



Validation of ZEUS common ntuples

Mark Baber, University of Bristol, UK

September 6, 2012

Abstract

Although the ZEUS experiment finished data taking in 2007, the data it recorded is still highly relevant to the development of the understanding particle physics. To ensure that with reduced resources this data remains accessible for future analyses, efforts have undertaken to convert the entire dataset to a common ntuple (CN) format. It is of high priority to ensure the validity of the conversion process in maintaining the integrity of the data. This report details the conversion and development of a validation tool to perform checks for consistency across ntuple versions of ZEUS data. A selection of the physics measured at the ZEUS experiment is also presented.

Contents

1	Introduction	2
2	Experimental theory	3
2.1	HERA	3
2.2	ZEUS	3
2.2.1	Tracking and PID	5
2.2.2	DIS	6
3	Common ntuple validation	8
4	ZEUS physics	13
4.1	Strange physics and glueballs	13
4.2	Charm physics	14
5	Conclusion	17
6	Acknowledgements	17

1 Introduction

Beginning its operation in 1992, the ZEUS experiment has provided a greater understanding of the inner structure of the proton. Based at the only ep collider ever built, HERA, the ZEUS experiment was able to perform measurements of precision inaccessible to other conventional colliders and fixed target experiments, making the data ZEUS and its sister experiment H1 invaluable in testing the Standard Model. Operating until 2007, the ZEUS experiment was able to accumulate an integrated luminosity of 0.5 fb. Due to the uniqueness of the ZEUS experiment in colliding electrons and protons, the data record still provides some of the best measurements of the standard model to date.

With a reduction of man power and resources however, it is uncertain whether the collaboration will be able to maintain processing the data centrally. The previous reconstructed data format, MDST, is computationally intensive to produce and maintaining backwards compatibility with new software and operating systems are also a problem. For these reasons, efforts are being made to preserve the vast volume of data recorded by migrating the data to a ROOT common ntuple format, the dominant storage format used by many experiments including at the LHC. Storing the data as ROOT ntuples will therefore ensure backwards compatibility with new software for a significant period, enabling further analyses in the decades to come.

The common ntuples utilise the same scheme used in ORANGE (Overlying Routine for Analysis Ntuple GEneration), to ensure compatibility to compare code with previous analyses and the retention of skills developed in previous analyses. They also require significantly less processing power, with the goal of reducing the size and production time of the ntuples to 10% of that required producing MDSTs with ZEPHYR [4]. The reduced size of the common ntuples allow them to be stored on the DESY dcache in conjunction with tape storage, ensuring fast access and low maintenance. This will allow the old RAW, MDST and Monte Carlo (MC) tapes currently stored in tape racks to be retired and kept in storage, reducing the required storage space from 1 PB to 100 TB [5].

New versions of the common ntuples have been generated in recent years to fix bugs with previous versions, providing reconstruction with improved algorithms and introducing new quantities required by analyses. With the production of ntuples due to be frozen, it is imperative that these new versions be checked for consistency to avoid the introduction of errors and the loss of data that could affect future analyses. An effort to check data persistence was begun by Achim Geiser with a PAW program written in FORTRAN. Due to limitations with the capabilities of PAW however the code was required to be ported to ROOT in C++. This report details the work undertaken in converting the analysis program and the findings from its running.

2 Experimental theory

2.1 HERA

The Hadron Elektron Ring Anlage (HERA) is a 6.3 km circumference ep -synchrotron situated at DESY in Hamburg. In collisions of electrons and positrons with protons its collider experiments, ZEUS and H1, were able to probe the inner structure of the proton and study QCD and electroweak interactions. At its peak, HERA operated with 27.5 GeV electron and 920 GeV proton beams, providing a centre of mass energy of $\sqrt{s} = 318$ GeV. The lesser electron beam energy due to the inverse quartic dependence of synchrotron emission with mass. This asymmetry in the momentum of the proton and lepton beams means the ZEUS detector is also asymmetric, with more instrumentation in the forward region of the detector in the proton beam direction.

2.2 ZEUS

ZEUS has been instrumental in the development of the Standard Model, in particular QCD, with measurements of the scattering of electrons and positrons with protons. The detector utilises an asymmetric hermetic design, shown in Figure 1, with protons entering from the right colliding with electrons from arriving from the left. The momentum imbalance of the beams leads to an asymmetric event, with the majority of emissions directed in the forward region of the detector along the proton direction.

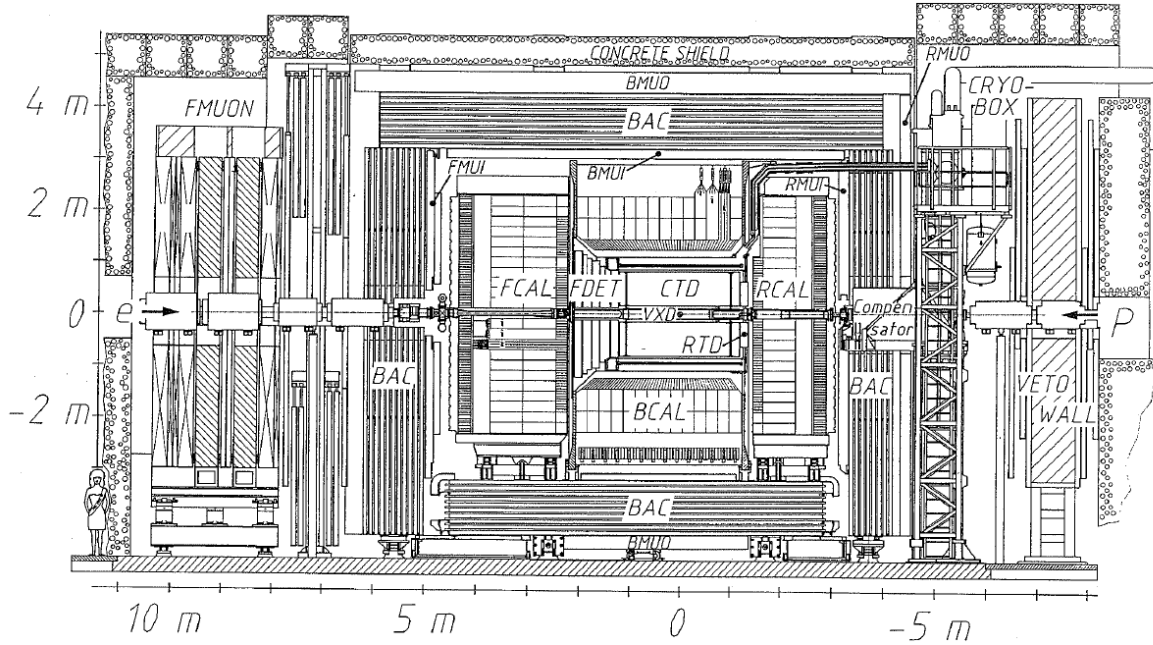


Figure 1: Longitudinal cross-section of the ZEUS detector.

The events in ZEUS are described by a right-handed coordinate system. Centred at the interaction point (IP) the orientation of the axes are chosen such that the x-axis is directed to the centre of the synchrotron ring, the y-axis directly upwards and the z-axis along the beamline in the proton direction. Angles in events are defined in respect to these axes, with the polar angle θ defined as the angle between the particle's direction and the z-axis in the zy -plane. A further angle, the azimuthal angle ϕ , is defined as the angle between the particle direction and the x-axis in the xy -plane. The polar angle is more commonly described by the quantity pseudorapidity η , which is defined: $\eta = -\ln\left(\tan\left(\frac{\theta}{2}\right)\right)$.

A brief overview of the detector components is presented below:

The beampipe exhibits an elliptical design, with the beamspot offset by several centimetres from the true detector centre. This is to allow the installation of photon absorbers in the beampipe to absorb the synchrotron emission from electrons that would otherwise be a significant background. The dimensions and position of the beamspot can be seen in the reconstructed secondary vertex track of Figure 2, which shows interactions with the beampipe enclosure and decays from the interaction point (IP).

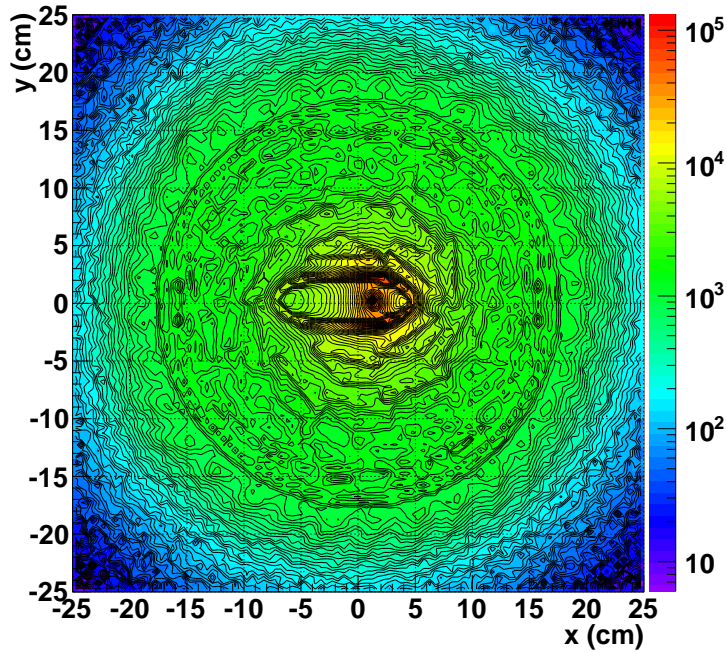


Figure 2: Reconstructed secondary tracks vertices in the xy -plane.

The closest tracking component to the interaction point (IP), the micro-vertex detector (MVD) performs measurements of the vertices of particle tracks with a spatial resolution of $35 - 70 \mu\text{m}$ [1]. This is particularly important in the measurement of secondary vertices to enable the reconstruction of decaying hadrons that contain charm and beauty

quarks. The MVD has three layers around the beamline, which can also be seen in Figure 2, with four additional silicon-strip 'wheels' in the forward region arranged perpendicularly to the beam line.

Enclosing the MVD the central track detector (CTD), a nine layer drift chamber, performs measurements of the tracks of charged particles far from the IP with a spatial resolution of $120 - 130 \mu\text{m}$. Operating in a 1.8 T field from the solenoid, the CTD is able to determine the momenta of particles and their rate of energy loss per propagation length, $\frac{dE}{dx}$, with high resolution [2]. In the forward proton direction the CTD only has a coverage of $\theta > 25^\circ$. In this very forward region the forward tracking detectors (FTD), which have a coverage: $6^\circ < \theta < 25^\circ$, perform measurements of particles. This includes straw tube trackers and drift chambers which enable the precise measurement of tracks in a high radiation environment and a transition radiation detector, which performs electron identification.

Beyond the tracking detectors, the calorimetry performs measurements of particle energies and the detection of neutral hadrons. ZEUS utilises Uranium attenuators in conjunction with scintillating sheets in its electromagnetic (EMC) and hadronic (HAC) calorimeters, to perform measurements of particle energies from hadronic and electromagnetic showers. Particles which fully traverse the inner calorimetry enter the backing calorimeter which utilises the return yoke of solenoid as an attenuator to fully absorb the particle energy. Highly penetrating muons however also traverse this calorimeter and are detected by muon systems that enclose the detector.

The 96 ns bunch crossing rate of the particle bunches in the HERA ring is far greater than the rate at which the electronics and computers can process an event and so the ZEUS detector utilises a three level trigger system [3]. The first level trigger (FLT) is a hardware trigger to select events which enter a pipeline, giving the higher level triggers sufficient time to process the event without detector deadtime. The second level trigger (SLT), performed by microprocessors, further analyses the triggered events and begins the event reconstruction for the final level, the third level trigger (TLT). Performed in software by a processor station, the TLT performs the final analysis on the reconstructed event to determine whether to retain the event and store to disk.

2.2.1 Tracking and PID

The precision of the measurements of a detector are limited by the ability of its subdetectors to perform. An important component of the ZEUS detector is the tracking provided by the CTD and MVD. As well as identifying particles by their particular responses to the individual detector systems such as the shape of energy deposits in the calorimeters, particle identification (PID) at ZEUS is achieved through measurements of a particle's energy loss as it propagates the detector, $\frac{dE}{dx}$, by measurements in the MVD and CTD. This is particularly important in distinguishing high energy particles of similar mass,

such as the pion and kaon, which exhibit almost an identical response in the detector. Shown in Figure 3 the energy loss for the different hadrons is distinctive with a clear separation for low momenta.

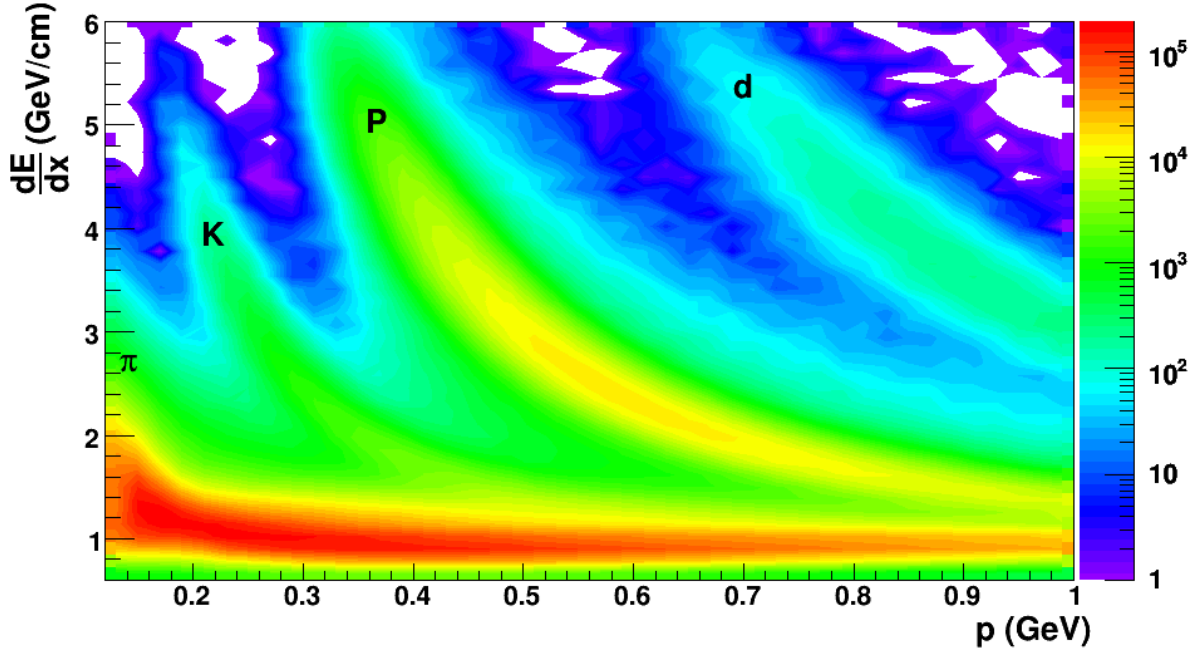


Figure 3: Combined $\frac{dE}{dx}$ from the CTD and MVD. Flatest to steepest: Pion, Kaon then Proton. Just like Bethe-Bloch equation predicts

2.2.2 DIS

At HERA, the ZEUS and H1 experiments studied the deep inelastic scattering (DIS) of the proton with electrons and positrons ($e + p \rightarrow e + X$), enabling the inner structure of the proton to be probed by the exchange of a gauge boson between the incident lepton and proton. Although this scattering can occur both via electromagnetic and weak forces, at the centre of mass energies studied at HERA this scattering is most commonly performed by the electromagnetic interaction with the exchange of a virtual photon.

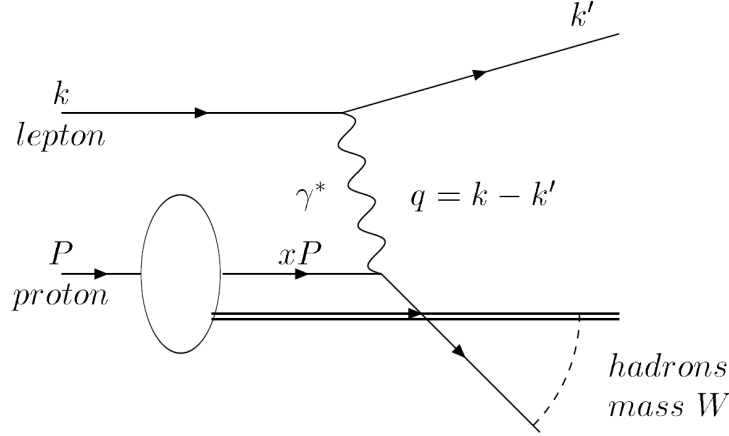


Figure 4: Feynman diagram of DIS

In DIS, high energy collisions between a lepton and hadron enables the incident lepton to resolve the inner structure of the hadron, probing its component quarks. Shown in the Feynman diagram in Figure 4, a lepton and proton with initial four-momenta k and p respectively interact by the exchange of a virtual photon, imparting a momentum of $q = k' - k$ to a quark with proton momentum fraction x . The resulting interaction results in a large deflection of the incident lepton and the fragmentation of the hadron, leading to a signature of a jet and lepton in the forward region of the detector.

The momentum transferred between the lepton and proton is more commonly expressed as its virtuality Q^2 , the negative of the square of the propagator's four-momentum: $Q^2 = -q^2 = -(k - k')^2$. The length scale that can be probed in an interaction is inversely related to the Q^2 value and so in DIS events where the virtuality is typically large with $Q^2 > 1 \text{ GeV}^2$, not only the valence quarks of the proton can be probed but also the sea quarks. Events where the propagator has small virtuality ($Q^2 \approx 0$) are known as photoproduction (PHP). In such interactions individual quarks cannot be resolved but only the proton as a whole.

From the measurement of the various kinematic parameters of the interaction, many features of the hadron can be determined including the parton distribution function (PDF) of the proton, the probability of interacting with a particular parton of the proton with a given x at a specific Q^2 . Calculations at low Q^2 cannot be performed by perturbative QCD theory but must rely on experimental measurement. The HERA experiments had the advantage over fixed target experiments of probing higher Q^2 , and lower x which is particularly important in the modelling of interactions at hadron colliders such as the LHC.

3 Common ntuple validation

The validation code to check ZEUS common ntuples is comprised of two parts: an analysis program and a comparison program. The analysis code reads the common ntuples via dcache and performs selection cuts, producing a myriad of distributions, from comparing machine data such as the trigger bits fired in an event to analysing physical quantities in interactions such as the reconstructed invariant mass. These distributions are stored in an ntuple which is read by a separate comparison program that produces overlay and ratio plots, enabling a visual comparison of different ntuple versions and also between data and MC.

To ensure good events in the comparison, the analysis code performs the following general selection cuts:

- ▷ EVTAK, MVDTAK and STTTAK requirement
- ▷ Cosmic muon cuts
- ▷ Vertex radius < 0.5 cm
- ▷ Cleaning cuts: $\frac{\text{Vertex tracks}}{\text{All tracks}} > 0.1$, Vertex $\chi^2/N_{\text{track}} < 50$
- ▷ Off momentum positron cut: at least 1 track or $E_T(\text{FCAL}) > 0.4$ GeV

Further cuts to separate events are then applied, grouping events in their respective processes including PHP and DIS, the selection cuts for the latter of which are presented below:

- ▷ Trigger: TLT SPP02//SPP09/HFL17/HPP31//HFL2,6,10//DIS01-06,11, SPP01,03,09
- ▷ Electron probability > 0.9 , $E_e > 10$ GeV, $Q_e^2 > 5$ GeV², $\theta > 1^\circ$
- ▷ $44 < E - p_z(\text{ZUFO}) < 65$ GeV
- ▷ $p_T^{\text{miss}}(\text{CAL}) < 8$ GeV
- ▷ Radius cut at 15 cm centred on (-1,0) + detector crack cuts

Where ZUFO is a ZEUS Unidentified Flow Object.

The analysis was run on data and MC samples produced by ARIADNE from the HERA-II 2003-2007 period, however only quantities from the 05e dataset are presented in this report. The results of interesting differences found in these comparisons are presented below.

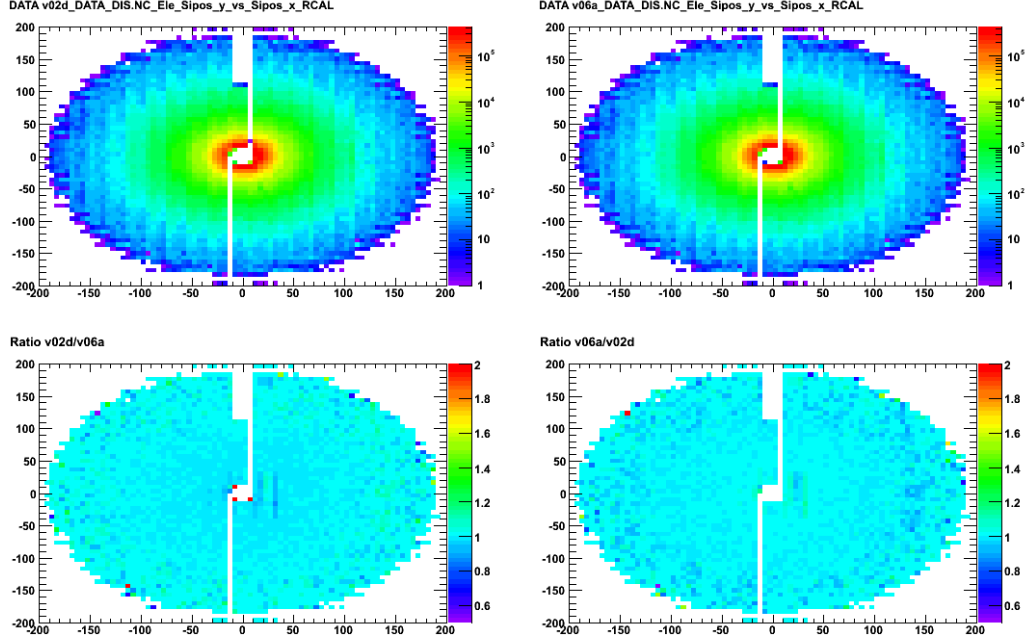


Figure 5: Comparison of electron x vs y distribution in centimetres in the RCAL between ntuple versions v02d v06a for the 05e dataset.

In Figure 5 we see the comparison of the x vs y position of electrons recorded in the rear calorimeter (RCAL) between CN v02 and v06, which illustrates the standard output of the comparison code for 2D plots, producing a visual comparison of the different versions in the top left and right regions respectively. A ratio and its inverse are displayed in the lower region, enabling features of the distribution that differ between versions to be easily detected. This comparison shows that the distributions are in good agreement over RCAL on a large scale between versions v02 and v06 however on the inspection of the inner RCAL, differences can be discerned. Seen in Figure 6 there is a visible difference between the v02 and v06 versions, with bins being 'displaced'. which is likely due to the change in position calibration of the RCAL that occurred between versions v02 and v06.

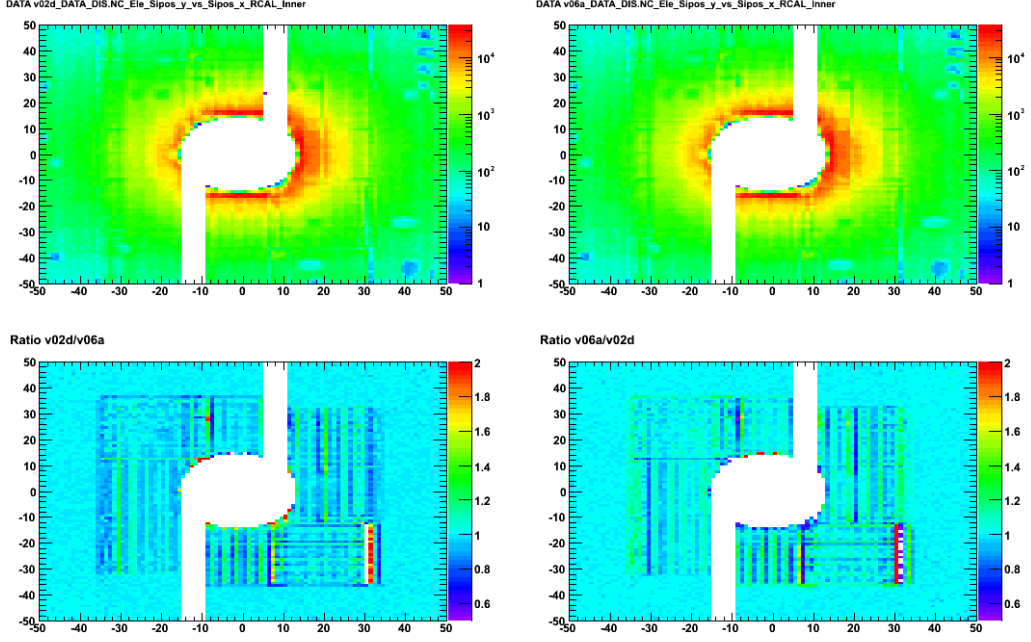


Figure 6: Comparison of electron x vs y distribution in centimetres in the inner RCAL for ntuple versions v02d v06a for the 05e dataset.

Another interesting difference found in the comparison was the difference in the track $\phi - \eta$ distributions between ntuple versions. This distribution was very similar for standard tracks however for tracks found in the CTD without matching hits in the MVD there was a noticeable difference as can be seen in Figure 7. It can be seen that in v06, the number of tracks with no MVD track is significantly smaller in some regions of the detector whilst other regions are unchanged. This difference can be explained by the improved CTD-MVD matching implemented in v06, which results in v06 having fewer tracks with no MVD association, reducing the number of events in the v06 distribution. This also explains the features seen in the distribution as arising to regions of the detector where MVD matching is better, partially owing to the asymmetric positioning of the MVD around the beam pipe resulting in some areas having three layers of MVD module coverage and others having just two. Similar features can also be observed in Figure 8 with the comparison of v06 with the latest ntuple version, v07. This further reduction in the number of tracks without MVD hits could also be due to further improvements in MVD matching in v07 although this is not properly understood.

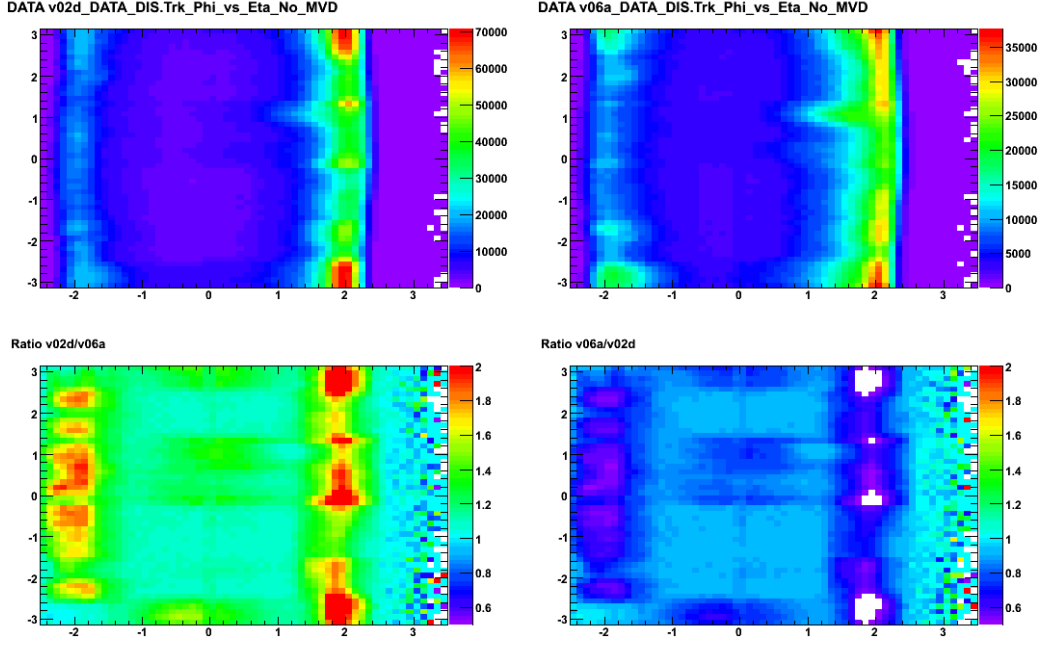


Figure 7: Comparison of reconstructed track ϕ vs η with no MVD hits between ntuple versions v02d v06a for the 05e dataset.

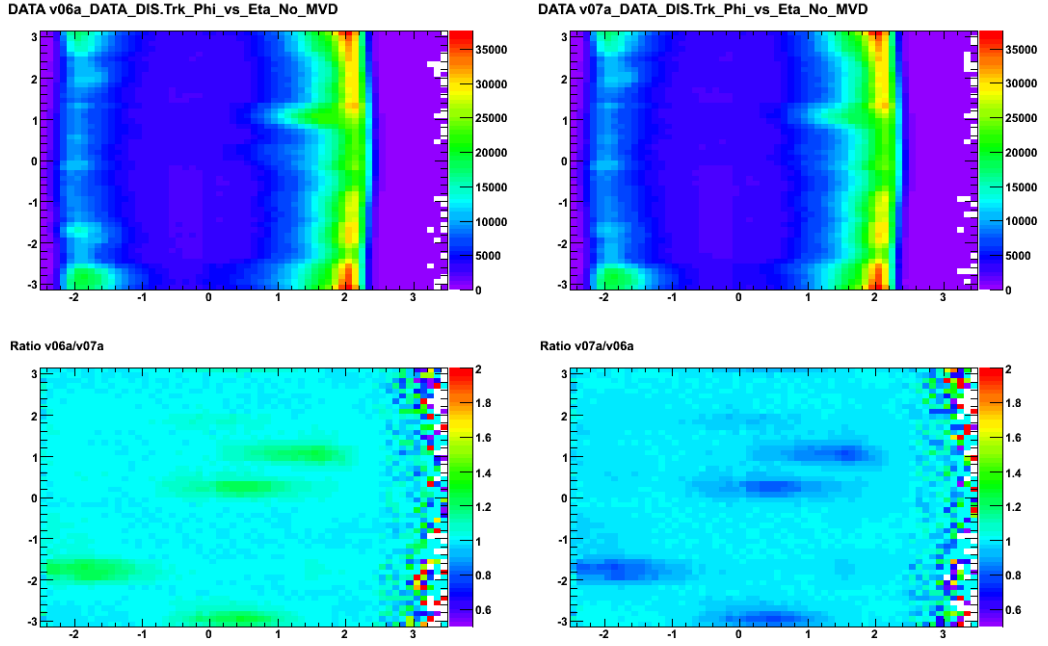


Figure 8: Comparison of reconstructed track ϕ vs η with no MVD hits between ntuple versions v06a v07a for the 05e dataset.

Non-vertexed tracks were also found to display unusual differences between versions v02 and v06, with more tracks in v06 being reconstructed with low ϕ . This appears to be due to a bug in the definition of ϕ in v02 as the tracks appear to be displaced by π in the later version. The effect of this change appears to improve the $\phi - \eta$ distribution in v06, which forms a more continuous distribution than in v02 where most events are grouped at large $|\phi|$. A comparison the distributions with MC would have to be performed to determine whether this has indeed improved the quality of reconstruction.

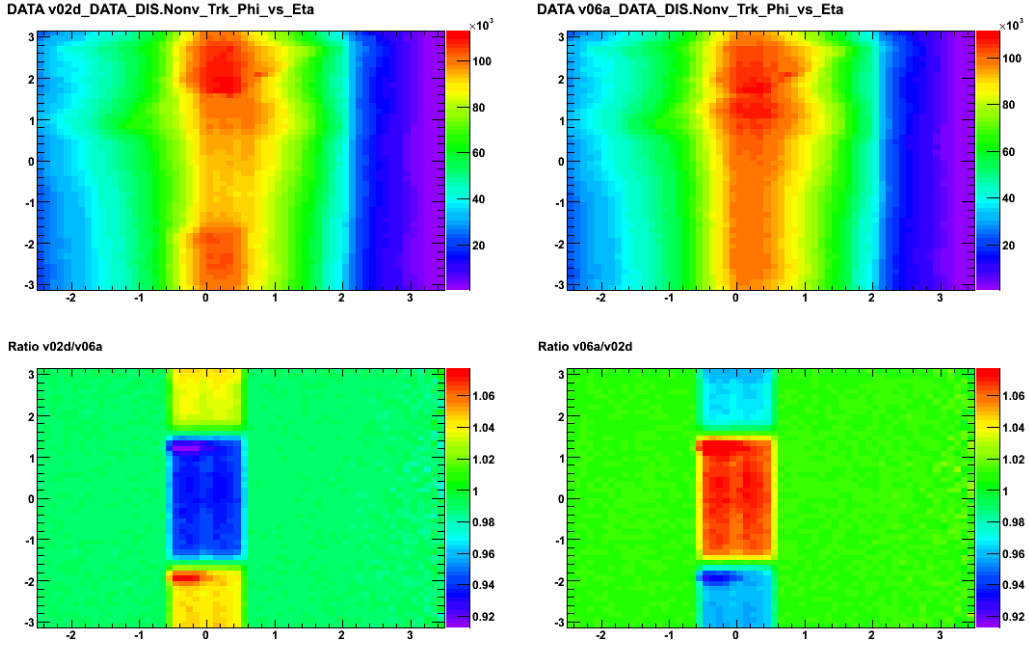


Figure 9: Comparison of reconstructed track ϕ vs η with no reconstructed vertex between ntuple versions v02d v06a for the 05e dataset.

In general the distributions compared in the CN were found to be in good agreement between different versions, with some differences arising due to known changes and improvements to the reconstruction and calibration. There were however some differences that could not be fully explained which should be investigated to determine whether they arise due to some known change in the ZEUS software reconstruction or a bug in CN production.

4 ZEUS physics

Another interesting aspect of the validation code is the underlying physics of the distributions that are compared. Below is a small selection of the wide field of physics that the validation software probes.

4.1 Strange physics and glueballs

K_S^0 mesons are particularly interesting in the search of glueballs, a hypothetical colourless bound state of gluons predicted by lattice QCD. The lightest such candidate is a scalar in the mass range: 1450 - 1750 MeV and so may decay to two K_S^0 mesons. For this reason resonance searches in the $K_S^0 K_S^0$ system are of utmost interest from which studies have produced a candidate, the $f_0(1710)$. This resonance however cannot be a pure glueball state as it has been found to also couple photons [6]. Being able to reconstruct the K_S^0 efficiency is therefore very important in the search for resonances from $K_S^0 K_S^0$ pairs.

The K_S^0 candidates in the code are selected from $\pi^+ \pi^-$ pairs that originate from a secondary vertex. The mass of the mother was reconstructed by taking the invariant mass of this pion pair, $M(\pi^+ \pi^-)$. To suppress contributions from backgrounds, the following selection cuts were applied [6]:

- ▷ $M(e^+ e^-) > 50 \text{ MeV}$
- ▷ $M(p\pi) > 1121 \text{ MeV}$
- ▷ $p_T(K_S^0) > 0.25 \text{ GeV}, \quad |\eta(K_S^0)| < 1.6$
- ▷ $\theta_{2D} < 0.12 \text{ rad}, \theta_{3D} < 0.24 \text{ rad}$

Where $M(e^+ e^-)$ is the invariant mass of the candidate tracks when assigned with electron masses, a cut designed to suppress photon $e^+ e^-$ pair production. Similarly $M(p\pi)$ is the invariant mass of the tracks with the proton and pion masses substituted which is designed to remove Λ and $\bar{\Lambda}$ contamination, which have masses 1116 GeV. The final cuts are selections on the collinearity of the K_S^0 momentum and propagation vectors, with θ_{2D} and θ_{3D} defined as the angles between the K_S^0 momentum vector and the vector from the IP to the secondary vertex of the K_S^0 in xy -plane and in three dimensions respectively.

The reconstructed K_S^0 invariant mass spectrum prior to cuts can be seen in Figure 10 which shows a clear peak at $\approx 500 \text{ GeV}$ due to K_S^0 . There is also a peak at $\approx 300 \text{ GeV}$ which is attributed to two low energy background pions. With the application of the selection cuts, seen in Figure 11, the background is significantly reduced and the signal peak becomes more distinct yielding a consistent with measurements by the PDG.

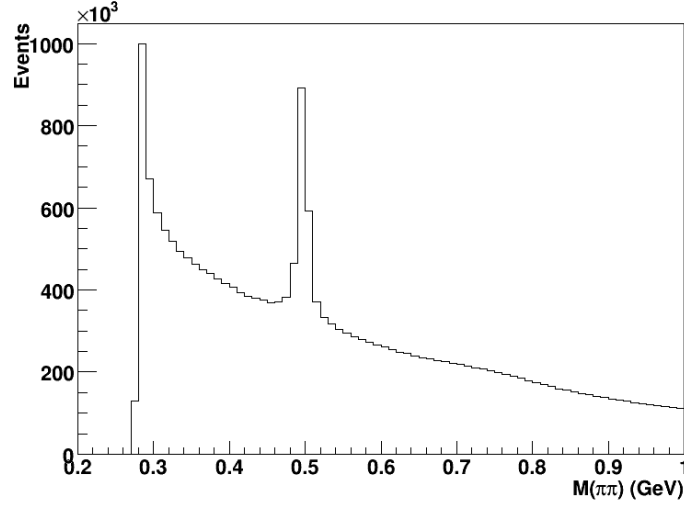


Figure 10: Invariant mass of reconstructed $\pi\pi$ candidates prior to selection cuts.

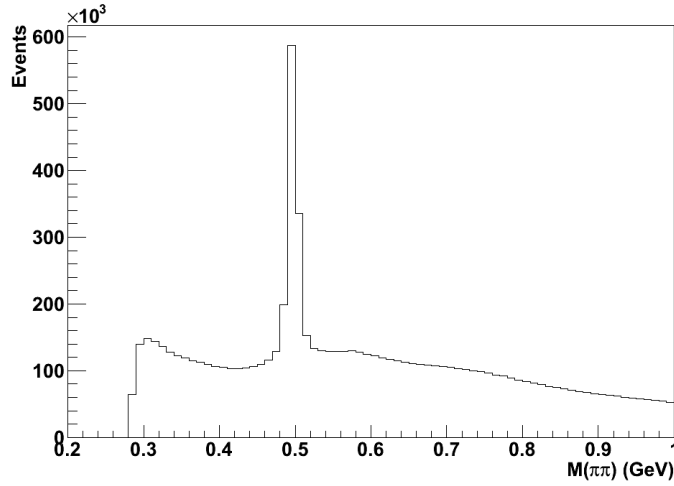


Figure 11: Invariant mass of reconstructed $\pi\pi$ candidates after selection cuts.

4.2 Charm physics

Another crosscheck that can be performed between CN versions is the reconstruction of the mass difference between the $D^{*\pm}(2010)$ and D^0 mesons, which due to unique kinematics in the decay of the D^* can be achieved to high precision. When produced, the D^* rapidly undergoes the decay:

$$D^{*+} \rightarrow D^0 \pi_s^+ \quad (1)$$

Due to the small mass difference between D^{*+} and D^0 , the pion created has low energy and is known as a soft pion, π_s^+ . The very low momentum imparted to the pion means that the mass difference between the two mesons can be ascertained very accurately. The mass of the D^0 can be obtained by reconstructing events with the either of the subsequent decays (or their respective charge conjugates):

$$\text{Channel (1)} \quad D^0 \rightarrow K^- \pi^+ \quad (2)$$

$$\text{Channel (2)} \quad D^0 \rightarrow K^- \pi^- \pi^+ \pi^+ \quad (3)$$

Therefore by reconstructing the invariant mass of final states with $K\pi\pi$ and $K\pi\pi\pi\pi$ with zero total charge, the mass of the D^0 can be obtained and hence the mass difference between D^* and D^0 mesons, $\Delta M = M(D^0 \pi_s^+) - M(D^0)$, can be determined. As pions and kaons are produced copiously in events, any combination of background pions and kaons can satisfy this criterion. To reduce random backgrounds therefore, the following selection cuts are applied:

$$\triangleright p_T(K) > 0.5 \text{ GeV}, p_T(\pi) > 0.5 \text{ GeV}$$

$$\triangleright p_T(\pi_s) > 0.125 \text{ GeV}$$

$$\triangleright 0.1435 < M(D^0) < 0.1475 \text{ GeV}$$

$$\triangleright |\eta(D^*)| < 1.6$$

Additional cuts on the properties of the D^* are also applied according to the respective decay channel:

$$\triangleright \text{Channel (1)} : p_T(D^*) > 1.5 \text{ GeV}, p_T(D^*)/E_T^{\theta > 10} > 0.15$$

$$\triangleright \text{Channel (2)} : p_T(D^*) > 2.0 \text{ GeV}, p_T(D^*)/E_T^{\theta > 10} > 0.20$$

Where $E_T^{\theta > 10}$ is the measured transverse energy in the calorimeters excluding the energy of the DIS electron and energy recorded in the FCAL, which covers the forward cone: $\theta < 10$.

The results of the preceding analysis can be seen in Figures 12 and 13. The reconstructed invariant mass of the $M(D^0)$ reconstructed by the validation code produces a resonance consistent with the PDG average: $M(D^0) = 1864.91 \pm 0.17 \text{ MeV}$ and a distinct peak in the ΔM distribution also in agreement with the PDG averaged mass difference of: $\Delta M = 145.421 \pm 0.010 \text{ MeV}$.

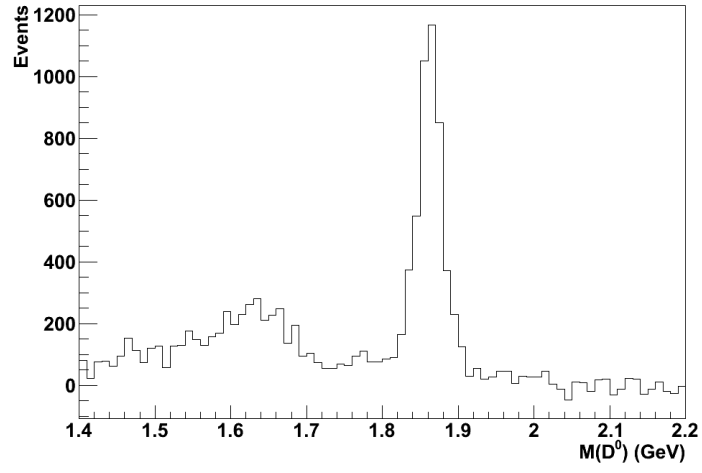


Figure 12: Reconstructed invariant mass plot of the D^0 .

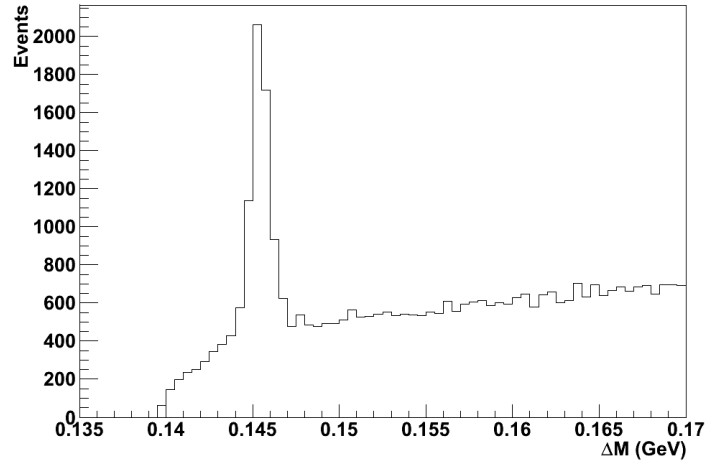


Figure 13: Distribution of the measured mass difference between the D^* and D^0 mesons.

5 Conclusion

Over the duration of the summer student project I have developed a common ntuple validation tool that enables differences between CN production versions to be observed, an important task for the ZEUS offline team. The comparisons made show that the different ntuple versions are generally in good agreement with each other with some differences arising both to known changes in the CN production and also some factors that are yet to be determined. With further development by Andrii Gizhko and Achim Geiser this code should enable a thorough test of CN versions, which should prove particularly useful in the testing of the final common ntuple version, v08, which is due to start production soon.

6 Acknowledgements

The author would like to thank Achim Geiser for his guidance throughout the project and expertise of the ZEUS experiment, which was crucial in interpretation the results. Also thanked are the summer student coordinators for providing the opportunity to work with the ZEUS collaboration and participate in the summer student program.

References

- [1] E. N. Koffeman, Experience with the ZEUS micro-vertex detector, Nucl. Instrum. Meth. A **582** (2007) 705.
- [2] D. Bailey and R. Hall-Wilton, Experience with the ZEUS central tracking detector, Nucl. Instrum. Meth. A **515** (2003) 37.
- [3] H. Boterenbrood, V. O'Dell, H. A. J. R. Uijterwaal, J. C. Vermeulen, L. W. Wiggers, R. van Woudenberg, M. Hazumi and M. Nakao *et al.*, The ZEUS first and second level trigger, In Orlando 1992, Proceedings, Nuclear science and medical imaging, vol. 1 335-337. and Amsterdam NIKHEF - NIKHEF-H-93-01-B (93,rec.Jun.) 3 p. C
- [4] http://www-zeus.desy.de/~ifhuta/ZEUS_ONLY/zephyr/
- [5] J. Szuba [DESY Data Preservation Group Collaboration], HERA data preservation plans and activities, J. Phys. Conf. Ser. **331** (2011) 072032.
- [6] S. Chekanov *et al.* [ZEUS Collaboration], Inclusive $K_0(S)K_0(S)$ resonance production in ep collisions at HERA, Phys. Rev. Lett. **101** (2008) 112003 [arXiv:0806.0807 [hep-ex]].
- [7] A. Aktas *et al.* [H1 Collaboration], Evidence for a narrow anti-charmed baryon state, Phys. Lett. B **588** (2004) 17 [hep-ex/0403017].