Computational Methods in High-Energy Physics: Z boson-like Particle Detection in CMS

Emilia Veikkola, Nestor Mancilla

May 22, 2025

1 Introduction

This project studies the detection of the Z boson-like particle [1] via the Z $\rightarrow \mu\mu$ decay channel in the CMS experiment during LHC Run 3 at 13.6 TeV, with $t\bar{t}$ production as the primary background. Using simulated events generated with PYTHIA8, we analyze trigger efficiencies, apply muon selection criteria, reconstruct the invariant mass, and estimate statistical significance. The results demonstrate rapid detection of the Z boson. Lastly we discuss improvements to enhance realism and scientific credibility.

2 Event Generation

Events were generated using `PYTHIA8` with the following criteria:

- **Center-of-mass energy:** 13.6 TeV
- **Physics processes:** `WeakSingleBoson:ffbar2gmZ` (signal), `Top:gg2ttbar` and `Top:qqbar2ttbar` (background)
- **Trigger:** `HLT_DoubleIsoMu20_eta2p1` ($p_T > 20$ GeV, $|\eta| < 2.1$ for muons)
- **Generated events:** 10 million for Z$\rightarrow \mu\mu$ (signal), 10 million for $t\bar{t}$ (background)
- **Cross-sections:** 62 pb (signal), 924 pb (background)

Ten million events were generated to ensure that statistical uncertainties remain low ($<$ 1%); this choice is validated by the post-selection counts described later in the Section 3.2.

The cross section obtained from PYTHIA for $t\bar{t}$ was negligible compared to the one for the signal. When normalizing the events to their respective cross sections, the background contribution to the invariant mass spectrum was insignificant compared to the signal. However, since the analysis requires studying both signal and background distributions, relying solely on the signal contribution would make the background statistically insignificant in the plot. For this reason, instead of using the PYTHIA cross sections, we adopted the reference cross sections from the Particle Data Group (PDG) [1] and the recommendations from ATLAS and CMS [2], which are more accurate and ensure the presence of the background for a better comparison.

[1] a particle with the same properties as Z boson but with weaker couplings

3 Analysis

3.1 Trigger Efficiency

Both signal and background samples were generated with 10^7 events each; their trigger efficiencies are obtained by comparing the number of events that pass the trigger with the total generated events:

$$\epsilon_{\text{trigger}} = \frac{\mathrm{N_{pass}}}{\mathrm{N_{generated}}},$$

Applying this definition gives the following efficiencies:

	Passed events	Trigger efficiency
Signal (Z$\to \mu\mu$)	769 958	7.70 %
Background ($t\bar{t}$)	177 773	1.78 %

3.2 Muon Selection

Muon candidates were selected with the following criteria:

- Gaussian smearing: 1% (momentum), 2 mrad (θ, ϕ).
- $p_T > 30$ GeV
- Track isolation: Sum of p_T of charged pions within $\Delta R < 0.3$ less than 1.5 GeV

The number of events passing these selections:

	Selected events	Selection efficiency
Signal (Z$\to \mu\mu$)	554 258	5.54 %
Background ($t\bar{t}$)	79 567	0.796 %

Next we want to test the statistical uncertainties to verify the number of generated events. With pure Poisson counting the *relative* statistical uncertainty is simply $1/\sqrt{N_{\text{sel}}}$, giving

$$\delta_{\text{rel}}^{\text{sig}} = \frac{1}{\sqrt{554\,258}} \approx 0.13\%, \qquad \delta_{\text{rel}}^{t\bar{t}} = \frac{1}{\sqrt{79\,567}} \approx 0.36\%.$$

Both are well below the one-percent level, confirming that the sample size is more than sufficient for the analysis that follows.

3.3 Invariant Mass Reconstruction

The invariant mass is calculated from the four-momenta of opposite-sign muon pairs. The explicit formula involves energy and the momentum:

$$m_{\mu\mu} = \sqrt{(E_1 + E_2)^2 - (\vec{p_1} + \vec{p_2})^2}.$$

However, the input variables are the experimentally measured quantities: p_T, η, azimuthal angle ϕ, and the energy E. The conversion is done internally by TLorentzVector::SetPtEtaPhiE, as defined by CERN ROOT [3].

Invariant mass distributions ($M_{\mu\mu}$) were plotted and fitted to distinguish signal from background. Figures 1 and 2 show these distributions.

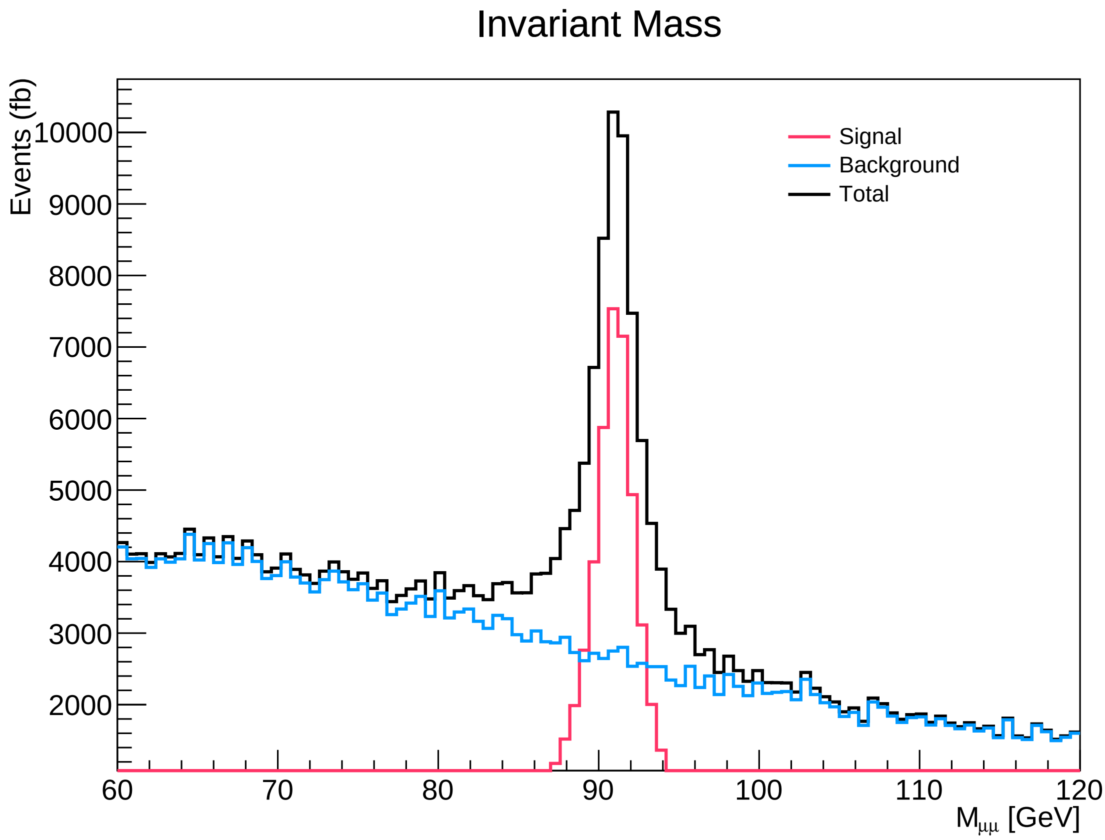

Figure 1: Invariant mass before selection.

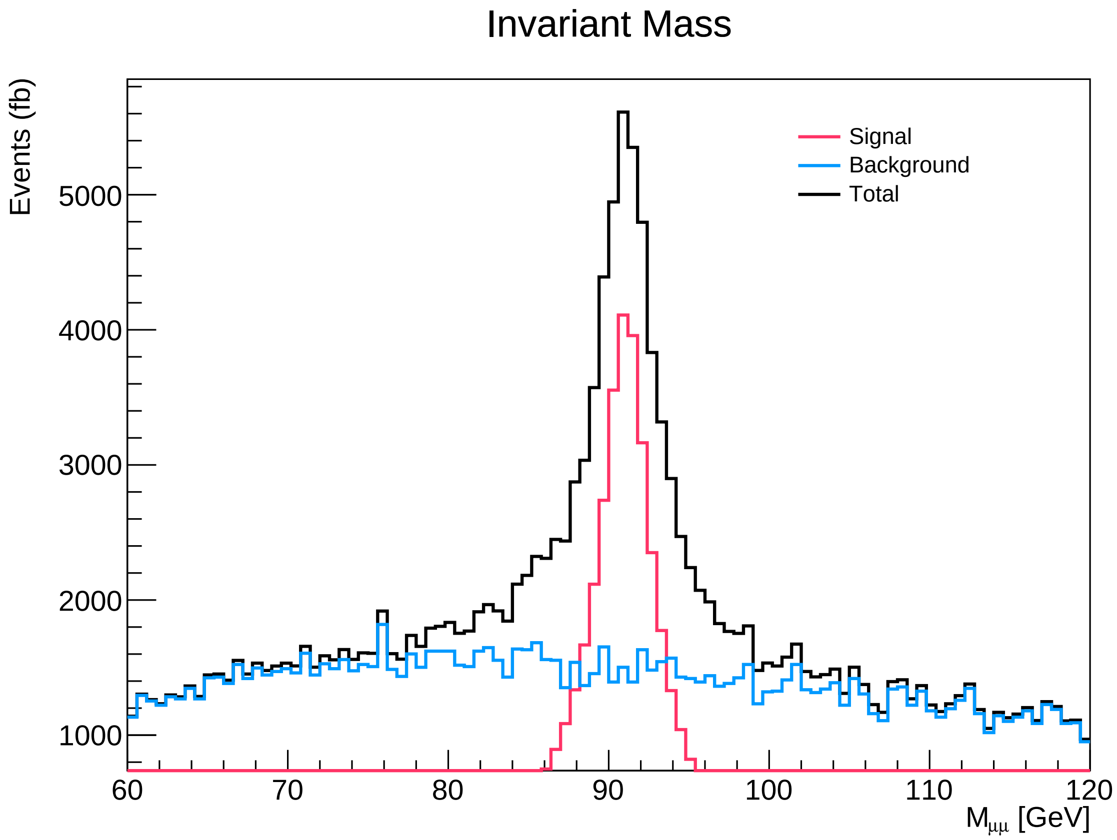

Figure 2: Invariant mass after selection.

4 Results

The invariant mass of muon pairs was reconstructed to identify Z-boson events. To model the distribution and extract the signal yield, three fit functions were applied to the combined signal and background histogram in the range 60 GeV to 120 GeV, shown in Figure 3:

- **Gaussian + linear background:** A Gaussian peak atop a linear continuum.
- **Voigtian + linear background:** A Voigtian function, which is "the convolution of a Lorentzian and a Gaussian" [4], plus a linear background.
- **Voigtian + Chebyshev(1&2) background:** A Voigtian signal with first- and second-order Chebyshev polynomial backgrounds to better capture curved backgrounds.

The signal window was defined as 80 GeV to 100 GeV, and the significance was calculated as $N_S/\sqrt{N_B}$ where S and B are the fitted signal and background entries within this range. The significance of the signal was calculated from the fitted distributions (here Voigtian + Chebyshev):

$$\text{Significance at 1 fb}^{-1} = \frac{N_S}{\sqrt{N_B}} = \frac{21855}{\sqrt{31288}} \approx 124\ \sigma$$

Fit Model	N_S [fb]	N_B [fb]	Significance [σ]	Peak [GeV]
Gaussian + linear	14 954	35 261	79.6	91.00
Voigtian + linear	17 244	35 261	91.8	91.01
Voigtian + Chebyshev(1&2)	21 855	31 288	123.6	91.03

Table 1: Comparison of fit models for the $M_{\mu\mu}$ distribution.

The Voigtian-based fits show a better match to the resonance line shape, especially when combined with a Chebyshev background.

Time needed at the LHC. Because the assumed significance grows with the square root of the collected integrated luminosity,

$$S(L) = \frac{N_S}{\sqrt{N_B}}\sqrt{L}\ , \qquad \frac{N_S}{\sqrt{N_B}} = 123.6\ \sigma,$$

the luminosity that yields a 5σ observation is

$$L_{5\sigma} = \left(\frac{5}{123.6}\right)^2 \simeq 1.64 \times 10^{-3}\ \text{fb}^{-1}.$$

As of September 2, 2024, the CMS experiment had collected approximately $88.9\,\text{fb}^{-1}$ of of high-quality data over ~107 days of of proton-proton collisions [5], achieving an average luminosity rate of about:

$$\left\langle \frac{dL}{dt} \right\rangle \simeq \frac{88.9\,\text{fb}^{-1}}{107\ \text{days}} \approx 0.83\ \text{fb}^{-1}/\text{day}.$$

At that rate the time to accumulate $L_{5\sigma}$ is therefore

$$t\ =\ \frac{1.64 \times 10^{-3}\ \text{fb}^{-1}}{0.83\ \text{fb}^{-1}/\text{day}} = 0.00197...\ \text{day}\ \approx\ 3\ \text{minutes}.$$

Even with realistic machine efficiency the $\text{Z} \to \mu\mu$ signal reaches 5σ well within the first hour of LHC data taking.

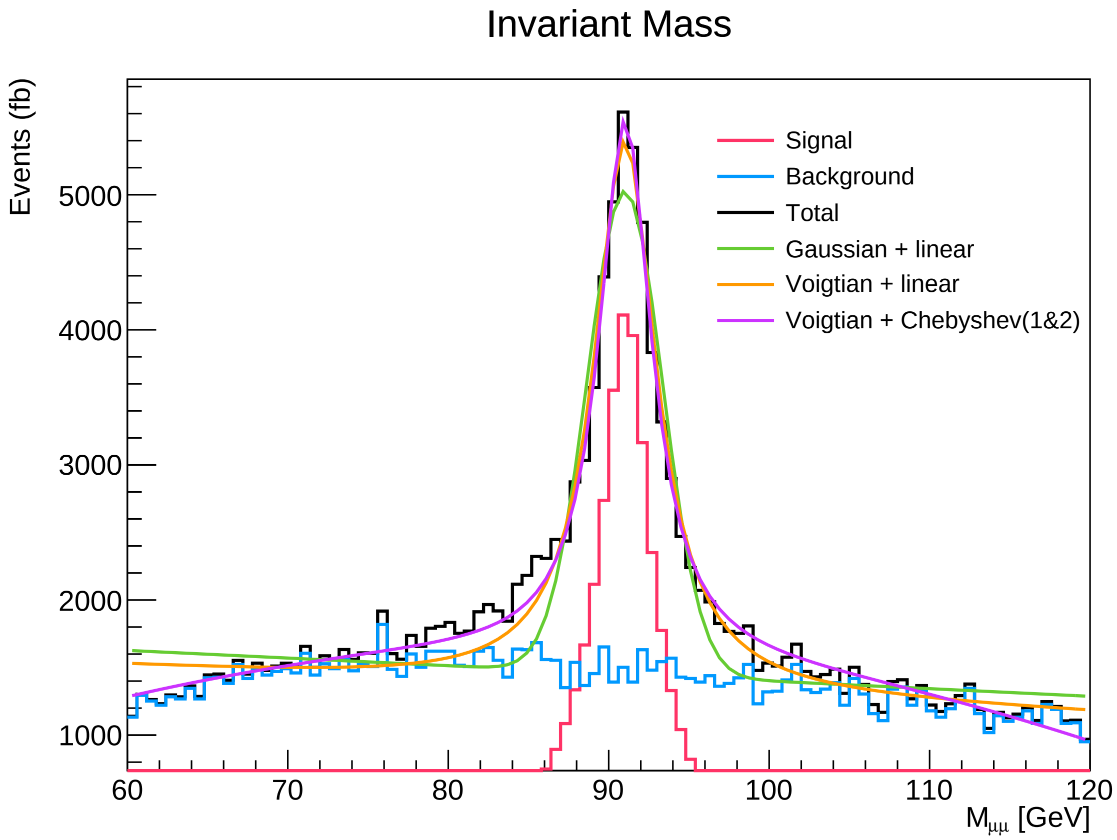

Figure 3: Invariant mass with different fits to signal and background.

5 Discussion and Improvements

The analysis identifies the Z $\rightarrow \mu\mu$ signal with very high statistical significance ($> 100\sigma$), primarily due to the distinctive Z resonance peak at approximately 91 GeV and the high Z boson production cross-section at the LHC. However, this significance should be interpreted cautiously, as the current simulation uses simplified event selection, fixed cross-sections (not directly from `PYTHIA8`), simulated detector effects by smearing, and an arbitrary number of simulated events that directly scales the significance. Real experimental conditions, including detector inefficiencies, systematic effects, and additional backgrounds, would likely reduce the observed significance.

For improved realism and scientific credibility, future analyses should:

- **Enhance detector realism** by incorporating detailed CMS detector simulations (e.g., GEANT4).
- **Include comprehensive background modeling**, particularly electroweak diboson production and the Drell–Yan continuum.
- **Use measured cross-sections** and associated uncertainties for more accurate results.
- **Assess systematic uncertainties** such as luminosity, muon reconstruction efficiency, and background normalization.
- **Validate with data-driven methods**, including sideband analyses and control region studies.

The current project is instructive but insufficient for a discovery claim. However, the Z peak passes basic sanity checks, so we can trust the results. Without the suggested improvements for validation, this remains a demonstration.

References

[1] M. Tanabashi et al. (Particle Data Group). Review of Particle Physics: Z Boson. *Phys. Rev. D*, 98:030001, 2018. `https://pdg.lbl.gov/2018/listings/rpp2018-list-z-boson.pdf`.

[2] M. Czakon, A. Mitov, et al. Top quark pair production cross-section at NNLO at the LHC. `https://twiki.cern.ch/twiki/bin/view/LHCPhysics/TtbarNNLO`, 2019. Accessed: 2025-05-14.

[3] ROOT Collaboration. *ROOT Data Analysis Framework: TLorentzVector Class Reference*. CERN, Geneva, Switzerland, 2023. Version 6.28/00.

[4] The CMS Collaboration. Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV. *Journal of Instrumentation*, 7(10):P10002–P10002, October 2012. `https://arxiv.org/pdf/1206.4071`.

[5] R. Steerenberg. Accelerator Report: LHC Run 3 achieves record-breaking integrated luminosity. `https://home.cern/news/news/accelerators/accelerator-report-lhc-run-3-achieves-record-breaking-integrated-luminosity`, September 2024. Accessed: 2025-05-21.