Event-level classification of Electroweak SUSY events using ML techniques

Matt Sullivan
matthew.james.sullivan@cern.ch

Supervisors: Yanyan Gao, Monica D'Onofrio

YETI2019

Introduction

- Electroweak SUSY searches at ATLAS
  - Motivation
  - $\tilde{\chi}_1^{\pm}\tilde{\chi}_2^0$ pair-production
- $\tilde{\chi}_1^{\pm}\tilde{\chi}_2^0$ prospects at HL-LHC
  - HL-LHC prospects overview
  - Signal selection with BDTs
  - HL-LHC projected limits
- Current R&D - XGBoost & DNNs:
  - Improving signal-background discrimination
  - Per-background classification

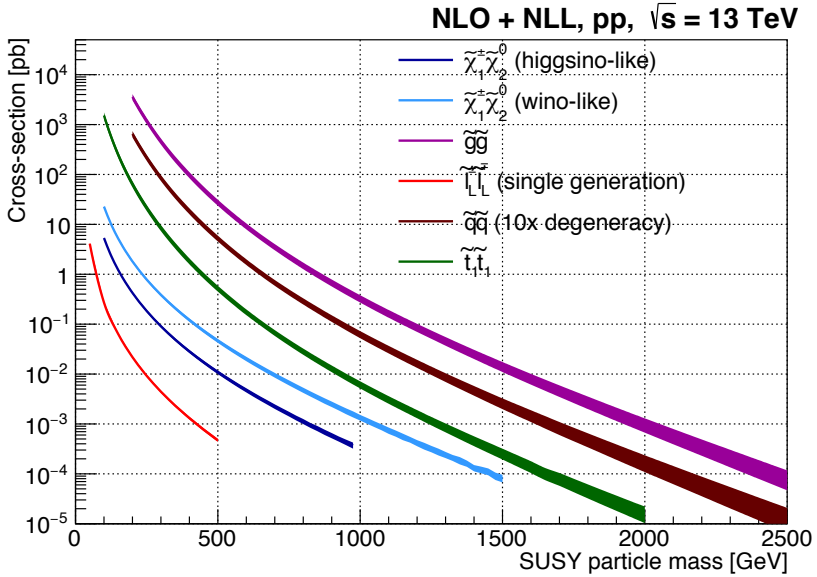
Motivation

Electroweak SUSY searches usually involve direct production of 'EWK-inos' and sleptons:

- EWK-inos: mixtures of SUSY partners to EWK gauge bosons and Higgs

Motivation:

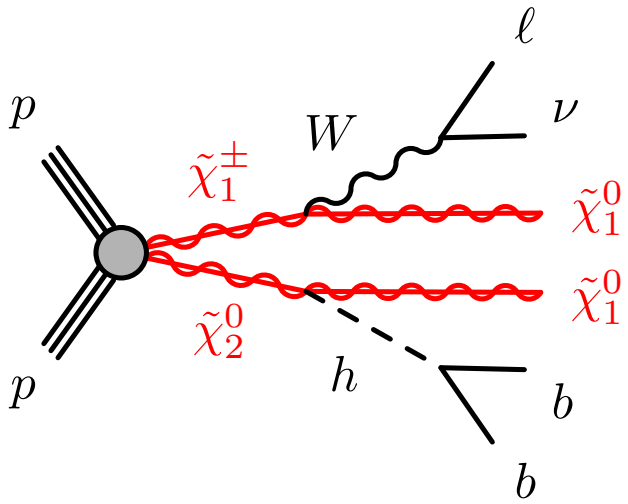
- Low mass EWK-inos may address various open questions (DM, hierarchy problem...)
- EWK SUSY could be dominant SUSY process at LHC if squarks & gluinos are decoupled
- Large CoM energies and datasets → ideal conditions for EWK SUSY searches



Cross-sections for a number of SUSY pair-production processes at $\sqrt{s} = 13$ TeV

$\tilde{\chi}_1^{\pm}\tilde{\chi}_2^0$ pair-production

Simplified model of $\tilde{\chi}_1^{\pm}\tilde{\chi}_2^0$ pair-production

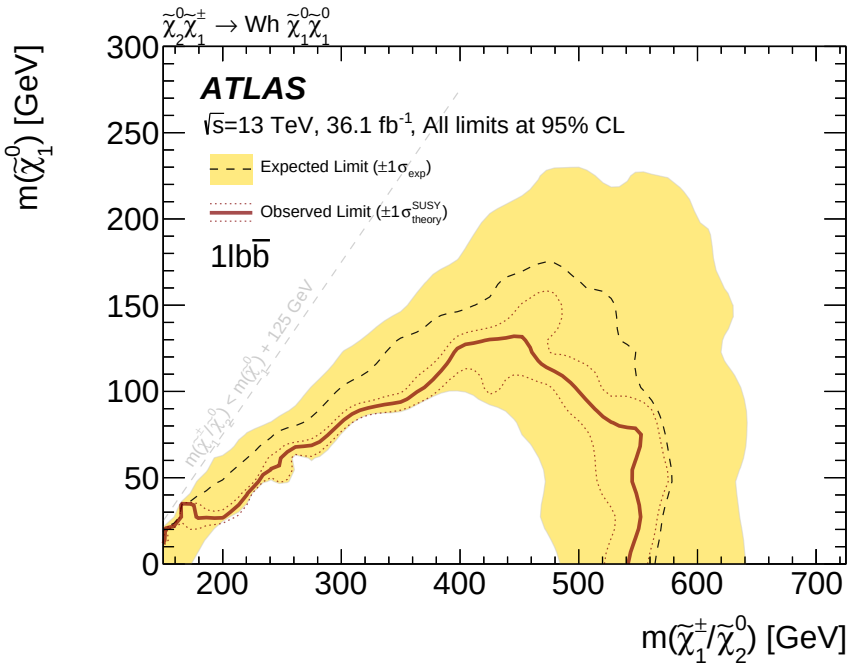


Model assumptions:

- BR($\tilde{\chi}_1^{\pm} \to W\ \tilde{\chi}_1^0$) = 100%
- BR($\tilde{\chi}_2^0 \to h\ \tilde{\chi}_1^0$) = 100%
- $m(\tilde{\chi}_1^{\pm}) = m(\tilde{\chi}_2^0)$
- The Higgs, h, is the lightest SUSY Higgs, with SM-like mass and couplings:
  - $m_h = 125$ GeV
  - BR($h \to b\bar{b}$) = 58%

For 500 GeV $\tilde{\chi}_1^{\pm}/\tilde{\chi}_2^0$, massless $\tilde{\chi}_1^0$:
$\sigma = 46.4$fb

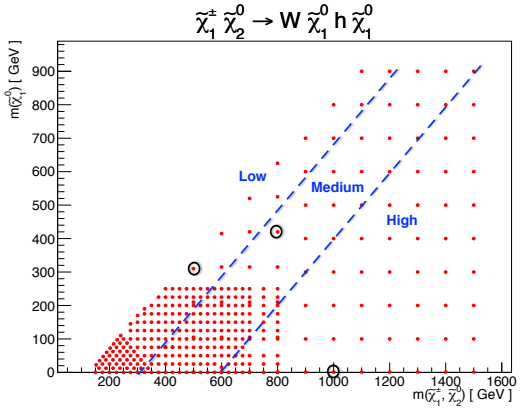
- Current result excludes this process up to m($\tilde{\chi}_1^{\pm}/\tilde{\chi}_2^0$) $\sim$ 550 GeV



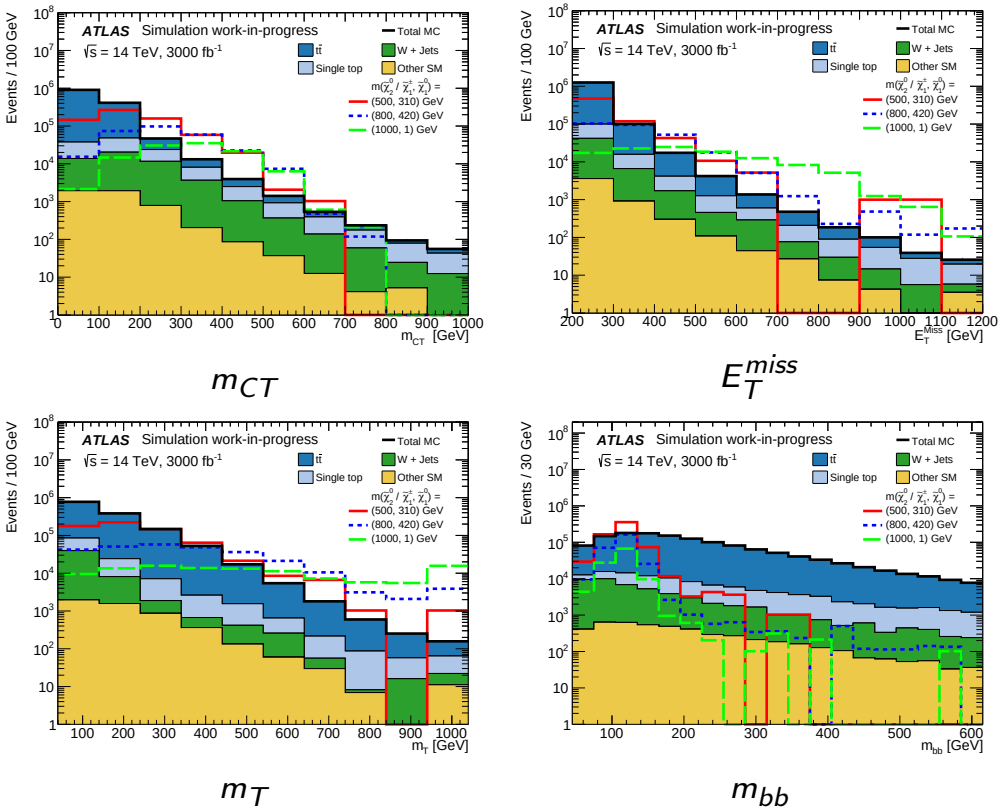
- Combined paper with other final states: ArXiv

HL-LHC prospects overview

- HL-LHC projection: derive expected exclusion (2σ) and discovery (5σ) limits for this channel at ATLAS with HL-LHC dataset!
- Make preliminary selection requiring 1 lepton, 2 jets from b-quarks + $E_T^{miss} > 200$ GeV
- Separate signal models into 3 regions:



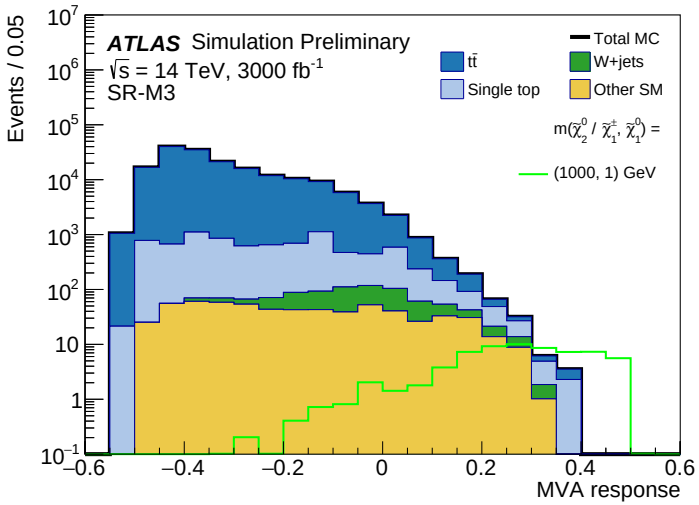
- Train a Boosted Decision Tree (BDT) on signal in each region and $t\bar{t}$ background



Signal yield $\times 1000$. All backgrounds estimated directly from Monte Carlo.

Signal selection with BDTs

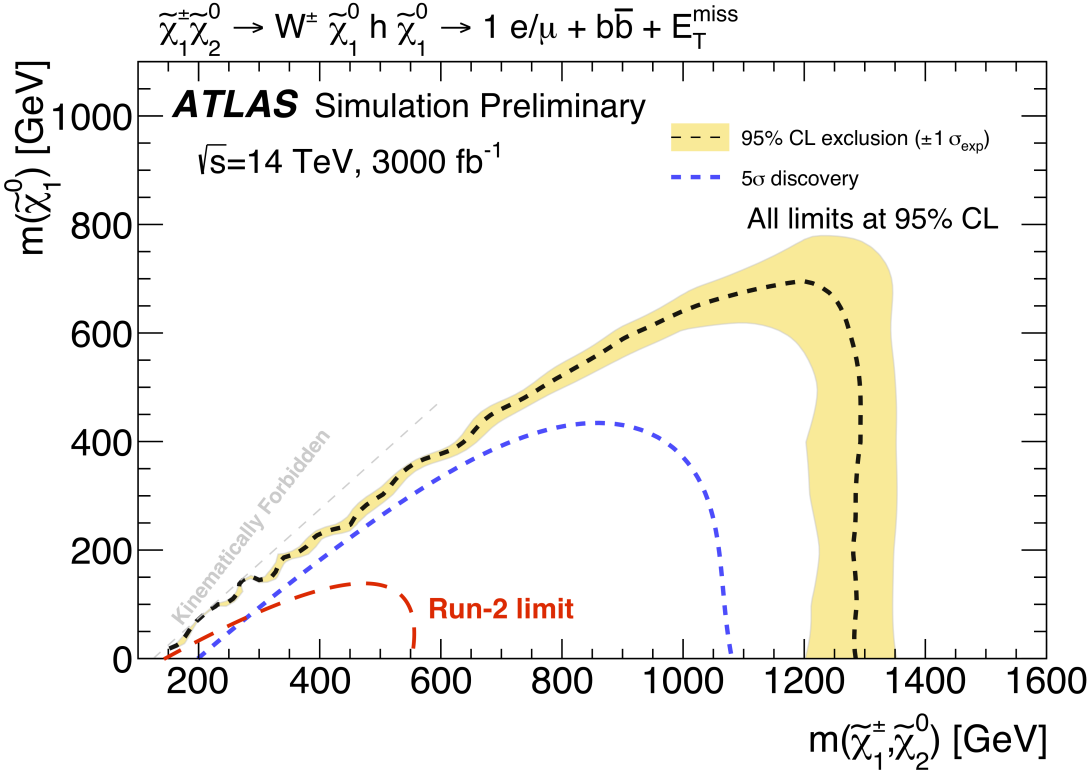
- BDT classifies event as signal or background with prediction $\in [-1, 1]$
  - Prediction $\rightarrow$ 1 = signal-like
  - Prediction $\rightarrow$ -1 = background-like
- Optimise cut on output of BDT using significance (assume 15% uncertainty)



Processes	SR-M1 (>0.25)	SR-M2 (>0.35)	SR-M3 (>0.30)
$t\bar{t}$	38.9 ± 8.4	8.7 ± 3.3	2.5 ± 1.8
single top	28.3 ± 4.8	10.7 ± 3.2	5.4 ± 2.5
W+jets	22.2 ± 5.4	3.0 ± 2.0	2.0 ± 1.8
ttV	5.1 ± 2.4	2.0 ± 1.4	1.0 ± 1.0
Diboson	2.0 ± 2.0	-	-
total background	96.5 ± 11.8	24.4± 5.2	10.9 ± 3.4
$m(\tilde{\chi}_1^{\pm}/\tilde{\chi}_2^0, \tilde{\chi}_1^0) = (500, 300)$ GeV	20.7 ± 4.8	4.6 ± 2.3	1.0 ± 1.0
$m(\tilde{\chi}_1^{\pm}/\tilde{\chi}_2^0, \tilde{\chi}_1^0) = (800, 420)$ GeV	44.3 ± 2.3	33.6 ± 2.0	21.2 ± 1.6
$m(\tilde{\chi}_1^{\pm}/\tilde{\chi}_2^0, \tilde{\chi}_1^0) = (1000, 1)$ GeV	32.2 ± 1.8	31.9 ± 1.8	28.9 ± 1.7

HL-LHC projected limits

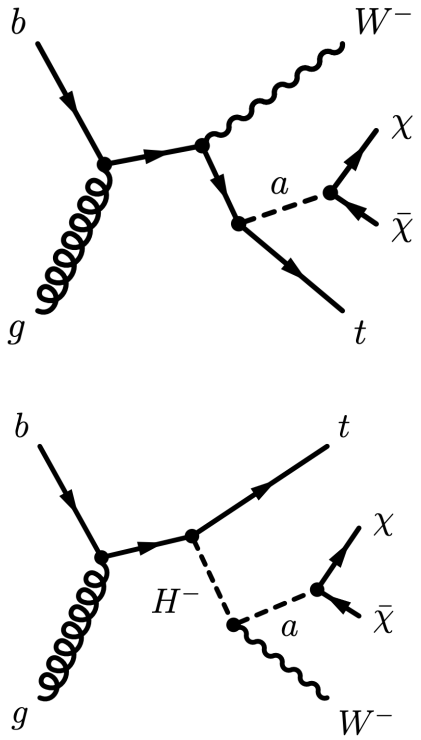
- Systematic uncertainties extrapolated from Run-2:
  - Theory modelling $\times 1/2$
  - Experimental $\times 1/3 - \times 1$
- HL-LHC expected exclusion limit (2σ) up to $m(\tilde{\chi}_1^{\pm}/\tilde{\chi}_2^0) = \sim 1280$ GeV
- HL-LHC expected discovery limit (5σ) up to $m(\tilde{\chi}_1^{\pm}/\tilde{\chi}_2^0) = \sim 1080$ GeV



This result included in CERN HL/HE-LHC Yellow Report: ArXiv

R&D: New techniques

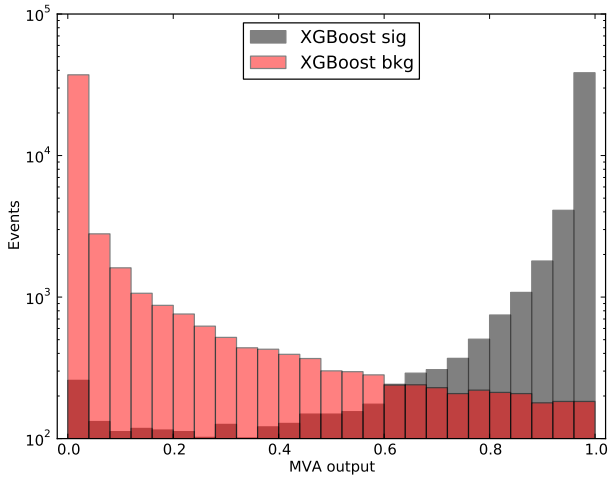
- HL-LHC analysis used BDTs implemented with TMVA
  - Quick, easy, established in HEP
- Many new packages & techniques gaining interest from HEP community:
  - XGBoost: eXtreme Gradient Boosted decision trees
  - Keras: High-level Python interface to Tensorflow neural networks
- Use these techniques to build better classifiers:
  - $1\ell b\bar{b}$ analysis: target low mass splitting region (experimentally difficult!)
  - $Wt + E_T^{miss}$ 2HDM+a analysis: Stronger background rejection through classification

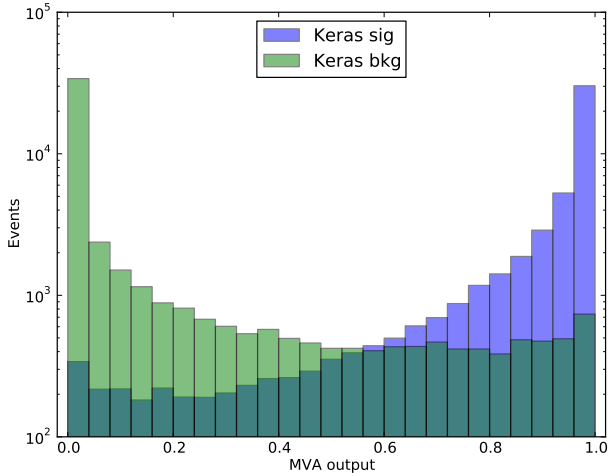


$Wt + E_T^{miss}$ 2HDM+a diagrams

R&D: Signal/background classification

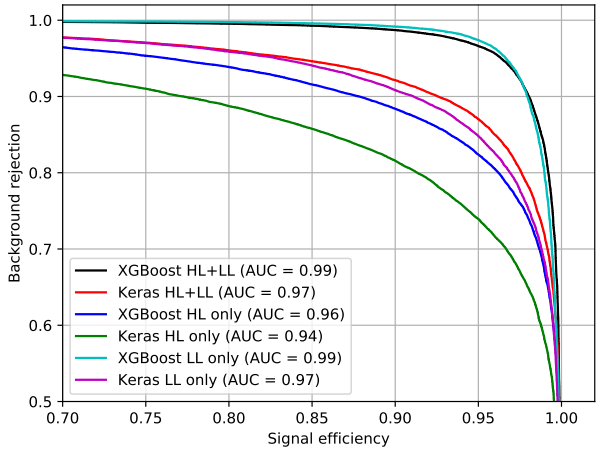
- Binary classification:
  - 1: m($\tilde{\chi}_1^{\pm}/\tilde{\chi}_2^0$, $\tilde{\chi}_1^0$) = (400, 200) GeV
  - 0: $t\bar{t}$ background
- Train an XGBoost classifier and a Deep Neural Network (DNN)
  - 50k events from each class split 50:50 for train/test
- Use following variables as inputs:
  - High-level (HL): m_{bb}, m_{CT}, m_T, E_T^{miss}, $\Delta R(b1, b2)$
  - Low-level (LL): $p_T^{\ell 1}$, $\eta^{\ell 1}$, $\phi^{\ell 1}$, p_T^{b1}, η^{b1}, ϕ^{b1}, p_T^{b2}, η^{b2}, ϕ^{b2}
- 3 sets of trainings: HL only, LL only, <u>HL+LL</u>

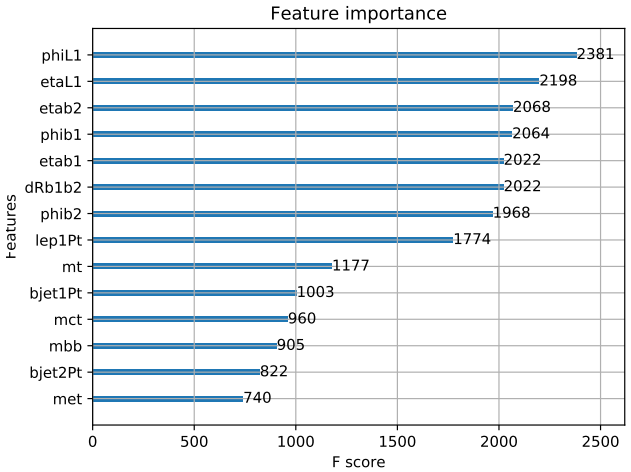




R&D: Classifier performance & feature importance

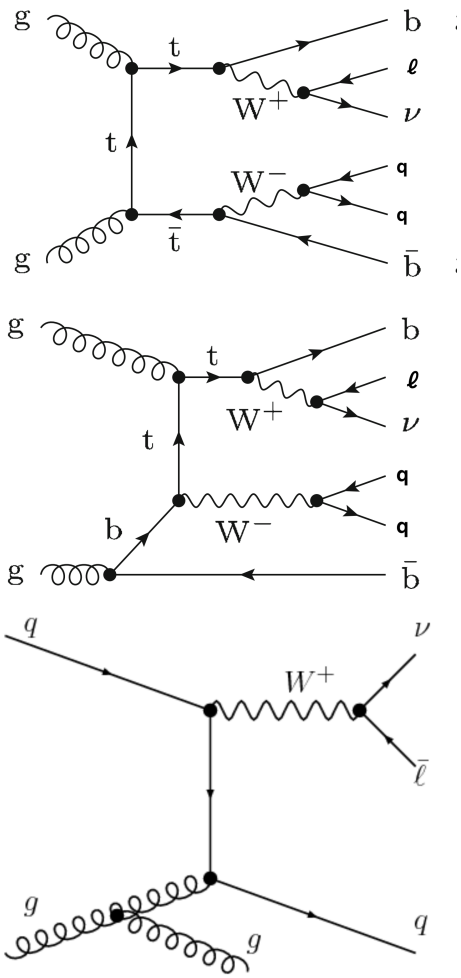
- All trainings converge quickly w/GPU acceleration
  - XGBoost $\mathcal{O}(10s)$
  - DNN $\mathcal{O}(100s)$
- XGBoost classifier has higher AUC than DNN - stats limited by signal MC
- Low-level variables carry much discrimination power
  - XGBoost feature importance = times feature is in final decision tree
- Additional features: $\phi^{E_T^{miss}}$, $m_{\ell b1}$, $m_{\ell b2}$





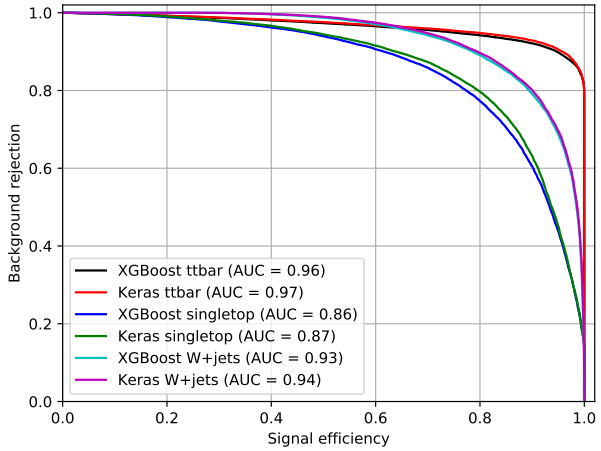
R&D: Classification of multiple backgrounds

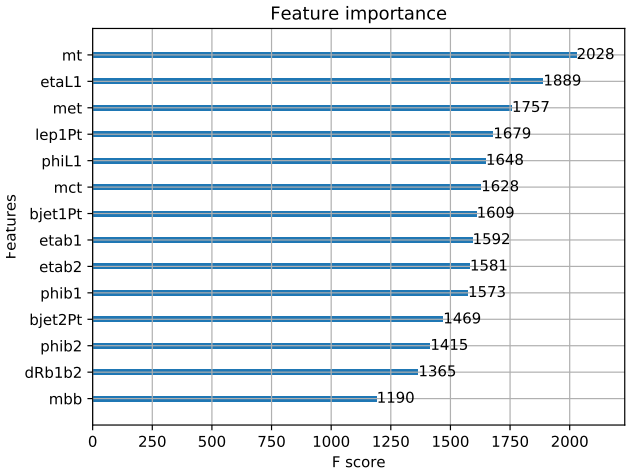
- Signal vs $t\bar{t}$ idealised scenario
  - Low $m_T \rightarrow W$+jets dominant
  - High $m_T \rightarrow Wt$ dominant
- Initial studies into per-background classification:
  - Same features as previously shown used to train
  - 50k events from each background class split 50:50 for train/test
- 3 classes:
  - $t\bar{t}$, Wt, W+jets



R&D: Multiclass performance

- First attempt at classifying different backgrounds underway
- What we learn:
  - Input features allow accurate classification of $t\bar{t}$
  - Wt classification not as accurate - additional features, more training events?
- Feature importance very different!
  - m_T & m_{CT} used in previous analyses as background discriminators





Conclusions

- Strong motivations for $\tilde{\chi}_1^{\pm}\tilde{\chi}_2^0$ searches
- HL-LHC $\tilde{\chi}_1^{\pm}\tilde{\chi}_2^0$ prospects analysis:
  - Signal selection done with BDTs
  - Exclusion limit: 1280 GeV
  - Discovery limit: 1080 GeV
- Currently studying potential improvements to Run-2 analysis:
  - Signal selection using XGBoost/DNN
  - Classifying individual backgrounds using multiclass XGBoost/DNN

Thanks for listening!

Any questions?

Backup

Kinematic variable definitions

$$\Delta R = \sqrt{(\Delta\eta^2 + \Delta\phi^2)} \tag{1}$$

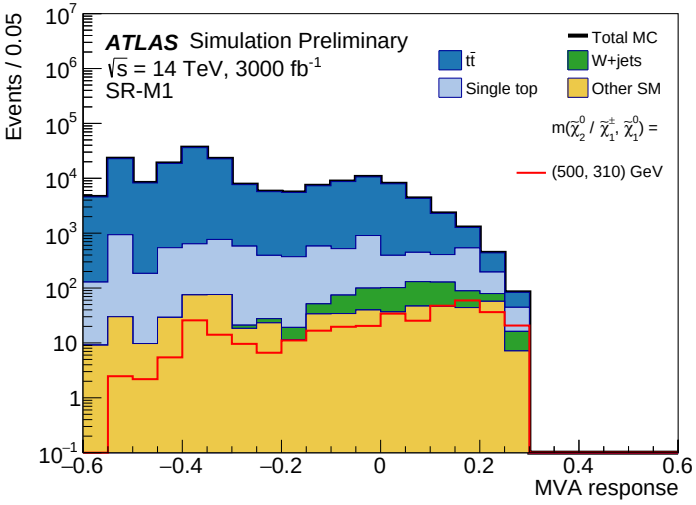
$$m_T = \sqrt{2p_T^{\ell} E_T^{miss} (1 - \cos\Delta\phi)} \tag{2}$$

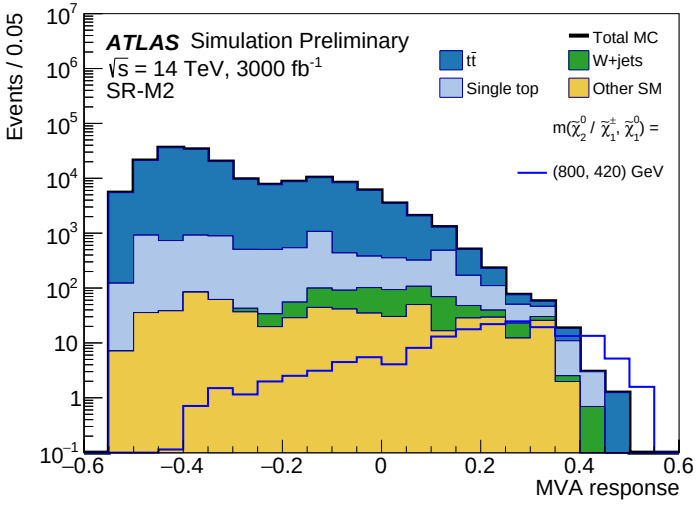
$$m_{CT} = \sqrt{\left(E_T^{b_1} + E_T^{b_2}\right)^2 - \left(\mathbf{p}_T^{b_1} - \mathbf{p}_T^{b_2}\right)^2} \tag{3}$$

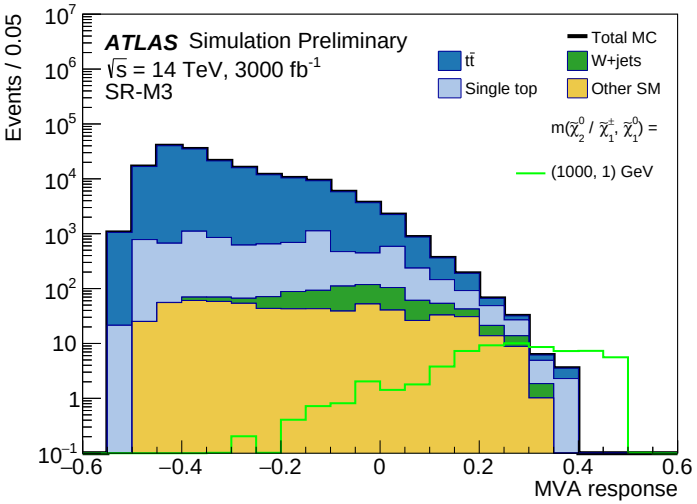
Boost-corrected m_{CT}: https://arxiv.org/abs/0910.0174

$$E_T^{miss} = \left|\mathbf{p}_T^{miss}\right| \tag{4}$$

BDT outputs


ATLAS Simulation Preliminary
√s = 14 TeV, 3000 fb-1
SR-M1
tt̄
Single top
Total MC
W+jets
Other SM
m(χ̃02 / χ̃1±, χ̃01) =
(500, 310) GeV
Events / 0.05
MVA response


ATLAS Simulation Preliminary
√s = 14 TeV, 3000 fb-1
SR-M2
tt̄
Single top
Total MC
W+jets
Other SM
m(χ̃02 / χ̃1±, χ̃01) =
(800, 420) GeV
Events / 0.05
MVA response


ATLAS Simulation Preliminary
√s = 14 TeV, 3000 fb-1
SR-M3
tt̄
Single top
Total MC
W+jets
Other SM
m(χ̃02 / χ̃1±, χ̃01) =
(1000, 1) GeV
Events / 0.05
MVA response

DNN parameter scan