$t\bar{t}H$ at CMS in the l+jets channel

9th Annual Meeting of the Helmholtz Alliance “Physics at the Terascale”, Hamburg

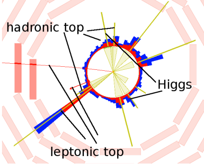
Hannes Mildner | November 17, 2015

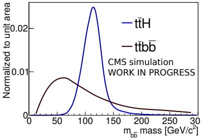
INSTITUT FÜR EXPERIMENTELLE KERNPHYSIK (IEKP)

Lepton+jets analysis strategy

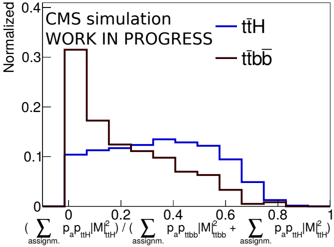
1. Selection
   - 1 isolated lepton
   - At least 4 jets and 2 b-tags
   - Mostly $t\bar{t}$+jets background left
2. Categorization
   - Events split according to jet- and b-tag multiplicities
   - Different background composition in categories
   - Different topologies – different discriminating variables
3. Multivariate analysis
   - Identify differences between signal and backgrounds
     - b-tagging
     - Event shape (HT, sphericity, ...)
     - Reconstruction of resonances
   - Train and optimize BDT in all categories
4. Fit
   - Build statistical model for signal and background-only hypothesis considering systematic uncertainties
   - Fit BDT-output simultaneously in all categories

A discr. variable: $t\bar{t}H/t\bar{t}b\bar{b}$ likelihood





- Interpret jets as quarks, MET as neutrino p_T
- Calculate $t\bar{t}H$ and $t\bar{t}b\bar{b}$ likelihood-ratio $p_{ttH}|M_{tth}|^2/p_{ttbb}|M_{ttbb}|^2$, containing
  - p_{ttH} / p_{ttbb}, the probabilities of the invariant $b\bar{b}$ mass to come from $t\bar{t}b\bar{b}$ / $t\bar{t}H$
  - $|M_{ttH}|^2$ /$|M_{ttbb}|^2$ MadGraph matrix elements, describing whether the $t\bar{t}b\bar{b}$-kinematics are signal- or background-like

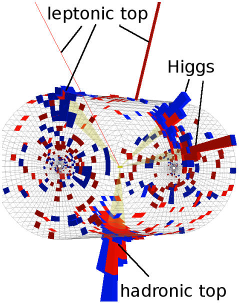


- Sum up all possible assignments
  - Weighted by probability p_a that they are correct
  - Correct assignments have W/top resonances – p_a is, similar to p_{ttH} and p_{ttbb}, evaluated using top and W resonance
- Variable separates signal from $t\bar{t}b\bar{b}$

Boosted category

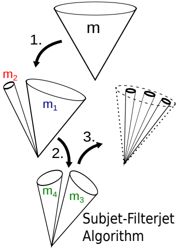
Motivation

- Combinatorial problem in final state
- $t\bar{t}$ + jets background very similar to signal
- Plehn, Salam, Spannowsky (2009): *Fat Jets for a light Higgs*
  - Search for top/Higgs with high transverse momentum
  - Decay products collimated and combinatorics are simplified
  - Background is suppressed



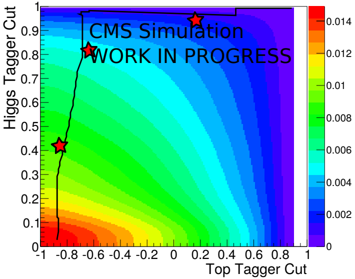
- Our approach
  - Cluster event into C/A 1.5 fat jets
  - Try to identify Higgs or top decay products in jet substructure using
    - **HEP Top-tagger v2** for top-quark
    - **Subjet-filterjet** algorithm for Higgs
  - Use additional category for events with boosted top and Higgs candidate

Boosted category – jet substructure



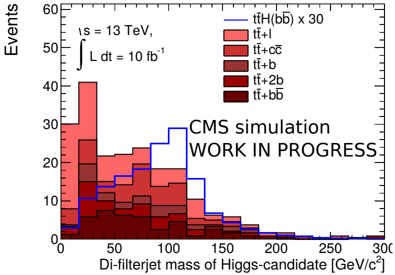
- Both subjet algorithms
  - *Decluster* jets until heavy-particle decay is identified
  - *Filter* by reclustering into filter jets with small radii, removing soft filter jets, and analyzing harder filter jets
  - Construct top/Higgs candidate from filtered jets

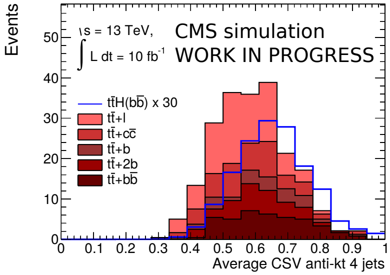
- Top identification: Subjet information (filtered W and top masses, b-tags, Nsubjettiness) combined in BDT top-tag
- Higgs is identified by 2 subjet b-tags
- Cuts on both tags are chosen so that background is minimized for given signal efficiency
- Different working points for boosted category possible



Boosted category – final discrimination

- Boosted category profits from
  - Reduced $t\bar{t}b\bar{b}$ background
  - Good identification of Higgs and top ($\approx$ 40% correct assignment instead of $\approx$ 20% for classic approach with anti-k_t 0.4 jets)
- Two types of variables used for final BDT

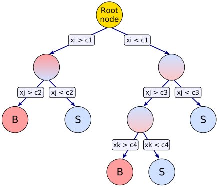


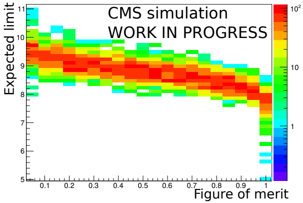


- Fatjet and subjet variables like the Higgs-candidate mass
- Event shape variables and properties of anti-k_t 0.4 jets

MVA analysis – optimization

- Wanted: Multivariate discriminant (BDT) with best sensitivity (expected limit)
- Problem: Limit calculation slow, many possible configurations (BDT parameters, set of variables)





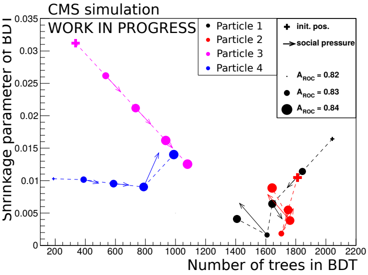
BDT optimization procedure

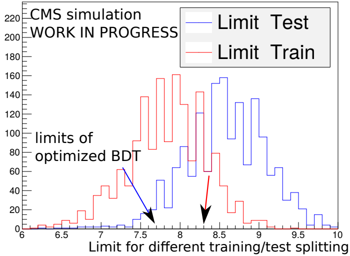
- Choose figure of merit (FOM) correlated to limit (e.g. integral under ROC of BDT)
- Train initial BDTs on a training sample
- Optimize FOM for BDT output on test sample by simultaneously
  - Changing the set of variables (add/remove variables so that FOM improves for resulting BDT)
  - Optimizing BDT configuration (nTrees, shrinkage, nCuts, maxDepth)

MVA analysis – optimization

Karlsruher Institut für Technologie

- Space of BDT configurations and variable sets is scanned using particle swarm algorithm
  (J. Kennedy and R. Eberhart, 1995)
- Every “particle” corresponds to BDT in configuration-space
- BDT configuration is adapted according to best BDT of current and all particles



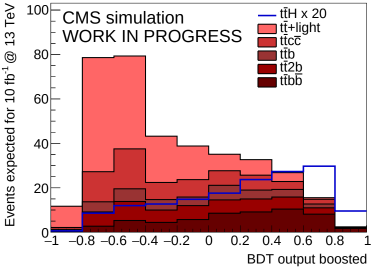


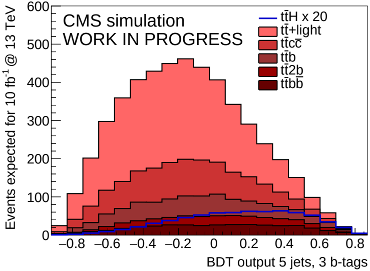
Problem: Over-training/optimization

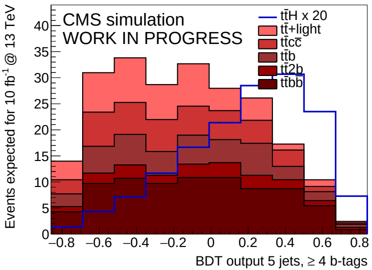
- Overtraining: BDT performs better on training sample than on test sample
- Overoptimization: BDT configuration only yields good result for test sample it was optimized for
- We are careful to avoid both

MVA analysis – final BDTs

- Examples for BDTs
- New: boosted category
- Different background compositions for different jet/b-tag multiplicities

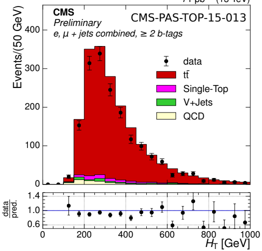






Final words

- Analysis relying on good description of data by MC simulation
- Decent out of the box agreement between 13 TeV data and simulation
- Studying systematic uncertainties and correcting remaining differences



Conclusion

- Direct measurement of top-Higgs coupling in $t\bar{t}H$ important
- Involved in $t\bar{t}H$ search in l+jets channel
- Contributions to analysis with
  - Advanced $t\bar{t}H$ reconstruction techniques
  - Analysis in boosted regime
  - Optimizing BDT analysis
- Preparing analysis of 13 TeV data