Updates on Higgs boson invisible decay studies

中国科学院高能物理研究所
Institute of High Energy Physics
Chinese Academy of Sciences

Geliang Liu (刘格良)

Jun. 4th, 2025

Outlines

- **Changes on analysis strategies**
  - New baseline selection and kinematic selection
  - Use XGBoost models to distinguish signals v.s. backgrounds
- **New results**

Samples

➢ Samples produced with CEPCSW 25.3.6

	Processes	Location	Events
Signal	Z(→ee/μμ/qq)H(→4v)	/cefs/higgs/liugeliang/CEPC/202503/Production/Hinvi	100k per final state
4-fermion bkg	single-Z, single-W, Z-or-W, ZZ, WW, ZZ-or-WW	/cefs/higgs/zhangkl/Production/25036/4fermions	400k per final state
2-fermion bkg	ll, qq	/cefs/higgs/zhangkl/Production/25036/E240_e1e1(e2e2, e3e3, qq)	100k – 400 k per final state
ZH	Z(→ee/μμ/ττ/vv/qq) H(→incl)	/cefs/higgs/zhangkl/Production/25036/E240_*HX	1M per sample

Baseline selection

Requirements

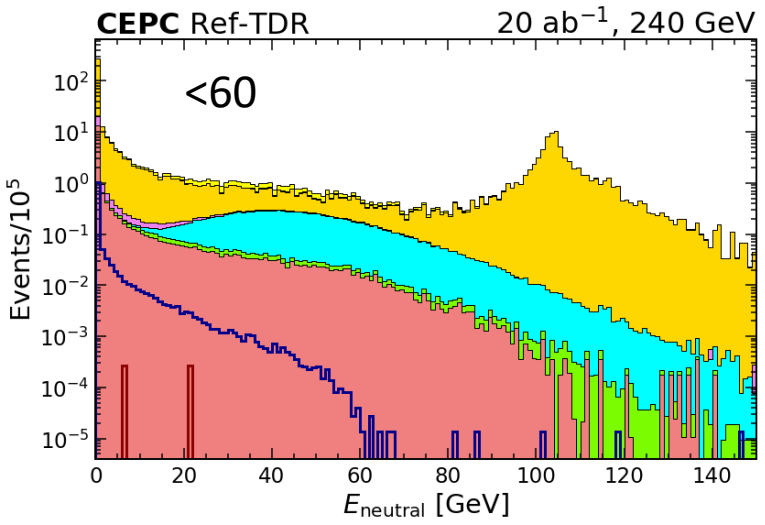
- Ensure high efficiency.
- Orthogonality between different channels (2mu, 2e, 2q).
- Low migration of signals to the wrong channels.

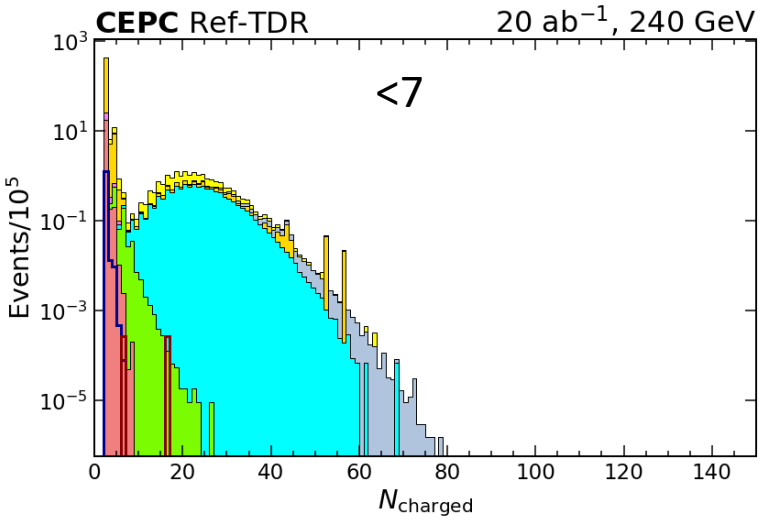
Criteria

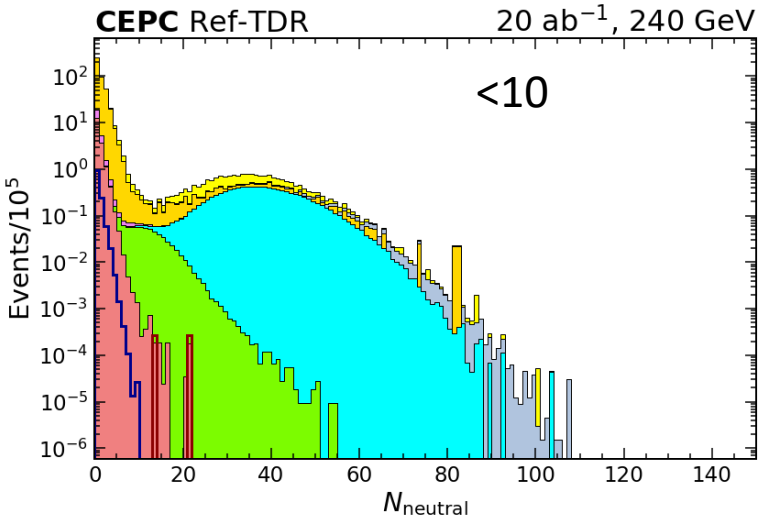
- Select **2mu** channel first:
  - Exactly two muons passing the BestWP, with $|\cos\theta|<0.99$, opposite charge.
  - The invariant mass of the two muons should be between 40 and 120 GeV.
- For events not entering 2mu channel, select **2e** channel:
  - Exactly two electrons passing the BestWP, with $|\cos\theta|<0.99$, opposite charge.
  - The invariant mass of the two electrons should be between 40 and 120 GeV.
- For events not entering 2mu or 2e channel, select **2q** channel:
  - The visible mass should be between 30 and 130 GeV.
  - The visible momentum should be between 10 and 80 GeV.

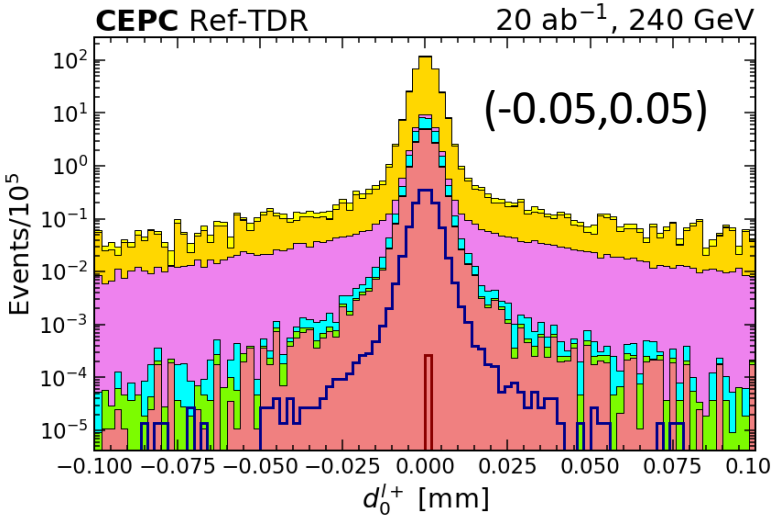
Kinematic selection: 2mu

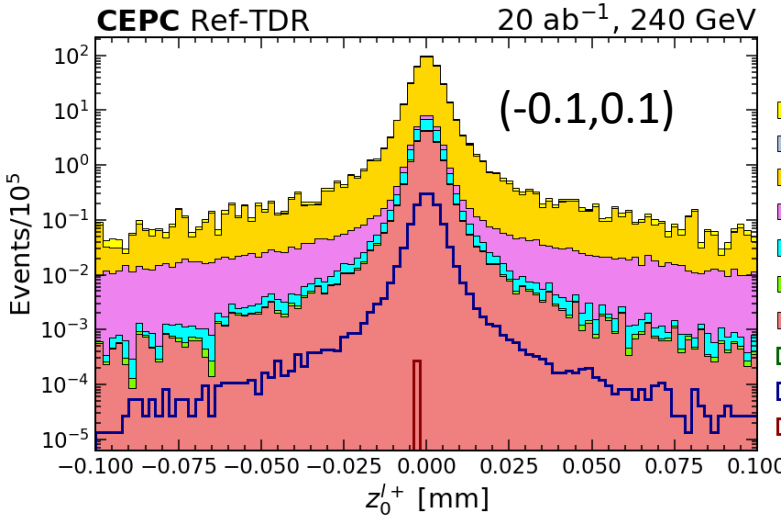
Requirements: ensure high signal efficiency (>98%); on condition of that, reduce backgrounds.





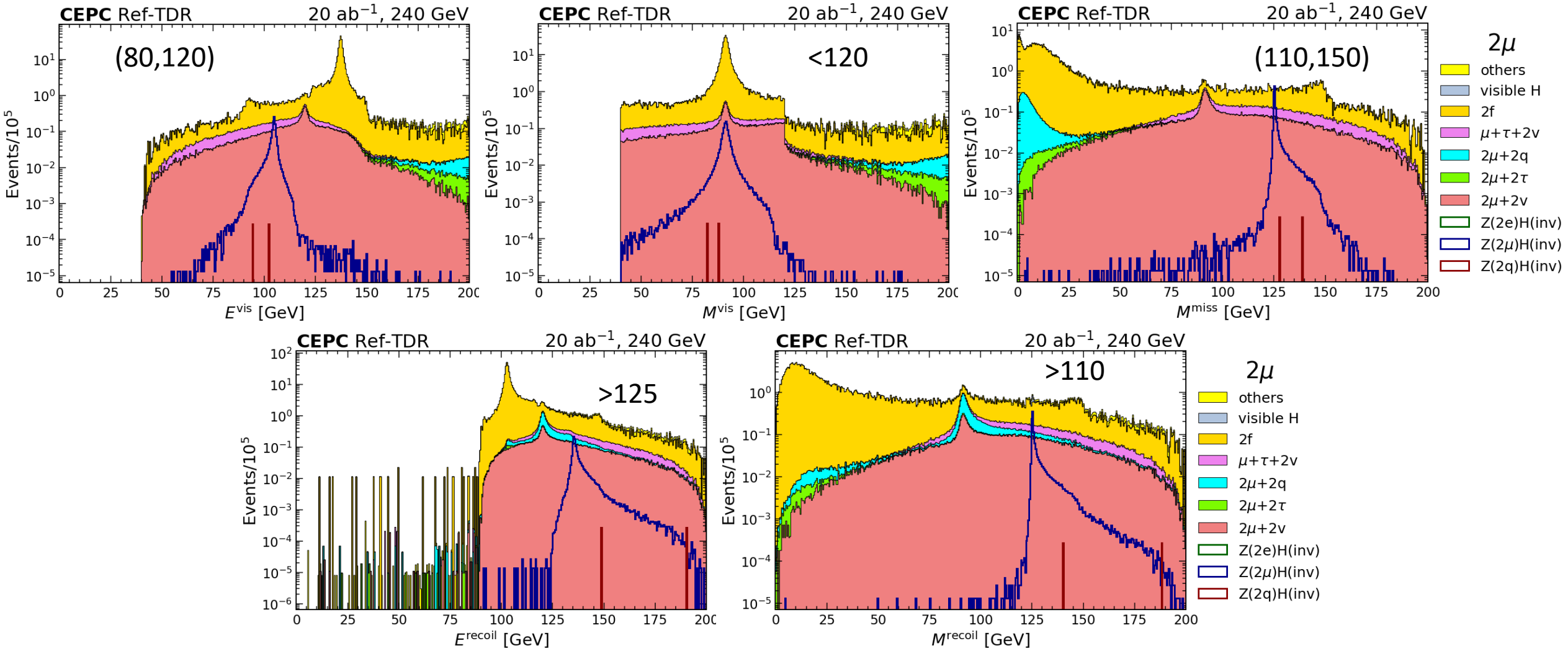






Kinematic selection: 2mu

Requirements: ensure high signal efficiency (>98%); on condition of that, reduce backgrounds.



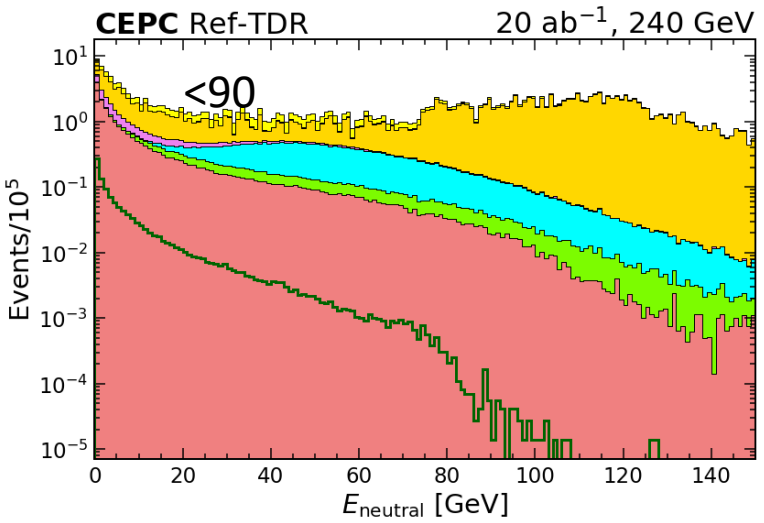
Kinematic selection: 2mu

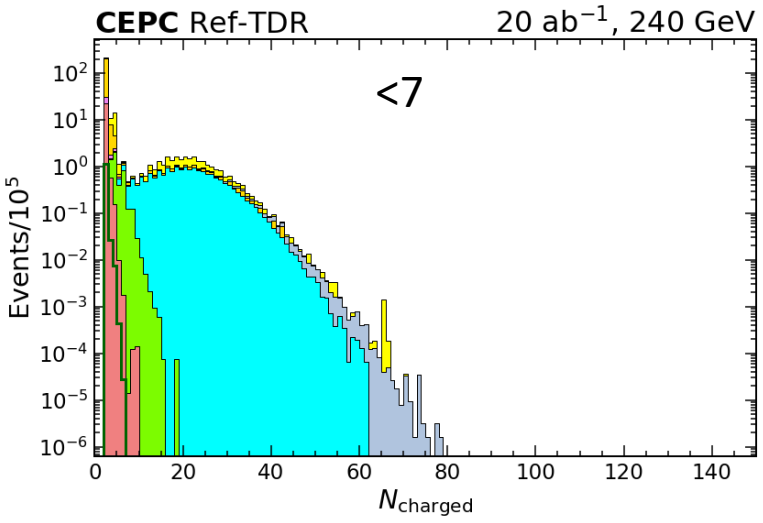
Cutflow

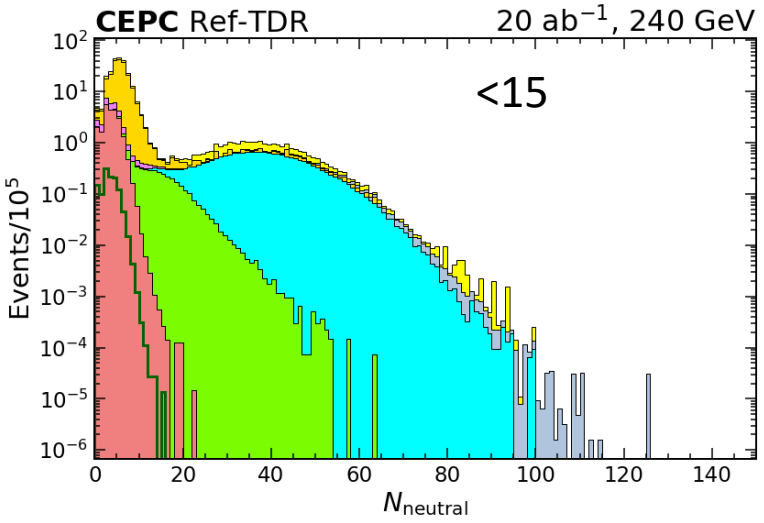
```
---------2mu channel----------
process          | total      | base eff | kinsel eff | selected
Z(2mu)H(inv)     | 1.435e+02 | 96.109 % |   99.194 % |      137
Z(2e)H(inv)      | 1.492e+02 |  0.000 % |    0.000 % |        0
Z(2q)H(inv)      | 2.900e+03 |  0.002 % |    0.000 % |        0

2mu+2v           | 5.678e+06 | 32.010 % |   25.335 % |   460478
2mu+2tau         | 3.730e+05 | 21.103 % |    1.100 % |      866
2mu+2q           | 4.471e+06 | 24.432 % |    0.000 % |        0
mu+tau+2v        | 8.073e+06 | 10.679 % |   23.368 % |   201471
2f               | 1.779e+09 |  2.348 % |    4.530 % |  1892732
hx               | 4.073e+06 |  2.547 % |    0.569 % |      590
others           | 3.658e+08 |  0.358 % |    1.208 % |    15804
```

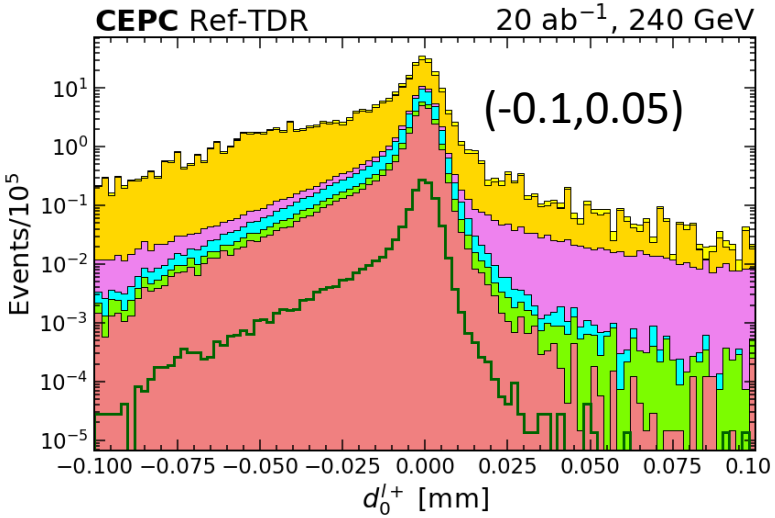
Kinematic selection: 2e

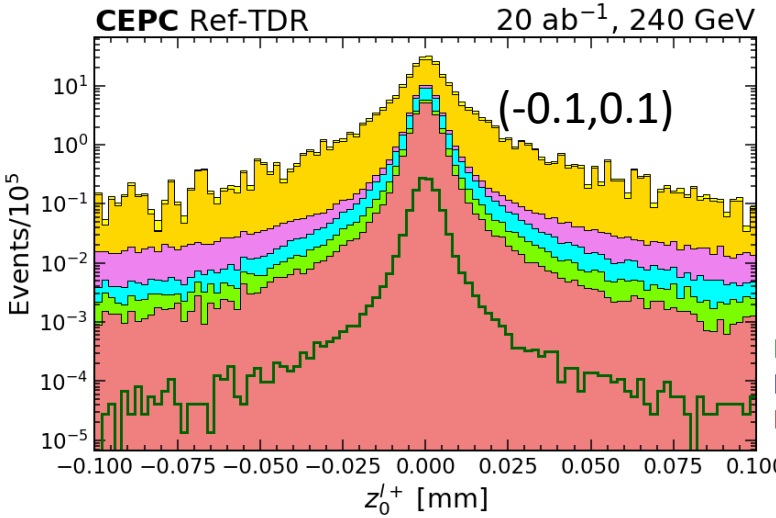
Requirements: ensure high signal efficiency (>98%); on condition of that, reduce backgrounds.

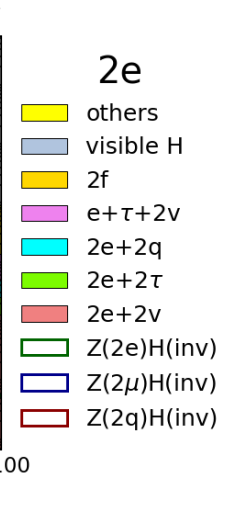






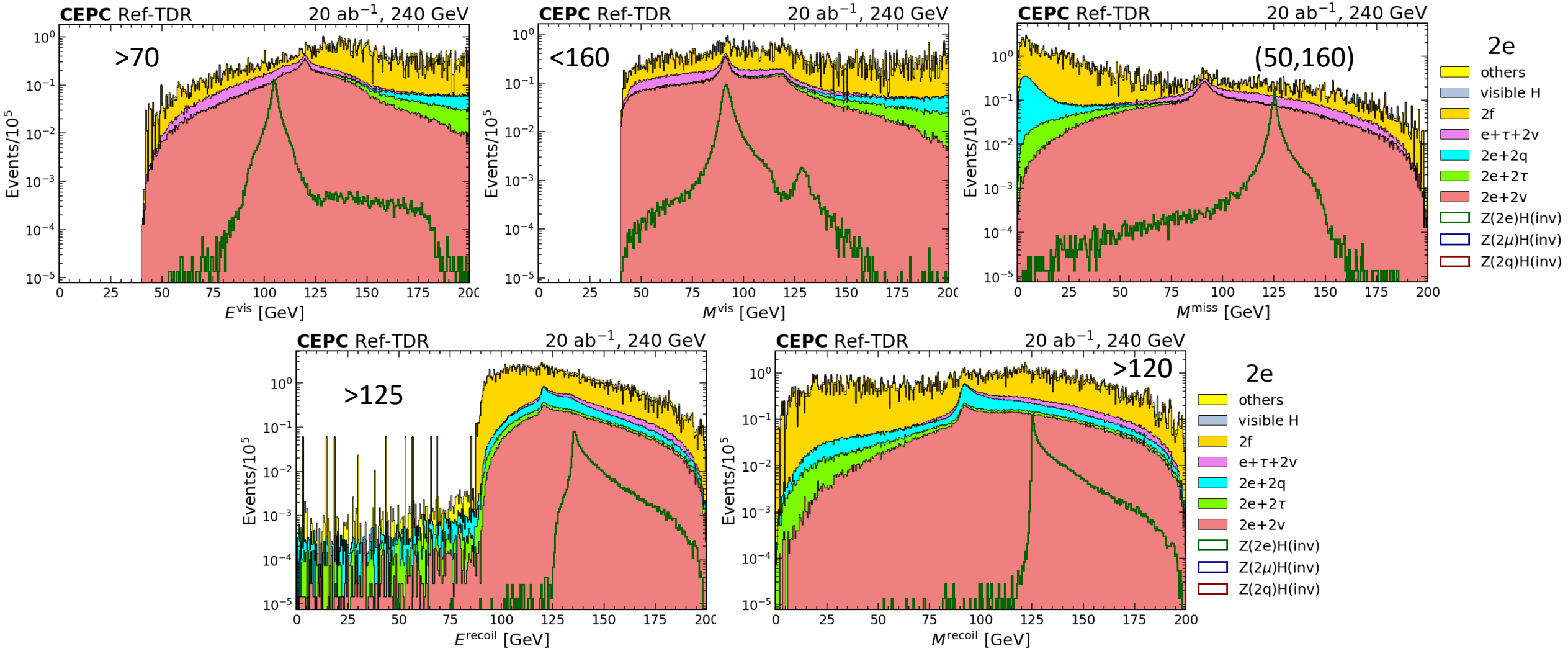






Kinematic selection: 2e

Requirements: ensure high signal efficiency (>98%); on condition of that, reduce backgrounds.



Kinematic selection: 2e

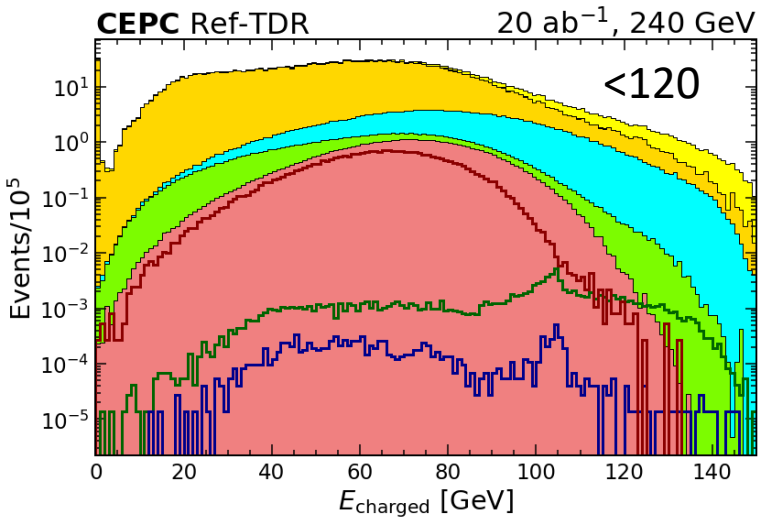
Cutflow

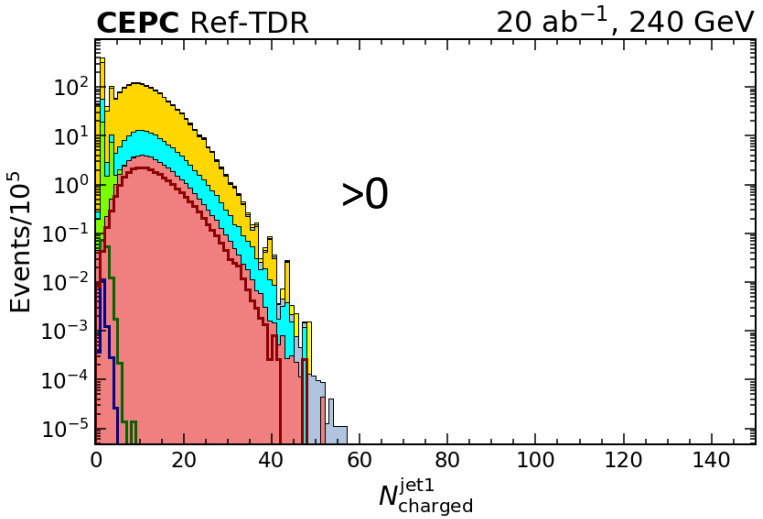
---------2e channel----------

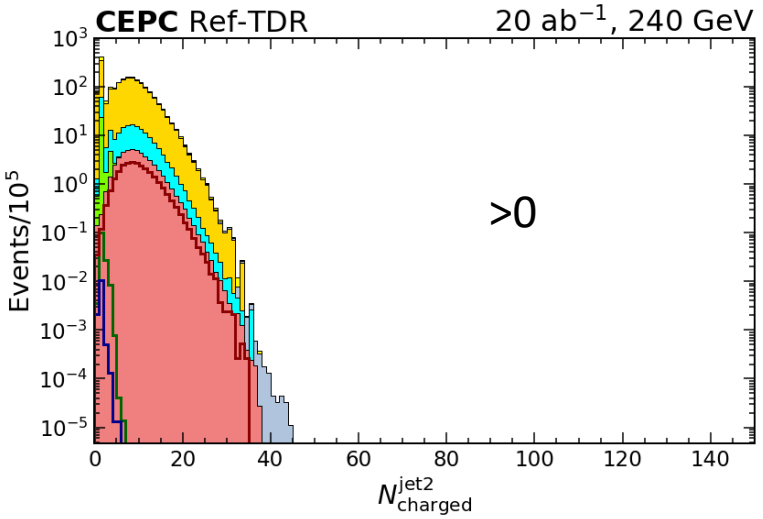
process	total	base eff	kinsel eff	selected
Z(2mu)H(inv)	1.435e+02	0.000 %	0.000 %	0
Z(2e)H(inv)	1.492e+02	83.747 %	97.591 %	122
Z(2q)H(inv)	2.900e+03	0.000 %	0.000 %	0
2e+2v	5.568e+06	41.675 %	33.039 %	766710
2e+2tau	2.946e+06	15.539 %	7.520 %	34421
2e+2q	6.321e+06	29.459 %	0.013 %	239
e+tau+2v	8.719e+06	9.921 %	29.786 %	257635
2f	1.779e+09	1.031 %	5.742 %	1053087
hx	4.073e+06	1.960 %	2.479 %	1980
others	3.609e+08	0.795 %	7.024 %	201625

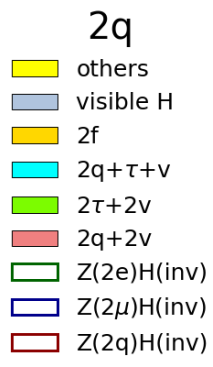
Kinematic selection: 2q

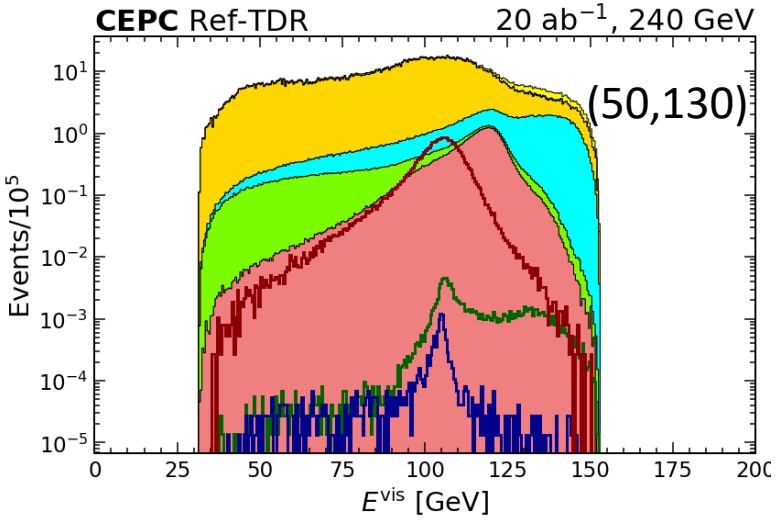
Requirements: ensure high signal efficiency (>98%); on condition of that, reduce backgrounds.

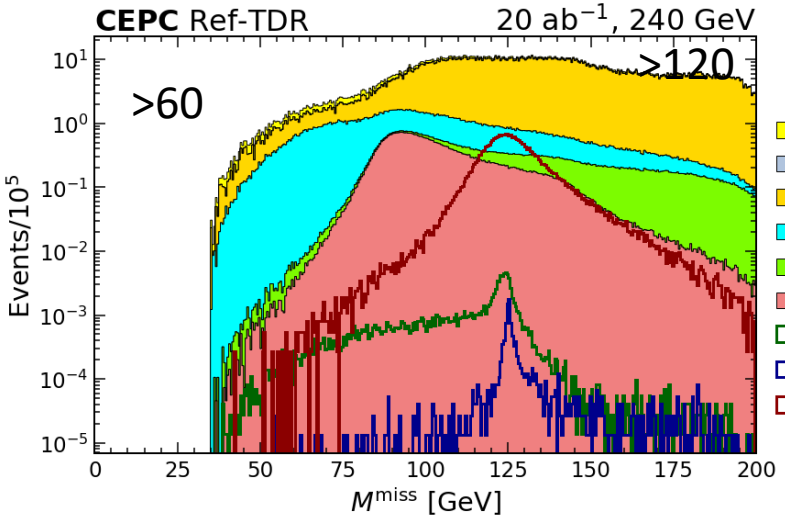


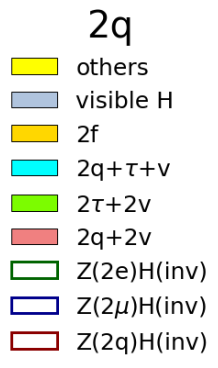












Kinematic selection: 2q

Cutflow

---------2q channel----------

process	total	base eff	kinsel eff	selected
Z(2mu)H(inv)	1.435e+02	0.984 %	77.541 %	1
Z(2e)H(inv)	1.492e+02	10.146 %	69.929 %	11
Z(2q)H(inv)	2.900e+03	98.967 %	99.366 %	2852
2q+2v	7.394e+06	66.069 %	95.304 %	4655834
2tau+2v	4.707e+06	60.975 %	83.916 %	2408589
2q+tau+v	5.654e+07	25.990 %	55.474 %	8152441
2f	1.779e+09	9.236 %	84.399 %	138692035
hx	4.073e+06	19.758 %	55.287 %	444932
others	3.158e+08	4.403 %	61.924 %	8609893

XGBoost models

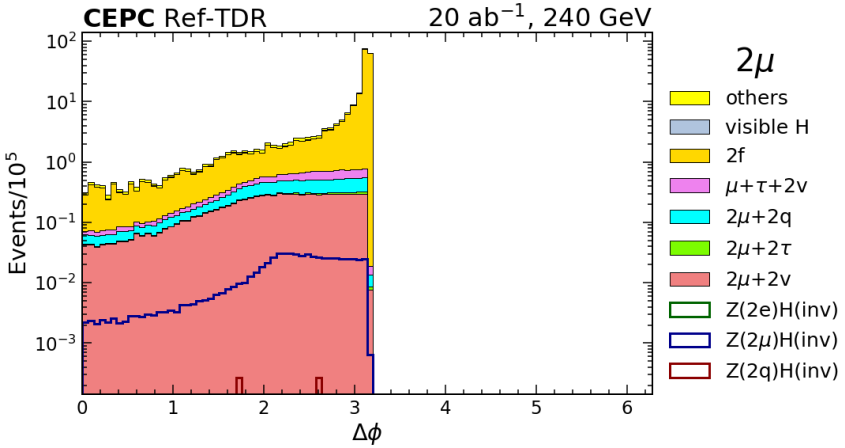
Discriminate between signals and backgrounds

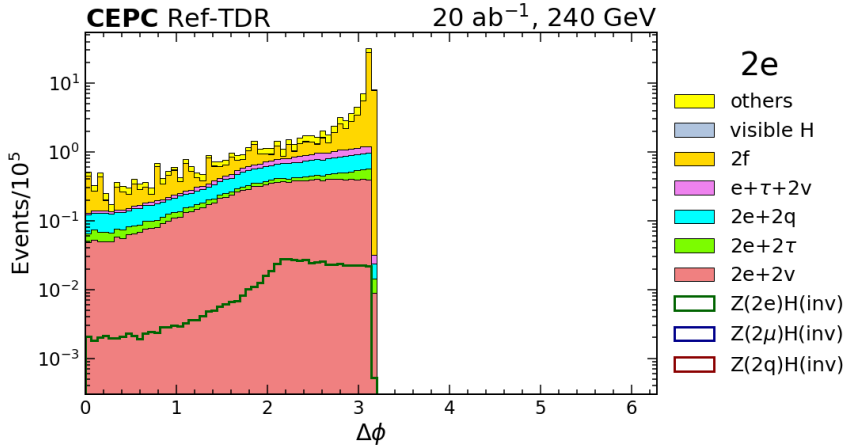
Input features

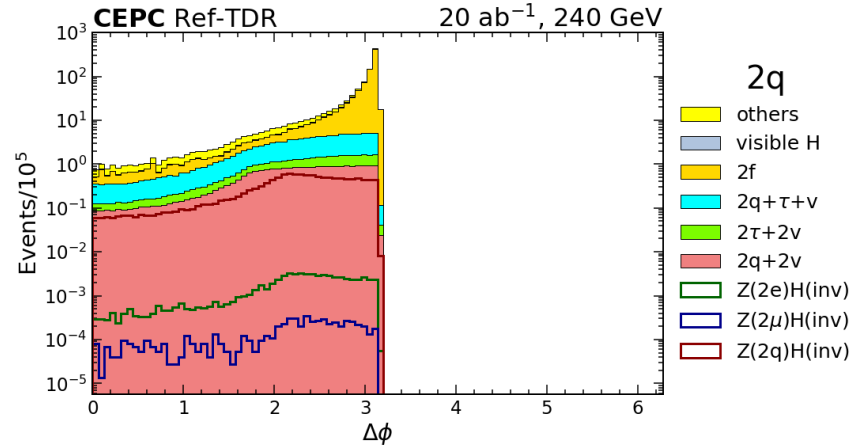
- **2mu / 2e:** nChargedParticles, nNeutralParticles, E_NeutralParticles, visible pt/E/M, missing M, ll pt/E/M, recoil M, $\Delta\phi(\mathrm{ll})$, $\Delta\mathrm{R}(\mathrm{ll})$, lepton D0 / Z0
- **2q:** nChargedParticles, nNeutralParticles, visible pt/p/E/M, missing M, $\Delta\phi(\mathrm{jj})$, $\Delta\mathrm{R}(\mathrm{jj})$, jet ECF2 / τ_{31} / nElectrons / nMuons / nChargedHadrons, jet ymerge2, ymerge 5.

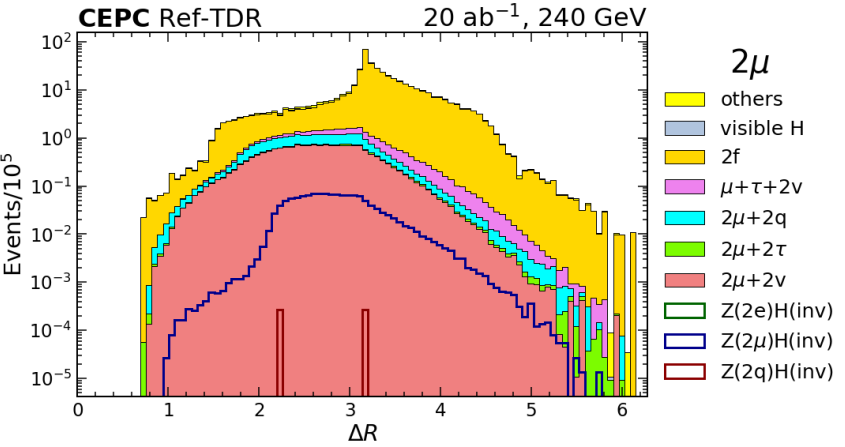
XGBoost models

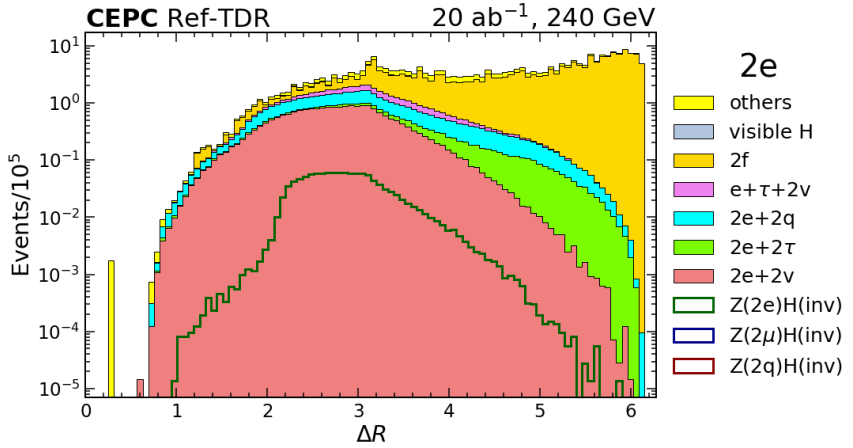
$\Delta\phi$, ΔR distributions

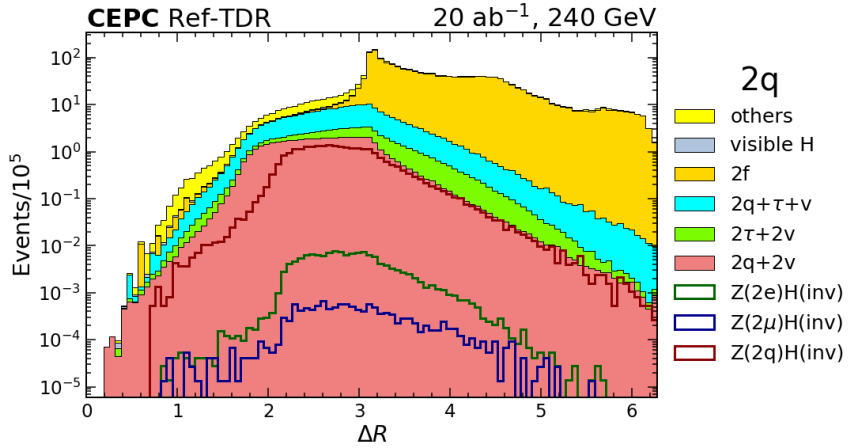






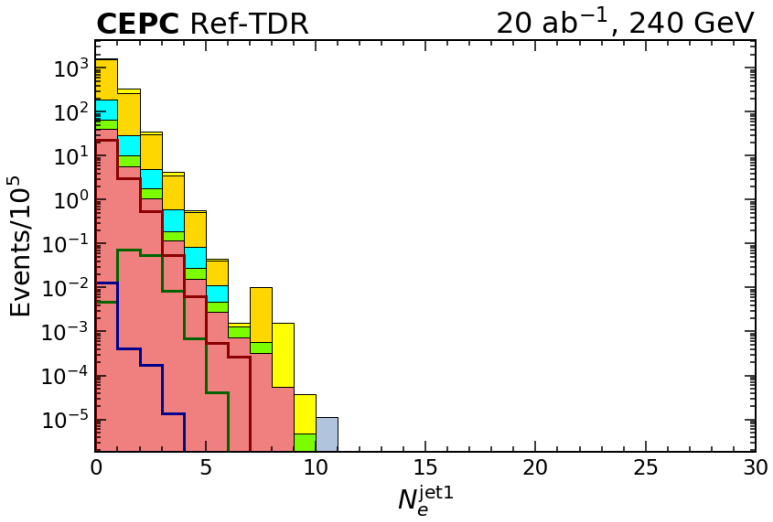


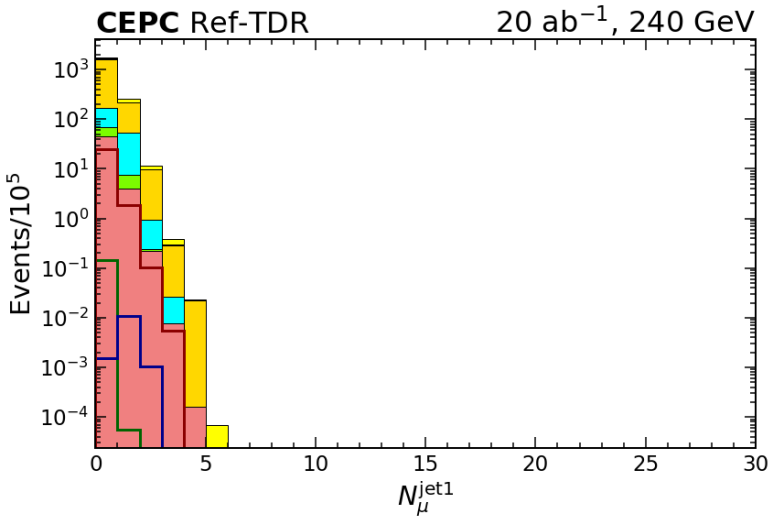


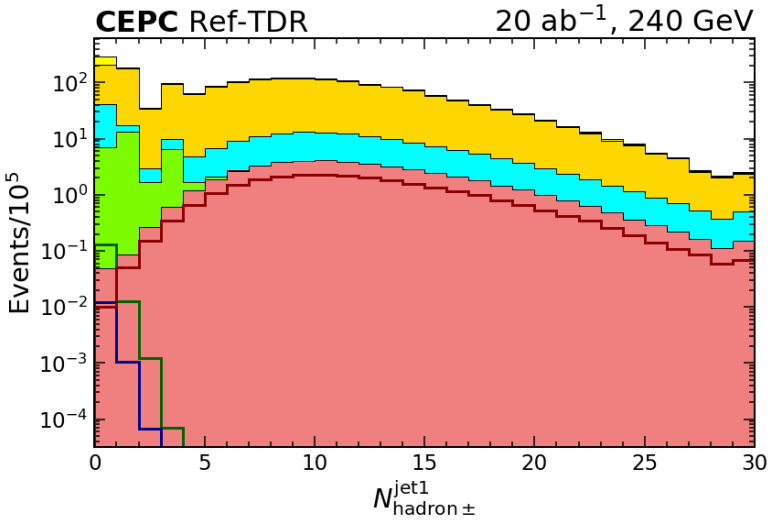


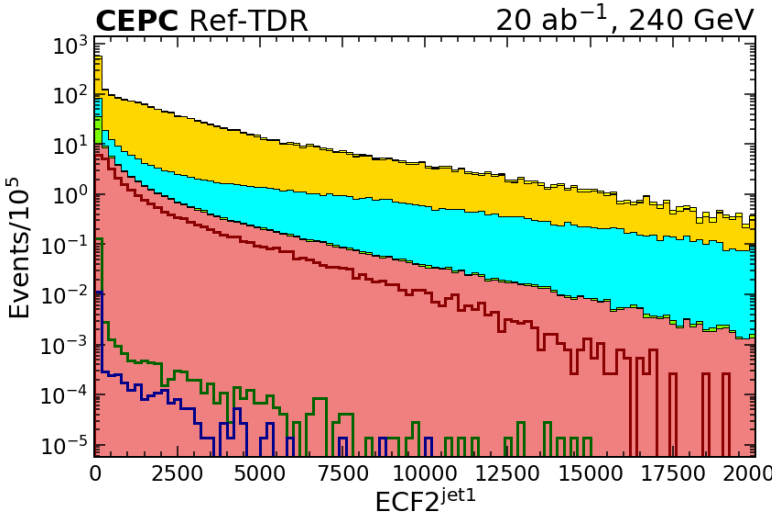
XGBoost models

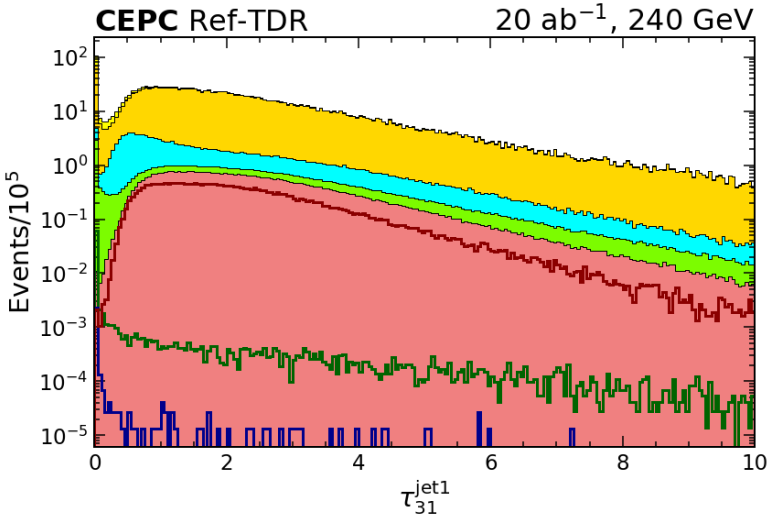
Jet substructure

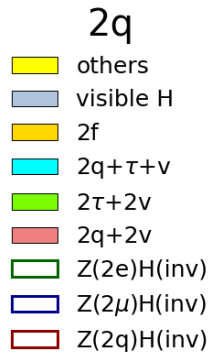


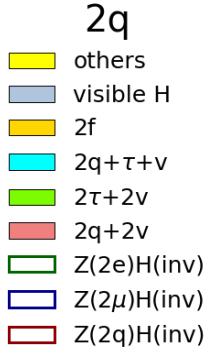






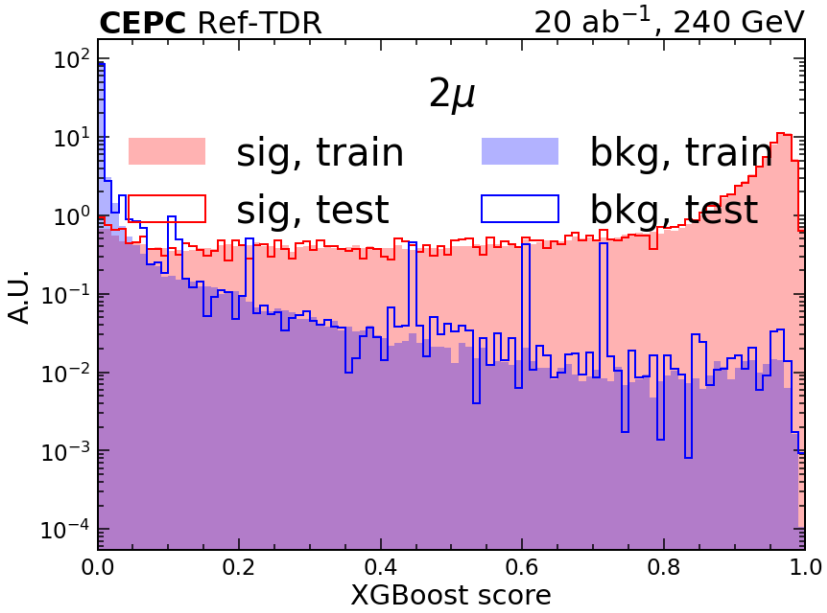


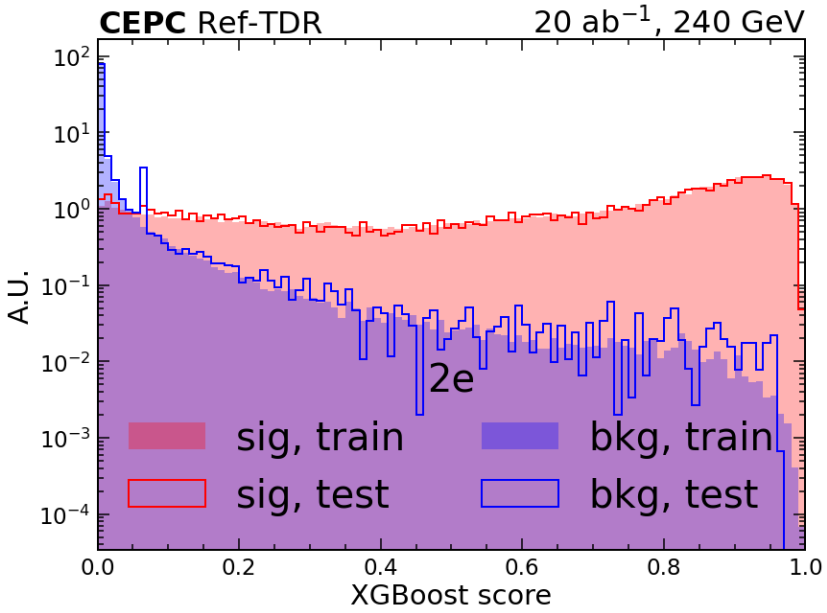


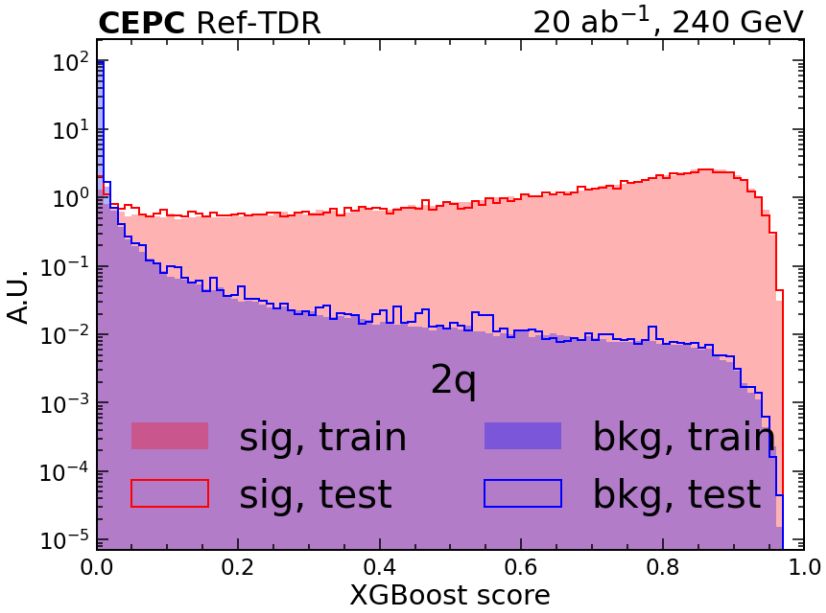


XGBoost models

Distribution of XGBoost score



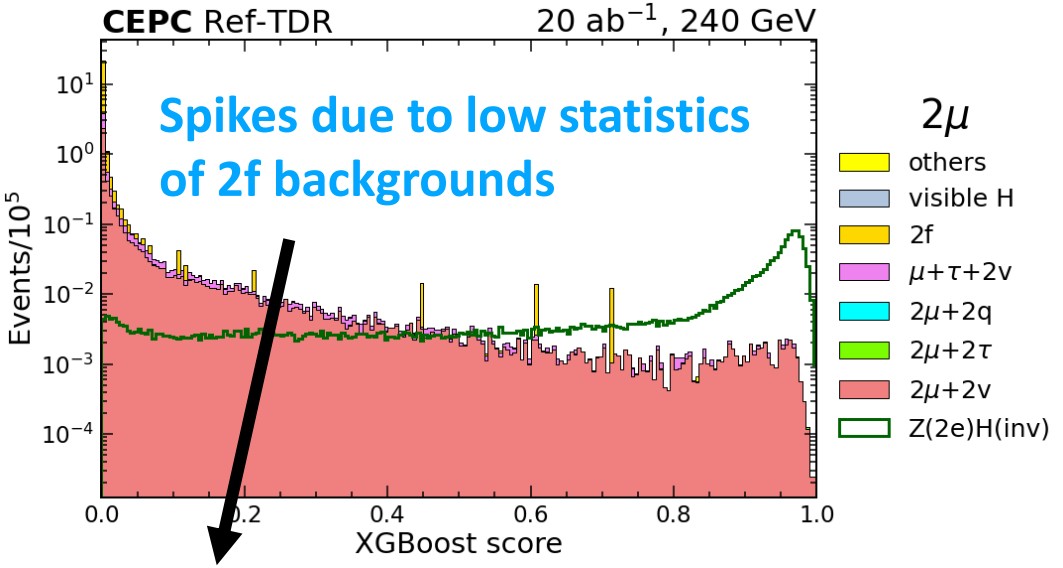


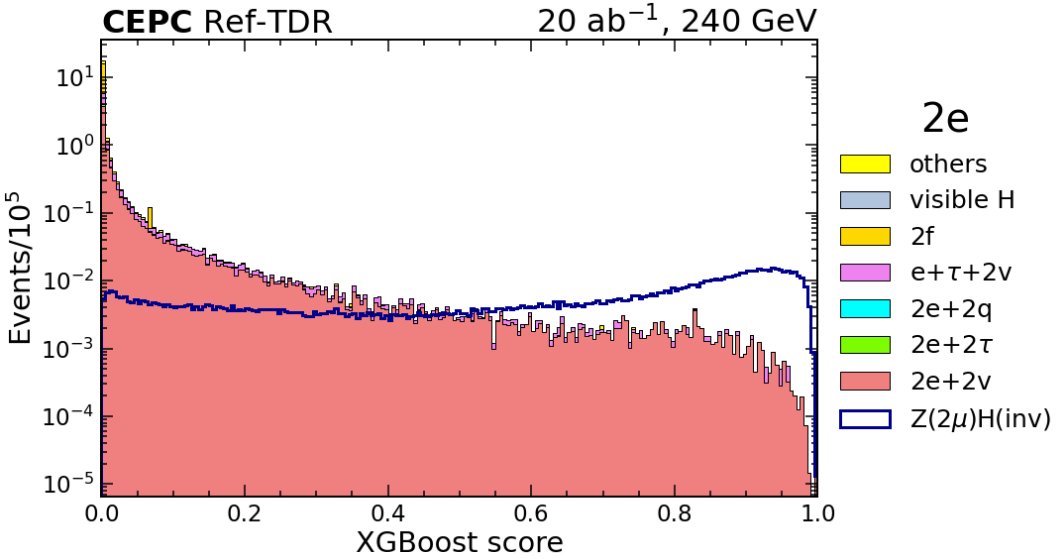


Very gentle overtraining effects.

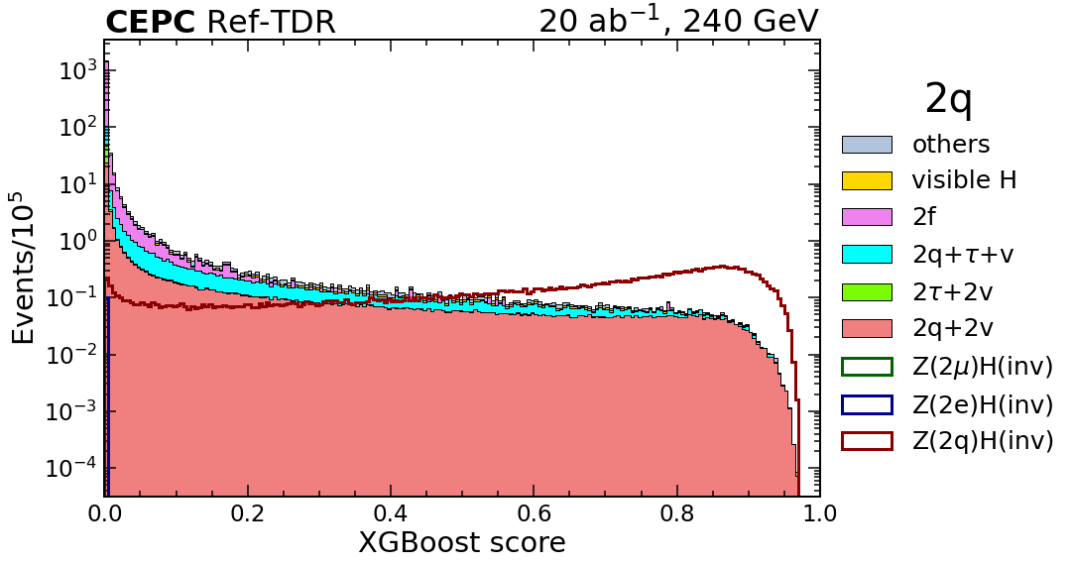
XGBoost models

Distribution of XGBoost score





If trained on the full dataset, spikes will disappear.



Highest contamination: irreducible backgrounds

New results

With 5.6 ab-1 data

Final state	New result	Last version	CDR
ee	$(1.00^{+2.42}_{-1.00}) \cdot 0.1\%$	$(1.00^{+4.68}_{-1.00}) \cdot 0.1\%$	$\pm 4.54 \cdot 0.1\%$
μμ	$(1.00^{+1.48}_{-1.00}) \cdot 0.1\%$	$\left(1.00^{+2.53}_{-1.00}\right) \cdot 0.1\%$	$\pm 2.36 \cdot 0.1\%$
qq	$(1.00^{+1.10}_{-1.00}) \cdot 0.1\%$	$(1.00^{+1.21}_{-1.00}) \cdot 0.1\%$	$\pm 0.95 \cdot 0.1\%$
All	$(1.00^{+0.82}_{-0.80}) \cdot 0.1\%$	$\left(1.00^{+1.05}_{-0.92}\right) \cdot 0.1\%$	$\pm 0.87 \cdot 0.1\%$

Compared to last version:

- Significant improvements in leptonic channels.
- Very little improvement in 2q channel.

Compared to CDR:

- Results in leptonic channels much better.
- Results in 2q channel worse.
- Overall result better.

Need to explore more in the 2q channel, because we expect the XGBoost model to make more improvements.