Search for top Squarks Using Multivariate Methods

Jonas Graw

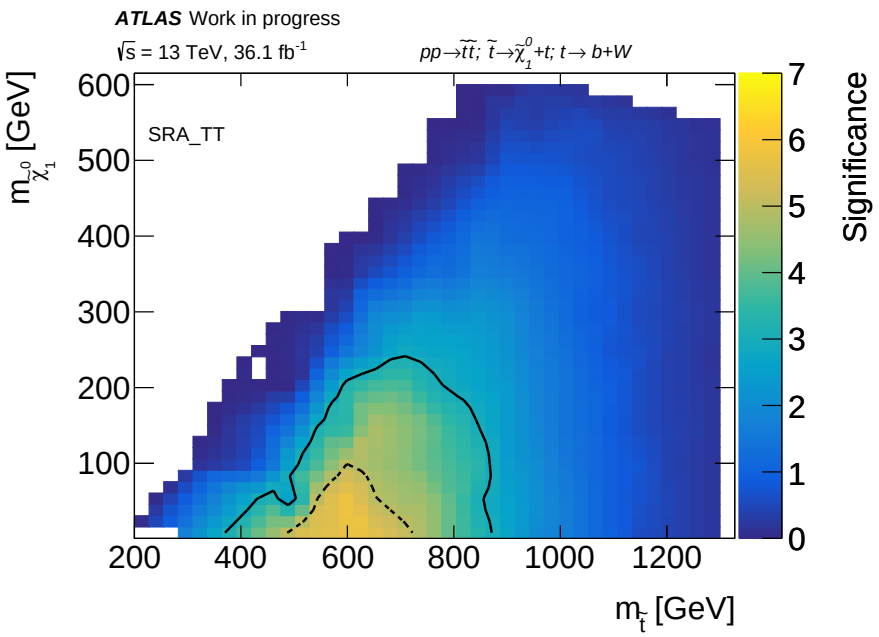
Max Planck Institute for Physics
(Werner-Heisenberg-Institut)

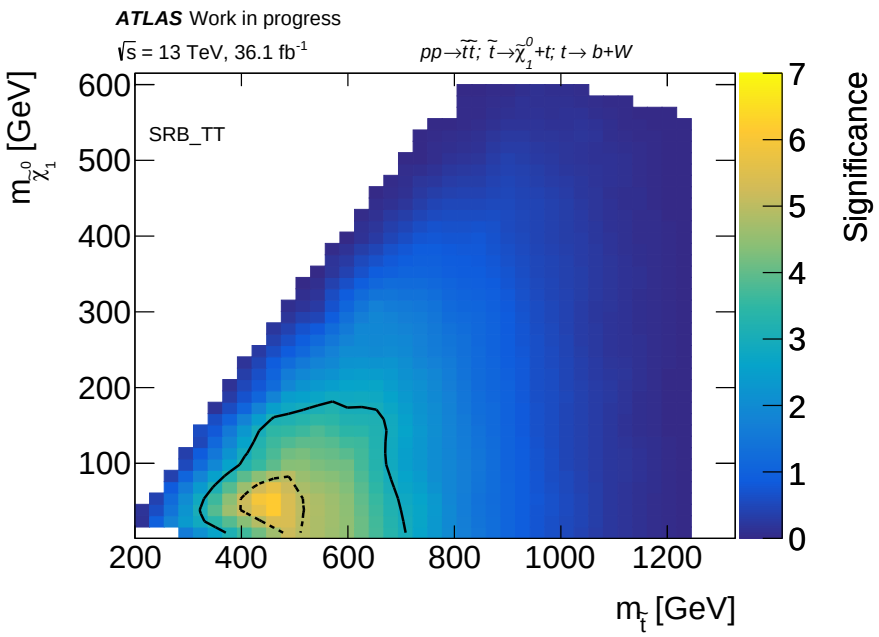
Thursday 20th July, 2017

MAX-PLANCK-GESELLSCHAFT

Motivation

- Standard Strategy: Cut-based analysis

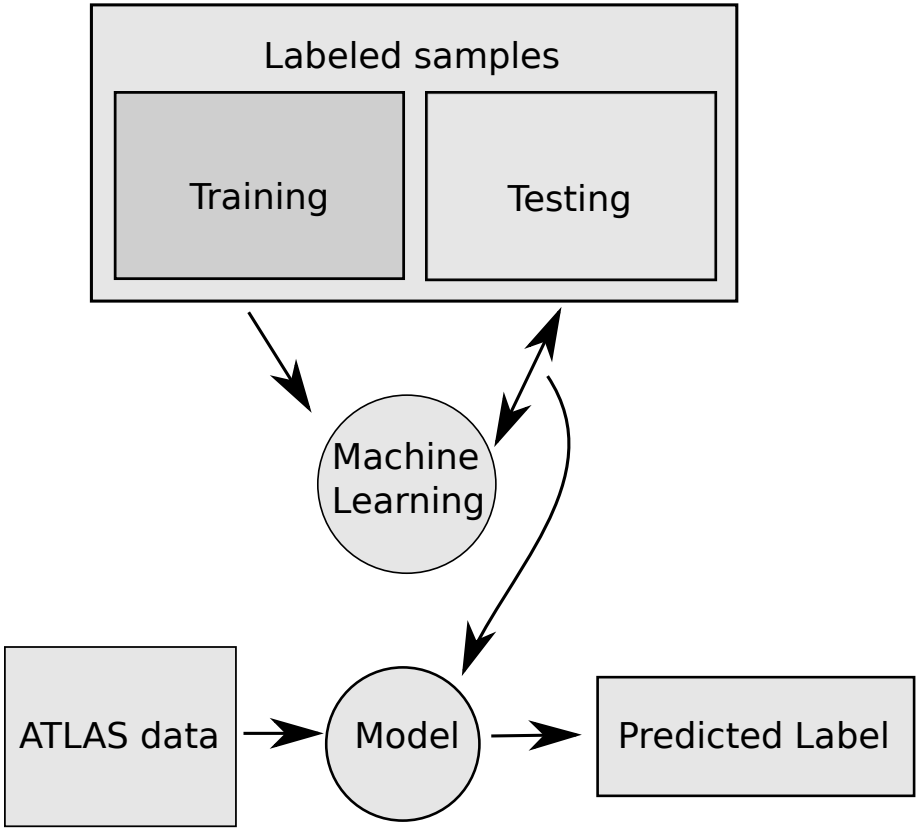




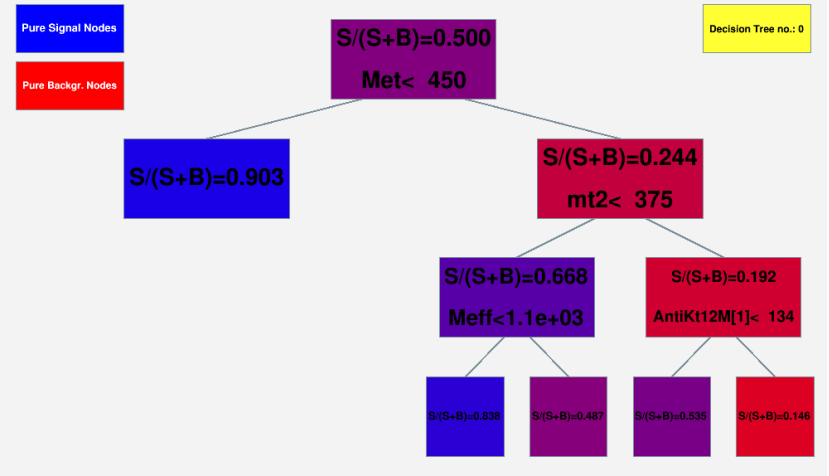
$\longrightarrow$ Try to look at multivariate methods using Monte Carlo-samples

Machine Learning

- Classification: Signal or Background
- Supervised learning: Training done with *labeled* simulated events
- Events divided into training and testing (e. g. 50%-50%)
- Overtraining (”learning by heart”) needs to be avoided



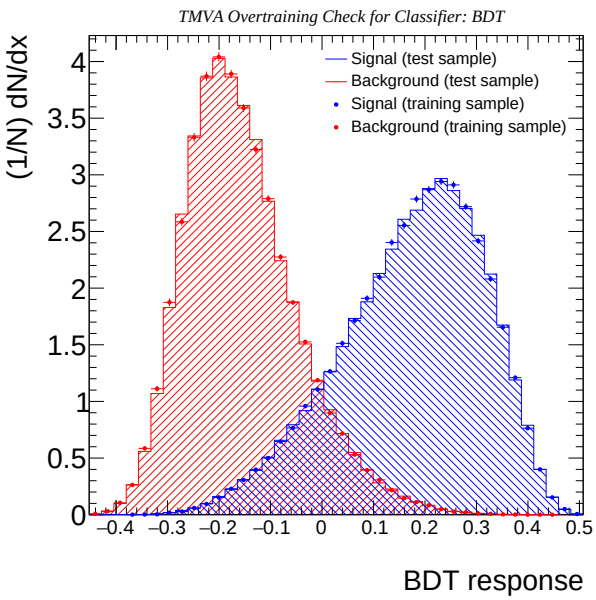
Boosted Decision Tree (BDT)



- Division into two processes: Signal and Background
- Decision which variable to take is done by exactly one discriminating variable (cut)
- Chosen discriminating variable gives the best possible signal background separation
- **Boosting**: Training of a new tree, for which falsely classified events get a bigger weight
- BDT response $r(i) \in [-1, 1]$ of an event i: Classification measure dependent on the trees with limits $\left\{ \begin{array}{l} r(i) = +1 \\ r(i) = -1 \end{array} \right\}$: All trees classify i as $\left\{ \begin{array}{l} \text{signal} \\ \text{background} \end{array} \right\}$

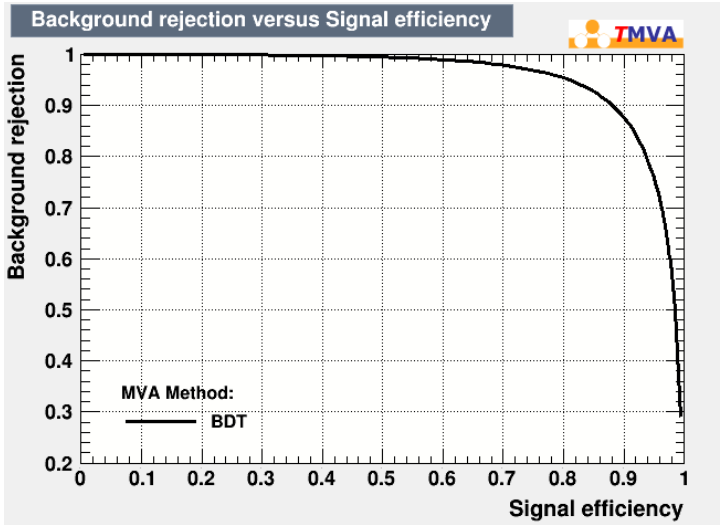
Multivariate Methods: Boosted Decision Tree

- Boosted Decision Tree (BDT) utilized to optimize the usage of discriminating variables
- Training on region with 0ℓ, $E_{\mathrm{T}}^{\mathrm{miss}} >$ 250 GeV, $\geq$ 4 Jets & $\geq$ 1 b-Jet



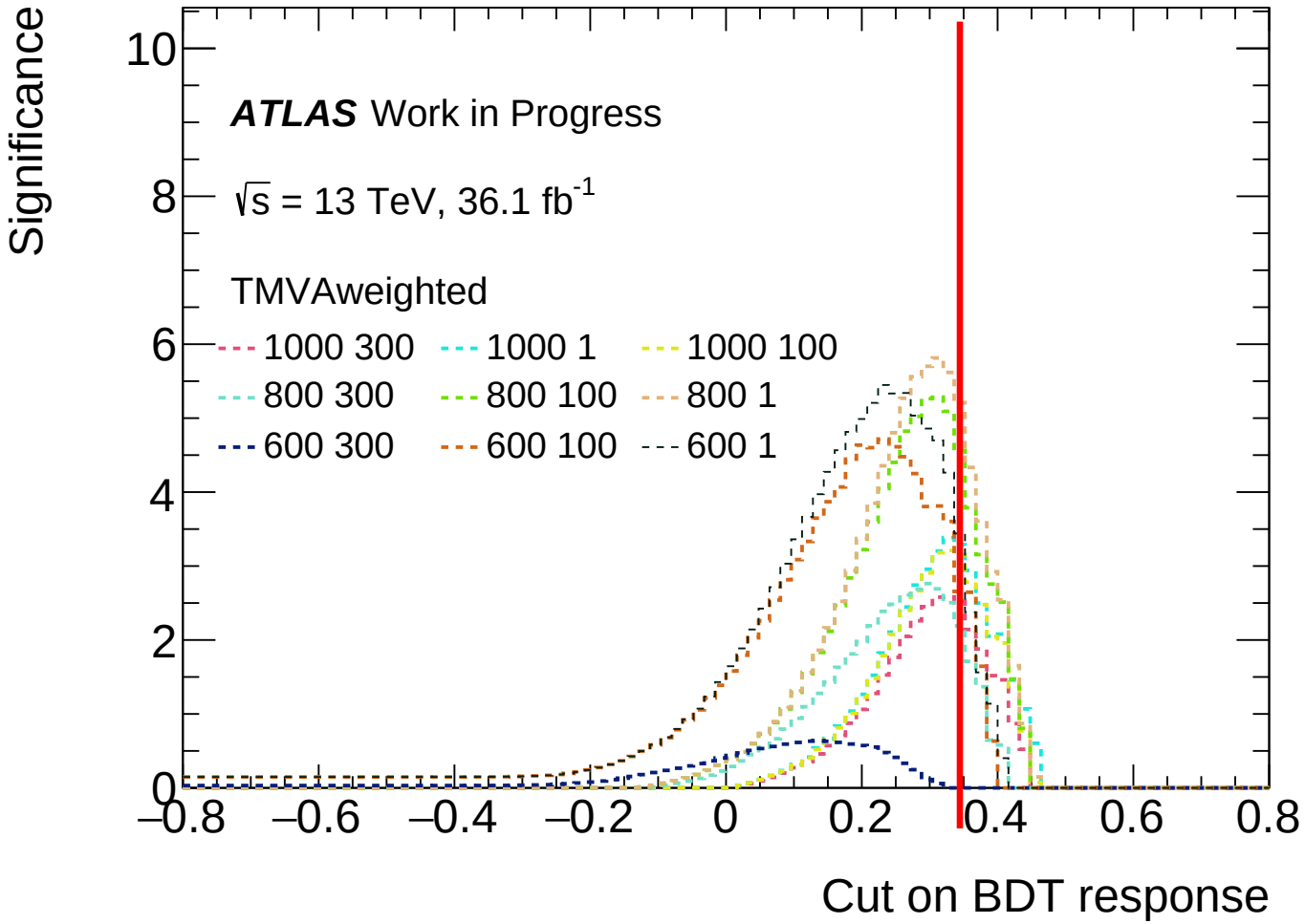
- BDT-response for MC-test- and MC-training data is in very good accordance $\rightarrow$ No overtraining

- Training for models with $m_{\tilde{t}} \geq$ 1TeV reaches better significances than Training for model with $m_{\tilde{t}} =$ 1TeV and $m_{\tilde{\chi}^0}$ =1 GeV
- ROC-Curve is an indicator of how good the training was



- The area under the curve is a good measure of the training

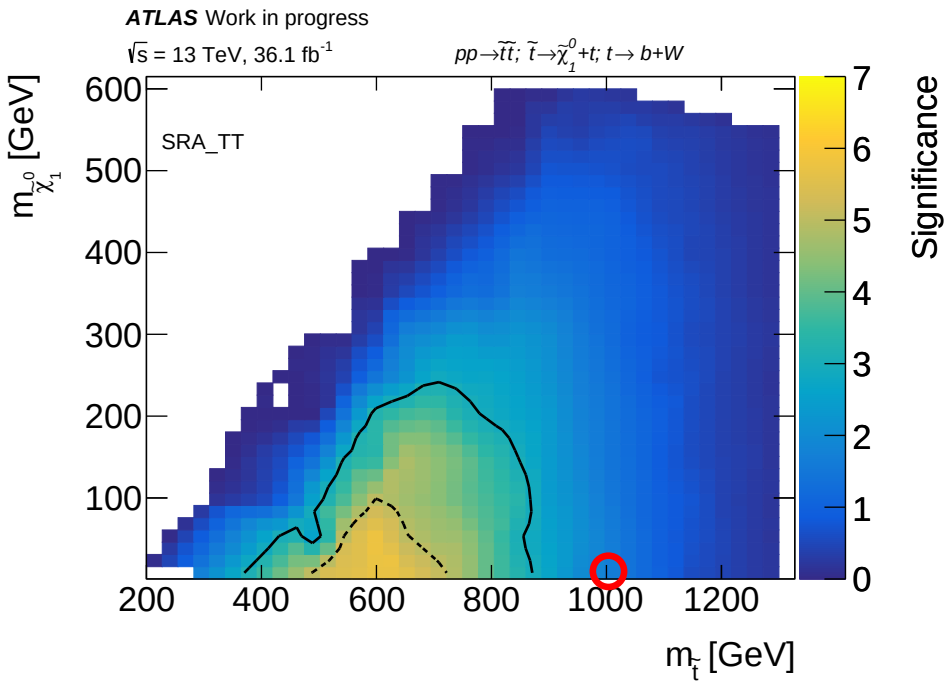
Cut on BDT-respones



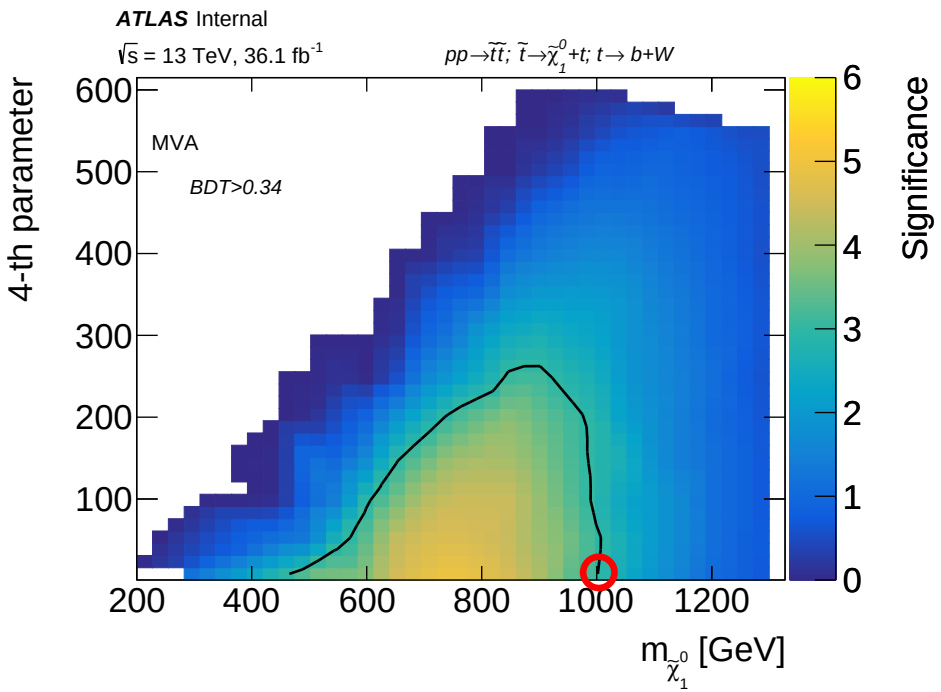
- Partially very high significances
- Important variables: m_{T2}, $p_T(top)$, E_T^{miss}
- Cut on BDT-response $\geq$ 0,34

Expected Significances in $m_{\tilde{t}}$-$m_{\tilde{\chi}_1^0}$-Parameter Space

Cut-based method



BDT method



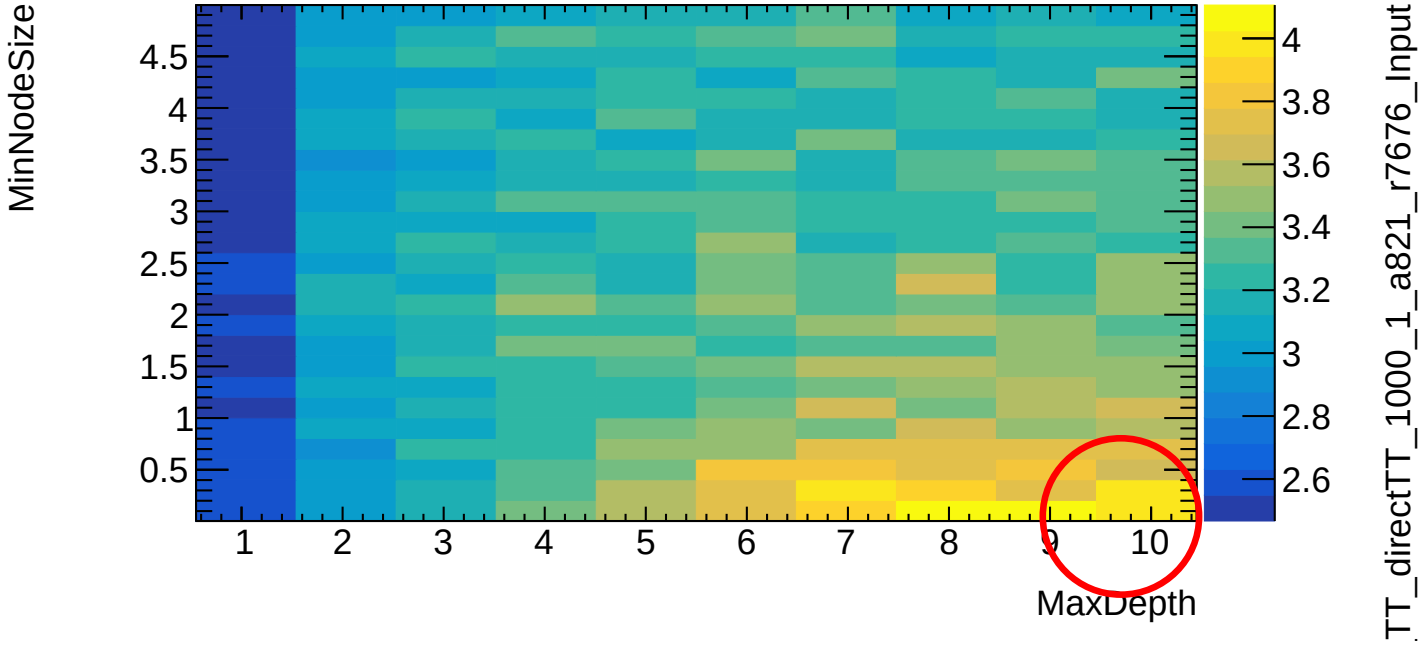
Cut-based	BDT
1.7σ	3.0σ

$\Rightarrow$ Expected significance of 3σ up to $m_{\tilde{t}} = 1$ TeV

$\Rightarrow$ Can we increase this by changing the BDT settings?

Optimization of the BDT-Settings

- Expected significances of a sample with $m_{\tilde{t}} = 1$ TeV and $m_{\tilde{\chi}^0} = 1$ GeV

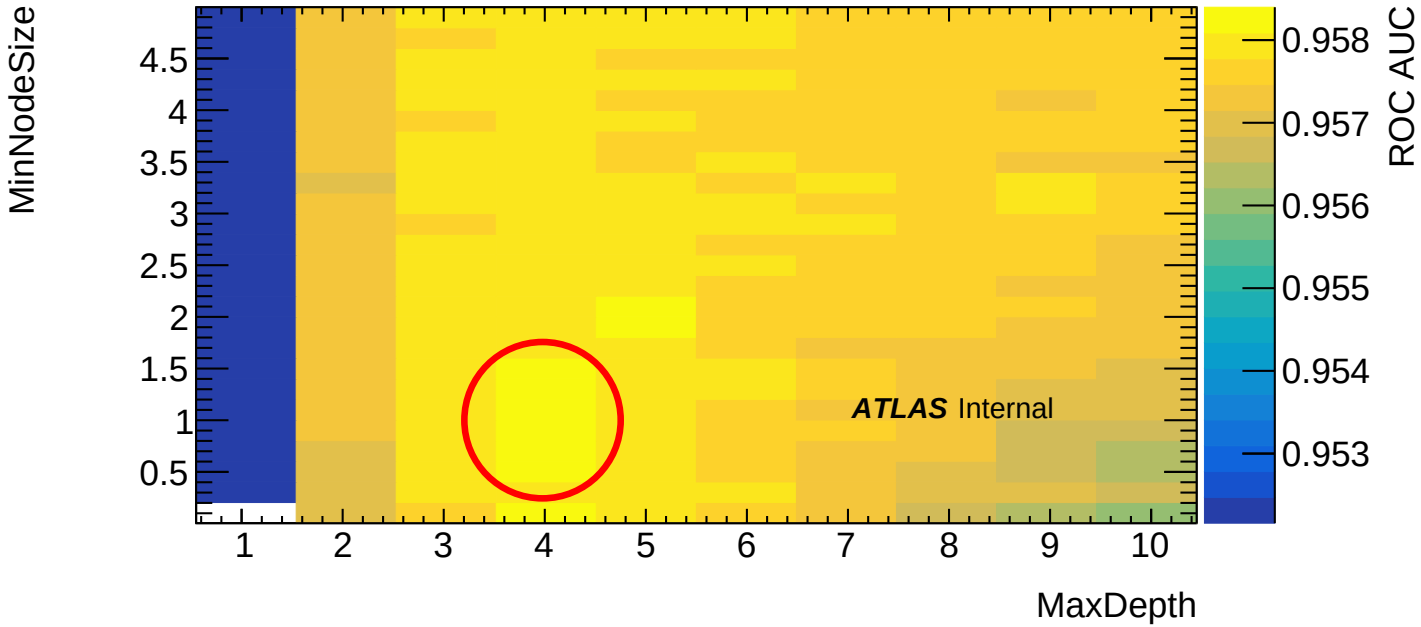


- **Maximal Depth**: How many different layers can an event surpass?
- **Minimal Node Size**: Fraction (%) of events required to be in a leaf

$\Rightarrow$ Are these really the best settings?

Optimization of the BDT-Settings

- Area under ROC-Curve of a sample with $m_{\tilde{t}} = 1$ TeV and $m_{\tilde{\chi}^0} = 1$ GeV



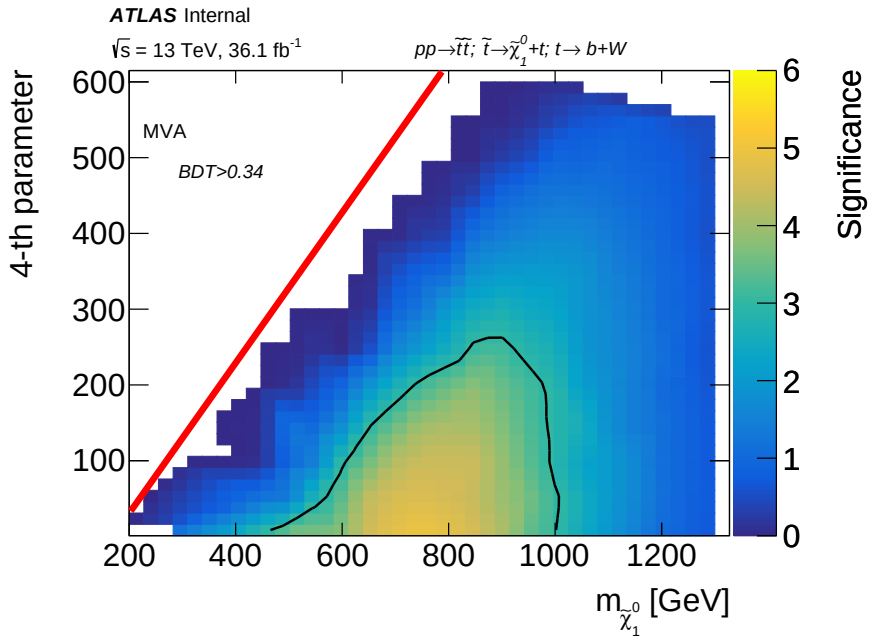
- Area under ROC-Curve tests against overtraining
- High Maximal Depth and Small Minimal Node Size minimizes ROC-area

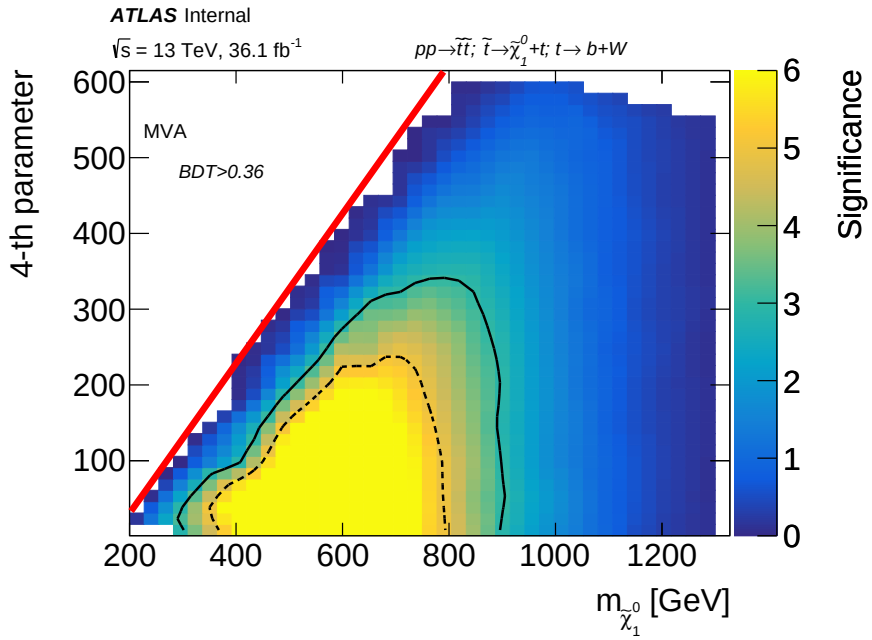
$\Rightarrow$ Sign for overtraining!

- Settings: *MaxDepth* $= 4$, *MinNodeSize* $= 1.5\%$

BDT Training for different models

- Training for model with $m_{\tilde{t}} \geq 1$ TeV
- Training with all samples



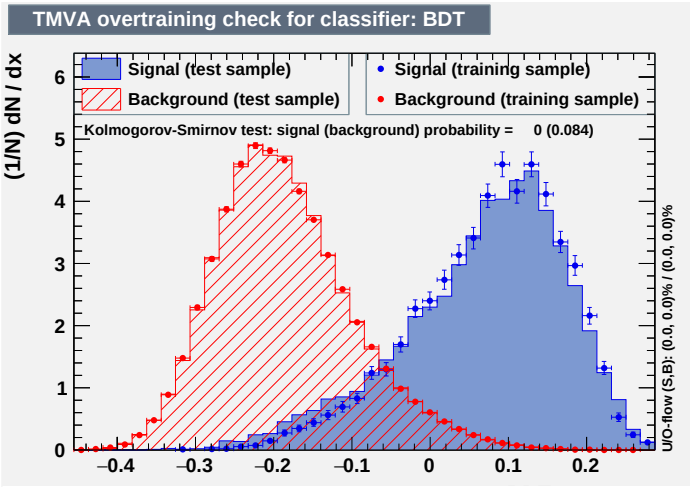


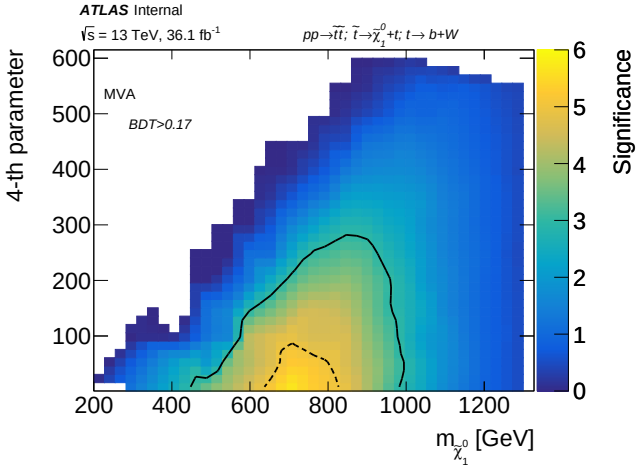
- Significant increase of the sensitive regions in the paramter space, especially in direction towards kinematic border $\left(m_{\tilde{t}} = m_{\tilde{\chi}^0} + m_t\right)$
- For large $m_{\tilde{t}}$: Training with MC-data with $m_{\tilde{t}} \geq 1$ TeV only more promising

Comparing against other MVA methods

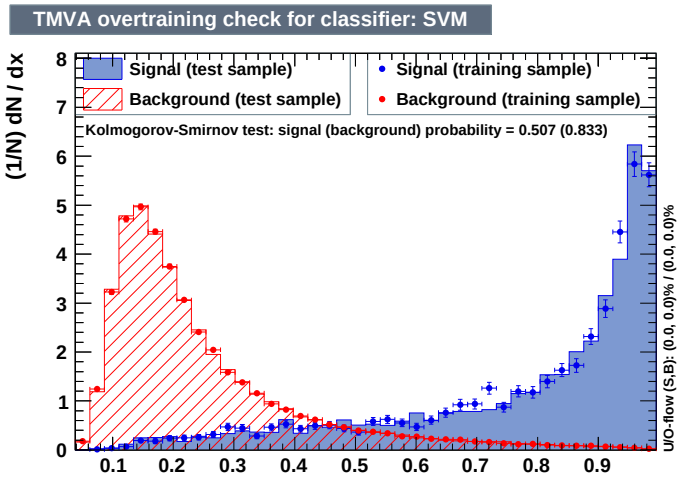
- Training for model with $m_{\tilde{t}} = 1$ TeV and $m_{\tilde{\chi}^0} = 100$ GeV

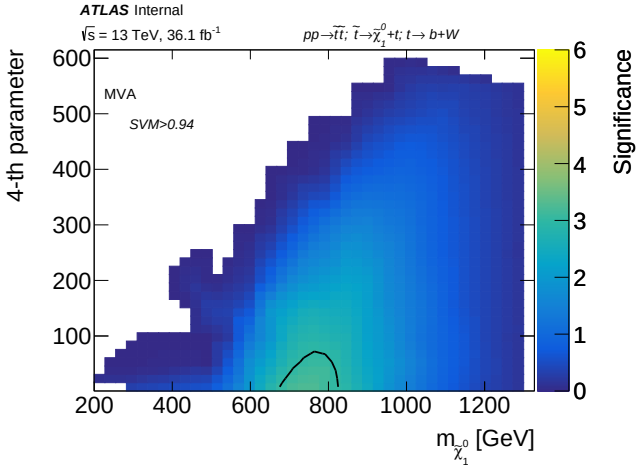
- BDT





- Support Vector Machine (SVM)





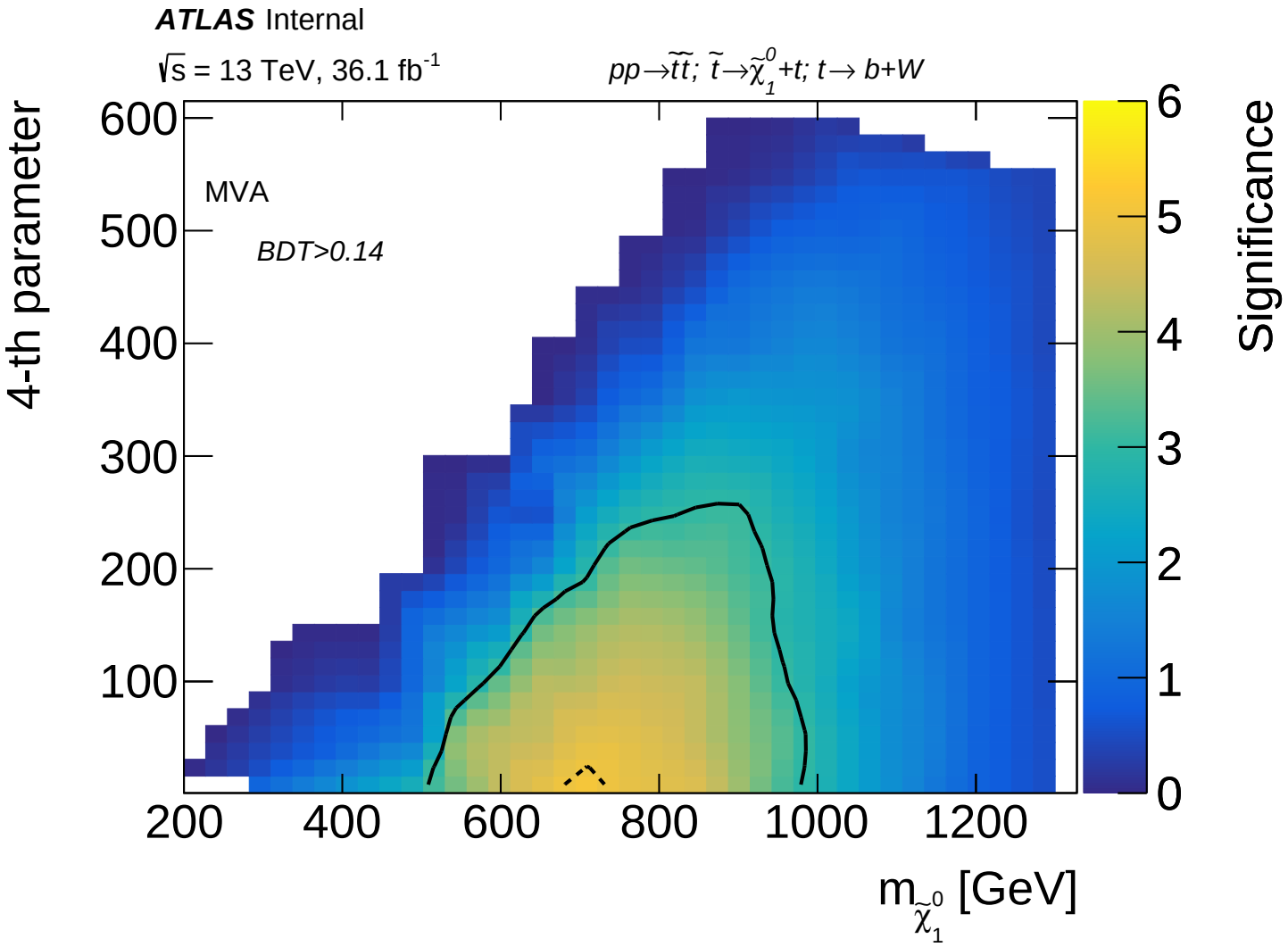
- BDT achieves better significances

Summary & Outlook

- Started to look at MVA methods for increasing sensitivity for $\tilde{t} \to 0\ell$ analysis
- Training with several signal samples seems reasonable
- Current Steps:
- Checking dependency on BDT settings and BDT input variables
- Look at neural networks

BACKUP

Training for model with $m_{\tilde{t}} = 1$ TeV and $m_{\tilde{\chi}^0} = 1$ GeV

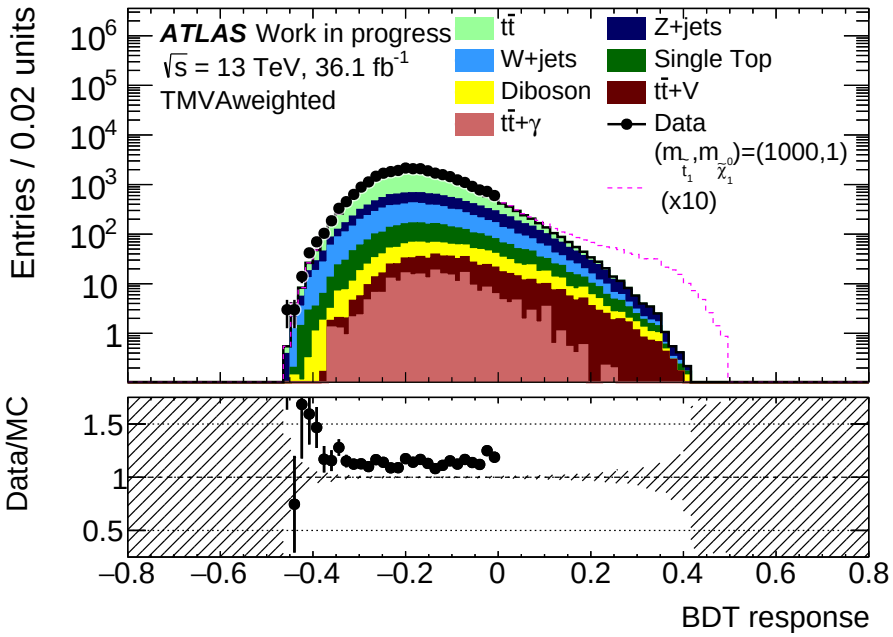


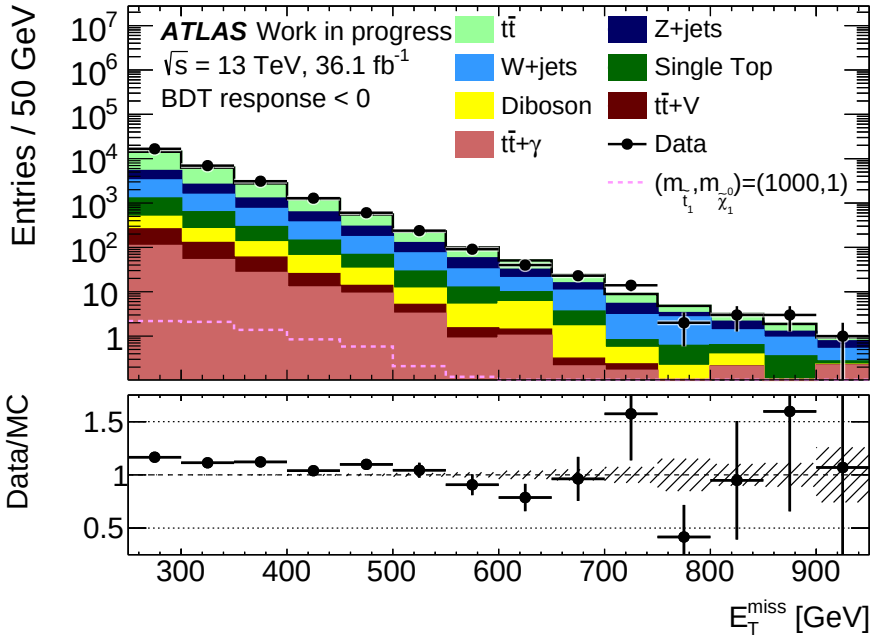
BDT response

- After training: Compare true values y_i with forecast s_i. w_i: weights, with $\sum_i w_i = 1$
- error fraction: $e = \sum_i w_i 1_{s_i \neq y_i}$
- boost factor $\alpha = \beta \cdot ln\left(\frac{1-e}{e}\right)$ with β constant, usually $\beta \in [0, 1]$
- New weights: $w_i \rightarrow w_i \cdot exp\left(\alpha \cdot 1_{s_i \neq y_i}\right)$
- BDT response of event i: $r_i = \frac{\sum_m \alpha_m \left(s_i\right)_m}{\sum_m \alpha_m}$, where

$$(s_i)_m = \begin{cases} 1 & \text{if tree } m \text{ predicts signal} \\ -1 & \text{if tree } m \text{ predicts background} \end{cases}$$

Data-MC agreement





- Data only for BDT-response < 0
- Tiny deviation in normalization between data and MC-forecast

- Good agreement of physical observables in region BDT response < 0