

ATLAS Open Data project: HEP for everyone

Meirin Oan Evans, on behalf of the ATLAS Collaboration^{*†}

University of Manchester

E-mail: meirin.oan.evans@cern.ch

We explore the many ways that public High Energy Physics resources are employed to teach and outreach particle physics and computer science. The current ATLAS model of Open Access to recorded and simulated data offers the opportunity to access datasets with a focus on education, training and outreach. This mandate supports the creation of platforms, projects, software, and educational products used all over the planet. We describe the overall status of ATLAS Open Data <http://opendata.atlas.cern> activities, from core ATLAS activities and releases to individual and group efforts, as well as educational programs, and final web or software-based (and hard-copy) products that have been produced or are under development. The relatively large number of heterogeneous use cases currently documented is driving an upcoming release of more data and resources for the ATLAS Community and anyone interested to explore the world of experimental particle physics and the computer sciences through data analysis.

*ICHEP 2018, International Conference on High Energy Physics
4–11 July 2018
Seoul, South Korea*

^{*}Speaker.

[†]With thanks to the ATLAS Outreach Data & Tools group

1. Introduction

Through its outreach group, the ATLAS Collaboration [1] has formed the ATLAS Open Data project. Broadly, the primary aim is to make High Energy Physics (HEP) data and tools more accessible to everyone. The project is not simply about providing public data; software, platforms and analysis examples are also available as educational tools. With these provisions, users analyse real ATLAS data from the Large Hadron Collider (LHC) [2], whilst learning HEP techniques. Key target audiences and focus groups are university students, as well as advanced high school students with basic HEP knowledge. Documentation, analysis tools, apps, datasets, software and virtual machines are all contained on the ATLAS Open Data web platform [3], and are also accessible from the CERN Open Data portal [4]. The first ATLAS Open Data release was of proton-proton collision data at a center-of-mass energy $\sqrt{s} = 8$ TeV, the second will be of data at $\sqrt{s} = 13$ TeV.

2. ATLAS Open Data communications

Two reviews of the 8 TeV ATLAS Open Data datasets have been written. ‘‘Review of ATLAS Open Data Dataset’’ [5] was published alongside the 8 TeV dataset release. Feedback was then gathered from multiple users over a year to write ‘‘Review of ATLAS Open Data 8 TeV datasets, tools and activities’’ [6]. Figure 1 shows a plot that could be made using the 8 TeV datasets. ATLAS Open Data have been used in two master’s theses in Venezuela [7, 8], and one in Great Britain [9].

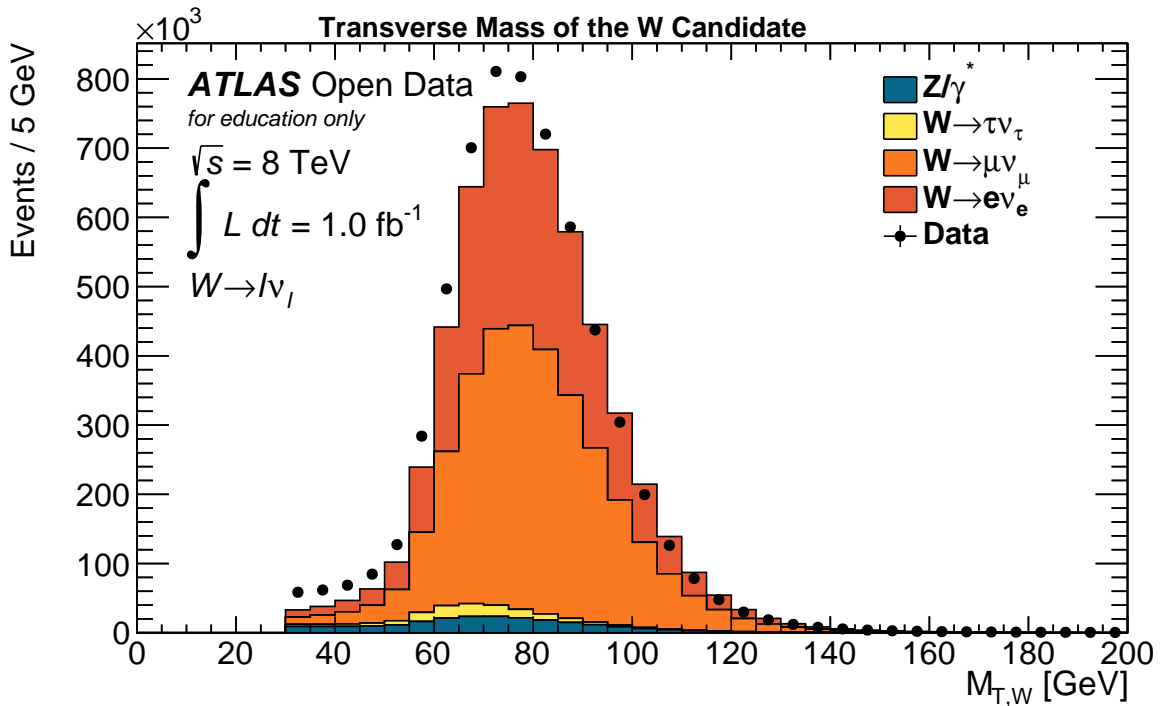


Figure 1: The number of events selected in a typical lepton + Missing Transverse Momentum (W boson) analysis as a function of the W boson candidate transverse mass, using 8 TeV ATLAS Open Data. Z/γ^* describes Drell-Yan production processes in which the gauge boson then decays leptonically. The $W \rightarrow \tau\nu_\tau$, $W \rightarrow \mu\nu_\mu$ and $W \rightarrow e\nu_e$ contributions together simulate all W leptonic decays.

3. ATLAS Open Data uses

The groups involved with ATLAS Open Data are worldwide. CEVALE2VE [10] is a project using ATLAS Open Data to train new analysers in Latin-America. The University of Oslo has developed the “ZPATH” measurement for International Masterclass high school students, to introduce the invariant mass concept then identify particles. They also use ATLAS Open Data with undergraduate level university students as a continuation of this effort within an independently funded project for student driven research. Laboratory courses for university students have been developed and run by universities in Manchester, Athens, Dortmund, Göttingen and Dresden, amongst others.

With 13 TeV ATLAS Open Data currently under development, users will be able to create high statistics studies of kinematic distributions of a variety of final states, such as the Z boson candidate invariant mass in Figure 2. After plotting such a histogram, users could practice using ROOT [11] by making fits to the distribution, extracting the Z boson mass, estimating the Z boson lifetime, optimising Z boson selection requirements or calculating the $Z \rightarrow \ell^+ \ell^-$ process cross section, to name a few possibilities. Derived quantities could then be compared to accepted values [12].

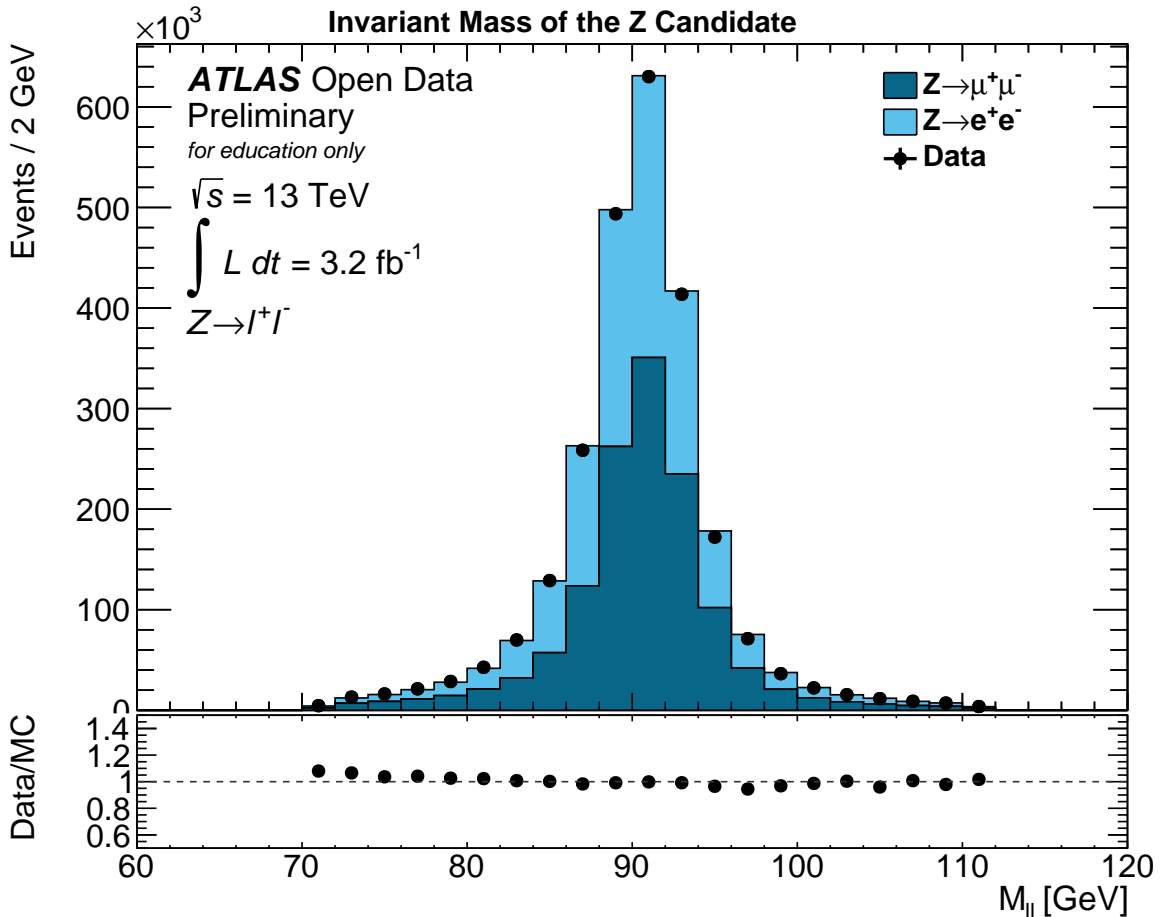


Figure 2: The number of events selected in a typical di-lepton (Z boson) analysis as a function of the Z boson candidate invariant mass, using 13 TeV ATLAS Open Data under development. Both opposite-sign di-electron and di-muon event selections are included.

4. Future work

Going forward, the ATLAS Open Data group is seeking members and collaborators to get involved. The group is developing a 13 TeV dataset package, with aim to release during 2019. Additionally to those available with the 8 TeV datasets, the 13 TeV package will include new MC, physics objects, variables and analysis examples, such as the Higgs di-photon decay channel. Inclusion of information to evaluate relevant systematic uncertainties and MC truth is being considered.

5. Summary

8 TeV ATLAS Open Data from the LHC are being used. Many tools and extensive documentation help users make their own discoveries. ATLAS Open Data are targeted at university students, as well as advanced high school students. Through projects, training, classes and courses for these students, there have already been successful use cases in education, outreach, public engagement and science communication. The aim is to release 13 TeV ATLAS Open Data during 2019.

Acknowledgements

The author wishes to highlight financial support from the School of Physics and Astronomy, University of Manchester, and the ICHEP Organising Committee for presentation of this work.

References

- [1] ATLAS Collaboration, *The ATLAS Experiment at the CERN LHC*, 2008 *JINST* **3** S08003.
- [2] L. Evans and P. Bryant (editors), *LHC Machine*, 2008 *JINST* **3** S08001.
- [3] The ATLAS Open Data portal - <http://opendata.atlas.cern>
- [4] The CERN Open Data portal - <http://opendata.cern.ch>
- [5] ATLAS Collaboration, *Review of ATLAS Open Data Dataset* - [ATL-OREACH-PUB-2016-001](#)
- [6] ATLAS Collaboration, *Review of ATLAS Open Data 8 TeV datasets, tools and activities* - [ATL-OREACH-PUB-2018-001](#)
- [7] M. R. Di Domenico Franco, *Reconstruction of the invariant masses of bosons of the Standard Model using public data from ATLAS Open Data* - [CERN-THESIS-2017-239](#)
- [8] I. Garcia, *Perspectives and Evaluation of Dark Matter production in association with a light quark, a heavy quark (*b*-quark) or an electroweak boson in particle colliders at a centre-of-mass energy of $\sqrt{s} = 8$ TeV* - [CERN-THESIS-2017-217](#)
- [9] M. O. Evans, *Enabling Open Science with the ATLAS Open Data project at CERN* - [CERN-THESIS-2018-099](#)
- [10] A. R. Sanchez Pineda, *The CEVALE2VE case* - [ATL-OREACH-PROC-2017-001](#)
- [11] R. Brun and F. Rademakers, *ROOT - An Object Oriented Data Analysis Framework*, *Proceedings AIHENP'96 Workshop, Lausanne, Sep. 1996, Nucl. Inst. & Meth. in Phys. Res. A* **389** (1997) 81-86. See also <http://root.cern.ch>
- [12] M. Tanabashi *et al.* (Particle Data Group), 2018 *Phys. Rev. D* **98**, 030001.