Multivariate Analysis Using an Articial Neural Network to Improve Sensitivity in the ATLAS $t\bar{t}H\left(b\bar{b}\right)$ Search

Simon Kirchhof

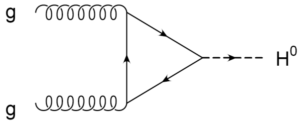
DESY Hamburg – ATLAS group
University of Heidelberg

September 10, 2015

Outline

Motivation for ttH

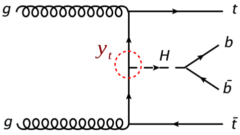
Higgs processes confirmed @LHC:



$$\mathrm{H} \longrightarrow \gamma\gamma$$
$$\mathrm{H} \longrightarrow \mathrm{ZZ}^{(*)} \longrightarrow 4\mathrm{l}$$
$$\mathrm{H} \longrightarrow \mathrm{WW}^{(*)} \longrightarrow \mathrm{l}\nu\mathrm{l}\nu$$
$$\mathrm{H} \longrightarrow \tau\tau$$

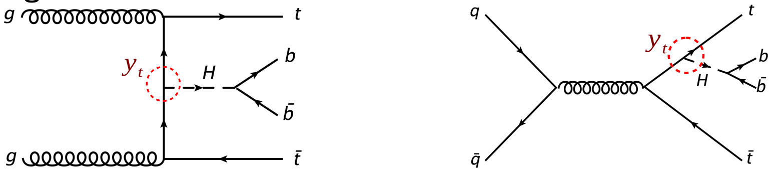
Interesting: Yukawa coupling $q\bar{q}H$

- Not directly observed so far
- $t\bar{t}H\,(b\bar{b})$ is most probable:
  - $Y_t \propto m_t$
  - $\mathrm{BR}\,(\mathrm{H} \to b\bar{b})$ high
- Associate top production "proofs" coupling

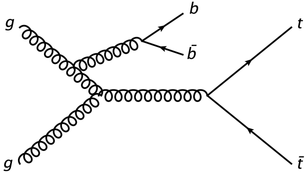


Signal & Background

Signal:



Dominant background:



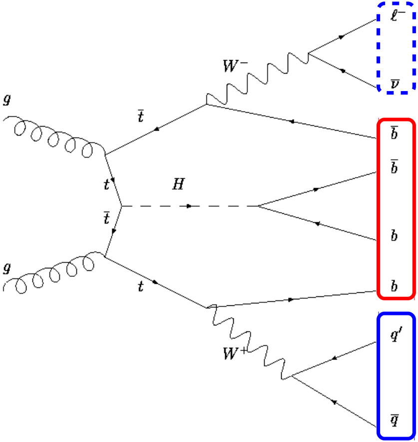
Other backgrounds:

- $t\bar{t}V$
- non-$t\bar{t}$

Outline

Basic Event Selection

Considerations: **Good to trigger** & **signal-like**



- Focus on single-lepton events:
  - e/μ with $p_T > 25$ GeV
- At least 4 *reconstructed* **jets**
- At least 2 *reconstructed* b-jets

- Classify events according to...
  - Number of **jets**
  - Number of b-jets

Categorisation into Analysis Regions

Table: Nine analysis regions.

	2 b-jets	3 b-jets	$\geq$4 b-jets
4 jets			
5 jets			rich
$>=$6 jets		rich	rich

"Signal rich":

- $\frac{S}{B} > 1\%$
- $\frac{S}{\sqrt{B}} > 0.3$

Remark:

- Technical definition of b-jet is ambiguous

Outline

Discriminating Variables

Find differences between signal and background, possible discriminators:

Event-shape variables

- Centrality
- Aplanarity
- ...

Global event variables

- Scalar sum of the jets p_Ts
- ...

Object pair properties

- ΔR of jet pair with $\cdots$
  - $\cdots$ largest $\sum \mathrm{p}_T$
  - $\cdots$ largest invariant mass
- ...

There are ~ 20 in the analysis. I'll only show two.

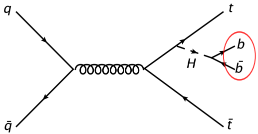
High Transverse Momentum Jets (1)

Definition

$$N_{40}^{\mathrm{jet}} := N_{\mathrm{jet}}\left(|\mathrm{p}_T| > 40\ \mathrm{GeV}\right)$$

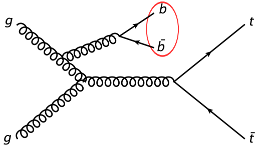
Signal:

- Two b-jets from **higgs** $\Longrightarrow$ $M_{\mathrm{b\bar{b}}} = M_{\mathrm{H}}$
- $|\mathrm{p}_T|$ likely to be high



Background:

- Two b-jets from **gluon**
- Gluon radiation tends to be “soft”
- $|\mathrm{p}_T|$ likely to be lower



$\Longrightarrow$ On average, signal has more high transverse momentum jets!

High Transverse Momentum Jets (2)

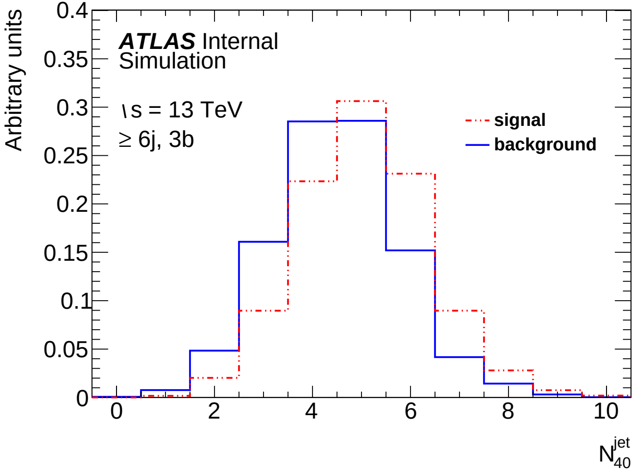


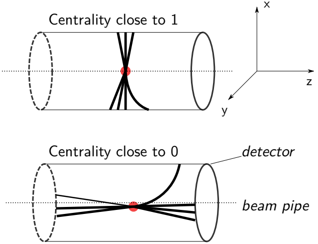
Figure: N_{40}^{jet} in the $\geq$6 jets 3 tight b-jets region.

Centrality (1)

Definition

$$\text{Centrality} := \frac{\sum_i |\mathrm{p}_T^i|}{\sum_i \mathrm{E}_i}$$

$$i = \text{Jets, Leptons}$$



- **Signal:**
  - $\mathbf{t\bar{t}H}$ on-shell: $\geq$ **475** $\mathbf{GeV}$
  - High $\hat{s}$ more likely for $x_1 = x_2$ (Bjorken x)
  - Less z-boosted $\approx$ more central
- **Background:**
  - $\mathbf{t\bar{t}}$ on shell: $\geq$ **350** $\mathbf{GeV}$
  - Lower $\hat{s}$ are possible
  - Possibly more z-boosted events

Centrality (2)

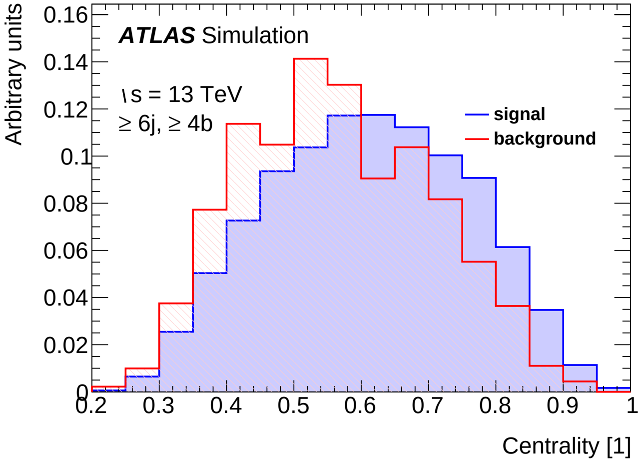


Figure: Centrality in the $\geq$6 jets $\geq$4 tight b-jets region.

Motivation for Multivariate Analysis

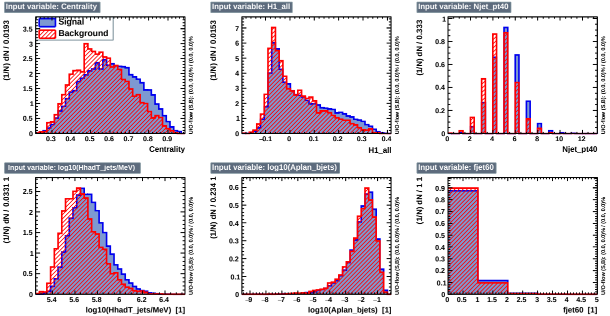
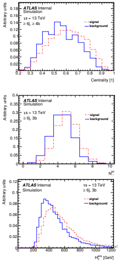


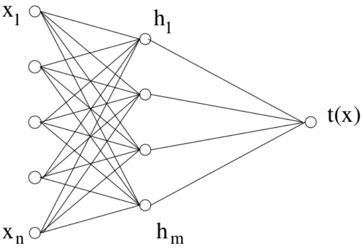
Figure: Some of the "better" variables.

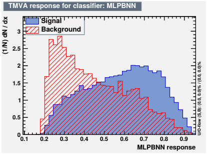
Figure: Me when I look at them.

Individual cuts won't do the job. Need more complex methods!

Artificial Neural Network







- Inspired by biology (i.e. the brain)
- Process information through Multilayer Network
  - Input layer: variables
  - "Hidden layer": neurons
  - Output layer: **t**
- **Parameters ("weights")** $\vec{w}$: Strength of connections
- **Important:** $\mathrm{t}\,(\tilde{\mathrm{x}})$ is non-linear in general!

Artificial Neural Network - Training (simplified)

Define error function ϵ for an input vector $\vec{x}$, e.g.

$$\epsilon\left(\vec{x},\vec{w}\right) = \left(\mathrm{t}\left(\vec{x},\vec{w}\right) - \hat{\mathrm{t}}\right)^2$$
$$\hat{\mathrm{t}}\left(\mathrm{sig}\right) = 1$$
$$\hat{\mathrm{t}}\left(\mathrm{bkg}\right) = -1$$

Iterative training with Monte-Carlo samples:

- Start with random/guessed weights $\vec{w}$
- Calculate error for training event
- Adjust weights $\vec{w}$ to minimize ϵ (i.e. backpropagation)

Outline

Overview of My Work

Software:

- **TMVA** (4.2.1), comes with **ROOT** (6.04/02)
- **T**oolkit for **M**ulti**V**ariate **A**nalysis
- Set up a dedicated software tool to train & test ANNs

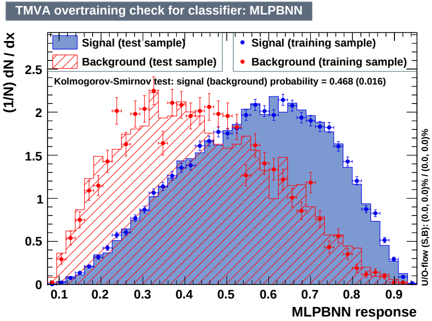
Variables & training settings:

- Adapted from LHC run 1 analysis
- Started optimisation for run 2

Impact on Sensitivity of ttH

- Calculated signal significance for estimated luminosity

ANN with Variables and Settings of Run 1



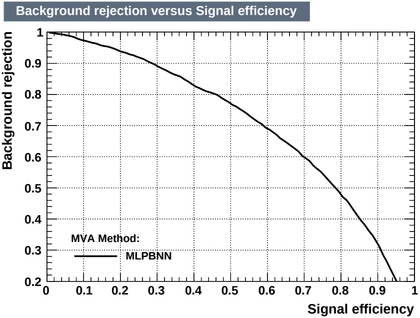


Figure: ANN output in the ≥6 jets 3 tight b-jets region.

Receiver **O**perating **C**haracteristic := Curve on the right, showing how efficient different cuts are w.r.t. keeping signal and rejecting background.

Signal Significance: Category + ANN

"New" category: $\geq$6 jets, $\geq$3 tight b-jets, $\geq$4 loose b-tags

- $\frac{S}{B} = 1.7\%$
- $\frac{S}{\sqrt{S+B}} = 0.9$

For a Luminosity of $\mathcal{L} = 5\ \mathrm{fb}^{-1}$:

- Signal: 45 events
- Background: 2577 events

After applying optimal Neural Network cut:

- $\frac{S}{\sqrt{S+B}} = 1.0$

Outlook

- Future analysis should include **all different background samples**
- ANN training with **more MC statistics**
- **New MVA variables** will possibly be added
- Inefficient: Cut on ANN output
- Better: **Likelihood fit** to ANN output distribution

Special thanks to my supervisors:

- **John Stakely Keller**
- **Judith Katzy**

Many thanks to the whole ATLAS group here at DESY!

THANK YOU FOR YOUR ATTENTION!

Questions?

Appendix: Simulated Data

Two Monte-Carlo samples

Type	Description	Generator	Events
$t\bar{t}$ (bkg.)	Semi-/dileptonic	`Powheg+Pythia`	$\sim$ 2 million
$t\bar{t}H$ (sig.)	Inclusice higgs decay, semileptonic top decay	`aMcAtNlo+Herwig++`	$\sim$ 2 million

- Minor bkg. contributions missing ($t\bar{t}V$, non-$t\bar{t}$)
- Number of available events (after skimming):
  - Signal: ca. 400000
  - Background: ca. 450000

Appendix: Pre-Selection Cuts

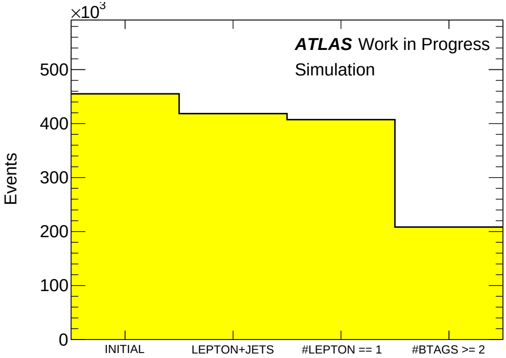


Figure: Cutflow of skimmed **bkg.** sample.

- $N_{\mathrm{jets}} \geq 4$
- One lepton with $|\mathrm{p}_T| > 25\ \mathrm{GeV}$
- Exactly one Lepton
- Require $N_{\mathrm{b-tags}} \geq 2$

Appendix: Categories for ANN

Reminder:

- Catogerisation according to number of jets & number of b-jets
- Different b-tagging working points possible:
  - Loose: 85% efficiency
  - Tight: 77% efficiency

Most promising regions:

- Good signal purity ($\frac{S}{B} > 1\%$) & good signal significance ($\frac{S}{\sqrt{B}} > 0.3$)
- As many events as possible

Raw MC events in examined categories

	old		new (Hyoyin)
	tight b-tags: 3	tight b-tags: ≥ 4	tight b-tags: ≥ 3 loose b-tags: ≥ 4
≥ 6 jets	**sig: $\sim$ 46000** **bkg: $\sim$ 7800**	**sig: $\sim$ 22600** **bkg: $\sim$ 900**	**sig: $\sim$ 35600** **bkg: $\sim$ 2300**

Appendix: MVA in Fomulas

- Each event (signal or background) has D measured variables $\vec{x}$
- "Feature space" in $\mathbb{R}^D$:
  - $x_0 = \mathrm{Centrality}$
  - $x_1 = N_{40}^{\mathrm{jet}}$
  - $\ldots$
- Find mapping:

$$\boxed{\mathrm{t}\left(\vec{x}\right) : \mathbb{R}^D \longrightarrow \mathbb{R}}$$

- **t** is new "classifier", desired properties:
  - Seperation power, i. e. $\langle \mathrm{t}_{\mathrm{sig}} \rangle \neq \langle \mathrm{t}_{\mathrm{bkg}} \rangle$ and $\sigma_{\mathrm{t}_{\mathrm{sig}}}$, $\sigma_{\mathrm{t}_{\mathrm{bkg}}}$ small
  - Good generalization properties when applied to "unkown" events

Appendix: Artificial Neural Network

Feed-Forward Network

$$h_i\left(\vec{x}\right) = s\left(w_{i0} + \sum_{i=1}^{n} w_{ij} x_i\right)$$

$$\mathrm{t}\left(\vec{x}\right) = s\left(a_0 + \sum_{i=1}^{n} a_i h_i\right)$$

ANN as referred to in this talk:

- MLP architecture
- Feed-forward

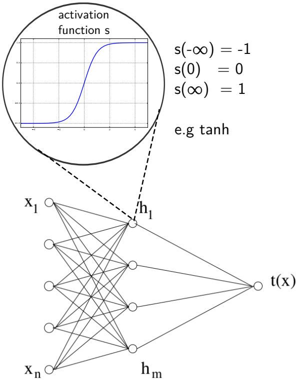
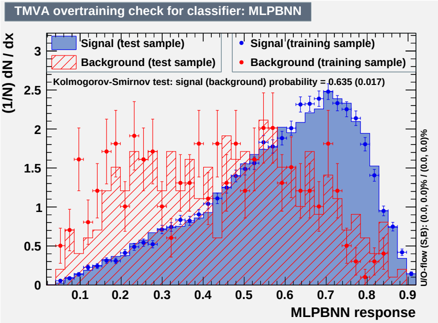


Figure: **M**ulti**L**ayer **P**erceptron with one hidden layer.

Appendix: ANN with Variables and Settings of Run 1



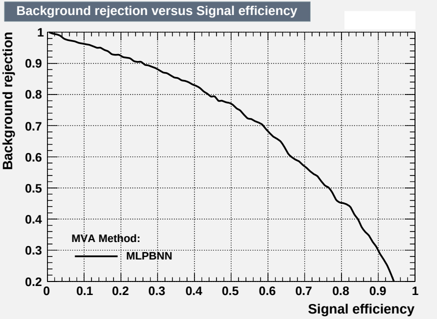
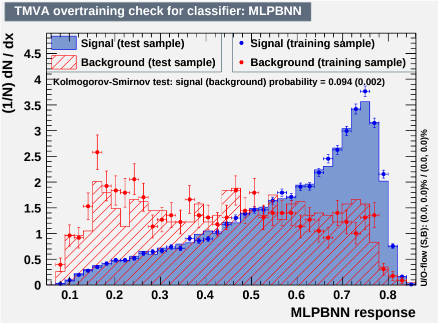


Figure: ANN output in the ≥6 jets ≥4 tight b-jets region.

Appendix: ANN with Variables and Settings of Run 1



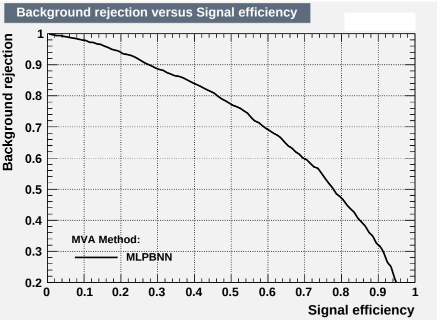


Figure: ANN output in the ≥6 jets ≥3 tight & ≥4 loose b-jets region.

Appendix: Convergence of the ANN

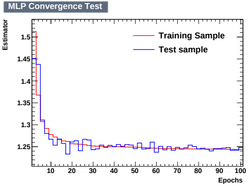


Figure: Error estimator vs. training cycles (6j 3b).

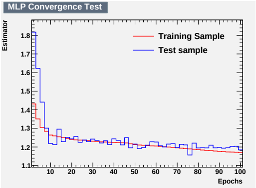


Figure: Error estimator vs. training cycles (6j 4b).

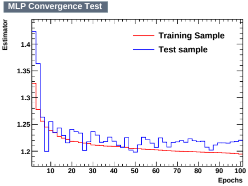


Figure: Error estimator vs. training cycles (6j 3/4b).

Appendix: Ranking of Variables – Naive

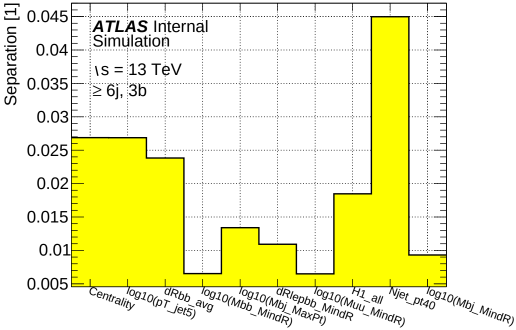


Figure: Separation power of all variables used in the $\geq$6 jets 3 tight b-jets region.

Separation power of a discriminating variable ξ:

$$\int \frac{\left(N_{\mathrm{sig}}\left(\xi\right) - N_{\mathrm{bkg}}\left(\xi\right)\right)^2}{\left(N_{\mathrm{sig}}\left(\xi\right) + N_{\mathrm{bkg}}\left(\xi\right)\right)} d\xi$$

Ranking

1. N_{40}^{jet}
2. Centrality
3. $\mathrm{p}_T^{\mathrm{jet5}}$
4. ...

Appendix: Ranking of Variables

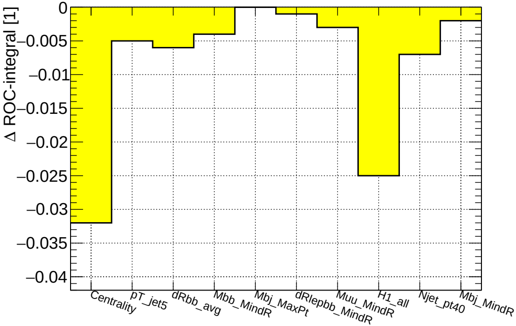


Figure: Separation power of all variables used in the ≥6 jets 3 tight b-jets region.

- Train ANN 10 times
- Each time, leave out one variable
- Check how quality of ANN "degrades"

Ranking

1. Centrality
2. H_1 (Fox-Wolfram)
3. N_{40}^{jet}
4. ...

Appendix: Ranking – Run 1 vs. Run 2 Simulation (6j3b)

Run 1

1. N^{jet}_{40}
2. Centrality
3. H_1 (Fox-Wolfram moment)
4. $M^{\mathrm{min}\Delta R}_{\mathrm{bj}}$
5. $\Delta R^{\mathrm{avg}}_{\mathrm{bb}}$
6. $\mathrm{p}^{\mathrm{jet5}}_{T}$
7. $M^{\mathrm{max}\ \mathrm{p}_T}_{\mathrm{bj}}$
8. $M^{\mathrm{min}\Delta R}_{\mathrm{uu}}$
9. $\Delta R^{\mathrm{min}\Delta R}_{\mathrm{lepbb}}$
10. $M^{\mathrm{min}\Delta R}_{\mathrm{bj}}$

Run 2 simulation

1. Centrality
2. H_1 (Fox-Wolfram moment)
3. N^{jet}_{40}
4. $\Delta R^{\mathrm{avg}}_{\mathrm{bb}}$
5. $\mathrm{p}^{\mathrm{jet5}}_{T}$
6. $M^{\mathrm{min}\Delta R}_{\mathrm{bb}}$
7. $M^{\mathrm{max}\ \mathrm{p}_T}_{\mathrm{bj}}$
8. $\Delta R^{\mathrm{min}\Delta R}_{\mathrm{lepbb}}$
9. $M^{\mathrm{min}\Delta R}_{\mathrm{bj}}$
10. $M^{\mathrm{min}\Delta R}_{\mathrm{uu}}$