Project proposal

Statistical treatment of the AGC results with RooFit.

Valerii Kholoimov[1]

[1]Taras Shevchenko National University of Kyiv, Ukraine

1 Project Summary

1.1 Introduction

The IRIS-HEP Analysis Grand Challenge (AGC) is a realistic environment for investigating how high energy physics data analysis workflows scale to the demands of the High-Luminosity LHC (HL-LHC). The project offers a blueprint for HEP analysis applications that can be implemented using different tools and approaches. One of the implementations offered is done with ROOT, the tool for storing, processing and data analysis used by LHC experiments. In particular, it demonstrates usage of the RDataFrame high-level interface for data analysis in the CMS ttbar OpenData application. At the same time, it lacks the final steps of the AGC workflow, which involve the estimation of physics model parameters from the output histograms using the maximum likelihood method. The objective of this project is adding those steps via RooFit, the tool provided by ROOT for statistical analysis and advanced fitting, showcasing the use of such tool in a Python environment[1].

1.2 Analysis Grand Challenge

The Analysis Grand Challenge includes both integration of software components for analyzing the data as well as the deployment of the analysis software at analysis facilities. The vertical slice implements the functionality needed for a prototypical analysis use case with a moderately complex analysis with multiple event selection requirements, observables to be histogrammed, and systematic uncertainties that must be taken into account. The image below gives an overview of the software tools that will be integrated for this vertical slice[2].

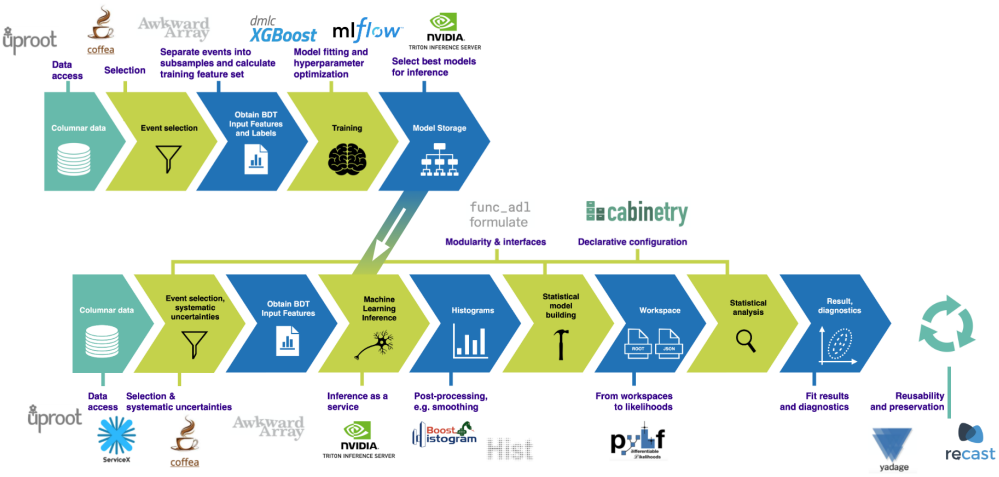

Figure 1: AGC schematic workflow

1.3 Purpose of the work

The estimation of physics model parameters from the output histograms using the maximum likelihood method already included in AGC workflow, but it uses an external frameworks for this purpose.

Goal of the project: understand the idea of fitting process, which in AGC reference implementation is made by cabinetry[3] and pyhf[4]. Implement same process via RooFit[5] and HistFactory[6] package, that creates a RooFit

probability density function from ROOT histograms. Expected fitting parameters could be provided using two different input types: JSON and internal c++ configuration. Additionally, it is very crucial that a user-friendly interface is created to help new users easily integrate this into their analyses, in addition to the documentation.

2 Timeline

The anticipated duration of the project is a three-month period, June–Aug 2024. Supervision of this project will be provided by Jonas Rembser (CERN) and Alexander Held (University of Wisconsin–Madison). A timeline with deliverables is provided below.

- **Week 1:** Studying the AGC workflow process. Producing input data for further calculations. Introduction to Coffea.
- **Week 2** Studying existing fitting algorithms, maximum likelihood methods, statistical analysis, theory of uncertainty.
- **Week 3** Investigating existed fitting process in RooFit. Studying the processes organization inside RooFit.
- **Week 4** Studying HistFactory. Introduction to cabinetry, which used for building and steering binned template fits and pyhf - a pure-python implementation of the HistFactory. Processing reference data, using pyhf and HistFactory, for intermediate comparison during development.
- **Week 5-6** Integration estimation process via HistFactory and RooFit, using maximum likelihood method. Code improvement and refactoring.
- **Week 7-8** Comparison of AGC reference implementation results with RooFit integrated fit results.
- **Week 8-10** User-friendly framework development. Integrating possibility to use JSON options files.
- **Week 11-12** Documentation preparing. Presenting the results. Final tests.

References

[1] "IRIS-HEP Projects Page." http://research-software-collaborations.org/projects_irishep.

[2] O. S. Alexander Held, "The iris-hep analysis grand challenge," in *ICHEP2022 (Bologna)*, 2022. https://agenda.infn.it/event/28874/contributions/169204/.

[3] "CABINETRY library." https://pypi.org/project/cabinetry/.

[4] "PYHF library." https://pyhf.github.io/pyhf-tutorial/IntroToHiFa.html.

[5] "ROOFIT webpage." https://root.cern/manual/roofit/.

[6] "HistFactory." https://root.cern/doc/master/group$_{HistFactory.html}$.