2nd Year PhD Annual Report

PhD Student: Lucrezia Rambelli (XXXVIII Cycle)
Supervisors: Andrea Coccaro, Francesco Armando Di Bello

September 8, 2025

The projects on which I have worked on during this second year of PhD are mainly three, where the first two can be considered as the continuation of projects started during the first year:

1. Calibrations of the ATLAS boosted tagger
2. ARIDAJE: A differentiable tracking approach
3. Spurious signal studies for Zbb + jets ATLAS analysis

1 Calibrations of the ATLAS boosted tagger

The study of the Higgs boson is a central objective of the Large Hadron Collider (LHC) research program due to its pivotal role in the Standard Model of particle physics. The Higgs boson provides a critical explanation for electroweak symmetry breaking and the origin of mass for elementary particles. Investigating its properties is crucial for testing the limits of the Standard Model and for probing potential new physics beyond it.

Since the discovery of a boson consistent with the Standard Model Higgs, subsequent measurements of its properties have significantly deepened our understanding of these fundamental processes. Recently, there has been a growing interest in studying Higgs boson production at high transverse momenta, as this regime can act as a sensitive probe for physics beyond the Standard Model, particularly through momentum-dependent anomalous couplings. In this regime, where jets from the final states become highly collimated, traditional jet-tagging techniques struggle to perform effectively, and the final state is reconstructed building a single large radius jet around the decay products.

In this context, the ATLAS Collaboration has been developing a specialized tagger, GN2x, designed to select boosted Higgs boson decays into pairs of heavy quarks, specifically b- and c-quarks. The GN2x tagger operates by directly analyzing the charged particle tracks within the large-radius jet, allowing to select $H - b\bar{b}/ - c\bar{c}$ final states and rejecting at the same time the QCD and the other backgrounds with high performance.

Since GN2x is a machine learning-based model trained on Monte Carlo (MC) samples, it requires calibration to ensure that its performance is aligned between MC simulations and real data. As part of my ATLAS Qualification Project during the first year of my PhD, I began working on these calibrations, and in the second year, I completed the project, earning official authorship within the ATLAS Collaboration.

Specifically, my work focused on calibrating GN2x for topologies that cannot be directly calibrated using real data, relying instead on *adjusted* MC samples. The adjusted MC method follows a bottom-up approach, where auxiliary measurements at the track level are propagated to the jet level, modifying the GN2x score to more closely match real collision data.

Following guidelines from the ATLAS Tracking Group, I developed a pipeline for generating multiple adjusted versions of the same MC sample, each incorporating different tracking systematic uncertainties. After performing various checks on how these modifications impacted key variable distributions (e.g., smearing of the impact parameter distribution core), I deployed a comprehensive Python-based workflow. This workflow was designed to compute the tagging efficiency for each adjusted sample and compare it with the nominal (original) sample, ultimately yielding the scale factor values needed to align the tagger's performance between data and MC samples.

Throughout the year, I presented my results to the Flavour Tagging group, refining the project direction based on feedback. The final results demonstrate that the inclusion of all available tracking systematic uncertainties leads to significant changes in the tagger's response only in final states without b- or c-hadrons. In these cases, it is essential to collaborate with other ATLAS groups to incorporate additional relevant track systematics for further refinement.

2 ARIDAJE: A differentiable tracking approach

In the second year of my PhD, I worked on developing a machine learning-based differentiable tracking algorithm, focusing on charged particle tracking in the Muon Spectrometer (MS). The approach uses differentiable programming techniques to incorporate physics-based information into a graph-based model. This is achieved through the use of the Google public library JAX, which provides key tools such as just-in-time (JIT) compilation, vectorization, and automatic differentiation. These capabilities are essential for injecting physics constraints, such as a circular fit term representing the curved trajectory of charged particles in the spectrometer's magnetic field, directly into the model's training. This circular fit term is included in the loss function, allowing the optimization of both the model parameters and the fit parameters throughout the training process.

In the first year, the project focused on preliminary development, including selecting the data preprocessing method, designing the graph-based structure, and choosing the neural network architecture for processing. Each event's signal track was split into two segments due to memory limitations. The *outer* part of the event was represented as a fully connected graph, where each node corresponded to a spectrometer hit, with hit coordinates serving as features.

To classify the nodes in the *outer* graph, a Graph Attention Network (GAT) was developed and trained, with performance evaluated in terms of classification accuracy.

Building on this, a series of differentiable functions were implemented over the first and second years. These functions perform tasks such as clustering external hits based on node predictions, applying a weighted circular fit to the clusters, and extrapolating track information to the innermost detector layers. A differentiable function was also created to select the nearest hits to the extrapolated track, whose coordinates are used as features in another graph that represents the inner part of the event.

Initially, during the first year, these *inner* graphs were processed by a separate Multi-Layer Perceptron (MLP) model for hit classification. However, in the second year, a unified model was developed to process both the outer and inner parts of each event using a shared set of parameters. This optimized approach has shown promising results, improving both classification accuracy and the precision of transverse momentum (p_T) reconstruction, which can be estimated from the fitted track's radius.

3 Spurious signal studies for Zbb calibration

At the end of this year I started to work on a data analysis within the ATLAS Collaboration. In particular I'm starting to work on the spurious signal test for an analysis in which events in which a Z boson decaying in a couple of b-quarks is produced together with a jet in a boosted regime.

In this context, the spurious signal test is important as it makes possible to quantify the modeling bias-related to the chosen MC background description model.

This bias can be estimated using $Z \to b\bar{b}$ + jets and $Z \to l^+l^-$ + jets events, and allows to validate the SF values for the boosted tagger obtained before with the Adjusted MC method (explained in Section 1).

Under the assumption that the cross-section for the two processes is the same for each Z-candidate p_T bin, SF values can be computed comparing tagging efficiencies on data events and MC ones, and it can be shown that SFs can be expressed as the ratio betwee n the signal strength pre- and post-tag.

The $\mu_{post-tag}$ value can be extracted fitting the mass distribution of events with at least a large-R jet identified as a Z-boson candidate and modelling the large amount of background given by multijet events.

Due to the overwhelming multijet background, a $\mu_{pre-tag}$ measurement with $Z \to b\bar{b}$ events is unfeasible. For this reason, the pre-tag signal strength is obtained using $Z \to ll$ events in data and MC samples, and SF values are then computed for each Z-boson candidate p_T bin as $\mu_{post-tag}/\mu_{pre-tag}$.

Regarding this work I am currently doing some preliminary studies looking at how the samples are produced and comparing the MC ones (for signal and background) with the available data.

Attended Courses and Exams Given:

- *The double trouble of the missing matter in the Universe*: Oral presentation on MOdified Newtonian Dynamics (MOND) Theories
- *Theoretical physics*: Oral exam on massive neutrino theories
- *Advanced Statistics for Data Analysis*: exam not given yet
- *2024 European School of High-Energy Physics*: will participate in September 2024
- *Fisica 1 and ALGA Tutoring*