Implementation of a Rivet Multijet Analysis for Comparing Sherpa with H1 Data

September 16, 2011

Summer Student Programme 2011

Yunfan Zhang (Princeton University. yunfanz@princeton.edu)
Supervisor: Dr. Roman Kogler (H1 collaboration, DESY. kogler@desy.de)

Abstract

I participated in the 2011 DESY summer student program under the supervision of Dr. Roman Kogler of the H1 collaboration. For my main task, I wrote a Rivet analysis for checking MC generators Sherpa with recently analyzed data taken in the HERA2 running phase at the H1 experiment. Generated results from Sherpa are compared to data and presented in the results section. In this report, I will attempt to lay out the relevant background I learned, as well as present some results and documentation of my work.

Contents

1 Introduction

1.1 Deep Inelastic Scattering

Rutherford and Geiger-Marsden's α particle scattering experiments showed that atoms have hard centers, the nuclei. Likewise electrons scattering off protons showed that the proton is a composite particle consisting of what Feynman called partons. A typical such scattering is depicted in Fig. 1.

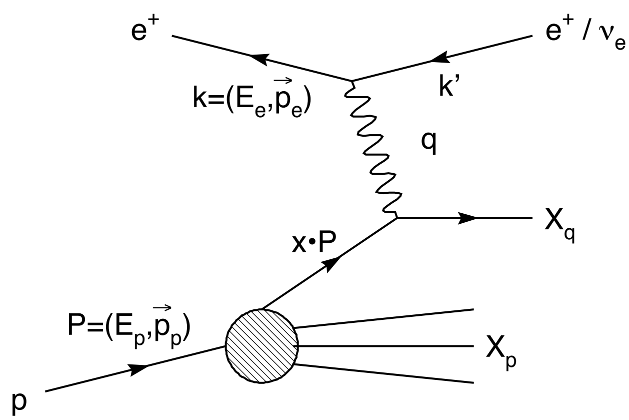

Fig. 1 Feynman graph of leading order DIS, where the incoming lepton interacts with a parton of momentum $x \cdot P$.

P is the momentum of the orginal proton with mass m_p, and X_q and X_p combined is the hadronic final state.

(Courtesy of HERA Physics Feynman Diagram Gallery by Gerhard Brandt)

Here we see that the momentum transfer $q = k^{'} - k$ can be calculated from the initial and final states of the electron. The commonly used kinematic variables, which are also used in this report, are given as follows.

$W^2 = (q + P)^2$ is the mass of the hadronic final state.

$Q^2 = -q^2$ is the momentum transfer of the process, characterises the interaction. For Deep Inelastic Scattering, we require $Q^2 \gg m_p$. Higher Q^2 corresponds to shorter wavelength of the virtual boson and makes probing subprotonic structure possible. Today, DIS is a major method to study the parton structure. We also require the mass of the hadronic final state to be large. i.e. $W \gg m_p$.

$s = (k + P)^2$ is the center of mass energy of such an ep scattering event.

x is the Bjorken scaling variable, the fraction of the proton's momentum carried by the interacting parton.

$y = \frac{P \cdot q}{P \cdot k}$ is the inelasticity, and measures the energy loss of the lepton in the proton's rest frame, and so $0 \leq y < 1$.

For DIS ($Q^2 \gg m_p$), we can ignore all particle masses and so $x = \frac{Q^2}{2P \cdot q} = \frac{Q^2}{W^2+Q^2-m_p^2}$, from which we see that $0 \leq x < 1$. In the same regime, we also have $Q^2 = sxy$, such that only two of the above variables are independent.

Another commonly useful variable is p_T, the momentum transverse to the direction of the incoming proton and electron beams.

And finally, the pseudo-rapidity is defined as $\eta = - \left[tan \left(\frac{\theta}{2} \right) \right]$, where θ is the angle between the particle's trajectory and the beam axis. η is thus a frame dependent parameter that measures the direction of the outgoing particle.

1.2 Jets

The force responsible for the inter-quark interaction is called the strong force and is mediated by the massless bosons called gluons. Analogous to the electric charge q of the electro-magnetic force, the strong force charge is called color. However, the strong force is described by QCD, which has SU(3) symmetry, incomparison to QED and its U(1) symmetry. As a result there are 3 color charges while only one electric charge is responsible for QED. Also, unlike its counterpart the photons, the gluons carry color charges themselves and can thus interact with each other. Unlike previously discovered "elementary" particles, the quarks cannot be observed as isolated particles. Instead, when two quarks are separated, their potential energy increases, the so-called long distance effects caused by gluon-gluon interactions. And if the original quarks have sufficient energy, new $q\bar{q}$ pairs and baryons, particles consist of three quarks, are created. In a series of such processes, a well collimated "jet" of hadrons is formed and observed. Studying jets gives insight into the QCD interactions that formed them, and it is for the simulation of these jets that my work is involved.

However, to study parton dynamics, we are only interested in those jets directly proportional to the strong coupling constant, which means that we need to eliminate processes like the one in Fig. 1. There only electro-weak coupling is involved. To do this, we boost the system to the Breit Frame, defined by:

$$2x\vec{p} + \vec{q} = 0$$

Where p and q are the momentum vectors of the incoming proton and the exchanged virtual boson. In the Breit Frame, the parton involved in the hard process is exactly backscattered. And thus, by requiring the transverse momentum of the jets to be higher than a certain value we can eliminate the jets we are not interested in.

1.3 The H1 Experiment

The data used in this report was taken by H1[1] during the 2nd running phase of HERA between 2003 and 2007, with an integrated luminosity of 351.6 pb^{-1}. The phase space cuts of the measurements are $150 < Q^2 < 15000\ GeV^2$, and $0.2 < y < 0.7$. Jets are required to the have lab frame pseudo-rapidity in the range $-1.0 < \eta_{lab} < 2.5$. For inclusive jets, it is required that $P_T > 7\ GeV$, while for dijets and trijets $P_T > 5\ GeV$. Finally, for the multijets events the mass of the two jets of the highest P_T (in Breit Frame) is required to be greater than 16 GeV.

1.4 Technical Introduction

1.4.1 Rivet Analysis Framework

Rivet (Robust Independent Validation of Experiment and Theory)[2] is a analysis framework written in C++ for validating Monte Carlo calculations of high energy physics. It is an successor of HZTool in C++. Rivet implements calculation packages called "projections" and the abstract histogram interface AIDA. Run on the generic "HepMC" event record, Rivet enables its users to tune and validate MC generators with succinct code and minimal effort. My main task was to write a Rivet analysis that compares results from MC Generators such as Sherpa with data taken from the HERA2 running phase at H1.

1.4.2 MC Generators and Sherpa

In high energy physics, data taken from the detector are typically compared to numerically calculated results from theory. The complicated multi-dimensional integrals and the complexity of phase space geometries make exact solution often impossible and provide difficulties for traditional numerical methods. Monte Carlo methods, on the other hand, work by sampling pseudo-random points in different regions of phase space and assigning each event a proper weight, and are so able to calculate high energy events with arbitrarily complex phase space geometry. Event generators such as Sherpa (Simulation of High Energy Reactions of PArticles)[3] use the "divide et impera" strategy and separates HEP events into different stages according some characteristic energy scales.

1.4.3 Global Structure of the MC Generator Validation process

Ther versions of major sofware used are Sherpa 1.3.0, Rivet 1.5.1, HepMC 2.06 and AIDA 3.2[4]. Events generated by Sherpa are recorded in HepMC format, which is then read into Rivet and fillled into histograms via AIDA.

2 Implementation Procedures

Sherpa is run with the steering file called Run.dat. In the steering file, one can set parameters such as the output format (HepMC), beam type and energy, the matrix element generator used(Comix), as well as numerous parameters regarding how the generation is performed.[1] The final file used is one of 2 million events at $Q^2 > 150^2\ GeV^2$.

[1] See the Appendix for an example of such a steering file

To analyse the events generated with Sherpa, one needs to write a Rivet analysis, which was my main task. Skeletons of Rivet analyses are automatically generated with the command rivet-mkanalysis, and all the code implementation are done in the .cc file thus generated. The code structure consist of a void constructor, which sets the need for crosssection, set the beam types etc., a no-argument init() method, which books histograms and add projections, an event by event analyze(Event&), which fills the histograms, and a no-argument finalize() method, which scales the histograms for the final cross sections to compare with the data.

The jet analyses are performed in the Breit Frame as described above. However, the Rivet version used, Rivet-1.5.1, does not yet have a final state class in the Breit Frame. And so we changed the center of mass frame class FinalStateHCM into a boost into the Breit frame. The Rivet Breit Frame Lorentz transform boostBreit() was tested against the relation $2x\vec{p}+\vec{q}=0$ to be correct at least to $10^{-10}GeV$ in the transverse directions. In the entire analysis, all but the pseudo-rapidity cut was done in the Breit frame, while for the pseudo-rapidity criterion the inverse transform was used.

In filling the histograms, I noticed that according to the AIDA documentation, the fill(value, weight) method throws illegal argument exception if the weight is not in the range $0 \leq weight \leq 1$. However, I made several tests including a test analysis filling a 1D histogram with 100 values of 5 and weight 10 each and the method works as wished and no exception was thrown.

3 Analysis and Results

The generated Sherpa events had a total cross section of $\sigma = 5164.65 \pm 117.163\ pb$. Each event generated is assigned a weight, and the sum of all the weights is equivalent to the effective total number of events, which is related to the total cross section via the relation $N = L\sigma$, where L is the integrated luminosity. Thus, with σ and the sum of weights we can get the integrated luminosity. And with the luminosity we can obtain the simulated differential cross sections as functions of the dynamic variables of interest. In the analysis, the generated histograms are scaled by $\frac{\sigma}{sumofweight}$ in the finalize() method to compare to the H1 data.

3.1 Event Histograms

The event by event histograms for x, y, Q^2, W^2, as well as jet multiplicities are shown in Fig. 2-6[2]:

[2] All these histgrams are filled after phase space cuts. Unless otherwise noted, all units of energy are GeV

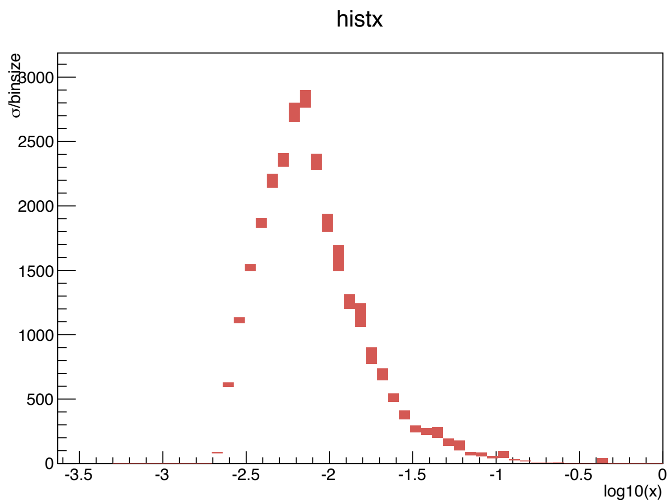

Fig. 2 The x histogram.

We see that x satisfies $0 < x < 1$ as explained in the introduction, and the x distribution peaks around $10^{-2.2}$

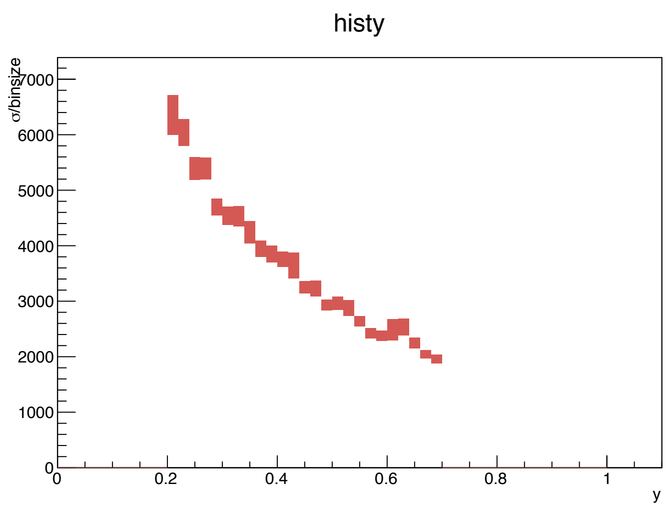

Fig. 3 The y histogram. This histogram is made with the phase space cut $0.2 < y < 0.7$.

We see that the inelasticity, and hence the energy loss of the lepton in the proton's rest frame drops off.

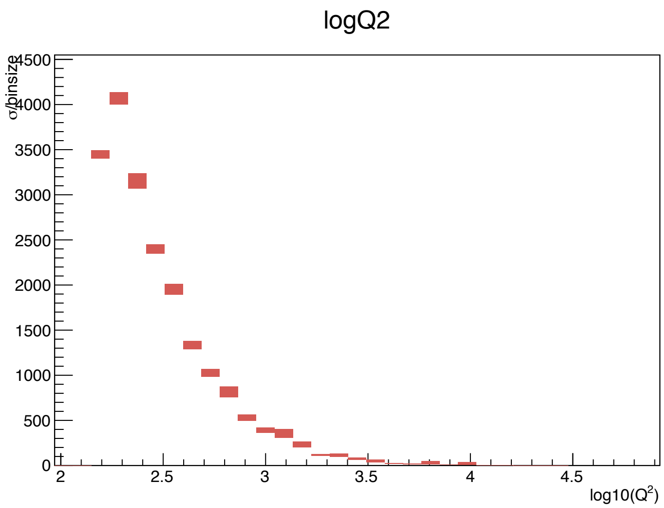

Fig. 4 Q^2 histogram, with the cut $Q^2 > 150^2\ GeV^2$. We see that the distribution drops off sharply,

despite the irregularity in the first bin.

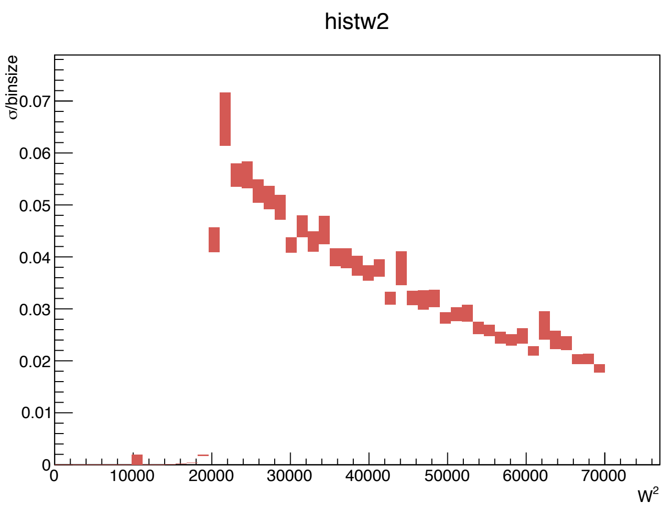

Fig. 5 The W^2 Histogram

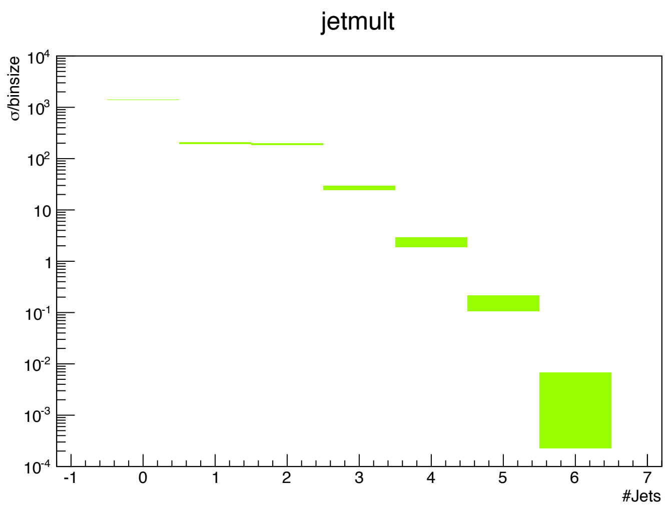

Fig. 6 The Jet Multipliccity histogram. We see a sharp drop-off of with the multiplicities of the jets,

but the number of dijet events is almost the same as the number of single jet events

3.2 Weight Histogram

The calculation of weights is a major part of a Monte Carlo Generator. In past H1 analyses, event weights range wildly from 10^{-10} to 10^{10}, and occasional large weights can lead to irregularities in the histograms, and hence it is important to see how the overall event weight distribution.

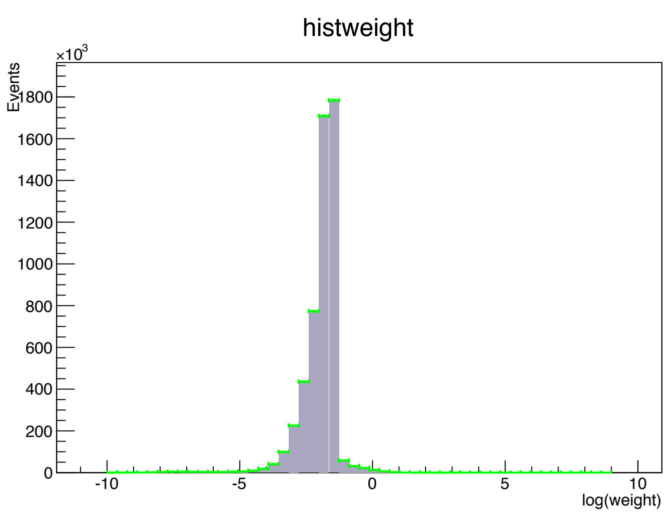

Fig. 7 The weight histogram. Notice that most of the weight are between 10^{-1} and 10^{-4}, which means it's fairly well-behaved.

3.3 Jet Histograms

The lab frame pseudo-rapidity histogram for the jets are as below:

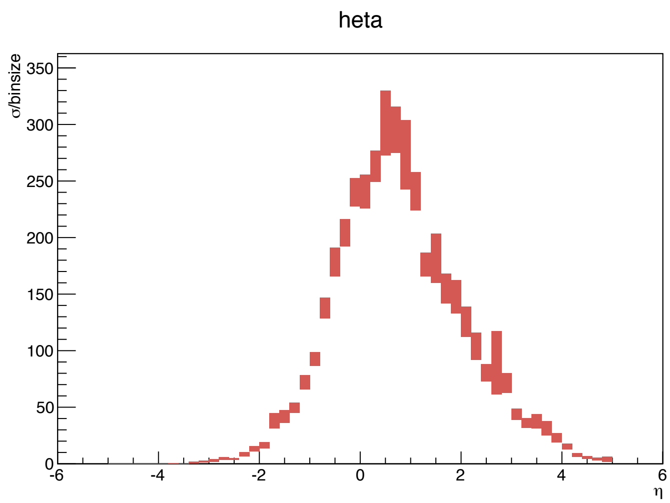

Fig. 8 The η_{lab} histogram. All the jets found are shown here, before the cut on the pseudo-rapidity.

For the main results, we realized that AIDA handles histograms by dividing the counts by the bin size. The standard bins for the H1 analysis are Q^2 : (150, 200), (200, 270), (270, 400), (400, 700), (700, 5000), (5000, 15000) GeV^2. And for Pt: (7, 11), (11, 18), (18, 30), (30, 50) GeV. To compare the data with the AIDA generated histograms, it is therefore necessary to divide the data by the corresponding bin sizes, resulting in the data below, where the errors, given in pb/GeV and pb/GeV^2, are obtained from the non-correlating sum of the systematic and statistical error percentages of the data file:

Table. 1 Ploted data of P_T histograms

Bins (GeV)	IncJets (pb/GeV)	Error (pb/GeV)	DiJets	Er	TriJets	Er
(7, 11)	70.04	2.48	24.22	0.960	4.605	0.438
(11,18)	21.36	0.821	9.385	0.401	1.858	0.177
(18, 30)	3.602	0.167	1.654	0.0827	0.2387	0.028
(30, 50)	0.2986	0.020	0.137	0.00978		

Table. 2 Ploted data of Q^2 histograms

Bins (GeV^2)	Inc (pb/GeV^2)	Er (pb/GeV^2)	DiJets	Er	TriJets	Er
(150, 200)	2.263	0.0879	0.9895	0.0446	0.1677	0.0160
(200, 270)	1.339	0.0520	0.5836	0.0252	0.09506	8.27E-3
(270, 400)	0.7209	0.0282	0.3160	0.0136	0.05564	4.90E-3
(400, 700)	0.2855	0.0109	0.1267	0.00550	0.02302	2.19E-3
(700, 5000)	0.02029	0.000812	0.009151	0.000424	1.640E-5	1.75E-4
(5000, 15000)	0.0005413	3.51E-05	0.00025	2.13E-5	4.618E-5	8.10E-6

These data are shown in Fig. 9-14, where the dot represents the data and the blue rectangles represent the generated values with errors. However, because in this analysis the generated total cross section is not already implemented in Rivet and needs to be manually fed in, the uncertainty of its value is not taken into account on these graphs.

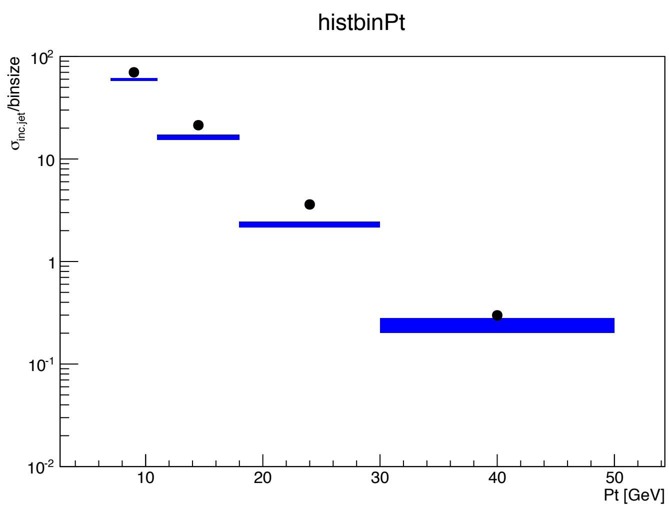

Fig. 9 The Inclusive Jets P_T histogram. We see that the histogram is falling off as expected for hard QCD interactions in Breit frame.

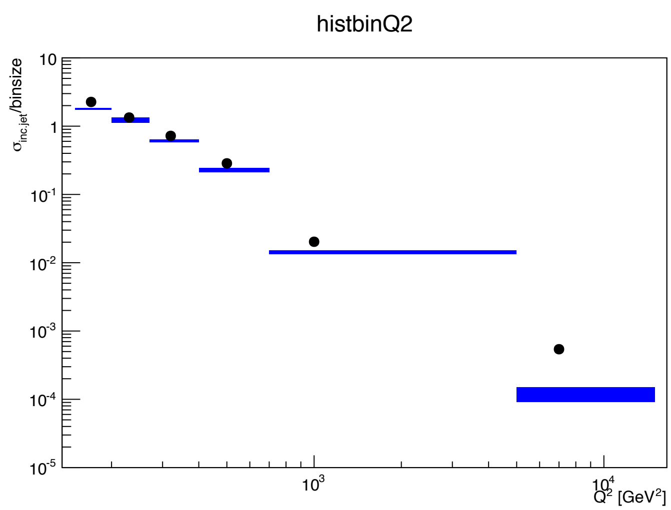

Fig. 10 The Inclusive Jets Q^2 histogram

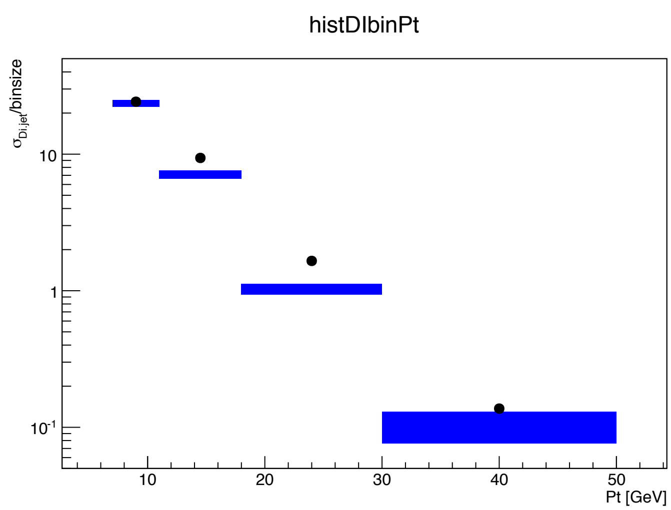

Fig. 11 The Dijet P_T histogram

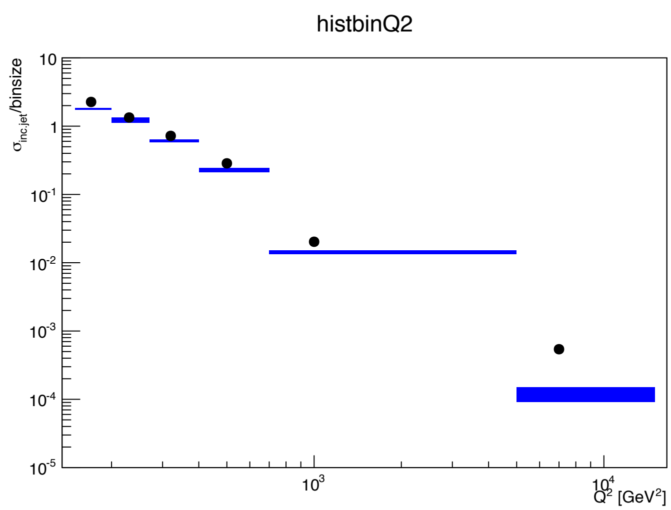

Fig. 12 The Dijet Q^2 histogram

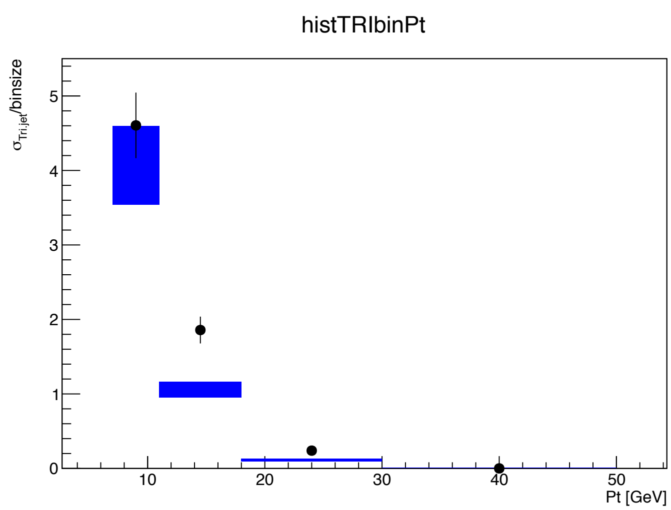

Fig. 13 The Trijet P_T histogram

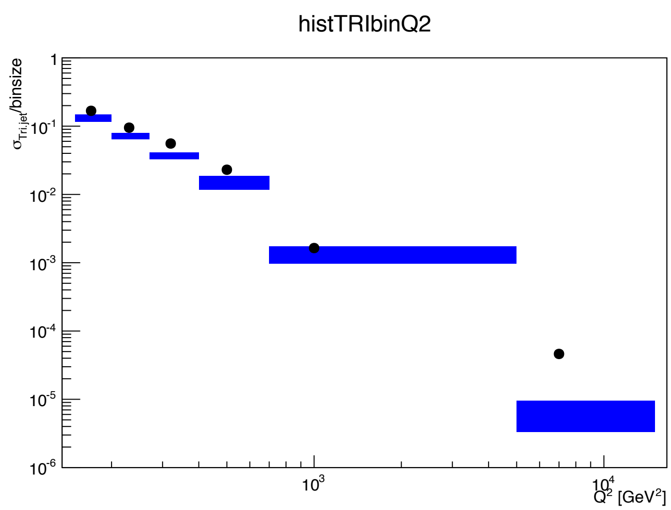

Fig. 14 The Trijet Q^2 histogram

Notice that for the majority of the bins the generated values are smaller than the data values. This may be due to the imprecisions of the generated total cross section. In particular, we observe that for the Q^2 histograms, the generated values are farthest away from the data for the last bin, while for the P_T histograms the middle two bins are described with the least accuracy.

4 Conclusion

As a result of my work at the summer student program I wrote a Rivet analysis to calculate the total cross sections from the Monte Carlo generator Sherpa. Sherpa describes the data with fair accuracy, but the total cross sections are somewhat too small. Many other tests were made to best ensure the correctness of the analysis and the same analysis could also be used to compare the data to other generators such as Herwig.

Appendix

Fine-binned Jet Histograms

The jet histograms in section 3.3 are presented with the standard H1 binning as described in that section. As a reference guide the fined binned histograms are presented in Fig. 15-20, where the histograms are plotted with 50 bins.

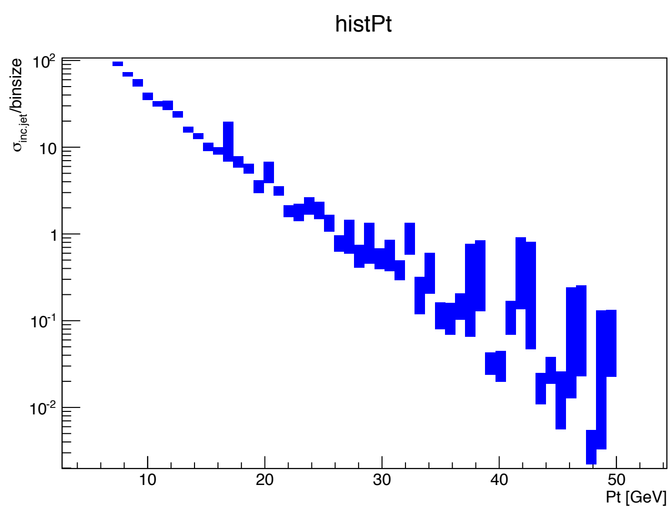

Fig. 15 Inclusive Jets P_T falls off as expected in Breit Frame

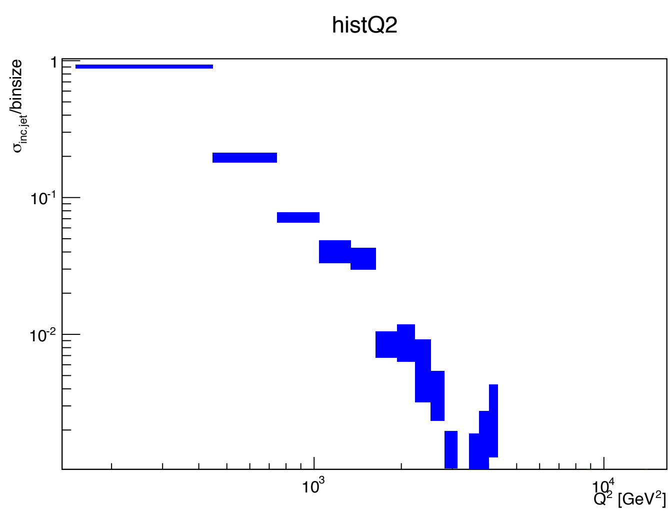

Fig. 16 Inclusive Jet Q^2

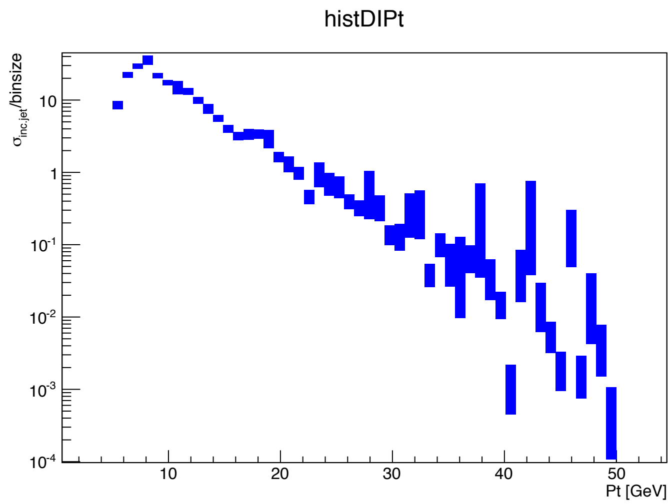

Fig. 17 Dijet P_T

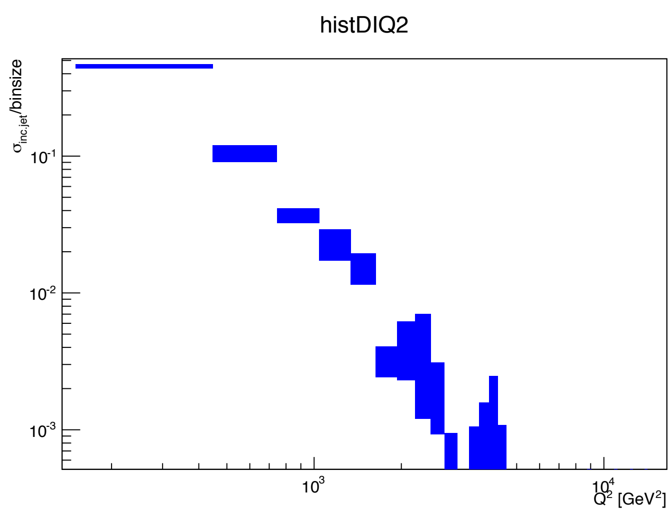

Fig. 18 Dijet Q^2

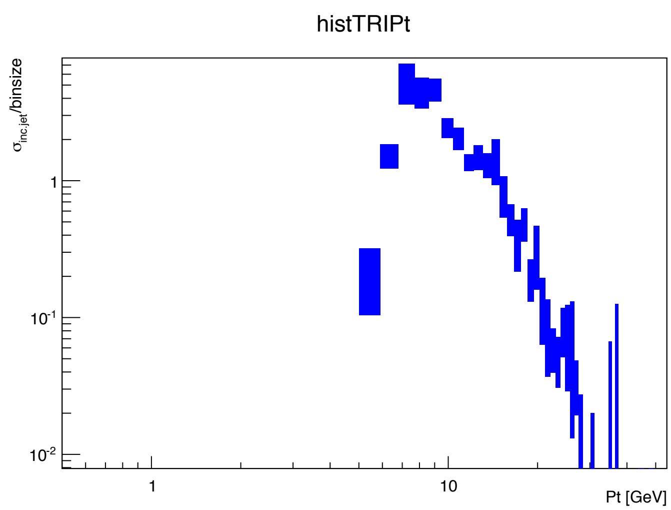

Fig. 19 Trijet P_T Notice the fall-off is not precise

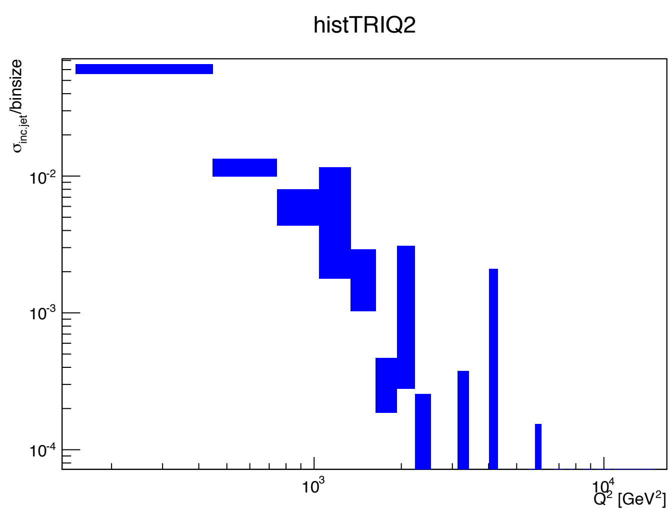

Fig. 20 Trijet Q^2

Acknowledgement

I would like to thank my supervisor Dr. Roman Kogler, other group members including Dr. Guenter Grindhammer, Daniel Britzger, Aziz Dossanov for kindly answering my questions and providing a comfortable working and studying environment. I would also like to thank my fellow summer student Junwu Wang for being such a great office-mate. Overall, I would like to thank everyone in the group as well as DESY Summer Student Programme for this wonderful experience and I am really glad that opportunities like this are and will be available for undergraduate students like us.

References

[1] R. Kogler. Measurement of Jet Production in Deep-Inelastic ep Scattering at Hera. Hamburg 2010, ISSN 1435-8085

[2] Rivet 1.5.1 Hepforge documentation (http://projects.hepforge.org/rivet/)

[3] Sherpa 1.3.0 documentation (http://www.sherpa-mc.de/)

[4] AIDA 2.6 documentationn (http://aida.freehep.org/doc/v3.3.0/api/index.html)