Machine learning at CERN: ATLAS, LHCb, and more

Steven Schramm,

on behalf of the ATLAS and LHCb collaborations

ICHEP 2018, Seoul

July 5, 2018

Overview

- Machine Learning (ML) is ubiquitous in modern HEP
  - Historically used for final analysis selections
  - Increasingly finding its way into lower-level tasks
- Modern HEP is also starting to use more recent ML developments
  - Traditionally: Boosted Decision Trees (BDTs), Neural Networks (NNs)
  - More recently: "deep learning" explosion (DNNs, RNNs, CNNs, ...)
- I will focus on ML usage outside of final analysis selections
  - At the end, I will give a few examples ML used in key analyses

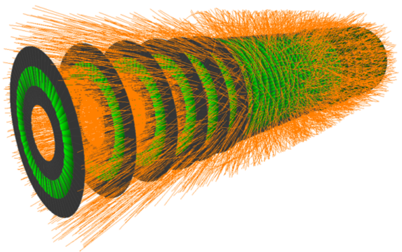
- **Object reconstruction, calibration, and similar**
- Object identification (tagging)
- Simulation
- Automation
- ML usage in key results

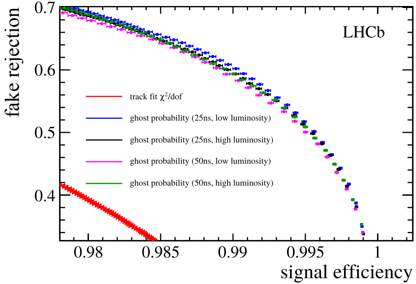
Track reconstruction

Left: TrackML challenge

Right: PUB-2017-011

- ATLAS and LHCb track reconstruction both investigating ML
- ATLAS: prepare for extreme expected pileup conditions
    - Recent TrackML challenge to speed up track reconstruction
    - Discussed more in the next talk
- LHCb: Efficiently suppress fake tracks within the trigger
    - Significant gains in fake track rejection reduce combinatorics
    - Key piece enabling use of offline track algorithms in trigger

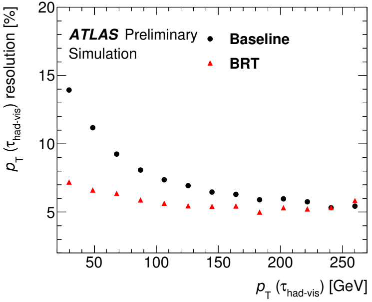


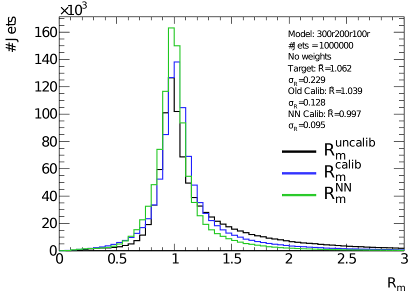


Object calibration (regressions)

Left: CONF-2017-029
Right: Ennis (2017)

- ATLAS hadronic τ calibration is now ML-based
  - Boosted Regression Tree (BRT) significantly improves p_{T}^{τ} resolution
  - BRT adds track and decay mode info, removes most of p_{T} dependence
- Jet mass calibration can also be improved with ML
  - Comparing a normal calibration vs neural network regression
  - NN calibration has superior mass resolution $\Longrightarrow$ many benefits

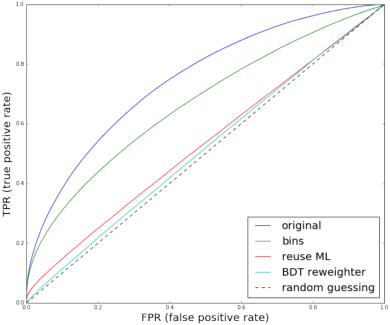


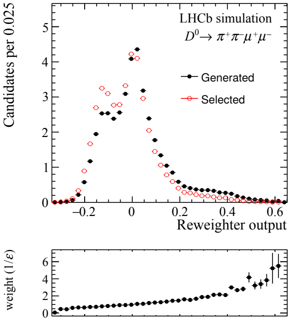


Differences between data and MC

Left: Rogozhnikov (2016)
Right: PAPER-2018-020

- Data and MC can often be told apart based on modelling differences
  - Train a BDT to reweight MC to remove the differences!
  - After this is done, large majority of discrimination power removed
- Used in many LHCb publications to fix modelling differences
  - Example of the BDT reweighter applied to $D^0 \to \pi^+\pi^-\mu^+\mu^-$





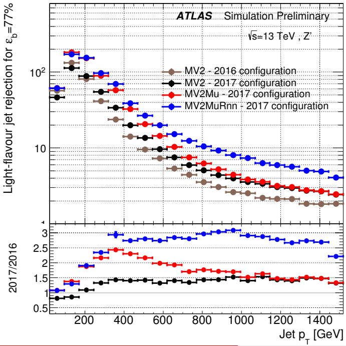
- Object reconstruction, calibration, and similar
- **Object identification (tagging)**
- Simulation
- Automation
- ML usage in key results

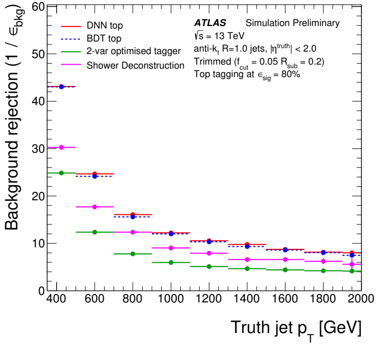
Taggers with high-level inputs

Left: PUB-2017-013

Right: CONF-2017-064

- Long history of using BDTs for b-tagging in ATLAS
  - Done using "high-level" variables = derived quantities (not raw tracks)
  - Latest iteration continues to improve identification capabilities
- Both BDTs and DNNs tested for top-jet tagging
  - Similar performance for BDTs and DNNs using high-level features
  - Consistent gain of factor of ~ 2 with respect to simple taggers

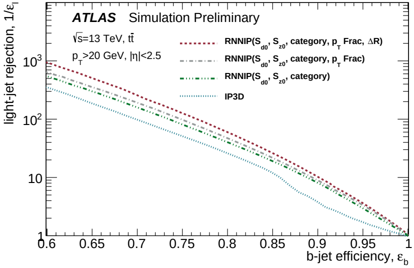


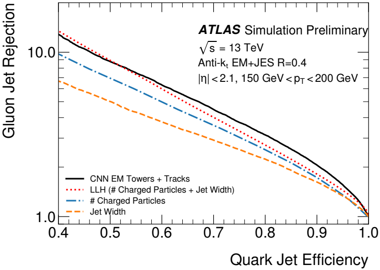


Tagging with lower-level inputs

Left: PUB-2017-003
Right: PUB-2017-017

- "Deep learning" promises gains by using low-level inputs
- Use of tracks as input to RNN for impact parameter b-tagging
  - Recurrent NN considers up to 15 tracks for each jet (for training speed)
  - ~50% gain using same variables vs likelihood, more with extra variables
- Use of jet images and CNNs for quark/gluon tagging
  - Convolutional NN uses tracks and calorimeter towers as inputs
  - Moderate gain over likelihood combination for high quark efficiency



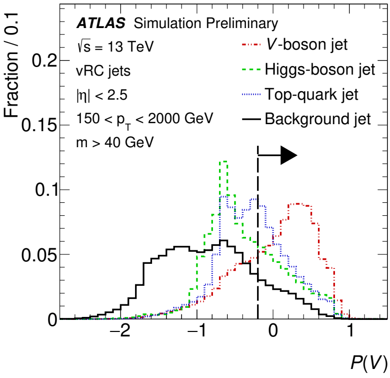


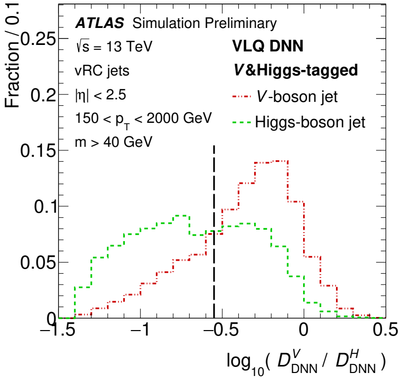
Multi-class tagging

Left: EXOT-2017-14 ATLAS EXPERIMENT

Right: EXOT-2017-14 ATLAS EXPERIMENT

- Sometimes there are multiple similar objects to differentiate
  - Boosted W/Z, Higgs, and top jets have non-negligible overlap
- Multi-class DNN trained on a mix of low-level and high-level variables
  - First stage: discriminate vs QCD for each signal type independently
  - Second step: likelihoods of discriminants for signal ambiguity resolution



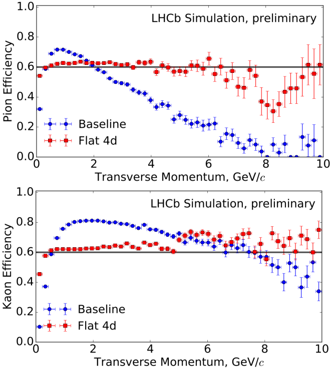


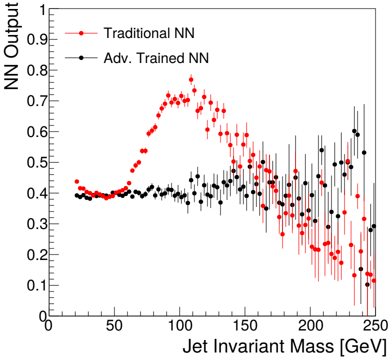
Decorrelated taggers

Left: Lucio (2018)

Right: Shimmin et al (2017)

- ML can remove correlations / flatten distributions [methods: 1, 2, 3]
- Uniform BDT (uBoost) flattens vs 4 variables at once
  - Provides unbiased background determination for different channels
- Adversarial NN (ANN) puts two NNs against each other
  - Removes sculpting of the jet mass distribution when rejecting QCD

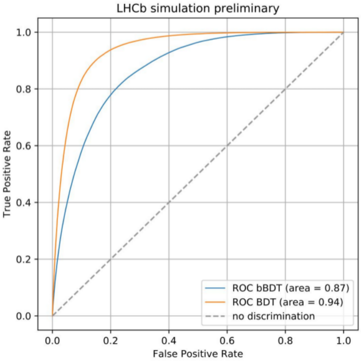


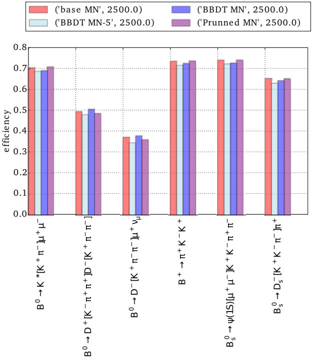


Compressing for the trigger

Left: Dendek (2018)
Right: Likhomanenko et al (2015)

- BDTs and NNs are typically "fast" to evaluate, but "fast" is relative
  - In the trigger, every bit of speed is needed
- Bonsai BDTs binarize the nodes for faster evaluation
  - Left: some degradation in performance from this simplification
  - Right: loss in performance is minimal in real trigger use-case



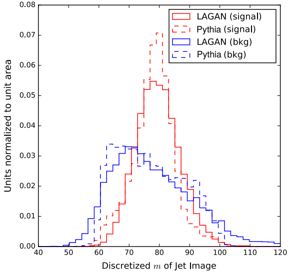


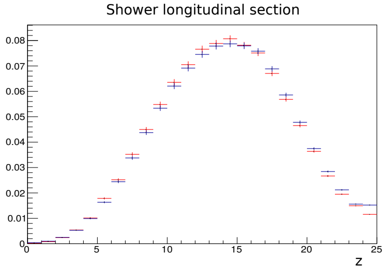
- Object reconstruction, calibration, and similar
- Object identification (tagging)
- **Simulation**
- Automation
- ML usage in key results

Speeding up simulations

Left: de Oliveira et al (2017)
Right: Vallecorsa (2017)

- Rigorous calorimeter simulation is very computationally expensive
  - Typically a "fast simulation" exists for where lower precision is needed
  - ML is a promising means of extending/improving fast simulation
- Idea: calorimeter shower is a 3D image, use ML to generate images
  - Left: jet mass for W vs QCD in independent studies
  - Right: electron 3D longitudinal depth for Geant4 vs ML in GeantV
- Two dedicated talks later on LHCb and ATLAS ML-based simulation
  - Also a poster by ATLAS on this topic



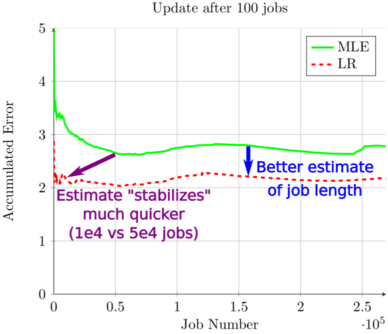


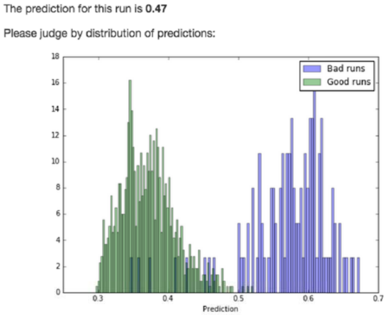
- Object reconstruction, calibration, and similar
- Object identification (tagging)
- Simulation
- **Automation**
- ML usage in key results

Job scheduling and data quality

Left: Rauschmayr et al (2014)
Right: Adinolfi et al (2017)

- Investigated ML for job submission and resource demand
  - ML (linear regression) does better job than likelihood average
  - Not used in production as dominant factor is MC or data
- ML being used for automated data quality monitoring
  - Running in production but not yet used for decisions
  - Not enough bad runs yet to validate that it is working as intended



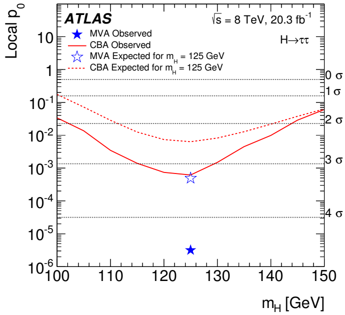


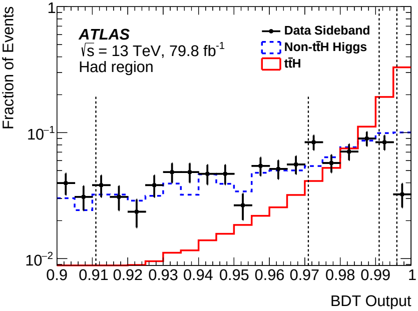
- Object reconstruction, calibration, and similar
- Object identification (tagging)
- Simulation
- Automation
- **ML usage in key results**

ATLAS results

Left: HIGG-2013-32
Right: HIGG-2018-13

- BDTs and NNs are used throughout many high-profile ATLAS results
  - Two NNs were used in the Higgs boson discovery paper (not shown)
- BDTs played a key role in the first evidence for $H \to \tau\tau$
  - BDT increases significance by $\sim 1\sigma$ compared to cut-based
- BDTs are also key to the recent *ttH* observation
  - Two separate BDTs are used, with multiple channels and categories

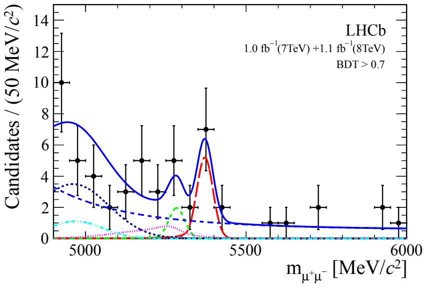


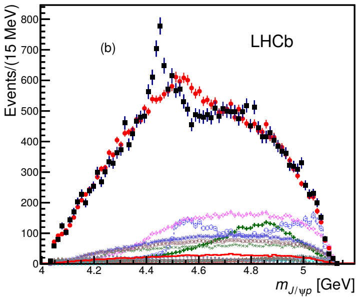


LHCb results

Left: PAPER-2012-043
Right: PAPER-2015-029

- The first evidence for $B_s^0 \to \mu^+\mu^-$ made use of BDTs
  - Two-stage BDT discriminant for high signal efficiency
- The first observation of pentaquarks also used BDTs
  - Separate Λ_b^0 signal from backgrounds
- These are only two of many examples of LHCb results using ML
  - BDTs are frequently used and 2/3 of results use the bonsai BDT trigger





Summary

- ML usage is increasing for both ATLAS and LHCb
  - Historically used to improve sensitivity in flagship analyses
  - Increasingly used for other lower-level tasks
    - Reconstruction, tagging, simulation, automation, and more
- The set of ML techniques used is also growing
  - BDTs and simple NNs are still quite common
  - However, deep learning is increasingly used in many forms
    - DNNs, CNNs, RNNs, GANs, ANNs, VAEs, ...
- This trend is likely to continue as datasets grow
  - The HL-LHC in particular will deliver a huge dataset for ML analysis

- LPCC Inter-experimental Machine Learning working group (IML) is open to all who are interested in discussing the usage of ML in HEP
  - $>$650 people on the mailing list, $>$300 registrants at the 2018 workshop

Backup Material

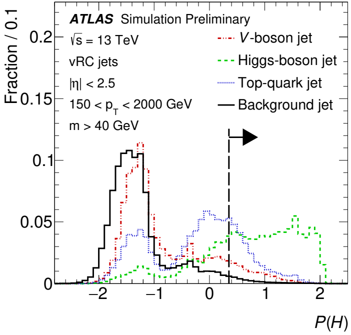
Multi-class tagging, step 1

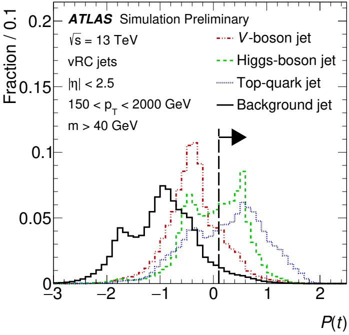
Left: EXOT-2017-14
Right: EXOT-2017-14

$$P(V) = \log_{10}\left(\frac{D^{V}_{\mathrm{DNN}}}{0.9\cdot D^{\mathrm{background}}_{\mathrm{DNN}} + 0.05\cdot D^{t}_{\mathrm{DNN}} + 0.05\cdot D^{H}_{\mathrm{DNN}}}\right)$$

$$P(H) = \log_{10}\left(\frac{D^{H}_{\mathrm{DNN}}}{0.9\cdot D^{\mathrm{background}}_{\mathrm{DNN}} + 0.05\cdot D^{V}_{\mathrm{DNN}} + 0.05\cdot D^{t}_{\mathrm{DNN}}}\right)$$

$$P(V) = \log_{10}\left(\frac{D^{t}_{\mathrm{DNN}}}{0.9\cdot D^{\mathrm{background}}_{\mathrm{DNN}} + 0.05\cdot D^{V}_{\mathrm{DNN}} + 0.05\cdot D^{H}_{\mathrm{DNN}}}\right)$$





Multi-class tagging, step 2

Left: EXOT-2017-14
Right: EXOT-2017-14

$$P(V) = \log_{10}\left(\frac{D^{V}_{\mathrm{DNN}}}{0.9 \cdot D^{\mathrm{background}}_{\mathrm{DNN}} + 0.05 \cdot D^{t}_{\mathrm{DNN}} + 0.05 \cdot D^{H}_{\mathrm{DNN}}}\right)$$

$$P(H) = \log_{10}\left(\frac{D^{H}_{\mathrm{DNN}}}{0.9 \cdot D^{\mathrm{background}}_{\mathrm{DNN}} + 0.05 \cdot D^{V}_{\mathrm{DNN}} + 0.05 \cdot D^{t}_{\mathrm{DNN}}}\right)$$

$$P(V) = \log_{10}\left(\frac{D^{t}_{\mathrm{DNN}}}{0.9 \cdot D^{\mathrm{background}}_{\mathrm{DNN}} + 0.05 \cdot D^{V}_{\mathrm{DNN}} + 0.05 \cdot D^{H}_{\mathrm{DNN}}}\right)$$