Study of Bin Migration in the Z Boson Rapidity Measurement at CMS

Nathan E. Mirman

University of Minnesota

May 2010

Abstract

In the Z boson rapidity measurement [2], bin migration effects are currently added to the total measurement error. Unfolding is an alternative approach in which bin migration effects are reduced in order to lower the measurement uncertainties. Since our knowledge of the PDFs at LHC is limited, it is important for the unfolding process to be insensitive to variations in the assumed PDFs. We demonstrate that unfolding exhibits minimal PDF-dependence, and is a viable tool for reducing measurement uncertainties in the analysis [2].

1 Introduction

The Large Hadron Collider (LHC) is a 14 TeV proton-proton collider based at CERN near the Franco-Swiss border. This massive 27 km accelerator currently holds the world record for collision energy, surpassing the Tevatron's opposing 1 TeV proton, antiproton beams. Two general-purpose detectors, CMS and ATLAS, are positioned on opposite ends of the ring – these experiments are expected to answer many long-standing questions in elementary particle physics. Most notably, CMS and ATLAS are involved in the search for the Higgs boson, the last yet-unobserved constituent of the Standard Model of particle physics[1], and the search for physics beyond the Standard Model, such as supersymmetry (SUSY) and extra dimensions.

[1]The Standard Model is the leading theory describing the properties and interactions of elementary particles. It includes three of the four fundamental forces (strong, weak, and electromagnetic), and the three generations of matter (quarks, leptons, and force carrying bosons). The theory has undergone extensive experimental verification since its development in the 1960s and 1970s.

1.1 Z Boson Production

A crucial step in the search for new physics at LHC is the reproduction and study of well-understood phenomena. In addition to polishing our understanding of these phenomena, such a study serves as a check of our detection and analysis techniques before we enter the realm of new physics. The Z boson, first observed in 1983 at the Super Proton Synchrotron (SPS) at CERN [6], is produced copiously at LHC and has a clean dilepton signature in the detector. For this reason, the Z boson is an excellent subject to study before moving to more exotic physics. Also, Z production is a background for many other physics analyses, and must be understood well to improve the quality of our measurements [10].

The study of Z boson production at LHC is intrinsically tied to the proton's structure. Historically, the internal parton structure of the nucleon was established via deep inelastic scattering (DIS) experiments. The quark and gluon structure of hadrons was ascertained at the Stanford Linear Accelerator (SLAC) with a high-energy electron beam. Later, DIS experiments at CERN and DESY aided in the development of Quantum Chromodynamics (QCD) theory describing the strong interaction [8]. The Livingston plot shows the historical trend of increasing collision energy since the first scattering experiments in the early 1930s (Figure 1). After a long tradition of scattering, LHC experiments will continue to probe the proton's inner structure at the TeV scale.

1.2 Parton Distribution Functions

In order to understand the physics behind DIS, it is crucial to treat the parton distribution functions (PDFs). These functions describe how the momentum of a hadron is distributed among its parton constituents. The PDFs can be conveniently represented in terms of the kinematic variables Q^2, the four-momentum transfer, and x, the Bjorken scaling variable. For the interaction $\ell N \to \ell' X$ (Figure 2), these variables are given by

$$Q^2 \equiv -q^2 \tag{1}$$

$$x = Q^2/(2p \cdot q) \tag{2}$$

where p is the hadron four-momentum and $q = k - k'$ is the four-momentum transfer [4]. A review of relativistic kinematics can be found in Appendix A. The four-momentum transfer (Q^2) is essentially a measure of probe fineness – a large value of Q^2 corresponds to a small de Broglie wavelength, and thus a high resolution. The Bjorken scaling variable (x) can be approximated as the fraction of four-momentum

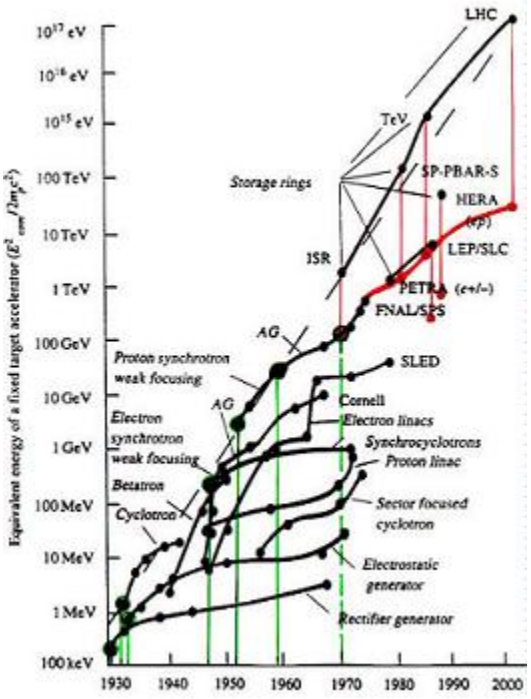


Figure 1: Livingston plot shows the center of mass collision energies of major accelerators built in the last century. The y-axis on this plot corresponds to the beam energy of an equivalent fixed-target accelerator. As of 2010, the LHC is at the high energy frontier [7].

carried by the struck parton. Thus, it describes properties of the target proton. It is worth noting here that a DIS event has two degrees of freedom, whereas an elastic scattering event has only one. Although energy is conserved in both instances, kinetic energy is not conserved in the former. In DIS, a portion of the collision energy is used to excite or alter the scattering products.

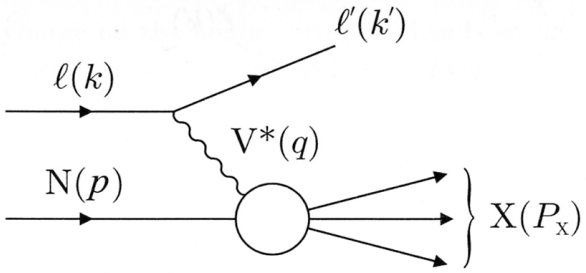


Figure 2: Feynman diagram for lepton-hadron deep inelastic scattering process [4].

The kinematic variables x and Q^2 allow us to describe the PDFs of a hadron. We say that the probability of finding a parton of flavor i and four-momentum fraction x in a hadron is given by $f_i(x, Q^2)$. The PDFs for a proton are illustrated in Figure 3. The Q^2 evolution of the function can be calculated by perturbative QCD. However, it is not possible to find the x evolution with this method. To evaluate $f_i(x, Q^2)$, we parametrize the x-evolution for some initial $Q^2 = Q_0^2$ that is sufficiently large for pQCD to remain valid, and then apply the DGLAP equations[2] to evolve $f_i(x, Q_0^2)$ to some $f_i(x, Q^2)$. Then, $f_i(x, Q^2)$ undergoes a χ^2 fit to the data, typically in a 10-20 parameter space. Once the fit parameters are known, we have the x-evolution of $f_i(x, Q^2)$ for all Q^2, including the initial Q_0^2. Because this method of evaluating $f_i(x, Q^2)$ involves various theoretical assumptions, there is an inherent uncertainty in the resulting form. The LHC operates in a currently unexplored (x, Q^2) kinematic regime (Figure 4), and will provide further constraints on the proton PDFs. However, the current PDF uncertainties need to be accounted for in the analysis of LHC data. A more detailed discussion of this topic follows in Section 2.3.

In hadronic DIS, both the probe particle and the target particle are hadrons – here the probe particle is not pointlike. This complicates the interpretation of scattering data in several ways. Unlike leptonic DIS, a soft interaction between individual partons in the scattering hadrons may modify the hard scattering cross-section. Also, parton-parton interactions may alter the scattering products, forming jets or new particles in the process. In this case, the PDFs can be reconstructed by careful analysis of the scattering products.

At LHC, the Drell-Yan process ($q\bar{q} \to \gamma^*/W/Z \to \ell^+\ell^-$) is a prominent channel for Z production [4].

[2]Named for Dokshitzer, Gribov, Lipatov, Altarelli, and Parisi.

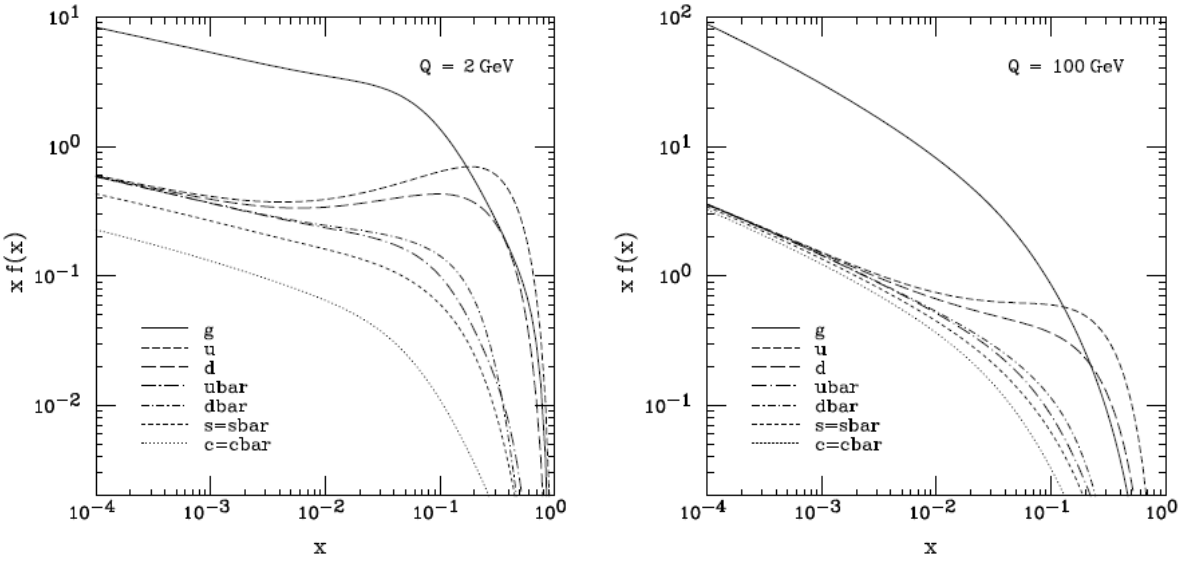


Figure 3: Overview of the CTEQ6M parton distribution functions at $Q = 2$ and $Q = 100$ GeV [9].

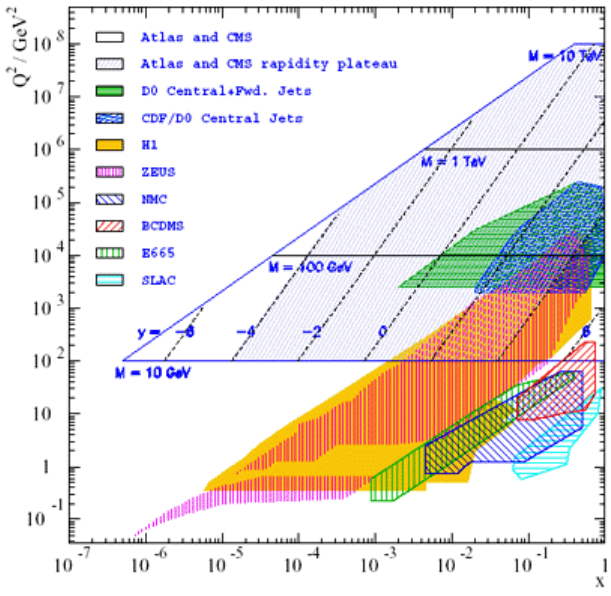


Figure 4: Kinematic coverage of the DIS and collider $pp - p\bar{p}$ experiments. The LHC will allow for the study of an unexplored low-x, high-Q^2 regime [5].

Here, two partons annihilate to create a virtual photon, W boson, or Z boson, which in turn decays into a lepton, antilepton pair. Although this process is responsible for a minute portion of the total hadron-hadron cross section, its clean dilepton signature is a convenient feature for scattering studies. The cross-section for this process is given by

$$\frac{d\hat{\sigma}}{d\cos\theta^*}(q_i\overline{q}_i \to \ell^+\ell^-) = \frac{\pi\alpha^2}{6\hat{s}}\left[(1+\cos^2\theta^*)A + \cos\theta^* B\right] \tag{3}$$

where A and B are constants determined by the flavor of i and ℓ, and θ^* is the angle of ℓ^+ in the Z^0 rest frame with respect to the incident proton beam direction [4]. A measurement of the angular distribution of the lepton, anti-lepton pair provides information about the Z production cross section, as well as the underlying PDFs. In the study [2], the positions of Drell-Yan electron, positron pairs in the CMS detector are reconstructed and the shape of the resulting Z boson rapidity distribution is measured at a center of mass energy of 10 TeV.

2 Z Boson Rapidity Measurement

Here we give an overview of topics from the analysis [2] which are relevant for our study. In the analysis, estimates are given for the measurement of the shape of the rapidity distribution for Z bosons produced in proton-proton collisions at a center of mass energy of 10 TeV. In this measurement, the angular distribution of the Z boson in the CMS detector is reconstructed from the Drell-Yan electron, positron pair, which is created in the interaction $pp \to Z/\gamma^* \to e^+e^- + X$. Angular distribution is measured in the rapidity variable, denoted Y (see Appendix B). The measurement seeks to provide constraints on the proton PDFs produced at LHC.

To measure the normalized differential cross section of the Z boson rapidity distribution, the following expression is evaluated:

$$\frac{1}{\sigma}\frac{d\sigma(Z \to e^+e^-)}{dY_i} = \frac{\epsilon \times A}{N - B} \cdot \frac{N_i - B_i}{\Delta_i(\epsilon \times A)_i} \tag{4}$$

where σ is the total cross section, N_i is the number of Z candidates observed in data, B_i is the estimated number of background candidates, Δ_i is the bin width, ϵ and A are the efficiency and acceptance for detecting and fully reconstructing a Z boson with rapidity Y, and the index i indicates the bin of the rapidity distribution. The results of the measurement are given in Figure 5. There are several sources of measurement uncertainty which play a role in the analysis – here we focus on bin migration errors and

theoretical errors due to parton distribution function models.

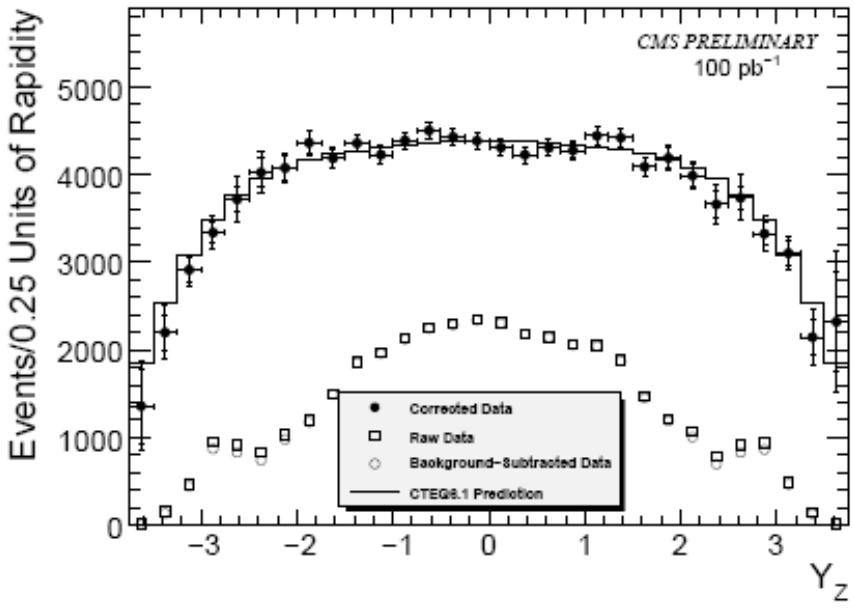


Figure 5: The final results for the rapidity measurement. The raw full simulation (data) distribution is shown with empty circles and the distribution corrected by $(\epsilon \times A)$ is shown with solid circles. The errors on the solid circles are shown for statistical and statistical+systematic separately. The prediction of CTEQ6.1 is shown for comparison. Results are shown for an integrated luminosity of 100 pb^{-1} [2].

2.1 Bin Migration and Smearing

Bin migration is a statistical effect due to limited measurement resolution which causes an alteration to the rapidity distribution obtained from data. There are several possible physical causes for bin migration, but they all share a common property – the processes at work here act unevenly across the different regions of the rapidity range which is covered by the measurement. In one possible scenario, a final state radiation photon is emitted at a large angle from the Drell-Yan electron and altering the rapidity of the reconstructed Z boson. Other scenarios include detector effects, such as the emission of bremsstrahlung photons, energy loss in the tracker, limitations in the detector energy resolution, and errors in the tracker position measurement. By understanding the dominant sources of bin migration, we can account for them in the analysis to reduce measurement errors.

A Monte Carlo full-simulation was carried out to quantify the effects of final state radiation and limited detector resolution on the Z boson rapidity measurement. The PYTHIA event generator version 6.416 [11] and CMSSW version 2_2_10 were used to generate a $Z \to e^+e^-$ signal sample at 7 TeV. The parton distri-

bution function CTEQ6.1 [9] [12] with 40 error sets was used in the generation (see Section 2.3).

Bin migration effects were introduced into the simulation via a parametrized Monte Carlo simulation which applied simple smearing to generator-level electron candidate objects. Given a "true" probability density function $f_{true}(y)$, the measured (smeared) function is given by a convolution [3]:

$$f_{meas}(x) = \int R(x|y) f_{true}(y)\, dy \tag{5}$$

where $R(x|y)$ is the response function. The generator-level objects were defined as the Lorentz-invariant sum of a final-state electron and all other final-state particles in a ΔR cone of 0.08 around the electron. The size of the ΔR cone was chosen to approximate the size of an ECAL cluster. Smearing was applied with resolution functions which have been determined from the full-simulation.

Because bin migration effects differ for various subdetectors in CMS, the smearing functions were chosen uniquely for the Electromagnetic Calorimeter (ECAL) and the Hadron Forward Calorimeter (HF). For electrons reconstructed within ECAL, we define the Crystal Ball function for smearing:

$$f(x;\alpha,n,\overline{x},\sigma) = \begin{cases} \exp\left(-\frac{(x-\overline{x})^2}{2\sigma^2}\right) & \text{for } \frac{x-\overline{x}}{\sigma} > -\alpha \\ \left(\frac{n}{|\alpha|}\right)^n \exp\left(-\frac{|\alpha|^2}{2}\right)\left(\frac{n}{|\alpha|} - |\alpha| - \frac{x-\overline{x}}{\sigma}\right)^{-n} & \text{for } \frac{x-\overline{x}}{\sigma} \leq -\alpha \end{cases} \tag{6}$$

$$\frac{\sigma_{EB}}{E} = c + \frac{a}{\sqrt{E}}(1 - b_1|\eta_d| + b_2\eta_d^2) \tag{7}$$

$$\frac{\sigma_{EE}}{E} = \frac{a}{E_T}(1 - b_1|\eta_d| + b_2\eta_d^2) \tag{8}$$

Here the four primary parameters are: the mean ($\overline{x}$), width (σ), power-law magnitude (n), and the point of changeover between the power-law and Gaussian (α). The values of the parameters were determined independently for the ECAL endcap (EE) and ECAL barrel (EB) by conducting a χ^2 analysis on the Z invariant mass distribution. They are given in Table 1.

For HF electrons, the smearing function is a Gaussian parametrized by its mean value (m), and a width given by:

$$\frac{\sigma}{E} = \frac{a}{\sqrt{E}} \oplus c \tag{9}$$

where a is the classic stochastic term and c is the constant term of the calorimeter energy (E) resolution. The parameters here were determined by a χ^2 analysis, as in the ECAL case. The HF smearing parameters

Parameter	ECAL Barrel Value		Error	ECAL Endcap Value		Error
$\overline{x}$	0.9960	$\pm$	0.0001	0.993	$\pm$	0.001
α	1.91	$\pm$	0.01	1.62	$\pm$	0.03
a	0.040	$\pm$	0.001	1.096	$\pm$	0.012
c	0.040	$\pm$	0.005	(N/A)		
n	3.0			3.0		
b_1	1.06			0.72		
b_2	4.78			0.15		

Table 1: ECAL Smearing Parameters

are given in Table 2.

Parameter	Value		Error
a (stochastic)	2.58	$\pm$	0.06
c (constant)	0.050	$\pm$	0.03
m (mean)	0.997	$\pm$	0.003

Table 2: HF Smearing Parameters

2.2 Unfolding

Given a measured distribution undergoing bin migration effects, we unfold, or unsmear, the data to "remove" the bin migration effects from the underlying distribution. Here, we approach this task by inversion of the response matrix [3]. The response matrix is given by:

$$R_{ij} = \text{Prob(observed in bin } i \mid \text{true value in bin } j) \tag{10}$$

The efficiency, ϵ_j, of some bin j can then be obtained by the relation

$$\sum_{i=1}^{N} R_{ij} = \text{Prob(observed anywhere} \mid \text{true value in bin } j) = \epsilon_j \tag{11}$$

The measured data, $\mathbf{n}$, with expectation values $\boldsymbol{\nu}$, is then related to the "true" data, $\boldsymbol{\mu}$, by the relation

$$E[\mathbf{n}] = \boldsymbol{\nu} = R\boldsymbol{\mu} + \beta \tag{12}$$

where β is the expected number of background events. We unfold the data by inverting the response matrix:

$$\boldsymbol{\mu} = R^{-1}(\boldsymbol{\nu} - \beta) \tag{13}$$

An illustration of the machinery described above is given in Figure 6.

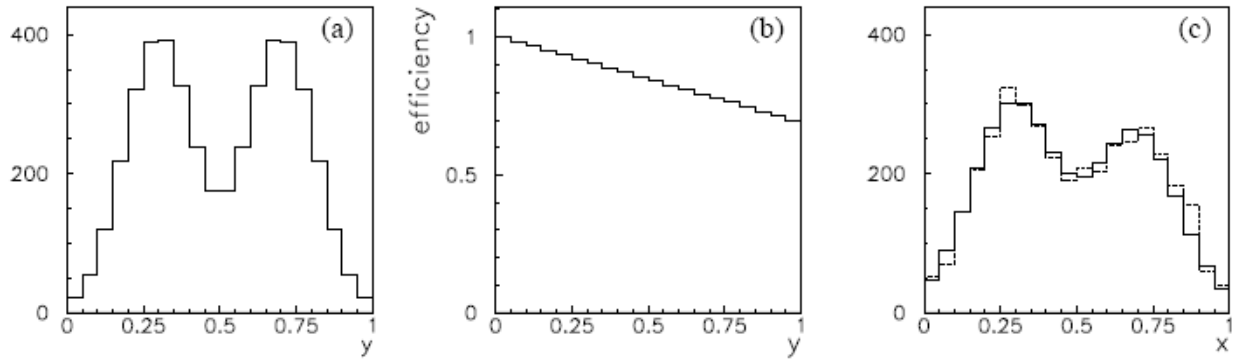


Figure 6: Illustration of ingredients for unfolding: (a) a 'true histogram' $\boldsymbol{\mu}$, (b) a possible set of efficiencies ϵ, and (c) the observed histogram $\mathbf{n}$ (dashed) and corresponding expectation values $\boldsymbol{\nu}$ (solid) [3].

2.3 Parton Distribution Function Uncertainties

The task of unfolding a measured distribution depends on our knowledge of the response matrix for the measurement at hand. The response matrix is, in turn, sensitive to the underlying PDFs. Given that the predicted PDFs have inherent theoretical uncertainties (Section 1.2), the "effectiveness" of unfolding may vary depending on the assumed PDFs. Figure 7 demonstrates the effect of PDF uncertainties on the $(\epsilon \times A)$ measurement – the current approach is to include the PDF-induced uncertainties into the overall measurement error. In our study, we wish to quantify the effect of PDF uncertainties on the unfolding process.

The PDFs used in our study are provided by the CTEQ (Coordinated Theoretical-Experimental Project on QCD) collaboration. The CTEQ6.1 PDF set includes a base PDF and 40 variations corresponding to 20 parameters used in the definition of $f_i(x, Q_0^2)$. The input PDFs are parametrized by:

$$xf_i(x, Q_0^2) = A_0 x^{A_1}(1-x)^{A_2} e^{A_3 x}(1+e^{A_4}x)^{A_5} \tag{14}$$

The parton flavor combinations $u_v \equiv u - \overline{u}$, $d_v \equiv d - \overline{d}$, g, and $\overline{u} + \overline{d}$ have unique sets of parameters A_k, hence, we have 20 theoretical parameters in total. The values of A_k are determined by evolving $f_i(x, Q_0^2)$ to some $f_i(x, Q^2)$, and then conducting a χ^2 fit to the data.

The strength of the CTEQ6.1 set lies in its treatment of correlated systematic uncertainties. For the χ^2

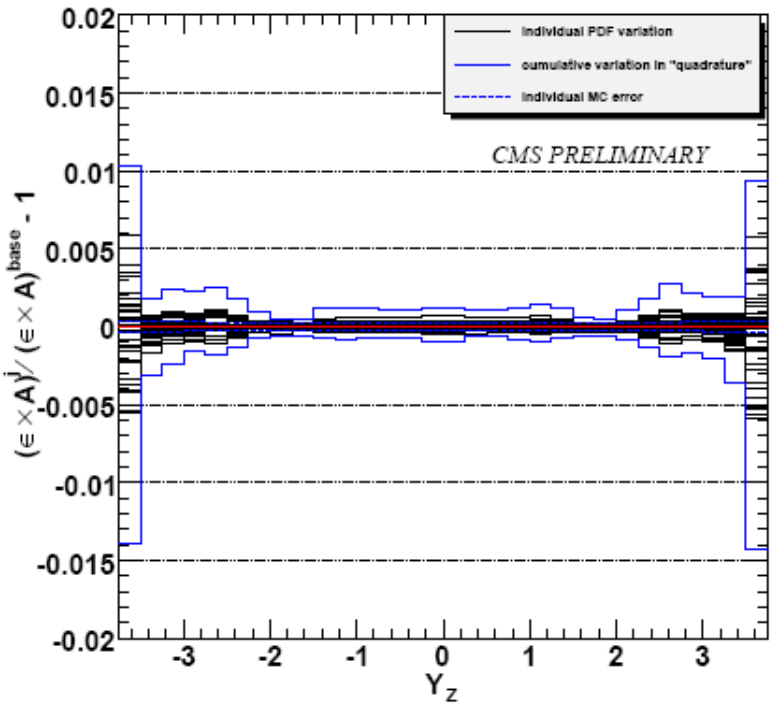


Figure 7: Fractional difference of PDFs $(\epsilon \times A)$ from median $(\epsilon \times A)$, as a function of Z boson rapidity, individual and combined [2]. Statistical error of MC is shown for reference.

fit of the parametrized PDFs to data, we compute [4]:

$$\chi^2 = \sum_i \sum_j \left[F_i^{QCD}(p) - F_i(meas)\right] V_{ij}^{-1} \left[F_i^{QCD}(p) - F_i(meas)\right] \tag{15}$$

where $F_i^{QCD}(p)$ is the prediction from QCD for data point i given the theoretical parameters p. The covariance matrix V_{ij} is given by:

$$V_{ij} = \delta_{ij}\sigma_i^2 + \sum_\lambda \Delta_{i\lambda}^{sys} \Delta_{j\lambda}^{sys} \tag{16}$$

Here σ_i^2 is the one standard deviation uncorrelated error on data point i from both statistical and systematic sources, and $\Delta_{i\lambda}^{sys}$ represents the one standard deviation correlated systematic error on data point i due to the correlated error source λ. Then, we can write:

$$\chi^2 = \sum_i \frac{\left[F_i^{QCD}(p) - F_i(meas) + \sum_\lambda s_\lambda \Delta_{i\lambda}^{sys}\right]^2}{\sigma_i^2} + \sum_\lambda s_\lambda^2 \tag{17}$$

where s_λ are systematic uncertainty parameters with zero mean and unit variance. Also, we define the

Hessian matrix:

$$M_{jk} = \frac{1}{2}\frac{\partial^2 \chi^2}{\partial p_j \partial p_k} \tag{18}$$

where p_j and p_k are theoretical parameters. The standard propagation of errors to some distribution F is given by:

$$\langle \sigma_F^2 \rangle = T \sum_j \sum_k \frac{\partial F}{\partial p_j} V_{jk} \frac{\partial F}{\partial p_k} \tag{19}$$

where T is the χ^2 tolerance. The PDFs in CTEQ6.1 are fit to data via the "Hessian method", which includes the systematic uncertainty parameters s_λ together with the theoretical parameters p in the fit. In this fitting procedure, the theoretical prediction is not fit to the central values of the published data. Instead, the fit data points are allowed to vary according to their correlated systematic uncertainties. In effect, the systematic shifts between two given data sets are correlated in the fit. Thus, given several sets of measured data, each carrying its own statistical and systematic uncertainties, the Hessian method determines the most consistent fit to all the data sets.

The covariance and Hessian matrices determined by the Hessian method are usually not diagonal because of correlations between the parameters. We diagonalize the covariance matrix by computing its eigenvector PDFs and its eigenvalues, representing the corresponding parameter errors added in quadrature. This procedure is illustrated in Figure 8. Uncertainty bands for the up and down-quark PDFs are shown in Figure 9.

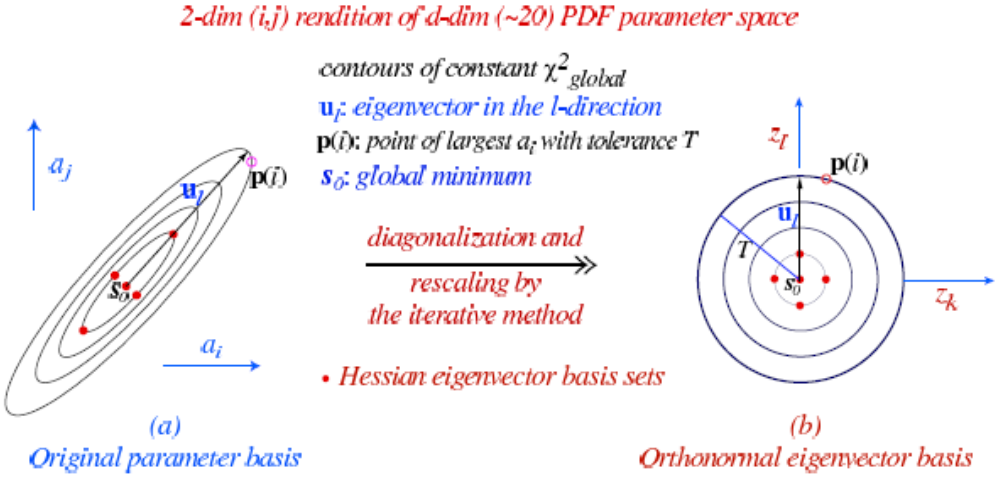


Figure 8: Illustration summarizing the diagonalization of eigenvector PDFs obtained by the Hessian method [9]. Two PDF variations extracted per eigenvector correspond to $\pm u_l$.

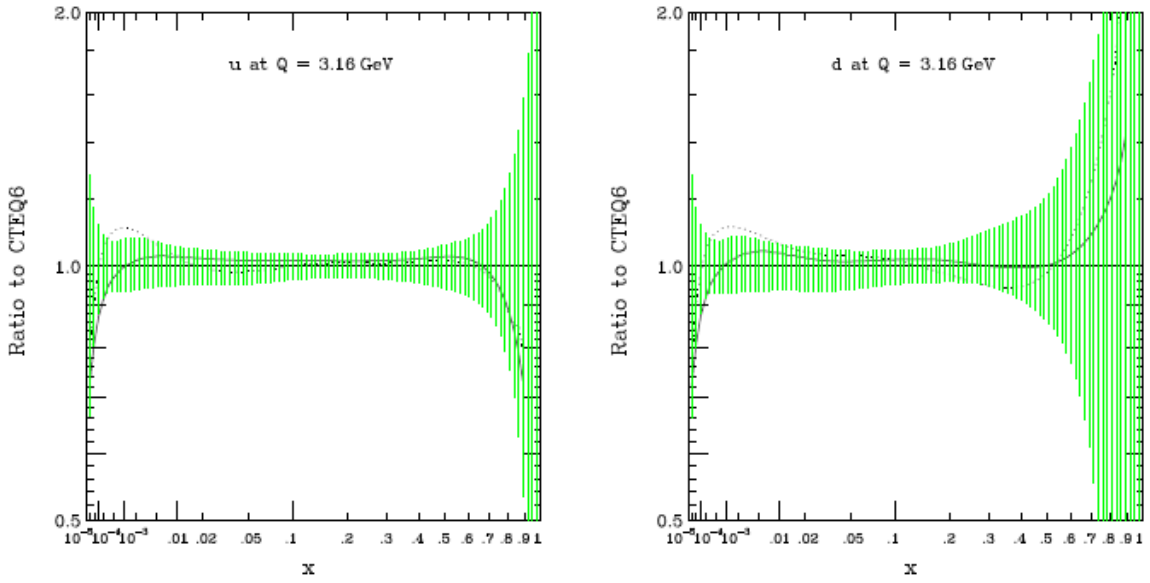


Figure 9: Uncertainty bands for the u- and d-quark PDFs at $Q^2 = 10$ GeV^2. The solid line is the CTEQ5M1 PDF and the dotted line is the MRST2001 PDF [9].

3 Results

To estimate the PDF dependence of unfolding in the Z boson rapidity distribution, response matrices were computed for each of the 40 CTEQ6.1 error PDF sets and the base set. These matrices were computed with 9 million events for each of the 40 error PDF sets, and 49 million events for the base set. The unfolding process was tested on 1 million events from each of the 40 error sets and the base set, in which the Z boson rapidity distribution was smeared according to the method described in Section 2.1. The events used to determine the response matrices were separated from those events used to test the unfolding process in order to avoid correlations between these two sets. The response matrix computed for the base PDF set is represented in Figure 10. The smeared and unfolded distributions for PDF variation 20 are shown in Figure 11.

The "effectiveness" of unfolding was quantified via the Kolmogorov-Smirnov Test (see Appendix C). Given a measured (smeared) distribution $\boldsymbol{\nu}$, we unfold it and determine the maximum KS distance, D_{KS}, between the unfolded distribution and the Monte Carlo truth. There are several factors at work here – the PDF variation v_R used to determine the response matrix R, and the PDF variation v_S used to determine the smeared data vector $\boldsymbol{\nu}$. We say that a set of vectors $\boldsymbol{\nu}$ are unfolded with a variable PDF set if $v_R = v_S$ for each variation, and unfolded with the base PDF set if $v_R = 0$ for each variation (variation 0 is the base PDF). Unfolding in these two cases is shown for $v_S \in [0, 40]$ in Figure 12. Finally, we examine the cases where $v_R \in [0, 40]$ and $v_S \in [0, 40]$, and compare to the two cases above by appropriately normalizing the

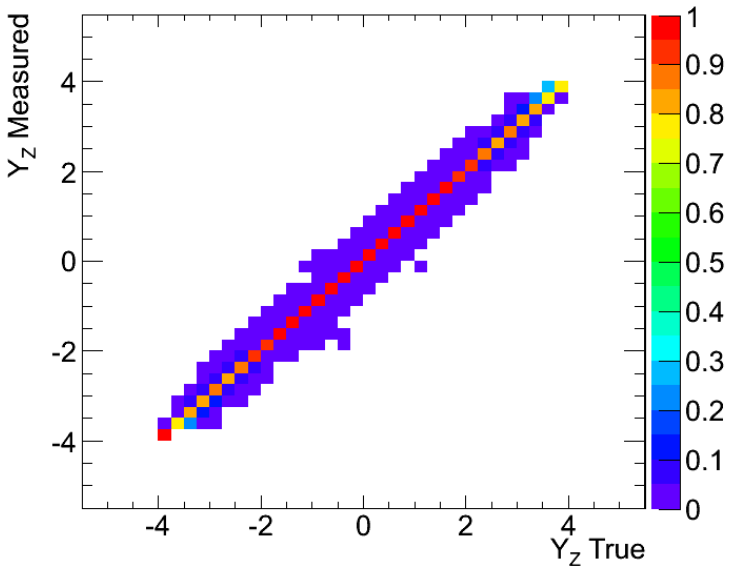
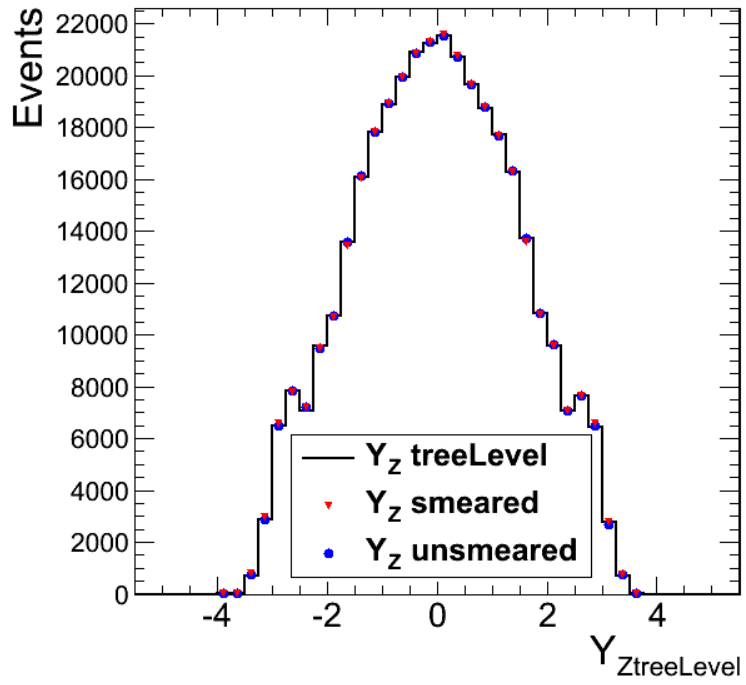


Figure 10: Response matrix determined for base PDF set (49 million events). Off-diagonal entries are responsible for bin migration in the Z boson rapidity distribution.

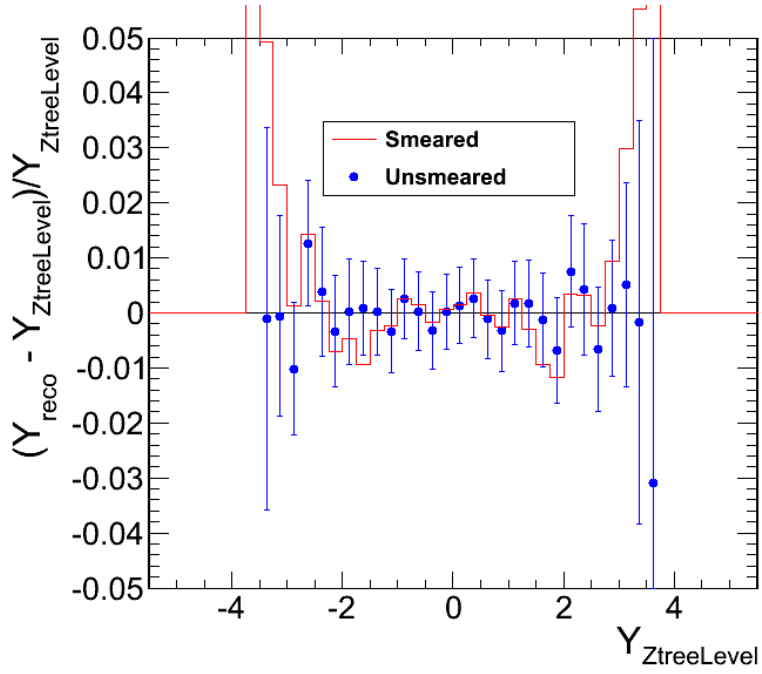
results (Figure 13). We see that the distributions are relatively uniform – thus unfolding does not have a large PDF-dependence. We also investigate the behavior of these distributions when only the rapidity range $|Y| > 2$ is considered. Here the distributions show larger fluctuations in D_{KS}, but the mean normalized KS distance remains near unity.

4 Conclusion

We have investigated the PDF-dependence of unfolding in the Z boson rapidity measurement. The unfolding process shows minimal dependence on PDF variations – by Figure 12 we see that unfolding with a variable PDF set performs only slightly better than unfolding with the base PDF set. This is also supported by the minute difference between both normalization methods in Figure 13. At high rapidities, the normalized KS distances show larger fluctuations, perhaps due to low statistics (Figure 13). Overall, there are no trends which indicate a strong dependence of unfolding on the underlying PDF. Thus, unfolding is a viable approach to reduce measurement errors in the Z boson rapidity measurement.



(a)



(b)

Figure 11: Shown is the Z boson rapidity distribution before normalization by $(\epsilon \times A)$ which underwent unfolding, taken from PDF variation 20. Bin migration effects are visible in the fractional plot 11(b). Errors bars are the allowed Poisson counting errors for each bin, shown for reference.

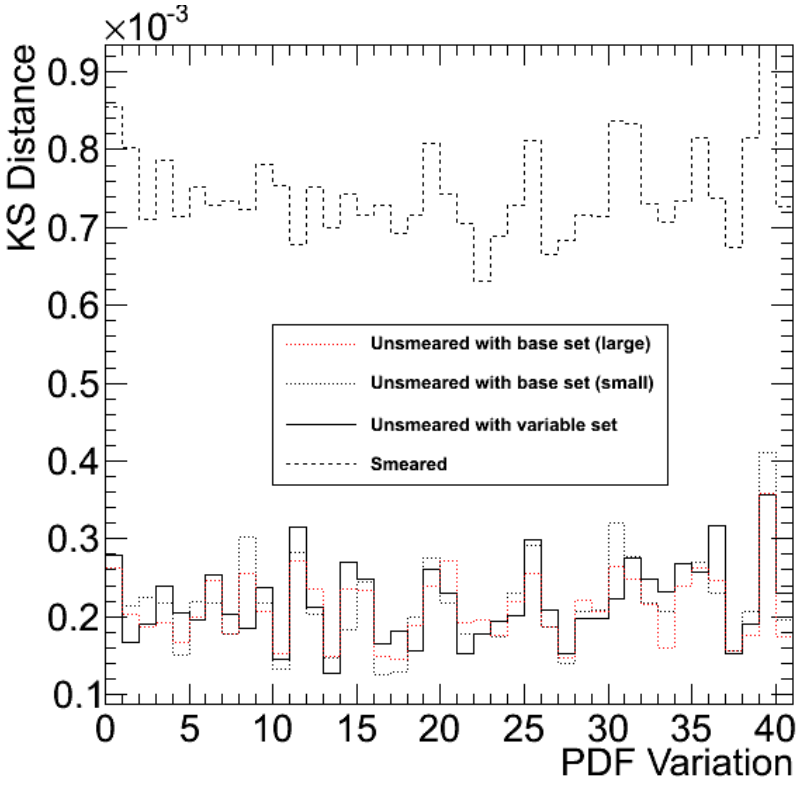
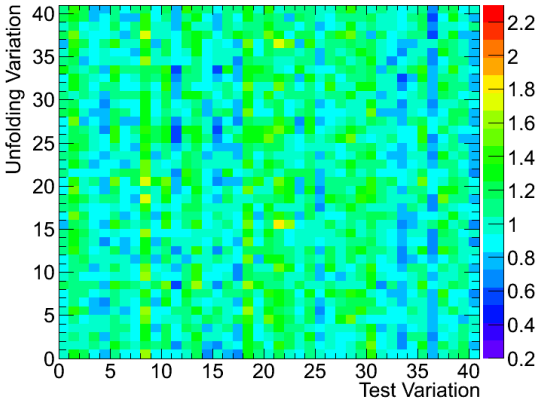
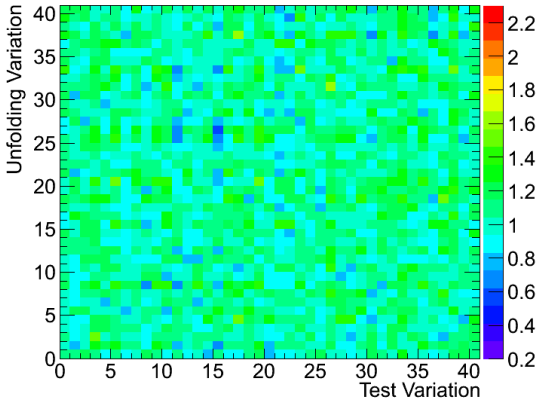


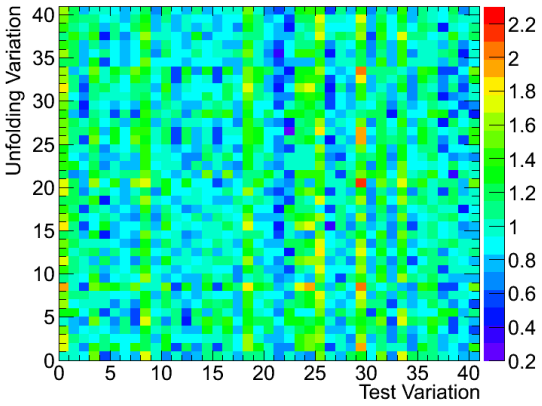
Figure 12: Shown for each PDF variation is the maximum KS distance between unfolded distribution and Monte Carlo truth. The cases with $v_R = 0$ (dotted lines) and $v_R = v_S$ (solid line) are represented along with the case before unfolding, for reference. Large base set contains 49 million events, and small base set contains 9 million events.



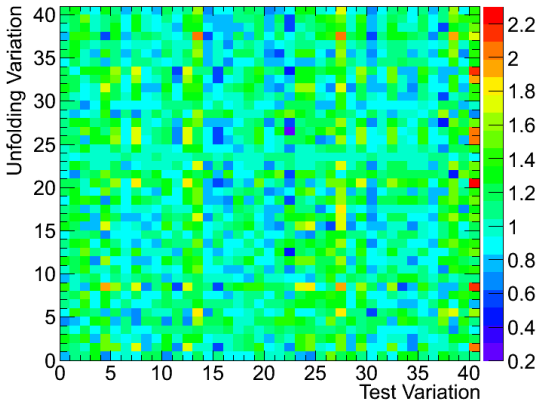
(a) Normalized via $D_{ij\,norm} = D_{ij}/D_{ii}$



(b) Normalized via $D_{ij\,norm} = D_{ij}/D_{ij}$(base).



(c) $|Y| > 2$. Normalized via $D_{ij\,norm} = D_{ij}/D_{ii}$



(d) $|Y| > 2$. Normalized via $D_{ij\,norm} = D_{ij}/D_{ij}$(base).

Figure 13: Shown are the maximum KS distances for $v_R \in [0, 40]$ and $v_S \in [0, 40]$. Bin migration effects are accented in 13(c) and 13(d) by testing only bins with $|Y| > 2$ in the rapidity distribution – these effects are stronger at high rapidities in the detector. The measured KS distances are normalized by KS distances obtained in the cases $v_R = 0$ (13(b), 13(d)) and $v_R = v_S$ (13(a), 13(c)).

Acknowledgments

I would like to thank my advisor, Professor Jeremy Mans, for his guidance and support, and for three awesome years of research at Minnesota CMS. Thank you to Kevin Klapoetke for all his help when I was starting out, Jason Haupt for running my Grid jobs and for his assistance with various coding issues, Bryan Dahmes for his guidance in the HF trigger project, and Giovanni Franzoni for his involvement in the work leading up to this thesis. I would also like to acknowledge Phillip Dudero and Vladimir Rekovic for their work in the Z shape analysis. Finally, thank you to the CMS group – I had a blast working with everyone.

References

[1] C. Amsler et al. (Particle Data Group). The review of particle physics. *Physics Letters*, **B667**(1), 2008.

[2] CMS Collaboration. Study of the Z→ee differential cross section as a function of z rapidity at sqrts = 10 TeV. (CMS-PAS-EWK-09-005).

[3] G. Cowan. A survey of unfolding methods for particle physics. Advanced Statistical Techniques in Particle Physics, Durham, 2002.

[4] R. Devenish and A. Cooper-Sarkar. *Deep Inelastic Scattering*. Oxford University Press, New York, 2004.

[5] S. Glazov. Measurement of DIS cross section at HERA. *Brazilian Journal of Physics*, **37**(2C), June.

[6] D. Griffiths. *Introduction to Elementary Particles*. WILEY-VCH, Weinheim, 2004.

[7] E. Hand. Lepton photon 2009: An LHC in every home, August 2009.

[8] B. Povh, K. Rith, C. Scholz, and F. Zetsche. *Particles and Nuclei*. Springer-Verlag, Berlin, 6th edition, 2008.

[9] J. Pumplin, D. R. Stump, J. Huston, H.-L. Lai, P. Nadolsky, and W.-K. Tung. New generation of parton distributions with uncertainties from global QCD analysis. *Journal of High Energy Physics*, **2002**(07):012, 2002.

[10] M. Schott. Z boson production at LHC with first data. *Journal of Physics: Conference Series*, **110**(4):042024, 2008.

[11] T. Sjostrand, S. Mrenna, and P. Skands. PYTHIA 6.4 physics and manual. *JHEP*, **0605**:026, 2006.

[12] D. Stump, J. Huston, J. Pumplin, W.-K. Tung, H.-L. Lai, S. Kuhlmann, and J. F. Owens. Inclusive jet production, parton distributions, and the search for new physics. *Journal of High Energy Physics*, **2003**(10):046, 2003.

A Relativistic Kinematics

In the relativistic regime, a particle's momentum ($\vec{p}$) and energy (E) no longer obey the classical equations $\vec{p} = m\vec{v}$ and $E = p^2/2m$. Instead, we define these quantities as follows:

$$\mathbf{p} = \gamma m \mathbf{v} \tag{20}$$

$$E = \gamma m c^2 \tag{21}$$

where c is the speed of light, and

$$\gamma = \frac{1}{\sqrt{1 - v^2/c^2}} \tag{22}$$

Here, the factor γ is responsible for the particle's relativistic behavior at high velocities. It is convenient to combine these quantities into the energy-momentum four-vector:

$$p = \left(\frac{E}{c}, p_x, p_y, p_z\right) = (p^0, \mathbf{p}). \tag{23}$$

Recalling that the classical dot product of any two three-vectors, say $\vec{a} \cdot \vec{b}$, is invariant under a rotation of coordinates, we define the following scalar product for energy-momentum four-vectors:

$$p \cdot q = (p^0 q^0 - \mathbf{p} \cdot \mathbf{q}) \tag{24}$$

This operation is invariant under Lorentz transformations in spacetime. Taking advantage of this fact, we can write

$$p \cdot p = \frac{E^2}{c^2} - \mathbf{p}^2 = \frac{E_{rest}^2}{c^2} = m^2 c^2 \tag{25}$$

Hence, we have the relation

$$E^2 - \mathbf{p}^2 c^2 = m^2 c^4 \tag{26}$$

In relativistic collisions, all four components of the four-vector are conserved.

B Rapidity

At CMS, the scattering angle of a particle is measured in rapidity, defined by the relation [1]:

$$Y = \frac{1}{2} \ln \frac{E + p_z}{E - p_z} \tag{27}$$

where E is the energy of the particle and p_z is its momentum along the beam axis. The energy and momentum can then be written as

$$E = m_T \cosh Y,\, p_x,\, p_y,\, p_z = m_T \sinh Y, \tag{28}$$

where m_T is the transverse mass, given by:

$$m_T^2 = m^2 + p_x^2 + p_y^2 \tag{29}$$

The rapidity can also be written as:

$$Y = \ln \left(\frac{E + p_z}{m_T} \right) = \tanh^{-1} \left(\frac{p_z}{E} \right) \tag{30}$$

The rapidity variable is used in collider experiments primarily because it is invariant under Lorentz boosts. For example, a boost in the z-direction to a reference frame with velocity β alters the rapidity by an additive constant: $Y \to Y - \tanh^{-1} \beta$. Thus, the shape of the rapidity distribution, dN/dY, is independent of the reference frame used.

Also a convenient variable for scattering measurements, the pseudorapidity η, is defined as follows:

$$\eta = -\ln \left[\tan \left(\frac{\theta}{2} \right) \right] = \frac{1}{2} \ln \left(\frac{|\mathbf{p}| + p_z}{|\mathbf{p}| - p_z} \right) \tag{31}$$

We can then define the transverse momentum and energy – the components perpendicular to the beam axis:

$$p_T = |\mathbf{p}| / \cosh \eta \tag{32}$$

$$E_T = E / \cosh \eta \tag{33}$$

Experimentally, these quantities can be deduced by measuring a particle's total momentum (or energy) and its trajectory in the detector.

C Kolmogorov-Smirnov Test

The Kolmogorov-Smirnov test is a non-parametric goodness-of-fit test which can be employed to compare two distributions. For a discrete data set of n events X_i, we compute the empirical distribution function:

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} I_{X_i \leq x} \tag{34}$$

where $I_{X_i \leq x}$ is defined by:

$$I_{X_i \leq x} = \begin{cases} 1 & \text{if } X_i \leq x \\ 0 & \text{if } X_i > x \end{cases} \tag{35}$$

The function $F_n(x)$ is analogous to the integral $\int_{-\infty}^{x} f(x')\, dx'$, also known as the cumulative distribution function, for a continuous probability density $f(x)$. The Kolmogorov-Smirnov statistic is given by:

$$D_{n,n'} = \sup_x |F_n^1(x) - F_{n'}^2(x)| \tag{36}$$

where $F_n^1(x)$ and $F_{n'}^2(x)$ are empirical distribution functions of the samples we wish to compare. Given a KS statistic $D_{n,n'}$ and a sample of size n, we can obtain a confidence level comparing the correlation between $F_n^1(x)$ and $F_{n'}^2(x)$ with that of the null hypothesis.