

CERN Summer Student Programme - Work Project Report

Energy Resolution Study for Particle Flow Jets at the ATLAS Experiment

Alexander Pedersen Lind

CERN E-mail: alexander.pedersen.lind@cern.ch

University E-mail: lind@nbi.dk

Main Supervisor: Christopher Young

2nd Supervisor: Heather Gray

August 28, 2015

Abstract

This project presents a study on the energy resolution of topo-cluster jets compared to particle flow jets at the ATLAS detector. The data used is pp -events at $\sqrt{s} = 13$ TeV recorded at the ATLAS detector in 2015. Data is compared to dijet 2015 Monte Carlo samples.

Included is a comparison of the relative scale in 2015 data and Monte Carlo samples. We find that data and MC is behaving similarly. The relative scales between the topo-cluster jets and the particle flow jets agree between data and MC by 2% for jets with $p_T > 40$ GeV.

Also included is a comparison of the resolution between topo-cluster jets and particle flow jets in 2015 data. For the MC samples a comparison with the truth jets has also been done. The reconstructed topo-cluster jets has a better resolution than the truth jets for the MC samples. This indicates a bias in our method which might be caused by non-closure in the jet calibration. Several sources of the possible bias are investigated but further work is required to understand it completely.

1 Introduction

The reconstruction of hadronic jets with good resolution is a key ingredient in the ATLAS physics program. For example the jet resolution directly enters the reconstruction of Higgs particles in the channel $H \rightarrow b\bar{b}$.

There are several on-going studies in ATLAS to improve the jet resolution, one of which uses the inner detector tracker to assist in the measurement of low p_T charged particles and the calorimeter to measure high p_T charged particles and neutral particles, in a particle flow algorithm.

1.1 The ATLAS Detector

ATLAS (A Toroidal LHC ApparatuS) is a multi-purpose detector and is one of the seven particle detector experiments at the Large Hadron Collider (LHC) at CERN.

Closest to the collision vertex is the **Inner Detector (ID)**. It spans a pseudorapidity range of $|\eta| < 2.5$ and has full azimuthal coverage. The ID has a length of 6.2 m and a diameter of 2.1 m. The ID is contained in the central solenoid, which provides a nominal magnetic field of 2 T. The magnetic field ensures that the trajectory of charged particles is bent which allows the measurement of the particle momentum (p) and the determination of the charge of the particle.

The ID is surrounded by **electromagnetic (EM) sampling calorimeters** which are in turn surrounded by **hadronic sampling calorimeters**. The calorimeters span a pseudorapidity range of $|\eta| < 4.9$ and has hermetic coverage of the azimuthal plane. Reconstruction of jets is done by energy deposited in the calorimeter system.

The trigger system for the ATLAS detector consists of a hardware-based Level 1 (L1) trigger, as well as a software-based higher-level trigger (HLT). Dijet and multijet events are retained for further analysis using this trigger system.

For a full and detailed explanation of the ATLAS detector, see reference [1].

2 Particle Flow

The particle flow paradigm proposes the use of all information (tracking and calorimeter) in order to achieve the best energy resolution possible. The particle flow algorithm aims at identifying and reconstructing individually each particle arising from the LHC proton-proton collision, by combining the information from all the subdetectors.

The resulting particle-flow event reconstruction leads to an improved performance for the reconstruction of jets and missing transverse energy (MET), and for the identification of electrons, muons and taus. It has the potential to provide improvements on current calorimetric methods for reconstructing jet energies.

2.1 The Particle Flow Algorithm

The particle flow algorithm was optimised using single pion samples, double particle samples and in Pythia8 dijet Monte Carlo samples without pile-up. In this section a brief description of the structure of the algorithm is given.

First tracks are selected where the measurement of the energy of particle by the ID tracker is expected to be superior to that of the calorimeter. These tracks are then matched to a single topological cell cluster in the calorimeter. The expected energy in the calorimeter can then be computed based on the position of the calorimeter cluster and the track momentum. It is relatively common that a single particle will deposit energy in more than one cluster. This results in the need to add additional

clusters to the system to recover energy deposited in other clusters. Finally the expected energy in the calorimeter is subtracted cell-by-cell. If the energy remaining in the system is consistent with the expected resolution of the energy deposited by that particle then the remnant cluster is removed.

The algorithm is illustrated in figure 1.

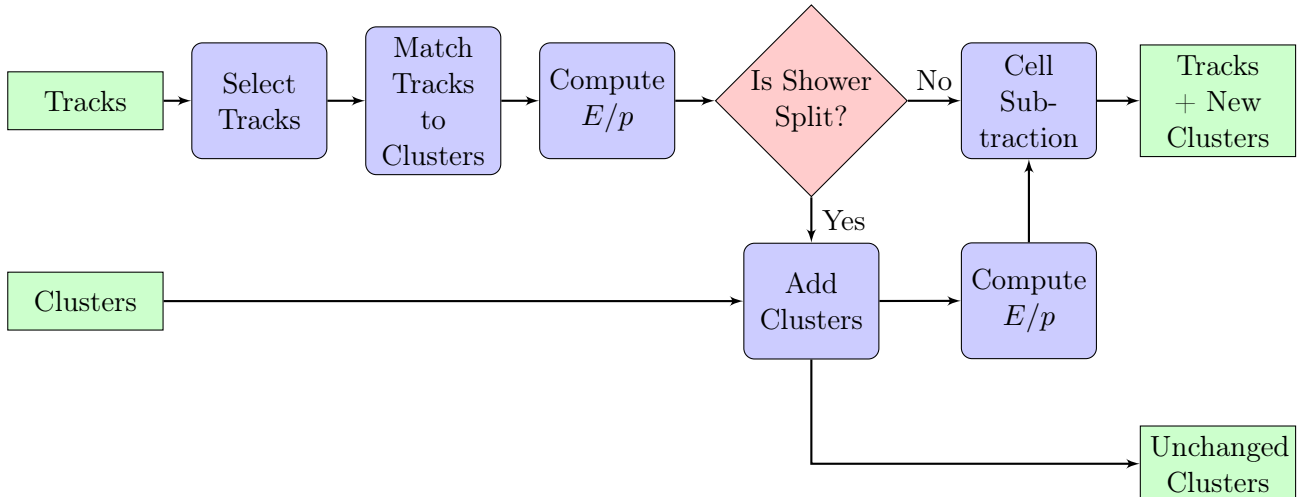


Figure 1: A flow chart describing the particle flow algorithm.

3 Dataset and Monte Carlo Samples

The dataset consists of proton-proton (pp) collisions at a center-of-mass energy of $\sqrt{s} = 13$ TeV, recorded at the ATLAS detector on July 13, 2015 (ATLAS run 271516). All subdetectors were required to be operational and events were rejected if they didn't meet certain data quality criteria. The events that passed these criteria is recorded in a Good Runs List (GRL). The peak luminosity was $1.4 \cdot 10^{33} \text{ cm}^{-2}\text{s}^{-1}$, the peak interactions per crossing was 22.2 and the total luminosity after GRL requirements was 19.72344 pb^{-1} .

The data was compared to Pythia8 [2] dijet Monte Carlo samples interfaced with EvtGen. The samples are divided up into slices depending on the generated jet p_T range. Slices JZ0W, JZ1W, JZ2W and JZ3W have been used for a full p_T range of 0 - 160 GeV. The samples are for $\sqrt{s} = 13$ TeV pp -events and uses the NNPDF23L0 parton distribution functions (PDFs) [3]. The tune used is the ATLAS tune A14 [4]. Pile-up events are overlaid on top of the hard scatter signal using the same generator and process with 50 ns bunch spacing. The events are then fed through a full simulation of the ATLAS detector [5] based on a Geant4 simulation tool-kit [6] using FTFP_BERT for simulation of hadronic showers.

4 Jet Reconstruction and Calibration

The jets used in this study are reconstructed using the anti- k_t algorithm [7] with an angular radius parameter $R = 0.4$. The topological cluster jets used are at the electromagnetic (EM) scale [8], Topo EM jets, as opposed to Topo LC jets, i.e. jets calibrated with the Local Cluster (LC) calibration scheme, where the clusters are calibrated before being reconstructed by the anti- k_t algorithm [9]. The electromagnetic (EM) energy scale correctly measures the energy deposited by the electromagnetic showers in the calorimeter.

The Particle Flow algorithm does not use LC clusters as they appear to make the resolution of low p_T pions worse, it makes the $\frac{E}{p}$ distribution lumpy and non-gaussian and it makes the low p_T Particle

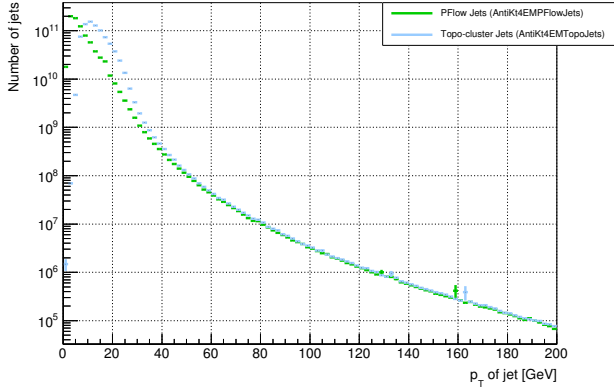


Figure 2: The transverse momentum p_T for both jet collections for data.

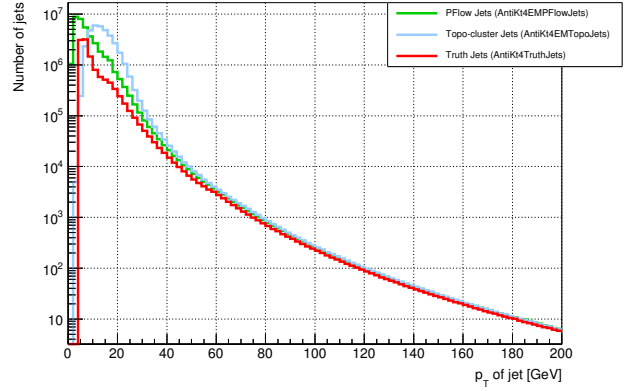


Figure 3: The transverse momentum p_T for both jet collections and the truth jets for MC.

Flow jet resolution significantly worse. As such only clusters at EM scale are used as input. The jets are then built from the remaining clusters in the calorimeter at EM scale after subtraction and the tracks associated with the primary vertex.

Both jet collections, TopoEM and the particle flow (PFlow) jets, have been calibrated with the 2015 recommendations for data and MC15 xAODs and DxAODs [10].

5 Comparison of relative scale between data and MC

To compare the relative scale between the data and the MC samples, we first do a jet selection. For each event we make sure that the Topo EM jets have more than one jet (i.e. a dijet or multijet event). We then find the leading and the sub-leading jet in p_T , i.e. the jets with the highest and second-highest p_T . The p_T criteria for the two found jets are $p_T > 10$ GeV and the pseudorapidity criteria is $|\eta| < 1.0$. The two jets are also checked for jet vertex fraction (JVF) [11], $|JVF| > 0.2$, to ensure that they come from the same interaction vertex. To ensure the two jets are back-to-back we cut at $d\phi > \pi - 0.3$ (see appendix A.1). The two found Topo EM jets are then randomly assigned as A and B . It is then possible to find the PFlow jet nearest to jet A in the $\eta - \phi$ plane given the condition that $dR < 0.4$.

The ratio of transverse momentum $\frac{p_{T,PFlow}}{p_{T,TopoEM}}$ between the PFlow jet near A and jet A will take a Gaussian shape around 1. In figure 4 and 5 this ratio has been plotted for the data samples and in figure 6 and 7 for the MC samples for 8 different bins of the p_T of the jet A (20 - 25 GeV, 25 - 30 GeV, 30 - 40 GeV, 40 - 50 GeV, 50 - 60 GeV, 60 - 80 GeV, 80 - 100 GeV and 100 - 120 GeV).

The histograms have been fitted with a log-likelihood Gaussian fit over the entire range. The width σ of this fit has been extracted and then used as a range $\pm\sigma$ around the mean of the histogram for the final log-likelihood Gaussian fit shown in the figures. This is to ensure that the fit is not biased by the tails of the distributions. The mean value from the fit has been plotted as a function of the p_T range for the data in figure 8 and for the MC in figure 9.

As seen from figure 8 and 9, data and MC is behaving similarly. The relative scales between the topo-cluster jets and the particle flow jets agree between data and MC by 2% for Topo EM jets with $p_T > 40$ GeV.

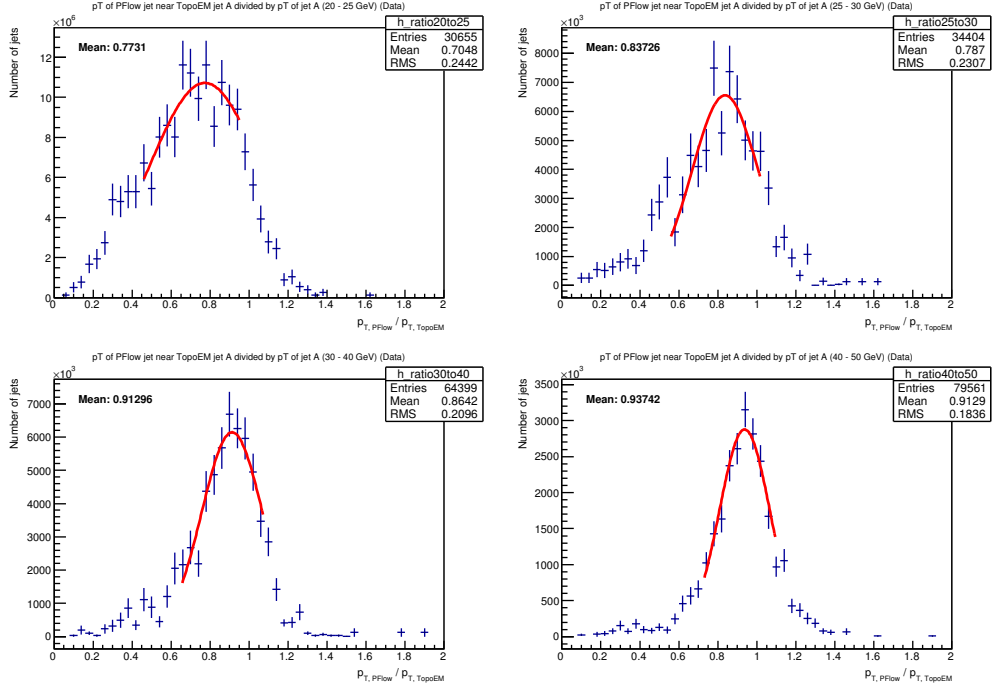


Figure 4: The ratio $p_{T,PFLOW}/p_{T,TopoEM}$ for the four first p_T bins for the data samples. Each bin has been fitted with a log-likelihood Gaussian distribution over the range of $\pm$ root-mean-square around the mean of the histogram.

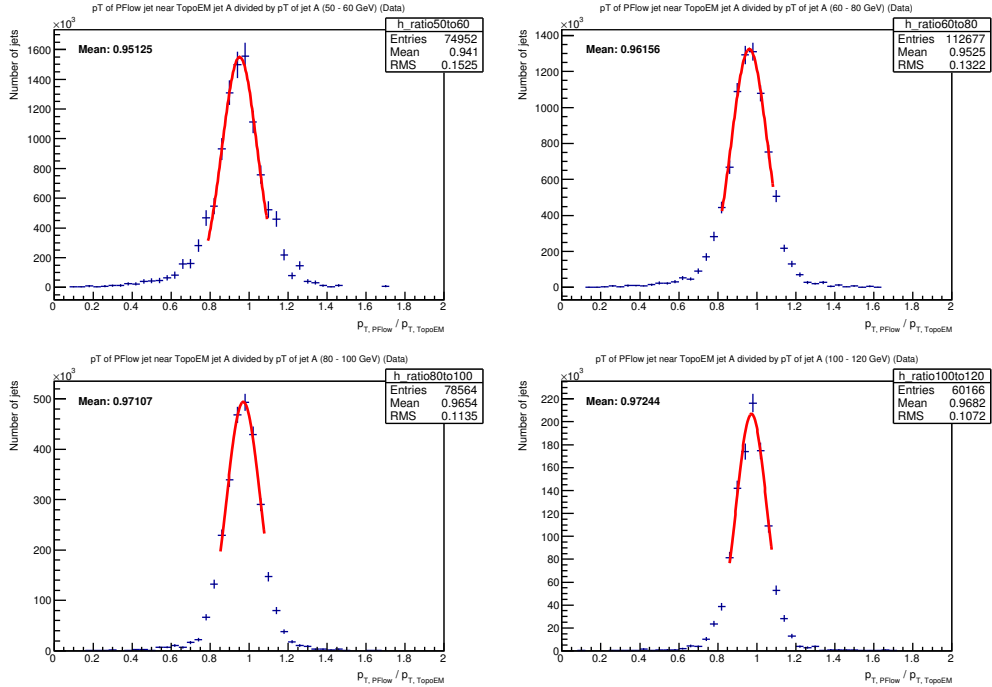


Figure 5: The ratio $p_{T,PFLOW}/p_{T,TopoEM}$ for the four last p_T bins for the data samples. Each bin has been fitted with a log-likelihood Gaussian distribution over the range of $\pm$ root-mean-square around the mean of the histogram.

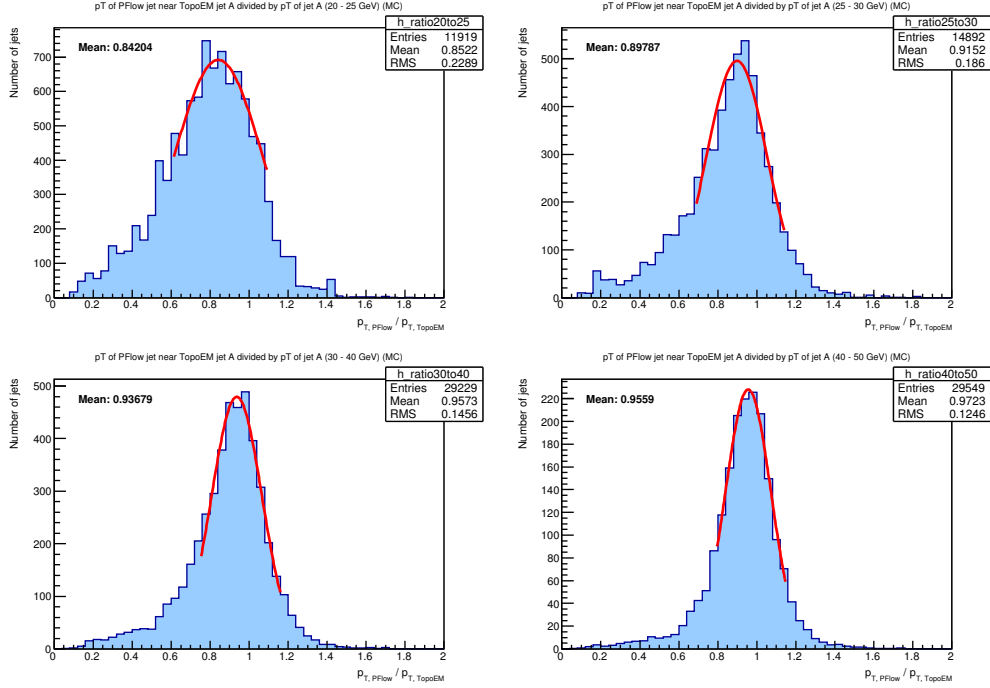


Figure 6: The ratio $p_{T,PFlow}/p_{T,TopoEM}$ for the four first p_T bins for the MC samples. Each bin has been fitted with a log-likelihood Gaussian distribution over the range of $\pm$ root-mean-square around the mean of the histogram.

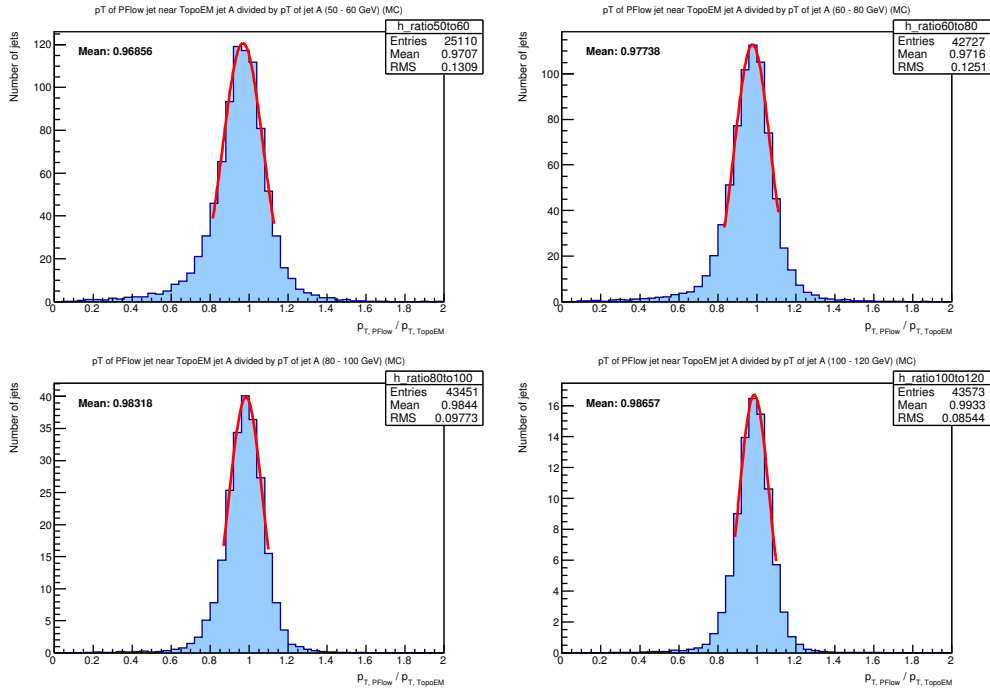


Figure 7: The ratio $p_{T,PFlow}/p_{T,TopoEM}$ for the four last p_T bins for the MC samples. Each bin has been fitted with a log-likelihood Gaussian distribution over the range of $\pm$ root-mean-square around the mean of the histogram.

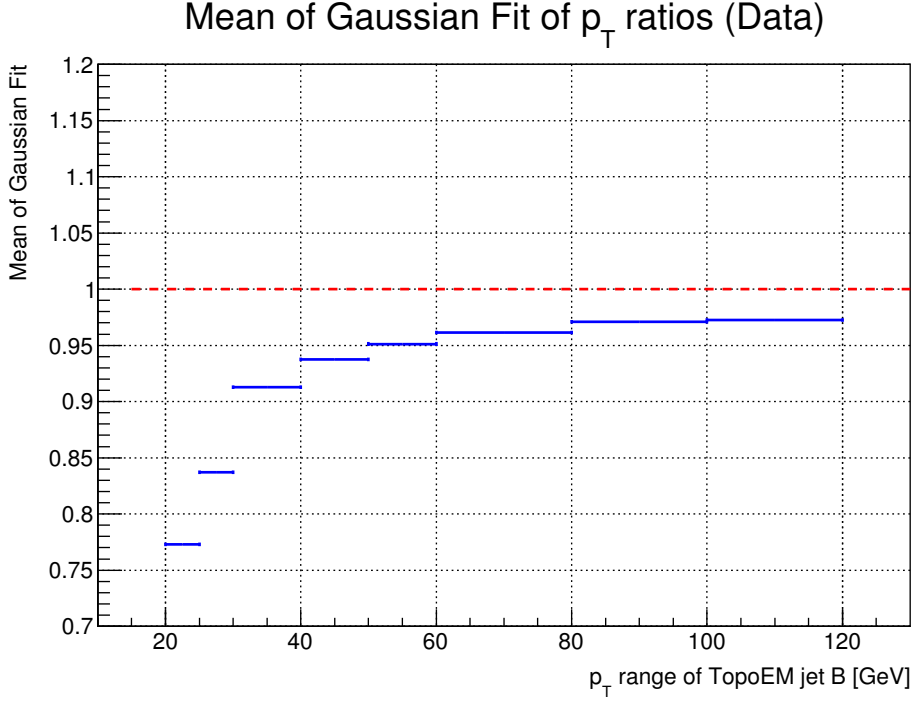


Figure 8: The mean of the Gaussian fits to the ratio between the p_T of the PFlow jet nearest to jet A and the p_T of jet A as a function of the p_T range of jet A . Done for the data samples.

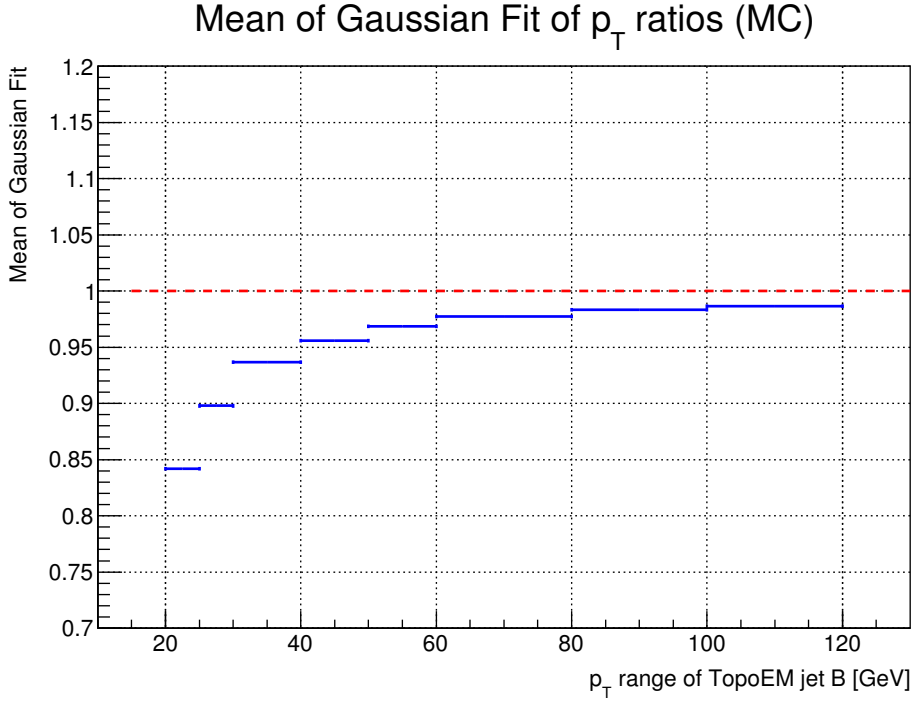


Figure 9: The mean of the Gaussian fits to the ratio between the p_T of the PFlow jet nearest to jet A and the p_T of jet A as a function of the p_T range of jet A . Done for the Monte Carlo (MC) samples.

6 Comparison of resolution between PFlow jets and Topo EM jets

To compare the resolution between particle flow jets and the topo-cluster jets, two back-to-back jets A and B for the particle flow jet collection has been found with the same jet selection criteria as outlined for the Topo EM jet collection in the previous section.

With a jet A and B found for both jet collections it is then possible to plot the dijet balance

$$\frac{p_{T,A} - p_{T,B}}{p_{T,A} + p_{T,B}}$$

for both PFlow jets and Topo EM jets. This is done in the same 8 p_T bins as before but now for the average p_T for A and B , $\frac{p_{T,A} + p_{T,B}}{2}$. These distributions will take a Gaussian shape centered around 0. Fitting the distributions with a log-likelihood Gaussian fit and finding the width of this Gaussian fit will provide us the resolution of the respective jet collection.

The original samples were missing jets with $p_T < 20$ GeV, which is a result of a cut of $p_T > 5$ GeV on both jet collections after the pile-up (PU) correction in the jet calibration and the subsequent Monte-Carlo based jet energy scaling where the scale can be 4 for the Topo EM jets (for central jets). All results presented in this report is with samples where the reconstruction have been rerun with a $p_T > 1$ GeV cut after PU correction instead of the previous $p_T > 5$ GeV cut. This attempted to remove some of the bias in the method, discussed at the end of this section, that could be caused by these missing jets, however, as some bias is still observed this was not the only source of bias.

Additional cuts have been introduced as well. For the MC samples the p_T of the leading reconstructed jet within $|\eta| < 2.8$ and the p_T of the leading truth jet within $|\eta| < 2.8$ have been compared. If $p_{T(\text{reco.})} > 1.5 \times p_{T(\text{truth})}$ then the reconstructed jet deviates too much from the truth jet and the MC event is considered bad and is not included in the analysis. This ensures better statistics for the high p_T bins for the MC samples.

For both data and MC, before the jet selection, a JVF cut for the Topo EM jets at $|\text{JVF}| > 0.25$ (for the `JvtJvfcorr` correction) and a $|\eta| < 2.3$ cut on both jet collections have been introduced.

The dijet balance is shown for both jet collections for the data samples in figure 10 and 11 and is shown for both jet collections and the truth jets for the MC samples in figure 12 and 13. The histograms have been fitted with log-likelihood Gaussian fit over the entire range. The width σ of this fit has been extracted and then used as a range $\pm\sigma$ for the final log-likelihood Gaussian fit shown in the figures.

The width of the final fits provides the resolution. The width of the final Gaussian fit as a function of the p_T range for data is shown in figure 14 and for MC samples in 15.

From figure 12 a quite large bias in the Topo EM jets can be seen and from figure 15 it is noticeable that the reconstructed Topo EM jets has better resolution than the truth jets for the MC samples. This indicates a bias in our method. The shape of the width of this resolution as a function of p_T shown in Figures 14 and 15 is unexpected. The width of the distribution would be expected to grow with decreasing p_T as the jets become softer making them harder to measure and out-of-cone effects larger. The fact that this unexpected shape is seen indicates a bias in the method selecting events where the jets are better balanced than on average. This is further seen when looking at the width in true jets (figure 15). The width of the balance is seen to be wider in truth jets than reconstructed jets indicating that in the reconstruction we are introducing a bias. It was checked that the pile-up jet suppression (JVF) cuts, as well as the sub-leading jet p_T cuts were not the source of the bias. As discussed earlier it was also checked that the usual threshold cuts in the central reconstruction were not a source of the bias. At the time of writing the source of the bias is still not fully understood and further work is required to remove the bias to allow the measurement to extend down to low p_T . One possible source is the non-closure seen in the jet energy scaling (JES) as a result of the functional forms used in it's derivation and the slight non-gaussian shape of the response functions.

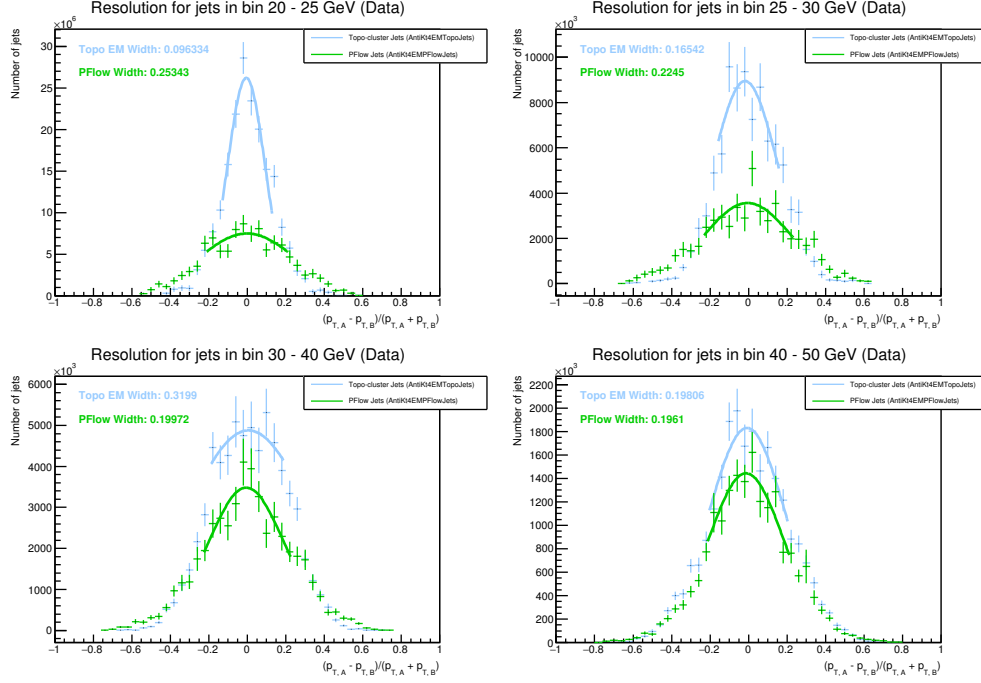


Figure 10: The dijet balance $\frac{p_{T,A} - p_{T,B}}{p_{T,A} + p_{T,B}}$ plotted for both jet collections in the four lowest p_T bins for the data samples. The final log-likelihood Gaussian fit for each bin and jet collection is shown along with the width of the fit.

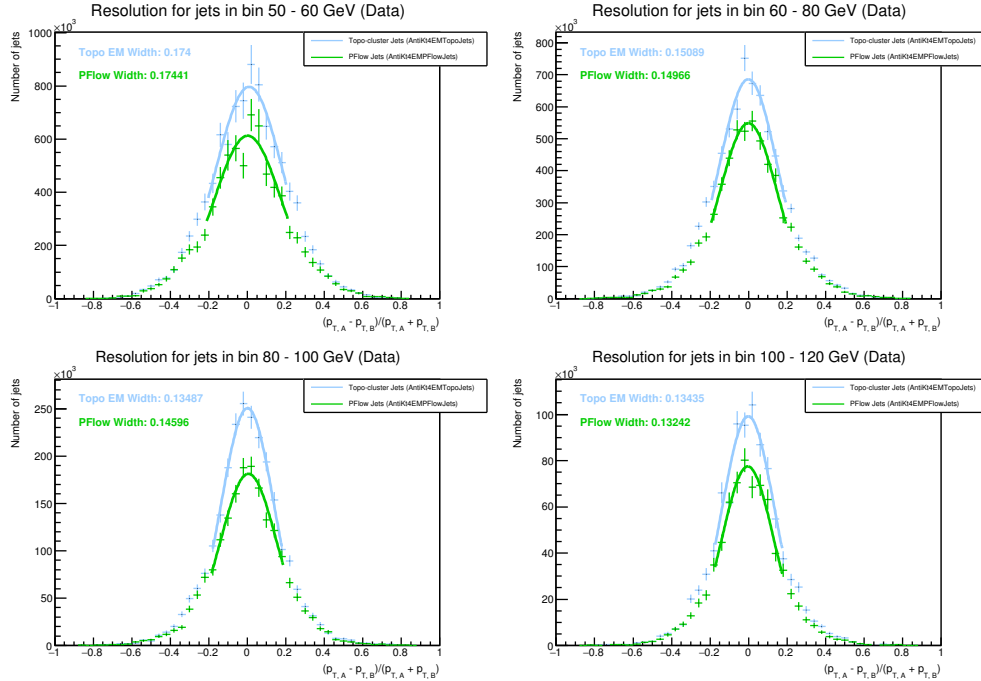


Figure 11: The dijet balance $\frac{p_{T,A} - p_{T,B}}{p_{T,A} + p_{T,B}}$ plotted for both jet collections in the four highest p_T bins for the data samples. The final log-likelihood Gaussian fit for each bin and jet collection is shown along with the width of the fit.

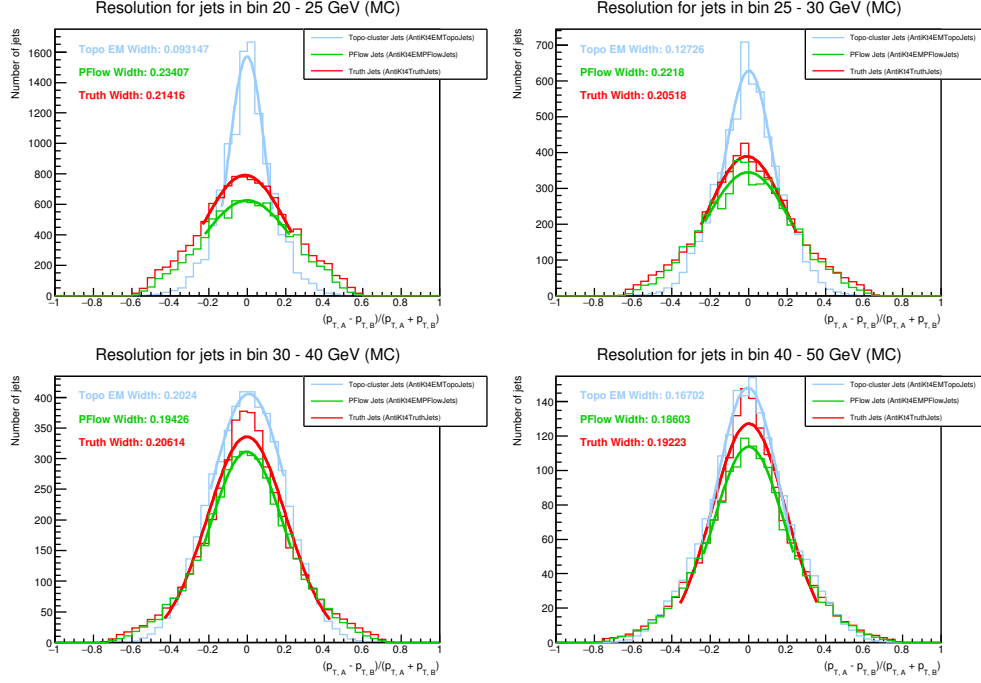


Figure 12: The dijet balance $\frac{p_{T,A} - p_{T,B}}{p_{T,A} + p_{T,B}}$ plotted for both jet collections and truth jets in the four lowest p_T bins for the MC samples. The final log-likelihood Gaussian fit for each bin and jet collection is shown along with the width of the fit.

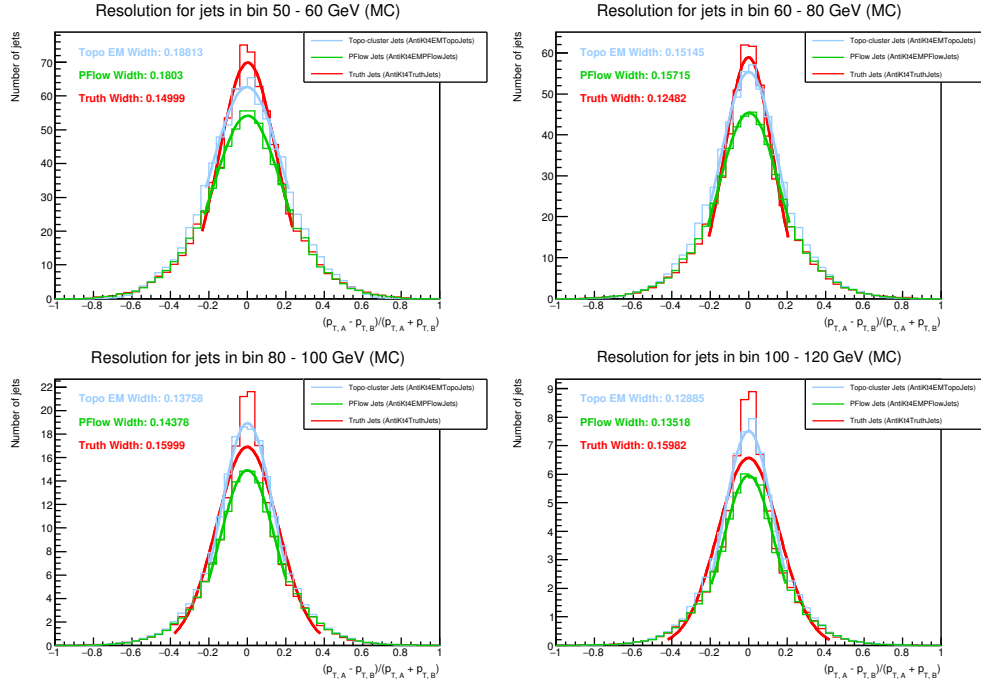


Figure 13: The dijet balance $\frac{p_{T,A} - p_{T,B}}{p_{T,A} + p_{T,B}}$ plotted for both jet collections and truth jets in the four highest p_T bins for the MC samples. The final log-likelihood Gaussian fit for each bin and jet collection is shown along with the width of the fit.

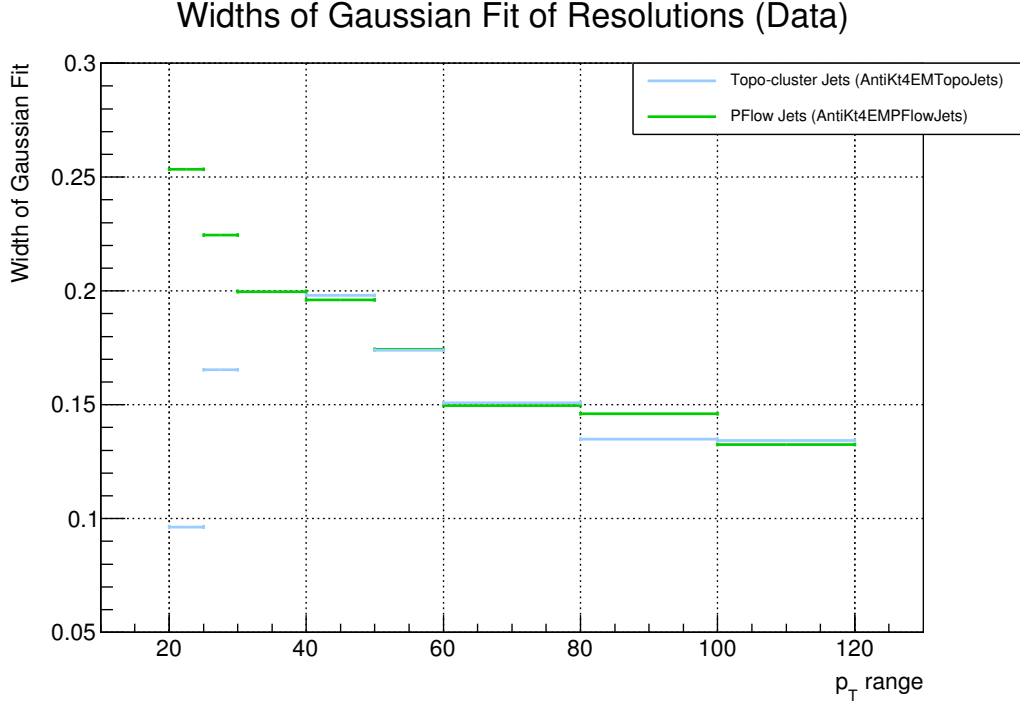


Figure 14: The width of the final Gaussian fits to the dijet balance as a function of the p_T range of the average p_T of the jet A and jet B . Done for both jet collections in data.

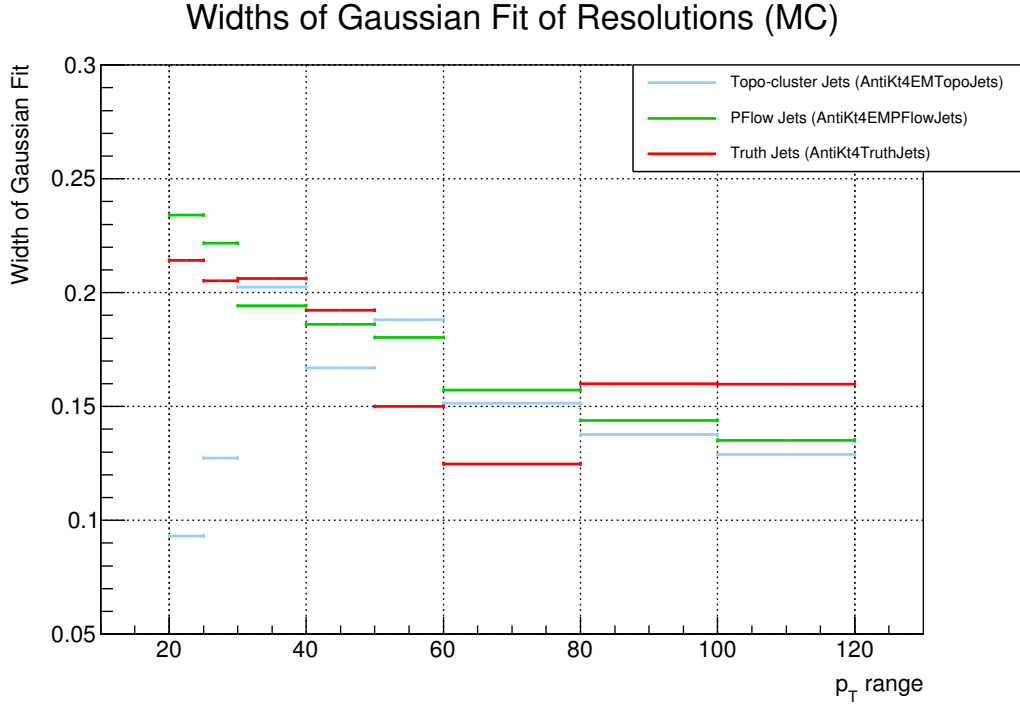


Figure 15: The width of the final Gaussian fits to the dijet balance as a function of the p_T range of the average p_T of the jet A and jet B . Done for both jet collections and for the truth jets in MC.

7 True resolution

Topo EM jets have been matched with the truth jets in the MC samples. They have been matched given that they both fulfill $|\eta| < 1.0$ and that they are within $dR < 0.3$ of each other. An additional requirement that no other truth jet with $p_T > 10$ GeV is within $dR < 1.0$ of the matched reconstructed Topo EM jet.

The p_T ratio $\frac{p_T(\text{reco.})}{p_T(\text{truth})}$ have been plotted in the same 8 p_T bins of the denominator. The plots are shown in figure 16 and 17. The histograms have been fitted with log-likelihood Gaussian fits over the entire range. The width σ of these fits have then been used as the range $\pm\sigma$ around the mean of the histogram for a final log-likelihood Gaussian fit as shown in the figures. The widths of this final Gaussian fit provides the true resolution and is shown in figure 18.

Using the true resolution can give a closure of method. The true resolution σ_{true} for a given p_T should be:

$$\sigma_{\text{true}} = \frac{1}{\sqrt{2}} \sqrt{\sigma_{\text{TopoEM,MC}}^2 - \sigma_{\text{truth}}^2} \quad (1)$$

where $\sigma_{\text{TopoEM,MC}}$ is the Topo EM jet resolution for the MC samples while σ_{truth} is the truth jet resolution for the MC both from figure 15. For the bins where (1) is the case we can claim that our method works and apply it to data.

However due to the bias as mentioned in the previous section, the truth jet resolution is higher (worse) than the Topo EM jet resolution for most p_T bins meaning that the right-hand side of eq. (1) takes on a complex value. The two bins where this is not the case is bin 50 - 60 GeV and bin 60 - 80 GeV. For these bins the right-hand side is 0.08 and 0.06 respectively while the left-hand side is 0.11 and 0.10.

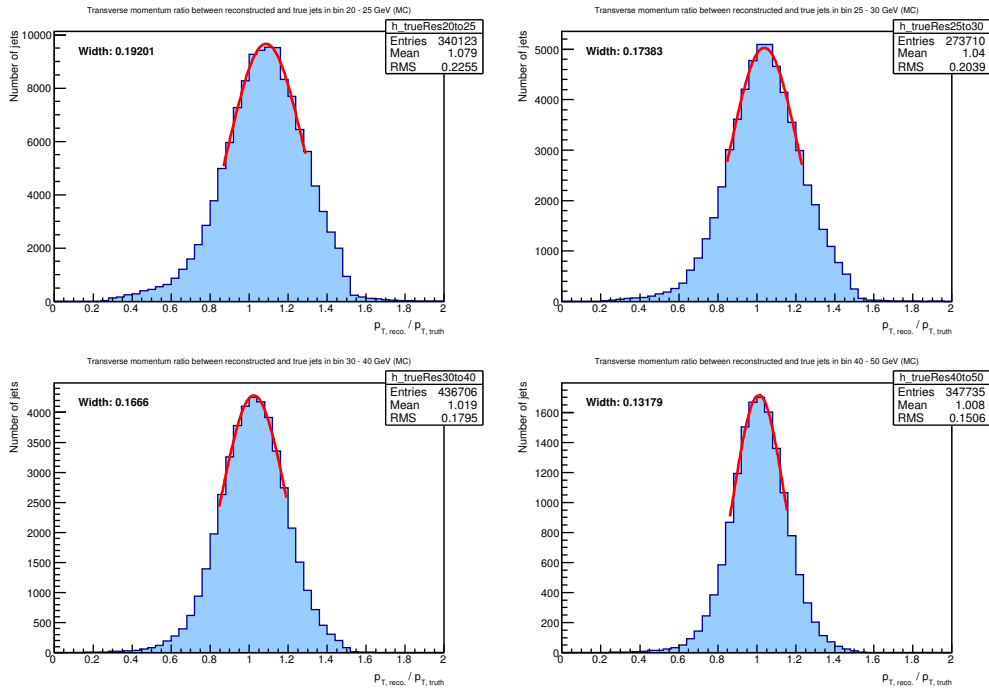


Figure 16: The ratio $\frac{p_T(\text{reco.})}{p_T(\text{truth})}$ for the four lowest p_T bins of the denominator.

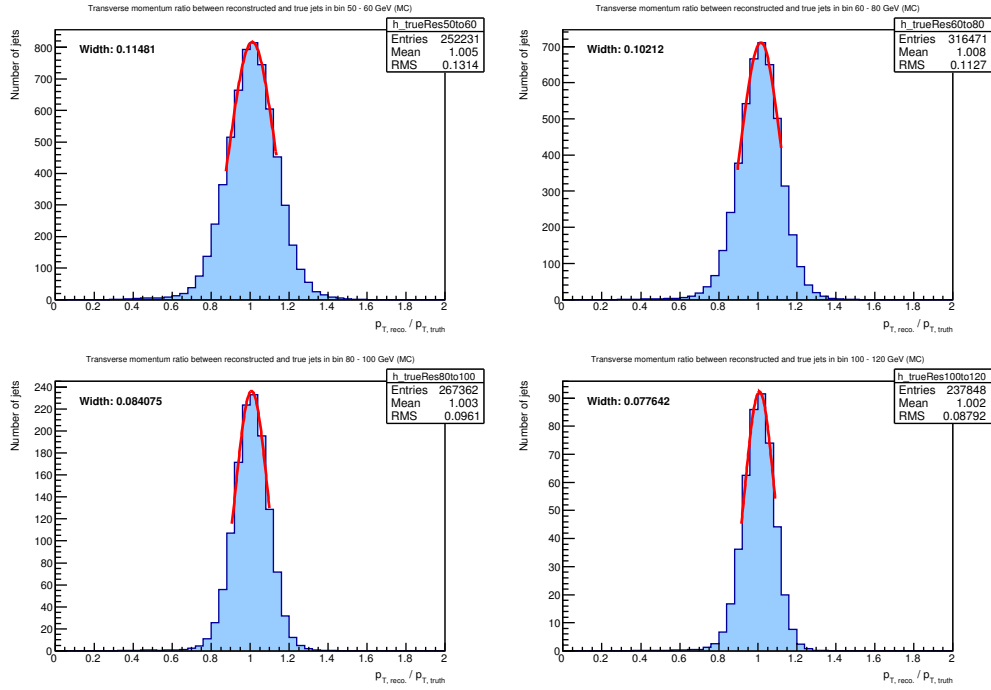


Figure 17: The ratio $\frac{p_T(\text{reco.})}{p_T(\text{truth})}$ for the four highest p_T bins of the denominator.



Figure 18: The widths of the final Gaussian fits to the p_T ratios $\frac{p_T(\text{reco.})}{p_T(\text{truth})}$ as a function of the p_T range of the denominator.

8 Conclusion and Outlook

The data and the MC is behaving similarly. The relative scales between the topo-cluster jets and the particle flow jets agree between data and MC by 2% for Topo EM jets with $p_T > 40$ GeV. But there is a bias that we are trying to remove. There is quite a large bias in the method of measuring the Topo EM jet resolution. The reconstructed jets can't be better than truth jets. This bias can be caused by the non-closure in the jet calibration.

This bias needs to be removed and has not yet been removed as of the writing of this report. The PFlow jets appears to show less bias than the Topo EM jets which could give a clue to where the bias comes from.

References

- [1] ATLAS Collaboration, *The ATLAS experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003.
- [2] T. Sjostrand, S. Mrenna and P. Z. Skands, *A Brief Introduction to PYTHIA 8.1*, Comput.Phys.Commun. 178 (2008) 852–867, arXiv: 0710.3820 [hep-ph].
- [3] J. Pumplin et al., *New generation of parton distributions with uncertainties from global QCD analysis*, JHEP 0207 (2002) 012, arXiv: hep-ph/0201195 [hep-ph].
- [4] ATLAS Collaboration, *ATLAS Run 1 Pythia8 tunes*, ATL-PHYS-PUB-2014-021 (2014), url: <http://cds.cern.ch/record/1966419>.
- [5] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, Eur.Phys.J. C70 (2010) 823–874, arXiv: 1005.4568 [physics.ins-det].
- [6] S. Agostinelli et al., *GEANT4: A Simulation toolkit*, Nucl.Instrum.Meth. A506 (2003) 250–303.
- [7] M. Cacciari, G. P. Salam and G. Soyez, *The Anti- $k(t)$ jet clustering algorithm*, JHEP 0804 (2008) 063, arXiv: 0802.1189 [hep-ph].
- [8] ATLAS Collaboration, *Jet energy scale and its systematic uncertainty in proton-proton collisions at $\sqrt{s}=7$ TeV with ATLAS 2011 data*, ATLAS-CONF-2013-004 (2013), url: <https://cds.cern.ch/record/1509552>.
- [9] ATLAS Collaboration, *Jet energy measurement with the ATLAS detector in proton-proton collisions at $\sqrt{s}=7$ TeV*, Eur.Phys.J. C73.3 (2013) 2304, arXiv: 1112.6426 [hep-ex].
- [10] ATLAS Collaboration, *Monte Carlo Calibration and Combination of In-situ Measurements of Jets in ATLAS*, ATL-COM-PHYS-2015-071 (2015), url: <https://cds.cern.ch/record/1987296>.
- [11] ATLAS Collaboration, *Tagging and suppression of pileup jets with the ATLAS detector*, ATLAS-CONF-2014-018 (2014), url: <https://cds.cern.ch/record/1700870>.

A Appendix

A.1 Cut on $d\phi$ in jet selection

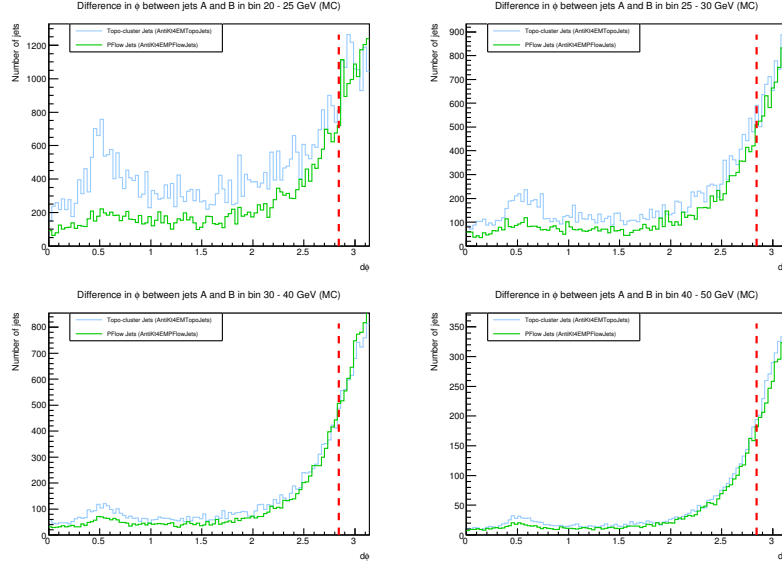


Figure 19: The difference in the azimuthal angle ϕ between Topo EM jets A and B for MC for the four lowest p_T bins. The red stippled lines indicates the cut at $\pi - 0.3$.

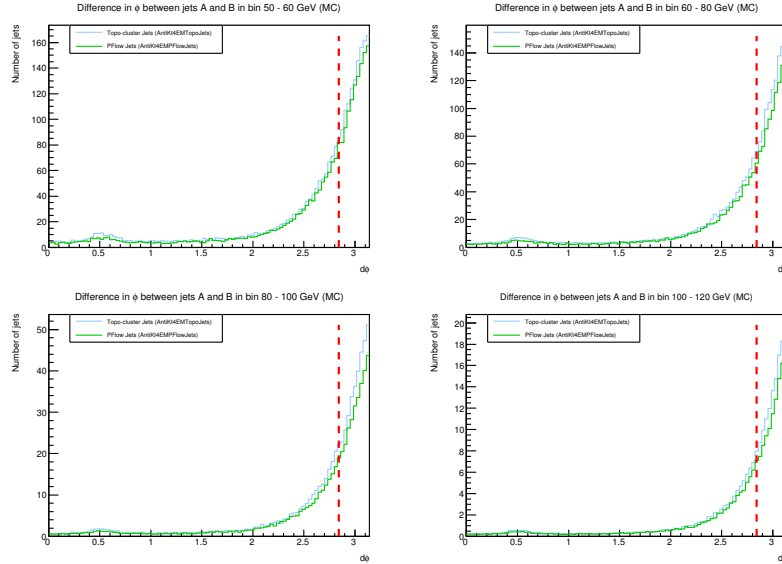


Figure 20: The difference in the azimuthal angle ϕ between Topo EM jets A and B for MC for the four highest p_T bins. The red stippled lines indicates the cut at $\pi - 0.3$.

From the figures above, the cut at $d\phi > \pi - 0.3$ seems reasonable for both jet collections.