Proton Inelastic XS: MC Study

Ryan Bouabid

November 25, 2018

Abstract

This note will go over the final methodology for the MC proton-Argon inelastic cross-section measurement. This analysis is meant to take LArIAT proton events and identify inelastic scatter candidates. To make a final measurement on data we also need to evaluate the accuracy of such a selection as well as identify and correct sources of systematic uncertainty associated with the selection and the detector. The strategy to make this measurement and the appropriate corrections is described here. We also highlight the strategy for estimating statistical and systematic uncertainty in this MC analysis.

1 Introduction

Future LAr neutrino experiments will rely on an understanding of charged particle behavior inside the TPCs. Protons are a common final state particle to come out of a neutrino-Ar interaction vertex. Particle identification algorithms on protons in liquid argon rely on the full reconstruction of a Bragg peak at the end of the protons track. However, before coming to a stop the proton may interact inelastically with an argon nucleus and make it impossible to identify the particle by its signature rapid energy deposition. By measuring the cross-section of inelastic scatters on argon, we can better tune these particle identification algorithms to take into account this potential inefficiency.

This analysis focuses on LArIAT MC data. The main objectives of this analysis is to come up with a measurement method that is able to reproduce Geant4 predictions within MC, and to estimate the uncertainty on this method that will be applied on data.

2 XS Calculations

2.1 Thin Slab Method

In a traditional cross-section measurement experiment we consider a thin slab of thickness dz, a large area relative to an incoming beam of particles, and scattering centers along the surface of the slab with density n and cross-section σ. The probability that a particle from our beam scatters off of one of these centers is equal to

$$dp = \sigma n dz. \tag{1}$$

If we call the total number of particles that hit the front face of the target N, then the number of particles that go through without scattering is $N - dN$, where

$$dN = dpN. \tag{2}$$

And so the ratio of interacting to incident particles can be described as

$$\frac{dN}{N} = \sigma n dz. \tag{3}$$

This differential equation has the solution

$$N(z) = N_0 e^{-\sigma n dz}. \tag{4}$$

In other words, the probability of a particle in the incident beam surviving through the slab is equal to $e^{-\sigma n dz}$. For a slab of finite thickness z [cm]. The probability of interacting then, is

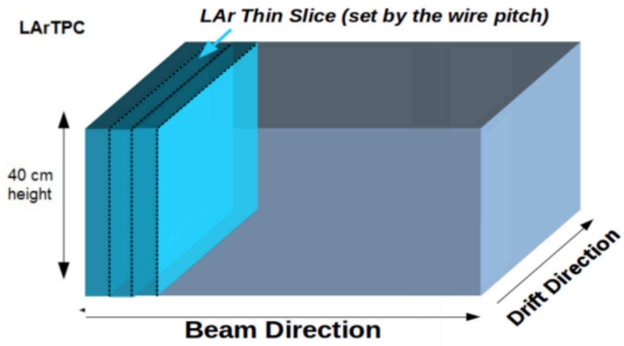

Figure 1: LArIAT detector visualized as a series of thin slabs of argon.

$$\frac{N_{interacting}}{N_{incident}} = 1 - e^{-\sigma n z}. \tag{5}$$

And so for sufficiently small z we can approximate the cross-section as

$$\sigma \approx \frac{1}{nz} \frac{N_{interacting}}{N_{incident}}. \tag{6}$$

2.2 Calculation in LArIAT

In LArIAT we use this method of calculating particle cross-sections by dividing the detector into multiple thin targets. As the incoming charged particle passes through each thin slab it loses energy due to ionizing argon atoms. As a result from one initial incoming particle, we can accumulate many separate incidences to test for scatters at ranges of energies, as seen in Figure 1. The technical challenges in this measurement rely on identifying the energy of the primary proton at the front face of each slab, and identifying whether or not the proton scattered inelastically on a particular slab.

3 Simulation Information

This analysis centers on a MC simulation of LArIAT data and cross-section measurements calculated within this simulation. The simulations used assumes an underlying theoretical proton-argon inelastic cross-section. The sample generated using a single particle that is not data driven. We have not included pile-up, beam kinematics, or other data related effects. It is a flat energy sample that covers a greater range than the data kinematics.

3.1 Geant4 Cross-Section and Models

The Geant4 simulations used in this study sample interaction probabilities from default cross-section tables. For this analysis this is the protonInelastic table. After an interaction the final state kinematics is dependent on the hadronic model loaded in the simulation. There are two hadronic models to compare at LArIAT energies: Bertini Cascade (default) and Binary Ion Cascade. Neither of these models effect the underlying cross-section at the truth level.

Thanks to work by Hans Wensel and Elena Gramellini we have an extraction of the proton-argon inelastic cross-section straight from Geant4 [1] as seen in Figure 2.

3.2 Measuring XS Using G4 Info

Within our LArIAT simulation we have access to Geant4 information for every particle in each event. This includes position and energy of the primary proton - the necessary components to calculate our cross-section.

[1] https://github.com/hanswenzel/G4HadStudies

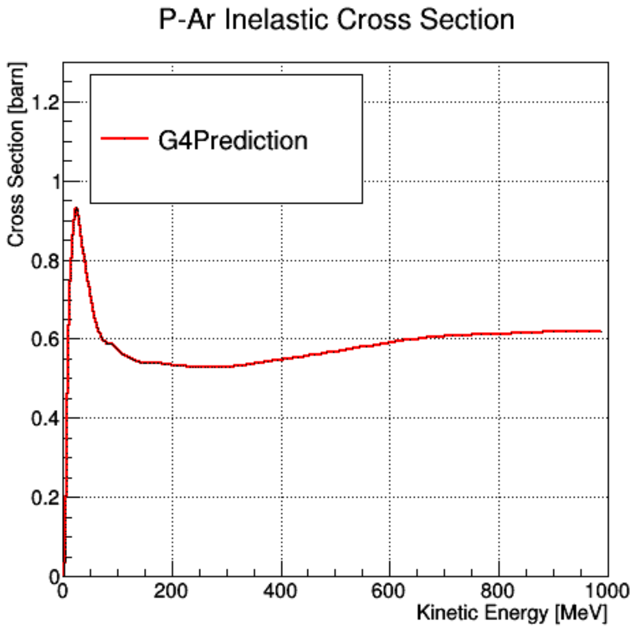

Figure 2: Bertini Cascade proton inelastic cross-section model.

3.2.1 General Method

We generate a sample of single proton events entering the LArIAT TPC from the front face. Geant4 keeps track of particle trajectories, which are made of "steps". Each step contains member objects of position, energy, and process type (either an ionization process or an interaction with an argon nucleus).

We iterate over steps of a primary proton and at each step we enter the energy of the proton into a histogram of incident energy. At each step we also ask if the step process is an inelastic interaction. If it is, then the energy at that step is entered into a histogram of interacting energy. The cross-section, in bins of kinetic energy is then calculated as follows:

$$\sigma(KE) = \frac{A}{\rho z N_A} \frac{N_{interacting}(KE)}{N_{incident}(KE)}. \quad (7)$$

Where

- $N_{interacting}(KE)$: Number of entries in given bin of kinetic energy for inelastically interacting protons
- $N_{incident}$: Number of entries in a given bin of kinetic energy for incident protons
- A: The molar mass of argon
- ρ: The density of argon
- z: The thickness of our target slabs. This is set by the Geant4 step size

3.2.2 Technical Details

One key feature of this measurement at the truth level (using Geant4 information alone), is access to near continuous information about the primary proton track. A default feature in LArIATSoft is to throw away the majority of Geant4 steps for which the process is ionization alone. But to make this truth level measurement effectively, it is necessary know the energy of the primary proton at a regular and small spacing (our slab size). Because of this constraint a small change in the LArSim package was made to retain the full list of steps in a proton trajectory.

The change was made in this file: larsim/LArG4/ParticleListAction.cxx

The change is shown in figure 3.

Figure 3: Piece of the ParticleListAction module, with the change in place to retain all spacepoints. Default line is commented.

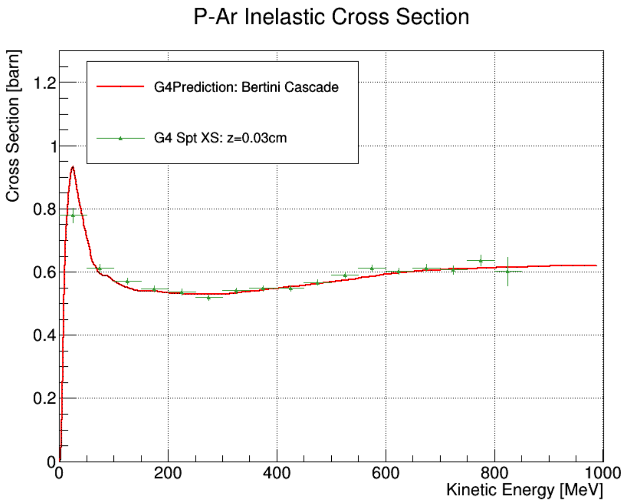

Figure 4: Comparison of Geant4 inelastic cross-section to cross-section calculated using thin slab method in LArIAT simulation using Geant4 steps 0.03cm apart.

3.2.3 Results and Slab Size

This calculation method reproduces the model that this simulation uses. A result of this study is that we observed a dependence on the value of the calculated cross-section and the slab size used in the measurement, which we expect as our formula for σ is only true to first order. For a step size of 0.03 cm, a comparison is shown in Figure 4.

Since such granularity is not available to us in reconstruction we also produce a cross-section at truth level using a larger step size that we will use in LArIAT data. This comparison with the addition of a cross-section calculated with a step size of 0.5 cm is shown in Figure 5.

4 Measuring XS Using Reco Info

4.1 General Method

We also produce a cross-section calculation using LArIAT reconstruction variables in MC. The procedure is generally the same, we step over pieces of a reconstructed tracks looking for interactions while populating interacting and incident distributions. The differences between this calculation and the Geant4 calculation described above are as follows:

- Instead of knowing whether or not a slab entry is an interaction or not, we select interaction and incident candidates and apply selection corrections based on truth information.
- The slab size at the reconstructed level has a minimum possible value tied to the spatial granularity of the detector (roughly the spacing of the wires on collection plane).

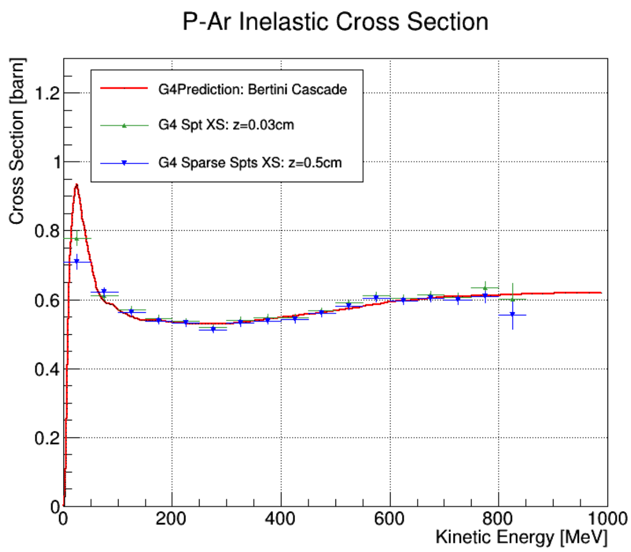

Figure 5: Comparison of Geant4 inelastic cross-section to cross-section calculated using thin slab method in LArIAT simulation. Both dense and sparse spacings are compared (0.03cm and 0.5cm).

- The energy that is reconstructed is convolved by detector effects and reconstruction algorithm systematics. We unfold these values to recover true kinetic energy.

4.2 Cross-Section Formula

The new cross-section formula that we use as a result for the reconstructed MC (and that we will use on data) is as follows:

$$\sigma = \frac{A}{\rho z N_A} \frac{\mathbf{U^{ij}}(N_{selected} - B_{selected})}{\mathbf{U^{ij}}(N_{incident} - B_{incident})} \frac{\epsilon_{incident}}{\epsilon_{selected}} \quad (8)$$

- $N_{selected}$: Number of inelastic scatter candidates at a given energy
- $N_{incident}$: Number of total events at a given energy
- $B_{selected}$: Number of background events within interacting candidates
- $B_{incident}$: Number of background events in the incident entries
- U^{ij}: Energy unfolding matrix
- $\epsilon_{selected}$: Efficiency of selection
- $\epsilon_{incident}$: Efficiency of the incident entries

4.3 Results and Agreement

Using appropriate corrections we are able to recover the original Geant4 calculation of the cross-section. The comparison between the model, truth level cross-section, and reconstructed cross-section is shown in Figure 6.

5 Cross-Section Cookbook

We've seen so far that this analysis method correctly reproduces the underlying Geant4 cross-section curve. This closure test is crucial to analyzing data properly. To reproduce this method in similar LArIAT analyses we'll go over each of the ingredients in the cross-section equation we described.

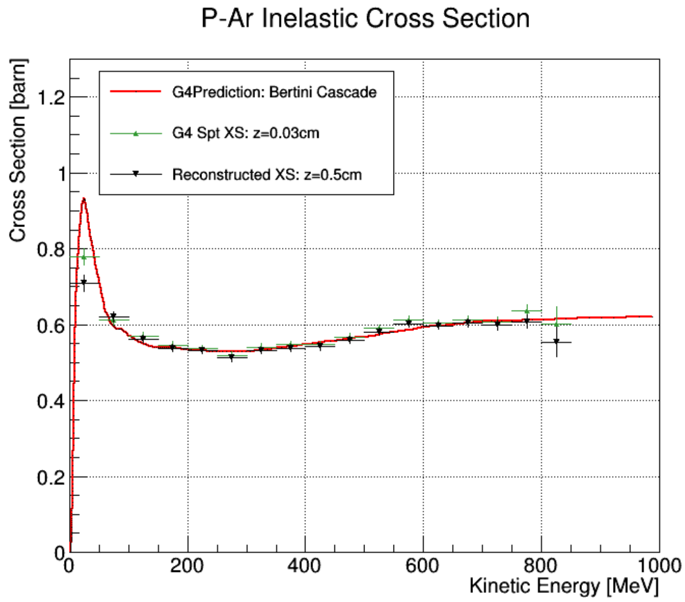

Figure 6: Comparison of model inelastic cross-section to cross-section calculated using thin slab method in LArIAT simulation using Geant4 steps as well as cross-section calculated using reconstructed variables.

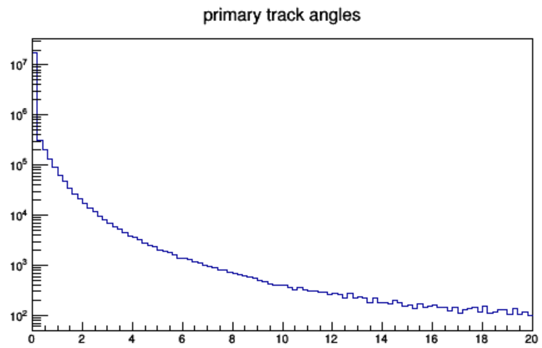

Figure 7: Reconstructed angular distribution of primary tracks.

5.1 N Selected

We use a set of reconstructed parameters to identify possible interacting primary protons. Depending on event topology we have specific cuts to look for inelastic scatters. If these cuts are flagged in an event, this event will contribute one entry to an "interacting histogram". The entry is at the energy of the proton at the front face of the slab of argon it interacted in.

5.1.1 Cuts

There are three main cuts that we look for as we step over the length of a primary proton entering the TPC.

- Whether or not there is a kink in the primary track greater than 6 degrees. The total distribution of reconstructed angles is seen in Figure 7.
- Whether nor not there are any secondary tracks emanating from the primary proton.
- Whether or not there is a characteristic proton Bragg peak at the end of the primary track if the previous two cuts are not flagged. The end mean dE/dx is shown in Figure 8.

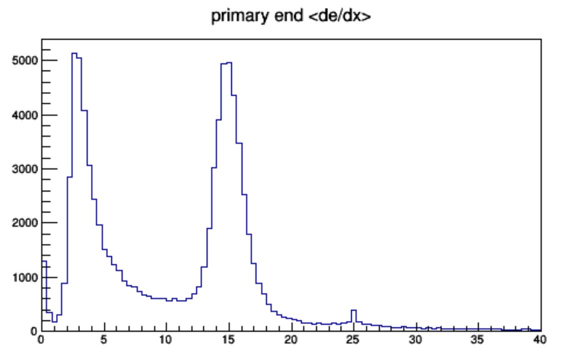

Figure 8: End mean dE/dx across last 2.5 cm.

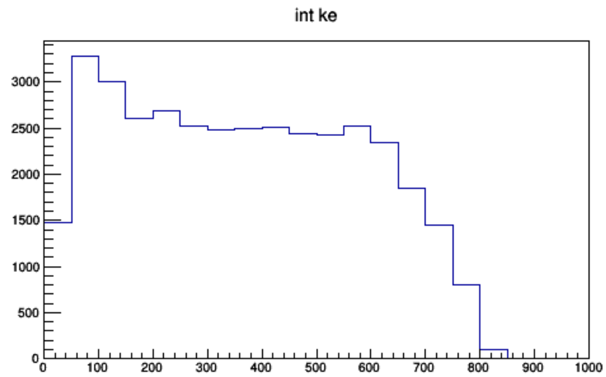

Figure 9: Distribution of kinetic energy at interaction point for inelastic scatter candidates.

5.1.2 Distribution of Selection

If one of these cuts is flagged we take the energy of the proton at the front face of the slab of argon nearest the interaction vertex to be the candidate kinetic energy. This total distribution is shown in Figure 9.

5.2 N incident and z

Just like at the truth level measurement, we keep track of the energy of the primary proton at the front face of each slab. Since we have divided the LArTPC into multiple sequential thin slab scatter experiments, the energy lost due to ionization in each slab of argon is subtracted from the initial kinetic energy to get this incident kinetic energy.

This distribution population is closely related to the slab size and requires some extra care. It's important to note that the total number of entries in this histogram multiplied by the slab size must give the amount of argon that was traversed by the primary proton. Messing up this bookkeeping will introduce bias in our measurement.

5.2.1 Calculation Method

Our strategy will be fix a permanent slab size for the entire measurement of 0.5cm. Depending on the pitch of the primary track the distance between calorimetric hits (projected onto wire planes of fixed spacing) will vary. But we will use these hits to make linear fits of energy loss across fractions of a centimeter. This allows us to calculate the energy of the proton at the front face of each slab 0.5 cm apart. A one dimensional toy example is shown in Figure 11.

5.2.2 Distribution

The kinetic energy distribution of incident entries is shown in Figure 12. Note that this distribution includes the interacting candidate distribution. This is because every slab that a proton goes

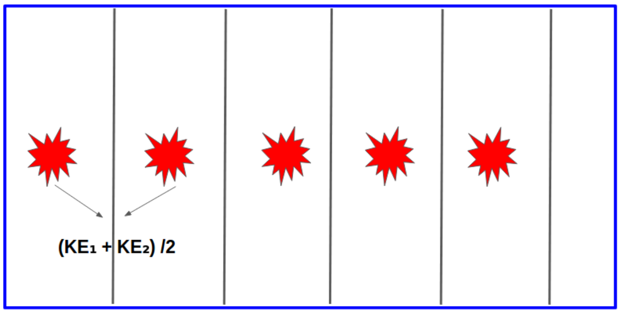

Figure 10: 1D cartoon of how we extrapolate variable hit position to a fixed thin slab spacing.

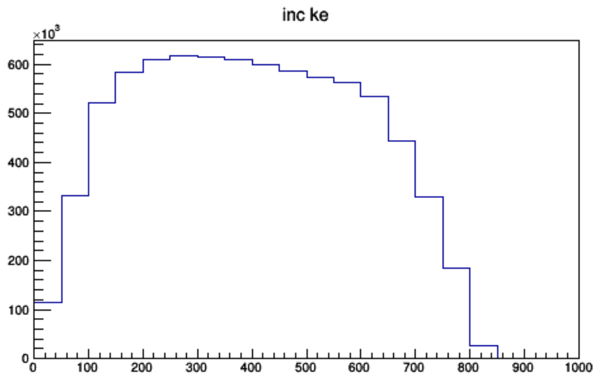

Figure 11: Distribution of kinetic energy of primary proton at the front face of every slab.

through shows up as an entry in this distribution.

5.3 Backgrounds

Both the interacting distribution and the incident distribution are susceptible to impurities in the selection procedure. As we'll see with efficiency corrections, it's important that we correct both the numerator and denominator of this measurement.

5.3.1 Interacting Background

Backgrounds to the interaction candidates are defined as follows:

- If we identify an Geant4 interaction correctly but it is not a proton inelastic collision we call it a background.
- If our interaction candidate vertex is greater than one slab distance (0.5cm) from a proton-argon interaction we call it a background.

The breakdown of the source of background is shown in Figure 12. Failure to identify the vertex within one slab of an interaction includes scenarios where there is a proton-argon inelastic collision inside the TPC and we miss reconstruct it, and scenarios where there is no interaction whatsoever. Typically the latter is due to reconstruction pathologies. The signal and background is compared in Figure 13.

5.3.2 Incident Background

The background distribution in the incident selection is less obvious, but closely related. It is easiest explained with an example.

In Figure 14 we have an event display of a LArIAT MC event from our sample. This event has an inelastic interaction that our selection procedure missed (an inefficiency). As a result, the total

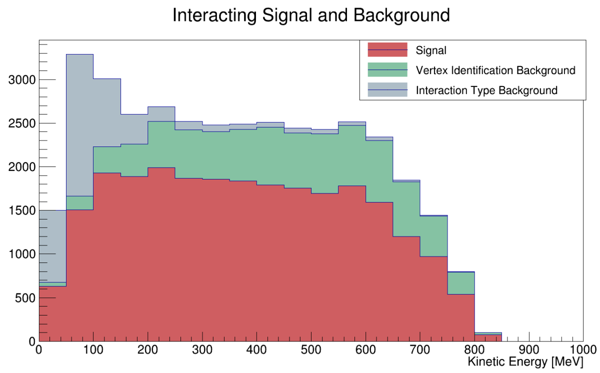

Figure 12: Figure background breakdown.

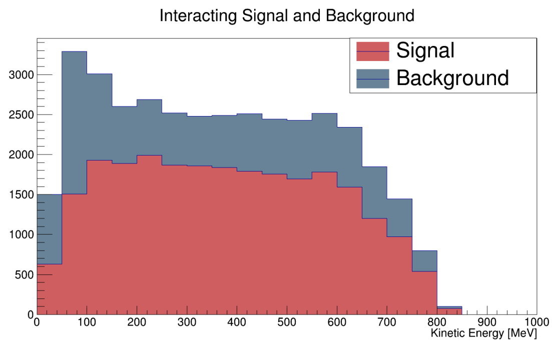

Figure 13: Figure showing background distribution

number of entries in the reconstructed incident distribution from this event includes entries passed the true interaction point. This is an over count because the proton scattered inelastically off an argon nucleus after going through a fixed amount of argon, and everything reconstructed down stream of this interaction point is not our primary proton. But by assuming this was all the same particle, we over counted the denominator for this particular event. In general, any inefficiency in the selection of inelastic scatters manifests itself as an impurity in the incident distribution that must be subtracted off.

5.4 Unsmearing Matrices

This measurement technique requires knowing the energy of the primary proton at the front face of every slab of argon. As the particle goes through the LArIAT detector it loses energy due to ionizing argon electrons. The LArIAT reconstruction of that energy loss is, in general, very accurate. However, as the proton loses energy, the rate of energy of energy loss increases rapidly leading to small regions of argon containing large clouds of ionization electrons. Misreconstructions of energy loss in these scenarios lead to a bias in the reconstructed distributions of interacting and incident energies. To transform our reconstructed kinematic distributions to true physical distributions we use a matrix inversion method.

5.4.1 Construction

The construction of this unfolding operator relies on effectively matching truth level information in the simulation with reconstructed variables. For every slab entry at the reconstructed level, we check if its position is lined up with the position of the proton at the truth level. If it is (meaning it is a signal entry in the denominator), then we add an entry into a 2d histogram with the value (reconstructed kinetic energy, true kinetic energy). This 2d distribution is shown in Figure 15, where we see very strong agreement for most of the kinematic range between the reconstruction and truth information. The low energy bins have significant off diagonal terms from improperly reconstructed proton Bragg peaks.

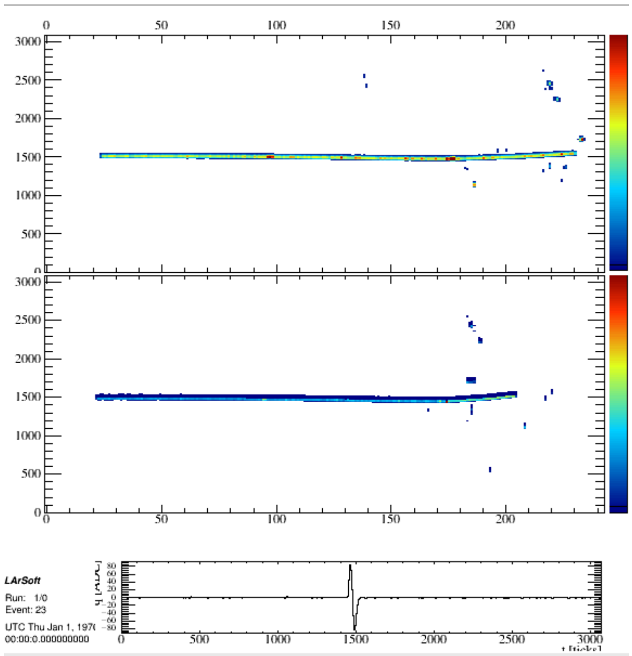

Figure 14: LArIAT MC event with a true inelastic proton-argon collision that was missed by our event selection.

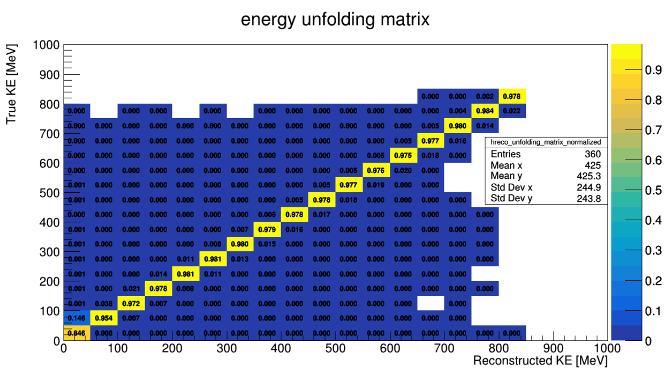

Figure 15: Matrix inversion transformation.

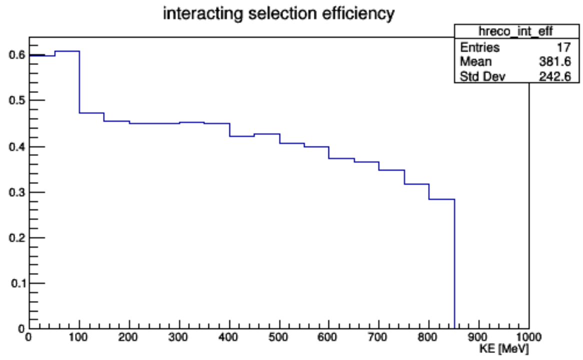

Figure 16: Interacting efficiency.

5.4.2 Linear Algebra

The columns in the matrix shown in Figure 15 are then normalized to 1, so that each bin in a column becomes a weight to be applied in the transformation from reconstructed energy to true energy. A bin from the the true distribution is calculated as follows:

- We take a bin in the reconstructed distribution.
- We multiply that value by the weights in a column in our matrix.
- We add those values to a true distribution along the true axis of our matrix.
- We repeat this process for every reco bin until we've built our entire true distribution.

Over the MC sample used to construct this unfolding matrix, this method recovers the exact same true distribution that was used to populate the inversion.

5.5 Efficiencies

Using the MC we come up with an estimate for the inefficiency in our selection for both the interacting and incident distributions.

5.5.1 Interacting Efficiency

Interaction inefficiencies come from true Geant4 inelastic scatters being missed by the cuts used to flag interaction candidates described above. An example of an inefficiency in this selection is a straight reconstructed track with a true inelastic scatter somewhere along its path. Without kinks or secondary tracks reconstructed our selection procedure will miss this event and so we will under count the interaction distribution.

The selected interacting distribution, after background subtraction and after unfolding, is shown as a ratio of the true interacting distribution from the same sample in Figure 16.

5.5.2 Incident Efficiency

The inefficiency in the incident distribution is closely related to the impurity in the interacting distribution. We'll go over this section through an example as well.

Figure 17 shows another LArIAT MC event display. This time our selection procedure has mistakenly identified an interaction candidate which is in fact not an inelastic scatter. As a result we fill no incident entries after this slab at the reconstructed level. However at the truth level, we continue past this point for a larger total of incident entries for the same event. For this event we have under counted our denominator.

In general, every contribution to the background distribution in the numerator will manifest itself as an inefficiency in the denominator that must be accounted for and corrected. The ratio of the selected distribution and the true distribution for the incident histogram are shown in Figure 18.

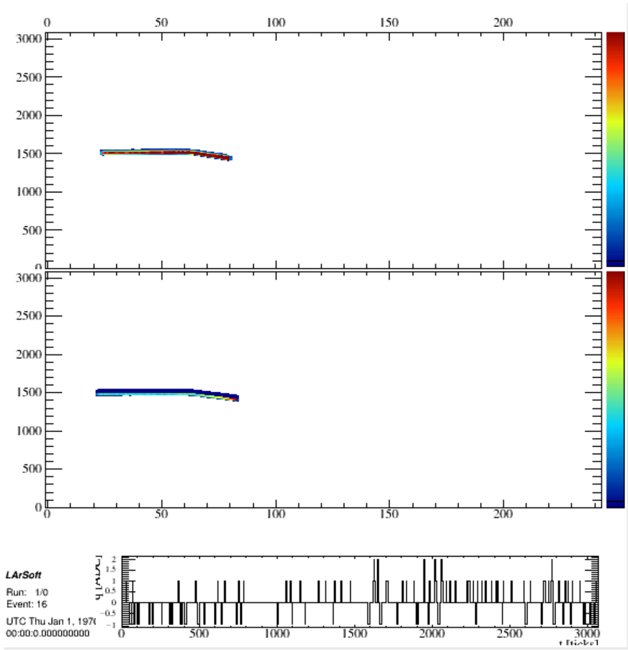

Figure 17: LArIAT MC event display of an event that our selection procedure has incorrectly flagged as an inelastic collision.

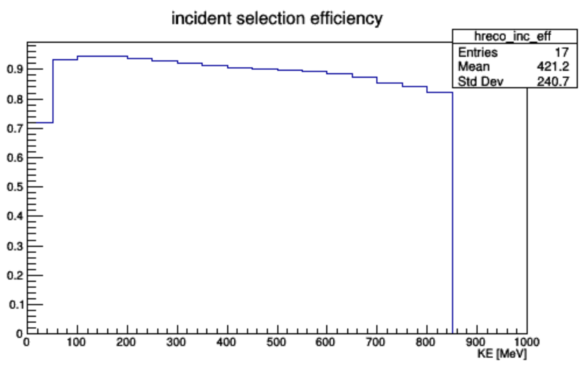

Figure 18: Incident Efficiency.

6 Statistical Uncertainty

The statistical uncertainty in this measurement requires some care. Recall our measurement comes from the following equation:

$$\sigma = \frac{A}{\rho z N_A} \frac{\mathbf{U^{ij}}(N_{selected} - B_{selected})}{\mathbf{U^{ij}}(N_{incident} - B_{incident})} \frac{\epsilon_{incident}}{\epsilon_{selected}} \tag{9}$$

To get a full picture of the uncertainty in the cross-section we need to take into account the following terms:

- The uncertainty associated with background subtraction. Every scatter candidate is either signal or background, the exact populations of each distribution introduces a standard binomial error.
- The ratio of scatters to incident similarly introduces a binomial error term.
- The efficiencies for both the interacting and incident distributions both introduce binomial error.

An individual error binomial error terms goes like:

$$\delta x = \frac{1}{N}\sqrt{k(1-x)} \tag{10}$$

where N is the sample size, k is the number events that pass a cut in question, and x is the ratio k/N.

To take into account the scatter/incident error as well as the background subtraction error we add those terms in quadrature as follows:

$$\delta ratio = \sqrt{(\delta scatter)^2 + (\frac{\delta B_{int}}{B_{int}})^2 + (\frac{\delta B_{inc}}{B_{inc}})^2} \tag{11}$$

Where each term is itself a binomial error. The final total statistical uncertainty is this $\delta ratio$ added in quadrature with the efficiency uncertainties:

$$\delta xs = \sqrt{(\frac{\delta ratio}{ratio})^2 + (\frac{\delta \epsilon_{int}}{\epsilon_{int}})^2 + (\frac{\delta \epsilon_{inc}}{\epsilon_{inc}})^2} \tag{12}$$

This estimate of the uncertainty is an overestimate. This is because we have not taken into consideration correlations in the correction terms that we know are built into the measurement from Section 5. Rather than build covariance matrices to more accurately asses the MC uncertainty in this analysis our goal is to produce a large enough MC sample to drive the correction uncertainty down to zero. In data the statistical uncertainty will depend only on the binomial scatter uncertainty.

7 Systematic Uncertainty

There is an additional source of uncertainty in this measurement that comes from the potential bias of using a particular Geant4 model in estimating the corrections we use to calculate the cross-section. Since different models have different kinematics after an interaction they may give us different estimates for the efficiency and backgrounds present. In reality there is some true distribution for each of these corrections and these models serve as estimates of those true curves. We can get a sense of the model dependence in our measurement by swapping the two leading models at this energy range. The strategy we'll use is to come up with a set of corrections using an alternative model and apply it to our original sample. The deviation is some combination of statistical and systematic uncertainty. Our claim will be that deviation greater than statistical uncertainty is the systematic uncertainty.

Figure 19 shows the same sample used to calculate a cross-section with two different sets of corrections from two different models. The two models we compare are the Bertini Cascade and the Binary Ion Cascade.

From the previous section we know that across most bins the statistical uncertainty floats around three percent. The ratio of these two curves stays within six percent for almost all bins. We'll quote 3 percent as the systematic uncertainty in data.

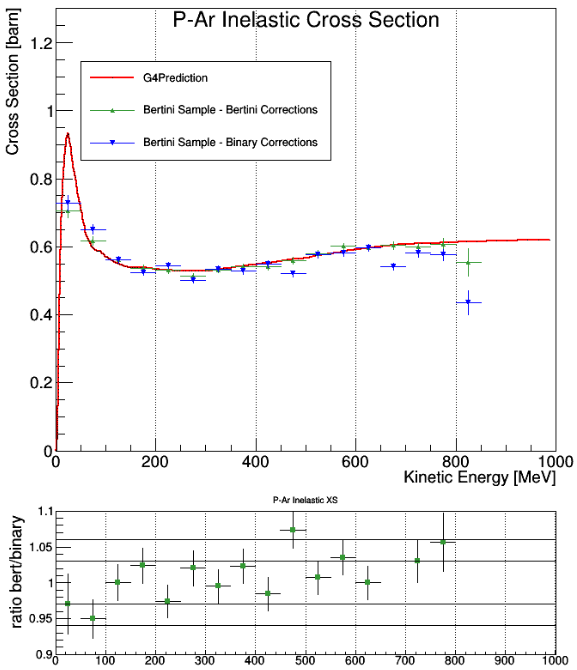

Figure 19: Cross-section calculated with two different sets of corrections.

8 Future Work

This concludes the MC analysis for calculating a cross-section in LArIAT data. This analysis will need to be reproduced with the final reconstruction and Data Driven MC sample to get final corrections we'll apply on data.

There is also an opportunity to validate the models used in this analysis independent of the cross-section we calculate. This study will require comparing final state kinematics after an inelastic collision in MC and data. Additionally, we can evaluate the performance of uboone and SBND PID algorithms after an inelastic collision in LArIAT data and MC.