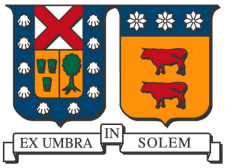
Study of Hadronization Process using Positive Pions with CLAS in Jefferson Lab



Sebastián Morán Vásquez
Universidad Técnica Federico Santa María
Physics Department

A thesis submitted for the degree of
Master in Physics

Valparaiso , 2021

To my family, my parents Maria Alejandra and Manuel Patricio and my brothers Daniel, and Franco. Also to my dear “Tía Carmen” and “Tía Hilda” .

Acknowledgements

I would like to thank first the person that is responsible for me doing this project, my advisor Hayk Hakobyan, from now on "El Hayko". El Hayko has always been very generous with his knowledge, open to discussion, supportive, and a very good friend in general. I also have to thank Orlando Soto, from now on "Orly", who help me immensely with my first steps in the field. Orly is one of the few people that knows how dumb I can really get. If I want to make justice explaining the benefits of working for someone like professor Will Brooks this thesis will be a thousand pages long, so I summarize it just by saying that we all in the group are better because of him, personally and professionally. Special thanks to Taisiya Mineeva for being always so nice to me, and for being so open to helping me when I need it (and for the tea of course). Thanks to Ahmed El Alaoui for sharing his infinite knowledge with me, especially with the horrible process of making a program compile. Thanks to Yuri Ivanov for helping me a lot with all the different nuances and mysteries of the university cluster, without him this thesis would not be possible. Thanks to my EG2 friends Esteban (who made me realize that that old saying "there are no stupid questions" is false), Jose, the other Jose, and another one (they are not EG2 members really (one of the Jose was but he was fired) but I put them here anyway), Antonio, Andrés and Gabriela for helping me with ... well I don't think they help me at all with this really, but they are fun to talk to (sometimes). Special thanks to my friend Esteban (this is another one) for helping me in so many ways since we meet back in 2013. Last but not least I have to thank Akira Toriyama for creating the best thing that is ever happened in the world and also to Eiichiro Oda for creating such a perfect series, it was the perfect 20 minutes (sometimes 4 hours) break during the making of this thesis.

And finally, really looking at the bright side of life, thanks to the coronavirus for forcing me to stay home working on this. I'm three pandemics away from the Nobel price.

Abstract

The study of pion electro-production can provide valuable information about the inner structure of the nucleus. Data obtained in the DIS regime provide direct information on hadron formation from quark's hadronization.

The work presented in this thesis is based on the EG2 experiment performed on CLAS detector at Hall B in the Jefferson Laboratory (Newport News, Virginia, USA), where 5 GeV incident electrons interact with different types of heavy nuclear targets (carbon, iron, or lead). The results are compared with interaction on deuterium target. Experimentally this was achieved through specially built double target system which guarantees the cancellation of several systematic effects.

The main observable discussed here is the hadronic multiplicity ratio for π^+, which is defined as the ratio between the heavy nuclei differential semi-inclusive cross sections to the one of deuterium. This observable allows to study the impact of nuclear medium on the hadronization process. To explore initial state effects, the EMC effect is also obtained. The experimental results may be compared with theoretical models.

Contents

List of Figures

Chapter 1

Introduction

The nuclear physics' birth dates back to the early 1900s, when Rutherford's experiments on the scattering of α particles on matter revealed the existence of a dense core at the centre of the atoms called nucleus. Since then, understanding the structure of the atomic nucleus have been a major concern to the scientific community. Many great contributions by the physicist of the past allows us to have a better knowledge of the atomic nucleus and the dynamics of its component parts. In the late 1960s, a major revelation change the current status of the nucleus, they themselves have internal constituents called quarks, whose existence must be inferred, since they are not observed directly in particle detectors. Extensive work has focused on studying this deeper underlying structure. The Quantum Chromodynamics (QCD) theory was introduced in order to describe the strong interaction between quarks. QCD is a very important part of the Standard Model (SM).

Despite the incredible accuracy and success of QCD, there are some areas in which the understanding is still very limited. Short distance process can be treated with a perturbative approach in QCD, but a long distance process has to be based on phenomenological models. Such as the hadronization process. This is the connection with the work presented in this thesis.

Hadronization is the process in which hadrons are formed, that is how a quark or gluon forms into a hadron before being detected as a free particle. The phenomenon is a consequence of confinement. This is the central puzzle in hadronic nuclear physics in Quantum Chromo-dynamics and cannot be studied with inclusive DIS where the final state hadrons are not observed. Another facet of confinement is that colored objects -such as, isolated quarks or gluons- may not be observed directly, so their properties must be inferred through indirect means. Semi-inclusive processes of hadron electro-production have been recognized long ago as an important tool for

testing QCD predictions of nucleon structure. This is they admit to get information about quark distributions in the nucleon for each flavour separately. Detecting hadrons in the final step of a DIS event (SIDIS) allow us to calculate an experimental observable called Multiplicity Ratio (MR) which is a measure of the attenuation of hadrons. This attenuation for different nuclear sizes enabled the experimental study of hadronization.

The goal of the present work is to explore the different nuances of the hadronic multiplicity ratio for positive pions using data collected in Hall B during an experiment called EG2, which consisted in a double target system exposed simultaneously to an electron beam of 5.014 GeV. The initial state effects are highlighted in the ratio of DIS yield (EMC effect) which is also studied here.

This thesis is segmented in different chapters. The first chapter is called **Physics Motivation** and is dedicated to a theoretical subject's introduction. The second chapter called **Experiment** is dedicated to explain the experimental setup, CEBAF accelerator, CLAS detector, with a brief explanation of its components and major features. In addition, the experiment, called EG2, is explained, the data comes from this experiment.

Next is the chapter called **Data Analysis** which goals are to explain the data taking procedure (how the information is stored); the particle identification scheme for electrons and pions; Defining the observables to calculate; some additional variables to used in the analysis; present and describe the simulation set and the corrections implemented in the observables. Then, the chapter called **Results and discussions** is the one dedicated to explore the result of the observables; effect of some cuts; corrections and comments about current state of the analysis; and the future perspective. Additionally, a chapter called **Analyses Comparison** is presented as part of this thesis in which an important and historical discrepancy between two independent analysis is explore and solved.

Chapter 2

Physics Motivation

2.1 Introduction

The goal of this chapter is to provide a theoretical review over the subject of interest. The main references used are [1], [2], [3], and [4].

In order to understand the structure of microscopic objects, such as an atomic nuclei, it is necessary to bombard them with small object and watch how these objects get deflected. This was the way Lord Rutherford discovered the nucleus. Since the 1950s, energetic electron beams are used in scattering experiments. This beam has nice properties [1] that allow us to obtain an atomic image.

From **Q**uantum **C**romo-**D**ynamics (QCD) we can calculated the expression for the scattering rate. The interaction takes place via virtual photon (γ^*) exchange [2] between the electron and the proton, which results in electron scattering due to momentum transfer. If the energy transfer is large enough, the proton may breaks apart.

In figure 2.1 is shown an schematic diagram of the electron-proton interaction for different relations between the wavelength of the probing photon (λ) and the *radius* of the proton (r) [3].

[1] for example the fact that they are easy to accelerate, manipulate and control the polarization. Interaction with other electrons is via EM (small probabilities of weak interaction). Protons and hadrons interact via EM and strong interactions. The fact that electrons are fundamental particles, truly point like particles, means that they won't break into smaller pieces. There is also weakness of the electron choice, such as the difficulty to reach high energies (tens of GeV).

[2] real photons have no mass, but a virtual photon have non-zero mass. This is called a off-shell photon, because it does not satisfy the energy momentum relation ($P^2 = P_\mu P^\mu = m^2$). In addition, virtual photons (or any virtual particle) are not detectable.

[3] It can be shown that $\lambda \sim \frac{1}{\sqrt{-q^2}}$

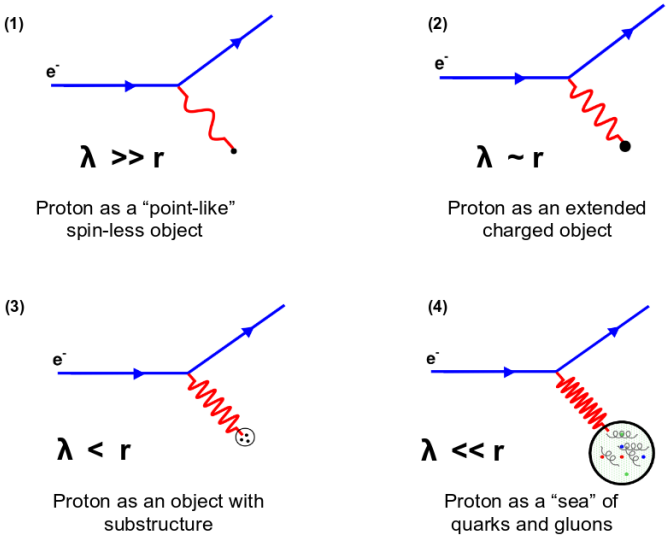


Figure 2.1: In $e^- p \to e^- p$, the scattering nature of the reaction for different values of the virtual photon's wave length.

2.2 QCD, pQCD and Coupling Constant α_S

QCD is the theory of strong interactions. It is formulated in terms of elementary fields (of quarks and gluons), whose interactions obey the principles of a relativistic QFT, with a non-abelian gauge invariance SU(3). However it is far from being complete, there are still some issues to address. Experimentally, there is a large number of phenomena that lack of a detailed qualitative and quantitative explanation, for example the fact of color confinement[4], which may be described as the central puzzle of hadronic nuclear physics.

In a typical collision at high energies (short distance) -of two protons for example- we have a large range of distance scales, all the way from $\sim 10^{-15}$ (the size of a proton) to $\sim 10^{-18}$ (the size of a quark). In addition, we have to understand the physics as it evolves through that three order of magnitudes change in scale.

Another key concept about QCD is how it behaves at different distance scales. One way to understand the fact that sometimes is suitable pQCD and sometimes not is by understanding that the coupling strength depends on distance. This goes by the name of what is called the *running coupling*.

[4]confinement is that colored objects such as isolated quarks or gluons may not be observed directly, so that their properties must be inferred through indirect means.

In QCD the resolution scale μ is very important, because the strong coupling depends on it:

$$\alpha_S = \frac{g^2}{4\pi} = \alpha_S(\mu) \tag{2.1}$$

this means that the coupling changes value as we change the resolution that were are probing.

This g coupling is just a parameter in Quantum Field Theory (QFT). Every time that there is a parameter in QFT, it goes through a process called *Renormalization*, meaning that the parameter must be defined because the field theory is divergent. There are some ultraviolet divergences that need to be absorbed in that parameter and, even though there is some ambiguity about how to absorb finite contributions, there is some inherent definition in the parameters of the QFT. In order to do that, we must pick a renormalization scheme for the parameters.

The normalization scheme used is called $\overline{MS}$, Modified Minimal Substraction. Once the scheme is picked, there is a scheme parameter that comes along with the definition; this is the μ parameter. After the process of renormalization, there is an equation that arises called the *betta function equation*, that tell us how α_S evolves with the scale μ. See Figure 2.2.

In the region where the strong coupling is high, it is also where the confinement process occurs. Here the coupling is so strong that does not let the quarks fly apart to each other, and it binds them into hadrons. In the region where the coupling is weak, pQCD is suitable, and if we go up to infinity in the plot, the strong coupling vanishes this means that at very very short distances the quarks are free. This is the idea of what is called *Asymptotic Freedom*. Additionally, there is some particular finite value where the coupling blows up. This is called the QCD Scale Λ_{QCD}.

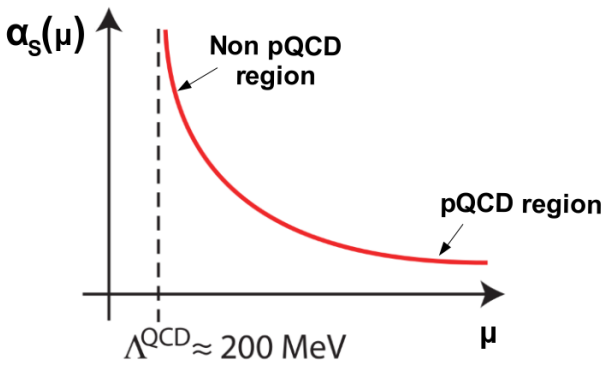


Figure 2.2: Strong coupling as a function of parameter μ.

- If $\alpha_S \to 1$: lower resolution, low Q^2, large scales, large distance, low energy. **Non pQCD** (example: confinement, hadronization.) Only static systems can be approximated in the non perturbative regime.

- If $\alpha_S \ll 1$: short distance, high energies, large Q^2 , **pQCD** (example: asymptotic freedom). Expansion in powers of $\alpha_S \ll 1$.

2.3 Elastic Scattering

Before getting into Deep Inelastic Scattering, let's start with the simpler case of an elastic scattering of an spinless electron colliding with a proton [5]. The process is :

$$e^- p \to e^- p \tag{2.2}$$

An schematic view of the process is depicted in figure 2.3.

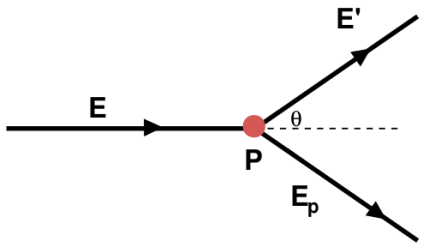


Figure 2.3: Schematic representation of a elastic scattering process. Electron- proton scattering $e^- p \to e^- p$

Using the Dirac equation to include the spin of the electron we find that the differential cross section $d\sigma$ per solid angle element $d\Omega$ is:

$$\frac{d\sigma}{d\Omega} = \frac{\alpha^2}{4E^2 sin^4(\theta/2)} cos^2(\theta/2) \tag{2.3}$$

this is called **Mott Cross Section**. Here $\alpha = \frac{e^2}{4\pi} \approx \frac{1}{137}$ is the fine structure constant (electromagnetic coupling). The Mott cross section formula is in the limit where the target recoil is neglected and the scattered particles are relativistic.[6] One

[5] neglecting the recoil of the proton.

[6] The "real" factor in this expression should be $1 - \beta^2 sin^2(\theta/2)$, but, if $\beta \to 1$, this turns into $cos^2(\theta/2)$. The case where the particle is **not** relativistic is called the Rutherford Scattering and is

$$\left(\frac{d\sigma}{d\Omega}\right)_{Rutherford} = \frac{\alpha^2}{16E_K^2 sin^4(\theta/2)},$$

important thing is that the Mott cross section is equivalent to scattering of spin 1/2 electrons in a fixed electrostatic potential.

We can consider the recoil of the proton and the spin-spin interaction between the proton and electron. The result of the differential cross section $d\sigma$ per solid angle element $d\Omega$ for an elastic scattering, assuming point-like Dirac spin half particles, is :

$$\frac{d\sigma}{d\Omega} = \frac{\alpha^2}{4E^2 sin^4(\theta/2)} \frac{E'}{E} \Big(\underbrace{cos^2(\theta/2)}_{\textbf{electric effects}} - \underbrace{\frac{q^2}{2M^2} sin^2(\theta/2)}_{\textbf{magnetic effects}} \Big) \tag{2.4}$$

in red are the new terms added to the previous Mott cross section, where the term E'/E is due to the recoil of the proton. The whole idea of this cross sections is that if we multiply this by an incoming flux, this allows us to figure out how frequently our electron detector should tick if we place it at an angle θ off the incoming electron beam's direction.

The second term within the brackets is the magnetic effect term -due to the proton spin- and depends on q^2. This means that if the electron passes by at low energies (low q), this term effectively vanishes, as it should since the electron then only feels the effect of a static charge.

One important observation of this differential cross section is that it depends on a single parameter. For a fixed angle θ, it can be shown that:

$$\frac{E'}{E} = \frac{M}{M + E(1 - cos(\theta))} \quad ; \quad q2 = -\frac{2ME^2(1 - cos(\theta))}{M + E(1 - cos(\theta))} \tag{2.5}$$

These provides a precise prediction for the behavior of scattered electrons. The proton is not point-like, it has an structure. Not knowing how to describe the inner structure of the proton, we can insert some mathematical functions of unknown form and value into equation 2.4. With the experiment we can discriminate some hypothetical functional forms for these unknown functions.

In this context, the so-called *Form Factors*, $G_E(q^2)$ and $G_M(q^2)$, appear. These are terms included in the cross section expression in order to allow the proton electric and magnetic structure to have their say on the electron scattering behavior. The form factors cannot be calculated from first principles. Up to now, they remain empirical functions, extracted from measured differential cross sections. The physical meaning of the form factors is that it measures the amplitude that, under an impact,

in this limit only the interaction between the electric charge of the particles matters.

the proton remains intact.[7] The result of thinking about the proton as an extended object is called the *Rosenbluth Formula*:

$$\frac{d\sigma}{d\Omega} = \frac{\alpha^2}{4E^2 sin^4(\theta/2)} \frac{E'}{E} \left(\frac{G_E^2 + \tau G_M^2}{1+\tau} cos^2(\theta/2) + 2\tau G_M^2 sin^2(\theta/2) \right) \tag{2.6}$$

where τ is a Lorentz invariant quantity defined as $\tau = -\frac{q^2}{4M^2}$. Also, the form factors are normalized so that $G_E(0) = 1$ and $G_M(0) = \mu_p = +2.79$ [8]. In figure 2.4 on the left experimental data show the form factors (extracted from the scattering rate) as a function of the square of the momentum transfer. They behave like a simple *dipole formula*. In figure 2.4 on the right is shown the high q^2 result, when the electric contribution becomes negligible and only the magnetic form factor G_M can be measured with reasonable accuracy. This shows the marked contrast with the hypothetical point proton idea.

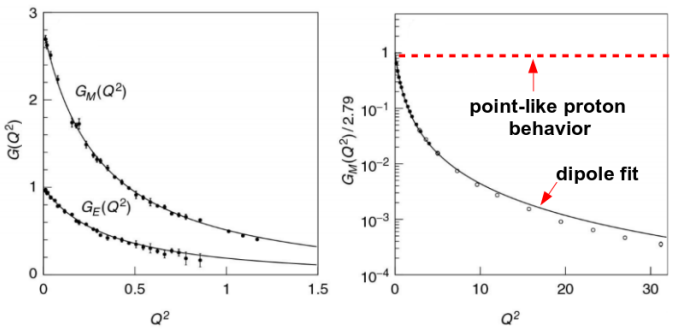


Figure 2.4: Form factors as a function of q^2, in $(\mathrm{GeV/c})^2$. On the right magnetic form factor, G_M, high q^2 result.

Until now, we only have explore the elastic scattering. In order to resolve the inner structure of the proton higher-energy interactions are required. As the momentum transfer yielded by the electron is brought up to about a 1 GeV and above, the virtual photon behaves like a very short-wavelength probe (see Figure 2.1), capable of penetrating deep in the core of the proton. In that case we are going to talk about **inelastic** scattering processes, which is the topic of the next section.

[7] As the impact gets larger, the probability that the proton remains a proton gets smaller, see figure 2.4 on the left.

[8] When expression 2.6 is derived, we assume that the proton was a spin half Dirac particle (that means $\vec{\mu} = \frac{e}{M}\vec{S}$), but the experimentally measured value of the proton magnetic moment is larger than expected for a point-like Dirac particle $\vec{\mu} = 2.79\frac{e}{M}\vec{S} = 2.79\vec{\mu}$. This anomalous magnetic moment is evidence that the proton is not point-like.

2.4 Deep Inelastic Scattering (DIS)

To "see" what is inside the proton, we need a high value of q^2 .We take that limit in equation 2.7:

$$\left(\frac{d\sigma}{d\Omega}\right)_{elastic} \overset{=}{\tau \gg 1} \left(\frac{d\sigma}{d\Omega}\right)_{Mott} \frac{E'}{E}\left(\frac{q^2}{2M^2}G_M^2 sin^2(\theta/2)\right) \tag{2.7}$$

where, from measurements, it holds that $G_M(q^2) \propto q^{-4}$, which means that the cross section goes like q^{-6}. This result tells us that at high q^2 elastic scattering reactions are unlikely and inelastic reactions, where the proton breaks up, dominate.

In figure 2.5, we can see the dramatic drop for the elastic case and the weakly q^2 dependence for the inelastic case. For DIS cross sections are almost independent of q^2, which means that the form factors $\to 1$. This is very important because a form factor close to unity implies a scattering from a point like-object, but this time that object is inside the proton.

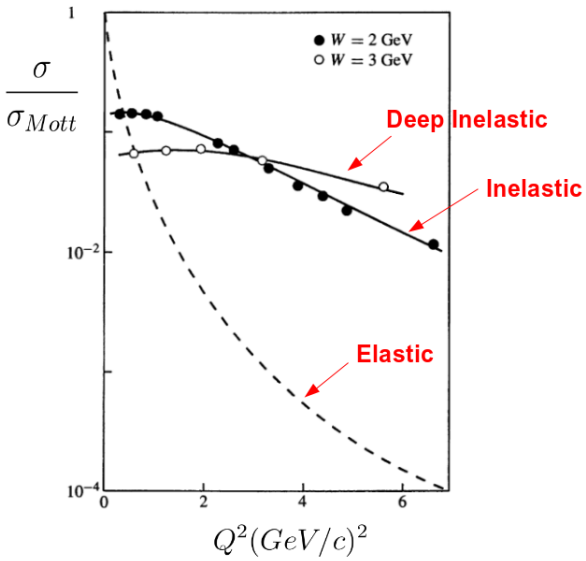


Figure 2.5: Early SLAC measurements of the inelastic electron-proton scattering cross-section divided by the Mott (point-like) cross-section, for two values of the invariant mass W of the hadronic final state.

In the inelastic scattering regime, the process is, at Born approximation (only one photon exchange), as depicted in figure 2.6[9]:

[9]note that in a general DIS case, the incoming lepton is not necessarily an electron, it may be a muon or a neutrino [15]. Since leptons can only participate in electroweak interactions, the

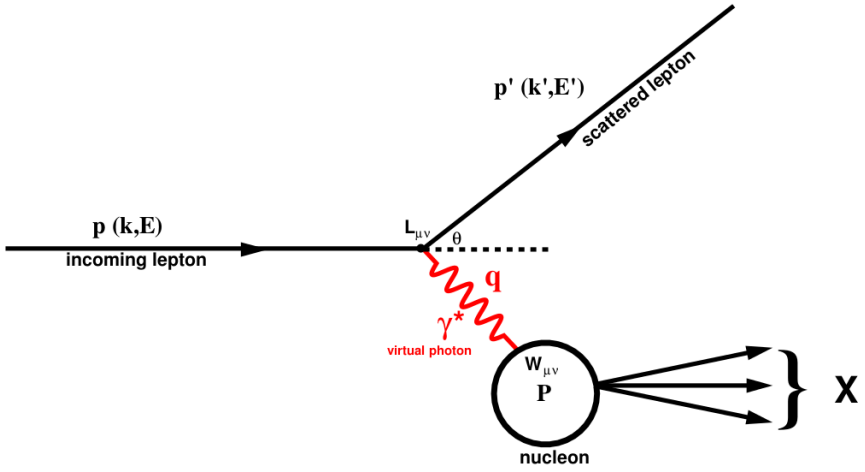


Figure 2.6: Schematic representation of a SIDIS process, at first order (one photon exchange).

where X is the hadronic final state. In DIS, the mass of the system X is not fixed, and the final state contains at least one baryon to conserve quantum numbers.

When we talk about measurement of DIS processes, it is necessary to make a distinction about what is detected at the end. There are three categories:

- Inclusive : when **only** the scattered lepton is detected.
- Semi-Inclusive : when the scattered lepton **and** one hadron are measured.
- Exclusive : when **all** products are identified.

In this thesis, we are going to work with the first case for the EMC ratio observable and with the second case for the hadronic multiplicity ratio, which is the main observable treated in this work. The main process considered here is called *SIDIS* (**S**emi-**I**nclusive **D**eep **I**nelastic **S**cattering) [28] [29], where at least one hadron is detected in coincidence with the scattered electron. The hadron carries information on the flavour of the struck quark. The reaction in SIDIS is:

$$p + P \rightarrow p' + h + X \tag{2.8}$$

exchanged particle is a vector boson V, such as $W^{\pm}$, γ or Z^0. The center of mass energy in our case -Jlab EG2 experiment- is of the scale of a few GeV, so the production of the heavy weak interactions carriers, $W^{\pm}$ and Z^0 is suppressed.

In this case, the hadron h measured is the positive pion π^+ [10] and X represents the remaining unobserved final state.

For inelastic scattering, there are four kinematic variables to consider: X_b, Y_b, ν and Q^2. Now, when we add the measurement of the π^+, additional variables are used. These are: Z_h, P_t^2 and ϕ_{PQ}. In the lab frame[11], the four momentum of the particles are given by:

$$P = (M,0,0,0) \tag{2.9}$$
$$p = (E,0,0,E) \tag{2.10}$$
$$q = \left(E-E',\vec{p}-\vec{p'}\right) \tag{2.11}$$
$$p' = \left(E',\vec{p'}\right) \tag{2.12}$$
$$P_h = (E_h,\vec{p_h}) \tag{2.13}$$

Note that this work will make use of natural units $\hbar = 1$ and $c = 1$. If we want to include the c factors in the calculation, simply replace $p \to pc$, $m \to mc^2$ and $v \to v/c$.

The first variable is the Q^2 and it is defined as:

$$Q^2 \equiv -q^2 \quad ; \quad Q^2 > 0 \ \text{ (probe resolution)} \tag{2.14}$$

where[12] :

$$Q^2 = -(q\cdot q) \tag{2.15}$$
$$= -\left((E-E')^2 - (\vec{p}-\vec{p'})\right) \tag{2.16}$$
$$= -\left(E^2 + E'^2 - 2EE' - \vec{p}^2 - \vec{p'}^2 + 2\vec{p}\vec{p'}\right) \tag{2.17}$$
$$= -\left((E-\vec{p}^2) + (E'^2 - \vec{p'}^2) - 2(EE' - \vec{p}\vec{p'})\right) \tag{2.18}$$
$$= -\left(\cancelto{0}{2m_e^2} - 2EE'\left(1 - \frac{\cancelto{1}{pp'}}{\cancel{EE'}}\, cos(\theta)\right)\right) \tag{2.19}$$
$$\approx 4EE' sin^2\left(\frac{\theta}{2}\right) \tag{2.20}$$

[10] this is not the only choice for our data. In addition to leading positive pions, leading negative and neutral pions and protons could be measured to the extent permitted by particle identification. Exploratory measurements with charged kaons could also be carried out with lower statistics and limited kinematic coverage.

[11] where the proton is at rest.

[12] for this calculus we are neglecting the electron mass ($k^2 \approx k'^2 \approx 0$, ultrarelativistic approximation), and using the relation $2sin^2(x) = 1 - cos(\theta)$.

Q^2 is known as the *virtuality* of the exchanged boson. The ν variable is defined as:

$$\nu \equiv \frac{P \cdot q}{M} \overset{lab}{=} E - E' \quad ; \quad \text{(probe depth)} \tag{2.21}$$

so, ν is the energy lost by the incoming particle.

The scaling variable Bjorken X, X_b or simply x is given by:

$$X_b \equiv \frac{Q^2}{2P \cdot q} \overset{lab}{=} \frac{Q^2}{2M(E - E')} = \frac{Q^2}{2M\nu} \quad ; \quad \text{(probe depth)} \tag{2.22}$$

which is the fraction of the proton momentum carried by the struck quark[13].

The Y_b, or simply y, inelasticity scaling variable is defined as:

$$Y_b \equiv \frac{P \cdot q}{P \cdot p} \overset{lab}{=} \frac{\nu}{E} \tag{2.23}$$

so , Y_b is the fractional energy loss of the incoming particle (or the fractional energy transfer in the lab). It holds that $0 < Y_b < 1$.

Now, let's introduced the variables that take into account the hadron. The first one is called Z_h and it is defined as:

$$Z_h \equiv \frac{P \cdot P_h}{P \cdot q} \overset{lab}{=} \frac{M E_h}{M(E - E')} = \frac{E_h}{\nu} \tag{2.24}$$

where the energy of the hadron is effectively calculated as $E_h = \sqrt{M_h^2 + (\vec{p_h})^2}$. Z_h, or just z, is the fraction of the virtual photon energy transferred to the observed hadron. It holds that $0 < Z_h < 1$.

The ϕ_{PQ} variable, or just ϕ is the azimuthal angle between the hadronic and leptonic production planes of the reaction. The planes are defined following the Trento convention. In figure 2.7, there is an schematic view of the planes. The leptonic plane is the one formed by the virtual photon, while the incoming lepton and the hadronic plane are formed by the virtual photon and the detected hadron.

[13] this interpretation is only true in the Breit frame, not the lab frame.

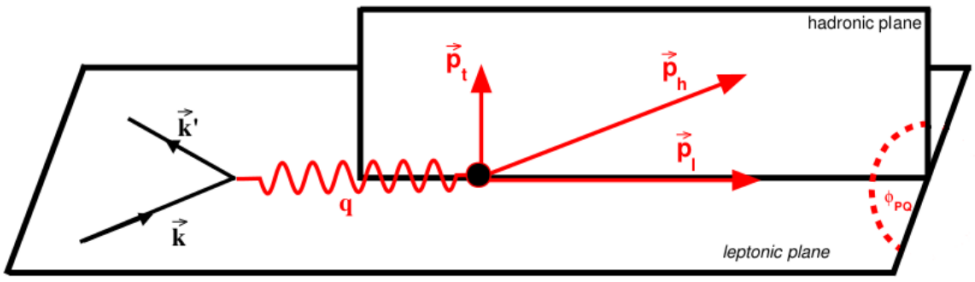


Figure 2.7: Depiction of Leptonic and Hadronic planes in SIDIS in the target rest frame.

where $\vec{p}_t$ is the transverse momentum. $\vec{p}_l$ is the longitudinal momentum, and $\vec{p}_h$ is the momentum of the measured hadron.

Thus, the azimuthal angle of the hadron relative to the virtual photon is given by:

$$\phi_h = cos^{-1}\left(\frac{(\vec{q}\times\vec{p})\cdot(\vec{q}\times\vec{p}_h)}{|\vec{q}\times\vec{p}||\vec{q}\times\vec{p}_h|}\right) \tag{2.25}$$

according to this definition $\phi_h \in [0, 180°]$, but for this analysis a different and equivalent choice is used, where $\phi_h \in [-180°, 180°]$. [14]

Finally, the P_t^2 variable, which is the transverse momentum of the hadron. If we defined the angle between $\vec{p}_{hadron}$ and $\vec{q}$ as θ_{PQ}, then (in lab frame):

$$P_t^2 = \left(\frac{|\vec{q}\times\vec{p_h}|}{|\vec{q}|}\right)^2 \tag{2.26}$$

$$P_t^2 = p_h^2(1 - cos^2(\theta_{PQ})) \tag{2.27}$$

The transverse momentum can be further written as:

$$P_t^2 = p_t^2 + Z_h^2 k_T^2 \tag{2.28}$$

where k_T is the intrinsic momentum of the struck quark, which produces a jet, and p_t is the momentum of the final hadron in the quark jet. These two observables are necessary in order to study the orbital motion of the quarks [7] [8] [9] [15].

At the end of the DIS process, the mass of the final state is no longer the proton mass (M) but it is W. To find the value of the invariant mass of the electron-nucleon interaction is straightforward:

[14] The relation is simply: $|\phi_{h(-180\to180)}| = 180 - \phi_{h(0\to180)}$.

[15] in this case, we can think of the PDF's as representing the probability of finding a parton q, with momentum fraction X_b and intrinsic momentum k_T.

$$W^2 = (P+q)^2 \tag{2.29}$$
$$= P^2 + q^2 + 2P\cdot q \tag{2.30}$$
$$= M^2 - Q^2 + 2M\nu \tag{2.31}$$

where $W > M$. Note that in the elastic case $W = M$, because the proton remains intact. In figure 2.8 there is the DIS differential cross section as a function of the invariant mass W. When the value of Q^2 is increased then the proton can be excited, for example:

$$e + p \rightarrow e + \Delta^+ \rightarrow e + p + \pi^0 \tag{2.32}$$

in this case the Δ^+ particle is called a *resonance* [16], and the final hadron comes from nuclear resonance decay. In this case the proton is not destroyed; it just got excited to a higher state. This could be interpreted as the proton having a structure and, when the electron gives energy, these constituents are excited to higher energies. Therefore, these states are formed.

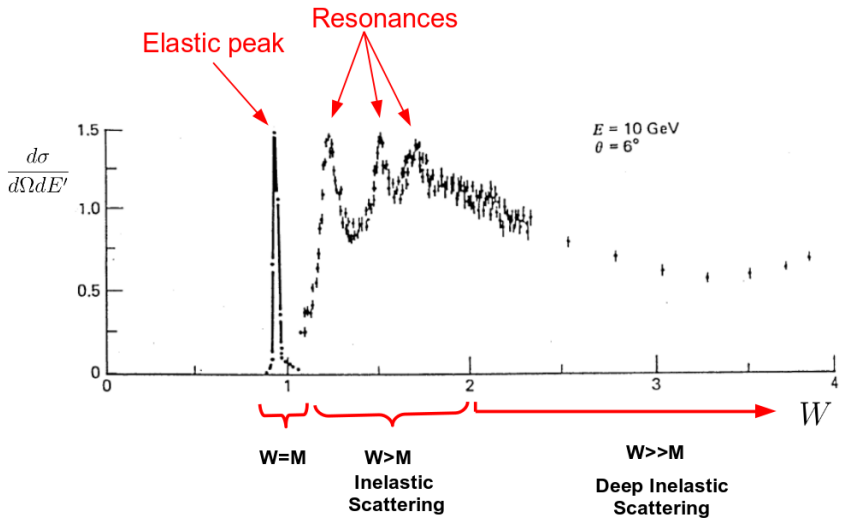


Figure 2.8: DIS differential cross section as a function of the invariant mass W (GeV/c^2). Note that E and angle is fixed.

[16] resonances are particles that have an extremely short lifetime, this is, a broad mass distribution. In this case, Δ^+ have the same structure than the proton (uud quarks), but a higher mass (1230 [MeV/c^2] instead of the 940 [MeV/c^2] of the proton). The Δ^+ decays very quickly into the proton and the π^0, taking about $\sim 10^{-23}$ seconds.

With all this Lorentz invariant kinematic variables, the Rosenbluth (Eq. 2.6) formula can be rewritten as:

$$\frac{d\sigma}{dQ^2} = \frac{4\pi\alpha^2}{Q^4}\left(\left(1-y-\frac{M^2y^2}{Q^2}\right)f_2(Q^2)+\frac{1}{2}y^2f_1(Q^2)\right) \tag{2.33}$$

$$f_2(Q^2) = \frac{G_E^2+\tau G_M^2}{1+\tau} \quad ; \quad f_1(Q^2) = G_M^2 \tag{2.34}$$

Now for DIS, we have not one independent variable (as in elastic), but two. Therefore, a double differential cross section is needed. It can be shown that the form of it is remarkable similar to the elastic case:

$$\frac{d^2\sigma}{dxdQ^2} = \frac{4\pi\alpha^2}{Q^4}\left(\left(1-y-\frac{M^2y^2}{Q^2}\right)\frac{F_2(x,Q^2)}{x}+y^2F_1(x,Q^2)\right) \tag{2.35}$$

these $F_2(x,Q^2)$ and $F_1(x,Q^2)$ are taking the place of the form factors, but this time are called *structure functions* (SF). It is convenient to express this in terms of the results obtained before from consider the lab frame of reference, that means, in terms of E' and the angle θ, that is what is measured. Taking the high energy limit ($Q^2 \gg M^2y^2$):

$$\frac{d^2\sigma}{dE'd\Omega} = \frac{\alpha^2}{4E^2sin^4(\theta/2)}\left(\frac{1}{\nu}F_2(x,Q^2)cos^2(\theta/2)+\frac{2}{M}F_1(x,Q^2)sin^2(\theta/2)\right) \tag{2.36}$$

where F_2 is the electromagnetic structure function, and F_1 is the pure magnetic structure function[17]. This cross section formula accounts only for single hard (highly energetic) photon exchange and thus, neglects multi-photon exchanges[18]. Now, it is clearer why in figure 2.8 the plot has to be done fixing two variables -the energy and the angle- and why there is a double differential equation. This structure functions are only determined experimentally, and it is observed that both are (almost) independent of Q^2. This fact is called **Bjorken Scaling**, and it is strongly suggestive of scattering

[17] Similar to the previous case, the term with the $cos^2()$ is the electromagnetic, and the term with the $sin^2()$ is the magnetic.

[18] As the electron beam scatters from a heavy nucleus, it interacts with the quark of a nucleon through the exchange of a hard photon and is accelerated and decelerated by the Coulomb interactions through the exchange of soft (less energetic) photons. This extends the standard one photon exchange model to a one hard photon and several soft photon exchanges.

from point-like constituents within the proton.[19]

$$F_1(x, Q^2) \to F_1(x) \quad ; \quad F_2(x, Q^2) \to F_2(x) \tag{2.37}$$

Also it has been observed that these structure functions are not independent but satisfy the **Callan-Gross Relation**:

$$F_2(x) = 2xF_1(x) \tag{2.38}$$

this relation confirms the fact that the quarks have spin one half. Also, note that if the quarks were spin zero particles, we would expect $F_1(x) = 0$ since this is purely magnetic function. These two facts of the structure functions can be explained thinking that DIS is dominated by the scattering of a single virtual photon from point-like spin half constituents of the proton. [20]

The relation between the structure functions can be explained with the *Parton Model*[21]. This says that the DIS is dominated by the scattering of a single virtual photon off one point-like spin half parton; the other partons are spectators. The scheme in this context is different than the one shown in figure 2.6. In the parton model, we should think in terms of interaction between γ^* and the parton. This photon is not an ordinary particle that can exchange an arbitrarily small amount of energy and momentum with the target, instead, it must be completely absorbed by the target. Thus, a single quark is projected forward sharply isolated in momentum space from the remaining partons. This leaves the latter with a distribution such as the unperturbed initial distribution in the target hadron with a hole in it.

[19] The Bjorken Scaling was proposed in 1968 by J. Bjorken. He said that the structure functions should depend only on the ratio ν/q^2 (proportional to x) in the limit $q^2 \to \infty$ and $\nu \to \infty$. This behavior could be understood by noticing that the wavelength of the virtual photon $\lambda \sim 1/Q$, but the resolving power is irrelevant if the object is point-like, hence the independence of Q^2.

[20] Feynman introduce the concept of *Partons* (in 1969) as the point-like constituents of the proton: quarks and gluons. Quarks are point-like fermions like leptons, but unlike leptons they take part in strong interaction as well as electromagnetic and weak. Gluons are massless spin 1 bosons, they carry the strong interaction. Partons are not physical particles, they can not propagate freely. We, therefore, need to describe the transition of the quarks and gluons in our perturbative calculations into the hadrons which can propagate freely.

[21] this model is an approximation and it can be viewed as the first term in an expansion in α_S. It is not an exact QCD prediction.

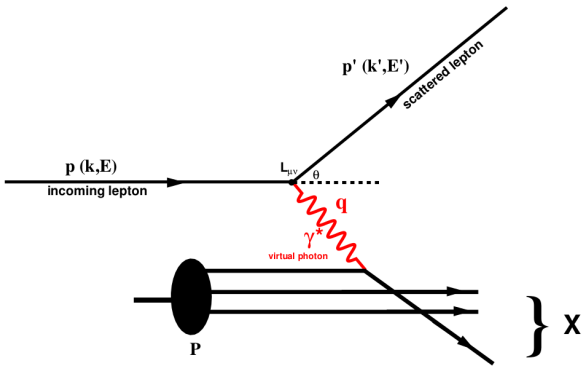


Figure 2.9: Representation of the process ep → eX in the Parton Model. The virtual photon interacts with **one** of the quarks constituents of the proton.

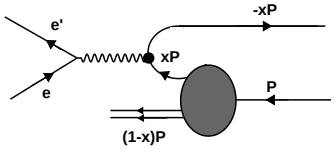
In the parton model, the basic interaction is elastic, and the quark is treated as a free particle (quasi-free). Also, this model is usually formulated in the *infinite momentum frame*[22]. The struck quark can be thought of as carrying a fraction of the proton's momentum. Only in this frame, this fraction is x, the X_b variable. The struck fast moving quark recoils and eventually fragments into hadrons in a characteristic way, independent of the other partons. The spectator partons also evolve into some final system of hadrons.

The cross section for this case (elastic $e^-\mathbf{q}$ scattering from a quark that carries x of the proton momentum) is:

$$\frac{d\sigma}{dQ^2} = \frac{4\pi\alpha^2 e_q^2}{Q^4}\left((1-y)+\frac{y^2}{2}\right) \tag{2.39}$$

where e_q is the charge of a quark of type q ($q = u, d, s, \bar{u}, \bar{d}, \bar{s}$). Note that is not a double differential cross section. In first order approximation, we can think of quarks as reduced protons. Lets say that the proton has (E,P,M), then the quark has (xE,xP,xM), that means *Scaling*. This is based on the idea that the proton has almost

[22] a frame where the proton has very high energy. We can neglect the proton and quarks mass, and any momentum transverse to the direction of the proton, i.e seeing the proton moving at a very large momentum towards the photon. The adjacent figure is an schematic view of this process (the parton comes with a fraction x of the proton's momentum) This frame is also called the Breit frame [55] or Infinite momentum frame (IMF) or *brickwall*. There is no energy transferred to the proton in this frame, and it behaves like if it had bounced off a brick wall. The formal definition of this frame is $2xP + q = 0$. Two important things happen to an hadron in this frame: first, it is Lorentz contracted in the direction of the collision, and second, the processes connected to its internal structure are time-dilated.



light-like momentum along the collision axis. Therefore, the quarks inside the proton also has light-like momentum, which is basically collinear to the proton's momentum. The only way the quark can acquired large transverse momentum is through the exchange of hard gluons, but here we are dealing with perturbation theory in QCD, which means small coupling α_S that suppressed this effect.

Now, each parton inside the proton carries a different fraction x of the proton's momentum. The probability that the struck parton carries a fraction x of the proton momentum is usually called **Parton Distribution Function** or just **PDF** [23]. Let's define:

- $f_q(x)dx$: number of quarks of type q, within the proton carrying a fraction x of the proton's momentum P, with momentum between $x \to x + dx$.
- $p = xPf_q(x)dx$: the total momentum carried by this quarks.
- $\frac{p}{P} = xf_q(x)dx$: probability that a quark carries a momentum fraction between x and $x + dx$.

Now, equation 2.39 gives the cross section for one particle. If we multiply for the total number of quarks of type q with momentum between $x \to x + dx$, and then sum over all types of quarks, we end up with:

$$\frac{d^2\sigma}{dQ^2dx} = \frac{4\pi\alpha^2}{Q^4}\left((1-y) + \frac{y^2}{2}\right)\sum_q e_q^2 f_q(x) \tag{2.40}$$

if we take equation 2.35 with the limit $Q^2 \gg M^2y^2$:

$$\frac{d^2\sigma}{dQ^2dx} = \frac{4\pi\alpha^2}{Q^4}\left((1-y)\frac{F_2(x,Q^2)}{x} + y^2F_1(x,Q^2)\right) \tag{2.41}$$

making the comparison between the two, we can remarkably say:

$$F_2(x,Q^2) = 2F_1(x,Q^2) = x\sum_q e_q^2 f_q(x) \tag{2.42}$$

the importance of this relation is that we can relate measured structure functions to the underlying quark distributions. At the present, PDF cannot be calculated from QCD, and is not possible to use perturbation theory due to a large coupling constant α_S. Therefore, the structure functions help us to determine them.

In figure 2.10, is shown that for values of $x > 0.05$ only a weak dependence of F_2 on Q^2. This is consistent with the quark-parton model, but at low x there is

[23] this probability distributions are used for all types of quarks, antiquarks, and gluons.

some clear **Scaling Violations**. That is because of gluon radiation. In order to understand this a little further, we must recognized that at higher values of Q^2 there is a shorter wavelength, this means that the virtual photon is able to "see" better the inner structure of the parton. For example: at some particular value of Q^2, we see a quark with a momentum fraction x_0. If we increase the value, we can realize that in fact the quark was sharing that x_0 with gluons. Thus, at higher values of Q^2 we expect to observed low x values quarks.

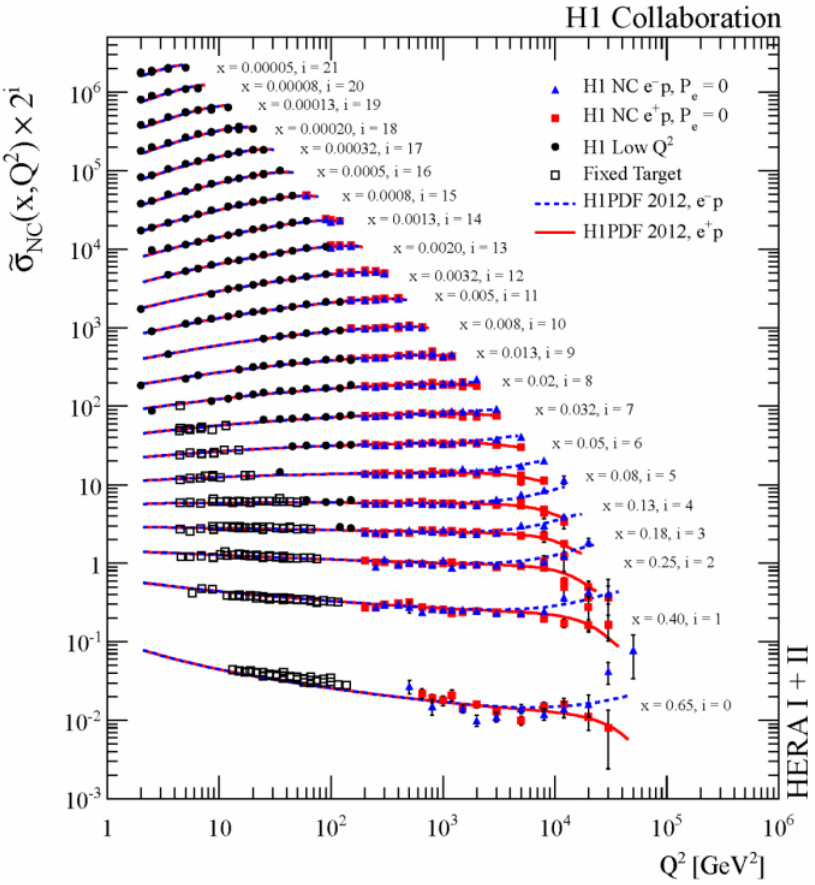


Figure 2.10: Scaling Violation example.

One important remark is that the integration of the quark distribution obtained from DIS gives ~ 0.5. There is some missing momentum. The cause is that we had not take into account the gluons[24] yet; they are the ones that carry the other half. The proton is made up of three quarks (uud) only at first order. At higher orders, the proton also contains anti-up and anti-down quarks. This is due to processes such as $g \to u\overline{u}$. It also contains heavier quarks, but those are usually neglected. Gluons play

[24] the gluons does not contribute to electron-nucleon scattering because they are neutral.

an important role in the Quark Parton Model. Nucleons are composed of **valence quarks** (the proton is made up of three valence quarks uud) that carry most of the nucleons momentum; **sea quarks** which are virtual quarks produced by gluons quark/anti-quark pair production (there is not a fixed number of sea quarks); and finally, gluons which carry fractionally less momentum. This fact of the sea being dominant at low momentum ranges can be explained in principle due to the gluon propagator, which goes $\sim 1/q^2$. All of this information can be combined into parton distribution functions. PDFs describe the makeup of the nucleon in terms of valence and sea quarks.

In figure 2.11 is shown the momentum distributions of the valence u and q quarks, quark-antiquark sea and gluons. The fits of three different groups are in good agreement. The data from all DIS experiments taken together in *Global Fits* using a QCD-based theoretical framework yield distributions of the individual partons.

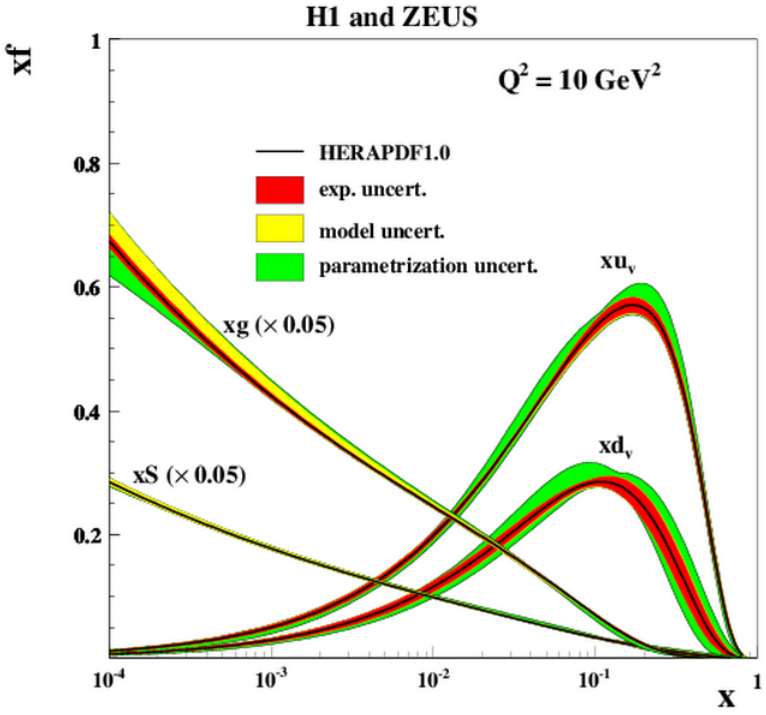


Figure 2.11: Example of a PDF set.

2.5 SIDIS Cross Section

The process when the proton breaks up into many hadrons is called Fragmentation. This takes place at low virtualities (small Q^2), and involves long time scales (long

distances) in comparison to the hard polarized lepton-quark interaction. For this reason pQCD, is not suitable for this process.

Theoretically, we can describe the DIS process using the **Factorization Theorem** [13]. This is a major assumption that allow us to introduce a **Fragmentation Function** (FF) $D_f^h(Q^2, Z_h)$, where:

- $D_f^h(Q^2, Z_h)$: probability that a quark of flavor **f** fragments into a hadron of type **h**, with a fraction $\mathbf{Z_h}$ of the virtual photon energy.

The basic assumption here is that the characteristic time of the parton-parton interaction is much shorter than any other long-distance interaction before or after the hard collision itself. In this case, we can treat the nucleus as a collection of free partons. Both the PDFs and the FFs are non-perturbative quantities, but in the collinear framework, where they are integrated over parton transverse momentum, they are believed to be universal, i.e., they does not depend on the particular type of process from which they were determined.[25] PdFs and FFs are treated as independent quantities. In the parton model, the inclusion of FF allows us to make a relation between the cross section for SIDIS and the cross section for a DIS process, that is:

$$\frac{d^3\sigma}{dxdQ^2dz}(eN \to ehX) = \frac{d^2\sigma}{dxdQ^2}(eN \to eX)\left(\frac{\sum_f e_f^2 q_f(x,Q^2) D_f^h(Q^2, Z_h)}{\sum_f e_f^2 q_f(x,Q^2)}\right) \tag{2.43}$$

this expression usually refers as:

$$\sigma_{SIDIS} = \sigma_{DIS} \times \left(\frac{\sum_f e_f^2 q_f(x,Q^2) D_f^h(Q^2, Z_h)}{\sum_f e_f^2 q_f(x,Q^2)}\right) \tag{2.44}$$

Note that factorization separate the cross section in three independent parts:

- The hard scattering cross section between the lepton and the parton.
- The parton distribution function.
- The fragmentation functions.

The PDFs and FFs are supposed to evolve with the scale Q^2. This is DGLAP (Dokshitzer-Gribov-Lipatov-Altarelli-Parisi) evolution [26].

[25] when we deal with hadron production in a nuclear system, universality is observed to breakdown. In other words, the details of the hadron production cross sections depends on the collision process that yields the final particles.

[26] the idea of the DGLAP evolution equations is that given the value of a function in a particular scale Q_0^2, we can perturbatively computed their value at any other scale. The gluon radiation produces the Q_0^2 evolution of parton densities.

2.6 Hadronization

Hadronization is the process of coloured partons' fragmentation into observed colourless hadrons, and it is a fundamental process in QCD. Because of the property of confinement, quarks can not propagate as free particles; they have to hadronize - which means they have to find a partner and form new hadrons.

In DIS, a quark is briefly liberated from any hadron and travels as a "free" particle. It is the mechanisms involved in hadron formation that enforces the colour charge neutrality and confinement into the final state hadron. This mechanism -which is assumed based on traditional treatments of confinement based on model potential or lattice QCD- can be studied. For example, the lifetime of the *free quark* may be inferred experimentally from the nuclear modification of hadron production on cold nuclei, which acts as detectors of the hadronization process.

It is possible to divide the hadronization process into two main parts, introduced as phenomenological tools:

- **Production Time:** process in which a highly virtual parton radiates gluons or splits into quark-anti-quark pair. Here, the parton can be described as an asymptotically free particle and it can be described using *perturbative* methods, such as, DGLAP equations.

- **Formation Time:** process in which the final hadron is formed. This process takes place at low virtualities (i.e $Q \sim \Lambda_{QCD}$).

  Hadronization is -at its last stage- both a dynamics and a low energy process, which means is dominated by non perturbative methods. Due to this, a number of approximate theoretical models were built.

Because of their complexity, first principle calculations modeling and phenomenology are often used to describe the hadronization process. That an experimental guidance is very important to support any theoretical development. In figure 2.12 there is a representation of the hadronization process divided into four steps:

- **Hard Scattering:** the process start with a hard scattering at time $t = 0$, where the quark absorbs the incident virtual photon. [27]

[27] Here $X_b > 0.1$ in order to avoid quark-pair production. When $X_b > 0.1$ the incoming lepton interacts incoherently with only one bound nucleon via a single virtual photon γ^* exchange. This region is dominated by the valence quarks. This means that we can treat DIS as e-q scattering. If $X_b < 0.1$, the situation is very different, because this region is dominated by the sea quarks. Here γ^* produces two jets: q and $\bar{q}$, those share the energy ν.

- **Production time:** after the hard scattering, the struck quark is deconfined, propagates and, since it has colour, loses energy via gluon Bremsstrahlung. This production time, in principle, should not be dependent on which final state the hadron is formed.

- **Colour neutralization time:** after the production time, the parton is likely to develop an inelastic cross section of the order of the final state hadron. Thus, a coloured pre-hadron is created. Here, the gluon emission continues for neutralising its colour.

- **Hadron Formation time:** here, the gluon radiation stops, and we have a colourless pre-hadron. The wave function of the pre-hadron collapses on the final hadron's wave function. The final hadron is formed. The magnitude of the formation time is expected to depend in detail on the hadron species being formed.

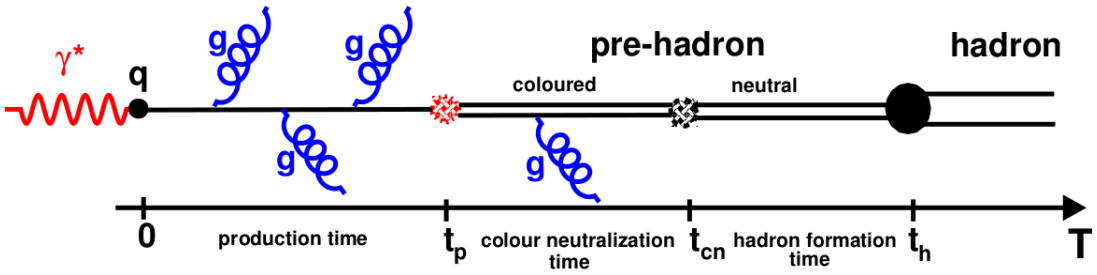


Figure 2.12: Schematic representation of the time evolution in the hadron formation process.

Here are some models for the hadronization process:

- Independent Fragmentation.

- Lund String Model.[28]

- Cluster Model.

[28]Lund model is a string model, where strings represents color fields and, when, sufficient energy is stored in the strings it breaks into a pair $q\bar{q}$. The fragmentation in PYTHIA is based on the Lund string model. For more details please see [68].

2.6.1 Hadronization in Nuclear Medium

The hadronization process is expected to be the same for partons travelling in QCD matter, but it will be modified by interactions with the surrounding medium. Medium is the nucleus itself, also called *cold QCD matter* [29]. The influence of the nuclear medium will have measurable consequences on the evolution of the hadronization process, such as the distance needed by the quark to neutralize its color and by the pre-hadron to expand into a hadron. The way we account for those modifications is by looking at the final hadron yield distribution on the nuclei as compared to those produced in "vacuum".

2.7 Observables

In the DIS regime, we start with an electron and, at the end, we measure -among other particles- a hadron. This process is the result of a sequence of components that can be summarized as follows:

- The leptonic part of the interaction (i.e γ^* emission, radiative corrections), described by **QED**.
- The hard scattering and parton evolution, described by **pQCD**.
- Hadronization and formation of final state hadrons, described by **phenomenological models**. This models rely heavily on experimental guidance.

The experimental observables -aiming at improving the description of nuclei- are highly desirable, since, after all nuclei represent the best QCD laboratory that nature might yield. It is important to take into account as many general principles as possible, that could be a key point in order to test or discard a theory. In this section, the aim is to explain, define, and motivate the observables calculated in this thesis.

2.7.1 Hadronic Multiplicity Ratio (MR)

The ideal case for an observable is that it meets the following requirements:

- Independent of initial states.

[29] In heavy ion collision (A + A), the produced parton must travel in addition the created hot and dense medium. This is called *hot QCD matter*. It could be a low temperature hadron gas or a quark gluon plasma (QGP) at high temperatures. In this cases, the hadronization process might be delayed until the medium cools down and comes closer to the confinement transition.

- Sensitive to the final states.
- Easily interpretable.[30]

In this thesis, the main observable is the hadronic **M**ultiplicity **R**atio (MR). One way to study fragmentation and hadronization is to perturb the environment surrounding the hard-scattered parton by introducing a nuclear medium. The nuclear medium provides a sensitive probe of parton evolution through the influence of initial-state (IS) and/or final-state (FS) interactions. Such IS and FS may result on modifications of the final hadron yield distributions compared to vacuum production and can help us understand, for example, the time-scale of the hadronization process.

Treating the nucleus as a collection of free partons, we can say that -in the absence of any initial state effect- the parton distribution function in a nucleus with mass number A is simply the superposition of A independent nucleons:

$$f_A(x,Q^2) = \mathbf{A} f_N(x,Q^2) \tag{2.45}$$

now, if in addition, there is no final state effects the parton fragments with universal fragmentation functions and the pQCD factorization theorem predicts that minimum-bias inclusive hard cross sections scale respectively as:

$$d\sigma\left(l,h+A\rightarrow h+X\right) = \mathbf{A} d\sigma\left(l,h+p\rightarrow h+X\right) \tag{2.46}$$

To test the validity of this equation, nuclear DIS is the cleanest environment for it because there are many kinematical variables controlled, and the medium -that is the nucleus- is well known. This represents the ratio of the handrons' number of type h produced per deep-inelastic scattering event on a nuclear target of mass A to that from a deuterium target. In equation 2.47 is defined this double ratio.

At leading order, this double ratio corresponds to a good approximation to the fragmentation functions' ratio in cold nuclear matter over that in *vacuum* (deuterium).

$$R_M^h(Z_h,\nu,Q^2,P_T^2) = \frac{\left(\dfrac{N_h(Z_h,\nu,Q^2,P_T^2)}{N_e(\nu,Q^2)}\right)_A}{\left(\dfrac{N_h(Z_h,\nu,Q^2,P_T^2)}{N_e(\nu,Q^2)}\right)_D} \tag{2.47}$$

[30] this was one of the biggest problems with early results on hadron production at SLAC (or EMC). Some of those issues were proton contamination in the hadron sample; observables sensitive to initial state nuclear effects; low statistical precision; among others. This may have hidden some important features of the phenomenon. Even then, a global picture emerged from the data and valuable information was obtained.

here N_h is the yield of semi-inclusive hadrons in a given (Z_h,ν, Q^2, P_T^2) bin and N_e is the yield of inclusive DIS leptons in the same (ν, Q^2) bin [31]. The attenuation of the hadron production on a nuclear target, relatively to deuterium, is given by 1-R_M^h.

It is the normalization with the numbers of electron which permits to cancel the initial state effects. Also, the particle multiplicity in the final state is low for nuclear DIS leading to precise measurements.

In the absence of any kind of attenuation produced by interactions in the medium, these two quantities would be identical. That means $R_M^h = 1$, but, if we look at Figure 2.13, there is a significant difference between what we expect and what we observed. This shows clearly a breakdown of universality for fragmentation functions. Produced particles interact with the medium and somehow get suppressed. According to the figure, almost 60% of the hadrons "disappear" from the high Z_h bin; this is a dramatically strong interaction (later in this thesis is shown that for Pb this numbers are even bigger, reaching 80% at high Z_h). The Zh dependence of the multiplicity ratio will be the most used in this work because is at the heart of the hadronization process. Hard parton scattering cross section is contained within the definition of fragmentation functions, which have known evolution with $Q2$, and depend primarily on Zh.

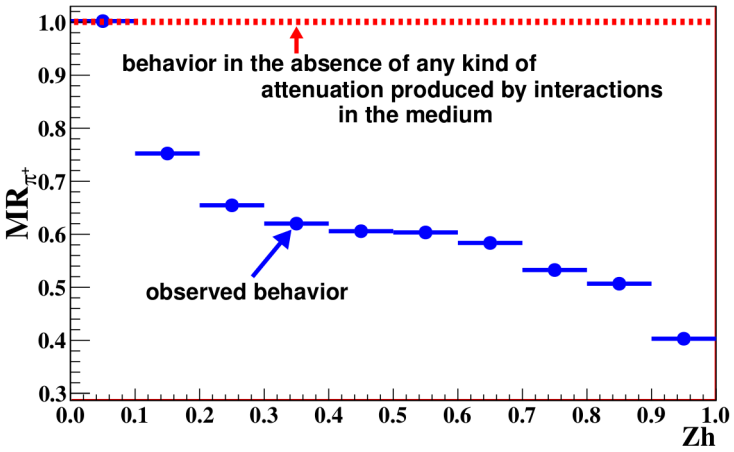


Figure 2.13: Multiplicity ratio curve, as a function of Z_h example, hadron attenuation in cold nuclear matter.

2.7.2 EMC Effect

Protons and neutrons act differently when they are inside an atom, versus floating freely through space. As it was mentioned before, the normalization by the number of

[31] the variables here can change depending on the case.

electrons in the R_M^h was to cancel out the initial state effects. However, if we now only take the ratio of the number of inclusive DIS electrons, this will represent the main initial state in nuclear lepton scattering. One of the most basic models to describe the nucleus represents it as a collection of free nucleons moving non-relativistically under the influence of a mean field. In this picture -in the rest frame of the nucleon- should hold that partonic structure functions (SF) of *bound* nucleons are identical to partonic structure functions of **free** nucleons. Since DIS experiments are sensitive to the partonic SF of the nucleon, it was generally expected (except for nucleon Fermi motion) that they give the same result for all nuclei. Instead, the measurements show a reduction in the structure function of nucleons bound in nuclei relative to nucleons bound in deuterium. This was called the EMC Effect[32]. See figure 2.14 as an example of this reduction in the SF.

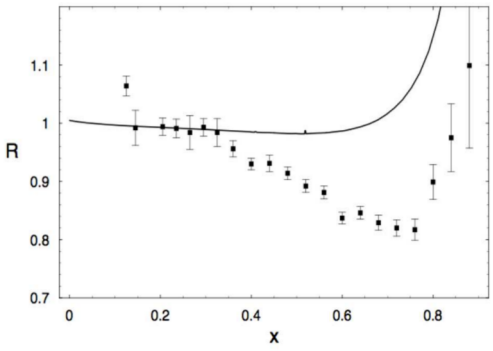


Figure 2.14: Measurements of the DIS cross section ratio of gold relative to deuterium as a function of X_b from SLAC. The solid black line is the expected ratio taking into account only Fermi motion of nucleons in Gold. Taken from [11].

The nuclear ratio $R_{F_2}^A(x, Q^2)$ is defined as the nuclear structure function per nucleon divided by the nucleon structure function:

$$R_{F_2}^A(x, Q^2) = \left(\frac{F_2^A(x, Q^2)/A}{F_2^{nucleon}(x, Q^2)} \right) \tag{2.48}$$

where A is the number of nucleons inside the nucleus. Now, the nucleon structure function is usually defined through measurements on deuterium:

$$F_2^{nucleon} = \frac{F_2^{deuterium}}{2} \tag{2.49}$$

[32]due to the **E**uropean **M**uon **C**ollaboration, which is the group that first discovered the effect, back in 1983.

this is the case assuming that the nuclear effects in deuterium are negligible. The behavior of this nuclear ratio -as a function of x, for a fixed Q^2- is shown schematically in figure 2.15. From this figure we can recognized four distinct regions:

- **Shadowing**: nuclear shadowing is caused by hadronic fluctuations of the intermediate boson. The fluctuated hadrons multiple scatter off surface nucleons, but are invisible to the nucleons bound deeper in the nucleus. This leads to an interference term in the cross section, which in turn reduces the bound nucleon structure functions. Nuclear shadowing occurs at low X_b because there is a coherence length associated with the hadronic fluctuations of the boson.

- **Antishadowing**: is the converse of shadowing. Anti-shadowing compensates for the missing quarks by enhancing the cross section at intermediate X_b values.

- **EMC Effect**: is the depletion of the bound nucleon structure functions in the intermediate X_b region of $0.3 < X_b < 0.7$. The fact that this happens in intermediate values of X_b tell us that the origin of the phenomenon emerges from the valence quarks in the nucleon. Larger nuclei exhibit a stronger EMC effect relative to deuterium than lighter nuclei. We will see more on this aspect later.

- **Fermi Motion**: in the region of high X_b, the valence quarks dominate the structure functions. The fact that the bound nucleon structure function begins to grow much larger than the deuterium is known as Fermi Motion. The cause of this is that, since the nucleons inside a nucleus may have an X_b up to $X_b = A$, [33] the nucleons have access to a larger X_b than those bound in deuterium (lighter nucleus).

[33] A is the number of nucleons inside a nucleus.

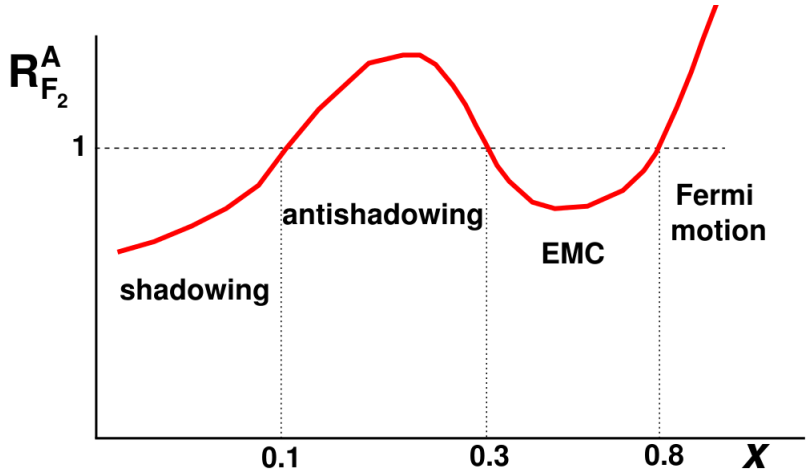


Figure 2.15: Schematic behavior of $R^A_{F_2}$ (x, Q^2) as a function of x for a given fixed Q^2. Taken from [10].

Quarks slow down massively once they are confined to a nucleus in a atom. The strong force between the quarks determines their speed, whereas forces that bind the nucleus are supposed to be small. Two particles in a nucleus are pulled together by a force of ~ 8 [MeV], quarks are bound together by ~ 1000[MeV]. Experimentally the apparent size of the nucleus can change by 10 up to 20 percent in comparison when they are free. Since the discovery of the phenomenon a lot of papers have been produced, but there is no definitive explanation available yet. In figure 2.15, we can see an example of the EMC curves obtained at our kinematics. In the y axis is the ratio of electrons in solid target C, Fe, and Pb divided by the deuterium case as a function of Xb. From this plot, we can see the stronger effect for heavier nuclei clearly.

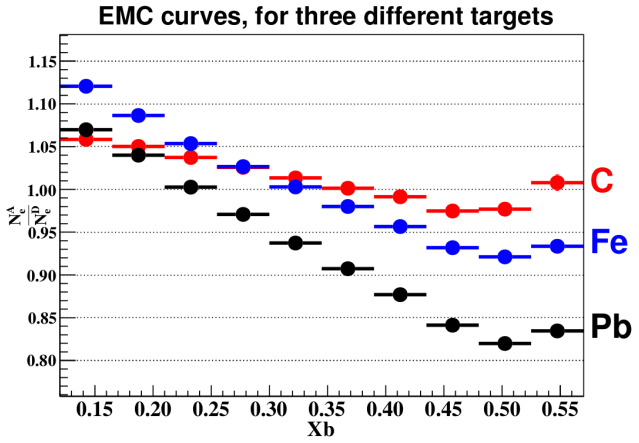


Figure 2.16: EMC curves for three different targets: carbon, iron, and lead.

Since the EMC effect is not completely explained by theory, it is necessary to ensure the experimental analysis as complete and accurate as possible. All non negligible corrections must be made as well as analysing the data, so that it may be compared to theory. In this thesis, two corrections are considered: Coulomb and radiative corrections.

2.8 Previous Measurements

This section's aim is to give a brief overview about the results in the multiplicities for hadrons in the past. Historically, the first measurements of the MR (not necessarily in the exact same form as presented in section 2.7.1) was done at SLAC (**S**tandford **L**inear **A**ccelerator **C**enter) in the late 1970s [14]. Apart from intrinsic problems at that stages of the study, the main global picture was concluded from their data. Their results can be summarized as that the attenuation of the forward component is observed and increases with A. As it was depicted in figure 2.13, the produced particles interact with the medium and get suppressed by some mechanism. The same main conclusion was confirmed by Fermilab and EMC collaboration. The biggest set of data were due to the HERMES collaboration, which, among many other things[34], observed the fact that there is more attenuation as Z_h increases. Refer to [25] and references therein for more details.

[34] like the fact that is less attenuation as ν increases, that there is no sensitive observed as a function of Q^2, X_b, or Φ. Also, the enhancement above unity of MR as a function of P_t^2, the so-called *Cronin effect*.

Chapter 3

Experiment

In this chapter the detector and the experiment are described. The **T**homas **J**efferson **N**ational **A**ccelerator **F**acility (TJNAF) -in Newport News, VA, USA- is the site of a recirculating linear electron accelerator, capable of delivering GeV electron range beams to four experimental Halls simultaneously. Such Halls are A, B, and C. In those Halls, different physics are studied. The data's experiment for this thesis is called *EG2*, and it took place in Hall B between January 9 and March 5, 2004 where a beam of 5 GeV was produced.

3.1 CEBAF

The **C**ontinuous **E**lectron **B**eam **A**ccelerator **F**acility, CEBAF (figure 3.1) is a pair of antiparallel superconducting radio-frequency (RF) linear accelerators (linacs) built 8 m below the earth's surface joined by two 180° arcs with a radius of 80 meters. The recirculating arcs are composed of five separate beam line sections, which permit the beam to recirculate in both linacs up to 5 times. For each linac, the gain in energy of the beam varies between 0.4 GeV to 0.6 GeV , giving a final pick energy of $\sim$ 6 GeV. CEBAF is designed for experimental investigations of the electromagnetic structure of mesons, nucleons, and nuclei through high energy electron and photon beams.

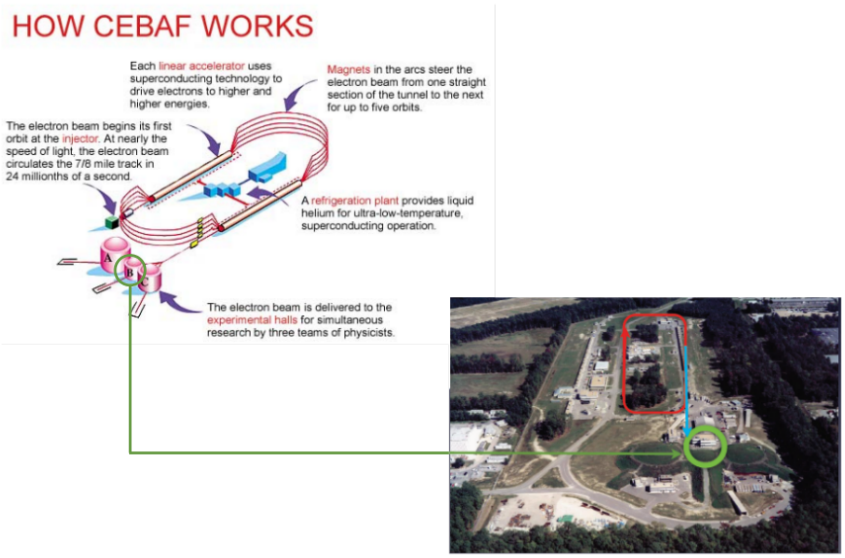


Figure 3.1: Schematic view of the CEBAF accelerator in JLab, Newport News, VA, USA.

3.2 CLAS

The main detector in Hall B was called **C**ebaf **L**arge **A**cceptance **S**pectrometer, or simply CLAS. The design of the CLAS detector is based on a toroidal magnetic field, which provides good momentum resolution; and a large angle acceptance of charged particles -almost 4π acceptance in solid angle and covers angles between $8°$ and $142°$ relative to the direction of the incoming beam from the accelerator. The target is located inside the detector on the beam axis. CLAS is well suited for experiments, that require the detection of two or more particles in the final state because it connects the polar angle range with the full azimuthal coverage. For Hall B the electron beam can reach currents of a few nA; this correspond to a luminosity of $\sim 10^{34}[\mathrm{cm}^{-2}s^{-1}]$. CLAS is composed of several detectors designed to measure properties of the particles from the reaction.

CLAS is divided into 6 identical sectors -see figure 3.2- each of one works as an independent spectrometer[1]. 5 meter long superconducting coils of toroidal shape magnets separate the sectors. The reason of the toroidal shape is this magnetic field will bend particles only in the polar direction, not in the azimuthal. For CLAS experiments the standard is that the direction of the field makes the negatively charged particles in-bending compared to the direction of the beam. This will be the case

[1] even though they share the target, trigger and data acquisition system.

of the experiment of this thesis (called EG2). The particle detector system consists of **D**rift **C**hambers (DC), **C**herenkov **C**ounters (CC), **T**ime-**O**f-**F**light (TOF) **S**cintillation **C**ounters (SC), and **E**lectromagnetic **C**alorimeters (EC.) Each one of these will be explained in the next section.

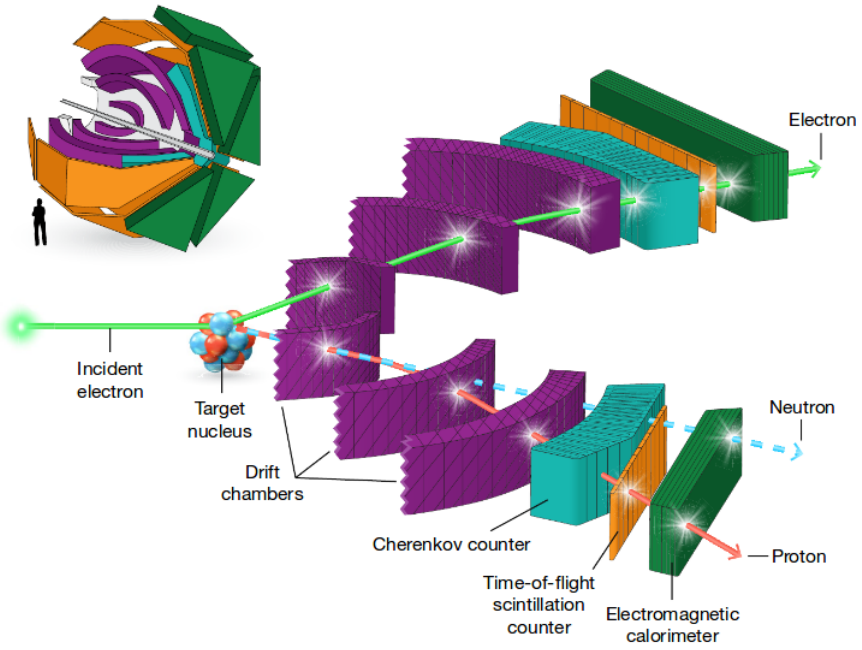


Figure 3.2: Schematic view of two segments of CLAS spectrometer. The trajectories of the charged particles are shown as they bend in a toroidal magnetic field. The neutron momenta are deduced from the time of flight until they interact with the electromagnetic calorimeter. The grey pipe in the upper figure is where the electron beam travels along, hitting a target near the centre of the spectrometer. Figure taken from [16].

3.2.1 Drift Chambers (DC)

As soon as the particles are scattered from the target, the first detector encountered are the Drift Chambers [17]. The main function of the DC are to determine the trajectory of charged particles of momentum -measuring the bending of the track- above 200 MeV. There are 3 DC in each sector, 18 DC in total. The polar angles (θ) that DC covers are between 8° and 142°.

There are three radial layers called *regions*, to take into account:

- **Region 1**: region of the chambers surrounding the target (innermost). The magnetic field here is weak. Here is determined the initial direction of the particle while passing through the DC.

- **Region 2**: region between the magnetic coils, where the field is the strongest. Here is determined the maximum curvature of the particle track.

- **Region 3**: region outside the magnetic field influence. It has the largest chambers. Here is determined the final direction of the particle while passing through the DC.

Each region of the DC consist of two superlayers of wires, and each superlayer consist of six hexagonal wire cells, see figure 3.3. Each cell has a sense wire in the center surrounded by six shared wires, located at the vertexes.

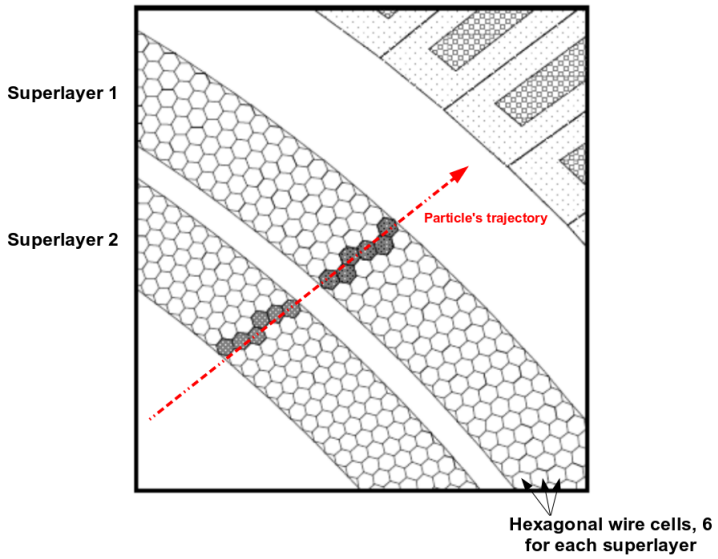


Figure 3.3: Schematic view of a Drift Chamber.

The main idea of the DC is based on the fact that the electron ionize matter, in this case the gas that filled the volume of each DC (90% Ar and 10% CO2.) The electrons produced by the ionization drift toward the central wire, and as they get closer to the wire they create an avalanche of electrons. This signal is amplified and recorded.

In figure 3.3, we can see that the shadowed regions are a track segment. the idea is to form the track linking these different track segments from each super-layer. It is important to notice that we need additional information. We need the distance of the particle to the cell's center. This information is obtained with the Scintillator Counters (SC.)

3.2.2 Cherenkov Counters (CC)

The main purpose of the CC [18] is to discriminate between electrons and negative pions after they have passed through the DC. The CC can collect information about a charged particle that travel faster than the speed of light in that medium. In CLAS, there are CC in each sector in the forward region covering polar angles between 8° and 45°, and covering the full azimuthal angle. This background of negative pions are mainly produced in quasi-real photo-production, when scattered electron goes at polar angle θ very close to zero and, therefore, not measured by CLAS.

The Cherenkov radiation produced by the electron as it passes through the gas inside the CC -C_4F_{10}, with a high refraction index ($n = 1.0014$) which results in a high photon yield and an acceptable high pion momentum threshold of ~ 2.5 Gev/c- is collected with mirrors, light-collecting cones, and Photo-multiplier tubes (PMTs), see figure 3.4. Electrons cross through approximately 1 m of gas, while the produced Cerenkov photons undergo between 3 and 4 bounces before they reach the PMT face. The CC is designed to just take into account the azimuthal direction due to the fact that placing the PMTs in the shadows of the magnetic coils does not directly influence the angular coverage). This is the case because the magnetic field just affect the polar angle of the trajectories of the particles, which can be thought of as taking place in a plane of constant ϕ. In this case, the polar angle information of the track is not affected.

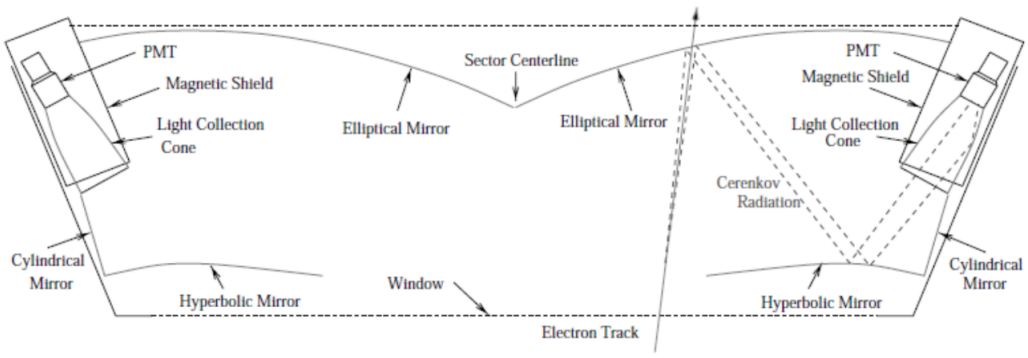


Figure 3.4: Schematic view of a Cherenkov Counter.

Each of the six CLAS sectors consist of 18 CC *segments* in θ. Each segment was divided into two modules about the symmetry plane bisecting each sector, see figure 3.5. This results in a total of 12 identical -except for an inversion symmetry- sub-sectors. For an upgraded version of this please, see [19].

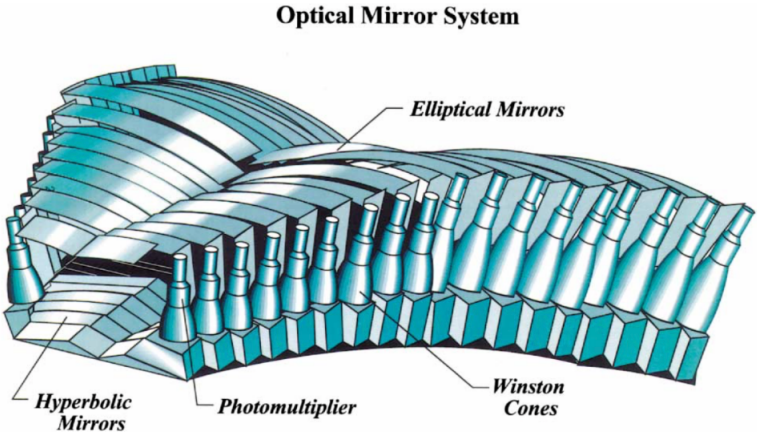


Figure 3.5: One sector consist of 6 elliptical and 36 hyperbolic mirrors focusing the Cherenkov light onto 36 PMTs. There are 6 of this Optical Mirror System in CLAS detector, one in each sector.

3.2.3 Scintillator Counters (SC), Time of Flight (TOF) System

The TOF detectors [20] consist of a group of scintillators positioned after region 3 of the DC between CC and EC covering polar angles between 8° and 142°, and full azimuthal extension. The TOF system serves as their main tool for particle identification -as will be shown in the π^+ identification-. The idea is that it measures the time between the point of interaction in the target to the external boundary of CLAS, which is where the SC are located.

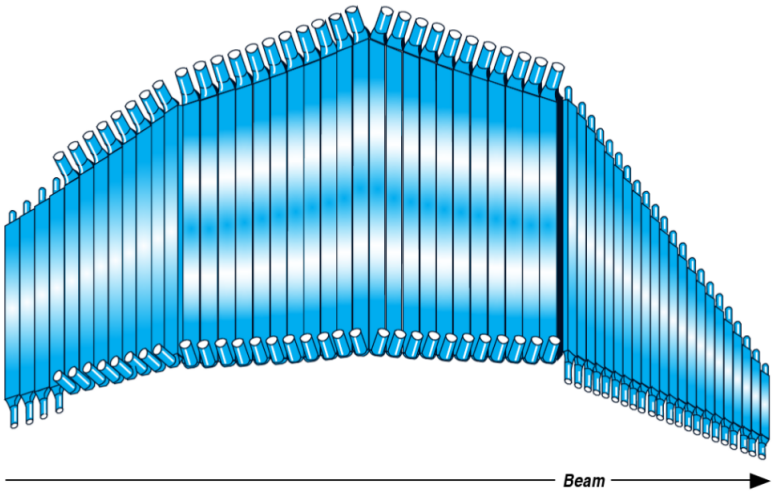


Figure 3.6: Schematic view of a TOF detector for one sector, in CLAS.

3.2.4 Electromagnetic Calorimeter (EC)

The forward sampling electromagnetic calorimeter (EC) [21] is installed in the forward region and consists of two main parts:

- Active Medium: scintillator strips, which produces the signal output.
- Passive Medium: lead sheets, which works as an absorber with the purpose of inducing a shower.

The EC covers an angle from $\theta = 8°$ and $\theta = 45°$ [2]. In two sectors, the coverage is extended up to $\theta = 75°$ by a second detector called **L**arge **A**ngle **C**alorimeter (LAC).

For energies of order of the GeV, photons lose energy mainly due to pair production. Electrons -and positron as well- lose it due to Bremstrahlung radiation.[3] These processes create photons that will cause an electromagnetic shower[4] which will be detected using the scintillators (active medium.) Hadrons are heavier particles, so they will lose energy mainly by ionization. The signal is very different compared to the case of electromagnetic showers given by photon and electrons or positrons. This feature will be important when we deal with pion plus identification because there exist a threshold value of momentum (2.7 GeV). If the pions are above this value, they start to fire the Cherenkov detector.

The main purpose of the EC, in CLAS, are:

- Detection of electrons with high energies above 0.5 GeV.
- Detection of photons with energies higher than 0.2 GeV. [5]
- Detection of neutrons.

To accomplish these, the detector is constructed from alternating layers of scintillators strips between lead sheets -as a sandwich- in the shape of an equilateral triangle, with a total thickness equal to 16 radiation lengths. The ratio between the

[2] the same angles as those of the Cherenkov Counters.

[3] this is also called *Breaking Radiation* and is shown for electrons when they are in the vicinity of a positively charged nucleus. In this case, the electron change its trajectory leading to the emission of a photon.

[4] the term electromagnetic shower refers when a high-energy electron interacts in a medium, it radiates a Bremsstrahlung photon, which in turn produces an e^+e^- pair. The process of Bremsstrahlung and pair production continues to produce a cascade of photons, electrons, and positrons, referred to as an electromagnetic *shower*. The number of particles in an electromagnetic shower approximately doubles after every radiation length of material traversed.

[5] this is used to the reconstruction of π^0 an η , via the decays $\eta \to \gamma\gamma$, and $\pi^0 \to \gamma\gamma$.

thickness of lead and scintillator is 0.24. In such a configuration $\sim 1/3$ of the total energy of the showering particles is deposited in the scintillators.

Each layer is composed of 36 scintillator strips parallel to one of the sides of triangle, such that the orientation of the strips is rotated by 120° in each successive layer. Thus, we can defined U,V, and W as the three views -each containing 13 layers- see figure 3.7. In order to improved hadron identification, a longitudinal sampling is implemented subdividing the entire detector into inner and outer stacks containing 5 and 8 layers for each view, respectively.

$$\underbrace{\text{Energy deposited in the 13 layers}}_{E_{tot}} = \underbrace{\text{Energy deposited in the 5 innermost layers}}_{E_{in}} + \underbrace{\text{Energy deposited in the 8 outermost layers}}_{E_{out}}$$

The distinction between the two different groups of super-layers are because of the thick of the scintillator. The first 5 super-layers are made of 15 cm thick scintillators and 2.3 thick scintillators lead shits. The 8 outermost super-layers are made of 24 cm thick scintillator and lead shits.

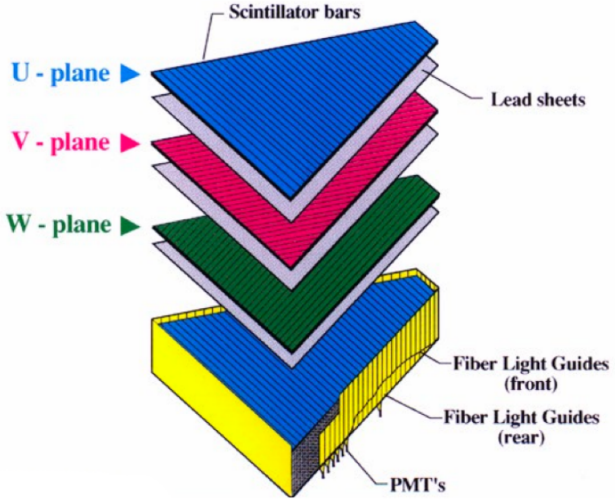


Figure 3.7: Schematic view of one of the EC modules.

The energy resolution of the EC is :

$$\frac{\sigma}{E} < \frac{10\%}{\sqrt{E}} \tag{3.1}$$

where E is in GeV. The position resolution $\delta r \approx 2.3$ cm at 1 GeV, e/π rejection is $\sim 99\%$ at energies greater than 1 GeV.

3.3 EG2 Experiment

The aim of this section is to give a brief description of the experiment. The details can be found in [22] and [30]. The EG2 experiment used two targets exposed to the 5 GeV beam simultaneously. The first target -seen by the beam- is the cryogenic deuterium, referred further to as the *liquid target* (LT). The second target is the *solid target* (ST), which can be chosen to be Carbon (C), Aluminium (Al), Lead (Pb), Iron (Fe), and Tin (Tn); see figure 3.9. The largest data set was accumulated for C, Fe, and Pb; those will be the three targets studied in this thesis.

The advantage of having a double target system is that several systematic effects related to beam and detector properties will cancel in the nuclear ratio. Both targets have the same luminosity ($\mathcal{L}$), this reduced the source of errors in the estimation of the cross sections' ratio.

The length of the cryo-target was 2 [cm]. The heavy target had a form of a circular disk with a radius of 0.15 [cm], see figure 3.8. The thickness of the different targets are in table 3.1. The dimensions of the targets were chosen to have approximately the same number of nuclei along the target length[6].

Thickness of Solid Targets	
Target	Thickness (cm)
C	0.17
Fe	0.04
Pb	0.014

Table 3.1: Thickness of the three solid targets under studied, in cm. The radius is the same for the three.

The separation distance between the two targets is 4 cm. The solid target was placed at $Z = -25$ cm away from the CLAS center ($Z = 0$)[7], while the center of the liquid target was at $Z = -30$ cm. All solid targets were supported in the target system by so-called *Aluminium fingers*. This aspect of the experiment could cause some distortions in the measurement of the observables, so to deal with this issue a detailed simulation of the Eg2 target is needed. The solid and liquid target were separated by an aluminium isolation foil, used for beam tuning.

[6] for lead it was reduced by a factor of two to avoid excessive background.

[7] the reason for positioning the target system backwards with respect to the CLAS center was motivated to increase the acceptance for the negatively charged particles (the CLAS torus polarity in this experiment was such that particles with negative charged were in-bending.).

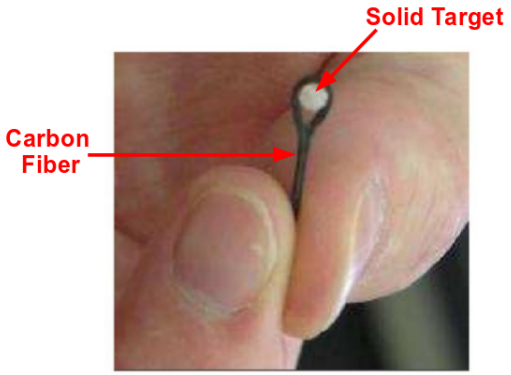


Figure 3.8: A photograph of one of the solid targets surrounded by a carbon fiber loop.

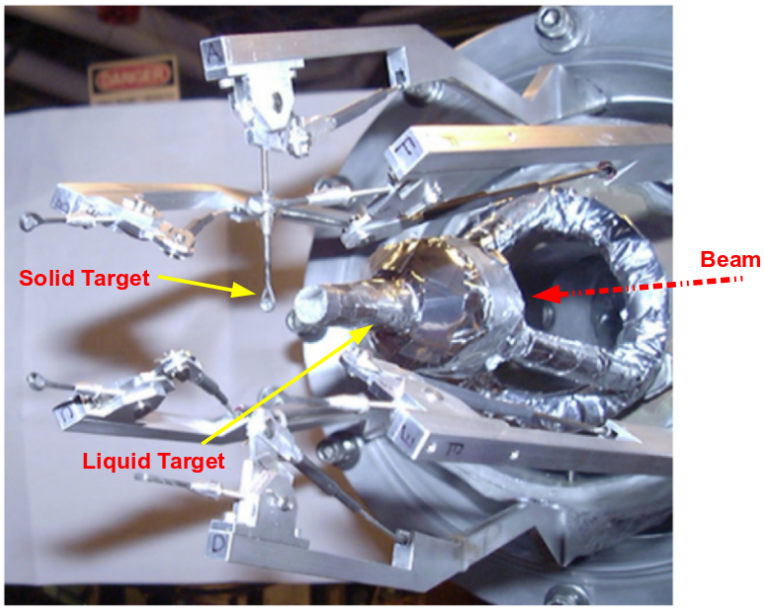


Figure 3.9: Double target assembly. There is one solid target exposed and five solid targets retracted.

The deuterium cryo-target was enclosed in the cell which has aluminium entrance and exit walls. These are called endcaps. The position of the end caps is well determined from the empty target runs [8], and their spatial resolution of z-vertex is the same as for all targets. This endcaps were made of $15\mu m$ aluminium in order to ensure the transmission of the beam, see figure 3.10.

[8] a special run where the cryo-target is full with cold deuterium gas.

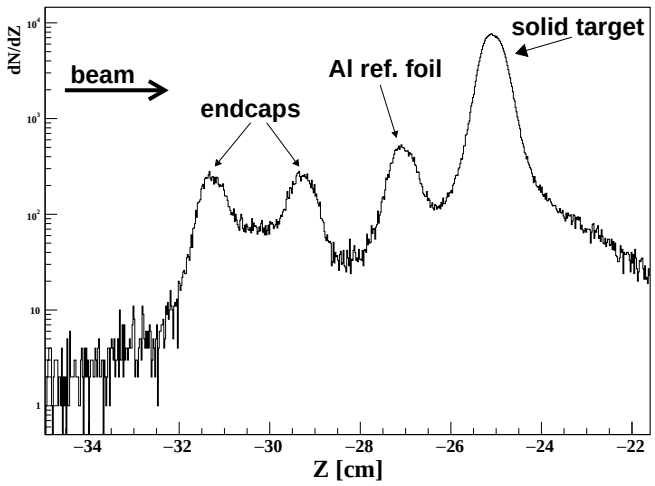


Figure 3.10: One-dimensional Z-vertex distribution, in cm, for electrons for an empty run. The electron beam goes from left to right.

Eg2 experiment was divided into three runs, labelled as a,b, and c depending on the energy of the incoming beam. This thesis was performed based on the data of the third run EG2c, which used an electron beam of 5.014[GeV]. Such electron beam is at the lowest practical energy in which we can study quark propagation. This run took place from January to March of 2004 (50 days).

Among the main experimental requirements for the target, we need a large acceptance for semi inclusive kinematics, approximately equal rates for the two targets simultaneously; less than 2-3% of radiation length of any material to suppress secondary electromagnetic processes; and a rapid target changes for the solid target -minimal mass in support structure-. To accomplish the latter, a support structure was designed to minimize the impact on the quality of the data, but some impact is unavoidable, and we need a high precision comparison of the targets. Recognizing this problem is necessary to implement a highly detailed simulation of the system into a GEANT simulation package for the CLAS detector, called GSIM -see figure 3.11 as a reference for the Eg2 real target assembly and the simulated picture of it, more on this in section 4.7.

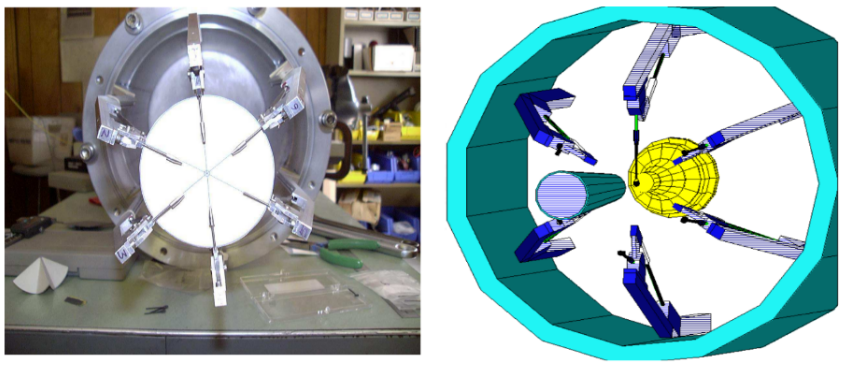


Figure 3.11: On the left, a photograph of the EG2 target during the assembling process, on the right the EG2 target input in GSIM.

Chapter 4

Data Analysis

4.1 Introduction

In order to extract observables from the data, the first thing we need is to identify the particles. In this thesis, the particles of interest are electrons and positive pions.

The way to select the particles is through a series of *cuts* ,i.e, to impose certain selection criteria on the data set in order to reject events that are not of our interest. These cuts are implemented on the basis of information stored in data *banks*.

The first part of this chapter aims to give a brief description of the data set features, such as the process of which the information is stored.

In the next section of this chapter is shown the method for particle selection using a series of cuts on the various detector outputs. The way of presenting the information is treating each cut separately by explaining the need for such cut and the impact in the particles candidates number. This is performed for electrons and then, for pions.

Once the particles are fully identified, we can extract the observables [1], that will be the main tool to test our results. Also some additional cuts are discussed, those are not PID cuts. However they are useful to clean the SIDIS sample.

After that, the Monte Carlo (MC) simulation set used in this thesis is introduced[2] with its own PID. Also some features of the simulations files are discussed. The final

[1] In this thesis the observables are the EMC ratio for inclusive DIS electrons and the Multiplicity Ratio for semi-inclusive DIS hadrons.

[2] A Monte-Carlo simulation is basically like running an experiment in your computer. The basic steps are to simulate the physics reaction with model inputs (target/reaction point). Then, simulate how particles travel in a EM field (spectrometer/solenoid). Next, simulate how detectors respond to particles (detectors). After that, simulate how detectors signals converted into different electric signals. Finally, reconstruct physics quantities after building in *actual effects*, such as energy-loss, detector resolution, background and so on. All the event generators split the simulation up into the same phases: Hard process → parton shower → secondary decays → hadronization → multiple scattering/soft underlying event → hadron decays.

section is dedicated to the applied corrections.

4.2 Data Taking

In order to extract the relevant information, the data acquired from different detectors (EC, SC, CC and DC) are stored in data banks[3], where the information is obtained. The reconstruction process[4] is very complex; the data reconstruction is performed with RECSYS. The *cooking* used in this thesis was performed by Taisiya Mineeva and is known as *pass2* [5]. The information extracted is stored in a database called BOS (**B**ank **O**bject **S**ystem).

The way the final data is organized is in term of *events*. By event we mean that an electron -the scattered electron- passed the triggering threshold, followed by all the possible particles reconstructed.

Among all the banks, the EVNT bank is special. The integration of all the information available from the detectors is called SEB (**S**imple **E**vent **B**uilder), and this one produces the EVNT bank. This contains the four momentum of the reconstructed particles and pointers to the tracks left in all the detectors.

To read the information the ROOT framework is used. The files are in a special format called *ClasTool* (created by Maurick Holtrop and Gagik Gavalyan) which is a ROOT based package for analysing CLAS data. It consists of a C++ classes' set in a number of libraries which you can use to build your own CLAS analysis program. The information is stored inside TTree objects with links between different banks included as pointers.

4.3 Particle Identification

In this section we discuss the selection criteria used in this analysis. The whole set of cuts can be, basically, divided into three main categories:

- Cuts that involve the correct signal in each detector (DC,EC,SC and CC). These are *data banks* cuts.
- Cuts aiming to reject some contamination from the sample.

[3] units of information corresponding to certain detector or algorithm output.

[4] in CLAS, the reconstruction process is called the "cooking".

[5] this was a recooking process of EG2 data-set that took place in 2009. The original cooking and calibration procedures was performed in 2005 by Lamiaa El Fassi and Lorenzo Zana.

- Cuts that consider the regions where the detectors have efficiency problems, so-called *fiducial cuts*.

This section is dedicated to explain in detail all the cuts and motivations for them.

4.3.1 Electron Identification

For identifying the scattered electrons, the first track in the EVNT bank is considered[6]. The general status must be positive; this is for rejecting particles that passed the Hit Base Tracking (HBT), but failed the Time Based Tracking (TBT). Next, the particle must leave a track in all detectors. This means that the number of rows in the DCPB, CCPB, SCPB, and ECPB banks must be different than zero[7].

Those will be all the cuts involving the data banks. Additionally, the charge[8] of the electron candidate must be negative. A cut in the momentum of the electron is necessary, $P > 0.75$ GeV. This cut is, at our energies, equivalent to the $Y_b < 0.85$.

Now, let's see one by one the remaining cuts.

4.3.1.1 Number of Photo-electrons Cut

Until now, one big concern is that we have selected possibly a lot of π^- in our electron candidates sample. The standard way to reduced pion contamination is using the information on CCPB, the Cherenkov Counter Particle Bank; and see the distribution of the number of photo-electrons (Nphe) collected [9]. A big peak at low values of the number of photo-electrons is observed. This peak is commonly refer as *single photo-electron peak* which is located at 1-2 photo-electrons. This peak is not expected to be produced by electrons crossing CC within the fiducial volume. It is worth noting that the energy fraction (E_{tot}/P) of this peak has much smaller released energy fraction than expected for electrons -~ 0.3 for proper electrons-. Therefore, we can conclude that the single photo-electron peak is produced by the tail of minimum ionizing particles. Those are low energy π^- which produce a knock on electron (δ-rays.)

[6]This is because the cooking procedure put the particle that most likely have triggered the event in the first position.

[7]There are two extra cuts regarding the status in the banks. The first one is that the general goodness status for the particle should be less than 100 -in addition of being bigger than zero- and than the status for the Scintillator Counter (SCPB) must be equal to 33. These two cuts, do not make any significant difference; they are a "just in case" cuts.

[8]In the CLAS reconstruction software, the electric charge is determined based on the bending direction of the tracks in drift chambers (DC).

[9]it is very common to present the Nphe$\times$10 distribution instead of just Nphe, for readability.

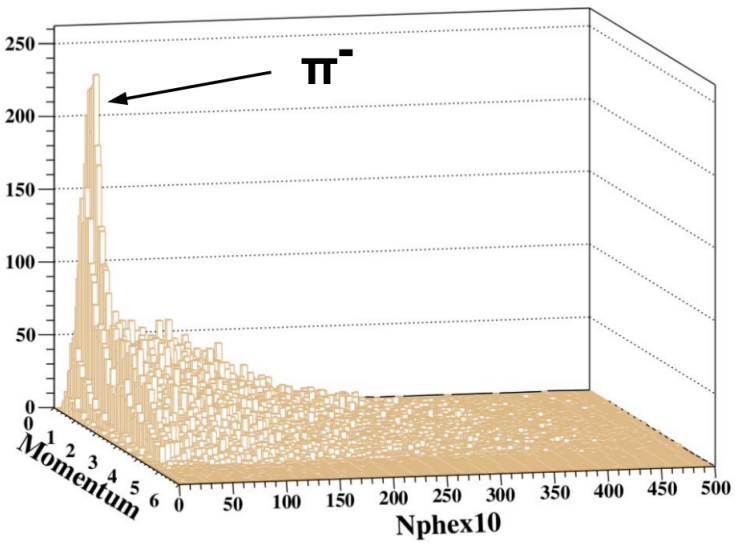


Figure 4.1: Number of photo-electrons ($\times 10$) as a function of momentum, in GeV. The distribution of Nphe represents a Poisson distribution centered around 8-10 photo-electrons. The pions give a signal centered around a few photo-electrons.

Mainly the most contaminated region is the one with momentum below 2 GeV, see figure 4.1. The reason for this is that the number of photo-electrons emitted increases with the velocity of the particle. For this reason, we expect that the electron should be at a very high speed, near the speed of light, so should not produce low numbers of photo-electrons. The π^-, being slower, produce fewer Nphe than electrons. To get rid of the peak we make a Nphe sector dependent cut, see figure 4.2.

4.3.1.2 Cut for Coincidence Time between SC and EC

Looking at figure 3.2, the scintillators are located radially outside the tracking system in front of the calorimeter. The electron first encounters SC and then EC. If a particle is accidentally misidentified as an electron, it may produce a signal in either EC or SC. Those events can be eliminated if we impose a cut in the time elapsed between SC and EC. This is a coincidence time, where we define ΔT as:

$$\Delta \mathrm{T} = (\mathrm{Time(EC)} - \mathrm{Time(SC)}) - \left(\frac{\mathrm{Path(EC)}}{\mathrm{Vel}} - \frac{\mathrm{Path(SC)}}{\mathrm{Vel}} \right) \tag{4.1}$$

the cut is $|\Delta \mathrm{T}| < 5 \times 0.35$, where this 0.35 is the σ (width of a Gaussian distribution) that comes from a fitting of the distribution of the electrons with all the cuts imposed.

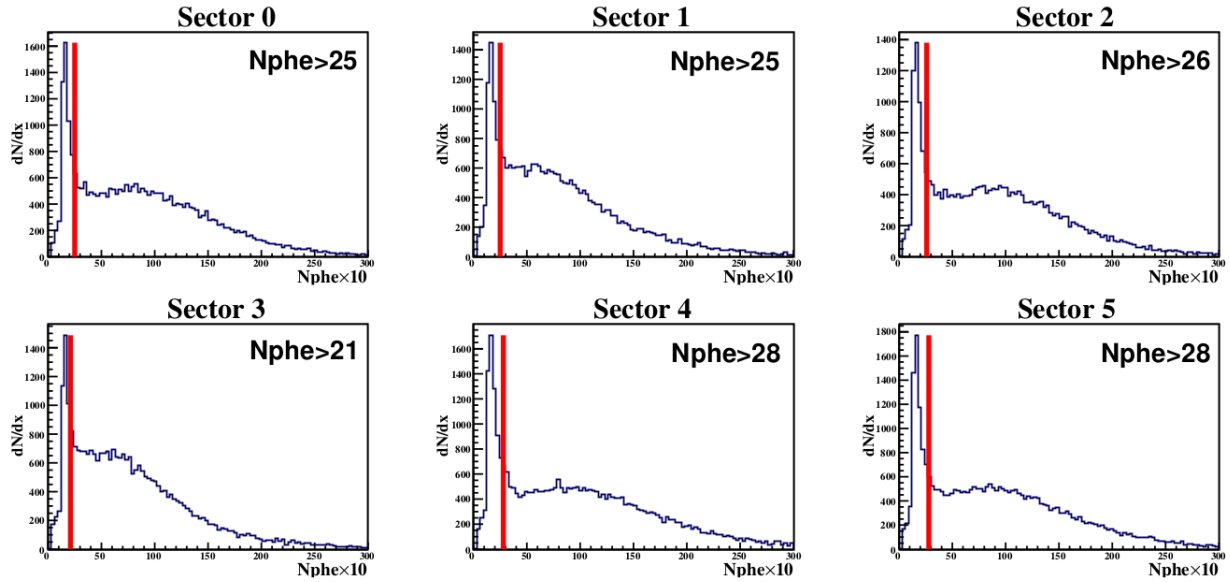


Figure 4.2: Number of photo-electrons distribution (×10) for each CLAS sector. The red lines correspond to the cut applied to reject π^- contamination.

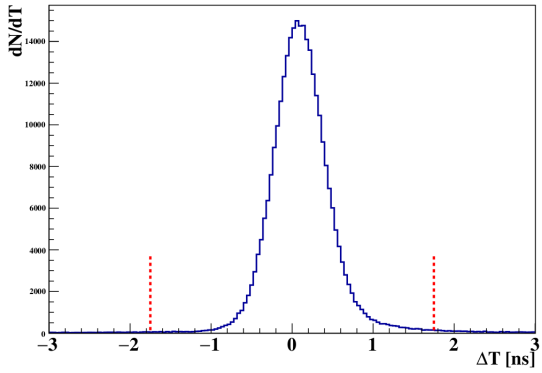


Figure 4.3: Cut in the ΔT variable, in nanoseconds. The dashed red lines are the cuts, all the particles outside this limits are rejected.

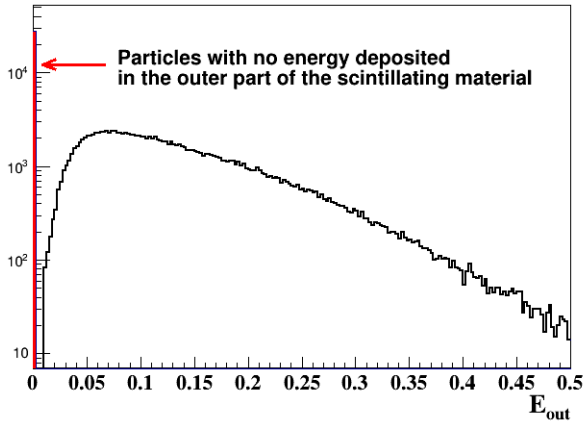


Figure 4.4: Cut in the energy deposited in the outer part of the EC.

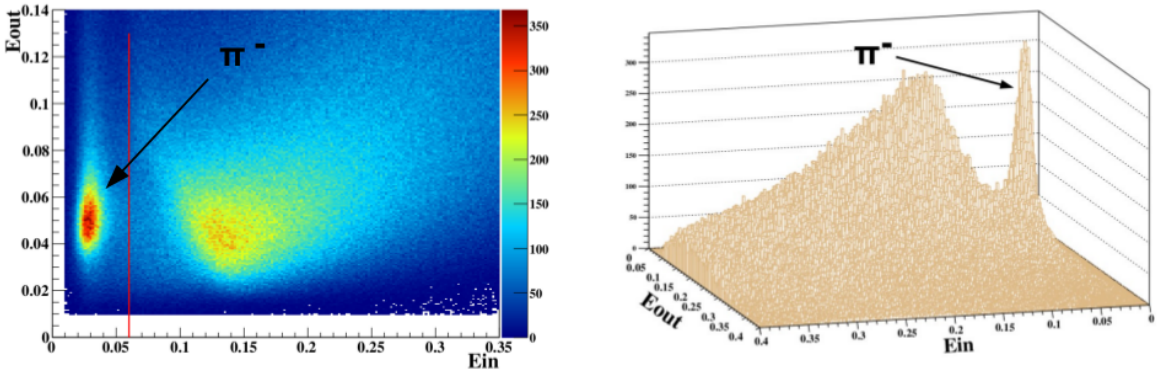


Figure 4.5: Energy deposited in the inner and outer layers of the calorimeter.

4.3.1.3 Cut for Energy Deposited in EC

To suppress negative pion contamination, a set of cuts in the energy deposited in the EC is imposed. First of all, energy deposition cut on EC outer layer is applied, demanding that is not null. See figure 4.4.

The scattered electron always travels in a forward direction while pions can have any direction. Since EC only covers forward angles, any track not producing a signal in the EC is immediately rejected. The pion contamination is visible as a sharp peak in the energy deposited in the inner part of the detector below some threshold value, see figure 4.5. A cut in the energy deposited in the inner part of the calorimeter -made of the first 5 super layers and the thick lead shit- is imposed, 60 MeV. This cut is useful because the pions are minimum ionizing, in other words, there will deposit a fixed amount of energy regardless of its momentum.

In figure 4.6 it can be seen , for each sector, the distribution of the total energy

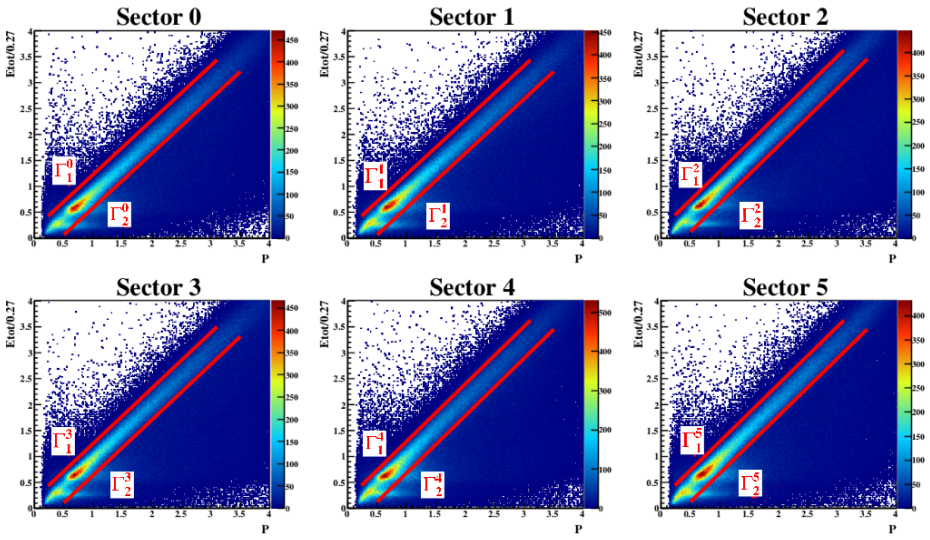


Figure 4.6: Cuts in the total energy deposited in the EC and the momentum of the particle. Particles outside the region between the lines are rejected.

deposited in the EC, normalized by the momentum of the track, as a function of the particle's momentum, for all the electron candidates with only the basic cuts in data-banks. The cuts are depicted with red lines, denoted as Γ_1^s and Γ_2^s, where the superscript s stands for different CLAS sectors. From equations 4.2 to 4.7 one can see the cuts applied. Only the particles within this two limits remain as electron candidates.

$$\Gamma_1^0: \quad \mathrm{E}'_{\mathrm{tot}} < 1.05 \times \mathrm{P} + 0.18 \quad ; \quad \Gamma_2^0: \quad \mathrm{E}'_{\mathrm{tot}} > 1.05 \times \mathrm{P} - 0.46 \tag{4.2}$$

$$\Gamma_1^1: \quad \mathrm{E}'_{\mathrm{tot}} < 1.05 \times \mathrm{P} + 0.18 \quad ; \quad \Gamma_2^1: \quad \mathrm{E}'_{\mathrm{tot}} > 1.05 \times \mathrm{P} - 0.46 \tag{4.3}$$

$$\Gamma_1^2: \quad \mathrm{E}'_{\mathrm{tot}} < 1.11 \times \mathrm{P} + 0.18 \quad ; \quad \Gamma_2^2: \quad \mathrm{E}'_{\mathrm{tot}} > 1.11 \times \mathrm{P} - 0.43 \tag{4.4}$$

$$\Gamma_1^3: \quad \mathrm{E}'_{\mathrm{tot}} < 1.07 \times \mathrm{P} + 0.18 \quad ; \quad \Gamma_2^3: \quad \mathrm{E}'_{\mathrm{tot}} > 1.07 \times \mathrm{P} - 0.43 \tag{4.5}$$

$$\Gamma_1^4: \quad \mathrm{E}'_{\mathrm{tot}} < 1.11 \times \mathrm{P} + 0.18 \quad ; \quad \Gamma_2^4: \quad \mathrm{E}'_{\mathrm{tot}} > 1.11 \times \mathrm{P} - 0.43 \tag{4.6}$$

$$\Gamma_1^5: \quad \mathrm{E}'_{\mathrm{tot}} < 1.11 \times \mathrm{P} + 0.18 \quad ; \quad \Gamma_2^5: \quad \mathrm{E}'_{\mathrm{tot}} > 1.11 \times \mathrm{P} - 0.43 \tag{4.7}$$

where $\mathrm{E}'_{\mathrm{tot}} = \frac{\mathrm{E}_{\mathrm{tot}}}{0.27}$

A set of sector dependent E_{in} and E_{out} cuts are shown in equations from 4.8 to 4.13. See figure 4.7 for the actual form of the cuts and the effect on the electron candidates.

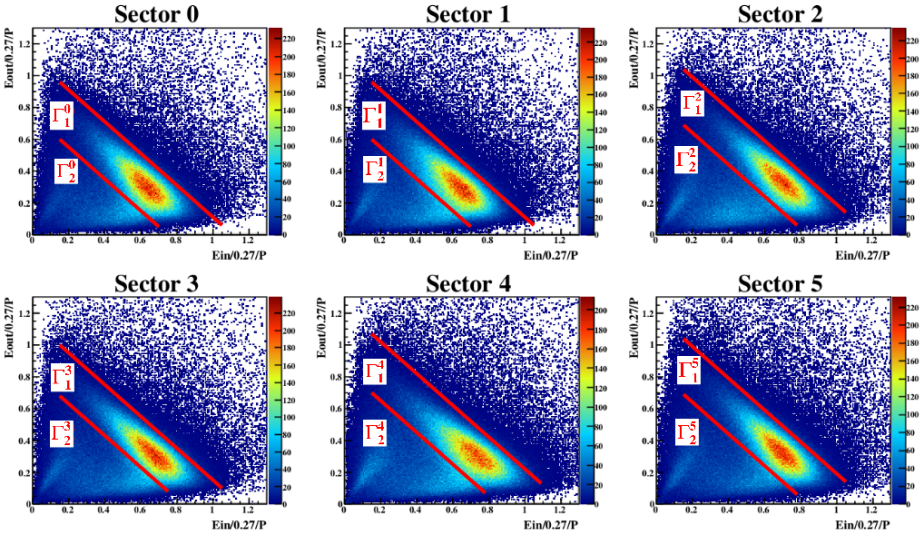


Figure 4.7: Cuts in the inner/outer energy deposited in the EC. Particles outside the region between the lines are rejected.

$$\Gamma_1^0: \quad E'_{in} + E'_{out} < 1.11 \times P \quad ; \quad \Gamma_2^0: \quad E'_{in} + E'_{out} > 0.75 \times P \tag{4.8}$$

$$\Gamma_1^1: \quad E'_{in} + E'_{out} < 1.11 \times P \quad ; \quad \Gamma_2^1: \quad E'_{in} + E'_{out} > 0.75 \times P \tag{4.9}$$

$$\Gamma_1^2: \quad E'_{in} + E'_{out} < 1.19 \times P \quad ; \quad \Gamma_2^2: \quad E'_{in} + E'_{out} > 0.84 \times P \tag{4.10}$$

$$\Gamma_1^3: \quad E'_{in} + E'_{out} < 1.15 \times P \quad ; \quad \Gamma_2^3: \quad E'_{in} + E'_{out} > 0.83 \times P \tag{4.11}$$

$$\Gamma_1^4: \quad E'_{in} + E'_{out} < 1.22 \times P \quad ; \quad \Gamma_2^4: \quad E'_{in} + E'_{out} > 0.85 \times P \tag{4.12}$$

$$\Gamma_1^5: \quad E'_{in} + E'_{out} < 1.19 \times P \quad ; \quad \Gamma_2^5: \quad E'_{in} + E'_{out} > 0.84 \times P \tag{4.13}$$

where $E'_{in/out} = \frac{E_{in/out}}{0.27}$

4.3.1.4 Electron Fiducial Cuts on DC

We apply a fiducial cut on the DC measurements following the procedure described in detail in Ref. [23]. As shown in figure 4.8, the two-dimensional phase space of the laboratory angles θ and ϕ for different P ranges presents some regions of low efficiency near the edges, which are removed after the fiducial cut, see figure 4.9.

4.3.1.5 Electron Fiducial Cuts on EC

The coordinates perpendicular to the borders of EC are labeled U, V, and W. A set of geometrical cuts are imposed in order to define a region of uniform efficiency (see

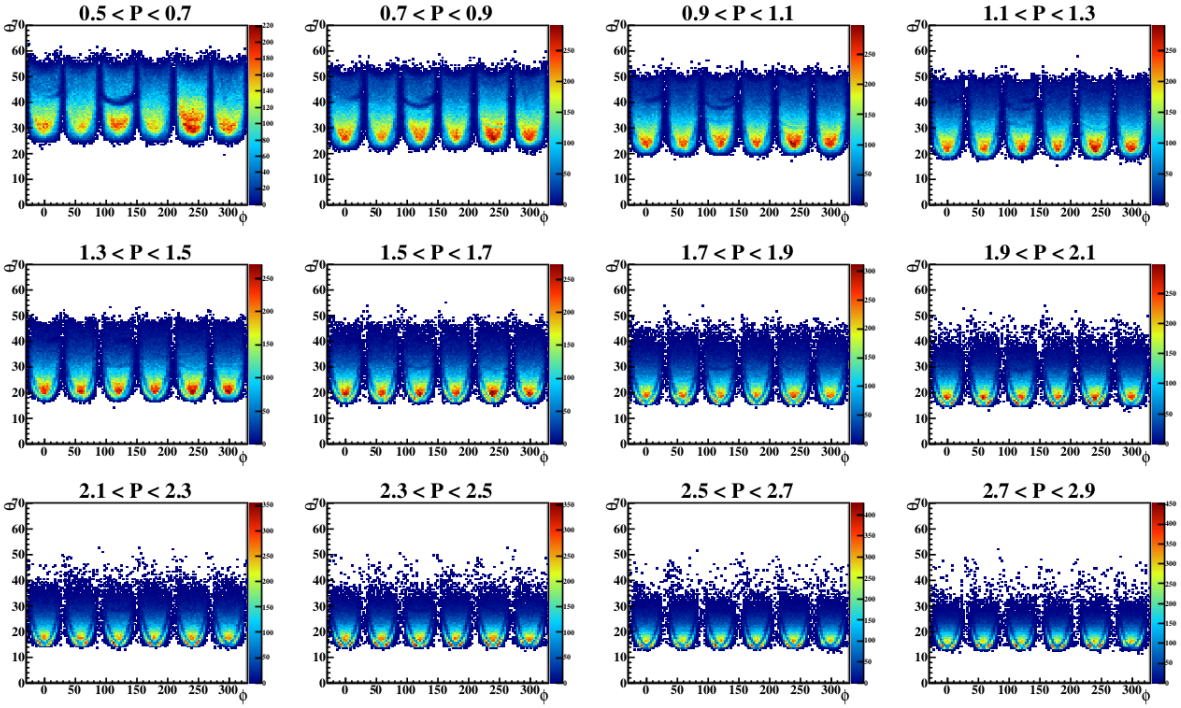


Figure 4.8: ϕ and θ laboratory angles phase space for different momentum ranges, for all electron candidates. The six distinct regions correspond to the different CLAS sectors.

Figure 4.10):

$$40 < U < 410 \text{ cm} \tag{4.14}$$

$$V < 370 \text{ cm} \tag{4.15}$$

$$W < 405 \text{ cm} \tag{4.16}$$

Another useful way to see the effect of this cuts is to focus on the distributions of the X and Y coordinates of EC, and see the distribution before and after the application of the cut, see figure 4.11. All the events in gray are rejected.

4.3.1.6 Sampling Fraction Cut

Figure 4.12 shows the measured energy-to-momentum ratio for electron candidates, which is not entirely constant as a function of momentum, this occurs because the attenuation of the light collected and the EC sampling of the shower. We use a sampling fraction cut that is described in detail in the EG2 π^0 Analysis Note [60]. The procedure entails fitting the mean value $\mu(P)$ and the width $\sigma(P)$ of the E_{tot}/P distribution for each P bin with the following forms:

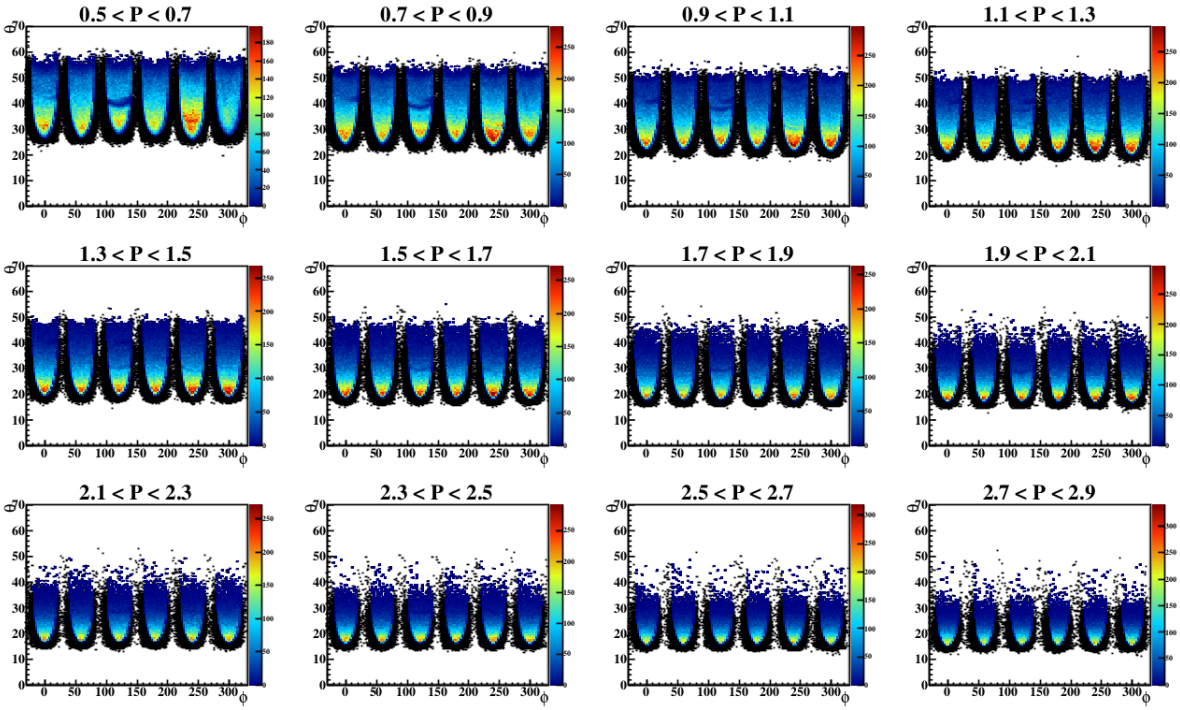


Figure 4.9: ϕ and θ laboratory angles phase space for different momentum ranges after the application of the DC fiducial cuts. The region in black are the particles that were rejected due to this cut.

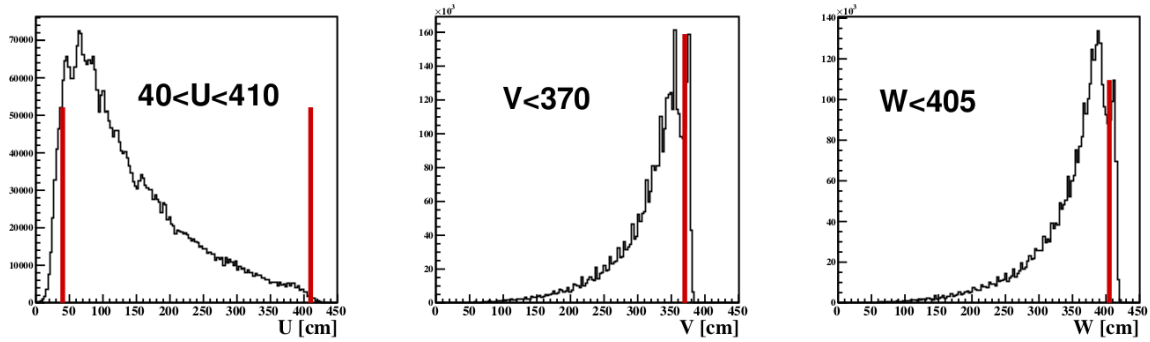


Figure 4.10: U,V, and W distributions for the electron candidates with the basic cuts (banks and charge.) The region outside the red line is rejected.

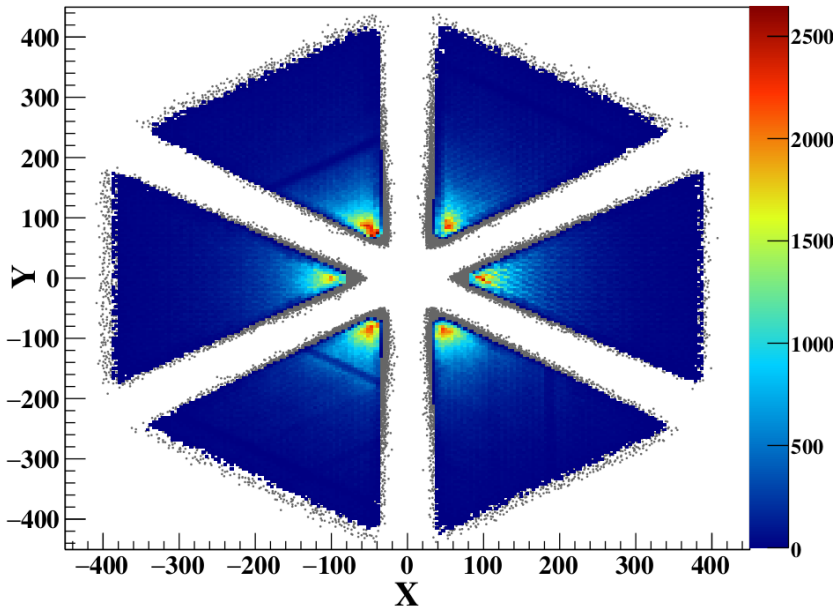


Figure 4.11: X and Y coordinates, in [cm], for all electron candidates with basic cuts in banks and charge (in gray), and over imposed the case with the fiducial cut in EC included.

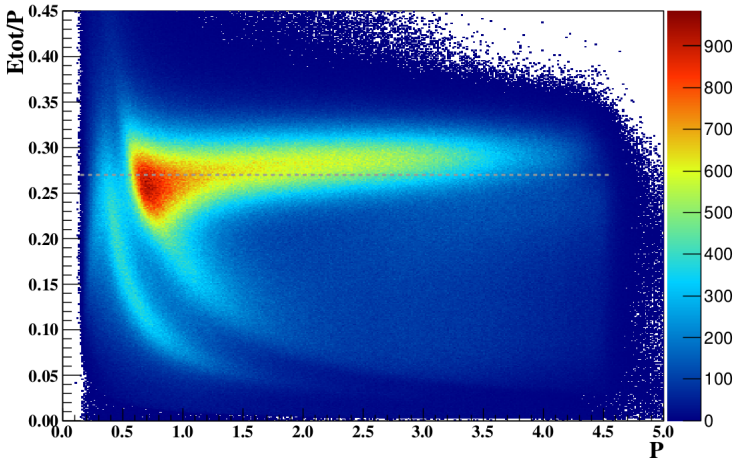


Figure 4.12: Total energy deposited in the calorimeter divided by the momentum as a function of the momentum (in GeV.)

Sampling Fraction Coefficients for **carbon**						
Coef	Sector 0	Sector 1	Sector 2	Sector 3	Sector 4	Sector 5
a_1	0.252164	0.278574	0.262079	0.251108	0.263396	0.255245
a_2	0.0122263	0.0187482	0.0230685	0.0201568	0.00955238	0.0232659
a_3	-7.939e-04	-0.0023821	-0.0035474	-0.0033236	-0.0010203	-0.0030479
a_4	9.5511e-03	1.3988e-02	9.3276e-03	8.2105e-03	2.2568e-02	1.1725e-02
a_5	3.4067e-02	3.7468e-02	2.9004e-02	2.9889e-02	3.0650e-02	3.6422e-02

Table 4.1: Parameters extracted from fit on carbon data.

Sampling Fraction Coefficients for **iron**						
Coef	Sector 0	Sector 1	Sector 2	Sector 3	Sector 4	Sector 5
a_1	0.222404	0.234623	0.252287	0.250946	0.271956	0.252613
a_2	0.0222688	0.0194985	0.024248	0.0208409	0.0118487	0.022819
a_3	-0.0024153	-0.0020835	-0.0033884	-0.0032682	-0.0018708	-0.0031124
a_4	9.2302e-03	8.6636e-03	1.0782e-02	7.2258e-03	1.8407e-02	4.1146e-03
a_5	2.9834e-02	3.0885e-02	2.6385e-02	2.9880e-02	3.4802e-02	3.5508e-02

Table 4.2: Parameters extracted from fit on iron data.

$$\mu(P) = a_1 + a_2 \times P + a_3 \times P^2 \tag{4.17}$$

$$\sigma(P) = \sqrt{a_4^2 + \frac{a_5^2}{P}} \tag{4.18}$$

where $a_i \to a_i(\text{sector})$, for $i = 1, ..., 5$. The value of the energy used here is:

$$\mathrm{E} = \mathrm{Max}\left(\mathrm{E_{tot}}, \mathrm{E_{in}} + \mathrm{E_{out}}\right) \tag{4.19}$$

With this, the final cut is:

$$\left|\frac{E}{P} - \mu\right| < 2.5 \times \sigma \tag{4.20}$$

The values of the parameters a_i are shown in tables: 4.1, 4.2, and 4.3, for each target separately.

In figure 4.13 are the distributions of E/P vs P for each CLAS sector. The black line is the second order polynomial fit to the mean value, and red lines are the cuts on the sampling fraction corresponding to $\pm 2.5\sigma$.

To have an idea about the effect of each cut of the electrons selection, in table 4.4 are shown , in percentage, the remaining fraction of electron candidates after each cut for each sector. The DIS cuts are included already (see below). At the end of the table, there are the final percentage of particles that are called truly electrons. It varies between sectors, but it is roughly of order of 18%.

Sampling Fraction Coefficients for **lead**						
Coef	Sector 0	Sector 1	Sector 2	Sector 3	Sector 4	Sector 5
a_1	0.253431	0.249059	0.254573	0.255589	0.276739	0.262587
a_2	0.0138251	0.0147784	0.022589	0.0190419	0.0111585	0.0191659
a_3	-0.001401	-0.0014869	-0.0030568	-0.0030526	-0.0017578	-0.002626
a_4	7.6740e-03	7.5279e-03	8.1324e-03	7.2030e-03	1.8084e-02	1.9922e-03
a_5	3.5439e-02	3.3837e-02	2.7730e-02	3.0362e-02	3.5302e-02	3.7617e-02

Table 4.3: Parameters extracted from fit on lead data.

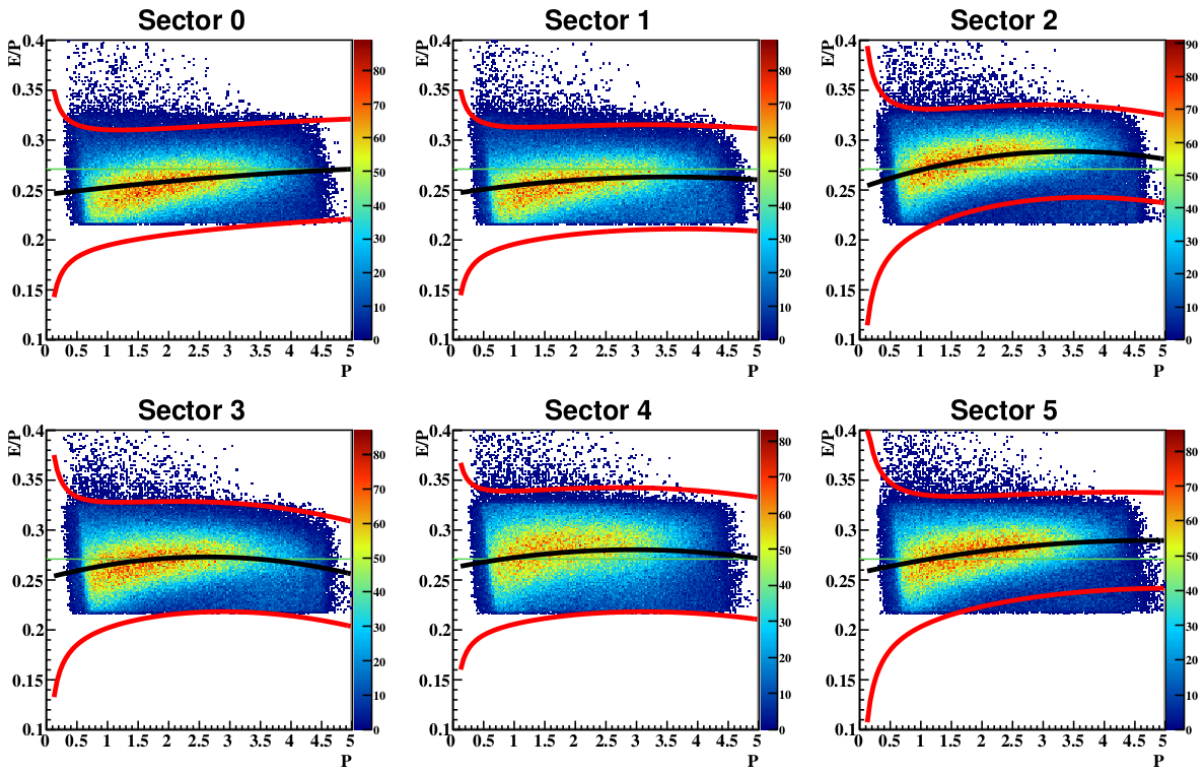


Figure 4.13: Sampling Fraction cuts. The region outside the red lines is rejected. The black line is the $\mu(P)$ given in equation 4.17.

Fraction of remaining particles in Electron Identification							
N^o	Cuts	Sector 0	Sector 1	Sector 2	Sector 3	Sector 4	Sector 5
0	No cuts	100	100	100	100	100	100
1	numberDC $\neq 0$	100	100	100	100	100	100
2	numberEC $\neq 0$	98.9597	98.8295	98.8938	98.805	98.8937	99.0965
3	numberSC $\neq 0$	98.909	98.7727	98.7976	98.7555	98.7572	99.0333
4	numberCC $\neq 0$	92.9599	92.7536	93.2418	91.9092	93.2033	93.4922
5	StatCC > 0	65.9483	62.3308	65.7212	61.5765	68.3371	66.6946
6	StatSC > 0	65.517	61.8882	65.0893	61.1589	67.6612	66.1937
7	StatDC > 0	65.517	61.8882	65.0893	61.1589	67.6612	66.1937
8	StatEC > 0	60.8397	57.4792	60.7786	56.9787	62.9669	61.5808
9	DCStatus > 0	57.1877	53.9531	56.8513	53.5676	59.0894	57.8291
10	SCStatus $= 33$	56.6115	52.0205	54.3027	53.1451	57.7305	57.1909
11	$0 <$ Status <100	56.6093	52.019	54.3018	53.1444	57.7305	57.1906
12	Charge=-1	44.721	41.4911	41.7353	42.7279	45.2292	44.0111
13	Ein$>$ 0.06	43.1521	40.3593	40.7783	41.9184	43.9294	43.0488
14	Nphe cut	34.3475	31.5955	31.4081	34.7273	33.3572	32.6459
15	ΔT	32.4433	29.0224	29.5019	32.8374	31.2609	30.8146
16	(Etot,P)	25.8467	22.8732	23.4116	25.7639	22.8204	23.9848
17	(Ein/Eout,P)	22.8954	20.4548	21.3617	22.5101	19.7545	20.8466
18	Eout $\neq$0	21.8388	19.7001	20.8035	21.8493	18.8527	20.2465
19	Fiducial (DC)	18.8133	16.7906	17.7869	18.2882	16.1003	18.2319
20	Fiducial (EC)	18.8079	16.7835	17.7843	18.2835	16.0951	18.2197
21	Sampling Fraction	18.8027	16.7826	17.7823	18.2835	16.0951	18.2197

Table 4.4: Fraction of remaining particles after the incremental application of each cut, for the electron selection.

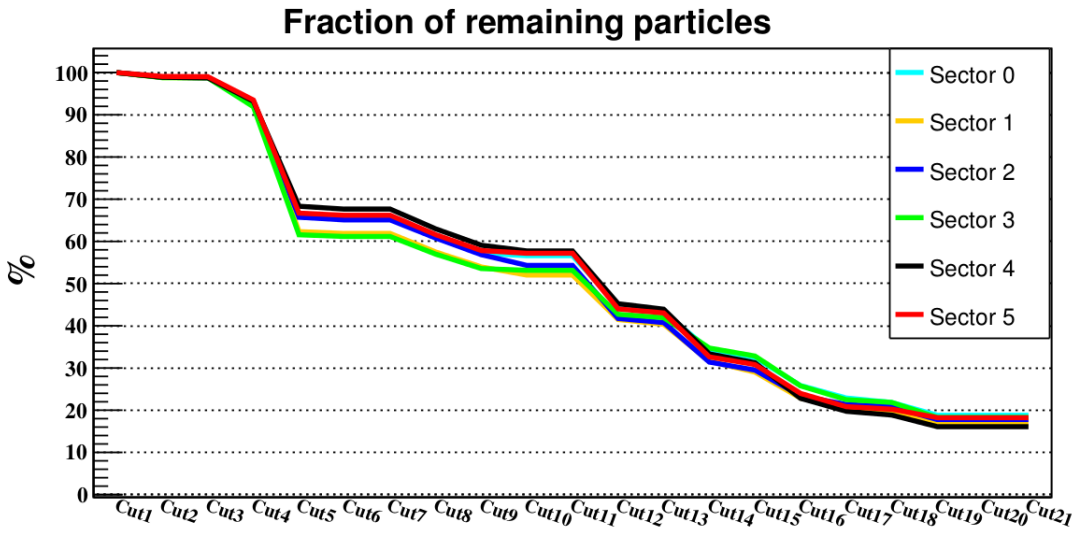


Figure 4.14: Fraction of remaining events in the electrons selection as the different cuts are imposed in an incremental way.

where Ein, Eout, and Etot are normalized by a factor of 0.27. These numbers were calculated for iron target.

Figure 4.14 shows a graphical representation of the content of table 4.4, where at the rightmost part of the plot are the fully select electrons with all the cuts applied. The final fraction is similar in all six sectors.

Since many of the cuts are not independent, seeing the effect in an incremental manner hides the effect of each cut separately. It is also useful to have an idea of the percentage of rejected electron candidates when we apply this cuts separately. In table 4.5, there are all cuts similar to the previous table, but this time with the fraction of remaining events after the application of each cut. For a graphical representation of the content of table 4.5, see figure 4.15.

4.3.2 DIS Kinematics

We select events with $Q^2 > 1$ GeV^2 (range of virtualities to resolve a parton), $W > 2$ GeV (to avoid the resonance region) and $y < 0.85$ (to avoid large contribution from radiative processes). Here W is the invariant mass of the photon-nucleon system; $y = \nu/E$ is the energy fraction of the virtual photon; and E is the electron-beam energy. Figure 4.16 shows the effect of these cuts on the phase space of (Q^2, X_b). Finally, figure 4.17 shows the final distributions of the electrons' E_{tot}/P vs P and E_{out}/P vs E_{in}/P after all the cuts previously mentioned, including the DIS cuts.

Fraction of rejected events in Electron Identification							
N^o	Cuts	Sector 0	Sector 1	Sector 2	Sector 3	Sector 4	Sector 5
1	numberDC $\neq 0$	0	0	0	0	0	0
2	numberEC $\neq 0$	1.04028	1.17055	1.10624	1.19499	1.10633	0.903454
3	numberSC $\neq 0$	0.0507453	0.0567779	0.0961523	0.0494762	0.136519	0.0632804
4	numberCC $\neq 0$	6.33513	6.64639	6.15569	7.55222	6.11531	5.93737
5	StatCC > 0	33.4508	37.2012	33.8032	38.0141	31.0931	32.817
6	StatSC > 0	0.629498	0.672434	0.948573	0.638464	0.984268	0.713019
7	StatDC > 0	0	0	0	0	0	0
8	StatEC > 0	18.0933	19.5715	17.3346	18.8558	17.5696	17.5994
9	DCStatus > 0	8.2429	8.11863	8.36072	8.31327	8.38089	8.36728
10	SCStatus $= 33$	3.2856	6.03841	6.1272	4.20264	5.1938	4.07698
11	$0 <$ Status <100	8.3097	8.16497	8.38241	8.31768	8.38176	8.36876
12	Charge=-1	34.0382	36.9259	36.0322	35.4956	34.0708	35.4338
13	Ein$>$ 0.06	25.9938	25.0354	20.4403	23.1416	21.9	22.2851
14	Nphe cut	55.7241	58.7172	57.5625	55.3721	56.3373	57.9441
15	ΔT	29.0536	34.5492	31.0646	28.6836	30.1681	29.3826
16	(Etot,P)	59.8769	62.0101	61.9043	61.3594	63.3232	62.1898
17	(Ein/Eout,P)	67.1081	68.7071	68.3831	69.383	70.8011	69.7433
18	Eout $\neq$0	18.9543	16.9505	16.4009	16.8471	19.7736	18.6232
19	Fiducial (DC)	36.4544	36.8805	37.0688	35.8019	35.2101	32.5151
20	Fiducial (EC)	26.6403	28.7167	23.7509	27.4681	26.0445	26.1247
21	Sampling Fraction	57.8281	61.2612	62.5459	61.5103	58.6951	59.1966

Table 4.5: Effect in percentage of each cut in terms of rejected particles for all different CLAS sectors.

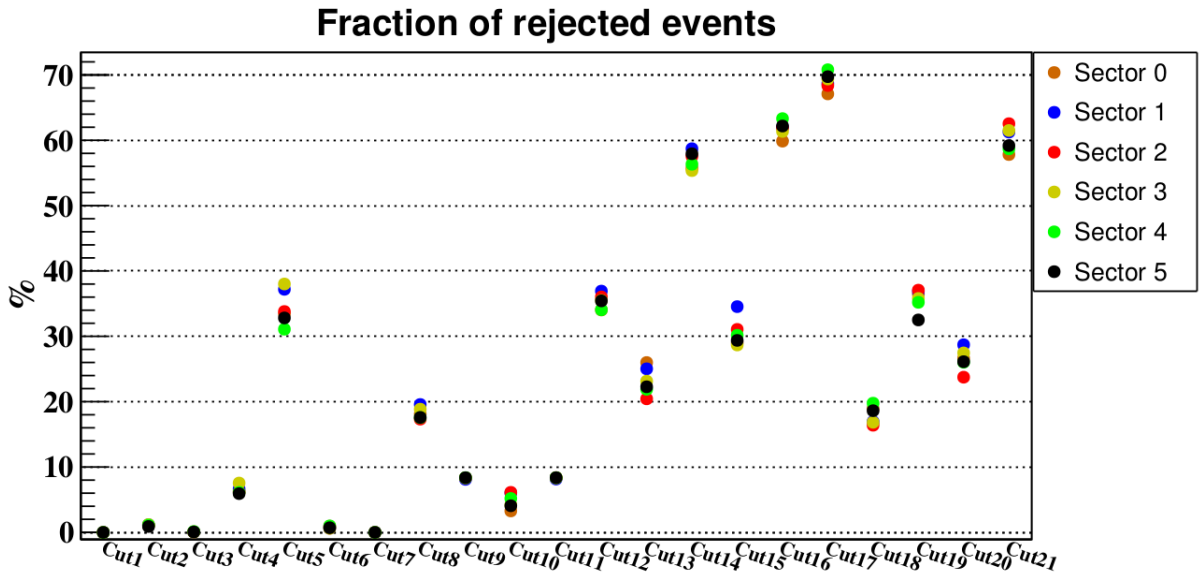


Figure 4.15: Fraction of rejected particles for each individual cut in the electron selection.

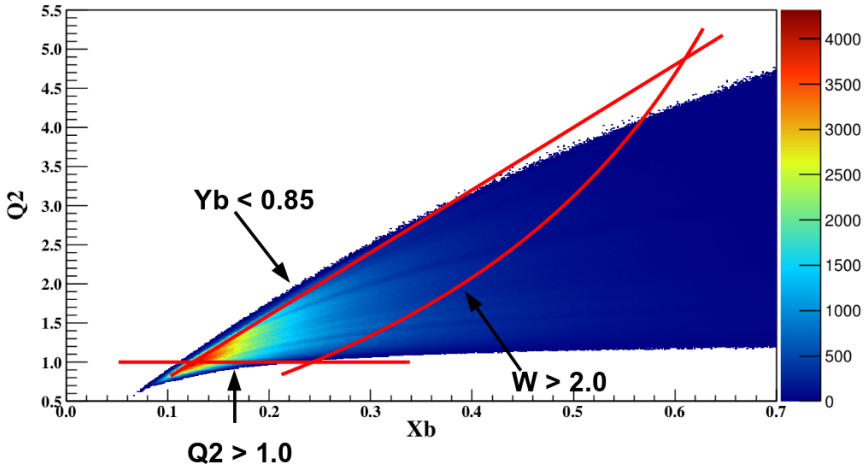


Figure 4.16: Depiction of the set of DIS cuts on the (X_b,Q^2) phase space for inclusive electron events ($ep \longrightarrow eX$). The cuts are shown with red lines.

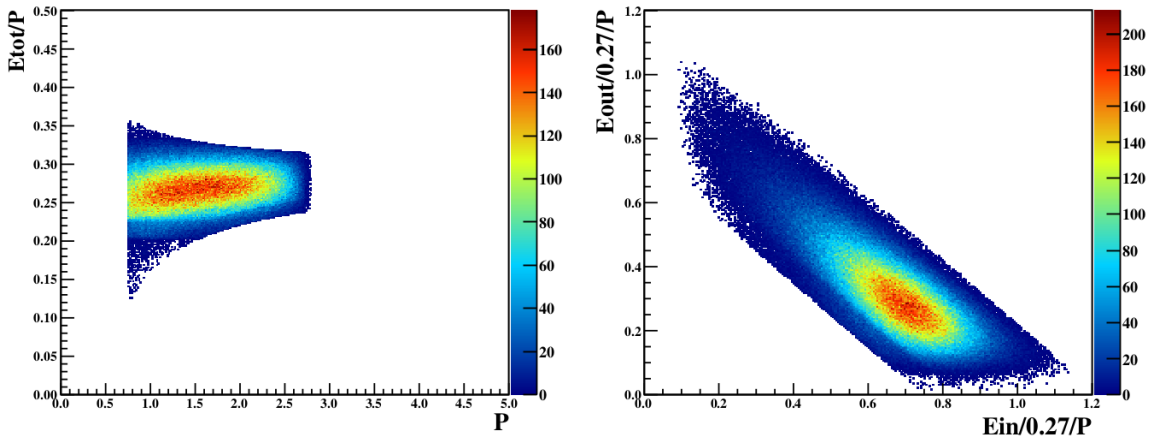


Figure 4.17: Final distributions for electrons candidates that passed all the particle identification cuts. The DIS cuts are applied here as well.

4.3.3 Positive Pion Identification

For hadron identification the general procedure will be analogous to the electron case. Let's start with the cuts associated with the data banks. The general status must be bigger than zero to reject particles that passed the HBT, but failed the TBT. A positive hit is necessary, so the charge of the particles are the one of the π^+.

In order to reject contamination, two-dimensional (ΔT, P) distribution is used, where ΔT [10] is a pure TOF variable [11] defined as:

$$\Delta \mathrm{T} \equiv \left(\frac{L^{e^-}_{flight}}{c} - \frac{L^{\pi}_{flight}}{v_\pi} \right) - (t_{e^-} - t_\pi) \tag{4.21}$$

where:

- $L^{e^-}_{flight}$ and L^{π}_{flight} are path lengths along the track from the vertex to the TOF counters. The term that involves $L^{\pi}_{\mathrm{flight}}$ is effectively calculated using the β variable and assuming the mass of the π^+:

$$\frac{L^{\pi}_{\mathrm{flight}}}{v_\pi} = \frac{PathSC_\pi}{c} \times \underbrace{\sqrt{\frac{m^2}{P^2} + 1}}_{1/\beta}$$

- c and v_π are the velocity of the electron and hadron respectively, in [cm/nsec][12].

- t_{e^-} and t_π are particle's time of flight from the interaction vertex to the scintillator plane.

Basically, this variable is the difference between the time of arrival of the pion candidate and the theoretical pion. The same term for electron pursues the objective to center the whole distribution around zero. In figure 4.18 is the ΔT distribution as a function of the momentum P for all positive tracks. See figure 4.19 for this distribution plotted for each CLAS sector. There are small upward shifts in sectors 0 and 5.

[10] not to confuse with the ΔT variable defined for electrons, should be clear which is which from the context.

[11] T.O.F and *RF* offset variable. The RF correction stand for Radio Frequency (historical jargon) which corresponds to the time shift of much less than 2 [ns] that aligns the beam bunches in time. The beam bunches arrived in the hall every 2 [ns] in the 6 [GeV] era. By using detector timing one can figure out which beam bunch is responsible for the event. Then, the RF correction is the time shif necessary to put that bunch at zero nanoseconds. When the accelerator beam is re-tuned, the trajectory through the machine can be a little bit different, so that the path length is a little bit different. Thus, the bunch timing is different. The calibration of the RF correction -also called RF offset or RF time- is done similar to all the calibration on a run-by-run basis. It is a single number. It is included in the calibration database.

[12] in this units the speed of light is ~ 30[cm/nsec].

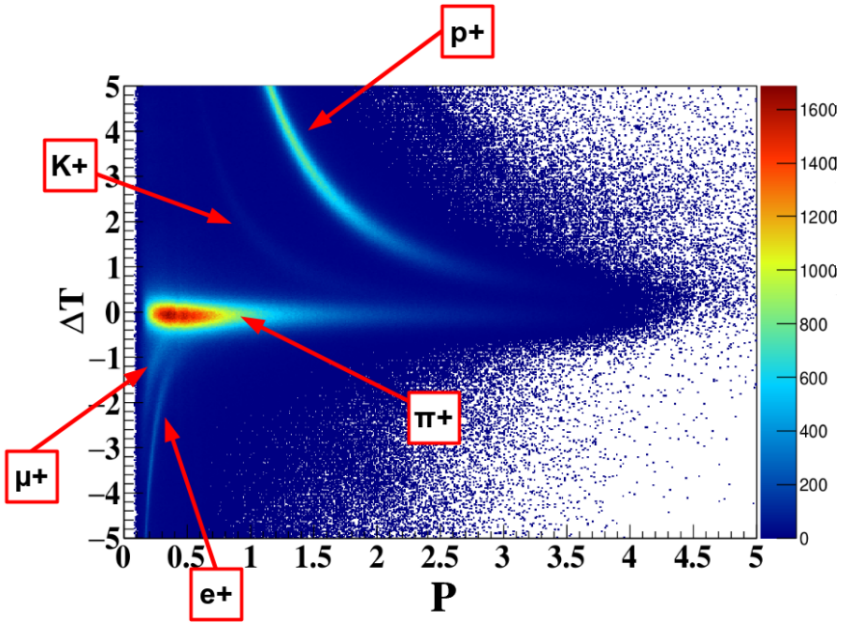


Figure 4.18: ΔT distribution as a function of momentum P (in [GeV/c]), for all positive particles.

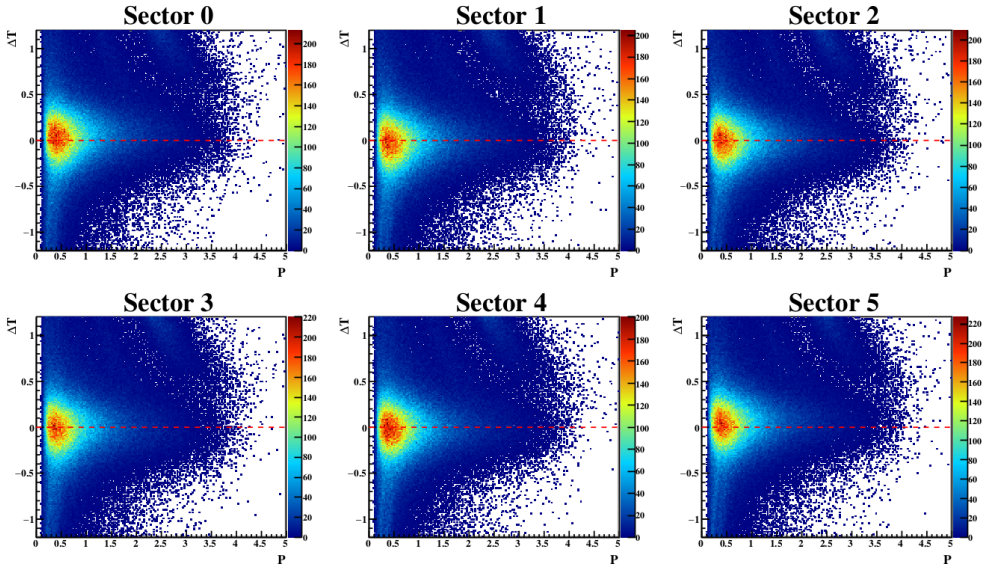


Figure 4.19: ΔT as a function of P, for all positive particles, for each CLAS sector.

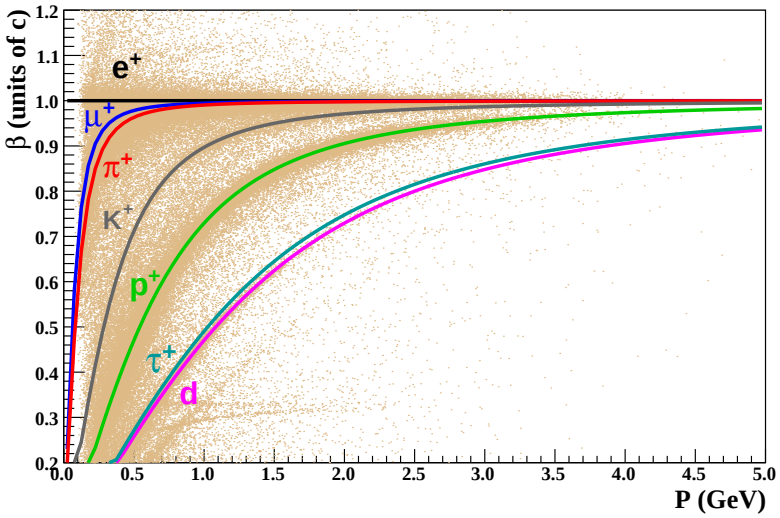


Figure 4.20: β distribution as a function of momentum P (in [GeV/c]), for all positive particles. The kinematical curves are, from top to bottom, positrons (black), μ^+ (blue), π^+ (red), K^+ (gray), p^+ (green), τ^+ (dark green) and deuterons (magenta), they are produced on the aluminum windows of the target cell (tritium as well.)

The β [13] distribution of the positive tracks as a function of the momentum are plotted in figure 4.20. The ideal kinematical curves -the ones obtained using the different masses in expression 4.22- are superimposed.

One problem that arises when we look at figure 4.20 (and figure 4.18 as well) is that at high P values it becomes more challenging to distinguish different type of particles. For example, the particle's velocity is calculated as:

$$\beta = \sqrt{\frac{1}{1+\left(\frac{m}{P}\right)^2}} \tag{4.22}$$

but if $P \gg m$ then $\beta \to 1$, independent of particle mass. So, there is a natural limit about the reliability of TOF technique for particle discrimination. For high P we know that particles like kaons and protons, that are more massive than pions ($m_{\pi^+} < m_{K^+} < m_{p^+}$), will not be able to surpass the speed of light in that medium (see section 3.2.2 for the details of CLAS Cherenkov detector). They will not produce a distinct Cerenkov signal, so CC information can be used to discriminate pions in that P region. Thus, the pion candidates are divided into two groups, depending on the momentum. For each case, a different technique is used to select them:

- Positive partciles with P < 2.7 [GeV] $\rightarrow$ **TOF**

[13] velocity of the particles in units of the speed of light.

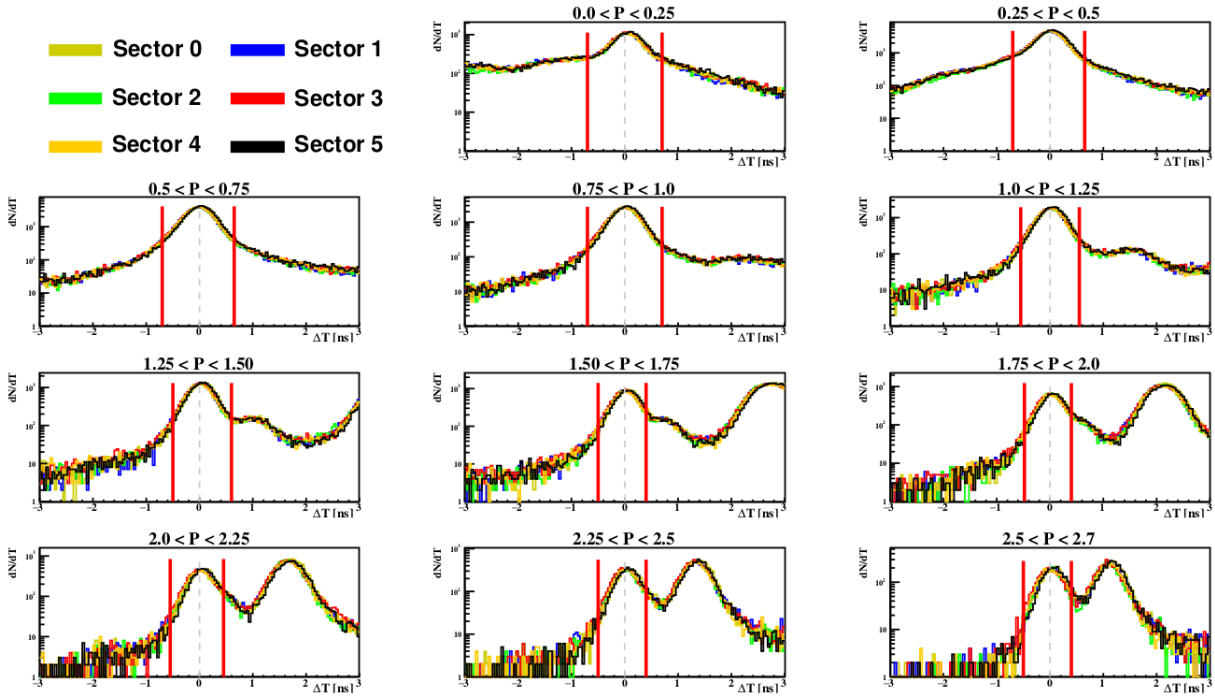


Figure 4.21: ΔT as a function of P, for all positive particles, for each CLAS sector.

- Positive particles with P $>$ 2.7 [GeV] $\rightarrow$ **CC**

4.3.3.1 TOF Technique, P $<$ 2.7 GeV

For particles with momentum below 2.7 GeV, first, a positive status in SC data bank is required. Then, a momentum dependent TOF cuts are applied. See figure 4.21 for a one-dimensional distribution of ΔT for different ranges in P and for all sectors. The red lines are the cuts imposed (explicitly in table 4.6). The cuts are wide enough to be sector independent.

To reduced kaon contamination at high P, a cut in the mass squared m^2 of the particles is applied. The expression for the m^2 is given by:

$$m^2 = p^2 \left(\frac{1}{\beta^2} - 1 \right) \tag{4.23}$$

the cuts are shown in table 4.7.

4.3.3.2 CC Technique, P $>$ 2.7

First a non null number of entries in the CC bank and a positive status in CC data bank are required. In addition a geometrical matching is required. The angle between CC hit and nearest SC hit should be less than 5°. A sector independent cut in the

ΔT limits for different P ranges		
P Ranges (GeV/c)	Lower Limit	Upper Limit
$[0.00-0.25]$	-0.70	0.70
$]0.25-0.50]$	-0.70	0.65
$]0.50-0.75]$	-0.70	0.65
$]0.75-1.00]$	-0.70	0.65
$]1.00-1.25]$	-0.55	0.55
$]1.25-1.50]$	-0.50	0.55
$]1.50-1.75]$	-0.50	0.40
$]1.75-2.00]$	-0.48	0.40
$]2.00-2.25]$	-0.50	0.40
$]2.25-2.50]$	-0.50	0.40
$]2.50-2.70]$	-0.50	0.40

Table 4.6: Limits for the ΔT variable (in [ns]), for different momentum ranges.

Mass Squared limits for different P ranges	
P Ranges(GeV/c)	Upper Limit
$[2.25-2.50]$	0.5
$]2.50-2.70]$	0.4

Table 4.7: Cut in mass squared, for two different momentum ranges.

number of photo-electrons is imposed: Nphe > 2.5. With this, a good selection of π^+ is achieved with some positron contamination. To reduce positron contamination the pion candidate must failed the positrons cuts. The positron is defined in the same way as the electron, but this time the charge must be positive. After the CC cuts there are a few events remaining with high values of ΔT, those are eliminated using $|\Delta T| < 0.35$.

The distributions of β and ΔT -as a function of the momentum of the particles- are depicted in figure 4.22 for fully selected pions. Note that after the threshold value of 2.7 GeV, there is drop in statistics.

In table 4.8 (and graphically in figure 4.23) there are fractions of remaining particles after the implementation of each PID cut in an incremental manner[14]. Out of all the positive particles, only about 7% of them are pions. In table 4.9 (and graphically in figure 4.24) are the effect of each cut individually.

In the positive pion selection, there is always a problem with possible contamination due to positrons and muons because the feasibility for separation of these three types of particles is strongly momentum dependent. Note in figures 4.18 and 4.20

[14]Note the factor 57.3 corresponds to the value in degrees of 1 radian, $57.3^\circ = 1Rd$.

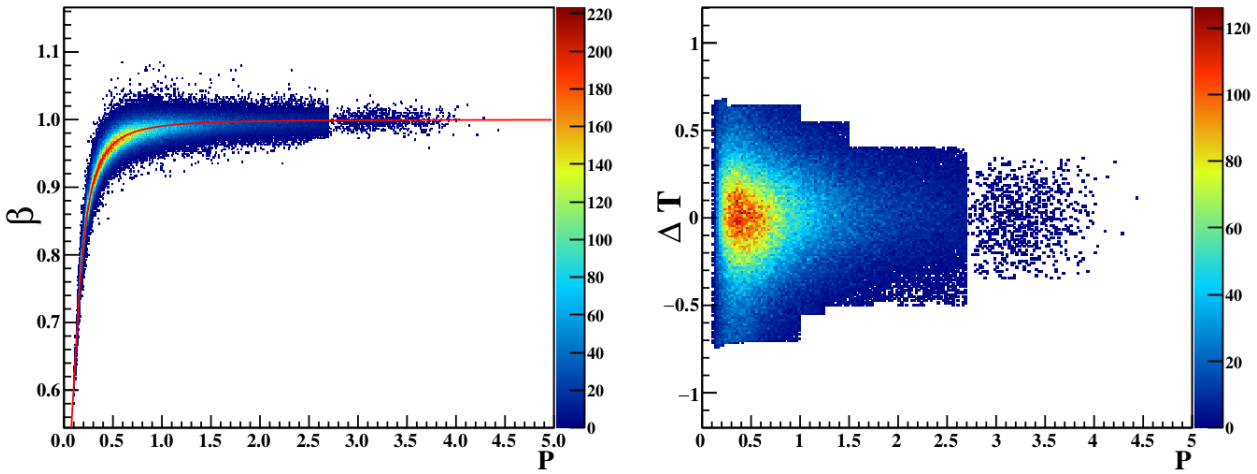


Figure 4.22: Distributions of β and ΔT as a function of the momentum, after all the identification cuts are imposed.

Fraction of remaining particles in Pion Identification							
N^o	Cuts	Sector 0	Sector 1	Sector 2	Sector 3	Sector 4	Sector 5
1	NumbDC !=0	100	100	100	100	100	100
2	DCStatus >0	34.744	37.118	34.874	36.009	35.236	36.555
3	StatDC >0	34.744	37.118	34.874	36.009	35.236	36.555
4	0<Status<100	32.825	35.534	33.436	33.893	33.964	35.397
5	NumbSC !=0, P<2.7	32.825	35.534	33.436	33.893	33.964	35.397
6	StatSC >0, P<2.7	32.595	35.200	32.401	33.560	32.938	35.025
7	NumbCC !=0, P$\geq$2.7	32.595	35.200	32.401	33.560	32.938	35.025
8	StatCC >0, P$\geq$2.7	32.595	35.200	32.401	33.560	32.938	35.025
9	Charge $=1$	28.074	30.433	28.212	29.022	28.499	30.488
10	No Positron	28.063	30.427	28.209	29.012	28.492	30.476
11	ΔT cut, P< 2.7	7.581	8.245	8.090	8.103	7.856	8.728
12	M2 cut, P< 2.7	7.578	8.235	8.083	8.094	7.846	8.713
13	ΔT cut, P$\geq$ 2.7	7.409	8.059	7.880	7.915	7.673	8.534
14	Nphe >25, P$\geq$ 2.7	7.309	7.939	7.788	7.809	7.583	8.439
15	χ^2_{CC} <5/57.3, P$\geq$ 2.7	7.303	7.929	7.768	7.807	7.571	8.429

Table 4.8: Effect in percentage of each cut in terms of remaining particles.

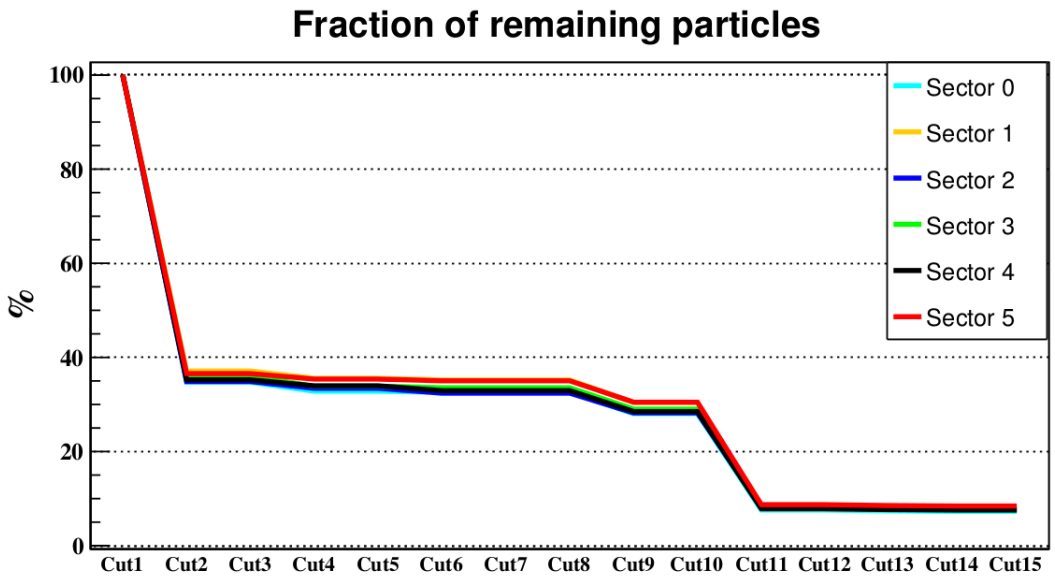


Figure 4.23: Remainig fraction of events after the incremental application of the cuts for positive pion selection.

Fraction of rejected events in Pion Identification							
N^o	Cuts	Sector 0	Sector 1	Sector 2	Sector 3	Sector 4	Sector 5
1	NumbDC !=0	0	0	0	0	0	0
2	DCStatus >0	65.2555	62.8816	65.1258	63.9908	64.7636	63.4447
3	StatDC >0	23.324	21.1352	24.1657	19.03	20.9575	21.0168
4	0<Status<100	52.3037	48.4642	49.4403	52.7903	52.0285	50.5306
5	NumbSC !=0, P<2.7	0	0	0	0	0	0
6	StatSC >0, P<2.7	19.6581	18.0882	20.4479	16.0778	18.322	17.8551
7	NumbCC !=0,$\geq$2.7	0	0	0	0	0	0
8	StatCC >0, P$\geq$2.7	0	0	0	0	0	0
9	Charge $=1$	44.6081	42.0084	44.7306	40.3378	42.1179	41.3098
10	No Positron	0.01137	0.006146	0.002637	0.01062	0.006919	0.01124
11	ΔT cut, P< 2.7	77.2082	76.5961	77.068	75.9168	76.0035	76.0296
12	M2 cut, P< 2.7	11.5252	10.2657	10.4658	12.2719	11.3613	10.9494
13	ΔT cut, P$\geq$ 2.7	26.4327	22.5265	25.5472	20.6977	23.5698	22.49
14	Nphe >25, P$\geq$ 2.7	2.3537	2.66437	2.86925	2.62221	2.51705	2.56324
15	χ^2_{CC} <5/57.3, P$\geq$ 2.7	0.751213	0.820516	0.955831	0.869519	0.948698	0.86191

Table 4.9: Effect, in percentage, of each cut in terms of rejected particles, for all different CLAS sectors.

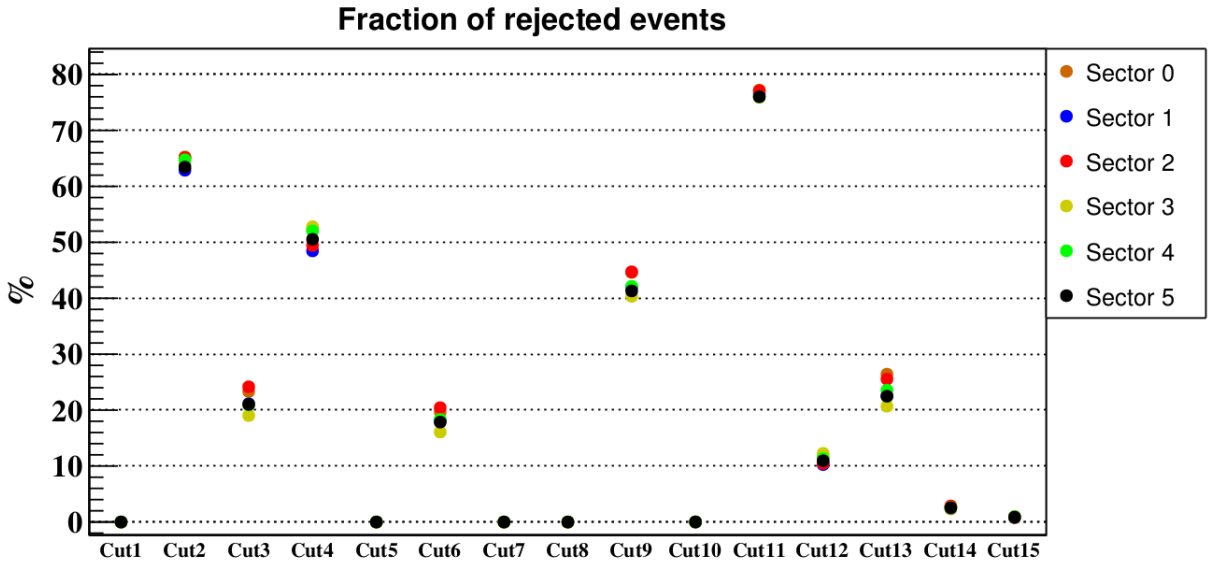


Figure 4.24: Remaining fraction of events after the individual application of the cuts for positive pion selection.

where e^+, μ^+, and π^+ -only at low values of momentum the peaks- are distinguished, while at higher momentum the mass resolution deteriorates and these peaks merge. For this reason, the default CLAS particle identification procedure always presumes all these particles to be pions. In figure 4.25, there is the mass squared distribution for all positive particles with momentum below 0.25 GeV; on the left the three particles are clearly visible . In order to enhance the positron peak, we can demand that the electron and the other charged particle are in the same sector (see right plot in figure 4.25.) This is due to the fact that the direction of the virtual photon would be towards a sector which is opposite to the electron's one. If it directly creates a pion from scattering off the proton (t-channel production [65]), then the pion would be almost in the same direction as the virtual photon. Positron would be produced through a multi-step process, thus the chances for positrons to be in the same sector as the electron are higher.

In figure 4.26 is depicted the mass squared distribution for fully selected π^+ with $P < 0.25$ GeV. The peaks for positrons and muons disappeared which gives us confidence in the selection criteria.

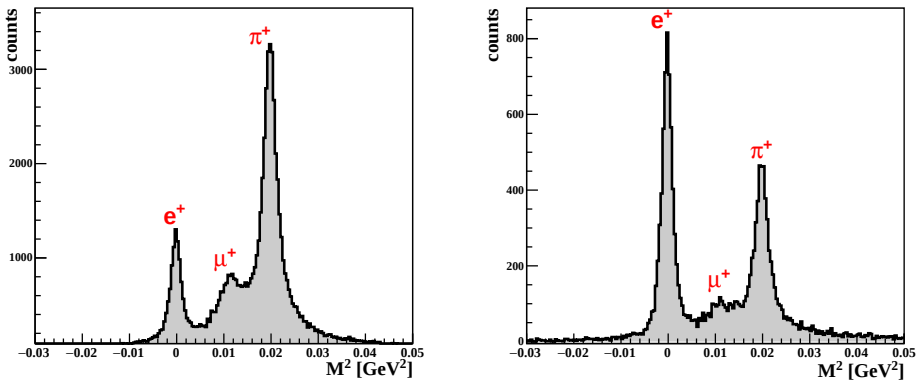


Figure 4.25: Mass squared (in GeV^2) distribution for positive particles with momentum below 0.25 GeV. On the left, it is without any sector cut. On the right, it is with imposing the same sector for electron and the positive particle.

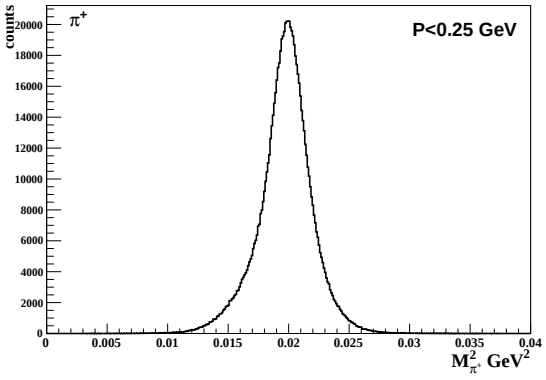


Figure 4.26: Mass squared (in GeV^2) distribution for fully selected π^+ with momentum below 0.25 GeV.

4.4 Extraction of Observables

4.4.1 Definitions

In term of yields the EMC ratio (**EMC**) is given by the following expression:

$$\mathrm{EMC} = \frac{\left(\mathrm{N}_{el^-}^{\mathrm{DIS}}\right)_A}{\left(\mathrm{N}_{el^-}^{\mathrm{DIS}}\right)_D} \tag{4.24}$$

where a normalization factor related to the target thickness must be included.

The hadronic multiplicity ratio (**MR**) is given by:

$$\mathrm{MR} = \frac{\left(\frac{\mathrm{N}_{\pi^+}^{\mathrm{SIDIS}}}{\mathrm{N}_{el^-}^{\mathrm{DIS}}}\right)_A}{\left(\frac{\mathrm{N}_{\pi^+}^{\mathrm{SIDIS}}}{\mathrm{N}_{el^-}^{\mathrm{DIS}}}\right)_D} \tag{4.25}$$

where $N_{el^-}^{\mathrm{DIS}}$ is the yield of DIS electrons and $N_{\pi^+}^{\mathrm{SIDIS}}$ is the yield of SIDIS pions, in a given bin for a solid target A or a deuterium target D.

4.4.2 Vertex Determination

Due to the nature of the double target system and the definition of the multiplicity ratio, it is necessary to impose a distinction between the events coming from the solid and cryo-target. A cut on the z-vertex of the electrons is applied. It must be a sector dependent cut because these distributions can be shifted, due to a misalignment of the beam.

In figure 4.27 it is presented the reconstructed vertex of all particles identified as electrons along the beam direction relative to the center of CLAS ($Z = 0$). The region outside the coloured one is rejected. This selection follows closely Ref [25].

The procedure to extract the vertex cuts is based on the mean value of each target, depicted in table 4.10, and a sector dependent shift, depicted in table 4.11. The detected electrons are associated with the solid target, if their vertex position is less than 1.5 cm ($\sim 3\sigma$) from the value of table 4.10. For the cryo-target, the cut is larger: 2 cm, in order to take into account the size of the target. In addition to those values, the sector dependent shift must be applied.

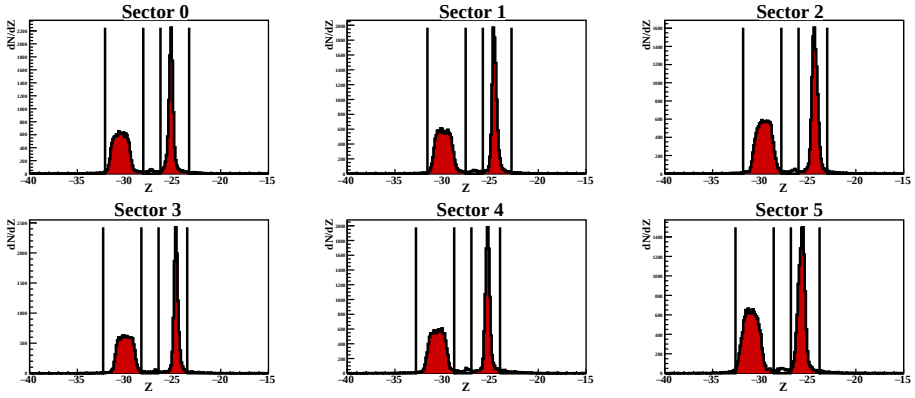


Figure 4.27: Z distribution for electrons in [cm] for each sector. The red region represents the applied vertex cut. The left red area corresponds to events coming from the deuterium target. The right red area corresponds to electrons coming from the solid target.

Vertex position			
Target	Carbon	Iron	Lead
Liquid	-30.1	-30.2	-30.1
Solid	-24.7	-24.9	-24.9

Table 4.10: Mean vertex position of the target

Sector dependent shift						
Sector	0	1	2	3	4	5
Shift (cm)	+0.1	-0.4	-0.6	-0.1	+0.4	+0.6

Table 4.11: Shifts (in cm) for each sector.

4.5 Additional Cuts

Besides the cuts for PID, some other cuts are considered in this analysis. The definition and motivation for each one is discussed in this section.

4.5.1 Scaling Variable Feynman X

The information on the transverse momentum distributions of quarks and gluons inside of the proton is encoded in the Transverse Momentum Dependent distribution functions (TMDs) [35], which are probed mainly in SIDIS. We would hope the measured hadrons to be described in terms of the *fragmentation* of the quark and gluon constituents of the nucleon involved in lepto-production processes. So, in a SIDIS interaction we expect the high energy hadrons in the final state X to be the fragments

of the struck quark, and the low energy hadrons to be products of the spectator or di-quark system. Thus, different kinematical regions of SIDIS processes correspond to different underlying partonic pictures. Let's define briefly the main two regions considered in this work:

- **C**urrent **F**ragmentation **R**egion (CFR): if the measured hadron belongs to the struck quark in the nucleon, it is current fragmentation region. In this region, a factorization picture with fragmentation functions is appropriate (to useful accuracy)[15]. This is the region we are interested in and the one that has received the most theoretical attention.

- **T**arget **F**ragmentation **R**egion (TFR): if the hadron is produced in the fragmentation process of the target remnants, it is target fragmentation region. If the hadron comes from the di-quark or anything else that is not the struck quark, it is TFR. Here, the SIDIS process can be described trough the so-called *fracture functions* [64].

In order to distinguish those regions, the Feynman X variable (X_f) is introduced. The use of this variable for this purpose is common (see, for example [54], [53], [32], [75] or [56]), but is not the only choice used. Other variables, such as rapidity (we will comment on this later) or the z scaled multiplicity, are useful in some cases. Those are related (see [67]).

In the $\gamma^* N$ center of mass (CM) frame [16] the hadron momentum is $P_h = (E_h, \mathbf{P}_{hT}, P_{hL})$. X_f is defined as the fraction of the maximum longitudinal momentum carried by the observed hadron, i.e:

$$X_f \equiv \frac{\mathrm{P}_l^{\mathrm{CM}}}{\mathrm{P}_l^{\mathrm{CM}}(\mathrm{max})} \qquad ; \qquad |X_f| < 1 \tag{4.26}$$

, where $\mathrm{P}_l^{\mathrm{CM}}$ is the true longitudinal momentum of the particle and $\mathrm{P}_l^{\mathrm{CM}}(\mathrm{max})$ is the maximum momentum that the particle could ever have in this frame. It is clear that X_f is dimensionless and it goes between -1 and 1. This variable identifies the CFR ($X_f > 0$) and TFR ($X_f < 0$).

In order to calculate the expression for $\mathrm{P}_l^{\mathrm{CM}}$, let's consider a boost in the longitudinal direction:

[15] this means that the appropriate theoretical framework for describing this picture is TMD factorization, with TMD parton distribution functions (PDFs) as well as TMD fragmentation functions (FFs).

[16] contrary to Breit frame, the interaction of the virtual photon with vacuum fluctuations are not suppressed in this frame.

$$\begin{pmatrix} E_\pi \\ P_\pi \end{pmatrix}_{\text{lab}} = \begin{pmatrix} \gamma_{cm} & \beta_{cm}\gamma_{cm} \\ \beta_{cm}\gamma_{cm} & \gamma_{cm} \end{pmatrix} \begin{pmatrix} E_\pi \\ P_\pi \end{pmatrix}_{\text{CM}}$$

where γ is the usual Lorentz time-dilation factor (for more details of this refer to [66]):

$$\gamma_{cm} = \frac{1}{\sqrt{1-\beta_{cm}^2}} = \frac{\nu + M_p}{W} \tag{4.27}$$

$$\beta_{cm} = \sqrt{1-\frac{1}{\gamma_{cm}^2}} = \frac{\sqrt{\nu^2+Q^2}}{\nu + M_p} \tag{4.28}$$

thus, the final expression is:

$$\mathrm{P}_l^{\mathrm{CM}} = \left(\mathrm{P}_l^{\mathrm{Lab}} - \frac{\sqrt{Q^2+\nu^2}Z_h\nu}{\nu+M_p}\right)\left(\frac{\nu+M_p}{W}\right) \tag{4.29}$$

It is worth noting the way to calculate the $\mathrm{P}_l^{\mathrm{CM}}(\max)$ term. Clearly this term attains its maximum value when the undetected collection of particles X consists of a single particle with a rest mass M_X. This corresponds to the minimum value of the rest mass of X allowed by the conservation laws. Due to baryon number and charge conservation it can be found that the particle should be a neutron.

$$e + p \longrightarrow e' + \pi^+ + n \tag{4.30}$$

From the energy conservation condition, where $\sqrt{s}$ is the center of mass energy of the collision:

$$\sqrt{M_{\pi^+}^2 + p_{\pi^+}^2} + \sqrt{M_n^2 + p_n^2} = \sqrt{s} \qquad /(\ )^2 \tag{4.31}$$

$$2\sqrt{\left(M_{\pi^+}^2 + p^2\right) + \left(M_n^2 + p^2\right)} = s - \left(\left(M_{\pi^+}^2 + p^2\right) + \left(M_n^2 + p^2\right)\right) \qquad /(\ )^2 \tag{4.32}$$

$$4\left(M_{\pi^+}^2 M_n^2 + p^2 M_{\pi^+}^2 + p^2 M_n^2 + p^4\right) = s^2 + \left(M_{\pi^+}^2 + M_n^2 + 2p^2\right)^2 - 2s\left(M_{\pi^+}^2 + M_n^2 + 2p^2\right) \tag{4.33}$$

$$\begin{aligned} 4M_{\pi^+}^2 M_n^2 + \cancel{4p^2 M_{\pi^+}^2} + \cancel{4p^2 M_n^2} + \cancel{4p^2} &= s^2 + M_{\pi^+}^4 + M_n^4 + 2M_{\pi^+}^2 M_n^2 \\ &\quad + \cancel{4p^2} + \cancel{4p^2 M_{\pi^+}^2} + \cancel{4p^2 M_n^2} \\ &\quad - 2sM_{\pi^+}^2 - 2sM_n^2 - 4sp^2 \end{aligned} \tag{4.34}$$

thus, the expression for p is :

$$p^2 = \frac{s^2 + M_{\pi^+}^4 + M_n^4 - 2\left(M_{\pi^+}^2 M_n^2 + s\left(M_{\pi^+}^2 + M_n^2\right)\right)}{4s} \tag{4.35}$$

considering $\sqrt{s} = W$:

$$P_l^{CM}(max) = \frac{\sqrt{W^4 + M_{\pi^+}^4 + M_n^4 - 2\left(M_{\pi^+}^2 M_n^2 + W^2\left(M_{\pi^+}^2 + M_n^2\right)\right)}}{2W} \tag{4.36}$$

The constraint on X_f $(X_f > 0)$ will reduce possible contributions from target fragmentation. The X_f distribution is depicted in figure 4.28, on the left and, on the right, is the energy distribution of the pion $(=Z_h\nu)$ for the TFR (red) and CFR (blue) compared to the total case, where no distinction is made (gray). Note the dramatic drop in the statistics after the X_f cut, which less than $\sim 60\%$ of the remained events. This explains why some studies prefer higher statistics over a strict cut to reduce fragmentation contamination in the data.

Note that this distinction of regions -based on the sign of X_f- is not a final solution to the problem of selecting just the hadrons from the struck quark, but it is simplistic and rather arbitrary. Associating individual hadrons with a struck quark or recoiling di-quark involves numerous pitfalls, assumptions, and uncertainties. Moreover, we can think of three regions instead of just two, adding a central (or soft) fragmentation region in which the other two overlap for some cases. A unified description with optimal accuracy requires matching of the factorization properties of the individual regions. These considerations are beyond the scope of this thesis. We are going to identy the regions based on the sign of X_f.

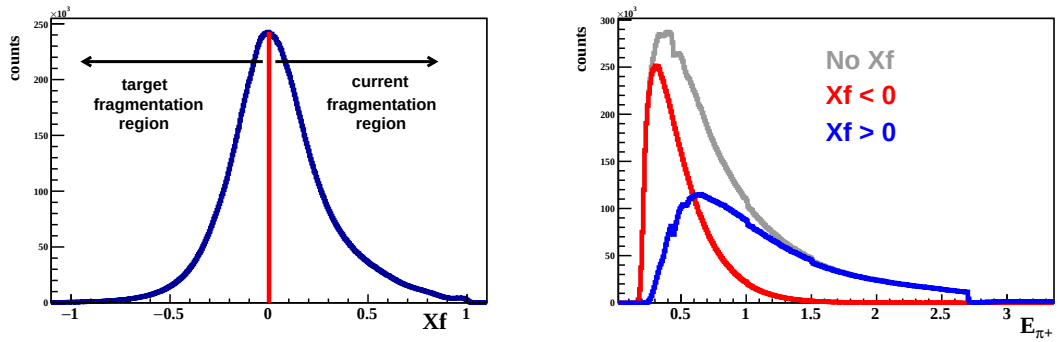


Figure 4.28: Distribution of Xf (left). The red line represent the cut applied. On the right is the energy of the observed hadron with and without the X_f cut.

In literature it is very common the use of rapidity variable (y_h) of the observed hadron to define these regions (see [51] or [52]). In the $\gamma^* N$ center of mass reference frame it holds that:

$$y_h = \frac{1}{2} ln \left(\frac{E_h^{CM} + P_L^{CM}}{E_h^{CM} - P_L^{CM}} \right) \quad ; \quad E_h^{CM} = \frac{W^2 + M_\pi - M_p}{2W} \tag{4.37}$$

In figure 4.29 there is a two-dimensional plot of X_f and y_h. It is clear the corelation between the two. For more details about the relation between X_f and y_h and some other variables see [57].

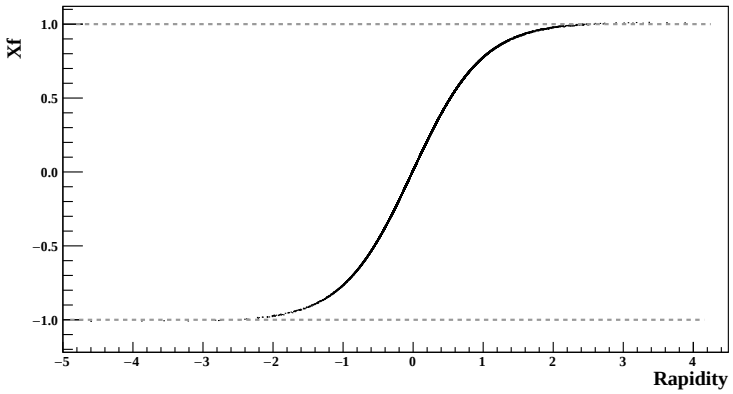


Figure 4.29: Comparison of the X_f distribution and the y_h distribution. Both are calculated in the $\gamma^* N$ center of mass frame.

4.5.2 ΔZ Variable

The pion primary decay mode is into a muon and a muon neutrino:

$$\pi^+ \longrightarrow \mu^+ + \nu_\mu \tag{4.38}$$

Some GEANT-based MC studies showed that about 18% of the positive pions decay in flight into $\mu^+ \nu_\mu$ [26]. Most of the momentum of the original pion is carried by the muon, which is, often detected and reconstructed as a π^+ with a significantly different momentum vector.

In order to reduce the number of events with decaying pions, the vertex for pions is checked against the electron one. It can be seen that the targets are well separated, in figure 4.30.

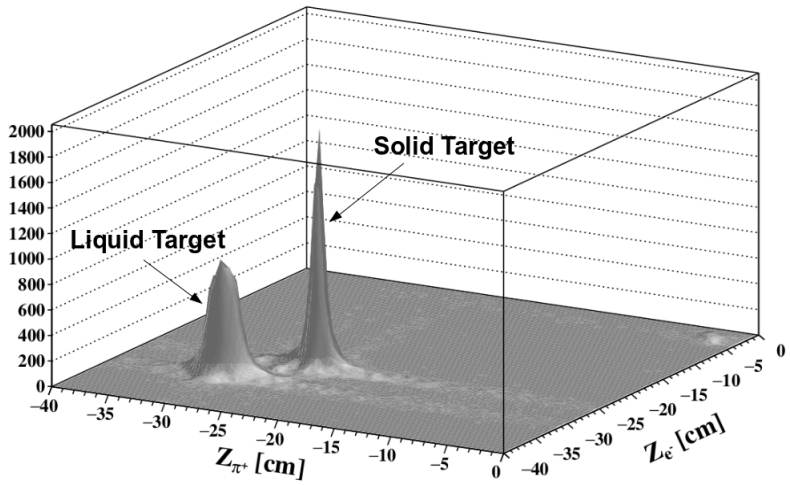


Figure 4.30: Distribution of Z vertex position, for the electron and the π^+.

A vertex difference may eventually be applied, for that purpose we define the ΔZ as:

$$\Delta Z \equiv Z_{\pi^+} - Z_{el^-} \tag{4.39}$$

the cut applied is:

$$|\Delta Z| < 3[cm] \tag{4.40}$$

In figure 4.31, we can see the distribution of ΔZ and the cut applied. All events outside those limits are rejected which out of the total sample are $\sim 15\%$. This selection follows Ref [25]. For a more restrictive choice see [26].

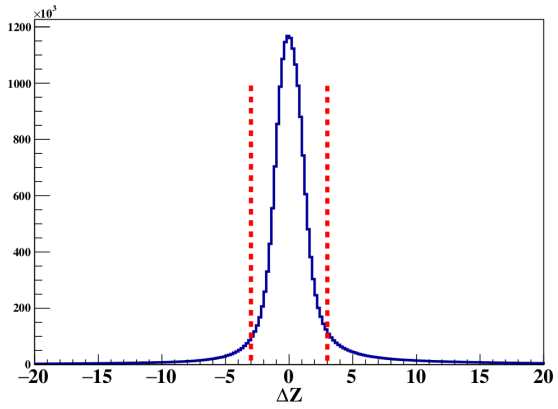


Figure 4.31: Distribution of ΔZ. The red lines represent the cut applied.

4.5.3 YC Variable

In vertex determination (section 4.4.2), a cut in the Z coordinate of the electron was applied. One of the problems is that, a cut in Z will not remove all aluminium signals. In order to get a clearer sample we need to explore how to eliminate this background.

For background subtraction purposes, a cut in the corrected value of Y coordinate is applied. This correction was performed by Taisiya Mineeva [59], and it is done taking into account the fact that the real beam position is not $(x, y) = (0, 0)$. There is an offset, and it was determined to be $(x, y) = (-0.043, 0.33)$. In order to eliminate the sector dependence, the track was extrapolated linearly along the momentum to the plane containing the real beam position. This was done for X, Y, and Z coordinates. The corrected version of these coordinates are called XC, YC, and ZC respectively. See figure 4.32 to observe the effect in the Z coordinate for each sector. Here it can be seen the double target system before and after the correction. The small peak between the targets correspond to the thin aluminium reference foil.

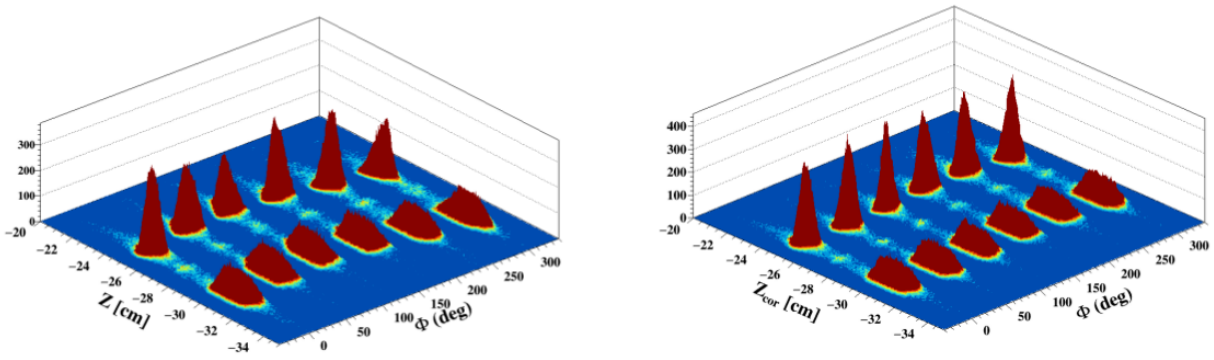


Figure 4.32: Z coordinate before the correction (left) and after the correction (right) as a function of Φ_{lab} angle of the electron.

The reconstruction of Y is worse than the one of Z due to the positioning of the stereo wires. That is why it is necessary a cut in YC, which is identical for solid and cryo-targets.

To see how the YC vertex cut affects Z resolution, look at the figure 4.33 (left). This is a one-dimensional plot of the Z distribution (using empty targets run) for three different YC regions (normalized to unity), the gray case is without any cut in YC. For $|YC| > 0.1$[cm] the two peaks are resolved, but for the case of $|YC| > 1.0$[cm] the small peaks are indistinguishable while the Z resolution for $|YC| > 2.0$ is extremely wide. This means that a cut in Z vertex is worthless for selecting targets in $YC > 1.5 - 2$[cm]. The background under the target peak comes from all Z values.

So, if we do not apply a YC based cut, we are effectively mixing events from the two targets.

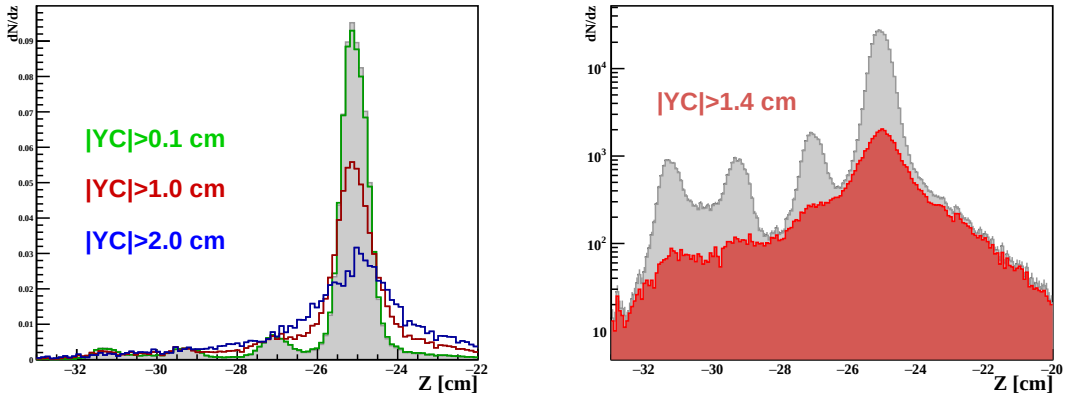


Figure 4.33: On the left, there is the Z vertex distribution for electrons tracks, for three different cuts in YC coordinate in [cm]. This is for lower Q^2. On the right, there is the Z distribution but this time in red there are the events with values greater than 1.4 cm or less than -1.4 cm.

Based on a study performed by Orlando Soto and Will Brooks [60] about empty target fits, we can use a cut in the YC at 2σ, 1.4 cm, which removes $\sim 22\%$ of events (see figure 4.33 on the right). With this cut, the background is nearly zero at $Z = -28$cm. This is our nominal cut for the YC variable. Note that, even with background subtraction, the aluminium foil is still there. Therefore, a vertex cut in the Z coordinate is still necessary. A cut in the X coordinate is not necessary, due to the fact that it is determined based on the reconstruction of the Y vertex.

4.5.4 Missing Mass Cut, Contribution of Exclusive Events.

In SIDIS events, the missing mass (W_x or M_x) is defined as the invariant mass of the recoil system, i.e, the undetected state X. Let's consider q, P, and P_h as the 4-momenta of the virtual photon; the initial state of the proton; and the detected pion, respectively. Then:

$$W_x^2 = (q + P - P_h)^2 \tag{4.41}$$

taken into account that:

$$P_h = (E_h, \vec{p_h}) \tag{4.42}$$
$$P_h = (P + p - q) \tag{4.43}$$
$$P_h = (M + E', -(\vec{p} - \vec{p'})_x, -(\vec{p} - \vec{p'})_y, E - (\vec{p} - \vec{p'})_z) \tag{4.44}$$
$$q^2 = \nu^2 - (\vec{p} - \vec{p'})^2 = -Q^2 \tag{4.45}$$
$$||\vec{p} - \vec{p'}|| = \sqrt{\nu^2 + Q^2} \tag{4.46}$$

thus, we can find the final expression for W_x:

$$W_x^2 = (P+q)^2 + P_h^2 - 2(P+q)\cdot P_h \tag{4.47}$$
$$= W^2 + M_h^2 - 2P\cdot P_h - 2q\cdot P_h \tag{4.48}$$
$$= W^2 + M_h^2 - 2P\cdot(p + P - q) - 2(E - E', \vec{p} - \vec{p'})\cdot(E_h, \vec{p_h}) \tag{4.49}$$

where:

$$2P\cdot(p + P - q) = 2(P\cdot p + P\cdot P - P\cdot q) \tag{4.50}$$
$$= 2(ME + M^2 - M\nu) \tag{4.51}$$
$$2(E - E', \vec{p} - \vec{p'})\cdot(E_h, \vec{p_h}) = 2(\nu E_h - p_h\sqrt{Q^2 + \nu^2}cos(\theta)) \tag{4.52}$$

replacing all of this into equation 4.41, the final expression is:

$$W_x^2 = \left(W^2 + M_h^2 - 2\nu^2 Z_h - 2M\nu Z_h + 2p_h\sqrt{Q^2 + \nu^2}cos(\theta)\right) \tag{4.53}$$

Note that in the final expression of W_x, we assumed the pion mass, i.e, $M_h = M_{\pi^+}$. Strictly speaking, a cut in W_x is not part of the identification cuts for π^+, but sometimes it is used to ease the π^+/p separation at high momentum [35].

In our SIDIS sample, there might be some contribution from exclusive reactions ($ep \to e'\pi^+ n$) and/or resonances particles [17]. In figure 4.34 (left), it is shown the distribution of the missing mass of the X system, and in red it is shown the region of $W_x < 1.4$ GeV [18]. The two shown peaks in red are from exclusive reactions (mass of the neutron is 0.9395 GeV in the leftmost peak) while the other one is due to a

[17] In general, the main reactions are $ep \to e'(\pi^+ n, \pi^0 p, \Delta^0, \rho_0)$ and $en \to e'(\pi^- p, \pi^0 n, \Delta^{++}, \rho_0)$

[18] This choice of threshold value for the missing mass is very common for these processes, see [36]. There are other choices tough, such as in [35].

resonance (mass of $\Delta^0 = 1.232$ GeV). This exclusive contribution is highlighted in deuterium basically because there are fewer nuclear effects than in the heavy nuclei.

A constant value of W_x creates a contour in the $P_t^2 - Zh$ space which can be seen in figure 4.34, on the right. From this plot, we can see that the exclusive processes dominates the SIDIS process at high Zh and P_t^2. Note that this plot compares three different cuts in W_x, from outermost to innermost are cuts at the M_p (mass of the target nucleon, in this case this would be a exclusive reaction), at 1.4 GeV and finally a more extreme cut at 1.5 GeV. As we increase the value of the cut the higher values of Zh disappear. A cut in the missing mass is not part of the nominal cuts presented in this thesis. The effect of this cut on the final observable will be discussed in the next chapter.

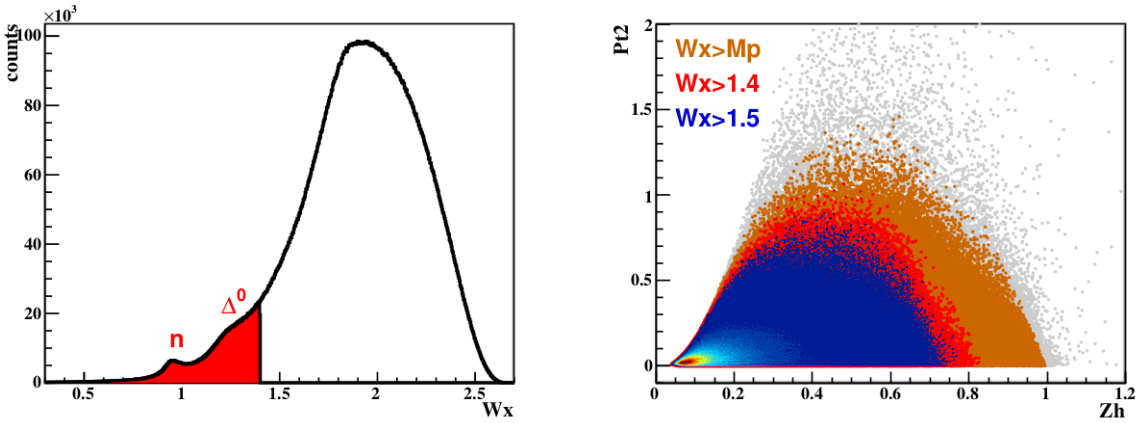


Figure 4.34: Missing mass distribution for semi-inclusive π^+ in the left and (P_t^2,Zh) space for different cuts in the missing mass in the right.

4.6 Binning

As it is discussed in section 4.8.1 for single pion electro-production at fixed beam energy, one needs to specify five independent kinematical variables to uniquely determine all other kinematical quantities. Here are presented a set of bins used in the five variables. The idea behind this choice is to have an equal bin statistics. Since mainly the MR is going to be as a function of Zh, the width of that variable remains fixed; the same goes for Φ_{PQ}. All the corrections are implemented in a bin by bin basis, so this choice of binning also applies to the simulation set. See tables 4.12 and 4.13.

Binning				
Variable	N° Bins	Low Limit	Up Limit	Width
Q^2	6	1.0	4.0	Not fixed
Z_h	10	0.0	1.0	0.1
X_b	5	0.12	0.57	Not fixed
P_T^2	5	0.0	1.0	Not fixed
Φ	12	-180°	180°	30°

Table 4.12: Binning choice information.

Binning	
Variable	Edges
Q2	(1.0, 1.17, 1.33, 1.51, 1.75, 2.12, 4.0)
Xb	(0.12, 0.19, 0.23, 0.27, 0.33, 0.57)
Pt2	(0.0, 0.03, 0.06, 0.1, 0.18, 1.0)

Table 4.13: Bin edges for three different kinematical variables.

4.7 Simulation Set

The CLAS detector has some unavoidable difficulties due to the complexity of the detector and high particle yields. This ends up in a reduction of the data obtained from the experiment. The results of the experiments ideally should be detector independent. For this reason we must correct our results for acceptance effects [19].

We have mainly two issues to address: the fact that the geometry of CLAS does not cover the full momentum space range of the phenomena, and the limited efficiency of CLAS detectors and the reconstruction protocol. At the end, we only have dataset distributions *after* the experiment, with all these problems already there. In order to determine these corrections, we need to know, based on our experimental distributions, what should be the dataset's shape *before* passing through all the detectors. The method used to find the response of CLAS to our experiment is determined by the behavior observed on simulated data.

The task associated with simulations can be divided into two parts: generation of physical events, and reconstruction in detector components. The next sections discuss both.

4.7.1 Generation of Events

The MC generator used in this thesis is Pythia 6.319. This generator contains a model of non-perturbative and perturbative DIS processes, and spans the whole region

[19] *detector acceptance* is the probability that an event is *accepted* by a detector.

from real to quasi-real photo-production to high Q^2. To run it, the ROOT C++ version 5.08 package was used. The simulation is a multi-stage process with a large number of parameters and options, which were tuned on in the generator. This is covering the semi-inclusive region of CLAS. In this thesis, the simulation set consists in approximately 100 million events for each target.

4.7.2 Reconstruction of Events

The generated events -also called "thrown" or "true"- serves as an input to GSIM. GSIM model is an idealized program which simulates the response of the CLAS spectrometer to MC generated events. The features of the EG2 experiment has been implemented by Hayk Hakobyan [61]. Such features are built on the base of the GEANT simulation package of CERN software, and allow to model the response of the spectrometer as the particles pass trough it. This will take care of the acceptance of CLAS. The output of GSIM is then passes through GPP, **G**SIM **P**ost **P**rocessing. This is a program that permits a more realistic behavior of the simulations, considering signal from dead channels, dead wires in the DCs and bad tubes in SCs. The output of GPP will pass to the reconstruction process, using the RECSIS program. The same libraries for processing the actual data from EG2 run period was used here for the simulated events.

As a summary, the reconstructed events are those who pass the routine:

$$\text{MC} \longrightarrow \text{GSIM} \longrightarrow \text{GPP} \longrightarrow \text{RECSIS}$$

4.7.2.1 Particle Identification Cuts Applied to Simulations

Since the GSIM package contains the ideal response of the detector to the passage of the particle due to an idealized description of the geometry, it is necessary to look at the particle identification cuts for the reconstructed MC events. If the cuts need some tuning for the distribution in the simulations, we have to take that into account to obtain optimal results. This section's aim is to see all the relevant distributions for electron and pion selection, and also redo the cuts if it is necessary. Here, we are not going into a full description of each cut since the idea is the same as it was for the data set.

4.7.2.2 Electron PID

First, the data banks, charge and fiducial cuts remain the same as in data for the reconstructed events. The number of photo-electron cuts remains the same as in data

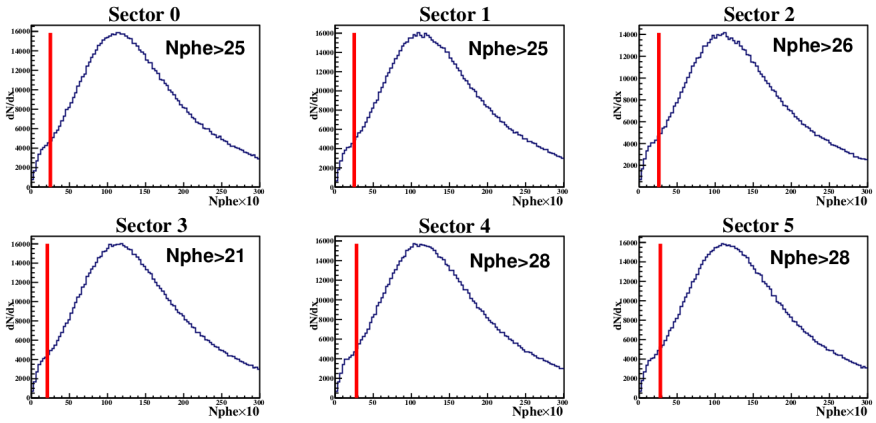


Figure 4.35: Nphe cut, applied in the reconstructed events, for each CLAS sector.

as well, as it is shown in figure 4.35. One thing to notice is that the CC efficiency is much worse in the real data than in the simulations. The CC had some inefficient zones that could not be simulated by the MC technique as being too dependent on specific features of the CC design. Signals from these zones, -being depleted of photo-electrons-, shifted the measured CC spectrum toward zero, and added up to the few photo-electron peak. Thus, the inefficient zones are associated with a more pronounced few photo-electron peak and the efficient ones with a smaller peak. For this reason, the Nphe spectrum in the simulated data does not have a big peak at 1-2 photo-electron. The cuts are imposed here just for consistency with data.

The cuts based on EC energy are slightly modified. The new cuts are in equations 4.54 to 4.59 (see figure 4.36) and in equations 4.60 to 4.65 (see figure 4.37). In figures 4.38 and 4.39 are shown the fiducial cuts, based on DC and EC respectively. Figure 4.40 show the cut in the elapsed time between SC and EC.

$$\Gamma_1^0: \quad E'_{tot} < 1.05 \times P + 0.21 \quad ; \quad \Gamma_2^0: \quad E'_{tot} > 1.03 \times P - 0.42 \tag{4.54}$$

$$\Gamma_1^1: \quad E'_{tot} < 1.05 \times P + 0.21 \quad ; \quad \Gamma_2^1: \quad E'_{tot} > 1.03 \times P - 0.42 \tag{4.55}$$

$$\Gamma_1^2: \quad E'_{tot} < 1.05 \times P + 0.21 \quad ; \quad \Gamma_2^2: \quad E'_{tot} > 1.05 \times P - 0.46 \tag{4.56}$$

$$\Gamma_1^3: \quad E'_{tot} < 1.03 \times P + 0.23 \quad ; \quad \Gamma_2^3: \quad E'_{tot} > 1.03 \times P - 0.44 \tag{4.57}$$

$$\Gamma_1^4: \quad E'_{tot} < 1.06 \times P + 0.20 \quad ; \quad \Gamma_2^4: \quad E'_{tot} > 1.06 \times P - 0.46 \tag{4.58}$$

$$\Gamma_1^5: \quad E'_{tot} < 1.04 \times P + 0.22 \quad ; \quad \Gamma_2^5: \quad E'_{tot} > 1.04 \times P - 0.47 \tag{4.59}$$

where $E'_{tot} = \frac{E_{tot}}{0.27}$.

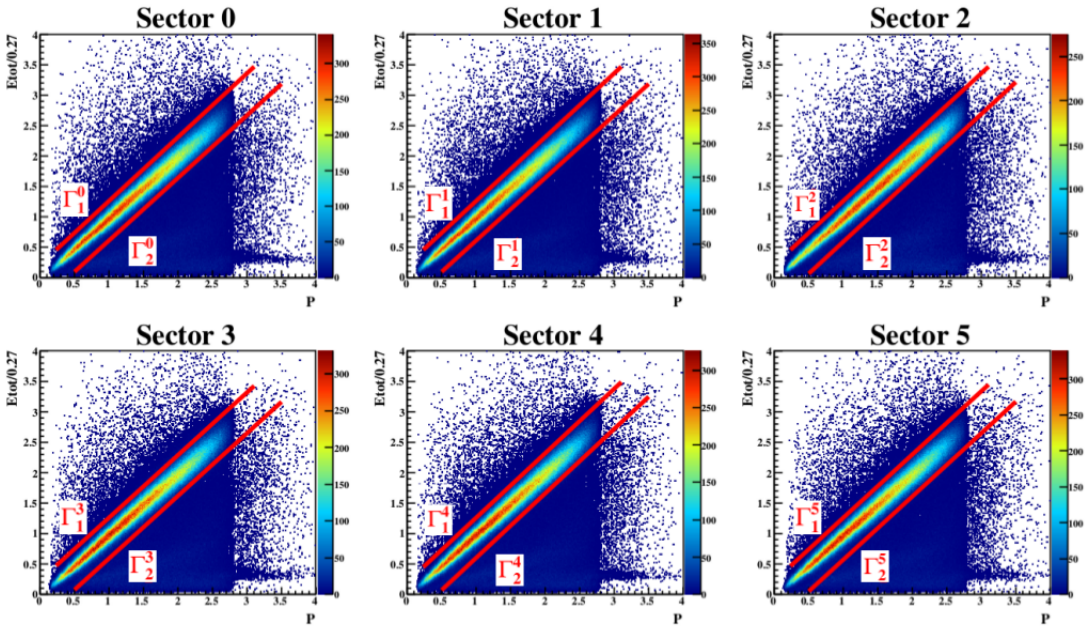


Figure 4.36: Etot/P as a function of the momentum for all electron candidates with only banks cuts and charge in the reconstruction. The red lines represents the cuts applied

$$\Gamma_1^0: \quad E'_{in} + E'_{out} < 1.13 \times P \quad ; \quad \Gamma_2^0: \quad E'_{in} + E'_{out} > 0.79 \times P \tag{4.60}$$

$$\Gamma_1^1: \quad E'_{in} + E'_{out} < 1.13 \times P \quad ; \quad \Gamma_2^1: \quad E'_{in} + E'_{out} > 0.79 \times P \tag{4.61}$$

$$\Gamma_1^2: \quad E'_{in} + E'_{out} < 1.14 \times P \quad ; \quad \Gamma_2^2: \quad E'_{in} + E'_{out} > 0.79 \times P \tag{4.62}$$

$$\Gamma_1^3: \quad E'_{in} + E'_{out} < 1.14 \times P \quad ; \quad \Gamma_2^3: \quad E'_{in} + E'_{out} > 0.79 \times P \tag{4.63}$$

$$\Gamma_1^4: \quad E'_{in} + E'_{out} < 1.15 \times P \quad ; \quad \Gamma_2^4: \quad E'_{in} + E'_{out} > 0.79 \times P \tag{4.64}$$

$$\Gamma_1^5: \quad E'_{in} + E'_{out} < 1.14 \times P \quad ; \quad \Gamma_2^5: \quad E'_{in} + E'_{out} > 0.79 \times P \tag{4.65}$$

where $E'_{in/out} = \frac{E_{in/out}}{0.27}$

The sampling fractions case, a downward shift of the distributions was observed. Taking that into account all the fits were recalculated for all targets. The new coefficients are in tables 4.14, 4.15, 4.16 and 4.17. In figure 4.41 there is an example of the application of these new cuts.

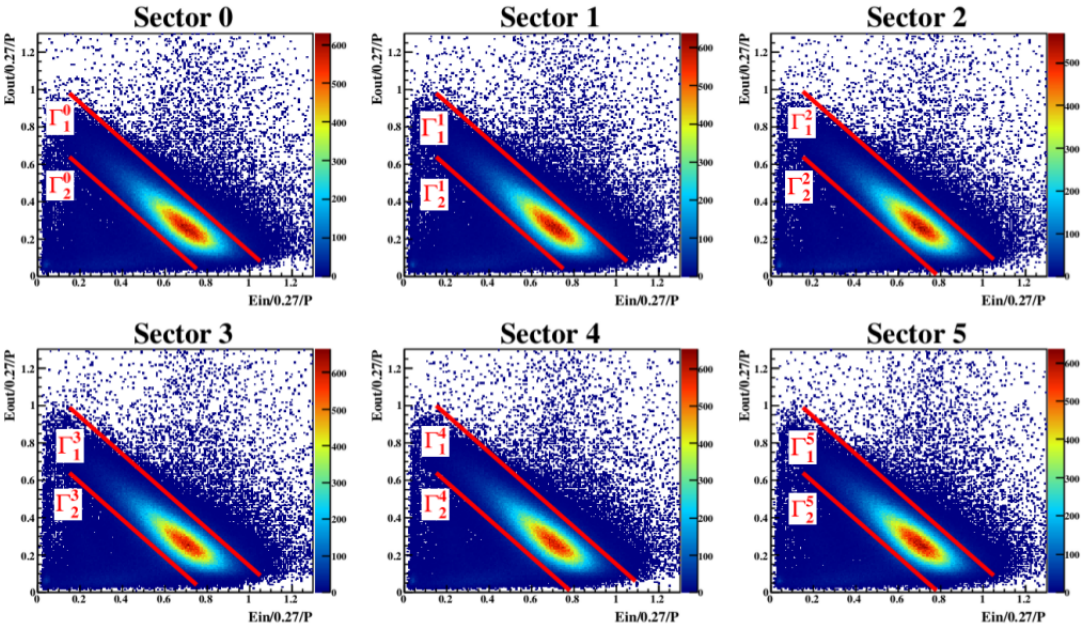


Figure 4.37: Ein and Eout as a function of the momentum for all electron candidates with only banks cuts and charge in the reconstruction. The red lines represents the cuts applied.

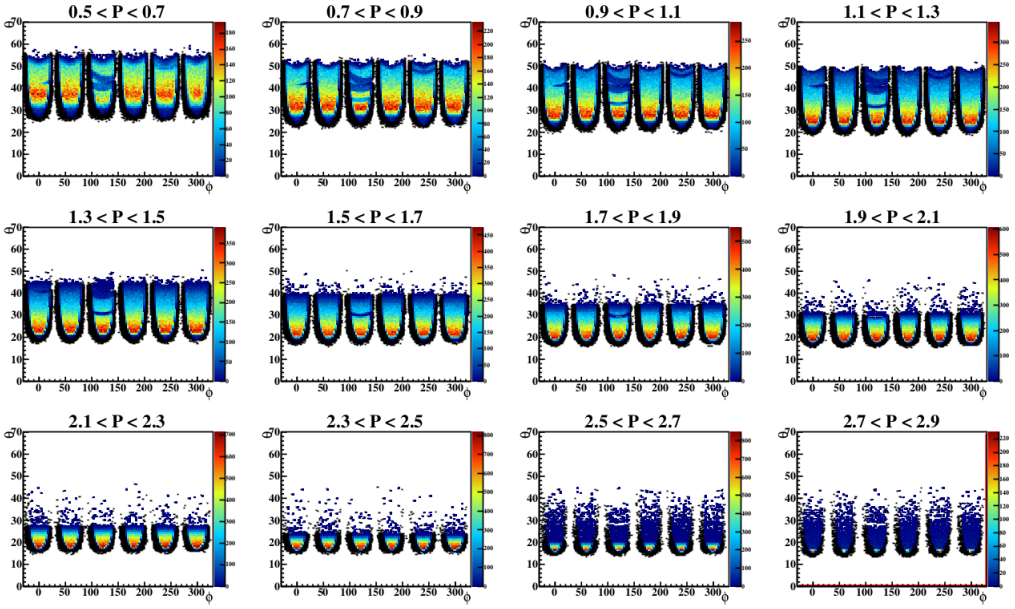


Figure 4.38: Fiducial cuts applied in the reconstructed events. Two dimensional view of Φ and Θ angles for different P ranges. The black region is rejected.

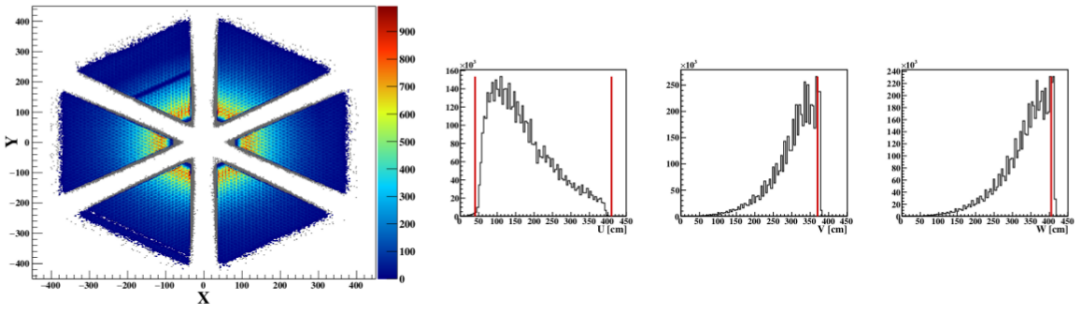


Figure 4.39: Fiducial cuts on EC, in reconstructed events. On the left, is a view of X and Y coordinate (gray region is rejected) while on the left there is the one-dimensional view of U,V and W coordinate. The red lines represents the cuts applied.

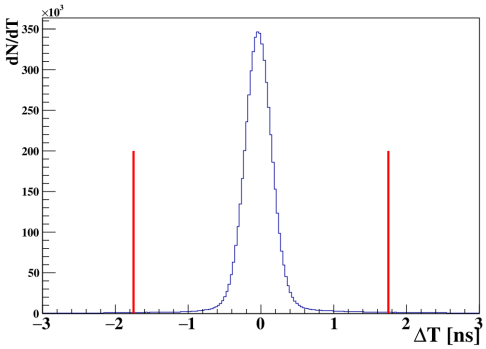


Figure 4.40: Elapse time based cut in the reconstruction.

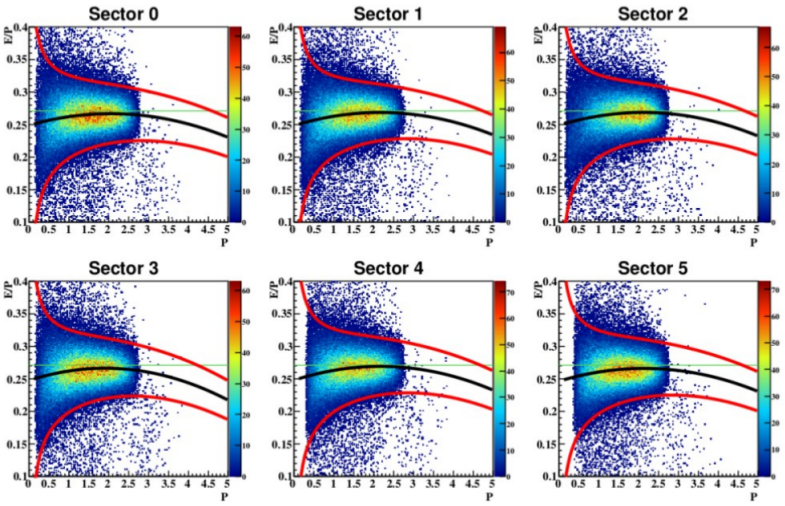


Figure 4.41: Example of a fit process applied in the simulations for sampling fraction cut coefficients. The same procedure as in data.

Sampling Fraction Coefficients for **Carbon**						
Coef	Sector 0	Sector 1	Sector 2	Sector 3	Sector 4	Sector 5
a_1	0.250469	0.251445	0.250031	0.25105	0.25027	0.249087
a_2	0.015964	0.0159875	0.0178331	0.0152426	0.01694	0.0150609
a_3	-0.003894	-0.003758	-0.004224	-0.003909	-0.003813	-0.00343
a_4	0.00510237	0.00431604	8.4798e-08	0.00255027	0.00399372	0.00552327
a_5	0.0248174	0.0244632	0.024983	0.0254091	0.0249107	0.0246847

Table 4.14: Parameters extracted from fits on Carbon target for reconstruction data.

Sampling Fraction Coefficients for **Iron**						
Coef	Sector 0	Sector 1	Sector 2	Sector 3	Sector 4	Sector 5
a_1	0.249113	0.248658	0.249016	0.247588	0.247964	0.246463
a_2	0.0172746	0.0185365	0.0185937	0.0196735	0.0196745	0.0184054
a_3	-0.004206	-0.004289	-0.004387	-0.005154	-0.004551	-0.004314
a_4	0.00431529	0.00540396	0.00417432	0.00431868	0.0038769	0.00444093
a_5	0.0251379	0.0241749	0.0249617	0.0250086	0.0251642	0.025143

Table 4.15: Parameters extracted from fits on Iron target for reconstruction data.

Sampling Fraction Coefficients for **Lead**						
Coef	Sector 0	Sector 1	Sector 2	Sector 3	Sector 4	Sector 5
a_1	0.247887	0.247159	0.247663	0.249099	0.249669	0.247958
a_2	0.0188033	0.0206521	0.0197017	0.0173088	0.0175364	0.0165727
a_3	-0.004632	-0.004939	-0.004592	-0.004369	-0.003922	-0.003811
a_4	0.00358101	0.00417953	1.8082e-07	0.00339505	0.00482576	0.00332387
a_5	0.0253111	0.0245174	0.0253269	0.0254005	0.0250319	0.0253196

Table 4.16: Parameters extracted from fits on Lead target for reconstruction data.

Sampling Fraction Coefficients for **Deuterium**						
Coef	Sector 0	Sector 1	Sector 2	Sector 3	Sector 4	Sector 5
a_1	0.24876	0.248957	0.248164	0.250525	0.248031	0.249069
a_2	0.0179894	0.0181533	0.0188304	0.0161461	0.0196832	0.015629
a_3	-0.004429	-0.004166	-0.004274	-0.004166	-0.004568	-0.003634
a_4	0.00399225	0.00538627	0.00335732	0.00346977	0.00354094	0.00274538
a_5	0.0249298	0.0242045	0.0247301	0.0251755	0.0252441	0.0254185

Table 4.17: Parameters extracted from fits on Deuterium target for reconstruction data.

Analogous to the data section in tables 4.18 and 4.19 (also in figures 4.42 and 4.43) are the information about the effects of the each cut.

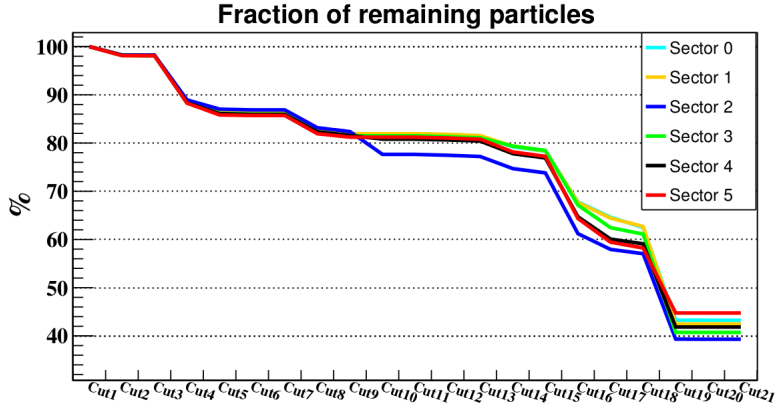


Figure 4.42: Fraction of rejected particles after the icremental application of each cut for selecting electron in the reconstructed events.

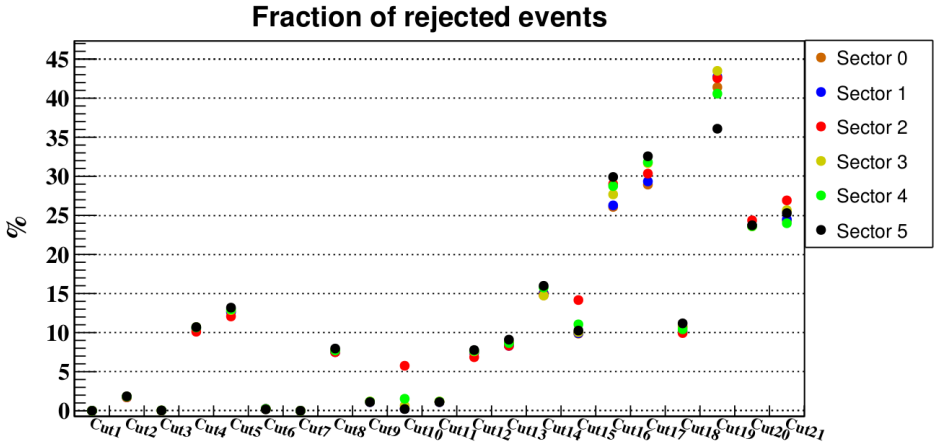


Figure 4.43: Fraction of rejected events for each cut in the electron selection individually for each sector.

Fraction of remaining particles in Electron Identification							
N^o	Cuts	Sector 0	Sector 1	Sector 2	Sector 3	Sector 4	Sector 5
0	No cuts	100	100	100	100	100	100
1	numberDC $\neq 0$	100	100	100	100	100	100
2	numberEC $\neq 0$	98.2725	98.2524	98.28	98.1615	98.1994	98.1423
3	numberSC $\neq 0$	98.2468	98.2145	98.2411	98.1219	98.1563	98.1045
4	numberCC $\neq 0$	88.6088	88.5294	88.952	88.2904	88.3726	88.268
5	StatCC > 0	86.4645	86.4852	87.0313	86.1479	86.0907	85.8217
6	StatSC > 0	86.2841	86.3059	86.8765	86	85.9338	85.7103
7	StatDC > 0	86.2841	86.3059	86.8765	86	85.9338	85.7103
8	StatEC > 0	82.6856	82.7124	83.1713	82.3038	82.3332	81.9375
9	DCStatus > 0	81.9095	81.9465	82.3762	81.501	81.5716	81.2222
10	SCStatus $= 33$	81.9095	81.9465	77.6457	81.501	80.8035	81.2163
11	$0 <$ Status <100	81.9095	81.9465	77.6457	81.501	80.8035	81.2163
12	Charge=-1	81.7534	81.7935	77.473	81.3281	80.6545	81.0526
13	Ein$>$ 0.06	81.4868	81.5195	77.2093	81.0857	80.399	80.7892
14	Nphe cut	79.199	79.3441	74.6997	79.3687	77.7966	78.1587
15	ΔT	78.2808	78.3501	73.8129	78.4729	76.9036	77.2463
16	(Etot,P)	67.8033	67.6716	61.2083	67.1215	64.7125	64.3464
17	(Ein/Eout,P)	64.6135	64.4294	57.9387	62.4505	60.0838	59.4362
18	Eout $\neq$0	62.4803	62.692	57.0565	61.1166	59.1083	58.2378
19	Fiducial (DC)	43.2492	42.4836	39.3402	40.7315	41.8475	44.7609
20	Fiducial (EC)	43.2485	42.4821	39.3122	40.73	41.846	44.7565
21	Sampling Fraction	43.2411	42.4814	39.3114	40.73	41.846	44.7565

Table 4.18: Effect in percentage of each cut in terms of remaining particles.

Fraction of rejected events in Electron Identification							
N^o	Cuts	Sector 0	Sector 1	Sector 2	Sector 3	Sector 4	Sector 5
1	numberDC $\neq 0$	0	0	0	0	0	0
2	numberEC $\neq 0$	1.72746	1.74756	1.71998	1.83854	1.80063	1.8577
3	numberSC $\neq 0$	0.0257721	0.0378954	0.038896	0.0395544	0.0430808	0.037838
4	numberCC $\neq 0$	10.4583	10.5735	10.1153	10.7112	10.668	10.7125
5	StatCC > 0	12.6364	12.6541	12.0725	12.8896	12.9878	13.1982
6	StatSC > 0	0.242257	0.247049	0.22793	0.216084	0.232928	0.175365
7	StatDC > 0	0	0	0	0	0	0
8	StatEC > 0	7.56079	7.5113	7.49681	7.81272	7.71584	7.96199
9	DCStatus > 0	1.13765	1.11135	1.15288	1.19615	1.16391	1.11258
10	SCStatus $= 33$	0.242257	0.250692	5.76127	0.643125	1.5385	0.224117
11	$0 <$ Status <100	1.13765	1.11135	1.15288	1.19615	1.16391	1.11258
12	Charge=-1	7.42309	7.46684	6.85736	7.60396	7.68663	7.77571
13	Ein$>$ 0.06	8.59388	8.31657	8.31518	8.80744	8.65851	9.10077
14	Nphe cut	15.0767	14.9847	14.7851	14.7486	15.7763	15.9909
15	ΔT	9.87364	9.93223	14.1682	10.0644	11.0601	10.2599
16	(Etot,P)	26.0828	26.2841	29.0203	27.6873	28.7553	29.9189
17	(Ein/Eout,P)	28.9435	29.3499	30.3482	31.694	31.8236	32.5676
18	Eout $\neq$0	10.7145	10.1057	9.94881	10.8005	10.4321	11.1884
19	Fiducial (DC)	41.4047	42.7423	42.5654	43.4966	40.5806	36.0916
20	Fiducial (EC)	23.6691	23.6656	24.3668	23.6015	23.6258	23.7346
21	Sampling Fraction	24.3737	24.4957	26.9222	25.6737	23.9975	25.2969

Table 4.19: Effect in percentage of each cut in terms of rejected particles, for all different CLAS sectors.

4.7.2.3 Positive Pion PID

For the pion case, TOF and CC based cuts remain the same as in data. The distribution of ΔT as a function of P for each sector is shown in figure 4.44. The final distributions of π^+ can be seen in figure 4.45.

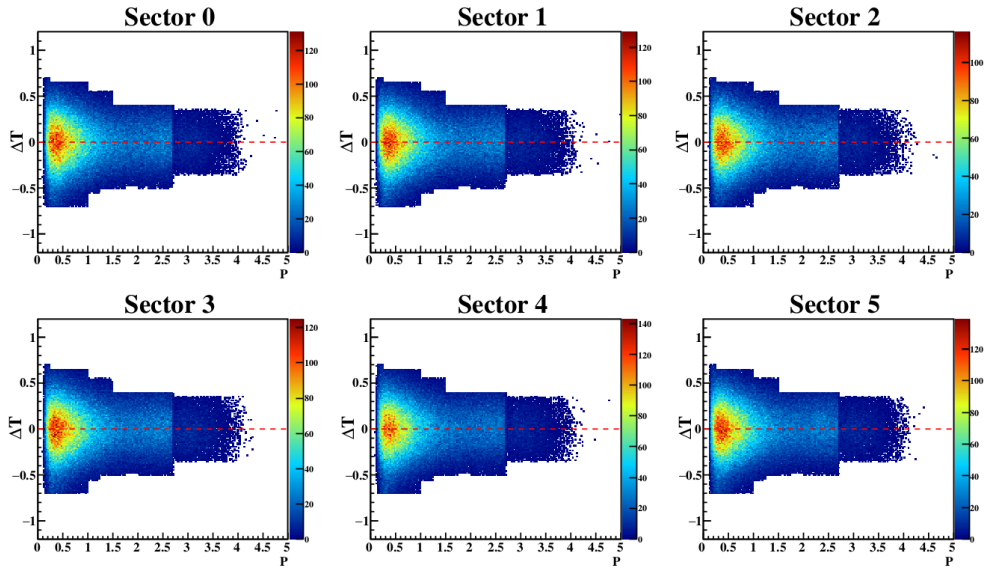


Figure 4.44: Positive pions in reconstruction, with all cuts applied (DIS included).

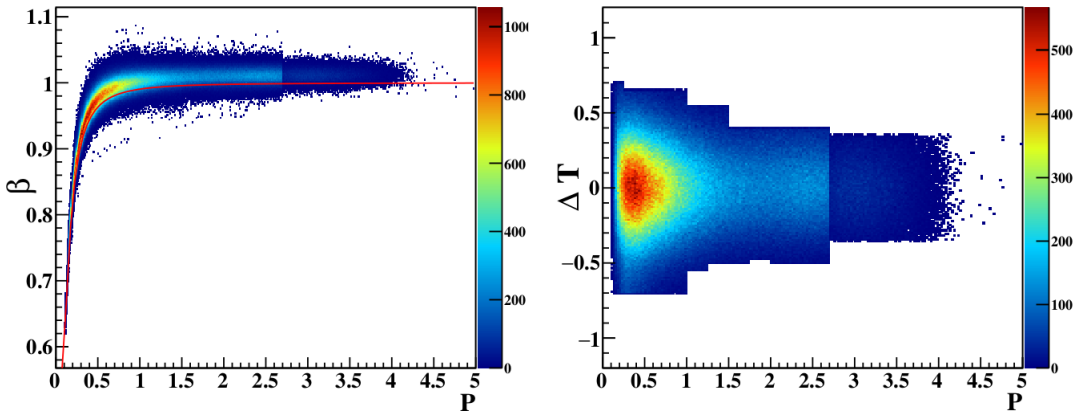


Figure 4.45: Two-dimensional view of βv/s P (left) and ΔT v/s P (right) for all the fully selected positive pions in the reconstruction.

4.7.2.4 Y Coordinate Shift and Vertex Determination

During the making of simulations there was an implementation of a shift on the Y position of the target for some runs. The Y shift corresponds to the $Y_{beam} = 2$[mm] instead of $Y_{beam} = 0$[mm]. In table 4.20, it is shown the percentage of the total

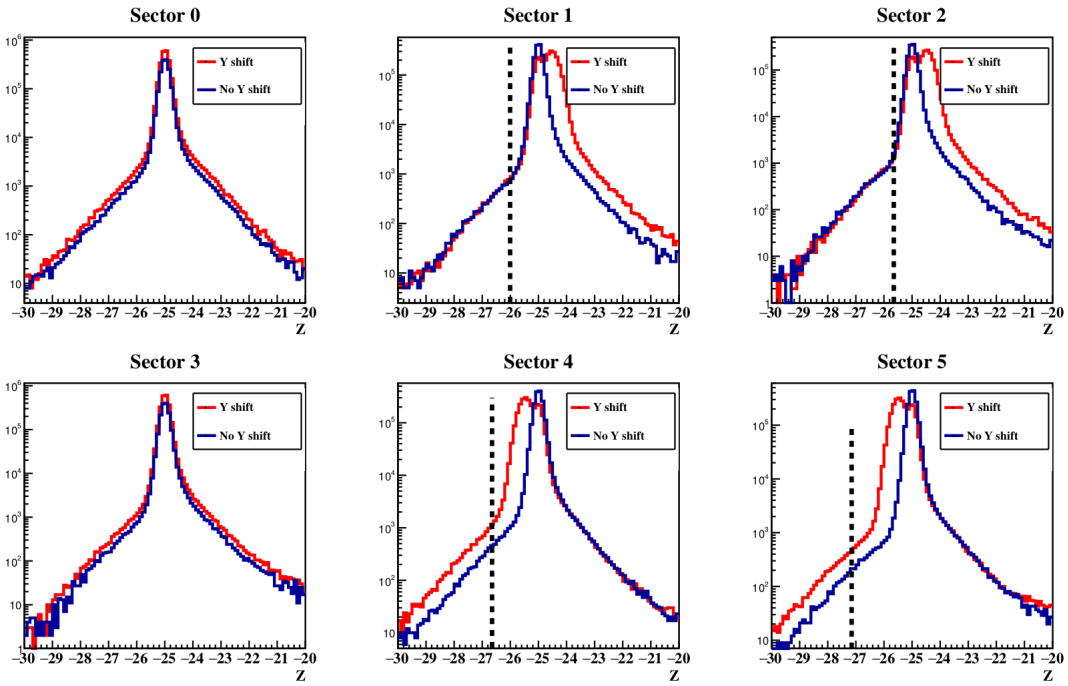


Figure 4.46: Simulated Z distributions in all sectors with and without 2 [mm] shift in Y coordinate. In blue is the case with no shift and in red is the case with the shift. The dashed black line is the lower limit in the vertex cuts for simulated data.

simulation set with and without this Y shift, for each target. Roughly speaking, there is $\sim 60\%$ of the simulations with the shift implemented.

Simulation Set		
Target	With Y shift	Without Y shift
C	63.93%	37.07%
Fe	56.31%	43.69%
Pb	60.08%	39.92%
D2	59.56%	40.44%

Table 4.20: Percentage of the total simulation set that contain the Y shift and the ones that does not for each target.

The distributions of the reconstructed Z vertex changes with and without the shift. In figure 4.46, there is a comparison for each sector of the Z distribution.

A vertex cut for electrons must be implemented in MC reconstructed events. In principle, the simulation has perfect alignment of beam with CLAS, but due to the Y shift implementation we need to change the vertex selection. For this purpose, a different set of vertex cuts are implemented in the simulation set. Such set is depicted in the table 4.21 and it is the one used in Ref [61]. In figures 4.47 and 4.48, there

Vertex Cuts for simulated data		
Sector	Liquid Target	Solid Target
0	$[-32.50\ ;\ -28.00]$	$[-26.50\ ;\ -20.0]$
1	$[-32.50\ ;\ -27.50]$	$[-26.00\ ;\ -20.0]$
2	$[-32.00\ ;\ -27.25]$	$[-25.65\ ;\ -20.0]$
3	$[-32.00\ ;\ -27.75]$	$[-25.85\ ;\ -20.0]$
4	$[-32.50\ ;\ -28.35]$	$[-26.65\ ;\ -20.0]$
5	$[-33.50\ ;\ -28.75]$	$[-27.15\ ;\ -20.0]$

Table 4.21: Vertex cuts edges, for the simulated electron tracks. Target independent.

are the Z distributions of electrons for simulated data in solid and liquid targets, respectively. The coloured lines represent the cut for each sector.

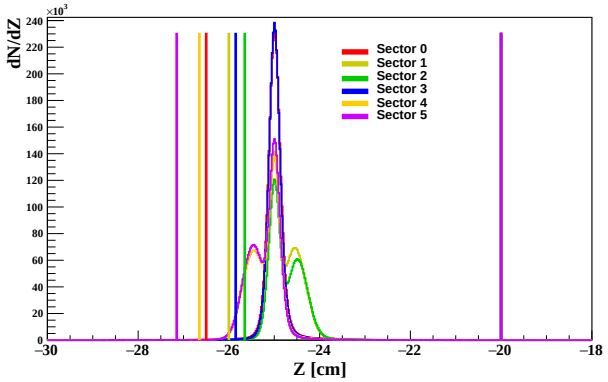


Figure 4.47: Solid target Z distribution of simulated electrons for each sector. The coloured vertical lines represent the vertex cuts according to table 4.21.

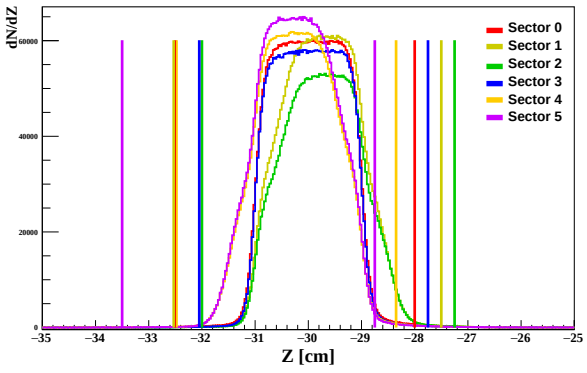


Figure 4.48: Liquid target Z distribution of simulated electrons for each sector. The coloured vertical lines represent the vertex cuts according to table 4.21. For the case when two or more sectors share the same edge, the lines were shifted slightly to make the plot easier to read.

4.8 Corrections

The aim of this section is to explore the different corrections implemented in this thesis. They are Acceptance Correction (AC), Radiative Corrections (RC), and Coulomb Corrections (CC).

4.8.1 Acceptance Correction (AC)

The Acceptance Correction consists in the application of a simulation based factor to each bin in experimental data to correct for the inefficiencies[20] of some parts of the detector. The geometry of CLAS does not cover the full momentum space range of the phenomena under study. Basically, the CLAS acceptance correction is a combination of:

- Geometrical acceptance.
- Detector efficiencies, drift chambers, and scintillator counters.
- Track reconstruction efficiencies.
- Event selection efficiencies.

In principle, acceptance could be calculated analytically. In practice, it is always calculated using MC calculations because of effects such as the complexity of detector geometries, magnetic field effects, and time-varying inefficiencies. Acceptance is a function of kinematic variable used to describe the observable.

The acceptance correction factor [21], **A**, is defined as the ratio of the number of events in the reconstruction [22] over the number of events in the generation, for each specific kinematical bin. Thus, the definition is:

$$\mathrm{A} = \frac{\mathrm{N_{rec}}}{\mathrm{N_{gen}}} \tag{4.66}$$

where N represents the number of counts in the bin for the respective case (reconstructed or generated). In the acceptance correction factor application, each single bin is treated as an independent identity. Therefore, for each bin in this n-dimensional

[20] Efficiency refers to the fact that the detectors that compose CLAS, the reconstruction protocol, and our analysis does not have a perfect efficiency in detecting each particle in the experiment.

[21] for the acceptance correction factor is used the word A; the same one to denote the mass number of the targets. The distinction will be evident from the context.

[22] "Reconstructed" or "found" events means generated events after passing through the Monte Carlo, GSIM, GPP, and RECSIS sequence.

space there must be a A factor. This means that the binning scheme used in the simulation to get the acceptance factors must be the same as in experimental data. This is called a *bin by bin basis* application. The main advantage of this procedure is that it should be, in principle, independent of the model used for the event generator if the chosen bins are as small as possible[23]. The data is corrected using weights defined as $\omega = 1/A$.

4.8.1.1 Independent Variables

In order to bin the data distribution and to be able to treat each bin as an independent entity, a complete set of independent variables -that fully delineate the reaction under study- must be chosen. With the independent variables determined, all other kinematical quantities can be constructed. For a single pion electro-production at fixed beam energy, one has:

- Incoming electron.
- Proton inside the nucleus.
- Outgoing (scattered) electron.
- Identified hadron.

For the incoming electron, it is fully known. There are no variables necessary to describe it (fixed energy and fixed direction). For the case of the proton inside the nucleus, we may consider it at rest in the target, so it is completely known, no variables needed[24]. For the scattered electron, we need three variables[25]. However, one of these can be chosen to be the electron azimuthal angle, ϕ_{el}, in the lab frame, and in the absence of any transverse polarization of the target or the beam, the cross section of the interaction is uniform in ϕ_{el}, so averaging over 2π does not introduced any uncertainty. Thus, the number of independent variable for the outgoing electron are just two. For the identified hadron, three variables must be chosen. In total, we end up with five variables needed to fully describe the interaction, which are:

[23] this is true only if the aceptance is 5-fold differential.

[24] we may consider the proton inside the nucleus in motion as well; this is known as Fermi Motion. If that is the case, then we need some variables to describe it. However, the kinematics of this particle can be consider as fully known if we average the acceptance over the Fermi motion. At the end, the result is the same, no additional variables were needed to described the proton inside the nucleus.

[25] In principle, it should be 4 because of the four-vector of the particle. Therefore, since the particle is physical, it must obey $p^2 = m^2$. This is the mass shell condition, which reduces the numbers of variables to three.

$$\underbrace{Q^2\ ,\ X_b}_{\textbf{Electron}} \quad \text{and} \quad \underbrace{P_t^2,\ Z_h,\ \phi_{PQ}}_{\textbf{Positive Pion}}$$

Technically, the acceptance can depend on more variables than there are independent variables in the event. One example of this could be the azimuthal ϕ angle in the lab frame. The physics does not typically depend on this, but the acceptance can. For example if there are non-functional channels. The kinematic distributions from the simulation are compared to the experimental data, in order to check the ability of the simulation to reproduce the experiment which will be essential for the acceptance correction.

4.8.1.2 Comparison of Real Data to Simulation

Comparisons between the MC and the data are shown as one dimensional plots of the distributions of some variables of interest in figures 4.49 (solid target) and 4.50 (deuterium). The data (in blue), the reconstructed (in red) and the generated (in green) distributions are normalized to unit area. Note the logarithmic scale of the plots. The agreement is reasonable, but not perfect. There are some important issues at the edges of some variables. See for example high Zh or high Xf.

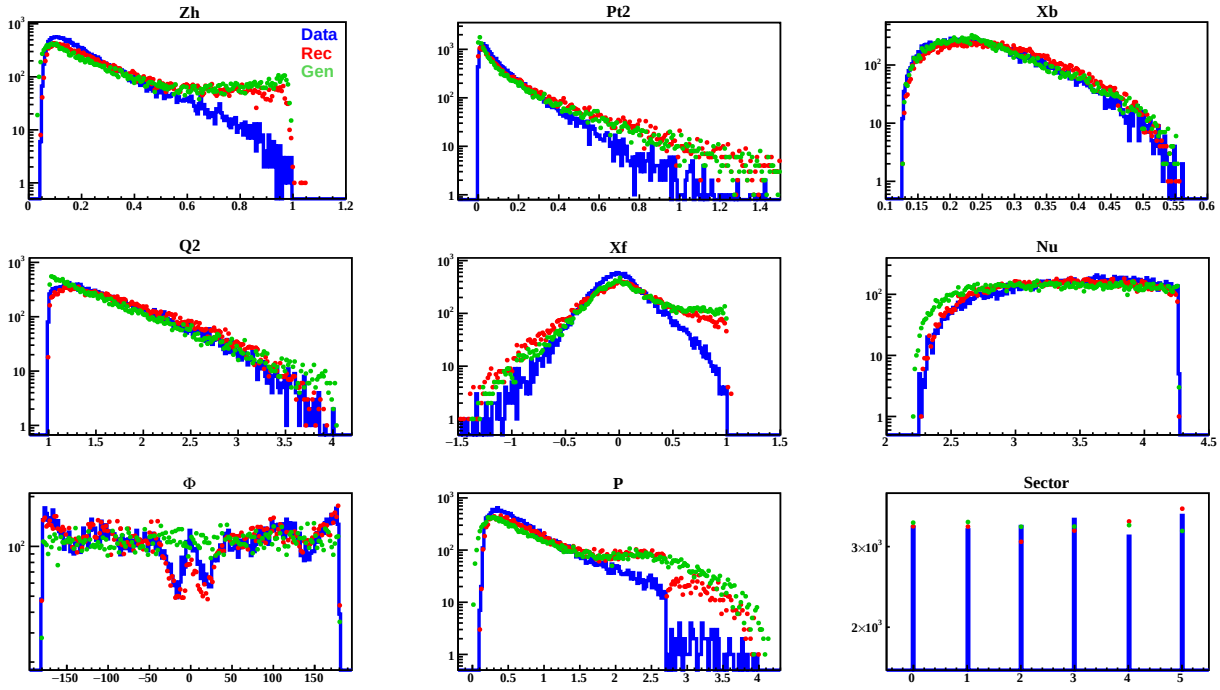


Figure 4.49: Comparison between the data events in blue, reconstructed events in red and generated events in green. Solid target.

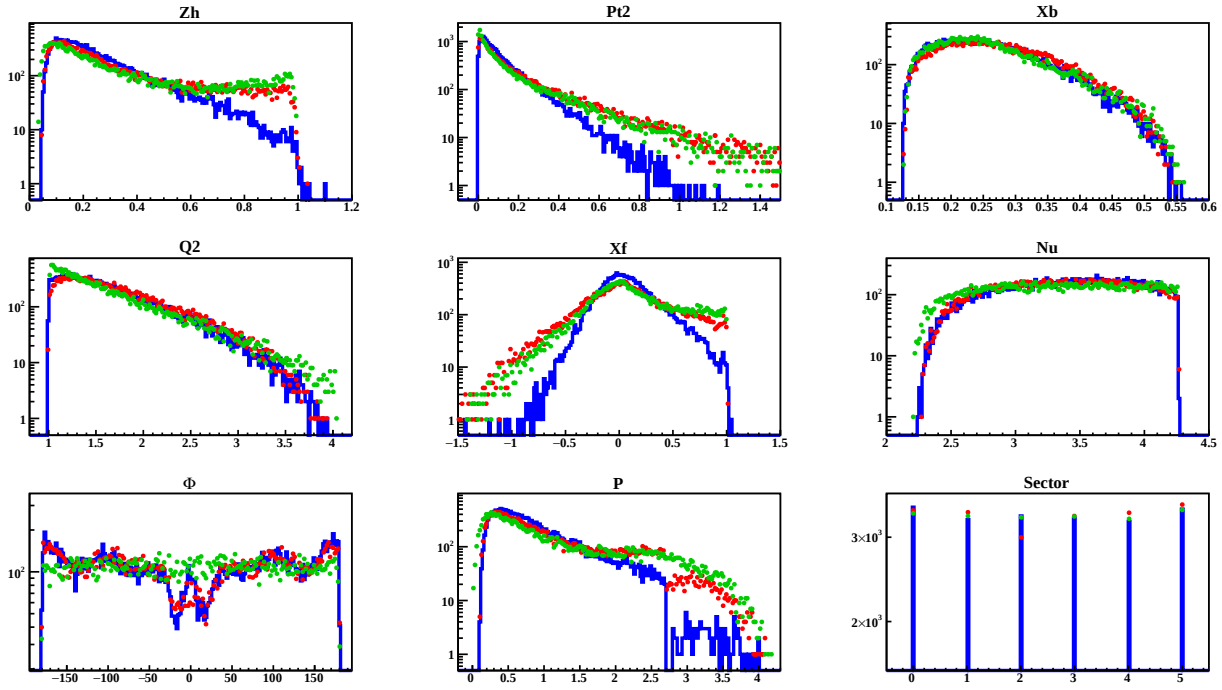


Figure 4.50: Comparison between the data events in blue, reconstructed events in red and generated events in green. Liquid target.

In figure 4.51, there are the distributions of the missing mass for the case of the simulation, the reconstructed events in blue, and the generated ones in red. The cut in 1.4 GeV is optimal to get ride of those events as it was for data. If the cut in W_x is imposed in data and in the simulations, we end up with a much better agreement between data, generated and reconstructed events. The problem at the edge of some variables in figures 4.49 and 4.50 vanishes as it is depicted in figure 4.52. The data in blue, reconstructed in red and generated in green events are compared for some kinematical variables with the cut at 1.4 GeV. In the same plot, it is the previous case (with no W_x cut) with light blue for data, orange for reconstruction, and light green for generation.

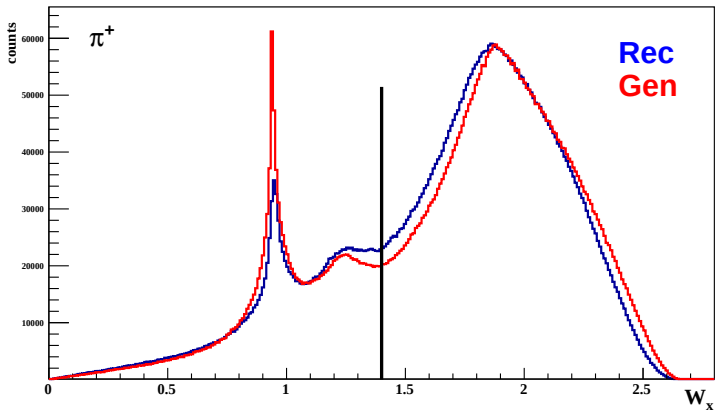


Figure 4.51: Missing mass distribution for simulations. Generated events are in red and reconstructed are in blue. The value of $W_x = 1.4$ GeV is represented by a black line.

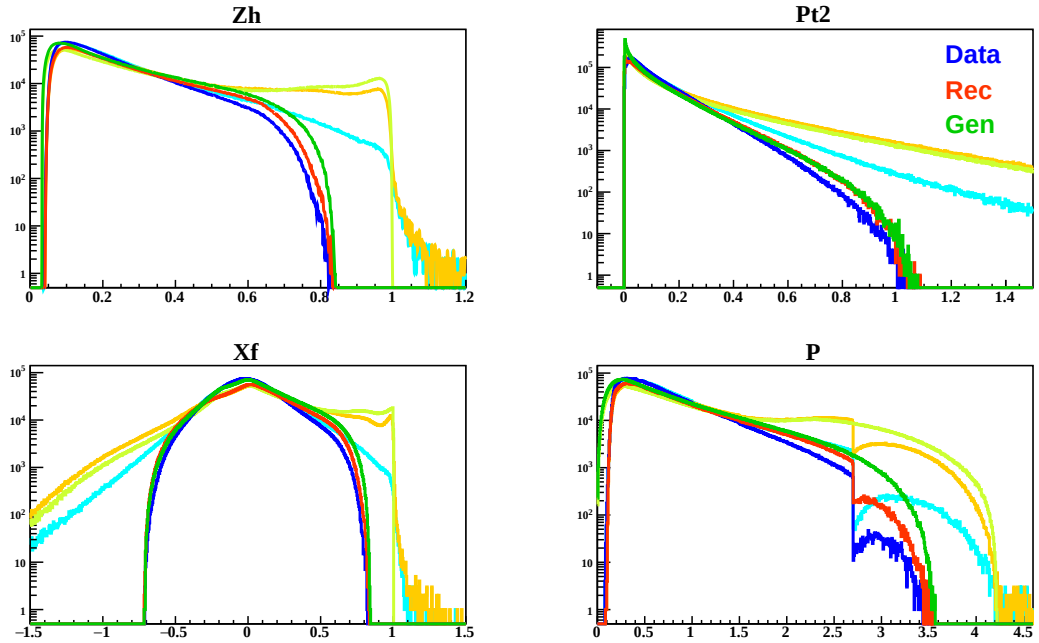


Figure 4.52: Comparison between data in blue, generated in green and reconstructed in red events with the cut in W_x applied. In dimmer colors are the case with no cut in the missing mass, just for comparison purpose.

The shape of the ϕ_{PQ} variable has some interesting features that are worth to highlight. Following the definition of this angle, we can characterize the different peaks by the relationship between the sectors in which the electron and pion were measured. In figure 4.53, there are the distributions of ϕ_{PQ} for data and for simula-

tions, reconstructed and generated events in gray. The central peak in green are those events in which the electron and the π^+ were measured in the same sector. The yellow peaks are those events in which the scattered electron and the pion were measured in adjacent sectors and so on, until we reach the extremes of the distribution were the electron and the pion have opposite sectors (in red). If we add up all the coloured plots, we obtain the gray one.

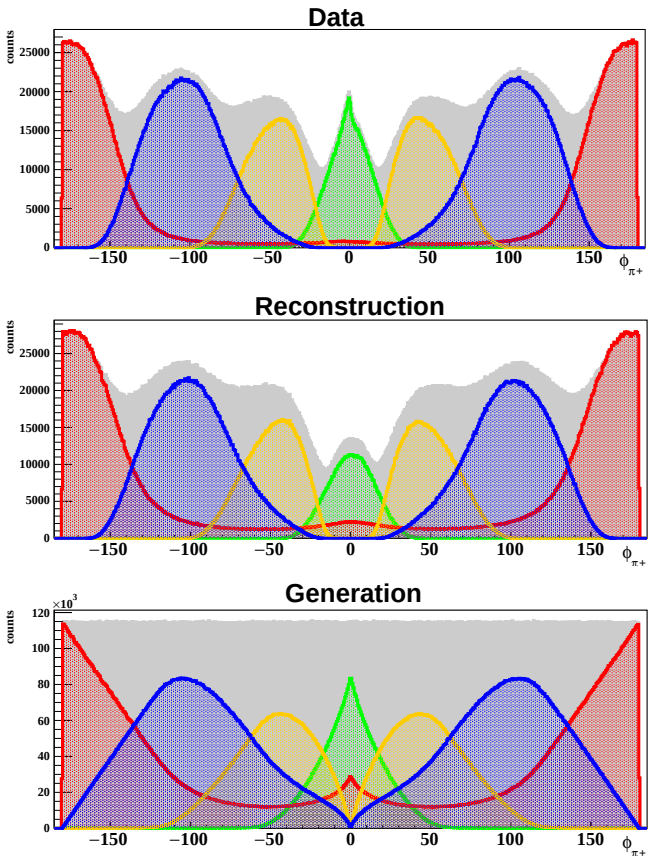


Figure 4.53: Distribution of the azimuthal angle in degrees between leptonic, and hadronic planes for data (top) and for simulations (middle for reconstruction and bottom for generated events). The gray plot is the usual distribution. The coloured plots are based on relationships between the sectors in which the scattered electrons and the produced pions were measured in.

4.8.1.3 Statistical Errors of the Acceptance Correction

Let's consider a generated event. This event, independent of others, will have two possible outcomes: being part of the reconstructed bin or not. This fact gives us the dichotomic feature which allows us to use the binomial statistical distribution to explain this behavior in the correction. Along with the notion that the probability

of success is the same from one trial to another. It can be shown that the maximum likelihood estimator of the probability of success, p, is:

$$\hat{p} = \frac{x}{N} \tag{4.67}$$

where x is the number of successes and N is the number of trials. For the acceptance case $\hat{p}$ is given by the acceptance factor itself:

$$\hat{p} = \frac{x}{N} = \frac{N_{rec}}{N_{gen}} = A \tag{4.68}$$

The goal now is to find an expression for the statistical error of A:

$$\delta A = \delta \left(\frac{N_{rec}}{N_{gen}} \right) \tag{4.69}$$

$$\delta A = |A| \sqrt{\left(\frac{\delta N_{rec}}{N_{rec}} \right)^2 + \left(\frac{\cancelto{0}{\delta N_{gen}}}{N_{gen}} \right)^2} \tag{4.70}$$

since the number of generated events is considered the number of trials, these events will not carry any error. Following the Central Limit Theorem [26] in order to see what is the distribution of $\hat{p}$, it can be shown that the variance is:

$$\sigma^2 \left(N_{rec} \right) = N_{gen} A (1 - A) \tag{4.71}$$

so, replacing this into the previous expression:

$$\delta A = \sqrt{\left(\frac{N_{rec}}{N_{gen}} \right)^2 \left(\frac{\sqrt{N_{gen} A (1 - A)}}{N_{rec}} \right)^2} \tag{4.72}$$

$$\delta A = \sqrt{\frac{A(1 - A)}{N_{gen}}} \tag{4.73}$$

gives the final expression for the error in the acceptance correction factor.[27]

[26] since the number of trials is big we can approximate this to a normal distribution.

[27] The expression used by ROOT in the TH1F::Divide() method with the option B uses normal error calculation using the variance of Maximum Likelihood Estimation (MLE) with weights, standard binomial statistics. The final expression is:

$$\delta \left(\frac{N_1}{N_2} \right) = \sqrt{\left| \frac{\left(1 - 2(N_1/N_2) \right) (\delta N_1)^2 + (N_1/N_2)^2 (\delta N_2)^2}{N_2^2} \right|} \tag{4.74}$$

which is equivalent to our previous expression.

The total statistical error on the acceptance corrected yields will be calculated with an account for statistical error on the number of events before correction (δN_{meas}) and statistical error on the acceptance (δA):

$$\frac{\delta N_{cor}}{N_{cor}} = \sqrt{\left(\frac{\delta N_{meas}}{N_{meas}}\right)^2 + \left(\frac{\delta A}{A}\right)^2} \tag{4.75}$$

4.8.2 Radiative Corrections (RC)

Radiative corrections (RC) must be applied to the measured cross sections to eliminate the effects of the radiation of photons by electrons. These corrections also remove higher order electro-dynamic contributions to the electron-photon vertex and the photon propagator [34]. The purpose of RC is to account for processes other than the one-photon exchange (Born cross section). This leads to the modification of the kinematical variables due to photon emission. Since this modification affects both leptonic and hadronic components of the MR [28], there are two distinct types of radiative corrections implemented in this thesis. One concerning the DIS electrons and the other concerning the SIDIS hadrons.

4.8.2.1 Radiative Corrections for Electrons

In the DIS process that we are considering, there is only one photon exchange. There are non negligible QED processes that are important to consider. The measured yields need to be corrected for radiative processes, see figure 4.54 for a schematic representation. As is depicted in the diagrams, for example top left and right, the electron can radiate energetic photons before the interaction. In this case, the energy of the incoming electron when interacts with the nucleon is going to be different then the energy measured in the detector. In the same way, after the interaction, the electron can lose some energy through the emission of an energetic photon leading again to a difference between the real electron energy and the one measured in the detector. For the two bottom diagrams, there are the vertex correction (left) and the Vacuum Polarization case (right)[29].

[28] the nuclear effects in solid and liquid target, in addition to the difference in radiation lengths. Each target leads to the fact that RC does not cancel in the ratio.

[29] this case account for loops, that means when the electron spontaneously splits in $e^{\pm}$, $\tau^{\pm}$, $\mu^{\pm}$, and $q\bar{q}$ pairs. Those charged pairs act as electric dipole, creating a partial screening of the field.

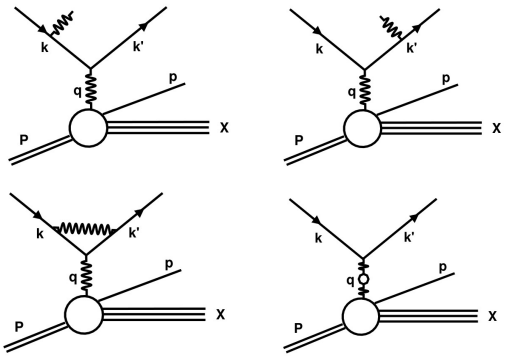


Figure 4.54: Feynman diagrams for higher order internal radiative corrections. On the top are Bremsstrahlung radiation and, at the bottom and in the left, is the vertex correction and, in the right, is the Vacuum Polarization case.

It is important to make a distinction though between two types of radiative processes, the one called "Internal" and "External":

- External radiation: associated with energy loss through the emission of a photon as it traverses a material. Materials such as the target, interactions in the target chambers and walls, and any component of the detector. This features are presented in the simulation set and are accounted through the acceptance correction factors.

- Internal radiation: associated with the scattering off nuclei process itself. These are the ones that we need to consider in the radiative correction factors.

The idea is to construct the factors, based on Born first approximation, and next, to leading order processes calculable in QCD. The radiative factors δ_{Rad} connect Born and measured, or radiated, cross sections as:

$$\sigma_{Rad} = \underbrace{(1+\delta_{Rad})}_{\delta_{RC}}\sigma_{Born} \tag{4.76}$$

where it is defined the Radiative Correction factor δ_{RC}, which will be eventually the number we are going to use to correct our data.

The corrected expression for the electron yield is:

$$N^e_{\text{Radiative Corrected}} = \frac{N^e_{\text{Measured}}}{\delta_{RC}} \tag{4.77}$$

$$N^e_{\text{Coulomb Corrected}} = N^e_{\text{Measured}} \times \delta_{CC} \tag{4.78}$$

The calculation of the δ_{RC} factors are done using a code called EXTERNAL. The code is based on Mo and Tsai work [38] [39] about the treatment of the radiative effects. Since the original version of EXTERNAL, developed in SLAC by S.Dasu, some important modifications have been implemented throughout the years by SLAC analysis groups. As a brief description of the process to calculate δ_{RC}, let's mention the two main approximations used in the code:

- Angular peaking approximation: the Bremsstrahlung photons are collinear to the initial and scattered electrons.
- Equivalent radiator method: the effect of internal Bremsstrahlung can be computed by using two hypothetical radiator placed before and after interaction vertex.

The nuclear target is included in the process by the empirical fit function f_{EMC}, which is defined as follows:

$$f_{EMC} = cA^{\alpha} \tag{4.79}$$

where α is a sum of 9 different coefficients and c is a Bjorken-x dependent parameter. This function is for $A > 2$ and is based on world data. The range of validity is for $0 < W < 3.2\text{GeV}$ and $0.2 < Q^2 < 5\text{GeV}^2$. For details of this refer to [40]. In figure 4.55, the f_{EMC} fit is illustrated for the three different targets.

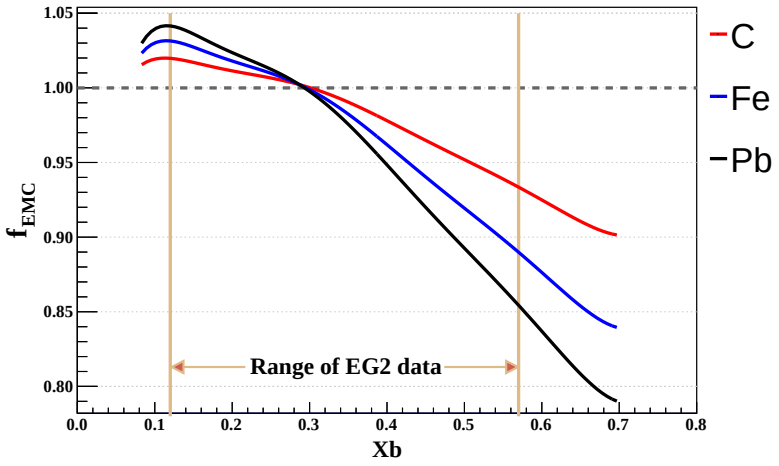


Figure 4.55: Parametrization function f_{EMC} for the three targets.

4.8.2.2 SIDIS Radiative Corrections for Positive Pions

The photon radiation from incident and scattered electron affects both hadronic and leptonic components in the multiplicity ratio. For this reason, we must have a radiative correction for pions as well.

$$e(k_1) + N(P) \longrightarrow e'(k_2) + h(P_h) + X(P_x) \tag{4.80}$$

$$e(k_1) + N(P) \longrightarrow e'(k_2) + \gamma(k) + h(P_h) + X(P_x) \tag{4.81}$$

The scattering cross section is dependent upon structure functions, which define aspects of the nucleons' internal structure including charge and momentum distributions of the quarks. For the SIDIS case, the cross section depends on four structure functions, as is depicted in equation 4.82. Assuming single photon exchange, the unpolarised SIDIS [27] cross section can be written in term of four structure functions as [30]:

$$\begin{aligned}\sigma_{SIDIS} &= \frac{d^5\sigma_{UU}}{dxdydzdP_t^2d\phi} \\ &= 2\pi \times \left(\frac{\alpha^2}{xyQ^2}\right) \times \left(\frac{y^2}{2(1-\epsilon)}\right) \times \left(1+\frac{\gamma^2}{2x}\right) \\ &\times \left(F_{UU,T} + \epsilon\, F_{UU,L} + \sqrt{2\epsilon(1+\epsilon)}F_{UU}^{cos\phi_h} cos(\phi_h) + \epsilon\, F_{UU}^{cos2\phi_h} cos(2\phi_h)\right)\end{aligned} \tag{4.82}$$

[30] this expression is after integrating over θ_{lab} angle, which is perfectly uniform for this experiment. The results of that is the extra 2π factor.

The first subscript on the structure functions (SF) represents the polarization of the beam. The second subscript represents the target polarization. Finally, the third subscript -if there is one- represents the virtual photon polarization, where U, L, and T stand for **U**npolarized, **L**ongitudinally polarized, and **T**ransversely polarized, respectively. ϵ is the ratio of longitudinal and transverse virtual photon flux. $F_{UU,T} + \epsilon\, F_{UU,L}$ are the experimentally measured SF and are ϕ_h independent. The other two SF , $F_{UU}^{cos\phi_h}$ and $F_{UU}^{cos2\phi_h}$ are $cos\phi_h$ and $cos2\phi_h$ modulations respectively.

Based on [44], the HAPRAD code calculates the radiative corrections for hadrons estimating the structure functions directly from the data through a fitting procedure on the ϕ_{PQ} spectrum, for each target separately.

$$RC factor = \frac{\sigma_{Rad}}{\sigma_{Born}} \rightarrow \frac{\text{lowest QED effects + vacuum polarization}}{\text{Born}} \tag{4.83}$$

The Born cross section has the following form:

$$\sigma_0 = \frac{N}{Q^4}\left(A + A_c cos(\phi_h) + A_{cc} cos(2\phi_h)\right) \tag{4.84}$$

In order to run the code the structure functions $\mathcal{H}_1$, $\mathcal{H}_2$, $\mathcal{H}_3$ and $\mathcal{H}_4$ are needed. These are derived from the A, A_c, and A_{cc} constants previously determined through the fitting on ϕ_{PQ}.

$$\begin{aligned} A &= 2Q^2\mathcal{H}_1 + SX - M^2Q^2 \\ &\quad + 4a^1a^2 + 2b^2 - M_h^2Q^2\mathcal{H}_3 \\ &\quad + 2Xa^1 + 2Sa^2 - zS_xQ^2\mathcal{H}_4 \end{aligned} \tag{4.85}$$

$$A_c = 2b2a^+\mathcal{H}_3 + S_p\mathcal{H}_4 \tag{4.86}$$

$$A_{cc} = 2b^2\mathcal{H}_3 \tag{4.87}$$

The final RC factor is applied to the data on a bin by bin basis. All the details of this procedure can be found in [44] and references therein. In this thesis a $HAPRAD_cpp$ version was used, this is a $C++$ version of the original code written in FORTRAN

4.8.3 Coulomb Corrections (CC)

In DIS experiments, the particles that approach the vicinity of a target are affected by the Coulomb field produced by the target nuclei. It is important to remove/correct

for apparent changes in the cross section due to Coulomb effects, mainly at relatively low energies of a few GeV -our case is ~ 5 GeV- and medium-weight to heavy nuclei.

The electron's wave function is distorted due to the effect of the electrostatic field V. The incoming electrons momentum is enhanced as it approaches the nuclei. The scattered electrons momentum gets reduced as it leaves the nuclei. The process is depicted in figure 4.56. Here, k_i is the incoming four momentum of the electron. k_i' is the enhanced four momentum of the incoming electron as it approaches attractive potential V; the wavelength of the electron becomes shorter near the nucleus. k_f is the momentum of the outgoing electron. Finally, k_f' is the decelerated outgoing electron's momentum as it leaves the potential V.

The wave function of the electron is modified, hence its kinematics. This also produces an effect in the produced hadron π^+,[31] but this correction is not include it in this thesis[32]. In scattering interactions, the primary interaction is via the exchange of a virtual photon between electrons and struck quark. As the electron passes through the positively charged target, soft photons are exchanged. This extends the standard one photon exchange to a one hard photon and several soft photon exchanges.

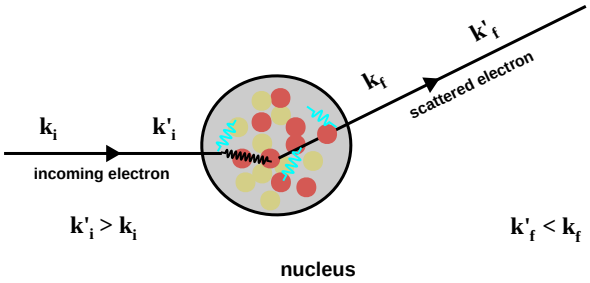


Figure 4.56: Representation of the interaction of the incoming electron off a nuclei A. As the electron scatters from a heavy nucleus, it interacts with the quark of a nucleon through the exchange of a hard photon (black). It is accelerated and decelerated by the Coulomb interactions through the exchange of soft photons (blue).

The electrostatic potential inside the nucleus can be approximated as that of the potential inside of a charged sphere:

$$V(r) = -\frac{3\alpha(Z-1)}{2R} + \frac{\alpha(Z-1)}{2R}\left(\frac{r}{R}\right) \tag{4.88}$$

[31] it also applied for the negative pions, but no to the π^0, since it has no charge.
[32] Coulomb corrections for SIDIS data are expected to be negligible, except at low Z_h.

where Z is the atomic number[33], r is the distance from the center of the charge sphere. R is the radius, given by: $R \approx 1.1A^{1/3} + 0.86A^{1/3}$, where A is the atomic number of the nucleus.[34]

Because most the nucleons of heavy nuclei are located in the nucleus peripherical region, taking the electrostatic potential at the center of the nucleus will be an overestimate of the Coulomb effect. The effective Coulomb potential seen by the electron can be approximate, according to [41], as $\bar{V} \sim (0.75...0.8)V_0$ [35]. This additional potential must be added to the incident and scattered energies of the electron [37].

$$\overline{E} \longrightarrow E + \overline{V} \tag{4.89}$$

$$\overline{E}' \longrightarrow E' - \overline{V} \tag{4.90}$$

This result is using the **E**ffective **M**omentum **A**pproximation (EMA) [36] tuned to agree with **D**istorted **W**ave **B**orn **A**pproximation (DWBA)[37]. To have a number in mind, for the heaviest nuclei ($A = 208$, $Z = 82$, Pb) the value of $V_0 \sim 25$ [MeV]. In most experiments CC are not considered because of the high values of the incident beam energies. In our case, the energy is rather small and the energy of the scattered particles are even smaller. Thus, $\overline{V}$ could be a non-negligible fraction of the beam energy and this could impact our results in an important way.

The implementation of this correction is based on the EXTERNAL code, as for the RC case, based on [41]. The code uses models to obtain a Coulomb corrected cross section, which is σ_{CC}, and a non corrected cross section, which is σ_{Born}. For the deuterium case, the cross section is not affected by Coulomb distortions. At the end, as an output of the code, we end up with a single factor called δ_{CC} and is given by:

$$\delta_{CC} = \frac{\sigma_{Born}}{\sigma_{CC} \cdot f^2} \tag{4.91}$$

[33] $Z-1$ is used instead of just Z because the acceleration is being calculated for the $A-1$ spectator nucleus.

[34] Note that in order to the units to make sense there is a factor of $\hbar c$ in the expression of $V(r)$. So, the potential has units of energy and the momentum k$(= V/c)$ has units of momentum.

[35] $V_0 = V(r = 0)$.

[36] In this approach, the electron's wave function is treated at the lowest order expansion in αZ, where α is the fine structure constant and Z is the atomic number. Only a change in the electron momentum magnitude is considered in this approach, not the change in the direction of it.

[37] the cross sections can not be described by the **P**lane **W**ave **B**orn **A**pproximation (PWBA) due to the changes in the initial and final kinematics and the wave functions of the lepton [43].

where f is the *focusing factor*, $f = \left(\frac{E+\overline{V}}{E}\right)$. The focusing factor accounts for the nucleus, acting like a lens to focus the wave function of the electron. This factor is implemented quadratically. Note that $\sigma_{CC} = \sigma_{Born}(k'_i, k'_f)$. The Coulomb correction is a ratio of the cross section model with experimental kinematics to the cross section model with shifted kinematics multiplied by a focusing factor [42]. Some of the main features of this correction is that its effect increases with the number of protons inside the nucleus. The effect is expected to have the opposite sign with a positron beam [45].

Chapter 5

Results and Discussions

The aim of this chapter is to show and discuss the observables. The effect of the cuts and corrections implemented are treated separately. Final results for EMC and multiplicity ratios with all cuts and corrections implemented are reported. At the end of the chapter multi-dimensional multiplicity ratios are reported and discussed.

5.1 Data and Simulation Sets

5.1.1 Run Numbers (RN)

During the EG2 experimental period, the data collected in each run was saved in separate files of about 10 million events each. Each run has a specific number associated; those are the *Run Numbers* (RN). Special runs -such as *cosmic* runs- are excluded from this analysis. Each run was studied separately in order to avoid runs with any kind of problems, such as, test runs, which only have a couple of events in them or runs with only deuterium and so on. The final list gives 118 files for carbon, 262 files for iron, and 169 files for lead. The size of each run varies. The explicit numbers are in tables 5.1, 5.2, and 5.3, for each target separately.

The ratio of scattered electron yield from the solid and liquid target was compared for every run as a data quality/stability check. The idea is that this ratio will give us a clear way to identify problematic runs. For example, if the ratio is off, -compared to the rest- it can indicate a problem with the detector or the fact that the beam could be hitting other materials than the target itself. The result of this ratios, for each target, are presented in figure 5.1 as a function of the run number.

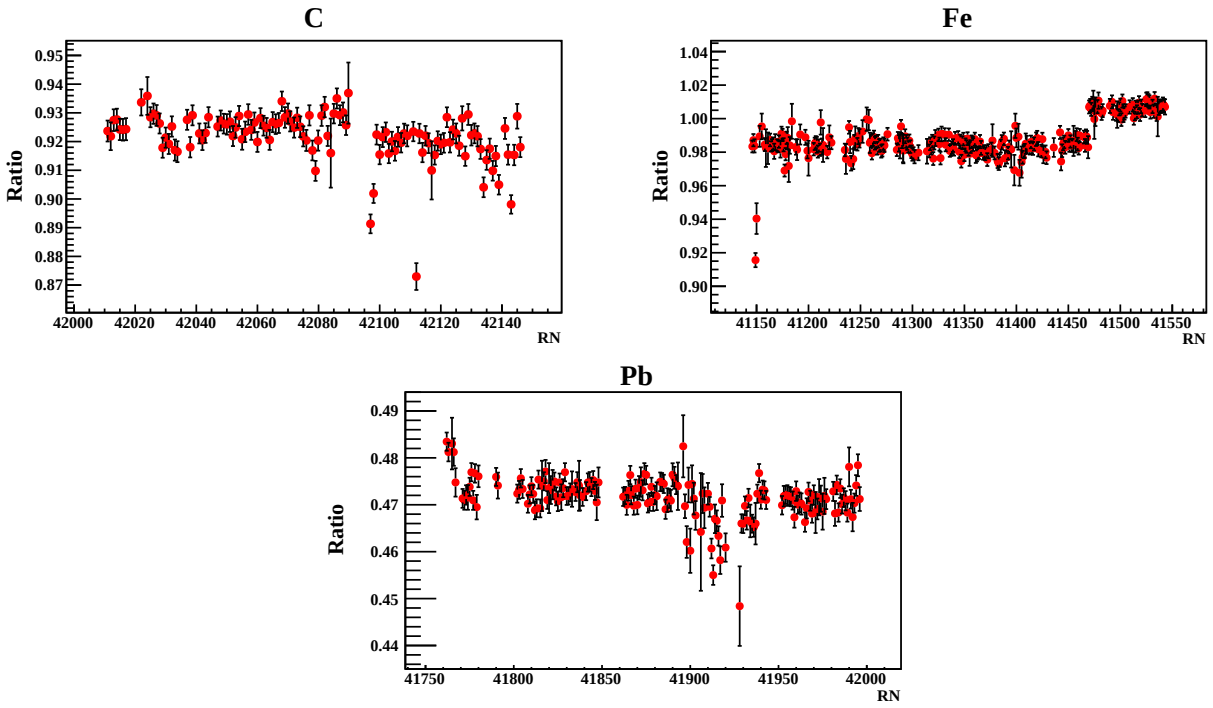


Figure 5.1: Ratio of the scattered electron yield from the solid and liquid target for each run number (RN).

Carbon Run Numbers									
42011	42012	42013	42014	42015	42016	42017	42022	42024	42025
42026	42027	42028	42029	42030	42031	42032	42033	42034	42037
42038	42039	42041	42042	42043	42044	42047	42048	42049	42050
42051	42052	42053	42054	42055	42056	42057	42058	42059	42060
42061	42062	42063	42064	42065	42066	42067	42068	42069	42070
42071	42072	42073	42074	42075	42076	42077	42078	42079	42080
42081	42082	42083	42084	42085	42086	42087	42088	42089	42090
42097	42098	42099	42100	42101	42102	42103	42104	42105	42106
42107	42108	42109	42111	42112	42113	42114	42115	42116	42117
42118	42119	42120	42121	42122	42123	42124	42125	42126	42127
42128	42129	42130	42131	42132	42133	42134	42135	42136	42137
42138	42139	42141	42142	42143	42144	42145	42146		

Table 5.1: Run numbers for carbon.

Iron Run Numbers									
41146	41147	41148	41149	41150	41153	41155	41158	41159	41161
41163	41164	41165	41166	41167	41168	41170	41171	41172	41173
41174	41175	41176	41177	41178	41179	41180	41181	41182	41184
41188	41189	41192	41197	41199	41200	41205	41206	41207	41208
41209	41210	41211	41212	41213	41214	41218	41220	41222	41235
41236	41239	41240	41241	41243	41244	41245	41247	41248	41249
41252	41256	41258	41259	41260	41261	41262	41263	41264	41265
41266	41267	41268	41269	41270	41271	41272	41273	41276	41285
41286	41287	41288	41289	41290	41291	41292	41293	41294	41295
41296	41297	41298	41299	41300	41301	41302	41306	41314	41316
41317	41318	41319	41320	41321	41322	41325	41326	41327	41328
41331	41332	41335	41336	41337	41338	41339	41340	41341	41344
41345	41346	41347	41348	41349	41350	41351	41352	41353	41354
41355	41356	41357	41358	41359	41360	41361	41362	41363	41364
41365	41366	41368	41369	41370	41371	41372	41373	41376	41377
41379	41382	41383	41384	41385	41386	41388	41389	41391	41392
41393	41395	41398	41399	41400	41402	41403	41405	41406	41407
41408	41409	41410	41413	41414	41415	41416	41417	41419	41420
41421	41424	41425	41426	41428	41429	41436	41442	41443	41444
41445	41446	41447	41450	41451	41452	41453	41454	41455	41457
41458	41459	41460	41461	41465	41466	41467	41468	41469	41470
41471	41472	41473	41474	41475	41476	41478	41479	41482	41483
41490	41492	41493	41496	41497	41498	41499	41500	41501	41502
41503	41504	41505	41509	41512	41513	41514	41515	41516	41517
41518	41519	41520	41521	41524	41525	41526	41527	41528	41529
41531	41532	41533	41535	41536	41537	41538	41539	41540	41541
41542	41543								

Table 5.2: Run numbers for iron.

Lead Run Numbers									
41762	41763	41765	41766	41767	41771	41772	41773	41774	41775
41776	41777	41778	41779	41780	41790	41791	41802	41803	41804
41805	41808	41809	41810	41811	41812	41813	41814	41815	41816
41817	41818	41819	41820	41821	41822	41823	41824	41825	41826
41827	41828	41829	41830	41831	41832	41833	41835	41836	41837
41838	41839	41841	41842	41843	41844	41845	41846	41847	41848
41862	41863	41864	41865	41866	41867	41868	41869	41870	41871
41872	41873	41874	41875	41876	41877	41878	41879	41880	41881
41883	41884	41885	41886	41887	41888	41889	41890	41891	41892
41893	41896	41897	41898	41899	41900	41901	41902	41903	41906
41907	41908	41910	41911	41912	41913	41914	41915	41916	41917
41918	41920	41928	41929	41930	41931	41932	41933	41934	41935
41936	41937	41938	41939	41940	41941	41942	41943	41952	41953
41954	41955	41956	41957	41958	41959	41960	41961	41962	41963
41965	41966	41967	41968	41969	41970	41971	41972	41973	41974
41975	41976	41977	41981	41982	41983	41984	41985	41986	41987
41988	41989	41990	41991	41992	41993	41994	41995	41996	

Table 5.3: Run numbers for lead.

5.1.2 Simulation Set

We can make a stability test for each file in the simulation set. For that purpose let's calculate the ratio between the reconstructed and generated events for each file. The results can be seen in figure 5.2. It was found that approximately 10% of the files for Fe, Pb, and D2 present a ratio lower with respect to the others. For consistency, all the results in this thesis by default are calculated without the first files in figure 5.2, even for carbon target.

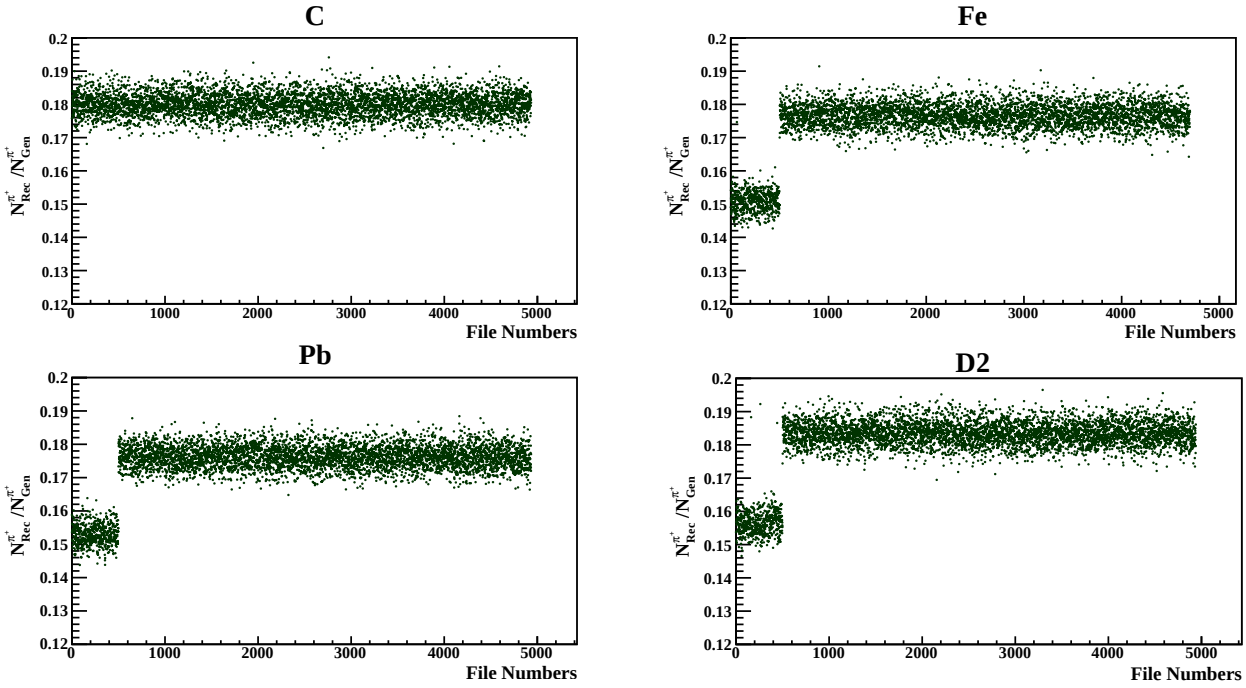


Figure 5.2: Ratio of reconstructed and generated events for each file of the simulation set.

5.2 Presentation of Results

The effect of some corrections, or some specific modification in general, is easily highlighted in terms of a percentage difference plot. Such a plot presents the difference , in %, between the two cases. Let's consider N_1, N_2, δN_1, and δN_2 the bin content of histograms 1 and 2, and the errors of the bins 1 and 2 respectively. We can plot a quantity Δ define as:

$$\Delta = \left(\frac{N_1 - N_2}{N_2}\right) \times 100 \tag{5.1}$$

Assuming random and uncorrelated errors [12] the error of Δ are calculated. For convenience, let's define $\Delta' = \Delta/100$. [1]

[1] In the error calculation, a constant term, c, just follows the following rule: if $R = cX$. Then $\delta R = |c|\delta X$, so at the end is just multiply the final expression by 100.

Target	Factor
C	1.118
Fe	1.054
Pb	2.090

Table 5.4: Normalization factors for each target. This should be included in the final EMC ratio, it is the ratio of target thickness (D/A).

$$\Delta' = \left(\frac{N_1 - N_2}{N_2}\right) \quad / \quad \delta \tag{5.2}$$

$$\left(\frac{\delta\Delta'}{\Delta'}\right)^2 = \left(\frac{\delta\left(N_1 - N_2\right)}{N_1 - N_2}\right)^2 + \left(\frac{\delta N_2}{N_2}\right)^2 \tag{5.3}$$

$$\frac{\delta\Delta'}{|\Delta'|} = \sqrt{\left(\frac{\left(\delta N_1\right)^2 + \left(\delta N_2\right)^2}{N_1 - N_2}\right)^2 + \left(\frac{\delta N_2}{N_2}\right)^2} \tag{5.4}$$

Thus, the final expression for the error is:

$$\delta\Delta = 100 \times \left|\left(\frac{N_1 - N_2}{N_2}\right)\right| \times \sqrt{\left(\frac{\left(\delta N_1\right)^2 + \left(\delta N_2\right)^2}{N_1 - N_2}\right)^2 + \left(\frac{\delta N_2}{N_2}\right)^2} \tag{5.5}$$

5.3 EMC Ratio

In this section, the results for the EMC ratio are presented. The main corrections involved in this analysis are acceptance corrections (AC), Coulomb corrections (CC) and radiative corrections (RC). Each one of those are treated separately to see their individual effect in the final results[2]. The EMC ratio is reported as a function of X_B, integrated over Q2. The corrections are applied for a grid of 10×10 equal-width bins. The target dependent normalization factors (related to target thickness) are shown in table 5.4. The ratio of yields must be multiplied by these factors [3].

5.3.1 Acceptance Correction (AC).

The expression for the acceptance corrected EMC ratio is given by equation 5.6:

[2]Every correction is applied in a bin by bin basis. The correction is a weight applied in each bin of the grid before the integration procedure. Other choices of implementation are possible, for example, a correction could be done in an event by event basis, in which is the event weighted instead of the bin. However, those options are not treated in this thesis.

[3]These values for the target thickness ratio are in the target paper [30], except for the case of carbon, which was incorrect there.

$$\left(\frac{\mathrm{N}_e^{\mathrm{A}}(\mathrm{Xb,Q2})}{\mathrm{N}_e^{D}(\mathrm{Xb,Q2})}\right)_{\text{Acceptance Corrected}} = \left(\frac{\mathrm{N}_e^{\mathrm{A}}(\mathrm{Xb,Q2})}{\mathrm{N}_e^{\mathrm{D}}(\mathrm{Xb,Q2})}\right)_{\text{Measured}} \times \left(\frac{\frac{\mathrm{N}_{\mathrm{Rec}}^{\mathrm{D}}(\mathrm{Xb,Q2})}{\mathrm{N}_{\mathrm{Gen}}^{\mathrm{D}}(\mathrm{Xb,Q2})}}{\frac{\mathrm{N}_{\mathrm{Rec}}^{\mathrm{A}}(\mathrm{Xb,Q2})}{\mathrm{N}_{\mathrm{Gen}}^{\mathrm{A}}(\mathrm{Xb,Q2})}}\right) \tag{5.6}$$

The AC on EMC ratios are found to be between -7% and +3.8% depending on X_b and target type, see figure 5.3. The acceptance effect is basically zero in carbon target for most part, meanwhile the other two targets show a similar behavior, the acceptance is a few percent effect, on average 1.78% for Fe and a 2.26% for Pb. One-dimensional projection of all the acceptance factors involved in a 2-fold differential acceptance-corrected EMC ratio is depicted in figure 5.4, for each target.

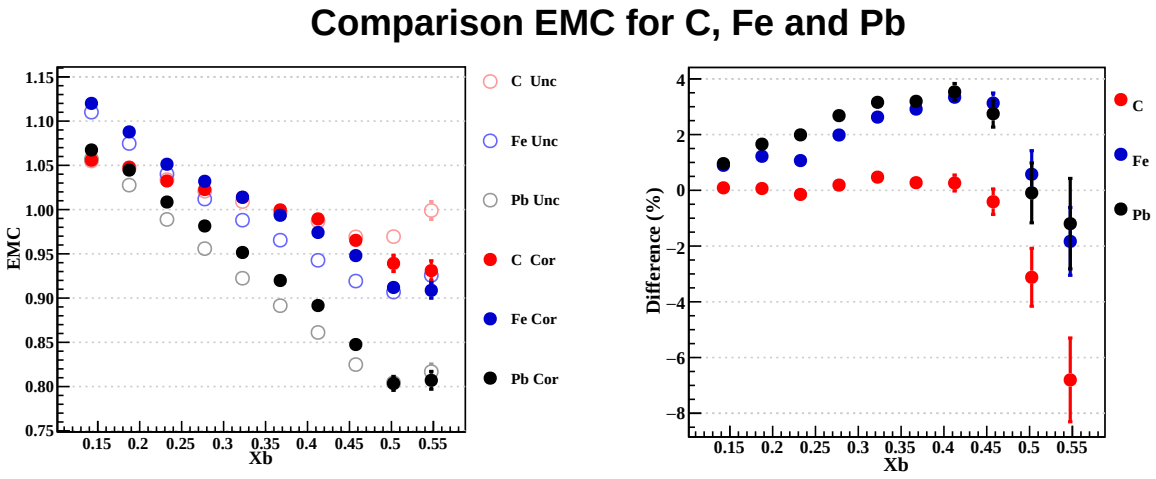


Figure 5.3: EMC ratio comparison, for all three targets, with (full circles) and without (hollow circles) acceptance correction.

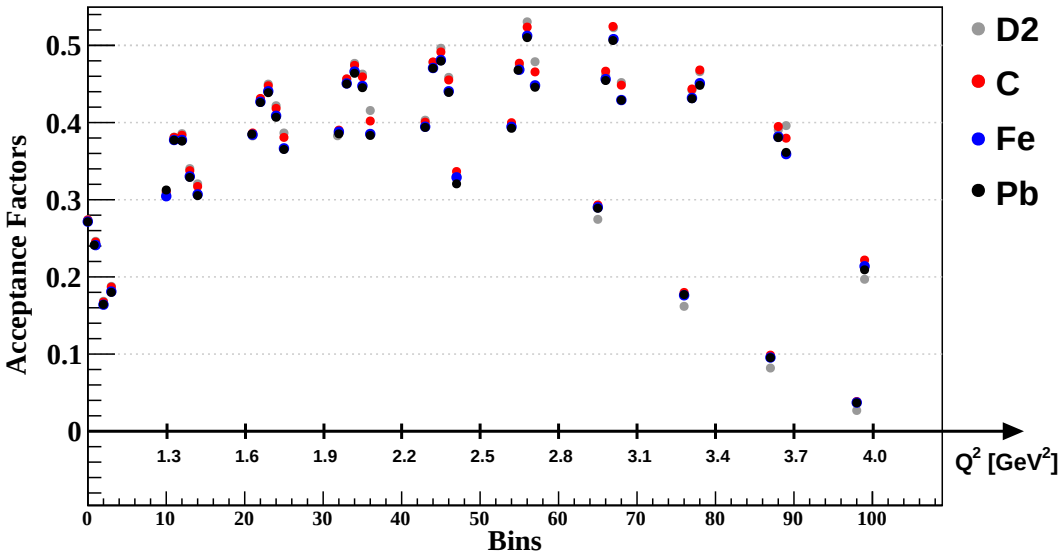


Figure 5.4: Acceptance correction factors for C, Fe, Pb and D2. The erors are smaller than the marker size.

An additional cut is imposed, when the AC is calculated, to avoid bins with too few events. The cut is $N_{Rec} > 4$ (used in [25]). Keeping a minimum value in the reconstructed events avoid to end up with very large statistical errors.

The potential bias due to mis-modeling of the detector response is assessed by exploiting the redundancy of the six CLAS sectors to perform independent measurements of the EMC ratio, see figure 5.5. The results are found to be consistent within $\pm 2\%$, except at the highest values of X_b , where the deviations are within $\pm 6\%$.

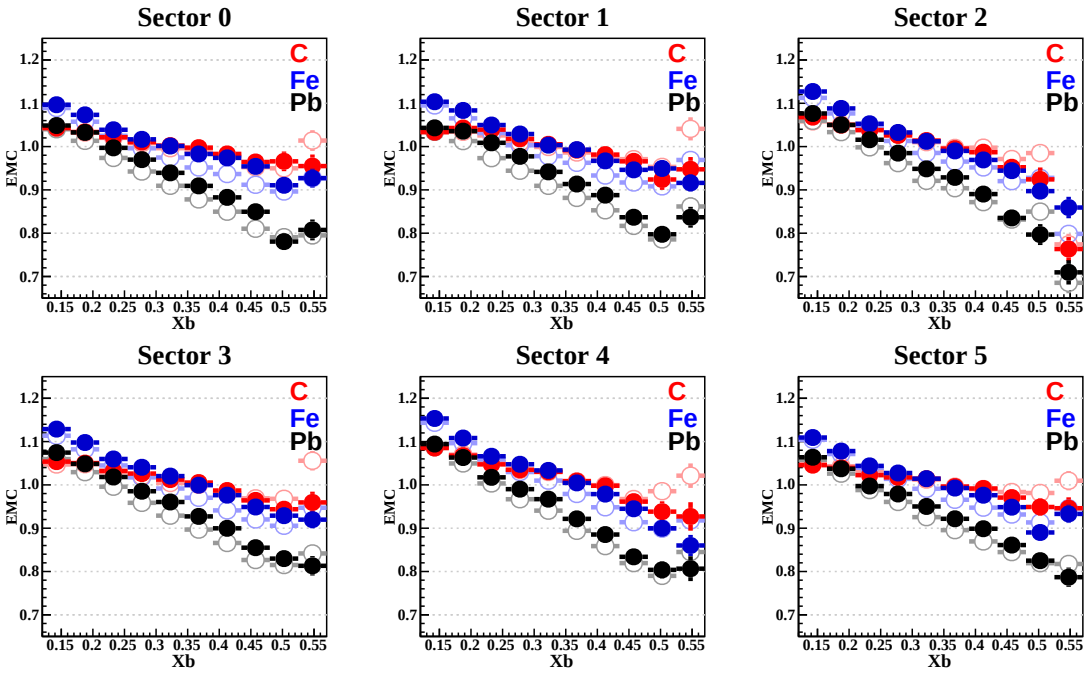


Figure 5.5: EMC ratio comparison, for all three targets, with (full circles) and without (hollow circles) acceptance correction, for each sector.

5.3.2 Background Subtraction (YC)

To explore the effect of a YC cut on the EMC ratio we discussed two possibilities. A cut in YC between -2.2 and 2.0 cm (used in [59]), see figure 5.6, and the nominal cut $|\mathrm{YC}| < 1.4$cm, see figure 5.7. The effect of the cut is found to be smaller than 1% except at high X_b.

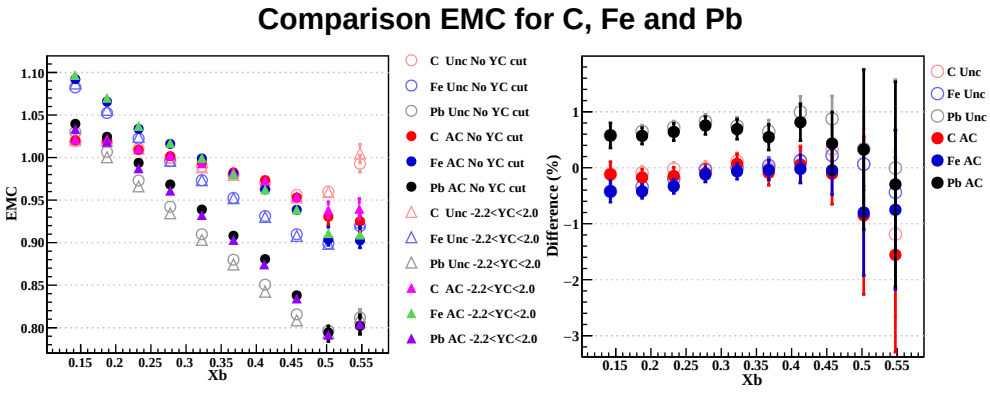


Figure 5.6: EMC ratio comparison, for all three targets, with (full symbols) and without (hollow symbols) acceptance correction, and with and without YC cut.

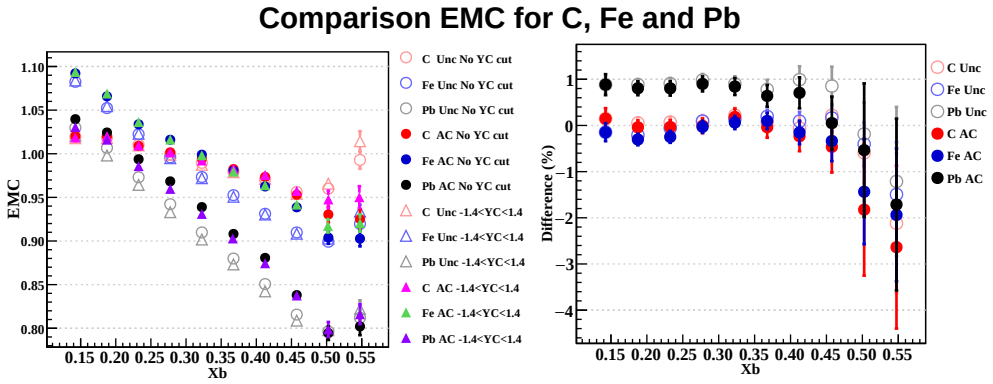


Figure 5.7: EMC ratio comparison, for all three targets, with (full symbols) and without (hollow symbols) acceptance correction, and with and without YC cut.

5.3.3 Vertex Cuts

To study z-vertex dependency in the EMC ratio 3 different set of vertex cuts are used. Vertex cuts by Hayk Hakobyan (**HH**), Raphael Dupre (**RD**) and Taisiya Mineeva (**TM**) are chosen to be compared. The reason to pick these set of cuts is due to the different features among them. Here are a brief description of each case:

- **HH vertex cuts:** are loose cuts, target independent, based on cuts in the Z distribution of the electrons. Solid target always goes up to $Z = -20$cm. They are sector dependent. See table 4.21. This are the vertex cuts applied by default in this thesis for simulated data. After the application of these cuts, 9.64% (C), 8.46% (Fe), and 11.44% (Pb) of the electrons from data are rejected. In this approach, the cuts in the simulation are the same as in data. Rejected electrons are less than 1% for all cases.

- **RD vertex cuts:** are narrower than HH's, target dependent, based on cuts in the Z distribution of electrons. They are sector dependent. These vertex cuts are used in this thesis by default on data. See section 4.4.2. After the application of these cuts, 11.43% (C), 10.17% (Fe), and 12.71% (Pb) of the electrons from data are rejected. In this approach, the cuts in the simulation are different than in data. It is just -30 ± 1.5[cm] for liquid target and -25 ± 1.5[cm] for solid target. Rejected electrons are less than 1% for all cases.

- **TM vertex cuts:** are even narrower than HH's and RD's. They are based not in the Z distribution, but rather in the corrected version of it (ZC), see table

5.5. They are sector independent. After the application of this cuts, 15.23% (C), 13.46% (Fe), and 15.87% (Pb) of the electrons from data are rejected. In this approach, there are no vertex cuts in the reconstructed simulation.

Target	TM Vertex cut
C	-25.33 < ZC < -24.10
Fe	-25.65 < ZC < -24.26
Pb	-25.54 < ZC < -24.36
D2	-31.80 < ZC < -28.40

Table 5.5: Vertex cuts according to TM criteria. ZC is the correct Z distribution for electrons, in [cm].

In figure 5.8 is a comparison between all the above cases, HH and RD cases are denoted as straight lines (green and red respectively) and the TM case, since it is not a cut directly in Z distribution, is super-imposed on the Z distribution when the TM vertex cuts are applied.

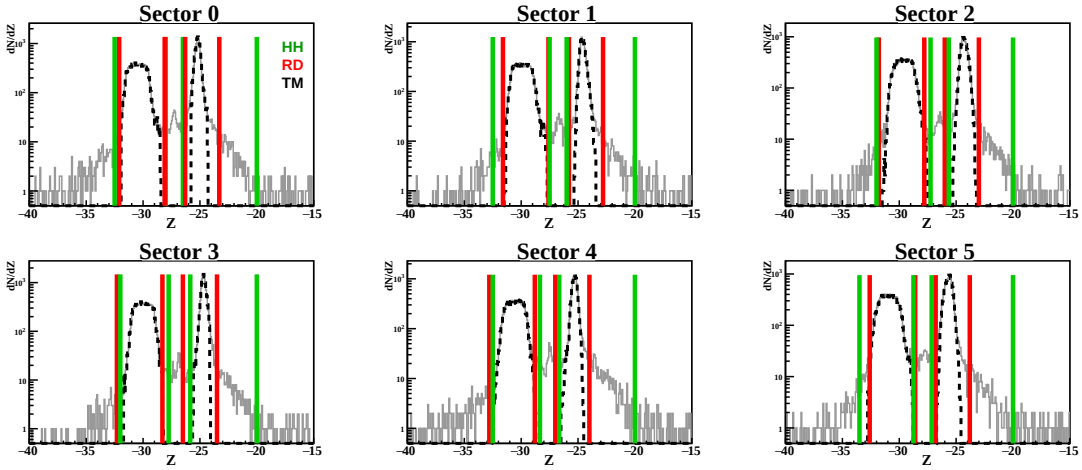


Figure 5.8: Z distribution for electrons, in cm. Green lines are HH vertex cuts, red lines are RD vertex and the dashed black line is TM.

In order to see the difference between using one set of vertex cuts or another, the EMC ratio was calculated for the three cases separately. The average difference in % with respect to RD case is presented in tables 5.6 and 5.7. The information is presented for all the targets, the uncorrected and acceptance corrected cases, and for three different cuts in YC coordinate. The YC cut implementation is included because the difference between the three vertex cuts depend highly on whether the YC cut is applied or not. As we increase the background subtraction, the events

on the tail of the distribution of Z are fewer. This makes the external edges of the vertex cuts irrelevant. The last detail to discussed is the ability of each set of cuts to efficiently remove the aluminum foil which separates the solid and cryo targets. This is fullfilled by all three set of cuts.

	HH - RD		
Target	**No YC**	$-2.2 <$ **YC** < 2.0	$-1.0 <$ **YC** < 1.0
C (Unc)	2.37%	0.55%	0.33%
Fe (Unc)	1.78%	0.36%	0.19%
Pb (Unc)	1.72%	0.22%	0.04%
C (AC)	4.35%	2.45%	2.26%
Fe (AC)	3.75%	2.26%	2.12%
Pb (AC)	3.66%	2.11%	2.00%

Table 5.6: Difference, in %, between results using HH and RD vertex cuts. Three different cases of YC cuts are considered.

	TM - RD		
Target	**No YC**	$-2.2 <$ **YC** < 2.0	$-1.0 <$ **YC** < 1.0
C (Unc)	-6.77%	-3.05%	-1.92%
Fe (Unc)	-5.24%	-2.19%	-1.42%
Pb (Unc)	-8.13%	-3.77%	-2.45%
C (AC)	-4.87%	1.24%	-0.05%
Fe (AC)	-3.21%	-0.26%	0.52%
Pb (AC)	-6.25%	-1.99%	-0.61%

Table 5.7: Difference, in %, between results using TM and RD vertex cuts. Three different cases of YC cuts are considered.

5.3.4 Coulomb Corrections (CC)

The Coulomb corrected EMC ratio is given by:

$$\left(\frac{N_e^A}{N_e^D}\right)_{\text{Coulomb Corrected}} = \left(\frac{N_e^A}{N_e^D}\right)_{\text{Measured}} \times \left(\frac{\delta_{CC}^A}{\delta_{CC}^D}\right) \tag{5.7}$$

where δ_{CC}^D is unity.

To extract the CC factors from EXTERNAL code the procedure is straightforward. First, we extract the centroid of each bin (instead of just the geometrical center), see figure 5.9, with the red dots representing the centroids of each bin.

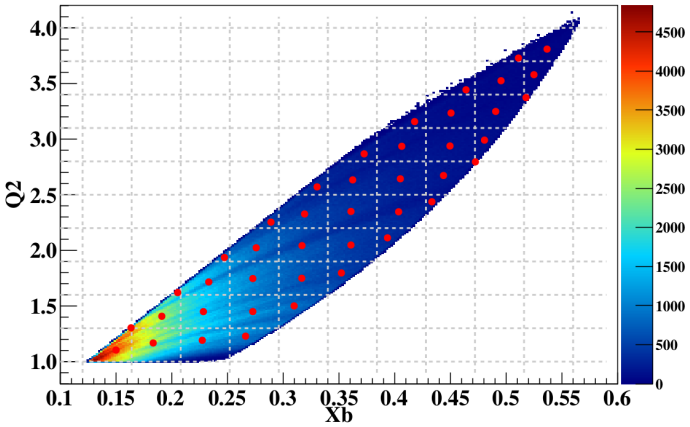


Figure 5.9: Centroids for each bin in (Q2,Xb) phase space. This is the input for EXTERNAL code.

Next step is to transform the centroid of (Q2,Xb) into (E',Θ) variables, through the expressions given in equation 5.8.

$$E' = E_b - \frac{Q^2}{2M_p X_b} \quad ; \quad \Theta = 2sin^{-1}\left(\sqrt{\frac{Q^2}{4E_b\left(E_b - \frac{Q^2}{2M_p X_b}\right)}}\right) \tag{5.8}$$

In figure 5.10 is shown how the distribution transforms when passing from one phase space to another.

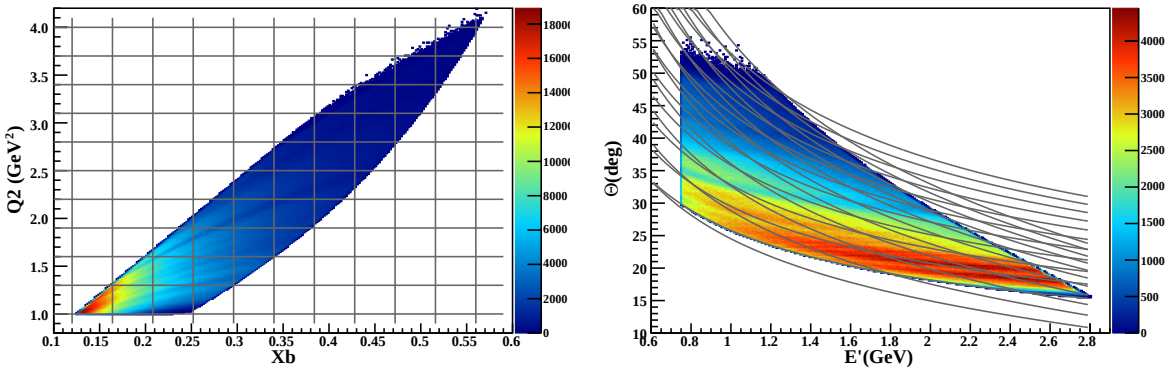


Figure 5.10: Transformation of the distribution from the (Q2,Xb) space to (E',Θ). The grid also transforms accordingly.

Those are all the elements needed for the code to do the calculations and to obtain the final δ_{CC} factors. The result is depicted in figure 5.11. The results clearly show a bigger factors as the nuclear target gets heavier, and as we increase the value of Q^2.

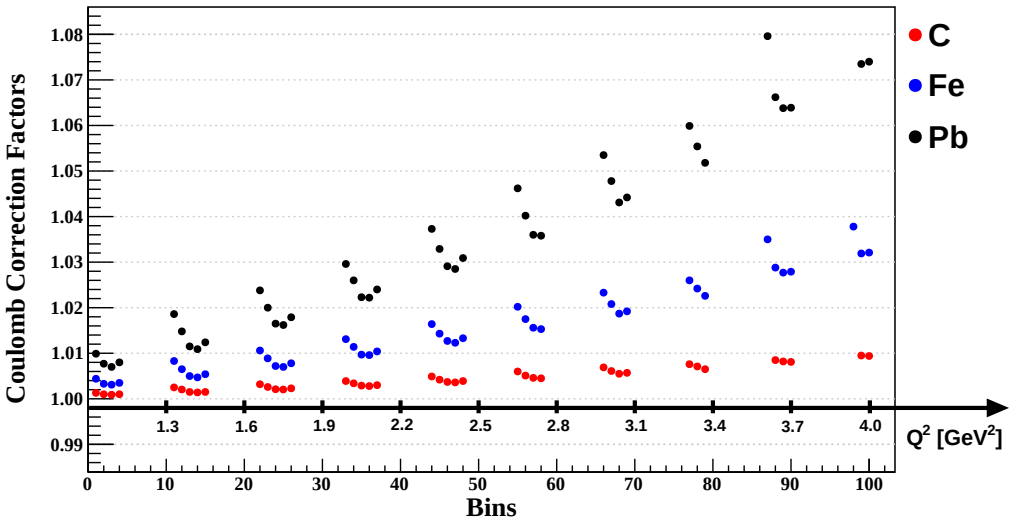


Figure 5.11: Coulomb Correction factors, for C, Fe and Pb.

Figures 5.12 and 5.13 show the CC factors as a function of X_b and Q^2 for five different polar angle ranges. Note that to improve the visualization the binning was increased. The factors increase their value for the largest polar angles.

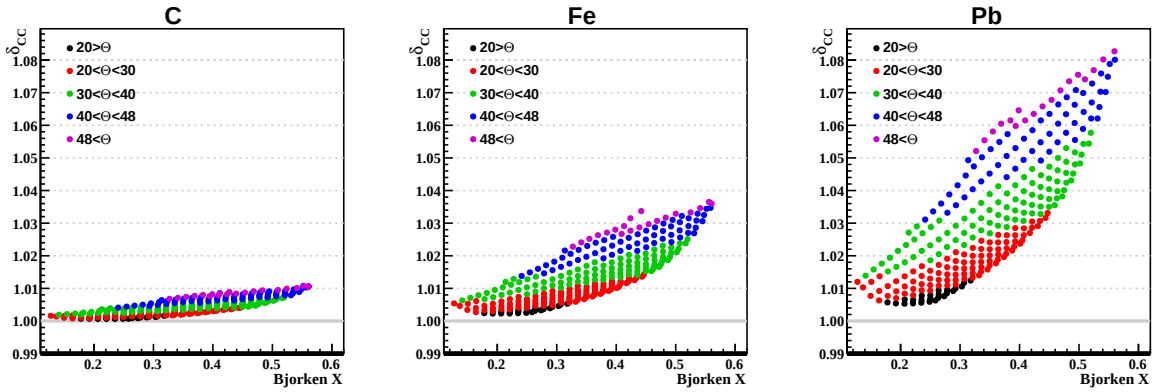


Figure 5.12: Coulomb Correction factors for C, Fe and Pb as a function of Xb and for different Θ ranges.

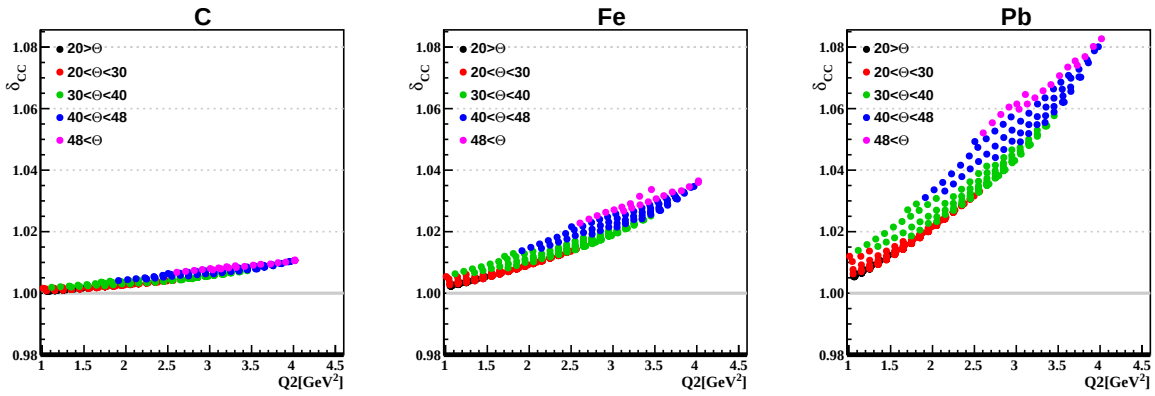


Figure 5.13: Coulomb Correction factors, for C, Fe and Pb as a function of Q^2 and for different Θ ranges.

Figure 5.14 shows the EMC ratio with and without CC applied, for each target independently. It can be clearly seen that neglecting Coulomb distortions at our energies imply an overestimate of the EMC effect, which is more dramatic for heavier nuclei. Upward shifts are observed for the three targets, for C the effect is small, below +1% for all X_b, for Fe it goes between +1% and +3% and for Pb it goes from +1% up to +7% increasing with X_b. The magnitude and sign of the correction is similar to what is reported by the JLab E03-103 experiment[4] [45] [46]. We have also checked that we can reproduce the results presented in ref. [48] with a selection of $W > 1.8$ GeV. In figure 5.15 is depicted the distribution of X_b and Q^2. Note that the largest polar angles are related to the highest CC factors. With a cut in W at 1.8 GeV instead of 2 GeV we are increasing the data available at high X_b for smaller porlar angles (smaller CC factors), therefore the contribution in that region will bring the net effect of the correction down. The CC in that region will be smaller than the one reported in this work.

[4] the heaviest target in this experiment was gold (Z=79).

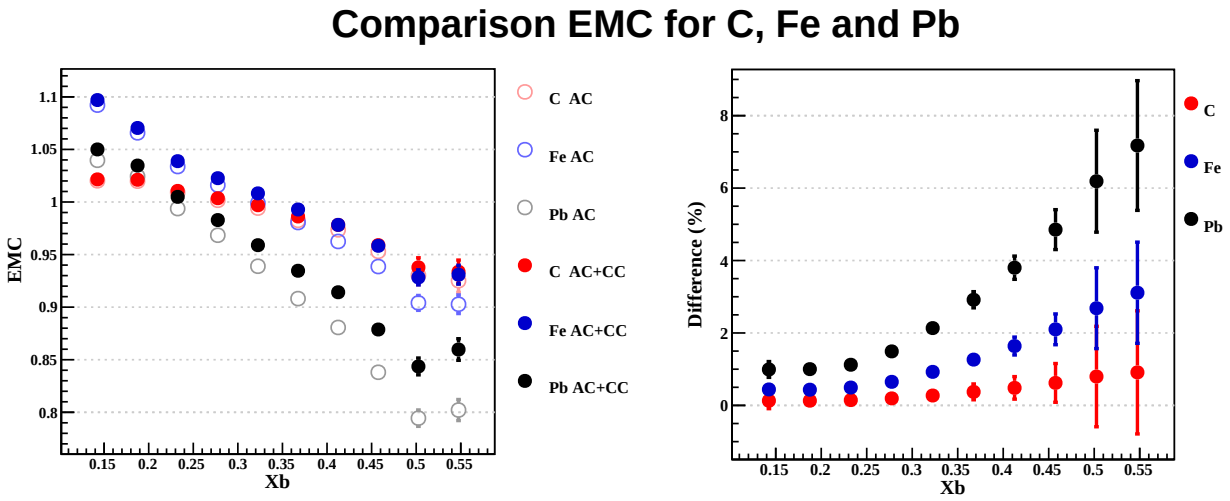


Figure 5.14: Comparison EMC ratio with and without CC applied, for C, Fe and Pb.

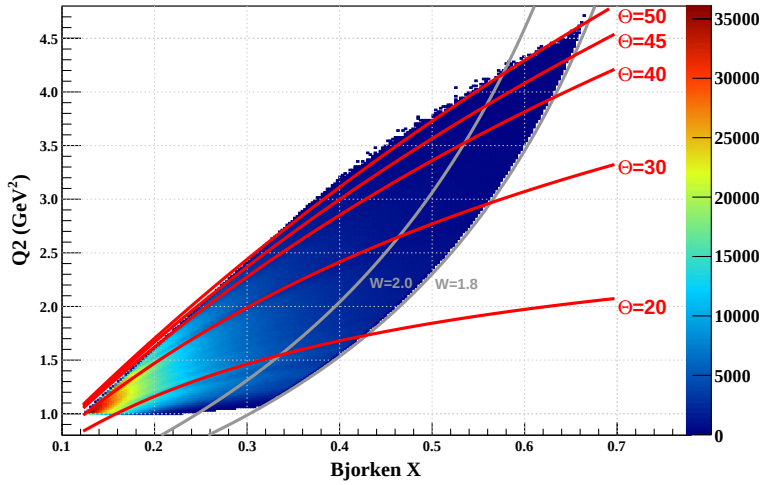


Figure 5.15: Two-dimensional visualization of the Q^2 and X_b variables. Red lines are constant values of polar angles. Gray lines are constant values of W.

5.3.5 Radiative Corrections (RC)

The expression of the radiative corrected EMC ratio is given by equation 5.9.

$$\left(\frac{N_e^A}{N_e^D}\right)_{\text{Radiative Corrected}} = \left(\frac{N_e^A}{N_e^D}\right)_{\text{Measured}} \times \left(\frac{\left(\frac{\sigma_{\text{Rad}}}{\sigma_{\text{Born}}}\right)_D}{\left(\frac{\sigma_{\text{Rad}}}{\sigma_{\text{Born}}}\right)_A}\right) \tag{5.9}$$

Figure 5.16 shows the RC factors for all targets (this time the factor for deuterium is not unity). Figure 5.17 shows the ratio -which eventually will be the number to use in the correction- between solid and cryo targets.

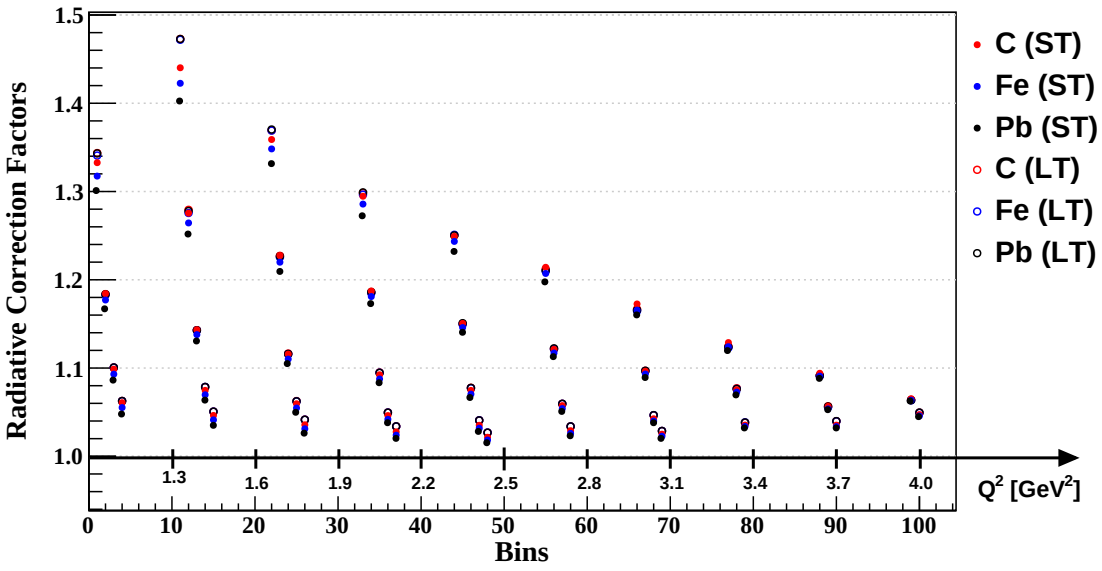


Figure 5.16: Radiative Correction factors, using EXTERNAL.

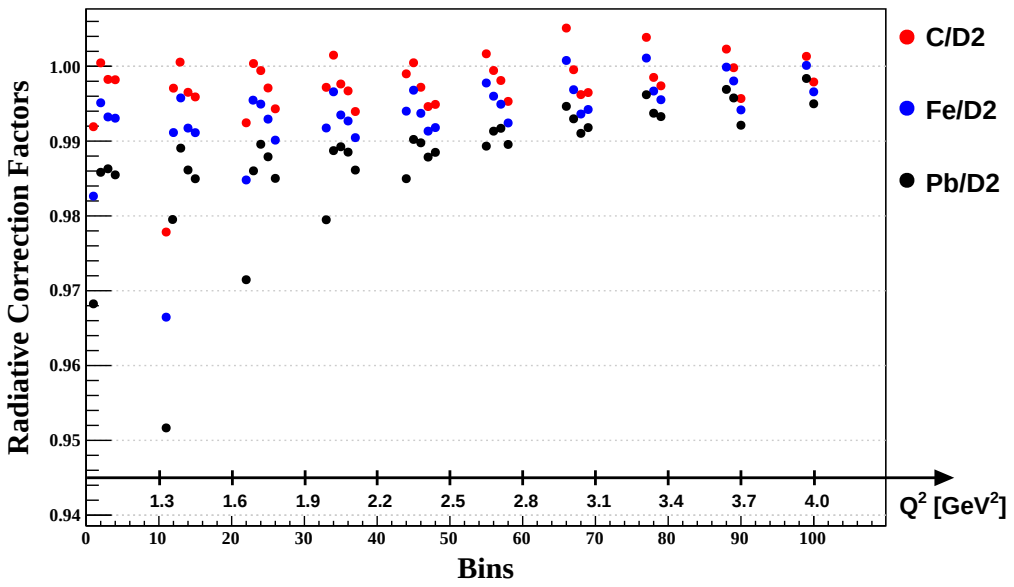


Figure 5.17: Radiative Correction factors, using EXTERNAL.

Figures 5.18 and 5.19 show the RC factors as a function of X_b and Q^2 for five

different polar angle ranges. Note that to improve the visualization, the binning was increased.

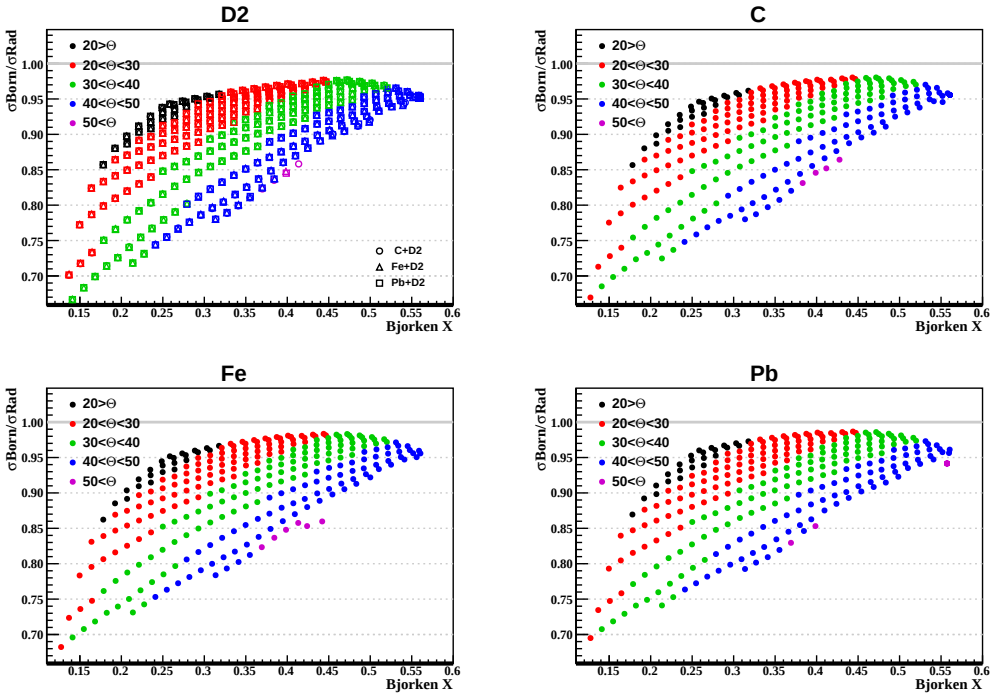


Figure 5.18: Radiative Correction factors for D2, C, Fe and Pb as a function of Xb and for different Θ ranges.

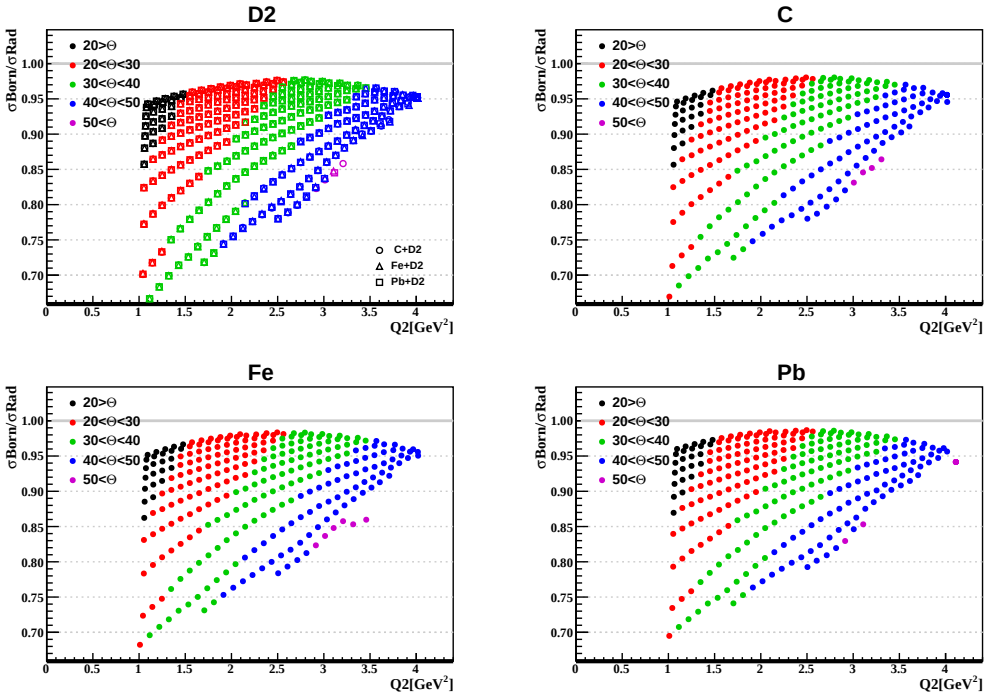


Figure 5.19: Radiative Correction factors for D2, C, Fe and Pb as a function of $Q2$ and for different Θ ranges.

Figure 5.20 shows the EMC ratio with and without RC applied for each target independently. Upward shifts are observed for the three targets, for C the effect is small, below +1% for all X_b, for Fe it goes up to +1.8% and for Pb it goes from +0.5% up to +3.3% decresing with X_b. On average the effect of the RC are +0.16% for carbon, +0.62% for iron and +1.45% for lead.

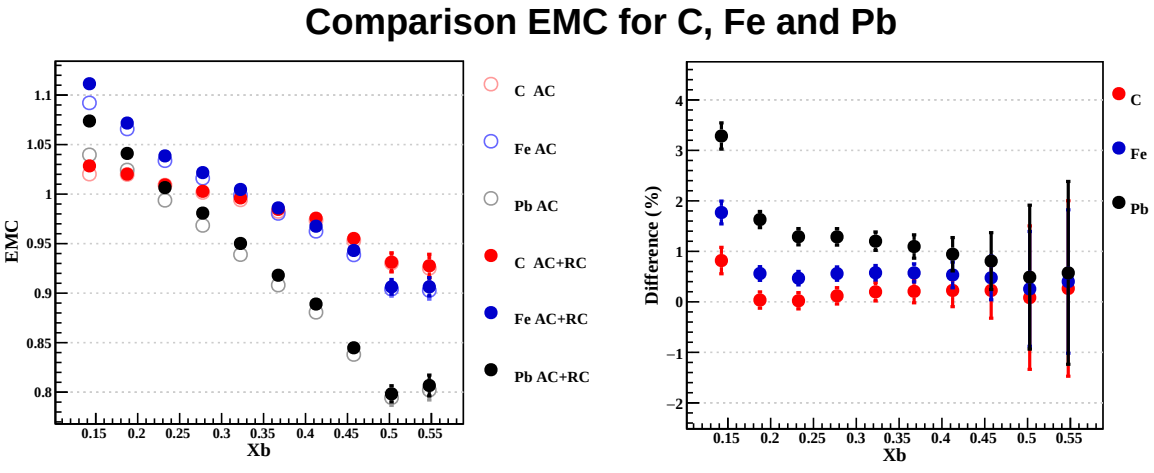


Figure 5.20: Comparison EMC ratio with and without RC applied, for C, Fe and Pb.

5.3.6 CC Mirror Matching

In section 3.2.2, the Cherenkov Counter (CC)[5] in CLAS was introduced as a detector aimed to separate electrons from pions. In the electron PID section (section 4.3.1.1), a sector dependent cut in the number of photo-electrons was applied for π^- rejection. Even though this cut is used extensively, it is not the only approach to use CC information to get rid of pion contamination (see section 4.3.1.1). A procedure called *CC mirror matching* [76] [77] is used to remove the single photo-electron peak. First, information of the *Status* word from CCPB data bank is used:

$$\text{Status (CC)} = 10 \times (\text{CC segment number}) + 1000 \times (1+\phi) \quad ; \quad \phi = -1, 0, +1 \quad (5.10)$$

where the variable ϕ indicates which PMT side was fired. $\phi = +1$ is for right PMTs (2,4,6,...36), $\phi = -1$ is for left PMTs (1,3,5,...35) and $\phi = 0$ is for both left and right, i.e, electron near mid-plane. See figure 5.21.

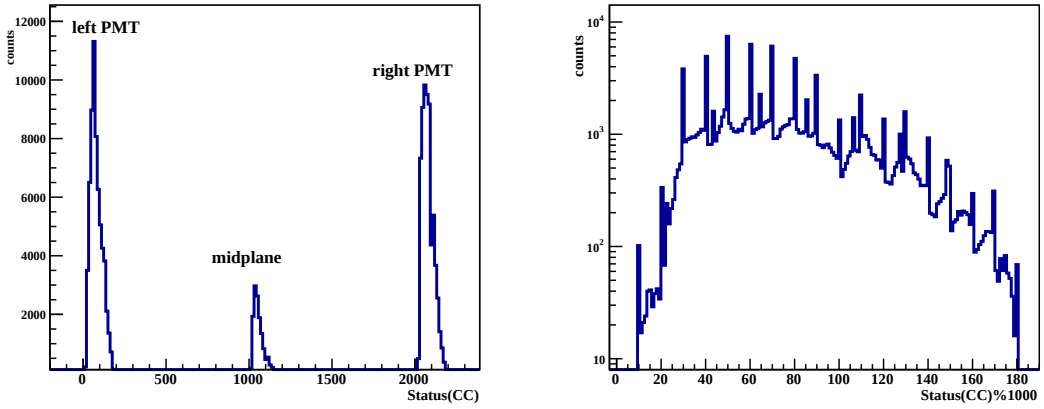


Figure 5.21: Information obtained from the status word in CCPB. On the left, is the status variable; the three peaks correspond to the three possible values of ϕ, related to the PMT in which the electron was measured. On the right, there is the mirror number, calculated as the status module (the % sign) 1000. Note that it can be clearly seen the 18 segments.

This cut is defined in the CC plane, so an imaginary projectile plane behind the CC is defined. This plane is where Cherenkov radiation would arrive in the case it

[5] not to be confused with Coulomb corrections, also known as *CC*. This is an unfortunate, but widely used notation. The meaning should be clear from the context.

propagated the same distance from emission point to PMT, but without any reflection in the mirrors. This plane, in the sector reference coordinate system, is given by:

$$1 - 7.84 \times 10^{-4} x - 1.68 \times 10^{-3} z = 0 \tag{5.11}$$

Angles in the CC plane are not directly defined in the *BOS* data banks, some calculations are necessary in order to construct them. A new set of projected variables are defined (θ_p, ϕ_p). For that purpose, we have to use the coordinates and direction of the track information with SC plane as measured in DC[6]. The polar and azimuthal angles in projective plane are calculated making a crossing of the straight line with the plane, see figure 5.22. Note that the track after hitting the CC plane moves along the $\vec{t}$ vector because there is no magnetic field in the region between SC and CC planes.

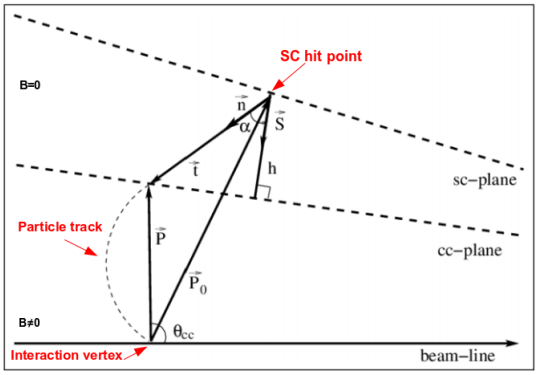


Figure 5.22: Diagram of the SC and CC planes needed for the calculation of θ_p (θ_{CC} in the figure) and ϕ_p. Taken from [78].

In figure 5.23, there are the projective angles distributions: θ_p v/s ϕ_p (in degrees) for each CLAS sector for the left PMTs (left), the mid-plane (middle), and for the right PMTs (right). The different segments are highlighted alternating red and black colors; it should be 9 red strips and 9 black strips in each plot. The first two segments and the last one have very low number of photo-electrons. Looking at the distributions, there are regions with efficiency issues (like when the CC mirrors joined). Those regions were not specifically considered in the PID for electrons in this analysis. Look at [78] for a detailed study of a CC based cut for electron selection.

[6] they are $\vec{P_0} = (x_{SC}, y_{SC}, z_{SC})$ and the direction $\vec{t}(cx_{SC}, cy_{SC}, cz_{SC})$, according to figure 5.22.

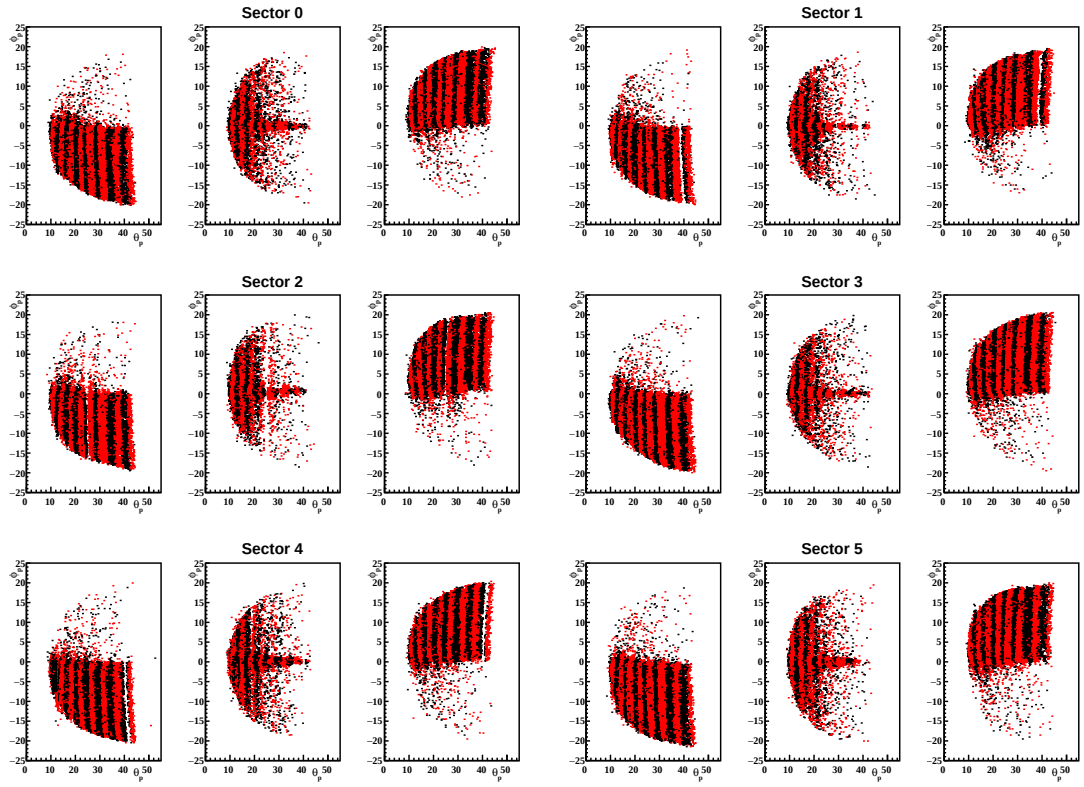


Figure 5.23: Polar and azimuthal angles in the CC *special plane* distributions for each CLAS sector. In each sector, the leftmost plot is the case of left PMTs. The middle one is for the mid-plane tracks. The rightmost is for the right PMTs. The 18 different segments can be clearly identified due to the color scheme: black for segments 1,3,..17 and red for segments 2,4,..18.

The matching between the sector where the track was measured in DC, and the sector, where hits in EC and CC were detected, is a requirement for this approach. This means that if we have signals in coincidence in DC and CC noise that belongs to different sectors, those events are removed. A potential problem emerged when we realized that within one sector no geometrical matching between the track and CC hit was applied. Thus, a coincidence of the pion track in DC with CC noise, not related to the track, but located in the same CLAS sector is possible. The distribution of θ_p against the segment in CC (mirror number (MN)) is shown in figure 5.24 (a). To reject mismatches between the mirror number and the angle in CC plane (those indicates signal from π^-) a fitting procedure is performed on this distribution. Following [59], the fit is a second order polynomial to define a matching region:

$$|\theta_p - \left(a_1 + a_2 \times MN + a_3 \times MN^2\right)| < 2.5 \tag{5.12}$$

where a_i are the parameters of the fit. The result of this cut can be seen in figure 5.24 (b) and the effect in the Nphe distribution is shown in 5.24 (c) for a particular CLAS sector.

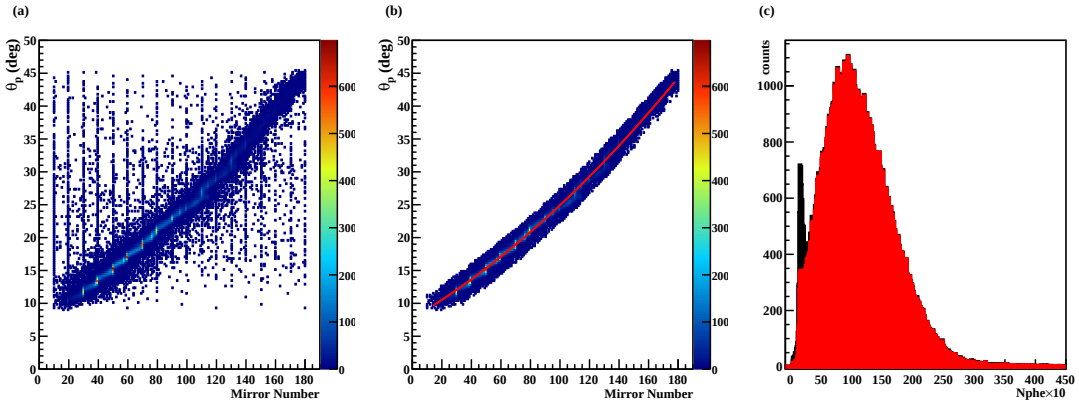


Figure 5.24: In (a), there is the projected polar angle, θ_p, as a function of the mirror number for all electrons with all Id cuts but the ones in Nphe. In (b), there is the same case as (a) but with the CC matching mirror cut applied, the red line is a fit based on a second order polynomial. In (c), there is the Nphe×10 distribution without the CC matching cut (black) and with the cut applied (red). It is clear that the single peak is gone after the cut is imposed. These plots are for sector 0.

The difference between *CC Matching Cut* and a photo-electron cut is depicted in figure 5.25. The difference is less than 1% for the whole range in X_b. No significant difference among target types are observed.

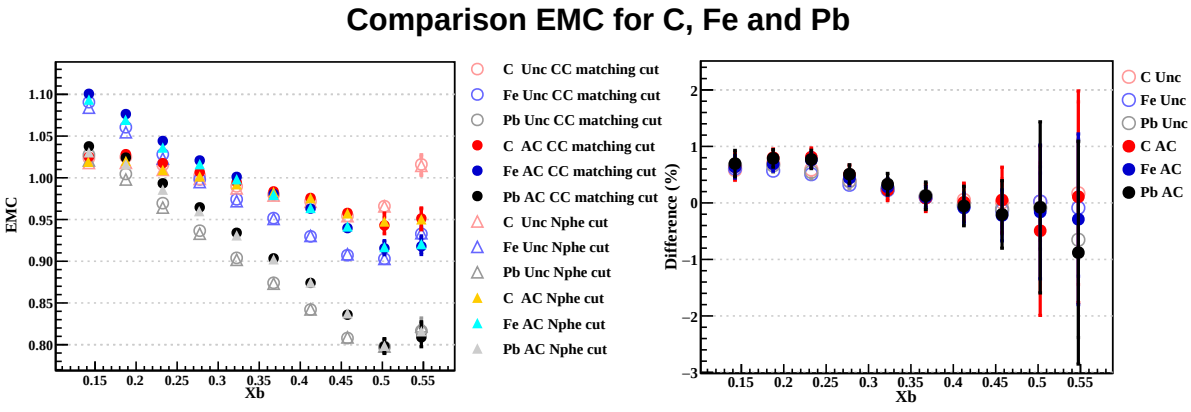


Figure 5.25: EMC ratio comparison for C, Fe and Pb for two different methods for removing negative pion contamination using CC information.

5.3.7 EMC Ratio Results

A comparison with three other experiments is depicted in figure 5.26, for the case of carbon target. For other targets see [73]. The major features of these experiments are listed below:

- The **N**ew **M**uon **C**ollaboration (NMC): performed with a 200 GeV muon beam. Radiative corrections were applied following a similar formalism we used. Deuterium was used to produce EMC ratio. For the C/D case the kinematic range covered is $0.0035 < X_b < 0.65$ and $0.5 < Q^2 < 90$ GeV^2.

- SLAC E139: performed with an electron beam of energy in the range of 8.0 Gev to 24.5 GeV. The targets exposed to the beam were He, Be, C, Al, Ca, Fe, Ag and Au. Deuterium was used to produce EMC ratio.

- JLab E03 103: performed at JLab using CEBAF in Hall C. The beam energy was 5.766 GeV. The electrons were scattered from solid and cryogenic targets of $He(3)$, $He(4)$, Be, C, Cu, and Au. Deuterium was used to produce EMC ratio.

We can see from figure 5.26 the results obtained in this thesis are in reasonable agreement with the rest of the experiments. See references [70], [73] and [74] and references therein for more details.

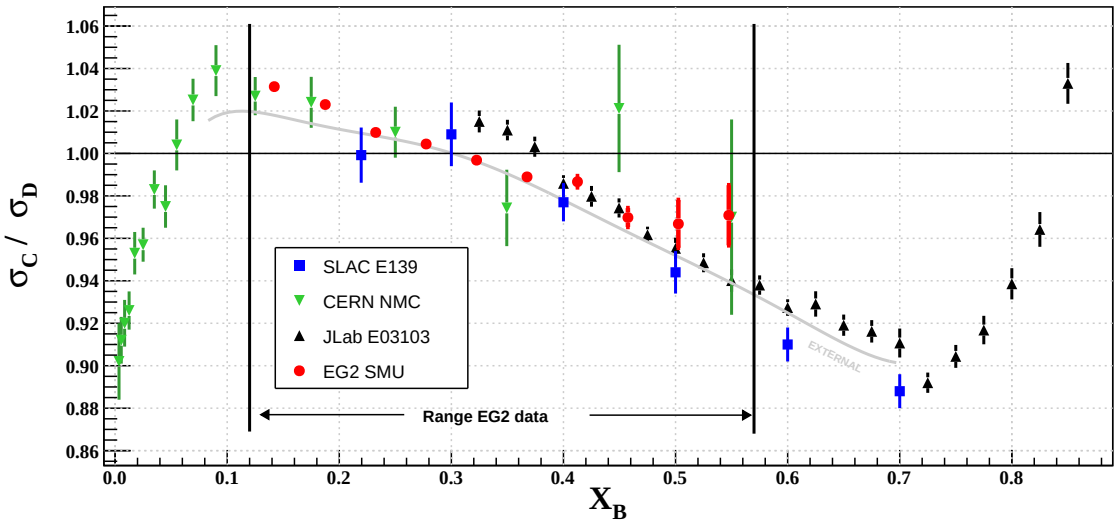


Figure 5.26: Ratios of the C to Deuterium DIS cross sections. The measurements shown are for SLAC (blue squares), NMC (green triangles), JLab E03-103 (black triangles) and for our results from JLab EG2 experiment (red circles).

In figure 5.27 there is a comparison for all three targets with all the corrections implemented. There is a consistent behavior for carbon and lead compared with the world-data based f_{EMC} function (see section 4.8.2.1). The iron is consistently above the parametrization function. Previous analyses observed some offsets for the iron case in some observables, this suggests that the origin of that could be at the electron level at low Xb, where the difference with the function is the largest.

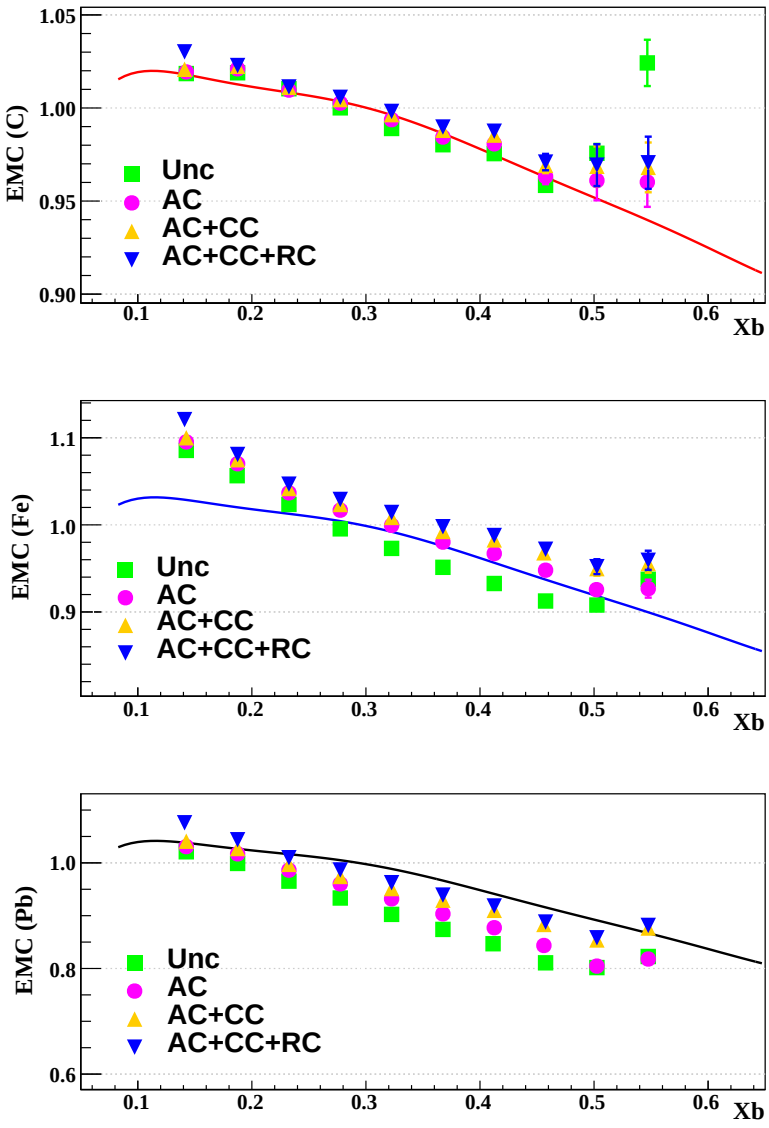


Figure 5.27: EMC ratio comparison for C (top), Fe (middle), and Pb (bottom) as a function of Bjorken x. *Unc* stand for uncorrected, AC stand for Acceptance correctionCC stand for Coulomb correction and RC for radiative correction. The continuous line is the parametrization function $f_{EMC}(A)$.

5.4 Multiplicity Ratio (MR)

In this section, the results for the multiplicity ratio are presented. The MR is reported as a function of several variables, but mainly focusing on Z_h dependence. The multiplicity ratio contains two factors: one is the nuclear-to-deuterium ratio of the inclusive electron yield, i.e. the EMC ratio, and the other is the nuclear-to-deuterium ratio of the yield of pions. The corrections for electron and pion yields are treated separately. One aspect to consider about the MR definition is that when MR is performed as a function of a variable that is not an electron variable, the ratio of inclusive DIS electrons in the kinematical bin comes out as a single number. In this thesis, for sake of simplicity, this number will be denominated as *Normalization factor*:

$$\mathrm{MR} = \frac{\left(\mathrm{N}_{\pi}^{\mathrm{DIS}}\right)_A}{\left(\mathrm{N}_{\pi}^{\mathrm{DIS}}\right)_D} \times \underbrace{\frac{\left(\mathrm{N}_{el}^{\mathrm{DIS}}\right)_D}{\left(\mathrm{N}_{el}^{\mathrm{DIS}}\right)_A}}_{\text{Normalization Factor}} \tag{5.13}$$

In table 5.8 are listed all the integrated normalization factors for all the targets. On the left is the ratio without any correction, in the second column is the acceptance corrected ratio and in the rightmost column is the effect of the correction in percentage.

Target	**Normalization Factors**		
	Uncorrected	**Acceptance Corrected**	**Effect(in %)**
C	1.0889	1.08594	-0.27
Fe	1.0245	1.00294	-2.16
Pb	2.1487	2.08732	-2.94

Table 5.8: Normalization factors

5.4.1 Acceptance Correction

The acceptance correction factors are evaluated in binning of four kinematic variables: Q^2, X_b, Z_h, and p_T^2. This case is called in this work *4-fold differential AC* or *4D acceptance*. The correction factors are calculated separately for each target type. See figure 5.28 to see the MR as a function of Z_h with and without AC. The range of the correction goes between -3% and +7% for the highest bin in Z_h. Figure 5.29 shows all the acceptance factors, which are a total of $6 \times 5 \times 5 \times 10 = 1500$ (in Q^2, X_b, p_T^2 and Z_h respectively) bins.

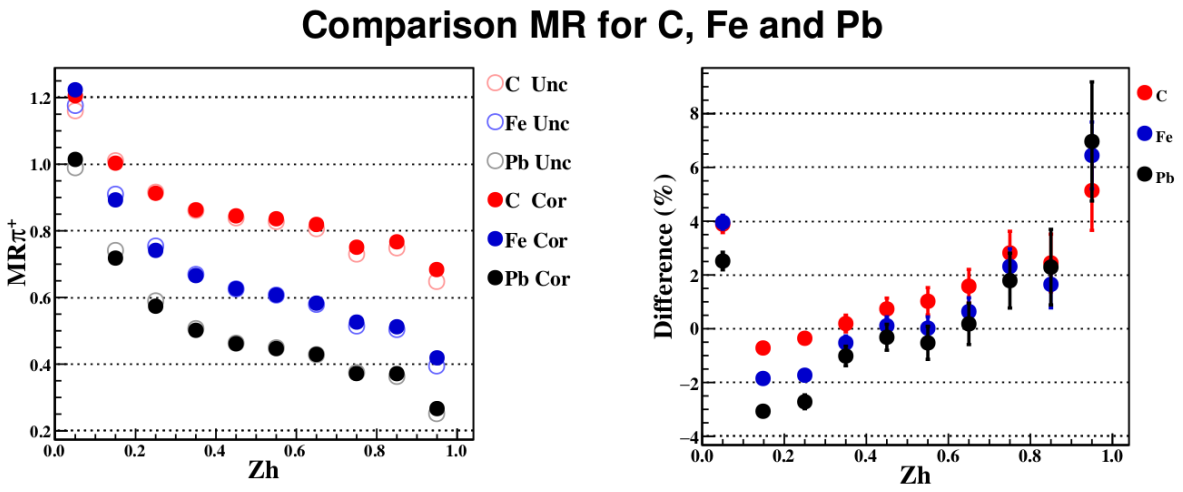


Figure 5.28: MR curves, the hollow circles are the uncorrected case and the full circles are with acceptance correction applied.

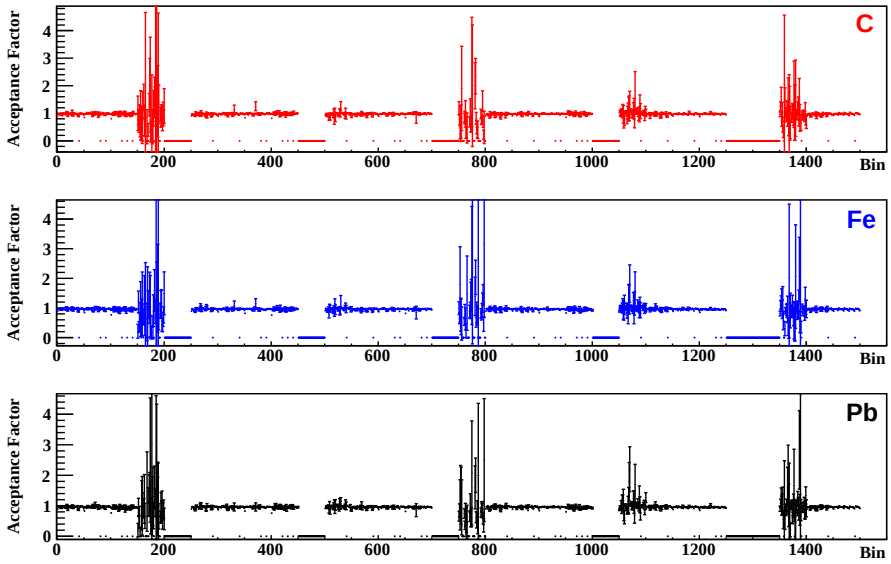


Figure 5.29: One-dimensional distribution of all the ratios of acceptance factors involved in a 4-fold differential acceptance corrected multiplicity ratio.

The same MR could be performed for each CLAS sector independently. In figure 5.30, there are the MR for all targets, as a function of Z_h for each sector. The hollow marker correspond to the uncorrected MR, and the full one is with acceptance correction. In table 5.9, there is a list of all the different normalization factors for all sectors and for each target.

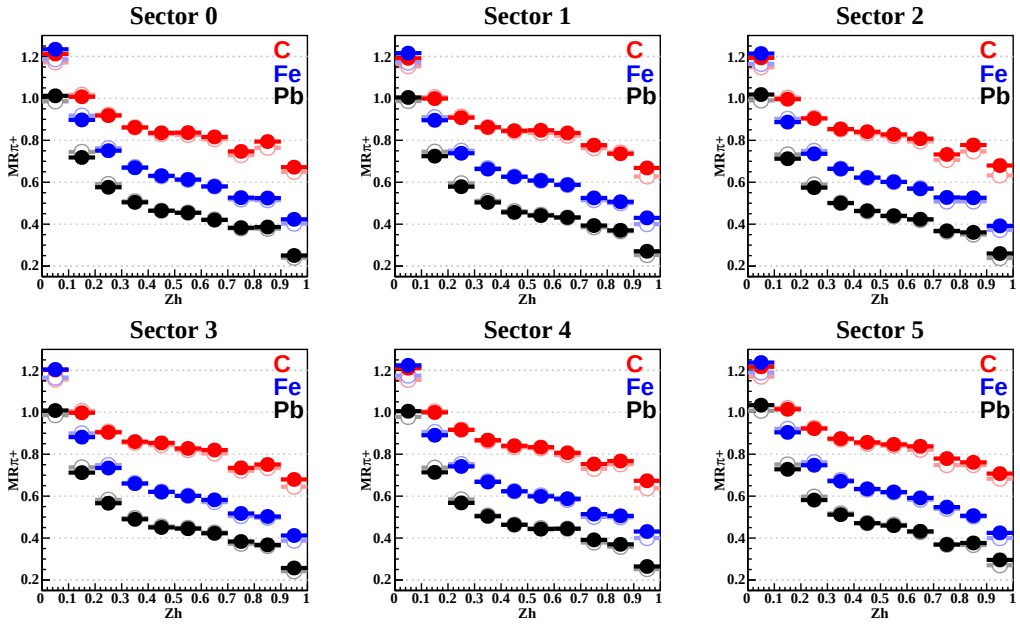


Figure 5.30: MR curves, for each sector, the hollow circles are the uncorrected case and the full circles are with acceptance correction applied.

Target	Sector	Normalization Factors		Effect (in %)
		Uncorrected	Acceptance Corrected	
C	0	1.103	1.099	-0.38
	1	1.102	1.088	-1.32
	2	1.081	1.079	-0.20
	3	1.091	1.084	-0.68
	4	1.067	1.069	+0.17
	5	1.090	1.094	+0.32
Fe	0	1.041	1.017	-2.44
	1	1.034	1.006	-2.82
	2	1.020	1.001	-1.81
	3	1.020	0.993	-2.68
	4	1.003	0.985	-1.81
	5	1.025	1.010	-1.48
Pb	0	2.183	2.114	-3.27
	1	2.187	2.101	-4.09
	2	2.126	2.071	-2.62
	3	2.142	2.071	-3.42
	4	2.109	2.060	-2.38
	5	2.146	2.102	-2.10

Table 5.9: Normalization factors for different targets and sectors.

5.4.2 Threshold P Value

The threshold for pion identification with the Cherenkov counter method was varied from the nominal 2.7 GeV to 2.5 GeV (see [23] as an example); the acceptance correction was updated accordingly. The deviation from the nominal result is depicted in figure 5.32 . No significant variation is observed for $z < 0.7$, as expected (see figure 5.31), while for $z > 0.7$ the results vary within -2% up to +1.0%, independently of the target type.

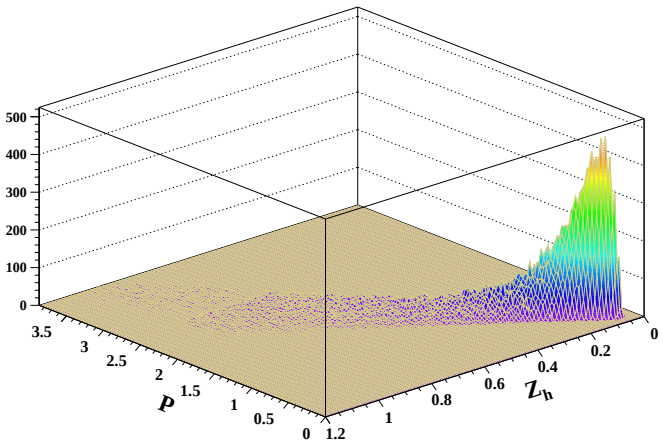


Figure 5.31: Three-dimensional representation of the (Z_h,P) phase space.

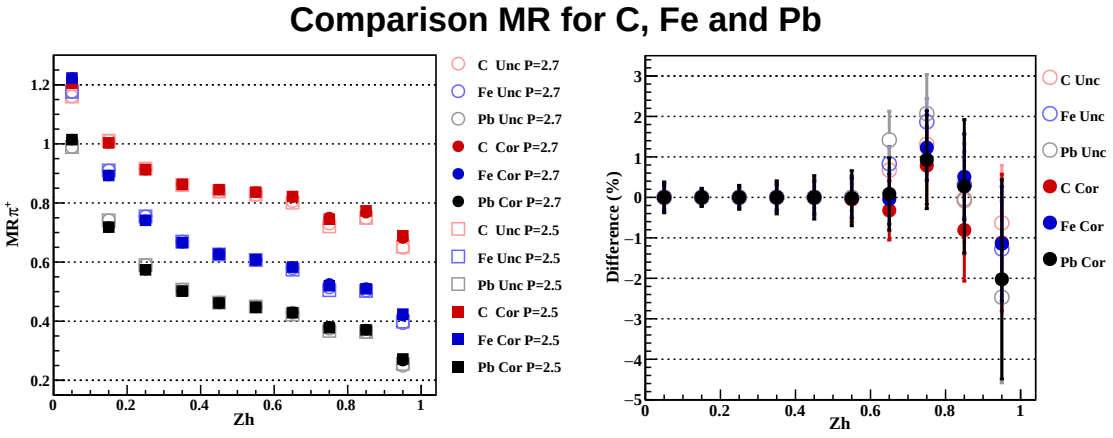


Figure 5.32: MR comparison for π^+ for two choices of threshold values for P in the pion selection cuts. Full symbols are with AC (4-fold) and hollow symbols are the uncorrected case.

5.4.3 ΔZ Cut Effect on MR

The impact of the application of the ΔZ cut is depicted in figure 5.33. Besides the nominal cut, two additional are considered. The effect of the ΔZ cut ranges from +1% down to -3% at high Z_h. The MR shows some systematic difference but do not change by more than ±0.5%, except at large Z_h where the statistical uncertainty is large.

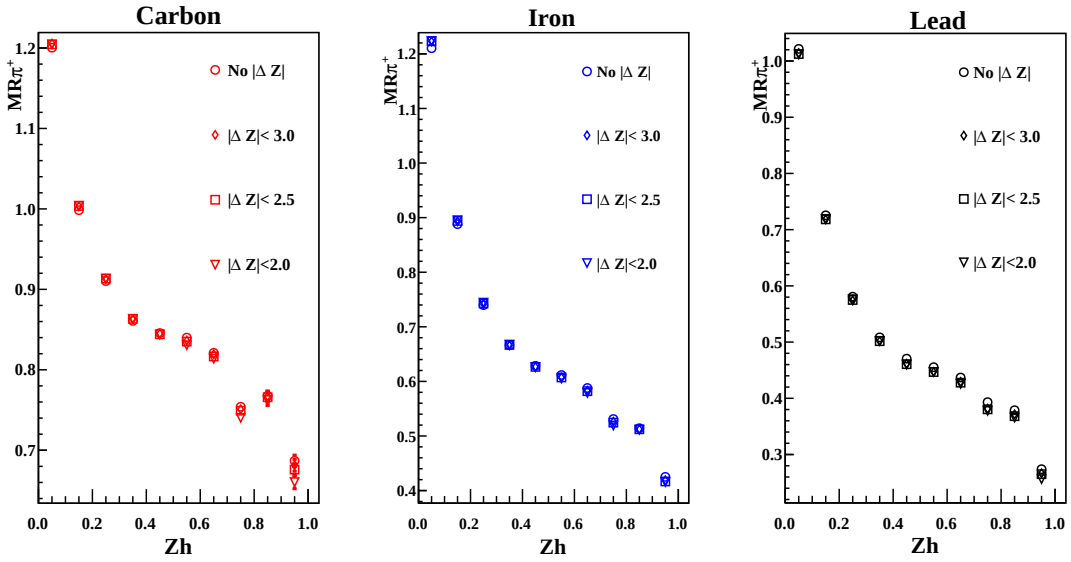


Figure 5.33: MR curves, for three different ΔZ cut, in [cm].

5.4.4 Feynman-X Cut on MR

The effect of a Xf cut is depicted in figure 5.34. The cut only affects the lower values of Z_h, reaching -8% for carbon and -18% for iron and lead. The effect is observed to increase with A.

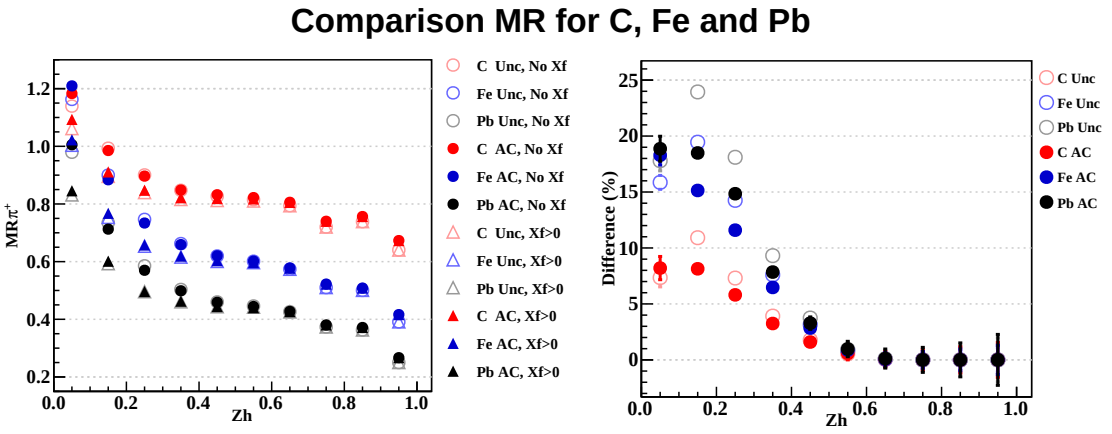


Figure 5.34: MR comparison as a function of Z_h, with AC (4-fold) for the full symbols and without AC for the hollow ones, with and without the $X_f > 0$ cut.

The hadrons measured with low energies (low Z_h) accumulate at values of X_f near zero (figure 5.35 left), exactly where target and current fragmentations joined. Following ref [24] we could select CFR at $Xf > 0.1$ to avoid the region of overlap. The MR comparison for such a case is depicted in figure 5.36.

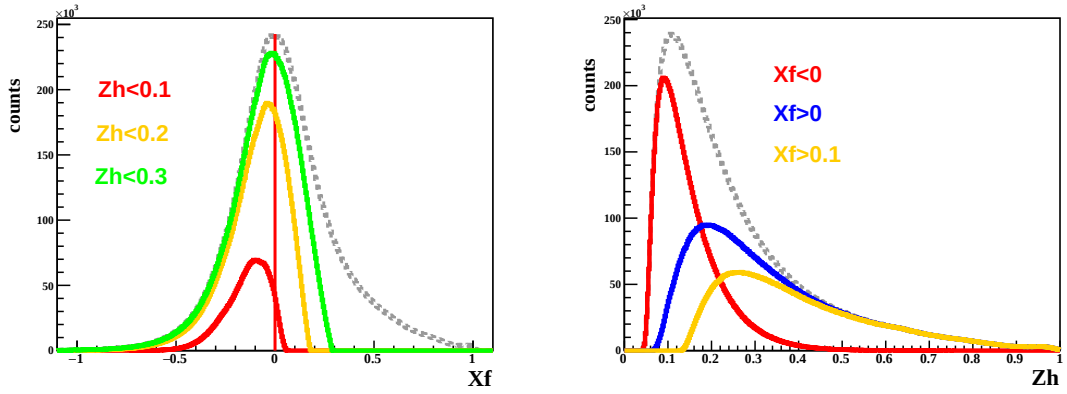


Figure 5.35: On the left, there is the one-dimensional X_f distribution for four different cases, No Z_h cuts (dashed gray line), $Z_h < 0.1$ (red), $Z_h < 0.2$ (orange) and $Z_h < 0.3$ (green). On the right, there is the Z_h distribution for four different cases, No X_f cut (dashed gray line), $X_f < 0$ (red), $X_f > 0$ (blue) and $X_f > 0.1$ (orange).

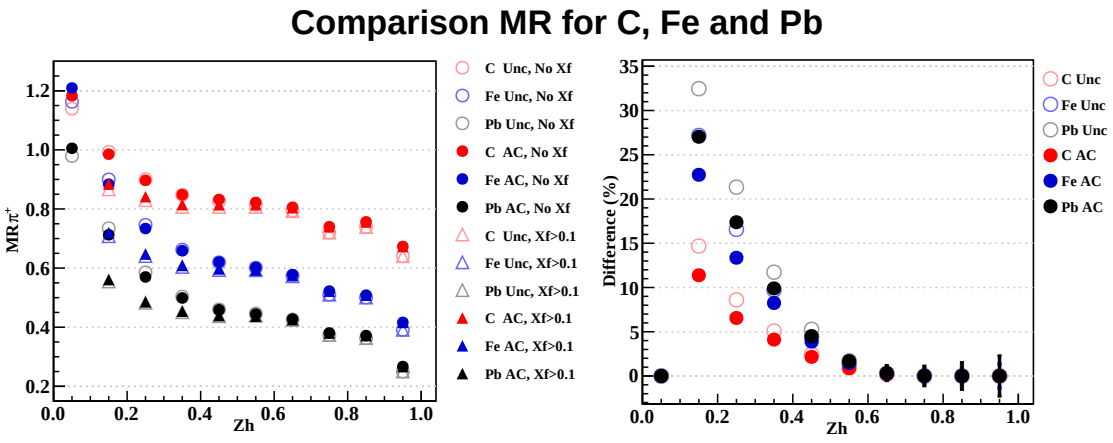


Figure 5.36: MR comparison for a CFR defined as $X_f > 0.1$.

5.4.5 Missing Mass Cut Effect on MR

To explore the effect on MR of a missing mass cut (section 4.5.4) we explore two different cuts. The proton mass and 1.4 GeV are used as limits for the cut. Figure 5.37 shows that the results vary less than 2% except at high Z_h. Figure 5.38 shows higer variations in the results with uncontroled fluctuations at high Z_h. Figures 5.39 and 5.40 show the same cuts in the MR, but considering only CFR. In this case no significant variation is observed for $Z_h < 0.6$. No target dependent effect is observed due to this cut.

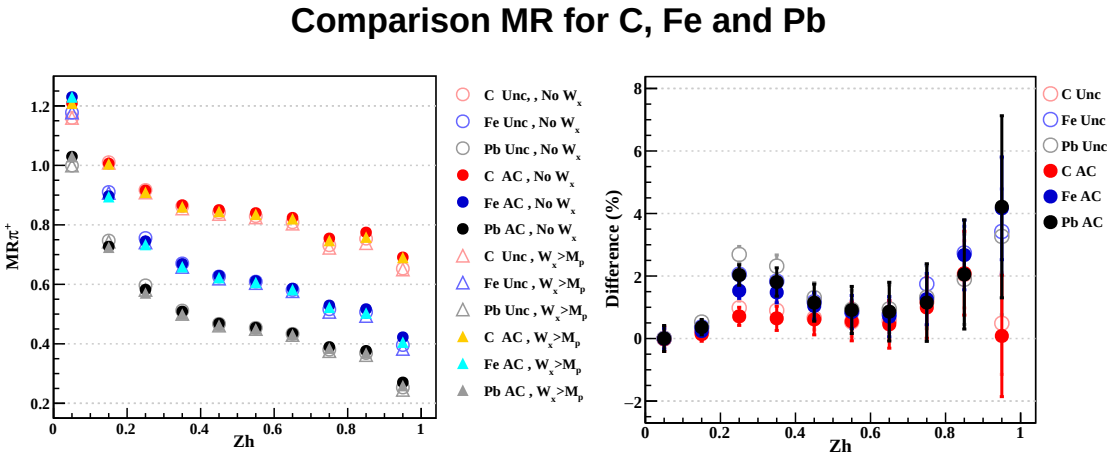


Figure 5.37: MR comparison between no cut in W_x and $W_x > M_p$ for C, Fe and Pb.

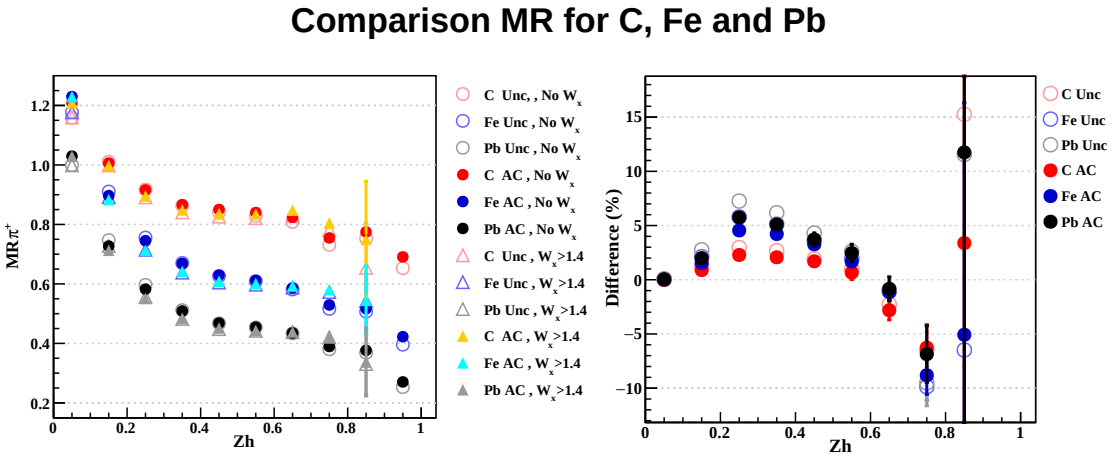


Figure 5.38: MR comparison between no cut in W_x and $W_x > 1.4$ GeV for C, Fe and Pb.

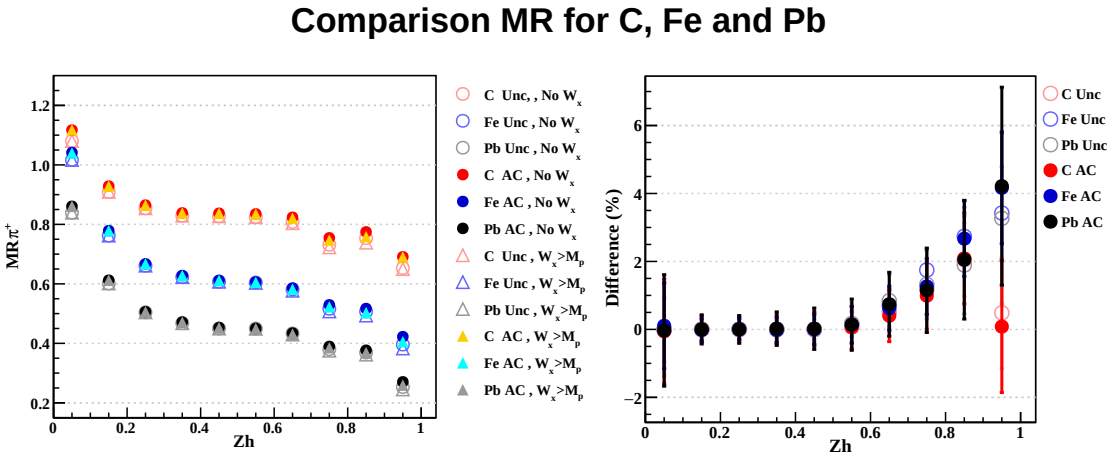


Figure 5.39: MR comparison between no cut in W_x and $W_x > M_p$ for C, Fe and Pb. CFR only.

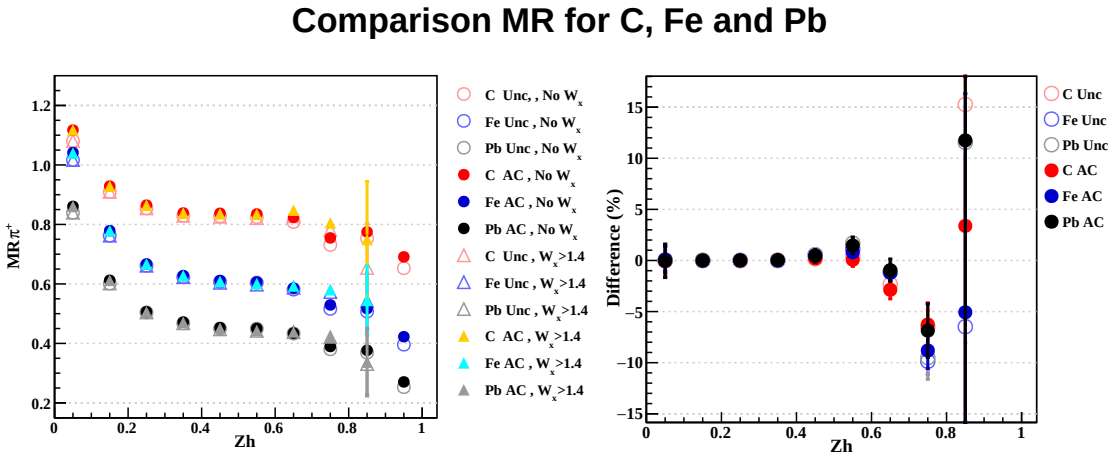


Figure 5.40: MR comparison between no cut in W_x and $W_x > 1.4$ GeV for C, Fe and Pb. CFR only.

5.4.6 Azimuthal Angle Relative to Photon Axis

According to figure 5.41 (a) our π^+ sample allows low P particles with big CC signals. The correlation between the electron and pion sectors is depicted in figure 5.41 (b). This can be thought of as a 6×6 matrix, where diagonal blocks represent pions and electrons measured in the same sector. Pions and electrons tend to be measured in opposite sectors. There should not be any big signal in CC for pions with momentum below the P threshold unless there is also an electron or positron in the same sector. Therefore, low P pions with big Nphe should pile up near $\phi_{\pi^+} = 0$ (see section 4.8.1.2). This is consistent with figure 5.41 (c) (red peak). We are going to refer to the red peak as *central peak* in ϕ_{π^+}.

There are individual mirrors in CCs in each sector, so we can apply mirror matching (section 5.3.6) to the pion track to make sure that we are isolating the mirrors corresponding to our pion, see figure 5.42. The cut only retains the pions that are in red in the top two plots. For low P pions, the CC cut basically removes all the pions in the region $Nphe > 2.5$. On the bottom of the figure, the (Nphe,P) space is plotted before and after the mirror matching. Notice the drop in the number of events for the bottom right plot.

Figure 5.43 left, shows a three-dimensional plot of the lab azimuthal angles for electrons and pions and the number of photo-electrons. The events in the diagonal position (highlighted in red) are low P π^+ with big CC signal . On the right, there is the ϕ_{π^+} distribution before and after the CC mirror cut; the central peak is almost totally suppressed.

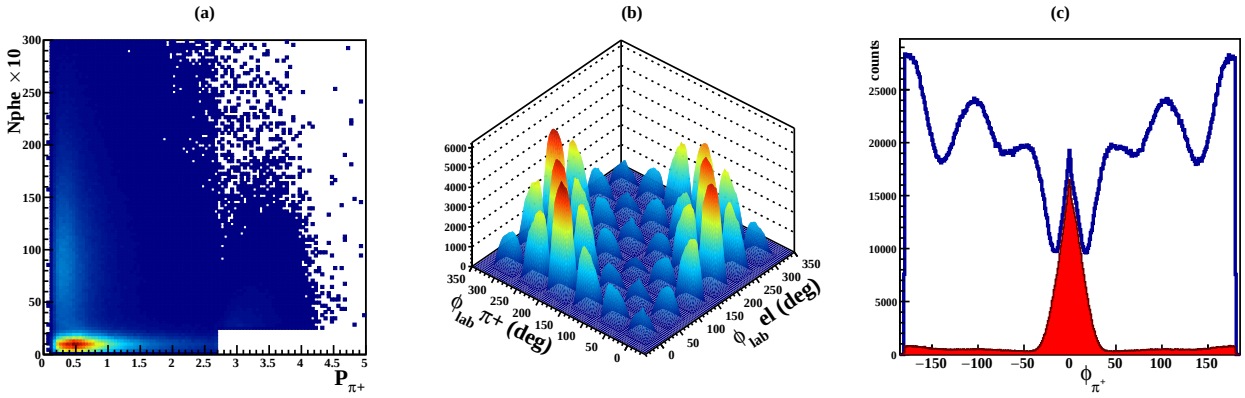


Figure 5.41: In (a), is the distribution of the number of photo-electrons (×10) v/s momentum of the pion (in GeV). In (b), there is the azimuthal angle for pions and electrons with respect to the lab frame. In (c), there is the azimuthal angle between leptonic and hadronic production plane for π^+: in blue, there is the total pion sample, and in red, there are the low momentum pions with number of photo-electrons bigger than 2.5.

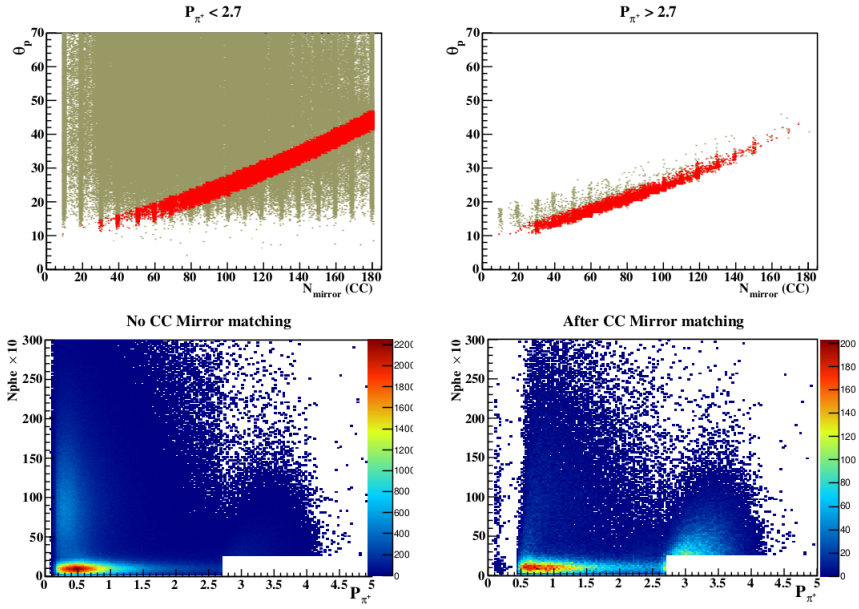


Figure 5.42: On the top, there are the projected polar angle in CC plane as a function of mirror number for low and high momentum values. The red strip are those events that pass the cut; the rest is removed. on the bottom, there is the Nphe×10 against the $P_{\pi+}$ (in GeV) before and after the CC mirror matching cut.

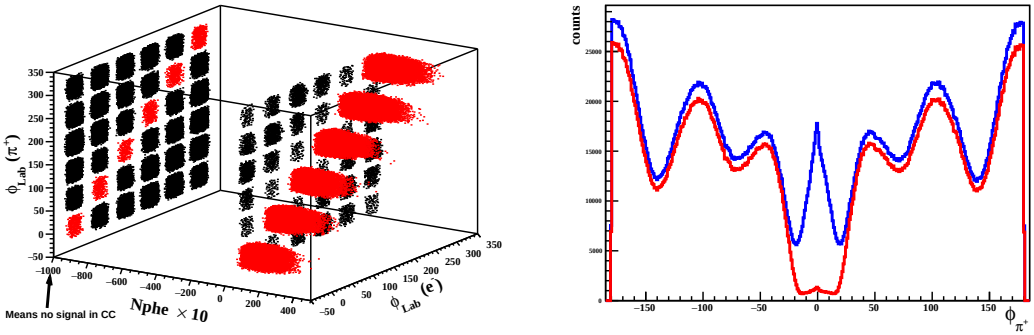


Figure 5.43: On the left, there is the azimuthal lab angles for electrons and pions as a function of Nphe×10. In red, there are the events that belong to the diagonal in the distribution of the angles. On the right, there are the ϕ_{π^+} distributions for two cases: before the CC cut in blue and after in red.

The remaining few events with low P π^+ and big CC signal most likely are positrons (from decay of π^0). Some authors use a cut in Nphe for hadrons. For example, in [79] a cut in number of photo-electrons below 1 is applied. Another approach to the central peak in ϕ_{π^+} is to implement a *fiducial* cut in ϕ_{π^+} claiming that this drop in the distribution is an *edge effect* due to acceptance [35]. There are certainly further studies needed on this subject.

The effect of the CC matching cut to the pion track on the 5-fold acceptance corrected MR is depicted in figure 5.44. The varitions observed are within ±2% for all Z_h range with bigger effect at low Z_h. A more drastic cut is shown in figure 5.45, where the central peak in ϕ_{π^+} is removed by definition, excluding events where electron and pions were measured in the same sector. The effect reach 3% at most at low Z_h.

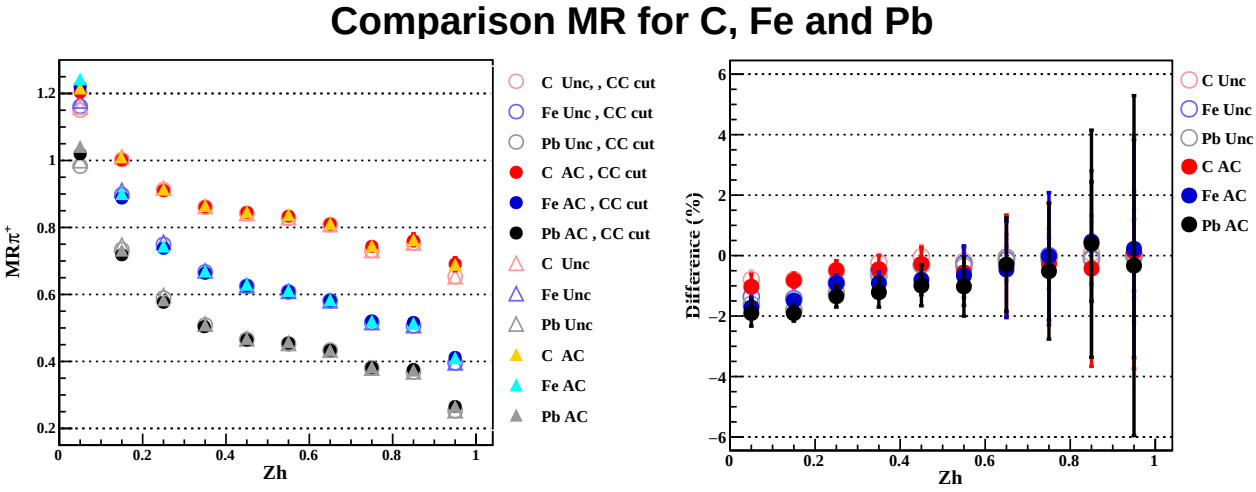


Figure 5.44: MR comparison for two cases, with and without CC mirror cut to the pion.

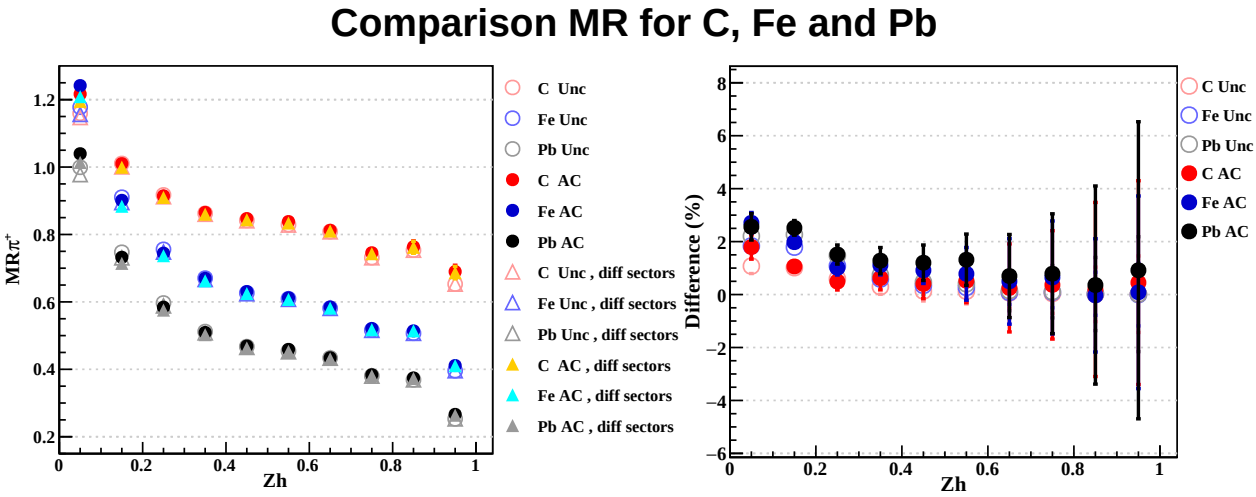


Figure 5.45: MR comparison between the previous case and with only events where the electron and the pion were measured in different sectors.

5.4.7 SIDIS Radiative Correction

In section 4.8.2.2 the radiative correction for π^+ using HAPRAD code are discussed. In that section is explained the fitting procedure to obtain structure functions. Figure 5.46 shows the reduced χ^2 obtained from those fits, which illustrates a reasonable goodness of fit. While the correction factors for the cross-sections are in the range

of 0.7–1.3, they largely cancel in the nuclei-to-deuterium ratio. Figures 5.47 and 5.48 show the nuclei-to-deuterium ratio of the radiative factors as a function of Z_h in different (p_T, ϕ_{pq}) intervals for the three targets. Figure 5.49 shows the impact of the SIDIS radiative corrections on the multiplicity ratio. The average effect is about -0.2% and it is not beyond -0.4% effect in the final result.

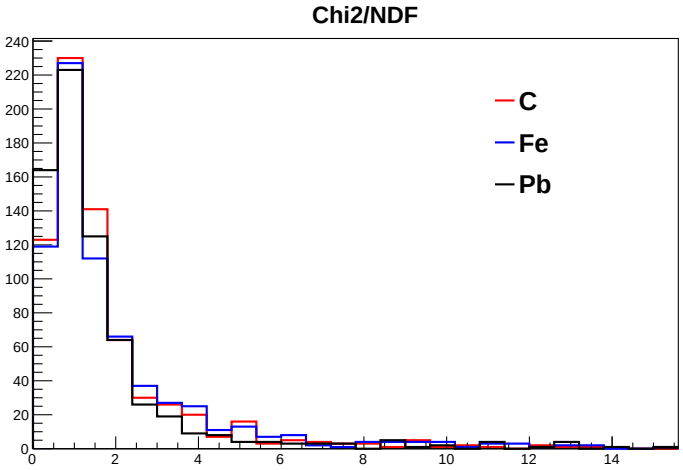


Figure 5.46: Reduced χ^2 for the fits used in the HAPRAD calculation.

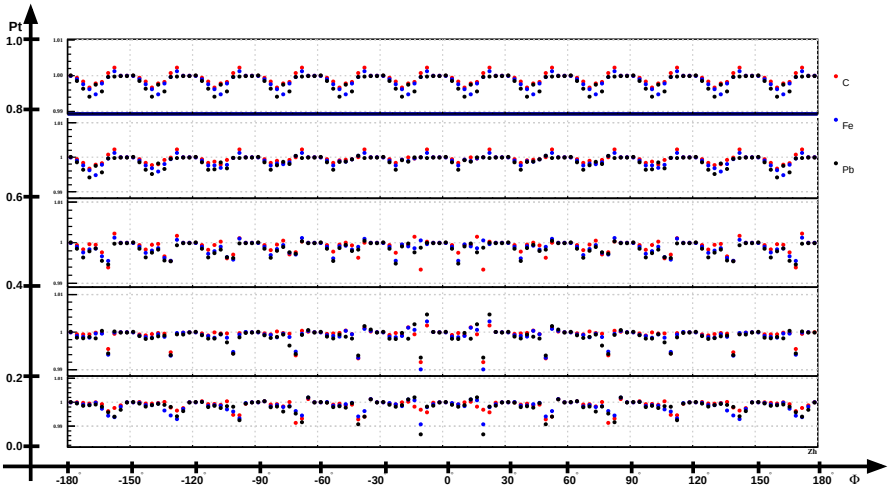


Figure 5.47: Radiative correction factors as a function of z for different ranges in p_T and ϕ_{pq}. This is for $1.0 < Q^2 < 1.5$ and $0.19 < X_b < 0.28$.

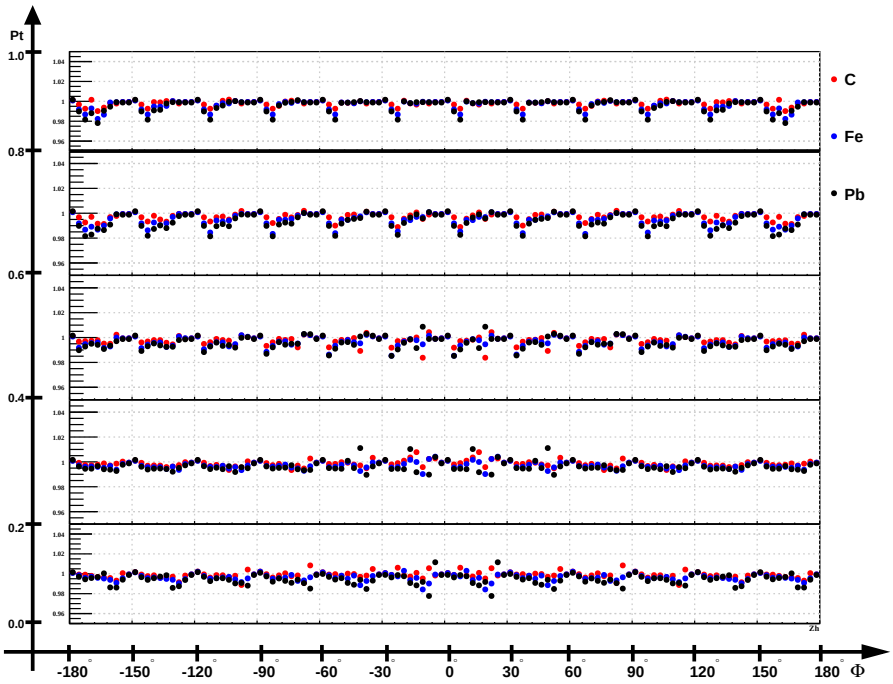


Figure 5.48: Radiative correction factors as a function of z for different ranges in p_T and ϕ_{pq}. This is for $1.5 < Q^2 < 2.0$ and $0.19 < X_b < 0.28$.

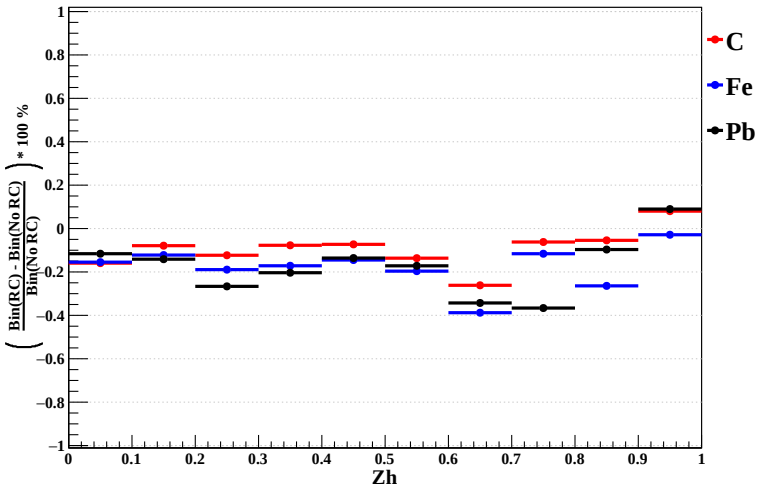


Figure 5.49: Effect of the radiative correction using HAPRAD code in the multiplicity ratio, for the three targets.

5.4.8 Electron Vertex Cuts Dependence on MR

The vertex cuts chosen to be compared were presented in section 5.3.3. The conclusions obtained for the MR are consistent with the ones obtained for the EMC ratio.

Figure 5.50 shows the MR for different vertex cuts before (left) and after (right) the YC cut. With YC cut applied, the MR for different vertex cuts are in agreement.

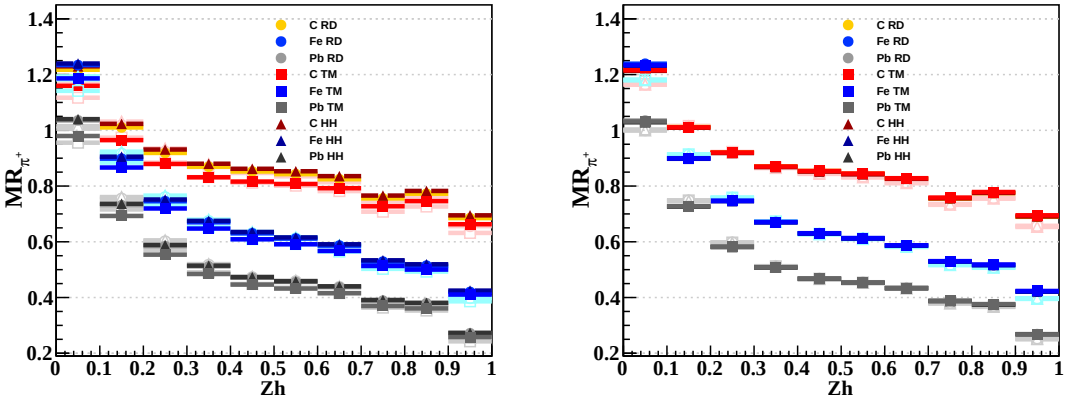


Figure 5.50: MR curves comparison, as a function of Z_h for three targets and for three different set of vertex cuts.

5.4.9 Bin Statistics

The aim of this section is to give information about the bin statistics. This information can tell us about problems related to bin migration or purity. Those topics were not discussed in detail in this work, but according to the information given in this section are expected to be small. For a 4-fold differential acceptance corrected MR there are $6 \times 10 \times 5 \times 5 = 1500$ bins in the grid, for Q2, Zh, Pt2 and Xb respectively. Getting rid of the bins out of the kinematics, that means all the bins with null entries in data, reconstruction and generation, and also the bins with only non null generated bins. The final number of bins to consider for each case are:

$$\begin{array}{lcll} \text{C} & : & 1500-354 & = \ 1146 \text{ bins} \\ \text{Fe} & : & 1500-355 & = \ 1145 \text{ bins} \\ \text{Pb} & : & 1500-356 & = \ 1144 \text{ bins} \\ \text{C+D2} & : & 1500-355 & = \ 1145 \text{ bins} \\ \text{Fe+D2} & : & 1500-355 & = \ 1145 \text{ bins} \\ \text{Pb+D2} & : & 1500-352 & = \ 1148 \text{ bins} \end{array}$$

In tables 5.10, 5.11, 5.12, 5.13, 5.14, and 5.15 are, in %, the number of empty/non empty bins in the data, reconstruction and generation for the three targets. As an example in table 5.10, the 2.79 value means that considering only the bins in data with no events, for carbon (solid target), there are 2.79% of the same bins but this time in generation that are not empty.

Empty Data Bins	**Carbon**			
	Solid Target		**Deuterium**	
	Empty Gen.	**Non Empty Gen.**	**Empty Gen.**	**Non Empty Gen.**
Empty Rec.	0%	0%	0%	0%
Non Empty Rec.	0.09%	2.79%	0.17%	1.75%

Table 5.10: Percentage of occupation for bins that are empty in real data. For carbon target.

Non Empty Data Bins	**Carbon**			
	Solid Target		**Deuterium**	
	Empty Gen.	**Non Empty Gen.**	**Empty Gen.**	**Non Empty Gen.**
Empty Rec.	0.09%	0.26%	0%	0.35%
Non Empty Rec.	0%	96.77%	0%	97.73%

Table 5.11: Percentage of occupation for bins that are not empty in real data. For carbon target.

Empty Data Bins	**Iron**			
	Solid Target		**Deuterium**	
	Empty Gen.	**Non Empty Gen.**	**Empty Gen.**	**Non Empty Gen.**
Empty Rec.	0%	0%	0%	0%
Non Empty Rec.	0%	3.14%	0.17%	1.31%

Table 5.12: Percentage of occupation for bins that are empty in real data. For iron target.

Non Empty Data Bins	Iron			
	Solid Target		Deuterium	
	Empty Gen.	Non Empty Gen.	Empty Gen.	Non Empty Gen.
Empty Rec.	0%	0.26%	0%	0.35%
Non Empty Rec.	0.09%	96.5%	0%	98.17%

Table 5.13: Percentage of occupation for bins that are not empty in real data. For iron target.

Empty Data Bins	Lead			
	Solid Target		Deuterium	
	Empty Gen.	Non Empty Gen.	Empty Gen.	Non Empty Gen.
Empty Rec.	0%	0%	0%	0%
Non Empty Rec.	0.09%	4.55%	0.17%	1.92%

Table 5.14: Percentage of occupation for bins that are empty in real data. For lead target.

Non Empty Data Bins	Lead			
	Solid Target		Deuterium	
	Empty Gen.	Non Empty Gen.	Empty Gen.	Non Empty Gen.
Empty Rec.	0%	0.09%	0.17%	0.44%
Non Empty Rec.	0%	95.28%	0%	97.3%

Table 5.15: Percentage of occupation for bins that are not empty in real data. For lead target.

In tables 5.16, 5.17 and 5.18 are the numbers of bins, in percentage, for 5 different ranges of numbers of events. The first case is the number of bins (in %) that have 0 events , the second is for five or less events in them, the third is for bins that have more than 5 events, but less than 10, and so on.

Carbon	**[0]**	**]0 - 5]**	**]5 - 10]**	**]10 - 15]**	**]15 - ... [**
Data	2.87958%	5.06108%	2.1815%	1.7452%	88.1326%
Rec.	0.34904%	4.71204%	1.48342%	1.57068 %	91.8848%
Gen.	0.17452%	0%	0.26178%	0.26178%	99.3019%
Data(D2)	1.9214%	5.93886%	1.74672%	0.960699%	89.4323%
Rec.(D2)	0.349345%	3.84279%	2.18341%	1.04803%	92.5764%
Gen.(D2)	0.174672%	0%	0.0873362%	0.262009%	99.476%

Table 5.16: Percentage of bins, for each case (real data, reconstructed data and generated data), with different ranges of events in them. For carbon target.

Iron	**[0]**	**]0 - 5]**	**]5 - 10]**	**]10 - 15]**	**]15 - ... [**
Data	3.1441%	3.93013%	1.65939%	0.960699%	90.3057%
Rec.	0.262009%	4.97817%	1.31004%	1.31004%	92.1397%
Gen.	0.0873362%	0%	0.349345%	0.262009%	99.3013%
Data(D2)	1.48472%	4.01747%	1.9214%	1.13537%	91.441%
Rec.(D2)	0.349345%	3.84279%	2.18341%	1.04803%	92.5764%
Gen.(D2)	0.174672%	0%	0.0873362%	0.349345%	99.3886%

Table 5.17: Percentage of bins, for each case (real data, reconstructed data and generated data), with different ranges of events in them. For iron target.

Lead	**[0]**	**]0 - 5]**	**]5 - 10]**	**]10 - 15]**	**]15 - ... [**
Data	4.63287%	5.94406%	2.18531%	2.01049%	85.2273%
Rec.	0.0874126%	5.06993%	1.04895%	1.48601%	92.3077%
Gen.	0.0874126%	0.0874126%	0.174825%	0.174825%	99.4755%
Data(D2)	2.09059%	4.18118%	2.09059%	1.21951%	90.4181%
Rec.(D2)	0.609756%	3.83275%	2.1777%	1.0453%	92.3345%
Gen.(D2)	0.348432%	0%	0.174216%	0.261324%	99.216%

Table 5.18: Percentage of bins, for each case (real data, reconstructed data and generated data), with different ranges of events in them. For lead target.

5.4.10 Fiducial Cuts for Pions

The potential mis-modeling of the acceptance edges in simulation is assessed by repeating the analysis with the fiducial cuts for the pion reconstruction. The method is analogous to the electron case, but with different parameters, given in detail in ref. [23]. These cuts are important for cross section measurements but for a double ratio the impact is smaller. The percentage of rejected events after the cut is in order

of 13% for π^+. The impact of this cut is illustrated in Figure 5.51. The multiplicity ratio results vary by less than $\pm 1\%$ within statistical uncertainty, without any significant trend with Z_h or target type.

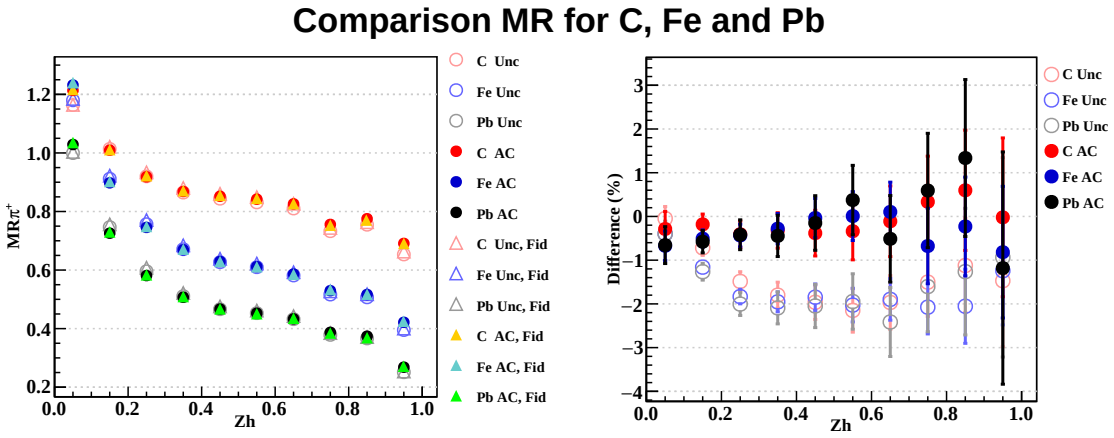


Figure 5.51: MR curves comparison, as a function of Z_h for three targets before and after the implementation of fiducial cuts for positive pions.

5.4.11 5-fold Differential Acceptance Correction

The potential bias due to dependence of the physics input on the simulation is assessed by repeating the analysis with an acceptance correction computed in a more differential way. As discussed previously the acceptance correction is nominally calculated in intervals of four variables (Q^2, X_b, Z_h and p_T^2) to minimize model dependency; to assess residual biases, the acceptance calculation is repeated using an extra kinematic variable, which is the azimuthal angle of the hadron relative to the photon axis: ϕ_{pq}. As shown in Figure 5.52, the MR do not change by more than about $\pm 1\%$ for $Z_h < 0.7$. For high Z_h the difference is bigger, reaching -3%. No significant trend with target type is observed.

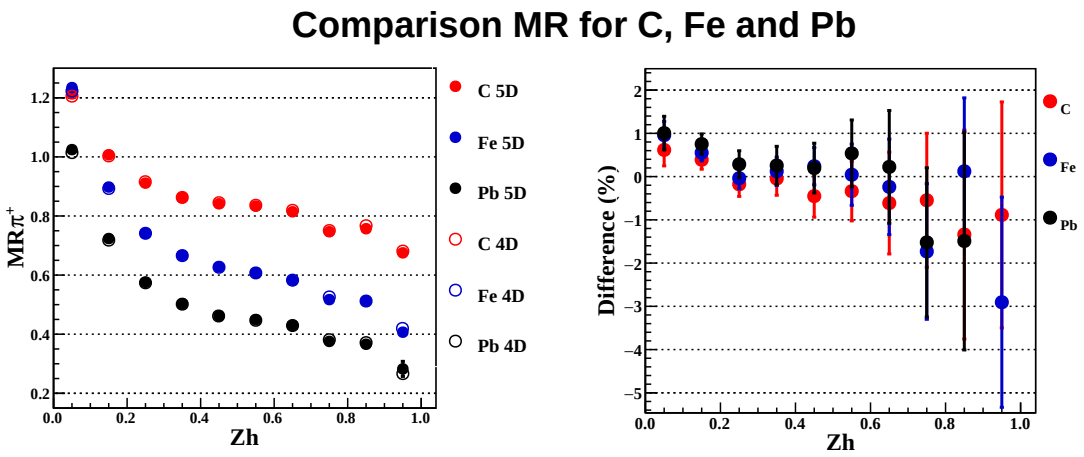


Figure 5.52: MR comparison between 4-fold and 5-fold acceptance correction.

Even though the data statistics drops considerably upgrading up to 5-fold dimensional acceptance correction there is still enough statistics to make a sector dependent MR, see figure 5.53. The results are consistent as it was for the 4-fold case. There is some issues at the edges. In principle the AC should be of the same size for all targets. The thicknesses of the targets are included in the simulation, and it makes a few-percent difference in the results. There is more multiple scattering for some targets than others. Therefore the fuzziness at the edges of the acceptance causes a few percent difference in the acceptances. One target is C, another is Pb at half thickness.

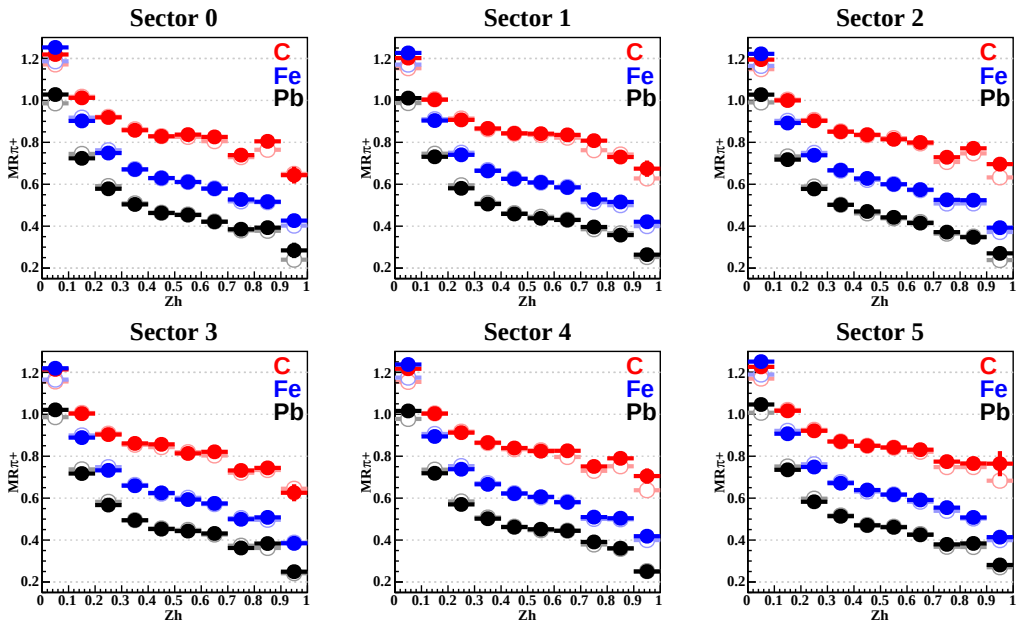


Figure 5.53: Comparison for 5-fold differential acceptance corrected MR (full circles), for each CLAS sector. The hollow circles are the uncorrected case.

5.5 Systematic Uncertainties

The double-target system of the EG2 experiment was designed to minimize systematic uncertainties. Exposing the two targets to the beam simultaneously causes effects due to dead channels, fluctuations of the beam, etc., to cancel out in the ratio since those effects are the same for both targets. Moreover, normalization uncertainties such as luminosity and trigger efficiency also cancel out.

The aim of this section is to present the identified systematic uncertainties. The main sources of systematic uncertainties for this measurement are related to vertex measurements, the reconstruction of electrons and charged pions, and radiative and Coulomb corrections. Many cases have been explored and discussed earlier in the chapter, therefore, this section will be a summary of some previous topics plus some new ones.

Additionally, in appendix A, there is a comparison between two independent analysis. The two analysis agree at $\pm 2\%$ level.

5.5.0.1 Vertex Selection and Target Identification

The uncertainty associated with the vertex selection was determined by repeating the analysis with three different selection and recalculating the correction factor in the

simulation. The deviation from the nominal result in the EMC ratio is plotted in Figure 5.54. A systematic trend is observed for the results with the TM cuts being lower than the nominal results over all the X_b range; the results with the HH cuts are above the nominal results at low X_b but consistent with the nominal results for $X_b > 0.2$. The deviation from the nominal result for the multiplicity ratio is plotted in figure 5.50. For the π^+ case, deviation from the nominal result is within $\pm 0.5\%$ for most of the z range except at low and high z where it is within $\pm 1.0\%$. The deviation from the nominal results is attributed to either a mis-modeling of the background level in the simulation and/or a potential mis-tagging of the target type. This effect is more significant for the EMC ratio than the multiplicity ratio.

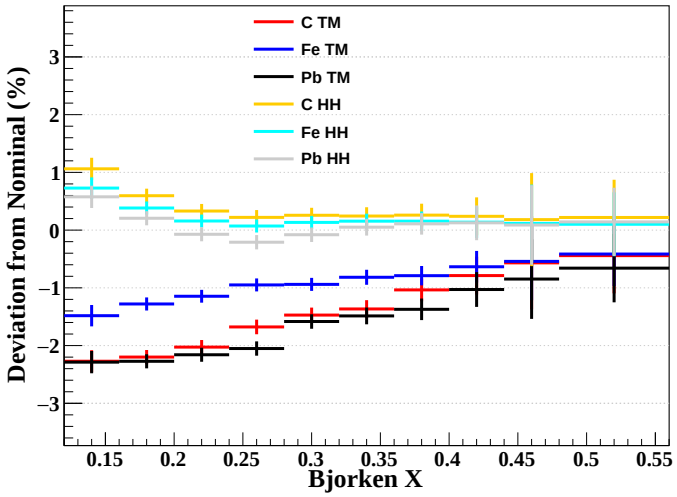


Figure 5.54: Deviation from the nominal case resulting from a variation in the vertex selection for the EMC ratio. The error bars represent the statistical uncertainty of the nominal case.

In addition, the selection on the longitudinal vertex separation between the electron and the charged pion is varied from the nominal $|\Delta z| < 3.0$, see figure 5.33. The calculation of the vertex correction with simulation is updated accordingly. The MR shows some systematic difference but do not change by more than $\pm 0.5\%$, except at large z where the statistical uncertainty is large.

Another source of uncertainty comes from the background of the cryo-target aluminum entrance and exit windows (so-called endcaps), which are illustrated in figure 3.10. The cryo-target walls are composed of aluminum with 15 μm thickness. The residual background in the deuterium yield is estimated from an empty-target run

using an analysis that also takes into account nuclear effects in the Al target, as described in detail in the analysis note of the recently-approved π^0 analysis that uses the EG2 dataset [60]. The effect in the final multiplicity ratio is $\leq 0.1\%$ for all target types. No correction is applied to the data. Rather, we include a $\pm 0.1\%$ normalization uncertainty on the multiplicity ratio results for all target types.

In summary, based on the sensitivity of the results to the variations of the vertex selection, we assign a normalization uncertainty on the EMC ratio that ranges from $\pm 2.5\%$ at low X_b to $\pm 1\%$ at large X_b; the same systematic uncertainty is assigned for all target types. For the multiplicity ratio case, we assign a normalization uncertainty that is $\pm 0.5\%$ results; the same systematic uncertainty is assigned for all target types.

5.5.0.2 Target Thickness and Stability

The solid targets' lengths are known to within 1 μm [61], which yields normalization uncertainties of $\pm 0.7\%$ for Pb, $\pm 0.25\%$ for Fe, and $\pm 0.1\%$ for C respectively. This uncertainty cancels in the multiplicity ratio but it propagates as a normalization uncertainty in the EMC ratio.

Given that the run was operated with a current of a few nA, no significant boiling of the cryo-target is expected. This is confirmed with computational fluid-dynamics simulations [49] [50]. No systematic uncertainty is assigned to this effect.

Studies of the charge-normalized yield of electrons scattering from over all the EG2 deuterium runs shows stability with a $\pm 0.65\%$ standard deviation, which was taken as a systematic uncertainty in ref. [48]. Given that we use the same dataset and a very similar electron selection, we take the same systematic uncertainty for the EMC ratio we report.

In summary, the following normalization uncertainties are assigned to the EMC ratios to account for uncertainty in target thickness and stability of the cryo-target: $\pm 1.2\%$ for Pb, $\pm 1.1\%$ for Fe, and $\pm 1.0\%$ for C.

5.5.0.3 Acceptance Correction

We consider two possible sources of systematic uncertainties on this correction: mis-modeling of the detector response and the physics-model input. The first one is assessed by performing independent measurmements for each CLAS sector, see figures 5.5, 5.30 and 5.53. For the EMC ratio the results are found to be consistent within $\pm 2\%$, except at the highest values of X_b, where the deviations are within $\pm 5\%$. For the multiplicty ratios the results are consistent with the average value within $\pm 2\%$, except at the $z \to 1$ limit where the deviations are within $\pm 5\%$. The mis-modeling is also

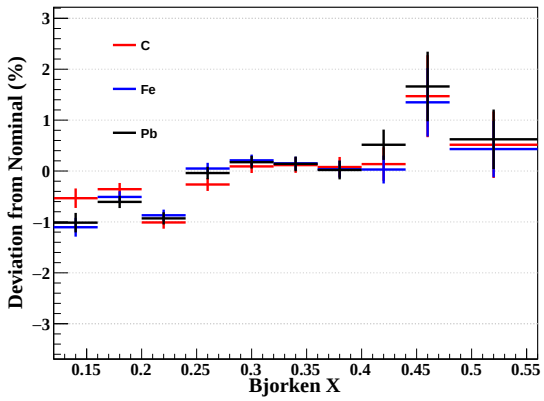


Figure 5.55: Deviation of the EMC ratio (in %) resulting from removing the fiducial cuts in the electron sample. The error bars represent the statistical uncertainty of the nominal case.

considered in the application of fiducial cuts for pions (see figure 5.51) and electrons (see figure 5.55). The physics-model input is assessed by repeating the analysis with an acceptance correction computed in a more differential way (see figure 5.52) and varying the choice of binning used in simulation(see figure 5.56 and table 5.19). The multiplicity ratios do not change by more than about ±2%, with some systematic patterns in Z_h.

Binning	
Variable	Edges
Q2	(1.0, 1.19, 1.38, 1.62, 2.0, 4.0)
ν	(2.2, 3.23, 3.55, 3.79, 4.0, 4.2)
Pt2	(0.0, 0.05, 0.11, 0.2, 0.36, 1.5)
Zh	(0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0)

Table 5.19: Bin edges.

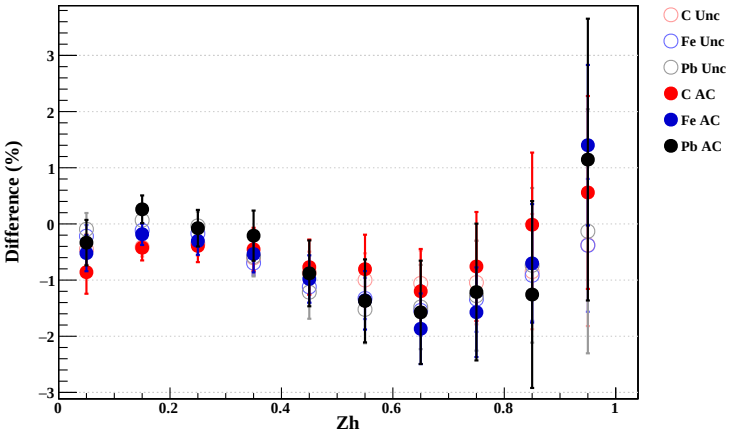


Figure 5.56: MR comparison for the three targets for two different binning choices.

In summary, the total systematic uncertainty of the acceptance correction for the EMC ratio is estimated to be between ±2% at low X_b and ±5% at high X_b. This is taken to be a normalization uncertainty for all target types. For the multiplicity ratio, the uncertainty is estimated to be ±3.0–6.0% depending on Z_h.

5.5.0.4 Pion Identification

The threshold for pion identification with the Cherenkov counter method was varied from the nominal 2.7 GeV to 2.5; the acceptance correction was updated accordingly, see figure 5.32. No significant variation is observed for $z < 0.7$, as expected, while for $z > 0.7$ the results vary within ±1.0%, independently of the target type. A systematic uncertainty of ±1% is assigned to the multiplicity ratios for $z > 0.7$. To estimate the kaon contamination we follow a conservative approach and take a ±3% [25] normalization uncertainty for the multiplicity ratio for $Z_h > 0.6$ for all target types.

5.5.0.5 Electron Identification

The systematic uncertainties related to electron identification are estimated by varying the sampling fraction selection and using an alternative selection to suppress π^- (using CC matching cut instead of Nphe, see section 4.3.1.1 and 5.3.6). The sampling-fraction selection is varied from the nominal ±2.5σ to ±2.0σ and ±3.0σ; the correction factors are recalculated accordingly. The resulting effect on the corrected EMC ratios is shown in figure 5.57. The variation with respect to the nominal result is mostly within ±0.1% for C and Pb and within ±0.2% for Fe. A normalization uncertainty of ±0.2% is assigned to the EMC ratio for all target types.

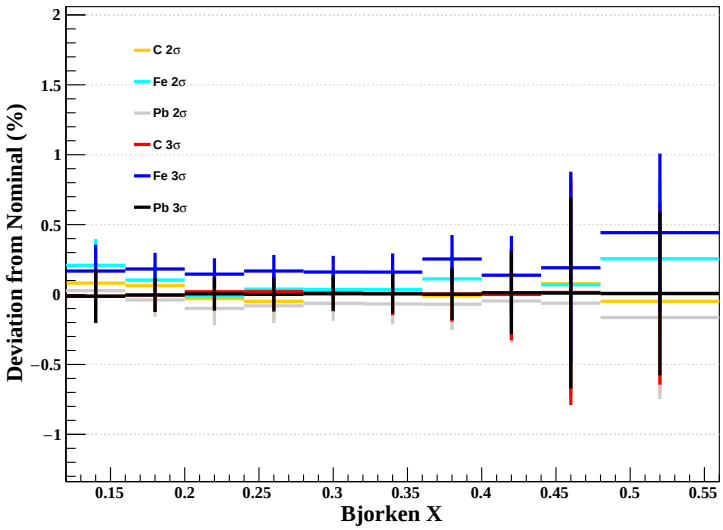


Figure 5.57: Variation of the corrected EMC ratio given by the variation on the sampling fraction cut for electron ID (nominally $\pm 2.5\sigma$). The error bars represent the statistical uncertainty of the nominal case.

The effect of the sampling-cut variation on the corrected multiplicity ratios is shown in figure 5.58. Some systematic patterns are observed but in general the results do not vary by more than $\pm 0.5\%$, except at $z \approx 1$ where observed variations can be attributed to statistical fluctuations. A normalization uncertainty of $\pm 0.5\%$ is assigned to the multiplicity ratio, for all target types.

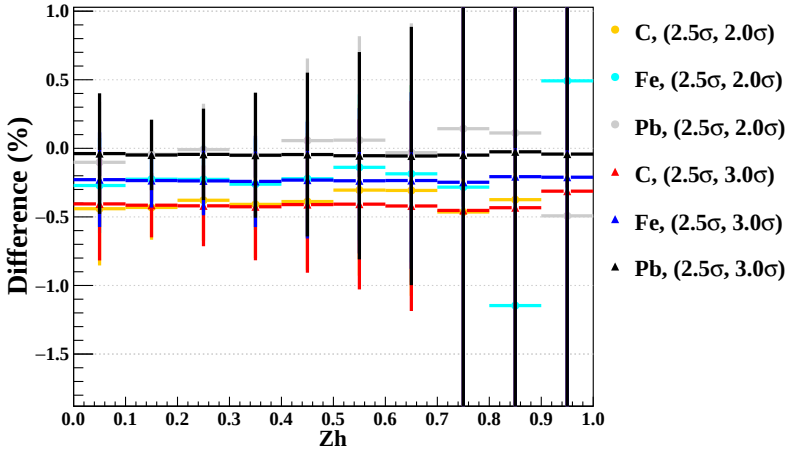


Figure 5.58: MR comparison for the three targets for different sampling fraction cuts.

Using CC mirror matching cut instead of Nphe for π^- rejection was also studied.

Figure 5.25 shows the impact of this variation on the EMC ratio. The deviation from the nominal result varies from ±1.0%; no significant difference among the target types is observed. Figure 5.59 shows the corresponding impact on the multiplicity ratio, which is on average about +0.5% without Z_h dependence, as expected; no significant difference between target types is observed.

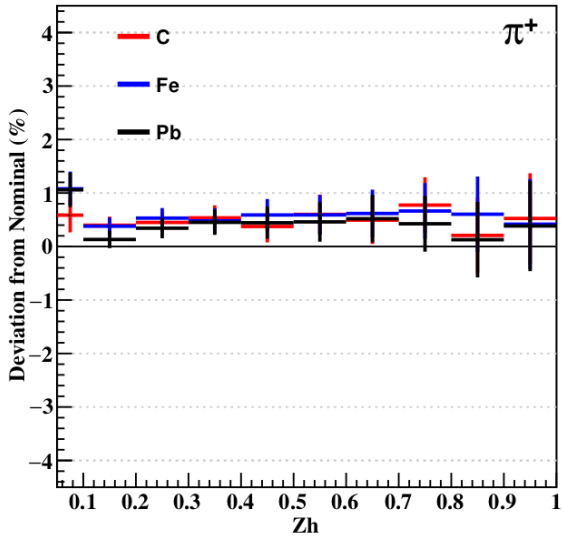


Figure 5.59: Deviation from the nominal multiplicity ratio measurement obtained by using the CC matching mirror technique instead of a hard cut in the number of photo-electrons for electron ID. The error bars represent the statistical uncertainty of the nominal case.

5.5.0.6 Radiative and Coulomb Corrections

Systematics associated to the CC and RC corrections are estimated by repeating the analysis using the INCLUSIVE [62] program rather than the nominal EXTERNAL program. Figures 5.60 and 5.61 show the RC and CC factors as a function of X_b and Q^2 for different values of the polar angles using INCLUSIVE. Figures 5.62 and 5.63 show the ratio between the corrected EMC ratio using one program to extract the factors or the other. The absolute difference between the EMC ratio results obtained with EXTERNAL and INCLUSIVE are 1% to 4% for Fe and Pb and between −4% and +2% for C. The largest deviations are observed at low X_b. The full absolute difference is taken as a symmetric uncertainty on the EMC ratio. This uncertainty considers the combined effect of the radiative and Coulomb corrections on the inclusive DIS yield. For the SIDIS RCs we follow the studies in [60], where a comparsion between the original HAPRAD code written in FORTRAN and a $C++$

version took place. The relative difference between the two codes bears an average 0.5% difference.

5.5.0.7 Summary

Tables 5.20 and 5.21 summarize the systematic uncertainties for the EMC ratio and multiplicity-ratio measurements.

	C	Fe	Pb
Electron identification	1.0%	1.0%	1.0%
Acceptance correction	2.0–5.0%	2.0–5.0%	2.0–5.0%
Vertex selection	1.0–2.5%	1.0–2.5%	1.0–2.5%
DIS Coulomb & rad. corr.	2–4%	1–4%	1–4%
Target thickness	1.0%	1.1%	1.2%
Luminosity	negligible.	negligible.	negligible.
Trigger efficiency	negligible.	negligible	negligible
Time-dependent effects	negligible	negligible	negligible.
Total systematic uncertainty	3.3–7.0%	2.8–7.0%	2.9–7.0%

Table 5.20: Summary of the systematic uncertainties on the EMC ratios. The quoted range spans the uncertainties across all kinematic intervals.

	C	Fe	Pb
Acceptance correction	3.0–6.0%	3.0–6.0%	3.0–6.0%
Vertex selection	0.5%	0.5%	0.5%
DIS Coulomb & rad. corr.	2–4%	1–4%	1–4%
SIDIS rad. corr.	0.5%	0.5%	0.5%
Pion identification	$< 3\%$	$< 3\%$	$< 3\%$
Electron identification	0.7%	0.7%	0.7%
Luminosity	negligible.	negligible.	negligible.
Trigger efficiency	negligible.	negligible	negligible
Time-dependent effects	negligible	negligible	negligible.
Total systematic uncertainty	4.7–7.8%	3.7–7.8%	3.7–7.8%

Table 5.21: Systematic uncertainties on the multiplicity ratios. The quoted range spans the uncertainties across all kinematic intervals.

5.6 Multidimensional Multiplicities and Discussion

One of the key points of EG2 data is to probe multidimensionality. The relatively high luminosity ($\mathcal{L}$) in combination with the large acceptance affords high statistical

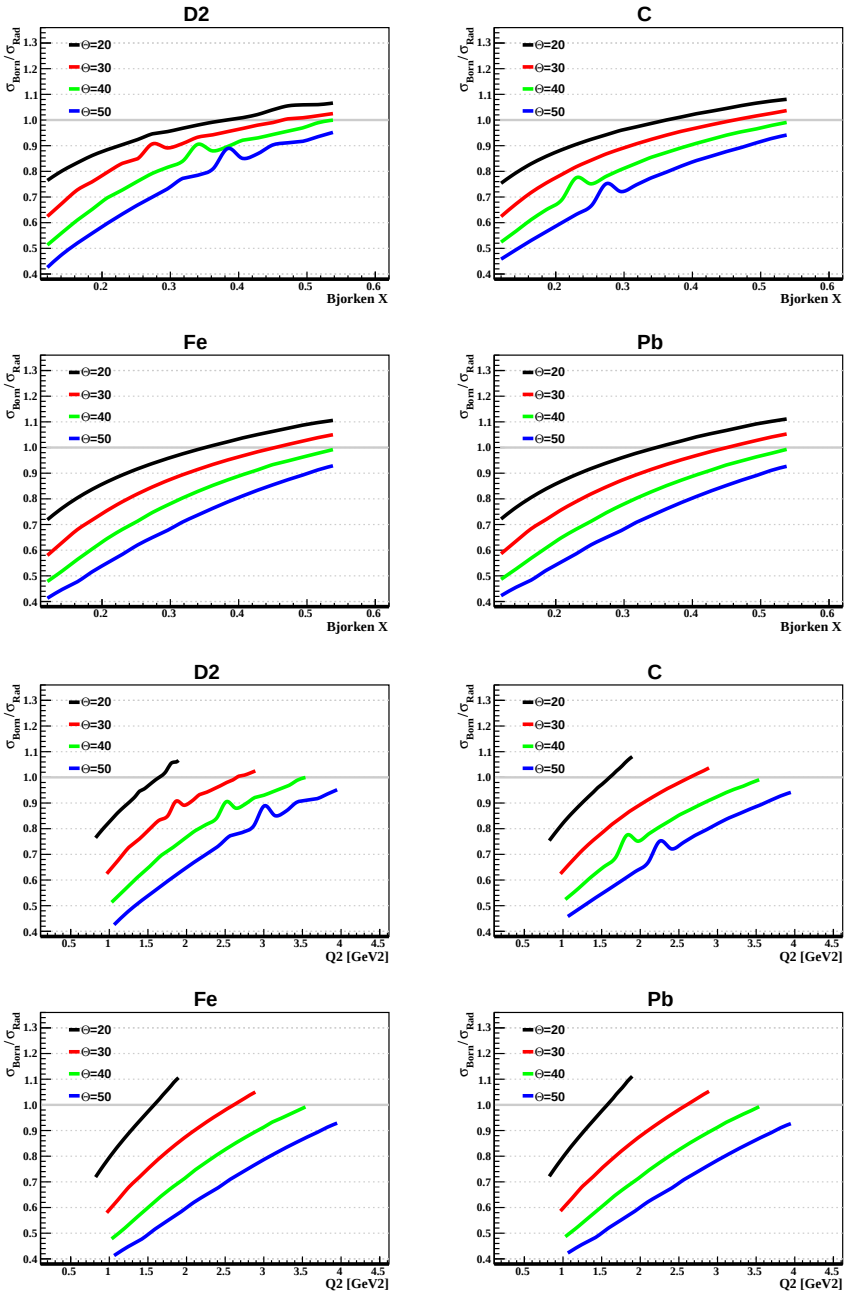


Figure 5.60: Radiative correction factors, for C, Fe and Pb as a function of X_b (upper 4 panels) and Q^2 (lower 4 panels) using INCLUSIVE code.

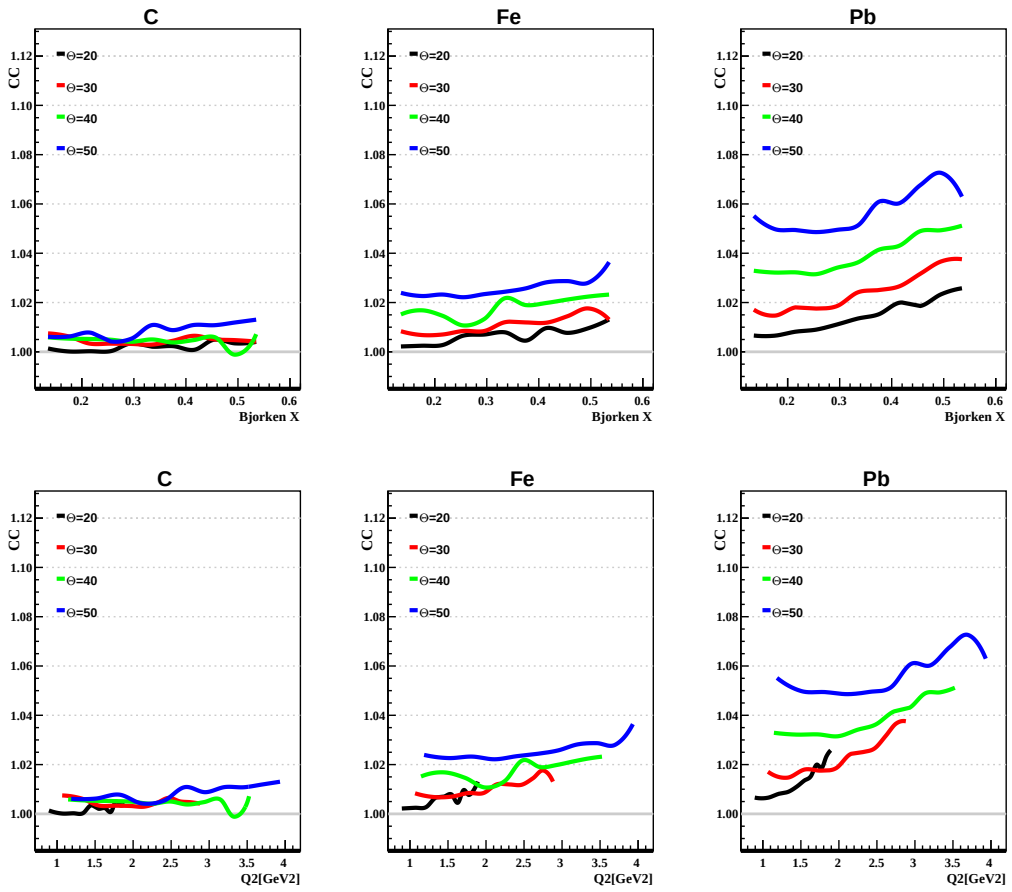


Figure 5.61: Coulomb correction factors, for C, Fe and Pb as a function of X_b (upper panels) and Q^2 (bottom panels). These correction factors have been obtained using INCLUSIVE code.

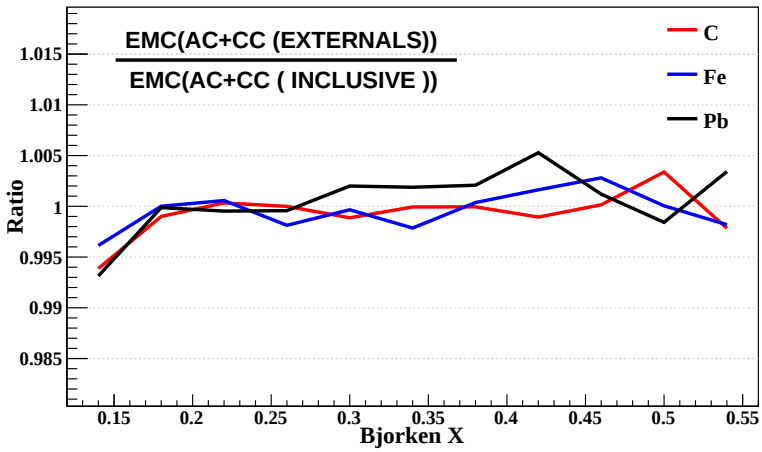


Figure 5.62: Ratio between the Coulomb corrected EMC ratio using EXTERNAL package to the Coulomb corrected EMC ratio using INCLUSIVE package, for the three targets.

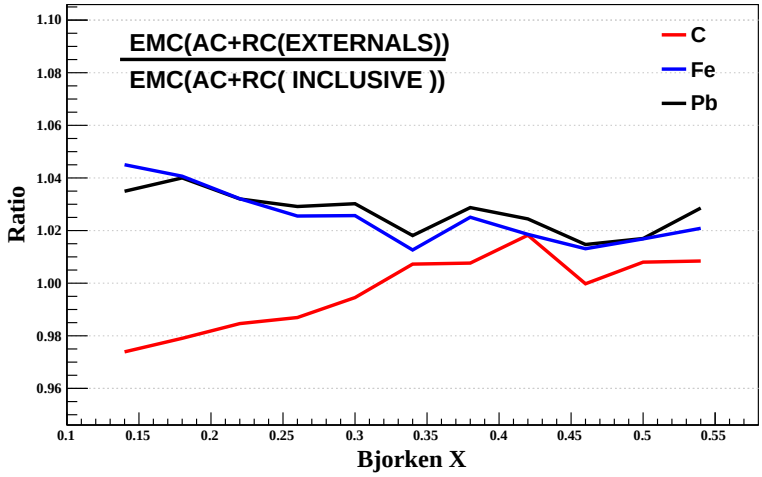


Figure 5.63: Ratio between the radiative corrected EMC ratio using EXTERNAL package to the radiative corrected EMC ratio using INCLUSIVE package, for the three targets.

accuracy over a wide range in kinematics. The largest statistics is for positive pions, so we can study up to 4-fold differentiated multidimensionality[7]

The aim of this section is to show the results of MR as a function of some kinematic quantities of interest and discuss their main features. The multidimensional studies of the MR are important because the integration over a wide range of kinematics and variables may introduce biases, false dependencies in the final result. In general the dependency of the MR on ν, Zh, Q^2 and P_t^2 does not factorize (that means $(\mathrm{MR}(\nu, Zh, Q^2, P_t^2) \neq \mathrm{MR}_1(Zh)\mathrm{MR}_2(Q^2)\mathrm{MR}_3(\nu)\mathrm{MR}_4(P_t^2)$).

For the purpose of multidimensional analysis following binning set is used(based on [60] [59]), the values are given in table 5.22: [8].

Binning	
Variable	Edges
Zh	(0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 1.0)
Pt2	(0.0, 0.1, 0.25, 0.4, 0.55, 0.75, 1.5)
ν	(2.20, 3.20, 3.73, 4.25)
Xb	(0.12, 0.20, 0.25, 0.57)
Q2	(1.0, 1.33, 1.76, 4.1)

Table 5.22: Bin scheme used for multidimensional multiplicities.

We know that the detected hadrons carry information on the flavor of the struck quarks from which they originate. One aspect that we haven't consider yet is the quantum fluctuations of the virtual photon γ^*. Consider the following reaction, this is called a diffractive [9] process:

$$\gamma^* + p \longrightarrow V^0 + p \qquad , \qquad V^0 = (\rho^0, \omega, \phi, ...) \tag{5.14}$$

where V^0 is a vector meson. Exclusive production of vector mesons (ρ^0, ω, or ϕ) can be described in the **V**ector **M**eson **D**ominance (VMD)[10] model as the fluctuation of the virtual photon into a $q\bar{q}$ pair before its interaction with the target nucleon. These vector mesons subsequently decay into lighter hadrons that are then found in the final hadronic state. They are contamination to our SIDIS sample, because

[7] negative pions has more complicated acceptance, but multidimensionality studies are still possible [59].

[8] in some cases there is change in the binning to not exclude any region of interest. The new values will be explicitly mentioned in the plot.

[9] The term diffraction was introduced in nuclear high energy physics in the Fifties. Apparently, the very first to use it were Landau and his school [69].

[10] VMD connects photon-nucleon with meson-nucleon amplitudes, for more details see, for example [63].

those hadrons do not involve the fragmentation of quarks originating from the target nucleon[11]. Diffractive events ([63], [31], [71], [72], [23]) are characterized by the property that the nucleus is left intact in the process and is surrounded by an angular region in which no particle is observed, called "gap" (there is a gap in the rapidity distribution of the produced particles [71]).

The major contribution due to exclusive vector mesons to the final state hadron sample arises in the form of pions originating from ρ^0 decay:

$$\rho^0 \longrightarrow \pi^+\pi^- \tag{5.15}$$

Due to its anisotropic decay-angle distribution, pions from ρ^0 decay are concentrated at low and high Z_h. For the low statistics high Z_h region near $Z_h = 1$, it is estimated that up to 50% (estimation by HERMES) of the charged pions originate from ρ^0 , while at low Z_h this fraction is negligible due to the much larger yield from other channels. See figure 5.64 for the reaction under study.

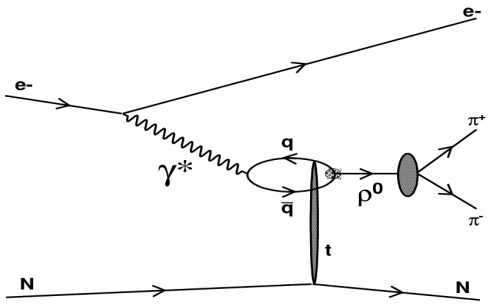


Figure 5.64: Diffractive reaction where the charge hadrons originate from a vector meson ρ^0. Taken from [23].

To see these contributions in our data, our events of interest require the detection of the scattered electron and the two oppositely charged pions from the ρ^0 decay. Let's calculate the square of the 4-momentum transferred to the nucleon (t mandelstam variable):

[11] If fragmentation functions were to be extracted from multiplicities that include such an exclusive production, they would be process dependent.

$$t \equiv (q - P_{\pi^+} - P_{\pi^-})^2 \tag{5.16}$$
$$= (q - P_{\pi^+})^2 + P_{\pi^-}^2 - 2(q - P_{\pi^+})P_{\pi^-} \tag{5.17}$$
$$= -Q2 + (Z_{\pi^+}\nu)^2 - (\vec{p}_{\pi^+})^2 - 2\nu^2 Z_{\pi^+} \tag{5.18}$$
$$+ 2p_{\pi^+}\sqrt{\nu^2 + Q^2}cos(\theta_{\pi^+\gamma^*}) + (Z_{\pi^-}\nu)^2 - (\vec{p}_{\pi^-})^2$$
$$- 2\left(\nu^2 Z_{\pi^-} - p_{\pi^-}cos(\theta_{\pi^-\gamma^*})\sqrt{\nu^2 + Q^2} - \nu^2 Z_{\pi^-}Z_{\pi^+} + p_{\pi^-}p_{\pi^+}cos(\theta_{\pi^-\pi^+})\right)$$

Here we only consider events with just one π^+ and one π^-. The way to identify the π^- is, in principle, very similar to that of the π^+, but this time the charge has to be negative and only a cut in $\Delta\beta$ is applied ([25] as an example):

$$|\Delta\beta| < 0.03 \quad , \quad \text{Negative pion selection cut} \tag{5.19}$$

According to [23] to select the diffractive and elastic processes and to exclude coherent production we must select the region: $0.1[\text{GeV}^2] \leq t \leq 0.4[\text{GeV}^2]$. In figure 5.65 it is shown the one-dimensional distribution of the t variable and the region in consideration is the one within the red lines. This region, viewed in the Z_h distribution is, as expected, at the highest values. This is a very important feature of our analysis, because the statistics in this region is considerably low and, at the same time being contaminated. That tells us that we must be very careful with the interpretation at high Z_h values.

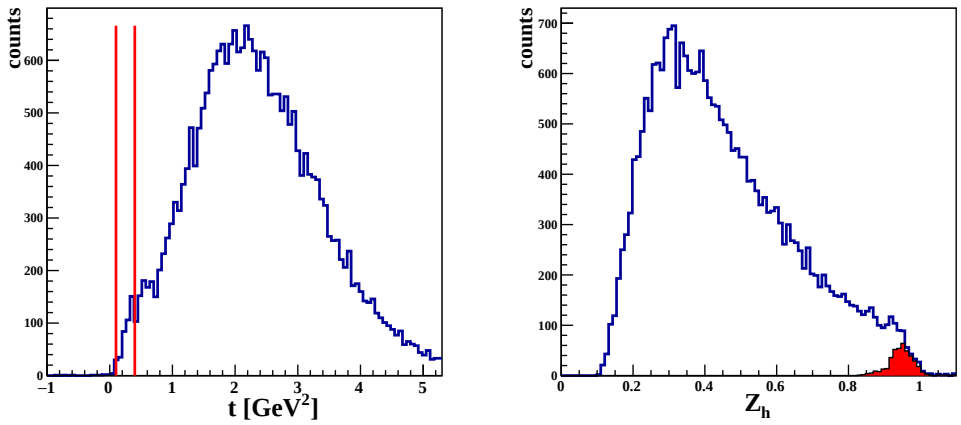


Figure 5.65: On the left is the t distribution, the red lines are the cuts for selecting the diffractive region. On the right is the one-dimensional distribution of the fractional energy of the hadrons, the area in red is the one that correspond to hadrons coming from a diffractive process.

5.6.0.1 Zh Dependence

In figure 5.66 are the MR in (Q^2,ν,Z_h) set plotted as a function of Zh integrated over Pt^2, for carbon iron and lead. For visual purposes the color scheme was chosen to be reddish for carbon, bluish for iron and blackish for lead. Also the gradient color darkens as the value of ν increases. Acceptance correction is the only correction implemented. In figure 5.67 there is a similar plot but the MR is in (Q^2,Xb,Z_h) set instead.

Many features emerge from this plots, first we observed that the MR decreases with Zh (more attenuation), for all targets. This is in agreement with other CLAS studies [61] [25] [81] and with HERMES results [32] [33]. Within the many models existing to date, in the absorption type models the decrease of the MR for increasing Zh is assumed to be due to a decrease in the production length in combination with pre-hadron absorption.

According to figures 5.66 and 5.67 we observed that the MR has a weak dependence on the electron kinematics (within the uncertainty of the measurement), i.e. we observed very similar results for different bins in $Q2$ and Xb. According to previous studies [61] [25] [59] the larger dependence is for ν, mainly the attenuation increases for lower values of ν. Therefore ν seems to be the only inclusive variable with a nuclear effect dependence. At our energies the range in ν goes basically between 2 and 4 GeV which is a tiny increase, so it makes our observation not conclusive. The explanation of this observation, according to most models, relies on a Lorentz boost contraction of the target's size (for high ν), this leads to a prehadron production time outside the nucleus i.e. less interaction with the nuclear medium, so, less attenuation (darker colors at the top and lighter colors in the bottom of figure 5.66).

The region $0.3 \lesssim Zh \lesssim 0.7$ is more stable (less fragmentation contribution and factorization requirement fulfilled) implying that the ratio of fragmentation functions is constant.

Another interesting feature of our results is the enhancement at low Zh, this is due to the redistribution of the events, shifting events at high Zh to low Zh. If the hadron (or pre-hadron) is formed inside the nucleus, it can experience hadronic interactions, so-called **F**inal **S**tate **I**nteraction (FSI) and the effect of those interactions is a loss of hadrons at a given value of Zh and an increase of the same hadron (possible also other hadrons) at lower Zh. For example a FSI may be the reescattering of the hadron from the nucleons in the nucleus, the hadron losing energy and one possibility is to generate another, mainly low energy meson. The hadron also could be absorbed, usually accompanied by the emission of other, again, mainly low energy hadron. Thus,

bin migration from higher Zh bins due primarily to hadronic cascades in the nucleus *pile up* in the low Zh region. This enhancement is reduced, as expected, with the application of $X_f > 0$ to select CFR.

The choice of MR in different ν or Xb bins is because they are the most interesting dependencies. ν carries information about the struck quark's initial energy and therefore regulates time-dilation, while Xb tells more about the initial state quark momentum before the scattering. Another aspect that is not studied directly in this thesis is Fermi motion, see [25] for a detailed discussion about it. This features are in agreement with other results from the group.

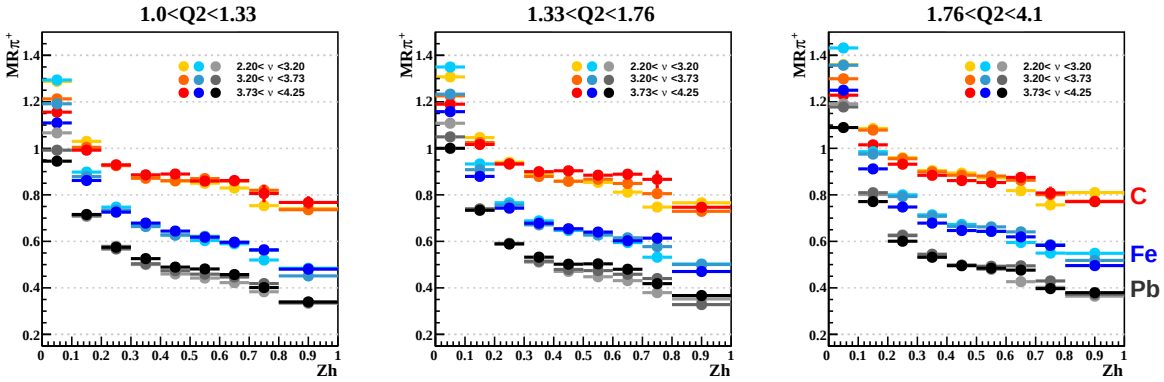


Figure 5.66: MR dependence on Zh for π^+ in different (ν, Q^2) bins integrated over P_t^2. Acceptance correction is applied. Only statistical errors are shown.

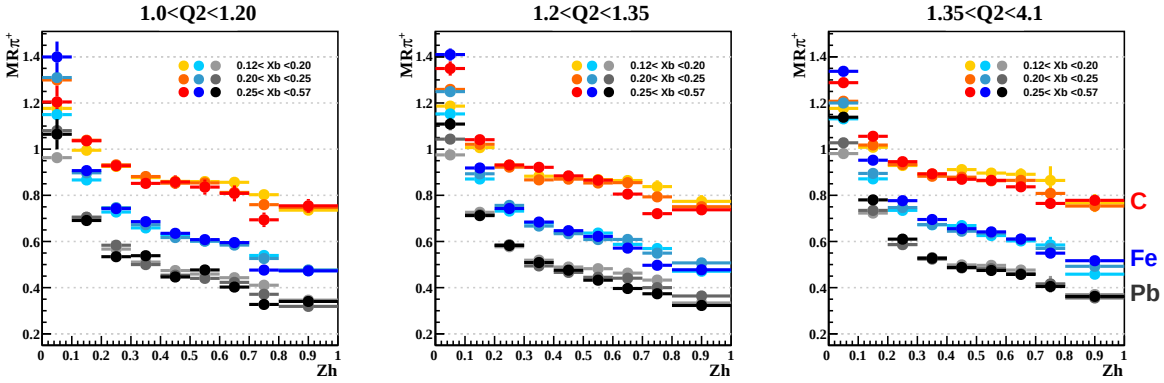


Figure 5.67: MR dependence on Zh for π^+ in different (Xb, Q^2) bins integrated over P_t^2. Acceptance correction is applied. Only statistical errors are shown.

5.6.0.2 P_t^2 Dependence

In order to study the P_t^2 dependency, an extra cut is applied. This is the so-called P_t^2 *cut-off cut*. The tails of the P_t^2 distributions are likely to be the product of mis-

recontruction due to, for instance, the poor resolution of this variable. For the MR observable is not too important[12], but it is applied here to reduce background. The idea of the cut is to remove the tail of the distribution, applying a cut based on a fit using an exponential distribution. It is worth noting that the effects are basically non negligible for high values of ν and Zh. For a more detailed study of this cut, see [82] (note that the fitting function is Gaussian instead of exponential in that case).

In figure 5.68, there are the MR curves in (Zh,ν,Pt^2) integrated over Q^2. In figure 5.69 are the MR curves in (Q^2,Xb,Pt^2) integrated over Zh. The y axis is in logarithmic scale. Acceptance correction is applied[13].

It is clear that the MR as a function of P_t^2 is characterized by a large increase at high values. This phenomenon is known from heavy-ion collisions as the *Cronin effect*[14], i.e., nuclear enhancement of high-p_T hadrons due to multiple interactions in nuclear matter [85]. Bound nucleons somehow cooperate producing high-p_T particles. The P_t^2 distribution for the π^+ -due to the multiple scattering of the propagating quark and hadron- is expected to be broader on a nuclear target than in deuterium. This increase is due to the small cross section at high P_t^2.

Looking at figure 5.68, we can notice that the raise of the MR is consistently the largest for low Zh (for all targets) which leads us to target fragmentation effects responsible for this enhancement. Another feature is that the raise at high P_t^2 is more modest for higher values of ν while for values of $P_t^2 \lesssim 0.7$ the Cronin effect is largely independent of ν [15]. Also, the effect vanishes -in qualitative agreement with [32] and [59]- for the largest Zh bin that is not the last one ($0.8 < Zh < 1.0$). This presents large enhancement at large P_t^2 and, also, a bump at small values of P_t^2. However, this region is highly contaminated, mainly with pions coming from ρ^0 decay. Thus, the interpretation is not clear. The fact that the largest Zh bin starts at the bottom for low P_t^2 (lowest MR) and then presents almost the biggest enhancement at higher values -breaking the rule of decreasing enhancement at high P_t^2 for higher Zh- was also observed for π^0 [60] [59], but much stronger. Also, for π^0 the enhancement observed is more modest than for π^+. This particular aspect of the highest Zh bin could be related to the hadron P_t broadening (not quark P_t broadening), since we are

[12] is particularly important for the transverse momentum broadening observable ($\Delta P_t^2 = P_t^2(A) - P_t^2(D2)$).

[13] in the case of a MR as a function of Pt^2, the number of inclusive DIS electrons remains as a normalization factor equal for all bins because Pt^2 is not an electron variable.

[14] named after James Cronin.

[15] When the MR is weakly dependent on a variable, this allow us to integrate over that variable or variables without introducing any kind of spurious correlation.

at high Zh, the hadron forms very fast and the color lifetime goes to zero. Further studies are necessary.

In figure 5.69, we can see that the attenuation is only noticeable for large P_t^2. The enhancement increases for higher values of Xb. This is consistently true for the three targets. This is in agreement with [61].

In figure 5.70, there are MR in different (Q^2, Xb, P_T^2, Zh) kinematical bins (four dimensional MR) with acceptance correction included. The three targets can be distinguished by the color scheme. In each block, there are 3 plots: at the top, there is carbon; in the middle, there is the iron; and at the bottom, there is lead. The colors becomes darker as we increase the value of P_T^2. From this plot, we can clearly notice that for high P_T^2 the value of the MR raises way beyond unity for low Zh, but comes closer to unity for high Zh. This fact is in qualitative agreement with [32]. The fact that the Cronin effect disappears at high Zh (excluding the last bin) is consistent with the idea that the raise of MR at high P_t^2 is of partonic origin. Notice that largest Zh means that $E_{\pi^+} \sim \nu$, so the parton is not allowed to have any energy loss, i.e., there is no room for partonic reescattering.

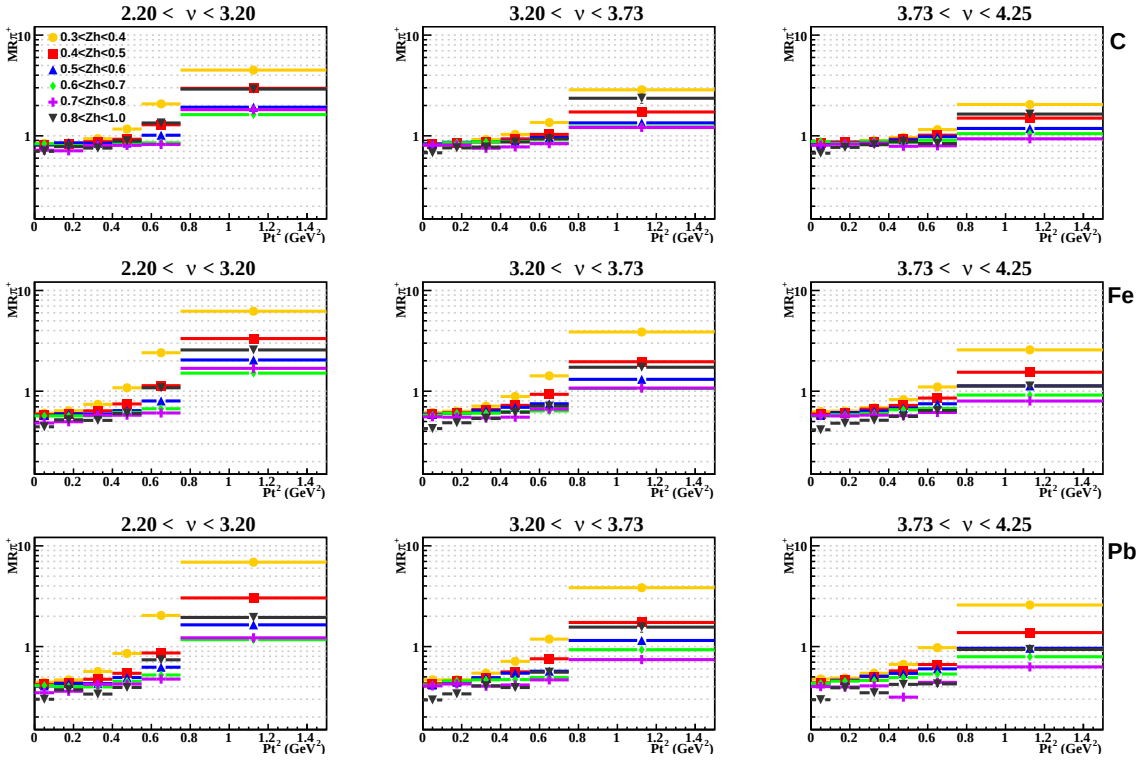


Figure 5.68: MR curves, in (ν, Z_h, P_t^2) integrated over Q^2. Acceptance correction is applied. Carbon is at the top in the middle iron and at the bottom lead. Only statistical errors are shown.

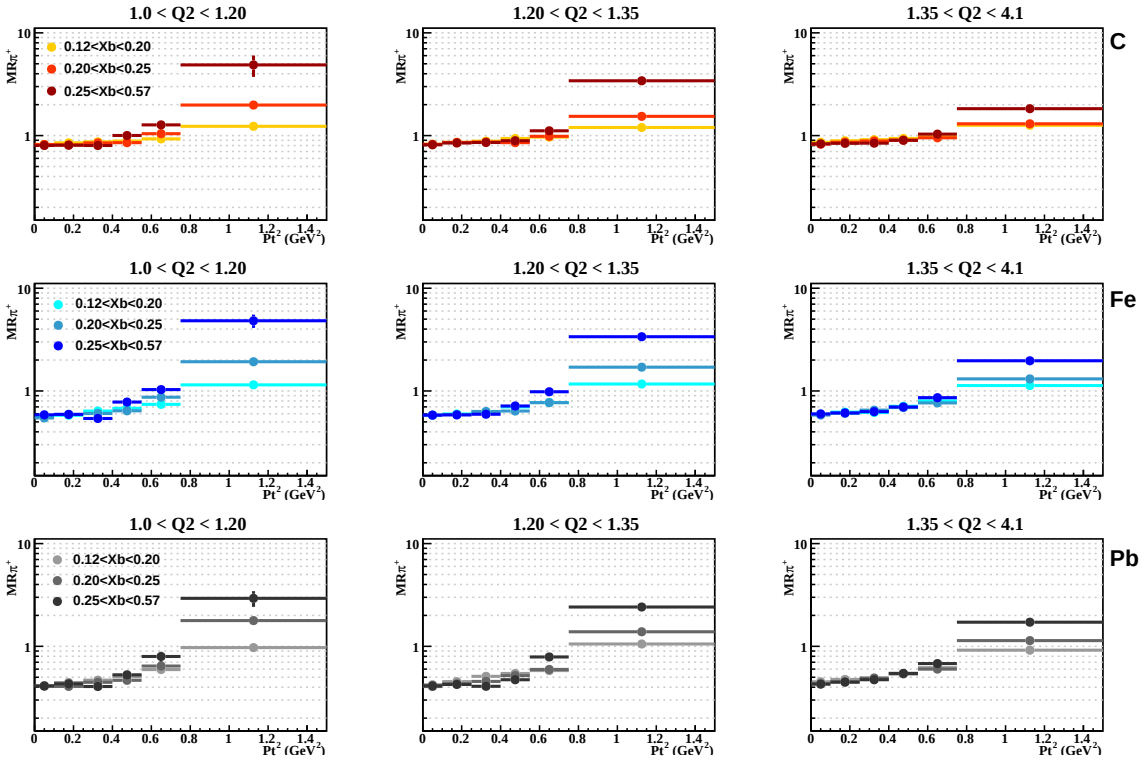


Figure 5.69: MR curves in (Q^2, X_b, P_t^2) integrated over Z_h. Acceptance correction is applied. Carbon is at the top; in the middle, there is iron; and at the bottom, there is lead. Only statistical errors are shown.

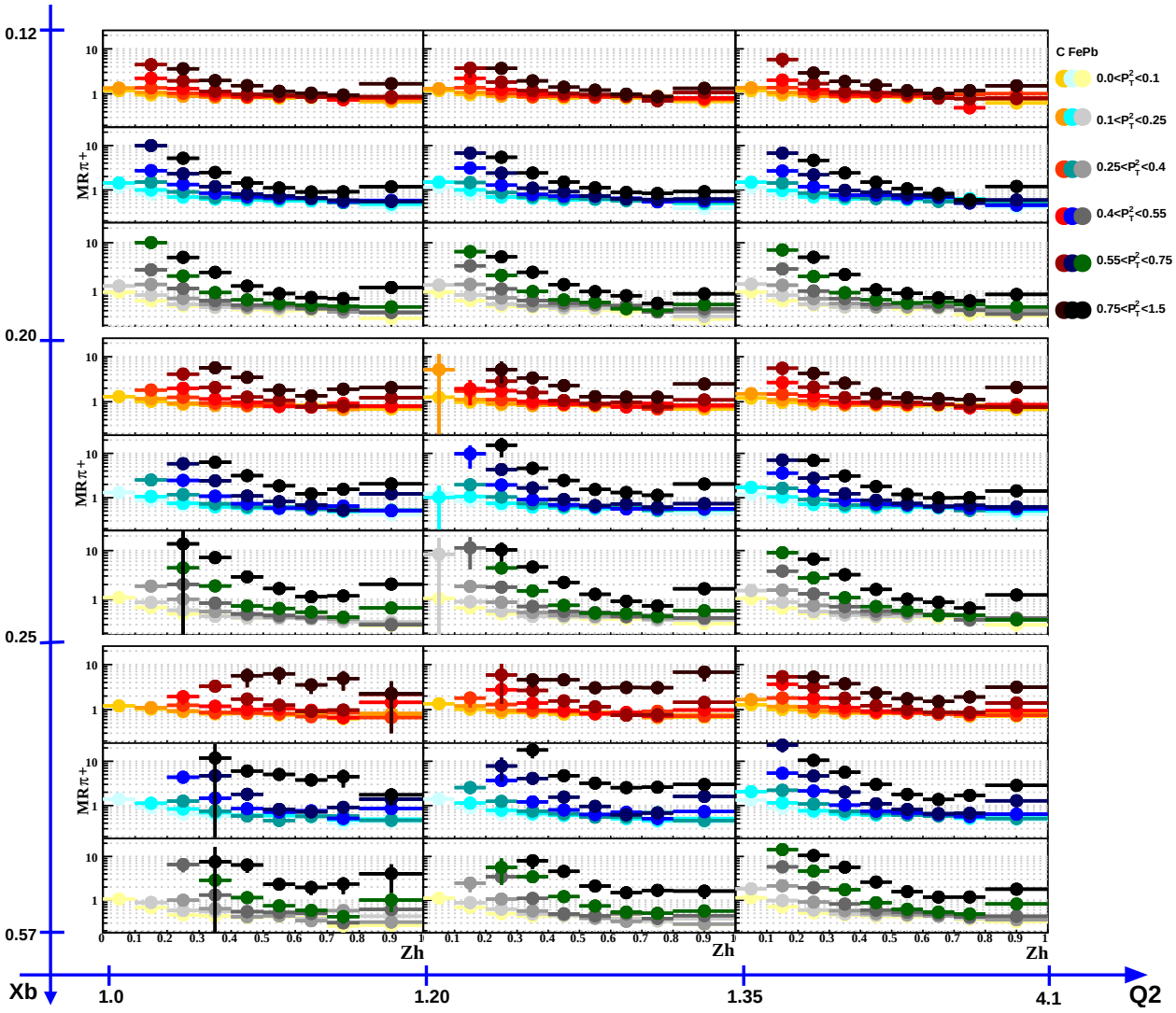


Figure 5.70: MR curves in $(Q^2, X_b, P_t^2,\ Zh)$ different kinematical bins for carbon, iron, and lead. Aceptance correction is applied. Only statistical errors are shown.

5.6.0.3 Q^2 Dependence

According to HERMES, there is a hint to claim that the MR increases with Q^2. For our data, the Q^2 dependence is depicted in figure 5.71 for different (ν, Zh) kinematical bins; acceptance correction is applied (integrated over P_t^2). No significant dependence on Q^2 is observed, mainly for the stable region of Zh. Note the different behavior for the last two bins in Zh. For our data, the MR dependence on Q^2 is weak and can be considered constant within the uncertainties of the measurement. This fact allows to integrate over Q^2 without worries of including some bias (false dependencies) in the distribution of our MR. This is in agreement with the studies of other members of

the group cited in this work .

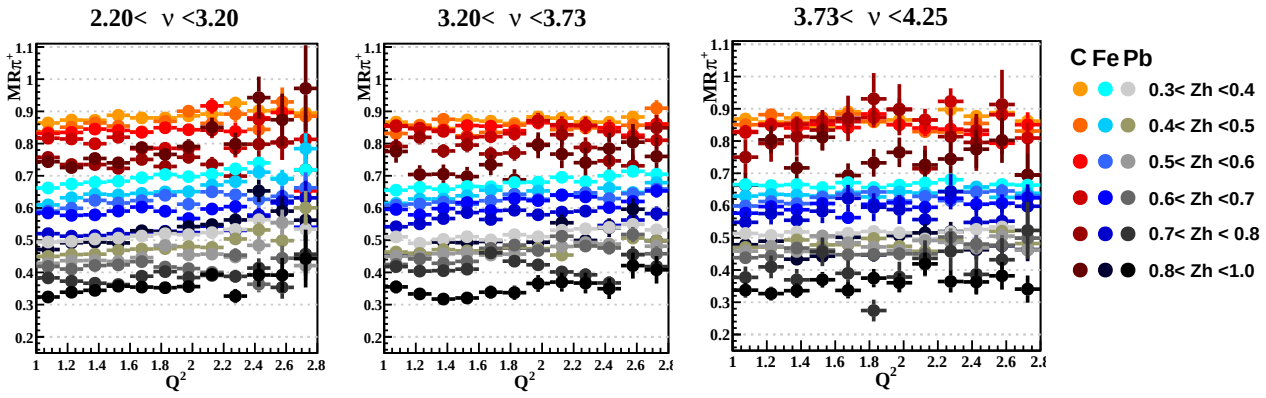


Figure 5.71: MR dependence on Q^2 for different (ν, Zh) kinematical bins, integrated over P_t^2. Acceptance correction is applied. Only statistical errors are shown.

5.6.0.4 ν Dependence

In figure 5.72, there is the MR as a function of ν in different $(Q2, Zh)$ kinematical bins. This time the results are not as flat as for the case of the Q^2 dependency of MR. A slight increase in the MR is expected with increasing values of ν. This is observed just for the stable region of Zh and is very weak, being the lead the one that shows a larger signal of that increment. The lower values of Zh show a flat behavior and, for the two highest bins in Zh the shape of the distribution, is not clear to interpret properly. The trend of larger nuclear suppression at low energies and vanishing suppression with increasing energy is well reproduced by the color dipole model.

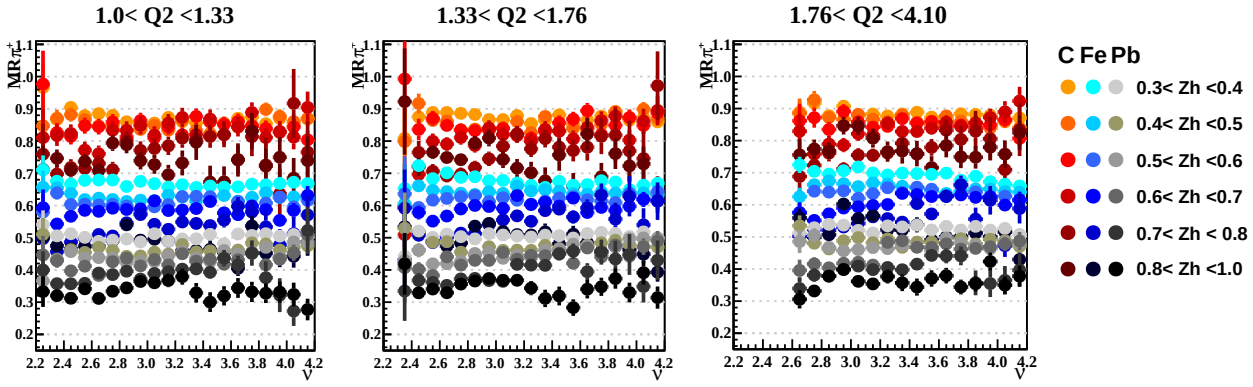


Figure 5.72: MR as a function of ν in different $(Q2, Zh)$ kinematical bins. Acceptance correction is applied. Only statistical errors are shown.

5.6.0.5 Xb Dependence

In figure 5.73, there is the MR as a function of Xb in different $(Q2, Zh)$ kinematical bins. The phase space of Xb v/s Q^2 leads to many empty bins in the distribution. The flat behavior is similar to the Q^2 dependence. Looking at the most stable region of Zh the MR is, within the uncertainty of the measurement, independent of Xb. This is in agreement with the fact that the only inclusive variable that seems to have a dependency on MR is ν.

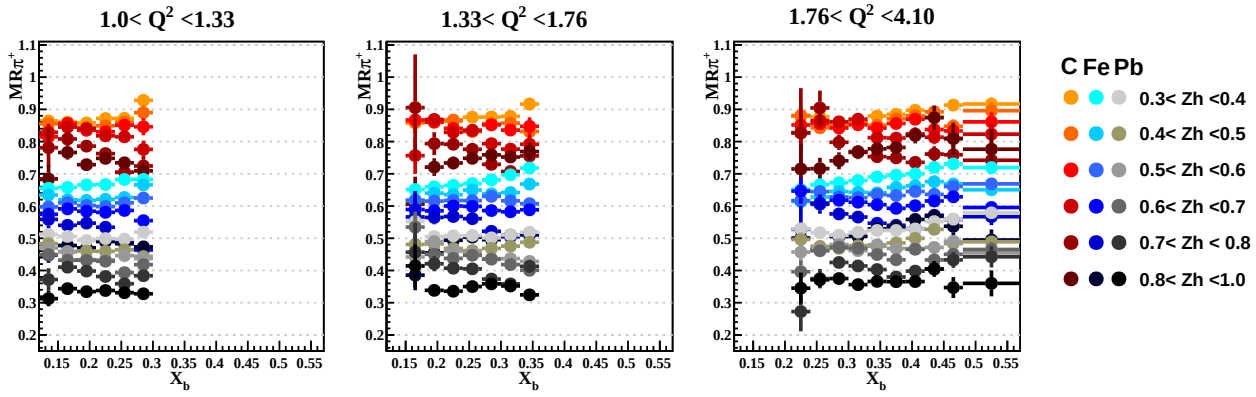


Figure 5.73: MR as a function of Xb in different $(Q2, Zh)$ kinematical bins. Acceptance correction is applied. Only statistical errors are shown.

Chapter 6

Conclusions

This thesis consisted of the analysis of positive pion electroproduction data from Hall B in Jefferson Lab for studies of hadronization. For that purpose, the hadronic multiplicity ratio was measured in four different targets of varying sizes. The use of various nuclear targets allows to study the space-time development of the hadronization process in nuclear medium. The ratio of inclusive electron yields was also studied (EMC ratio).

New particle identification cuts for electrons and pions were implemented to improved DIS and SIDIS samples. Several problems were addressed, such as non-optimal particle selection for reconstructed MC particles, problems regarding the simulation set, too large z-vertex dependence in the observables, unknown origin of the central peak in the azimuthal angle relative to virtual photon axis (ϕ_{π^+}), unknown behavior of the observables for different CLAS sectors and unrevised implementation of Coulomb and radiative corrections.

We reported results for the EMC ratio. Several corrections were applied, such as background subtraction, acceptance correction, radiative correction, and Coulomb corrections. A comparison with world-data, encoded in an implemented parametrization function, was performed as well, with which our results have, within uncertainties, a good agreement

We explored the effect of different cuts on the multiplicity ratio as a function of the fraction of the virtual photon energy transferred to the observed pion. Each cut or correction was presented, discussed, and applied separately. Different variations from the nominal cuts were studied and included in the systematic uncertainty section. Our results present systematic uncertainty that ranges between 3% and 7% for the EMC ratio and between 4% and 8% for the multiplicity, the numbers are similar for all target-types.

We reported multidimensional measurements of the multiplicity ratio for π^+ as functions of the virtuality of the exchanged boson, the fraction of the virtual photon energy transferred to the observed pion, the transverse momentum of the hadron, the fraction of the proton momentum carried by the struck quark and the energy lost by the incoming particle. Our results were compared with studies from other collaborations and with studies within our workgroup. We achieved 4-fold differentiated multidimensionality. Many features were in agreement with other results, such as enhancement and suppression trends, and some conflicts, such as clear dependencies on some variables or some quantitative aspects.

A long historical discrepancy between two independent analysis was studied and resolved, cleaning the path to further publications.

A future CLAS Analysis Note will be published about this thesis's subject with greater detail and also with the consideration of both charged pions.

The results presented in this thesis have been reported at conferences at JLab from 2018 to 2020.

This result also represents a benchmark for electron and pion identification with the EG2 dual-target design, which informs future measurements of multi-hadron final states with CLAS. Future measurements during the CEBAF 12 GeV era with the CLAS12 detector will provide enhanced luminosity, which will be critical to measure the multiplicity ratio of heavier mesons and baryons, as well as in larger kinematic ranges.

The combination of the present result with CLAS12, and the future Electron-Ion Collider, will provide a rather large lever arm in kinematic variables that will help reveal the origin of the suppression of hadron in nuclei, as well as to explore the interplay between the hadronic and partonic degrees of freedom.

Bibliography

[1] Francis Halzen, Alan D. Martin , *Quarks and Leptons: An Introductory Course in Modern Particle Physics.*

[2] Mark Thompson. *Modern Particle Physics.* University of Cambridge.

[3] Robin Devenish, Amanda Cooper-Sarkar, *Deep Inelastic Scattering.* OXFORD.

[4] Alberto Accardi et.al. , *Parton Propagation and Fragmentation in QCD Matter.* arXiv:0907.3534v1 [nucl-th] 21 jul 2009.

[5] Cheuk-Yin Wong. *Introduction to High-Energy Heavy Ion Collisions.* Oak Ridge national Laboratory, USA. 1994.

[6] Yu. L. Dokshitzer, V. A. Khoze, A. H. Mueller and S.I Troyan, *Basic of Perturbative QCD.* 1991.

[7] M. Anselmino et.al,*Understanding the role of Cahn and Sivers effects in Deep Inelastic Scattering.* arXiv:hep-ph/0507157v1 13 Jul 2005.

[8] M. Anselmino et.al, *The role of Cahn and Sivers effects in Deep Inelastic Scattering.* arXiv:hep-ph/0501196v2 23 May 2005.

[9] M. Anselmino et.al, *Comments on Cahn and Sivers effects in SIDIS.* arXiv:hep-ph/0412316v1 21 Dec 2004.

[10] Néstor Armesto, *Nuclear shadowing.* Departamento de Física de Partículas and IGFAE, Universidade de Santiago de Compostela, 15782 Santiago de Compostela, Spain. arXiv:hep-ph/0604108v2, 5 Jul 2006.

[11] Or Hen et.al, *The EMC Effect and High Momentum Nucleons in Nuclei.*

[12] William R. Leo. *Techniques for Nuclear and Particle Physics Experiments.*

[13] John C. Collins *Factorization of Hard Processes in QCD.* Penn State University,University Park PA 16802, U.S.A. arXiv:hep-ph/0409313.

[14] L.S. Osboorne. *Electroproduction of Hadrons from Nuclei.* Phys. Rev. Lett. 40, 1624 - Published 19 June 1978.

[15] Kevin McFarland, *Neutrino Neutrino-Nucleon Nucleon Deep Inelastic Scattering Deep Inelastic Scattering.* Lecture 6.

[16] M. Duer, O Hen, et. al., *Probing high-momentum protons and neutrons in neutron-rich nuclei.* CLAS Collaboration, 2018.

[17] M.D. Mestayer et.al., *The CLAS drift chamber system.*1999.

[18] G. Adams et.al., *The CLAS Cherenkov detector.*2001.

[19] S. Joosten et. al., *Enhanced UV light detection using a p-terphenyl wavelength shifter.* arXiv:1611.03467v4 , 2017.

[20] E.S Smith et. al., *The time-of-flight system for CLAS.* 1999.

[21] M. Amarian et. al., *The CLAS forward electromagnetic calorimeter.* 2000.

[22] Will Brooks et al. , *Quark Propagation Through Cold QCD Matter.* A proposal to Jefferson Lab PAC 22.

[23] Lorenzo Zana, *Search for the onset of color transparency through ρ^0 electroproduction on nuclei.* B.S. Physics, University of Turin (2002).

[24] A.B. Larionov, M. Strikman. *Slow neutron production as a probe of hadron formation inhigh-energy $\gamma^* A$ reactions.* https://arxiv.org/pdf/1812.08231.pdf

[25] Raphael Dupré, *Quark Fragmentation and Hadron Formation in Nuclear Matter*, THESE DE DOCTORAT. Université de Lyon, Faculté des Sciences et Technologies. École Doctorale PHAST. 2011.

[26] Hovanes Egiyan et.al.,*Single π^+ Electroproduction on the Proton in the First and Second Resonance Regions at $0.25GeV2 < Q2 < 0.65GeV2$ Using CLAS.* arXiv:nucl-ex/0601007v2 5 Jan 2006.

[27] A. Asaturyan et.al, *Transverse Momentum Dependence of Semi-Inclusive Pion and Kaon Production.* Update of Experiment E12-09-017. July 5,2011.

[28] Alessandro Bacchetta et.al. , *Semi-inclusive deep inelastic scattering at small transverse momentum.* arXiv:hep-ph/0611265v2 26 Feb 2007.

[29] H. Avakian et.al. *Measurement of Beam-Spin Asymmetries for π^+ Electroproduction Above the Baryon Resonance Region* CLAS Collaboration. arXiv:hep-ex/0301005v2 27 Apr 2004

[30] H.Hakobyan, W. Brooks et al. , *A double-target system for precision measurements of nuclear medium effects* . 2008

[31] A. Airapetian et.al, *Multiplicities of charged pions and kaons from semi-inclusive deep-inelastic scattering by the proton and the deuteron.* The HERMES Collaboration, arXiv:1212.5407v2 [hep-ex] 24 Apr 2013.

[32] A. Airapetian et.al, *Hadronization in semi-inclusive deep-inelastic scattering on nuclei.* The HERMES Collaboration. arXiv:0704.3270v1

[33] A. Airapetian et.al., *Multidimensional Study of hadronization nuclei.* The HERMES Collaboration. arXiv:1107.3496v3

[34] Henry W. Kendall *Deep Inelastic Scattering experiments on the proton and the observatgion of scaling.* Nobel Lecture, December 8, 1990. Massachusetts Institute of Technology, Cambridge, Massachusetts, USA.

[35] Nathan Andrew Harrison, *Exploring the Structure of the Proton via Semi-inclusive Pion Electroproduction.* University of Connecticut. Ph.D thesis.2010.

[36] Narbe Kalantarians, *Semi-Inclusive Single Pion Electroproduction From Deuteron and Proton Targets.* PhD thesis, 2008. University of Houston.

[37] Dave Gaskell - JLab. *Experimental Overview of Past and Future Studies of the EMC Effect.* INT Workshop on Nuclear Structure and Dynamics at Short Distance. February 14, 2013.

[38] L.W.Mo, Y.S.Tsai. *Radiative corrections to elastic and inelastic ep and μp scattering.* Reviews of Modern Physics Volume 41, 1 (1969).

[39] Y.S.Tsai, *Radiative corrections to electron scatterings.* SLAC-PUB-848 (1971).

[40] P.E. Bosted, V. Mamyan. *Empirical Fit to electron-nucleus scattering.* arXiv:1203.2262v2 [nucl-th] 19 Mar 2012.

[41] Andreas Aste, Cyrill von Arx, Dirk Trautmann, *Coulomb distortion of relativistic electrons in the nuclear electrostatic field.* Department for Physics and Astronomy, University of Basel, 4056 Basel, Switzerland. arXiv:nucl-th/0502074v3.

[42] Thomas Carroll, *Coulomb Corrections for Inclusive Electron Scattering Data.* Masters Theses. May 2019

[43] Andreas Aste. *Coulomb distortion effects in quasi-elastic (e,e') scattering on heavy nuclei.* Department of Physics, University of Basel, Klingelbergstrasse 82, 4056 Basel,Switzerland. arXiv:0710.1261v3 [nucl-th] 7 Mar 2008.

[44] I.Akushevich, N.Shumeiko, and A.Soroko.*Radiative effects in the processes of hadron electroproduction.* National Center of Particle and High Energy Physics, 220040 Minsk, Belarus. arXiv:hep-ph/9903325v1 11 Mar 1999.

[45] P. Solvignon et.al, *Coulomb distortion in the inelastic regime.* arXiv:09006.0512v1 [nucl-ex] 2 Jun 2009.

[46] J. Seely et.al, *New measurements of the EMC effect in very light nuclei.* arXiv:0904.4448v3 [nucl-ex] 27 Oct 2009.

[47] H.Hakobyan,S.Mancilla, *Migration of HAPRAD2 to C++.* $https :: //github.com/usm-data-analysis/HAPRAD$

[48] Schmookler, B. and others. *Modified structure of protons and neutrons in correlated pairs.* CLAS Collaboration, 2019

[49] Schaub, T. *CFD.* CLAS Note, 2016

[50] Schaub, Thomas. *Determinacion por medio de simulacion numerica del perfil de densidades en hidrogeno en nestado criogenico con haz de electrones como fuente de calor.* Universidad Tecnica Federico Santa Maria, 2015

[51] K. Gallmeister, T. Falter. *Space-Time Picture of Fragmentation in PYTHIA/JETSET for HERMES and RHIC.* arXiv:nucl-th/0502015v2 6 Sep 2005.

[52] M. Boglione, J. Collins, L. Gamberg, J. O. Gonzalez-Hernandez, T. C. Rogers,N. Sato *Kinematics of Current Region Fragmentation in Semi-Inclusive Deeply Inelastic Scattering* arXiv:1611.10329

[53] J.C Vander Velde *Properties of hadrons pproduced by neutrinos in hydrogen* University of Michigan, Ann Arbor, Michigan 48109, U.S.A. *Probing hadrons with leptons.Ettore Majorana International Science Series (pag 385).*

[54] Thomas Ferbel, *Techniques and Concepts of High-Energy Physics.* University of Rochester, Rochester, New York.

[55] Michail P. Rekalo and Egle Tomasi-Gustafsson.*Polarization effects in elastic electron-proton scattering (Lecture notes)* DAPNIA/SPhN, CEA/Saclay, 91191 Gif-sur-Yvette Cedex, France, October29,2018. arXiv:nucl-th/0202025v1.

[56] M. Anselmino, V. Barone, A. Kotzinian *SIDIS in the target fragmentation region: polarized and transverse momentum dependent fracture functions.*

[57] Akinori 0HSAWA. *Violation of the Feynman Scaling Law in the Forward Region* Institute for Cosmic Ray Research, University of Tokyo, Tanashi 188, May 12, 1994.

[58] V. Blobel. BOS: Bank Organization System. Dynamic Storage Organization with FORTRAN. 1977.

[59] Taisiya Mineeva. *Hadronization Studies via π^0 Electroproduction off D, C, Fe and Pb.* PhD thesis.

[60] Taisiya Mineeva. *Neutral Pion Multiplicity Ratios from SIDIS Lepton-nuclear Scattering.* CLAS Analyis Note, 2020.

[61] Hayk Hakobyan. *Observation of Quark Propagation Pattern in Nuclear Medium* PhD thesis. Yerevan State University, 2008.

[62] Sargsian,M. *Computer Code for Inclusive (e,e) Electro-production Reactions and Radiative Corrections.* 1990 CLAS Note. *https : //www.jlab.org/Hall − B/notes/clas$_n$otes90/note90 − 007.pdf*

[63] Maurice Lévy . *Weak and Electromagnetic Interactions at High Energies , Cargése 1975, Part B.*

[64] *Semi-Inclusive Λ electroproduction in the Target Fragmentation Region, A 12 GeV Proposal to Jefferson Lab PAC 42.*

[65] T. Mart. *Can we extract the pion electromagnetic form factors froma a t-channel diagram only?* arXiv:0805.1800v1 [hep-ph] 13 May 2008.

[66] Raghunath Sahoo, *Relativistic Kinematics.* Indian Institute of Technology Indore, India-452020. arXiv:1604.02651v1 [nucl-ex] 10 Apr 2016.

[67] K. Goulianos. *Universality of Particle Multiplicities.* FERMILAB-Conf-94/266-E

[68] K. Gallmeister, T. Falter. *Space-Time Picture of Fragmentation in PYTHIA/JETSET for HERMES and RHIC*, Institut fur Theoretische Physik, Universitat Giessen, Germany.

[69] E.Pedrazzi, *Diffraction: past, present and future.* Dipartimento di Fisica Teorica, Universit'a di Torino and INFN, Sezione di Torino, 10125 Torino, Italy. arXiv:hep-ph/9809454v1 18 Sep 1998.

[70] Stacy E. Karthas. *Re-analysis of the World Data on the EMC Effect and Extrapolation to Nuclear Matter.* May 15 , 2014.

[71] Stéphane Munier, *Diffractive patterns in deep-inelastic scattering and parton genealogy.* CPHT, École Polytechnique, CNRS, Université Paris-Saclay, Route de Saclay, 91128 Palaiseau, France. arXiv:1809.10697v1 [hep-ph] 27 Sep 2018.

[72] E.A. De Wolf. *Diffractive scattering.* CERN, European Laboratory for Particle Physics, 1211 Geneva 23, CH. arXiv:hep-ph/0203074v2 18 Mar 2002.

[73] P Amaudruz et al., The new Muon Collaboration (NMC). *A Re-Evaluation of the nuclear Structure Function Ratios for D, He, 6Li, C and Ca.* arXiv:hep-ph/9503291v1 10 Mar 1995.

[74] Simona Malace, David gaskell, Douglas W. Higinbotham, Ian Cloet *The Challenge of the EMC Effect: existing data and future directions.* arXiv:1405.1270v2 [nucl-ex] 25 Jun, 2014.

[75] M. Osipenko et. al., *Measurement of semi-inclusive π^+ electroproduction off the proton.* CLAS Collaboration.

[76] M. Osipenko, A. Vlassov, M Taiuti. *Matching between the eelctron candidate track and the Cherenkov counter hit.* June 18, 2004

[77] M. Osipenko, A. Vlassov. *Calibration of the Cherenkov counter TDCs.* May 1, 2006.

[78] G. Fedotov et. al., *Analysis report on the* $ep \to e'p'\pi^+\pi^-$ *reaction in the CLAS detector with a 2.039 GeV beam for* $0.4GeV^2 < Q^2 < 1.0GeV^2$ *and* $1.3GeV < W < 1.825GeV$. https://misportal.jlab.org/ul/Physics/Hall-B/clas/viewFile.cfm/2018-001.pdf?documentId=776

[79] P.E. Bosted et. al., *Target and Beam-Target Spin Asymmetries in Exclusive* π^+ *and* π^- *Electroproduction with 1.6 to 5.7 GeV Electrons.* The CLAS Collaboration. March 12, 2018. arXiv:1604.04350v3.

[80] Raphael Dupre, *Study of the hadronization of charged pions.* Analysis Note.

[81] Ahmed El Alaoui Hadronization analysis, EG2 data Webpage. http://atlasusr.fis.utfsm.cl/alaoui/dist3d.php

[82] webapge, $https://clasweb.jlab.org/wiki/index.php/PT2_CutOff$

[83] Cronin, J.W. and Frisch, Henry J. and Shochet, M.J. and Boymond, J.P. and Mermod, R. and Piroue, P.A. and Sumner, Richard L. *Production of hadrons with large transverse momentum at 200, 300, and 400 GeV*

[84] Falter, T. and Cassing, W. and Gallmeister, K. and Mosel, U. *Hadron attenuation in deep inelastic lepton-nucleus scattering.*

[85] B.Z. Kopeliovich et.al., *Cronin Effect in Hadron Production off Nuclei.* arXiv:hep-ph/0201010v2

[**?**]

Appendix A

Analyses Comparison

A.1 Introduction

The goal of this chapter is to address incompatibility between two independent analyses. The situation is the following: in the calculation of the hadronic Multiplicity Ratio for positive pions, two independent analyses were performed, one by Raphael Dupré (RD) and one by Hayk Hakobyan (HH). The results present important discrepancies, reaching 20%. A proposal for a forward publication of these results can only be achieved if this discrepancy is explained.

Since the two original analyses are old (2008 for HH and 2011 for RD), if every single detail of the procedure to obtain the final results are not properly documented, it is difficult to keep track of the necessary information to reproduce the original results. Throughout the years, some information got lost and some aspects of the analyses were modified. For these reasons, to work out the different nuances of the analysis of the discrepancy's source, a reproduction process took place. A flexible analysis framework that allows to use the requirements of either historical analysis or new requirements as needed has been developed and implemented by the author of this thesis.

Let's look at the problem: the Multiplicity Ratio, as a function of Zh for Iron, is in Figure A.1. The Labels are *RD* for the results of Raphael Dupré and *SM* for the results of Santa María University group, initially by Hayk Hakobyan[1]. The discrep-

[1]It will not be appropriate to label this result as *HH* because the result comes from an analysis that has been modified through the years by different members of the group in Santa Maria, this includes the author. They were small modifications; the most important part of the analysis can be found in detail in [61]. It is important to notice that the results labeled here as *SM* are not the same as the results presented in the previous chapters on this thesis. SM results are the state of the analysis back in 2017 with only a modification made by the author of this thesis, related to the identification of π^+. It consisted in make tighter cuts in ΔT in order to reduced kaon contamination from the sample, but such modification did not change the previous result in an important way.

ancy between the two cases at the uncorrected level is small. When the acceptance correction is applied, the discrepancy in the MR as Zh grows becomes larger. For RD the acceptance correction effect goes up to 20% in the last bin while for SM the acceptance effect, is only a few percent: it never goes beyond 4%.

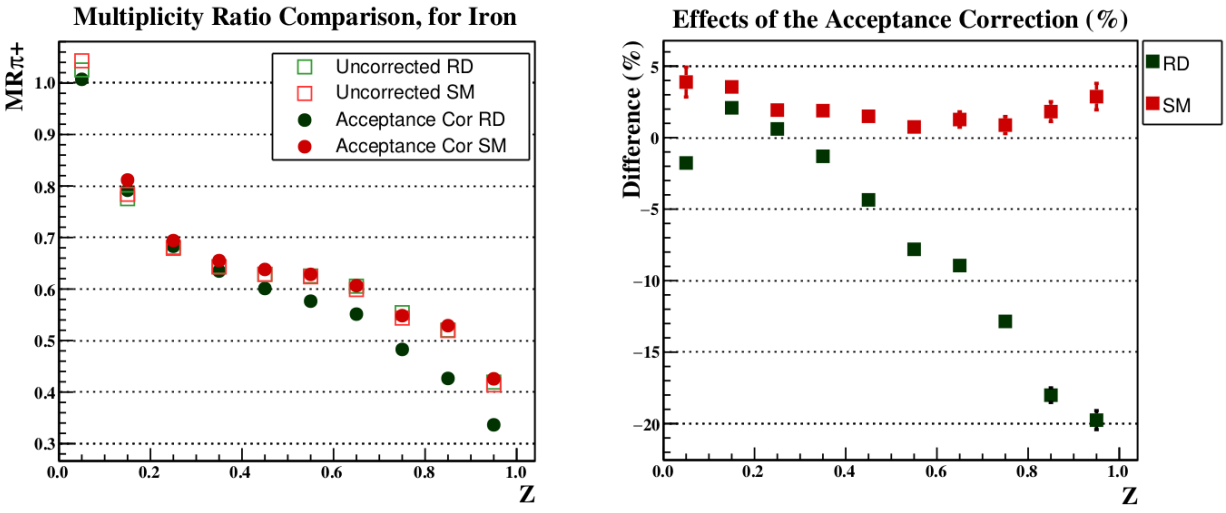


Figure A.1: On the left, there is the MR as a function of Zh for the two different analyses. The squares are the uncorrected case and the full circles are with acceptance correction. On the right hand side, there are the effects of the Acceptance Correction in %. $|\Delta Z| < 3[cm]$ and $Xf > 0$ are applied (CFR).

One important comment about the plot in figure A.1 is that the RD case are the results coming from RDs analysis. The only modification made on those numbers is to put the point in the center of the bin instead of the centroid -where they were originally- just to make comparisons easier to see, but everything else remains untouched.

The biggest discrepancy only arises when the acceptance correction is applied. At the uncorrected level, the two analyses agree, see figure A.2. Since we have two completely different particle identification cuts, we can mix them up and the final results are found to be consistent. In this case, the cuts $|\Delta Z| < 3$ and $X_f > 0$ are applied.

With the problem fully recognized, the next sections of the chapter are dedicated to explore the source of this discrepancy.

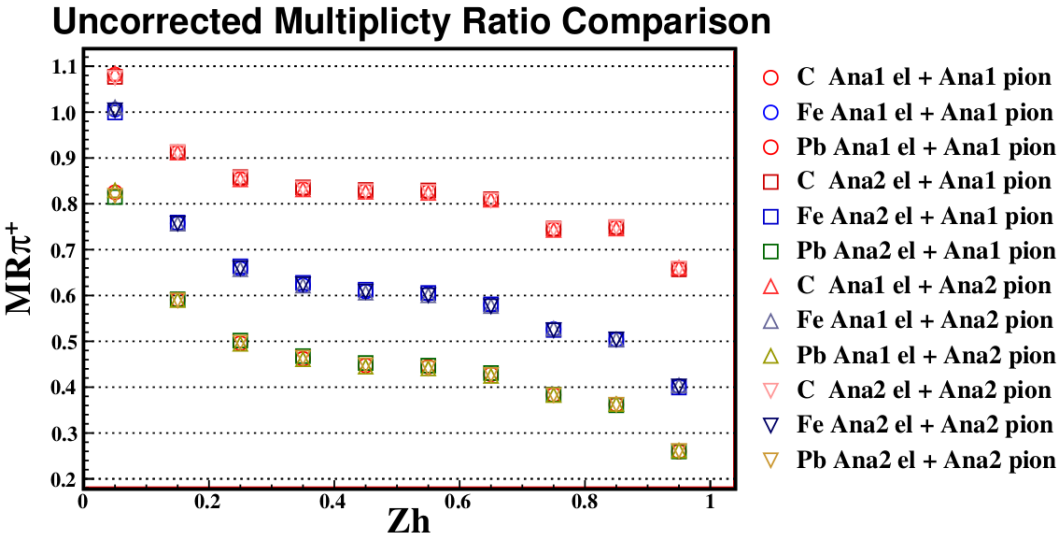


Figure A.2: MR comparison for all combinations of particle identification cuts, according to each analysis. *Ana1 el/pion* stand for the cuts of SM analysis for electron/pion and *Ana2 el/pion* stand for the cuts of RD analysis for electron/pion.

A.2 Exploring the Source of the Discrepancy

The first thing to do to solve this problem is to reproduce the results shown by RD with the code implemented by the author of this thesis. The full comparison between the original and the reproduced results is in figure A.3.

This reproduction process took into account the following features:

- Particles Identification Cuts for electrons and pions [2].
- Simulation Set, provided by RD.
- Isospin Correction included in the case of the iron target is 3% [3]. There is no isospin correction in SM case.
- Vertex Cuts for electrons. The vertex cuts choice here is narrower than SM's and it is target dependent. The comparison between the two cases is shown in figure A.4.

[2] following RD analysis note "Study of the hadronization of charged pions", from 2017 and personal comments during group meetings. Every detail of RD analysis not included in this thesis is fully explained in [25] [80].

[3] For Pb is 10% and there is no isospin correction for C.

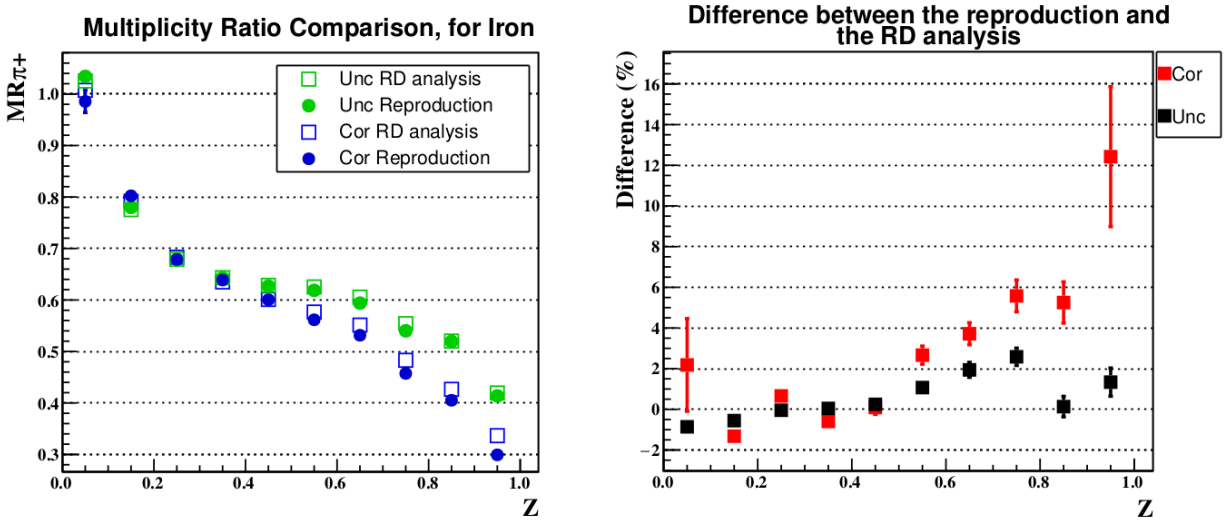


Figure A.3: Comparison between the original RD result and the reproduction of it in green. On the right, there is the difference in % between the two cases, for the uncorrected and corrected case.

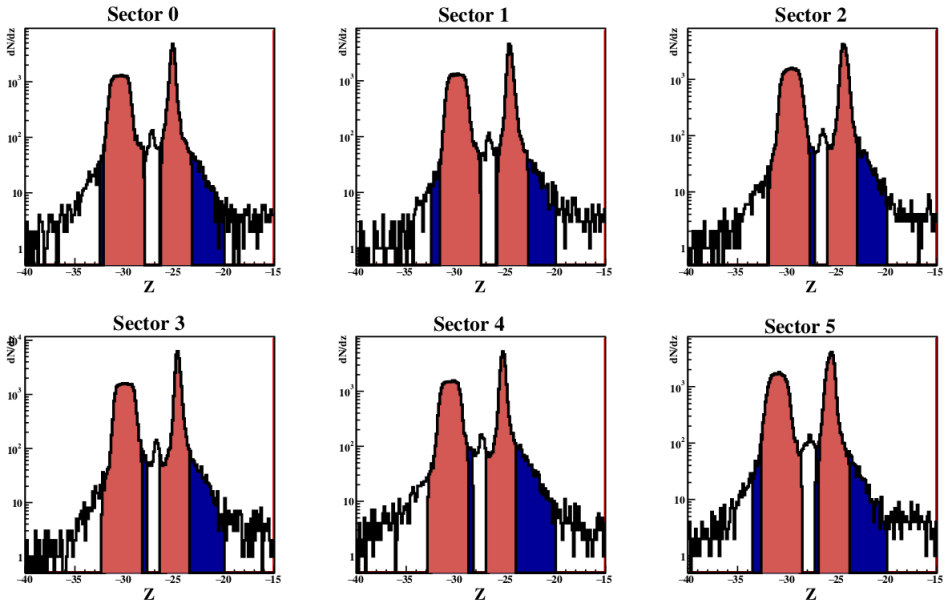


Figure A.4: Electron Z vertex cut comparison for each CLAS sector, for both analyses. In blue, there are the wider cuts belonging to SM vertex cut choice, and in light red are RD's.

- Additional cuts: $|\Delta Z| < 3[cm]$ and $Xf > 0$, with RD implementation of the Xf variable.

The reproduction is not perfect. The source of this imperfect reproduction relies on various things, such as the run numbers used in the analysis, the binning choice,

the variables involved in the integration for the acceptance, the way of implementing the acceptance correction (event by event instead of bin by bin), an additional cut for dealing with bin migration issues, the fact of demanding that the electron not only should be in the first place for the event to be considered a good event, among others. Each of these possibilities has been studied in the past. None of them make a difference of the magnitude's order we are interested in, however, as presumably, each different aspect of the analysis adds up some uncertainty at the end. Thus, a reproduction that differs from the original in 3 or 4% -except for the last bin- is good enough because it successfully reproduces the behavior we are looking for, as is depicted in figure A.3.

In this study, one aspect of the analysis was to apply the SM particle identification cuts in RD simulations. The results were unreasonable, hint that in the particle selection process involving the simulations we must pay special attention to the cuts to avoid sub-optimal results.

The first thing to do to work with simulations is to apply the cuts necessary to select the particle we are interested in. In this case, it is always the positive pion, and of the scattered electron. Let's start with the particle selection for the electron. The main cut for electron selection, according to RD analysis, is based on the Ein/Eout v/s P phase space and it follows the following expression:

$$0.156197 \leq \left(\frac{E_{in}}{P} + \frac{E_{out}}{P}\right) \leq 0.2713 \times \left(1 + \left(\frac{0.3}{\sqrt{b}}\right)\right) \tag{A.1}$$

the values of b vary for different ranges of P, according to table A.1. For more details about the procedure of this, look at RD analysis note [80]. This phase space is plotted in figure A.5, all electron candidates are in there and the red lines are the cuts for electron selection. On the left, there is the data case and on the right, there is the reconstruction. The dashed black line is at 0.27, just for reference. Looking at the reconstruction a downward shift is observed compared to the data case, by eye, roughly is a 17% shift.

If we compare these cuts to the ones explained in [61], we can found out that this is more relaxed with the electron candidates, roughly speaking in RD case, out of all the electron candidates, 50% of them are considered good electrons, meanwhile for SM case only 30% of them are called good electrons. In figure A.6, there are comparison of the distributions for each case.

For the positive pion selection, in RD analysis, the main cut for that is based on the $\Delta\beta$ variable, defined as:

Electron Identification Cuts	
P range	Parameter b
0.5 - 0.7	0.85
0.7 - 0.9	0.80
0.9 - 1.1	0.85
1.1 - 1.3	1.05
1.3 - 1.5	1.10
1.5 - 1.7	1.35
1.7 - 1.9	1.35
1.9 - 2.1	1.45
2.1 - 2.3	1.35
2.3 - 2.5	1.35
2.5 - 2.7	1.35
2.7 - 2.9	1.30
2.9 - 3.1	1.35
3.1 - 3.3	1.35
3.3 - 3.5	1.50
3.5 - 3.7	1.60
3.7 - 3.9	1.80
3.9 - 4.1	1.80
4.1 - 4.3	1.80
4.3 - 4.5	1.80

Table A.1: Values of the b parameter for different momentum ranges

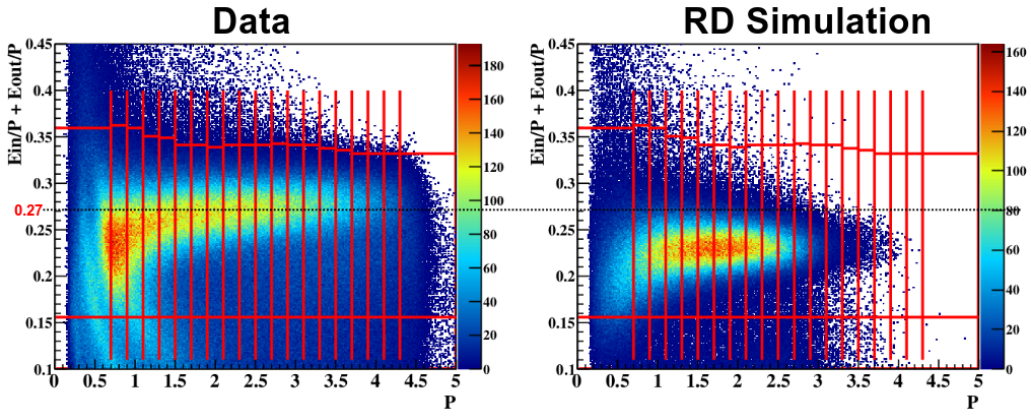


Figure A.5: Sum of the inner and outer energy of the EC, normalized by the momentum as a function of momentum. The red lines are the applied cuts for electron selection.

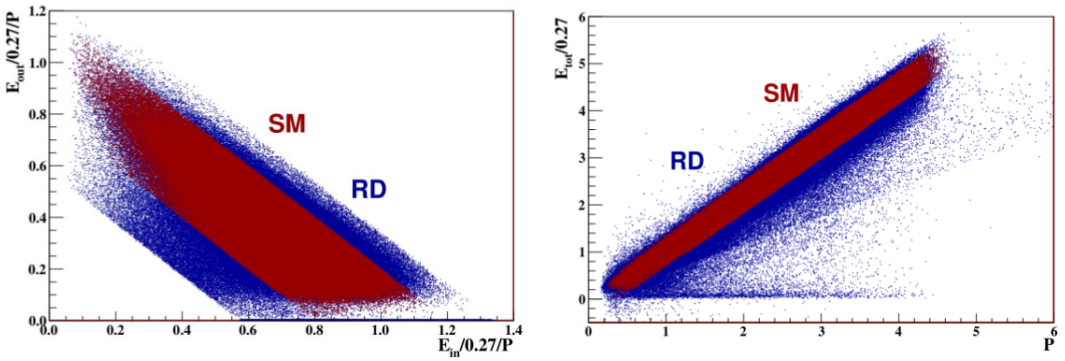


Figure A.6: Comparison between the SM and RD analysis regarding the selection cuts for electrons.

$$\Delta\beta = \beta - \frac{p}{\sqrt{p^2 + m_{\pi+}^2}} \tag{A.2}$$

where a cut is performed for different P ranges, analogous to the π^+ selection described before, using ΔT instead. The explicit cut is in equation A.3.

$$\Delta\beta(p)\ \text{Cut:} \begin{cases} \Delta\beta \leq +0.03 & ; \quad \text{All } p \\ \Delta\beta \geq -0.03 & ; \quad 0 \leq p \leq 1.487 \\ \Delta\beta \geq \dfrac{p}{\sqrt{p^2+0.4^2}} - \dfrac{p}{\sqrt{p^2+m_{\pi+}^2}} & ; \quad 1.487 < p \leq 1.838 \\ \Delta\beta \geq -0.02 & ; \quad 1.838 < p \leq 3.375 \\ \Delta\beta \geq \dfrac{p}{\sqrt{p^2+0.7^2}} - \dfrac{p}{\sqrt{p^2+m_{\pi+}^2}} & ; \quad 3.375 < p \leq 6 \end{cases} \tag{A.3}$$

The cut is shown in figure A.7, for solid target and for deuterium. A black dashed line at 0 gives us a reference point and, according to that, a shift is observed in both cases: an upward shift for the solid case and a downward for the liquid target.

The fact that there is a shift in the simulation distribution compared to the data could produce a bias if the cuts are not tuned for that case. To explore this aspect of the analysis, a set of shifts in the cuts applied to the simulations is implemented, for consistency, in electron and pions as well; and for all the four targets, C, Fe, Pb, and D2. In some cases, it is big, while in other cases, it is small.

If we do not consider the shifted distributions and tuned the cut, we end up with sub-optimal cuts such as the one depicted in figure A.8.

The shifts are in table A.2.

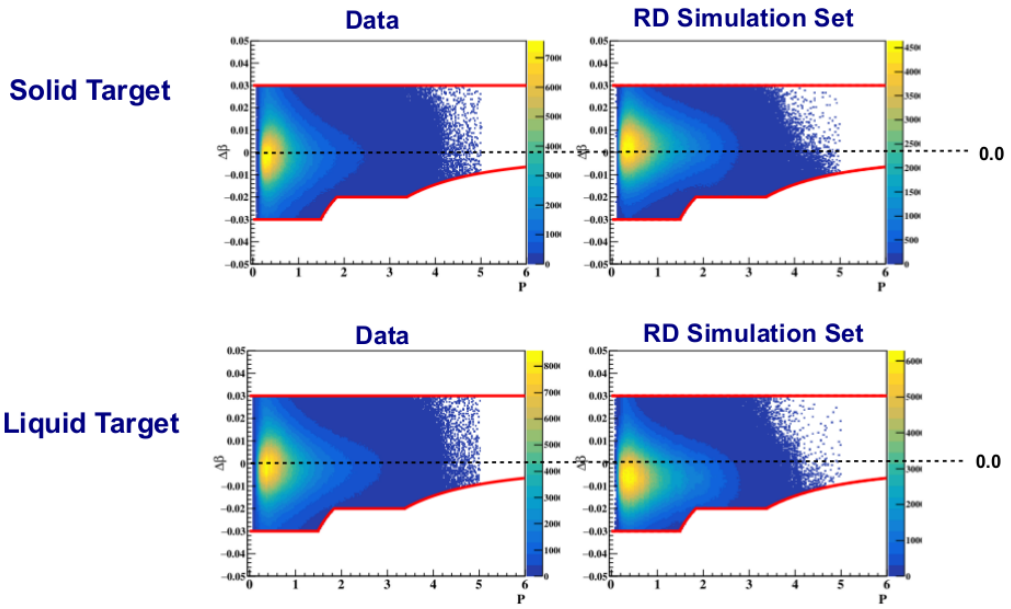


Figure A.7: Main cut for selecting π^+, according to RD analysis. On the left, there is the data ($\Delta\beta$,p) phase space, and on the right, the ($\Delta\beta$,p) for reconstruction. The red lines highlight the cuts from equation A.3.

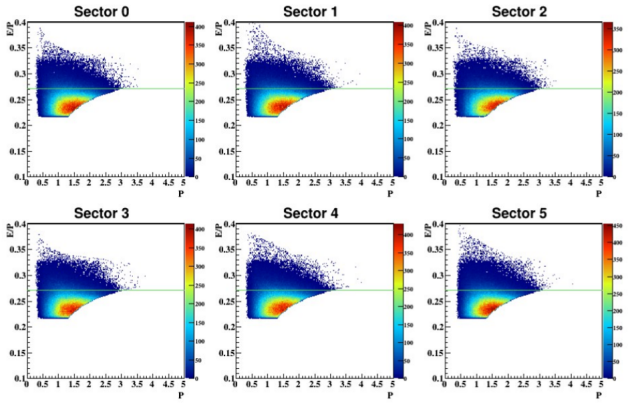


Figure A.8: First attempt to implement the sampling fraction cut, as implemented in SM analysis in RD simulations. The result is not optimal, if the shift in the distribution is not consider.

Shifts in Particle Identification Cuts		
Target	Electron	Positive Pion
C	-0.03	0.05
Fe	-0.03	0.04
Pb	-0.03	0.04
D2	-0.03	-0.05

Table A.2: Shifts on the cuts, according to RD analysis, applied in RD simulations.

Figures A.9 and A.10 shown the new and previous cuts (dashed lines).

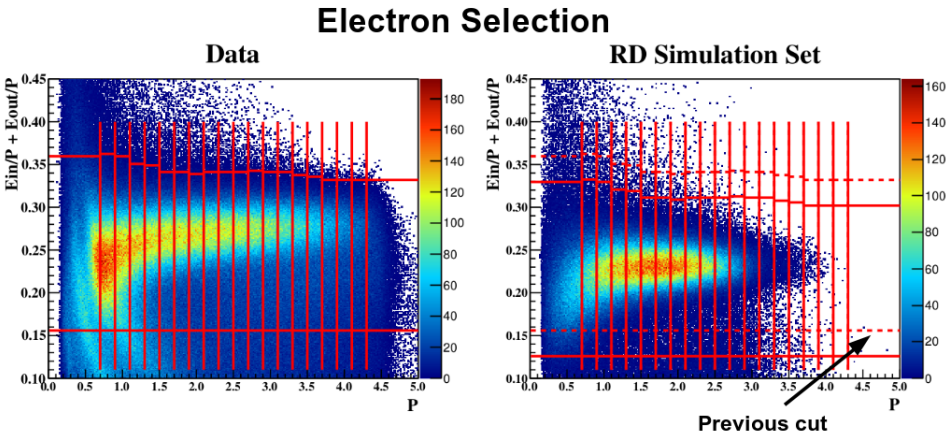


Figure A.9: Sum of the inner and outer energy of the EC, normalized by the momentum as a function of momentum. The red lines are the applied cuts for electron selection with the shift included, and the dashed red lines are the previous cuts. The cuts in data remains untouched.

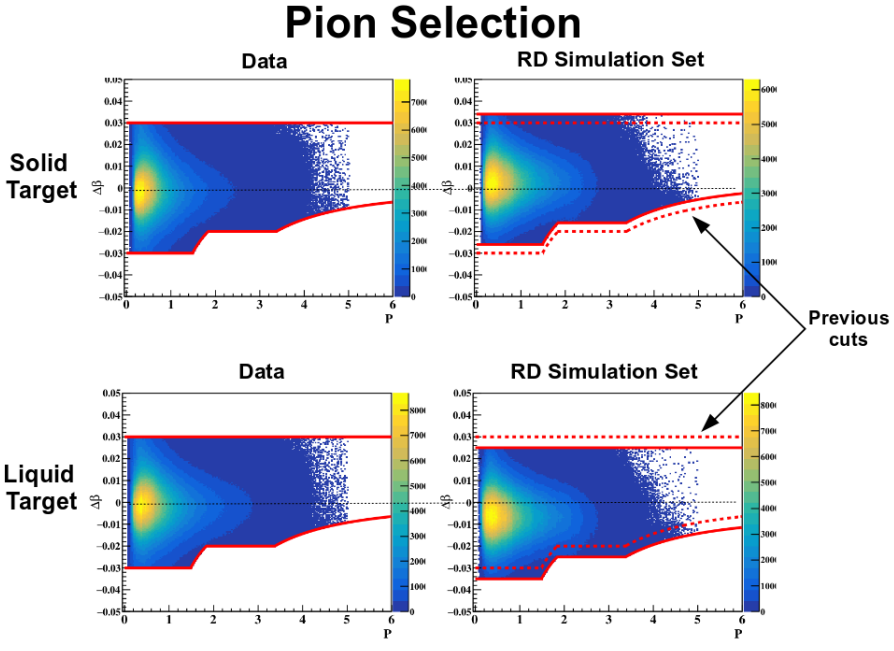


Figure A.10: Main cut for selecting π^+, the dashed lines are the previous cuts, and the full lines are the modified cuts. The cuts in data remains untouched.

After the implementation of this tuning process for reconstruction cuts, the result in the MR observable is shown in figure A.11. The previous case (blue dots in the plot) showed that the acceptance makes a difference in the MR up to 20%, but with

the modified cuts the new results (red dots in the plot) shows a behavior similar to SM case; the acceptance having an effect of a few percent. Since the only modification was in the simulation set, the uncorrected case remains the same. This procedure can be performed in all the remaining targets, following table A.2.

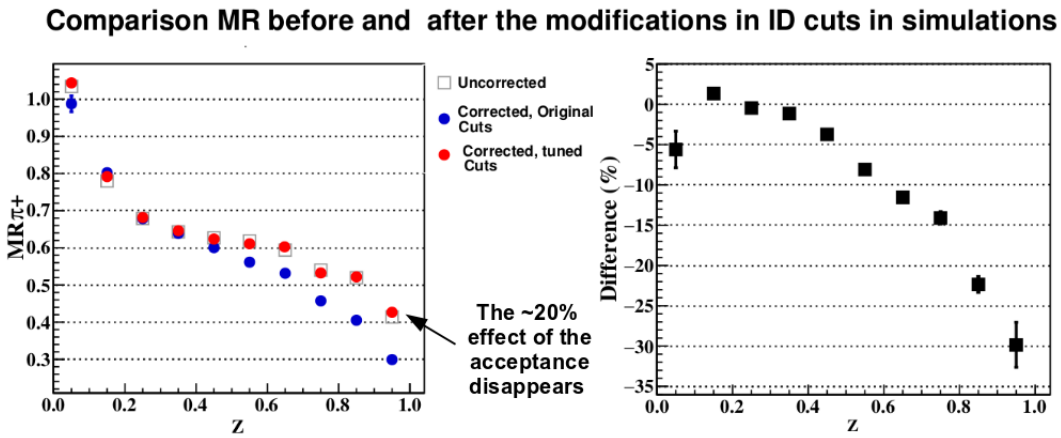


Figure A.11: Comparison of MR as a function of Zh for the previous result and, after all the particle identification cuts, were tuned in the reconstruction. On the right hand side, there is the difference (in %) between the two cases.

A.3 Results

The comparison for all the targets and the two analyses are plotted in figure A.12. As we can see on the right-hand side of the plot, the difference between the two analyses for the main part is not bigger than 2%. The label *Ana1* stands for the analysis following SM implementation of the cuts, and *Ana2* stand for Raphael's.

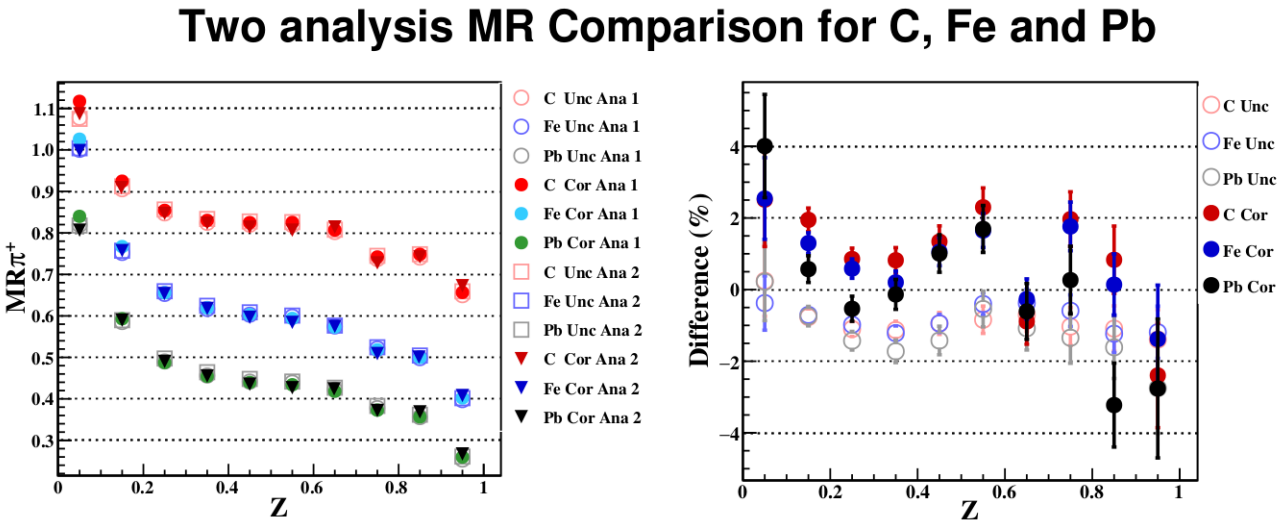


Figure A.12: MR Comparison for both analyses, uncorrected and acceptance corrected results. On the right hand side, there is the difference between the two uncorrected cases (hollow circles), and the difference between the acceptance corrected ones (full circles).

A.4 Conclusion

As this shows, the main source of discrepancy was that different CLAS simulations, which were run at different times by different people, produced different results from the data for the EC and $\Delta\beta$. At the same time, the cuts applied to the simulations were the same as were used on the data. Better quality control of the CLAS simulations could have completely avoided this problem. This should be implemented at the level of the CLAS collaboration, not at the level of the individual researcher.