Investigation of Most Sensitive Variables for HWW

Richard St. Denis, Stan Thompson University of Glasgow/CDF

4 February 2008

Outline

- Procedure
- Quality definitions (Ranking): Pass 0 best separation
- Pass 1: Pair Results
- Pass 2: Triplet Results
- Summary and Conclusions

Procedure

Goals:

- Develop understanding of the interplay of the input variables
- Develop understanding of the interaction of ME and non-ME
- Obtain the best combination of variables.
- Understand the Ranking estimators.

Procedure:

- Separately study LR, NonLR and combined, starting with Non-LR.
- Compute "Separation" of NonLR NN input variables and take largest.

- Do full analysis through to fitting for all Non-LR pairs. Pick the pair with the best sensitivity, $\frac{\sigma}{\sigma_{SM}}$. Examine Training, Sensitivity, (observed limit), NN distribution for all masses.
- Repeat with triplets, quadruplets, etc.
- Repeat with LR alone, then with mix.

Quality definitions

- Separation:

$$\langle S^2 \rangle = \frac{1}{2} \int \frac{\left(\hat{y}_s(y) - \hat{y}_b(y) \right)^2}{\left(\hat{y}_s(y) + \hat{y}_b(y) \right)} dy,$$

where $\hat{y}_s$ and $\hat{y}_b$ are the signal and background PDFs of y, respectively. This is used in Pass 0 to select the most powerful separated variable, R_{ll}. This has nothing to do with a neural net.

- Importance:

$$I_i = \bar{x}_i^2 \sum_{j=1}^{n_h} \left(w_{ij}^{(1)} \right)^2, i = 1, \ldots, n_{var},$$

where $\bar{x}_i$ is the sample mean of input variable i. This is defined in TMVA.

- Nσ: In Neurobayes, distributions are transformed to be equally populated on the interval from 0 to 1 and then this is transformed to a Gaussian. The separation of signal and background after this can then be computed in Gaussian σ.

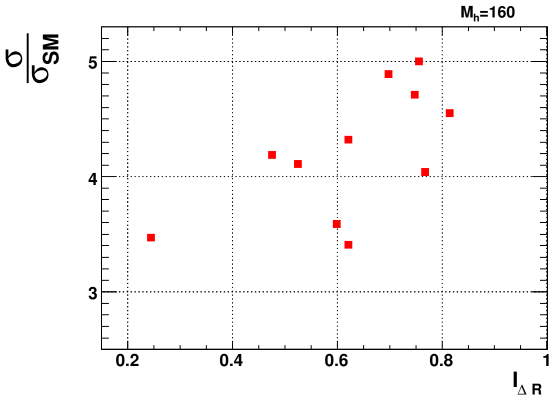
Rankings: Separation

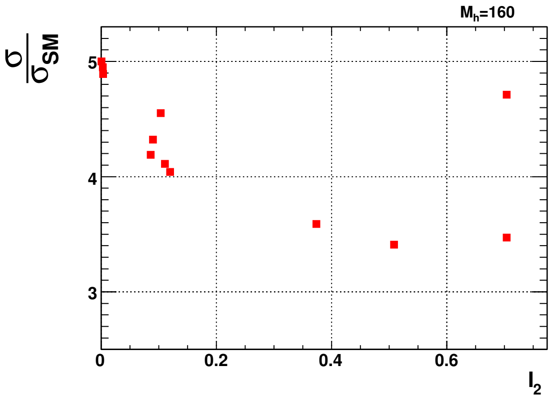
1	: LRHWW	: 3.531e-01		16	: LRWj	: 4.280e-02
2	: dRLeptons	: 3.176e-01		17	: LRWg	: 2.387e-02
3	: dPhiLeptons	: 2.519e-01		18	: jet1_Et	: 1.106e-02
4	: dimass	: 2.489e-01		19	: Njets	: 6.200e-04
5	: MetSpec	: 1.950e-01				
6	: Met	: 1.934e-01				
7	: addEt	: 1.582e-01				
8	: MetSig	: 1.359e-01				
9	: lep1_E	: 1.290e-01				
10	: MetDelPhi	: 1.279e-01				
11	: LRZZ	: 1.136e-01				
12	: Ht	: 9.369e-02				
13	: SumEt	: 8.591e-02				
14	: lep2_E	: 5.994e-02				
15	: LRWW	: 5.760e-02				

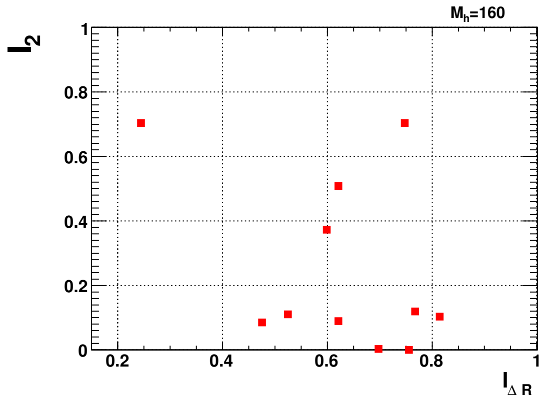
Sensitivity of Pairs at $M_h = 160$ GeV/c^2

$I_{\Delta R}$	Variable	I_2	$(\frac{\sigma}{\sigma_{SM}})$
0.6212	Ht	0.5080	3.4
0.2448	addEt	0.7036	3.5
0.5993	sumEt	0.3734	3.6
0.7674	lepE2	0.1196	4.0
0.5250	Met	0.1104	4.1
0.4754	MetSpec	0.0857	4.2
0.6213	dimass	0.0896	4.3
0.8142	lepE1	0.1032	4.6
0.7479	Njet	0.7036	4.7
0.6981	dphilepton	0.0032	4.9
1.1900	MetDelPhi	0.0025	5.0
0.7560	jet1_Et	0.0002	5.0

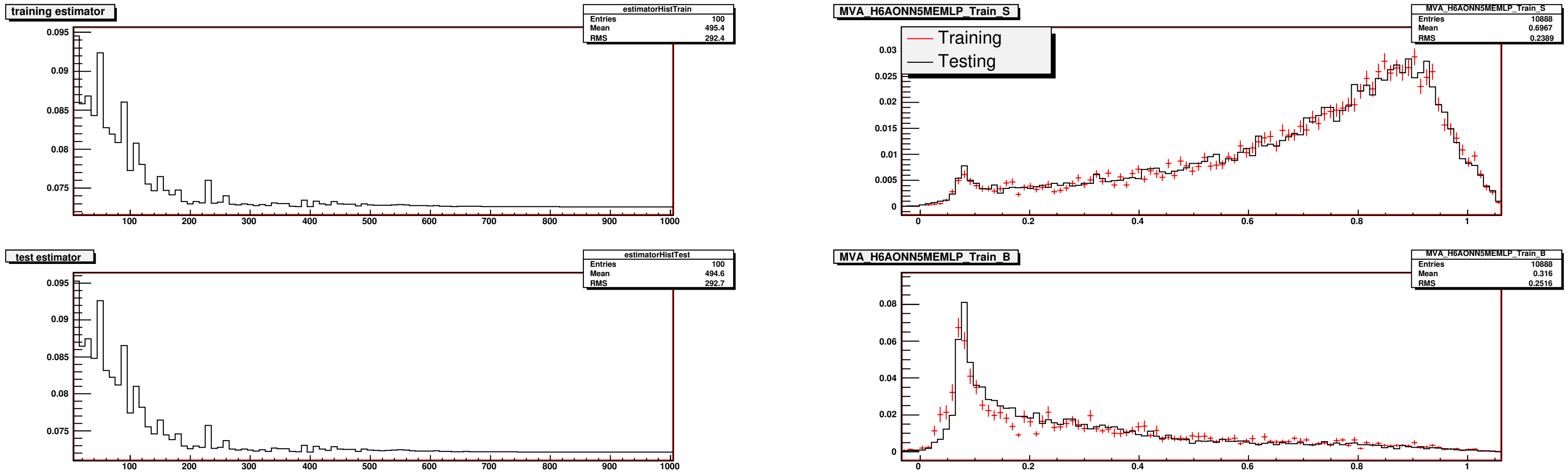
Importance: $\bar{x}_i^2 \sum_{j=1}^{n_h} \left(w_{ij}^{(1)}\right)^2, i = 1, \ldots, n_{var}$



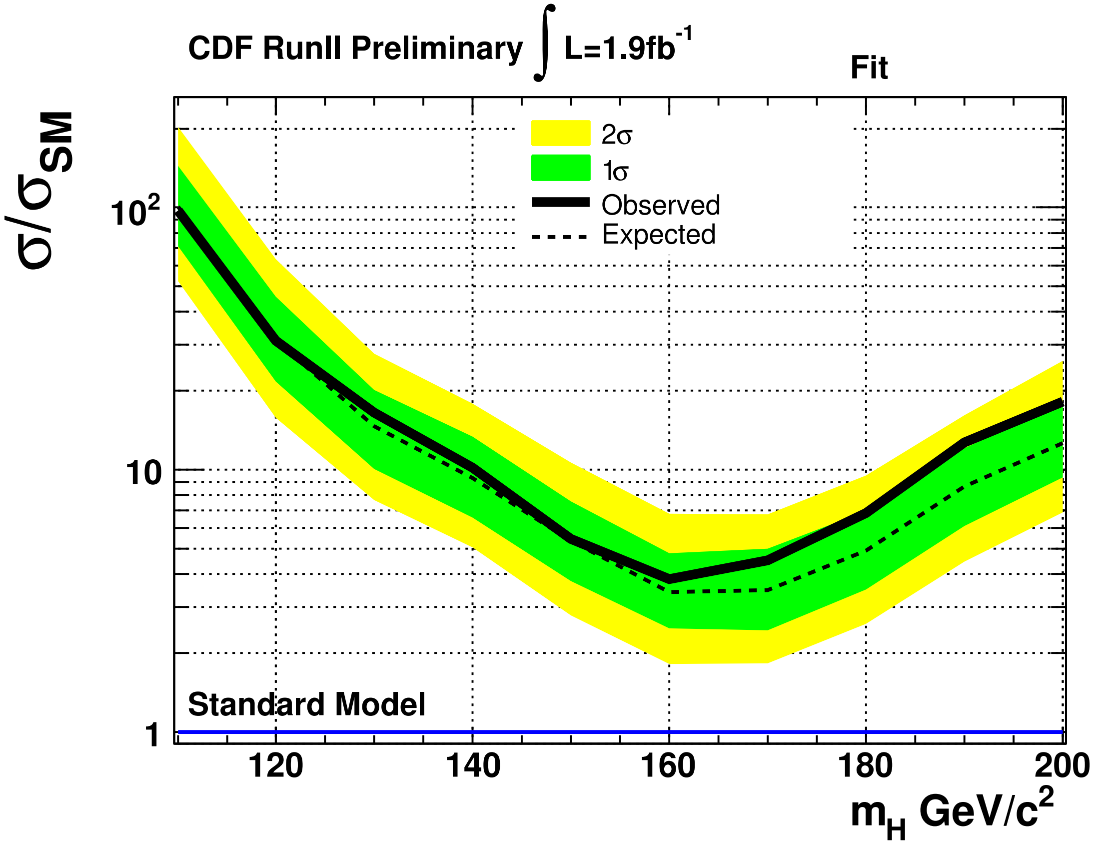


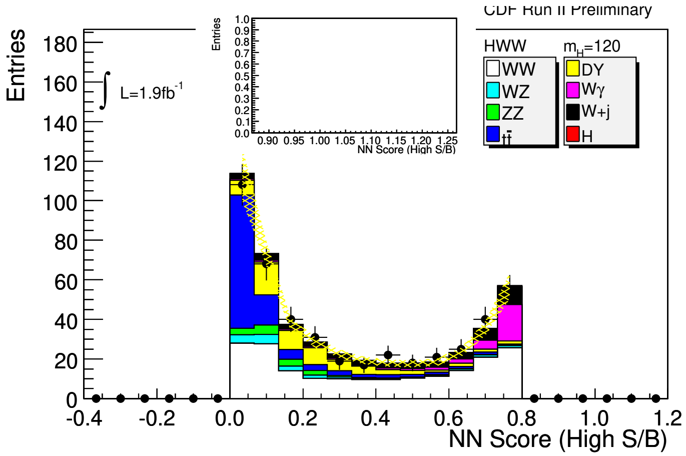


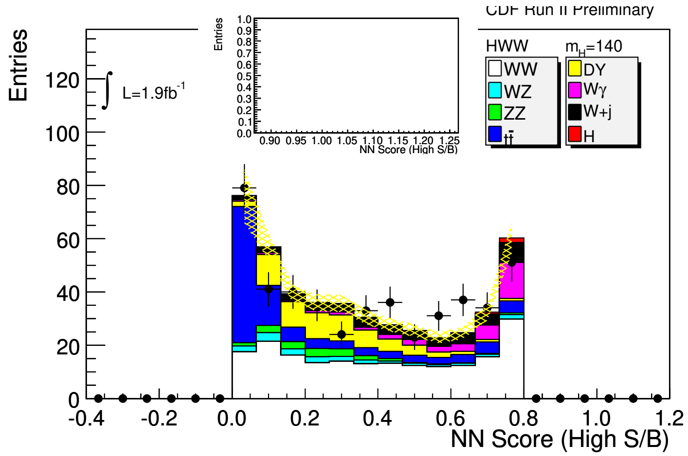
Training at 160

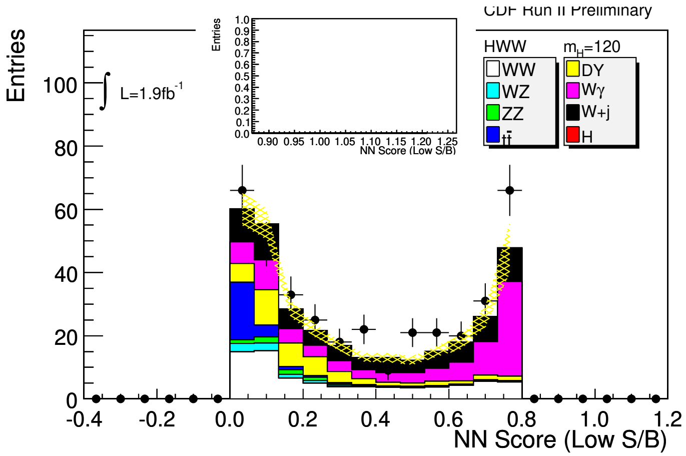


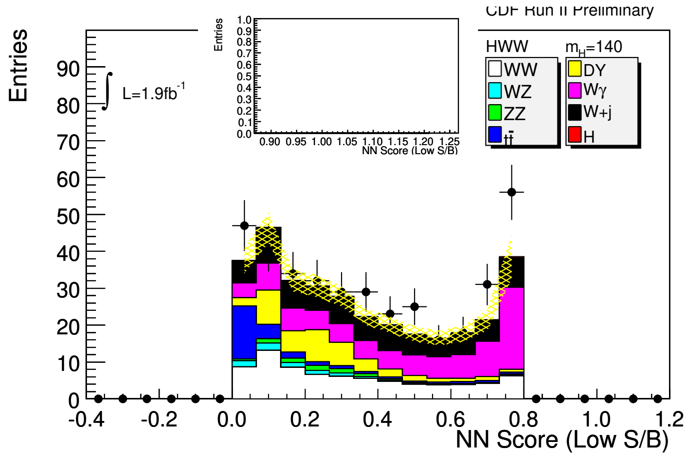
Fit of DRLeptons and Ht

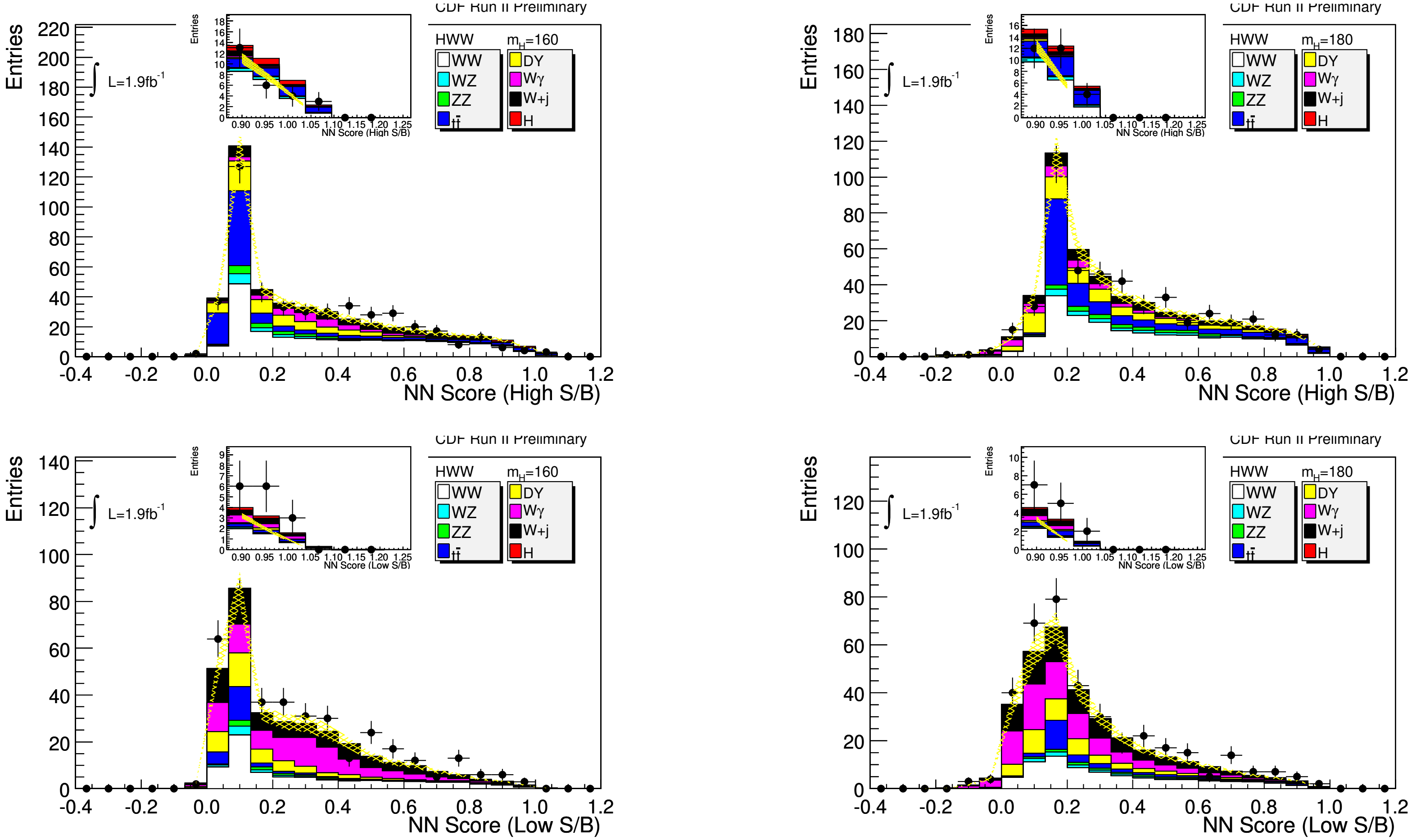



CDF Run II Preliminary
Entries
$\int$ L=1.9fb^{-1}
HWW
m$_H$=120
WW
WZ
ZZ
tt̄
DY
Wγ
W+j
H
NN Score (High S/B)


CDF Run II Preliminary
Entries
$\int$ L=1.9fb^{-1}
HWW
m$_H$=140
WW
WZ
ZZ
tt̄
DY
Wγ
W+j
H
NN Score (High S/B)


CDF Run II Preliminary
Entries
$\int$ L=1.9fb^{-1}
HWW
m$_H$=120
WW
WZ
ZZ
tt̄
DY
Wγ
W+j
H
NN Score (Low S/B)

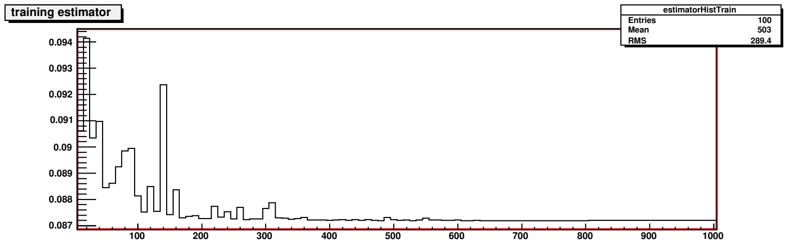

CDF Run II Preliminary
Entries
$\int$ L=1.9fb^{-1}
HWW
m$_H$=140
WW
WZ
ZZ
tt̄
DY
Wγ
W+j
H
NN Score (Low S/B)

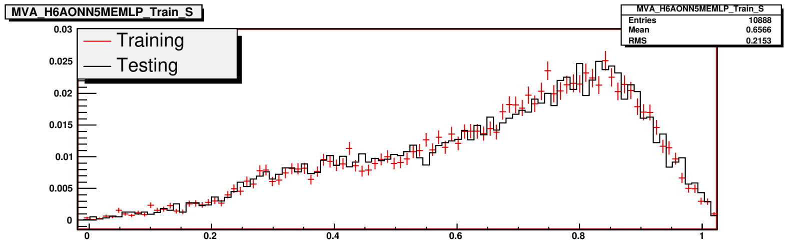

CDF Run II Preliminary
Entries
$\int$ L=1.9fb^{-1}
HWW
m_H=160
m_H=180
WW
WZ
ZZ
$t\bar{t}$
DY
Wγ
W+j
H
NN Score (High S/B)
NN Score (Low S/B)

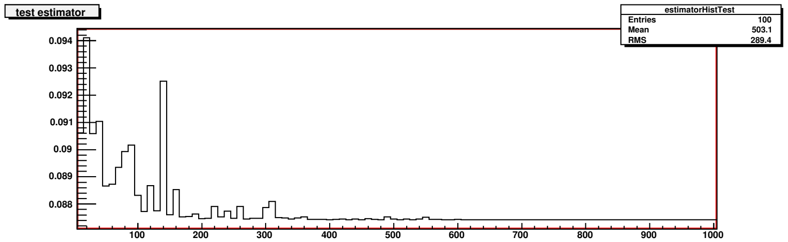
Triplets

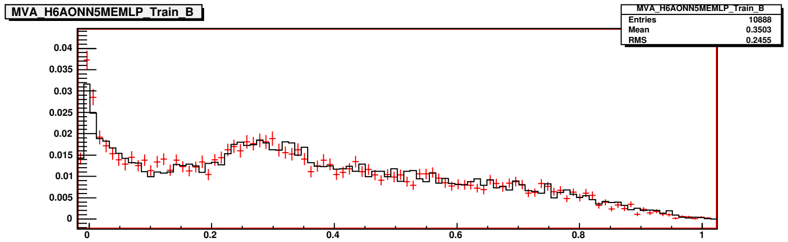
$I_{\Delta R}$	I_{Ht}	3rd Var	I_3	$\frac{\sigma}{\sigma_{SM}}(160)$	$\frac{\sigma}{\sigma_{SM}}(170)$
0.7000	0.0619	Lep2_E	0.0838	3.3	3.5
0.6610	0.0753	addEt	0.8912	3.4	3.6
0.3959	0.1452	lep1_E	0.0640	3.5	3.5
1.1120	0.1136	Met	0.0302	3.5	3.4
0.2779	0.0414	MetSpec	0.0414	3.6	3.6
0.9482	0.1242	MetDelPhi	0.0047	3.6	3.5
0.8241	0.3473	jet1_Et	0.0001	3.7	3.6
0.7105	0.1710	Njets	0.9989	3.7	4.0
0.2972	0.0917	Dimass	0.0505	3.7	3.7
1.1470	0.0727	SumEt	0.0505	3.8	3.8
0.7447	0.1149	dPhiLeptons	0.0025	3.9	3.5

Training at 160

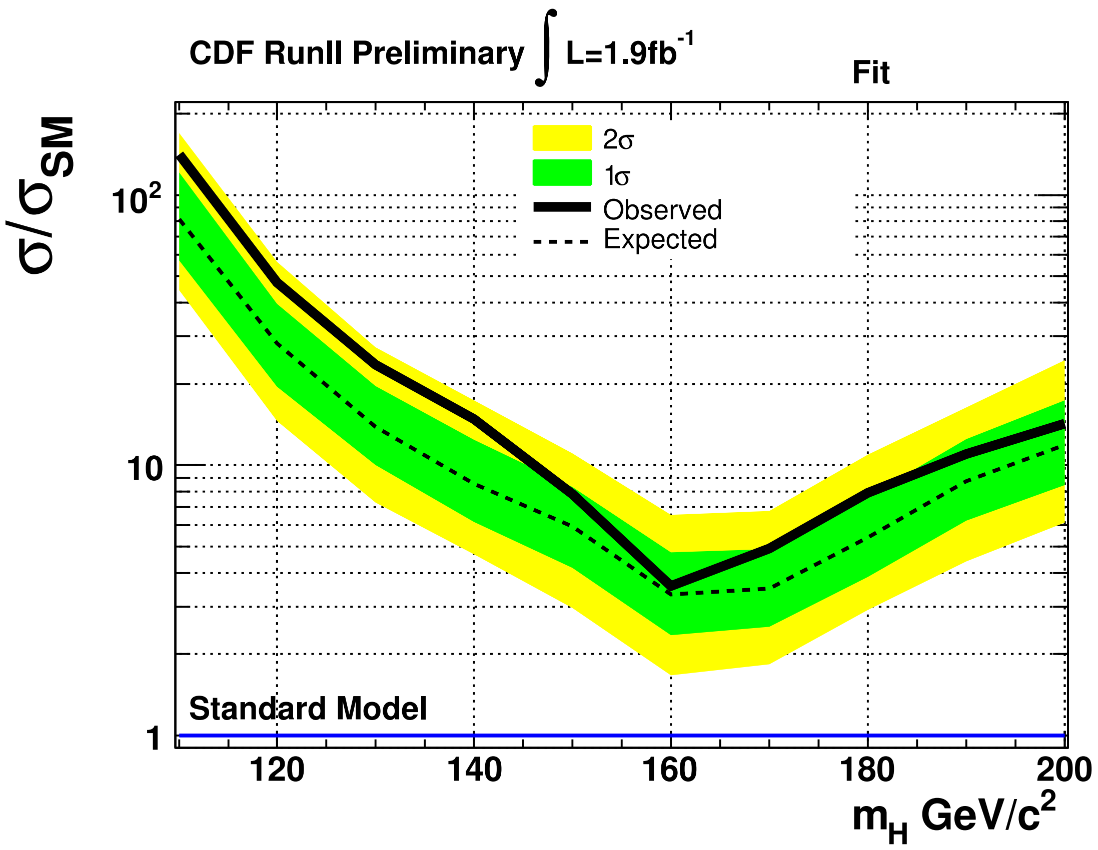


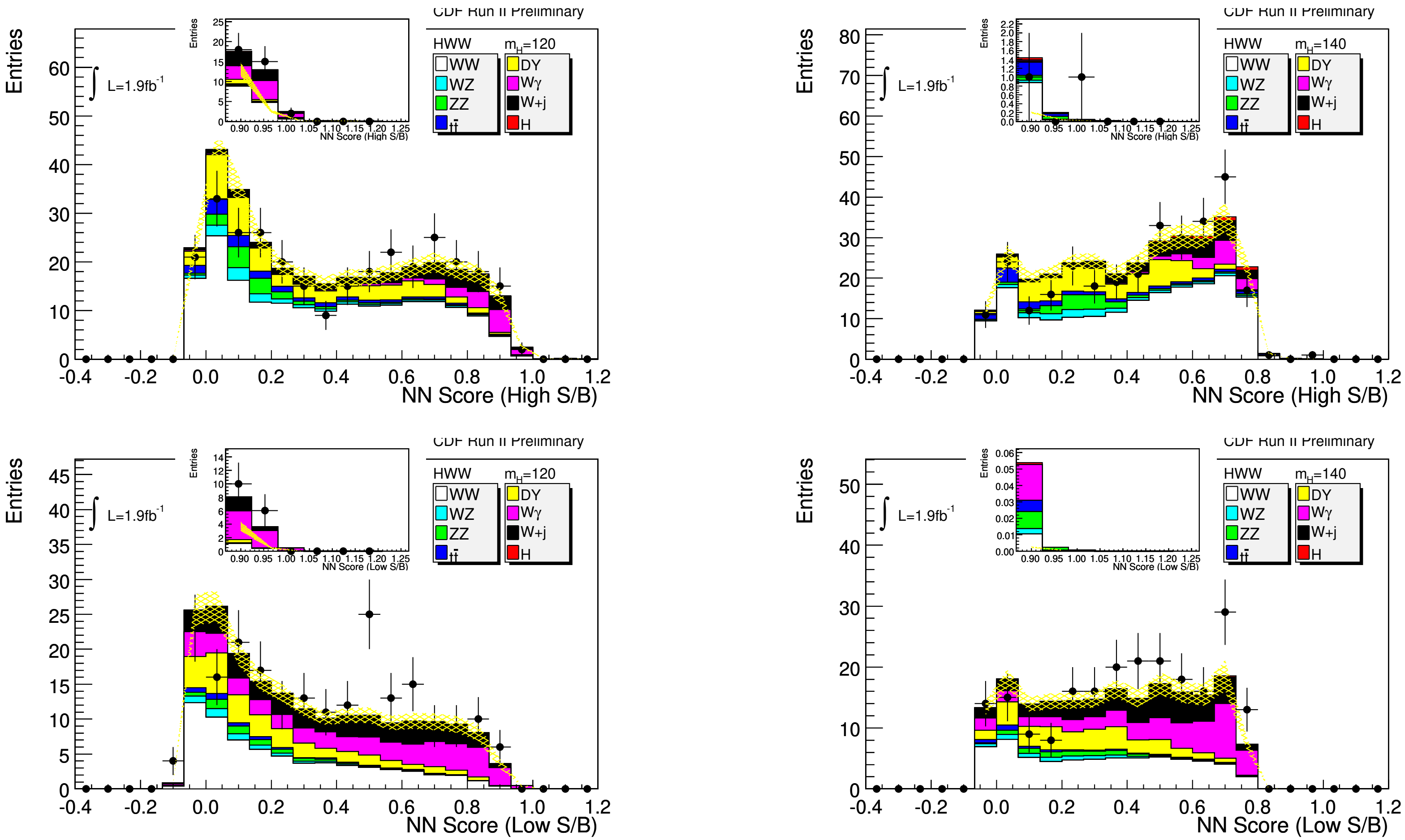


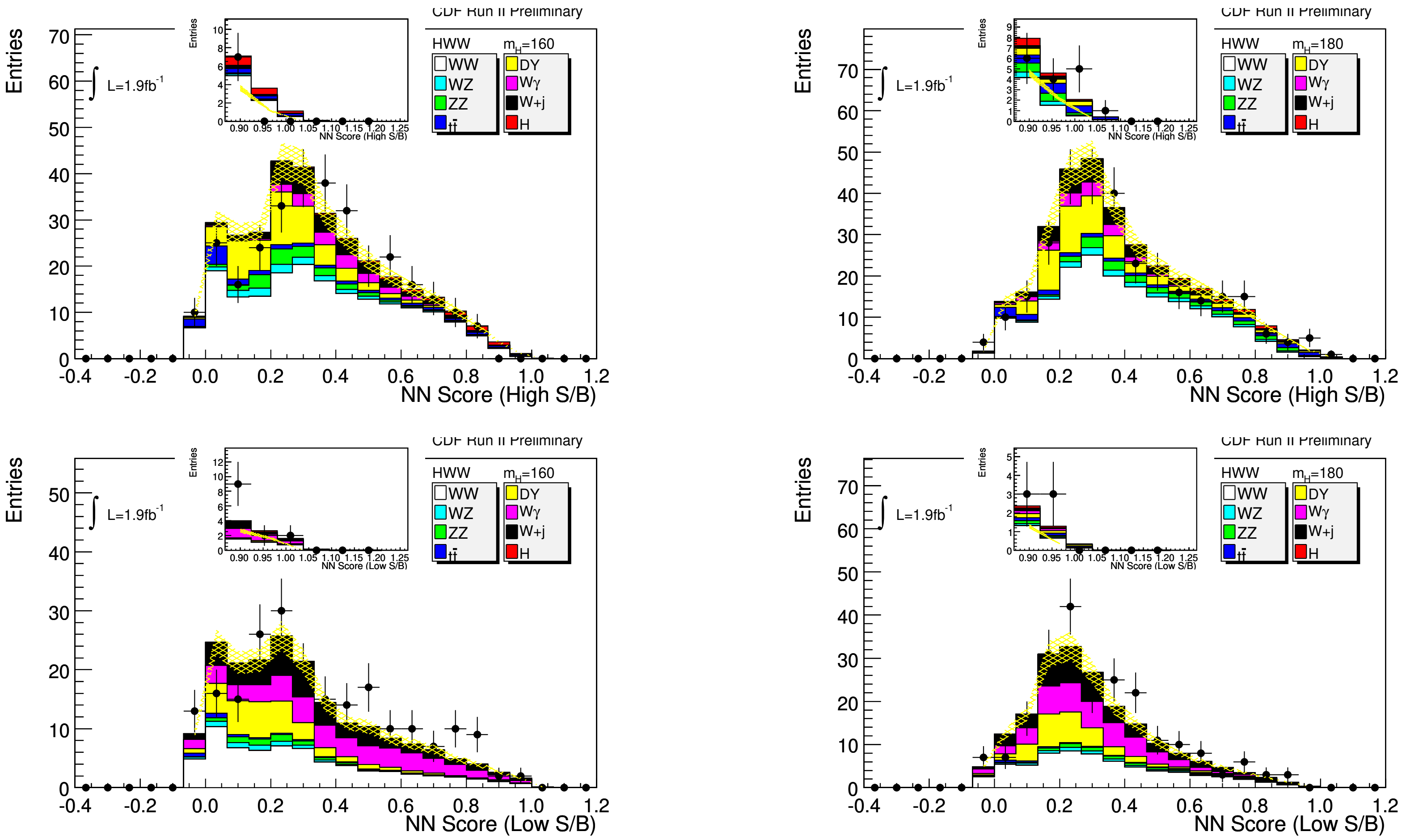




Fit of DRLeptons and Ht




CDF Run II Preliminary
Entries
$\int$ L=1.9fb^{-1}
HWW
m$_H$=120
WW
WZ
ZZ
$t\bar{t}$
DY
Wγ
W+j
H
NN Score (High S/B)
CDF Run II Preliminary
Entries
$\int$ L=1.9fb^{-1}
HWW
m$_H$=140
NN Score (High S/B)
CDF Run II Preliminary
Entries
$\int$ L=1.9fb^{-1}
HWW
m$_H$=120
NN Score (Low S/B)
CDF Run II Preliminary
Entries
$\int$ L=1.9fb^{-1}
HWW
m$_H$=140
NN Score (Low S/B)


CDF Run II Preliminary
HWW
m_H=160
m_H=180
WW
WZ
ZZ
$t\bar{t}$
DY
Wγ
W+j
H
$\int$ L=1.9fb^{-1}
Entries
NN Score (High S/B)
NN Score (Low S/B)

Summary and Conclusions (thusfar)

- Found best Separation is dRLeptons
- Got best Significance in 2 variables with dRLeptons, Ht. Noticable top background. The signficance is already 3.4 in these two variables, with a range up to 5 in all variables.
- Importance (I) and sensitivity are correlated, except for Njet. Perhaps Njet is not well handled.
- Best 3 variable significance was dRLeptons, Ht, Lep2E. The Lep2E removes the top background. The signficance decreases to 3.3, with range up to 3.9 (narrowed).
- Optimum for each mass may not be the same.
- Onwards to 4 variables, LR alone and a mix.