LIP-STUDENTS-20-18

Muon efficiency studies using Tag and Probe method

Allan Jales[1,a] and Thomas Gaehtgens[2,b]

[1] *Universidade do Estado do Rio de Janeiro, Rio de Janeiro, Brazil*
[2] *Instituto Superior Técnico, Universidade de Lisboa, Lisboa, Portugal*

Project supervisors: E. Melo[1], N. Leonardo[2], S. Fonseca[1]

October 2020

Abstract. We have carried out measurements of muon identification and reconstruction efficiency, using the data-driven Tag and Probe method. We have developed two implementations of the method, relying on two distinct approaches for extracting the signal muons from data. The method is applied to proton-proton collision data collected by the CMS experiment at the LHC (LHC Open Data). The efficiency results extracted from the datasets are compared against Monte Carlo simulation, thus providing calibration factors that can be used for correcting the simulation in physics analyses involving muons.

KEYWORDS: LHC, CMS, Muon, Sideband-subtraction, Likelihood Fitting, Detector Calibration, Open Data

1 Introduction

The Tag and Probe method is an experimental procedure commonly used in particle physics that allows to measure a process' efficiency directly from data. The procedure provides an unbiased sample of *probe* objects that can be then used to measure the efficiency of a particular selection criteria.

The paper is organized as follows. The CMS detector is introduced in Sec. 2, along with the muon reconstruction algorithms, and the datasets used in this work. The Tag and Probe method is presented in Sec 3, together with the particle resonances studied. The methods employed to extract the signal muons from data are explained in Sec. 4. The measured muon efficiencies are finally presented in Sec. 5.

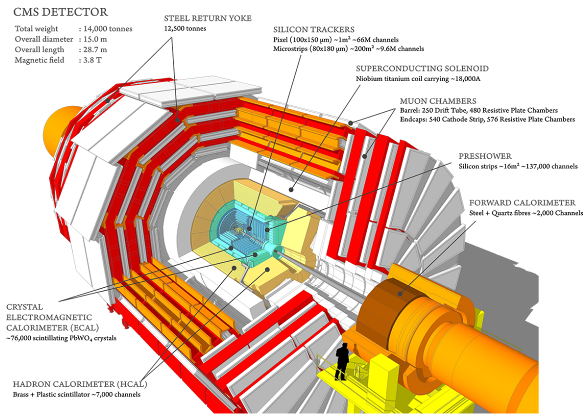


Figure 1: Schematic view of the CMS detector.

2 Muons at CMS

Muons are central physics objects employed in virtually any analysis. The measurement of the efficiency of their detection, identification and reconstruction is correspondingly an essential element of a physics measurement.

2.1 The CMS experiment

The CMS (Compact Muon Solenoid) detector, represented schematically in Fig. 1, is a general-purpose experiment at the LHC (Large Hadron Collider). Particles produced in the LHC collisions interact with several sub-detectors, allowing for their trajectories and energy depositions to be reconstructed, and their momenta to be precisely determined.

The CMS detector has an overall cylindrical shape, with its sub-detectors disposed in layers around the central region, where the collisions occur. In the innermost part, near the collision point, there is a silicon tracker that traces the passage of charged particles. It is followed by an electromagnetic calorimeter, where electrons and photons will lay their energy, and an hadronic calorimeter, where hadrons will deposit their energy. Further outwards, there is a superconducting solenoid that produces a 3.8 T magnetic field and, on the outermost part, there are the muon chambers, made of stations of gas-ionization detectors interleaved with layers of steel return yoke. A detailed description of the CMS detector may be found elsewhere [1].

2.2 Relevant variables

The main event observables, obtained from the reconstructed trajectories of charged particles in the detector, that are used in the analysis are:

- $\eta = -\ln(\tan\theta/2)$: pseudorapidity;
- ϕ : angle of the trajectory of the object in the plane transverse to the direction of the proton beams;
- p_T : transverse momentum: projection of the particle's momentum in the plane transverse to the direction of the proton beams;
- m : invariant mass of the muon pair.

These are illustrated in Fig. 2. The invariant mass variable m will be used as the discriminating variable, that will al-

[a] e-mail: jales.allan@cern.ch
[b] e-mail: thomas.gaehtgens@tecnico.ulisboa.pt

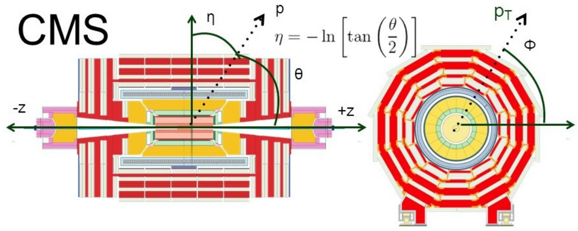


Figure 2: Schematic view of CMS main variables.

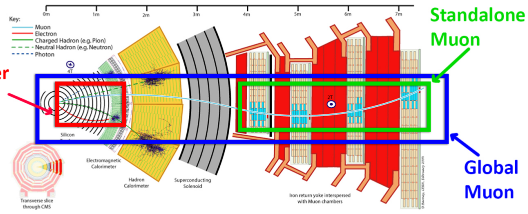


Figure 3: Schematic transverse view of the CMS detector shows muon identification algorithms.

low to distinguish signal from background. The momentum of a charged particle such as the muon can be inferred from the curvature of its trajectory, registered in the silicon tracker, in the presence of the magnetic field. The identification of the particle as indeed a muon is achieved with the muon chambers. This information allows to construct the 4-momentum of the particle, $p = (E, \mathbf{p})$. Here we will be interested in reconstructing the invariant mass of a resonance particle that decays into a pair of muons. This is achieved by summing the 4-momenta of the individual muons, and taking its norm,

$$m^2 = p_1 \cdot p_2 = (E_1 + E_2)^2 - \|\mathbf{p_1} + \mathbf{p_2}\|^2 . \quad (1)$$

The obtained invariant mass spectra, i.e. the histogram distribution of m reconstructed from the collision data, may be seen in Figs. 5a and 5b.

2.3 CMS Muon identification and reconstruction

In the standard CMS reconstruction for proton-proton collisions, tracks are first reconstructed independently in the inner tracker and in the muon system. Based on these objects, three reconstruction approaches are used:

- Tracker Muon reconstruction: all tracker tracks with $p_T > 0.5\ GeV/c$ and total momentum $p > 2.5\ GeV/c$ are considered as possible muon candidates, and are extrapolated to the muon system taking into account the magnetic field;
- Standalone Muon reconstruction: all tracks of the segments reconstructed in the muon chambers (performed using segments and hits from Drift Tubes in the barrel region, Cathode Strip Chambers and Resistive Plates Chambers in the endcaps) are used to generate "seeds" consisting of position and direction vectors and an estimate of the muon transverse momentum;
- Global Muon reconstruction: starts from a Standalone reconstructed muon track and extrapolates its trajectory from the innermost muon station through the coil and both calorimeters to the outer tracker surface.

These are illustrated in Fig. 3. The muon identification is given according to the muon reconstruction approach. More details concerning muon identification and reconstruction in CMS may be found in Ref. [2].

2.4 Datasets and triggers

The datasets used in the current analysis are specified in Table 1. These correspond to collision proton-proton data recorded by the CMS experiment, employing the specified real-time selection algorithms (triggers), as well as Monte Carlo simulated data (MC) generated and processed within the CMS software framework (CMSSW). The datasets are available as CMS Open Data [3].

Table 1: Datasets (DS) and triggers used in the analysis.

Resonance	Triggers	DS
J/ψ Data	HLT_Dimuon10_Jpsi_Barrel_v*	[4]
Υ Data	HLT_Dimuon0_Barrel_Upsilon	[4]
J/ψ MC	HLT_Dimuon10_Jpsi_Barrel_v*	[5]
$\Upsilon(1S)$ MC	HLT_Dimuon0_Upsilon	[6]

3 The Tag and Probe method

The tag and probe method (T&P) is a data-driven technique employed for measuring efficiencies. It is based on the decays of known resonances (e.g. J/ψ, Υ and Z) to pairs of the particles being studied.

The determination of a detector efficiency is a critical ingredient in any physics measurement. It accounts for the particles that were produced in the collision but escaped detection (did not reach the detector elements, were missed by the reconstructions algorithms, etc). It can be generally estimated using simulations, but simulations need to be calibrated with data. The T&P method described here provides a useful mechanism for extracting efficiencies directly from data.

In this method, resonances used to calculate the detector efficiencies decay to a pair of particles, which in our case are muons, and which will be labeled as *tag* and *probe*. The decay muons are labeled according to the following criteria:

- Tag muon: well identified, triggered muon (tight selection criteria).
- Probe muon: unbiased set of muon candidates (very loose selection criteria), either passing or failing the criteria for which the efficiency is to be measured.

The efficiency will be given by the fraction of probe muons that pass the criteria. The tag is employed to trigger the

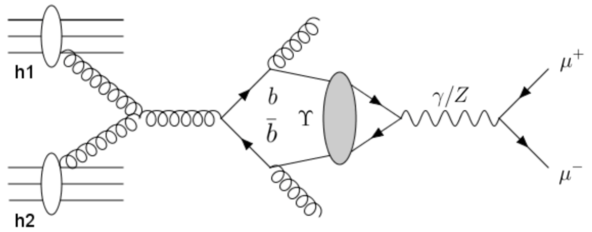


Figure 4: Feynman diagram for $b\bar{b}$ pair production in hadronic collisions followed by formation of meson state (resonance), via the strong force, and eventual decay into a muon pair, via the electroweak force [9].

presence of a resonance decay. Specifically, the invariant mass of the dimuon pair, i.e. the tag and the probe, is formed, and the distribution is employed to extract the signal by rejecting the background present in the data. Further details about the tag and probe method can be found in Ref. [7].

3.1 Resonances

In particle physics, resonances are particles that are extremely short lived and travel tiny distances [8]. Being unstable, they are not directly detected but reconstructed from their decay products. This study focuses on two known resonances decaying in dimuons of opposite charges. The production and decay processes for these particles are represented in Fig. 4.

The J/ψ meson is a hadron formed by a charm quark and a charm antiquark ($c\bar{c}$). It was first detected in 1974 by two groups independently at SLAC (Stanford Linear Accelerator Center) and BNL (Brookhaven National Laboratory) [10]. The Upsilon mesons (Υ) are a family of hadrons formed by a bottom quark and a bottom antiquark ($b\bar{b}$). They were first identified in 1977 at Fermilab, in an experiment where the existence of bottom quarks and bottom antiquark was also discovered [11].

The discovery of these particles marked important milestones in the history of particle physics and in the development of the Standard Model. They are denoted as heavy quarkonia, and provide ideal laboratories in which to study the strong force (QCD). In this work however we explore them as *standard candles*, to calibrate our detector. The physics aspects of these particles (bottomonium) are summarized in Ref. [9].

3.2 Efficiency Definition

The muon efficiency is given by

$$\varepsilon = \frac{\textit{Passing}\ \text{probe muon criteria}}{\textit{All}\ \text{probe muon}} \,. \tag{2}$$

The denominator corresponds to the number of resonance candidates (tag+probe pairs) reconstructed in the dataset. The numerator corresponds to the subset for which the probe passes the (muon identification) criteria.

4 Signal extraction methods

Detector reconstruction efficiencies are calculated using signal muons, that is, only true J/ψ and $Y(1S)$ candidates decaying to dimuons. This is achieved in this study by extracting signal from the data by the usage of two methods: fitting and sideband subtraction.

4.1 Sideband subtraction method

The sideband subtraction method involves choosing sideband and signal regions in invariant mass distribution for each tag+probe pair. In this study, the signal region is selected by finding the highest bin of the invariant mass distribution and defining its position as the resonance invariant mass value (M). The signal region is taken as a 3σ window around the maximum, $M \pm 3\sigma$, where the σ was obtained using full width at half maximum ($FWHM$) which is calculated from the same distribution by selecting the first and last bin that passes a certain high in the invariant mass histogram and getting the interval size between their centers. The relation between σ and $FWHM$ was given by $\sigma = \frac{FWHM}{2.355}$.

While the signal region contains both signal and background, the sideband region is chosen such as to have only background. The background region selected in this study is at least 4σ distant from resonance peak. These regions are shown in Figs. 5a and 5b, for the J/ψ and Υ datasets.

For each event category (passing criteria and passing+failing criteria) and for a given variable of interest (as probe p_T, η or ϕ), two distributions are obtained: one for signal region and other for background regions. To obtain the desired variable distribution only for signal, this method subtracts background distribution from the signal+background distribution, as:

$$N_{signal} = N_{signal\ region} - \alpha \times N_{sideband\ region} \,, \tag{3}$$

where the normalization factor α expresses the amount of background particles present in the signal region and is given by

$$\alpha = \frac{\text{yield of background in signal region}}{\text{yield of background in sideband region}} \tag{4}$$

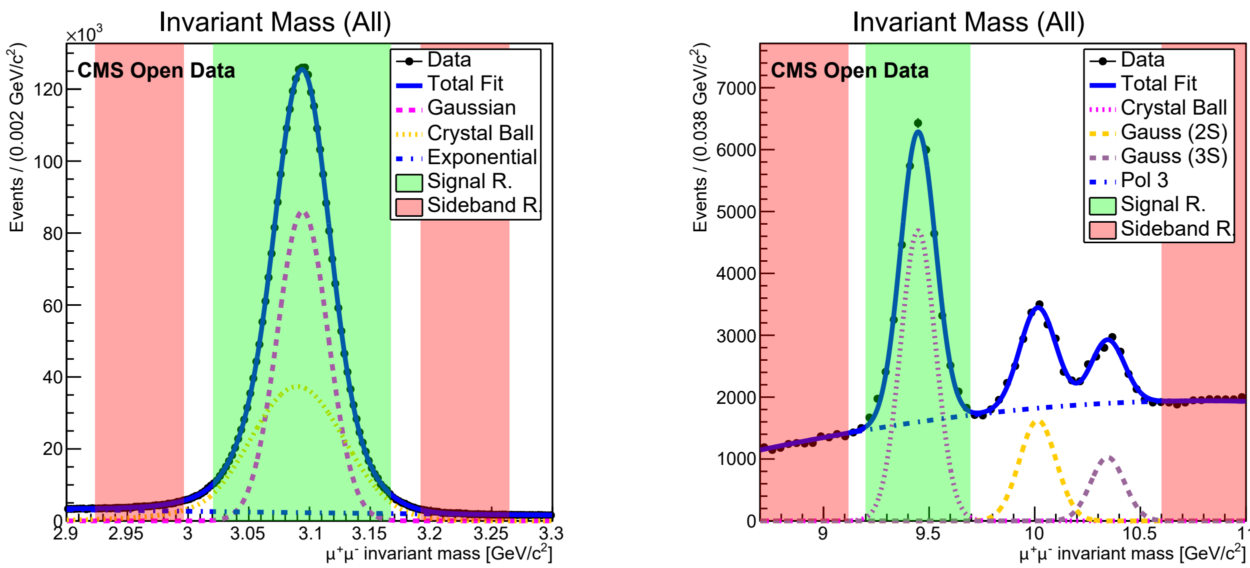
For the uncertainty of each bin, this method uses

$$\alpha_{signal} = \sqrt{N_{signal\ region} + N_{sideband\ region}} \tag{5}$$

Applying the sideband procedure to the probe muon p_T, as an example, one obtained the distributions shown in Fig. 6. The procedure starts from the total distribution (solid blue line) of all muon candidates in the signal region, and the corresponding distribution for candidates in the sideband region (dashed blue line); subtracting the latter from the former (Eq. 3), the signal distribution (solid magenta line) is finally obtained.

4.2 Fitting method

In this method, the signal is extracted not by histogram manipulation but by likelihood fitting. The procedure is



(a) J/ψ meson.

(b) Υ(nS) meson family.

Figure 5: Sideband (red) and signal (green) regions in the dimuon spectrum near the studied resonance states.

applied after splitting the data in sub-samples, corresponding to bins of the kinematic variable of interest (p_T, η, ϕ) of the probe objects. As such, the efficiency will be measured as a function of that variable. Each sub-sample contains signal and background events; the signal is accessed by fitting the invariant mass spectra.

The fit for each bin allows to statistically discriminate between signal and background. In particular, the fit yields the number of signal events. The J/ψ invariant mass distribution is described by the sum of a Crystal Ball function and a Gaussian function, while the background is modeled with an exponential function. For Υ, the $\Upsilon(1S)$ signal is modeled by a Crystal Ball function (first peak), and Gaussian functions are used for the two other resonances (second and third peaks), while a Chebyshev function parameterizes the background.

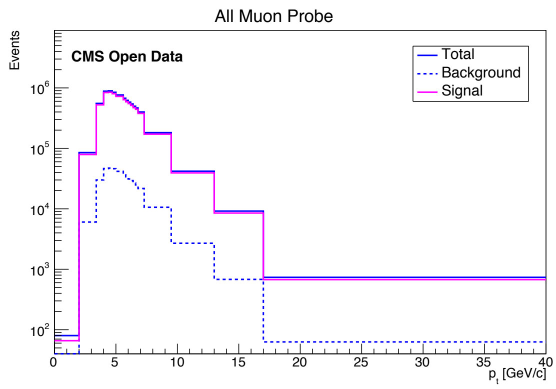


Figure 6: Distributions of the transverse momentum p_T of the probe muons, obtained with the sideband subtraction method applied to the J/ψ meson.

The efficiency is finally obtained by simply forming the ratio of the signal yield from the fit to the *passing* category by the signal yield from the fit of the inclusive *all* category. This approach is illustrated in Fig. 7.

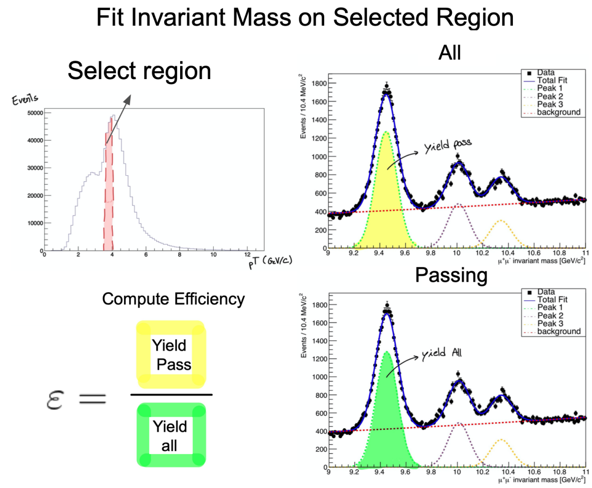


Figure 7: Illustration of the fitting method.

5 Efficiency measurements

The Tracker Muon identification efficiency is measured as function of the muon transverse momentum p_T, pseudorapidity η, and azimuthal angle ϕ. The results obtained with the J/ψ resonance are presented in Fig. 8, while Fig. 9 shows the same efficiencies obtained with Υ resonance.

The efficiencies measured for a same detector, physics object, and selection criteria should be expected to match. Muons originating from the decay of from different particles on the other hand have different kinematic distributions and are embedded in different backgrounds. These differential studies allow to probe and quantify such effects.

The results obtained with J/ψ and Υ are fairly consistent. The uncertainties in the measurements, displayed by the error bars in the plots, differ noticeably. Such is expected due to the fact that the J/ψ dataset is about six times larger than the Υ sample.

The figures display in addition a comparison of the results obtained with the two methods, that have been implemented for extracting the signals from data. These provide cross-checks and may be used to quantify possible systematic uncertainties associated to the method. In general, while the sideband subtraction method is more readily executed and less computational intensive, the likelihood fitting procedure achieves a more complete characterization of the data.

In Fig.10 shows the comparison between real data and Monte Carlo with the J/ψ resonance using the muon standalone identification, while in Fig.11 is presented same distributions for muon global identification. In the both figures there are some agreement between real data and Monte Carlo simulation for low p_T regions (4.0 - 7.0 GeV) (Fig.10a and Fig.11a). Also in Figs 10c and 11c shows a good agreement between them.

Additional preliminary results for other muon identification algorithms not here presented are also provided [12]. These allow further comparisons between algorithms, methods, resonances, and simulation. The tools developed in this work are shared [13, 14], and further documented in the accompanying tutorial [15].

6 Conclusions

We have devised and implemented a data-driven procedure that allows to evaluate the CMS detector efficiency for identifying muons. Our tag and probe implementation uses two quarkonium states, J/ψ and $\Upsilon(1S)$, that decay to pairs of muons. Two methods have been devised for extracting these resonance signals from the data, namely sideband subtraction and likelihood fitting. These procedures have been applied to proton-proton collision data collected by CMS, which are openly available. We measured the efficiency for several muon identification algorithms, and the obtained results extracted from data have been compared to simulation. The complementary methods and resonances explored have allowed for systematic studies of both the physics and method performance evaluations. While the obtained results are in general consistent, the likelihood method is found to offer increased robustness.

The developed implementation has been documented [15] and made available for users. Further work will result in the integration of the code and tools into the central CMS Open Data analysis framework.

Acknowledgements

Thanks to LIP Internship Program 2020 that allowed this work to be carried out and published. Thanks to Eliza Melo, Nuno Leonardo and Sandro Fonseca for guiding us on this particle physics analysis. Also thanks to all those who supported this work.

Part of the results obtained in this work were presented at the conference FÍSICA 2020 (SPF, Portugal). Following the internship, the group was also allowed to prepare and facilitate a workshop tutorial (FNAL, US) on the tools developed in this work [15]. We thank the organizers of these events for the opportunity to further present our work.

References

[1] CMS Collaboration, Journal of Instrumentation **3**, S08004 (2008)

[2] CMS Collaboration, Journal of Instrumentation **7**, P10002 (2012)

[3] *LHC Open Data*, `opendata.cern.ch`

[4] CMS Collaboration (2016), *MuOnia primary dataset from RunA of 2011*, http://opendata.cern.ch/record/27

[5] CMS Collaboration (2016), *Simulated dataset for* J/ψ, http://opendata.cern.ch/record/1335

[6] CMS Collaboration (2016), *Simulated dataset for* Υ*(1S)*, http://opendata.cern.ch/record/1522

[7] R.G. de Souza, *Tag And Probe 2019*, https://twiki.cern.ch/twiki/bin/view/Sandbox/TestTopic11111203

[8] C. Dudley, *What is a Resonance Particle?*, https://webhome.phy.duke.edu/~kolena/modern/dudley.html

[9] Z. Hu, N.T. Leonardo, T. Liu, M. Haytmyradov, International Journal of Modern Physics A **32**, 1730015 (2017)

[10] *The 1976 nobel prize in physics*, `https://www.nobelprize.org/prizes/physics/1976/press-release/`

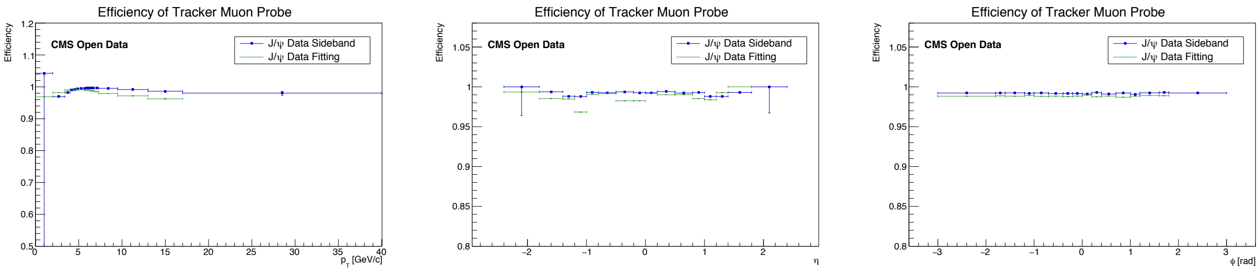
[11] L. Lederman, *The discovery of the Upsilon, bottom quark, and B mesons*, in *3rd International Symposium on the History of Particle Physics: The Rise of the Standard Model* (1992), pp. 101–113

[12] *Github: CMS Open Data efficiency studies*, `https://github.com/allanjales/Open_data_Extra_studies`

[13] A. da Silva Jales, *Github: sideband subtraction method*, `https://github.com/allanjales/tag-probe`

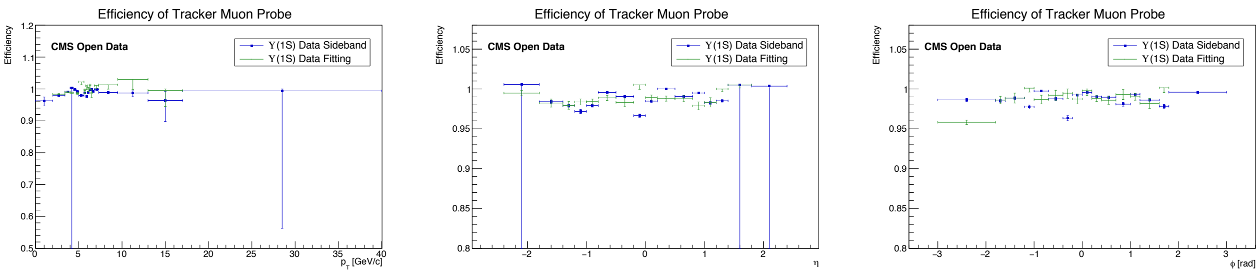
[14] T.F.D. Gaehtgens, *Github: fitting method*, `https://github.com/AthomsG/LIP_INTERNSHIP`

[15] *Efficiency studies using the tag and probe method*, `https://cms-opendata-workshop.github.io/workshop-lesson-tagandprobe/`



(a) p_T for J/ψ (b) η for J/ψ (c) ϕ for J/ψ

Figure 8: Efficiency of Tracker muons probe in J/ψ sample by sideband subtraction and fitting method.

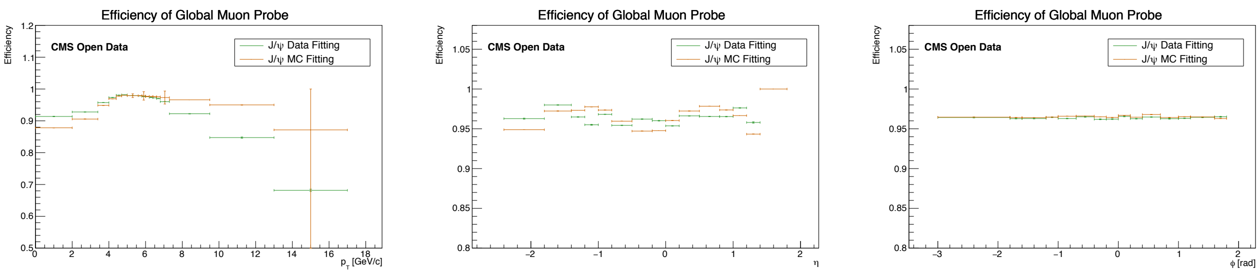


(a) p_T for Υ (b) η for Υ (c) ϕ for Υ

Figure 9: Efficiency of Tracker muons probe in Υ sample by sideband subtraction and fitting method.

(a) p_T for J/ψ (b) η for J/ψ (c) ϕ for J/ψ

Figure 10: Efficiency comparing between data and MC using the Standalone muon identification for J/ψ by fitting method.



(a) p_T for J/ψ (b) η for J/ψ (c) ϕ for J/ψ

Figure 11: Efficiency comparing between data and MC using the Global muon identification for J/ψ by fitting method.