Transformer Neural Networks for Large Radius Jet Classification and Regression for Boosted Higgs Bosons at the ATLAS Detector

Andrius Vaitkus

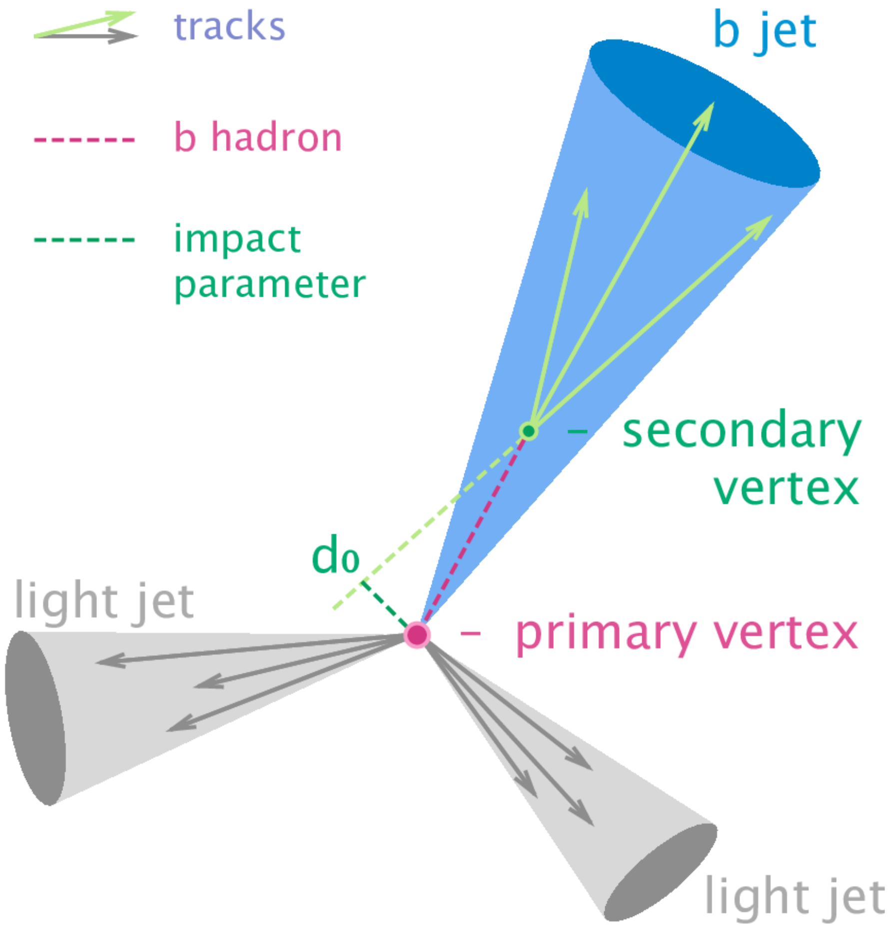
IoP APP, HEPP & NP 2024

andrius.vaitkus.16@ucl.ac.uk

b-tagging at ATLAS

- Identifying *b*-jets (***b*-tagging**) is important for studying heavy particles:
    - H→bb, t→Wb are main decay modes
    - Separating *b*-jets from overwhelming background is crucial

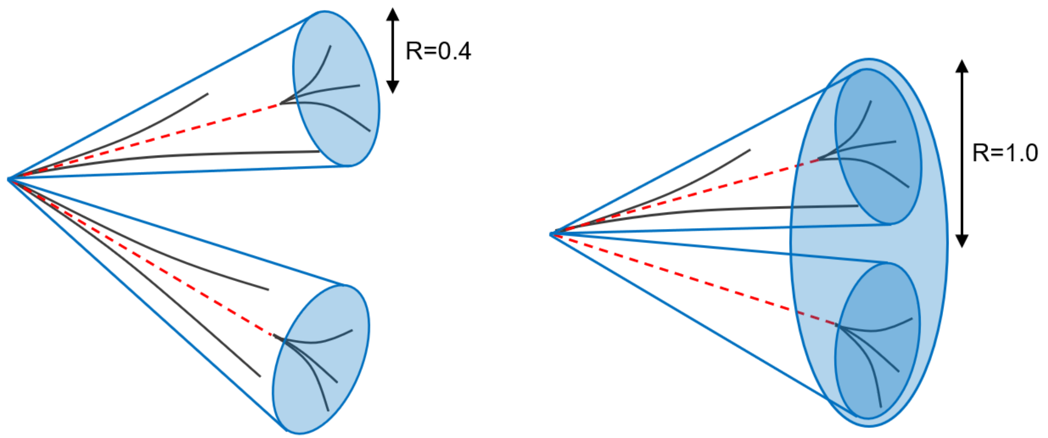
- *b*-tagging takes advantage of unique ***b*-quark properties**:
    - Long lifetime (~10^{-12} s) of *b*-hadrons
      → displaced vertices
    - High mass (~5 GeV)
      → higher p_T
      → higher multiplicity
      → large transverse impact parameter



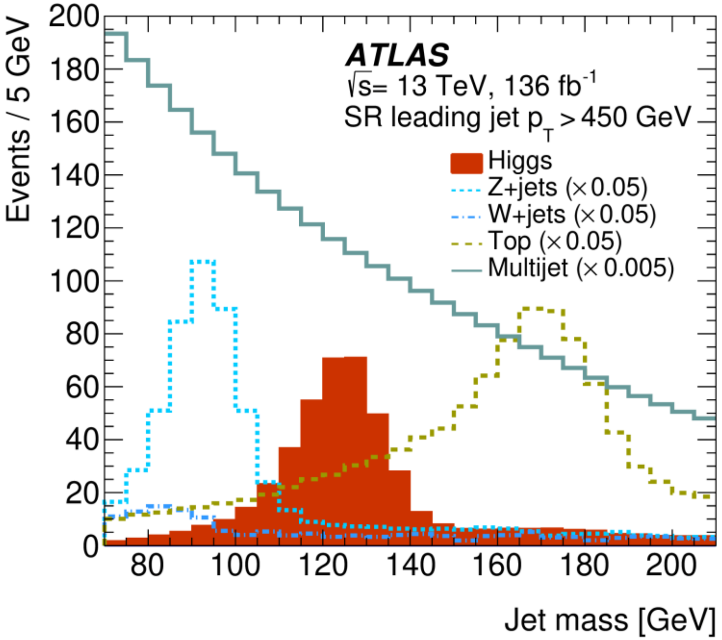
Large-R Jet H→bb Tagging

- H→bb tagging is especially important for identifying **boosted Higgs bosons** decaying to $b\bar{b}$
  - Boosted H→bb events could uncover signs of BSM physics
  - For high p_T, *b*-jets are more collimated
    → treated as one Large-R jet
- Goal: **separate $\mathbf{H}(\boldsymbol{b\bar{b}})$ signal** from $H(c\bar{c})$, hadronic top decays, QCD
  - H→cc tagging work also ongoing but not covered in this talk

- Improving our measurements of **jet kinematic variables** is also crucial
  - Can be used in calibration efforts or searches
  - Most important variables to learn: jet mass, p_T

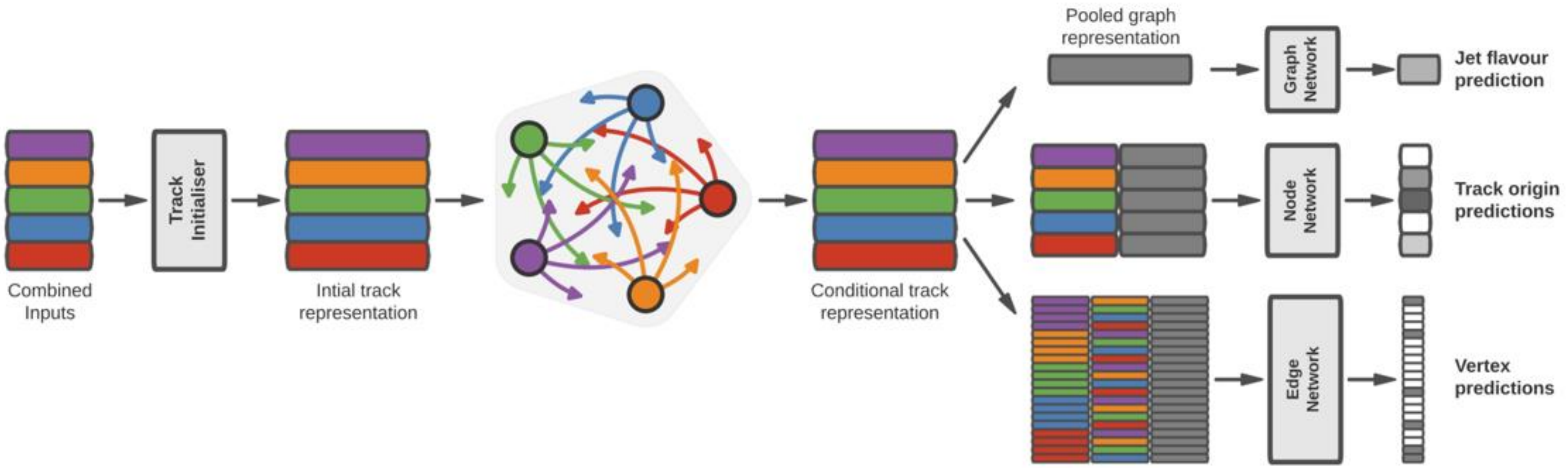


Low p_T $b\bar{b}$ Boosted (high p_T) $b\bar{b}$



Jet mass distributions for jets from SM predictions (from H→bb study)

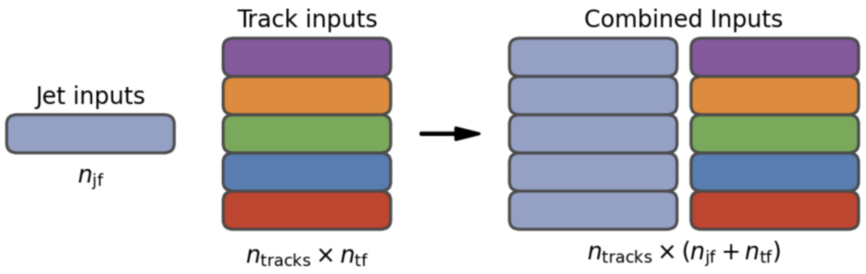
GN2X Model for Large-R Jet Tagging



Schematic network architecture of GN2/GN2X (ref)

- **GN2**:
  - Consituent based tagger
  - Uses jet and track information
  - Combines transformer architecture with auxiliary training objectives

- **GN2X** modifies existing GN2 architecture for Large-R Jets
  - Trained on 60M jets in 4 categories: H→bb, H→cc, Top, QCD



jet- and track-level information concatenation

GN2X Performance for H→bb

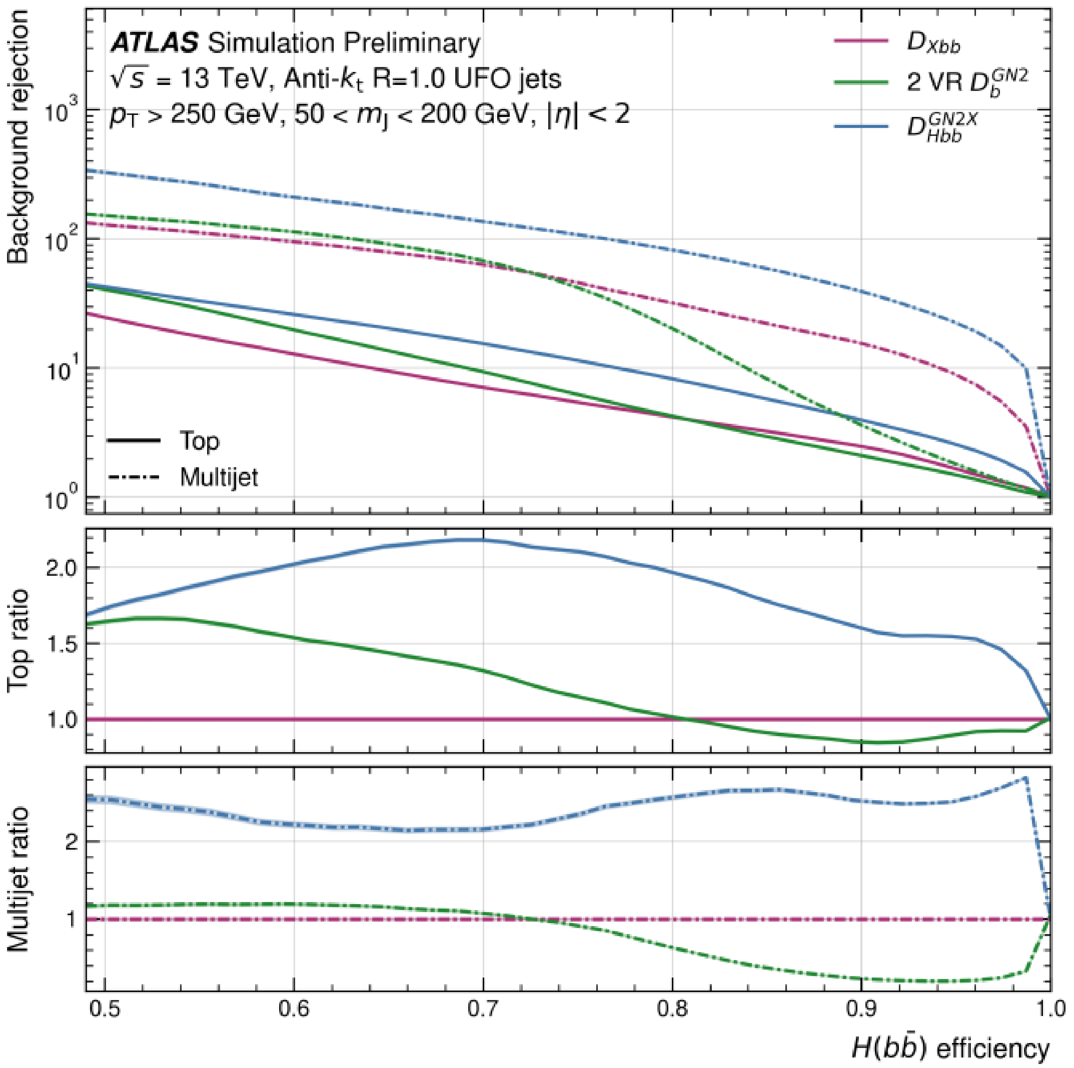
Discriminant: $D_{\text{Hbb}}^{\text{GN2X}} = \ln\left(\frac{p_{\text{Hbb}}}{f_{\text{Hcc}} \cdot p_{\text{Hcc}} + f_{\text{top}} \cdot p_{\text{top}} + (1 - f_{\text{Hcc}} - f_{\text{top}}) \cdot p_{\text{QCD}}}\right)$

where $f_{\text{Hcc}} = 0.02$, $f_{\text{top}} = 0.25$

- D_{Xbb}:
  - Previous state-of-the-art Xbb tagger
- 2 VR D_b^{GN2}
  - GN2 architecture
  - 2 leading VR jets are b-tagged
- $D_{\text{Hbb}}^{\text{GN2X}}$
  - New large-R jet tagger

Performance:

- GN2X outperforms both previous taggers across all $H(b\bar{b})$ efficiencies
- At 50% $H(b\bar{b})$ signal efficiency, compared to D_{Xbb}:
  - **1.6x** increase in top rejection
  - **2.5x** increase in QCD (multijet) rejection

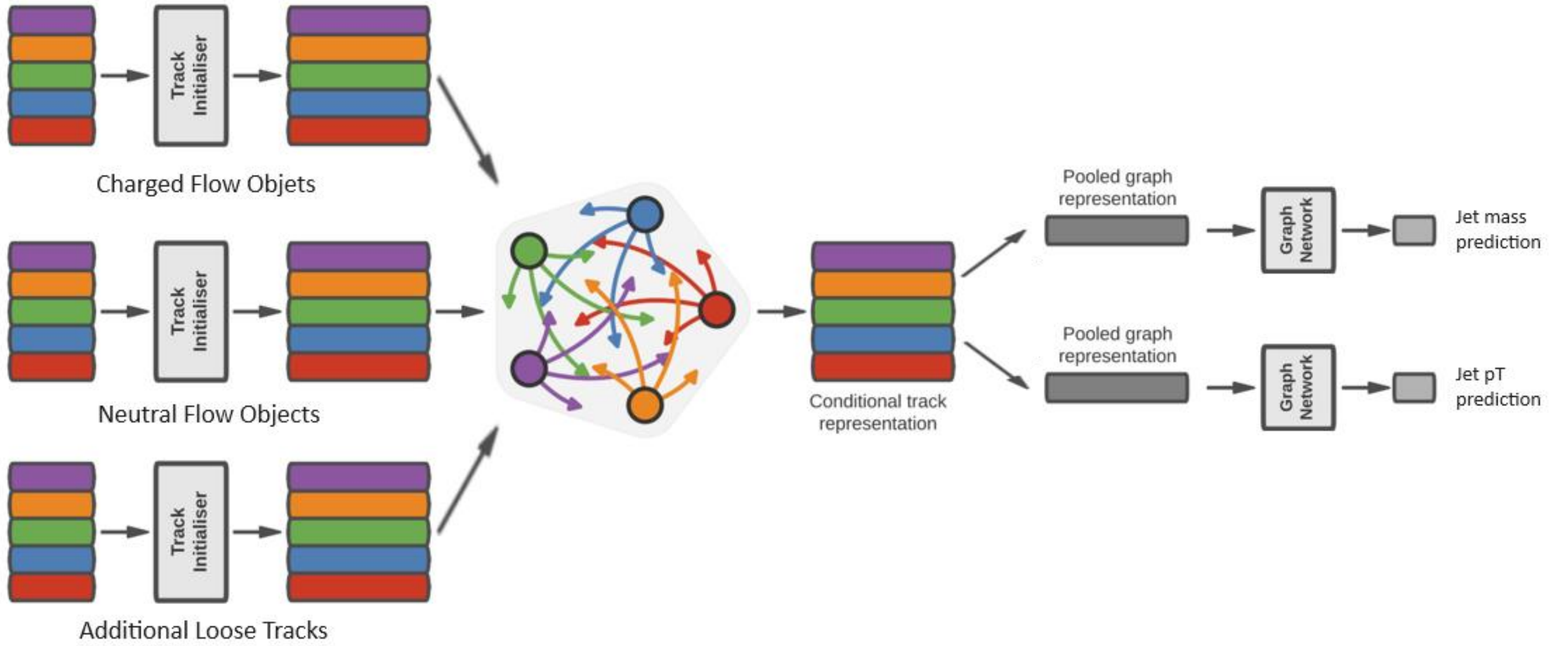


Adapting GN2X for Regression

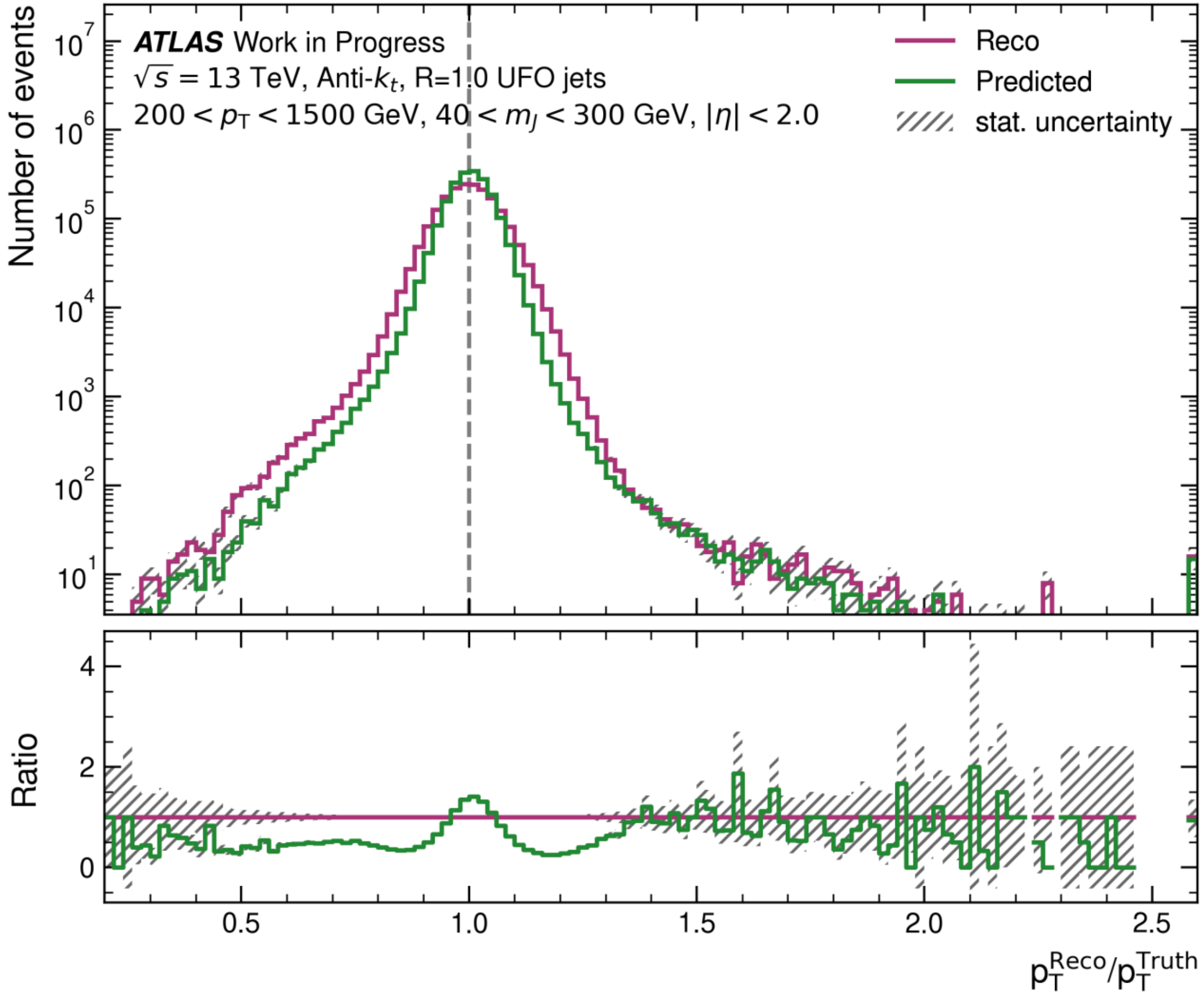
- Transformer-based models can also be used to improve **reconstruction of jet kinematic variables**
  - One potential use: improve calibration after tagging
  - In addition, help improve sensitivity for boosted Higgs studies

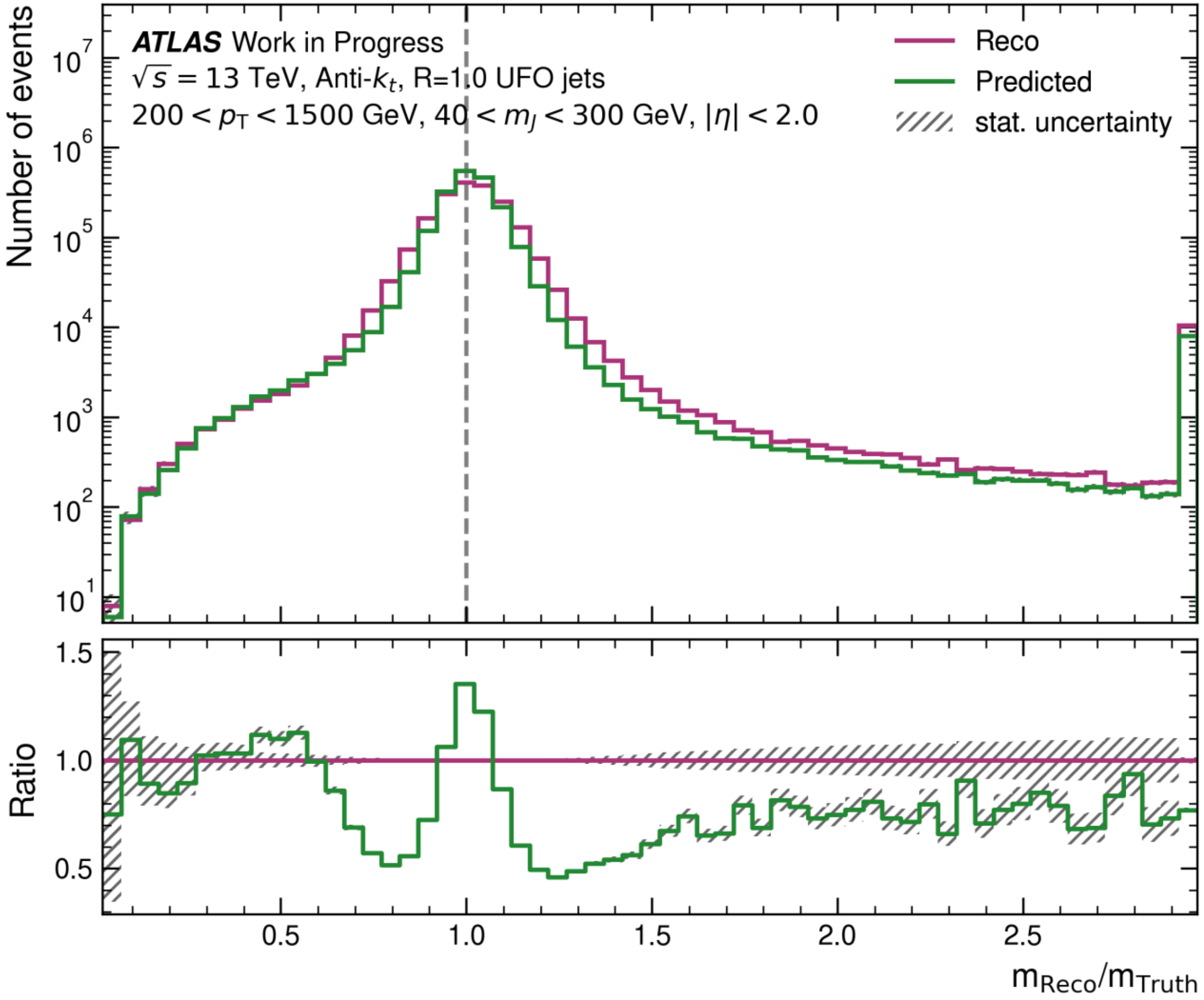
- For that, **modifications to GN2X**:
  - Replaced all main and auxiliary tasks with jet mass, p_T regressions
  - Use same training data, but remove resampling stage from preprocessing
  - Add calorimeter information (charged + neutral) instead of only using tracks

- Regression performance evaluated on jets that pass 70% WP $H \rightarrow bb$ tagger!



GN2X Performance for Regression



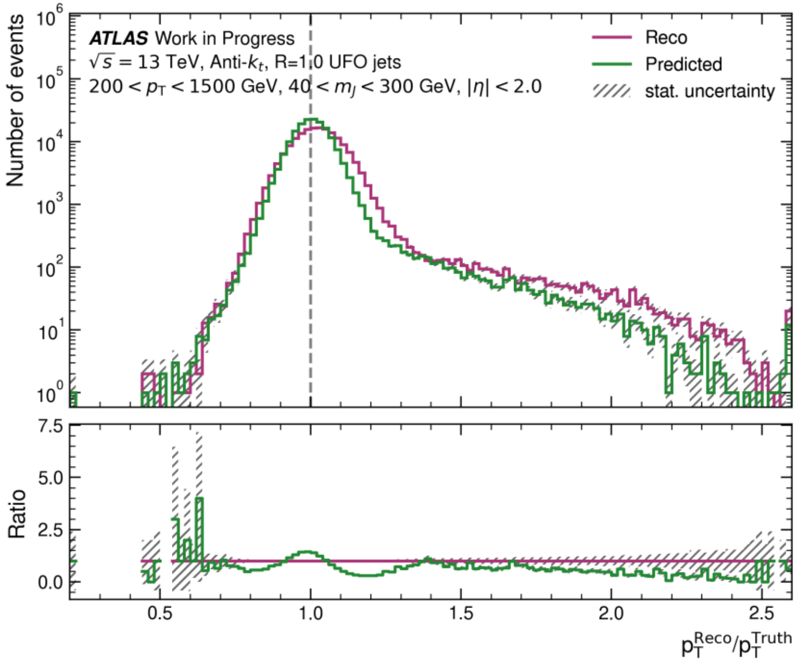


- Predicted peak is **28.5% narrower**
- Predicted peak is **0.2% further** from 1

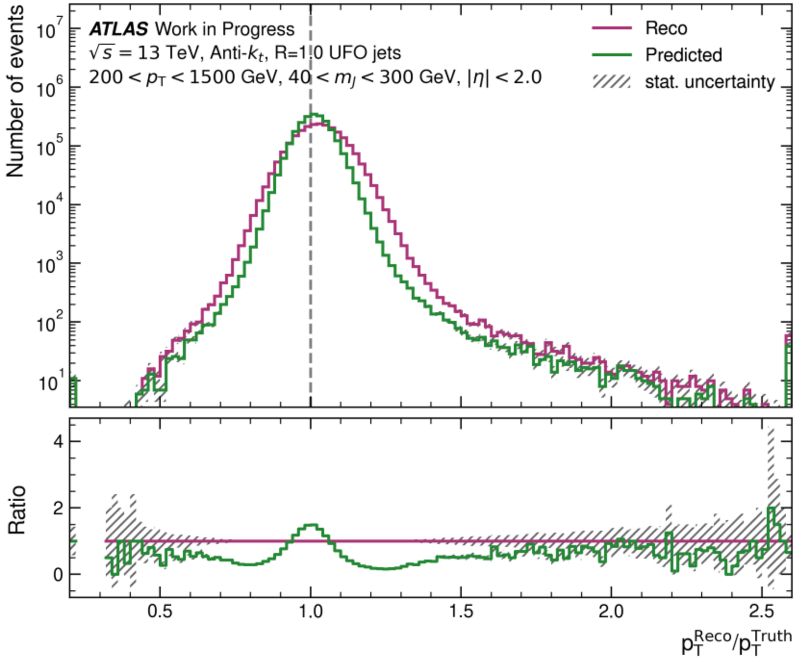
- Predicted peak is **26.6% narrower**
- Predicted peak is **0.3% closer** to 1

p_T Regression Performance for SM Samples

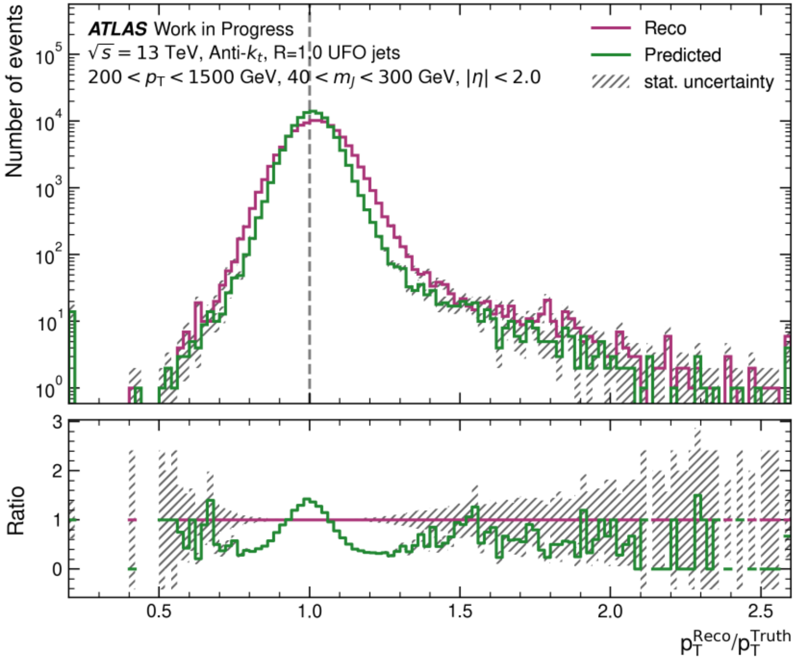
gg→H→bb



Z→bb



QCD

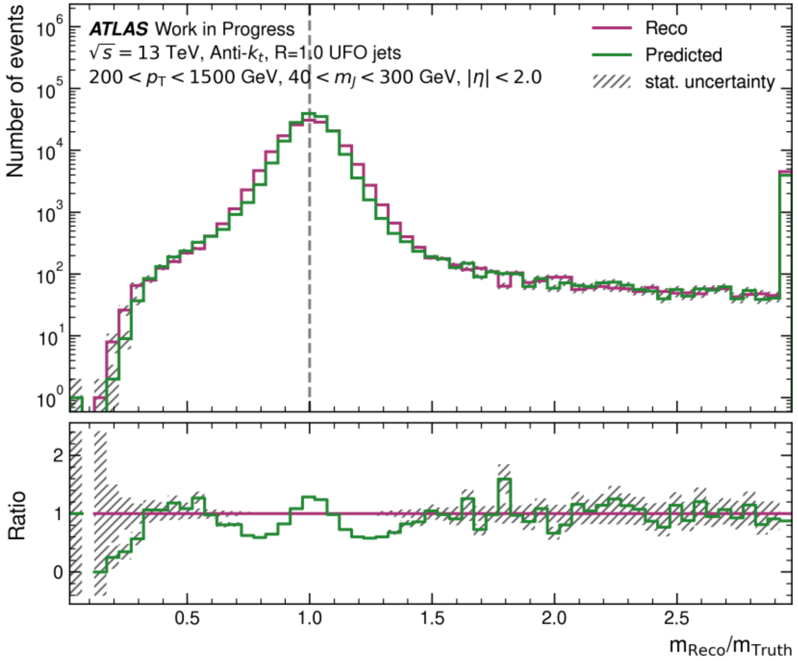


Great p_T regression performance on SM samples (not used in training)!

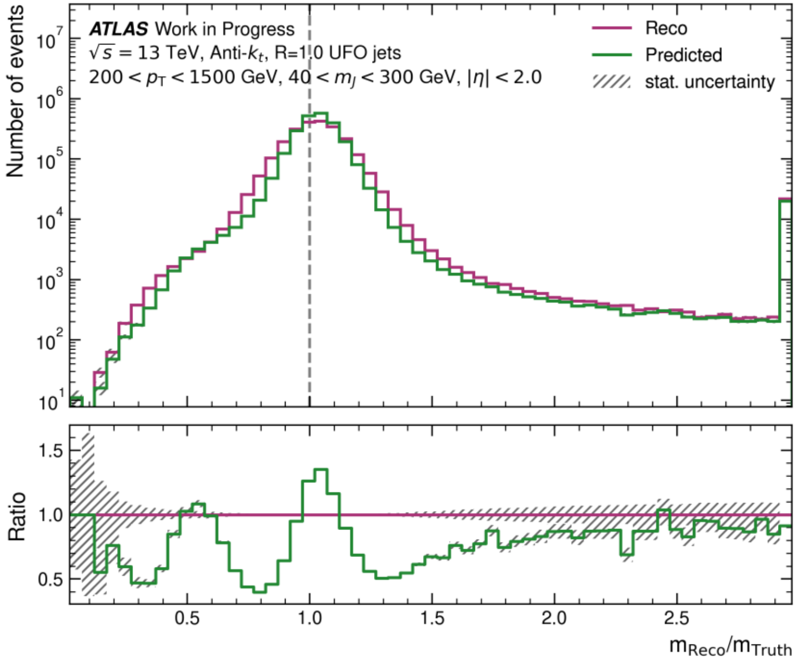
- Response resolution improvement from **26%** (QCD) to **30%** (Z→bb) compared to reco p_T
- Reduced response bias for all SM samples

Mass Regression Performance for SM Samples

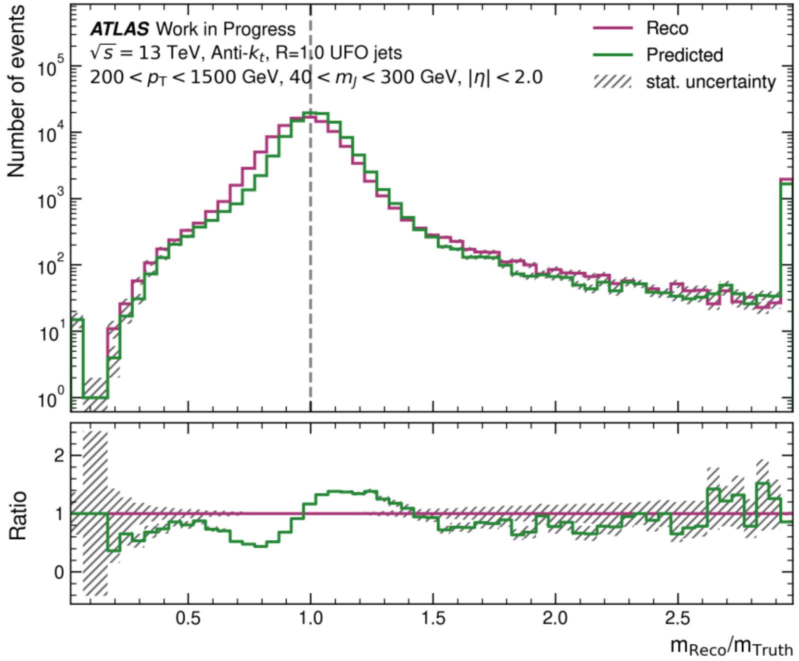
gg→H→bb



Z→bb



QCD



Mass regression improves resolution compared to reco, but introduces small bias:

- Response resolution improvement from **14%** (QCD) to **26%** (Z→bb) compared to reco mass
- Response peak median shifted away from 1
  - Worst for QCD, median at 0.99 for m_{reco}, at 1.02 for $m_{predicted}$
  - Not much, but still important to look into

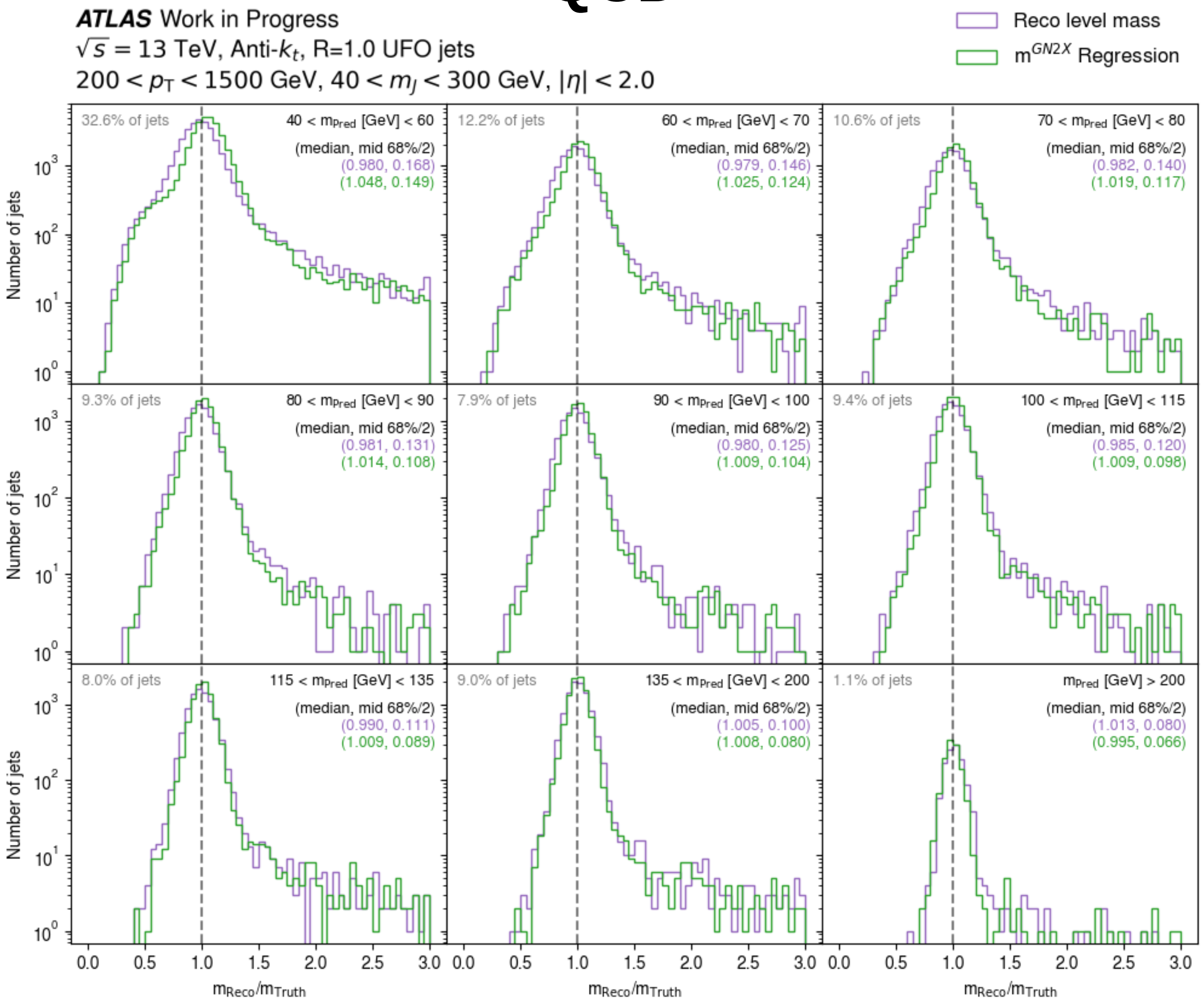
Closer Look at Mass Regression for QCD

- QCD jet mass regression is important for improving background modelling
  - For **analyses**, region around m_{Higgs} is most important
  - For **calibration** efforts, need consistent performance across all mass ranges

- Results:
  - Improved resolution across all masses
  - Similar bias for QCD jets close to m_{Higgs}
  - Increased bias in 40-60 GeV region

- Requires further model tuning, more QCD training jets (WIP)
  - Ideally get more QCD(bb)

QCD

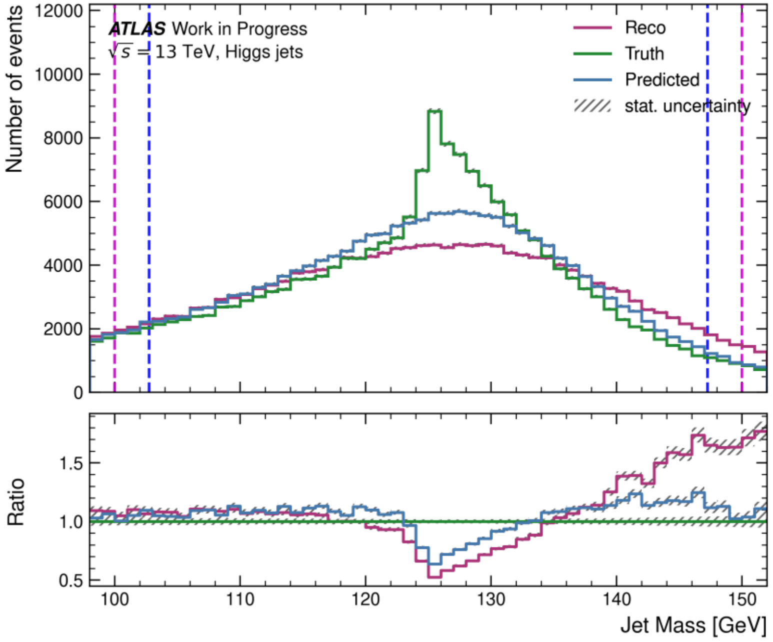


Potential Sensitivity Improvement

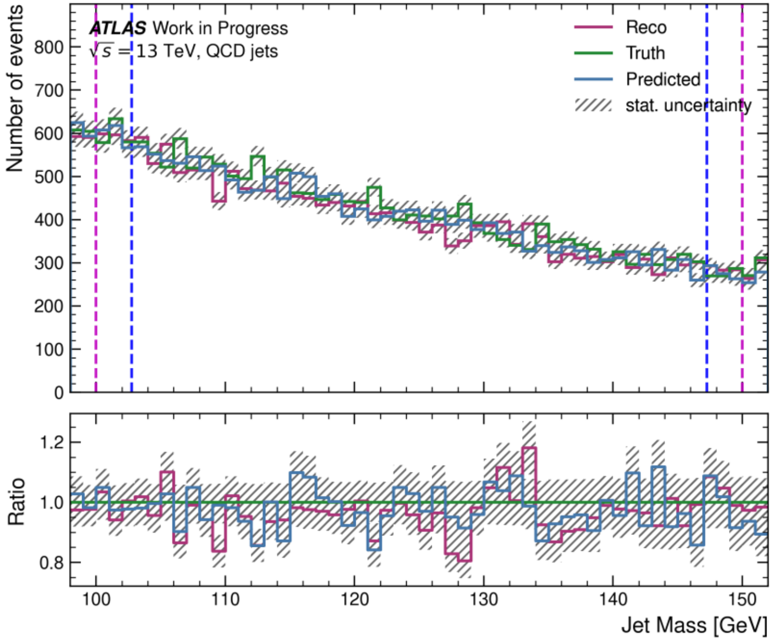
- To estimate potential sensitivity improvement, calculate reduction in QCD jets for fixed Higgs efficiency
  - Need to also verify that QCD distribution is unaffected

- Results:
  - For 75% Higgs jets efficiency, using mass regression results in **10.1% less QCD jets**
  - Correlates to approx. **5-6% sensitivity improvement**
  - Predicted mass peak could be further improved

Higgs



QCD



Mass window for 75% Higgs efficiency
- In pink for reco
  - (100-150 Gev)
- In blue for predicted
  - (102.7-147.3 GeV)

Summary

- Large-R jet classification using transformer-based architecture provides significant improvement over previous results
  - **1.6x** increase in top rejection
  - **2.5x** increase in QCD rejection

- Regression provides substantial improvement in resolution:
  - Over **25%** improvement in resolution for jet p_T compared to reco
  - **14%** improvement for QCD jet mass
  - **22%** improvement for Higgs jet mass

- Using new predicted masses allows to reduce QCD background by **10%** for 75% Higgs efficiency
  - Translates roughly to **5-6%** sensitivity improvement

- Work on H→bb tagger ongoing, with plan to provide update model for Run 3 data
- Aiming to make regression model available to entire collaboration and calibrations soon!

Thank you for listening!

Backup

Technical info

Jet Collections:

- Reco: AntiKt10UFOCSSKSoftDropBeta100Zcut10
- Truth: AntiKt10TruthSoftDropBeta100Zcut10Jets

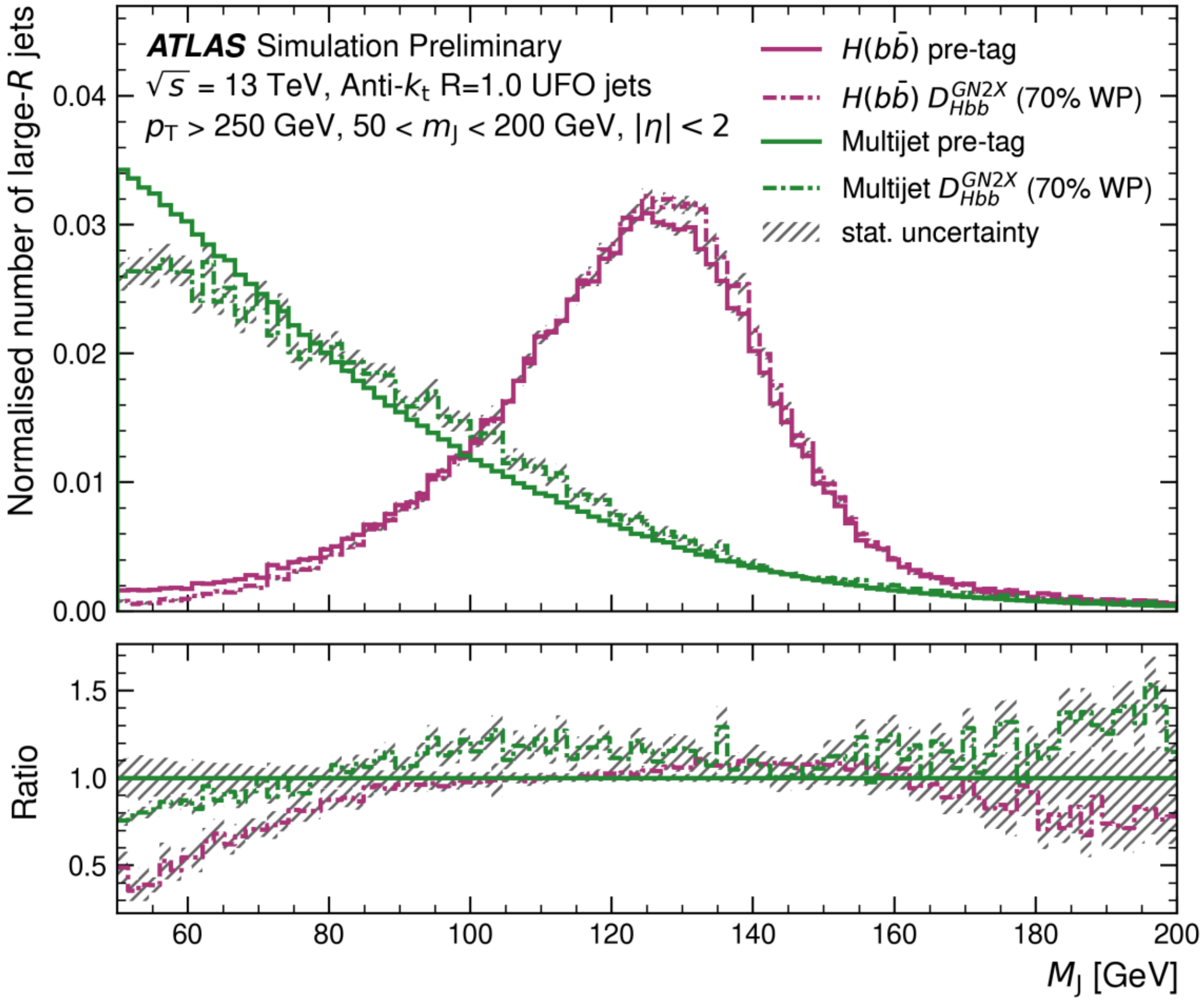
Frameworks used:

- DAODs processed with training-dataset-dumper
  - Produces intermediate ntuples

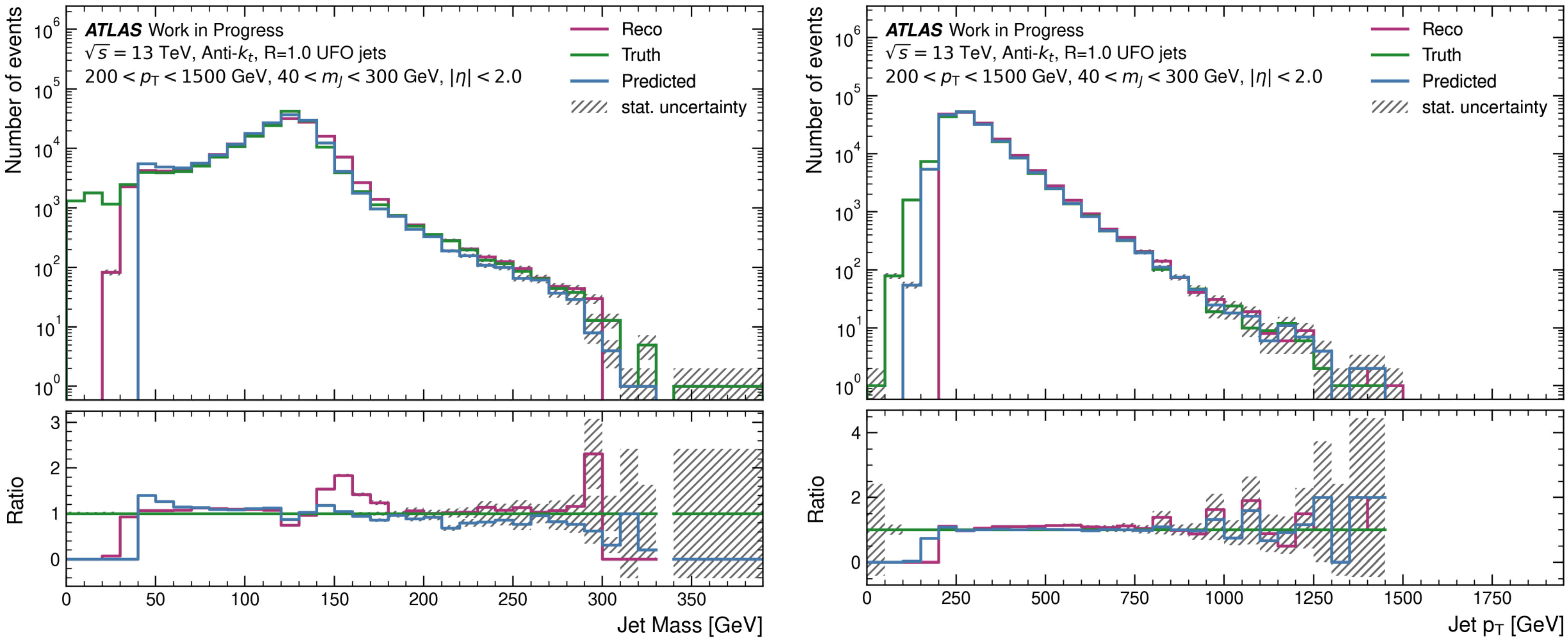
- Training files created with umami-preprocessing (UPP)
  - Modular preprocessing pipeline for jet tagging
  - Uses atlas-ftag-tools package extensively
  - Data prep, resampling and splitting data

- Training done using Salt
  - General-purpose framework to train state-of-the art jet flavour tagging algorithms
  - Model architecture and training fully configured via YAML config files / CLI
  - Up-to-date documentation, docker image support, and extensive CI tests
  - Flexible support for many types of input objects/formats and network architectures

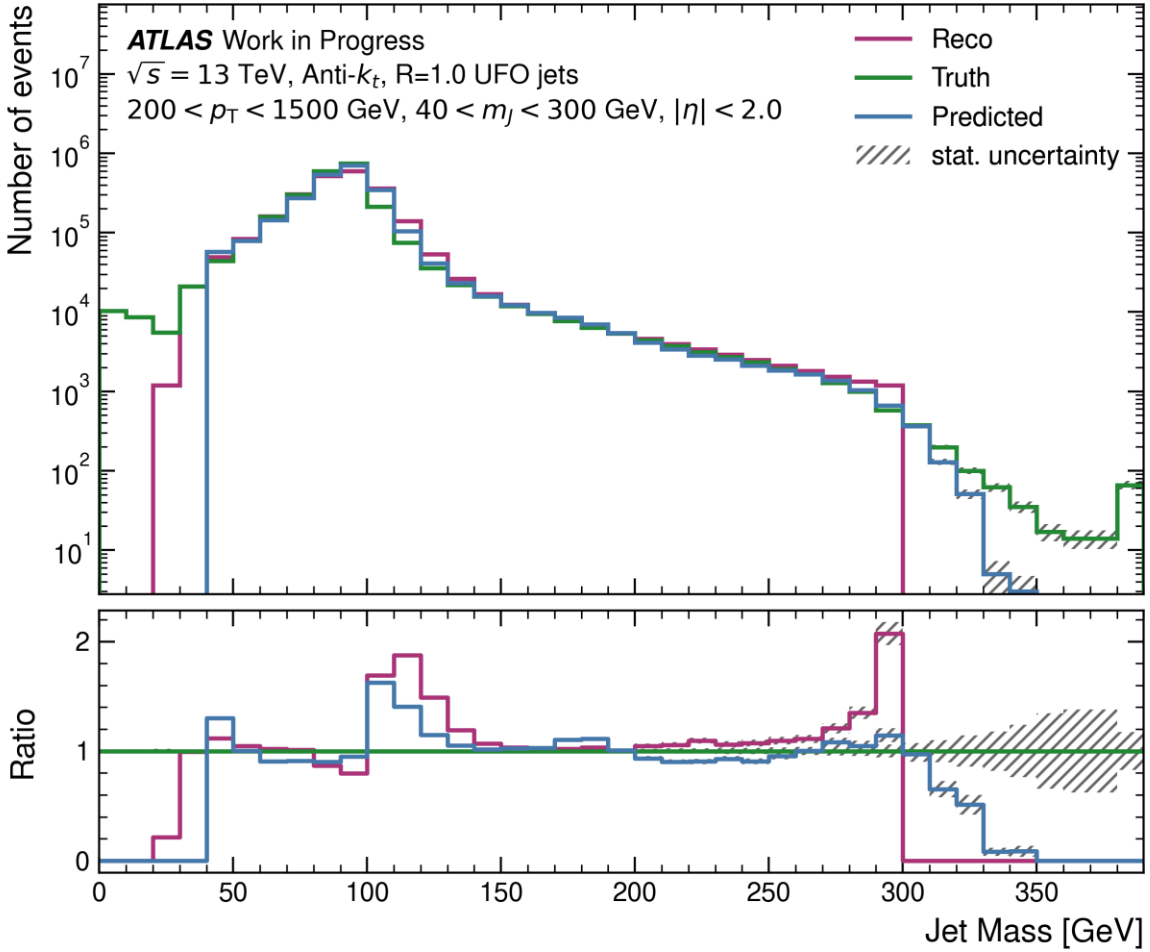
GN2X Mass Sculpting

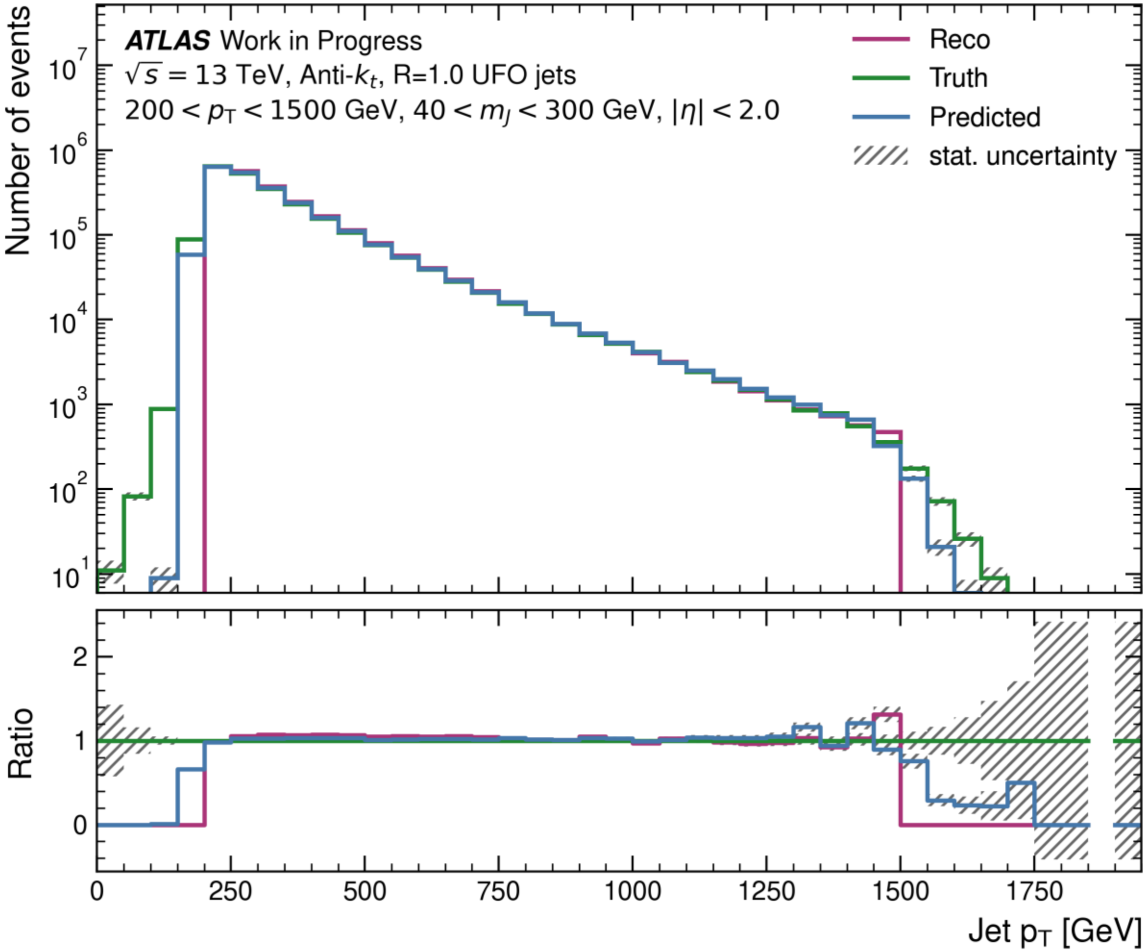


Variable Distributions (gg→H→bb)

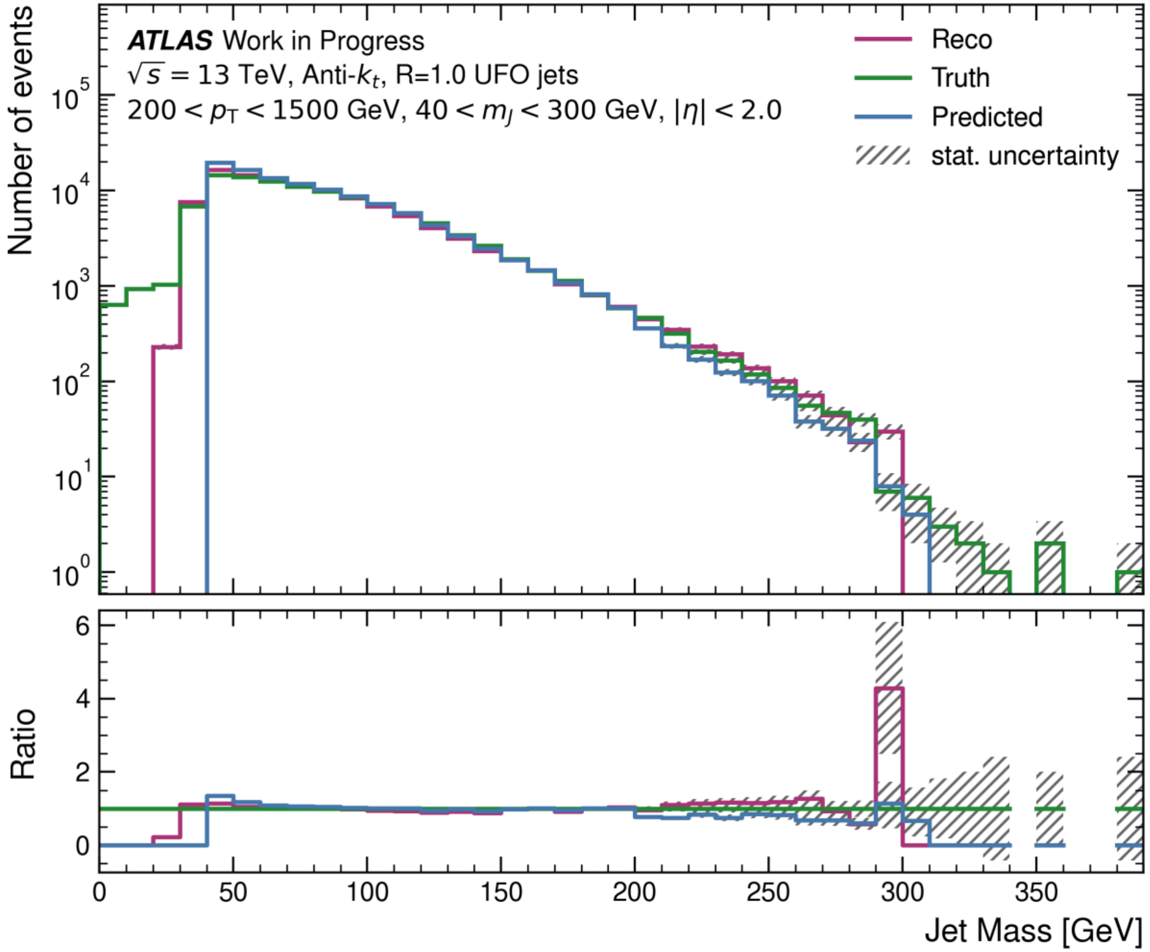


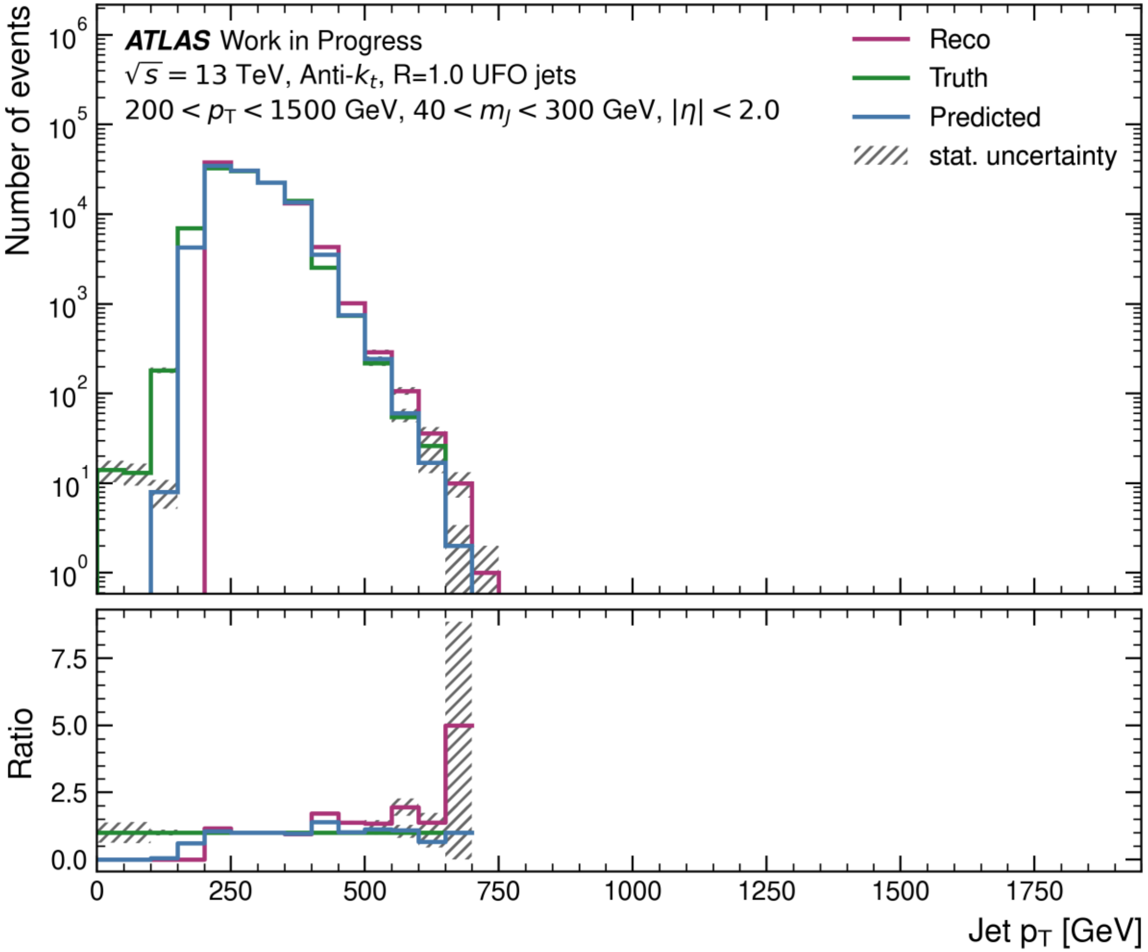
Variable Distributions (Z→bb)





Variable Distribution (QCD)





Closer Look at Mass Regression for Z→bb