Study of α_s in $Z+1$ Jet and $Z+2$ Jets Channels

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

Shalini Das

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

April, 2025

Supervisor: Dr. Shilpi Jain
Co-supervisor: Prof. Sourabh Dube

Certificate

This is to certify that this dissertation entitled Study of α_s in $Z + 1$ Jet and $Z + 2$ Jets Channels, towards the partial fulfilment of the BS-MS dual degree programme at the Indian Institute of Science Education and Research, Pune represents work carried out by Shalini Das at Tata Institute of Fundamental Research Mumbai under my supervision in the department of High Energy Physics (DHEP), during the academic year 2024-2025.

Dr. Shilpi Jain

Committee:

Dr. Shilpi Jain

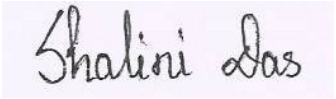
Prof. Sourabh Dube (cosupervisor)

Prof. Sachin jain (expert member)

This thesis is dedicated to my parents Shyamal Das and Shyamali Das

Declaration

I hereby declare that the matter embodied in the report entitled Study of α_s in $Z + 1$ Jet and $Z + 2$ Jets Channels are the results of the work carried out by me for the partial requirement of fulfilling my degree at the Department of Physics, Indian Institute of Science Education and Research, Pune, under the supervision of Dr. Shilpi Jain (TIFR Mumbai) and the same has not been submitted elsewhere for any other degree.



Shalini Das

Acknowledgments

I would like to thank my supervisor Dr. Shilpi Jain for guiding me in the project throughout. The fact that I could complete my project nicely because of her constant support. She has always made me feel that it is our project and not just mine. Hence, I always knew that if something goes wrong in it I will have her support. I would like to express my deepest gratitude for her constant guidance, patience, and unwavering support throughout this research. She provided me all the necessary research environment for growth. Her insightful feedback and encouragement motivated me in the completion of this project. She guided me in the analysis and in the basics understanding of Physics. I got to learn so many new things which motivated me to pursue a carrier in the field. She has an important role to generate a scientific rigor in me. I am always grateful to her.

I would like to thank my co-supervisor Prof. Sourabh Dube for guiding me and helping my throughout my BSMS journey. He is my first guide who introduced me to high energy particle Physics and since then he's been a big support to me in terms of Academia and mental health. He continuously observes the progress of the project and I express my sincere gratitude to him for being a constant support and helping me figure out a way whenever I am stuck in my studies. He has also guided me in couple of management activities and taught me useful life lessons. I express my sincere gratitude to him for nurturing my scientific temper throughout and I am thankful to him for believing in myself when I could not.

I thank my expert member Prof. Sachin Jain for being a part of my TAC commitee and giving me valuable suggestions.

I am thankful to Dr. Rajdeep Mohan Chatterjee for helping me in many aspects of the project. I express my sincere gratitude to Prof. Gobinda Majumder for helping me develop curiosity from the group meetings and Prof. Monoranjan Guchait for giving me the opportunity to learn from his classes. I express my sincere gratitude to Prof. Sunanda Banerjee for believing in me and giving me his valuable time and suggestions

to improve. I am thankful to TIFR CMS-B collaboration for hosting me during my MS theis and giving me the environment to learn new aspects of particle Physics. I thank Prachurya, Riya and Arnab for helping me in some crucial aspect of the experimental setup. Thanking Avik, my first collaborator who helped in several aspects of the project and helped me grow an interest in the field through discussion.

For giving me an opportunity to learn and nurturing my scientific interest in particle Physics, I am grateful to my other guides, Prof. Manjit Kaur and Dr. Ritu Aggarwal. They have guided me for a long time and helped me follow my research journey in so many aspects.

I thank Ranjan Kishor Sahu for helping me overcome the technical difficulties in courseworks in almost every semester. He had been a great help to all the students and suggests valuable inputs in designing a semester whenever he gets a chance.

I am grateful to my parents for believing in me and blessing me with a beautiful life and I thank Lubdhak Mondal for his support in my life.

I thank my friends Ananya, Anwesha, Manjima, Sneha, Orishma, Soumya, Barish, Aniket and Haraprasad for being there with me and helping me throughout. They made my college life beautiful and helped me overcome several difficulties.

I thank my friends from TIFR HECR hall, Ayush, Ashish, Animeash, Avik, Gursharan, Lokesh, Supriyo and Pravesh for providing me a nice friendly environment to work during my thesis.

Abstract

The QCD coupling constant (α_s) is today the least accurately known of all interaction couplings in nature. The coupling runs with energy and, by convention, it is given at the Z Boson mass scale. Currently, uncertainty on the measurement of α_s contributes significantly to the uncertainties in the calculations of key Standard Model (SM) decay and cross-section processes, such as the production and hadronic decay of the Higgs Boson.

Hence, precise determination of this quantity is of utmost importance to the high energy physics (HEP) community. Currently, its value is extracted from a number of processes in hadron colliders. However, what has not been attempted is its estimation in ratios of cross-sections where common uncertainties cancel out. Therefore, the aim of this study is to understand observables that are sensitive to α_s in the ratios of Z + n jet to Z + (n-1) jet cross-section (or differential cross-section) ratios. We will also study the observables in Z + n jet processes sensitive to α_s. This study will serve as input to the precise extraction of α_s from Z + n jet processes and their ratios.

Contents

Introduction

The $Z+$jets process plays a crucial role in understanding the interactions of electroweak and strong forces. We can test our theoretical predictions from quantum chromodynamics (QCD) and electroweak theory by studying this. It gives us insights into the structure of the proton. This analysis focuses on comparing experimental data from the 2018 UL datasets with Monte Carlo (MC) simulations to verify the theory predictions and finally leads to evaluate systematic uncertainties.

The study involves selecting events with at least two well-identified electrons to reconstruct the Z-boson and categorize events on the basis of jet multiplicity. By analyzing key distributions, such as the Z-boson transverse momentum (p_T) and jet multiplicities, the analysis captures how theoretical systematics like renormalization (μ_R) and factorization (μ_F) scales, parton shower effects, and parton distribution functions (PDFs) influence the results.

This detailed comparison of data and simulation helps refine the theoretical models, improve Monte Carlo generators, and ensure accurate predictions for future experiments. By this study, one can hope to better understand the uncertainties in $Z+$jets processes and identify event categories that are less sensitive to these uncertainties, paving the way for more precise physics measurements.

Part I

Theoretical motivation

Chapter 1

QCD running coupling

1.1 Introduction to QCD and the running coupling

Along with electromagnetic force, weak force, and gravity, the strong force is also a fundamental interaction in nature, it binds quarks together to form hadrons and nucleons into nuclei. The coupling constant for strong force is denoted as α_s. The strength of a force is determined by its coupling constant, which sets how strongly it acts at a given distance and a quantity of matter. The theoretical framework that describes the strong force is called quantum chromodynamics (QCD), this in some aspect, is analogous to how electromagnetism is governed by quantum electrodynamics (QED).

1.1.1 Color charges and quark interactions

The strong force sources in QCD are known as color charges and they come in three varieties: red, green, and blue. Quarks carry a single color charge, whereas anti-quarks carry the inverse colors. In contrast to the electrically neutral photons in electromagnetism, gluons, the strong force mediators, carry both a color and an ant-icolor. The strong force occurs when quarks exchange color charges through gluons, with α_s regulating the probability of gluon emission. Since the gluons are also colored, they can interact with each other which gives rise to complexities in QCD.

1.1.2 Quantum Effects and Running Coupling Constant

In classical physics, coupling constants have set values. However, in QCD quantum anomalies and loop effects cause them to become scale-dependent. Quantum loops involve the brief formation and annihilation of particle-antiparticle pairs and it can alter the force over very short distances. This alters the basic inverse-square law, necessitating renormalization to keep calculations limited and understandable. As a result, coupling constants have a scale dependence, known as their *running*.

The electromagnetic coupling constant, α, only increases by roughly 10 percent at large energy scales. In comparison, the running of α_s is considerable, varying by many folds. Importantly, while α increases at shorter distances, α_s declines, eventually vanishing at zero distance. This phenomena is known as quark asymptotic freedom. This feature facilitates the use of perturbation theory for QCD at short distances, resulting in precise theoretical predictions in this area.

1.2 Challenges and Uncertainties in Measuring α_s

Despite advancements, challenges remain in understanding α_s:

1. **Complexity of QCD:** Quantum Chromodynamics, the theory that describes the strong force is not very simple. Unlike simpler forces like electromagnetism, it is non-abelian which means that the force carriers of QCD (gluons) can interact with each other, leading to a lot of non-linearity in the equations. As a result, when researchers try to calculate things using perturbative methods, it becomes increasingly difficult to handle higher-order terms(NLO, NNLO etc) in these calculations.
2. **Experimental Limitations:** Though the theory of QCD predicts the *running of* α_s, it does not directly give the absolute value of α_s. To figure out its exact value at a given energy scale, various experiments or non-perturbative methods (like lattice QCD) are required. Still the precise determination of α_s difficult because these experimental methods come with their own challenges and uncertainties.

These issues lead to uncertainties in measuring α_s. For example, the precision for α is roughly one part in 10 billion. Higher accuracy is necessary for experiments at facilities like CERN's LHC, where exact α_s values are crucial for verifying the Standard Model and investigating physics phenomena (eg. Higgs cross section analysis).

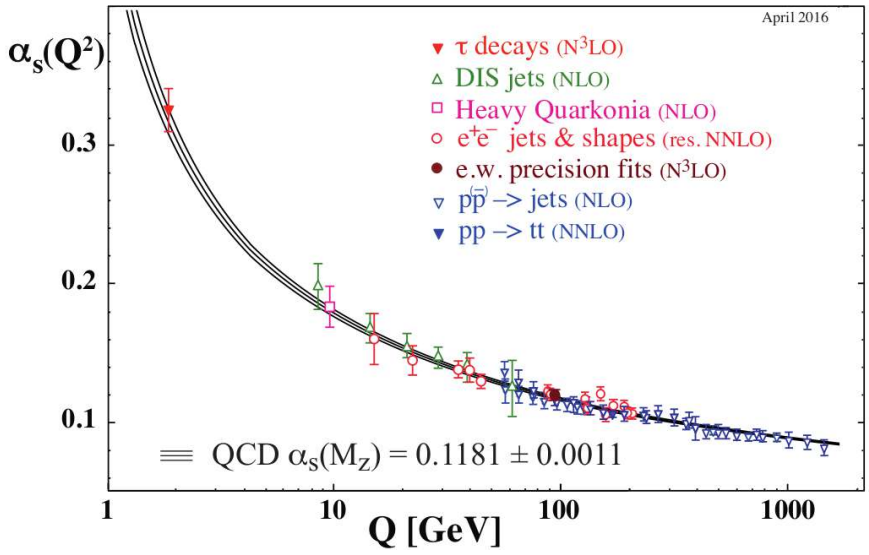


Figure 1.1: Summary of measurements of α_s as a function of the energy scale Q.

1.2.1 The Long-Distance Behavior of α_s

At high energies, the strong coupling constant α_s decreases which indicates that quarks and gluons interact weakly. This leads to *asymptotic freedom*, which permits quarks to travel autonomously within hadrons when probed at extremely small distances or high energy. The reason for this is the non-abelian nature of QCD, which allows gluons (the force carriers) to interact with one another. With the increasing energy, self-interactions becomes weak (the force between quarks), resulting in a drop in α_s. This property has been experimentally validated in high-energy particle collisions (e.g., LHC).

1.2.2 The Short-Distance Behavior of α_s

At low energies, the strong coupling constant α_s grows significantly, increasing the interactions between quarks. This causes the *confinement*, in which quarks are strongly bonded within hadrons (like protons and neutrons). Thus a single quark can not be found. The increase in α_s at low energies is also due to gluon self-interactions, it increase the force between quarks as the energy scale drops. This explains why quarks are always detected in bound states rather than in isolation, which is very peculiar to QCD.

1.3 Why and How Force Couplings Run

The running of force couplings is a cornerstone of quantum field theory (QFT) which describes how the interaction strength of fundamental forces varies with the energy scale Q^2. This phenomenon arises due to quantum corrections, vacuum polarization, and the nature of gauge interactions. Here is a classical and quantum insights for the same.

1.3.1 Classical Analogy

In classical electromagnetism, Coulomb's law describes the force between two charges:

$$F = \frac{q_1 q_2}{4\pi\epsilon_0 r^2}.$$

At large distances r, the observed charge q_{eff} remains constant, as there is no modification to the field. However, if a medium surrounds the charge, the effective charge is screened due to the polarization of the medium.

This idea can be extended to quantum mechanics, where the vacuum itself behaves like a medium. In QFT, the vacuum is not empty but it is filled with virtual particle-antiparticle pairs that interact with the field. These virtual particles polarize the vacuum and effectively screens or modifies the charge as observed at different distances. This provides an intuitive bridge to the quantum mechanical concept of "running" couplings, where the strength of interactions is energy dependent.

1.3.2 Quantum Corrections

In QFT, the interaction strength (or coupling constant) becomes scale-dependent due to quantum corrections from higher-order Feynman diagrams. These corrections arise from loop diagrams, which contribute to the propagators and vertices of the interactions.

The key mechanisms are as follows.

- **Vacuum Polarization in QED:** In Quantum Electrodynamics (QED), virtual electron-positron pairs in the vacuum act as a screening medium. At larger distances (or smaller Q^2), these pairs shield the bare charge, reducing the effective charge observed. As a result, the electromagnetic coupling constant increases with the

decrease in energy scale.

- **Color Charge Anti-Screening in QCD:** In Quantum Chromodynamics (QCD), the scinario is different. Here, gluon self-interactions dominate, leading to an anti-screening effect. Instead of shielding the color charge, these self-interactions spread it spatially. This causes the strong coupling constant α_s to decrease as the shorter distances (or Q^2 becomes larger). This unique property of QCD is called *asymptotic freedom*, where quarks interact more weakly at high energies.

Mathematical Formulation

The scale dependence of a coupling constant is governed by the renormalization group equation (RGE):

$$\frac{d\alpha}{d\ln Q^2} = \beta(\alpha),$$

where $\beta(\alpha)$ is the beta function. For small values of α, the running coupling can be approximated as:

$$\alpha_s(Q^2) = \frac{1}{\beta_0 \ln(Q^2/\Lambda^2)},$$

where Λ is the QCD scale parameter, and β_0 is a constant determined by the number of active quark flavors.

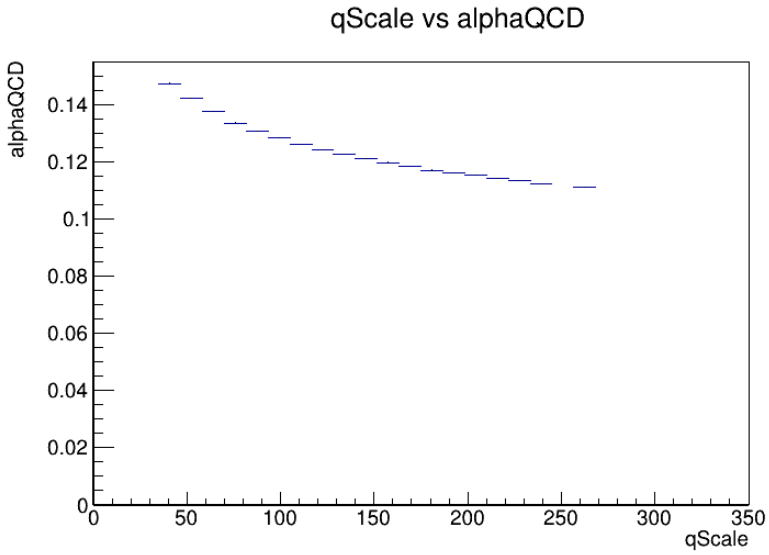


Figure 1.2: Running of strong coupling with respect to energy.

1.4 Importance of α_s

The motivation to study α_s lies in its significance as the least precisely known coupling constant, determined only up to the 3rd decimal place, compared to other couplings known to the 7th–10th decimal place. The recent ATLAS measurement in 2023 ($\alpha_s(M_Z^2) = 0.1183 \pm 0.0009$) represents the most precise determination yet, but uncertainties in α_s still significantly affect cross-section predictions for processes like $H \to b\bar{b}$, $H \to c\bar{c}$, and $H \to gg$.

Since α_s evolves with energy, studying its impact on sensitive observables at different scales provides valuable insights into precision tests of the Standard Model (SM). This study complements CMS jet cross section measurements by focusing on Z+jets processes, using ratio cross sections to reduce systematic uncertainties and achieve more accurate extraction of α_s. The improved precision in α_s improves the predictions of the SM and increases the potential to discover new physics beyond the SM.

Chapter 2

α_s determination at CMS

Jet Inclusive

Jets are produced in abundance, which makes them ideal for α_s cross section measurements. Cross sections at 7 TeV and 8 TeV are measured and compared with the theoretical calculations from pQCD at NLO. The renormalization scale (μ_R) and factorization scale (μ_F) are varied to determine their impact on the determination of $\alpha_s(m_Z)$. Theoretical calculations are improved by including non-perturbative effects and electroweak contributions.

The observations span a wide range of momentum transfer (Q) up to the TeV scale, allowing for the investigation of the running of α_s. It was found that α_s ranges between 0.1164 and 0.1192 with an uncertainty of 1.3%–3%. These uncertainties are preliminarily driven by truncation of perturbative expansions, through the use of NNLO calculations. At NLO, μ_R and μ_F are significant sources of errors. NNLO has the potential to reduce uncertainties.

Dijet Production

Dijet production involves selecting events with two jets and measuring the differential cross section as a function of Q, which is commonly specified by the dijet system's invariant mass or the transverse momentum of the leading jet. NLO QCD calculations are compared with the measured cross sections. These calculations take into account

higher-order corrections to improve accuracy.

It is observed that the dijet p_T spectrum and rapidity distribution are sensitive to α_s.

Three-Jet Production

Observable ratios, such as $R(3/2)$, are used to decrease susceptibility to experimental errors and renormalization and factorization scale selection. These ratios are sensitive to α^3 and α^2, making them powerful tools to probe. To match the experimental data's precision, advanced computations such as NNLO and logarithmic correction resummation are used. Uncertainties are reduced but care must be taken with the jets produced in the final state.

PDF Uncertainties

PDFs represent the probability distributions of quarks and gluons within a proton. Uncertainties in PDFs, particularly the gluon distribution, are a substantial cause of inaccuracy in α_s measurements. Fitting different PDF sets with α_s simultaneously reduces some uncertainties.

From pair production of $t\bar{t}$

From the $t\bar{t}$ channel, the top quark mass is a source of uncertainties. Considering the rapidity and invariant mass of $t\bar{t}$, the cross section is calculated. These distributions are used to extract α_s. This channel has uncertainties around 1%–2% from the top quark mass and PDFs.

Scale Variations

To evaluate theoretical uncertainty owing to missing higher-order corrections, the scales μ_R and μ_F (Renormalization and factorization, respectively) are modified separately by two factors around a central value.

Extreme Cases

To understand the influence on the extracted α_s, extreme changes of the scales ($\mu_R/\mu_F = 1/4$ and $\mu_R/\mu_F = 4$) are studied.

Non-Perturbative Effects

Hadronization

Non-perturbative effects such as hadronization (the formation of hadrons from quarks and gluons) are examined. These impacts are often studied using Monte Carlo simulations and factored into the uncertainty estimations.

Combination of Results

The results of several measurements (including jets, dijets, three-jets, and top-quark pairings) are integrated. To confirm the reliability of the α_s determination, the measurements are reviewed for consistency. To validate measurements, the extracted values of $\alpha_s(m_Z)$ are compared to the world average (e.g., from the Particle Data Group). Other than these, the DIS experiments (e.g., H1, ZEUS) also attempted to measure α_s.

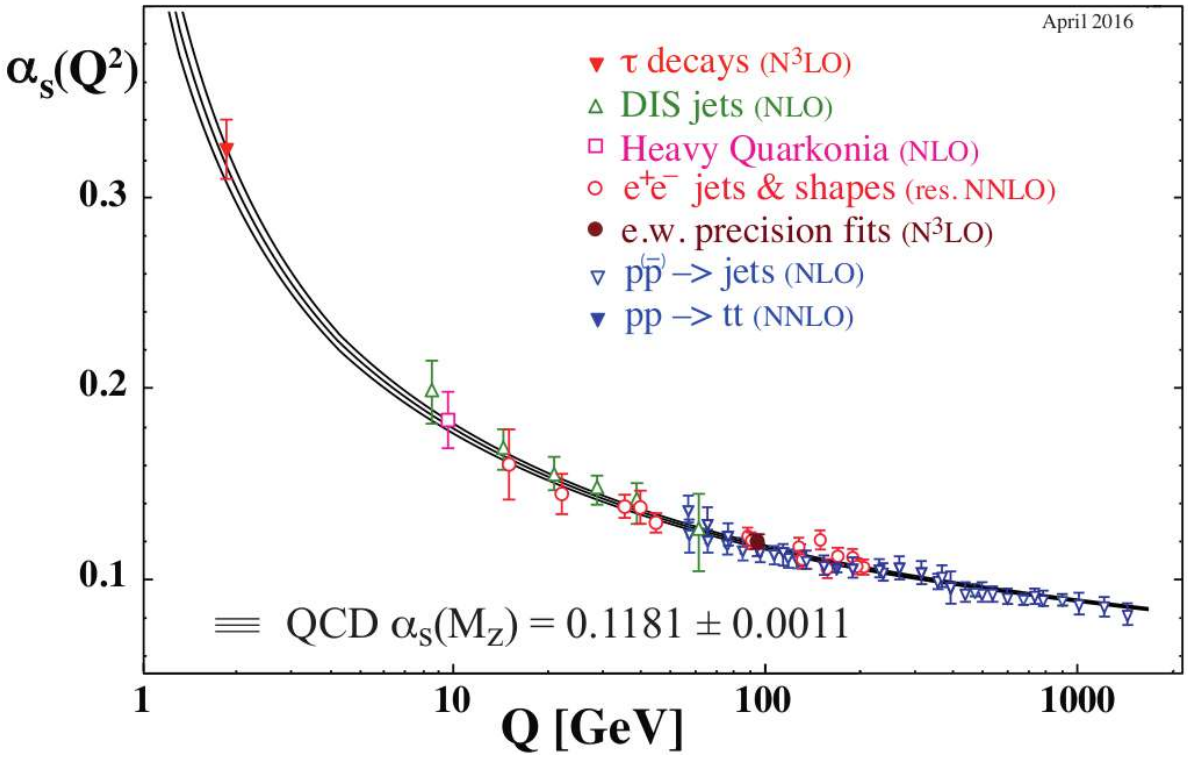


Figure 2.1: Summary of measurements of α_s as a function of the energy scale Q.

2.0.1 Extraction of $\alpha_s(m_Z)$ from Inclusive $W^{\pm}$ and Z Cross Sections at the LHC

This study extracts the QCD coupling constant, $\alpha_s(m_Z)$, by comparing 28 measurements of $W^{\pm}$ and Z boson production cross sections in proton-proton collisions at the LHC with theoretical predictions at NNLO in QCD, supplemented by NLO electroweak corrections. Experimental data were taken from ATLAS, CMS, and LHCb at 7, 8, and 13 TeV. Four parton distribution function (PDF) sets were used: CT14, HERAPDF 2.0, MMHT14, and NNPDF 3.0.

Theoretical cross sections were computed using MCFM (NNLO QCD) and corrected for electroweak effects using MCSANC. The input $\alpha_s(m_Z)$ was varied over the range 0.111–0.121. Cross sections were calculated for experimental kinematic cuts, and systematic uncertainties were included. PDF uncertainties ($\sim$ 2–3%) were dominant, followed

by scale ($\sim 1\%$) and numerical uncertainties ($\sim 0.7\%$).

To determine $\alpha_s(m_Z)$, a first-order polynomial fit was used to model the sensitivity of theoretical cross sections to variations in $\alpha_s(m_Z)$. The preferred $\alpha_s(m_Z)$ values were determined by the intersection of experimental and theoretical predictions for each system. These were combined for each PDF set using the `convino` tool, which accounts for correlations and uncertainties.

The results are:

- CT14: $\alpha_s(m_Z) = 0.1181 \pm 0.0016$
- MMHT14: $\alpha_s(m_Z) = 0.1209 \pm 0.0015$
- NNPDF 3.0: $\alpha_s(m_Z) = 0.1163 \pm 0.0019$

HERAPDF 2.0 showed instability and did not converge. Despite differences, all values agree with the world average (0.1181) within uncertainties ($\sim 1.5\%$). Stability tests demonstrated robustness, with results remaining consistent under variations in data and assumptions.

This precise approach highlights the sensitivity of inclusive $W^{\pm}$ and Z cross sections to $\alpha_s(m_Z)$, contributing valuable insights into QCD.

Part II

Experimental Setup

Chapter 3

LHC and the CMS detector

The Large Hadron Collider (LHC) is the world's most powerful particle accelerator and a groundbreaking tool for studying fundamental particles and forces in nature. The LHC collides protons to research phenomena including the Higgs boson, dark-matter candidates, and the environment of the early cosmos. The process starts with the generation of protons, which are acquired by ionizing hydrogen gas. The resultant protons are then accelerated through a succession of mechanisms that steadily boost their energy to unprecedented levels.

The protons are accelerated by an electric field in a linear accelerator(Linac 2). By the time the packet of protons leaves Linac 2, it travels with a speed of one third of light speed. In order to maximize the intensity of the proton beam, the proton packet is divided into four chunks, one for each of four booster rings or circular accelerators that have a circumference of 157 m. In the four coils, the protons are accelerated with an electric pulse and magnetic field. The protons achieve a speed of 91.6 percent of light speed and are squeezed closer together. The protons then flung on to stage three, the proton synchrotron, which is 628 m in circumference. 99.9 percent of light velocity is achieved at this stage. This is also the transition point, a point where the energy added to the protons by the pulsating electric field cannot increase their speed; it manifests itself as the increasing mass of proton. The protons are now channeled into stage four, the super proton synchrotron; a huge ring of 7 km circumference, designed specifically to increase the energy of the protons. The protons are then launched into the orbits of gigantic large hadron collider(LHC), lying deep underground with a circumference of 27 km. There are two vacuum pipes within the LHC that contain protons in opposite directions. The counter rotating beams cross over in the four detectors of LHC where

they can be made to collide. The detectors are CMS, ATLAS, ALICE, and LHCb.

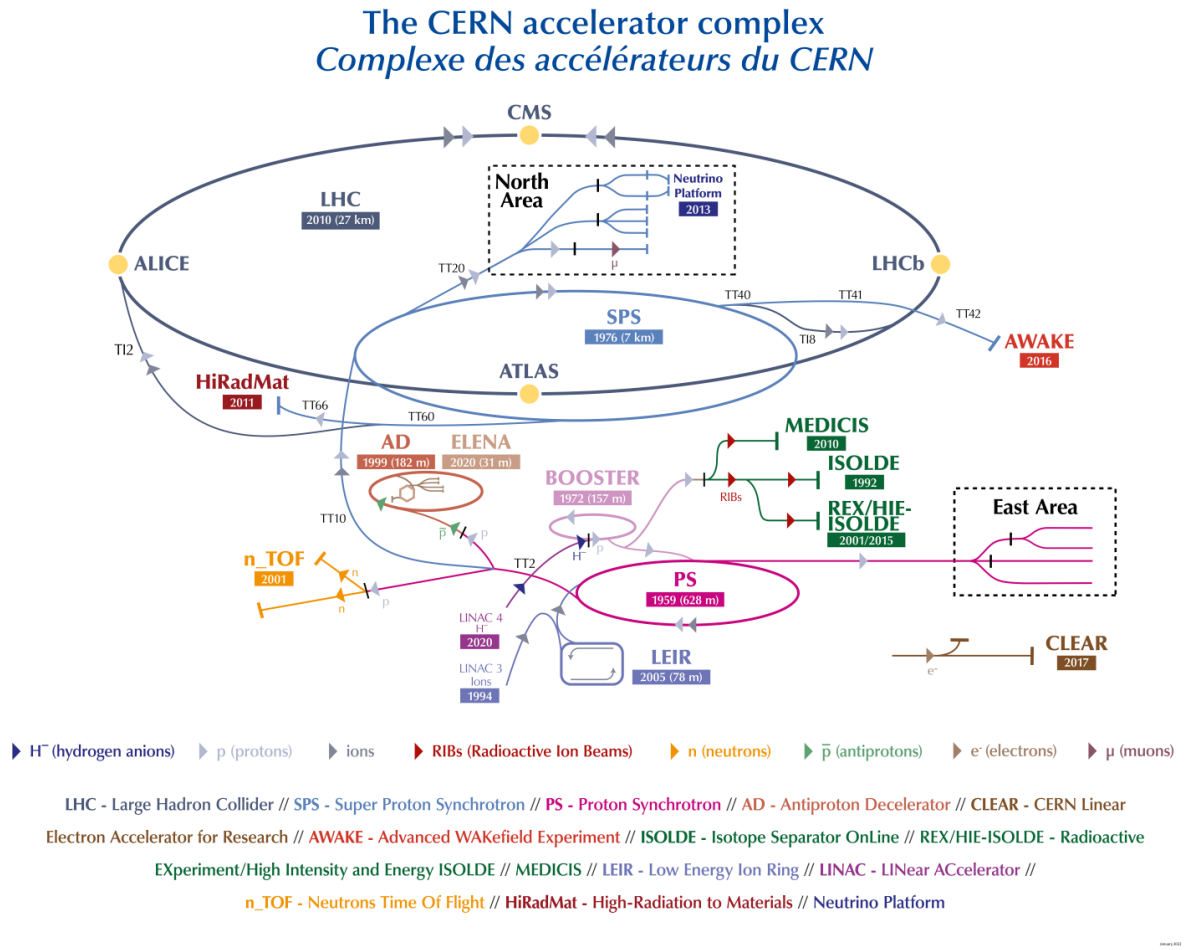


Figure 3.1: detectors of LHC

The CMS detector:

The Compact Muon Solenoid (CMS) detector is located at one of these four collision locations. It is a general-purpose detector, meaning it is intended to observe any novel physics phenomena that the LHC may discover.

CMS functions as a massive, high-speed camera, taking 3D "photographs" of particle collisions from every angle up to 40 million times per second. Although the majority of the particles created in the collisions are "unstable," they quickly change into stable particles detectable by CMS. By recognizing (almost) all the stable particles created in each collision, measuring their momenta and energies, and then piecing together the information of all these particles like the pieces of a puzzle, the detector can recreate a "image" of the collision.

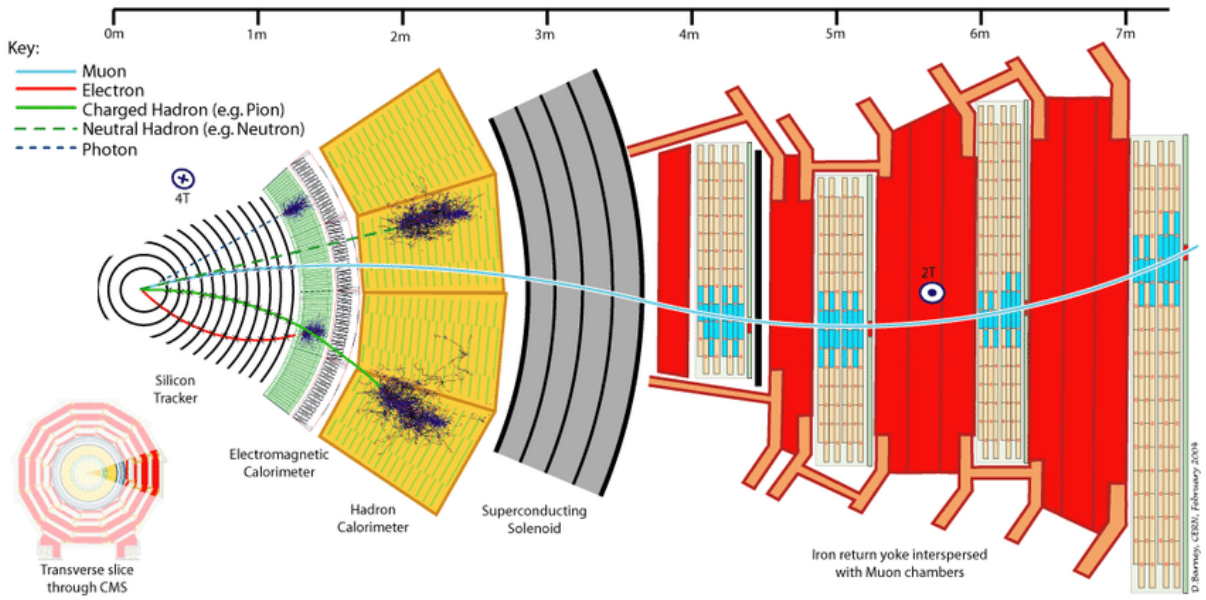


Figure 3.2: cross section of the Compact Muon Solenoid

From center to outwards the CMS decector consists of a **Superconducting Solenoid** for bending particle trajectories, the **Inner Tracking System** for tracking charged particles, the **ECAL** and **HCAL** for energy measurements, and the **Muon Detectors** for identifying muons. Systems like **PPS**, **BRIL**, and the **Trigger and Data Acquisition Systems** further enhance precision and event selection.

1. Superconducting Solenoid

Component	Materials/Composition	Function
Solenoid	6 m internal diameter, 12.5 m length, superconducting material	Provides 3.8 T magnetic field, 2.2 GJ stored energy, bends charged particles for momentum measurement

Table 3.1: Superconducting Solenoid

2. Inner Tracking System

Component	Materials/Composition	Function
Pixel Detector	Four barrel layers and three disks of silicon sensors on each side, 124 million readout channels	Measures trajectories of charged particles, provides precise position measurements, closest to interaction point
Strip Tracker	Ten layers in barrel, nine disks in endcaps, silicon strip sensors	20 m hit resolution for charged particles, detects particles with p_T as low as 50 MeV, provides impact parameter resolution of ∼10 m

Table 3.2: Inner Tracking System

3. Electromagnetic Calorimeter (ECAL)

Component	Materials/Composition	Function
ECAL Barrel (EB)	61,200 lead tungstate ($PbWO_4$) crystals, crystal depth: ∼23 cm (25 radiation lengths)	Identifies electrons and photons, measures their position and energy, covers $\|\eta\| < 1.48$
ECAL Endcap (EE)	7,324 lead tungstate crystals in each endcap, preshower detectors with silicon sensors and lead absorbers	Extends coverage to $\|\eta\| < 3.0$, discriminates single photons from π^0 decays, measures arrival time of electrons and photons

Table 3.3: Electromagnetic Calorimeter

4. Hadronic Calorimeter (HCAL)

Component	Materials/Composition	Function
HCAL Barrel (HB)	Plastic scintillators, brass absorbers	Measures hadronic showers and energy, covers $\|\eta\| < 1.3$
HCAL Endcap (HE)	Plastic scintillators, brass absorbers	Extends coverage to $\|\eta\| < 3.0$
HCAL Forward (HF)	Quartz fibers, steel absorbers	Measures energy and position of hadrons and jets in forward region, covers $3.0 < \|\eta\| < 5.0$

Table 3.4: Hadronic Calorimeter

5. Muon System

Component	Materials/Composition	Function
Drift Tubes (DT)	Four stations in barrel region, layers of drift cells	Detects muons in barrel, provides position and momentum measurement
Cathode Strip Chambers (CSC)	Four stations in endcap, multi-wire chambers	Measures muons in endcap, provides precise position measurements in high-radiation areas
Resistive Plate Chambers (RPC)	Covers entire barrel and endcap	Provides timing information and additional muon detection

Table 3.5: Muon System

6. Trigger and Data Acquisition System

Component	Materials/Composition	Function
Level-1 Trigger	Custom electronics, fast hardware processors	Reduces event rate from 40 MHz to 100 kHz
High-Level Trigger	Processor farm running algorithms	Further reduces event rate to ∼1 kHz for permanent storage

Table 3.6: Trigger and Data Acquisition System

3.1 Reconstruction

3.1.1 electron reconstruction

Electrons are reconstructed by merging data from the silicon tracker and electromagnetic calorimeter. The tracker reconstructs charged particle tracks using hits from silicon sensors, taking into account variations induced by bremsstrahlung radiation, particularly in the high-radiation forward zone. Energy clusters in the ECAL generated by electromagnetic showers are matched to extrapolation paths. Critical identifying criteria include the energy-to-momentum ratio (E/p), energy deposit form, and track-cluster matching quality. To reduce background noise from jets and other particles, isolation is used by calculating the total of transverse energy (E_T) in a cone surrounding the electron candidate. Energy calibration processes accommodate for detector non-uniformities and material interactions, adjusting the observed energy based on simulation and data.

3.1.2 jet reconstruction

Jets are reconstructed using particle-flow (PF) algorithms, which combine tracker, ECAL, and HCAL readings to identify and reconstruct individual particles. The anti-k_T algorithm clusters energy deposits from charged and neutral hadrons, photons, and other particles into jets. The distance parameter (R) is often 0.4 or 0.8, depending on the analysis. Jet energy corrections are used to account for pile-up effects and detector response nonlinearities. These corrections include offset corrections for pile-up, relative corrections for pseudorapidity (eta) dependence, and absolute corrections for overall energy scale calibration. Jets are recognized and cleaned based on the proportion of energy transported by neutral and charged components, which reduces the impact of detector noise and misreconstructed objects. DeepCSV techniques use displaced secondary vertices, impact parameter significance of tracks, and soft lepton features from heavy-flavor decays to identify b-jets with great efficiency and purity.

Chapter 4

Event Generation

An **event generator** is a computational tool used in particle physics to simulate collision events between particles, such as those produced in colliders like the Large Hadron Collider (LHC), Large Electron Positron Collider (LEP) etc. It provides a theoretical framework to predict the outcomes of particle interactions based on models like the Standard Model or extensions beyond it.

In high-energy physics, simulating events that mimic particle collisions observed in experiments is essential for theoretical prediction and data analysis. Monte Carlo (MC) event generators play a critical role in achieving this by producing events which are based on quantum mechanical principles and the dynamics of particle interactions. The process involves sampling the Lorentz Invariant Phase Space (LIPS) and weighting the configurations using the underlying physics processes, encoded in the matrix element $|M|^2$.

Here is a brief explaination of the theory and implementation of event generation, highlighting the role of phase space selection, random number generation, and the matrix element calculation in creating realistic simulations.

4.1 Lorentz Invariant Phase Space (LIPS)

The phase-space density for n-particle final states is defined as:

$$d\Phi_n(P; p_1, p_2, \ldots, p_n) = \left(\prod_{i=1}^{n} \frac{d^3 p_i}{(2\pi)^3 2p_i^0} \right) (2\pi)^4 \delta^{(4)} \left(P - \sum_{i=1}^{n} p_i \right), \tag{4.1}$$

where:

- P is the total four-momentum of the system,
- p_i are the outgoing four-momenta of the final-state particles.

The delta function ensures the conservation of energy and momentum. To simplify the analysis, the calculations are often performed in the center-of-mass (COM) frame, where $P = (\sqrt{s}; 0)$.

4.2 Random Number Generation and Phase Space Sampling

The foundation of Monte Carlo methods lies in random number generation. These random numbers are mapped to physical quantities such as particle energies, angles, momenta etc. This allows the generation of trial configurations that satisfy the kinematic constraints of the system.

4.2.1 Mapping Random Numbers to Kinematic Variables

For each event:

1. Random numbers $R_1, R_2, \ldots$ are generated from a uniform distribution between 0 and 1.
2. These numbers are transformed into physical quantities using appropriate map-

pings. For instance:

$$\phi = 2\pi R_1, \tag{4.2}$$

$$\cos\theta = 2R_2 - 1. \tag{4.3}$$

3. The magnitudes of energy and momentum are sampled based on the phase-space boundaries.

4.2.2 Acceptance-Rejection Method

To ensure that the generated configurations are consistent with the physics of the process, the acceptance-rejection method is employed:

1. First, the weight of each configuration is computed which is proportional to $|M|^2$, where $|M|^2$ is the squared matrix element which describes the interaction dynamics.
2. Generate a random number r uniformly in the range $[0, 1]$.
3. Accept the configuration if its weight exceeds $r \times |M|^2_{\text{max}}$, where $|M|^2_{\text{max}}$ is the maximum possible weight. Otherwise, reject the configuration and repeat the process.

Here, r introduces a probabilistic element, ensuring that configurations with higher weights are selected more frequently, aligning with their likelihood in nature.

4.3 Role of the Matrix Element $|M|^2$

The matrix element $|M|^2$ encapsulates the quantum mechanical probability of a specific interaction occurring. It depends on the underlying theory, such as Quantum Chromodynamics (QCD) or the Standard Model, and includes:

- Coupling constants,
- Propagators for intermediate particles,
- Kinematic factors.

For example, in the Drell-Yan process $q\bar{q} \to Z \to \ell^+\ell^-$, the matrix element includes terms related to the quark-antiquark annihilation, the Z-boson propagator, and the leptonic decay vertex. Its magnitude determines how likely a given phase-space configuration is to occur.

4.4 Event Generation Workflow

The process of event generation involves several interconnected steps that model the interactions of particles in high-energy collisions. Each stage is critical in ensuring that the generated events accurately represent the physical processes under investigation. The workflow is outlined below:

4.4.1 Initial State Configuration

The event generation begins with the definition of the initial state. This includes specifying the initial-state particles, such as protons in a collider experiment, and their corresponding energies. The momenta of the colliding partons are then sampled using parton distribution functions (PDFs). These PDFs describe the probability distributions of finding the quark and gluon constituents within the protons and it is a function of their momentum fractions. This configuration ensures that the input parameters reflect the conditions of the physical system being simulated.

4.4.2 Phase Space Sampling

In this stage, the phase space of the final-state particles is explored by generating random configurations of their momenta and angles. Techniques like Monte Carlo integration are employed to uniformly sample the phase space. Each configuration is constrained to satisfy fundamental conservation laws, such as energy and momentum conservation. This step creates a diverse set of possible outcomes for the particle interactions, laying the groundwork for the subsequent calculations.

4.4.3 Weighting by $|M|^2$

The next step involves calculating the weight of each phase space configuration. This weight is proportional to the square of the matrix element ($|M|^2$), which represents the quantum mechanical amplitude of the process under study. The acceptance-rejection method is applied to select configurations with probabilities corresponding to their weights. This ensures that the generated events accurately reflect the dynamics of the physical process.

4.4.4 Unweighting

After weighting, the events are processed to produce an unweighted sample suitable for practical analysis. Unweighting assigns an equal probability to each selected event, making the dataset statistically manageable. This is achieved by retaining events with weights above a certain threshold while discarding the others. This ensures the uniformity in distribution.

4.4.5 Showering and Hadronization

The selected events are then subjected to additional processes that occur after the primary interaction. Parton showering is simulated to account for gluon radiation and other secondary emissions, resulting in a cascade of particles. Next step is Hadronization where the partons form color-neutral hadrons. These steps are modeled using specialized algorithms, such as those implemented in software like PYTHIA.

4.4.6 Detector Simulation

The final step is simulating the experimental detector's response to the generated particles. This involves modeling the detector's efficiencies, resolutions, and acceptance to produce simulated data that mimic real experimental observations. The output of this step provides a direct comparison with actual measurements.

Each of these steps plays an essential role in producing reliable and realistic simulated events for particle physics research. By following this workflow, researchers can study complex phenomena and make meaningful comparisons with experimental data.

4.5 Conclusion

Monte Carlo event generators are indispensable tools in high-energy physics which enables detailed simulations of particle collisions and decays. By combining random number generation, phase-space sampling, and the physics encoded in the matrix element, they provide a robust framework for studying complex processes. The careful implementation of these methods provides accurate and efficient simulations and bridges the gap between theory and experiment.

Part III

Physics Analysis

Chapter 5

Problem statement

The goal of this project is to study the effect of the theoretical systematics on α_s. Since, Z boson p_T changes with the change in α_s, we are taking it as an observable to study the effects of uncertainties. We are analyzing the transverse momentum (p_T) distribution of the Z boson in $Z +$ jets processes. Specifically, across different jet multiplicity channels, including:

- exactly one jet and at least one jet channel
- exactly two jets, and at least two jets.

Observable and Systematics

Observable: The Z boson p_T is sensitive to the dynamics of jet production, it is used as a primary variable for being a probe of perturbative QCD.

Systematics: The theoretical uncertainties studied in this analysis are:

1. **LHE Scale Weight:** Uncertainty arising from variations in the renormalization scale (μ_R) and factorization scale (μ_F), which affect the cross section and higher-order corrections.
2. **Parton Shower (PS) Weight:** Uncertainty introduced by the parton shower model used to simulate the evolution of partons into jets.

3. **PDF Scale Weight:** Uncertainty associated with the parton distribution functions (PDFs), which describe the momentum distribution of quarks and gluons inside protons.

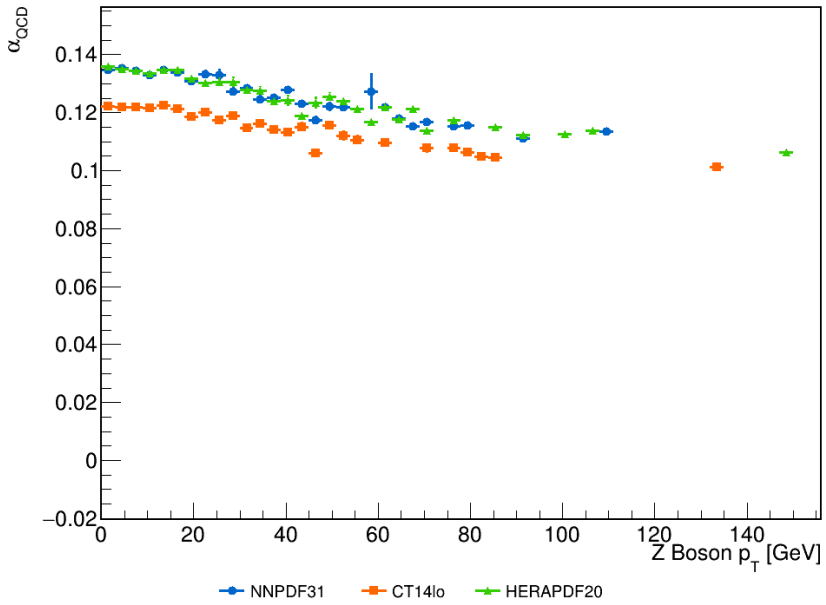


Figure 5.1: α_s as a function of Z p_T (generated by me using Madgraph)

Objective

The main objective of this project is to:

- Quantify the size of theoretical uncertainties for the Z boson p_T distribution in each jet multiplicity channel and select the channel which is less affected by the uncertainties for α_s related analysis.
- Selection of the channel which is least affected by the uncertainties for α_s related analysis.

Importance of the Study

- Z + jets processes are essential for testing perturbative QCD predictions and for understanding the jet production dynamics.

- The results serve as a benchmark for theoretical models and are crucial for reducing background uncertainties in searches for new physics.
- By identifying the jet multiplicity channel with the least systematic impact, this study contributes to improving the accuracy of cross section measurements and theoretical predictions.

Chapter 6

Data and Monte Carlo Samples

The data used for the analysis were collected from Runs A, B, C and D of the 2018 LHC run. The data used here is Egamma data set. The Monte Carlo samples used here are from 2018 UL simulations. The analysis is based on the file format `NANOAODSIM` using CMSSW release `CMSSW_12_4_11`.

For the simulation of the signal (`DYJetsToLL`), generated by me, I have used the sample generated with MadGraph5_aMC@NLO (shortened as MG5_aMC) using the ickkw merging scheme. The parton shower and hadronization is performed with Pythia. The matrix element includes Z + 0,1,2 jets at LO. The LO NNPDF31 (315000), HERAPDF20 (61000) and CT14LO_NF6 (13203) LHAPDFs are used for the ME calculation.

Label	**Full Name of data set**	**nEvents**
EGamma_A	/Run2018A-UL2018_MiniAODv2_NanoAODv9-v1/NANOAOD	339013231
EGamma_B	/Run2018B-UL2018_MiniAODv2_NanoAODv9-v1/NANOAOD	153792795
EGamma_C	/Run2018C-UL2018_MiniAODv2_NanoAODv9-v1/NANOAOD	147827904
EGamma_D	/Run2018D-UL2018_MiniAODv2_NanoAODv9-v3/NANOAOD	751683150

Table 6.1: Summary of EGamma datasets used in the analysis.

Monte Carlo sample Name	**nEvents**	**xsec(pb)**
/DYJetsToLL_M-50	1.3157245e+08	5675
/Single-top_t-channel_AntiTop_InclusiveDecays	90022642	67.9300
/TTToSemiLeptonic	4.7255763e+08	365.3400
/WWTo2L2Nu	9994000	11.0900
/WZTo3LNu	6482815	5.1970
/ZZTo2Q2L	29357938	3.7030

Table 6.2: Summary of Monte Carlo samples used in the analysis.

The total luminosity of the full data set is 59970 pb^{-1}

For Drell yan, Single top, TTBar, WZ samples weighted sum is used for the luminosity calculation purpore whereas for the other MC samples only sum of events are used due to the current unavailability of the samples in CMS DAS(Data Aggregation System). The drell yan process is the signal in my analysis.

6.1 Why the Z Boson is the Best Candidate for Measuring the Strong Coupling Constant (α_s)

The Z Boson is a pivotal particle in the study of fundamental interactions in particle physics, and its use in measuring the strong coupling constant (α_s) is particularly advantageous for several reasons:

6.1.1 Clean Signal Characteristics

The Z boson is produced in *pp* collisions (here *Drell yan* process is studied particularly), providing a well-defined initial state for the experiment. This setup leads to precise kinematic predictions for the decay products, resulting in what is referred to as a *clean signal*. Key aspects of this clean signal include:

1. **Dominance of Specific Decay Channels:** The Z Boson decays primarily into specific final states, such as lepton pairs (e.g., e^+e^-, $\mu^+\mu^-$) or quark-antiquark pairs, allowing for distinctive signatures that are easily isolated from background processes.

2. **Low Background Interference:** With a narrow width compared to other resonances, the Z Boson is less susceptible to background noise, enhancing the clarity of the observed signal. This allows for more accurate measurements and less uncertainty in data analysis.

6.1.2 Bridging Weak and Strong Interactions

While the Z boson mediates weak interactions, its decay into quarks and gluons makes it an important tool for studying strong interactions. The Z Boson interacts with quarks via electroweak coupling, and when it decays into quark-antiquark pairs, the resulting quarks and the gluons (often called **jet**) participate in strong interactions. This methodology offers a straightforward way to investigate QCD effects and obtain α_s.

6.1.3 Event Shape Variables

The hadronization of quarks generated in Z Boson decays may be studied using event shape variables, which are sensitive to the dynamics of QCD. These variables shed light on the strong coupling constant, allowing for its extraction using theoretical models that characterize the energy and momentum distribution of the ensuing hadronic final state.

6.1.4 Perturbative QCD Applicability

The energy scale associated with the Z Boson mass (about 91 GeV) is sufficient for perturbative quantum chromodynamics (pQCD) to be used. This permits precise theoretical predictions and comparisons with experimental data, allowing physicists to estimate α_s at this energy scale.

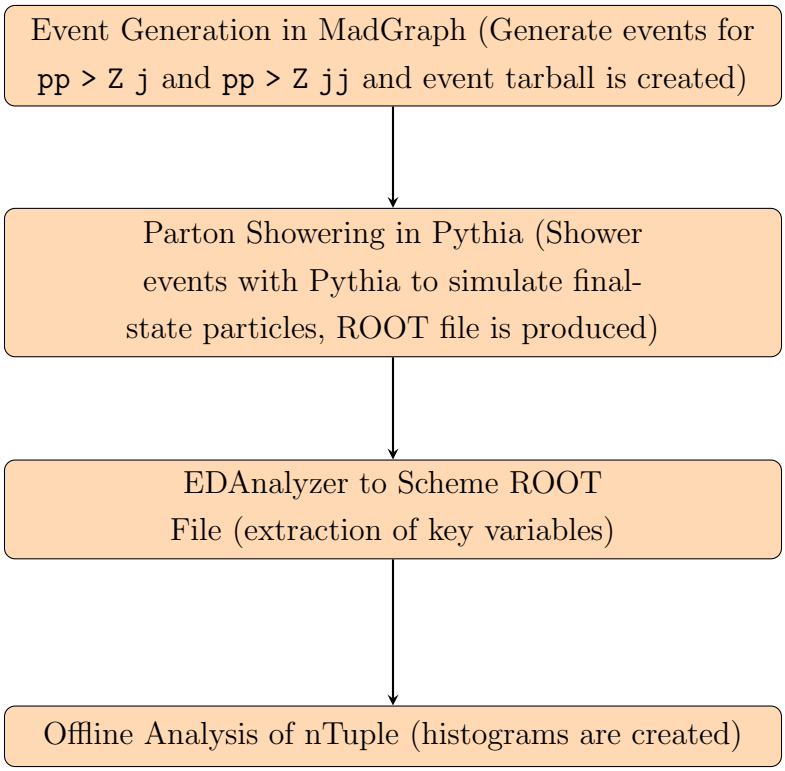
6.1.5 High Precision Measurements

Experiments using the Z Boson, especially at colliders like the LHC (massive Hadron Collider), have access to massive datasets, resulting in high accuracy measurements of α_s. The combination of clear signals and controlled systematic errors makes these experiments critical for proving the Standard Model's consistency and discovering novel physics.

6.1.6 Sample generation using Madgraph

To start with the analysis, CERN lxPlus (Linux Public Login User Service) account was created. Within lxPlus, CMSSW (Compact Muon Solenoid Software) environment was set up. CMSSW is a software framework developed for the CMS experiment at the Large

Hadron Collider. It is primarily used for data processing, simulation, and analysis of particle physics experiments. Afterwards, Madgraph is installed in lxPlus. Drell yan process is generated. The event generation takes around 10-15 minutes for 50000 events. Pythia is used after this step to hadronize the final state particles, this process also takes around four hours. Here we get a root file which is further schemed by CMSSW EDAnalyzer (Event Display Analyzer) which is C++ based script. Compilation and running of the EDAnalyzer takes around 5-7 minutes in total. after this step we get a reduced root file containing necessary information of the events, which is further analyzed by an offline C++ based root analysis code.



50000 events are generated at LO for each $pp \to Zj$ and $pp \to Zjj$ processes (j represents jet) using Madgraph NLO event generator at COM energy 13 TeV. Drell yan process is being considered here and electrons, positrons are studied as final state particles along with the jets. We get a Madgraph tarball that is later used for showering with pythia.

Since Hadronization is not done in Madgraph, the need to use another event generator, pythia comes into picture. After pythia showering we get hadronized final state particle and jets. All the information of the events are stored in a ROOT file which occupies a larger space. Hence, to scheme all the necessary information we use EDAnalyzer in

CMSSW framework.

EDAnalyzers are used to process and analyze the events recorded in the data files. They can be used to extract physics information from the event data, such as measuring particle properties, calculating cross-sections, or producing histograms for visualization. After this process, using CMSSW we create a ROOT nTuple (Monte Carlo file) with desired data and reduced size.

Now, an offline C++ analysis code is developed to analyse the nTuple (eg. applying some cuts, constructing another variable or function etc.) and get the desired histograms as output.

In this analysis to study the theory systematics, different nTuples are created for different processes and different parameters. NNPDF31 (two variations, α_s 0.118 and α_s 0.13), CT14lo and HERAPDF sets are used and for each of the PDF sets two nTuple files had been generated for one jet processes and two jet processes.

6.2 set up for sample generation

Below here the setup and configuration for simulating Drell-Yan processes in proton-proton collisions at the LHC using MadGraph5_aMC@NLO is described that I used. The primary focus is dilepton production with associated jets.

6.2.1 Process Card

The ‘proc_card.dat‘ defines:

- **Model:** Standard Model with CKM lepton masses.
- **Processes:** $pp \to \ell^+\ell^-$ (dilepton production), $pp \to \ell^+\ell^- j$, and $pp \to \ell^+\ell^- jj$
- **Output:** MLM matching is used to handle jet multiplicity without double-counting events.

6.2.2 Run Card

Key parameters in ‘run_card.dat‘:

- **Collision Energy:** 13 TeV proton-proton collisions.
- **Kinematic Cuts:**
  - $p_T > 20\,\text{GeV}$ for leptons and jets.
  - Invariant mass $M_{\ell\ell} > 50\,\text{GeV}$.
  - Pseudorapidity: $|\eta| < 5$.
- **PDF:** NNPDF31 (2 variations), HERAPDF, CT14LO
- **Event Count:** 50,000 unweighted events.

6.2.3 Applications

The simulation supports:

- Study of dilepton production and jet kinematics.
- Analysis of systematic uncertainties (PDF, scale variations and parton shower).
- Differential cross-section measurements.

6.2.4 Conclusion

The configurations in ‘proc_card.dat‘ and ‘run_card.dat‘ provide a robust framework for simulating Drell-Yan processes at the LHC, focusing on dilepton production and associated jets. This setup facilitates systematic studies and data comparison.

Chapter 7

Data and Monte Carlo comparison

7.1 Introduction

The comparison of experimental data with Monte Carlo (MC) simulations plays a vital role in validating theoretical models and understanding detector effects. This ensures the robustness of the analysis and helps quantify systematic uncertainties.

7.2 Event Selection and Cuts

The following selection criteria are applied to ensure high-quality events and reduce background contamination:

7.2.1 Electron Selection

Electrons are selected based on the following criteria:

- The event must contain at least two electrons.
- The transverse momentum (p_T) of each electron must exceed 20 GeV.
- Electrons must fall within the detector acceptance: $|\eta| < 1.4442$ (barrel region) or $1.566 < |\eta| < 2.5$ (endcap region).

7.2.2 Jet Selection

Jets are selected with the following requirements:

- Jets must have a transverse momentum (p_T) greater than 30 GeV.
- The pseudorapidity ($|\eta|$) of jets must be less than 2.4.
- Jets must satisfy the jet ID criterion (`Jet_jetId` ≥ 2).
- Jets must not overlap with selected electrons ($\Delta R > 0.4$).

7.2.3 Jet Multiplicity

Events are categorized based on jet multiplicity:

- Events with fewer than two jets or fewer than one jet are rejected.
- Exactly two jets and exactly one jet events are classified separately for specific analyses.

7.2.4 Z-Boson Reconstruction

The Z-boson is reconstructed from the two selected electrons. The invariant mass of the reconstructed Z-boson must lie within the range 60–120 GeV to ensure the signal region is isolated.

7.2.5 Weighting

For data events, a unit weight is assigned. For MC events, weights are computed using:

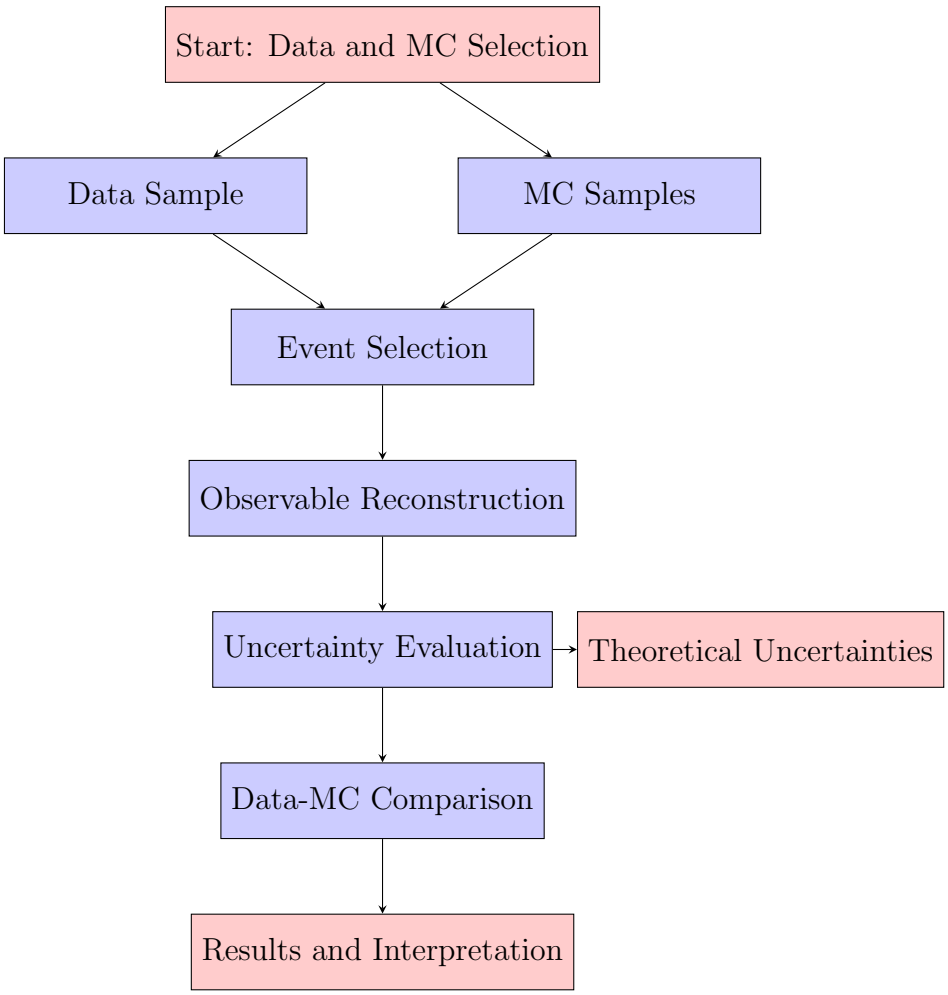
$$\text{Weight} = \pm \frac{\text{Data Luminosity}}{\text{MC Luminosity}}. \tag{7.1}$$

7.3 Comparison of Observables

Key observables, including Z-boson p_T, jet p_T, and jet multiplicities, are compared between data and MC simulations. The agreement validates the theoretical models and provides insight into systematic uncertainties. Any discrepancies are studied further to refine the analysis.

7.4 Conclusion

The data and MC comparison is a fundamental step in ensuring the reliability of the analysis. It validates the theoretical predictions, assesses detector effects, and identifies sources of systematic uncertainties, thereby contributing to precise and accurate measurements. Here is a simple flowchart for describing the whole process.



Chapter 8

Histogram binning in high energy particle Physics

Histograms are important tools in data analysis because they provide a visual depiction of data distributions. They divide the data range into intervals known as bins and count how many data points are in each. The height of each bin corresponds to the frequency of data points falling within that interval. The efficiency of a histogram in transmitting information about the underlying distribution is highly dependent on how the bins are defined.

Details of for the tested and best suited methods are explained herewith.

8.0.1 Sturges' Method

Mathematical Formula

Sturges' rule suggests that for a dataset with n observations, the optimal number of bins k is given by:

$$k = \lceil \log_2 n + 1 \rceil \tag{8.1}$$

where $\lceil \cdot \rceil$ denotes the ceiling function.

Implementation

Algorithm 1: Sturges' Method for Histogram Binning

[1] SturgesMethoddata $n \leftarrow$ length of data $k \leftarrow \lceil \log_2 n + 1 \rceil$ Calculate number of bins $x_{min} \leftarrow \min(\text{data})$ $x_{max} \leftarrow \max(\text{data})$ $h \leftarrow \frac{x_{max} - x_{min}}{k}$ Calculate bin width **return** k, h

Advantages and Limitations

Sturges' method is simple and widely implemented but tends to produce relatively few bins, particularly for large datasets. It assumes an approximately normal distribution of the data, making it less effective for skewed or multimodal distributions.

8.0.2 Doane's Formula

Mathematical Formula

Doane extended Sturges' rule to account for non-normal distributions by incorporating a skewness measure:

$$k = \left\lceil 1 + \log_2(n) + \log_2\left(1 + \frac{|g_1|}{\sigma_{g1}}\right) \right\rceil \tag{8.2}$$

where g_1 is the sample skewness and σ_{g1} is the estimated standard error of skewness:

$$\sigma_{g1} = \sqrt{\frac{6n(n-1)}{(n-2)(n+1)(n+3)}} \tag{8.3}$$

Implementation

Advantages and Limitations

Doane's formula improves upon Sturges' rule by accounting for non-normal distributions. It increases the number of bins for skewed distributions, making it more adaptive to

Algorithm 2: Doane's Formula for Histogram Binning

[1] DoaneMethoddata $n \leftarrow$ length of data Calculate skewness g_1 of data $\sigma_{g1} \leftarrow \sqrt{\frac{6n(n-1)}{(n-2)(n+1)(n+3)}}$ Standard error of skewness $k \leftarrow \lceil 1 + \log_2(n) + \log_2(1 + |g_1|/\sigma_{g1}) \rceil$ $x_{min} \leftarrow \min(\text{data})$ $x_{max} \leftarrow \max(\text{data})$ $h \leftarrow \frac{x_{max} - x_{min}}{k}$ **return** k, h

various data shapes. However, it remains less effective for multimodal distributions and very large datasets.

8.0.3 Scott's Normal Reference Rule

Mathematical Formula

Scott's rule focuses on determining the optimal bin width h rather than the number of bins:

$$h = 3.49 \times \sigma \times n^{-1/3} \tag{8.4}$$

where σ is the sample standard deviation. The number of bins k can then be calculated as:

$$k = \left\lceil \frac{x_{max} - x_{min}}{h} \right\rceil \tag{8.5}$$

Implementation

Algorithm 3: Scott's Normal Reference Rule for Histogram Binning

[1] ScottMethoddata $n \leftarrow$ length of data $\sigma \leftarrow$ standard deviation of data $h \leftarrow 3.49 \times \sigma \times n^{-1/3}$ Calculate bin width $x_{min} \leftarrow \min(\text{data})$ $x_{max} \leftarrow \max(\text{data})$ $k \leftarrow \lceil \frac{x_{max} - x_{min}}{h} \rceil$ Calculate number of bins **return** k, h

Theoretical Foundation

Scott's rule is derived by minimizing the integrated mean squared error (IMSE) between the histogram and the underlying probability density function, under the assumption that the true density is normal:

$$\text{IMSE}(h) = E\left[\int (\hat{f}(x) - f(x))^2 dx\right] \tag{8.6}$$

where $\hat{f}$ is the histogram estimate and f is the true density.

Advantages and Limitations

Scott's rule is optimal for normally distributed data and scales appropriately with sample size. It offers a statistically principled approach that adapts to the variability in the data. However, it may not perform optimally for non-normal or multimodal distributions.

8.0.4 Freedman-Diaconis Rule

Mathematical Formula

The Freedman-Diaconis rule determines bin width based on the interquartile range (IQR):

$$h = 2 \times \text{IQR} \times n^{-1/3} \tag{8.7}$$

where IQR is $Q_3 - Q_1$, the difference between the third and first quartiles. The number of bins is then:

$$k = \left\lceil \frac{x_{max} - x_{min}}{h} \right\rceil \tag{8.8}$$

Algorithm 4: Freedman-Diaconis Rule for Histogram Binning

[1] FDMethoddata $n \leftarrow$ length of data Sort data Calculate first quartile Q_1 and third quartile Q_3 IQR $\leftarrow Q_3 - Q_1$ Interquartile range $h \leftarrow 2 \times \text{IQR} \times n^{-1/3}$ Calculate bin width $x_{min} \leftarrow \min(\text{data})$ $x_{max} \leftarrow \max(\text{data})$ $k \leftarrow \lceil \frac{x_{max} - x_{min}}{h} \rceil$ Calculate number of bins **return** k, h

Implementation

Advantages and Limitations

The Freedman-Diaconis rule is robust against outliers due to its use of the IQR instead of standard deviation. It works well for skewed and non-normal distributions. However, it may produce a large number of bins for complex multimodal distributions and requires slightly more computation than some other methods.

8.0.5 Knuth's Rule

Mathematical Approach

Knuth's rule selects the number of bins k that maximizes the posterior probability of the histogram model:

$$\log(p(k|x)) = n\log(k) - k\log(n) + \log(\Gamma(k)) - \log(\Gamma(n+k)) + \sum_{i=1}^{k} \log(\Gamma(n_i + 1)) \quad (8.9)$$

where $p(k|x)$ is the posterior probability of k bins given the data x, n is the sample size, n_i is the count in the i-th bin, and Γ is the Gamma function.

Implementation

In practice, Knuth's rule requires searching across a range of possible bin numbers to find the one that maximizes the posterior probability. This is typically done using numerical optimization.

Algorithm 5: Simplified Knuth's Rule for Histogram Binning

[1] KnuthMethoddata $n \leftarrow$ length of data $x_{min} \leftarrow \min(\text{data})$ $x_{max} \leftarrow \max(\text{data})$ PosteriorProbability$k$, data Create k equal-width bins covering $[x_{min}, x_{max}]$ Count data points in each bin: $n_1, n_2, \ldots, n_k$ $\log p \leftarrow n \log(k) - k \log(n) + \log \Gamma(k) - \log \Gamma(n+k)$ **for** $i = 1$ *to* k **do**

$\log p \leftarrow \log p + \log \Gamma(n_i + 1)$ **return** $\log p$ $k_{optimal} \leftarrow \arg\max_k \text{PosteriorProbability}(k, \text{data})$ for k in range $h \leftarrow \frac{x_{max} - x_{min}}{k_{optimal}}$ Calculate bin width **return** $k_{optimal}, h$

Advantages and Limitations

Knuth's rule is data-driven and makes no assumptions about the underlying distribution. It adapts well to complex distributional features and is especially good at identifying multimodality. However, it is computationally intensive and more complex to implement than other methods.

8.0.6 Rice Rule

Mathematical Formula

The Rice rule is a simple rule of thumb:

$$k = \lceil 2 \times n^{1/3} \rceil \tag{8.10}$$

where n is the sample size.

Implementation

Algorithm 6: Rice Rule for Histogram Binning

[1] RiceMethoddata $n \leftarrow$ length of data $k \leftarrow \lceil 2 \times n^{1/3} \rceil$ Calculate number of bins $x_{min} \leftarrow \min(\text{data})$ $x_{max} \leftarrow \max(\text{data})$ $h \leftarrow \frac{x_{max} - x_{min}}{k}$ Calculate bin width **return** k, h

Advantages and Limitations

The Rice rule is simple to calculate and generally produces reasonable bin counts. It works adequately for moderately sized datasets but does not adapt to the specific characteristics of the distribution.

8.0.7 Square Root Choice

Mathematical Formula

The square root choice is one of the simplest binning rules:

$$k = \lceil \sqrt{n} \rceil \tag{8.11}$$

where n is the sample size.

Implementation

Algorithm 7: Square Root Choice for Histogram Binning

[1] SqrtMethoddata $n \leftarrow$ length of data $k \leftarrow \lceil \sqrt{n} \rceil$ Calculate number of bins $x_{min} \leftarrow \min(\text{data})$ $x_{max} \leftarrow \max(\text{data})$ $h \leftarrow \frac{x_{max} - x_{min}}{k}$ Calculate bin width **return** k, h

Advantages and Limitations

The square root choice is extremely simple to calculate and provides a quick default. However, it lacks theoretical justification and does not account for distribution characteristics, potentially producing too many bins for large datasets.

8.0.8 Equal Population (Quantile-Based) Binning

Mathematical Approach

Unlike width-based methods, equal population binning creates variable-width bins that contain approximately the same number of data points:

1. Sort the data in ascending order: $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$ 2. For a desired number of bins k, select bin edges at quantiles: $q_j = x_{(\lfloor jn/k \rfloor)}$ for $j = 1, 2, \ldots, k-1$ 3. Use x_{min} and x_{max} as the first and last edges

Implementation

Algorithm 8: Equal Population Binning for Histograms

[1] EqualPopMethoddata, k $n \leftarrow$ length of data Sort data bin_edges $\leftarrow [x_{min}]$ Initialize with minimum value **for** $j = 1$ *to* $k - 1$ **do**

index $\leftarrow \lfloor j \times n/k \rfloor$ Append data[index] to bin_edges Append x_{max} to bin_edges Add maximum value **return** bin_edges

Advantages and Limitations

Equal population binning ensures balanced representation across the histogram and prevents bins with very few observations. It is robust against outliers and useful for skewed distributions. However, variable bin widths can be harder to interpret and may obscure the true shape of the distribution.

8.0.9 Bayesian Blocks

Mathematical Approach

Bayesian Blocks is an adaptive binning algorithm that identifies structure in the data by optimizing a fitness function derived from Bayesian principles. The algorithm:

1. Starts with the finest binning possible (one bin per unique value) 2. Iteratively merges adjacent bins to maximize a fitness function 3. Uses dynamic programming to find the optimal partitioning

A commonly used fitness function is:

$$\text{Fitness} = \sum_{i=1}^{k} n_i \log\left(\frac{n_i}{w_i}\right) \tag{8.12}$$

where n_i is the count in bin i and w_i is the width of bin i.

Implementation

The implementation of Bayesian Blocks is complex and typically involves dynamic programming. Below is a simplified algorithm outline:

Algorithm 9: Simplified Bayesian Blocks for Histogram Binning

[1] BayesianBlocksdata Sort unique data values: $x_1, x_2, \ldots, x_m$ Initialize fitness array F and last changepoint array L $F[0] \leftarrow 0$ Base case: fitness of empty data **for** $i = 1$ *to* m **do**

$F[i] \leftarrow -\infty$ Initialize with worst fitness **for** $j = 0$ *to* $i - 1$ **do**

Compute fitness of bin $(x_j, x_i]$: $f_{j,i}$ $f_{current} \leftarrow F[j] + f_{j,i}$ **if** $f_{current} > F[i]$ **then**

$F[i] \leftarrow f_{current}$ $L[i] \leftarrow j$ Record last changepoint Reconstruct optimal bin edges using L array **return** optimal_bin_edges

Advantages and Limitations

Bayesian Blocks is adaptive to complex structure in the data and can identify multimodality and sharp transitions. It provides statistically optimal bin locations without assuming uniformity within bins. However, it is computationally intensive, complex to implement, and may be sensitive to noise in the data.

Here in my analysis, I have tried all the different methods listed above and selected the Bayesian block method for binning. Here is a comparative depiction of different binning methods on the observable I am using in my analysis.

Here is a comparative analysis of different methods applied on Z p_T for my analysis.

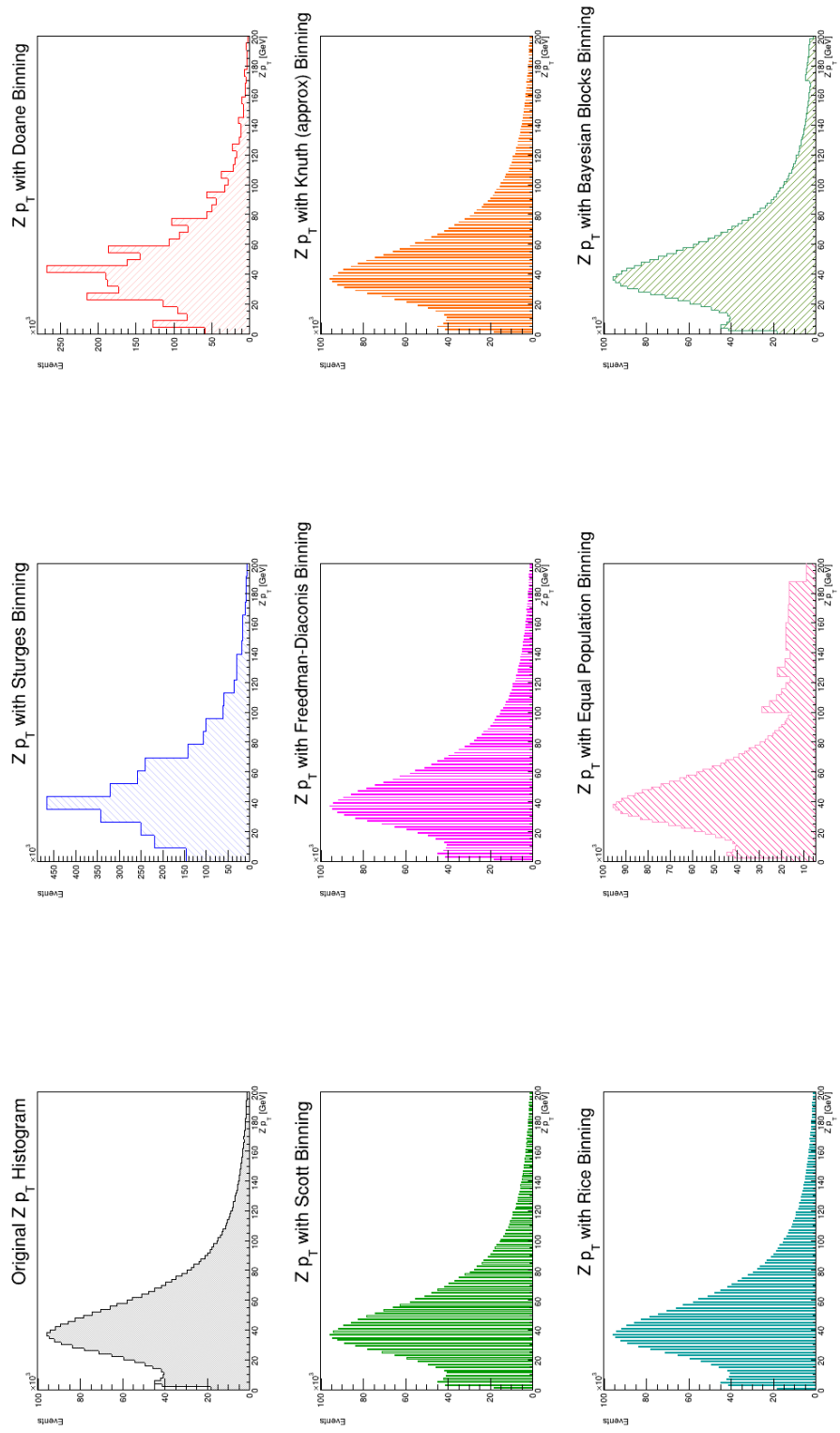


Figure 8.1: Comparison of Z p_T histograms using different binning techniques.

Chapter 9

Introduction to Event Weights

In particle physics simulations, event weights are crucial for accurately representing theoretical distributions. Each simulated event corresponds to a specific process, and the event weight scales the contribution of the event to ensure that the results are normalized to the expected physical distribution. This normalization takes into account the theoretical cross section of the process, the desired luminosity, and any additional factors such as systematic uncertainties.

9.1 An Easy-to-Understand Example

To better understand the concept of weights, consider the following example.

Rolling a Weighted Die

Imagine simulating the process of rolling a die. However, instead of a fair die, each number from 1 to 6 appears with a different probability. These probabilities and observed frequencies from 1000 rolls are shown in Table 9.1.

Applying Weights

To accurately represent the expected probabilities of each outcome in a histogram, you can assign weights to each result. The weight for each outcome is calculated as:

Outcome	True Probability (%)	Observed Frequency
1	10	90
2	20	180
3	30	300
4	10	110
5	20	200
6	10	120

Table 9.1: Probabilities and Observed Frequencies for a Weighted Die

$$w = \frac{\text{True Probability}}{\text{Observed Frequency}}$$

For each number:

$$\text{Weight for rolling a 1} = \frac{0.1}{\frac{90}{1000}} = 1.11$$
$$\text{Weight for rolling a 2} = \frac{0.2}{\frac{180}{1000}} = 1.11$$
$$\text{Weight for rolling a 3} = \frac{0.3}{\frac{300}{1000}} = 1.0$$

And so on.

Filling the Histogram

To fill the histogram, multiply the observed count for each outcome by its corresponding weight:

- Contribution of 1: $90 \times 1.11 = 100$
- Contribution of 2: $180 \times 1.11 = 200$
- Contribution of 3: $300 \times 1.0 = 300$

This process ensures that the histogram reflects the true probabilities of the weighted die, even though the observed frequencies may differ.

9.2 Relevance in Particle Physics

In particle physics, weights play a similar role. Each simulated event has a specific weight that scales its contribution to observables such as histograms or distributions. These weights account for theoretical cross sections, systematic variations, and normalization to experimental conditions. By applying weights correctly, physicists ensure a robust and accurate comparison between data and theoretical predictions. The event weight is multiplied by the variable's contribution to ensure proper scaling.

9.3 Types of Weights

In high-energy physics, different types of weights are used to capture various aspects of the simulation and uncertainties. Below are some commonly used weights:

9.3.1 Generator Weight (genWeight)

This is the baseline weight assigned to an event during the generation step. It ensures proper normalization of events to match the total cross section of the process. For next-to-leading-order (NLO) simulations, `genWeight` can also account for negative weights, which arise to cancel unphysical regions and improve theoretical accuracy.

9.3.2 PDF Weights (PDFscaleweight)

PDF weights encode uncertainties in the parton distribution functions (PDFs), which describe the momentum distributions of quarks and gluons inside protons. By varying the eigenvectors of the PDF set, multiple weights are provided for each event, allowing the estimation of theoretical uncertainties.

9.3.3 Scale Weights (LHEScaleWeight)

Scale weights are associated with variations in the renormalization (μ_R) and factorization (μ_F) scales. These weights capture the uncertainty in the perturbative QCD calculations

by varying μ_R and μ_F up or down by a factor of two. For example, typical variations include $(\mu_R, \mu_F) = (1, 1), (0.5, 0.5), (2, 2)$.

9.3.4 Parton Shower Weights (PS Weights)

Parton shower weights capture uncertainties in the modeling of the parton shower, which describes the evolution of partons into observable hadrons. These weights allow variations in parameters such as the strong coupling constant (α_s) and matching scales.

9.4 Combined Weights

The total weight of an event is typically a combination of these individual weights:

$$W_{\text{total}} = \text{genWeight} \times W_{\text{PDF}} \times W_{\text{scale}} \times W_{\text{PS}}$$

Proper handling of these weights ensures accurate normalization and robust uncertainty estimation.

9.5 Conclusion

Weights are an essential component of Monte Carlo simulations, enabling precise normalization and the inclusion of theoretical uncertainties. By understanding and applying these weights correctly, physicists can achieve more accurate comparisons between data and theoretical predictions, ultimately leading to a better understanding of particle physics processes.

Chapter 10

LHE scale uncertainties

This chapter provides a detailed explanation of the systematic uncertainties Z + jets analysis, focusing on the roles of LHE weights, μ_F and μ_R scales, and their associated variations. It also explains how these systematics contribute to uncertainties in my case and why they are important to study.

10.1 LHE Weights

Les Houches Event (LHE) weights are additional weights stored in the event record to estimate the impact of variations in theoretical parameters. These weights allow you to evaluate systematic uncertainties without regenerating events for different parameter configurations.

LHE weights are calculated at the **matrix-element (ME) level**, representing the probability amplitude for a given scattering process. The primary variations include:

- **Factorization scale (μ_F)**: Related to parton distribution functions (PDFs).

- **Renormalization scale (μ_R)**: Governs the QCD coupling strength (α_s).

10.2 μ_F and μ_R Scales

10.2.1 Factorization Scale (μ_F)

Definition: Factorization is a technique, used to separate different contribution to a process particularly when dealing with hadrn collision. Such processes include short distance and long distance interactions. Short distance is calculated by QCD and the long distance is modeled using PDEs (Parton Distribution Functions).

μ_F scale is a scale where we decide to cut off the perturbative part and start using PDFs for the non-perturbative parts.

Physical Meaning:

- At μ_F, soft and collinear QCD radiation is absorbed into the PDFs, ensuring finite cross-sections.
- The choice of μ_F determines how much radiation is treated as part of the PDFs versus the explicit matrix element (ME) calculation.

10.2.2 Renormalization Scale (μ_R)

Definition: Renormalization is a systematic procedure to deal with the infinities in QCD calculations.It absorbs infinities in the redefinition of the theory parameters like mass, coupling constant, fied etc.

μ_R is where renormalized coupling constant is defined. Changing this parameter changes α_s reflecting the running of the QCD coupling constant (α_s).

Physical Meaning:

- At μ_R, the perturbative calculation is truncated at a given order, leaving a residual dependence on μ_R.
- The value of μ_R determines the effective strength of α_s, which governs gluon emission probabilities.

10.2.3 Nominal Scales and Variations

- **Nominal Scales**: Typically, μ_F and μ_R are set to the transverse momentum (p_T) of the hard process or the invariant mass of the system (e.g., Z-boson mass).
- **Scale Variations**: To estimate theoretical uncertainties, these scales are varied independently or simultaneously by factors of 0.5 and 2. Common combinations include:

$$(\mu_F, \mu_R) = (1,1), (0.5,0.5), (2,2), (0.5,1), (1,0.5), (2,1), (1,2).$$

  (0.5, 2) and (2, 0.5) are often excluded to avoide extreme scinarios.

10.3 Systematic Uncertainties in My Analysis

The systematic uncertainties in this analysis arise primarily from variations in the factorization (μ_F) and renormalization (μ_R) scales, which influence multiple aspects of the $Z+$jets process. These scales affect QCD radiation, modifying cross-sections and altering predictions for jet multiplicities and Z-boson p_T distributions. The transverse momentum of the Z-boson, reliant on jet recoil, shifts with variations in μ_F and μ_R, while higher μ_R values promote stronger gluon emissions, increasing jet counts, and lower μ_F values suppress emissions, impacting predictions. Additionally, these scale variations affect both the normalization and the shapes of key distributions, such as Z-boson p_T, jet p_T, and event shape variables, contributing to systematic uncertainty bands in the analysis.

10.4 Sources of Uncertainty

The primary sources of uncertainty associated with the factorization (μ_F) and renormalization (μ_R) scales arise from several factors. A significant source is the truncation of perturbative QCD calculations, which leads to residual dependence on μ_F and μ_R due to the absence of higher-order corrections. Additionally, scale variations influence the matching between matrix elements and parton showers, affecting the accuracy of the event generation process. These variations also impact event kinematics, particularly the jet properties and the Z-boson p_T, which are crucial observables in this analysis. Furthermore, the number of jets produced in the event, known as jet multiplicities, is highly

sensitive to QCD radiation modeling, adding another layer of uncertainty to the analysis. Together, these factors contribute to the systematic uncertainties in the study of $Z+$jets processes.

10.5 Why Study These Systematics?

Studying these systematic uncertainties is essential for making accurate predictions in $Z+$jets processes and ensuring meaningful comparisons between Monte Carlo simulations and experimental data. Scale variations help establish error bands, quantify systematic effects, and identify jet multiplicity channels with minimal sensitivity. This analysis also aids in improving Monte Carlo generators, enhancing their predictive accuracy and alignment with experimental observations.

10.6 Results

It is observed that the "at least one jet" ($N_{\text{jets}} \geq 1$) channel is less affected by systematic uncertainties compared to the "exactly one jet" ($N_{\text{jets}} = 1$) channel. But there is no great difference between the "at least two jets" ($N_{\text{jets}} \geq 2$) channel and the "exactly two jets" ($N_{\text{jets}} = 2$) channel. This is because the "at least one" channel includes a broader range of events, which helps average out extreme variations caused by theoretical uncertainties such as renormalization and factorization scale variations or parton shower effects. In contrast, the "exactly one or two jet" channel, being more exclusive, depends heavily on specific configurations, making it relatively more sensitive to uncertainties in QCD modeling and higher-order corrections.

here is a reference table for the scale variations those are mentioned in the plots

Weight Index	μ_F (MUF)	μ_R (MUR)	Description
0	0.5	0.5	$\mu_F = 0.5$, $\mu_R = 0.5$
1	1.0	0.5	$\mu_F = 1.0$, $\mu_R = 0.5$
2	2.0	0.5	$\mu_F = 2.0$, $\mu_R = 0.5$
3	0.5	1.0	$\mu_F = 0.5$, $\mu_R = 1.0$
4	1.0	1.0	$\mu_F = 1.0$, $\mu_R = 1.0$
5	2.0	1.0	$\mu_F = 2.0$, $\mu_R = 1.0$
6	0.5	2.0	$\mu_F = 0.5$, $\mu_R = 2.0$
7	1.0	2.0	$\mu_F = 1.0$, $\mu_R = 2.0$
8	2.0	2.0	$\mu_F = 2.0$, $\mu_R = 2.0$

Table 10.1: LHE Scale Variation Weights and Corresponding μ_F and μ_R Values

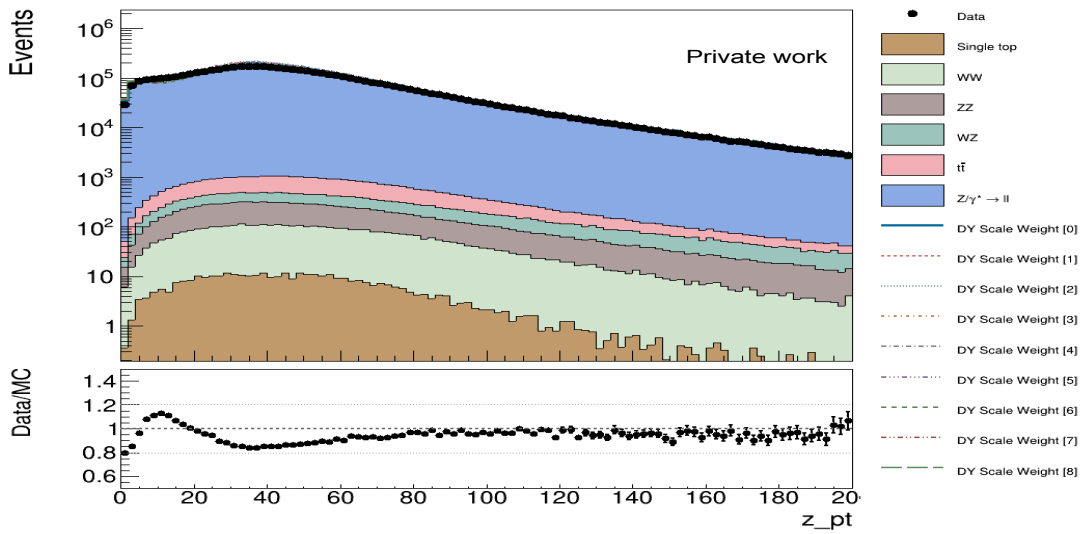


Figure 10.1: Z $_T$ at least one jet plot with uncertainties for DY sample

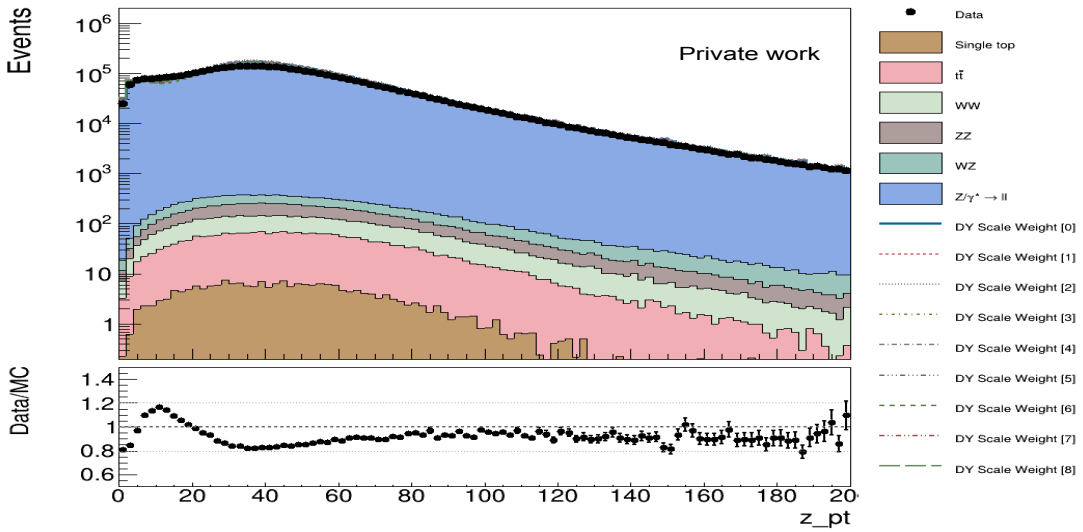


Figure 10.2: Z $_T$ exactly one jet plot with uncertainties for DY sample

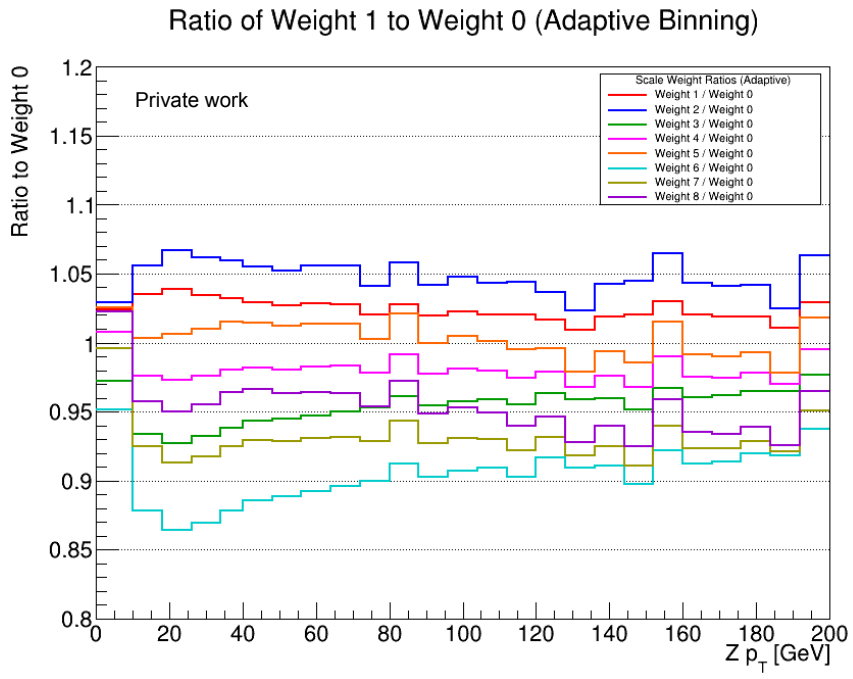


Figure 10.3: Ratio plot of the uncertainties with respect to the nominal value for exactly one jet channel DY sample

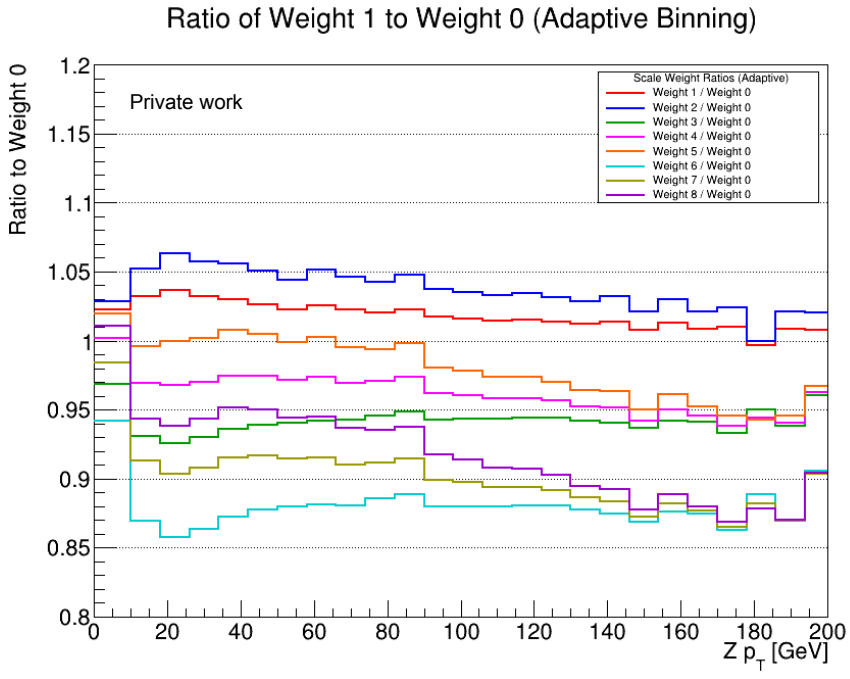


Figure 10.4: Ratio plot of the uncertainties with respect to the nominal value for at least one jet channel DY sample

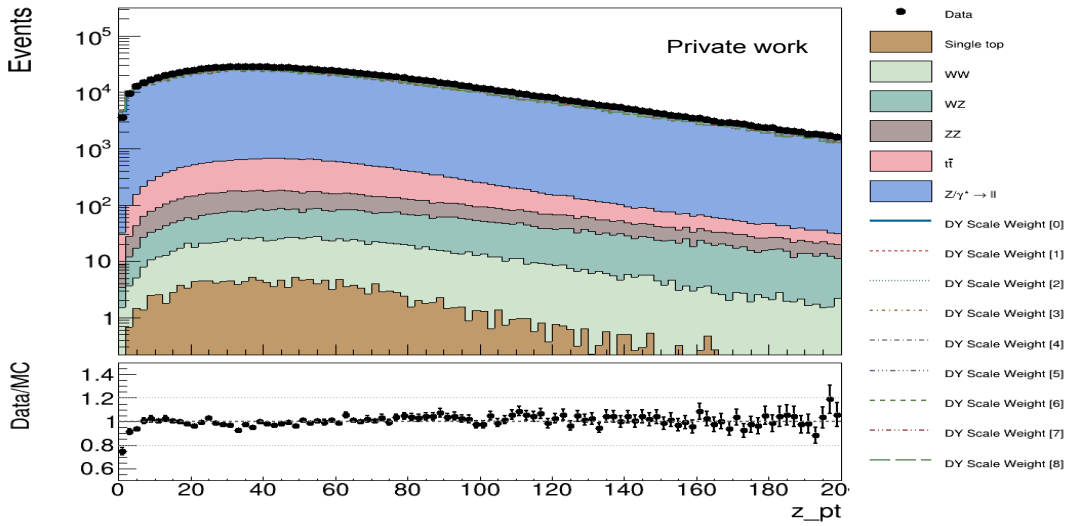


Figure 10.5: Z $_T$ at least two jets plot with uncertainties for DY sample

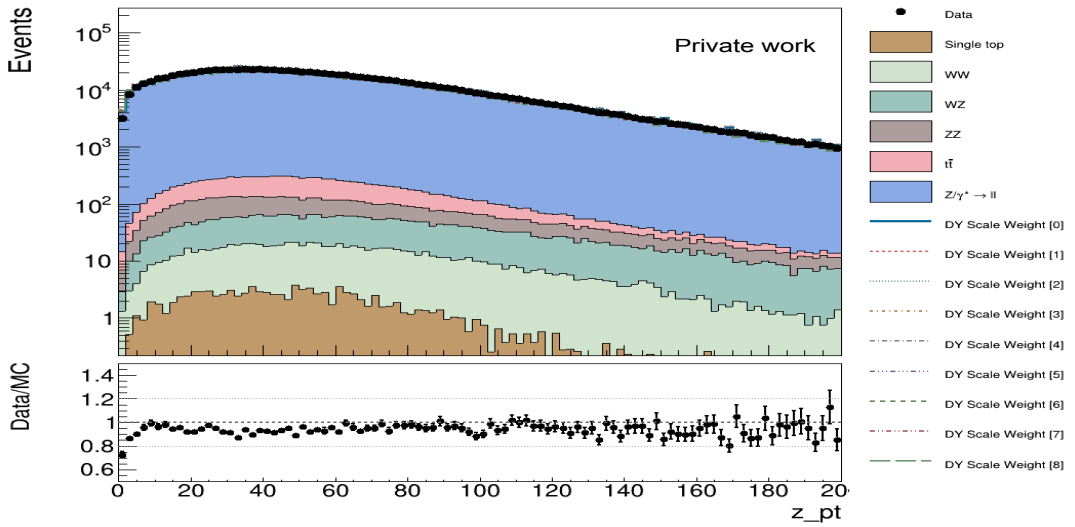


Figure 10.6: Z $_T$ exactly two jets plot with uncertainties for DY sample

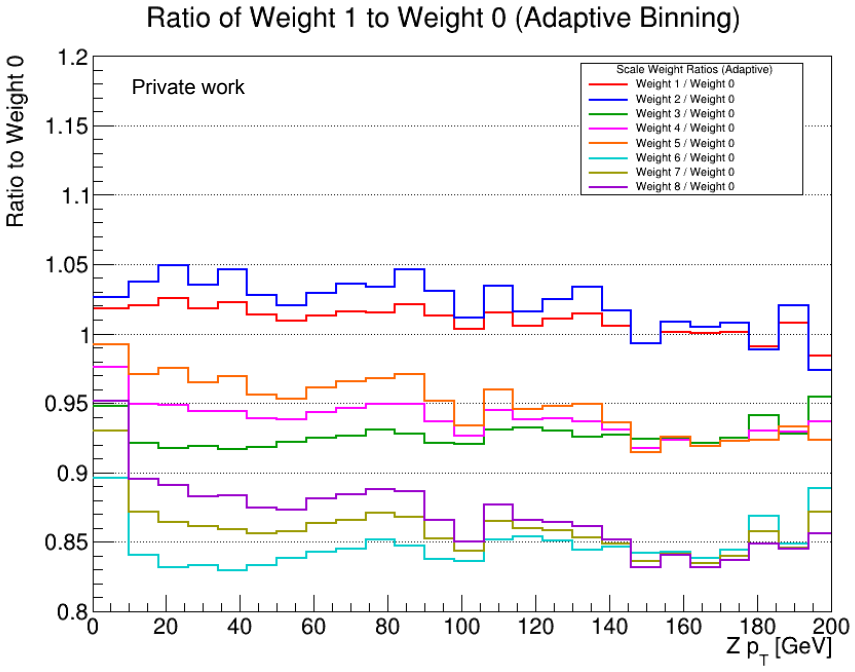


Figure 10.7: Ratio plot of the uncertainties with respect to the nominal value for exactly two jets channel for DY sample

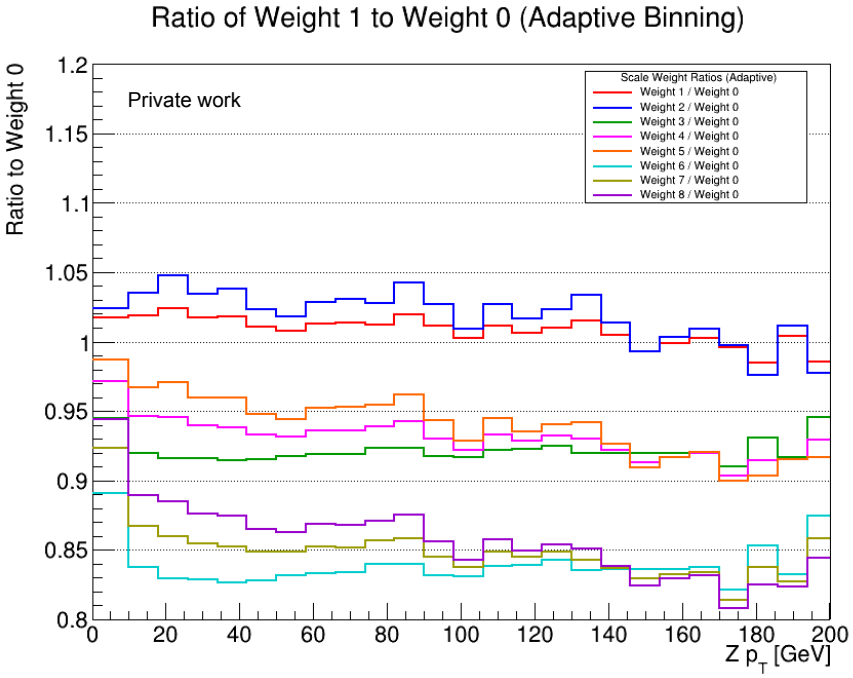


Figure 10.8: Ratio plot of the uncertainties with respect to the nominal value for at least two jets channel for DY sample

Chapter 11

PS scale weight uncertainties

Parton showers simulate the radiation of quarks and gluons from partons during high-energy collisions, bridging the gap between matrix-element calculations and the formation of jets observed in detectors. The parton shower consists of:

- **Initial State Radiation (ISR):** Radiation emitted by incoming partons before the hard scattering, affecting the transverse momentum (p_T) balance of the event.

- **Final State Radiation (FSR):** Radiation emitted by outgoing partons after the hard scattering, influencing jet shapes and energy distributions.

11.0.1 Scale Variations and Their Role

To estimate the uncertainties associated with parton shower modeling, systematic variations in the Initial State Radiation (ISR) and Final State Radiation (FSR) scales are applied. In this analysis, the following combinations of scale factors are considered: (ISR, FSR) = $(2, 1), (1, 2), (0.5, 1), (1, 0.5)$. These variations reflect changes where either the ISR or FSR scale is increased or decreased while the other is held constant. Such variations enable a systematic study of their effect on key observables like the Z-boson transverse momentum (p_T) and jet distributions.

11.0.2 Sources of Uncertainty

The primary sources of uncertainty in this context include: **ISR and FSR Variations:** These influence the radiation patterns, jet structures, and the p_T distribution of the Z-boson and associated jets. ISR variations mainly affect the momentum balance and initial system configurations, while FSR variations impact the energy distribution and jet multiplicity.

11.0.3 Systematics in This Analysis

In this analysis, ISR (Initial State Radiation) and FSR (Final State Radiation) variations are examined as a significant source of systematic uncertainty impacting the $Z+$jets process. These variations arise from the modeling of QCD radiation during parton showers, which simulate the emission of gluons and quarks before and after hard scattering events. By varying ISR and FSR scales, systematic uncertainty bands are generated for key observables, such as the Z-boson p_T distribution and jet multiplicities.

The scale variations considered, such as (ISR, FSR) combinations of (2,1), (1,2), (0.5,1), and (1,0.5), allow for the evaluation of uncertainties associated with the softness or hardness of QCD radiation. These systematic effects are particularly relevant for precision measurements and comparisons between Monte Carlo predictions and experimental data.

11.0.4 Why Study These Systematics?

Parton shower (PS) uncertainties are crucial for understanding the reliability of theoretical predictions in $Z+$ jets processes. PS models simulate QCD radiation, including Initial State Radiation (ISR) and Final State Radiation (FSR), which significantly influence key observables such as the Z-boson p_T and jet multiplicities. Variations in ISR and FSR scales, such as (2,1), (1,2), (0.5,1), and (1,0.5), are used to estimate systematic uncertainty bands, providing error margins for distributions.

Studying PS uncertainties complements LHE scale uncertainties by addressing soft and collinear QCD effects. It helps identify channels with reduced sensitivity to these variations, such as "at least 1 jet" or "at least 2 jets," ensuring robust comparisons between data and simulations. Additionally, this analysis highlights areas for improving PS models in Monte Carlo generators, enhancing the accuracy of theoretical predictions.

11.0.5 Results

In this analysis, it was observed that the "at least one jet" ($N_{\text{jets}} \geq 1$) and "at least two jets" ($N_{\text{jets}} \geq 2$) channels are less sensitive to PS scale uncertainties compared to the "exactly one jet" ($N_{\text{jets}} = 1$) and "exactly two jets" ($N_{\text{jets}} = 2$) channels. This is because inclusive channels average out variations over a broader event sample, reducing the impact of ISR and FSR scale uncertainties. The study of these uncertainties helps improve the reliability of Monte Carlo predictions and ensures meaningful comparisons with experimental data.

Here is a reference for different ISR, FSR scale factor combinations those are used in the plot

Weight Index	ISR	FSR	Description
0	2	1	ISR = 2, FSR = 1
1	1	2	ISR = 1, FSR = 2
2	0.5	1	ISR = 0.5, FSR = 1
3	1	0.5	ISR = 1, FSR = 0.5

Table 11.1: PS Weight Variations for ISR and FSR Values

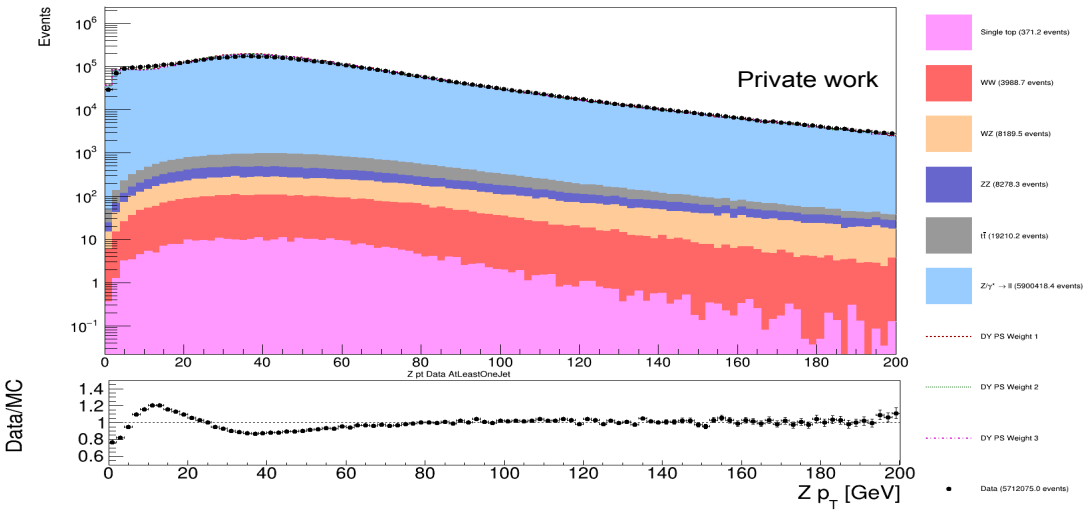


Figure 11.1: Z $_T$ at least one jet plot with uncertainties for DY sample

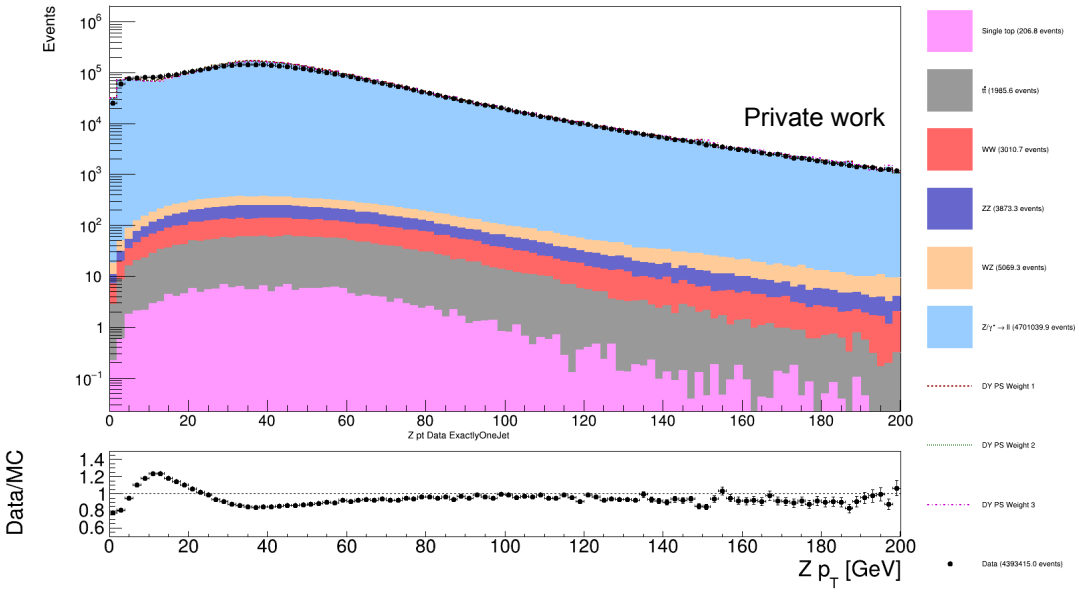


Figure 11.2: Z $_T$ exactly one jet plot with uncertainties for DY sample

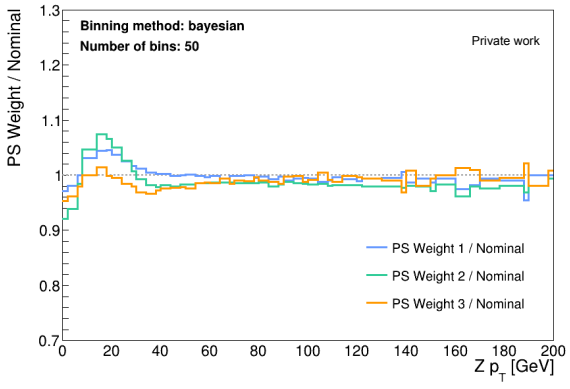


Figure 11.3: Ratio plot of uncertainties for at least one jet channel DY sample

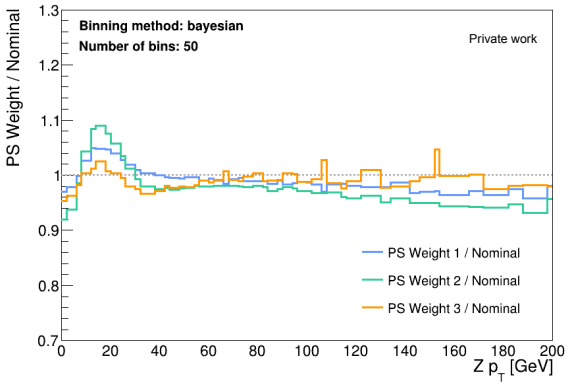


Figure 11.4: Ratio plot of uncertainties for exactly one jet channel DY sample

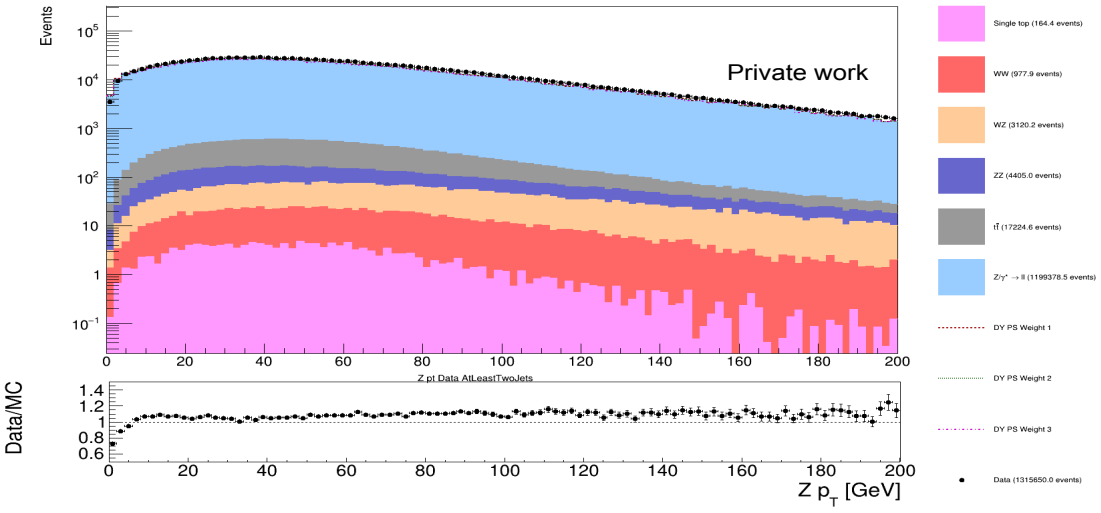


Figure 11.5: Z $_T$ at least two jets plot with uncertainties for DY sample

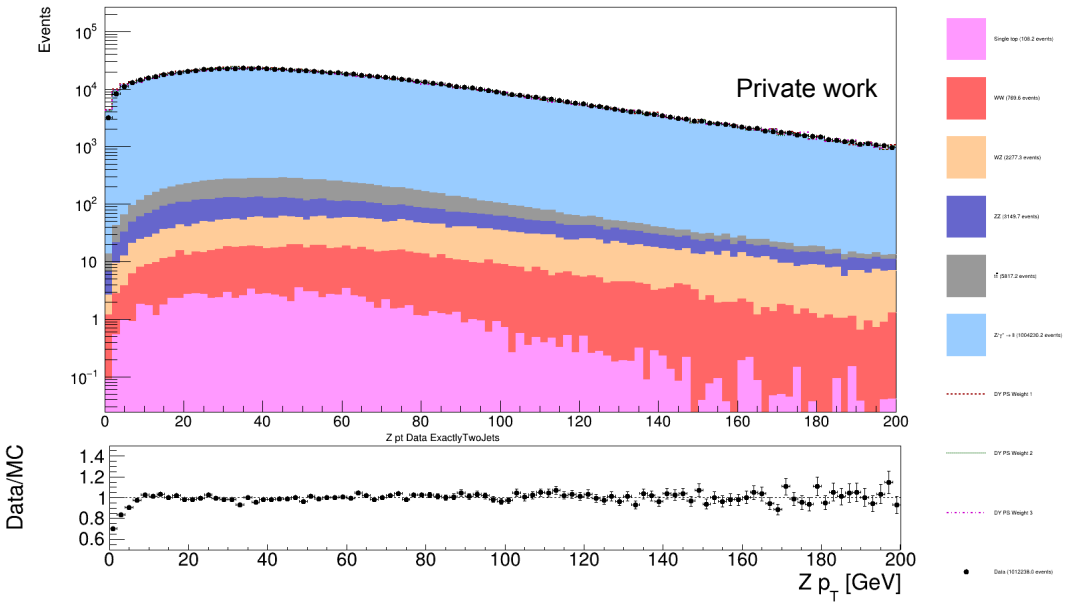


Figure 11.6: Z $_T$ exactly two jets plot with uncertainties for DY sample

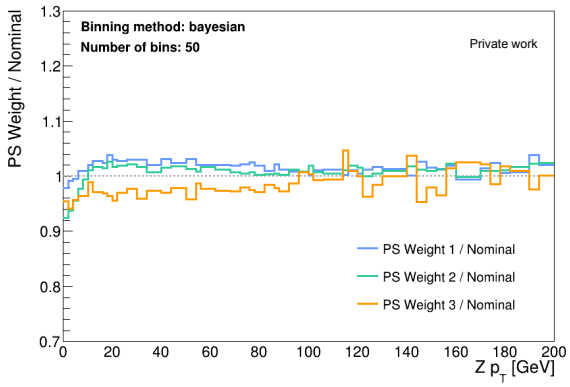


Figure 11.7: Ratio plot of uncertainties for at least two jets channel DY sample

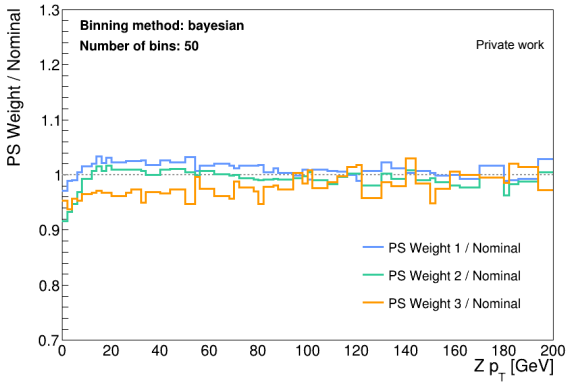


Figure 11.8: Ratio plot of uncertainties for exactly two jets channel DY sample

Chapter 12

PDF scale uncertainties

12.1 Introduction

In this chapter, the systematic approach to incorporating theoretical uncertainties into the Z boson transverse momentum (p_T) distribution is discussed. The uncertainties which are considered here are:

- **PDF Uncertainty ($\sigma_{\mathbf{PDF}}$)**: This represents the variations in the parton distribution functions.
- α_s **Uncertainty (σ_{α_s})**: This represents the variations in the value of strong coupling α_s that runs with energy.

These uncertainties might ensure robust agreement between the data and the Monte Carlo (MC) simulation.

12.1.1 Parton distribution Function (PDF)

Parton Distribution Functions (PDFs) describe the probability of finding a parton (quark or gluon) inside a proton carrying a specific fraction (x) of the proton's momentum at a given energy scale (Q^2). They are essential inputs for predicting cross sections and event rates in high-energy particle collisions, such as those studied in Z + jets processes.

The PDFs are extracted from global fits to experimental data obtained from various processes, including deep inelastic scattering, Drell-Yan processes, and other high-energy interactions.

The mathematical expression for the differential cross-section in terms of PDFs is:

$$\sigma = \int dx_1\, dx_2\, f_i(x_1, Q^2)\, f_j(x_2, Q^2)\, \hat{\sigma}_{ij}(x_1, x_2, Q^2), \tag{12.1}$$

where:

- $f_i(x_1, Q^2)$ and $f_j(x_2, Q^2)$: PDFs for partons i and j, representing the probability densities of finding the partons inside the proton with momentum fractions x_1 and x_2, respectively, at the energy scale Q^2.

- $\hat{\sigma}_{ij}(x_1, x_2, Q^2)$: Partonic cross-section for the subprocess involving partons i and j at energy scale Q^2.

- dx_1, dx_2: Integration over the momentum fractions x_1 and x_2.

This formula shows how PDFs connect the partonic cross-section to the observable hadronic cross-section, incorporating the internal structure of the proton.

12.2 Sources of Uncertainties

These uncertainties in my analysis arise from the `LHEPdfWeight` branch which is stored in the ROOT file:

- **PDF Variations**: Correspond to LHE IDs 325301 to 325400 (eigenvector variations). In total, 100 such variations are there for the PDF set.

- α_s **Variations**: Corresponding to the LHE IDs (eigenvector variations) 325401 ($\alpha_s = 0.116$) and 325402 ($\alpha_s = 0.120$).

12.3 Methodology

The uncertainties are calculated and propagated into the Z boson p_T histogram by the following method:

12.3.1 PDF Uncertainty

For each bin of the Z boson p_T distribution:

1. **Central Value**: The observable or yield in the bin, calculated using the nominal weight w_0 which is the 0th eigenvector.

2. **Variations**: For each eigenvector ($i = 1$ to 100):

$$\Delta_i = Y_i - Y_0 \quad \text{(difference from central value)},$$

   Y_i is the yield for the i-th eigenvector variation.

3. **Total Uncertainty**: Combine variations using:

$$\sigma_{\text{PDF}} = \sqrt{\sum_{i=1}^{100} \Delta_i^2}.$$

12.3.2 α_s Uncertainty

1. **Calculate Variations**: Using yields for $\alpha_s = 0.116$ ($Y_{\alpha_s,\text{down}}$) and $\alpha_s = 0.120$ ($Y_{\alpha_s,\text{up}}$):

$$\Delta_{\alpha_s} = \frac{Y_{\alpha_s,\text{up}} - Y_{\alpha_s,\text{down}}}{2}.$$

2. **Total α_s Uncertainty**: Directly take Δ_{α_s} as the uncertainty.

12.3.3 Combined Uncertainty

The total uncertainty is the quadratic sum of PDF and α_s uncertainties:

$$\sigma_{\text{combined}} = \sqrt{\sigma^2_{\text{PDF}} + \Delta^2_{\alpha_s}}.$$

12.4 Bin-by-Bin Filling of Histograms

1. For each event, considering the nominal weight w_0.
2. Looping over PDF and α_s weights to compute the corresponding yields in each bin.
3. For each bin of the Z boson p_T:
    - **Central Value**: Use the yield calculated using w_0.
    - **Up/Down Variations**: Computing separate yields for PDF and α_s variations.
    - **Combined Up/Down**: Using the central yield $Y_0 \pm \sigma_{\text{combined}}$.

12.5 Results

This process ensures that theoretical uncertainties are rigorously incorporated into the Z boson p_T analysis. The inclusion of PDF and α_s uncertainties enables a more robust comparison between exclusive jet channel and inclusive jet channel. The effects of uncertainties are more visible on the 2 jet channel than one jet channel whereas at least two jet is less effected by the uncertainties than exactly two jet channels and at least one jet channel is less effected than exactly one jet channel. Here are the Data and MC comparison plots with ratio of Data/total MC with the MC uncertainties displayed as red bands around this central ratio line

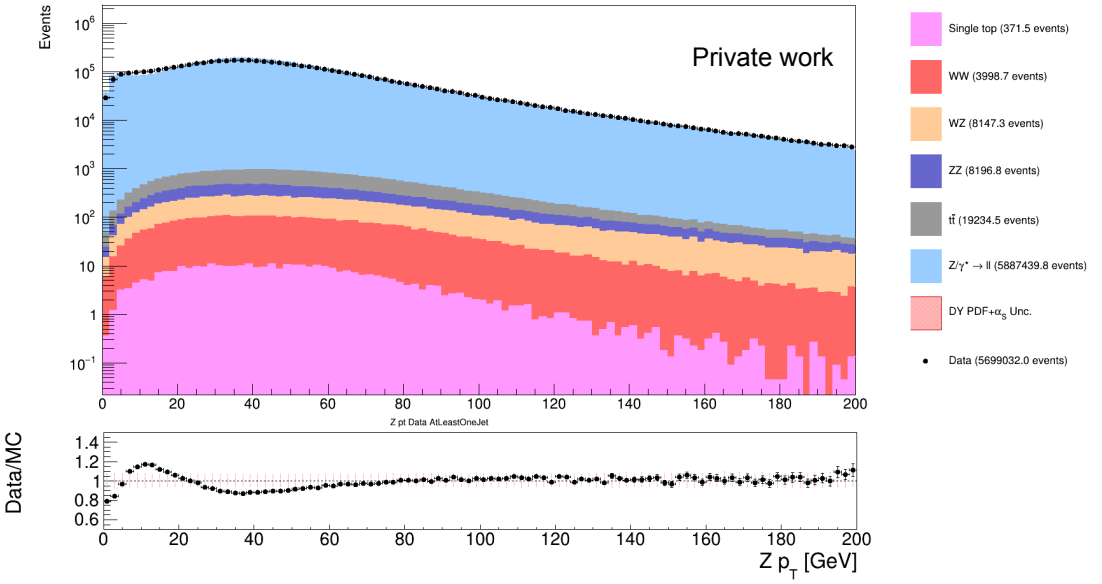


Figure 12.1: Z $_T$ at least one jet plot with uncertainties for DY sample

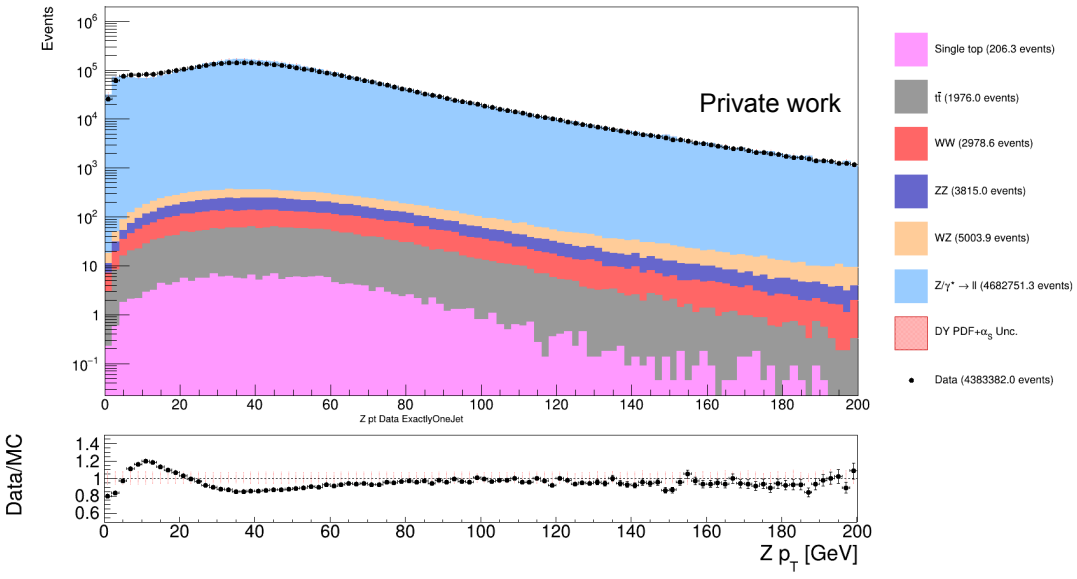


Figure 12.2: Z $_T$ exactly one jet plot with uncertainties for DY sample

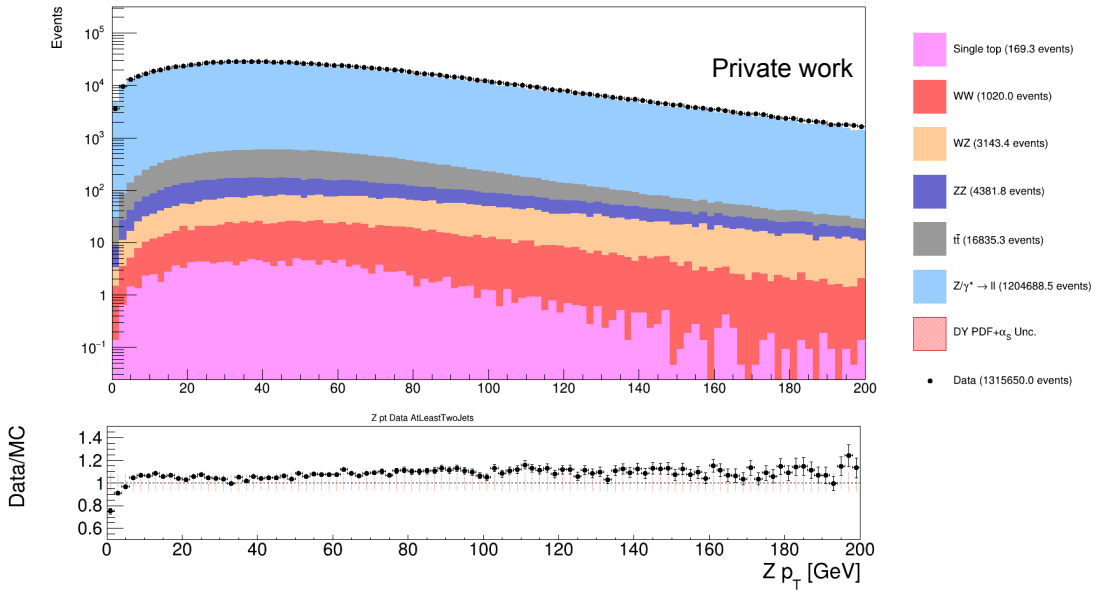


Figure 12.3: Z $_T$ at least two jets plot with uncertainties for DY sample

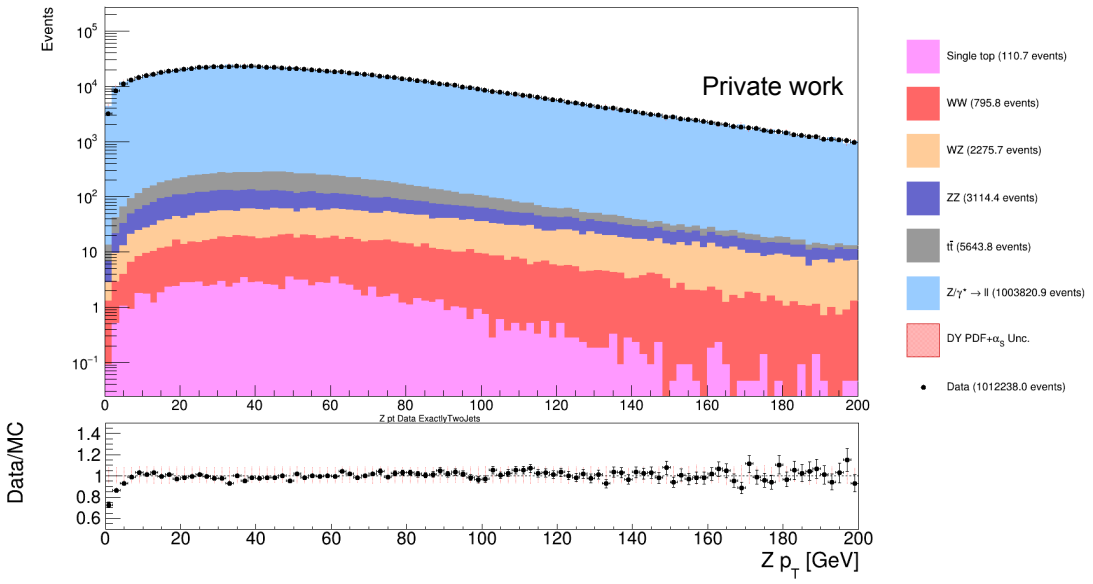


Figure 12.4: Z $_T$ exactly two jets plot with uncertainties for DY sample

Chapter 13

Data nfolding

Unfolding is a crucial technique that is used to follow the true distributions from a measured data. Measured distributions are always effected by detector effects, including finite resolution and inefficiencies. These distortions must be corrected to compare the experimental results with theoretical predictions or Monte Carlo simulations.

The need for unfolding arises because:

- Detector smearing effects cause deviations in reconstructed observables from their true values.
- Background contributions distort the true signal.
- Comparisons with theoretical models require distributions at the generation level, that is, free from detector biases.

Methodology

The unfolding process involves the following steps:

1. Response Matrix Construction

The response matrix describes the relationship between true(generation level) and reconstructed distributions. It is constructed using a signal MC sample where the generator-

level and reconstructed-level events are matched. Basically the probability of a event getting reconstructed give a corresponding true event(gen) is calculated.

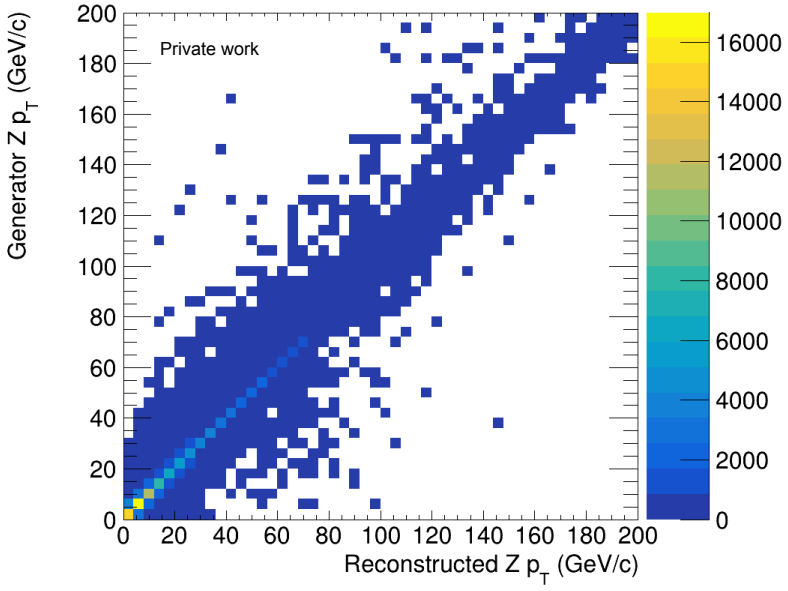


Figure 13.1: Unnormalized Response matrix calculated from DY dataset

2. Background Subtraction

Backgrounds from various processes (e.g., $t\bar{t}$, Single top, WW, WZ, ZZ) are simulated using Monte Carlo (MC) generators and subtracted from the measured data. It is made sure that no negative bins are introduced after subtraction.

3. Bayesian Unfolding

RooUnfold is a C++ framework designed for unfolding problems. RooUnfold's Bayesian method is employed for unfolding, which iteratively updates the unfolded distribution based on the response matrix and background-subtracted data. The number of iterations is optimized to balance statistical fluctuations and resolution.

Analysis

Data and Monte Carlo Samples

All the cuts are kept same as the analysis.

- Data: Measured Z boson p_T distribution for events with at least two jets and atleast one jet.
- Signal: Z + jets events simulated with DY samples.
- Backgrounds: Contributions from $t\bar{t}$, single top, WW, WZ, and ZZ processes.

Results

Two unfolded distributions were obtained:

1. Unfolded Z p_T from data.
2. Unfolded Z p_T from DY simulation.
3. Two more DY samples (generated by me at LO) are also plotted along with the above mentioned two. The newly added samples have different α_s values.

Discussion

Since at least two jets and at least one jet channels have less systematic effects, unfolding has been done with these channel. The observed discrepancy between unfolded data and DY simulation could arise from:

- Uncertainties in background estimation.
- Limitations in detector resolution modeled in the response matrix.
- Theoretical uncertainties in the MC simulations.

The analysis emphasizes the importance of using accurate MC models and understanding systematic uncertainties.

Future Work

Further a study on the impact of different systematic uncertainties (PDFs, μ_F and μ_R scales and Parton showers) and validation of unfolding using pseudo-data studies will be done.

Results

Since inclusive jets channels are less effected by the systematics than the exclusive jets channels, here the unfolded data is compared with the inclusive channels. Here the dark blue is CMS centrally produced samples and the brick red and sea green samples are produced by me using Madgraph. We can see the Blue sample is showing better agreement than the other two samples, this is beacuse the CMS centrally produced samples are generated at NLO and my samples are generated at LO. This leads to better precision with the use of the NLO level.

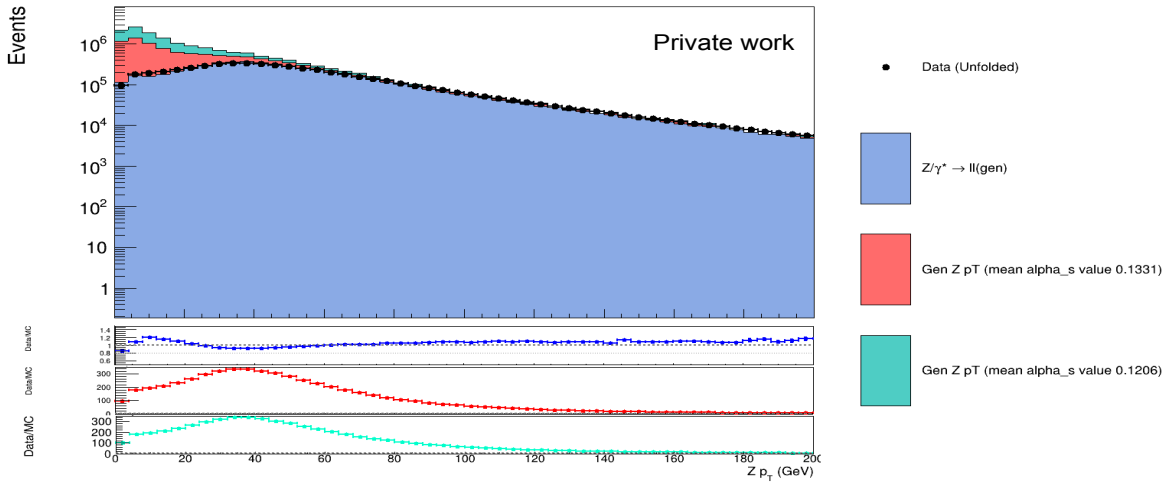


Figure 13.2: unfolded data with DY samples (gen level) for atleast one jet

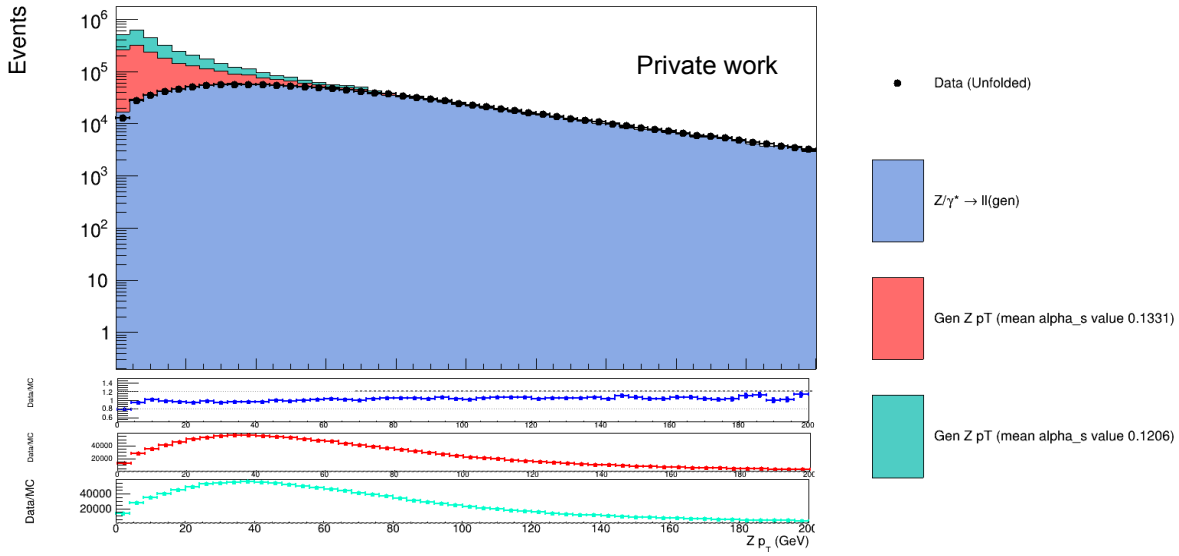


Figure 13.3: unfolded data with DY samples (gen level) for atleast two jets

Bibliography

[1] CERN, *LHC Particle Acceleration In-depth Explanation*, [Online Video], Available: `https://home.cern/resources/video/acceleration`, Accessed: March 2025.

[2] The CMS Collaboration, *The CMS Detector*, [Online], Available: `https://cms.cern/detector`, Accessed: March 2025.

[3] CMS Collaboration, *Detector Cross Section and Related Studies*, `https://arxiv.org/pdf/2309.05466`, Accessed: March 2025.

[4] Chatrchyan S. et al., *Reconstruction Performance of the CMS Detector*, J. Instrum. 3 S08004, 2008. Available: `https://iopscience.iop.org/article/10.1088/1748-0221/3/08/S08004/pdf`.

[5] CMS Collaboration, *Electron Reconstruction and Identification*, CMS Physics Analysis Summary, 2014. Available: `https://cds.cern.ch/record/1988091/files/cms-egm-13-001_arxiv.pdf?version=1`.

[6] CMS Collaboration, *Jet Energy Scale Corrections and Reconstruction in CMS*, CMS PAS JME-16-003, 2017. Available: `https://cds.cern.ch/record/2256875/files/JME-16-003-pas.pdf`.

[7] Sjöstrand, T., Ask, S., Christiansen, J. R., et al., *An Introduction to PYTHIA 8.2*, Comput. Phys. Commun., vol. 191, 2015, 159-177. doi:10.1016/j.cpc.2015.01.024.

[8] CMS Collaboration, *Measurement of differential cross sections for Z boson production in association with jets in proton-proton collisions at* $\sqrt{s} = 13$ *TeV*, Eur. Phys. J. C 78, 965 (2018), `https://arxiv.org/abs/1804.05252`.

[9] LHAPDF Collaboration, *LHAPDF: Parton Distribution Function Choices*, [Online], Available: `https://lhapdfsets.web.cern.ch/current/`.

[10] T. Adye, *RooUnfold: Software for Data Unfolding*, [Online], Available: `https://hepunx.rl.ac.uk/~adye/software/unfold/RooUnfold.html`.

[11] OpenAI, ChatGPT Language Model, 2023, *An AI-based tool for language generation and assistance*, `https://openai.com/chatgpt`.

[12] CERN, *FxFx Merging for NLO+PS Simulations*, [Online], Available: `https://amcatnlo.web.cern.ch/FxFx_merging.htm`.

[13] Rappoccio, S., *TUnfold Systematics Example*, GitHub Repository, Available: https://github.com/rappoccio/TUnfoldExamples/blob/master/tunfoldsys_example.ipynb.

[14] CERN, *CMS Generator Introduction*, CERN Twiki, Available: https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookGenIntro.

[15] Anthropic, *Claude AI - Conversational AI Model*, Available: https://www.anthropic.com/index/claude-ai.

[16] Sjöstrand, T., et al., *PYTHIA 8 - A Brief Introduction*, Pythia Official Documentation, Available: https://www.pythia.org/latest-manual/Welcome.html.

[17] Frederix, R., et al., *MadGraph Event Generation: Lecture Notes*, NTU-MG5 Lectures, Available: https://cp3.irmp.ucl.ac.be/projects/madgraph/raw-attachment/wiki/SchoolNTU/NTU-MG5-lectures.pdf.

[18] Giele, W., *Parton Showers: CTEQ-MCnet 2021 Lecture Notes*, Available: https://indico.cern.ch/event/1005703/contributions/4221947/attachments/2304165/3924201/CTEQ-MCnet_2021_PartonShowers.pdf.

[19] Guzzi, M., *PDF Theory and Applications*, Presented at the 2019 Workshop on PDFs, Available: https://indico.cern.ch/event/819261/contributions/3423306/attachments/1883709/3104375/MGUZZI-PDF1-F_compressed.pdf.

[20] Weinberg, S., *Quantum Field Theory*, Rep. Prog. Phys. **45**, 1167 (1982). Available: https://iopscience.iop.org/article/10.1088/0034-4885/45/11/002.