

Analysis and Interpretation of the Four-Lepton Invariant Mass Spectrum Using CERN Open Data

Project Report

January 2026

Abstract

This report details the analysis of real CERN Open Data to reconstruct the four-lepton invariant mass spectrum. The study focuses on the decay channel $H \rightarrow ZZ^* \rightarrow 4\ell$, known as the "golden channel." We present the dataset description, the physics observable, the analysis workflow, and a detailed interpretation of the resulting spectrum in the context of Standard Model expectations.

Contents

1	Introduction	2
2	Dataset Description	2
3	Physics Observable: Four-Lepton Invariant Mass	2
4	Data Processing and Analysis Workflow	3
5	Expected Physics Contributions	3
5.1	Background: Continuum ZZ Production	3
5.2	Signal: Higgs Boson	3
6	Interpretation of the Invariant Mass Spectrum	3
6.1	What the Distribution Shows	4
6.2	Absence of Resonant Peaks	4
6.3	Physical Interpretation	5
7	Conclusion	5
8	Future Improvements	5

1 Introduction

The discovery of the Higgs boson in 2012 by the ATLAS and CMS collaborations marked a major milestone in particle physics. One of the cleanest and most important discovery channels is the decay:

$$H \rightarrow ZZ^* \rightarrow 4\ell \ (\ell = e, \mu) \quad (1)$$

commonly referred to as the *golden channel*. Despite its very small branching ratio, this channel provides excellent mass resolution and low background contamination, making it ideal for precision studies.

In this project, real CERN Open Data are analyzed to reconstruct the four-lepton invariant mass spectrum. The goal is **not** to rediscover the Higgs boson, but to perform a realistic small-statistics analysis, understand background-dominated data, and demonstrate proper experimental reasoning consistent with LHC analyses.

2 Dataset Description

The data used in this project come from publicly released CERN Open Data in ROOT format

`ODE0_FEB2025_v0_exactly4lep_data15_periodD.exactly4lep.root`

. The file contains reconstructed lepton-level information for proton–proton collision events recorded at the LHC.

Link for the root file: <https://opendata.cern.ch/record/93924/>

Each event consists of:

- Four reconstructed leptons (electrons or muons)
- Lepton four-momenta (E, p_x, p_y, p_z)
- Event-level information suitable for invariant mass reconstruction

A total of **3424 four-lepton events** pass the basic selection criteria and are used in this analysis.

3 Physics Observable: Four-Lepton Invariant Mass

The central observable is the four-lepton invariant mass, defined as:

$$m_{4\ell} = \sqrt{\left(\sum_{i=1}^4 E_i\right)^2 - \left|\sum_{i=1}^4 \vec{p}_i\right|^2} \quad (2)$$

This quantity is Lorentz invariant and directly sensitive to resonant particle decays, such as:

- $Z \rightarrow \ell\ell$
- $H \rightarrow ZZ^* \rightarrow 4\ell$

4 Data Processing and Analysis Workflow

The ROOT file is analyzed using a Jupyter Notebook (`Analysis-1.ipynb`) with Python-based HEP tools. The main steps are:

1. Reading the ROOT file using appropriate Python libraries
2. Extracting lepton four-momenta for each event
3. Constructing four-vectors and summing them event-by-event
4. Computing the invariant mass $m_{4\ell}$
5. Filling a histogram over the range 70–200 GeV

This workflow closely mirrors the structure of real LHC analyses, at a simplified and educational scale.

5 Expected Physics Contributions

Two main Standard Model processes contribute to the four-lepton invariant mass spectrum:

5.1 Background: Continuum ZZ Production

$$pp \rightarrow ZZ^{(*)} \rightarrow 4\ell \quad (3)$$

This process produces a **smooth, falling invariant mass distribution** and dominates the event yield across the full mass range.

5.2 Signal: Higgs Boson

$$pp \rightarrow H \rightarrow ZZ^* \rightarrow 4\ell \quad (4)$$

The Higgs signal would appear as a **narrow resonance** around:

$$m_H \approx 125 \text{ GeV} \quad (5)$$

However, due to the extremely small branching ratio and limited dataset size, only a very small number of Higgs events are expected.

6 Interpretation of the Invariant Mass Spectrum

This section details the interpretation of the reconstructed spectrum shown in the analysis results. The figure (Figure 1) displays the four-lepton invariant mass $m_{4\ell}$ obtained from the dataset.

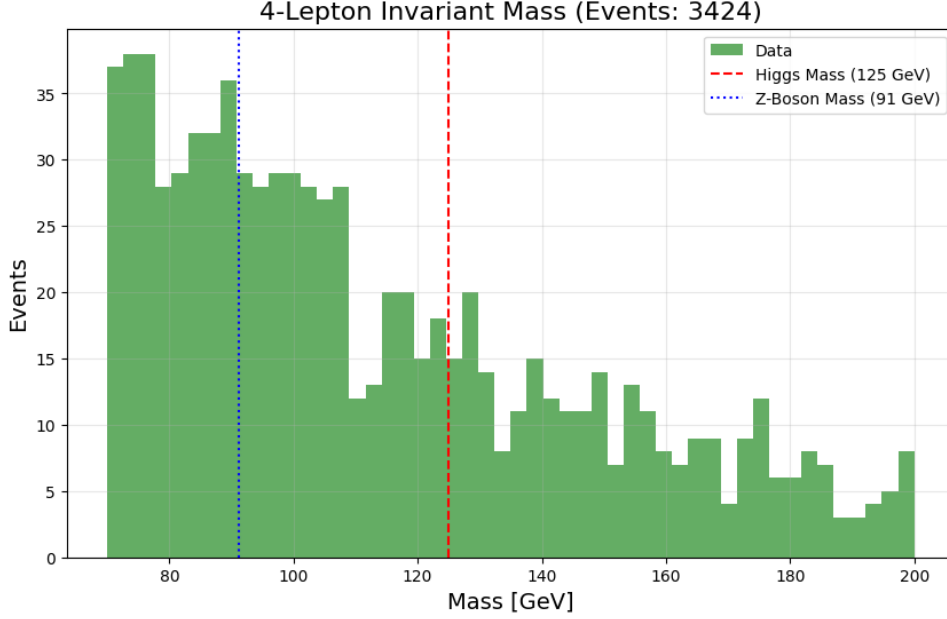


Figure 1: The reconstructed four-lepton invariant mass distribution.

The invariant mass is calculated from the combined four-momenta of the leptons:

$$m_{4\ell} = \sqrt{\left(\sum_{i=1}^4 E_i\right)^2 - \left|\sum_{i=1}^4 \vec{p}_i\right|^2} \quad (6)$$

This observable plays a crucial role in Higgs boson studies in the golden channel $H \rightarrow ZZ^* \rightarrow 4\ell$, as it provides good mass resolution and a clean experimental signature.

6.1 What the Distribution Shows

The invariant mass spectrum exhibits a **smoothly falling distribution** across the full mass range. This behavior is characteristic of the dominant **Standard Model background**, primarily arising from the process:

$$pp \rightarrow ZZ^{(*)} \rightarrow 4\ell \quad (7)$$

Two vertical reference lines are typically shown on such plots:

- **Z boson mass (≈ 91 GeV):** This line indicates the on-shell Z boson mass. In a four-lepton final state, a sharp peak is not expected at this value, since the observable corresponds to the combined mass of four leptons rather than a dilepton pair. Nevertheless, it provides a useful physical scale and marks the region where one Z boson becomes on-shell in $ZZ^{(*)}$ production.
- **Higgs boson mass (≈ 125 GeV):** This line indicates the expected position of a Higgs resonance decaying via $H \rightarrow ZZ^* \rightarrow 4\ell$.

6.2 Absence of Resonant Peaks

No clear excess is observed near either of the reference masses. In particular:

- No visible Higgs peak is present near 125 GeV.
- The region around 91 GeV follows the expected smooth background behavior for a four-lepton invariant mass distribution.

The absence of a visible Higgs peak is expected because the decay $H \rightarrow ZZ^* \rightarrow 4\ell$ has a very small branching ratio, and the available CERN Open Data sample is limited. Any Higgs contribution is therefore much smaller than the dominant $ZZ \rightarrow 4\ell$ background and cannot be identified by visual inspection alone.

6.3 Physical Interpretation

The measured four-lepton invariant mass spectrum is consistent with:

- Standard Model expectations
- A background-dominated sample of $ZZ^{(*)} \rightarrow 4\ell$ events

The lack of a statistically significant Higgs signal is a consequence of the limited dataset size rather than a deficiency in the analysis.

7 Conclusion

A complete four-lepton invariant mass analysis was performed using real CERN Open Data. The observed spectrum is dominated by Standard Model background processes, and no statistically significant Higgs signal is observed. This outcome is expected and reflects realistic experimental conditions.

This project demonstrates:

- Correct reconstruction of a key LHC observable
- Sound physical interpretation
- Proper use of statistical reasoning

8 Future Improvements

Possible extensions of this work include:

- Signal and background fitting
- Optimization of event selection criteria
- Combination with simulated datasets
- Extension to other decay channels

These steps would improve sensitivity and more closely resemble full-scale LHC analyses.