ATLAS PUB Note
ATL-PHYS-PUB-2024-015
27th July 2024

Transformer networks for constituent-based *b*-jet calibration with the ATLAS detector

The ATLAS Collaboration

The precise measurement of a jet's kinematics is a critical component of the physics program based on proton–proton collision data recorded by the ATLAS detector at the Large Hadron Collider. The determination of the energy and mass of jets containing bottom quarks *b*-jets is particularly difficult as, for example, they have different radiation patterns compared to the average jet and can contain heavy-flavour decays into a charged lepton and an unobserved neutrino. This document reports on a novel calibration technique for jets focusing on *b*-jets using transformer-based neural networks trained on simulation samples to correct reconstructed jet properties to the true values. Separate simulation-based regression methods have been developed to estimate the transverse momentum of small-radius jets and the transverse momentum and mass of large-radius jets. In both cases, the regression methods move the median measurement closer to the true value. A relative resolution improvement with respect to the nominal calibration between 18% and 31%, depending on the transverse momentum, is demonstrated for small-radius jets. Both the large-radius jet transverse momentum and mass resolution are shown to improve by 25–35%.

1 Introduction

Hadronic jets are abundantly produced in proton–proton (pp) collisions at the Large Hadron Collider (LHC). Many studies of the Standard Model (SM) and searches for physics beyond the SM rely on precisely measuring these jets. Several key signatures contain b-jets, jets produced by the hadronisation of a b-quarks, such as the production of top quarks or heavy flavour decays of the Higgs (H) and Z bosons.

In the ATLAS experiment [1], particle-flow algorithms are used to reconstruct jets and determine their kinematic properties such as the transverse momentum p_{T} and the mass m [2, 3]. The objects from which a jet is made are referred to as *constituents*. Depending on the physics object being captured, ATLAS utilises small-radius (small-R) or large-radius (large-R) jets built with the anti-k_t algorithm [4] using a radius parameter of $R = 0.4$ and $R = 1.0$, respectively. Small-R jets are appropriate to capture the hadronic activity of jets formed from the hadronisation and fragmentation of high-p_{T} partons (quarks and gluons), while large-R jets capture the hadronic decay products of boosted massive particles, such as a high-momentum $H \to b\bar{b}$ decay, where the large-R jet mass m_{J} corresponds to the mass of the initial particle.

The current calibration scheme, referred to as the *nominal calibration* and documented in Refs. [5] and [6], is used to measure jet energy and mass from detector signals in the ATLAS detector. The calibration aims to adjust the reconstructed jet energies such that for a fixed true energy, the most probable calibrated energy matches that true energy. Central to the calibration are studies of the *response*, the ratio of a jet observable from jets reconstructed after the detector simulation divided by the value from jets reconstructed using generated particles. The width of the response is a measure of the jet resolution. A response of unity and resolution of zero is the ideal behaviour. In the current ATLAS calibration scheme, simulation-based corrections are derived from studies of response as a function of a limited number of jet parameters. Further corrections for jets in data are obtained from collider data using known physical processes or well-reconstructed objects.

Machine learning methods based on deep neural networks (DNNs) have been employed to perform a particular step of the energy calibration for small-R jets [7] and have been shown to improve the large-R jet energy and mass calibrations [8]. In the latter, the DNN method enhances the large-R jet response by incorporating inter-dependent summary parameters of the reconstructed jets, such as jet substructure variables and sums of charged and neutral energy constituents, which cannot be easily added to the current calibration methods. The regression networks explored in this work, based on the transformer architectures used to identify b-jets (flavour-tagging) [9, 10], utilise jet constituents and charged particle tracks within the jet to improve the response and the resolution.

The ATLAS jet calibration is primarily derived from light quark and gluon jets [5, 6]. Due to the large b-quark mass and the relatively large semileptonic branching fraction of b-hadrons, around 20%, the radiation patterns of b-jets significantly differ from those of light quarks or gluons. Furthermore, in a semileptonic decay containing a muon-neutrino pair, the muon's energy is not accounted for in the jet clustering, and the neutrino traverses the ATLAS detector undetected. Therefore, the estimated energy for a b-jet can deviate considerably from its true underlying energy, motivating further corrections.

So far, several ATLAS analyses have included such corrections. *Muon-in-jet* and *PtReco* described in Ref. [11] are two common methods used to correct the energies of small-R b-tagged jets. The muon-in-jet correction adds the muon four-momentum to the jet four-momentum and removes the energy deposited by the muon in the calorimeter. For small-R jets, the correction utilises the closest muon to the jet axis within

$\Delta R = 0.4$[1] with a p_{T} greater than 5 GeV. The PtReco method applies a correction based on the jet p_{T} from the residual difference in the response from unity versus jet p_{T}. These two corrections can improve the mass resolution of the reconstructed Higgs boson mass in its decay to a $b\bar{b}$ pair by about 20% if the two b-quarks are reconstructed as two small-R jets [11]. The muon-in-jet method is also regularly used to correct energies of large-R b-tagged jets, utilising multiple muons within the jet to improve the reconstructed Higgs mass resolution by between 5% and 12%, depending on the applied kinematic selection [12–14].

In the CMS experiment, a DNN-based regression method has been implemented to correct for the energies of the small-R b-tagged jets [15]. For large-R jets, a flavour-aware jet energy and resolution regression has been recently included within the flavour-tagging approach [16].

The agreement between data and simulation for the energy measurement of small-R b-jets has been studied in ATLAS using top-quark-pair events [17] and balanced against a well-calibrated photon [7]. Good agreement was reported in both studies.

This note presents a regression method to predict a correction to the calibrated jet momentum or mass, focusing on b-jets. The two jet regression networks, one for small-R and one for large-R jets predict an additional simulation-based correction for b-jets on top of the nominal ATLAS jet calibration to improve the response. While both networks had similar origins, the adaptations to the regression problem were performed in separate yet complementary directions. Both regression algorithms use the kinematic properties of the jet along with charged constituents, similar to the flavour-tagging networks. The small-R jet network explored semileptonic b-hadrons decays by adding lepton information to the input features and modifying the truth jet definition to account for the neutrinos and muons from the b-hadron decay. The large-R network attempted to capture all the activity within the jet by using both charged and neutral constituents along with additional tracks in the jet that were not included in the clustering. Both models predict the ratio of the true value to the calibrated value of the transverse momentum p_{T} (and mass m_J for large-R jets) based on training with simulated samples enriched in true b-jets. The prediction is then applied as a correction to the calibrated jet to obtain an updated estimate of the true jet kinematics. All methods and their presented performance in this note are based on simulations with the conditions of the ATLAS Run 2 data-taking (2015–2018) of pp collisions at a centre-of-mass energy $\sqrt{s}$ of 13 TeV.

This note is organised as follows. The ATLAS experiment and the simulation samples used in these studies are described in Sections 2 and 3. Object definitions are given in Section 4. The methods to derive the b-jet energy and mass corrections are detailed in Section 5, and their performances on simulation are shown in Section 6. Concluding remarks are given in Section 7.

2 The ATLAS detector

The ATLAS experiment [1] at the LHC is a multipurpose particle detector with a forward–backward symmetric cylindrical geometry and a near 4π coverage in solid angle. It consists of an inner tracking detector surrounded by a thin superconducting solenoid providing a 2 T axial magnetic field, electromagnetic

[1] ATLAS uses a right-handed coordinate system with its origin at the nominal interaction point (IP) in the centre of the detector and the z-axis along the beampipe. The x-axis points from the IP to the centre of the LHC ring, and the y-axis points upwards. Polar coordinates (r, ϕ) are used in the transverse plane, ϕ being the azimuthal angle around the z-axis. The pseudorapidity is defined in terms of the polar angle θ as $\eta = -\ln\tan(\theta/2)$ and is equal to the rapidity $y = \frac{1}{2}\ln\left(\frac{E+p_z c}{E-p_z c}\right)$ in the relativistic limit. Angular distance is measured in units of $\Delta R \equiv \sqrt{(\Delta y)^2 + (\Delta\phi)^2}$.

and hadronic calorimeters, and a muon spectrometer. The inner-detector system (ID) provides charged-particle tracking in the range $|\eta| < 2.5$. The high-granularity silicon pixel detector covers the interaction region and typically provides four measurements per track, the first hit generally being in the insertable B-Layer (IBL) followed by the B-Layer. It is followed by the SemiConductor Tracker (SCT), which usually provides eight measurements per track. These silicon detectors are complemented by the transition radiation tracker (TRT), Lead/liquid-argon (LAr) sampling calorimeters provide electromagnetic (EM) energy measurements with high granularity within the region $|\eta| < 3.2$. A steel/scintillator-tile hadronic calorimeter covers the central pseudorapidity range ($|\eta| < 1.7$). The endcap and forward regions are instrumented with LAr calorimeters for EM and hadronic energy measurements up to $|\eta| = 4.9$. The muon spectrometer (MS) surrounds the calorimeters and is based on three large superconducting air-core toroidal magnets with eight coils each. The field integral of the toroids ranges between 2.0 and 6.0 T m across most of the detector. The muon spectrometer includes a system of precision tracking chambers up to $|\eta| = 2.7$ and fast detectors for triggering up to $|\eta| = 2.4$. The luminosity is measured mainly by the LUCID–2 [18] detector, which is located close to the beampipe. A two-level trigger system is used to select events [19]. The first-level trigger is implemented in hardware and uses a subset of the detector information to accept events at a rate below 100 kHz. This is followed by a software-based trigger that reduces the accepted event rate to 1 kHz on average depending on the data-taking conditions. A software suite [20] is used in data simulation, in the reconstruction and analysis of real and simulated data, in detector operations, and in the trigger and data acquisition systems of the experiment.

3 Simulated samples

The training and evaluation of the jet regression networks presented are performed on a per-jet basis, using jets from simulated pp collisions at $\sqrt{s} = 13$ TeV. A separate mix of samples is used for small-R and large-R jets motivated by kinematic differences. Both include a *training* dataset of samples used consistently for training, validation, and testing and a separate dataset of *evaluation* samples to test the performance on leading simulations of SM resonances. In an analysis setting, flavour-tagging results in a sample dominated by b-jets with a varying degree of non-b-jets depending on the exact requirements applied. Instead of applying requirements on the continuously improving ATLAS flavour-tagging discriminants, the training samples include a mixture of jet flavours, including those without any b-hadron decays, to ensure good performance in various analysis settings.

All simulation samples used in this study are passed through the ATLAS full detector simulation [21] based on GEANT4 [22] and reconstructed using detector geometry and algorithms corresponding to the 2015–2018 data-taking period of the LHC. The effect of multiple interactions in the same and neighbouring bunch crossings (pile-up) was included by overlaying the simulated hard-scattering event with inelastic pp collisions generated with Pythia 8.186 [23] using the NNPDF2.3lo PDF [24] and the A3 set of tuned parameters [25]. The decays of b- and c-hadrons are modelled by EvtGen [26], except for samples generated with the Sherpa [27] event generator where the internal hadrons++ module is used [28].

3.1 Small-R jet samples

The training dataset used in the small-R regression algorithms consists of jets from $t\bar{t}$ production with at least one top quark decaying leptonically, as well as from $Z(\mu\mu)$+jets production enriched in b-jets. The $Z(\mu\mu)$+jets sample increases the statistics of low-p_{T} jets. The dataset is composed of 266 million

b-jets from $t\bar{t}$ decays and 19 million jets from $Z(\mu\mu)$+jets production. A flavour ratio of 9:1:9 is used for light:charm:bottom jets.

The evaluation samples include jets from Higgs boson or Z boson decays. The performance on H decays to two b-jets $H(b\bar{b})$ is evaluated on simulated SM Higgs boson production in association with a leptonically decaying Z boson. Similarly, the SM ZZ production process where one Z boson decays to $b\bar{b}$ and the other decays leptonically is considered to evaluate the performance on a resonance not included in the training dataset.

The details on versions of the MC event generators, tunes and parton distribution functions (PDF) used in the simulation samples for training and evaluating small-R regression algorithm are summarised in Table 1.

Table 1: Details of the simulation samples used for training and evaluating the small-R jet calibration algorithm. $\ell = e, \mu, \tau$. † Enriched in b-jets.

Process	**Generator**	**Parton shower**	**PDF set**
Training, validation and test samples			
$pp \to t\bar{t}$ fully/semileptonic	Powheg [29–31]	Pythia 8.230 [23] with A14 [32]	NNPDF3.0nlo [24]
$pp \to Z(\mu\mu)$+jets†	MadGraph5_aMC@NLO [33] + FxFx [34]	Pythia 8.245 with A14	NNPDF3.0nlo
Evaluation samples			
$pp \to Z(\ell\ell)H(b\bar{b})$	Powheg Box v2 [29–31] + MiNLO [35–37]	Pythia 8.230 with AZNLO [38]	NNPDF3.0nlo
$pp \to Z(\ell\ell)Z(b\bar{b})$	Sherpa [27]	Sherpa 2.2.11	NNPDF3.0nnlo

3.2 Large-R jets samples

The training and evaluation datasets include multiple simulated Higgs boson samples. Jets that contain a Higgs boson decay to $b\bar{b}$ or $c\bar{c}$ pairs are sourced from the simulation of a H boson produced in association with a Z boson (ZH) decaying to muons. To avoid artificial mass peaks being created when evaluating the large-R jet regression network, the training sample is generated with a biased phase-space sampling by choosing the Higgs boson width to be 400 GeV and a Higgs boson mass range between 25–200 GeV. This achieves an approximately uniform, or flat, distribution of the jet mass. In contrast, the evaluation sample consists of SM ZH production where the Higgs boson mass and width are set to SM values; the jet p_{T} has a smoothly falling physical distribution, and the jet mass is sharply peaked. Production of ZH events was simulated at next-to-leading-order (NLO) accuracy in quantum chromodynamics (QCD) using the Powheg program [29–31] as discussed in Ref. [39].

Another process used in both training and evaluation is jet production from purely QCD processes, or multijet production, which includes a mixture of jet flavours. In order to ensure a sufficient population of *QCD jets* at high momenta, the multijet process simulation is performed in slices of leading truth jet p_{T} [40] (further defined in Section 4.) Multijet production, dictated by QCD, contains a small fraction of b-jets. To maximise the number of QCD jets with two b-hadrons for training, a multijet sample, denoted as *multijet* $(b\bar{b})$, is created and requires the presence of four small-R truth jets with $p_{\mathrm{T}} > 15$ GeV of which at least two are b-jets.

Boosted SM resonances, which are not included in the training sample, provide a robust test case for evaluating the network's generalisation and performance on unseen data. Top-quark jets are produced in the decay of a hypothetical Z' boson ($Z' \to t\bar{t}$) of mass 4 TeV. This ensures the population of boosted top-quark jets in the high-p_{T} region is large. For the same reason, $Z(b\bar{b})$ jets are taken from a sample of Z decays to $b\bar{b}$ where the Z boson is produced with a p_{T} above 200 GeV.

The training dataset contains 80 million large-R jets, which consists of 15 million flat-mass $H(b\bar{b})$ jets, 15 million flat-mass $H(c\bar{c})$ jets, 25 million QCD jets from multijet production without flavour selection, and 25 million QCD jets containing two b-hadrons from the multijet ($b\bar{b}$) sample. The number of jets in the evaluation sample is approximately 15 thousand SM $H(b\bar{b})$ jets, 575 thousand $Z(b\bar{b})$ jets, 725 thousand top-quark jets, and 800 thousand QCD jets. The versions of the MC event generators, tunes and parton distribution functions (PDF) used are detailed in Table 2 for training and evaluation samples.

Table 2: Details of the simulation samples used for training and evaluating the large-R jet calibration algorithm. $\ell = e, \mu$. † QCD jets refer to multijet jet production with heavy-flavour hadron content determined from quantum chromodynamics. ‡ The multijet ($b\bar{b}$) sample provides QCD jets with two b-hadrons, QCD ($b\bar{b}$), as described in the text.

Jet type	**Process**	**Event generator and tune**	**PDF set**
	Training, validation and test samples		
$H(b\bar{b})$	$q\bar{q} \to ZH, Z \to \mu^+\mu^-$	Pythia 8.306 [23] with A14 [32]	NNPDF3.0nlo [24]
$H(c\bar{c})$	$q\bar{q} \to ZH, Z \to \mu^+\mu^-$	Pythia 8.306 with A14	NNPDF3.0nlo
QCD †	Multijet	Pythia 8.235 with A14	NNPDF2.3lo
QCD ($b\bar{b}$) ‡	Multijet ($b\bar{b}$), $N_{\text{jet}} \geq 4$, $N_{b-\text{jet}} \geq 2$	Pythia 8.235 with A14	NNPDF2.3lo
	Evaluation samples		
$H(b\bar{b})$	$q\bar{q}/gg \to ZH$, $Z \to \ell\bar{\ell}/\nu\bar{\nu}/q\bar{q}$	Powheg v2 +Pythia 8.212 [30] with AZNLO [38]	NNPDF3.0nlo
Top	$Z' \to t\bar{t}$	Pythia 8.235 with A14	NNPDF2.3lo
$Z(b\bar{b})$	$Z \to b\bar{b}$	Sherpa 2.2.11 [27]	NNPDF3.0nnlo
QCD †	Multijet	Pythia 8.235 with A14	NNPDF2.3lo

4 Object definitions

Multiple types of jet definitions are considered for this study: truth, small-R, and large-R constructed from charged and neutral objects. All jets are created using the anti-k_t algorithm [4] implemented in fastjet [41]. The movement of charged particles through the ID is reconstructed as tracks from individual hits in tracking layers [42]. Neutral and charged particles, including electrons, photons, and hadrons, deposit their energy in the electromagnetic and hadronic calorimeters and are reconstructed as massless topological clusters of calorimeter cells [43]. Muons penetrate the calorimeters, typically depositing only 3 GeV before traversing the MS [44]. Neutrinos pass through the detector without interacting and are not reconstructed.

Tracks and calorimeter clusters are combined to form higher-level flow objects that are utilised to construct jets. Particle flow objects (PFOs) represent a single particle and were designed to improve jet performance at low p_{T}. Charged PFOs are created through the association of tracks and topoclusters to leverage superior energy measurement from the ID at low p_{T} and not double-count energy deposits. This improves the

accuracy of the charged-hadron measurement while retaining the calorimeter measurements of neutral PFOs [2]. Track-calo clusters (TTCs) [45] are flow objects designed to improve reconstruction at high p_{T}, where the excellent energy resolution of the calorimeter complements the precise angular resolution of the tracking system. Unified flow objects (UFOs) are created from the union of the PFO and TCC algorithms to provide optimal performance across a wide kinematic range [3]. UFOs can be charged or neutral depending on the presence of a track object. Tracks used to create muons passing the *Medium* [46] identification criteria are removed from the set of tracks used to create PFO and UFO objects.

4.1 Truth jets

Truth-jets are created by clustering *stable* particles originating from the hard-scatter interaction in the simulation event record with a lifetime τ in the particle rest frame such that $c\,\tau_0 > 10\,\mathrm{mm}$. In particular, b-hadrons are not stable, and only the stable decay products of b-hadrons are used in the clustering algorithm. Particles that do not leave significant energy deposition in the calorimeter (i.e. muons and neutrinos) are generally excluded. For the small-R regression, an alternative to the truth-jet definition used to derive the nominal calibration is employed. This alternative includes leptons from resonance decays, specifically those expected from the decay of a b-hadron. The small-R truth jet clustering algorithm uses the anti-k_t algorithm with radius parameter $R = 0.4$ implemented in FASTJET [41]. Large-R truth jets use as radius parameter $R = 1.0$ and are groomed as those reconstructed from the detector information, incorporating the grooming procedure within the jet definition described below.

4.2 Small-R jets

Jets used in the small-R regression are reconstructed using the same algorithms applied to the corresponding truth jets. They are created by clustering PFOs using the anti-k_t algorithm with radius parameter $R = 0.4$ implemented in FASTJET [41]. Jets must have a Jet Vertex Tagger discriminant of values greater than 0.5 to suppress pile-up contamination [47]. Jets are matched to a truth jet, where the truth jet with the largest p_{T} within $\Delta R < 0.4$ of the reconstructed jet is selected. If no truth jet is found, the jet is discarded. Jets entering the training are required to have reconstructed and truth p_{T} above 10 and 7 GeV, respectively. The selection requirement on the truth p_{T} is looser to avoid biases in the calibration.

Tracks passing the *Loose* [48] quality criteria and vertex-association requirement given in Table 3 are used as input for the regression model. The selection is looser than that used to create charged PFOs. Neutral PFOs are not incorporated into the small-R jet regression network.

Small-R jets are assigned a flavour label depending on the number and type of hadrons with $p_{\mathrm{T}} > 5$ GeV found within $\Delta R < 0.3$ of the jet in question. A b-jet is defined by the presence of a truth b-hadron, c-jets are those with a truth c-hadron and no b-hadrons, while the light jets do not have any heavy-flavour hadrons near the jet axis.

Muons passing the Medium identification criteria are considered for the muon-in-jet correction with their energy deposited in the calorimeter removed to avoid double-counting. For the regression model, rather than applying the muon-in-jet correction directly, *soft-electrons* and *soft-muons* information is provided as input features to the regression network. An electron with $1 < p_{\mathrm{T}} < 50$ GeV and $|\eta| < 2.5$ within the jet cone, which has the highest probability of being from a heavy flavour decay based on the ATLAS Electron Identification tool [50] is identified as a soft electron. Similarly, a muon within the jet cone is identified as *soft* based on the output of the ATLAS Run 2 *Soft Muon Tagger* tool [51].

Table 3: Loose [48] track selection requirements used for the small-R jet calibration algorithm, where d_0 is the transverse impact parameter (IP) of the track, z_0 is the longitudinal IP with respect to the primary vertex and θ is the track polar angle. Shared hits are hits used in the reconstruction of multiple tracks which have not been classified as split by the cluster-splitting neural networks [49]. A hole is a missing hit, where one is expected, on a layer between two other hits on a track.

Parameter	**Requirement**
Track selection	
p_{T}	> 500 MeV
Silicon hits	≥ 8
Shared silicon hits	≤ 1
Silicon holes	< 2
Pixel holes	< 1
Track-to-vertex association	
$\lvert d_0 \rvert$	< 3.5 mm
$\lvert z_0 \sin\theta \rvert$	< 5 mm

4.3 Large-R jets

Large-R jets are built from UFOs using the anti-k_t algorithm with radius parameter $R = 1.0$ implemented in FASTJET [3]. Pile-up and underlying event contributions are removed via grooming with the Soft-Drop algorithm [52, 53] along with Constituent Subtraction [54] and SoftKiller [55]. Jets are matched to a truth jet where the truth jet with the largest p_{T} within $\Delta R < 0.75$ of the reconstructed jet is selected. The truth definition used for the large-R jet studies does not include leptons from resonance decays. Large-R jets are required to have $p_{\mathrm{T}} > 200$ GeV and $|\eta| < 2$.

UFO constituents are the primary regression model input and must satisfy the selection criteria based on their properties. Tracks used to construct UFOs must pass the *Tight Primary* [48] criteria and be associated with the primary vertex (PV) using a working point that is designed to select all prompt and non-prompt tracks from a given vertex. The Run 2 adaptive multi-vertex finder [56] uses a weighted Kalman filter to minimise the sum of standardised distances ($\chi = d/\sigma_d$) of tracks to the vertex. Track compatibility is evaluated through a χ^2-like measurement called *weight*. The "MaxWeight" working point assigns tracks used in any vertex fit to the vertex for which they have the highest weight. If a track is not used in any vertex fit, the selection reverts to impact parameter (IP) requirements ($|d_0| < 5$ mm and $|z_0 \sin\theta| < 5$ mm).

Additional tracks that are not associated with the selected UFOs and are ghost-associated [7, 57] to the large-R jet are also used as model inputs if they satisfy a looser track selection criteria and are compatible with the PV. Both the UFO and looser track selection criteria are summarised in Table 4.

Large-R jets are assigned a type and flavour label depending on the number and type of truth particles matched to the jet using ghost association. Jet type is defined by the presence of and compatibility with a massive truth particle checked in order from top quark, W, Z, then H [58, 59]. A jet failing to match any massive truth particle is labelled as a QCD jet. Top-quark jets are required to contain the subsequent hadronic W boson decays and a b-hadron. H jet flavour is $H(b\bar{b})$ if two b-hadrons are associated with the jet. Otherwise, it can be an $H(c\bar{c})$ jet if two c-hadrons are found. A similar definition is used for $Z(b\bar{b})$ jets. The QCD jet flavour label is determined by first counting the number of b-hadrons. If there are fewer than two, c-hadrons are counted. If there are no b- and c-hadrons, the jet is labelled as a light jet.

Table 4: UFO constituents and additional track selection requirements, where d_0 is the transverse impact parameter (IP) of the track, z_0 is the longitudinal IP with respect to the primary vertex and θ is the track polar angle. Shared hits are hits used in the reconstruction of multiple tracks which have not been classified as split by the cluster-splitting neural networks [49]. A hole is a missing hit, where one is expected, on a layer between two other hits on a track. N_{IBL} and $N_{\text{B-Layer}}$ are the number of hits a track has on the IBL and B-Layer, respectively. UFO selection corresponds to *Tight Primary* [48] track selection and "MaxWeight" track-to-vertex operating point [56]. †IP requirements are used for tracks not included in any vertex fit. See the text for more information.

Parameter	**UFO constituents requirement**	**Additional track requirement**
	Track selection	
p_{T}	> 500 MeV	> 500 MeV
Silicon hits	$\geq 9(11)$ if $\lvert\eta\rvert \leq 1.65(\geq 1.65)$	≥ 8
Shared silicon hits	≤ 1	–
Silicon holes	≤ 2	< 3
Pixel holes	0	< 2
$N_{\text{IBL}} + N_{\text{B-Layer}}$	> 0	–
	Track-to-vertex association	
Tracks in vertex fits	MaxWeight	–
$\lvert d_0\rvert$	< 5 mm†	< 5 mm
$\lvert z_0 \sin\theta\rvert$	< 5 mm†	< 5 mm

For training, all jets with p_{T} in the range 200 GeV to 1.5 TeV, $|\eta| < 2.0$ and reconstructed invariant mass in the window $20\,\text{GeV} < m_{\text{J}} < 300\,\text{GeV}$ when applying the nominal calibration are considered. The relatively low mass cut for the training sample ensures the truth mass distribution is unbiased above 50 GeV, allowing for a smooth response from the neural network predictions. For evaluation, the mass window is changed to $40\,\text{GeV} < m_{\text{J}} < 300\,\text{GeV}$ after the jet regression network correction has been applied.

5 Neural network architecture

5.1 Network architecture

The model architecture used for jet calibration is based on the GN2 and GN2X flavour-tagging models for small- and large-R jets [9, 10]. Both models utilise the same overall architecture as the GN1 model described in Ref. [9], but use the Transformer network architecture [60] as in the GN2-type models instead of a Graph Neural Network. In the following, the architecture is detailed, and schematics can be found in Appendix A.

The models take as input a set of jet kinematic properties and variables associated with each charged flow object or track and neutral flow object (later referred to as "constituent-level"). The full list of features is outlined in Section 5.2. Jet- and constituent-level inputs are concatenated, and the combined jet-constituent sequence vectors are fed into a per-constituent initialiser network. In the case of large-R jets, additional charged and neutral UFO information is fed into separate initialiser networks to allow separate representations to be learned for each input type. Each initialiser network uses a Deep Sets style [61] architecture but does not contain a reduction operation over the output constituent representations. The initialiser network for each input type consists of a single dense layer of size 256 and a rectified linear unit (ReLU) activation function [62–64] projecting the input representations to an embedding dimension of 256.

The constituent representations are fed into a Transformer Encoder where the transformer architecture utilised follows that introduced in Ref. [65]. Multiple Layer Normalisation layers [66] are used to aid in providing stability during training along with residual connections. The small-R jet (large-R jet) regression model employs 4 (3) encoder blocks with 8 (2) attention heads. The model does not use separate transformer encoders for each input type due to the significant increase in the number of parameters required.

The output representation of each constituent is then combined to form a global representation of the jet to be used for calibration. This global representation is formed by a weighted sum over the constituent representations, where the attention weights for the sum are learned during training.

The training targets for each regression are the ratio of truth and reconstructed level values of a kinematic variable. Each regression task takes the global jet representation as input, and the small-R jet (large-R jet) model consists of a dense neural network with four layers of size 128 (256), 128, 64, 32 neurons, followed by a ReLU (Mish [67]) activation layer. The regression loss function used is the Mean Absolute Error (MAE), also known as L1 loss. The total number of trainable parameters in the small-R jet (large-R jet) model is approximately 1.7 (2.4) million.

The training datasets are split 80/10/10 into training, validation, and test samples. The training and validation samples are used to train a network and monitor for overtraining. In contrast, the test sample is used to evaluate the model's performance on a per-jet basis.

5.2 Input features

The input objects for the small-R jet regression network are tracks within the jet that pass the requirements in Table 3 as discussed in Section 4.2. Therefore, the jet regression is trained on jet-level kinematic features and the constituent-level features from tracks listed in Table 5 as well as soft lepton features listed in Table 6.

The input objects and features used for the large-R jet regression network are the same as the GN2X flavour-tagging network [10]. The input objects are all the UFO jet constituents and additional tracks within the jet that pass the requirements in Table 4 as discussed in Section 4.3. Jet-level kinematic features, constituent-level features from tracks and both types of UFOs listed in Table 5 are used. Two separate tasks are used for mass and p_{T} regression.

The effect of simplifying the set of input features on the regression performance was studied. Four soft muon variables measure significance of compatibility between ID and MS tracks, as well as the track and the PV location; these variables are not highly correlated to jet energies but rather represent object quality. The track q/p is highly correlated to another input feature, p_{T}. The calorimeter variables for soft electrons help distinguish electrons from other objects and any derived improvement to the electron energy measurement is small compared to the total jet energy. Removing all of these variables increases the width of the response calculated with the small-R jet regression network by approximately 2%. A suite of input feature ablation studies is left for future studies.

Adopted from the flavour-tagging models, a comprehensive description of each track or charged flow object – including the two IPs and their associated significances, the track momentum and angular distance from the jet with the associated uncertainties, and nine hit multiplicity counts – is provided to the networks. The optimal set of input features can differ for classification and regression networks. For example, track hit content influences the IP resolution and, thus, flavour-tagging performance, but it does not strongly correlate with jet kinematics. Removing all nine track hit multiplicity variables from the large-R network

degrades the model's achieved mass resolution by about 1%. Removing the variables related to the IPs results in an approximate 3% performance loss. Other feature ablations would eliminate essential energy and angular information and are therefore left for future studies.

Table 5: Input features to the regression models. Features are separated into jet features, track and charged UFO constituent (flow) features, and charged and neutral UFO constituent features. Tracks and charged UFO constituents have a common set of input features related to the ID and tracking. Charged and neutral UFO constituents have a separate set of common features related to the calorimeter energy measurements. Only features associated with jets, tracks and ones marked with † are used for small-R jets. Jet features, constituent-level features from tracks and both types of UFOs, and ones marked with ‡ are used for large-R jets.

Jet feature	**Description**
p_{T}	Transverse momentum
η	Signed pseudorapidity
m ‡	Jet mass
Track & charged UFO feature	**Description**
q/p	Track charge divided by reconstructed momentum
$\mathrm{d}\eta$	Pseudorapidity of track relative to the jet η
$\mathrm{d}\phi$	Azimuthal angle of the track, relative to the jet ϕ
d_0	Transverse IP: Closest distance from track to beam-line in the transverse plane
$z_0 \sin\theta$	Longitudinal IP: Closest distance from track to PV in the longitudinal plane
$\sigma(q/p)$	Uncertainty on q/p
$\sigma(\theta)$	Uncertainty on track polar angle θ
$\sigma(\phi)$	Uncertainty on track azimuthal angle ϕ
$s(d_0)$	Significance of transverse IP
$s(z_0 \sin\theta)$	Significance of longitudinal IP times the sin of the polar angle
nPixHits	Number of pixel hits
nSCTHits	Number of SCT hits
nIBLHits	Number of IBL hits
nBLHits	Number of B-layer hits
nIBLShared	Number of shared IBL hits
nIBLSplit	Number of split IBL hits
nPixShared	Number of shared pixel hits
nPixSplit	Number of split pixel hits
nSCTShared	Number of shared SCT hits
LeptonID †	Information on if the track was used in lepton reconstruction
Charged & neutral UFO feature	**Description**
$p_{\mathrm{T}}^{\mathrm{Flow}}$ ‡	Transverse momentum of charged flow constituent
E_{Flow} ‡	Energy of charged flow constituent
$\mathrm{d}\eta_{\mathrm{Flow}}$ ‡	Pseudorapidity of track relative to the large-R jet η
$\mathrm{d}\phi_{\mathrm{Flow}}$ ‡	Azimuthal angle of the track, relative to the large-R jet ϕ
$\mathrm{d}r_{\mathrm{Flow}}$ ‡	Angular distance of the track from the large-R jet direction

6 Regression performance

The performance of small-R and large-R jet regression networks is presented in the relevant phase space, with the response defined as the ratio of a reconstructed jet quantity over the corresponding truth jet quantity. The jet p_{T} response is presented for both small-R and large-R jets, and the jet mass response is

Table 6: Additional soft muon and electron input features used for the small-R jets regression [50, 51].

Soft Muon Input	**Description**
p_{T}	Transverse momentum
η	Signed pseudorapidity
ϕ	Azimuthal angle
dR	Angular distance of the soft muon from the small-R jet axis
q/p	Muon charge divided by the reconstructed momentum
Momentum Balance Significance	Ratio of the difference in momentum measured by the ID and MS to the uncertainty on the energy loss measured by the calorimeters
Scattering Neighbour Significance	Sum of the significances of the angular difference $\Delta\phi$ between pairs of adjacent hits along the track, multiplied by the particle charge
$p_{\mathrm{T}}^{\mathrm{rel}}$	Orthogonal projection of the muon p_{T} onto the jet axis
d_0	Transverse IP: Closest distance from track to beam-line in the transverse plane
z_0	Longitudinal IP: Closest distance from track to PV in the longitudinal plane
$\sigma(d_0)$	Uncertainty on measurement of transverse IP
$\sigma(z_0)$	Uncertainty on measurement of longitudinal IP
$d_0/\sigma(d_0)$	Significance of transverse IP
$z_0/\sigma(z_0)$	Significance of longitudinal IP
Soft Electron Input	**Description**
$p_{\mathrm{T}}^{\mathrm{r}}$	Relative p_{T} of the electron with respect to the jet
dR	Angular separation between electron and jet axis
$p_{\mathrm{T}}^{\mathrm{iso}}$	Isolation variable
$\lvert\eta\rvert$	Absolute value of pseudorapidity
$s(d_0)$	Transverse IP: Closest distance from track to beam-line in the transverse plane
$z(d_0)$	Longitudinal IP: Closest distance from track to PV in the longitudinal plane
$s(d_0/\sigma_{d_0})$	Significance of the transverse IP
$\Delta\phi^{\mathrm{res}}$	The azimuthal angle difference $\Delta\phi$ between the cluster position in the middle layer and the track.
E/p	Ratio of the cluster energy to the track momentum
R_{had}	Ratio of E_{T} in the hadronic calorimeter to E_{T} of the EM cluster
R_{had1}	Ratio of transverse energy E_{T} in the first layer of the hadronic calorimeter to E_{T} of the EM cluster
E_{ratio}	Ratio of the energy difference between the largest and second-largest energy deposits in the cluster over the sum of these energies
$w_{\eta 2}$	Lateral shower width
R_{η}	Ratio of the energy in 3×7 cells over the energy in 7×7 cells centered at the electron cluster position
f_1	Ratio of the energy in the strip layer to the total energy in the EM accordion calorimeter
f_3	Ratio of the energy in the back layer to the total energy in the EM accordion calorimeter
p_{HF}	Probability of being from heavy flavour decay

presented for large-R jets. Key performance metrics include the median response and the relative response resolution, defined as one-half the inter-quantile range (IQR) between 15.9% and 84.1% percentiles, divided by the median. Achieving a median response close to one and uniform across the phase space is crucial for minimising the complexity and uncertainties in the ultimate calibration procedure. Resolution improvements are quantified using the root square difference (RSD) defined as $\mathrm{sgn}\,(\sigma' - \sigma)\sqrt{|\sigma'^2 - \sigma^2|}$, where σ (σ') denotes the nominal calibration (regression) relative resolution. Resolutions closer to zero are indicative of better performance as the reconstructed values are closer to the true values. The uncertainty on the median and IQR are both calculated as the standard deviation of the same quantity calculated with 100 sub-samples.

6.1 Small-R jet regression network performance

This section describes the performance of the small-R regression model trained on the training dataset defined in Section 3.1 using the Transformer network and baseline inputs defined in Section 5. With the simulation-based corrections of the nominal calibration applied [5], the small-R regression network performance is compared to the nominal calibration with and without the muon-in-jet and PtReco corrections. Performance is only shown for jets matched to truth b-jets and satisfying $p_{\mathrm{T}}^{\mathrm{truth}} > 20$ GeV, imitating the minimum kinematic selection generally applied to b-jets in physics analyses.

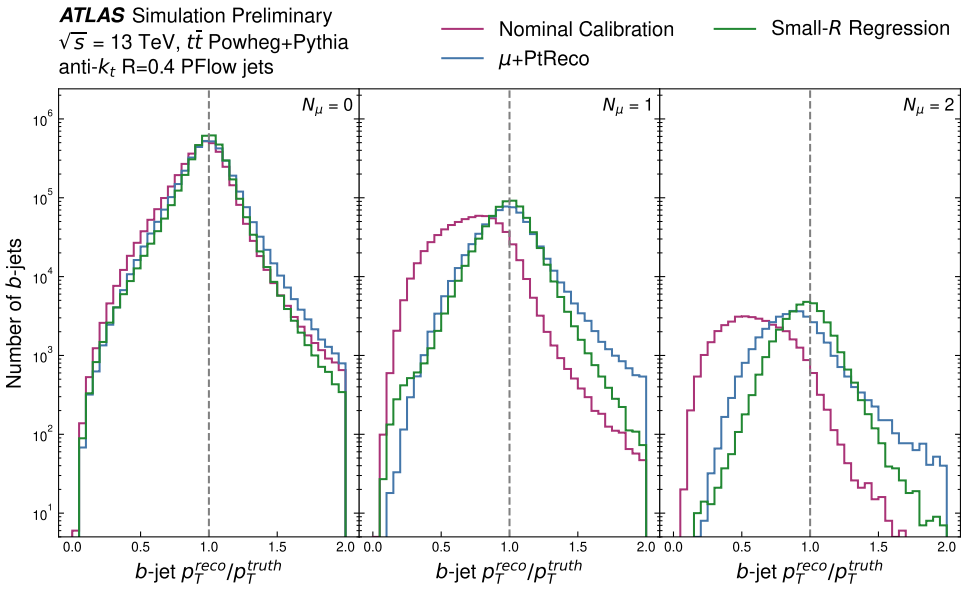

Figure 1: Distribution of b-jet p_{T} response split by the number of muons in the jet. The performance is evaluated on the b-jets in the $t\bar{t}$ samples from the test set of the training samples listed in Table 1. Uncertainties are not shown.

Figure 1 shows the distribution of b-jet p_{T} response split by the number of muons in the jet. The absolute number of jets is presented, indicating the roughly 15% rate of finding a muon within the jet. Jets without muons typically have a hadronically decaying b-hadron, leading to a response distribution for the nominal calibration centred at unity with an expected upward shift from the PtReco correction. However, the nominal calibration response for jets with at least one muon shows a decrease in median response of around 25%.

Figure 2 shows the b-jet response as a function of the truth jet p_{T}. The median p_{T} response of the nominal calibration is within 5% of unity. It changes non-linearly with jet p_{T}, while the median response after the PtReco and muon-in-jet corrections varies by up to 20%, and the median response of the regression changes by up to 10% over the p_{T} range studied. A smaller relative p_{T} resolution translates to a better resolution on the mass of heavy particles decaying to two separate, or resolved, b-jets. For b-jets with p_{T} within 40-100 GeV, typical for decays of heavy SM particles ($W/Z/H/t$), the PtReco correction improves the relative resolution by 20% while the regression model improves by around 30%.

Figure 3 shows the b-jet response as a function of the number of muons in the jet. The response is stable for jets with at least one muon, while the nominal calibration response is degraded for jets with muons. The regression model improves the relative resolution for all muon multiplicities by 5% for jets with exactly one muon and 15% for jets with two or more muons relative to the PtReco correction.

The relative scalar sum of track p_{T} over the jet p_{T} ($\Sigma p_{\mathrm{T}}^{\mathrm{track}}/p_{\mathrm{T}}^{\mathrm{jet}}$) is used within the jet calibration pipeline as part of the simulation-based corrections known as the global sequential calibration. This correction improves the jet p_{T} resolution and associated uncertainties by removing the dependence of the reconstructed jet response on jet observables [5]. The DNN replacement uses similar information [7] but was not available for the studies presented here. Figure 4 shows the b-jet response as a function of the relative scalar sum of track p_{T}. There is a substantial reduction in the median response of nominal calibration p_{T} when the track p_{T} sum is greater than 40% of the calibrated jet p_{T}. The PtReco procedure over-corrects the jet p_{T} when the sum of the track p_{T} is less than 80% of the jet p_{T} and mitigates the strong reduction in the median response seen in the nominal calibration alone. When the regression model is used, the median response is stable and always within 2% of unity. The largest region of improvement from the regression model is when the scalar sum of track p_{T} is greater than the jet momentum.

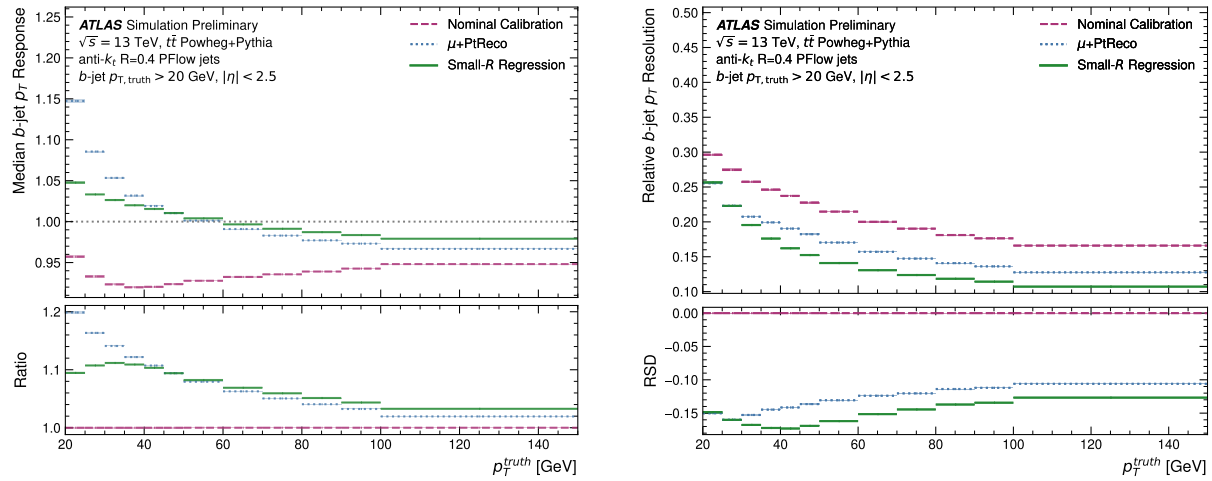

Figure 2: The median (left) and relative resolution (right) of the b-jet p_{T} response are plotted as a function of truth jet p_{T}. The performance is evaluated on b-jets in the $t\bar{t}$ samples from the test set of the training samples listed in Table 1. The width of the coloured lines represents the statistical uncertainty. In the bottom left panel, the ratio of the muon-in-jet and PtReco (μ+PtReco) median and the jet regression network median to the nominal calibration median is shown. In the bottom right panel, the RSD compares the relative resolution of the μ+PtReco and the jet regression network to the nominal calibration.

To further evaluate the performance of small-R regression, the invariant mass of either the Higgs or the Z boson is reconstructed in SM $Z(\ell\ell)H(b\bar{b})$ or $Z(\ell\ell)Z(b\bar{b})$ evaluation samples, respectively. The invariant

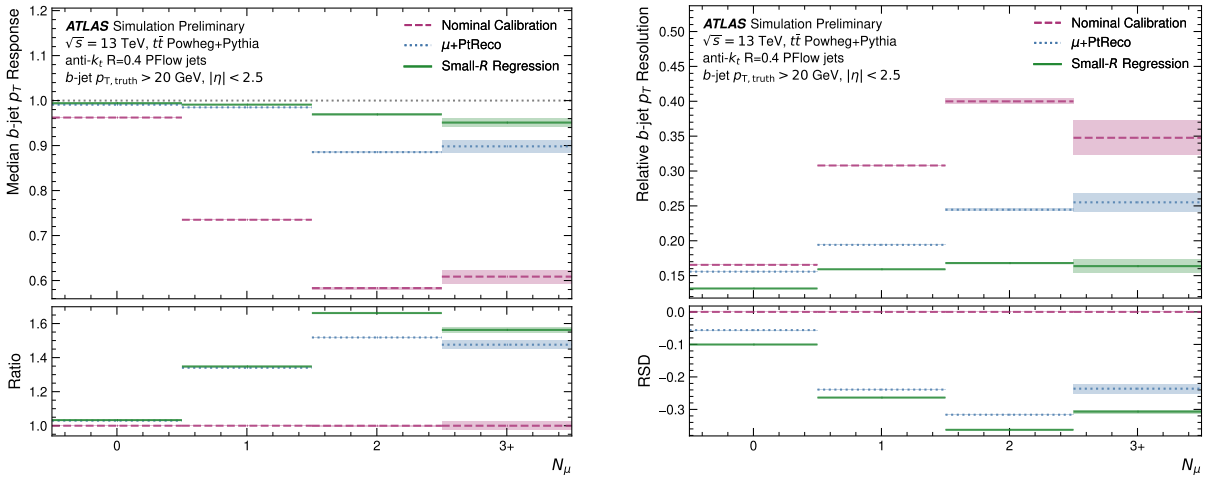

Figure 3: The median (left) and relative resolution (right) of the b-jet p_{T} response are plotted as a function of number of muons found in the jet cone. The performance is evaluated on b-jets in the $t\bar{t}$ samples from the test set of the training samples listed in Table 1. The width of the coloured lines represents the statistical uncertainty. In the bottom left panel, the ratio of the muon-in-jet and PtReco (μ+PtReco) median and the jet regression network median to the nominal calibration median is shown. In the bottom right panel, the RSD compares the relative resolution of the μ+PtReco and the jet regression network to the nominal calibration.

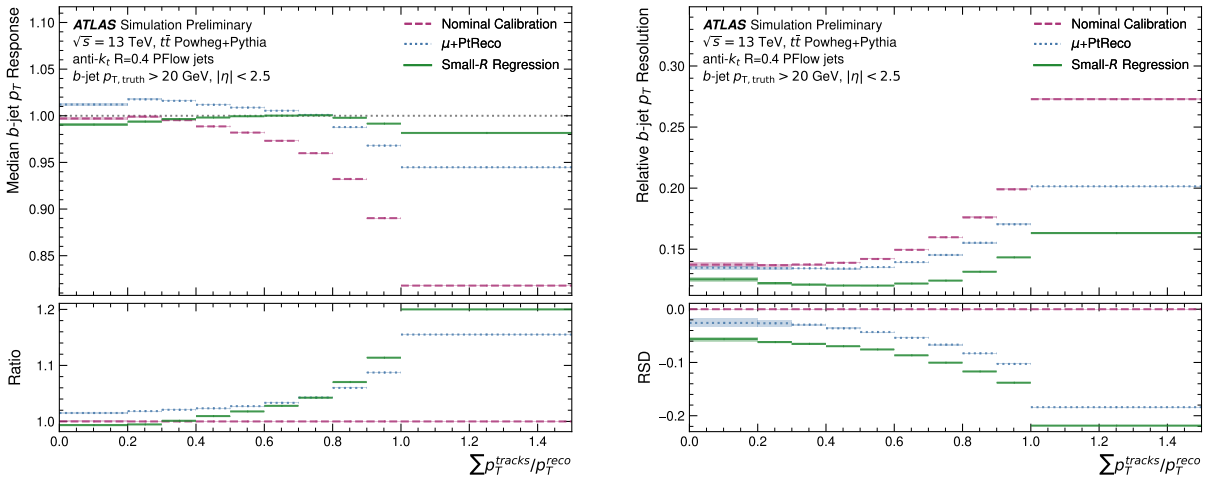

Figure 4: The median (left) and relative resolution (right) of the b-jet p_{T} response are plotted as a function of relative scalar sum of track p_{T} ($\Sigma p_{\mathrm{T}}^{\mathrm{track}}/p_{\mathrm{T}}^{\mathrm{jet}}$). The performance is evaluated on b-jets in the $t\bar{t}$ samples from the test set of the training samples listed in Table 1. The width of the coloured lines represents the statistical uncertainty. In the bottom left panel, the ratio of the muon-in-jet and PtReco (μ+PtReco) median and the jet regression network median to the nominal calibration median is shown. In the bottom right panel, the RSD compares the relative resolution of the μ+PtReco and the jet regression network to the nominal calibration.

mass $m_{b\bar{b}}$ of the parent particle reconstructed using the two b-jets after applying nominal calibration, muon-in-jet and PtReco corrections, and small-R regression calibration are compared. Figure 5 shows the $m_{b\bar{b}}$ distribution from $Z(b\bar{b})$ and $H(b\bar{b})$ decays in the mass range $m_{b\bar{b}} < 200$ GeV. Only events with exactly two b-jets are selected to reconstruct the shown invariant mass.

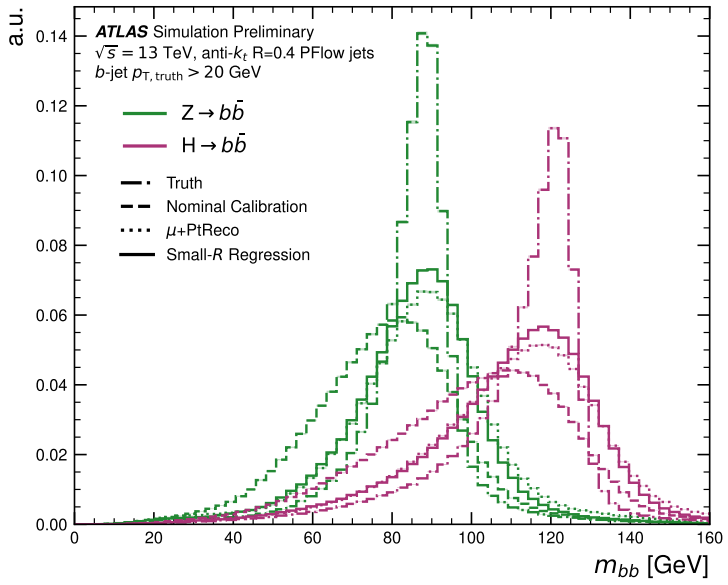

Figure 5: Invariant mass of the two b-jets using truth b-jets, b-jets reconstructed using the nominal calibration, muon-in-jet and PtReco corrections, as well as small-R regression calibration using a loose event selection. The truth-particle heavy-flavour hadron content defines a b-jet. The width of the coloured lines represents the statistical uncertainty.

Figure 6 presents the individual mass distributions for different samples and processes. For the $Z(b\bar{b})$ process, the median values of the mass distribution are 79.0 GeV for the nominal calibration and 87.6 GeV for the muon-in-jet plus PtReco corrections. When using the small-R regression, the median $m_{b\bar{b}}$ shifts to 86.5 GeV. This adjustment results in a 22% improvement in the relative response resolution compared to the nominal calibration and a 6% improvement over the muon-in-jet plus PtReco corrections. Similarly, for the $H(b\bar{b})$ process, the median $m_{b\bar{b}}$ values are 103.0 GeV for the nominal calibration and 112.8 GeV for the muon-in-jet plus PtReco corrections. With the regression-calibrated b-jets, the median $m_{b\bar{b}}$ shifts to 112.9 GeV. This shift corresponds to a 23% enhancement in resolution compared to the nominal calibration and a 6% enhancement over the muon-in-jet plus PtReco corrections.

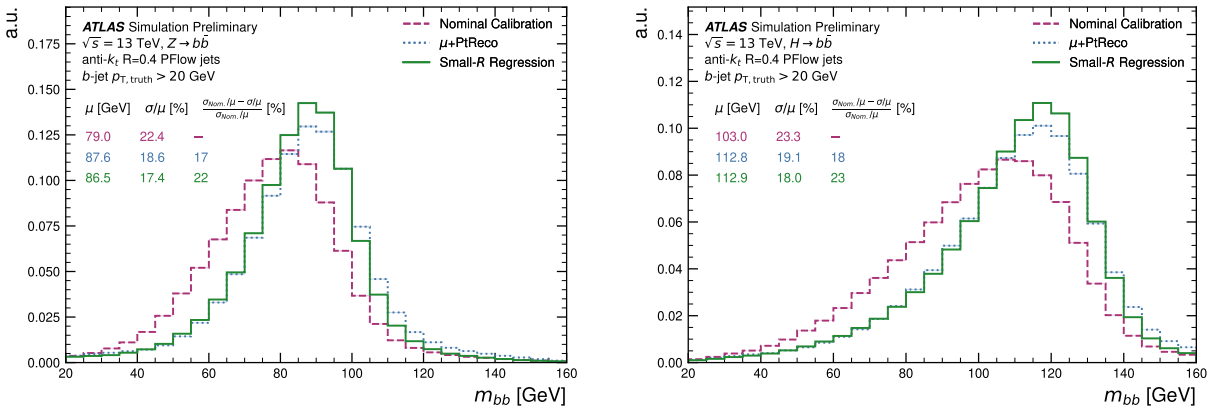

Figure 6: Reconstructed invariant mass of the two b-jets with a loose event selection using the nominal calibration, muon-in-jet and PtReco corrections, as well as small-R regression calibration in $Z(b\bar{b})$ (left) and $H(b\bar{b})$ (right) events. The μ and σ/μ show the median and the relative resolution of the plotted $m_{b\bar{b}}$ distributions. The relative resolution is defined as the half of the central 68.2% quantile ratio to the median of the response. The width of the coloured lines represents the statistical uncertainty. The truth-particle heavy-flavour hadron content defines a b-jet.

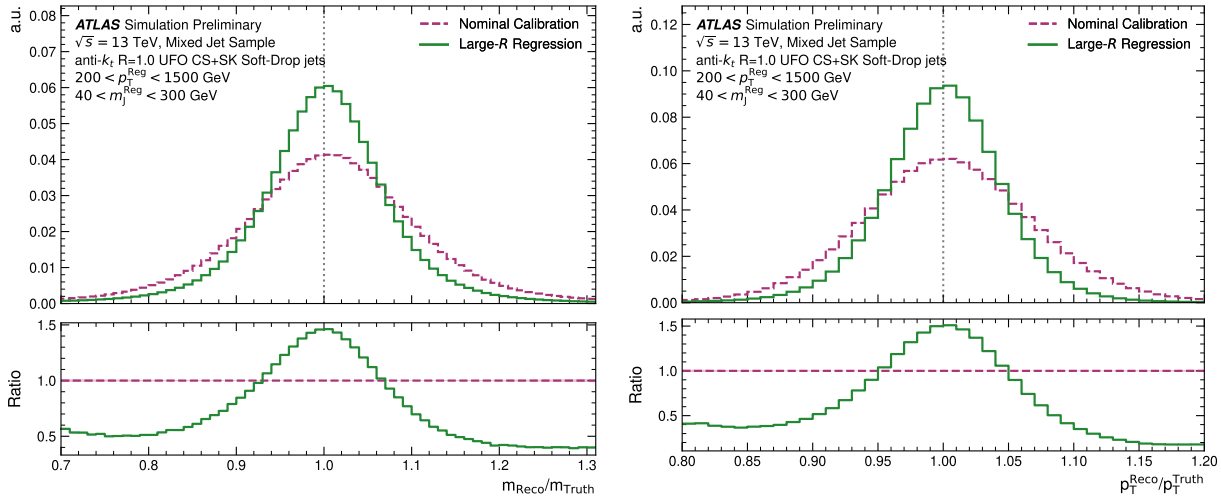

Figure 7: The mass (left) and p_{T} (right) large-R jet calibration response distributions. Response is defined as the ratio of truth to reconstructed level jet mass and p_{T}, respectively. The mixed jet sample denotes the test set of samples listed in the first half of Table 2 and is a mixture of $H(b\bar{b})$, $H(c\bar{c})$ and QCD jets. The bottom panels show the ratio of the jet regression network response to the nominal calibration response.

6.2 Large-R jet regression network performance

This section describes the performance of the large-R regression model trained on the training dataset defined in Section 3.2 using the Transformer network and inputs defined in Section 5. The performance is evaluated using several metrics. Mass and p_{T} regressions are considered separately from each other. As leptons are not included in the large-R jet truth inputs, the muon-in-jet correction is not considered.

The response, median and relative resolutions are defined above in Section 6 for both the jet p_{T} and mass. Response distributions can be found in Figure 7, which show the model performance for the test set (from the training samples listed in Table 2). For mass regression, the response peak of the large-R jet model predictions offers a nearly 30% resolution improvement compared to the nominal calibration. The median of the mass response is also shifted closer to the optimal value. However, the median for both distributions is less than a percentage away from one, so the improvement is relatively small. Similarly, the p_{T} regression achieves 33% better response resolution and slightly improves the median.

A more detailed look at the calibration performance in different p_{T} regimes can be found in Figure 8 and Figure 9 for mass and p_{T}, respectively. For both, the relative resolution is more uniform across all p_{T} bins, ensuring consistent performance. The median of the mass response shows clear improvements for p_{T} $>$ 500 GeV. For 200 $< p_{\mathrm{T}} <$ 400 GeV jets, the relative deviation from the mean is of the same magnitude but in opposite directions. The deviation in the median of the p_{T} response is reduced by the regression for 200 $< p_{\mathrm{T}} <$ 700 GeV and is comparable for higher values. The relative p_{T} resolution is consistently 25-35% better across the studied p_{T} range.

To evaluate the large-R jet mass calibration further, the mass distributions between the truth jet and reconstructed jets with either the nominal calibration or regression model predictions are compared for the evaluation samples. Figure 10 shows an overview of all samples in the mass range 50 $< m_{\mathrm{J}} <$ 230 GeV, while Figure 11 shows each sample and mass peak individually. In both figures, the jet p_{T} range is 400–1500 GeV to ensure the SM Higgs jets have sufficient Lorentz boost to capture both b-hadrons within

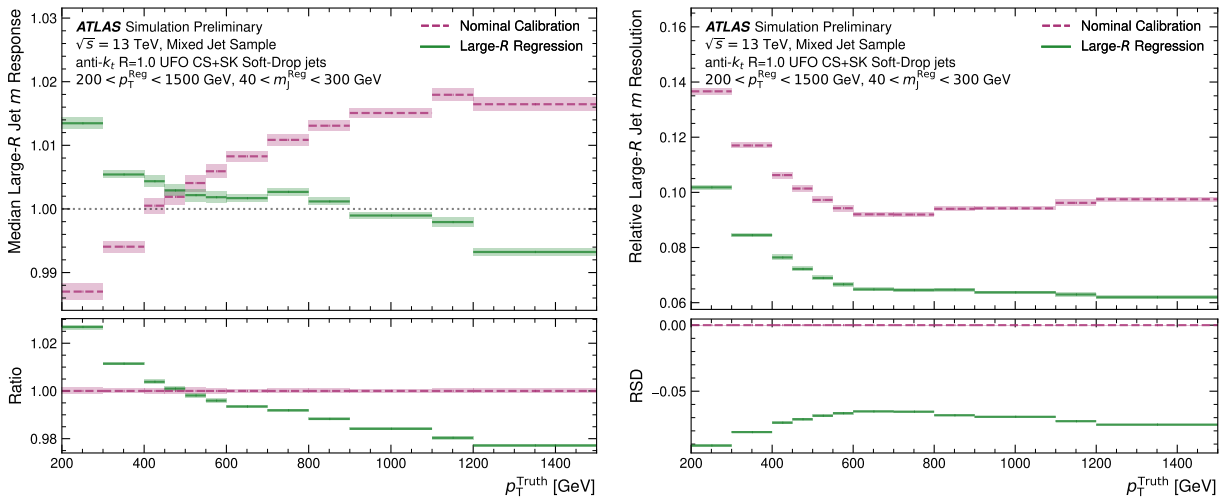

Figure 8: (left) The jet mass response median as a function of truth jet p_{T}. Response is defined as the ratio of truth to reconstructed level jet mass. The bottom panel shows the ratio of the jet regression network median to the nominal calibration median. (right) The relative jet mass resolution as a function of truth jet p_{T}. Relative resolution is defined as the half of central 68.2% quantile ratio to the median of the response. The RSD in the bottom panel compares the relative resolution of the nominal calibration and large-R regression. In both plots, the mixed jet sample denotes the test set of samples listed in the first half of Table 2 and is a mixture of $H(b\bar{b})$, $H(c\bar{c})$ and QCD jets. The width of the coloured lines represents the statistical uncertainty.

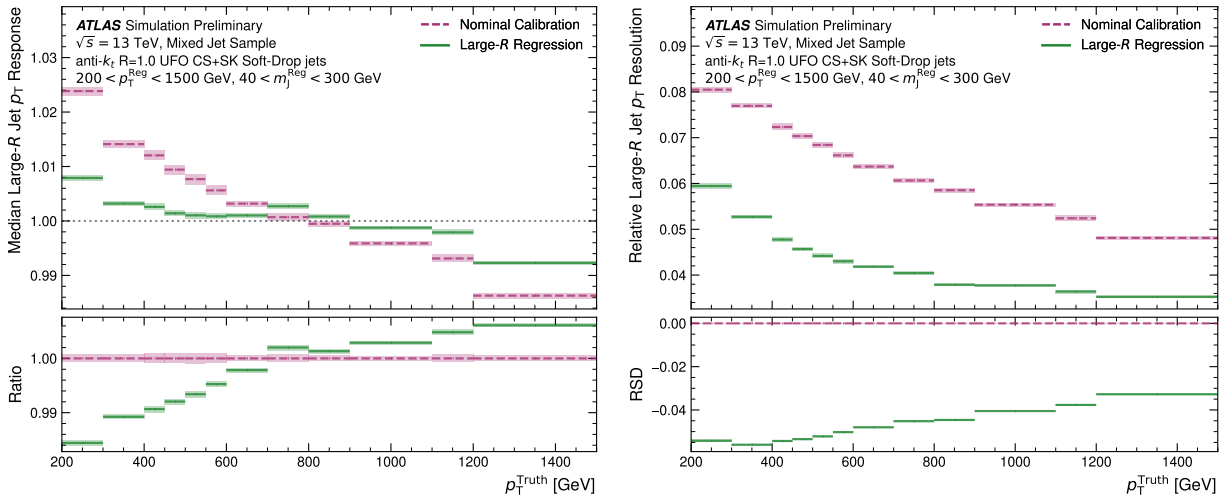

Figure 9: (left) The p_{T} response median as a function of truth jet p_{T}. Response is defined as the ratio of truth to reconstructed level jet mass. The bottom panel shows the ratio of the jet regression network median to the nominal calibration median. (right) The relative p_{T} resolution as a function of truth jet p_{T}. Relative resolution is defined as the half of central 68.2% quantile ratio to the median of the response. The RSD in the bottom panel compares the relative resolution of the nominal calibration and large-R regression. In both plots, the mixed jet sample denotes the test set of samples listed in the first half of Table 2 and is a mixture of $H(b\bar{b})$, $H(c\bar{c})$ and QCD jets. The width of the coloured lines represents the statistical uncertainty.

a jet. Only jets within the limits of the plots in Figure 11 are considered for the median and resolution estimations.

The large-R jet regression model achieves an improvement in resolution for all samples of around 10–15%, even for $Z(b\bar{b})$ and top-quark jets that were not explicitly present in the training dataset. Defining the peak position as the median of the mass distributions in Figure 11, the differences between the truth jets, the nominal jet calibration, and the regression network are on the order of 1–3 GeV. Artificial peaks are not created within the QCD sample.

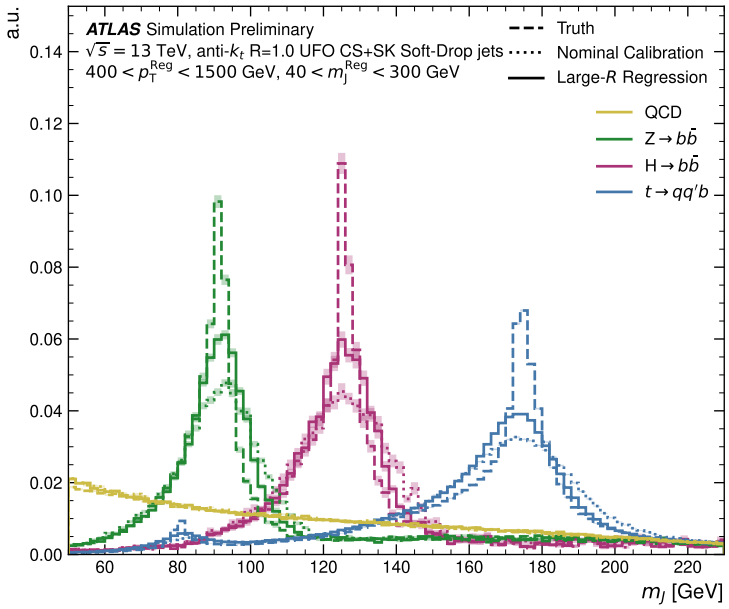

Figure 10: Mass distributions for truth, jets reconstructed using the nominal calibration, as well as large-R regression model predictions. The distributions are shown for the evaluation samples listed in Table 2 where QCD jets are from multijet jet production with the heavy-flavour hadron content determined from quantum chromodynamics. The width of the coloured lines represents the statistical uncertainty.

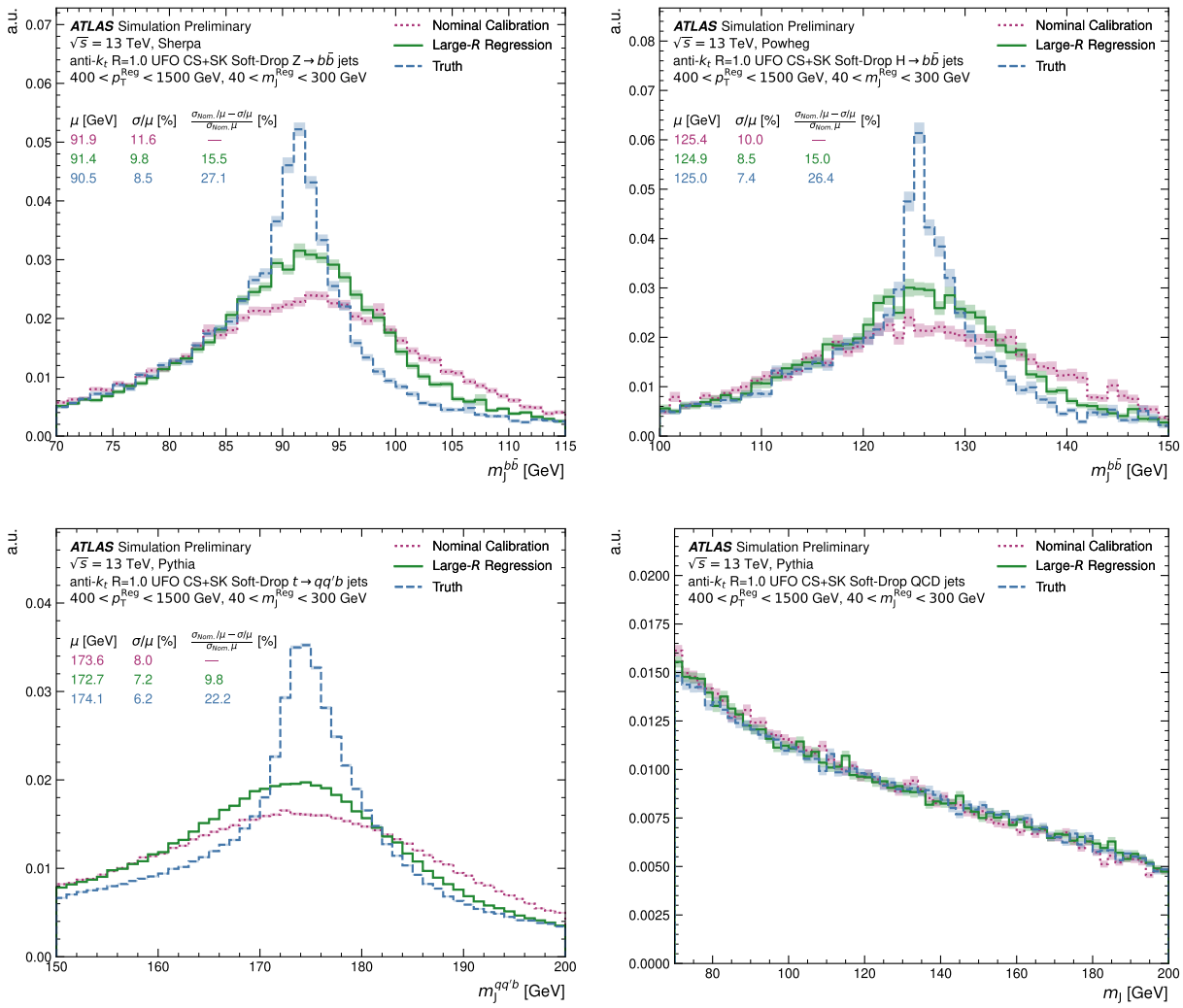

Figure 11: Mass distributions for truth, jets reconstructed using the nominal calibration, as well as large-R regression model predictions. The distribution is shown for the evaluation samples listed in Table 2: Z boson (top left), Higgs boson (top right), top quark (bottom left) and QCD jets (bottom right). QCD jets are from multijet jet production with the heavy-flavour hadron content determined from quantum chromodynamics. Besides the inherent differences between a scalar and a boson, the difference between the shape of the $H(b\bar{b})$ and $Z(b\bar{b})$ truth mass peaks can be attributed, in part, to the difference from MC generators [68–70]. The width of the coloured lines represents the statistical uncertainty.

7 Conclusion

New methods to calibrate the energy of b-jets are presented for both small-R and large-R jets. The new methods are based on transformer encoder neural networks first deployed by ATLAS in the context of flavour tagging. This note presents the application of these models to regress the transverse momentum and, in the case of large-R jets, also the mass of jets. The training of these algorithms uses simulated samples that are enriched in b-jets. The input information used in these algorithms includes low-level features such as jet constituents and tracks associated with the jets. The algorithms show significant improvements in the jet p_{T} response and relative resolution over the nominal calibrations employed in ATLAS.

In the case of small-R jets, evaluating the performance on true b-jets, the response is closer to the true jet p_{T} and the improvements in relative resolution range between 18% and 31% depending on the transverse momentum regime. The largest improvements are achieved for jets with a true jet p_{T} between 40 GeV and 100 GeV. The method also shows improvements to the methods previously explored by ATLAS analyses, muon-in-jet and PtReco corrections, in the true jet p_{T} regime between 30 GeV and 150 GeV and comparable performance in regime with lower or higher p_{T}. Considering the presence of at least one muon inside the reconstructed small-R jet, the presented regression method shows considerable improvements over nominal calibration methods, clearly demonstrating the ability to correct for the energy of muons and neutrinos that are not clustered within the jet and are not considered in nominal calibration methods. The regression model improves the resolution of the reconstructed Higgs and Z boson mass distributions by around 22% and 6% when compared to the nominal, and muon-in-jet and PtReco calibrations, respectively.

Similar significant improvements are reported for the regression model developed for large-R jets. The response of the jet p_{T} is improved, with 25% to 35% better relative resolutions. The largest improvements observed are for jets with p_{T} between 400 GeV and 1.2 TeV. The model for the large-R jets is also trained to regress a jet mass correction. The model succeeds in regressing the mass and transverse momentum simultaneously, and the reported improvements for the mass resolution are 26% for jets with a p_{T} of 200 GeV and increase to 33% for jets with p_{T} exceeding 1.2 TeV. The regression model reduces the width of the jet mass distribution for simulated Z boson, Higgs boson, and top-quark jets by around 10%. Comparisons to the large-R jet DNN calibration in Ref. [8], which deliver significant improvements over the nominal calibration, are left to future work where the balance between performance and the level of information included can be studied in samples with similar jet flavour composition.

The regression models studied show significant enhancement in the jet p_{T} and the mass response of simulated jets selected using truth information to include b-hadron decays. Next, studies with collider data using jets after a selection on the flavour-tagging discriminants Refs. [9] and [10] are necessary before such models can be deployed within analyses.

Appendix

A Transformer Network Model Schematics

Schematics of the small-R and large-R jet regression network are shown in Figure 12 and Figure 13, respectively.

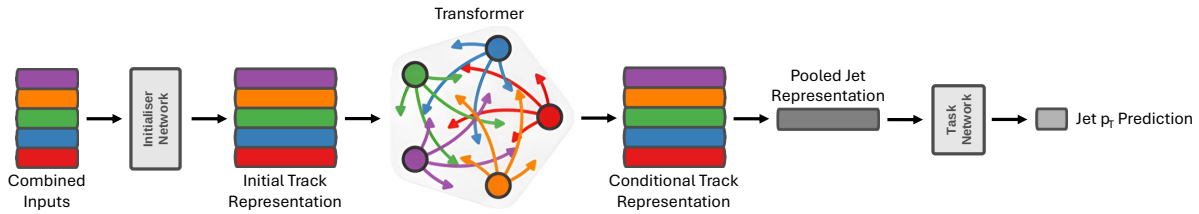

Figure 12: Schematics of the transformer-based small-R jet regression networks described in Section 5.

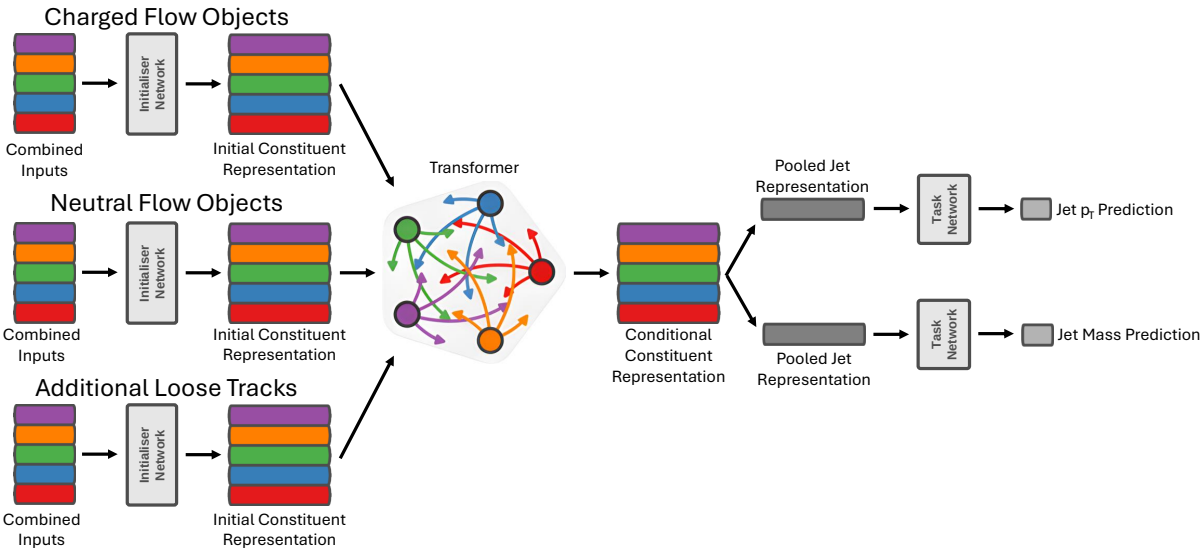

Figure 13: Schematics of the transformer-based large-R jet regression networks described in Section 5.

References

[1] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003 (cit. on pp. 2, 3).

[2] ATLAS Collaboration,
Jet reconstruction and performance using particle flow with the ATLAS Detector,
Eur. Phys. J. C **77** (2017) 466, arXiv: 1703.10485 [hep-ex] (cit. on pp. 2, 7).

[3] ATLAS Collaboration, *Optimisation of large-radius jet reconstruction for the ATLAS detector in* 13 *TeV proton–proton collisions*, Eur. Phys. J. C **81** (2021) 334, arXiv: 2009.04986 [hep-ex] (cit. on pp. 2, 7, 8).

[4] M. Cacciari, G. P. Salam and G. Soyez, *The anti-k_t jet clustering algorithm*, JHEP **04** (2008) 063, arXiv: 0802.1189 [hep-ph] (cit. on pp. 2, 6).

[5] ATLAS Collaboration, *Jet energy scale and resolution measured in proton–proton collisions at* $\sqrt{s}$ = 13 *TeV with the ATLAS detector*, Eur. Phys. J. C **81** (2021) 689,
arXiv: 2007.02645 [hep-ex] (cit. on pp. 2, 13, 14).

[6] ATLAS Collaboration, *In situ calibration of large-radius jet energy and mass in* 13 *TeV proton–proton collisions with the ATLAS detector*, Eur. Phys. J. C **79** (2019) 135,
arXiv: 1807.09477 [hep-ex] (cit. on p. 2).

[7] ATLAS Collaboration, *New techniques for jet calibration with the ATLAS detector*,
Eur. Phys. J. C **83** (2023) 761, arXiv: 2303.17312 [hep-ex] (cit. on pp. 2, 3, 8, 14).

[8] ATLAS Collaboration, *Simultaneous energy and mass calibration of large-radius jets with the ATLAS detector using a deep neural network*, (2023), arXiv: 2311.08885 [hep-ex]
(cit. on pp. 2, 22).

[9] ATLAS Collaboration, *Graph Neural Network Jet Flavour Tagging with the ATLAS Detector*,
ATL-PHYS-PUB-2022-027, 2022, URL: https://cds.cern.ch/record/2811135
(cit. on pp. 2, 9, 22).

[10] ATLAS Collaboration, *Transformer Neural Networks for Identifying Boosted Higgs Bosons decaying into $b\bar{b}$ and $c\bar{c}$ in ATLAS*, ATL-PHYS-PUB-2023-021, 2023,
URL: https://cds.cern.ch/record/2866601 (cit. on pp. 2, 9, 10, 22).

[11] ATLAS Collaboration, *Evidence for the $H \to b\bar{b}$ decay with the ATLAS detector*,
JHEP **12** (2017) 024, arXiv: 1708.03299 [hep-ex] (cit. on pp. 2, 3).

[12] ATLAS Collaboration, *Identification of boosted Higgs bosons decaying into b-quark pairs with the ATLAS detector at* 13 *TeV*, Eur. Phys. J. C **79** (2019) 836, arXiv: 1906.11005 [hep-ex]
(cit. on p. 3).

[13] ATLAS Collaboration, *Constraints on Higgs boson production with large transverse momentum using $H \to b\bar{b}$ decays in the ATLAS detector*, Phys. Rev. D **105** (2022) 092003,
arXiv: 2111.08340 [hep-ex] (cit. on p. 3).

[14] ATLAS Collaboration, *Study of High-Transverse-Momentum Higgs Boson Production in Association with a Vector Boson in the qqbb Final State with the ATLAS Detector*,
Phys. Rev. Lett. **132** (2023) 131802, arXiv: 2312.07605 [hep-ex] (cit. on p. 3).

[15] CMS Collaboration,
A Deep Neural Network for Simultaneous Estimation of b Jet Energy and Resolution,
Comput. Softw. Big Sci. **4** (2020) 10, arXiv: 1912.06046 [hep-ex] (cit. on p. 3).

[16] *A unified approach for jet tagging in Run 3 at $\sqrt{s}$=13.6 TeV in CMS*, (2024),
URL: https://cds.cern.ch/record/2904702 (cit. on p. 3).

[17] ATLAS Collaboration,
Energy scale calibration of b-tagged jets with ATLAS Run 2 data using $t\bar{t}$ lepton+jets events,
ATLAS-CONF-2022-004, 2022, URL: https://cds.cern.ch/record/2803523 (cit. on p. 3).

[18] G. Avoni et al., *The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS*,
JINST **13** (2018) P07017 (cit. on p. 4).

[19] ATLAS Collaboration, *Performance of the ATLAS trigger system in 2015*,
Eur. Phys. J. C **77** (2017) 317, arXiv: 1611.09661 [hep-ex] (cit. on p. 4).

[20] ATLAS Collaboration, *Software and computing for Run 3 of the ATLAS experiment at the LHC*,
(2024), arXiv: 2404.06335 [hep-ex] (cit. on p. 4).

[21] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, Eur. Phys. J. C **70** (2010) 823,
arXiv: 1005.4568 [physics.ins-det] (cit. on p. 4).

[22] S. Agostinelli et al., *GEANT4 – a simulation toolkit*, Nucl. Instrum. Meth. A **506** (2003) 250
(cit. on p. 4).

[23] T. Sjöstrand, S. Mrenna and P. Skands, *A brief introduction to PYTHIA 8.1*,
Comput. Phys. Commun. **178** (2008) 852, arXiv: 0710.3820 [hep-ph] (cit. on pp. 4–6).

[24] NNPDF Collaboration, R. D. Ball et al., *Parton distributions with LHC data*,
Nucl. Phys. B **867** (2013) 244, arXiv: 1207.1303 [hep-ph] (cit. on pp. 4–6).

[25] ATLAS Collaboration, *The Pythia 8 A3 tune description of ATLAS minimum bias and inelastic measurements incorporating the Donnachie–Landshoff diffractive model*,
ATL-PHYS-PUB-2016-017, 2016, URL: https://cds.cern.ch/record/2206965 (cit. on p. 4).

[26] D. J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth. A **462** (2001) 152
(cit. on p. 4).

[27] E. Bothmann et al., *Event generation with Sherpa 2.2*, SciPost Phys. **7** (2019) 034,
arXiv: 1905.09127 [hep-ph] (cit. on pp. 4–6).

[28] K. D. Tran,
Form factor models and branching fractions of B meson decays in the SHERPA simulation,
Presented 2016, TU Dresden, 2016, URL: https://cds.cern.ch/record/2815289
(cit. on p. 4).

[29] P. Nason, *A new method for combining NLO QCD with shower Monte Carlo algorithms*,
JHEP **11** (2004) 040, arXiv: hep-ph/0409146 (cit. on p. 5).

[30] S. Frixione, P. Nason and C. Oleari,
Matching NLO QCD computations with parton shower simulations: the POWHEG method,
JHEP **11** (2007) 070, arXiv: 0709.2092 [hep-ph] (cit. on pp. 5, 6).

[31] S. Alioli, P. Nason, C. Oleari and E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, JHEP **06** (2010) 043,
arXiv: 1002.2581 [hep-ph] (cit. on p. 5).

[32] ATLAS Collaboration, *ATLAS Pythia 8 tunes to 7 TeV data*, ATL-PHYS-PUB-2014-021, 2014,
URL: https://cds.cern.ch/record/1966419 (cit. on pp. 5, 6).

[33] J. Alwall et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, JHEP **07** (2014) 079,
arXiv: 1405.0301 [hep-ph] (cit. on p. 5).

[34] R. Frederix and S. Frixione, *Merging meets matching in MC@NLO*, JHEP **12** (2012) 061, arXiv: 1209.6215 [hep-ph] (cit. on p. 5).

[35] K. Hamilton, P. Nason and G. Zanderighi, *MINLO: multi-scale improved NLO*, JHEP **10** (2012) 155, arXiv: 1206.3572 [hep-ph] (cit. on p. 5).

[36] J. M. Campbell et al., *NLO Higgs boson production plus one and two jets using the POWHEG BOX, MadGraph4 and MCFM*, JHEP **07** (2012) 092, arXiv: 1202.5475 [hep-ph] (cit. on p. 5).

[37] K. Hamilton, P. Nason, C. Oleari and G. Zanderighi, *Merging H/W/Z + 0 and 1 jet at NLO with no merging scale: a path to parton shower + NNLO matching*, JHEP **05** (2013) 082, arXiv: 1212.4504 [hep-ph] (cit. on p. 5).

[38] ATLAS Collaboration, *Measurement of the Z/γ^* boson transverse momentum distribution in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, JHEP **09** (2014) 145, arXiv: 1406.3660 [hep-ex] (cit. on pp. 5, 6).

[39] K. Hamilton, P. Nason and G. Zanderighi, *Finite quark-mass effects in the NNLOPS POWHEG+MiNLO Higgs generator*, JHEP **05** (2015) 140, arXiv: 1501.04637 [hep-ph] (cit. on p. 5).

[40] ATLAS Collaboration, *Multijet simulation for 13 TeV ATLAS Analyses*, ATL-PHYS-PUB-2019-017, 2019, URL: https://cds.cern.ch/record/2672252 (cit. on p. 5).

[41] M. Cacciari, G. P. Salam and G. Soyez, *FastJet user manual*, Eur. Phys. J. C **72** (2012) 1896, arXiv: 1111.6097 [hep-ph] (cit. on pp. 6, 7).

[42] ATLAS Collaboration, *Software Performance of the ATLAS Track Reconstruction for LHC Run 3*, Comput. Softw. Big Sci. **8** (2023) 9, arXiv: 2308.09471 [hep-ex] (cit. on p. 6).

[43] ATLAS Collaboration, *Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1*, Eur. Phys. J. C **77** (2017) 490, arXiv: 1603.02934 [hep-ex] (cit. on p. 6).

[44] ATLAS Collaboration, *Studies of the muon momentum calibration and performance of the ATLAS detector with pp collisions at $\sqrt{s} = 13$ TeV*, Eur. Phys. J. C **83** (2023) 686, arXiv: 2212.07338 [hep-ex] (cit. on p. 6).

[45] ATLAS Collaboration, *Improving jet substructure performance in ATLAS using Track-CaloClusters*, ATL-PHYS-PUB-2017-015, 2017, URL: https://cds.cern.ch/record/2275636 (cit. on p. 7).

[46] ATLAS Collaboration, *Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s} = 13$ TeV*, Eur. Phys. J. C **76** (2016) 292, arXiv: 1603.05598 [hep-ex] (cit. on p. 7).

[47] ATLAS Collaboration, *Performance of pile-up mitigation techniques for jets in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector*, Eur. Phys. J. C **76** (2016) 581, arXiv: 1510.03823 [hep-ex] (cit. on p. 7).

[48] ATLAS Collaboration, *Early Inner Detector Tracking Performance in the 2015 Data at $\sqrt{s} = 13$ TeV*, ATL-PHYS-PUB-2015-051, 2015, URL: https://cds.cern.ch/record/2110140 (cit. on pp. 7–9).

[49] ATLAS Collaboration, *Performance of the ATLAS track reconstruction algorithms in dense environments in LHC Run 2*, Eur. Phys. J. C **77** (2017) 673, arXiv: 1704.07983 [hep-ex] (cit. on pp. 8, 9).

[50] ATLAS Collaboration,
Identification of electrons using a deep neural network in the ATLAS experiment,
ATL-PHYS-PUB-2022-022, 2022, URL: https://cds.cern.ch/record/2803878
(cit. on pp. 7, 12).

[51] A. Sciandra, *Development of a new Soft Muon Tagger for the identification of b-jets in ATLAS*,
PoS **EPS-HEP2017** (2017) 768, ed. by P. Checchia et al. (cit. on pp. 7, 12).

[52] A. J. Larkoski, S. Marzani, G. Soyez and J. Thaler, *Soft Drop*, JHEP **05** (2014) 146,
arXiv: 1402.2657 [hep-ph] (cit. on p. 8).

[53] M. Dasgupta, A. Fregoso, S. Marzani and G. P. Salam,
Towards an understanding of jet substructure, JHEP **09** (2013) 029, arXiv: 1307.0007 [hep-ph]
(cit. on p. 8).

[54] P. Berta, M. Spousta, D. W. Miller and R. Leitner,
Particle-level pileup subtraction for jets and jet shapes, JHEP **06** (2014) 092,
arXiv: 1403.3108 [hep-ex] (cit. on p. 8).

[55] M. Cacciari, G. P. Salam and G. Soyez, *SoftKiller, a particle-level pileup removal method*,
Eur. Phys. J. C **75** (2015) 59, arXiv: 1407.0408 [hep-ph] (cit. on p. 8).

[56] ATLAS Collaboration, *Development of ATLAS Primary Vertex Reconstruction for LHC Run 3*,
ATL-PHYS-PUB-2019-015, 2019, URL: https://cds.cern.ch/record/2670380
(cit. on pp. 8, 9).

[57] M. Cacciari, G. P. Salam and G. Soyez, *The catchment area of jets*, JHEP **2008** (2008) 005
(cit. on p. 8).

[58] ATLAS Collaboration,
Boosted hadronic vector boson and top quark tagging with ATLAS using Run 2 data,
ATL-PHYS-PUB-2020-017, 2020, URL: https://cds.cern.ch/record/2724149 (cit. on p. 8).

[59] ATLAS Collaboration,
Identification of hadronically-decaying top quarks using UFO jets with ATLAS in Run 2,
ATL-PHYS-PUB-2021-028, 2021, URL: https://cds.cern.ch/record/2776782 (cit. on p. 8).

[60] A. Vaswani et al., *Attention is all you need*,
Advances in neural information processing systems **30** (2017) (cit. on p. 9).

[61] M. Zaheer et al., *Deep Sets*, 2018, arXiv: 1703.06114 [cs.LG] (cit. on p. 9).

[62] R. H. R. Hahnloser, R. Sarpeshkar, M. A. Mahowald, R. J. Douglas and H. S. Seung,
Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit,
Nature **405** (2000) 947, URL: https://doi.org/10.1038/35016072 (cit. on p. 9).

[63] K. Jarrett, K. Kavukcuoglu, M. Ranzato and Y. LeCun,
'What is the best multi-stage architecture for object recognition?',
2009 IEEE 12th International Conference on Computer Vision, 2009 2146 (cit. on p. 9).

[64] V. Nair and G. E. Hinton, 'Rectified linear units improve restricted boltzmann machines',
Proceedings of the 27th International Conference on International Conference on Machine Learning, ICML'10, Haifa, Israel: Omnipress, 2010 807, ISBN: 9781605589077 (cit. on p. 9).

[65] S. Shleifer, J. Weston and M. Ott,
NormFormer: Improved Transformer Pretraining with Extra Normalization, 2021,
arXiv: 2110.09456 [cs.CL] (cit. on p. 10).

[66] J. L. Ba, J. R. Kiros and G. E. Hinton, *Layer Normalization*, 2016, arXiv: 1607.06450 [stat.ML] (cit. on p. 10).

[67] D. Misra, *Mish: A Self Regularized Non-Monotonic Activation Function*, 2020, arXiv: 1908.08681 [cs.LG] (cit. on p. 10).

[68] ATLAS Collaboration, *ATLAS measurements of the properties of jets for boosted particle searches*, Phys. Rev. D **86** (2012) 072006, arXiv: 1206.5369 [hep-ex] (cit. on p. 21).

[69] ATLAS Collaboration, *Measurement of the ATLAS Detector Jet Mass Response using Forward Folding with* $80\,fb^{-1}$ *of* $\sqrt{s} = 13$ *TeV pp data*, ATLAS-CONF-2020-022, 2020, URL: https://cds.cern.ch/record/2724442 (cit. on p. 21).

[70] ATLAS Collaboration, *Dependence of the Jet Energy Scale on the Particle Content of Hadronic Jets in the ATLAS Detector Simulation*, ATL-PHYS-PUB-2022-021, 2022, URL: https://cds.cern.ch/record/2808016 (cit. on p. 21).