Analysis of Br(H->bb,cc,gg) at neutrino channels

CIAE

SHOUYANG HU, HAO LIANG

Department of physics, CIAE

Content

- Physics motivation
- Brief review on Higgs decay (bb, cc, gg) in SM
- Simulation and event reconstruction
- Analysis step 1: pre-event selection
- Analysis step 2: MVA analysis
- Analysis step 3: template fitting
- Discussion

Physics motivation

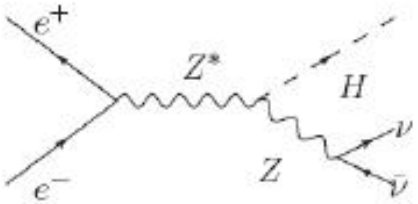
- Large statistics + clean background = High precision.
  - Best test on Higgs physics.
  - Precision measurement of Higgs interaction with particles.
- Basic problem in High energy physics: how many Higgs ? The key method to distinguish between different Higgs models involves study of the particles interaction and exact decay process. Which can be measure and tested experimentally in particle collisions: *CEPC*.

Brief review on Higgs decay (bb, cc, gg) in SM

Higgs production in CEPC:

Major Z pole: $e^-e^+ \to Z^*H \to \nu\nu H(bb, cc, gg)$

Minor WW fusion: $e^-e^+ \to W^*W^*H \to \nu\nu H(bb, cc, gg)$



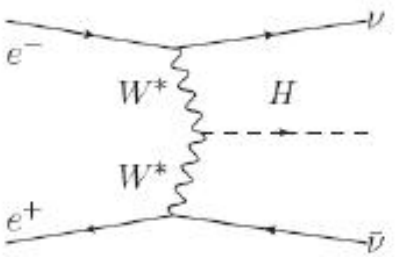
Higgs decay into quarks pair:

Yukawa interaction $L_{\mathrm{I}} = -\left(\sqrt{2}G_F\right)^{1/2} m\bar{q}qH$

First order approximation: decay rate$\sim m^2 (m \ll m_H)$

Higgs decay into gluon pair:

Quarks loops



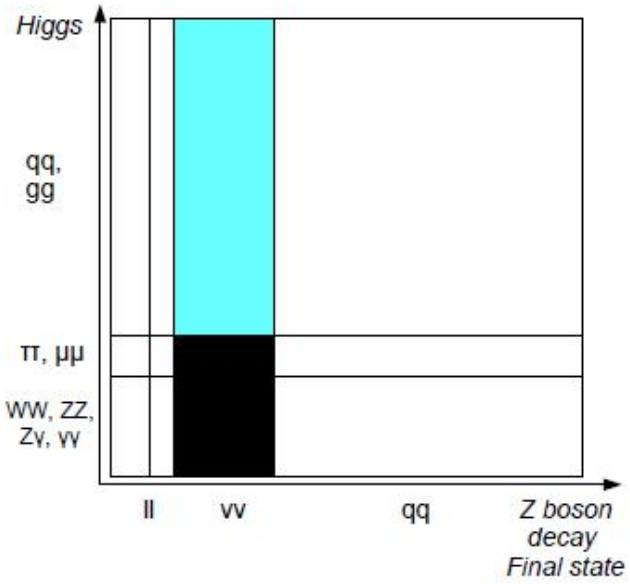
Mode	$b\bar{b}$	$c\bar{c}$	gg	WW*	$\mu^+\mu^-$	$\tau^+\tau^-$	ZZ*	$\gamma\gamma$	$Z\gamma$
BR (%)	57.8	2.7	8.6	21.6	0.02	6.4	2.7	0.23	0.16

Total: ~70%

Simulation and event reconstruction

- Simulation: *ilcsoft*
- Reconstruction: *marlin*
- Statistics: 5000fb^-1!
- We have 148K signal events (bb+cc+gg)!

Name	Statistics	weight	Note
$\nu\nu H$	5000fb^{-1}	1	Full simulation
$(qq, e^+e^-, \mu^+\mu^-)H$	5000fb^{-1}	1	Full simulation
$\tau^-\tau^+H$	0	0	Not available
2fermions/4fermions	500fb^{-1}	10	Fast simulation



We have many possible channels to analyze, Where is our position?

Analysis 1: pre-events selection

- 1. Number of particle > 20 (veto leptonic final states)
- 2. 110GeV <Total Energy <150GeV (veto hadronic final states)
- 3. Isolated electrons & isolated muons veto. (threshold energy 10GeV)
- 4. 100GeV < Invariant mass < 135GeV 70GeV < recoil mass < 125GeV (Higgs peak and Z peak)
- 5. Cuts on y12, y23, y34 (select double jets)
- 6. -0.98 < Cos(Included angle of two jets) <-0.4 (large Higgs mass)

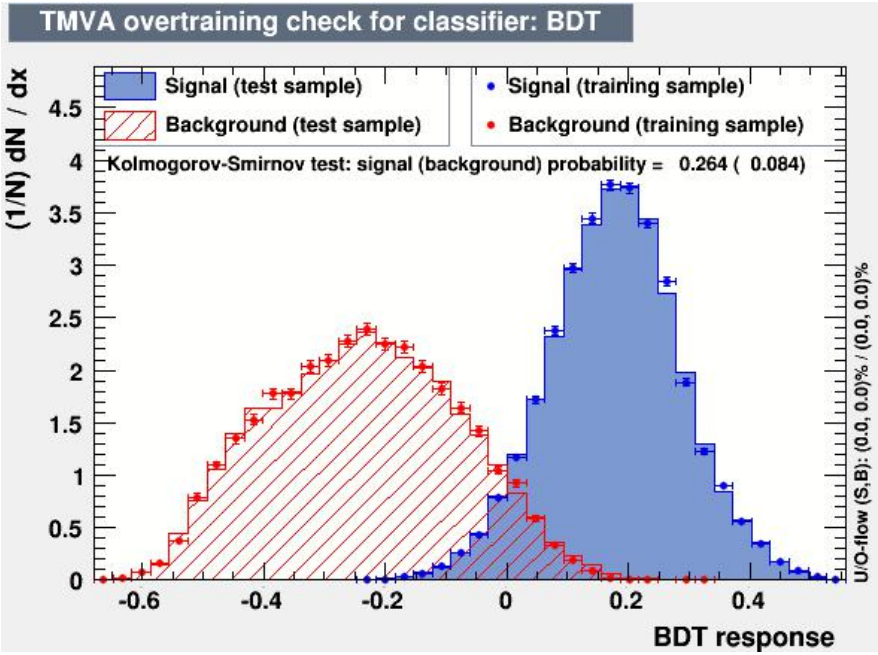
Analysis 1: pre-events selection

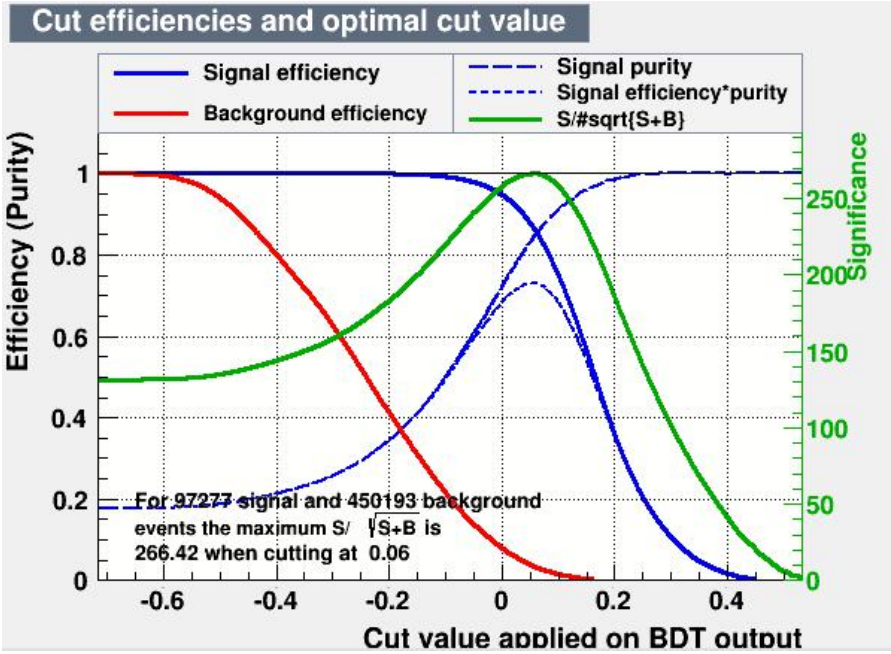
Cut Definition	*Sig.*	*qq*	*qqnn*	*qqln*	*nnh*
Generated	16260	25M	183K	3681K	
FSClasser output	16768	25M	183K	3681K	7485
$N_{\mathrm{PFO}(E>0.4\mathrm{GeV})} > 20$	16748	23M	163K	3439K	4889
$110 < E_{\mathrm{total}} < 150$	14689	10M	126K	705K	3311
$P_T > 19$	13687	34K	116K	627K	3101
Isolation lepton veto	13429	33775	115K	327K	2537
$100 < M_{\mathrm{inv}} < 135$	12827	9506	10420	162K	2269
$70 < M_{\mathrm{rec}} < 125$	12166	7521	10045	110K	2260
$0.15 < y_{12} < 1$	12093	7405	9702	101K	2211
$y_{23} < 0.06$	10902	6644	8456	69313	1220
$y_{34} < 0.008$	10377	6504	7878	58532	519
$-0.98 < \cos(\theta^{(2\mathrm{jets})}_{\mathrm{included}}) < -0.4$	10284	5766	5454	34823	485
$BDT > 0.04$	8705	381	465	267	230
Significance	**84.92**				
Efficiency	**53.5%**				

Analysis 2: Multivariate analysis (MVA)

- The software: *TMVA*(http://tmva.sourceforge.net/)
- Analysis method: boosted decision tree (BDT)
- Input variables: N_{PFO}, $P_{T,total}$, M_{inv}, M_{reco}, y_{12}, y_{23} , y_{34}, θ_{2jets}. (E_{total} was not included to reduce the overfitting.)
- Optimization: maximize the statistics significance (= $S/\sqrt{S+B}$).

Analysis 2: Multivariate analysis (MVA)





Analysis 2: Multivariate analysis (MVA)

- Integral luminosity 5000fb^-1,
- Total signal Efficiency 58.5%, Significance **266**
- Main non-Higgs SM backgrounds: qqln, qqnn, qq

Number	bb	cc	gg	Oth higgs	SM
before	125725	5853	17377	Didn't cal.	Didn't cal.
Efficiency	59.0%	55.8%	55.9%	Didn't cal.	Didn't cal.
Left	74191	3266	9710	3299	1556*10

Analysis 3: Template fitting

- Construct fitting variables: bb-likeliness, cc-likeliness

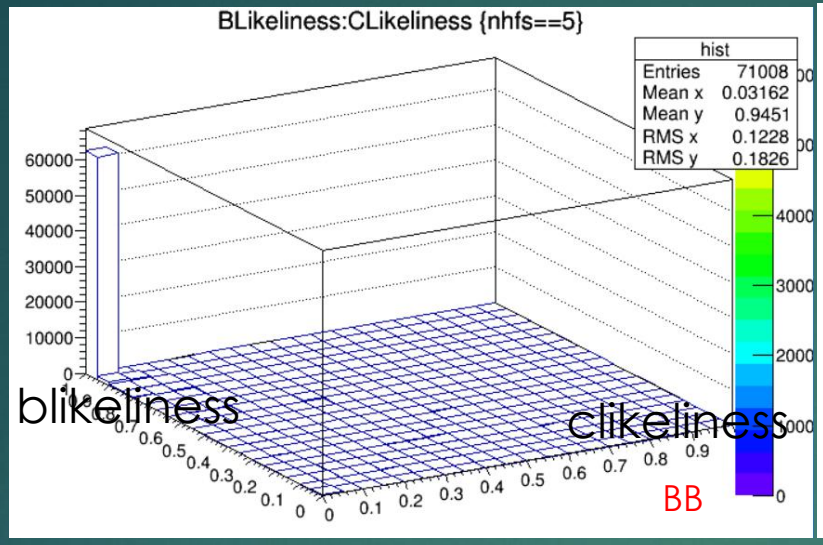
$$L_{qq} = \frac{qq\ pair}{qq\ pair\ +\ neither\ is\ \ q} = \frac{x_q^1\ x_q^2}{x_q^1 x_q^2 + (1-x_q^1)(1-x_q^1)}\ (qq = bb, cc)$$

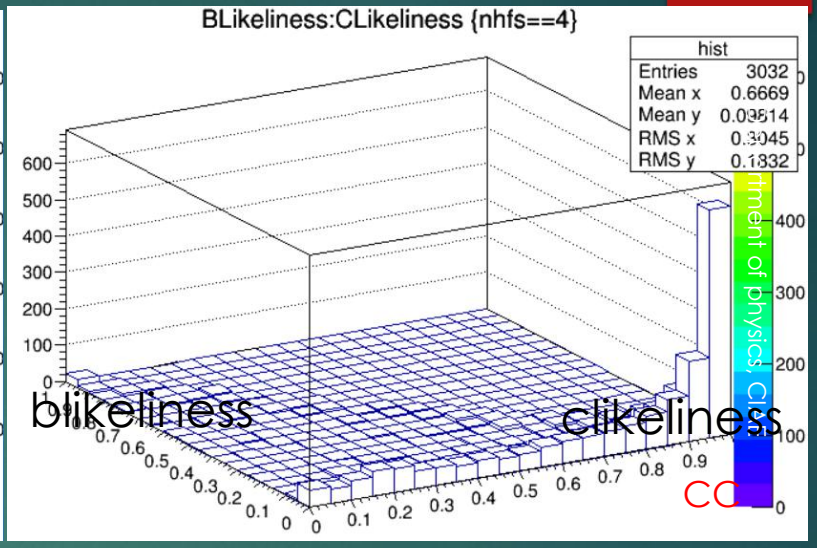
- Extract the N_{br}: Maximize the likelihood (assuming the Poisson distribution in each bins)

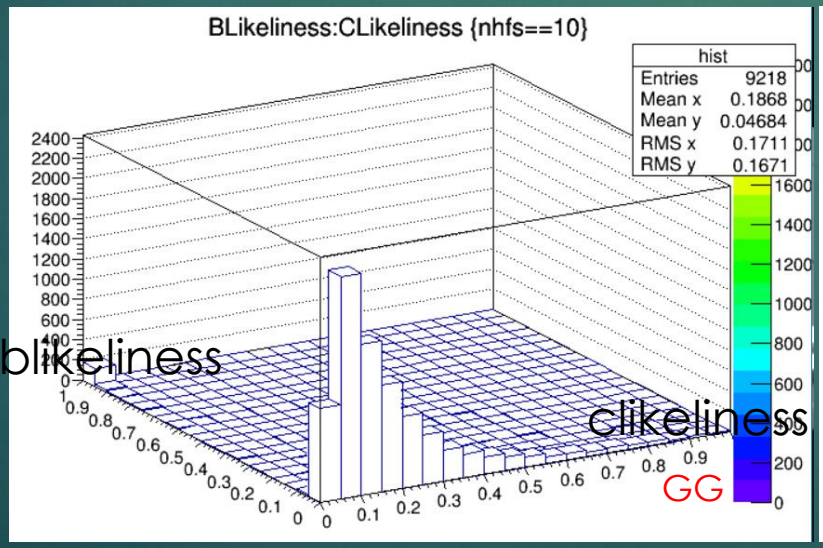
$$likelihood(N_{br})$$

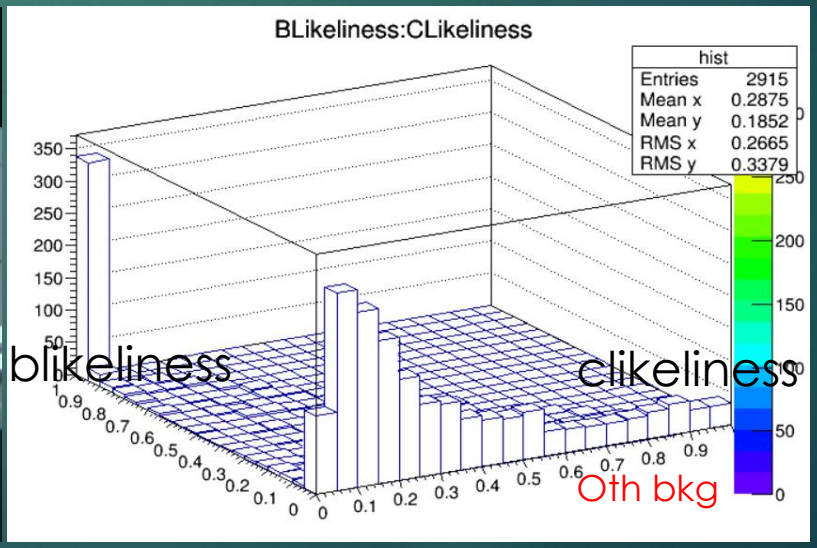
$$= Prob(hist\ of\ data \mid \sum_{br=bb,cc,gg,other,sm} N_{br} template\_hist_{br})$$

- However:
  - We would not count non-Higgs backgrounds in this work.
  - We would fix the $\nu\nu H(others)$ to be the constant of its truth-value. Because we had found this could improve the precision of branching ratio for gg.


BLikeliness:CLikeliness {nhfs==5}
hist
Entries 71008
Mean x 0.03162
Mean y 0.9451
RMS x 0.1228
RMS y 0.1826
blikeliness
clikeliness
BB

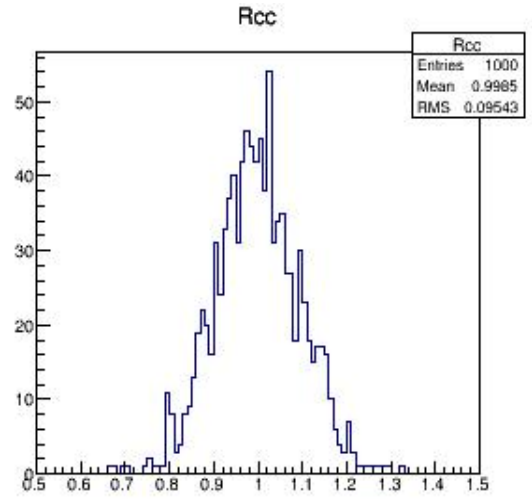

BLikeliness:CLikeliness {nhfs==4}
hist
Entries 3032
Mean x 0.6669
Mean y 0.03514
RMS x 0.3045
RMS y 0.1332
blikeliness
clikeliness
CC

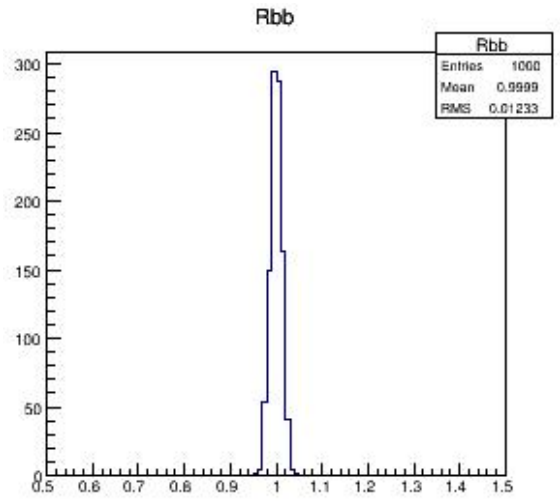

BLikeliness:CLikeliness {nhfs==10}
hist
Entries 9218
Mean x 0.1868
Mean y 0.04684
RMS x 0.1711
RMS y 0.1671
blikeliness
clikeliness
GG

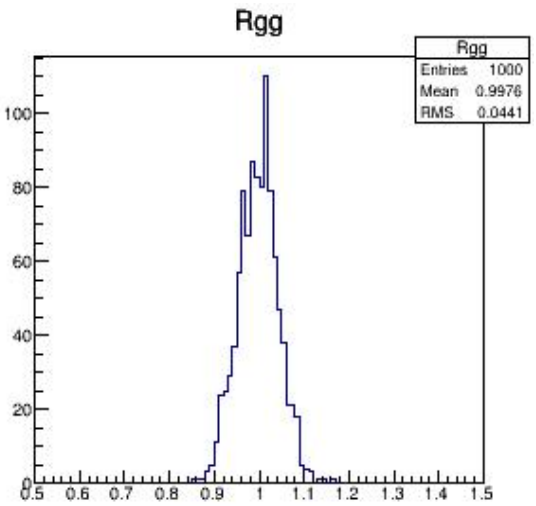

BLikeliness:CLikeliness
hist
Entries 2915
Mean x 0.2875
Mean y 0.1852
RMS x 0.2665
RMS y 0.3379
blikeliness
clikeliness
Oth bkg

Analysis 3: Template fitting

- We use ToyMC method to generate 1000 templates and 1000 data. So we have 1000 fitting results. Data size 500fb^-1, Template size 5000fb^-1.







Fitting result over truth for cc, bb, gg respectively

Analysis 3: Template fitting

Channel	bb	cc	gg	Oth Higgs
Truth	7419.1	326.6	971.0	329.9
Mean	7419.0	325.0	969.4	330.0
RMS (fitTo)	88.1	28.4	38.8	-
RMS (ToyMC)	87.1	30.1	39.6	-
Relative error	1.23%	9.3%	4.2%	-
1/sqrt(Truth)	1.16%	5.5%	3.2%	-

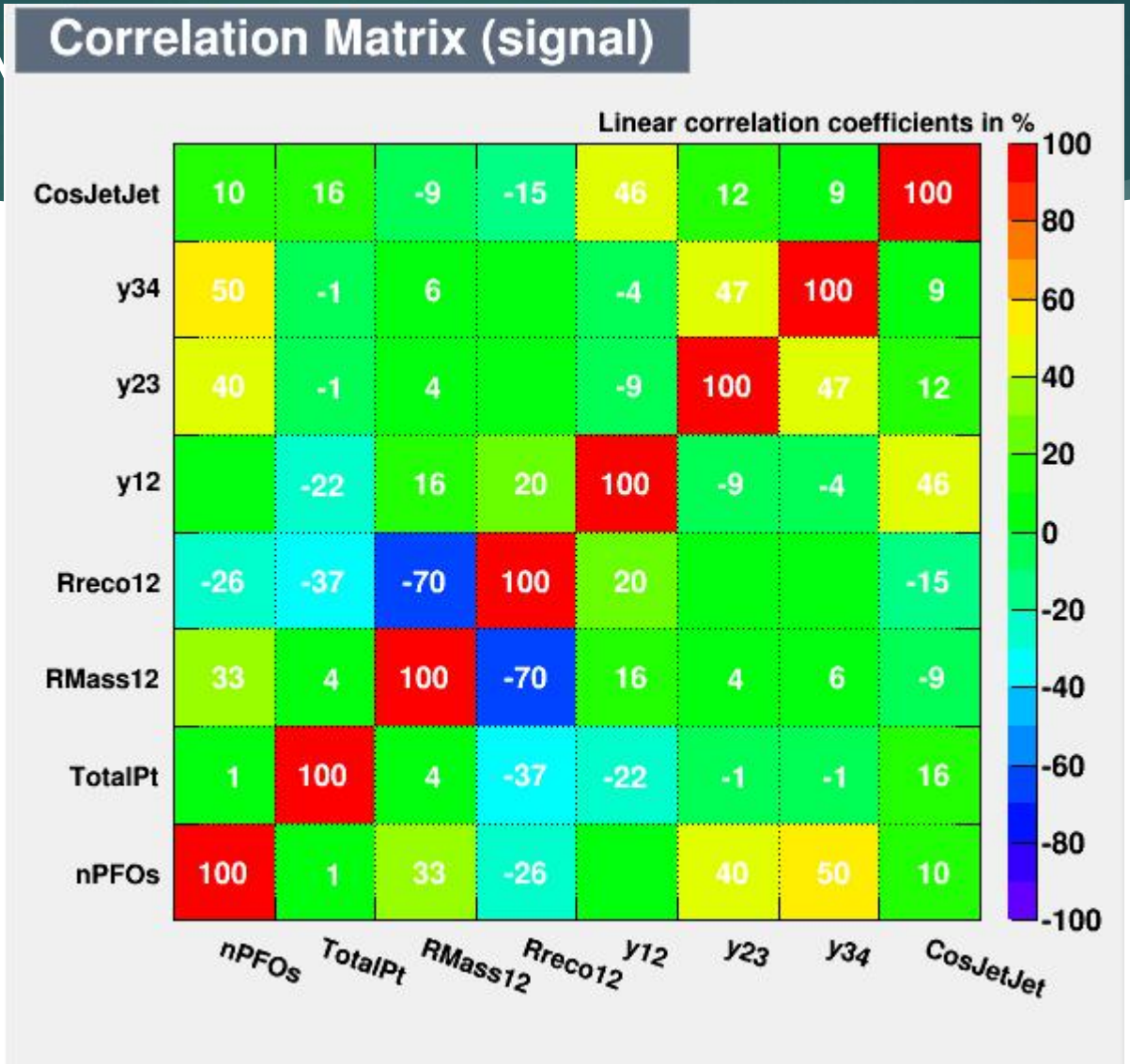
- Result:
  - 1.No biases was found!
  - 2.The precision of $H \to \bar{b}b$ branching ratio has achieve the statistics limit $\frac{1}{\sqrt{N}}$; For gg, cc, the errors are 30%-100% more than statistics limits.

What next?

- Optimize statistics significance for bb, cc, gg respectively and extract bb, cc, gg branching by three template fittings. How much improvement?
- How much background would come from $e^-e^+ \to \tau\tau H$? Will count it.
- How to improve the precision of gg if assuming the number of $\nu\nu H(others)$ is *unknown*. (very hard problem we met at present.)

Thanks for your attention!

MVA


Correlation Matrix (signal)
Linear correlation coefficients in %
CosJetJet 10 16 -9 -15 46 12 9 100
y34 50 -1 6 -4 47 100 9
y23 40 -1 4 -9 100 47 12
y12 -22 16 20 100 -9 -4 46
Rreco12 -26 -37 -70 100 20 -15
RMass12 33 4 100 -70 16 4 6 -9
TotalPt 1 100 4 -37 -22 -1 -1 16
nPFOs 100 1 33 -26 40 50 10
nPFOs TotalPt RMass12 Rreco12 y12 y23 y34 CosJetJet
100 80 60 40 20 0 -20 -40 -60 -80 -100