Performance of UFO and LC Topo jet collections for the VBF di-Higgs production in the HH → $b\bar{b}b\bar{b}$ channel in terms of response scale and resolution.

CERN Summer Student Programme (2023) Project Report

Titanilla Braun

School of Physics and Astronomy, The University of Manchester, Manchester M13 9PL, United Kingdom

This project was completed as part of the CERN Summer Student Programme under the supervision of Reina Camacho Toro (LPNHE-IN2P3/CNRS) and Dilia María Portillo Quintero (TRIUMF)

(Dated: August 4, 2023)

For performance comparison, scale and resolution graphs were created for the mass and energy response of large-radius jets reconstructed from Unified Flow Objects (UFOs) and jets reconstructed only from calorimeter topological clusters (referred to as LC Topo jets) using the SoftDrop and Trimmed groomed truth jet collections, respectively. The analysis was performed on two Monte Carlo simulated data samples based on LHC Run-2 conditions. One of these samples only considered the non-resonant standard model processes for vector boson fusion di-Higgs production, while the other solely contained the resonant process for a graviton decaying to two Higgs bosons. The analysis focused on the di-Higgs to 4b quarks channel exclusively. As a result for both samples, the mass scale showed a good closure of 2% for both jet collections, except at low p_T, and the mass resolution was either comparable between the two jet types or the UFO jet collection had a better performance, which is in line with expectations. The energy scale had a good closure of 2-3% and the energy resolution was better for the LC Topo jet collection, as expected due to the LCW calibration.

I. INTRODUCTION

Jet reconstruction is a crucial part of particle physics experiments. Quarks and gluons resulting from high energy collisions hadronise due to colour confinement and create a shower of stable particles that are called jets. To analyse the event that created the quarks and gluons, these jets need to be identified and reconstructed based on the tracks and energy deposits recorded by the detector. This is done using the anti-k_T clustering algorithm with a distance parameter R [1].

The inputs for this algorithm can be varied for different experiments and detectors. The ATLAS collaboration has been using topological clusters composed of calorimeter cells with corrections applied to the energy measurements, these are the LC Topo reconstructed jet collections [1]. However, recently a new particle-flow algorithm, which uses "Unified Flow Objects" (UFOs), was developed to improve the mass resolution throughout a wide p_T range by combining tracking information and calorimeter data [2]. The parameter R also varies across experiments. ATLAS uses a large value of R = 1.0 which is a good choice for the search for a new resonance decaying to two Higgs bosons, since the momentum of the Higgs particles will be much higher than their mass [1].

A current goal of particle physics research is the reconstruction of the Higgs field in order to gain a deeper understanding of the interaction which gives mass to fundamental particles. Studying the simultaneous production of two Higgs bosons can provide essential information about the Higgs potential. Amongst the decay channels to be considered, the HH → $b\bar{b}b\bar{b}$ channel has the highest branching ratio and is the one that can profit the most from improvements on the jet reconstruction performance. [1]

The aim of this project was to check the performance of the LC Topo and the UFO jet collections for the vector boson fusion (VBF) di-Higgs process in the HH → $b\bar{b}b\bar{b}$ channel. The UFO jet collection was expected to perform better in terms of mass resolution over a wider kinematic range compared to the LC Topo jet collection. These studies can provide important insight about possible improvements in preparation for the upcoming analysis that will be done during Run-3.

II. METHODOLOGY

ATLAS coordinate system and relevant variables

ATLAS uses a coordinate system centered at the interaction point of the detector and with the beamline defined as the z-axis. The x-axis points towards the middle of the LHC ring and the y-axis points upwards. Cylindrical coordinates (r, ϕ) are used in the plane perpendicular to the beamline where ϕ is the azimuthal angle around the z-axis. The transverse momentum p_T is measured on this plane. [3]

The pseudorapidity of a jet is defined in terms of the polar angle θ as [3]

$$\eta = -\ln tan(\theta/2). \quad (1)$$

Central jets have small absolute values, while forward jets have high absolute values of pseudorapidity.

To determine the distance of two jets, angular distance ΔR is used, which is given by [3]:

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}, \quad (2)$$

where $\Delta\eta$ and $\Delta\phi$ are the differences between the two jets in pseudorapidity and azimuthal angle, respectively.

Data samples

Two Monte Carlo simulated data samples were used in the analysis, both of which contained VBF HH $\rightarrow b\bar{b}b\bar{b}$ processes. The first sample included only the non-resonant standard model processes for VBF di-Higgs production, which can be seen in figure 1 as processes "a", "b", and "c". In contrast, the second sample considered the resonant process, Graviton $\rightarrow$ HH $\rightarrow b\bar{b}b\bar{b}$, which can be seen in figure 1 as process "d" with the X particle decaying to two Higgs particles being the graviton with a mass of 3000 GeV. These samples were processed using the Easyjet framework [4] to produce reduced datasets for this study and are both Monte Carlo simulations created based on LHC Run-2 conditions (data taking period between 2015 and 2018). LO MadGraph was used for the matrix element calculation, Pythia8 was used for the parton shower and hadronization step, and EvtGen was used as an afterburner that ensures the decays of B hadrons are correctly modeled [5].

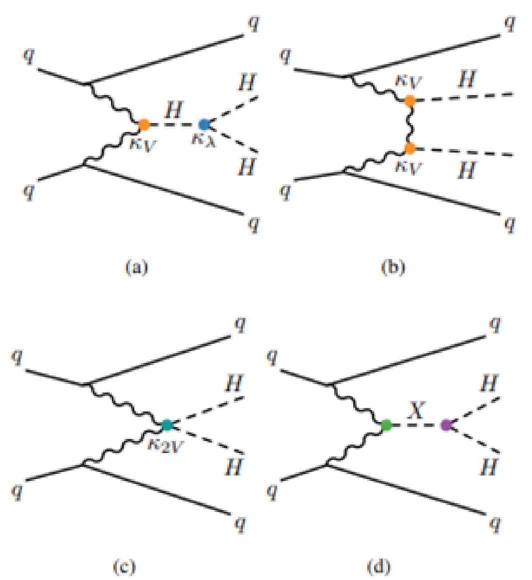

FIG. 1. Feynman diagrams contributing to VBF di-Higgs production. Diagrams a, b, and c represent the non-resonant processes (with scaling constants κ_V, κ_λ, and κ_{2V} equal to 1 for the standard model scenario), while diagram d shows the resonant process.[3]

Selection cuts and matching

Out of all the events in the data samples, only the ones with at least two high energy central reconstructed jets were selected for the analysis. A further selection then chose the events which all had at least two jets where both jets had a variable called TruthLabel_R22 equal to 11, meaning that both jets were created from an H $\rightarrow b\bar{b}$ process. A summary of these selection criteria can be seen in table I. It is important to mention that a jet energy and mass calibration was already applied to the reconstructed jets when producing the samples [6].

Variable	Selection applied
number of jets	≥ 2
p_T	>250 GeV
$\lvert\eta\rvert$	<2
TruthLabel_R22	= 11

TABLE I. Selection criteria applied on both data samples. The first criterion was applied on all the events in the samples, while the other criteria were applied on the leading and subleading jets in the events selected by the first criterion.

After passing the selection criteria, jets in the reconstructed jet collections, UFO and LC Topo jets, were matched to a corresponding jet in the truth jet collections. Truth jets are large-R jets reconstructed from stable particles, which are required to have a lifetime greater than 10 ps, with muons and neutrinos excluded as they only leave minimal energy deposits in the calorimeter. This matching was based on the angular distance between the reconstructed jets and the truth collection jets. The truth jet with the smallest angular distance relative to the reconstructed jet was selected. Truth jet collections used for UFO jets and LC Topo jets were the SoftDrop and the Trimmed truth jet collections, respectively, due to the lack of an ungroomed truth jet collection. A histogram of the angular distance values between the leading reconstructed jets and their matched truth jet can be seen in figure 2, this shows that the matching is excellent since the ΔR values are all quite close to 0.

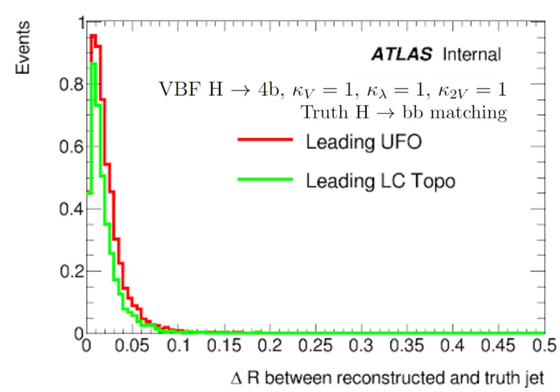

FIG. 2. Histogram showing the angular distance between the leading reconstructed jets and their corresponding truth jets. The red histogram shows the values between the leading UFO jets and their matched SoftDrop truth jet, while the green histogram shows the same for the leading LC topo jets and their Trimmed truth jet pair.

Responses

Responses in mass and energy are defined as

$$R_m = m_{reco}/m_{truth}, \tag{3}$$

and

$$R_E = E_{reco}/E_{truth}, \tag{4}$$

where m_{reco} and E_{reco} are the mass and energy of the reconstructed jet, and m_{truth} and E_{truth} are the mass and energy of the corresponding truth jet. Ideally the response is very close to one.

These responses were calculated for each event and then histograms were plotted showing the responses in different p_T bins. For the non-resonant sample the bins were: 250-300, 300-350, 350-400, 400-500, 500-600, and above 600 GeV. Similarly, there were six bins for the resonant sample as well: 500-800, 800-1000, 1000-1250, 1250-1500, 1500-1750, and above 1750 GeV.

A Gaussian function was fitted to each of these histograms using a fitting range of 0.65-1.5 for the non-resonant sample and a range of 0.8-1.2 for the resonant sample. Two examples of these fits can be seen in figure 3.

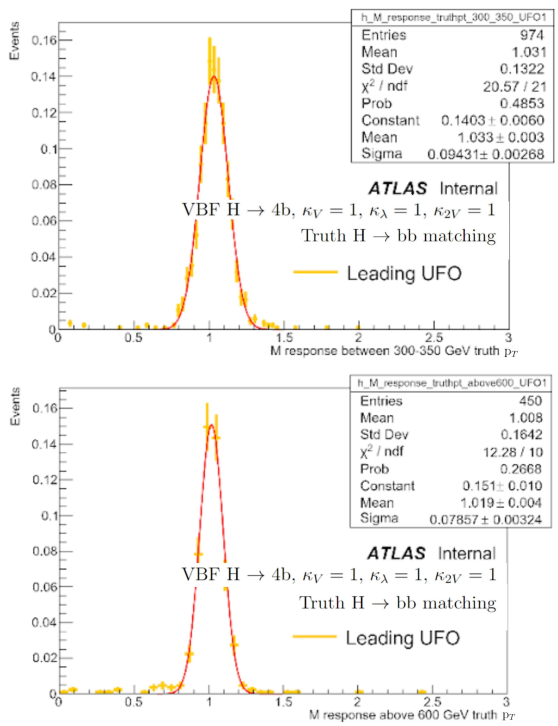

FIG. 3. Gaussian fits for the leading UFO reconstructed jets in the p_T bins 300-350 (on top) and above 600 GeV (at the bottom). The same fits were performed for the subleading UFO jets and the LC Topo jets as well.

The mean of these fits were used to create the mass and energy scale graphs showing the mean value relative to its p_T bin. The resolution graphs were created showing the mean value of the fit divided by the standard deviation of the fit relative to the relevant p_T bin.

III. RESULTS

The resulting mass scale, energy scale, mass resolution, and energy resolution graphs for both the leading and subleading jets and for both data samples can be seen in figures 4, 5, 6, and 7.

Both samples gave a good closure of about 2% for the mass scale, except at low p_T, and the UFO jet collection demonstrated an improved performance in regard to mass resolution compared to the LC Topo jet collection. Regarding the energy resolution, however, due to a Local Cell Weighting (LCW) calibration applied to the LC Topo jet collection, the latter performed better. For the energy scale, the closure was around 2-3%. All these results satisfy previous expectations.

Further studies to examine these scale and resolution metrics could be done using a similar truth jet collection for both reconstructed jet types, and repeating the analysis using other resonant samples or samples with Run-3 conditions. Studying the pile-up dependency could also provide important information.

IV. CONCLUSION

The final results align with prior predictions. In terms of mass resolution, the UFO jet collection, which combines calorimeter and tracker information, outperforms the LC Topo jet collection, which solely relies on tracking data. Possible extensions of this analysis include the use of an ungroomed truth jet collection for the matching of both reconstructed jet collections, examining the effects on the pile-up dependency, and the investigation of additional data samples.

V. ACKNOWLEDGEMENTS

I would like to express my deep gratitude to Reina Camacho Toro and Dilia María Portillo Quintero, my supervisors, for their patient guidance, positive encouragement and constructive critiques throughout this internship. I would also like to thank Arely Cortes and Daariimaa Battulga for generously providing the data samples and István Julinek for offering valuable advice on certain coding aspects.

VI. REFERENCES

[1] B. D. Micco, M. Gouzevitch, J. Mazzitelli, and C. Vernieri, Higgs boson potential at colliders: Status and perspectives, Reviews in Physics **5**, 100045 (2020).

[2] ATLAS Collaboration, Optimisation of large-radius jet reconstruction for the ATLAS detector in 13 TeV proton-proton collisions, The European Physical Journal C **81**, 334 (2021).

[3] ATLAS Collaboration, Search for the HH $\rightarrow b\bar{b}b\bar{b}$ process via vector-boson fusion production using proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector, Journal of High Energy Physics **7**, 108 (2020).

[4] Easyjet framework, `https://gitlab.cern.ch/easyjet/easyjet/-/tree/master`.

[5] ATLAS Collaboration, ATLAS Software Documentation, `https://atlassoftwaredocs.web.cern.ch/AnalysisSWTutorial/mc_generation/`.

[6] Calibrations derived for the Run-2 R22 conditions and using dijet samples.

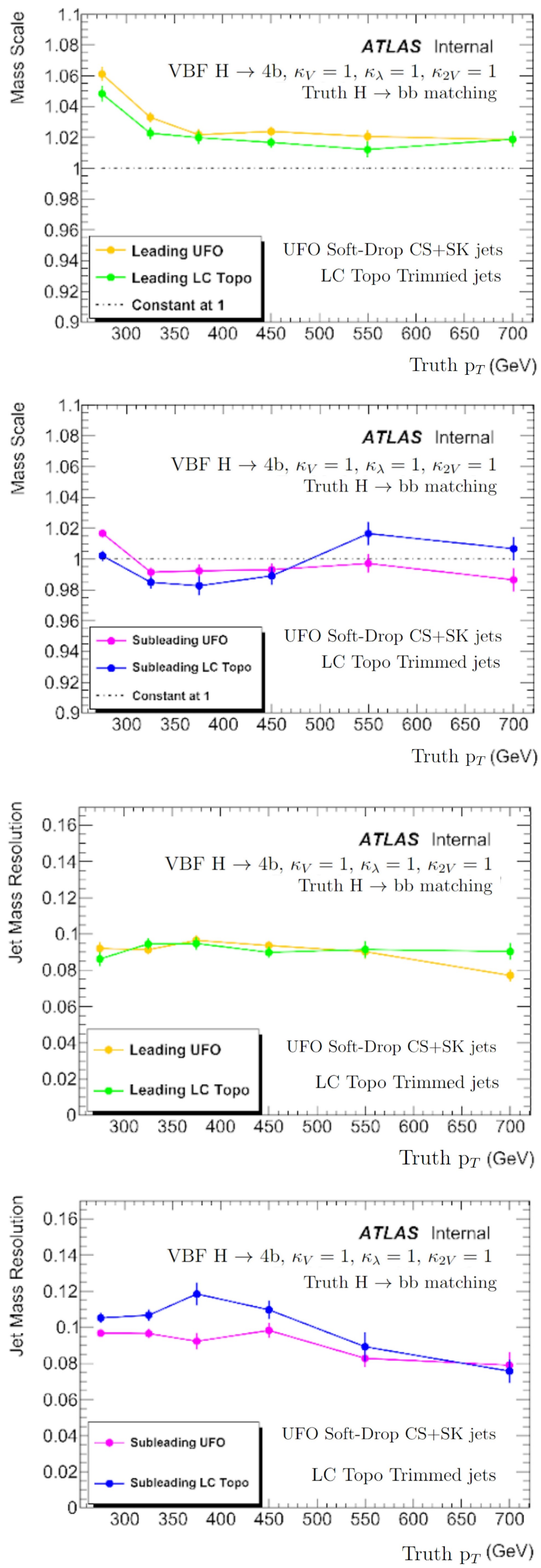

FIG. 4. Mass scale and mass resolution graphs for the leading and subleading UFO and LC Topo jets in the non-resonant data sample.

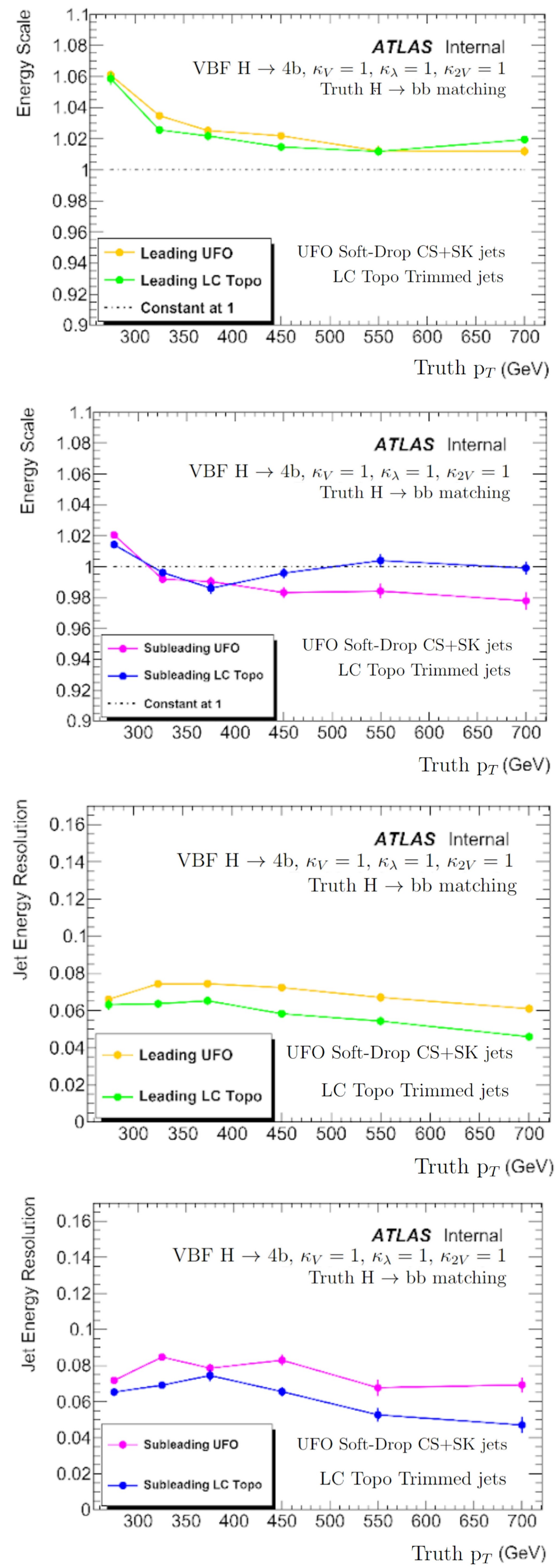

FIG. 5. Energy scale and energy resolution graphs for the leading and subleading UFO and LC Topo jets in the non-resonant data sample.

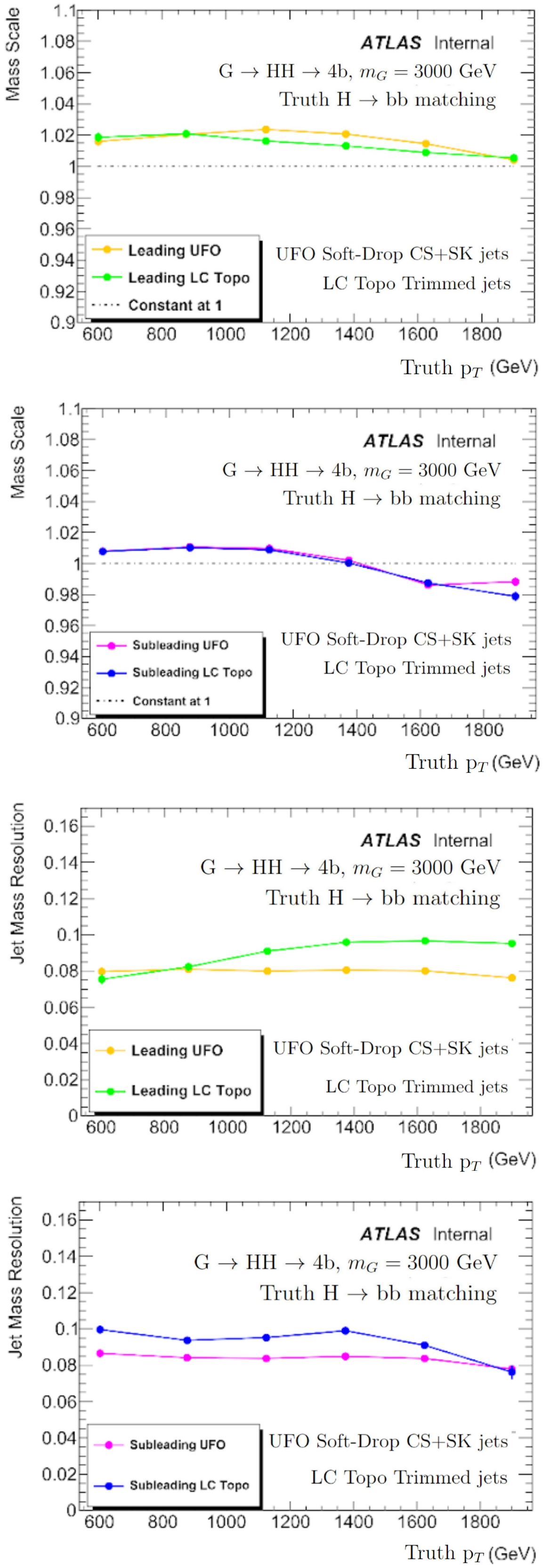

FIG. 6. Mass scale and mass resolution graphs for the leading and subleading UFO and LC Topo jets in the resonant data sample.

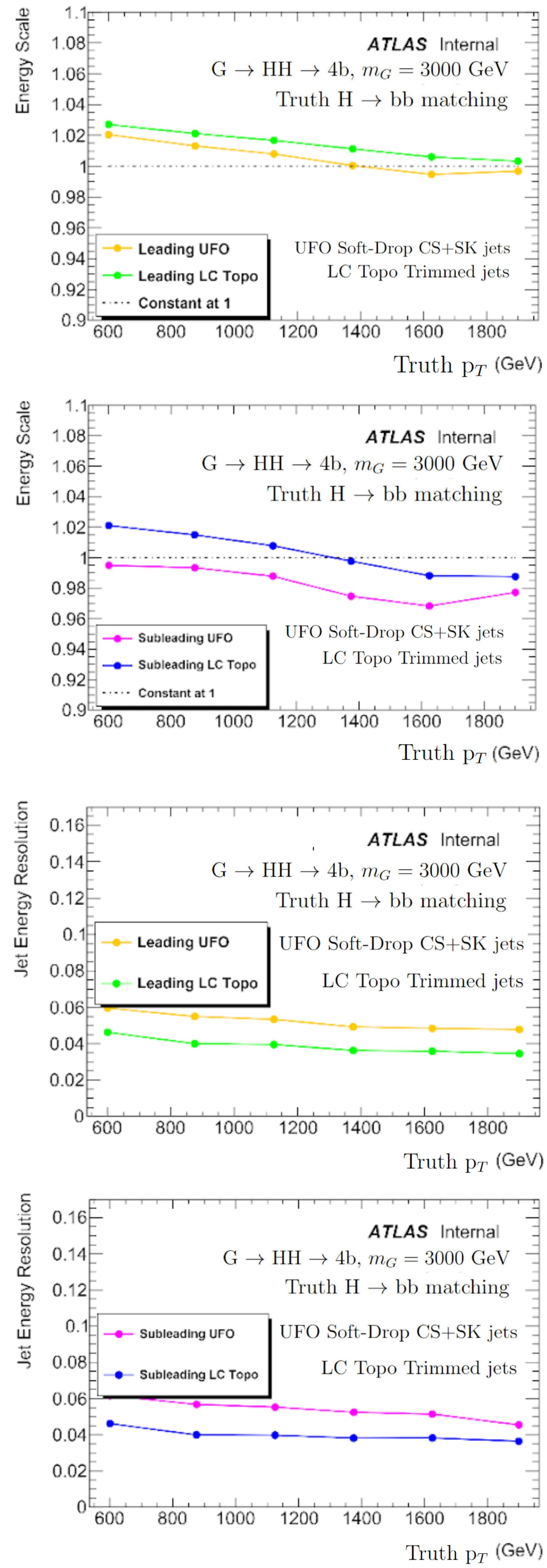

FIG. 7. Energy scale and energy resolution graphs for the leading and subleading UFO and LC Topo jets in the resonant data sample.