Reconstruction of the ttH(bb) single lepton channel using multivariate technique

Royer Edson Ticse Torres

Supervisor: Arnaud DUPERRIN

November 25, 2015







Royer Edson Ticse Torres,

PhD subject :

Search of the Higgs boson in the $t\bar{t}H(H \rightarrow b\bar{b})$ channel and identification of jets containing two B hadrons with the ATLAS experiment.

1st year

 "Tagging jets with double B-hadron using multiple secondary vertices"

CPPM - Séminaire étudiants (fin de la première année), 01/12/2014

2nd year

 "Reconstruction of the tīH(H → bb̄) single lepton channel using multivariate technique."

tTH Motivation

- ▶ Top quark coupling to the Higgs boson is the most strongest coupling in the SM (Y_t ~ 1).
- Indirect constrain on the top Yukawa coupling Y_t Current measurements of Higgs boson productions via gluon fusion are consistent with SM within experimental uncertainties.
- Direct measurement of Y_t in ttH production A measurement of the rate of ttH production provides a direct test of the coupling.

Observation of a significant deviation in the ttH production rate would be an indirect indication of unknown phenomena.



Single lepton $t\bar{t}H(H \rightarrow b\bar{b})$

Semileptonic t t H(bb) signal produces 6 jets, 4 of them b-jets and 1 leptons and 1 neutrino.



 Background processes to consider: tt+light, tt+cc, tt+bb, tt+V, W+jets, Z+jets, Diboson, Multijets.

Pairing problem:

Jets cannot be associated uniquely to the partons of the hard-scattering



To resolve the problem:

- Likelihood method
- KLFitter method
- BDT method

Run 1 : Analysis strategy

 Analysis channels defined splitting in jet and b-tag multiplicities:



- Signal-rich regions : Neural Network trained to separate tt
 from tt
 +jets
 - Event kinematics: H_T
 - Object kinematics: p_T^{jet5}
 - Object pair properties: ΔR_{bb}^{avg}
 - Event shape: centrality
 - Matrix element method.



The reconstruction of the ttH(bb) system was not used in the Run1

Run 2 : MVA approach in signal regions

Reconstruction

A BDT is used to find the good combination.

Signal discrimination from background.

• A BDT to discriminate $t\bar{t}H(b\bar{b})$ from $t\bar{t}+b\bar{b}$.



$t\bar{t}H(b\bar{b})$ MVA reconstruction

Aim: reconstruct the ttH(bb) system finding good discriminant variables to reject ttbb.

- Reconstruct the full ttH(bb) system using all combinations of jets.
- We use a BDT to find the good combination in the signal region: ≥6j ≥4b, ≥6j =3b and =5j ≥4b.
- We are training 2 different BDTs:
 - recoBDT : using variables uncorrelated with Higgs.
 - recoBDT_withH : adding variables: Higgs mass and dR(b,b) from Higgs.

 Signal : full good ttH(bb) combination
Bkg : all other different jets combinations.



Reconstruction in $\geq 6j \geq 4b$ - recoBDT

 Combination of jets with the highest recoBDT output is chosen as the good combination.





Reconstruction efficiency



recoBDT

recoBDT_withHiggs



blt bht wj1 wj2

matched objects

Discriminating variables from reconstruction : $\geq 6j \geq 4b$

- The highest recoBDT_withH output per event
- Higgs mass using combination from recoBDT
- ΔR(b,b) from Higgs using combination from recoBDT





BDT to discriminate $t\bar{t}H$ from $t\bar{t} : \ge 6j \ge 4b$

- Classification BDT trained with the same variables used in run 1 and adding variables from reconstruction.
- \blacktriangleright Adding variables from the reconstruction we can increase the separation in ${\sim}17\%$



BDT to discriminate $t\bar{t}H$ from $t\bar{t}$



Summary

- Preliminary studies show an approach to reconstruct the semileptonic ttH system and it shows promising results.
- ► Variables from the reconstruction improve the separation between tt̄H and tt̄
- ▶ Goal is include the reconstruction method in $t\bar{t}H(b\bar{b})$ analysis for ICHEP 2016

Next steps:

- Optimize the method.
- Define more variables from reconstruction.
- Test more ideas.

Backup

Matching events

- To estimate the efficiency of the method, reconstructed events are matched with the true partons.
- Truth matching: dR between final parton and jets < 0.3
- Fraction of events with the corresponding truth information.



Discriminating variables for reconstruction 5j,4b

- Variables uncorrelated with Higgs.
 - leptop_mass
 - incomplete_hadtop_mass
 - (qhadW+blepTop)_mass
 - (lepW+bhadTop)_mass
 - dR(qhadW, bhadTop)
 - dR(qhadW, blepTop)
 - dR(lep, blepTop)
 - dR(lep, bhadTop)
 - dR(lepTop, incomplete_hadTop)
 - dR(lep, blepTop) dR(bhadTop, qhadW)



Reconstruction in $\geq 6j \geq 4b$ - reco_BDT_withHiggs

 Combination of jets with the highest reco_BDT_withH output is chosen as the good combination.





BDT to discriminate $t\bar{t}H$ from $t\bar{t}$: $6\geq 6j \geq 4b$

 \blacktriangleright Adding variables from the reconstruction we can increase the separation in ${\sim}17\%$

Ranking from TMVA

Ranking from TMVA

		Rank	Variable
Rank	Variable	1	semilepMVAreco_RecoBDT_withH
1	dRbb_avg	2	dRbb_avg
2	Mbb_MindR	3	H1_all
3	H1_all	4	Centrality
4	pT_jet5	5	pT_jet5
5	Aplan_bjets	6	Aplan_bjets
6	Centrality	7	semilepMVAreco_higgs_mass
7	dRlepbb_MindR	8	Mbb_MindR
8	dRbb_MaxPt	9	dRbb_MaxPt
9	Mbj_MaxPt	10	dRlepbb_MindR
10	Muu_MindR	11	Mbj_MaxPt
		12	semilepMVAreco_bbhiggs_dR
		13	Muu_MindR

Linear correlations among the input variables (6j,4b)

Correlation Matrix (signal)

