Sensitivity studies using multivariate techniques for the search for fully hadronic decays of top squarks with the ATLAS detector

Nicolas Köhler

Max Planck Institute for Physics
(Werner-Heisenberg-Institut)

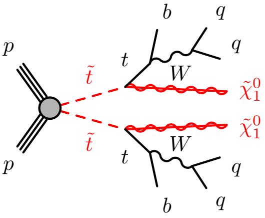
Wednesday 21st March, 2018

Bundesministerium
für Bildung
und Forschung

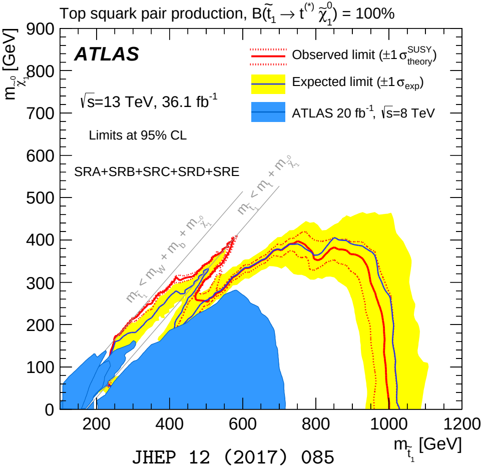
FSP 103
ATLAS

MAX-PLANCK-GESELLSCHAFT

The search for hadronically decaying top squarks



- Introduced by P. Mogg (T 50.9)
- Simplified models with $\tilde{t}_1$, $\tilde{\chi}_1^0$ (and eventually $\tilde{\chi}_1^{\pm}$)
- Cut&Count analysis with regions depending on decay kinematics
- Main backgrounds: Z+jets, $t\bar{t}$, single top quark production

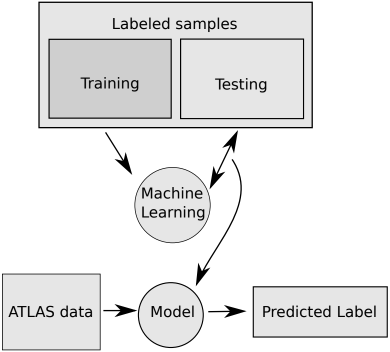


JHEP 12 (2017) 085

→ Exclusion of top squark masses up to $m_{\tilde{t}} = 1$ TeV

Increasing the sensitivity towards higher top squark masses

- Multivariate techniques of increasing popularity in HEP
- Supervised learning in order to classify signal and background events
- Sensitivity study done with J. Graw (MSc):
  - -- Is MVA really improving compared to Cut&Count?
  - -- Systematic study of machine learning algorithms / settings / inputs etc.
  - → BDT vs. Neural networks, ...
  - -- How to validate MVA outputs?
  - → Overtraining, systematics, ...
- Focus on $\tilde{t}_1 \rightarrow t + \tilde{\chi}_1^0$ ($m_{\tilde{t}_1} \gg m_{\tilde{\chi}_1^0}$) only



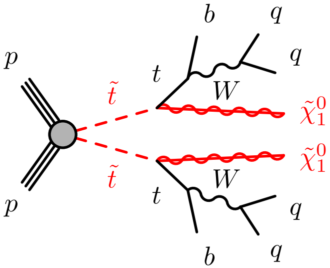
Analysis setup

Use preselection of Cut&Count analysis

- $E_{\mathrm{T}}^{\mathrm{miss}}$-trigger & $E_{\mathrm{T}}^{\mathrm{miss}} > 250$ GeV
- Lepton veto
- ≥ 4 jets ($p_{\mathrm{T}} > 80,80,40,40$ GeV)
- $\geq$ **1** *b*-jet
- QCD cleaning

+ At least 2 reclustered $R = 1.2$ jets (no additional requirement)

+ At least 1 reclustered $R = 0.8$ jet (no additional requirement)

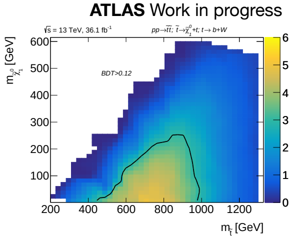


Variables available for training:
All SR A+B variables

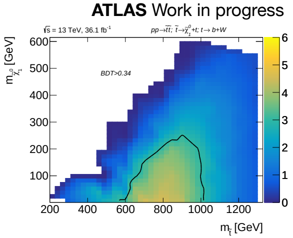
Training setup for $\tilde{t}_1 \rightarrow t + \tilde{\chi}_1^0$ decays

- Train Boosted decision tree (BDT) with all variables used in Cut&Count analysis
- Which signal sample to use for training?

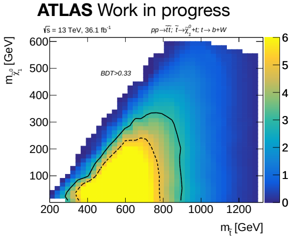
Use $m_{\tilde{t}} = 1000$ GeV,
$m_{\tilde{\chi}_1^0} = 1$ GeV only



Use $\tilde{t}_1 \rightarrow t + \tilde{\chi}_1^0$
simulation with
$m_{\tilde{t}} > 1000$ GeV



Use all $\tilde{t}_1 \rightarrow t + \tilde{\chi}_1^0$
simulation available

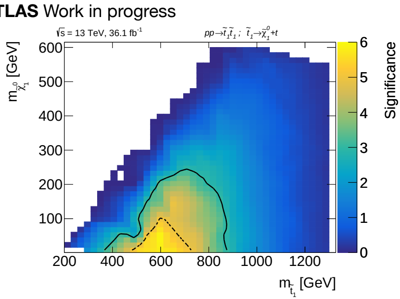


→ For boosted region, using simulation with $m_{\tilde{t}} > 1000$ GeV performs best

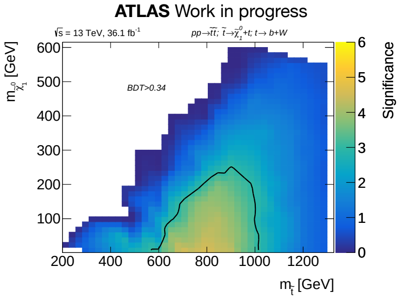
Training setup

Compare to Cut&Count:

Use Cut&Count SRA_TT



Use simulations with $m_{\tilde{t}} > 1000$ GeV



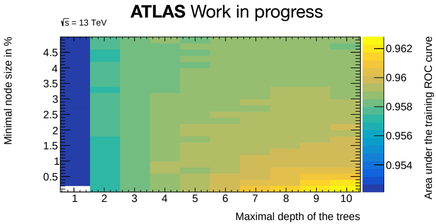
→ Expected 3σ line moves from 900GeV to 1TeV

BDT - settings

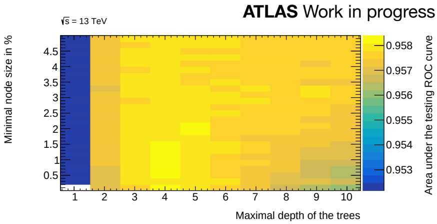
Performed trainings scanning through BDT setting parameters

→ Look at area under ROC-curve to get measure for MVA performance

Training



Testing



→ Overtraining in left plot observed when comparing to testing (right)

→ Best BDT parameters would be Maximal depth of 4,

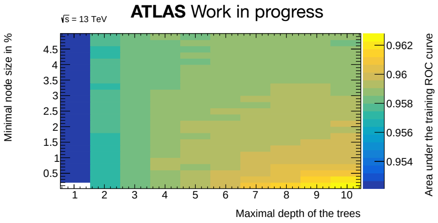
minimal node size of 1%

BDT - settings

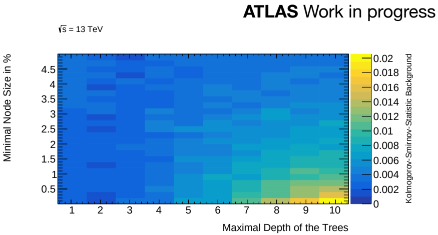
Performed trainings scanning through BDT setting parameters

→ Look at area under ROC-curve to get measure for MVA performance

Training



Kolmogorov-Smirnov-Score



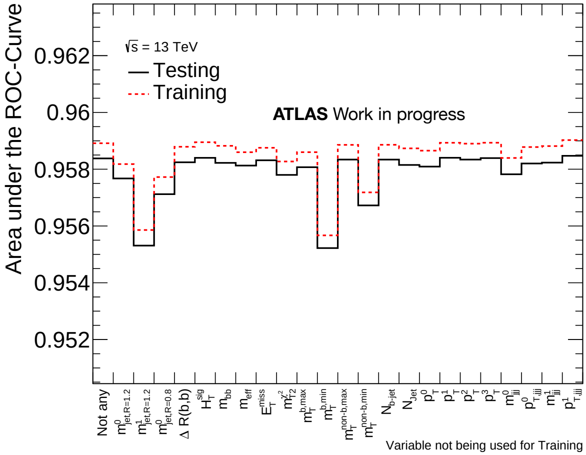
→ Overtraining in left plot observed when comparing to testing (right)

→ Best BDT parameters would be Maximal depth of 4,

minimal node size of 1%

BDT - input variables

Train with all but one variable with the optimized BDT settings

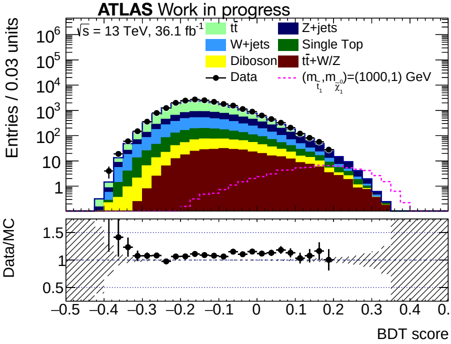


→ E.g.: $m_{\mathrm{T}}^{b,\mathrm{min}}$ is more important than $H_{\mathrm{T}}^{\mathrm{sig}}$

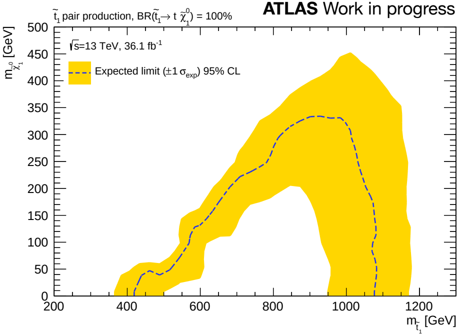
BDT - results

Run with optimal BDT settings and skip unimportant variables

BDT score (Data blinded)



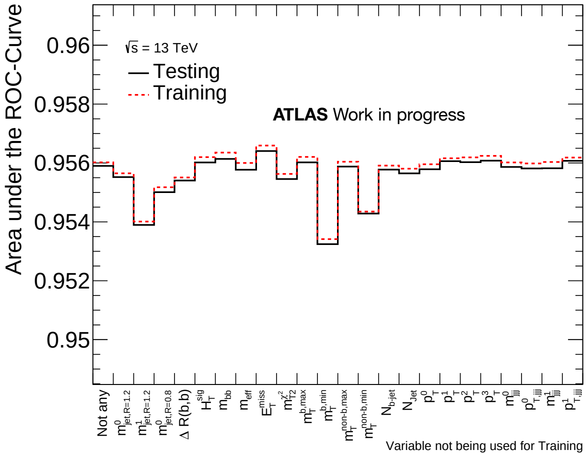
Expected limit (simple one-bin fit using $t\bar{t}$ & Z+jets 0L VRs with inverted BDT cut as CRs)



→ No real improvement (9%) compared to cut based results

Neural networks (NN) - input variables

- The same type of studies were done for neural networks
- Train with all but one variable with the optimized NN settings



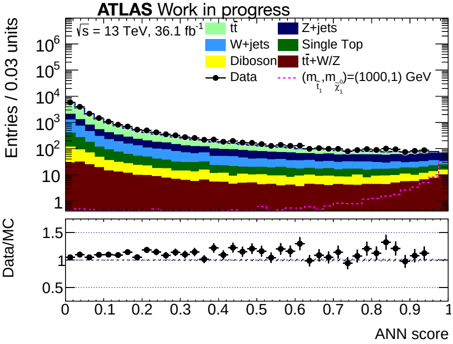
→ Again: $m_{\mathrm{T}}^{b,\mathrm{min}}$ is more important than $H_{\mathrm{T}}^{\mathrm{sig}}$

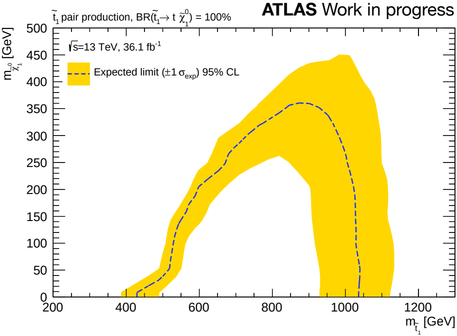
Neural networks - results

Run with optimal NN settings and skip unimportant variables

NN score (Data blinded)

Expected limit (simple one-bin fit using $t\bar{t}$ & Z+jets 0L VRs with inverted NN cut as CRs)





$\rightarrow$ No real improvement ($\sim$ 5%) compared to cut based results

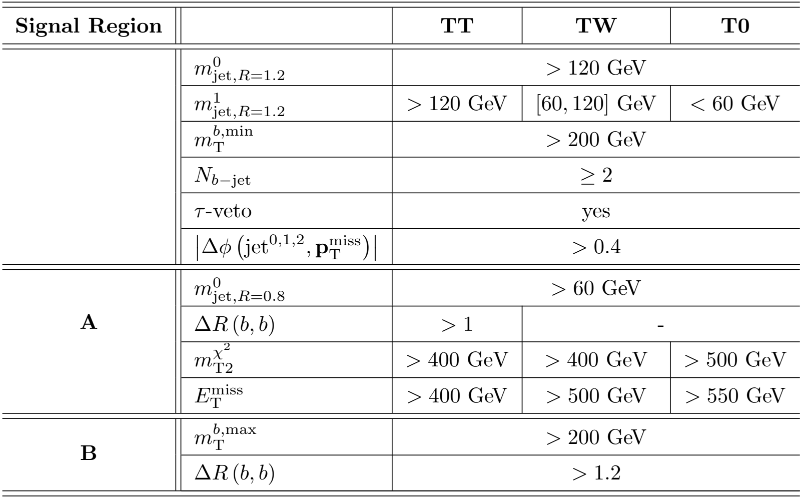
Summary

- Sensitivity of search for top squarks using multivariate techniques performed by J. Graw
- Looked at BDT and neural networks
- Performed systematic scans on MVA settings and input variables
- Calculated expected exclusion limits with simplified control regions

→ 10% / 5% improvement in expected sensitivity for BDT/NN compared to Cut&Count

BACKUP

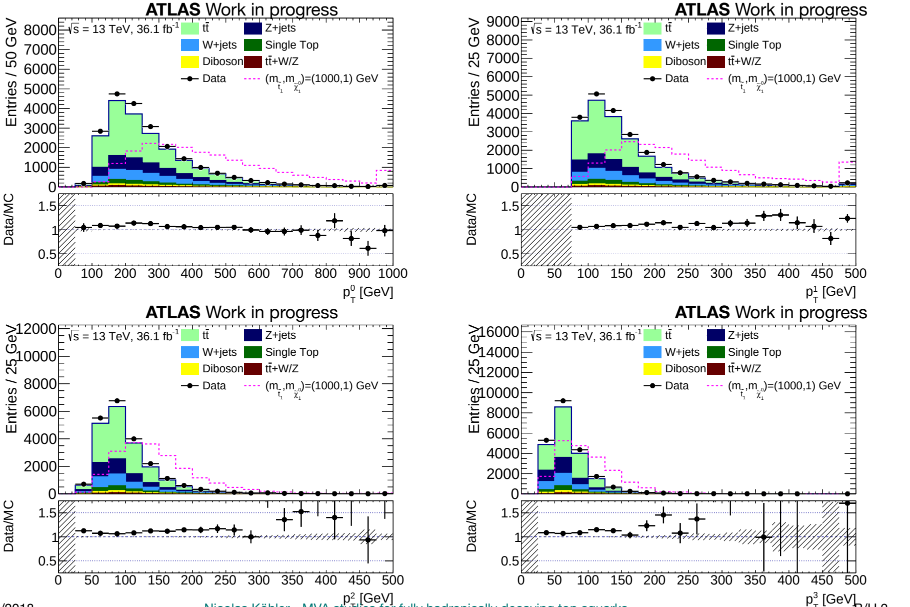
Discriminating variables for SR A+B



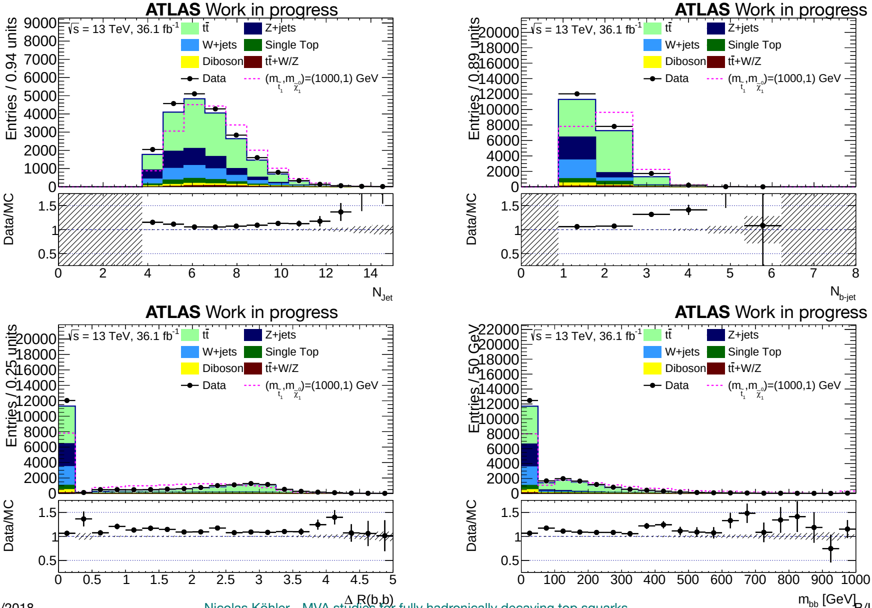
Signal Region		TT	TW	T0
	$m^{0}_{\mathrm{jet},R=1.2}$	> 120 GeV		
	$m^{1}_{\mathrm{jet},R=1.2}$	> 120 GeV	$[60, 120]$ GeV	< 60 GeV
	$m_{\mathrm{T}}^{b,\mathrm{min}}$	> 200 GeV		
	$N_{b-\mathrm{jet}}$	≥ 2		
	τ-veto	yes		
	$\left\lvert\Delta\phi\left(\mathrm{jet}^{0,1,2}, \mathbf{p}_{\mathrm{T}}^{\mathrm{miss}}\right)\right\rvert$	> 0.4		
A	$m^{0}_{\mathrm{jet},R=0.8}$	> 60 GeV		
	$\Delta R\left(b, b\right)$	> 1	-	
	$m_{\mathrm{T2}}^{\chi^2}$	> 400 GeV	> 400 GeV	> 500 GeV
	$E_{\mathrm{T}}^{\mathrm{miss}}$	> 400 GeV	> 500 GeV	> 550 GeV
B	$m_{\mathrm{T}}^{b,\mathrm{max}}$	> 200 GeV		
	$\Delta R\left(b, b\right)$	> 1.2		

JHEP 12 (2017) 085

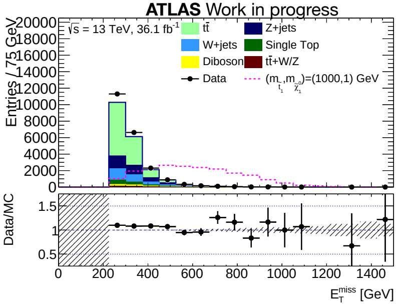
Input distributions: signal normalized to SM integral

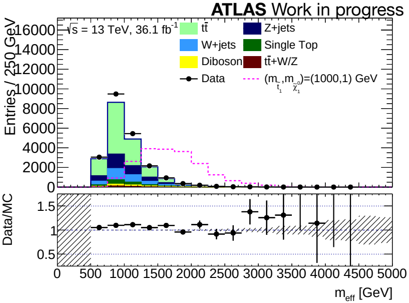


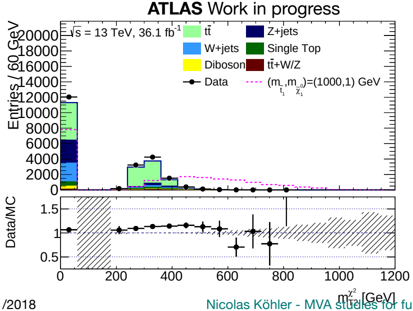
Input distributions: signal normalized to SM integral

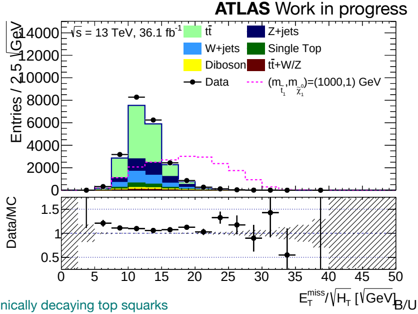


Input distributions: signal normalized to SM integral

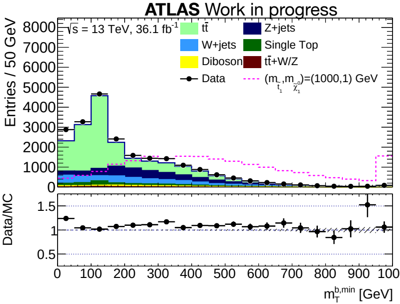


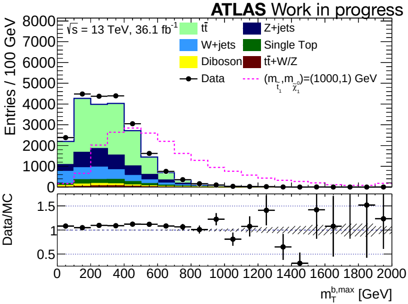


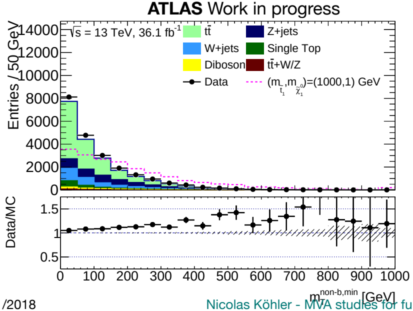


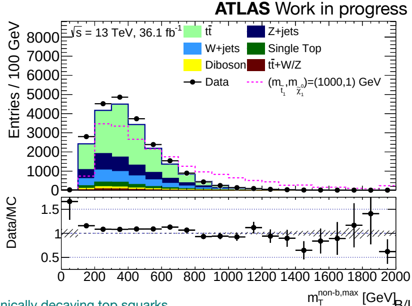


Input distributions: signal normalized to SM integral

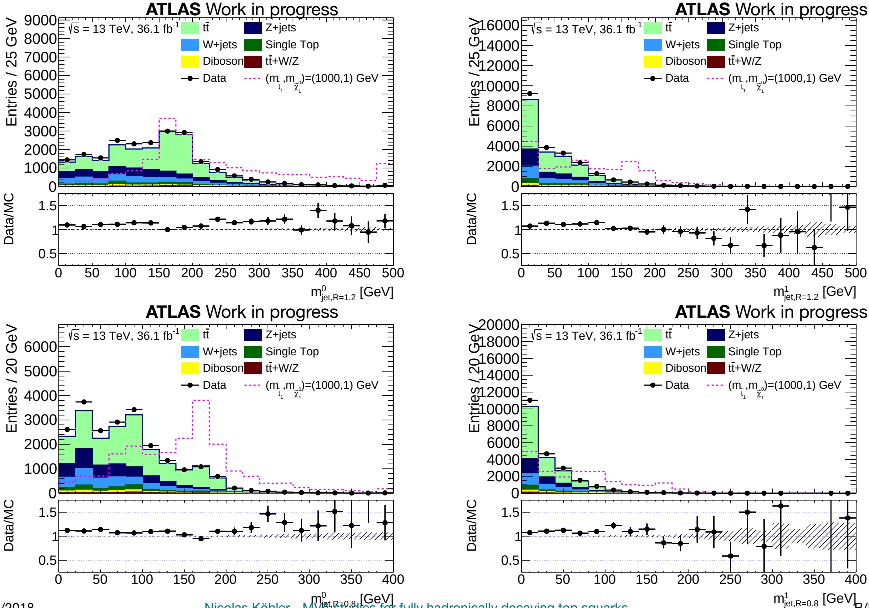




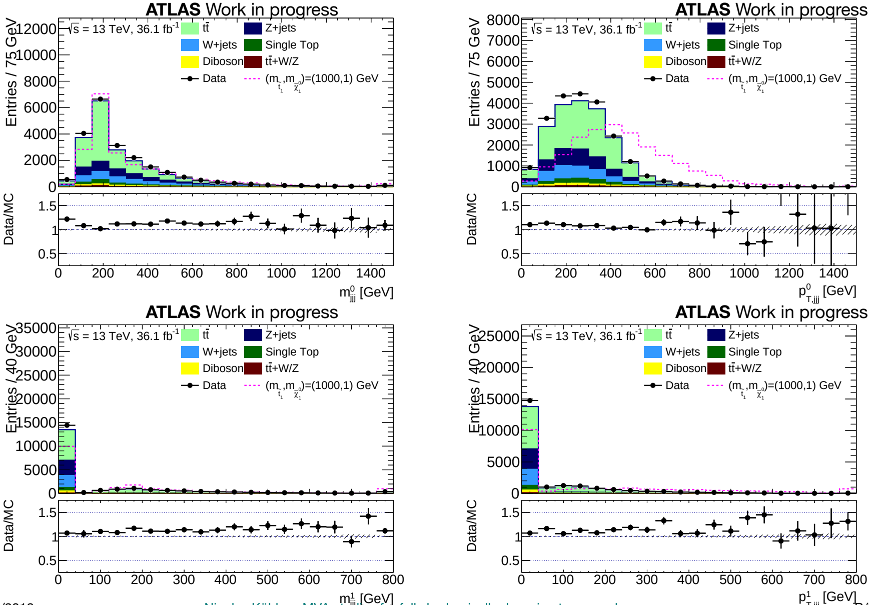




Input distributions: signal normalized to SM integral



Input distributions: signal normalized to SM integral

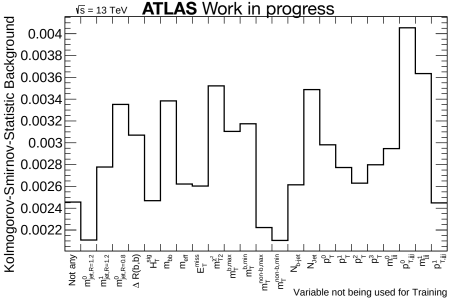


BDT - overtraining

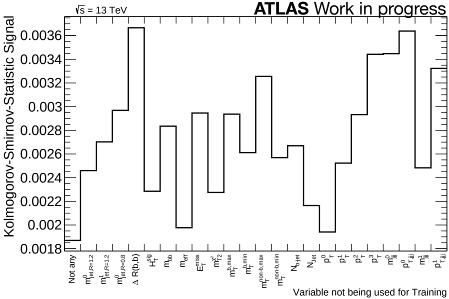
Train with all but one variable with the optimized BDT settings

Calulate Kolmogrov-Smirnov score comparing testing and training ROC-AUC

Background



Signal

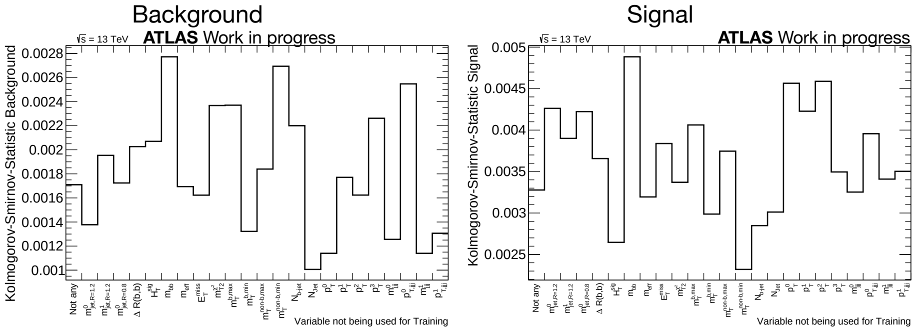


→ The bigger the score, the more overtraining

Neural networks - overtraining

Train with all but one variable with the optimized NN settings

Calulate Kolmogrov-Smirnov score comparing testing and training ROC-AUC



Background

Signal

→ The bigger the score, the more overtraining