

# Transverse Momentum in Nucleons, From Raw Data to TMD Extraction

David M. Riser

March 2018



# **Contents**



# Chapter 1

## Introduction

The majority of this thesis explains how we performed our experimental measurement. The role of this introductory chapter is to provide an explanation of why we performed this measurement, and what exactly it is that was measured. First, a brief statement of the measurement is given. Then, in order to understand why we perform the current measurement, a historical look back at the parallel development of nucleon structure in theoretical and experimental aspects is presented. After obtaining a historical perspective, it is evident why the current study is needed, and the scope of the present measurement is re-stated. Finally, a high-level discussion of how this measurement was performed is presented, leaving the details to the remainder of the document.

### 1.1 Statement of Purpose

This work aims to contribute to the understanding of nucleon structure in the framework of transverse momentum dependent parton distribution functions (TMDs) by measuring structure functions in semi-inclusive deeply inelastic scattering (SIDIS). By measuring the cross section for charged pi-mesons  $\pi^\pm$ , we measure ( $F_{UU,T} + \epsilon F_{UU,L}$ ,  $F_{UU}^{\cos\phi}$ , and  $F_{UU}^{\cos(2\phi)}$ ). By analyzing the beam spin asymmetry (BSA) we measure the ratio

$$A_{LU}^{\sin\phi} = \sqrt{2\epsilon(1-\epsilon)} \frac{F_{LU}^{\sin\phi}}{F_{UU,T} + \epsilon F_{UU,L}} \quad (1.1)$$

for positively charged k-mesons. Finally, we are able to use these structure functions to estimate the model parameters in TMD models.

## 1.2 Nucleon Structure from Experiment and Theory

Nucleon structure is currently a very active field of research which aims to explain the structure of protons and neutrons in terms of the fundamental particles known as quarks. This sub-field was born in 1958 following Hofstadters demonstration that the electron-proton scattering cross sections are not consistent with the predictions based on a pointlike proton theory [[physics-hofstadter](#)]. At the same time, an ever growing number of particles discovered by bubble and spark chamber experiments confused physicists.

Some theoretical progress was made during the 1960s as Murray Gell-Mann proposed the construction of this large number of observed particles from pairs or triplets of fundamental particles called quarks. The proposed quark came in 3 flavors called up, down, and strange. However, in order to correctly predict the charge of the observed particles the quarks had to be fractionally charged (up 2/3, down -1/3, and strange 1/3). Accordingly, the quarks had to be permanently confined somehow inside of bound states. This idea was troubling to many physicists, and support for the quark model only really established itself after the  $J/\psi$  ( $c\bar{c}$ ) particle was observed in 1974 and the inclusion of the charm quark into the existing set of 3 quark flavors accommodated its existence. This quark model picture enjoyed some success predicting the masses of hadrons, but gives no information about the dynamics of quarks in hadrons.

Almost all modern measurements that relate to quark dynamics are done by scattering electrons on protons or proton/neutron pairs in deuterium. Since this process proceeds via virtual photon exchange between the charged electron and the charged quarks, it is easier to understand using the tools of QED. If the virtual photon is of sufficiently high energy ( $\approx 2\text{GeV}$ ) its wavelength is an order of magnitude less than the diameter of the proton/neutron and likely interacts with just one part (quark or gluon) of the nucleon. This idea is known as the parton model, and was proposed by Feynman in 1969 [[physics-feynman-1969](#)].

For the 30 years following this development, the measurement of parton distribution functions (PDFs) was performed at experiments around the world (HERMES, COMPASS, and Jefferson Lab). The PDFs (at leading order) describe the probability to observe a quark with a given momentum fraction  $x$  in a hadron. There is one such function for each quark flavor in each hadron. The unpolarized PDF is now known quite well, and the polarized PDFs have also been measured. During the measurement of polarized PDFs in 1989 the European Muon Collaboration (EMC) observed that only 30% of the total spin of the proton appeared to be due to the spin of the quarks. This result came to be known as the proton spin crisis.

One possible resolution to this problem is that the quarks carry orbital angular momentum inside of the proton and this contributes to the total observed spin. In this case, measurements of the three-

dimensional momentum structure of the quarks inside of hadrons is expected to be very useful. The transverse momentum dependent parton distribution functions (TMD PDFs) describe the quark momenta in both the longitudinal direction  $x$  (defined by the hard momentum transfer direction) and the momentum in the plane transverse to that as well  $p_T$ .

### **1.3 Measurement of Semi-Inclusive Deeply Inelastic Scattering with CLAS**

Our purpose is re-stated clearly.

### **1.4 Overview of our Measurement**

Let's talk about detectors, accelerators, and software written in c++.



# Chapter 2

## Basic Analysis & Corrections

### 2.1 Introduction

This chapter discusses analysis procedures that are common to the subsequent data analyses of pions and kaons. These procedures can be divided into two groups. The first type of basic analysis described is the aggregation or calculation of scalar values over the run-period (examples include luminosity and helicity). The second type of analysis procedure described is a correction to measured values. Vertex corrections, timing corrections, and kinematic corrections will be discussed.

### 2.2 Luminosity Calculation

A useful concept in accelerator/collider physics is the luminosity  $\mathcal{L}$ . Luminosity is defined as the number of collisions per unit area per unit time that could lead to some process of interest. Consider as an example elastic scattering of electrons from protons, the luminosity is the number of electron-proton collisions per unit time per unit area. The rate  $\frac{dN}{dt}$  of the occurrence of events for some process  $X$  can be written in terms of this luminosity and the cross section for the process.

$$\frac{dN_X}{dt} = \mathcal{L}\sigma_X \quad (2.1)$$

For the fixed target case, the luminosity has a simple expression.

$$\mathcal{L} = \frac{j_e \rho_p l_T}{e} \quad (2.2)$$

Here  $l_t$  is the target length,  $\rho_p$  is the proton number density in the target, and  $j_e$  is the beam current. To find the total number of events which accumulate in some time  $t_{exp}$  the event rate is integrated with respect to time.

$$N_X = \int_0^{t_{exp}} \frac{j_e \rho_p l_T}{e} \sigma_X dt = \frac{\rho_p l_T}{e} \sigma_X \int_0^{t_{exp}} j_e dt = \frac{\rho_p l_T}{e} \sigma_X \Delta Q \quad (2.3)$$

Thus the experimentally observed cross section for some process  $X$  is,

$$\sigma_X = \frac{N_X}{\mathcal{L}_{int}} \quad (2.4)$$

where the number of events  $N_X$  is corrected for all effects and  $\mathcal{L}_{int}$  is the integrated luminosity as shown above.

Experimentally, the factor  $\Delta Q$  can be calculated from charge deposition measurements by the Faraday cup. The Faraday cup charge is a scalar value written periodically into the output event stream, not with every recorded event. This information is stored in the output BOS files in a bank called TRGS, the variable is named FCUP\_G2.

For this data analysis, the final ntuple (root files) used did not contain the Faraday cup charge information. For this reason, the authors used the BOS files directly and recorded the value of FCUP\_G2 for every scalar reading, as well as the event number directly after each scalar entry (from the HEAD bank). This event number correlates directly to the event number stored in the root files used for analysis.

The total accumulated charge over a run is simply the sum over consecutive differences in the Faraday cup charge.

$$\Delta Q = \sum_{i=1}^{n-1} q[i+1] - q[i] \quad (2.5)$$

Here  $n$  denotes the number of scalar entries for a given file. Due to the periodic nature of the scalar bank writing events are also recorded after the last reading of the file, and before the first scalar reading of the next file in the run. To account for this the difference between consecutive files last and first readings is added to the total.

For the E1-F dataset, a run typically contains around 20 files, each representing a raw file size of 2 gigabytes. These files are named by run number, and given an index from 0 to  $n_{files} - 1$ . It is not uncommon that a run will contain missing files in the middle of the range. If this occurs, the charge difference between last/first reading is not added to the total.

Any charge which accumulates in a period of time where the number of events did not change is not

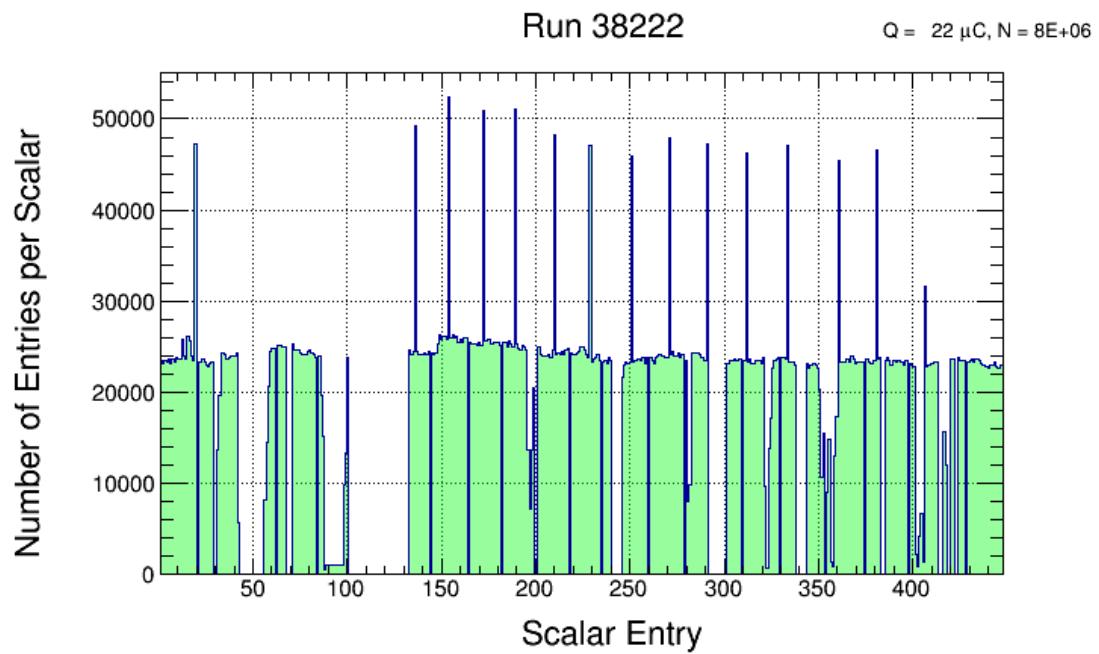


Figure 2.1: Entries per scalar reading for run 38222.

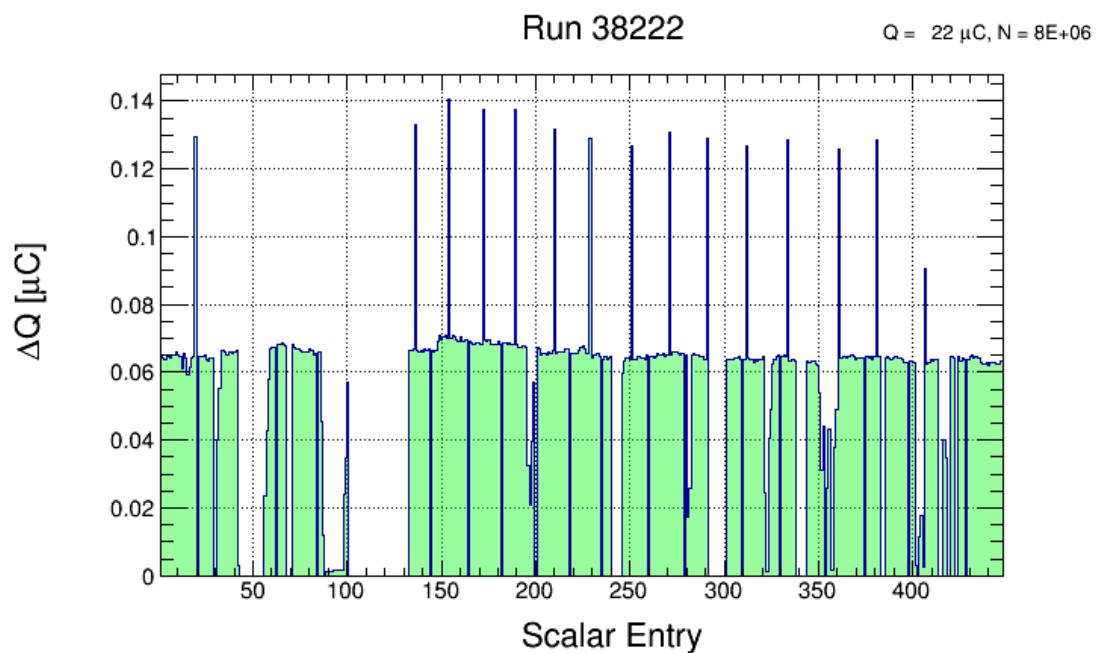


Figure 2.2: Charge per scalar reading for run 38222.

added to the total. Similarly, any events which occur within regions where no charge is recorded need to be discarded. This is accomplished by recording the bad event ranges for every file and removing these events from our analysis.

The result of this procedure is a numerical value of charge for each run. For practitioners, it is important to note that this value needs to be scaled by the DAQ scaling factor before it represents a value of charge. In our analysis, the numerical value of charge for a typical file is a few tens of micro-Coulombs.

## 2.3 Determination of Good Run List

The total dataset contains 831 runs. Due to the complex experimental setup, it is not uncommon for run conditions to change during a few of the runs such that the data collected are not of analysis quality. Imagine as a simple example that the liquid hydrogen target boils, the density is suddenly decreased, and the number of recorded events drops drastically (but the Faraday Cup charge would look the same). For this reason, a good run list is constructed and used in the analysis.

To construct this list, we simply count good electrons in every file and normalize that by the accumulated charge for that file. While the number of events collected varies from run to run the ratio defined above is a stable quantity – provided that the run conditions do not vary greatly. Runs which are within 3 standard deviations of the mean (calculated over the dataset) are used as good runs. The good run list used for this analysis contains 522 runs.

## 2.4 Helicity Determination

During the course of the E1-F run period the beam helicity convention was changed by the insertion of a half-wave plate at the injector. Our definition of + and – helicity must change in accordance with these wave-plate insertions. To monitor these changes, the value of  $A_{LU}^{\sin \phi}$  for  $\pi^+$  is recorded for every run. Whenever the asymmetry (which has a magnitude of around 3%) changes sign, we know that the sign convention has changed. These changes are then taken into account in the data analysis.

## 2.5 Vertex Corrections

Vertex information ( $v_x, v_y, v_z$ ) is calculated based on the intersection of a track with the midplane (the plane which contains the beamline and bisects the sector at  $\phi_{rel} = 0$ ). If the beam is not centered at

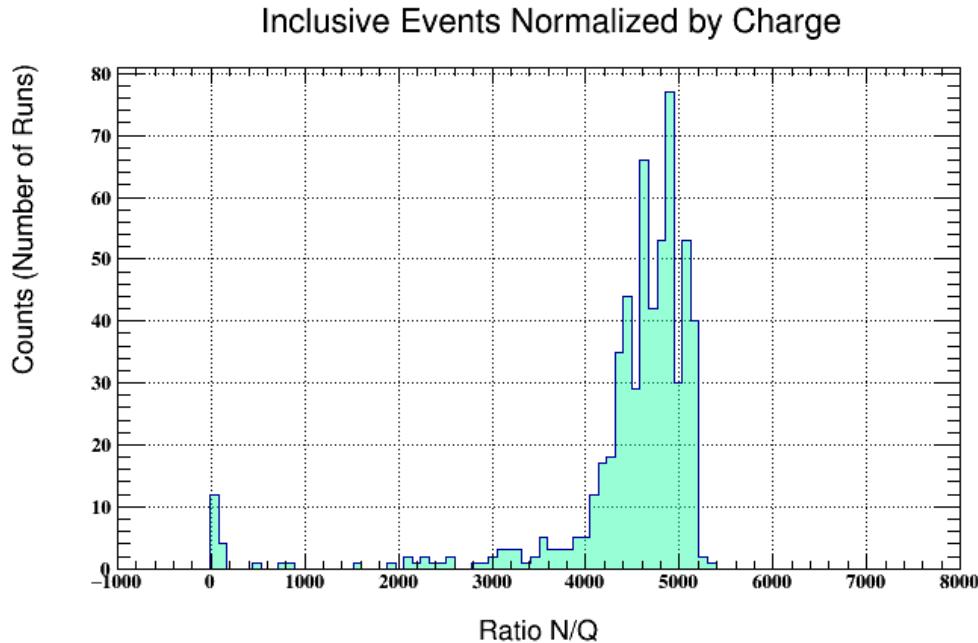


Figure 2.3: Inclusive electrons per file normalized by the total charge accumulated for the file. This quantity is used to make a good run list.

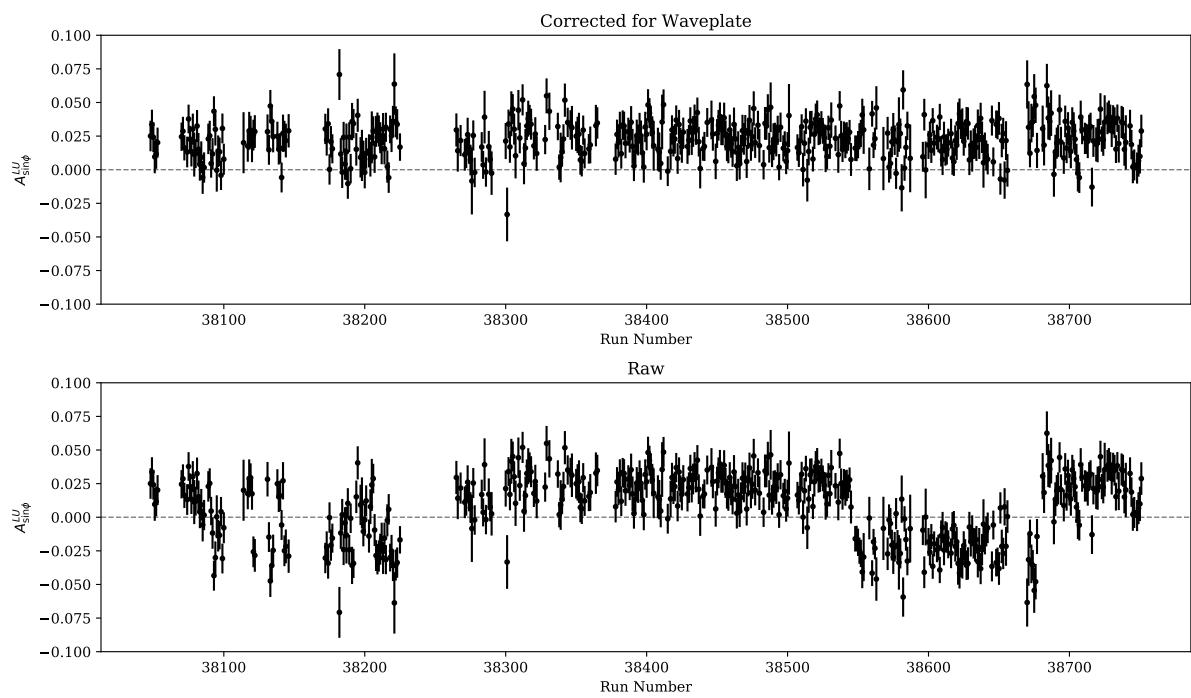


Figure 2.4: The waveplate position is determined and corrected by plotting the BSA for  $\pi^+$  mesons as a function of the run. The top panel shows the corrected results, the bottom shows the results before changing the helicity.

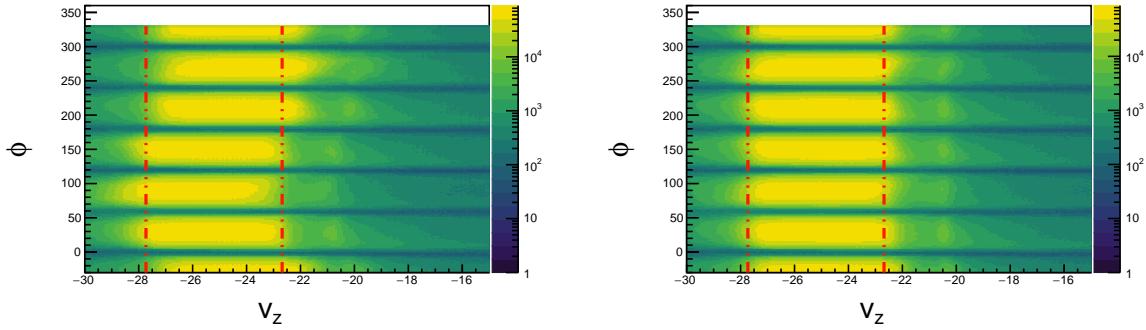


Figure 2.5: The z-vertex  $v_z$  position shown for different values of  $\phi$  the azimuthal angle in the hall. The left figure shows the distribution before corrections are applied, the right after. The vertical red lines bound the region which we define as acceptable for electrons in our analysis.

$(x, y) = (0, 0)$ , the calculation of vertex position needs to be corrected by shifting the midplanes in accordance with the target offset. This offset  $(x, y)$  is identified by plotting events from the control foil placed near the target, which has a well defined z position of  $-2$  cm.

For the E1-F run period, the beam position was  $(0.15, -0.25)$  cm.

## 2.6 Timing Corrections

Timing information comes from the time-of-flight detector system. After normal calibration, small offsets in timing between time of flights paddles still exist for the E1-F dataset. These biases can be removed on a run-by-run and paddle-by-paddle basis by adding a small shift  $t_{corr}$ . In order to determine this shift  $t_{corr}$  for each paddle, charged pions are used.

Using momentum information from the drift chambers the value of  $\beta$  can be predicted and the difference  $\Delta\beta$  can be determined for each electron/pion pair.

$$\Delta\beta = \beta_{obs} - \beta_{pred} = \frac{d}{ct_{obs}} - \sqrt{1 + (m/p)^2} \quad (2.6)$$

Here m is assumed to be  $m_\pi$ . The offset  $\Delta\beta$  from 0 is used to define the value of  $t_{corr}$  for each paddle. If this value is exceedingly small, no correction is applied. For some paddles with low statistics a reasonable value for  $t_{corr}$  cannot be obtained and these paddles are excluded from the analysis.

In the method described above, the calibrated paddle is the one which is struck by the pion. The electron paddle which was struck could also require calibration. In practice the magnitude of the correction term  $t_{corr}$  is small, and the paddle offset is (likely) randomly distributed about 0 when considering

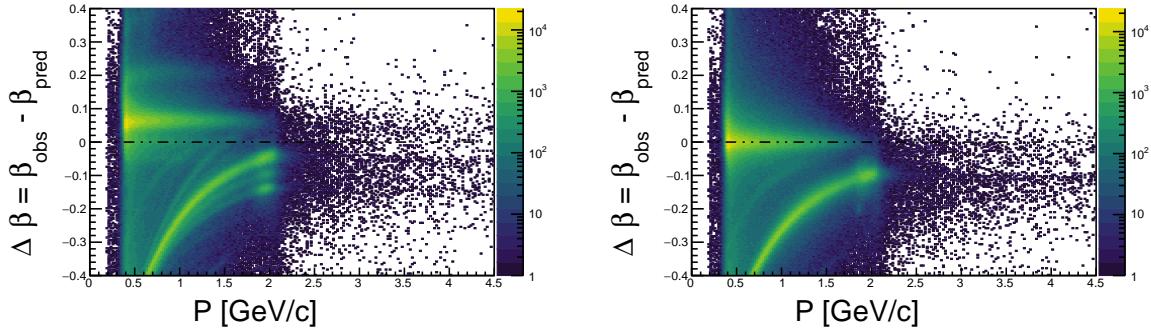


Figure 2.6: Timing corrections are shown for paddle 24 of sector 1. The left image shows the  $\Delta\beta$  distribution before corrections. On the right the same is shown after correction of the timing for this paddle. We assume the mass of the track to be the pion, these show up as the green band. Heavier protons are visible below the pion band.

all paddles. By including events from many different (electron) paddles, miscalibration effects from the electron side cease to be important. This is demonstrated by the success of the technique in centering the  $\Delta\beta$  distributions.

The work described in this section was performed by Nathan Harrison for the E1-F dataset.

## 2.7 Kinematic Corrections

The magnetic field map used in reconstruction to swim particle tracks cannot perfectly match the real magnetic field of the hall. As a result of this the reconstructed momentum of particles is often slightly off (on order of 1%). Small misalignments in detector positions also contribute to this effect. In order to correct for these small differences, the kinematics ( $p, \theta$ ) of charged tracks is corrected.

Various procedures exist for the correction of kinematic variables of measured particles, and they all rely on energy and momentum conservation applied to standard processes (such as elastic scattering). The procedure used to derive corrections for the E1-F dataset was developed and described by Marco Mirazita in [\[physics-mirazita\]](#).

In this analysis the need for correction to  $\theta$  (the polar angle measured from the beamline) arises from misalignments in the drift chambers. This implies that the correction will be the same for positives and negatives, and this assumption is used in the correction algorithm. First elastic ( $ep \rightarrow ep$ ) events are selected by identifying electron/proton events and using a missing mass around 0. The kinematics of the event are then calculated.

$$k^\mu = (k, 0, 0, k) \quad (2.7)$$

$$p^\mu = (M_p, 0, 0, 0) \quad (2.8)$$

$$k'^\mu = (k', k' \sin \theta, 0, k' \cos \theta) \quad (2.9)$$

$$p'^\mu = (E_p, -p' \sin \alpha, 0, p' \cos \alpha) \quad (2.10)$$

Applying energy and momentum conservation to the equations above yeilds 3 equations.

$$k + M_p = k' + \sqrt{M_p^2 + p'^2} \quad (2.11)$$

$$k' \sin \theta = p' \sin \alpha \quad (2.12)$$

$$k = k' \cos \theta + p' \cos \alpha \quad (2.13)$$

Using these equations, the electron angle  $\theta$  and the proton angle  $\alpha$  can be predicted by using the momenta  $(k', p')$ . These values are compared with measured values and iteratively corrected by tuning the parameters of a phi-dependent 2nd order polynomial.

$$\cos \theta = 1 - M_p \frac{k - k'}{kk'} \quad (2.14)$$

$$\tan \alpha = \frac{1}{p'} \frac{k' \sin \theta}{k - k' \cos \theta} \quad (2.15)$$

After  $\theta$  corrections are applied, the momentum of the electrons are corrected by using an analogous procedure for  $k'$  instead of  $\theta$  and  $\alpha$ . The momentum corrections are calculated as functions of  $\phi$  for each sector in one degree bins of  $\theta$ . Finally, the positively charged particles momenta are corrected by selecting the exclusive event ( $ep \rightarrow e\pi^+N$ ). In this reaction the scattered electron and pion are detected and the neutron is selected using a missing mass cut. Assuming the electron momentum, electron angle, and pion angle to be correct, the pion momentum correction is then calculated by iteratively improving the central position of the neutron mass peak to coincide with  $M_N$ . Marco Mirazita shows in his note that these corrections can be satisfactorily applied to all negative (coming from electron) and positive (coming from positive pi-meson) particles.

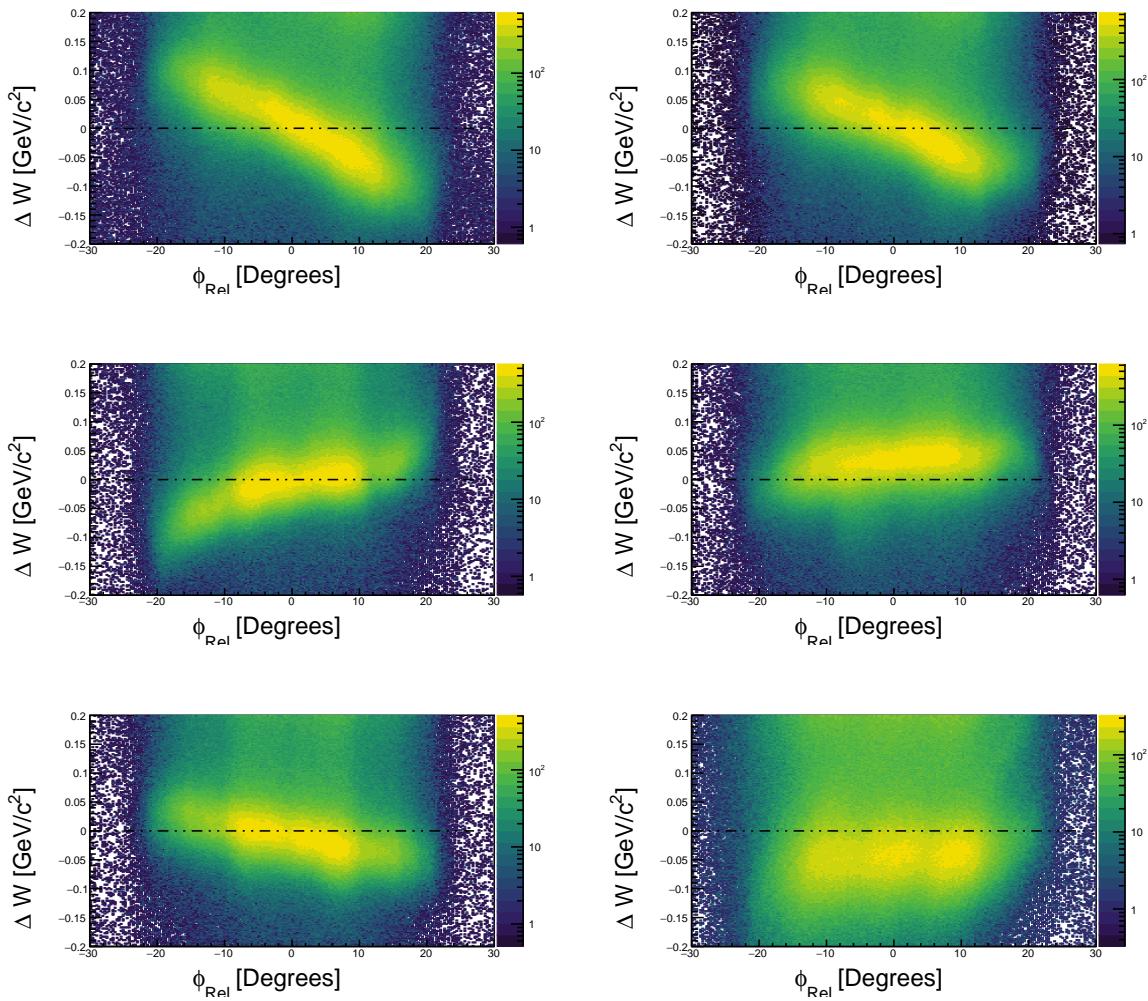


Figure 2.7: This figure shows the deviation from  $M_p$  of the  $W$  spectrum peak for elastic  $ep \rightarrow ep$  events (before corrections).

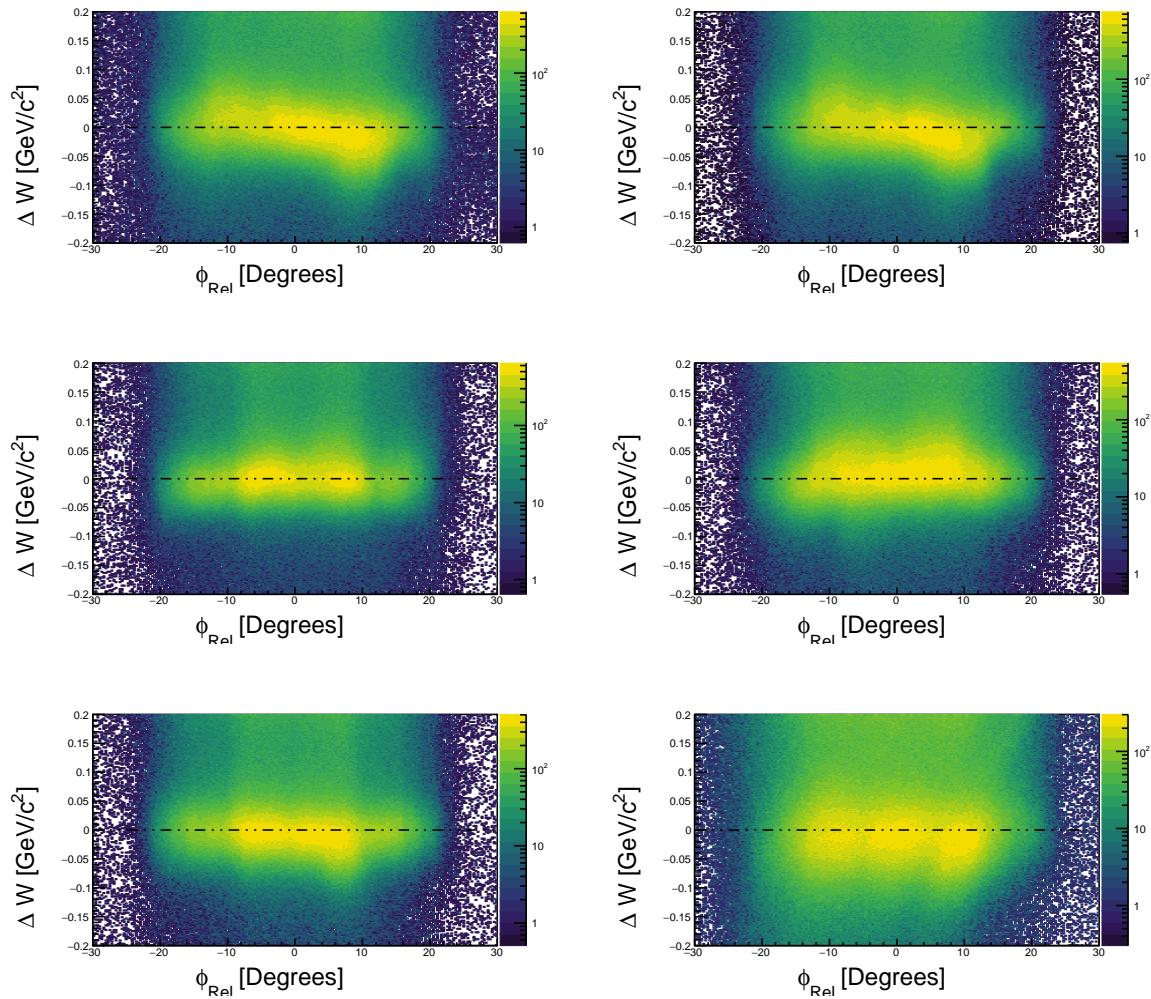


Figure 2.8: This figure shows the deviation from  $M_p$  of the  $W$  spectrum peak for elastic  $ep \rightarrow ep$  events (after  $\phi$ -dependent corrections).

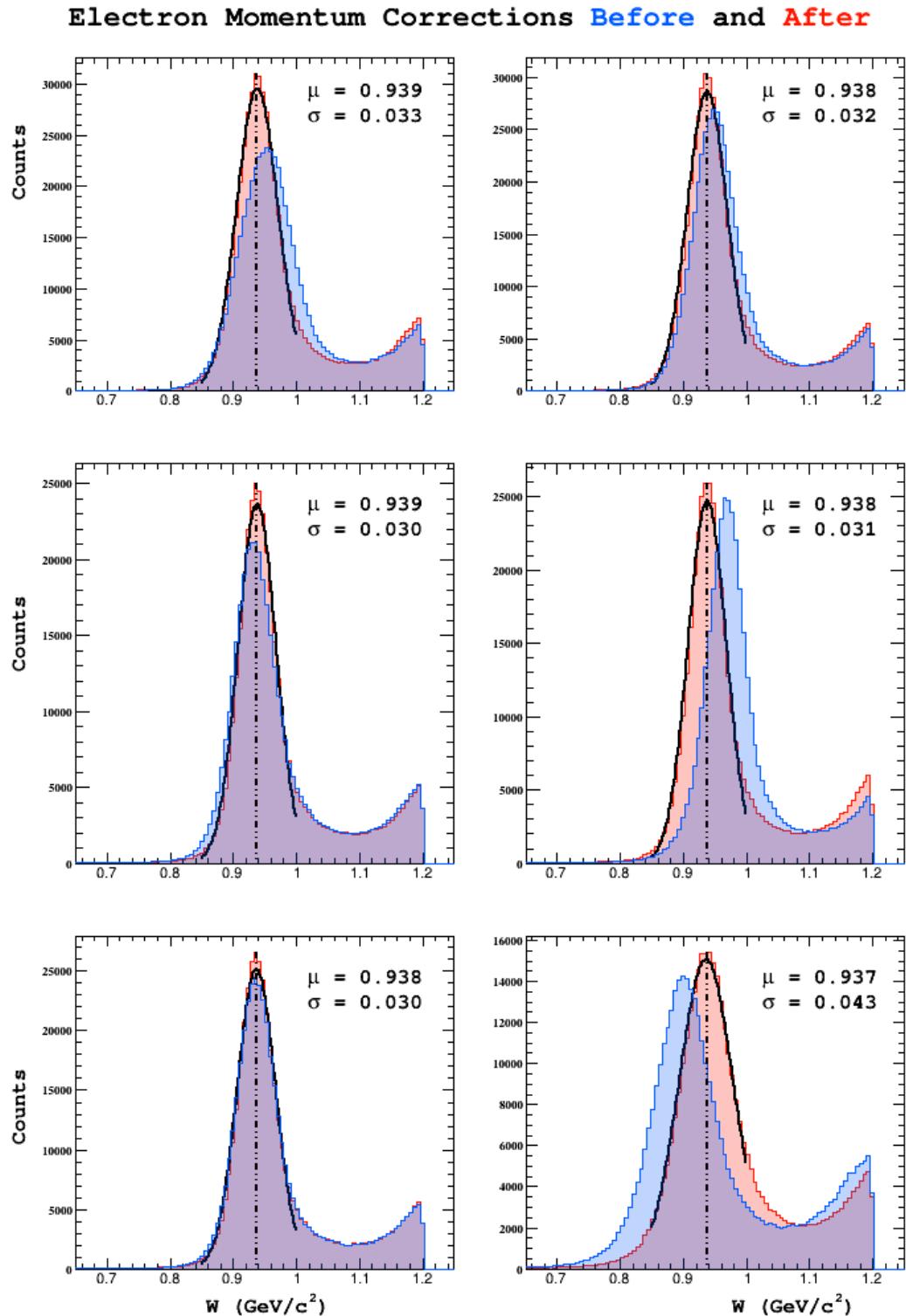


Figure 2.9: Elastic events shown in the spectrum of  $W$  before and after momentum corrections are applied.



# Chapter 3

## Particle Identification

### 3.1 Introduction

Particle identification (PID) is the process of classifying tracks as known particles. After reconstruction and matching of detector responses to each track, the reconstruction package `reccsis` assigns a preliminary particle identification based on loose selection criteria. In this analysis, tracks are classified based on a more stringent criteria. This chapter discusses the methodology used by the authors to classify particles.

### 3.2 Electron Identification

Electrons in CLAS are abundant, and the detection of an electron is a basic necessity for every event that will be analyzed. Each negative track is considered a possible electron, a series of physically motivated cuts is applied. If a track passes all cuts, it is identified as an electron. All track indices which pass electron identification are saved, and the one with the highest momentum is used in the analysis.

#### 3.2.1 Electron ID Cuts

The cuts used to select electrons are enumerated below.

- Negative charge
- Drift chamber region 1 fiducial
- Drift chamber region 3 fiducial
- Electromagnetic Calorimeter fiducial (UVW)

- EC minimum energy deposition
- Sampling Fraction (momentum dependent)
- z-vertex position
- Cherenkov counter  $\theta_{cc}$  matching to PMT number
- Cherenkov counter  $\phi_{rel}$  matching to PMT (left/right)

Each cut will now be described in more detail.

### Negativity Cut

Each track is assigned a charge based on the curvature of its trajectory through the magnetic field of the torus. This is done during the track reconstruction phase. The tracks are eliminated as electron candidates if they are not negatively charged.

### Drift chamber fiducial

Negative tracks which pass geometrically close to the edges of the drift chamber are, from a tracking perspective, more difficult to understand. Often these tracks originate from downstream background, or are otherwise unacceptable. Additionally, tracks which fall outside of the fiducial region of the drift chambers are likely to fall outside of the fiducial region of the downstream detectors as well. For these reasons, it is common to remove tracks which are geometrically close to the boundaries of the drift chambers in region 1 as well as region 3 coordinate systems.

To implement this cut the  $(x, y)$  coordinates of the drift chambers are rotated into one sector. Then boundaries  $y_{left}, y_{right}$  are defined as linear functions of  $x$ . The boundary lines are parametrized by an offset  $h$  and an angle of the boundary line with respect to the center of the sector at  $x = 0$ . The slope of these lines is  $\pm \cot(\theta)$ .

$$y_{right} = h + \cot(\theta) \quad (3.1)$$

$$y_{left} = h - \cot(\theta) \quad (3.2)$$

Tracks passing this criterion are those which have  $y > y_{left}(x)$  and  $y > y_{right}(x)$ .

### Electromagnetic Calorimeter fiducial (UVW)

As tracks traverse the electromagnetic calorimeter they develop electromagnetic showers. If the track passes close to the edges of the detector, there is a chance that the shower will not be fully contained

Region	Height $h$ (cm)	Angle $\theta$ (degrees)
1	22	60
3	80	49

Table 3.1: Cut parameters used for the DC fiducial cut.

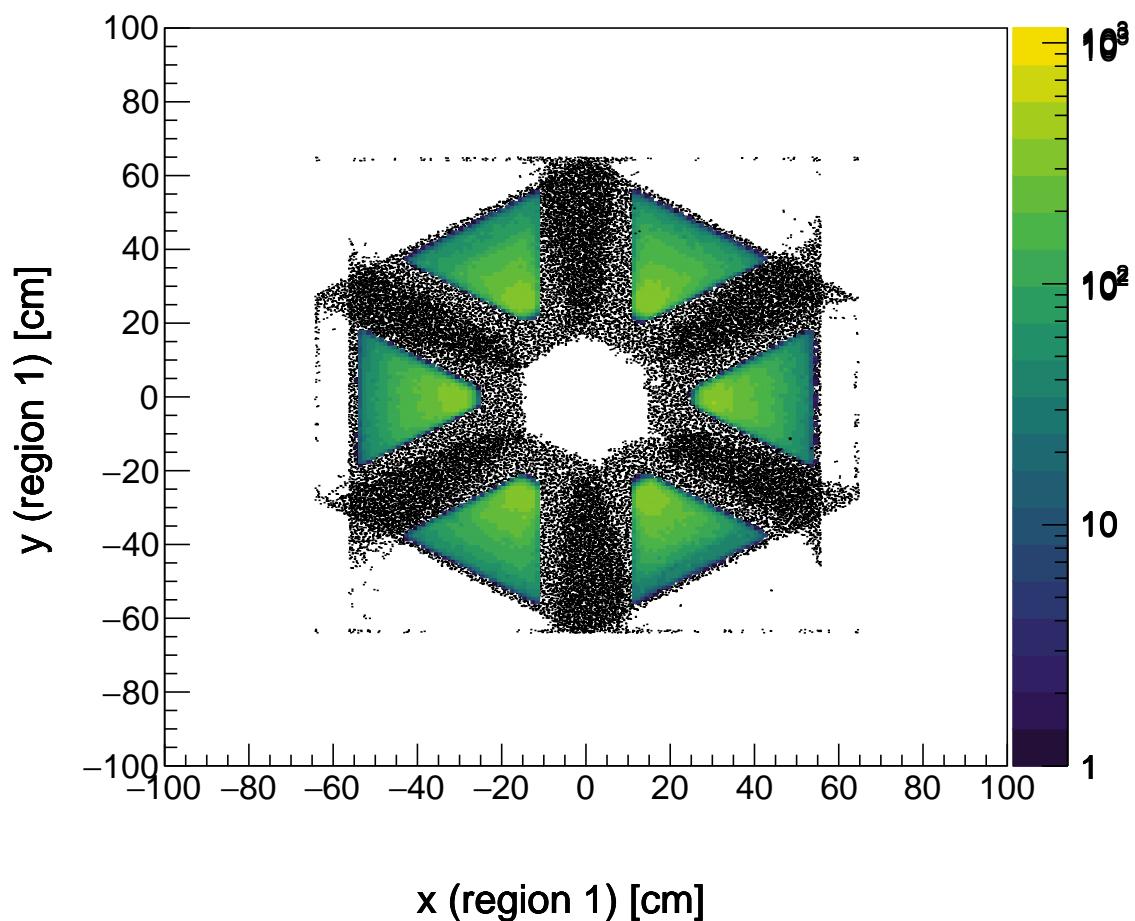


Figure 3.1: Tracks shown in color remain after the application of drift chamber region 1 fiducial cuts to all cuts, shown here as black points.

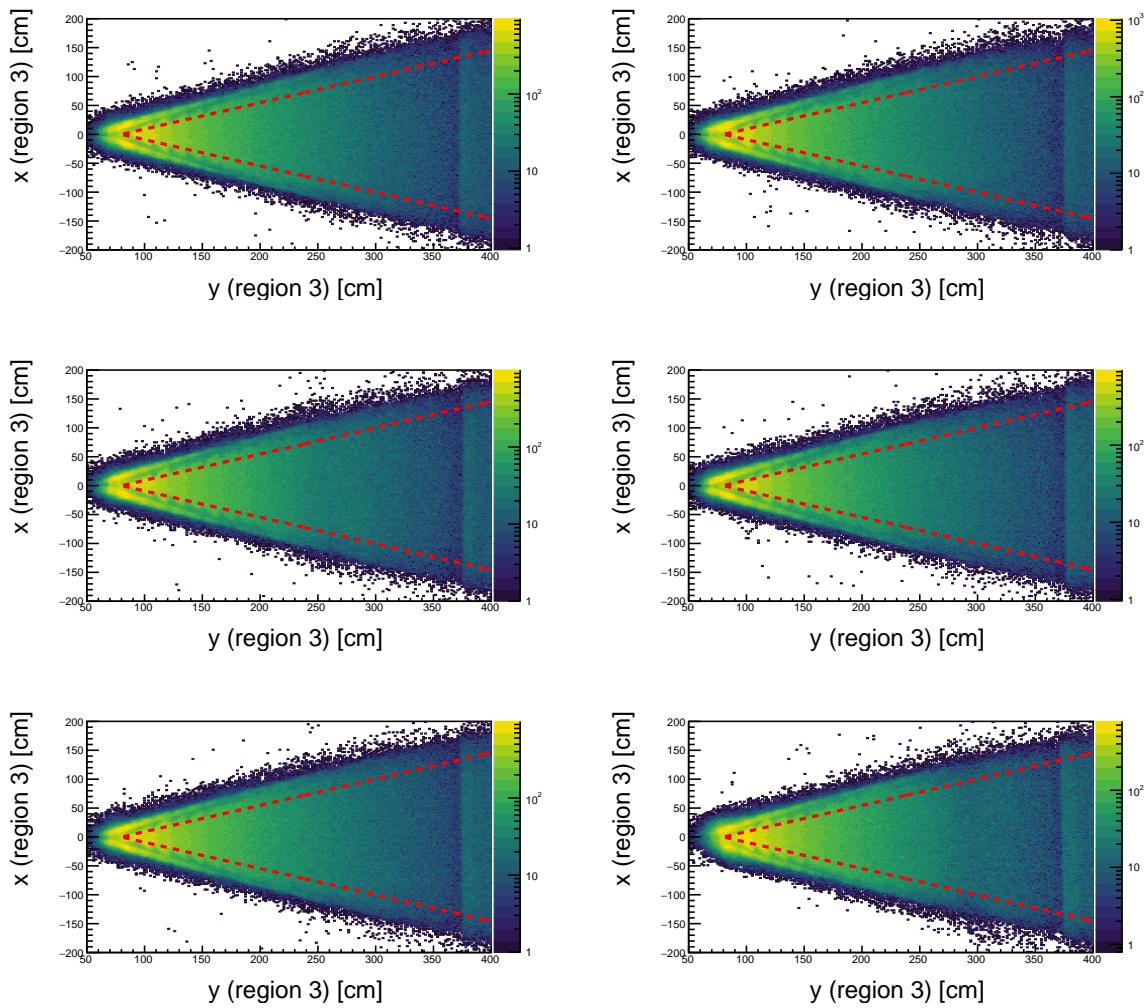


Figure 3.2: The selection criteria shown in red is applied to drift chamber region 3.

EC Coordinate	Min (cm)	Max (cm)
U	70	400
V	-	362
W	-	395

Table 3.2: Cut parameters used for the EC fiducial cut.

within the calorimeter volume (it spills out the edges). For this reason, it has become standard to remove the hits which fall within the outer 10 centimeters of each layer of the EC (10 centimeters is the width of a scintillator bar). This cut is applied in the U, V, W coordinate system.

### EC minimum energy deposition

The negative tracks that start out as electron candidates are primarily composed of electrons and negative  $\pi$  mesons. One way to differentiate between these two species is to exploit the difference in energy deposition between the two in the electromagnetic calorimeter. Electron typically develop a much larger more energetic shower than  $\pi$  mesons, which minimally ionize the calorimeter material. The result is that the total energy deposition is typically larger for electrons than  $\pi$  mesons. In this analysis we require that at least 60 MeV was deposited in the inner calorimeter for electron candidates.

### Sampling Fraction (momentum dependent)

The electromagnetic calorimeter is designed such that electrons will deposit  $E_{dep}/p \approx 0.3$  approximately one-third of their energy, regardless of their momentum. In contrast to this, the ratio  $E_{dep}/p$  for  $\pi$  mesons decreases rapidly with momentum. To develop a momentum dependent cut for this distribution, all negative candidates are first filled into a two-dimensional histogram of  $E_{dep}/p$  vs.  $p$ . The histogram is then binned more coarsely in momentum, and projected into a series of 40 slices. Each of these slices is fit with a Gaussian to extract the position  $\mu_i$  and width  $\sigma_i$  of the electron peak. Finally, the authors choose a functional form for the mean and standard deviation of the distributions to be a third order polynomial in momentum.

$$\mu(p) = \mu_0 + \mu_1 p + \mu_2 p^2 + \mu_3 p^3 \quad (3.3)$$

$$\sigma(p) = \sigma_0 + \sigma_1 p + \sigma_2 p^2 + \sigma_3 p^3 \quad (3.4)$$

Boundaries are constructed from this information by adding (subtracting)  $n_\sigma$  from the mean. In the nominal case, we use  $n_\sigma = 2.5$ .

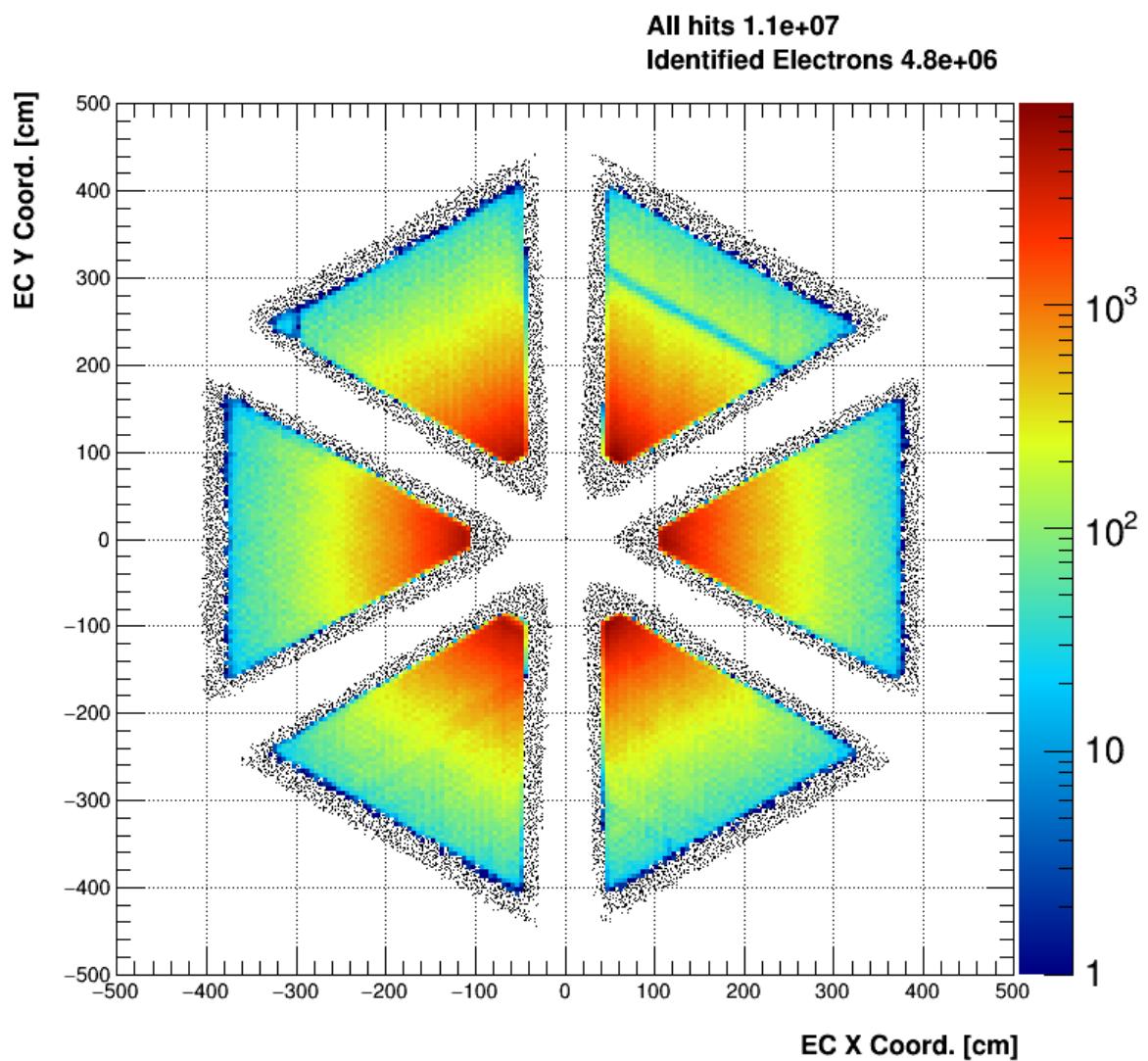


Figure 3.3: All negative tracks are shown here in black. In color, the tracks which pass the EC fiducial cut are shown.

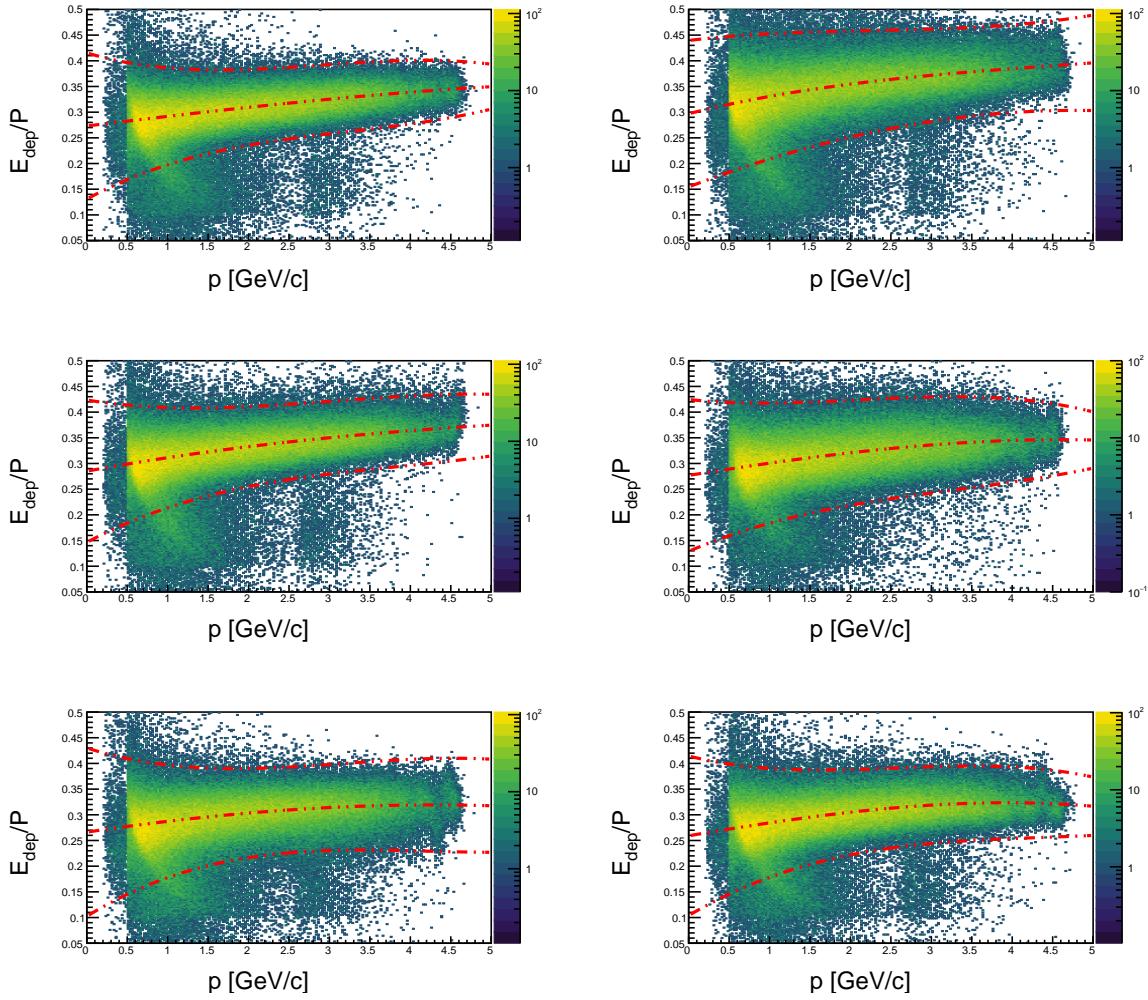


Figure 3.4: The sampling fraction selection boundary is shown here for the nominal value of  $N_{sigma} = 4$ .

$$f_{max}(p) = \mu(p) + n_\sigma \sigma(p) = (\mu_0 + n_\sigma \sigma_0) + (\mu_1 + n_\sigma \sigma_1)p + (\mu_2 + n_\sigma \sigma_2)p^2 + (\mu_3 + n_\sigma \sigma_3)p^3 \quad (3.5)$$

$$f_{min}(p) = \mu(p) - n_\sigma \sigma(p) = (\mu_0 - n_\sigma \sigma_0) + (\mu_1 - n_\sigma \sigma_1)p + (\mu_2 - n_\sigma \sigma_2)p^2 + (\mu_3 - n_\sigma \sigma_3)p^3 \quad (3.6)$$

Due to slight differences between the 6 sectors of the CLAS detector, the authors choose to calibrate and apply this cut for each sector individually. The results are shown in table ??.

### **z-vertex position**

Electrons can be produced as part of  $e^+e^-$  pairs. For this analysis, these are not of interest. The authors choose to select only electrons which originate from the target and are believed to be the scattered incoming electron. For this reason the authors accept only electron candidates which have a z-vertex  $v_z \in [-27.7302, -22.6864]$ . This cut is applied after the vertex position has been corrected (this correc-

Parameter	Sector 1	Sector 2	Sector 3	Sector 4	Sector 5	Sector 6
$\mu_3$	-8.68739e-05	0.000459313	9.94077e-05	-0.000244192	-7.65218e-05	-0.000392285
$\mu_2$	-0.000338957	-0.00621419	-0.00267522	-0.00103803	-0.00222768	-0.00105459
$\mu_1$	0.0191726	0.0393975	0.02881	0.0250629	0.0233171	0.0265662
$\mu_0$	0.2731	0.296993	0.285039	0.276795	0.266246	0.25919
$\sigma_3$	-0.000737136	0.000189105	-0.000472738	-0.000553545	-0.000646591	-0.000633567
$\sigma_2$	0.00676769	-0.000244009	0.00493599	0.00434321	0.00717978	0.00626044
$\sigma_1$	-0.0219814	-0.00681518	-0.0180929	-0.0140827	-0.0246181	-0.022029
$\sigma_0$	0.0474188	0.0475098	0.0461743	0.0492728	0.0546257	0.0517508

Table 3.3:  $\mu$  and  $\sigma$  values used to construct the momentum dependent sampling fraction cut.

tion will be discussed in a subsequent chapter).

### Cherenkov counter $\theta_{cc}$ and $\phi_{rel}$ matching to PMT

The placement of photo-multiplier tubes (PMT) in the Cherenkov counter allows for additional consistency conditions to be applied. The placement of 18 PMTs increasing in polar angle away from the beamline means that the PMT segment number is correlated to the angle which the electron has with the beamline at the Cherenkov counter  $\theta_{cc}$ . Additionally, PMTs that are placed on the left and right of the detector can be used to check consistency with the azimuthal angle the track forms with the central line of the detector (ie  $\phi_{rel} > 0$  means the track was in the right half of the sector,  $\phi_{rel} < 0$  means the track was in the left half of the sector). An integer code is used to describe the PMT associated with the track. The left PMT is assigned value -1, the right 1, and a signal in both PMTs is assigned 0. If both PMTs have a signal, the track is allowed to pass. If the left PMT was the one that had a signal, only events with  $\phi_{rel} < 0$  are allowed to pass. Similarly if the right PMT fired (code = 1), only events with  $\phi_{rel} > 0$  are allowed to pass. Technical note: the integers in question can be obtained from the ntuple22 format tree by doing the following.

---

```
for (int index = 0; index < event.gpart; index++) {
    int pmt = event.cc_segm[index]/1000 - 1;
    int segment = event.cc_segm[index]%1000/10;
}
```

---

## 3.3 Hadron Identification

Hadron identification in CLAS is done by correlating particle momentum from the drift chambers with timing information supplied by the time of flight detector. In this analysis some quality assurance cuts are applied preliminarily, but they do not discriminate between different species of particle. The likelihood methodology described in this section is based on the discussion provided by the BES collaboration in [bes-physics].

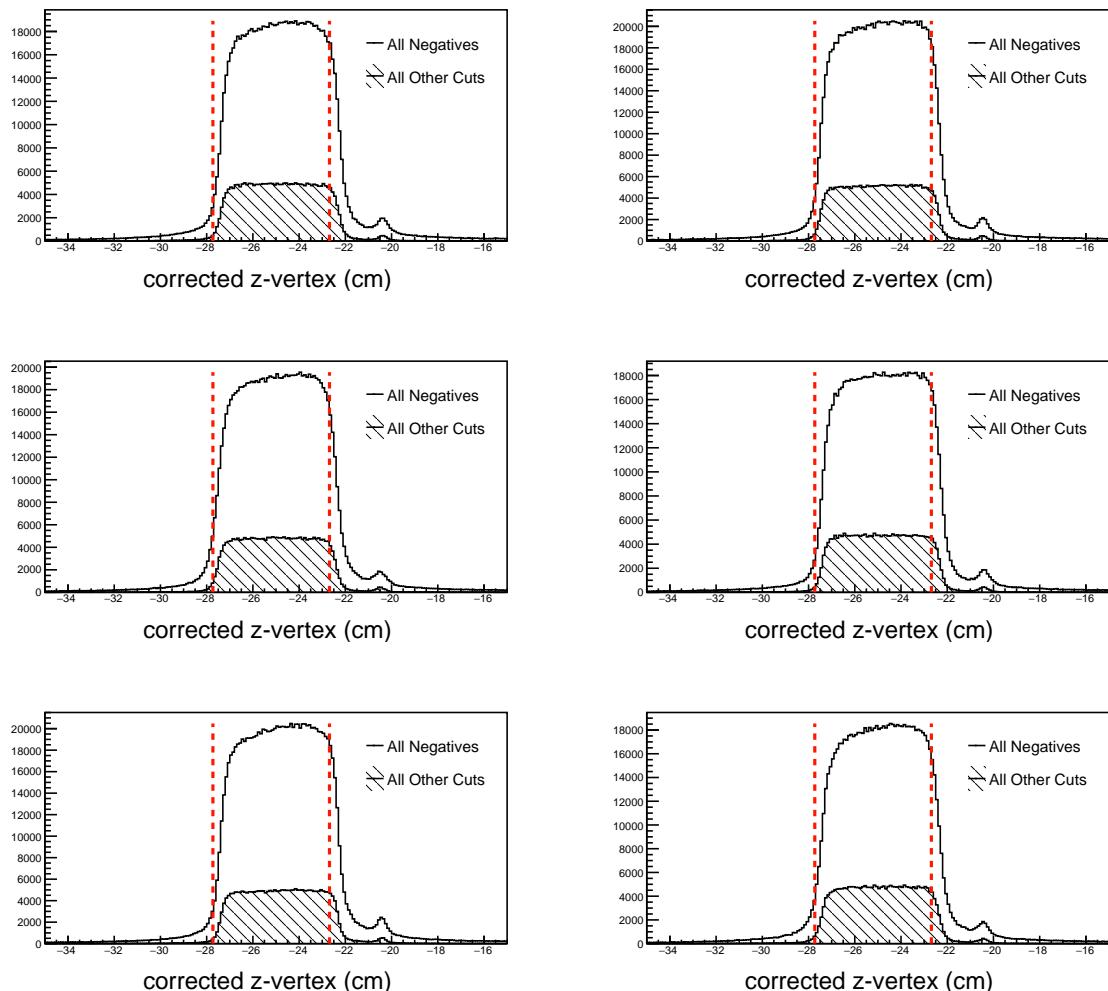


Figure 3.5: The track vertex cut is shown above. All negative tracks are shown in white, while the tracks passing all other criteria are shown in black hatch. The cut boundary is displayed as red lines. For E1-F the target center was located at -25 cm.

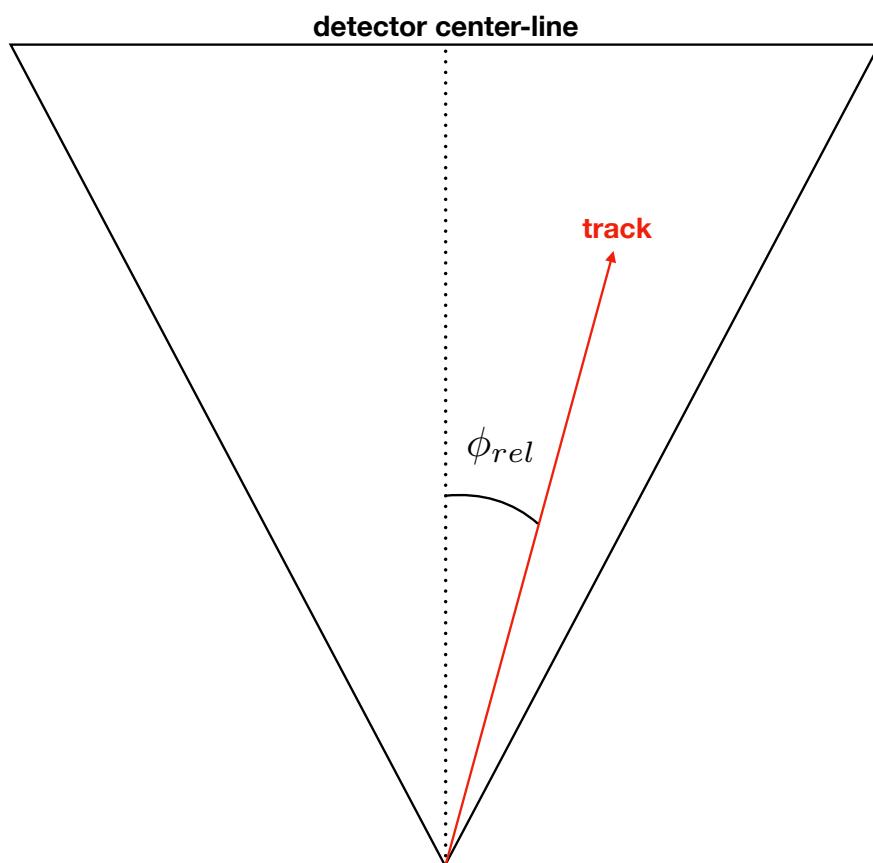


Figure 3.6: The angle  $\phi_{rel}$  is the azimuthal angle between the central line of the detector and the track.

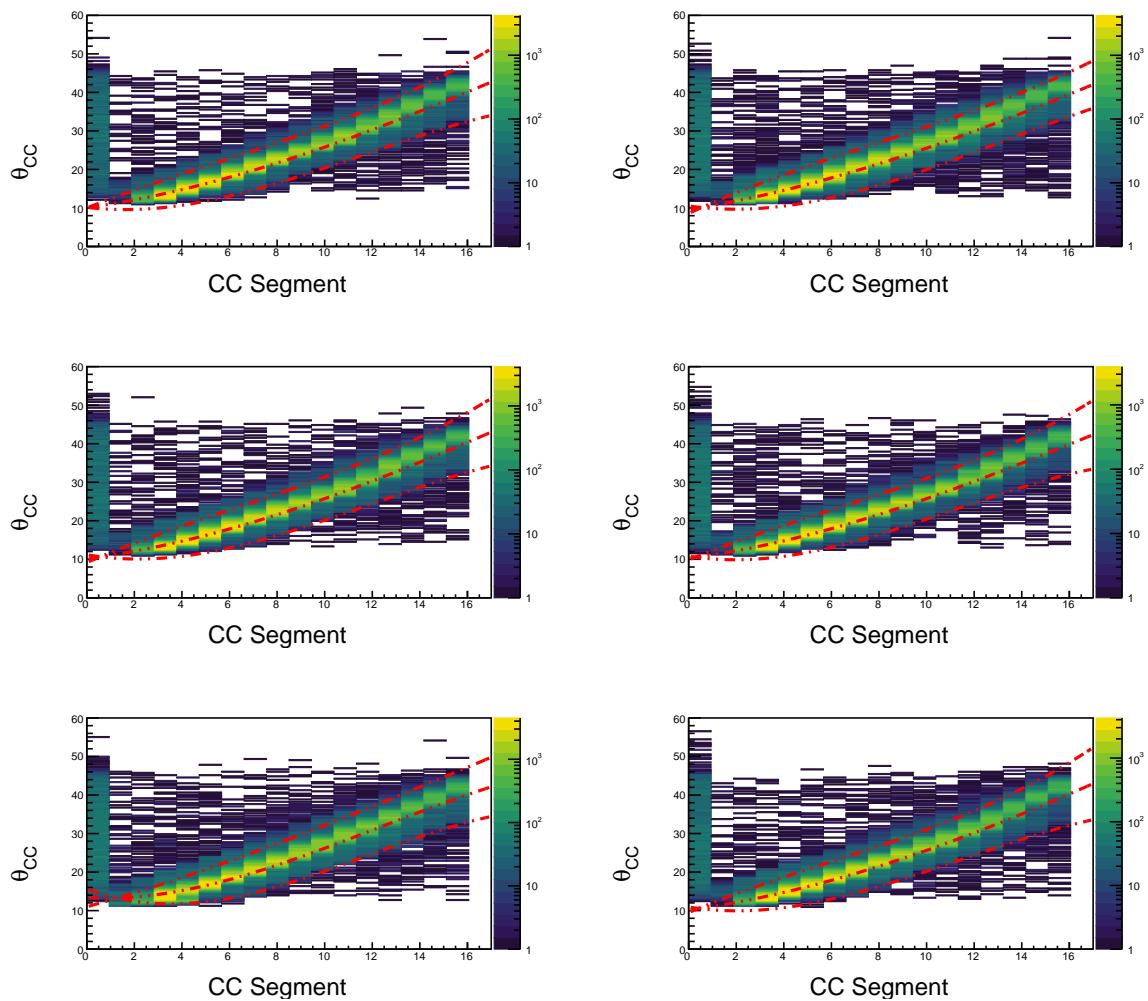


Figure 3.7: Correlation between  $\theta_{CC}$  and the CC segment is shown above, with our selection boundaries overlaid in red.

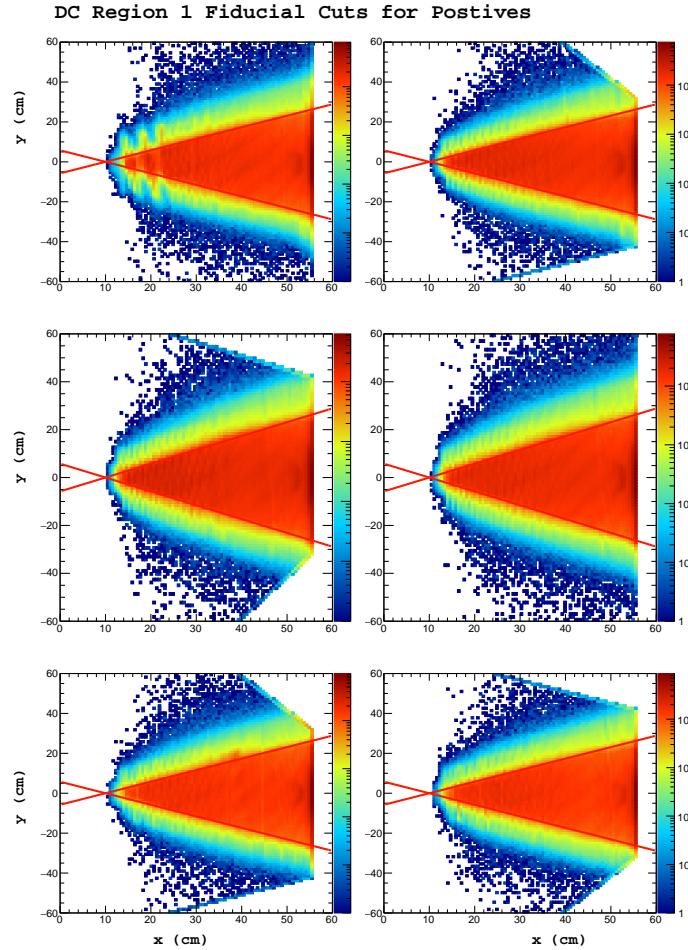


Figure 3.8: Shown above: Positive track hits on the region 1 drift chamber, events falling between the red lines are kept for analysis.

### 3.3.1 Hadron ID Cuts

The cuts used by the author for hadron classification are enumerated below.

- Drift chamber fiducial
- Hadron-electron vertex difference
- Likelihood maximization of  $\beta(p, h)$

#### Drift chamber fiducial

Drift chamber fiducial cuts are applied (only region 1) using the same procedure as described for electrons. The parameters are for negative hadrons are those which are used for the electron. The parameters used for positive tracks are  $h = 10, \theta = 60$ .

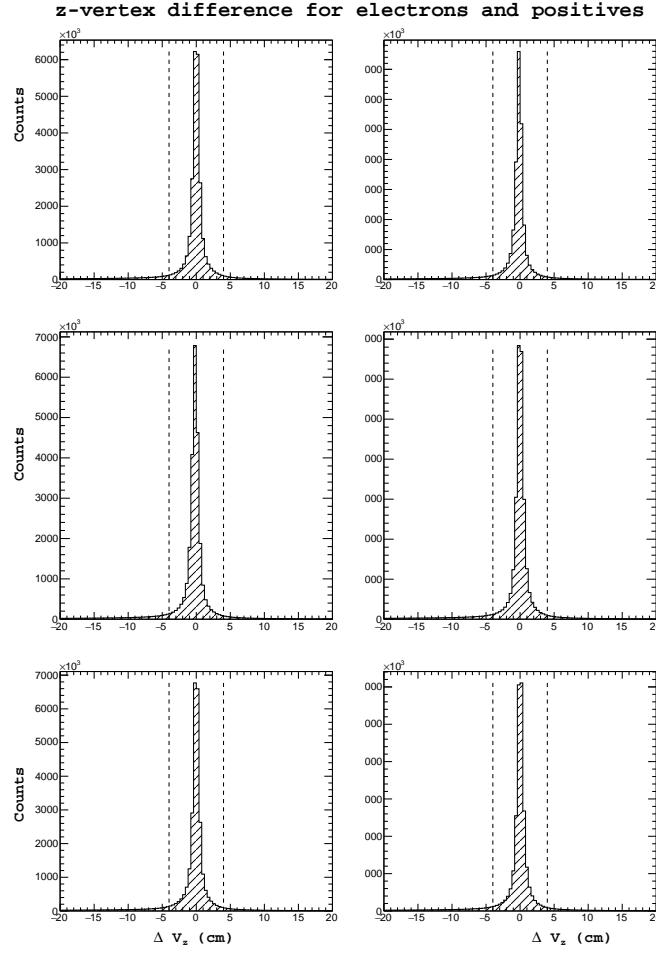


Figure 3.9: Shown above: The difference between the z-vertex position between detected electrons and positive tracks.

### Hadron-electron vertex difference

The distance between the electron vertex and the hadron candidate track vertex is computed ( $\delta v_z = v_z^e - v_z^+$ ). This distance is constrained to be within the length of the target (5 cm) see figure ???. If the analyst desires to look at events where the hadron is produced as the result of a decaying hadron, this cut should be removed.

### Likelihood maximization of $\beta(p, h)$

In this section, positive hadrons are used as an example. The same method is applied to the negative hadrons. For each particle species considered, a normalized probability density function  $P(x; p, h)$  is constructed for each input into the likelihood analysis. Here,  $x$  corresponds to the feature being used to categorize different particles (in our case,  $x$  is the  $\beta$  value measured by CLAS time-of-flight),  $p$  is the particle momentum, and  $h$  is the hadron being hypothesized (eg: in our case the possible values for

positive hadrons are pion, kaon, proton). In general if one uses a set of  $N$  variables  $x = (x_1, x_2, \dots, x_N)$ , the likelihood for a hypothesis  $h$  is defined below.

$$\mathcal{L}_h = \prod_{i=1}^N P_i(x_i; p, h) \quad (3.7)$$

In our case, the only random variable we consider is  $\beta$ , and the likelihood is just the PDF. Here, and in many cases where the choice is statistically appropriate, it is possible to use a Gaussian PDF for the variable  $x_i (\beta)$ .

$$P(\beta; p, h) = \frac{1}{\sqrt{2\pi}\sigma_\beta(p, h)} \exp \left\{ -\frac{1}{2} \left( \frac{\beta - \mu_\beta(p, h)}{\sigma_\beta(p, h)} \right)^2 \right\} \quad (3.8)$$

The identity is assigned by choosing the particle hypothesis  $h$  which maximizes the likelihood ratio.

$$\frac{\mathcal{L}_h}{\mathcal{L}_\pi + \mathcal{L}_K + \mathcal{L}_p} \quad (3.9)$$

Using this method, every positive track is assigned a particle identification. However, at times the likelihood value is quite small when compared with the maximum likelihood for that species. This is the case for positrons which are classified by this method as positive pions, because they are the closest particle for which a hypothesis has been provided. To avoid these situations, the confidence level  $\alpha$  of each track is calculated and a cut is applied on the minimum confidence. This cut can be easily varied to see how it changes the analysis result.

$$\alpha = 1 - \int_{\mu - \beta_{obs}}^{\mu + \beta_{obs}} P(\beta; p, h) d\beta \quad (3.10)$$

This quantity represents the probability to observe a value of  $\beta$  as far from the mean as  $\beta_{obs}$ . Confidence levels of 0 then correspond to tracks which are poorly identified as the class  $h$ . In the case that the PDF is Gaussian, the standard 1, 2, and 3 sigma cuts on  $\beta$  vs.  $p$  can be understood simply as confidence levels of approximately  $0.32 = 1-0.68$ ,  $0.05 = 1-0.95$ , and  $0.01 = 1-0.99$ .

### Determination of probability density functions for likelihood method

The most important and most difficult part of constructing the likelihood ratio identification is the ascertainment of the mean and standard deviation of the probability density function (which depends on momentum) for the different particle hypothesis. In the case where exceptionally accurate monte carlo (MC) simulations of the detector are available, one can use the truth information and track matching to construct the  $\beta$  vs.  $p$  2-dimensional histograms, and fit the  $\mu(p)$  and  $\sigma(p)$ . In the absence of high quality MC, analysts typically fit directly the spectrum of  $\beta$  vs.  $p$  and extract the mean and variance. In this work, the authors chose to create an enhanced sample of candidates for each of the three positive

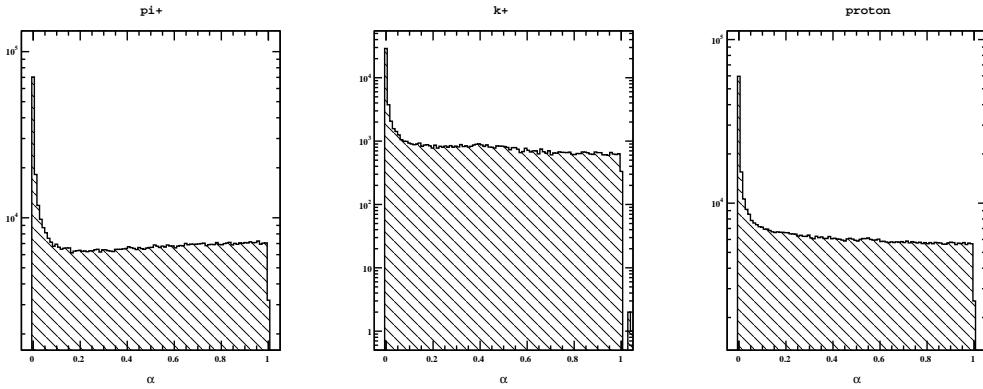


Figure 3.10: Shown above: The distribution of confidence level for all positive tracks after being classified by the likelihood ratio.

particles in question before doing the fitting. In this way, we hope to get a better quality fit of the true mean, and resolutions for the different species. For fitting of pion and proton resolutions, positive tracks are assumed to be pions and the missing mass of the event is calculated. Then, a cut is placed around the neutron mass. In doing so, we are selecting mainly two types of exclusive events. The first is  $ep \rightarrow e\pi^+N$ , and the second is  $ep \rightarrow ep\pi^0$ . In this way most positrons, and positive kaons are removed from the sample prior to fitting. The mean and variance are fit using a third order polynomial in  $p$  (MINUIT  $\chi^2$  minimization is used). The negative tracks  $\pi^-$ ,  $K^-$  are fit directly as is normally done.

The parametrization used for the mean  $\mu(p, h)$  and resolutions  $\sigma(p, h)$  are shown below.

$$\mu(p, h) = \mu_{theory} + \Delta\mu \quad (3.11)$$

$$\mu_{theory} = \frac{1}{\sqrt{1 + (m_h/p)^2}} \quad (3.12)$$

$$\Delta\mu = \mu_0 + \mu_1 p + \mu_2 p^2 \quad (3.13)$$

$$\sigma(p, h) = \sigma_0 + \sigma_1 p + \sigma_2 p^2 \quad (3.14)$$

The values are displayed in the table below.

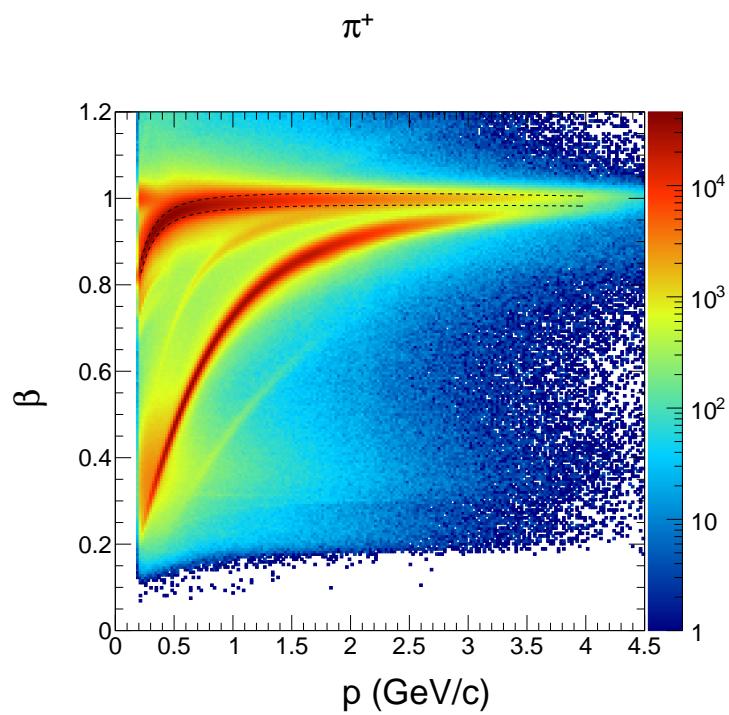


Figure 3.11: Shown above: All positive tracks overlaid with our determination of  $\mu(p) \pm \sigma(p)$  for  $\pi^+$

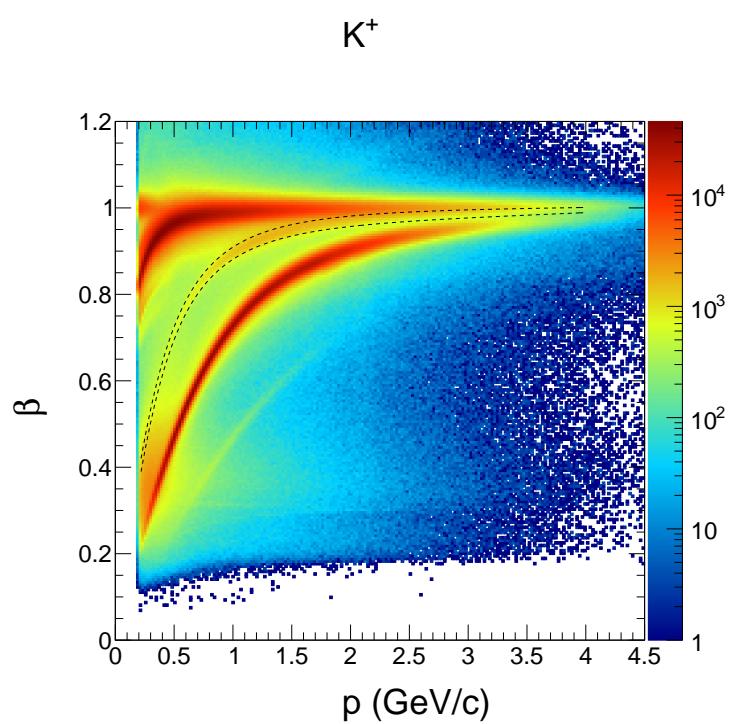


Figure 3.12: Shown above: All positive tracks overlaid with our determination of  $\mu(p) \pm \sigma(p)$  for  $K^+$

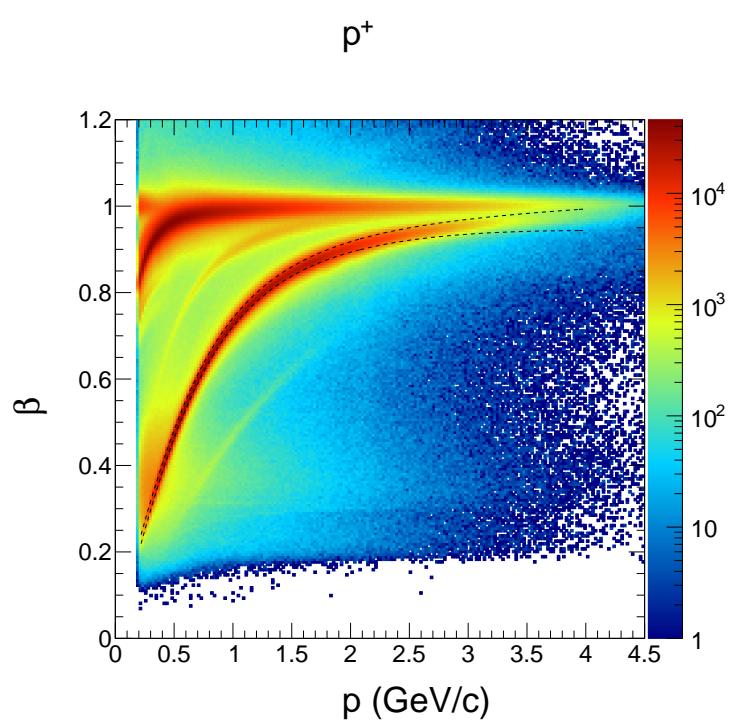


Figure 3.13: Shown above: All positive tracks overlaid with our determination of  $\mu(p) \pm \sigma(p)$  for  $p^+$

Hadron	Parameter	Sector 1	Sector 2	Sector 3	Sector 4	Sector 5	Sector 6
$K^+$	$\mu_2$	0.00111554	-8.97687e-05	4.78796e-05	0.000376425	-0.00204856	0.000652209
$K^+$	$\mu_1$	-0.00468038	6.19414e-05	-0.00081741	-0.00107931	0.00629181	-0.00264143
$K^+$	$\mu_0$	0.00361012	0.00134921	0.00299674	0.00220194	0.000117821	0.00162582
$K^+$	$\sigma_2$	-0.000331838	-0.00105807	-0.000712404	-0.000573934	-0.000259289	0.000508389
$K^+$	$\sigma_1$	-0.00105857	0.00236686	0.000509169	0.000163467	-0.00233617	-0.00461598
$K^+$	$\sigma_0$	0.0154964	0.0117702	0.0140748	0.0143761	0.0184055	0.0180945
$\pi^+$	$\mu_2$	-0.000962041	-0.000300602	-0.000306326	-3.2245e-05	-0.00226511	-0.000330818
$\pi^+$	$\mu_1$	0.00296349	0.0016512	0.0021962	0.00176045	0.00750862	0.00126443
$\pi^+$	$\mu_0$	-0.00225794	-0.00047045	0.000370406	0.000435526	-0.000449409	-0.00131045
$\pi^+$	$\sigma_2$	-0.000127659	0.000691895	-0.000289961	0.000315041	-0.000936521	-0.000131269
$\pi^+$	$\sigma_1$	-0.000489092	-0.0033948	0.00196853	-0.00197841	0.00212778	-0.000339411
$\pi^+$	$\sigma_0$	0.0155195	0.0167998	0.0124066	0.0157476	0.0145571	0.0141728
$p^+$	$\mu_2$	-0.00039358	-0.000701003	-0.000347651	0.0004854	-0.00121666	0.000563786
$p^+$	$\mu_1$	-0.000295423	0.00170899	0.000794901	-0.000744446	0.00376887	-0.00353545
$p^+$	$\mu_0$	0.00227353	0.00231676	0.00364672	0.00276859	0.00128827	0.00439605
$p^+$	$\sigma_2$	0.001429	0.00144256	0.00124456	0.00190709	0.00141039	0.0011516
$p^+$	$\sigma_1$	-0.0021472	-0.00262226	-0.00196308	-0.00385218	-0.00186708	-0.00186749
$p^+$	$\sigma_0$	0.0107541	0.0109091	0.0104381	0.0115449	0.0109969	0.0107759
$\pi^-$	$\mu_2$	3.28823666e-04	-1.30673670e-05	-2.32502052e-04	-9.75619848e-04	-5.89834444e-04	5.27496718e-04
$\pi^-$	$\mu_1$	-3.94924663e-03	-2.66028661e-03	-1.28565631e-03	9.09410075e-04	-2.01610684e-03	-4.42276918e-03
$\pi^-$	$\mu_0$	9.48011169e-04	1.55078786e-03	1.43431985e-03	1.35056935e-03	4.59833580e-03	2.30751866e-03
$\pi^-$	$\sigma_2$	4.37635504e-04	4.38306224e-04	5.32057510e-04	3.36999845e-04	7.74135462e-04	1.36515196e-04
$\pi^-$	$\sigma_1$	-3.28011836e-03	-3.28456104e-03	-3.82847286e-03	-3.11749323e-03	-4.63110728e-03	-2.21229710e-03
$\pi^-$	$\sigma_0$	1.63296567e-02	1.62229164e-02	1.59769911e-02	1.58803427e-02	1.74670064e-02	1.51753145e-02
$K^-$	$\mu_2$	-2.72020947e-03	-5.21081786e-03	-2.13868763e-02	-4.45600034e-03	-7.60703841e-03	-5.27074813e-03
$K^-$	$\mu_1$	1.78610401e-02	2.30787460e-02	9.49357818e-02	1.95764575e-02	3.63245785e-02	2.92417500e-02
$K^-$	$\mu_0$	-2.26190100e-02	-2.22562379e-02	-1.02704771e-01	-2.25931014e-02	-5.10484618e-02	-3.19918187e-02
$K^-$	$\sigma_2$	1.76905114e-02	1.62989708e-02	3.60928130e-02	1.51270521e-02	1.91308107e-02	2.38470033e-02
$K^-$	$\sigma_1$	-7.74901862e-02	-7.33041628e-02	-1.57454534e-01	-7.26870393e-02	-9.23654247e-02	-1.02397836e-01
$K^-$	$\sigma_0$	1.07082820e-01	1.00573410e-01	1.93148260e-01	1.00993689e-01	1.26963814e-01	1.30057621e-01

Table 3.4: Values used to calculate the mean and resolutions for hadron likelihood based identification.

# Chapter 4

## Beam Spin Asymmetry Analysis

### 4.1 Introduction

Measurement of the beam spin asymmetry is carried out for the positively charged k-meson. As discussed in the introduction, the beam spin asymmetry theoretically depends on  $F_{UU,L}$ ,  $F_{UU,T}$ ,  $F_{UU}^{\cos\phi}$ ,  $F_{UU}^{\cos 2\phi}$ , and  $F_{LU}^{\sin\phi}$ . By dividing the electron-kaon events into several bins of SIDIS kinematic variables, beam spin asymmetry measurements can be taken at different average values of the kinematic variables. Finally, the structure function ratios  $A_{LU}^{\sin\phi}$ ,  $A_{UU}^{\cos\phi}$ , and  $A_{UU}^{\cos 2\phi}$  can be extracted from each bin. In this chapter, we discuss the selection of SIDIS events, the binning used in this analysis, our measurement with associated systematic uncertainties, and the extraction of structure function ratios using the  $\phi$  dependence in each kinematic bin.

### 4.2 Event Selection and Binning

#### Event Selection

After particle identification is performed on the event, those events which have a trigger electron and a positive kaon are kept for analysis. We discard events that do not have  $W > 2$  and  $Q^2 > 1$ , because they are not considered to be part of the deeply inelastic region. Additionally, to avoid exclusive resonances in the  $ep \rightarrow eK^+X$  spectrum, the authors impose a cut on the missing mass of the final state  $X$ . For this analysis we use  $M_X(ep \rightarrow eK^+X) > 1.25$ . Finally, we attempt to perform our measurement in the current fragmentation region where factorization has been proven. This is done by excluding events with  $z_h < 0.25$ . We also require that  $z_h < 0.75$  to avoid exclusive events. This restriction on  $z_h$  is not applied to the  $z_h$  axis, where we measure across the entire experimentally observed range. After these selection criteria have been applied, the data is sorted into kinematic bin.

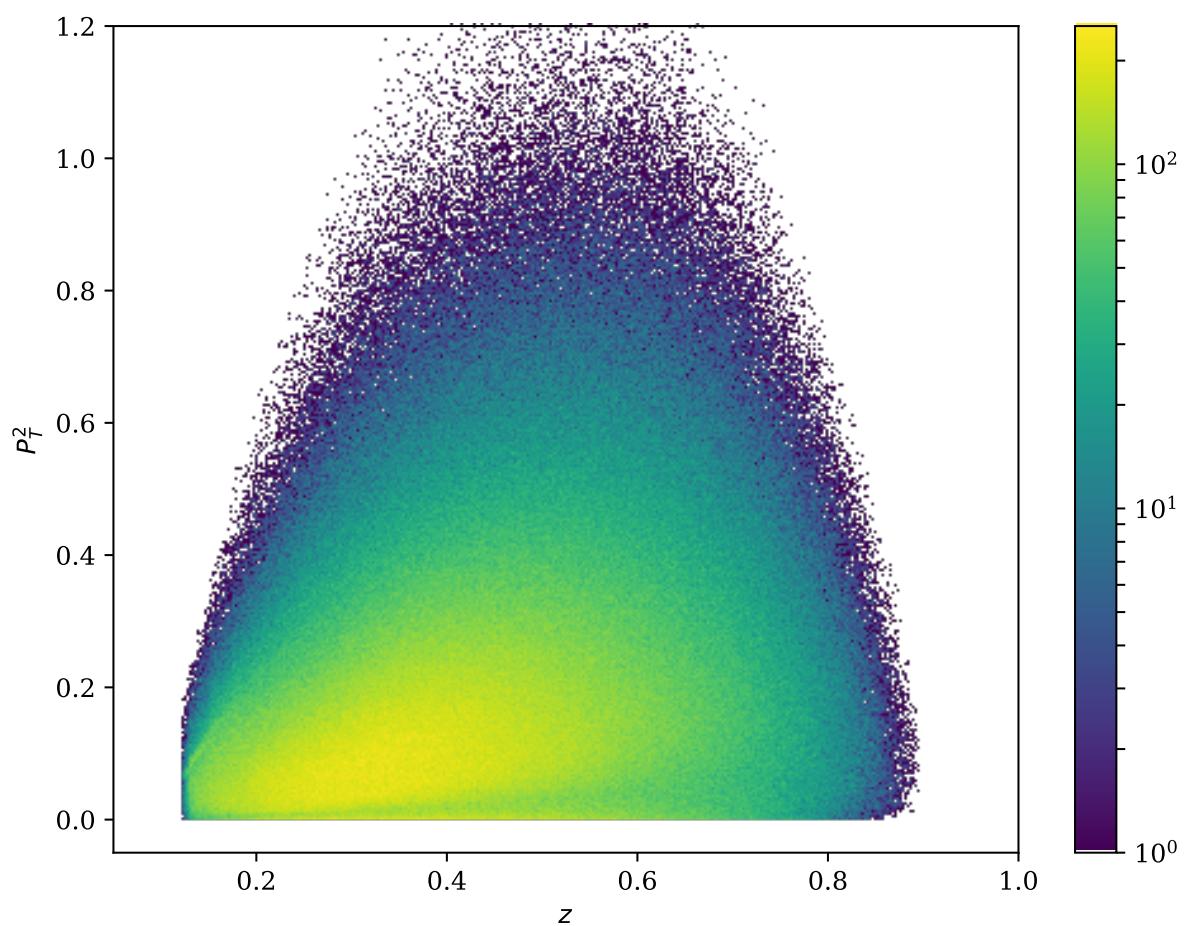


Figure 4.1: Correlation between  $z_h$  and  $P_T^2$  for each event in our analysis sample.

## Binning

For this study, the authors chose to measure the integrated beam spin asymmetry. This simply means that for a given axis ( $P_T$  for example), the events included have all observed values of the other kinematic variables (in this example  $x, z_h, Q^2$ ). The axes studied are  $x, Q^2, z_h$ , and  $P_T$ . We chose to use 12 bins in  $\phi$  and 10 bins of the other kinematic variables for a total of 120 analysis bins.

The bins were chosen using a simple method to ensure equal statistics in each bin. The procedure will be described using the axis  $x$  as an example. First, all events are sorted by their  $x$  value from smallest to largest. Then, the smallest and largest values are recorded, which are just  $x_1$  and  $x_N$  if there are  $N$  events in the sample. Next, the target number of bins  $M$  is chosen (this choice is done by the analyst based on what he/she believes to be the best choice). Finally, the limits of each bin can be chosen simply by calculating the number of events per bin  $N/M$  and then using the value of  $x$  which corresponds to multiples of  $N/M$  in the sample.

$$\vec{b} = (x_1, x_{N/M}, x_{2N/M}, \dots, x_N) \quad (4.1)$$

Here, the symbol  $\vec{b}$  denotes a vector of  $(M+1)$   $x$  values which represent bin limits. The binning in  $\phi$  is chosen to be regularly spaced between -180 and 180 degrees.

## 4.3 $\phi_h$ Distributions

### Measured Asymmetry Values

In each bin  $i$  the beam spin asymmetry (here  $A_i$ ) is calculated according to,

$$A_i = \frac{1}{P_e} \frac{n_i^+ - n_i^-}{n_i^+ + n_i^-} \quad (4.2)$$

where  $P_e$  is the average beam polarization over the dataset. The symbols  $n_i^\pm$  refer to the number of events counted in bin  $i$  with helicity  $\pm$ .

### Statistical Uncertainties

The uncertainty on the measured value of  $A_i$  can be attributed to statistical uncertainty on the counts  $n_i^\pm$ , and the uncertainty associated with the measurement of  $P_e$ . The treatment of the statistical uncertainty reported on the measurement includes the contribution from counts, but not from the uncertainty in  $P_e$  which is included in the systematic errors. The uncertainty in a measured observable  $\mathcal{O}$  depends on the uncertainty of the parameters used to construct it  $\vec{\theta}$  in the following way (see appendix).

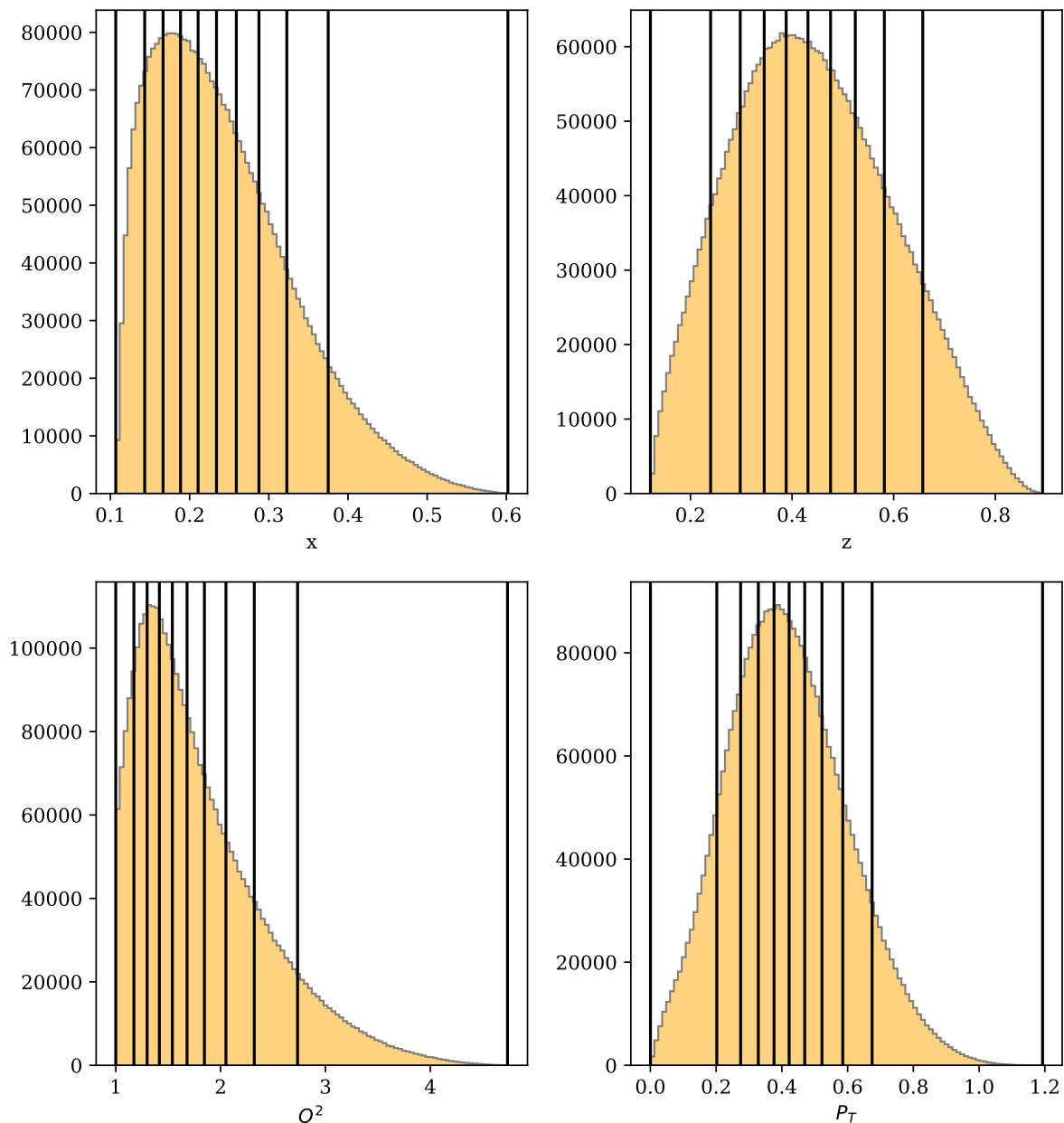


Figure 4.2: The binning used for each of the kinematic axes.

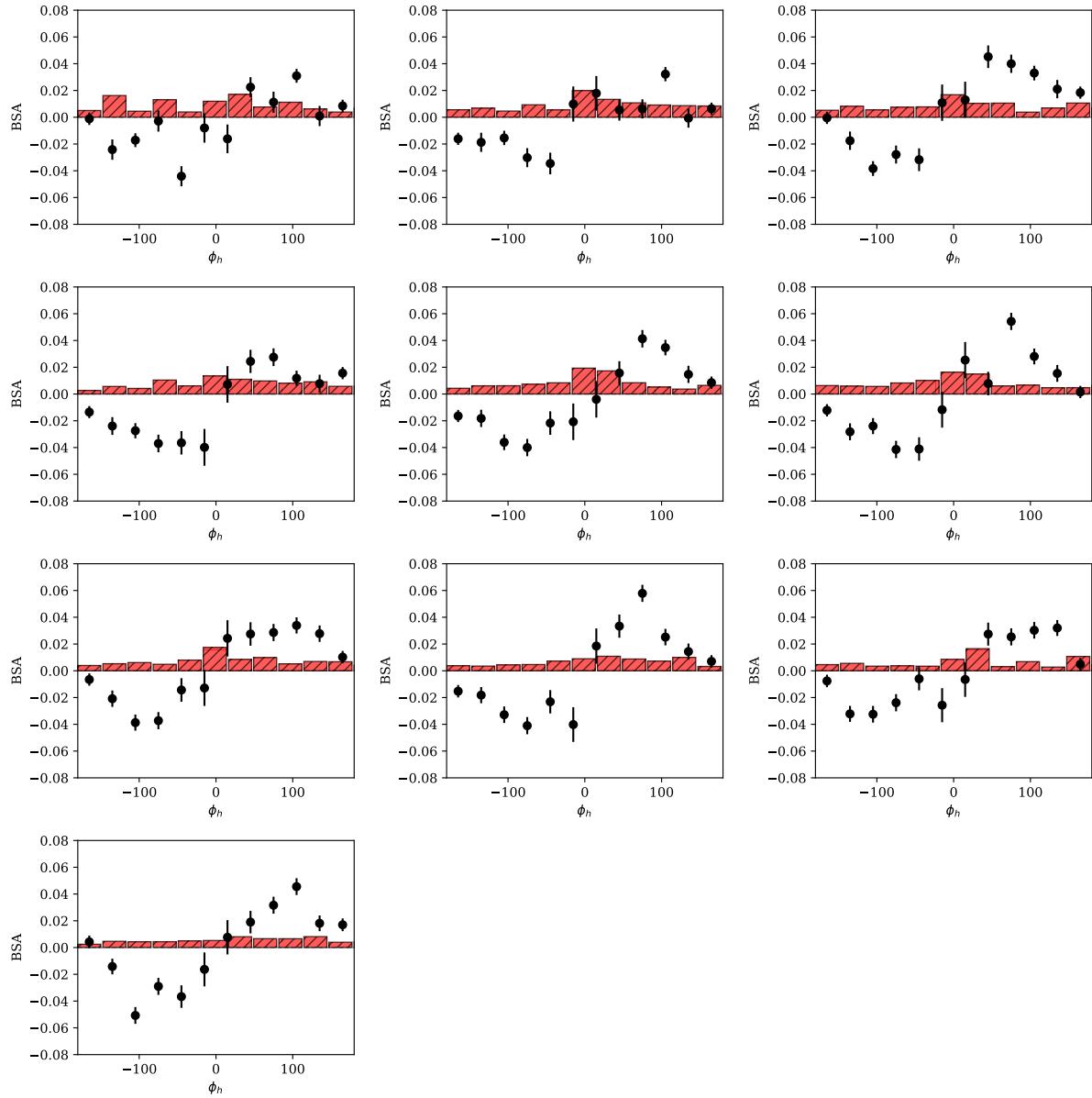


Figure 4.3: The  $\phi_h$  dependence is shown for each bin of  $x$ , increasing in value from the top left to the bottom right. The statistical uncertainty is shown as black error bars on each point. The total systematic uncertainty is shown as a red bar centered at zero.

$$\sigma_{\mathcal{O}}^2 = \sum_{i=1}^N \sum_{j=1}^N \frac{\partial \mathcal{O}}{\partial \theta_i} \frac{\partial \mathcal{O}}{\partial \theta_j} \rho_{ij} \sigma_i \sigma_j \quad (4.3)$$

For the beam spin asymmetry in the  $i^{th}$  bin  $A_i$  one finds that without correlations ( $\rho_{ij} = \delta_{ij}$ ) the error propagation proceeds as shown below.

$$\sigma_A^2 = \frac{A^2}{P_e^2} \sigma_{P_e}^2 + \frac{4(n_-^2 \sigma_+^2 + n_+^2 \sigma_-^2)}{P_e^2 (n_+ + n_-)^4} \quad (4.4)$$

The first term which is the contribution from the variance in the measurements of beam polarization will be included as a systematic error. The second term is used as the statistical error bars shown through the analysis. The counts  $n_{\pm}^i$  for the  $i^{th}$  bin are assumed to be Poisson in nature, and therefore have a variance equal to the expected number of counts  $\sigma_{\pm}^2 = n_{\pm}^i$ . With this expression for the statistical uncertainty on the counts, and dropping the beam polarization term for now, the expression becomes:

$$\sigma_A^2 = \frac{4n_+ n_-}{P_e^2 (n_+ + n_-)^3} \quad (4.5)$$

## Systematic Uncertainties

### Formalism

Systematic effects are shifts or biases in the measured result of some observable as a result of the procedure used in the measurement. Systematic effects can typically be identified and corrected for, or removed all together from the measurement. In the cases where an effect cannot be completely removed, the degree to which the correction for the effect is uncertain is included in the result of the measurement as a systematic uncertainty.

Sources of systematic effects can include background events from different processes which enter the sample, calibrations of different detector systems, misalignments in detector geometry, and biases in selection criteria. Each of the systematic sources mentioned here has at least one associated procedure for correcting it's effect on the analysis. As an example consider momentum corrections in CLAS. These corrections are performed to remove the effect of slight mis-alignments in detector geometry from what is in reconstruction, as well as slight differences between the true magnetic field and the field map used in reconstruction. These physical effects introduce a systematic effect, the particle 4-momenta reconstructed are shifted away from the true values. Standard reactions (elastic scattering) can be used to develop corrections for the 4-momenta of particles, and these corrections typically depend on a set of parameters  $\vec{\theta}$ , which have an associated parameter uncertainty described by a covariance matrix  $V_{ij}$ . It is these parameter uncertainties that propagate through to the final observables, and the assignment of

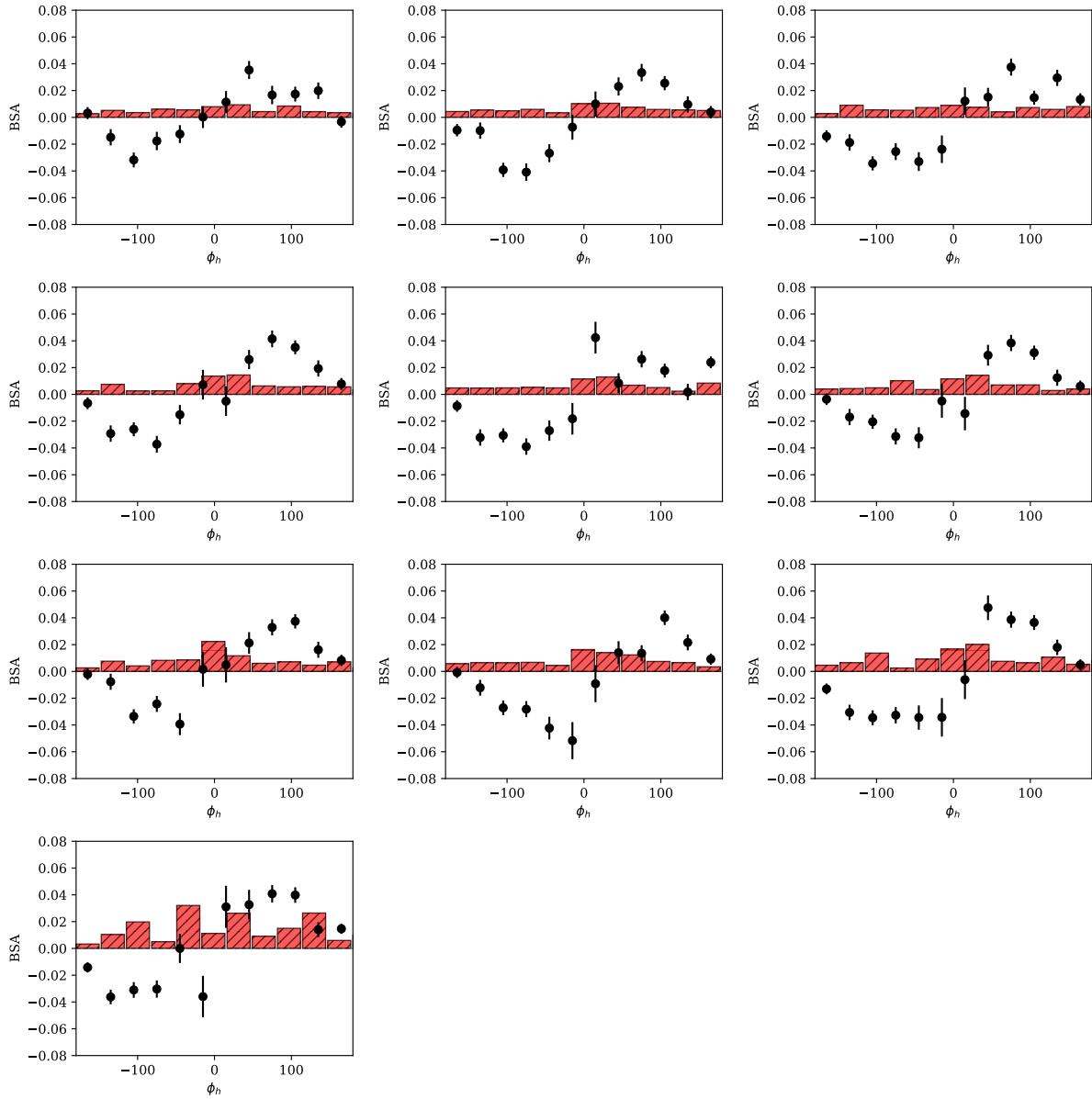


Figure 4.4: The  $\phi_h$  dependence is shown for each bin of  $z_h$ , increasing in value from the top left to the bottom right. The statistical uncertainty is shown as black error bars on each point. The total systematic uncertainty is shown as a red bar centered at zero.

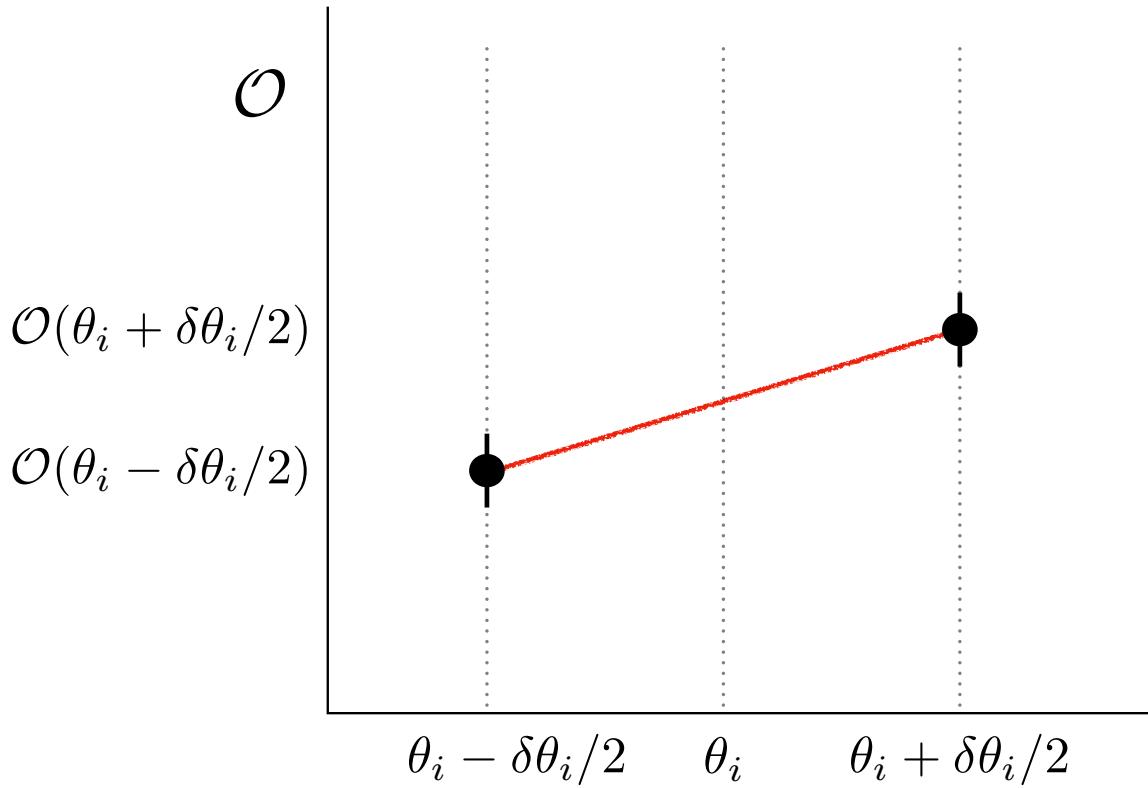


Figure 4.5: The analysis is run for variations in the input parameters  $\theta_i$  to calculate the dependence of the result  $\mathcal{O}$  on each parameter, as described in this section.

the magnitude of such effects is then what is referred to as systematic uncertainty.

Systematic uncertainties can be included using the standard equation for error propagation. In some cases it is possible to analytically find the derivatives needed to calculate the dependence of the observable on a source of systematic uncertainty. This is the case for effect of the variance of the beam polarization on the beam spin asymmetry observable. However in many cases, it is not possible to analytically calculate the effect of some analysis parameter  $\theta_i$  on the observable  $\mathcal{O}$ . Since the observable is usually calculated using some computational chain which starts with the input parameters  $\vec{\theta}$ , it is possible to find the dependence of the observable  $\mathcal{O}$  on the inputs numerically.

$$\frac{\partial \mathcal{O}}{\partial \theta_i} \approx \frac{\mathcal{O}(\theta_i + \sigma_{\theta_i}/2) - \mathcal{O}(\theta_i - \sigma_{\theta_i}/2)}{\sigma_{\theta_i}} \quad (4.6)$$

After inserting the above into equation ?? one finds,

$$\sigma_{\mathcal{O}}^2 = \sum_{i=1}^n \sum_{j=1}^n \rho_{ij} (\mathcal{O}(\theta_i + \sigma_{\theta_i}/2) - \mathcal{O}(\theta_i - \sigma_{\theta_i}/2)) (\mathcal{O}(\theta_j + \sigma_{\theta_j}/2) - \mathcal{O}(\theta_j - \sigma_{\theta_j}/2)) \quad (4.7)$$

Source	Variation	Magnitude
Beam polarization	0.024	0.000672
DC Region 1 Fid.	1 (cm)	0.001344
DC Region 3 Fid.	3 (cm)	0.001821
EC-W	12 (cm)	0.000948
EC-V	12 (cm)	0.000797
EC-U	12 (cm)	0.002487
Kaon Confidence ( $\alpha$ )	0.01-0.07	0.001827
$\theta_{cc}$ Matching	$\sigma$	0.001152
EC Energy Deposition	0.01 (GeV)	0.001644
$p_{K^+}$	$2.5-E_{beam}$	0.002360
EC Sampling Fraction	$0.5\sigma$	0.001240
Z-Vertex	0.5 (cm)	0.002581
Statistical	-	0.007494
MC Estimate	-	0.002917

Table 4.1: Different sources of systematic effect considered in this analysis. The magnitude of the effect is shown here averaged over all bins. The units of the shift are just the same units of the asymmetry value.

where  $\rho_{ij}$  is the correlation  $V_{ij}/\sigma_i\sigma_j$ . In most cases, these correlations are assumed to be zero. In some cases, when the parameters  $\theta_i$ ,  $\theta_j$  come from a fit one may have a correlation provided by the covariance matrix and it should be used. In the case where correlations are assumed to be zero, the total systematic uncertainty is simply the quadrature sum of the shift in the observable within the uncertainty window on each parameter.

$$\sigma_{\mathcal{O}}^2 = \sum_{i=1}^n \left[ \mathcal{O}(\theta_i + \sigma_{\theta_i}/2) - \mathcal{O}(\theta_i - \sigma_{\theta_i}/2) \right]^2 \quad (4.8)$$

Another approach exists that takes into account possible correlations between the analysis parameters  $\theta_i$ . This approach has not yet been widely used, and probably requires a thorough understanding of systematics using the previously described method before its application. The approach consists of generating monte carlo  $M$  sets of parameters  $\vec{\theta}$  and obtaining  $M$  results for the observable  $\mathcal{O}$ . The results are then interpreted probabilistically and the observable value and total systematic error are reported as the mean and standard deviation of the results.

$$\langle \mathcal{O} \rangle = \frac{1}{M} \sum_{i=1}^M \mathcal{O}_i \quad (4.9)$$

$$\sigma_{\mathcal{O}}^2 = \frac{1}{M-1} \sum_{i=1}^M (\mathcal{O}_i - \langle \mathcal{O} \rangle)^2 \quad (4.10)$$

## Sources of Systematic Uncertainty

The table ?? below summarizes the sources of systematic effects considered in this analysis.

Except for the beam polarization and the momentum of the kaon track, all parameters listed in the table are treated using the formalism outlined above. The beam polarization uncertainty quoted at 2.4%

contains contributions from the standard deviation of the Moller polarimetry measurements (0.2%), residual target polarization effects (1.4%), and atomic motion/finite acceptance corrections (0.8%).

Because of the inability to distinguish kaon, pion, and proton tracks at higher momentum, the maximum kaon track momentum is varied between 5.5 (no maximum) and 2.5 GeV. The difference between these results is quoted as a systematic uncertainty and added in quadrature with the other sources. This source of systematic uncertainty has a larger effect on the  $z_h$ -axis, because  $z_{max}$  is limited by limiting  $p_{max}$ . While for the large  $z_h$ -bins this contribution is dominant, its size is comparable with other systematic sources throughout the remainder of the bins.

## 4.4 Extraction of Modulations

The motivation to measure the beam spin asymmetry in several kinematic bins as well as bins of  $\phi_h$  is to perform an estimate of the value of structure functions at the kinematic points (or the average value of the structure functions over the range of values included in a point). To do this, the authors perform parameter estimation on the  $\phi_h$  distributions taking as a model the theoretical dependence of the beam spin asymmetry on  $\phi_h$ .

$$f(\phi_h, \vec{a}) = \frac{a_0 \sin \phi_h}{1 + a_1 \cos \phi_h + a_2 \cos(2\phi_h)} \quad (4.11)$$

The parameters  $\vec{a}$  are the structure function ratios we wish to extract. The simplest way to extract these parameters is to use a standard fitting package like `Minuit` or `scipy.optimize.minimize`. In these approaches,  $\chi^2$  is defined as the square difference between the observed data values and those predicted by the model, normalized by the error. If the fluctuation between the data and theory predictions is on the order of the error, the  $\chi^2$  is simply on the order of the number of data points. The parameters  $\vec{a}$  which best describe the data are those which make the  $\chi^2(\vec{a})$  assume it's minimum value. This minimization is done in practice with gradient descent or quasi-Newton's method based algorithms, and we will not discuss details of these here. It is sufficient to say that these methods produce the parameters  $\vec{a}$ , and an estimate of the covariance matrix  $V$ . These parameters and their errors become the extracted value and uncertainty on the extracted values of the structure function ratio in each bin.

Unfortunately, applying the standard single-fit procedure described above does not always produce stable results. In some cases, the resulting parameter sets are reasonable, in other cases however the parameters in the denominator become unphysically large and oppose each other. This effect has motivated previous analysts to search for other means of extracting the dominant  $\sin \phi_h$  behaviour from the distributions. One common technique is to assume that  $a_1$  and  $a_2$  of above are small compared to

1. The analyst can then fit the  $\phi_h$  distribution with just one linear parameter  $a_0$ . This produces a stable result, but has the disadvantage that one needs to introduce a systematic uncertainty associated with the difference observed between using the full model (with a restricted range for the parameters in the denominator) and the results obtained using the single parameter model. In order to avoid this, the authors choose to use a Monte Carlo method of replicas. The replica method consists of generating  $N_{rep}$  psuedo-data  $\phi_h$  distributions which have a normal distribution located at the observed value, and with a variance equal to the statistical errors on the associated data point.

$$\vec{A}_{rep} = \mathcal{N}(\vec{A}, \sigma_A) \quad (4.12)$$

Where here  $\vec{A}$  is a vector of length  $n_{phi}$  bins, representing the measured beam spin asymmetry for each value of  $\phi_h$  in a given kinematic bin. Each of these distributions is fit with the full model, and the resulting parameter values are saved. The final reported value for each fit parameter, as well as its uncertainty can be reported as the mean, and standard deviation of the fit results. This procedure which is similar to bootstrapping, can be seen as an attempt to fit the underlying distribution that generated the data while avoiding the statistical noise. This technique has been discussed in [replica-fitting].

$$\langle a_j \rangle = \sum_{i=1}^{N_{rep}} a_j^{(i)} \quad (4.13)$$

$$\sigma_{a_j}^2 = \frac{1}{N_{rep} - 1} \sum_{i=1}^{N_{rep}} (a_j^{(i)} - \langle a_j \rangle)^2 \quad (4.14)$$

## Results

We observe that the term  $A_{LU}^{\sin\phi}$  is weakly dependent on the kinematic variables. The value is around +3% for most bins.

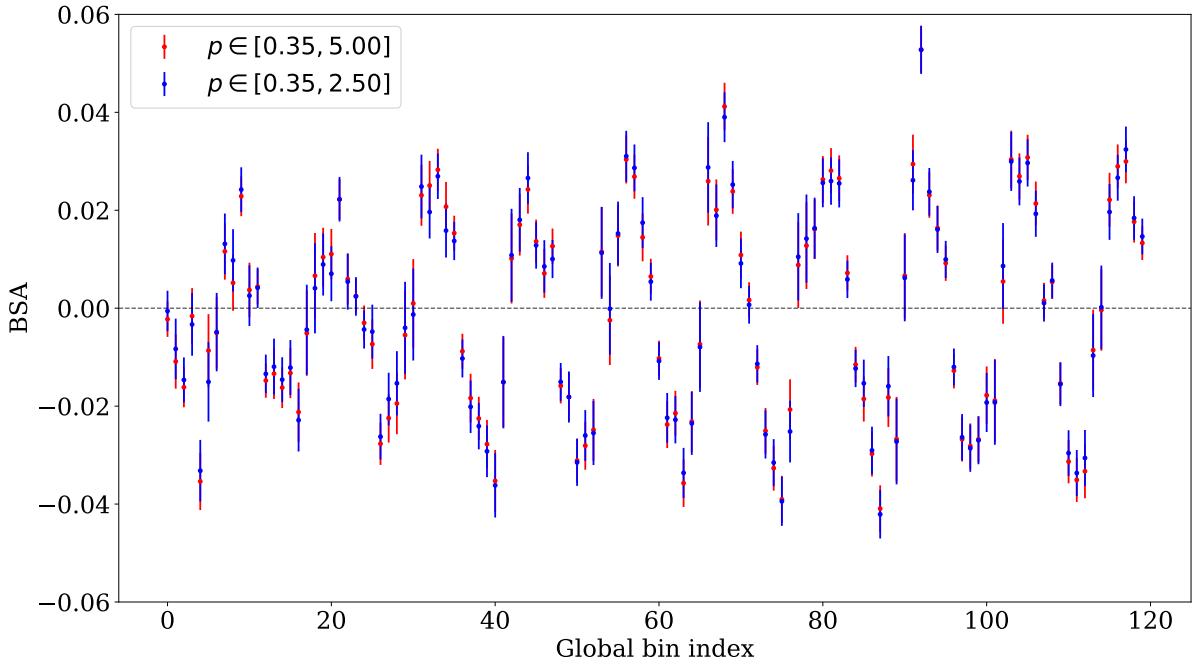


Figure 4.6: The BSA for each bin of  $x, \phi_h$  (all plotted together on the x-axis) is compared with and without a maximum momentum (no tracks exceed 5 GeV) for the kaon track. The global bin coordinate on the x-axis is  $i + n_\phi * j$  where  $i$  is the  $\phi_h$  bin,  $j$  is the  $x$  bin, and  $n_\phi$  is the number of  $\phi$  bins (12). Both  $i$  and  $j$  start at 0.

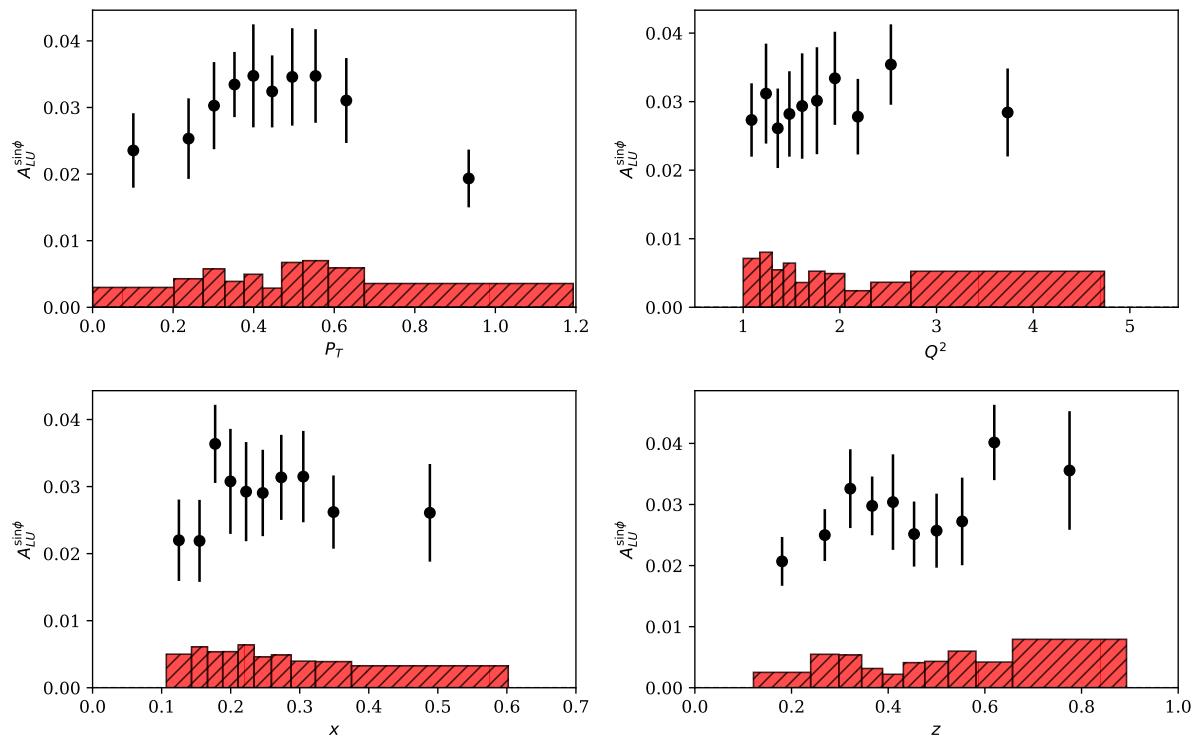


Figure 4.7: Our extraction of  $A_{LU}^{\sin\phi}$  for the kinematic bins described above. The black error bars represent uncertainty in the extraction of the parameter value. Red error bars are systematic uncertainties.

# Chapter 5

## SIDIS Cross Section

This chapter discusses the analysis of inclusive and semi-inclusive deeply inelastic events. First, the important work on  $\pi^\pm$  SIDIS that was done before this analysis is described. Then, motivation is provided to measure the inclusive cross section. Our analysis of inclusive scattering in the resonance region is then described in detail. Finally, the SIDIS cross section is calculated and the results are discussed.

### 5.1 Introduction

The primary goal of this work is to provide the SIDIS cross section for charged pions ( $\pi^\pm$ ) over a large kinematic range ( $0.1 < x < 0.6$ ,  $1.0 < Q^2 < 4.7$ ,  $0.0 < z < 0.9$ ,  $0.0 < P_T^2 < 1.0$ ,  $-180^\circ < \phi_h < 180^\circ$ ). Fortunately these channels were studied by Nathan Harrison and the CLAS collaboration [analysis-harrison] using the E1-F dataset (2010-2015). By writing the cross section as,

$$\frac{d^5\sigma}{dx dQ^2 dz dP_T^2 d\phi_h} = A_0 \left[ 1 + A_{UU}^{\cos \phi_h} \cos \phi_h + A_{UU}^{\cos(2\phi_h)} \cos(2\phi_h) \right] \quad (5.1)$$

Harrison et al. measured the un-normalized quantities  $A_0$ ,  $A_{UU}^{\cos \phi}$ , and  $A_{UU}^{\cos(2\phi)}$  which are defined below.

$$A_0 = \frac{\pi \alpha^2 y (1 + \gamma^2/2x)}{2EM_p x^2 Q^2 (1 - \varepsilon)} (F_{UU,T} + \varepsilon F_{UU,L}) \quad (5.2)$$

$$A_{UU}^{\cos \phi_h} = \sqrt{2\varepsilon(1 + \varepsilon)} \frac{F_{UU}^{\cos \phi_h}}{F_{UU,T} + \varepsilon F_{UU,L}} \quad (5.3)$$

$$A_{UU}^{\cos(2\phi_h)} = \varepsilon \frac{F_{UU}^{\cos(2\phi_h)}}{F_{UU,T} + \varepsilon F_{UU,L}} \quad (5.4)$$

In order to measure the structure functions  $F_{UU}^{\cos \phi_h}$  and  $F_{UU}^{\cos(2\phi_h)}$  directly, the integrated luminosity is needed. The calculation of this quantity for E1-F is described in detail during chapter 2 of this

document. Experimentally, the cross section in the  $i^{th}$  bin is given as,

$$\frac{d\sigma}{dx dQ^2 dz dP_T^2 d\phi_h} = \frac{1}{\Delta x \Delta Q^2 \Delta z \Delta P_T^2 \Delta \phi_h} \frac{N_{obs}^{(i)}}{\mathcal{L} A^{(i)} R^{(i)}} \quad (5.5)$$

where the superscript  $(i)$  reminds the reader that these quantities are calculated for every bin. Throughout this chapter the symbols  $A^{(i)}$ ,  $R^{(i)}$ , and  $B^{(i)}$  refer to the acceptance correction, radiative correction, and bin centering correction respectively. These factors will be described in more details in this chapter. Finally, the  $\Delta$  factors here denote the width of each bin in 1 dimension of the 5-dimensional space (non-uniform sized bins may be used, in which case this factor also carries an index  $(i)$ ).

The integrated luminosity obtained in chapter 2 can be directly applied to the measurement of Harrison et al. to produce 5-dimensional differential cross sections. This procedure is carried out, but only after the luminosity factor is independently verified by calculating the cross section for inclusive inelastic electron scattering in the resonance region (here  $1.1 < W < 2.1 \text{ GeV}/c^2$ ). Accurate models for the inclusive cross section exist based on phenomenological fits to existing datasets. For verification, we compare the cross section from E1-F to a model created by Cynthia Keppel [**physics-keppel**].

The calculation of the inelastic scattering cross section as described here is non-trivial, and (together with the phenomenological analysis presented in chapter 6) constitutes the main original effort exerted by the authors.

## 5.2 Inclusive Cross Section

Inclusive electron scattering is the process  $ep \rightarrow eX$ , where only the final state electron is detected and the rest of the event is not (anything apart from the electron that is detected is not analyzed). As a function of  $W$  (the invariant mass of the final state ( $\gamma^* + p$ ) system) the region below 2 GeV contains resonances and is often referred to as the resonance region. Resonance structures are difficult to detect higher than about 2 GeV, and this region is typically called the *deeply inelastic* region. While the deeply inelastic region is used extensively for measurements in nuclear/particle physics, the goal of luminosity verification is more easily achieved in the resonance region. This fact is due principally to the excess of Bethe-Heitler events which collect in the  $2 < W < 3$  region for  $E_{beam} = 5.498$  (such events are difficult to remove when detecting only the final state electron).

It is true that the elastic scattering cross section for  $E_{beam} = 5.498$  is small compared to the inclusive cross section, but a significant number of electrons radiate photons before colliding with the target with  $E_i < E_{beam}$ . These lower energy electrons then have a significantly higher probability to scatter elastically and for our beam energy collect in the region of  $2 < W < 3$ .

Variable	N	Min.	Max	Width
$W$	35	1.1	2.1	0.286
$Q^2$	10	1.7	4.2	0.25

Table 5.1: Summary of  $W$  and  $Q^2$  binning used for the inclusive cross section.

In this section, our procedure for extracting the inclusive cross section and limiting the impact of background events is discussed.

### 5.2.1 Event Selection and Binning

A simple choice of 10 bins in  $Q^2$  and 35 bins in  $W$  is used. This choice is mainly driven by the desire to keep bin migration effects small. Events are generated and reconstructed in some bins  $R^{(j)}$  and  $G^{(i)}$  respectively. Due to finite detector resolution, it is not always the case that  $i = j$ . This effect is known as bin migration, and negatively impacts the acceptance calculation.

The only kinematic restriction that is imposed is applied to the *inelasticity*  $y = 1 - E'/E < 0.7$ . This restriction is applied because events with large- $y$  have a significantly higher probability to be Bethe-Heitler events. This cut is equivalent to enforcing a minimum energy for the scattered electron.

$$E_{min} = E_{beam}(1 - y_{max}) \approx 1.6 \text{ GeV} \quad (5.6)$$

### 5.2.2 Simulation

All processes that CLAS measures are observed through the combination of signals from several sub-detectors. During analysis all sub-systems are calibrated accurately, but such a complicated system still often produces distributions that do not look like the true physical distribution. This discrepancy arises from the combination of several effects.

1. Holes, barriers, obstructions, shadows of other detectors, and any other physical effects that prevent events from being measured in some range of  $\theta, \phi$  are known as geometrical acceptance effects. An important geometrical acceptance effect is the presence of the torus coils in between every sector. These represent a complete loss of information for a small range of  $\phi$  between each sector.
2. Inefficiencies due to the probabilistic nature of particle interaction in the detector subsystems also lower the overall acceptance.
3. Detectors have finite spatio-temporal resolutions.

In order to understand and limit the impact of these effects on the physics extracted from the experiment, a mock experiment is simulated. In the simulated experiment everything is modelled as

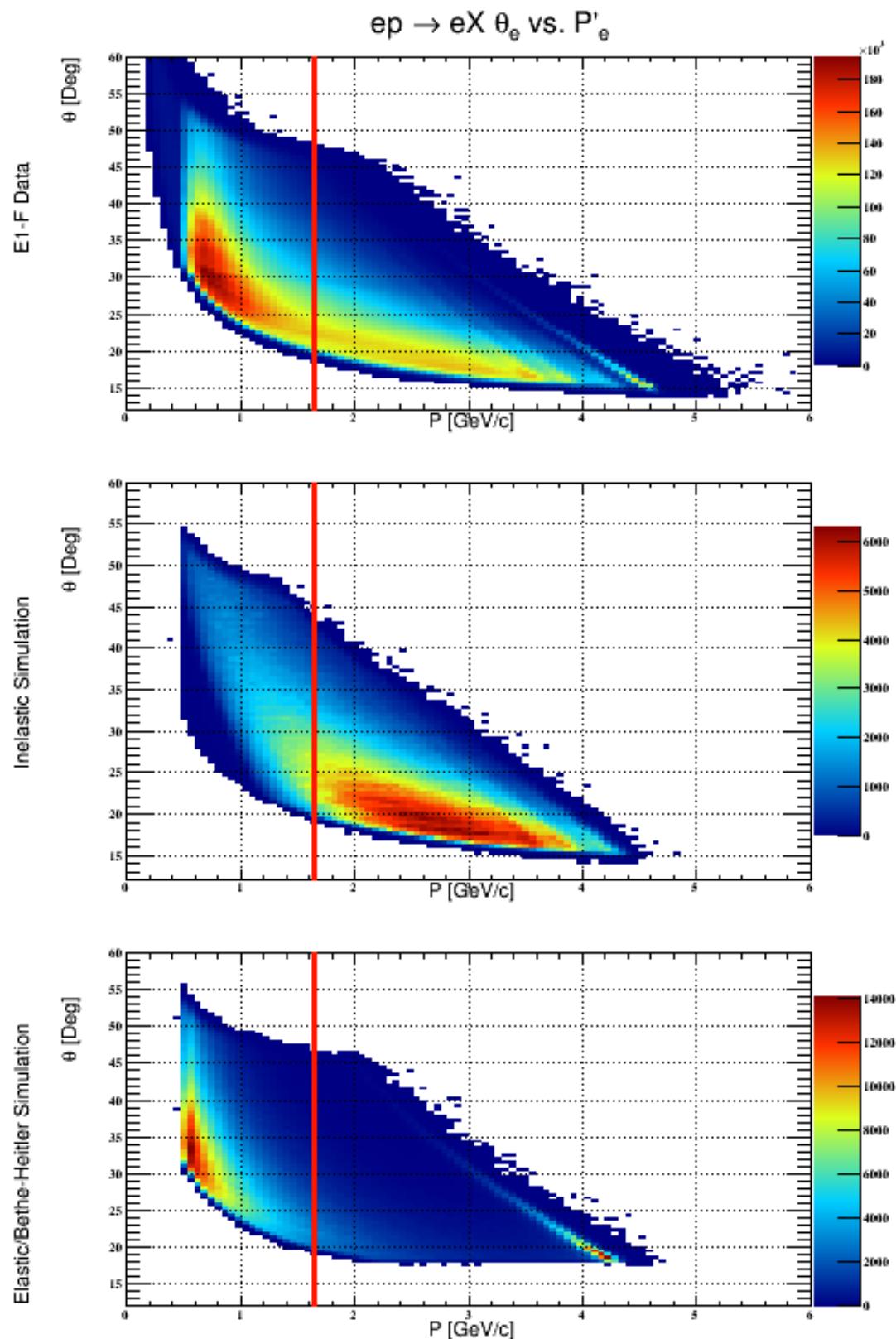


Figure 5.1: Event distributions ( $\theta_e$  vs  $p$ ) for data, simulated inelastic events, and simulated elastic events with radiation are shown. The red line indicates the momentum cut applied by restricting  $y < 0.7$ .

realistically as possible. The simulation used for CLAS is called GSIM and is based on the CERN package GEANT3 (GEometry ANd Tracking).

In this controlled environment, control samples of events can be generated and fed into the simulation.

The output of GSIM is a bos file that is similar to the raw data from the data aquisition system. This is then reconstructed using the same reconstruction algorithm that is applied to data (`userana`).

By retaining the truth information for all particles that are generated, the effect of the detector can be studied completely. These concepts can be stated more formally by considering the true  $t(x')$  and measured  $m(x)$  distributions of some observable. In the absence of background processes, the relationship between these distributions is expressed as a Fredholm integral equation of the first kind.

$$m(x) = \int_{\Omega} K(x, x') t(x') dx' \quad (5.7)$$

Here  $K(x, x')$  is a kernel which encodes information about detector acceptance due to the effects described above. The goal of the Monte Carlo simulation is then to *unfold* the measured distribution  $m(x)$  by providing an estimate of  $K(x, x')$  and finally corrected the data to get  $t(x)$ .

Observed events are usually aggregated into bins and the problem is naturally discretized and written in vector-matrix form.

$$\mathbf{A}\mathbf{x} = \mathbf{y} \quad (5.8)$$

In this notation  $\mathbf{A}$  represents the responce matrix, a discretized version of the kernel function  $K$ , the vector  $\mathbf{y}$  represents the measured distribution in the bins, and the vector  $\mathbf{x}$  is the true distribution over the bins. The matrix elements  $A_{ij}$  can be estimated by using generating events, passing them through a monte-carlo detector simulation, and then counting the number of events that are reconstructed in bin  $i$  when generated in bin  $j$ . This quantity is then normalized by the total number of events generated in the  $j^{th}$  bin. In the absence of bin migration and with perfect acceptance this matrix is the identity matrix  $I^n$  where  $n$  is the total number of bins.

$$A_{ij} = \frac{n_{rec=i,gen=j}}{n_{gen=j}} \quad (5.9)$$

The binned true distribution can be recovered by inverting the responce matrix and correcting the observed distribution.

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{y} \quad (5.10)$$

In the absence of bin migration, the matrix becomes diagonal with efficiency elements  $\epsilon_i$  that repre-

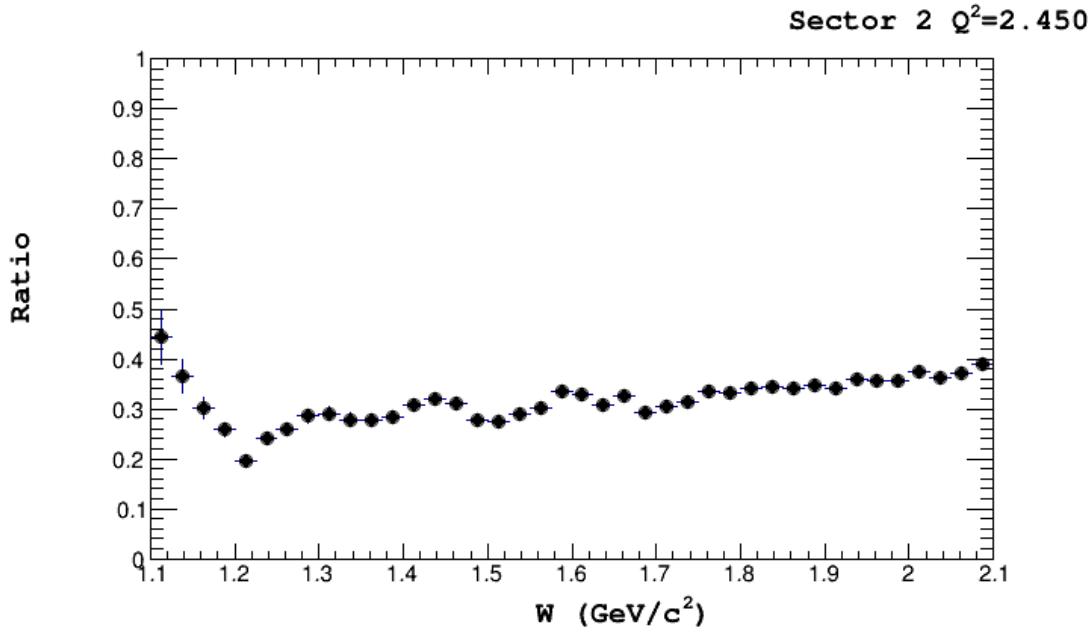


Figure 5.2: Acceptance for one bin of  $Q^2$  and one sector of CLAS.

sent the fraction of events reconstructed in the bin  $i$ .

$$\mathbf{A} = \begin{pmatrix} \epsilon_0 & 0 & 0 \\ \vdots & \ddots & \\ 0 & & \epsilon_n \end{pmatrix} \quad (5.11)$$

The inverse is,

$$\mathbf{A}^{-1} = \begin{pmatrix} 1/\epsilon_0 & 0 & 0 \\ \vdots & \ddots & \\ 0 & & 1/\epsilon_n \end{pmatrix} \quad (5.12)$$

and the corrected observation for the  $i^{th}$  bin is simply given by the observation over the efficiency.

$$t_i = \frac{m_i}{\epsilon_i} = m_i \frac{n_{gen=i}}{n_{rec=i}} \quad (5.13)$$

This is the simple *bin-by-bin* acceptance correction method, which is widely used and produces accurate results provided that bin migration is not significant. In this analysis the simple bin-by-bin acceptance correction is used.

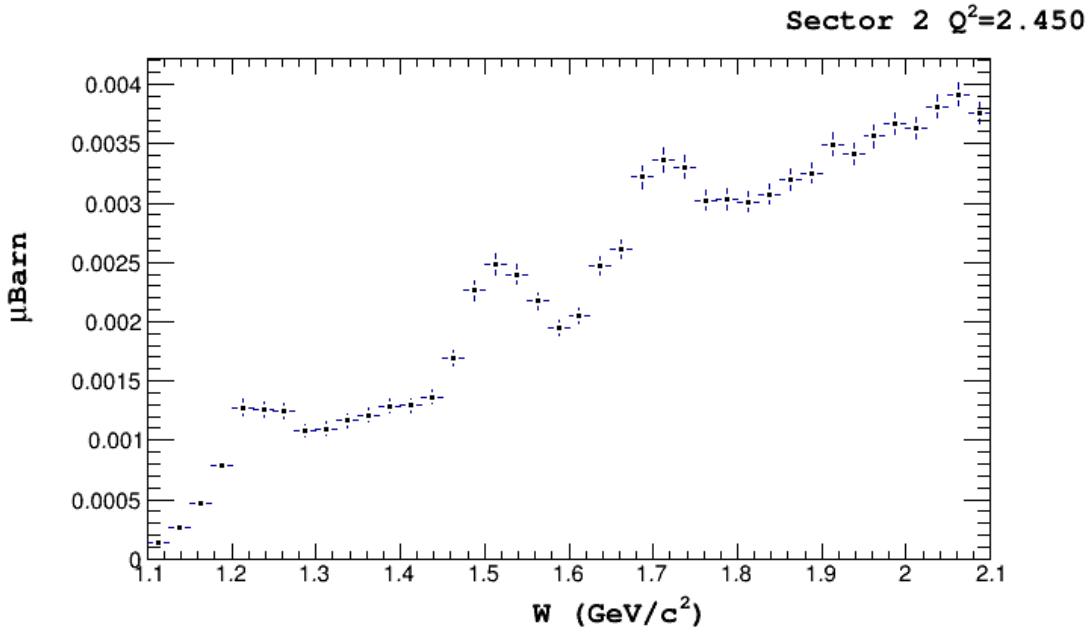


Figure 5.3: Inclusive cross section in the resonance region shown for one bin of  $Q^2$  in one sector of CLAS. Our calculation is compared with a trusted model.

### 5.2.3 Radiative Corrections

The removal of radiative effects from the measured distribution is a similar unfolding problem as described above for acceptance corrections. For this work we use two monte carlo event generators. The first generator produces inelastic events in the resonance region with radiative effects included that alter the kinematics. The second generator is the same parametrization but without radiative effects. The radiated generator is used to calculate the acceptance corrections, and both are used to try to remove the radiative effects on the cross section. The ratio  $R^{(i)}$  is defined for the  $i^{th}$  bin as shown below.

$$R^{(i)} = \frac{n_{unrad}^{(i)}}{n_{rad}^{(i)}} \quad (5.14)$$

This factor can be estimated without passing events through the simulation and we use the results directly from the output of the event generator to correct the cross section.

### 5.2.4 Model Comparison

## 5.3 Results for SIDIS

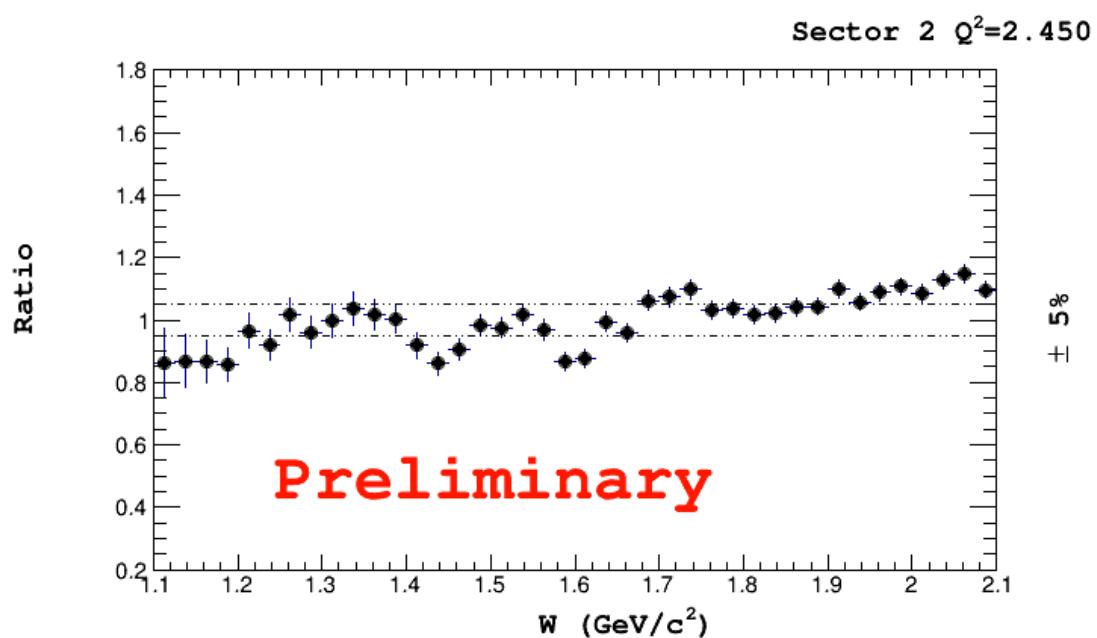


Figure 5.4: The ratio of cross section/model is shown here for one bin of  $Q^2$ .

# **Chapter 6**

## **TMD Extraction**

### **6.1 Introduction**

### **6.2 EVA**

#### **6.2.1 Wandzura Wilzcheck Approximation**

#### **6.2.2 Parametrization of TMD Functions**

### **6.3 Results for $\cos(2\phi_h)$ Modulation**

### **6.4 Predictions for $\cos\phi_h$ Modulation**

## Appendix A: Derivation of formulas related to errors

### Propagation of errors

Let  $\vec{x}$  be a set of  $n$  random variables  $\vec{x} = (x_1, x_2, \dots, x_n)$  and known mean  $\mu_i = \langle x_i \rangle$  and covariance  $V_{ij} = \langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle$ . Suppose that we measure a function  $f(\vec{x})$  that depends on the variables  $\vec{x}$  and we want to understand how the covariances  $V_{ij}$  on  $\vec{x}$  will show up manifest themselves as errors on our measurement of  $f(\vec{x})$ . We can start by expanding our function around the expected value of  $x_i$ .

$$f(\vec{x}) \approx f(\vec{\mu}) + \sum_{i=1}^n \frac{\partial f}{\partial x_i} \Big|_{x_i=\mu_i} (x_i - \mu_i) \quad (1)$$

We can then take the expectation value of our function.

$$\langle f(\vec{x}) \rangle = \langle f(\vec{\mu}) \rangle + \sum_{i=1}^n \frac{\partial f}{\partial x_i} \Big|_{x_i=\mu_i} \langle x_i - \mu_i \rangle \quad (2)$$

Where here the term  $\langle x_i - \mu_i \rangle$  is zero.

$$\langle x_i - \mu_i \rangle = \langle x_i \rangle - \mu_i = \mu_i - \mu_i = 0 \quad (3)$$

It is apparent then that the expectation value of our function  $f$  evaluated close to the expected values of our variables  $\vec{x}$  is just the function evaluated at the expectation value of the random variables  $\vec{x}$ .

$$\langle f(\vec{x}) \rangle = \langle f(\mu) \rangle = f(\mu) \quad (4)$$

We can also consider the second moment  $\langle f^2(\vec{x}) \rangle$ ,

$$\langle f^2(\vec{x}) \rangle \approx \left\langle \left( f(\vec{\mu}) + \sum_{i=1}^n \frac{\partial f}{\partial x_i} \Big|_{x_i=\mu_i} (x_i - \mu_i) \right)^2 \right\rangle \quad (5)$$

which is,

$$= \langle f^2(\mu) \rangle + \sum_{i=1}^n \sum_{j=1}^n \frac{\partial f}{\partial x_i} \Big|_{x_i=\mu_i} \frac{\partial f}{\partial x_j} \Big|_{x_j=\mu_j} \langle (x_i - \mu_i)(x_j - \mu_j) \rangle + 2 \left\langle f(\mu) \sum_{i=1}^n \frac{\partial f}{\partial x_i} \Big|_{x_i=\mu_i} (x_i - \mu_i) \right\rangle \quad (6)$$

and the last expectation value vanishes due to the same logic used when calculating the first moment. We recognize the term  $\langle (x_i - \mu_i)(x_j - \mu_j) \rangle$  as the element of the covariance matrix  $V_{ij}$ . Our second moment is then complete as follows.

$$\langle f^2(\vec{x}) \rangle = f^2(\mu) + \sum_{i=1}^n \sum_{j=1}^n \frac{\partial f}{\partial x_i} \Big|_{x_i=\mu_i} \frac{\partial f}{\partial x_j} \Big|_{x_j=\mu_i} V_{ij} \quad (7)$$

We can then calculate the variance of the function.

$$\sigma_f^2 = \langle f^2(\vec{x}) \rangle - \langle f(\vec{x}) \rangle^2 \quad (8)$$

$$= (f^2(\vec{\mu}) - f^2(\vec{\mu})) + \sum_{i=1}^n \sum_{j=1}^n \frac{\partial f}{\partial x_i} \Big|_{x_i=\mu_i} \frac{\partial f}{\partial x_j} \Big|_{x_j=\mu_i} V_{ij} \quad (9)$$

$$= \sum_{i=1}^n \sum_{j=1}^n \frac{\partial f}{\partial x_i} \Big|_{x_i=\mu_i} \frac{\partial f}{\partial x_j} \Big|_{x_j=\mu_i} V_{ij} \quad (10)$$

This is the standard error propagation formula which is widely used. These correlations  $\sigma_{ij}$  can arise from several sources.

- Common measurement uncertainties.
- Correlations in  $x_i x_j$  leading to correlations in  $\sigma_i \sigma_j$ .