
THIRD YEAR LABORATORY

ANALYSING EVENTS WITH W AND Z BOSONS: AN INTRODUCTION TO PARTICLE PHYSICS DATA WITH ATLAS

Version 1.06 November 2018

This analysis experiment and the associated documentation was produced by Manchester summer student **Nathan Simpson** and Prof. **Terry Wyatt**.

School of Physics & Astronomy
The University of Manchester

1 Aims

This project aims to give you some appreciation of the physics processes that can occur in high energy proton-proton collisions at the LHC and to introduce you to event selection and measurement methods used in particle physics data analysis.

2 Objectives

1. Design criteria to select the different sorts of particles and types of events produced at the LHC and eliminate unwanted “backgrounds”. The initial focus is on events containing the decays of Z and W bosons to electrons and muons.
2. Compare real data collected by the ATLAS experiment with computer-simulated “Monte Carlo” data.
3. Measure the cross sections for the production of Z and W bosons at the LHC and estimate statistical and systematic uncertainties on the results.
4. Gain experience of using `ROOT` — a C++ framework designed at CERN to analyse large data sets.
5. (Optional) Extend the event selections and measurements to other sources of events at the LHC, such as those containing tau leptons, pairs of vector bosons, top quarks, or potential sources of new physics from “beyond the standard model”.

	I	II	III
charge = -1	electron (e)	muon (μ)	tau (τ)
charge = 0	electron neutrino (ν_e)	muon neutrino (ν_μ)	tau neutrino (ν_τ)

Table 1: The three generations of leptons.

3 Introduction

This document is designed to be used electronically. Websites/sections of this document are appropriately hyperlinked. Hyperlinks are indicated by underlined text. For example, here is a link to the ROOT users' guide. So, until e-paper becomes standard at educational establishments, printing of this document should not be considered as absolutely necessary. For ease of reference all the hyperlinks distributed throughout the document are also collected together in subsection 7.1.

4 Background reading

4.1 A femto-course in particle physics

Since many students (especially those doing this lab. in the 1st semester) have not had a formal introduction to particle physics at this point in the academic year, a brief introduction to particle physics is given here.

The most well-tested and robust theory of particle physics that we have today is called the **standard model**. In this model, all of matter is made up of fundamental constituent particles that come in two categories: quarks and leptons. We sort these particles into **three generations of matter** (I, II, III). Each successive generation contains two quarks and two leptons, making up twelve particles in total. If we then factor in the existence of **antiparticles** — particles with opposite charge to their counterparts, denoted with a bar (e.g. the up antiquark, \bar{u}) — we have twelve additional particles (six antiquarks and antileptons, three of which are charged and three of which are neutral antineutrinos).

4.1.1 Leptons

Leptons are spin- $\frac{1}{2}$ fundamental particles. The term spin refers to the particles' intrinsic angular momentum, which is given in terms of \hbar . They can have charges of -1 , $+1$ (for antiparticles), and 0 in units of the electron charge $e \approx 1.6 \times 10^{-19}$ C. Each generation of leptons has a flavour — electron, muon, or tau — and consist of a charged lepton and a neutrino. These leptons are shown in Table 1. "Flavour" can simply be thought of as a way of categorizing particles into species.

	I	II	III
charge = $\frac{2}{3}$; up-type	up (u)	charm (c)	top (t)
charge = $-\frac{1}{3}$; down-type	down (d)	strange (s)	bottom (b)

Table 2: The three generations of quarks.

Particle	Force mediated	Mass (GeV)	Charge (e)
Z	Weak	91.1876 ± 0.0021	0
W^+/W^-	Weak	80.385 ± 0.015	$+1/-1$
Gluon (g)	Strong	0	0
Photon (γ)	Electromagnetic	0	0
Graviton (hypothetical)	Gravity	0	0

Table 3: The force-mediating bosons. The hypothetical graviton is included for completeness, although we have so far found no experimental evidence of its existence.

4.1.2 Quarks

Quarks are also spin- $\frac{1}{2}$ fundamental particles, but differ from leptons in a number of aspects, including charge, mass, etc. There are two categories of quarks: **up-type** quarks with charge $\frac{2}{3}$ and **down type** quarks with charge $-\frac{1}{3}$. As far as we know, there are twelve quarks in total — three up-type, three down-type, and their corresponding antiparticles. A summary of the known quarks can be found in Table 2. Quarks can bind together under the influence of the strong interaction to form different types of composite particles called **hadrons**. These can be **baryons** (containing three quarks, or three antiquarks for antibaryons) or **mesons** (containing a quark and an antiquark).

4.1.3 Bosons

In addition to quarks and leptons, we have **bosons**, a category of particles that are responsible for mediating the forces of nature. These are listed in Table 3.

4.2 Interactions & Feynman diagrams

The main interactions of interest for this experiment are the production and decay of Z and W bosons. The most easily identified decay mode of the Z boson is to a pair of charged leptons. We write this as $Z \rightarrow l^+l^-$, where l denotes a particular lepton flavour ($l = e, \mu, \tau$). An analogous decay mode of the W^- boson is to a lepton and an antineutrino. We write this as $W^- \rightarrow l^- \bar{\nu}_l$. We can represent such processes with **Feynman diagrams**, as shown in Figure 1. Note that **electric charge is conserved at each vertex**. To learn more about Feynman diagrams see, e.g., [this resource](#).

Although the beams of particles accelerated in the LHC are protons, the collisions we are most interested in here occur between individual quarks or gluons within the colliding protons. We use the collective term **parton** to describe a constituent of a hadron (be it a quark, an antiquark, or a gluon).

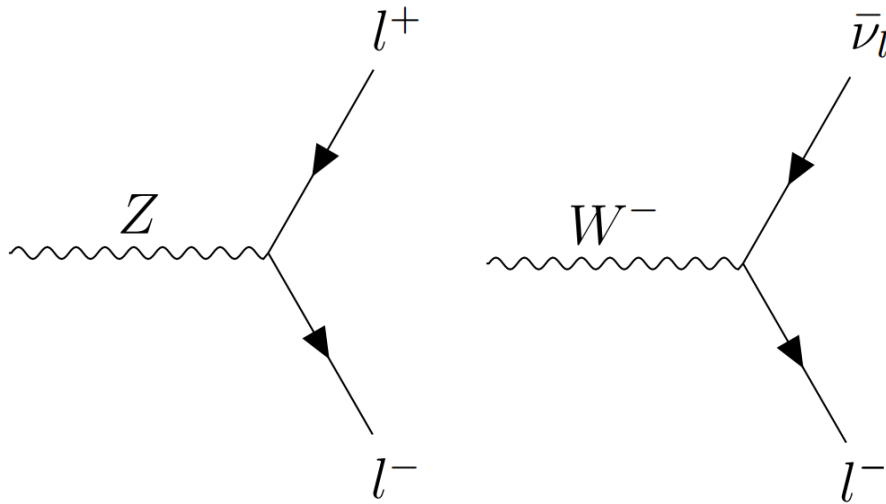


Figure 1: The Feynman diagrams for the decay processes $Z \rightarrow l^+l^-$. (left) and $W^\pm \rightarrow l^\pm \nu_l$ (right).

4.3 The experiment

Information about the ATLAS collaboration, the detector, the physics etc. can be found [here](#). It is worth knowing your way around the detector and the experiment, as it will add some valuable context to the work you do.

A journal paper describing the measurements of Z and W production in the very early ATLAS data at the LHC can be found [here](#). Hopefully you can manoeuvre your way around some of the the technical jargon and get an idea of the kind of analysis that can be performed on the data.

4.4 About ATLAS OpenData

ATLAS OpenData is a project designed by the outreach team at ATLAS to enable the curious individual to explore the techniques used in modern experimental particle physics. They have produced several (largely unmodified) data sets from the actual 2012 ATLAS run for use in educational analysis.

For details on the data sets you are using, you can consult the [OpenData online software book](#) for reference. In addition, links to relevant pages will be given throughout this document where appropriate.

5 Getting started

To log in to your Linux machine, you should use the username and password given to you by your demonstrator¹. Once you are logged on, you just need to follow a couple simple steps to get started.

¹I recommend you log out at the end of each lab. day and log back in at the beginning of the next day. The machines occasionally behave erratically when users stay logged in for weeks at a time!

Throughout this section, the following convention will be used:

Here will be a Linux command in bold that you should type (or copy and paste) into your terminal...

...and here will be the explanation of its function.

5.1 Setting up the code — *To do once at the start of the experiment:*

- Open a terminal window by clicking on the icon at the top of your homescreen, or click on Applications → System tools → Terminal.
- Type in (or copy and paste) the following lines:

cp -r /opt/ATLAS-Software .	Copies starting material to your local directory.
cd ATLAS-Software	Changes your “current directory” to ATLAS-Software.
ln -s /data/ATLAS-data/Data Data	Makes a “symbolic link” in your local directory to the ATLAS data files...
ln -s /data/ATLAS-data/MC MC	...and again for the Monte Carlo data sets.
mkdir OutputFiles	Makes a sub-directory called <i>OutputFiles</i> , which is where your histograms will be written by ROOT.
mkdir OutputPlots	Makes a sub-directory called <i>OutputPlots</i> , which is where you will save pictures of selected histograms for archiving and printing.
ls -l	Displays the files/folders in your current directory and their permissions.

In your directory ATLAS-Software, you should now be able to see the following files and sub-directories:

atlas-experiment-lab-script.pdf (this lab script) || *MyAnalysis.C* || the directory ‘Headers’ || *atlas-setup.sh* || *LayerHist.C* || *mini.C* || *MultiRun.py* || *MultiRun.txt.py* || *AddHist.C* || the sub-directories *OutputFiles* and *OutputPlots* || Symbolic links to the directories ‘Data’ and ‘MC’.

Task 5.1. *Try to get yourself set up using the above instructions. Talk to your demonstrator at any point if something is not clear. Remember that you need to do this only once, at the beginning of your first lab. day!*

5.2 To do every time you open a new terminal window:

<code>cd ATLAS-Software</code>	Changes your directory to ATLAS-Software
<code>source atlas-setup.sh</code>	Installs latest version of ROOT, the data analysis software used in this experiment.

Tip 1. *The above commands are pretty much the only bits of Linux you will need to complete this lab. successfully.*

A few simple Linux tricks can help minimise the amount of repetitive typing you have to do:

*You can use the “**tab**” key on the keyboard to “auto-complete” the names of commands, files, and directories. For example, if you type ‘**source at**’ and then press “**tab**”, it should complete the filename for you and say ‘**source atlas-setup.sh**’.*

*You can use the “**up arrow**” key to recall commands that you have typed previously.*

At the end of the day in the lab. we suggest you log off. Ensuring a fresh start by logging on at the beginning of each new lab. day is recommended.

5.3 Using MyAnalysis.C

The only files you should need to edit in this directory are *MyAnalysis.C* and *MyAnalysis.h* (the latter is found in the ‘Headers’ folder). The former contains three functions to book, fill, and write to file histograms, whilst the latter is where you declare the pointers to your histograms to let the program know about them.

It is worth mentioning at this point that **this is not a computing project**. Your knowledge of C/C++ for this experiment need not extend past the use of `if` and `for` statements, which you should remember from last year. For those who need a refresher, please see the resources available on, e.g., cplusplus.com.

You should now open *MyAnalysis.C* and *MyAnalysis.h* with your text editor of choice — included on these computers are ‘*gedit*’ and ‘*Emacs*’.

Task 5.2. *Try and get familiar with the layout of the code in MyAnalysis.C. Make sure to add your name and the date at the top in a comment to help you keep records!*

As you can see, the code contains some very simple histograms. The code uses the vector `lep_pt` — defined as the vector of the lepton momenta in the plane transverse to the beam direction² — and has one instance per event, e.g. for some event with three leptons, `lep_pt` will contain the three³ leptonic momenta at `lep_pt[0]`, `lep_pt[1]`, and `lep_pt[2]`. The units in which momenta and energies are given on the ntuple is MeV.

²The plane transverse to the beam direction is the ‘*x-y* plane’ in cartesian coordinates. The *z* coordinate points along the beam direction.

³Remember that in C++ the first element in a vector is at the location `[0]` and not `[1]`.

The definition of all the available variables on the ATLAS OpenData data sets for this experiment can be found in [this web-page](#). The variable names to use in your *MyAnalysis.C* are given in the column labelled **ntuple branchname**. In the code the variable names are all declared in the file *mini.h* in the folder *Headers*. The values of energies and momenta are all given in units **MeV**!

5.4 Running the code and examining the results

Before you try adding to the code, you should try running the existing code and check you can view the output. Enter the following commands:

python MultiRun.py	In terminal, you run a python script by typing python filename , where in our case, we are running the analysis macro <i>MultiRun.py</i> .
Zee	The code will ask you to list a set of decay chains, separated by a comma. Here we just want to analyse the Monte Carlo simulation of $Z \rightarrow ee$ events.
root	Starts a new ROOT session.
new TBrowser	Opens the ROOT object browser.

Tip 2. ‘Zee’ was the string code you entered to tell *MultiRun.py* which decay chain to analyse. For future reference, the full set of “string codes” for different data sets are shown in Table 4.

Tip 3. You can tell *MultiRun.py* to analyse more than one data set by responding to the prompt with a list of string codes separated by commas — crucially, **you should not put spaces between the commas!**

ROOT should then display a directory structure in the *TBrowser* window. You should now see a file in your directory *OutputFiles* called *Zee.root*. To view the histograms you just created, double click on this file — inside you should see three histograms. You can view these in the browser by double clicking.

At the end of your ROOT session you can exit by typing:

.q	Ends ROOT session and returns to the Linux terminal shell.
-----------	--

If all this was smooth sailing, you are ready to move on and try and make your own histograms!

5.5 More information about the code in *MyAnalysis.C* and *MyAnalysis.h*

The file *MyAnalysis.C* contains three functions:



<u>Monte Carlo data sets</u>	
physics process	string code
$Z \rightarrow ee$	'Zee'
$Z \rightarrow \mu\mu$	'Zmumu'
$Z \rightarrow \tau\tau$	'Ztautau'
Drell-Yan ee $8 < m_{\ell\ell} < 15$ GeV	'DYeeM08to15'
Drell-Yan ee $15 < m_{\ell\ell} < 40$ GeV	'DYeeM15to40'
Drell-Yan $\mu\mu$ $8 < m_{\ell\ell} < 15$ GeV	'DYeeM08to15'
Drell-Yan $\mu\mu$ $15 < m_{\ell\ell} < 40$ GeV	'DYeeM15to40'
$W \rightarrow e\nu$ (b veto, no jets)	'WenuNoB'
$W \rightarrow e\nu$ (b veto, with jets)	'WenuNoBJets'
$W \rightarrow e\nu$ (with b , with jets)	'WenuB'
$W \rightarrow \mu\nu$ (b veto, no jets)	'WmunuNoB'
$W \rightarrow \mu\nu$ (b veto, with jets)	'WmunuNoBJets'
$W \rightarrow \mu\nu$ (with b , with jets)	'WmunuB'
$W \rightarrow \tau\nu$ (b veto, no jets)	'WtaunuNoB'
$W \rightarrow \tau\nu$ (b veto, with jets)	'WtaunuNoBJets'
$W \rightarrow \tau\nu$ (with b , with jets)	'WtaunuB'
$t\bar{t} \rightarrow leptons$	'ttLep'
$t\bar{t} \rightarrow hadrons$	'ttHad'
WW pair production	'WW'
WZ pair production	'WZ'
ZZ pair production	'ZZ'
$H \rightarrow ZZ \rightarrow leptons$ (gluon–gluon)	'ggH4lep'
$H \rightarrow ZZ \rightarrow leptons$ (vector boson fusion)	'VBFH4lep'
<u>ATLAS real data sets</u>	
lepton type selected	string code
Electron data	'Electrons'
Muon data	'Muons'

Table 4: The string code corresponding to each data set that you should type in *MultiRun.py* when you want to analyse that particular data set.

- Function *Book* is where you create or ‘book’ new histograms. This function is called only once per input file of events. It is called BEFORE any events are read in.
- Function *FillHist* is where you select events and enter information into (or ‘fill’) your histograms. This function is called ONCE PER EVENT.
- Function *Style* is where you can modify your histograms in various ways and write them to an output file. This function is called only once per input file. It is called AFTER all the events have been processed by Function *FillHist*.

The file *MyAnalysis.h* is where you declare the C++ pointers to your histograms to let the program know about them.

Task 5.3. By modifying the examples in the code, try to plot histograms of lepton p_T in the $Z \rightarrow ee$ data set separately for positively charged and negatively charged leptons and separately for electrons and for muons. Remember, you will need to declare the pointer to each new histogram in *MyAnalysis.h*. In *MyAnalysis.C* you will need to “book” or create the histograms and write them to the output file as well as filling them.

In order to keep a permanent record of a set of histograms for future reference it can be useful to rename the output file so that it does not get overwritten the next time you run through the same data set. This can be achieved using the Linux command **mv**

mv OutputFiles/Zee.root OutputFiles/ZeeNewHist.root	Rename the file <i>Zee.root</i> in the directory <i>OutputFiles</i> to <i>ZeeFirstNewHistograms.root</i>
--	--

In a similar vein, you may want to consider keeping an ‘archive’ version of your code at points at which you have completed one chunk of the analysis and are about to embark on major code modifications.

It can be useful to create a hard copy of key plots. This can be done in the following way:

1. From root save the plot as a .pdf file to the subdirectory *OutputPlots*
2. Open the .pdf file with the program *Adobe Reader*.
3. Click *print* and select:
 - i-p00-sr-7835-0452 for colour printer
 - i-p00-sr-5955-0056 for black and white

Both printers are on the 6th floor (research wing).

For more information on booking, filling, and formatting histograms, please see the relevant sections in the [online ROOT documentation](#).

Tip 4. You may have noticed that in the directory “ATLAS-Software” there is a python script called *MultiRun_txt.py* — this works exactly like *MultiRun.py*, but asks for the name of a .txt file as input. You can use it by creating a .txt file in the same directory with a list of string codes separated by a comma (and ending with a comma) — crucially, **you should not put**

6 Analysis

Now we can get started with some actual physics analysis!

6.1 General Introduction

Each of the separate “Monte Carlo” data sets given in Table 4 represents a computer simulation of a specific physics process. For example, in every event in the file corresponding to the string code ‘*Zmumu*’ a Z boson has been produced and has decayed to an opposite-charge pair of muons ($Z \rightarrow \mu\mu$). In contrast, the files corresponding the real data from the ATLAS experiment contain all relevant physics processes mixed up together within the one file. One of the main challenges of this analysis project is to figure out how to select a sample of events corresponding to a specific physics process from the ATLAS data. The difference between the two real data files is that one (‘*Electrons*’) has been “pre-selected” by requiring at least one electron candidate with high p_T and the other (‘*Muons*’) by requiring at least one muon candidate with high p_T .

We shall use the individual Monte Carlo samples to allow us to find out about the characteristic properties of events that correspond to particular physics processes. This will enable us to design “selection cuts” that we can apply to select a sample of events in the ATLAS data that correspond to the desired particular physics processes (our “signal”). In choosing our selection cuts there is always a trade off between the conflicting aims of achieving as high as possible a selection efficiency for the signal and rejecting as much as possible of the “background” from other physics processes. There is no unique “right answer” to such problems. Whilst it is important to achieve a reasonable compromise in the choice of selection cuts, it is even more important to make estimates of the selection efficiency and level of residual background and to come up with credible statistical and systematic uncertainties on these estimates.

In each event in a particular Monte Carlo data set the specified physics process (e.g., the $Z \rightarrow \mu\mu$) will have been generated. Note, however, that the generated particles are not always actually observed in the ATLAS detector. Note also that the events may contain additional particles (e.g., jets of hadrons produced by radiation from the incoming partons before the annihilation that produced the Z boson).

6.2 Analysing Z bosons

Using relativistic kinematics, one can derive an expression for the invariant mass of a parent particle that decays into multiple daughter particles. For example, in the simple decay mode of a Z boson to two muons ($Z \rightarrow \mu\mu$), we can use 4-momentum conservation and a little bit of algebra to find out the invariant mass of the lepton pair system $m_{\ell\ell}$.

ATLAS uses a right-handed coordinate system, with its origin at the nominal interaction point in the centre of the detector and the z -axis along the beam pipe. In the transverse plane, the x -axis points from the interaction point to the centre of the LHC ring, the y -axis points upwards.

In the data, the kinematic variables you are given for each lepton are the transverse momentum (p_T), the azimuthal angle between the p_T and the x -axis (ϕ), and the pseudorapidity (η). Pseudorapidity is defined as

$$\eta = -\ln \left(\tan \left(\frac{\theta}{2} \right) \right), \quad (1)$$

where θ is the usual polar angle. For massless particles, this is equivalent to the rapidity (y), defined as

$$y = \frac{1}{2} \ln \left(\frac{|\mathbf{p}| + p_L}{|\mathbf{p}| - p_L} \right). \quad (2)$$

where p_L is the longitudinal momentum and \mathbf{p} is the three-momentum of the particle. Since you will most likely be far more comfortable working in terms of p_x , p_y and p_z , you will need to make the appropriate change of variables in order to plot a histogram with the data given.

Task 6.1. Derive a conversion from the kinematic variable set (p_T, ϕ, η) to (p_x, p_y, p_z) . You may find it useful to draw a diagram!

Tip 5. For working out an expression for p_z you may want to remind yourself of the half-angle formula for $\tan \theta$.

Task 6.2. Derive an expression for $m_{\ell\ell}$ for the decay $Z \rightarrow \mu^+ \mu^-$.

Tip 6. You may want to remind yourself of the formula from your first year relativity lectures $m^2 = E^2 - p^2$. Remember that here E will be the total energy of the $\mu^+ \mu^-$ system and p is the total momentum of the system (i.e., the **magnitude** of the **vector sum** of the momenta of the μ^+ and μ^-).

Tip 7. Alternatively, if you are comfortable with 4-vectors you may want to use the formula that $m^2 = \tilde{p}^2$, where \tilde{p} is the sum of the 4-vectors of the μ^+ and μ^- .

Tip 8. You may neglect the mass of a muon in your final expression. **Question:** Why can we do this?

You should now attempt to either rewrite your formula for $m_{\ell\ell}$ in terms of p_T , η and ϕ using your conversion, or (probably simpler) define variables for p_x , p_y and p_z in your program and use these.

Tip 9. It is good practice to split your calculation into steps by defining appropriate variables within your code. This enhances readability and ease of use of your program. Make sure to add comments throughout your program as you go!

Task 6.3. Produce a histogram of $m_{\ell\ell}$ by running your code on the $Z \rightarrow \mu\mu$ MC data set. Look up the mass of the Z boson to verify your plot. Afterwards, run your code on the ATLAS 'Muons' data set and produce a second plot. Discuss with your lab partner whether or not, and if so why, you think these plots are different.

6.3 Making event selection cuts

To select the highest quality sample of signal events, particle physicists impose “selection criteria”, or “cuts” on each event. For the decay $Z \rightarrow \mu^+ \mu^-$, we are clearly only interested in events with **oppositely charged** leptons of the **same type** or “flavour” (that is electrons or muons). If you have not included these restrictions in your code already, you should do so — you may see a small change in your mass distribution, especially in the ATLAS data.

Something else we can use to restrict the events we plot is the p_T of the leptons. In general, we always want to use the particles with the highest p_T .

Task 6.4. You will find some events in which there are two or more leptons of the **same flavour** with the **same charge**. In each such case, select the lepton that has the highest p_T . Reproduce histograms of $m_{\ell\ell}$ by running your code on the $Z \rightarrow \mu\mu$ and the muon data sets. Compare your results with your previous analyses — again, you may see a small change in your mass distribution.

Task 6.5. Now, run your analysis code to find the invariant mass of $Z \rightarrow ee$ on the appropriate MC data set and on the ‘Electrons’ data set. Discuss with your lab partner other ways you think you could reduce the background.

Leptons produced in the decay of W and Z bosons tend to be “isolated” from other particles produced in proton-proton collisions. In contrast, leptons from “background” processes (such as the decay of b quarks) tend to be accompanied by a “jet” of other particles. The ntuple variable `ptcone30` contains a sum over the p_T of all tracks contained within a cone of half-width 0.3 in ΔR around the lepton direction. ΔR is given by

$$\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2},$$

where $\Delta\phi$ and $\Delta\eta$ are, respectively, the differences in azimuthal angle and pseudorapidity between the lepton and any additional tracks. The ntuple variable `etcone20` contains a sum over the p_T of all calorimeter clusters contained within a cone of half-width 0.2 in ΔR around the lepton direction. The above sums exclude the tracks and clusters associated with the lepton candidate itself.

Task 6.6. Plot graphs of the variables `ptcone30` and `etcone20` for the leptons that are members of the pairs of oppositely charged, same flavour leptons that you have selected above. Do this for each of the four data sets you have been using so far (‘Zee’, ‘Zmumu’, ‘Electrons’, ‘Muons’). These are representative of the “signal” leptons from Z decay you are targeting. Can you think of any way of looking at the distributions of the variables `ptcone30` and `etcone20` for leptons that are more likely to be “background” than signal? Use the resulting histograms to help you decide whether or not, and if so where, to apply selection cuts using these variables.

6.4 Event weights for Monte Carlo data sets

The two provided ATLAS data sets correspond to an “integrated luminosity”, $\int L dt$, of 1000 pb^{-1} or “inverse picobarns”. Equivalently, the integrated luminosity can be expressed as $\int L dt = 1 \text{ fb}^{-1}$ or “inverse femtobarns”. The integrated luminosity is a measure of the number of proton–proton collisions in ATLAS to which the data sets correspond. The provided Monte Carlo data sets correspond in general to higher integrated luminosities and thus have to be “scaled down” in order

to be compared with the ATLAS data. In addition, various corrections need to be applied to the Monte Carlo to account for deficiencies in the simulation of the performance of the detector. This is achieved by means of “event weights” that must be used whenever a histogram is filled or a counter for the number of selected events is incremented. The relevant variable in the code is called *weight*; it is calculated for you event by event in the function *FillHist* in *MyAnalysis.C*.

Tip 10. *In order to get useful additional information about the contents of your histograms in the “statistics box” that is usually displayed in the top right hand corner of the ROOT “Canvas” it is useful to execute the ROOT command `gStyle->SetOptStat(1111111)`.*

Tip 11. *Note that for histograms of the Monte Carlo data sets it is important to distinguish between the number of *Entries* and the value *Integral* displayed in the statistics box. Technically speaking, the number of *Entries* corresponds to the total number of times the ROOT function *Fill* has been called for the histogram in question from within your job. In contrast, the value *Integral* corresponds to the sum of weights for all of the calls of *Fill* for which the plotted value lies between the lower and upper bounds you have specified for the histogram in question. When comparing histograms of ATLAS data and Monte Carlo, it is value of the *Integral* that should be compared, not the number of *Entries*.*

6.5 Cross Sections, backgrounds and efficiencies

The “cross section” σ for a given process, such as $Z \rightarrow \ell\ell$, can be evaluated as follows:

$$\sigma(pp \rightarrow Z \rightarrow \ell\ell) = \frac{N^{\text{selected}} - N^{\text{background}}}{\epsilon \int L dt}, \quad (3)$$

where N^{selected} is the total number of events in the ATLAS data that pass your final selection cuts, $N^{\text{background}}$ is your estimate of the number of background events in the selected data sample. Therefore, $N^{\text{selected}} - N^{\text{background}}$ is your estimate of the number of “signal” events in the ATLAS data for the targeted physics process.

The quantity ϵ is the efficiency for selecting the “signal” events. The value of ϵ can be estimated using the Monte Carlo event sample corresponding to the targetted signal process. The sum of weights for all signal Monte Carlo events that pass your selection cuts needs to be divided by that corresponding to all generated events for the relevant sample. This information is provided in Table 5.

The fractional uncertainty on $\int L dt$ is 1.9%. The determination of the luminosity and its uncertainty in ATLAS at 8 TeV is described in the journal paper Eur. Phys. J. C (2016) 76: 653. It is usual to quote the uncertainty on the cross section arising from the luminosity separately from the uncertainties arising from other sources.

Other systematic uncertainties may arise, for example, from:

- backgrounds that you have not accounted for,
- disagreements between the simulation of the $Z \rightarrow \ell\ell$ signal in the Monte Carlo and the real ATLAS data.

A possible “catch-all” method to estimate the size of such effects is to re-calculate the cross sec-

	string code	sum of weights
$Z \rightarrow ee$	'Zee'	1241207
$Z \rightarrow \mu\mu$	'Zmumu'	1241207
$Z \rightarrow \tau\tau$	'Ztautau'	1240899
$W \rightarrow e\nu$ (b veto, no jets)	'WenuNoB'	12070498
$W \rightarrow e\nu$ (b veto, with jets)	'WenuNoBJets'	
$W \rightarrow e\nu$ (with b , with jets)	'WenuB'	
$W \rightarrow \mu\nu$ (b veto, no jets)	'WmunuNoB'	12072346
$W \rightarrow \mu\nu$ (b veto, with jets)	'WmunuNoBJets'	
$W \rightarrow \mu\nu$ (with b , with jets)	'WmunuB'	

Table 5: The sum of weights corresponding to all generated events for various Monte Carlo data sets. Use these as the denominator in the calculation of selection efficiencies.

tion having changed your event selection cuts. With regard to $N^{\text{background}}$ you might want to look back at Table 4; can you think of any possible sources of background in your selected $Z \rightarrow \ell\ell$ samples and how you might investigate these? Can you think of any other sources of statistical and systematic uncertainty on your measurements of $\sigma(pp \rightarrow Z \rightarrow ee)$ and $\sigma(pp \rightarrow Z \rightarrow \mu\mu)$?

Present your measurements of the cross section for Z boson production $Z \rightarrow \ell\ell$ in the form:

$$\sigma(pp \rightarrow Z \rightarrow \ell\ell) = ??? \pm XXX(\text{stat.}) \pm YYY(\text{syst.}) \pm ZZZ(\text{lumi.}), \quad (4)$$

where XXX is the statistical uncertainty, YYY is the systematic uncertainty, and ZZZ is due to the luminosity uncertainty. Please make sure you keep a good record in your lab. book of your calculations of cross sections and the various uncertainty components! What kind of questions should you be asking yourself about your measured cross sections and uncertainties?

6.6 Analysing W bosons

W bosons are somewhat more difficult to select and analyse than Z bosons, as they decay to produce a single charged lepton and a neutrino, $W \rightarrow \ell\nu$. Neutrinos do not interact in the detector and lead to an apparent imbalance in the vector sum of the p_T of all the observed particles in an event — this is called the “missing momentum” or “**missing energy**”. The magnitude (E_T^{miss}) and azimuthal angle (ϕ^{miss}) of the missing energy are interpreted as the transverse components of the neutrino momentum⁴. Because very energetic particles may escape detection if they are produced very close to the beam direction, only the transverse components of the neutrino momentum can be measured in ATLAS. Because of this limitation we are unable to calculate the total invariant mass of the decay products of W bosons. Instead, we use a quantity called the “**transverse mass**”, m_T , which uses only the transverse components of the lepton and neutrino momenta. It is defined as follows:

$$m_T = \sqrt{2E_T^{\text{miss}}p_T^l(1 - \cos(\Delta\phi))}, \quad (5)$$

where $\Delta\phi = \phi^{\text{miss}} - \phi^l$, and ϕ^l and p_T^l are the ϕ and p_T of the lepton.

⁴On the ntuples these variables are called `met_et` and `met_phi` respectively.

Tip 12. When calculating $\Delta\phi$, one should be careful to ensure that the smallest angle between the two ϕ values is used. It should also be noted that for all the particles provided on the input data sets $-\pi \leq \phi \leq \pi$. As a cross check of your calculation of m_T you might want to start from your formula for the invariant mass from subsection 6.2 and eliminate the z components (including replacing the total momenta with the corresponding transverse momenta).

Task 6.7. Produce a plot of m_T for the $W \rightarrow \mu\nu$ Monte Carlo data set ‘WmunuNoB’. Compare with the peak in $m_{\ell\ell}$ that you found for Z bosons. How do you interpret the differences in shape between the two distributions?

Task 6.8. Now produce a plot of m_T for the ATLAS ‘Muons’ data set. Compare this with the results from the $W \rightarrow \mu\nu$ Monte Carlo. How do you interpret the differences in shape between the two distributions?

The biggest challenges in analysing W decays are:

- to reduce the level of background from other processes in the selected event sample
- to evaluate the residual numbers of background events in the selected event sample, together with an estimate of the uncertainty on that evaluation.

Both of these challenges can be addressed by the approach of plotting the distribution of each of the variables that most effectively discriminate between signal and background **without making a cut on the variable in question, but having applied all other selection cuts**. For each of these discriminating variables can you understand the distribution you see in the ATLAS data as the sum of contributions from the $W \rightarrow \ell\nu$ signal and those from the various backgrounds?

A technical note: an irritating feature of the ATLAS OpenData for the $W \rightarrow \ell\nu$ Monte Carlo is that there are three data sets. (For example, for decays to electrons these are ‘WenuNoB’, ‘WenuNoBJets’, and ‘WenuB’). In order to calculate the overall event selection efficiency the numbers selected from these three data sets need to be added.

6.7 Some thoughts on extensions to your analysis

This is a very open-ended project. Whilst the main aim is to measure the cross sections for Z and W boson decays to leptons, if you have time the provided ATLAS data and Monte Carlo allow many other interesting investigations of the production of Z and W bosons to be made. In addition, you get credit for spotting interesting/unexpected features in the data and following up on these; feel free to discuss your ideas and observations with your demonstrator! It may also be possible to select and study in the ATLAS data candidate event samples corresponding to some of the other physics processes for which Monte Carlo samples are listed in Table 4. For example, you might consider events containing tau leptons, pairs of vector bosons, top quarks, or potential sources of new physics from “beyond the standard model”.

6.8 Your feedback

We welcome your feedback/constructive criticism on the experiment, software, and documentation. Please feel free to talk to your demonstrator and/or send email to Terry Wyatt at twy-

7 Appendices

7.1 Collected hyperlinks

- The [ROOT users' guide](#).
- To learn more about Feynman diagrams see, e.g., [this resource](#).
- Information about the ATLAS collaboration, the detector, the physics etc. can be found [here](#).
- A journal paper describing the measurements of Z and W production in the very early ATLAS data at the LHC can be found [here](#).
- For details on the data sets you are using, you can consult the [OpenData online software book](#) for reference.
- For a refresher on C++, please see the resources available on, e.g., [cplusplus.com](#).
- The definition of all the available variables on the ATLAS OpenData data sets for this experiment can be found [in this web-page](#).
- For information on booking, filling, and formatting histograms, see the relevant sections in the [online ROOT documentation](#).

The determination of the luminosity and its uncertainty in ATLAS at 8 TeV is described in the journal paper Eur. Phys. J. C (2016) 76: 653.

7.2 Remote access to the computers in the 3rd year lab.

If you wish to use the computers in the 3rd year lab. outside normal lab. days you are welcome to do so, if the 5th lab. is open. If the 5th floor lab. is not open it is possible to access the computers remotely from any University of Manchester computer cluster. The procedure to follow is given here.

1. Pick any unused machine in a University of Manchester cluster.
2. Switch the machine OFF and then back ON again using the power button on the box.
3. As the machine boots up you will be presented with a list of possible boot options. Select: *"EPS Linux (Scientific Linux 7.3) [Home directory = P drive]"* and hit **RETURN**.
4. Once the machine has finished booting into Scientific Linux log in with **your personal University of Manchester account**. (N.B. NOT the "atlaslab" account at this stage.)
5. In the pull-down menu "Applications" select "Utility" and then open a "Terminal" application.

6. In the Terminal window type

```
ssh -X -Y atlaslab $\mu$ @heplabpc $\mu$ .hep.manchester.ac.uk,
```

where the two occurrences of “ μ ” should be replaced by the number of the atlaslab account you were allocated at the beginning of the lab.

7. You should now be able to work just as if you were physically sitting in front of the machine in the 5th floor lab.

8. At the end of your session type **exit** to log out.

9. This is a new feature. Please let us (twyatt@fnal.gov and sabah.salih@manchester.ac.uk) know if you try this and whether or not it works!

If you have access to your own laptop/desktop computer running a Unix/Linux operating system you should also be able to access the computers in the 3rd Year laboratory remotely with the command:

```
ssh -X -Y atlaslab $\mu$ @heplabpc $\mu$ .hep.manchester.ac.uk.
```

N.B. You are expected to be physically present in the lab. from 9:00 to 17:00 on lab. days and remote access is not a substitute!

7.3 A few hints on archiving your work and keeping a logbook

It can be tempting to think that because you are making lots, and lots, and lots, of plots these can replace the need to keep a logbook. Please do not fall into this trap! You need to keep a systematic record of the investigations you have carried out, the calculations you have performed and, most important of all, a summary of conclusions you have reached. I can guarantee that part of the question and answer session following your talk will require you to look up in your logbook additional information that is not contained in your talk. Students who can quickly and efficiently find the required information and answer questions in a clear and concise fashion gain extra marks, and conversely . . . ;-)

Particularly in regard to the conclusions, I suggest sitting down at the end of each lab. day and carefully reviewing your logbook, writing down any details you forgot to record, and making sure you have written down appropriate conclusions. I strongly suggest that as part of this daily review you write down what you plan to do and what you hope to achieve during the next lab. day! Five minutes taken at the end of one lab. day can enormously improve your productivity during the next lab. day!

You should feel free to keep either a paper or an electronic logbook. The choice is yours, but . . .

- If you keep a paper logbook then make sure you print out and stick into your logbook copies of the absolutely vital plots that you have archived in the sub-directory *OutputPlots*.
- If you keep an electronic logbook then make sure you do this on a laptop that you **must** bring to the interview. Make sure that you can add plots, details of calculations, etc. to your electronic logbook, and that under the time pressure of the interview you are able to find information as quickly and efficiently as if you'd kept a paper logbook. If you have any

doubts on this score then the safer option may be to keep a paper logbook.

7.4 A few hints for preparing your talk

You should be aiming to deliver your talk in ten minutes. This is a quite a small amount of time: spend most of it telling us about **your** analysis, results and conclusions. Be specific: show us actual plots, numbers, uncertainties, and not just generalities. Describe the cuts you have employed and show plots to justify why you chose those cuts.

Whilst giving a small amount of “background” and “introductory” information is fine, do not waste large amounts of time just telling us what we already know from this lab. script!

Please add page numbers to your talk! It helps us ask questions about specific items in your talk.

Please add your names to the title page of your talk!

Consider adding some “Backup” slides to your talk, that contain additional plots, details of your calculations, etc., that might be useful in answering our questions.