ATLAS Open Data
software for visualization and physics analysis

Ya-Feng LO

Arturo Rodolfo SANCHEZ PINEDA, Leonid SERKIN
Outline

• ATLAS Open Data
• Software's description
• Data analysis concept
• Example: HWW analysis
• Other analysis result
• Conclusion
ATLAS Open Data

• Aim to release real and simulated data, together with software resources to analyze those samples.

• Provide detailed documentation and software training for people all around the world to access high energy physics at CERN.

Check out the webpage! http://opendata.atlas.cern
Software's description

• Build a framework based on ROOT to analyses high energy physics datasets.

• Use C++ programming language
  ➢ One of the main programming languages for high energy physics.
  ➢ This is one important piece still missing in the set of ATLAS Open Data public analysis code.
  ➢ Improve the speed and the ability by running in multiple-cores at once.

• The framework contains:
  ➢ All the needed pieces to run, edit and create physics analysis.
  ➢ Contain six cut-and-count physics analysis.
  ➢ Documentation to guide user on how to include new analysis using the same datasets.
Data analysis concept

• The dataset we use:
  - Samples coming for the ATLAS Experiment.
  - Real data: integrated luminosity of 1.0 fb⁻¹.
  - MC: center-of-mass energy of 8 TeV.

• The histograms we get from the analysis framework contain the real data and MC.
  • By comparing them we can know the quality of the modelling of MC.
  • We can check the physical phenomenon we are interested.
Take $H \rightarrow WW$ for example

- No jets with $p_T > 25$ GeV;
- If leptons have same flavour:
  - $m_{\ell\ell}^{\text{vis}} > 12$ GeV;
  - $|m_{\ell\ell}^{\text{vis}} - m_Z| > 15$ GeV;
  - $E_T^{\text{miss}} > 40$ GeV;
- Else:
  - $m_{\ell\ell} > 10$ GeV;
  - $E_T^{\text{miss}} > 20$ GeV;
- $p_T, \ell \ell > 30$ GeV;
- $\Delta\phi(\ell\ell, E_T^{\text{miss}}) > \pi/2$;
- $m_{\ell\ell} < 55$ GeV;
- $\Delta\phi($leadlep, traillep$) < 1.8$ radians.

In order to reduce the background, we will apply some cut selections.
**Other analysis result**

- **W Analysis**
  - Data and MC have good agreement.
  - Show the stability of the samples and the code.

- **Z Analysis**

- **Top Analysis**

- **WZ Analysis**

- **Z’ Analysis**

---

2018/8/8

Ya-Feng LO . Summer Student Presentation
Conclusion

• To provide a C++ code based on ROOT to compute high energy physics analysis for education and professional training.
• All the development was performed using a versioning system: GitLab.
• Will be ready for public release in the ATLAS and CERN Open Data platforms.
  ➢ http://opendata.atlas.cern
  ➢ http://opendata.cern.ch
Backup
Steps to use framework

• **Download** the data set from Open Data webpage.
  - Data set contain 42 root files for different process MC samples and 2 root files for real data.
  - Real data: integrated luminosity of 1.0
  - MC: center-of-mass energy of 8 TeV with pre-selections below.

<table>
<thead>
<tr>
<th></th>
<th>electrons</th>
<th>muons</th>
<th>jets</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>reconstruction</td>
<td>Muid combined</td>
<td>antiKt4LCTopo</td>
</tr>
<tr>
<td></td>
<td>author 1</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>medium++ quality</td>
<td>$p_T &gt; 5$ GeV</td>
<td>$p_T &gt; 5$ GeV</td>
<td>$p_T &gt; 25$ GeV</td>
</tr>
<tr>
<td></td>
<td>$</td>
<td>\eta</td>
<td>&lt; 2.47$ w/o crack</td>
</tr>
<tr>
<td>Object Quality is Good</td>
<td>MCP Hit requirement.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$</td>
<td>z_0</td>
<td>&lt; 2.0$ mm</td>
</tr>
<tr>
<td></td>
<td>not Converted</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Steps to use framework

• **Download** the framework

• Framework contain:
  - Analysis code:
    - ✓ Main code: read in all the root file.
    - ✓ One source file: where we make cut to do the analysis with TSelector.
    - ✓ Two header files: where we define and book the histograms we are interested.
  - Plotting code:
    - ✓ Read in the output root file from analysis code and apply weight to each process
    - ✓ Output all the histograms we need.
Steps to use framework

• Take W analysis for example:
  – time ./analysis.sh main_W 0
    ✓ 1: parallel mode
    ✓ 0: linear mode
  – Get the output folder: output_W
  – Include the output folder in the plotting code

```cpp
void Plotting_HW::readFiles()
{
    // THIS IS THE DIRECTORY FROM WHERE HISTOS ARE TAKEN
    std::string readname = "/afs/cern.ch/user/y/yl0/Plotting_HWW/Input_HWW_HWW 402 2013 Thermal Top_023 2";
    std::string readname = "/afs/cern.ch/user/y/yl0/Plotting_HWW/Input_HWW_0730";
    std::string readname = "/afs/cern.ch/user/y/yl0/Plotting_HWW/Input_HWW_0730";
    std::string readname = "/afs/cern.ch/user/y/yl0/Plotting_HWW/Input_HWW_0730";
    std::string readname = "/afs/cern.ch/user/y/yl0/Plotting_HWW/input Top 0801";
    std::string readname = "/afs/cern.ch/user/y/yl0/Plotting_HWW/input WS 0901";
}
```

• Run the plotting code:
  ✓ make
  ✓ ./plot
• Get the output folder: histogram

$ ls histograms_W_0802/
hist_etmiss.png
hist_leadjet_eta.png
histLeadjet_jv.png
hist_leadjet_m.png
hist_leadjet_MV.png
hist_leadjet_pt.png
hist_leadlep0.png
hist_leadlept.png
hist_leadlepteta.png
hist_leadleptetc.png
hist_leadleptID.png
hist_leadleptphi.png
hist_leadleptpt.png
hist_leadlepZ0.png
hist_m.png
hist_njets.png
hist_pvxp.png
hist_vismass.png
hist_vxp.png

W analysis

- Exactly one good lepton$^1$ with $p_T > 25$ GeV;
- $E_{T\text{miss}} > 30$ GeV;
- $M_T^W > 30$ GeV.
Z analysis

- Exactly two good leptons with $p_T > 25$ GeV;
- Leptons have opposite charge;
- Leptons have same flavour;
- $|m_{\ell\ell} - m_Z| < 20$ GeV with $m_Z = 91.18$ GeV.

![Graph with ATLAS Open Data and various datasets like Data, Z+jets, W+jets, Diboson, tt, Single top]
Top analysis

- Exactly one good lepton with $p_T > 25$ GeV;
- At least four good jets;
- At least two b-tagged jets (MV1@70%);
- $E_{T}^{\text{miss}} > 30$ GeV;
- $m_T^W > 30$ GeV.
WZ analysis

- Exactly three good leptons with $p_T > 25$ GeV;
- WZ candidate is chosen by finding the Z boson candidate closest to the nominal Z mass;
- $|m_{\ell\ell} - m_Z| < 10$ GeV with $m_Z = 91.18$ GeV;
- $m_T^W > 30$ GeV.
Z' analysis

- Exactly one good lepton with \( p_T > 25 \text{ GeV}; \)
- At least four good jets;
- At least one b-tagged jet (MV1@70%);
- \( E_{\text{T}}^{\text{miss}} > 30 \text{ GeV}; \)
- \( m_T^W + E_{\text{T}}^{\text{miss}} > 60 \text{ GeV}. \)