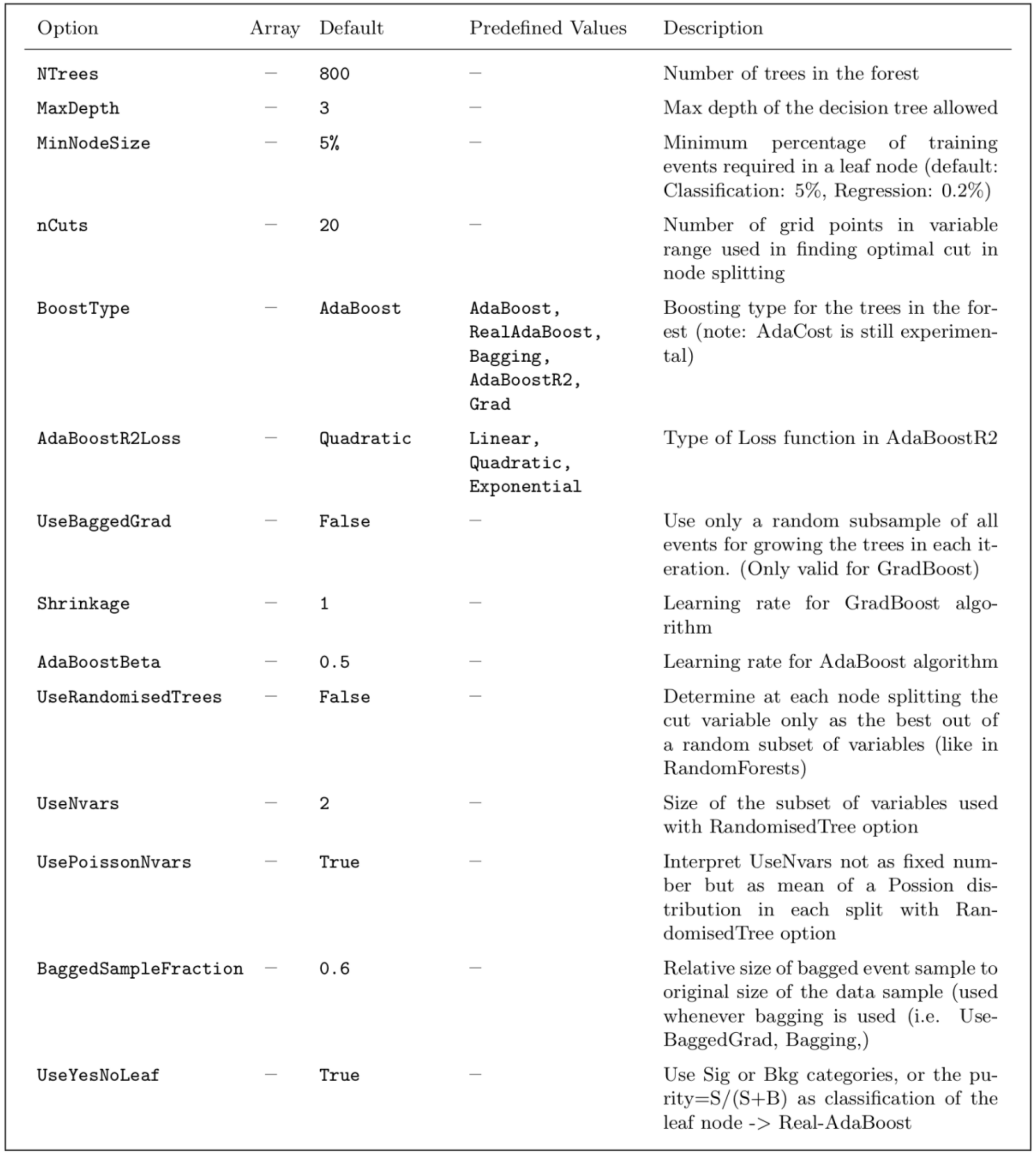
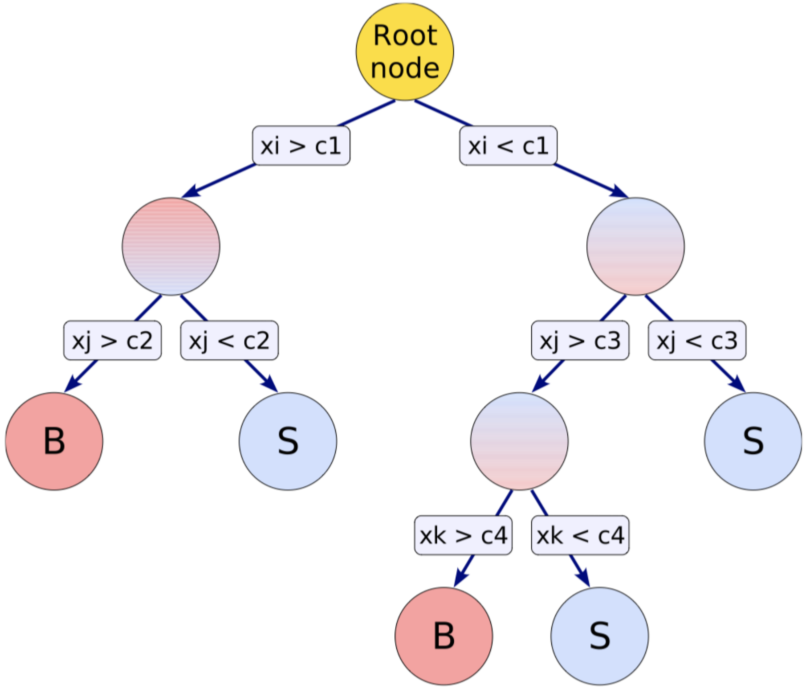
Ejets Update 28/05/2020

BDT (TMVA) Implementation for Selection



Option	Array	Default	Predefined Values	Description
NTrees	–	800	–	Number of trees in the forest
MaxDepth	–	3	–	Max depth of the decision tree allowed
MinNodeSize	–	5%	–	Minimum percentage of training events required in a leaf node (default: Classification: 5%, Regression: 0.2%)
nCuts	–	20	–	Number of grid points in variable range used in finding optimal cut in node splitting
BoostType	–	AdaBoost	AdaBoost, RealAdaBoost, Bagging, AdaBoostR2, Grad	Boosting type for the trees in the forest (note: AdaCost is still experimental)
AdaBoostR2Loss	–	Quadratic	Linear, Quadratic, Exponential	Type of Loss function in AdaBoostR2
UseBaggedGrad	–	False	–	Use only a random subsample of all events for growing the trees in each iteration. (Only valid for GradBoost)
Shrinkage	–	1	–	Learning rate for GradBoost algorithm
AdaBoostBeta	–	0.5	–	Learning rate for AdaBoost algorithm
UseRandomisedTrees	–	False	–	Determine at each node splitting the cut variable only as the best out of a random subset of variables (like in RandomForests)
UseNvars	–	2	–	Size of the subset of variables used with RandomisedTree option
UsePoissonNvars	–	True	–	Interpret UseNvars not as fixed number but as mean of a Possion distribution in each split with RandomisedTree option
BaggedSampleFraction	–	0.6	–	Relative size of bagged event sample to original size of the data sample (used whenever bagging is used (i.e. UseBaggedGrad, Bagging,)
UseYesNoLeaf	–	True	–	Use Sig or Bkg categories, or the purity=S/(S+B) as classification of the leaf node -> Real-AdaBoost



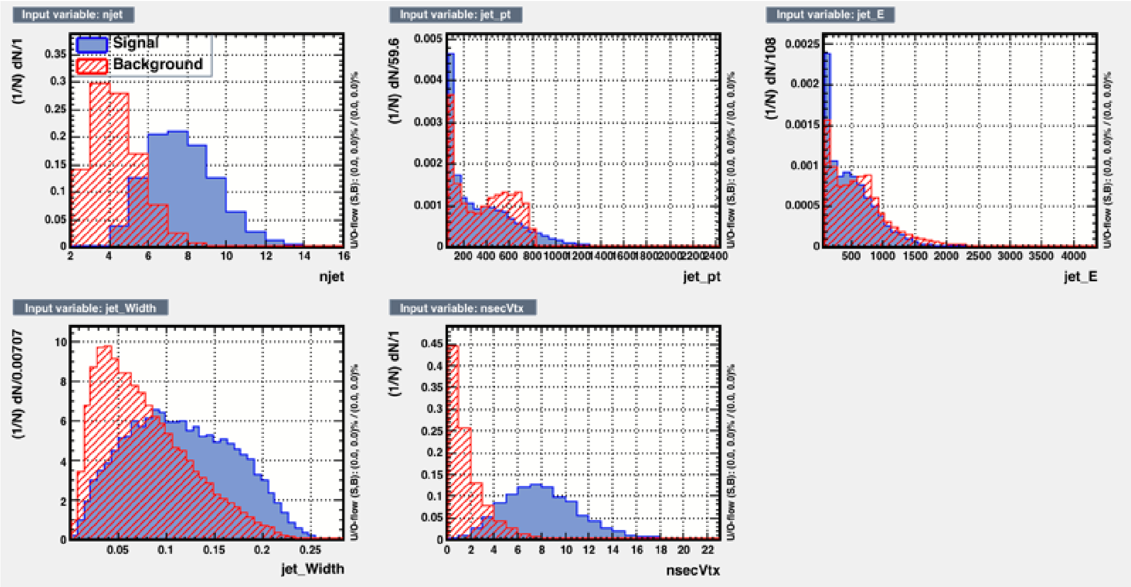
Summary

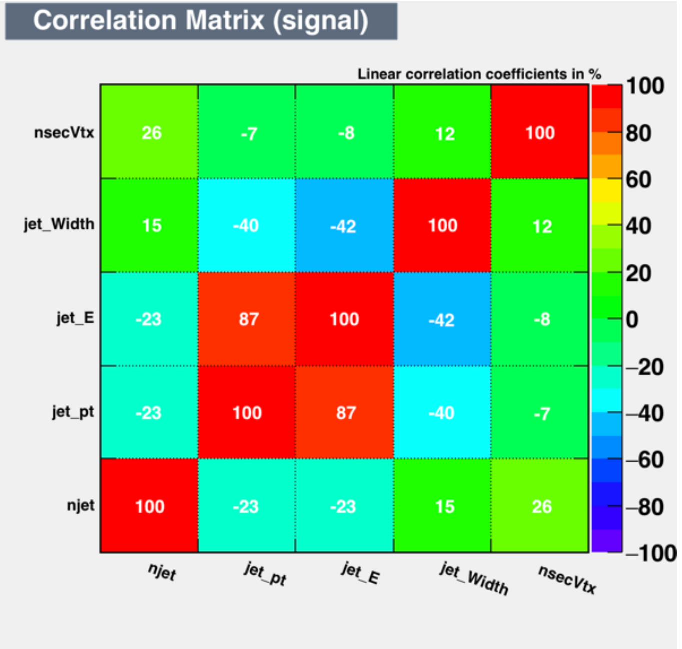
- have implemented the ROOT TMVA BDT tool on signal/background MC
- used all events from files with no cuts
- did not apply event weighting
- treated all variables as ‘event level’ variables
- used default BDT settings
- ran over cases:
  ModelA_1400_20
  ModelA_1000_150
  ModelA_600_1
  ModelB_1400_20
  ModelB_1000_5
  ModelB_600_300
  ModelE_1400_75
  ModelE_1000_150
  ModelE_600_0p5

- Used variables: **njet jet_pt jet_E jet_Width nsecVtx**

This is not a real selection/analysis…just proof-of-principle test…

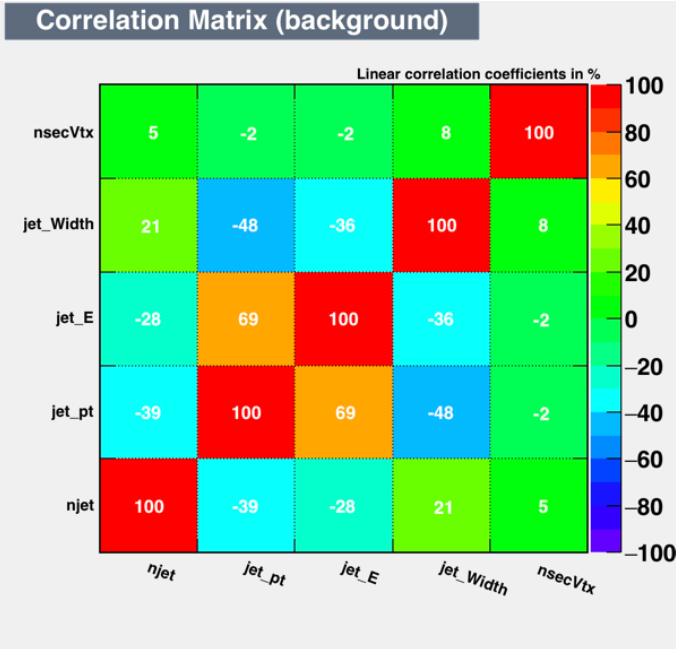
ModelA_1400_20




Correlation Matrix (signal)

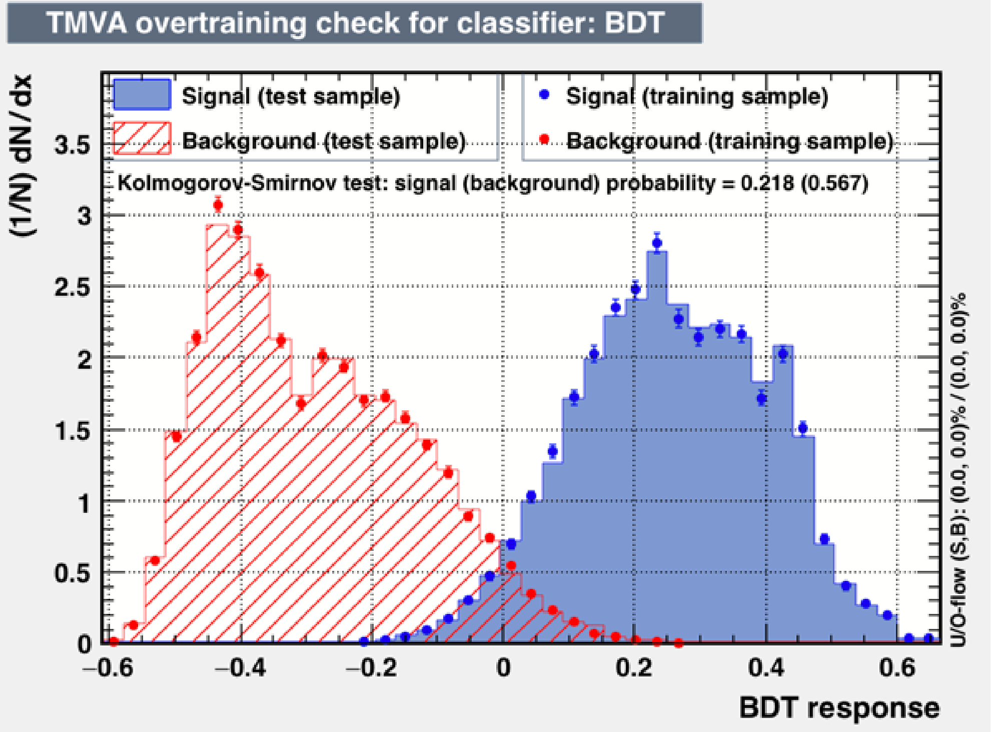
Linear correlation coefficients in %

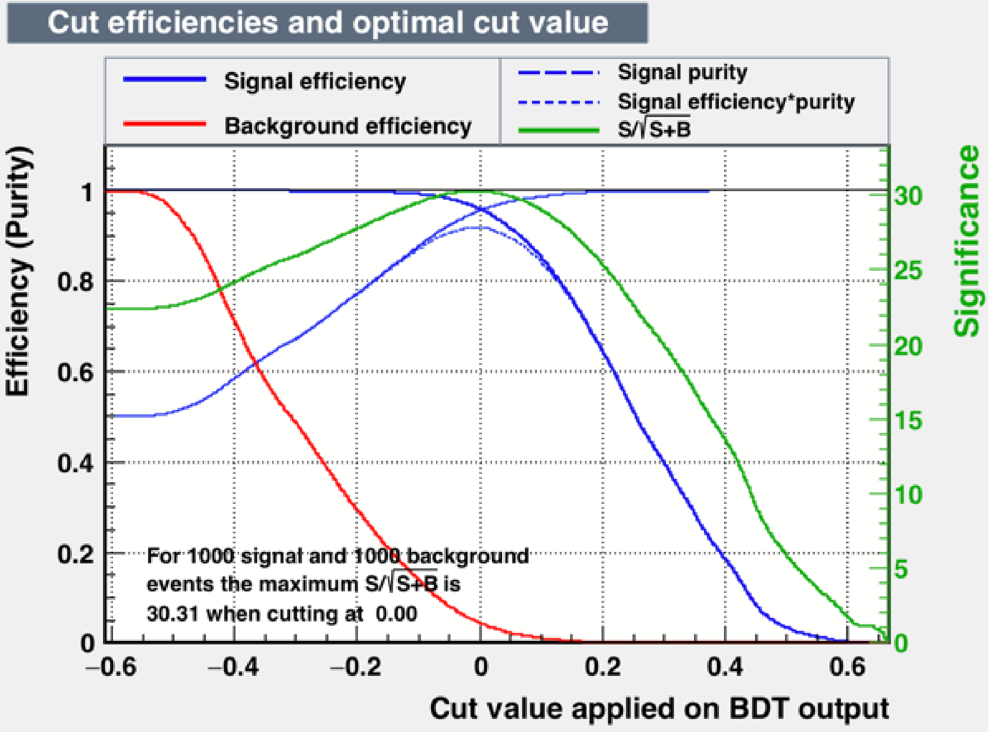
	njet	jet_pt	jet_E	jet_Width	nsecVtx
nsecVtx	26	-7	-8	12	100
jet_Width	15	-40	-42	100	12
jet_E	-23	87	100	-42	-8
jet_pt	-23	100	87	-40	-7
njet	100	-23	-23	15	26


Correlation Matrix (background)

Linear correlation coefficients in %

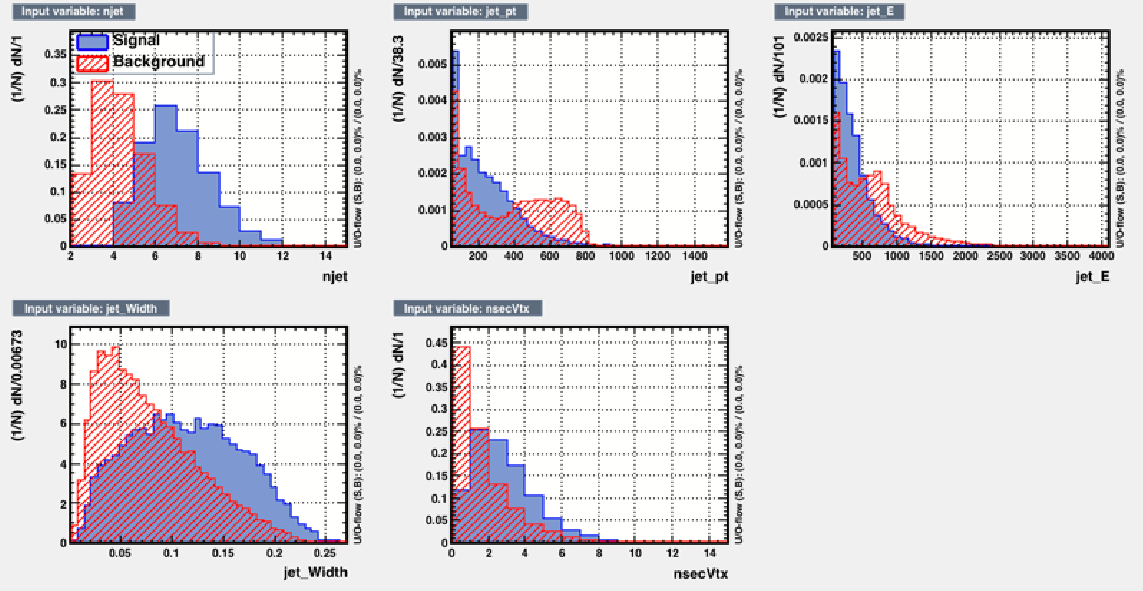
	njet	jet_pt	jet_E	jet_Width	nsecVtx
nsecVtx	5	-2	-2	8	100
jet_Width	21	-48	-36	100	8
jet_E	-28	69	100	-36	-2
jet_pt	-39	100	69	-48	-2
njet	100	-39	-28	21	5


TMVA overtraining check for classifier: BDT

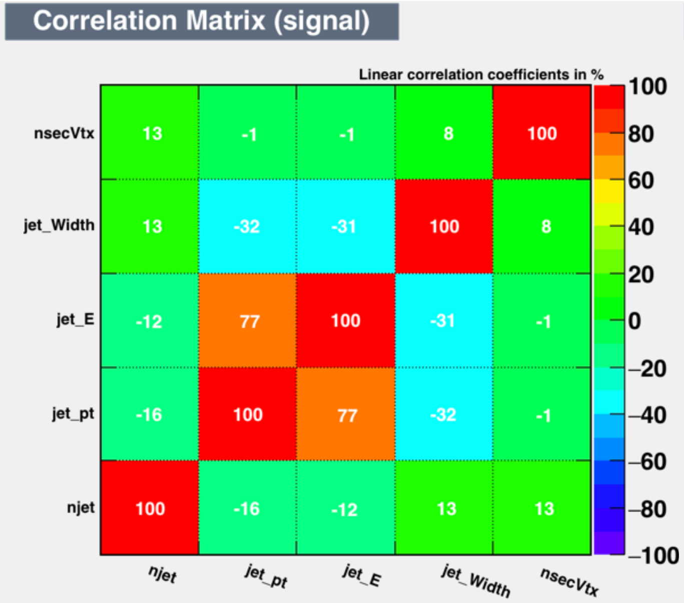

Cut efficiencies and optimal cut value

For 1000 signal and 1000 background events the maximum $S/\sqrt{S+B}$ is 30.31 when cutting at 0.00

ModelA_600_1

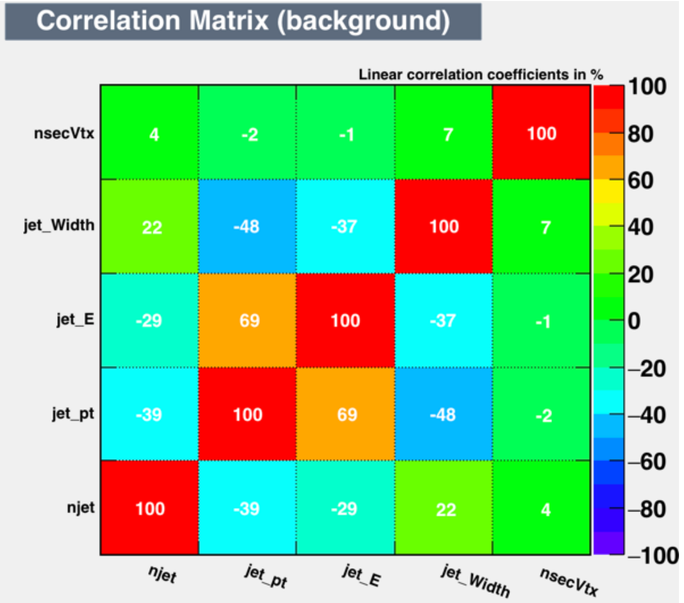


Correlation Matrix (signal)


Linear correlation coefficients in %
100
80
60
40
20
0
−20
−40
−60
−80
−100

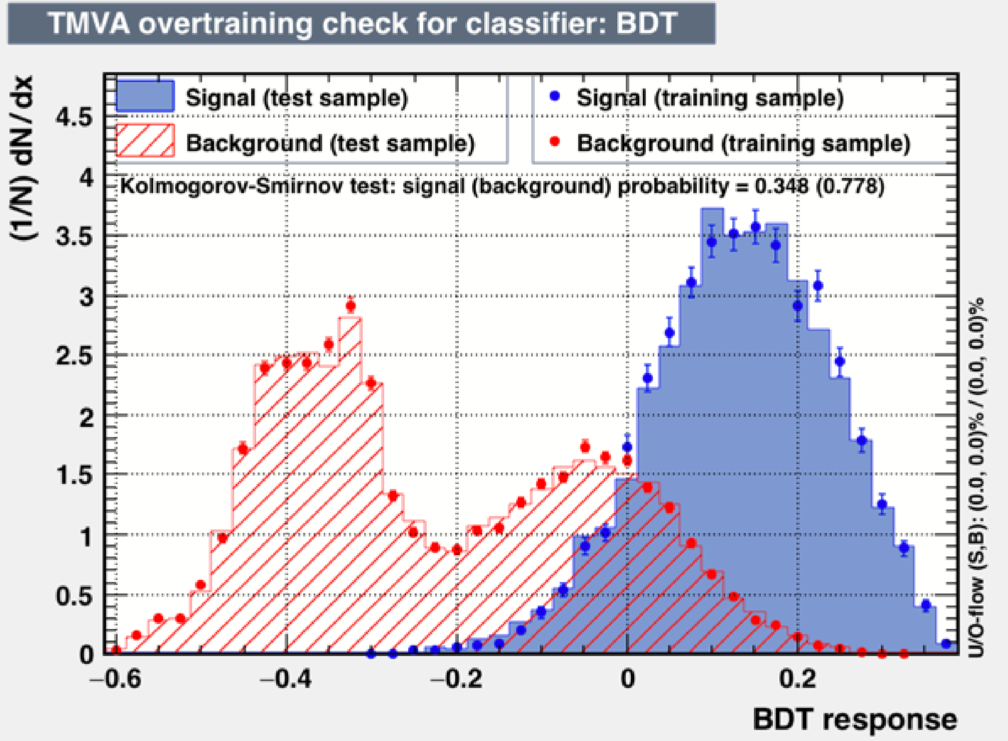
	njet	jet_pt	jet_E	jet_Width	nsecVtx
nsecVtx	13	-1	-1	8	100
jet_Width	13	-32	-31	100	8
jet_E	-12	77	100	-31	-1
jet_pt	-16	100	77	-32	-1
njet	100	-16	-12	13	13

Correlation Matrix (background)


Linear correlation coefficients in %
100
80
60
40
20
0
−20
−40
−60
−80
−100

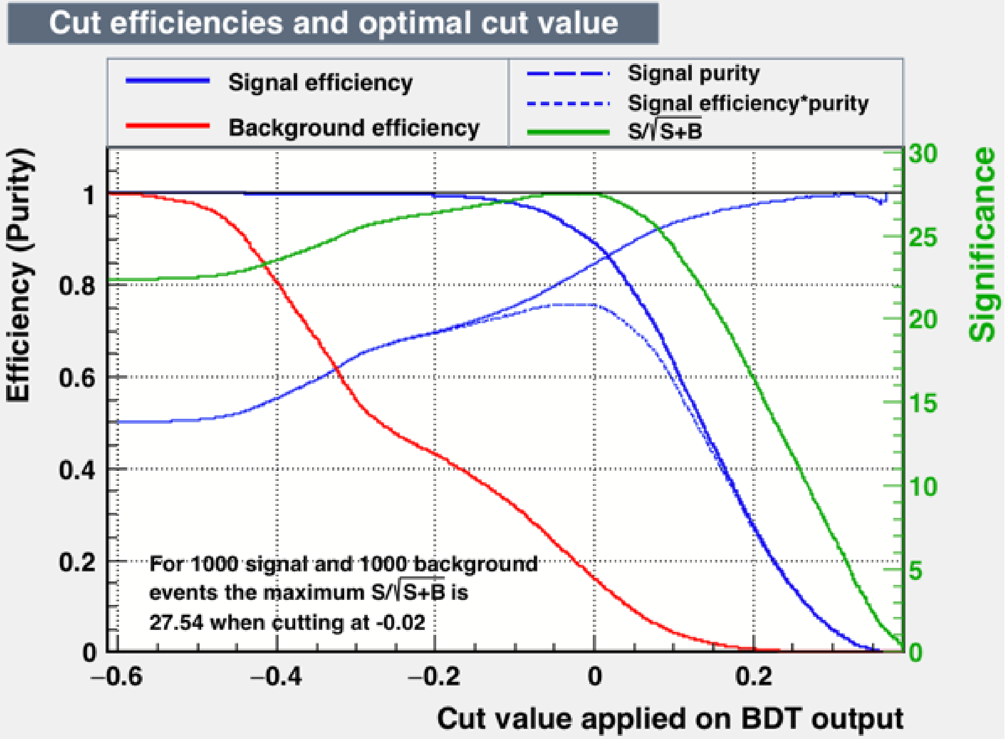
	njet	jet_pt	jet_E	jet_Width	nsecVtx
nsecVtx	4	-2	-1	7	100
jet_Width	22	-48	-37	100	7
jet_E	-29	69	100	-37	-1
jet_pt	-39	100	69	-48	-2
njet	100	-39	-29	22	4

TMVA overtraining check for classifier: BDT



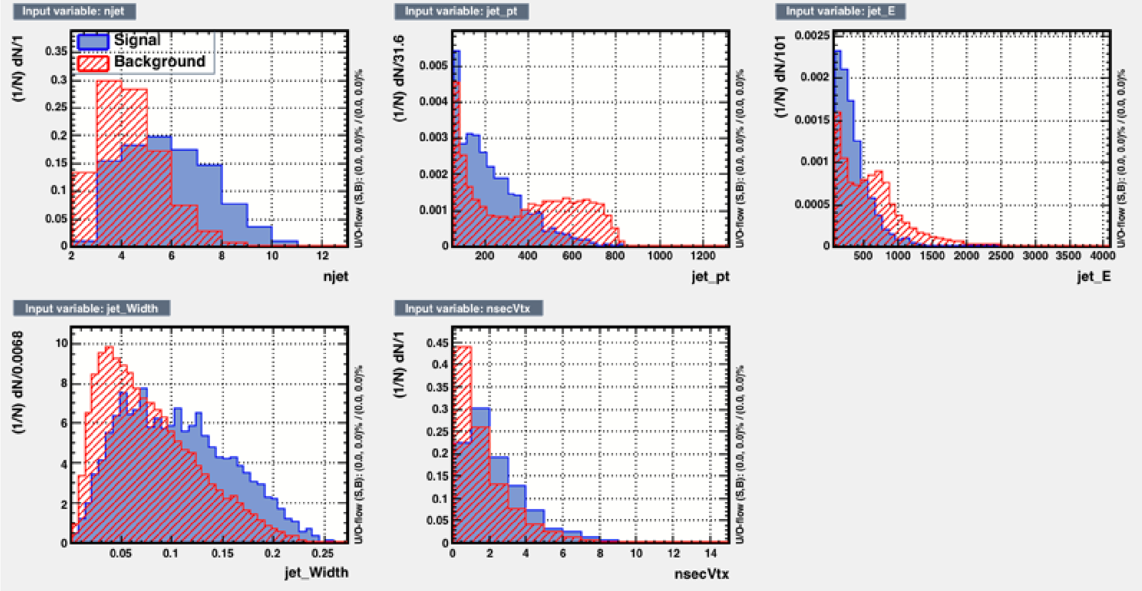
Kolmogorov-Smirnov test: signal (background) probability = 0.348 (0.778)

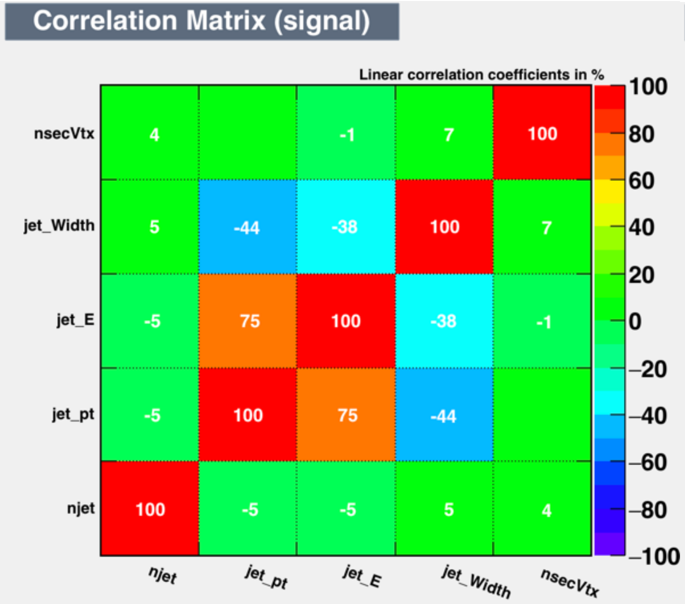
Cut efficiencies and optimal cut value



For 1000 signal and 1000 background events the maximum S/√S+B is 27.54 when cutting at -0.02

ModelB_600_300

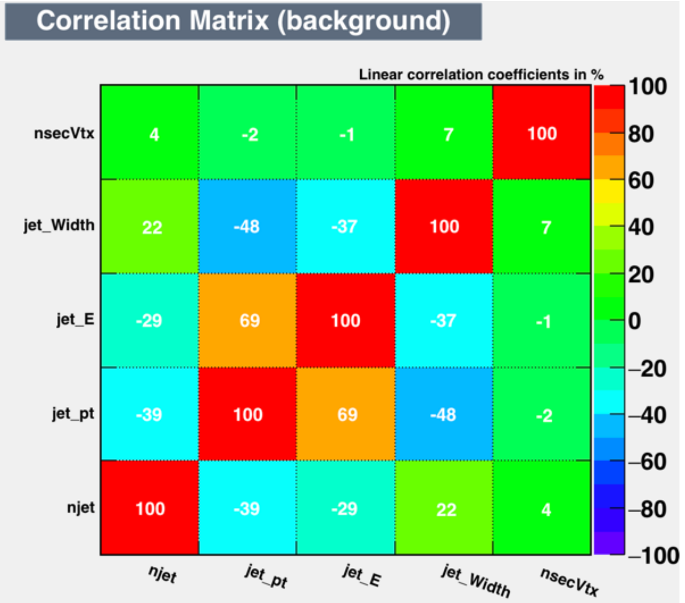




Correlation Matrix (signal)

Linear correlation coefficients in %

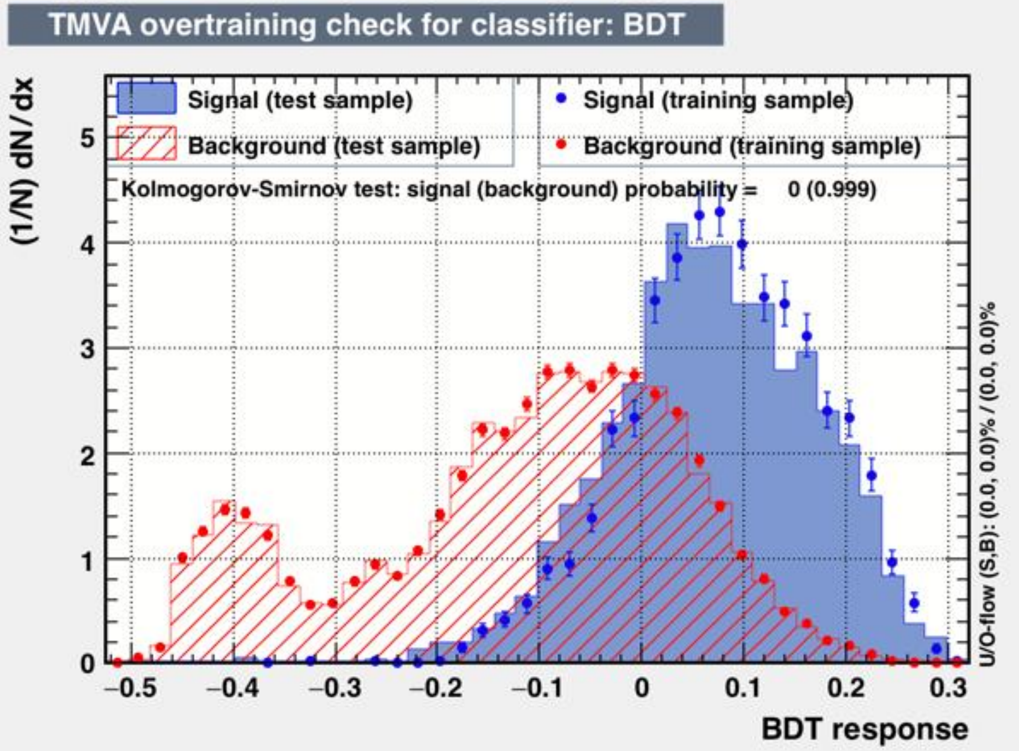
	njet	jet_pt	jet_E	jet_Width	nsecVtx
nsecVtx	4		-1	7	100
jet_Width	5	-44	-38	100	7
jet_E	-5	75	100	-38	-1
jet_pt	-5	100	75	-44	
njet	100	-5	-5	5	4



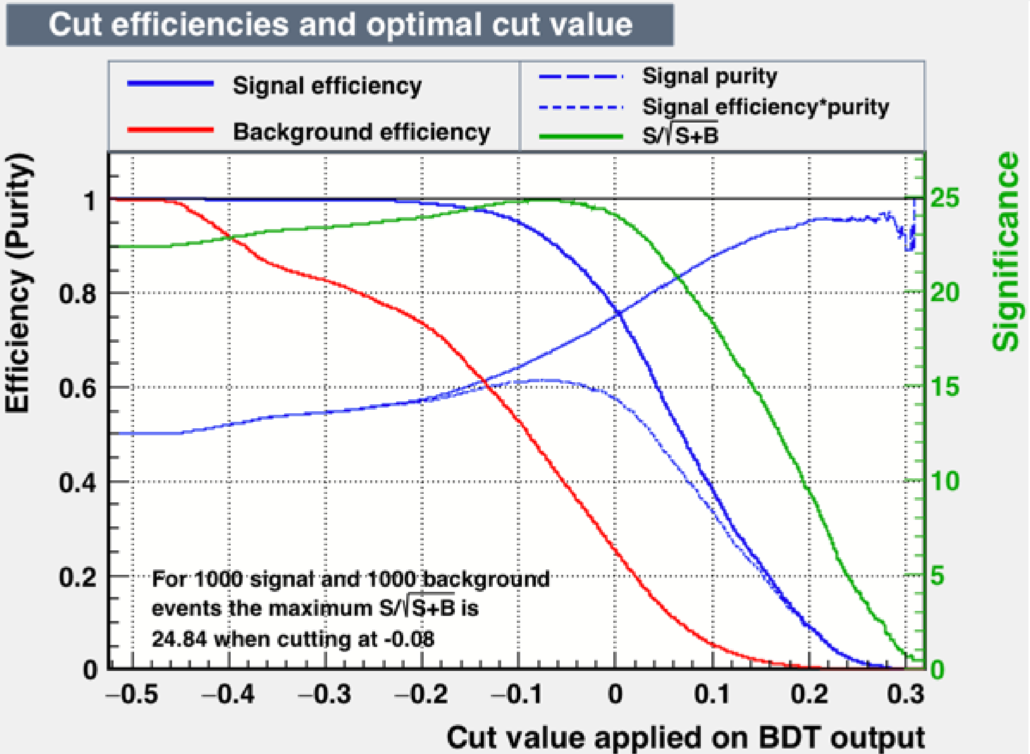
Correlation Matrix (background)

Linear correlation coefficients in %

	njet	jet_pt	jet_E	jet_Width	nsecVtx
nsecVtx	4	-2	-1	7	100
jet_Width	22	-48	-37	100	7
jet_E	-29	69	100	-37	-1
jet_pt	-39	100	69	-48	-2
njet	100	-39	-29	22	4



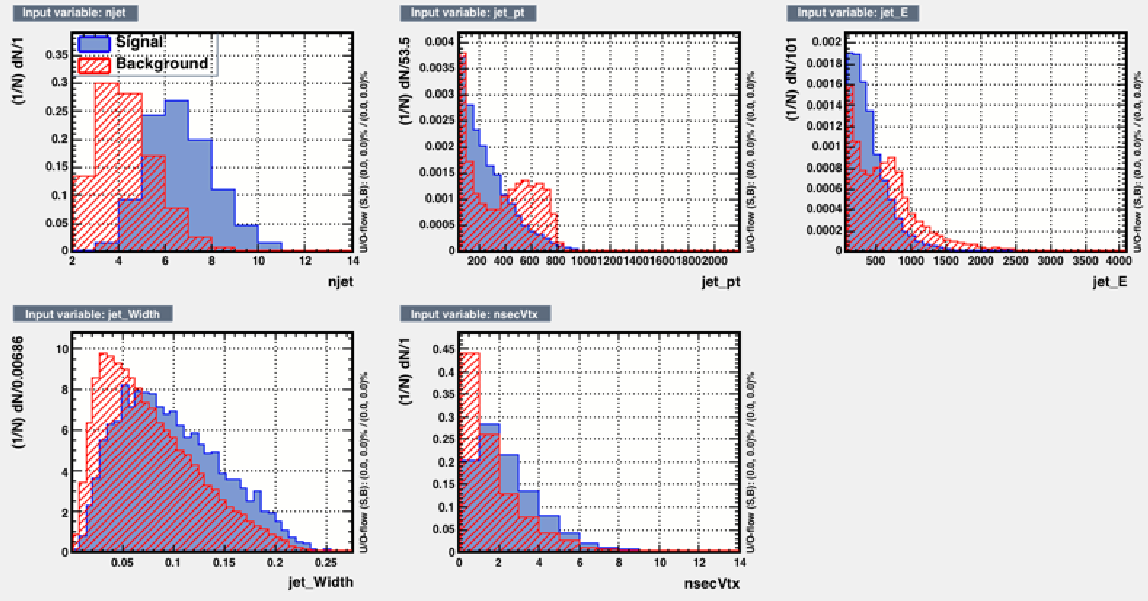
TMVA overtraining check for classifier: BDT

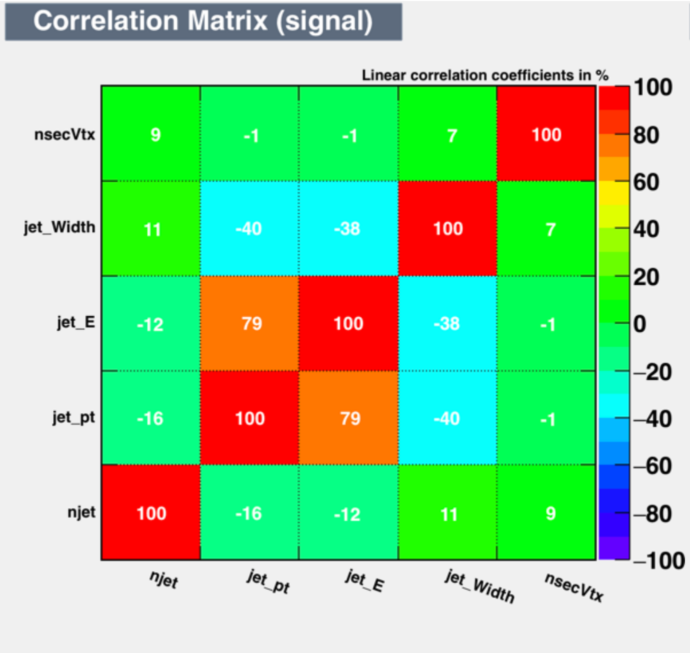


Cut efficiencies and optimal cut value

For 1000 signal and 1000 background events the maximum S/$\sqrt{S+B}$ is 24.84 when cutting at -0.08

ModelE 1000 150

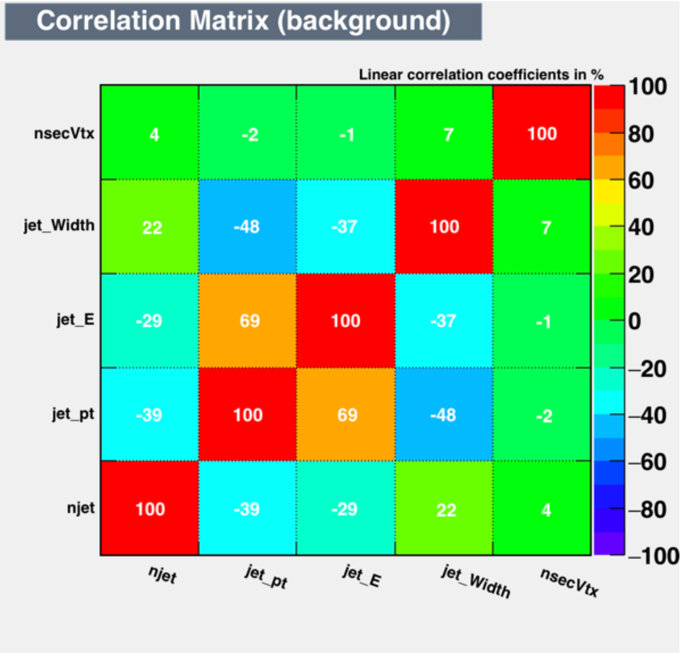




Correlation Matrix (signal)

Linear correlation coefficients in %

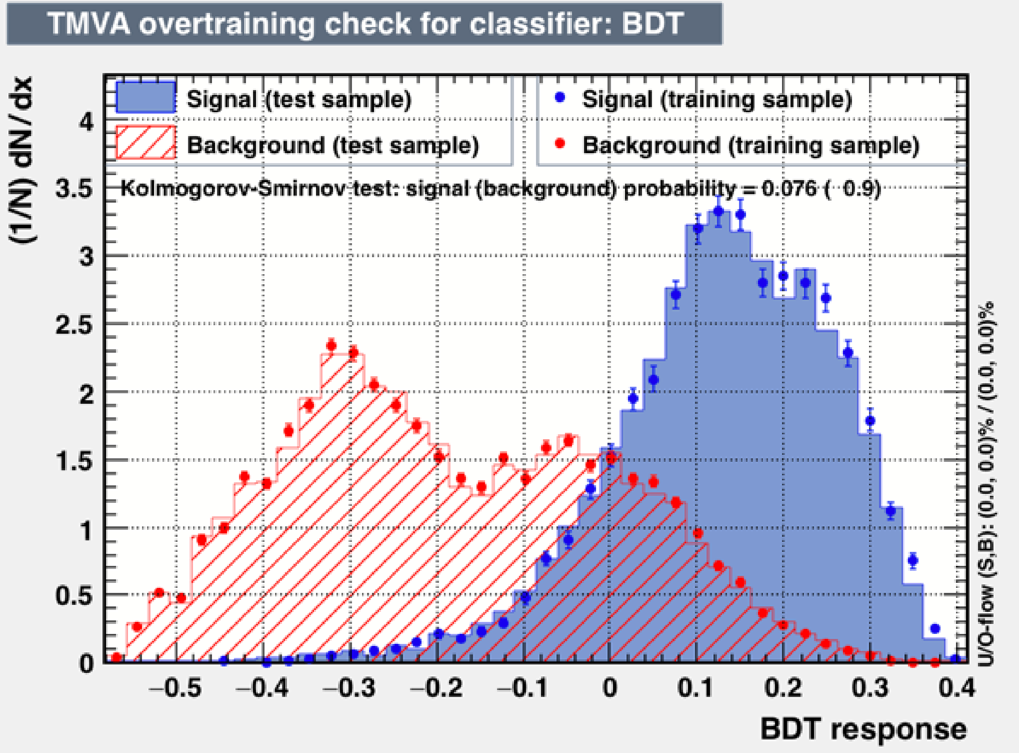
	njet	jet_pt	jet_E	jet_Width	nsecVtx
nsecVtx	9	-1	-1	7	100
jet_Width	11	-40	-38	100	7
jet_E	-12	79	100	-38	-1
jet_pt	-16	100	79	-40	-1
njet	100	-16	-12	11	9



Correlation Matrix (background)

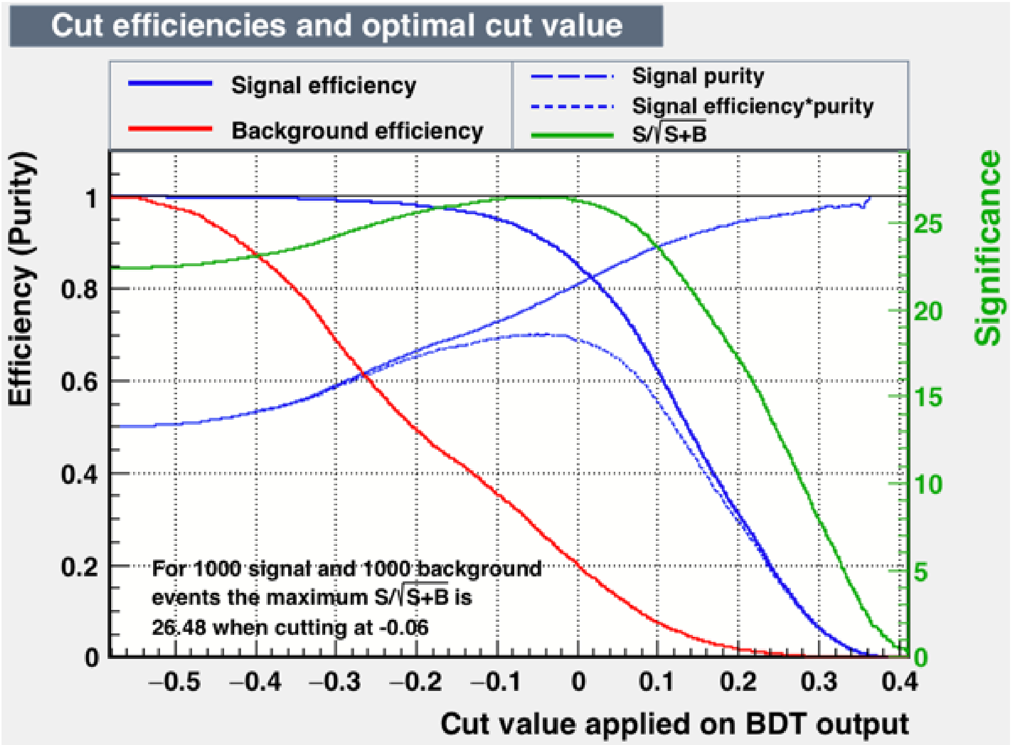
Linear correlation coefficients in %

	njet	jet_pt	jet_E	jet_Width	nsecVtx
nsecVtx	4	-2	-1	7	100
jet_Width	22	-48	-37	100	7
jet_E	-29	69	100	-37	-1
jet_pt	-39	100	69	-48	-2
njet	100	-39	-29	22	4



TMVA overtraining check for classifier: BDT

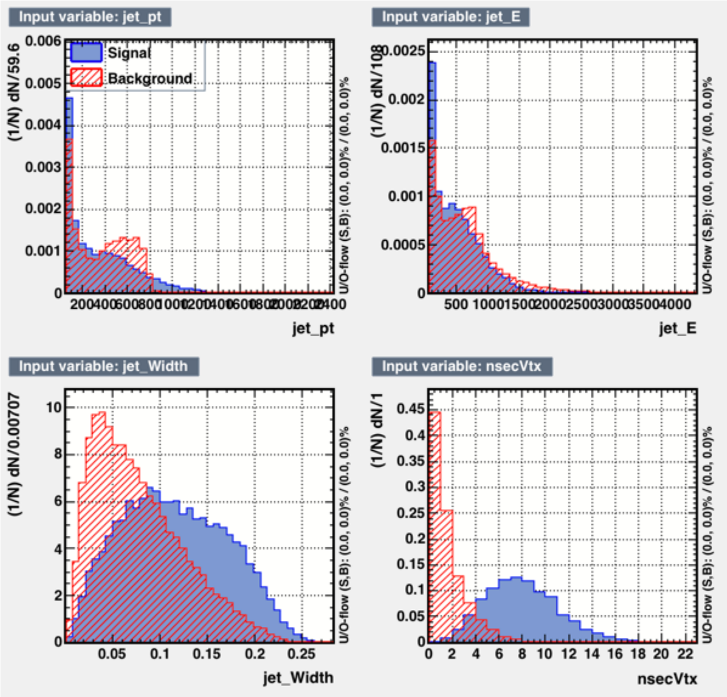
Kolmogorov-Smirnov test: signal (background) probability = 0.076 (0.9)

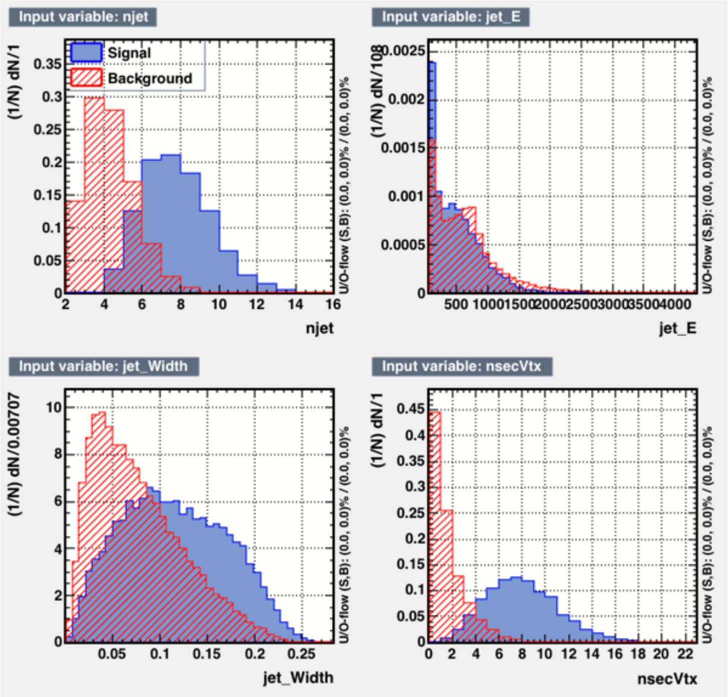


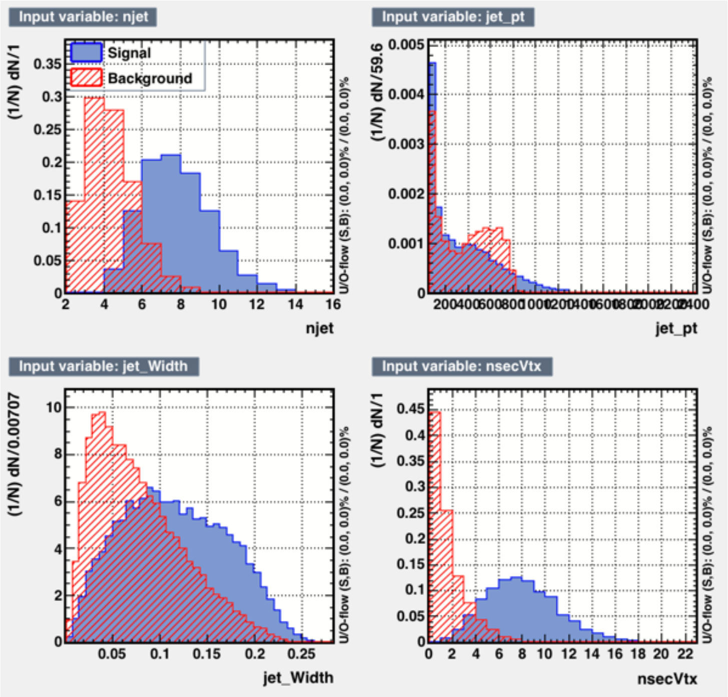
Cut efficiencies and optimal cut value

For 1000 signal and 1000 background events the maximum $S/\sqrt{S+B}$ is 26.48 when cutting at -0.06

Now try removing variables…again with Model A 1400 20

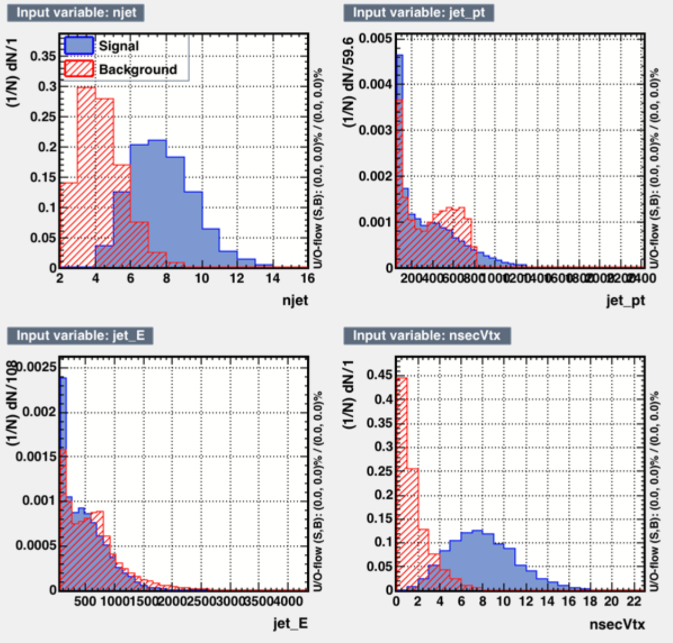


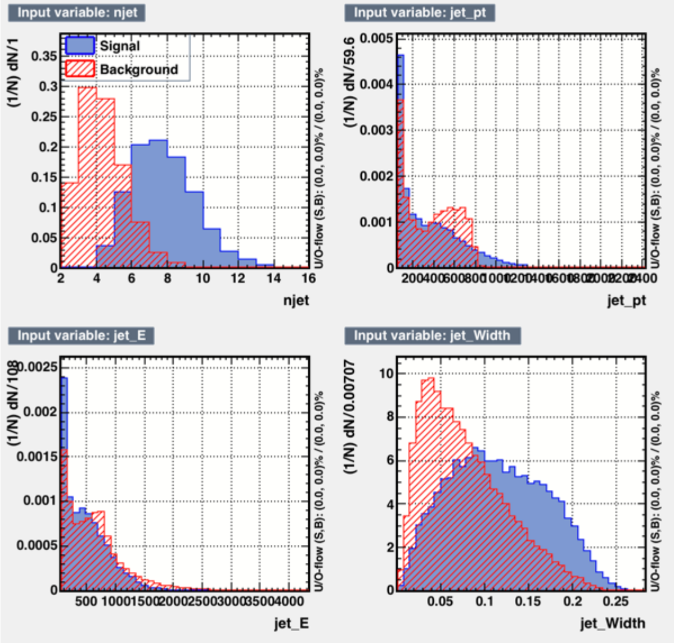




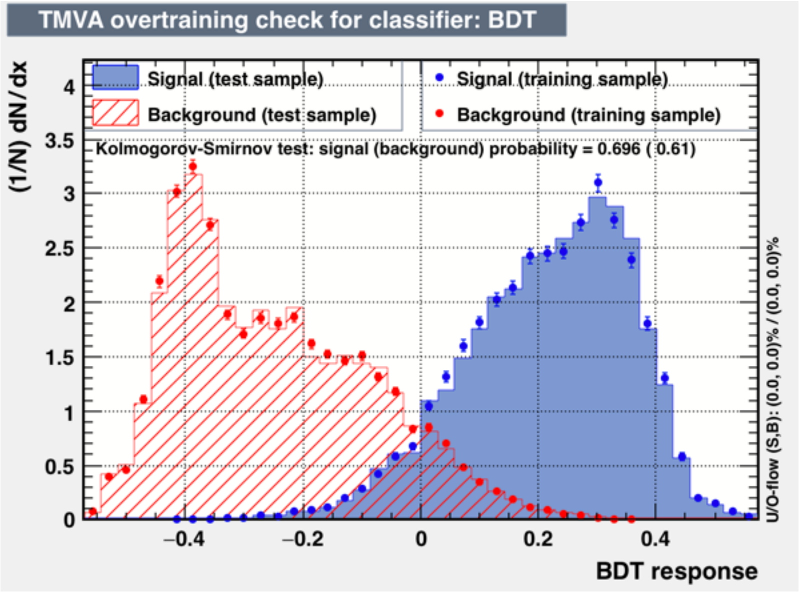
Drop:

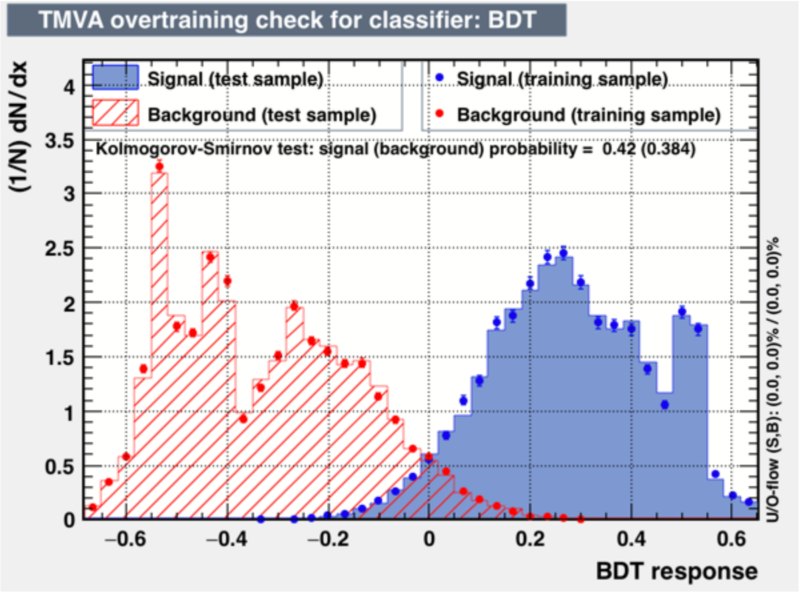
1. njet
2. jet_pt
3. jet_E
4. jet_Width
5. nsecVtx

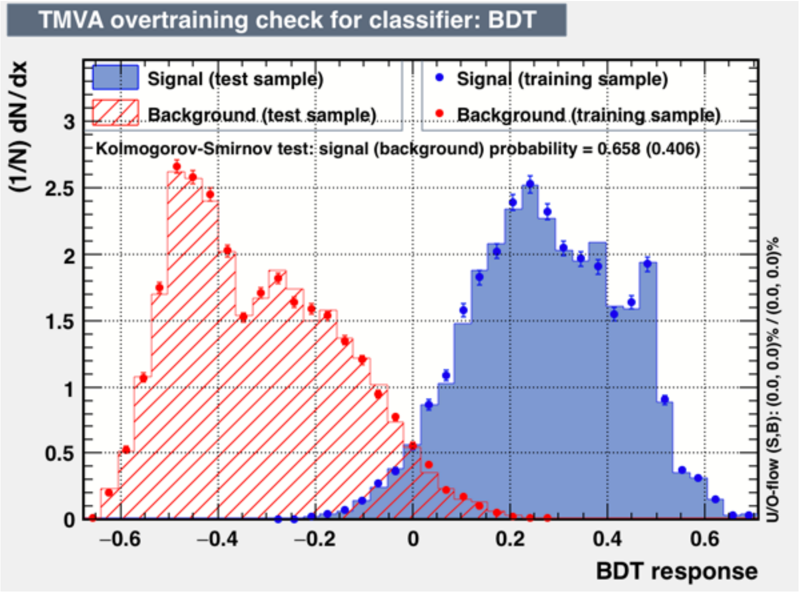




Now try removing variables…again with Model A 1400 20

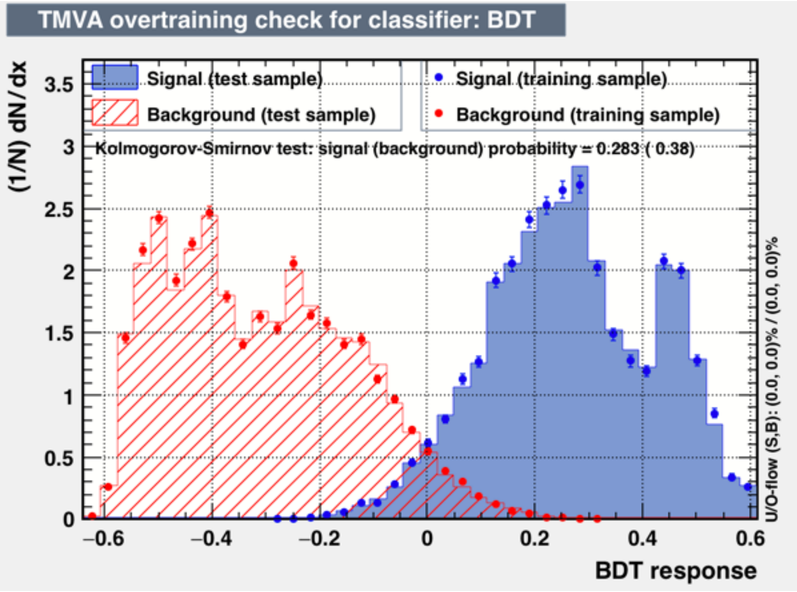


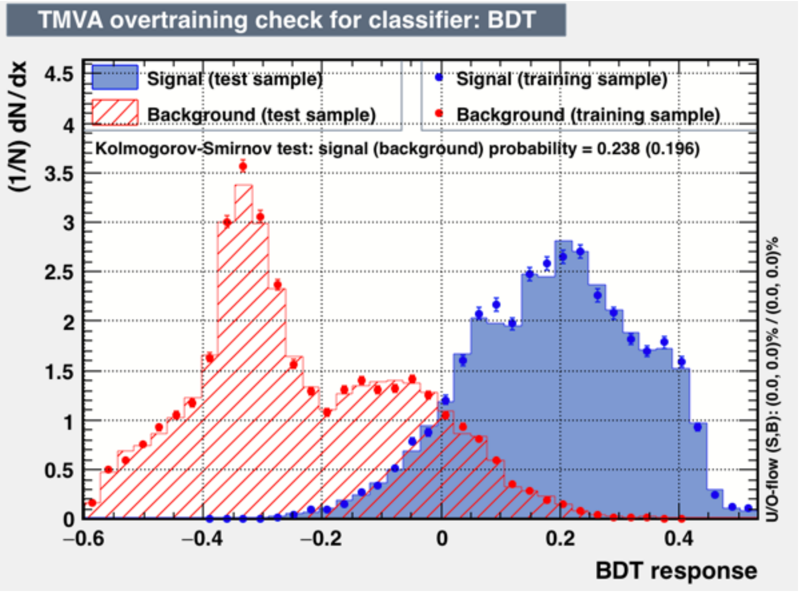




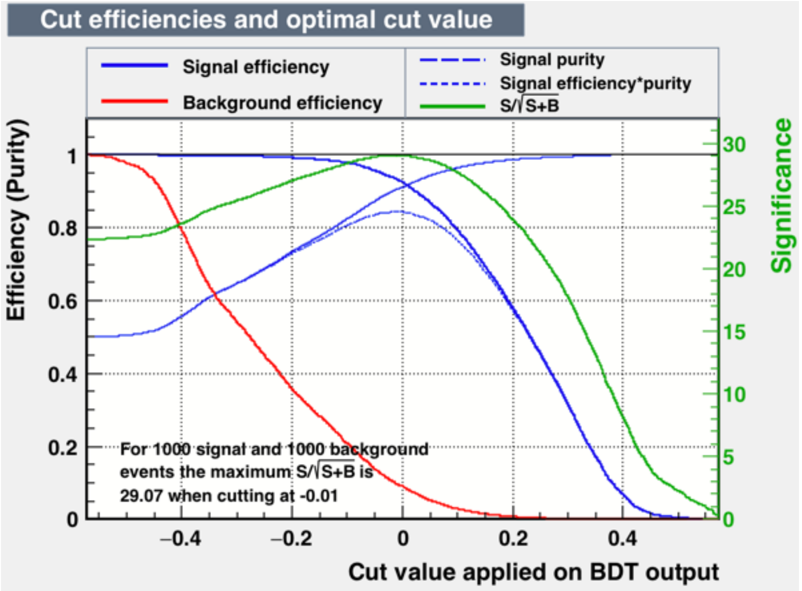
Drop:

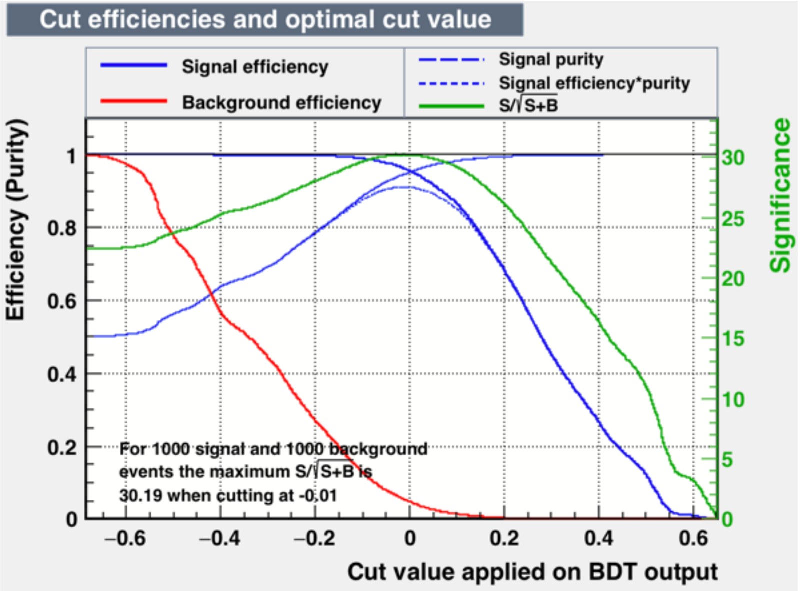
1. njet
2. jet_pt
3. jet_E
4. jet_Width
5. nsecVtx

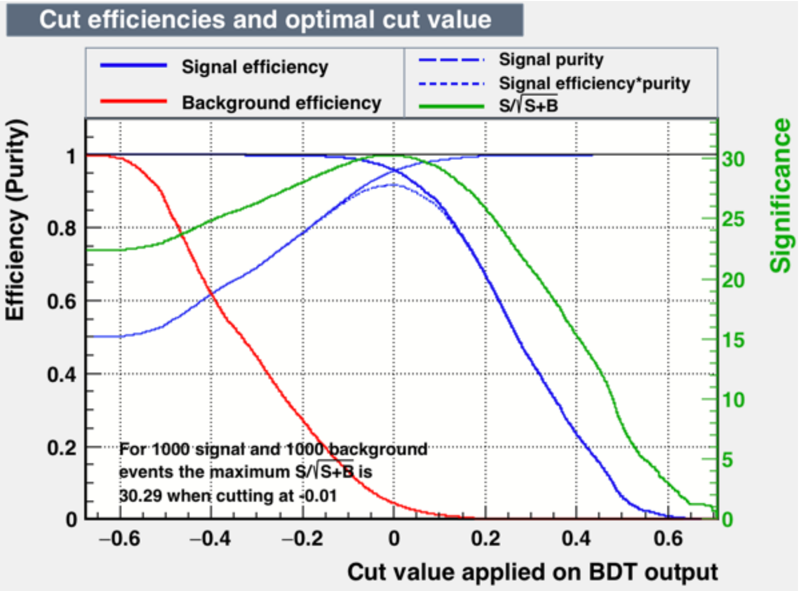




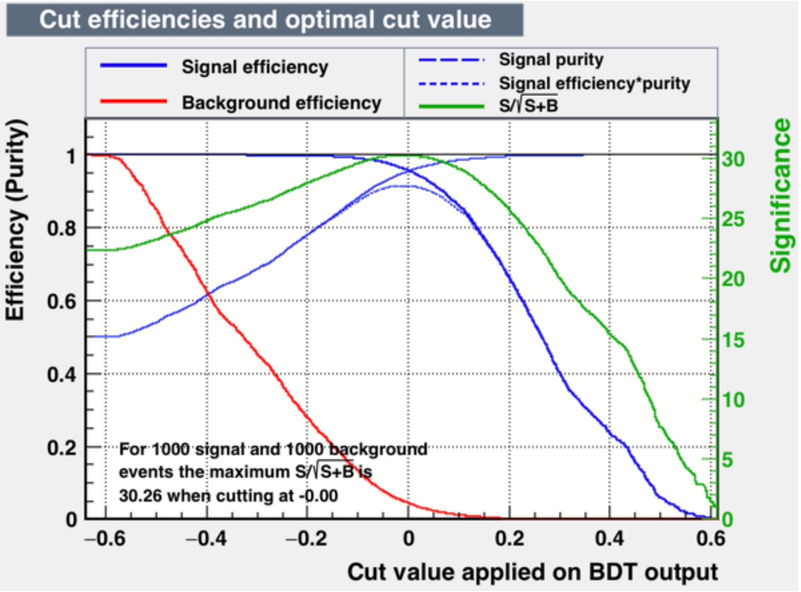
Now try removing variables…again with Model A 1400 20

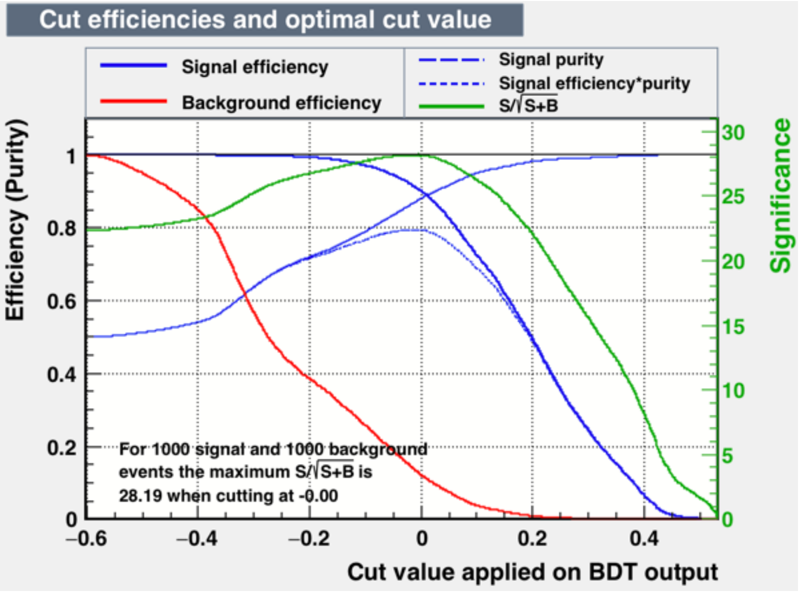




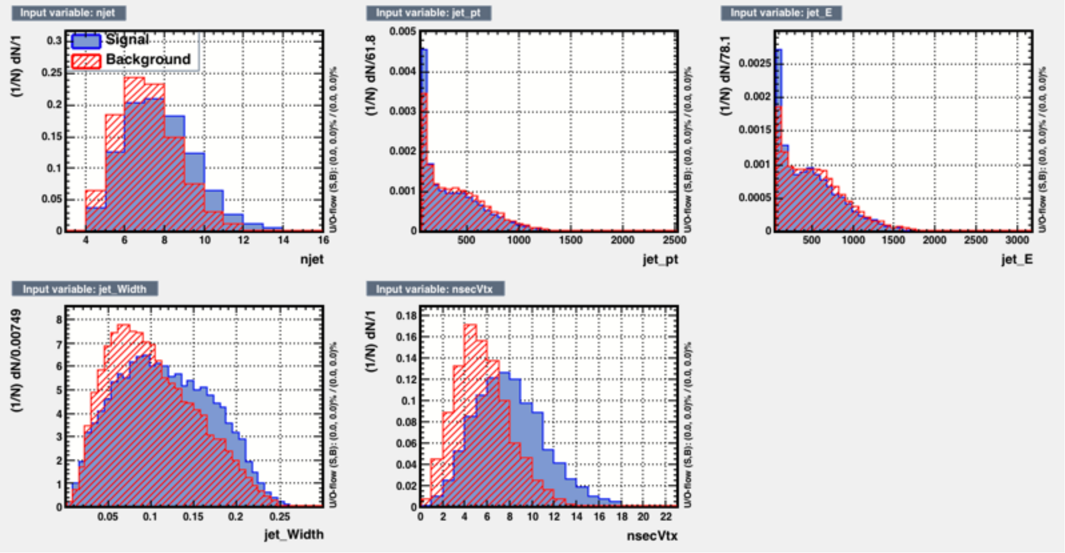


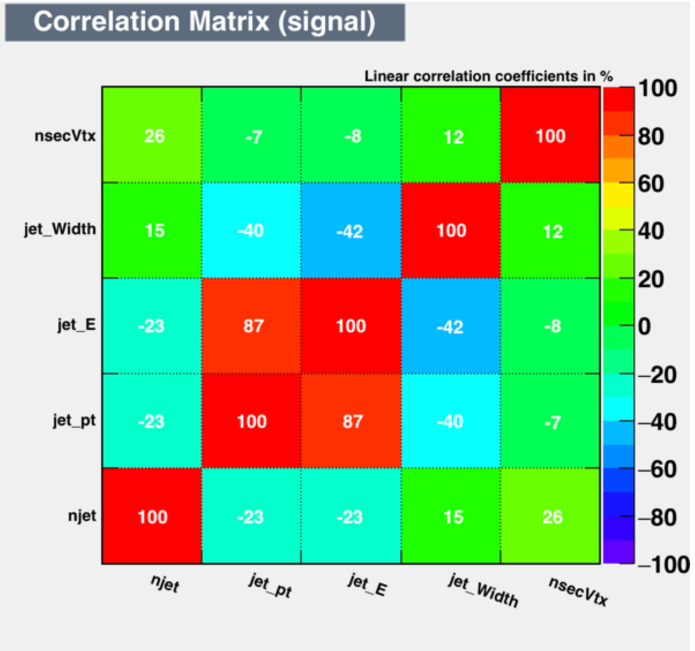
Drop:
1. njet
2. jet_pt
3. jet_E
4. jet_Width
5. nsecVtx





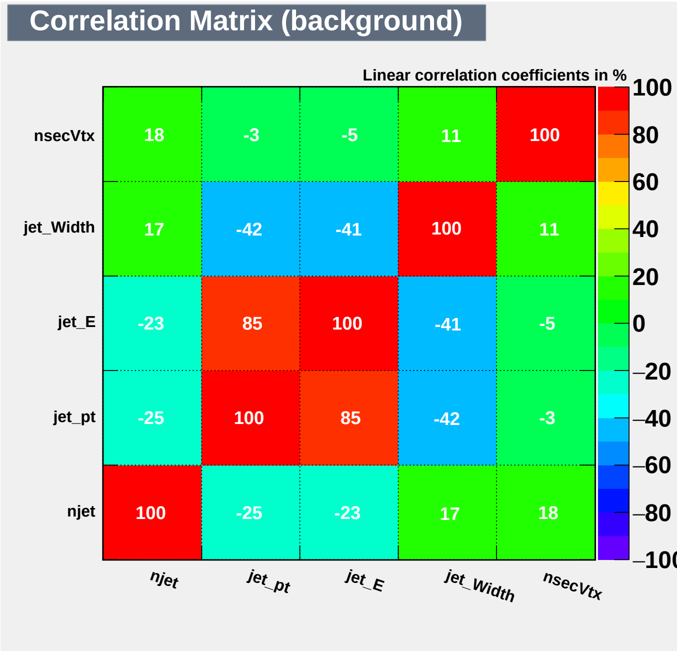
Try A_1400_20 against B_1400_20




Correlation Matrix (signal)

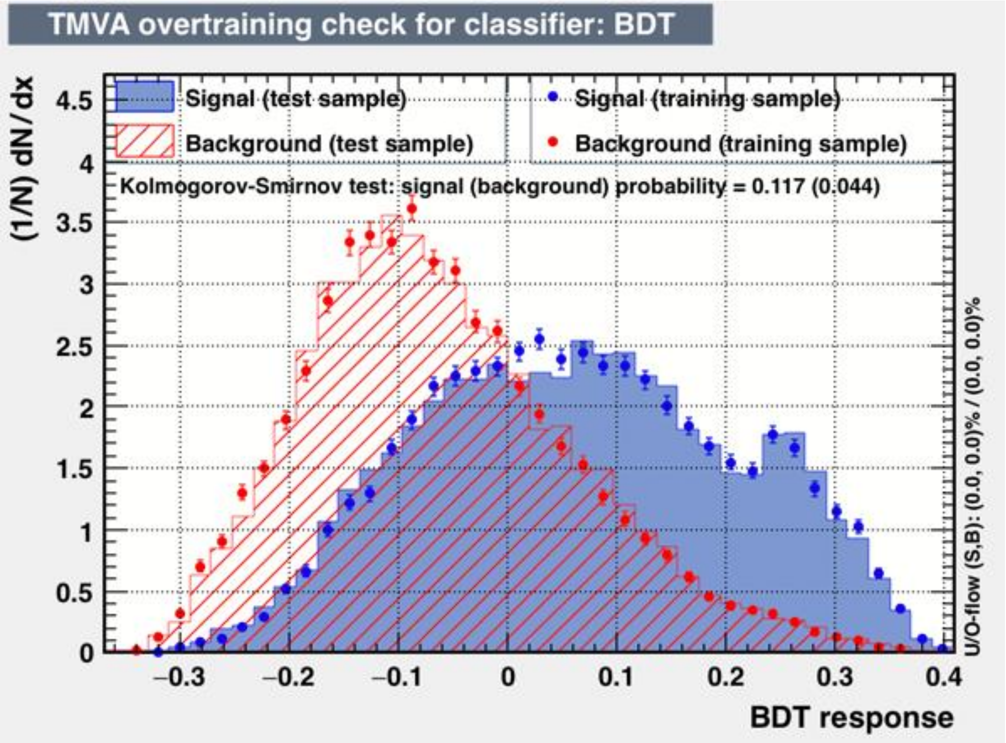
Linear correlation coefficients in %

	njet	jet_pt	jet_E	jet_Width	nsecVtx
nsecVtx	26	-7	-8	12	100
jet_Width	15	-40	-42	100	12
jet_E	-23	87	100	-42	-8
jet_pt	-23	100	87	-40	-7
njet	100	-23	-23	15	26

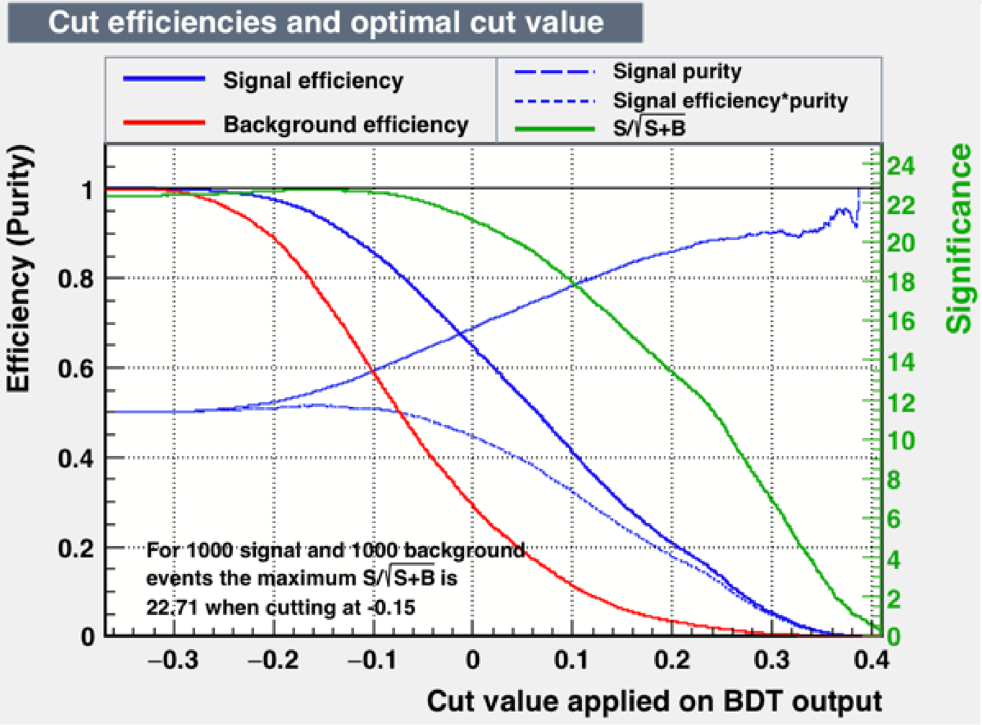

Correlation Matrix (background)

Linear correlation coefficients in %

	njet	jet_pt	jet_E	jet_Width	nsecVtx
nsecVtx	18	-3	-5	11	100
jet_Width	17	-42	-41	100	11
jet_E	-23	85	100	-41	-5
jet_pt	-25	100	85	-42	-3
njet	100	-25	-23	17	18


TMVA overtraining check for classifier: BDT

Kolmogorov-Smirnov test: signal (background) probability = 0.117 (0.044)


Cut efficiencies and optimal cut value

For 1000 signal and 1000 background events the maximum $S/\sqrt{S+B}$ is 22.71 when cutting at -0.15

Comments

- wrong to treat variables as event variables…just a test
  - plan to have sub-jet-level variables, jet-variables, event variables…multi-step approach
  - also need to apply pre-selection cuts and correct event weightings
  - MC statistics are on the low side

- use as a tool to test sensitivity of variables…but need to be aware of modelling fidelity
  - need to connect to sensitivity as benchmark…maybe S/sqrt(S+B) OK though?

- can check variable correlations

- can isolate events with ‘distinct’ BDT space features…should map to reconstructed space features

- easy to try alternate TMVA implementations and compare (e.g. NN, Fischer…)

- how to test systematics? re-training with shifted distributions…but not necessarily easy to do..

- model dependence of selection is a challenge

- **I think this can be useful…but needs the implementation needs to go up in sophistication a few orders of magnitude…**