Consistency check of common nTuple versions

Teresa Hönigl, University College London, UK

September 3, 2013

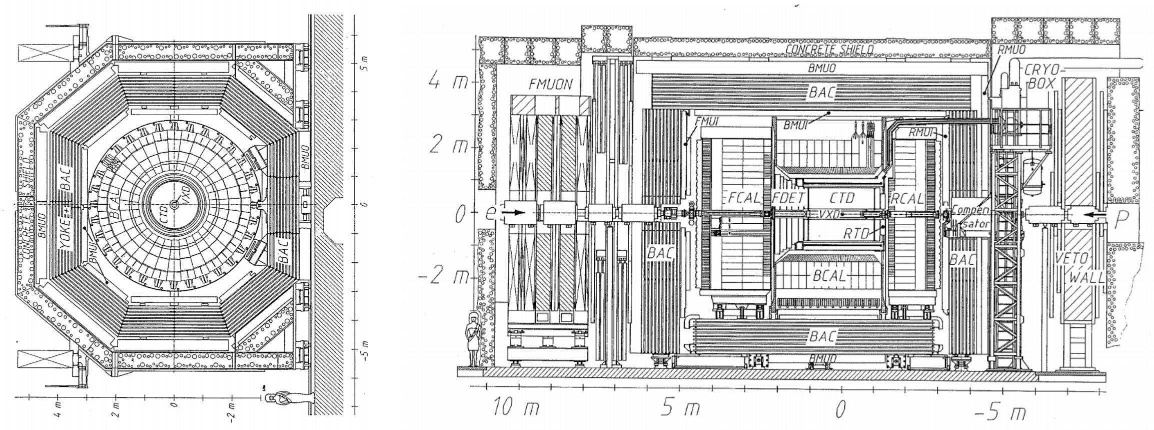
Abstract

The Hadron Electron Ring Accelerator, HERA, is the only storage ring in the world that collided protons with electrons. The electron proton interactions are very useful to probe the structure of matter and therefore provide tests on QCD. Although HERA data taking has stopped in 2007, data analysis is still on-going. Common nTuples are the main and only supported tool for analysis in ZEUS.

The latest version, v08, of the ZEUS nTuple has been checked for consistency. This was done by comparing data with data, Monte Carlo with Monte Carlo and data to Monte Carlo. The Monte Carlo used was the Ariadne Inclusive DIS sample. A selection of the output histograms is presented in this report.

It was concluded that version v08 is ready to use for analysis.

Contents



(a) Cross section of the ZEUS detector perpendicular to the beam

(b) Cross section of the ZEUS detector along the beam

Figure 1: Overview of the Zeus detector

1 Introduction

Although HERA has been shut down in 2007, analysis is still on going. Since the man power available is further decreasing, data storage is of great importance and should be made easily accessible.

1.1 Experimental set up

1.1.1 HERA

The Hadron Electron Ring Accelerator, HERA, is the only storage ring in the world that collided 820 GeV protons with 30 Gev electrons. The electron proton interactions are very useful to probe the structure of matter and therefore provide tests on QCD.

1.1.2 The Zeus detector

Since the installation of the ZEUS detector in 1992 there have been several updates on the components of the detector.
The essential elements are a vertex detector (VXD), a central track detector (CTD) plus transition radiation detector (TRD), and planar drift chambers (FTD, RTD) in the field of a thin magnetic solenoid (SOLENOID), an electromagnetic (EMC) and a hadronic calorimeter (HAC) surrounding the coil over the full solid angle, a backing calorimeter (BAC), barrel and rear muon detector (MU), and a forward muon spectrometer (FMU) [2].
In figure 1 a cross section of the ZEUS detector is shown.

1.2 ZEUS analysis - Introducing ZEUS nTuples

The infrastructure for running EAZE jobs and using ADAMO/ZEBRA/zsmsm on a computer farm, reading MDST's is no longer centrally maintained.[3]. Problems with this previos reconstructed data format, MDST, arise with maintaining backwards compatibility with new software and operating systems due to it being computationally intensive. Hence, a common ZEUS nTuple was designed to replace this functionality in the long term.

The object oriented framework used is ROOT which is a C++ replacement of the popular PAW program developed at CERN and the preferred choice for large scale data analysis [4]. NTuples are a set of files containing variables for every event in a sample of events and are a common storage format also used in a number of other experiments such as the LHC.

Migrating data to a ZEUS NTuples format provides several advantages. Since ROOT is the main analysis tool for large scale data analysis, the ZEUS nTuple is expected to be easily maintained in the long term. Furthermore, the Ntuples are based on the existing ORANGE (Overlying Routine for Analysis NTuple GEneration) ensuring backwards compatibility and making use of the expertise and knowledge accumulated in the ZEUS analyses over the last decade.

It is of great importance that the ZEUS nTuples contain enough information to allow any kind of physics analysis. Therefore the involvement of all physics groups was essential in the design of the ZEUS nTuples.

Storing root ntuples on dCache is a long term storage scheme with minimal maintenance. The size and the production time of the ZEUS nTuple is of the order 10% of MDST production. It is possible to produce mini-tuples containing a subselection of events and information according to the user's interests. These mini-ntuples can be stored on private disks of the user.

Migrating to a common ZEUS nTuple is therefore an economic use of resources and an efficient and extremely versatile tool for physics analysis.

2 Consistency Check of common nTuple versions

Common nTuples are the main and soon the only supported tool for analysis. Over recent years, versions of common nTuples have been updated and newly generated in order to fix bugs, to provide reconstruction with improved algorithms and to introduce new quantities required by analyses.

With the production of nTuples due to be frozen, it is crucial to check these common nTuple versions for consistency to avoid the introduction of errors and the loss of data that could affect future analyses [5]. Efforts to check data persistence with a PAW program written in FORTRAN were made by Achim Geiser. Due to limitations with the capabilities of PAW however the code was converted to ROOT in C++ by Mark Barber, last year's summer student.

This report details further consistency checks of common nTuple versions v06 and v08 and discusses the differences found between the two versions.

The validation code to inspect the ZEUS common nTuples consists of an analysis program and a comparison program. The analysis code (main.C or readTuples.C for HeraI) reads the common nTuple via dcache and performs a standard low Q^2 DIS selection:

General cuts (both data and MC)

- require EVTAKE and MVDTAKE (no STTTAKE, MBTAKE,...)
- ($|zvtx| < 30$ cm)
- vertex tracks/all > 0.1

DIS cuts

- trigger: SPP02//SPP09//HFL17/HPP31//HFL2,6,10
- sinistra p > 0.9, $E_e > 10$ GeV, $Q^2{}_e > 5\text{GeV}^2$, theta > 1
- v06: $44 <$ E-pz (zufo) < 64 GeV
- v08: $47 <$ E-pz (zufo) < 68 GeV

The EVTAKE vairable holds information about quality of data in a particular run and about the performance of particular detector components.

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, is then produced. These distributions are stored in an ntuple which is read by a separate comparison program (compareFiles.C) that produces overlay and ratio plots, enabling a visual comparison of different ntuple versions.

2.1 Data vs Data Comparison

Data-data comparisons of each year from 2003 up to 2007 were made. Having examined the consistency of the visualised results of each year seperately, all data sets have been merged allowing a full comparison between data from versions v06d and v08b to be made.

A selection of the histograms produced by the comparison program are shown below. For reference, a greater selection of plots can be viewed in the presentation slides of the talks that have been hold in the weekly ZEUS physics meetings.

In figure 2 the expected difference in the E-pz cut from ZUFO's (Zeus Unidentified Flying Object) is shown. This cut was altered between version v06 and v08 due to changes in the electron correction.

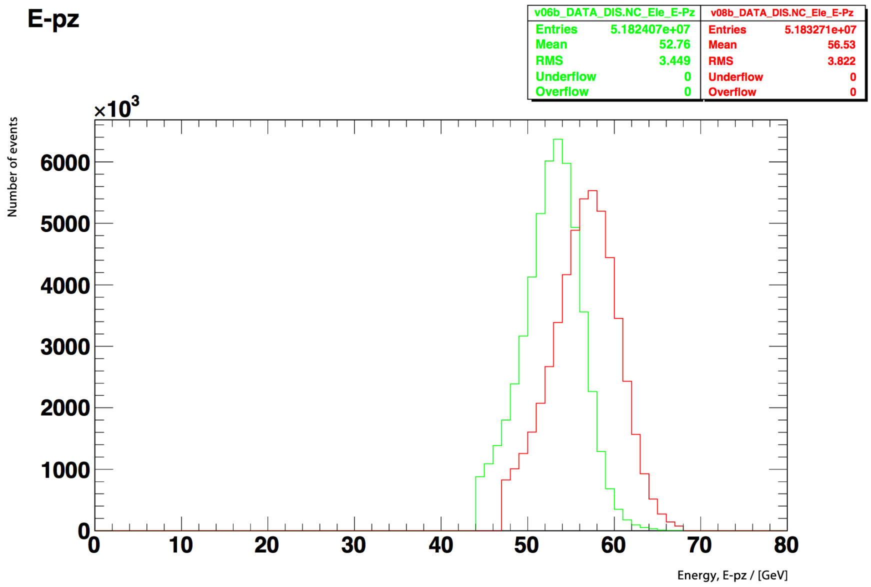


Figure 2: is illustrating the difference in the E-pz cut from zufos that is due to the change in electron correction

Figure 3 shows the secondary vertex tracks and illustrates the standard output of the comparison program of 2D histograms. On the top left and top right a visual presentation of versions v06 and v08 are displayed respectively. For ease of comparison ratios of the two top plots are shown on the bottom. From the ratio plots it can be seen that the number of secondary vertices in version v06d is slightly smaller than in version v08b implying improvements in tracking in the latter version.

Another difference between the two versions is illustrated in Figure 4 where the chi squared per number of degrees of freedom was plotted against the number of entries. Due to the change in zufos the number of jets in version v08b has increased leading to a change in normalisation and a greater number of vertices in version v08b
In figure 5 the unlike dimuon mass was plotted against the number of candidates where dimuon just means that we are dealing with a muon pair and where unlike refers to the opposite charge of the muons. The J/ψ peak at a mass of about 3.1 GeV and the ψ' are

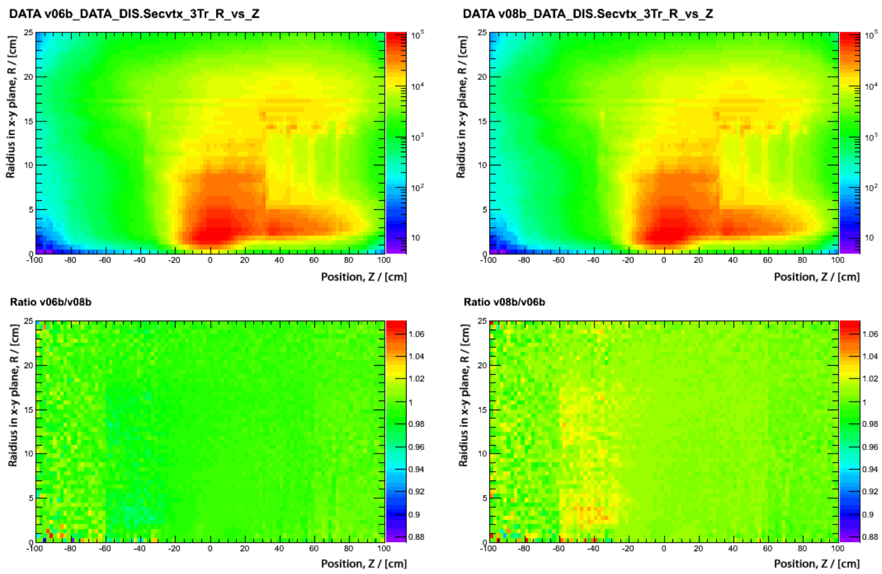


Figure 3: is showing the secondary vertex tracks. It can be seen that there are slightly more secondary vertices in version 8. Thus, it can be concluded that there have been improvements in tracking.

visible and the two versions are found to be in good agreement with each other.

To summarise, differences in version v06 and v08b include the number of jets, improvements in tracking and the energy correction in zufos. Other than the differences detailed above version v06d and v08b are in good agreement with each other.

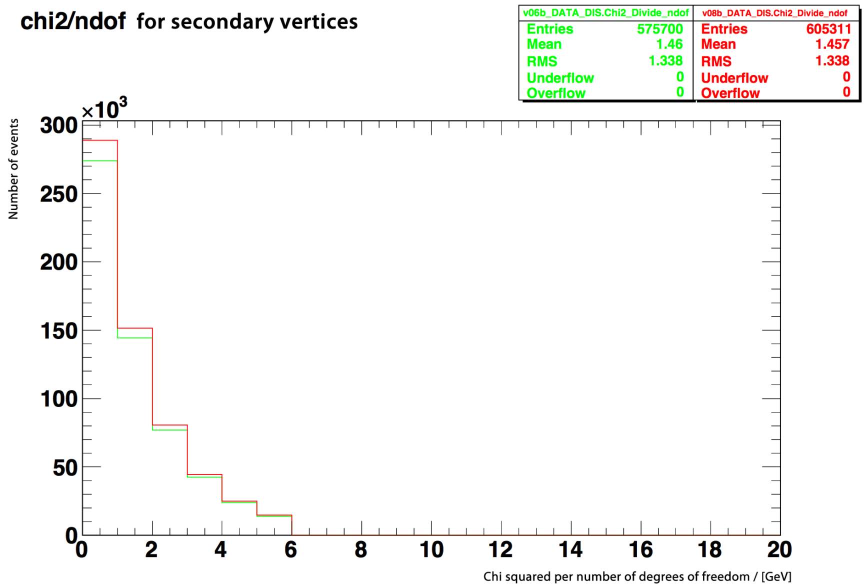


Figure 4: shows that there are more vertices for v08. Due to the change in zufos the number of jets in v08 is greater leading to a change in normalisation.

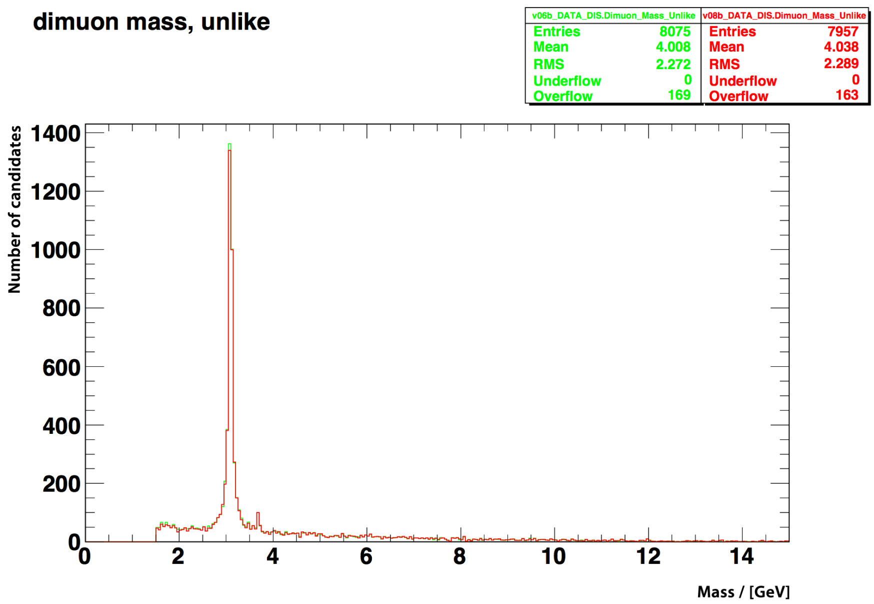


Figure 5: Dimuon mass was plotted against the number of candidates. The J/ψ and the ψ' peaks are visible and the two versions are in agreement with each other.

2.2 Monte Carlo vs Monte Carlo Comparison

The Ariadne Inlcusive DIS Monte Carlo samples were used in all the comparisons. A similar method as before was implemented in the Monte Carlo - Monte Carlo comparison: Each year of version v06b was compared seperately to each year of version v08b to be able to spot discrepancies between years before merging all years. This method allows to deduce the origin of differences between the two versions more easily.

Again, figure 6 illustrates the different E-pz cut from zufos in version v06b and v08b due to changes in the electron correction.

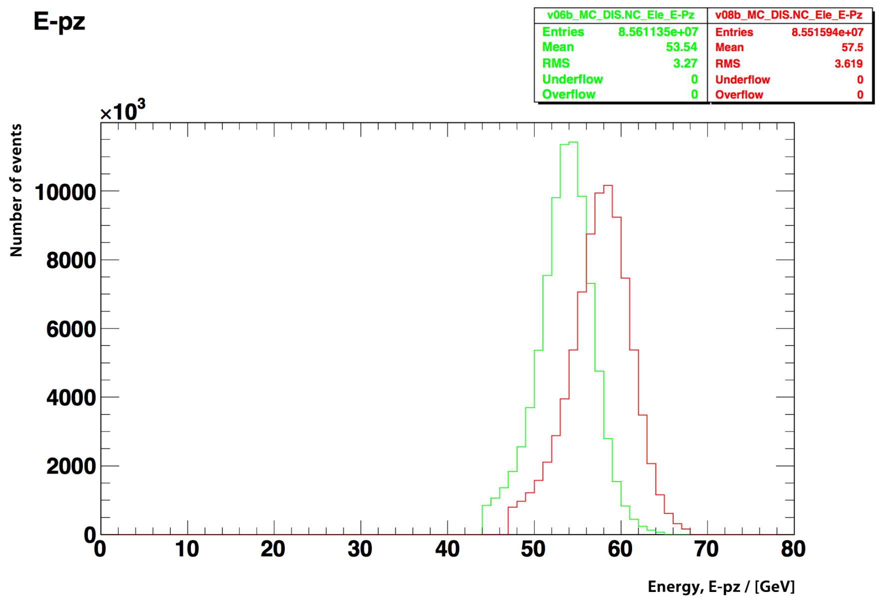


Figure 6: is illustrating the difference in the E-pz cut in zufos that is due to the change in electron correction.

Figure 7 is showing the secondary vertex tracks. In the data-data comparison we have observed an improvement in tracking in version v08b. However, this improvement has not been simulated in the Monte Carlo. As it can be seen in figure 7 both versions are in good agreement with each other.
Figure 8 demonstrates that there have been more vertices simulated for v08b for the same reasons as before.

To conclude, the Ariadne Inlcusive DIS Monte Carlo samples from years 2003 up to 2007 were merged and a full comparison between version v06b and v08b was made. Similar qualitativ difference between the data-data and Monte Carlo - Monte Carlo comparisons were found.

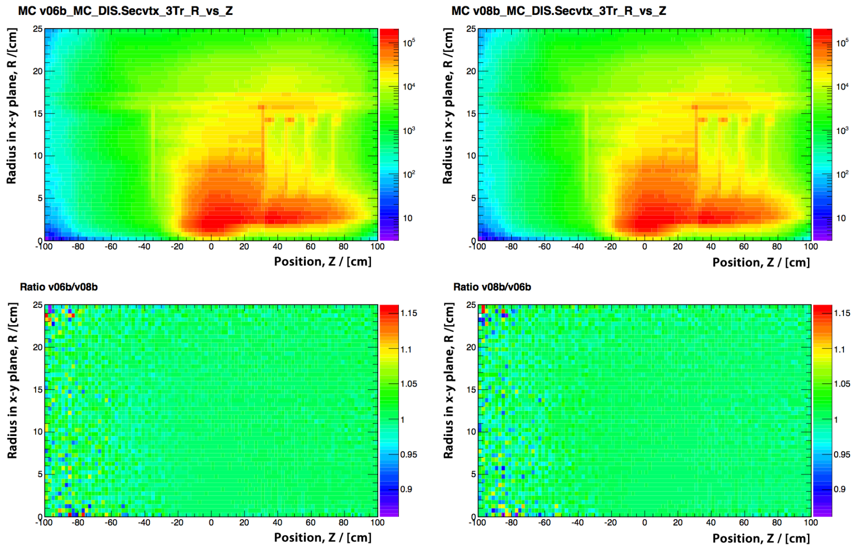


Figure 7: is showing the secondary vertex tracks

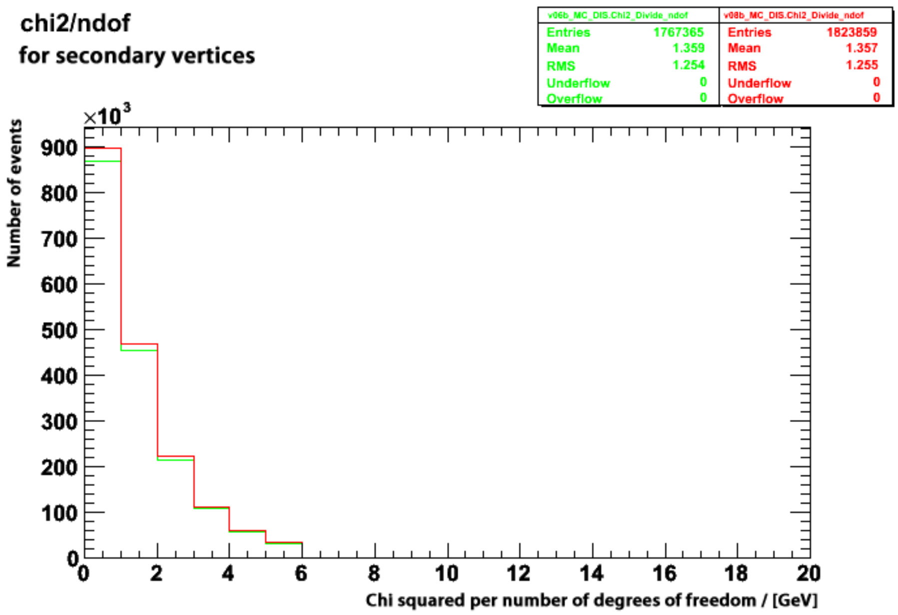
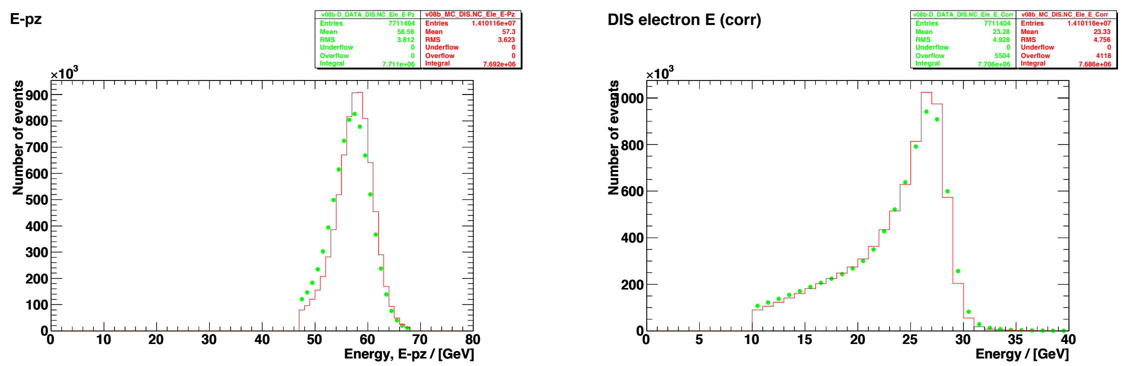


Figure 8: shows that there are more vertices for v08. Due to the change in zufos the number of jets in v08 is greater leading to a change of normalisation.



(a) is showing a difference of about 1% in the E-pz energy scale from zufos

(b) Electron energy from the calorimeter after corrections

Figure 9: is showing the energy scale from zufos and from electrons from the calorimeter after corrections

2.3 Data vs Monte Carlo Comparison

Finally, data was compared to the Ariadne Inclusive DIS sample. Data and Monte Carlo from each year of version v06 and version v08 was compared to each other. There have been histogram overflow and DIS trigger simulation problems that are still being investigated. Thus, year 2006e of version v08b was selected to showcase the data - Monte Carlo comparison results since these effects have not been that distinctive in year 2006.

Standard low Q^2 DIS Selection cuts have been applied as before. Data is being displayed as green marker dots and for the Monte Carlo a red line has been drawn.

In figure 9, it can be seen that there is a difference of about 1% in the E-pz energy scale from zufos. After corrections the data and Monte Carlo of the electron energy scale from the calorimeter are in good agreement. Hence, it rather should be cut on this quantity than the E-pz energy scale from zufos.
Another interesting plot is shown in figure 10 where Eta has been plotted against the number of entries. From this histogram, it can be inferred that the forward and backward efficiency is better in the Monte Carlo than in the data.
Figure 11 illustrates that there have been more secondary vertices simulated in the Monte Carlo.
It was found that Evtake, Tpoltake, Lpoltake and Tag6take are not correctly simulated and/or filled. This can be seen in figure12.
The exclusive pipi mass, meaning that there are exactly two tracks and there have not been any cuts on elasticity, is shown in figure 13. This plot features some interesting physics. The first peak is the ρ meson peak and can be seen in both data and Monte Carlo whereas the reflection of the J/ψ peak is not simulated in this Monte Carlo. There

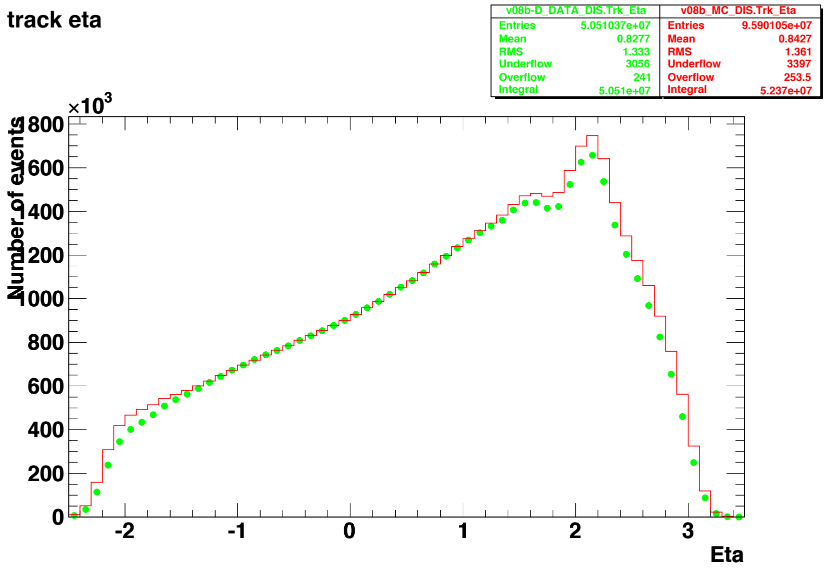


Figure 10: is showing that the forward/backward tracking efficiency is better in the Monte Carlo

are further interesting bumps in the data between the log mass of 0 and 0.4 that could be analysed and identified.
The data-Monte Carlo comparison provides the most interesting comparison since one can see how well the simulation describes the actual data. Some of the differences observed were expected since the data is not described perfectly by the Monte Carlo. Overall it was found that v08 is consistent and that is is ready to use for analysis.

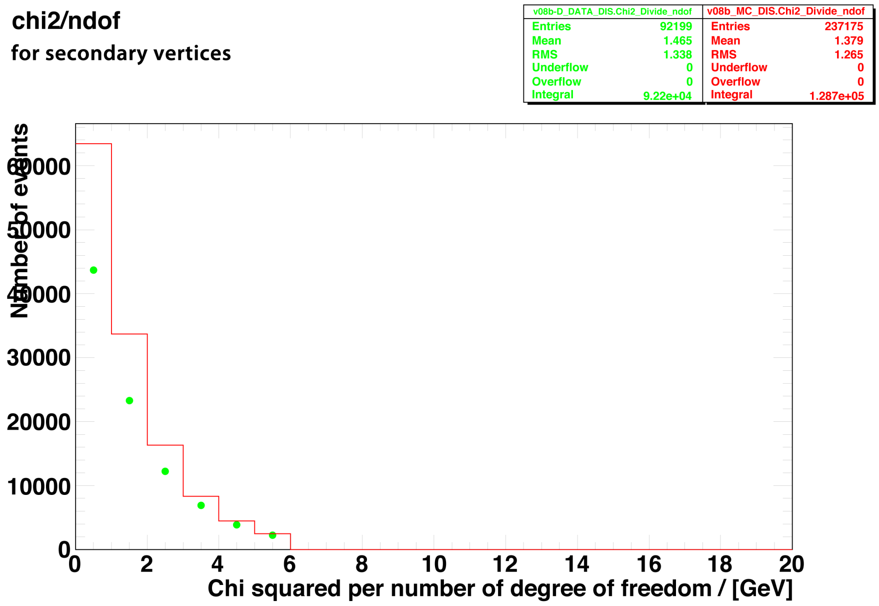


Figure 11: shows that there are more vertices simulated in the Monte Carlo.

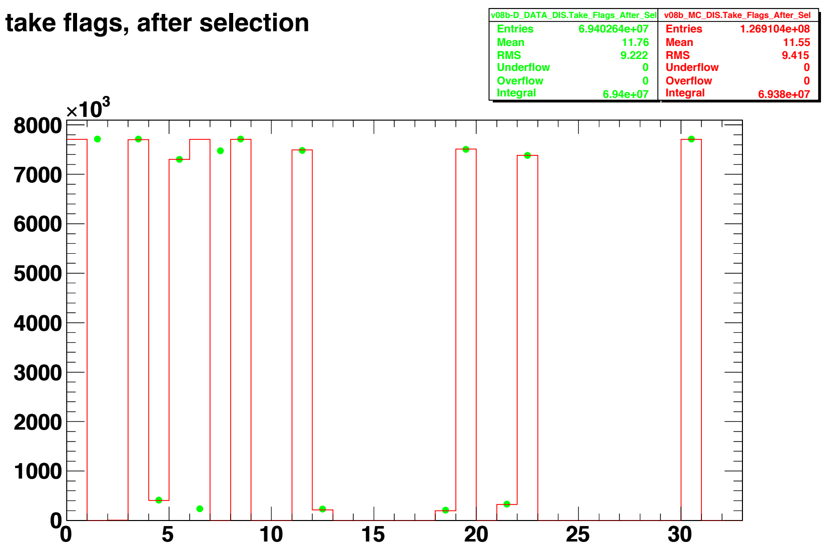


Figure 12: is showing Evtake from 0-1, Mvdtake from 2-3, Sttake from 4-5, Tpoltake from 6-7, Lpoltake from 8-9, Evtake iwant from 10-16, Mbtake from 17-19, Fmutake from 20-23 and Tag6take from 30-31.

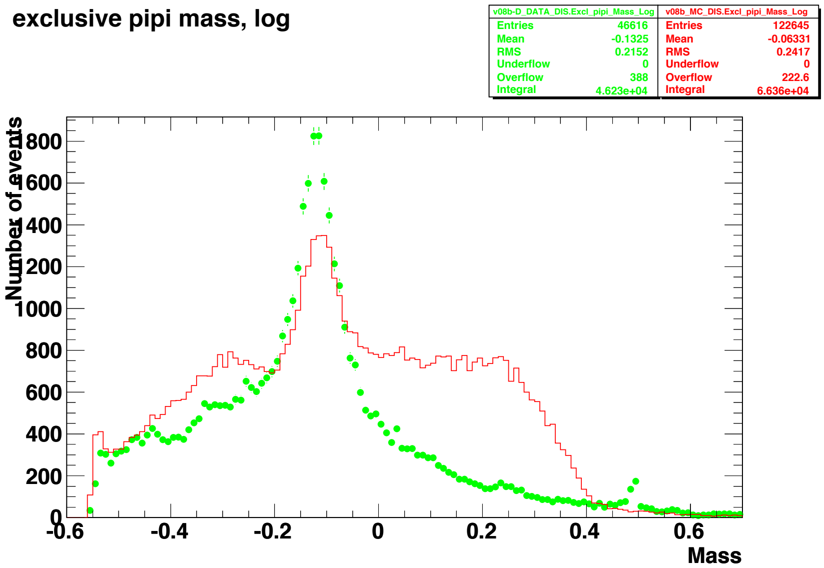


Figure 13: is showing the exclusive pipi mass (exactly two tracks, no elasticity cut). The first peak is the ρ meson peak and can be clearly seen in both data and Monte Carlo whereas the reflection of the J/ψ peak is not simulated in this Monte Carlo.

3 Conclusion

Since common nTuples are soon the only supported tool for analysis it is key to check the latest version, v08, for bugs and consistency. Data-data, Monte Carlo-Monte Carlo and data-Monte Carlo comparisons have been conducted.
The Monte Carlo sample used was the Ariadne Inclusive DIS sample. Data and Monte Carlos of years 2003 up to 2007 have been merged to allow a comparison with larger statistics. Similar qualitativ differences between the data-data and Monte Carlo-Monte Carlo comparisons were found. Differences in version v06 and v08 include the number of jets, improvements in tracking and the energy correction in zufos.
In the data-Monte Carlo comparison problems with histogram overflow and DIS trigger simulation arose. These issues are still being investigated.
Despite those minor issues, v08 is ready to use for analysis.

4 Acknowledgements

I'd like to thank my supervisor, Achim Geiser, and PhD student, Andrii Gizhko, for guidance throughout the project. They have never been reluctant to asnwer any questions and their expertise in the field have been crucial in the interpretation of histograms and results.
Also a big thank you to the summer student programme coordinators for providing the opportunity to work within a research group at Desy and for making our time at Desy such a pleasurable one.

References

[1] Measurement of Charm and Beauty Production in Deep Inelastic Scattering at HERA. *P. Roloff*

[2] ZEUS status report 1993: $http : //www - zeus.desy.de/bluebook/scanned_bluebook.pdf$

[3] ZEUS web page: $http : //www - zeus.desy.de/ZEUS_ONLY/analysis/comntp/$

[4] Root web page: $root.cern.ch$

[5] Validation of ZEUS common nTuples. *Mark Barber*