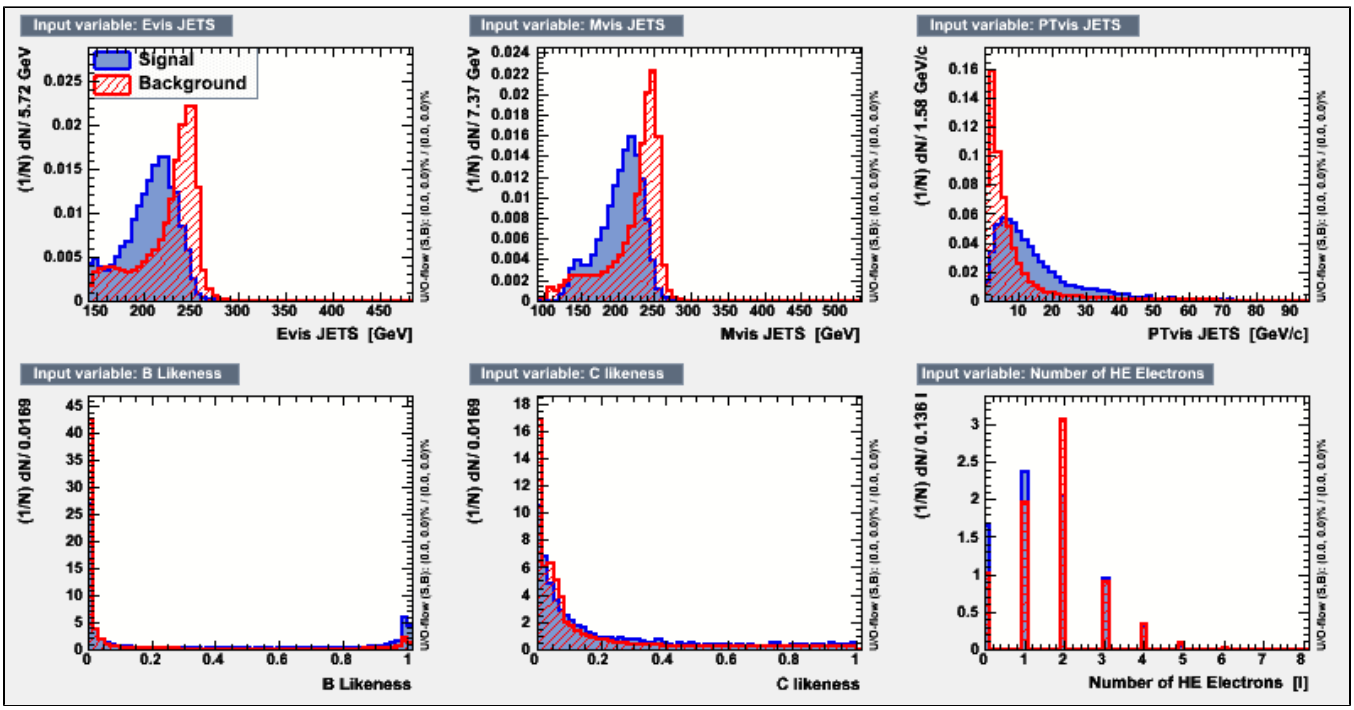
HZZ 250 GeV Analysis - Update 2 - 17-19 September 2013

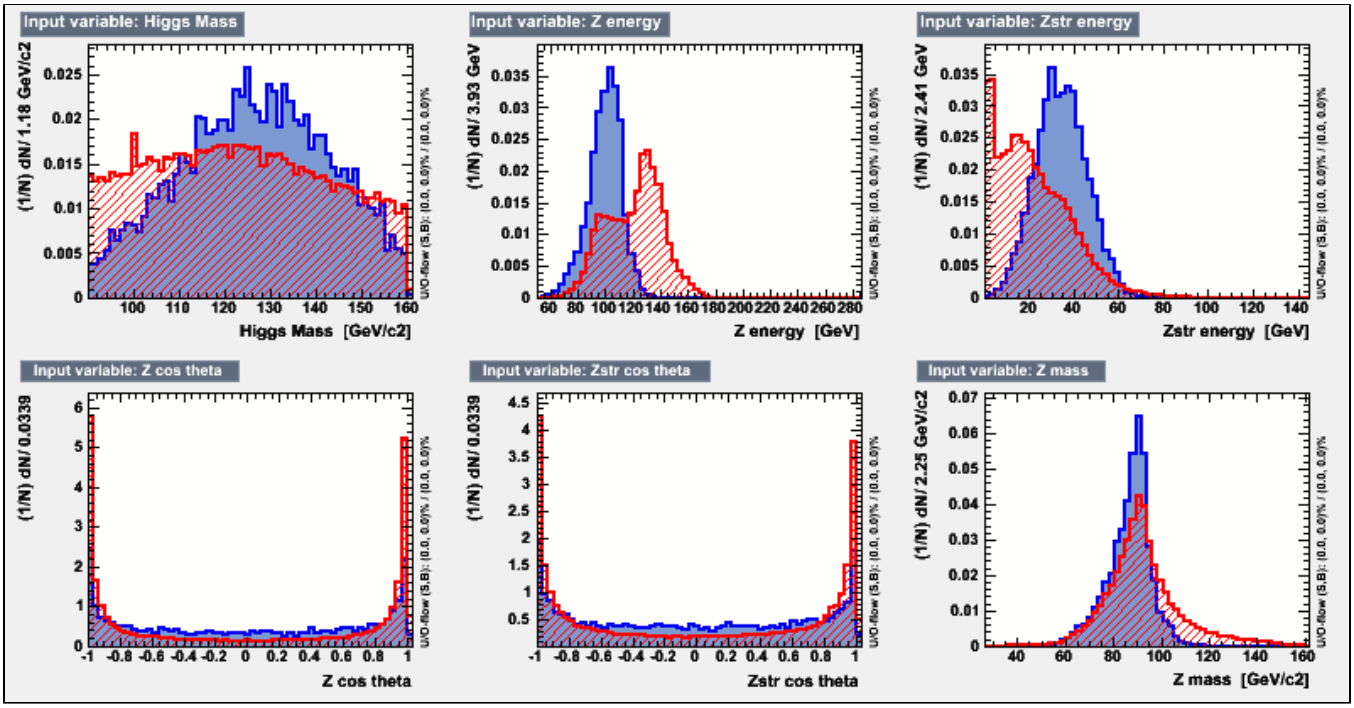
Update on the HZZ 250 GeV +80/-30 250/fb Analysis for 17-19 September 2013:

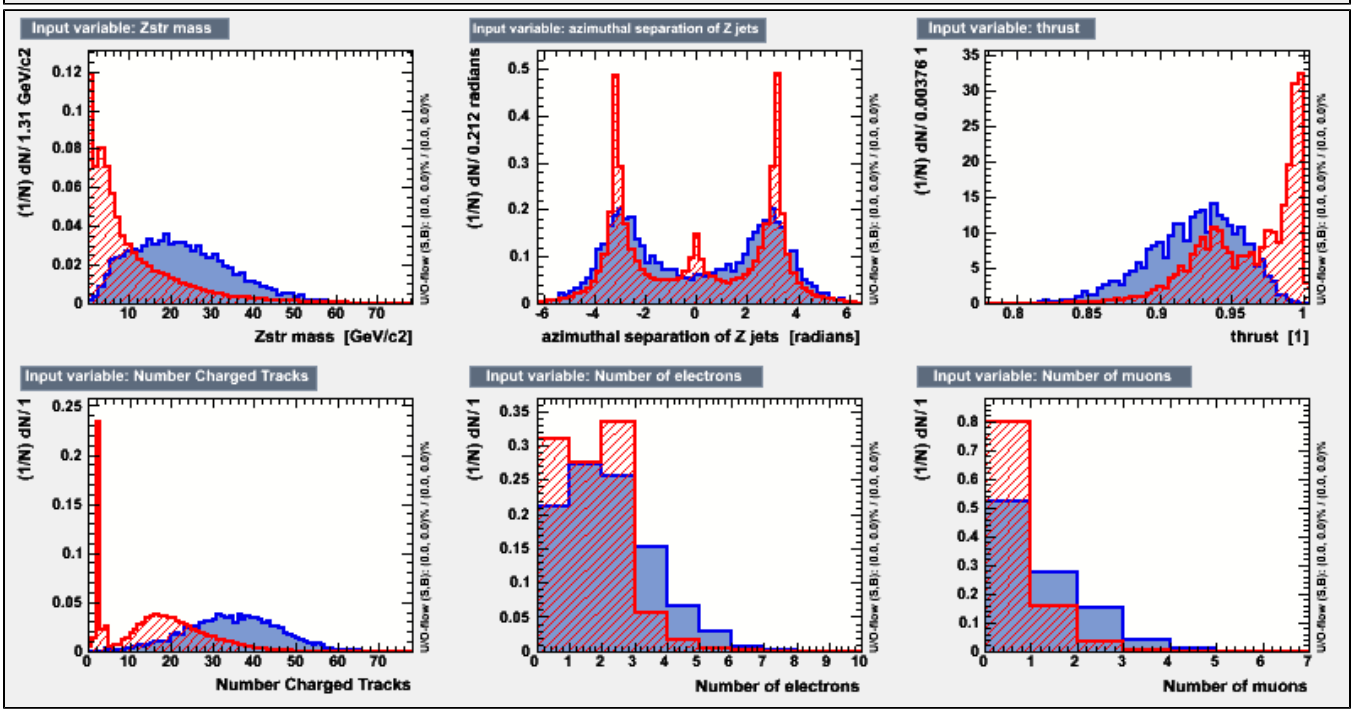
Analysis Procedure:

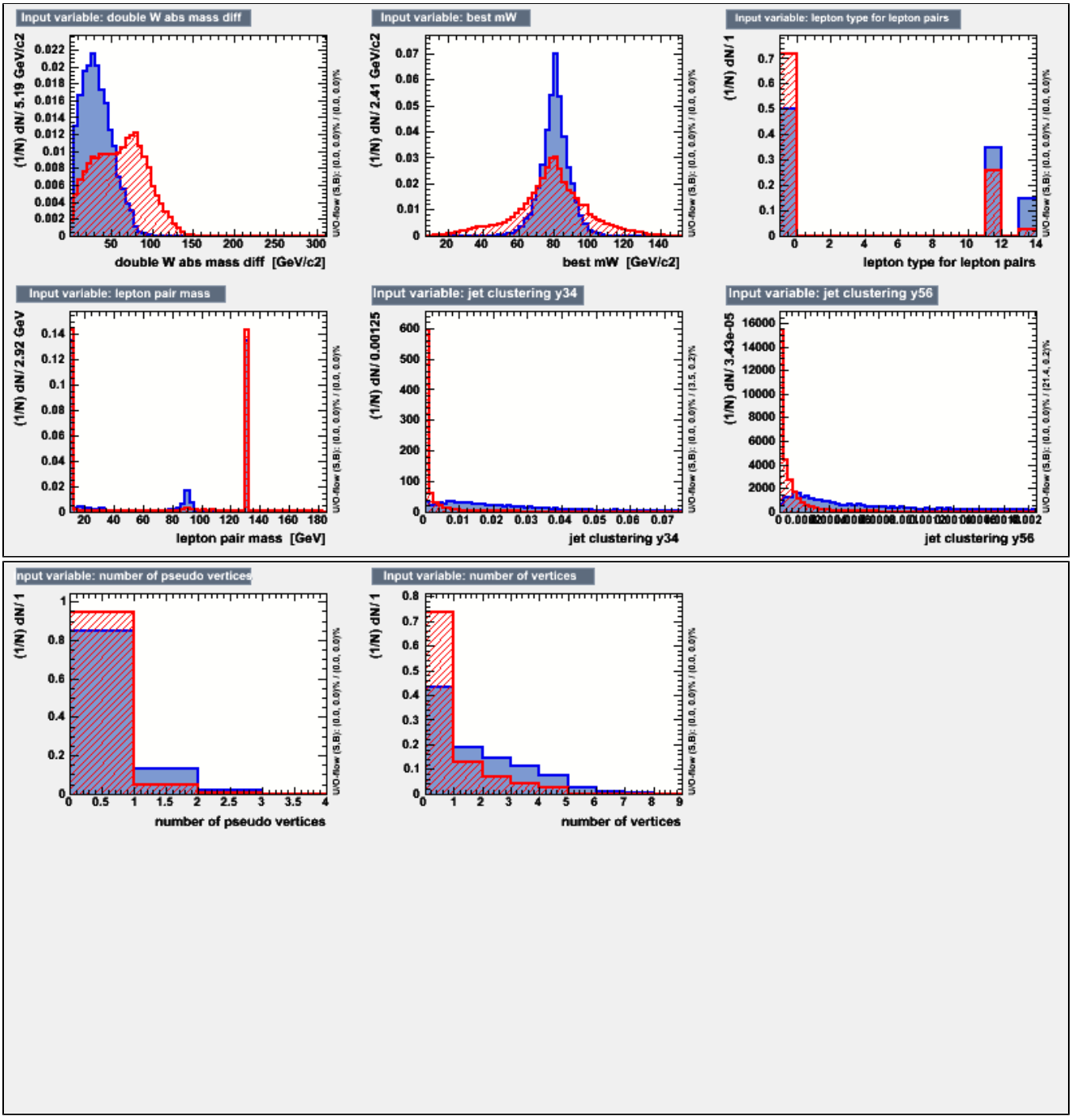
- divide into 4 jet / 6 jet topologies
  - this is ZH with H->ZZ* , Z-> nn,ll,qq
- apply preselection depending on topology
- train/apply TMVAs
- validate with cut table
- check remaining backgrounds

Distributions before preselection with only a cut on the reconstructed Higgs mass:




Input variable: Higgs Mass
(1/N) dN/ 1.18 GeV/c2
Higgs Mass [GeV/c2]
Input variable: Z energy
(1/N) dN/ 3.93 GeV
Z energy [GeV]
Input variable: Zstr energy
(1/N) dN/ 2.41 GeV
Zstr energy [GeV]
Input variable: Z cos theta
(1/N) dN/ 0.0339
Z cos theta
Input variable: Zstr cos theta
(1/N) dN/ 0.0339
Zstr cos theta
Input variable: Z mass
(1/N) dN/ 2.25 GeV/c2
Z mass [GeV/c2]
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%


Input variable: Zstr mass
(1/N) dN/ 1.31 GeV/c2
Zstr mass [GeV/c2]
Input variable: azimuthal separation of Z jets
(1/N) dN/ 0.212 radians
azimuthal separation of Z jets [radians]
Input variable: thrust
(1/N) dN/ 0.00376 1
thrust [1]
Input variable: Number Charged Tracks
(1/N) dN/ 1
Number Charged Tracks
Input variable: Number of electrons
(1/N) dN/ 1
Number of electrons
Input variable: Number of muons
(1/N) dN/ 1
Number of muons
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%

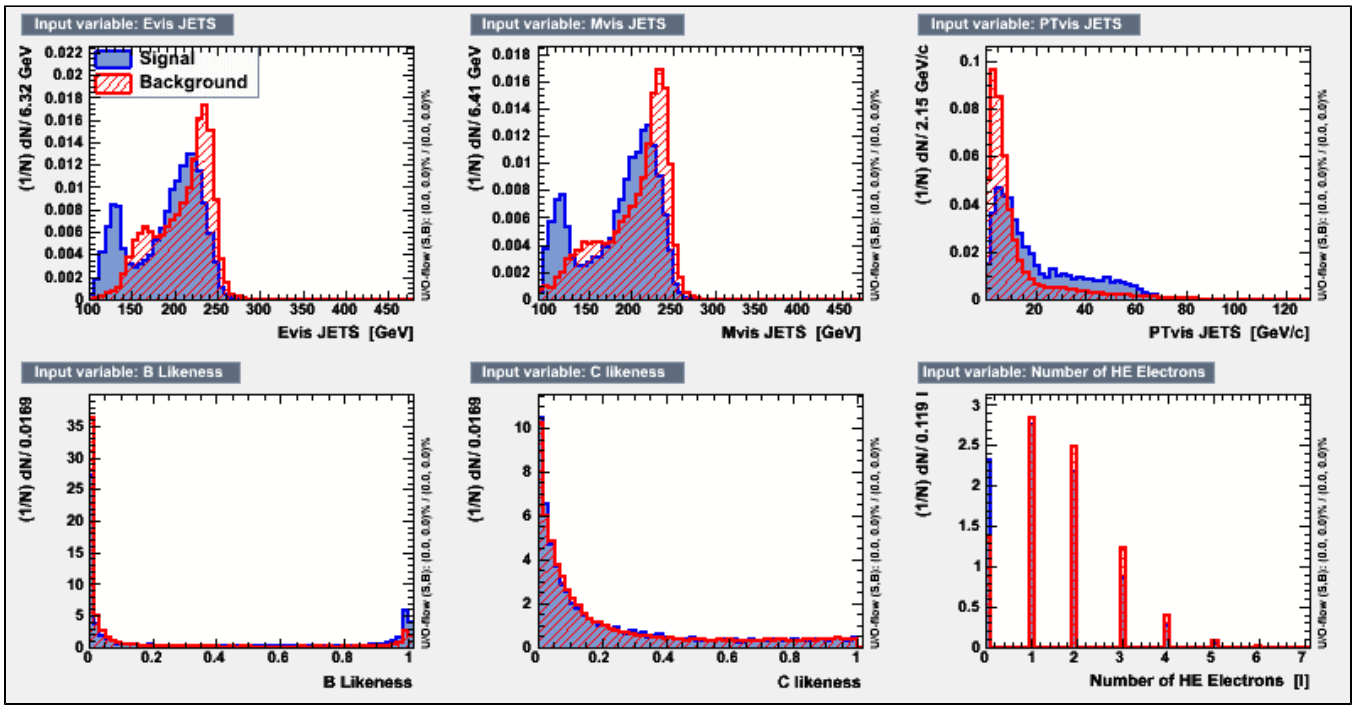


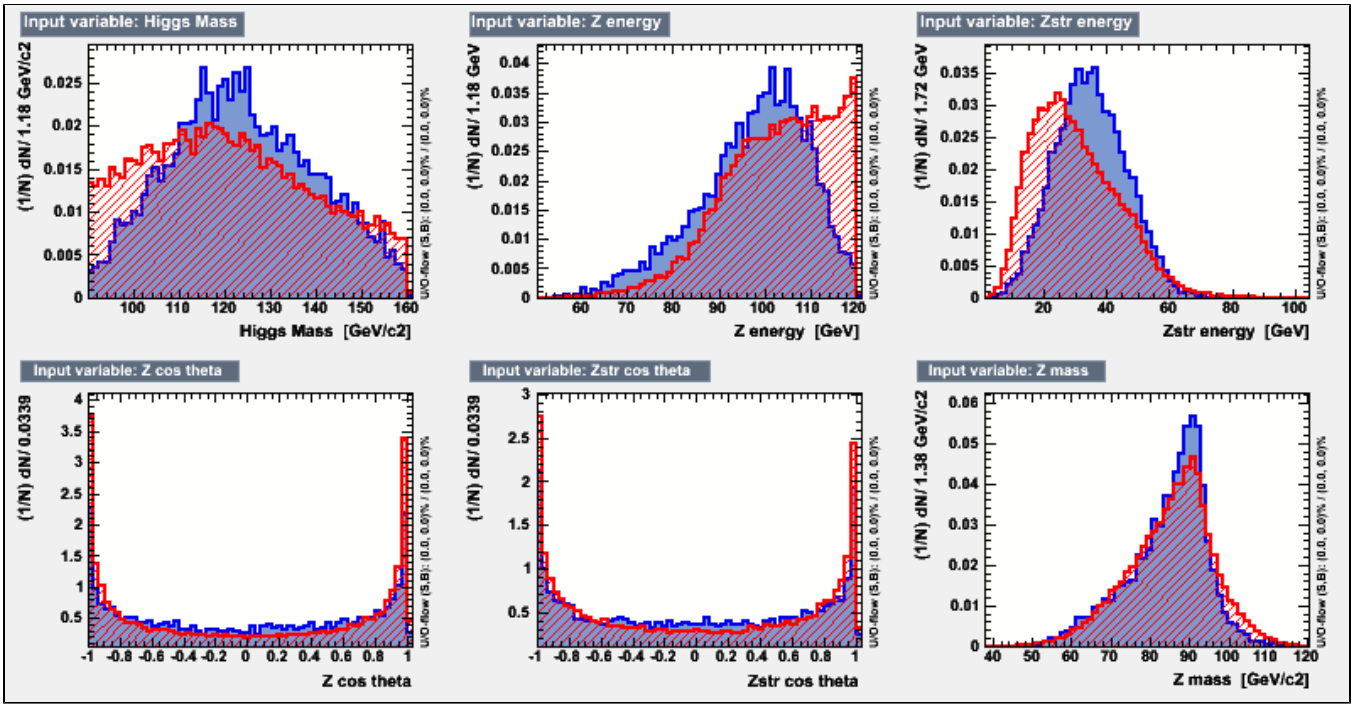
Preselection:

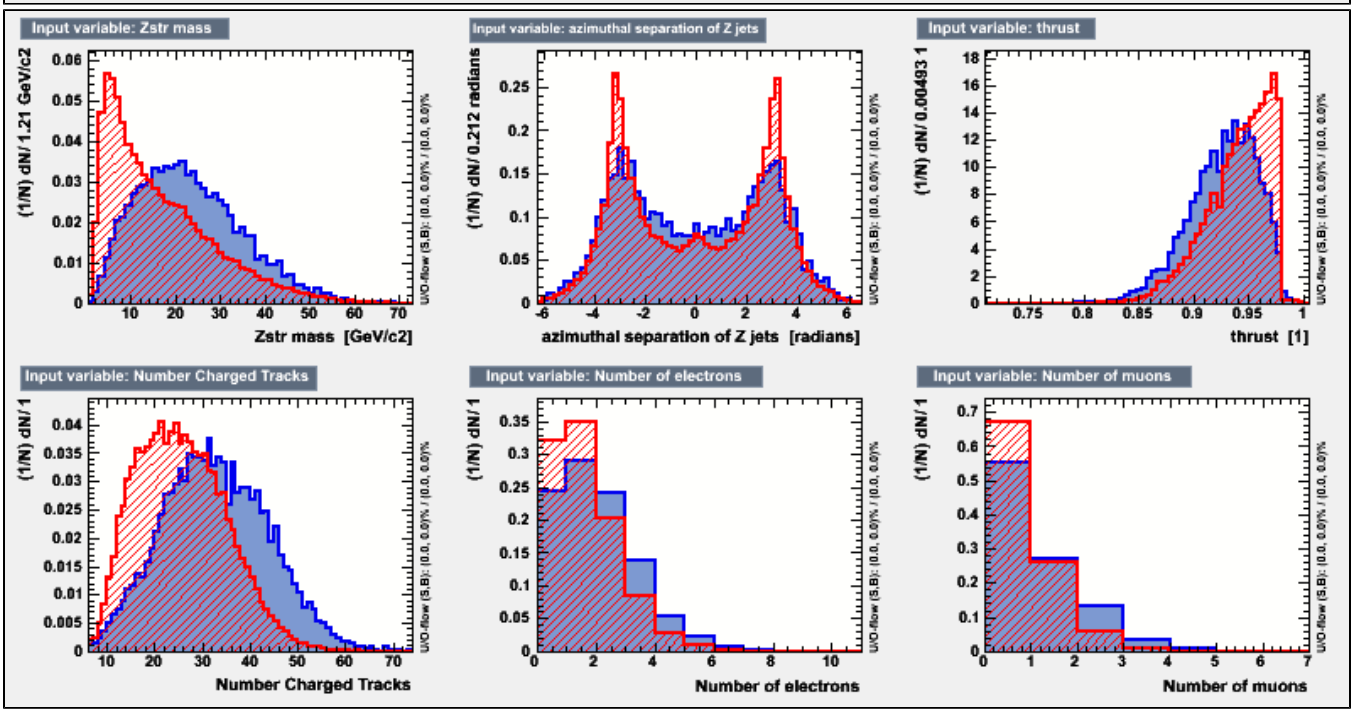
Simplified and making only clean cuts ...

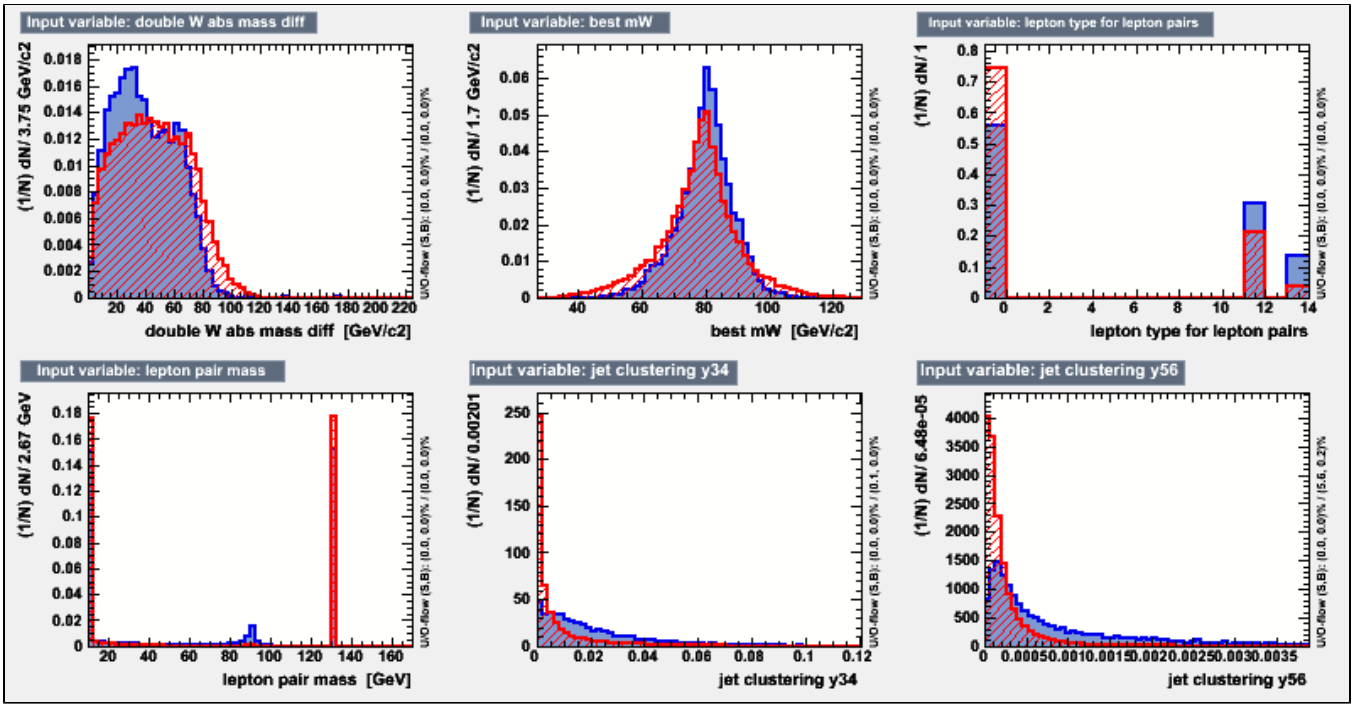
```
Evis<140.0: (4 Jet category)
#y34>0.0#hmass>95. && hmass<140.
#PTvisJETS>25.0 && PTvisJETS<70.0
#nTrks>5.
#ej1<120.

Evis>140: (6 jet category)
#y34>0.0
#hmass>90. && hmass<160.
#ej1<120.
#jetthrust<0.98
#nTrks>5
```

Distributions after preselection:

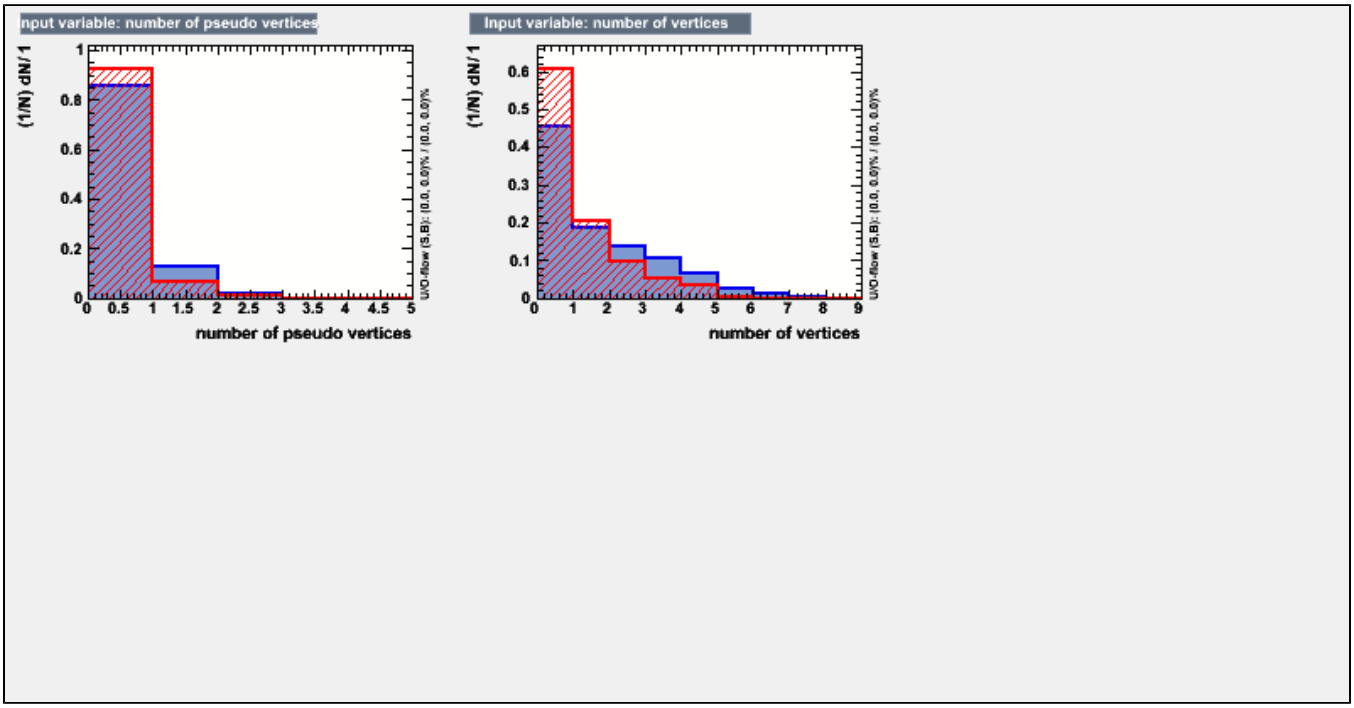



Input variable: Higgs Mass
(1/N) dN/ 1.18 GeV/c2
Higgs Mass [GeV/c2]
Input variable: Z energy
(1/N) dN/ 1.18 GeV
Z energy [GeV]
Input variable: Zstr energy
(1/N) dN/ 1.72 GeV
Zstr energy [GeV]
Input variable: Z cos theta
(1/N) dN/ 0.0339
Z cos theta
Input variable: Zstr cos theta
(1/N) dN/ 0.0339
Zstr cos theta
Input variable: Z mass
(1/N) dN/ 1.38 GeV/c2
Z mass [GeV/c2]
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%


Input variable: Zstr mass
(1/N) dN/ 1.21 GeV/c2
Zstr mass [GeV/c2]
Input variable: azimuthal separation of Z jets
(1/N) dN/ 0.212 radians
azimuthal separation of Z jets [radians]
Input variable: thrust
(1/N) dN/ 0.00493 1
thrust [1]
Input variable: Number Charged Tracks
(1/N) dN/ 1
Number Charged Tracks
Input variable: Number of electrons
(1/N) dN/ 1
Number of electrons
Input variable: Number of muons
(1/N) dN/ 1
Number of muons
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%



(Note: the lepton pair mass plot has the entries at less than 10 GeV randomly set to 10 and 130 GeV so that the signal peak can be seen.)



The TMVA variables:

1. Evis JETS
2. Mvis JETS
3. PTvis JETS
4. B Likeness
5. C likeness
6. Number of HE Electrons
7. Higgs Mass
8. Z energy
9. Zstr energy
10. Z cos theta
11. Zstr cos theta
12. Z mass

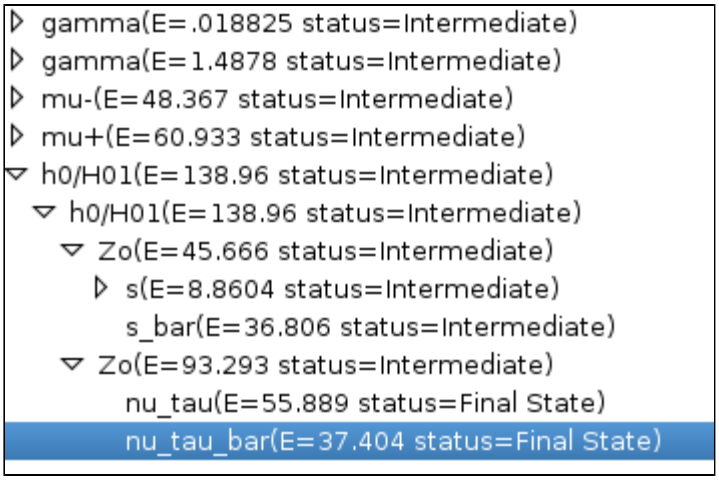
13. Zstr mass
14. azimuthal separation of Z jets
15. thrust
16. Number Charged Tracks
17. Number of electrons
18. Number of muons

----NEW VARIABLES:

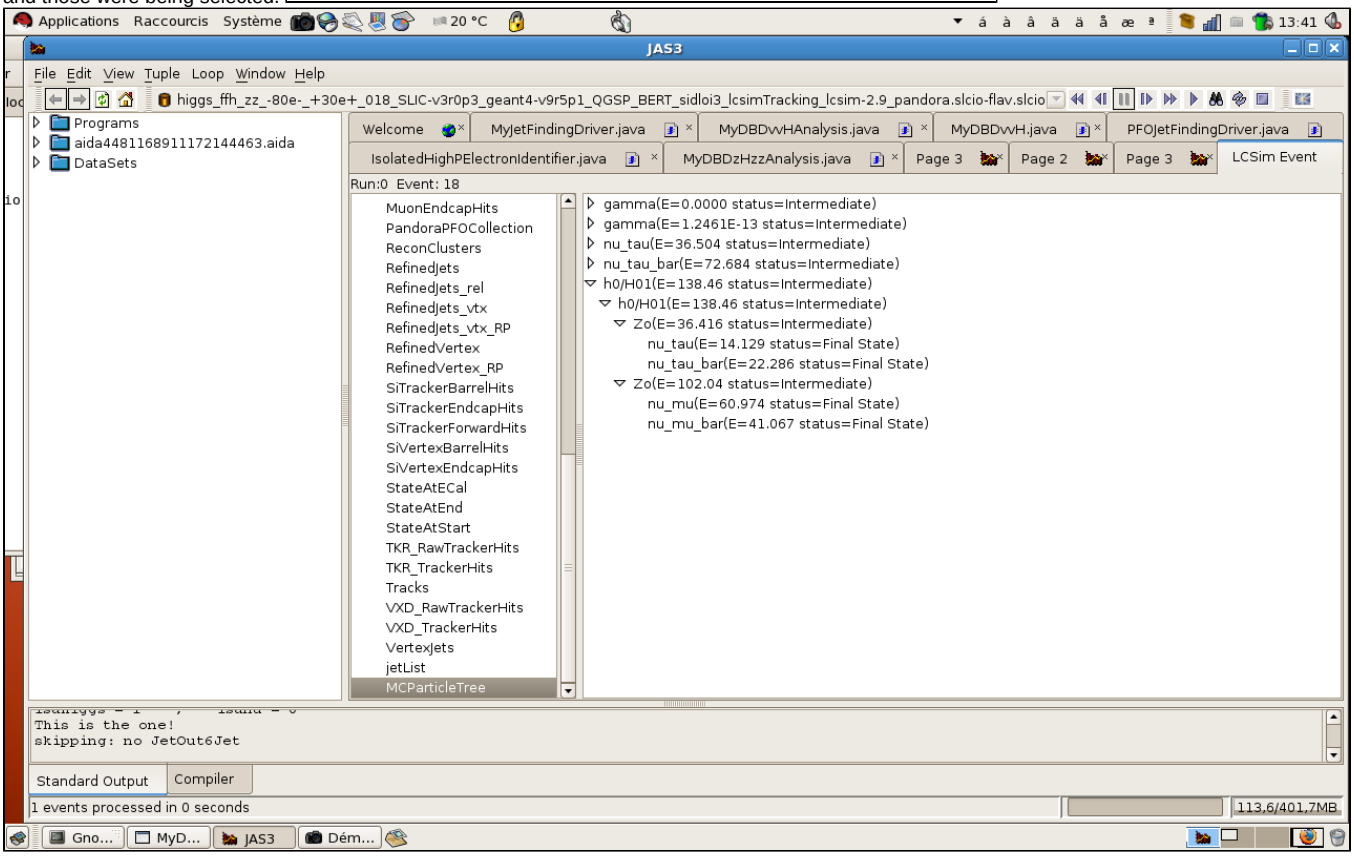
1. y34
2. y56
3. lepton pair (PDG ID1 = -ID2) mass closest to Mz
4. jet pair mass closest to mW
5. | jet pair1 - mW | + | jet pair2 - mW |
   |---|---|---|

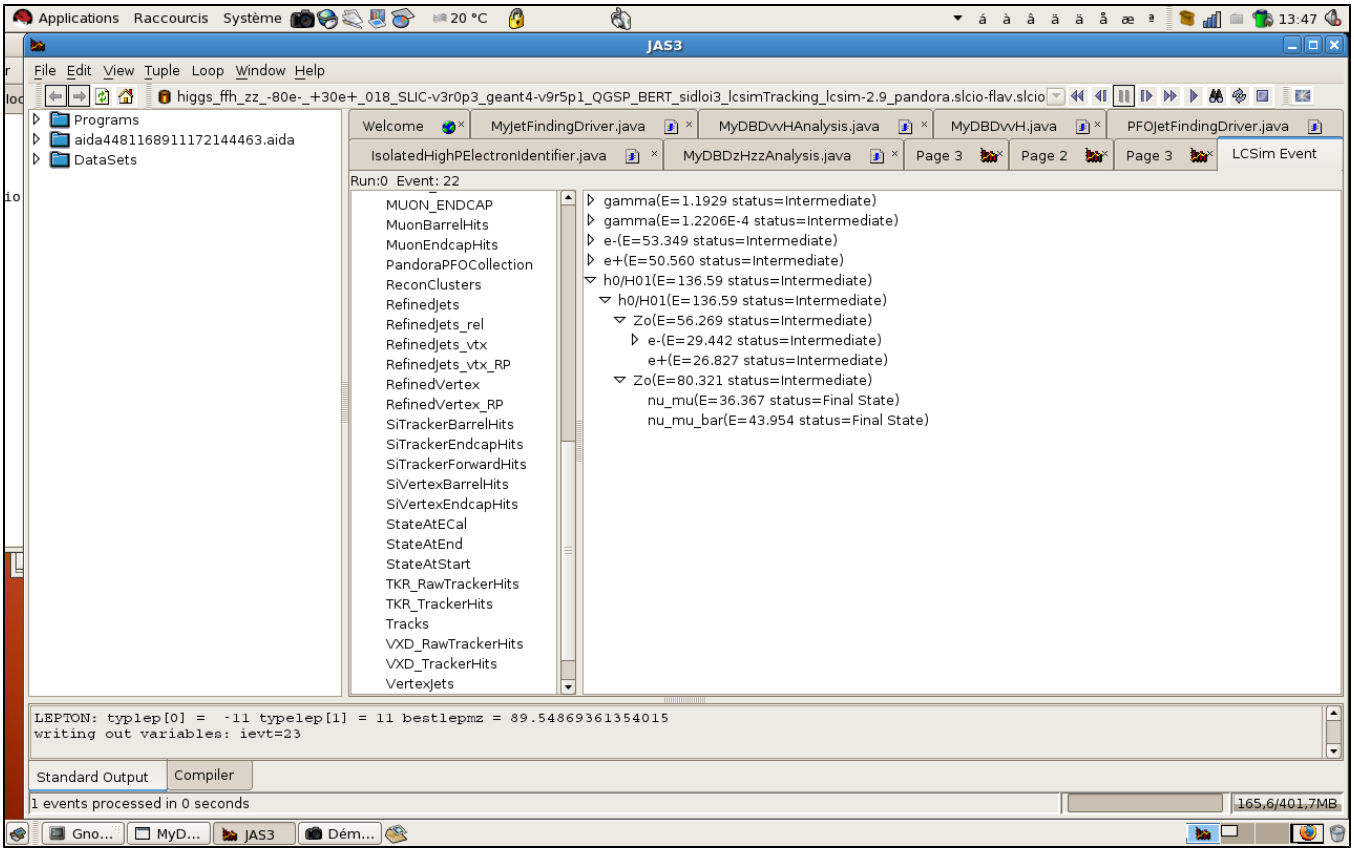
Some signals are just not reasonable to try to select:

1) decays like l+l-H where H->ffnunu
Even in cases where the ff are leptons the initial l+l- are more energetic

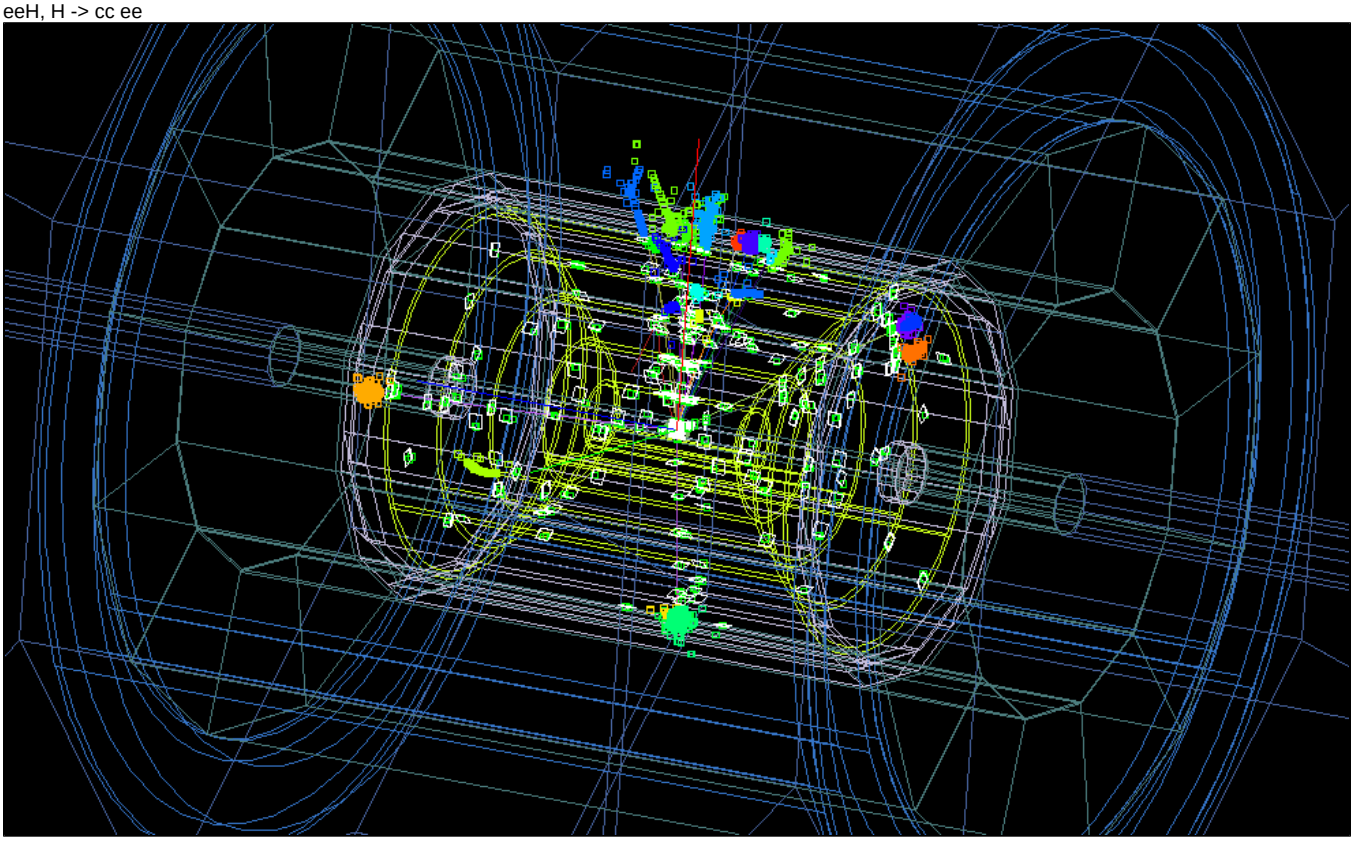


and those were being selected.

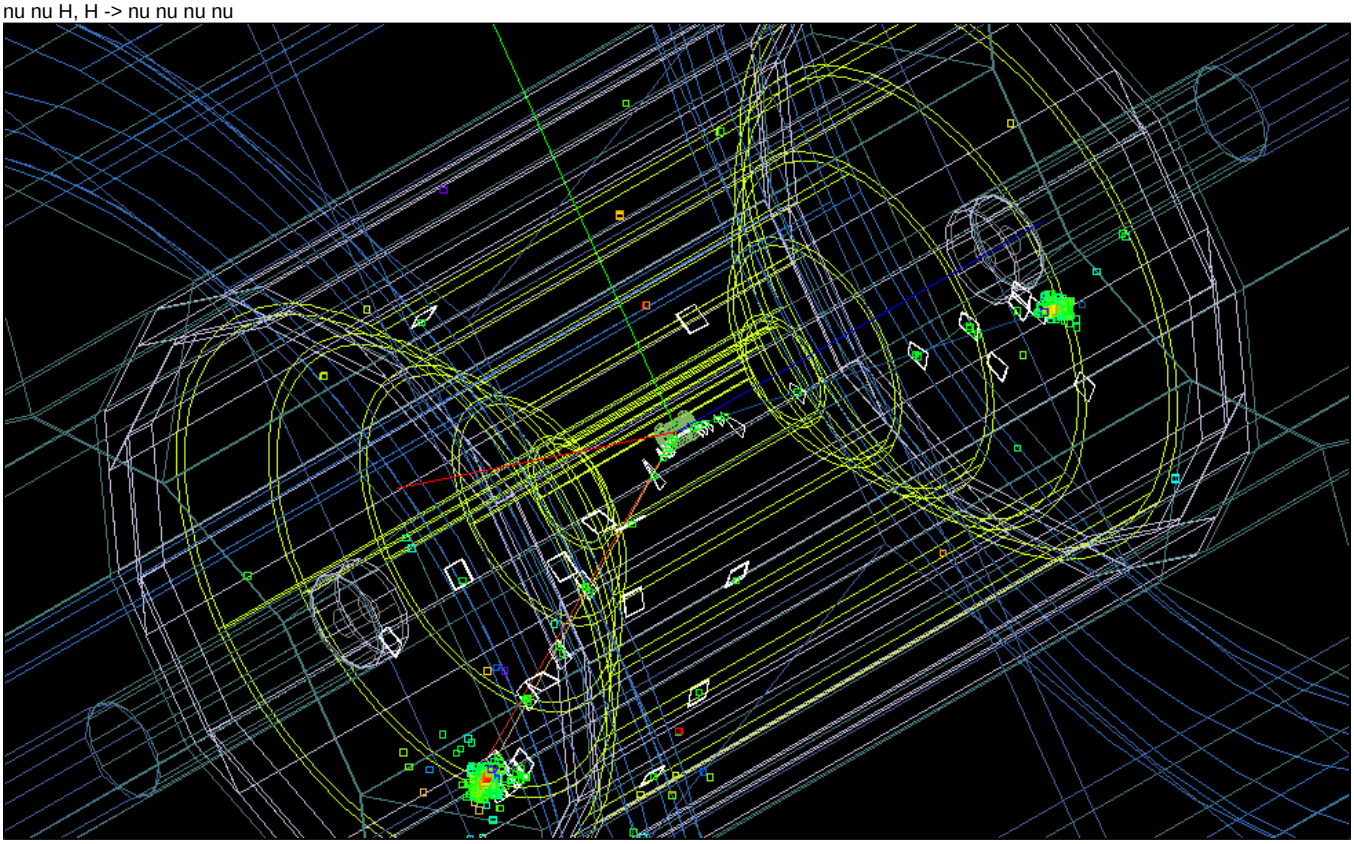



Applications Raccourcis Système
20 °C
13:47
JAS3
File Edit View Tuple Loop Window Help
Programs
aida4481168911172144463.aida
DataSets
Welcome
MyJetFindingDriver.java
MyDBDvvHAnalysis.java
MyDBDvvH.java
PFOJetFindingDriver.java
IsolatedHighPElectronIdentifier.java
MyDBDzHzzAnalysis.java
Page 3
Page 2
Page 3
LCSim Event
Run:0 Event: 22
MUON_ENDCAP
MuonBarrelHits
MuonEndcapHits
PandoraPFOCollection
ReconClusters
RefinedJets
RefinedJets_rel
RefinedJets_vtx
RefinedJets_vtx_RP
RefinedVertex
RefinedVertex_RP
SiTrackerBarrelHits
SiTrackerEndcapHits
SiTrackerForwardHits
SiVertexBarrelHits
SiVertexEndcapHits
StateAtECal
StateAtEnd
StateAtStart
TKR_RawTrackerHits
TKR_TrackerHits
Tracks
VXD_RawTrackerHits
VXD_TrackerHits
VertexJets
gamma(E=1.1929 status=Intermediate)
gamma(E=1.2206E-4 status=Intermediate)
e-(E=53.349 status=Intermediate)
e+(E=50.560 status=Intermediate)
h0/H01(E=136.59 status=Intermediate)
h0/H01(E=136.59 status=Intermediate)
Zo(E=56.269 status=Intermediate)
e-(E=29.442 status=Intermediate)
e+(E=26.827 status=Intermediate)
Zo(E=80.321 status=Intermediate)
nu_mu(E=36.367 status=Final State)
nu_mu_bar(E=43.954 status=Final State)
LEPTON: typlep[0] = -11 typelep[1] = 11 bestlepmz = 89.54869361354015
writing out variables: ievt=23
Standard Output
Compiler
1 events processed in 0 seconds
165,6/401,7MB
Gno...
MyD...
JAS3
Dém...

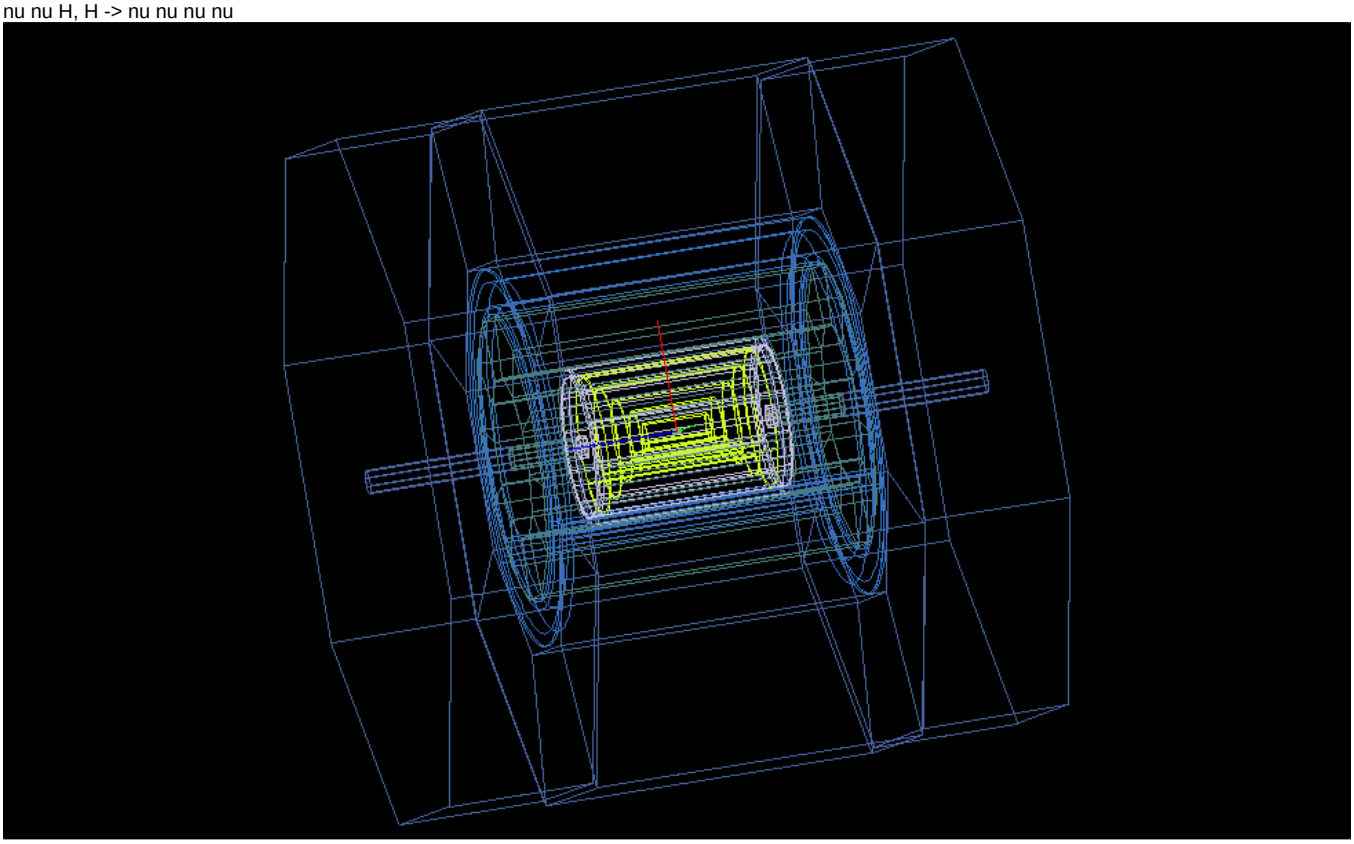
eeH, H -> cc ee



nu nu H, H -> nu nu nu nu



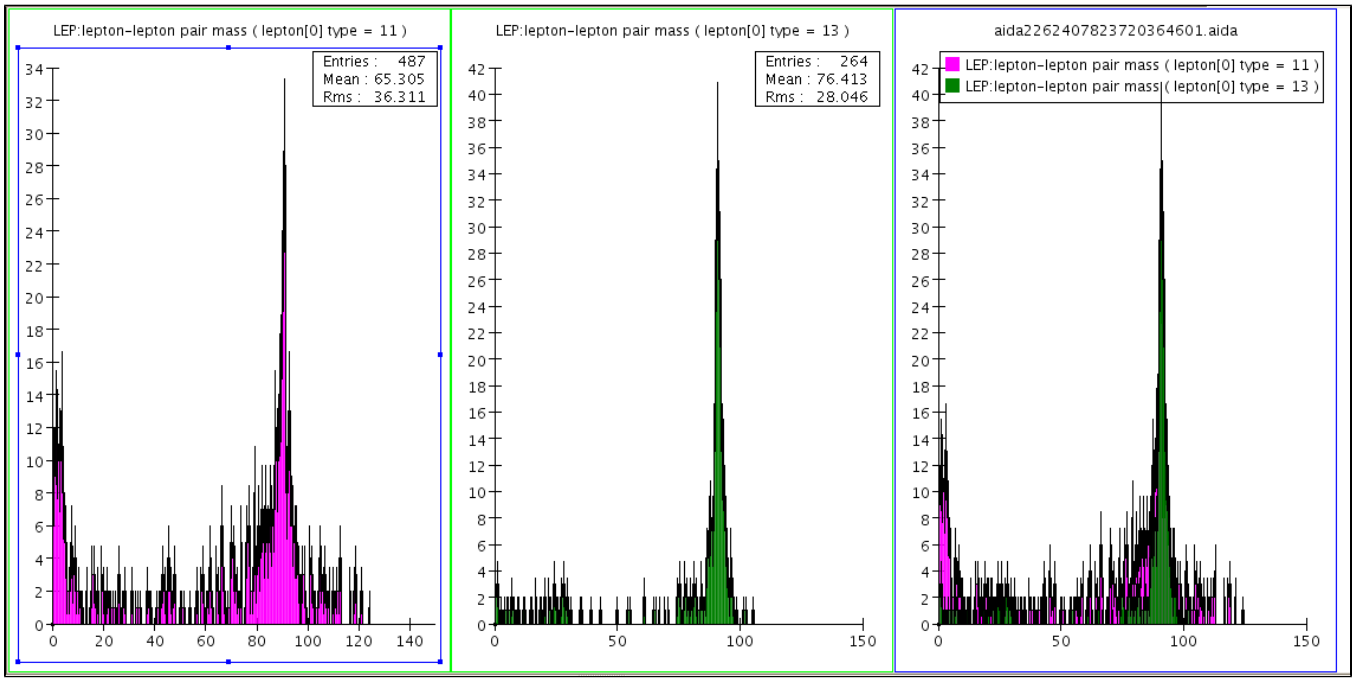
nu nu H, H -> nu nu nu nu



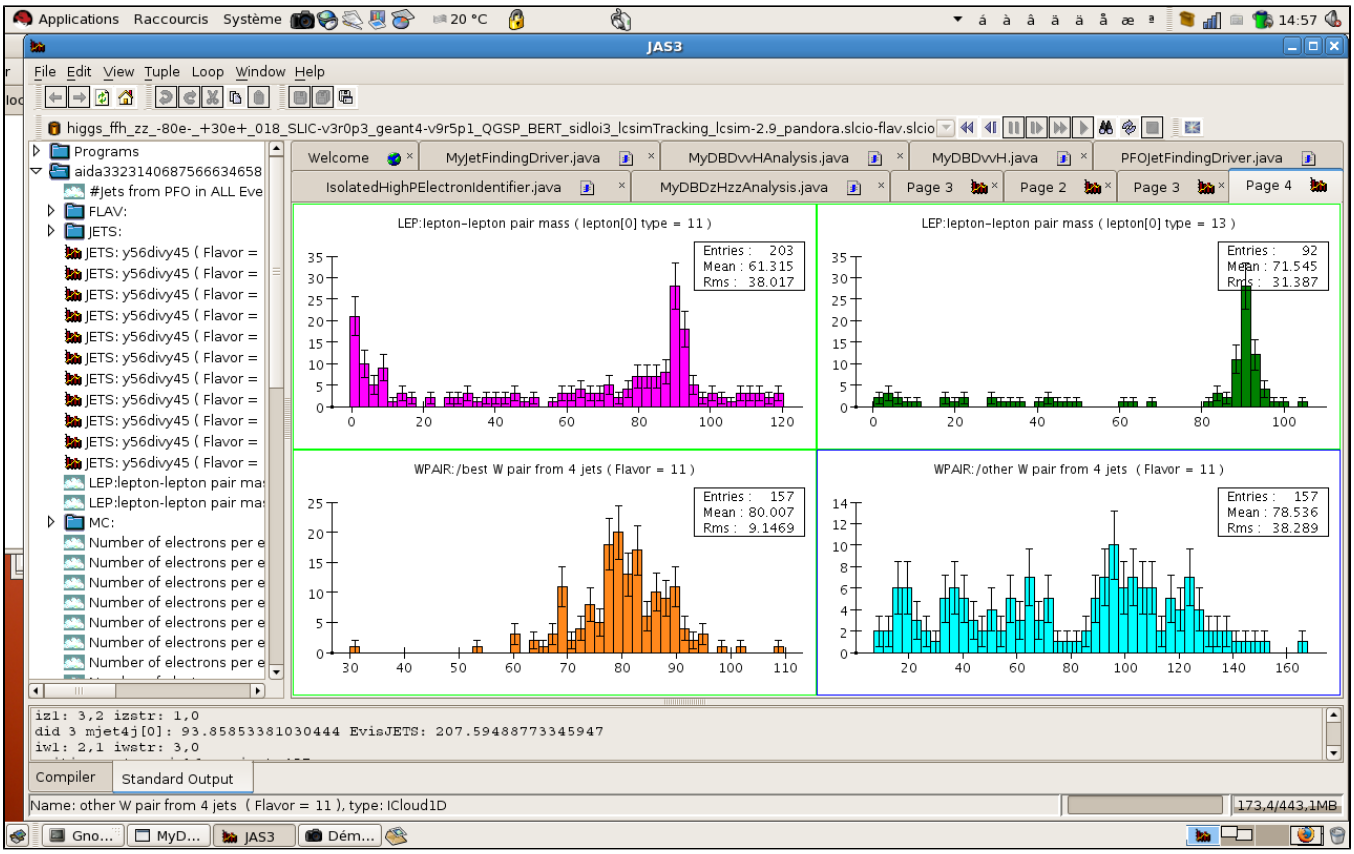
		HZZ generated sample stats		

		events	weight	
I108056.001.stdhep	10000	20000	1.29E-004	e1e1h_zz.Gwhizard-
I108055.001.stdhep	10000	20000	4.36E-003	e1e1h_zz.Gwhizard-
I108054.001.stdhep	10000	20000	4.11E-004	e1e1h_zz.Gwhizard-
I108053.001.stdhep	10000	20000	2.70E-005	e1e1h_zz.Gwhizard-
I108058.001.stdhep	10000	20000	4.29E-003	e2e2h_zz.Gwhizard-
I108057.001.stdhep	10000	20000	4.00E-004	e2e2h_zz.Gwhizard-
I108060.001.stdhep	10000	20000	4.28E-003	e3e3h_zz.Gwhizard-
I108059.001.stdhep	10000	20000	4.00E-004	e3e3h_zz.Gwhizard-
		Sum =	1.43E-002	10.40%
I108062.001.stdhep	10000	20000	2.54E-002	nnh_zz.Gwhizard-1_
I108061.001.stdhep	10000	20000	3.01E-003	nnh_zz.Gwhizard-1_
		Sum =	2.84E-002	20.68%
I108064.001.stdhep	10000	20000	8.67E-002	qqh_zz.Gwhizard-1_
I108063.001.stdhep	10000	20000	8.08E-003	qqh_zz.Gwhizard-1_
		Sum =	9.47E-002	68.93%
		All =	1.37E-001	

Mass of lepton pair with mass closest to mZ:



The following also shows the jet pair mass for those that have mass closest to mW:



Note: Compared to last week plots and tables there was also a simple programming error which I've already fixed.

Performance of different MVA options:

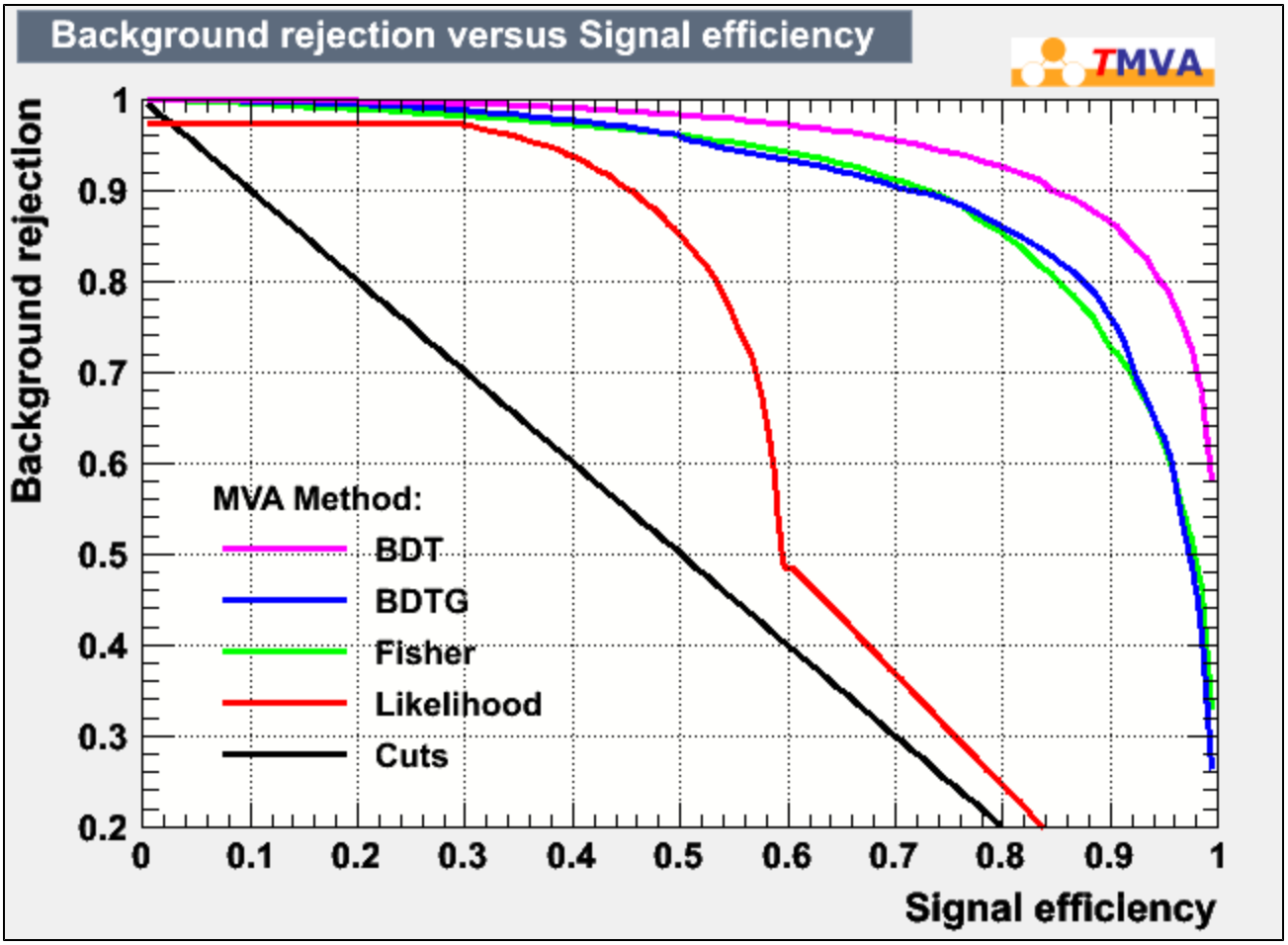
BEFORE

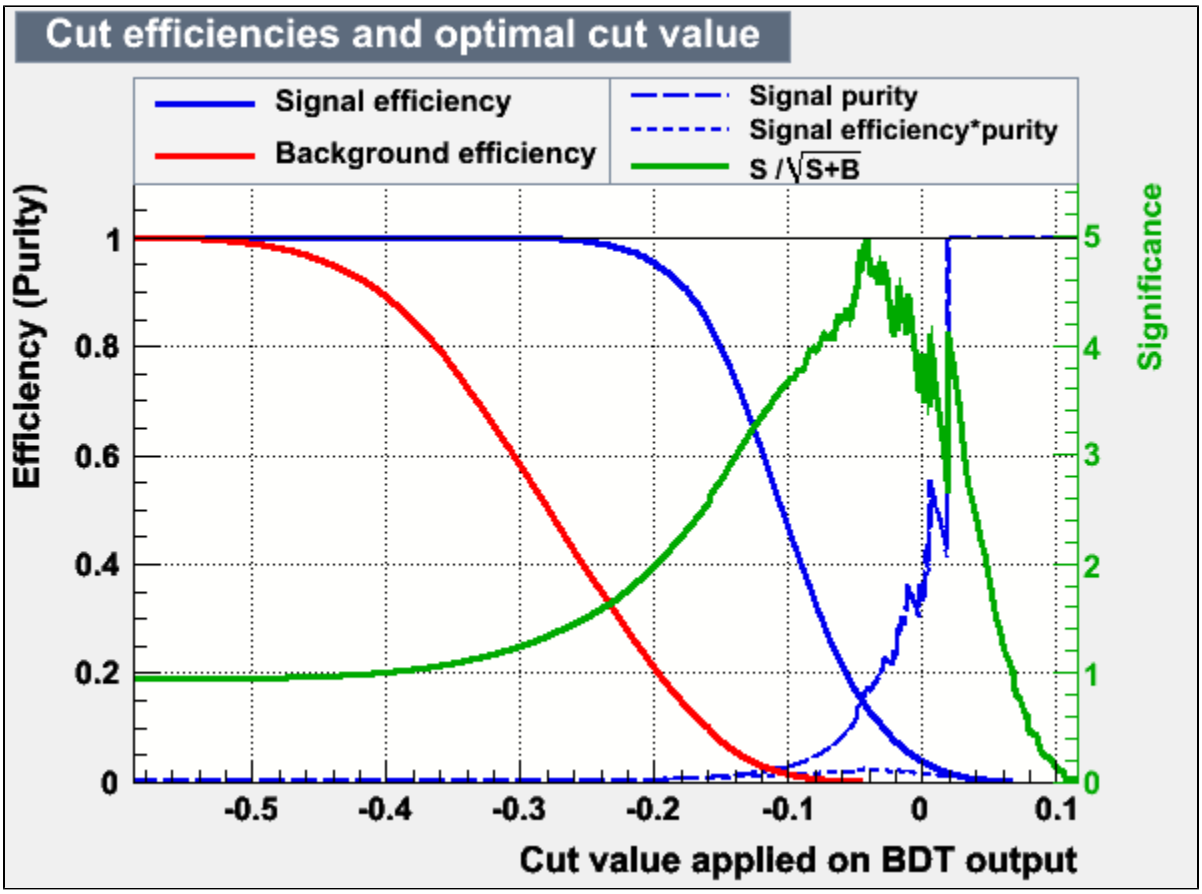
```
[neal@localhost weights]$ grep -A 8 -i optimal- ../hzzv14-presel-v4-new-common0410results.txt
--- Classifier   (  #signal, #backgr.)  Optimal-cut  S/sqrt(S+B)      NSig      NBkg   EffSig   EffBkg
-- -------------------------------------------------------------------------------------------------
---       Cuts:  ( 664.0744,1312202.5)      -0.0050            0         0         0        0        0
--- Likelihood:  ( 664.0744,1312202.5)       1.0000     0.728991  644.0848  779980.1   0.9699   0.5944
---     Fisher:  ( 664.0744,1312202.5)       0.0120      1.70335  258.2099  22721.13   0.3888  0.01732
---       BDTG:  ( 664.0744,1312202.5)      -0.9823      1.17754  538.5353  208621.3    0.811    0.159
---        BDT:  ( 664.0744,1312202.5)      -0.0462      3.42709  100.6751  762.2891   0.1516 0.0005809
-- -------------------------------------------------------------------------------------------------
```

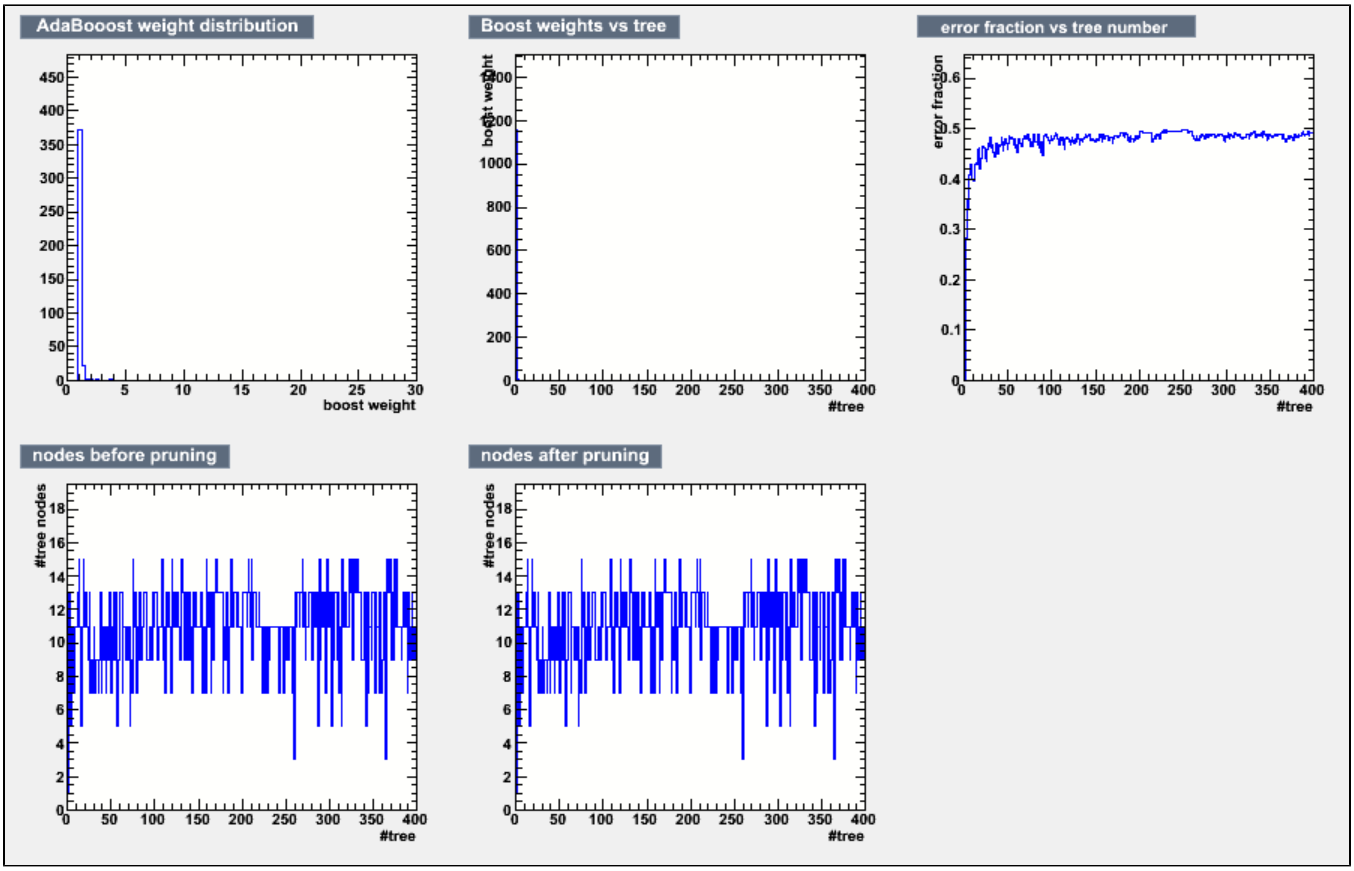
NOW:

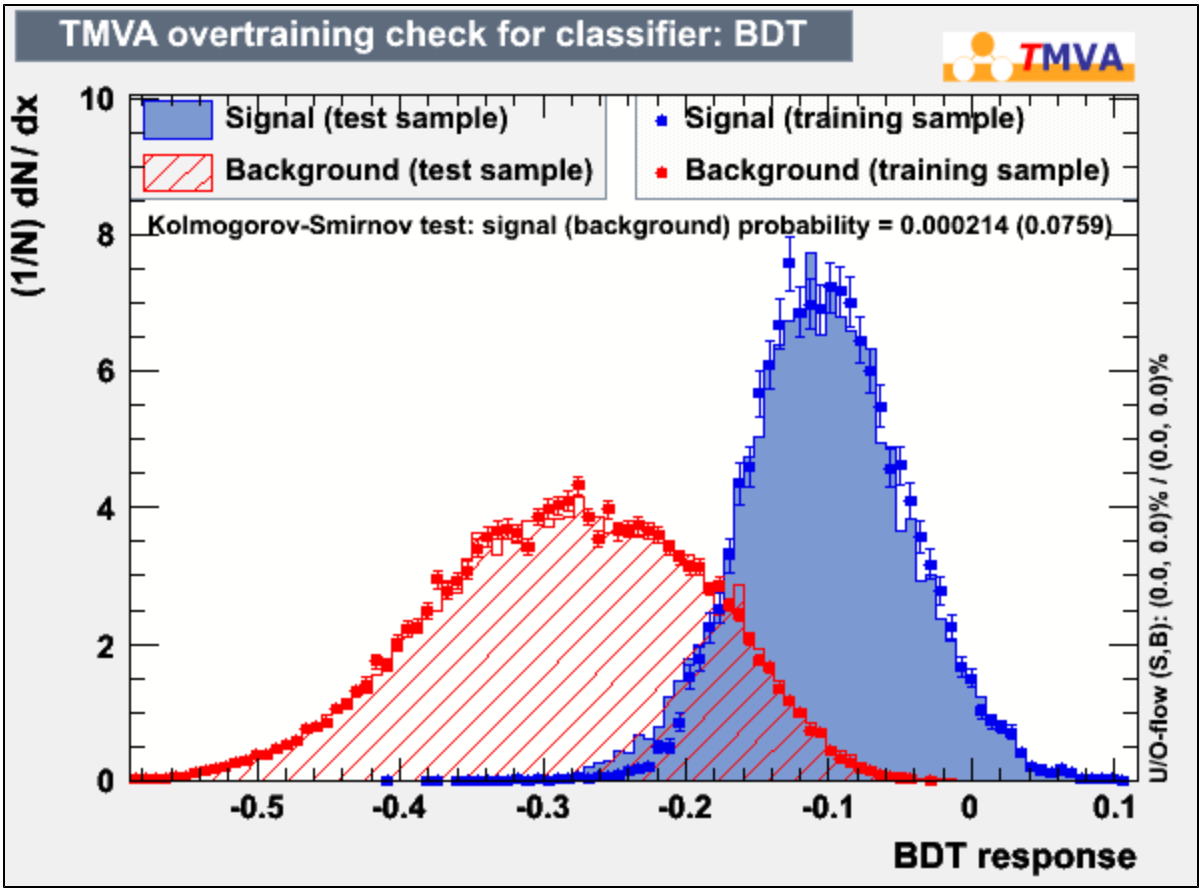
```
    if ((fp = fopen("HZZ-TMVA-vars-ffh_zz_all_SM_background_+80e-_-30e+-v6-mini.txt","r"))==NULL) exit(0);
[neal@localhost test]$ source storeresultsHZZv4.sh hzz-withlep-withmW-n2-Mhiggspresel-plusptntrkp8g16y34nvtx


--- ============================================================================================================
--- Classifier   (  #signal, #backgr.)  Optimal-cut  S/sqrt(S+B)      NSig      NBkg   EffSig   EffBkg
--- ------------------------------------------------------------------------------------------------------------
---       Cuts:  (1035.8644,  1196975)      -0.0050           0         0         0        0         0
--- Likelihood:  (1035.8644,  1196975)       1.0000      1.66886  315.8824  35510.98   0.3049  0.02967
---     Fisher:  (1035.8644,  1196975)       0.0047      2.38122  586.8862  60157.76   0.5666  0.05026
---       BDTG:  (1035.8644,  1196975)      -0.9795        2.625  293.8675  12238.81   0.2837  0.01022
---        BDT:  (1035.8644,  1196975)      -0.0427       4.9889  146.0711  711.2005    0.141 0.0005942
--- ------------------------------------------------------------------------------------------------------------
                                                    ^^^^^^
```

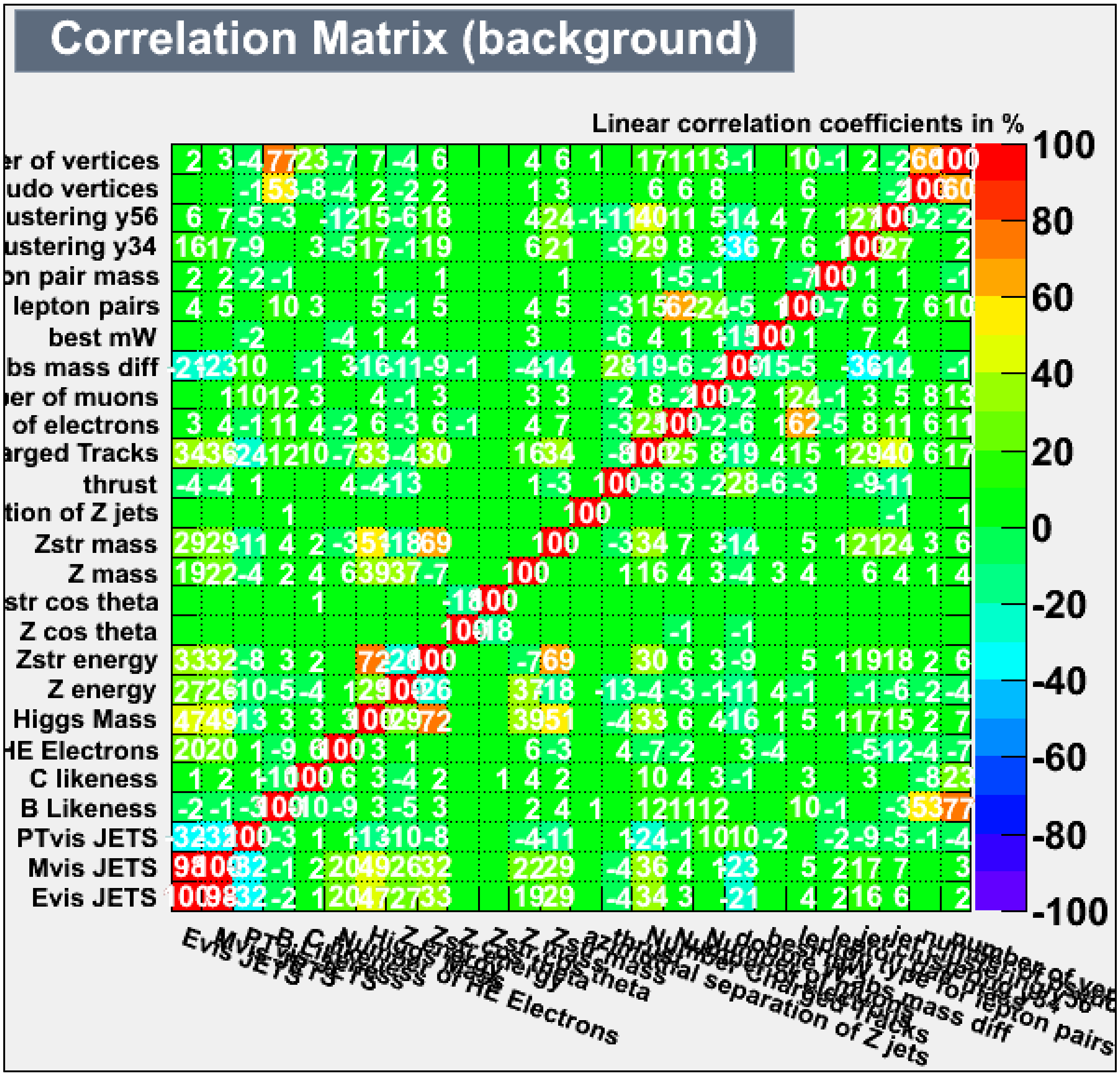
BDT plots:



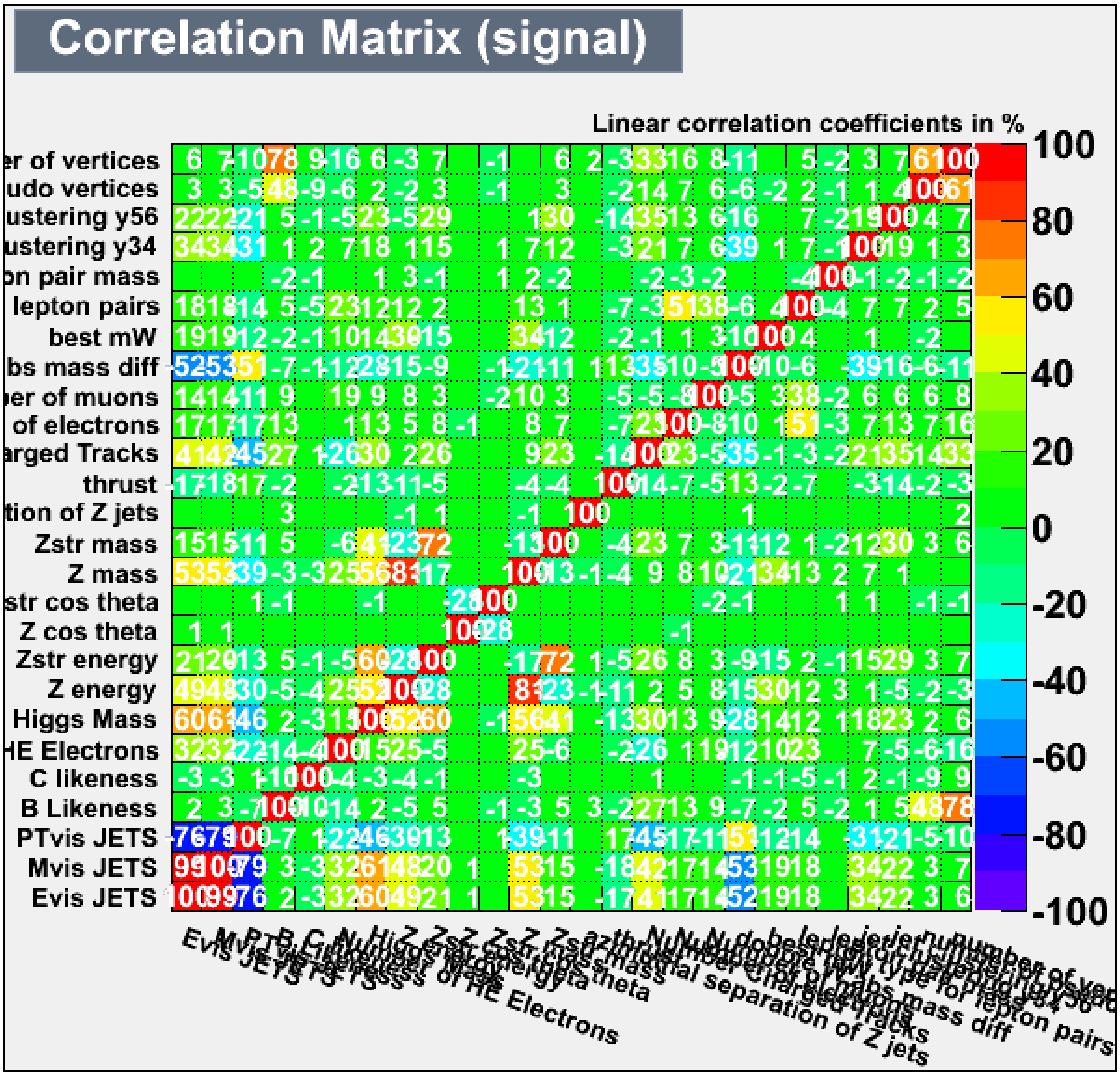

Cut efficiencies and optimal cut value
Signal efficiency
Background efficiency
Signal purity
Signal efficiency*purity
S /√S+B
Efficiency (Purity)
Significance
Cut value applied on BDT output


AdaBooost weight distribution
boost weight
Boost weights vs tree
boost weight
#tree
error fraction vs tree number
error fraction
#tree
nodes before pruning
#tree nodes
#tree
nodes after pruning
#tree nodes
#tree


TMVA overtraining check for classifier: BDT
TMVA
Signal (test sample)
Background (test sample)
Signal (training sample)
Background (training sample)
Kolmogorov-Smirnov test: signal (background) probability = 0.000214 (0.0759)
(1/N) dN / dx
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%
BDT response
10
8
6
4
2
0
-0.5
-0.4
-0.3
-0.2
-0.1
0
0.1



Correlation Matrix (background)
Linear correlation coefficients in %
er of vertices
udo vertices
ustering y56
ustering y34
on pair mass
lepton pairs
best mW
bs mass diff
ber of muons
of electrons
arged Tracks
thrust
tion of Z jets
Zstr mass
Z mass
str cos theta
Z cos theta
Zstr energy
Z energy
Higgs Mass
HE Electrons
C likeness
B Likeness
PTvis JETS
Mvis JETS
Evis JETS
100
80
60
40
20
0
-20
-40
-60
-80
-100



Cut table for BDT: (NEW)NEW:

```
 cuts__ (Cut Name                          ):       all   others
2f       4f       6f       aa       1f       3f       5f        q      e+mu      tau       nu  Signif.
 cut #0 (all                               ): 7.769e+07   625000  14507062  1019938        0
54073784        0  7464875        0       948       91       45      262      0.15  Delta(sig*BR) = 6.548594
+/-  0.046879
 cut #1 (y34>0.y34>0.                      ): 7.705e+07   625000  14449800  1018312        0
53568784        0  7387500        0       947       91       45      262      0.15  Delta(sig*BR) = 6.523157
+/-  0.046724
 cut #2 (95.<hmass<140. 95.<hmass<140.     ): 1.257e+07    150000
10264438   825238        0   387992        0   938000
0      901       86       43      213      0.35  Delta(sig*BR) =  2.854636 +/- 0.021075
 cut #3 (25.<PTvisJETS<70 PTvisJETS no cut ): 5.637e+06    25000
4545812   744212        0     8750        0   312125        0
866       84       37      180      0.49  Delta(sig*BR) = 2.033046 +/-  0.015436
 cut #4 (nTrks>5 nTrks>5                   ): 3.826e+06        0
2957175   679838        0     8750        0   179500        0
866       81       36      176      0.59  Delta(sig*BR) = 1.688536 +/-  0.012688
 cut #5 (jetthrust no cut jetthrust < 0.98 ): 3.826e+06        0
2957175   679838        0     8750        0   179500        0
866       81       36      176      0.59  Delta(sig*BR) = 1.688536 +/-  0.012688
 cut #6 (ej1<120. ej1<120.                 ): 2.144e+06        0
1480162   539312        0     8750        0   114500        0
844       78       35      176      0.77  Delta(sig*BR) = 1.292336 +/-  0.009952
 cut #7 (MVA                               ): 9.252e+02        0
50      712        0        0        0        0        0       79        27        3       53      5.35  Delta
(sig*BR) = 0.186991 +/- 0.010382
```

Remaining backgrounds: (NEW)

```
$ sed 's/\./,/g' zzhpassing.dat | awk '{if (NF==18) print $3,$7,$15}' | sort -n | uniq -c | awk '{if (NF==4)
aa+=($1*$4);print $1*$4"\t"$0} END{print "Sum of weights = "aa}' | sort -k 1,1 -n
Sum of weights = 762,5
12,5          1 106561 4f_sz 12,500000
12,5          1 106564 4f_sw 12,500000
12,5          1 106573 4f_zz 12,500000
12,5          1 106580 4f_zz 12,500000
12,5          1 106607 2f_z_ 12,500000
25            2 106551 4f_ww 12,500000
25            2 106562 4f_sz 12,500000
37,5          3 106575 4f_zz 12,500000
37,5          3 106608 2f_z_ 12,500000
50            4 106578 4f_ww 12,500000
62,5          5 106572 4f_sz 12,500000
100           8 106574 4f_zz 12,500000
125          10 106577 4f_ww 12,500000
237,5        19 106576 4f_zz 12,500000  <----
```

Plans:

- just noticed that the preselection nTrks cut is on the PFO count and not the #charged tracks distribution shown so I will switch to using the charged tracks count in the preselection
- may still be able to do better with the leptonic Z decay selection
    - high evis events with well reconstructed leptonic Z decays and a clear 6 jet topology (using y56 cut)
- write report