ATLAS Open Data

software for visualization and physics analysis

Ya-Feng LO

Arturo Rodolfo SANCHEZ PINEDA, Leonid SERKIN

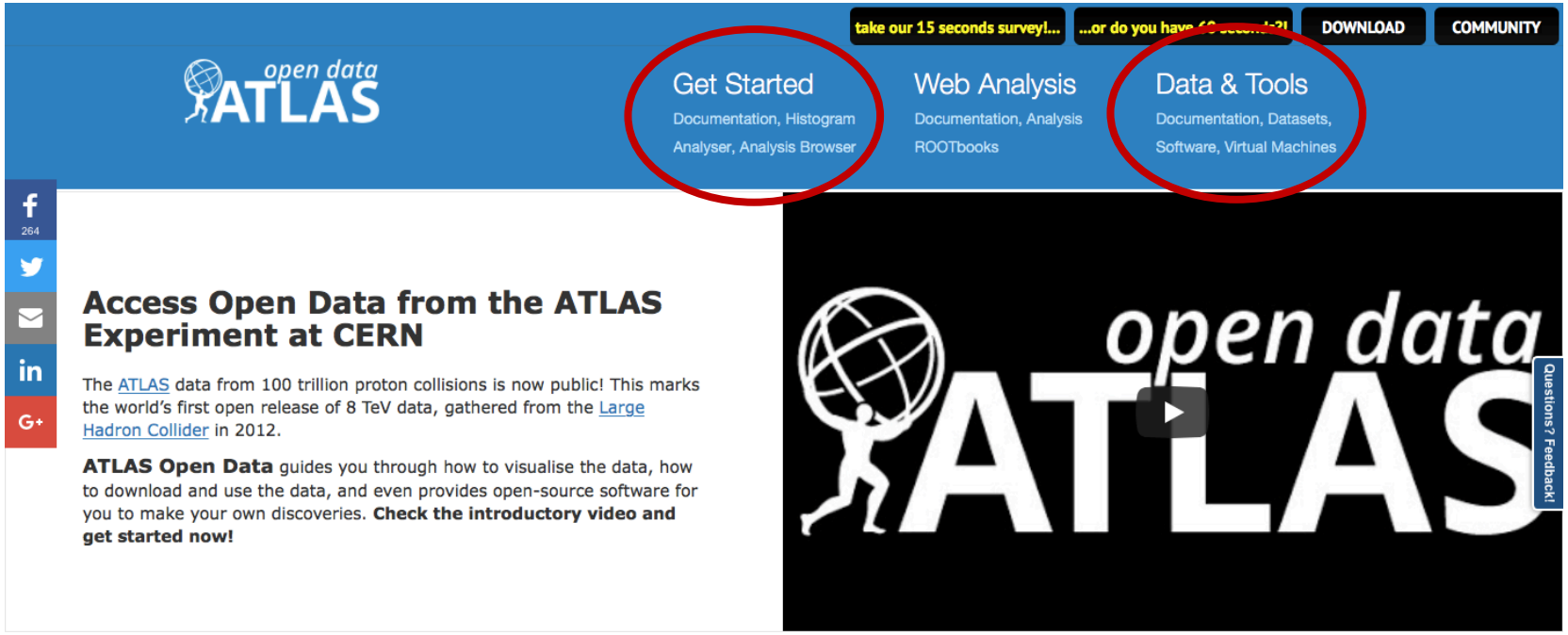
Outline

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

ATLAS Open Data

Check out the webpage!
http://opendata.atlas.cern

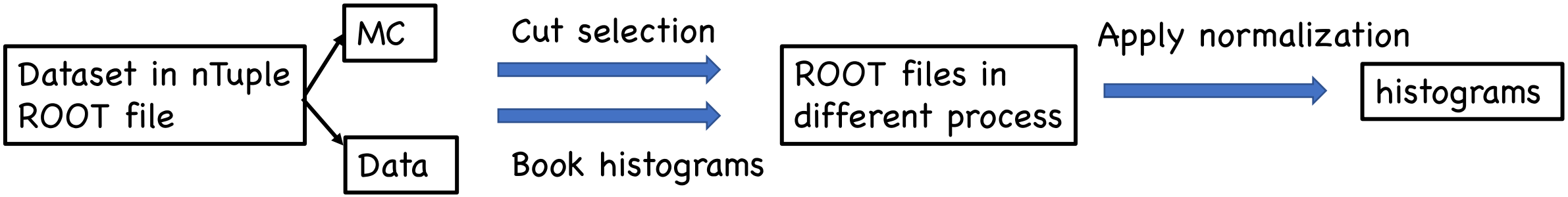
- Aim to release real and simulated data, together with software resources to analyze those samples.
- Provide detailed documentation and software training for people all around the world to access high energy physics at CERN.



Software's description

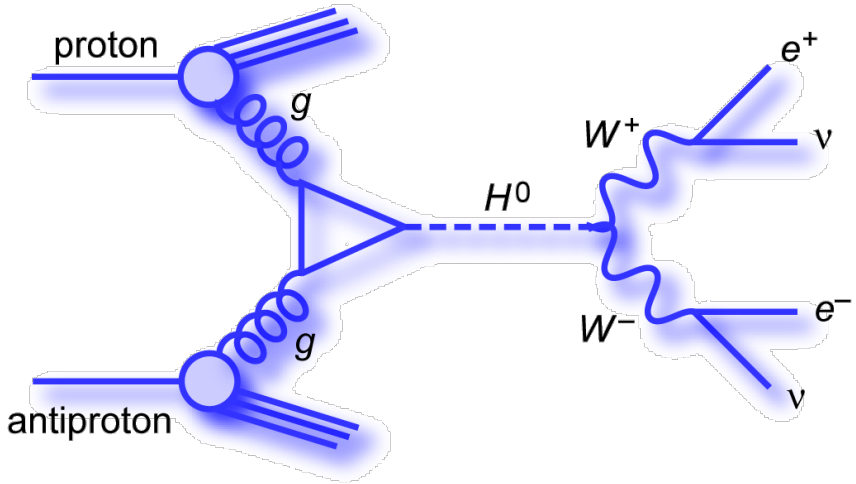
- Build a framework based on ROOT to analyses high energy physics datasets.
- Use C++ programming language
  - One of the main programming languages for high energy physics.
  - This is one important piece still missing in the set of ATLAS Open Data public analysis code.
  - Improve the speed and the 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



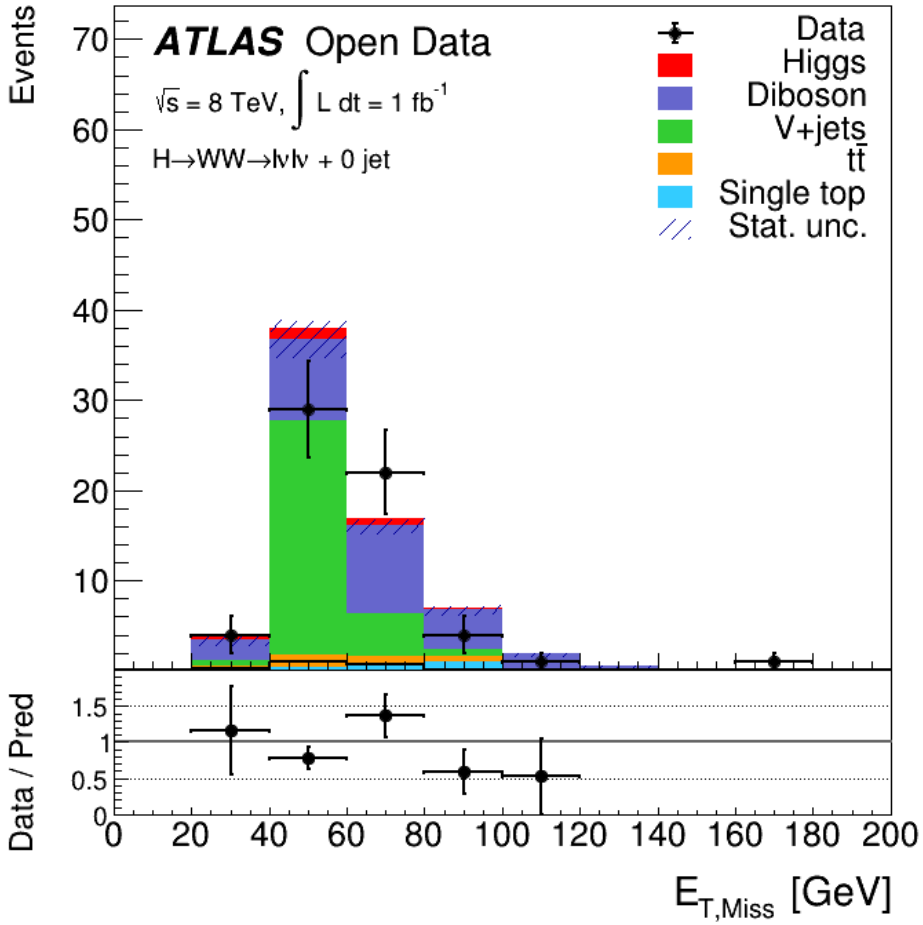
- The dataset we use :
  - Samples coming for the ATLAS Experiment.
  - Real data: integrated luminosity of 1.0 fb^{-1}.
  - MC: center-of-mass energy of 8 TeV.
- The histograms we get from the analysis framework contain the real data and MC.
  - By comparing them we can know the quality of the modelling of MC.
  - We can check the physical phenomenon we are interested.

Take H->WW for example



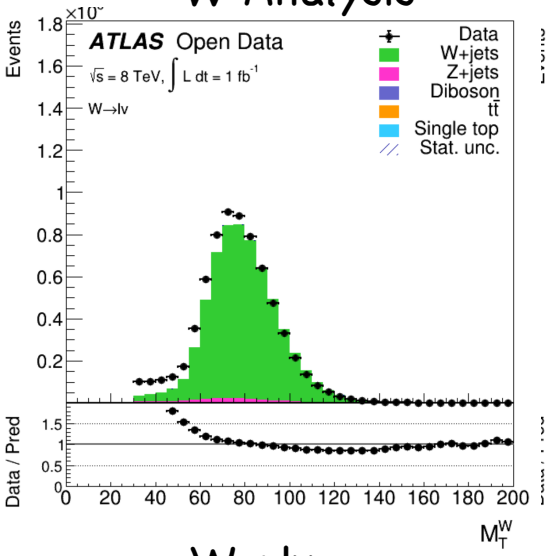
In order to reduce the background, we will apply some cut selections ⇨

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



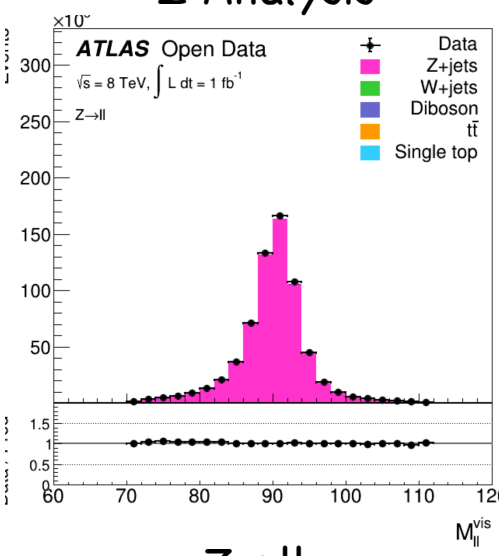
Other analysis result

W Analysis



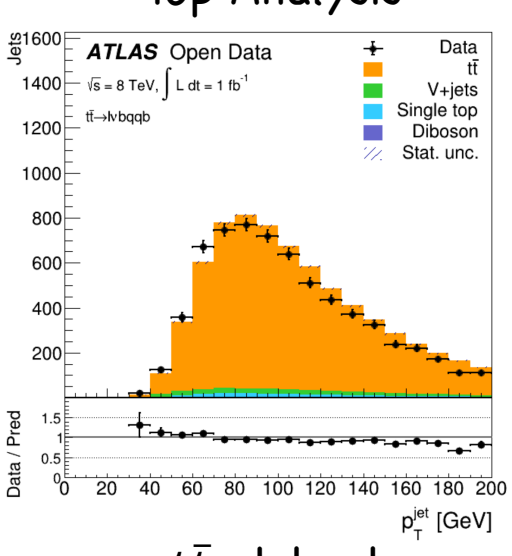
W->lv

Z Analysis



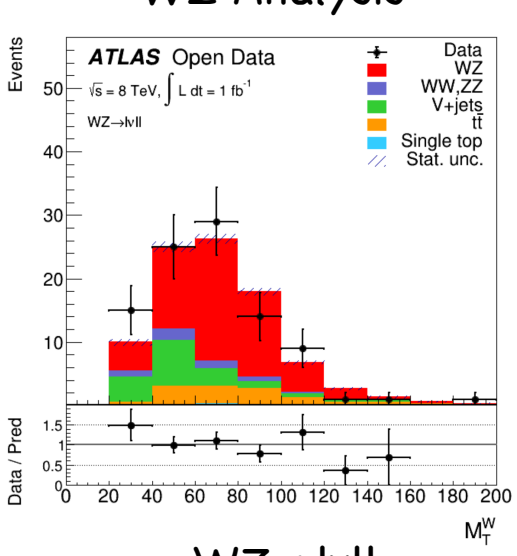
Z->ll

Top Analysis



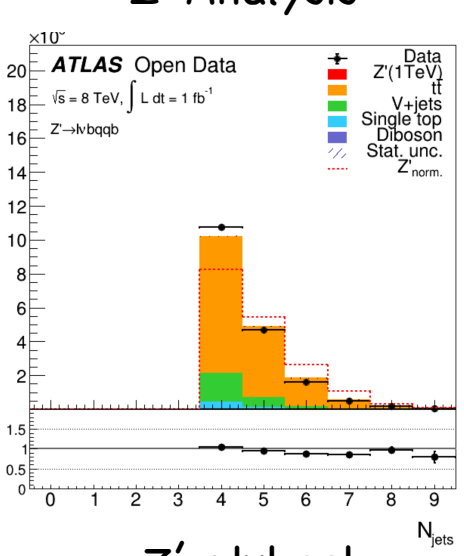
$t\bar{t}$->lvbqqb

WZ Analysis



WZ->lvll

Z' Analysis



Z'->lvbqqb

- Data and MC have good agreement.
  - ➢Show the stability of the samples and the code.

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.

electrons	muons	jets
reconstruction author 1||3	Muid combined	antiKt4LCTopo
medium++ quality	tight quality	jet cleaning (veto BadLooseMinus)
$p_T > 5$ GeV	$p_T > 5$ GeV	$p_T > 25$ GeV
$\|\eta\| < 2.47$ w/o crack	$\|\eta\| < 2.5$	$\|\eta\| < 2.5$
Object Quality is Good	MCP Hit requirement.	
$\|z_0\| < 2.0$ mm	$\|z_0\| < 2.0$ mm	
not Converted		

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 out put 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 out put folder in the plotting code

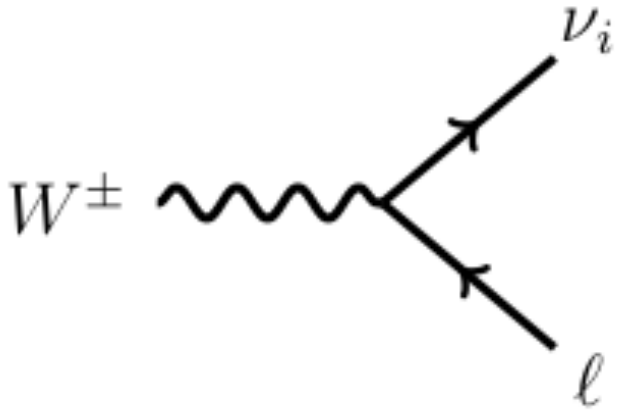
```
///
void Plotting_HWW::readFiles(){

  // THIS IS THE DIRECTORY FROM WHERE HISTOS ARE TAKEN
  std::string readname = "/afs/cern.ch/user/y/ylo/Plotting_HWW/Input_HWW_0723_2";
  //std::string readname = "/afs/cern.ch/user/y/ylo/Plotting_HWW/Input_W_0730";
  //std::string readname = "/afs/cern.ch/user/y/ylo/Plotting_HWW/Input_Z_0730";
  //std::string readname = "/afs/cern.ch/user/y/ylo/Plotting_HWW/Input_Top_0801";
  //std::string readname = "/afs/cern.ch/user/y/ylo/Plotting_HWW/Input_WZ_0801";
```

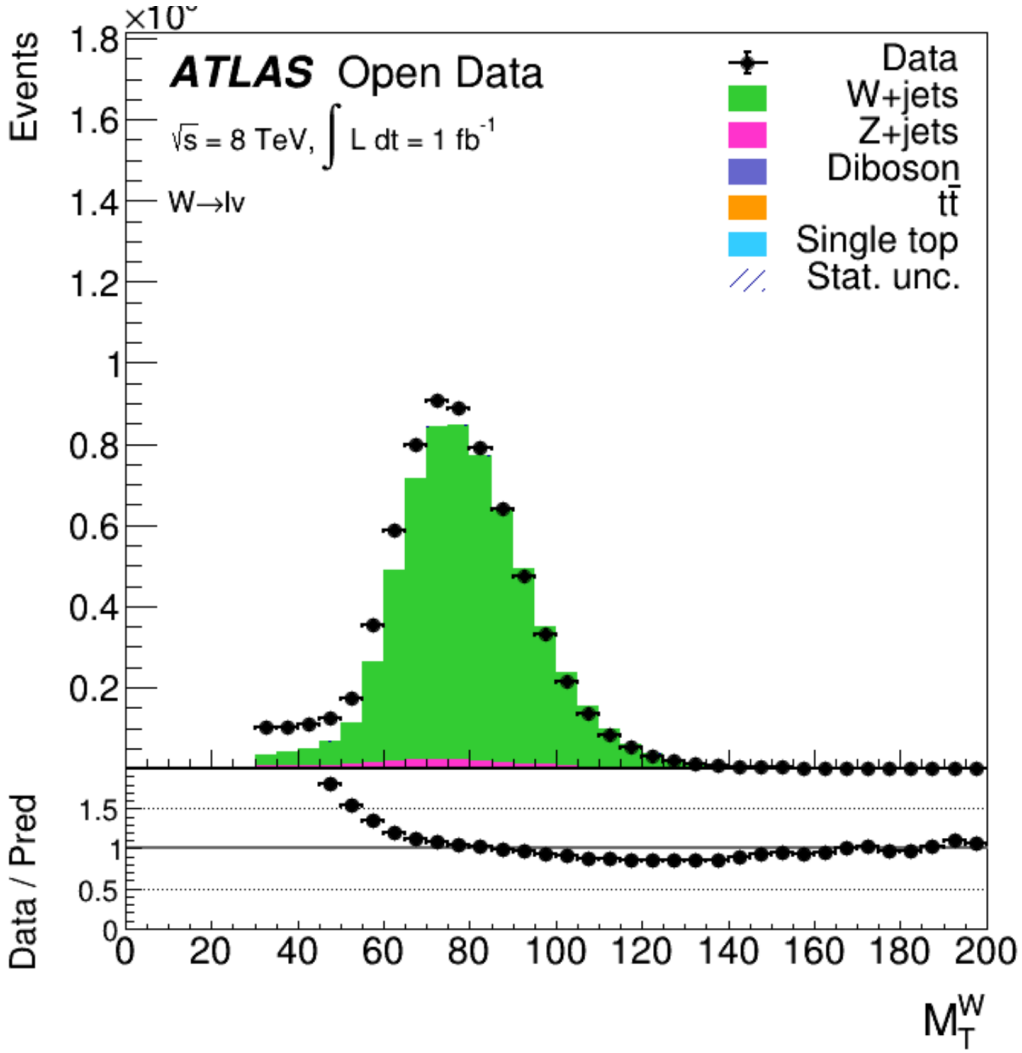
  - Run the plotting code:
    - ✓ make
    - ✓ ./plot
  - Get the output folder: histogram

```
$ ls histograms_W_0802/
hist_etmiss.png
hist_leadjet_eta.png
hist_leadjet_jvf.png
hist_leadjet_m.png
hist_leadjet_MV1.png
hist_leadjet_pt.png
hist_leadlepd0.png
hist_leadleptch.png
hist_leadleptE.png
hist_leadlepteta.png
hist_leadleptetc.png
hist_leadleptID.png
hist_leadleptphi.png
hist_leadlept_ptc.png
hist_leadleptpt.png
hist_leadlepz0.png
hist_mt.png
hist_n_jets.png
hist_pvxp_n.png
hist_vismass.png
hist_vxp_z.png
```

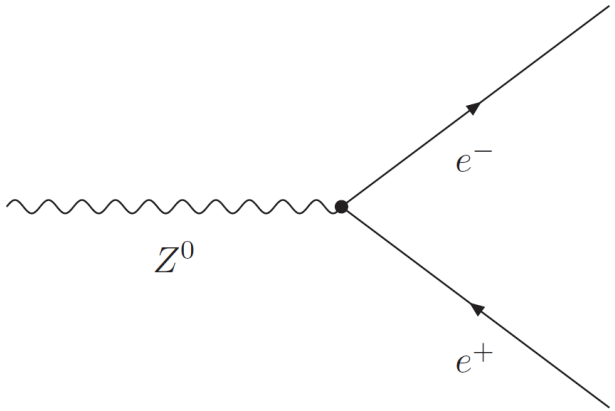
W analysis



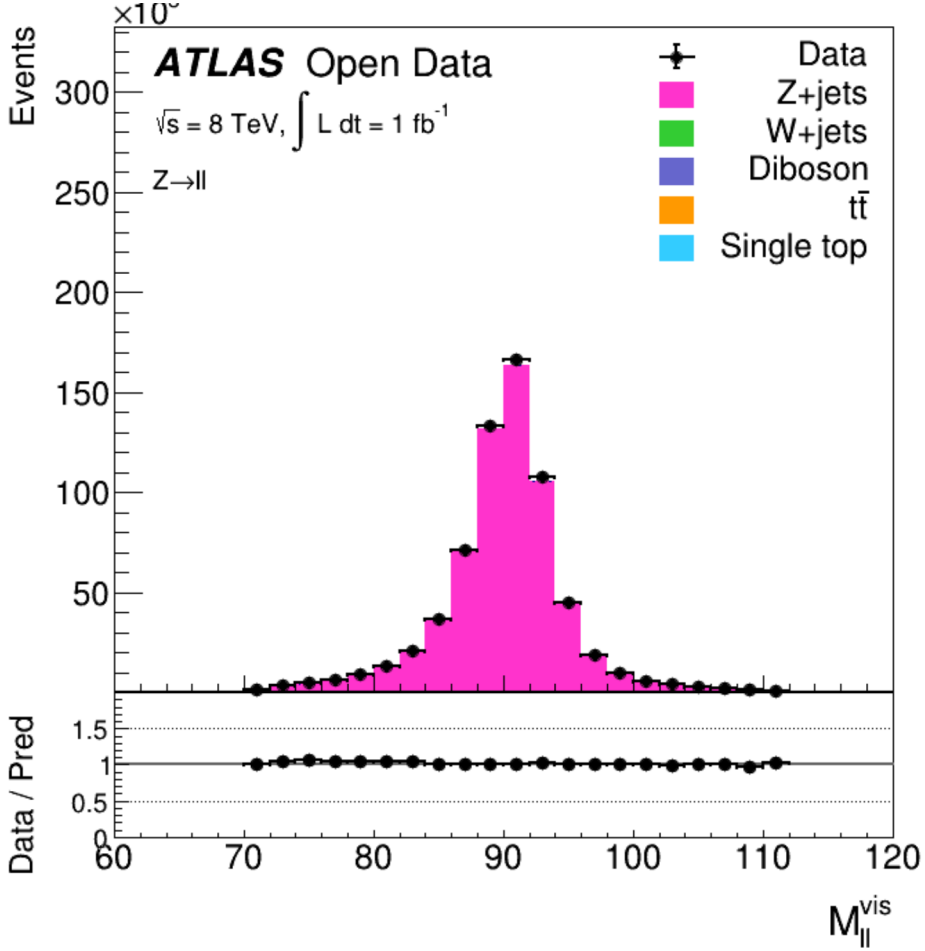
- Exactly one good lepton[1] with $p_{\mathrm{T}} > 25$ GeV;
- $E_{\mathrm{T}}^{\mathrm{miss}} > 30$ GeV;
- $M_{\mathrm{T}}^{W} > 30$ GeV.



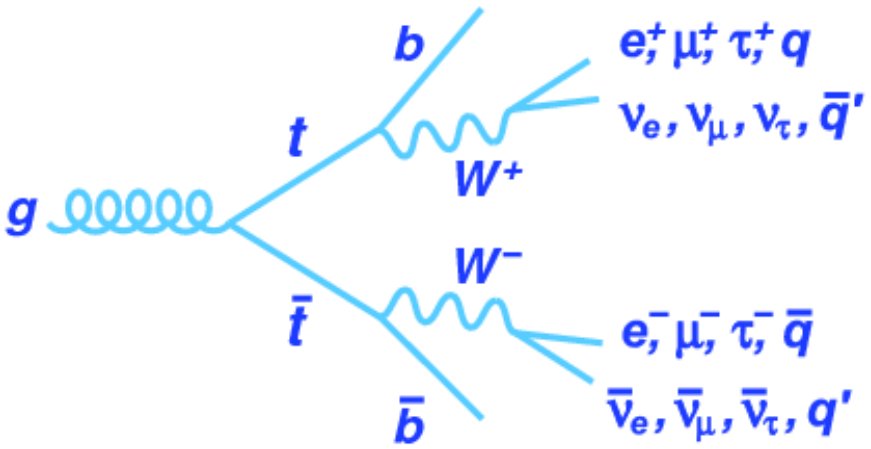
Z analysis



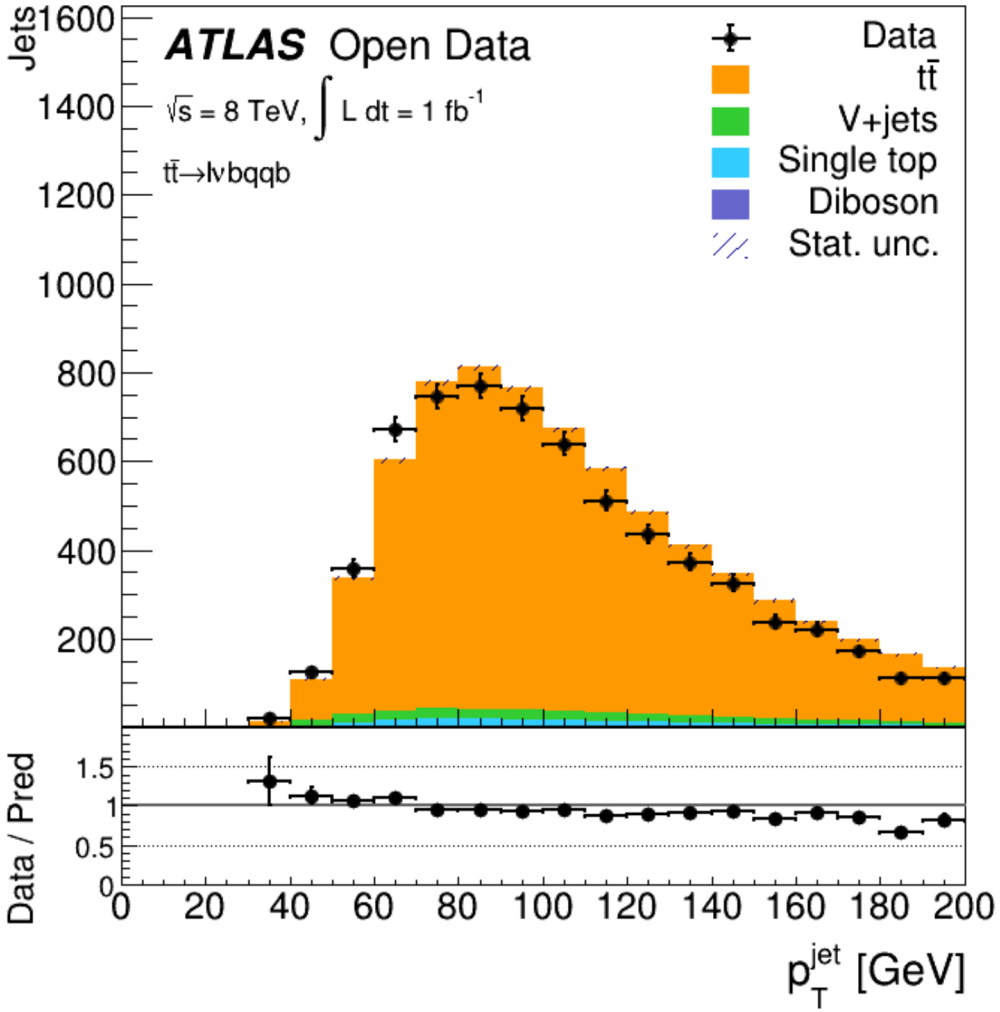
- Exactly two good leptons with $p_T > 25$ GeV;
- Leptons have opposite charge;
- Leptons have same flavour;
- $|m_{\ell\ell} - m_Z| < 20$ GeV with $m_Z = 91.18$ GeV.



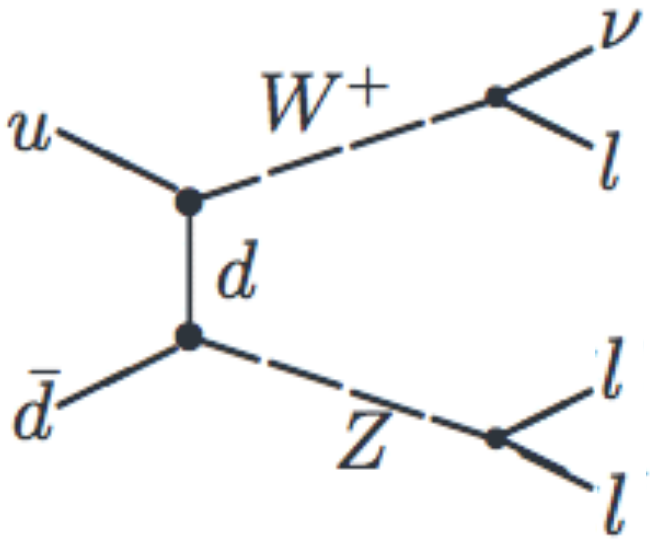
Top analysis



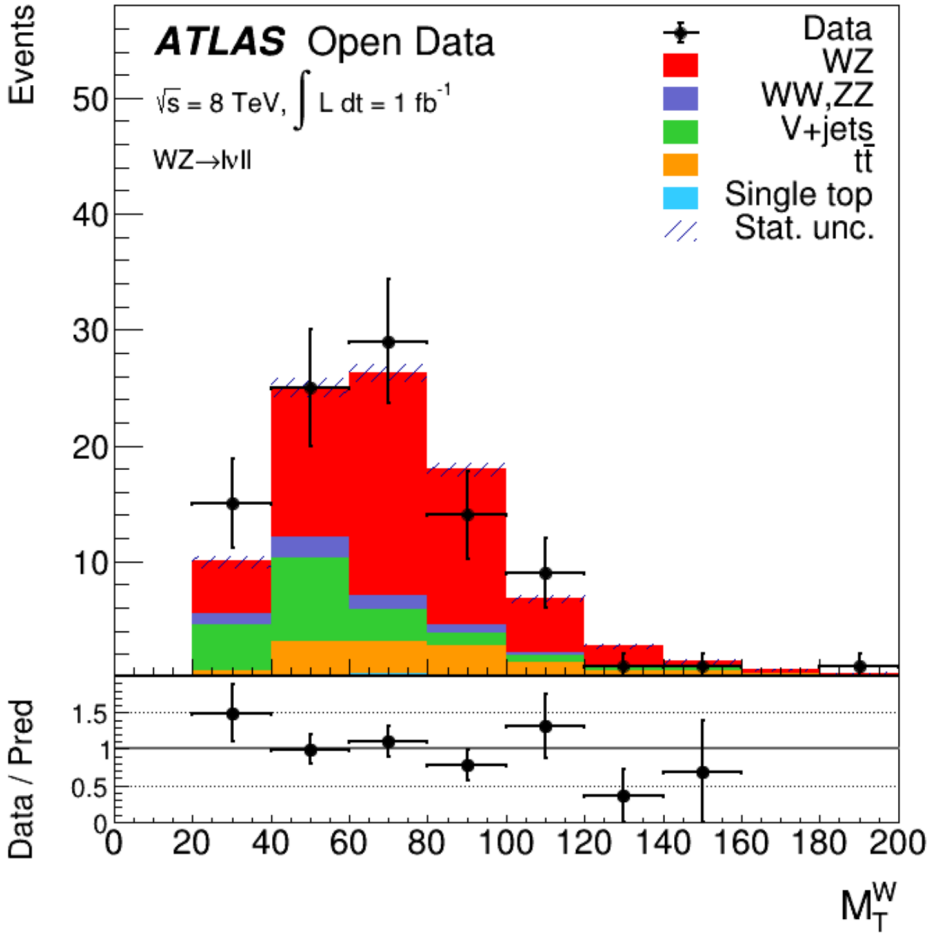
- Exactly one good lepton with $p_{\mathrm{T}} > 25$ GeV;
- At least four good jets;
- At least two b-tagged jets (MV1@70%);
- $E_{\mathrm{T}}^{\mathrm{miss}} > 30$ GeV;
- $m_{\mathrm{T}}^{W} > 30$ GeV.



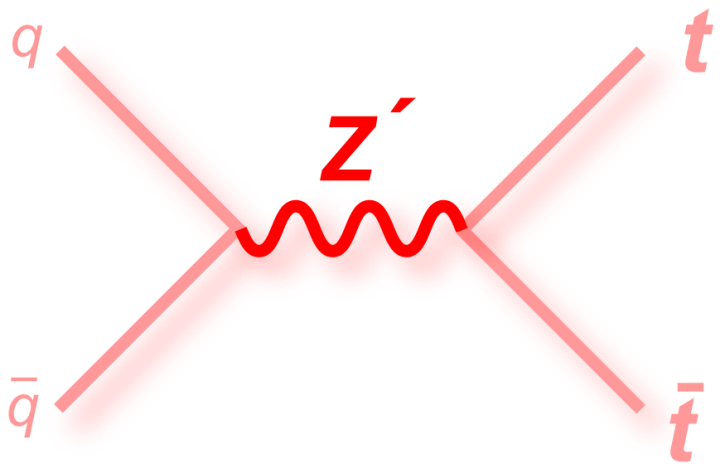
WZ analysis



- 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 $p_{\mathrm{T}} > 25$ GeV;
- At least four good jets;
- At least one b-tagged jet (MV1@70%);
- $E_{\mathrm{T}}^{\mathrm{miss}} > 30$ GeV;
- $m_{\mathrm{T}}^{W} + E_{\mathrm{T}}^{\mathrm{miss}} > 60$ GeV.