ATLAS Open Data in Schools: Student, Teacher, and Researcher Experiences

Seth Zenz and Nigel Sharma

Thursday 14 May 2026

Overview

- Year-long activity for 15-18 year olds
  - ATLAS Open Data Project
  - Physics Research in School Environments (PRiSE)
- Lessons Learned
  - Skills involved can be underestimated, for <u>all</u> of:
    - Students
    - Teachers
    - University-based Researchers
  - Clear, step-by-step self-guided activities are essential for success
- Ideas for the Future

About Us

Seth Zenz

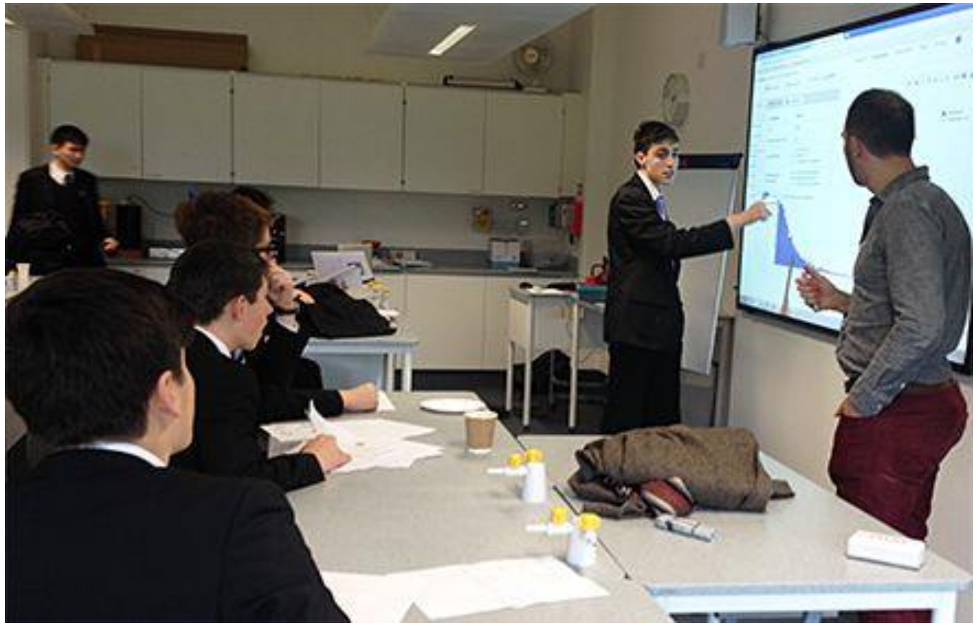
- Reader in Experimental Particle Physics, Queen Mary University of London
- Physics Research in School Environments - ATLAS Open Data (since 2019)
- Higgs Boson Dominoes
- STFC Advisory Panel for Public Engagement
- User of ATLAS Open Data, but not a developer

Nigel Sharma

- Former Secondary School Physics and Science Teacher since 2004
- Focused entirely on Enrichment since 2017
- Involved in running PRiSE projects in schools since 2017
- Substantial experience running Science Enrichment activities in schools including Science Clubs, CREST projects, EPQs, Residential Science Weeks, Science competitions, Expos, Field trips

PRiSE Programme

- Year-long programme
  - Kickoff in September-October
  - Final conference in March
- Typical participants:
  - Enthusiastic 15-to-18 year old school students
  - No specific particle physics coursework (yet)
  - Teachers have exceptional STEM experience and “spare” time
- Goals: science skills, STEM self-efficacy

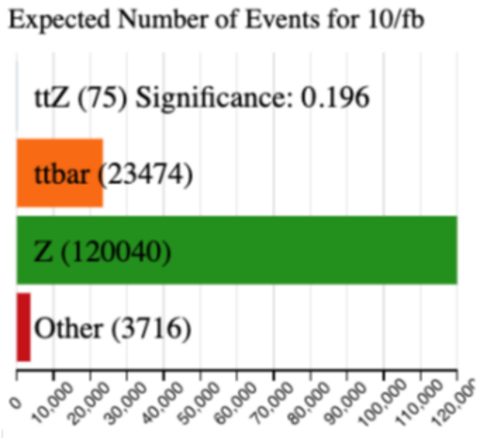


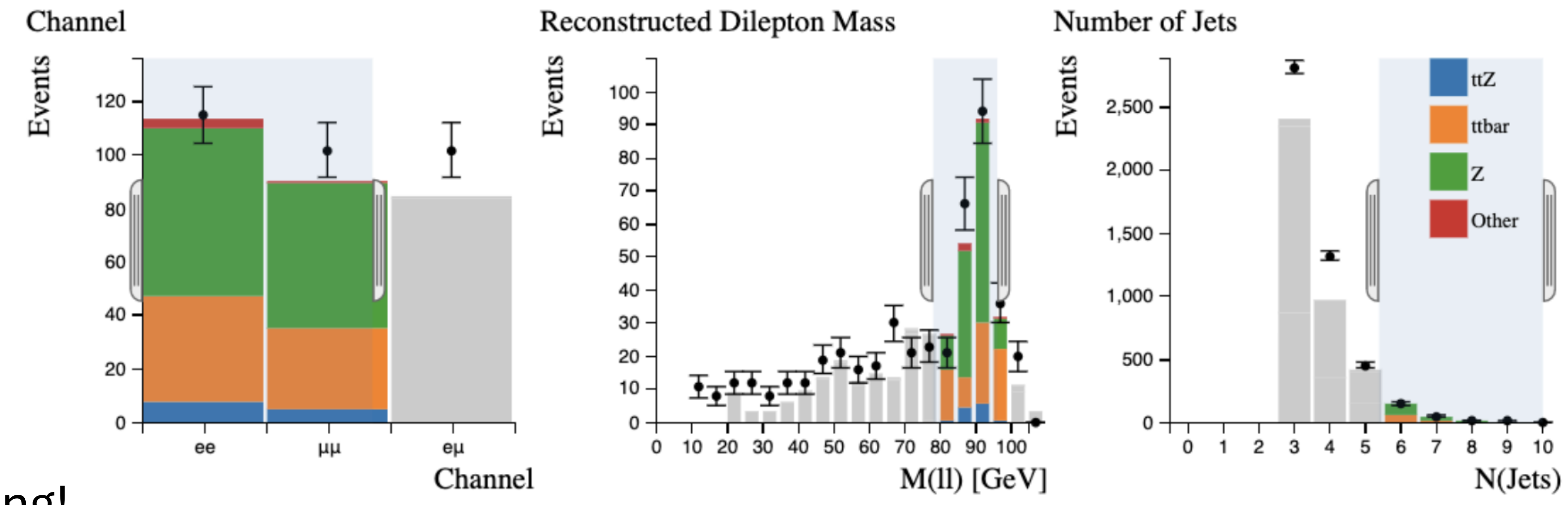
Why 15-to-18 year olds?

Opportunity to develop their understanding of what “real science” at university level will look like!

ATLAS PRiSE - Overview

- We use resources from ATLAS Open Data: https://atlas.cern/Resources/Opendata
- Typically we use the interactive histogram analyzer examples as a launch point



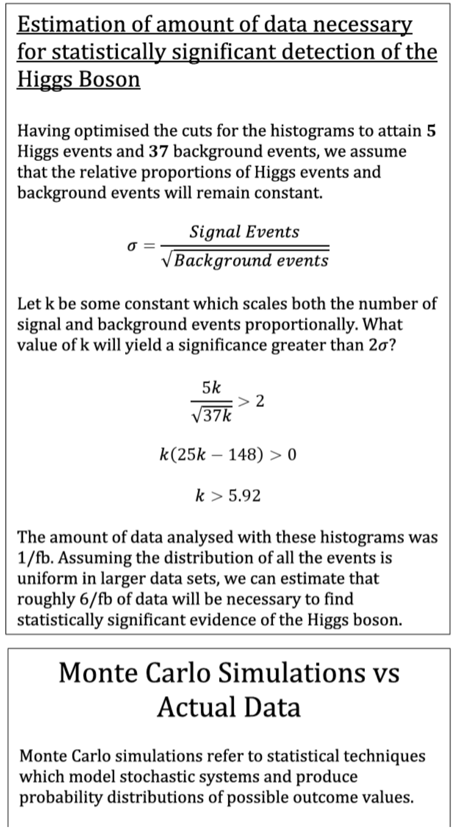


ttZ events:
best-explained analysis

But physics context is daunting!

ATLAS PRiSE - Overview

- We use resources from ATLAS Open Data: https://atlas.cern/Resources/Opendata
- Typically we use the interactive histogram analyzer examples as a launch point
- Final projects usually <u>explain</u> significance, signals and backgrounds, variables, why certain selections work – often without further data analysis


<u>Estimation of amount of data necessary for statistically significant detection of the Higgs Boson</u>

Having optimised the cuts for the histograms to attain **5** Higgs events and **37** background events, we assume that the relative proportions of Higgs events and background events will remain constant.

$$\sigma = \frac{Signal\ Events}{\sqrt{Background\ events}}$$

Let k be some constant which scales both the number of signal and background events proportionally. What value of k will yield a significance greater than 2σ?

$$\frac{5k}{\sqrt{37k}} > 2$$

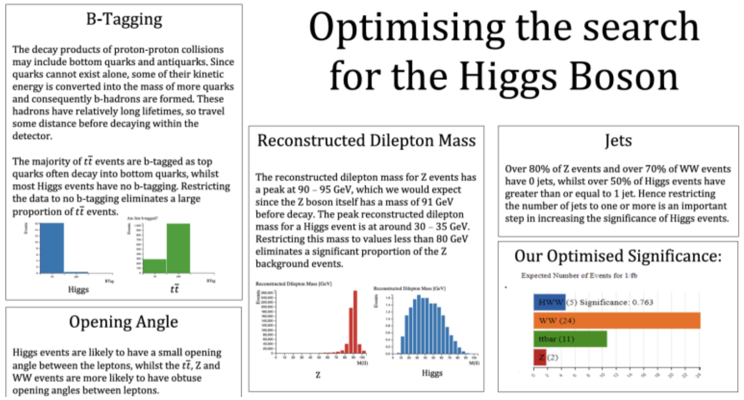
$$k(25k - 148) > 0$$

$$k > 5.92$$

The amount of data analysed with these histograms was 1/fb. Assuming the distribution of all the events is uniform in larger data sets, we can estimate that roughly 6/fb of data will be necessary to find statistically significant evidence of the Higgs boson.

Monte Carlo Simulations vs Actual Data

Monte Carlo simulations refer to statistical techniques which model stochastic systems and produce probability distributions of possible outcome values.


Optimising the search for the Higgs Boson

B-Tagging

The decay products of proton-proton collisions may include bottom quarks and antiquarks. Since quarks cannot exist alone, some of their kinetic energy is converted into the mass of more quarks and consequently b-hadrons are formed. These hadrons have relatively long lifetimes, so travel some distance before decaying within the detector.

The majority of $t\bar{t}$ events are b-tagged as top quarks often decay into bottom quarks, whilst most Higgs events have no b-tagging. Restricting the data to no b-tagging eliminates a large proportion of $t\bar{t}$ events.

Opening Angle

Higgs events are likely to have a small opening angle between the leptons, whilst the $t\bar{t}$, Z and WW events are more likely to have obtuse opening angles between leptons.

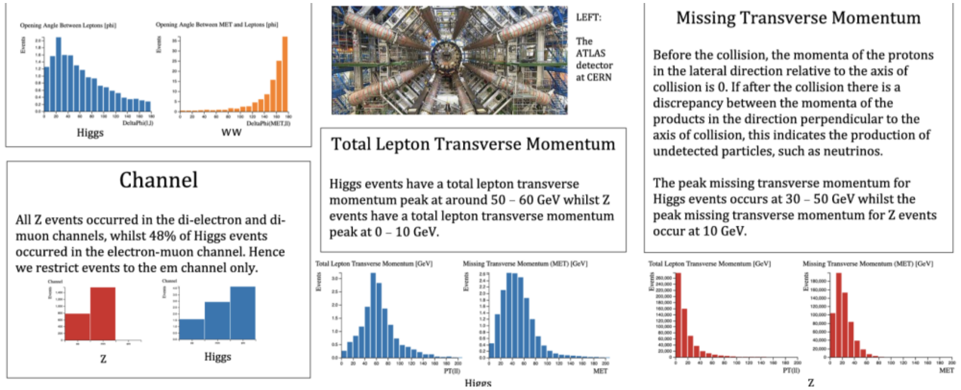
Reconstructed Dilepton Mass

The reconstructed dilepton mass for Z events has a peak at 90 – 95 GeV, which we would expect since the Z boson itself has a mass of 91 GeV before decay. The peak reconstructed dilepton mass for a Higgs event is at around 30 – 35 GeV. Restricting this mass to values less than 80 GeV eliminates a significant proportion of the Z background events.

Jets

Over 80% of Z events and over 70% of WW events have 0 jets, whilst over 50% of Higgs events have greater than or equal to 1 jet. Hence restricting the number of jets to one or more is an important step in increasing the significance of Higgs events.

Our Optimised Significance:



Channel

All Z events occurred in the di-electron and di-muon channels, whilst 48% of Higgs events occurred in the electron-muon channel. Hence we restrict events to the em channel only.

Total Lepton Transverse Momentum

Higgs events have a total lepton transverse momentum peak at around 50 – 60 GeV whilst Z events have a total lepton transverse momentum peak at 0 – 10 GeV.

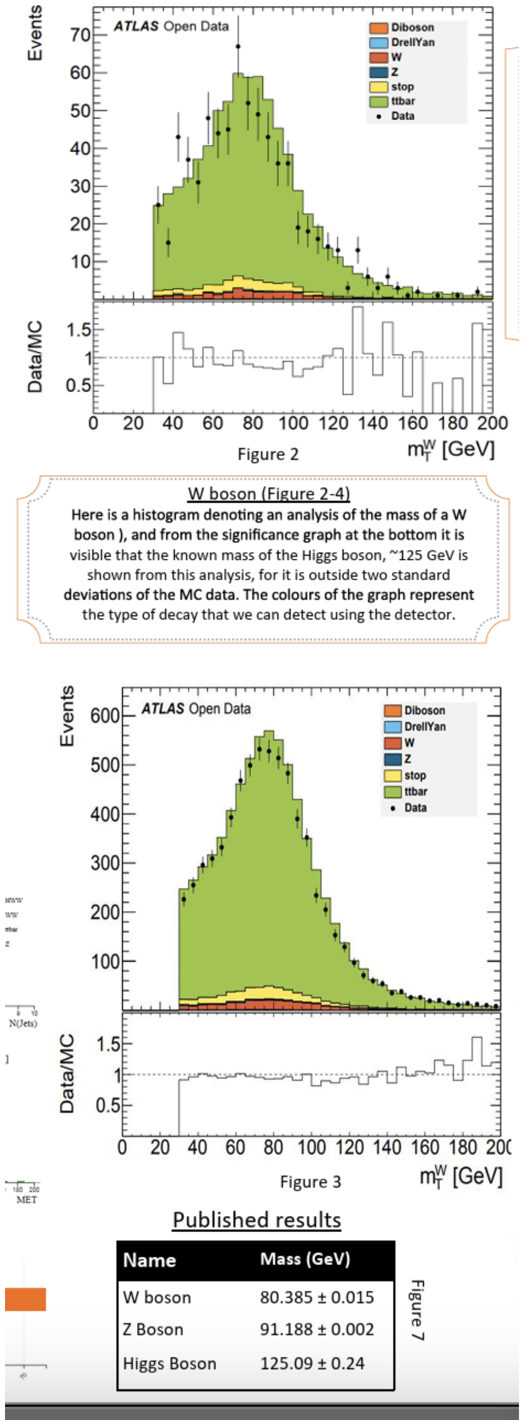
Missing Transverse Momentum

Before the collision, the momenta of the protons in the lateral direction relative to the axis of collision is 0. If after the collision there is a discrepancy between the momenta of the products in the direction perpendicular to the axis of collision, this indicates the production of undetected particles, such as neutrinos.

The peak missing transverse momentum for Higgs events occurs at 30 – 50 GeV whilst the peak missing transverse momentum for Z events occur at 10 GeV.

ATLAS PRiSE - Overview

- We use resources from ATLAS Open Data: https://atlas.cern/Resources/Opendata
- Typically we use the interactive histogram analyzer examples as a launch point
- Final projects usually explain significance, signals and backgrounds, variables, why certain selections work – often without further data analysis
- Rare exception at right: direct use of Open Data in ROOT format to make histograms
  - *This has happened once in ~10 years as far as I know!*



Skills Developed

- Creative thinking
- Working independently from adults
- Teamwork
- Problem solving
- Analysis of complex data
- Research skills
- Presentation skills
- Project and time management
- Inspiration

How can we do more to <u>develop</u> these skills through the project?

Challenges Found

- Guided activities are often more useful than providing “raw” data
  - Multivariable analysis frameworks are complex for secondary level, and would require detailed examples
- Activities may be hard to follow if website is complex
  - Example: ATLAS Open Data website has multiple datasets and strands
    - Not always clear how to navigate between related pages without encountering extraneous pages
    - It may be difficult to identify whether explanatory pages are specific to the current activity
    - → External guide to website may be developed by academics or teachers
  - Tiny changes to website can break existing guides/instructions!
- Expertise in subject matter and website content needed to support independent student exploration → very little time for teachers to develop this!
- Tension between science as an approach/skillset and science as a “fixed” collection of knowledge

How to get more out of the programme?

- Teacher CPD
  - Particle physics background
  - Guiding students through independent projects
    - Useful training for researchers and academics as well!
- Engage teachers in development
  - Guides to websites and choice of activities
  - Initial definition of guided activities, with datasets as integral components of those activities
- Actively study student journeys through datasets/activities
  - What are the realistic self-guided projects that are available?
  - What features do they have?

Conclusions

- ATLAS Open Data and activities built around it are effective for 15 year olds, with positive outcomes, but challenging to use
- Skills involved are complex – and useful if we can develop them
  - Multivariate analysis: often not taught explicitly but essential for many modern careers. How do we teach it better?
  - Independent extended projects:
    - New to school students
    - Mostly new to teachers
    - University researchers and academics may need thinking and training on how to support them at this level!
- “Student Journeys” need more systematic attention
- Better connections to A Levels
- Broader Geographic Range

Extra Info: Overview and Questions

This talk describes lessons learned from using ATLAS Open Data in the Physics Research in School Environments (PRiSE) Programme at Queen Mary University of London, which teaches physics and STEM skills through year-long projects to 15-18 year old GCSE and A-Level students. We focus here on identifying the range of skills involved in completing such projects, and the challenges experienced by the various people involved (i.e. students, teachers, and university researchers).

- **A question regarding a particular challenge in your work for which other participants may provide input:**
  How do we realistically enable teachers and university researchers to support projects for GCSE and A-Level students using complex Open Data?
- **A statement on the future of Open Data in Higher Education, what we should do:**
  Focus on robust presentation that is usable in a self-guided way by specific realistic target audiences (e.g. specific student year(s), teachers, etc.). Think of datasets as integral parts of specific activities with learning objectives and maximise involvement of educators at the target level in developing both the activities and the datasets themselves.
- **A suggestion for how this work or expertise may be useful to a broader audience or to colleagues in other fields:**
  Multivariate analysis underpins an increasing number of jobs, notably in AI, but it is often underestimated as a taught skill and way of thinking. Even at UG project and PhD level we ask students to "learn on the job", which is not always effective; working with younger audiences forces us to identify what to actively teach and how.

Extra Info – Skills Detail Notes

- **Creative thinking**: The hardest part for nearly all students is deciding what to do their project on. Up till now they have almost always been told what to do, what to research, etc. They find it extremely difficult to cope with a blank slate and have to think of what they would like to research and what their project should be about.
- **Working independently from adults:** This is necessitated, not only because teaching staff simply don't have time to drill down into all the projects, but also the students are often going beyond the teacher's own knowledge on their particular project. The physics teachers themselves often learn something from their students' presentations!
- **Teamwork:** I have seen team sizes from 2 up to 7 (and individuals) but in my personal opinion 4 is an optimum team size (depending on the number of tasks required). There are often a range of tasks that can be divided amongst the team such as research, calculations, programming, creating presentations, etc. I try to encourage the students to form teams where all participants have a common free period at least once a week, and this can be when they get together to discuss the project.
- **Problem solving:** Speaks for itself. These projects always involve some degree of problem solving.
- **Analysis of complex data:** Certainly up to GCSE and even up to A Level, students are taught that science experiments always involve one independent variable, one dependent variable and lots of control variables. They almost never have situations with multiple variables (that are not control variables) and yet this is often the reality in real life science experiments.
- **Research skills:** The project will usually involve research to at least some degree. The students are encouraged to compare the research with their actual findings and to consider why the two may be different. This also encourages scientific integrity. In many school experiments it can be relatively easy to make the results fit what you want them to be, or think they should be!
- **Presentation skills:** Presentation via posters as per normal scientific conferences forces the students to distil their project to a level that others can comprehend it. Some projects are presented via powerpoint at the Cosmic Con. I actually encouraged all my groups to produce powerpoint presentations that could be shared within the school, even if they were not chosen to present at Cosmic Con. Those that are chosen certainly get to think carefully about and hone their presentation skills. When my groups were chosen, I got them to present internally first so as to help polish their presentations.
- **Project and time management:** The Cosmic Con deadline necessitates this. There are various deadlines within the deadline, e.g. for submitting posters, etc which forces groups to set their own internal deadlines; e.g. finishing research, finishing work on the project, analysing results, revisiting results, preparing the poster, etc.
- **Inspiration:** The fact that the students are working on real data from real scientific experiments such as CERN and the Kepler Space Telescope is very inspiring to the students. The exoplanets they find are real exoplanets that actually exist out there in space. The students get a real taste of what university level research is like.