6th Inter-experiment Machine Learning Workshop - Poster Session
CERN 27 November 2023

Calibration of large-R jets measured with the ATLAS detector using a DNN

Introduction

The energy and mass measurements of jets are crucial tasks for the Large Hadron Collider (LHC) experiments. This paper presents a new calibration method to simultaneously calibrate these quantities for large-radius jets measured with the ATLAS detector using a deep neural network (DNN). To address the specificities of the calibration problem, special loss functions and training procedures are employed, and a complex network architecture, which includes feature-annotation and residual connection-like layers, is used. The DNN-based calibration is compared to the standard numerical approach in an extensive set of tests. The DNN approach is found to perform significantly better in almost all the tests and over most of the covered phase space. In particular, it consistently improves the energy and mass resolutions, with a 30% better energy resolution obtained for $p_T > 500$ GeV.

DNN calibration principle

- Goal: simultaneous calibration of the energy and mass of large-R jets ($R = 1$)
- Method:
  - use one DNN to predict the E and m response ($r_E = \frac{E^{reco}}{E^{true}}$, $r_m = \frac{m^{reco}}{m^{true}}$) for any large-R jet
  - DNN predicts the **mode** of the response distribution:
    - $R_{DNN} = \text{mode}(r = \frac{X_{reco}}{X_{true}})$ as a function of $(\vec{x}_{reco}, \theta)$, for $X = E$ or m and $\vec{x}_{reco}$ = input reconstructed variables, θ = DNN trainable parameters
    - then $X_{calib} = \frac{X_{reco}}{R_{DNN}}$
- Why mode and not individual response ? → Not possible to predict individual responses : to one true jet corresponds a distribution of possible responses.
- Mode and not mean ? → To avoid biases from possible asymmetric distributions, is the most probable value of the distributions.

	Name	Definition
Jet level	E	Energy of the jet in GeV, the *log* of E is taken to reduce the spread of its distribution
	m	Mass of the jet in GeV, the *log* of m is taken to reduce the spread of its distribution
	η	Jet pseudorapidity
Substructure level	groomMRatio	Mass ratio between groomed and ungroomed jets
	Width	$\sum_i p_{Ti}\Delta R(i, jet)/(\sum_i p_{Ti})$ where ΔR is the angular distance (sum over the jet constituents)
	Split12,Split23	Splitting scales at the 1st and 2nd exclusive k_T declusterings
	C2, D2	Energy correlation ratios
	τ_{21}, τ_{32}	N-Subjettiness ratios using WTA axis
	Qw	Smallest invariant mass among the proto-jets pairs of the last 3 steps of a k_T reclustering sequence
Detector level	EMFrac	Energy fraction deposited in the electromagnetic calorimeter
	EM3Frac	Energy fraction deposited in the third layer of the electromagnetic calorimeter
	Tile0Frac	Energy fraction deposited in the 1st layer of the hadronic calorimeter
	EffNConsts	$(\sum_i E_i)^2/(\sum_i E_i^2)$ (sum over the jet constituents)
	NeutralFrac	Energy fraction from neutral constituents
	ChargedPTFrac	p_T fraction from charged constituents
	ChargedMFrac	Mass fraction from charged constituents
Event level	μ	Mean number of interactions per bunch crossing
	NPV	Number of primary vertices per event

Table 1 : Input features

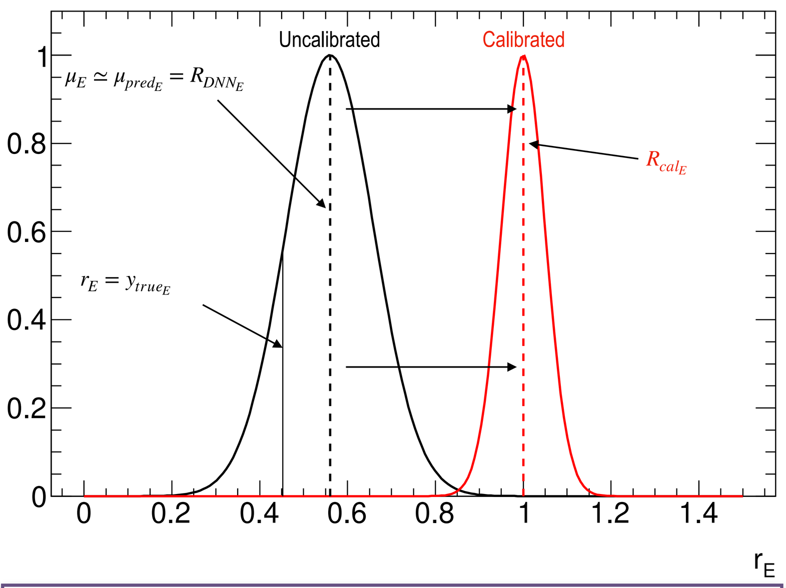
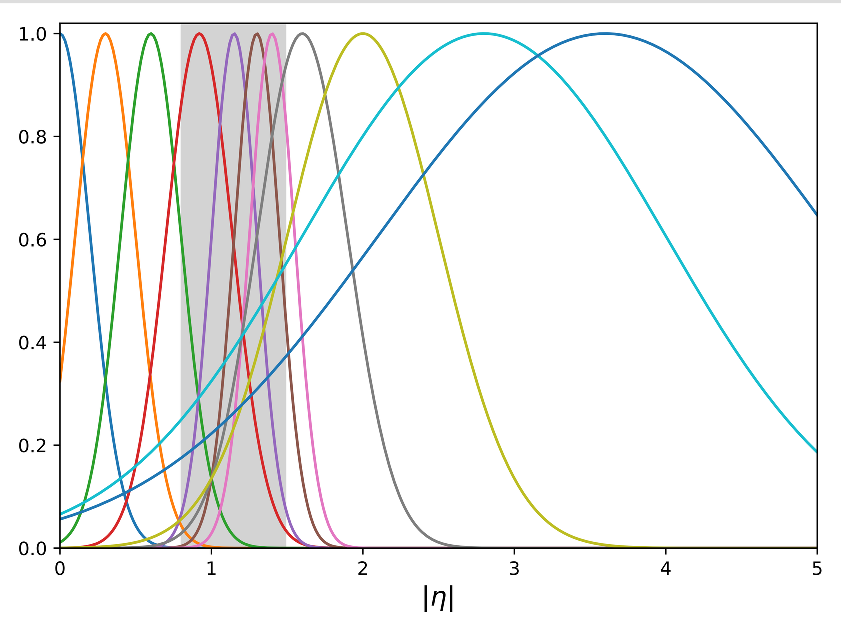


Fig. 1: Idealised uncalibrated energy response distribution and calibration principle

η annotation

- E or m response dependency on η is very complex due to the segmentations of the calorimeters
- Leads to sharp variations of the response in small intervals of η
  → Important difficulty for the DNN to adapt its predictions
- Specific solution: addition of an input preparation step called η-annotation
- Computes extra features based on η to encode the proximity of the jet to the different regions of the detector
  → Helps the DNN to adapt to the complex response dependency on η



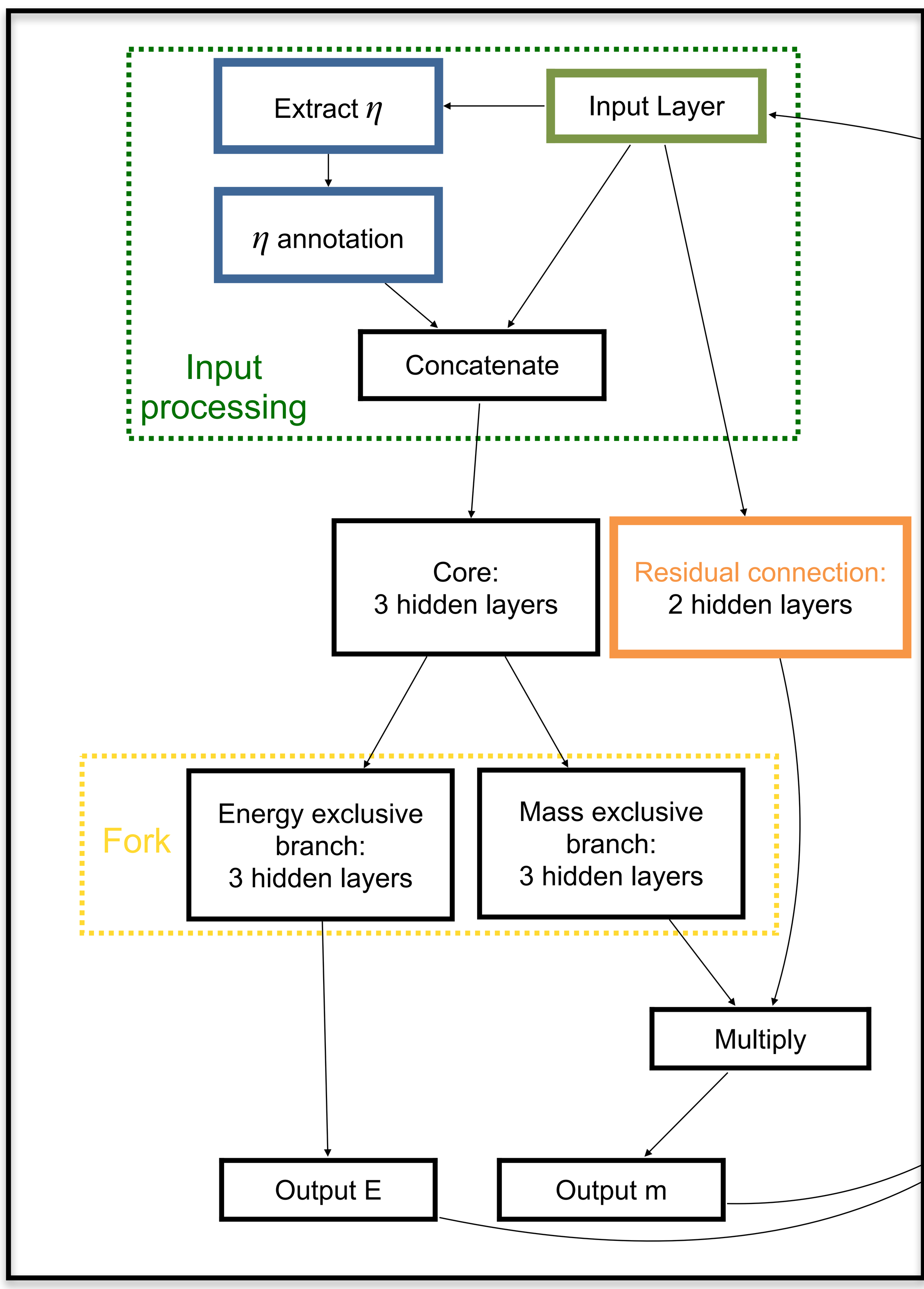


Fig. 2: DNN architecture

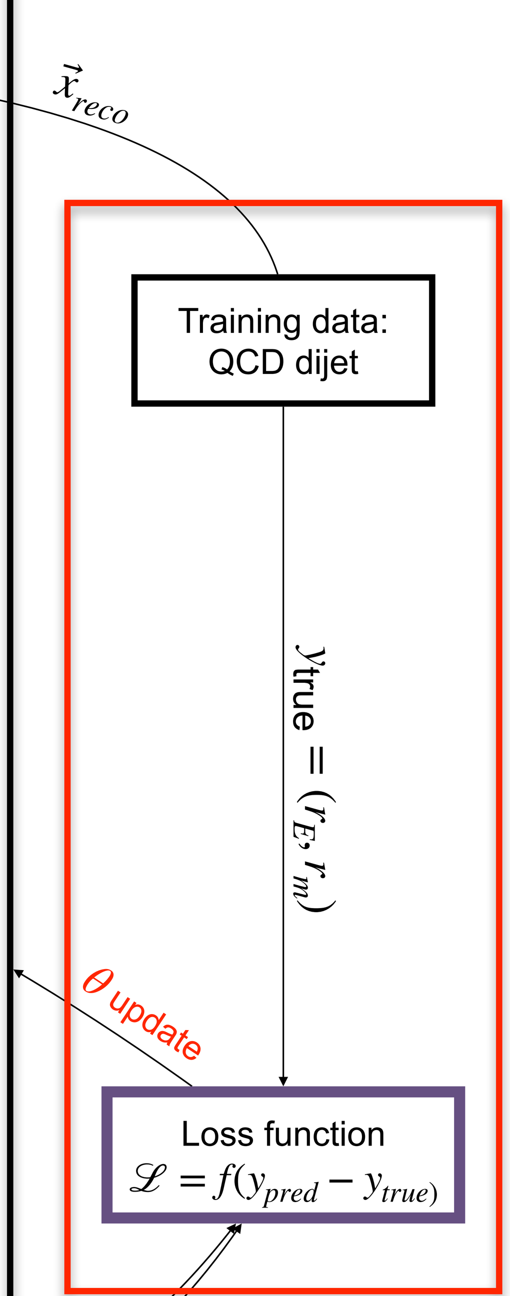


Fig. 3: Training principle

Residual connection

- Additional *multiplicative residual connection* for the mass prediction
- Extra layers linking the input layer directly to the mass output
- Makes the DNN learn which inputs are the most important for the mass calibration
- Only necessary for the mass calibration, energy calibration already shows good performance.

Loss function

- DNN needs to learn the mode of r : mode$(r = y_{true})$
- Assumption: r distribution is close to a gaussian:

$$P(y_{true}, (\mu, \sigma)) \propto \frac{e^{-(y_{true}-\mu)^2/2\sigma^2}}{\sigma}$$

- Estimators of μ (and σ) can be obtained from:

$$LH(\mu, \sigma) = \prod_{y_{true}} P(y_{true}, (\mu, \sigma))$$

- Then a DNN trained with this loss:

$$\mathcal{L}_{MDN} = \log(\sigma_{pred}) + \frac{1}{2}\frac{(y_{true} - \mu_{pred})^2}{\sigma^2_{pred}}$$

Will minimise the log-likelihood and predict $\hat{\mu}$ and $\hat{\sigma}$
→ This is called a Mixture Density Network (MDN)

- DNN then outputs: $y_{pred} = (\mu_{pred}, \sigma_{pred})$ for both E and m. μ_{pred} is then interpreted as R_{DNN}

- In practice r is not perfectly gaussian, training uses:
  - truncated MDN loss to avoid non-gaussianities in the distributions tails
  - asymmetric MDN loss (MDNA) to avoid baises from asymmetric distributions

Performance and validation procedure

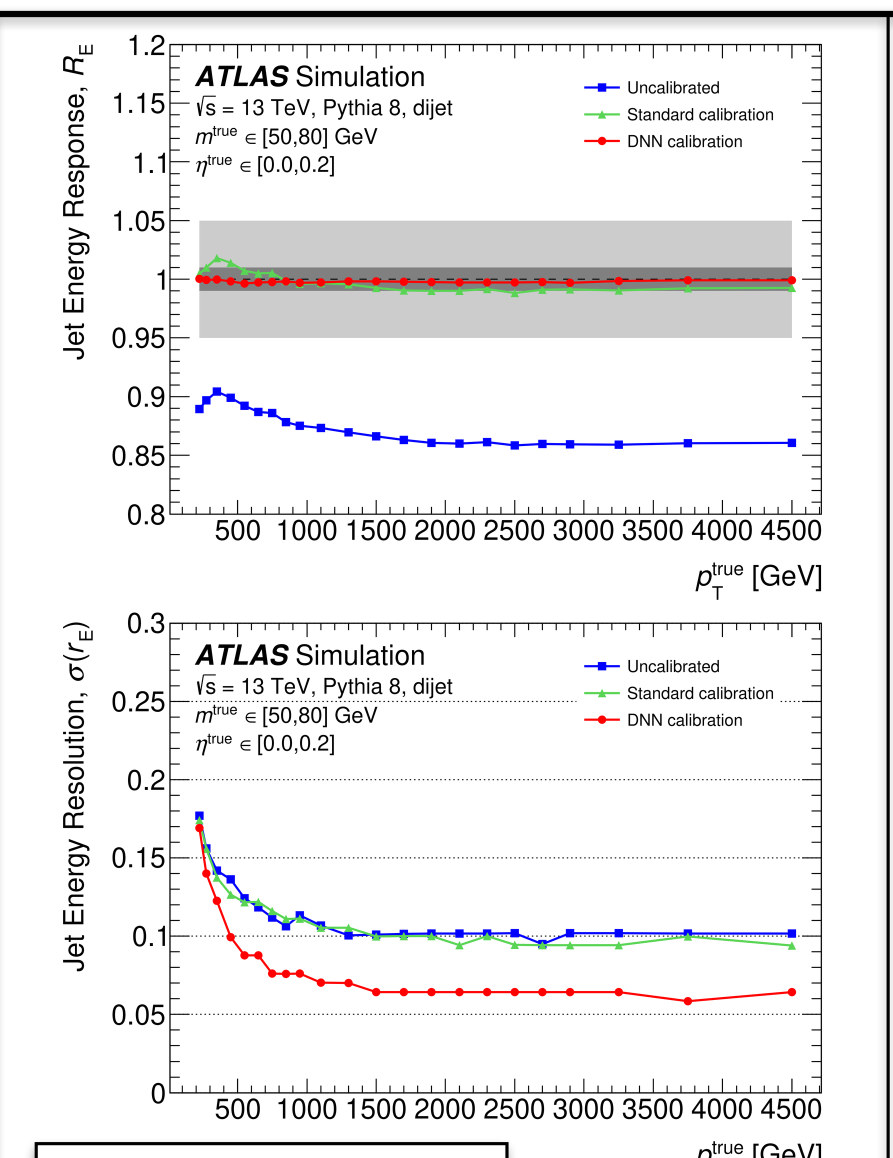
- Two measurements:
  - response closure: in bins of $(p_T^{true}, \eta^{true}, m^{true})$, $R_E = \text{mode}(r_E) \to 1$
  - response resolution: $IQR(68\%)/R_E \to 0$

- Compared with the standard calibration method, the DNN calibration:

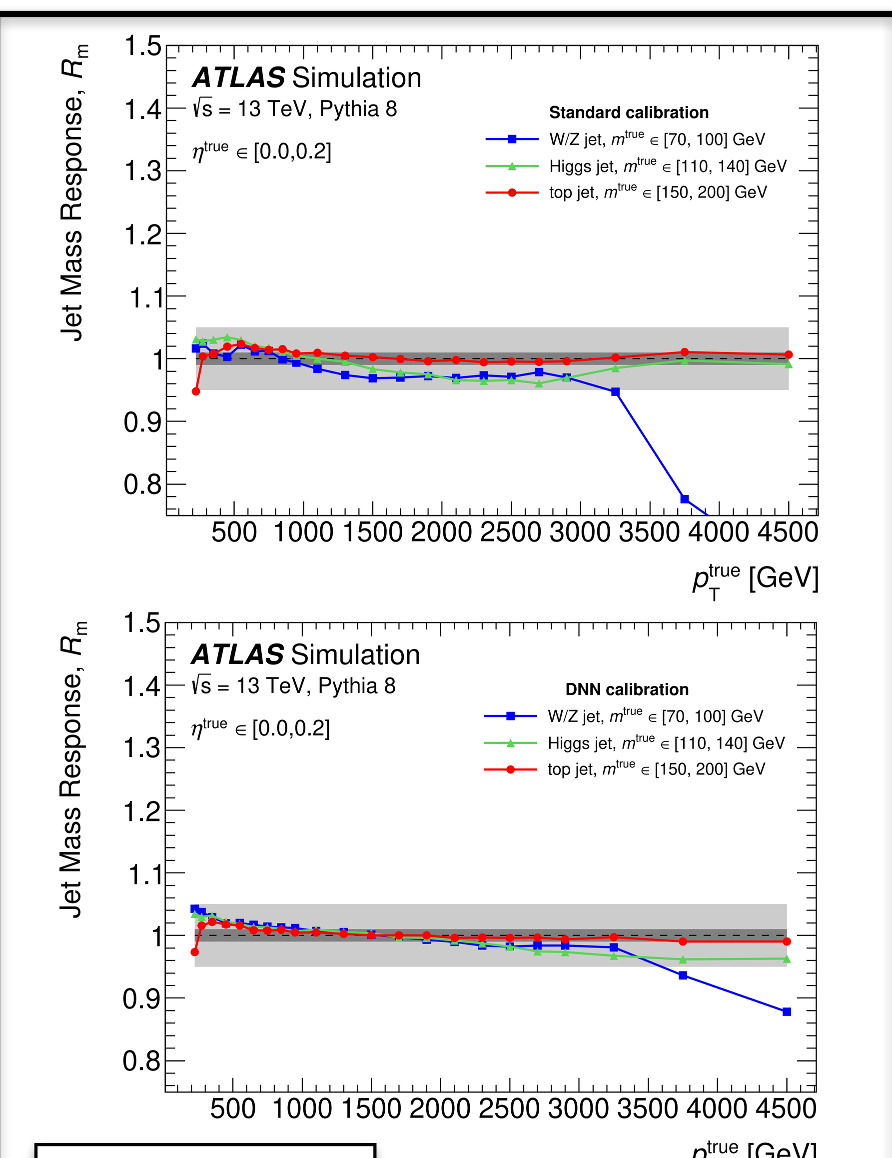
  - performs better in almost all the phase space:
    - in terms of closure and resolution
    - on QCD jets and boosted topologies

  - shows a lower dependency on:
    - pile-up
    - jet flavour → smaller systematic uncertainties
    - jet spectrum → no biases from input distributions
    - MC generators → can be applied to other MC generators and data

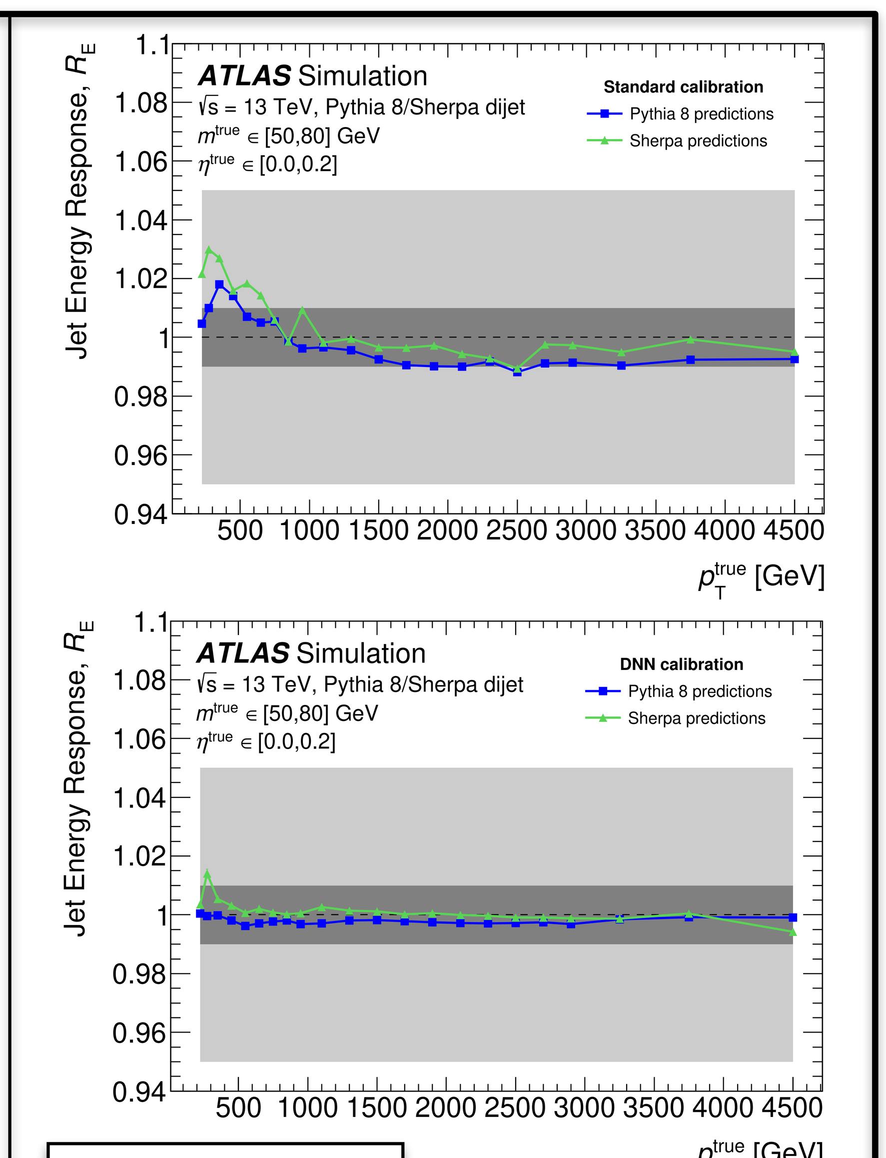
- DNN calibration should replace the actual numerical method for the consolidated recommendations of Run-3



Energy calibration performance



Topology dependency



Generator dependency

Pierre-Antoine Delsart (CNRS, LPSC-IN2P3 Grenoble) on behalf of the ATLAS Collaboration

Main reference: CERN-EP-2023-250, arXiv:2311.08885