



# Updates of the Flavor Tagging study with Z sample and Higgs sample at the CEPC

Tong LI, Lianliang MA

Shandong University

2017.06.05

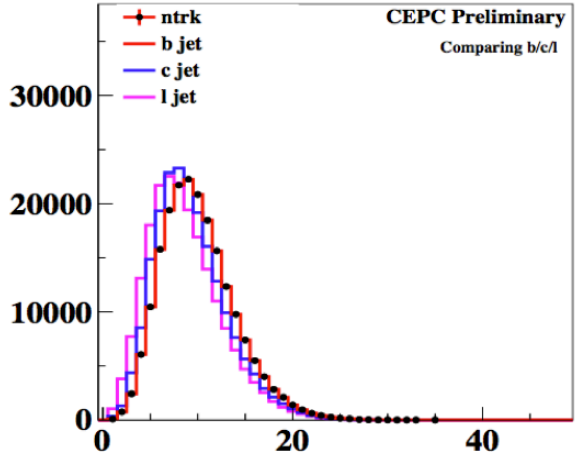
# Outline

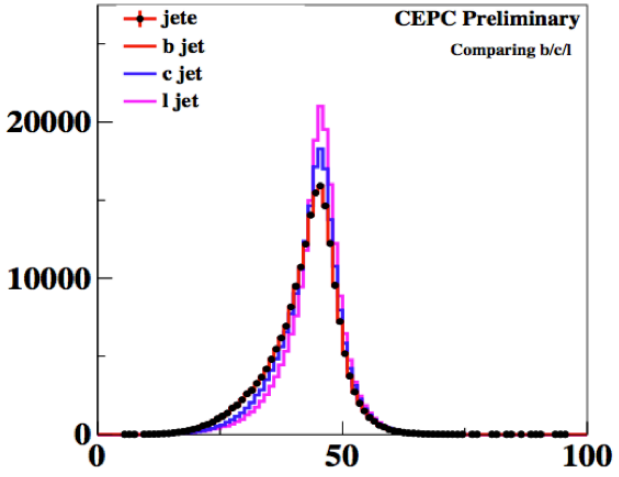
- Introduction
- Training with Z sample and checking performance with Higgs sample
- Training with Higgs sample directly and checking performance

  - bb, cc and ss  - bb, cc and ss+GG  - bb, cc and GG

- Summary

# Introduction




Distribution of variable ntrk and jete. We choose them to optimize the flavor tagging performance for Z sample in last talk.

Reconstructed algorithm: Vertex Finding, Jet Clustering and Jet Vertex Refiner.

http://indico.ihep.ac.cn/event/6877/contribution/8/material/slides/0.pdf

# Introduction – comparison

For b tag, the improvement of c jet and light jet rejection power is around 20%-50% and 10%-35% respectively versus the signal rate from 0.6 to 0.92.

For c tag, we can improve about 10%-20% for b jet rejection power. The improvement for light jet rejection power can be negligible.

| b tag | with c-jet | | | with light-jet | | |
|---|---|---|---|---|---|---|
| sig. rate | 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% |
| 0.78 | 12.18 | 8.54 | 47.60% | 122.62 | 89.93 | 36.40% |
| 0.69 | 40.16 | 29.32 | 37% | 383.49 | 288.83 | 32.80% |
| 0.6 | 144.11 | 119.67 | 20.40% | 975.61 | 887.3 | 10% |
| (reference is the default performance) | | | | | | |

| c tag | with b-jet | | | with light-jet | | |
|---|---|---|---|---|---|---|
| sig. rate | rej. pow. | refer. | improv. | rej. pow. | refer. | improv. |
| 0.9 | 2.05 | 1.87 | 9.60% | 1.07 | 0.98 | 9.50% |
| 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% |
| 0.27 | 37.91 | 31.68 | 19.70% | 287.09 | 278.76 | 3% |
| (reference is the default performance) | | | | | | |

# Checking performance with Higgs sample

```
zpole_v00_c0_bdt.class.C
zpole_v00_c0_bdt.weights.xml
zpole_v00_c0.root
zpole_v00_c1_bdt.class.C
zpole_v00_c1_bdt.weights.xml
zpole_v00_c1.root
zpole_v00_c2_bdt.class.C
zpole_v00_c2_bdt.weights.xml
zpole_v00_c2.root
zpole_v00_c3_bdt.class.C
zpole_v00_c3_bdt.weights.xml
zpole_v00_c3.root
```

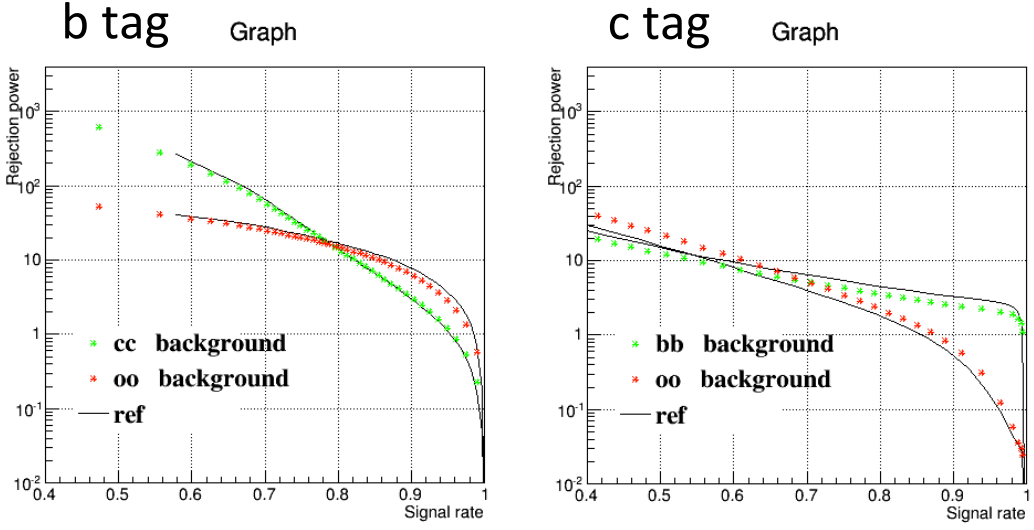
```
nnh_bb.01.001.slcio  nnh_bb.01.059.slcio  nnh_cc.01.017.slcio  nnh_cc.01.075.slcio
nnh_bb.01.002.slcio  nnh_bb.01.060.slcio  nnh_cc.01.018.slcio  nnh_cc.01.076.slcio
nnh_bb.01.003.slcio  nnh_bb.01.061.slcio  nnh_cc.01.019.slcio  nnh_cc.01.077.slcio
nnh_bb.01.004.slcio  nnh_bb.01.062.slcio  nnh_cc.01.020.slcio  nnh_cc.01.078.slcio
nnh_bb.01.005.slcio  nnh_bb.01.063.slcio  nnh_cc.01.021.slcio  nnh_cc.01.079.slcio
nnh_bb.01.006.slcio  nnh_bb.01.064.slcio  nnh_cc.01.022.slcio  nnh_cc.01.080.slcio
nnh_bb.01.007.slcio  nnh_bb.01.065.slcio  nnh_cc.01.023.slcio  nnh_cc.01.081.slcio
nnh_bb.01.008.slcio  nnh_bb.01.066.slcio  nnh_cc.01.024.slcio  nnh_cc.01.082.slcio
nnh_bb.01.009.slcio  nnh_bb.01.067.slcio  nnh_cc.01.025.slcio  nnh_cc.01.083.slcio
nnh_bb.01.010.slcio  nnh_bb.01.068.slcio  nnh_cc.01.026.slcio  nnh_cc.01.084.slcio
nnh_bb.01.011.slcio  nnh_bb.01.069.slcio  nnh_cc.01.027.slcio  nnh_cc.01.085.slcio
nnh_bb.01.012.slcio  nnh_bb.01.070.slcio  nnh_cc.01.028.slcio  nnh_cc.01.086.slcio
nnh_bb.01.013.slcio  nnh_bb.01.071.slcio  nnh_cc.01.029.slcio  nnh_cc.01.087.slcio
nnh_bb.01.014.slcio  nnh_bb.01.072.slcio  nnh_cc.01.030.slcio  nnh_cc.01.088.slcio
nnh_bb.01.015.slcio  nnh_bb.01.073.slcio  nnh_cc.01.031.slcio  nnh_cc.01.089.slcio
nnh_bb.01.016.slcio  nnh_bb.01.074.slcio  nnh_cc.01.032.slcio  nnh_cc.01.090.slcio
nnh_bb.01.017.slcio  nnh_bb.01.075.slcio  nnh_cc.01.033.slcio  nnh_cc.01.091.slcio
nnh_bb.01.018.slcio  nnh_bb.01.076.slcio  nnh_cc.01.034.slcio  nnh_cc.01.092.slcio
nnh_bb.01.019.slcio  nnh_bb.01.077.slcio  nnh_cc.01.035.slcio  nnh_cc.01.093.slcio
nnh_bb.01.020.slcio  nnh_bb.01.078.slcio  nnh_cc.01.036.slcio  nnh_cc.01.094.slcio
nnh_bb.01.021.slcio  nnh_bb.01.079.slcio  nnh_cc.01.037.slcio  nnh_cc.01.095.slcio
nnh_bb.01.022.slcio  nnh_bb.01.080.slcio  nnh_cc.01.038.slcio  nnh_cc.01.096.slcio
nnh_bb.01.023.slcio  nnh_bb.01.081.slcio  nnh_cc.01.039.slcio  nnh_cc.01.097.slcio
nnh_bb.01.024.slcio  nnh_bb.01.082.slcio  nnh_cc.01.040.slcio  nnh_cc.01.098.slcio
nnh_bb.01.025.slcio  nnh_bb.01.083.slcio  nnh_cc.01.041.slcio  nnh_cc.01.099.slcio
nnh_bb.01.026.slcio  nnh_bb.01.084.slcio  nnh_cc.01.042.slcio  nnh_cc.01.100.slcio
nnh_bb.01.027.slcio  nnh_bb.01.085.slcio  nnh_cc.01.043.slcio  nnh_GG.01.001.slcio
nnh_bb.01.028.slcio  nnh_bb.01.086.slcio  nnh_cc.01.044.slcio  nnh_GG.01.002.slcio
nnh_bb.01.029.slcio  nnh_bb.01.087.slcio  nnh_cc.01.045.slcio  nnh_GG.01.003.slcio
```

Weight files created by training of Z pole sample

vvH sample

Then we will check the performance with Higgs sample.
90k events (30k for bb, 30k for cc and 30k for GG) are used.

# Checking performance – rejection power




Rejection power for different signal rates, the ref (black line) corresponds to the default weight files and the dots is corresponding to the weight files after adding jete and ntrk.

# Checking performance – comparison

Comparison of rejection power for b tag and c tag respectively.

The performance become worse compared with the default result except the light jet rejection power for c tag.

| b tag | with c-jet | | | with light-jet | | |
|---|---|---|---|---|---|---|
| sig.rate | rej.pow. | refer. | improv. | rej.pow. | refer. | improv. |
| 0.9 | 2.96 | 3.08 | -3.90% | 6.09 | 7.91 | -23.01% |
| 0.85 | 7.12 | 6.52 | 9.20% | 10.72 | 11.98 | -10.52% |
| 0.79 | 16.03 | 16.88 | -5.04% | 16.02 | 18.01 | -11.05% |
| 0.69 | 65.52 | 71.82 | -8.77% | 25.79 | 28.06 | -8.09% |
| 0.67 | 93.79 | 94.54 | -0.79% | 28.77 | 30.7 | -6.29% |

| c tag | with b-jet | | | with light-jet | | |
|---|---|---|---|---|---|---|
| sig.rate | rej.pow. | refer. | improv. | rej.pow. | refer. | improv. |
| 0.87 | 2.73 | 3.64 | -25.00% | 1.08 | 0.9 | 20.00% |
| 0.71 | 5.01 | 6.16 | -18.67% | 4.86 | 3.63 | 33.88% |
| 0.53 | 10.57 | 12.9 | -18.06% | 17.95 | 13.01 | 37.97% |
| 0.39 | 21.73 | 26.66 | -18.49% | 44.94 | 32.58 | 37.94% |
| 0.27 | 53.64 | 57.25 | -6.31% | 95.16 | 71.59 | 32.92% |

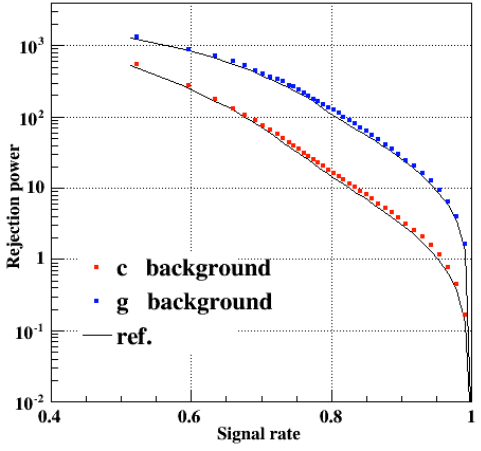
# Training with Higgs sample

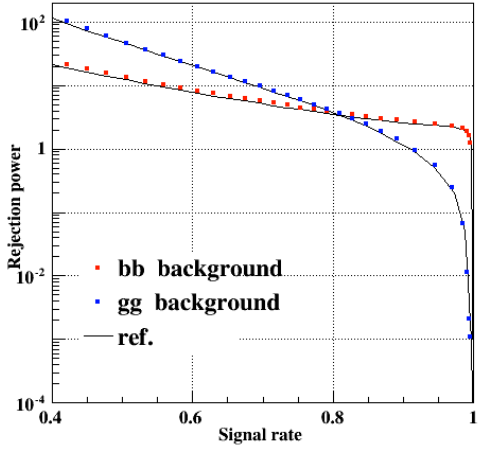
| | |
|---|---|
| bb-nnh.root | 200k entries |
| cc-nnh.root | 200k entries |
| GG-nnh.root | 200k entries |
| ss-nnh.root | 198k entries |

In vvH sample, the light jet will consist of gluon jet, which is different from Z pole sample.

# Training Higgs – bb, cc and ss

b tag




c tag

Rejection power for different signal rates, the ref. (black line) is for default variables and the dots is the result after adding jete and ntrk.

# Comparison - bb, cc and ss

The performance is similar with Z sample.

For b tag, the improvement is around 10%-25% when signal rate is large than 0.7.

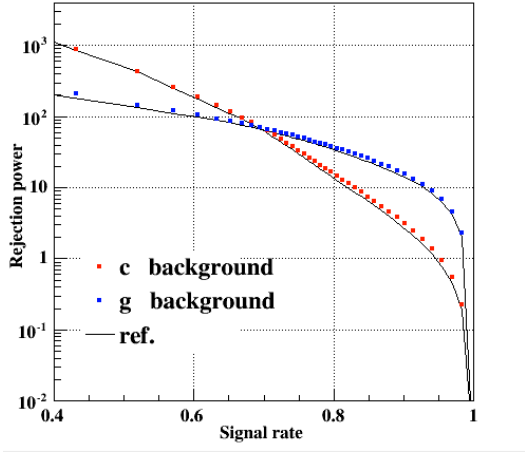
For c tag, the improvement is much less than b tag.

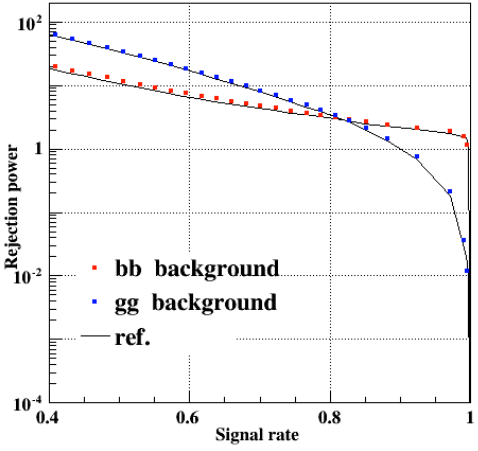
| b tag | with c-jet | | | with light-jet | | |
|---|---|---|---|---|---|---|
| sig. rate | rej. pow. | refer. | improv. | rej. pow. | refer. | improv. |
| 0.9 | 3.68 | 2.94 | 25.17% | 28.96 | 24.75 | 17.01% |
| 0.85 | 7.78 | 6.77 | 14.92% | 60.5 | 54.32 | 11.38% |
| 0.79 | 19.71 | 15.94 | 23.65% | 146.43 | 117.8 | 24.30% |
| 0.69 | 85.49 | 85.34 | 0.18% | 429.77 | 437.39 | -1.74% |
| 0.52 | 524.29 | 535.59 | -2.11% | 1285.7 | 1302.63 | -1.30% |

| c tag | with b-jet | | | with light-jet | | |
|---|---|---|---|---|---|---|
| sig. rate | rej. pow. | refer. | improv. | rej. pow. | refer. | improv. |
| 0.83 | 3.33 | 3.22 | 3.42% | 2.88 | 2.85 | 1.05% |
| 0.81 | 3.54 | 3.42 | 3.51% | 3.45 | 3.43 | 0.58% |
| 0.58 | 8.86 | 8.45 | 4.85% | 23.34 | 23.71 | -1.56% |
| 0.53 | 11.24 | 10.7 | 5.05% | 35.76 | 36.07 | -0.86% |
| 0.39 | 24.19 | 22.97 | 5.31% | 131.28 | 129.19 | 1.62% |

# Training Higgs – bb, cc and ss+GG

b tag




c tag

Rejection power for different signal rates, the ref. (black line) is for default variables and the dots is the result after adding jete and ntrk.

# Comparison - bb, cc and ss+GG

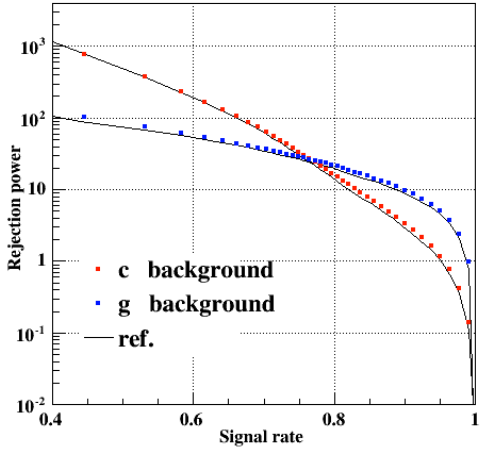
The improvement is not obvious compared with the case for bb, cc and ss.

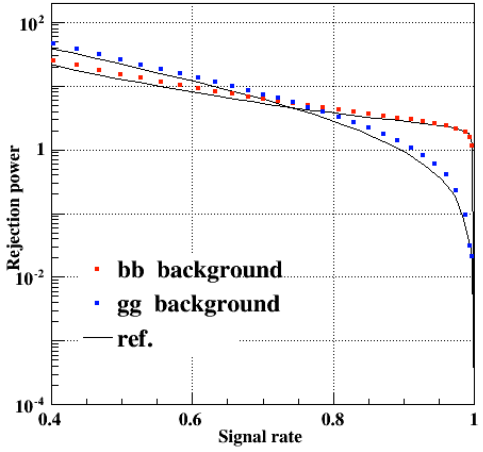
| b tag | with c-jet | | | with light-jet | | |
|---|---|---|---|---|---|---|
| sig.rate | rej.pow. | refer. | improv. | rej.pow. | refer. | improv. |
| 0.9 | 3.02 | 2.87 | 5.23% | 14.9 | 14.79 | 0.74% |
| 0.85 | 7.25 | 6.19 | 17.12% | 24.81 | 22.69 | 9.34% |
| 0.78 | 20.07 | 18.84 | 6.53% | 40.78 | 39.6 | 2.98% |
| 0.68 | 79.88 | 81.08 | -1.48% | 71.92 | 71.55 | 0.52% |
| 0.57 | 247.27 | 249.77 | -1.00% | 115.6 | 111.18 | 3.98% |

| c tag | with b-jet | | | with light-jet | | |
|---|---|---|---|---|---|---|
| sig.rate | rej.pow. | refer. | improv. | rej.pow. | refer. | improv. |
| 0.83 | 2.79 | 2.73 | 2.20% | 2.67 | 2.64 | 1.14% |
| 0.72 | 4.17 | 4.02 | 3.73% | 6.72 | 6.75 | -0.44% |
| 0.66 | 5.41 | 5.13 | 5.46% | 10.99 | 11.26 | -2.40% |
| 0.55 | 8.87 | 8.38 | 5.85% | 24.17 | 24.74 | -2.30% |
| 0.33 | 31.77 | 29.24 | 8.65% | 104.33 | 100.45 | 3.86% |

# Training Higgs – bb, cc and GG

b tag




c tag

Rejection power for different signal rates, the ref. (black line) is for default variables and the dots is the result after adding jete and ntrk.

# Comparison - bb, cc and GG

For b tag, the improvement can be negligible.

For c tag, the improvement for b jet rejection power is around 10% and for light jet rejection power it is around 10%-25%.

| b tag | with c-jet | | | with light-jet | | |
|---|---|---|---|---|---|---|
| sig. rate | rej. pow. | refer. | improv. | rej. pow. | refer. | improv. |
| 0.9 | 3.3 | 3.18 | 3.77% | 9.48 | 9.23 | 2.71% |
| 0.83 | 10 | 8.86 | 12.87% | 16.98 | 15.71 | 8.08% |
| 0.78 | 20.74 | 19.44 | 6.69% | 23.29 | 22.39 | 4.02% |
| 0.64 | 124.38 | 126.14 | -1.40% | 46.22 | 45.25 | 2.14% |
| 0.53 | 365.93 | 372.81 | -1.85% | 71.72 | 66.9 | 7.20% |

| c tag | with b-jet | | | with light-jet | | |
|---|---|---|---|---|---|---|
| sig. rate | rej. pow. | refer. | improv. | rej. pow. | refer. | improv. |
| 0.91 | 2.85 | 2.76 | 3.26% | 1.04 | 0.86 | 20.93% |
| 0.77 | 4.7 | 4.14 | 13.53% | 4.47 | 3.61 | 23.82% |
| 0.66 | 7.17 | 6.28 | 14.17% | 9.57 | 8.16 | 17.28% |
| 0.58 | 9.91 | 8.68 | 14.17% | 15.14 | 13.47 | 12.40% |
| 0.47 | 16.99 | 15.22 | 11.63% | 29.53 | 26.8 | 10.19% |

# Summary

- Flavor tagging performance of Higgs sample is different from Z sample with the same training result.
- Variables jete and ntrk are sample-dependent.
- Flavor tagging (especially for b tag) performance may get worse when light jet consists of gluon jet.
- Gluon jet is similar with b jet.

# backups

# Training with vvH sample

bb-nnh.root
cc-nnh.root
GG-nnh.root
ss-nnh.root

In vvH sample, the light jet will consist of gluon jet, which is different from Z pole sample.

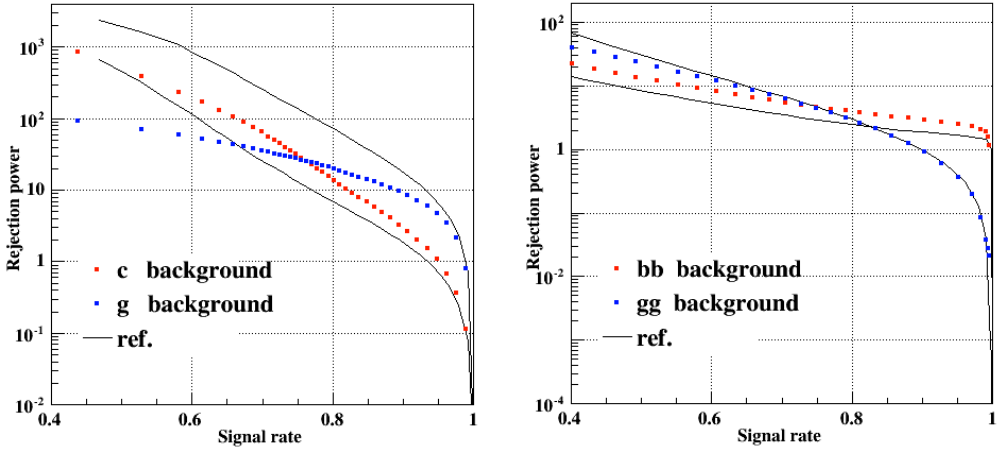
```
trk1d0sig trk2d0sig trk1z0sig trk2z0sig trk1pt_jete trk2pt_jete jprobr5sigma jprobz5sigma
d0bprob d0cprob d0qprob z0bprob z0cprob z0qprob nmuon nelectron trkmass
```

```
trk1d0sig trk2d0sig trk1z0sig trk2z0sig trk1pt_jete trk2pt_jete jprobr jprobz
vtxlen1_jete vtxsig1_jete vtxdirang1_jete vtxmom1_jete vtxmass1 vtxmult1 vtxmasspc vtxprob
d0bprob d0cprob d0qprob z0bprob z0cprob z0qprob
trkmass nelectron nmuon
```

```
trk1d0sig trk2d0sig trk1z0sig trk2z0sig trk1pt_jete trk2pt_jete jprobr jprobz
vtxlen1_jete vtxsig1_jete vtxdirang1_jete vtxmom1_jete vtxmass1 vtxmult1 vtxmasspc vtxprob
1vtxprob vtxlen12all_jete vtxmassall
```

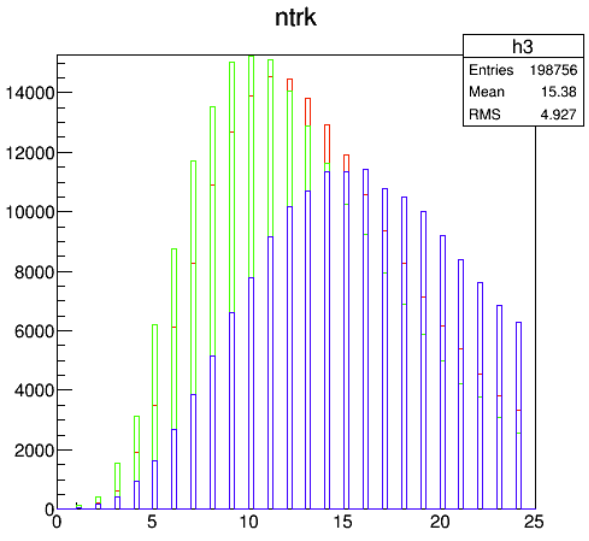
```
trk1d0sig trk2d0sig trk1z0sig trk2z0sig trk1pt_jete trk2pt_jete jprobr jprobz
vtxlen1_jete vtxsig1_jete vtxdirang1_jete vtxmom1_jete vtxmass1 vtxmult1 vtxmasspc vtxprob
vtxlen2_jete vtxsig2_jete vtxdirang2_jete vtxmom2_jete vtxmass2 vtxmult2
vtxlen12_jete vtxsig12_jete vtxdirang12_jete vtxmom_jete vtxmass vtxmult
1vtxprob
```

# Training – comparison

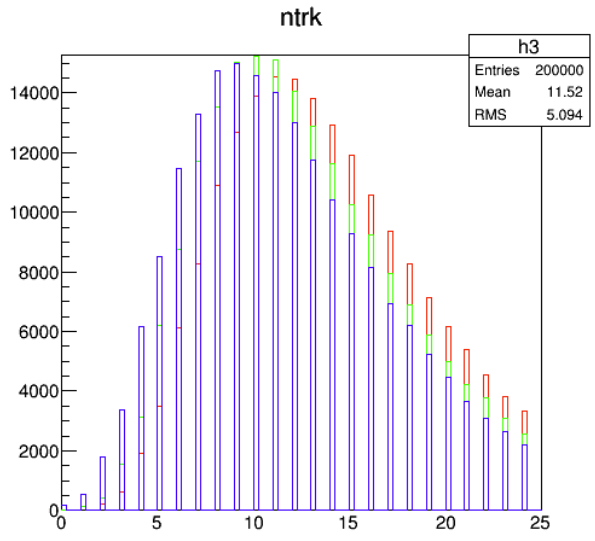



Rejection power for different signal rates, the ref. (black line) is corresponding to Z pole sample and the dots is the result of vvH training.
Default variables are used in these 2 cases.

# Variable distribution of Higgs sample

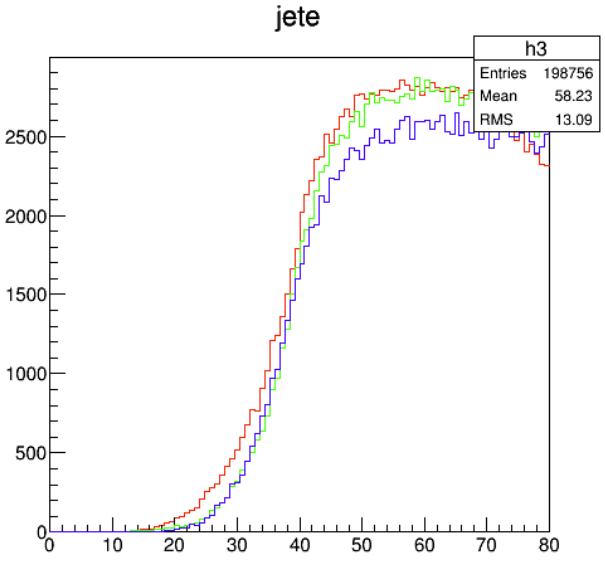


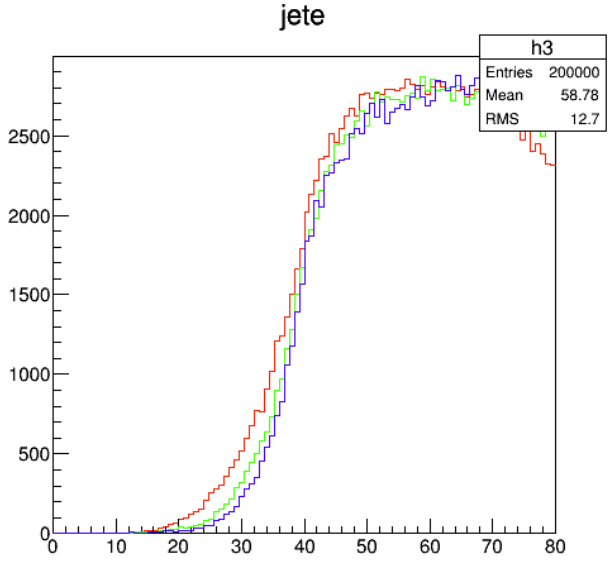

Red: b jet
Green: c jet
Blue: gluon jet




Red: b jet
Green: c jet
Blue: light jet

# Variable distribution of Higgs sample

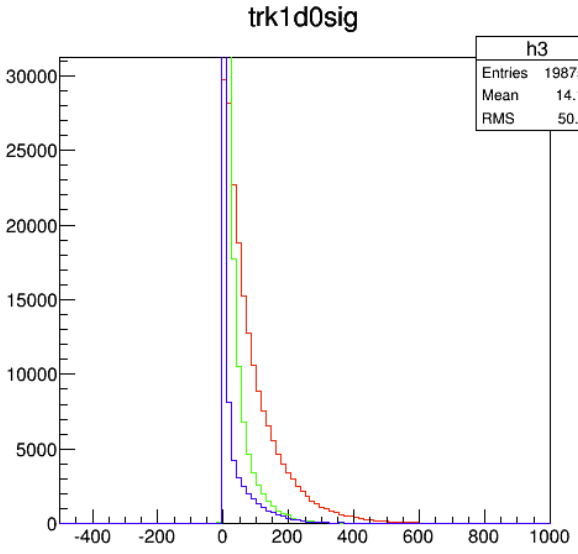


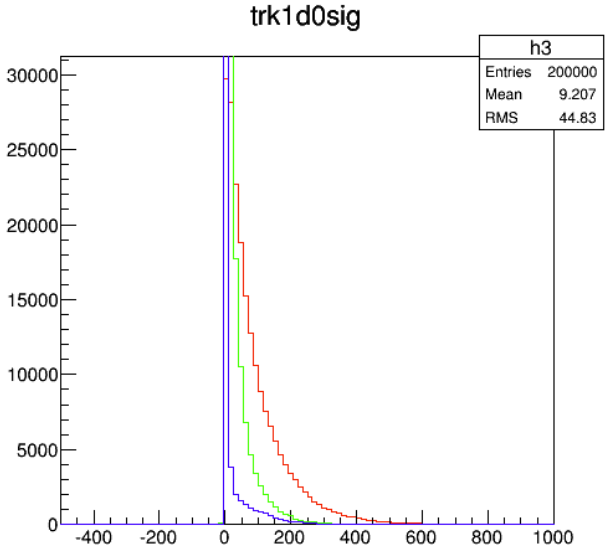




Red: b jet
Green: c jet
Blue: gluon jet

Red: b jet
Green: c jet
Blue: light jet

# Variable distribution of Higgs sample

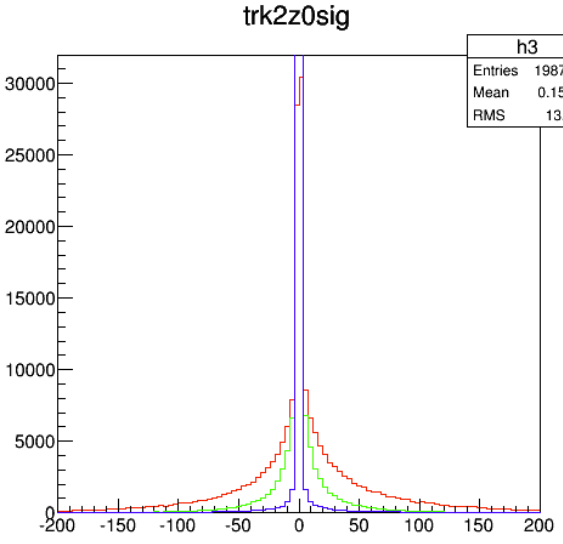


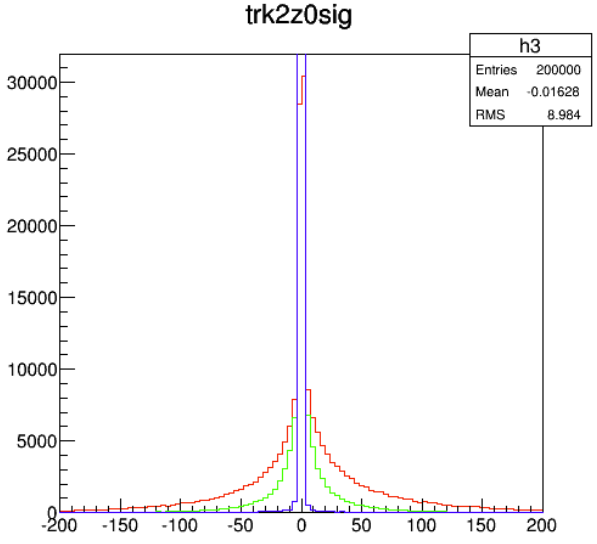




Red: b jet
Green: c jet
Blue: gluon jet

Red: b jet
Green: c jet
Blue: light jet

# Variable distribution of Higgs sample




Red: b jet
Green: c jet
Blue: gluon jet

Red: b jet
Green: c jet
Blue: light jet