Analysis of the Dimuon Mass Spectrum in CMS Open Data

Liutauras Pocius
Fizika, Fizikos fakultetas, Vilniaus Universitetas
2025-12-08

Data and method. I analysed a public CMS dimuon dataset from the 2011 run at $\sqrt{s} = 7$ TeV (DoubleMu primary dataset, about 10^5 events) made available via the CERN Open Data Portal.[1] Each event contains two reconstructed muon candidates with four-momentum components, transverse momentum p_T, pseudorapidity η, azimuthal angle ϕ and charge Q, as well as the dimuon invariant mass M in GeV.

Using Python, `numpy` and `pandas` I recomputed the invariant mass from the four-vectors according to

$$m_{\mu\mu}^2 = (E_1 + E_2)^2 - (p_{x1} + p_{x2})^2 - \\ -(p_{y1} + p_{y2})^2 - (p_{z1} + p_{z2})^2.$$

A consistency check shows that the difference $M_{\mathrm{CSV}} - M_{\mathrm{reco}}$ has mean $\langle \Delta M \rangle = 2.6 \cdot 10^{-5}$ GeV and RMS $\sigma_{\Delta M} = 4 \cdot 10^{-3}$ GeV, i.e. the reconstructed masses match the values provided in the dataset within numerical precision. I then filled histograms of $m_{\mu\mu}$: a wide range [0, 125 GeV] displayed on a logarithmic y-axis, and a narrow window [80 GeV, 100 GeV] around the Z peak. The latter is fitted with a simple Gaussian model using `scipy.optimize.curve_fit`.

Results. Figure 1 displays the overall dimuon spectrum. On a single logarithmic plot we observe QCD and electroweak physics across two orders of magnitude in mass: the J/ψ and $\psi(2S)$ charmonia, the Υ bottomonium family and a clear Z boson peak, all sitting on top of the Drell-Yan background. Integrating fixed mass windows I found, for example, $N_{J/\psi} = 8422$, $N_{\Upsilon} = 6359$ and $N_Z = 5174$ events in the J/ψ, Υ and Z regions, respectively. Figure 2 focuses on the $Z \to \mu^+\mu^-$ region and overlays a Gaussian fit. The extracted mass is $\mu_Z = 90.81$ GeV and the effective width $\sigma_Z = 2.22$ GeV, where the uncertainties are taken from the fit covariance. Within the statistical precision and simplistic line-shape model, μ_Z is compatible with the world-average Z boson mass $M_Z \approx 91.2$ GeV. The fitted σ_Z is dominated by detector resolution and selection effects rather than the physical Z width.

[1] CMS Open Data, dimuon sample `Dimuon_DoubleMu.csv`, record 545, CERN Open Data Portal, `http://opendata.cern.ch/record/545`.

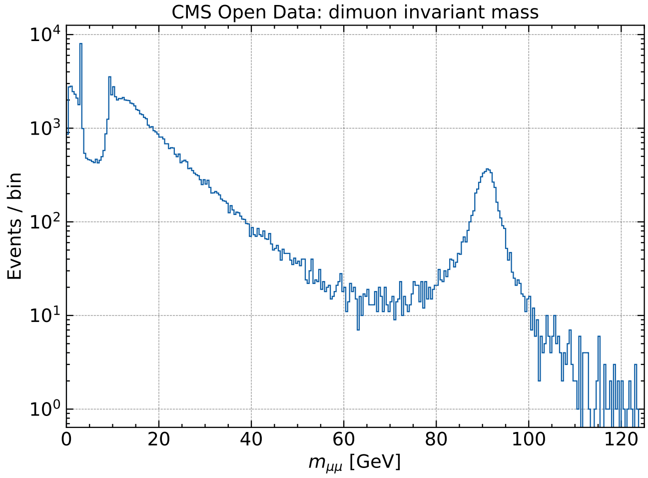

Fig. 1: Dimuon invariant mass spectrum in CMS open data. The plot shows narrow charmonium peaks (J/ψ and $\psi(2S)$) at a few GeV, the Υ region around 10 GeV, the broad Drell-Yan continuum and the dominant $Z \to \mu^+\mu^-$ resonance near 91 GeV.

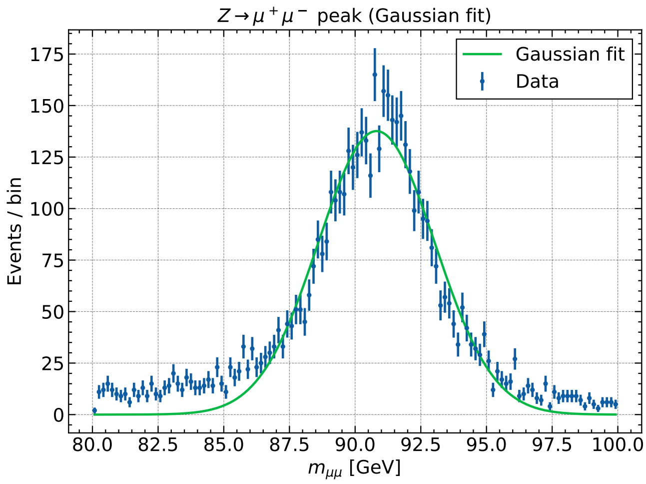

Fig. 2: Dimuon invariant mass in the Z region with a Gaussian fit. Points show data with Poisson uncertainties; the curve is the best-fit model.

Conclusion. Using a small open dataset and a compact, modular Python analysis we can reproduce the characteristic resonant structure of the dimuon spectrum at the LHC and obtain a reasonable estimate of the Z mass. The code is organised into separate modules for I/O, physics logic, fitting and plotting, making it straightforward to extend this study to more refined selections or alternative models.