From Raw data to Physics Results (3/3)

Dr Paul Laycock
The particle physics cycle
Course outline

• Lecture 1
  • The journey of raw data from the detector to a publication

• Lecture 2
  • How we reconstruct fundamental physics processes from raw detector data

• Lecture 3
  • How we extract our signals from the mountain of data, finding needles in the haystack
Standard Model Total Production Cross Section Measurements

**ATLAS** Preliminary

Run 1,2 $\sqrt{s} = 7,8,13$ TeV
Measuring cross sections

\[ \sigma = \frac{N}{L} \]

• The cross section for a process is defined as the number of events divided by luminosity
Measuring cross sections

$$\sigma = \frac{N}{L_{\text{int}}}$$

• The cross section for a process is defined as the number of events divided by the integrated luminosity, $L_{\text{int}}$, which measures how much data we have collected.
ATLAS Luminosity

• Question: Why does ATLAS record less data than the LHC delivers?

• How do we know the integrated luminosity delivered?
• The LHC accelerates **bunches of $10^{11}$ protons** separated by 25ns gaps
Measuring Luminosity at the LHC

• Ingredients for a measurement of the luminosity

  • Measuring the size of the beams (for a certain LHC configuration)
    • This requires a dedicated measurement where we scan the beams across each other in the horizontal and vertical directions - a van der Meer scan

  • Measuring the beam currents in each bunch
    • This is done during collisions, integrating all of the bunch currents and knowing their size, we can calculate the luminosity

• Make many cross checks because this is such a crucial measurement
Measuring cross sections

\[ \sigma = \frac{N}{L_{\text{int}}} \]

- The cross section for a process is defined as the number of events divided by the integrated luminosity, \( L_{\text{int}} \), which measures how much data we have collected.

\[ \sigma = \frac{N_{\text{obs}}}{A \cdot \epsilon \cdot L_{\text{int}}} \]

- \( N_{\text{obs}} \) in data needs to be corrected for the detector acceptance, \( A \), for selecting those events. The reconstruction efficiency, \( \epsilon \), is a product of all of the efficiencies that we need to measure and ensure that they are the same in our data and simulation.

Did I mention that simulation is important?
Before the detector, came the simulation

- When designing detectors, we *simulate detector response* to physics of interest
- Adding a *solenoid magnet* makes it possible to measure momentum (and charge) in our tracker by measuring curvature in the transverse plane
- Interesting physics is often at *high momentum*, e.g. four high momentum muon tracks here

No magnet
Before the detector, came the simulation

- When designing detectors, we **simulate detector response** to physics of interest
- Adding a **solenoid magnet** makes it possible to measure momentum (and charge) in our tracker by measuring curvature in the transverse plane
- Interesting physics is often at **high momentum**, e.g. four high momentum muon tracks here
Before the detector, came the simulation

• When designing detectors, we **simulate detector response** to physics of interest

• Adding a **solenoid magnet** makes it possible to measure momentum (and charge) in our tracker by measuring curvature in the transverse plane

• Interesting physics is often at **high momentum**, e.g. four high momentum muon tracks here

What detector technology might this example motivate?
Simulation and understanding detectors

- We use simulations to model the detector as accurately and precisely as possible.
- We then test that our simulations are accurate using real data.
- We correct our simulations if necessary.
- Once our simulation is an accurate model of our detector, we can use it to correct the data for detector response.

**Inner Detector Material**

- Detailed mapping of material in Pixel detector using
  - Hadronic interactions
  - Photon conversions
- Simulation updated after comparison to improve geometry description.

---

**Graphs**

- **ATLAS Preliminary**
  - Data 2015
  - Pythia8 Simulation (Updated Geometry)
  - Pythia8 Simulation (Default Geometry)

- **Vertex X [mm]**
  - **Vertex Y [mm]**

- **Graphs**
  - **Vertex X [mm]**
  - **Vertex Y [mm]**
  - **Vertices / mm² / Event**
  - **Hadronic Interaction Radius [mm]**

---

**December 2, 2015**

A. Polini, 124th LHCC Open Session
Exabyte-scale physics analysis

Exabytes of Data

Exabytes of Simulation

Publish!
Ingredients to the ATLAS physics program

- We compare data with simulation
Ingredients to the ATLAS physics program

- We make a LOT of comparisons of data and simulation
Ingredients to the ATLAS physics program

Compare data and simulation in every conceivable way!

We need to check that the simulation does a good job of describing the data

That’s more important than the underlying physics model, in fact it should be independent

WHY?
Measuring cross sections

\[ \sigma = \frac{N}{L_{int}} \]

- The cross section for a process is defined as the number of events divided by the integrated luminosity, \( L_{int} \), which measures how much data we have collected.

\[ \sigma = \frac{N_{obs}}{A \cdot \epsilon \cdot L_{int}} \]

- \( N_{obs} \) in data needs to be corrected for the detector acceptance, \( A \), for selecting those events. The reconstruction efficiency, \( \epsilon \), is a product of all of the efficiencies that we need to measure and ensure that they are the same in our data and simulation.

\[ \sigma = \frac{N_{obs} - N_{bkg}}{A \cdot \epsilon \cdot L_{int}} \]

- Finally, we need to measure and subtract background events that are not part of our signal process.
Discovering the Higgs Boson: \( H \rightarrow ZZ \rightarrow 4l \)

- We will (nearly) always have some irreducible background to the signal process that we are trying to measure
Measuring cross sections

\[ \sigma = \frac{N}{L_{int}} \]

- The cross section for a process is defined as the number of events divided by the integrated luminosity, \( L_{int} \), which measures how much data we have collected.

\[ \sigma = \frac{N_{obs}}{A \cdot \epsilon \cdot L_{int}} \]

- \( N_{obs} \) in data needs to be corrected for the detector acceptance, \( A \), for selecting those events. The reconstruction efficiency, \( \epsilon \), is a product of all of the efficiencies that we need to measure and ensure that they are the same in our data and simulation.

\[ \sigma = \frac{N_{obs} - N_{bkg}}{A \cdot \epsilon \cdot L_{int}} \]

- Finally, we need to measure and subtract background events that are not part of our signal process.

Now we can compare this to the theoretical cross section!
Physics model builders
Physics event generators

- There are lots of different physics models implemented in physics event generators, depending on the type of physics that you’re interested in.
- We want to see if reality looks like theory (and which one!)

(Courtesy: Z. Marshall)
Are we ready to do some exabyte-scale physics analysis?

Exabytes of Data

Exabytes of Simulation

Publish!
First - measuring the Z boson

**Z->ee in UA1**

Two EM clusters with $E_T > 25\text{GeV}$.

As above plus a track with $p_T > 7\text{GeV}$ pointing to the cluster. Hadronic and track isolation requirements applied.

A second cluster has also an isolated track.
Measuring the Z boson at ATLAS

\[ \sigma = \frac{N_{\text{obs}} - N_{\text{bkg}}}{A \cdot \epsilon \cdot L_{\text{int}}} \]

- Select events with (here) two muons

- Question: what other selections can we apply to the muons?

- Here I have only considered events with two muons

- Question: is this the cross section for Z boson production?
Measuring the Z boson at ATLAS

\[ \sigma = \frac{N_{\text{obs}} - N_{\text{bkg}}}{A \cdot \epsilon \cdot L_{\text{int}}} \]

• Backgrounds are small but still need to be measured and subtracted

• **We will quote a fiducial cross section corresponding to good detector acceptance**

• After making the event selection, applying the same selection to all of the simulations of background processes, and measuring my acceptance and efficiencies (and knowing the luminosity) - *am I done?*
Measuring the Z boson at ATLAS

\[ \sigma = \frac{N_{obs} - N_{bkg}}{A \cdot \epsilon \cdot L_{int}} \]

• No! You would like to publish with the smallest uncertainties possible

• Every ingredient to the analysis comes with an uncertainty

• \(N_{obs}\) has a statistical uncertainty

• \(N_{bkg}\) is typically composed of several sources (different physics processes) with corresponding statistical and systematic contributions to the final uncertainty

• \(A\) and particularly \(\epsilon\) have many systematic components stemming from each reconstruction algorithm that we used

• Finally, \(L_{int}\) also has an uncertainty that dictates how well we know the absolute scale of the measurement - a normalisation uncertainty
Standard Model Total Production Cross Section Measurements

ATLAS Preliminary
Run 1,2 $\sqrt{s} = 7,8,13$ TeV

LHC pp $\sqrt{s} = 7$ TeV
- Data $4.5 - 4.6$ fb$^{-1}$

LHC pp $\sqrt{s} = 8$ TeV
- Data $20.2 - 20.3$ fb$^{-1}$

LHC pp $\sqrt{s} = 13$ TeV
- Data $3.2 - 79.8$ fb$^{-1}$
Elements of a search
Elements of a search

Like $Z\rightarrow\text{ee}$ but at higher mass.

\[ \int L \, dt = 167 \, \text{pb}^{-1} \]
\[ \sqrt{s} = 7 \, \text{TeV} \]
Elements of a search

LIKE Z->ee but at higher mass.

ATLAS Preliminary

\[ \int L \, dt = 167 \, \text{pb}^{-1} \]
\[ \sqrt{s} = 7 \, \text{TeV} \]
Elements of a search

Like $Z \rightarrow ee$ but at higher mass.

Select 2 electron candidates and plot their invariant mass for:

1. Data

$\int L \, dt = 167 \, \text{pb}^{-1}$

$\sqrt{s} = 7 \, \text{TeV}$
Elements of a search

Like $Z \rightarrow ee$ but at higher mass.

Select 2 electron candidates and plot their invariant mass for:
1. Data
2. Simulated background events
Elements of a search

Like $Z\rightarrow ee$ but at higher mass.

Select 2 electron candidates and plot their invariant mass for:
1. Data
2. Simulated background events
3. Simulated signal with different masses
Elements of a search

Like $Z\rightarrow ee$ but at higher mass.

ATLAS Preliminary

\[ \int L \, dt = 167 \text{ pb}^{-1} \]
\[ \sqrt{s} = 7 \text{ TeV} \]

Select 2 electron candidates and plot their invariant mass for:
1. Data
2. Simulated background events
3. Simulated signal with different masses

Data inconsistent with a 1TeV $Z'$
Elements of a search

Like $Z\rightarrow ee$ but at higher mass.

Select 2 electron candidates and plot their invariant mass for:
1. Data
2. Simulated background events
3. Simulated signal with different masses

Cross-section decreases with mass (higher the mass of the $Z'$, the more data needed to discover it)
Elements of a search

© And similar for muons

Select 2 muon candidates and plot their invariant mass for:
1. Data
2. Simulated background events
3. Simulated signal with different masses

Data inconsistent with a 1TeV Z’
Elements of a search

Why is the resolution worse in the muon channel?

Differences in:

- Resolution
- Background composition
- Dataset
Data analysis
Needles in haystacks

• We record billions of events

• The data are structured but each event is different - unique data science challenge

• Data reduction proceeds via a two-pronged approach:

  • Select only the events that you are interested in
    • e.g. events with two photons

  • Keep only the information you need
    • Throw away the rest!

• Final statistical inference is only performed on the reduced data
There are two dimensions to our data challenge, one is the (billions) of individual events, the other is the properties of each event.

Simplified picture, in reality the event properties depend on the event content found by event reconstruction.
Slice and dice - data reduction

Event Properties

\[ \begin{array}{cccccc}
  x_1 & x_2 & x_3 & x_4 & \cdots & x_i \\
\end{array} \]

\[ N_{\text{event}} \]

- We can reduce data by selecting only our interesting events
Slice and dice - data reduction

Event Properties

\[ N_{\text{event}} \]

- And we can reduce data by selecting only the properties needed for our analysis
### Slice and dice - data reduction

- **Event Properties**

<table>
<thead>
<tr>
<th>Event Properties</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x_1 )</td>
</tr>
</tbody>
</table>

- Data reduction usually aims for factors of 100 or more (more than shown here!)
Ingredients to the ATLAS physics program

• We make lots of reduced samples of both data and simulation, which all need to be replicated around the world - a computing challenge!
The best computing model

- How to most efficiently do this across the whole physics program making the best use of computing resources and the best use of people’s time is an important question
Now you know how to do exabyte-scale physics analysis!
Now it’s over to you!

- Our future computing needs outstrip our computing resources
  - and computing gets more heterogeneous and complicated
  - and we want to be as environmentally-responsible as possible
- So you have work to do - good luck and have fun!
Contact details

- I am usually based at Geneva Observatory in Versoix, but will be here at CERN Wednesday 28th through Friday 30th June.
  - I will be available for Q&A every afternoon from 3pm-4pm in restaurant 1, feel free to send questions to my email

- email: paul.laycock@unige.ch