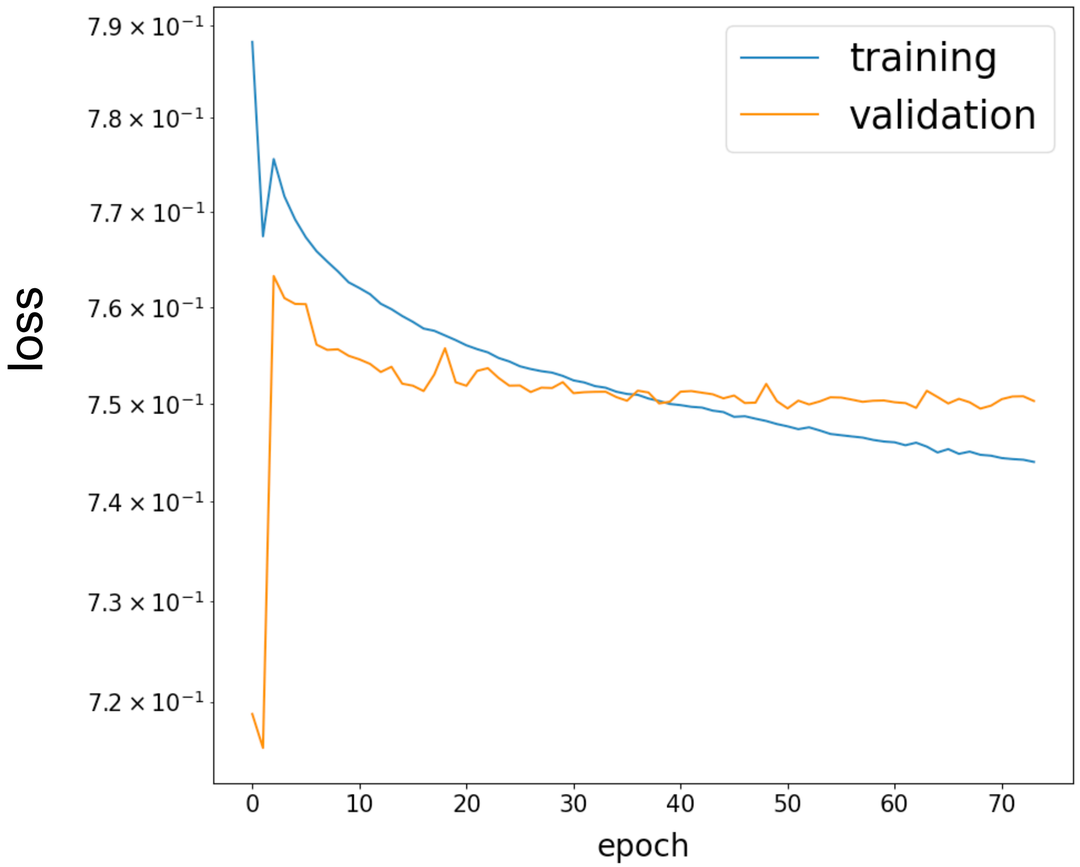
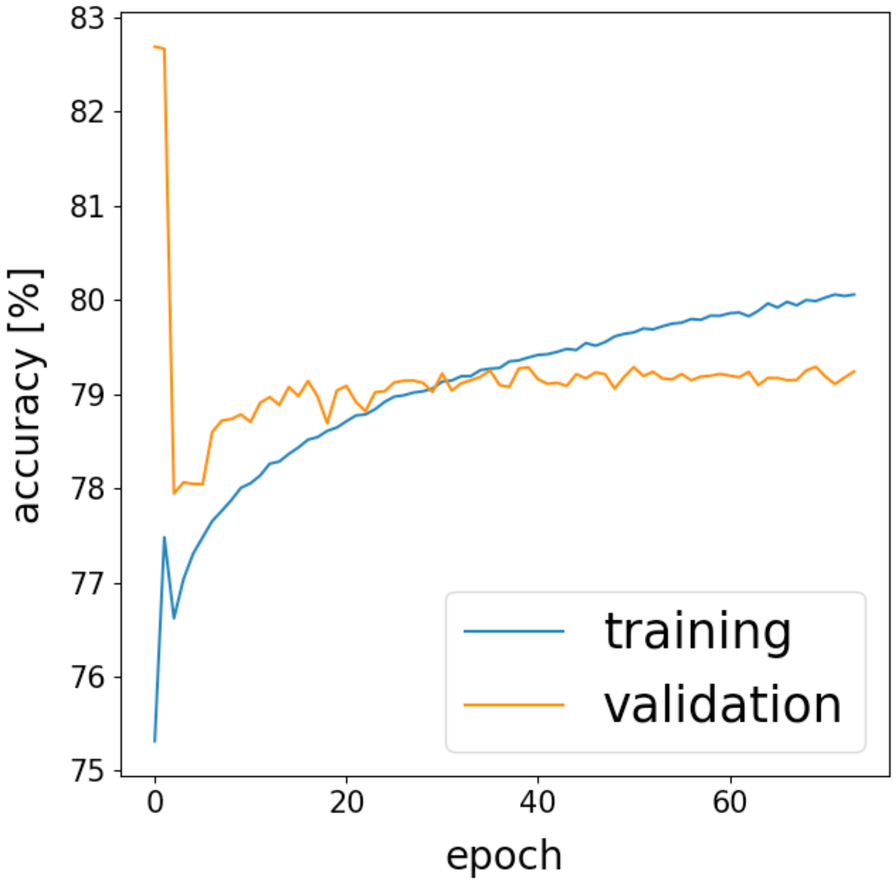
News

- found bug in forward pass of NN
  - data : (batch size, length of sequence, number of features)
  - batch norm and CNN layers: (batch size, length of sequence, number of features)
  - LSTM: (batch size, number of features, length of sequence)
- started to study influence of composition of training and validation sample on results (my results so far: 1:1:1 for b:c:light, LCFIPlus: 1:1:3)

- many new plots:
  - ROC curves with stat. uncertainties
  - AUC for each epoch
  - ROC curves for different epochs
  - NN output for different epochs
  - validation sample split in two samples
  - LCFIPlus: studied different binnings, ROC curves with uncertainties, ROC determined with efficiencies not by BDT output

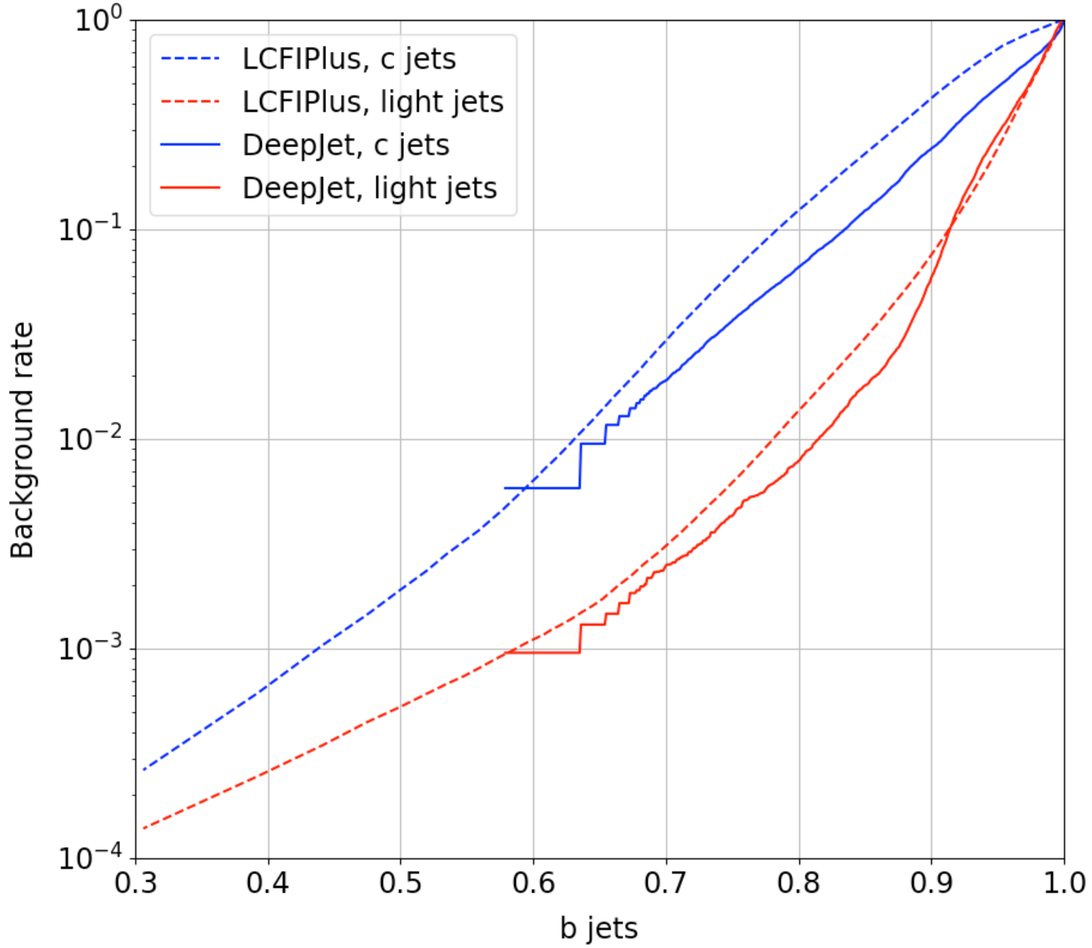
Loss and accuracy

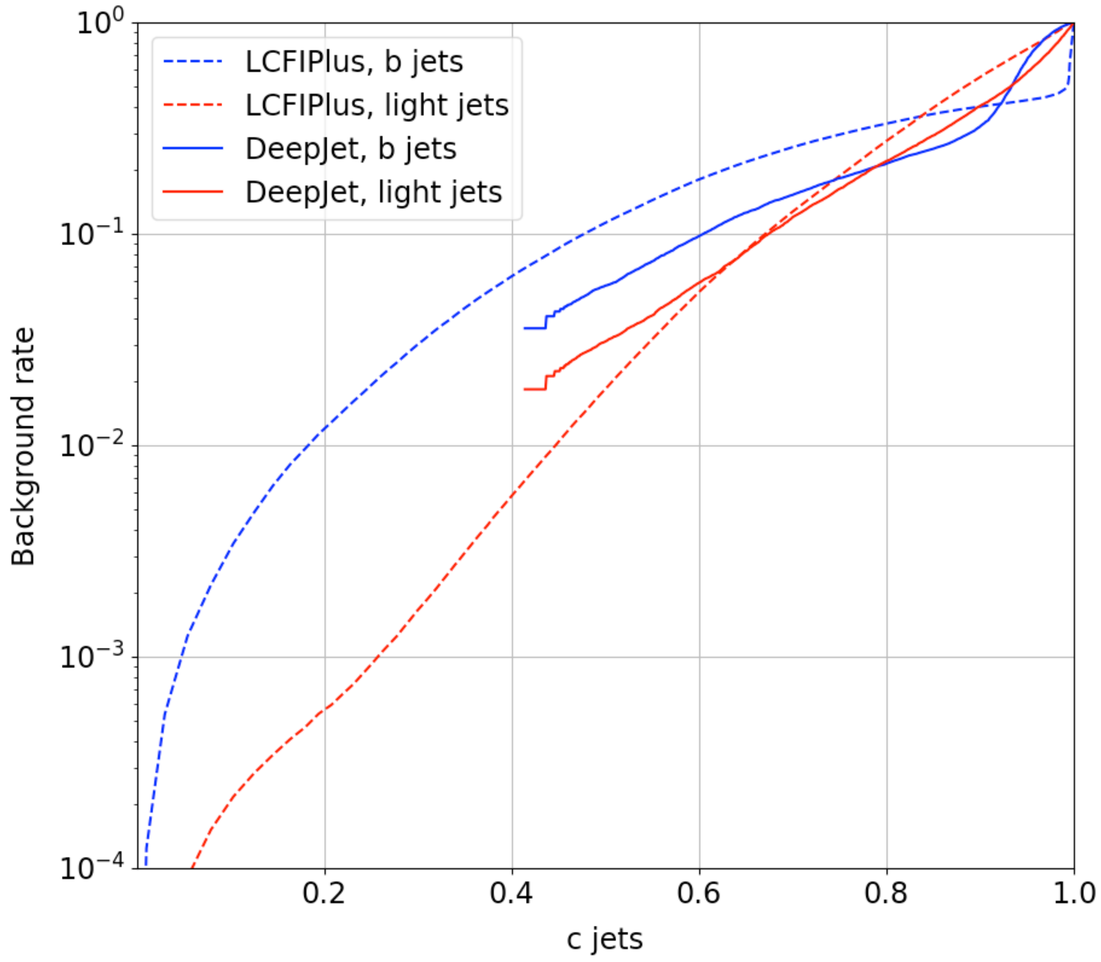




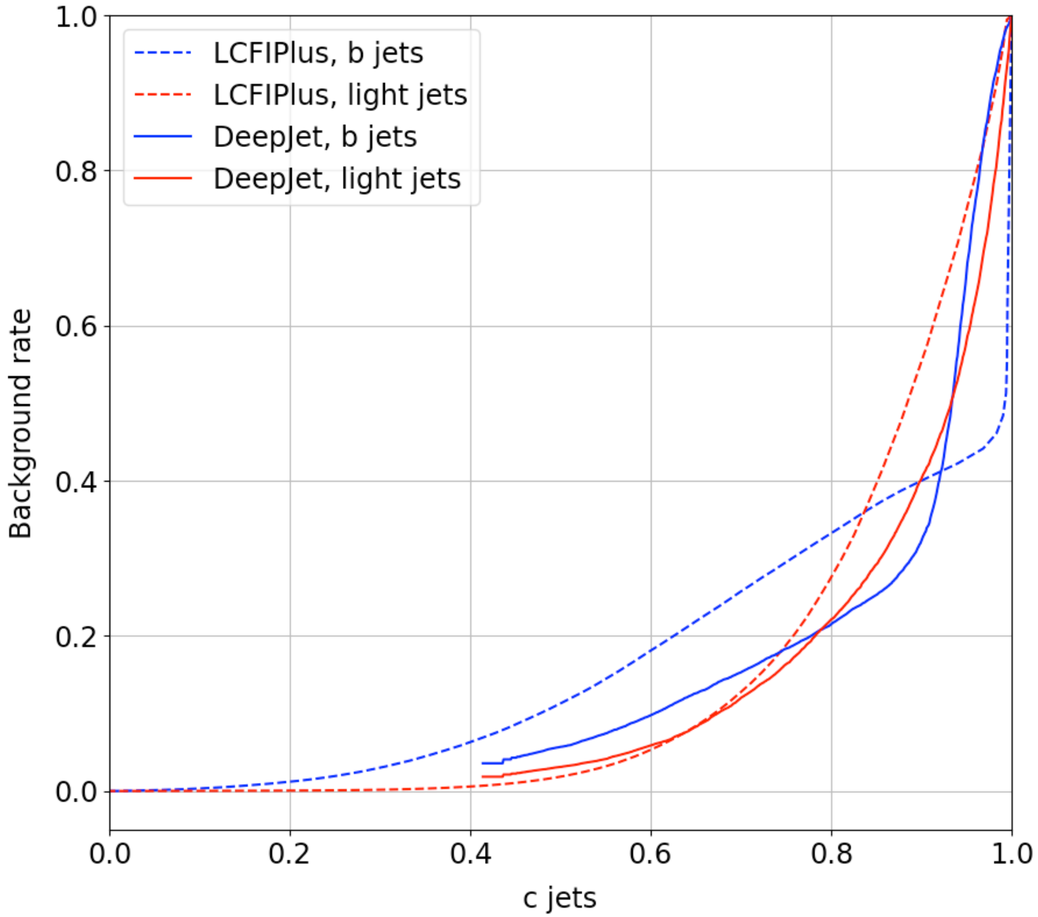
- down-sampling of light and c jets to get same number of b jets in training and validation data
- over-training

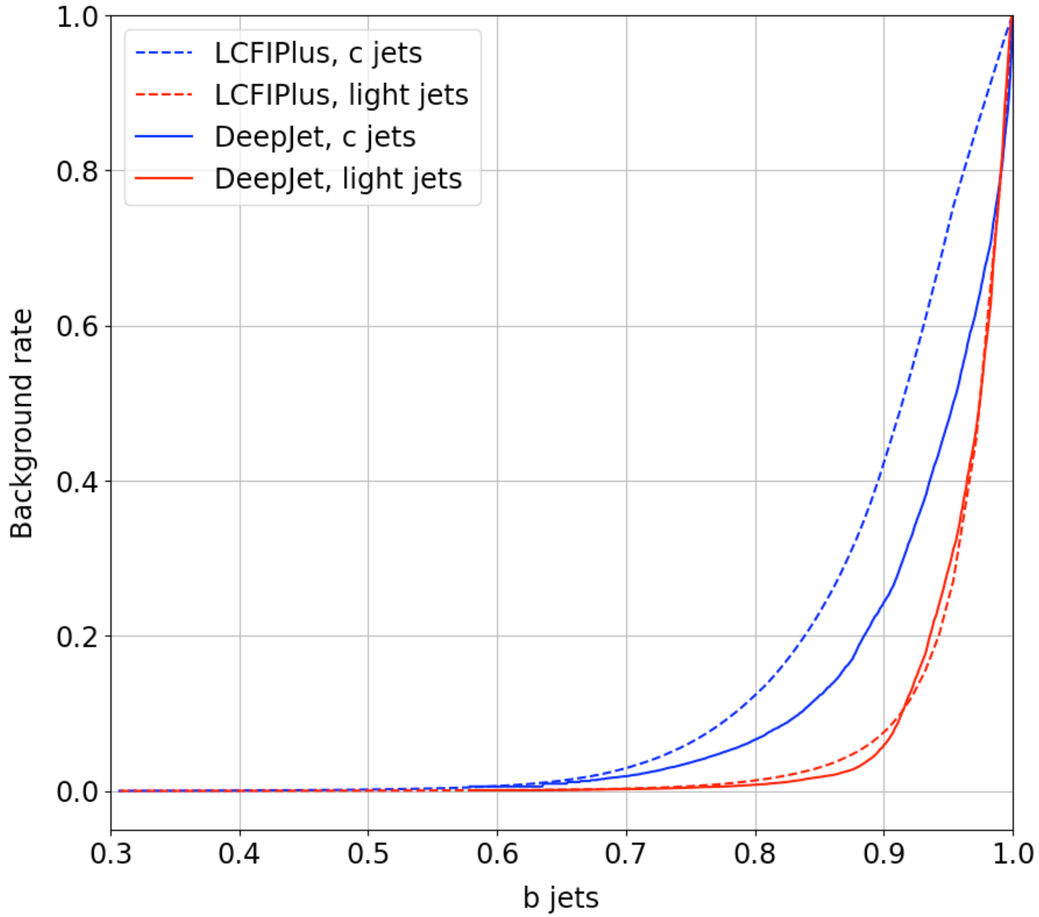
ROC curves


LCFIPlus, c jets
LCFIPlus, light jets
DeepJet, c jets
DeepJet, light jets
Background rate
b jets
10^0
10^{-1}
10^{-2}
10^{-3}
10^{-4}
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0

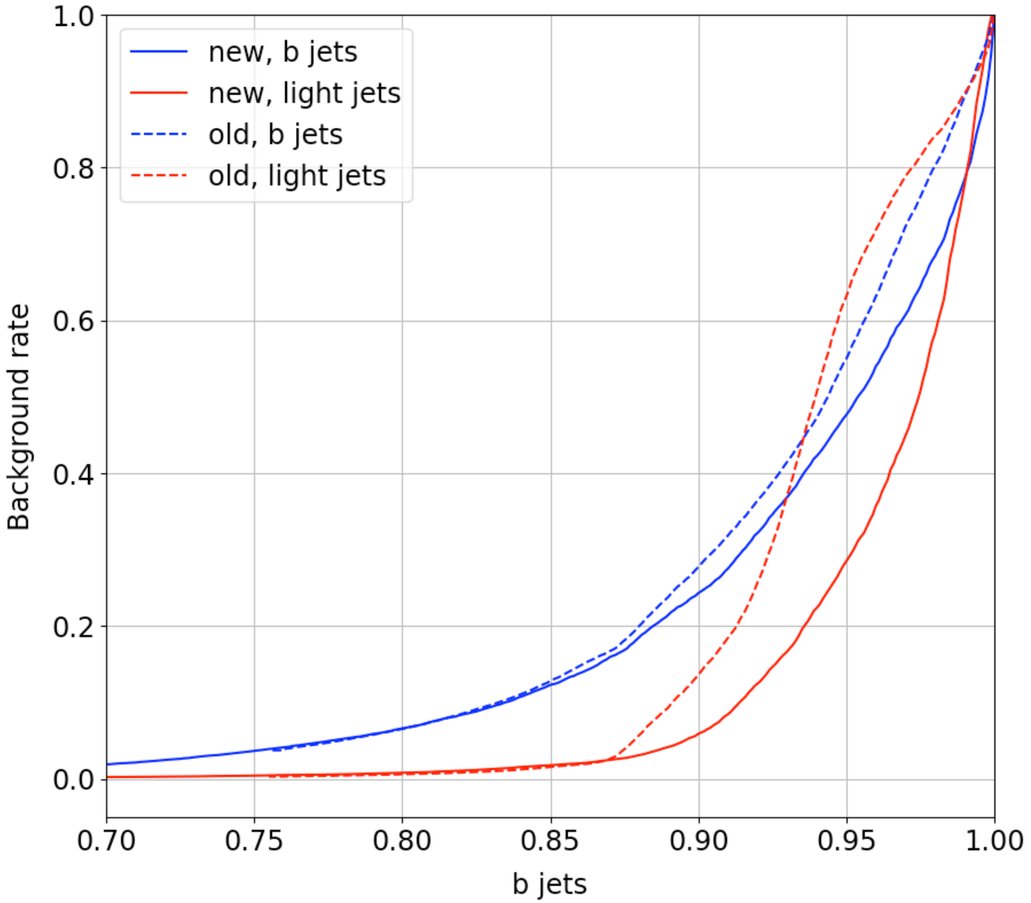

LCFIPlus, b jets
LCFIPlus, light jets
DeepJet, b jets
DeepJet, light jets
Background rate
c jets
10^0
10^{-1}
10^{-2}
10^{-3}
10^{-4}
0.2
0.4
0.6
0.8
1.0

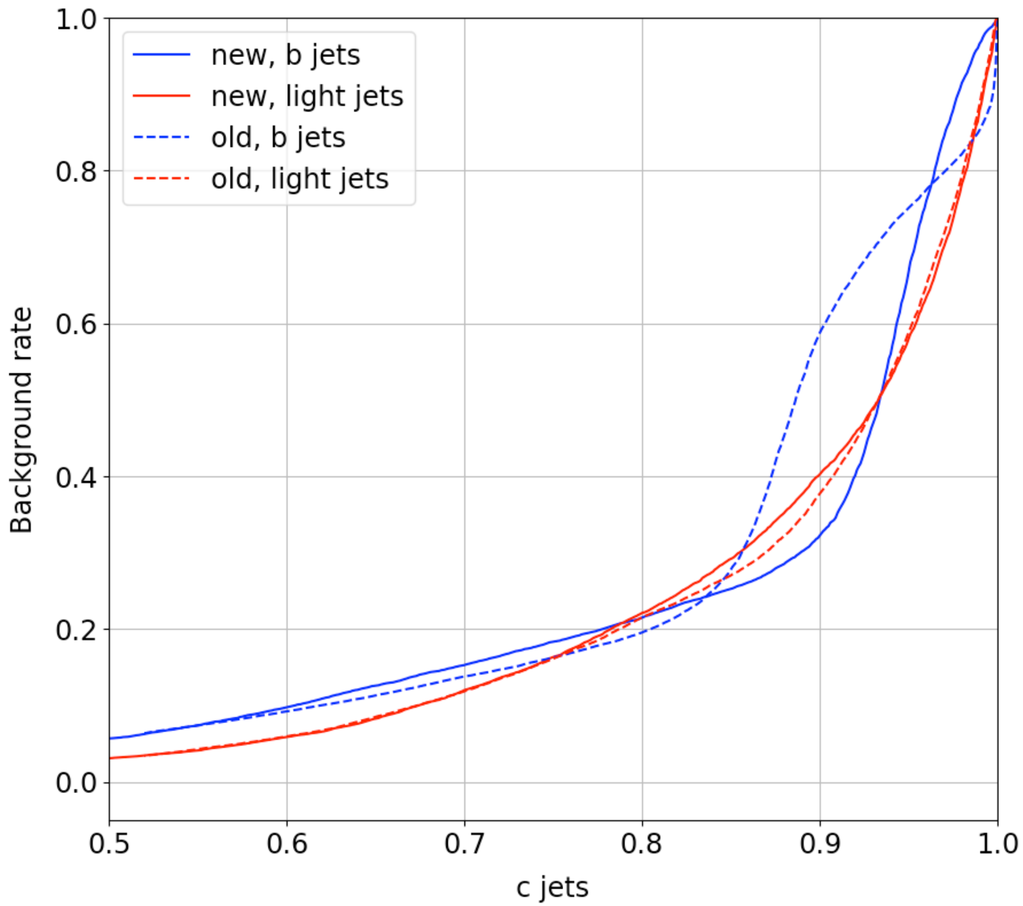
ROC curves


LCFIPlus, b jets
LCFIPlus, light jets
DeepJet, b jets
DeepJet, light jets
Background rate
c jets
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

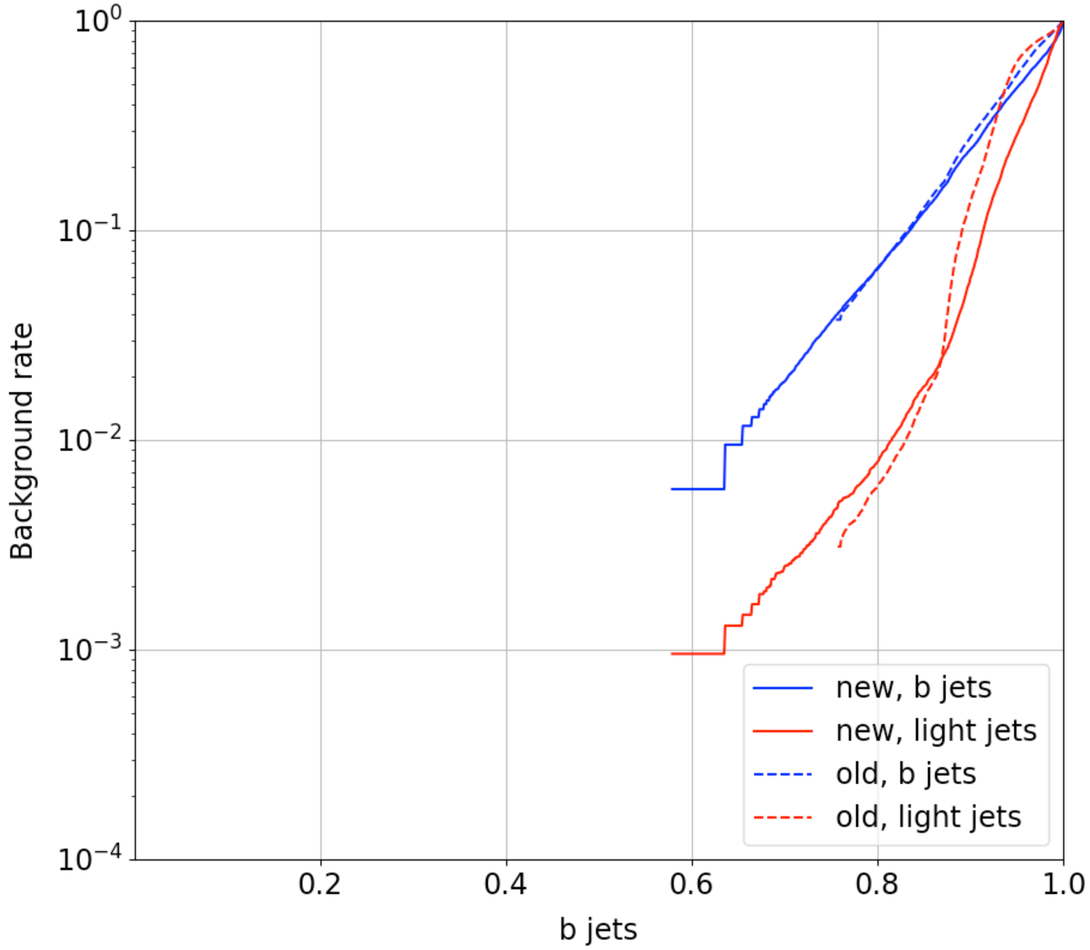

LCFIPlus, c jets
LCFIPlus, light jets
DeepJet, c jets
DeepJet, light jets
Background rate
b jets
1.0
0.8
0.6
0.4
0.2
0.0
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0

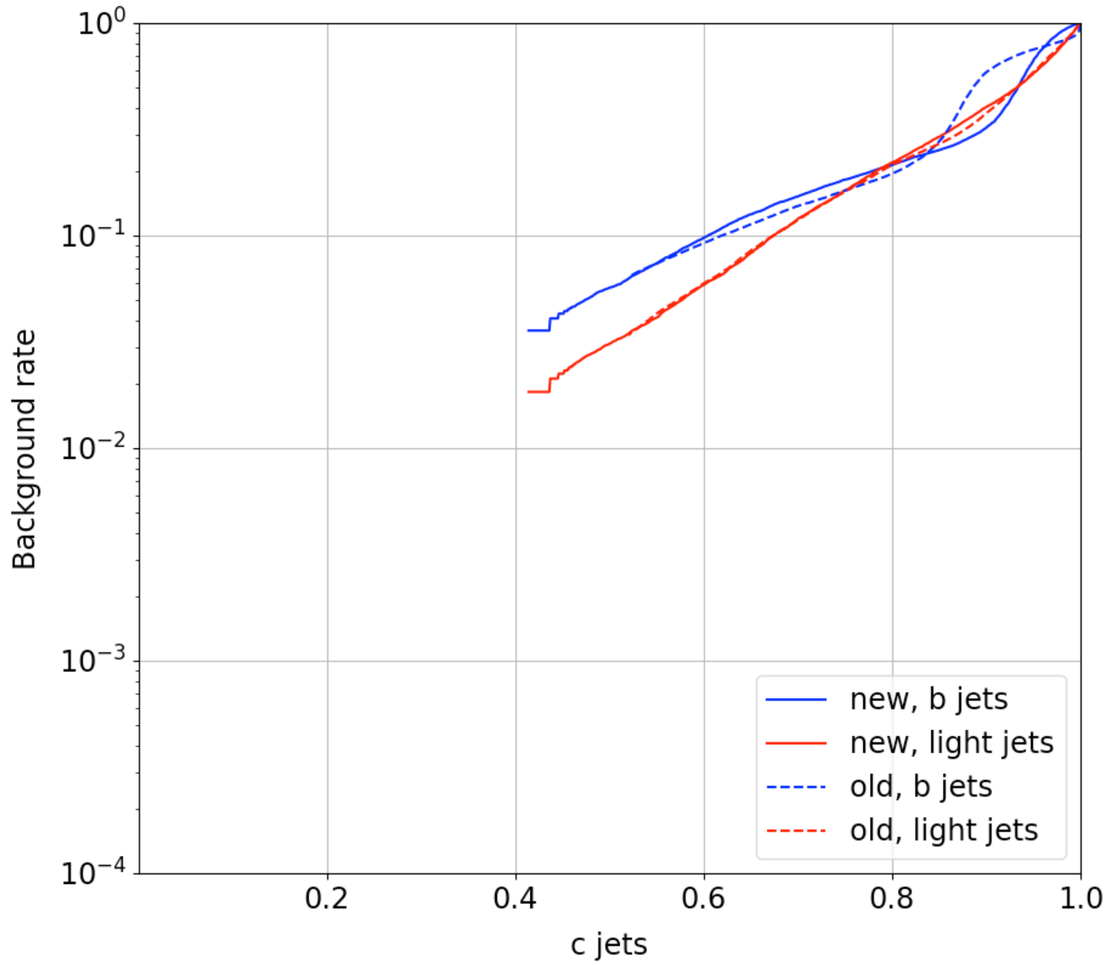
Comparison to results with bug



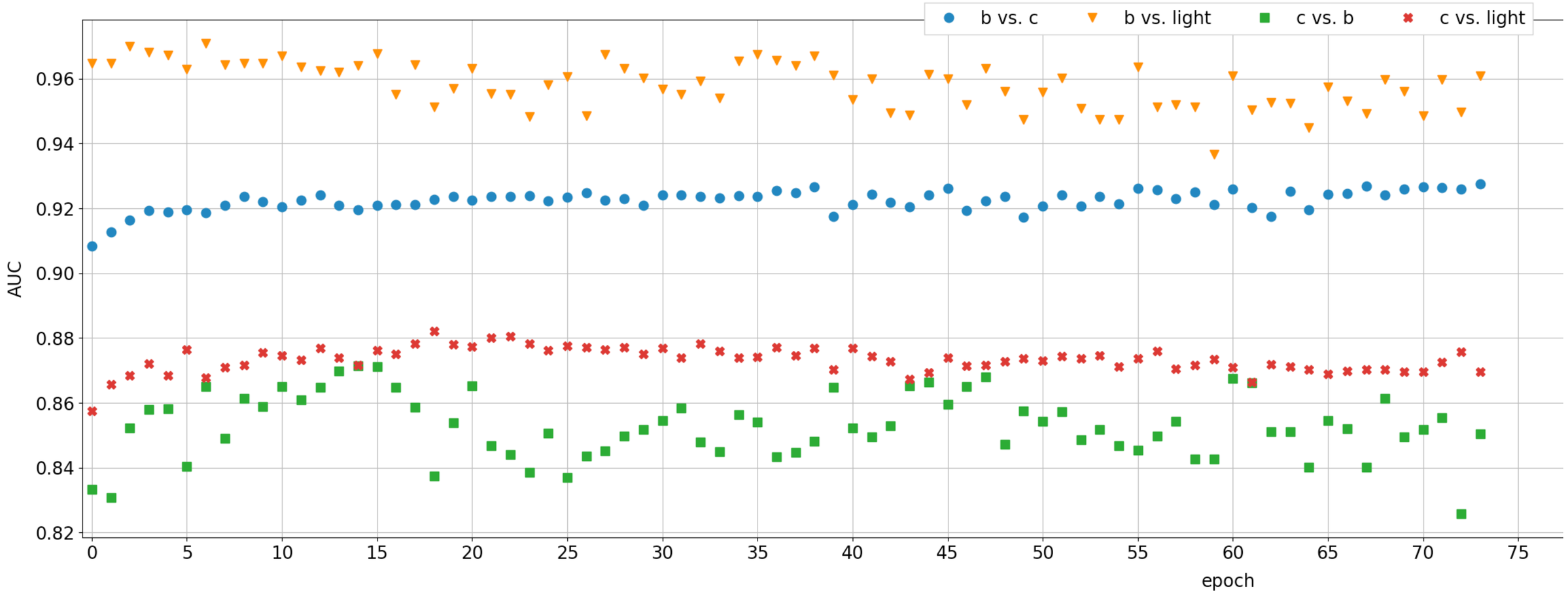


Comparison to results with bug



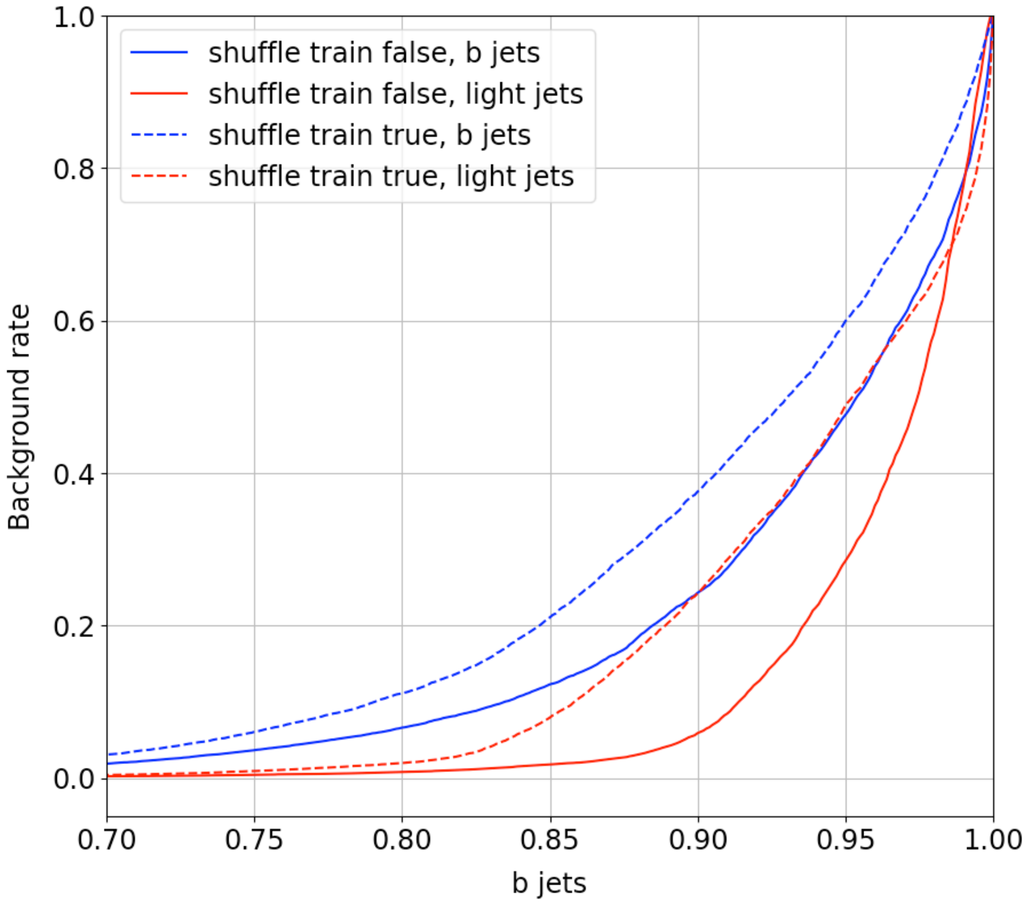


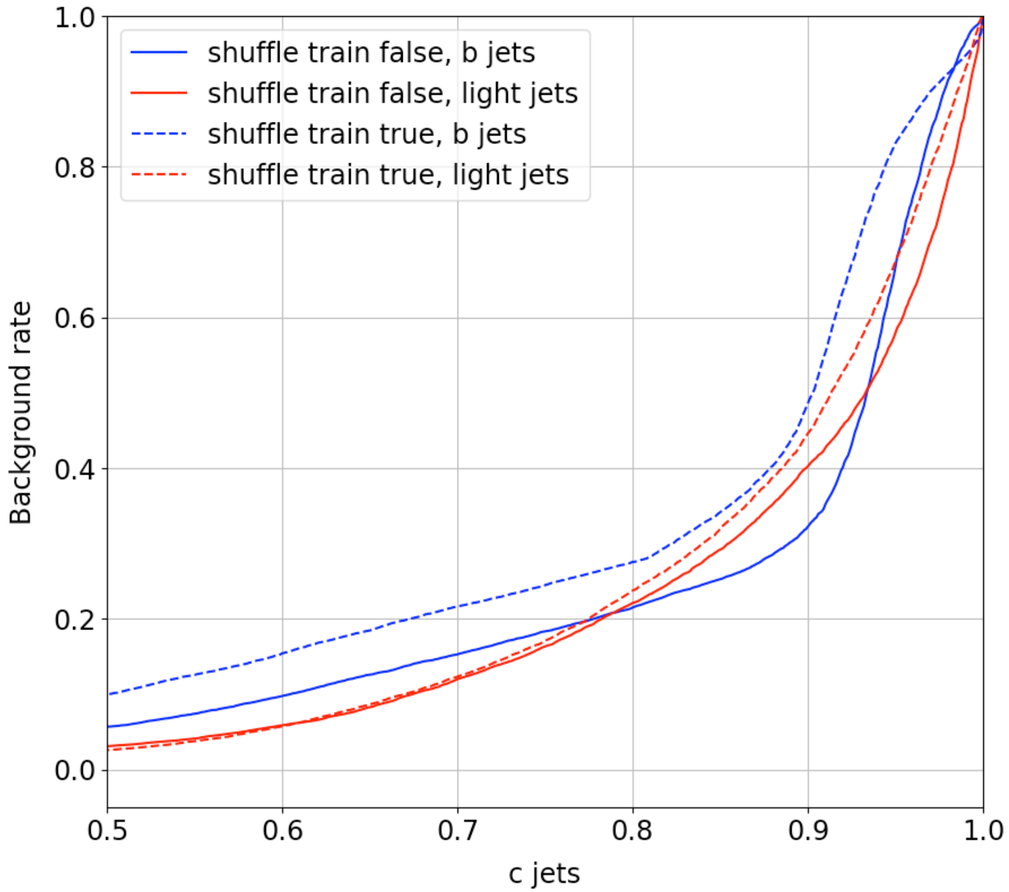
AUC for each epoch



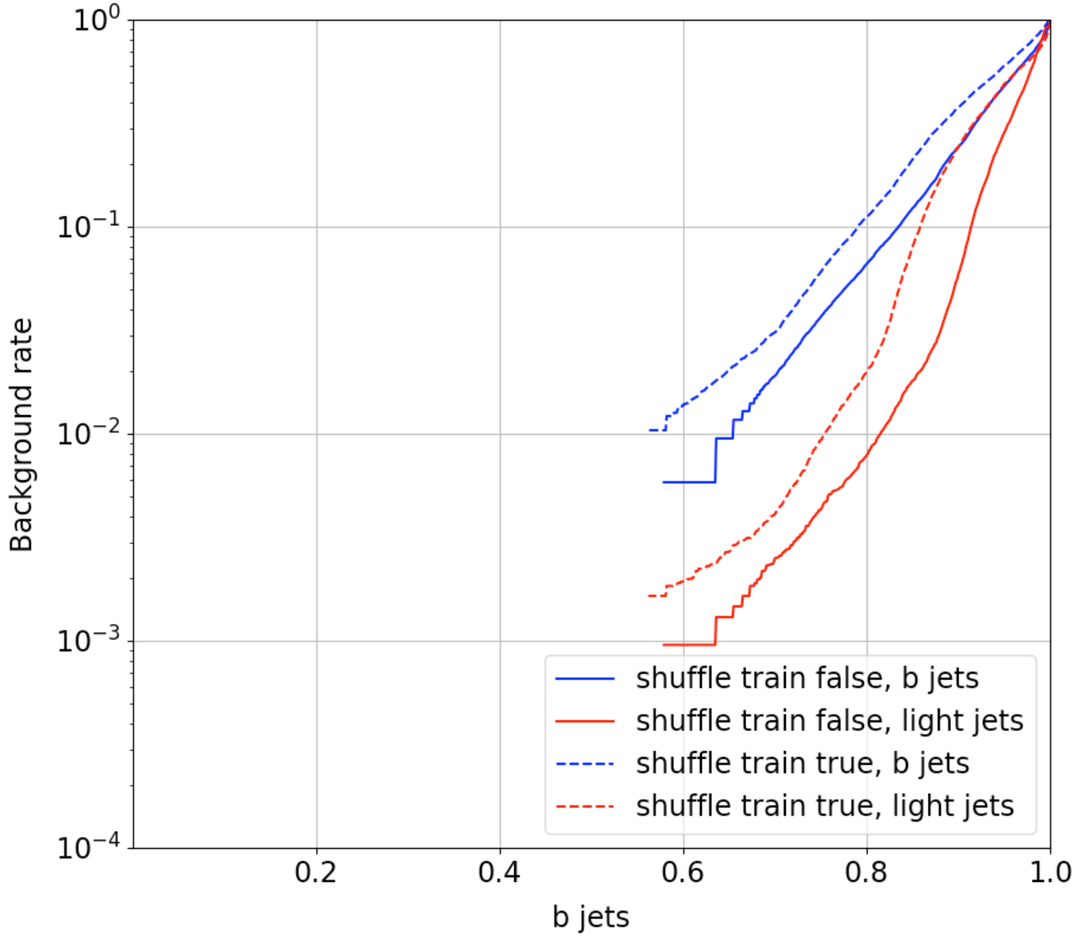
- max. AUC b vs. c at epoch: 73
- max. AUC b vs. light at epoch: 6
- max. AUC c vs. b at epoch: 14
- max. AUC c vs. light at epoch: 18

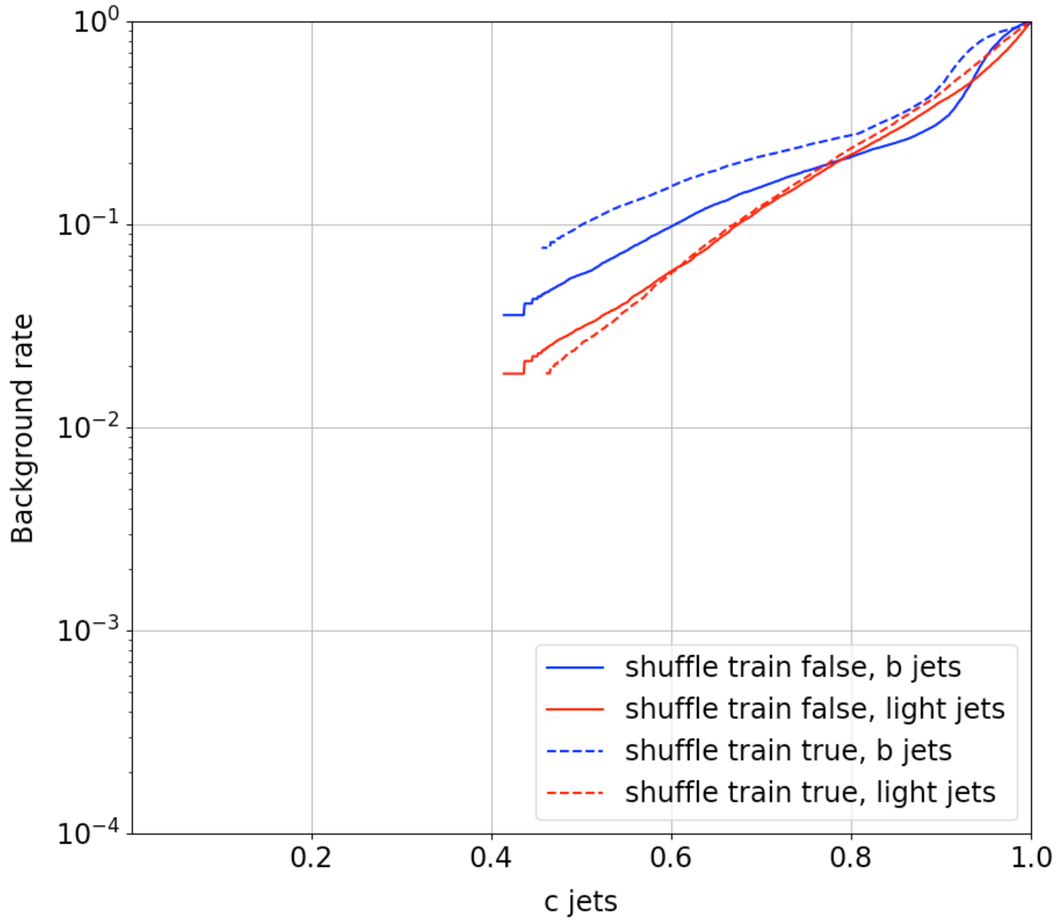
Shuffle training data before each epoch



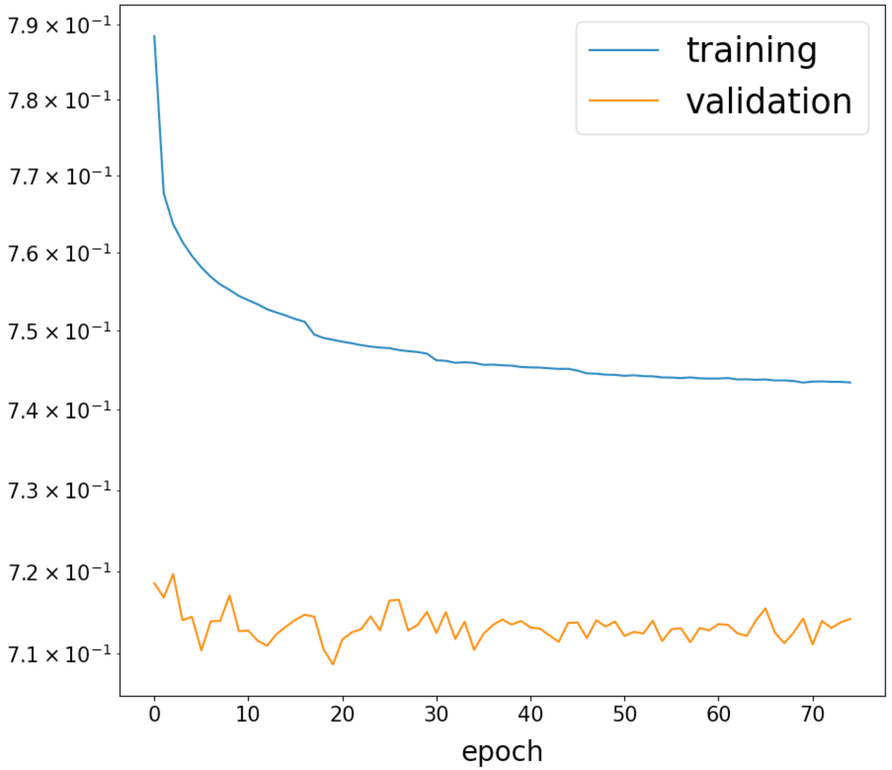


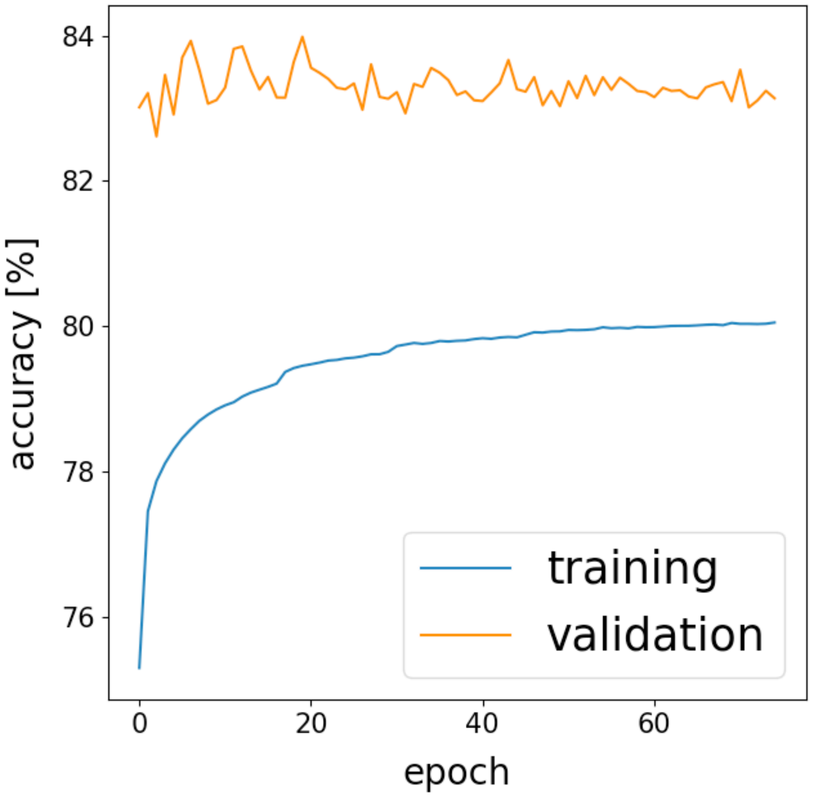
Shuffle training data before each epoch



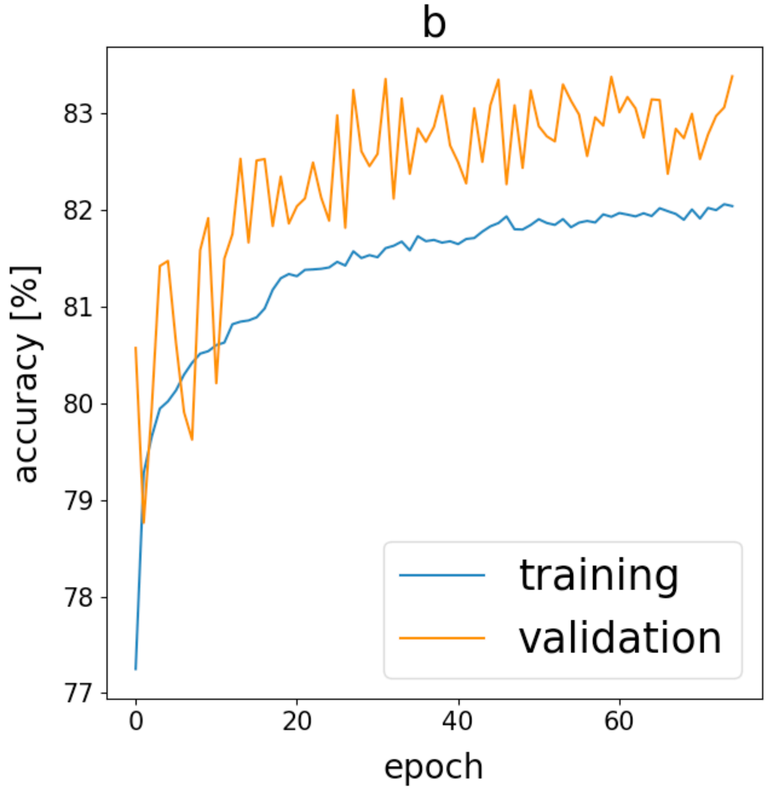


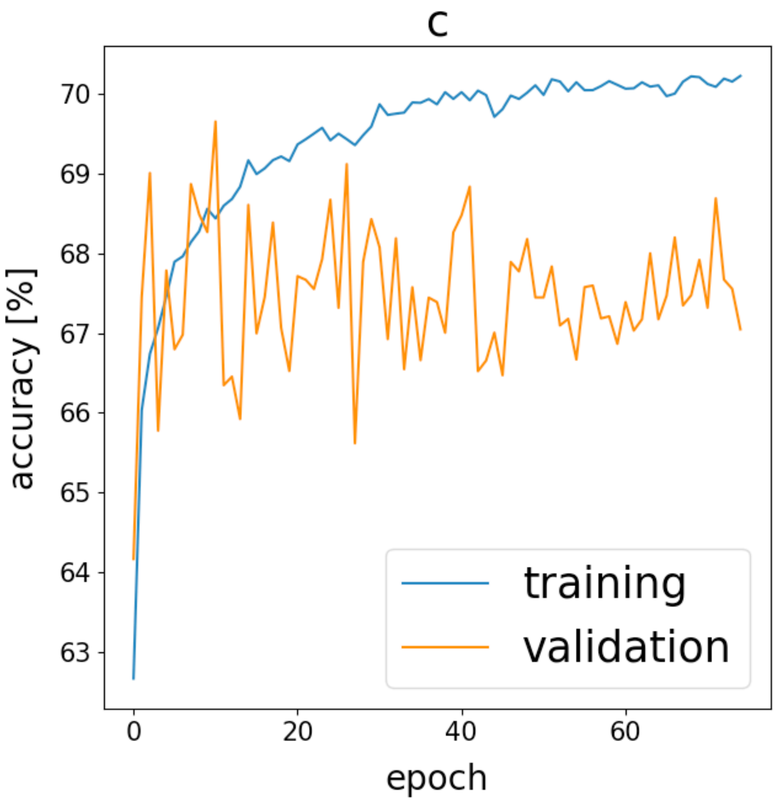
Loss and accuracy

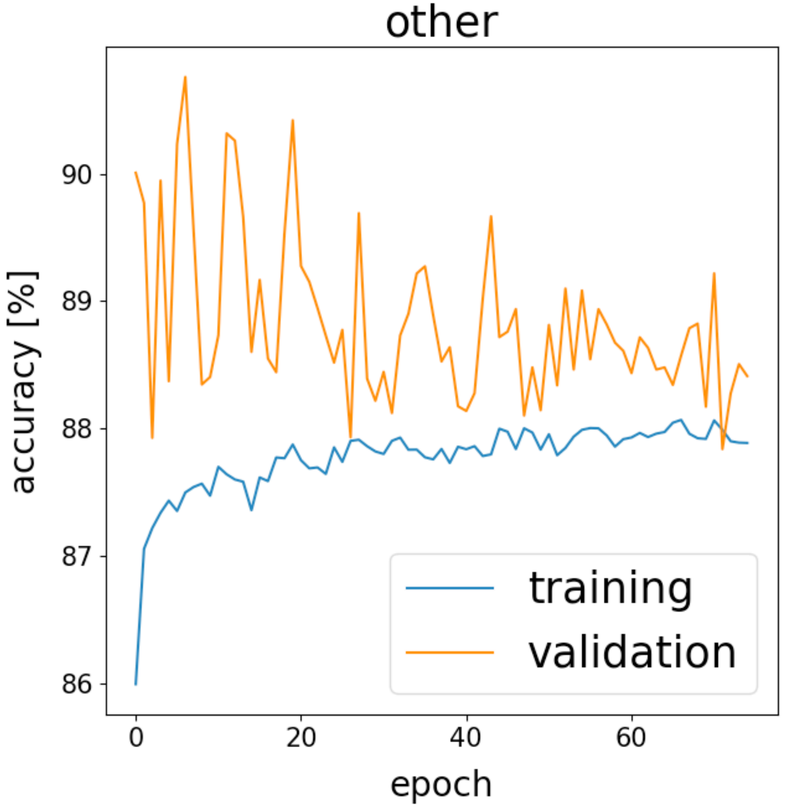




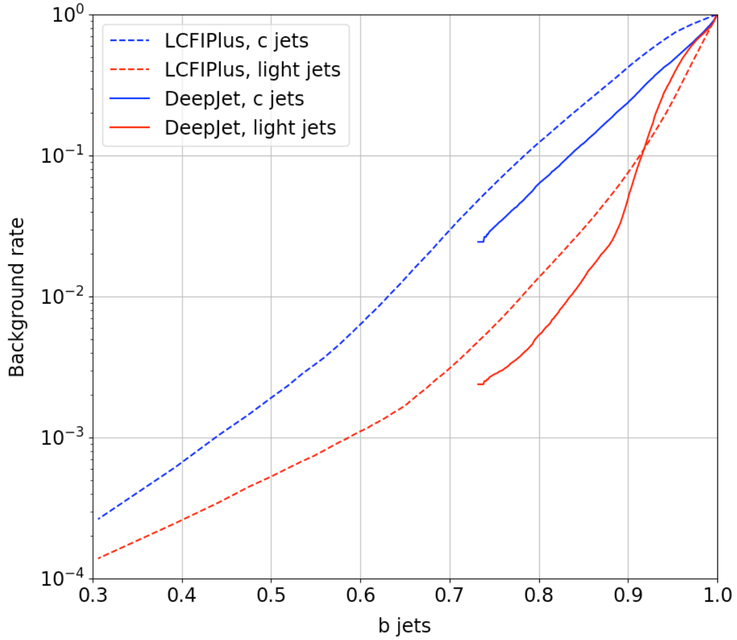
Accuracies

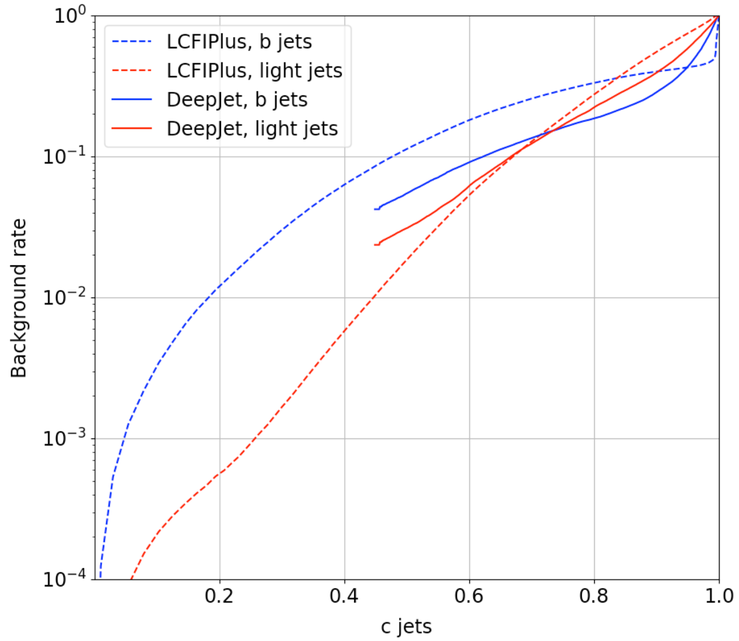

b
accuracy [%]
epoch
training
validation

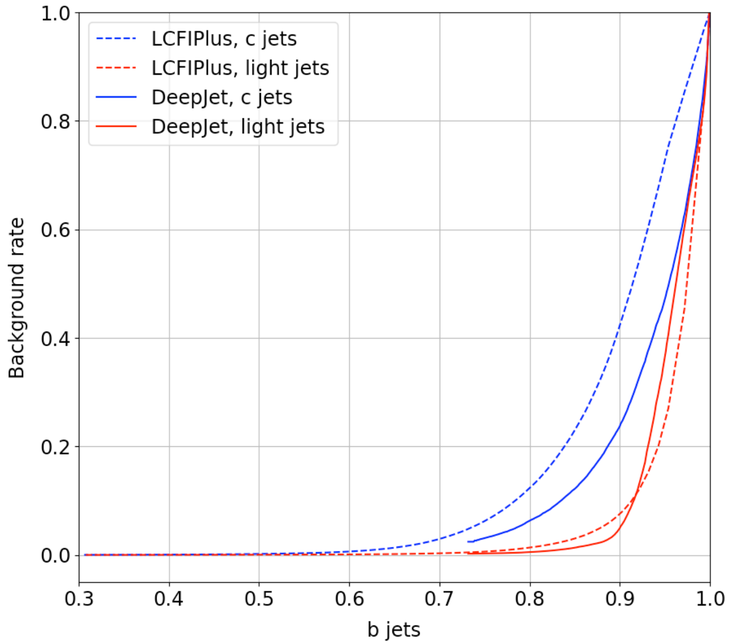

c
accuracy [%]
epoch
training
validation

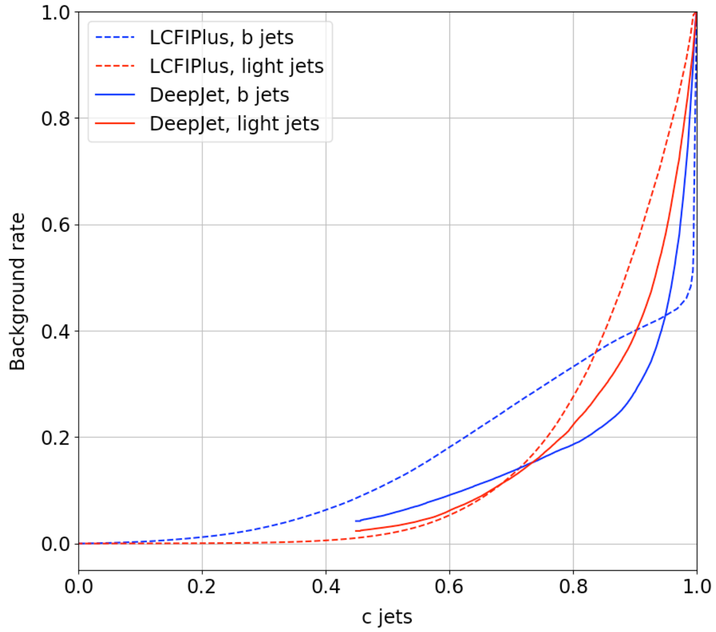

other
accuracy [%]
epoch
training
validation

ROC

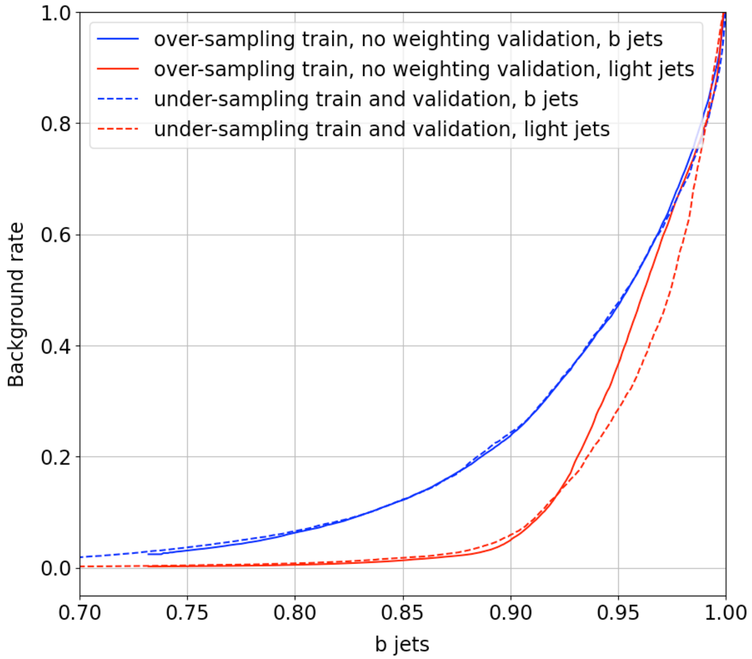

LCFIPlus, c jets
LCFIPlus, light jets
DeepJet, c jets
DeepJet, light jets
Background rate
b jets

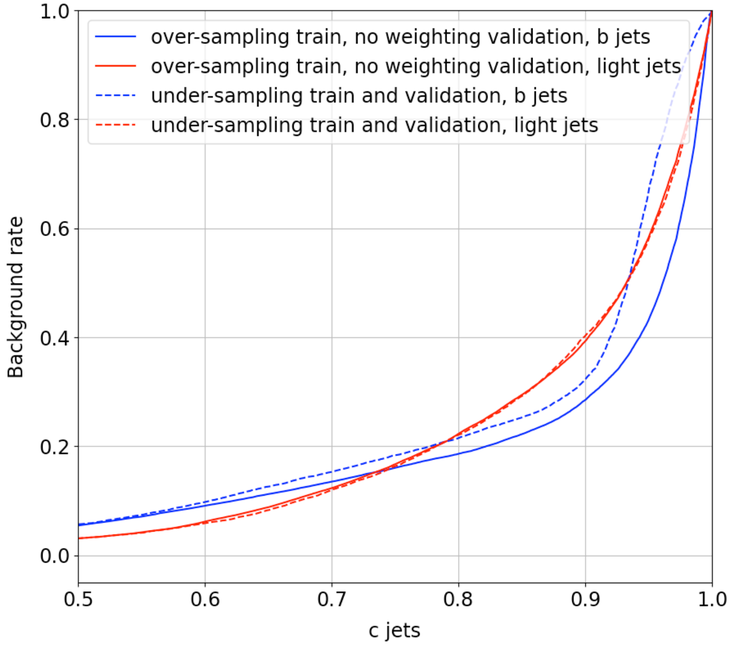

LCFIPlus, b jets
LCFIPlus, light jets
DeepJet, b jets
DeepJet, light jets
Background rate
c jets

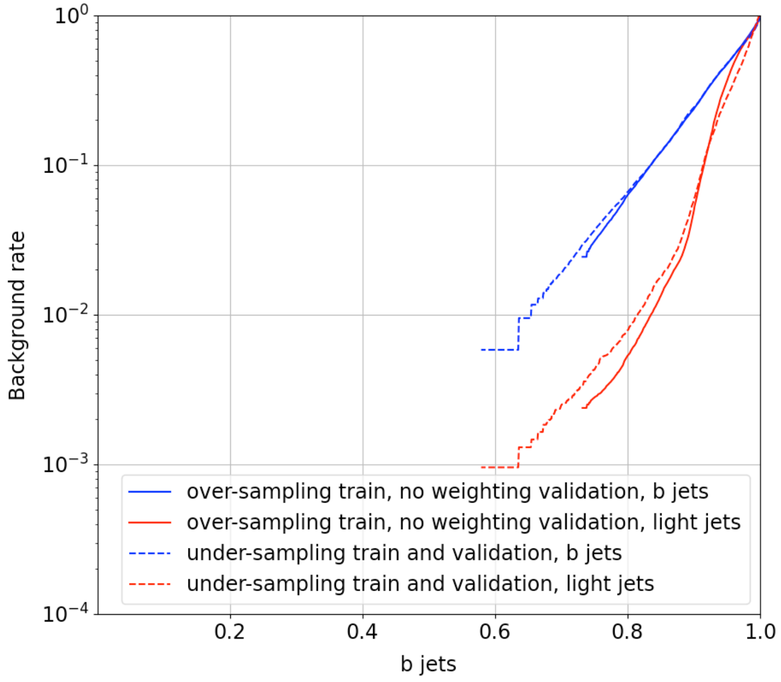

LCFIPlus, c jets
LCFIPlus, light jets
DeepJet, c jets
DeepJet, light jets
Background rate
b jets

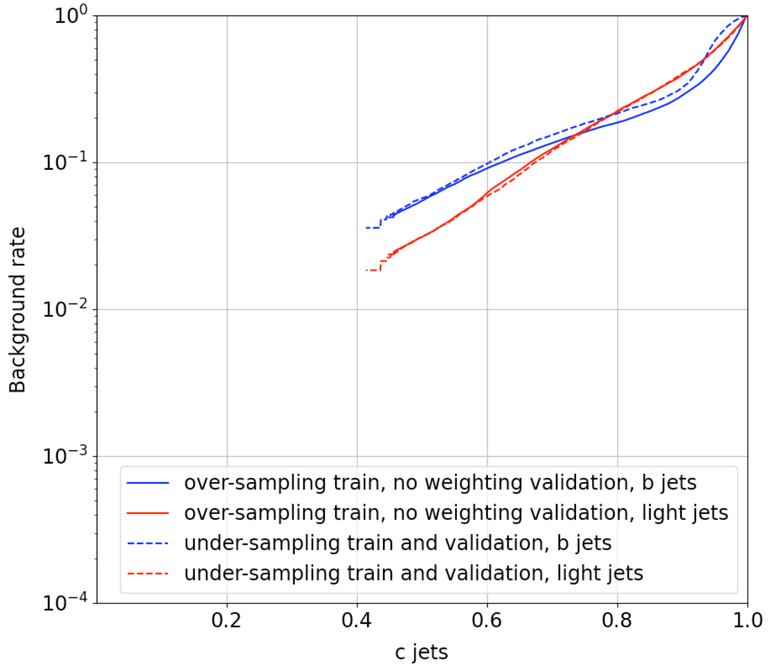

LCFIPlus, b jets
LCFIPlus, light jets
DeepJet, b jets
DeepJet, light jets
Background rate
c jets

ROC


over-sampling train, no weighting validation, b jets
over-sampling train, no weighting validation, light jets
under-sampling train and validation, b jets
under-sampling train and validation, light jets
Background rate
b jets

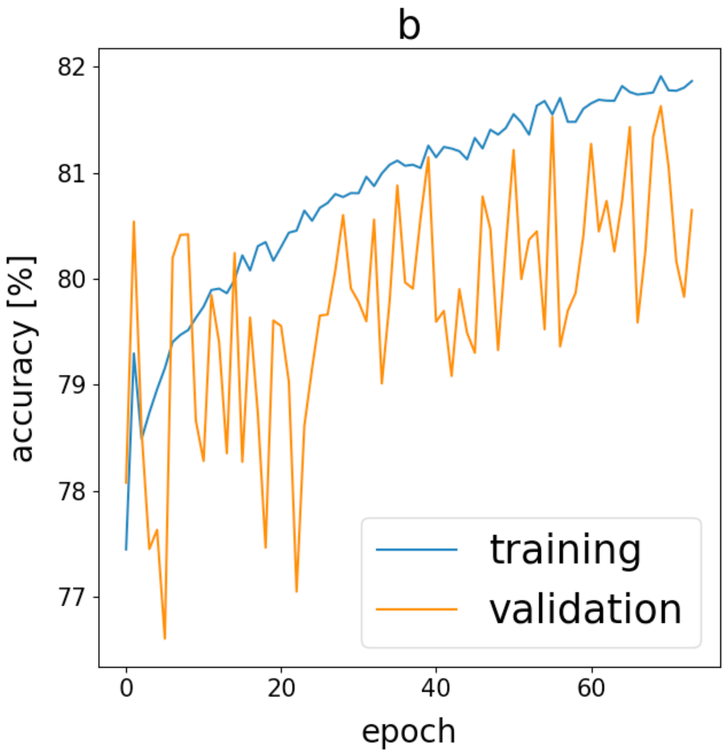

over-sampling train, no weighting validation, b jets
over-sampling train, no weighting validation, light jets
under-sampling train and validation, b jets
under-sampling train and validation, light jets
Background rate
c jets

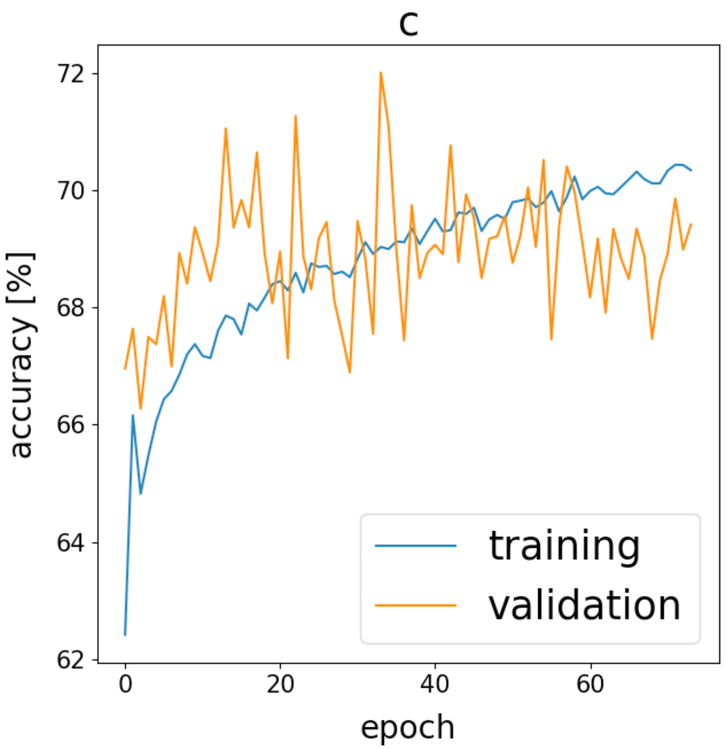

over-sampling train, no weighting validation, b jets
over-sampling train, no weighting validation, light jets
under-sampling train and validation, b jets
under-sampling train and validation, light jets
Background rate
b jets

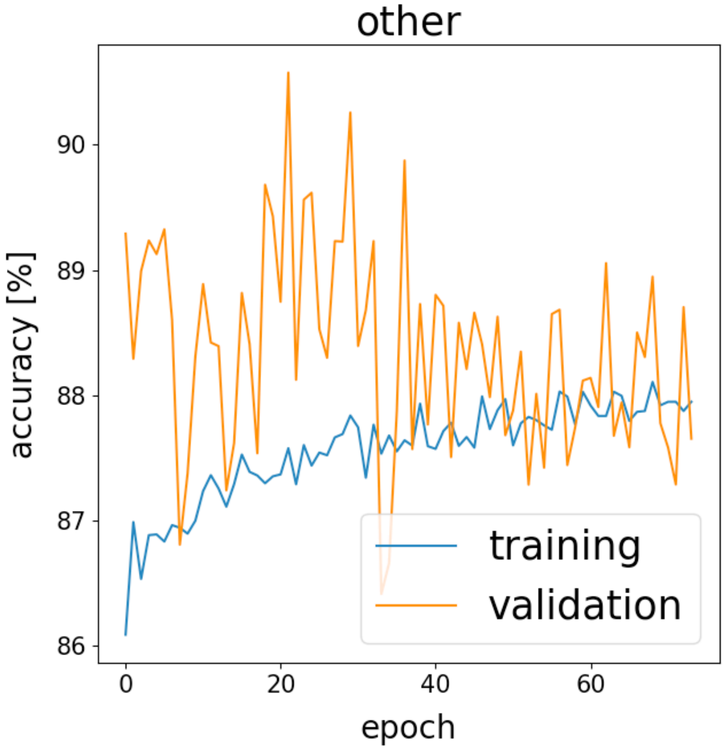

over-sampling train, no weighting validation, b jets
over-sampling train, no weighting validation, light jets
under-sampling train and validation, b jets
under-sampling train and validation, light jets
Background rate
c jets

Backup

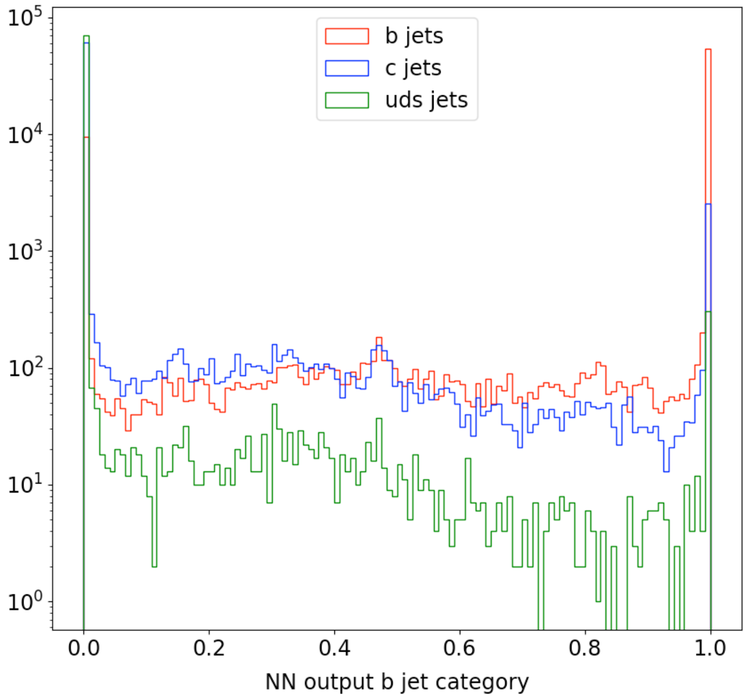
Accuracies in the different categories

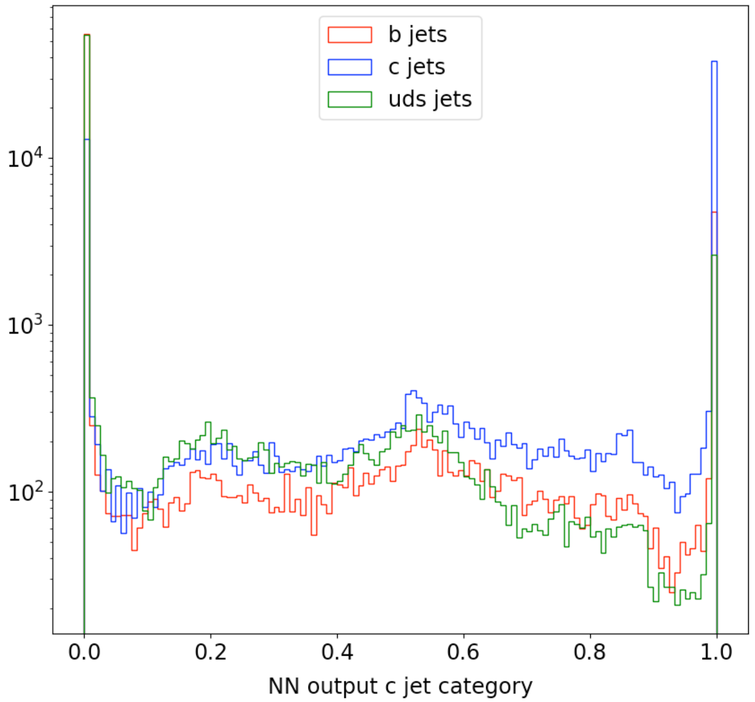


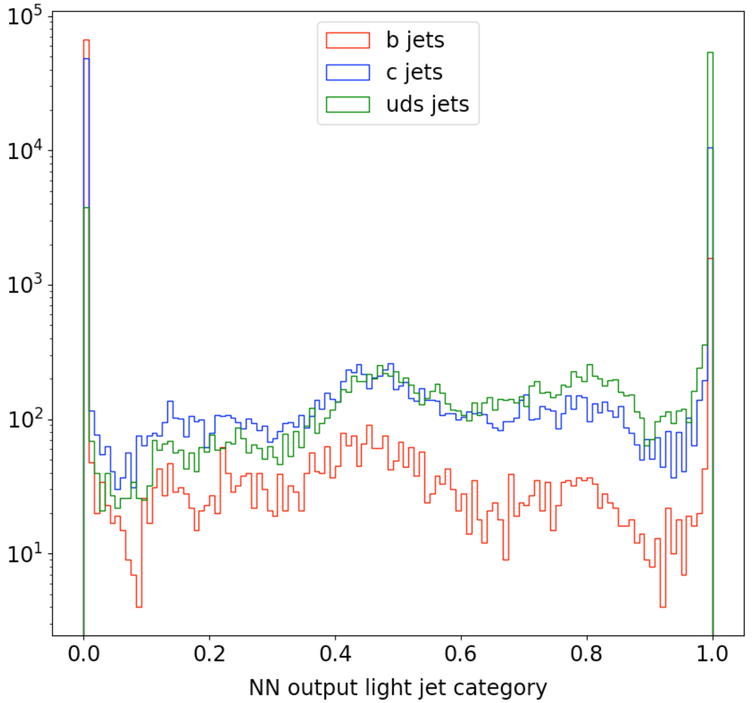




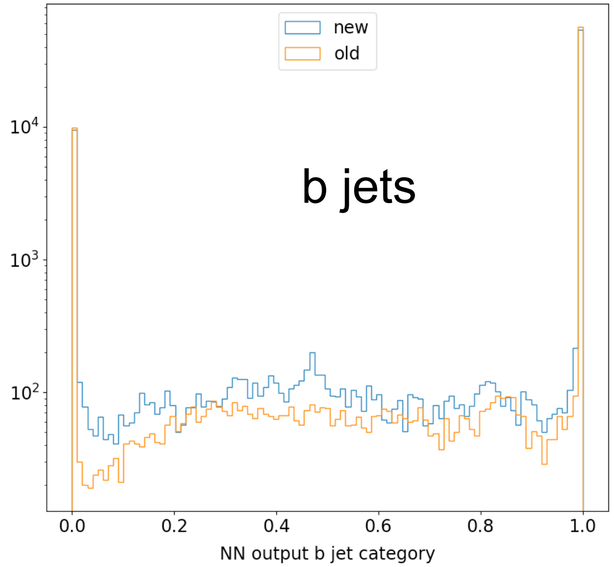
NN output

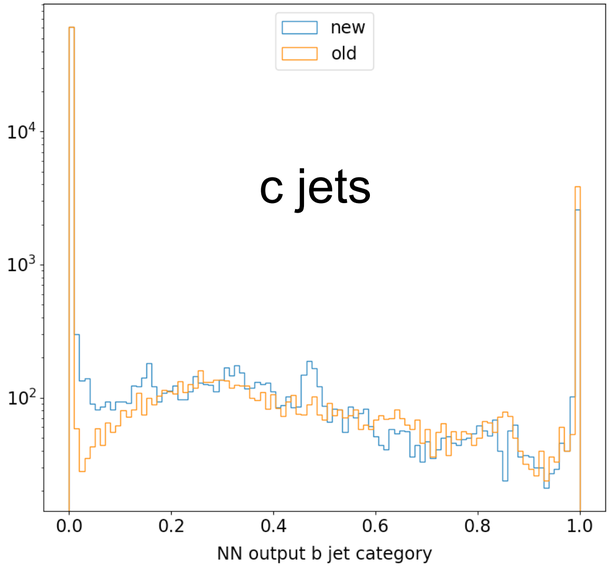

b jets
c jets
uds jets
NN output b jet category

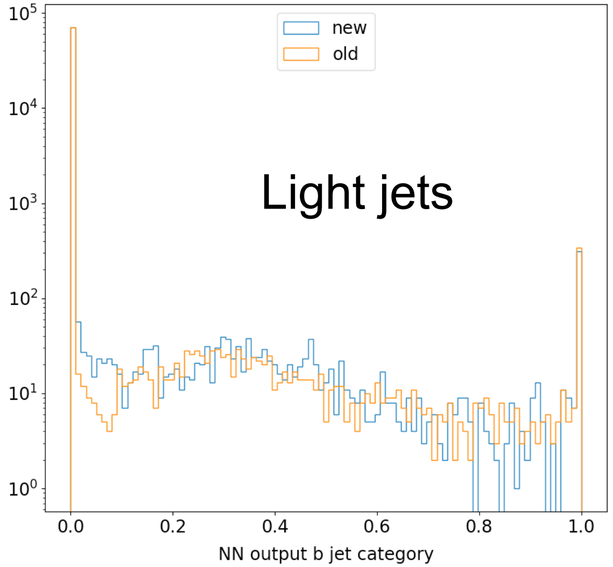

b jets
c jets
uds jets
NN output c jet category

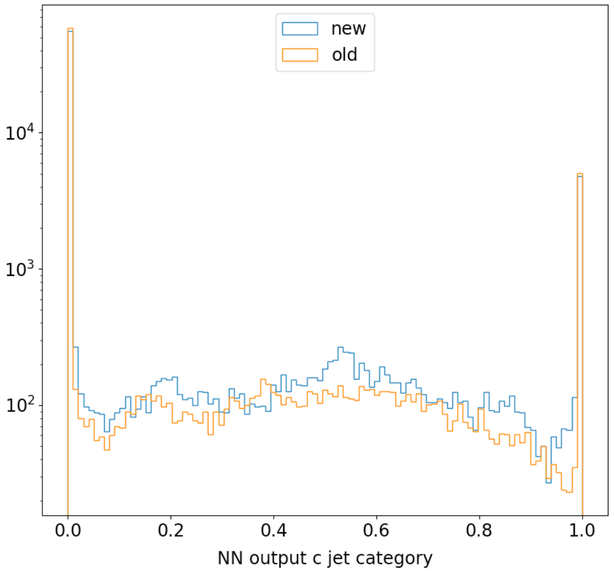

b jets
c jets
uds jets
NN output light jet category

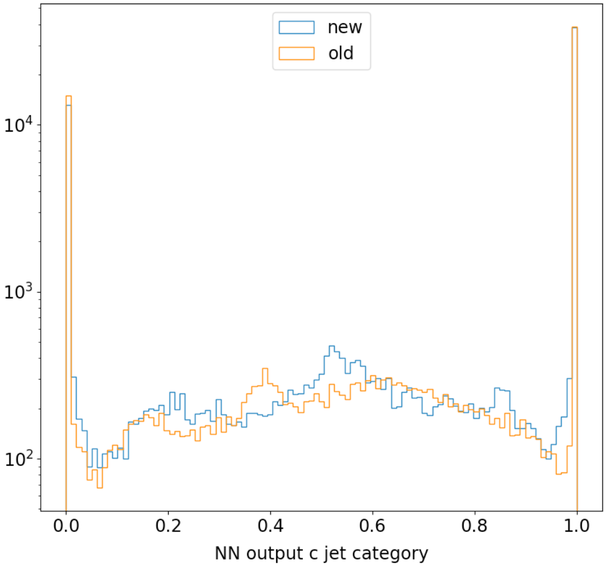
Comparison to results with bug: NN output

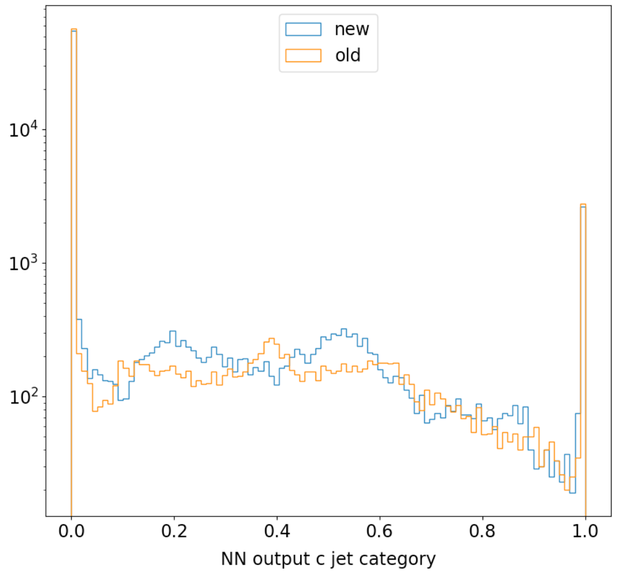




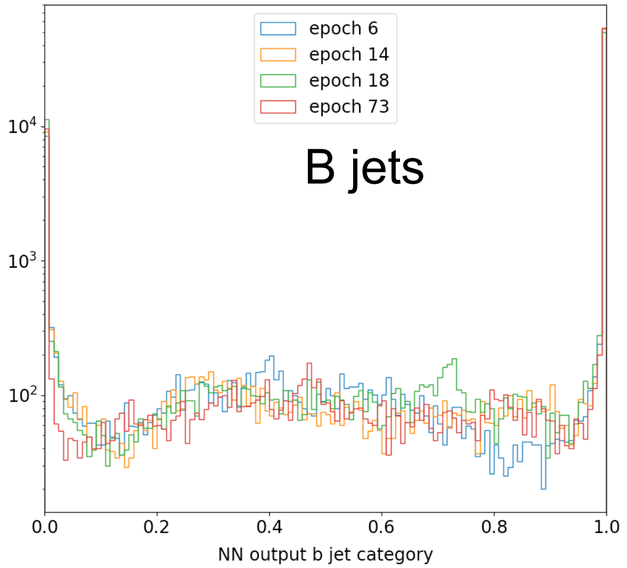


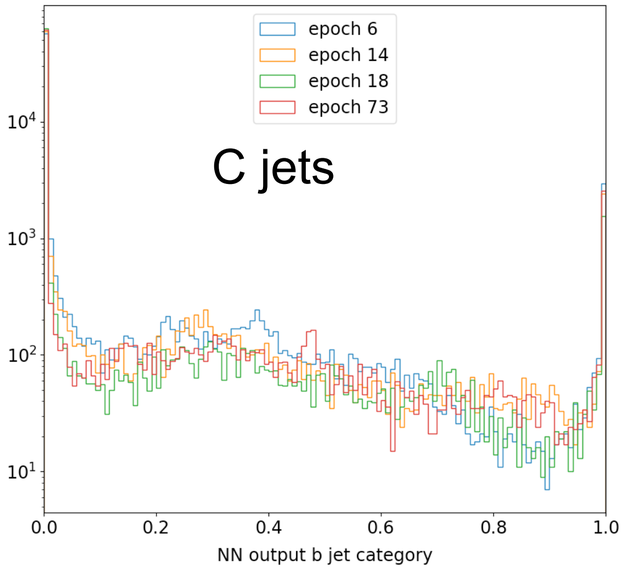


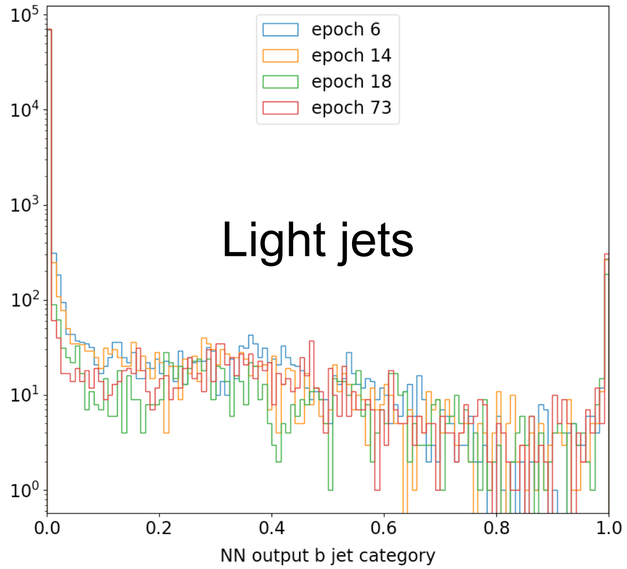


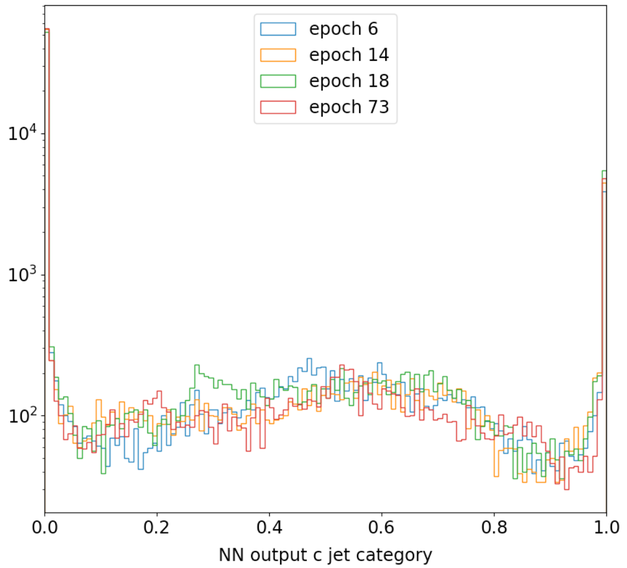


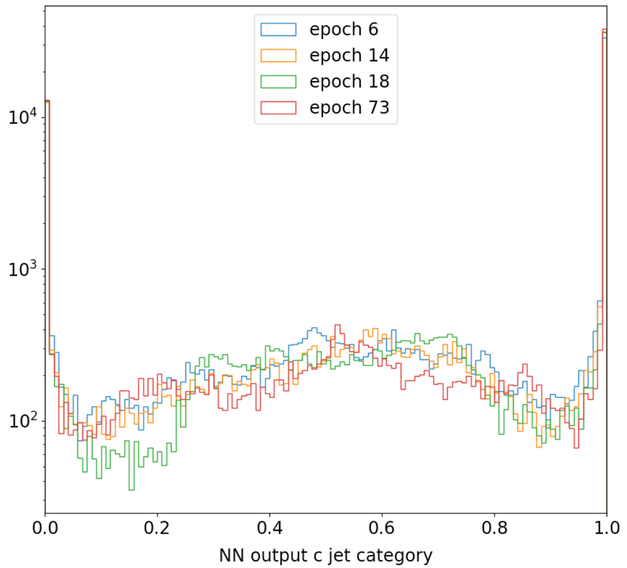
Comparison NN output for different epochs

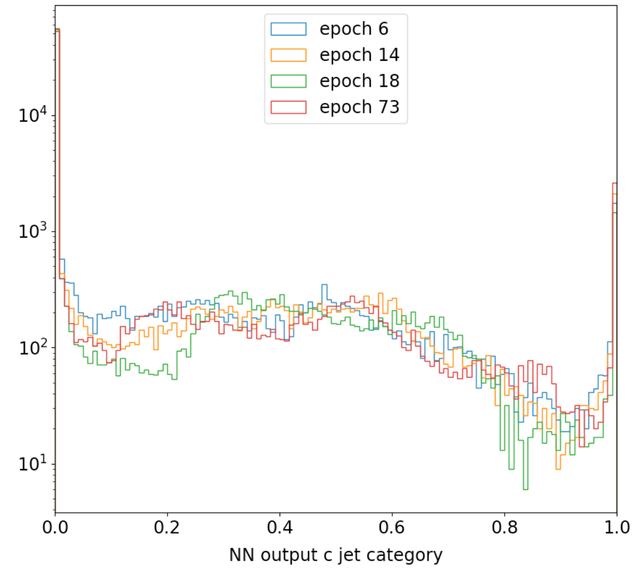




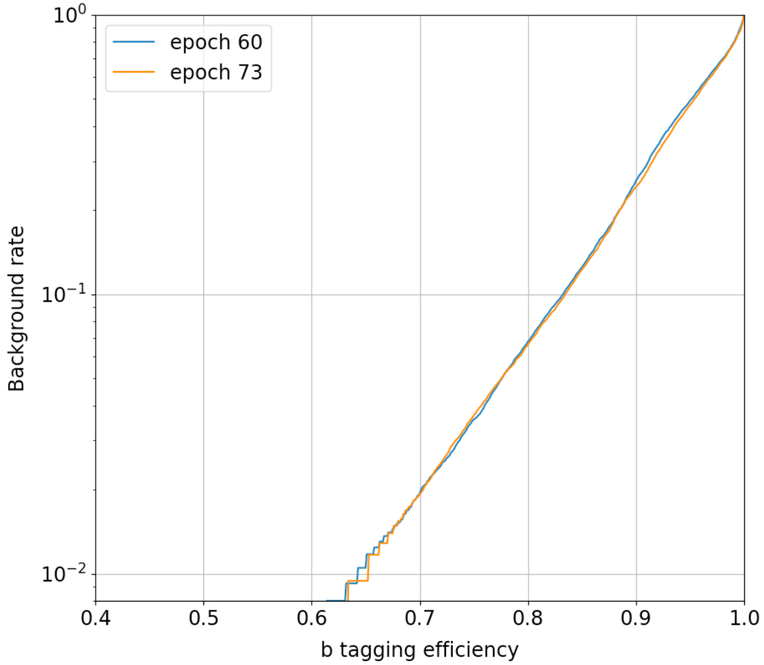


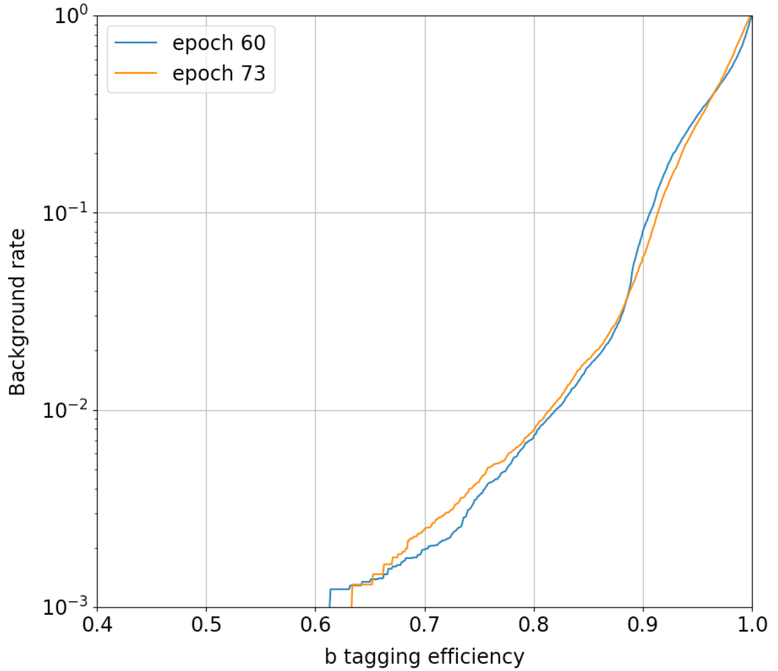


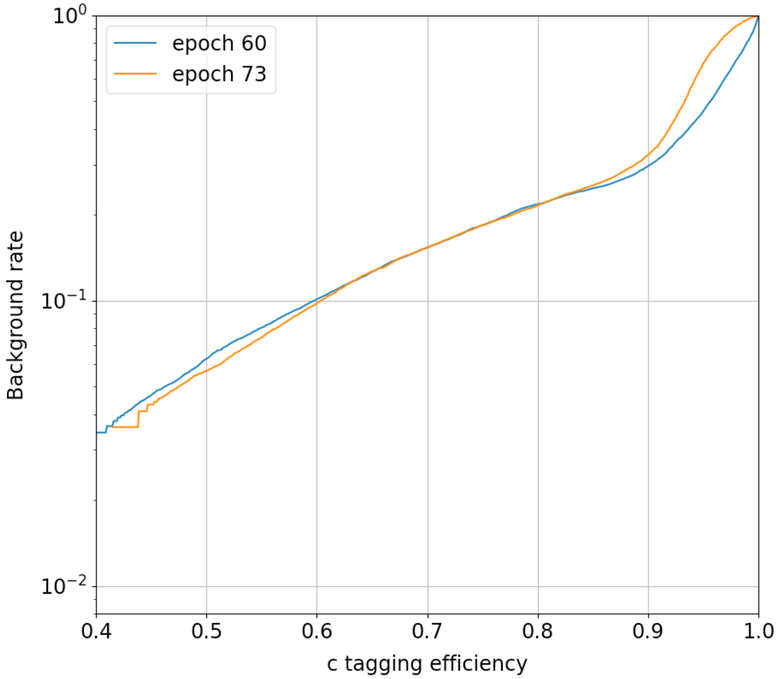


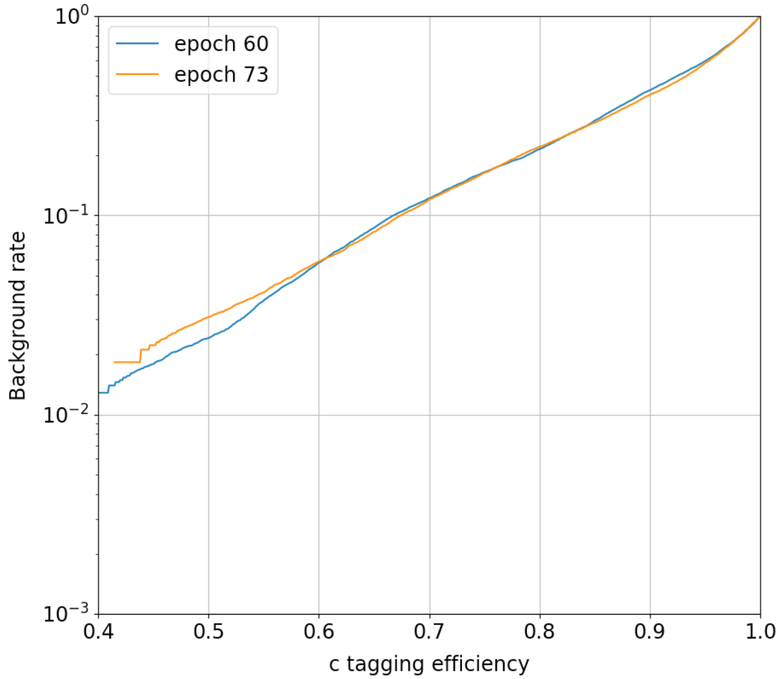


ROC comparison

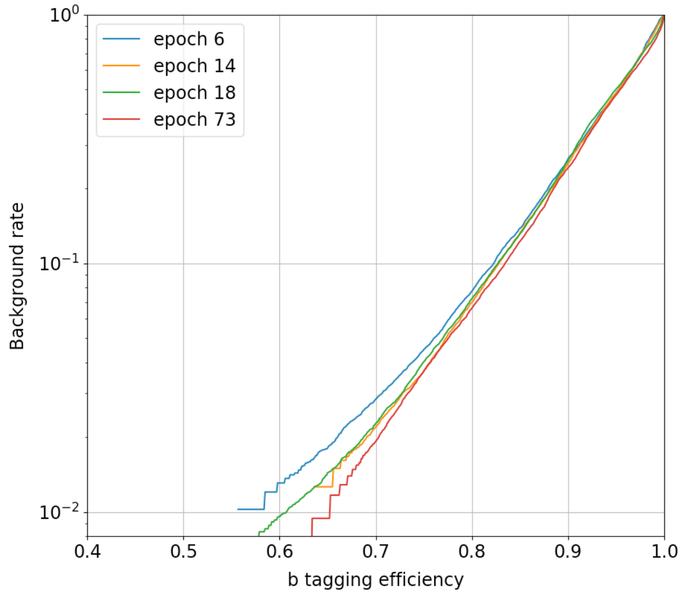


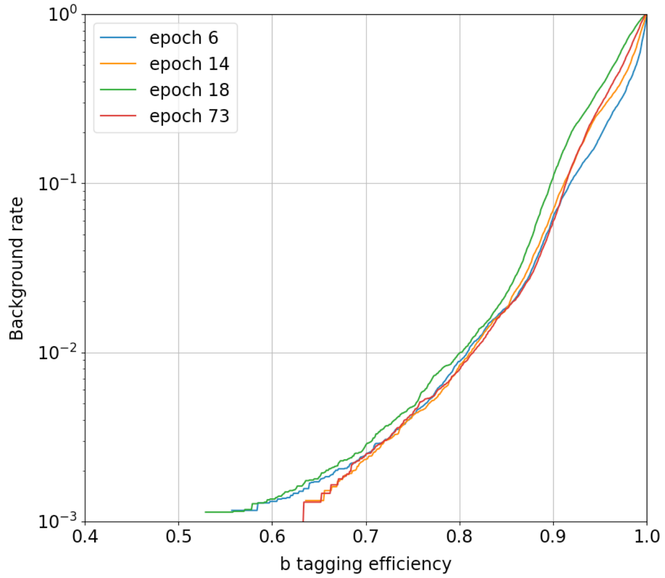


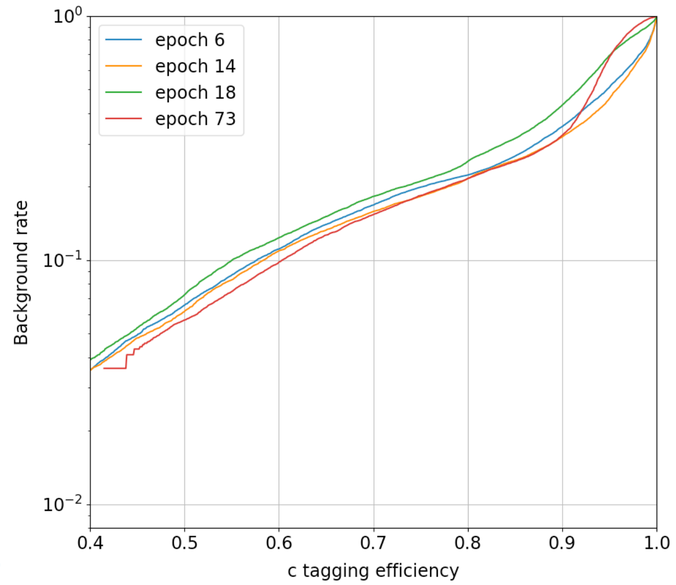


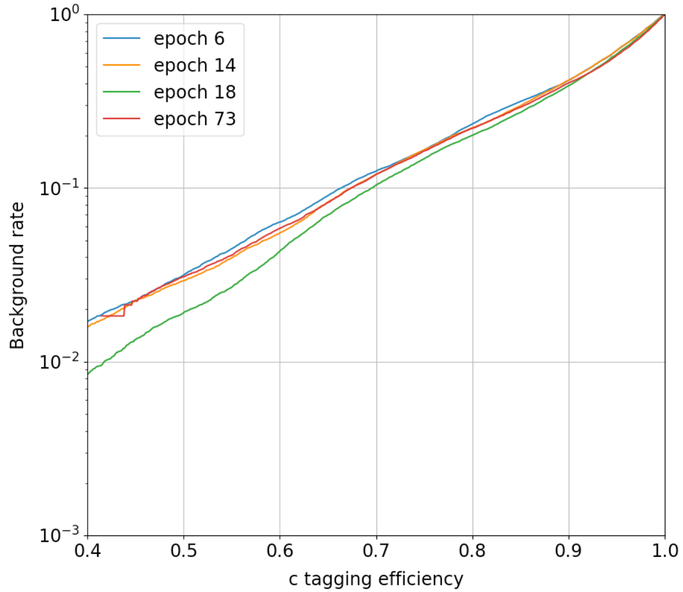


Comparison ROCs different epochs

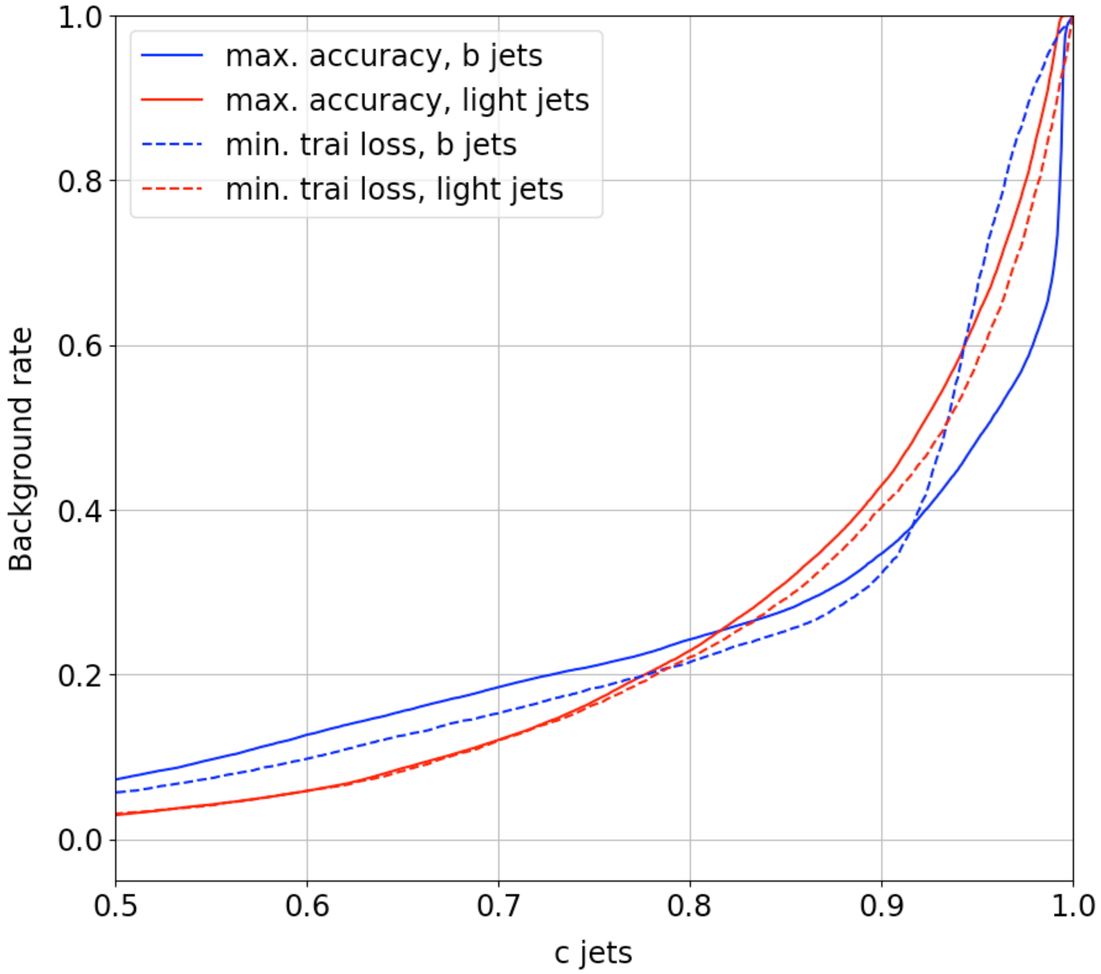


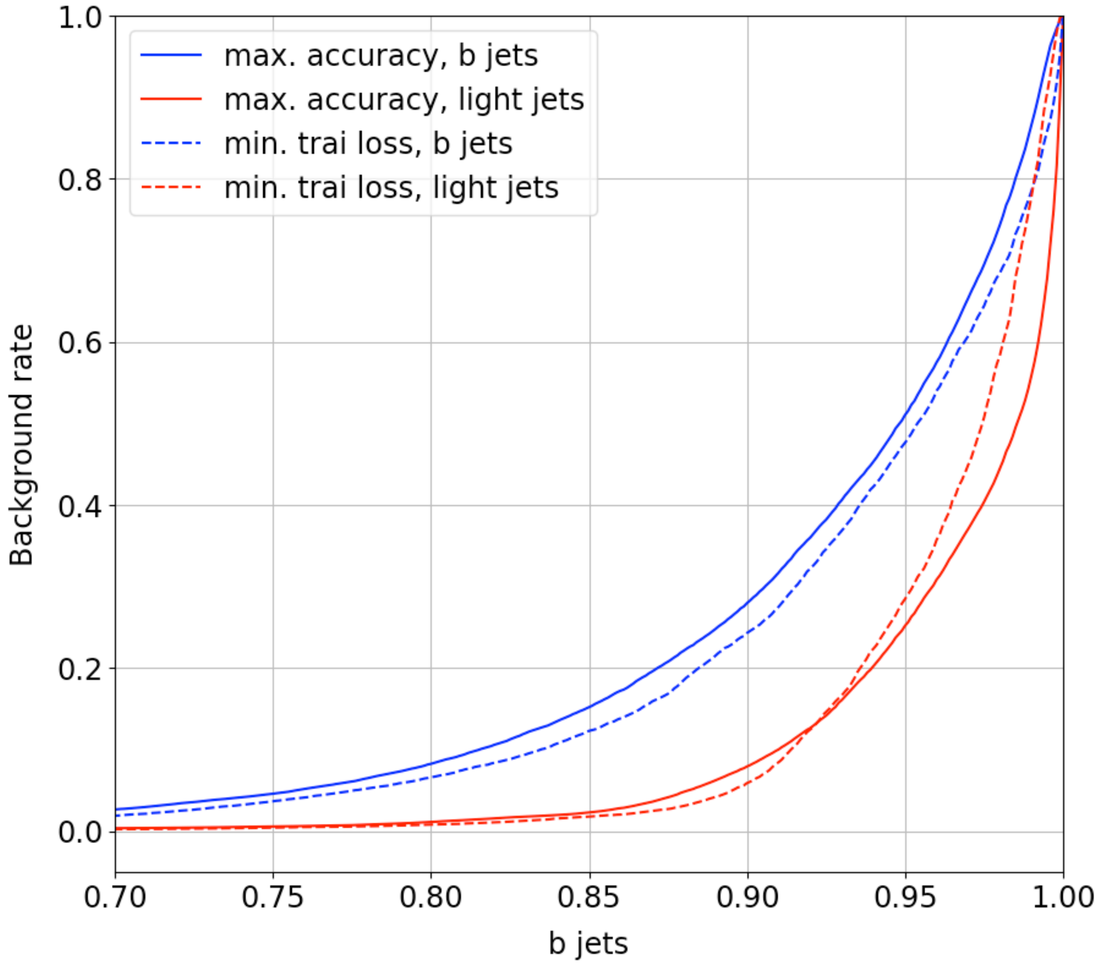




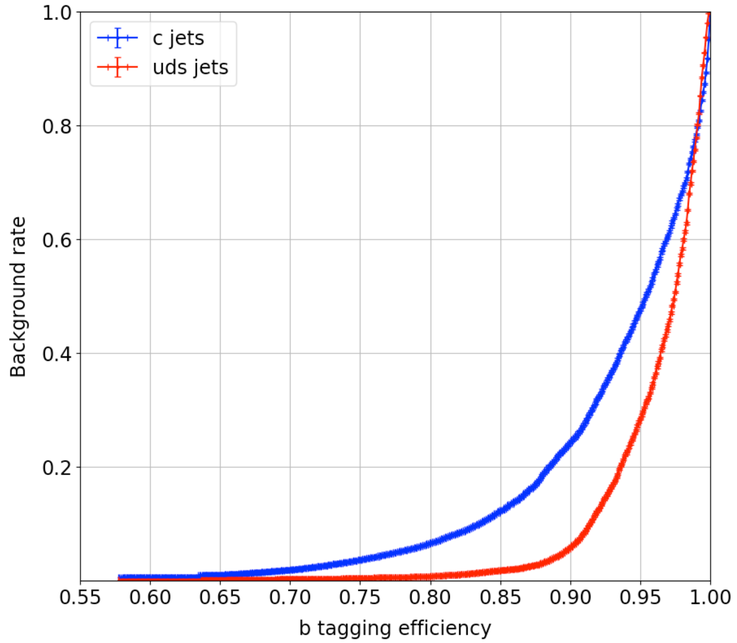


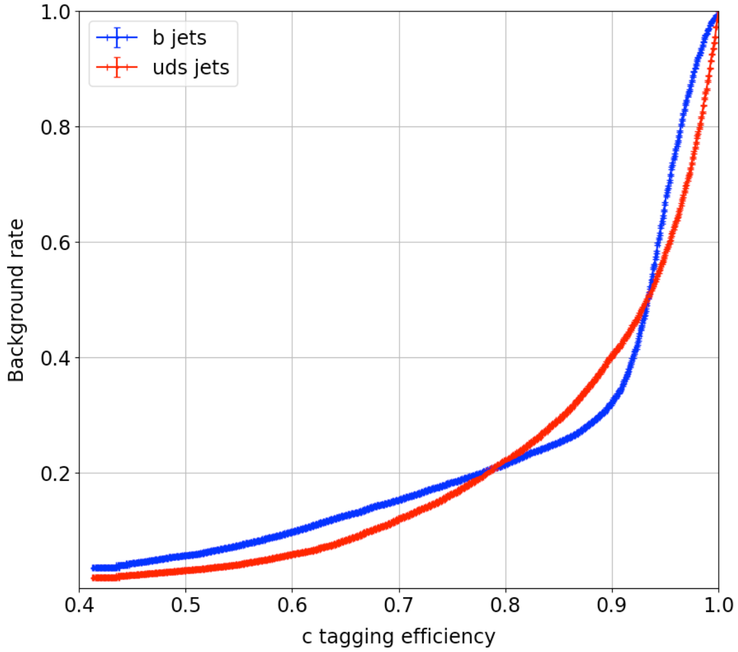
Min train loss vs. max accuracy

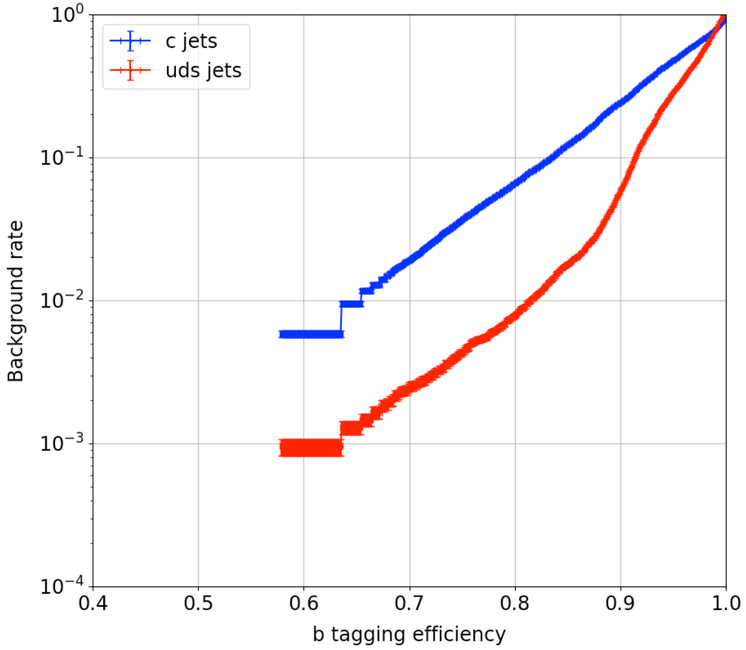


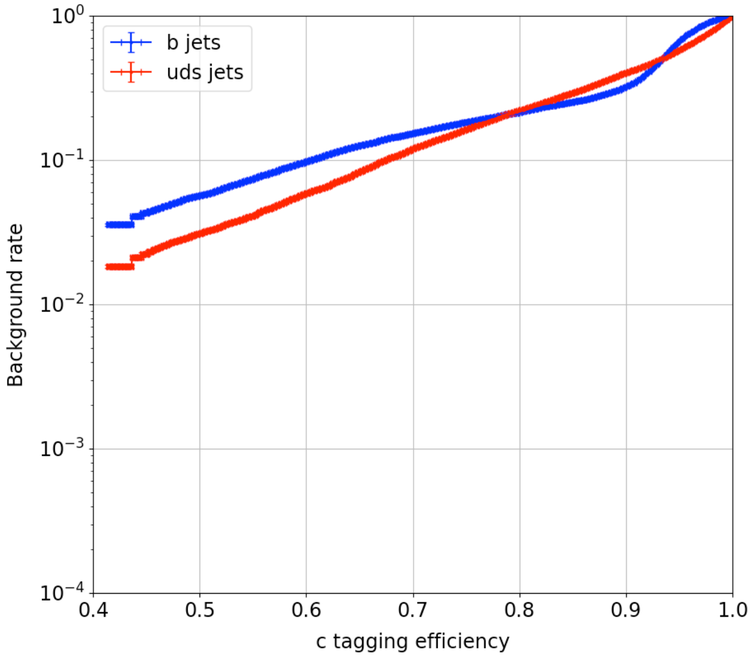


ROCs

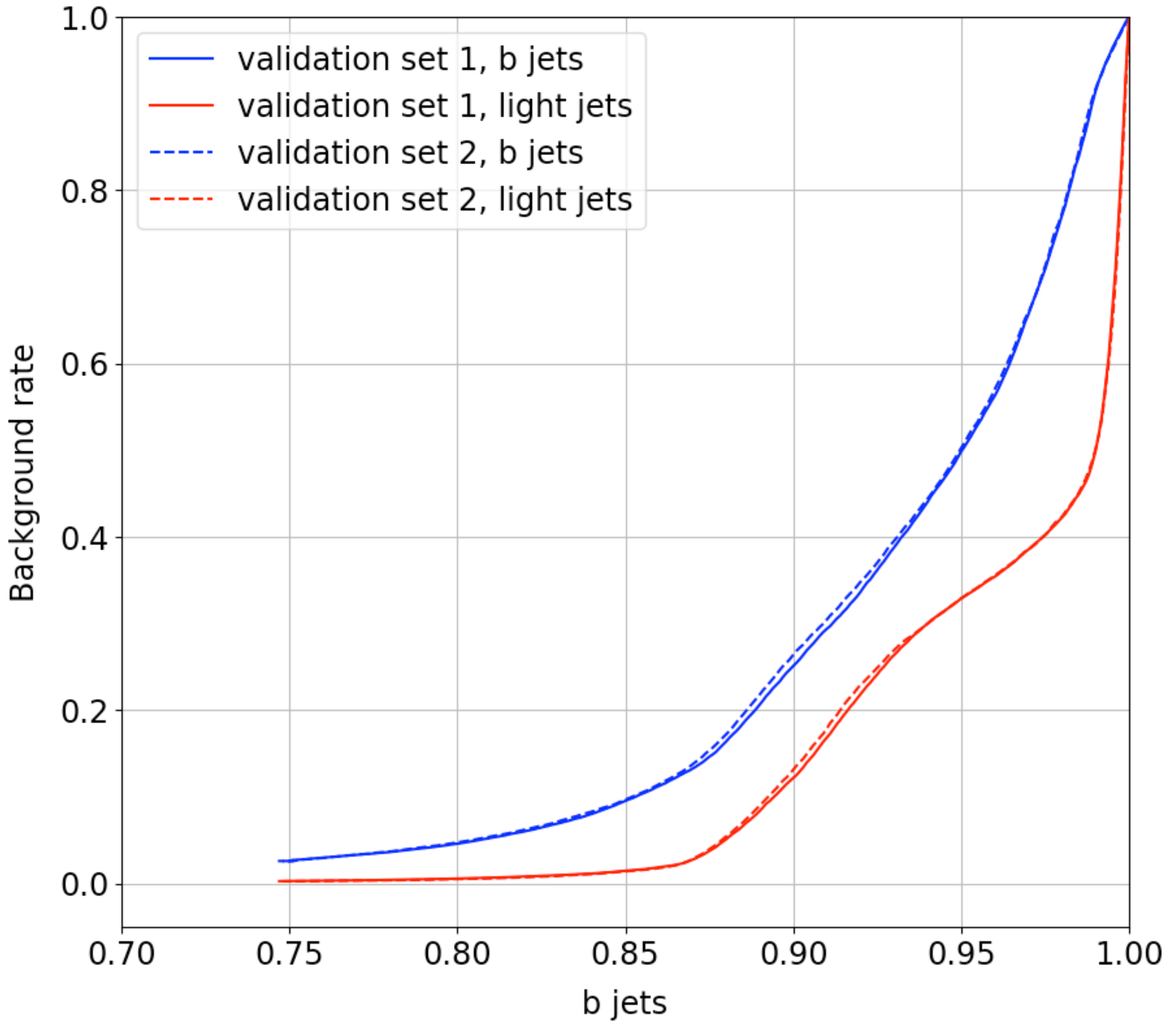

c jets
uds jets
Background rate
b tagging efficiency

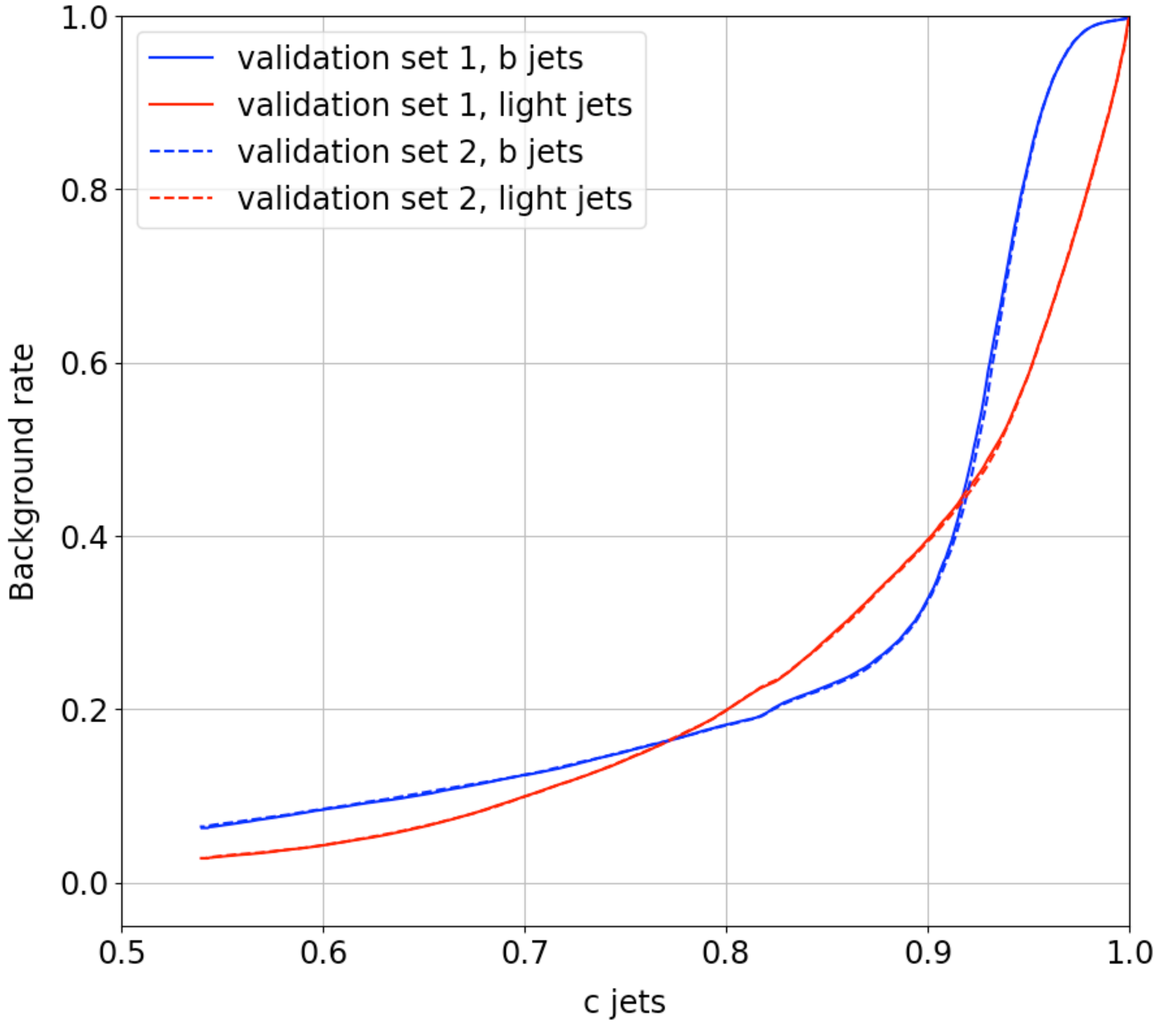

b jets
uds jets
Background rate
c tagging efficiency


c jets
uds jets
Background rate
b tagging efficiency

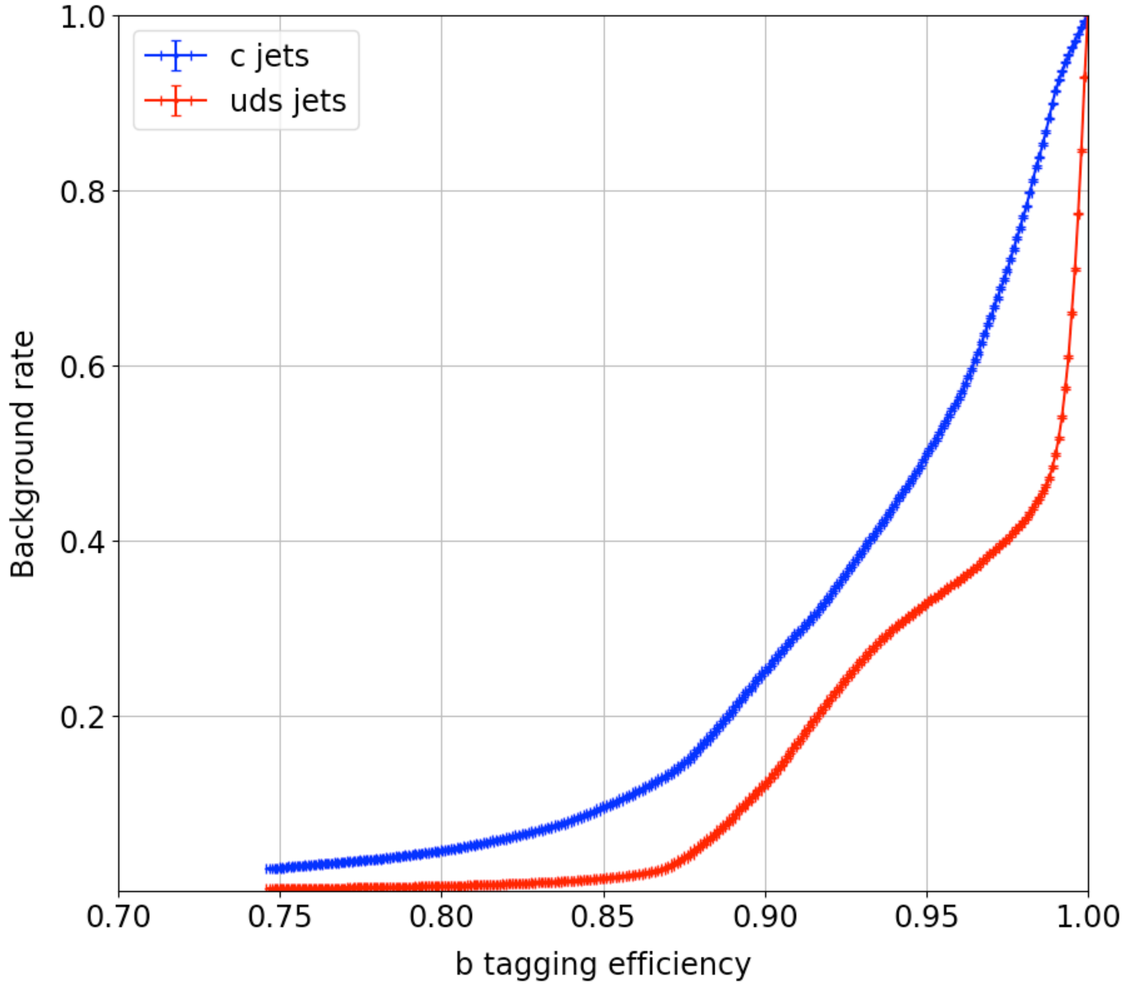

b jets
uds jets
Background rate
c tagging efficiency

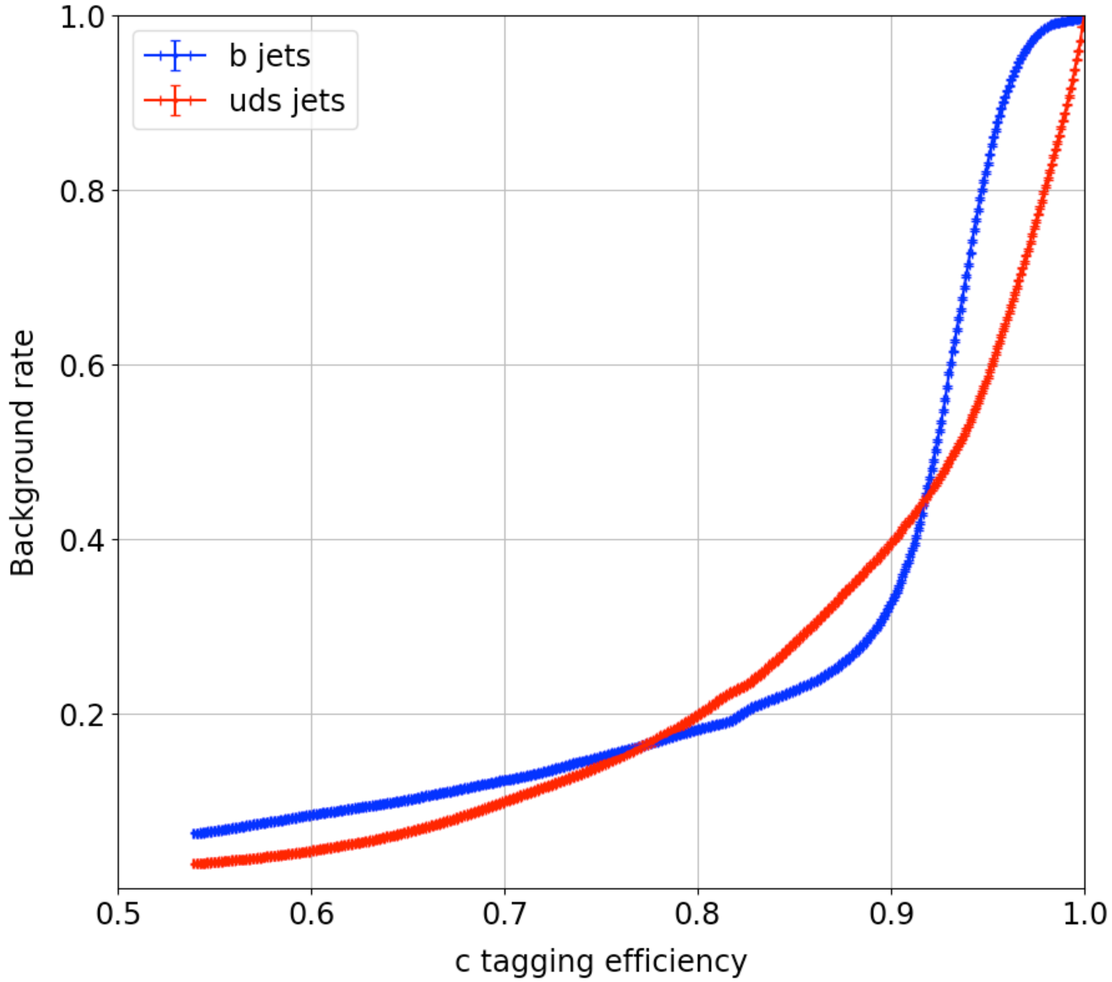
Split validation sample in two



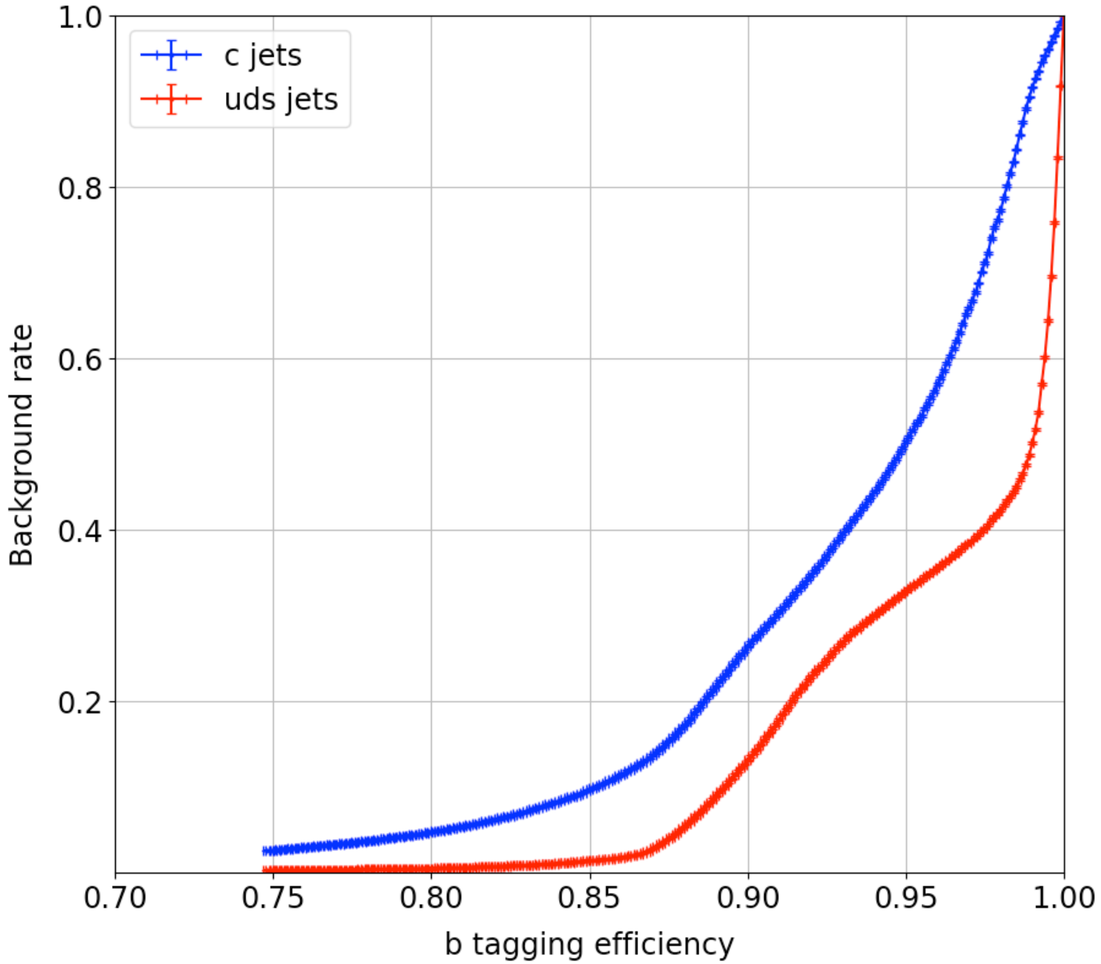


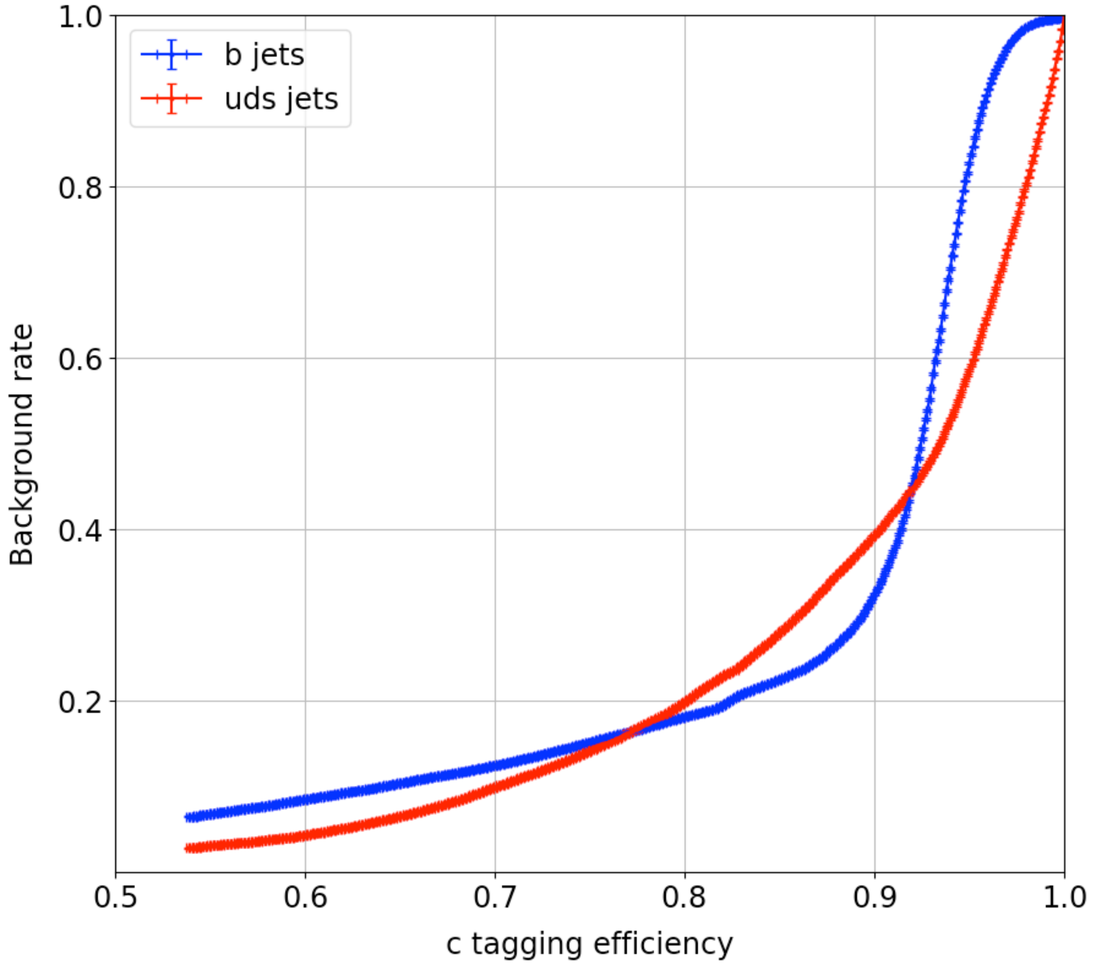
ROC with errors validation set 1



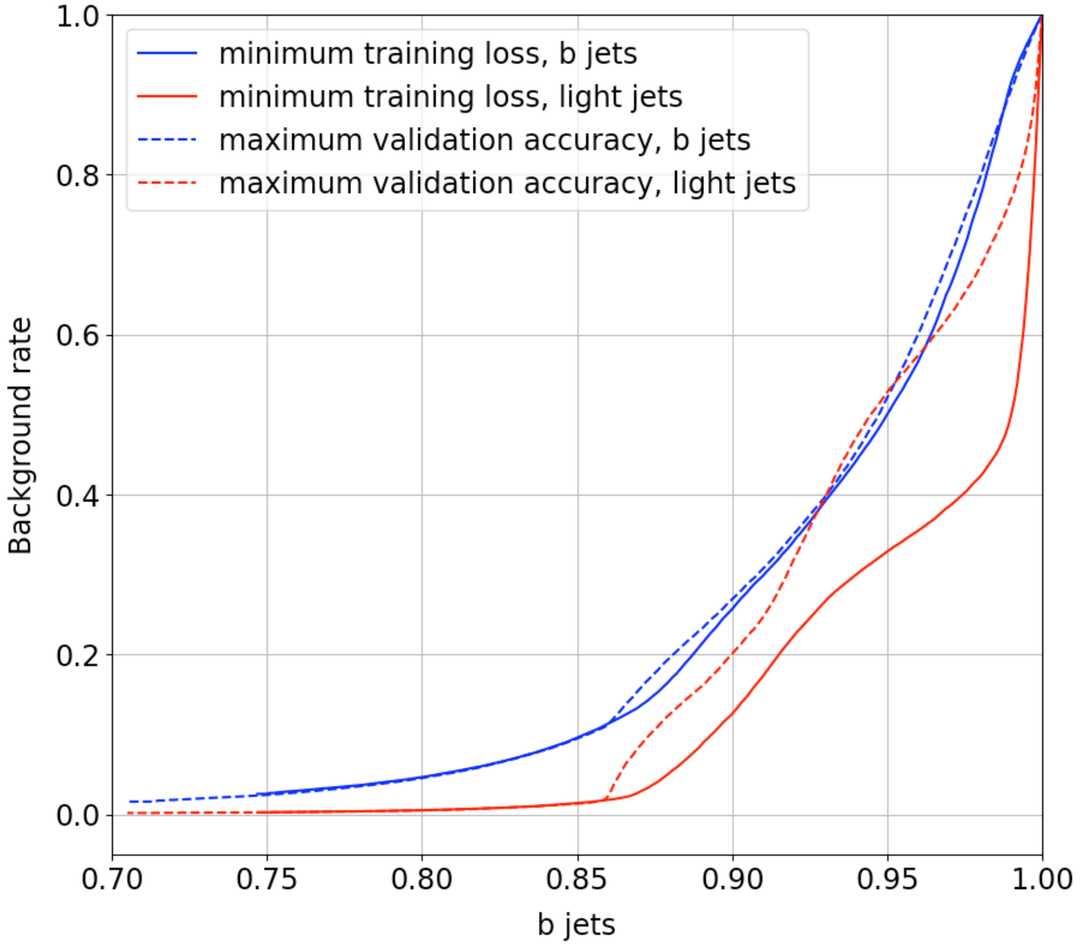


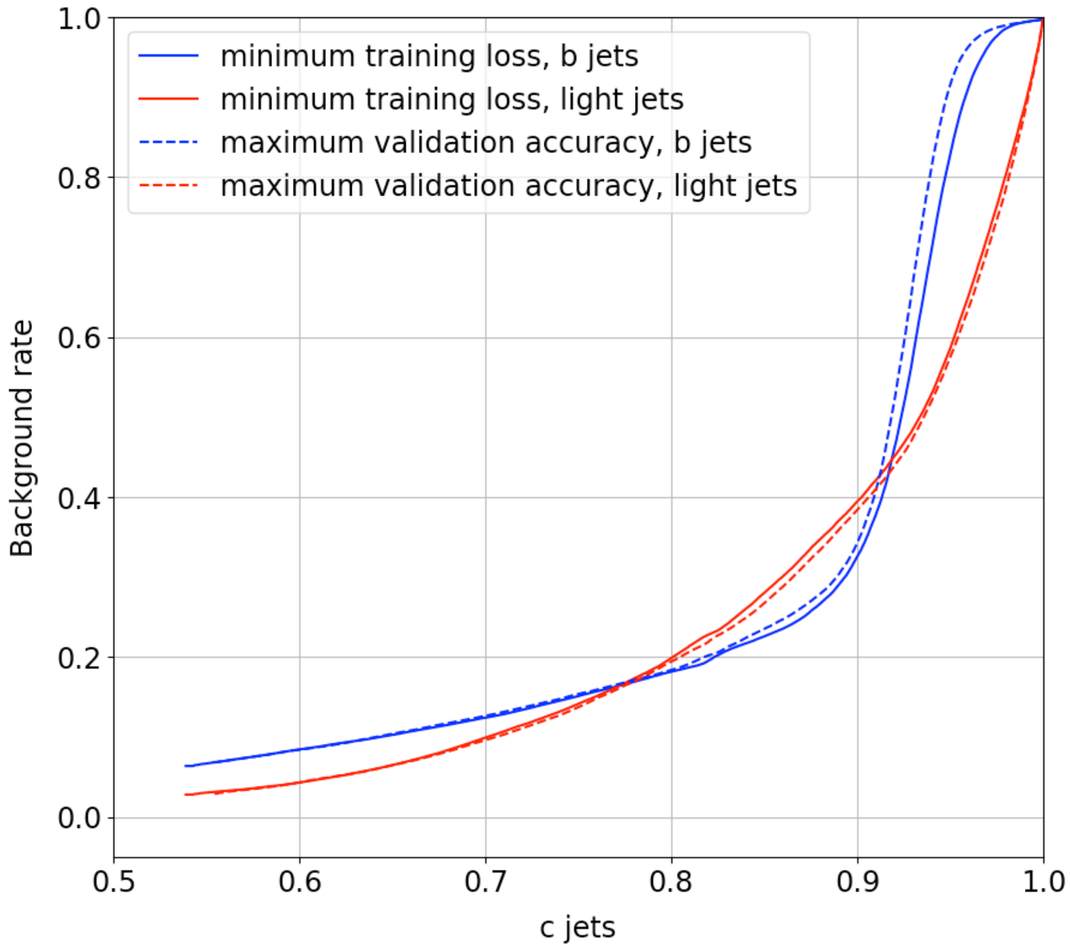
ROC with errors validation set 2



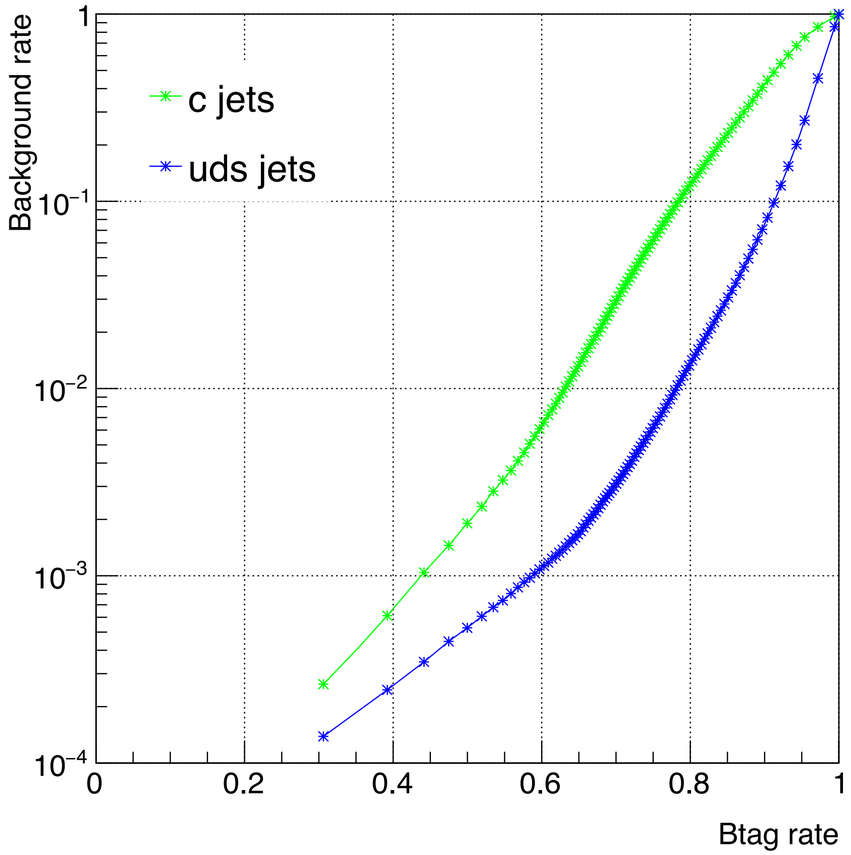


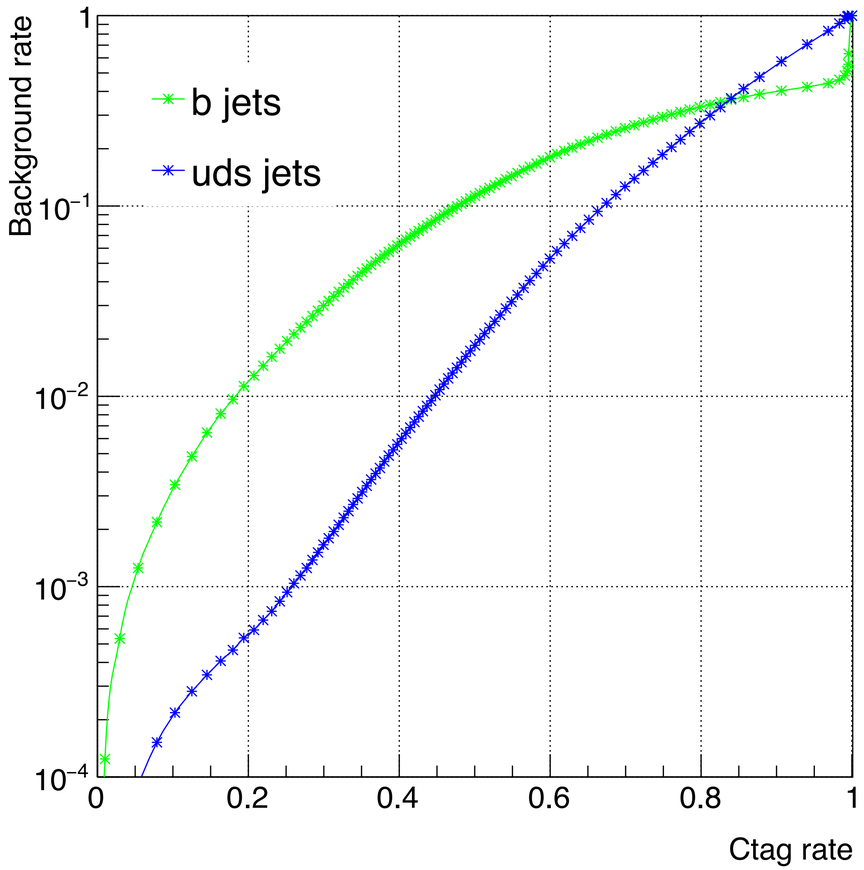
Min. training loss vs. max validation loss



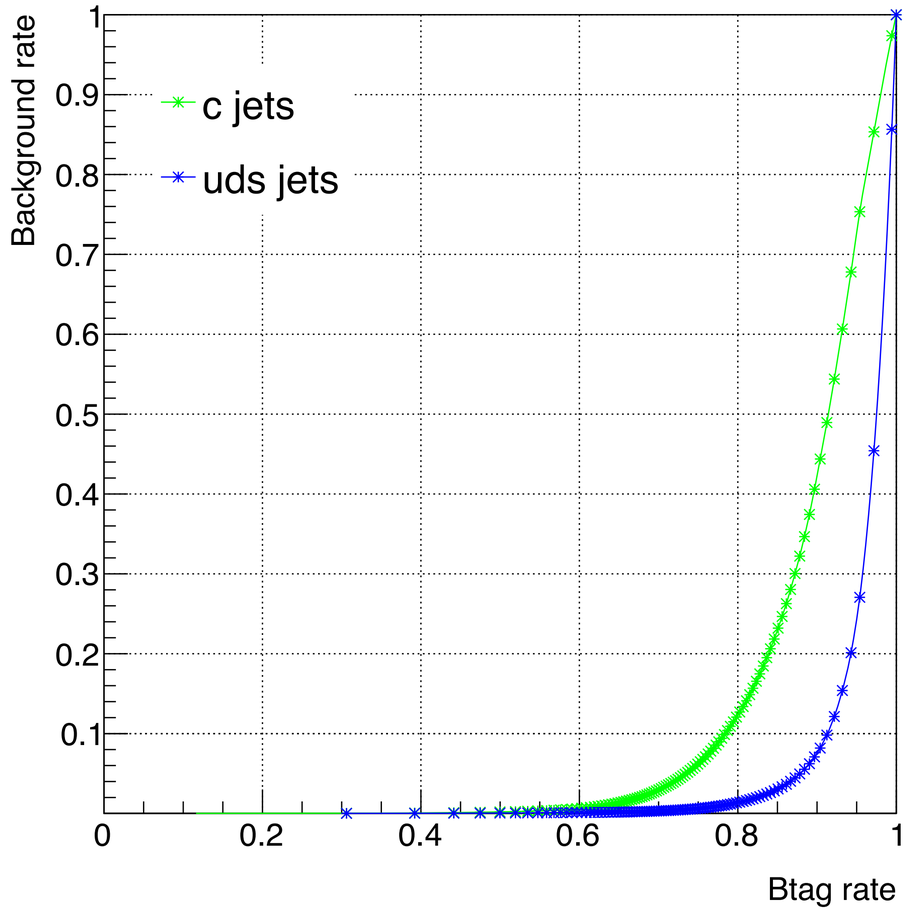


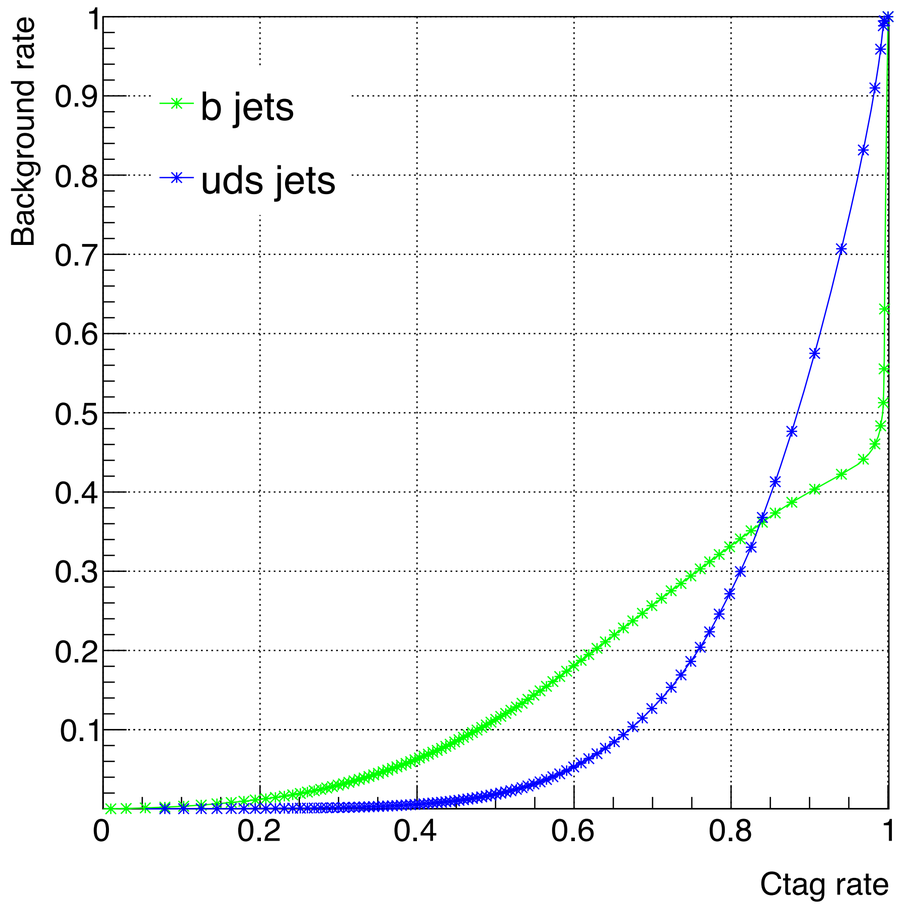
LCFIPlus - ROC from BDT output (100 bins)



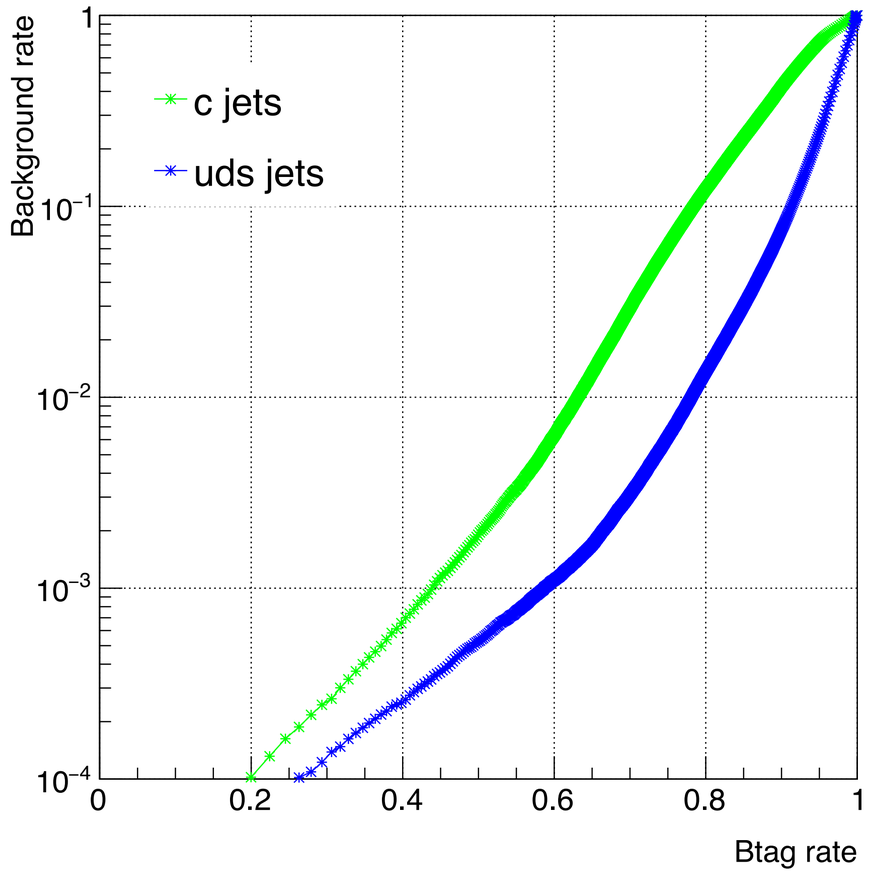


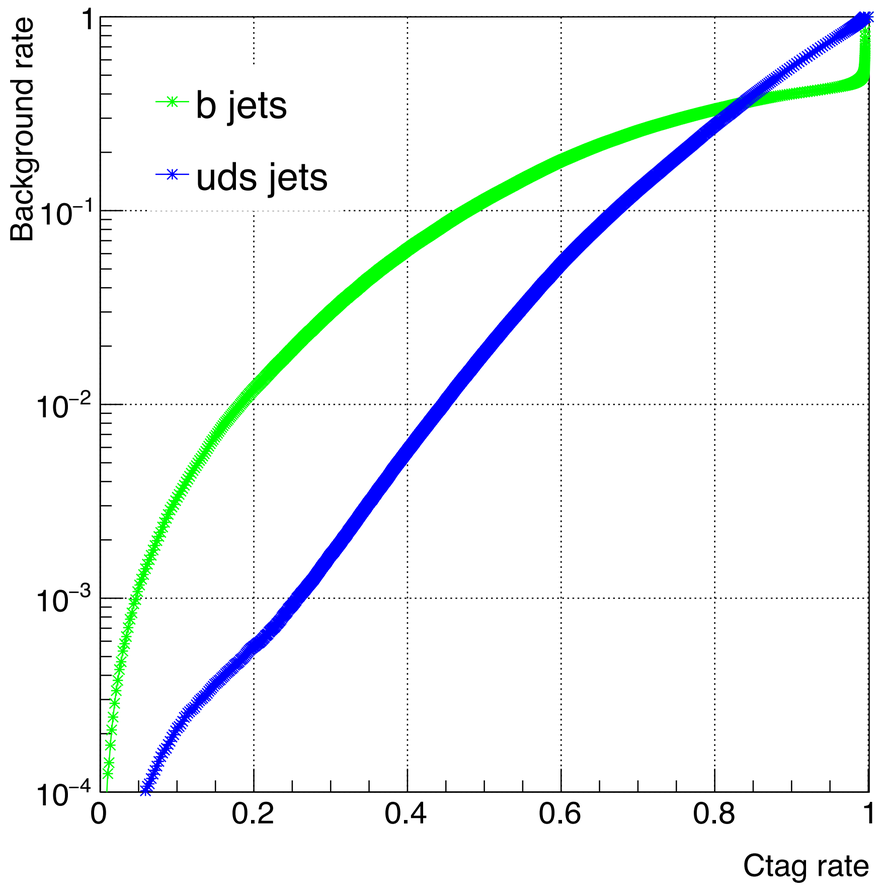
LCFIPlus - ROC from BDT output (100 bins)



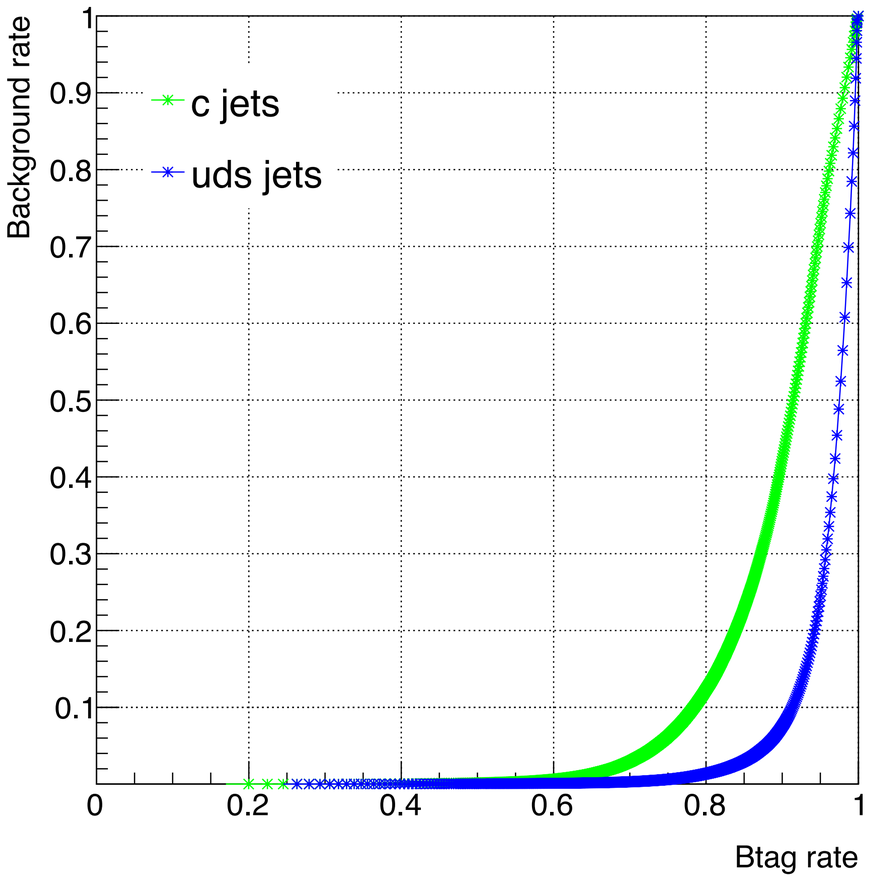


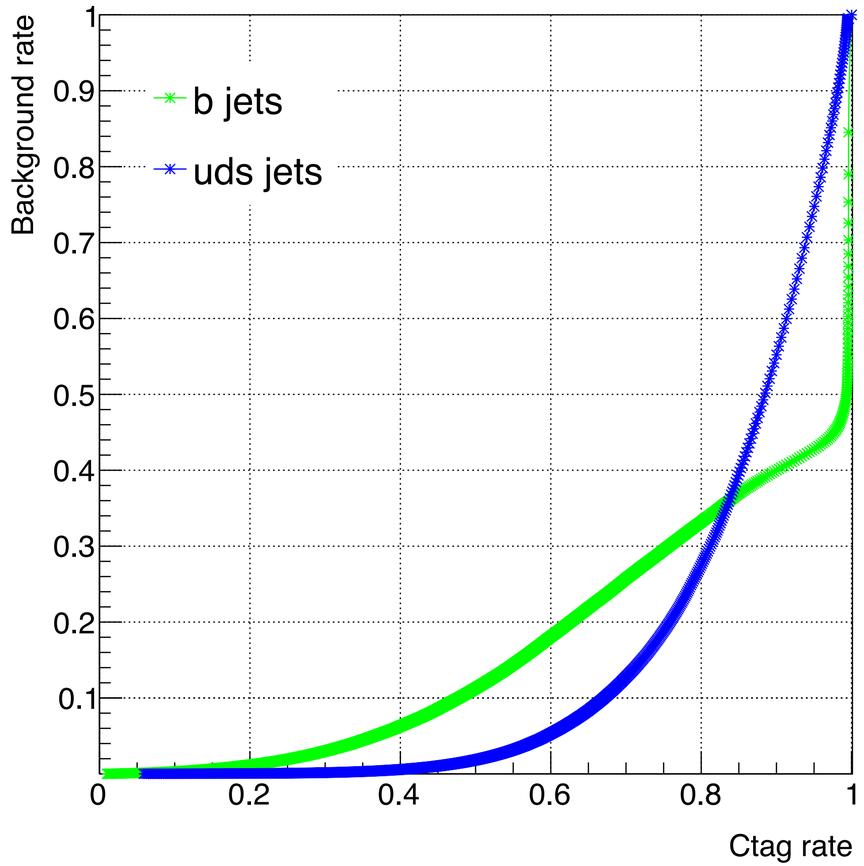
LCFIPlus - ROC from BDT output (1000 bins)



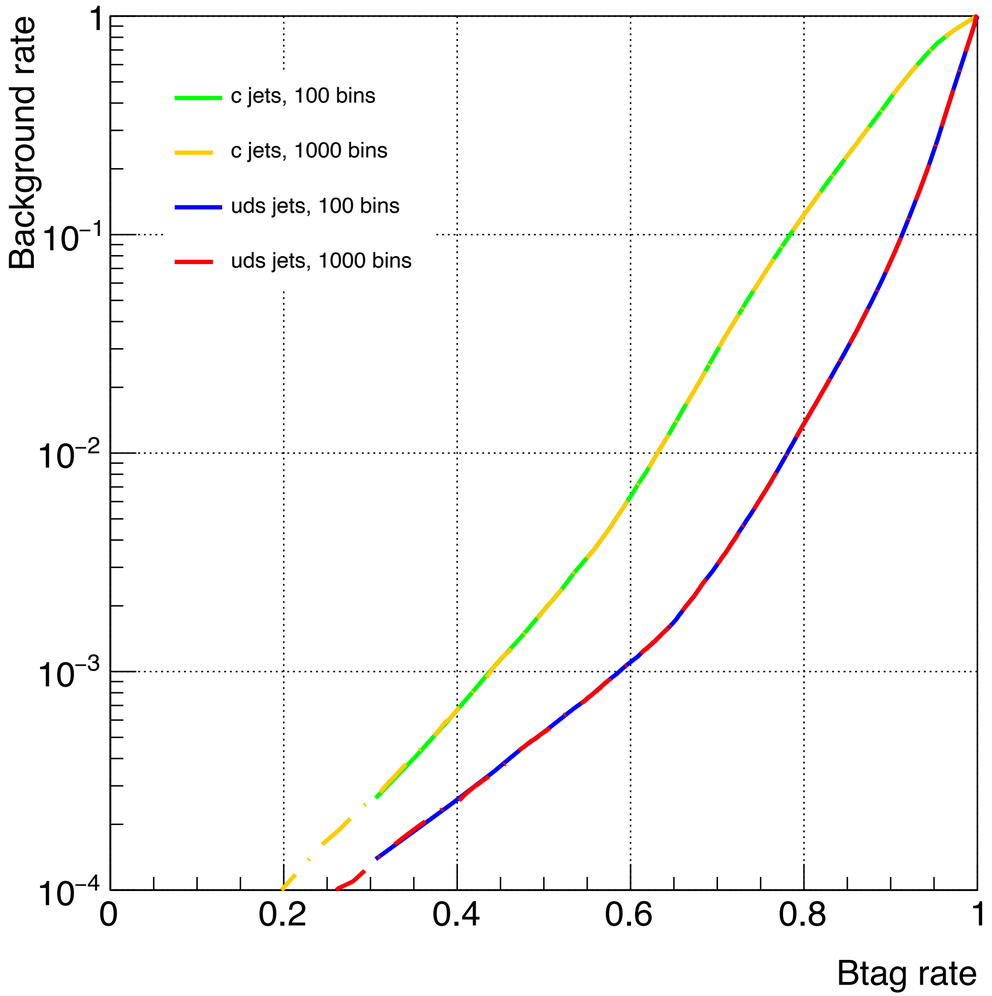


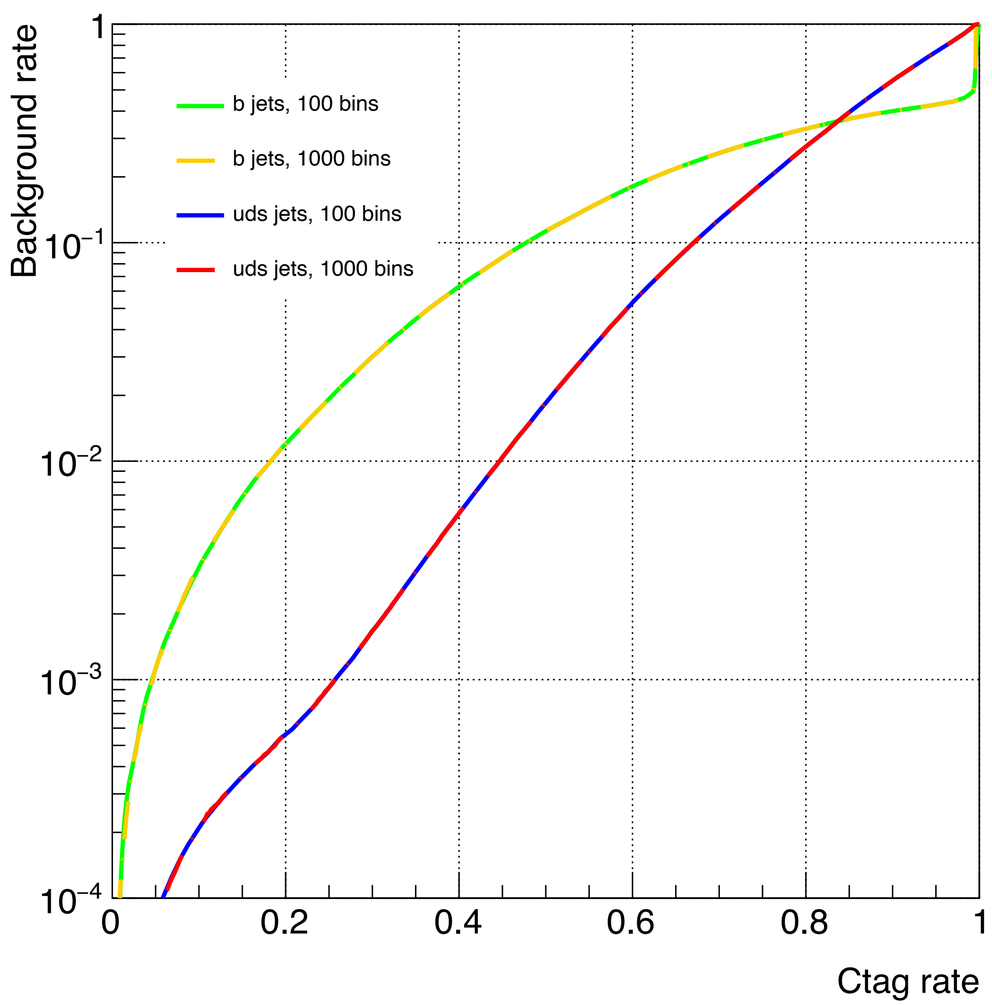
LCFIPlus - ROC from BDT output (1000 bins)



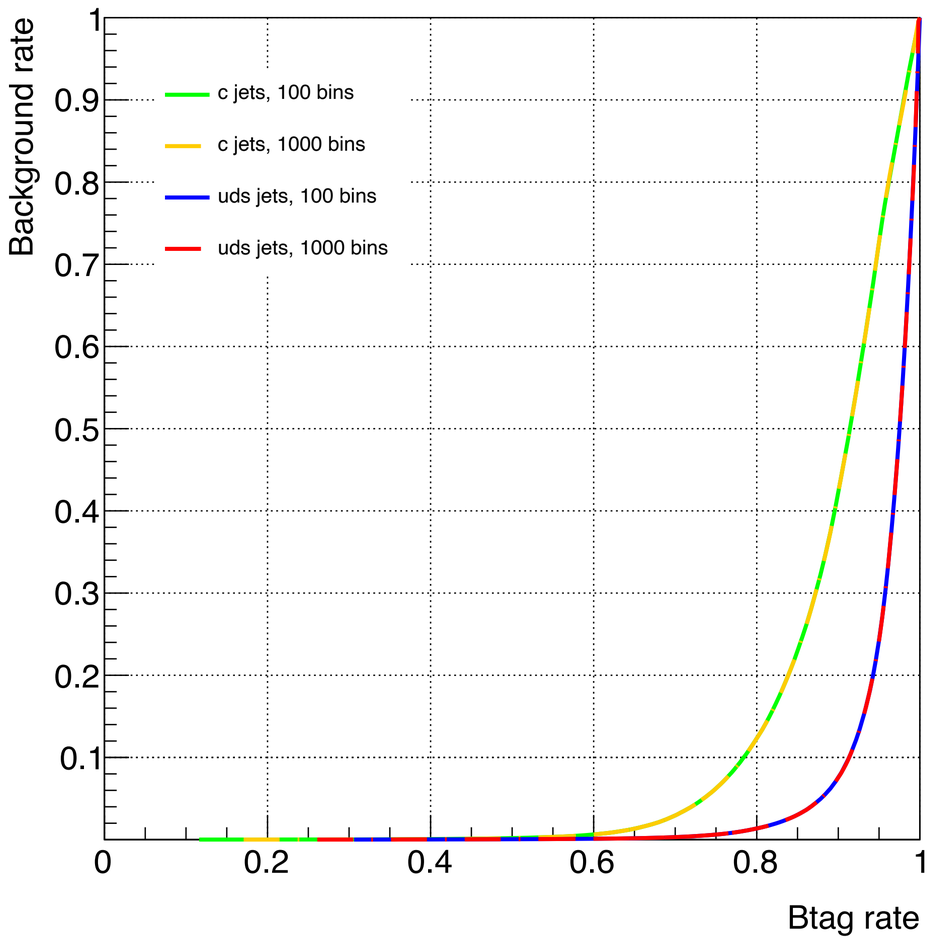


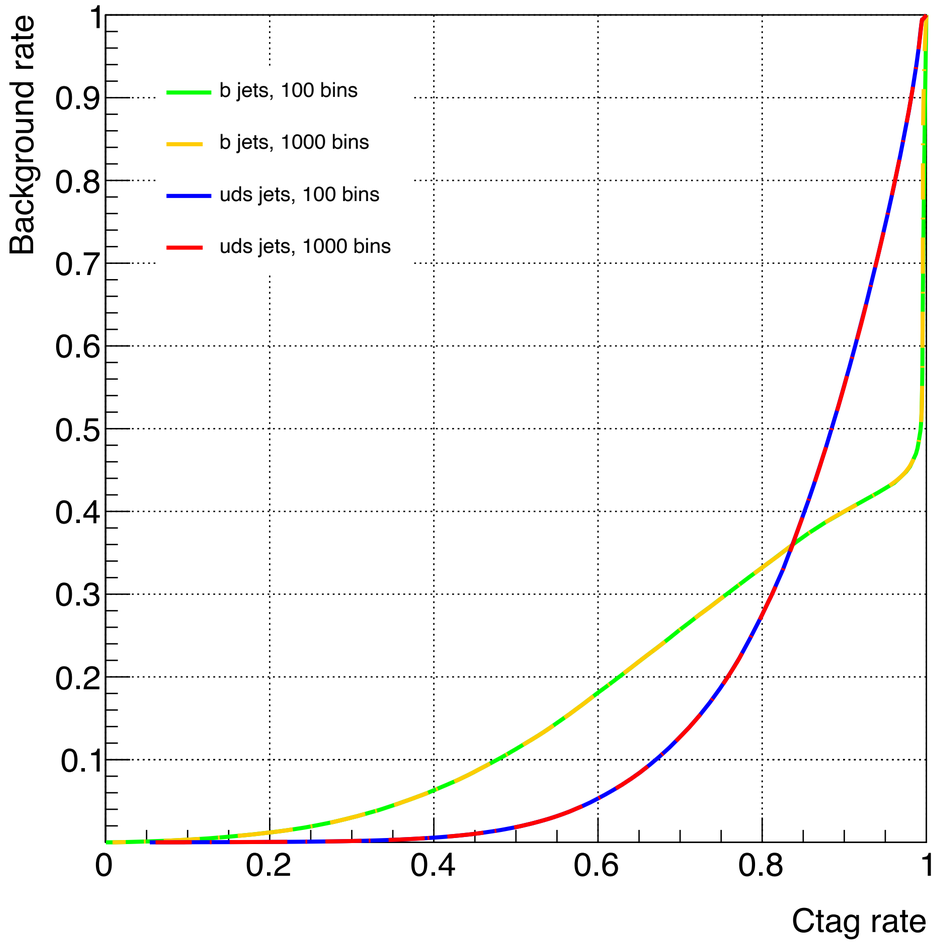
LCFIPlus - ROC 100 bins vs. 1000 bins



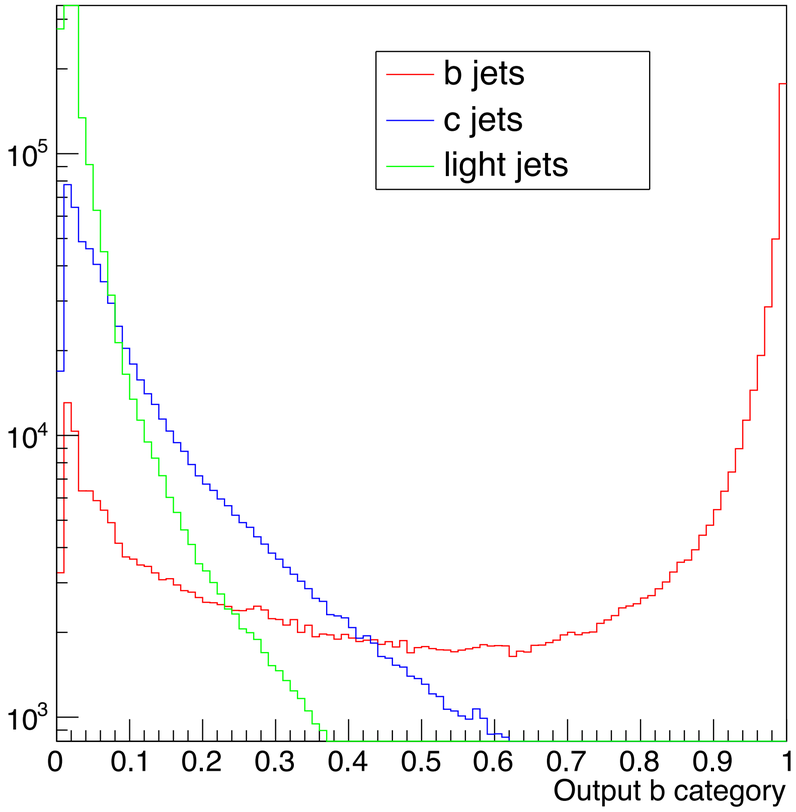


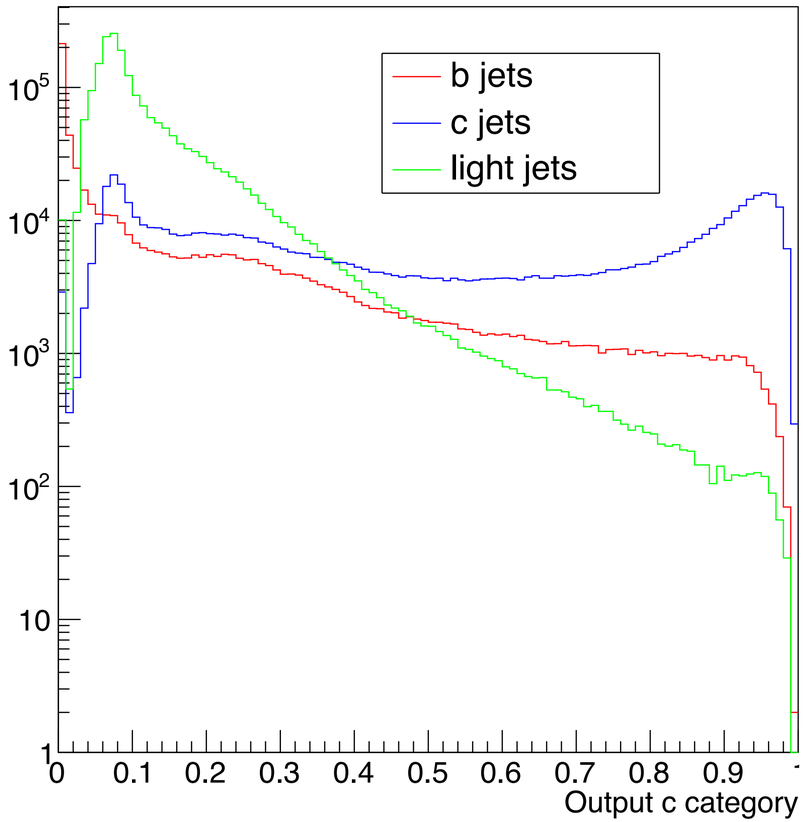
LCFIPlus - ROC 100 bins vs. 1000 bins

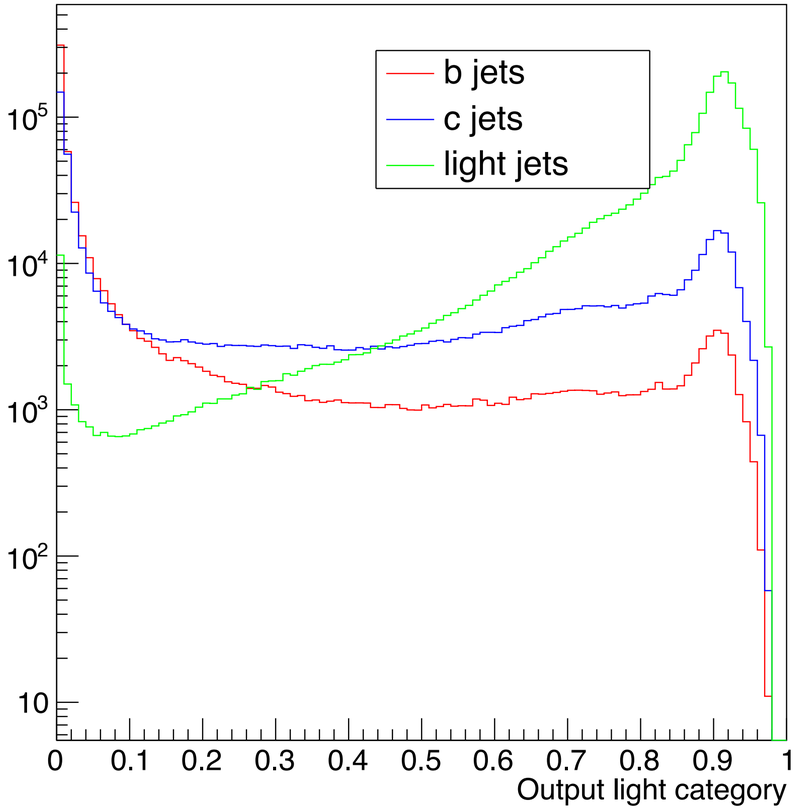




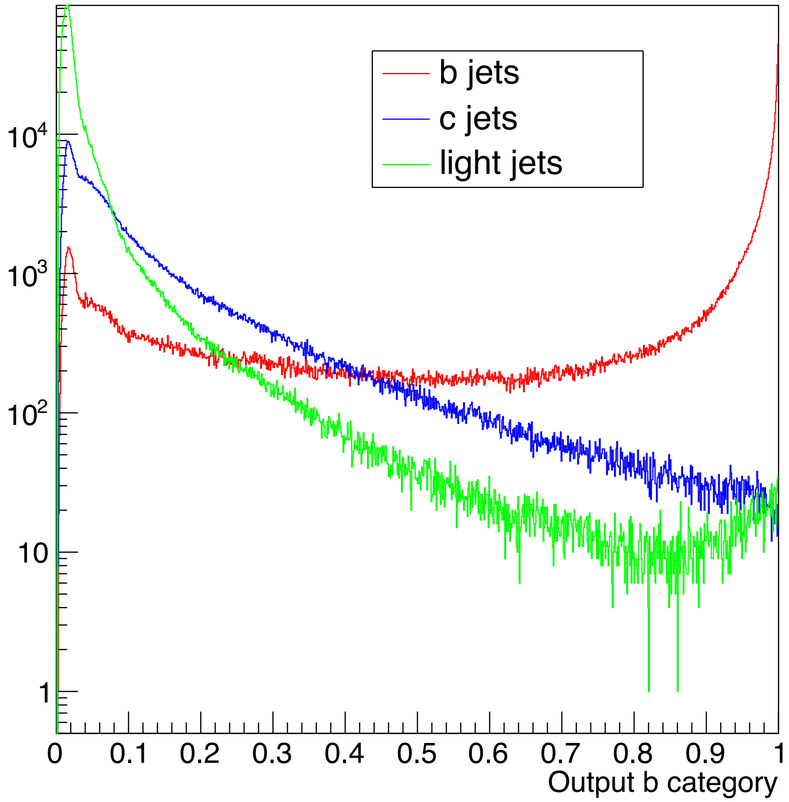
LCFIPlus - BDT output 100 bins

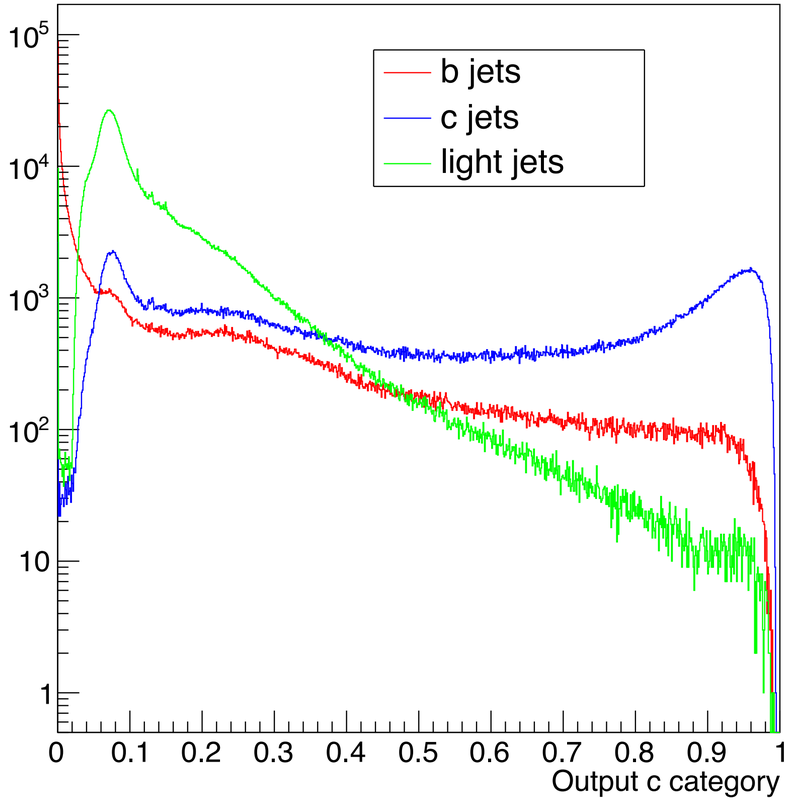


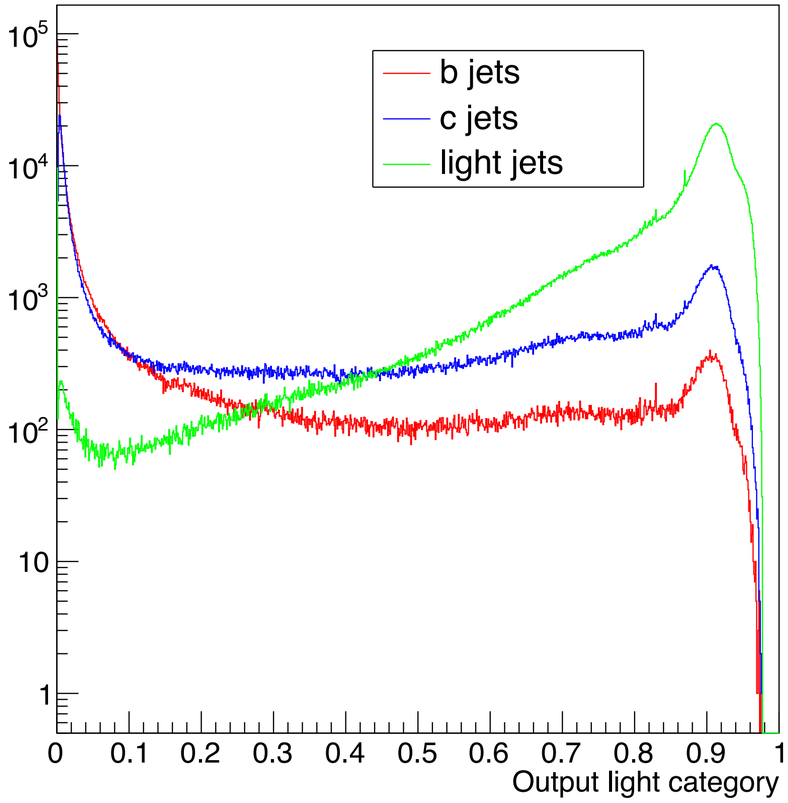




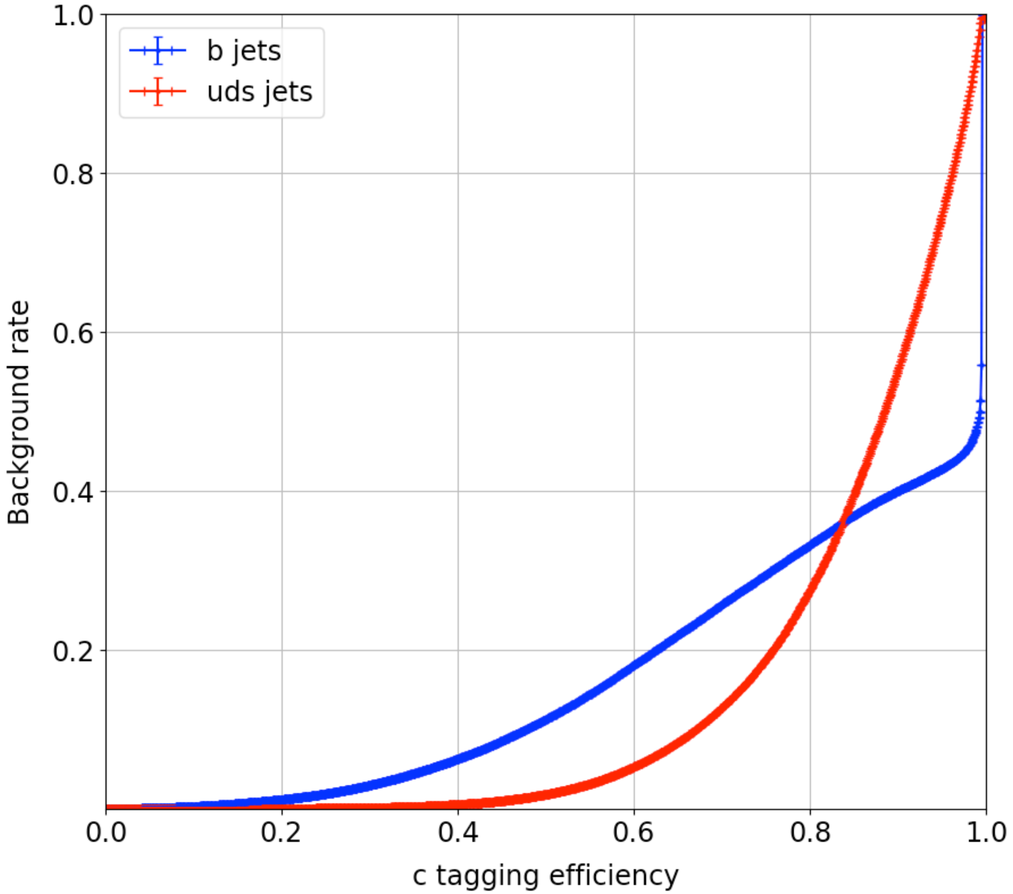
LCFIPlus - BDT output 1000 bins

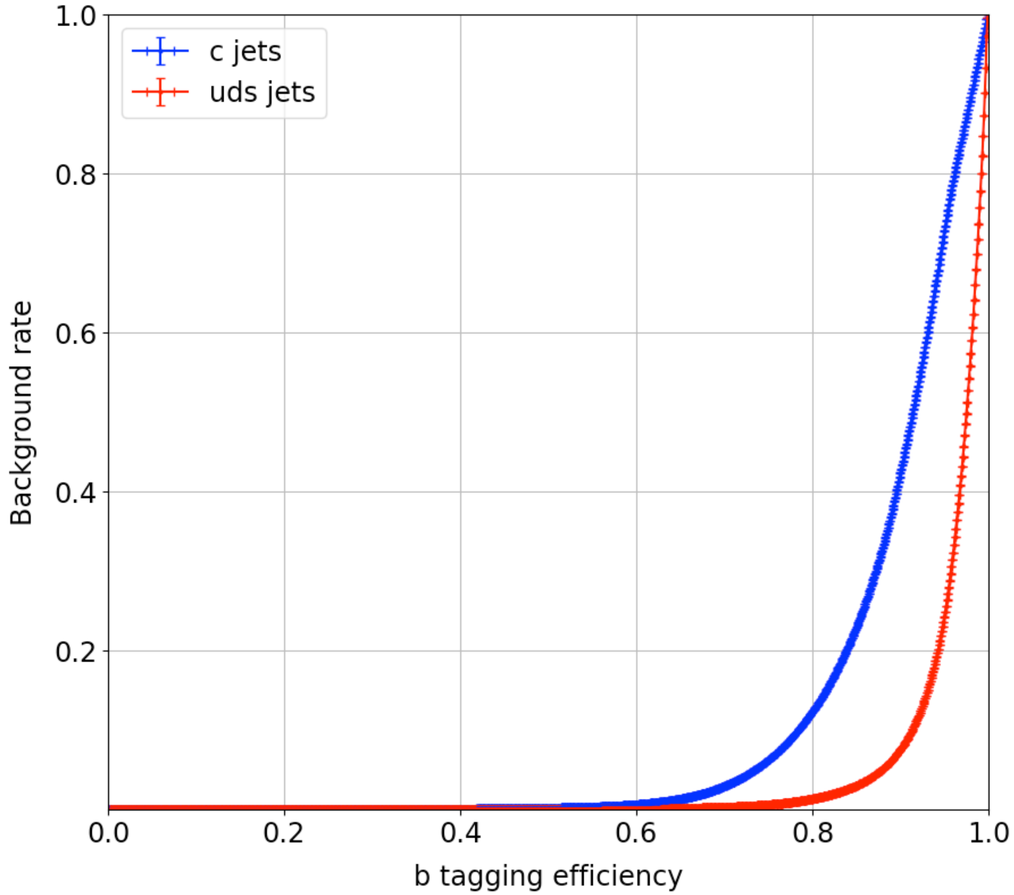




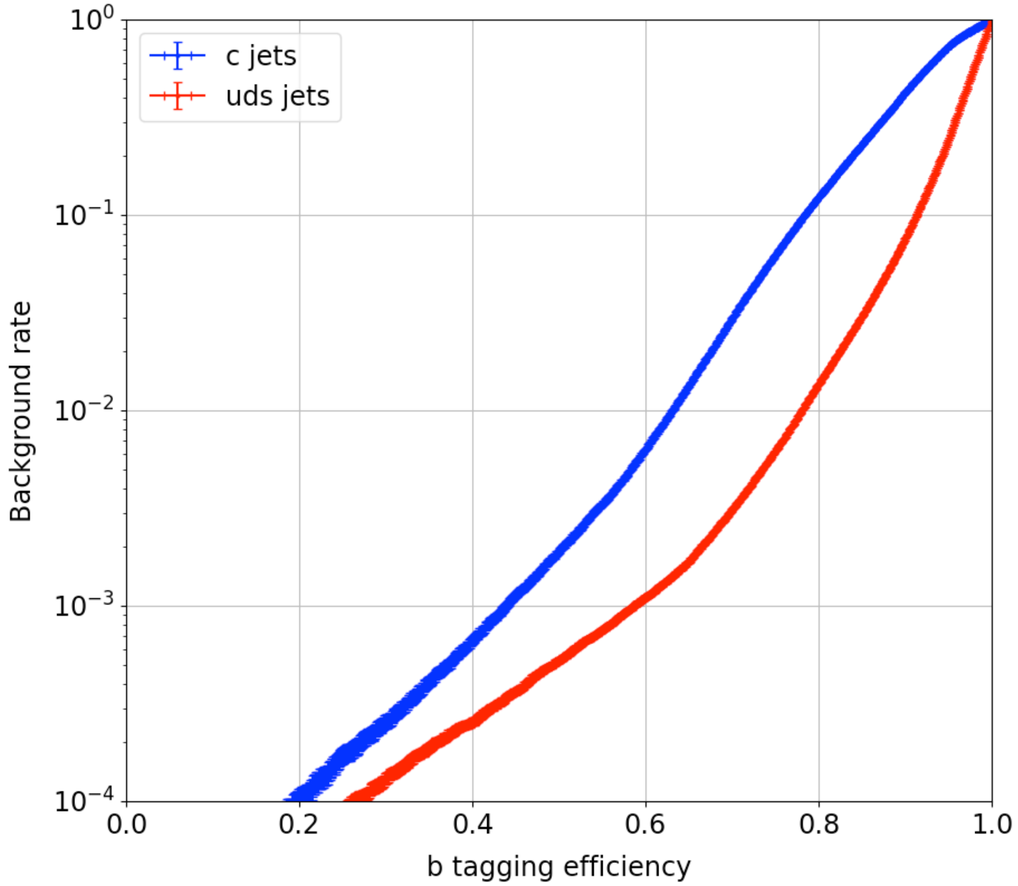


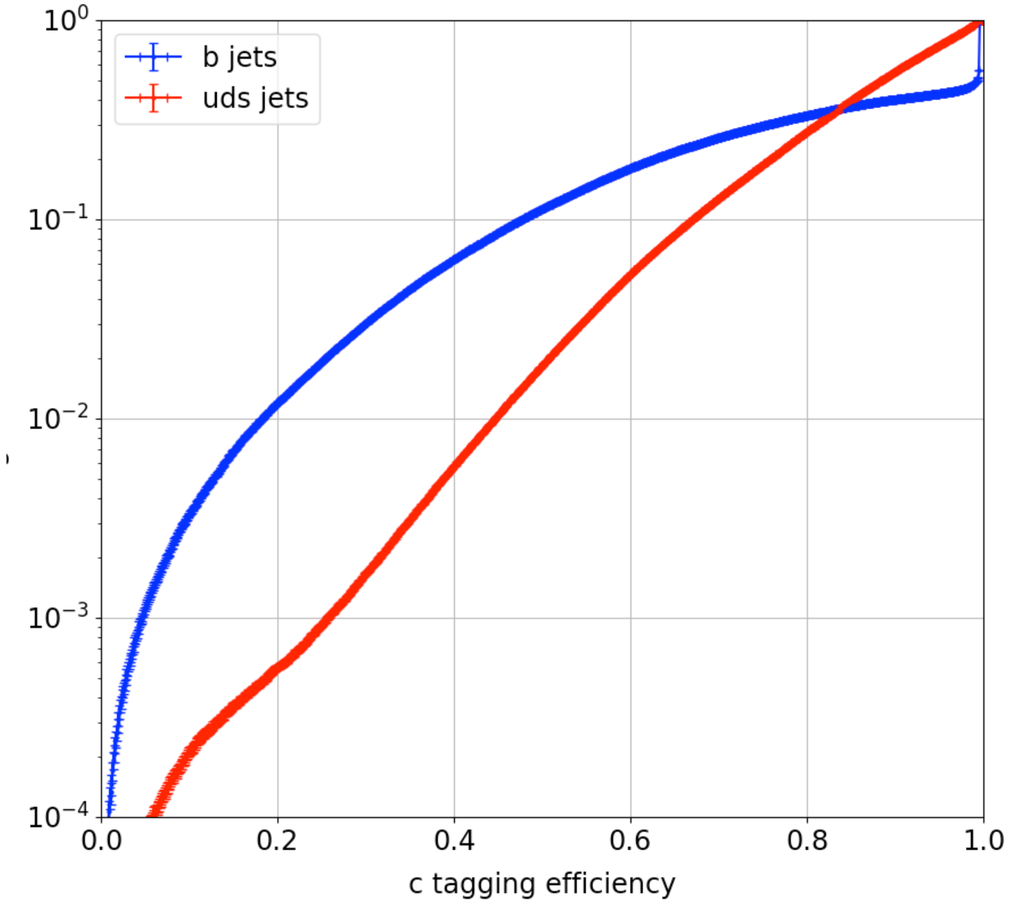
LCFIPlus - ROC from eff, with error bars



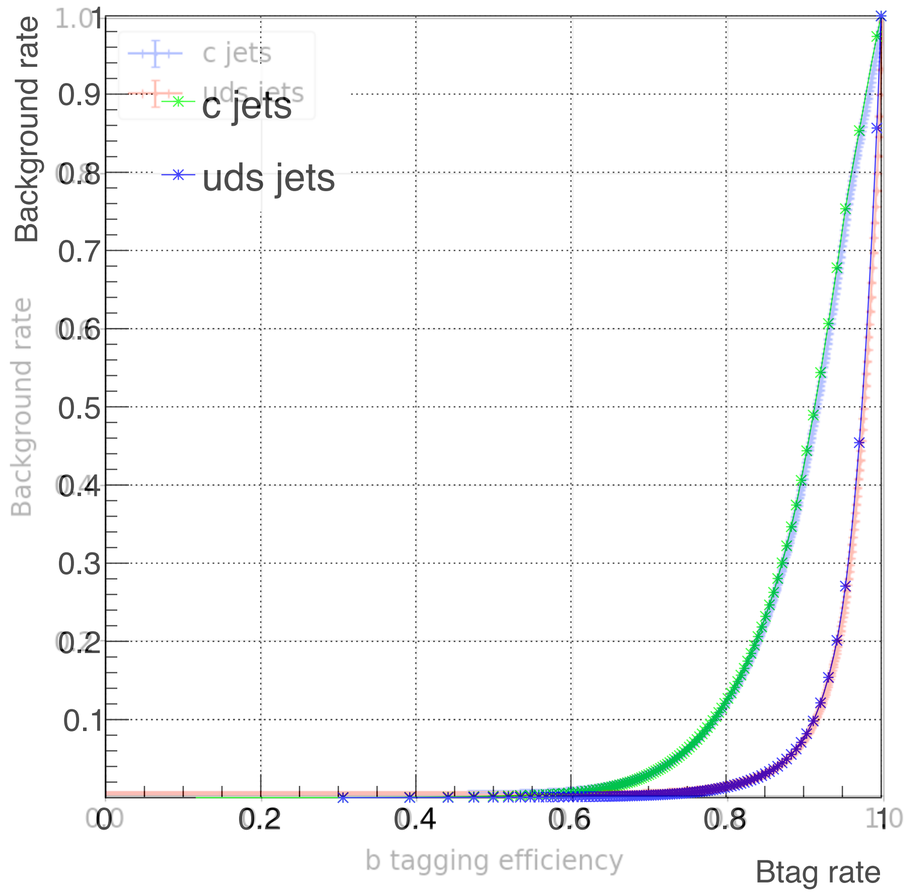


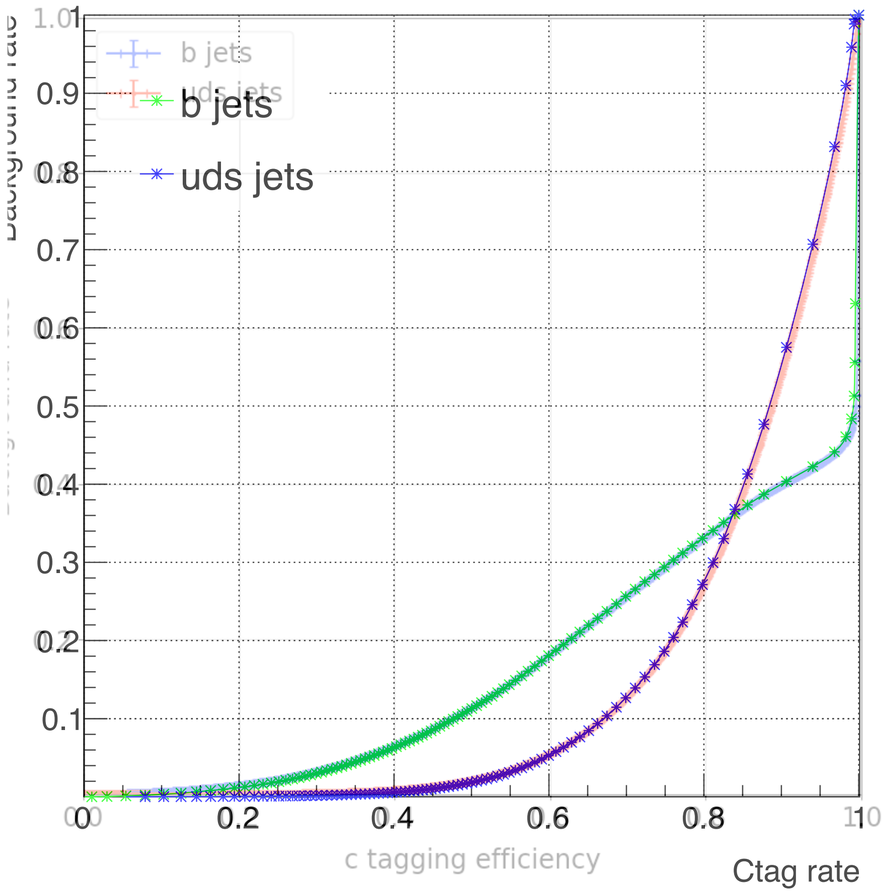
LCFIPlus - ROC from eff, with error bars



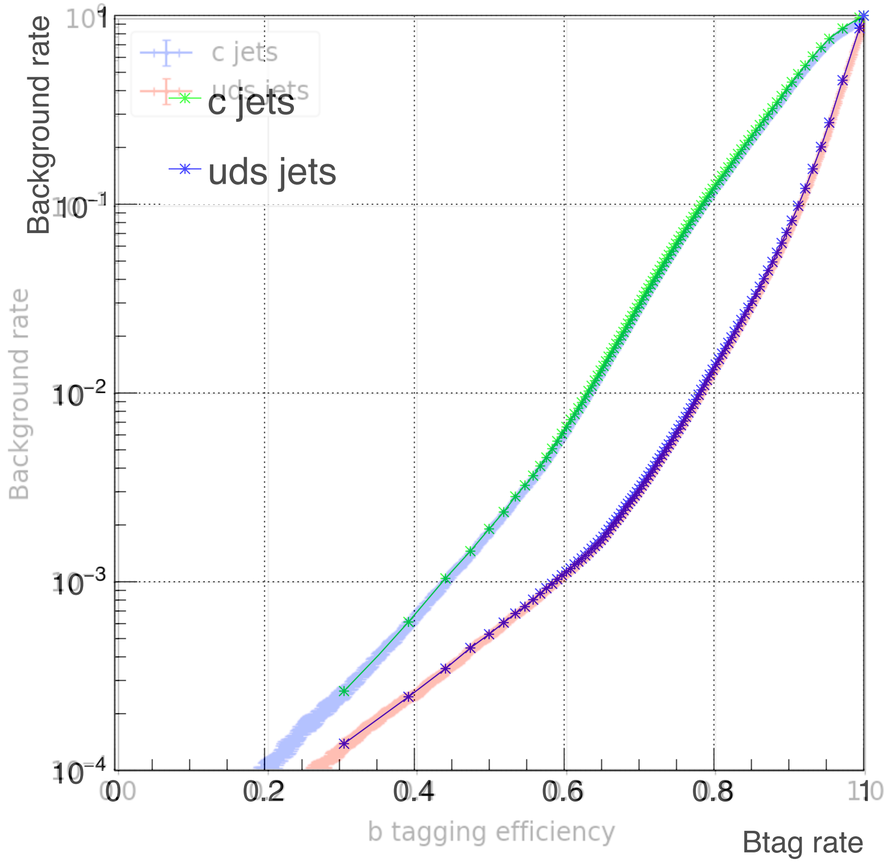


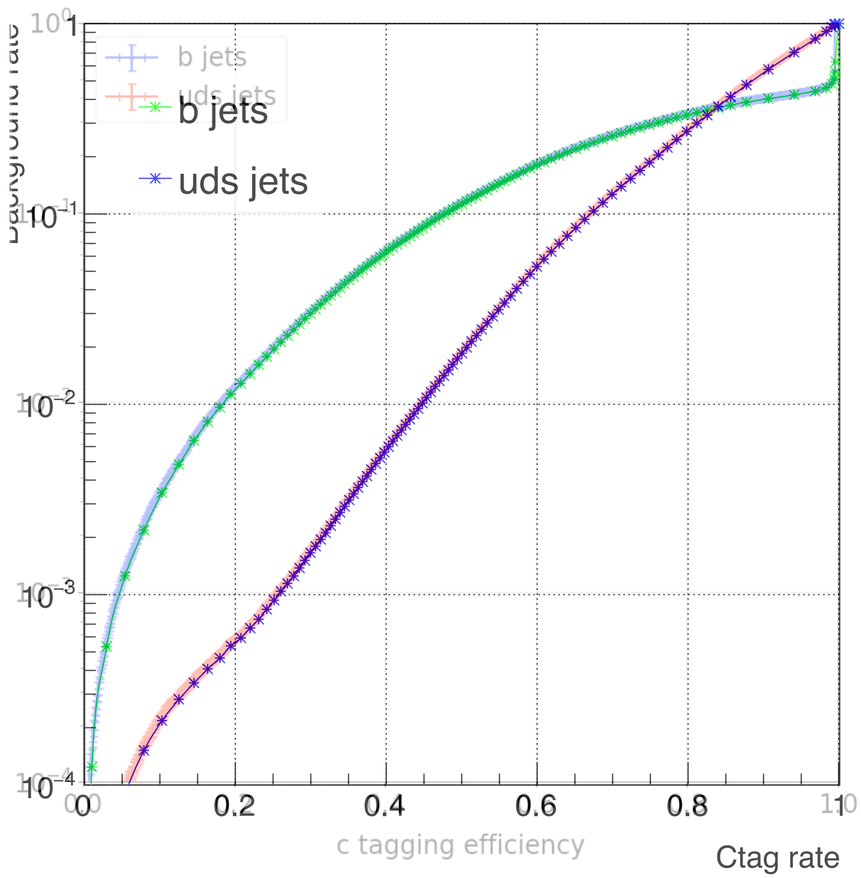
LCFIPlus - ROC from eff, with error bars





LCFIPlus - ROC from eff, with error bars





LCFIPlus variables

Name	Description	Normalization factor	Used by category
trk1d0sig	d0 significance of track with highest d0 significance	1	A, B, C, D
trk2d0sig	d0 significance of track with second highest d0 significance	1	A, B, C, D
trk1z0sig	z0 significance of track with highest d0 significance	1	A, B, C, D
trk2z0sig	z0 significance of track with second highest d0 significance	1	A, B, C, D
trk1pt	transverse momentum of track with highest d0 significance	$1/E_{\text{jet}}$	A, B, C, D
trk2pt	transverse momentum of track with second highest d0 significance	$1/E_{\text{jet}}$	A, B, C, D
jprobr	joint probability in the r-phi plane using all tracks	1	A, B, C, D
jprobr5sigma	joint probability in the r-phi plane using all tracks having impact parameter significance exceeding 5 sigma	1	A, B, C, D
jprobz	joint probability in the z projection using all tracks	1	A, B, C, D
jprobz5sigma	joint probability in the z projection using all tracks having impact parameter significance exceeding 5 sigma	1	A, B, C, D
d0bprob	product of b-quark probabilities of d0 values for all tracks, using b/c/q d0 distributions	1	A, B, C, D
d0cprob	product of c-quark probabilities of d0 values for all tracks, using b/c/q d0 distributions	1	A, B, C, D
d0qprob	product of q-quark probabilities of d0 values for all tracks, using b/c/q d0 distributions	1	A, B, C, D
z0bprob	product of b-quark probabilities of z0 values for all tracks, using b/c/q z0 distributions	1	A, B, C, D
z0cprob	product of c-quark probabilities of z0 values for all tracks, using b/c/q z0 distributions	1	A, B, C, D
z0qprob	product of q-quark probabilities of z0 values for all tracks, using b/c/q z0 distributions	1	A, B, C, D
nmuon	number of identified muons	1	A, B, C, D
nelectron	number of identified electrons	1	A, B, C, D
trkmass	mass of all tracks exceeding 5 sigma significance in d0/z0 values	1	A, B, C, D

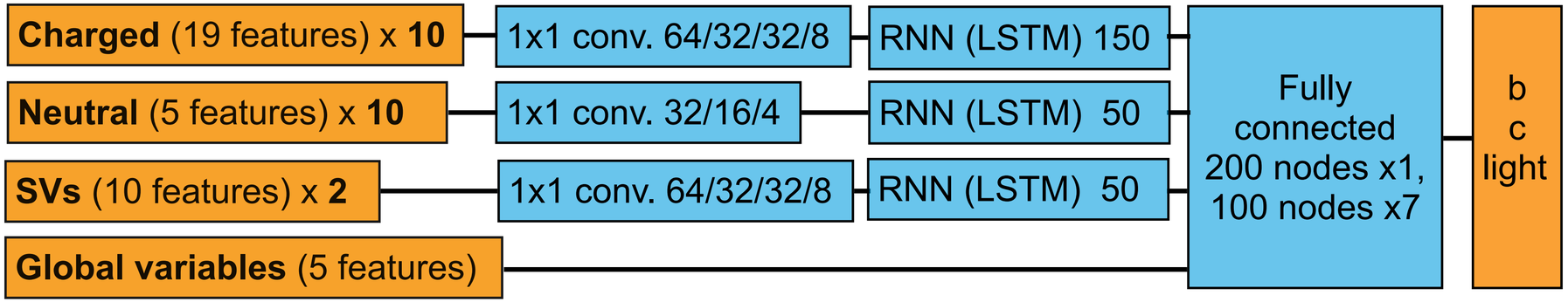
used as global variables

LCFIPlus variables

Name	Description	Normalization factor	Used by category
1vtxprob	vertex probability with all tracks associated in vertices combined	1	B, C, D
vtxlen1	decay length of the first vertex in the jet (zero if no vertex is found)	$1/E_{\text{jet}}$	B, C, D
vtxlen2	decay length of the second vertex in the jet (zero if number of vertex is less than two)	$1/E_{\text{jet}}$	D
vtxlen12	distance between the first and second vertex (zero if number of vertex is less than two)	$1/E_{\text{jet}}$	D
vtxsig1	decay length significance of the first vertex in the jet (zero if no vertex is found)	$1/E_{\text{jet}}$	B, C, D
vtxsig2	decay length significance of the second vertex in the jet (zero if number of vertex is less than two)	$1/E_{\text{jet}}$	D
vtxsig12	vtxlen12 divided by its error as computed from the sum of the covariance matrix of the first and second vertices, projected along the line connecting the two vertices	$1/E_{\text{jet}}$	D
vtxdirang1	the angle between the momentum (computed as a vector sum of track momenta) and the displacement of the first vertex	E_{jet}	B, C, D
vtxdirang2	the angle between the momentum (computed as a vector sum of track momenta) and the displacement of the second vertex	E_{jet}	D
vtxmult1	number of tracks included in the first vertex (zero if no vertex is found)	1	B, C, D
vtxmult2	number of tracks included in the second vertex (zero if number of vertex is less than two)	1	D
vtxmult	number of tracks which are used to form secondary vertices (summed for all vertices)	1	D
vtxmom1	magnitude of the vector sum of the momenta of all tracks combined into the first vertex	$1/E_{\text{jet}}$	B, C, D
vtxmom2	magnitude of the vector sum of the momenta of all tracks combined into the second vertex	$1/E_{\text{jet}}$	D
vtxmass1	mass of the first vertex computed from the sum of track four-momenta	1	B, C, D
vtxmass2	mass of the second vertex computed from the sum of track four-momenta	1	D
vtxmass	vertex mass as computed from the sum of four momenta of all tracks forming secondary vertices	1	B, C, D
vtxmasspc	mass of the vertex with minimum pt correction allowed by the error matrices of the primary and secondary vertices	1	B, C, D
vtxprob	vertex probability; for multiple vertices, the probability P is computed as 1-P = (1-P1)(1-P2)...(1-PN)	1	B, C, D

used as global variables

Architecture & data pre-processing



- classify jets into **three classes**: b jets, c jets & light jets
- **ordering of input particles** by (as applied in CMS)
  - 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
- if a value of a features is not available, the value is set to -10
- **normalize input features** to mean 0, std 1