Data Analysis of $(e, e'p)$ Argon and Titanium Electron Scattering

Austin Batz
The College of William and Mary, Williamsburg, Virginia, 23187, USA

July 24, 2019

Abstract

This analysis addressed comparing the data of $(e, e'p)$ observation in Jefferson Lab Hall A experiment E12-14-012 to Monte Carlo (MC) models. The results include comparisons of the data and the MC distributions of various kinematical variables where the background has been removed from the data. The distributions are scaled based on the charge of the incident electron beam, as well as the efficiency of the detector. The background subtraction improved the data-MC agreement, and the absolute scaling revealed that factors not yet accounted for such as final state interaction (FSI) depend in part on the shell of the nucleus struck by the incident electron. Completion of the analysis of the efficiency and full systematics will provide a viable nuclear model of neutrino scattering on argon, which will benefit future long-baseline neutrino experiments such as DUNE.

1 Introduction

1.1 The $(e, e'p)$ Reaction

Current models of lepton-nucleus interactions involving complex nuclei are limited by systematic uncertainties. To address this, experiment E12-14-012 in Hall A of Jefferson Lab (JLab) was performed using electron scattering off of various nuclei. The argon and titanium nuclei were of particular interest because of future long-baseline neutrino experiments investigating neutrino and anti-neutrino scattering off of the argon nucleus such as the Deep Underground Neutrino Experiment (DUNE). These experiments will measure charge-parity symmetry violation, which will provide insight into matter-antimatter asymmetry [7]. The data we analyzed concerned the $(e, e'p)$ reaction, wherein a single electron knocks off a single proton from the target nucleus as shown in Fig. 1, and the electron and proton are detected in coincidence. The argon nucleus has Z=18, A=40. The inequality of protons and neutrons makes argon an isospin-asymmetric nucleus, so lepton-proton interactions would yield different results than lepton-neutron interactions. Neutrino-neutron interactions in the argon nucleus are also of interest for future experiments, but electron scattering is not a viable method for measuring neutron knockout because the probability of the reaction is extremely small. The titanium nucleus, with Z=22, A=48, has the same number of protons as argon has neutrons, so $(e, e'p)$ in titanium serves as a useful proxy for neutron reactions in argon. The exclusive $(e, e'p)$ reaction is a significant part of the inclusive (e, e') reaction, which includes all electron scattering reaction mechanisms independent from the final state particles. The inclusive analysis has been completed, and the exclusive analysis, which is currently ongoing, will yield an even greater understanding of lepton-nucleus interaction mechanisms [6].

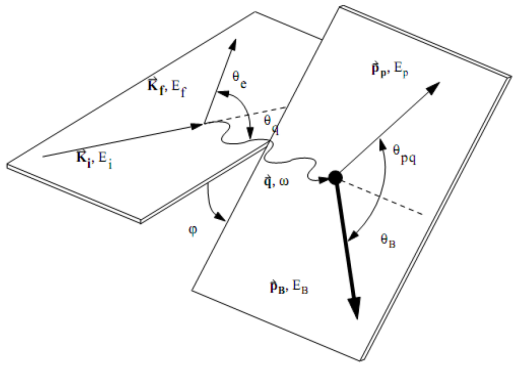

Figure 1: This is a diagram of the $(e, e'p)$ reaction, with θ as the in-plane angle and ϕ as the out-of-plane angle.

1.2 Jefferson Lab Hall A

The experiment took place in February and March of 2017 in Hall A of Jefferson Lab in Newport News, Virginia. The incident electrons came from JLab's continuous electron beam accelerator facility (CEBAF) in the form of a 2.222 GeV beam. The high resolution 4 GeV/c spectrometers in Hall A were specifically designed for experiments probing nuclear structure, with $(e, e'p)$ being a commonly used mechanism [3]. The beam struck either a solid titanium foil target or gaseous argon closed cell. For the argon target, pictured in Fig. 2, the beam passed through an aluminum entry and exit windows. An aluminum dummy target was used to characterize accidental events due to aluminum scattering that contaminated the argon scattering data. Since argon is a gaseous target, it was subject to a boiling effect, wherein the density near the beam path was lower than the density near the cell walls because the beam deposited heat in the gas and caused it to expand locally. At a beam current of 20 μA, the boiling effect is 27.5% on average. This is important to account for because the density of the target is a major factor in the probability of a scattering event [9].

After the reaction took place in the target, the final state electron and proton were separated by magnets and entered different spectrometers, with the electron and proton detected respectively by the left and the right high resolution spectrometers (HRS). Since not every interaction was $(e, e'p)$, not every particle in the detectors was an electron or proton, so different parts of the detectors aided in tracking and identifying which events were good and which were not. Good events created only one track through the vertical drift chambers (VDC), produced a minimum amount of Čerenkov light detected by the threshold Čerenkov counters, passed through both layers of the scintillators, and deposited energy within a specific range in the calorimeters. Using the final state electron and proton energy and momentum, one can use conservation laws to reconstruct the energy and momentum of the final state nucleus, known as the missing energy and missing momentum. This was not possible in the inclusive analysis, since only the final state electron was measured [1, 2]. The detectors had different settings to detect events with different energies, momenta, and scattering angles. These settings were called kinematics, and there were five distinct kinematic ranges to analyze. Tab. 1 contains the values the detector was set to for several kinematical variables. For the most part, we investigated kinematic #4, which corresponds to a relatively high missing momentum range.

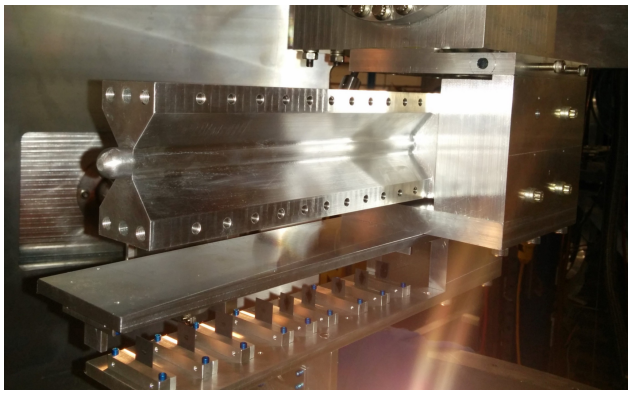

Figure 2: Argon target in sealed cell [9].

2 Methods

We compared the kinematic #4 on both the argon and titanium targets to the preliminary models provided by the MC. We produced plots of missing energy, missing momentum, scattering angles (in plane and out of plane), and other variables with the data and MC distributions superimposed. The kinematic #4 data included a higher rate of background events (such as pions in the detector) than lower kinematics due to the

	E_e MeV	$E_{e'}$ MeV	θ_e deg	P_p MeV/c	θ_p deg	$\|\mathbf{q}\|$ MeV/c	P_m MeV/c
kin1	2222	1777	21.5	915	−50.0	857.5	57.7
kin2	2222	1716	20.0	1030	−44.0	846.1	183.9
kin3	2222	1799	17.5	915	−47.0	740.9	174.1
kin4	2222	1799	15.5	915	−44.5	658.5	229.7
kin5	2222	1716	15.5	1030	−39.0	730.3	299.7

Table 1: This is a table of the kinematic settings. E_e is the energy of the incident electron, $E_{e'}$ is the energy of the outgoing electron, θ_e is the in-plane angle of the electron, P_p is the measured momentum of the proton, θ_p is the in-plane angle of the proton, $|\mathbf{q}|$ is the magnitude of momentum transferred to the proton, and P_m is the missing momentum.

high missing momentum, so we characterized the number of background events and their kinematical distributions and subtracted them from the data. We also scaled the distributions based on the charge of the beam and the efficiency of the detector (i.e. the number of events recorded per event that occurred). The analysis is ongoing, but these developments will be useful for determining the systematic uncertainties in the model.

2.1 Preparing the Data

We did all of our analysis work via remote login to the JLab servers. The data was too complicated in its raw form to analyze directly, so it had to go through multiple steps of preparation. The events were grouped in many runs, each corresponding to about an hour of the beam on the target and data being collected. Each run contained between a few hundred thousand events and several million events. For the plotting, we used the ROOT [8] data analysis framework built on C++, so we had to replay each run to extract the data from the raw files and create ROOT files. Even these files were too large and complicated to use directly, so we had another macro to extract only the relevant information from the large files and make new ROOT files. Every so often, the macro would encounter a bad event that caused it to crash, so many runs had several of these reduced ROOT files which do not include those bad events. After generating the reduced files and the simulation data, we could run the comparison macro.

2.2 Kinematical Variables and Cuts

Our ROOT comparison macro takes in the information from events in the selected ROOT files and simulation data and outputs histogram plots displaying the distributions of several variables with the data and MC overlaid. The variables are Ldp/p, Rdp/p, $L\theta$, $R\theta$, $L\phi$, $R\phi$, Z, β, E_m, and P_m. For dp/p, θ, and ϕ, the L or R corresponds to the left- or right-hand spectrometer, respectively. For each kinematic, the detectors were set to detect a particular momentum, and dp is the difference between the this setting and the actual detected momentum. The dp/p is the ratio of this difference to the setting. The angles θ and ϕ are used to describe the direction of the outgoing particle with respect to the incident beam, with θ being the in-plane and ϕ being the out-of-plane angle. The position Z (in meters) is where the reaction took place along the length of the target. β is speed of the outgoing proton as a fraction of the speed of light. E_m is the reconstructed missing energy (in GeV/c^2), and P_m is the reconstructed missing momentum (in GeV/c).

There are many different cuts on the data that filter out bad events. One type of cut is the acceptance cut. These are ranges over which dp/p, θ, ϕ, Z, and β are plotted where the MC is expected to model the data well. Other cuts are meant to filter out background (also known as accidental) events such as pions. The left arm of the detector needs to have observed a minimum amount of Čerenkov light, both arms can only have ob-

served one track through the VDC, a minimum amount of energy needs to have been deposited in the calorimeters, the beam current needed to be in a particular range, the event had to occur within a particular range of time, and the difference in the event being observed in the left and right arms had to be within a particular range. The ranges for these cuts were determined by observing the data-MC agreement and the distributions of the relevant variables with and without them. For example, the upper plot in Fig. 3 shows the distribution of energy deposited on the two layers of the calorimeter with the only cut being one track through the VDC. The lower plot is the same information with cuts one track, minimum Čerenkov, and minimum β. These cuts remove the accidental events in the small peak in the lower-left corner of the upper plot and the dark region near the actual peak of interest. Applying a cut results in a loss of efficiency and an increase in statistical uncertainty. The acceptance cuts were determined using the MC and finding where the efficiency distributions were flat [4]. The tradeoff is between using the best data to determine the systematic uncertainties and having enough data to make that determination.

2.3 Background Subtraction

The events that survive all the cuts fill histograms to show the distributions of the variables of interest. However, not all of these events are $(e, e'p)$. We could estimate the number of background events and their kinematical distributions using the time_diff variable, which represents the difference in time between the start of the event in the left and right spectrometer.

The peak in the time_diff plot in Fig. 4 is the coincidence peak. The peak is fit to a Gaussian distribution, from which we could extract its mean and σ. The events inside the coincidence peak include both good events and accidentals, but the events on either side of the peak (the anti-coincidence region) are all accidentals. The background distribution in the anti-coincidence region is almost exactly flat, so we assume the background distribution inside the peak is also flat. We made an anti-coincidence cut that accepts only events in the anti-coincidence region and applied all the other cuts the data is subject to. To reduce statistical uncertainty, we made the anti-coincidence region as wide as possible. We scaled the number of events by the ratio of the width of the peak ($2\ \sigma$) to the width of the anti-coincidence region. This scaled number of events is the the number of background events expected to be inside the peak.

Besides the number of background events, we could also find the distributions of background events for other variables, such as dp/p, θ, ϕ, and E_m. Subtracting the background distributions from the data resulted in a better comparison with the MC expectations due to the fact that the MC does not model any backgrounds. This has been especially important for analyzing higher P_m kinematics, such as kinematic #4, which will have significant contribution from background.

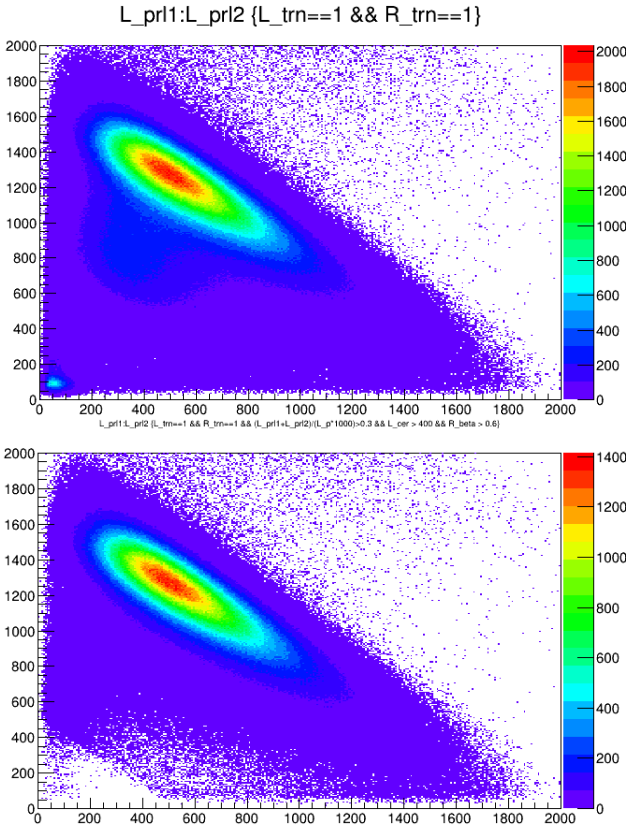

Figure 3: These plots are the distribution of energy deposited on the two-layered calorimeter with the energy in the first layer on the y-axis and the second layer on the x-axis. The lower plot has several cuts to remove the background events in the small peak seen in the lower-left corner of the upper plot.

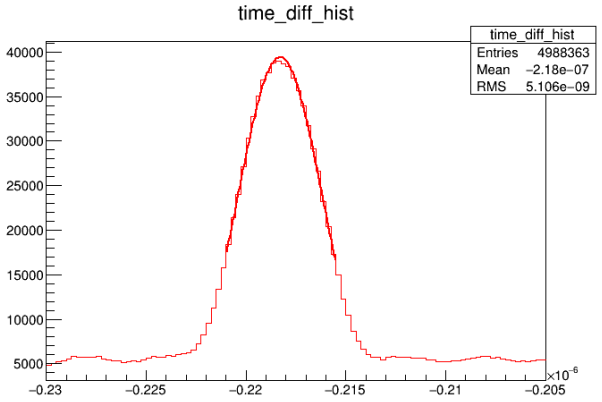

Figure 4: This is a plot of time_diff fit to a Gaussian. The events inside the peak are signal plus background, and the events on either side of the peak are all background. The background events outside of the peak can be used to estimate the behavior of the background events inside the peak.

2.4 Normalization

In order to compare the data and MC, both sets of distributions must be normalized to the total charge (i.e. scaled) to account for the assumptions made in creating the MC. At first, we normalized the distributions to have equal total number of events. This was useful for comparing the relative shape of the data and MC histograms, but it takes away important information, such as the effects of efficiency and the final state interaction. Efficiency plays a role because the MC was not subject to the real-world limitations of a detector or cuts that remove background. Different contributions to efficiency loss have complex interdependencies, and the calculations are ongoing, so the plots we have made factor in only preliminary efficiency data. The boiling effect was another scale factor for argon.

The other normalization factors to consider are total charge, and the final state interaction (FSI). The charge of events in the MC is set to 1 milli-Coulomb, but the charge of events in the data is on the order of micro-Coulombs and varies by run. Charge affects normalization, so the data is scaled down by the total charge in terms of milli-Coulombs, and the MC is scaled down by its number of events. FSI is when the final state particles interact with other nucleons in the final state nucleus after the $(e, e'p)$ reaction. This effect distorts the data distributions, and the MC does not account for it. For kinematic #1, we expected a normalization factor due to FSI of about 70-80% and a shift in the E_m and P_m distributions of 3-5 MeV, which is what we currently find. Kinematic #1 events predominantly correspond to events in the $1s$ and $1p$ nuclear shells, which are deep inside the nucleus. Kinematic #4 events are mostly in higher shells – the $1d$ and $2s$. This means FSI should have a smaller effect in kinematic #4 than kinematic #1 since the final state proton is less likely to interact with the inner final state nucleons [5].

3 Preliminary Results

3.1 Argon

We used our comparison macro to produce plots with signal-plus-background with the background distributions overlaid, as well as background-subtracted plots including subplots of the data-to-MC ratio distributions. The ratio subplots exclude bins where the corresponding bins of data or MC is less than 10% of the height of the tallest bin in that plot. This is meant to avoid uninformative spikes in the ratio distribution where it divides by a quantity close to zero. Most of the plots referenced in this paper can be found in Appendix A.

Fig. 10 contains the Ldp/p and Rdp/p distributions for argon. The data (in red) is above the MC, but the effect of FSI is expected to make the normalization of the data plots lower than the MC. However, the background distributions are fairly bin-independent, and the height of the background is greater than the difference between the data and MC. The same is true for $L\theta$ and $R\theta$ in Fig. 11. The $L\phi$ and $R\phi$ distributions in Fig. 12 show why it was important to use the background in the anti-coincidence region to get distributions of the background because they are bin-dependent. In fact, the regions where the data and MC disagree the most are also the regions with the most background.

In Fig. 13, we see that the Z background distribution is fairly bin-independent, but there is a slight offset between the data and MC. The β

data peak is much wider than the MC. The Z and β comparisons are limited by the fact that the data represents real-world events and detectors, and the MC is designed to make mathematically perfect peaks. The disagreement in these variables should be expected given the disagreement in other variables such as dp/p and ϕ.

In some of these plots, the shape of the MC is depends greatly on the theory prediction. We have seen this in dp/p and ϕ, and now also in E_m in Fig. 5. The model has four sharp peaks, with the first two overlapping. The data has a single sharp peak, then a gentler slope downward. The background is mostly bin-indenpendent, so it cannot account for the disagreement. FSI likely plays a role, but we cannot quantify what that would be. The P_m plot appears to have an offset, but the background can account for it. The height of the background plots in kinematic #4 is noteworthy. Nearly 3.8% of the events were accidental. For reference, we looked at kinematic #1 data and found only 0.019% background. The kinematic #1 $L\phi$ distribution before background subtraction is in Fig. 6, and the background distribution is almost completely invisible. This is why subtracting the background was especially important in kinematic #4. As seen in Fig. 7, the background for kinematic #4 titanium was lower than that of argon, but still significant (0.17%).

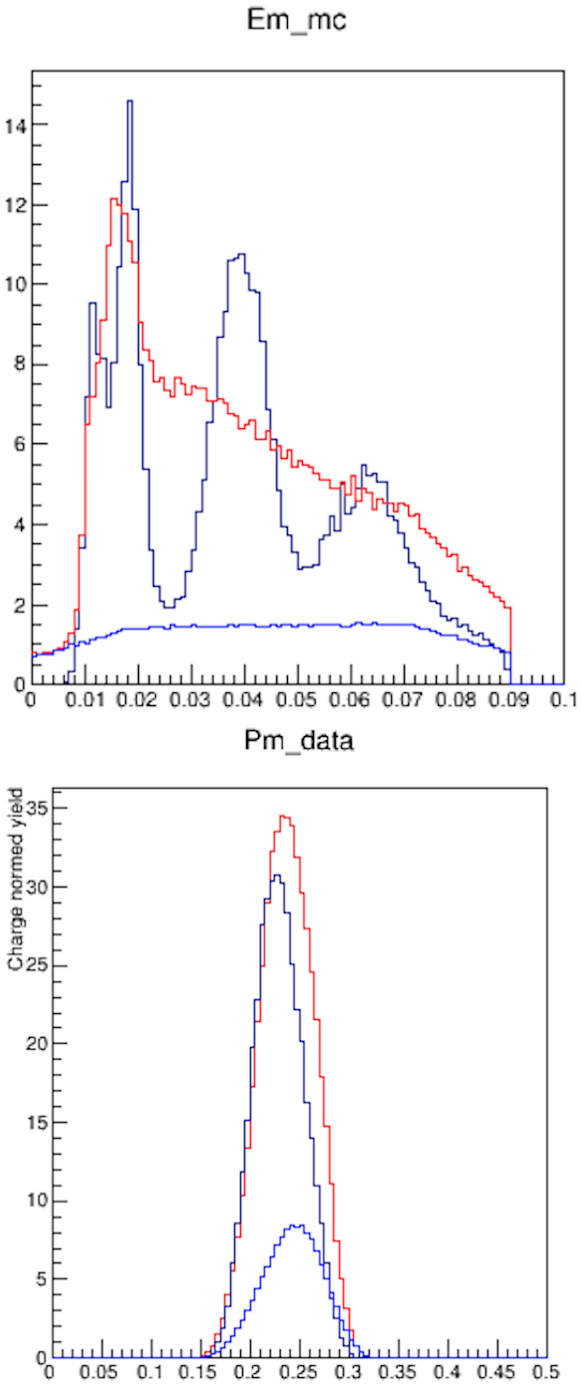

Figure 5: E_m (in GeV/c^2) and P_m (in GeV/c), respectively, for argon. The data without background subtraction is in red, the background distribution is blue, and the MC is black. The E_m plot was cut off at 0.09.

Figs. 8 and 14-17 show the argon plots after background subtraction with the corresponding data-to-MC ratio plots in the insets. Figs. 9 and 18-21 shows the same for the titanium target. The ratio plots have a red dashed line corresponding to a ratio of 1 for reference. The background subtraction greatly improves the data-MC agreement, and the normalization is closer to what we would expect considering the effect of FSI.

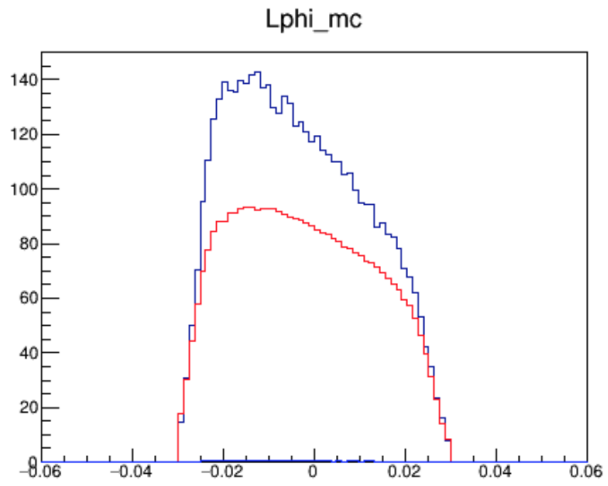

Figure 6: $L\phi$ for kinematic #1 argon. The data without background subtraction is in red, the background distribution is blue, and the MC is black. Note that the background plot is nearly invisible.

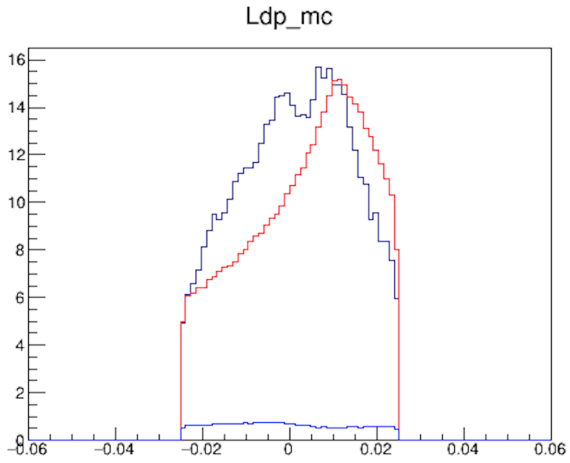

Figure 7: Ldp/p for titanium. The data without background subtraction is in red, the background distribution is blue, and the MC is black. Note that the background is smaller than kinematic #4 argon but larger than kinematic #1 argon.

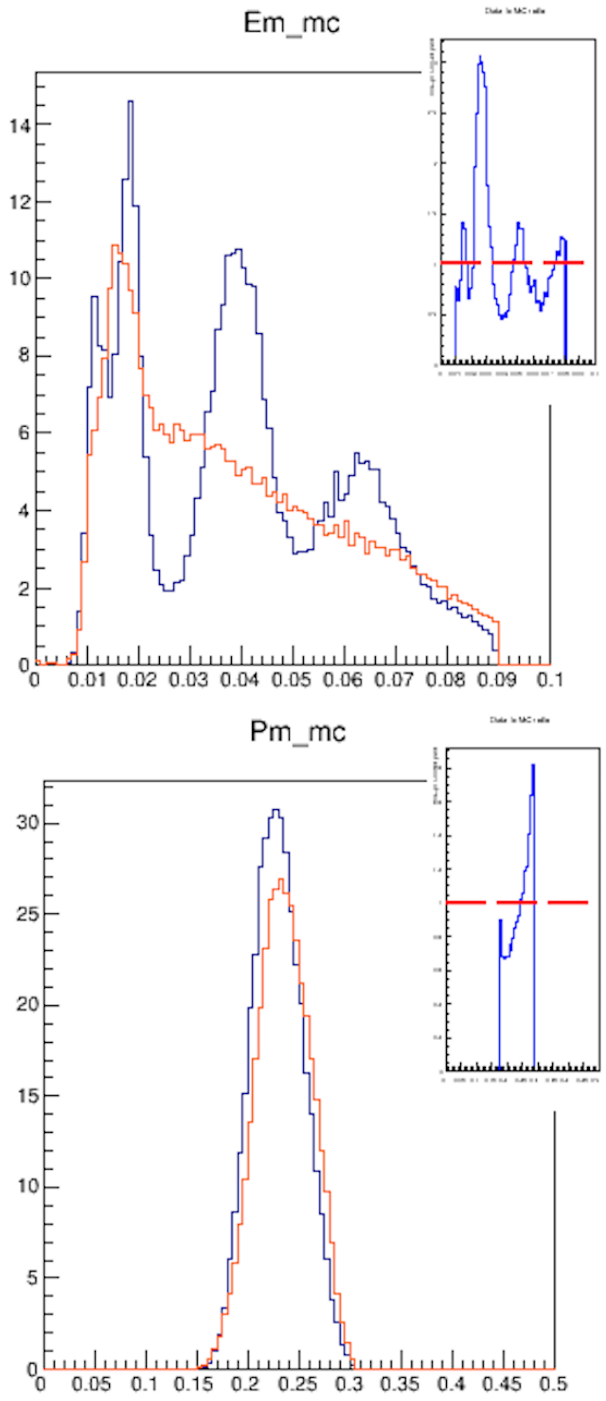

Figure 8: E_m (in GeV/c^2) and P_m (in GeV/c), respectively, for argon. The data after background subtraction is in orange, and the MC is black. The E_m plot was cut off at 0.09.

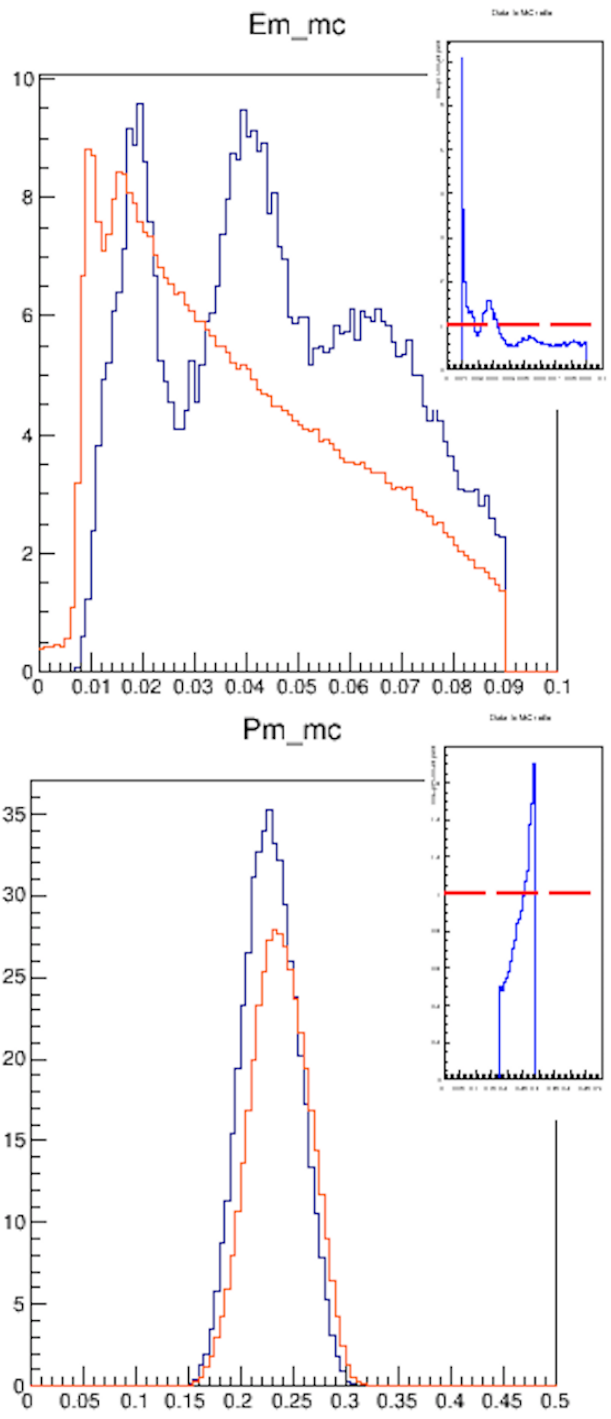

Figure 9: E_m (in GeV/c^2) and P_m (in GeV/c), respectively, for titanium. The data after background subtraction is in orange, and the MC is black. The E_m plot was cut off at 0.09

3.2 Titanium

The contribution of background to the titanium data may be smaller than that of argon, but the effect of FSI is similar. Both sets of distributions show the data below the MC, an offset of the E_m peak of about 5 MeV/c^2, and an offset of the P_m peak of about 2-3 MeV/c. Another point to consider is the dependence of FSI on the P_m range. Figs. 22 and 23 contain Ldp/p distributions cut on low and high ranges of reconstructed missing momentum for argon and titanium, respectively. For both targets, the data peak is lower and shifted to the right relative to the MC in the low P_m range. In the high P_m range, the data is higher than the MC, and the peaks align fairly well. This means that FSI has the greatest impact in low P_m regions.

3.3 Discussion

Even after background subtraction, none of the data plots exactly overlap the MC. However, this is reasonable because we have not accounted for the shape and scaling distortion due to FSI. The idea that FSI has the greatest effect in lower P_m ranges is reasonable, since low P_m corresponds to events that occurred deeper in the nucleus, so the final state particles had more opportunity to interact with other nucleons. These comparisons, while imperfect, are enough for the analysis to proceed in refining efficiency calculations and describing systematic uncertainties.

4 Conclusion

In this paper, we have reported our contribution to the data analysis of the exclusive $(e, e'p)$ nuclear interaction in argon and titanium experiment in Jefferson Lab Hall A. Our role was to implement background subtraction and normalization to charge and efficiency on distribution plots of various variables in order to compare them to the MC model. We found that the background subtraction improved the data-MC agreement, and the results of normalization are consistent with the effects of FSI. The next steps in the analysis are implementing revised efficiency calculations and describing systematic uncertainties of the nuclear model. These systematics will benefit

future long-baseline neutrino experiments such as the Deep Underground Neutrino Experiment.

References

[1] H. Dai *et al.*, (Jefferson Lab Hall A) arXiv:1803.01910v2 [nucl-ex] (2018).

[2] H. Dai *et al.*, (Jefferson Lab Hall A) arXiv:1810.10575v2 [nucl-ex] (2019).

[3] J. Alcorn *et al.*, Nucl. Instrum. Methods Phys. Res. A **522**, 294 (2004).

[4] J. Arrington *et al.*, Phys. Rev. Lett. **82**, 2056 (1999).

[5] J. Mougey *et al.*, Nuclear Phys. A, 262 (1976) p. 482.

[6] M. Murphy *et al.* (Jefferson Lab Hall A Collaboration) Al_Working_Draft_v3.pdf. (2019) (in progress).

[7] R. Acciarri *et al.* (DUNE Collaboration),719 arXiv:1512.06148.

[8] Rene Brun and Fons Rademakers, *ROOT - An Object Oriented Data Analysis Framework*, Proceedings AIHENP'96 Workshop, Lausanne, Sep. 1996, Nucl. Inst. & Meth. in Phys. Res. A 389 (1997) 81-86. See also http://root.cern.ch/

[9] S. N. Santiesteban *et al.* arXiv:1811.12167v2 [physics.ins-det] (2019).

A Other Plots

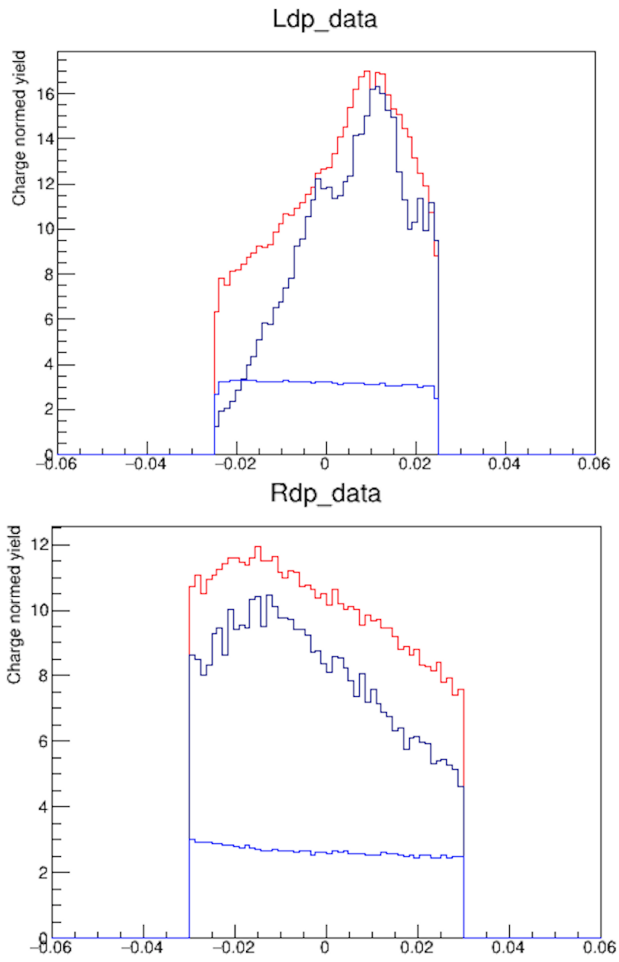

Figure 10: Ldp/p and Rdp/p, respectively, for argon. The data without background subtraction is in red, the background distribution is blue, and the MC is black. The acceptance cuts are $|Ldp/p| < 0.025$ and $|Rdp/p| < 0.03$.

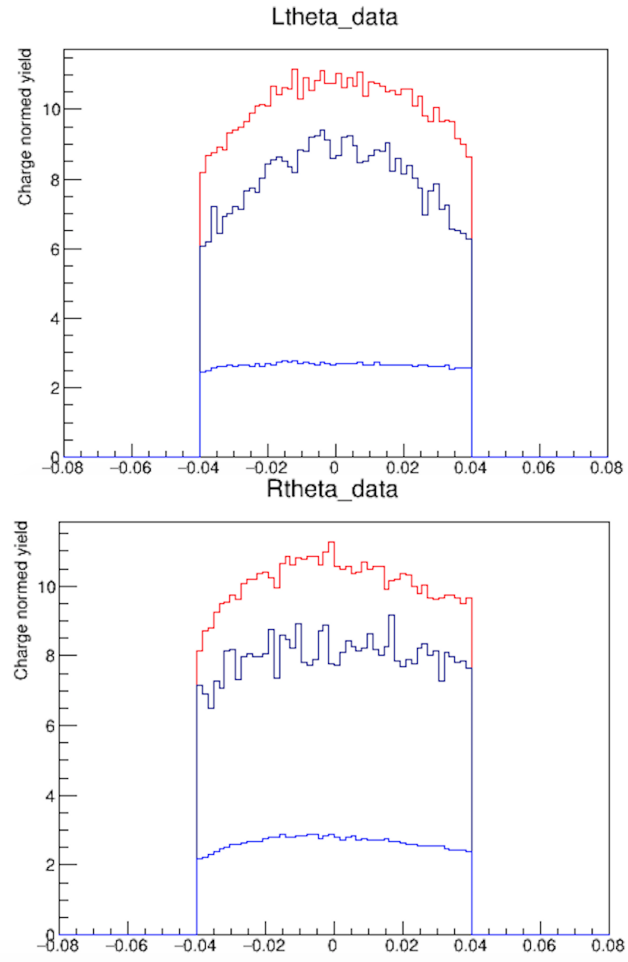

Figure 11: $L\theta$ and $R\theta$, respectively, for argon. The data without background subtraction is in red, the background distribution is blue, and the MC is black. The acceptance cuts for both the right and left arm are $|\theta| < 0.04$.

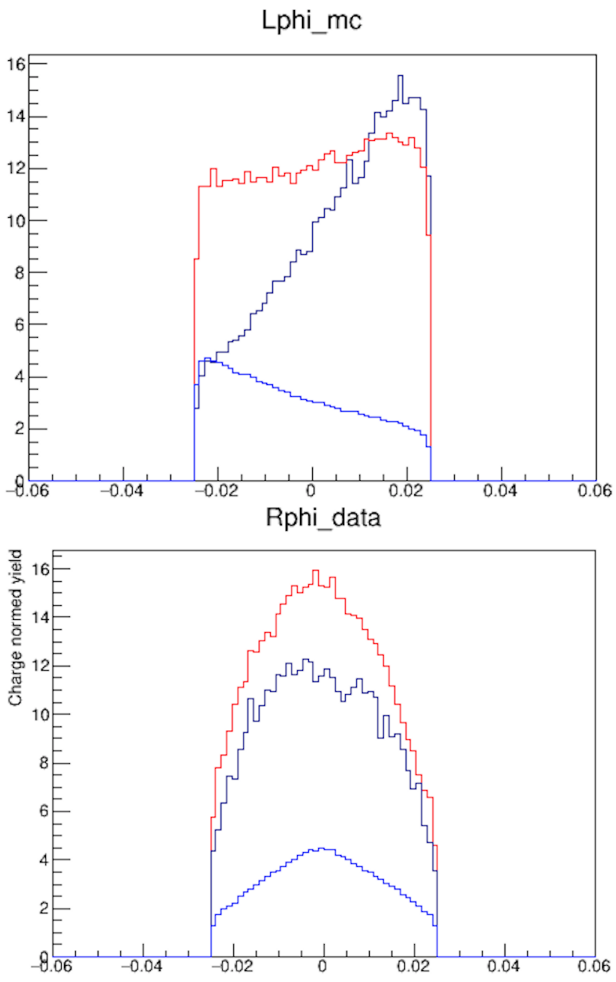

Figure 12: $L\phi$ and $R\phi$, respectively, for argon. The data without background subtraction is in red, the background distribution is blue, and the MC is black. The acceptance cuts for both the right and left arm are $|\phi| < 0.025$.

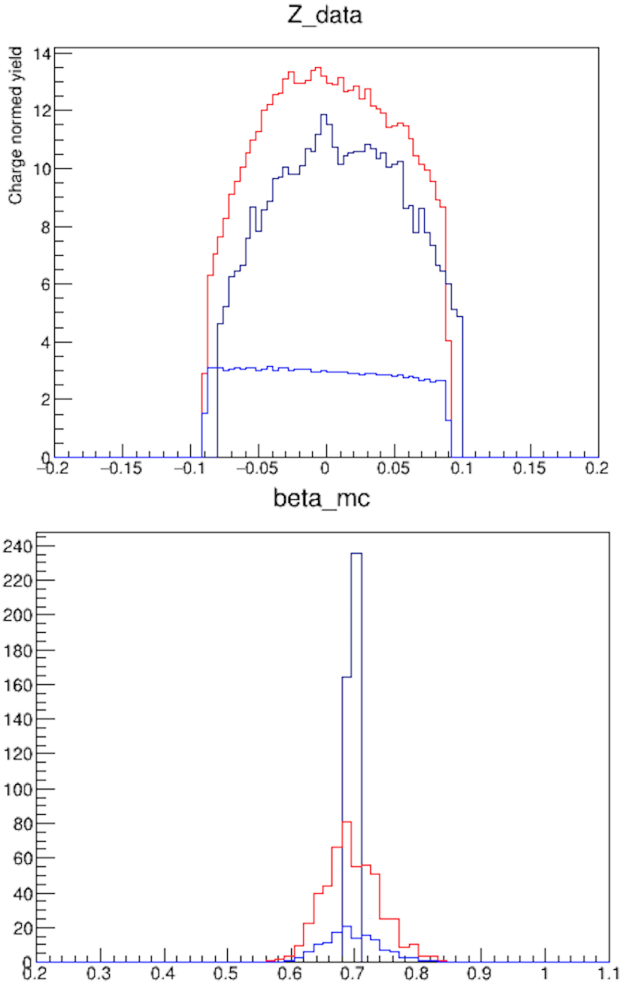

Figure 13: Z (in meters) and β, respectively, for argon. The data without background subtraction is in red, the background distribution is blue, and the MC is black. The acceptance cuts are $|Z| < 0.09$ and $0.55 < \beta < 0.85$.

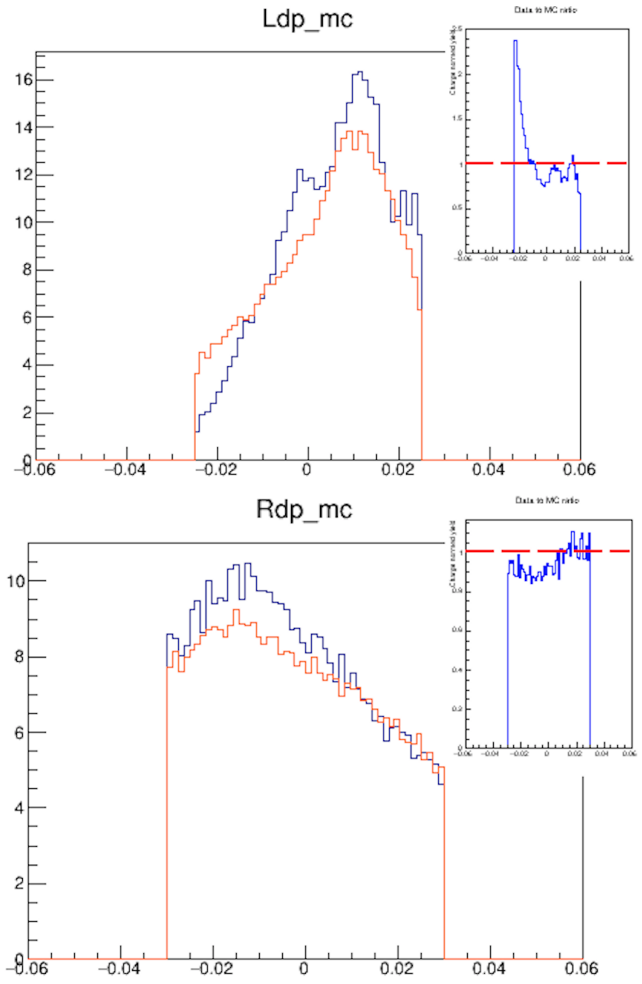

Figure 14: Ldp/p and Rdp/p, respectively, for argon. The data after background subtraction is in orange, and the MC is black. The acceptance cuts are $|Ldp/p| < 0.025$ and $|Rdp/p| < 0.03$.

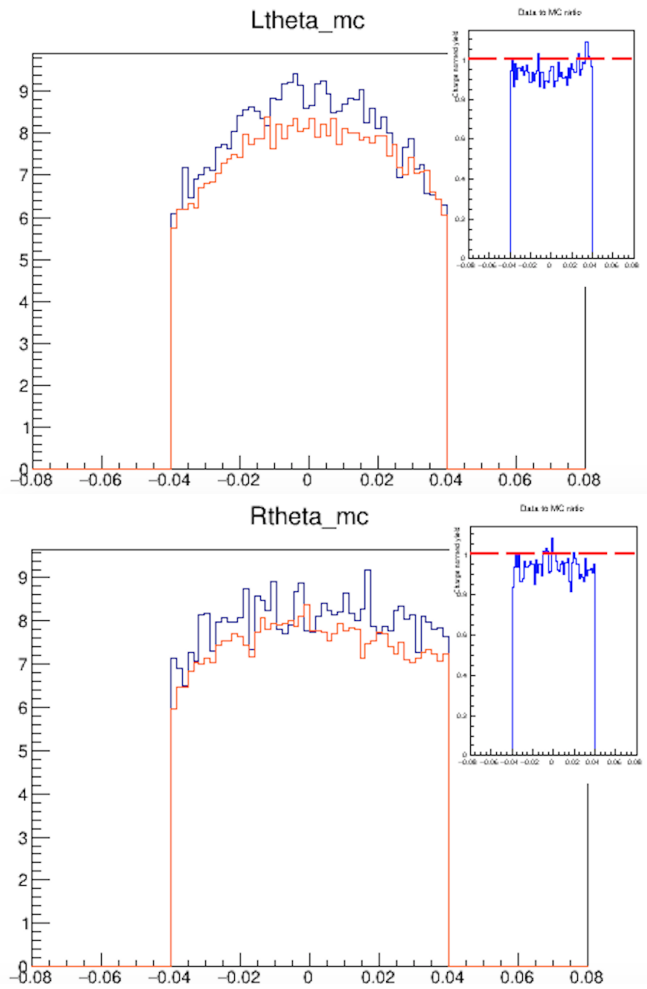

Figure 15: $L\theta$ and $R\theta$, respectively, for argon. The data after background subtraction is in orange, and the MC is black. The acceptance cuts for both the right and left arm are $|\theta| < 0.04$.

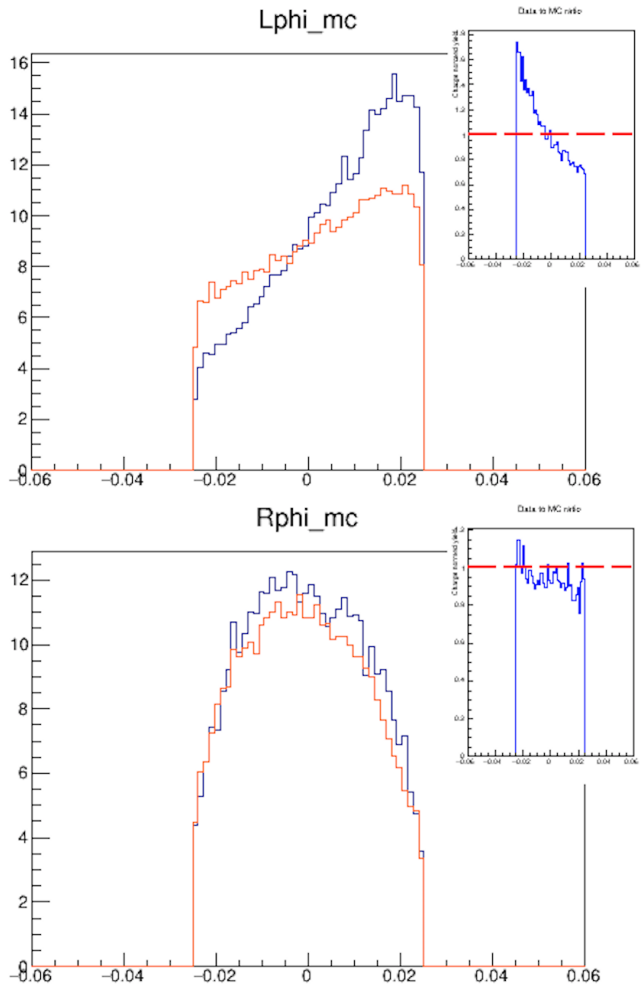

Figure 16: $L\phi$ and $R\phi$, respectively, for argon. The data after background subtraction is in orange, and the MC is black. The acceptance cuts for both the right and left arm are $|\phi| < 0.025$.

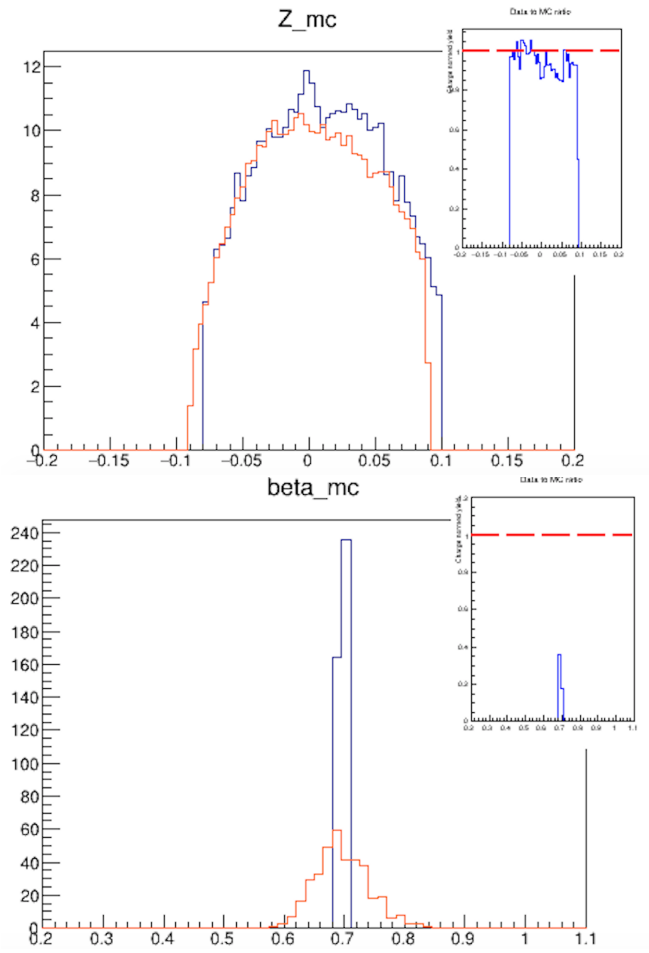

Figure 17: Z (in meters) and β, respectively, for argon. The data after background subtraction is in orange, and the MC is black. The acceptance cuts are $|Z| < 0.09$ and $0.55 < \beta < 0.85$.

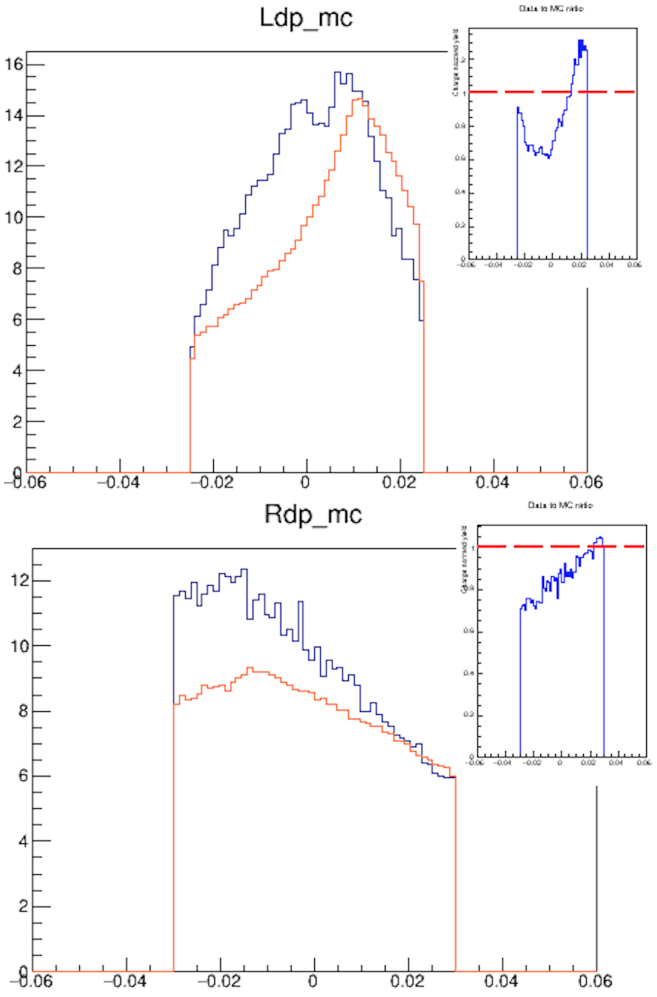

Figure 18: Ldp/p and Rdp/p, respectively, for titanium. The data after background subtraction is in orange, and the MC is black. The acceptance cuts are $|Ldp/p| < 0.025$ and $|Rdp/p| < 0.03$.

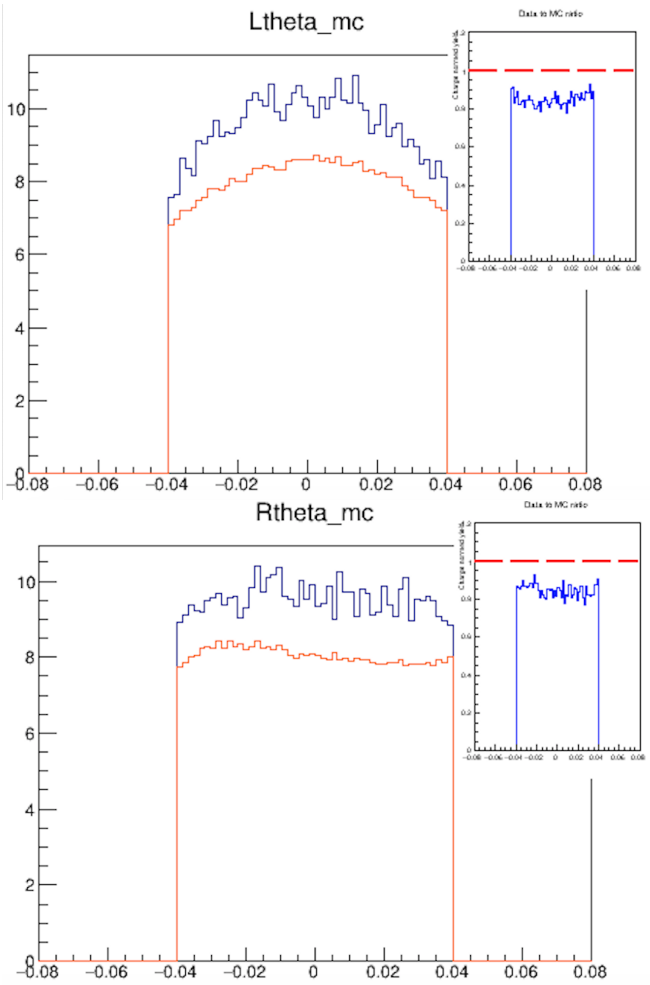

Figure 19: $L\theta$ and $R\theta$, respectively, for titanium. The data after background subtraction is in orange, and the MC is black. The acceptance cuts for both the right and left arm are $|\theta| < 0.04$.

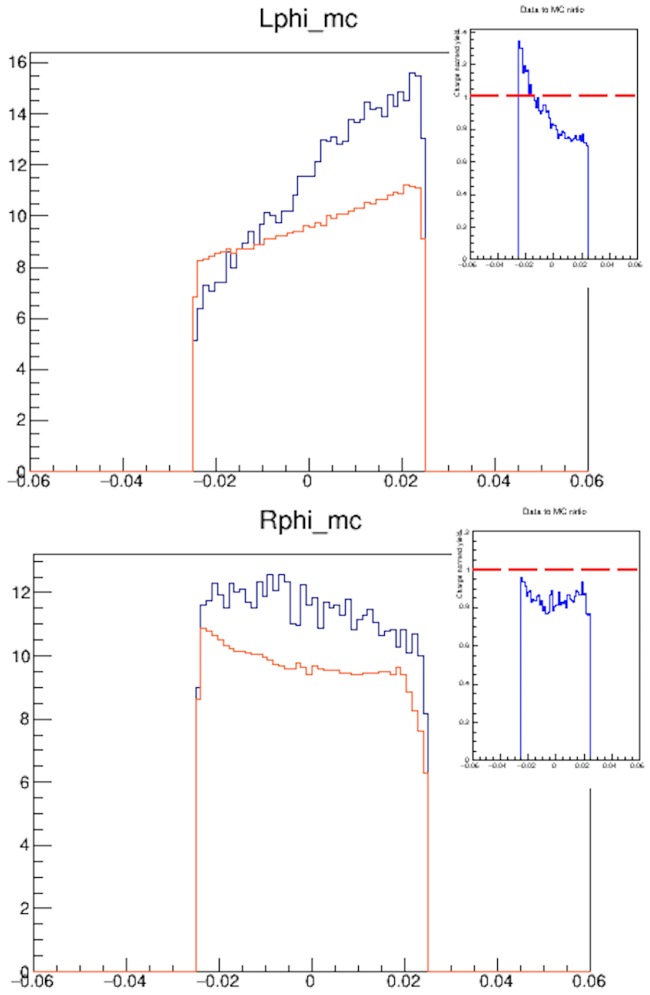

Figure 20: $L\phi$ and $R\phi$, respectively, for titanium. The data after background subtraction is in orange, and the MC is black. The acceptance cuts for both the right and left arm are $|\phi| < 0.025$.

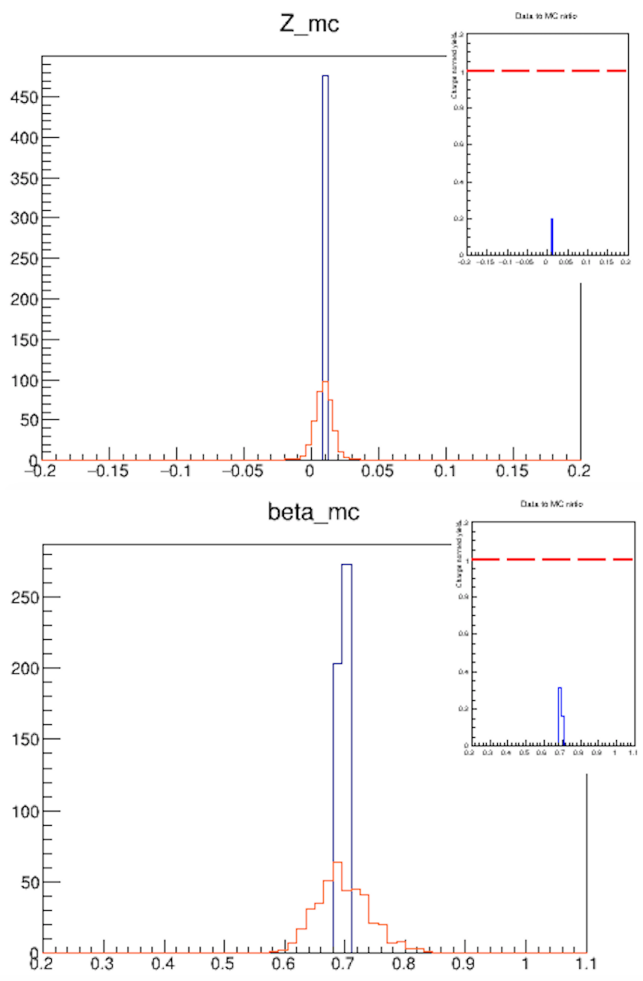

Figure 21: Z (in meters) and β, respectively, for titanium. The data after background subtraction is in orange, and the MC is black. The acceptance cuts are $|Z| < 0.09$ and $0.55 < \beta < 0.85$.

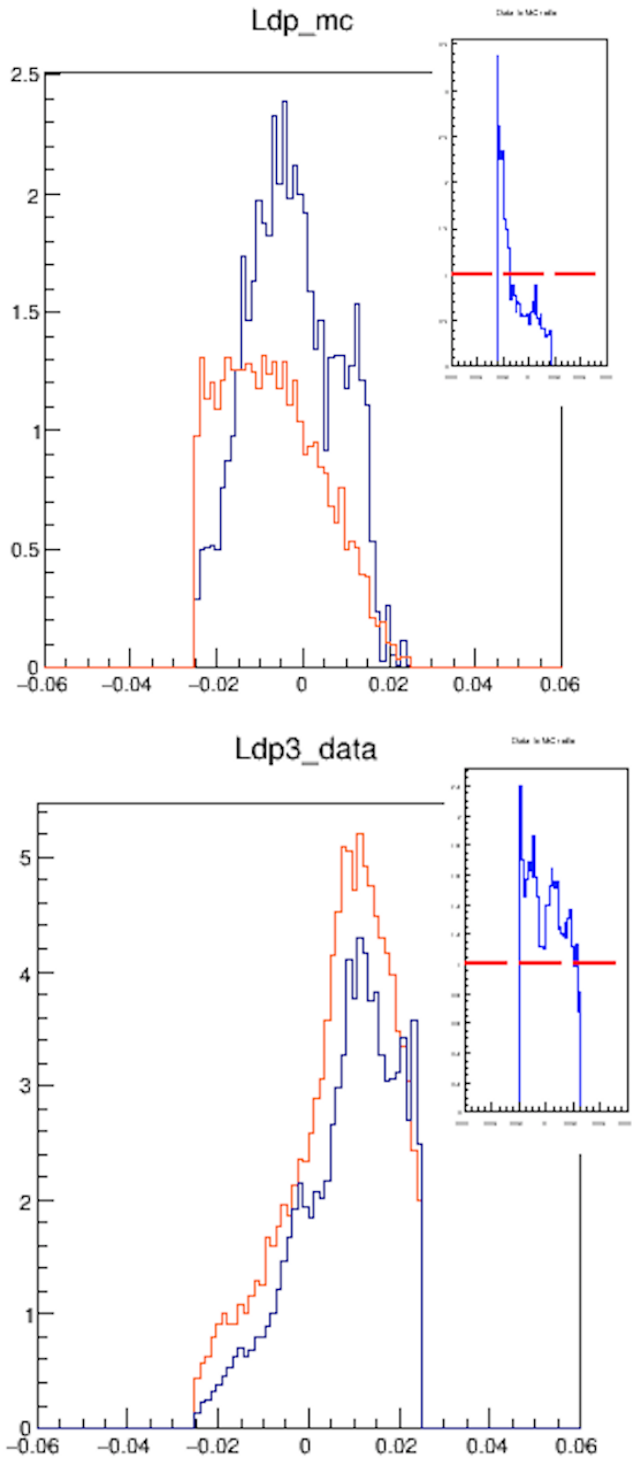

Figure 22: Ldp/p in low and high P_m ranges, respectively, for argon. The data after background subtraction is in orange, and the MC is black. The low range corresponds to $0.15 < P_m < 0.2$, and the high range corresponds to $0.25 < P_m < 0.3$ (in GeV/c).

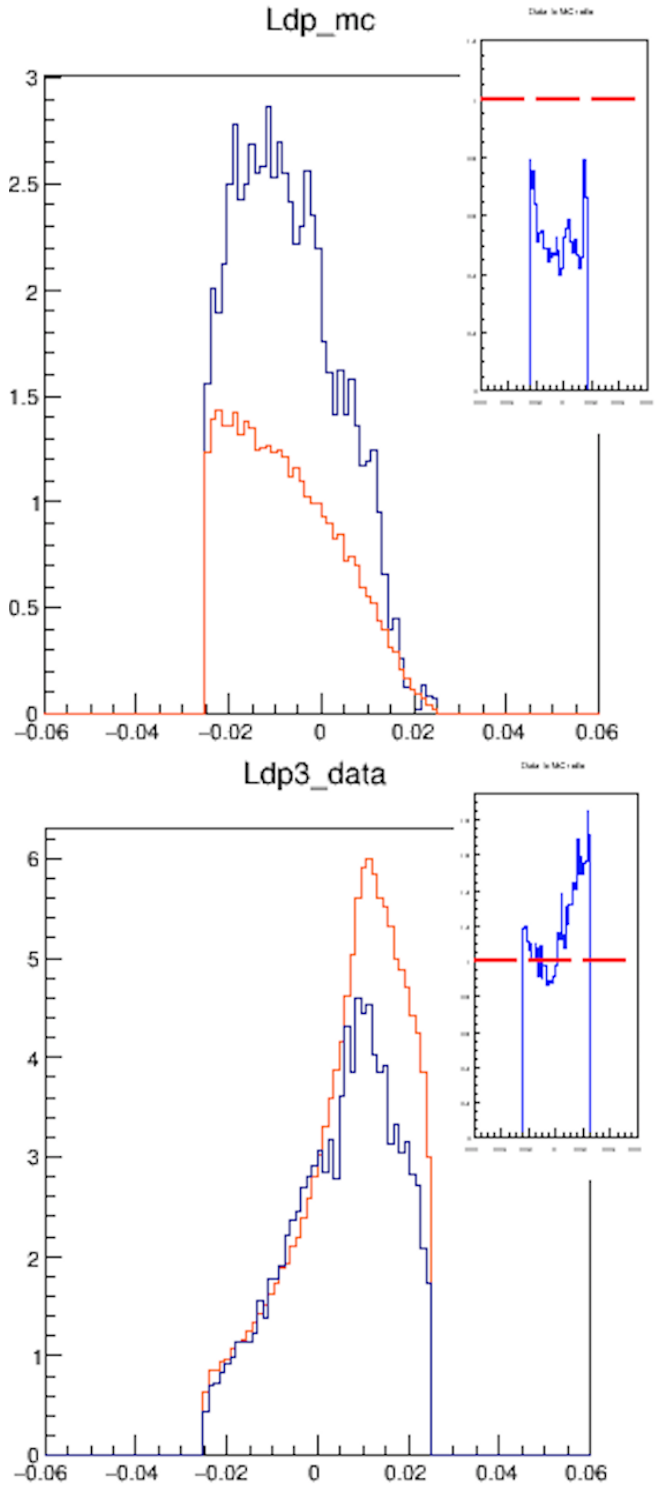

Figure 23: Ldp/p in low and high P_m ranges, respectively, for titanium. The data after background subtraction is in orange, and the MC is black. The low range corresponds to $0.15 < P_m < 0.2$, and the high range corresponds to $0.25 < P_m < 0.3$ (in GeV/c).