School of Physics and Astronomy

ATLAS Detector Commissioning and Underlying Event Physics
Second Year Report

Benjamin Wynne
Jun 2010

Supervisors: Dr. Phil Clark, Prof. Peter Clarke

Contents

1 Progress this Year

In my First Year Report I mentioned three aims for my work in the coming year: working on my own physics analysis; continuing the development of RapidFit; and ongoing service work for the ATLAS collaboration. My progress in each of these areas is described below.

1.1 Physics Analysis

Since the course of my work was dependent on the performance of the LHC, I should note that the LHC successfully collided protons at 900 GeV during December 2009, producing initial data for my analysis. In March 2010, 7 TeV collisions were achieved and have been continuing since then, with some improvements in luminosity.

Rather than working on the ATLAS minimum bias analysis as suggested by my first year report, I am instead working on a different aspect of soft-QCD: the Underlying Event (UE). The term 'minimum bias' refers to collision data with (essentially) no selection criteria, dominated by low energy interactions. The 'underlying event' is the low energy particle background to any collision, regardless of the centre-of-mass energy. The distinction is subtle but it requires a different analytical approach, primarily looking at data from the regions of the detector that register the *least* activity in a given event.

1.1.1 Bayesian Iterative Unfolding

One particular area of interest for me is the deconvolution of detector effects from the reconstructed data, i.e. accounting for statistical uncertainty caused by finite detector resolution and removal of systematic effects caused by the detector structure. In particular, I have produced a tool based on a Bayesian iterative unfolding method [1]. Comparing Monte Carlo (MC) truth events to those that have been passed through a simulation of the ATLAS detector (in Geant4) and then reconstructed produces a 'Smearing Matrix' that describes the detector effects. Using this matrix and the MC truth as a prior distribution, applying Bayes' Theorem produces the correct physical distribution without the detector effects. Inaccuracies in the MC model can be accounted for by using the corrected distribution as a new prior and iterating.

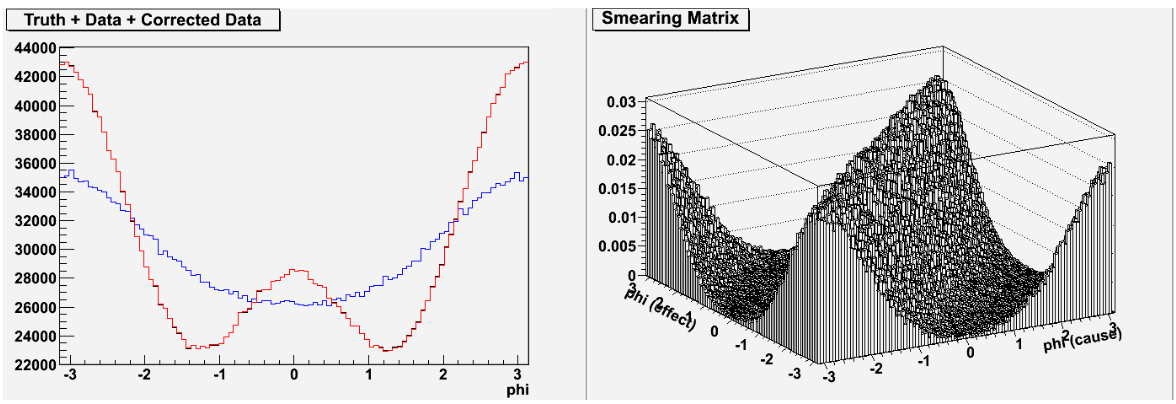

Figure 1: An example underlying event distribution. The uncorrected data (blue) is unfolded to produce the corrected distribution (red), which is almost identical to the true distribution (black). On the right is the smearing matrix used.

1.1.2 Workflows and the Grid

Using the PANDA workflow tool, and specifically the PANDArun (prun) software, I made it possible to run the UE analysis code on the Grid. I used prun to create a simple workflow in which the analysis source code is uploaded to grid worker nodes, compiled, and then run over a subset of the available input data. I also produced corresponding code to retrieve and combine analysis results from the Grid.

1.1.3 Underlying Event Paper

The first UE analysis - on 900 GeV data - is in the ATLAS approvals process, on the way to becoming a conference note, and eventually a paper. I am credited as a primary author of this paper (not just via the general ATLAS author list). Analysis of early 7 TeV data will also be included in the paper. Currently the analysis examines individual charged particle tracks, but future work I am involved in will use reconstructed jets instead.

1.2 RapidFit

I have continued development of RapidFit, so that it is currently fully-featured and in use by members of the group. To give a quick recap, RapidFit is a program for fitting a mathematical function to a set of data using the minimisation tool MINUIT. The intention is to provide a simple interface to this process, so that a user need not write any C++ code. Significant new features are discussed below.

1.2.1 Projection Plots

One of the most useful features I have added is the ability to plot the fitted analytical function on a histogram of the data, providing a visual confirmation of the quality of a fit result. While this is trivial for 1D functions, for an N-dimensional function the plotting tool must integrate over the $N-1$ dimensions that are not being plotted. The RapidFit plotting tool performs this integration and superimposes the resulting 1D function on the corresponding data histogram. Axis labels, units, ranges, and histogram binning are also handled automatically by the tool.

1.2.2 Contour Plots

The quality of a fit can also be judged by looking at the values of the minimised function in the region of parameter space around the minimum. By drawing contours at (for example) 1σ and 2σ uncertainty, the shape of the minimum can be found. The assumption is that the minimum is roughly parabolic in cross-section, and so deviation from this could indicate an unreliable fit result and the need for more detailed calculation of the errors on fitted parameters.

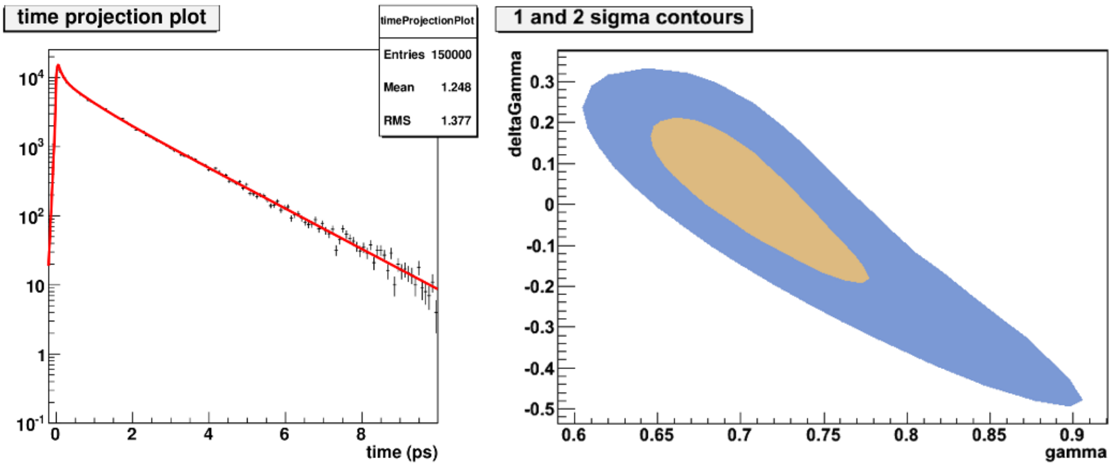

Figure 2: Using RapidFit, Monte Carlo events were used to fit a 4D function describing a decay of B-mesons. Afterwards the function was projected onto the time axis to show the lifetimes of the B-mesons, and the agreement between the fitted function (red) and the events (black). NB: there is also a simulated background contribution included in the fit. A contour plot for the Γ and Δ_Γ fit parameters is shown on the right.

1.2.3 Numerical Integration

In order to perform the integrations described above a method of numerical integration is required (either that or the function to fit would need to be accompanied by $N-1$ integral expressions). I use the (Quasi-Newtonian) numerical integration libraries provided by ROOT for the plotting as described above, and also to verify that the integral expression provided to normalise the function to fit is correct. The numerical and analytical integrals over the whole function are compared, and in case of discrepancy numerical integration is used throughout the fitting process, with automatic caching of results to minimise the performance penalty (since numerical integration is very time-consuming).

1.2.4 Modifying the Fit

While none of the features mentioned above is particularly novel, the central aim of RapidFit is to make routine fitting tasks trivial for the user, rather than to innovate in particular. In this regard it is a great success, as numerical integration is handled completely automatically, and the plots need only be requested by the user, with no additional input required. However, there has been the opportunity to add more unique features as well:

- I have added the ability to place external constraints on fit parameters. This could be an existing measurement of a parameter conducted by another experiment, or another measurement related to one or more fit parameters by a mathematical expression.
- I have implemented a method of background subtraction [2]. By finding some observable - independent of those used in the fit - which differentiates the signal and background, you can weight the events in the fit to emphasise the signal and diminish the background. A simple example would be the reconstructed particle mass in a given decay, presumably a Gaussian peak for the signal, and a flat distribution for the background. By comparing the reconstructed mass in each event to the expected mass distribution for signal and background, that event is given a weight which is then used in the fit.

1.3 ATLAS Service Work

This year I have continued my service work maintaining and improving the ATLAS Detector Control Systems (DCS). I also undertook a separate task comparing SCT data quality between different versions of the reconstruction software. As a result of this service work I am now quoted as on the official ATLAS author list to be used for all publications. I am also continuing with my 'Expert-on-call' role for the cooling system, to help fill the Edinburgh group's shift quota. Since the introduction of shift categories - separating Control Room roles from remote roles - I have begun taking luminosity monitoring shifts as well.

1.3.1 Detector Control Systems

I am still responsible for the environmental sensor system (ENV), and now have also taken on the system that operates the cooling loops themselves (EVCOOL). For each of these systems I have added new alerts and safety features in response to problems that have occurred during the ATLAS long physics run so far. For example, in response to an incident where communication between the control software and the hardware was lost, I have implemented code to draw sufficient information from alternative sources to protect the detector. I have also created new panels to the user interface displaying information about the LHC beam, and have replaced or improved a number of inefficient pieces of code that were having a dramatic impact on the system performance. I recently attended the Vertex 2010 conference, where I presented a poster on the performance of the cooling system. This poster has been approved as an official ATLAS conference slide.

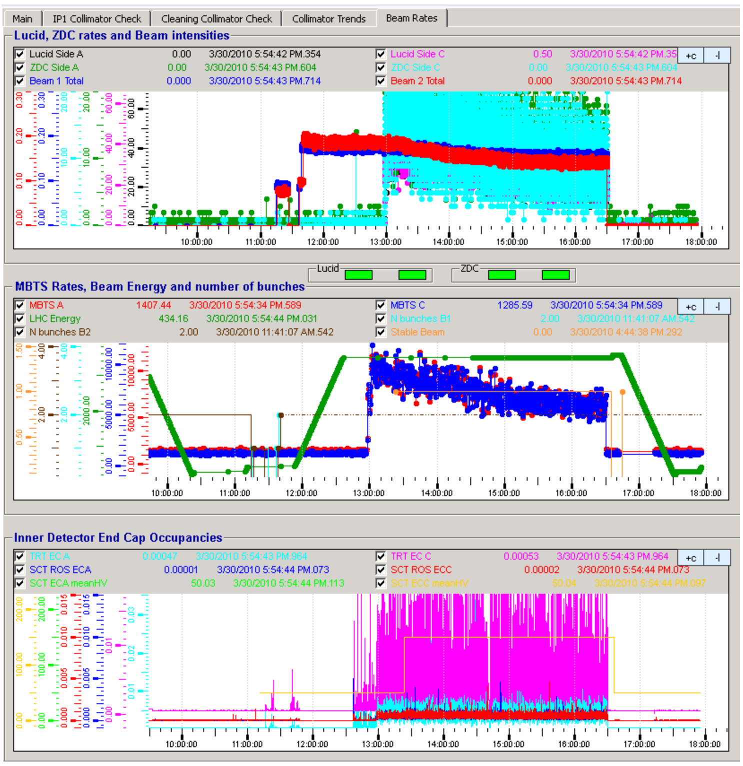

Figure 3: One of the new panels I have added to the DCS, showing data about the beam. Here the time axis covers the first 7 TeV collisions on the 30th of March, with beam injected at 11:30, magnet ramping beginning at 12:00, and then collisions from 13:00. The panel is for expert use, hence the high information density.

1.3.2 Thermal Enclosure Heaters

Surrounding the cooled silicon components of the Inner Detector are electric heater pads keeping the rest of the detector at its warmer operational temperature. The software controlling these Thermal Enclosure Heaters (TEH) was unreliable and unfinished, and so I took responsibility for this system for some months, in order to overhaul it. I made the following changes:

- Replaced the Finite State Machine (FSM), as the existing one tended to crash and had a poorly-defined set of states and transitions between states.
- Either improved or replaced inefficient code that was impairing system performance, such as the script for counting the number of active heaters, which now runs 100 times faster.
- Added new code to calculate average, minimum and maximum temperatures over regions of the TEH, and added readout of these values to the user interface.
- Integrated a new 3D display of the TEH system into the user interface (although I did not create the 3D display itself).
- Added mobile phone text message alerts for low temperatures.

After making these changes I passed responsibility for the system on to a new student.

1.3.3 SCT Data Quality

The software that reconstructs physics events from raw detector data (Athena) is a work-in-progress, and periodically there is a 'reprocessing' when the raw data stored at CERN is reconstructed using the latest version of Athena. This new reconstructed data is compared to that already available on the grid, and if the quality of the data has improved then the newer version is used. While judging the quality of the data is a task tackled by groups from each ATLAS subdetector, there had been little effort to develop a tool for comparing these results between the newer and older reprocessings. I produced such a tool for the SCT data quality group, which has the following features:

- Provides a simple configuration file allowing users to specify histogram names to search for, and how to test them.
- Uses PANDArun to upload source code to grid worker nodes, compile it, and then provide the appropriate data quality files to analyse.
- Automatically compensates for changes in histogram naming conventions between reprocessing versions.
- Retrieves a summary of the comparison, indicating where comparisons were unsatisfactory and providing superimposed "before and after" plots in such cases.
- Code structure designed to allow additional tests to be implemented with minimal changes to the rest of the program.

This tool was successfully used to examine both the April and May data reprocessing.

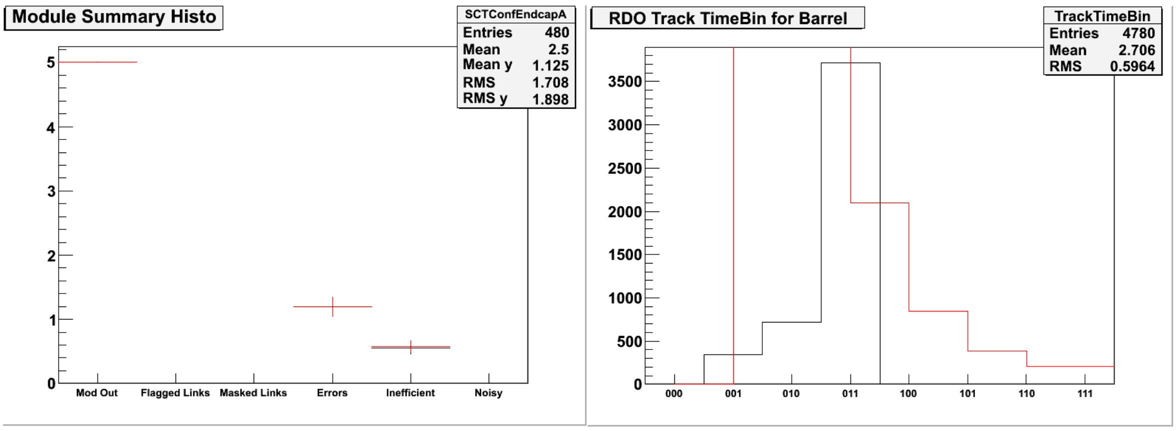

Figure 4: Two example outputs from my data quality comparison tool, superimposing the old histogram (black) and the corresponding new one (red). The plot on the left shows a slight change between two versions, compared to a major discrepancy on the right. The stringency of each test is set by the user.

2 Plan for Next Year

After 18 months in Geneva, on long term attachment to CERN, I will be returning to Edinburgh in October. Before then I need to do plenty of shift work to help fulfil Edinburgh's obligation to the ATLAS collaboration. I will also look to train a replacement for my DCS responsibilities, and to make the most of face-to-face meetings with the other members of the UE analysis group before I return to the UK.

Over the next year I plan to develop the underlying event analysis from the simple track-based analysis at present to use reconstructed jets instead. Ideally there will be production and use of common tools across all underlying event analysis efforts (hence my unfolding tool), and parallel implementation of the analysis code in Rivet to aid the effort to tune Monte Carlo generators.

2.1 Thesis Structure

- Detector chapter
  - Description of the ATLAS detector
  - Detector Control Systems and my work with the cooling system
- Theory chapter
  - Underlying event
  - Discussion of Monte Carlo generators and tuning
- UE analysis
  - Analysis methods and results, including those published
- Further analysis
  - Tuning with UE results?
  - Use of RapidFit?

2.2 Approximate Timescale

- Mid 2010: Publish track-based UE analysis
- Late 2010: Detector chapter
- Late 2010: Jet-based UE analysis
- Early 2011: Publish jet-based UE analysis
- Mid 2011: Analysis Chapter
- Mid 2011: Theory Chapter
- Late 2011: Further Analysis
- Late 2011: First draft
- Early 2012: Corrections and submit
- April 2012: Funding expires

References

[1] Giulio D'Agostini. A multidimensional unfolding method based on bayes' theorem. *Nucl. Instr. and Meth. in Phys. Res. A362*, 1995.

[2] Yuehong Xie. Sfit: a method for background subtraction in maximum likelihood fit. *arXiv:0905.0724v1*, 2009.