ZH studies with BDT selection

Lucien Sochard

Supervised by Gregorio Bernardi and Ang Li

Jamboree FCC, 1st June 2022

Introduction : BDT training

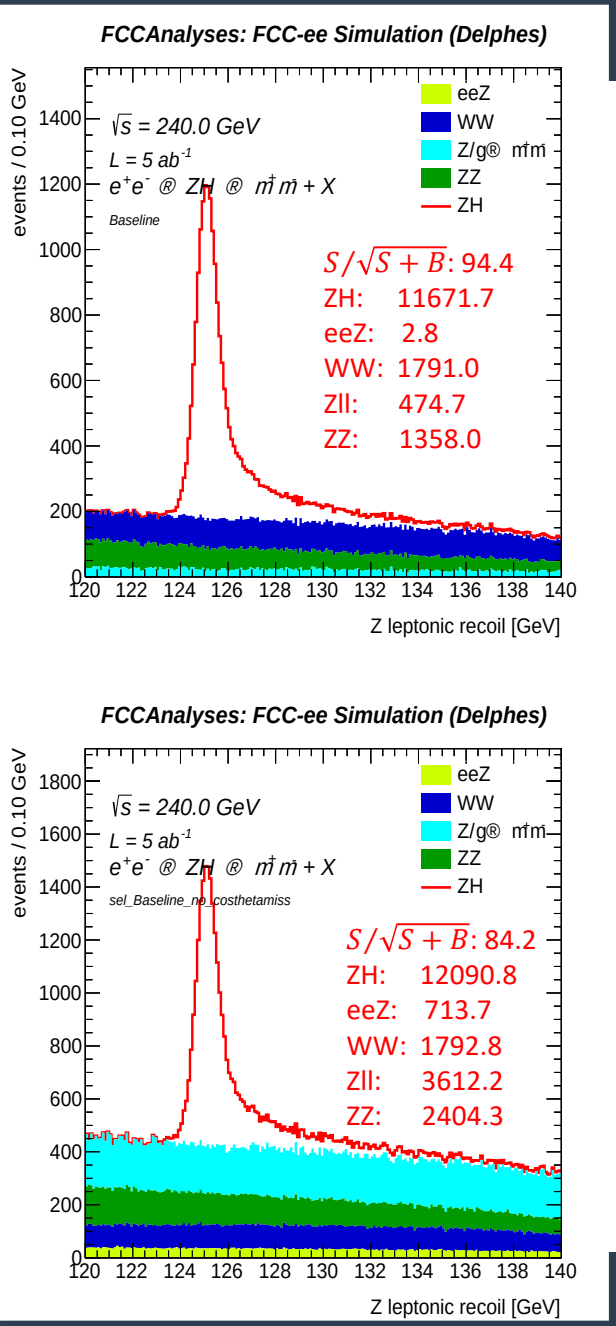
GOAL :

- Model independent approach
  → no $|\cos\theta_{missing}|$ cut
- Increase the signal statistics, while cutting as much background as possible

WORKING PRINCIPLE :

- Give the BDT a set of variables
- Try to optimize this set
- Try to optimize the other parameters of the BDT model

→ When we have the best background cut with a good signal statistics, fit the background to remove it



Remove $|\cos\theta_{missing}|$ cut

Set of variables

→ First draft : 25 variables

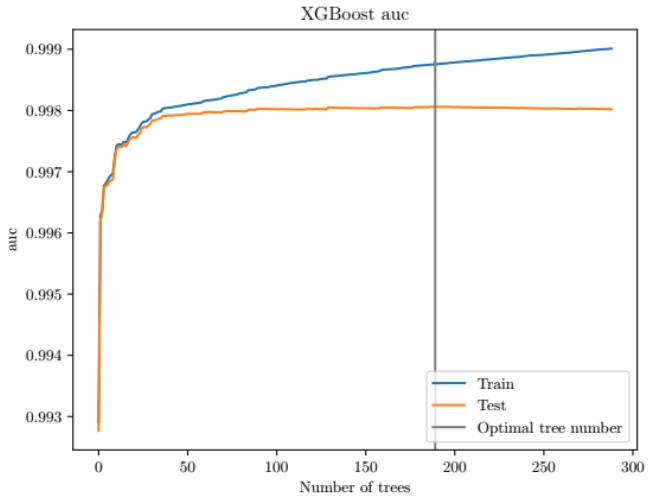
Maximal and minimal Angle between 2 muons in the event ←

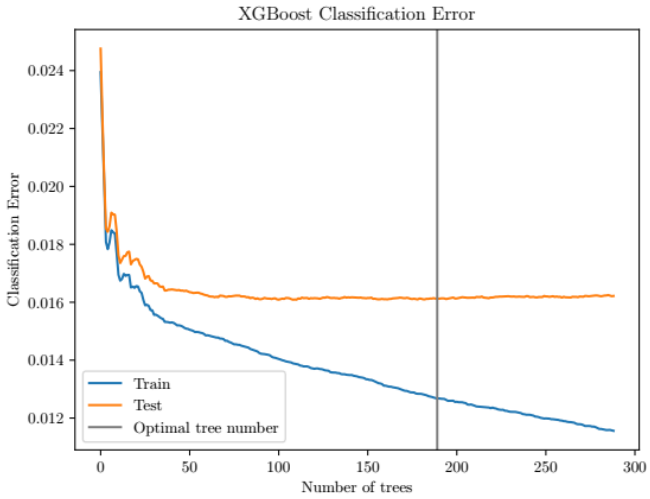
(When there are only 2 muons, delta_max=delta_min)

1. selected_muons_delta_max
2. selected_muons_delta_min
3. muon_leading_pt
4. muon_leading_px
5. muon_leading_py
6. muon_leading_pz
7. muon_leading_eta
8. muon_leading_e
9. muon_leading_theta
10. muon_subleading_pt
11. muon_subleading_px
12. muon_subleading_py
13. muon_subleading_pz
14. muon_subleading_eta
15. muon_subleading_e
16. muon_subleading_theta
17. Z_leptonic_m
18. Z_leptonic_pt
19. Z_leptonic_y
20. Z_leptonic_p
21. Z_leptonic_px
22. Z_leptonic_py
23. Z_leptonic_pz
24. Z_leptonic_eta
25. Z_leptonic_theta

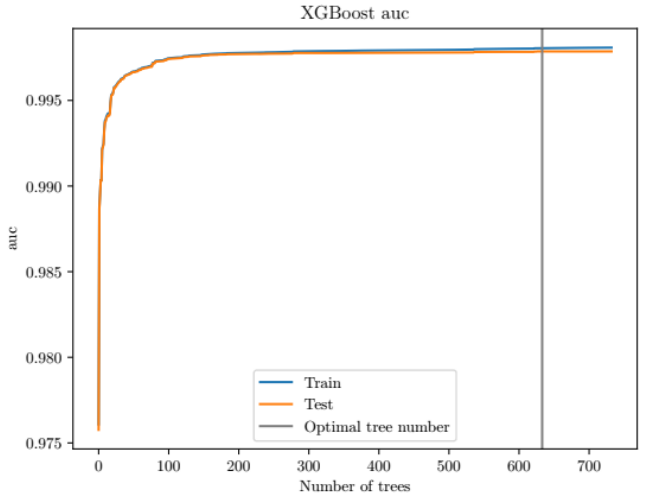
Max depth optimization

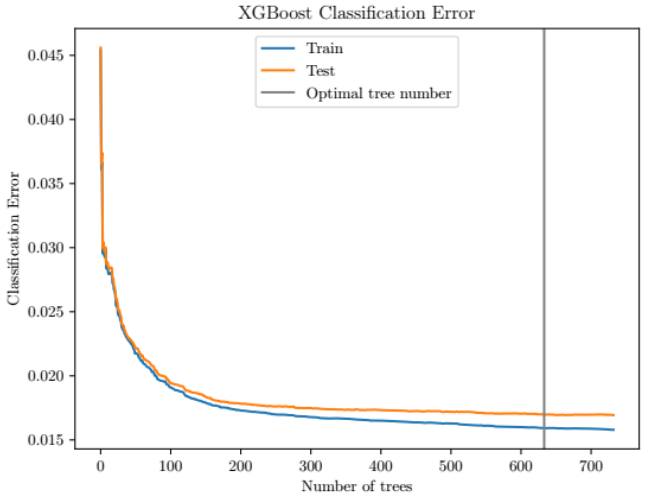
Max_depth=10



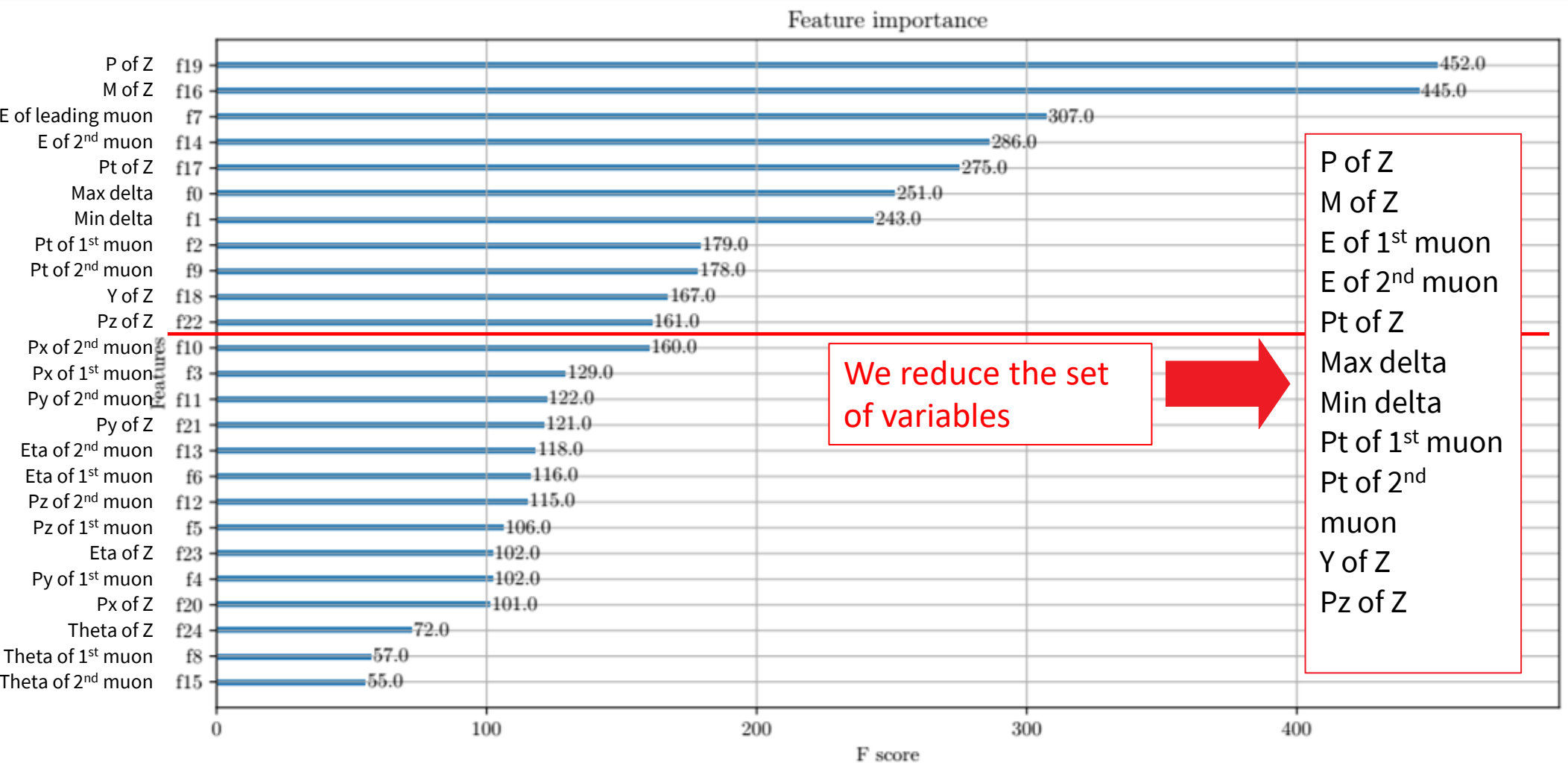


Max_depth=3

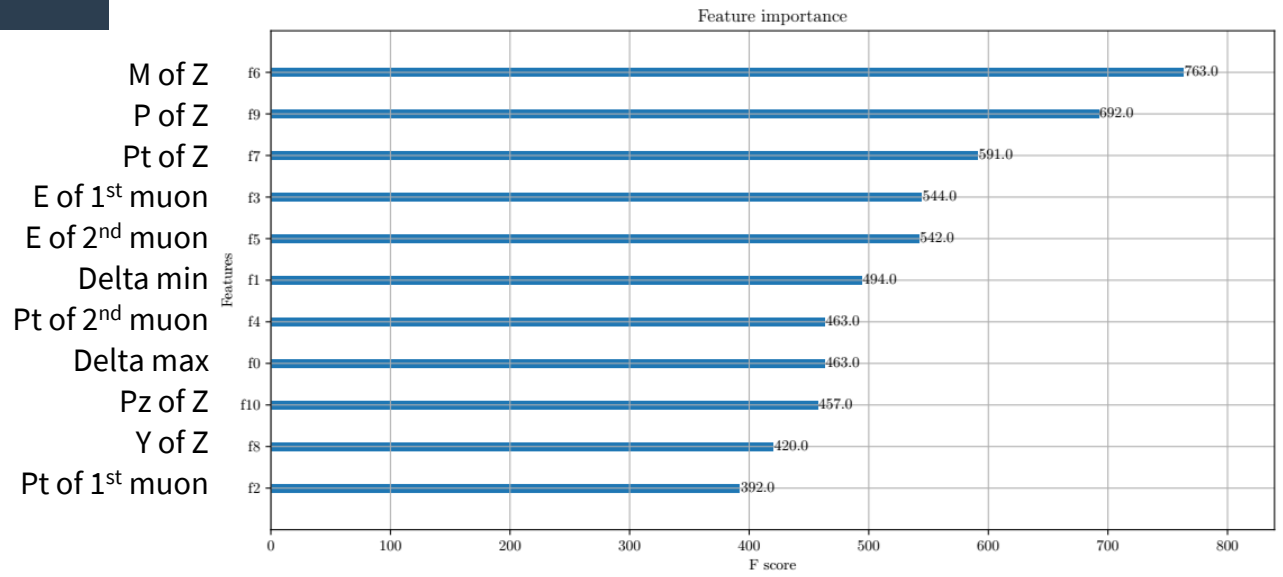




First set : 25 variables



Feature importance



P of Z	Pt of 1st muon
M of Z	Pt of 2nd
E of 1st muon	muon
E of 2nd muon	Y of Z
Pt of Z	Pz of Z
Max delta	
Min delta	

APC3_MVA08

ZH : 14490.8

Backgrounds : total : 20217.6

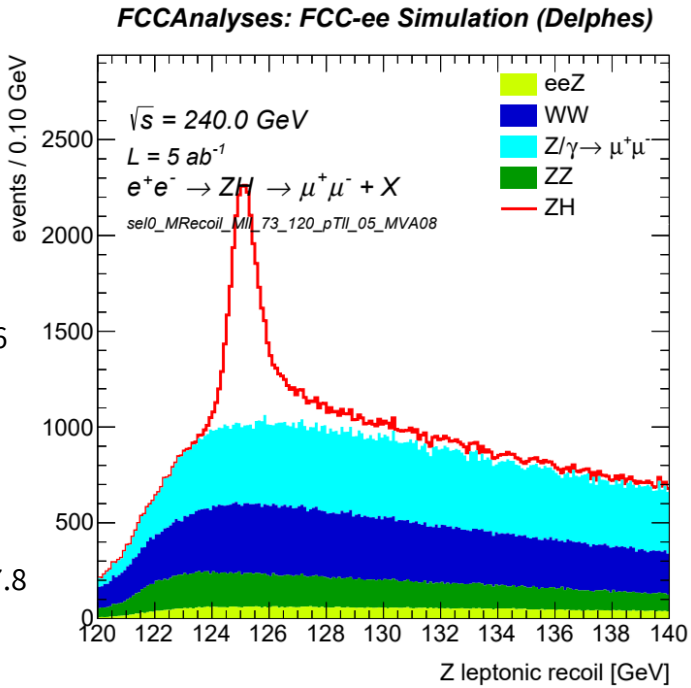
EeZ : 1214.7

WW : 7149.7

ZZ : 3541.2

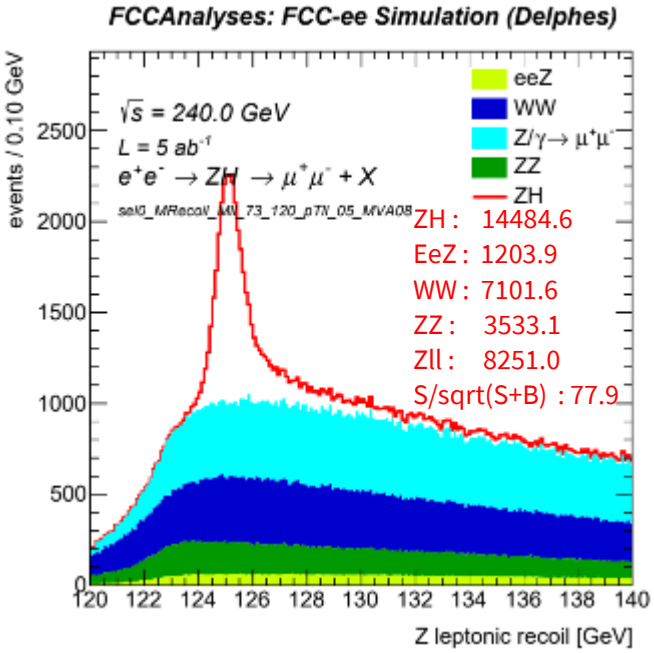
Zll : 8312.0

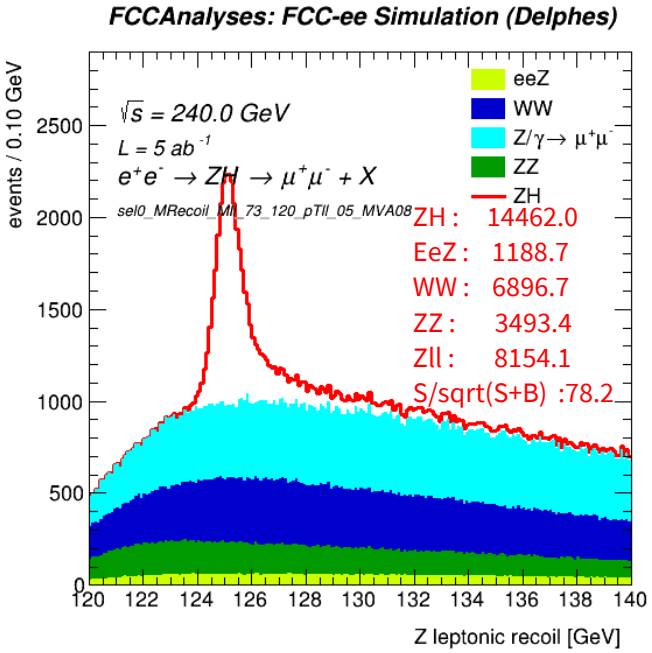
Significance S/sqrt(S+B) : 77.8



→We can remove Y of Z and Pz of Z

Z recoil plots, reduced variable set without pz and Y of Z





Don't want this massive cut on the background on the tail

The BDT calculates the recoil mass and choses the events around it.
P of Z is in the formula :

$$M^2_{recoil} = (\sqrt{s} - E_{l\bar{l}})^2 - p^2_{l\bar{l}}$$

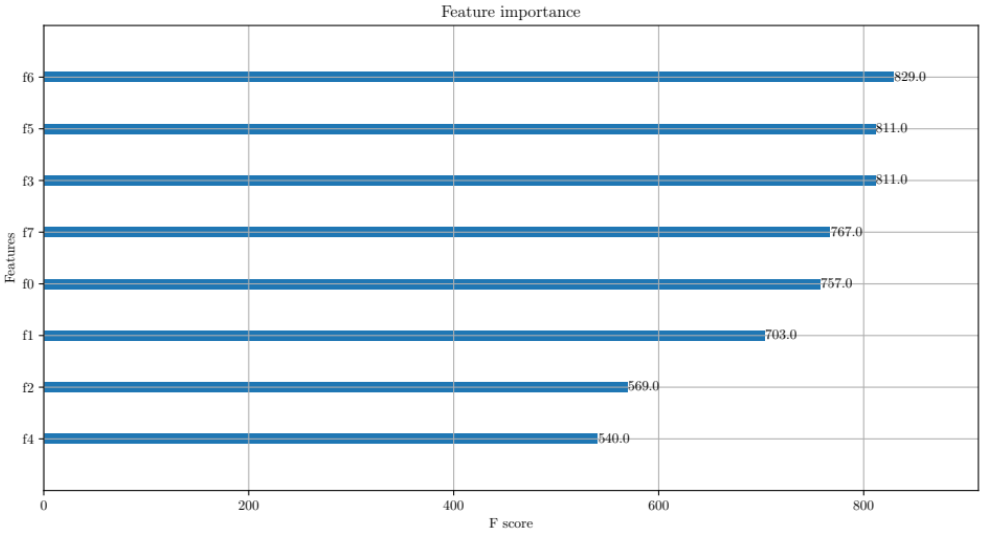
→ Remove P of Z

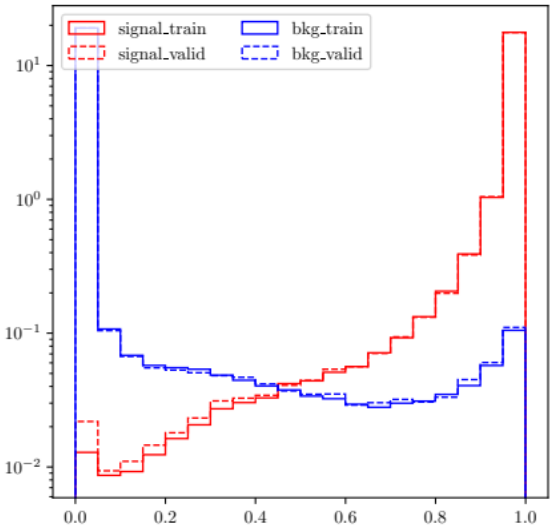
P of Z	Max delta
M of Z	Min delta
E of 1st muon	Pt of 1st muon
E of 2nd muon	Pt of 2nd muon
Pt of Z	~~Y of Z~~
	~~Pz of Z~~

~~P of Z~~	Max delta
M of Z	Min delta
E of 1st muon	Pt of 1st muon
E of 2nd muon	Pt of 2nd muon
Pt of Z	~~Y of Z~~
	~~Pz of Z~~

Final set : without P of Z

M of Z
E of 2nd muon
E of 1st muon
Pt of Z
Delta max
Delta min
Pt of 1st muon
Pt of 2nd muon





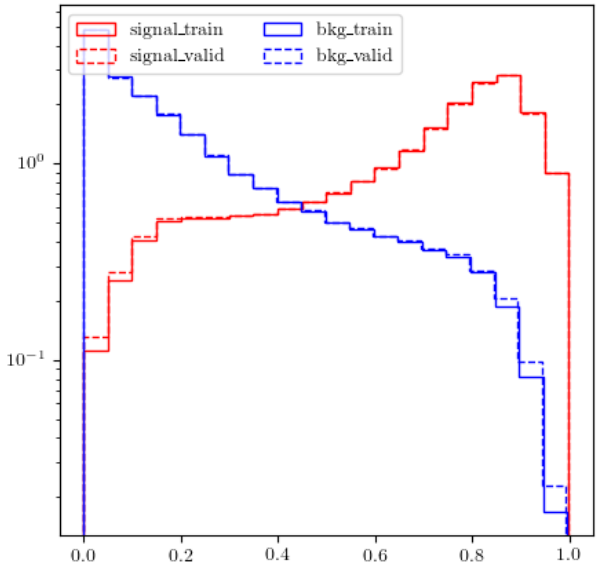
→ Pb : The background goes up at high BDT scores.

(Here, the training is done on the simulated events only requesting the presence of a Z → μμ)

→ Train the BDT on samples that have stricter selection

Training on selected samples

M of Z	Max delta
E of 1st muon	Min delta
E of 2nd muon	Pt of 1st muon
Pt of Z	Pt of 2nd muon

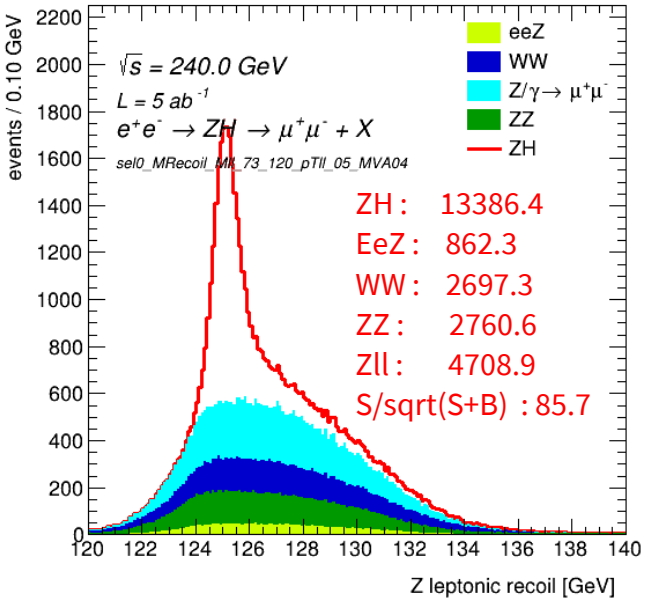


Applied Selection before BDT training :

APC-3-Selection

1. At least one Z boson from a μ+μ- pair
2. M_{μ+μ-} in [73,120] GeV
3. M_recoil in [120,140] GeV
4. Pt_{μ+μ-} in [5,+∞] GeV

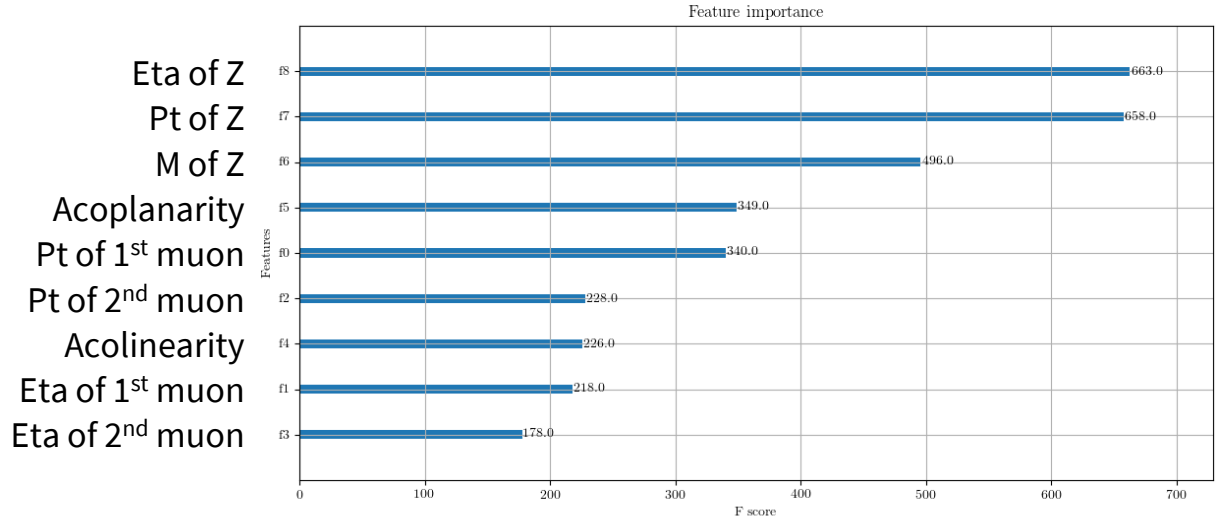
APC-3-MVA04

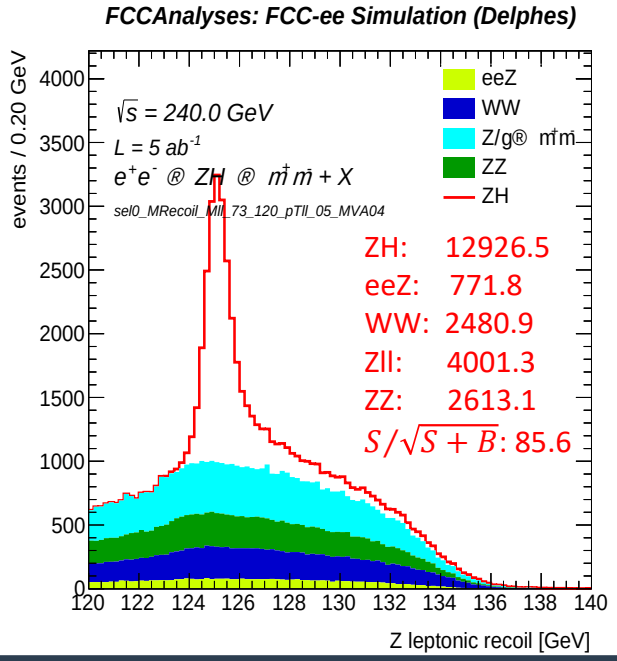


New variables

M of Z
eta of 1st muon
eta of 2nd muon
eta of Z
Pt of Z
→ acolinearity
→ acoplanarity
Pt of 1st muon
Pt of 2nd muon

(Set used by Ang)





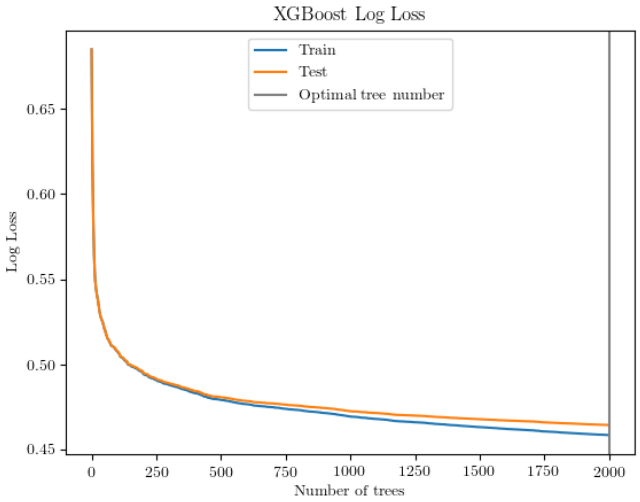
APC-3-MVA04

Better shape of the background on the tails

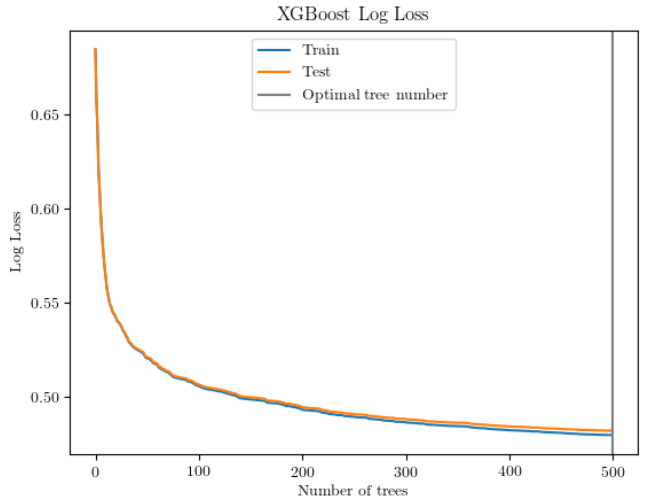
→ Replace delta min and delta max by acolinearity and acoplanarity in the previous set

Number of trees optimization : reduce over-training

M of Z
E of 1st muon
E of 2nd muon
Pt of Z
acolinearity
acoplanarity
Pt of 1st muon
Pt of 2nd muon

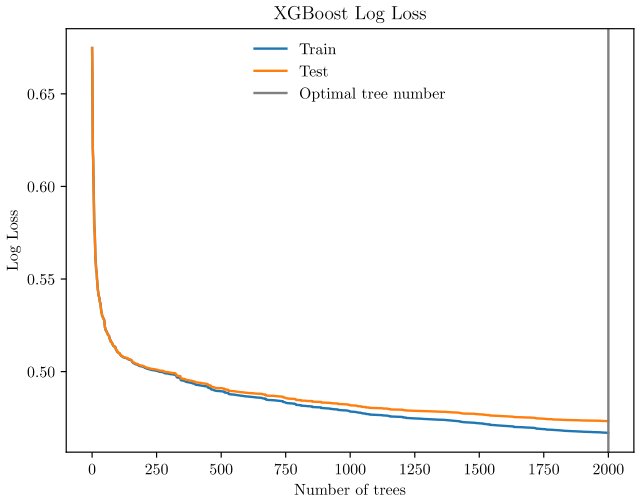


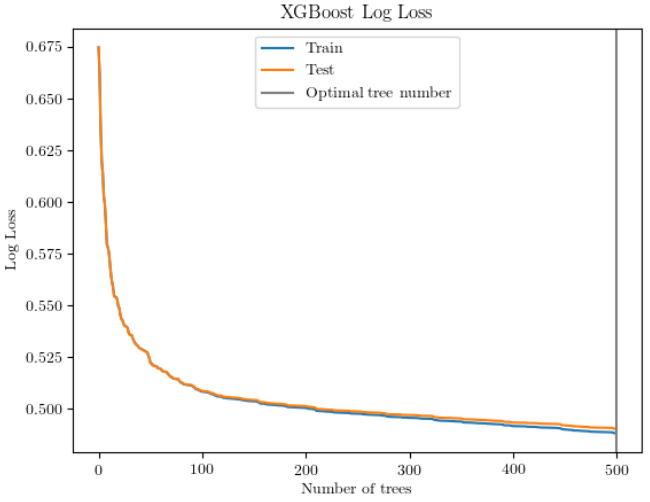
Max number of trees = 2000



Max number of trees = 500

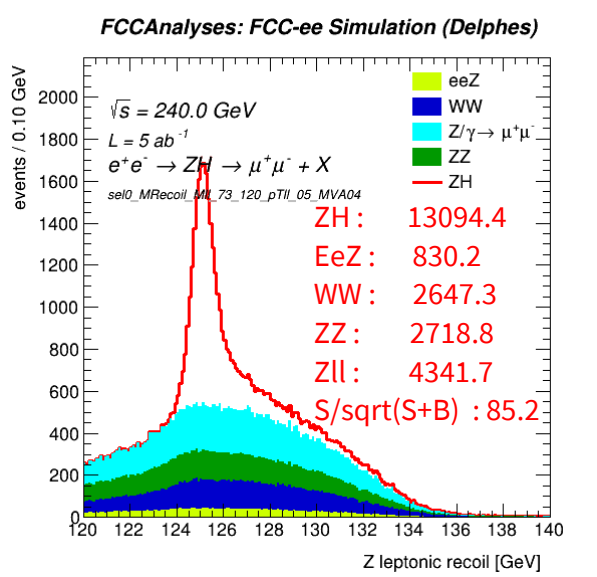
M of Z
eta of 1st muon
eta of 2nd muon
eta of Z
Pt of Z
acolinearity
acoplanarity
Pt of 1st muon
Pt of 2nd muon

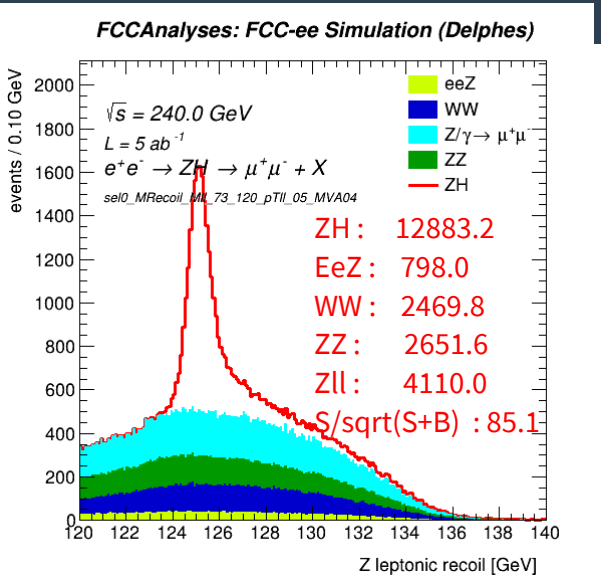




Number of trees optimization : Z recoil plots

M of Z
E of 1st muon
E of 2nd muon
Pt of Z
acolinearity
acoplanarity
Pt of 1st muon
Pt of 2nd muon

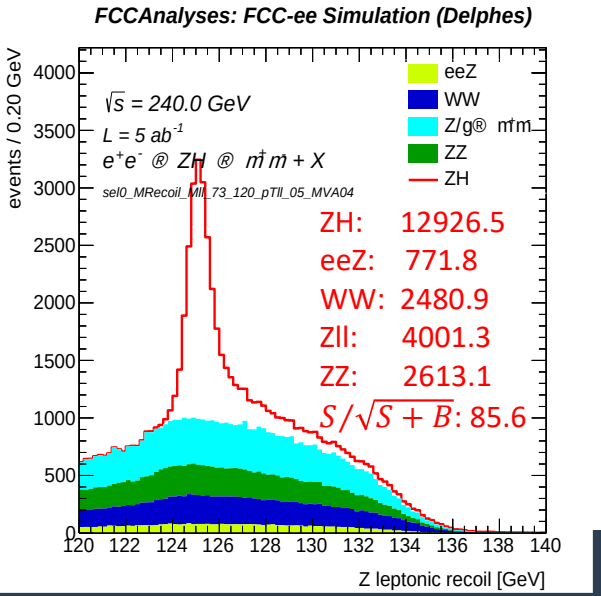


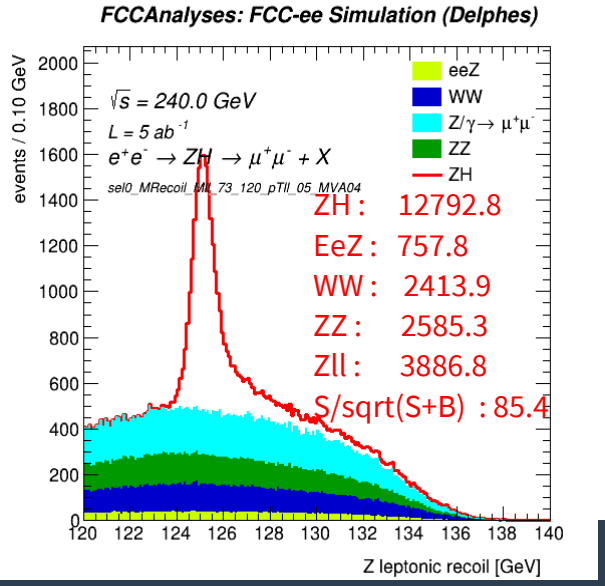


Max number of trees = 2000

Max number of trees = 500

M of Z
eta of 1st muon
eta of 2nd muon
eta of Z
Pt of Z
acolinearity
acoplanarity
Pt of 1st muon
Pt of 2nd muon



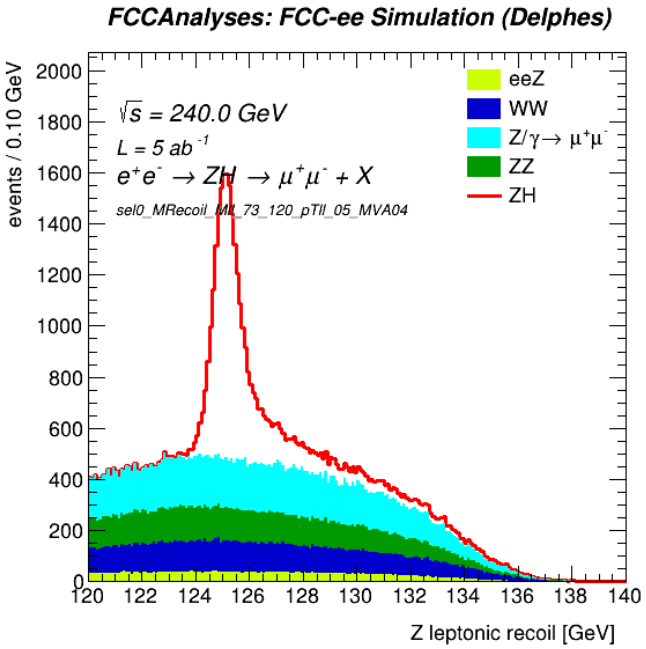


Current point : Z recoil plots

M of Z
eta of 1st muon
eta of 2nd muon
eta of Z
Pt of Z
acolinearity
acoplanarity
Pt of 1st muon
Pt of 2nd muon

+

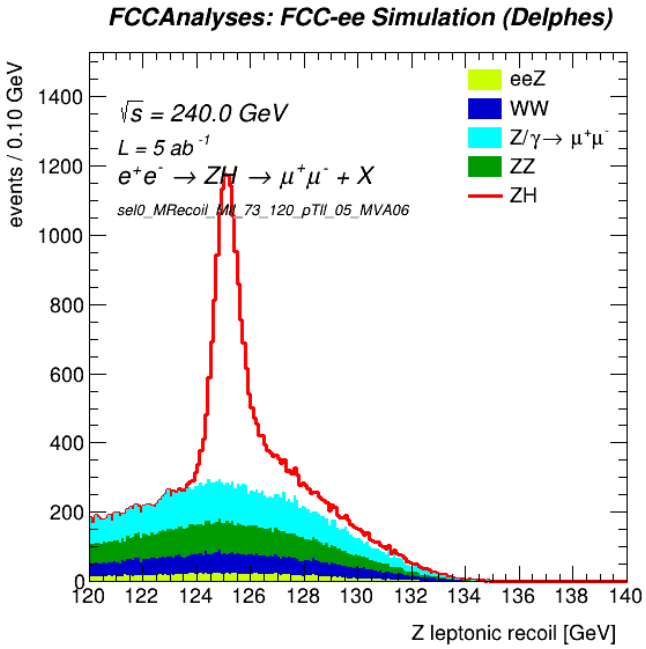
Max_depth = 3,
Max number of trees = 500
and events pre-selected with
APC-3-selection



APC-3-MVA04
ZH : 12792.8

Backgrounds : total : 9643.7
EeZ : 757.8
WW : 2413.9
ZZ : 2585.3
Zll : 3886.8

Significance S/sqrt(S+B) : 85.4



APC-3-MVA06
ZH : 10193.5

Backgrounds : total : 5604.4
EeZ : 453.1
WW : 1155.1
ZZ : 1745.1
Zll : 2251.1

Significance S/sqrt(S+B) : 81.1

Conclusion and perspective

SUMMARY

- Parameters of the BDT model optimized
  - maximum depth = 3
  - max number of trees = 500 (for now)
- Set of variables reduced : 2 different sets with respectively 8 and 9 variables
- With training after APC-3-selection, background tends to have a shape that we can try to fit

PERSPECTIVE

- Adjust the set of variables and number of trees to avoid over-training
- Train the BDT with $M_{\text{recoil}} \in [110, 150]$ GeV to better constrain the background with the tails