Techniques avancées dans l'analyse WH

Jonathan Brown
(LPNHE, Paris)

DØ France (Grenoble)
01/042009

Outline

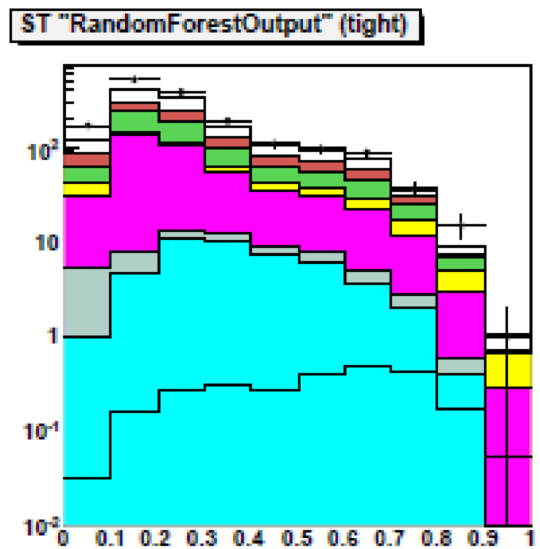
- Testing a new NN structure, trying to discriminate all backgrounds from WH signal,
- M_H = 115 GeV, 2 jets bin,
- Previous results using WH vs. Wbb NN & WH vs. ttbar NN,
- Impact on limit calculation,
- This study has been done with p17 muon channel.
- Use of Matrix Element discriminant (developped by Michigan University) is not used here,
- BDT & Random Forest for WH is recently under development.

Random Forest

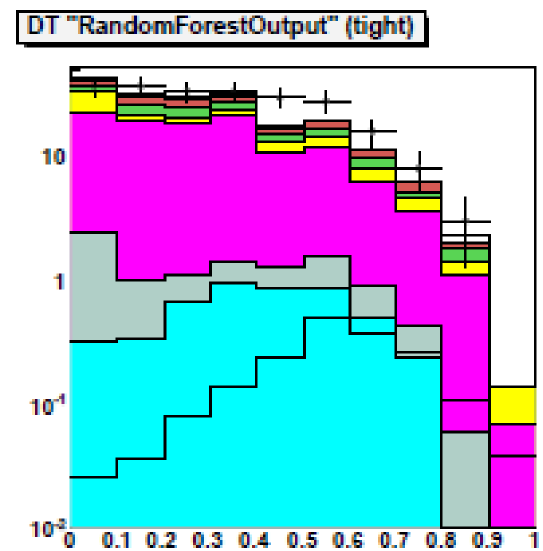
Trained with:

- 7 input variables (jet1_pt, jet2_pt, DRjj, Dphijj, ptjj, ellnu_pt, mjj)
- p20 muon samples

ST



DT



CS x BR / SM Exp (Obs) limits CLFAST(no systematics) : 9.14 (6.90)

WORK IN PROGRESS

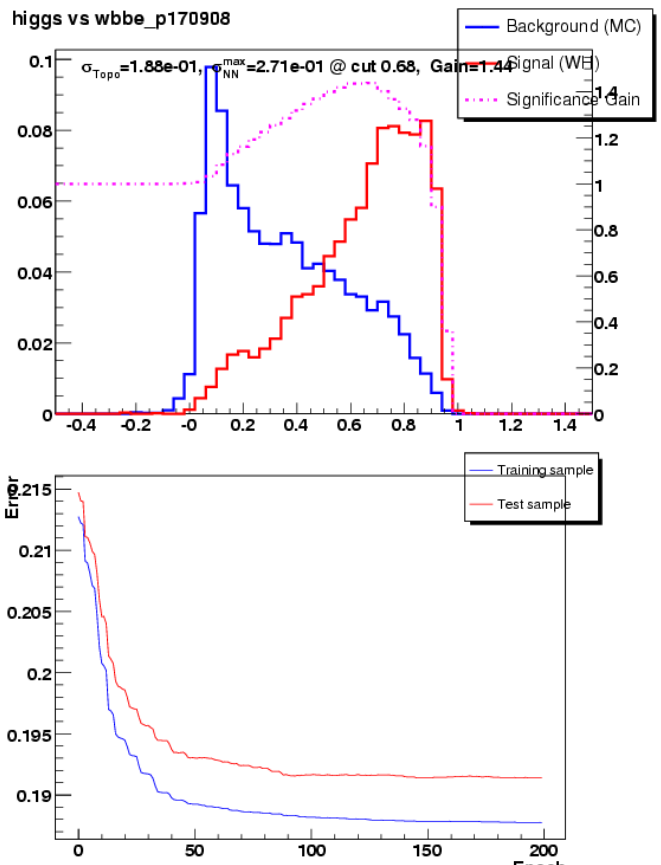
Tests with SPR and TMVA are performed in parallel by
Ken Herner & Hatim Hegab

Training of NN Wbb vs. WH

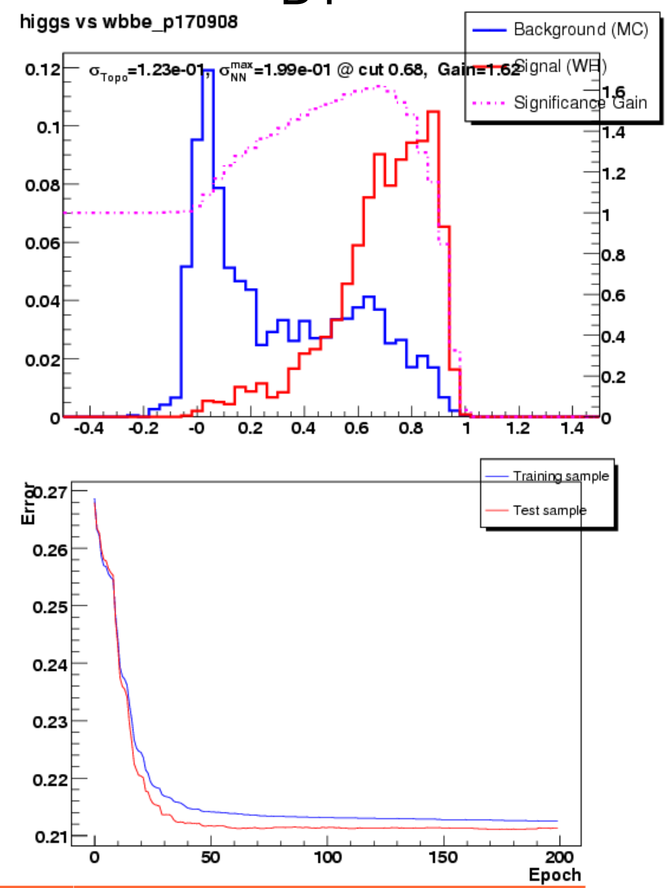
Trained with:

- 7 input variables (jet1_pt, jet2_pt, DRjj, Dphijj, ptjj, ellnu_pt, mjj),
- 7 hidden nodes,
- 200 epochs,
- Publication NN
- This configuration was chosen because it provides the best limit calculation based on CLFAST

ST

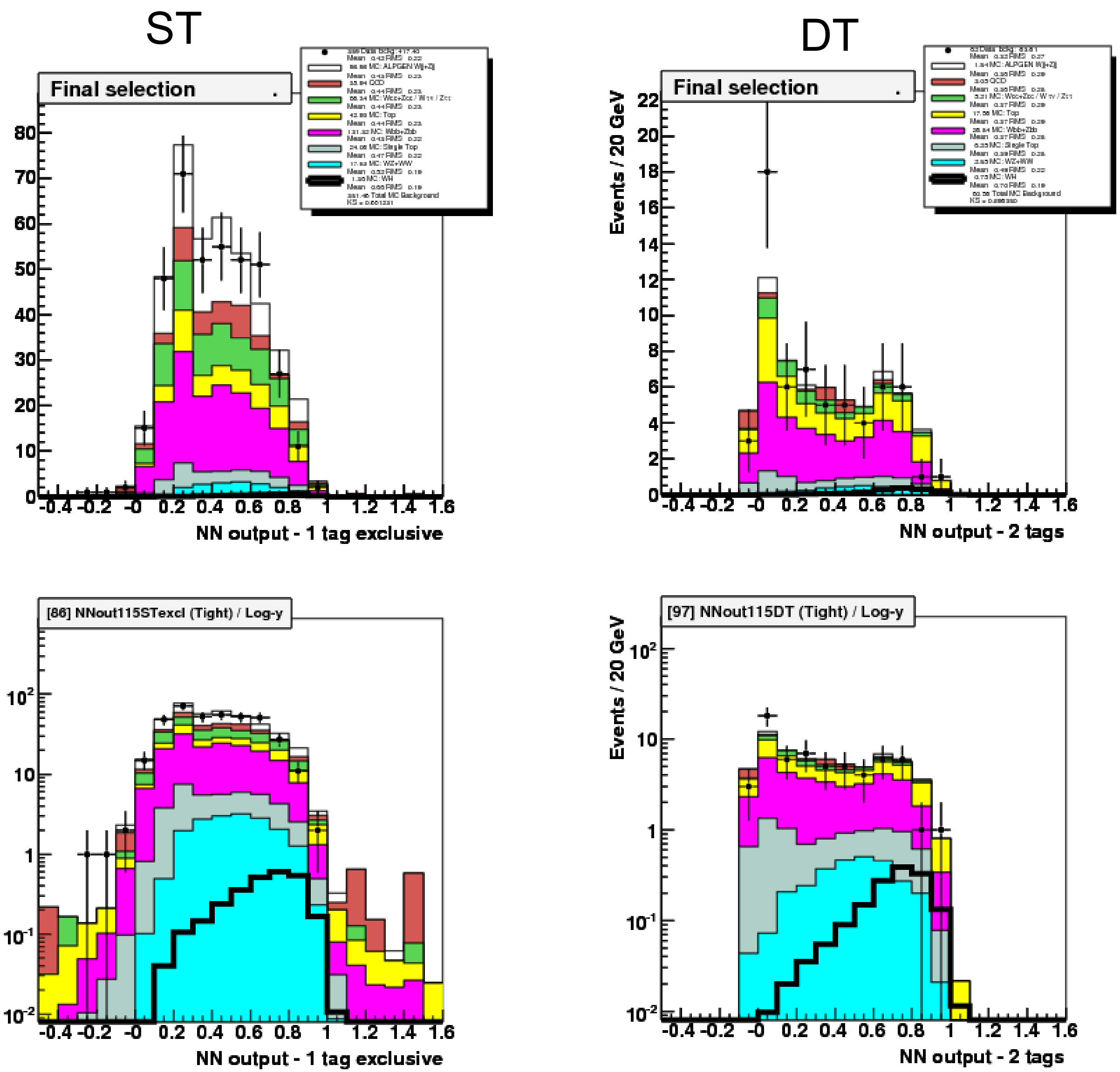


DT



	# of training events (ST)	# of training events (DT)
signal	6162	2752
background	9401	1935

NN Wbb vs. WH



Wbb
ttbar
WH

M_H = 115 GeV:

CS x BR / SM (Exp limits CLFIT2)	
Dijet Mass	17.5
NN	15.7

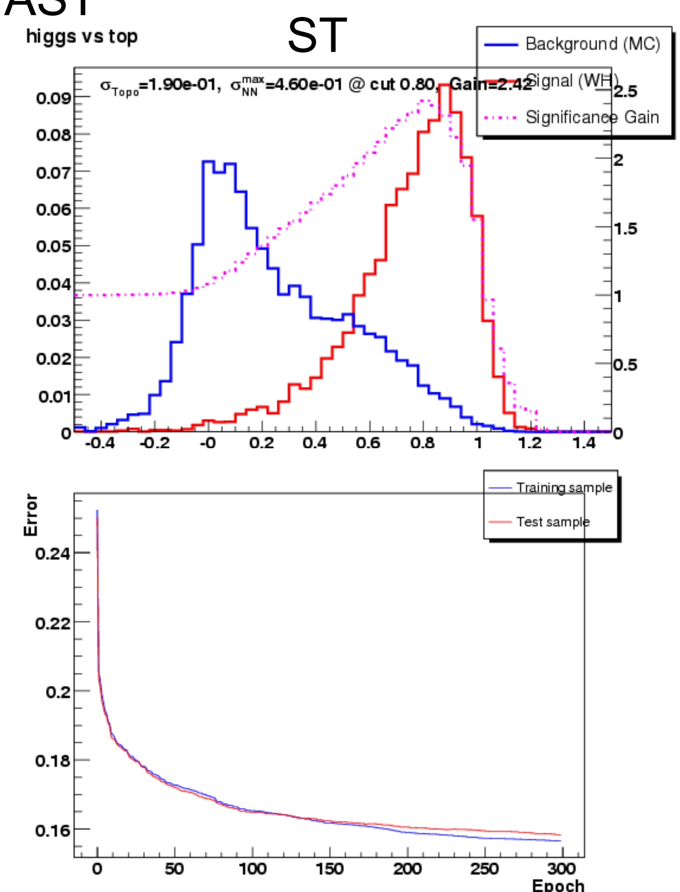
5% improvement by adding the ME as an input variable

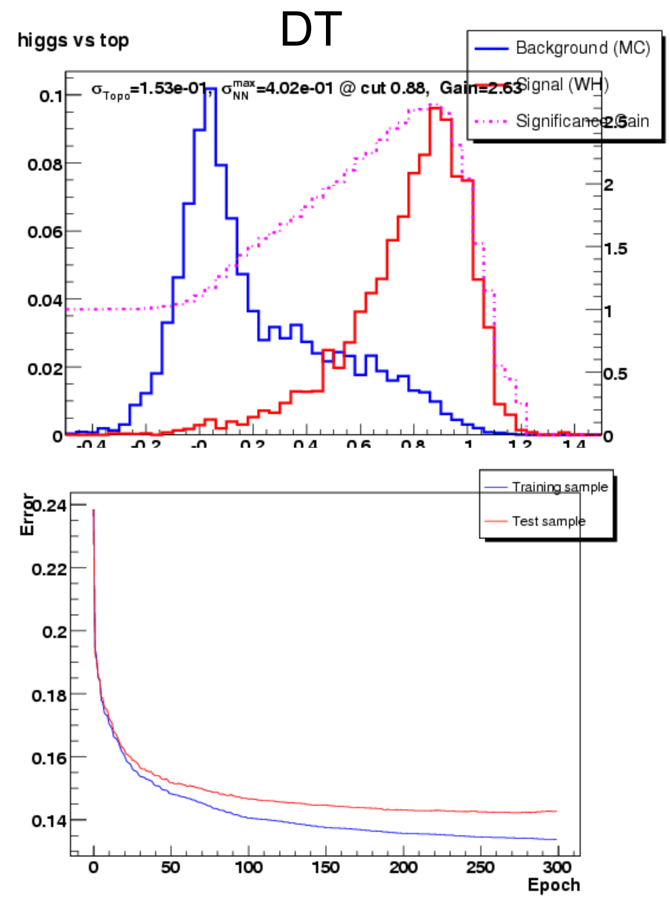
- Good agreement between Data & MC

Training of NN ttbar vs. WH

Trained with:

- 17 input variables (Hz, mjj, DRjj, Aplan_jetslep, Dphi_lepjet1, jet1_pt, jet2_pt, ellnu_mt, jet2_E, lep_qeta, Etmiss, sqrts1, sqrts2, DR_ellnujj1, DR_ellnujj2, ptjj, ellnu_pt),
- 20 hidden nodes,
- 200 (300 for ST) epochs,
- This configuration was chosen because it provides the best limit calculation based on CLFAST

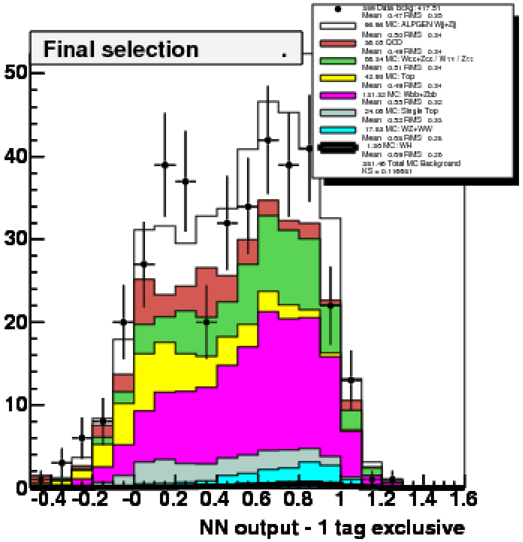




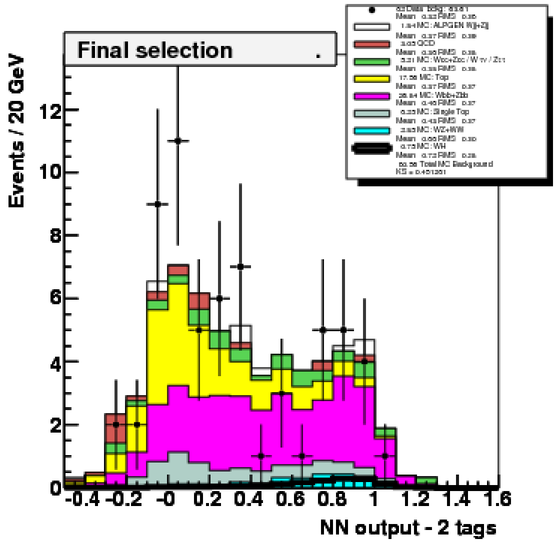
	# of training events (ST)	# of training events (DT)
signal	6162	2752
background	17168	6575

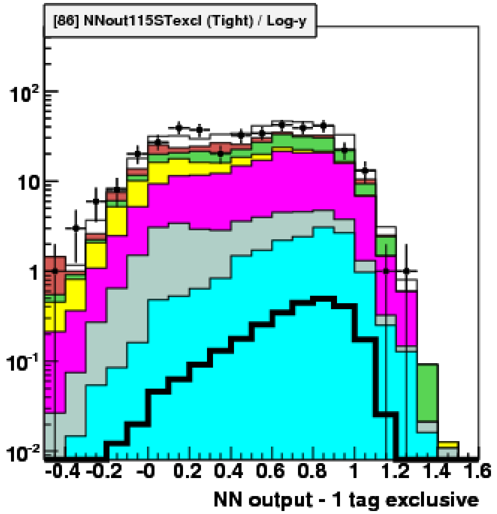
NN ttbar vs. WH

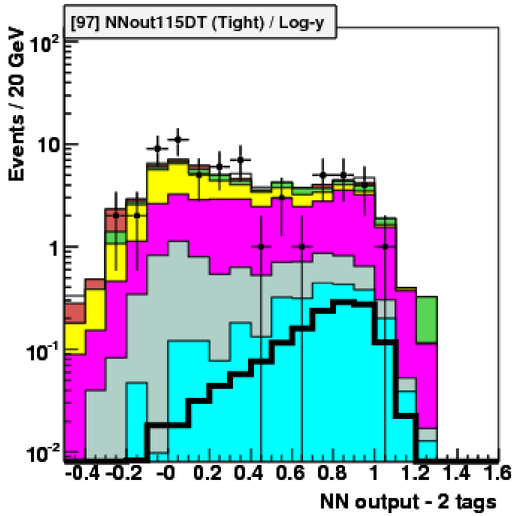
ST



DT







Wbb
ttbar
WH

M_H = 115 GeV:

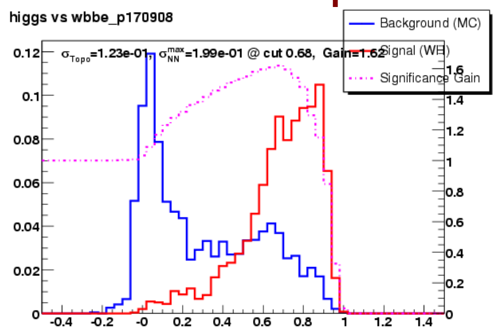
CS x BR / SM (Exp limits CLFIT2)	
Dijet Mass	17.5
NN	22.6

- Good agreement Data vs. MC but,
- ttbar is well rejected but,
- Wbb is not discriminated → no improvement on limit calculation.

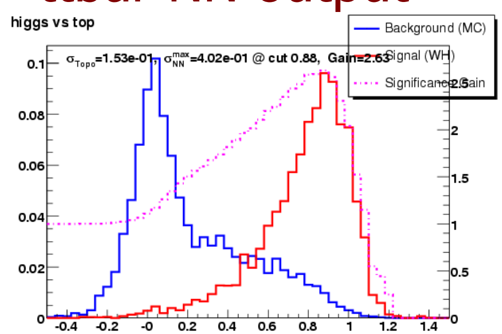
NN Wbb+ttbar vs. WH
a.k.a. "SUPERNN"

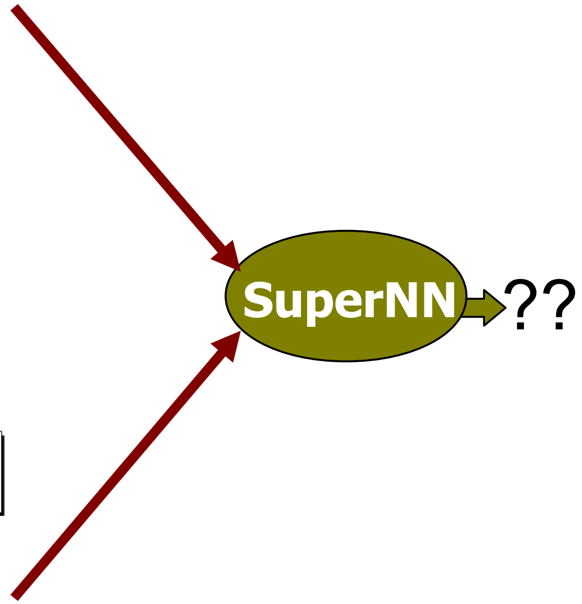
Trying to discriminate Wbb & ttbar vs. WH, outputs of the Wbb NN and ttbar NN (calculated for each event) are inputs to a new structure.

Wbb NN output



ttbar NN output



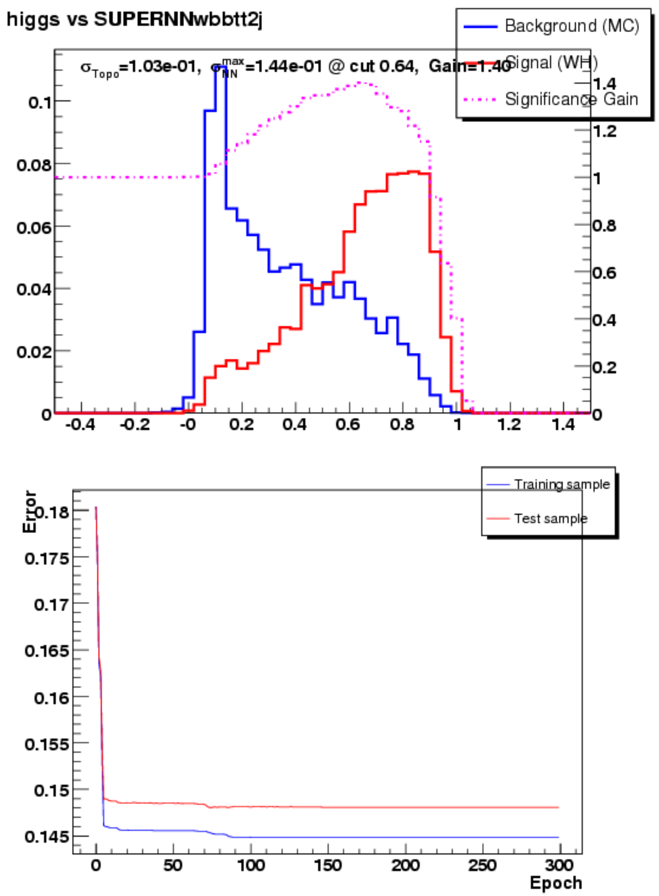


Training of SUPERNN (1)

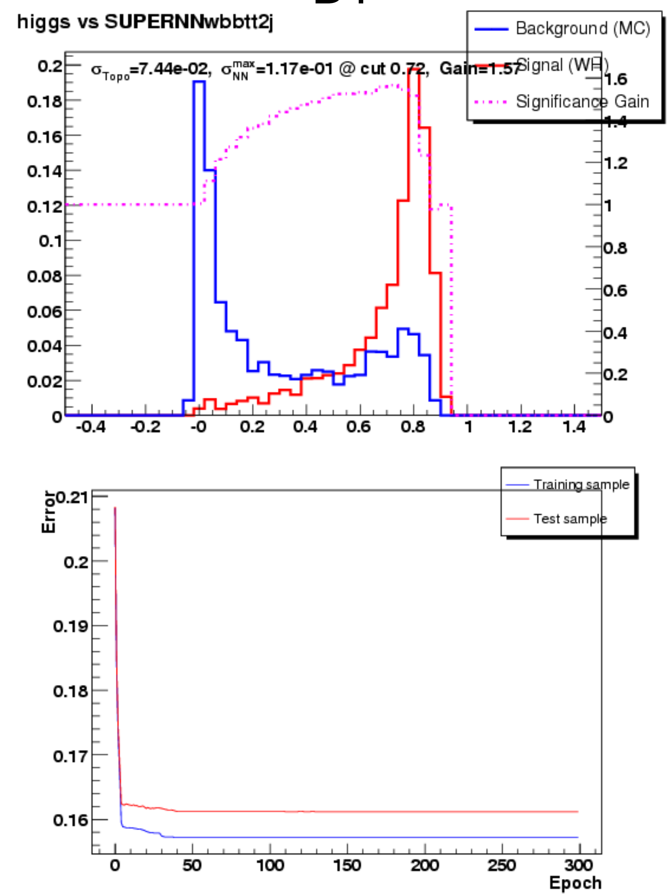
Trained with:

- 2 input variables (Wbb NN output {mass} {btag}, tt NN output {mass} {btag}),
- 2 hidden nodes,
- 300 epochs

ST



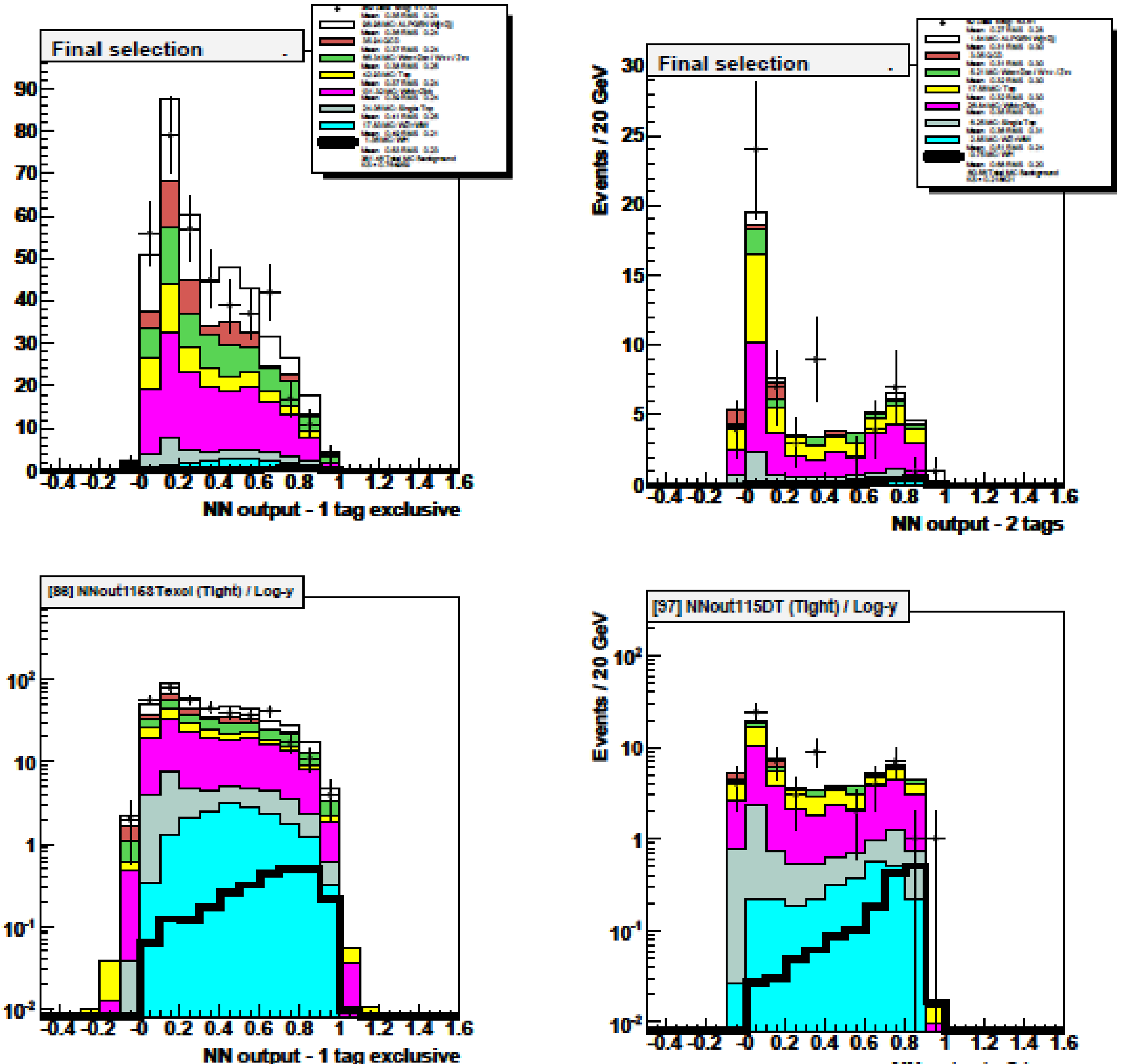
DT



	# of training events (ST)	# of training events (DT)
signal	5046	2256
background	17845	6528

SUPERNN (1)

ST DT



Wbb
ttbar
WH

M_H = 115 GeV:

CS x BR / SM (Exp limits CLFIT2)	
Dijet Mass	17.5
Wbb NN	15.7
SuperNN	15.4

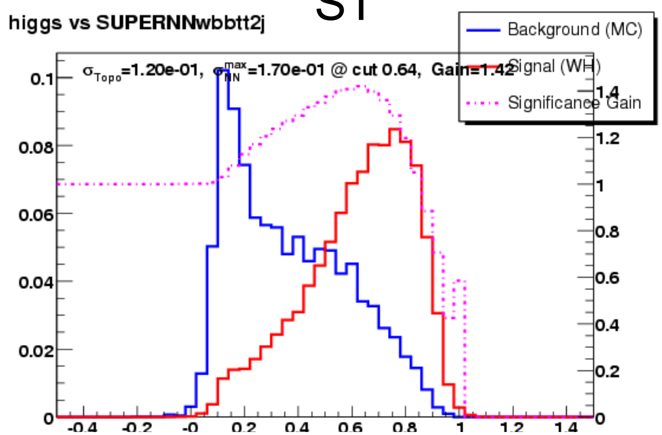
- Agreement OK,
- Pretty nice shape for background distributions,
- Minimal improvement on the limit

Training of SUPERNN (2)

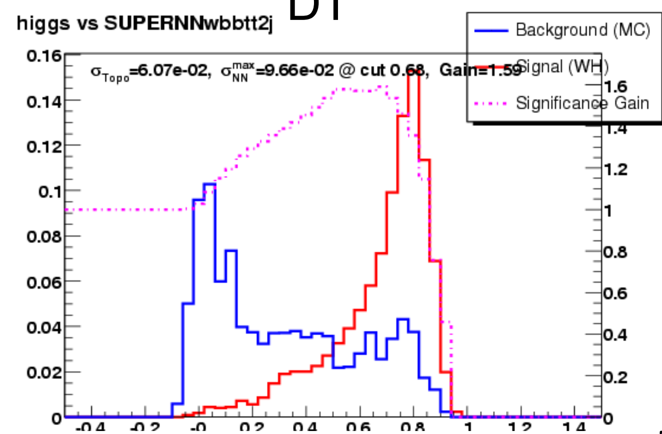
Trained with:

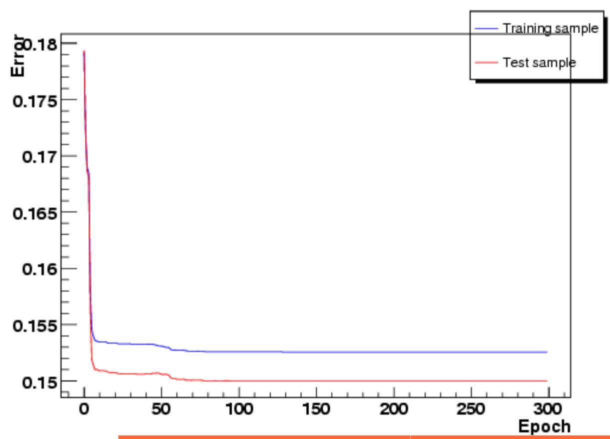
- 2 input variables (Wbb NN output {mass} {btag}, tt NN output {mass} {btag}),
- 2 hidden nodes,
- 300 epochs,
- 4 mass points for WH are used in the training for M_H = 115 GeV: WH105, WH110, WH120 & WH125

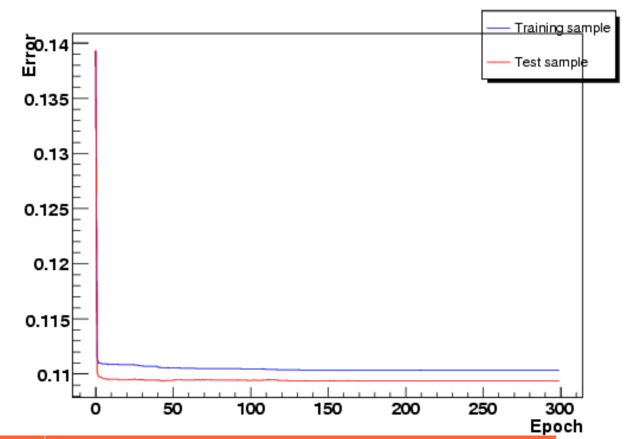
ST



DT





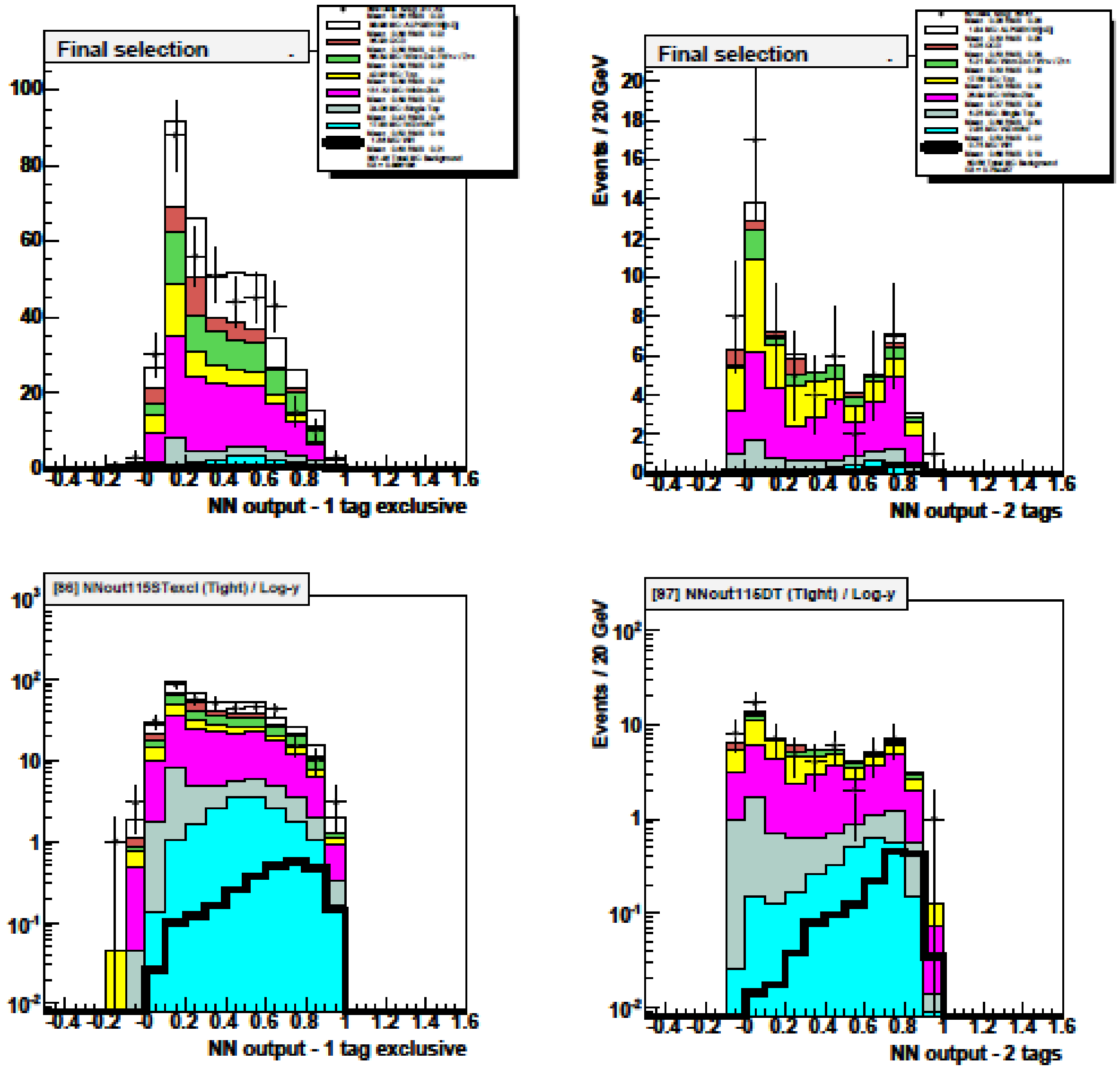


More statistics for the training & for the analysis as M_H = 115 GeV is not used In the training

	# of training events (ST)	# of training events (DT)
signal	28149	12598
background	17845	6528

SUPERNN (2)

ST DT



Wbb
ttbar
WH

M_H = 115 GeV:

CS x BR / SM (Exp limits CLFIT2)	
Dijet Mass	17.5
Wbb NN	15.7
SuperNN	14.4

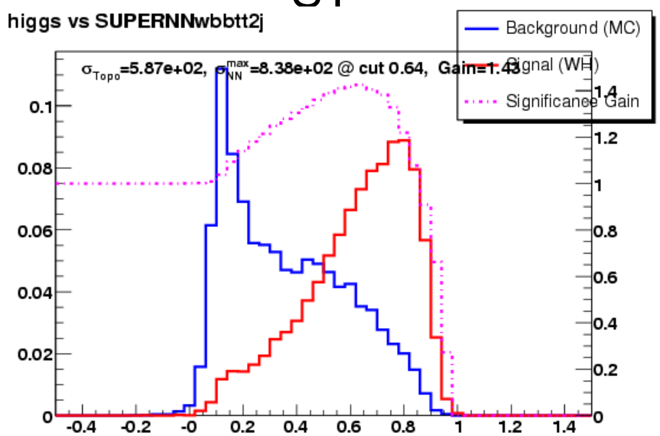
- Agreement OK,
- Pretty nice shape for background distributions,
- Improvement on the limit (~10% compared to Wbb NN only)

Training of SUPERNN (2)

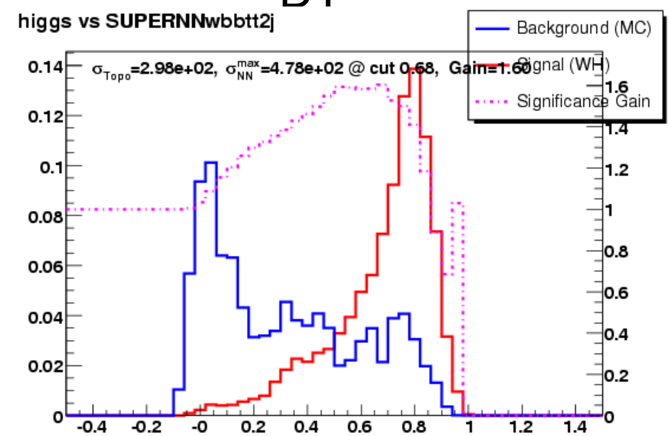
Trained with:

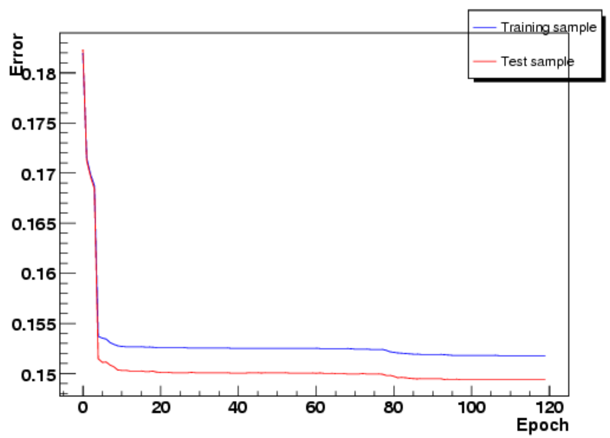
- Same as before +
- Weights for WH samples in training are set to 1 so as not to bias towards low masses cross section ($\sigma \uparrow$ when $M_H \downarrow$)

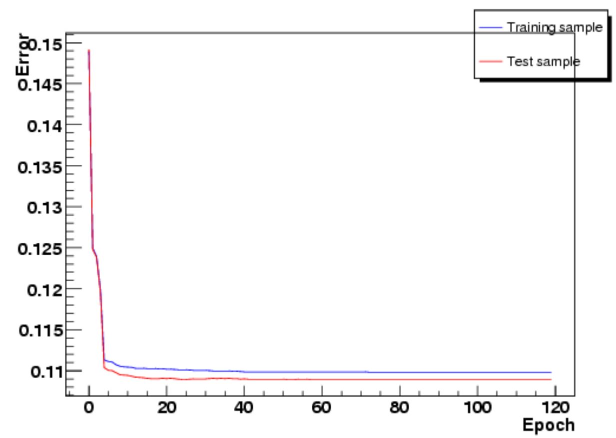
ST



DT



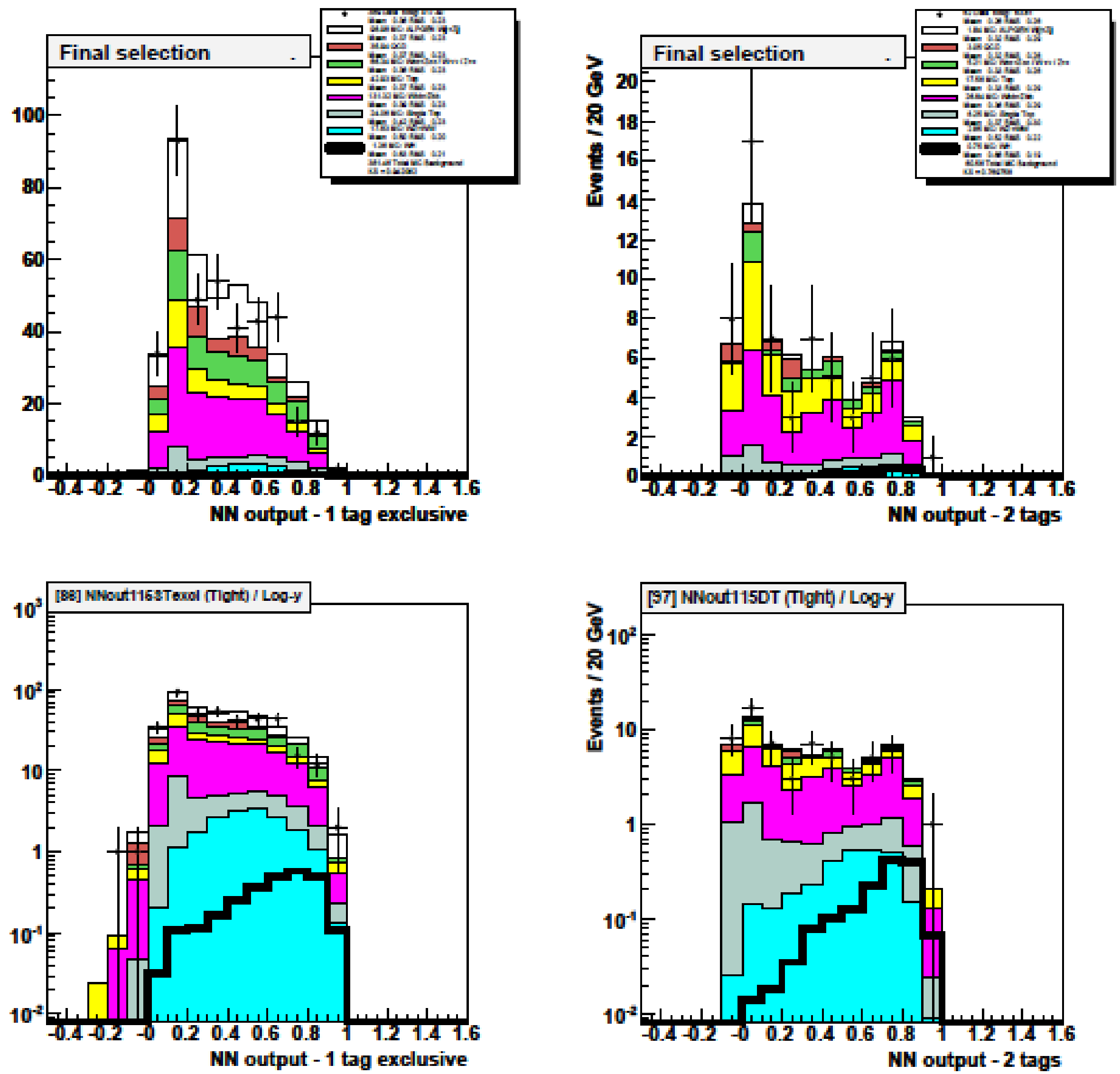




	# of training events (ST)	# of training events (DT)
signal	28149	12598
background	17845	6528

SUPERNN (2)

ST DT



Wbb
ttbar
WH

M_H = 115 GeV:

CS x BR / SM (Exp limits CLFIT2)	
Dijet Mass	17.5
Wbb NN	15.7
SuperNN	14.4

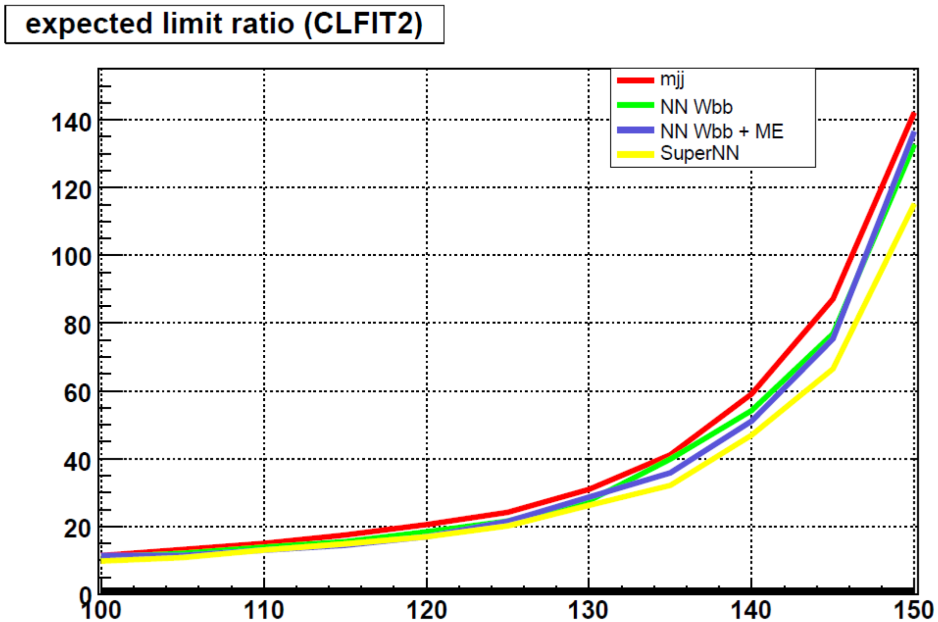
- Agreement OK,
- Pretty nice shape for background distributions,
- Improvement on limit (~10% compared to Wbb NN only), no changes wrt SuperNN1

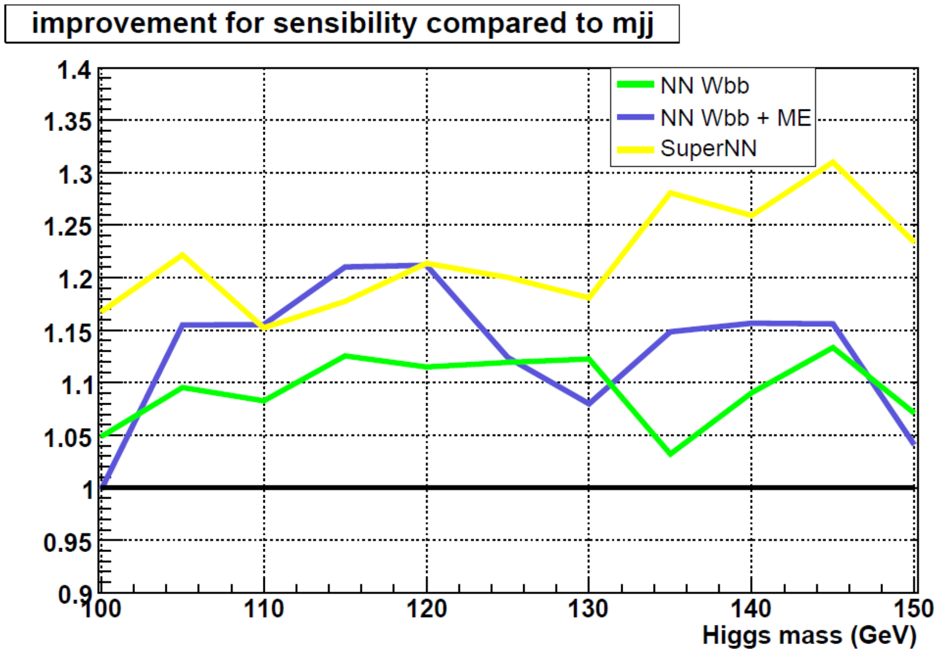
Summary

	CS x BR / SM (CLFIT2)	
	expected	observed
Dijet mass	17.5	12.4
NN Wbb	15.6	12.7
NN tt	22.6	15.7
SuperNN (1)	15.4	13.1
SuperNN (2)	14.4	10.4

~25% compared to mjj
~15% compared to Wbb NN

Expected limits with systematics (CLFIT2)





<u>SuperNN:</u>

~25% compared to mjj

~15% compared to Wbb NN

Conclusion

- Gain obtained by this new structure is ~25% compared to mjj, ~15% compared to Wbb NN, *without adding the Matrix Element yet,*

- Next steps :
  - Include the Matrix Element discriminant as an input to the SuperNN,
  - Propagate this tool to p17 electron and p20 (mu + ele),

- Many configurations can be tested once we'll trying to discriminate all backgrounds at the same time against WH,
- Will be tried for the 3 jets sample,
- An alternative from making a 2D histogram ttNN vs. WbbNN,
- Combine NNs, BDT & Matrix Element discrimant