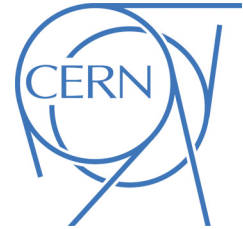**ATLAS NOTE**

ATLAS-CONF-2013-083

August 9, 2013

Pile-up subtraction and suppression for jets in ATLAS

The ATLAS Collaboration

Abstract

An improved pile-up correction for jets in ATLAS, based on the concept of the jet area, is described and evaluated. The method corrects the jet transverse momentum and inherently accommodates jet-by-jet variations in pile-up sensitivity as well as event-by-event fluctuations in pile-up activity. The parameters for the method are optimized in simulation, and its performance is validated using in-situ methods in 2012 data, from which systematic uncertainties are also estimated. The resulting fractional uncertainty in the jet p_T scale is typically less than 2% for jets with $p_T > 50$ GeV across the full coverage of the detector, with significantly lower uncertainties for central jets. A track-based method for the suppression of pile-up jets is also presented, using the jet vertex fraction (JVF). The dependence of JVF on jet kinematics, jet flavor, and pile-up is discussed, and the performance of JVF as a discriminating variable is assessed in both 2012 data and simulation. Finally, a method to address mis-modeling in JVF is presented, with recommended systematic variations in specific JVF cut values.

1 Introduction

Jets are the dominant final state objects of high-energy proton-proton (pp) interactions at the Large Hadron Collider (LHC) at CERN. They are key ingredients for many physics measurements and for searches for new phenomena. Jets are observed as groups of topologically-related energy deposits in the ATLAS [1] calorimeters, most of which are associated with tracks of charged particles as measured in the inner detector. They are reconstructed with the anti-k_t [2] or Cambridge/Aachen [3, 4] algorithm and calibrated using a combination of methods based on Monte Carlo (MC) simulation and data-driven techniques. All relevant details of the jet reconstruction in ATLAS can be found in Ref. [5].

In 2012, ATLAS collected signals from pp collisions at a center-of-mass energy of $\sqrt{s} = 8$ TeV, with total statistics corresponding to approximately 20 fb^{-1} of integrated luminosity. Such a large data set is essential for analyses searching for evidence of new phenomena, as well as for precision measurements of the Standard Model. However, with increased luminosity the beam conditions in 2012 were more challenging than in 2011. The mean number of inelastic pp interactions per bunch crossing, $\langle \mu \rangle$, is related to the instantaneous luminosity, L:

$$\langle \mu \rangle = \frac{L \times \sigma_{\text{inel.}}}{N_{\text{bunch}} \times f_{\text{LHC}}}, \qquad (1)$$

where $\sigma_{\text{inel.}}$ is the total inelastic pp cross-section and $N_{\text{bunch}} \times f_{\text{LHC}}$ is the average frequency of bunch crossings in the LHC [6]. The instantaneous luminosity in 2012 reached values as high as 7.7×10^{33} cm^{-2} s^{-1}, and the average pile-up activity in 2012 was $\langle \mu \rangle \approx 20.7$ interactions per bunch crossing. The presence of additional interactions in the same bunch crossing as the triggered event (*in-time pile-up*) produced additional signals in the ATLAS calorimeters, while the LHC bunch spacing of 50 ns allowed for further signal modulation from multiple interactions in surrounding bunch crossings to which the calorimeter is sensitive (*out-of-time pile-up*).

Reconstruction and calibration of jets from calorimeter signals are particularly sensitive to the effects of pile-up, due mainly to the fact that the particles associated with a jet tend to extend across a wide area of the detector. The larger the area, the more likely it becomes that particles from pile-up interactions overlap with the jet and affect the reconstructed jet kinematics. Furthermore, pile-up contributes additional jets to any event, both genuine QCD particle jets from individual additional interactions and stochastic fluctuations of soft activity from several interactions.

The first part of this note describes the implementation in ATLAS of a method to directly estimate event-by-event pile-up activity and jet-by-jet pile-up sensitivity, originally proposed in Ref. [7]. These estimates allow for an improved pile-up subtraction technique, in which the four momentum of the jet is corrected for event-by-event fluctuations in pile-up, and where jet-by-jet variations in pile-up sensitivity are automatically accommodated. The performance of this pile-up correction is assessed, in comparison to a previous pile-up correction based on the number of reconstructed primary vertices and the instantaneous luminosity [8]. As before, the new correction is the first of several steps in the standard jet calibration chain in ATLAS.

The second part of this note describes the use of tracks to associate jets to the hard-scatter interaction. By matching tracks to jets, one obtains a measure of the fraction of the jet energy associated with a particular primary vertex: the jet vertex fraction (JVF). JVF is the fraction of summed track p_T for all tracks matched to a given jet and associated with the primary vertex, relative to the total summed p_T for all tracks matched to the jet (see Section 7). A cut on JVF can be used to reject spurious calorimeter jets resulting from local fluctuations in pile-up activity, as well as real QCD jets originating from single pile-up interactions, resulting in improved stability of the reconstructed jet multiplicity against pile-up. The performance of three JVF cut values is assessed, and associated systematic uncertainties are presented based on comparisons between data and Monte Carlo simulation.

2 The ATLAS detector

The ATLAS experiment features a multi-purpose detector designed to precisely measure individual particles and particle jets produced in the pp collisions at the LHC. From inside out, the detector consists of an inner tracking detector (ID), surrounded by electromagnetic sampling calorimeters. Those in turn are surrounded by hadronic sampling calorimeters and, behind those, an air-toroid muon spectrometer (MS). The ATLAS calorimeters provide coverage hermetic in azimuth in the pseudorapidity range $|\eta| < 4.9$. A detailed description of the ATLAS detector can be found in Ref. [1].

For the studies presented in this note, in which the effect of pile-up on reconstructed jets is described in detail, it is important to note that different sub-detectors have different sensitivity to pile-up. The ID is affected only by in-time pile-up, due to the fast response of the silicon detectors and the TRT. Also, due to high granularity in the ID, tracking signals from different interactions are generally well separated. The effects of pile-up in the ATLAS calorimeters are fundamentally different: due to coarser granularity, there is significant probability that signals from multiple interactions will overlap, and a longer response time results in sensitivity to out-of-time as well as in-time pile-up.

The ATLAS liquid argon calorimeters in particular are sensitive to the bunch crossing history. In any liquid argon calorimeter cell, the reconstructed energy is sensitive to the pp interactions occurring in approximately 24 preceding and one immediately following bunch crossings (assuming the design value of 25 ns bunch spacing), in addition to pile-up interactions in the current bunch crossing. This is due to the relatively long charge collection time in these calorimeters (typically 400 ns), as compared to the bunch crossing intervals at the LHC (design 25 ns and actually 50 ns in 2012 data). To reduce this sensitivity, a fast, bi-polar shaped signal is used with net zero integral over time.

The signal shapes in the liquid argon calorimeters have been optimized for this purpose, thus on average resulting in cancellation of in-time and out-of-time pile-up in any given calorimeter cell. By design of the shaping amplifier, the most efficient suppression is achieved for 25 ns bunch spacing in the LHC beams. The 2012 beam conditions, with 50 ns bunch spacing, do not allow for full pile-up suppression by signal shaping, in particular in the central calorimeter region because of its high granularity. Additionally, bunch intensities in the LHC are not entirely uniform: there are gaps of various sizes between filled bunch trains, and within bunch trains the individual bunch intensity may vary significantly between adjacent bunch crossings. In 2012, the reconstructed calorimeter cell energies are stabilized against non-uniform bunch intensity by subtraction of a reference pedestal value obtained in each event from the mean cell energy in narrow slices of pseudorapidity.

Pile-up suppression is further limited by large fluctuations in the number of additional interactions from bunch crossing to bunch crossing, and the energy flow patterns of the individual collisions in the time window of sensitivity[1] of approximately 600 ns. Therefore, the shaped signal extracted by digital filtering shows a residual sensitivity to in-time and out-of-time pile-up. This sensitivity is further enhanced by the noise suppression inherent in topological clustering (see Sec. 4.2). All details of the ATLAS liquid argon calorimeter readout and signal processing can be found in Ref. [9].

3 Simulation

For the performance studies of the jet area based correction in simulated data and the derivation of the residual correction, simulated dijet events were used. The events were produced with the PYTHIA8 generator [10] (version 8.160) with the CT10 PDF set [11] using the underlying-event tune AU2 CT10 [12]. The dijet event generation was made in slices of leading jet p_T, with the corresponding samples being weighted[2] so as to achieve a uniform distribution in the leading jet p_T. Different minimum jet p_T

---

[1] The shaped pulse has a duration exceeding the charge collection time.
[2] The samples corresponding to the highest and lowest jet p_T slice are unweighted.

cuts were applied to the different jet p_T slices, in order to avoid the divergence of the cross section at low p_T. For the lowest jet p_T slice a cut on the momentum fraction of the incoming partons was used instead: $x_{1,2} > 0.01$.

The Z-boson samples used for the validation studies were produced with the POWHEG generator [13–15], which provides NLO matrix elements for inclusive Z production. The CT10 [11] PDF set was used in the matrix-element calculation. The modeling of the parton shower, multi-parton interactions and hadronization was provided by Pythia 8.163 with the underlying-event tune AU2 CT10. Pythia was in turn interfaced with PHOTOS [16] for the modeling of QED final-state radiation.

In order to study the event-topology dependence of the jet area based correction, additional simulated event samples were considered. More specifically, the dijet sample was also interfaced to the Herwig++ generator [17] (version 2.5.2), using the CTEQ6L1 [18] PDF set and the UE-EE-3 tune [19]. The Z-boson samples were also produced with the Alpgen generator [20], using the CTEQ6L1 PDF set. This sample was interfaced to Herwig [21, 22] (version 6.5.20) for modeling of the parton shower and hadronization and to the JIMMY generator [23] for modeling of multi-parton interactions, using the AUET2 tune [24]. The topology studies also made use of two γ+jet samples produced either with Pythia8 (version 8.165) with the CTEQ6L1 PDF and the AU2 tune or with Herwig++ (version 2.5.2) with the CTEQ6L1 PDF and the UE-EE-3 tune.

All the MC samples are simulated taking into account the pile-up conditions in the 2012 data, including both the effects of in-time and out-of-time pile-up. The pile-up is modeled by overlaying simulated hits from events with exactly one inelastic (signal) collision per bunch crossing with hits from minimum-bias events that are produced with Pythia 8.160 using the A2M tune [12] and the MSTW2008LO PDF [25]. The number of overlaid minimum-bias events is sampled according to a Poisson distribution with mean $\langle\mu\rangle$. The bunch-filling pattern surrounding the signal interaction is used to modulate this number, so that only filled bunch-crossings are simulated. The number of simulated bunch crossings is different for each detector subsystem, taking into account their different sensitivities to out-of-time pile-up.

The generated particles are propagated through a full simulation of the ATLAS detector [26] using Geant4 [27].

4 Object reconstruction

4.1 Tracks and vertices

Tracks are reconstructed within the full acceptance of the ID $|\eta| < 2.5$. Track reconstruction begins from "hits", namely energy deposits from charged particles in the ID sub-detectors. A sequence of algorithms is used to build tracks from individual hits [28]. The baseline algorithm uses 3-point seeds in the silicon detectors (Pixel and SCT) to form track candidates, which are then extrapolated to include TRT measurements. Tracks are required to have transverse momentum of at least 400 MeV, in addition to further quality criteria relating to impact parameters and numbers of hits in the different ID sub-detectors.

Primary vertices are reconstructed from tracks using an iterative vertex finding algorithm [29]. The vertex reconstruction is organized in two steps: i) the association of reconstructed tracks to vertex candidates, and ii) the vertex fitting, i.e. the calculation of the vertex position and its associated error matrix. Tracks used in vertex reconstruction are required to pass a set of selection criteria designed to minimize the fraction of tracks originating from secondary interactions. Vertex candidates are reconstructed starting from seeds from the reconstructed tracks' z-position at the beam line and fitting the nearby tracks. Tracks displaced by more than 7σ from the vertex are used to seed a new vertex and the procedure is repeated until no additional vertices can be found. The primary vertex with the highest Σp_{T}^2 of constituent tracks is chosen to correspond to the hard-scatter vertex.

The total number of reconstructed primary vertices with 2 or more tracks, N_{PV}, provides a measure of in-time pile-up activity in each event. As shown in Figure 1, N_{PV} is strongly correlated with the mean number of interactions per bunch crossing, $\langle\mu\rangle$. However, $\langle\mu\rangle$ is obtained from the measurement of instantaneous luminosity and is therefore sensitive to both in-time and out-of-time pile-up.[3] If N_{PV} is kept fixed, thus constraining the amount of in-time pile-up, one can probe the effects of out-of-time pile-up through variations in $\langle\mu\rangle$.
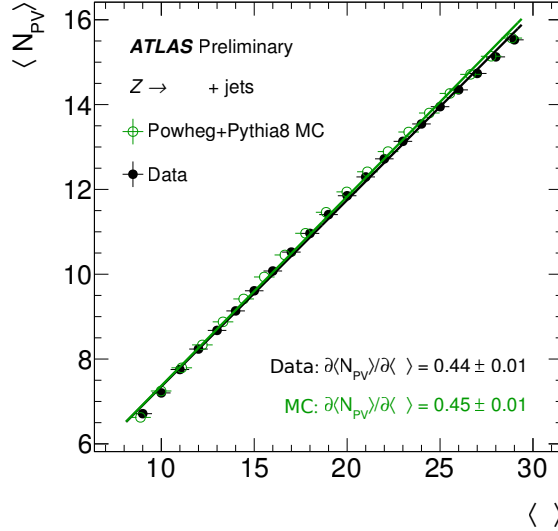

Figure 1: Mean N_{PV} as a function of $\langle\mu\rangle$ for Z+jet events in data and MC. Linear fits are carried out and the resulting gradients shown. Some non-linearity is also evident, due to the merging of multiple interactions into a single reconstructed vertex.

Also apparent in Fig. 1 is the effect of pile-up on primary vertex reconstruction. Some non-linearity is seen in the relation between N_{PV} and $\langle\mu\rangle$, due to multiple close-by interactions being reconstructed as a single vertex. With higher pile-up, the probability for vertex merging increases, causing the slope of N_{PV} vs. $\langle\mu\rangle$ to decrease. This effect illustrates one limitation in the use of N_{PV} to describe pile-up activity.

4.2 Topological clusters

The energy depositions in the cells of the calorimeters are the basis for cluster formation. The 4/2/0 topological clustering scheme [30] is used in which clusters are seeded by cells with signal greater than 4σ, where σ is the sum in quadrature of the standard deviations of the distributions of electronic and estimated pile-up noise. The cluster is then built from all 2σ cells surrounding the seed in three dimensions, and finally a single layer of all cells immediately adjacent to the 2σ region. A cluster splitting algorithm is then used to handle cases where multiple particles were clustered together. The topological cluster (topo-cluster) energy is either calibrated to the electromagnetic scale (EM)[4] or compensated for the presence of hadronic activity using local cluster weights (LCW)[5] [31].

---

[3]The "instantaneous" luminosity is actually an average computed over a time period of one minute, containing a large number of bunch crossings.

[4]The EM scale is defined by unit response to electromagnetic particles, i.e. photons and electrons.

[5]The LCW calibration method classifies clusters as electromagnetic or hadronic, and applies calibration accordingly, taking energy lost due to the non-compensation and non-instrumented regions of the calorimeters, as well as noise suppression, into account.

4.3 Jets

Topological clusters are used as input to FASTJET 2.4.3 [32,33] to form calorimeter jets with either the anti-k_t [2] or Cambridge/Aachen [3,4] algorithms. Truth jets are similarly reconstructed in MC simulation from the stable, interacting[6] particles in the final state of the the hard-scatter interaction in a simulated event. Track jets are reconstructed in both data and MC events from tracks originating from the hard-scatter vertex [34].

Unless otherwise specified, the jets used in these studies were calibrated using a similar multistage calibration as used for 2011 data [5, 8], replacing the previous average pile-up correction with the improved pile-up subtraction presented in Sec. 6. The pile-up subtraction is the first stage in the calibration procedure; it serves to remove the impact of pile-up from the jet p_T, but it does not correct for the difference in calorimeter response between single particles and jets. A jet-energy-scale (JES) response correction immediately follows the pile-up subtraction, resulting in average kinematical agreement between reconstructed calorimeter jets and simulated truth jets. Before the JES correction, the jet is said to be at the same energy scale as its constituent clusters (either EM or LCW); the application of the JES correction is indicated by the label EM+JES or LCW+JES.

It is important to note that the JES correction is not independent of the pile-up subtraction. Given a new pile-up subtraction procedure, the JES correction factors are updated in order to preserve the average calibrated energy scale resulting from the sequence of pile-up subtraction and JES correction.

4.4 Association of tracks and truth particles to jets

Calorimeter jets contain very little information on the origin of their constituent particles, so it is often useful to associate tracks or simulated truth particles to jets in order to determine the contribution to the jet from particles associated with a particular pp interaction. This association is done using the directions of tracks or truth particles at the point of closest approach to their originating vertex, so magnetic bending of trajectories in the ID does not affect the association. There are two standard methods of association: the simplest procedure is ΔR matching[7], while the more correct approach is known as *ghost association* [7].

For studies comparing reconstructed calorimeter jets to truth particle jets in simulation, calorimeter jets are required to geometrically match truth jets within a certain angular distance ΔR of the calorimeter jet axis. Matching is performed in order of decreasing jet p_T, discarding jets that have already been matched; ambiguities are resolved by choosing the jet with the highest p_T as the match.

Ghost association provides a more correct matching of tracks to calorimeter jets. In this technique, tracks are treated as infinitesimally soft, low-p_T particles by scaling their p_T by a very small number, such as 10^{-100}. These tracks are then added to the list of inputs to the jet algorithm. The low scale means the tracks do not affect the reconstruction of calorimeter jets. However, after the jet algorithm, it is possible to identify which tracks were clustered into which jets. This approach properly accounts for jets with irregular cross-sectional shapes, which would lead to incorrect association in the case of simple ΔR matching.

4.5 Electrons and muons

Electron reconstruction starts from energy deposits in the EM calorimeter (clusters). In the central region, $|\eta| < 2.47$, the clusters are matched to tracks reconstructed in the ID. The matching is done by comparing the η and ϕ coordinates of the impact point of the track to the corresponding parameters of

---

[6]Truth particles are considered stable if their decay length $c\tau$ is greater than 1 cm. A truth particle is considered to be interacting if it is expected to deposit most of its energy in the ATLAS calorimeters; muons and neutrinos are considered to be non-interacting.

[7]ΔR is the sum in quadrature of the difference in pseudorapidity and the difference in azimuth: $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$.

the seed cluster, after extrapolating the track from its last measurement point to the second layer of the electromagnetic calorimeter. In order to account for energy losses due to bremsstrahlung, tracks assigned to electrons are refitted using a Gaussian Sum Filter, as described in Ref. [35]. The 4-momenta of electron candidates is computed from both cluster and track parameters. More specifically, the energy of the electron candidate is given by the cluster energy, while η and ϕ are taken from the respective track parameters. In the forward region, where there is no tracking coverage, only the information from the calorimeter clusters is used. There, the direction of the electron is defined by the barycenter of the cells belonging to the cluster in the calorimeter and the energy of the electron is given by the sum of the cluster cells [36].

Electron identification in the central region is based on a cut-based selection, which uses a combination of calorimeter and tracking variables to provide a separation of isolated electrons from jets misidentified as electrons. Information from the EM shower shapes, track quality and cluster-track matching of the electron candidates are used to define three levels of electron-identification quality with increasing background rejection power: loose, medium and tight. The identification criteria for forward electrons are based on cluster moments and shower shapes which provide efficient identification due to the good transverse and longitudinal segmentation of the calorimeters. More details of the performance of electron reconstruction and identification can be found in Ref. [35].

Muon candidates are built from tracks reconstructed in the ID and in the muon spectrometer. Different classes of muon candidates are defined according to whether only tracks from the MS are used or a combination of tracks from both the ID and the MS. The studies described here use combined muons [37]. These are built from tracks reconstructed independently in the ID and the MS, with the muon momentum given by the weighted average of the independent momentum measurements in the ID and the MS [37].

5 Event selection

The studies presented in this note used different samples of both MC simulated events and recorded data, spanning several event topologies. Unless otherwise specified, reconstructed calorimeter jets were always required to satisfy loose quality criteria, including a pseudorapidity cut of $|\eta| < 4.5$. When calorimeter jets were matched to truth jets, they were also required to be truth-isolated, meaning that there is no other truth jet ($p_T > 4$ GeV) within ΔR of 1.5 times the jet radius.

5.1 Selection of dijet events from simulation

For the derivation of the new pile-up correction and optimization of associated parameters, as well as the MC-based assessments of performance, a sample of simulated dijet production was used. To avoid kinematic bias, no additional event selection criteria were used for this portion of the studies presented here.

5.2 $Z \to \ell\ell$ + jets selection

Events (both data and MC) with the Z+jets topology were used extensively for validation of the pile-up correction, for the assessment of systematic uncertainties, and in the jet-vertex-fraction studies of Sec. 7.

5.2.1 Dataset and triggers

The dataset used for Z+jets studies corresponds to the full integrated luminosity in 2012 of 20.3 fb^{-1}. The dataset was selected requiring all subdetectors to be active and ready. Electron events were triggered

by an isolated electron with $p_T \geq 24$ GeV (60 GeV if not isolated). Similarly, muon events were triggered by a 24 (36) GeV muon.

5.2.2 *Z* candidate

Each event used for *Z*+jets studies is required to contain a *Z* decay candidate consisting of two opposite-sign electrons or muons, with dilepton invariant mass in the range $66 < m_{\ell\ell} < 116$ GeV. Electrons are accepted within $|\eta| \leq 2.47$, excluding the crack region $1.37 \leq |\eta| \leq 1.52$, and required to satisfy medium++ quality criteria, an ID requirement reoptimised from those in Ref. [38]. Muons are accepted within $|\eta| < 2.4$ and required to satisfy quality criteria based on a 2012 reoptimisation of the definition in Ref. [37]. Muons are further required to be isolated by restricting the scalar sum of the transverse momenta of other tracks around the muon candidate. Both electrons and muons are required to have $p_T > 25$ GeV. Events are vetoed if there is a third lepton satisfying these criteria.

5.2.3 *Z* + jet balance criteria

In order to exploit the p_T balance between *Z* boson and jet for in-situ validation of the pile-up correction, further event selection criteria are necessary beyond the mere existence of a *Z* candidate. Jets are required to be isolated from leptons in the event: $\Delta R(\text{jet}, \text{lepton}) > 0.5$ for $R = 0.6$ jets and $\Delta R(\text{jet}, \text{lepton}) > 0.35$ for $R = 0.4$ jets. The dilepton invariant mass is required to be within $80 < m_{\ell\ell} < 116$ GeV. The leading (sub-leading) jet is defined as the most energetic (second most energetic) jet in the event with $p_T > 10$ GeV. If any jet is within $|\eta_{\text{det}}| < 2.5$, then it is further required to satisfy $|JVF| > 0.25$ (see Section 7). Events are vetoed if the sub-leading jet $p_T \geq \max(10 \text{ GeV}, 0.2 p_T^{\text{ref}})$, where p_T^{ref} is the projection of the p_T of the *Z* boson onto the leading-jet axis, defined by $p_T^{\text{ref}} \equiv p_T^Z |\cos \Delta\phi(\text{jet}, Z)|$. This reference p_T is used instead of p_T^Z in order to correct for additional parton radiation [39].

6 Pile-up subtraction

The standard pile-up correction in the ATLAS jet calibration procedure for the 2011 dataset was an average correction parametrized by N_{PV} and $\langle\mu\rangle$. The correction was derived by comparison of reconstructed calorimeter jets to truth particle jets in simulated dijet events [8]. It was derived separately for each jet definition, since the sensitivity of a jet to pile-up generally depends on jet algorithm and the distance parameter *R* used. Although there can be significant fluctuations in pile-up activity even between events with the same N_{PV} and $\langle\mu\rangle$, the average offset correction could not account for this. Furthermore, full reliance on N_{PV} meant that the correction was sensitive to inefficiency of vertex reconstruction, which is itself dependent on the pile-up (see Sec. 4.1).

In response to the limitations in the average offset correction and the significant increase in pile-up activity in 2012 as compared to 2011, an improved pile-up correction has been developed, using the jet area and the median p_T density ρ (described in Sec. 6.1 and Sec. 6.2, respectively), as well as N_{PV} and $\langle\mu\rangle$. The jet area is a measure of the susceptibility of the jet to pile-up, and is measured jet-by-jet. The density, ρ, is a measure of the pile-up activity in the event. Using the jet area based method for the jet pile-up correction provides an estimate of the amount of pile-up contained in each jet, and a way to correct the jet four-momentum accordingly. The underlying assumption is that the pile-up can be treated as a uniform, diffuse background, adding signal to the jets. In this case a pile-up density times an area gives the amount of pile-up inside the jet.

Comparing this correction to the average offset correction used previously in ATLAS, there are four main advantages:

- ρ captures event-by-event fluctuations in pile-up that are not described by N_{PV} and $\langle\mu\rangle$, leading to resolution improvement and improved rejection of pile-up jets.

- Any mis-modelling of the pile-up activity described by ρ is automatically corrected on an event-by-event basis; the correction does not rely on a proper modelling of the relationship between ρ and N_{PV} or $\langle\mu\rangle$.

- The same correction may be applied to any jet regardless of clustering algorithm or radius, as the jet area provides a direct estimate of pile-up sensitivity for each jet; only a small residual correction is necessary that remains specific to each jet definition.

- The method relies less on the correct identification of tracks and vertices, as the bulk of the correction is based only on calorimeter information. The correction is therefore largely insensitive to the effects of pile-up on vertex reconstruction.

6.1   The jet area

The jet area [40] is measured using an algorithm implemented in FASTJET. It is evaluated by overlaying a dense, uniform population of neutral, infinitesimally soft ($p_T = O(10^{-100}$ GeV)) "ghost" particles onto the event and including these ghosts in the jet clustering. The ghost constituents do not affect the jet clustering of an infrared safe algorithm, but the number of ghost constituents associated to each jet serves as a measure of its area.

The concept of jet area can be taken one step further, using the fact that the ghost constituents all have four-momenta. Summing over the ghost four-momenta g_i belonging to a jet j, one obtains the jet area four-momentum A_j. This is defined as:

$$A_j = \frac{1}{\nu_g\langle g_t\rangle} \sum_{g_i \in j} g_i,$$

(2)

where $\nu_g\langle g_t\rangle$ is the transverse momentum density of the ghosts. In practice, the ghost transverse momentum g_t is constant: the area is intended to describe sensitivity to a background of uniform p_T density within the angular extent of the jet. The transverse component of the (more exact) area four-momentum will, in the limit of small enough jets, be equal to the more intuitive "scalar area" obtained from counting the number of ghosts associated to a jet and dividing by the ghost number density.

A theoretically sound estimate of the area of any jet (not only circular jets) is one of the advantages of the method. For example, jets reconstructed using the Cambridge/Aachen algorithm will generally be of irregular shape. This is illustrated in Fig. 2, showing areas normalised to πR^2 (the expected area of a circular jet with distance parameter R) for two different jet algorithms: the Cambridge/Aachen algorithm and anti-k_t algorithm. In the case of anti-k_t, leading or isolated jets will be close to circular, with an area close to πR^2. This is also shown in Fig. 2. The reason is that the anti-k_t algorithm starts out with the hardest contributions, and clusters situated within R of several jets will in this algorithm be attributed to the harder jet. Non-isolated jets with lower p_T will thus be crescent shaped (since the intersection with a circular jet is cut out from some of its area), or even more irregular. This is reflected in the low-area tail of the anti-k_t distributions.

6.2   The median p_T density ρ

The median p_T density measured in the event is defined as

$$\rho = \mathrm{median}\left\{\frac{p_{T,i}^{\mathrm{jet}}}{A_i^{\mathrm{jet}}}\right\},$$
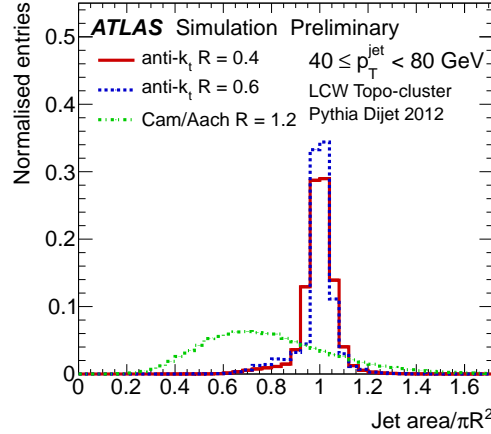
(3)

8

Figure 2: Jet area distributions for two different jet clustering algorithms in the same set of events: anti-k_t and Cambridge/Aachen.

where the index i enumerates the jets found when clustering the event with the k_t algorithm [41, 42]. The term jet applies to any clustered entity of output from the jet algorithm, and is not necessarily related to a collimated shower of particles produced in an interaction.

The use of the k_t algorithm for the ρ calculation is motivated by its sensitivity to soft radiation; for more details see Ref. [7], the original paper proposing the method. No lower p_T cut is used in the calculation of ρ since the goal is to quantify the predominantly soft pile-up background. The transverse energy density from pile-up is roughly uniform in pseudorapidity, which motivates the use of p_T in the ρ calculation.

ρ is calculated from all positive energy calorimeter topo-clusters within $|\eta| \leq 2.0$, clustered together with a uniform population of infinitesimally soft ghost particles, using a k_t distance parameter $R_{k_t} = 0.4$ and the Voronoi area definition (for more details on area definitions see [40]). Two choices of calibration scale (EM and LCW) are provided, matching the available choices of calibration scale of the jets that will eventually be corrected.

Although in this note ρ is calculated using calorimeter information, in principle ρ can be calculated from any suitable input to a jet algorithm, for example tracks. A track based evaluation of ρ is only sensitive to the charged component of in-time pile-up. When based on calorimeter information, ρ also measures the contributions from neutral particles and to a lesser extent out-of-time pile-up.

The jets used to calculate ρ are generally not the same jets as those reconstructed as physics objects for an analysis, which are often reconstructed using another jet algorithm. The hard scatter jets can however also be present in the clustering for determining ρ, and they will have a large $\rho_{\text{jet}} \equiv \frac{p_T^{\text{jet}}}{A^{\text{jet}}}$. Using the event median greatly suppresses the influence of these jets on ρ, as they will be in the high-density tail of the distribution for the event.

Figure 3 illustrates the general properties of ρ. In particular, ρ increases with N_{PV} and for a given combination of N_{PV} and μ there are large fluctuations in ρ. These fluctuations are larger for higher N_{PV}, indicating that event fluctuations will be an even more important effect for higher luminosity where in-time pile-up will increase. This illustrates the need for a method which quantifies pile-up using calorimeter information from the actual event.

The choice of η range used for calculating ρ is motivated by the calorimeter occupancy (topo-cluster density in η) as illustrated in Fig. 4. Figure 4(a) shows the calorimeter positive cluster distribution in η, overlaid with the calorimeter granularity [30]. Figure 4(b) in turn shows the average ρ when calculated using narrow strips in η. The strips are $\Delta\eta = 0.7$ wide and shifted in steps of $\delta\eta = 0.1$ going from
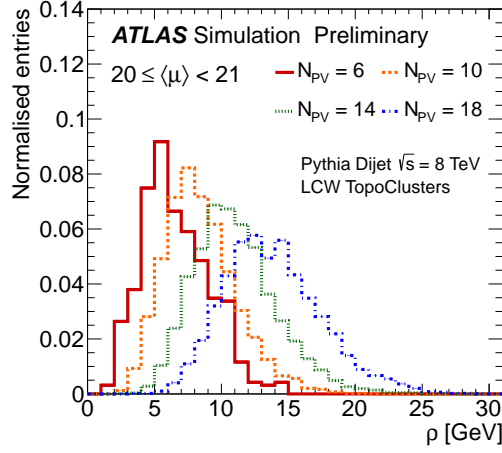
Figure 3: The event p_T density, ρ, distribution for an average number of interactions $20 < \langle\mu\rangle < 21$, drawn for four different values of N_{PV}.

$\eta = -4.9$ to 4.9.

(a) Granularity and positive cluster distribution vs η
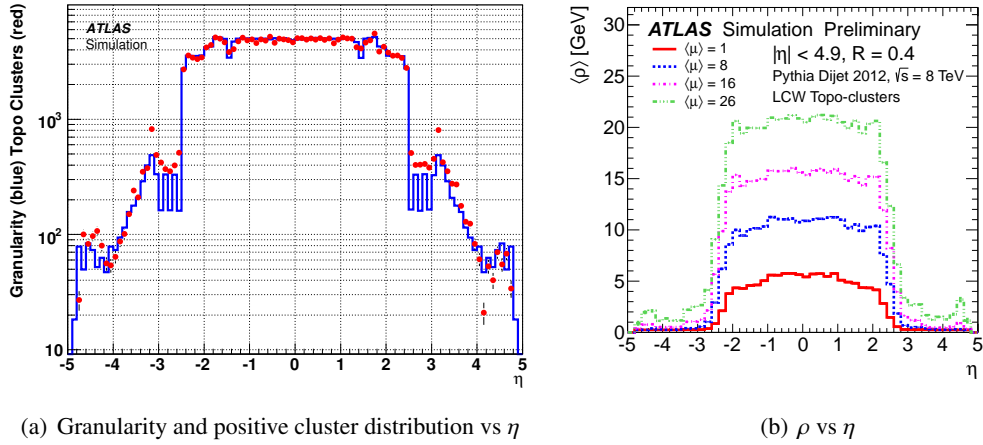

(b) ρ vs η

Figure 4: (a) Calorimeter topo-cluster distribution (red points) overlaid with calorimeter granularity (blue lines) [30]. (b) The median p_T density, ρ, evaluated as a function of η using a sliding η interval of width $\Delta\eta = 0.7$ in the range $|\eta| < 4.9$.

The cause of the low (pile-up) occupancy in the forward region is two-fold. The first and physically motivated reason is that the cross-section as a function of pseudorapidity of particles of a given p_T produced in inelastic collisions is falling at higher values of η. The second and dominant reason is, as is clear from Fig. 4(a), the coarser calorimeter granularity at higher η, coupled with the noise suppression inherent in topological clustering. Since clusters are seeded on significance relative to (electronic and pile-up) noise rather than an absolute threshold, having a larger number of cells (finer granularity) increases the probability that one cell will fluctuate up to a significant value due to (electronic or pile-up) noise. With the coarser granularity in the forward region, which notably sets in around $|\eta| = 2.5$, this probability becomes smaller, and clusters will predominantly be seeded only by the hard scatter signal.

Figure 4(b) shows that as $\langle\mu\rangle$ increases, so does $\langle\rho\rangle$ in the central region of the calorimeters, while it stays close to 0 at larger $|\eta|$. A calculation of ρ in the central region thus gives a more meaningful

10

measure of the pile-up activity than the median over the entire η range, or an η dependent ρ calculated in slices across the calorimeter.

In the absence of signal in a region, the jet clustering algorithm can form jets that are made up entirely of ghost constituents, and thus have $p_T = 0$. When instead using the Voronoi area definition, the area of jets becomes very large in sparse events. Both of these cases lead to a very small p_T density, ρ_{jet}, of the individual jets. In the central calorimeter region, the cluster density is generally large enough for most k_t jets to have non-zero ρ_{jet}. In the forward region, only clusters that are part of high-p_T jets will have non-zero p_T, and there will be many jets made up entirely of ghost constituents. Similarly, when using the Voronoi area as in the ρ calculation, the k_t jet area will be very large, leading to small ρ_{jet}. Calculating ρ locally here will give $\rho \sim 0$; since ρ is the median, it will be dominated by jets with small ρ_{jet}. For $|\eta| < 2.0$, the ρ curves are non-zero and flat vs η. If the regions of the detector with very little pile-up signal are included in the ρ calculation, detector effects will thus have a large impact on the result. Such a calculation will give a value of ρ that does not reflect the real pile-up activity. From these considerations, the range $|\eta| < 2.0$ is used for the calculation of ρ.

Fig. 4(b) also shows the result of the ρ calculation in the presence of very little pile-up ($\langle\mu\rangle = 1$). In the central region, the median p_T density remains non-zero because ρ is sensitive to low-p_T activity originating from the hard-scatter interaction. In other words, ρ describes not only the event-by-event pile-up activity, but also the underlying event of the hard-scatter interaction. This leads to some amount of topology dependence in the absolute scale of ρ, though the change in ρ as a function of pile-up is unaffected. In Sec. 6.7, the modelling of this topology dependence is taken into account in the assessment of systematic uncertainties for pile-up subtraction.

6.3   Subtracting $\rho \cdot A$

The median p_T density ρ provides a direct estimate of the global pile-up activity in any given event, while the jet area provides an estimate of a jet's sensitivity to pile-up. By the multiplication of these two quantities, an estimate of the effect of pile-up on the jet is obtained. As a first-pass correction this estimate can be subtracted from the jet p_T to achieve significantly reduced dependence on pile-up:

$$p_T^{\mathrm{jet,corr}} = p_T^{\mathrm{jet}} - \rho \cdot A, \tag{4}$$

where A is the transverse component of the area four-momentum A_j.

Figure 5 shows the effect of $\rho \cdot A$ subtraction on the in-time pile-up dependence of the jet p_T for four different jet definitions: anti-k_t jets with distance parameters $R = 0.4, 0.6$ and 1.0, and Cambridge/Aachen jets with distance parameter $R = 1.2$. The slopes of linear fits to $\langle p_T^{\mathrm{reco}} - p_T^{\mathrm{true}}\rangle$ vs. N_{PV}, for a fixed value of $\langle\mu\rangle$, are shown as functions of η. The dependence on in-time pile-up is reduced from several hundred MeV per vertex to $\lesssim 200$ MeV per vertex in all cases, demonstrating that $\rho \cdot A$ subtraction is able to effectively reduce the impact of in-time pile-up regardless of jet definition, though a small residual dependence on N_{PV} remains. The larger sensitivity to in-time pile-up of larger area jets is also seen. Similar performance was found in Z+jets samples and across a wide range of values for $\langle\mu\rangle$.

Finally, it should be noted that the pile-up activity in any given event may have significant local fluctuations of similar angular size to jets, and a global correction such as this cannot account for them. A variable such as JVF may be used to reject jets that result from such fluctuations in pile-up density, as described in Sec. 7.
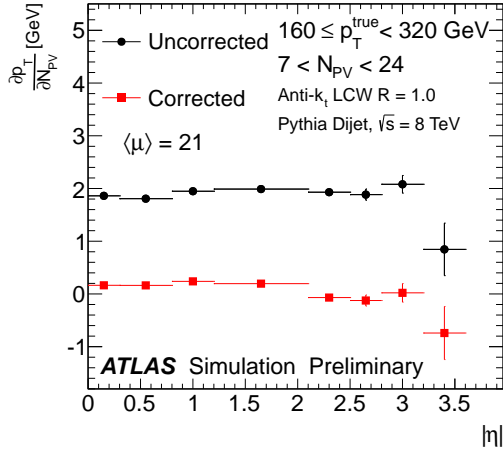
6.4   Residual correction

There are two ways in which pile-up can contribute energy to an event: either by forming new clusters, or by overlapping with signal from the triggering event. Because of the noise suppression inherent in
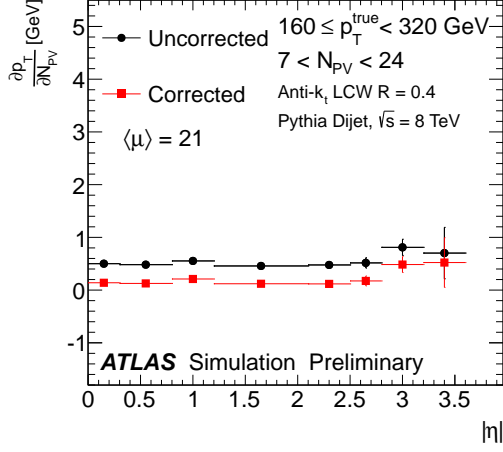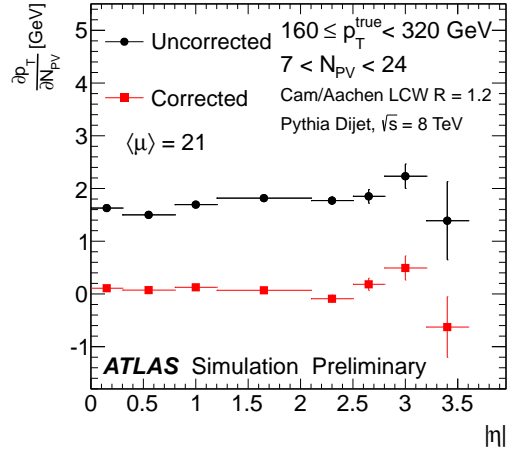
(a) Dependence on N_{PV}, vs η, anti-k_t $R = 0.4$
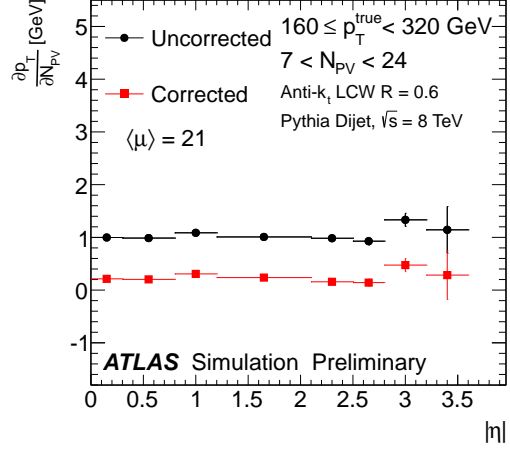
(b) Dependence on N_{PV}, vs η, anti-k_t $R = 0.6$

(c) Dependence on N_{PV}, vs η, anti-k_t $R = 1.0$

(d) Dependence on N_{PV}, vs η, Cambridge/Aachen $R = 1.2$

Figure 5: Dependence of jet transverse momentum p_{T} on in-time pile-up, represented by the number of primary vertices N_{PV} for a given average number of interactions $\langle\mu\rangle$. The dependence is drawn before (black points) and after (red squares) correction, as a function of the pseudorapidity η. Different jet algorithms and distance parameters R are shown.

topological clustering, only pile-up signal above a certain threshold can form separate clusters. Low energy pile-up deposits can thus only contribute signal if they overlap with other deposits which survive noise suppression. The probability of overlap is dependent on the transverse size of EM and hadronic showers in the calorimeter, relative to the size of the calorimeter cells. In the central regions of the calorimeter, due to fine granularity, pile-up mainly contributes extra clusters. This is the region where ρ is calculated.

As discussed in Sec. 2, the bipolar shaping of the liquid argon calorimeter pulses can result in negative-energy signals associated with out-of-time pile-up activity. If isolated from in-time signals, they may form negative clusters, which are excluded from jet reconstruction and the calculation of ρ. However, overlap between positive jet signal and out-of-time activity will result in a negative modulation of the jet energy. This could be seen as a negative dependence in the jet p_T on out-of-time pile-up. Such overlap is more probable at higher η, due to coarser granularity relative to the transverse shower size. In addition, the length of the bipolar pulse is shorter in the forward calorimeters, resulting in a negative tail of larger amplitude. As a result, forward jets have enhanced sensitivity to out-of-time pile-up.

Since the ρ calculation is dominated by low-occupancy regions in the calorimeter, the sensitivity of ρ to pile-up does not fully describe the pile-up sensitivity of the high-occupancy region at the core of a high-p_T jet. Furthermore, the effects of pile-up in the forward region are not well described by the median p_T density as obtained from positive clusters in the central region. A residual correction is therefore necessary to obtain an average jet response that is insensitive to pile-up across the full η range.
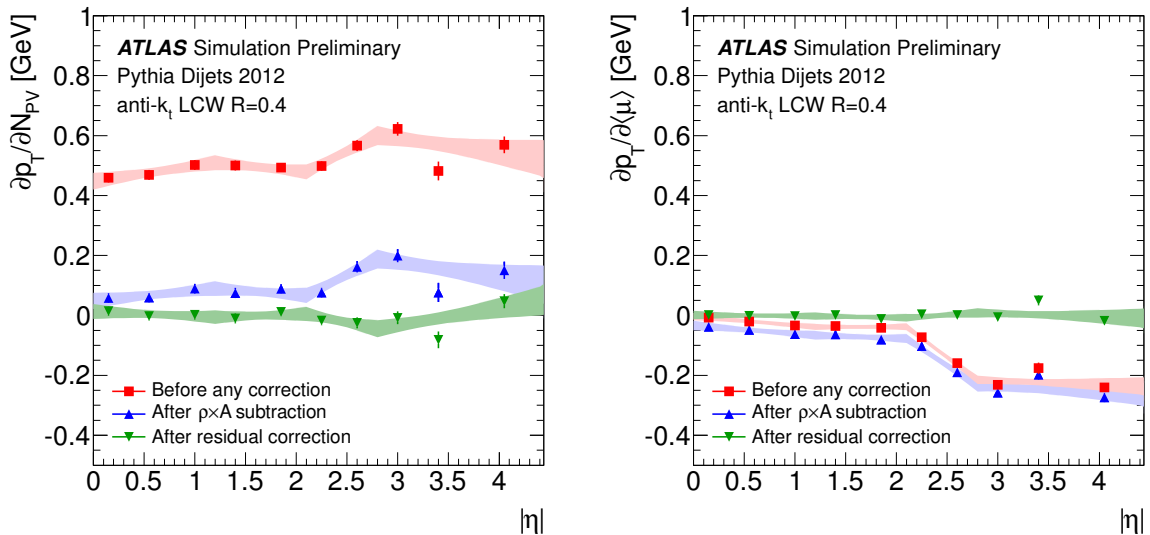
(a) Dependence of jet p_T on N_{PV} averaged over all $\langle\mu\rangle$ bins.   (b) Dependence of jet p_T on $\langle\mu\rangle$ averaged over all N_{PV} bins.

Figure 6: Dependence of the reconstructed jet p_T (anti-k_t, $R = 0.4$, LCW scale) on in-time pile-up (a) and out-of-time pile-up (b), at various correction stages: before any correction, after $\rho \cdot A$ subtraction, and after the residual correction. The dependence is shown in bins of jet $|\eta|$ and fit using the same functional form as in the residual correction itself. The error bands show the 68% confidence intervals of the fits. The dependence was obtained by comparison to truth particle jets in simulated dijet events, and corresponds to a truth-jet p_T of 25 GeV.

Figure 6 shows the dependence of anti-k_t $R = 0.4$ LCW jets on (a) N_{PV} for fixed $\langle\mu\rangle$ and (b) $\langle\mu\rangle$ for fixed N_{PV}, probing separately the effects of in-time and out-of-time pile-up, respectively. The subtraction of $\rho \cdot A$ removes most of the sensitivity to in-time pile-up, though some residual dependence on N_{PV} remains. Fig. 6(b) shows that $\rho \cdot A$ subtraction has very little effect on the sensitivity to out-of-time pile-

13

(a) N_{PV} term
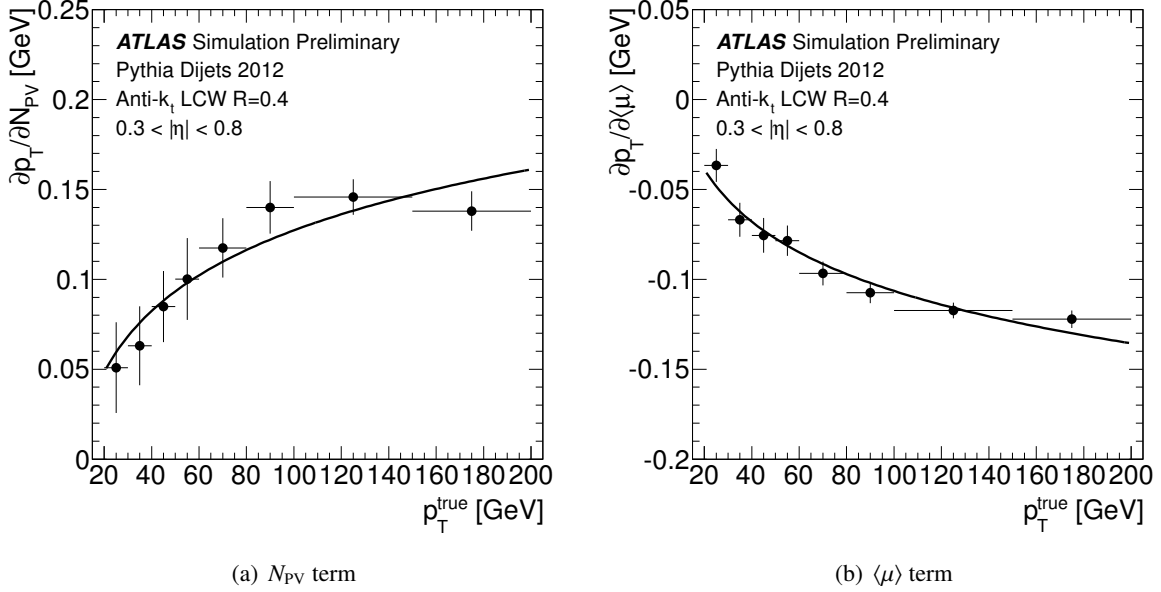
(b) $\langle \mu \rangle$ term

Figure 7: Dependence of the coefficients of the residual pile-up correction on the p_T of matched truth particle jets in simulated dijet events, for anti-k_t $R = 0.4$ jets at the LCW scale and within the pseudorapidity range $0.3 < |\eta| < 0.8$. The points are consistent with a logarithmic functional form.

up, which is particularly significant in the forward region. These results were obtained from linear fits to $p_T^{reco} - p_T^{true}$ vs. N_{PV} ($\langle \mu \rangle$) in bins of p_T^{true}, η, and $\langle \mu \rangle$ (N_{PV}). Within each (p_T^{true}, η) bin, it was verified that the slope of the linear fit did not depend significantly on $\langle \mu \rangle$ (N_{PV}); in other words, there was no statistically significant evidence for non-linearity or cross-terms in the sensitivity of the jet p_T to in-time or out-of-time pile-up for the values of $\langle \mu \rangle$ seen in 2012 data.

After subtracting $\rho \cdot A$ from the jet p_T, there is an additional subtraction of a residual term proportional to the number of reconstructed pile-up vertices $N_{PV} - 1$ (to account for in-time pile-up), as well as a residual term proportional to $\langle \mu \rangle$ (to account for out-of-time pile-up). This residual correction is derived by comparison to truth particle jets in simulated dijet events, and it is completely analogous to the average pile-up offset correction used previously in ATLAS [8]. Due to the preceding $\rho \cdot A$ subtraction, the residual correction is generally quite small for jets with $|\eta| < 2.1$. In the forward region, the negative dependence in jets on out-of-time pile-up results in a significantly larger residual correction. The $\langle \mu \rangle$-dependent term of the residual correction is approximately the same size as corresponding term in the average offset correction used previously in ATLAS, but the N_{PV}-dependent term is significantly smaller. This is true even in the forward region, which shows that ρ is a useful estimate of pile-up activity even beyond the region in which it is calculated.

Due to the fact that pile-up sensitivity is related to local occupancy in the calorimeter, the effect of pile-up on the reconstructed jet p_T is itself dependent on the jet p_T. As the jet p_T increases, so does the signal occupancy at the core of the jet, leading to increased pile-up sensitivity. This is observed during the derivation of the residual correction, as the coefficients of the N_{PV} and $\langle \mu \rangle$ terms are seen to vary as a function of the p_T of the matched truth particle jet.

Figure 7 shows this dependence for both the N_{PV} term and the $\langle \mu \rangle$ term, for anti-k_t $R = 0.4$ jets at the LCW scale in the pseudorapidity range $0.3 < |\eta| < 0.8$. A logarithmic functional form fits the data well for both terms, though the p_T dependence of the $\langle \mu \rangle$ term is opposite to that of the N_{PV} term. The fits are evaluated at 25 GeV to obtain the nominal coefficients of the correction.

The p_T dependence shown in Fig. 7 is neglected in the residual correction as applied to jets in ATLAS.
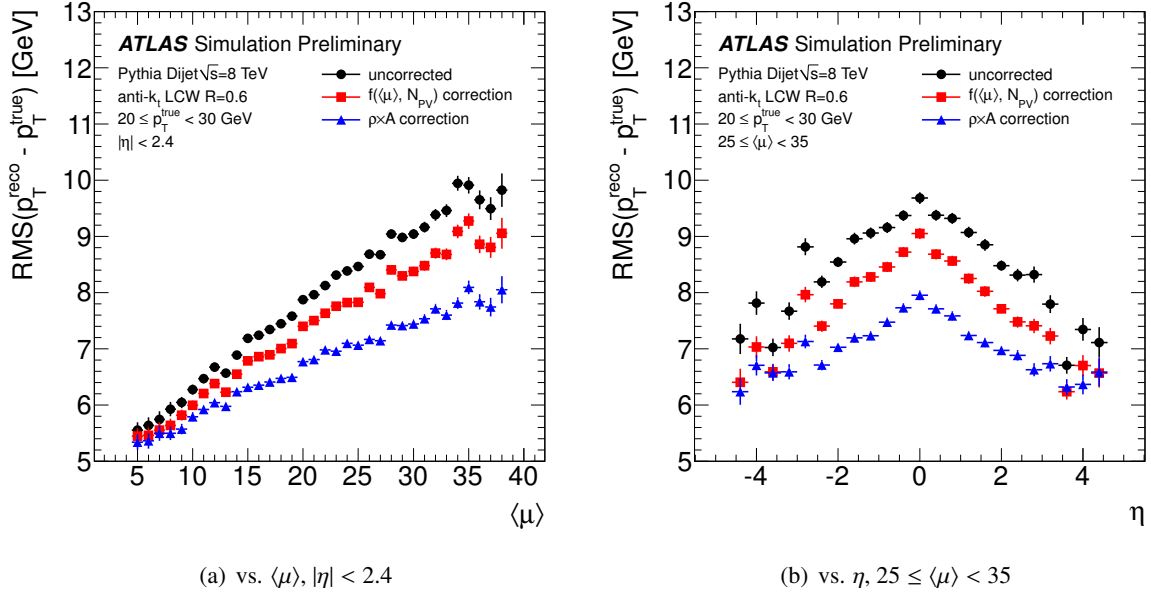
(a) vs. $\langle\mu\rangle$, $|\eta| < 2.4$

(b) vs. η, $25 \leq \langle\mu\rangle < 35$

Figure 8: RMS width of the $p_T^{\mathrm{reco}} - p_T^{\mathrm{true}}$ distribution versus $\langle\mu\rangle$ (a) and versus pseudorapidity η (b), for anti-k_t $R = 0.6$ jets at the LCW scale matched to truth particle jets satisfying $20 < p_T^{\mathrm{true}} < 30$ GeV, in simulated dijet events. The new pile-up subtraction method based on jet areas is seen to improve the resolution of the reconstructed jet p_T over the previous subtraction method.

This uncorrected dependence is taken as a source of systematic bias: the logarithmic fits are used to obtain a p_T-dependent systematic uncertainty, defined by the coefficients $\alpha_{N_{\mathrm{PV}}}$ and $\alpha_{\langle\mu\rangle}$ from

$$\frac{\partial p_T}{\partial N_{\mathrm{PV}}}(p_T) = \frac{\partial p_T}{\partial N_{\mathrm{PV}}}(25\ \mathrm{GeV}) + \alpha_{N_{\mathrm{PV}}} \times \log\left(\frac{p_T}{25\ \mathrm{GeV}}\right) \tag{5}$$

and

$$\frac{\partial p_T}{\partial \langle\mu\rangle}(p_T) = \frac{\partial p_T}{\partial \langle\mu\rangle}(25\ \mathrm{GeV}) + \alpha_{\langle\mu\rangle} \times \log\left(\frac{p_T}{25\ \mathrm{GeV}}\right). \tag{6}$$

6.5 Resolution improvement

Figure 8 shows the RMS width of the $p_T^{\mathrm{reco}} - p_T^{\mathrm{true}}$ distribution in simulated dijet events, as a function of the amount of pile-up as characterized by $\langle\mu\rangle$ and versus jet η. The impact of pile-up on the jet p_T resolution is evident from the linear rise observed in the uncorrected points in (a). When compared to the previous pile-up subtraction method, the new pile-up correction based on jet areas further alleviates the degradation in jet p_T resolution due to pile-up. However, some dependence of the RMS width on $\langle\mu\rangle$ remains, which may be attributed to local fluctuations in the pile-up activity. Figure 9 shows even more clearly the advantage of the new correction over the old, as the old correction resulted in almost no improvement within any given bin of N_{PV}.

6.6 In-situ validation

Two methods of in-situ validation of the pile-up correction are employed to study the dependence of jet p_T on N_{PV} and $\langle\mu\rangle$. The first method uses track-jets to provide a pile-up independent measure of jet p_T. This requires the presence of track-jets and so can only be used in the most central region of the detector for $|\eta| < 2.1$. It is not statistically limited. The second method exploits the p_T balance between a
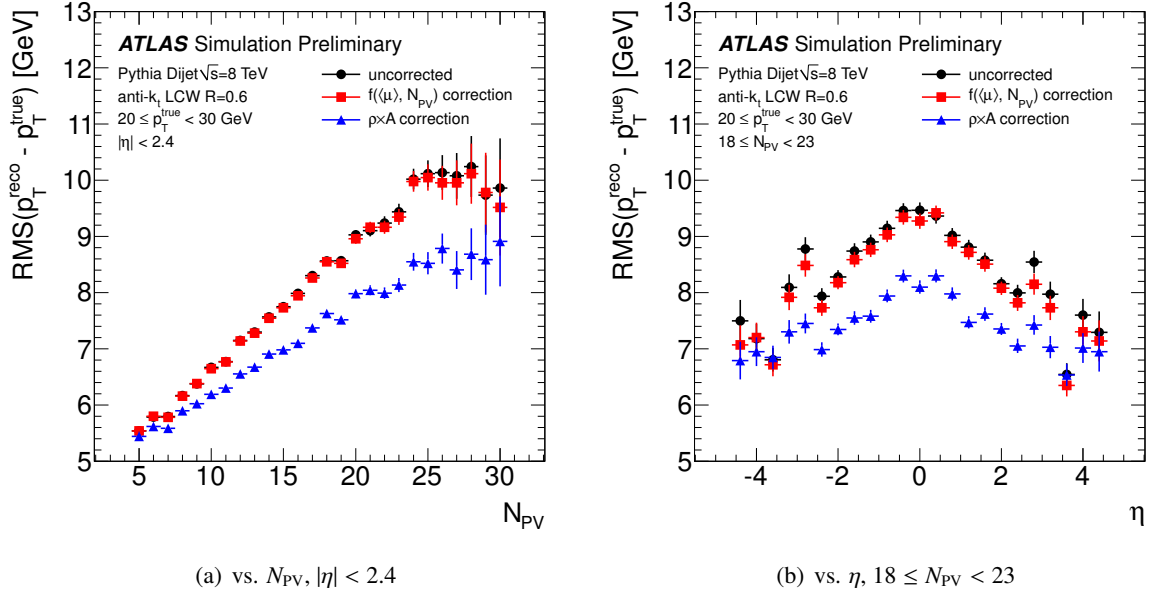
15

(a) vs. N_{PV}, $|\eta| < 2.4$

(b) vs. η, $18 \le N_{PV} < 23$

Figure 9: RMS width of the $p_T^{reco} - p_T^{true}$ distribution versus N_{PV} (a) and versus pseudorapidity η (b), for anti-k_t $R = 0.6$ jets at the LCW scale matched to truth particle jets satisfying $20 < p_T^{true} < 30$ GeV, in simulated dijet events. Binning in N_{PV}, one can see clearly the advantage of the new correction over the previous subtraction method.

reconstructed jet and a Z boson, using the Z p_T as a measure of the jet p_T. This enables an analysis over the full ($|\eta| < 4.5$) range of the detector, but the extra cuts which must be applied to the jet and Z boson mean that it is statistically limited. The N_{PV} dependence must therefore be evaluated inclusively in $\langle\mu\rangle$ and vice versa. This results in a degree of cross talk between the measured N_{PV} and $\langle\mu\rangle$ dependence.

While the pile-up correction was derived from simulated dijet events, the in-situ validation is done entirely using Z+jets events, aside from studies of topology dependence (see Section 6.7). In the track-jet validation, although the kinematics of the Z boson candidate are not used directly, the dilepton system is relied upon for triggering, thus avoiding any potential bias from jet triggers.

6.6.1 Track-jet validation

The presence of a Z boson as outlined in Sec. 5 is used to select an unbiased sample of jets in both MC and data. All reconstructed jets in the event are then matched to track-jets using the method outlined in Sec. 4.4 and the reconstructed calorimeter jet p_T distribution is plotted in bins of width 1 unit in both $\langle\mu\rangle$ and N_{PV} and width 5 GeV in p_T^{track}. The reconstructed p_T distribution is fitted with a Gaussian to extract a mean reconstructed p_T in each bin. These values are used to find the dependence of reconstructed p_T on $\langle\mu\rangle$ by linearly fitting the $\langle\mu\rangle$ dependence of the reconstructed p_T and averaging over the gradients found in each N_{PV} bin. The N_{PV} dependence is extracted similarly by averaging over $\langle\mu\rangle$ bins. The average slopes are found by averaging over the slopes found in each p_T^{track} bin.

This method assumes the pile-up independence of track-jet p_T. Figure 10 demonstrates this independence by matching track-jets to truth-jets and using the same method to extract the $\langle\mu\rangle$ and N_{PV} dependence as discussed above. The p_T range of track-jets used in this study is from 15 to 30 GeV which corresponds to the lower p_T^{true} bins in Fig. 10 where the points are mostly consistent with being pile-up independent.

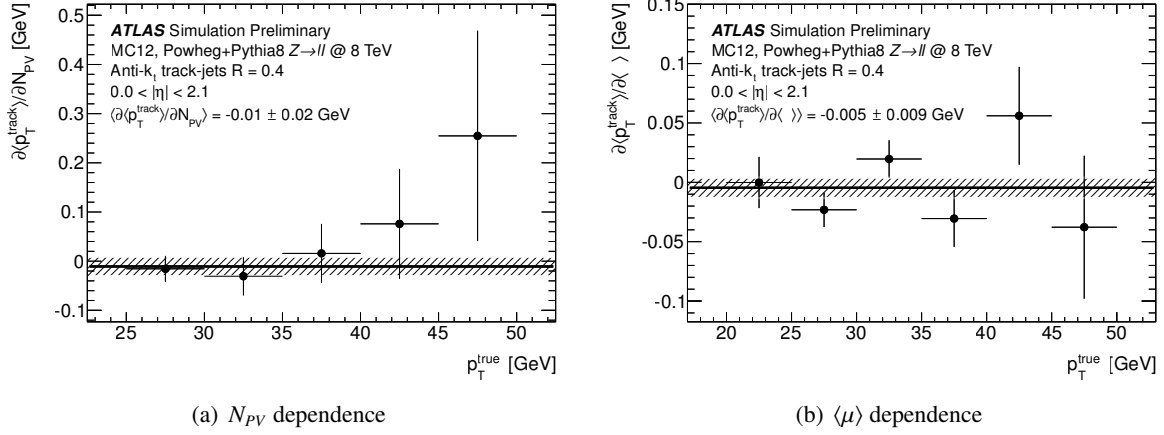Figure 11 shows the results obtained when matching anti-k_t, R=0.4, LCW reconstructed jets to anti-

16

(a) N_{PV} dependence

(b) $\langle\mu\rangle$ dependence

Figure 10: N_{PV} and $\langle\mu\rangle$ dependence of the p_T of anti-k_t, R=0.4 track-jets as a function of truth-jet p_T. Consistency of the average dependence with zero demonstrates the independence of p_T^{track} from pile-up.

k_t, R=0.4 track-jets. Good agreement is observed between data and MC; however, a small overcorrection is observed in the N_{PV} dependence of both MC and data. In the in-situ uncertainty combination outlined in Sec. 6.6.3 this non-closure of the correction is taken as an uncertainty on the jet p_T dependence on N_{PV}.

(a) N_{PV} dependence
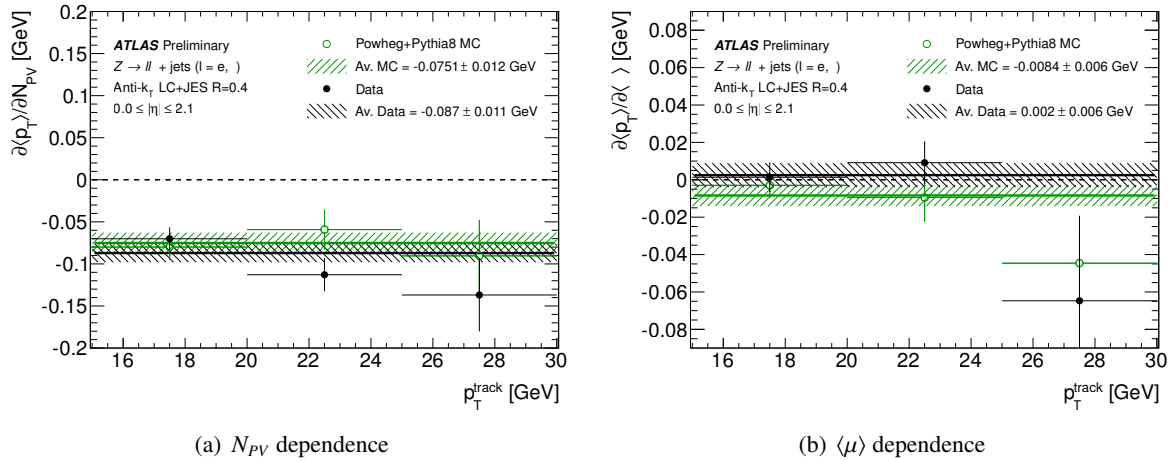
(b) $\langle\mu\rangle$ dependence

Figure 11: N_{PV} and $\langle\mu\rangle$ dependence of the reconstructed p_T of anti-k_t, R=0.4, LCW jets as a function of track-jet p_T.

6.6.2 Z + jet validation

In events where a Z boson is produced in association with one jet, momentum conservation ensures balance between the Z boson and the jet in the transverse plane. In the direct p_T balance method, this principle is exploited by using the p_T of the Z boson as a proxy for the true jet p_T. In the case of a perfect measurement of lepton energies and provided that all particles recoiling against the Z boson are included in the jet cone, the jet is expected to balance the Z boson. In practice, the balance is spoiled by the following factors:

- imperfect measurement of lepton energies;

- presence of additional parton radiation in the hard scattering event;

- energy associated with the leading parton that is deposited outside of the jet cone, as a result of the shower development in the calorimeter, magnetic field bending etc.;

- contributions to the jet coming from underlying-event activity;

- contributions to the jet coming from in-time and out-of-time pile-up.

The event selection outlined in Sec. 5.2.3 is designed to mitigate the effect of the first factors, while the jet area based correction should subtract contributions to the leading jet from the underlying event and pile-up. In the case of a perfect subtraction scheme, one expects $\Delta p_T \equiv p_T^{\text{jet}} - p_T^{\text{ref}} = p_T^{\text{jet}} - p_T^Z |\cos \Delta\phi(\text{jet}, Z)|$ to be independent of the amount of pile-up in the event, i.e. Δp_T should be independent of N_{PV} and $\langle \mu \rangle$.

The arithmetic mean of Δp_T is computed in bins of p_T^{ref} and η^{jet}. Iterating over all values of $\langle \mu \rangle$, one obtains $\langle \Delta p_T \rangle$ as a function of $\langle \mu \rangle$ for a given p_T^{ref} and η^{jet} bin, as shown in Fig. 12.
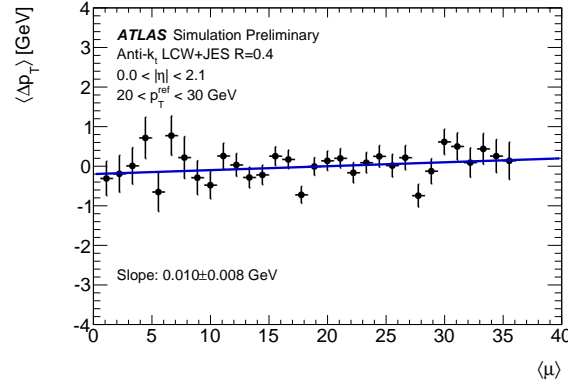

Figure 12: $\langle \Delta p_T \rangle$ as a function of $\langle \mu \rangle$ for simulated Z+jet events plotted in the range $|\eta| < 2.1$ for $20 < p_T^{\text{ref}} < 30$ GeV for anti-k_t, R=0.4, LCW jets. The points represent central values of the arithmetic mean of Δp_T and the error bars correspond to the associated standard error.

Taking the mean of the $\langle \Delta p_T \rangle$ distribution, the slope $\frac{\partial \langle \Delta p_T \rangle}{\partial \langle \mu \rangle}$ is extracted and plotted as a function of p_T^{ref}, as shown in Fig. 13. A small residual slope is observed after the jet-areas correction. As shown in Fig. 13, the data is well modeled by Monte Carlo. Any non-closure or mis-modeling, quantified by the difference of $\frac{\partial \langle \Delta p_T \rangle}{\partial \langle \mu \rangle}$ between data and MC, is accounted for by an uncertainty as described in Sec. 6.6.3.

6.6.3 Combination of in-situ results

The results from the Z balance and track-jet in-situ validation studies are combined to produce final pile-up uncertainty terms, $\Delta\left(\frac{\partial p_T}{\partial N_{PV}}\right)$ and $\Delta\left(\frac{\partial p_T}{\partial \langle \mu \rangle}\right)$, which quantify the uncertainty on the N_{PV} and $\langle \mu \rangle$ dependence of the reconstructed jet p_T. Track-jet measurements provide a fully decoupled measure of the N_{PV} and $\langle \mu \rangle$ dependences, so the results from track-jet validation alone are used in the central region ($|\eta| < 2.1$). Measurements of $\frac{\partial p_T}{\partial N_{PV}}$ and $\frac{\partial p_T}{\partial \langle \mu \rangle}$ are use to assign values to $\Delta\left(\frac{\partial p_T}{\partial N_{PV}}\right)$ and $\Delta\left(\frac{\partial p_T}{\partial \langle \mu \rangle}\right)$ to form an envelope by taking the maximum of:

- The difference between MC and data;

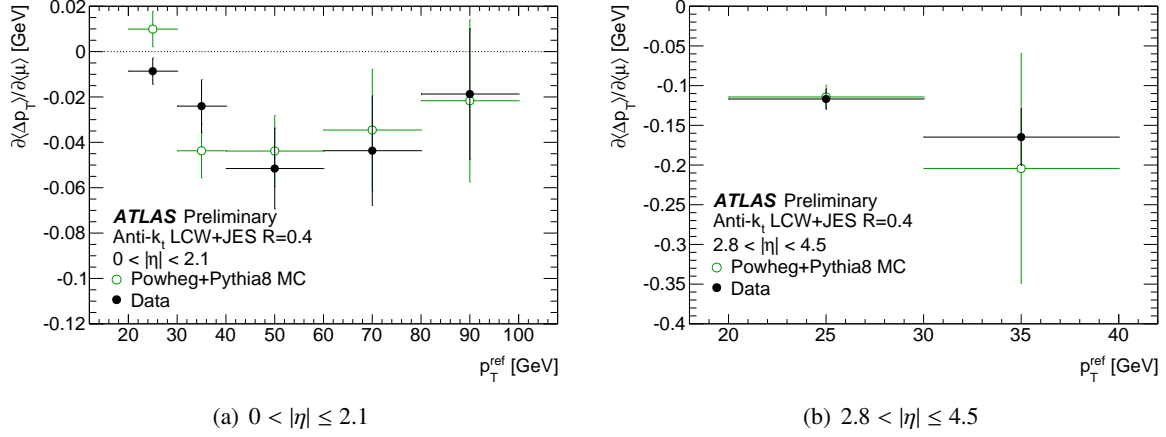- The statistical error on the difference between MC and data;

(a) $0 < |\eta| \leq 2.1$

(b) $2.8 < |\eta| \leq 4.5$

Figure 13: Validation results from Z+jet events: $\frac{\partial\langle\Delta p_T\rangle}{\partial\langle\mu\rangle}$ for anti-k_t, R=0.4, LCW in the central (left) and forward (right) region. The points represent central values of the linear fit to $\frac{\partial\langle\Delta p_T\rangle}{\partial\langle\mu\rangle}$ and the error bars correspond to the associated fitting error.

- The non-closure observed in MC;

- The non-closure observed in data.

The statistical error is considered in the envelope to provide a conservative estimate of our level of knowledge of the N_{PV} and $\langle\mu\rangle$ dependence of jet p_T in the event that both closure and data/MC agreement are observed. In this way the statistical accuracy of the test is included as the limit of our knowledge on the pile-up dependence of jet p_T when only a single in-situ measurement is available. In the future we would look to incorporate non-closure or biases as a further correction step or introduce additional methods of in-situ validation to further reduce the uncertainties.

In the more forward regions of the detector with $2.1 < |\eta| < 2.8$ and $2.8 < |\eta| < 4.5$, only the Z balance technique provides validation but suffers from not being binned exclusively in N_{PV} and $\langle\mu\rangle$, meaning that the effects of in-time and out-of-time pile-up cannot be fully decoupled. For these regions the N_{PV} uncertainty is assumed to be η-independent and so extended forward from the central bin. Z balance results are used to assign forward values to $\Delta\left(\frac{\partial p_T}{\partial\langle\mu\rangle}\right)$ by comparing forward and central η bins so that any η dependence is fully encapsulated by the $\langle\mu\rangle$ dependent uncertainty. Analogous to the above derivation of uncertainties from track-jet studies, forward $\langle\mu\rangle$ uncertainties are defined as the maximum of a non-closure term and a data/MC bias term. The data/MC bias term is defined as

$$\sigma_{\text{bias}} = \left| \left(\frac{\partial(p_T - p_T^{\text{ref}})}{\partial\langle\mu\rangle}^{\text{forward}}_{\text{data}} - \frac{\partial(p_T - p_T^{\text{ref}})}{\partial\langle\mu\rangle}^{\text{forward}}_{\text{MC}} \right) - \left(\frac{\partial(p_T - p_T^{\text{ref}})}{\partial\langle\mu\rangle}^{\text{central}}_{\text{data}} - \frac{\partial(p_T - p_T^{\text{ref}})}{\partial\langle\mu\rangle}^{\text{central}}_{\text{MC}} \right) \right| \quad (7)$$

and the non-closure term as

$$\sigma_{\text{closure}} = \max\left(\left| \frac{\partial(p_T - p_T^{\text{ref}})}{\partial\langle\mu\rangle}^{\text{forward}}_{\text{data}} - \frac{\partial(p_T - p_T^{\text{ref}})}{\partial\langle\mu\rangle}^{\text{central}}_{\text{data}} \right|, \left| \frac{\partial(p_T - p_T^{\text{ref}})}{\partial\langle\mu\rangle}^{\text{forward}}_{\text{MC}} - \frac{\partial(p_T - p_T^{\text{ref}})}{\partial\langle\mu\rangle}^{\text{central}}_{\text{MC}} \right| \right). \quad (8)$$

The final value of $\Delta\left(\frac{\partial p_T}{\partial\langle\mu\rangle}\right)$ is then taken as the envelope of σ_{bias}, σ_{closure} and associated statistical uncertainties.

6.7 Final Uncertainties

As defined above $\Delta\left(\frac{\partial p_T}{\partial N_{PV}}\right)$ and $\Delta\left(\frac{\partial p_T}{\partial \langle\mu\rangle}\right)$ represent uncertainties on the N_{PV} and $\langle\mu\rangle$ dependence of the reconstructed jet p_T. To evaluate the effect of such uncertainties on final analyses the effects are propagated as systematic shifts of the reconstructed jet p_T. The shifts are defined relative to reference pile-up conditions given by the 2012 mean conditions of $N_{PV}^{ref} = 11.8$ and $\langle\mu\rangle^{ref} = 20.7$. These systematic shifts have opposite sign for events on opposite sides of the reference conditions and are given by:

1. The systematic shift due to the uncertainty on the N_{PV} dependence

$$\Delta_{p_T}^{N_{PV}} = \pm\Delta\left(\frac{\partial p_T}{\partial N_{PV}}\right) \times \left(N_{PV} - N_{PV}^{ref}\right). \tag{9}$$

2. The systematic shift due to the uncertainty on the $\langle\mu\rangle$ dependence

$$\Delta_{p_T}^{\langle\mu\rangle} = \pm\Delta\left(\frac{\partial p_T}{\partial \langle\mu\rangle}\right) \times \left(\langle\mu\rangle - \langle\mu\rangle^{ref}\right). \tag{10}$$

3. The p_T dependence on the N_{PV} and $\langle\mu\rangle$ terms of the residual correction comes from the same source; namely, the increase in signal occupancy at the jet core with increased jet p_T. Therefore the systematic shift derived from the p_T dependence is defined as a combination of the N_{PV} and $\langle\mu\rangle$ terms of the p_T dependence

$$\Delta_{p_T}^{p_T} = \pm\left(\alpha_{N_{PV}} \times \left(N_{PV} - N_{PV}^{ref}\right) + \alpha_{\langle\mu\rangle} \times \left(\langle\mu\rangle - \langle\mu\rangle^{ref}\right)\right) \times \log\left(\frac{p_T}{25\ \text{GeV}}\right). \tag{11}$$

This results in partial cancellation of the systematic shift for events with similar amounts of in-time and out-of-time pile-up.

Potential mis-modeling of the sample dependence introduced by the use of ρ for pile-up subtraction is also addressed by a systematic uncertainty. The difference between ρ in MC and data is evaluated in dijet, photon+jet [43] and Z+jet events. The dependence of $\langle\rho\rangle$ on $\langle\mu\rangle$ is evaluated using a linear fit. The differences in slope and intercept of the fit are then studied to evaluate the degree to which the simulation models the sample dependence of ρ. The effect of the slope is studied using

$$\Delta_S = \left(\frac{\partial\langle\rho\rangle}{\partial\langle\mu\rangle}\bigg|_{\text{data}}^{\text{channel 1}} - \frac{\partial\langle\rho\rangle}{\partial\langle\mu\rangle}\bigg|_{\text{data}}^{\text{channel 2}}\right) - \left(\frac{\partial\langle\rho\rangle}{\partial\langle\mu\rangle}\bigg|_{\text{MC}}^{\text{channel 1}} - \frac{\partial\langle\rho\rangle}{\partial\langle\mu\rangle}\bigg|_{\text{MC}}^{\text{channel 2}}\right) \tag{12}$$

where channel 1 and channel 2 denote all possible pairings of the three channels studied. Similarly the modelling of the intercept is evaluated as

$$\Delta_I = \left(\rho(20.7)_{\text{data}}^{\text{channel 1}} - \rho(20.7)_{\text{data}}^{\text{channel 2}}\right) - \left(\rho(20.7)_{\text{MC}}^{\text{channel 1}} - \rho(20.7)_{\text{MC}}^{\text{channel 2}}\right) \tag{13}$$

where $\rho(20.7)$ denotes the value of $\langle\rho\rangle$ for $\langle\mu\rangle = 20.7$ as taken from the linear fit.

Figure 14 and Table 1 show the $\langle\mu\rangle$ dependence of ρ in Z+jet events. Figure 15 shows the results of such a linear fit in Z, photon and dijet topologies. The dominant tension is observed in the intercept of the Z+jet selection in which both Monte Carlos predict a larger ρ value than observed in data. Two systematic uncertainties are considered using the average area of anti-k_t jets (πR^2) and the mean jet energy calibration factor at $p_T = 25$ GeV averaged over each $|\eta|$ bin, C^{JES}, to propagate $\max(\Delta_I)$ and $\max(\Delta_S)$ through to a systematic shift of the calibrated jet p_T. These are:

1. The systematic shift due to the uncertainty on the slope of the dependence of ρ on $\langle\mu\rangle$ using

$$\Delta_{p_T}^{S} = \pm\max(\Delta_S) \times C^{\text{JES}} \times \pi R^2 \times \left(\langle\mu\rangle - \langle\mu\rangle^{ref}\right). \tag{14}$$
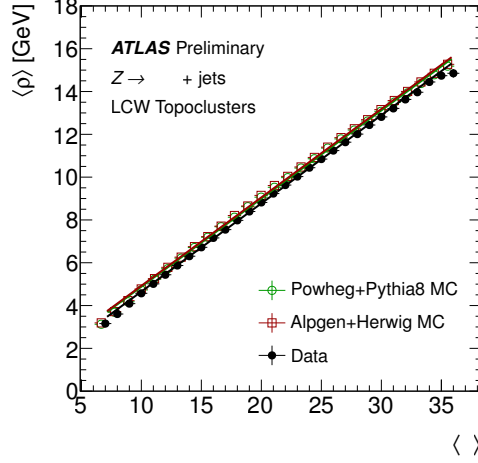
Figure 14: The dependence of ρ on $\langle\mu\rangle$ in Z events for data and two different generators. The gradients and intercepts of the linear fits are shown in Table 1.

Generator	Fit Parameter	Value
Powheg+Pythia8	$\partial\langle\rho\rangle/\langle\mu\rangle$	0.4119 ± 0.0003
	$\rho(\langle\mu\rangle = 20.7)$	8.998 ± 0.007
Alpgen+Herwig	$\partial\langle\rho\rangle/\langle\mu\rangle$	0.4131 ± 0.0003
	$\rho(\langle\mu\rangle = 20.7)$	9.053 ± 0.008
Data	$\partial\langle\rho\rangle/\langle\mu\rangle$	0.4121 ± 0.0003
	$\rho(\langle\mu\rangle = 20.7)$	8.773 ± 0.005

Table 1: The fit parameters extracted from the linear fits to $\langle\rho\rangle$ vs. $\langle\mu\rangle$ in Figure 14.

2. The systematic shift due to the uncertainty on the value of ρ using

$$\Delta_{p_{\mathrm{T}}}^{I} = \pm \max(\Delta_I) \times C^{\mathrm{JES}} \times \pi R^2. \tag{15}$$

The $\Delta_{p_{\mathrm{T}}}^{S}$ and $\Delta_{p_{\mathrm{T}}}^{\langle\mu\rangle}$ systematic shifts have the same functional dependence on $\langle\mu\rangle$. It is seen that $\Delta_{p_{\mathrm{T}}}^{\langle\mu\rangle}$ is dominant in all cases so $\Delta_{p_{\mathrm{T}}}^{S}$ is not considered further. The $\Delta_{p_{\mathrm{T}}}^{I}$ term is dominated by the tension observed in Z+jet events and included in the final pile-up uncertainty as a conservative estimate of topology dependence. Analyses particularly affected by this uncertainty will rederive an analysis specific topology dependence using their particular samples and event selection to reduce this conservative uncertainty.

Figures 16 and 17 show the resulting systematic shifts for $p_{\mathrm{T}} = 25$ GeV and $p_{\mathrm{T}} = 40$ GeV anti-k_t, R=0.4, LCW jets as a function of N_{PV} and $\langle\mu\rangle$ for positive variations of the uncertainty terms. In these figures the N_{PV}, $\langle\mu\rangle$ and p_{T} terms are seen to be odd above and below the reference values. The p_{T} term is seen to be 0 at $p_{\mathrm{T}} = 25$ GeV as this is the point at which the residual correction is defined whilst it becomes non-zero for $p_{\mathrm{T}} = 40$ GeV. Furthermore the p_{T} term shows anticorrelation between the $\langle\mu\rangle$ and N_{PV} dependence which leads to partial cancellation of this systematic when propagated through an analysis. Figures 18 through 23 show the same systematic shifts for other jet collections used in ATLAS.

The four components of the jet pile-up systematics are defined as independent. To properly account for all correlations in the systematic shifts between events, particularly conserving the anti-correlation between events either side of the reference pile-up conditions, all four components must be propagated individually through an analysis.

21

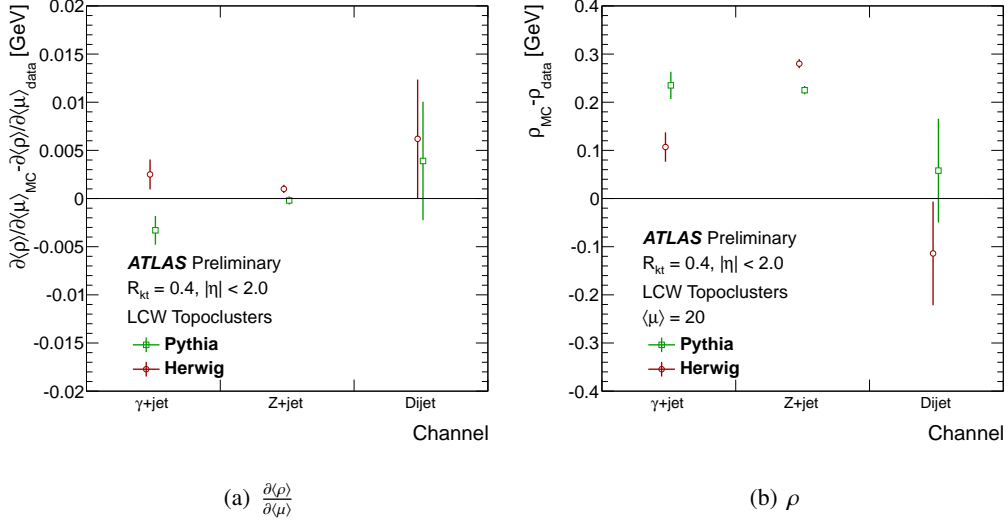(a) $\frac{\partial \langle \rho \rangle}{\partial \langle \mu \rangle}$ (b) ρ

Figure 15: The difference between data and MC in the fits demonstrated in Figure 14 for three different processes, Z+jet, photon+jet and dijet production shown for data and two different generators. (a) shows the differences in the slope of the fits and (b) shows the differences in the value of ρ at $\langle \mu \rangle = 20.7$. The greatest tension is observed in the Z+jet selection where both Monte Carlos predict a larger ρ than observed in data and do not agree with one another. This tension drives the uncertainty derived from these results.

7 Suppression of jets originating from pile-up

The number of reconstructed jets tends to increase with the pile-up activity, as shown in Fig. 24. Pile-up subtraction based on jet areas removes the majority of pile-up jets by shifting their p_T below threshold. This has the effect of improving the level of agreement between recorded data and MC simulation; the phenomenon of pile-up jets is in general poorly modeled. However, some pile-up jets remain even after pile-up subtraction, due to localized fluctuations in pile-up activity. Information from the tracks associated to each jet may be used to further reject any jets not originating from the hard-scatter interaction.

7.1 The jet vertex fraction

The jet vertex fraction (JVF) is a variable used in ATLAS to identify the origin vertex of a given jet. A cut in the JVF variable can help to filter jets coming from additional pp collisions in the event. Using tracks reconstructed from the ID information, the JVF variable can be defined for each jet with respect to each identified primary vertex (PV) in the event, by identifying the PV of origin of charged particle tracks pointing towards the given jet . Once the hard-scatter PV is selected, the JVF variable can be used to select jets having a high likelihood of originating from that vertex. Tracks are associated to calorimeter jets following the ghost association procedure. Then, the JVF is calculated as the ratio of the sum of transverse momentum of matched tracks that originate from a chosen PV to the sum of transverse momentum of all matched tracks in the jet, independently of their origin.

JVF is defined for each jet with respect to each PV. For a given jet$_i$, its JVF with respect to the primary vertex PV$_j$ is given by:

$$\text{JVF}(\text{jet}_i, \text{PV}_j) = \frac{\sum_k p_T(\text{track}_k^{\text{jet}_i}, \text{PV}_j)}{\sum_n \sum_l p_T(\text{track}_l^{\text{jet}_i}, \text{PV}_n)}, \tag{16}$$
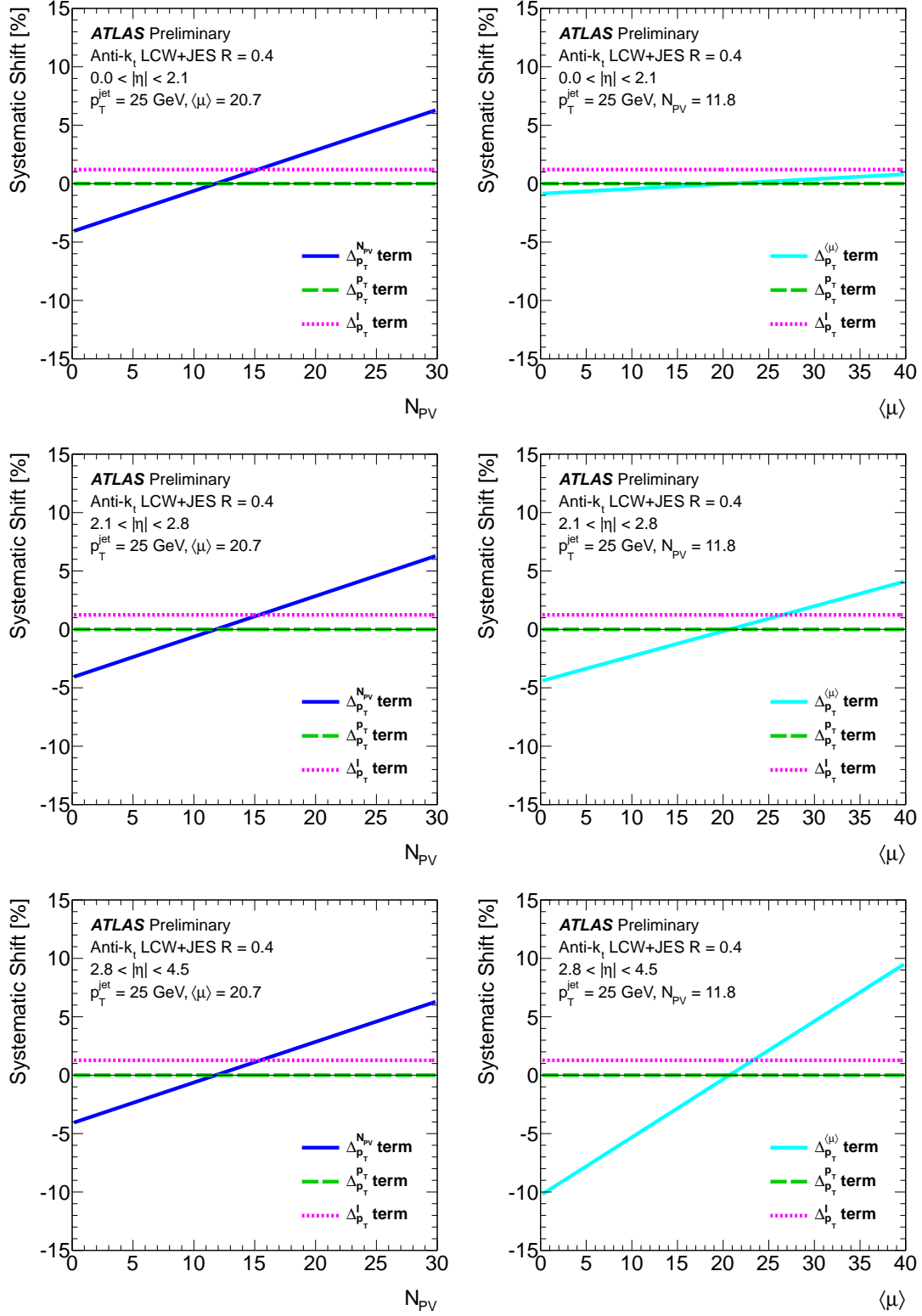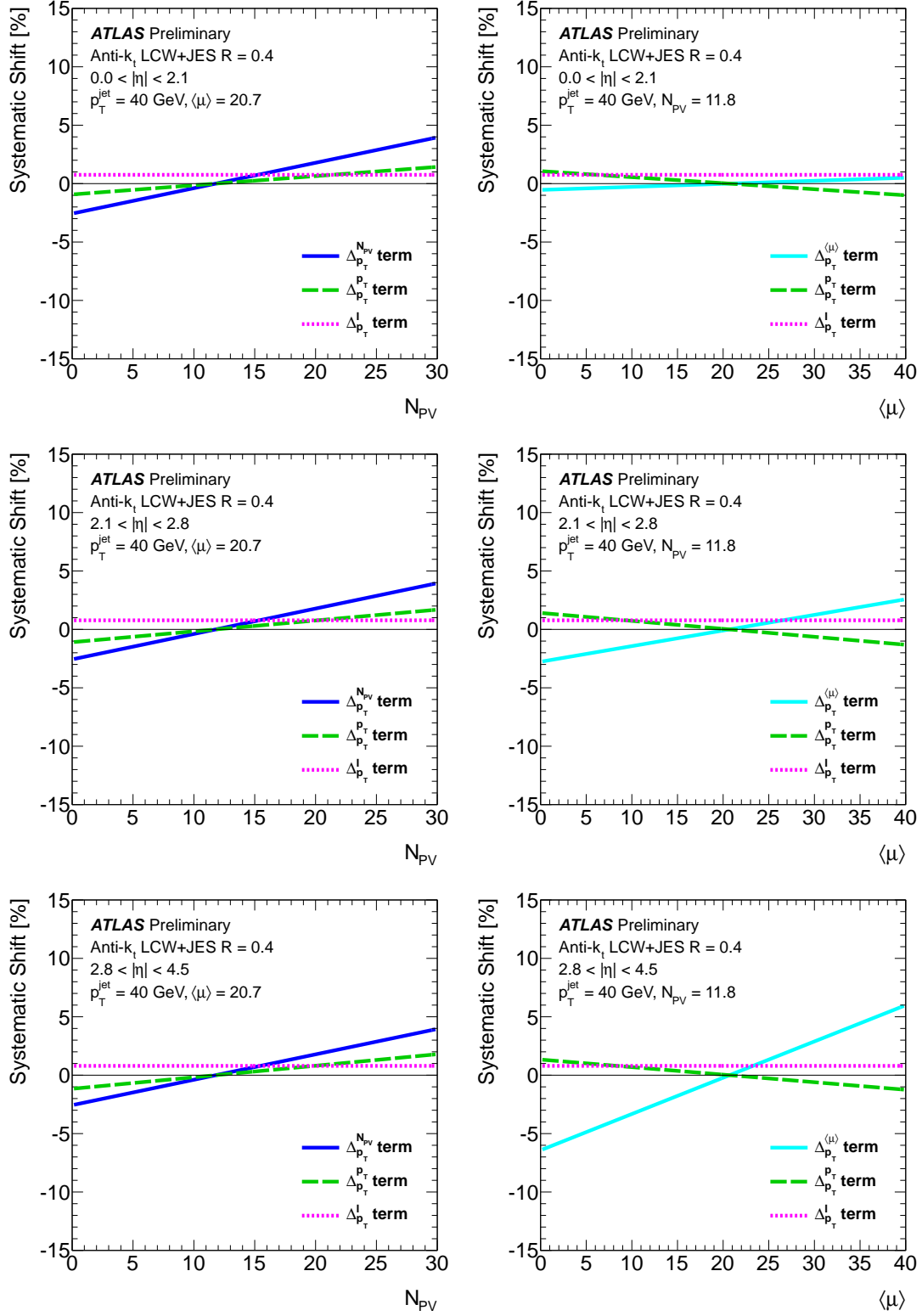
Figure 16: Systematic shifts arising from the four jet pile-up uncertainty terms for anti-k_t, $R = 0.4$, LCW jets shown as a function of N_{PV} (left) and $\langle\mu\rangle$ (right). The N_{PV} dependence is evaluated for $\langle\mu\rangle = 20.7$ and the $\langle\mu\rangle$ dependence for $N_{PV} = 11.8$. Plots are shown for $p_T = 25$ GeV jets in the three $|\eta|$ regions considered. Only the shifts associated to a positive variation of the uncertainty terms are shown.
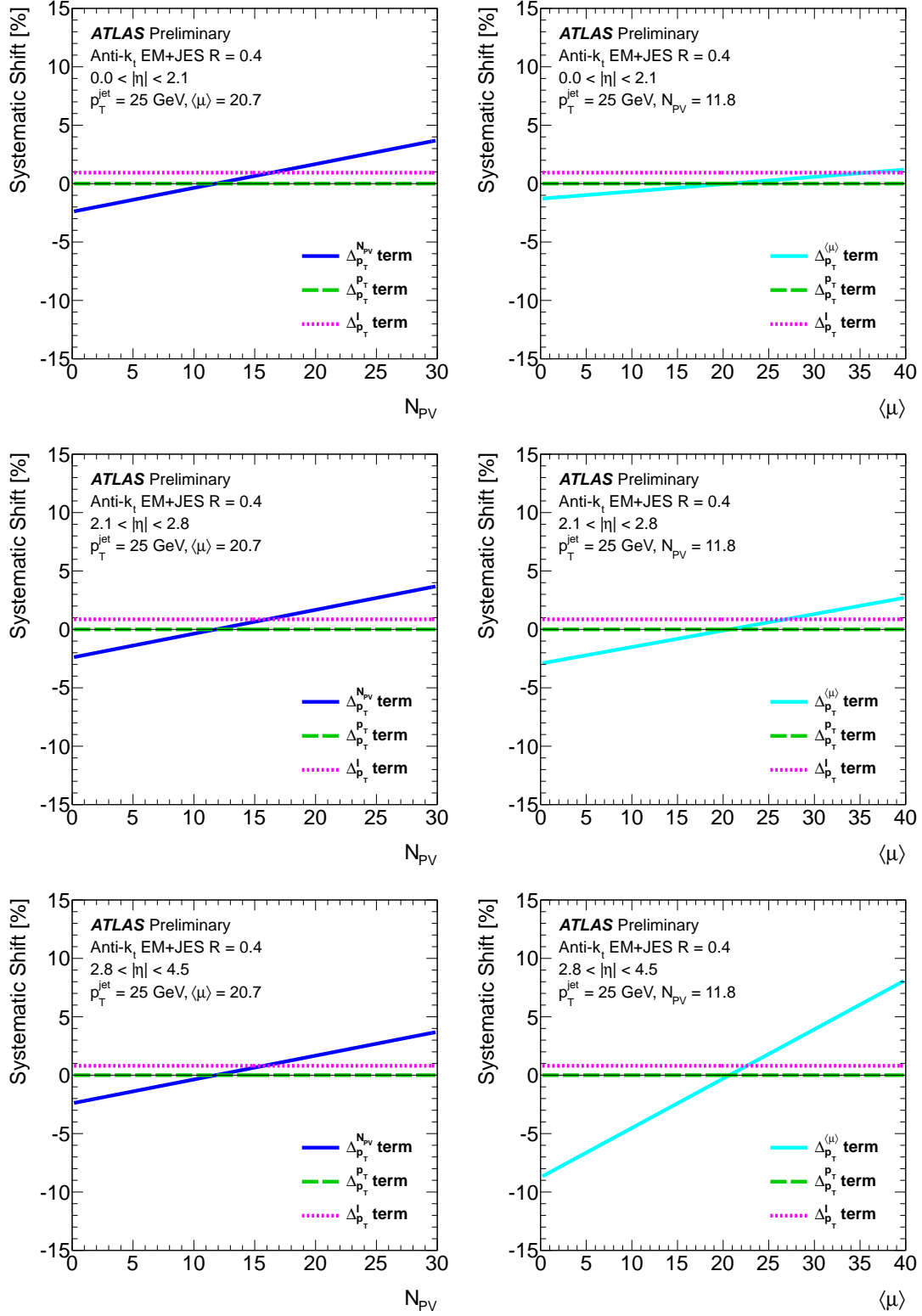
Figure 17: Systematic shifts arising from the four jet pile-up uncertainty terms for anti-k_t, $R = 0.4$, LCW jets shown as a function of N_{PV} (left) and $\langle \mu \rangle$ (right). The N_{PV} dependence is evaluated for $\langle \mu \rangle = 20.7$ and the $\langle \mu \rangle$ dependence for $N_{PV} = 11.8$. Plots are shown for $p_T = 40$ GeV jets in the three $|\eta|$ regions considered. Only the shifts associated to a positive variation of the uncertainty terms are shown.
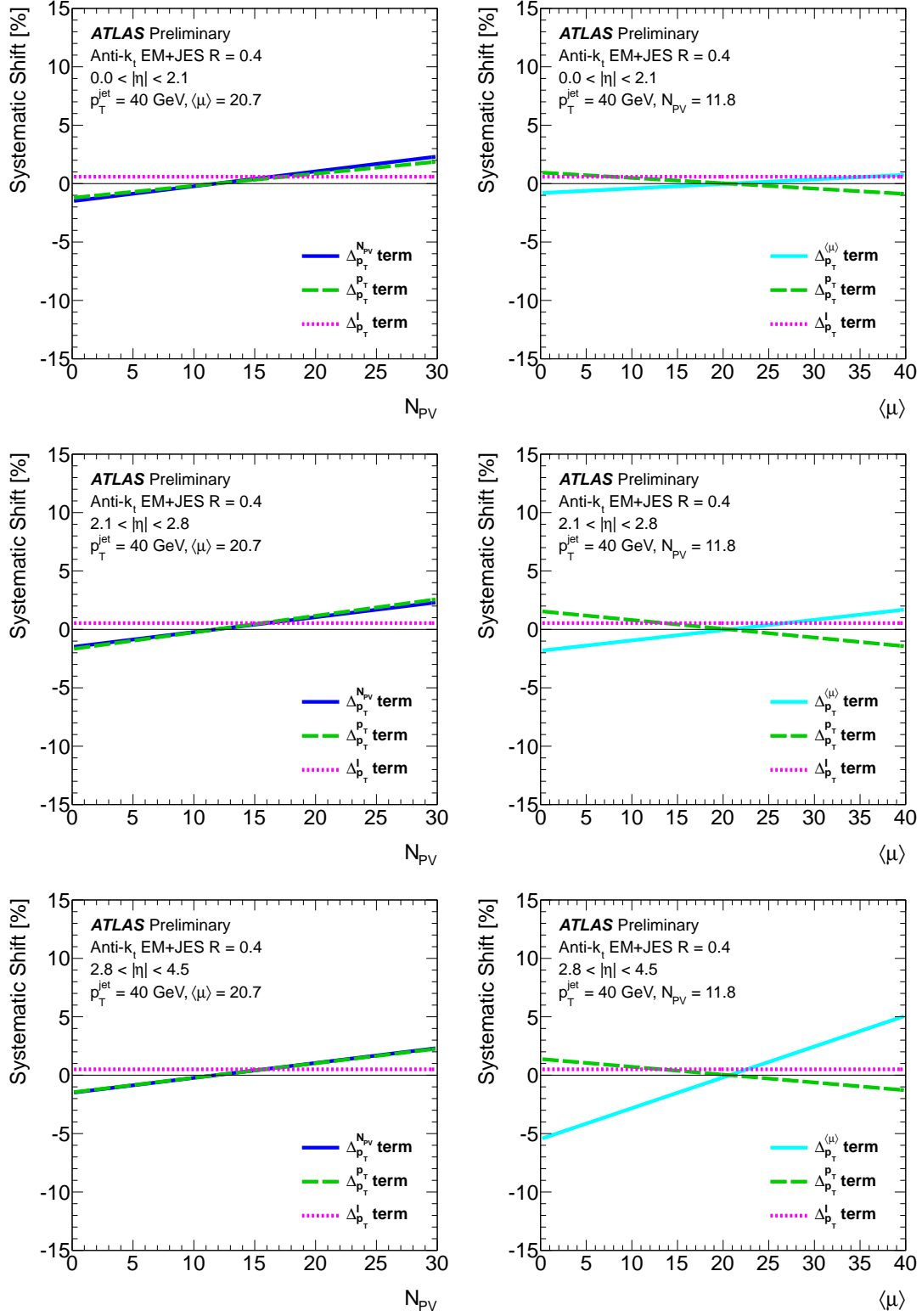
Figure 18: Systematic shifts arising from the four jet pile-up uncertainty terms for anti-k_t, $R = 0.4$, EM jets shown as a function of N_{PV} (left) and $\langle\mu\rangle$ (right). The N_{PV} dependence is evaluated for $\langle\mu\rangle = 20.7$ and the $\langle\mu\rangle$ dependence for $N_{PV} = 11.8$. Plots are shown for $p_T = 25$ GeV jets in the three $|\eta|$ regions considered. Only the shifts associated to a positive variation of the uncertainty terms are shown.
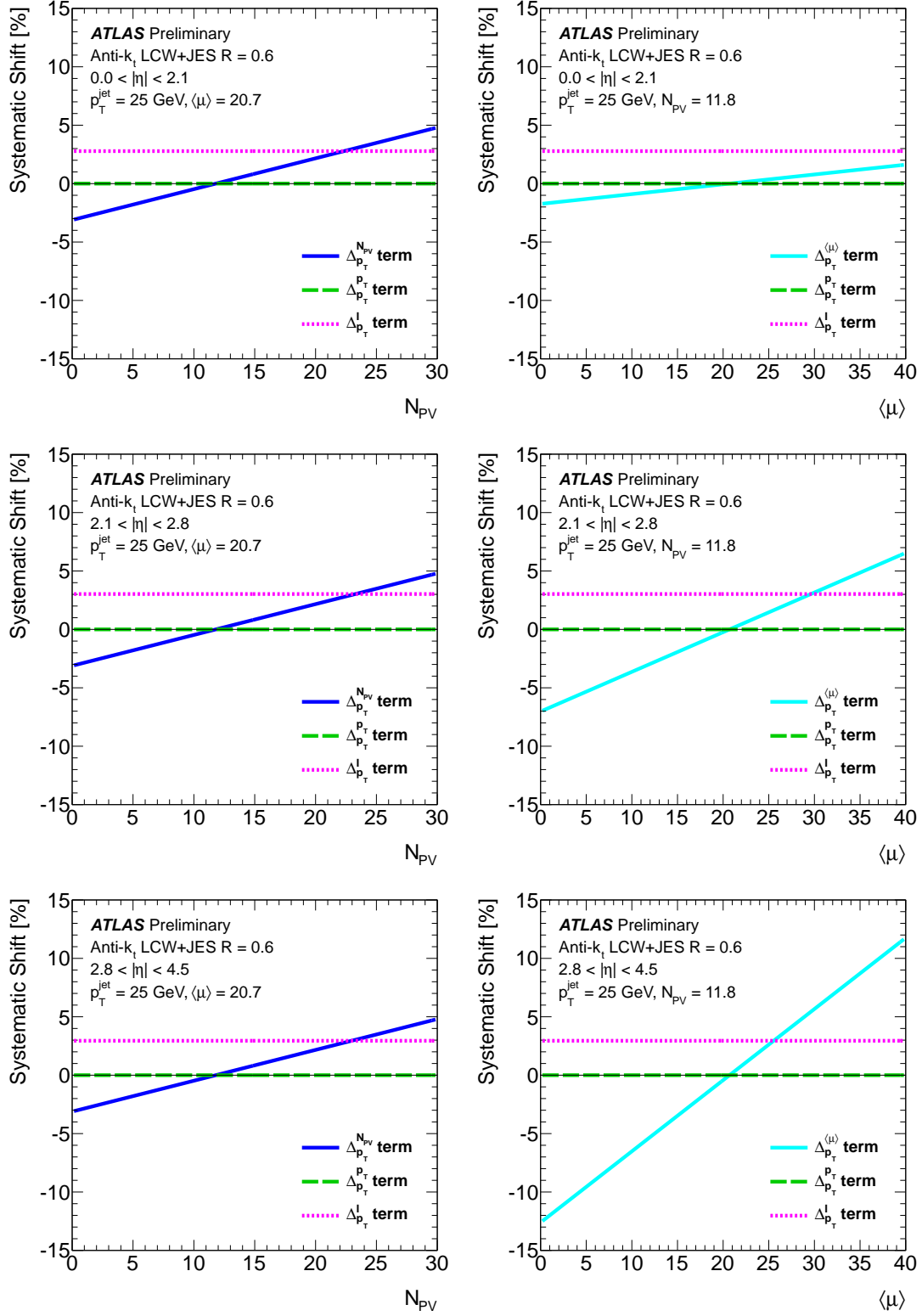
Figure 19: Systematic shifts arising from the four jet pile-up uncertainty terms for anti-k_t, $R = 0.4$, EM jets shown as a function of N_{PV} (left) and $\langle\mu\rangle$ (right). The N_{PV} dependence is evaluated for $\langle\mu\rangle = 20.7$ and the $\langle\mu\rangle$ dependence for $N_{PV} = 11.8$. Plots are shown for $p_T = 40$ GeV jets in the three $|\eta|$ regions considered. Only the shifts associated to a positive variation of the uncertainty terms are shown.
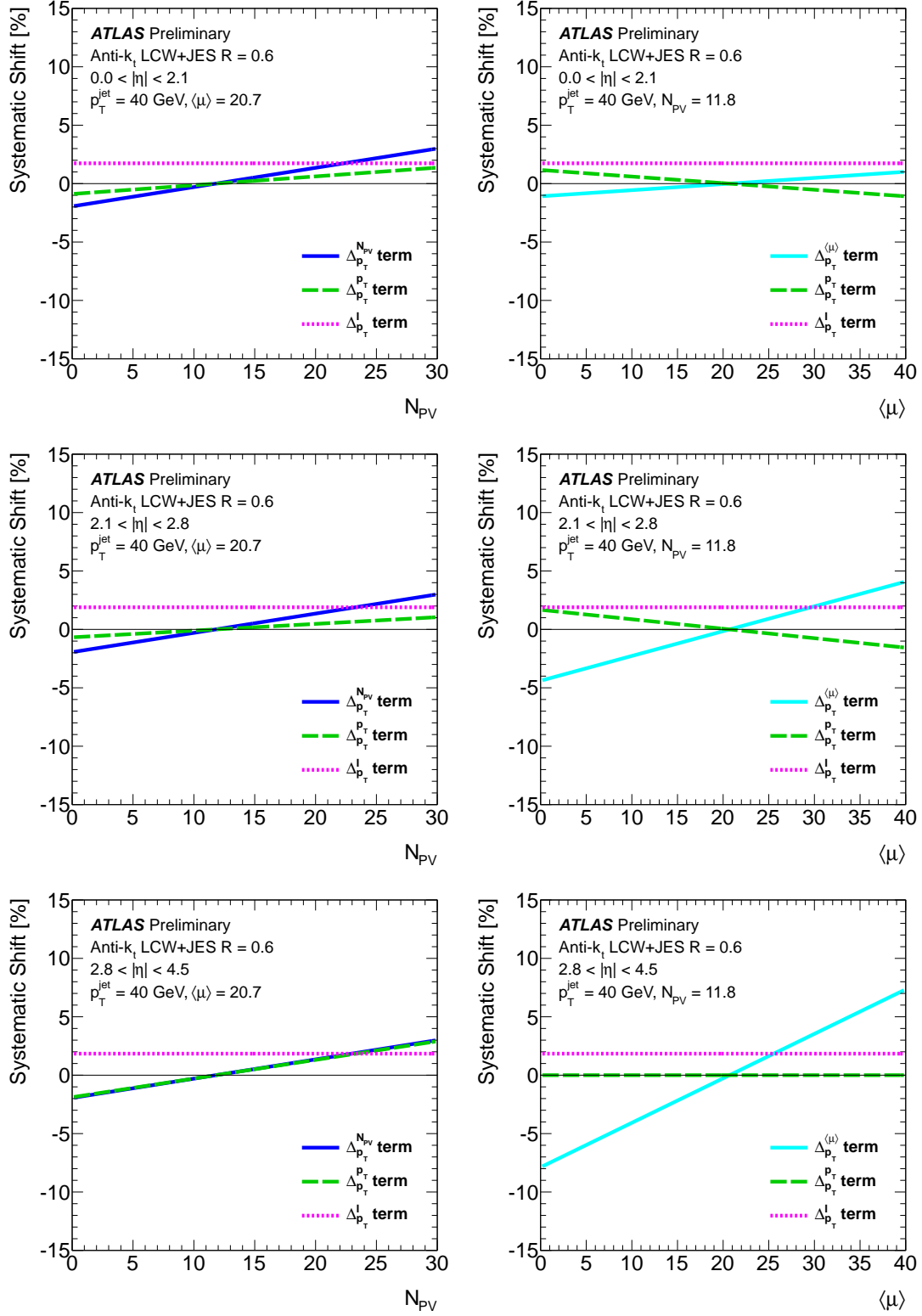
Figure 20: Systematic shifts arising from the four jet pile-up uncertainty terms for anti-k_t, $R = 0.6$, LCW jets shown as a function of N_{PV} (left) and $\langle\mu\rangle$ (right). The N_{PV} dependence is evaluated for $\langle\mu\rangle = 20.7$ and the $\langle\mu\rangle$ dependence for $N_{PV} = 11.8$. Plots are shown for $p_T = 25$ GeV jets in the three $|\eta|$ regions considered. Only the shifts associated to a positive variation of the uncertainty terms are shown.

Figure 21: Systematic shifts arising from the four jet pile-up uncertainty terms for anti-k_t, $R = 0.6$, LCW jets shown as a function of N_{PV} (left) and $\langle \mu \rangle$ (right). The N_{PV} dependence is evaluated for $\langle \mu \rangle = 20.7$ and the $\langle \mu \rangle$ dependence for $N_{PV} = 11.8$. Plots are shown for $p_T = 40$ GeV jets in the three $|\eta|$ regions considered. Only the shifts associated to a positive variation of the uncertainty terms are shown.
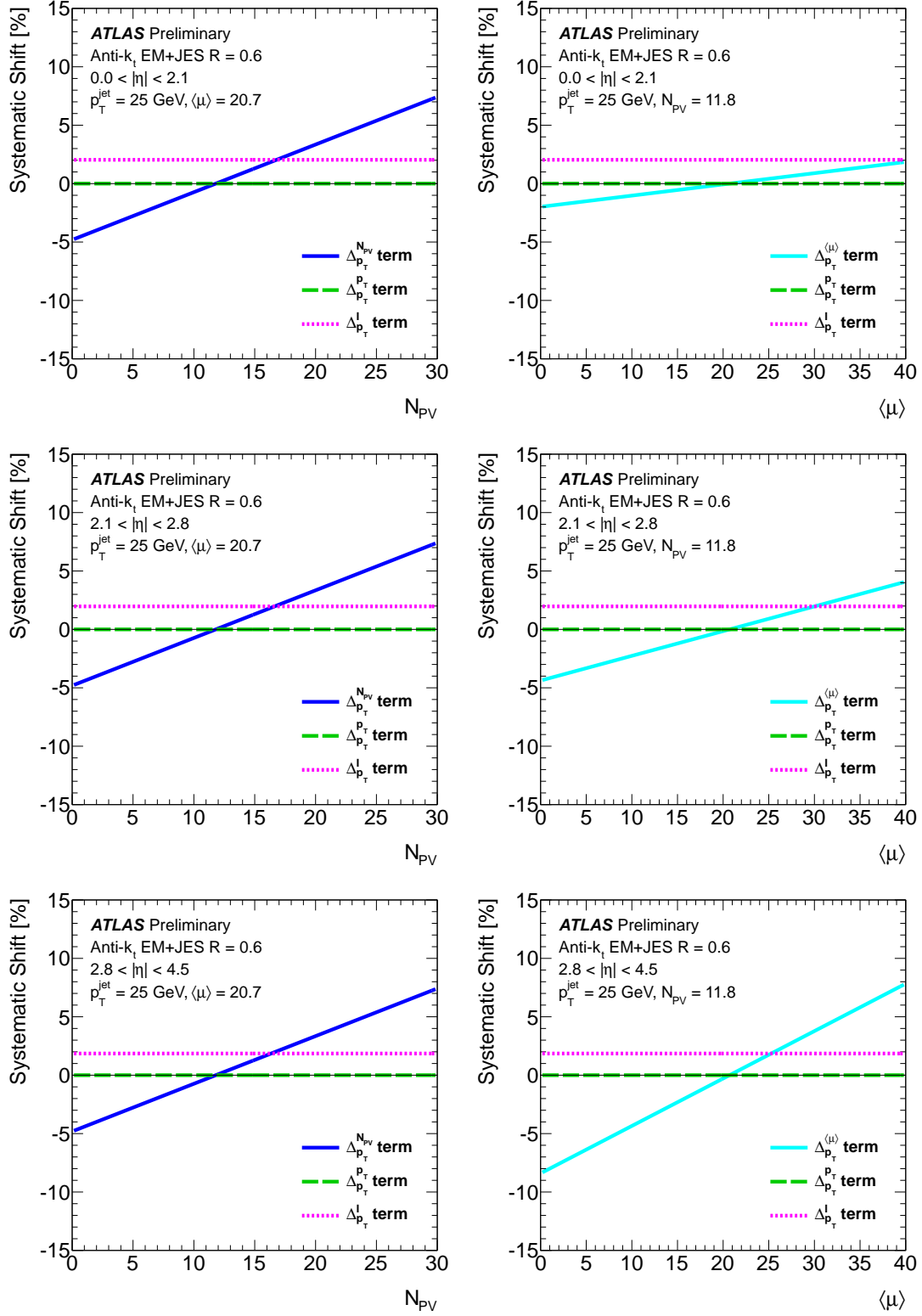
Figure 22: Systematic shifts arising from the four jet pile-up uncertainty terms for anti-k_t, $R = 0.6$, EM jets shown as a function of N_{PV} (left) and $\langle\mu\rangle$ (right). The N_{PV} dependence is evaluated for $\langle\mu\rangle = 20.7$ and the $\langle\mu\rangle$ dependence for $N_{PV} = 11.8$. Plots are shown for $p_T = 25$ GeV jets in the three $|\eta|$ regions considered. Only the shifts associated to a positive variation of the uncertainty terms are shown.
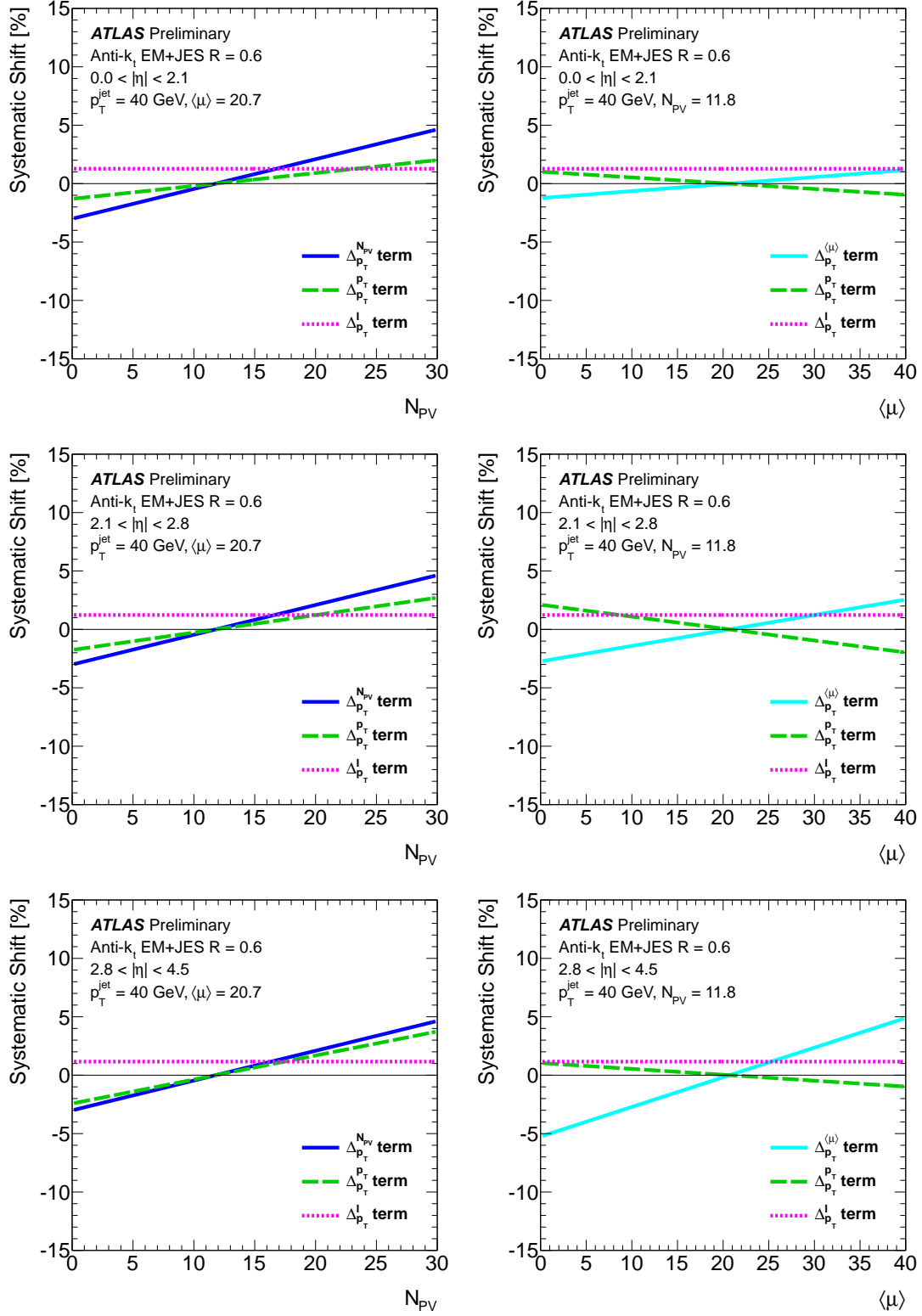
Figure 23: Systematic shifts arising from the four jet pile-up uncertainty terms for anti-k_t, $R = 0.6$, EM jets shown as a function of N_{PV} (left) and $\langle\mu\rangle$ (right). The N_{PV} dependence is evaluated for $\langle\mu\rangle = 20.7$ and the $\langle\mu\rangle$ dependence for $N_{PV} = 11.8$. Plots are shown for $p_T = 40$ GeV jets in the three $|\eta|$ regions considered. Only the shifts associated to a positive variation of the uncertainty terms are shown.
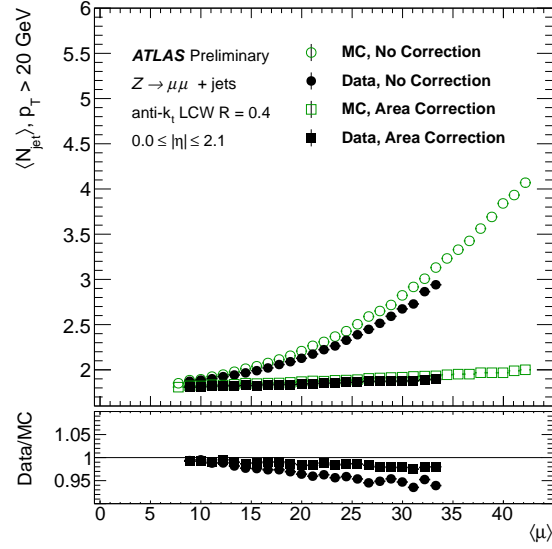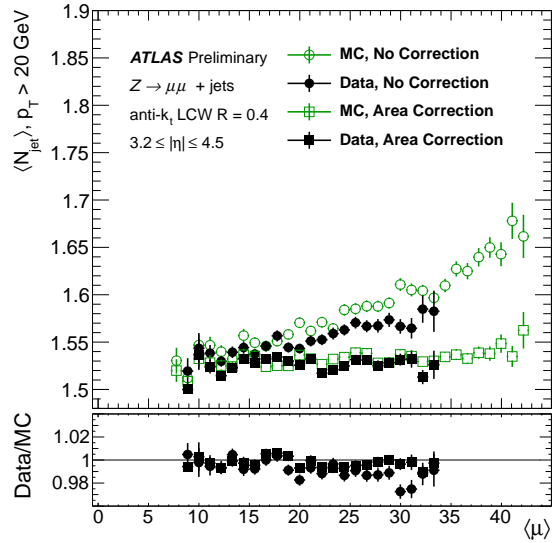
(a) Central Region
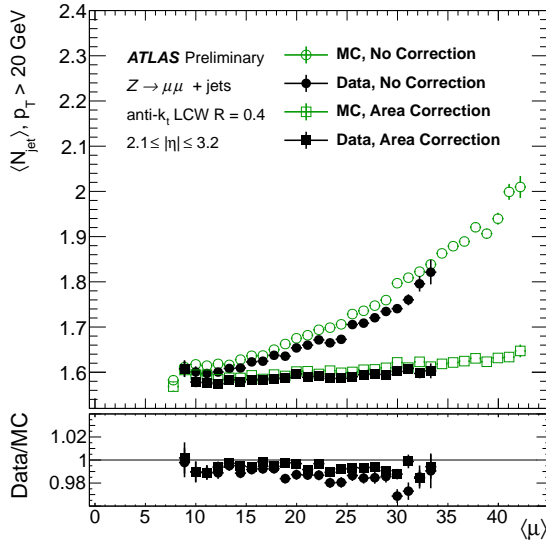
(b) Endcap Region

(c) Forward Region

Figure 24: The mean anti-k_t, R=0.4, LCW jet multiplicity against $\langle\mu\rangle$ in Z + jet events for jets with $p_T >$ 20 GeV and $|\eta| < 2.1$, $2.1 < |\eta| < 3.2$ and $3.2 < |\eta| < 4.5$ requiring at least 1 jet both before and after the application of the pile-up correction.

where k runs over all tracks originating from PV$_j$[8] matched to jet$_i$, n over all primary vertices in the event and l over all tracks originating from PV$_n$ matched to jet$_i$. Only tracks with $p_\mathrm{T} > 500$ MeV are considered in the JVF calculation.

For the purposes of this note, JVF will be defined with respect to the event hard-scatter vertex, which is selected as the primary vertex with the highest $\sum_\mathrm{tracks}(p_\mathrm{T}^2)$. In the Z+jets events used for these studies of pile-up suppression, this vertex selection criteria was found to be correct in at least 98% of events. JVF may be thus interpreted as an estimate of the fraction of energy in the jet that can be associated with the hard-scatter interaction. The principle of the JVF variable is shown schematically in Fig. 25.
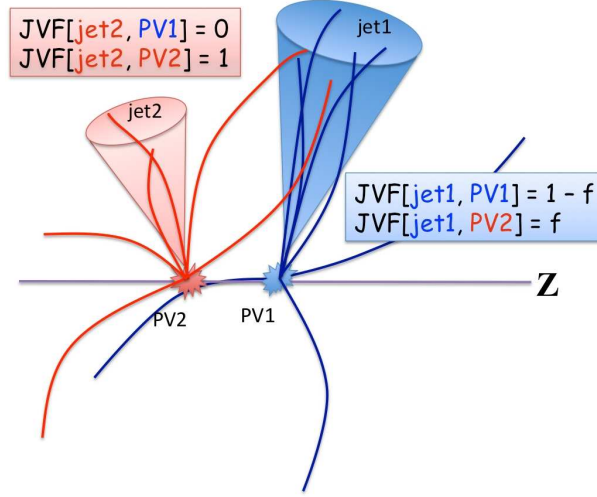

Figure 25: Schematic representation of the JVF principle.

Figure 26 shows the JVF distribution for hard-scatter jets and for pile-up jets with $p_\mathrm{T}^\mathrm{jet} > 20$ GeV after pile-up subtraction and JES correction in a $Z(\to ee)$+jets sample. It shows the discriminating power of the JVF variable. In Monte Carlo simulation, hard-scatter and pile-up jets are defined by ΔR association (see Sec. 4.4) to truth jets[9] as follows:

- *Hard-scatter jets*: calorimeter jets matched to truth jets from the hard-scatter ($\Delta R \leq 0.4$).

- *Pile-up jets*: calorimeter jets not matched to truth jets from the hard-scatter ($\Delta R > 0.4$).

Four distinct populations can be distinguished in the JVF distribution:

- JVF = −1 is assigned to calorimeter jets which do not have associated tracks.

- JVF = 0 indicates that all associated tracks originate from pile-up vertices.

- 0 < JVF < 1 indicates that some associated tracks originate from the hard-scatter vertex, while others come from pile-up.

- JVF = 1 indicates that all associated tracks come from the hard scatter.

While JVF is highly correlated with the actual fraction of hard-scatter activity in a reconstructed calorimeter jet, it is important to note that the correspondence is imperfect. For example, a jet with significant

---

[8]Tracks are associated with vertices by requiring $|\Delta z \times \sin\theta| < 1$ mm. In cases where more than one vertex satisfies this criterion, ambiguity is resolved by choosing the vertex of higher $\sum_\mathrm{tracks}(p_\mathrm{T}^2)$.

[9]Truth jets with $p_\mathrm{T} > 10$ GeV and $|\eta| < 5$ are considered.

neutral pile-up contributions may receive JVF = 1, while JVF = 0 may result from a fluctuation in the fragmentation of a hard-scatter jet such that its charged constituents all fall below the track p_T threshold. JVF also relies on the hard-scatter vertex being well separated from pile-up vertices. In some events, a pile-up jet may receive a high value of JVF because its origin interaction is very close to the hard-scatter interaction. While this effect is quite small in 2012 pile-up conditions, it will become more important at higher $\langle \mu \rangle$, as the average distance between interactions decreases as $1/\langle \mu \rangle$.
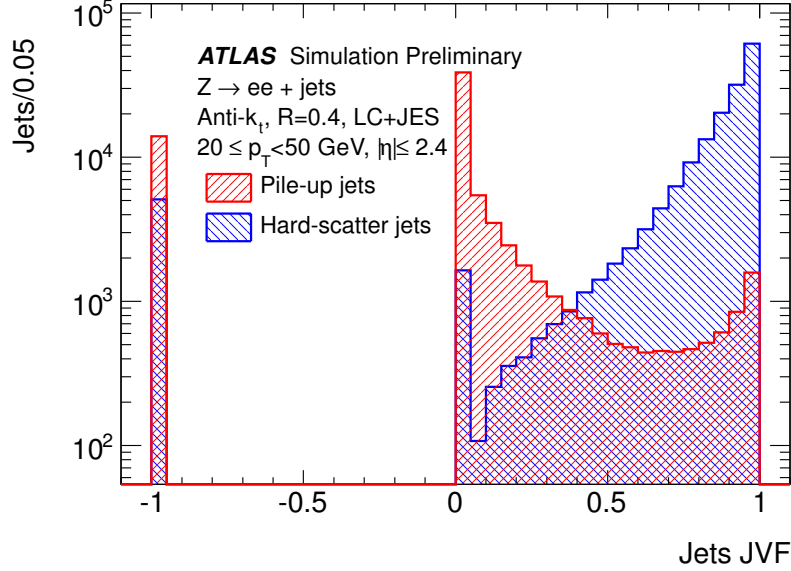

Figure 26: JVF distribution for hard-scatter (blue) and pile-up (red) jets with 20 GeV $\leq p_T^{\text{jet}} <$ 50 GeV and $|\eta| <$ 2.5 in simulated Z+jets events. Using JVF directly as a discriminating variable provides a way to separate both classes of jets.

7.2 Recommended JVF Cuts

In 2012, three JVF cuts are recommended for analyses in which pile-up jets are otherwise problematic. The loosest recommended cut is |JVF| > 0, which rejects only those jets that have zero matched tracks from the hard scatter. A somewhat tighter cut is |JVF| > 0.25, requiring that at least a quarter of all associated track p_T originates from the hard scatter, while the tightest recommended cut is |JVF| > 0.5. The cuts are applied to the *absolute value* of JVF, to avoid rejecting jets with zero matched tracks from any vertex.

Each analysis applying a JVF cut must choose an optimal cut value among the three recommended cuts, based on analysis-specific figures of merit. For example, given some definition of signal jets appropriate to an analysis, one could choose the JVF cut value that maximizes the signal jet efficiency divided by the rate for non-signal jets to pass the cut. Alternatively, one could choose the cut value that results in stability of jet multiplicity against pile-up, or devise a more sophisticated optimization procedure based on expected limits or measurement precision.

Figure 27 shows the dependence of JVF on the amount of pile-up, as characterized by the average number of interactions, $\langle \mu \rangle$. The denominator of JVF grows larger with increased pile-up, while the numerator remains unchanged. As a result, the optimal JVF cut value is expected to depend on pile-up conditions, which further emphasizes the need for analysis-specific JVF cut optimization. The ratio between data and MC is shown at the bottom.
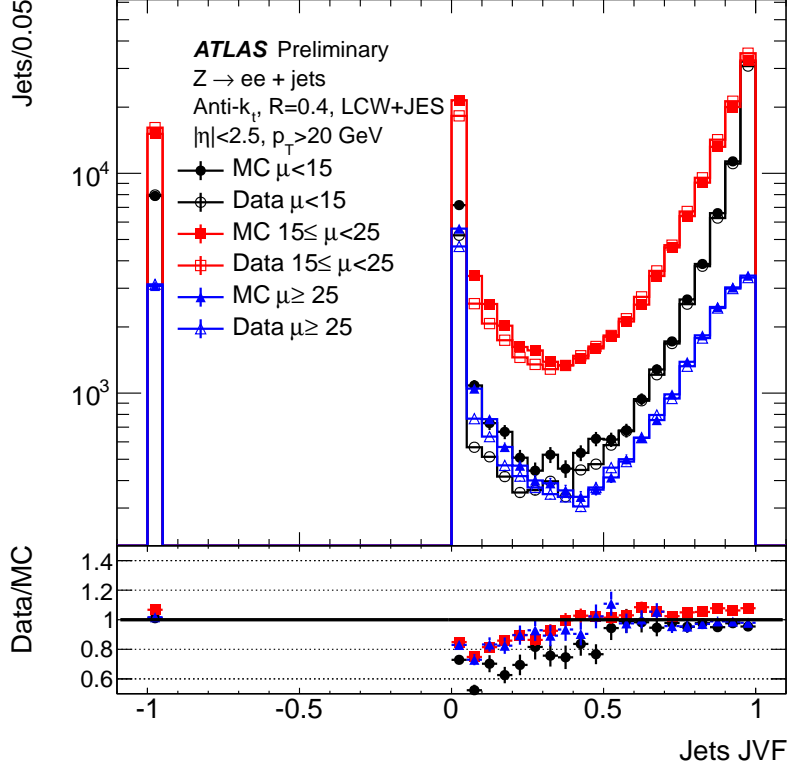
Figure 27: JVF distribution for different $\langle\mu\rangle$ bins for jets with $p_T^{\text{jet}} > 20$ GeV and $|\eta| < 2.5$. JVF is shifted closer to zero as the level of pile-up activity increases.

It is also necessary to define the kinematic range over which to apply the JVF cut. Figure 28 demonstrates that the number of jets with JVF = 0 increases significantly in the $2.1 \geq |\eta| < 2.8$ range, due to the hard-scatter core of the jet being located outside the ID acceptance while some part of the jet periphery remains inside, resulting in zero hard-scatter tracks but one or more pile-up tracks successfully associated to the jet. Figure 28 also shows that most of the jets with JVF = -1 are jets outside the ID acceptance. To avoid a loss in signal efficiency due to this effect, it is recommended to apply the JVF cut only to jets with $|\eta| < 2.4$. Additionally, studies performed in different topologies of di-jet, top and Z+jets events have shown that 99% of pile-up jets with $p_T^{\text{jet}} > 20$ GeV have $p_T < 50$ GeV. Therefore, applying a JVF cut above $p_T = 50$ GeV is not necessary in most analyses.

7.3 Modeling of JVF

The modelling of JVF is investigated in events where a boosted Z boson ($p_T^Z > 30$ GeV) and a jet ($|\eta| < 2.5$ and $p_T > 20$ GeV) were produced back-to-back ($\Delta\phi(Z, \text{leading jet}) > 2.9$), thus providing a nearly pure sample of hard-scatter jets (HS region). By comparison to truth jets in MC, it was found that the level of pile-up jet contamination in this sample is close to 2% at low p_T and almost zero at high p_T. The JVF distribution for the jet balanced against the Z boson in these events is shown in Fig. 29. It is observed that the JVF distribution is well modelled for hard-scatter jets.

However, the total jet multiplicity in these events is somewhat over-estimated in simulated events, due to mis-modeling of pile-up jets. This is shown in Fig. 30, for several different choices of minimum p_T cut applied at the fully calibrated LCW+JES scale (including jet-area based pile-up subtraction). The
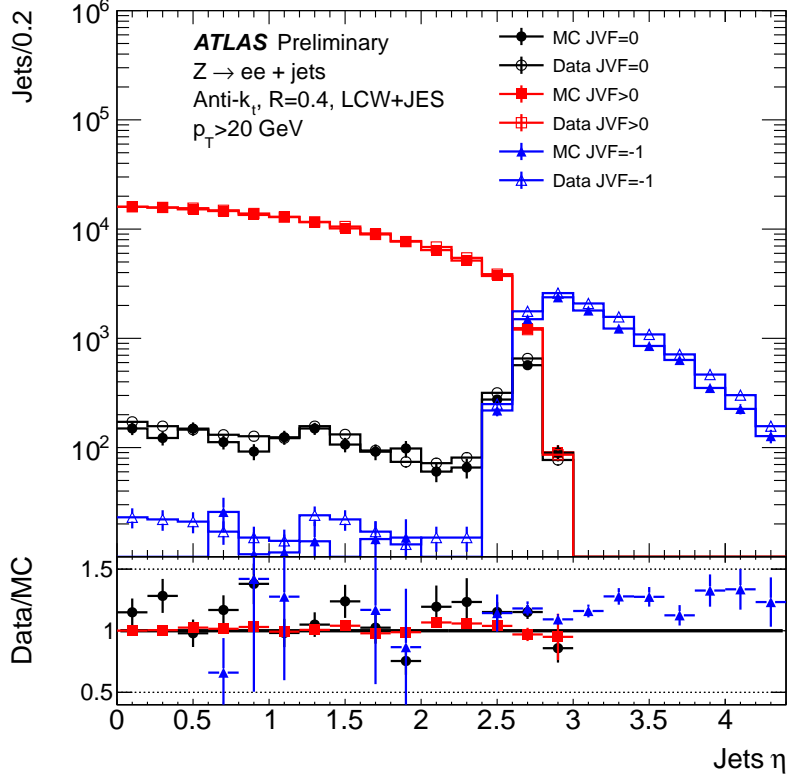
Figure 28: Jet η distribution for jets with $p_T^{jet} > 20\,\text{GeV}$ and JVF = 0 (black), positive (red) and JVF = -1 (blue) in Z+jet events with a jet well balanced against the Z boson candidate.

application of a JVF cut significantly improves the data/MC agreement, as the vast majority of pile-up jets fail the JVF cut: across all p_T bins, data and MC are seen to agree within 1% following the application of a JVF cut. It is also observed that the application of a JVF cut results in stability of the mean jet multiplicity against pile-up. Furthermore, for a minimum p_T of 50 GeV, the jet multiplicity is seen to be independent of pile-up even before a JVF cut, in both data and MC. This confirms the choice of 50 GeV as the maximum p_T up to which a JVF cut is recommended.

Fig. 30 also shows the systematic uncertainty bands. These uncertainties are based only on the extent to which the JVF distribution is mis-modeled for hard-scatter jets. The efficiency of a nominal JVF cut of X is defined as the fraction of jets, well balanced against the Z boson, passing the cut, denoted as $\mathcal{E}_{MC}^{nom}$ and $\mathcal{E}_{data}^{nom}$ for MC and data respectively. The systematic uncertainty is derived by finding two JVF cuts with $\mathcal{E}_{MC}$ differing from $\mathcal{E}_{MC}^{nom}$ by $\pm(\mathcal{E}_{MC}^{nom} - \mathcal{E}_{data}^{nom})$. The JVF uncertainty band is then formed by rerunning the analysis with these up and down variations in the JVF cut value.

Figure 31 shows $\mathcal{E}_{MC}$ and $\mathcal{E}_{data}$ for three different JVF operating points: 0.0, 0.25, 0.50 as a function of p_T and $\langle\mu\rangle$. Table 2 summarises the systematic JVF up/down cut variations for these three JVF operating points measured in the hard scatter enriched region. The variations are small, as expected from the agreement between data and simulation as observed in Fig. 29.

8 Summary and conclusion

We have demonstrated the implementation of a jet area based pile-up subtraction method in ATLAS. This method uses a direct measure of the pile-up activity in the calorimeter on an event-by-event basis (ρ) as
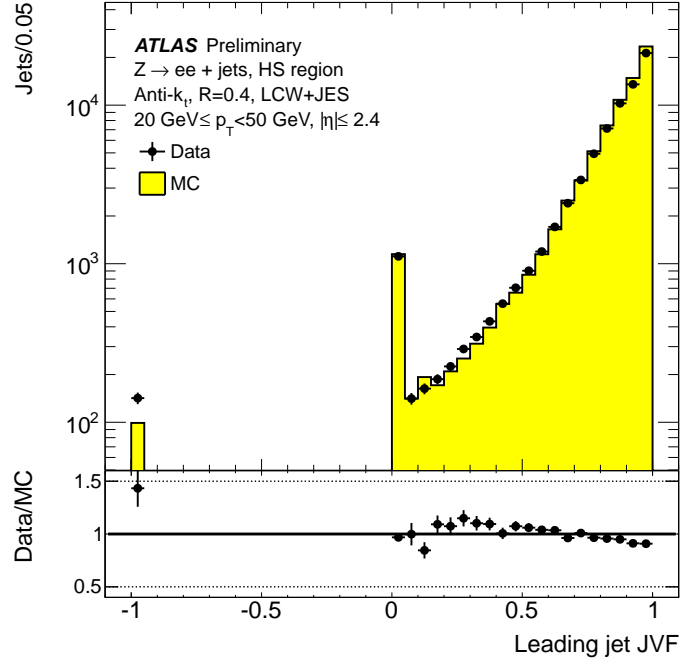
Figure 29: JVF distribution for jets (anti-k_t $R = 0.4$ LCW+JES) well balanced against $Z \to ee$ candidates, satisfying $20 < p_T^{jet} < 50$ GeV and $|\eta| < 2.4$. The distribution is strongly peaked at $JVF = 1$, and good data/MC agreement is observed.
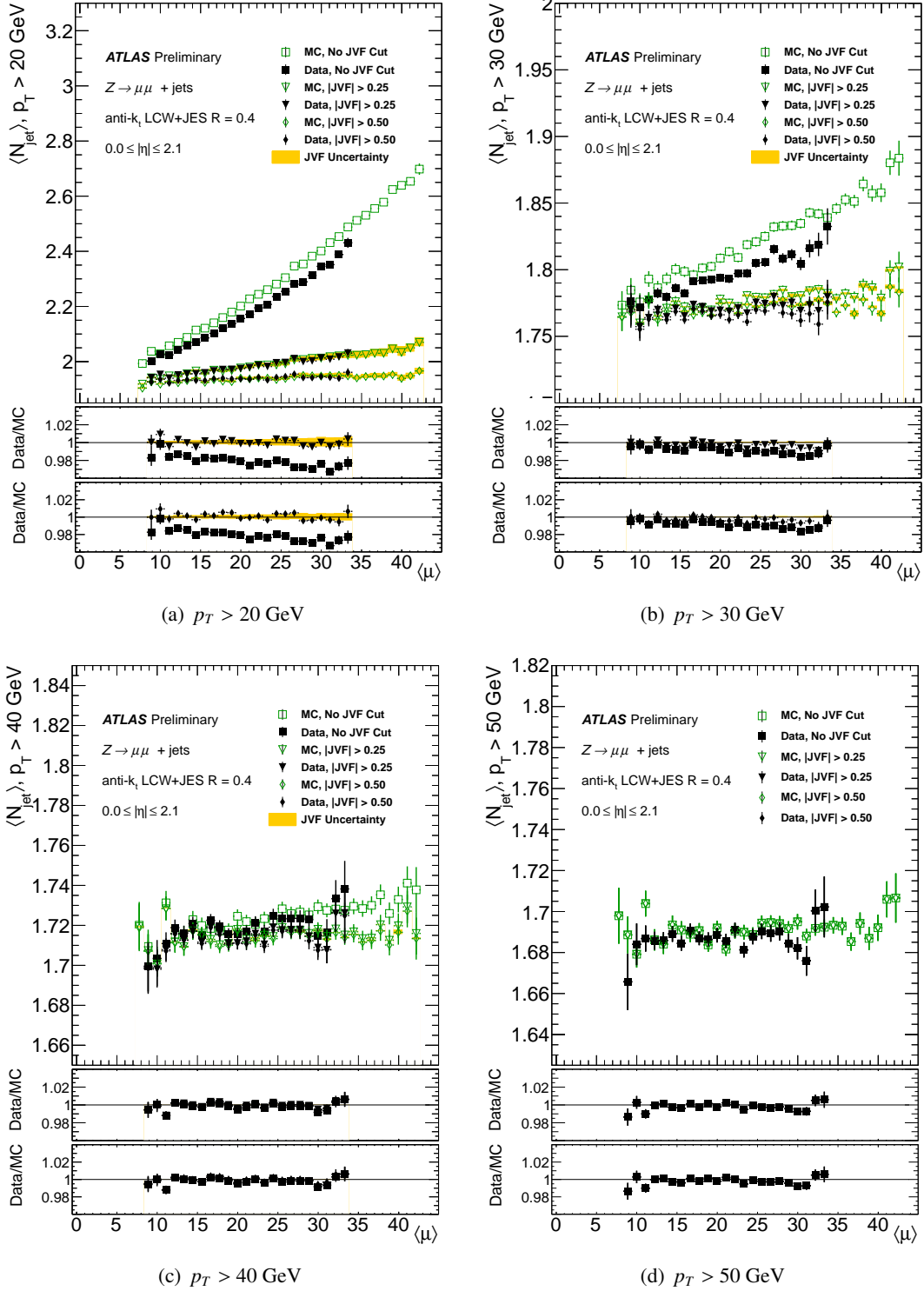
|JVF| > 0.00		|JVF| > 0.25		|JVF| > 0.50	
JVF up	JVF dw	JVF up	JVF dw	JVF up	JVF dw
0.1	-	0.28	0.21	0.53	0.47

Table 2: JVF up/down cut variations for three JVF operating points: 0.0, 0.25 and 0.50. Only jets with $|\eta| < 2.4$ and $20 < p_T < 50$ GeV are considered.

well as a jet-by-jet measure of pile-up susceptibility (the jet area, A). A residual pile-up correction is necessary to fully accommodate the impact of pile-up on jet p_T, as the high occupancy jet core contributes some extra sensitivity to both in-time and out-of-time pile-up, and the effects of pile-up on forward jets are not fully described by the median p_T density as calculated from topological clusters in the central calorimeter.

The combination of $\rho \cdot A$ subtraction and residual correction results in a stable jet p_T response across the full range of pile-up conditions in 2012, and it significantly reduces the degradation in jet p_T resolution associated with fluctuations in pile-up. It also improves the stability against pile-up of the jet multiplicity, shifting the majority of pile-up jets below the minimum jet p_T threshold. For $p_T > 50$ GeV, the pile-up subtraction procedure alone is sufficient to make the jet multiplicity stable vs pileup within statistical errors.

Systematic uncertainties have been estimated using two independent *in situ* validation techniques: Z + jet balance, and track-jet validation. Additional uncertainties are provided to cover mis-modeling in the observed topology dependence of ρ, as well as p_T dependence seen in the susceptibility of jets to

(a) $p_T > 20$ GeV

(b) $p_T > 30$ GeV

(c) $p_T > 40$ GeV

(d) $p_T > 50$ GeV

Figure 30: The mean anti-k_t, $R = 0.4$, LCW+JES jet multiplicity against $\langle \mu \rangle$ in Z+jet events for jets with $|\eta| < 2.1$ before and after selected $|JVF|$ cuts as applied to any jets with $p_T < 50$ GeV. Results for jets with $p_T > 20$, 30, 40 and 50 GeV are shown requiring at least one jet of that p_T. To remove effects of hard scatter modeling the dependence on $\langle \mu \rangle$ was fit and the MC shifted so that data and MC agree at zero pile-up, $\langle \mu \rangle = 0$. The upper ratio plots show before and after an application of a $|JVF|$ cut of 0.25 and the lower ratio plots show the same for a cut of 0.50. No JVF uncertainty is visible when counting jets with $p_T > 50$ as JVF is only applied to jets below 50 GeV.
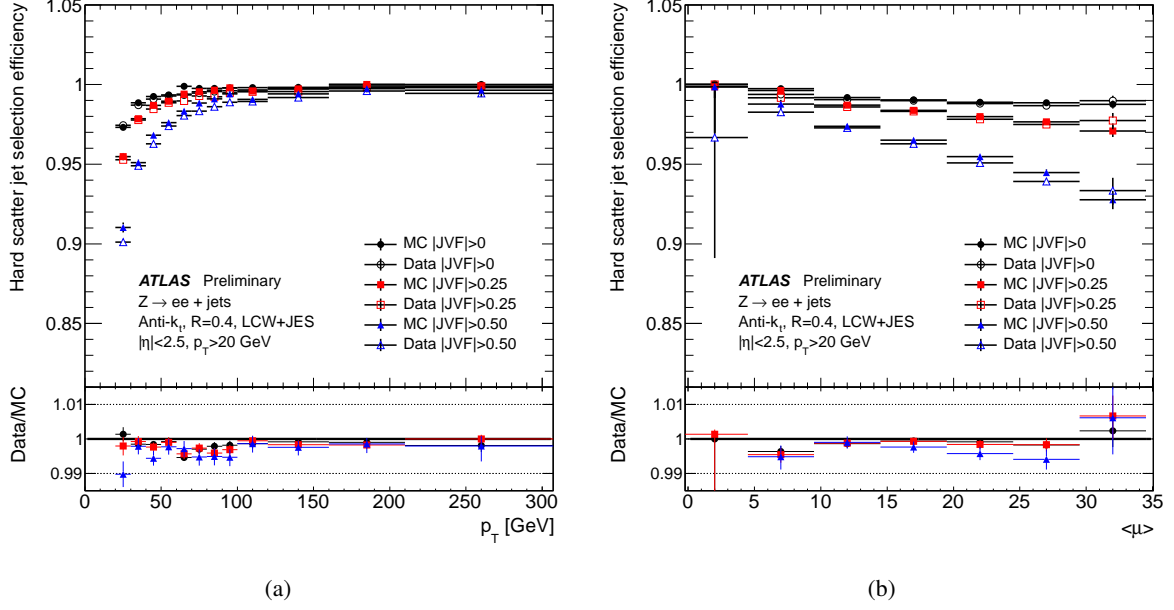
Figure 31: $\mathcal{E}_{\text{MC}}$ and $\mathcal{E}_{\text{data}}$ in $Z(\rightarrow ee)$+jets events as a function of (a) p_{T} and (b) $\langle\mu\rangle$. Three different JVF operating points: 0.0, 0.25, 0.50 are shown. Jets with a $p_{\text{T}}^{\text{jet}} > 20$ GeV and $|\eta| < 2.4$ are considered.

pile-up. All systematic uncertainties are presented as shifts in the fully calibrated jet p_{T} as a function of η and p_{T}, for several jet collections. While typically below 2% for $R = 0.4$ anti-k_t jets with $p_{\text{T}} > 40$ GeV in the central region of the calorimeters, they reach up to 5.6% at low p_{T} and higher η. In the future we will extend the uncertainty description, providing uncertainties on the uncertainties to enable sophisticated treatment of these systematic effects in analyses.

The use of the jet vertex fraction as a discriminating variable for rejection of pile-up jets was also presented, with recommended cut values and a kinematic range in which to apply the JVF cut. Systematic variations in the JVF cut value were suggested, based on modeling of the efficiency for hard-scatter jets to pass each cut in well-balanced Z+jet events. These systematic variations do not cover the observed overestimation of pile-up jet multiplicity in simulation, but the application of a JVF cut effectively removes this mis-modeling by rejecting the vast majority of pile-up jets.

References

[1] ATLAS Collaboration, *The ATLAS experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003.

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

[3] Y. L. Dokshitzer, G. Leder, S. Moretti, and B. Webber, *Better jet clustering algorithms*, JHEP **9708** (1997) 001, `arXiv:hep-ph/9707323 [hep-ph]`.

[4] M. Wobisch and T. Wengler, *Hadronization corrections to jet cross sections in deep- inelastic scattering*, `arXiv:hep-ph/9907280 [hep-ph]`.

[5] ATLAS Collaboration, *Jet energy measurement with the ATLAS detector in proton-proton collisions at $\sqrt{s}$ = 7 TeV*, accepted by EPJ (2011) , `arXiv:1112.6426 [hep-ex]`.

[6] ATLAS Collaboration, *Luminosity determination in p-p collisions at $\sqrt{s}$ = 7 TeV using the ATLAS detector at the LHC*, Eur. Phys. J. **C 71** (2011) 1630, `arXiv:1101.2185 [hep-ex]`.

[7] M. Cacciari and G. P. Salam, *Pileup subtraction using jet areas*, Phys. Lett. **B659** (2008) 119–126, `arXiv:0707.1378 [hep-ph]`.

[8] ATLAS Collaboration, *Pile-up corrections for jets from proton-proton collisions at $\sqrt{s}$ = 7 TeV in ATLAS in 2011*, ATLAS-CONF-2012-064.

[9] ATLAS Collaboration, *Readiness of the ATLAS Liquid Argon Calorimeter for LHC Collisions*, Eur. Phys. J. **C70** (2010) 723–753, `arXiv:0912.2642 [physics.ins-det]`.

[10] 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]`.

[11] H.-L. Lai, M. Guzzi, J. Huston, Z. Li, P. M. Nadolsky, et al., *New parton distributions for collider physics*, Phys. Rev. **D82** (2010) 074024, `arXiv:1007.2241 [hep-ph]`.

[12] ATLAS Collaboration, *Summary of ATLAS Pythia 8 tunes*, ATL-PHYS-PUB-2012-003.

[13] 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 **1006** (2010) 043, `arXiv:1002.2581 [hep-ph]`.

[14] S. Frixione, P. Nason, and C. Oleari, *Matching NLO QCD computations with Parton Shower simulations: the POWHEG method*, JHEP **0711** (2007) 070, `arXiv:0709.2092 [hep-ph]`.

[15] P. Nason, *A New method for combining NLO QCD with shower Monte Carlo algorithms*, JHEP **0411** (2004) 040, `arXiv:hep-ph/0409146 [hep-ph]`.

[16] P. Golonka and Z. Was, *PHOTOS Monte Carlo: A Precision tool for QED corrections in Z and W decays*, Eur. Phys. J. **C45** (2006) 97–107, `arXiv:hep-ph/0506026 [hep-ph]`.

[17] M. Bahr et al., *Herwig++ physics and manual*, Eur. Phys. J. **C 58** (2008) 639–707, `arXiv:0803.0883 [hep-ph]`.

[18] P. M. Nadolsky et al., *Implications of CTEQ global analysis for collider observables*, Phys. Rev. **D 78** (2008) 013004, `arXiv:0802.0007 [hep-ph]`.

[19] S. Gieseke, D. Grellscheid, K. Hamilton, A. Papaefstathiou, S. Platzer, et al., *Herwig++ 2.5 Release Note*, `arXiv:1102.1672 [hep-ph]`.

[20] M. L. Mangano, M. Moretti, F. Piccinini, R. Pittau, and A. D. Polosa, *ALPGEN, a generator for hard multiparton processes in hadronic collisions*, JHEP **07** (2003) 001, `arXiv:0206293 [hep-ph]`.

[21] G. Corcella et al., *HERWIG 6.5 release note*, `arXiv:0210213 [hep-ph]`.

[22] G. Marchesini et al., *A Monte Carlo event generator for simulating hadron emission reactions with interfering gluons*, Comput. Phys. Commun. **67** (1991) 465–508.

[23] J. M. Butterworth, J. R. Forshaw, and M. H. Seymour, *Multiparton interactions in photoproduction at HERA*, Z. Phys. C **72** (1996) 637–646, `arXiv:9601371 [hep-ph]`.

[24] ATLAS Collaboration, *New ATLAS event generator tunes to 2010 data*, ATL-PHYS-PUB-2011-008.

[25] A. Martin, W. Stirling, R. Thorne, and G. Watt, *Parton distributions for the LHC*, Eur. Phys. J. **C63** (2009) 189–285, `arXiv:0901.0002 [hep-ph]`.

[26] ATLAS Collaboration, *The ATLAS simulation infrastructure*, Eur. Phys. J. **C 70** (2010) 823–874, `arXiv:1005.4568 [physics.ins-det]`.

[27] GEANT4 Collaboration, S. Agostinelli et al., *GEANT4: A simulation toolkit*, Nucl. Instrum. Meth. **A 506** (2003) 250–303.

[28] ATLAS Collaboration, *Performance of the ATLAS Inner Detector Track and Vertex Reconstruction in the High Pile-Up LHC Environment*, ATLAS-CONF-2012-042.

[29] ATLAS Collaboration, *Performance of primary vertex reconstruction in proton-proton collisions at $\sqrt{s}$ =7 TeV in the ATLAS experiment*, ATLAS-CONF-2010-069.

[30] W. Lampl et al., *Calorimeter clustering algorithms: description and performance*, ATL-LARG-PUB-2008-002.

[31] T. Barillari et al., *Local hadronic calibration*, ATL-LARG-PUB-2009-001.

[32] M. Cacciari, G. P. Salam, and G. Soyez. http://fastjet.fr/.

[33] M. Cacciari and G. P. Salam, *Dispelling the N^3 myth for the k_t jet-finder*, Phys. Lett. **B641** (2006) 57, `arXiv:0802.1189 [hep-ph]`.

[34] ATLAS Collaboration, *Properties of jets measured from tracks in proton-proton collisions at center-of-mass energy $\sqrt{s}$ = 7 TeV with the ATLAS detector*, Phys. Rev. **D84** (2011) 054001, `arXiv:1107.3311 [hep-ex]`.

[35] ATLAS Collaboration, *Improved electron reconstruction in ATLAS using the Gaussian Sum Filter-based model for bremsstrahlung*, ATLAS-CONF-2012-047.

[36] ATLAS Collaboration, *Expected electron performance in the ATLAS experiment*, ATL-PHYS-PUB-2011-006.

[37] ATLAS Collaboration, *Muon reconstruction efficiency in reprocessed 2010 LHC proton-proton collision data recorded with the ATLAS detector*, ATLAS-CONF-2011-063.

[38] ATLAS Collaboration, *Electron performance measurements with the ATLAS detector using the 2010 LHC proton-proton collision data*, Eur. Phys. J. **C72** (2012) 1909, `arXiv:1110.3174` `[hep-ex]`.

[39] ATLAS Collaboration, *Probing the measurement of jet energies with the ATLAS detector using Z+jet events from proton-proton collisions at* $\sqrt{s}$ = 7 *TeV*, ATLAS-CONF-2012-053.

[40] M. Cacciari, G. P. Salam, and G. Soyez, *The Catchment Area of Jets*, JHEP **0804** (2008) 005, `arXiv:0802.1188` `[hep-ph]`.

[41] S. D. Ellis and D. E. Soper, *Successive combination jet algorithm for hadron collisions*, Phys. Rev. **D48** (1993) 3160, `arXiv:hep-ph/9305266` `[hep-ph]`.

[42] S. Catani, Y. L. Dokshitzer, M. Seymour, and B. Webber, *Longitudinally invariant $k_\perp$ clustering algorithms for hadron hadron collisions*, Nucl. Phys. **B406** (1993) 187.

[43] ATLAS Collaboration, *Probing the measurement of jet energies with the ATLAS detector using photon-jet events in proton-proton collisions at* $\sqrt{s}$ = 7 *TeV*, ATLAS-CONF-2012-063.