CDF/PHYS/EXOTIC/CDFR/XXXX
May 29, 2007
V2.11

Comparison of the NN and ME H$\rightarrow W^+W^-$ Analysis[1]

D. Benjamin[1], P. Bussey[2], D. Hidas[1], S. Hsu[3],M. Kruse[1]
E. Lipeless[3],M. Neubauer[3], A. Robson[2],
R. St. Denis[2], A.S. Thompson[2], F. Wuerthwein[3],W. Yao[5]
[1]Duke University
[2]Glasgow University
[3]University of California- San Diego
[4]University of Illinois- Urbana-Champaign
[5]Lawrence Berekely Laboratories

Abstract

Comparison of the analyses for H$\rightarrow W^+W^-$based on 1 fb^{-1} of integrated luminosity is made. Approximately 70-80% of the Monte Carlo events containing dibosons selected in the Neural Net analysis are also selected by the Matrix Element analysis, and 63% of the data events are also selected by the Matrix Element analysis. Details on all these overlaps are presented in tables. The is a high degree of correlation between the Matrix Element Likelihood Ratios and the Neural Net scores. However for WW events, although there is a large contribution for very low scores, when the score is above 0.1, there is almost no correlation between the NN and ME scores.

1 Introduction

This note describes a comparison of the analyses of 1fb^{-1} of data in looking for a contribution of H$\rightarrow W^+W^-$ to high p_t dileptons. The Matrix Element

[1]Hww/doc/CompareBlessed1fb/CompareBlessed.tex

analysis uses likelihood ratios as described in [1] and the neural net analysis described in [2] trains artificial neural nets to distinguish the Higgs contribution. The two scores are compared to one another for the common sample of Data and Monte Carlo events described in the corresponding references.

2 Analysis

The analysis was conducted on files that were produced from the Matrix Element (ME) and the Neural Net analyses (NN). Only the LR and NN discriminants for a Higgs mass, $M_h = 160$ GeV/c^2 were considered. All inputs event to these analysis came from the same Stntuples. The lepton identification and event selections differed. Of particular note is the fact that the ME analysis did not use events with jets. The NN analysis used a Drell-Yan neural net to reject Drell-Yan events and the ME analysis used a missing energy variable, "MetSigSpecial". In the NN analysis scores were computed for all events; however, due to the time that it takes to compute the ME for the Likelihood Ratio (LR), only a subsample of events in Monte Carlo have LR computations. The lists of events for the ME analysis consists then of two categories. First, there are events that are selected for the final plot of the ME values and have LR computations available. This is called the ME-LR sample. Next there are events that would have been selected but for which no LR computations have been done. This is called the ME-NoLR sample.

In comparing the analysis, the events in both the ME-LR and ME-NoLR samples are compared to those in the NN sample. This gives the overlap of events since the inputs are the same. Then events in the ME-LR sample that overlap with those in the NN sample are studied and six plots are presented:

- ME-LR vs NN Score (box style)
- ME-LR vs NN Score (lego style)
- ME-LR vs NN Score (box style, high scores)
- ME-LR vs NN Score (box style, low scores)
- NN Score - ME-LR vs NNScore
- NN Score - ME-LR vs NNScore + ME-LR

3 Results

Table 1 shows the overlap of events and the number compared. The WW sample was only done for Pythia as the NN analysis has not been done for MC@NLO. For the ME analysis, the $W\gamma$ sample is summed over the electron, muon and tau modes.

For Higgs Monte Carlo events selected by the Matrix Element analysis, 46% overlap with the Neural Net analysis. The WW, WZ and ZZ events have about 30% of the Matrix Element analysis events overlap with the Neural Net analysis and there is only 12% of the Matrix Element top events overlap with the Neural Net analysis. There is a small overlap in Fakes and the data overlap by about 24%. For Higgs Monte Carlo Events selected by the Neural Net analysis, 72% of the events overlap with the Matrix Element analysis. For WW, ZZ, WZ, and top 82%, 78% , 62% and 41% of the events in the Neural Net analysis overlap with the Matrix Element analysis. For fakes 10% of the events in the Neural Net analysis overlap with the Matrix Element analysis. For the data, 63% of the Neural Net analysis events overlap with the Matrix Element Analysis.

The plots for each sample of events are shown in Figures 1 to 10. The is a high degree of correlation between the Matrix Element Likelihood Ratios and the Neural Net scores. However for WW events, although there is a large contribution for very low scores, when the score is above 0.1, there are a significant number of events (to be quantified) that are not on the diagonal of the plot where ME and NN scores are the same.

The MC@NLO analyses is missing from the NN and the up and down variations in the fakes are available but not provided for fitting yet.

Events that are in the NN analysis but not in the ME analysis are likely to be due to:

- The lower $M_{\ell\ell}$ cut in the NN analysis
- The looser jet cut, specifically 2-jet events
- Events with low missing E_t but small dilepton angle can pass the DY-NN but fail in high MetSpecial.

Events that are in the ME analysis but not in the NN analysis are likely to be due to the extra acceptance gained by including more lepton categories.

4 Things to Do

A number of desirable studies are to be done:

- Study the dilepton categories in the common and different samples.
- Determine if the same leptons are actually used in the events that are common.
- Study the NJet dependence of the scores.
- Examine the region outside the peak at low scores in the simluated diboson events.
- Study summed weights (events) in overlap region.
- Put in plots for other masses for higgs
- Put in plots for weights and weighted scores for non-overlapping regions

5 Conclusion

The ME and NN analyses show a large overlap in the scores but there is likeley to be some gain in combining the methods. Understanding why there is any spread in the scores and what the origin is would be quite useful. It is particularly curious that the WW scores that are not very low show very little correlation between the two analyses.

	hww	WWP	ZZ	WZ	$t\bar{t}$
Total NN	143280	187869	42726	71668	2938
Total ME	222056	483313	121442	182755	10448
NN not ME	40502	33894	9410	27276	1727
ME not NN	119278	329338	88126	138363	9237
Overlap	102778	153975	33316	44392	1211
Total ME-LR	60684	62277	26123	7330	4036
ME-LR not NN	32525	42392	18928	5541	3588
Overlap ME-LR	28159	19885	7195	1789	448

	Zee	Zmm	Ztt	Wg	Fake	Data
Total NN	231	279	477	193	331	123
Total ME	349	582	227	587	1839	323
NN not ME	197	249	475	76	299	46
ME not NN	315	552	225	470	1807	246
Overlap	34	30	2	117	32	77
Total ME-LR	46	130	227	587	1839	323
ME-LR not NN	42	122	225	470	1807	246
Overlap ME-LR	4	8	2	117	32	77

Table 1: Comparison of numbers of events in the ME and NN analysis for each process that produces high p_t dileptons.

	hww	WWP	ZZ	WZ	$t\bar{t}$
NN not ME	0.28	0.18	0.22	0.38	0.59
ME not NN	0.54	0.68	0.73	0.76	0.88
Overlap/NN	0.72	0.82	0.78	0.62	0.41
Overlap/ME	0.46	0.32	0.27	0.24	0.12
ME-LR not NN	0.54	0.68	0.72	0.76	0.89
Overlap ME-LR	0.46	0.32	0.28	0.24	0.11

	Zee	Zmm	Ztt	Wg	Fake	Data
NN not ME	0.85	0.89	1	0.39	0.9	0.37
ME not NN	0.9	0.95	0.99	0.8	0.98	0.76
Overlap/NN	0.15	0.11	0.0042	0.61	0.097	0.63
Overlap/ME	0.097	0.052	0.0088	0.2	0.017	0.24
ME-LR not NN	0.91	0.94	0.99	0.8	0.98	0.76
Overlap ME-LR	0.087	0.062	0.0088	0.2	0.017	0.24

Table 2: Comparison of overlap fractions in the ME and NN analysis for each process that produces high p_t dileptons. For NN not ME, the fraction is taken relative to the total in NN. Similarly for ME not NN. For the overlap of the entire sample, the fraction is taken relative to the NN.

	110	120	130	140	150
Total NN	43164	50134	85237	87144	100867
Total ME	67792	76560	136361	137082	158914
NN not ME	17250	17631	26819	25474	28552
ME not NN	41878	44057	77943	75412	86599
Overlap	25914	32503	58418	61670	72315
Total ME-LR	22006	24911	54017	28826	29871
ME-LR not NN	13593	14400	30776	15779	16469
Overlap ME-LR	8413	10511	23241	13047	13402

	160	170	180	190	200
Total NN	143280	119432	150552	138177	134653
Total ME	222056	187114	258875	253214	231647
NN not ME	40502	30979	34596	28747	71328
ME not NN	119278	98661	142919	143784	168322
Overlap	102778	88453	115956	109430	63325
Total ME-LR	60684	31851	32627	32954	66879
ME-LR not NN	32525	16684	17882	18666	48610
Overlap ME-LR	28159	15167	14745	14288	18269

Table 3: Comparison of numbers of events in the ME and NN analysis for various Higgs masses.

	110	120	130	140	150
NN not ME	0.4	0.35	0.31	0.29	0.28
ME not NN	0.62	0.58	0.57	0.55	0.54
Overlap/NN	0.6	0.65	0.69	0.71	0.72
Overlap/ME	0.38	0.42	0.43	0.45	0.46
ME-LR not NN	0.62	0.58	0.57	0.55	0.55
Overlap ME-LR	0.38	0.42	0.43	0.45	0.45
	160	170	180	190	200
NN not ME	0.28	0.26	0.23	0.21	0.53
ME not NN	0.54	0.53	0.55	0.57	0.73
Overlap/NN	0.72	0.74	0.77	0.79	0.47
Overlap/ME	0.46	0.47	0.45	0.43	0.27
ME-LR not NN	0.54	0.52	0.55	0.57	0.73
Overlap ME-LR	0.46	0.48	0.45	0.43	0.27

Table 4: Comparison of overlap fractions in the ME and NN analysis for various higgs massses. For NN not ME, the fraction is taken relative to the total in NN. Similarly for ME not NN. For the overlap of the entire sample, the fraction is taken relative to the NN.

	CDF8719			Computed		
Channel	ME-HiBkg	ME-LowBkg	Total	All	ME-Only	Overlap
WW	96.8	36.1	132.9	121.0	81.4	39.5
WZ	6.4	3.1	9.5	9.09	6.83	2.26
ZZ	8.4	3.3	11.7	11.7	8.45	3.24
$t\bar{t}$	7.9	1.8	9.7	9.47	8.41	1.07
Zee				21.3	19.5	1.83
Zmm				25.1	23.4	1.69
Ztt				12.0	11.9	0.11
DYTot	27.7	27.7	55.4	58.4	54.8	3.63
Fake	17.4	25.0	42.4	42.2	41.7	0.545
Wγ	5.5	19.2	24.7	24.7	19.2	5.52
Subtotal	170.1	116.0	286.1	276.6	220.8	55.8
Data	187	136	323	323	246	77

Table 5: The summed weights for the ME analysis: non-Higgs sources.

	CDF8700	Computed		
Channel	Total	All	NN-Only	Overlap
Ww	48.67	48.8	8.74	40.00
WZ	3.63	3.63	1.35	2.29
ZZ	3.76	3.85	0.848	3.00
Ttbar	2.69	2.67	1.57	1.10
Zee		11.8	9.99	1.76
Zmm		12.8	11.5	1.36
Ztt		12.4	12.4	0.0474
DYTot	37.15	37.0	33.9	3.17
Fake	12.01	16.6	14.8	1.85
Wγ	9.60	6.37	2.42	3.94
hline Subtotal	117.51	119	63.6	55.3
Data	121	123	46	77

Table 6: The summed weights for the NN analysis: non-Higgs sources.

Source	Category	Higgs Mass GeV/c^2									
		110	120	130	140	150	160	170	180	190	200
CDF8719	HiBkg	0.1	0.5	1.1	1.9	2.5	3.1	3.1	2.6	1.9	1.5
	LoBkg	0.0	0.1	0.3	0.5	0.7	0.8	0.8	0.7	0.5	0.4
	Total	0.1	0.6	1.4	2.4	3.2	3.9	3.9	3.3	2.4	1.9
Computed	Total	0.165	0.611	1.43	2.43	3.23	3.90	3.84	3.26	2.37	1.97
	ME-Only	0.0999	0.344	0.79	1.29	1.72	2.03	1.95	1.74	1.31	1.41
	Overlap	0.0653	0.268	0.635	1.14	1.50	1.87	1.89	1.51	1.06	0.559
CDF8700		0.2	0.3	0.9	1.5	1.9	2.5	2.4	1.8	1.3	0.9
Computed	Total	0.106	0.397	0.875	1.51	2.02	2.49	2.38	1.84	1.25	0.972
	NN-Only	0.0422	0.139	0.274	0.44	0.568	0.702	0.615	0.42	0.258	0.515
	Overlap	0.0637	0.258	0.601	1.07	1.45	1.79	1.76	1.42	0.998	0.457

Table 7: The summed weights for the ME and NN analysis: Higgs sources.

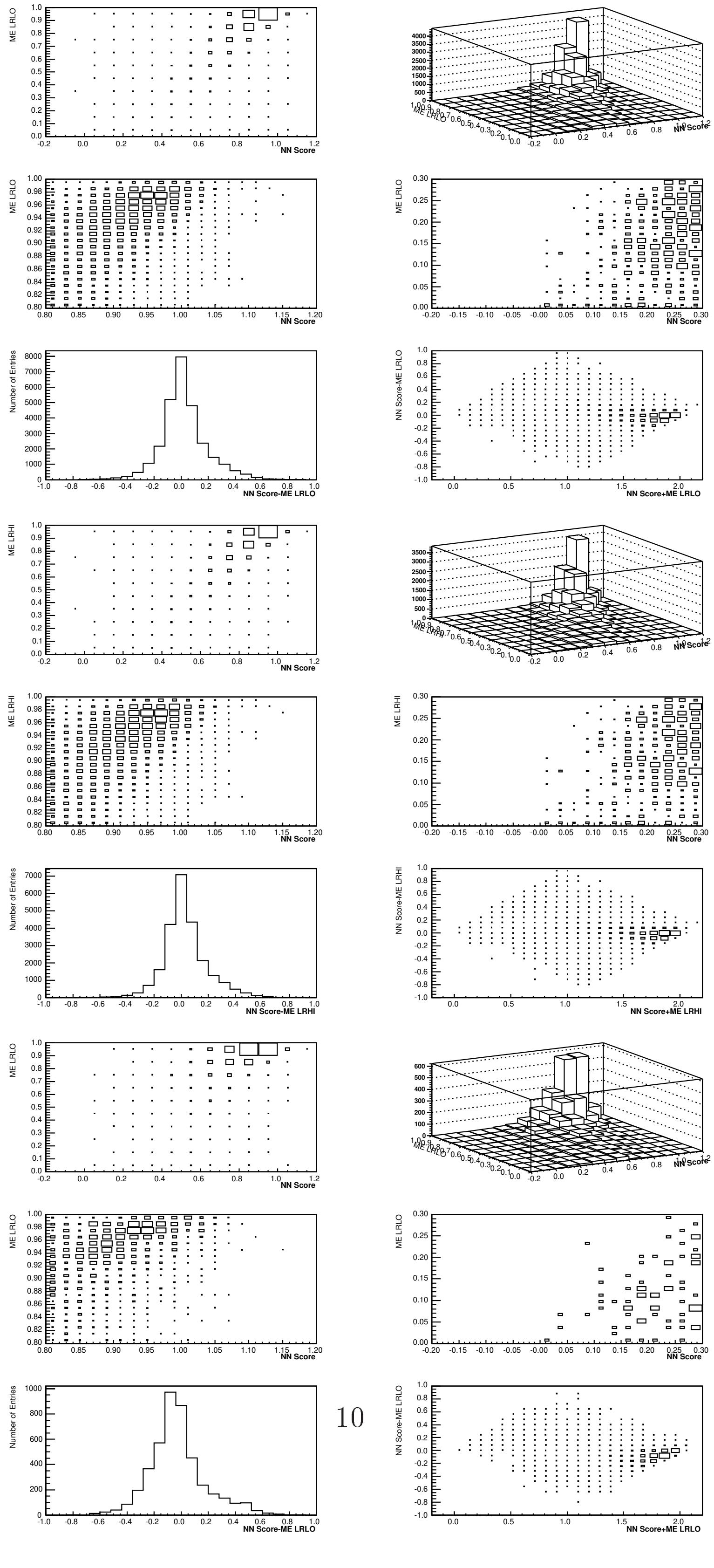

Figure 1: Plots for Higgs at $M_h = 160$ GeV/c^2: All, High,Low Signal/Background.

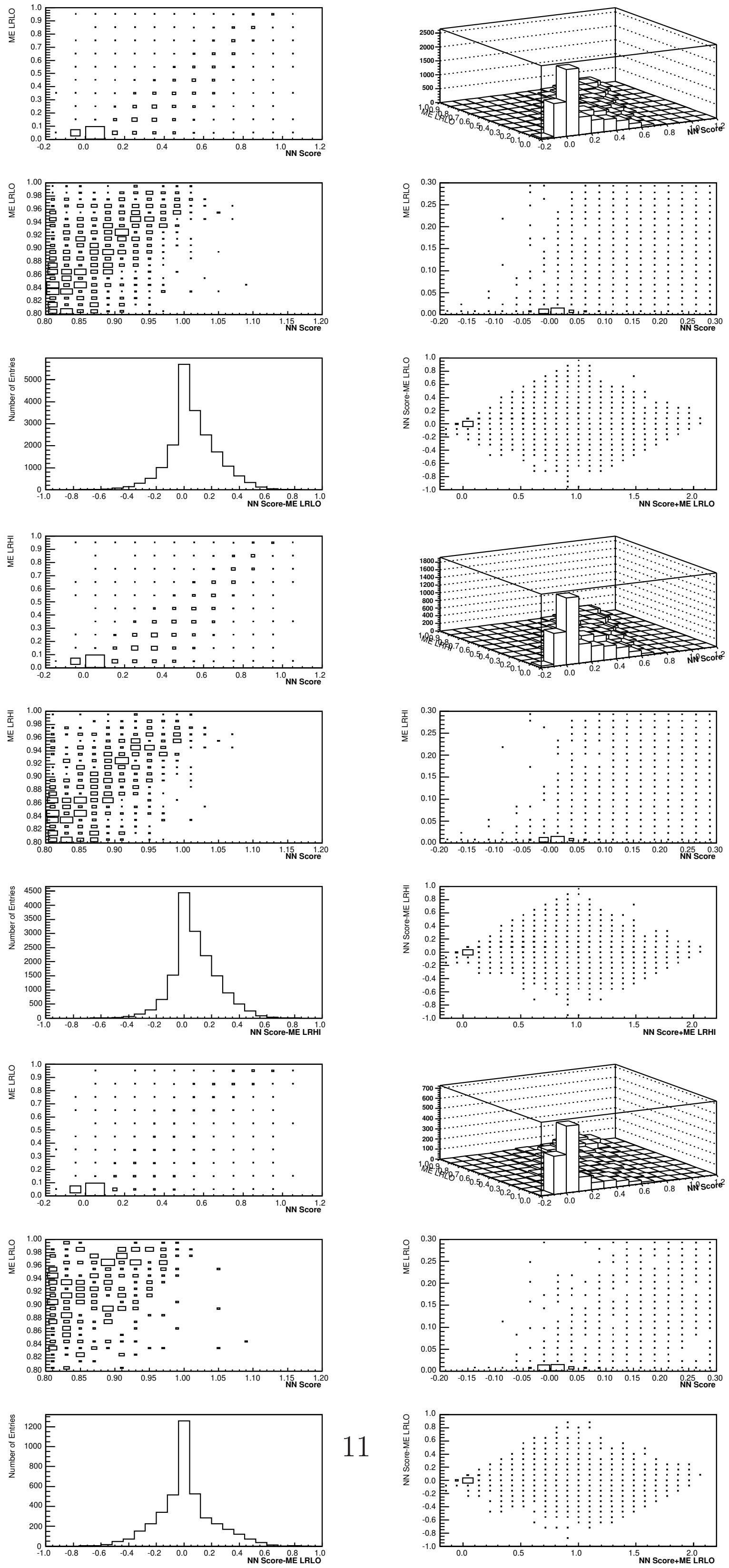

Figure 2: Plots for WW from Pythia based on Higgs at $M_h = 160$ GeV/c^2: All, High,Low Signal/Background.

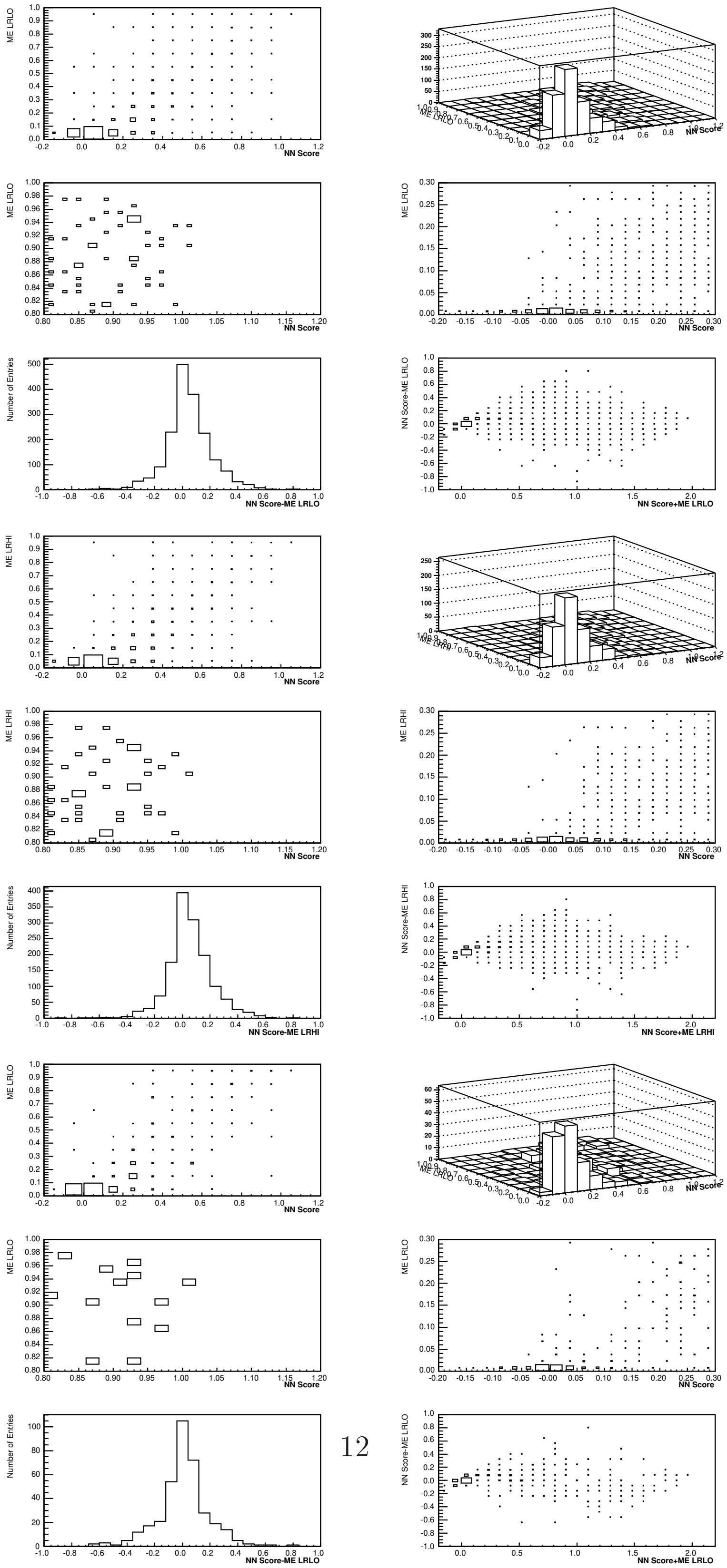

Figure 3: Plots for WZ based on Higgs at $M_h = 160$ GeV/c^2: All, High,Low Signal/Background.

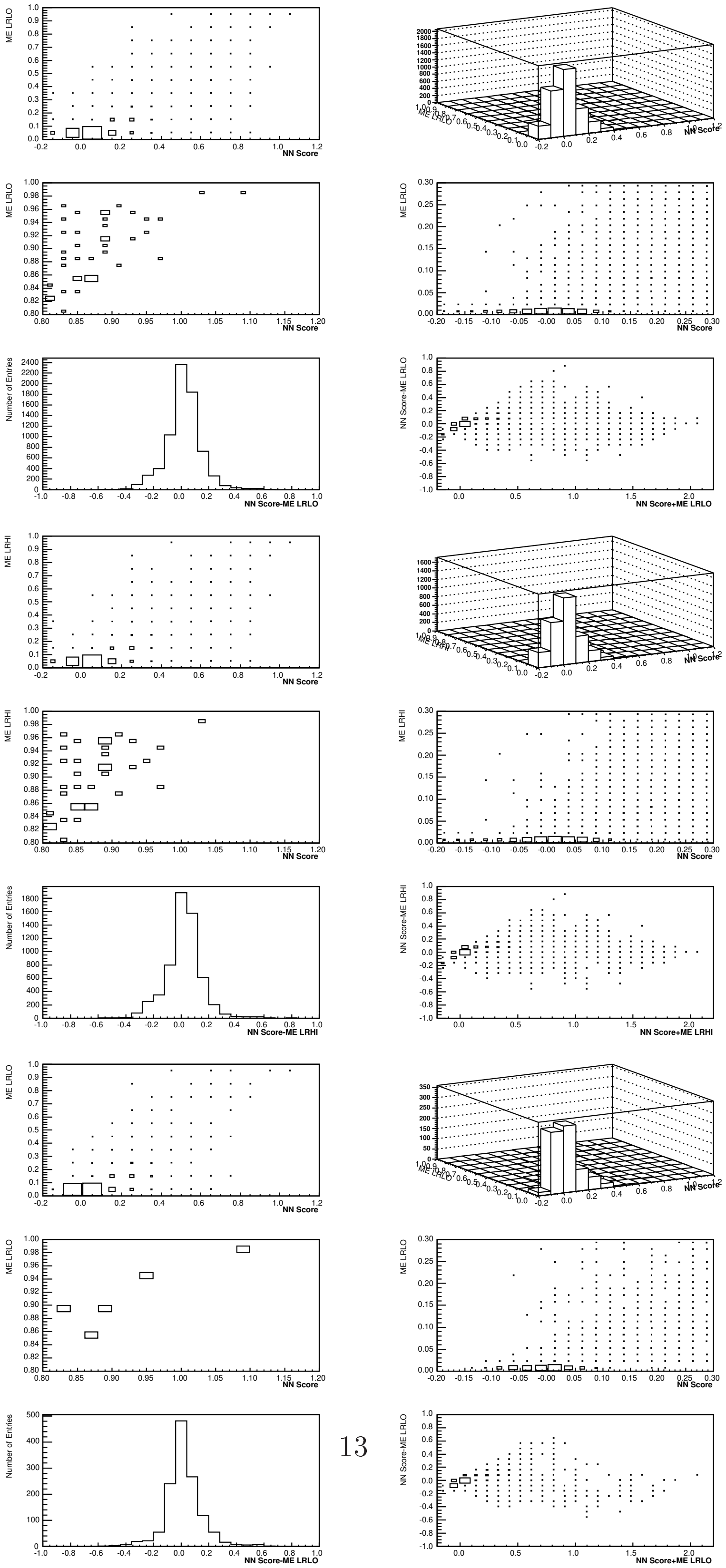

Figure 4: Plots for ZZ based on Higgs at $M_h = 160$ GeV/c^2: All, High,Low Signal/Background.

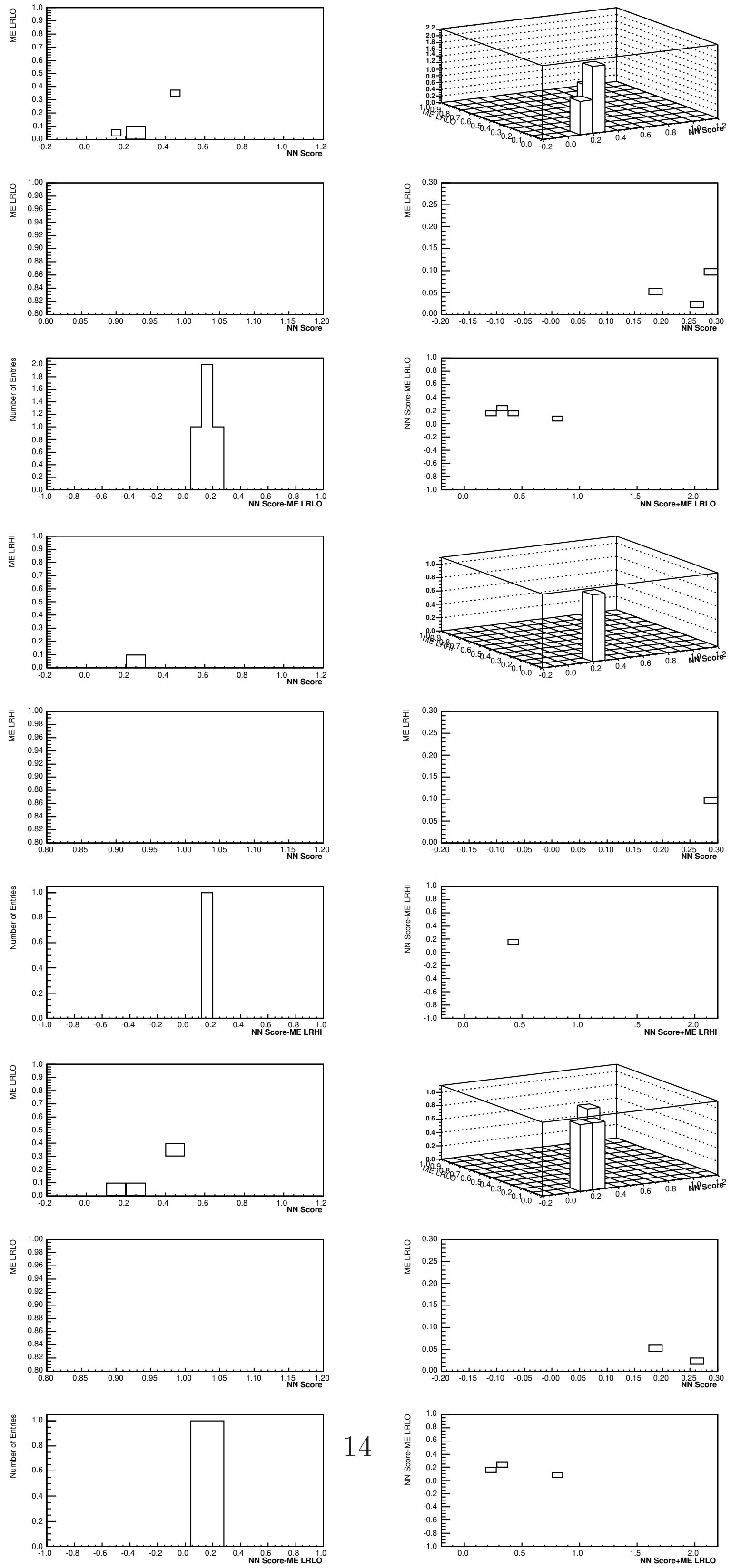

Figure 5: Plots for $Z^0 \to e^+e^-$ based on Higgs at $M_h = 160$ GeV/c^2: All, High,Low Signal/Background.

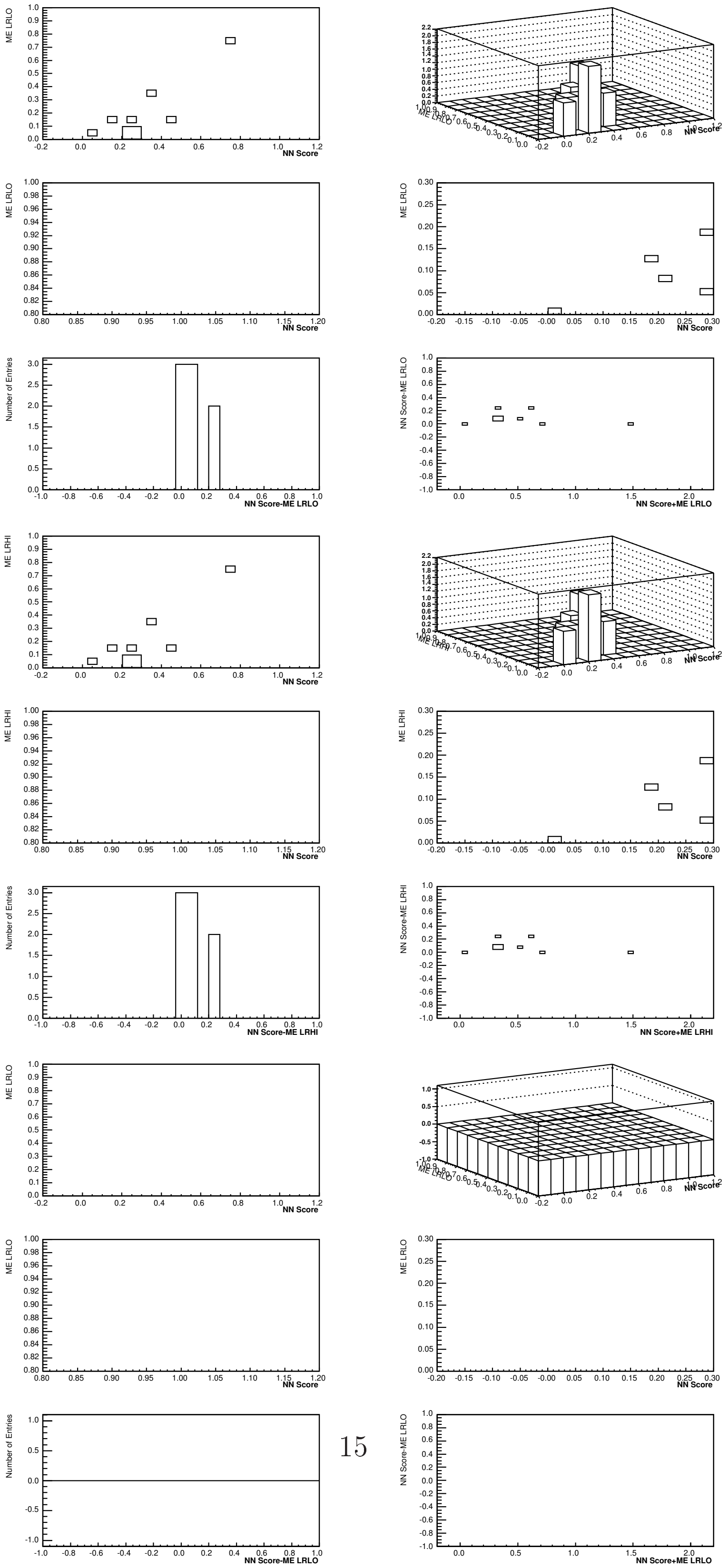

Figure 6: Plots for $Z^0 \to \mu^+\mu^-$ based on Higgs at $M_h = 160$ GeV/c^2: All, High,Low Signal/Background.

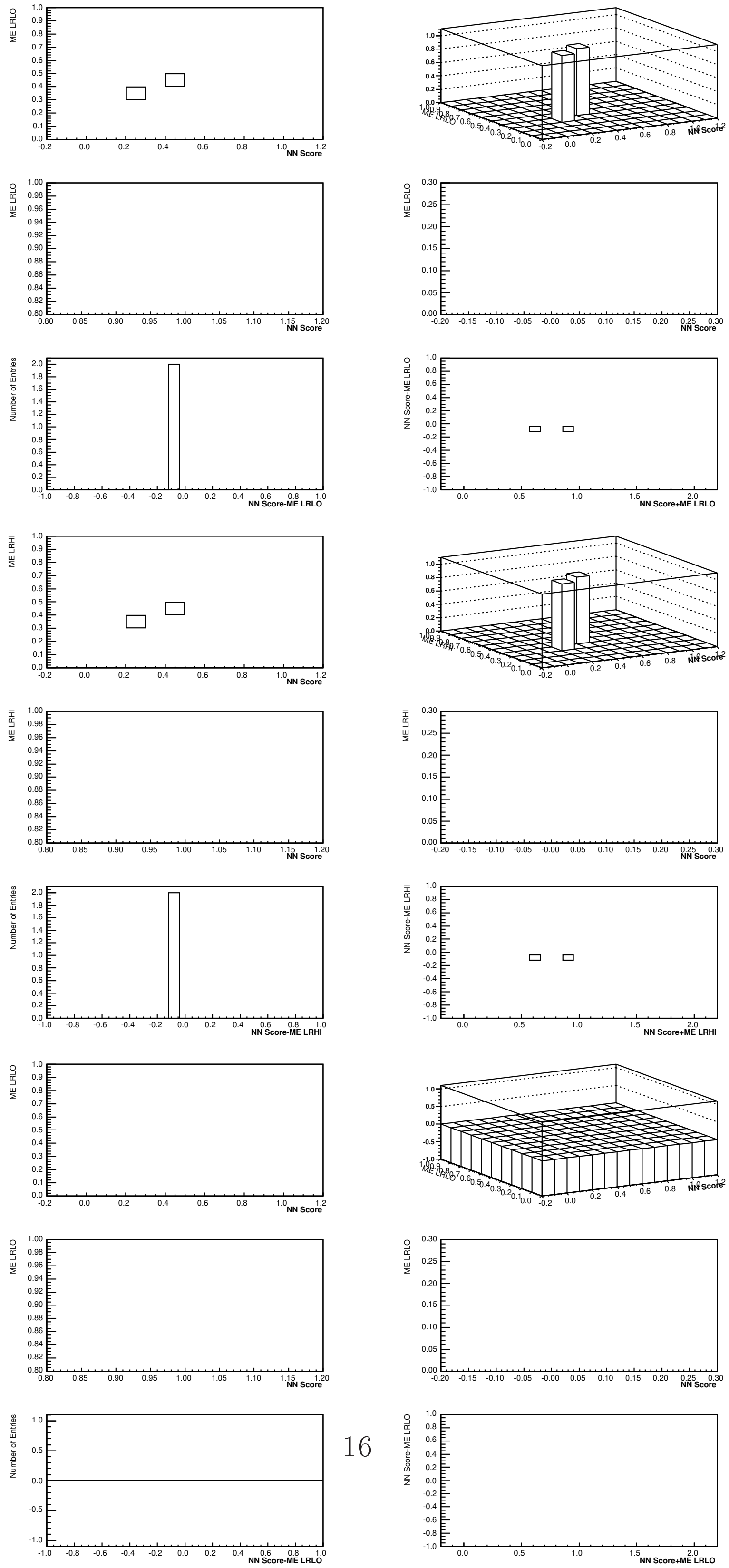

Figure 7: Plots for $Z^0 \to \tau^+\tau^-$ based on Higgs at $M_h = 160$ GeV$/c^2$: All, High,Low Signal/Background.

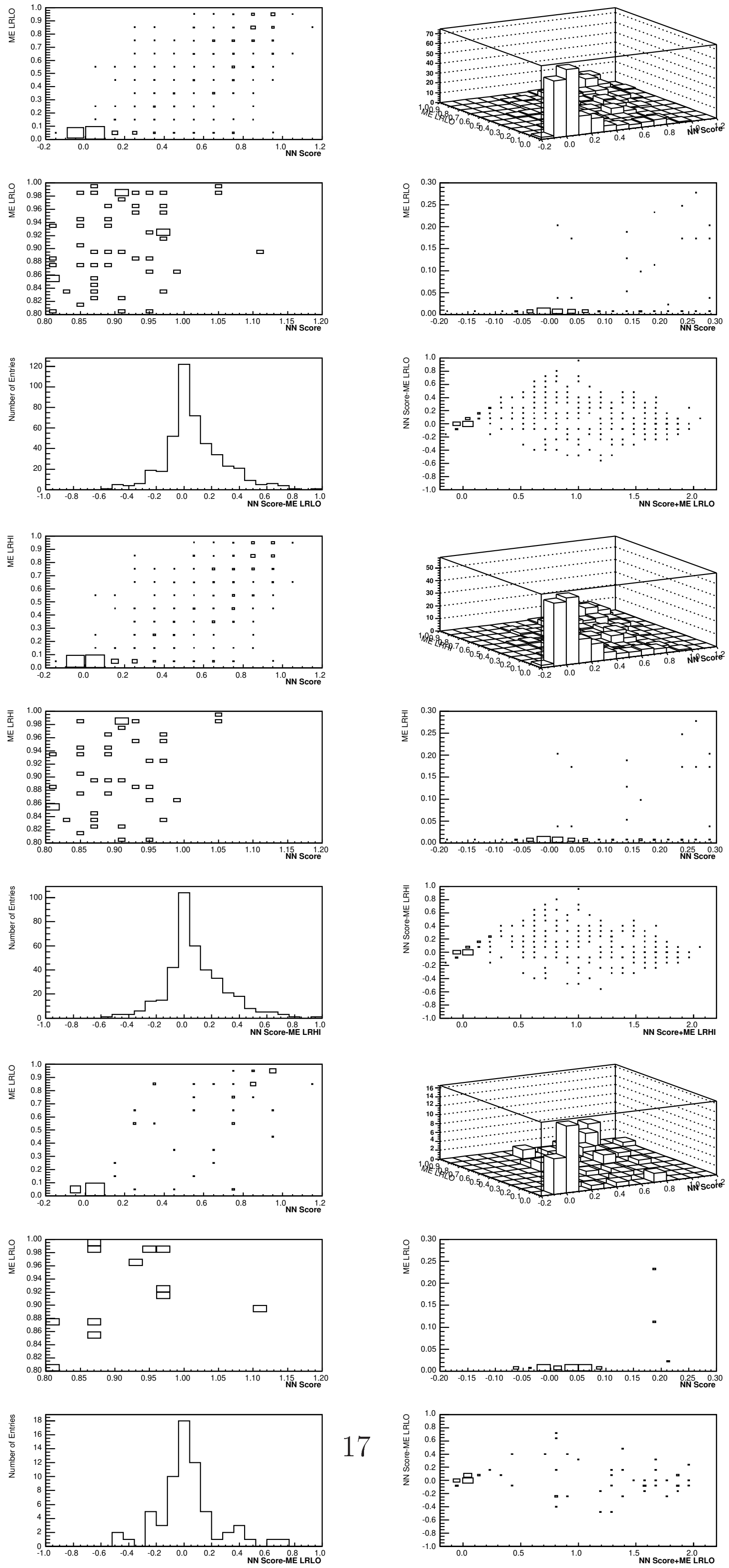

Figure 8: Plots for $t\bar{t}$ based on Higgs at $M_h = 160$ GeV$/c^2$: All, High,Low Signal/Background.

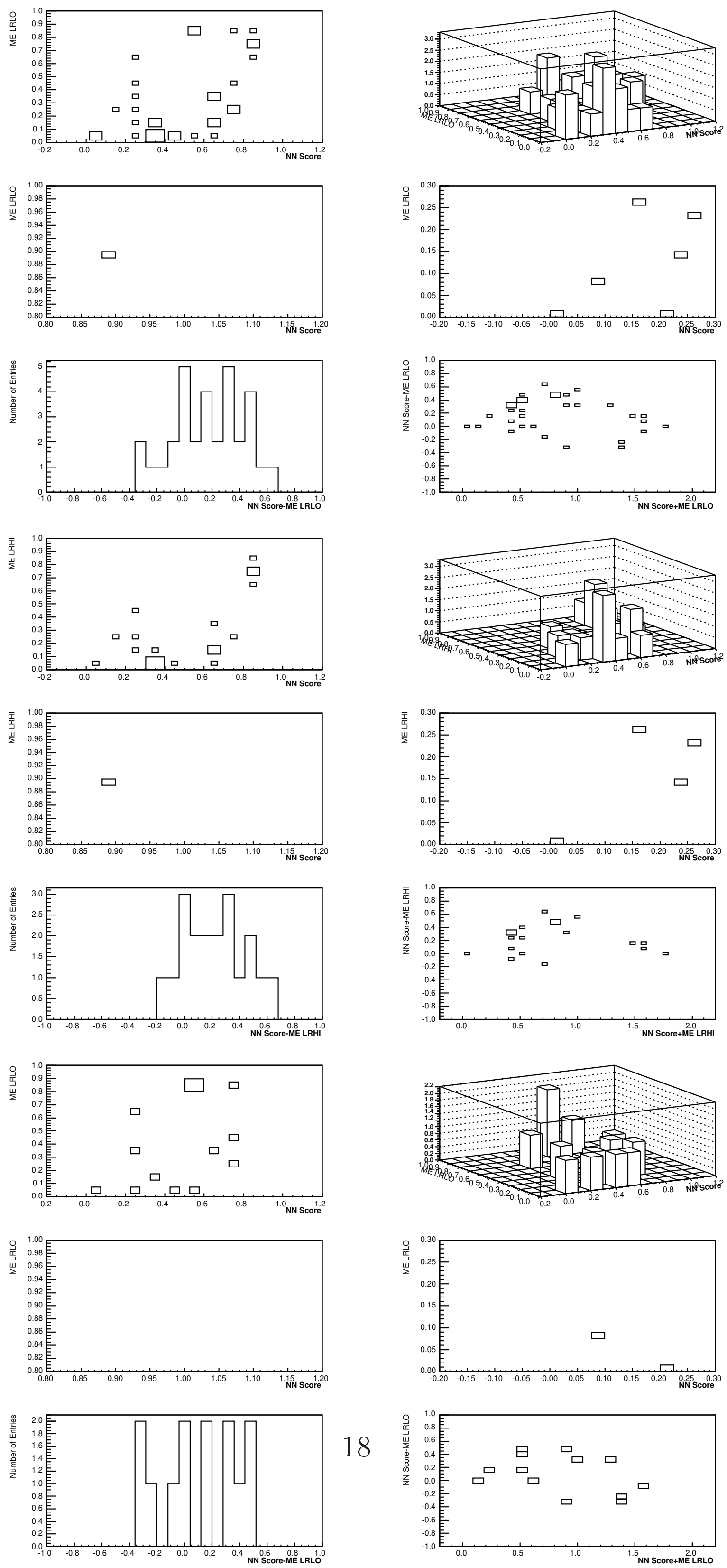

Figure 9: Plots for Fakes based on Higgs at $M_h = 160$ GeV$/c^2$: All, High,Low Signal/Background.

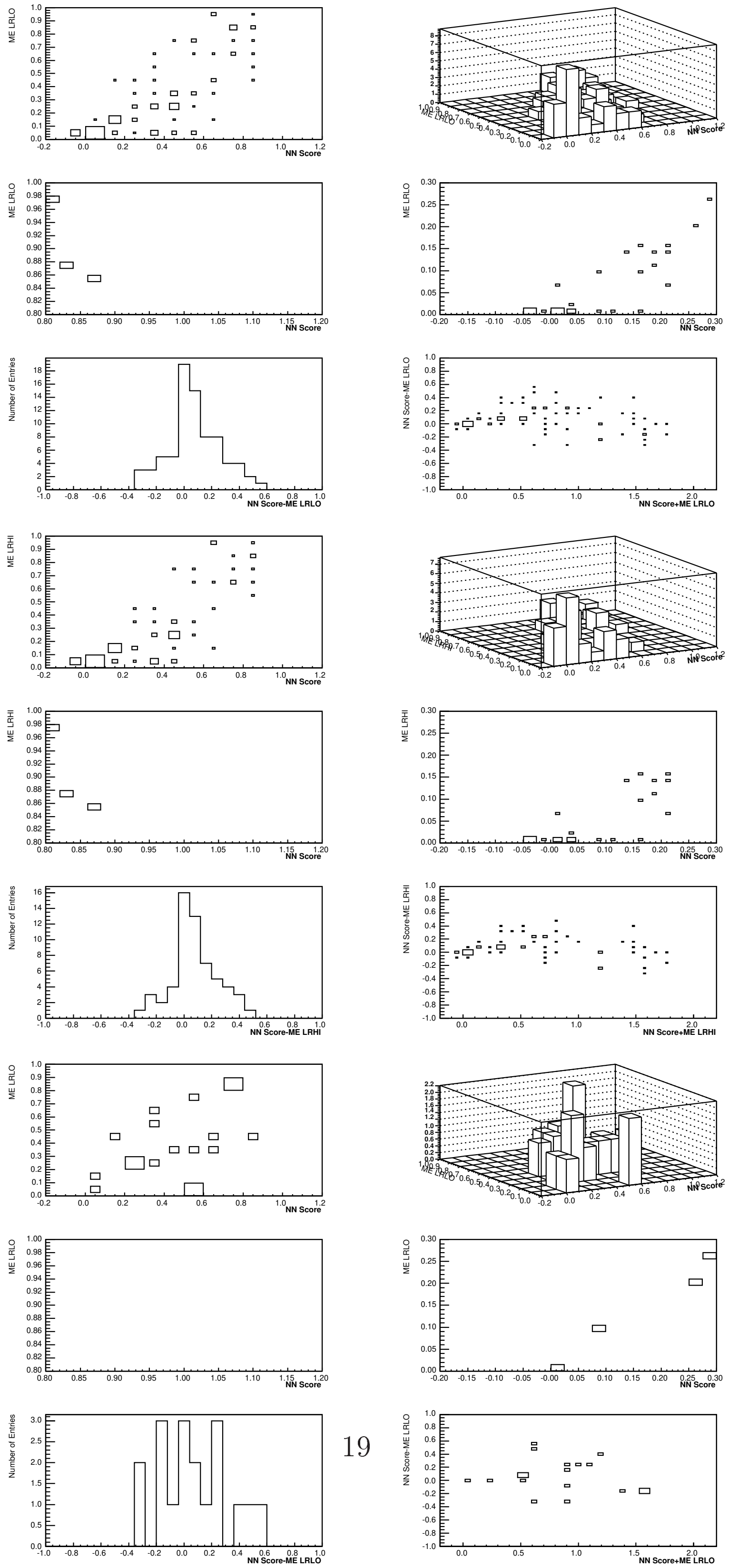

Figure 10: Plots for Data based on Higgs at $M_h = 160\,\mathrm{GeV}/c^2$: All, High,Low Signal/Background.

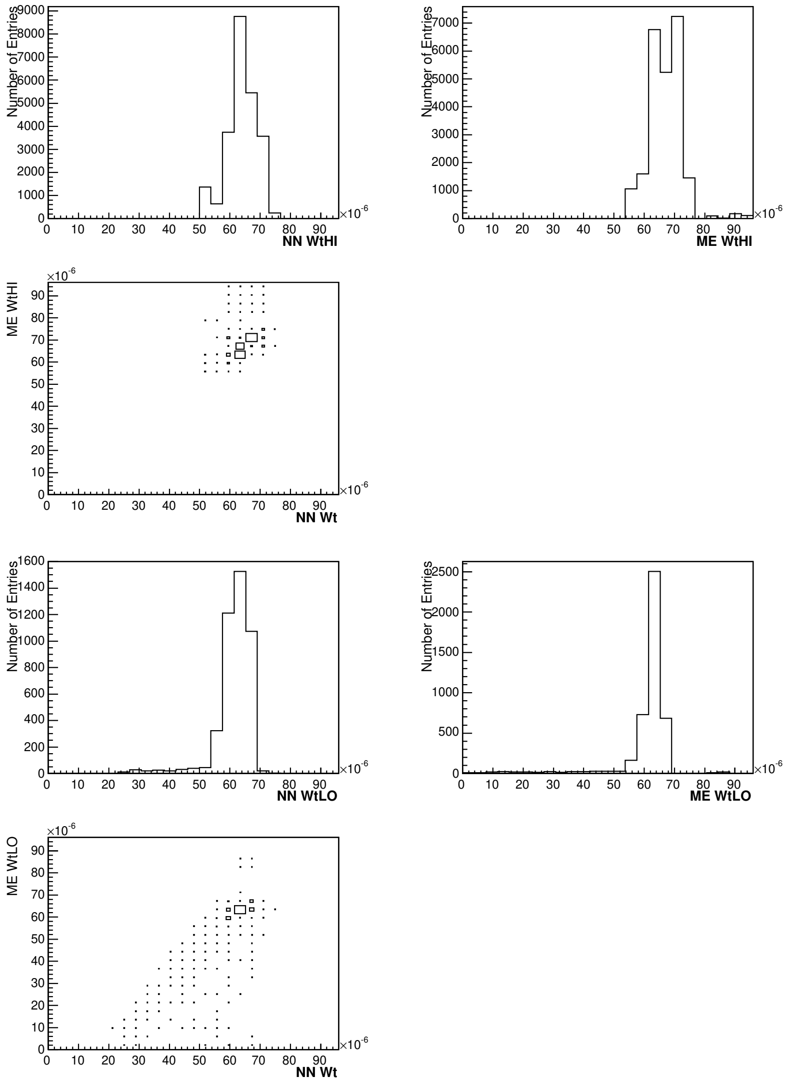

Figure 11: Weights Comparisons for Higgs at $M_h = 160$ GeV/c^2: High,Low Signal/Background.

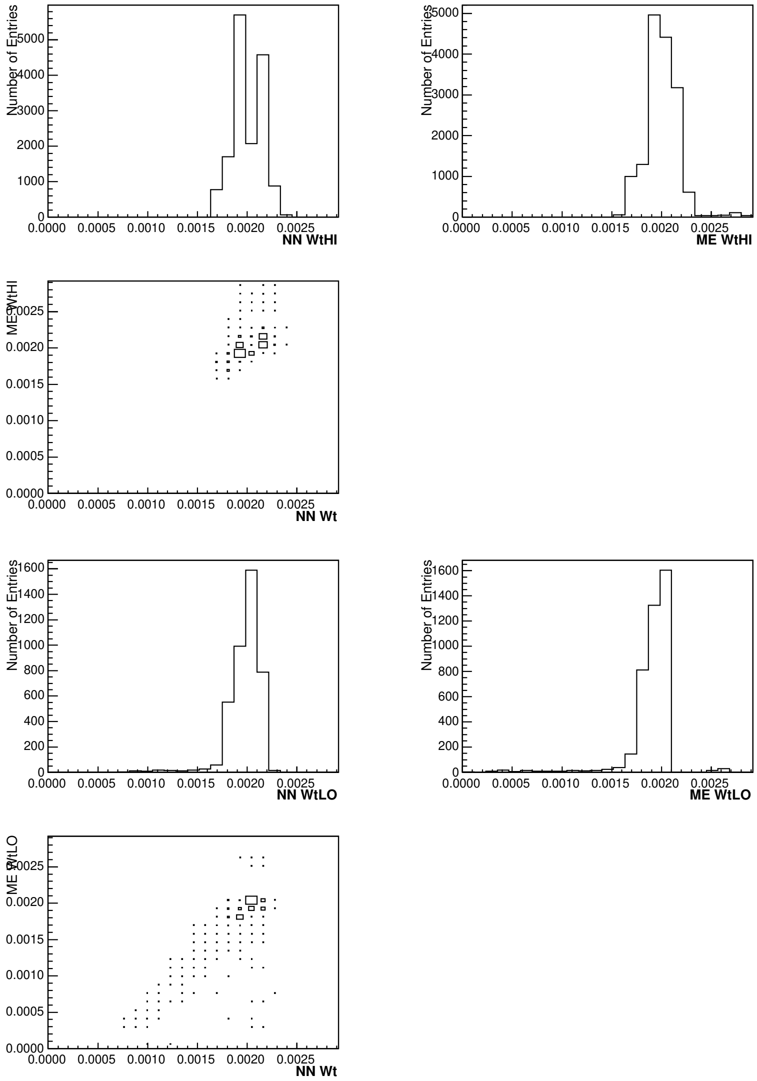

Figure 12: Weights Comparisons for WW from Pythia based on Higgs at $M_h = 160$ GeV/c^2: High,Low Signal/Background.

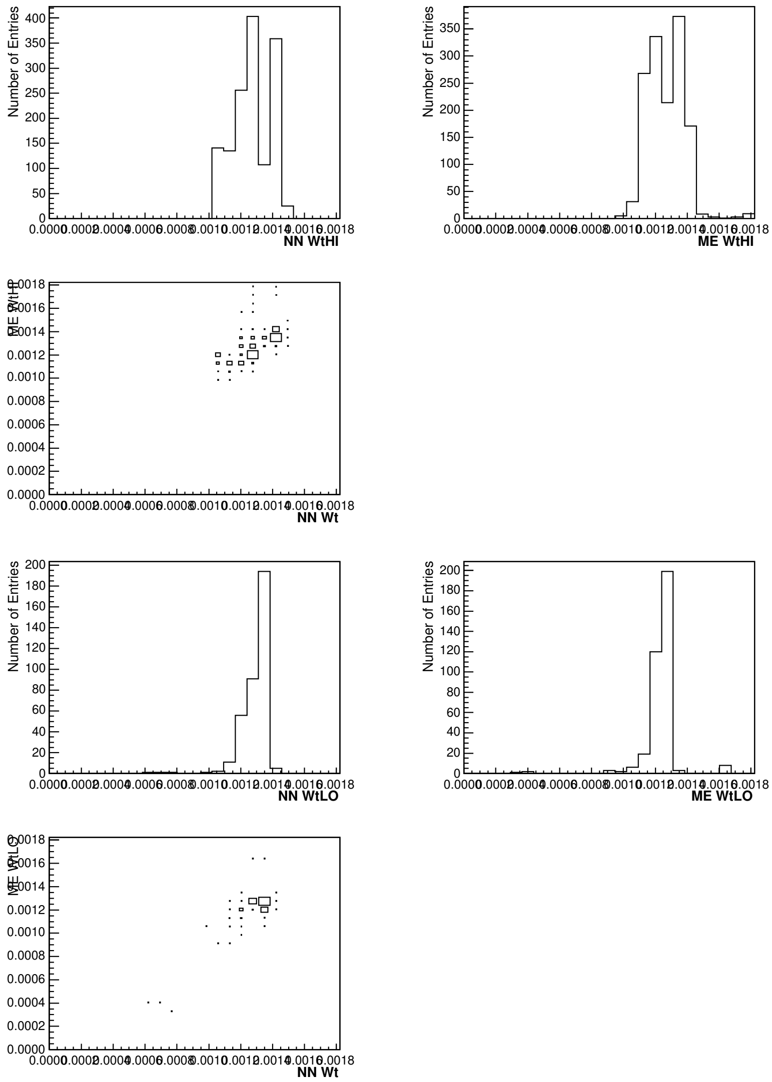

Figure 13: Weights Comparisons for WZ based on Higgs at $M_h = 160$ GeV$/c^2$: High,Low Signal/Background.

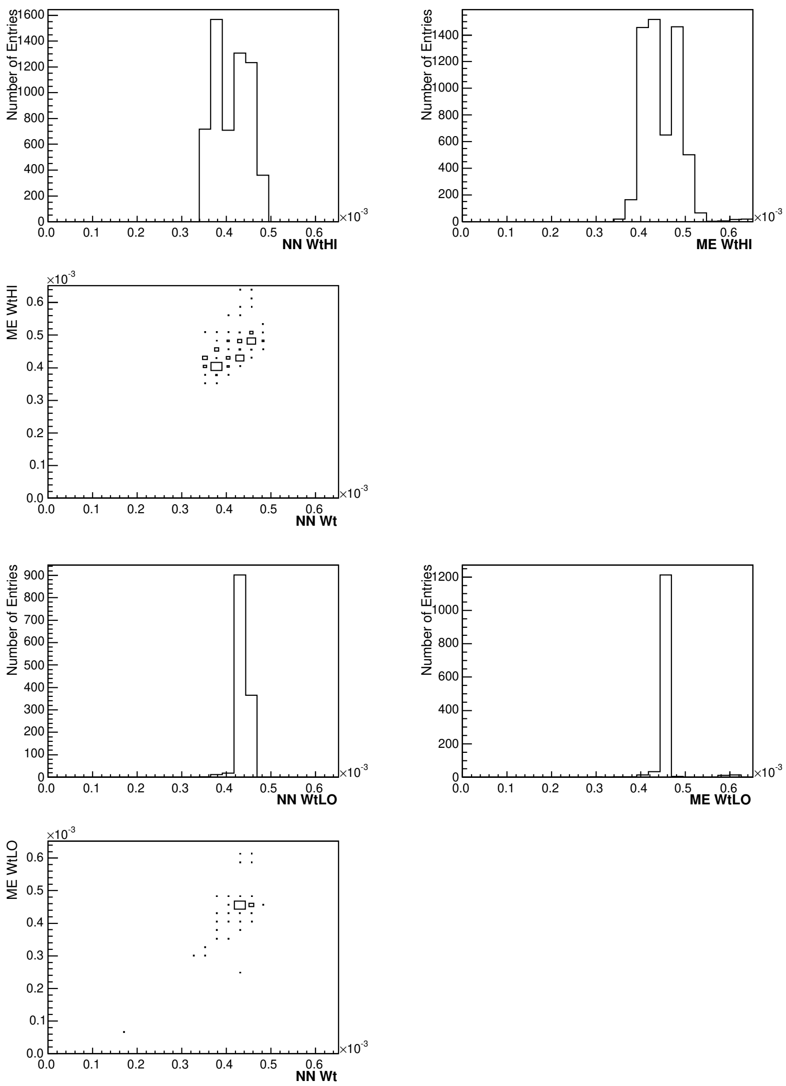

Figure 14: Weights Comparisons for ZZ based on Higgs at $M_h = 160$ GeV$/c^2$: High,Low Signal/Background.

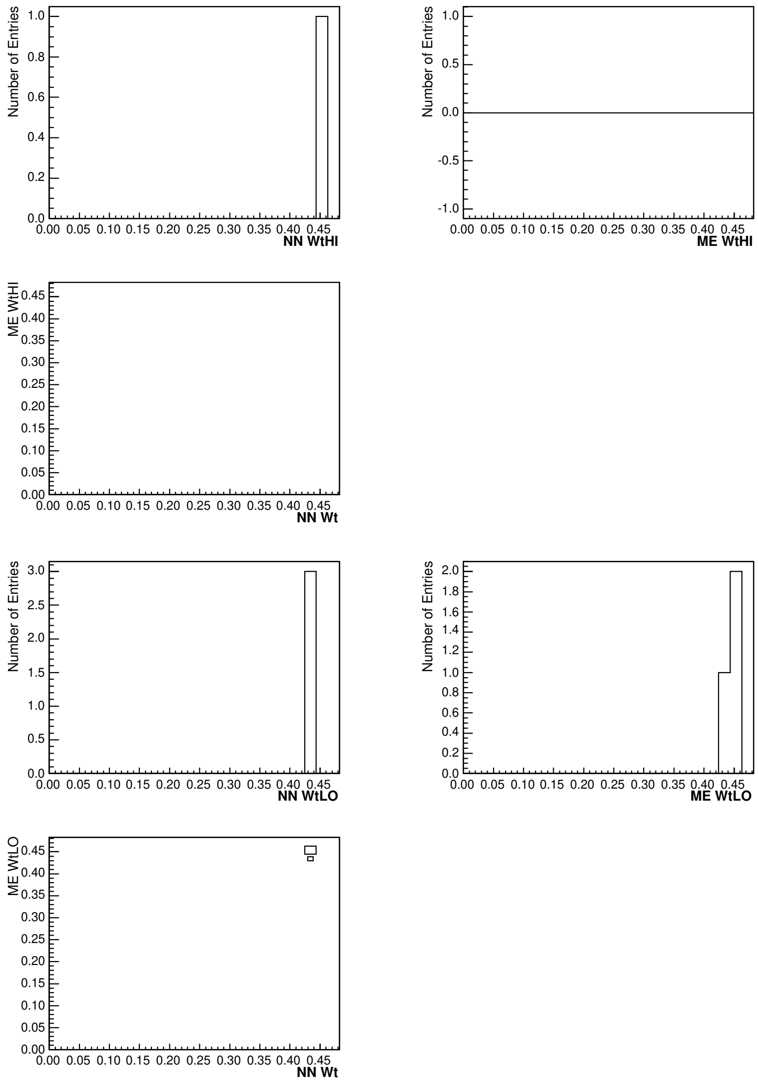

Figure 15: Weights Comparisons for $Z^0 \to e^+e^-$ based on Higgs at $M_h = 160$ GeV/c^2: High,Low Signal/Background.

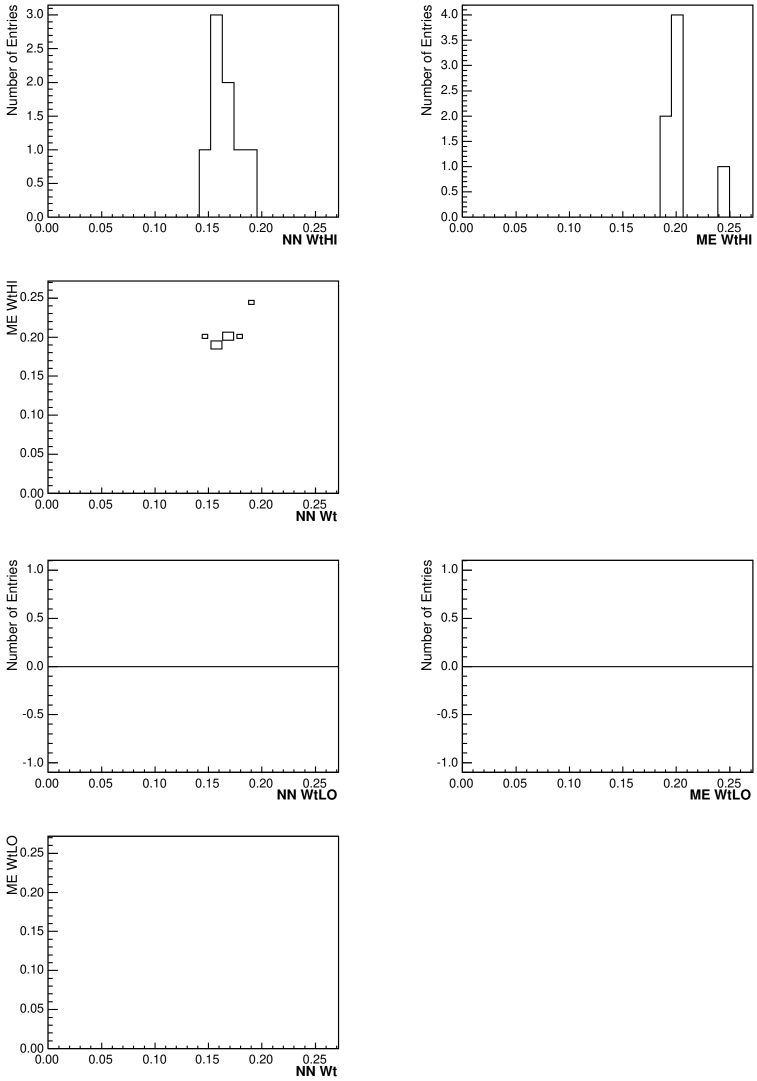

Figure 16: Weights Comparisons for $Z^0 \to \mu^+\mu^-$ based on Higgs at $M_h = 160$ GeV/c^2: High,Low Signal/Background.

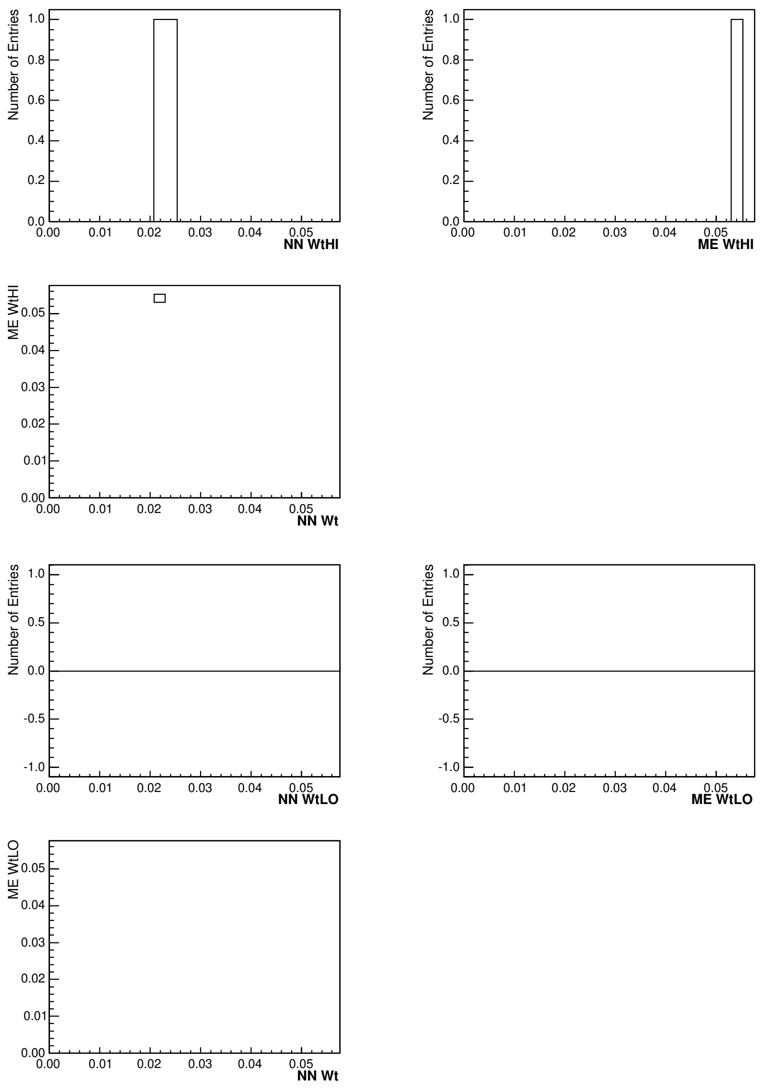

Figure 17: Weights Comparisons for $Z^0 \to \tau^+\tau^-$ based on Higgs at $M_h = 160$ GeV/c^2: High,Low Signal/Background.

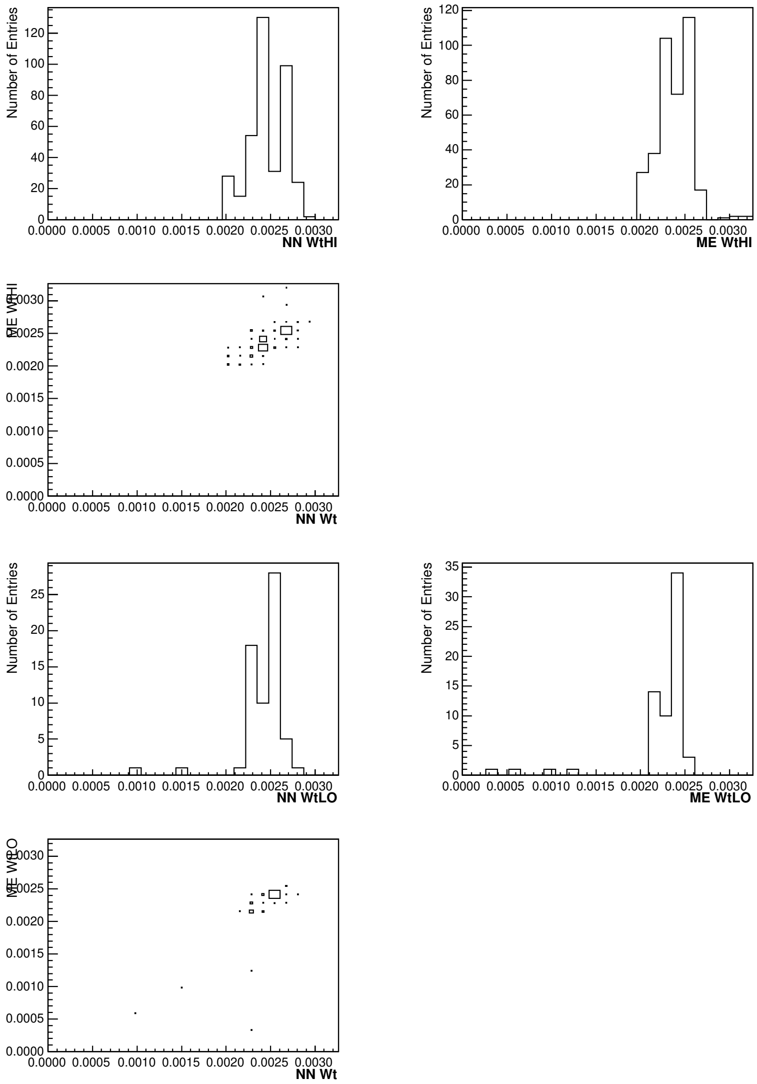

Figure 18: Weights Comparisons for $t\bar{t}$ based on Higgs at $M_h = 160$ GeV/c^2: High,Low Signal/Background.

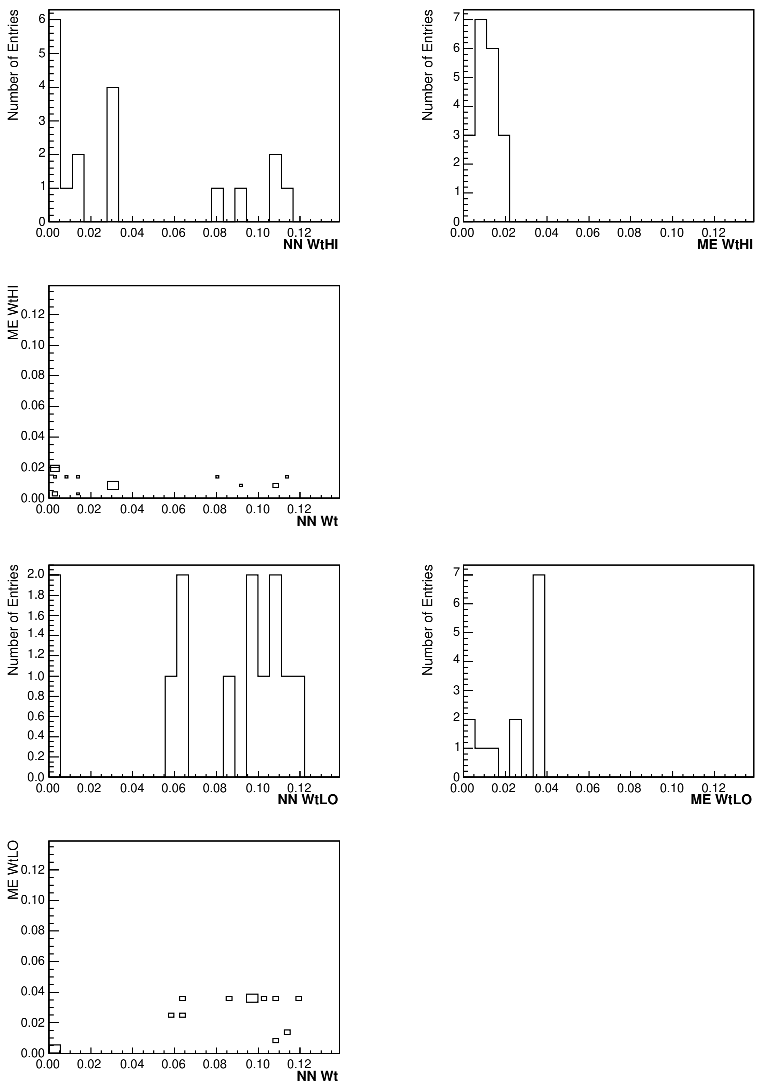

Figure 19: Weights Comparisons for Fakes based on Higgs at $M_h = 160$ GeV/c^2: High,Low Signal/Background.

References

[1] CDF 8719

[2] CDF 8700

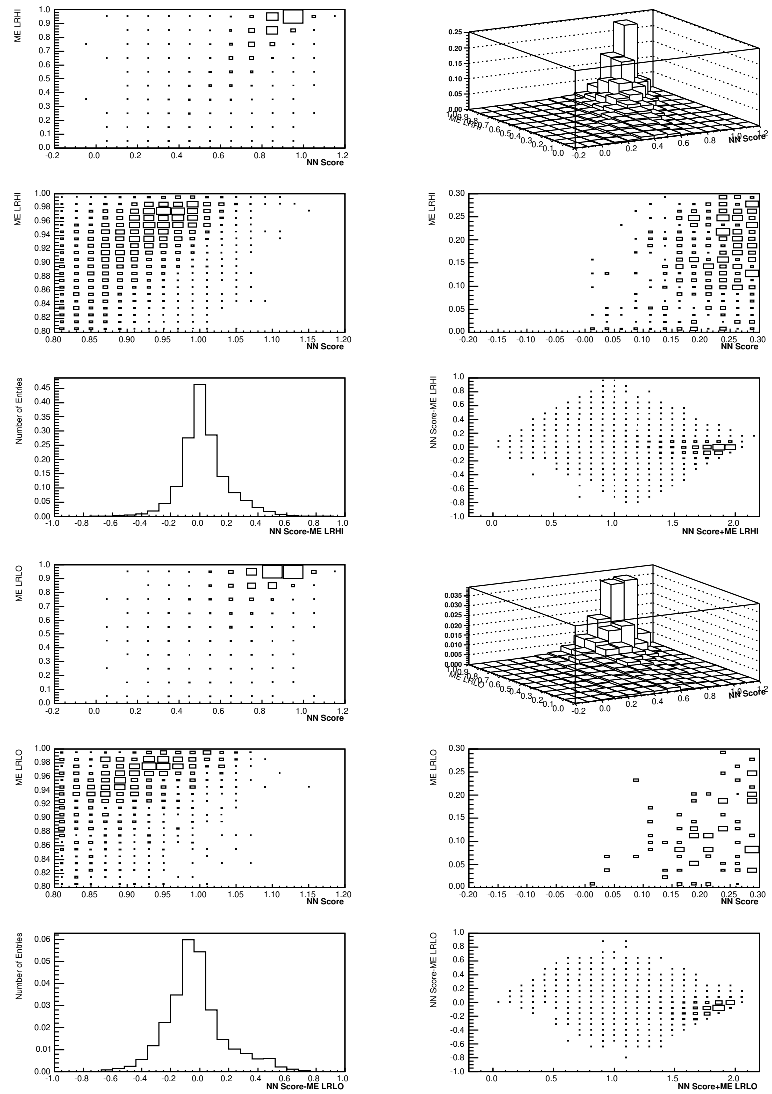

Figure 20: Weighted Score Comparisons for Higgs at $M_h = 160$ GeV$/c^2$: High,Low Signal/Background.

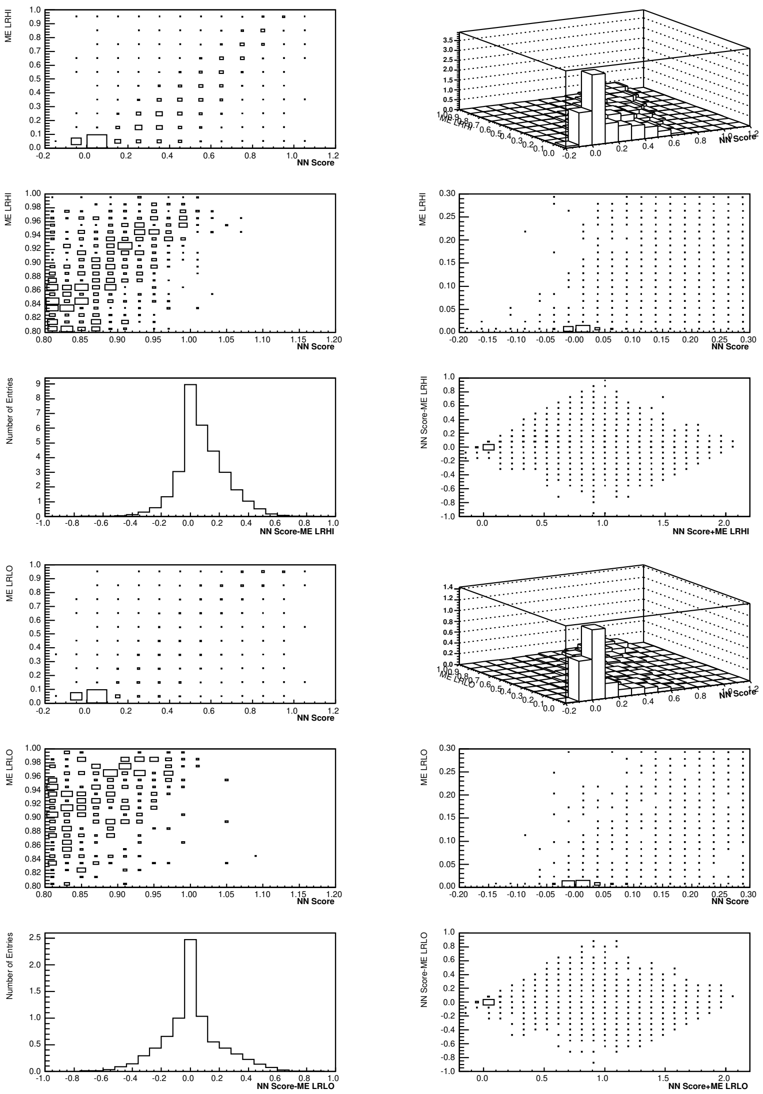

Figure 21: Weighted Score Comparisons for WW from Pythia based on Higgs at $M_h = 160$ GeV$/c^2$: High,Low Signal/Background.

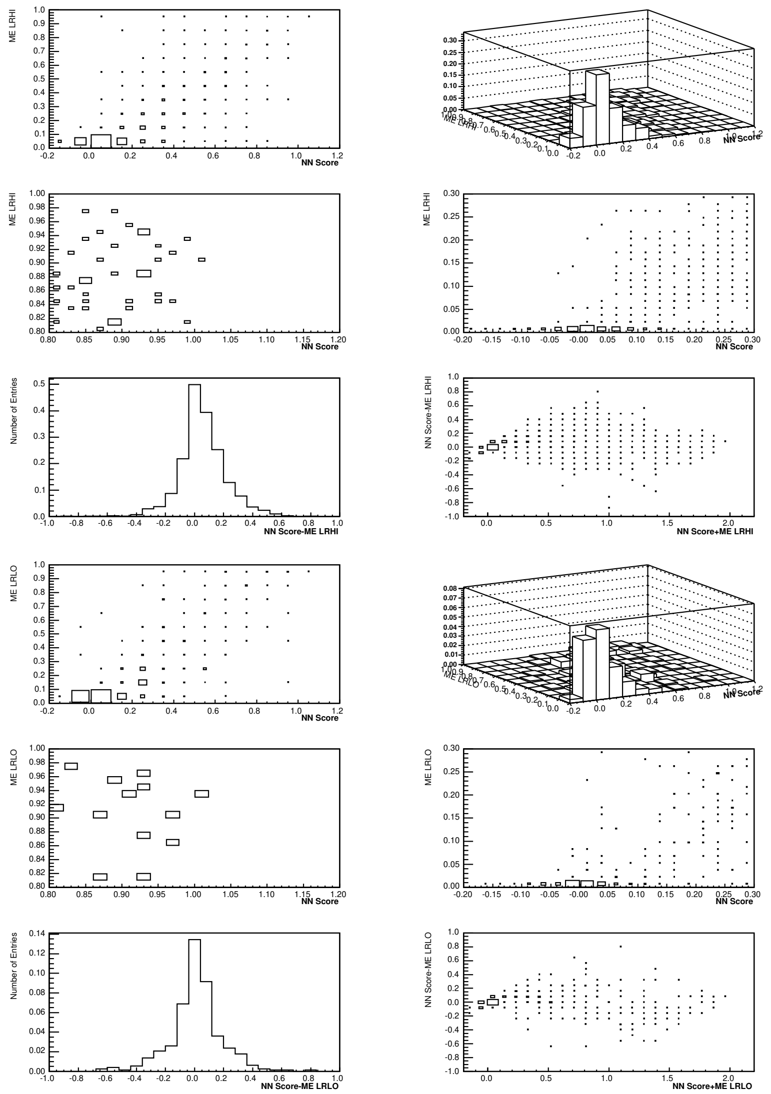

Figure 22: Weighted Score Comparisons for WZ based on Higgs at $M_h = 160$ GeV/c^2: High,Low Signal/Background.

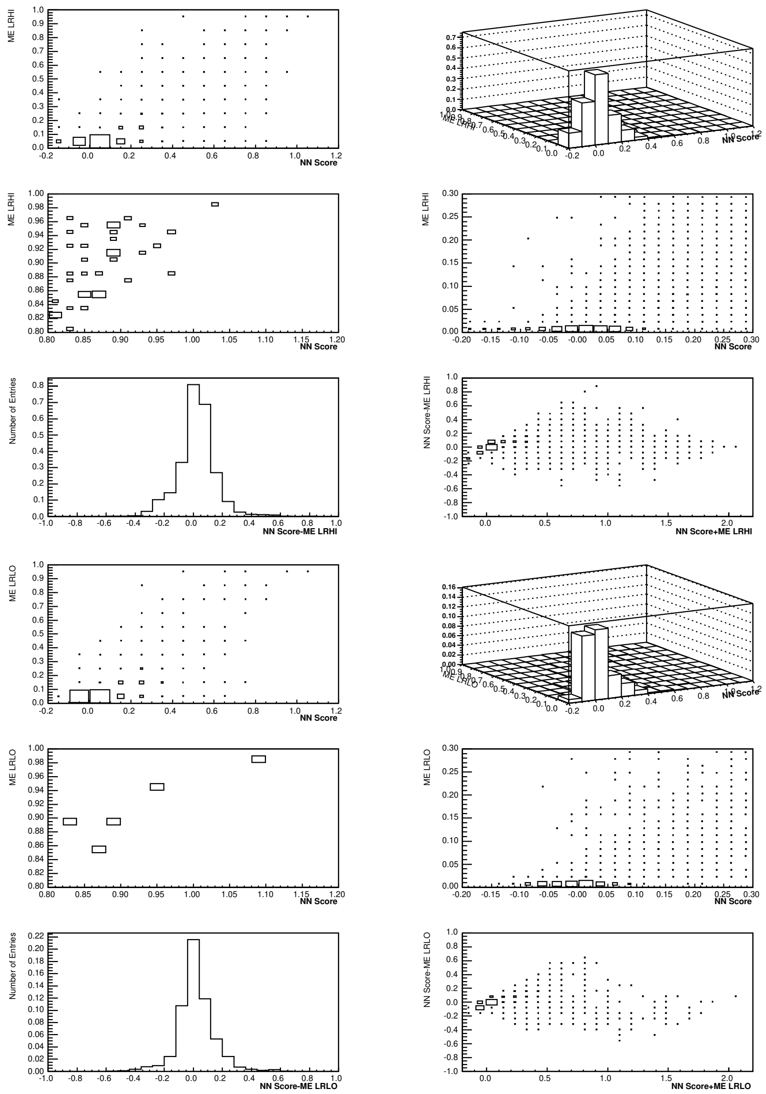

Figure 23: Weighted Score Comparisons for ZZ based on Higgs at $M_h = 160$ GeV/c^2: High,Low Signal/Background.

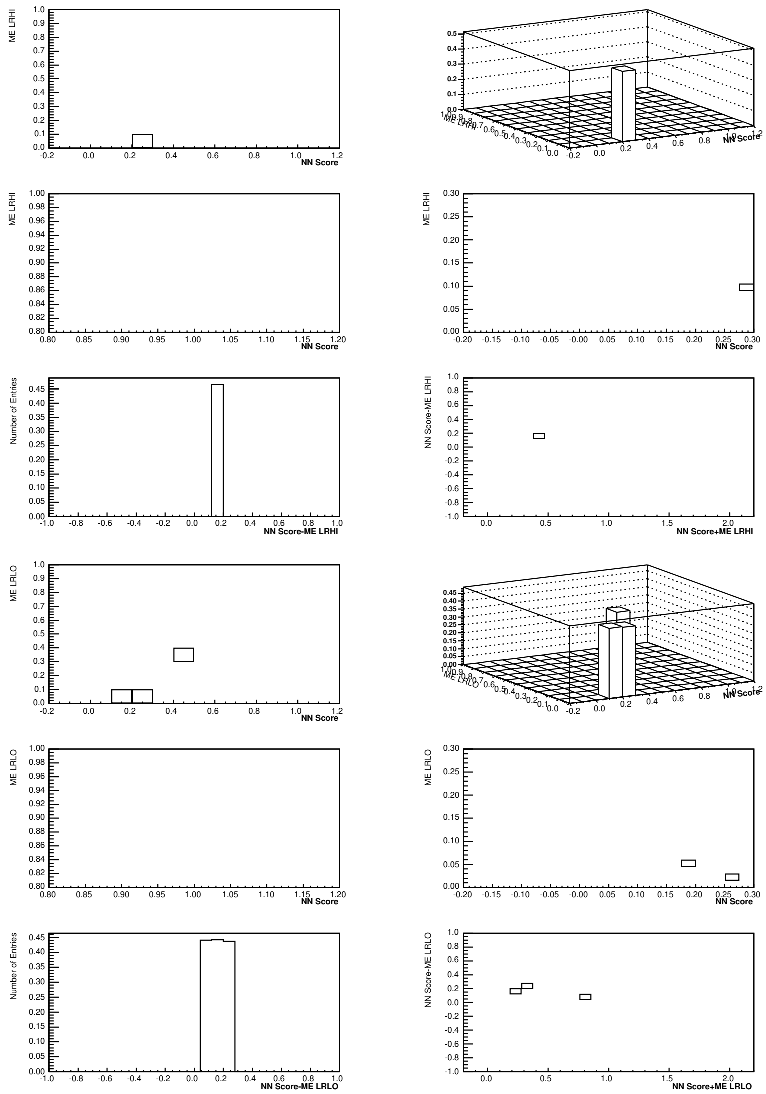

Figure 24: Weighted Score Comparisons for $Z^0 \to e^+e^-$ based on Higgs at $M_h = 160$ GeV/c^2: High,Low Signal/Background.

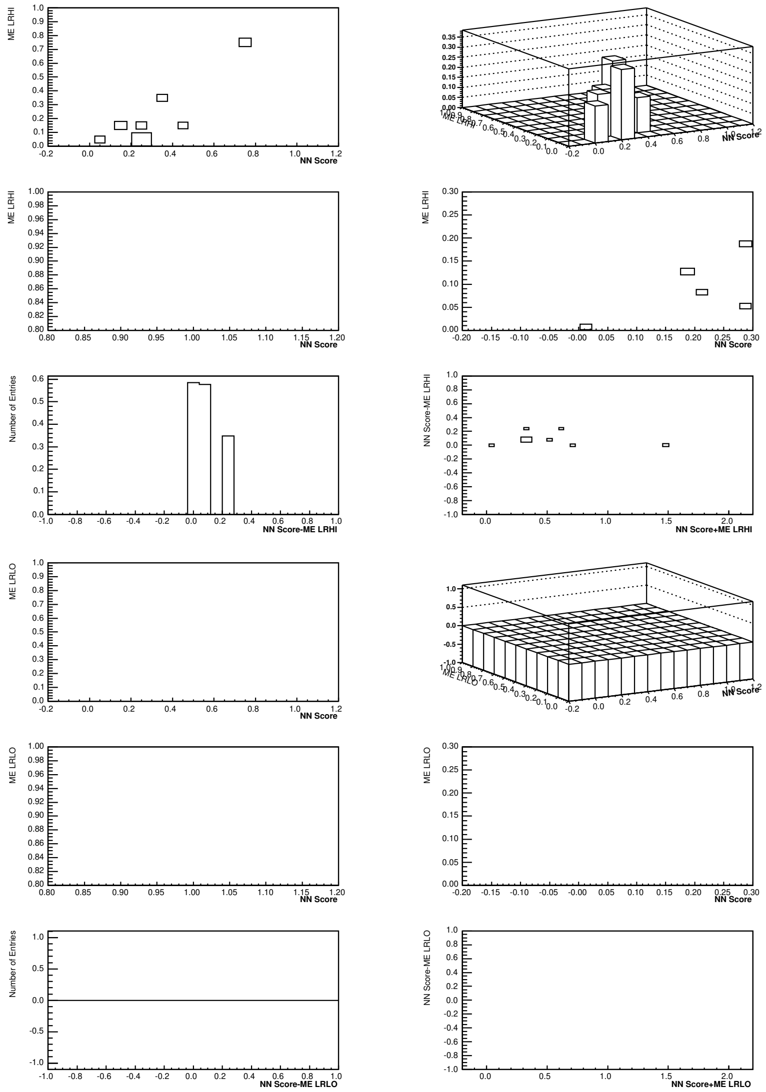

Figure 25: Weighted Score Comparisons for $Z^0 \to \mu^+\mu^-$ based on Higgs at $M_h = 160$ GeV$/c^2$: High,Low Signal/Background.

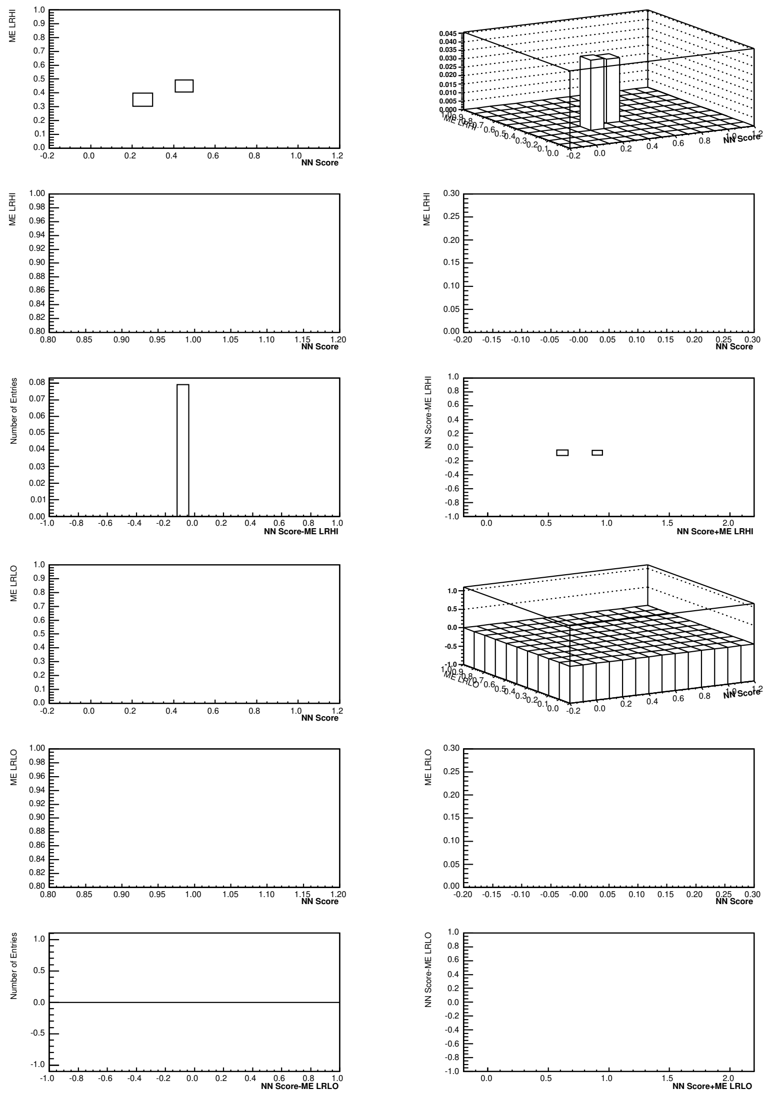

Figure 26: Weighted Score Comparisons for $Z^0 \to \tau^+\tau^-$ based on Higgs at $M_h = 160$ GeV$/c^2$: High,Low Signal/Background.

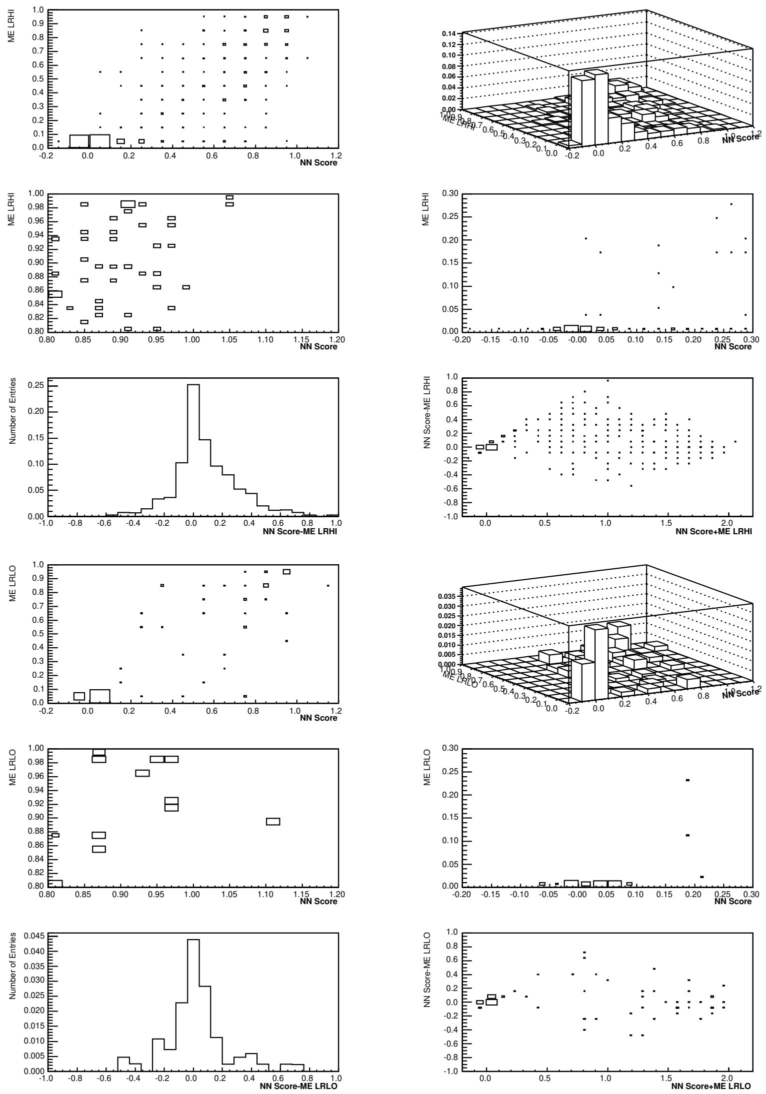

Figure 27: Weighted Score Comparisons for $t\bar{t}$ based on Higgs at $M_h = 160$ GeV/c^2: High,Low Signal/Background.

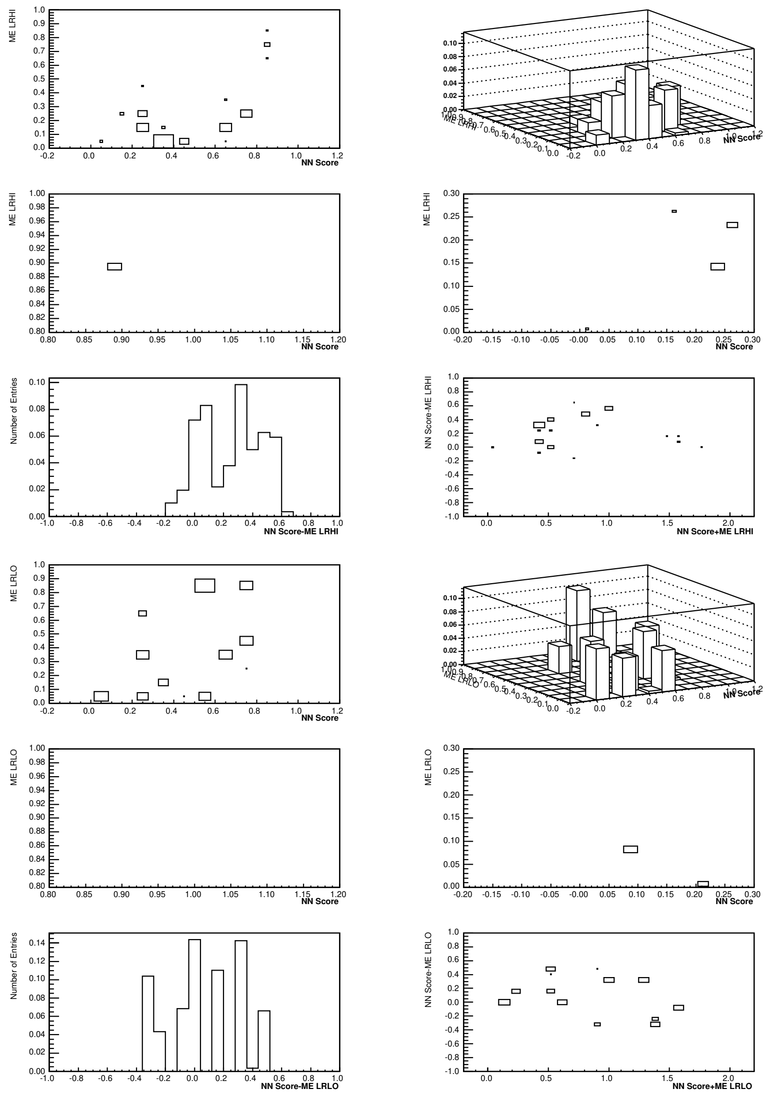

Figure 28: Weighted Score Comparisons for Fakes based on Higgs at $M_h = 160$ GeV$/c^2$: High,Low Signal/Background.