EUROPEAN ORGANIZATION FOR NUCLEAR RESEARCH

October 6, 2023

Summer Student Report

Study of prompt and non prompt quarkonium production at forward rapidity with Standard and Machine Learning techniques

Author:
Lukáš Viceník
lukas.vicenik@cern.ch

Supervisors:
Luca Micheletti
Martino Gagliardi

Abstract

Charmonia, bound states of charm (c) and anticharm ($\bar{c}$), represent an important tool to understand quantum chromodynamics (QCD). In particular, the J/ψ meson production has been investigated by many LHC experiments. The ALICE experiment has been upgraded recently, aiming to improve spatial resolutions at mid and forward rapidity. As a part of the upgrade, the Muon Forward Tracker (MFT), a new detector at forward rapidity, was installed with the goal of improving the vertexing resolution of ALICE's muon spectrometer. As a consequence it is now possible to separate the production cross section into the prompt (produced at the primary vertex) and non-prompt (produced in the decay of B hadrons) components for this vector meson.
The goal of this project is to determine the non-prompt J/ψ decay fraction using the standard template fit technique. In addition, an exploratory study of the possible Machine Learning (ML) application for this study will be discussed.

Keywords: ALICE, Charmonium, Run 3, J/ψ, Simultaneous fit, Machine learning, Gradient boosted decision tree, Prompt, Non prompt, Invariant mass, Pseudo proper decay time, Muon Forward Tracker

Contents

1 Introduction

The J/ψ meson is a bound state of charm (c) and anticharm ($\bar{c}$) quarks and it can be produced via three different processes: 1) direct production, 2) decay of heavier states ($\psi(2S)$, χ_c, ...), 3) decay of long-lived beauty flavored hadrons (b-hadrons). The first two contribute to the J/ψ prompt fraction, the third one represents the non-prompt fraction [1]. The figure 1 visualizes the process.

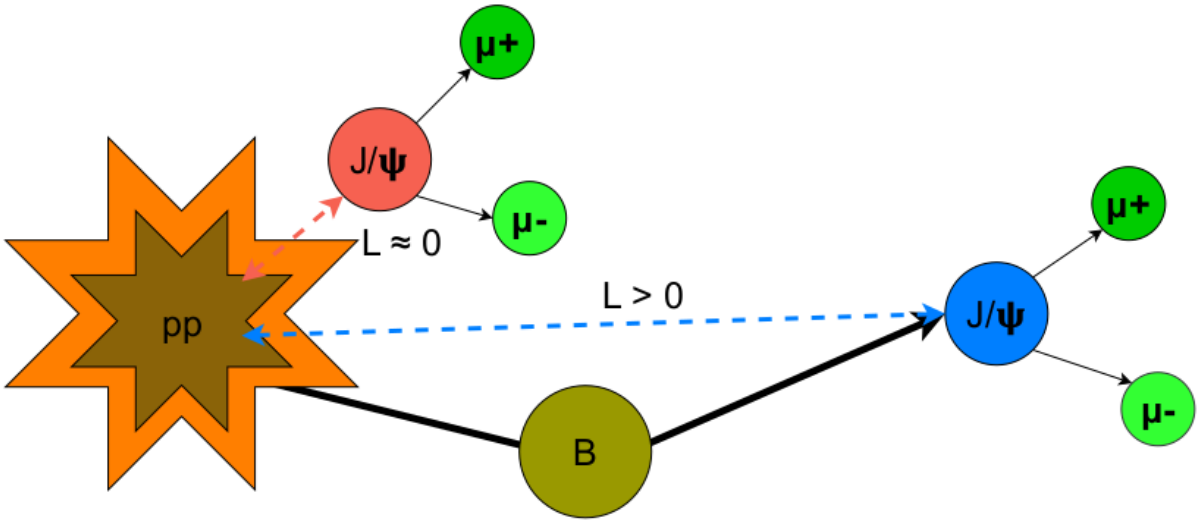

Figure 1: The prompt and non prompt production of J/ψ particle.

For both prompt and non-prompt fraction invariant mass observable has the same shape, whereas for the pseudo-proper decay time the shapes are significantly different. The latter are used to determine the non-prompt fraction.

First, let us properly define the used variables. The invariant mass is defined as:

$$m^2 = E^2 - p^2, \tag{1}$$

where E corresponds to the total energy and p to total momentum. The second one is a quantity called pseudo-proper decay time (τ_z), which is defined as:

$$\tau_z = \frac{L_z \cdot m}{p_z \cdot c}, \tag{2}$$

where L_z is the distance between the primary and secondary vertex [2].

To better understand the shape of these observables the normalized distributions are shown in Fig. 2. It is important to notice that the both invariant mass distributions have compatible shape, whereas the τ_z non-prompt distribution is significantly more asymmetric for non-prompt than for the prompt distribution.

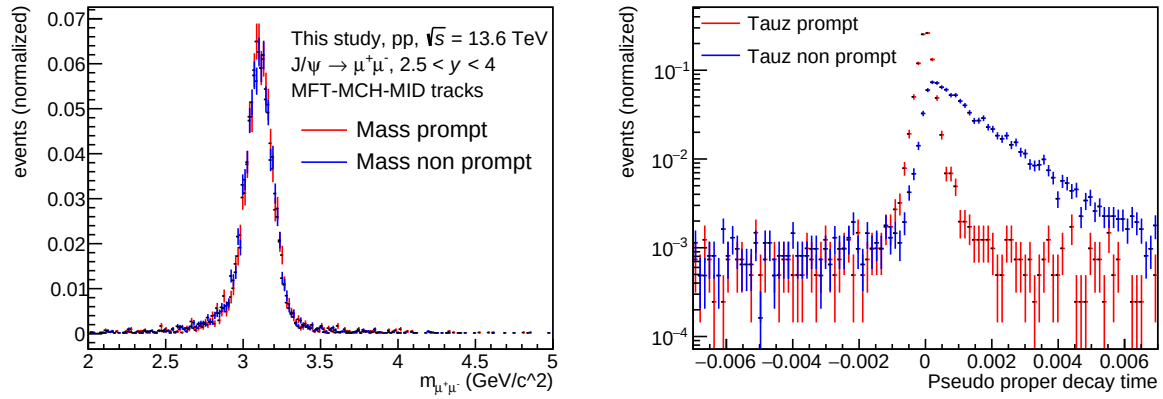

Figure 2: Left: Comparison of invariant mass Monte Carlo shapes for prompt and non-prompt distribution. Right: Comparison of τ_z Monte Carlo shapes for prompt and non-prompt distributions.

The goal of this project is to evaluate the J/ψ non-prompt fraction (f_b) and optimize the selection of signal candidates using two different approaches:

1. Template simultaneous fit: performing a template fit to the mass and τ_z distributions the f_b is evaluated.

2. Machine learning: exploiting a gradient-boosted decision tree algorithm we try to find features with the highest separation power and use them for the classification signal and background candidates.

2 Alice detector – RUN 3 upgrade

Hard, penetrating probes, such as heavy quarkonium states, provide an essential tool to study the early and hot stage of heavy-ion collisions. In particular they are expected to be sensitive to Quark-Gluon Plasma formation. In the presence of a deconfined medium (Quark Gluon Plasma) with high enough energy density, quarkonium states are dissociated because of color screening. This leads to a suppression of their production rates. At the high LHC collision energy, both the charmonium states (J/ψ and ψ') as well as the bottomonium states (Υ, Υ' and Υ'') have been studied [3]. The visual model of the ALICE detector after the upgrade for run 3 is shown in figure 3

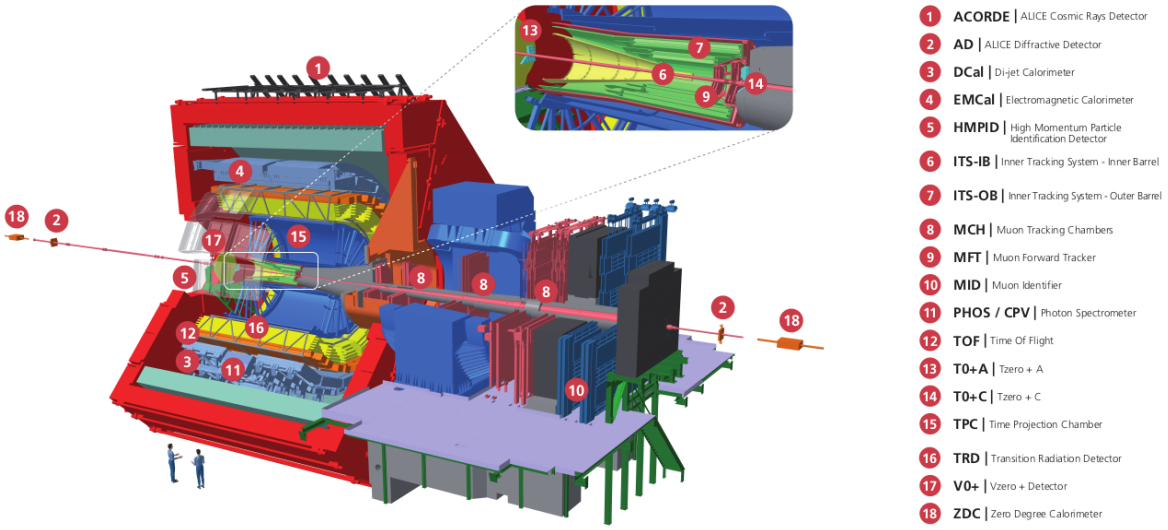

Figure 3: The ALICE detector detailed visualization after the RUN 3 upgrade. Taken from [4].

The Muon Spectrometer is located at forward rapidity, covering the pseudorapidity range $-4 < \eta < 2.5$, and has the capability to detect particles down to $p_T = 0$.

The four main components forming the Muon Spectrometer before the upgrade were:

- absorber: a 4 m long carbon and concrete structure that suppresses all particles coming from the interaction point (and from subsequent decays) but muons. It's the main contributor of multiple scattering;
- tracking system: 10 cathode pad/strip chambers arranged in 5 stations of 2 chambers each. The radiation length is less than 3% per chamber in order to reduce the scattering of muons and reach a spatial resolution of 100 mm;
- trigger system: 4 planes of resistive plate chambers behind a muon filter provide the transverse momentum of each muon, with the objective of selecting heavy quark decays;
- dipole magnet: one of the biggest warm dipoles in the world, can generate a nominal magnetic field up to B = 0.7 T and field integral along beam axis of 3 Tm.

Figure 4 shows a schematic view of the Muon Spectrometer after the proposed upgrades, which includes the upgrade of the trigger system to the Muon Identifier (MID). The Muon Forward Tracker is also represented.

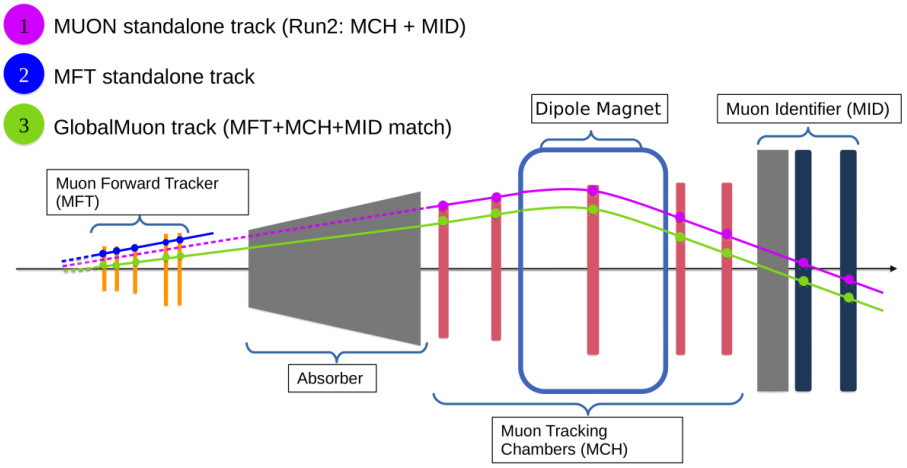

Figure 4: The absorber, Tracking chamber, Magnet, Filter, and Trigger chambers schematic representation visualizing the whole functionality. It also shows how particle tracks are reconstructed. Taken from [4].

The Muon Forward Tracker (MFT) is a new tracker detector of ALICE project installed during the Long Shutdown 2 (LS 2) as a part of the experiment upgrade for the third round of data taking (RUN 3). It was installed in a frontal geometry near the interaction point, complementing the Muon Spectrometer. MFT aims to increase the resolution on vertex determination of the particle trajectories observed by the Muon Tracking Chambers (MCH), allowing, for instance, to separate prompt and non-prompt components of J/ψ peaks. Figure 5 shows the MFT. More information about the run 3 upgrade can be acquired from [4].

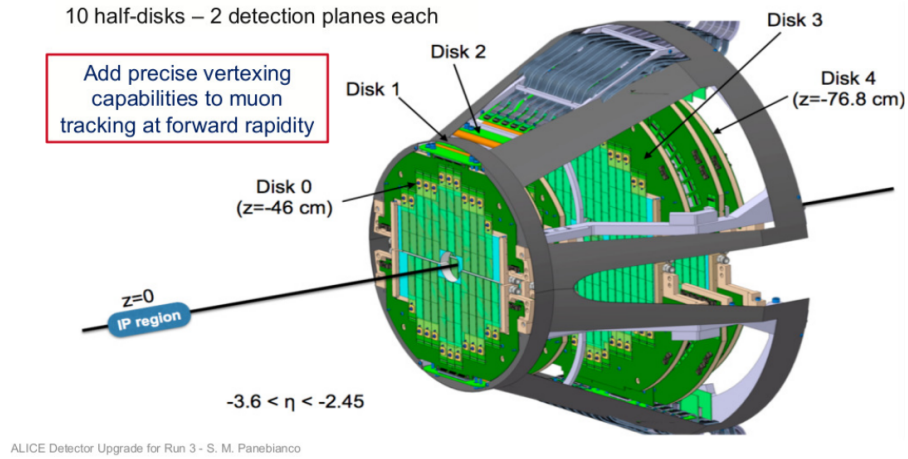

Figure 5: Schematic visualization of MFT. MFT counts on 936 silicon pixel sensors on 240 ladders of 2 to 5 sensors each. Taken from [4].

3 Used tools

3.1 Root language

ROOT is an open-source framework built for analyzing high-energy physics data at CERN. Data from Monte Carlo simulations are stored in files constructed by this framework and are called N-tuples. The N-tuples are root files built similarly to algorithmic structures called trees where the leaves of the tree correspond to the observables measured in the detector. More details can be found in [5].

3.2 Roofit

Roofit is a ROOT library providing tools for modeling the expected distributions of events in a physics analysis. Models can be used to perform unbinned maximum likelihood fits, create plots, and generate "toy Monte Carlo" samples for various studies. More details can be found in [6]

3.3 Hipe4ml

Hipe4ml is a machine learning environment for heavy ion physics used by ALICE experiment which provides easy access to important hyperparameters and produces graphs for model performance visualization [7].

4 Template simultaneous Fitting

In order to evaluate properly the non-prompt fraction (f_b) the simultaneous fit of the mass and the pseudo-proper decay time is performed. The reason behind this choice is related to the correlation among the fitting parameters, in particular:

- the invariant mass distribution allows to separate signal and background, since the J/ψ peak is significant and can be distinguished from the background. Therefore this fit allows to constrain the J/ψ raw yield.
- the pseudo-proper decay time allows to separate prompt and non-prompt J/ψ due to the difference in shape among these two components (see right panel of Fig. 4). Therefore this fit allows to constrain the non-prompt fraction (f_b).

The following formulas relate properties mentioned before:

$$N_p = N_{J/\psi} \cdot (1 - f_b) \tag{3}$$

$$N_n = N_{J/\psi} \cdot f_b, \tag{4}$$

where N_p and N_n correspond to number of prompt and non prompt particles. $N_{J/\psi}$ is number of J/ψ and f_b is the non-prompt fraction [1].

4.1 Template construction

The simultaneous fit template has different components that can be either directly histograms or fitted PDFs. Fig. 6 shows the PDF fit to the τ_z prompt distribution with a combination of a Breit-Wigner and a Gaussian function, while Fig. 7 shows the fit to the non prompt distribution with a superposition of a Landau function and Chebyshev polynomials.

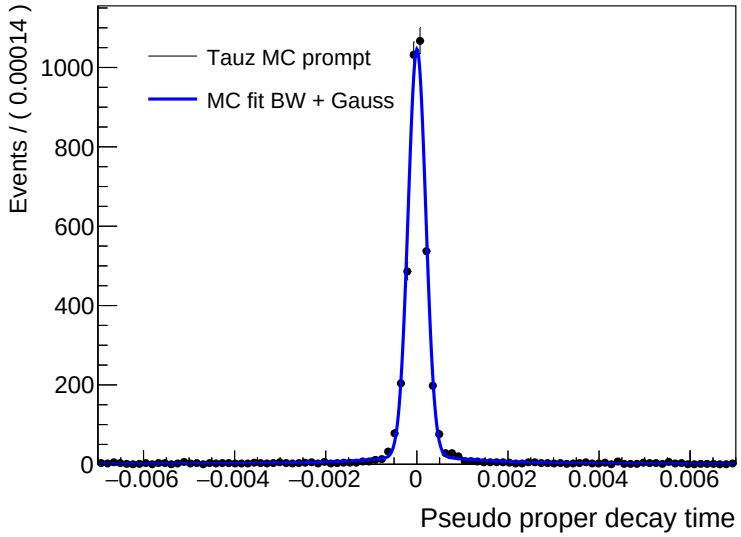

Figure 6: Fit to the τ_z prompt distribution with Breit-Wigner and Gaussian for $4 < p_T < 6$ GeV.

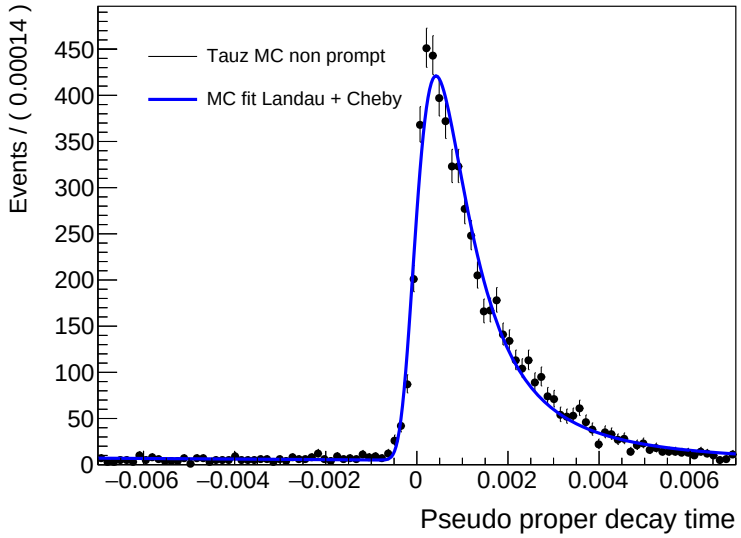

Figure 7: Fit to the τ_z non prompt distribution with Landau and Chebyshev polynomials for $4 < p_T < 6$ GeV.

4.2 Monte Carlo validation

Before performing the simultaneous fit it is necessary to validate the whole procedure with a toy Monte Carlo and for this purpose the following steps are performed:

1. Fit the prompt and non prompt distributions with fixed non-prompt fraction
2. Generate a toy MC sample from the distributions with a given non-prompt fraction
3. Fit the toy data with the model used for the generation of the sample, keeping free the non-prompt fraction

4. Compare the non-prompt fraction from the fit with the one used to generate the toy data sample

The result of this procedure is shown in Fig. 8. It should be noted that only for the first bin the value of f_b is exceeding the statistical uncertainty, while the overall agreement is very good.

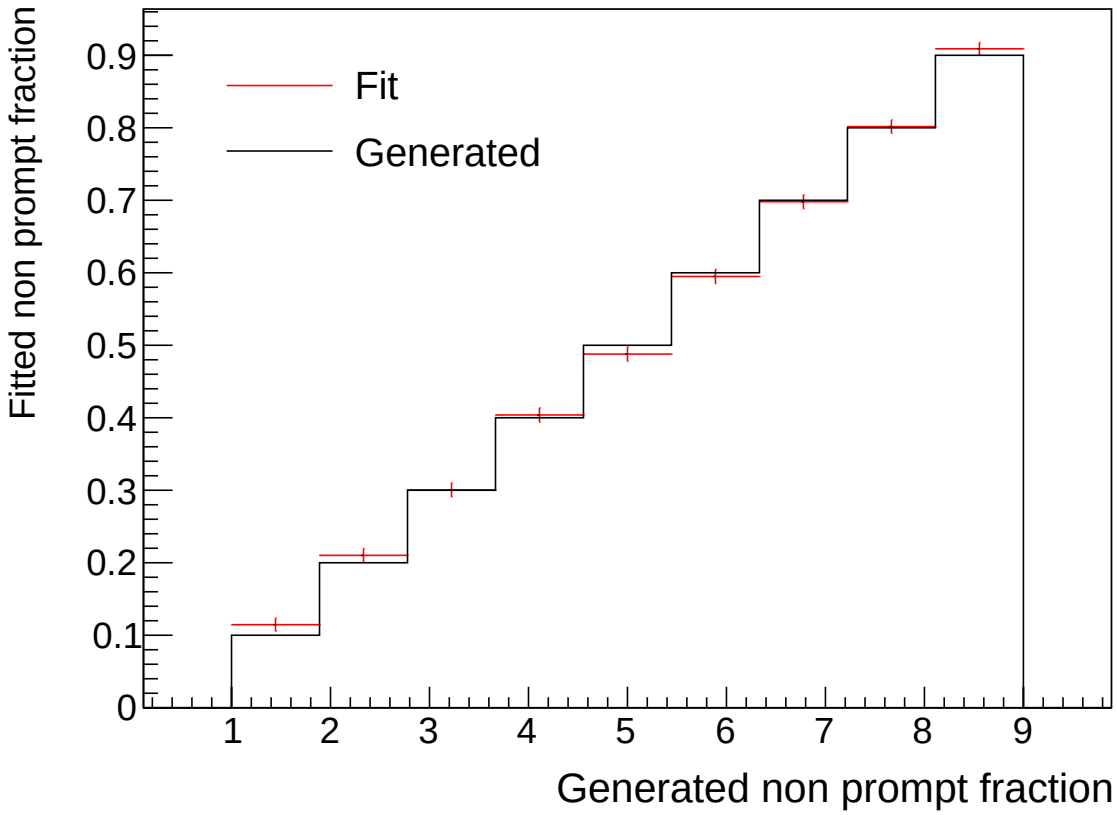

Figure 8: Comparison of expected and fitted non prompt fraction.

4.3 Fit with the background from side-bands

At this point it is possible to perform the simultaneous fit. The invariant mass distribution is fitted with two PDFs, the Crystal ball function for the signal peak and the Chebeyshev polynomials for the background. For τ_z the fit is composed of several histogram components. For prompt and non-prompt fractions, the Monte Carlo is used. For the background approximation, it is needed to extract it from the data with the following steps:

1. Search for the events that are far enough from the signal peak in the invariant mass spectrum.
2. Select events in the τ_z spectrum that correspond to the selection.
3. Use the extracted distribution as the background component in the template.

Afer this procedure the background template is extracted and it is possible to fit the data combining the various components. An example of the fit for the 2 to 4 GeV region is shown in the figure 9, while the non-prompt fractions after the fit are shown in the table 1 as a function of p_T. It should be noted that

the result is very unstable and far from the expectations from previous measurements, where f_b ranges around 10-30%, increasing with p_T. The remaining plots are show in the appendix A.1 - A.3

Table 1: Non prompt fractions with side-band background

p_T region	Non prompt [%]
[0,2]	32.1
[2,4]	22.6
[4,6]	0.0
[6, 10]	0.0

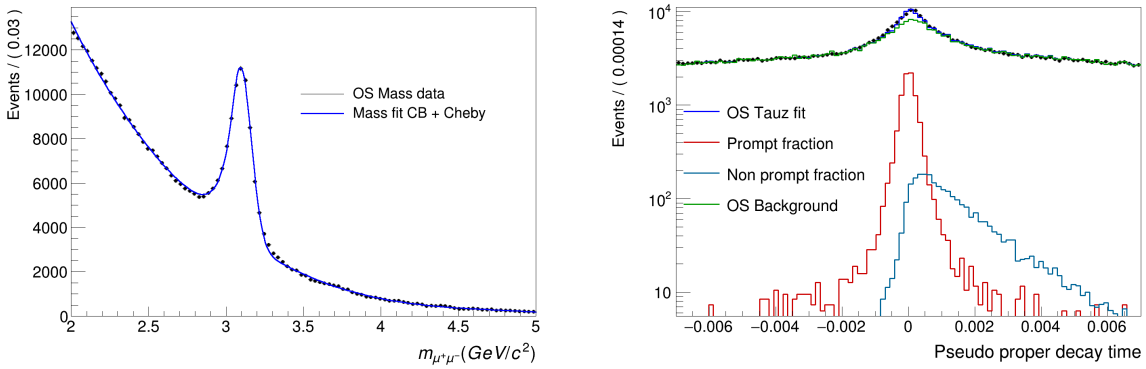

Figure 9: A basic simultaneous fit to two observables. Left: A fit to the invariant mass spectrum. Right: A fit to the τ_z spectrum. Both for $2 < p_T < 4$ GeV.

4.4 Like sign background

In order to check the stability of the result, the background shape is extracted using like-sign distribution. By definition the signal could not be included in this sample, nevertheless this distribution can estimate properly the background shape. The latter is evaluated as the geometrical average:

$$LS_{bkg} = 2 \cdot \sqrt{PP \cdot MM}, \tag{5}$$

where PP and MM are dimuon candidates build with only positive and negative charges respectively. In Fig. 10 it is shown a comparison between the different backgrounds for the transverse momentum region between 4 and 6 GeV. The rest of the p_T regions are shown in the appendix A.4 - A.6. If all the regions are compared it can be observed that the left side of the spectrum is very similar. On the other hand, a mild discrepancy for the right side grows with higher p_T.

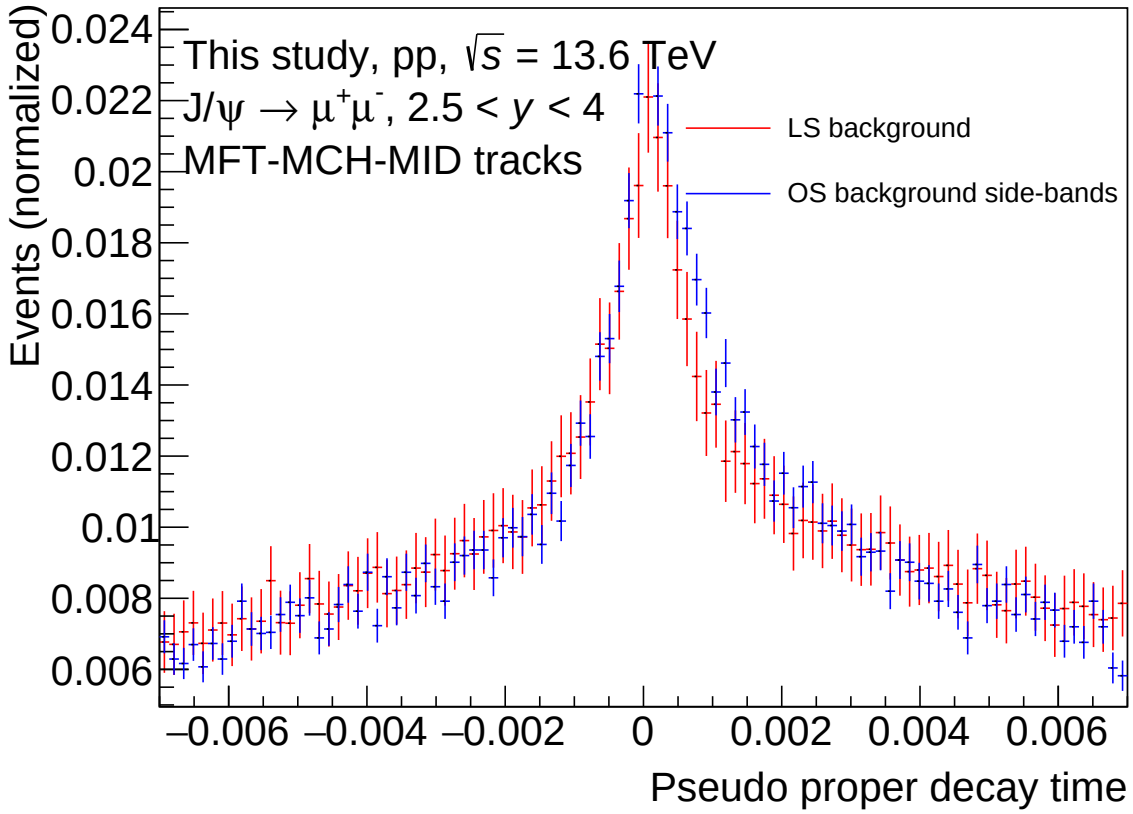

Figure 10: The like sign and opposite sign background comparison for $4 < p_T < 6$ GeV.

After the implementation of this background, non-prompt fraction stabilized between 40% to 55% for all the transverse momentum regions. The table 2 shows the detailed result.

Table 2: Non prompt fractions with side-band background

p_T region	Non prompt [%]
[0,2]	54.0
[2,4]	50.1
[4,6]	49.0
[6, 10]	39.7

4.5 Prompt distribution extraction from the data

Another important component of the template to be checked is the τ_z distribution for prompt J/ψ. For this purpose we performed a comparison of this distribution in data and Monte Carlo. In order to extract the prompt fraction from the data, the following procedure is followed.

1. Compute the opposite sign signal according to equation

$$OS_{sig} = OS_{data} - LS_{bkg}, \tag{6}$$

where OS_{sig} corresponds to opposite sign background, OS_{data} corresponds to opposite sign dataset and LS_{bkg} is the like sign background computed in previous section.

2. The opposite sign now contains just the prompt fraction, non-prompt fraction, and the residual background which is supposed to be negligible.

3. The left side of the spectrum does not contain the non-prompt distribution, and the residual background can be assumed to be symmetrical.

4. The left side of the spectrum can be therefore mirrored to obtain an approximation of prompt fraction from the data.

The mirroring visualization is shown in Figure 11. At this point, we can compare the extracted τ_z prompt distribution with the Monte Carlo τ_z prompt distribution. The rest of the p_T regions are shown in the appendix A.10 - A.12. We can see a wider peak for lower p_T regions.

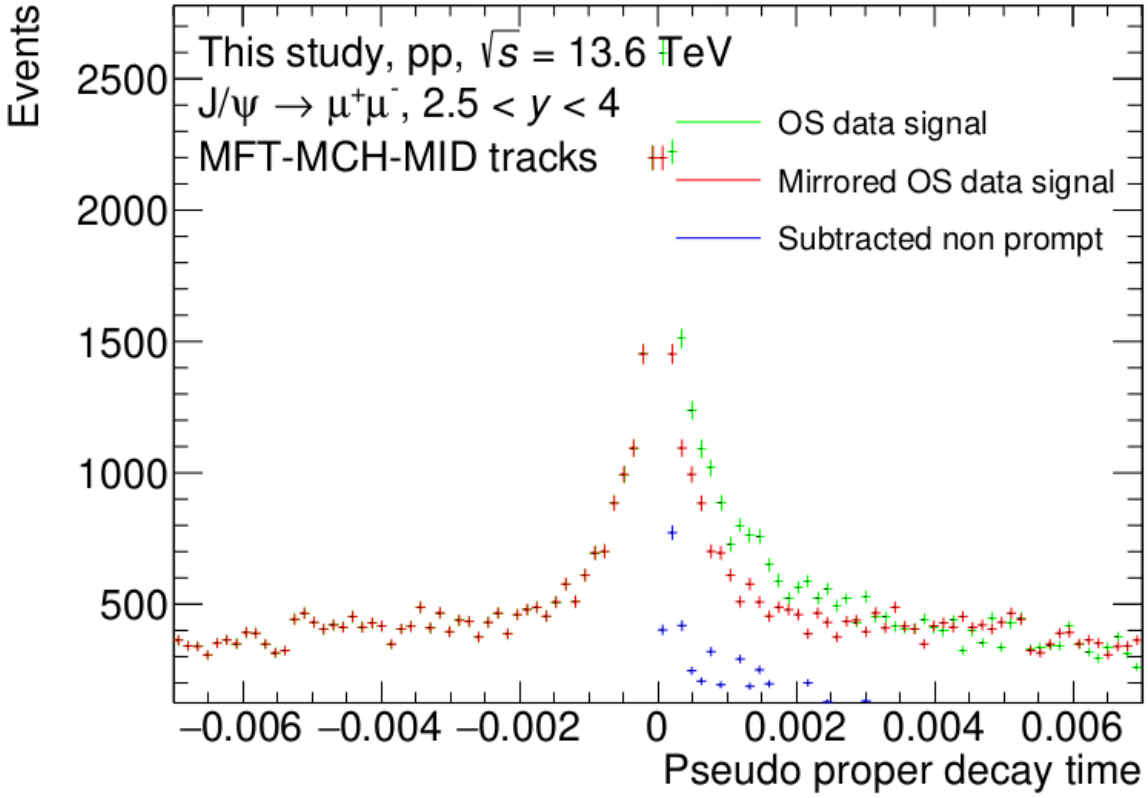

Figure 11: The green histogram corresponds to the original OS signal. The red histogram is the mirrored left side of the prompt distribution. The blue histogram is extracted non prompt fraction. Shown for or $2 < p_T < 4$ GeV.

4.6 Artificial smearing

When the comparison is performed, it is visible that differences in the σ parameter are significant, especially for lower $p_T s$. Table 3 shows the differences for all the $p_T s$ regions.

Table 3: Difference between MC and data extracted distribution before smearing

p_T region	σ difference [%]
[0,2]	1.0540694340018975e-03
[2,4]	9.804238209381367e-04
[4,6]	3.930027233032485e-04
[6, 10]	1.2194102447741221e-04

The smearing is applied with different intensities for every bin. It was tried to apply smearing with the gaussian distribution, but the results were not optimal. Then the Breit-Wiegner distribution was used, which resulted in a more natural shape of the smeared distribution. In Fig. 12 difference in the shape between the distributions is shown. It is important to point out that the extracted distribution still contains a residual background and its effect should be tested in the future to achieve better reliability of the smearing procedure. The rest of the p_T regions are shown in the appendix A.7 - A.9. We can see that for lower p_T ranges smearing needs to be more significant to obtain comparable widths.

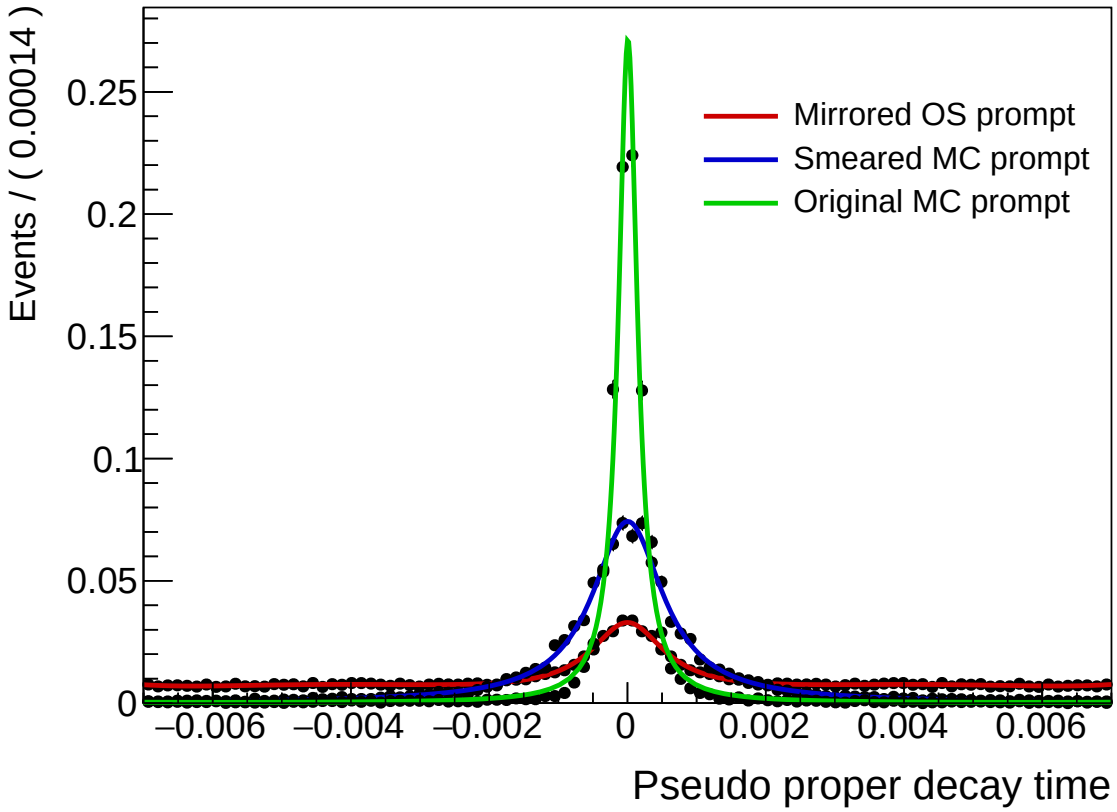

Figure 12: Comparison of the prompt distribution from the data (Red), the MC prompt distribution (Green) and the smeared MC prompt distribution (Blue) for $4 < p_T < 6$ GeV.

The adjusted sigmas are shown in the table 4.

Table 4: Difference between MC and data extracted distribution after smearing

p_T region	σ difference [%]
[0,2]	2.600438607131038e-06
[2,4]	-5.67925580409411e-05
[4,6]	-2.8974447170654317e-05
[6, 10]	3.129966814133803e-05

4.7 Final simultaneous fit

At the beginning of this section let us summarize what we have done so far

- the Like sign background template is extracted from like sign dimuons
- the τ_z prompt distribution was extracted from the data with the use of mirroring.
- the original Monte Carlo prompt distribution was smeared in such a way its σ parameter corresponded with the extracted distribution

At this point, we are ready to perform the simultaneous fit again with the listed improvements. Figure 13 shows the comparison of the original simultaneous fit and the fit after the application of the three mentioned procedures. Results for all the p_T regions are shown in appendix A.13 - A.16.

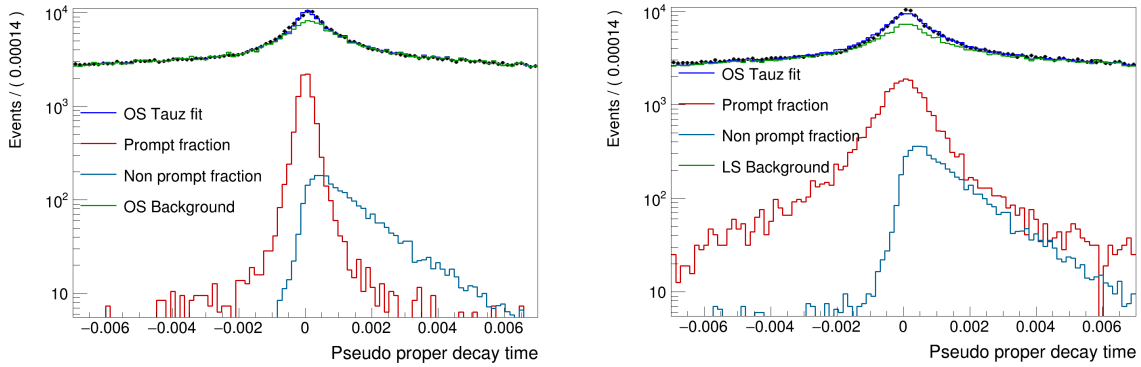

Figure 13: Left: Original simultaneous fit to invariant mass and τ_z with OS background from the side bands. Right: Simultaneous fit with LS background and smeared MC prompt τ_z distribution. Both for $2 < p_T < 4$ GeV.

The final prompt fractions are shown in table 5. The range is approximately $11.6\% - 31.9\%$ and the tendency is increasing, which is something that corresponds with our expectations. The results are visualized in figure 14 where the tendency is clearly visible.

Table 5: The non prompt fractions after the application of improving procedures.

p_T region	σ difference [%]
[0,2]	11.6
[2,4]	16.8
[4,6]	26.5
[6, 10]	31.9

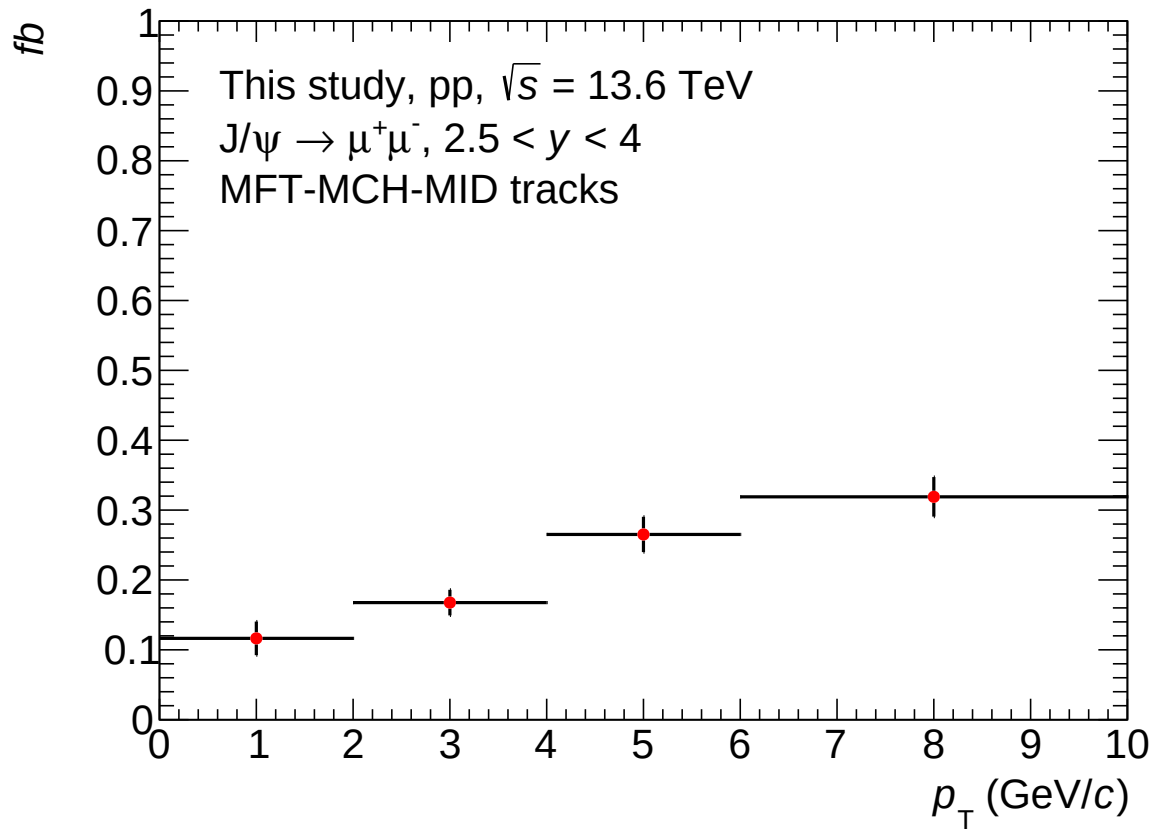

Figure 14: Final non prompt fraction for every p_T bin from 0 to 10 GeV.

5 Machine learning

For this project, which involves the improvement of the selection of signal candidates, we decided to use algorithms for the gradient-boosted decision tree (GBDT) family. GBDTs are usually the preferred type of machine learning algorithm regarding small-size to middle-size tabular data.

5.1 Decision trees

A decision tree is a structure in which each internal node represents a test on an attribute, each branch represents the outcome of the test, and each leaf node represents a class label. The paths from root to leaf represent classification rules.

5.2 Boosting

Boosting is an ensemble learning method where we sequentially fit multiple weak learners. Each model in the sequence is fitted, giving more importance to observations in the dataset that were badly handled by the previous models in the sequence. Each new model therefore focuses its efforts on the most difficult observations to fit up to now, so that we obtain a better learner with lower bias. The models are fit and added to the ensemble sequentially such that the second model attempts to correct the predictions of the first model, the third corrects the second model, and so on. More details can be found in [8].

5.3 Gradient boosting

Models are fit using any arbitrary differentiable loss function and gradient descent optimization algorithm. This gives the technique its name, as the loss gradient is minimized as the model is fit, much like a neural network. Let's take a look at how gradient boosting works in more detail.

We are given some data $(x_i, y_i)_{i=1}^n$ where x is vector of input variables and y is output variable. Then we need some differentiable loss function. Types of loss functions differ according to the task we want to perform. For explanatory purposes, let us use the general term

$$\mathscr{L}(y, F(x)), \tag{7}$$

where y is observed and $F(x)$ is a function equivalent to predicted value. Then we want to minimize it as

$$F_0(x) = \arg\min_p \sum_{i=1}^{n} \mathscr{L}(y, p), \tag{8}$$

where $F_0(x)$ is initial model value and p corresponds to predicted value. To minimize the loss function, we compute

$$r_{im} = -\left[\frac{\partial \mathscr{L}(y_i, F(x_i))}{\partial F(x_i)}\right]_{F(x)=F_{m-1}(x)} \text{for } i = 1, ..., n, \tag{9}$$

where r_{im} is the pseudo-residual after derivative for sample i and tree m that we are building. This step corresponds approximately to subtracting of predicted value from the observed value. We can also point out that this step is the gradient that Gradient Boost is named after.
Now we have to fit CART (Classification and Regression Trees) to pseudo-residuals and label terminal regions, R_{jm} where j is a number of regions that are leaves of the tree. To determine the output values, we compute,

$$p_{jm} = \arg\min_p \sum_{x_i \in R_{ij}} \mathscr{L}(y_i, F_{m-1}(x_i) + p), \tag{10}$$

which is a similar equation to 8. The difference is that the previous prediction is now considered. To compute the output value for each leaf, we pick only y values that correspond to this region. Then we are looking for p that minimizes the equation.
The final step is to make a new prediction, as

$$F_m(x) = F_{m-1}(x) + \nu \sum_{j=1}^{J_m} p_{jm} I(x \in R_{jm}), \tag{11}$$

where ν is a learning rate. This equation means that we take the original prediction and add values of corresponding leaves from our newly built tree multiplied by the learning rate. Then, we continue with the next iteration. It is desirable to mention that this step represents the boosting method described in [8]. More information can be found in [9].

5.4 XGBoost

XGboost is a scalable end-to-end tree boosting system that allows to achieve state-of-the-art results on many machine learning tasks. Novel techniques used are sparse-aware algorithm for sparse data and weighted quantile sketch for approximate tree learning. More insight into cache access patterns is achieved to build a scalable tree-boosting system. Full explanation in [9].

5.5 Machine learning model performance analysis

For machine learning tasks, a tabular dataset is composed. Rows of this dataset correspond to individual events and the columns are represented by the chosen observables. To choose a correct subset of observables that has the most significant impact on the GBDT performance, feature importance is computed. Two of the most important features are fTauxy and fTauz. The importance of the first one is mostly based on background separation that can be artificial due to our MC. The second feature core is based on the difference between prompt and non prompt fractions and therefore seems to be correct. An example of the result is shown in figure 15 for p_T region from 2 to 4 GeV.

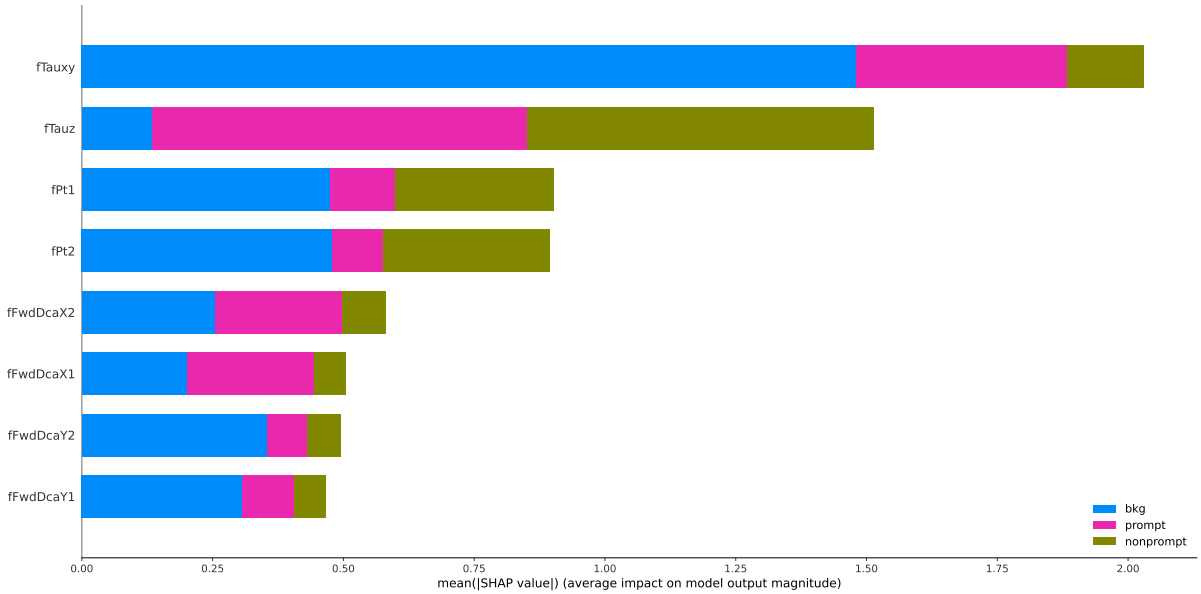

Figure 15: Example of the feature importance graph plotted for $2 < p_T < 4$ GeV region.

To verify the model performance, the so-called ROC curve [9] is evaluated. It shows the performance of a classification model at all classification thresholds. The higher the integral under the ROC curve, the higher the score. The figure 16 shows the results for p_T region from 2 to 4 GeV.

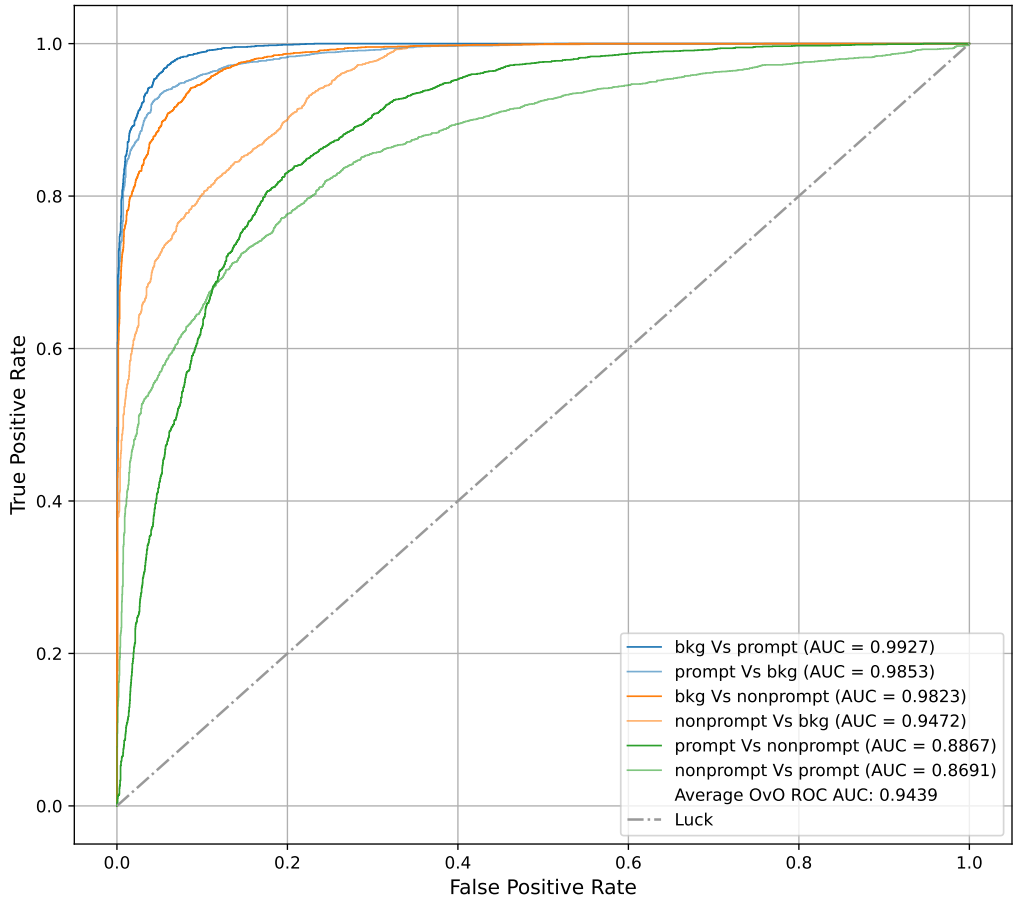

Figure 16: Particular curves correspond to separation power between two chosen components for for $2 < p_T < 4$ GeV.

5.6 GBDT output

After the training procedure, the output of the classifier is visualized. Our goal is to separate the chosen components. Figure 17 shows a separation of the MC signal and the background from the data. The higher value of the background is therefore in the vicinity of 1 and the MC signal is close to 0.

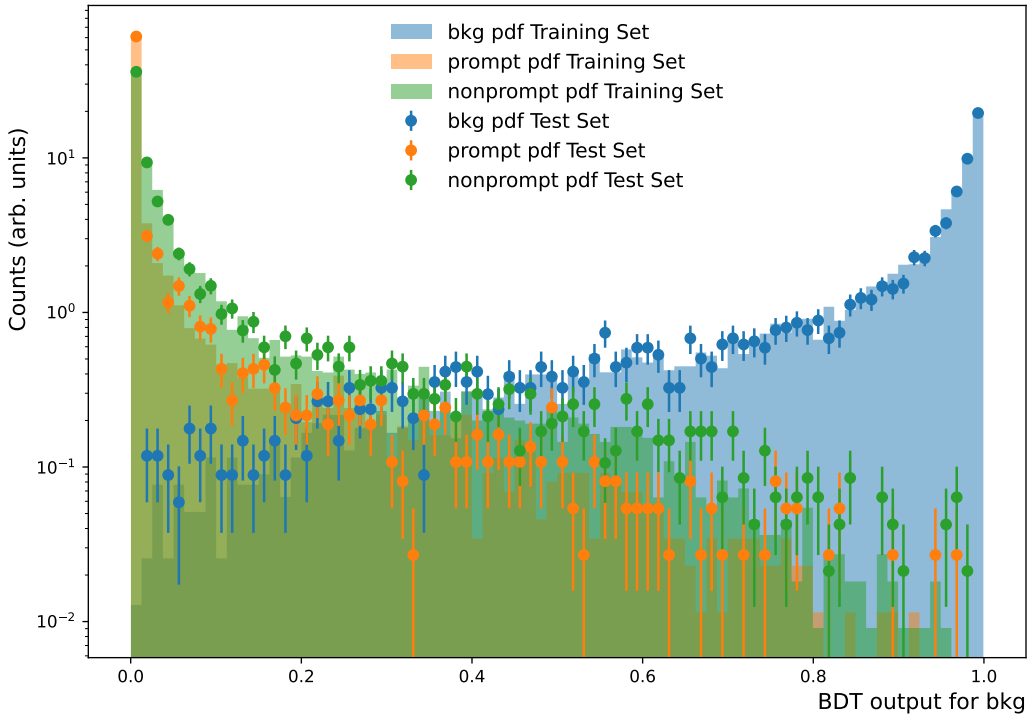

Figure 17: Separation of MC signal (prompt and non prompt) and the background from the data.

5.7 Cut on output

After the separation is made, the cut on the output is applied. By the cut, we want to separate as much background from the signal but still keep as much signal as possible. In our case, we chose the $0.0 - 0.9$ for the cut to achieve the best results. Figure 18 shows the Original data, data after the separation of the background and the separated background. This particular cut was performed for one combined $0 < p_T < 10$ GeV region.

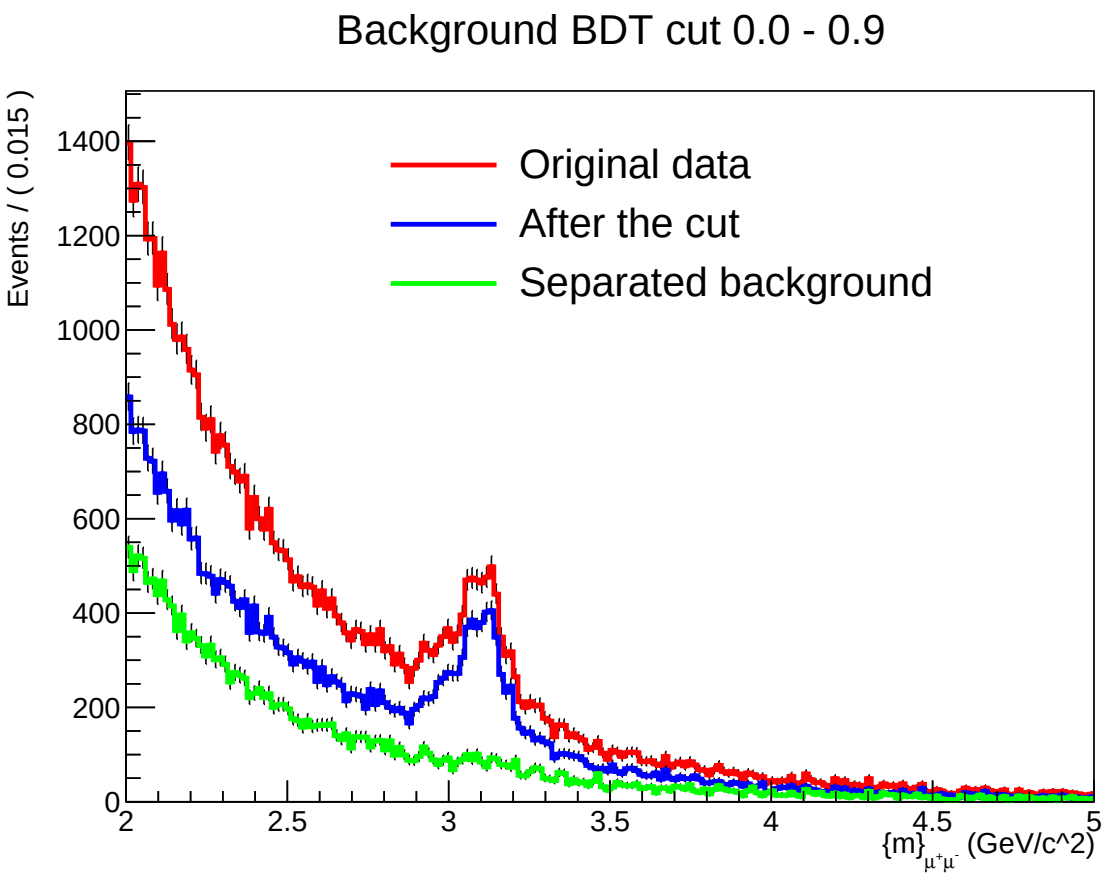

Figure 18: The red distribution corresponds to the original invariant mass data. The blue histogram is the Invariant mass data after subtraction of the background. The green histogram is the background separated by the cut.

6 Conclusion and Outlook

This summer student project has been separated into to parts. The first one is related to the template simultaneous fit. With this approach, the following techniques were performed to evaluate the J/ψ non-prompt fraction.

- Extraction of the LS combinatorial background from the data.
- Extraction of the τ_z prompt fraction from the data.
- Smearing of the original MC τ_z prompt fraction to achieve comparable distribution width.

With this approach the non-prompt fraction is extracted as a function of the transverse momentum. Even if the measurement requires further corrections, it is in reasonable agreement with published results. The second part describes the machine learning technique used to improve the signal to background separation for J/ψ:

- The performance of the model was verified by the computation of feature importance and ROC curves.

- The data was separated by the classification into two subsets based on demand.
- A cut was applied to the data to separate the background from the signal.

In this case the result show how powerful machine learning techniques are, even if this results is for the moment limited by the remainign differences among data and simulation.

References

[1] **Suraj Prasad, Neelkamal Mallick, Raghunath Sahoo** Collaboration, "Inclusive, prompt and non-prompt J/ identification in proton-proton collisions at the Large Hadron Collider using machine learning", *arxiv* (2023) .

[2] **Tamar Zakareishvili and on behalf of the ATLAS collaboration** Collaboration, "Journal of Physics: Conference Series Paper • The following article is Open access Measurement of the production cross-section of J/ and (2S) mesons at high transverse momentum in pp collisions at $\sqrt{s} = 13\,\text{TeV}$ with the ATLAS detector", *J. Phys.: Conf. Ser. 1690 012160* (2020) .

[3] **ALICE** Collaboration, *ALICE dimuon forward spectrometer: Technical Design Report*. Technical design report. ALICE. CERN, Geneva, 1999. `https://alice-collaboration.web.cern.ch/menu_proj_items/Muon-Spect`.

[4] L. N. Lopes, "Forward track reconstruction and quality assessment for the ALICE experiment of the Large Hadron Collider", 2022. `http://hdl.handle.net/10183/252032`. Presented 2022-11-30T04:53:59Z.

[5] "ROOT Data analysis Framework." `https://root.cern/`.

[6] "RooFit." `https://root.cern/manual/roofit/`.

[7] **ALICE** Collaboration, "hipe4ml",. `https://doi.org/10.5281/zenodo.5070131`.

[8] "A Gentle Introduction to Ensemble Learning Algorithms." `https://machinelearningmastery.com/tour-of-ensemble-learning-algorithms/`.

[9] L. Vicenik, "Machine Learning for the Leptoquark Search Using CERN ATLAS Data", 2022. `https://cds.cern.ch/record/2812370`. Presented 14 Jun 2022.

A Appendix

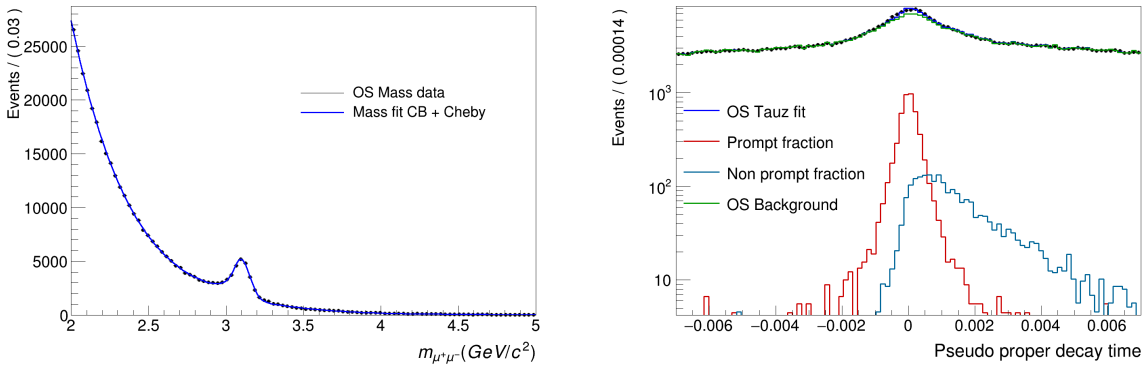

Figure A.1: A basic simultaneous fit to two observables. Left: A fit to the invariant mass spectrum. Right: A fit to the τ_z spectrum. For $0 < p_T < 2$ GeV.

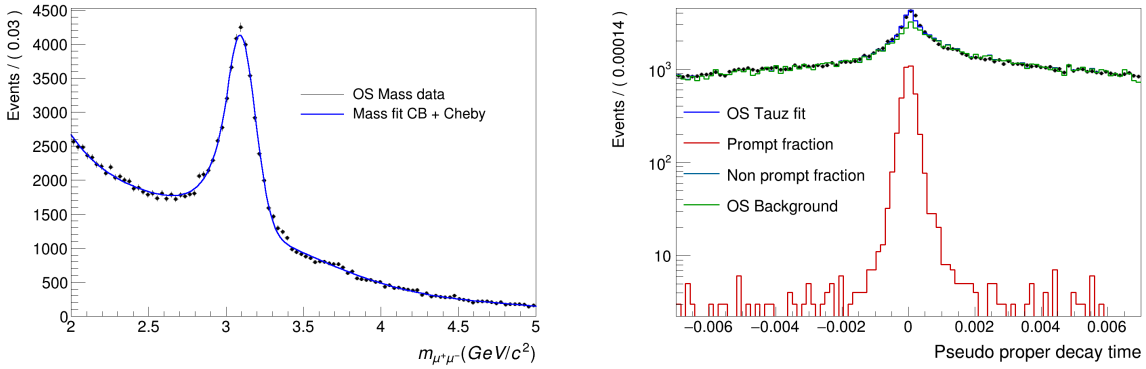

Figure A.2: A basic simultaneous fit to two observables. Left: A fit to the invariant mass spectrum. Right: A fit to the τ_z spectrum. For $4 < p_T < 6$ GeV.

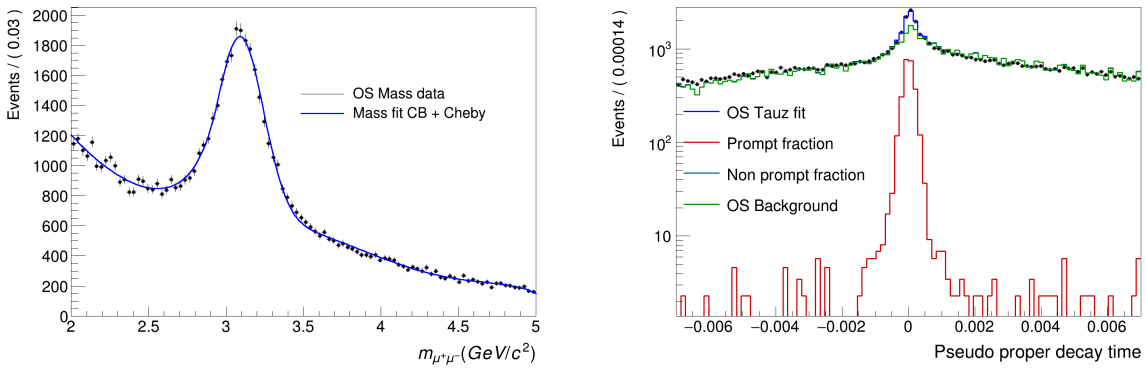

Figure A.3: A basic simultaneous fit to two observables. Left: A fit to the invariant mass spectrum. Right: A fit to the τ_z spectrum. For $6 < p_T < 10$ GeV.

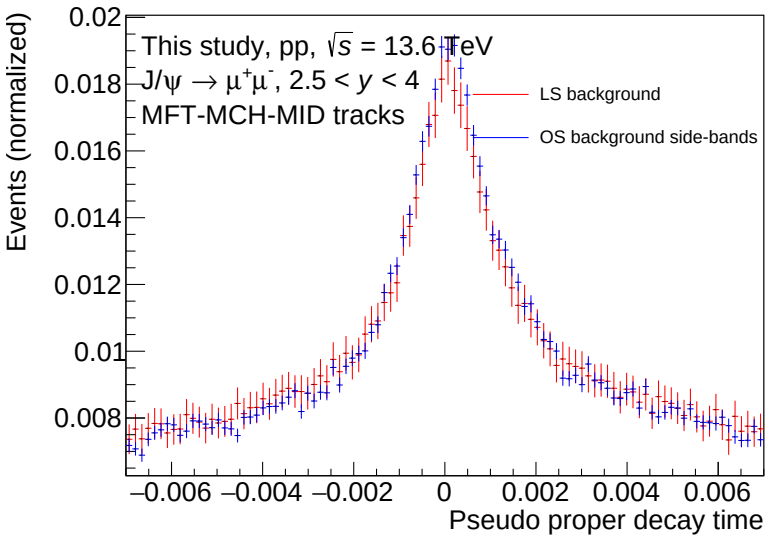

Figure A.4: The like sign and opposite sign background comparison for $0 < p_T < 2$ GeV.

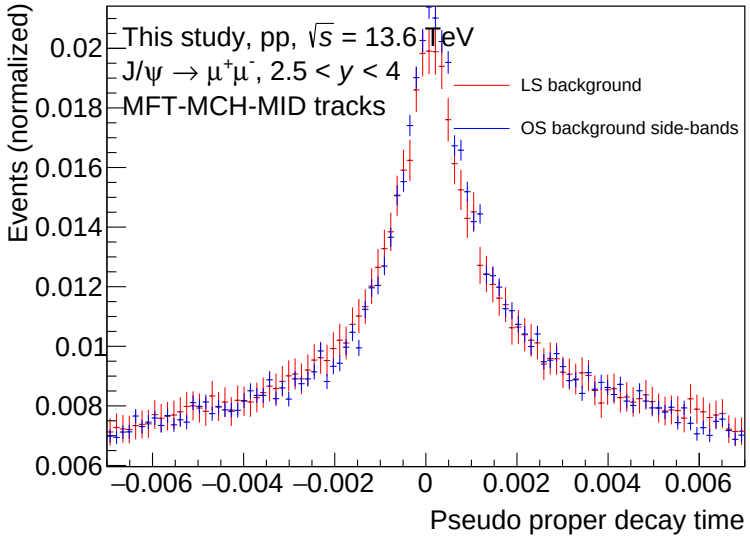

Figure A.5: The like sign and opposite sign background comparison for $2 < p_T < 4$ GeV.

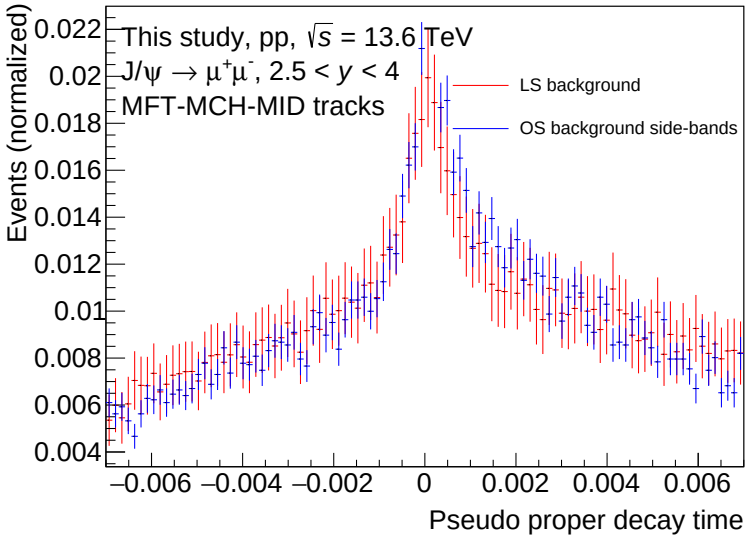

Figure A.6: The like sign and opposite sign background comparison for $6 < p_T < 10$ GeV.

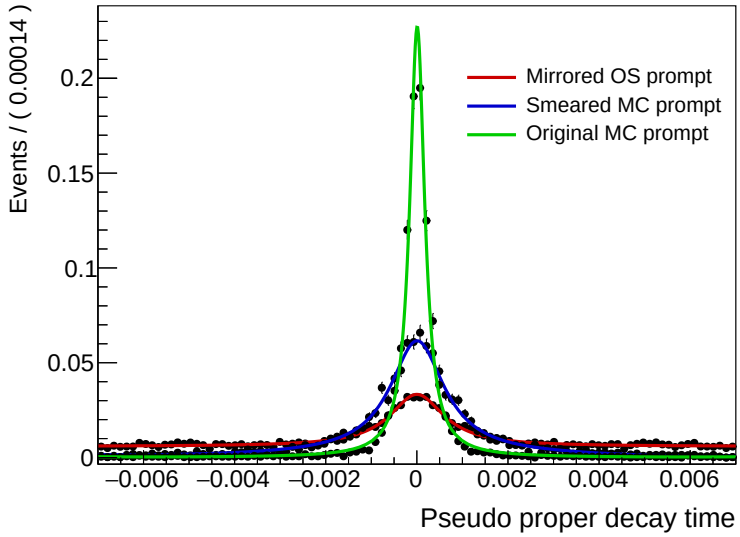

Figure A.7: Comparison of the prompt distribution from the data (Red), the MC prompt distribution (Green) and the smeared MC prompt distribution (Blue) for $0 < p_T < 2$ GeV.

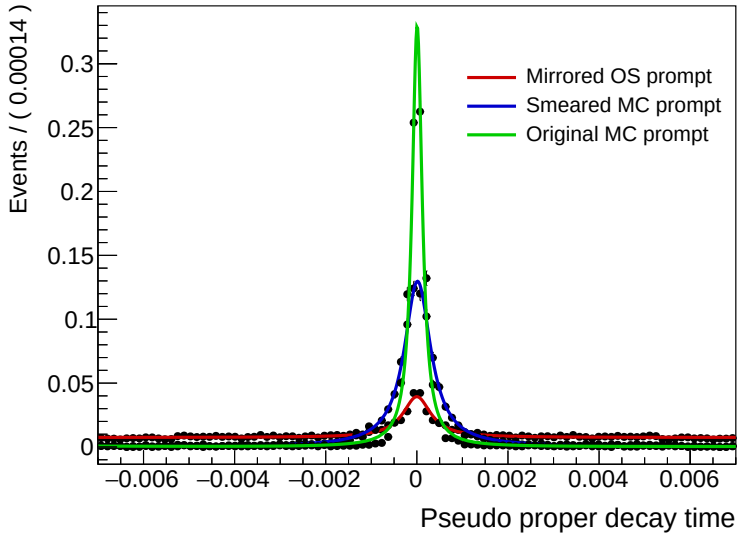

Figure A.8: Comparison of the prompt distribution from the data (Red), the MC prompt distribution (Green) and the smeared MC prompt distribution (Blue) for $4 < p_T < 6$ GeV.

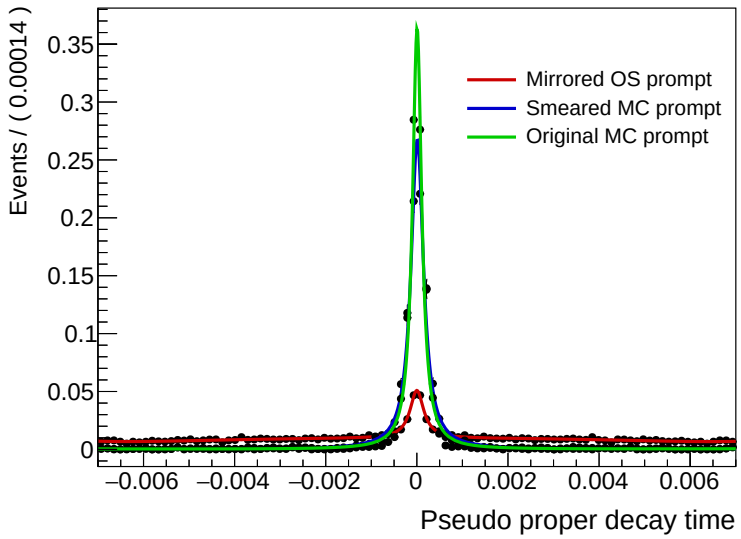

Figure A.9: Comparison of the prompt distribution from the data (Red), the MC prompt distribution (Green) and the smeared MC prompt distribution (Blue) for $6 < p_T < 10$ GeV.

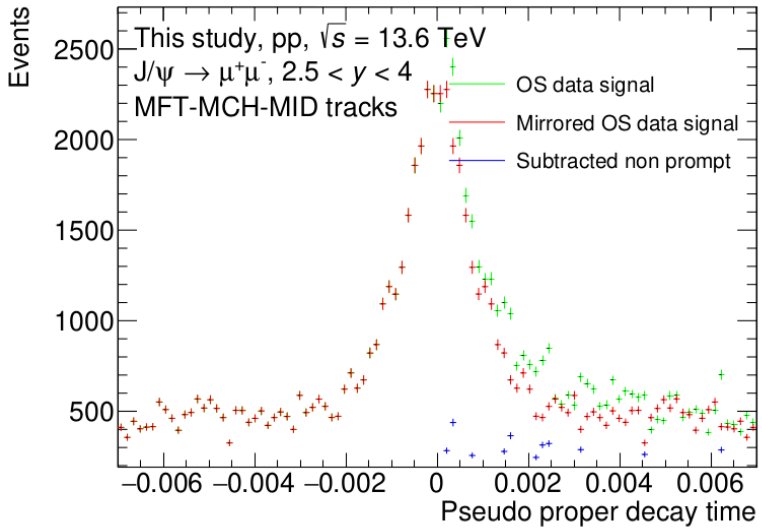

Figure A.10: The green histogram corresponds to the original OS signal. The red histogram is the mirrored left side of the prompt distribution. The blue histogram is extracted non prompt fraction. Shown for $0 < p_T < 2$ GeV.

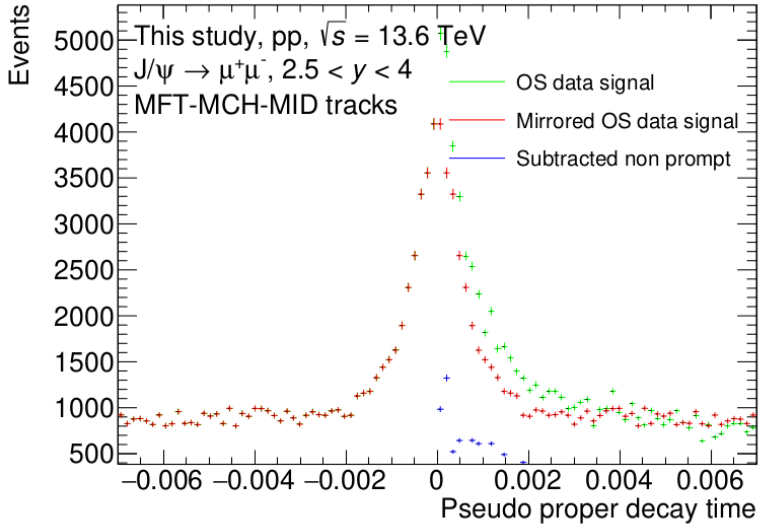

Figure A.11: The green histogram corresponds to the original OS signal. The red histogram is the mirrored left side of the prompt distribution. The blue histogram is extracted non prompt fraction. Shown for $2 < p_T < 4$ GeV.

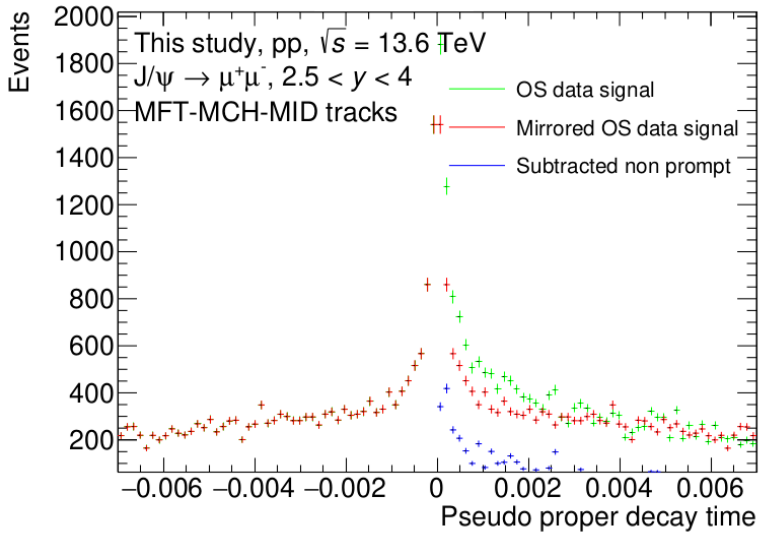

Figure A.12: The green histogram corresponds to the original OS signal. The red histogram is the mirrored left side of the prompt distribution. The blue histogram is extracted non prompt fraction. Shown for $6 < p_T < 10$ GeV.

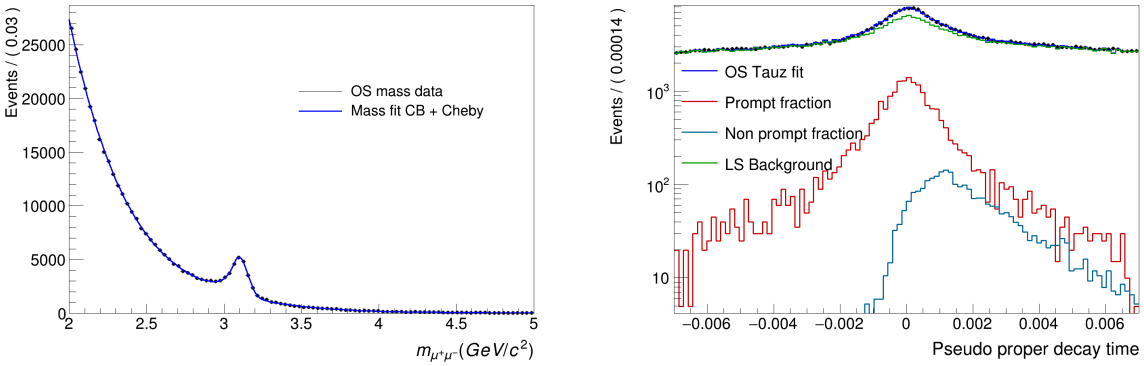

Figure A.13: Left: Fit to the invariant mass Right: Fit to the τ_z with LS background and smeared MC prompt distribution. For $0 < p_T < 2$ GeV.

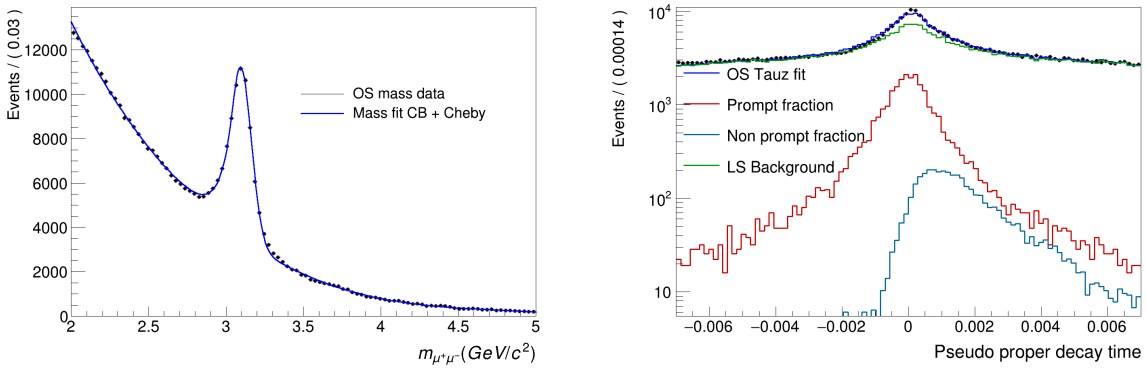

Figure A.14: Left: Fit to the invariant mass Right: Fit to the τ_z with LS background and smeared MC prompt distribution. For $2 < p_T < 4$ GeV.

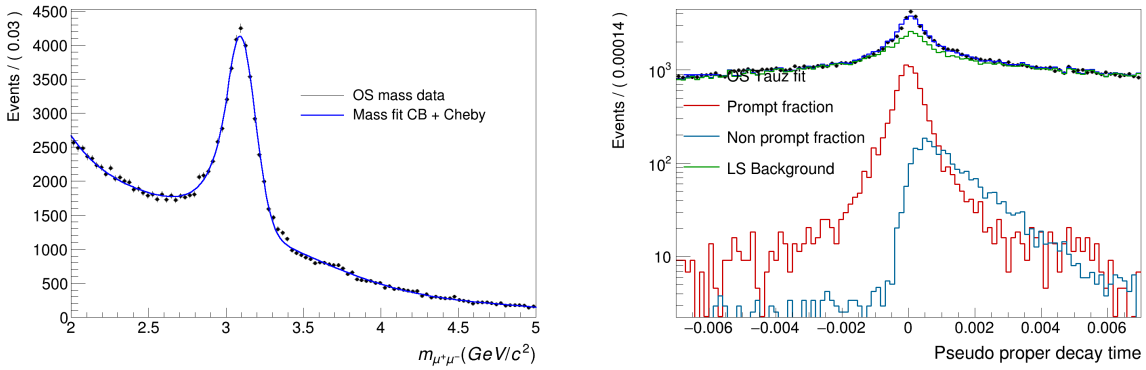

Figure A.15: Left: Fit to the invariant mass Right: Fit to the τ_z with LS background and smeared MC prompt distribution. For $4 < p_T < 6$ GeV.

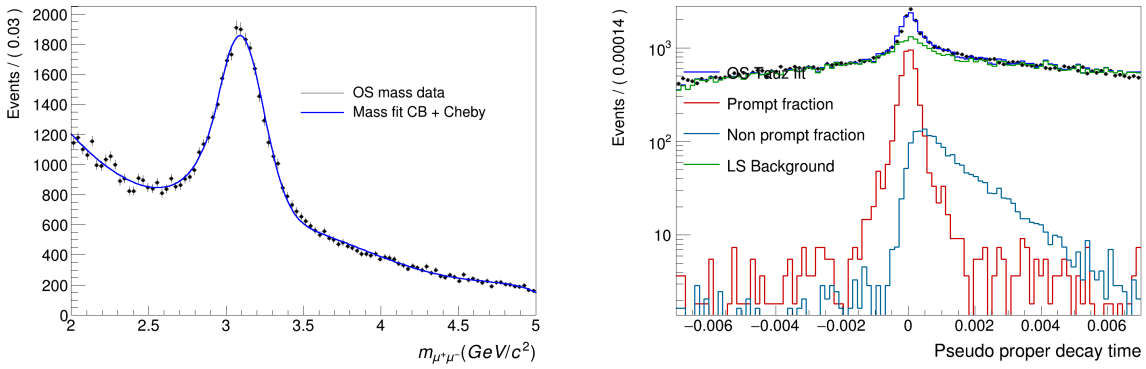

Figure A.16: Left: Fit to the invariant mass Right: Fit to the τ_z with LS background and smeared MC prompt distribution. For $6 < p_T < 10$ GeV.