13 TeV ATLAS Open Data: using open education resources effectively

Leonid Serkin
for the ATLAS Collaboration

CERN and INFN Gruppo Collegato di Udine and ICTP Trieste
• How can we overcome geographical distances and allow anyone interested in experimental particle physics to learn remotely?

✓ ATLAS Collaboration launched a comprehensive educational platform to guide university-level students and teachers on how to use the data and analysis tools
✓ Provide a straightforward interface to replicate the procedures used by high-energy-physics researchers and enable users to experience the analysis of particle physics data in educational environments
NEW TODAY: The ATLAS Collaboration is releasing a new set of proton-proton collision data to the public for educational purposes

- data collected by the ATLAS detector at the Large Hadron Collider at a centre-of-mass energy of \(13 \text{ TeV}\) during the year 2016
- corresponds to an integrated luminosity of \(10 \text{ fb}^{-1}\) (1 fb\(^{-1}\) released at 8 TeV)
- dataset accompanied by simulated events describing both several Standard Model (SM) processes, as well as hypothetical Beyond SM signal production
- 12 physics-analysis examples inspired by published ATLAS Collaboration results presented to demonstrate the wide range of scenarios
- associated computing tools to make the analysis of the dataset accessible
• 13 TeV ATLAS Open Data:
  ✓ 61 runs from the first 4 periods of the 2016 proton-proton data-taking
• **13 TeV ATLAS Open Data:**
  - 61 runs from the first 4 periods of the 2016 proton-proton data-taking
  - releasing to the public 10 fb\(^{-1}\) of pp collision data (~ 270 million collision events)
• **13 TeV ATLAS Open Data:**
  ✓ 61 runs from the first 4 periods of the 2016 proton-proton data-taking
  ✓ releasing to the public 10 fb⁻¹ of pp collision data (~ 270 million collision events)

• **Events are selected by applying several event-quality and trigger criteria, and classified according to the type and multiplicity of reconstructed objects**
  ✓ subjected to a loose event preselection to reduce processing **time**
- 13 TeV ATLAS Open Data reconstructed objects contain:
  - electrons, muons, photons, hadronically decaying tau-leptons, small-R jet and large-R jet candidates (and MET) reconstructed with the ATLAS detector

<table>
<thead>
<tr>
<th>Electron ((e))</th>
<th>Muon ((\mu))</th>
<th>Photon ((\gamma))</th>
</tr>
</thead>
<tbody>
<tr>
<td>InDet &amp; EMCAL rec. loose identification loose isolation (p_T &gt; 7) GeV (</td>
<td>\eta</td>
<td>&lt; 2.47)</td>
</tr>
<tr>
<td>Hadronically decaying (\tau)-leptons ((\tau_h))</td>
<td>Small-R jets</td>
<td>Large-R jets</td>
</tr>
<tr>
<td>InDet &amp; EMCAL rec. medium identification (p_T &gt; 20) GeV (</td>
<td>\eta</td>
<td>&lt; 2.5) 1 or 3 associated tracks</td>
</tr>
</tbody>
</table>
• Selected events classified into separate final-state collections
  • depending on the number of final state objects and their energy, and triggers used (single-lepton, diphoton,..)

• Not only pp collision data recorded with the ATLAS detector
  • also **MC simulation** samples describing several SM processes used to model the expected distributions of different signal and background events (top quark pair, single top quark, Z+jets, W+jets, WW/WZ/ZZ, SM Higgs and BSM signals)
  • both data and MC provided in a **simplified data format** reducing the information content of the original data analysis format used within ATLAS
  • ROOT tuple with more than **80 branches**, optimised to reduce the complexities encountered in a full-scale analysis (~**150 GB** of storage)
12 examples of physics analysis using 13 TeV ATLAS Open Data

- inspired and following as closely as possible the procedures and selections taken in already published ATLAS Collaboration results
### Analyses

<table>
<thead>
<tr>
<th>Physics processes</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>$W \rightarrow l\nu$, $Z \rightarrow (ee/\mu\mu)$, $t\bar{t}$ top-quark-pair</td>
<td>high event yields to study the SM processes in detail</td>
</tr>
</tbody>
</table>

**4 high statistics**
Analyses | Physics processes | Purpose
--- | --- | ---
**3 low statistics** | Single-top-quark, WZ and ZZ diboson | illustrate the statistical limitations of the released dataset
<table>
<thead>
<tr>
<th>Analyses</th>
<th>Physics processes</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 Higgs boson</td>
<td>H→WW, H→ZZ, H→γγ</td>
<td>“re-discover” the production of the SM Higgs boson</td>
</tr>
</tbody>
</table>

- Analyses
  - 3 Higgs boson

- Physics processes
  - H→WW
  - H→ZZ
  - H→γγ

- Purpose
  - “re-discover” the production of the SM Higgs boson
<table>
<thead>
<tr>
<th>Analyses</th>
<th>Physics processes</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 BSM physics</td>
<td>SUSY, heavy boson</td>
<td>searching for new physics using different physics objects</td>
</tr>
</tbody>
</table>

13 TeV ATLAS Open Data physics analysis examples

![Graph](image1)

**ATLAS Preliminary Open Data**

- **Data**
- **Z+Jets**
- **Diboson**
- **t#bar{t}**
- **Single Top**
- **W+Jets**
- **Stat. unc.**

$m(t#bar{t})=600,300$ GeV

![Graph](image2)

**ATLAS Preliminary Open Data**

- **Data**
- **t#bar{t}**
- **V+Jets**
- **Single Top**
- **Diboson**
- **Stat. unc.**
- **m_{t\bar{t}}=1$ TeV}

$m_{Top-tagged Large-R jet, Small-R jet, Lepton}$ [GeV]
13 TeV ATLAS Open Data can be used for educational purposes with different levels of task difficulty:

- **beginner level**: visualise the content and produce simple distributions
- **intermediate-level**: making histograms with collision data after selection
- **advanced-level**: measuring event properties and physical quantities
- simplified single-component systematic-uncertainty estimate included
13 TeV ATLAS Open Data can be used for educational purposes with different levels of task difficulty:

- **beginner level**: visualise the content and produce simple distributions
- **intermediate-level**: making histograms with collision data after selection
- **advanced-level**: measuring event properties and physical quantities
- simplified single-component systematic-uncertainty estimate included

13 TeV ATLAS Open Data educational tools provided:

- analysis framework: reading, analysis selection, histograms and plotting
- **ROOTbooks**: data analysis in a web browser using ROOT with Jupyter notebooks
- virtual machine: Linux, ROOT, framework and the released datasets
- online documentation platform
• Running the 13 TeV ATLAS Open Data $H\to\gamma\gamma$ analysis: an example

Based on:

PRD 98 (2018) 052005

11th January 2019

Measurements of Higgs boson properties in the diphoton decay channel with 36 fb$^{-1}$ of $pp$ collision data at $\sqrt{s} = 13$ TeV with the ATLAS detector

The ATLAS Collaboration

Simplified selection:

• 2 photons with $p_T > 35, 25$ GeV
• $E_T / m(\gamma\gamma) > 0.35(0.25)$
• $105$ GeV $< m(\gamma\gamma) < 160$ GeV

Figure to be produced:

H$\gamma\gamma$ Analysis
• **A new set** of pp collision data **is released** by the ATLAS Collaboration to the public for educational purposes, hosted on the [CERN](https://cern.ch) and [ATLAS](https://atlas.ch) Open Data online portals

• Data has been collected by the ATLAS detector at the LHC at $\sqrt{s} = 13$ TeV during the year 2016 and corresponds to an integrated luminosity of $10 \text{ fb}^{-1}$

• **First public release** of pp collision-data samples recorded at 13 TeV released by an LHC experiment

• The pp collision data is **accompanied** by a set of **MC simulated samples** describing several processes which are used to model the expected distributions of different signal and background events

• Several **physics-analysis examples** inspired by published ATLAS Collaboration results are presented to demonstrate the wide range of final-state scenarios provided within the 13 TeV ATLAS Open Data
A new set of pp collision data is released by the ATLAS Collaboration to the public for educational purposes, hosted on the CERN and ATLAS Open Data online portals.

Data has been collected by the ATLAS detector at the LHC at $\sqrt{s} = 13$ TeV during the year 2016 and corresponds to an integrated luminosity of $10$ fb$^{-1}$.

First public release of pp collision-data samples recorded at 13 TeV released by an LHC experiment.

The pp collision data is accompanied by a set of MC simulated samples describing several processes which are used to model the expected distributions of different signal and background events.

Several physics-analysis examples inspired by published ATLAS Collaboration results are presented to demonstrate the wide range of final-state scenarios provided within the 13 TeV ATLAS Open Data.

Have fun playing and learning!
BACK-UP