Jet Data Quality at ATLAS

Ian Moult
IPP Summer Student
May 2010-September 2010

Abstract

The application of automated checks for offline jet data quality based on the comparison of jet variables with reference distributions has been investigated at the ATLAS experiment. These studies were performed with jets from $\sqrt{s}$=7 TeV proton-proton collisions during the early running of the Large Hadron Collider (LHC). The stability of jet variables under the changing run conditions of the LHC, and the effect of event filters on the data were studied. A set of algorithms has been implemented, and is currently being used by the offline jet data quality shifters, who will evaluate their performance for a large number of runs. It is hoped that under stable running conditions, these algorithms can be used to fully automate the jet data quality analysis.

1 Introduction

1.1 LHC

The LHC is a proton-proton synchrotron currently operating at a center of mass energy of $\sqrt{s}$=7 TeV. The commissioning of the LHC, which began with the first collisions at $\sqrt{s}$=7 TeV in March 2010, focused initially on increasing the luminosity of the machine by increasing the bunch intensity and decreasing the β^* of the beam. This focus has now shifted to increasing the number of colliding bunches. These efforts have resulted in a highest instantaneous luminosity of $4\text{x}10^{30}cm^{-2}s^{-1}$ and a total integrated luminosity of $1.6pb^{-1}$ delivered to the experiments under stable running conditions. The LHC will continue running to the end of 2010 at $\sqrt{s}$=7 TeV, and stable running at $\sqrt{s}$=7 TeV is planned for 2011 with the goal of reaching $1fb^{-1}$ by the end of the year.

ATLAS is one of four detectors located around the LHC ring. Both ATLAS and CMS are general purpose detectors designed to maximize the discovery potential for new physics at the TeV scale. Two other more specialized experiments have also been designed to take advantage of the LHC energy scale. The LHCb detector is specialized to study b quarks, which will be produced copiously at the LHC energies, whereas the ALICE experiment is designed to study heavy ion collisions.

1.2 The ATLAS Detector

The ATLAS detector is a multi-purpose detector designed specifically for the high energy and luminosity conditions of the LHC. The design of the ATLAS detector is that of a standard hermetic detector, and is shown in Figure 1. It consists of an inner tracking system in a solenoidal magnetic field, surrounded by electromagnetic and hadronic calorimeters. The impressive size of the ATLAS detector is due to the toroid magnets and muon detection system which surround the entire detector giving it an outer diameter of 25m.

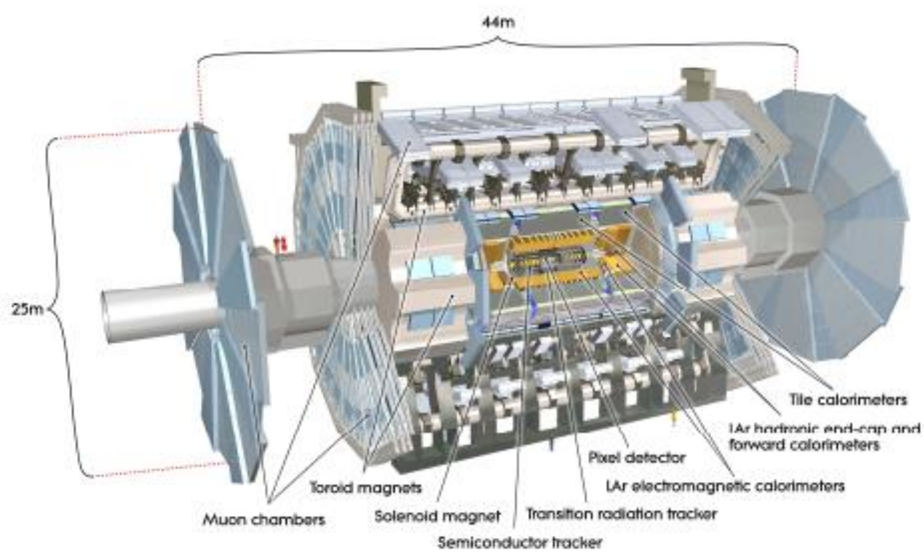


Figure 1: The ATLAS detector and subsystems. (Image from [1])

1.3 Atlas Calorimeters

This report will focus on the study of jets in the ATLAS detector, and it is thus essential to consider in more detail the ATLAS calorimeter system. The ATLAS detector has both electromagnetic (EM) and hadronic calorimeters. EM calorimeters are used to measure the energy of photons and electrons, whereas hadronic calorimeters are used to measure the energy of hadrons. A cutaway view of the ATLAS calorimeter system is shown in Figure 2. It provides coverage in the range $|\eta|<4.9$.

The ATLAS EM calorimeter is a liquid Argon detector, and is divided into three sections. The barrel calorimeter surrounds the inner tracking system and covers the pseudorapidity range $|\eta|<1.475$. Two end caps (EMEC) cover the range $1.375<|\eta|<3.2$. These calorimeters use stainless steel, or lead as the absorbing material, and liquid Argon as the active material. Each of the ATLAS EM calorimeters is further divided into three layers, and a pre-sampler, giving a total thickness of 22 radiation lengths in the barrel, and 24 radiation lengths in the endcaps.

The ATLAS hadronic calorimeter consists of the Tile barrel for $|\eta|<1.0$, and extended barrel for the range $0.8<|\eta|<1.7$. The Tile calorimeter is a sampling calorimeter using scintillating tiles separated by steel absorbers. Two hadronic end cap calorimeters (HEC) cover the pseudorapidity range $1.5<|\eta|<3.2$, and a forward calorimeter (FCAL) covers the high pseudorapidity range $3.1<|\eta|<4.9$. Both the HEC and FCAL are based on liquid Argon technology. The ATLAS hadronic calorimeters are also divided into layers. The three layers of the Tile calorimeter give it a total thickness of 9.7 interaction lengths, and thicknesses of 10 interaction lengths in both the FCAL and HEC are provided by three and four layers respectively.

To relate the quantity of energy deposited in the calorimeter with the initial energy of the incident particle a scale factor is used. This scale factor is a function of η, ϕ, and p_T, the component of momentum transverse to the beamline. The interactions of particles in EM calorimeters are mainly elastic, whereas the interactions in hadronic calorimeters are nuclear, and suffer from inelastic interactions, and neutrino and slow neutron production, leading to a loss in the measured energy. The ATLAS detector therefore has different scale factors for the EM and hadronic calorimeters. This is referred to as a non-compensating calorimeter.

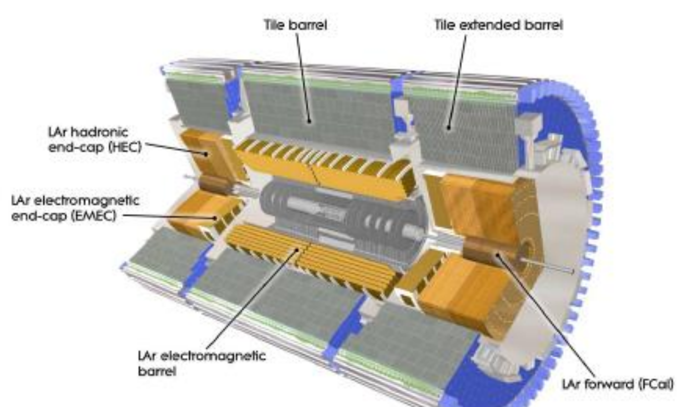


Figure 2: Schematic of the ATLAS calorimeter system providing coverage in the range $|\eta|<4.9$. (Image from [1])

In the early running of the LHC the ATLAS Liquid Argon calorimeters have had a variety of noise problems. Noise bursts in whole partitions of the EM barrel, or endcap have been observed, as well as large (~1TeV) energy spikes occurring in isolated cells in the HEC calorimeters [6]. Sporadically noisy single cells are also present in the calorimeters.

The ATLAS calorimeters have also suffered from electronics problems. This includes the failure of optical transmission fibers which couple the scintillator in the calorimeters to the electronic readouts, and the failure of front end boards. These problems, along with the masking of calorimeter cells due to noise, produce inactive regions in the detector. Currently only 2% of channels have been affected [7].

2 Jets at Hadron Colliders

2.1 Quantum Chromodynamics

Jet production is described by the theory of Quantum Chromodynamics (QCD). QCD is the SU(3) color component of the Standard Model describing objects with color, namely quarks and gluons. QCD processes have been less precisely measured experimentally, and are more difficult theoretically than electroweak processes. This is due to the fact that perturbation theory cannot be used at low energy scales, where the strong coupling constant, α_s, becomes large. At high energy scales α_s is small, which allows for the use of perturbation theory in hard scattering processes. These properties of QCD, namely α_s becoming large at low energy and small at high energy are referred to as confinement and asymptotic freedom respectively. These phenomena can be seen in Figure 3, which shows a plot of α_s as a function of the momentum transfer, Q, as predicted by QCD, and as measured in experiments.

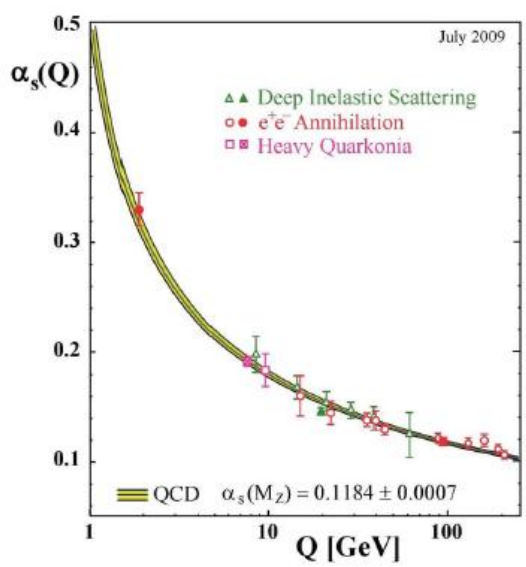


Figure 3: Experimental measurements and QCD prediction of the strong coupling constant, α_s, as a function of the momentum transfer Q. (Image from [2])

2.2 Jets and Hadronization

Jets and hadronization are the experimentally observed consequences of confinement and asymptotic freedom. When a color charge is produced in an event it will emit collinear partons (gluons and quarks). These partons in turn emit further partons in a so called shower. Showering is described by a fragmentation function, giving the probability of finding a parton, p_i, with a certain momentum fraction, z, in a shower initiated by a parton, p_j, starting with energy scale μ. The motion of the partons is decoupled at small distance scales due to asymptotic freedom. This allows two partons created together to be emitted in different directions without influencing one another's motion.

Due to QCD confinement, the color charges of a shower, if sufficiently separated, will have enough energy to pull a quark anti-quark pair from the vacuum. This will continue until all the quarks have formed bound hadronic states. This process is called hadronization, and from an experimentalist's point of view produces a collimated spray of hadrons which is observed in the detector. Such a spray of particles is called a jet. An example di-jet event at $\sqrt{s}$=7 TeV in the ATLAS detector is shown in Figure 4, where two collimated sprays of particles are clearly visible in the event reconstruction, and two large energy deposits are seen in the calorimeters.

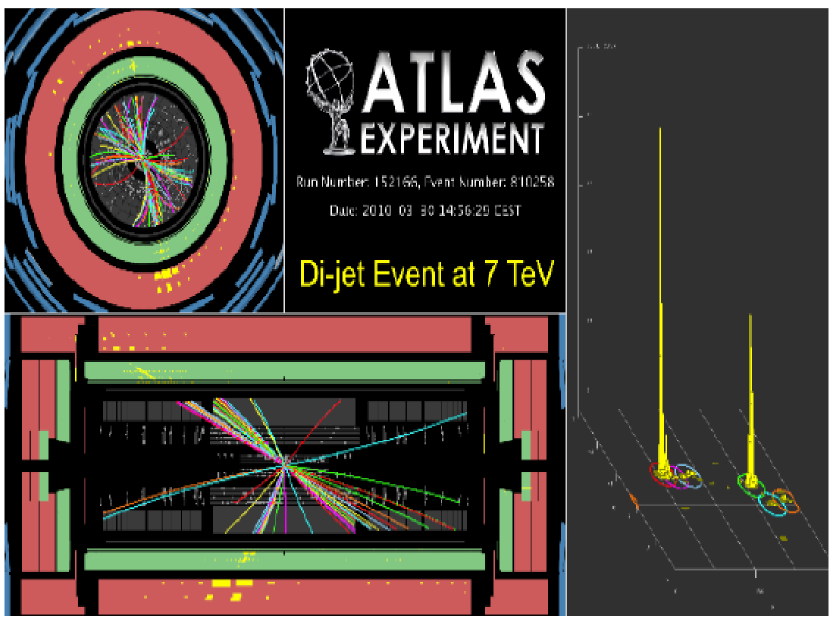


Figure 4: A reconstructed di-jet event in the ATLAS detector at $\sqrt{s}$=7 TeV. The two observed jets have energies of 500GeV, and 315 GeV. (Image from [3])

2.3 Importance of Jets in Physics Analysis at the LHC

Jets are important in hadron collider physics for many reasons. As is shown in Figure 5, the jet cross section at hadron colliders is orders of magnitude higher than that for the electroweak processes. This alone necessitates a good understanding of jet events. This also means that jets are present as underlying events in physics events of interest.

Jets are also important in physics processes in which a parton is produced as a final state. Due to hadronization, this parton will be observed experimentally as a jet. An example of such a process that will be studied at ATLAS is top quark decay, which occurs primarily into a b quark and a W boson. Analysis of such physics channels relies on, and is limited by, the understanding of jets, which is more difficult than leptons.

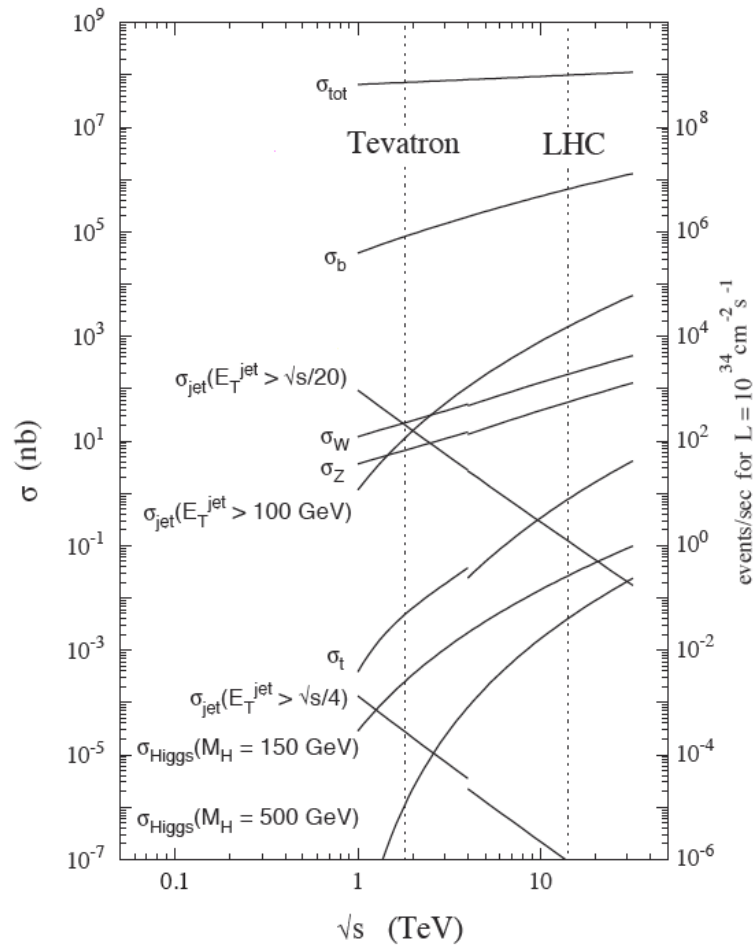


Figure 5: Cross-sections for physics processes as a function of $\sqrt{s}$. The discontinuity in the cross-sections is due to the fact that the Tevatron is a proton-antiproton collider, whereas the LHC is a proton-proton collider. The jet cross-section is orders of magnitude larger than that of the largest electroweak process.

3 Jet Reconstruction

Jet reconstruction is the process of reconstructing jet properties from the energy deposited in individual cells of the calorimeter. In the first step, the energy in clusters of cells is summed to give so called “proto-jets”. A jet reconstruction algorithm is then used to group these proto-jets to reconstruct the complete jet topology of the event.

3.1 Clustering

Tower Clustering:
Tower clustering is the simplest clustering method. This method is based on the physical segmentation of the calorimeters into an η-ϕ grid with granularity 0.1x0.1. Since the calorimeter is also segmented in the radial direction, a tower is defined as the stack of calorimeter cells with a fixed η-ϕ coordinate.
In tower clustering, the energy deposited in each of the cells in a tower are summed to give the energy of a proto-jet associated with the η-ϕ direction of the tower. A 4-momentum is then assigned to each proto-jet by defining it to be a massless pseudo-particle. Although this method is simple, it suffers from the fact that all cells in the calorimeter are included in the energy sums. This makes the algorithm maximally sensitive to noise and baseline fluctuations.

Topological Clustering:
The topological clustering technique attempts to reduce the effect of noise when defining proto-jets by only including cells which are deemed relevant for the event. The topological clustering technique first selects seed cells with an energy 4σ above the background noise level. Neighboring cells that are 2σ above background are then found, and finally a guard ring of 0σ cells around the cluster is included. The total energy of the cluster is then the sum over all cells included in the cluster, and the η-ϕ coordinates are defined by the center of energy of the cluster. Again, the 4-momentum is assigned by defining the cluster to be a massless pseudo-particle. A disadvantage of topological clustering is a loss of the shape information of the jet. This occurs when a cluster, which is typically larger than a single 0.1x0.1 η-ϕ cell, is assigned a single η-ϕ coordinate based on its center of energy.

TopoTower:
A combination of the above two methods can also be used. In the TopoTower method, topological clusters are first formed. The proto-jets are then defined by summing over the cells in each η-ϕ tower, however, only the cells which were found to be part of a cluster are included in the sum. This has the advantage of reducing the sensitivity to noise without a loss of the shape information of the jet [5].

Both the topological, and TopoTower clustering methods are currently used at ATLAS.

3.2 Jet Reconstruction Algorithms

Jet reconstruction algorithms are used to combine the proto-jets, as defined using one of the methods described above, into a reconstructed jet which can be used for physics analysis. A variety of algorithms exist for this purpose, however, only the Anti-Kt algorithm will be described here as it is currently being used at ATLAS[4][5]. The Anti-Kt method is preferred as it is a non-seeded algorithm, and thus does not suffer from the theoretical issues which seeded algorithms do[1].

The Anti-Kt algorithm is an iterative algorithm, which takes each pair of proto-jets, i, j, and either combines them into a single proto-jet, or leaves them separated, based on the measure functions D_{ij}, and D_i, defined as:

$$D_{ij} = \frac{\min\,({P_{t_i}}^{-2}, {P_{t_j}}^{-2})\Delta R_{ij}}{\mathrm{D}}$$

$$D_i = {P_{t_i}}^{-2}$$

where $\Delta R = \sqrt{(\Delta\varphi)^2 + (\Delta\eta)^2}$, and D is a parameter, typically chosen to be either 0.4, or 0.6 at ATLAS. Although the Anti-Kt algorithm is not based on a fixed cone size, the parameter D effectively acts as a jet radius, and is therefore chosen depending on the type of jets one is looking for.

At each stage, the two proto-jets are combined if $D_{ij}<D_i$ and otherwise are left separate. To combine the proto-jets, the 4-momenta are simply added. If the two proto-jets are not collinear, the addition of the 4-momenta will give a mass to the combined proto-jet, even if the two initial proto-jets were massless. This process is continued until a final stable configuration is reached. The reconstructed jets are then the proto-jets, or combinations of proto-jets, which remain separated in the final configuration.

4 Jet Data Quality Monitoring

4.1 Data Quality Monitoring

All sub-detectors of ATLAS have data quality monitoring to ensure that the detector is working correctly and to identify any problems such as noisy or dead channels. There is also data quality monitoring for high level physics objects such as Missing E_T, electrons, muons, taus and jets. ATLAS has both online data quality monitoring, which looks at the data live as it is coming out of the detector, as well as offline data quality monitoring, which looks at the processed data from the entire run. A special stream of data, which will be discussed later, is used for this purpose. Processing of this data stream is complete 1-2 hours after the run has finished.

Offline data quality monitoring is performed by shifters who have access to the run information through a series of histograms available on the ATLAS data quality website. Automatic algorithms are used to alert the offline shifters to potential problems which need to be followed up on. These algorithms are based either on self consistency checks, or comparisons with reference data. A screen shot of the ATLAS data quality website is shown in Figure 6. All the different data quality groups are shown in the list on the left, along with their status for this particular run based on the automatic checks. The jet data quality channel is currently marked as undefined, as the algorithms are currently in the testing stage.

[1] For example, Anti-Kt is infrared safe.

Figure 6: The ATLAS data quality website, showing a list of the data quality groups.

4.2 Jet Data Quality

The ATLAS jet data quality group focuses specifically on the quality of jet data in the detector. Jet data quality is especially important due to its relation to the calorimeters. Although there are multiple data quality groups focusing solely on the calorimeters, the jet data quality gives a probe of the effect of the problems in the calorimeters on the measurement of physics processes. Jets are particularly sensitive to the noise problems in the calorimeters, as they can be seeded by noise from a single calorimeter cell.

For jet data quality, each run is also split into groups of 10 luminosity blocks (A luminosity block is approximately two minutes of data), each of which can be analyzed individually. This makes it possible to determine the time of a noise burst during a run, or to correlate problems in the data with high voltage trips or other detector failures that occurred.

4.3 Overview of Jet Parameters

A variety of jet variables are available on the ATLAS data quality website to study jet properties. For simplicity these are divided into six groups, each of which will be discussed separately. Figure 7 shows a screen shot of the website for a particular jet data stream, with a list of the six groups. In this example, four groups have been flagged as red by the algorithms.

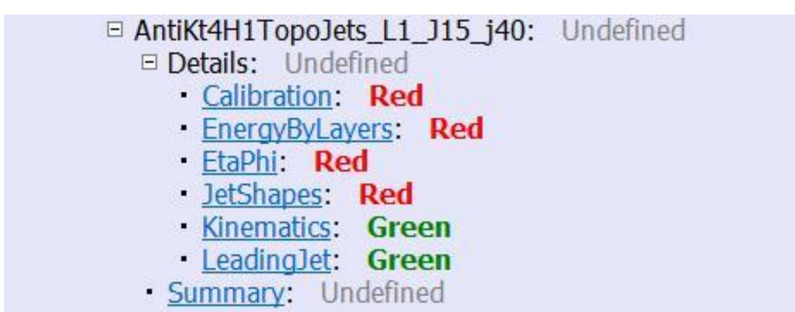


Figure 7: The six groups of variables used for jet data quality analysis.

The “Energy by Layers” histograms show the distributions of the jet energy fractions deposited in all the different calorimeters, and in the different layers of these calorimeters. These variables are important as they can be used to detect noise in a particular layer of the calorimeter. An example plot of the energy deposited into the first layer of the EM calorimeter is shown in Figure 8.a).

The “Jet shapes” directory includes distributions of the jet width, the number of constituents per jet, as well as jet profile information. The jet profile plots show the fraction of the jet constituents, or p_T of the jet, as a function of the jet radius R. Figure 8.b) shows an example of the average integrated jet profile with respect to p_T: $\int \frac{dp_T}{dr} dr / p_T$.

The Calibration directory contains plots of the ratio of the hadronic scale to the EM scale as functions of η, ϕ, and p_T. These plots are important as changes in the calibration scale between runs will lead to changes in the jet distributions. A sample plot of the calibration scale as a function of η is shown in Figure 8.c).

An important set of plots for offline data quality analysis are 2D η-ϕ plots of jet properties. These plots provide an easy method to spot noisy cells, or noisy regions of the calorimeter, and to determine their η-ϕ location. This is important due to the noise problems which are currently being experienced in the ATLAS calorimeters, as was previously discussed. Figure 8.d) shows an example of such a noisy run, which was flagged by the offline shifters for noise bursts in the calorimeter.

A full set of kinematic variables is also available for the jets. These provide distributions for properties such as mass, p_T, E_T, E, etc. These kinematic distributions are also available for the leading (highest p_T) jet.

4.4 Data streams

Two different data streams are considered in this analysis, and thus their triggers and purposes will be briefly discussed here. The express stream is a data stream specifically designed for data quality analysis. It begins processing shortly after a run begins, and is finished processing into AOD format 1-2 hours after the run is complete, allowing for rapid data quality analysis. To achieve this fast processing time, the calibration and alignment constants from the previous run are used in the data processing.

The express stream has a rate of 10-15Hz and contains a pre-scaled mixture of a variety of data triggers. It contains an even distribution of the jets, muons, electrons and Missing E_T physics triggers, the calibration triggers (in particular the calorimeter calibration triggers) and the MinBias triggers.

A trigger filtered version of the express stream is also used for jet data quality, and will be referred to as the _L1_J15_j40 stream. This stream uses the L1_J15 trigger, which is designed to trigger on events with at least one high p_T jet. This trigger is applied at level 1, where the jet energy is not accurately known and thus a high p_T threshold is not well defined. The trigger reaches an efficiency greater than 99% for jets with an offline p_T>80GeV.

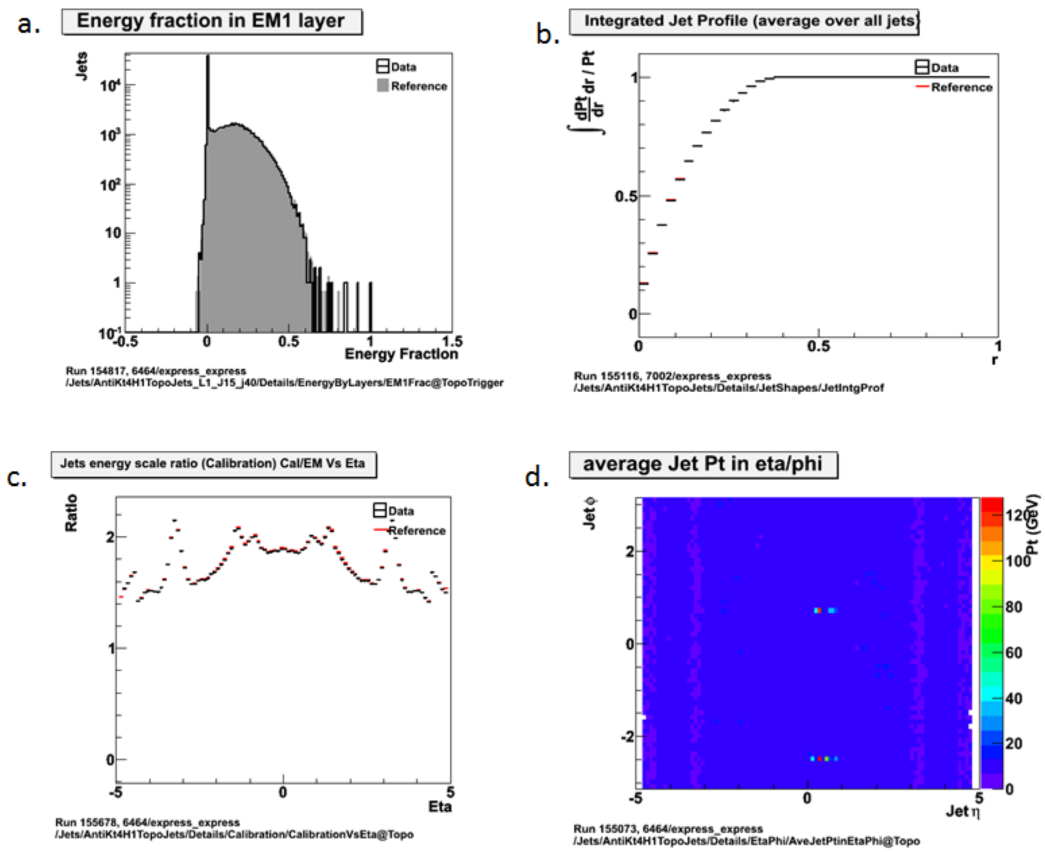
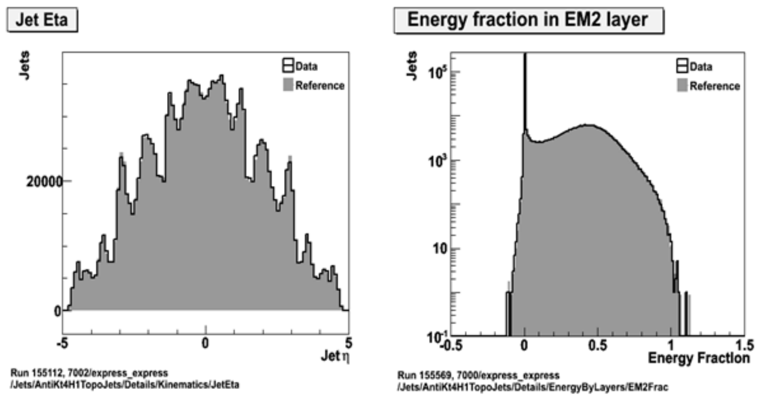


Figure 8: Examples of jet variables. a) Distribution of the jet energy fraction deposited in the first layer of the EM calorimeter. b) Integrated jet profile. c) Jet energy scale ratio as a function of η. d) Jet p_T distribution as a function of η and ϕ for run 155073, which was flagged for noise bursts in the calorimeters.

4.5 Stability of distributions

To determine the applicability of references for jet data quality, it is first necessary to study the stability of the jet variables from run to run, and also to investigate the effect of changing run conditions on the jet variables. This is especially important in the early running of the LHC, when the run parameters are being changed constantly. Some examples of run parameters that could have an effect on the jet distributions (assuming a fixed energy) are the bunch intensity, the number of bunches, and the β^* of the beam.

To begin with, the stability of the distributions under similar run conditions was considered. It was found that under similar run conditions there is, for the majority of variables, excellent agreement between distributions. For these distributions, the χ^2/ndf was used to assess the quality of the agreement. A few examples are shown in Figure 9, along with the χ^2/ndf values. This stability is promising as it suggests that simple comparisons between distributions using a χ^2/ndf test may be possible under stable running conditions.



a. χ^2/ndf=1.3 b. χ^2/ndf=1.1

Figure 9: Examples of the stability of the jet variables for two runs with similar run conditions. a) The jet η distribution for run 155112. b) The energy fraction in the EM2 layer for run 155569. In both cases the reference is run 155116. The interesting shape of the jet η distribution arises due to gaps between the different calorimeters, where the efficiency is reduced.

The study of the effect of changing run conditions on the jet parameter distributions is considerably more difficult. Due to the limited number of available runs, multiple parameters were changed together, making it difficult to determine the actual cause of any variations in the jet variables. Despite this fact, it is still possible to study the variation in the distributions under changing run conditions and to determine which variables are the most sensitive, and what types of changes are observed in the distributions.

The main parameter that was being changed in the runs that were available for this study was the bunch intensity, which was changed from $20\text{x}10^9$ protons per bunch to $2\text{x}10^{11}$ protons per bunch in a number of steps. The number of bunches was also increased, however due to the large bunch spacing, this is not considered to have an effect. The β^* of the beam was also varied between runs, but this effect was not separable from the changing bunch intensity.

The kinematic variables proved to be sensitive to changing run conditions. Significant variations were observed in both the average kinematic variables, and the distributions of each variable. Figure 10.a) shows an example of the effect of changing run conditions on the average Jet E_T. The average Jet E_T is seen to be considerably higher in the central region for run 155678, shown in black, compared with run 155116, shown in red. This is attributed to the fact that run 155678 has a higher bunch intensity ($2\text{x}10^{11}$) than run 155116 ($60\text{x}10^9$). In general, the average kinematic properties were found to be unstable under changing run conditions, which is manifest in a shifting of the central section of the average kinematic distributions.

The distributions for the kinematic jet properties, especially p_T, E_T, E, and mass are also quite unstable under changing run conditions. Figure 10.c) shows an example of this effect for Jet p_T, again for the runs 155678 (data) and 155116 (reference). The Jet p_T distribution for run 155678, with the higher bunch intensity, falls off more slowly than for run 155116. In general, changing run conditions caused variations in the tails of the distributions for the jet kinematic properties.

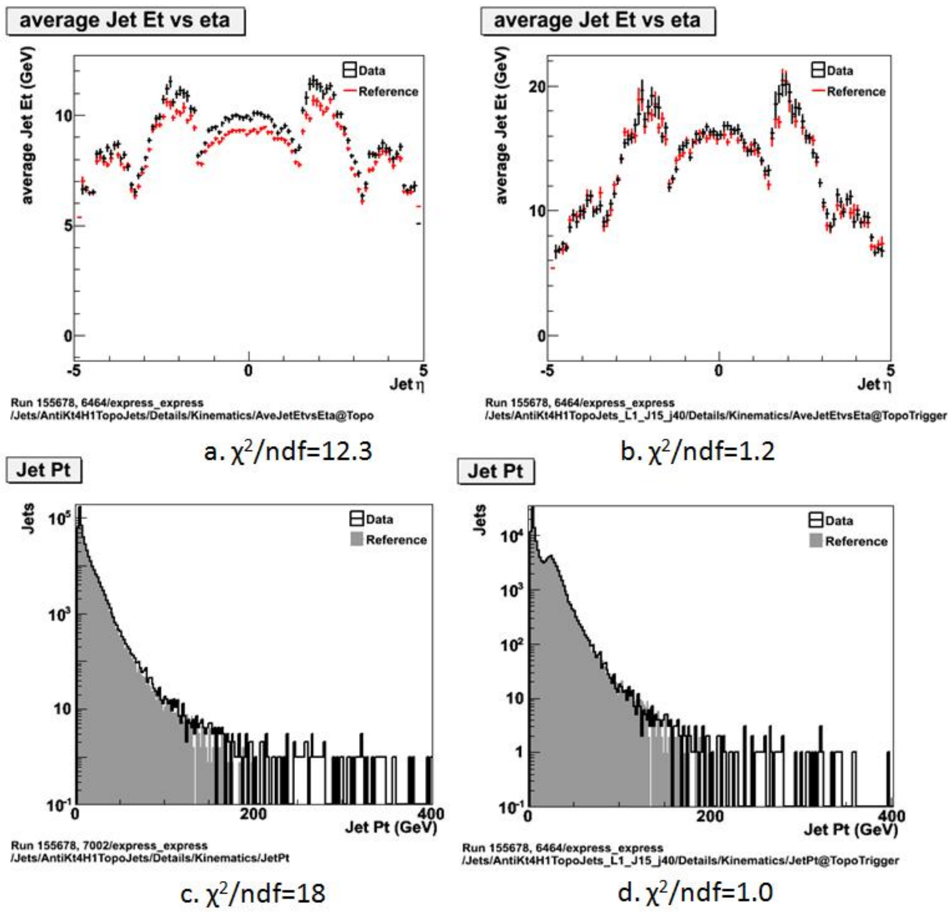
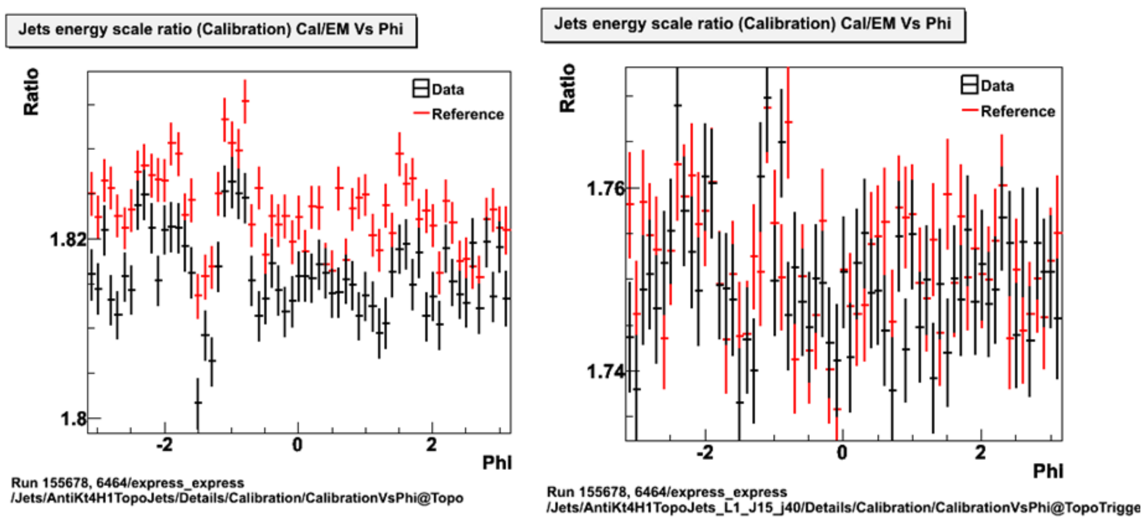


Figure 10: a) Average jet E_T as a function of η for run 155678 (black) with a bunch intensity of $2x10^{11}$ and run 155116 (red) with a bunch intensity of $60x10^9$. b) Jet E_T distributions for the same runs after the L1_J15 filter has been applied. c) Jet p_T distributions for run 155678 (data) and run 155116 (reference). d) Jet p_T distributions after the L1_J15 filter has been applied. In both cases better agreement is observed in the _L1_J15_j40 data stream.

In both of the cases described above, the discrepancies in the distributions are dominated by differences in the low p_T jets. This can be demonstrated by considering the _L1_J15_j40 event filter stream for these same two runs. This is shown for the two different distributions in Figures 10.b.) and 10.d.), respectively. As is clear both visually, and from the χ^2/ndf value, there is better agreement between the different runs in the _L1_J15_j40 stream. The reason that variations in the low p_T jets have an effect on the tails of the kinematic distributions is that these distributions are normalized to the total number of jets. Due to the fact that the number of low p_T jets is orders of magnitude larger than the number of high p_T jets, the low p_T jets dominate the normalization. Therefore, changes in the number of low p_T events between runs have an effect, through normalization, on the high p_T tails of such distributions. The same is true for the E_T, E and mass distributions.

Similarly to the average kinematic variables, the calibration scales also exhibit fairly large shifts when the run conditions are changed. An example of this for the scale ratio as a function of ϕ, again for the runs 155678 (shown in black), and run 155116 (shown in red), is shown in Figure 11.a). As with the kinematic variables, this effect is again due to variations in the low p_T jets, as can be demonstrated by considering the _L1_J15_j40 data stream, which is shown in Figure 11.b). A similar shift is also observed in the calibration scale as a function of η. The calibration scale as a function of p_T is however quite stable, especially at high p_T. This again agrees with the understanding that the changes in the distributions are dominated by variations in the low p_T jets.

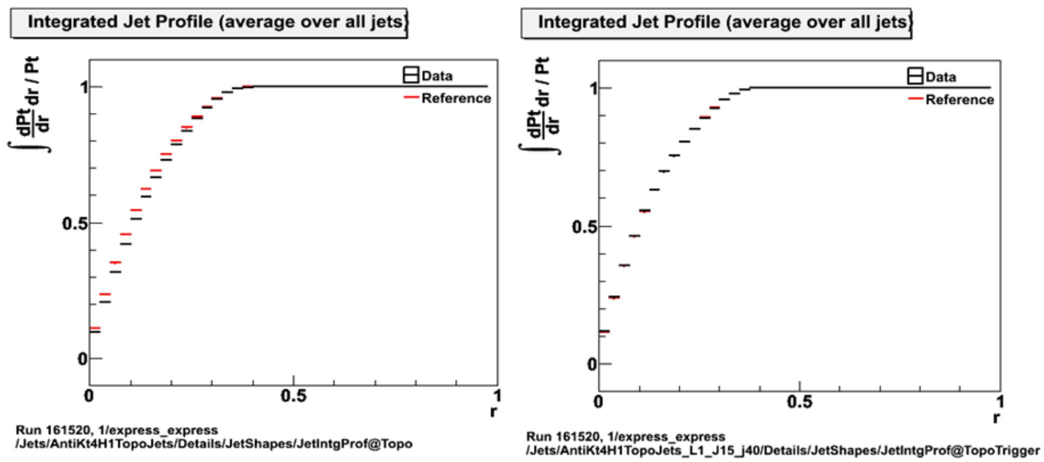


a. χ^2/ndf=5.0

b. χ^2/ndf=0.9

Figure 11: a) Jet energy calibration scale as a function of ϕ for run 155678 (black) and run 155116 (red). b) The same distributions as in a) but for the _L1_J15_j40 data stream. Better agreement is observed in the _L1_J15_j40 stream.

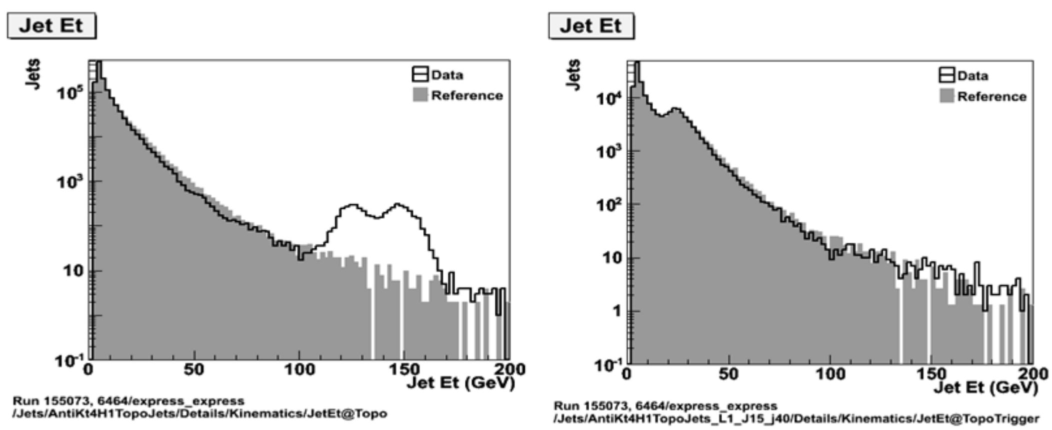
One final category of plots that are particularly unstable are the jet profile variables. The jet profile variables are problematic, as they tend to be unstable not only under changing run conditions, but also between runs with similar run conditions. An example of an integrated jet profile plot is shown in Figure 12.a) for two runs with similar bunch intensities. It is clear visually that the agreement between the distributions is not good. These distributions are difficult to compare using statistical algorithms due to their extremely small error bars. For example, a comparison of the two distributions in Figure 12.a) gives a χ^2/ndf of 852. It is expected that once the LHC has reached a stage where the run conditions are nearly identical for each run, more effective comparisons will be possible for these distributions. Fortunately, the L1_J15 trigger greatly reduces the instability of the jet profile distributions allowing for comparisons of runs with similar conditions. An example of this is shown in Figure 12.b), where the χ^2/ndf has been reduced from 852 to 0.98 by applying the L1_J15 trigger. Even in the L1_J15_j40 data stream, there is still in general not good agreement between the jet profile plots for runs with different run conditions, with typical χ^2/ndf values in the range of 5-15.



a. χ^2/ndf=852 b. χ^2/ndf=0.98

Figure 12: a) Integrated jet profile for two runs with similar run conditions showing the instability of this variable. b) The L1_J15 filtered version of the same run. A large improvement in the agreement is observed.

Although the _L1_J15_j40 data stream shows promise for use when comparing distributions under changing run conditions, it suffers from one drawback in that it also tends to remove the obvious noise problems. Figure 13 shows an example of the jet E_T distribution for run 155073 before and after the L1_J15 filter. The obvious noise was largely removed by the filter. Such noise will however be picked up by the online shifters, and thus comparisons of the _L1_J15_j40 data stream still provide a means of studying the more subtle variations in the distributions.



a. χ^2/ndf=59 b. χ^2/ndf=3.3

Figure 13: a) Jet E_T distribution for run 155073, which was flagged for noise in the calorimeter, compared with run 155116. b) Jet E_T distribution for the same runs, but with the L1_J15 filter applied. The effect of the noise has been reduced.

Based on the stability of the distributions in the runs analyzed, the implementation of the comparison of jet distributions using simple algorithms seems quite promising. The distributions proved to be extremely stable under similar running conditions. Under changing run conditions, a variety of distributions were unstable, most notably the kinematic properties, calibration scales, and jet profile variables. However, with the exception of the jet profile variables, these distributions change in a predictable manner, dominated by variations in the low p_T jets. These effects can therefore be largely removed by the L1_J15 filter, allowing for direct comparison of the distributions even for different run conditions.

4.6 Implementation of algorithms

Due to the stability of the χ^2/ndf algorithm found in the previous sections, this has been implemented for the majority of the distributions. To use the χ^2/ndf a variety of cuts are established, typically one for each of the data sections, and a few special cuts for problematic variables. If the χ^2/ndf of a certain variable when compared with the reference is higher than these cuts, then this variable is flagged for closer inspection by the offline shifters. Typical cuts under stable running conditions are between 3 and 5.

The χ^2/ndf algorithm was found to be more robust than, for example, Kolmogorov-Smirnov, or other statistical algorithms which proved to be too sensitive for the amount of change observed run to run. The χ^2/ndf provides a relatively stable measure of the discrepancies between the distributions and a reference.

For the 2D η-ϕ plots it is essential to be able to pick out the η-ϕ coordinates of noisy cells, and to flag a distribution even if there is only one noisy cell. Due to the large number of degrees of freedom in a 2 dimensional histogram, a χ^2/ndf is not ideally suited. A large variation in one cell will be masked by the large number of other cells. Because of this, two other algorithms were developed. One algorithm takes advantage of the fact that the distributions of kinematic variables are approximately constant with respect to ϕ. This allows one to take an average over a strip of bins with constant η, and look for bins in that strip which are significantly different than the strip average. This algorithm has the advantage that it does not rely on a reference, and thus is unaffected by changing run conditions.

The method of using a strip average to search for outliers cannot be used on a plot of the number of jets as a function of η-ϕ which does not exhibit such uniformity, and has significantly more substructure. An example distribution of the number of jets as a function of η-ϕ is shown in Figure 14.a), and a map of the problematic calorimeter sections is shown in Figure 14.b). The effect of the problematic regions of the calorimeter is clearly visible in the number of jets plots. Due to this substructure, an algorithm which detects the largest outliers with respect to a reference histogram and returns their η-ϕ coordinates was implemented. The χ^2/ndf is also computed to compare the more global properties of the distribution. A potential problem with this method is that it is quite sensitive to noise in the reference histogram, and if there are noisy cells which are present in the reference, then these will either not be detected by the algorithm, or will be incorrectly flagged as noisy.

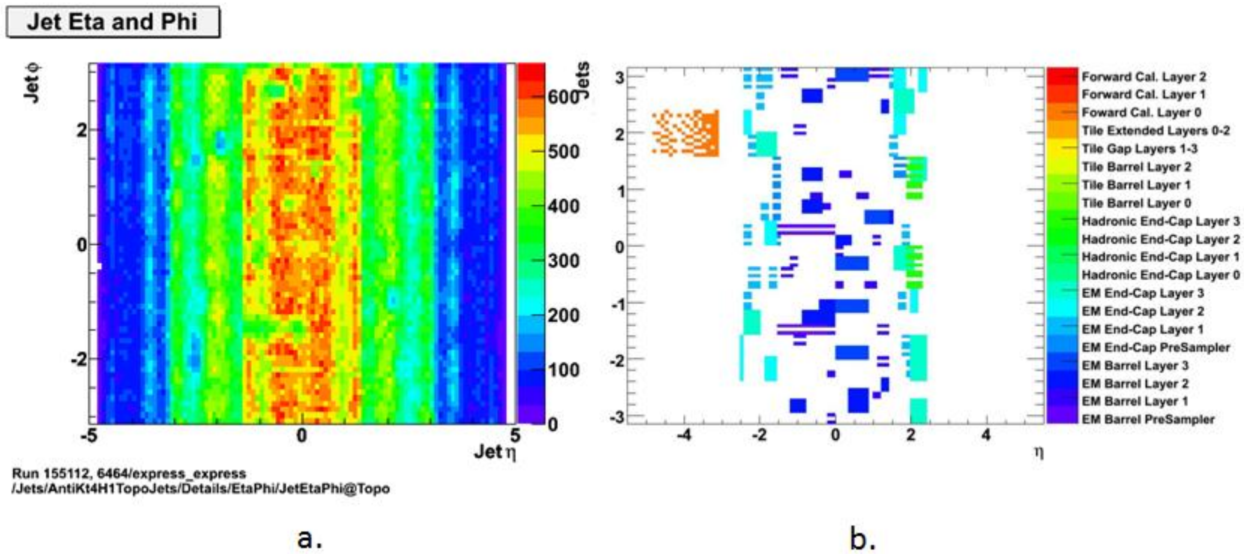


Figure 14: a) The number of jets observed in the detector as a function of η and ϕ.
b) A map of the problematic regions of the detector. A clear correspondence is observed between problematic regions and regions where fewer jets are observed. (Image b. from [7])

Figure 15 shows an example of the effectiveness of the χ^2/ndf method. Run 155280 was flagged for noise in the tile calorimeter, but has otherwise the same running conditions as the reference, run 155116. When the two runs are compared, a χ^2/ndf of 15.3 is found. This is considerably higher than the typical value of 1-2 for the energy distributions and thus this distribution would be immediately flagged for a closer look by the shifters.

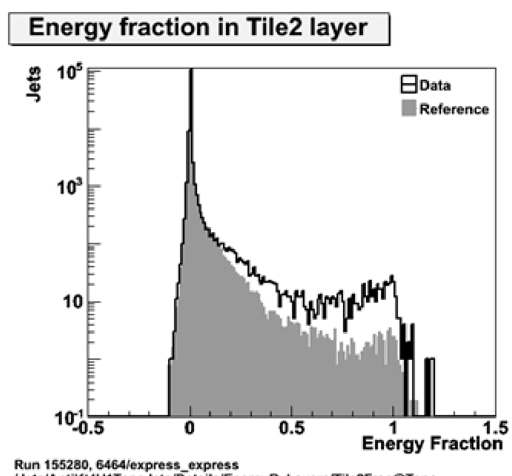


Figure 15: Energy fraction in Tile Layer 2 for run 155280 compared with run 155116. Run 155280 was flagged for noise in the Tile calorimeter.

A second example shows the effectiveness of the strip averaging algorithm for the 2D η-ϕ distributions for the run 155678. Figure 16 shows a screen shot of the output of the algorithm to the data quality website. The algorithm successfully flags the noisy bins, which are clearly visible in the image due to their red colour. Their η-ϕ coordinates are also printed in a list on the left, along with how many sigma the bin is above the strip average . The 2D η-ϕ distributions were found to be quite stable, even under changing run conditions. By using a cut of 5σ one can manage to flag only bins that seem to be caused by noise, and typically less than 10 bins are flagged per run.

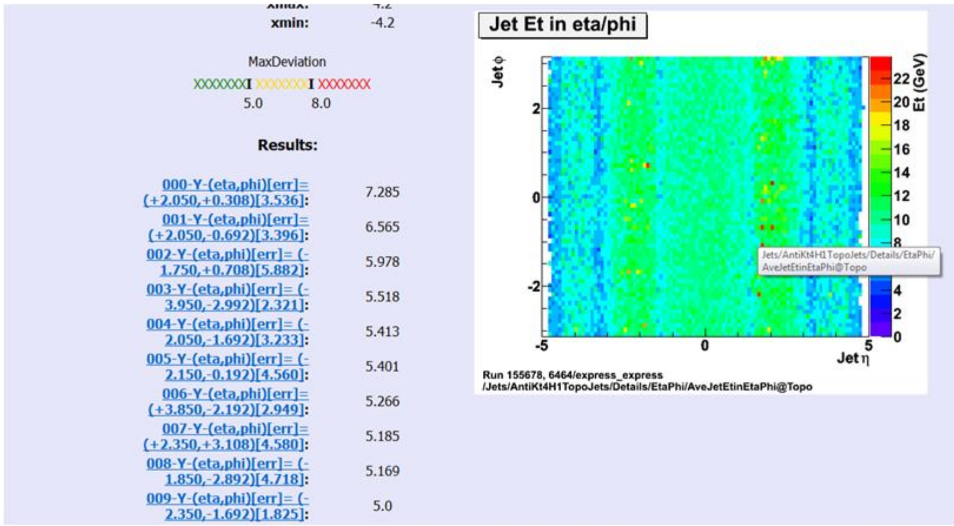


Figure 16: Output to the data quality website showing the flagging of noisy η-ϕ bins based on a strip average.

These algorithms have been implemented on the ATLAS data quality website. They are currently being evaluated by the offline jet data quality shifters on a larger number of runs to study their performance and effectiveness. This is important, as only a limited number of runs were available during the testing of these algorithms, and it is essential to see if the results extend to the newer runs. This evaluation period will also allow for adjustment of the cuts used by the algorithms to flag problematic runs.

Conclusion

The application of automated checks for offline jet data quality based on the comparison of jet variables with reference distributions has been investigated at the ATLAS experiment. Under similar running conditions jet variables have been found to be stable, with a typical χ^2/ndf of 1-3 when compared with a reference. Changing run conditions, as are currently being experience during the commissioning of the LHC, cause significant changes in the jet distributions, making comparison with a reference difficult. With the exception of the jet profile variables, these changes have been well characterized, and are primarily due to variations in the low p_T jets, whose effect can be reduced by applying an offline jet trigger. This trigger allows for effective comparison between runs with different conditions.

Due to the stability of the jet variables, a set of algorithms has been developed to compare jet variables with reference distributions. These algorithms, which are based on simple statistical comparisons proved to be effective for comparing jet variables for runs with similar run conditions. These algorithms have been implemented on the ATLAS jet data quality website for further testing, and it is hoped that once the LHC has achieved stable running conditions, they can be used to fully automate the offline jet data quality analysis.

References:

1. ATLAS Collaboration, *The ATLAS Experiment at the Large Hadron Collider,*2008 *JINST* 3 S08003
2. Stephen Ellis, *What is a Jet?*, CERN Summer Student Lectures, 2010
3. Mario Martinez, *Observation of Energetic Jets*, PLHC Approval Meeting, 2010
4. Peter Loch, *Jet Reconstruction in ATLAS*, PPD Seminar, 2007
5. Michele Petteni, *Jet Reconstruction,* ATLAS Canada Meeting, 2009
6. Mathieu Plamandon, *LAr Status*, ATLAS Canada Workshop, 2010
7. Ian Nugent and Lorraine Courneyea, *Confronting Detector Induced Fake MET*, ATLAS Canada Workshop, 2010