PrimeTagSvc

Jets, samples and Wednesday working meeting

Kaili Zhang

zhangkl@ihep.ac.cn

Latest: v2.2

- https://code.ihep.ac.cn/zhangkl/PrimeTagSvc

- Code reconstructed;
  - Ensure PFO sorted by Energy, length up to 50.

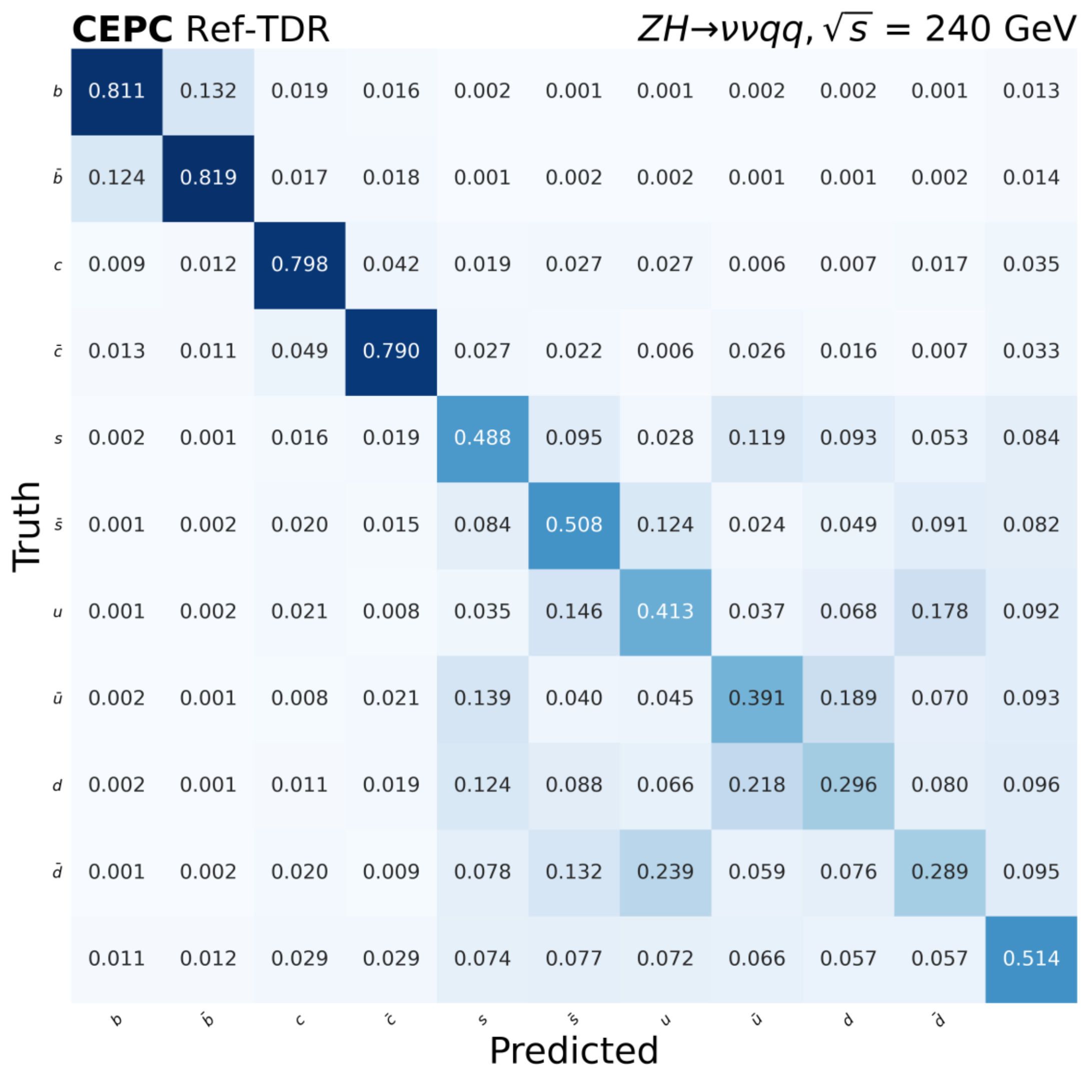
- Follow the tutorial JetDump
  - Mode: Zjet(M10) and Hjet(M11).

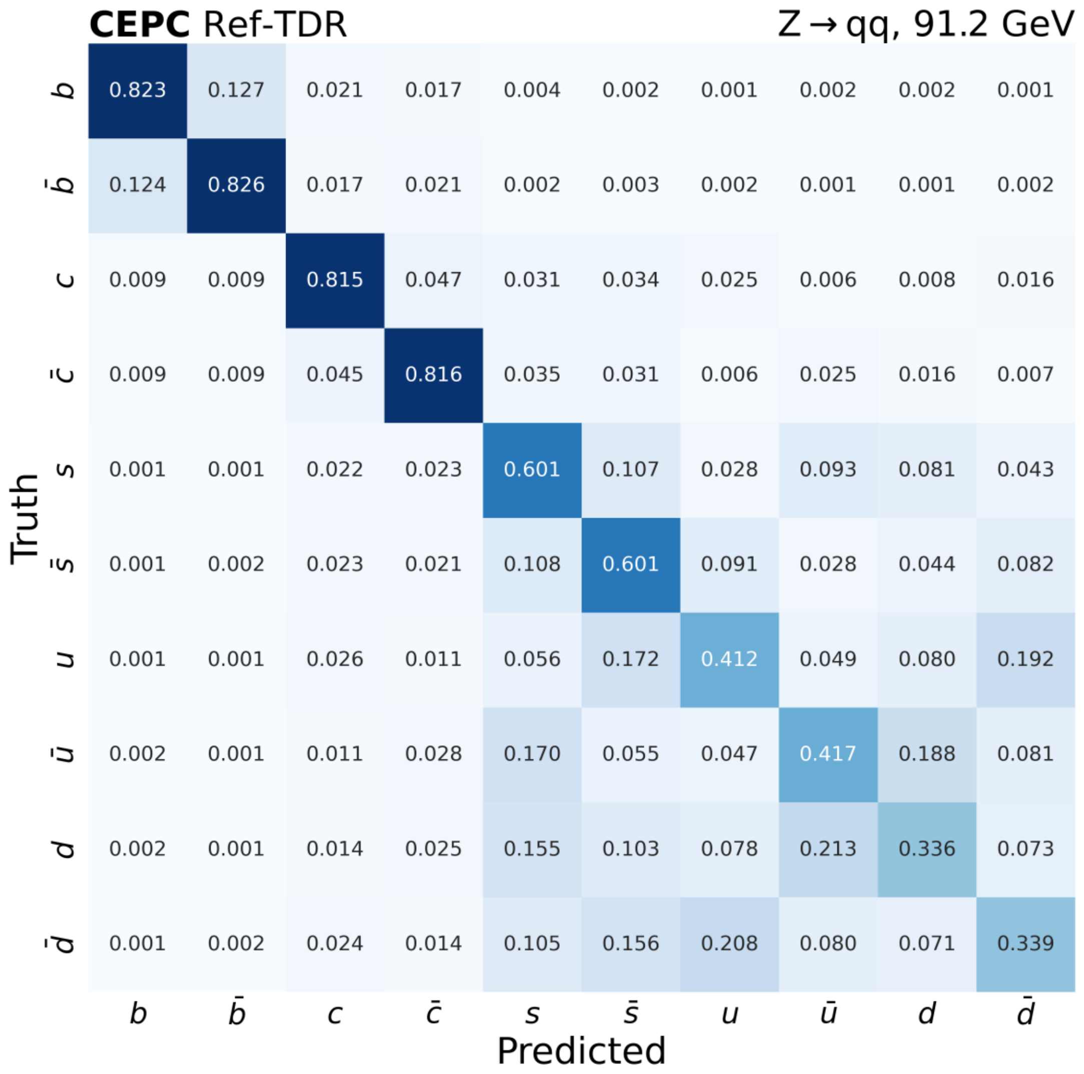
Interface for output;

```
std::vector<float> GetProb() const override;
int get_type_M11() const override;
int get_type_M6() const override;

// M11 Boolean functions (11 total)
bool is_b_quark() const override;
bool is_b_bar_quark() const override;
bool is_c_quark() const override;
bool is_c_bar_quark() const override;
bool is_s_quark() const override;
bool is_s_bar_quark() const override;
bool is_u_quark() const override;
bool is_u_bar_quark() const override;
bool is_d_quark() const override;
bool is_d_bar_quark() const override;
bool is_gluon() const override;

// M6 Boolean functions (6 total)
bool is_b_jet() const override;
bool is_c_jet() const override;
bool is_s_jet() const override;
bool is_u_jet() const override;
bool is_d_jet() const override;
bool is_gluon_jet() const override;
```

Default: M11/M10 model

In weaver environment we got results like this;



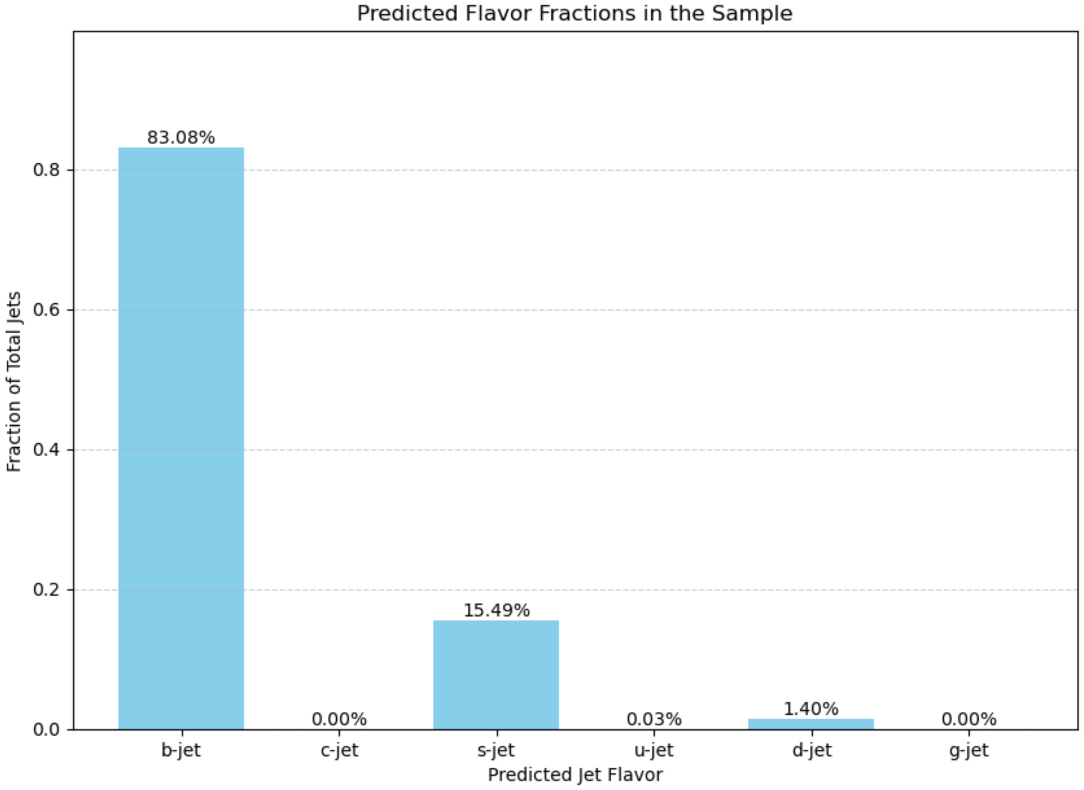


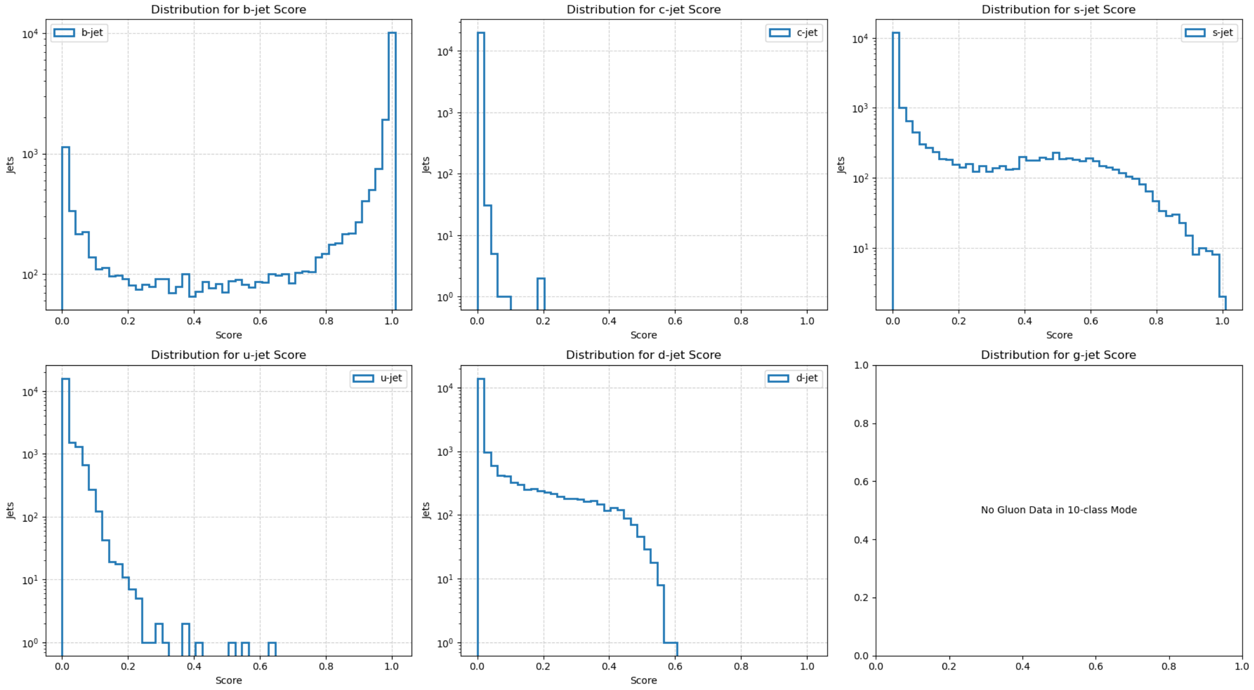
Current output in onnx, test in 10k.

1. b and bbar charge flip error;
2. incorrect response like s jet;

……

Current b eff ~83%.





Onnx consistence

```
--- Final Layer Bias Values ('mod.fc.0.bias') ---
-------------------------------------------------
b          : -0.351605
bbar       : -0.376052
c          :  0.080361
cbar       : -0.041093
s          :  0.030752
sbar       :  0.122631
u          :  0.060856
ubar       : -0.010073
d          :  0.097829
dbar       :  0.065422
-------------------------------------------------
```

- By checking the bias value and input/output structure, we can confirm that the pt in pytorch and onnx in C++ is exactly same.

Possible reason for errors

For low eff:
ill-behavior jets.
With small energy/PFOs and far away from quark direction. (DR>1.2)

This jet has only 3 tracks in final.
Model didn't see weird jets before.

```
PrimeTagSvc          DEBUG --- Flavor Tagging Scores ---
  b        : 0.001500
  b_bar    : 0.025265
  c        : 0.000655
  c_bar    : 0.003506
  s        : 0.387091
  s_bar    : 0.321368
  u        : 0.009419
  u_bar    : 0.021320
  d        : 0.030352
  d_bar    : 0.199525

JetDumpAlg           DEBUG Jet 1, E: 28.82, PFOs: 15, M11_type:  3, M6_type: 3, is_b: No , is_c: No , is_light: Yes
```

For model energy transition:
Hjet and Zjet model application difference;

- For charge flip and wrong response like s jets:
- Same Onnx model -> Inputs must be different.
- Float/Double, decimal changes between weaver/C++…
- Plan to retrain one new model to test.

To do

- Shap
- Variable Validation
- Delphes Validation
- Event multihead

Current Input variables

```yaml
pf_features:
    length: 50
    pad_mode: wrap
    vars:
        - [pfo_c_pt, null]
        - [pfo_c_energy, null]
        - [pfo_c_rel_energy, null]
        - [pfo_c_rel_pt, null]
        - [pfo_c_deltaR, null]
        - [pfo_charge, null]
        - [pfo_c_Ecal, null]
        - [pfo_c_Hcal, null]
        - [pfo_c_EClusize, null]
        - [pfo_c_HClusize, null]
        - [pfo_Rbest_isChargedKaon, null]
        - [pfo_Rbest_isPion, null]
        - [pfo_Rbest_isProton, null]
        - [pfo_Rbest_isElectron, null]
        - [pfo_Rbest_isMuon, null]
        - [pfo_d0, null]
        - [pfo_d0err, null]
        - [pfo_z0, null]
        - [pfo_z0err, null]
        - [pfo_deta, null]
        - [pfo_dphi, null]
pf_vectors:
    length: 50
    pad_mode: wrap
    vars:
        - [pfo_px, null]
        - [pfo_py, null]
        - [pfo_pz, null]
        - [pfo_Rbest_E, null]
```

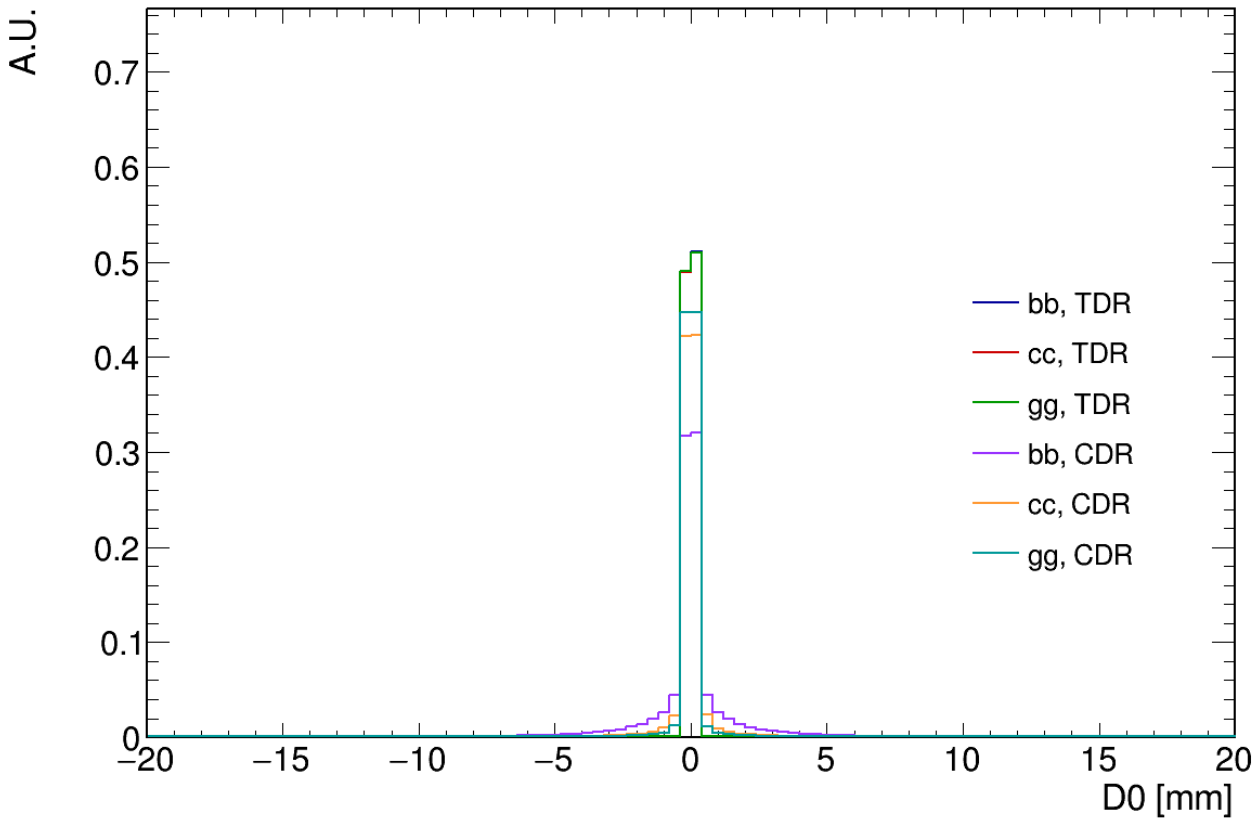
- Impact parameter d0 and log(d0):
    - To test.

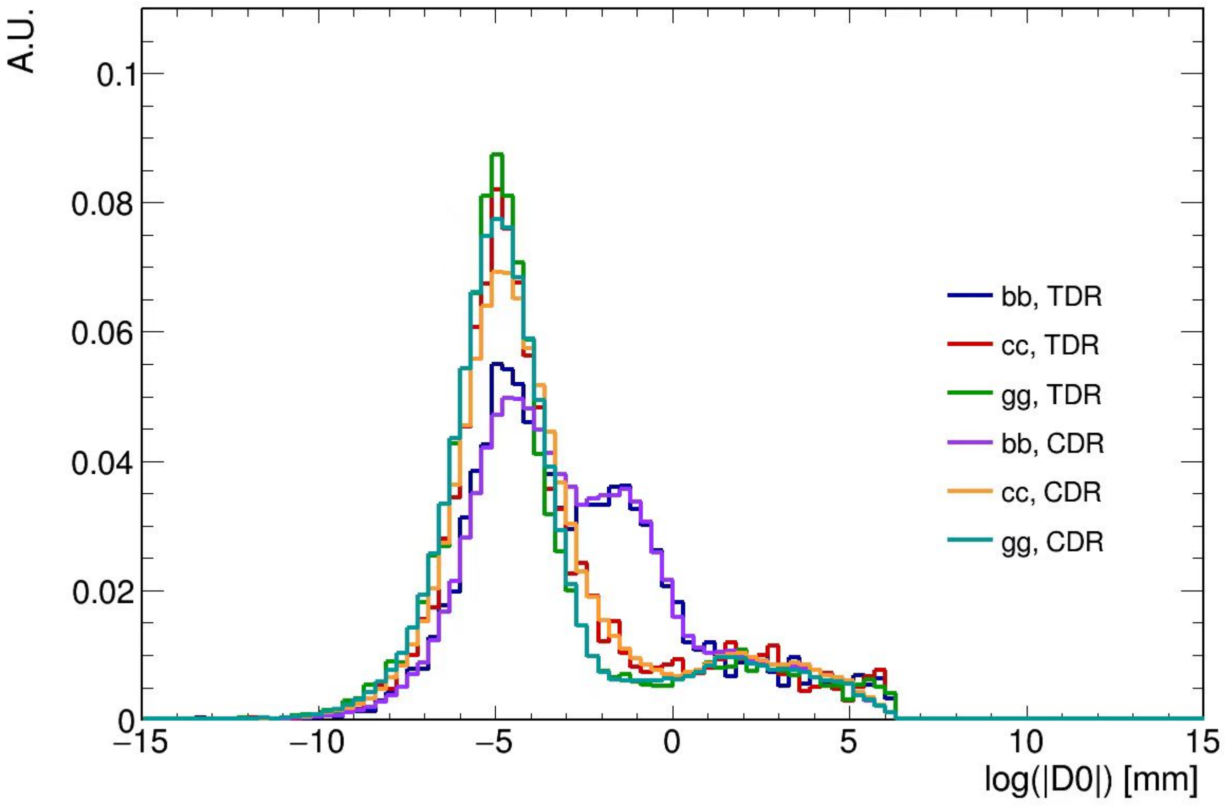
Argument on feature engineering d0 and log(d0):

- The goal is to make the learning task easier for the optimizer;
- By pre-computing a non-linear basis, the model now learns an easy linear combination instead of a hard logarithmic function.
- Possibly: linear d0 information lost in tanh() normalization (tanh(9)=1 in float), while log(d0) complete it;
- Standard Practice in CS:
  - ResNet, core idea: F(x)-x magic;
  - Positional Encodings: The "sin(pos)" trick give the information model already knows;
- Would try a new training to test its impact.

D0 distribution

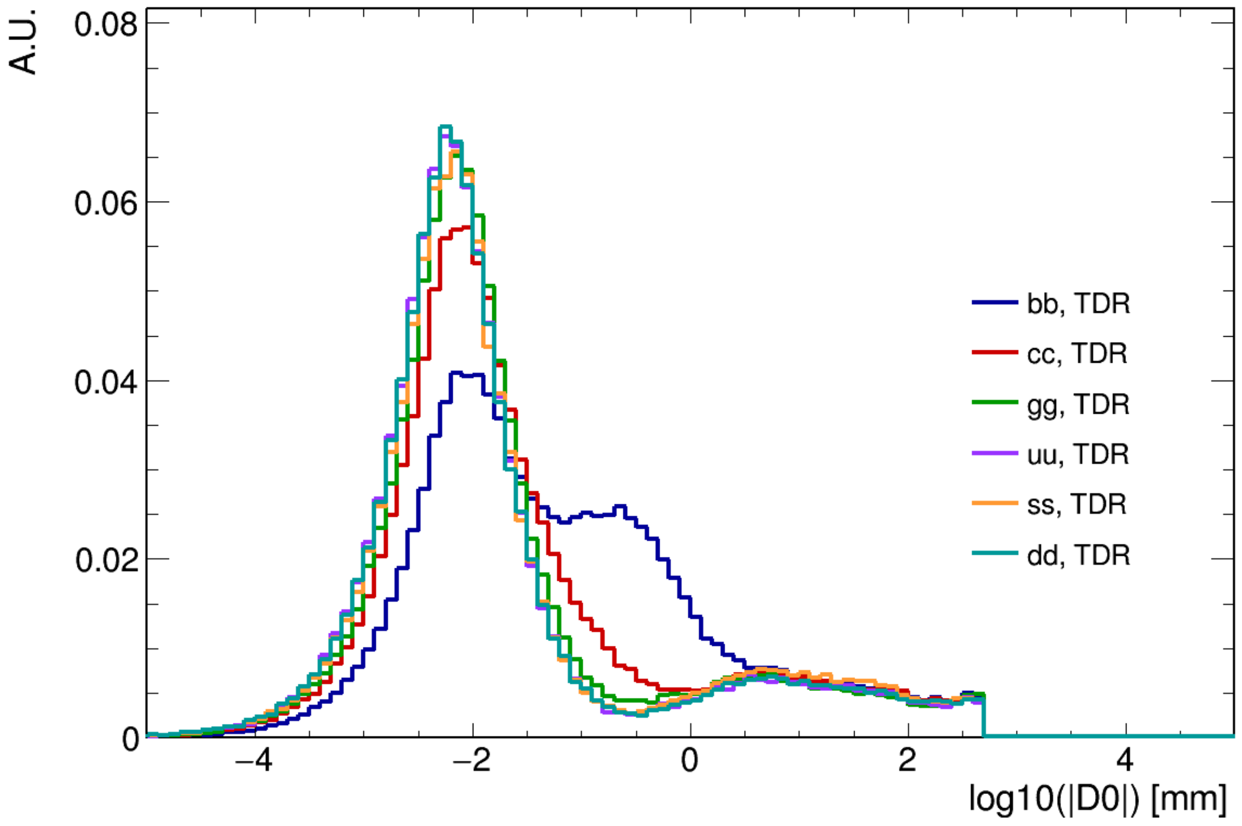
Both TDR/CDR full simulation has similar D0/Z0 performance.
And showing patterns in both D0 and log(D0) distribution.

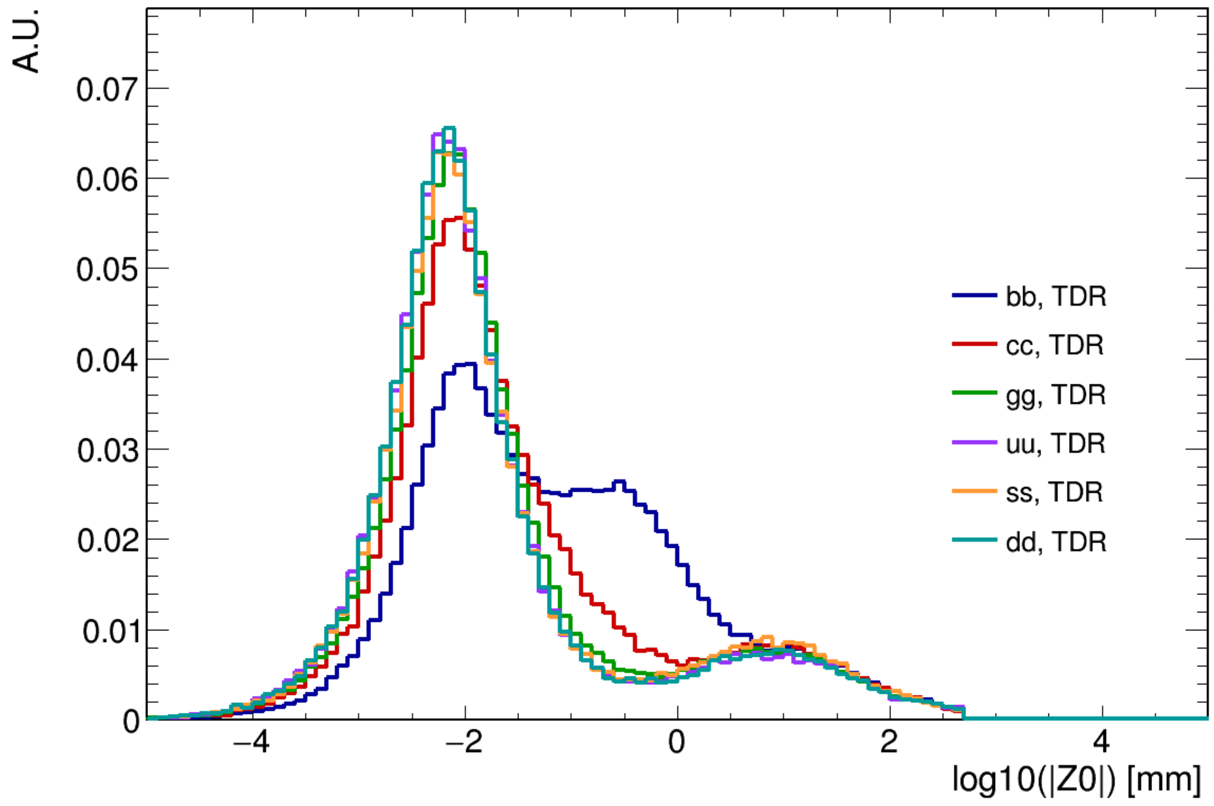




Log10(d0) pattern

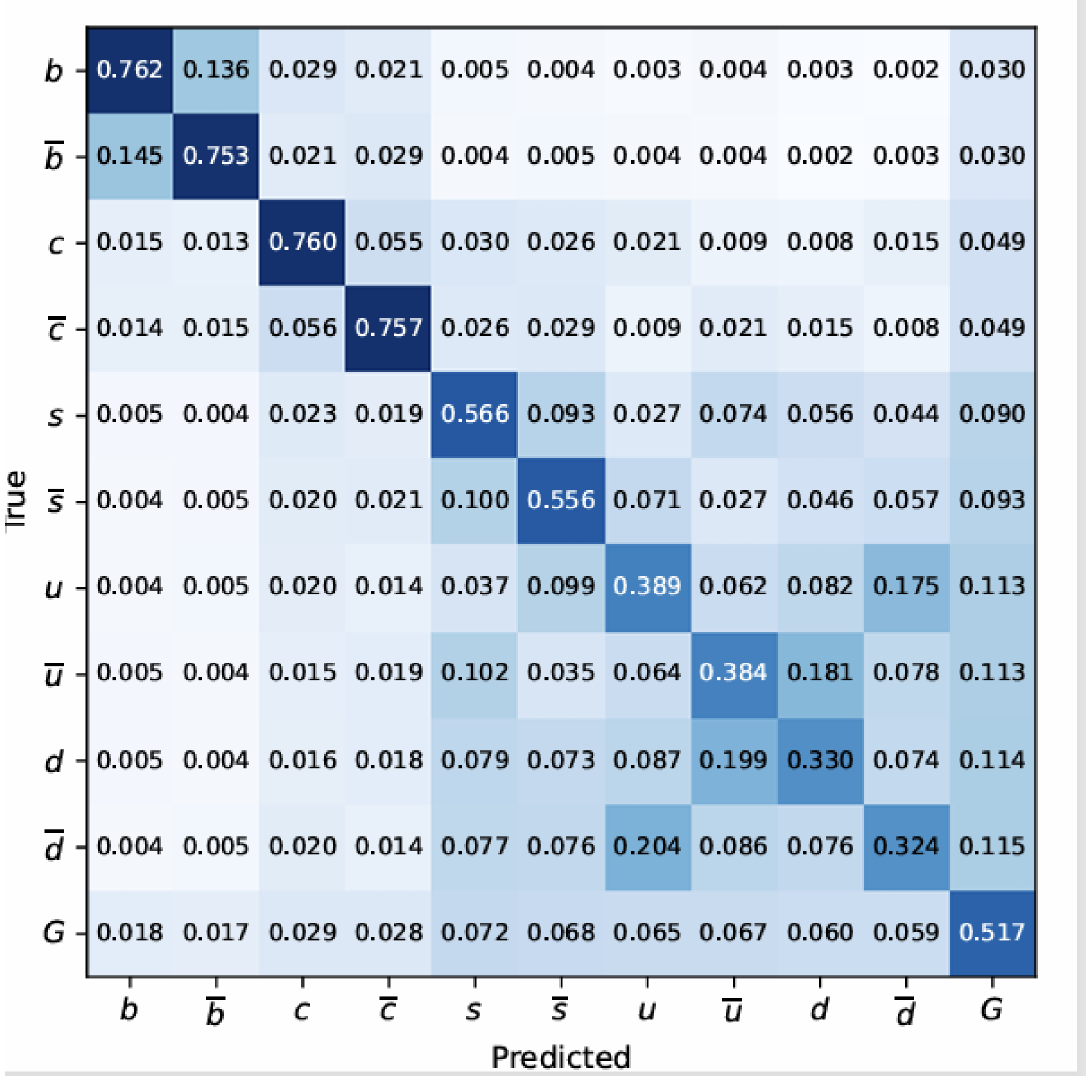
In D0 and Z0 plot, 3 pattern can be seen:
IP(Primary Vertex), while track in jet E~1GeV, the d0/z0 precision ~5um. (-2.5~3um)
Secondary/Thirdary Vertex(From b decay. Length~100um.)
and Long-Live Decay Vertex(From Kshort, Lambda……, length~cm.)

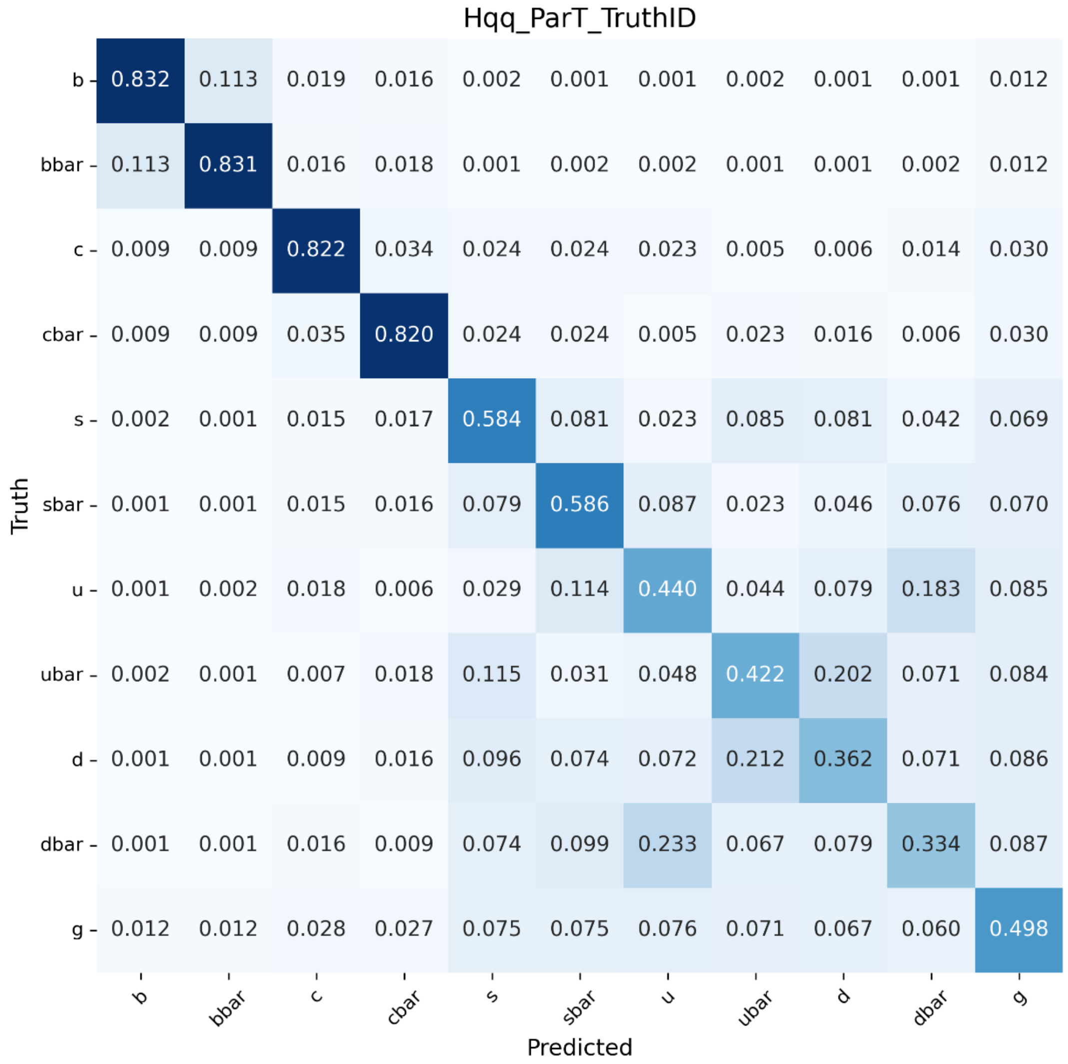




Fast/Full discrepancy

TDR truthID, fast(left)/full(right) JOI.
Fast overall worse, esp. for b eff.





PN/ParT difference

By PN to ParT we have 2% upgrade in Fast, 5% in Full.

True \ Predicted	b	$\bar{b}$	c	$\bar{c}$	s	$\bar{s}$	u	$\bar{u}$	d	$\bar{d}$	G
b	0.743	0.147	0.032	0.025	0.005	0.005	0.003	0.004	0.002	0.002	0.031
$\bar{b}$	0.151	0.739	0.025	0.033	0.004	0.006	0.003	0.003	0.002	0.003	0.031
c	0.014	0.014	0.748	0.059	0.032	0.029	0.021	0.009	0.007	0.014	0.052
$\bar{c}$	0.014	0.014	0.057	0.751	0.028	0.032	0.009	0.021	0.014	0.008	0.052
s	0.004	0.005	0.020	0.020	0.555	0.103	0.024	0.073	0.053	0.043	0.099
$\bar{s}$	0.004	0.005	0.020	0.021	0.099	0.558	0.072	0.024	0.041	0.054	0.102
u	0.004	0.005	0.019	0.014	0.038	0.110	0.375	0.056	0.080	0.180	0.119
$\bar{u}$	0.004	0.005	0.015	0.020	0.108	0.040	0.058	0.378	0.171	0.083	0.119
d	0.005	0.005	0.015	0.019	0.083	0.081	0.081	0.206	0.304	0.081	0.121
$\bar{d}$	0.004	0.005	0.019	0.014	0.080	0.086	0.202	0.078	0.076	0.313	0.121
G	0.019	0.018	0.029	0.029	0.072	0.072	0.063	0.066	0.056	0.058	0.518

True \ Predicted	b	$\bar{b}$	c	$\bar{c}$	s	$\bar{s}$	u	$\bar{u}$	d	$\bar{d}$	G
b	0.762	0.136	0.029	0.021	0.005	0.004	0.003	0.004	0.003	0.002	0.030
$\bar{b}$	0.145	0.753	0.021	0.029	0.004	0.005	0.004	0.004	0.002	0.003	0.030
c	0.015	0.013	0.760	0.055	0.030	0.026	0.021	0.009	0.008	0.015	0.049
$\bar{c}$	0.014	0.015	0.056	0.757	0.026	0.029	0.009	0.021	0.015	0.008	0.049
s	0.005	0.004	0.023	0.019	0.566	0.093	0.027	0.074	0.056	0.044	0.090
$\bar{s}$	0.004	0.005	0.020	0.021	0.100	0.556	0.071	0.027	0.046	0.057	0.093
u	0.004	0.005	0.020	0.014	0.037	0.099	0.389	0.062	0.082	0.175	0.113
$\bar{u}$	0.005	0.004	0.015	0.019	0.102	0.035	0.064	0.384	0.181	0.078	0.113
d	0.005	0.004	0.016	0.018	0.079	0.073	0.087	0.199	0.330	0.074	0.114
$\bar{d}$	0.004	0.005	0.020	0.014	0.077	0.076	0.204	0.086	0.076	0.324	0.115
G	0.018	0.017	0.029	0.028	0.072	0.068	0.065	0.067	0.060	0.059	0.517

Fast simulation:

Current fast simulation with overall worse performance including BMR, diphoton resolution and so on.
Under check.