad i \in \mathcal{P}(\{x_j\})$$

sum over test sample of size N

i^{th} Taylor coefficient

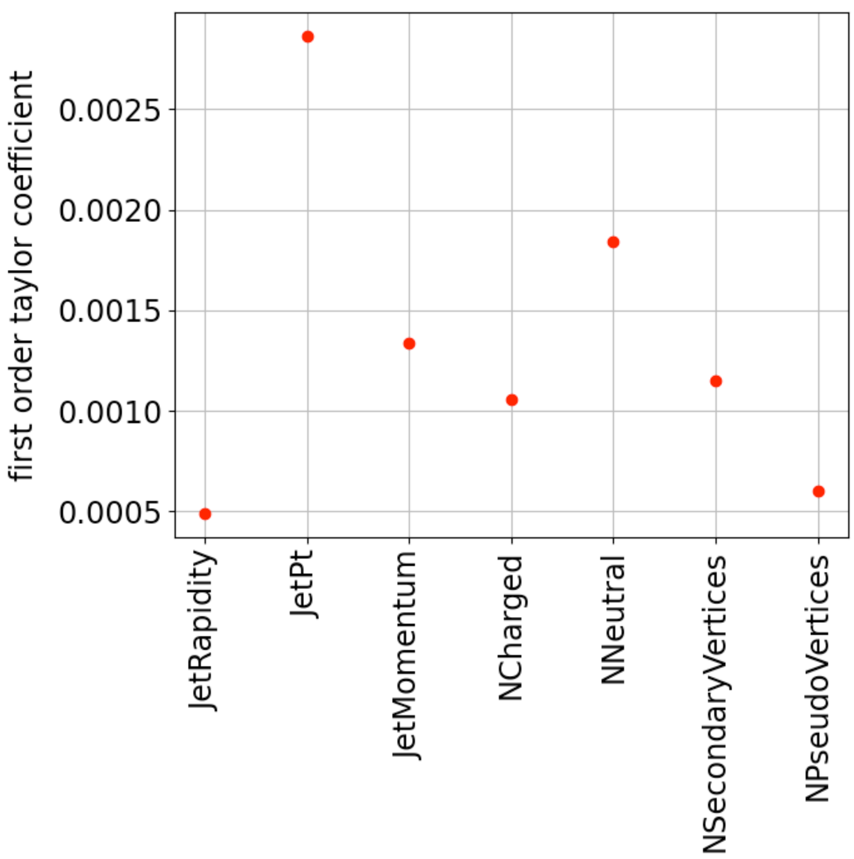
- **first order** Taylor coefficient: **influence of single input elements**
- **second order** Taylor coefficent: **influence of pair-wise or auto-correlations**

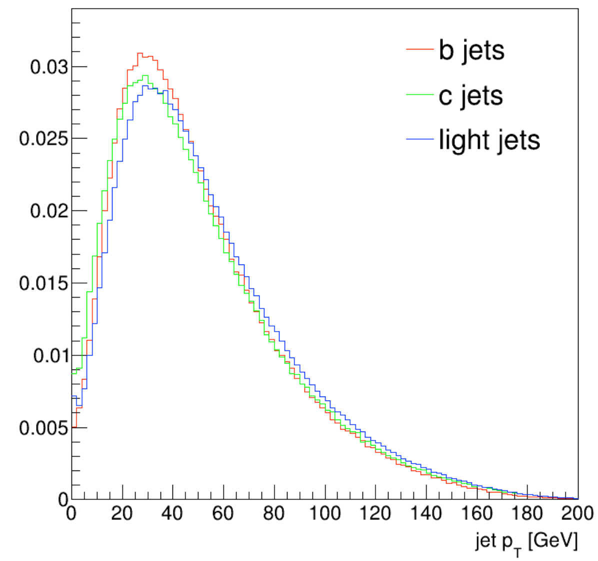
Relevance propagation: idea

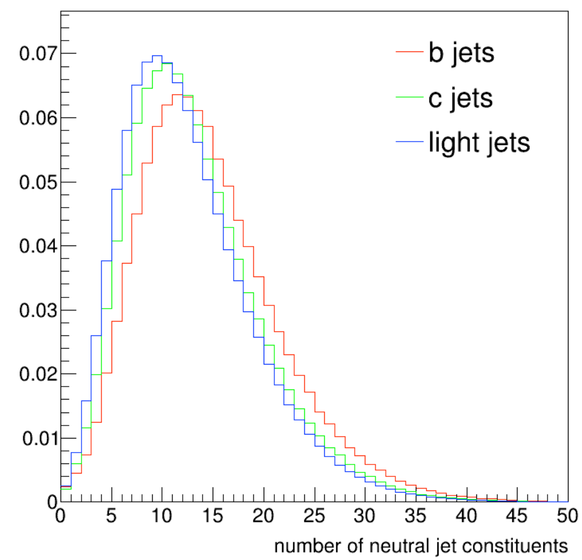
- **we can learn which variables were important for the NN to make the classification**
- presents only the view of the trained NN on information in training data, training might be not optimal, there might be more information in the training data that was not picked up

- can be used as guideline which variables can be tested to be taken out

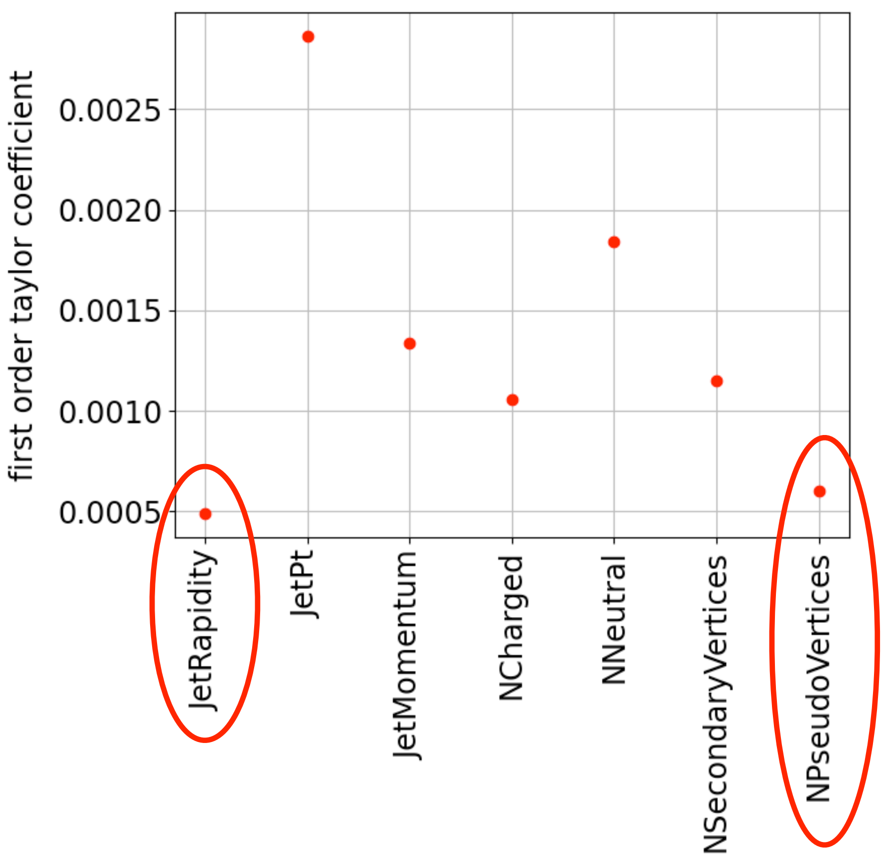
First order taylor coefficients

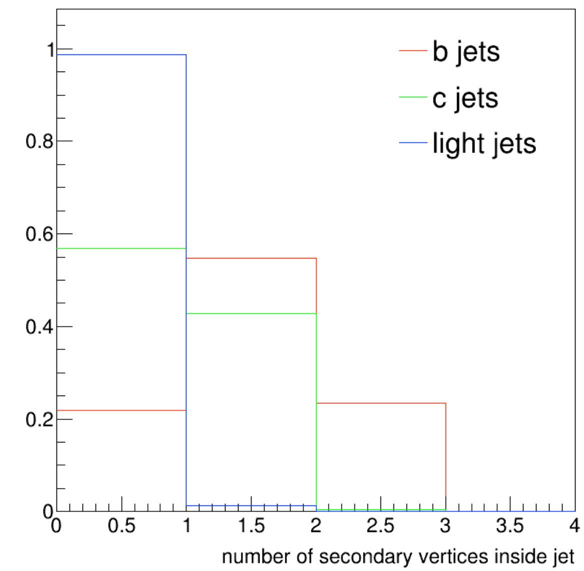


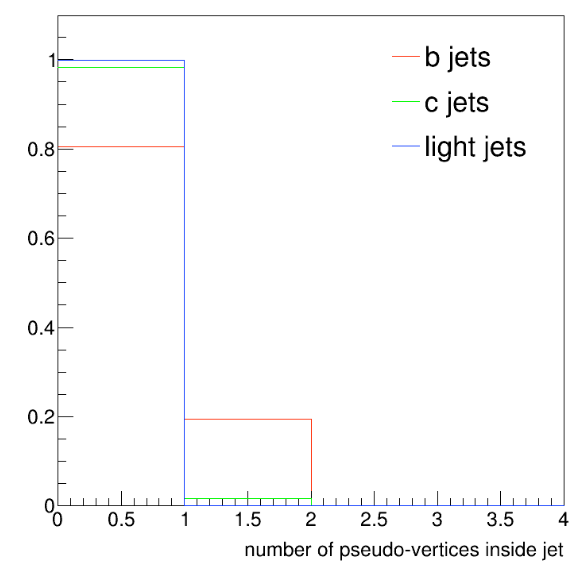




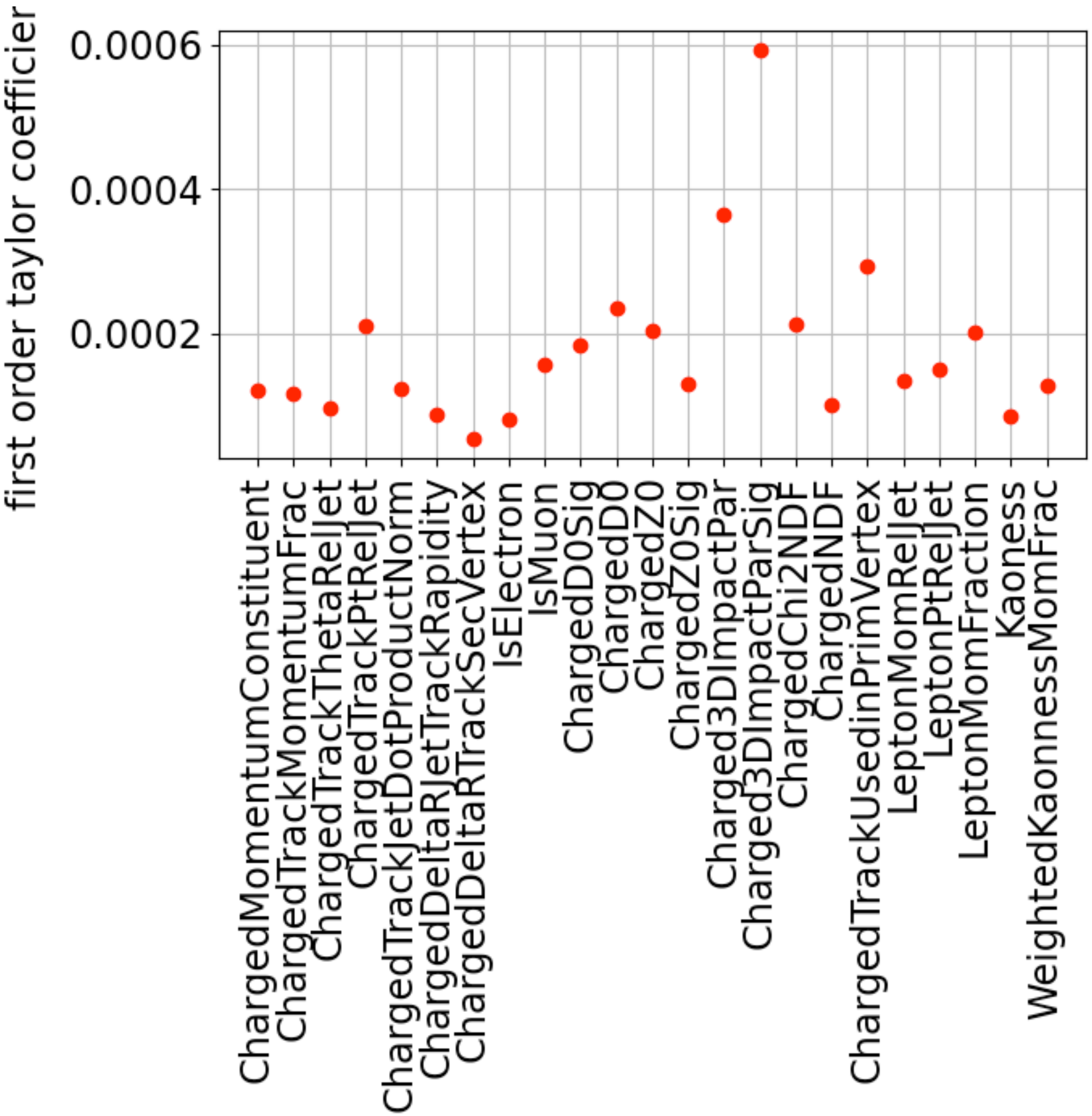
First order taylor coefficients

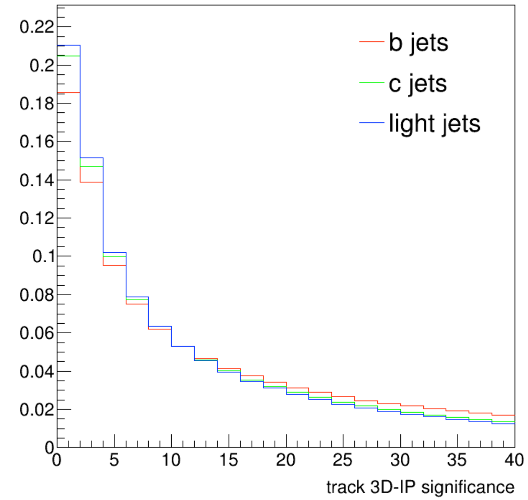


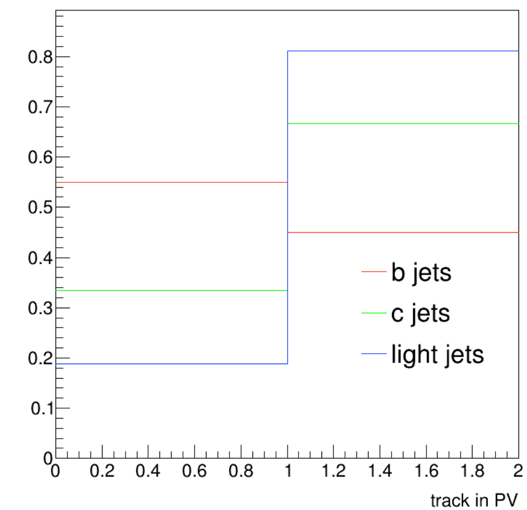




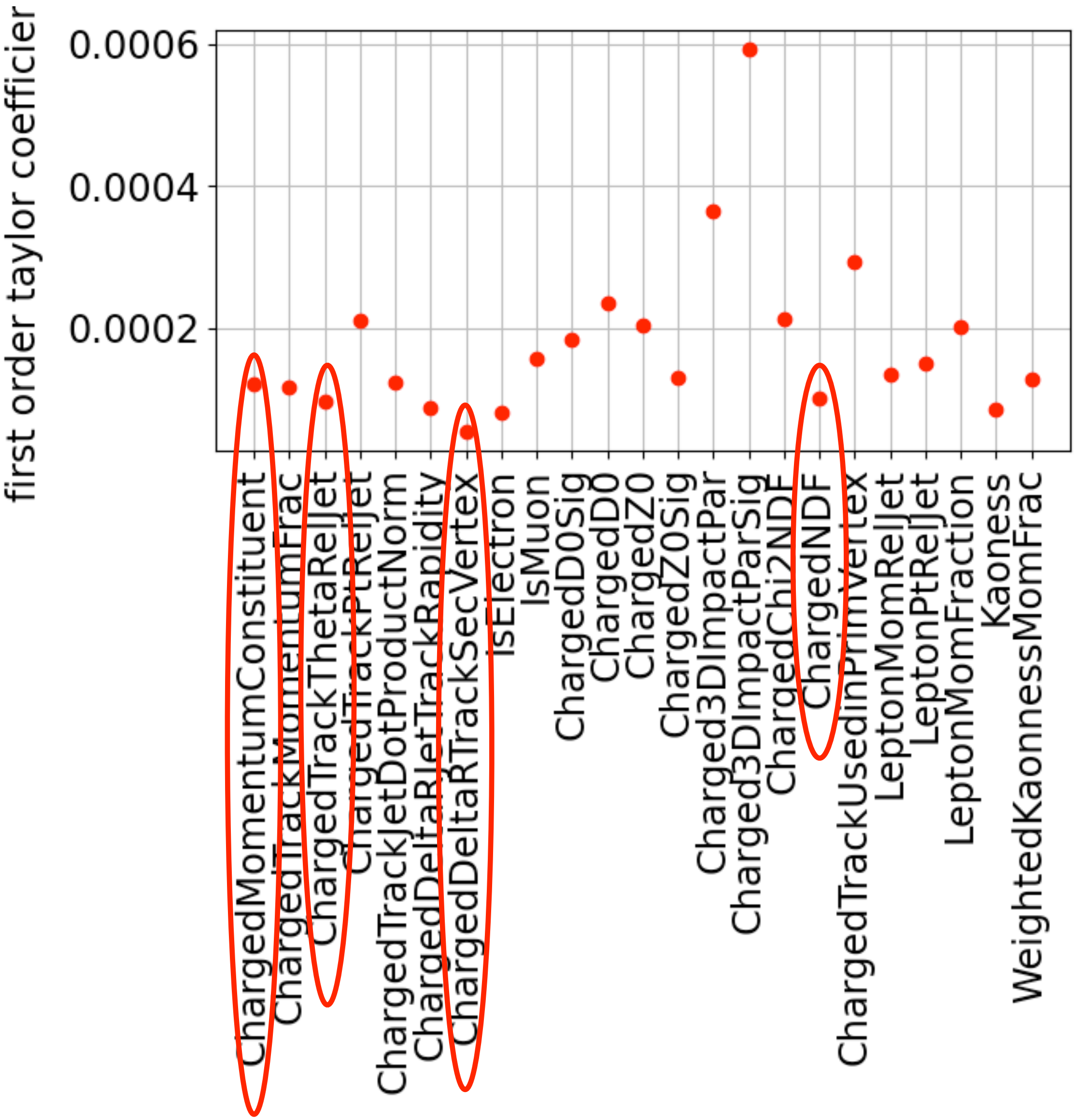
First order taylor coefficients

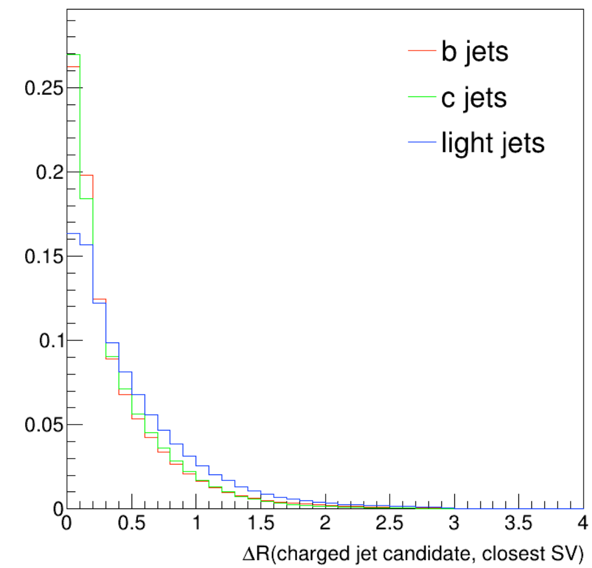


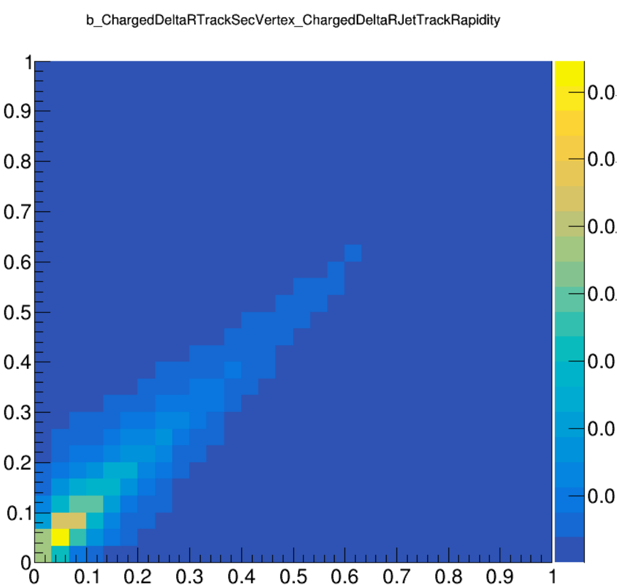




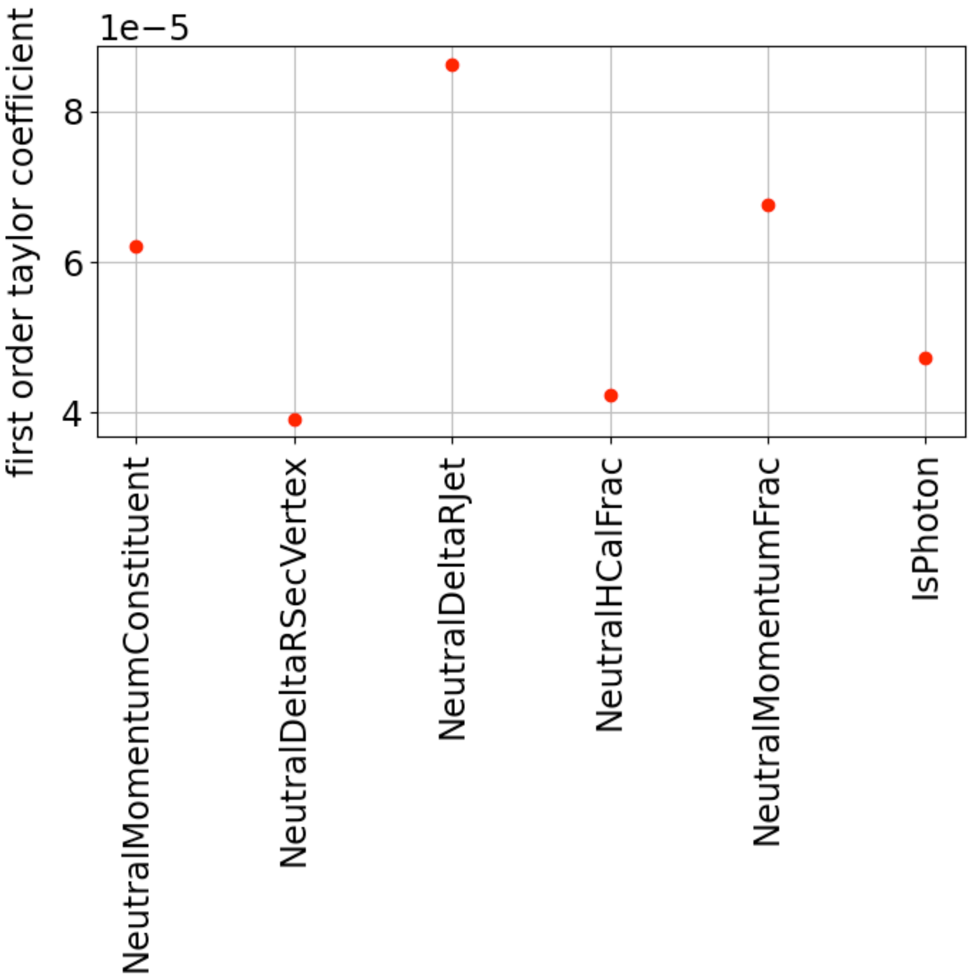
First order taylor coefficients

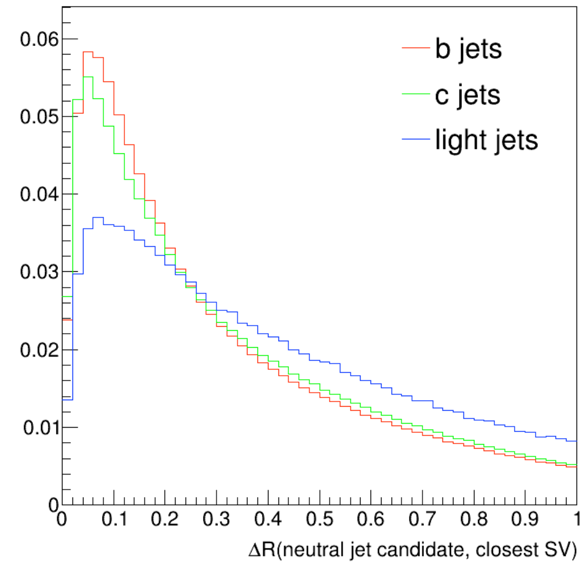


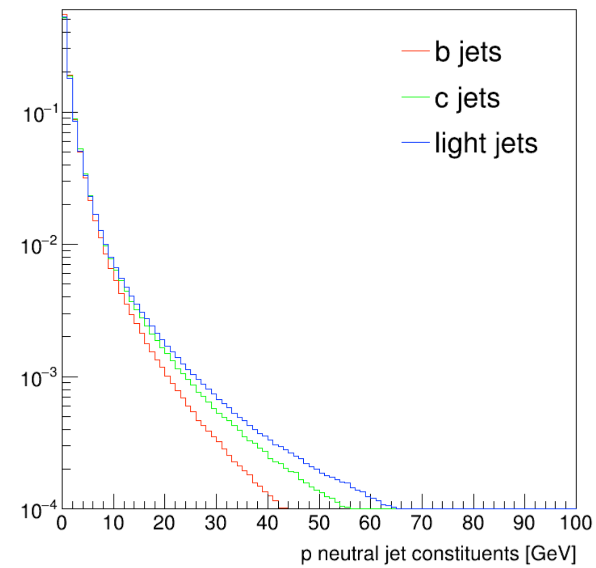




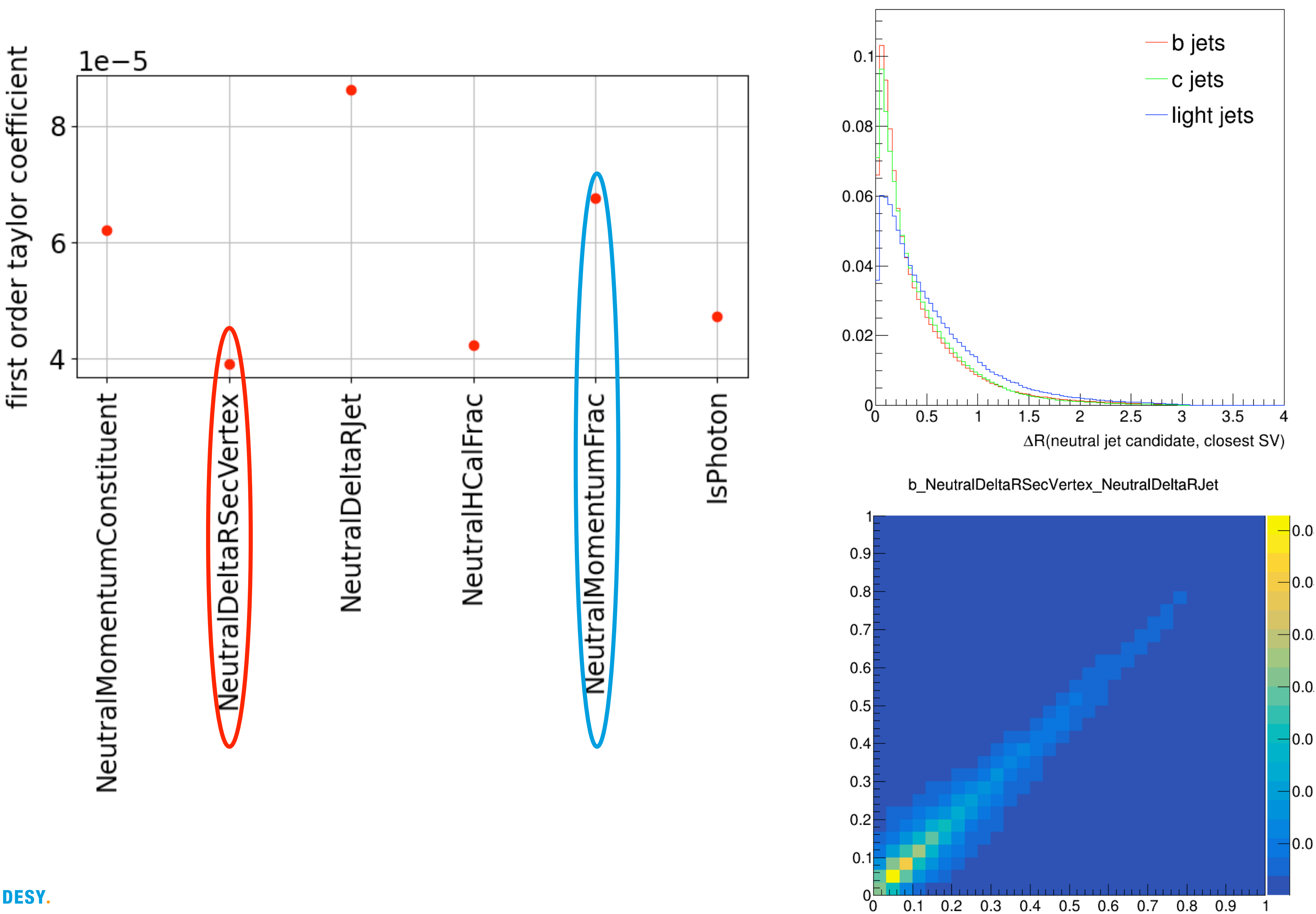
First order taylor coefficients



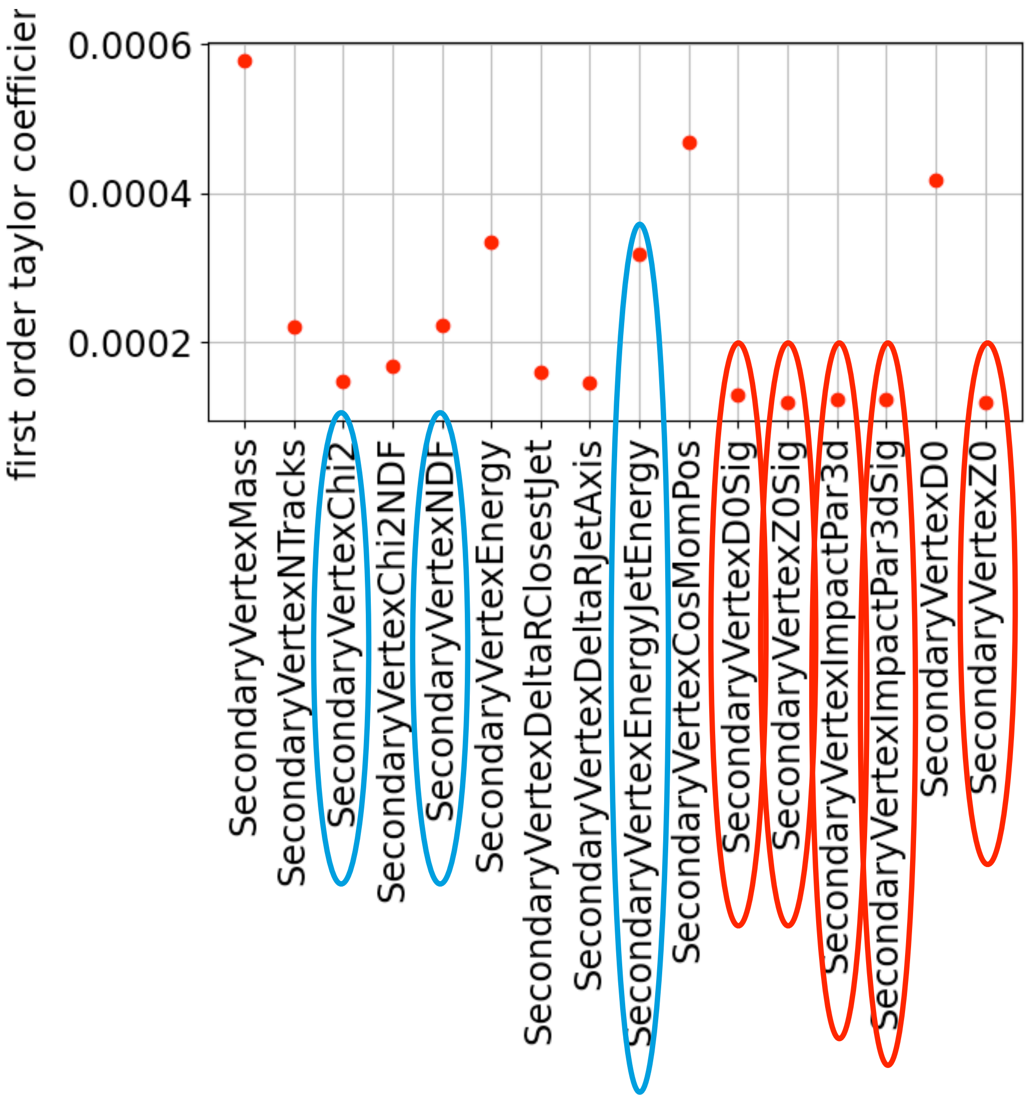


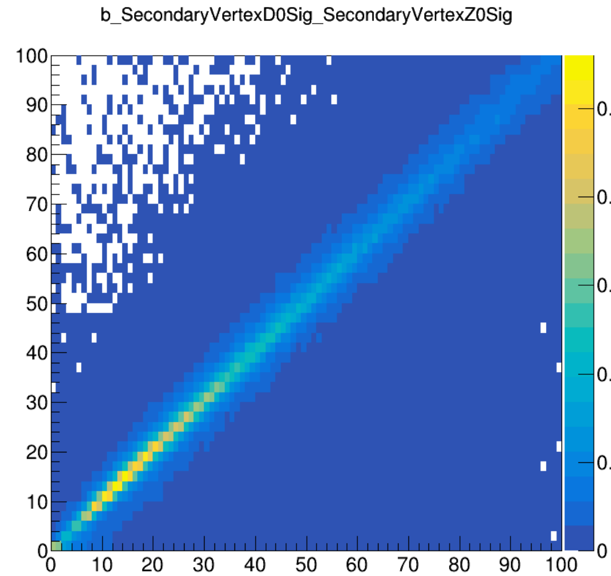


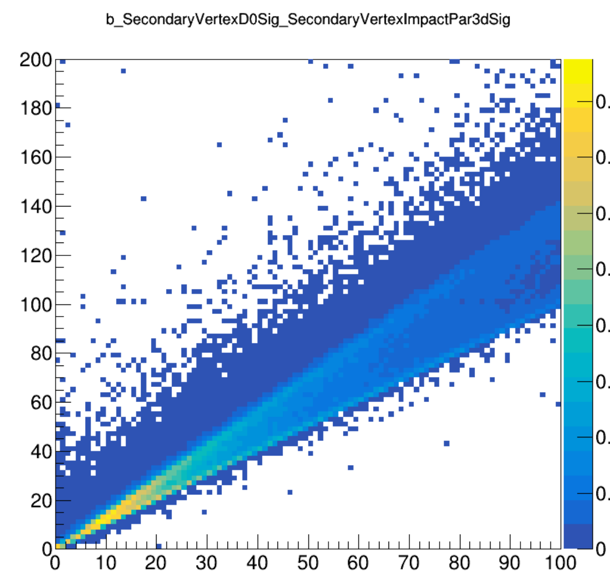
First order taylor coefficients



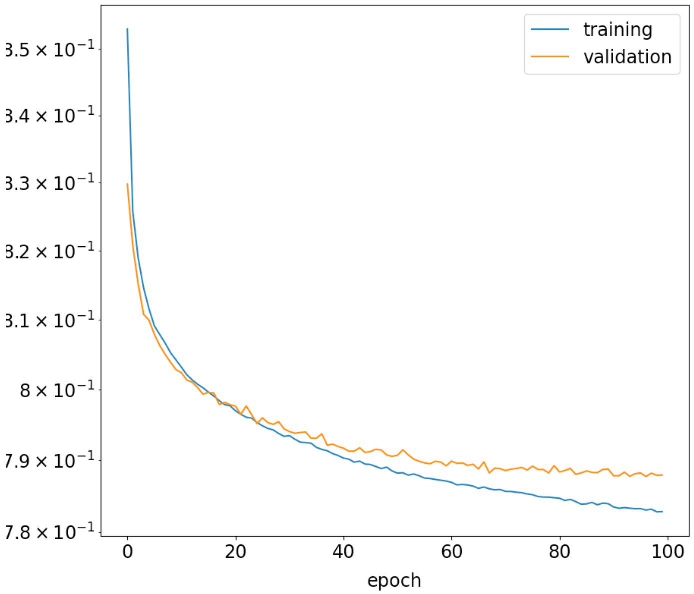
First order taylor coefficients

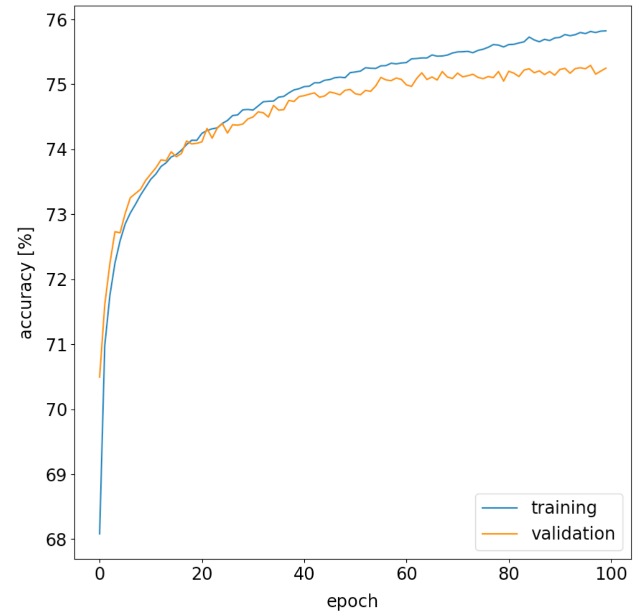


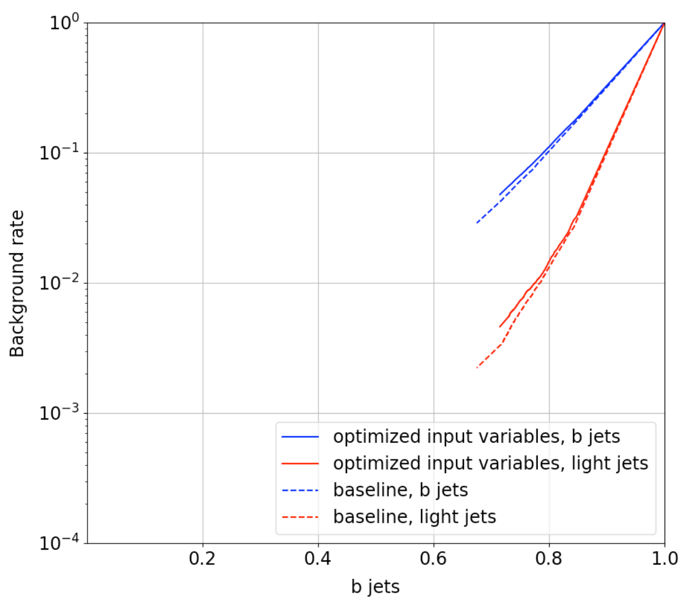


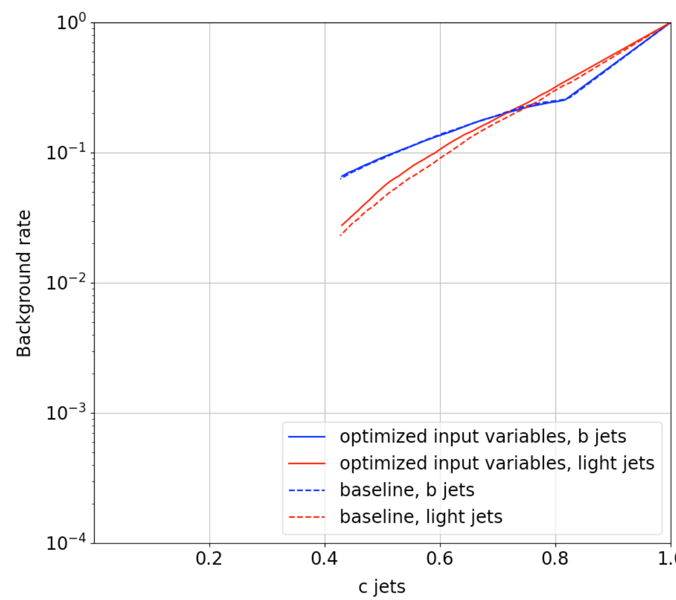


Results


training
validation
epoch
3.5 × 10^-1
3.4 × 10^-1
3.3 × 10^-1
3.2 × 10^-1
3.1 × 10^-1
8 × 10^-1
7.9 × 10^-1
7.8 × 10^-1
0 20 40 60 80 100


accuracy [%]
epoch
training
validation
68 69 70 71 72 73 74 75 76
0 20 40 60 80 100


Background rate
b jets
optimized input variables, b jets
optimized input variables, light jets
baseline, b jets
baseline, light jets
10^0
10^-1
10^-2
10^-3
10^-4
0.2 0.4 0.6 0.8 1.0


Background rate
c jets
optimized input variables, b jets
optimized input variables, light jets
baseline, b jets
baseline, light jets
10^0
10^-1
10^-2
10^-3
10^-4
0.2 0.4 0.6 0.8 1.0

Backup

Next steps

- optimize number of input constituents
- compare to LCFIPlus input variables
- comparison up-sampling, down-sampling, sample weights in loss function
- regularization
- hyperparameter optimization
- layer-norm instead of batch-norm

Toy studies

S.Wunsch, R. Friese, R.Wolf, and G. Quast, Computing and Software for Big Science 2 (Sep, 2018) 5, doi:10.1007/s41781-018-0012-, arXiv:1803.08782

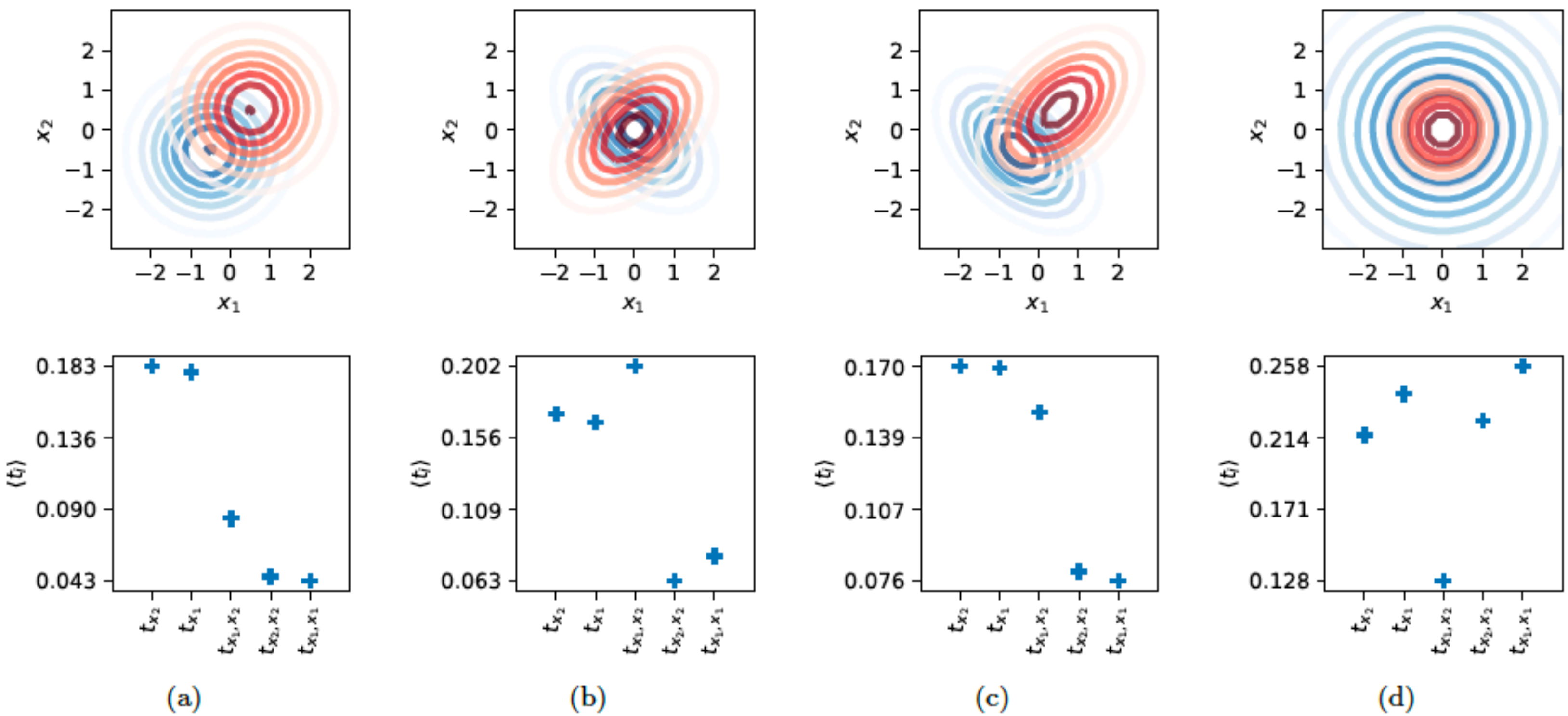
- simple task for illustration
- Keras, TensorFlow
- simple fully connected feed-forward NN
  - one hidden layer with 100 nodes
  - activation function: tanh / sigmoid
  - cross-entropy loss, Adam optimizer
- binary classification
- two inputs: x_1 and x_2 (Gaussian distributions for signal and background)

Task	Mean value				Covariance matrix	
	Signal (x_1, x_2)		Background (x_1, x_2)		Signal	Background
Fig. 1a	0.5	0.5	−0.5	−0.5	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$
Fig. 1b	0	0	0	0	$\begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & -0.5 \\ -0.5 & 1 \end{pmatrix}$
Fig. 1c	0.5	0.5	−0.5	−0.5	$\begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & -0.5 \\ -0.5 & 1 \end{pmatrix}$
Fig. 1d	0	0	0	0	$\begin{pmatrix} 0.5 & 0 \\ 0 & 0.5 \end{pmatrix}$	$\begin{pmatrix} 3 & 0 \\ 0 & 3 \end{pmatrix}$

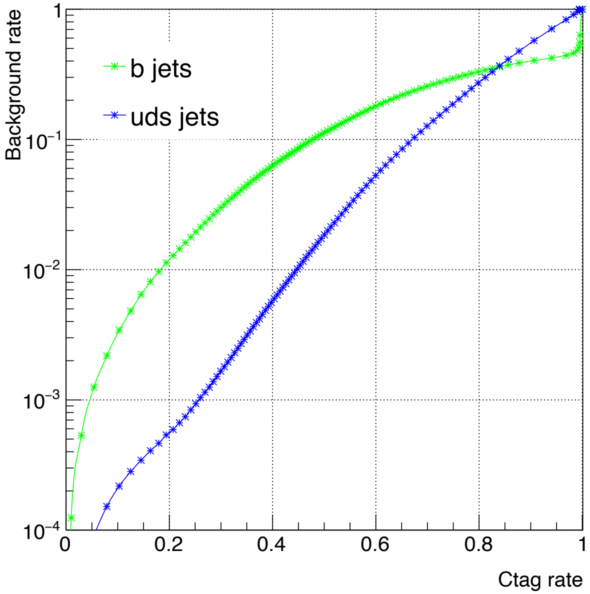
Toy studies: results

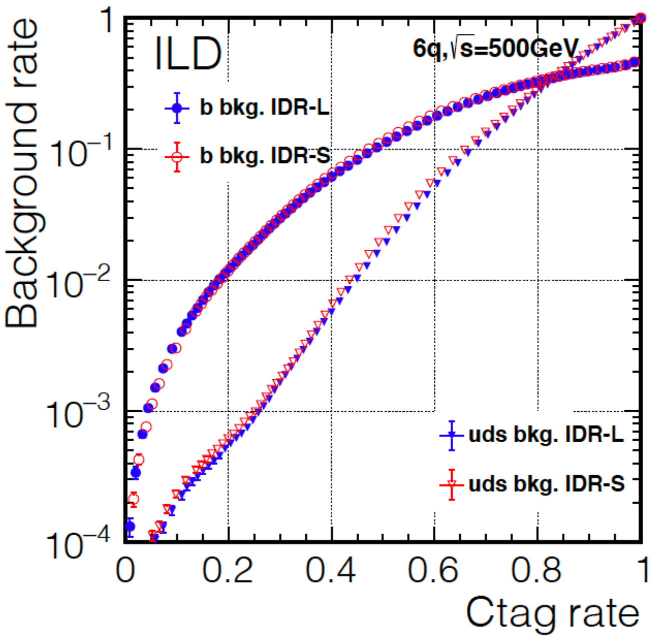
S.Wunsch, R. Friese, R.Wolf, and G. Quast, Computing and Software for Big Science 2 (Sep, 2018) 5, doi:10.1007/s41781-018-0012-, arXiv:1803.08782

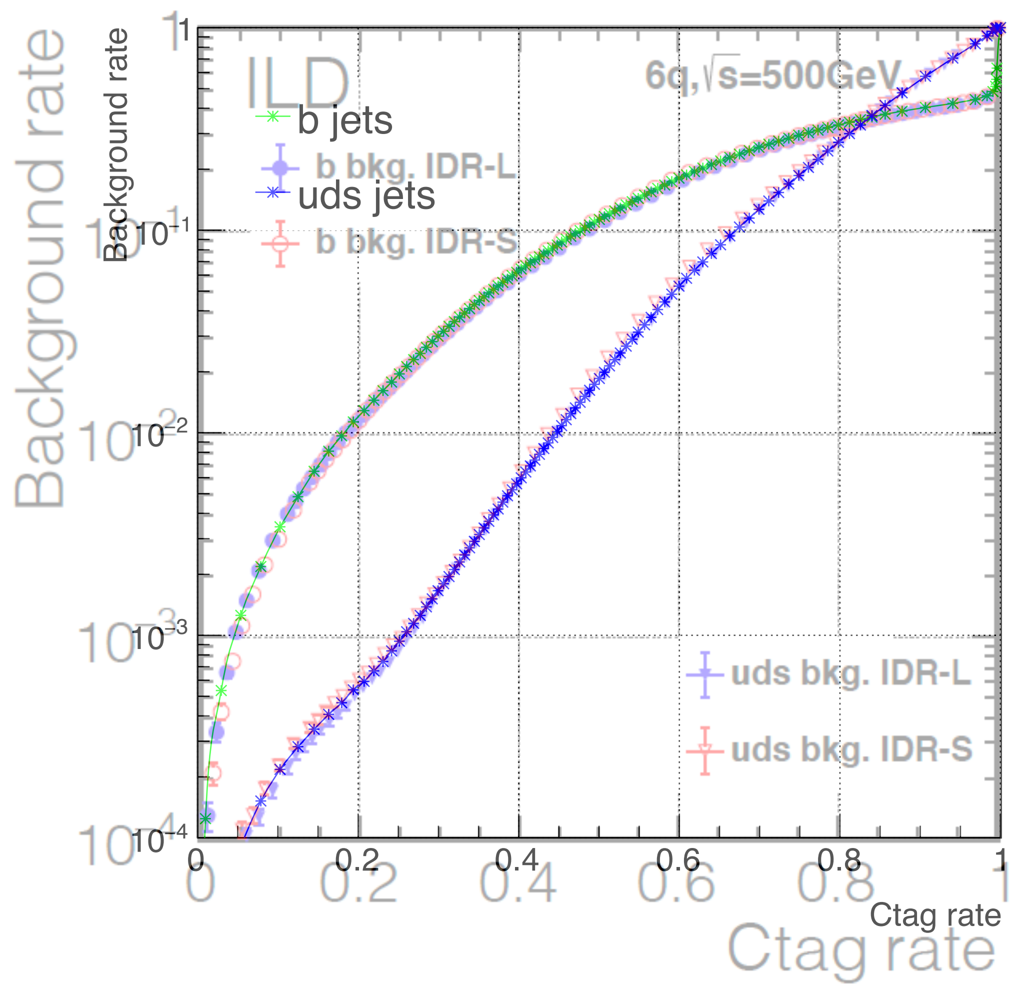
- <t_{x1}>, <t_{x2}>: influence of 1d distributions of x_1, x_2
- <$t_{x1,x2}$>: influence of the correlation of x_1 and x_2
- <$t_{x1,x1}$> , <$t_{x2,x2}$> : indicating the influence of the auto-correlation



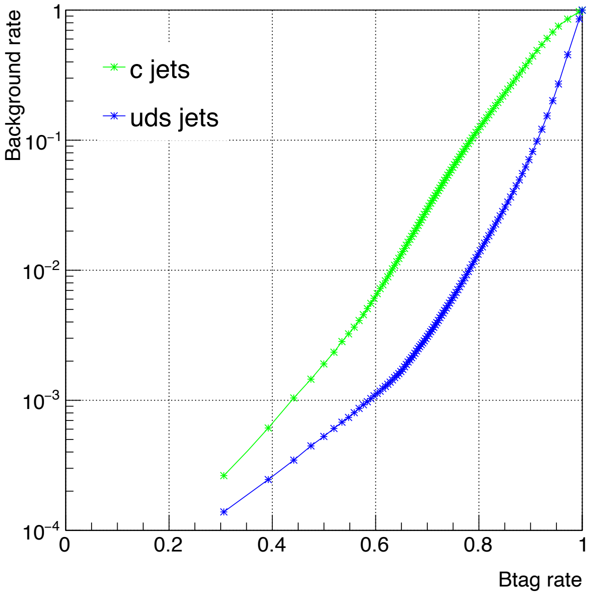
Performance LCFIPlus

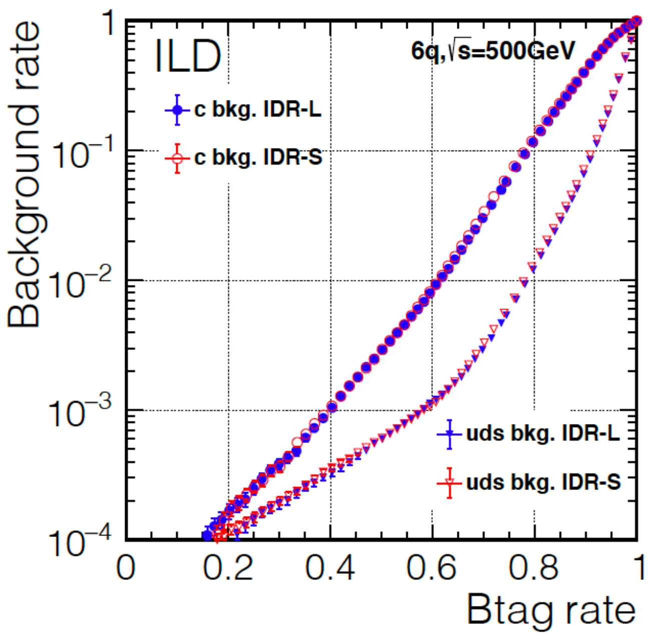


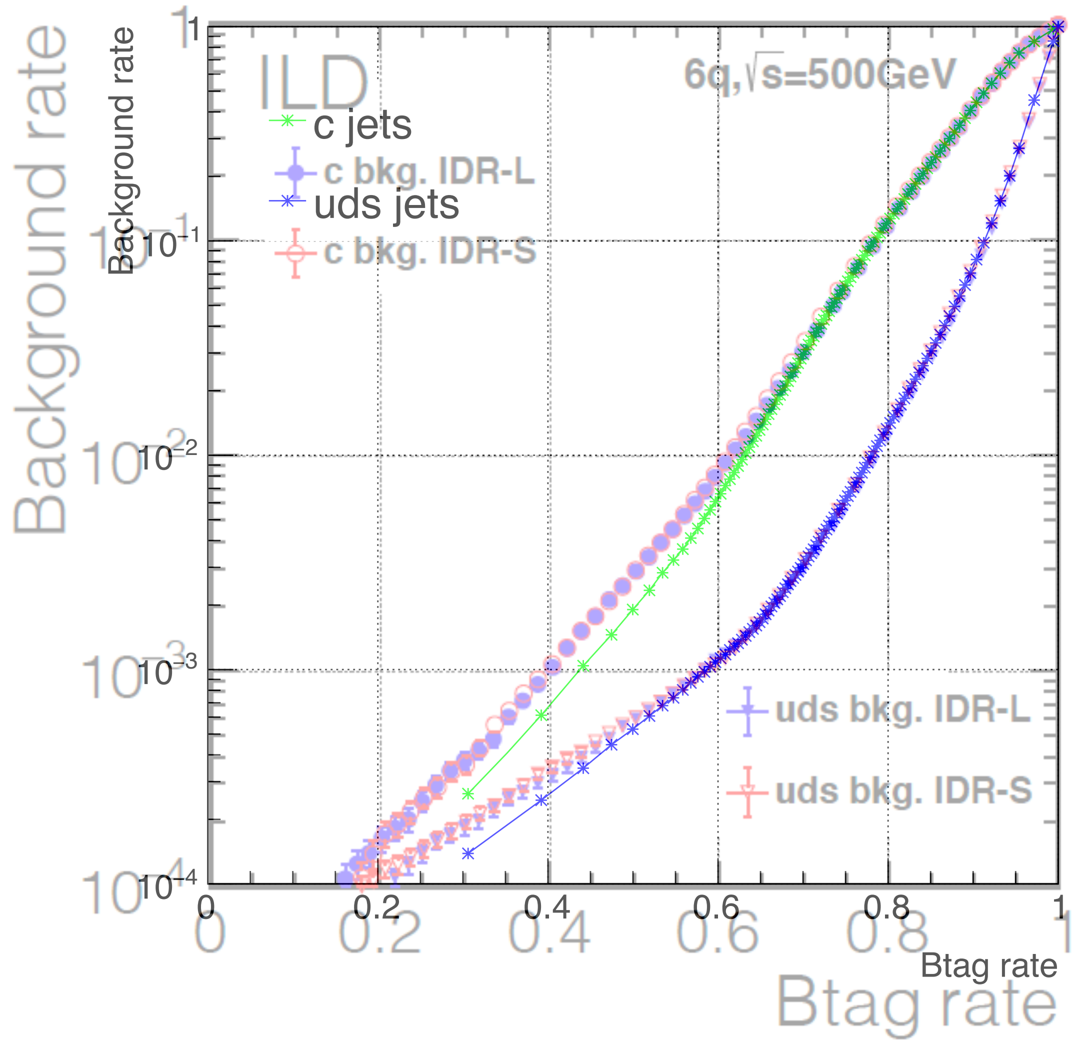


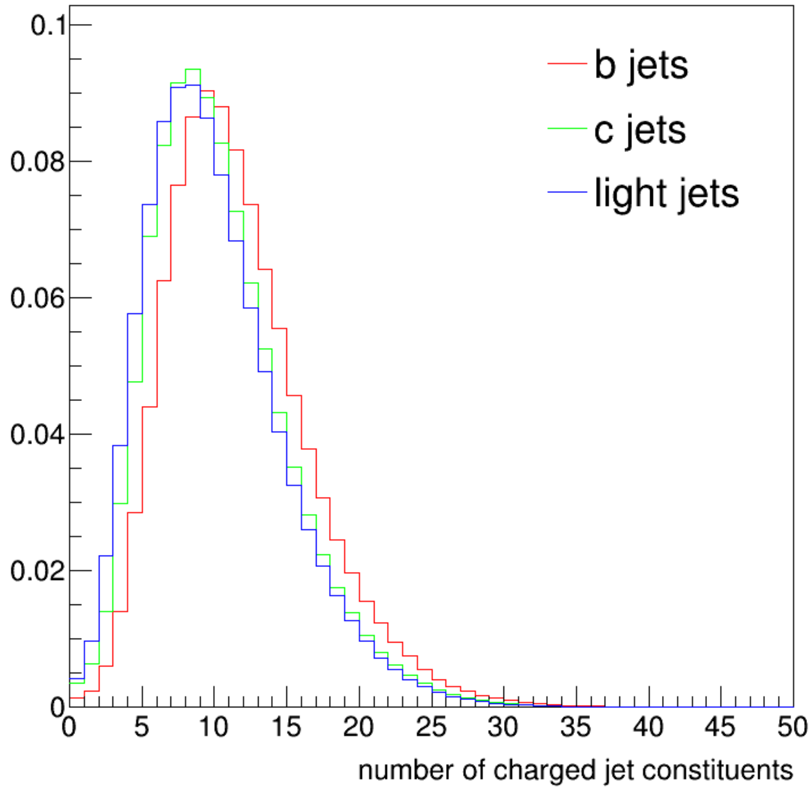


Performance LCFIPlus

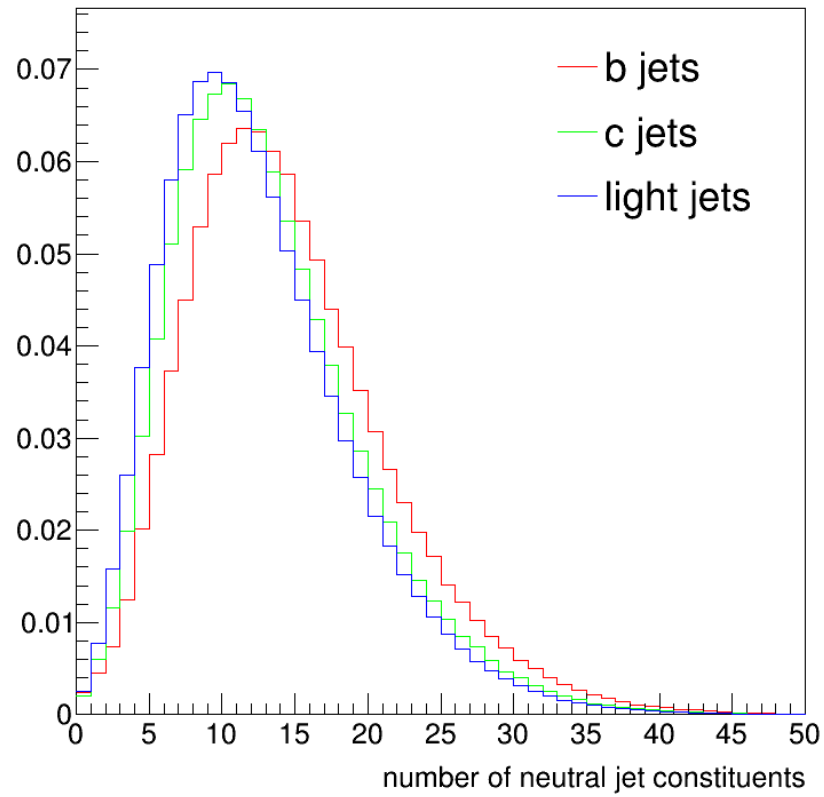






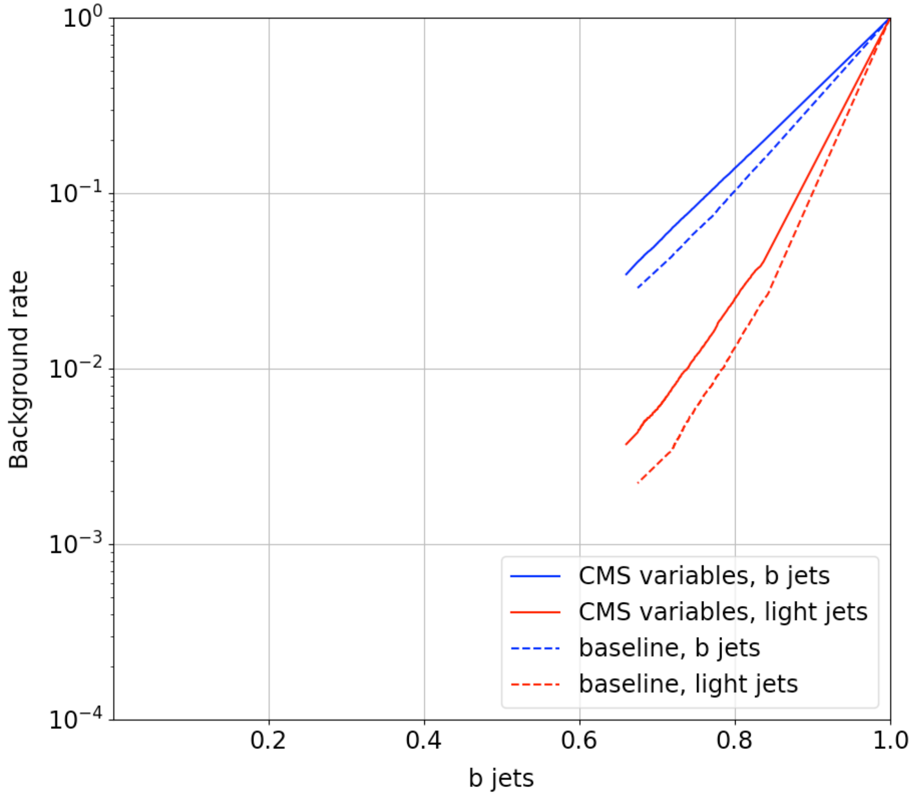


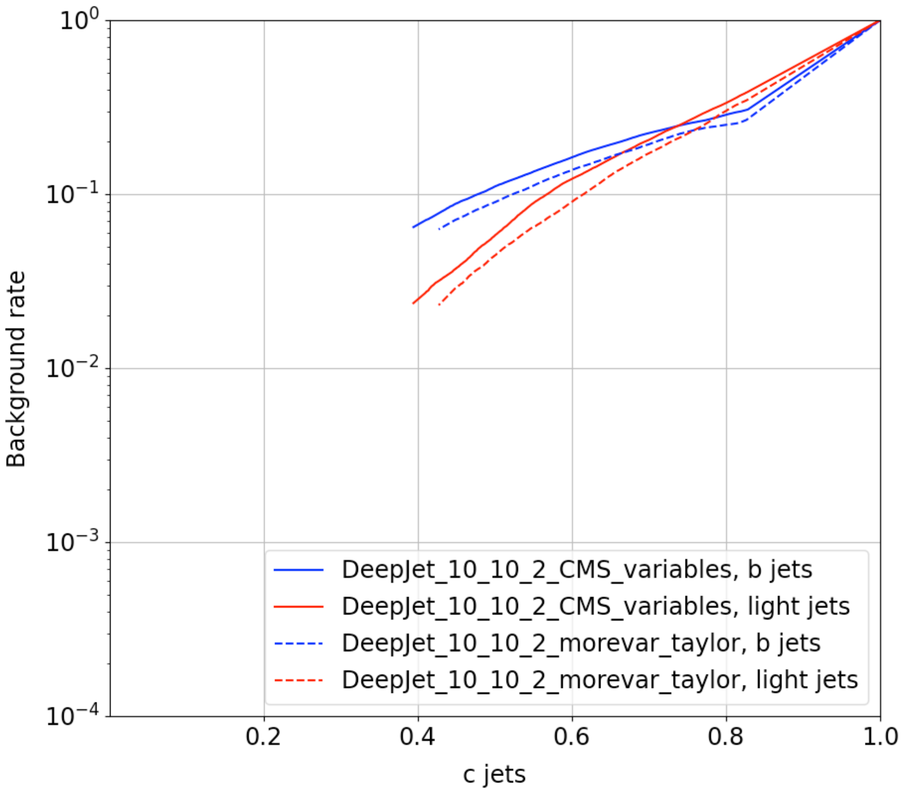
0.1
0.08
0.06
0.04
0.02
0
b jets
c jets
light jets
0
5
10
15
20
25
30
35
40
45
50
number of charged jet constituents



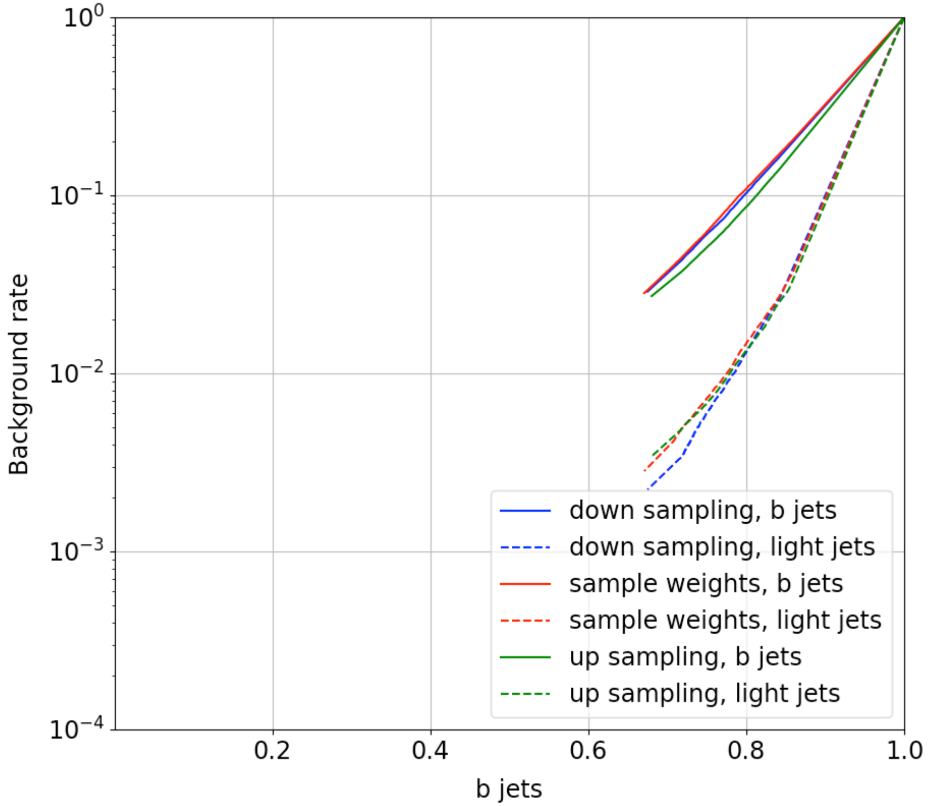
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
b jets
c jets
light jets
0
5
10
15
20
25
30
35
40
45
50
number of neutral jet constituents

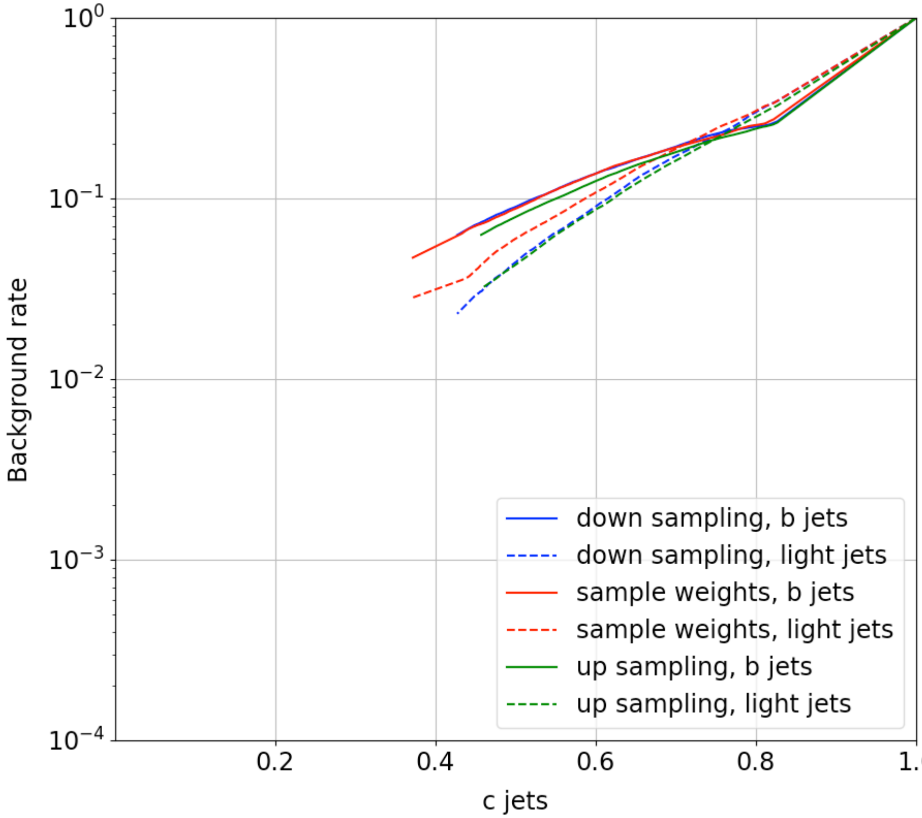
Training with CMS variables



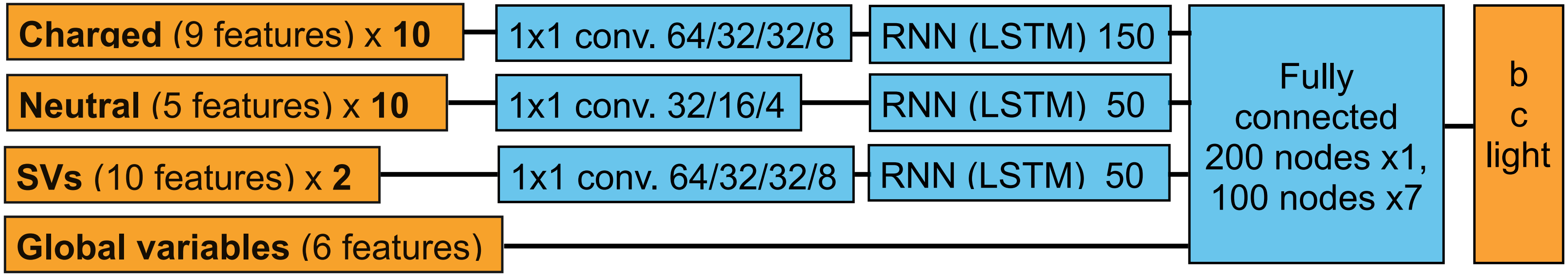


Up-sampling, down-sampling, sample weights



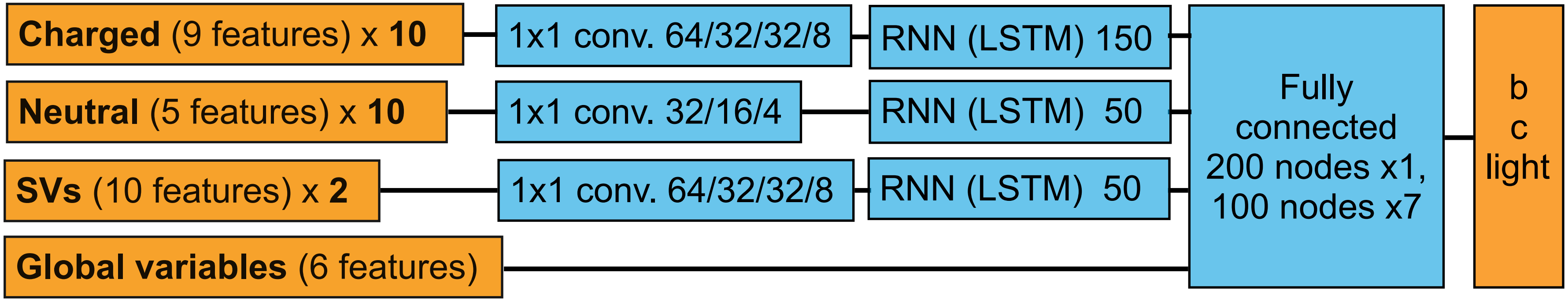


- reminder architecture:



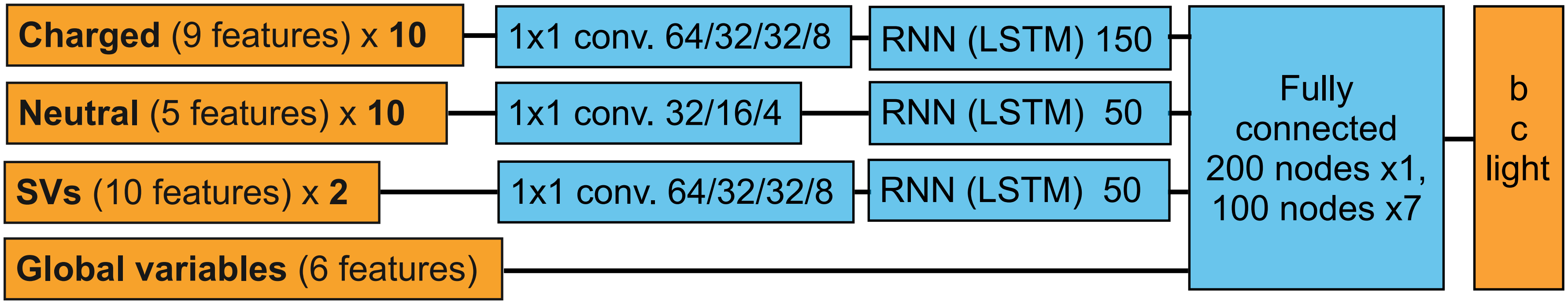
- **new trainings:**
  - DeepJet architecture with more constituents
    - **Charged** (9 features) x **25**
    - **Neutral** (5 features) x **25**
    - **SVs** (10 features) x **2**
    - **Global variables** (6 features)
  - DeepJet architecture with more variables
    - **Charged** (23 features) x **10**
    - **Neutral** (6 features) x **10**
    - **SVs** (16 features) x **2**
    - **Global variables** (7 features)

DeepJet: architecture & training



- ordering of particle objects by
  - impact parameter significance for charged jet constituents
  - shortest angular distance to a secondary vertex (by momentum if there is no secondary vertex) for neutral jet constituents
  - flight distance significance for secondary vertices

DeepJet: architecture & training



- batch normalization, dropout (0.1)
- activation functions: relu / softmax (last layer)
- cross entropy loss
- optimizer: Adam
- batch size: 50
- learning rate: 0.0003
- 75% training, 12.5% validation, 12.5% test
- number of epochs: 100
- balanced input: select c/light jets randomly to get same number of b,c,light jets
- Xavier weight initialization

Input features

jets:

- jet rapidity, jet momentum, number of charged jet constituents, number of neutral jet constituents, number of secondary vertices, number of pseudo-vertices

charged jet constituents:

- momentum,Θ(track,jet), ΔR(track, SV), d0 significance, Z0 significance, 3D IP significance, track reconstructed in PV, lepton momentum relative to jet, kaon-ness of charged particles

neutral jet constituents:

- ΔR(neutral,SV), ΔR(neutral,jet), neutral HCAL fraction,momentum/jet momentum, is photon?

secondary vertices:

- secondary vertex mass, number of tracks in SV, χ2/ndf, ΔR(SV,closest jet), SV energy/ jet energy, cosine of the angle between the secondary vertex flight direction and the direction of the secodary vertex momentum, d0 significance, Z0 significance, 3D IP, 3D IP significance

First order taylor coefficients