ATLAS Software Tutorials and Other Training

Jason Veatch
Nils Krumnack

Outline

- ATLAS Analysis Software Tutorial
- ATLAS Advanced Software Tutorials
- SLAC Software Tutorial
- ATLAS Lecture Series

ATLAS Analysis Software Tutorial

Overview

- Tutorial targeting PhD students and other Early Career Scientists in the collaboration
- Teaches common tools for major ATLAS analysis steps
- Offered 3 times per year
- Separate in-person tutorials at CERN and asynchronous online tutorials are offered
  - Material structured to work through offline
- Exercises available online
- Lectures are recorded and made available offline
- Exercise material is updated at each tutorial to reflect latest recommendations
  - Archived material is available from the documentation git history

Upcoming/Latest Tutorial
Online Dec 2024: 2 - 6 Dec 2024
CERN Nov 2024: 19 - 22 Nov 2024
Induction Day + Account Setup: 18 Nov 2024
Previous Tutorials
Online June 2024: 24 - 28 June 2024
CERN June 2024: 11 - 14 June 2024
Induction Day + Account Setup: 10 June 2024
Online March 2024: 25 - 29 March 2024
CERN March 2024: 12 - 15 March 2024

Philosophy and Curriculum

- Participants are expected to learn the broad strokes of performing an analysis from start to finish
  - Stronger focus on the available software tools and less focus on the physics motivation
- Consists of lectures followed by hands-on exercises spread over 4 days
- Each step has exercises that allow participants to try out the tools without requiring too much time
  - Time-consuming processing is skipped and inputs for each step are provided
- Exercises follow a leptoquark analysis

1. MC production and validation
2. Write a simple C++ Algorithm for analysis
3. Accessing, calibrating, and using physics objects
4. Introduction to and practice using the grid and finding datasets
5. Plotting the Z peak from ntuples
6. Optimizing a cut-based analysis
7. Implementing systematic uncertainties
8. Boosted decision trees
9. Statistical analysis/interpretation

Knowing what information is in the xAOD

Last update: 30 Aug 2024 [History] [Edit]

One question everyone will have is: how do I know what physics object information/variables are actually stored in my xAOD for each type? You an be sure for "particles" (inheriting from `IParticle`) you will have things like p_T, η, and ϕ. But what other variables are associated to the different object types? We'll try to answer that question...

Containers and key names

In order to "retrieve" the information stored in the xAOD containers we need to know the container type and the container key name. We will use a handy script called `checkxAOD.py`. If you have an xAOD file called `xAOD.pool.root` and want to know the containers and associated key names, do the following:

```
checkxAOD.py xAOD.pool.root
```

Note: You need to replace the fake `xAOD.pool.root` with the full path to an xAOD sample, for example:

```
checkxAOD.py $ALRB_TutorialData/mc20_13TeV.312276.aMcAtNloPy8E
```

Alternatively you can make a symbolic link using:

```
ln -s $ALRB_TutorialData/mc20_13TeV.312276.aMcAtNloPy8EG_A14N3
```

> If you have not set `ALRB_TutorialData` in your current shell nor set it in your `.bashrc` file, you will need to set it before using it. For more details, see xAOD Samples.

The last column will show you the xAOD container names and types. When you are retrieving information you usually need to know the container type (for example `xAOD::ElectronContainer`) and the key name for the particular instance of that container you are interested in (for example `"Electrons"`). In your analysis you can ignore the `Aux` containers (for Auxiliary store), these hold some behind-the-scenes magic. You can also "mostly" ignore the versions like `_v1`. Most information in the xAOD is stored and retrieved via the Auxiliary store. The user doesn't need to worry about this Auxiliary store, and only retrieves the interface from the event store (e.g. `TEvent` for ROOT standalone analysis). So now you should know the container type and key name. If you use the wrong key name the code will compile, but it will fail at run-time.

Statistical Analysis Exercises

Last update: 16 Aug 2024 [History] [Edit]

This is a summary of the statistical analysis exercises to be done using `TRexFitter`. More details are available in the introductory talk linked from the tutorial agenda and in the statistics tutorial repository.

You can access the tutorial material on gitlab.

The exercises you will perform are:

1. Read through the tutorial and run the example leptoquark fit
   - Setup the environment and inspect the `TRExFitter` config file (use documentation for reference)
   - Run the fit and produce / inspect the common plots from the Asimov fit (pre-fit, post-fit, tables, correlation matrix, pull plot, gammas, normalization factors, ranking plot, ...)
   - **Unblind the analysis** and re-run the fit and plots; compare to the results from the Asimov fit
2. Using the plots from 1. demonstrate that Drell–Yan and Top background normalization cannot be constrained simultaneously with the default fit setup
   - Find a solution (e.g. only one normalization factor) and compare results
3. Create `TRExFitter` config files for the remaining signal mass points (400, 700, 1800 GeV)
   - Run the fits / inspect the plots for all mass points
   - Calculate the observed and expected upper limit on signal strength for each mass point
   - Determine the lower limit on the LQ mass
   - **Bonus**: make a plot of all limits (via `TRExFitter` multi-fit functionality)
4. Instead of fitting the LQ mass, fit a BDT distribution obtained yesterday
   - Compare the results to the mass fit
5. Be creative and explore other capabilities of `TRExFitter` (e.g. likelihood scan, signal significance)

Tools Covered

- Not all available tools are covered - selected for ease of learning and broader applicability
- MC Production: MadGraph
- Finding Input Data: Run Query/COMA/AMI
- Analysis: Athena, CP Algorithms
- Grid Processing: RUCIO, PanDA
- NTuple Analysis: Jupyter Notebooks, Coffea
- Statistical Analysis: TRExFitter

Format/Schedule

In-person

- Monday: Induction Day and Account Setup
    - ID teaches participants basic details about working in ATLAS
- Tuesday: MC Production and Analysis Tools
- Wednesday: Triggers, Objects, Grid, Z-peak
- Thursday: Analysis Optimization, Systematics
- Friday: Statistical analysis
- Lectures followed by hands-on exercises with tutors in the room for immediate help
- Participants encouraged to work together

Remote

- Held 2 weeks after in-person tutorial
- Participants are asked to watch recorded lectures and work through exercises on their own
    - Lectures from the most recent in-person tutorial
- Tutors are available for assistance on Discord
    - Tutors in multiple time zones to ensure coverage

Evolution of Curriculum

- The tutorials are a living endeavor that are constantly evolving and growing
- Exercises have evolved to be more pedagogical:
  - Copy/paste commands/code -> Create a file containing this code -> Start with this code and then modify it
  - Disconnected exercises to showcase available tools -> Connected exercises using the same code base -> Analysis walkthrough using a selection of tools, many of which are coherent
- New software developments are incorporated into the exercises before each tutorial event
- Feedback is collected at the end of every tutorial event
- Material is regularly updated based on participant feedback and tutor observations
  - This feedback is often incorporated into software development as well

Planned Development

- Move to more Python and columnar analysis
- Add additional analysis tracks
- Make material more portable for locally hosted tutorial events

ATLAS Advanced Software Tutorials

Overview

- More advanced topics are covered in dedicated tutorials
  - Athena development, multi-threading, machine learning, etc.
- These are held less frequently (once every 1-2 years)
- Typically last 3-5 days
- Organized and run by topic experts in ATLAS

SLAC Software Tutorial

- An example of a local ATLAS tutorial

Overview

- Dedicated USATLAS tutorial held annually at SLAC
  - Make use of SLAC and other USATLAS computing infrastructure
- Based on the ATLAS Analysis Software Tutorial
  - Curriculum somewhat truncated to fit the 3-day schedule
- Funding provided by DOE for participant travel and lodging

Lecture Series

Overview

- An introductory lecture series aimed at early career scientists
- Lectures are given by ATLAS experts with occasional outside speakers
    - Preference given to postdocs
- Topics cover anything and everything that is relevant to people working in ATLAS
- Lectures designed with pedagogy in mind
    - Aim to give non-experts enough foundational knowledge to be able to ask questions from experts to learn more
    - 45-60 minutes of lecture at 30-45 minutes of questions/discussion
- Series began in November 2023
    - Expected to continue indefinitely recycling previous topics and giving opportunity to new lecturers

Topics covered so far

Software and Computing

Data Processing Workflows
Monte Carlo Simulation
Machine Learning
Use of GPUs in ATLAS

Physics

Top Physics
Supersymmetry
PDFs
Heavy Ion Physics
SM Higgs
DiHiggs
Long-Lived Particles
Dark Matter

Combined Performance

Tracking and Vertexing
Jet/EtMiss
Electrons and Photons
Muon Performance
Large-R Jets
Flavor Tagging
Tau CP

Detector Physics

Phase II Upgrades
Muon Spectrometer
Calorimetry
TDAQ
Triggers
Luminosity
CERN Accelerator Complex
Data Quality

Miscellaneous

CERN Knowledge Transfer
History of ATLAS
Outreach and Education

Lecture recordings and closed captioning

- Lectures are recorded and made available for asynchronous viewing
  - Synchronous attendance is encouraged
- Recordings are hosted on videos.cern.ch
- Links to videos are posted on the relevant Indico pages
- Videos are uploaded and linked within 24 hours

- Closed captioning will be provided by rev.com
- Currently finalizing logistics and funding
  - Difficulties due to CERN policies
  - Funding is provided by USATLAS
- Captioning will be made available for full catalog of recordings and all future recordings

Speakers

- Group conveners/project leaders help to identify high-quality lecturers
- Wide range of speakers have shared their expertise with the collaboration

Moderators

- The lectures could not happen without moderators who support in-person presence at CERN and remote discussion!
- Moderators run the room, introduce the speaker, and manage discussion/questions
  - We find that many students prefer to submit questions to an anonymous document
  - Moderators contribute to questions that drive discussions
- Thanks to our rockstar moderators from the ECSB for making the lectures possible!


Carolina Rossi

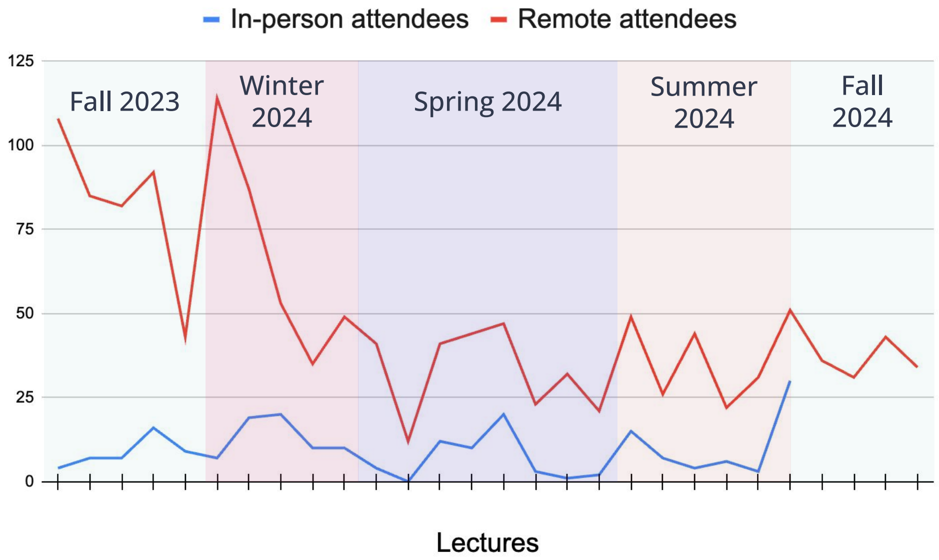
Laura Bruce


Kevin Greif

Participation over time

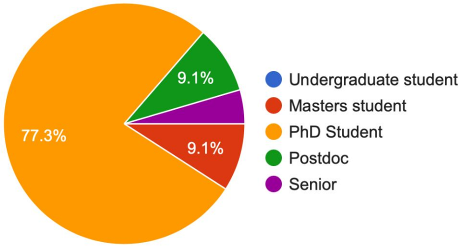
- Generally see larger remote attendance than in person
  - Remote attendance was very strong in the beginning, stabilized at 25-50
  - Happened around the time that the lecture recordings organization/posting improved

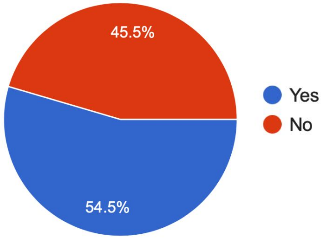
- Participation not correlated with subject area!
  - Students are showing up for detector, performance, and physics topics
- Starting in Fall 2024, lectures have moved to an entirely online format

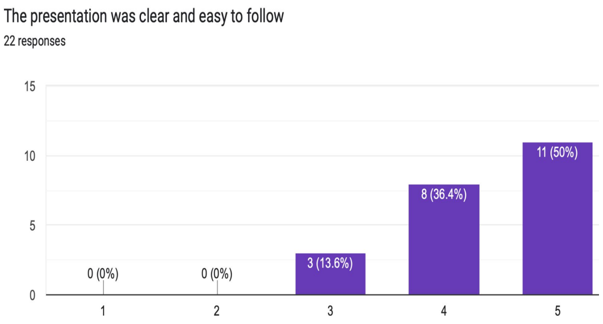


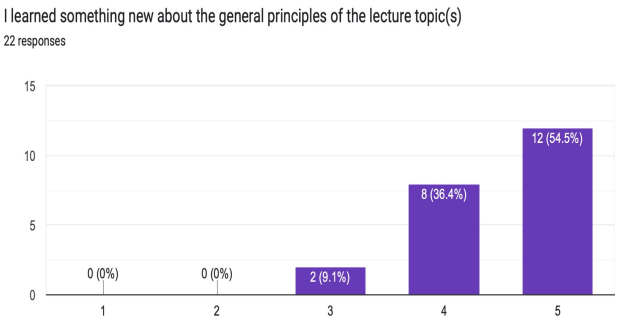
Feedback from attendees

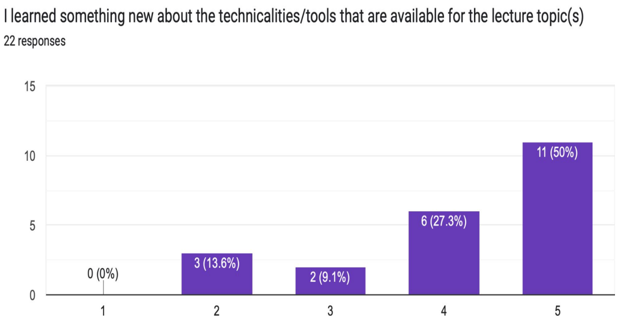
- Actively soliciting feedback on quality and effectiveness of the lectures via [survey]
  - Received 22 survey responses - very strong positive feedback (material, clarity, etc)!
  - 50% of participants have not yet completed their ATLAS Qualification Project











Selected feedback

What did/do you like about the lecture(s)?

“I liked that the presenter's slides "told a story" and it wasn't just a list of facts.”

“The speakers clear presentation and thoughtful selection of what to show in 45 mins to summarize such a large topic.”

“All of the lectures are very beginner friendly, but are also still useful for senior PhDs as well”

The lecture recordings are useful to help onboard new graduate students

What could still be improved?

Competing preference for either more detailed/more high-level lectures

Focus on less technical aspects, rather focus on pedagogical aspects

“When looking at everyone who presented a topic so far, I wonder if we could improve the diversity of the speakers a little bit?”