A New Tool For Measuring Detector Performance in ATLAS

- Arno Straessner – TU Dresden
  Matthias Schott – CERN

on behalf of the ATLAS Collaboration

Computing in High Energy and Nuclear Physics
Prague
March 21-27, 2009
Outline

• The Physics: Detector Performance
• An Infrastructure For Performance Data
• Current Implementation in ATLAS
• Outlook
The Physics: Detector Performance

- Complex detector with main subsystems:
  - Inner tracking
  - Calorimetry
  - Muon tracking

- 2500 physicists perform analyses

- All need detailed information about detector performance
  - General performance
    - Trigger, reconstruction and identification efficiencies
    - Resolutions of energy, momenta, angles, …
    - Energy and momentum scales
  - Time-dependent performance

detector response
Detector Performance Tool

• Commonly applied method to measure performance in data: → Tag & Probe

• Example: measure muon trigger efficiency in Z→µµ events
  • Identify triggered and well measured muon: → Tag
  • Use Z decay kinematics to find the 2\textsuperscript{nd} muon: → Probe
  • Check if 2\textsuperscript{nd} muon was triggered → efficiency

• Many more examples:
  • Z→ ee, J/psi → ee, Z+jets for tau fake rates, …

• Common infrastructure to
  • Implement the object tag and event selection
  • Store the collection of probe objects
  • Analyse the probes → efficiencies and detector response
  • Store matrices with efficiency, resolution and scale information

Clients:
• Performance groups
• Fast Monte Carlo simulation
• Individual Physicists
Application: Performance Groups

• Early version of the package was used in ATLAS physics book

• Estimation of muon identification and trigger efficiencies:

• Performance Tool is the ATLAS solution for:
  • Direct use in Standard Model physics $\rightarrow$ closely related to performance groups
  • Benchmark comparisons for reconstructed objects used in searches etc.
The Procedure

- ATLAS data storage
  - Event Summary Data (ESD)
  - Analysis Object Data (AOD)
  - Derived Physics Data (DPD)

- 1st step:
  - Signal selection and object tag
  - Input: ESD, AOD, DPD
  - Grid task

  → Performance-DPD with probe objects

- 2nd step:
  - Performance determination
  - Executed on local cluster
  - Overall fits might be necessary (background shape, …)

  → Performance database
Probe Object Data

- Probe objects are usually: tracks, reconstructed leptons, jets

- Only parameters necessary for further analysis are stored → can be freely defined

- Example for electron calorimeter identification efficiency:

  Electron reconstruction category
  Electron quality cut
  Electron trigger information
  Reconstructed Z mass from tag+probe electron

  Z boson $p_T$
  Angular difference ($\Delta R$) to next jet
  Jet multiplicity
  Sum of jet $p_T$
  Max of jet $p_T$
  Electron isolation variables

  Matching angle to generator truth electron

  To calculate number of trials and successes for efficiency
  Interesting variables for differential efficiencies
  Optional for Monte Carlo verification of the method

- Map<tag,float> with string tags in meta-data → flexible, user-friendly structure
- Small data size: $\sim$ 0.33 kB per event in DPD file
Performance Database

- Objects stored:
  - N-dimensional matrices to map detector areas and physics
  - Typically 4-dimensional, not larger because of:
    - data statistics per matrix entry
    - storage space
  - Matrix defined by N axes objects with free binning

- Matrix entries:
  - For “simple” efficiency calculations:
    - Number of trial and success counts
  - With side-band subtraction of background:
    - Data and background histograms
    - Pre-defined fitting functions
    - Background subtraction assuming predicted S/B ratio
Performance Database

- Matrix entries for detector response:
  - Resolution/scale histogram for Monte Carlo → data reference
  - Smearing functions for generator level → fully simulated Monte Carlo

\[
\begin{array}{|c|c|c|c|c|c|}
\hline
\eta & \ddots & & & & \\
& & \ddots & & & \\
& & & \ddots & & \\
& & & & \ddots & \\
& & & & & \\
\hline
\end{array}
\]

- Once ATLAS data is available:
  - Smearing functions for generator level → data projections
  - To be used in fast detector simulations

- Matrix entries for resolution and scaling:
  - ROOT Histograms → in future: RooWorkspace of RooFit
Matrix Objects

- Methods to calculate efficiency and resolution from stored data
  - Classical and Bayesian efficiencies and uncertainties
  - Efficiencies and detector response using fits to data and background
  - Caching: fit result is stored and then directly accessed

- All matrix entry objects are additive:
  - Distributed analysis
  - Averaging over different run periods

- Projections to any number of axes is supported:
  - Projections in form of matrices
  - Projections into histograms
  - Slices of matrices → cuts
  - Not trivial for resolutions

- Hide complexity from user
Implementation in ATHENA

- ATHENA = ATLAS software framework

- “Algorithm” is executed at each event
- Uses “Tool” to perform signal selection

- Intermediate data is stored in
  - DPD
  - Transient memory “Storegate”

- Performance determination

- Output to ATLAS official conditions database:
  - LCG product: COOL database
  - With Interval-Of-Validity (IOV)
Supported Database Structures

- ATLAS official conditions database: LCG Pool/COOL → for collaborative use
- Plain ROOT file → for online Tier-0 applications, development, debugging and individual use
- Unique database identifier:

<table>
<thead>
<tr>
<th>Database ID</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Object</td>
<td>Physics Parameter</td>
<td>Muon, Electron, Tau, Jet, …</td>
</tr>
<tr>
<td>Type</td>
<td>Performance Parameter</td>
<td>Efficiency, Scale/Resolution, Fake-Rate</td>
</tr>
<tr>
<td>Channel</td>
<td>Physics Channel</td>
<td>Z→μμ, J/ψ→ee, tt, …</td>
</tr>
<tr>
<td>Author</td>
<td>Author’s name</td>
<td>MuonPerformanceGroup, PJenni, …</td>
</tr>
<tr>
<td>RecoSWV</td>
<td>Software version (ATLAS Metadata Interface tag)</td>
<td>14.5.2.1</td>
</tr>
<tr>
<td>IOV</td>
<td>Interval of validity (runs), For MC: simulation software release</td>
<td>Run 1000-2000 13.0.1</td>
</tr>
</tbody>
</table>
Database Object Structure

Matrix

Payload:
- Vectors of integers and floats
- Data vectors
  - Matrix entries
  - Info vectors
  - Matrix dimensions and axes definitions

Unique ID
Database ID

COOL – Database

<table>
<thead>
<tr>
<th>DATA</th>
<th>IOV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data [float, float, int, int, float, ...]</td>
<td>10–23</td>
</tr>
<tr>
<td>Data [float, int, int]</td>
<td>12–35</td>
</tr>
<tr>
<td>...</td>
<td></td>
</tr>
<tr>
<td>Data [Reference to ROOT-File]</td>
<td>11–18</td>
</tr>
<tr>
<td>...</td>
<td></td>
</tr>
</tbody>
</table>

Pool / COOL / ROOT
Application: Fast Monte Carlo Simulation

- Fast Monte Carlo simulation
  - Correction of the generated 4-vectors
  - Smearing and efficiency correction

- Ideal place for input from performance tool → currently “hand-made” input tables
- Communication via COOL database → to be implemented

→ A better Monte Carlo simulation

→ Reference numbers from data for systematic detector studies

→ towards a “realistic” Monte Carlo: possibility to map on run-time effects
File Sizes

- Probe DPD files are much smaller than typical AOD:
  - 200 kB/event in AOD → 0.3-0.5 kB/event in DPD

- DPD files produced on grid and collected on local storage

- Matrix files eventually stored in COOL
  - Depends on number of dimensions
  - Potentially large – for a database – if full histograms are stored
  - For 2x50 bin histograms in 20x20x20 matrix ~ 3 MB
Current Status

• All underlying functionalities are implemented
  • Tagging framework
  • DPD creation
  • Matrix representations and operations
  • COOL and ROOT database operations
    → converters

• Working use-cases:
  • Electron reconstruction and identification efficiency
  • Muon reconstruction and identification efficiency
  • Muon trigger efficiency
  • Inner detector tracking efficiency with Z and J/psi tags
Summary and Outlook

- Performance tool provides useful and standardized service to the collaboration

- Performance data can be distributed via ATLAS central database

- Interesting use cases:
  - Individual physics analysis
  - Performance groups
  - Realistic Monte Carlo simulation

- Full working analyses planned to be available in a month’s time

- “Online” exercise on Tier-0 center to be done

- Many more performance analyses to be integrated

- Currently evaluating structures for dealing with systematic uncertainties
Current Analysis Structure

Intermediate Analysis DPD

Performance Analyses:
tag & probe

Baseline Representation:
efficiency matrix
efficiency matrix entry
resolution matrix
resolution matrix entry

Tools:
probe collection
probe collector
trigger association

Pool/COOL Interface:
converters

Performance Analyses:
tag & probe

Baseline Representation:
efficiency matrix
efficiency matrix entry
resolution matrix
resolution matrix entry

1D and 2D
histograms

as histogram (Tier 0):
converters

probes collection

fills

projections

get stored

1D and 2D
histograms

as histogram (Tier 0):
converters