



ATLAS Data Preservation and Access

Roger Jones





Data Preservation & Access

- ◆ Opening data access
 - ◆ Preparatory discussions with “management”, CB chair, authorship and Pubcom chairs
 - ◆ Has clear implications for authorship/membership rules
 - ◆ Needs CB-level discussion
 - ◆ Past experience says these topics provoke long discussion in the CB!
 - ◆ Common principles proposed by LHC experiment Data Policy Harmonization Group straw man
 - ◆ This has been reviewed by the SIPB and taken to CERN Council to become a “policy suggestion”
- ◆ A draft policy is with the management for discussion & has been seen by the ICB



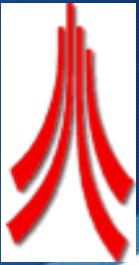
ATLAS DMP Organization



- ◆ Data Preservation now included as part of the upgrade activity planning
 - ◆ May increase the funding options – some evidence already
 - ◆ Data Management Planning is now required by some funders for upgrade grants
 - ◆ Looking at the cost/benefit of various strategies
 - ◆ Resource tensioning with other upgrade activities



Principles for preservation & access



- ◆ General agreement RAW data is preserved for the experiment and future – open data access is not usually possible even to the collaboration members (level 4 data) and is not proposed for general use
- ◆ Full reconstruction outputs for analysis might be made available after an embargo period – tbd, but clearly embargo of several years. The resource implications to make this useful are high. (Level 3 data)
- ◆ We support limited access of samples in simple formats for outreach and teaching (level 2 data) – but these are best integrated to our presenter tools
- ◆ Techniques like Recast may make data (information) usefully available, although it does not meet all the open access criteria for levels 2 & 3
- ◆ We already make data from papers and supporting information available through HEPDAT/Inspire (Level 1 data)



Data Preservation Policies

◆ Data Preservation

- ◆ There are DP policies implied in the Computing TDRs
 - ◆ conserve all raw data during the lifetime of the experiment
 - ◆ All formats & code used for paper analyses to be archived
 - ◆ Tier 0/1s responsible for the physical preservation
- ◆ Some tacit belief that older sets may be 'retired'
 - ◆ Retired data no longer to be on disk or under active analysis
 - ◆ This may need to be revised e.g. if external access is then granted
 - ◆ Obvious resource implications
- ◆ First priority to preserve data for active use by the collaboration



ATLAS DP Practical Steps

- ◆ Making sure raw data can be reprocessed long-term (Level 4)
 - ◆ Identifying key datasets for ‘unique data’ preservation
 - ◆ Setting up regular reprocessing and validation
 - ◆ This has been underway as a test case for the 2009 data, but progress is slow
 - ◆ Forward/backward compatibility issues illustrated in John Chapman’s talk on simulation release plans 14/3/13
- ◆ Ensure the capability to run old trigger selections offline
- ◆ AODfixing will help (reprocessing at analysis format level)
 - ◆ This means level 4 operations can be applied to level 3 AOD format



Digesting validation results

- ◆ Must display the results of the validation in a comprehensible way: web based interface
- ◆ The test must determine the nature of the results
 - ◆ Could be simple yes/no, plots, ROOT files, text-files with keywords or length, ...
- ◆ Need for semi-automated, detailed physics validation
- ◆ David South is on ATLAS and was central to the DESY SP and DPHEP activities
 - ◆ Identify the useful common components
 - ◆ Identify the ATLAS-specific elements
 - ◆ Set up CERN-based instance for ATLAS (and others?)



Existing open datasets

- ◆ The CB has authorized various datasets in (level-2) outreach formats for open use in education/outreach
 - ◆ Event displays for interactive analysis (MINERVA/HYPATIA/LPPP/CAMELIA)
 - ◆ JIVE-XML, root format data
 - ◆ Absolutely not intended for any serious analysis, but illustrative



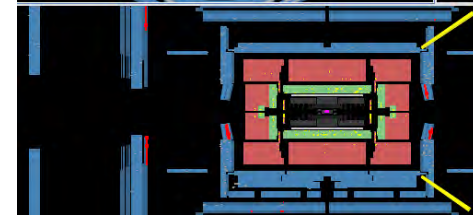
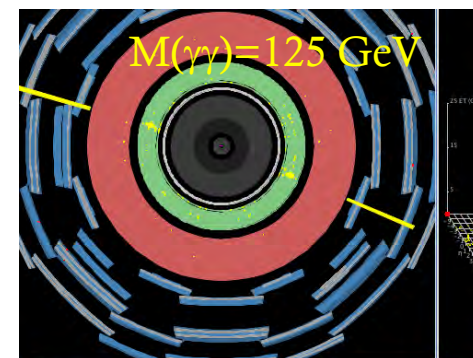
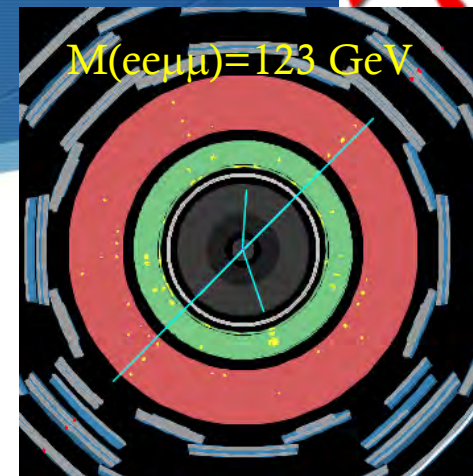
ATLAS Zpath



Master the invariant mass technique

- to study and measure the (Z, J/ψ , Y) decaying to l^+l^-
- to search for new physics (Z')
- And Higgs boson in $\gamma\gamma$ and $l^+l^-l^+l^-$

- HYPATIA using the ATLANTIS event display
- Data from 2011
 - 13000 events ~ 2.5 GB (password protected, 100 open)
 - 13 data groups/directories, 20 subgroups (A-T), and 50 events/mixed sample/2 students
 - 50% Z, 30% $\gamma\gamma$, 10% (J/ψ , Y), 5% Z' , 5% $l^+l^-l^+l^-$
 - Higgs candidate events:
 - 1 fb^{-1} and cuts according to ATLAS publication
 - 125 GeV Higgs MC signals ready to upload (1 fb^{-1} , 10 fb^{-1} , 25 fb^{-1})



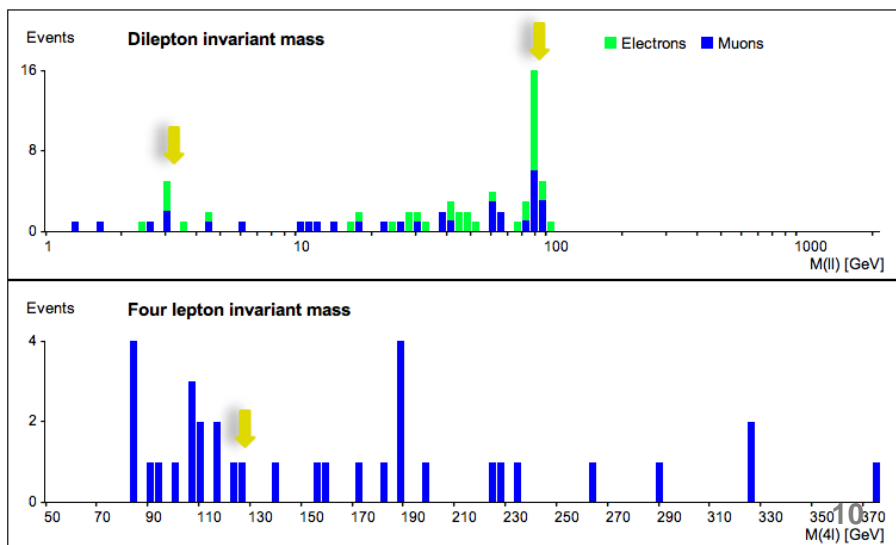
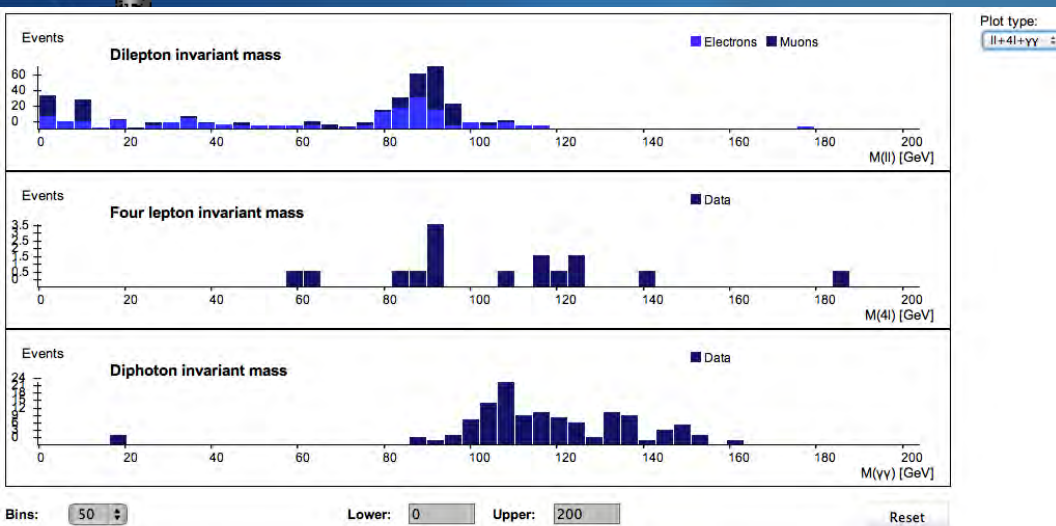


ATLAS Zpath tests

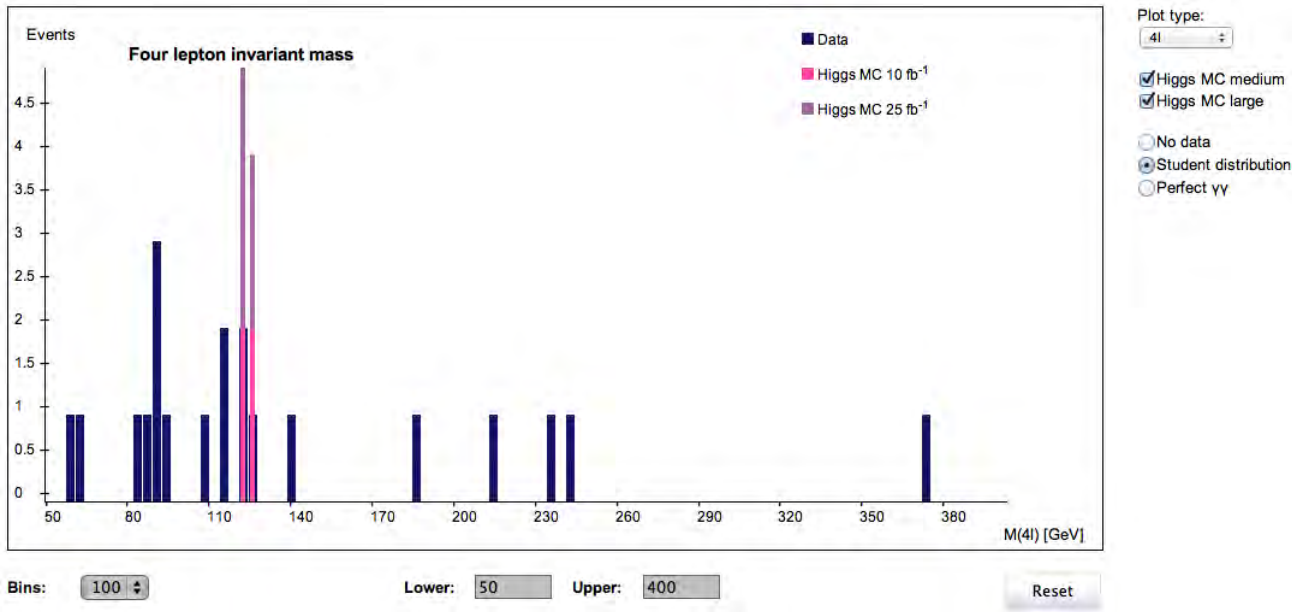
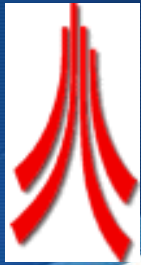


OPloT:

- M_{ll} and/or $M_{\gamma\gamma}$ and/or $M_{ll\gamma\gamma}$ to be discussed locally
- Moderator: 1 slide with 3 invariant masses; Invariant mass as a tool to identify particles, to discover new particles, and to search for exotic particles
- Web pages updated and measurement ready
- <http://www.physicsmasterclasses.org/exercises/ATLAS-2013/en/zpath.htm>
- Introduced Higgs
- Described new measurements
- Prepared material for instructors, moderators, for discussions, ...

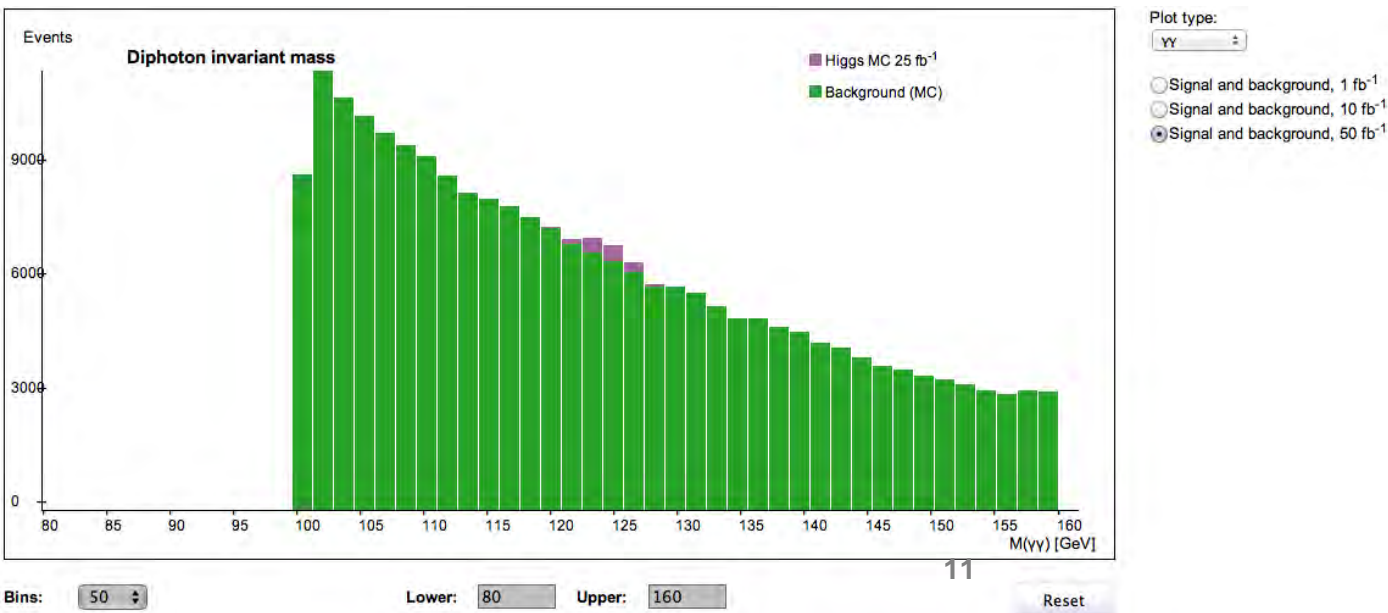


OPloT Tests 2013



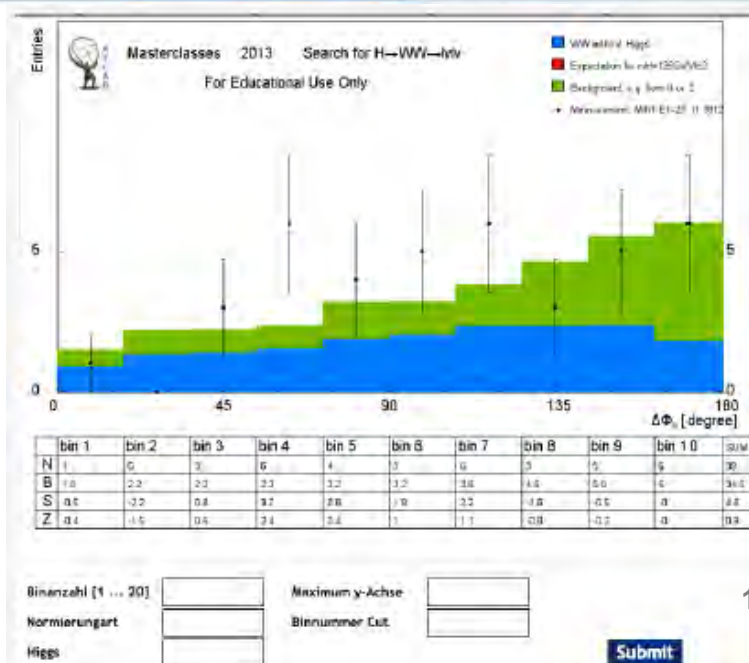
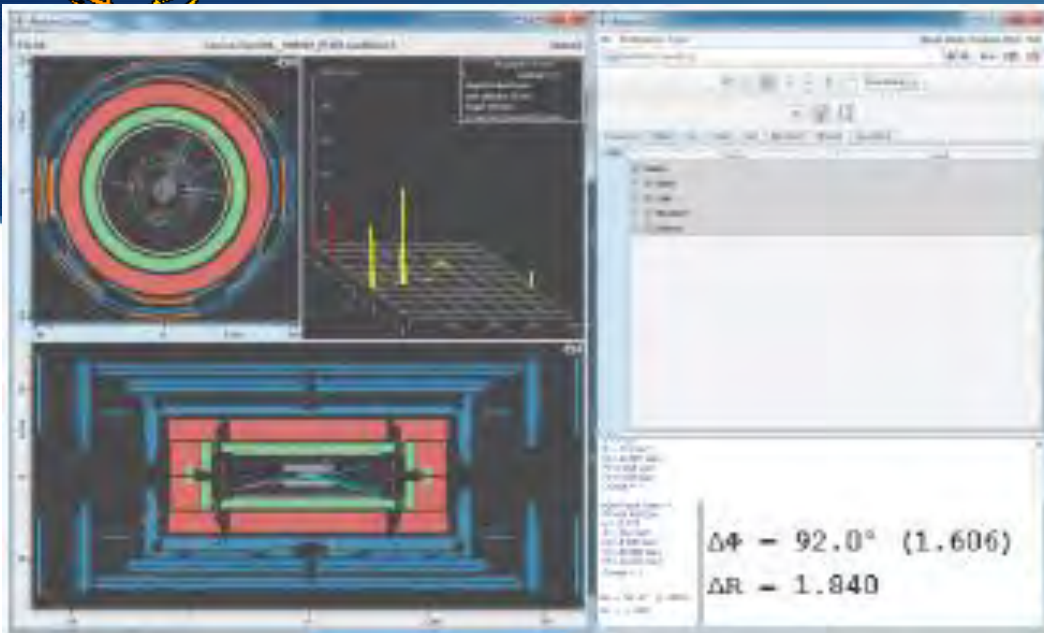
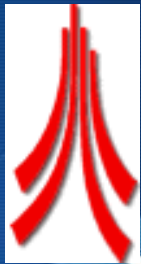
Higgs comments

- 4l provided without requiring 2l from Z, with lower cut on other pair
- $\gamma\gamma$ provide MC with 125 Higgs and background
- Upload 125 Higgs MC ((1)&10 & 25 fb^{-1})





ATLAS W-path with real WW (+H) events



Measurements

- W → lν
- W+/W- ratio
- Angular distribution between leptons in WW events

MINERVA program using the ATLANTIS event display

- 2011 real data: 693 WW/Higgs candidates (from released 1fb⁻¹) mixed with 5307 W and other background events

Histogram tool

- spreadsheet and histogram websites connected with database

New measurement tested