



山东大学
SHANDONG UNIVERSITY

Optimization of Flavor Tagging Performance with LCFIPlus at CEPC

Tong LI Supervisor: Lianliang MA
Shandong University

2017.03.27

OUTLINE

- Introduction
- Default performance
- Optimization and comparison
 - Adding 2 variables
 - Adding all of the non-constant variables
- Summary

Introduction

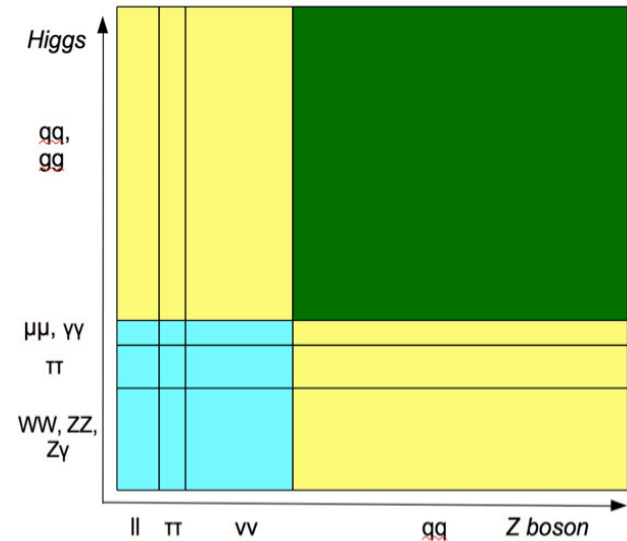
Identification of jets is essential for HEP data analysis.

On CEPC, the dominant Higgs production process is $e^+e^- \rightarrow ZH$.

The final states contained many kinds of jets.

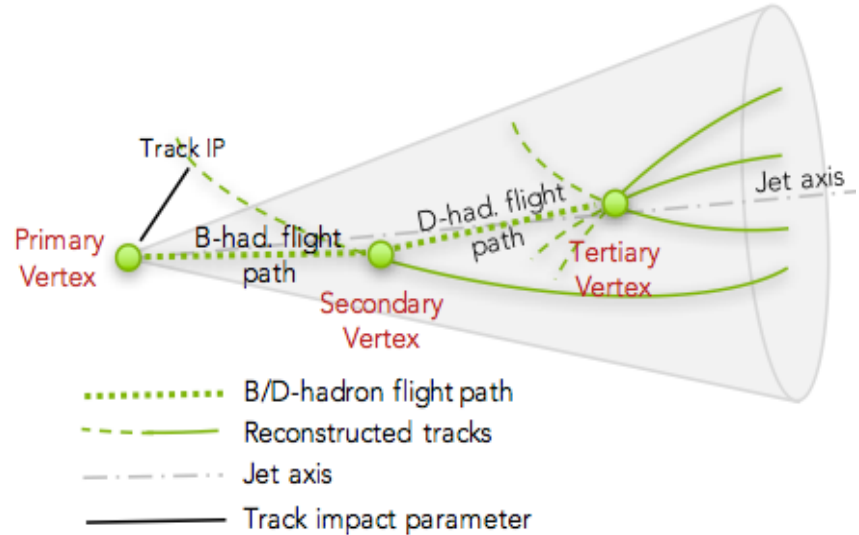
Process	Cross section (fb)	Nevents in 5 ab^{-1}
Higgs boson production		
$e^+e^- \rightarrow ZH$	209	1×10^8
$e^+e^- \rightarrow \nu\bar{\nu}H$	6.9	3.5×10^4
$e^+e^- \rightarrow e^+e^-H$	0.6	3.0×10^3

[G. LI]



Introduction - jet

Jet is a collimated bunch of hadrons flying roughly in the same direction.



We will take the advantages of the information of vertices, tracks and impact parameters.

Introduction - samples

bb-91.2GeV.root
cc-91.2GeV.root
qq-91.2GeV.root

Category	A	B	C	D
Number of vertices	0	1	1	2
Number of single-track pseudovertrices	0-2	0	1	0

3 samples are used, each of them has 396k entries, we divided them into 4 categories according to vertex.

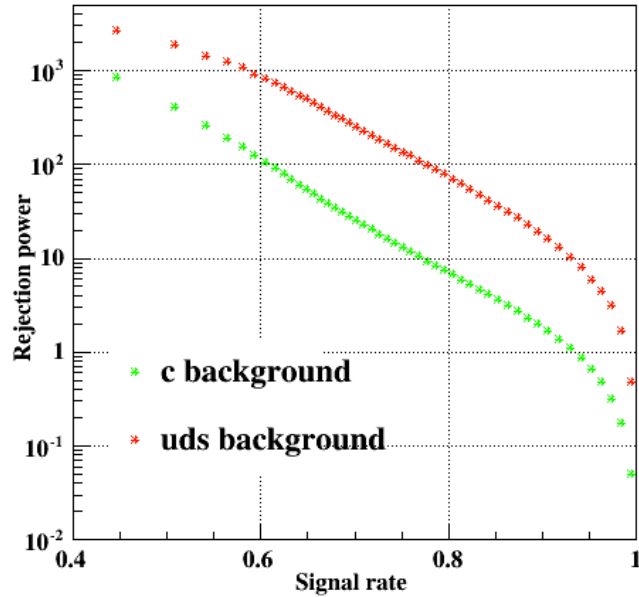
LCFIPlus is a framework for jet analysis. BDT in TMVA is used in this work.

Default performance - variables

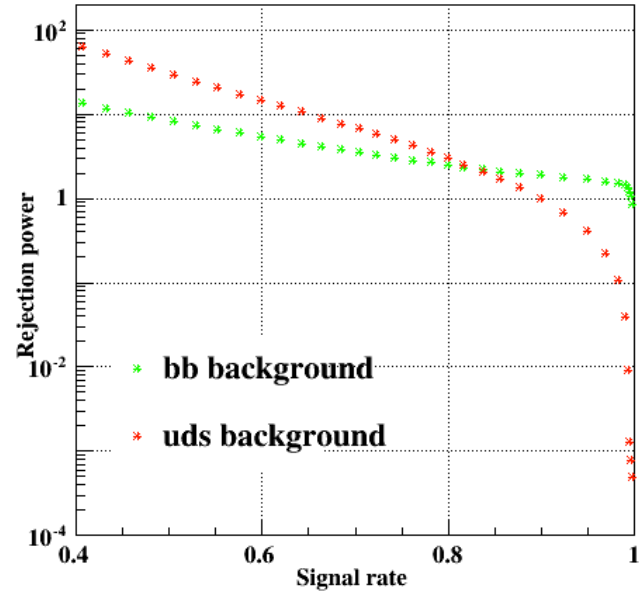
Name	Description	Normalization factor	Used by category
trk1d0sig	d0 significance of track with highest d0 significance	1	A, B, C, D
trk2d0sig	d0 significance of track with second highest d0 significance	1	A, B, C, D
trk1z0sig	z0 significance of track with highest d0 significance	1	A, B, C, D
trk2z0sig	z0 significance of track with second highest d0 significance	1	A, B, C, D
trk1pt	transverse momentum of track with highest d0 significance	$1/E_{jet}$	A, B, C, D
trk2pt	transverse momentum of track with second highest d0 significance	$1/E_{jet}$	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

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)	$1/E_{jet}$	B, C, D
vtxlen2	decay length of the second vertex in the jet (zero if number of vertex is less than two)	$1/E_{jet}$	D
vtxlen12	distance between the first and second vertex (zero if number of vertex is less than two)	$1/E_{jet}$	D
vtxsig1	decay length significance of the first vertex in the jet (zero if no vertex is found)	$1/E_{jet}$	B, C, D
vtxsig2	decay length significance of the second vertex in the jet (zero if number of vertex is less than two)	$1/E_{jet}$	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	$1/E_{jet}$	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	$1/E_{jet}$	B, C, D
vtxmom2	magnitude of the vector sum of the momenta of all tracks combined into the second vertex	$1/E_{jet}$	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

Default performance

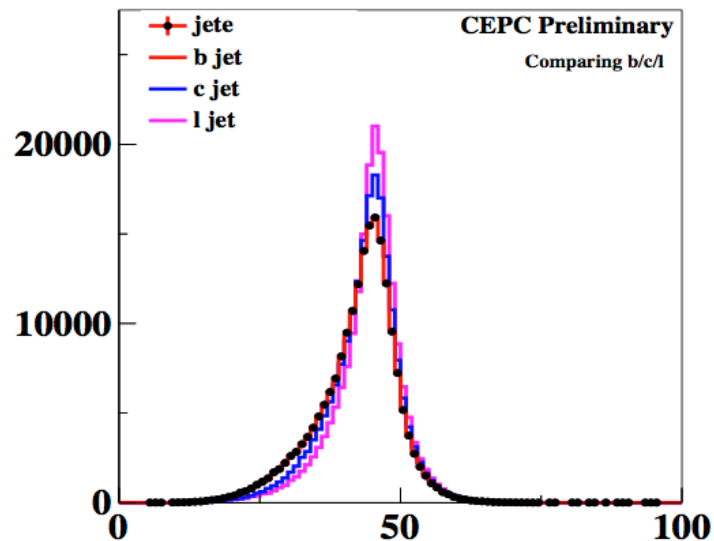
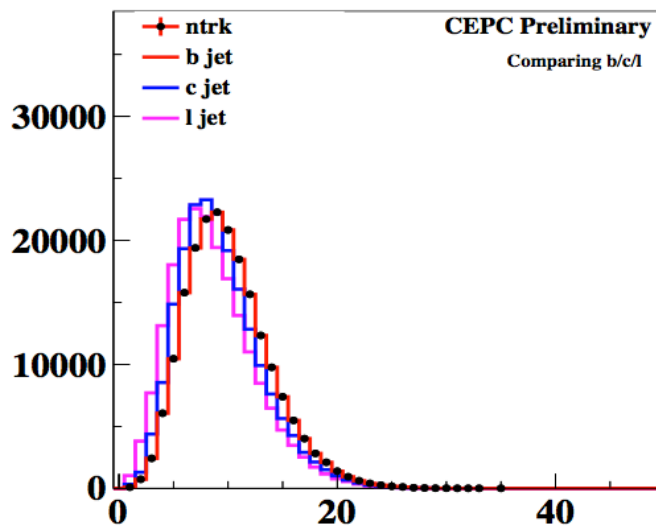


b tag



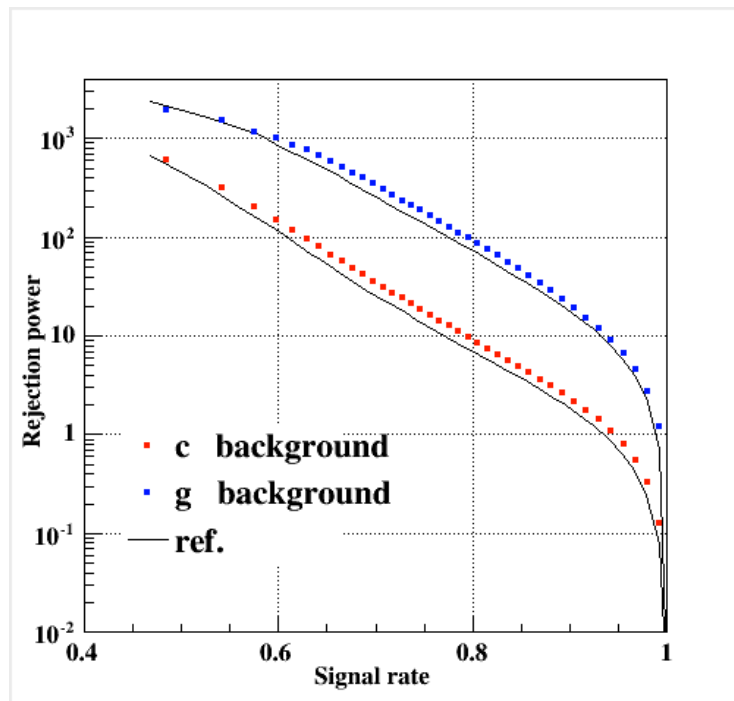
c tag

Optimization – adding 2 variables

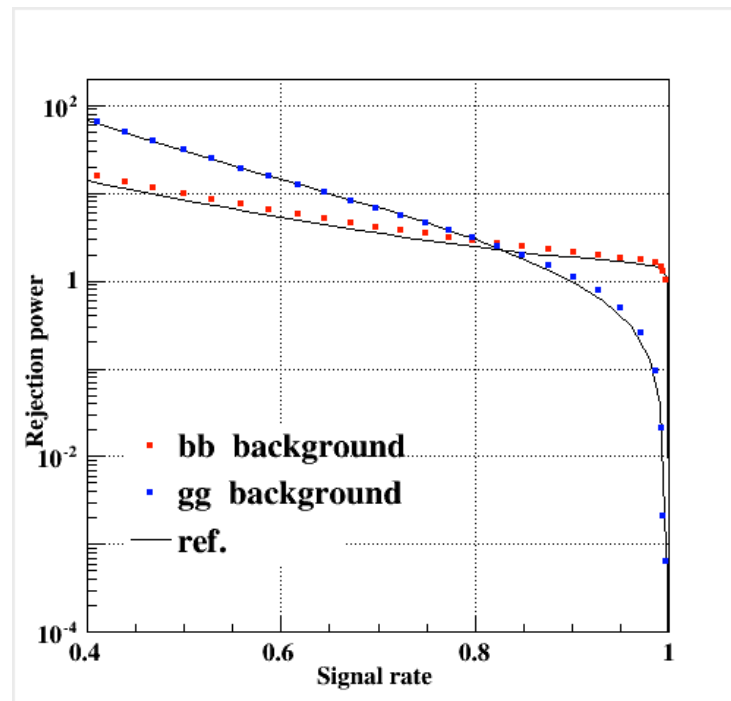


Distributions of variables ntrk and jete.
Then added them into each of the 4 categories.

Optimization - performance



b tag



c tag

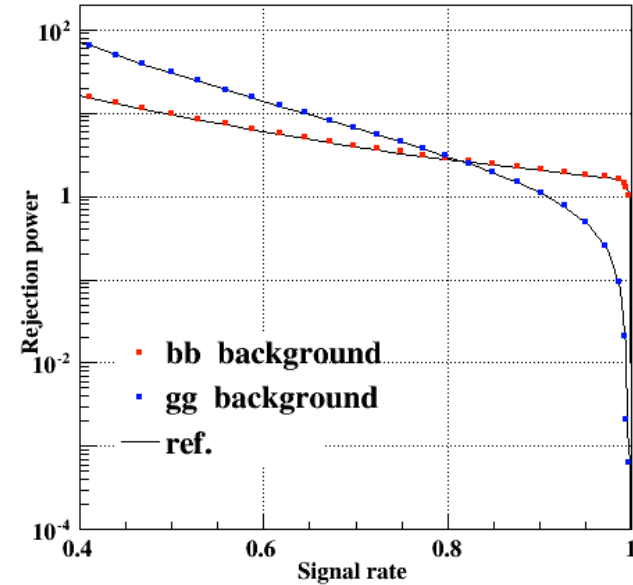
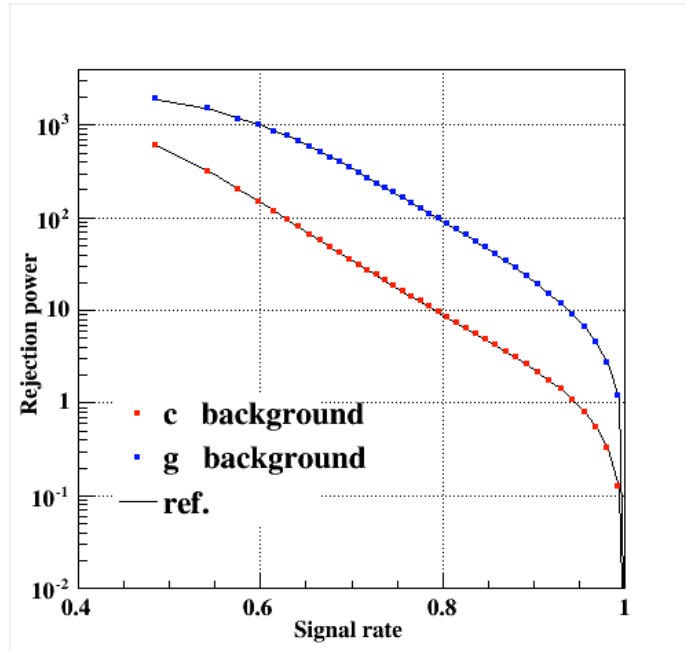
Comparison - improvement

b tag sig. rate	with c-jet			with light-jet		
	rej. pow.	refer.	improv.	rej. pow.	refer.	improv.
0.92	1.72	1.24	39%	14.82	11.71	26.60%
0.88	3	2.41	24.50%	27.76	23.59	17.70%
0.81	8.24	6.43	28.10%	83.92	67.4	24.50%
0.78	12.18	8.54	47.60%	122.62	89.93	36.40%
0.74	20.42	14.81	37.90%	204.39	196.93	33.20%
0.72	26.54	19.56	35.70%	253.99	196.93	29%
0.69	40.16	29.32	37%	383.49	288.83	32.80%
0.64	77.31	58.73	31.60%	641.51	521.25	25.60%
0.6	144.11	119.67	20.40%	975.61	887.3	10%
(reference is the default performance)						

c tag sig. rate	with b-jet			with light-jet		
	rej. pow.	refer.	improv.	rej. pow.	refer.	improv.
0.9	2.05	1.87	9.60%	1.07	0.98	9.50%
0.85	2.38	2.13	11.70%	1.92	1.85	3.80%
0.82	2.57	2.29	12.20%	2.43	2.4	1.20%
0.8	2.79	2.47	13%	3.01	2.97	1.30%
0.75	3.31	2.9	14.10%	4.48	4.55	1.50%
0.36	21.37	17.44	22.50%	104.24	99.73	4.50%
0.33	25.77	20.95	23%	143.15	136.56	4.80%
0.3	30.74	25.48	20.60%	192.58	192.12	0.20%
0.27	37.91	31.68	19.70%	287.09	278.76	3%
(reference is the default performance)						

Great improvements of rejection power have been achieved especially for b tagging.

Optimization - adding all non-constant vars



Reference is the performance after adding 2 variables.

All non-constant variables in 4 categories

Cate. A

```
trk1d0sig trk2d0sig trk1z0sig trk2z0sig trk1pt_jete trk2pt_jete jprobr5sigma jprobrz5sigma  
d0bprob d0cprob d0qprob z0bprob z0cprob z0qprob nmuon nelectron trkmass  
jete ntrkwithoutv0 vtxmassall nvtxall  
jetcosine jeteta jprobr jprobrz
```

Cate. B

```
trk1d0sig trk2d0sig trk1z0sig trk2z0sig trk1pt_jete trk2pt_jete jprobr jprobrz  
vtxlen1_jete vtxsig1_jete vtxdirang1_jete vtxmom1_jete vtxmass1 vtxmult1 vtxmasspc vtxprob  
d0bprob d0cprob d0qprob z0bprob z0cprob z0qprob  
trkmass nelectron nmuon  
jete ntrkwithoutv0 ntrk vtxmassall vtxmult  
Fd jeteta jetcosine jprobr5sigma sphericity trk1pt trk2pt vtxmass vtxmom1 vtxmom vtxmom_jete  
vtxsig1
```

Cate. C

```
trk1d0sig trk2d0sig trk1z0sig trk2z0sig trk1pt_jete trk2pt_jete jprobr jprobrz  
vtxlen1_jete vtxsig1_jete vtxdirang1_jete vtxmom1_jete vtxmass1 vtxmult1 vtxmasspc vtxprob  
1vtxprob vtxlen12all_jete vtxmassall  
jete ntrkwithoutv0 ntrk vtxmass vtxmult  
Fd d0bprob d0cprob d0qprob jetcosine jeteta jprobr5sigma jprobrz5sigma nelectron nmuon sphericity  
trk1pt trk2pt trkmass vtxlen12all vtxlen1 vtxmom1 vtxmom vtxsig1 z0bprob z0cprob
```

Cate. D

```
trk1d0sig trk2d0sig trk1z0sig trk2z0sig trk1pt_jete trk2pt_jete  
vtxlen1_jete vtxsig1_jete vtxdirang1_jete vtxmom1_jete vtxmass1 vtxmult1 vtxmasspc vtxprob  
vtxlen2_jete vtxsig2_jete vtxdirang2_jete vtxmom2_jete vtxmass2 vtxmult2  
vtxlen12_jete vtxsig12_jete vtxdirang12_jete vtxmom_jete vtxmass vtxmult  
1vtxprob  
jete ntrkwithoutv0 ntrk vtxmassall  
Fd d0bprob d0cprob d0qprob jetcosine jeteta jprobr5sigma jprobr jprobrz5sigma jprobrz nelectron nmuon  
sphericity trk1pt trk2pt trkmass vtxdirang12 vtxlen12 vtxlen12all vtxlen12all_jete vtxlen1 vtxlen2  
vtxmom1 vtxmom2 vtxmom vtxsig12 vtxsig1 vtxsig2 z0bprob z0cprob
```

Comparison - improvement

b tag	with c-jet			with light-jet		
	rej. pow.	refer.	improv.	rej. pow.	refer.	improv.
0.86	4.23	4.09	3.40%	37.95	39.26	-3.30%
0.82	7.34	7.2	1.90%	72.1	72.68	-0.80%
0.78	12.08	12.18	-0.80%	119.94	122.62	-2.20%
0.72	25.7	26.54	-3.20%	253.66	253.99	0.00%
0.69	38.91	40.16	-3.10%	375.34	383.49	-2.10%
0.64	87.43	77.31	13.10%	730.54	641.51	13.90%
0.6	140.71	144.11	-2.40%	985.47	975.61	1.00%
0.58	190.31	193.89	-1.80%	1175.64	1134.59	3.60%
0.49	556.98	586.71	-5.10%	1932.88	1895.3	2.00%

(reference is the performance after adding 2 variables)

c tag	with b-jet			with light-jet		
	rej. pow.	refer.	improv.	rej. pow.	refer.	improv.
0.88	2.27	2.21	2.70%	1.5	1.46	2.70%
0.85	2.45	2.38	2.90%	1.94	1.92	1%
0.8	2.88	2.79	3.20%	3.05	3.01	1.30%
0.75	3.42	3.31	3.30%	4.53	4.48	1.10%
0.7	4.13	3.99	3.50%	6.51	6.53	-0.30%
0.62	5.65	5.59	1.10%	11.52	11.96	-3.70%
0.36	21.96	21.37	2.80%	129.56	104.24	24.00%
0.3	32.56	30.74	5.90%	270.6	192.58	40.50%
0.25	50.21	48.02	4.60%	582.05	430.17	35.30%

(reference is the performance after adding 2 variables)

Great performance have been obtained for c-tag especially when the signal rate is less than 0.36
The slight decrease is mainly because of statistical fluctuation.

Summary

- Great improvement is achieved compared with default performance by adding variables.
- Appropriate variables should be selected for b or c tag to satisfy different experimental requirements.
- New variables?



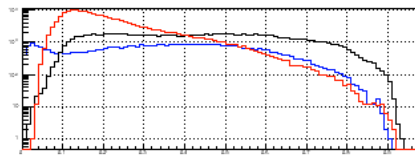
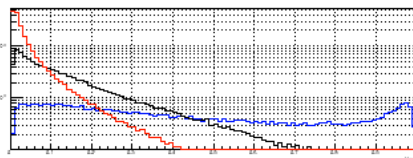
BACK UP

Optimization - BDT output

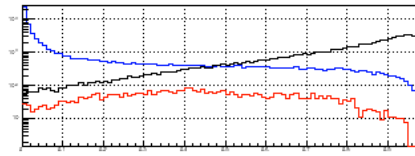
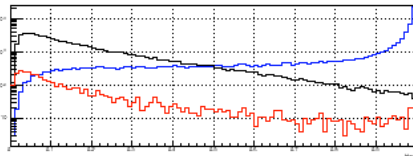
b tag

c tag

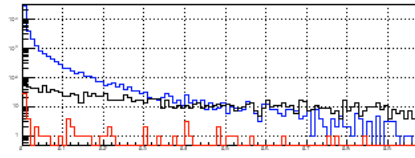
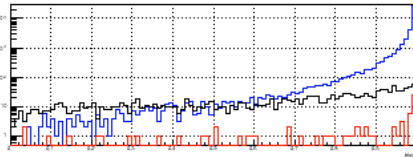
Cate. A



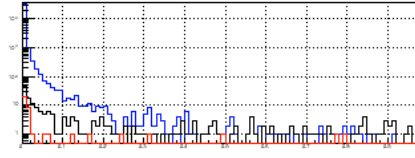
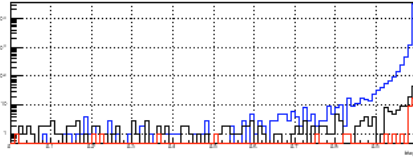
Cate. B



Cate. C

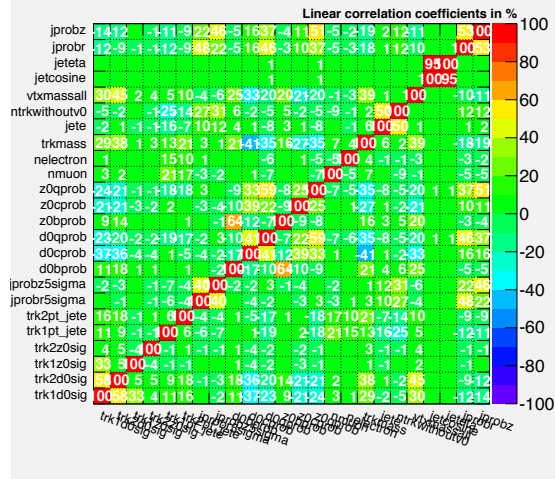


Cate. D

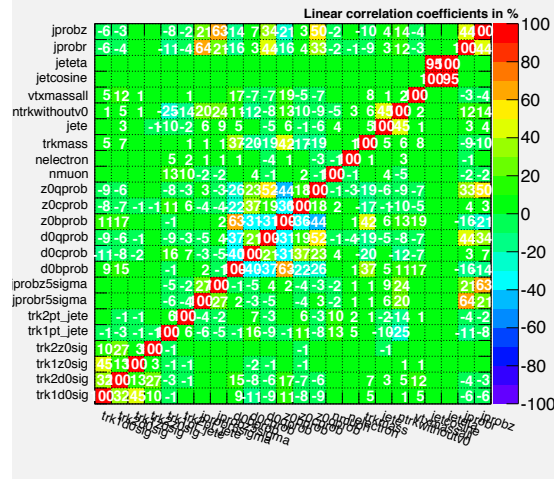


Optimization - Correlation Matrix

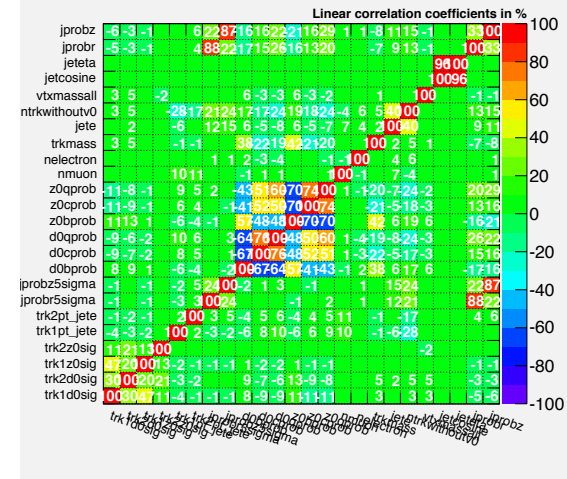
Correlation Matrix (jetB)



Correlation Matrix (jetC)



Correlation Matrix (jetO)



Variable importance index of category A

```
:  
: Rank : Variable      : Variable Importance  
:-----  
: 1 : jete              : 8.402e-02  
: 2 : z0qprob          : 7.203e-02  
: 3 : d0qprob          : 7.034e-02  
: 4 : trk2d0sig        : 6.787e-02  
: 5 : jprobr5sigma     : 6.195e-02  
: 6 : trkmass          : 6.112e-02  
: 7 : jeteta           : 5.576e-02  
: 8 : ntrkwithoutv0   : 5.303e-02  
: 9 : jprobz5sigma     : 4.765e-02  
: 10 : jetcosine        : 4.722e-02  
: 11 : z0cprob          : 4.567e-02  
: 12 : trk2z0sig        : 4.023e-02  
: 13 : trk1d0sig        : 3.727e-02  
: 14 : nelectron        : 3.417e-02  
: 15 : trk1z0sig        : 3.347e-02  
: 16 : d0cprob          : 3.080e-02  
: 17 : trk2pt_jete     : 2.936e-02  
: 18 : vtxmassall       : 2.804e-02  
: 19 : jprobz           : 2.766e-02  
: 20 : trk1pt_jete     : 2.535e-02  
: 21 : z0bprob          : 1.738e-02  
: 22 : d0bprob          : 1.133e-02  
: 23 : nmuon            : 9.980e-03  
: 24 : jprobr           : 8.299e-03  
: 25 : nvtxall          : 0.000e+00  
:-----
```

Variable
importance
index of
category B

1	: vtxmassall	: 8.074e-02	:	1	: vtxmassall	: 7.048e-02
2	: vtxmom1	: 6.208e-02	:	2	: vtxmasspc	: 5.044e-02
3	: jete	: 5.873e-02	:	3	: vtxdirang1_jete	: 4.679e-02
4	: d0qprob	: 5.405e-02	:	4	: trkmass	: 4.545e-02
5	: jprobr5sigma	: 5.041e-02	:	5	: vtxmom1	: 4.304e-02
6	: trkmass	: 3.839e-02	:	6	: vtxlen1	: 4.195e-02
7	: z0cprob	: 3.767e-02	:	7	: d0cprob	: 4.166e-02
8	: d0cprob	: 3.677e-02	:	8	: jete	: 4.074e-02
9	: z0qprob	: 3.499e-02	:	9	: lvtxprob	: 3.494e-02
10	: vtxmass1	: 3.464e-02	:	10	: vtxlen12all	: 3.435e-02
11	: vtxdirang1_jete	: 3.396e-02	:	11	: d0qprob	: 3.398e-02
12	: Fd	: 3.266e-02	:	12	: z0cprob	: 3.395e-02
13	: vtxsig1	: 3.217e-02	:	13	: jprobz	: 3.200e-02
14	: vtxmult1	: 3.167e-02	:	14	: jprobr5sigma	: 3.175e-02
15	: vtxprob	: 3.074e-02	:	15	: ntrkwithoutv0	: 2.582e-02
16	: jprobz	: 3.071e-02	:	16	: trk2d0sig	: 2.535e-02
17	: trk2d0sig	: 3.014e-02	:	17	: jprobr	: 2.357e-02
18	: nelectron	: 2.513e-02	:	18	: jprobz5sigma	: 2.281e-02
19	: jetcosine	: 2.466e-02	:	19	: trk1d0sig	: 2.222e-02
20	: vtxmom1_jete	: 2.363e-02	:	20	: vtxlen1_jete	: 2.222e-02
21	: jeteta	: 2.169e-02	:	21	: vtxmult1	: 2.184e-02
22	: trk2pt	: 2.131e-02	:	22	: jeteta	: 2.152e-02
23	: trk1pt	: 1.952e-02	:	23	: vtxmom1_jete	: 2.132e-02
24	: trk2z0sig	: 1.705e-02	:	24	: vtxmass1	: 2.114e-02
25	: trk1d0sig	: 1.488e-02	:	25	: vtxlen12all_jete	: 2.091e-02
26	: ntrk	: 1.465e-02	:	26	: jetcosine	: 1.886e-02
27	: ntrkwithoutv0	: 1.396e-02	:	27	: Fd	: 1.793e-02
28	: vtxlen1_jete	: 1.153e-02	:	28	: nmuon	: 1.601e-02
29	: trk1pt_jete	: 1.135e-02	:	29	: sphericity	: 1.488e-02
30	: trk1z0sig	: 1.113e-02	:	30	: trk2pt	: 1.483e-02
31	: d0bprob	: 1.023e-02	:	31	: vtxprob	: 1.298e-02
32	: trk2pt_jete	: 1.013e-02	:	32	: trk1pt	: 9.382e-03
33	: vtxsig1_jete	: 9.679e-03	:	33	: vtxsig1	: 9.246e-03
34	: z0bprob	: 9.075e-03	:	34	: trk2z0sig	: 8.138e-03
35	: nmuon	: 8.088e-03	:	35	: trk2pt_jete	: 7.956e-03
36	: jprobr	: 6.813e-03	:	36	: z0bprob	: 7.419e-03
37	: sphericity	: 4.956e-03	:	37	: vtxsig1_jete	: 6.828e-03
38	: vtxmassall	: 0.000e+00	:	38	: ntrk	: 6.588e-03
39	: vtxmult	: 0.000e+00	:	39	: d0bprob	: 5.648e-03
40	: vtxmass	: 0.000e+00	:	40	: trk1pt_jete	: 5.642e-03
41	: vtxmom	: 0.000e+00	:	41	: trk1z0sig	: 4.717e-03
42	: vtxmom_jete	: 0.000e+00	:	42	: nelectron	: 2.691e-03
			:	43	: vtxmass	: 0.000e+00
			:	44	: vtxmult	: 0.000e+00
			:	45	: vtxmom	: 0.000e+00

Variable
importance
index of
category C

1	: vtxmasspc	: 1.049e-01	32	: vtxmass1	: 8.990e-03
2	: vtxmass	: 9.437e-02	33	: jprobrz5sigma	: 7.919e-03
3	: trkmass	: 4.622e-02	34	: vtxprob	: 7.686e-03
4	: vtxdirang12_jete	: 4.449e-02	35	: vtxlen2_jete	: 7.241e-03
5	: d0qprob	: 4.392e-02	36	: vtxmult1	: 7.031e-03
6	: vtxdirang2_jete	: 4.273e-02	37	: trk1d0sig	: 7.019e-03
7	: vtxdirang1_jete	: 4.198e-02	38	: vtxlen2	: 6.749e-03
8	: vtxdirang12	: 3.751e-02	39	: d0bprob	: 6.677e-03
9	: vtxmom_jete	: 3.342e-02	40	: vtxlen1_jete	: 5.995e-03
10	: jete	: 2.983e-02	41	: sphericity	: 5.814e-03
11	: vtxmom	: 2.930e-02	42	: Fd	: 5.619e-03
12	: vtxmom1_jete	: 2.676e-02	43	: vtxsig1	: 5.468e-03
13	: jeteta	: 2.409e-02	44	: trk1pt_jete	: 5.267e-03
14	: jetcosine	: 2.297e-02	45	: vtxsig1_jete	: 3.908e-03
15	: vtxsig12_jete	: 2.119e-02	46	: trk2pt_jete	: 3.717e-03
16	: vtxmom2	: 2.094e-02	47	: trk2pt	: 3.549e-03
17	: vtxmom1	: 2.070e-02	48	: ntrk	: 3.174e-03
18	: trk2d0sig	: 1.788e-02	49	: trk1pt	: 2.833e-03
19	: vtxmult	: 1.680e-02	50	: vtxsig2	: 2.579e-03
20	: ntrkwithoutv0	: 1.632e-02	51	: trk1z0sig	: 2.542e-03
21	: vtxlen12	: 1.562e-02	52	: z0bprob	: 2.502e-03
22	: z0cprob	: 1.540e-02	53	: nelectron	: 2.463e-03
23	: d0cprob	: 1.468e-02	54	: trk2z0sig	: 2.189e-03
24	: vtxsig12	: 1.462e-02	55	: vtxmult2	: 4.485e-04
25	: vtxlen12_jete	: 1.380e-02	56	: vtxmassall	: 0.000e+00
26	: vtxsig2_jete	: 1.360e-02	57	: jprobr	: 0.000e+00
27	: vtxmass2	: 1.322e-02	58	: jprobrz	: 0.000e+00
28	: vtxlen1	: 1.244e-02	59	: nmuon	: 0.000e+00
29	: 1vtxprob	: 1.195e-02	60	: vtxlen12all	: 0.000e+00
30	: jprobr5sigma	: 1.115e-02	61	: vtxlen12all_jete	: 0.000e+00
31	: vtxmom2_jete	: 9.833e-03			
32	: vtxmass1	: 8.990e-03			

Variable importance index of category D