

Z boson analysis for physics laboratory courses using CERN Open Data

Bachelorarbeit
von

Philip Barthelmä

am Institut für Experimentelle Teilchenphysik

Reviewer: Prof. Dr. G. Quast
Second Reviewer: Dr. J. Wolf

Bearbeitungszeit: 15.05.2024 – 15.11.2024

Contents

1. Introduction	1
2. Fundamentals of resonances in the dimuon spectrum	3
2.1. Mathematical description of the resonances	4
2.2. Systematic uncertainties and biases affecting resonance shapes	5
2.2.1. Final State Radiation	6
2.2.2. Isolation of detected muons	6
3. The present state of the performed analysis	7
3.1. Estimation of the detector resolution	7
3.2. Calculation of the decay width Γ_Z of the Z boson	8
3.3. Potential for improvement in the current analysis strategy	9
3.3.1. Data set	9
3.3.2. Approximation of the transverse momenta and invariant mass	10
3.3.3. Determination of the detector resolution	10
4. The modernized analysis	11
4.1. The used tools	11
4.2. The used dataset	11
4.3. Estimating the detector resolution	12
4.3.1. Model function for the J/Ψ resonance	15
4.3.2. Estimating the contribution k_η	16
4.3.3. Estimating the momentum resolution $\Delta p_{T,i}$	18
4.4. Analysis of the Z resonance	23
4.4.1. Model Function for the Z Resonance	23
4.4.2. Reducing the background in the Z resonance	25
4.4.3. Fitting the Z resonance	26
4.4.3.1. Testing the accuracy of the mass resolution extrapolation .	27
4.4.3.2. Determining the decay width Γ_Z	29
5. Implementation of the analysis in a lab course	33
Appendix	35
A. NanoAOD to CSV conversion	35
B. Partitions of the Z resonance	36
C. Influence of the choice of partitions for the resolution $\Delta p_{T,1}$	38
D. Fits to the Z resonance	38

List of Figures

2.1. Feynman diagram of the Drell-Yan process. It describes the annihilation of a quark-antiquark pair and the creation of a lepton-antilepton pair via the exchange of a virtual photon or Z boson.	3
2.2. Plot of a normalized Breit-Wigner distribution with $\Gamma = 1$ and $M = 0$, a normalized Gaussian distribution with $\sigma = 1$ and $M = 0$ as well as a normalized Voigt profile with the same parameters.	5
3.1. Resonances of the J/Ψ meson and Z boson from the data of run 2010B	7
3.2. Splitting of the J/Ψ resonance for detector resolution estimation	8
3.3. Fit of the transverse momentum Resolution with respect to the average transverse momentum in each bin	9
4.1. Resonances of the J/Ψ meson and Z boson from the data of runs 2012B and 2012C	13
4.2. Characteristic dimuon invariant mass spectrum calculated with data from the runs 2012B and 2012C	13
4.3. Partitioning of the J/Ψ resonance to isolate the influence of k_η on the detector resolution	16
4.4. Gaussian fit to the reconstructed mass histogram with $p_{T,i} \in [10, 20]$ GeV for the $0 < \eta < 0.25$ bin.	17
4.5. Estimated contribution $k_\eta(\eta)$ to the detector resolution for different values of η	18
4.6. Different subsets of the J/Ψ resonance used to isolate the effect of $p_{T,2}$ on the mass resolution.	19
4.7. Fit of a Gaussian distribution to the subset of of data on the diagonal for $p_T = 12$ GeV to determine the constant c	20
4.8. Linear fit to the calculated relative uncertainties of $p_{T,2}$	21
4.9. Linear fit to the calculated relative uncertainties of $p_{T,1}$	22
4.10. Scatter plot of the transverse momentum of J/Ψ muons. The partitions used to determine the transverse momentum resolution $\Delta p_{T,1}$ are highlighted.	22
4.11. Filters used to reduce the background noise in the Z resonance.	26
4.12. Examples of different partitionings of the Z resonance based on the amount of data available in each η bin.	27
4.13. Different choices for J/Ψ partitions and their effect on the extrapolation on the transverse momentum resolution $\Delta p_{T,2}$	28
4.14. Comparison between the extrapolated and estimated true resolution of the detector	29
4.15. Fit results of the partition in the $\eta \in [0,0.25]$ bin as part of the multi fit to the Z resonance.	31
B.1. Partitions of the Z resonance in the $\eta \in [1.5,1.75]$ bin	36
B.2. Partitions of the Z resonance for the η bins $[1.75,2]$ and $[2,2.5]$	36
B.3. Partitions of the Z resonance for the η bins up to $\eta \in [1.25,1.5]$	37

C.4. Different choices for J/Ψ partitions and their effect on the extrapolation on the transverse momentum resolution $\Delta p_{T,1}$	38
D.5. Fit results of the partition in the $\eta \in [0,0.25]$ bin as part of the multi fit to the Z resonance.	39
D.6. Fit results of the partition in the $\eta \in [0.25,0.5]$ bin as part of the multi fit to the Z resonance.	39
D.7. Fit results of the partition in the $\eta \in [0.5,0.75]$ bin as part of the multi fit to the Z resonance.	39
D.8. Fit results of the partition in the $\eta \in [0.75,1]$ bin as part of the multi fit to the Z resonance.	39
D.9. Fit results of the partition in the $\eta \in [1,1.25]$ bin as part of the multi fit to the Z resonance.	40
D.10. Fit results of the partition in the $\eta \in [1.25,1.5]$ bin as part of the multi fit to the Z resonance.	40
D.11. Fit results of the bottom left partition in the $\eta \in [1.5,1.75]$ bin as part of the multi fit to the Z resonance.	40
D.12. Fit results of the top right partition in the $\eta \in [1.5,1.75]$ bin as part of the multi fit to the Z resonance.	40
D.13. Fit results of the bottom left partition in the $\eta \in [1.75,2]$ bin as part of the multi fit to the Z resonance.	41
D.14. Fit results of the top right partition in the $\eta \in [1.75,2]$ bin as part of the multi fit to the Z resonance.	41
D.15. Fit results of the bottom left partition in the $\eta \in [2,2.5]$ bin as part of the multi fit to the Z resonance.	41
D.16. Fit results of the top right partition in the $\eta \in [2,2.5]$ bin as part of the multi fit to the Z resonance.	41

List of Tables

4.1. Fit results for the detector resolution σ and signal data N_s for different bins of η of the J/Ψ resonance.	19
4.2. Fit results for data subsets along the $p_{T,1} = 12$ GeV axis shown in figure 4.6b	22
4.3. Fit results for data subsets along the $p_{T,2} = 12$ GeV axis shown in figure 4.10	23
4.4. Fit Results of the transverse momentum resolution extrapolation $\Delta p_{T,2}$ for different choices of partitions	28
4.5. Results for the extrapolated and fitted mass resolution for each partition of the Z resonance.	30
C.1. Fit Results of the transverse momentum resolution extrapolation $\Delta p_{T,1}$ for different choices of partitions	38

1. Introduction

Laboratory courses are an elementary component of a physics degree. Not only do they serve as a revision of the contents of lectures, but they also teach the basics of working with measuring instruments, safety equipment and help to build skills in the practical application of the material. They offer an initial insight into the methodologies of various fields of physics and thus serve as an orientation aid for further studies and careers. In addition, students learn how to write scientific papers in the form of protocols and how to analyze and interpret measured values and their uncertainties. Large amounts of data, non linear dependencies and correlated uncertainties require the help of a computer, the use of which for scientific purposes is already taught in the first semester. Ideally, the knowledge and skills imparted on the students in the lab courses transfer to further studies and prepare them for their future work. This is particularly difficult when it comes to the analysis of data. While central or historically relevant experiments in a field remain essentially unchanged, modern data analysis is constantly evolving and is largely detached from the actual experimental setup. Particularly in particle physics, large amounts of data are generated, the handling and processing of which require modern software packages. For example, the CMS detector produces 500 terabytes of data per second[1], the acquisition, selection and evaluation of which involves considerable effort, which would not be possible without powerful computers and efficient algorithms. The ROOT analysis framework[2] was developed at CERN specifically for this purpose. However, the learning effort required for the use of specialized software is too high in the context of a laboratory course.

Since 2017, the Karlsruhe Institute of Technology (KIT) has been offering students the opportunity to analyze data from the CMS detector and gain an initial insight into modern methods of data analysis as part of the lab course *Properties of Elementary Particles*. Using code libraries already known from previous experiments, the distribution of the invariant mass of $\mu^+\mu^-$ -pairs is investigated. The aim is to determine the decay width of the Z boson Γ_Z . However, from today's point of view, the experiment already contains outdated methods of analysis. Thanks to advances in software and faster computers, manual treatment of the signal and background, as well as averaging of the muon momenta, is no longer necessary. Lower computing power limited the scope of the analysis, which means that smaller data sets and simple assumptions are used. Nevertheless, this is associated with considerable programming effort due to the lack of suitable and widely adopted software packages at the time.

The experiment is modernized as part of this work. The necessary programming effort is reduced, and the focus of the experiment is directed towards the physical questions and the methods used. Furthermore, this lab course gives an insight into the handling of large amounts of data as well as widely used libraries in particle physics that simplify typical calculations and are designed for use on large data sets. In addition, the assumptions made are reduced to a practicable minimum to ensure a realistic insight into modern particle physics.

The provided analysis follows a data-only approach. Therefore, no simulations are used to find dependencies of variables or to correct for biases and systematic errors in the data. This not only keeps the experiment in a manageable scope, but also provides a more effective learning experience as students can conduct the entire analysis themselves. The analysis shown in this thesis is more comprehensive than what is manageable for a lab course. However, suggestions for a suitable implementation are provided.

Additionally, this thesis, along with the accompanying code and the up-to-date implementation of this lab course at KIT are provided on GitHub:

<https://github.com/PhilipBarthelmae/zlab>

2. Fundamentals of resonances in the dimuon spectrum

The starting point for the analysis are two data sets provided on the CERN Open Data Portal containing muon events from the CMS detector at the Large Hadron Collider (LHC). At the LHC, protons were accelerated along the 27 km long ring accelerator [3] to a center-of-mass energy of $\sqrt{s} = 8$ TeV and collided within the CMS detector. During the collision, partons can interact with each other and generate new particles and cause particle showers. The $\mu^+\mu^-$ -pairs (hereinafter referred to as dimuon pairs) relevant for this experiment are generated in the Drell-Yan process shown in figure 2.1.

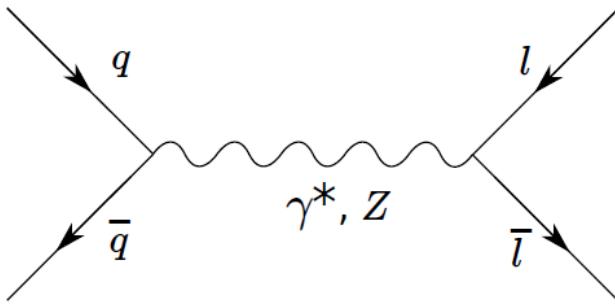


Figure 2.1.: Feynman diagram of the Drell-Yan process. It describes the annihilation of a quark-antiquark pair and the creation of a lepton-antilepton pair via the exchange of a virtual photon or Z boson.

This results in the annihilation of a quark-antiquark pair, creating a virtual photon or a Z boson, which then decays into a lepton-antilepton-pair, e.g. muons. Mesons, short-lived bound states from a quark-antiquark pair, can also decay into leptons in this way.

In special relativity, the invariant mass (or rest mass) of a particle is given by

$$m = \sqrt{E^2 - \vec{p}^2}. \quad (2.1)$$

Since the mass m of a particle is a scalar quantity, it is invariant under Lorentz transformation. Therefore it can be calculated from measurements of its energy E and momentum \vec{p} in any inertial reference frame. During decay, energy and momentum are conserved. The mass M of the decaying particle can be calculated using the four-momenta of the decay products via

$$M = \sqrt{\left(\sum_i P_i\right)^2} = \sqrt{\left(\sum_i E_i\right)^2 - \left(\sum_i \vec{p}_i\right)^2}, \quad (2.2)$$

where P_i , E_i and \vec{p}_i are the four-momentum, energy and momentum of the i -th decay product respectively. For a decay into two particles, the equation simplifies to

$$M^2 = m_1^2 + m_2^2 + 2(E_1 E_2 - \vec{p}_1 \cdot \vec{p}_2) \quad (2.3)$$

2.1. Mathematical description of the resonances

The characteristic dimuon invariant mass spectrum is the histogram of the recorded invariant masses of dimuon pairs. In the spectrum, several peaks, so called resonances, can be seen. The resonances correspond to different particles, such as mesons and the Z boson, from whose decay the dimuon pairs originate. These resonances have a decay width Γ and the associated particles a mean lifetime τ . In theory, the resonances have the shape of a Breit-Wigner distribution

$$B_W(m) = \frac{\frac{\Gamma}{2}}{\pi \left((m - M)^2 + \left(\frac{\Gamma}{2} \right)^2 \right)}. \quad (2.4)$$

The Breit-Wigner distribution is also known as the Cauchy distribution. Here, M describes the mean value of the distribution, corresponding to the mass of the decaying particle. The decay width Γ is given by the full width at half maximum (FWHM) of the distribution. The mean lifetime τ of system of unstable particles is defined as the time that elapses, after which the number of particles present has been reduced to $\frac{1}{e}$ times its original size. The relationship between these values can be derived from the energy-time-uncertainty relation. It yields

$$\Gamma \cdot \tau = \hbar. \quad (2.5)$$

However, when plotting the dimuon spectrum from invariant masses calculated from real world data, the resonances differ greatly from their theoretical form. This is primarily the result of the finite detector resolution. The resolution of the CMS detector can be assumed to be gaussian

$$G(m) = \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left(-\frac{1}{2} \frac{(m - M)^2}{\sigma^2}\right) \quad (2.6)$$

where M refers to the mean and σ the standard deviation of the distribution. The theoretical Breit-Wigner shape of the resonances is ‘smeared out’ by the detector resolution. Mathematically, this is described by a convolution. The resulting distribution is referred to as a Voigt profile

$$V(m) = B_W(m) * G(m). \quad (2.7)$$

For comparison, all three functions can be seen in Figure 2.2. In principle, the decay width of the Z boson Γ_Z could thus be determined by fitting a Voigt profile to the corresponding resonance. However, since both the detector resolution σ and the decay width Γ_Z contribute to the actual width of the resonance seen in the dimuon spectrum, these two quantities would be strongly correlated. The determined value for Γ_Z would therefore only be of limited use. A better result can be achieved by first obtaining an expression for the detector resolution and constraining it to the value found in the fit of the Z resonance. The resonance of the J/Ψ meson is suitable for this purpose.

With a decay width of

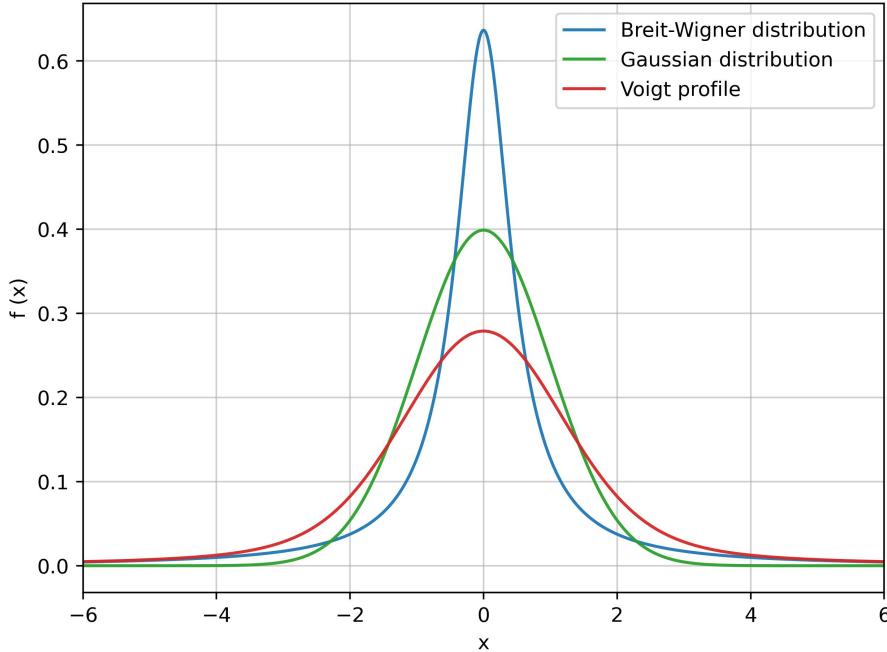


Figure 2.2.: Plot of a normalized Breit-Wigner distribution with $\Gamma = 1$ and $M = 0$, a normalized Gaussian distribution with $\sigma = 1$ and $M = 0$ as well as a normalized Voigt profile with the same parameters.

$$\Gamma_{J/\Psi} = (92.6 \pm 1.7) \text{ MeV}[4] \quad (2.8)$$

this resonance is extremely narrow. Its width is therefore dominated by the detector resolution and its shape can thus be regarded as Gaussian. An expression for the resolution σ can therefore be determined by performing suitable fits to the resonance. With a mass of

$$m_{J/\Psi} = (3096.900 \pm 0.006) \text{ MeV}[4] \quad (2.9)$$

the J/Ψ meson is significantly lighter than the Z boson with

$$m_Z = (91.1880 \pm 0.0020) \text{ GeV}[4]. \quad (2.10)$$

Consequently, significantly less rest energy is distributed to the decay products during the decay of a J/Ψ meson. Accordingly, the resulting muons have smaller momenta and transverse momenta than those produced by the decay of a Z boson. Since the resolution of the CMS detector depends on the four momenta of the recorded particles, a mass resolution σ determined at the J/Ψ resonance must first be extrapolated to the higher energies and momenta of Z decays.

2.2. Systematic uncertainties and biases affecting resonance shapes

As mentioned previously, the shapes of resonances found in the dimuon spectrum are affected by the detector resolution. This section will mention additional factors that alter the shape of the resonances for completeness sake. However, they will not be accounted for in this analysis as it would go beyond the scope of this thesis as well as the corresponding laboratory course. Additionally, the extensive use of simulations required to properly handle such effects would go against the data-only approach set out for this analysis.

2.2.1. Final State Radiation

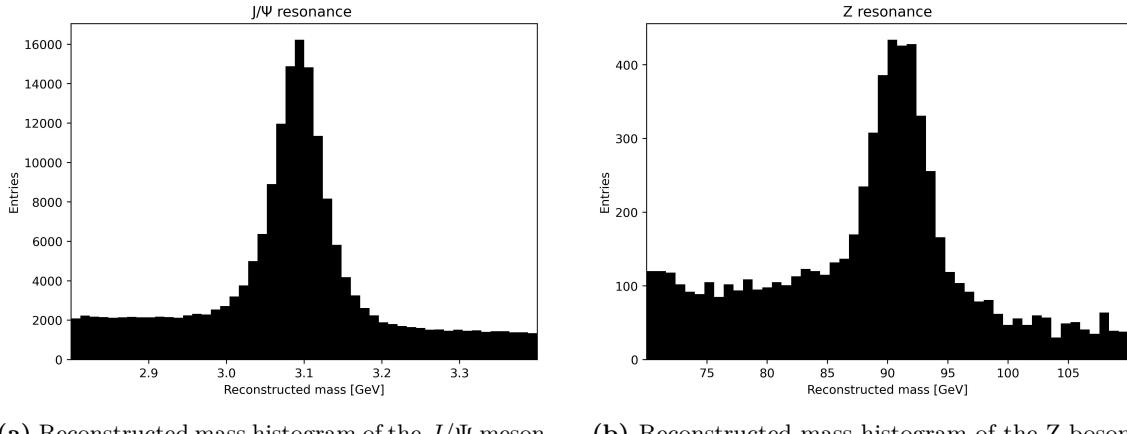
When the final state of a process emits photons, it is referred to as final state radiation. The muons created in the decay of the J/Ψ or Z boson can therefore radiate away energy even before they are measured in the calorimeter. This lowers the reconstructed mass of the decay products, biasing the resonance towards lower values. Even if the final state radiation is detected, it cannot be traced back to the particles it was emitted from. This effect can only be accounted for with the use of simulations or a significantly more complicated model function, whose shape has to be determined separately.

2.2.2. Isolation of detected muons

The resonances can also be distorted by insufficient isolation of the muons in the detector. This occurs particularly frequently with muons originating from the J/Ψ meson. These muons are often part of a jet and have a small opening angle, which increases the probability that they will not be registered in the detector in isolation. This can lead to misidentification of muons and false reconstructions of their momentum. In contrast, such problems occur much less frequently with the Z boson, as the muons are emitted roughly isotropically and can therefore generally be detected in isolation.

3. The present state of the performed analysis

The analysis is currently based on a data set of dimuon events from the 2010B run of the CMS detector containing around 32.2 million entries. The data set was previously filtered for the students so that only dimuon events with at least one global muon in the range from $M = 2$ GeV to $M = 110$ GeV are included. This reduces the data set to around 1.2 million events. Here, global refers to muons that were detected in the inner tracker as well as the muon system at the outer edge of the detector. After filtering, the J/Ψ resonance shown in 3.1a consists of around 186,000 events in the $M [2.8, 3.4]$ GeV range. The Z resonance in Figure 3.1b in the range $M [70, 110]$ GeV consists of only 6566 events. The following is an outline of the currently performed analysis. Afterwards, potential improvements are highlighted.



(a) Reconstructed mass histogram of the J/Ψ meson (b) Reconstructed mass histogram of the Z boson

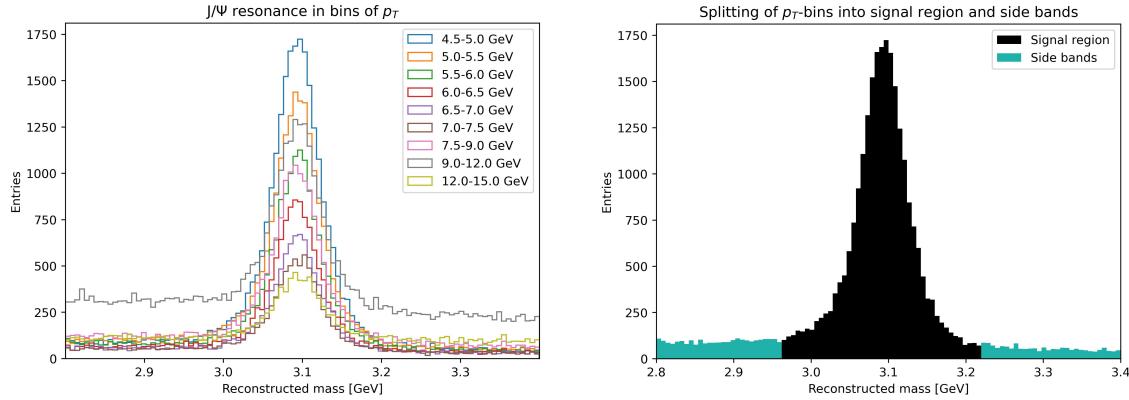
Figure 3.1.: Resonances of the J/Ψ meson and Z boson from the data of run 2010B

3.1. Estimation of the detector resolution

The J/Ψ resonance is now used as described in chapter 2 to estimate the mass resolution of the detector for the Z resonance. For this purpose, the transverse mass M_T is considered. It is given by

$$M_T^2 = 2p_{T,1}p_{T,2}(1 - \cos \theta). \quad (3.1)$$

Here, θ describes the angle between the muon momenta. Assuming that the transverse momentum of the two muons are of comparable size, both momenta can be approximated by a mean transverse momentum p_T , leading to



(a) Splitting of the J/Ψ resonance into different bins of p_T .

(b) Additional division of p_T -bins into signal region and side bands using the example of the $p_T \in [4.5, 5]$ GeV bin.

Figure 3.2.: Splitting of the J/Ψ resonance for detector resolution estimation

$$M_T^2 = 2p_T^2(1 - \cos \theta). \quad (3.2)$$

Using Gaussian error propagation, a simple relationship between the mass resolution σ or ΔM and the transverse momentum resolution Δp_T can be derived. It follows that

$$\frac{\Delta M}{M} = \frac{\Delta p_T}{p_T}. \quad (3.3)$$

Equation 3.3 can be used to convert the mass resolution of the J/Ψ resonance into a corresponding transverse momentum resolution. This resolution is then extrapolated to the mean transverse momentum of the Z resonance and can be converted back into the mass resolution of the Z boson using the same equation. For this purpose, the J/Ψ resonance is divided into different intervals of p_T (referred to as p_T bins in the following), as shown in Figure 3.2a. Each of the p_T bins is additionally divided into a signal region and two side bands containing only background noise. An example of the splitting into different regions is shown in figure 3.2b. By fitting a straight line to the side band regions, the background component in the signal region can be estimated and then subtracted from the data. The standard deviation of the remaining signal data then serves as an estimate of the detector resolution in the respective p_T bin. The resulting mass resolutions are then converted into corresponding momentum resolutions Δp_T using equation 3.3. The curve $\Delta p_T(\bar{p}_T)$, where \bar{p}_T is the mean transverse momentum of each p_T bin, is well described by a straight line, as shown in Figure 3.3, allowing for a simple extrapolation to the higher transverse momenta of muons present in the Z resonance.

With an average muon transverse momentum of $\bar{p}_{T,Z} = 39.31$ GeV, the extrapolated resolution comes out to $\Delta p_T = (1.10 \pm 0.94)$ GeV, leading to an estimated mass resolution of $\Delta M_Z = (2.47 \pm 0.17)$ GeV at the Z resonance.

3.2. Calculation of the decay width Γ_Z of the Z boson

Using the extrapolated mass resolution ΔM_Z from section 3.1, the half-width of the Z boson is now determined by fitting a Voigt profile to the Z resonance. Similar to figure 3.2b, the Z resonance is subdivided into side bands and signal region. The decay width Γ_Z is determined using a multi-fit of the signal region and side bands. This means that the fit of the signal and the background are combined into a single likelihood function,

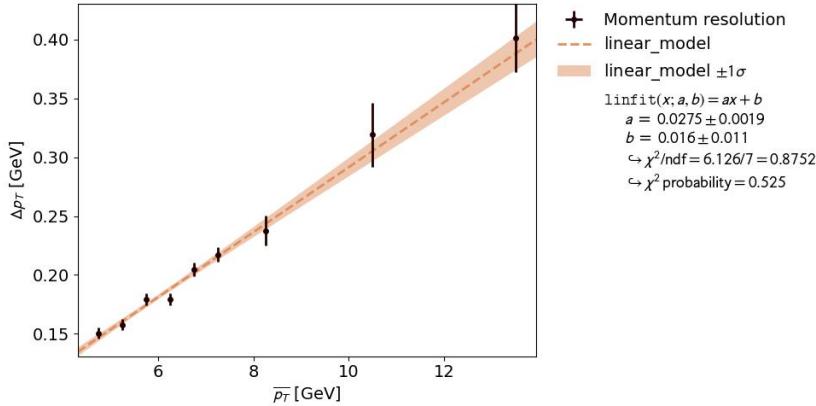


Figure 3.3.: Fit of the transverse momentum Resolution with respect to the average transverse momentum in each bin

which is subsequently optimized. The fit of a Voigt profile to the signal data with the mass resolution fixed to the value found above yields a final result of

$$\Gamma_Z = (1.870 \pm 0.074) \text{ GeV}. \quad (3.4)$$

However, upon revisiting the analysis for the purpose of this thesis, I found that the model function for the Z resonance uses the parameter γ in its definition of the Voigt profile, where $\Gamma = 2\gamma$. This is how the Voigt profile is commonly parameterized, but the analysis does not include the factor of 2 to convert between the two quantities, leading to the final result being half its intended value. Rerunning the code with an adjusted model function yields the result

$$\Gamma_Z = (3.740 \pm 0.148) \text{ GeV}. \quad (3.5)$$

3.3. Potential for improvement in the current analysis strategy

Properties of Elementary Particles is an important lab course in the masters degree at KIT that introduces students to the central content of modern experimental particle physics using real world data. However, simplifications had to be made in its design due to a lack of computing power and software packages which are no longer necessary from today's perspective. Therefore, the analysis method used does not provide a representative insight into modern data processing. The analysis strategy learned by students in the experiment is therefore only of limited use outside this lab course. In the following, the main potential for improving the analysis, which is to be implemented in the course of this work, is discussed.

3.3.1. Data set

The data set used contains only approximately 6500 events in the Z resonance, which is particularly relevant for the experiment. A larger number of data points would lead to a significantly improved statistical accuracy and reduce the uncertainties in the determination of the decay width. Since the introduction of the experiment, several data sets have been published on the CERN Open Data Portal that have significantly better statistics in the Z region and provide additional information about the muons, such as their isolation, that help to reduce background noise.

3.3.2. Approximation of the transverse momenta and invariant mass

The assumption that the transverse momenta of the two muons can be combined to their mean value simplifies the analysis considerably, but only represents a rough approximation. This will become especially apparent in the revised analysis, as it makes heavy use of the wide distribution of transverse momenta. A separate consideration of the $p_{T,i}$ reduces systematic errors and increases the accuracy of the result. Additionally, when calculating the mass resolution, only the transverse mass is considered. This approximation holds well for muons with a low pseudorapidity as $p_T \approx p$. However, for muons of higher η , the full invariant mass as well as uncertainties in their pseudorapidity must be considered for an accurate result.

3.3.3. Determination of the detector resolution

The current approach to calibrating the detector involves dividing the J/Ψ resonance into different bins of p_T , which are then additionally split into side bands and signal regions. This involves considerable programming effort and leads to a involved program structure. The mass resolution is then determined by arithmetically calculating the standard deviation of the data within the signal region. However, a parametrization of the background must first be found in order to remove it from the signal region. With the help of modern analysis software, this can be replaced by a single fit to the data, which performs a simultaneous adjustment to the background and signal. This significantly reduces the programming effort and allows astudents to focus on the physics at hand.

4. The modernized analysis

This chapter introduces the new analysis method for determining the decay width Γ_Z of the Z boson, along with the necessary tools and datasets. The analysis aims to incorporate as many effects to the final result as possible, while still providing clear and manageable approach for students to follow. Depending on the experience of the students, the analysis shown here might be too comprehensive for the scope of a lab course. Therefore, suggestions on how to simplify the analysis without significantly compromising on the accuracy of the final result are presented in a later chapter.

4.1. The used tools

The analysis is carried out using the programming language `python`, as it is taught to students at KIT starting from the first semester and is used in many lab courses throughout the physics degree. While `python` is slow compared to other programming languages, the many powerful software library's available and easy-to-read syntax make it a fitting choice for use in physics and scientific computing in general. This analysis incorporates widely used software packages like `numpy` [5][6], for handling collections of numbers, `matplotlib`[7] [8] for easy visual display of data, `pandas` [9] [10] for efficient handling of large amounts of data as well as `scipy` [11] [12], for their implementation of the voigt profile, for which no closed form solution exists. Additionally, some more specialized library's are used. These include:

scikit-hep vector [13] A package for handling 2D, 3D and relativistic four-vectors. It contains efficient, numerically stable and well tested implementations for many of the commonly performed calculations in high energy particle physics.

uncertainties [14] A package that allows for automatic propagation of uncertainties. It has integration with `numpy` arrays and can handle the propagation of correlated uncertainties in complex calculations with little change to existing code.

kafe2 [15] Karlsruhe Fit Environment 2 is a data fitting framework developed at KIT and designed for use in undergraduate physics lab courses. Most importantly, kafe2 offers the functionality to fit multiple histograms simultaneously to produce a single result. This is valuable, as different regions of the data set must be treated separately when accounting for uncertainties in measurements.

In addition to these software packages, the CMS Event Display [16] is used to introduce students to the lab course. It can be used to visualize data recorded by the CMS detector. This allows students to get more familiar with the detector as well as the different signatures muons can produce depending on their origin.

4.2. The used dataset

The basis for the modernized analysis is a combined dataset from events captured in the Run 2012B [17] and Run 2012C [18] at the CMS detector. Both are available for download

through the CERN Open Data Portal. While originally intended for the analysis of the Higgs boson, they contain enough dimuon events from decays of the J/Ψ meson and Z boson to be suitable for this analysis. In total, they consist of about 61.5 million events. For each event, numerous characteristics have been recorded. The characteristics used for this analysis are

Muon_pt Transverse momentum of the muons.

Muon_eta Pseudorapidity of the muons.

Muon_phi Azimuth angle of the muons.

Muon_charge Charge of muons (either +1 or -1).

Muon_dxy Distance of the muons to the primary vertex in the transverse plane.

Muon_dz Distance of the muons to the primary vertex in parallel to the beam pipe.

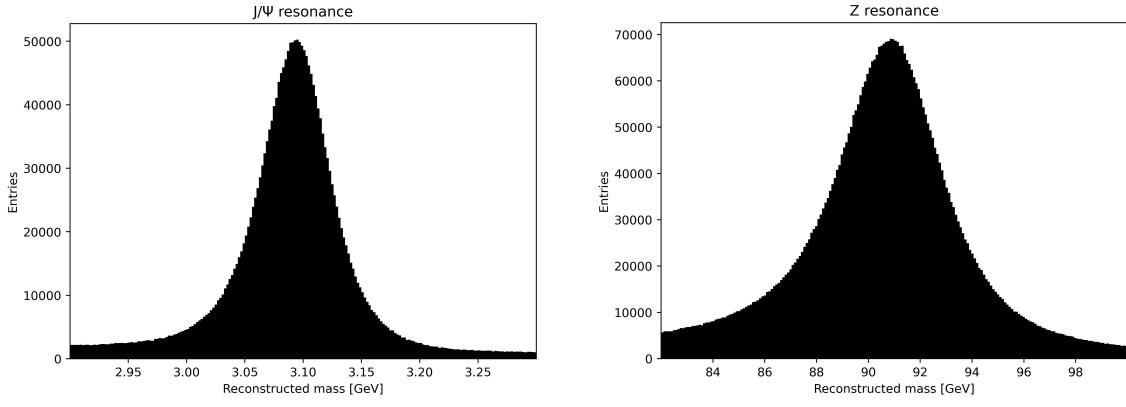
Muon_pfRelIso03_all Muon isolation divided by the transverse momentum in the $R = 0.3$ cone.

The transverse momentum p_T , pseudorapidity η and azimuth angle ϕ of the muons are used to calculate the invariant mass of the particle that decayed into muons. While the distances to the primary vertex d_{xy} and d_z as well as the isolation $pf_RelIso03_all$ (from here on simply referred to as $iso3$) are not strictly necessary to carry out the analysis, they help to reduce background noise in the dimuon spectrum. Since muons originating from particle decay are emitted near the primary vertex, restricting the analysis to muons with a short distance to it excludes those that coincidentally hit the detector at a similar time, but are unrelated to the event of interest. Similarly, $iso3$ can be used to restrict the analysis to muons that hit the detector in isolation, which reduces the chance of misidentification of particles or the reconstruction of their momenta, as described in section 2.2.2.

The data sets used from the CERN Open Data Portal are provided in the `.root` file format. While `.root`-files can be worked on in python using the `uproot` package, the data itself is stored in the NanoAOD (Analysis Object Data) format. Requiring students to learn to work with this format is not practicable within a lab course. Therefore, the data is provided to the students in the familiar CSV-format. The code used for the conversion is included in appendix A. Additionally, it only saves the characteristics listed above and only for events that contain exactly 2 muons. This helps to reduce the file size and filters out most events that are not needed for this Analysis. However, events that consist of two muons with the same charge are kept, as removing them from the data set is used as an introductory exercise. After applying these filters, including the removal of same-charge muon pairs, about 21 million dimuon events are left. Of these, around 2.2 million stem from the area around the J/Ψ resonance and about 4.6 million form the Z resonance. This is a major improvement from the currently used dataset and allows for a much more in depth analysis. Both resonances can be seen individually in figure 4.1 and the entire dimuon spectrum is shown in figure 4.2. In addition to the J/Ψ resonance at ≈ 3 GeV and the Z resonance at ≈ 90 GeV, the full dimuon spectrum also contains additional resonances corresponding to other particles. The peak immediately to the right of the J/Ψ , located at ≈ 3.6 GeV, corresponds to $J/\Psi(2S)$, the first excited state of charmonium. The collection of peaks at ≈ 10 GeV belong to different states of bottomonium, a particle consisting of a $b\bar{b}$ -pair. The large hump at ≈ 30 GeV does not correspond to a particle. Rather, it is an artifact of the data selection criteria used at the CMS experiment.

4.3. Estimating the detector resolution

Obtaining an estimate for the detector resolution is a critical part of the analysis. Directly fitting a Voigt profile on to the Z resonance leads to a high correlation between the detector



(a) Reconstructed mass histogram of the J/ψ meson (b) Reconstructed mass histogram of the Z boson

Figure 4.1.: Resonances of the J/ψ meson and Z boson from the data of runs 2012B and 2012C

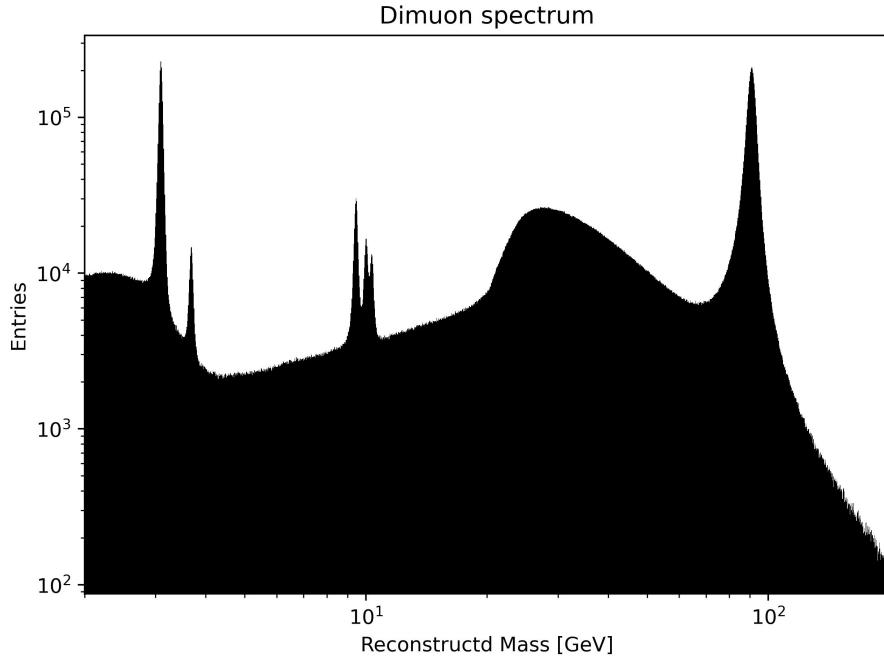


Figure 4.2.: Characteristic dimuon invariant mass spectrum calculated with data from the runs 2012B and 2012C

resolution σ or ΔM and the decay width Γ_Z , as both parameters influence the width of the Z resonance in the spectrum.

The detector resolution is not constant, but changes with the transverse momenta p_T , pseudorapidity η and azimuth angle ϕ of the detected muons. Since the Z resonance is comprised of the reconstructed mass of many different dimuon pairs whose muons range widely in these properties, the detector resolution depends on six different parameters. Therefore, finding a suitable expression is difficult, so some simplifications are made. Nevertheless, estimating the detector resolution comprises the bulk of the analysis.

The detector resolution can be determined using Gaussian error propagation. As mentioned in chapter 2, for a two particle decay, the invariant mass of a decaying particle can be calculated from the four-momenta of the decay products using the relationship

$$M^2 = m_1^2 + m_2^2 + 2(E_1 E_2 - \vec{p}_1 \cdot \vec{p}_2). \quad (4.1)$$

For particles with sufficiently high momenta, as is the case for the muons that originate from the decay of J/Ψ mesons and Z bosons, the rest mass contribution is negligible. The muons can therefore be assumed to be mass-less. For $m_1 = m_2 \approx 0$, the energy is given by $E_i = p_i$ and the equation simplifies to

$$M^2 = 2p_1 p_2 (1 - \cos \theta) = 2 \frac{p_{T,1}}{\sin \theta_1} \frac{p_{T,2}}{\sin \theta_2} (1 - \cos \theta), \quad (4.2)$$

where θ describes the angle between the muon momenta and θ_i is the angle between the momentum p_i and the positive direction of the beam line. The angle θ_i is related to the particles pseudorapidity η_i via

$$\eta_i = -\ln \left[\tan \left(\frac{\theta_i}{2} \right) \right]. \quad (4.3)$$

The mass resolution could now be determined by applying Gaussian error propagation on equation 4.2. However, if all dependencies on $p_{T,i}$, η_i and ϕ_i were accounted for, the resulting equation would become unmanageable. Therefore, some simplifications are made to reduce the parameter space. Firstly, it is assumed that the resolution does not depend on the azimuth angle ϕ_i as the CMS detector has cylindrical symmetry. Now, the error propagation is performed. However, only uncertainties in $p_{T,i}$ are considered, as their uncertainty is the dominating contribution to the detector resolution. This results in the expression

$$\frac{\Delta M}{M} = \frac{1}{2} \sqrt{\left(\frac{\Delta p_{T,1}}{p_{T,1}} \right)^2 + \left(\frac{\Delta p_{T,2}}{p_{T,2}} \right)^2}. \quad (4.4)$$

Ideally, the influence of η_1 and η_2 would be considered separately. However, even the new dataset does not contain enough statistics to reliably determine the contributions from both parameters individually. Therefore, to account for the dependency in η , an additional term k_η^2 is added separately, that combines the dependency from both pseudorapidities. Finally, the expression for the mass resolution of the detector is given by

$$\frac{\Delta M}{M} = \frac{1}{2} \sqrt{\left(\frac{\Delta p_{T,1}}{p_{T,1}} \right)^2 + \left(\frac{\Delta p_{T,2}}{p_{T,2}} \right)^2 + k_\eta^2}. \quad (4.5)$$

In order to be able to use this equation to calculate the mass resolution at the Z boson, the uncertainties $\Delta p_{T,i}$ as well as k_η have to be determined. To isolate the individual contributions, the data set is divided in such a way, that two of the three terms can either assumed to be constant or neglected entirely. Then, a fit to resonance is performed to determine its width ΔM . This result can then be used to calculate the respective uncertainty term. By repeating this process multiple times for suitable partitions of the data set, the dependencies $\Delta p_{T,i}(\bar{p}_{T,i})$ as well as $k_\eta(\bar{\eta})$ can be determined. Finally, the resolution ΔM_Z is then given by

$$\Delta M_Z = \frac{M_Z}{2} \sqrt{\left(\frac{\Delta p_{T,1}(\bar{p}_{T,1})}{\bar{p}_{T,1,Z}} \right)^2 + \left(\frac{\Delta p_{T,2}(\bar{p}_{T,2})}{\bar{p}_{T,2,Z}} \right)^2 + k_\eta^2(\eta_Z)}. \quad (4.6)$$

4.3.1. Model function for the J/Ψ resonance

Since the analysis relies heavily on fitting functions to determine key quantities, such as the detector resolution and the decay width Γ_Z of the Z boson, it is important to carefully consider the parametrization of the model functions used. In addition to accurately modeling the resonance, the model function must also account for the surrounding background. Failing to do so can lead to poor fit convergence or inaccurate parameter estimates. Therefore, a parametrization of the background must be included in the model function.

When performing histogram fits, it is helpful to parameterize the model functions as normalized probability density functions (PDFs), which are then scaled by an additional parameter N . In this approach, the PDF defines the shape of the distribution, while N provides an estimate of the number of entries that correspond to that distribution. Thus, the overall model function $f(x)$ can be expressed as:

$$f_{J/\Psi}(x) = N_s \cdot R_{J/\Psi}(x) + N_b \cdot B_{J/\Psi}(x), \quad (4.7)$$

where $N_s \cdot R_{J/\Psi}(x)$ captures the resonance (or signal) and $N_b \cdot B_{J/\Psi}(x)$ models the background. As mentioned in chapter 2, the decay width $\Gamma_{J/\Psi}$ of the J/Ψ meson is extremely small compared to the detector resolution. Since the detector resolution can be approximated as a Gaussian, the entire J/Ψ resonance in the dimuon spectrum can also be modeled as such:

$$R_{J/\Psi}(x) = \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left(-\frac{1}{2} \frac{(x - \mu)^2}{\sigma^2}\right) \equiv G(x). \quad (4.8)$$

For the background, no universal PDF is available. However, as seen in 4.1a, the background near the resonance is largely flat. Therefore, a second-order polynomial $p_{J/\Psi}(x)$ is used as the basis for $B_{J/\Psi}(x)$. To minimize the correlation between the fit parameters, an orthogonal polynomial centered around the mean of the distribution is used. Specifically the sum of the first three Legendre polynomials:

$$p_{J/\Psi}(x) = \frac{a}{2} \left(3(x - \bar{m})^2 - 1 \right) + b(x - \bar{m}) + c \quad (4.9)$$

where the parameters a , b , and c are determined by the fit and \bar{m} is the mean of the distribution, which can be estimated from the histogram entries. Alternatively, a polynomial of a different order could be used, but higher-order polynomials introduce more complexity and lead longer execution time for the fitting process, while lower-order polynomials might not adequately capture the background. To serve as the probability density function of the background, the polynomial must be normalized over the region of the J/Ψ resonance. If the data is taken from the interval $M \in [m_0, m_1]$, the PDF of the background $B(x)$ is given by:

$$B_{J/\Psi}(x) = \frac{p_{J/\Psi}(x)}{\int_{m_0}^{m_1} p_{J/\Psi}(x) dx}. \quad (4.10)$$

The integral in the denominator evaluates to

$$\begin{aligned} \int_{m_0}^{m_1} p_{J/\Psi}(x) dx &= \frac{1}{2} (m_1 - m_0) \left(am_1^2 + (a(m_0 - 3\bar{m}) + b)m_1 + am_0^2 \right. \\ &\quad \left. - (3a\bar{m} - b)m_0 + 3a\bar{m}^2 - 2b\bar{m} + 2c - a \right). \end{aligned} \quad (4.11)$$

Finally, the model function $f_{J/\Psi}(x)$ is given by

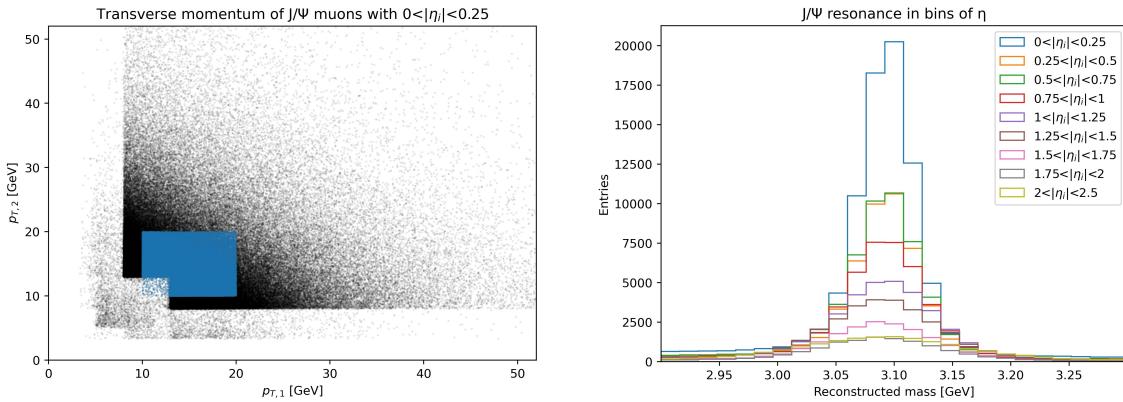
$$f_{J/\Psi}(x) = N_s \cdot G(x) + N_b \cdot \frac{p_{J/\Psi}(x)}{\int_{m_0}^{m_1} p_{J/\Psi}(x) dx}. \quad (4.12)$$

4.3.2. Estimating the contribution k_η

In order to determine $k_\eta(\eta)$, the dependency on $\Delta p_{T,i}$ needs to be eliminated from equation 4.5. For this purpose, the data comprising the J/Ψ resonance is divided into different bins of η . The bin edges are chosen to be $[0, 0.25, 0.5, \dots, 1.75, 2, 2.5]$. This results in 9 different bins. Each bin has a width of 0.25, with the exception of the last bin, whose bin width has been increased to 0.5 due to a lack of data for large values of η . Other choices for bin edges are possible, but a balance must be struck between having enough bins to capture $k_\eta(\eta)$ and each bin containing enough data to reliably fit a Gaussian distribution to the mass histogram. Figure 4.3b shows the resulting binned J/Ψ resonance.

Events are added to the corresponding bin only if the pseudorapidity of both muons of the respective dimuon pair fit within the bin edges. During this process, only the absolute value of the rapidity is taken into account. This is because the sign of η merely indicates the direction the muon is traveling in which has no bearing on the uncertainty of measurement.

For each bin, only the dimuon pairs with $p_{T,i} \in [10, 20]$ GeV are considered. Therefore, the transverse momenta of both muons are of roughly equal size. Figure 4.3a shows a scatter plot of the muon transverse momenta for the $0 < |\eta| < 0.25$ bin. The selected data is marked in blue. Additionally, it can be seen that there are seemingly two different distributions of muon transverse momenta layered on top of one another. This is because of the use of different triggers. Due to the enormous amount of data generated at the CMS experiment, not all of it can be committed to storage. Therefore, the incoming data is quickly examined and a trigger selects potentially interesting events that are then committed to memory for further analysis. The two overlapping distributions, as well as strict cut-offs seen in the scatter plot, are the result of different triggers selectively storing the data.



(a) Scatter plot of the transverse momentum of J/Ψ muons in the $0 < |\eta| < 0.25$ bin. The data used for the fit is marked in blue.

(b) J/Ψ resonance in bins of η

Figure 4.3.: Partitioning of the J/Ψ resonance to isolate the influence of k_η on the detector resolution

With $p_{T,1} \approx p_{T,2} \equiv p_T$ in the selected subset of data, the equation for the invariant mass simplifies to

$$M^2 = 2 \frac{p_T^2}{\sin^2 \theta'} (1 - \cos \theta). \quad (4.13)$$

Analogously to before, error propagation with respect to p_T is performed and k_η^2 is added to the resulting expression, giving

$$\frac{\Delta M}{M} = \sqrt{\left(\frac{\Delta p_T}{p_T}\right)^2 + k_\eta^2}. \quad (4.14)$$

For small values of η ($\eta << 1$), the muon momentum lies almost entirely in the transverse plane ($\theta' \approx 90^\circ$), and therefore $p_T \approx p$. In this limit, the influence of η on the detector resolution is negligible ($k_\eta \approx 0$). Therefore, for the $0 < |\eta| < 0.25$ bin shown in figure 4.3a, the mass resolution simplifies to

$$\frac{\Delta M}{M} = \frac{\Delta p_T}{p_T}. \quad (4.15)$$

This relationship is used to calculate the momentum resolution. Figure 4.4 shows the fit of a Gaussian distribution to the selected data for the $0 < |\eta| < 0.25$ bin. The mass of the J/Ψ meson can be estimated as the mean of the data of its resonance shown in figure 4.1a. This approach yields an estimate of

$$M_{J/\Psi} = 3.08939 \pm 0.00004 \text{ GeV}, \quad (4.16)$$

which is very close to the literature value provided by the Particle Data Group [4] of

$$M_{J/\Psi,\text{lit}} = 3.096900 \pm 0.000006 \text{ GeV}. \quad (4.17)$$

However, the estimate is slightly smaller than the literature value and the calculated statistical uncertainty can not account for the difference. This is caused by the systematic biases mentioned in chapter 2, primarily the final state radiation skewing the reconstructed mass towards lower values. The resulting asymmetry can also be seen in figure 4.1a, with the bins on the left side of the peak having consistently more entries. The effect is even more prominent for the Z resonance in figure 4.1b, as it has significantly more entries, making systematic biases even more apparent. As the asymmetry is not accounted for in the model function, the performed fits tend to have a poor goodness of fit (gof), indicated by the reduced likelihood $-2 \ln \mathcal{L}_R / \text{ndf}$. For a reduced likelihood of ≈ 1 , the fit is ideal.

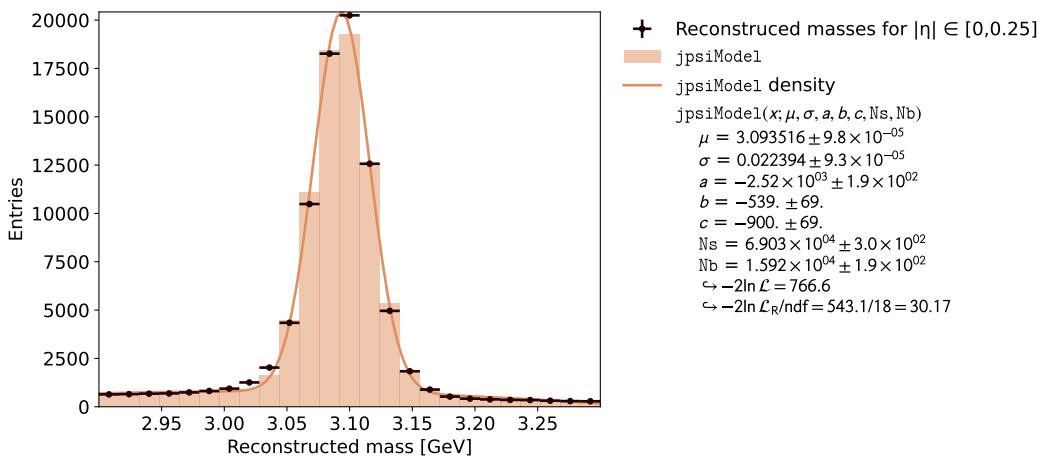


Figure 4.4.: Gaussian fit to the reconstructed mass histogram with $p_{T,i} \in [10, 20]$ GeV for the $0 < |\eta| < 0.25$ bin.

Using the resulting value for the mass resolution ΔM , given by the standard deviation σ of the fitted Gaussian, the momentum resolution is calculated to be

$$\frac{\Delta p_T}{p_T} = 0.007249 \pm 0.000030. \quad (4.18)$$

Since for every η bin, the same subset of data ($p_{T,i} \in [10, 20]$ GeV) is considered, the momentum resolution is assumed to be the same in all η bins. Therefore, for all subsequent η bins, a similar fit is performed and by rearranging equation 4.14 to

$$k_\eta = \sqrt{\left(\frac{\Delta M}{M}\right)^2 - \left(\frac{\Delta p_T}{p_T}\right)^2} \quad (4.19)$$

and using the momentum resolution determined above as well as the result of the fit, $k_\eta(\eta)$ is calculated. The relevant fit results for all η bins are shown in table 4.1 and the calculated values for $k_\eta(\eta)$ are shown in figure 4.5. The pseudorapidity of the muons has a noticeable influence on the detector resolution, as σ increases considerably with higher η bins. The table also illustrates an increase in the goodness of fit for bins of larger η . This is due to a decrease in the available data for each bin. Therefore, the systematic uncertainties introduced by the final state radiation are less pronounced and can be absorbed into the statistical uncertainty, leading a reduced likelihood of approximately 1.

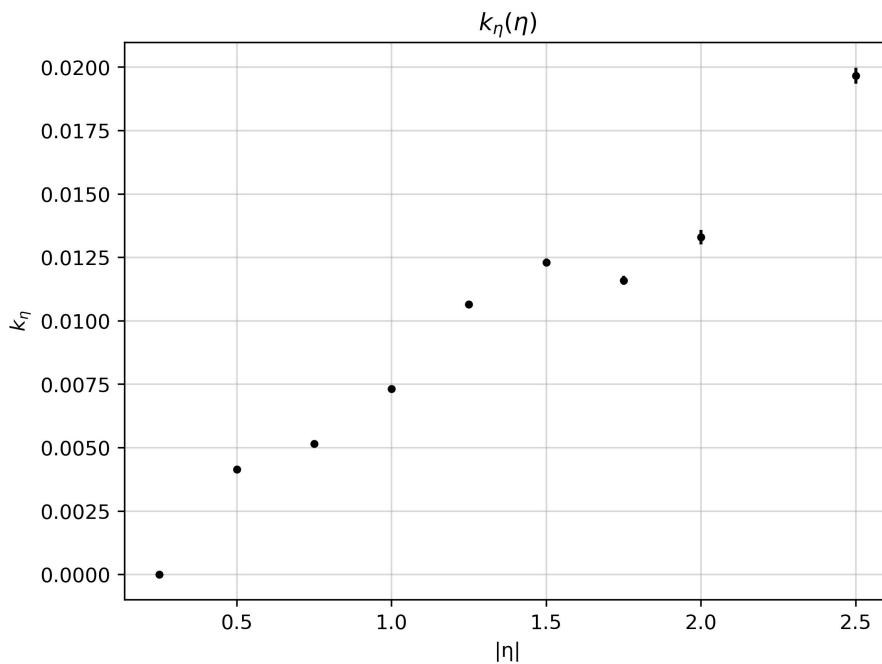


Figure 4.5.: Estimated contribution $k_\eta(\eta)$ to the detector resolution for different values of η

4.3.3. Estimating the momentum resolution $\Delta p_{T,i}$

The general approach to determine $\Delta p_{T,i}$ follows a similar methodology as described previously. By eliminating $\Delta p_{T,1}$ from equation 4.5, it can be rearranged to calculate $\Delta p_{T,2}$ and vice versa. However, because there are no data points in regions where either $p_{T,1}$ or $p_{T,2}$ are zero, both uncertainty terms will always contribute to the overall mass resolution, meaning neither term can be assumed to be negligible.

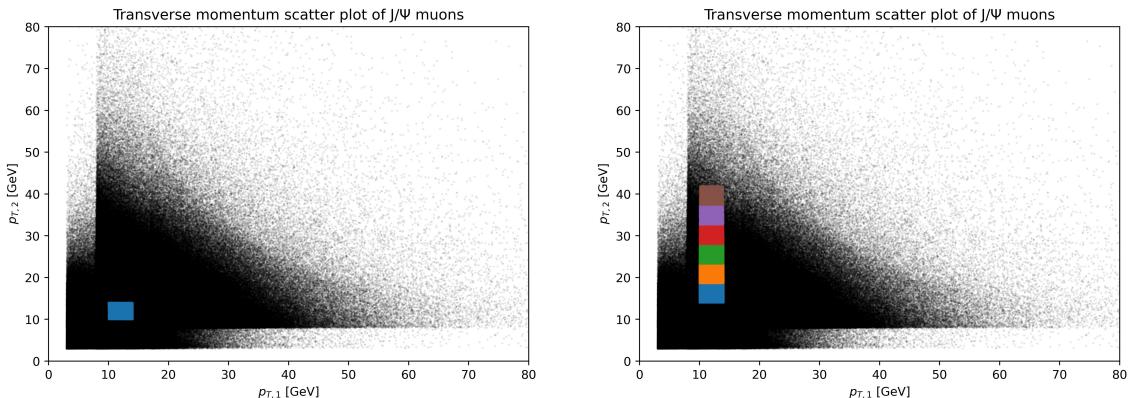
Ideally, to isolate the dependence on $\Delta p_{T,1}$, one would only consider data where $\Delta p_{T,2}$ remains constant. Since $\Delta p_{T,2}$ likely depends on both the transverse momentum $p_{T,2}$ and the pseudorapidity η_2 , this could theoretically be achieved by binning the J/Ψ resonance

η bin	σ [GeV]	$\Delta\sigma$ [GeV]	N_s	ΔN_s	$-2 \ln \mathcal{L}_R/\text{ndf}$
[0,0.25]	0.02239	0.00009	69035	296	30.17
[0.25,0.5]	0.02580	0.00014	41301	234	13.93
[0.5,0.75]	0.02748	0.00014	44253	243	12.41
[0.75,1]	0.03182	0.00028	36529	345	8.10
[1,1.25]	0.03980	0.00030	30008	261	3.06
[1.25,1.5]	0.04410	0.00041	25175	286	1.12
[1.5,1.75]	0.04226	0.00047	15439	205	1.08
[1.75,2]	0.04680	0.00076	10204	213	1.15
[2,2.5]	0.06475	0.00092	14331	220	1.19

Table 4.1.: Fit results for the detector resolution σ and signal data N_s for different bins of η of the J/Ψ resonance.

simultaneously in $p_{T,2}$ and η_2 . However, the available dataset does not provide enough data points within each bin to perform the necessary fits if both variables are binned simultaneously.

Given this limitation, the analysis is simplified by disregarding the potential dependence on pseudorapidity and it is assumed that the transverse momentum resolution depends solely on the transverse momentum p_T itself. This assumption allows for a more straightforward analysis, though it may introduce inaccuracy's if η has a significant effect on $\Delta p_{T,i}$. Therefore, the accuracy of the resulting detector resolution estimate is examined in a later chapter.



(a) Scatter plot of the transverse momentum of J/Ψ muons. The calibration partition is highlighted in blue.

(b) Scatter plot of the transverse momentum of J/Ψ muons. The partitions used to determine the transverse momentum resolution $\Delta p_{T,2}$ are highlighted.

Figure 4.6.: Different subsets of the J/Ψ resonance used to isolate the effect of $p_{T,2}$ on the mass resolution.

With these assumptions, equation 4.5 simplifies to

$$\frac{\Delta M}{M} = \frac{1}{2} \sqrt{c_1 + \left(\frac{\Delta p_{T,2}}{p_{T,2}} \right)^2 + k_\eta^2}, \quad (4.20)$$

if only data on the axis $p_{T,1} = \text{const.}$ is considered. Similarly, for an axis of constant $p_{T,2}$, the equation simplifies to

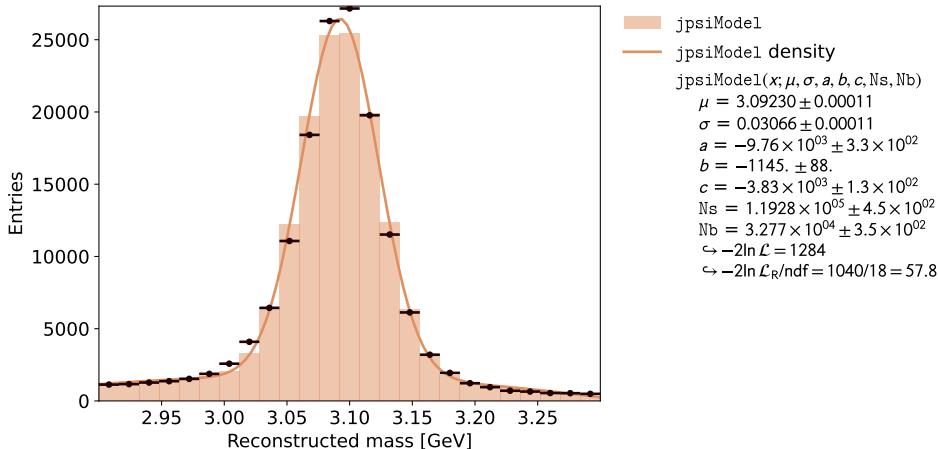


Figure 4.7.: Fit of a Gaussian distribution to the subset of data on the diagonal for $p_T = 12$ GeV to determine the constant c

$$\frac{\Delta M}{M} = \frac{1}{2} \sqrt{\left(\frac{\Delta p_{T,1}}{p_{T,1}} \right)^2 + c_2 + k_\eta^2}, \quad (4.21)$$

Now, all that is necessary is to find the transverse momentum resolution for one value of $p_{T,i}$ to determine the constant c_i . This can be achieved by considering a partition along the diagonal $p_{T,1} \approx p_{T,2} \equiv p_T$. This partition is now referred to as the calibration partition. As shown in section 4.3.2, the expression for the mass resolution then reduces to

$$\frac{\Delta M}{M} = \sqrt{\left(\frac{\Delta p_T}{p_T} \right)^2 + k_\eta^2} = \sqrt{c_{1,2} + k_\eta^2}. \quad (4.22)$$

Figure 4.6a shows a scatter plot of the transverse momenta of all dimuon pairs making up the J/Ψ resonance. The selected calibration partition used to calculate c_i is highlighted in blue. From here on out, the procedure to calculate $\Delta p_{T,1}$ and $\Delta p_{T,2}$ is identical. Therefore, the full analysis for $\Delta p_{T,2}$ will be shown in detail while only the results will be presented for $\Delta p_{T,1}$. Furthermore, the same subset of data is chosen to calculate both constants c_i , leading to $c_1 = c_2 \equiv c$.

Using equation 4.22 the constant c is calculated to be

$$c = \left(\frac{\Delta p_T}{p_T} \right)^2 = \left(\frac{\Delta M}{M} \right)^2 - k_\eta^2 = (-1.49 \pm 0.27) \cdot 10^{-5}. \quad (4.23)$$

Here, the mass resolution ΔM was again obtained by performing a Gaussian fit the selected subset of data. The fit results can be seen in figure 4.7. k_η has been determined in the previous section. However, since the data used for the fit was not binned in η due to a lack of events, it consists of muons with a wide variety of pseudorapidities. Consequently, no single value of k_η can perfectly capture the dependency on the pseudorapidity. As a compromise, the correction term of the center bin, $k_\eta(\eta = 1.25) = 0.0001134 \pm 0.0000026$, is used.

This results in small but negative value for the constant c . This is not physical, as c represents the squared relative momentum resolution, which cannot be imaginary. This is likely the consequence of the fact that one value of k_η had to be chosen to for a wide

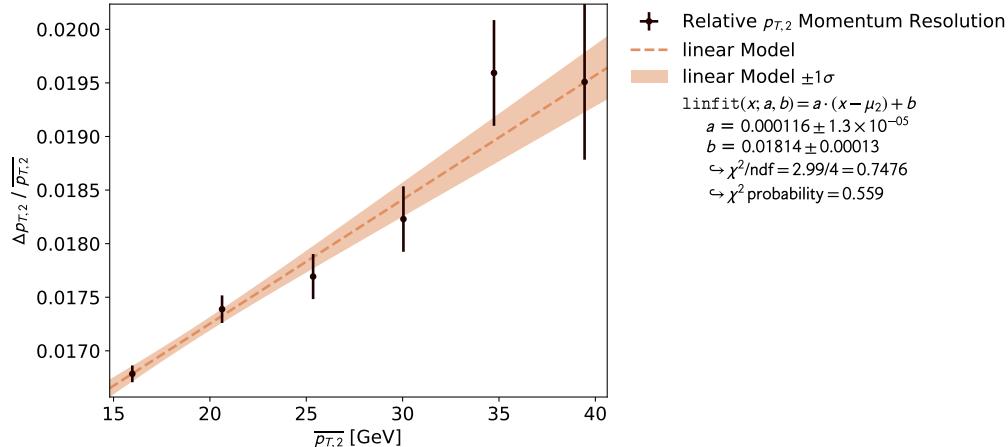


Figure 4.8.: Linear fit to the calculated relative uncertainties of $p_{T,2}$.

range of data as well as the simplifying assumptions made about the dependencies of the uncertainty terms.

However, despite this issue, the subsequent results obtained using this value of c still produce reasonable and consistent results. The negative value of c is minimal enough that its effect on the overall analysis is negligible.

With the value of c , the transverse momentum resolution $\Delta p_{T,2}$ can finally be determined. For this purpose, an additional six different subsets between $p_{T,2} = 14$ GeV and $p_{T,2} = 42$ GeV along the axis $p_{T,1} = 12$ GeV, shown in figure 4.6b, are considered. Each of these patches is of equal size with a height of 4.6 GeV and a width of 4 GeV. Again, this choice strikes a balance between having enough partitions to capture the trend of the transverse momentum resolution and each patch containing enough data to perform the necessary fit. For each subset, a Gaussian fit is performed as before to determine the mass resolution ΔM . Using the value of c , $k_\eta(\eta = 1.25)$ and the mass $M_{J/\Psi}$ determined above, the relative momentum resolution can be calculated for each partition by rearranging equation 4.20 to

$$\frac{\Delta p_{T,2}}{p_{T,2}} = \sqrt{4 \left(\frac{\Delta M}{M} \right)^2 - c - k_\eta^2(\eta = 1.25)}. \quad (4.24)$$

The fit results used for the calculation are summarized in table 4.2. As expected, the resolution gets worse with an increase in the transverse momentum $p_{T,2}$. Plotting the results against the average transverse momentum $\bar{p}_{T,2}$ for each bin shows that the trend of the relative momentum resolution is well approximated by a linear function. Figure 4.8 displays the corresponding linear fit. As a model function, a line centered around the mean of the data points $\mu_2 = 27.697$ GeV is used. The derived expression can now be used to calculate the momentum resolution $\Delta p_{T,2}$ for any given value of $p_{T,2}$.

As mentioned previously, the same approach is taken to calculate the transverse momentum resolution for $p_{T,1}$. The only change necessary is to now select partitions along the $p_{T,1}$ axis for a constant $p_{T,2}$ of 12 GeV, since the same patch of data is used to calculate the constant, $c_1 = c$. The same dimensions for each subset have been chosen, but with their width and height flipped accordingly. The different partitions are illustrated in figure 4.10. The patches appear to be larger compared to those used above. However, this is a consequence of the fact that the figures have a rectangular shape.

The respective linear fit to the transverse momentum resolution is shown in figure 4.9. Again, a linear model function centered around the mean of the data points is used. Due

$p_{T,2}$ range [GeV]	σ [GeV]	$\Delta\sigma$ [GeV]	N_s	ΔN_s	$-2 \ln \mathcal{L}_R/\text{ndf}$
[14.0,18.7]	0.03012	0.00010	151706	522	67.59
[18.7,23.3]	0.03093	0.00017	58313	330	27.16
[23.3,28.0]	0.03134	0.00028	23604	217	13.32
[28.0,32.7]	0.03206	0.00041	10370	143	2.16
[32.7,37.3]	0.03393	0.00068	5014	106	2.93
[37.3,42.0]	0.03381	0.00100	2458	79	1.64

Table 4.2.: Fit results for data subsets along the $p_{T,1} = 12$ GeV axis shown in figure 4.6b

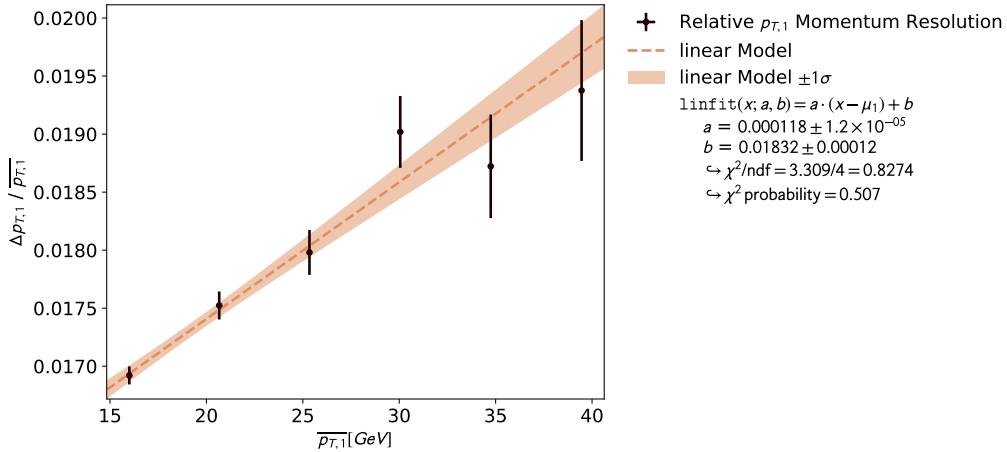


Figure 4.9.: Linear fit to the calculated relative uncertainties of $p_{T,1}$.

to the symmetric choice of bins, the mean $\mu_1 = 27.711$ GeV is almost identical to μ_2 . The fit results of the different partitions used are shown in table 4.3.

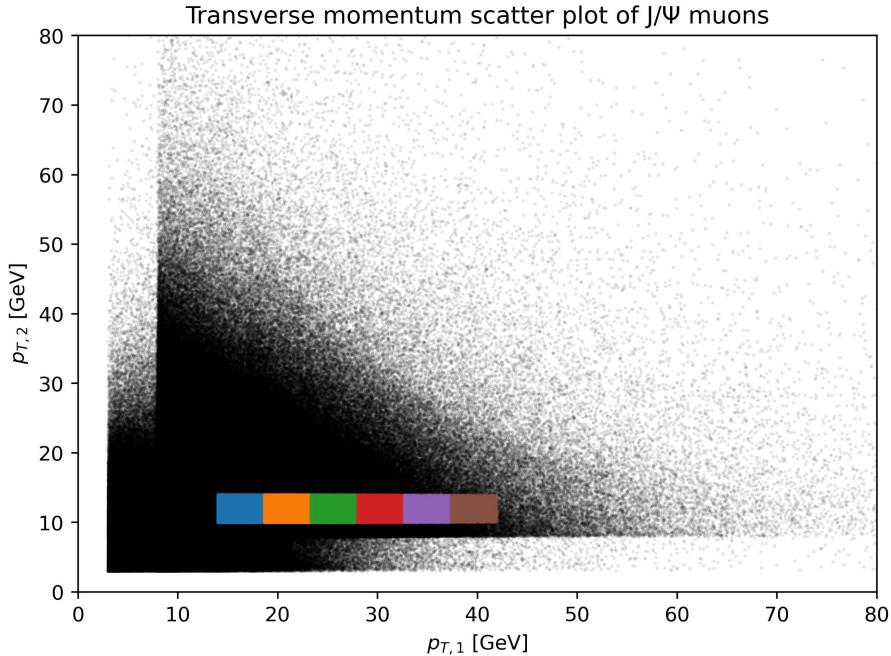


Figure 4.10.: Scatter plot of the transverse momentum of J/Ψ muons. The partitions used to determine the transverse momentum resolution $\Delta p_{T,1}$ are highlighted.

$p_{T,1}$ range [GeV]	σ [GeV]	$\Delta\sigma$ [GeV]	N_s	ΔN_s	$-2 \ln \mathcal{L}_R/\text{ndf}$
[14.0,18.7]	0.03030	0.00010	154870	549	68.85
[18.7,23.3]	0.03111	0.00016	63523	339	19.60
[23.3,28.0]	0.03172	0.00026	26770	229	10.66
[28.0,32.7]	0.03314	0.00042	11919	160	4.94
[32.7,37.3]	0.03273	0.00061	5726	110	3.52
[37.3,42.0]	0.03363	0.00083	2957	80	1.42

Table 4.3.: Fit results for data subsets along the $p_{T,2} = 12$ GeV axis shown in figure 4.10

Finally, the expressions for the transverse momentum resolution are given by

$$\begin{aligned} \frac{\Delta p_{T,1}}{p_{T,1}} &= 0.000118 \frac{1}{\text{GeV}} \cdot (p_{T,1} - 27.711 \text{ GeV}) + 0.0183 \text{ and} \\ \frac{\Delta p_{T,2}}{p_{T,2}} &= 0.000116 \frac{1}{\text{GeV}} \cdot (p_{T,2} - 27.697 \text{ GeV}) + 0.0181. \end{aligned} \quad (4.25)$$

For clarity, the parameter uncertainties have been omitted. Unsurprisingly, the expressions are almost identical, as both estimate the uncertainty in the transverse momentum measurement for muons. Which way the muons are labeled has no impact on the result.

Together with the correction term k_η^2 determined above, these results can now be used in conjunction with equation 4.5 to calculate the mass resolution of the CMS detector for any combination of transverse momenta of the muons as well as any pseudorapidity, so long as it is in the range $\eta \in [0, 2.5]$. However, some assumptions have been made in the process of obtaining these expressions. Therefore, the statistical uncertainties of the final results for $\Delta p_{T,i}$ and k_η^2 can not be fully trusted. The accuracy of the extrapolated mass resolution is assessed in a later section.

4.4. Analysis of the Z resonance

With the detector resolution determined, the Z resonance can finally be analyzed. The approach for the Z analysis makes use of the `MultiFit()` function provided by `kafe2`. It performs simultaneous histogram fits to different subsets of data by combining the likelihood functions of all fits into one and subsequently optimizing it. This allows the splitting of the Z resonance into several different partitions of data, for each of which the mass resolution is calculated with the expression derived above based on its average transverse momenta $\bar{p}_{T,1}$ and $\bar{p}_{T,2}$ as well as pseudorapidity η . For each partition, the assigned mass resolution is then held constant over the fitting procedure, but Γ_Z is fitted to all partitions simultaneously. Therefore, the multi fit returns a single value for Γ_Z as a final result, taking all partitions into account. To increase the accuracy, aid convergence and speed up the multi fit, the data comprising the Z resonance is first filtered to reduce background noise. Additionally, the accuracy of the extrapolated mass resolution is examined.

4.4.1. Model Function for the Z Resonance

Just as in the case of the J/Ψ resonance, fitting the Z resonance requires consideration of both the resonance shape and the surrounding background. The model function used for this purpose is again composed of two components. One describing the resonance and another accounting for the background.

The general structure of the model function is identical to that used for the J/Ψ resonance and is expressed as normalized probability density functions, scaled by an additional parameter N . The overall model function $f(x)$ can be written as:

$$f_Z(x) = N_s \cdot R_Z(x) + N_b \cdot B_Z(x), \quad (4.26)$$

where $N_s \cdot R_Z(x)$ captures the resonance (or signal) and $N_b \cdot B_Z(x)$ models the background.

As mentioned in chapter 2, due to the large decay width of the Z resonance, $R_Z(x)$ can not be approximated as a Gaussian. Rather, the Z resonance follows the shape of a Voigt profile. The Voigt profile is given by the convolution of a Gaussian distribution, reflecting the detector resolution, and a Breit-Wigner distribution describing the undistorted shape of the resonance. However, the integral

$$V(x) \equiv G(x) * B_W(x) = \int_{-\infty}^{+\infty} G(x') \cdot B_W(x - x') dx' \quad (4.27)$$

does not have an analytical solution. Therefore, the model function for the Z resonance uses `scipy`'s implementation of the Voigt profile. It is available under `scipy.special` with the function signature `voigt_profile(x,sigma,gamma)`. In its definition of the Voigt profile, `scipy` uses

$$B_W(x) = \frac{\gamma}{\pi(x^2 + \gamma^2)} \quad (4.28)$$

as the parametrization for the Breit-Wigner distribution. Here `gamma` (γ) represents the half width at half maximum (HWHM) and `sigma` (σ) the detector resolution. In particle physics however, the decay width is ordinarily given by $\Gamma = 2\gamma$, the *full* width at half maximum (FWHM). Additionally `scipy` does not include a separate parameter μ for the mean of the distribution in its definition. To be suitable for a fit to the Z resonance, the definition has to be changed accordingly. This is done with the substitutions

$$\begin{aligned} \gamma &\rightarrow 0.5 \cdot \Gamma \text{ and} \\ x &\rightarrow x - \mu, \end{aligned} \quad (4.29)$$

changing the function call to `voigt_profile(x-mu,sigma,0.5 * Gamma)`. Using these definitions, the fit returns the parameters of interest directly and no additional calculations have to be made.

For the background, the same quadratic model as used in the J/Ψ resonance could be implemented. However, given that the data used for the Z boson analysis is filtered to reduce background noise, a linear model is sufficient to capture the overall trend of the background without introducing unnecessary complexity. Since the Z boson is analyzed using a multi-fit, where the dataset is split into smaller subsets that are then fitted simultaneously, any additional parameters introduced to the model function can affect the execution time significantly. Since the multi-fit is performed using the `MultiFit()` function provided by `kafe2`, this process can not easily be sped up by using multithreading or parallel processing. Moreover, the simplicity of the linear model contributes to the stability and reliability of the fit results. With fewer parameters to estimate, the fitting process is less prone to issues such as over fitting or numerical instability and it is easier to estimate reasonable starting parameters for the fit, preventing multiple executions of the code.

In total, this allows students to more quickly iterate on different choices of starting parameters for the fit or the partitions used to estimate the detector resolution, as well as

find potential bugs in the code. The line is centered at the mean μ of the Z resonance to minimize parameter correlations.

$$p_Z(x) = a(x - \mu) + b \quad (4.30)$$

The background PDF $B_Z(x)$ is then given by the normalized polynomial $p_Z(x)$:

$$B_Z(x) = \frac{p_Z(x)}{\int_{m_0}^{m_1} p_Z(x) dx}. \quad (4.31)$$

The integral in the denominator evaluates to

$$\int_{m_0}^{m_1} p_Z(x) dx = -\frac{1}{2}(m_1 - m_0) \cdot (a(2\mu - m_1 - m_0) - 2b) \quad (4.32)$$

The model function is then given by

$$f_Z(x) = N_s \cdot V(x) + N_b \cdot \frac{p_Z(x)}{\int_{m_0}^{m_1} p_Z(x) dx}. \quad (4.33)$$

4.4.2. Reducing the background in the Z resonance

As mentioned in chapter 4.2, in addition to the momenta and charge of the muons, the dataset also contains the values dxy , dz and $iso3$. These quantities are not strictly necessary for the analysis as the dimuon spectrum can be calculated purely from the muon momenta. However, they do provide information about the quality and importance of the performed measurements. dxy and dz describe the distance of the muons to the primary vertex of the interaction in the transverse plane and in parallel to the beam pipe respectively. Muons that originate from the decay of Z bosons are emitted close to the primary vertex. Therefore, removing muons with a large distance to the primary vertex helps to reduce the background signal in the Z resonance, as they likely do not stem from an interaction of interest. Similarly, $iso3$ quantifies the isolation of muons. A small value of $iso3$ corresponds to an isolated muon. In any given event, numerous particles are produced, and particle jets can form due to quark confinement, leading to multiple particles striking the detector in close proximity. This can lead to a false reconstruction of a particles track through the detector and therefore a false estimate of its momentum. Since Z boson muons are emitted nearly isotropically and are not typically part of jet structures, removing events with poor isolation can effectively reduce background noise without significantly diminishing the dataset. J/Ψ mesons on the other hand are part of jet structures, which is why the same approach can not be used to reduce the background in its resonance.

Removing or reducing the background noise allows for a more accurate fit to the Z resonance as well as a simpler choice for the model function. Ideally, multiple different parameters would be considered simultaneously to determine whether a muon is part of the background noise or the resonance itself. However, to keep this analysis in the scope of a lab course, each parameter is considered individually. To filter out muons with a large distance to the primary vertex (PV), dxy and dz are combined to the total distance ds to the PV using the Pythagorean theorem:

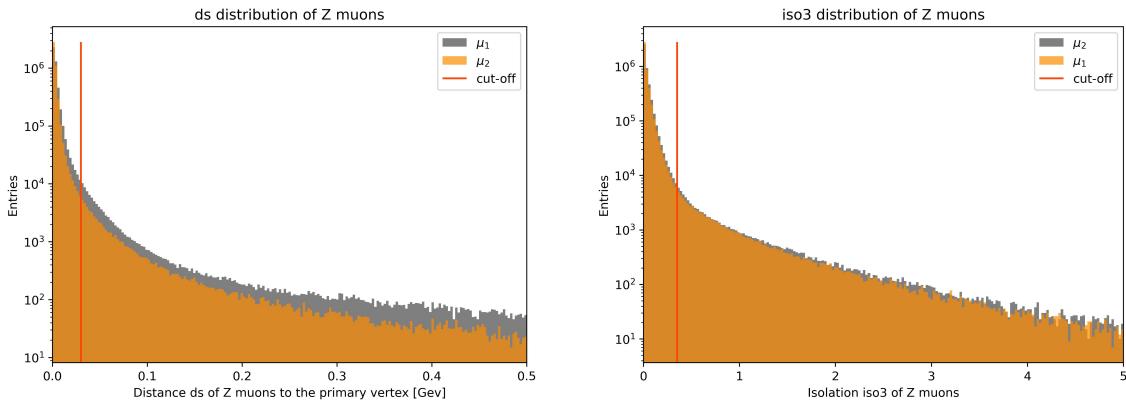
$$ds = \sqrt{(dxy)^2 + (dz)^2} \quad (4.34)$$

Figure 4.11a shows a histogram of the calculated distances ds of the muons μ_i from the Z resonance. The red line at $ds = 0.03$ GeV indicates the cut-off. All events containing at

least one muons with an isolation value greater than the cut-off will be discarded and not used in the subsequent analysis. Choosing a cut off is a balancing act, as a large cut-off barely helps to reduce the noise, while a small cut-off value greatly reduces the available data. Cut-offs are characterized by their efficiency ϵ , which represents the percentage of events kept.

$$\epsilon = 100 \cdot \frac{\text{number of events kept}}{\text{total number of events}} \quad (4.35)$$

The *ds* Filter shown in 4.11a has an efficiency of $\epsilon_{ds} = 93.71\%$. A similar filter is applied the isolation of the muons, removing all events where one or both muons have an *iso3* value greater than the cut-off. Figure 4.11b shows the corresponding histogram. This time, a cut-off of $iso3 = 0.35$ is chosen resulting in an efficiency of $\epsilon_{iso3} = 97.29\%$. Applying both filters simultaneously results in an efficiency of $\epsilon_{tot} = 91.85\%$, leaving around 4.3 out of the 4.6 million events for further analysis.



(a) Distribution of the distance from the primary vertex of Z muons in the range of $ds \in [0, 0.5]$ GeV. Events with ds greater than the cut-off are discarded.

(b) Distribution of the Isolation $iso3$ of Z muons in the range of $iso3 \in [0, 5]$. Events with an isolation greater than the cut-off are discarded.

Figure 4.11.: Filters used to reduce the background noise in the Z resonance.

4.4.3. Fitting the Z resonance

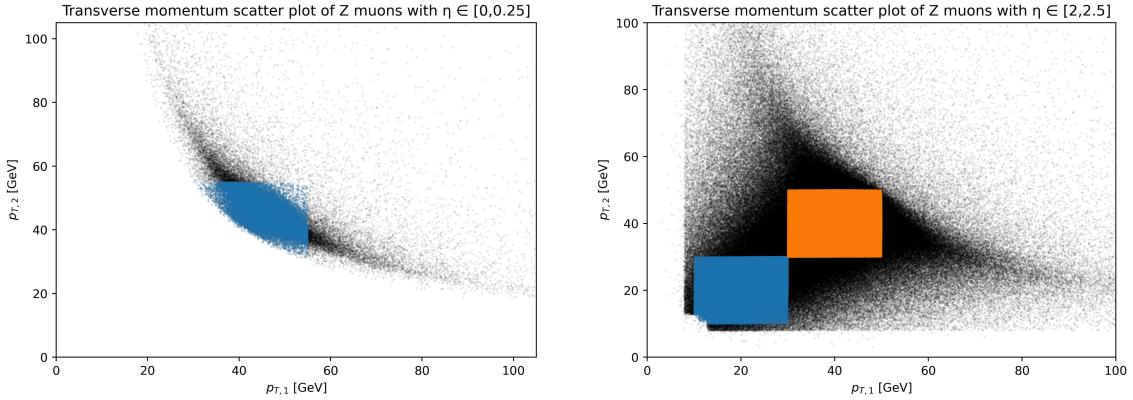
To determine the decay width Γ_Z , the Z resonance is divided into the same bins of η as the J/Ψ resonance in section 4.3.2. Each bin only contains events, in which both muons have a pseudorapidity within the bin edges. For every bin, one to two partitions are created depending on the amount of data available. Two such such partitionings are shown in figure 4.12. The partitions for all bins are given in appendix B.

For each partition, the mass resolution is calculated using

$$\Delta M_Z = \frac{M_Z}{2} \sqrt{\left(\frac{\Delta p_{T,1}(\overline{p_{T,1}})}{\overline{p_{T,1,Z}}} \right)^2 + \left(\frac{\Delta p_{T,2}(\overline{p_{T,2}})}{\overline{p_{T,2,Z}}} \right)^2 + k_\eta^2(\eta_Z)}, \quad (4.36)$$

where $\overline{p_{T,1}}$ and $\overline{p_{T,2}}$ are the average transverse momenta of the muons within the partition and $k_\eta^2(\eta_Z)$ is the appropriate correction term for the bin. The mass M_Z of the Z boson is again estimated by the average of the reconstructed masses of each dimuon pair making up the Z resonance. It is given by

$$M_Z = (90.6125 \pm 0.0015) \text{ GeV}, \quad (4.37)$$



(a) Partitioning of the $\eta \in [0, 0.25]$ bin of the Z resonance (b) Partitioning of the $\eta \in [2, 2.5]$ bin of the Z Resonance

Figure 4.12.: Examples of different partitionings of the Z resonance based on the amount of data available in each η bin.

which is noticeably smaller than the literature value of

$$M_{Z,lit} = (91.1880 \pm 0.0020) \text{ GeV} \quad [4], \quad (4.38)$$

provided by the Particle Data Group. This is again primarily the consequence of the final state radiation, biasing the reconstructed masses towards lower values. Since the Z resonance has significantly more entries than the J/Ψ resonance, this systematic bias is even more apparent here.

4.4.3.1. Testing the accuracy of the mass resolution extrapolation

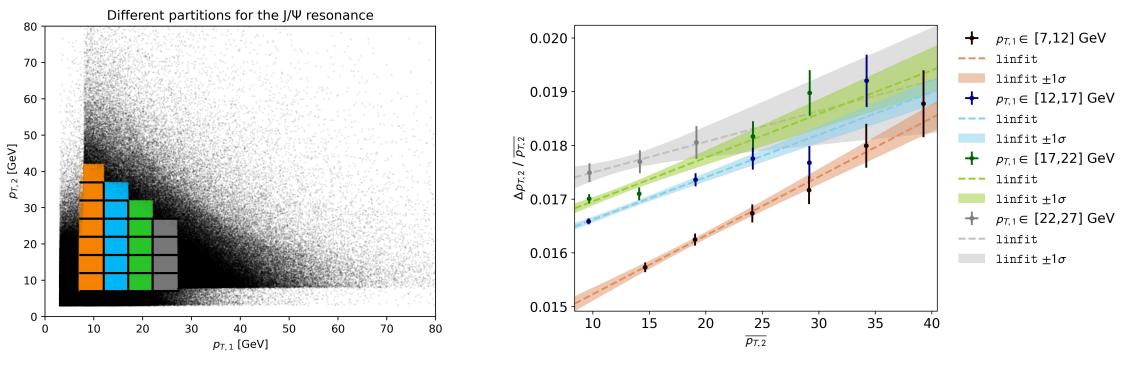
In the previous sections, all of the necessary terms to calculate the mass resolution of the detector have been determined. During this process, the uncertainties of the used fit parameters have been propagated via Gaussian error propagation to the final results of k_η^2 as well as the expressions for $\Delta p_{T,i}$. As can be seen in tables 4.2 and 4.3, the uncertainty for the parameter σ that has been used in the calculations is very small compared to its value. This is because each bin and partition of the data has specifically been chosen in such a way to contain lots of entries to ensure a reliable fit of the model function. With an increase in data, the statistical uncertainties get smaller, leading to the low parameter uncertainties seen in the tables above. However, several assumptions were made in the previous sections to simplify the analysis. Namely, the primary assumptions made are that the transverse momentum resolutions $\Delta p_{T,i}$ are purely a function of $p_{T,i}$ and that the $k_\eta(\eta)$ term properly captures any dependence on the pseudorapidity of the muons. Additionally, the systematic effects of final state radiation as well as distortions of the resonances caused by misidentified particles have been ignored to allow for a data-only approach to the analysis to make it suitable for a lab course. Therefore, the statistical uncertainties of the parameters obtained by fits are not a good indication of the accuracy of the mass resolution extrapolation.

In this section, the validity of this approach is tested. First, the influence of the choice of data partitions used to calculate the transverse momentum resolution is examined. As can be seen in figures 4.6b and 4.10, due to the necessity of one of the transverse momenta being held constant, much of the available dataset remains unused. The choice of the subsets used is a product of both convenience and accuracy. Given that the shape if the dataset in the transverse momentum scatter plots is roughly triangular, using only the

$p_{T,1}$ range [GeV]	a [1/keV]	Δa [1/keV]	$b \cdot 10^3$	$\Delta b \cdot 10^3$	μ [GeV]	$-2 \ln \mathcal{L}_R/\text{ndf}$
[7,12]	109.351	11.781	17.061	0.111	26.722	0.22
[12,17]	79.000	8.532	17.666	0.103	23.244	1.36
[17,22]	81.340	15.041	17.715	0.124	19.283	2.26
[22,27]	57.041	35.730	17.746	0.137	14.341	0.05

Table 4.4.: Fit Results of the transverse momentum resolution extrapolation $\Delta p_{T,2}$ for different choices of partitions

subsets along the legs of the triangle ($p_{T,i} = 12$ GeV) allows for the most possible partitions being created. Additionally the intersection point (or intersection partition) can be used to calculate the necessary constant c for both directions simultaneously. However, it can not be ruled out that a different choice of subsets leads to different result. Second, to ensure that the estimation of the mass resolution of the detector produces reasonable results at the transverse momenta of the Z boson, it is compared to suitable literature values.



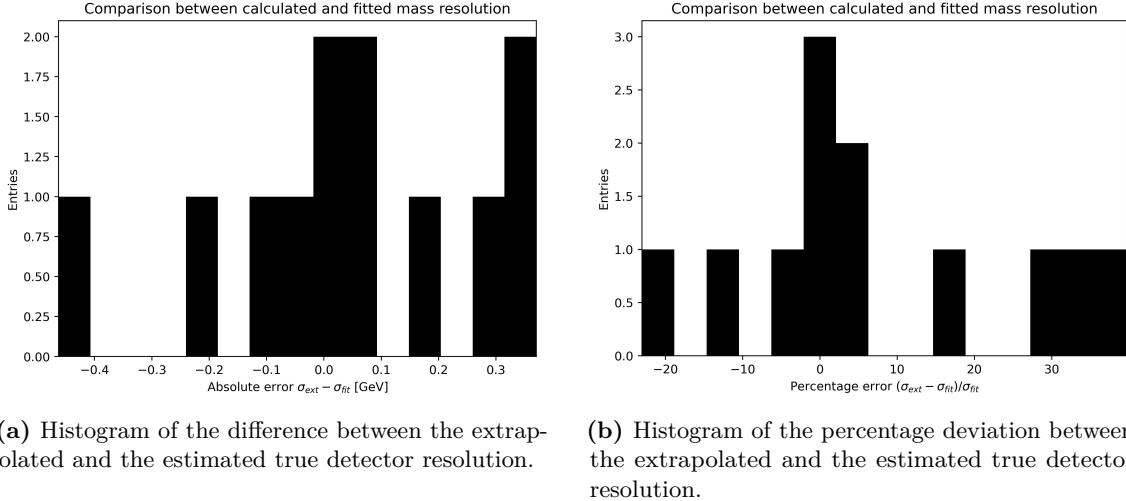
(a) Partitions for the J/Ψ resonance for different choices of $p_{T,1}$. (b) Transverse momentum extrapolation for different choices of partitions.

Figure 4.13.: Different choices for J/Ψ partitions and their effect on the extrapolation on the transverse momentum resolution $\Delta p_{T,2}$.

To test whether the selection of partitions influence the extrapolation of the transverse momentum resolution, different partitions are created over the J/Ψ dataset. Several choices for a constant $p_{T,1}$ can be seen in figure 4.13a. The different axes, indicated by different colors, are centered at $p_{T,1} = [9.5, 14.5, 19.5, 24.5]$ GeV. Each partition is given a height and width of 5 GeV. For each axis, the respective partition on the diagonal is used to calculate its constant c . Figure 4.13b shows the resulting transverse momentum extrapolation for each axis. Both plots are color coded, so each fit corresponds to the partitions of the same color. The fits seen in figure 4.13b differ slightly in their height and slope. With an increase in $p_{T,1}$, the slope of the line tends to decrease and it is slightly shifted upwards. However, the difference in the extrapolated resolution is roughly 0.3% at most. For the linear fit, the same model function is used as before. The fit results are shown in table 4.4. Since the partitions of Z muons are created between $p_T \in [10, 60]$ GeV, the exact choice of partitions used has little effect on the final result. However, the number of available partitions, and therefore data points for the extrapolation, decreases with higher transverse momenta. The same analysis is performed for $\Delta p_{T,1}$, giving similar results which are shown in appendix C.

To estimate the accuracy of the mass resolution extrapolation it can be compared to suitable literature values. For this purpose, a fit the each partition of the Z resonance is performed. By fixing Γ_Z to its literature value of $\Gamma_Z = (2.4955 \pm 0.0023)$ GeV [4] during the fitting process, an estimate of the true detector resolution σ can be obtained for each partition. Figure 4.14a shows the difference between the extrapolated detector resolution and the

ones obtained with the fit. Figure 4.14b shows the deviation in percent.



(a) Histogram of the difference between the extrapolated and the estimated true detector resolution.

(b) Histogram of the percentage deviation between the extrapolated and the estimated true detector resolution.

Figure 4.14.: Comparison between the extrapolated and estimated true resolution of the detector

As can be seen, many estimates are close to the true value, but there are outliers up to about -20% and +40%. This information can be used to gauge the accuracy of the extrapolation using the root mean squared error (RMSE), given by

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_i^N (y_i - \hat{y})^2}, \quad (4.39)$$

where y_i are the literature values and \hat{y}_i are the predicted values. Therefore, the error of the mass extrapolation is given by

$$\Delta\sigma = \sqrt{\frac{1}{N} \sum_i^N (\sigma_{fit,i} - \sigma_{ext,i})^2} = 0.22 \text{ GeV}, \quad (4.40)$$

which is significantly higher than the statistical uncertainties obtained above. The relative error can be calculated using the normalized root mean square error NRMSE.

$$\text{NRMSE} = \frac{\text{RMSE}}{\bar{y}} = \frac{1}{\sigma_{fit}} \cdot \sqrt{\frac{1}{N} \sum_i^N (\sigma_{fit,i} - \sigma_{ext,i})^2} = 0.176 \doteq 17.6\% \quad (4.41)$$

All values of σ_{fit} and σ_{ext} are listed in table 4.5, as well as which partition they were obtained from. The table shows that most outliers stem from low and high η bins, while the extrapolation is fairly accurate in the range of $\eta \in [1,2]$. The value of $\Delta\sigma$ is now passed to the fit as the uncertainty of the detector resolution σ .

4.4.3.2. Determining the decay width Γ_Z

For each partition, a histogram fit is created, but not yet executed, and the calculated value for the mass resolution, including its uncertainty obtained in the previous section, is fixed. The histograms each have 30 bins in the range of $M \in [82, 100]$ GeV. The list of fits is then passed to the `MultiFit()` function to be fitted. Only the parameters Γ and μ are fitted to every histogram simultaneously. The parameters N_s and N_b , quantifying the

η bin	$\overline{p_{T,1}}$ [GeV]	$\overline{p_{T,2}}$ [GeV]	σ_{fit} [GeV]	σ_{ext} [GeV]
[0,0.25]	45.528	45.505	0.931	1.301
[0.25,0.5]	44.625	44.632	0.990	1.308
[0.5,0.75]	43.049	43.062	1.018	1.304
[0.75,1]	41.235	41.235	1.113	1.312
[1,1.25]	39.867	39.874	1.269	1.348
[1.25,1.5]	39.591	39.610	1.366	1.375
[1.5,1.75]	27.983	27.995	1.208	1.283
[1.5,1.75]	42.385	42.376	1.395	1.382
[1.75,2]	21.927	21.949	1.303	1.276
[1.75,2]	40.510	40.518	1.491	1.400
[2,2.5]	21.556	21.661	1.648	1.433
[2,2.5]	40.387	40.387	2.008	1.546

Table 4.5.: Results for the extrapolated and fitted mass resolution for each partition of the Z resonance.

amount of data in each partition, as well as a and b , describing the background, are fitted only to their own respective partition.

However, `kafe2` does not provide the user the option to manually toggle which parameters are fitted only to one histogram and which variables are fitted over multiple histogram simultaneously. Instead, `kafe2` infers this information indirectly via the argument names of the python function that is assigned as a model function to each fit. If an argument of the same name appears in the model function of different fits, it is simultaneously fitted to all of them.

This is a problem, as the same model function for the Z resonance is supposed to be fitted to every partition. If the same python function containing the Z model were to be assigned as a model function for each partition, they would all share the same parameter names. Consequently, *all* parameters would be fitted to *all* partitions simultaneously.

To avoid this problem, each partition is assigned its own Python function with the same Z resonance model defined in the section 4.4.1. The parameter names for Γ and μ are shared over all python functions, to ensure they are fitted to every histogram simultaneously. However, every other parameter is assigned a unique name for each partition. This is achieved by adding the unique suffix `N_subXY` to the parameter name. Here, N indexes the different eta bins. $N = 0$ corresponds to the bin $\eta \in [0, 0.25]$, $N = 1$ corresponds to $\eta \in [0.25, 0.5]$ and so on. X and Y are the coordinates of the subset within the grid of partitions created in each bin. Therefore, the upper right partition shown in figure 4.12b is assigned the suffix `8_sub11`.

Figure 4.15 shows the resulting fit to the to the partition in the first η bin depicted in figure 4.12a. All fits are shown in the appendix D. The obtained Value for the Z decay width is

$$\Gamma_Z = (2.993 \pm 0.015) \text{ GeV}. \quad (4.42)$$

The literature value provided by the particle data group is

$$\Gamma_{Z,\text{lit}} = (2.4955 \pm 0.0023) \text{ GeV} [4]. \quad (4.43)$$

Therefore, the estimated decay width of the Z boson is about 0.4975 GeV higher than the literature value, which corresponds to a deviation of around 19.94%. This is an improvement over the current analysis, which yields a result of

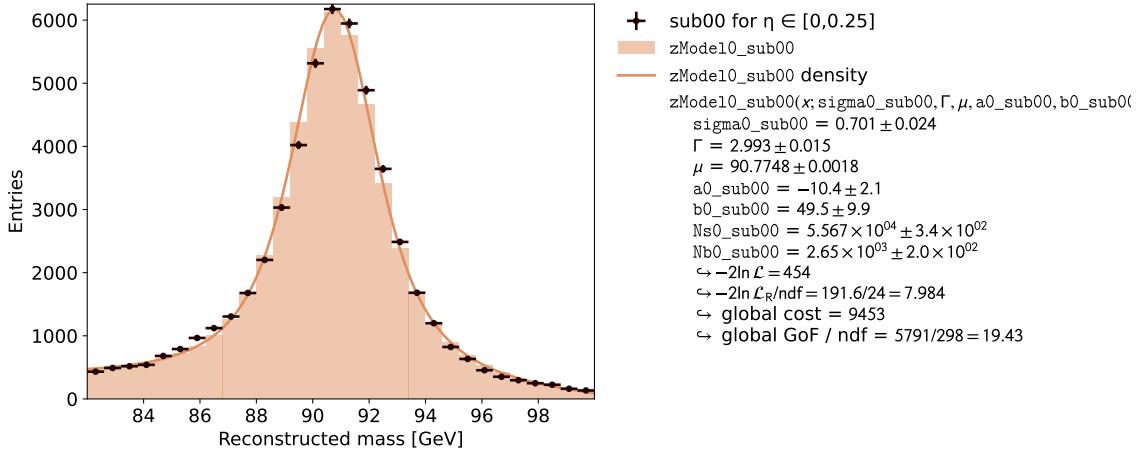


Figure 4.15.: Fit results of the partition in the $\eta \in [0, 0.25]$ bin as part of the multi fit to the Z resonance.

$$\Gamma_{Z,\text{curr}} = (3.740 \pm 0.148) \text{ GeV} \quad (4.44)$$

when corrected for the missing factor of two, but is still a significant deviation from the literature value. Changing the partitions used in the calculation of k_η , the transverse momentum resolution $\frac{\Delta p_T}{p_T}$, as well as the data used for the Z-fit does little to change this result. The fit always converges near the $\Gamma \approx 3$ GeV mark. Interestingly, when repeating the multi fit procedure using the ‘true’ values for the mass resolution along with their uncertainties, obtained with the literature value of Γ_Z in 4.4.3.1, the fit returns a result of

$$\Gamma_Z = (2.590 \pm 0.009) \text{ GeV}. \quad (4.45)$$

This indicates that the literature value of the Z decay width is not a good optimum for the likelihood, even when provided with ideal parameters with minimal uncertainties. This is likely due to the aforementioned final state radiation artificially broadening the Z resonance. The rather large uncertainties for the mass resolution obtained in 4.4.3.1 provide enough ‘wiggle room’ for the multi fit to accommodate this broadened peak, leading to large values for the Z decay width.

5. Implementation of the analysis in a lab course

The analysis presented in this thesis is quite exhaustive as it aims account for every parameter influencing the detector resolution and therefore the final result of Γ_Z . Consequently, a lot of code is required to create, manage and analyze the many partitions of the dataset necessary to perform the analysis as shown. Additionally, even with modern hardware, executing the code can take on the order of hours. This is largely due to the multi-fit performed at the end to determine the decay width of the Z boson. This fitting procedure can not easily be sped up via multithreading or parallel processing as it is performed via a single function call to `MultiFit()` from the library `kafe2`. Furthermore, the lab course is not intended purely as an exercise in calculating Γ_Z , but rather as a general introduction to data analysis in a physics framework. Consequently, some time is spent familiarizing the students with the data as well as the tools and libraries used to manipulate and perform calculations on large datasets. Therefore, some adjustment are necessary to fit this analysis within the scope of the lab course. Namely, the programming effort as well as the execution time of the analysis must be reduced.

To limit the programming effort, some concessions are necessary to the amount of parameters considered for estimating the detector resolution, as it constitutes the majority of the analysis. Since the extrapolation of the momentum resolution towards higher values of p_T is necessary to properly capture the mass resolution at the Z boson, the only way to cut down the analysis is to eliminate the need to incorporate the influence of the pseudorapidity of the muons. However, simply ignoring this dependency is likely to produce worse results for Γ_Z and give the students an incomplete understanding of the subject. A more suitable solution is to simply reduce the dataset to events, in which both muons have a negligible pseudorapidity. Comparing the size of the correction k_η , shown in figure 4.5, with the contributions from the transverse momentum resolution, shown in figures 4.8 and 4.9, the uncertainty in p_T is the dominating contribution for small pseudorapidities. Therefore, limiting the dataset to events with muons with small values for η_i eliminates the need to determine k_η , as its influence on the detector resolution can be neglected.

A reduction of the runtime of the analysis can be achieved in three main ways. First, a smaller dataset can be used. This significantly reduces the time required to perform the fits, as well as to load the dataset into python, which is especially relevant for computers with hard disk drives. Limiting the dataset to events with small η_i therefore serves two purposes. Second, simpler model functions can be chosen for the J/Ψ resonance. While each fit performed in the determination of the transverse momentum resolution does not take much time on its own, the aggregate time of all fits is significant. This is especially relevant for more complex model functions, as getting the fits to converge can take many iterations using different starting parameters. Additionally, whether a linear or quadratic polynomial is chosen to model the background noise has little impact on the result of the standard deviation σ obtained as the fit result. Therefore a linear model function is sufficient in a

physics lab course, as iterating on starting parameters for proper convergence does little to improve the students understanding of the subject matter. Third, and most important, is a reduction in execution time of the multi-fit procedure, as it single-handedly accounts for the majority of runtime. This is largely due to the fact that with each partition added to the multi-fit, its runtime roughly increases by a factor of 2. With the many partitions used in this analysis over multiple bins of η , the resulting fit to the Z resonance can take on the order of an hour, even on modern hardware. Consequently, the multi-fit can be sped up significantly by reducing the amount of subsets considered. This can of course be done by selecting only one partition for each bin of η , or simply by considering fewer bins of η to begin with. However, this would likely still result in a multi-fit runtime in the tens of minutes. The best improvement are likely to be seen when reducing the dataset to only include events with small values of η as mentioned previously, as it corresponds to only considering a single η bin. If the bin is chosen large enough, three to four partitions can be placed within it, reducing the runtime to a practicable minimum.

Appendix

A. NanoAOD to CSV conversion

The following code was used to convert the datasets from the Run2012B [17] and Run2012C [18] available on the CERN Open Data Portal from `.root` files storing the data in the nanoAOD format to human readable CSV files.

```
import pandas as pd
import uproot
import numpy as np
import awkward as ak

#open files
events1 = uproot.open("RootData/Run2012B_DoubleMuParked.root:Events")
events2 = uproot.open("RootData/Run2012C_DoubleMuParked.root:Events")

#create boolean mask to filter out events with exactly 2 myons
mask1 = events1['nMuon'].array(library ="np") == 2
mask2 = events2['nMuon'].array(library ="np") == 2

#create new df to copy values into
colNames = ['pt', 'eta', 'phi', 'Q', 'dxy', 'dz', 'Iso3']
nCols = len(colNames)
df = pd.DataFrame(columns = pd.MultiIndex.from_arrays([nCols*[ 'mu1' ]
+nCols*[ 'mu2' ], colNames+colNames]))

#dictionary to translate from root column names to df column names
#names from root file
rootCols = ['Muon_pt', 'Muon_eta', 'Muon_phi', 'Muon_charge', 'Muon_dxy',
'Muon_dz', 'Muon_pfRelIso03_all']
names = dict(zip(rootCols, colNames))

for c in rootCols:
    data1 = ak.to_numpy(ak.Array.__getitem__(events1[c].array(), mask1))
    data2 = ak.to_numpy(ak.Array.__getitem__(events2[c].array(), mask2))
    data = np.concatenate((data1, data2)).T
    df[ "mu1", names[c]] = data[0]
    df[ "mu2", names[c]] = data[1]
df = df[(df.mu1.Iso3 >= 0) & (df.mu2.Iso3 >= 0)]
df = df.sort_index(axis=1)

df.to_csv("CSVData/Run2012BC_DoubleMuons_prefiltered.zip", index = False)
```

In addition to converting the file, some basic filtering is performed. This includes the removal of all events which do not contain exactly two muons as well as removing events with negative $Iso3$ values. The `.root` files are opened with `uproot` [19] and `awkward` [20] is used to read the data.

B. Partitions of the Z resonance

For the analysis of the Z boson, the data comprising the Z resonance is split into different bins of η . Since the distribution of events within the transverse momentum scatter plot differs between bins, different partitions are created to adequately cover each bin.

For the first six bins of η ($[0,0.25]$, $[0.25,5]$, ..., $[1.25,1.5]$), a single partition is sufficient. It covers all events in the range $p_{T,i} \in [25,55]$ GeV. They are shown in figure B.3.

For the $\eta \in [1.5,1.75]$ bin, two partitions are created. The first contains all events in the range $p_{T,i} \in [20,35]$ GeV, the second all events within $p_{T,i} \in [35,50]$ GeV. It is shown in figure B.1.

For the bins $\eta \in [1.75,2]$ and $\eta \in [2,2.5]$, two partitions are created. The first in the range of $p_{T,i} \in [10,30]$ GeV, the second in the range of $p_{T,i} \in [30,50]$ GeV. They can be seen in figure B.2.

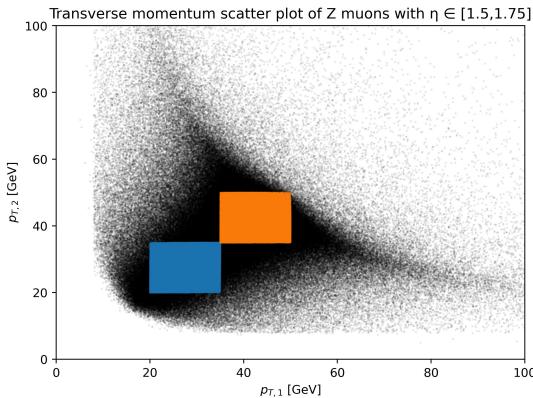
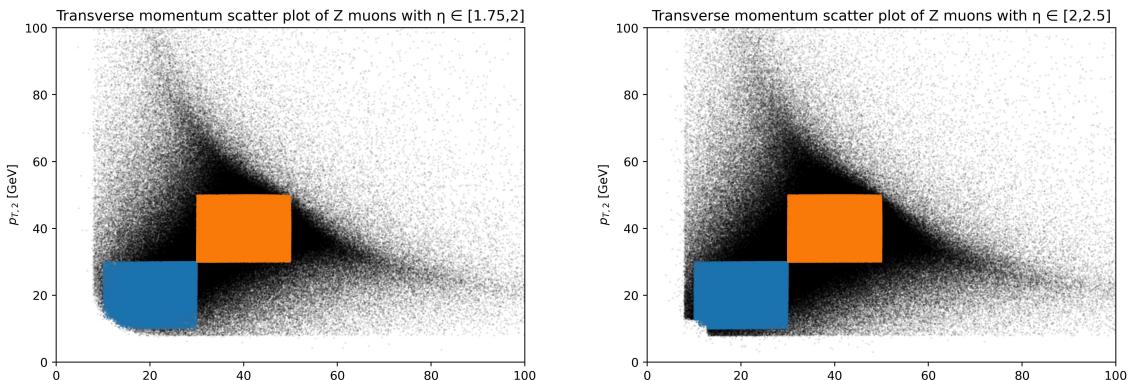


Figure B.1.: Partitions of the Z resonance in the $\eta \in [1.5,1.75]$ bin



(a) Partitions of the Z resonance in the $\eta \in [1.75,2]$ bin

(b) Partitions of the Z resonance in the $\eta \in [2,2.5]$ bin

Figure B.2.: Partitions of the Z resonance for the η bins $[1.75,2]$ and $[2,2.5]$

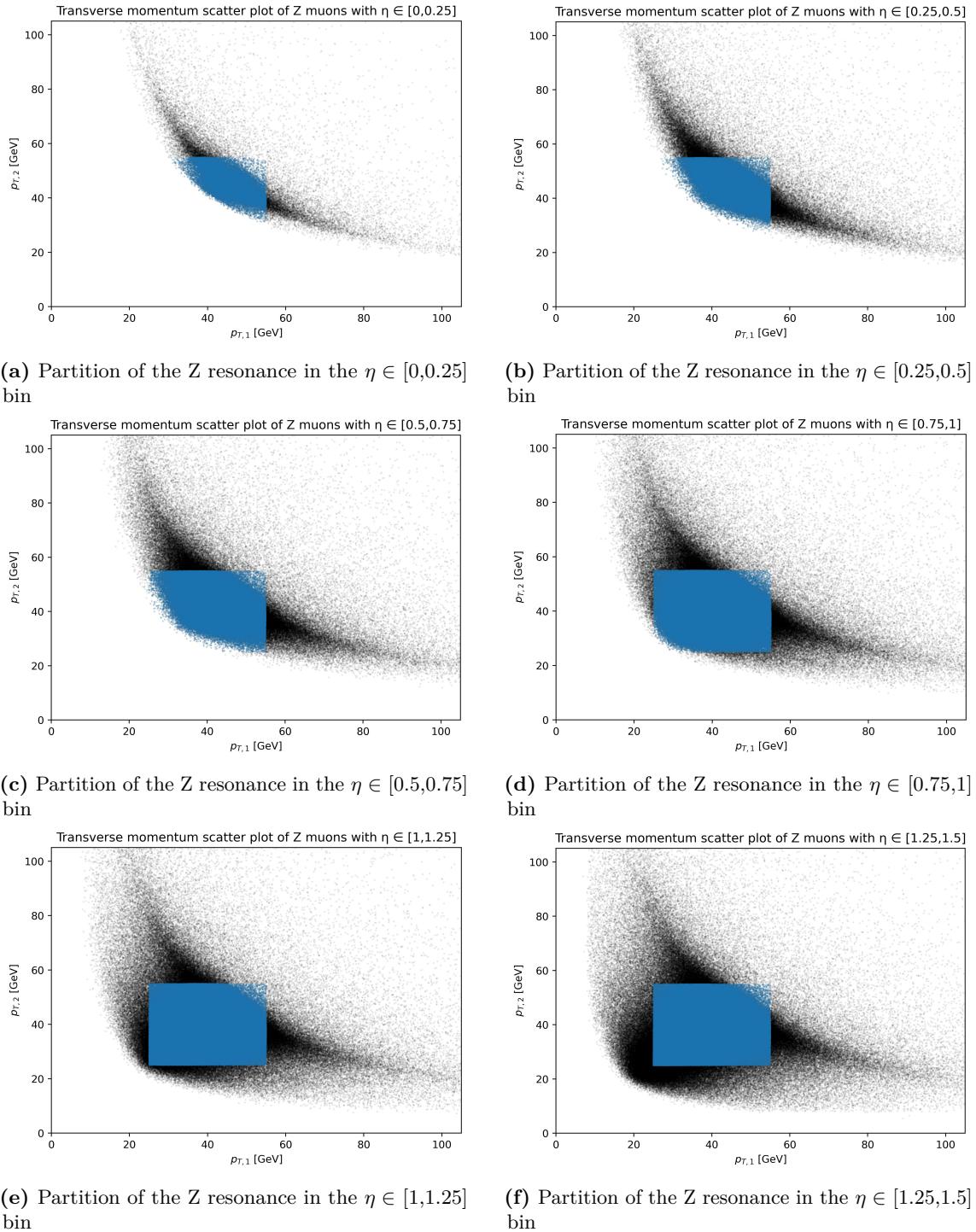


Figure B.3.: Partitions of the Z resonance for the η bins up to $\eta \in [1.25,1.5]$

C. Influence of the choice of partitions for the resolution $\Delta p_{T,1}$

The same analysis as performed in chapter 4.4.3.1 for the influence of the choice of partitions used to calculate the momentum extrapolation $\Delta p_{T,2}$ can be performed for $\Delta p_{T,1}$. The same sets of partitions are chosen but flipped on their side accordingly. The resulting extrapolations follow a similar pattern, with a decrease in the slope of the line accompanied by an upwards shift for larger values of $p_{T,2}$. The different partitions and resulting fits are shown in figures C.4a and C.4b respectively. The fit results are listed in table C.1.

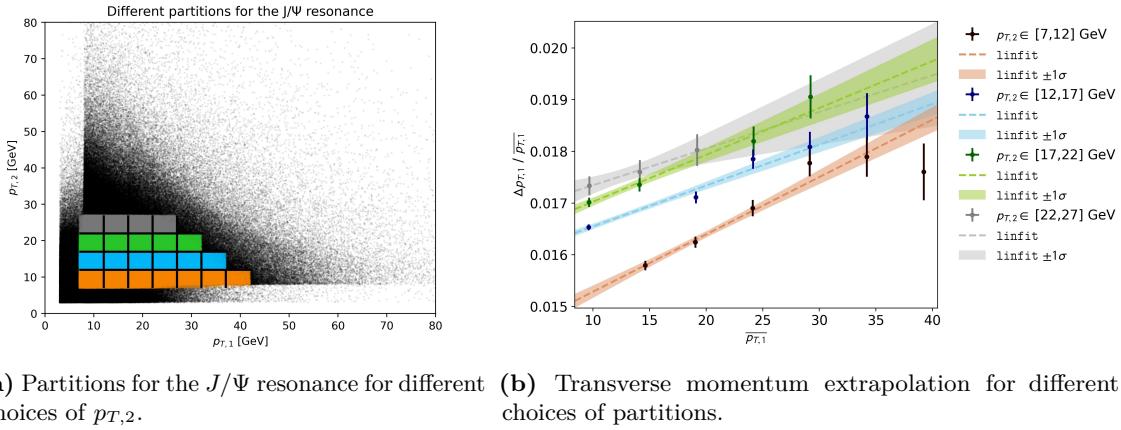


Figure C.4.: Different choices for J/Ψ partitions and their effect on the extrapolation on the transverse momentum resolution $\Delta p_{T,1}$.

$p_{T,1}$ range [GeV]	a [1/keV]	Δa [1/keV]	$b \cdot 10^3$	$\Delta b \cdot 10^3$	μ [GeV]	$-2 \ln \mathcal{L}_R/\text{ndf}$
[7,12]	110.886	11.292	17.139	0.106	26.723	1.30
[12,17]	78.988	8.087	17.594	0.096	23.244	1.04
[17,22]	90.646	14.998	17.868	0.125	19.289	0.40
[22,27]	70.894	36.273	17.652	0.141	14.338	0.04

Table C.1.: Fit Results of the transverse momentum resolution extrapolation $\Delta p_{T,1}$ for different choices of partitions

D. Fits to the Z resonance

Shown here are all the individual plots from the multi fit performed on the Z resonance. Since the decay width Γ_Z is fitted to all partitions, its final result the same across all fits.

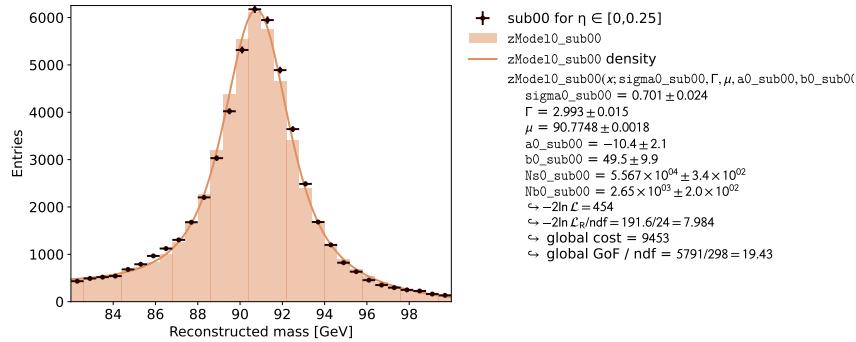


Figure D.5.: Fit results of the partition in the $\eta \in [0,0.25]$ bin as part of the multi fit to the Z resonance.

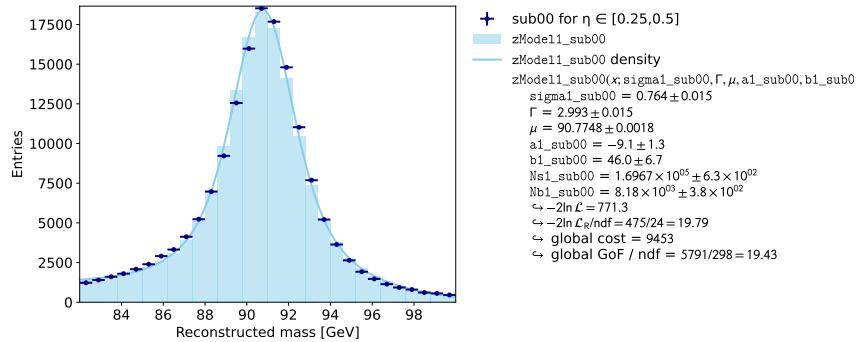


Figure D.6.: Fit results of the partition in the $\eta \in [0.25,0.5]$ bin as part of the multi fit to the Z resonance.

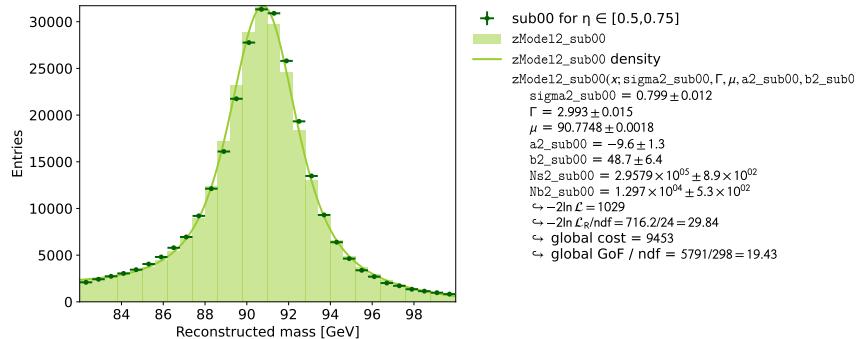


Figure D.7.: Fit results of the partition in the $\eta \in [0.5,0.75]$ bin as part of the multi fit to the Z resonance.

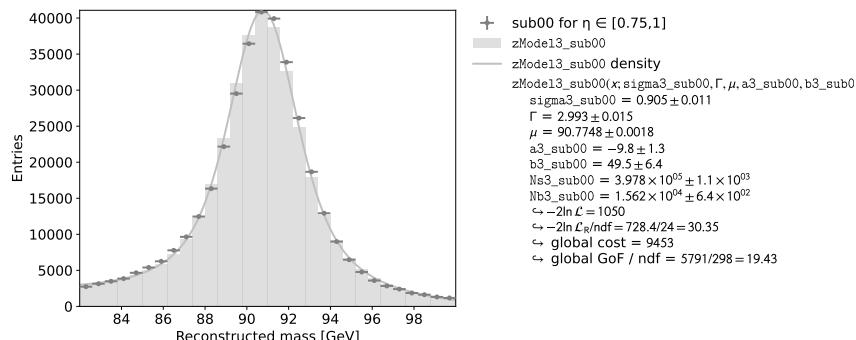


Figure D.8.: Fit results of the partition in the $\eta \in [0.75,1]$ bin as part of the multi fit to the Z resonance.

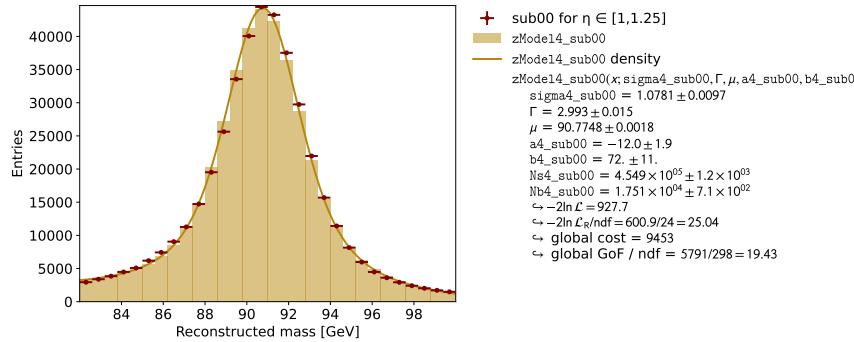


Figure D.9.: Fit results of the partition in the $\eta \in [1,1.25]$ bin as part of the multi fit to the Z resonance.

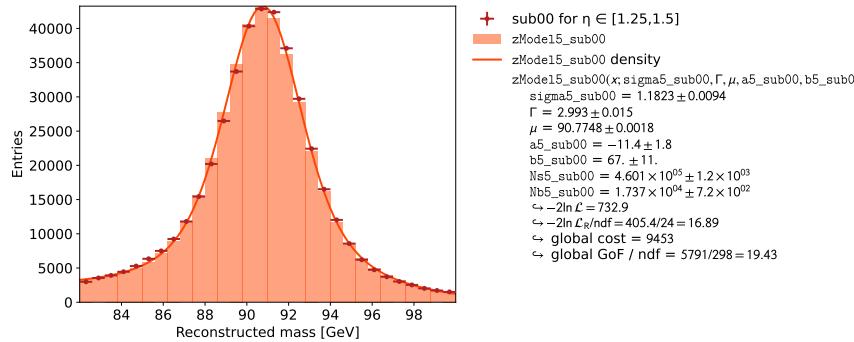


Figure D.10.: Fit results of the partition in the $\eta \in [1.25,1.5]$ bin as part of the multi fit to the Z resonance.

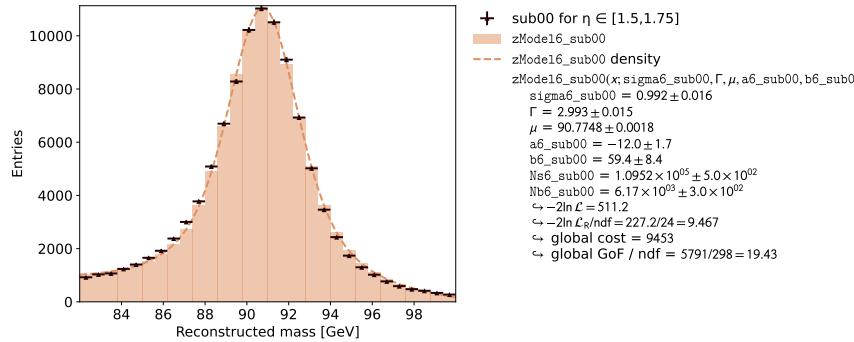


Figure D.11.: Fit results of the bottom left partition in the $\eta \in [1.5,1.75]$ bin as part of the multi fit to the Z resonance.

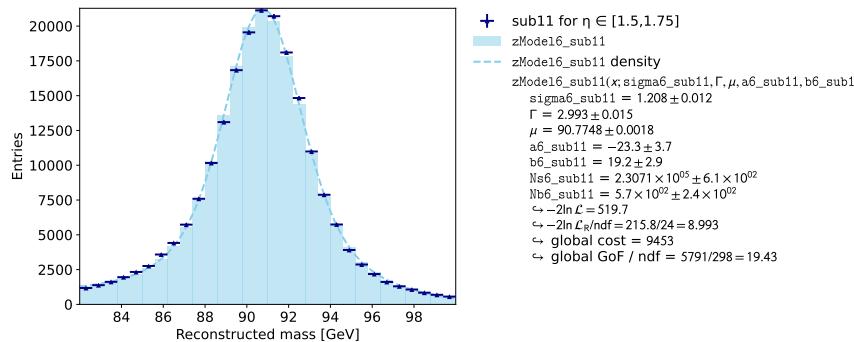


Figure D.12.: Fit results of the top right partition in the $\eta \in [1.5,1.75]$ bin as part of the multi fit to the Z resonance.

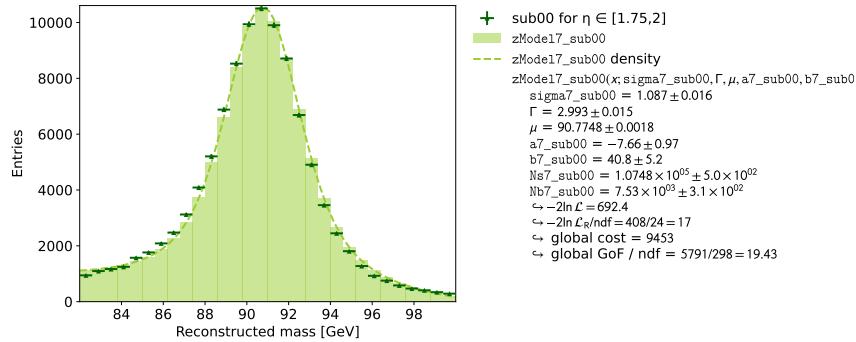


Figure D.13.: Fit results of the bottom left partition in the $\eta \in [1.75, 2]$ bin as part of the multi fit to the Z resonance.

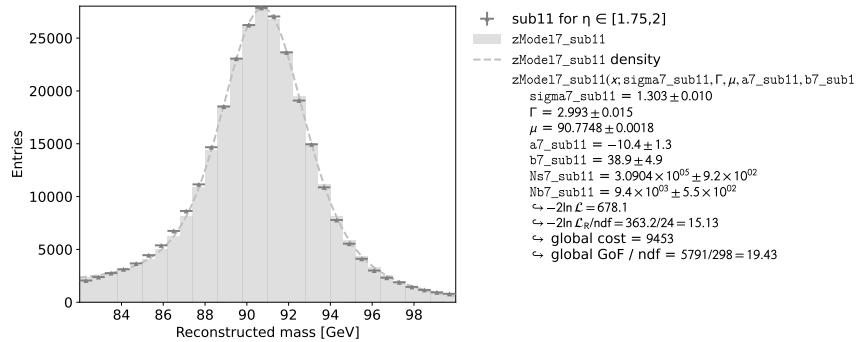


Figure D.14.: Fit results of the top right partition in the $\eta \in [1.75, 2]$ bin as part of the multi fit to the Z resonance.

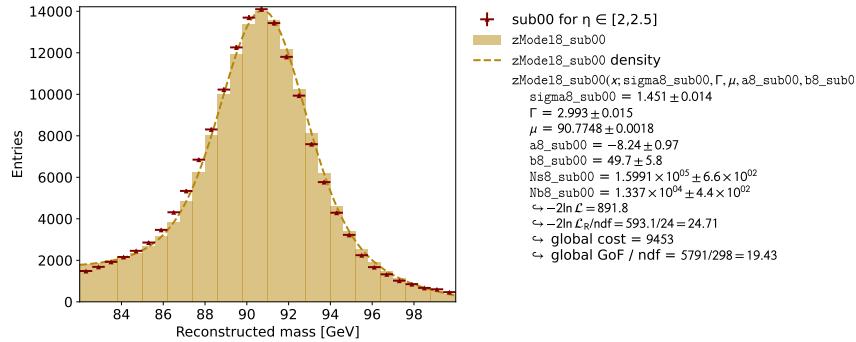


Figure D.15.: Fit results of the bottom left partition in the $\eta \in [2, 2.5]$ bin as part of the multi fit to the Z resonance.

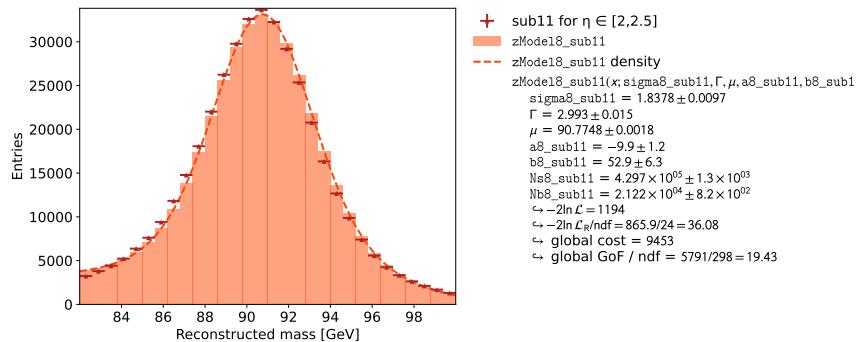


Figure D.16.: Fit results of the top right partition in the $\eta \in [2, 2.5]$ bin as part of the multi fit to the Z resonance.

Bibliography

- [1] Vlimant Jean-Roch. CMS and the machine learning at the forefront of data management. <https://cms.cern/news/cms-and-machine-learning-forefront-data-management>; last accessed on 09. July 2024.
- [2] Rene Brun and Fons Rademakers. Root - an object oriented data analysis framework, Proceedings AIHENP'96 Workshop, Lausanne, Sep. 1996. <https://root.cern/>; last accessed 09. July 2024.
- [3] The Large Hadron Collider. 2014. <https://cds.cern.ch/record/1998498>; last accessed 09. July 2024.
- [4] Particle Data Group and R L Workman et al. Review of Particle Physics. *Progress of Theoretical and Experimental Physics*, 2022(8):083C01, 08 2022. https://pdg.lbl.gov/2024/listings/contents_listings.html; last accessed 09. July 2024.
- [5] NumPy Developers. Numpy. <https://numpy.org/>. Version 1.26.4, NumPy.
- [6] Charles R. Harris et al. Array programming with NumPy. *Nature*, 585(7825):357–362, September 2020. DOI: 10.1038/s41586-020-2649-2.
- [7] J. D. Hunter. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9(3):90–95, 2007. DOI: 10.1109/MCSE.2007.55.
- [8] John D. Hunter. Matplotlib: A 2d graphics environment. <https://matplotlib.org>, 2007. Version 3.8.0, DOI: 10.5281/zenodo.8347255.
- [9] Wes McKinney. Data Structures for Statistical Computing in Python. In Stéfan van der Walt and Jarrod Millman, editors, *Proceedings of the 9th Python in Science Conference*, pages 56 – 61, 2010. DOI: 10.25080/Majora-92bf1922-00a.
- [10] The pandas development team. pandas-dev/pandas: Pandas. <https://pandas.pydata.org/>, February 2020. Version 2.1.4, <https://doi.org/10.5281/zenodo.3509134>.
- [11] Pauli Virtanen et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020. DOI: 10.1038/s41592-019-0686-2.
- [12] Pauli Virtanen et al. Scipy. <https://scipy.org>, 2020. Version 1.11.4, DOI: 10.5281/zenodo.10155614.
- [13] Jim Pivarski. Vector: A pure python library for physics vectors. <https://github.com/scikit-hep/vector>, 2020. Version 1.4.0 DOI: 10.5281/zenodo.7054478.
- [14] Eric O. Lebigot. Uncertainties: A python package for calculations with uncertainties. <http://pythonhosted.org/uncertainties/>. Version 3.1.7.
- [15] kafe2 Developers. Karlsruhe fit environment 2. <https://github.com/PhilFitters/kafe2>, 2020. Version 2.8.3.

- [16] L Barnard et al. CMS Event Display. <https://opendata.cern.ch/visualise/events/cms>.
- [17] Stefan Wunsch. Run2012B_DoubleMuParked dataset in reduced NanoAOD format for education and outreach. *CERN Open Data Portal*, 2021. DOI: 10.7483/OPENDATA.CMS.04XV.ESBR <https://opendata.cern.ch/record/12365>; last accessed 18. August 2024.
- [18] Stefan Wunsch. Run2012C_DoubleMuParked dataset in reduced NanoAOD format for education and outreach. *CERN Open Data Portal*, 2021. DOI: 10.7483/OPENDATA.CMS.86ZS.7G78 <https://opendata.cern.ch/record/12366>; last accessed 18. August 2024.
- [19] Jim Pivarski et al. Uproot. <https://github.com/scikit-hep/uproot>, 2020. Version 5.3.7, DOI: 10.5281/zenodo.1472436.
- [20] Jim Pivarski et al. awkward. <https://github.com/scikit-hep/awkward>, 2020. Version 2.6.5, DOI: 10.5281/zenodo.3952674.