Monte Carlo-Data validation for Electron Ion Collider Group

A Thesis

submitted to
Indian Institute of Science Education and Research Pune
in partial fulfillment of the requirements for the
BS-MS Dual Degree Programme

by

Akhil Mithran

Indian Institute of Science Education and Research Pune
Dr. Homi Bhabha Road,
Pashan, Pune 411008, INDIA.

July, 2021

Supervisor: Dr. Sadhana Dash

Certificate

This is to certify that this dissertation entitled Monte Carlo-Data validation for Electron Ion Collider Group towards the partial fulfilment of the BS-MS dual degree programme at the Indian Institute of Science Education and Research, Pune represents study/work carried out by Akhil Mithran, BS-MS student at Indian Institute of Science Education and Research under the supervision of Dr. Sadhana Dash, Associate Professor at Department of High energy physics during the academic year 2015-2021.

Sadhana Dash

Dr. Sadhana Dash

Committee:

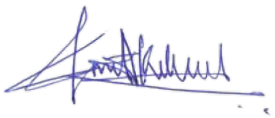
Dr. Sadhana Dash

Dr.Sourabh Dube

This thesis is dedicated to my Parents and Brother, for endless love and support during the whole course of the project

Declaration

I hereby declare that the matter embodied in the report entitled Monte Carlo-Data validation for Electron Ion Collider Group are the results of the work carried out by me at the Department of High energy physics under the supervision of Dr. Sadhana Dash and the same has not been submitted elsewhere for any other degree.



Akhil Mithran

Acknowledgments

I would like to express my sincere gratitude to Professor Sadhana Dash for providing me with an excellent opportunity to work on such an important and exciting project, which not only is of great significance but also has critical need for the future of High energy physics. I am also grateful to Dr Markus Diefenthaler for helping me setting up the work system at JLab during the final stages of the project and help me achieve a good workflow right after it's commencement. I once again thank Professor Sadhana for her ingenious planning and organizing the project to reach maximum efficiency and effectiveness. I am indebted to the Thomas Jefferson National laboratory for providing me system access during of the course of this project. I also convey my appreciation to Subhadeep for helping me with the resolve some issues for some components of the project. I am incredibly grateful to my parents and siblings for always being there for me with or their relentless support and encouragement. I am also grateful to my friends Adithyan, Harish and Nevin for providing some extra information and save time. Finally, I thank my friend Arya P V whom I could always count on during any difficult situations.

Abstract

The upcoming Electron Ion collider facility is aimed at providing more information on the structure inside the hadrons. As part of improving the detector development and integration of software into the main workflow of data analysis and prediction. Modern day general purpose generators like Herwig7 and Pythia8 are tested using data validation against Deep Inelastic Scattering data from HERA. Data for DIS was taken and compared in a comprehensible manner using rivet tool which is excellent for data comparison and generator validation. Overall performance of the generators are slightly inconsistent to be considered as more accurate than the other, for the list of experiments considered, The specific areas where the generators are more reliable than the other is given in the summary chapter. The possible reasons for the failure of a particular generator to reproduce the data properly in certain regions are also hypothesized. The results provide a useful input to the software working group of EIC when it comes to making decisions on integrating the software to workflow and use for data prediction and detector development.

Keywords; Phenomenology, High Energy Physics, Pythia8, Herwig, Rivet, Electron Ion Collider, Data validation, Monte Carlo, Simulation.

Contents

Introduction

This project work is carried out as part of the Software Working Group of the upcoming Electron Ion Collider at Brookhaven Lab. The main objective of the project lies towards providing a quantitative assessment of the detector capabilities and the associated physics involved so as to enhance the Technical Design Report (TDR) for the detectors involved. The main objective of the upcoming Electron-Ion collider (EIC) at Brookhaven National Laboratory is to look inside the nucleus. The internal structure of the nucleons like protons or neutron is studied or in their own words it's to take 3D snapshot of their internal structure. Such a large scale experiment involves various components to ensure the validity and accuracy of the data being produced and analysed. As such, one of the responsibilities of the Software Working group of the EIC group is to test the capabilities of the Monte Carlo event generator (MCEG) tools and analysis framework available in the context of reproducibility of the real experimental data. For the initial phase, data comparisons are carried out for deep inelastic collision data from the HERA experiment.

The deep inelastic scattering (DIS) experiments are quite important with respect to the tools they provide for testing Quantum Chromodynamics (QCD) by probing the structure function of the nucleon. These functions are necessary when it comes to precision measurements of the coupling constant $\alpha_s(M_Z^2)$ in QCD and the various distribution functions of the parton. In fact these quantities are necessary for all the precision calculations at hadron colliders. As such some of the important observables like charged particle multiplicity, transverse momentum spectra, vector meson production cross-sections etc will be looked at to valid the capability of various modern day MCEGs.

For this project, the latest software toolset at the time of project start, which were Pythia 8.303, Rivet 3.1.2 and Herwig 7.2.1 were used to simulate the data required for the validation as well as analysis and thus check the reliability and consistency in creating a

regular workflow for full detector simulations. The software validation not only contributes to the development of the future detector but also serve as a repository for reference which important when it comes to integrating the software in the EIC software collection and workflow.

During project duration we were able to compare a wide range of experiments with the simulated data and draw conclusions on the accuracy of reproduction. Of the Monte Carlo Event Generators compared, Herwig and Pythia, a single one cannot be recommended universal application out-of-the-box, i.e, with the default settings. This is because there are parts of the experimental subset which are reliably produced by one of the generators whereas the other one relatively requires improvement. These drawbacks and shortcomings of the various softwares are taken note of by the respective developer's team member present during the EIC India Software working group presentations.

Chapter 1

Deep Inelastic Scattering

1.1 Early years

The compositeness of matter was associated with the existence of indivisible structures called atoms from early on [1]. However, more constituent structures were being resolved after the advent of radioactivity from α, β and γ [2] rays which removed the limitations that came from the usage of visible spectra and beyond observation of matter.In a sense, it can be argued that the quest for exploring deep into atomic structures started off after the discovery of nucleus by Rutherford in 1911 using [3]α-particles scattered off gold, which undoubtedly raised curiosity on further sub-structures to the constituents. This atomic nucleus itself was exposed of its composite nature after the discovery of neutron by Chadwick [4] and the model for nuclear forces proposed by Yukawa [5]. In 1933 Frisch and Stern discovered the anomalous magnetic moment of the proton [6]and in 1939 Alvarez and Bloch discovered the anomalous magnetic moment of the neutron. These moments being different from that of electron or point-like particles [7] at the time were indicative of the existence of constituents in the nucleons. Further details pointing to substructure in nucleons was from the experiment by Hofstadter [8] in 1950 which revealed the distribution of charge inside the nucleon. The experiment had sufficient virtuality and shows a core and tail distribution that is positive for both proton and neutron, whereas the vector cloud is positive for proton and negative for neutron.

Quarks were proposed by Gell-Mann [9] and Zweig [10] in 1964 as the constituents of

hadrons so as to account for the various types of mesons and baryons. During this period, experiments conducted at Stanford Linear Accelerator (SLAC) on deep inelastic electron-nucleon scattering [11] - [17] for regions beyond the resonance and at relatively smaller distances showed that longitudinal structure functions were small and a scaling behaviour. This is also provided support for the predictions on scattering involving spin 1/2 particles by Callan and Gross [18]. The prediction of scaling with respect to the structure functions had also been predicted earlier by Bjorken with the help of algebraic methods [19]. All these new observation allowed Feynman to construct a model of point-like fermions in the nucleus that directly interacted with the electroweak gauge boson in the DIS process, the so called Parton model [20] [21]. This model brought about a new way for the understanding the composite nature of hadron with the confined point-like fermions and the involvement of strong interactions.

The theory of strong interaction itself took years of development before attaining its modern form. In 1965, a Yang-Mills SU(3) gauge theory [22] [23], built upon the 3-valued degrees [24] of freedom of charge, for the strong force was introduced by Nambu. Later on Hooft [25] had succeeded in renormalization of Yang-Mills theory, after the quantization in covariant gauges was possible with the help of formalism by Faddeev and Popov [26]. Finally adopting Quantum Chromodynamics (QCD) as a theory of the strong interactions was done in 1972 by Gell-Mann [27] and Fritzsch. Studies on the running strong coupling constant from Yang-Mills theory was done by Wilcsek and Politzer using color triplet quarks and observed asymptotic freedom [28] [29].

Overall, during the last 40-50 years, there has been major improvements in the measurements of DIS structure functions, with greater precision for both neutrino-nucleon as well as lepton-nucleon scattering. This is in tun allows more stringent precision tests for QCD. The accuracy of measurement of strong coupling α_s (M_Z^2) constant has also improved and same can be said for parton density functions, a essential component of hadron collider facilities at LHC. It thus, open up a path way for better studies and searches for new particles in the future.

1.2 Electron-Proton Scattering

1.2.1 Elastic Scattering

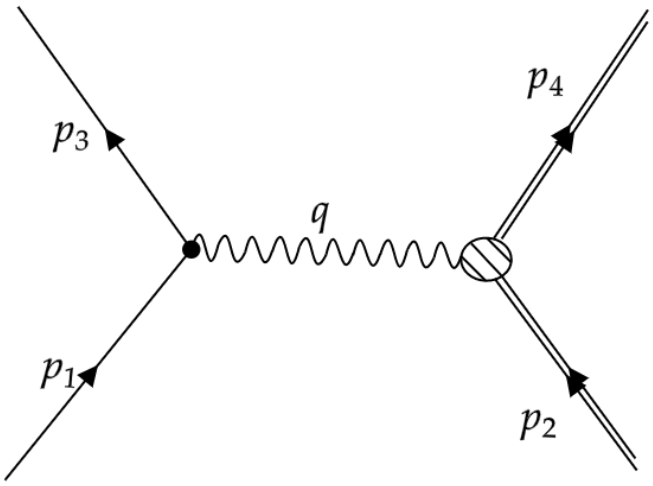


Figure 1.1

A good starting point would be to approximate the proton as a point-like particle. This enables to use the well studied equations for the electron-muon scattering with just the charge and mass of the muon replaced appropriately as that of the proton. We could then easily obtain the amplitude of the process averaged over spin to be :

$$\left\langle |M|^2 \right\rangle = \frac{g_e^4}{q^4} L^{\mu\nu}_{electron} L_{\mu\nu\ proton} \tag{1.1}$$

where $q = p_1 - p_3$ is the momentum of the virtual electroweak gauge boson, say photon here and

$$L^{\mu\nu}_{electron} = 2[p_1^\mu p_3^\nu + p_3^\mu p_1^\nu + g^{\mu\nu}(m_e^2 c^2 - p_1 \cdot p_3)]$$

Now the problem with this equation can be addressed starting with the assumption that proton is a point-like particle. So, the first step would be to replace $L_{\mu\nu proton}$ with a different function. Let's take this function to be $K_{\mu\nu proton}$. Some restrictions should be imposed on this function to make it real. The first of all the function should be a rank 2 tensor and the only variables that should be involved in the equation are q, p_2 and p_4. Of these variables,

the relation $q = p_4 - p_2$ allows to eliminate one the variables which is conventionally chosen to be p_4. Now we can equate the relation between the second-rank tensor and the variables involved as :

$$K^{\mu\nu}_{proton} \;=\; -K_1 g^{\mu\nu} \;+ \frac{K_2}{(m_p c)^2} p^\mu p^\nu \;+\; \frac{K_3}{(m_p c)^2} q^\mu q^\nu \;+\; \frac{K_2}{(m_p c)^2} (p^\mu p^\nu \;+\; p^\nu p^\mu$$

Here, the functions K_i are functions whose form is not yet identified. To breakdown the equation, we can see terms multiplied by $1/(m_p c)^2$ which is to ensure same dimensionality for the K_i functions. There is also a symmetric term present but no asymmetric term with respect to p and q variables. This is so because the $L^{\mu\nu}_{electron}$ is symmetric and so there won't be any corresponding contributions in $\langle |M|^2 \rangle$. In the scalar variables associated with K_i function, namely, q^2, p^2 and $p \cdot q$, the relation $p^2 \;=\; (m_p c)^2$ can be used along with the four-momentum conservation to obtain $q \cdot p \;=\; -q^2/2$

This shows that K_i have to be a function of q^2 only. Further limitations on K_i is imposed from the conservation of charge equation :

$$q_\mu K^{\mu\nu} \;=\; 0$$

which after expansion with q_μ and p_ν gives out the following two equations,

$$K_3 \;=\; \frac{(m_p c)^2}{q^2} K_1 \;+\; \frac{1}{4} K_2, \qquad\qquad K_4 \;=\; \frac{1}{2} K_2$$

. This implies that $K^{\mu\nu}$ are dependent on only the function of K_i which themselves are functions of q^2.

$$K^{\mu\nu}_{proton} \;=\; K_1 \left(-g^{\mu\nu} \;+\; \frac{q^\mu \nu}{q^2} \right) \;+\; \frac{K_2}{(m_p c)^2} \left(p^\mu + \frac{1}{2} q^\mu \right) \left(p^\nu + \frac{1}{2} q^\nu \right)$$

The K function can be determined experimentally from the cross-section for elastic electron-proton scattering and they are called the form factors. They are important components which provide useful information about the distribution of charge within the proton. Combining and simplifying the above equations, we can get the result of:

$$\left\langle |M|^2 \right\rangle = \left(\frac{2 g_e^2}{q^2} \right)^2 \left\{ K_1 [(p_1 \cdot p_3) - 2(m_e c)^2] + K_2 \left[\frac{(p_1 \cdot p)(p_3 \cdot p)}{(m_p c)^2} + \frac{q^2}{4} \right] \right\} \qquad (1.2)$$

This equation can be used for determining the cross section for elastic electron-proton scattering. After setting up certain assumptions, it can be shown that the four momenta relevant for the purpose at hand are:

$$p_1 = E_1/c(1, \hat{\mathbf{P}}_1) \quad p_2 = p = (M_p c, 0) \quad p_3 = E_3/c(1, \hat{\mathbf{p}}_3) \tag{1.3}$$

The core assumptions to be specified here is that the the calculation is in lab frame, the proton being t rest and the collision energy is moderate with $E_1 >> m_e c^2$ but sufficient enough that the electron mass can be neglected. Furthermore, the amplitude can be calculated to:

$$\left\langle |M|^2 \right\rangle = \frac{g_e^4 c^2}{4E_1 E_3 sin^4(\theta/2)} [2K_1 sin^2(\theta/2) + K_2 cos^2(\theta/2)] \tag{1.4}$$

Here, θ is the scattering angle between p_1 and p_3. Plugging this equation in the appropriate equation for the cross-section of the collision of a massless incident particle in the lab frame gives out :

$$\left\langle |M|^2 \right\rangle = \frac{\alpha h}{4m_p E_1 sin^2(\theta/2)} \frac{E_3}{E_1} [2K_1 sin^2(\theta/2) + K_2 cos^2(\theta/2)] \tag{1.5}$$

E_3 can be determined using conservation of four-momenta. This equation is known as Rosenbluth formula. It provides a way of determining the form factors K functions from the measurement of scattered particles in a certain solid angle as well as from the calculation of the differential cross-section.

By taking the form factors K_i to be a particular form, say :

$$K_1 = -q^2, \quad K_2 = 4(m_p c)^2 \tag{1.6}$$

it would be capable of describing the proton as if it were a point-like particle. More accurately it can be used in electron-muon collision instead of the electron-proton case at hand. However, due to the composite nature of the proton, this will be a poor approximation. But nonetheless it can be useful when dealing with the DIS formalism.

1.2.2 Inelastic *ep* Scattering

The scattering process that's been considered so far becomes more involved at higher energies when there's creation of a variety of hadrons or a so-called hadronic system. Such a scattering

process is known as inelastic.

$$e + p \rightarrow e + X \tag{1.7}$$

Here, X represents the final state hadronic system. The corresponding Feynman diagram is usually given as below:

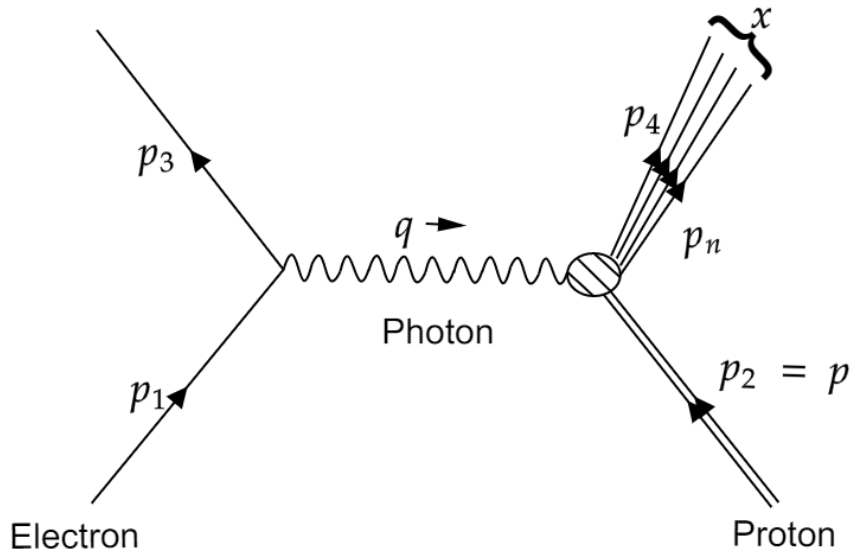


Figure 1.2

One thing that is common with the elastic scattering process is that the amplitude can taken in a similar form as:

$$\left\langle |M|^2 \right\rangle = \frac{g_e^4}{q^4} L^{\mu\nu}_{electron} K_{\mu\nu}(X) \tag{1.8}$$

This amplitude can be used through Fermi's golden rule to calculate scattering cross-section with the same conditions as the previous case, with the factor $K_{\mu\nu}$ depending only on q, p_2 and that of other outgoing particles. After proper substitutions, the equation left with is:

$$\begin{aligned} d\sigma \;=\; & \frac{h^2 \left\langle |M|^2 \right\rangle}{4q^2\sqrt{(p_1 \cdot p_2)^2 - (m_1 m_2 c^2)^2}} \left[\left(\frac{cd^3\mathbf{p}_3}{(2\pi)^3 2E_3} \right) \left(\frac{cd^3\mathbf{p}_4}{(2\pi)^3 2E_4} \right) \right. \\ & \left. \cdots \left(\frac{cd^3\mathbf{p}_n}{(2\pi)^3 2E_n} \right) \times (2\pi)^4 \delta^4(p_1 + p_2 - p_3 \cdots - p_n) \right. \end{aligned} \tag{1.9}$$

One of the advantages is that better detection capabilities when working with electron sig-

natures. As for the rest of the registers, an inclusive sum over all the possible configurations of the final state hadrons and integrating over the momenta of scattered particles is done. This inclusive cross-section can be formulated as :

$$d\sigma \;=\; \frac{h^2 g_e^4 L^{\mu\nu}}{4q^2\sqrt{(p_1\cdot p_2)^2-(m_1m_2c^2)^2}}\left(\frac{cd^3\mathbf{p}_3}{(2\pi)^3 2E_3}\right)4\pi m_p W_{\nu\mu} \tag{1.10}$$

The W factor is of the form,:

$$W_{\mu\nu}=\frac{1}{4\pi m_p}\sum_X\int\cdots\int K_{\mu\nu}(X)\left(\frac{cd^3\mathbf{p}_4}{(2\pi)^3 2E_4}\right)\cdots\left(\frac{cd^3\mathbf{p}_n}{(2\pi)^3 2E_n}\right)\times(2\pi)^4(2\pi)^4\delta^4(p_1+p_2-p_3\cdots-p_n) \tag{1.11}$$

As is the case for the subsection, there are certain changes in the frame of calculation to make things easier along with with certain assumptions. The calculation are done in the lab frame with the proton at rest and electron being incident on it. The energy of the incident electron is moderate enough to ignore the electron mass. There is a change of coordinates to spherical coordinates, $d^3\mathbf{p}_3=|\mathbf{p}_3|^2 d|\mathbf{p}_3|d\omega$, along with change to the energy with $|\mathbf{p}_3=E_3/c$. After all these we are left with,

$$\frac{d\sigma}{dE_3 d\Omega}=\left(\frac{\alpha h}{cq^2}\right)^2\frac{E_3}{E_1}L^{\mu\nu}W_{\mu\nu} \tag{1.12}$$

Unlike the corresponding equation for the elastic case, conservation of momentum cannot be used to determine E_3 due to the loss of momentum through extra hadrons produced in the collision. To overcome this, the differential cross-section is considered in a particular range dE_3.

The $W_{\mu\nu}$ factor here would carry similar properties as the $K_{\mu\nu}$ factor earlier, with it being a rank 2 tensor with q and p variables,

$$W^{\mu\nu}=-W_1 g^{\mu\nu}+\frac{W_2}{(m_pc)^2}p^\mu p^\nu+\frac{W_3}{(m_pc)^2}q^\mu q^\nu+\frac{W_4}{(m_pc)^2}(p^\mu p^\nu+p^\nu p^\mu) \tag{1.13}$$

Similar to previous notations, W_i are functions consisting of q^2, p^2 or $q\cdot p$. This time however, the dependency cannot be restricted to just q^2 by eliminating p^2 and $q\cdot p$ because the same constraints cannot be applied to the total momentum of the hadronic system X. So, except for p^2 which is a constant, the W_i has to be considered a function of q^2 and $q\cdot p$.

However, further reduction is possible with the relation between the W_i equations of the

form,

$$W_3 = \frac{(m_p c)^2}{q^2} W_1 + \left(\frac{q \cdot p}{q^2}\right) W_2, \qquad W_4 = -\left(\frac{q \cdot p}{q^2}\right) W_2 \tag{1.14}$$

which is possible with the constraint,

$$q_\mu W^{\mu\nu} = 0 \tag{1.15}$$

Using this relation between the W_i functions, the equation can be further reduced to

$$W^{\mu\nu} = W_1 \left(-g^{\mu\nu} + \frac{q^\mu q^\nu}{q^2}\right) + \frac{W_2}{(m_p c^2)^2} \left[p^\mu - \left(\frac{q\cdot}{q^2}\right)^2 q^\mu\right] \left[p^\nu - \left(\frac{q\cdot}{q^2}\right)^2 q^\nu\right] \tag{1.16}$$

The W functions here, namely $W_1(q^2, q \cdot p)$ and W_2 are the structure function. After substituting Eqn 1.16 and Eqn1.14, a more simplified form can derived as below:

$$\frac{d\sigma}{dE_3 d\Omega} = \left(\frac{\alpha h}{2E_1 sin^2\left(\theta/2\right)}\right)^2 \left[2W_1 sin^2\left(\theta/2\right) + W_2 cos^2\left(\theta/2\right)\right] \tag{1.17}$$

The Eqn 1.17 is one of the most fundamental result associated with inelastic scattering. It is also useful when it comes to determining the structure function W_1 and W_2 from experiments. As will be useful for better correlation later on, the dependence on q^2 and $q \cdot p$ can be changed to a dependence on q^2 and the Bjorken scaling variable x.

Considering the fact that elastic scattering is a special form of inelastic scattering, particular forms for W_1 and W_2 can be derived in order to convert the above Eqn 1.17 into the Rosenbuth formula formulated earlier. One particular form that satisfies all the conditions is as follows:

$$W_{1,2}(q^2, x) = -\frac{K_{1,2}(q^2)}{2m_p q^2} \delta(x_i - 1) \tag{1.18}$$

This was obtained after integrating over E_3. The equation comes with a delta function which specifies that the Bjorken scaling variable x should be equal to 1. Another way to look at this is by considering the formula for x in the lab frame, $x = -q^2/(q \cdot p)$ which translates to $q \cdot p = -q^2/2$ which is as expected for elastic scattering equation.

This equation can further be reduced if the proton is considered to point-like with the

corresponding form factor K_i derived earlier.

$$W_1^i = \frac{1}{2m_i}\delta(x_i - 1), \qquad W_2^1 = -\frac{2m_i c^2}{q^2}\delta(x_i - 1) \tag{1.19}$$

1.3 Deep Inelastic Scattering

With the formalism derived earlier it's now convenient to move into higher energy regimes and construct a model for the structure function W_i and compare it to the experimental data.

1.3.1 Quark Model

Here, the concept of proton consisting of quarks, the so-called point-like spin-1/2 particles is used to model the scattering process at energies and extra details regarding the proton structure functions. As such, the high energy inelastic electron proton scattering, i e , the deep inelastic scattering can be modeled as the sum of multiple elastic electron quark scattering.

Using the equations for the structure functions derived earlier Eqn 1.19. for elastic scattering off a point like particle and replacing the mass and charge terms appropriately with that of a quark of a particular flavour i, leads to:

$$W_1^i = \frac{Q^2}{2m_i}\delta(x_i - 1), \qquad W_2^1 = -\frac{2m_i c^2 Q_i^2}{q^2}\delta(x_i - 1) \tag{1.20}$$

repeating the same notations, m_i is the quark mass, p_i its momentum and the Bjorken variable $x_i = \frac{-q^2}{q \cdot p}$ for each of the corresponding quark. The Q_i is the quark charge which is fraction of the proton charge, namely, 2/3 for up quarks and -1/3 for down quarks.

To bring in the Bjorken variable into the structure function, the z_i is introduced which is the fraction of the 4-momentum carried by the struck quark.

$$p_i = z_i p \tag{1.21}$$

Using the Lorentz scalar variables, this can be converted into

$$m_i = z_i m_p \tag{1.22}$$

This implies that $x_i = x/z_i$. The functions W_i are then converted to

$$W_1^i = \frac{Q_i^2}{2m_p}\delta(x - z_1), \qquad W_2^i = -\frac{2m_p c^2 Q_i^2}{q^2}\delta(x - z_i) \tag{1.23}$$

To obtain the structure function the individual $W - i$ functions for all the flavours of quarks should be summed over and integrated over z_i. A factor measuring the probability of an ith quark carrying momentum fraction z_i should also be included. Overall, the structure function can be written as

$$W_1 = \sum_i \int_0^1 \frac{Q_i^2}{2m_p}\delta(x - z_1) f_i(z_i) = \frac{1}{2m_p}\sum_i Q_i^2 f_i(x) \tag{1.24}$$

$$W_2 = \sum_i \int_0^1 \frac{-2x^2 m_p c^2}{q^2} Q_i^2 \delta(x - z_1) f_i(z_i) = \frac{-2m_p c^2}{q^2} x^2 \sum_i Q_i^2 f_i(x) \tag{1.25}$$

Taking out the common factor involving q can be taken outside and multiplied by its reciprocal to remove it completely.

$$F_1(x) = m_p W_1 = \frac{1}{2}\sum_i Q_i^2 f_i(x) \tag{1.26}$$

$$F_2(x) = -\frac{q^2}{2m_p c^2 x} W_2 = x\sum_i Q_i^2 f_i(x) \tag{1.27}$$

This new functions, after removing the q dependence is only dependent on the Bjorken scaling variable x. This Independence on energy factor q^2 and Bjorken scaling variable. The obvious relation between $F_1(x)$ and $F_2(x)$ which is

$$F_2(x) = 2xF_1(x) \tag{1.28}$$

is known as the Callan-Gross relation. Before formulating the cross-section equation, it is useful to look at the implication of the Callan-Gross relation. The ratio of the F_i functions which is 1/2 implies that the the constituents that the electron interact within the proton

are spin 1/2 particles.Now, looking at the cross-section,

$$\frac{d\sigma}{dE_3 d\Omega} = \frac{F_1(x)}{2m_p}\left(\frac{\alpha h}{E_1 sin\,(\theta/2)}\right)^2\left[1 + \frac{2E_1E_3}{(E_1 - E_3)^2}cos^2\,(\theta/2)\right] \tag{1.29}$$

The effectiveness of the Bjorken scaling and Callan-Gross relation is that the Cross-section reduces to the dependence on just the function F_1 which itself is dependent the probability factors $f_i(x)$.

1.3.2 Observables of Interest

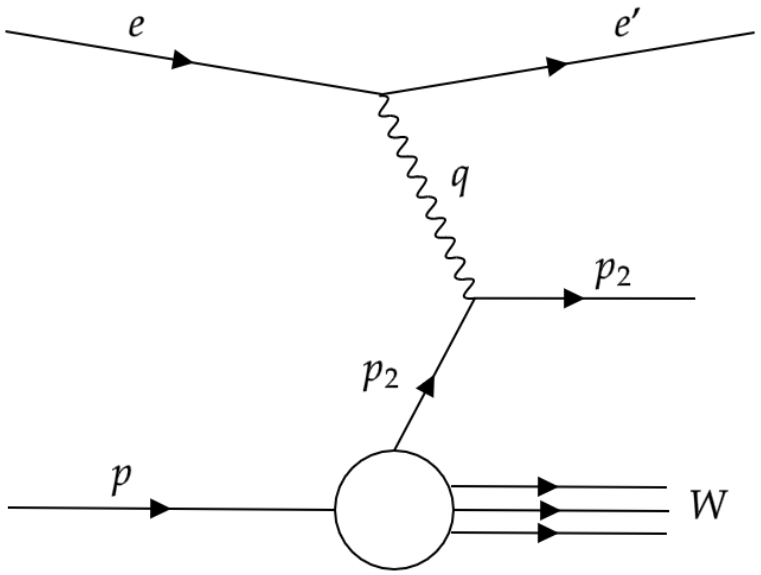


Figure 1.3

For this particular project, it will be convenient to establish clearly some of the kinematic variables involved so that the later cross-check and interpretations can be done easily. First of all, use the Figure 1.3 to visualize the a generalised form of the deep inelastic scattering process at the Born level. Here, the lepton, which can any of $e^{\pm}, \mu^{\pm}, \nu_i(\bar{\nu}_i)$ carrying a momentum k_1 is scattered of the nucleon of initial momentum p_1 with an exchange of elec-

troweak gauge boson, which can be any of the form $\gamma, Z^0, W^{\pm}$. The outgoing lepton carries a momentum $k_2 = K_1 - q$ and the other , shown as 'X' is the final state hadronic system. The effective process can be shown as a $2 \to 2$ process with $k_1 + p_1 \to k_2 + p_2$. Whenever the exchanged electroweak boson is a photon or Z-boson, the process is called the neutral current DIS and the other one is called charged current DIS. They are represented by:

$$e^- p \to e^- X$$

$$e^- p \to \nu_e X$$

From the corresponding figure for DIS process, the following variables, widely used in the context later on, can be extract from the labelled four momenta.

1. The CMS collision energy, $s = (k_1 + p_1)^2$
2. The inelasticity variable $y = \frac{p_1 \cdot (k_1 - k_2)}{p_1 \cdot k_1}$
3. Inelasticity using the Jacquet-Blondel method :- $y_{jb} = \frac{E_h - p_{zh}}{2E_e}$
4. The Bjorken scaling variable, $x = \frac{Q^2}{sy}$
5. The mass of the $\gamma - p$ system or the mass of the final state hadronic system:

$$W^2 = p_2^2 = (q + p_1)^2 = M^2 + 2q \cdot p_1 - Q^2 = M^2 + Q^2(1 - x)/x$$

The inclusive the DIS process are characterised by having $W > \dot{2}$ GeV and large values for the corresponding Q^2. For process with $Q^2 \approx 0$ is in the photoproduction regime and those with $Q^2 > 1$ GeV^2 is in the DIS regime.

Chapter 2

Experiments list for analysis

A short and detailed descriptions of experiments that were used for EIC data validation will be provided. The description will go through the importance of the kinematic variable under observation and the various phase space requirements.

2.1 Charged Particle multiplicity

One of the basic measure of the properties of the multiple final state particles is the multiplicity distribution[30] of the final state hadrons. The distribution pattern of the particles for a particular phase space domain of the provides useful information regarding the nature of correlations of the hadrons. The total event multiplicity is important but what is more of interest is the multiplicity in a particular subdomains of phase space region where the restriction of global conservation conditions are reduced as well as better exposing the correlation between the dynamics of the particles.

The data used for comparison was taken in the by the H1 detector of the HERA experiment during the running period of 1994 with an integrated luminosity of 1.3 pb^{-1}. The study was carried out in the hadronic centre of mass frame or the rest system of the electroweak gauge boson and the proton ($\gamma * -p$), in specific subdomains of the pseudorapidity space, tracking its relation to the total hadronic centre of mass energy W and the negative of the four momentum squared of the gauge boson, Q^2.

Multiplicity distribution refers to the set of probabilities(P_n) associated with the occurrence of the number of hadrons, n in the phase space region under observation. These measurements, especially in terms of the variations in the statistics provide a way to directly measure the correlation strength among the objects, here, the particles being observed. The first person to exploit this in the field of high energy physics was Mueller, for developing the formalism for short range order.

The data used for multiplicity distribution in specific pseudorapidity bins were corrected for several effects. One such adjustment comes from the limitations of tracking system with respect to the acceptance rates and resolution by which many particles and events may be lost. Others include limitations in in the efficiency of finding tracks, background noise from interactions with the detector material and initial state radiations (ISR) due to QED. The final results used for comparison takes care of all such corrections.

2.1.1 Kinematic selection

As mentioned earlier, the data used for the comparison was taken during the running period of 1994 of the HERA storage ring at DESY. The Centre of Mass collision energy of the system was 300 GeV with respective energies of the positrons and protons used for collision being 27.5 GeV and 820 GeV. The phase space was also selected in such a way to minimize the all other undesired effects.

The the energy of the scattered electron/positron was required to be $e` > 12$ GeV to select the neutral current DIS events. This also helps to reduce the photoproduction background to less than 1%. Rapidity-gap events are removed from the data sample by requiring that the total energy deposited between polar angles $4.4 < \theta < 15$ should be less than 0.5 GeV. Only those tracks which lie between the range $15 < \theta < 155$ for the polar angle are selected. The pseudorapidity, in the hadronic centre of mass frame, range for the for the experiment was taken in the $\eta* > 0$ or the so-called full current hemisphere and for the range $1 < \eta* < 3$. These domains were later separated into more finer regions. The main observables apart from charged multiplicity are the mean charged multiplicity $< n >$, the generalised dispersions, $D_q = \left[\overline{(n - \bar{n})^q}\right]^{1/q}$ and the normalised multiplicity moments, $C_q = \langle n^q \rangle / \langle n \rangle^q$. The virtuality cuts applied were in the range $10 < Q^2 < 1000$. The invariant mass squared of the hadronic system was required to be $80 < W < 220$ GeV.

2.2 Transverse momenta

The next data comparison comes from transverse momenta[32] of the charged particle distribution calculations. As compared to the transverse energy flow measurements the the charged particle spectra provides more information regarding the underlying parton dynamics. As shown through some QCD models, high-p_T tail is related to the parton radiation.

2.2.1 Kinematic selection

The data used for this analysis is obtained from the HERA running period 1994. same as previous experiment. As such the centre of mass energy, $\sqrt{s} = 300$ GeV and the energies of the incoming proton and electron/positron beams are also similar. The major difference comes with respect to the virtuality cuts applied, which is $5 < Q^2 < 50\text{GeV}^2$. As for the other differences, the polar range of the scattered electron was required to fall below $\theta < 153$. The energy deposited in the forward region (polar range $4.4 < \theta < 15$) should be larger than 0.5 GeV. This is done so that the large rapidity gap events are eliminated. The measurements were carried out for various regions of the virtuality the Bjorken scaling variable x, which was taken in the range (0.1 , 10). With respect to pseudorapidity, the data falls in the current region as well as the central fragmentation region.

2.3 $D*$ production cross-section

The motivation for this data comparison comes form the importance of information available form the charm production measurements for testing quantum chromodynamics (QCD). This is because the main process by which charm quarks are produced involves boson gluon fusion process, $\rightarrow c\bar{c}$ which in turn depends on the gluon distribution inside the proton. The data taken for the analysis comes[31] from the HERA running periods (1992-2000) and (2003-2007). It also provided useful information in terms of the charm quark mass m_c, improving the predictions for the cross-section of Z- and W- production. The data was extrapolated to full phase space for comparison purposes but to minimise dependence theoretical input, the comparison was properly done directly in the visible phase space region with respect to cuts on inclusive DIS variables like the photon virtuality and inelasticity, y. The data also

provides a good signal-to-background ratio with small uncertainties.

2.3.1 Kinematic selection

The centre-of-mass energy of the data sample is $\sqrt{s} = 318$ GeV. The data corresponds to single differential cross-section for $D* \pm$ production with respect to the kinematic variables, transverse momentum, $p_T(D*)$, pseudorapidity, $\eta(D*)$ and the elasticity, $z(D*) = (E(D*) - p_Z(D*))/(2E)$ both measured in the lab frame and global variables Q^2 and y.

2.4 J/ψProduction

The motivation for J/ψ cross-section[35] studies is that it can be used for determining the density of gluon in the proton. This is so because the the gluon density is part of the cross-section calculation. Two of the important contributions to the inelastic production for J/ψ in ep collision at HERA are colour-singlet (CS) and colour-octet (CO) contributions. Which of these dominate depends on the inelasticity variable, z for which the CS is more dominant when $z \leq 0.7$ and CO will be dominant at higher z. The problem with $z \approx 1$ is the background noise from elastic and diffractive proton-dissociative production.

There has been studies on J/ψ production in the photoproduction regime with $Q^2 \sim 0$ and this set of data is focused on the DIS regime with $Q^2 \geq 1\,\text{GeV}^2$. More specific kinematic cuts will be given in the next subsection. Overall, the production cross-section for this regime of interest in DIS is smaller than that of photoproduction and the CO contributions are expected to be more significant. But one advantage is that at high Q^2 the background from diffractive processes a can be reduced.

2.4.1 Kinematic selection

The data used for comparison is taken from the running period of 1996-2000 and at HERA with the ZEUS detector. There are two sets of data with centre-of-mass energy 300 GeV and 318 GeV during the specific running periods 1996-97 and 1998-2000. They cross-sections

were later correct to only $\sqrt{s} = 318$GeV using Monte Carlo simulation. The cuts on the photon virtuality Q^2/GeV^2 was required to be in the range [2,80]. The inelasticity variable y was calculated using the Σ method. The cuts on the invariant mass of the hadronic final state system was required to fall in the range $50 < W < 250$ GeV.

The requirement $= \Sigma_i (E_i - p_{z_i}) = 2E_i = 55$ GeV with the sum running over all the final state particles, is implied by the conservation of the energy and longitudinal momentum p_Z. To further reduce the photoproduction background, the inelasticity variable calculated using the scattered lepton and polar angle, $y_e < 0.95$. The $y_{JB} > 0.02$ was also placed on elasticity variable obtained through the Jacquet-Blondel method. The tracks selected for calculation was required to have a transverse momentum $p_T > 100$MeV. Since, the J/ψ mesons were identified using the dimuon signature, each of the track was required to have a track momentum $|p| > 1$ GeV. The remaining cuts included, $-1.6 < Y_{lab} < 1.3$ where, the Y_{lab} is the rapidity of J/ψ meson in the laboratory frame and is defined by,

$$Y_{lab} = 1/2 \ln (E_\psi + p_{z_\psi})/(E_\psi + p_{z_\psi})$$

. This cut was chosen mainly because the acceptance is higher in for that region. The cut on the inelasticity, $0.2 < z < 0.9$ was placed to reduce background from fake muon and to minimize the diffractive production from proton dissociation. The cross-sections were calculated with respect to z, virtuality, W, and the observables of J/ψ in the hadronic centre-of-mass frame, which are $Y*$ and p_T^{*2}. The differential calculations for a given variable O is given by the formulae:

$$\frac{d\sigma_i}{dO} = \frac{N_i}{BLA_i(O)}$$

Overall, the calculation are carried out in the regions $2 < Q^2 < 80\mathrm{GeV}^2$, $50 < W < 250$GeV, $0.2 < z < 0.9$ and $-1.6 < Y_{lab} < 1.3$ alongwith measurements in the region, $2 < Q^2 < 100\mathrm{GeV}^2$, $50 < W < 225$GeV, $0.3 < z < 0.9$ and $p_T^{82} > 1\mathrm{GeV}^2$.

2.5 Inclusive ϕ-meson production

The total content of quark in the proton is determined through the DIS experiment. But there still remains to more information of be known about the sea of quarks especially in terms of its flavour decomposition. The data compared here is from the study of ϕ meson

production[37] in neutral current DIS collisions.

There are several processes involved in the production of ϕ-meson. The ϕ meson can be produced after the hadronisation of the strange quark from the strange sea of quarks inside proton after hard scattering off from an incoming photon which can be shown by $\gamma^* s \rightarrow s$. This hard scattering process can be modelled either through zeroth order QCD or the so-called Quark Parton Model (QPM) or through a first order interaction which is called the QCD Compton (QCDC) process. Apart from this hard scattering process. Another means of ϕ meson production is through the boson gluon fusion (BGF), $\gamma^* g \rightarrow s\bar{s}$. The main difference between this type of production and production of hard scattering is that, here the main dependency comes from the gluon density and is not dependent on the whether or not strange quarks are involved in the hard scattering. Instead the ϕ mesons are produced from the strange quark produced in the hadronisation process. The other background process that can contaminate the data comes from diffractive scattering and from strange quarks produced from splitting of gluon which later hadronize or when a higher mass state like D_s decays into a ϕ. Out of the main methods for studying strange particle processes are from the K^0 mesons and baryons. Unlike these hadrons, which have contributions from higher mass state decays and fragmentation process, the ϕ meson is sensitive to the strange quarks involved in the hard scattering process. They also have have less contribution from the resonance decays. The contribution from QPM can also be improved by filtering for ϕ mesons with high longitudinal momentum in the Breit frame.

The main signature for studying the ϕ mesons produced comes from the $\phi \rightarrow K^+K^-$. The Breit frame is the best when it comes to studying the system separated from the proton remnants. The photon in this frame is space-like with a four four-momentum, q = (0,0,0,-Q). All the final state particles with transverse momentum in the Breit frame, $p_{Z_b} < 0$ are considered the current region and this particles are created from the fragmentation of the struck quark. All the remaining particles are called the target region.

2.5.1 Kinematic selection

The data consists of measurements from the ZEUS detector, with $\sqrt{s} = 300$ GeV. The kinematic region for for data selection was restricted to the following requirement. The energy of the scattered lepton was required to be greater than, $E_{e^`} \geq 10\,\text{GeV}$. The virtuality

of the gauge boson, photon here, was required to be $10 < Q_e^2 < 100\,\mathrm{GeV}^2$ The maximum Q^2 is limited to 100 to avoid complexities related to combinatorial background involved in the $\phi-$meson reconstruction.

As mentioned earlier, the ϕ candidates are identified from the charged Kaon pairs. The restriction on the tracks for the transverse momenta to be $p_T > 200$ MeV in the lab frame. Of these tracks all the ones satisfying the conditions are selected and the invariant mass calculated. For a successful candidate to be selected the requirement is, $0.99 < M(K^+K^-) < 1.06$ GeV. The restrictions on the transverse momentum and pseudorapidity of the ϕ meson in the lab frame were , $p_T^\phi > 1.7\,\mathrm{GeV}$ and $-1.7 < \eta^\phi < 1.6$.

2.6 Dijet analysis

. Similar to the previous experiments, jet production studies[33] at HERA provide useful tools for testing the theory of perturbative QCD. Here, the data used is for cross-section calculations for inclusive jet in the photoproduction regime. It gives a great measure of the strong coupling constant and its scale dependence. The various jet observables that can be used for the testing purposes include inclusive-jet, dijet and multijet cross-sections. The data compared is for neutral current DIS and the main objective was to extract information regarding the parton distribution functions or the PDFs of the proton. The data used corresponds to singe differential cross-sections of inclusive jet production, reconstructed using the algorithms of k_T, anti-k_T and SIScone, as a function of pseudorapidity η^{jet} and transverse energy of the jet E_T^{jet}

For *ep* collisions, as is for Neutral Current DIS, the k_T algorithm is sufficient enough to provide infrared and collinear safe cross-section calculations irrespective of the order of pQCD considered. However, in order for better comparison with hadron hadron collision data like that from the LHC, the new algorithms of anti-k_T and SIScone algorithms were used in the analysis.

Jet production in photoproduction regime can be associated with two main process which are namely, the direct process and the resolved process. The main difference between these two processes is that in the direct process, the photon directly interacts with the parton inside the proton whereas in the resolved process, the photon acts as a source of parton, one

of which interacts with a parton in the proton.

2.6.1 Kinematic selection

The data used for analysis was taken during the running period of 2005-2007 at HERA with a centre-of-mass energy, $\sqrt{s} = 318$ GeV. The events were selected as per the requirement that at least a single jet of $E_T > 10$ GeV was required and pseudorapidity, $\eta < 2.5$.

Charged current DIS were eliminated by requiring that the total missing transverse momentum, p_T^{miss} should be smaller as compared to total transverse energy, E_T^{jet}, i e, $p_T^{miss}/\sqrt{E_T^{jet}} < 2\sqrt{\text{GeV}}$. The neutral current DIS was eliminated using the presence of a high energy scattered lepton. The cuts on the invariant mass of the hadronic centre-of-mass system or the γp rest system was $142 < W_{\gamma p} < 293$ GeV where is $W_{\gamma p} = \sqrt{sy}$, with y being the inelasticity. As the measurements are done in the photoproduction regime, the cut on virtuality is required to be $Q^2 < 1\,\text{GeV}^2$

2.7 Event Shape variables

The event shape variables data has been compared with those available from H1 collaboration as well as the ZEUS collaboration[36]. Similar to all the previous experiments, the hadronic final states produced in *ep* collisions through the neutral current channel offer an opportunity to understand hadronisation process as well as measuring the strong coupling constant for a wide range of virtuality Q^2. Event shape variables in DIS can be used for studying the coupling constant for the strong force provided that it is possible to separate the particles can be divided into the current fragmentation region produced from the struck quark and the target fragmentation region produced from the proton remnant. As was specified in one of the experiments earlier, the Breit frame is a useful frame of reference where there is minimal contamination of the current region from the target region. Also, in this reference frame, we work with a purely space-like photon with momenta $q_\gamma = 0, 0, 0, -Q$ which collides with a quark carrying longitudinal momentum $p_{qz} = Q/2$ The struck quark is scattered back into the current region or hemisphere with longitudinal momentum $p_{qz} = -Q/2$ whereas the proton remnant undergoes fragmentation in the opposite hemisphere also called the target

region. This obviously implies that the energy for fragmentation available in the current hemisphere is $Q/2$ and it also makes comparison with collision data involving e^+e^- easier.

For the data validation purposes, certain event shape variables are considered in current region of Breit frame which are namely, Thrust, Jet Broadening and Jet mass.

Kinematic selection

The data used for the comparison comes from both the H1 detector and ZEUS detector at HERA for the running period 1994-1996. with a centre-of-mass energy of 300 GeV. The kinematic range of the phase space region used for data collection are as follows. The virtuality of the collision was restricted to the range $7 < Q < 100$ GeV. The energy of the scattered lepton, here electron has to be greater than , $E_{e'} > 10$ GeV. The polar angle of the scattered lepton should lie between the angular range $157 < \theta_e < 173$ for the low Q sample and between $30 < \theta_e < 150$. The energy in the forward region should be $E > 0.5$ GeV, whereas by the forward region it is meant that the polar angular range $4 < \theta_h < 15$. This is to eliminate the diffractive events as well as the ones with large rapidity gaps. Finally, the energy in the Breit frame is required to be greater than 0.1 Q. Now, that the phase space selection is completed, its proper to look at the equations for the event shape variables under consideration. The event shape variables that are infrared safe namely, Thrust which is basically the measurement of the collimation of the longitudinal component of a hadronic system and Jet Broadening which is the transverse or the complimentary aspect of what is represented by Thrust. In its most general for, thrust and broadening can be defined by the following equations,

$$T = \frac{\Sigma_i |\bar{p}_i \cdot \bar{n}|}{\Sigma_i |\bar{p}_i|}$$

for thrust and

$$B = \frac{\Sigma_i |\bar{p}_i \times \bar{n}|}{\Sigma_i |\bar{p}_i|}$$

where the sum runs over all the final state hadrons and p_i is the momentum of that final state particle. For the purpose of data validation the event shape variables are considered with respect to particular axis are:

$$T_c = \max \frac{\Sigma_h |\mathbf{p}_h \cdot \mathbf{n}_T|}{\Sigma_h |\mathbf{p}_h|}$$

where the axis in consideration is the thrust axis which is the any direction, $\mathbf{n}_T$ that maximises the thrust value.

$$T_z = \frac{\Sigma_h |\mathbf{p}_h \cdot \mathbf{n}|}{\Sigma_h |\mathbf{p}_h|} = \frac{\Sigma_h |\mathbf{p}_{zh}|}{\Sigma_h |\mathbf{p}_h|}$$

where the measurement is with respect to the current hemisphere axis which is $\mathbf{n}$ is 0,0,-1 and is in the same direction as the virtual boson direction.

Chapter 3

Monte Carlo Generators

The majority of the project is carried out through the Pythia 8.303 (Pythia8), Herwig 7.2.1 (Herwig7) and Rivet 3.1.2 event generators.

3.1 Simulation

One of the basis for simulation of the required DIS and pp collision events are generators with QCD models and providing a complete information about the final state particles like its four momenta etc. These are theoretical predictions for the experimental data measured which are critical when it comes to the calculation involving phase space where such calculations are difficult or cannot be used. It is also useful when it comes to estimation of hadronic corrections. For the purpose of DIS simulation, the e-p collision event can described as having the following core components:

- Calculation of the leading order QCD matrix elements for hard process.
- Simulation the parton cascade from the initial state system.
- Simulation of final state with nonperturbative hadronization and parton cascade from final state system.

3.1.1 Pythia8

Pythia was originally an event generator intended for the purpose of hadron-hadron collision. Over the years it became more of a general purpose event generator that can be used to simulate any type of process ranging from lepton-lepton, hadron-hadron, lepton-hadron and even heavy ion collisions. For this project however the main focus was on proper simulation and comparison of data from the HERA experiment. As such the important factors associated with the simulation of an event at HERA are:

? Pythia being a general purpose generator is unlike other generators like RAPGAP which has dedicated development for DIS. As parton showers are the core component when it comes to this simulation, it is useful to mention the two type of approaches implemented in Pythia8 for it, which are namely default shower and new dipole recoil. With respect to the incoming parton, both approaches have the same process but the main difference lies in the fact the outgoing parton has $k_T \neq 0$ in the dipole recoil approach.

3.1.2 Rivet

Rivet is the recommended analysis framework for EIC data analysis. It has been proposed to used for prediction and data estimation mainly due to one it's capability to adhere to physical and realistic measures. It is fully written in C++ and one of it's main objectives is to provide a easy and direct method for comparison with results from other experiments. The histograms can be set to have the same conditions as the limits as the experimental results used as a reference. As such it also a tool that fits the objectives of this project perfectly.

Chapter 4

Methodology

The main tools used are Pythia 8.303 (Pythia8), Herwig 7.2.1 (Herwig7) and Rivet 3.1.2 (Rivet). The general workflow throughout the project involved the following steps:

1. Configuring the input card with the necessary set of correction to the phase space cuts involved in the corresponding experiment.

2. generating the events and piping it

3. Read the events from the pipe using rivet.

4. Merging the simulate data and the experimental data sets for comparison using the `make-plots` script.

Most of the necessary tools, except for pythia8, was used through the docker images available at dockerhub website. These were also later used as singularity images. For Herwig and rivet, most of the requirement was satisfied through running the docker container in interactive mode and accessing the script files. The instruction for setting up and running these files were provided from the EIC Workshop in August. For the remaining two softwares, pythia and ROOT, instruction for getting started was taken from their home website.

4.1 Pythia 8.303

Since pythia comes with a inbuilt library of templates that can be used for simulating various collision experiments, it was relatively easy to understand the general syntax and overall workflow. However, the run card with the parameter settings, process selection and phase space cuts for was not present for DIS simulation and created with the help of reference from the Pythia tutorial given during the EIC workshop. The following package dependencies were required for running pythia for the purposes of the project,:

- HepMC2 or HepMC3
- ROOT

Although rivet was the main analysis tool recommended for the project, ROOT was used in scenarios were scatter plots were involved due to difficulty for generating the same with rivet. It was also useful for generating the 3D plots as can be seen for the results for event shape variables. The general workflow for pythia consisted of:

1. Configuring the run card with the appropriate phase space cuts and centre-of-mass energy of collision. Process selection was only changed when measurements in the photoproduction regime was simulated.
2. Compile an executable to read from the input card and writing the outputs to hepmc file
3. running the executable along with the proper names of the input run card and output hepmc file.
4. whenever restrictions due to storage space was a problem, the events were written into a pipe from which it was simultaneously read by rivet.

4.2 Herwig 7.2.1

During the initial stages of the project, Herwig was easier to work with as compared to pythia because of the integration with the Powheg and MC@NLO packages to generate not

only Leading order (LO) but Next-to-leading order (NLO) events. This made it easier to compare the effects of the NLO effects and how it effected as opposed to there being only LO events. Also, the switching on of the NLO calculations was rather straightforward with only including certain string in the event run card. Apart from this,

1. Herwig, similar to pythia comes with a wide set of example templates with preselected settings and parameter switches. This not only makes the whole simulation process convenient and faster, it also provides guidance to beginners on the necessary set of settings for a particular experiment.

2. The various run cards and default settings are all customizable and can be changed Accordingly as per specific experiment of interest. However, most of the default settings that come with the templates for DIS and the parameter values were sufficient to generate a closely resembling event sample.

3. For this project one of the important settings was modifying the phase space cuts provided along with changing the centre-of-mass energy of the collision (or the individual beam energies separately) which alternates between 300 GeV and 318 GeV depending on the running period of the experiment at HERA being used for comparison.

4. Similar to pythia, the files can be output as HepMC files. During the course of the project, storage issues wasn't of much concern to the relatively small size of DIS events written to HepMC file. In comparison, the file size of pp collision events from LHC consumes more space

5. For the project, the events were written to pipe, from which the events were read using rivet.

6. Apart from that, just like that in pythia, there's direct integration with rivet. As such, the particular analysis name of interest can be called from the event run card. This option was not considered for the project due to custom analysis shared libraries that were built on called during runtime from a particular working directory.

7. Due to ease of setting up the NLO+ parton shower matching within Herwig it was very convenient to observe the NLO corrections.

4.3 Rivet 3.1.2

One of the important aspects of this project is the reliable comparison with experimental data where by stacked plots on the same canvas is an essential aspect for quick and comprehensive interpretation of the results. Taking the data to be used for comparison has several options. Two of the viable options were in terms of importing the data as **ROOT** or **YODA** files. The former is special format exclusive to the ROOT analysis framework software by CERN and the other is a plain text file with a special syntax read by Rivet. These file type options are available in the hepdata website, which is a repository for almost all the high energy related experiment results. For this specific project, as per the instructions given to the Software Working Group of EIC India, :

1. the file format of YODA was chosen as the default analysis tool was assigned to be Rivet.

2. YODA files could be easily referenced in a particular analysis by changing to the same filename.

3. the references to be used while creating histograms with respect to the axis limits are given by specific codes defined according to the ID of the plots in the YODA file

4. The customization of the histogram parameters was done through the PLOT file with the same name as the analysis and the YODA file. This method however has its own limitations with respect to stacking of graphs and was only used in the initial stages of the project.

5. For plotting data along with the simulated data from multiple generators, the make-plots script that came with the Rivet package was used. It allowed for not only stacking the histograms but also for providing more information in the plots and more features regarding the plotting styles.

Now that the data for comparison was obtained in a usable manner, the important part is to carry out the analysis. In rivet analysis is carried out using various project classes which helps to extract particular observables from the event along with providing other useful tool such as 'boosting a specific reference frame' so the corresponding quantity can be obtained

in that frame, the use case here being the use of the variables in Breit frame and hadronic centre-of-mass frame or $\gamma - p$ rest system.

In the most general form, rivet analysis procedure can be divided into three stages which are namely, Initialization, Analysis and Finalization in the respective sections `init`, `analyze` and finalize. Majority of the analysis is carried out in the analysis stage with the initialization stage involved in creating the histogram and in the initialization of the necessary projections necessary for working the required observables of an event. It here that the init method is called and it is down only once. The next method called is the analyze method where the analysis and extraction of variables actually happen. This method is run for each and every event. After finishing up of these runs, the finalize method is called one final time at the end to carry out scaling and normalization of the histograms. There are commands to create a rivet analysis template with the classes and methods already defined inside it.

After the analysis file is configured, the shared object file is build and analysis is run for the required event sample. Rivet reads HepMC files which is how the analysis was mainly carried out during this project. During multiple file analysis run in parallel, the events were read from a pipe in order to overcome storage limitations. Tools like screen and `tmux` were also used for managing background process.

During the final stages of the project, one of the server farms at Thomas Jefferson Laboratory National Acceleration facility (Jlab) was used to carry out the project due to technical difficulties with the personal system. The jlab system was accessed through protocol of ssh. The details regarding this whole procedure has been documented as also a part of the project in a personal github repository, which is to be used for future references for beginners.

Chapter 5

Results

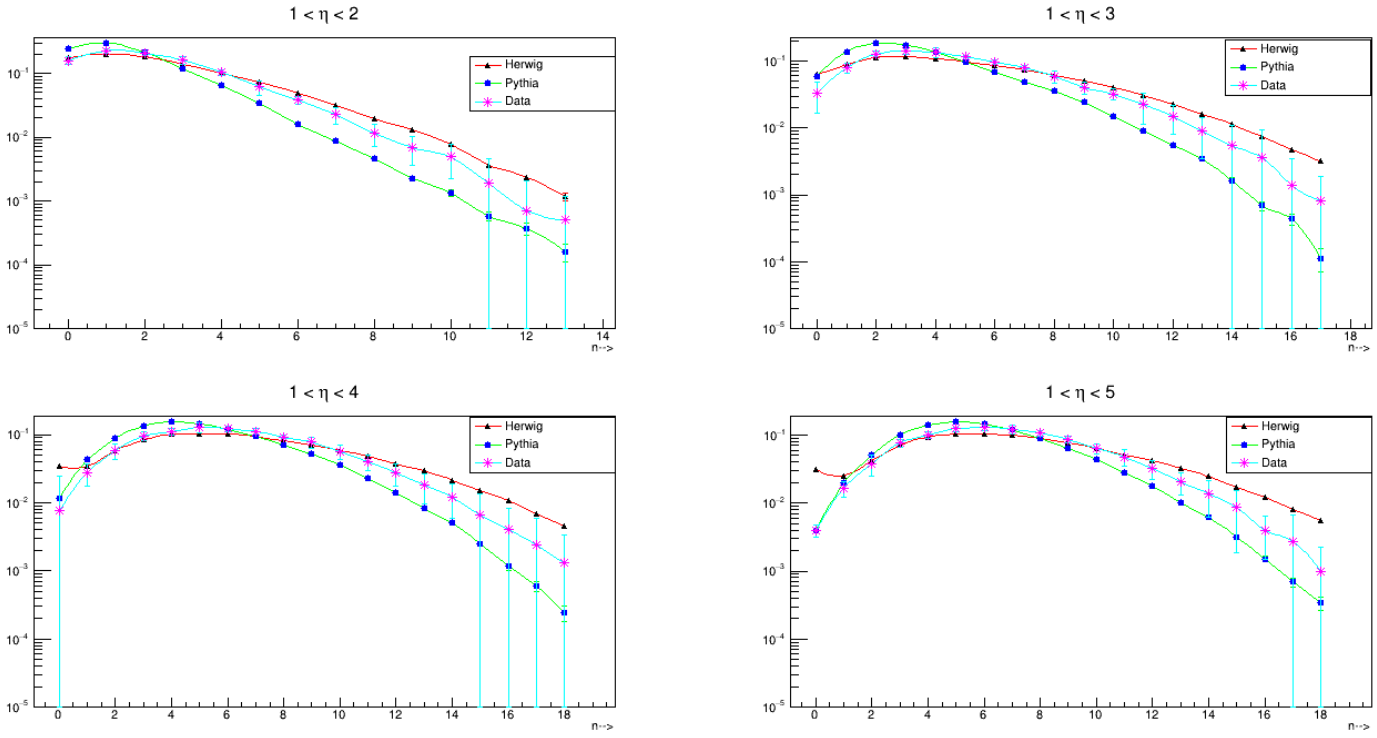
The data comparison results available are presented below with a brief mention of the phase space selection used for measurements. Due to vast number of plots, only a representative sample of plots are presented in this chapter and the remainder of the plots are provided towards the end of the thesis.

5.1 Charged Particle Multiplicity

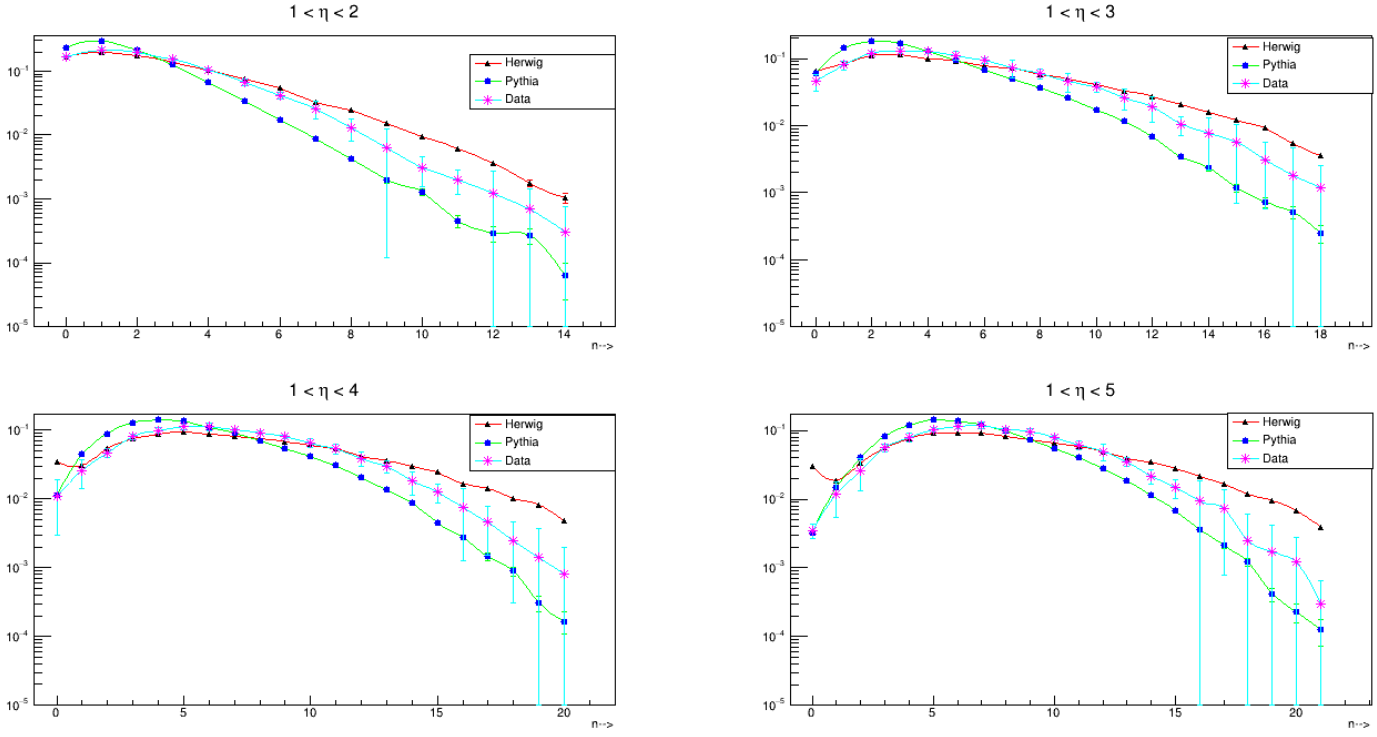
The multiplicity distribution has been measured, in the kinematic regions in W and Q^2 shown in the following figures, for charged particles with pseudorapidity in the domains $1 \leq \eta \leq \eta_c$ with $\eta_c = 2, 3, 4, 5$ and in intervals of unit pseudorapidity centered at $\eta = 2.5, 3.5, 4.5$, as well as for the full current hemisphere defined as the domain $\eta > 0$

Table 5.1: Phase space selection

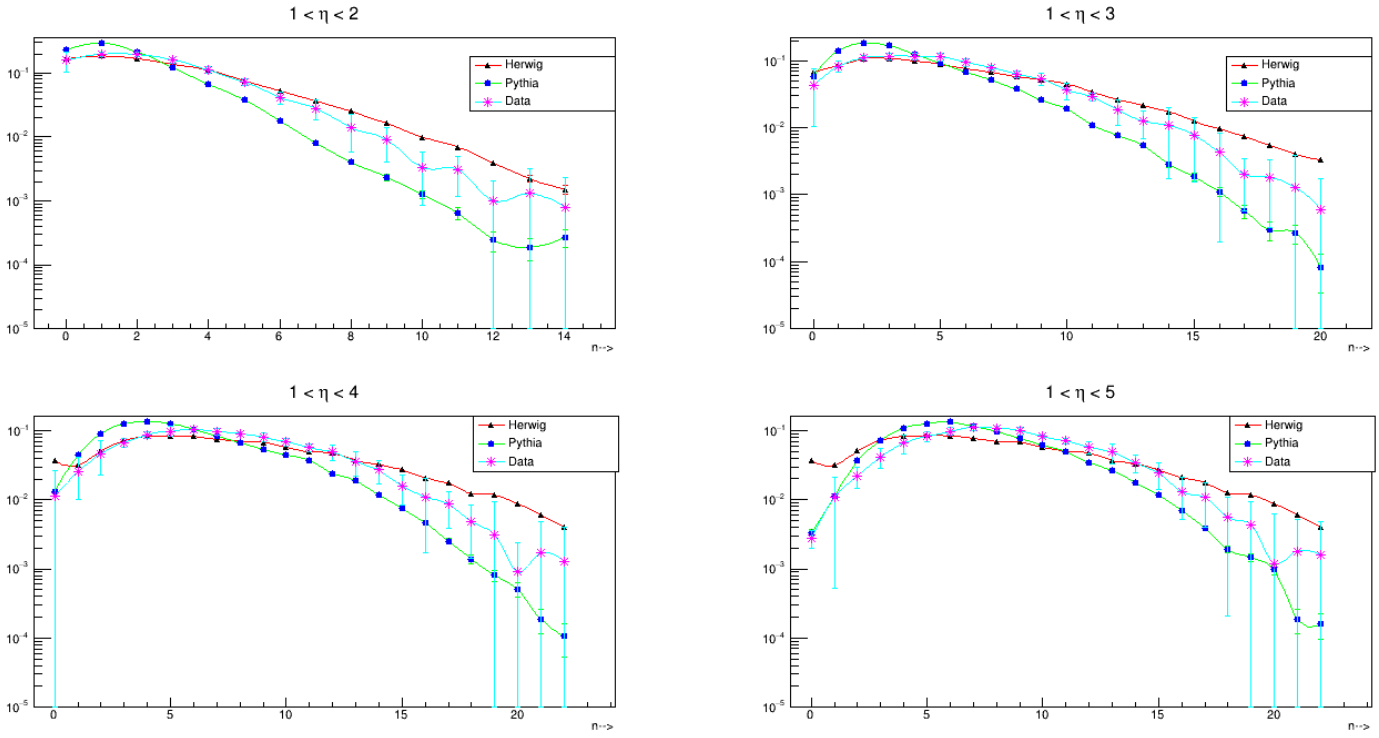
Range in	Single differential cross-section
$\sqrt{s}$ (GeV)	300
Q^2 (GeV^2)	10-1000
W (GeV)	80-220
$\|\eta\|$	0-3



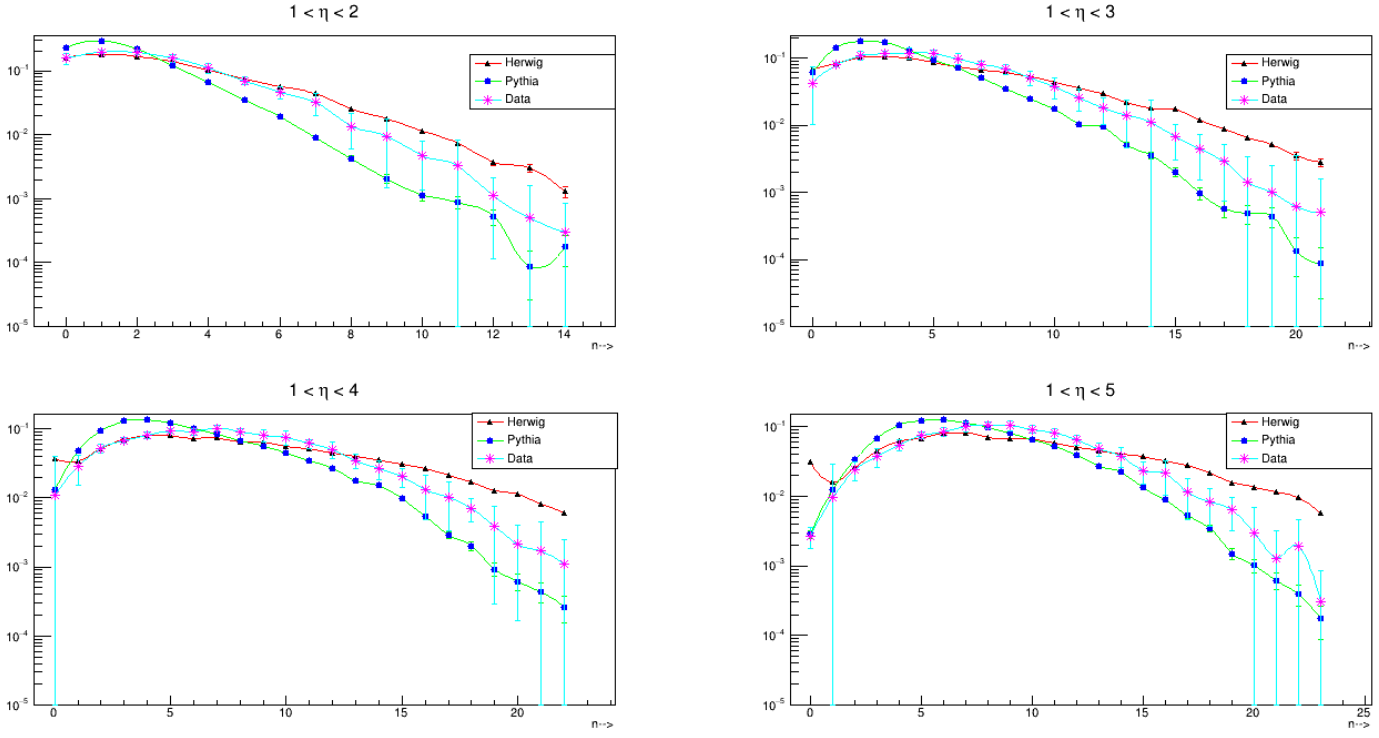
(a) 300k events are generated by Herwig7 and 1M events by pythia8. Charged Particle Multiplicity distribution for 80 GeV $< W <$ 115 GeV



(b) Charged Particle Multiplicity distribution for 115 GeV $< W <$ 150 GeV



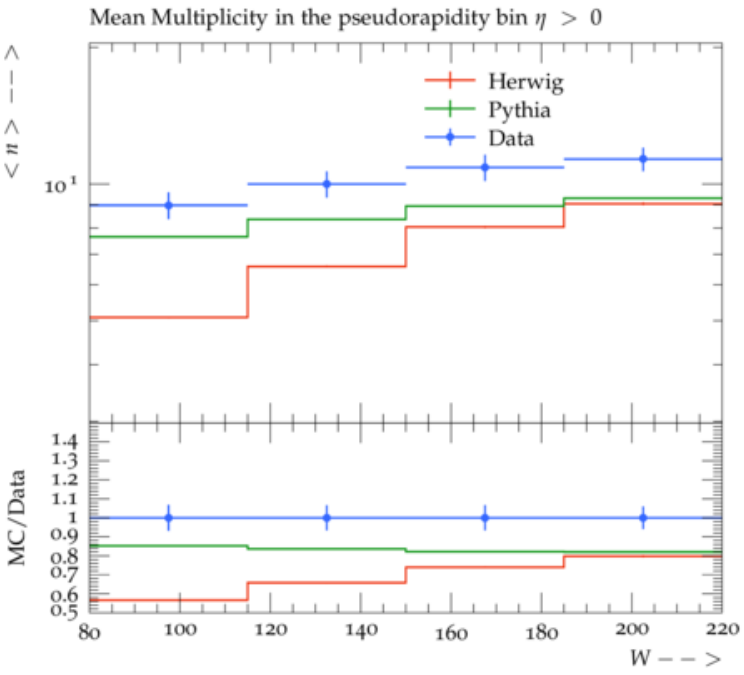
(a) Charged Particle Multiplicity distribution for 150 GeV $< W <$ 185 GeV



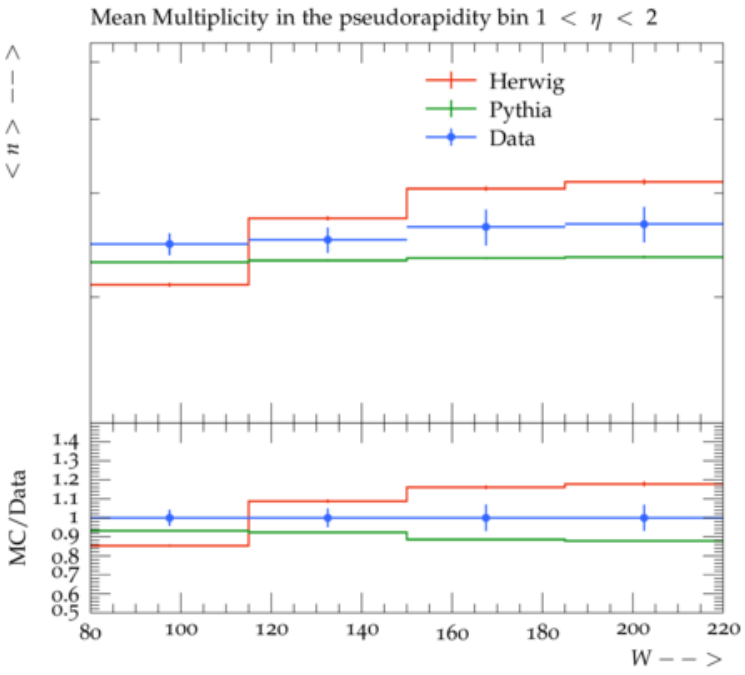
(b) Charged Particle Multiplicity distribution for 185 GeV $< W <$ 220 GeV

The results show that the reproduction of the kinematic variables are almost accurate in the low multiplicity regions whereas the comparison becomes worse in the higher multiplicity regions. The reason for this could be fluctuations from the uncertainties in the statistics. However, one thing that could be definitely concluded from this is that Herwig analysis is better at estimating the multiplicity distribution at lower regions of the multiplicity and the estimation comes of at higher regions.

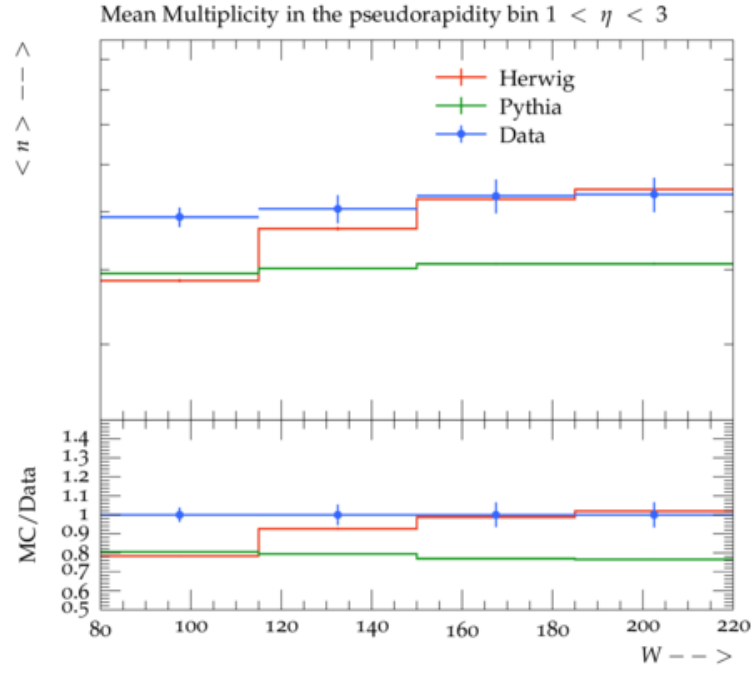
5.1.1 Mean Multiplicity, $\langle n \rangle$



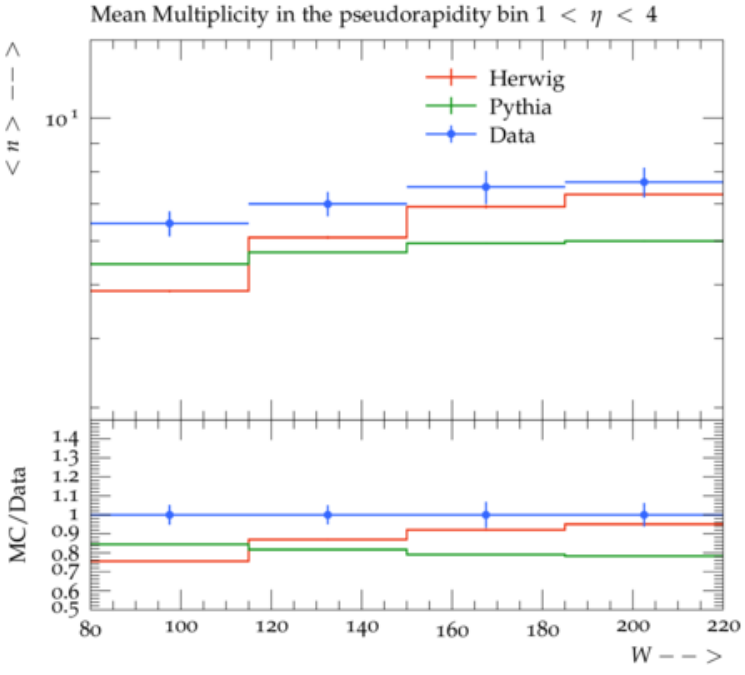
(a)



(b)



(c)



(d)

Figure 5.3: Mean of Charged Particle Multiplicity distribution

5.1.2 Dispersion, D_2 and higher orders

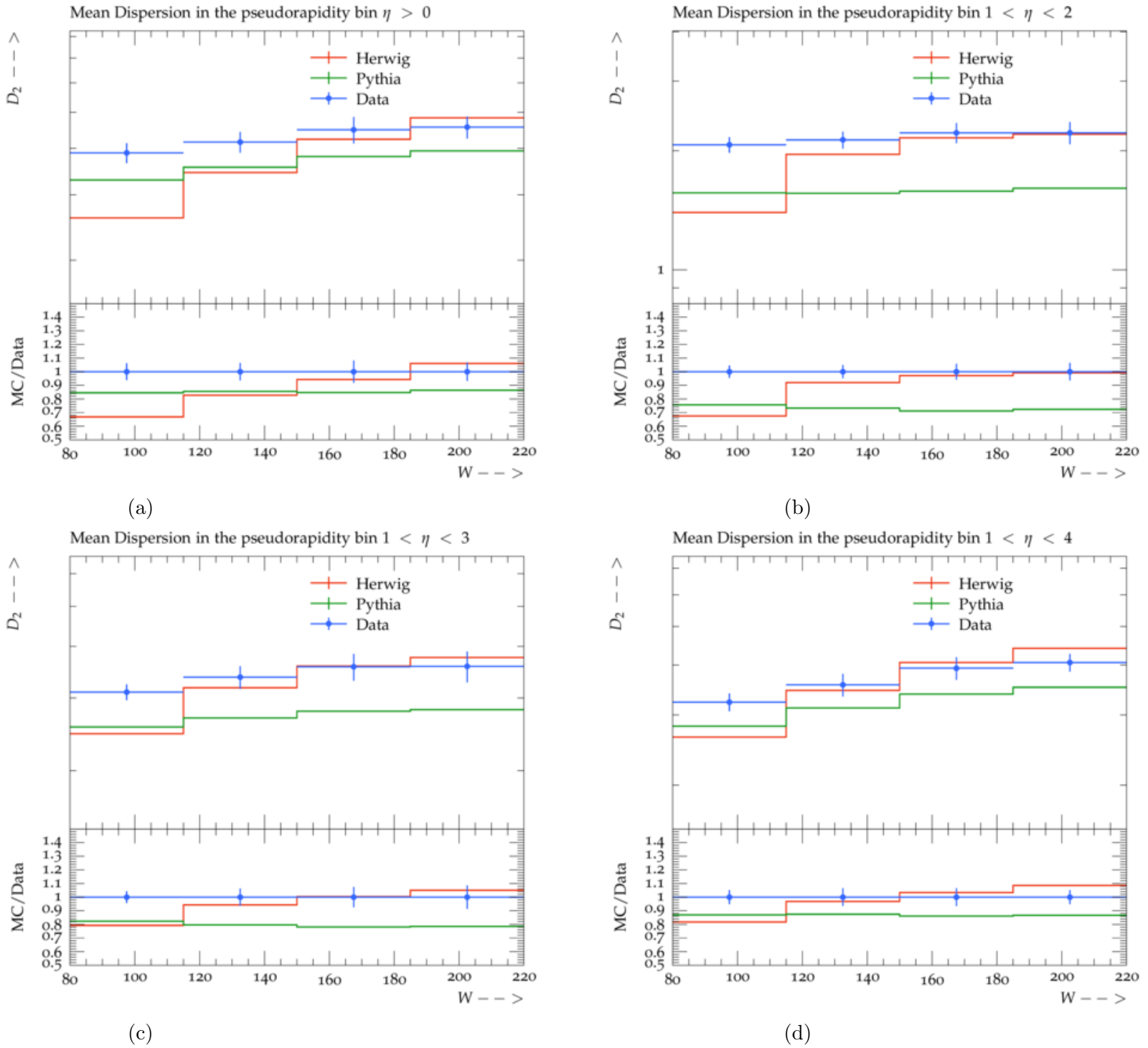


Figure 5.4: Dispersion of Charged Particle Multiplicity distribution

5.1.3 Generalised moment, C_2

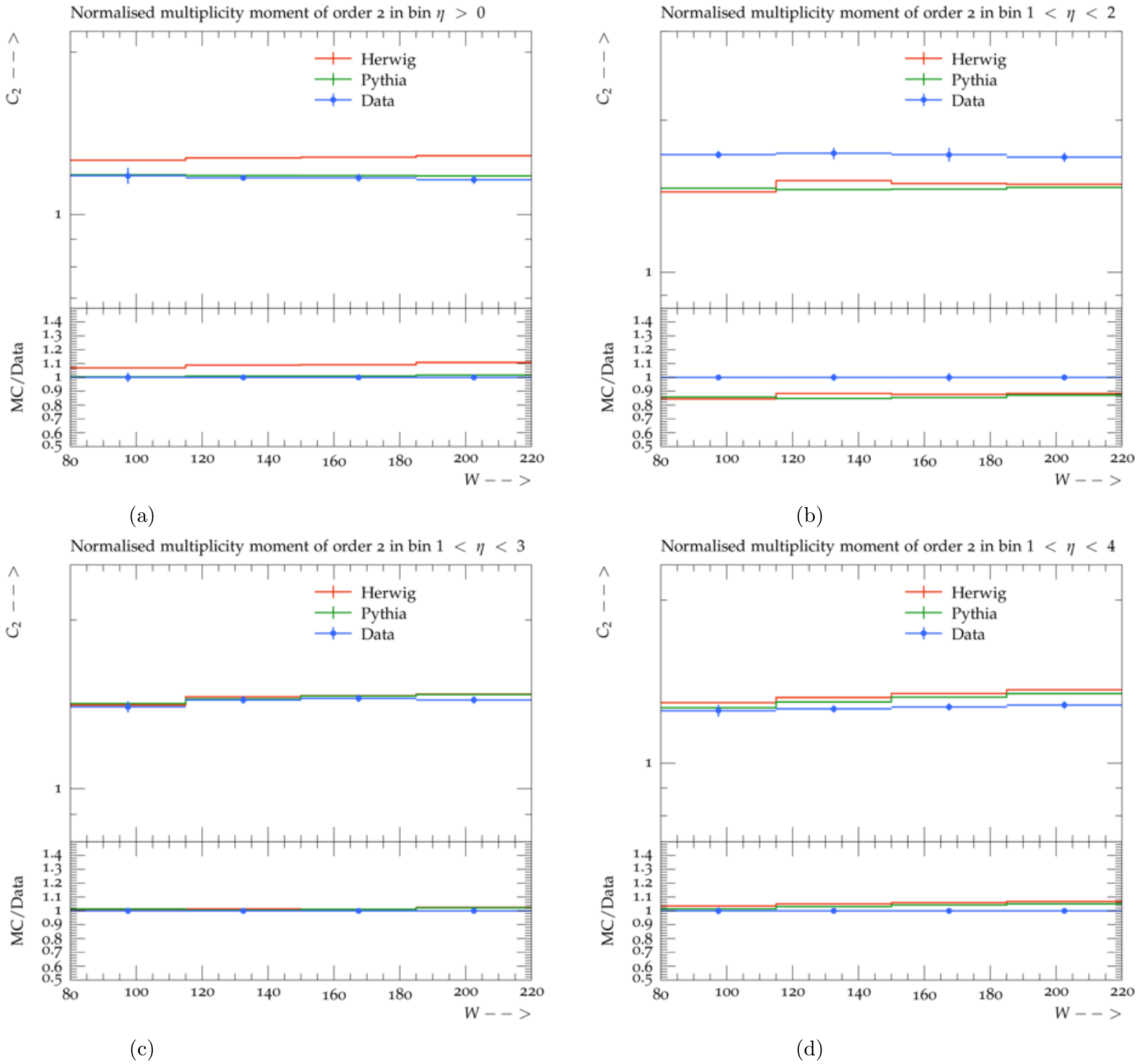


Figure 5.5: Generalized moment of Charged Particle Multiplicity distribution

Pythia does a good job in estimating the nature of the distribution but it fails when it comes to estimating the actual values quantitatively rather than qualitatively. Both the generators perform accurately at higher pseudorapidity regions. Herwig comes to accurate estimations for higher hadronic system invariant mass energies for all ranges of pseudorapidity but falls off when the pseudorapidity bins are low.

5.2 Transverse momentum spectra

These results also use ROOT like the previous due to a technical difficulty while analysing with rivet. Transverse momentum spectra of charged particles in deep inelastic scattering are measured as a function of x_B and Q^2 in the current and the central fragmentation regions. 300k events are generated by Herwig7 and 1M events by pythia8.

Table 5.2: Phase space selection

Variable	Limits
$\sqrt{s}$ (GeV)	300
$E_{e^{\prime}}$	$\geq$ 10 GeV
Q^2 (GeV^2)	5-50
$E_{forward}$	$>$ 0.5
$Bjorken\, x$	(0.0001,0.01)

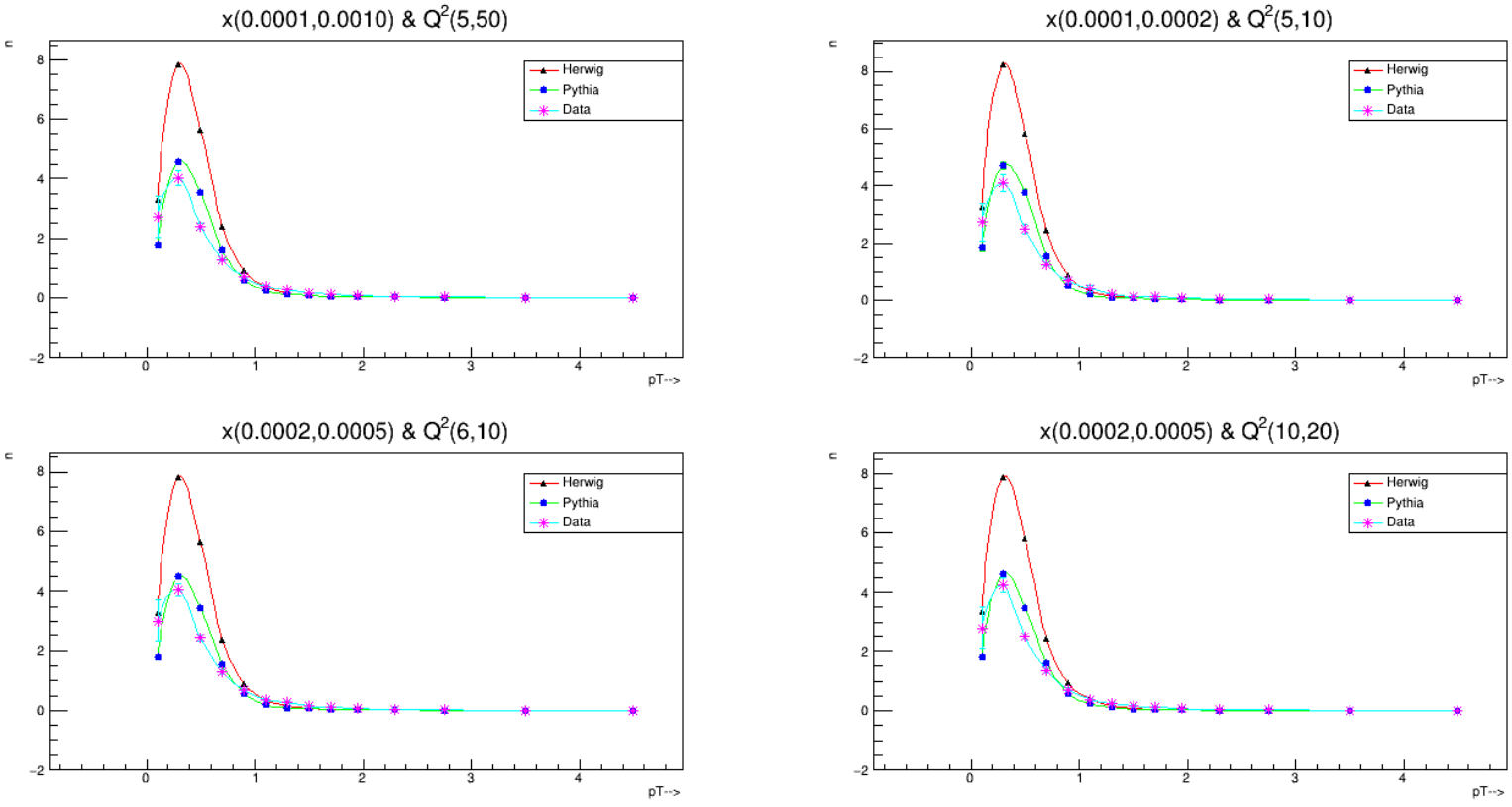


Figure 5.6: Charged Particle transverse momenta in respective bins of Q^2 and Bjorken scaling 'x'

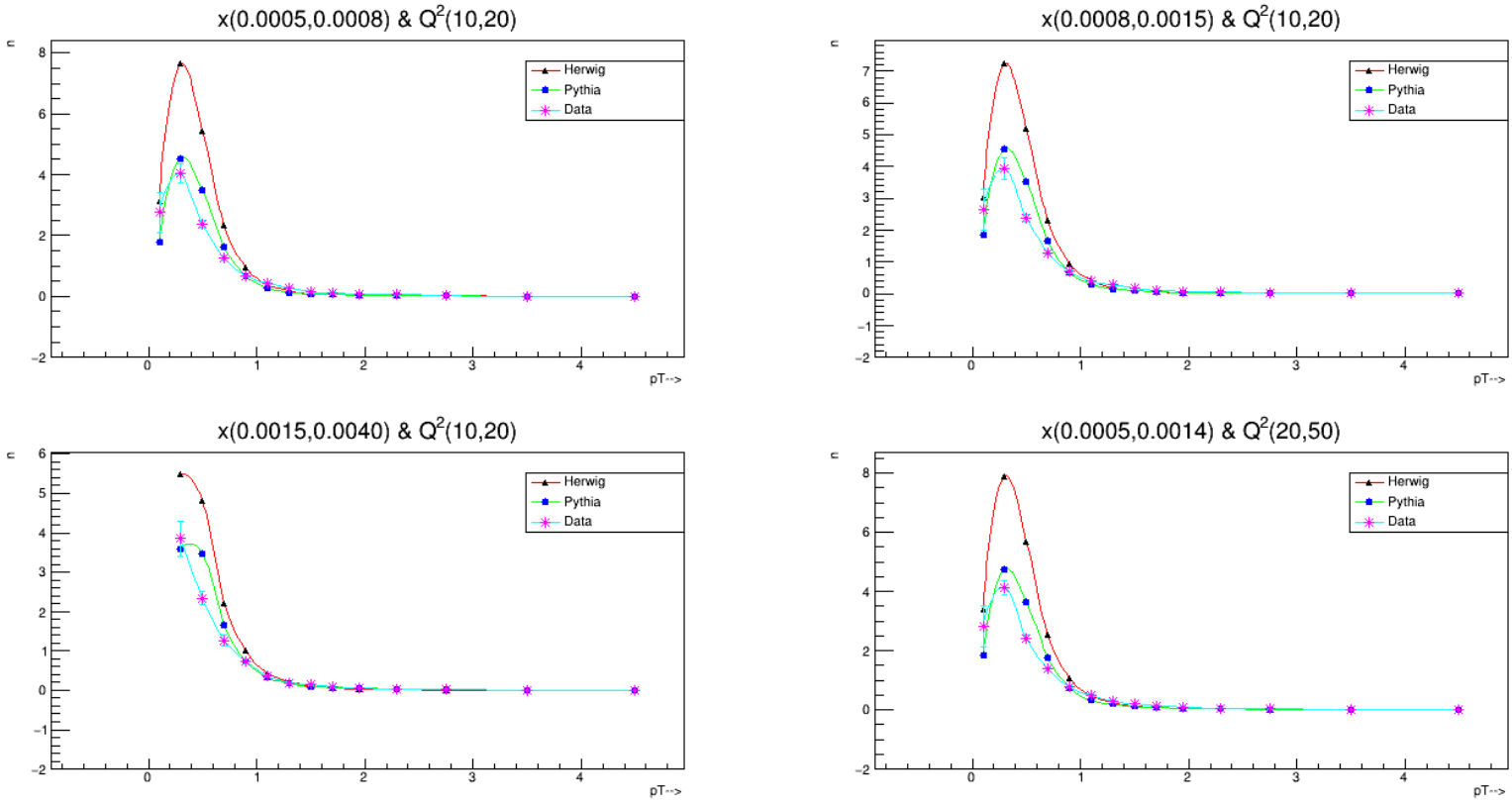


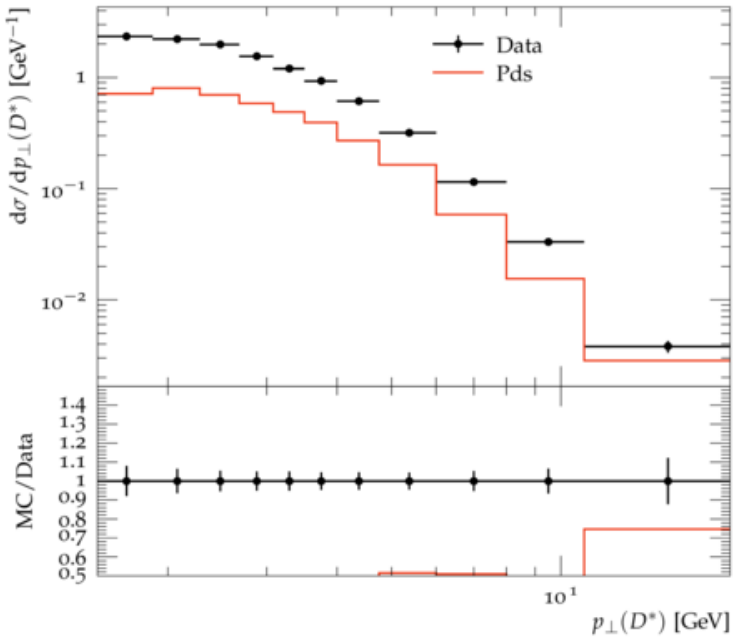
Figure 5.7

5.3 $D^*\pm$ production cross sections.

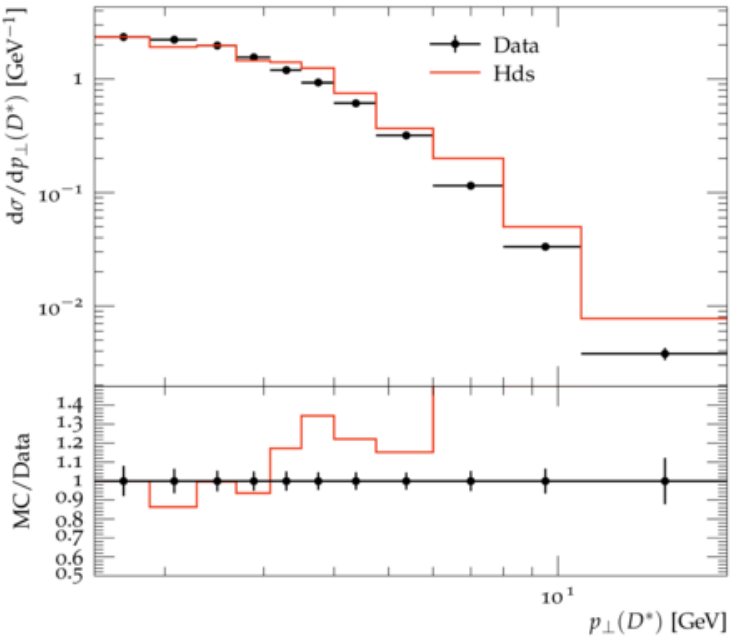
Table 5.3: Phase space selection

Range in	Single differential cross-section
Q^2 (GeV^2)	10-1000
y	0.02 - 0.7
$p_T(D*)(GeV)$	>1.5
$\|\eta(D*)\|$	<1.5

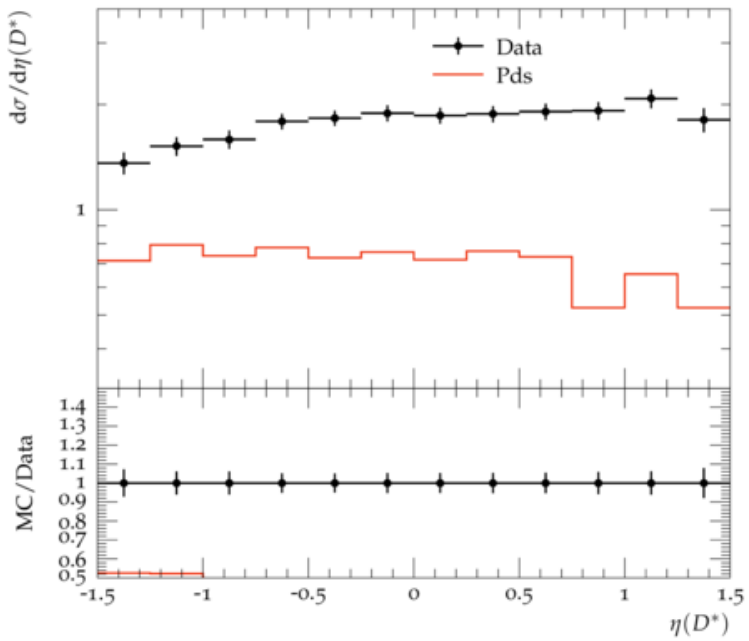
300k events are generated by Herwig7 and 1M events by pythia8.



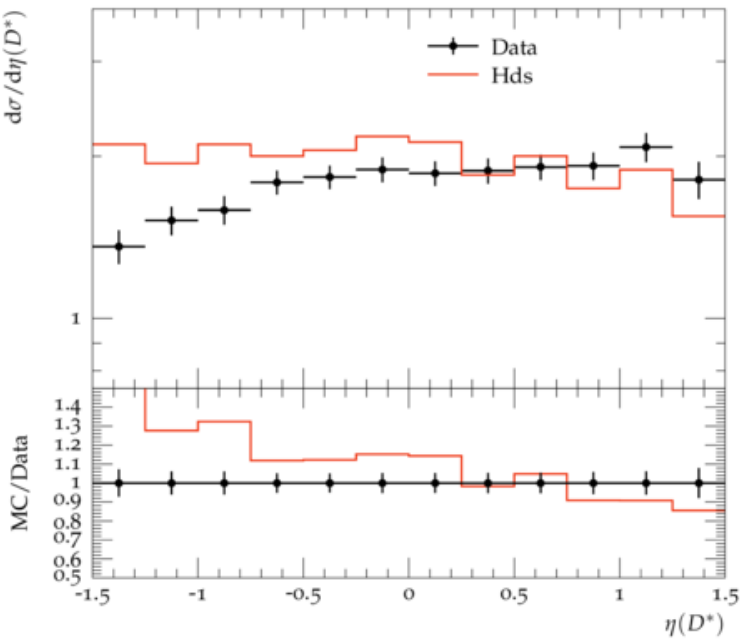
(a)



(b)

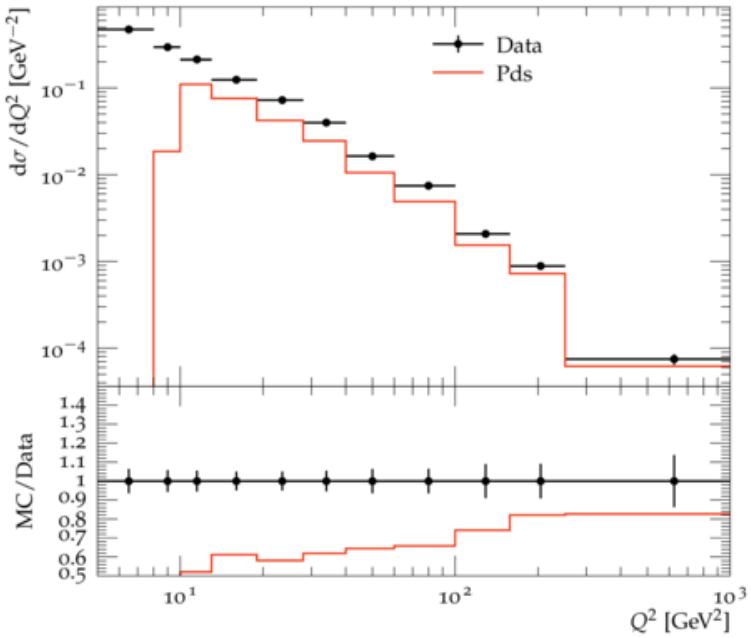


(c)

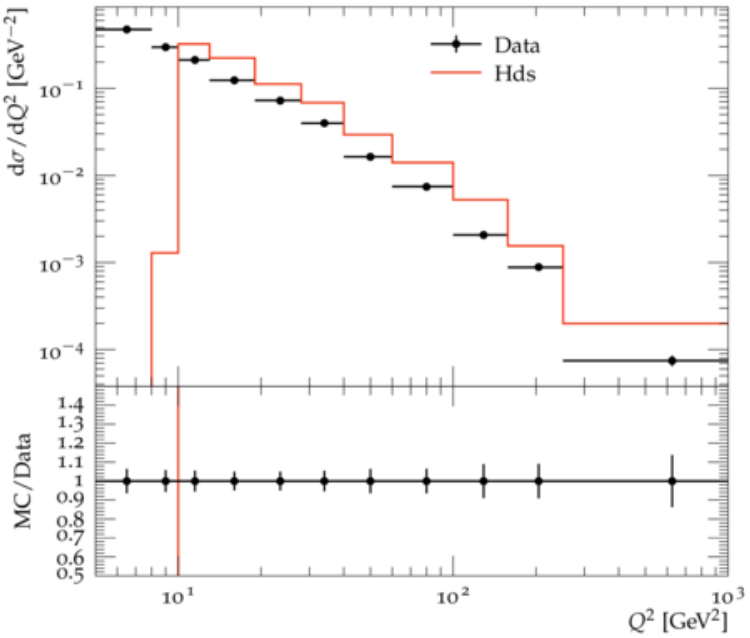


(d)

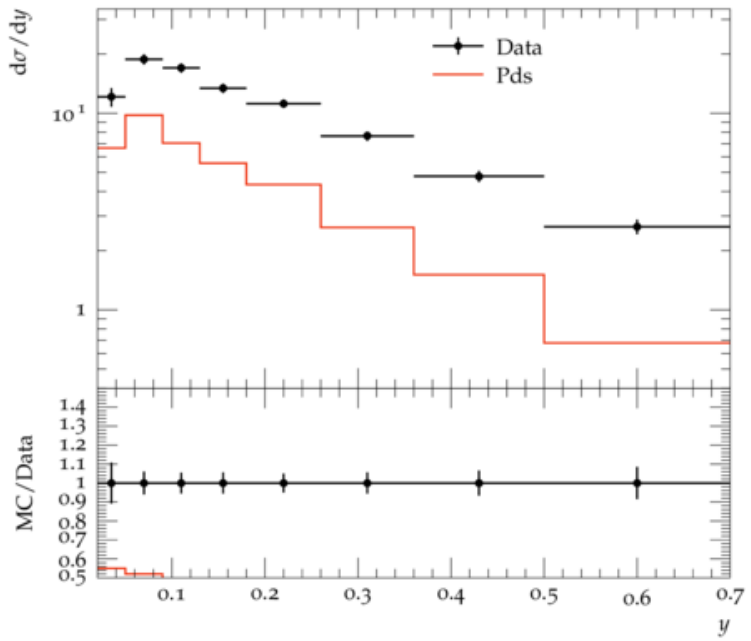
Figure 5.8: The data corresponds to single differential cross-section for $D* \pm$ production with respect to the kinematic variables, transverse momentum, $p_T(D*)$, pseudorapidity, $\eta(D*)$ and the elasticity, $z(D*) = (E(D*) - p_Z(D*))/(2E)$ both measured in the lab frame and global variables Q^2 and y. Here we have ones from Pythia 8.303(Left) and Herwig 7.2.1(right).



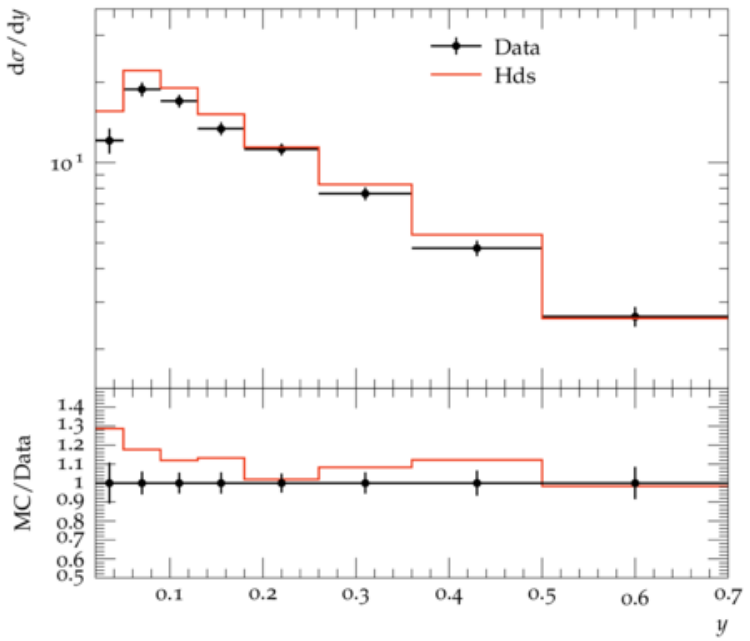
(a)



(b)



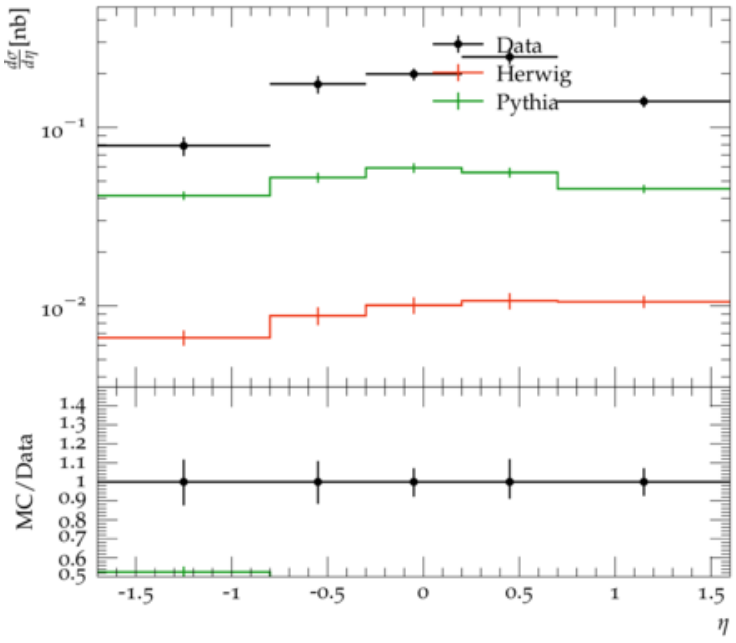
(c)



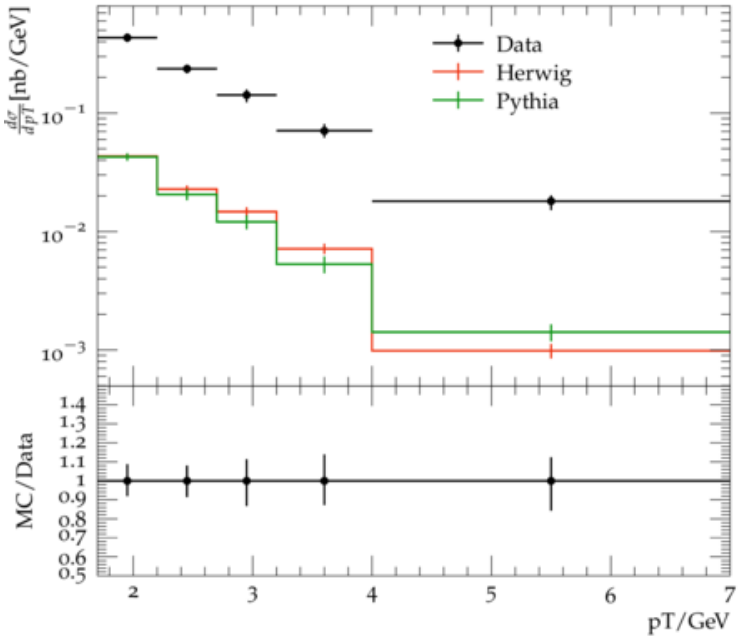
(d)

Figure 5.9

5.4 ϕ-meson production

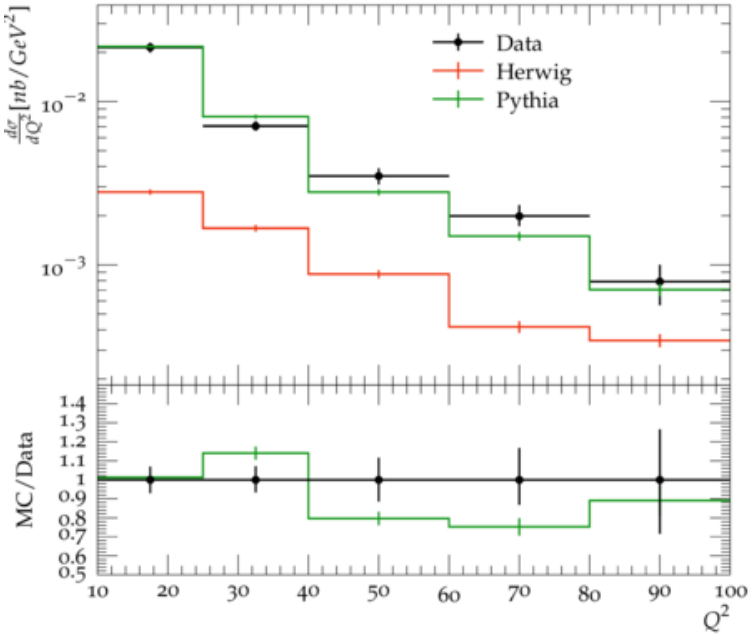


(a)

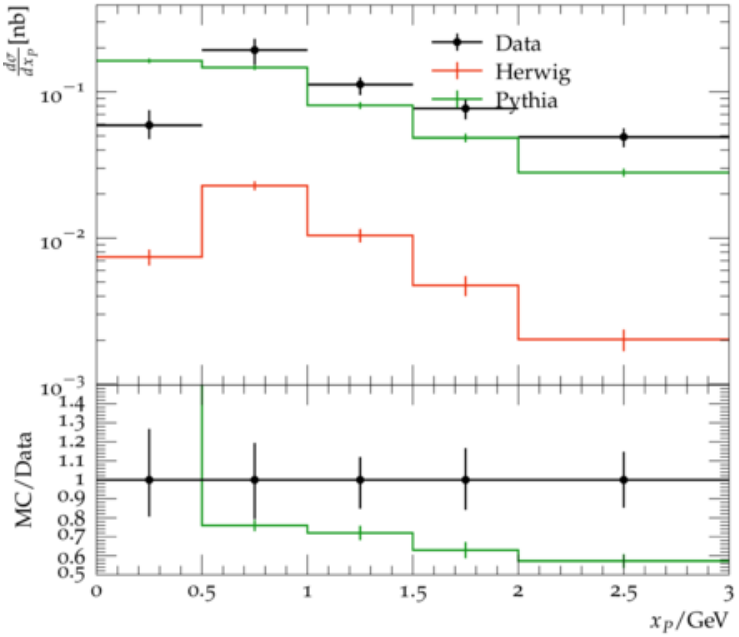


(b)

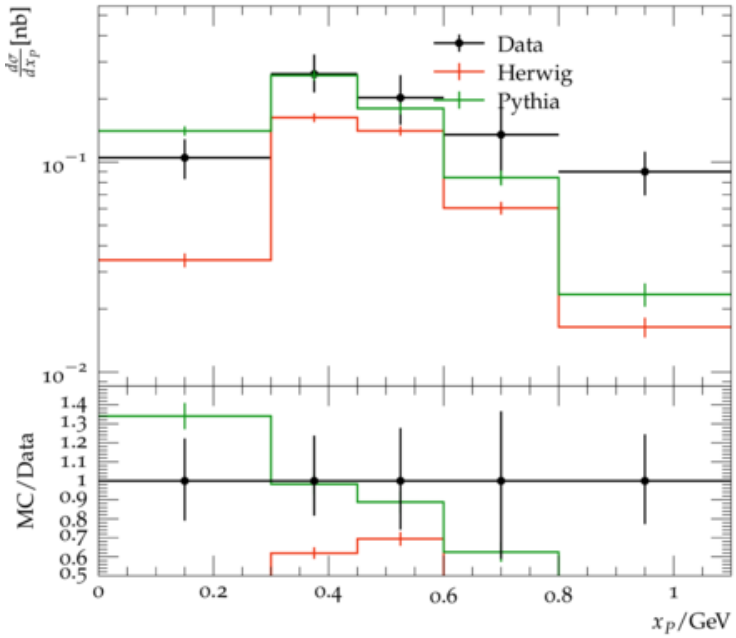
Figure 5.10



(a)

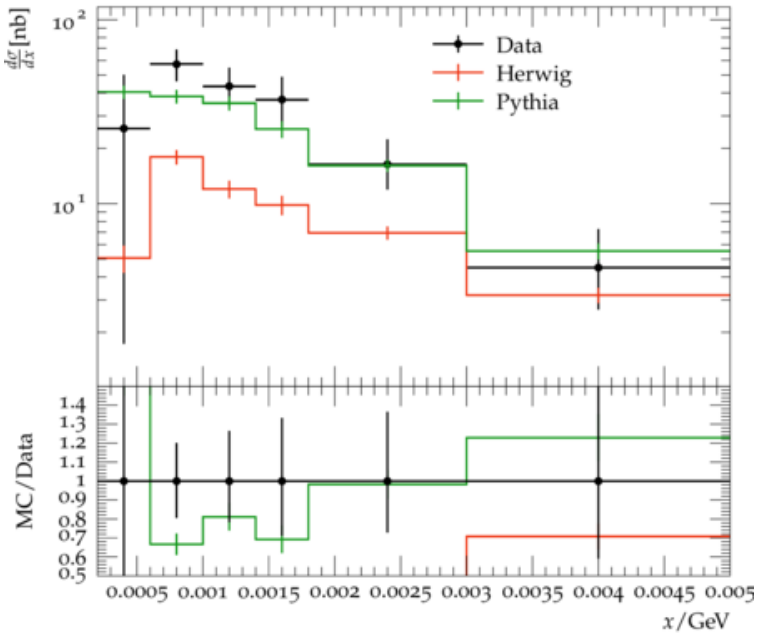


(b)

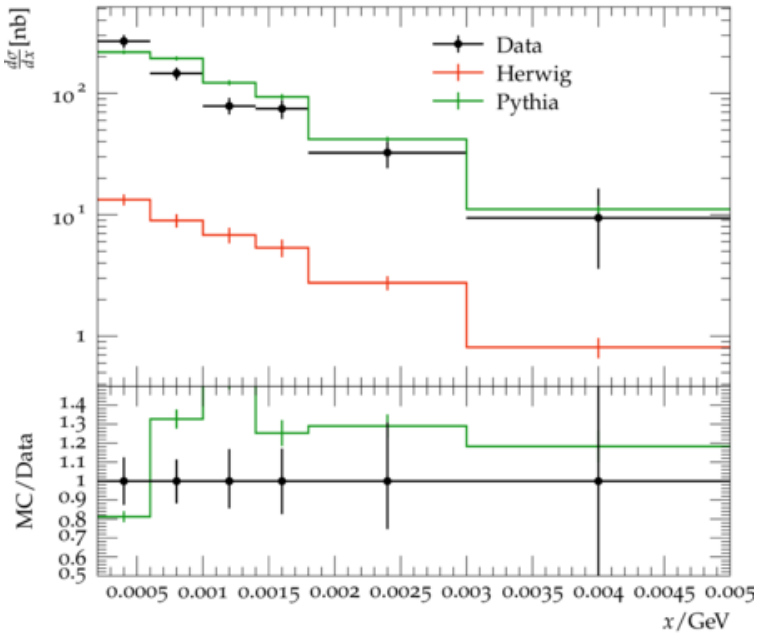


(c)

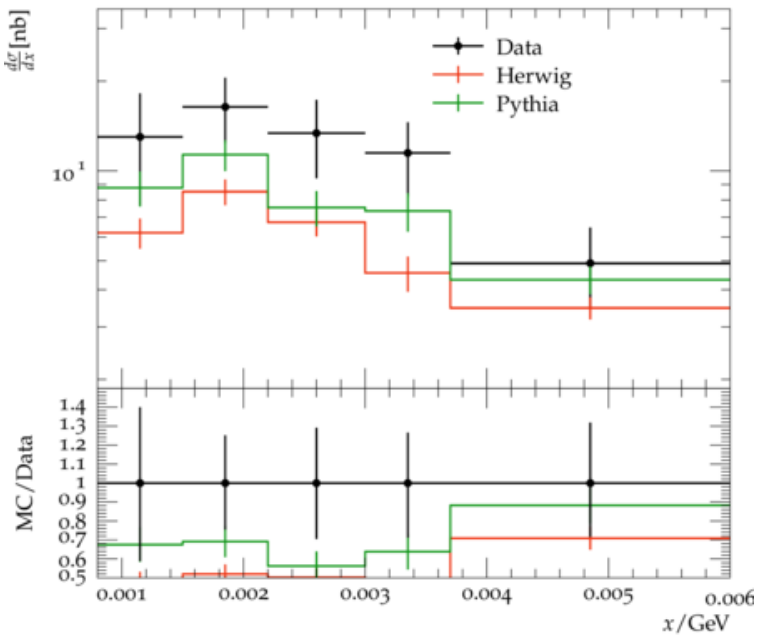
Figure 5.11: Differential ϕ-meson cross sections as functions of p_T^ϕ, η^ϕ , Q^2 and scaled momentum, x_P, compared to Herwig 7.2 and Pythia 8.3 simulations. The cross sections were measured in the kinematic region $10 < Q^2 < 100$ GeV^2 , $2 \cdot 10^4 < x < 10^2$, $1.7 < p_T^\phi < 7$ GeV and $1.7 < \eta_\phi < 1.6$.. The predictions are shown for $\lambda_s = 0.22$.



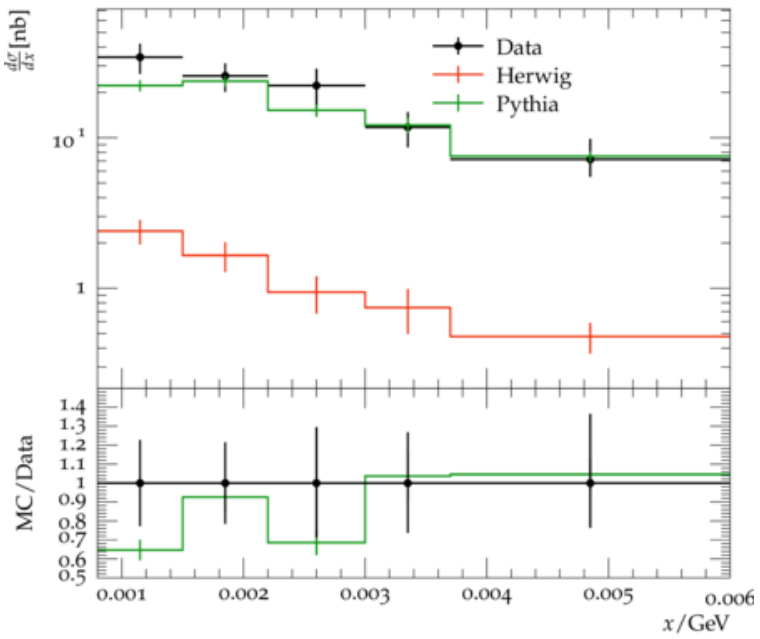
(a)



(b)



(c)



(d)

Figure 5.12: The inclusive cross-sections as a function of bjorken x for two Q^2 intervals, ($35 < Q^2 < 100\ GeV^2$) and ($10 < Q^2 < 35\ GeV^2$), for the current, a)-b), and the target, c)-d), regions of the Breit frame compared to the LO,NLO and LO+ Jet Matching for Pythia and Herwig with predictions for $\lambda_{\hat{s}}$= 0.22.

Table 5.4: Phase space selection

Variable	Limits
$\sqrt{s}$ (GeV)	300
$E_{e^{`}}$	$\geq$ 10 GeV
Q^2 (GeV^2)	10-100
p_T^{ϕ}	$>$ 1.7 GeV
$\lvert\eta(\phi)\rvert$	(-1.7,1.6)
$M(K^+K^-)$ GeV	(0.99 , 1.06) GeV

5.5 Event Shape variables

1. Thrust

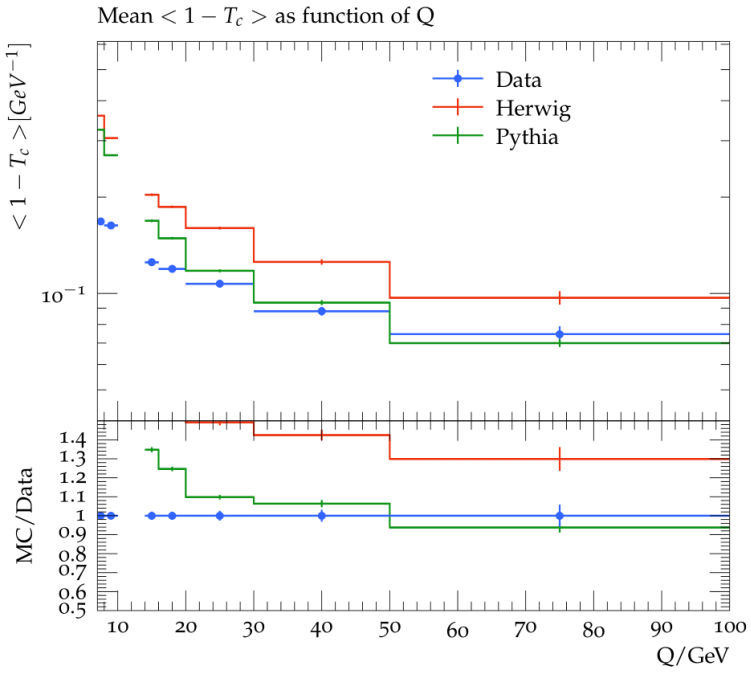
$$T = \frac{\Sigma_i |\bar{p_i} \cdot \bar{n}|}{\Sigma_i |\bar{p_i}|}$$

2. Jet broadening

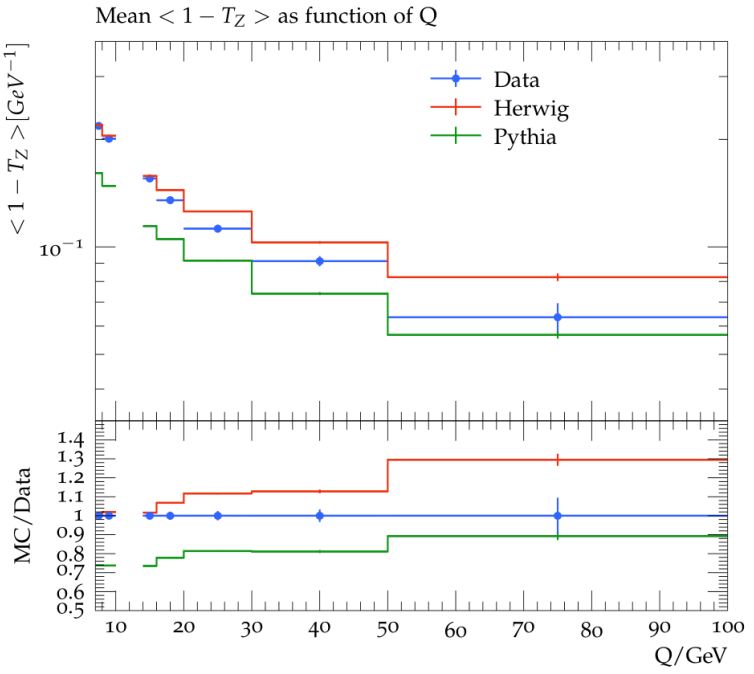
$$B = \frac{\Sigma_i |\bar{p_i} \times \bar{n}|}{\Sigma_i |\bar{p_i}|}$$

Table 5.5: Phase space selection

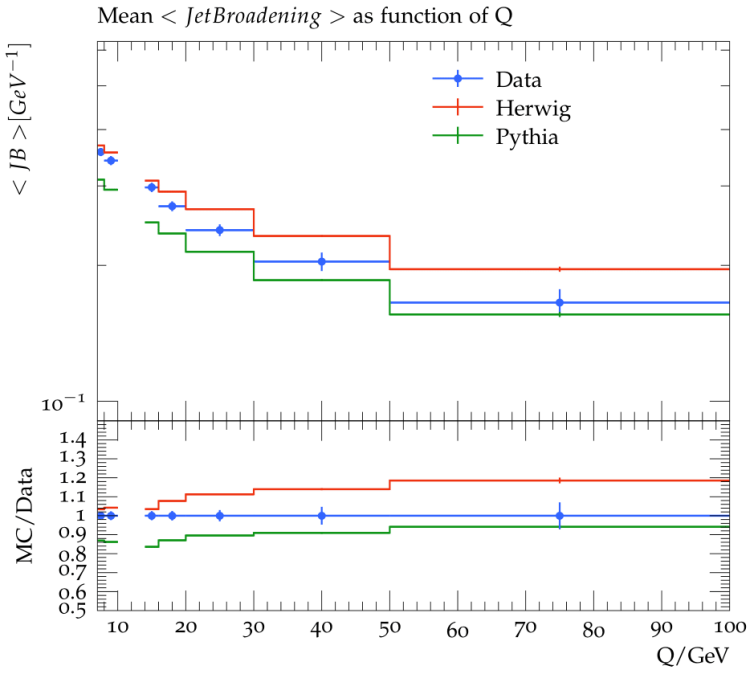
Variable	Limits
$\sqrt{s}$ (GeV)	300
$E_{e^{`}}$	$\geq$ 10 GeV
Q^2 (GeV^2)	49-10000
$E_forward$	$>$ 0.5
y	(0.05,0.8)



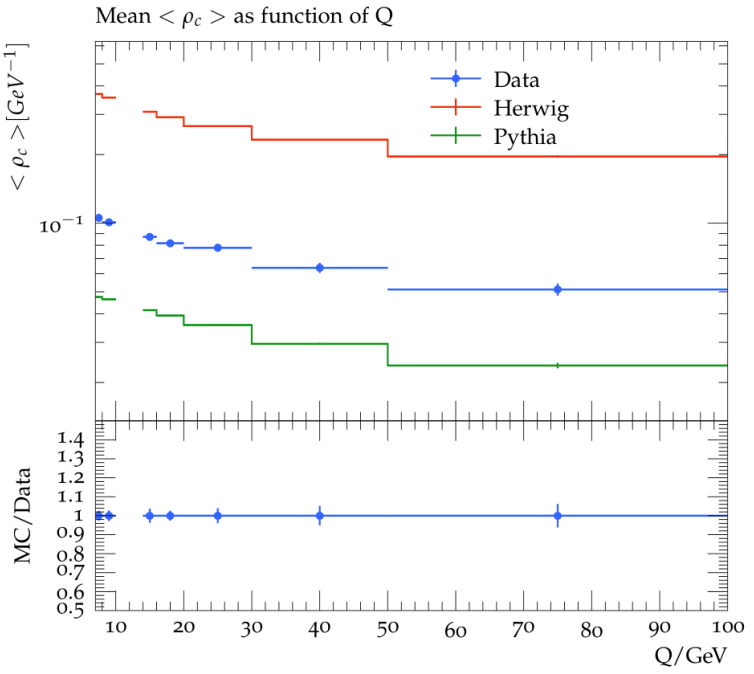
(a)



(b)



(c)



(d)

Figure 5.13: The mean value of $\langle 1 - T_C \rangle$ where thrust is the usual definition, similarly for jet broadening B_T, as well as with the axis of thrust taken with respect to the virtual photon direction. The kinematics cuts are for the ranges (7,100) in Q/GeV and (0.05, 0.8) in elasticity.

5.6 Inclusive jet cross sections

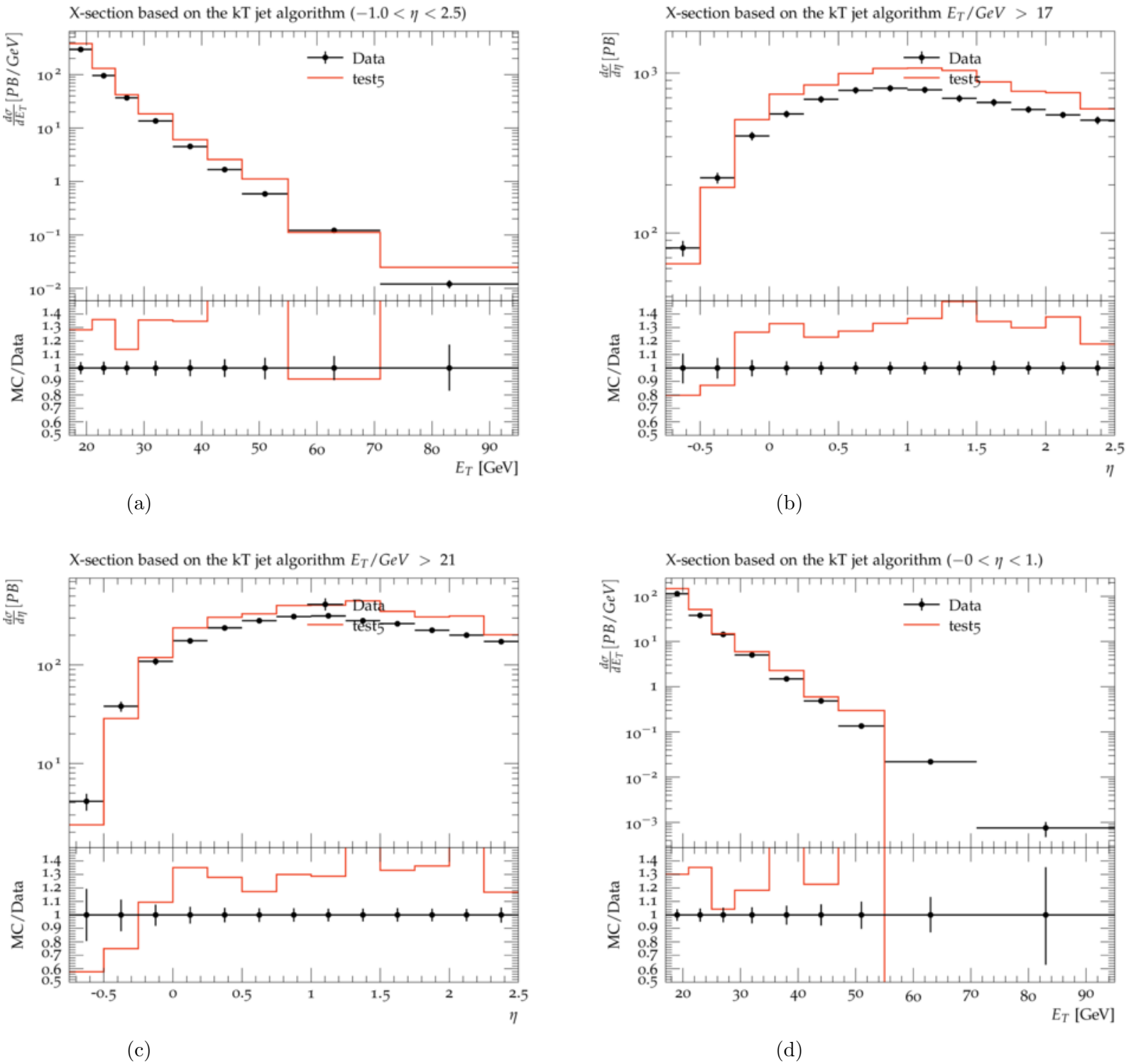
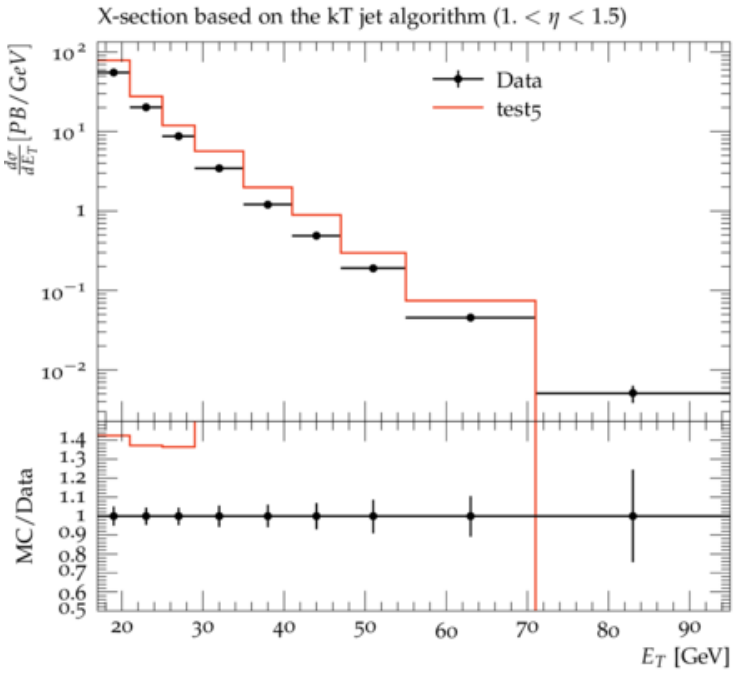
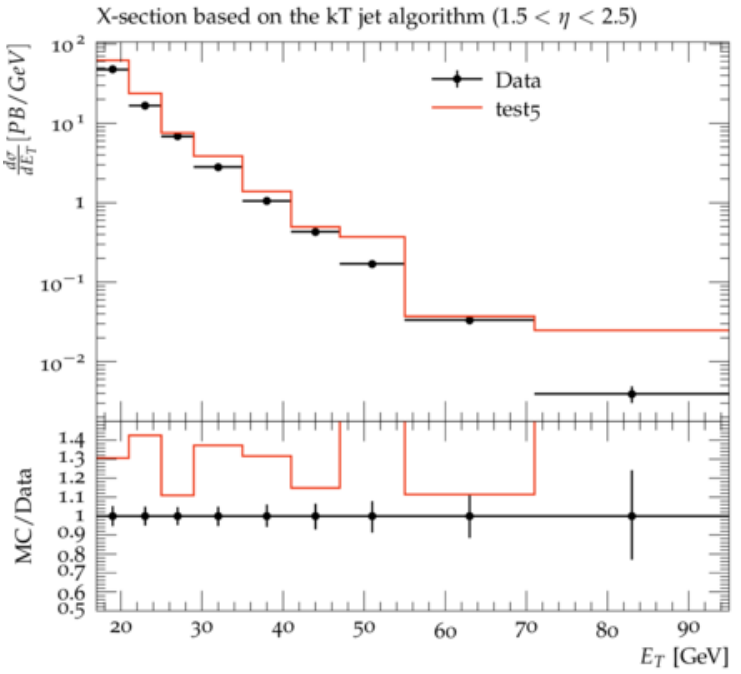


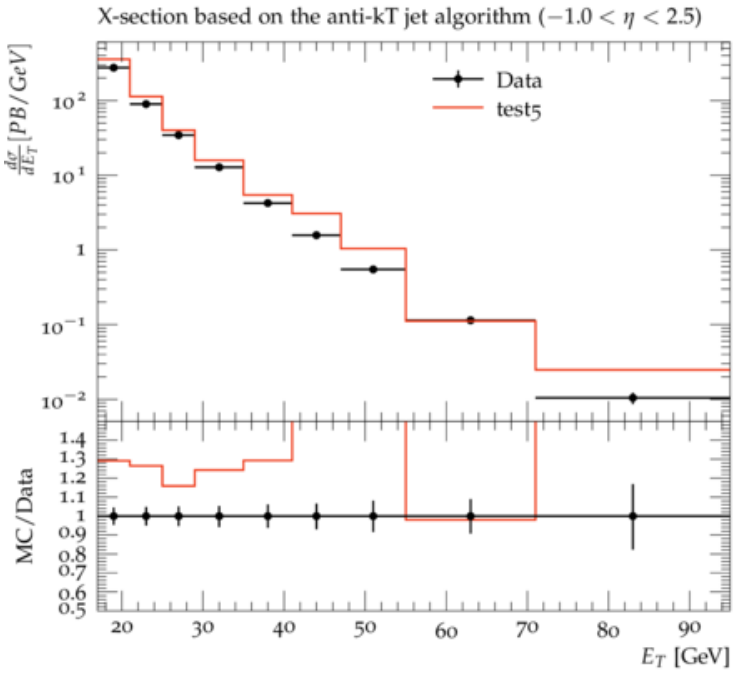
Figure 5.14: Inclusive jet production cross-sections



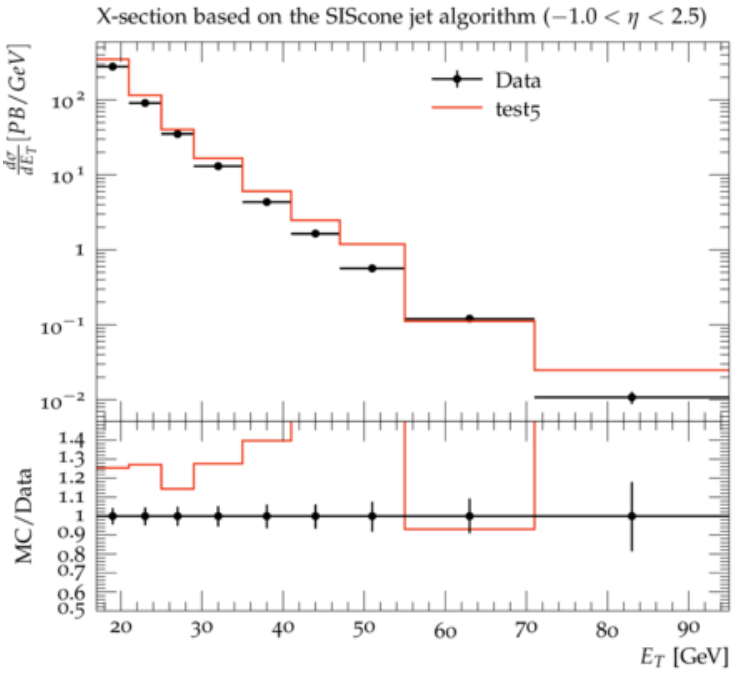
(a)



(b)



(c)



(d)

Figure 5.15: Inclusive Jet cross-sections

5.7 Discussion

The results obtained after the analysis show that there are specific areas where the Monte Carlo detectors are capable of reproducing the experimental data properly, whereas the it has its own areas were it fails. Due consideration has been put towards improving the future capabilities of the softwares in simulating the detector and collision event. The two event generators used are general purpose generators that are capable of generating a wide range of collisions form lepton-hadron and hadron - hadron. They are not specialized and doesn't cater to details of a particular experiment. This project helps in finding out the feasibility of the generators for testing in detector development and using it for documentation and user involvement.

The results covers a wide range of observables and provides a decent spectra for checking the consistency of the event generators and that it seems that certain regions can be mapped to the generators as areas where accurate reproduction is possible. The rivet package itself is a convenient and dependable analysis framework which allows for fair comparison among the generators.

Chapter 6

Summary

Overall the comparison with both pythia and Herwig for event simulation is rather complicated in the sense that both of them have certain regimes where the performance of one of them outweighs the other. But all in all in can be said for more accurate description of DIS observables it is better go with Herwig rather than Pythia which on its own does great job at capturing the trends of the variable dependency or cross-section's variations with the kinematics. However, this is when only DIS regime is considered and for the scenarios requiring simulation of photoproduction, Herwig has it's own limitations. Pythia on the other hand does it fairly accurately for some of the observables.

	Pythia	Herwig at LO
Charged Particle Multiplicity	1) Over estimates, 2) Captures nature of the curve 3) Doesn't produce accurate quantitative estimations	1) Over estimates, 2) Captures nature of the curve 3) Produce accurate quantitative estimations at lower multiplicity regions
Mean Multiplicity	1) Captures nature of the curve very precisely 2) Doesn't produce accurate quantitative estimations	1) Accurate estimation at higher pseudorapidity 2) Captures nature of the curve very precisely 3) Produce accurate quantitative estimations at lower W/GeV
Higher order averages in multiplicity	The MC data becomes more accurate as the pseudorapidity range increases	The MC data becomes more accurate as the pseudorapidity range increases
Transverse Momenta	Pythia represents the data relatively better as compared to herwig throughout pT range considered	The agreement is off in the lower transverse momenta regions but towards the higher range, we can see that it get relatively better.
Dstar Production	1) The nature of the distribution isn't exact but some parts are captured 2) Indications of missing scaling factor	1) The nature of the distribution alongwith quantitative estimations are fairly satisfactory within the uncertainty limits 2) Scaling not really required
ϕ meson production	Similar to the previous cases with following the nature but more accuracy as compared to Herwig for some of the global variables	The nature of the distribution alongwith quantitative estimations are not good at all The shape is okay for some plots
Event shape variables	Predicts the data relatively well alongwith Herwig	-
Inclusive jet production	Predicts the data fairly well 2) Produces accurate results in the photoproduction regime	NA

Table 6.1: Summarizes the performance of the Monte Carlo event generators used.

Bibliography

[1] J. Dalton
A New System of Chemical Philosophy
A. Bickerstaff, Strand, London, 1808.

[2] H. Becquerel *Comptes Rendus* 420; 501; 559; 689; 762.

[3] E. Rutherford *Phil. Mag. Ser. 6* (1911) 669.

[4] J. Chadwick *Proc. Royal Soc. Lond. Ser* (1932) 692.

[5] H. Yukawa *Proc. Phys.-Math. Soc. Japan* (1935) 48.

[6] R. Frisch and O. Stern *Z. Phys.* (1933) 4.

[7] R. Bacher *Phys. Rev* (1933) 1001. L. Alvarez and F. Bloch *Phys. Rev* (1940) 111.

[8] R. Hofstadter
Electron scattering and nuclear and nucleon structure. A collection of reprints with an intro- duction
(New York, Benjamin, 1963), 690 p. and references therein.

[9] M. Gell-Mann *Phys. Lett.* (1964) 214.

[10] G. Zweig CERN-TH-401 CERN-TH-412, (1964).

[11] W. K. H. Panofsky
Proc. 14th International Conference on High-Energy Physics, Vienna
1968, J. Prentki and J. Steinberger, eds., (CERN, Geneva, 1968), pp. 23.

[12] R. E. Taylor
Proc. 4th International Symposium on Electron and Photon Interactions at High Energies
Liverpool, 1969, (Daresbury Laboratory, 1969), eds. D.W. Braben and R.E. Rand, pp. 251.

[13] E. D. Bloom et al., Phys. Rev. Lett (1969) 930.

[14] M. Breidenbach et al., Phys. Rev. Lett. (1969) 935.

[15] R. E. Taylor Rev. Mod. Phys 573 (1991).

[16] H. W. Kendall Rev. Mod. Phys. 597 (1991).

[17] J. I. Friedman Rev. Mod. Phys. 615 (1991).

[18] C. G. Callan, Jr. and D. J. Gross Rev. Mod. Phys. (1969) 156.

[19] J. D. Bjorken Rev. Mod. Phys. (1969) 1547.

[20] R. P. Feynman, In: Proceedings of 3rd International Conference on High Energy Collisions, Stony Brook N.Y., 5-6 Sep 1969, pp 237.; Phys. Rev. Lett. 23 (1969) 1415.

[21] R. P. Feynman Photon-hadron interactions (Benjamin, Reading, MA, 1972), 282 p.

[22] Y. Nambu Preludes in Theoretical Physics, eds (North-Holland, Amsterdam, 1966), pp. 133.

[23] C. -N. Yang and R. L. Mills *RPhys. Rev* (1954) 191.

[24] M. Y. Han and Y. Nambu *Phys. Rev.* (1965) B1006.

[25] G. 't Hooft *Nucl. Phys. B* (1971) 173.

[26] H. Fritzsch and M. Gell-Mann
Proceedings of 16th International Conference on High-Energy Physics, Batavia, Illinois,
http://arxiv.org/abs/hep-ph/0208010

[27] H. Fritzsch, M. Gell-Mann and H. Leutwyler Phys. Lett. B (1973) 365.

[28] I. B. Khriplovich *Yad. Fiz.* (1969) 409.

[29] G. t'Hooft (1972)
Proc. Colloquium on Renormalization of Yang-Mills Fields and Appli- cation to Particle Physics, Marseille
(1985) 11.

[30] H1 Collaboration
Charged Particle Multiplicities in Deep Inelastic Scattering at HERA
arXiv: hep-ex/9608011v1.

[31] H1 and ZEUS Collaboration
Combination of Differential D± Cross-Section Measurements in Deep-Inelastic e p Scattering at HERA*
arXiv: 1503.06042v2 [hep-ex].

[32] H1 Collaboration
Measurement of Charged Particle Transverse Momentum Spectra in Deep Inelastic Scattering
arXiv: hep-ex/9610006v2.

[33] ZEUS Collaboration
Inlusive-jet photoproduction at HERA and determination of α_s
arXiv:1205.6153v1 [hep-ex] 28 May 2012

[34] H1 Collaboration
Inelastic Production of J/ψ Mesons in Photoproduction and Deep Inelastic Scattering at HERA
arXiv:1002.0234v1 [hep-ex] 1 Feb 2010

[35] ZEUS Collaboration
Measurement of inelastic J/ψ production in deep inelastic scattering at HERA
arXiv:hep-ex/0505008v1 5 May 2005

[36] H1 Collaboration
Measurement of Event Shape Variables in Deep Inelastic e p Scattering
arXiv:hep-ex/9706002v1 2 Jun 1997

[37] ZEUS Collaboration
Observation of the strange sea in the proton via inclusive ϕ-meson production in neutral current deep inelastic scattering at HERA
arXiv:hep-ex/0211025v2 13 Dec 2002