matteo.tresoldi0@gmail.com

Squeezing the proton: estimating and improving parton distribution uncertainties for precision electroweak measurements

Matteo Tresoldi
Supervisor: Emanuele Angelo Bagnaschi
CERN, CH-1211 Geneva, Switzerland

Keywords: W boson mass, PDF uncertainties, Electroweak interaction, Precision measurements

Summary

We consider the production of a lepton and a neutrino each with large transverse momentum at the LHC and the Tevatron, in the so called charged current Drell-Yan process. We propagate the uncertainties due to the knowledge of the proton structure to the prediction of several observables defined for this scattering process. We study the covariance with respect to variations of the proton parameterisation of the bins of the considered kinematical distributions. We introduce this covariance in the formulation of a fit that determines the preferred value of the W boson mass from the analysis of the experimental data and apply this procedure to different sets of proton parameterisations. We study the robustness of the determination of the PDF covariance matrix from Monte Carlo data. We study the covariance matrix under the variation of m_W to understand the difference in the dependence of the differential distributions between m_W and the PDF uncertainty.

Contents

1 Introduction

At the present time, the Standard Model of Elementary Particles describes the electromagnetic, strong and weak interactions and classifies all known elementary particles. The Standard Model has been enormously successful in providing theoretical predictions which match experimental results, but leaves unexplained many phenomena, most notably the matter-antimatter asymmetry and neutrino oscillations, which implies that at least two of the three neutrinos have a non-zero mass - in obvious contrast with the massless neutrino hypothesis of the Standard Model. Many theoretical physicists have spent a great deal of time trying to extend the Standard Model - which was completed in the mid 1970s - but up to this day none of these extensions are supported by experimental observations. The Standard Model now needs to be put under even more rigorous tests and that's why theoretical physicists are currently working on making very precise predictions in the hope of finding small discrepancies when compared to accurate experimental measurements. For example, the recent measurement of the anomalous magnetic dipole moment of the muon at Fermilab Muon g-2 [5] indicated a 3.7 standard deviations disagreement with respect to the current best Standard Model prediction [2], and may shed light where speculations have proliferated for too many years if proved to be a real discrepancy instead of just an unfortunate series of events that sometimes happen, like a random noise or an unlucky accumulation of little mistakes here and there.

Whether it be true or not, making very precise predictions through the Standard Model is, as of now, the only effective way physicists have to look for where tiny effects of New Physics may be at fault. In this regard, the electroweak sector of the Standard Model offers the possibility of making particularly precise tests. In fact, the mass of the W boson m_W

is connected to the masses of the top-quark and the Higgs boson (see for example Ref.[9]). Measuring m_W then tests this prediction and the overall self-consistency of the Standard Model (see Fig. 30 of Ref.[1]).

This work is meant to be a continuation of Ref.[3], whose authors have suggested that the uncertainties on the knowledge of the structure of the proton might not be an insurmountable limitation for high-precision measurements of the Standard Model parameters, but there are still some aspects that are not clear and need to be investigated thoroughly.

It has to be said that there are many sources of experimental and theoretical errors that need to be studied carefully, but in this report we will concentrate on the uncertainties due to the knowledge of the proton and ignore the others. For these reasons, we will study one of the many processes that experimental physicists have investigated for decades at CERN and Fermilab: the Drell-Yan process. Given a set of experimental data (pseudodata) we will employ a fitting technique to extract m_W from this data and propagate the uncertainties due to the knowledge of the proton to the measurement of the mass of the W boson. Of particular interest for this report is the recent high-precision measurement of the W boson mass with the CDF II detector [10]. The authors of this paper have asserted through their research that the mass of the W boson is significantly higher than the best Standard Model prediction with a discrepancy of 7 standard deviations. Were it to be undoubtedly true, a new revolution would have already started in physics, but in their work there are several critical aspects that have casted many doubts among the particle physics community. In this work, we study in more details the parton distribution uncertainties with the goal of trying to validate or confute part of their work.

2 The charged-current Drell-Yan process

In 1970 Drell and Yan [8] pointed out that Feynman's parton model of hadrons [11] - which he developed to study electron-proton and electron-neutron deep inelastic scatterings - could be extended to high-energy hadron-hadron collisions. Originally, they applied the parton model to account for the production of high-mass lepton-antilepton pairs as an electromagnetic process in which a quark and an anti-quark coming from a pair of interacting hadrons annihilates and then converts to a high-mass lepton-antilepton pair, namely the square of the invariant mass of the final state is $M^2 = (p_{\ell^-} + p_{\ell^+})^2 \gg 1\,\mathrm{GeV}^2$. Drell and Yan proposed that the quark-antiquark pair, after the annihilation, are at first converted into a Z boson or a virtual photon γ^* which then decays into a lepton-antilepton pair. In a simple and naive parton model, the cross section of such a process can be obtained by weighting the cross sections of the subprocesses $q\bar{q} \to \ell^-\ell^+$ with the proton collinear Parton Distribution Functions (PDFs) of the quark $f_q(x, Q^2)$ and anti-quark $f_{\bar{q}}(x, Q^2)$, where Q^2 is the squared invariant momentum transfer, and summing over all quark-antiquark combinations involved in the process. The Drell-Yan mechanism can be generalized and applied to the production of lepton-neutrino pairs each with large transverse momentum, at the CERN LHC, via proton-proton collisions, by simply replacing the differential cross section of the subprocess with that of the subprocess $d\hat{\sigma}(ab \to W \to \ell\nu_\ell + X, Q^2)$ - the sum and the PDFs have also to be changed accordingly.

$$d\sigma(pp \to W \to \ell\nu_\ell + X, Q^2) = \sum_{a,b} \int_0^1 dx_1 \int_0^1 dx_2 f_a(x_1, Q^2) f_b(x_2, Q^2) \\ \times d\hat{\sigma}(ab \to W \to \ell\nu_\ell + X, Q^2) \tag{1}$$

where a and b can be quarks, antiquarks, gluons or photons, and the summation is over all the possible parton flavors; X includes everything that we left out, x_1 and x_2 are the fraction of the longitudinal momentum of the respective parent hadron that is being carried by a quark or antiquark and $d\hat{\sigma}(ab \to W \to \ell\nu_\ell + X, Q^2)$ is the cross section of the subprocess. Drell and Yan derived this cross section by assuming that there are no parton-parton interactions, which stems from the hypothesis that the mass of the lepton-antilepton pair is assumed to be large when compared the mass of a nucleon. Thus, Heisenberg uncertainty principle applied to energy and time implies that the time of interaction between the quark and antiquark is short on the nuclear time-scale, in other words, the annihilating quark-antiquark pair doesn't interact with other components of the original hadrons.

The process of Eq.(1) is mediated by the W boson, and is called charged-current Drell-Yan process. Its counterpart, the one mediated by a virtual photon γ^* or a Z boson, is the neutral-current Drell-Yan process.

3 Kinematics

The previous section presented a general statement of the Drell-Yan process, now we need to define some observables that, if measured, may help us study this process and hopefully the W boson. The formulation we are going to provide is shared by other processes, so we will, at first, give a general recount and then return to the charged-current Drell-Yan process.

As an example, we will consider the Large Hadron Collider (LHC) at CERN. Inside the accelerator, two high-energy particle beams travel at a speed close to the speed of light. The beams travel in opposite directions in separate beam pipes kept at ultrahigh vacuum and are made to collide when physicists want to. The basic picture of such a collision is the collision between two protons traveling at the same speed, but in opposite directions. During the LHC Run 2, the center of mass energy was $\sqrt{s} = 13\,\text{TeV}$, so each proton carries a $6.5\,\text{TeV}$ momentum. It follows that these protons are ultrarelativistic and to simplify the calculations we have neglected the mass of the protons. Conventionally, the direction of their momentum is taken to be the z axis. When such a high-energy collision happens, it's the partons of which the protons are made of that are colliding against each other, and they carry a certain fraction of the momentum of the proton they come from with probability given by their respective PDF. So, now we concentrate on the partons. The two partons annihilates to form an intermediate particle which then decays. Generally, we are interested in measuring quantities associated to the products of the decay of the intermediate particle and then reconstruct quantities associated to this particle. Let's assume that the intermediate particle decays into two particles. In this case we might be interested in measuring the angles of the two particles with respect to the beam axis - which we assumed to be the z axis, but these angles are not always easy to measure, we need a substitute. We have chosen an axis of reference, so geometrically, it is now natural to ponder about which observables lie in the plane transverse to the beam axis. This is important, because it is impossible to know in which specific point the collision happened and then measure quantities with respect to that point, nor do we care. Moreover, since the transverse plane is, like the name says, orthogonal to the beam axis, Lorentz Boosts along the direction the beam axis don't affect observables defined on the transverse plane. Being Physicists, a universal quantity that we would like to measure is the momentum, specifically the transverse momentum, which is the momentum of a particle transverse to the beam axis:

$$\vec{p_\perp} \equiv (p_x, p_y), \quad p_\perp \equiv |\vec{p_\perp}| \tag{2}$$

A related observable is the transverse mass. If we don't known all the four components of the four-momenta, we cannot know the invariant mass of the particles that come from the decay of the intermediate particle, so the best we can do is consider their transverse mass, which is defined as:

$$m_\perp \equiv \sqrt{(E_{\perp,1} + E_{\perp,2})^2 - (\vec{p}_{\perp,1} + \vec{p}_{\perp,2})^2} \tag{3}$$

where $E_{\perp,i}$ for $i = 1, 2$ are the transverse energy of particle 1,2 and is defined by:

$$E_\perp^2 \equiv m^2 + p_\perp^2 \tag{4}$$

4 Proton Parton Distribution Functions

4.1 Physical background

In simple terms, the proton collinear Parton Distribution Functions express the fact that the proton is not an elementary particle. In the case of Eq.(1), the cross section of the process was factorized in terms of the cross sections for the hard interaction between two partons, coming from the initial hadrons, weighted by the respective PDF of the parton. The cross sections for the scattering of the partons can be computed in perturbative QCD, while the PDFs, their weights, are universal functions that depend only on the scale μ at which the processes is happening. This property is expressed by what is commonly called factorization theorem. The factorization of the cross sections or differential cross sections is a fundamental property of QCD with which we can calculate the cross sections of processes that involve the scattering of partons, and is therefore valid at high-energies, where high means that the characteristics Compton wavelength of the processes are of the order of the proton radius. At collider energies this condition is clearly satisfied, moreover such an high-energy scale makes it so that the proton and its constituents have almost a light-like momentum, and it also stands to reason that they are collinear. Indeed, processes in which a parton may acquire a large transverse momentum are suppressed at high-energies (see Ref.[12]). Concealed inside of the proton, partons interact with each other resulting in a complex system in which every parton's momentum follows a certain distribution. When the proton is probed at a scale μ_F - the factorization scale - the probability of finding a parton of type a inside the proton with a fraction of the longitudinal momentum of the proton falling in the infinitesimal interval $[x, x + dx]$ is given by $f_a(x, \mu_F^2)dx$. These distributions are what we call Parton Distribution

Functions. The dependence of the distributions $f_a(x, \mu_F^2)$ on the scale μ_F is known and given by:

$$\frac{df_a(x, \mu_F^2)}{d \log \mu_F^2} = \frac{\alpha_S(\mu_R^2)}{2\pi} \sum_b \int_x^1 \frac{dz}{z} f_b(z, \mu_F^2) P_{ab} \left(\frac{x}{z}\right) \tag{5}$$

where μ_R is called renormalization scale, P_{ab} are the splitting functions and $\alpha_S(\mu_R^2)$ is the strong coupling constant at the renormalization scale μ_R. Eq.(5) is a really famous set of integro-differential equations called Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) evolution equations or QCD evolution equations, and can be solved in perturbative QCD if the boundary conditions are known, it is thus necessary to know the set of distributions $\{f_a(x, \mu_0^2)\}$ at a reference scale μ_0 for every x. The PDFs x-dependence cannot be calculated from first principles and must be determined experimentally by comparing the cross sections of processes like that of Eq.(1), for which the cross sections for the hard interactions are known, from the theoretical standpoint, with a satisfactory degree of accuracy.

It goes without saying that experimental data comes with uncertainties, so how do we cope with these uncertainties in the formulation of a model that extracts PDFs from the data?

4.2 The imperfect knowledge of the proton

The knowledge of the proton structure plays an irreplaceable role in describing and understanding the physics of hadron colliders. For the last 50 years much of the information on the PDFs was obtained from lepton-hadron deep inelastic scatterings, especially at HERA before it was dismantled in 2007, but different processes depend on the proton structure in different ways. In high-energy events, the study of jets provides insight into the gluon PDF of the proton, while Drell-Yan processes are a sensitive tool to probe sea quark distributions at the LHC. Furthermore, at hadron colliders like the LHC and Tevatron, the production of W and Z bosons is mainly sensitive to up and down quarks and, depending on the characteristics of the process and the experimental uncertainties, it provides constraints on parton distributions or at least a cross-check. In the last years, many experimental improvements and theoretical developments have enriched the physical description of the proton. Still, our knowledge of the structure of the proton is imperfect and it may deeply affect other measurements.

As useful as they may be, PDFs come along with the particular point of view of the theoretical framework that is used to extract PDFs from experimental data and the experimental uncertainties altogether. In simple terms, PDFs are determined by **global fits** to all the available deep inelastic scatterings and proton-proton high-energy collisions data, but how do we proceed from here?

First of all, there is no commonly accepted unique set of PDFs that describes the proton. At the moment we have sets of experimental data points, each with an error, that indicate roughly the value of a PDF for a certain fraction of the longitudinal momentum. These points individuate, in a cartesian plane, regions where each of the real PDF of the proton is supposed to be. Therefore, every choice of a set of PDFs that lives inside these regions are equivalently good at a first approximation. We have established that given these limitations, extracting a single set of PDFs from the data might non to be a satisfactory answer to the problem of determining sensible PDFs. If extracting a single set of PDFs does not work, then maybe we could extract several sets. This idea stems from the desire to discretize the region in which PDFs can live. So, now the problem has been moved to discretizing this space of functions in terms of N replicas of PDFs, where N is a finite, and hopefully low, positive integer, while retaining the error band in the space of functions, which means that the discretization process must not be destructive towards statistical information. The space of functions needs to be sampled in a way that allows for statistical analysis of observables that depends on PDFs. Observables will therefore be affected by an uncertainty that depends on the PDFs, which is an uncertainty of theoretical origin like the uncertainties encountered in perturbation theory.

4.3 The Hessian matrix method

The basic idea of the Hessian matrix method is that we have to somehow take into account $N_{sources}$ sources of systematic errors in a best-fit paradigm. This can be done in several ways, but an important point is that given a best-fit method for the data, or in other words, given a χ^2 function, we can study its behavior around its global minimum - assuming that the aforementioned χ^2 function is sensible. The key point is that the goodness-of-fit that we want to minimize with respect to the free parameters has to be to useful to the assessment of the uncertainties of the extracted PDFs.

The two equivalent methods proposed in Refs.[15, 13] offer a way to exploit experimental systematic errors and estimating the uncertainties of the PDFs, but their details are out of the scope of this report. Nevertheless, we would like to present some of their aspects. What matters is that we know how the prescribed χ^2 changes in a neighborhood of its global minimum through an eigenvector basis - made of $N_{sources}$ vectors - of the Hessian matrix of χ^2 calculated at its global minimum. From this method, $2N_{sources} + 1$ replicas of PDF are extracted from the data. One is the best-fit PDF replica S_0, the central PDF replica, the other $2N_{sources}$ PDF replicas come from moving backwards and forwards in the direction of each of the eigenvectors. If $\mathcal{O}$ is an observable that is computed through PDF, then given the $2N_{sources} + 1$ replicas of PDF, the uncertainty of the observable is given by the following formula:

$$\Delta\mathcal{O} = \frac{1}{2}\sqrt{\sum_{k=1}^{N_{sources}} \left[\mathcal{O}(S_k^+) - \mathcal{O}(S_k^-)\right]^2} \tag{6}$$

where $\mathcal{O}(S_k^\pm)$ are the predictions for $\mathcal{O}$ computed through the two PDF replicas individuated by the k^{th} eigenvector. Additionally, if $\mathcal{O}_1$ and $\mathcal{O}_2$ are two observables that might be correlated or not, their covariance and correlation are given by:

$$\Delta(\mathcal{O}_1, \mathcal{O}_2) = \frac{1}{4}\sum_{k=1}^{N_{sources}} \left[\mathcal{O}_1(S_k^+) - \mathcal{O}_1(S_k^-)\right]\left[\mathcal{O}_2(S_k^+) - \mathcal{O}_2(S_k^-)\right] \tag{7}$$

$$\varrho(\mathcal{O}_1, \mathcal{O}_2) = \frac{\Delta(\mathcal{O}_1, \mathcal{O}_2)}{\Delta\mathcal{O}_1\Delta\mathcal{O}_2} \tag{8}$$

4.4 The Neural Network approach of NNPDF

In recent years, cutting-edge developments in artificial intelligence techniques have been applied to many fields in physics. In particular, they have aided the implementation of new modeling techniques in high-energy physics. Among these developments, neural networks have become indispensable to fit PDFs. Fixing the functional forms of the PDFs parameterisation introduces a bias and makes it difficult to even estimate the uncertainty due to this bias. Neural networks answer the question: is there a way to fit data points in a natural and unbiased approach? Indeed, neural networks provide an unbiased PDFs parameterisation and an estimate of the PDFs uncertainties. The basic idea is to implement machine learning fitting tools to set a neural network on a Monte Carlo replica of the data and then train it. The essence of this method is viewing the input experimental data along with its correlated uncertainties like a probability distribution from which it is possible to extract random replicas of the data. A single replica is not enough to sample the space of functions in a statistically significant way, so its necessary to extract as many replicas as needed to arrive at the desired accuracy. From each Monte Carlo replica of the data a PDFs set is extracted via the minimization of a suitable goodness-of-fit. This method provides an ensemble $\{S\}$ of PDF replicas, in particular, the central replica which we now call replica 0 is just the average over the ensemble of PDF replicas. Moreover, if $\mathcal{O}$ is an observable then the mean and uncertainty of the observable with respect to PDF variations are just the average and standard deviation over the ensemble of PDF replicas, and are given by the following formulas:

$$\langle \mathcal{O} \rangle_{PDF} = \langle \mathcal{O}[\{S\}] \rangle = \frac{1}{N_{rep}} \sum_{k=1}^{N_{rep}} \mathcal{O}(S_k) \tag{9}$$

$$\sigma(\mathcal{O})_{PDF} = \sigma(\mathcal{O}[\{S\}]) = \sqrt{\frac{1}{N_{rep}-1} \sum_{k=1}^{N_{rep}} \left(\mathcal{O}(S_k) - \langle \mathcal{O}[\{S\}] \rangle\right)^2} \tag{10}$$

where N_{rep} is the number of PDF replicas, and from now on we will use the label PDF to avoid having to specify that the quantity that we are considering is computed on the ensemble of PDF replicas $\{S\}$. Similarly, if $\mathcal{O}_1$ and $\mathcal{O}_2$ are two observables, their covariance and correlation are given by:

$$\sigma(\mathcal{O}_1, \mathcal{O}_2)_{PDF} = \frac{1}{N_{rep}-1} \sum_{k=1}^{N_{rep}} \left(\mathcal{O}_1(S_k) - \langle \mathcal{O}_1 \rangle_{PDF}\right)\left(\mathcal{O}_2(S_k) - \langle \mathcal{O}_2 \rangle_{PDF}\right) \tag{11}$$

$$\rho(\mathcal{O}_1, \mathcal{O}_2)_{PDF} = \frac{\sigma(\mathcal{O}_1, \mathcal{O}_2)_{PDF}}{\sigma(\mathcal{O}_1)_{PDF}\sigma(\mathcal{O}_2)_{PDF}} \tag{12}$$

5 The determination of m_W from a set of data

5.1 A templates fit technique

At hadron-hadron colliders like the CERN LHC or Fermilab Tevatron, m_W is determined from the measurements of observables of the charged-current Drell-Yan process like the transverse momentum of the charged-lepton $p_\perp^\ell$ and the the transverse mass of the W boson (of the lepton-neutrino pair) $m_\perp^W$. In this report, we concentrate on the differential cross sections with respect to the aforementioned kinematical observables, whose sensitivity to m_W is enhanced by the presence of a kinematical peak.

In the previous section, we explained that the experimental uncertainties of the PDFs are an essential aspect that cannot be ignored and that it is represented via a collection of PDF replicas that are equivalently good from a theoretical point of view, leading to a source of theoretical systematic error that has to be accounted for in a fitting technique meant to extract m_W from a set of data. The first step consists of using the mathematical tools of the Standard Model and Monte Carlo simulations to generate the theoretical predictions for some distributions of Drell-Yan observables. These theoretical distributions are called templates and are computed at a certain perturbative degree for some input parameters, one of which is the mass of the W boson. The problem is that the templates introduce a theoretical systematic error which is due to the degree at which we stopped the perturbative expansion, but also to modeling uncertainties, and PDF uncertainties. The distorsions on the templates make the measurement of m_W really challenging. Distortions at the per mille level are enough to induces shifts on the order of $\mathcal{O}(10\,\mathrm{MeV})$ on m_W. These uncertainties have been thought to be a bottleneck in the measurements of m_W, but, as already mentioned, all of the studies expect for Ref.[3] used a fitting procedure whose goodness-of-fit is the common χ^2. The intuition of the the author of Ref.[3] is that the uncertainty introduced by the PDFs is highly correlated and can be exploited to make precision measurements by minimizing a suitable goodness-of-fit. Following their footsteps, we introduce this systematic error in the definition of the following goodness-of-fit via a collection of free parameters:

$$\chi_k^2(\alpha) = \sum_{i \in bins} \frac{((\mathcal{T}_{0,k})_i - (\mathcal{D}^{exp})_i - \sum_{r\in\mathcal{R}} \alpha_r (\mathcal{S}_{r,k})_i)^2}{\sigma_i^2} + \sum_{r\in\mathcal{R}} {\alpha_r}^2 \tag{13}$$

where the parameters α_r are nuisance parameters that help describe the PDFs as sources of systematic errors. Every α_r is enumerated by the index r that takes values in $\mathcal{R}$, the space of PDF replicas. For every bin i of the distribution, $(\mathcal{D}^{exp})_i$ is the experimental value of the bin, $(\mathcal{T}_{0,k})_i$ is the corresponding value of the bin predicted by the Standard Model under the hypothesis that the mass of the W boson is $m_{W,k}$ computed with the replica 0 of the PDFs set. This is an important point. The Lagrangian of the Standard Model depends on m_W and so does every prediction that we wish to make about Drell-Yan cross sections. It is therefore necessary to make a choice for m_W if we wish to make predictions that have to be compared to experimental data. If the PDFs sets have been extracted by the NNPDF collaboration, then the differences between the predictions of the Standard Model with respect to PDFs set variation $\mathcal{S}_{r,k} = (\mathcal{T}_{r,k} - \mathcal{T}_{0,k})/\sqrt{N_{rep}-1}$ are the systematic errors of theoretical origin that we are trying to account for in this fitting procedure. Contrarily, in the case the PDFs sets have been extracted through the hessian method, then $\mathcal{S}_{r,k} = (\mathcal{T}^+_{r,k} - \mathcal{T}^-_{r,k})/2$. The theoretical distributions, which from now on we will call templates, are not only affected by the choice of m_W, but also by statistical Monte Carlo errors, moreover there may also be experimental errors, but for the sake of simplicity we will ignore these errors. The uncertainty σ_i of the i^{th} bin is therefore the Monte Carlo error of this bin, $\sigma_i = \sigma_i^{MC}$. Finally, the quadratic penalties $\sum_{r\in\mathcal{R}} {\alpha_r}^2$ represent a Gaussian penalty that we associate to the PDF replicas. A crucial point of this section are exactly the nuisance parameters α_r. Our final goal is extracting m_W from experimental data, but before that we need to make a choice for the nuisance parameters. In principles, we could choose them however we want, but it doesn't guarantee that they are the optimal choice that defines a reasonable χ^2. We are for this reason motivated to opt for minimizing every $\chi^2_k(\alpha)$ with respect to the nuisance parameters α_r. For every choice of $m_{W,k}$, $\chi^2_k(\alpha)$ is minimized by a collection of parameters $\alpha^k_{r,min}$ that depends on the choice of the W boson mass, but remarkably when substituted in $\chi^2_k(\alpha)$ they give the same result for every $m_{W,k}$ as proved in Ref.[7]. The final result is:

$$\chi^2_{k,min} = \sum_{(a,b)\in bins} (\mathcal{T}_{0,k} - \mathcal{D}^{exp})_a (C^{-1})_{ab} (\mathcal{T}_{0,k} - \mathcal{D}^{exp})_b \tag{14}$$

so, the minimum of $\chi^2_k(\alpha)$ with respect to the nuisance parameters is simply a quadratic form represented by the inverse of a matrix that contains the covariances with respect to PDF variations between bin a and bin b of the distribution. The preferred value for m_W is taken to be $m_{W,j}$, where j selects the smallest χ^2 among the sequence $\{\chi^2_{k,min}\}$. For the scope of this report it is also fundamental to associate an error to the preferred value of the W boson mass. The rule $\Delta\chi^2 = 1$ indicates the range of values of m_W in which we expect to find the true value of the mass of the W boson with a 68% confidence level. Half of this interval is the standard deviation.

In our case C is, for example, given by:

$$C = \Sigma_{PDF} + \Sigma_{MC} \tag{15}$$

where Σ_{MC} is a diagonal matrix whose elements are the Monte Carlo variances of every bin of the distribution, $(\Sigma_{PDF})_{ab}$ is the covariance between bin a and bin b with respect to PDF variations, which are given by Eq.(7) and Eq.(11).

We feel obliged to highlight that neglecting every other sources of errors, except for the PDF and Monte Carlo uncertainties, is not a limitation for the scope of our report. It is not just a mere simplification, and can be seen in a broader way. Imagine that we have at our disposal such a huge amount of experimental data that their statistical error is negligible and that the sources of systematic errors are under control and likewise negligible. Would it be possible to extract a very accurate measurement of m_W? The answer is no. The reason comes from Eq.(14): the templates are affected by theoretical uncertainties that are not under our control, in particular PDF uncertainties.

5.2 The role of pseudodata

In the previous section we presented a fitting technique to extract m_W from a set of experimental data, yet our real goal, in this report, is not measuring m_W, but rather presenting a new method with which the mass of the W boson may be extracted with a more accurate estimation of the PDF uncertainty. For this reason we let some theoretical distributions be the experimental data, meaning that we treated them as if they were real experimental data and used them to validate our method. In fact, if we generate templates with a given PDFs set for various hypothesis of the W boson mass and then let the template generated with the $m_{W,0}$ mass hypothesis act like experimental data, or in other words, if we treat a specific template as **pseudodata**, we expect the sequence $\chi^2_{k,min}$ to be, in an appropriate cartesian plane, a parabola centered at $m_{W,0}$ with value 0. This choice provides an immediate cross-check to the well-being of the data simulated through Monte Carlo simulations. More importantly, if we fit the same pseudodata with templates generated with other PDFs set we should not expect the χ^2 parabola to still be centered at $m_{W,0}$ with value 0. Templates generated with different PDFs sets induce shifts on the preferred value of the W boson mass, call $m_{W,j}$ the new value. The shifts $m_{W,j} - m_{W,0}$ are the representation of our imperfect knowledge of the proton structure as a source of theoretical systematic error. These shifts allows us to compare models via the theoretical distributions generated with their application and is, therefore, not limited to PDF uncertainties and can also be used to study how much certain types of corrections influence the final results, for example we could choose to ignore some radiative or QCD corrections. This approach is useful because it can give insights to the order of magnitude of the errors pertaining the corrections that we chose to exclude and then add them to the final results as another component of the theoretical systematic error. The method is also useful from the experimental standpoint. The shifts $m_{W,j} - m_{W,0}$ that one observes when comparing the pseudodata with several templates generated using different PDFs sets is a measure of the difference that experimental physicists would obtain if they fitted the data once with the templates generated with a PDFs set and then with another set. Studying the PDF uncertainty is therefore essential to make sure that the measured values of m_W are both sensible and not afflicted by an excessive PDF uncertainty.

Moreover, when pseudodata are fitted with templates with different PDFs sets, the $\Delta\chi^2 = 1, 4, 9$ rules become questionable and might not be able to estimate the uncertainty in a statistically significant way.

6 Numerical results for the PDF covariance matrix

6.1 Computational setup

We use the charged-current Drell-Yan Monte Carlo event generator of POWHEG-BOX [4] to simulate a sample of $\sim 100 \times 10^6$ pp events at the CERN LHC and $\sim 100 \times 10^6$ $p\bar{p}$ events at the Fermilab Tevatron and Pythia8.2 [14] for the parton shower at the center-of-mass energy of, respectively, $\sqrt{s} = 13\,\text{TeV}$ and $\sqrt{s} = 1.96\,\text{TeV}$. We generate the event sample using the `NNPDF31_nlo_as_0118` set, and then we reweight it to the PDF sets shown in Table 6.1 using the reweight framework of the POWHEG-BOX. We chose $m_{W,0} = 80.385\,\text{GeV}$ as our central value of the mass of the W boson, while the templates were generated for

$m_{W,k} \in [80.285, 80.485]$ GeV in 1 MeV step length.

	PDF Set
1	`NNPDF23_nlo_as_0118`
2	`NNPDF30_nlo_as_0118`
3	`NNPDF40_nlo_as_01180`
4	`MSTW2008CPdeutnlo68cl`
5	`MMHT2014nlo68cl`
6	`MSHT20_nlo_as_0118`
7	`CT10`
8	`CT14nlo`
9	`CT18NLO`
10	`cteq61`
11	`cteq66`
12	`CJ15NLO`
13	`ABMP16_5_nlo`

Table 1: PDF sets in which the `NNPDF31_nlo_as_0118` set has been reweighted to. Expect for the first three sets, all the others are Hessian sets.

6.2 PDF uncertainties on the distributions

The choice of PDFs influences at great lengths physical observables affecting the total cross section and distorting the shape of the differential cross section distributions. It is, for this reason, relevant to distinguish between the PDF uncertainty of a differential cross section and its normalized counterpart. For the sake of brevity, we shall refer to a differential cross section and its normalized counterpart respectively as absolute distribution and shape distribution. The PDF uncertainty of the former is generally bigger because both the shape and total area under the absolute distribution are affected by the choice of the PDFs of the proton, while the area under the latter has been normalized to 1, so the choice of PDFs of the proton influences only its shape. Moreover, it's the shape that is sensitive to the precise value of m_W through the location of its Jacobian Peak (kinematical peak), in this case the total cross section is mostly irrelevant. Since the area around a Jacobian Peak is the one that is sensitive to the precise value of m_W, we could then forget about the tails and redefine the shape distribution of, for example, the charged-lepton transverse momentum, on a shorter domain and ask that the distribution is normalized on the new domain I:

$$\left.\frac{d\sigma}{dp_\perp^\ell}\right|_I = \frac{1}{\int_{p^\ell_{\perp,min}}^{p^\ell_{\perp,max}} dp_\perp^{\ell\,\prime} \frac{d\sigma}{dp_\perp^{\ell\,\prime}}} \frac{d\sigma}{dp_\perp^\ell} \tag{16}$$

6.3 The correlation matrix

We show the correlation matrices with respect to PDF variations of the shape distributions. The observables that we considered are the transverse mass of the W boson and the charged-lepton transverse momentum with central cut at the LHC and the same observables at the Tevatron with cut on the transverse momentum of the W boson. The labels on the x and y axis indicate the observable and the bin and help identifying the sixteen matrices of which the joint correlation matrix is composed of. The distributions have been generated with the hypothesis $m_{W,0} = 80.385\,\text{GeV}$.

In these two figures, we can see that the correlation matrices exhibit a clear structure. In every block of the matrix there are well delimited zones where the correlation is either negative or positive separated by small regions where the change of sign of the correlation is abrupt. In the self-correlation matrices (the blocks on the diagonal) the Jacobian Peak plays the role of a dividing ridge. The bins before the peak are strongly correlated with each other just like those after the peak. Conversely, the bins before the Jacobian Peak are strongly anti-correlated with respect to those after the peak. A similar structure, although less defined, is also present in the cross-correlation matrices especially between observables of the same collider, while it's definitely weaker when we consider two observales from the two different colliders.

On the whole, this correlation pattern is also influenced by the choice of the PDFs set. It seems like `MSHT20nlo_as118` is enhancing every structure making them more defined and the observables more correlated.

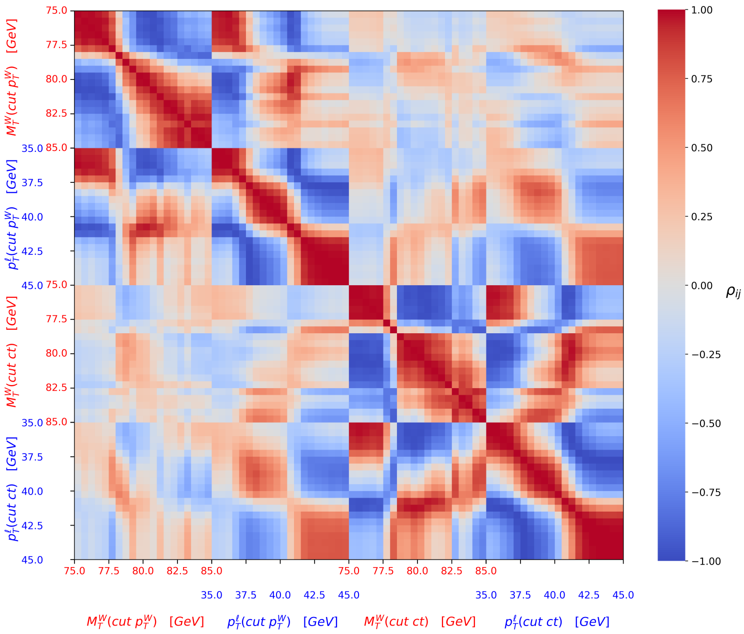

Figure 1: NNPDF31_nlo_as_0118

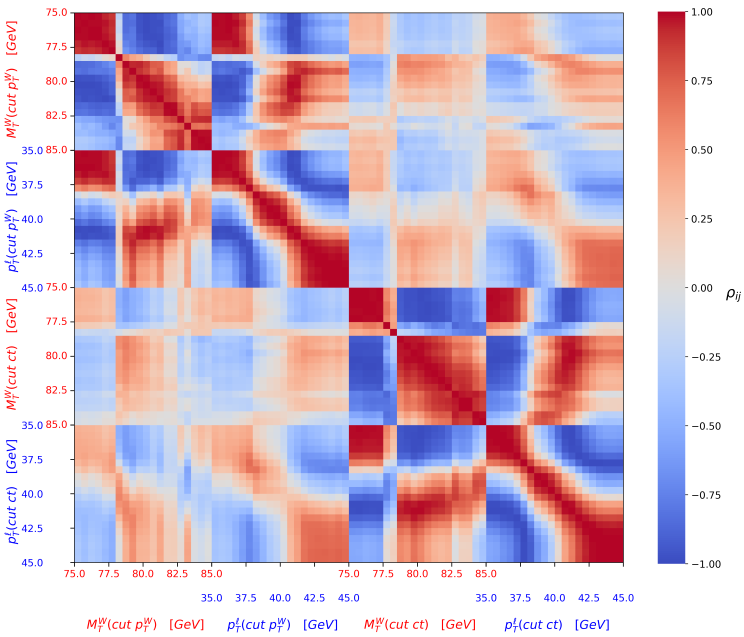

Figure 2: MSHT20_nlo_as_0118

6.4 Evolution of the eigenvalues of the PDF covariance matrix

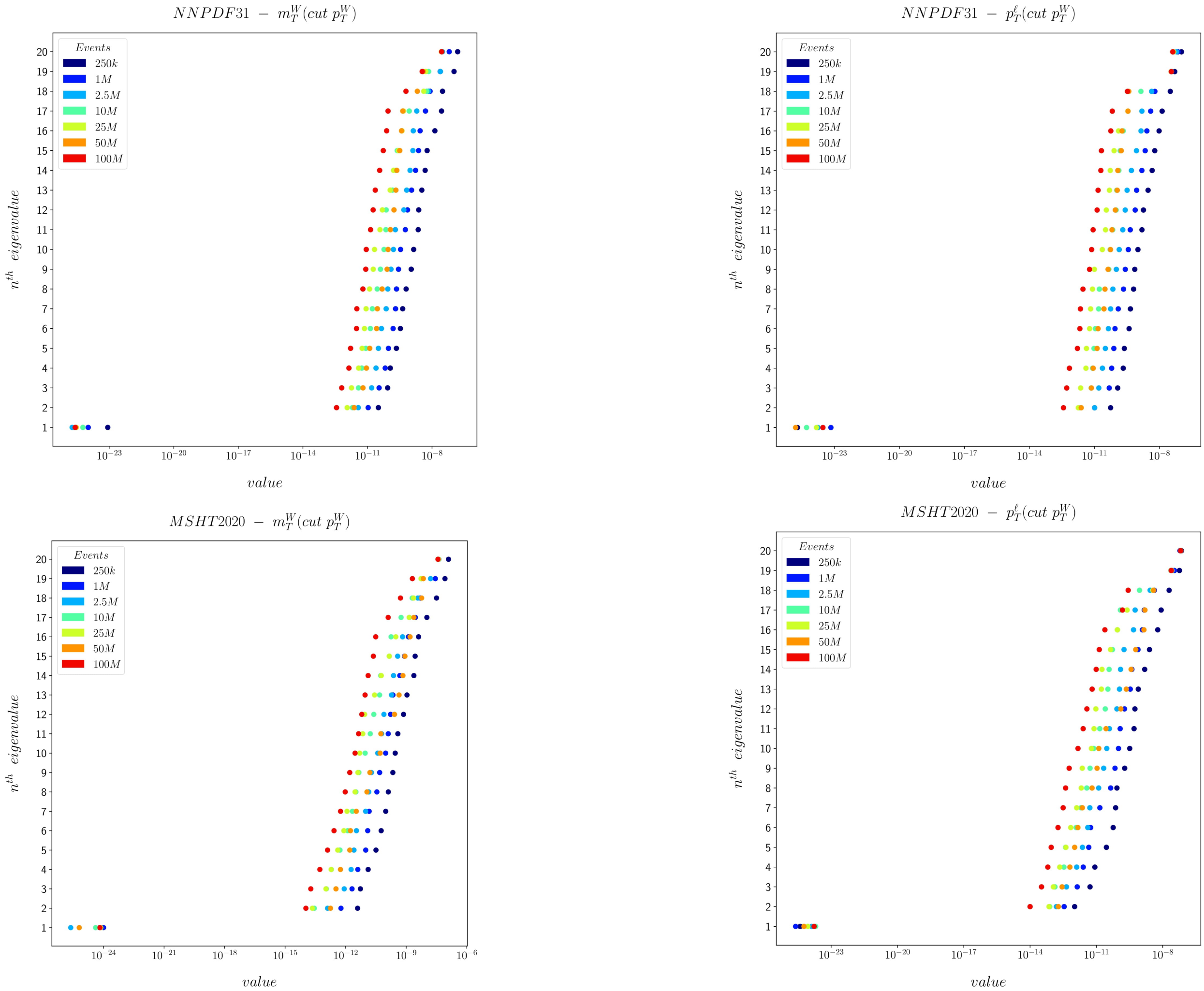

Figure 3: Evolution of the eigenvalues of the PDF covariance matrix with respect to the number of W boson production events. We considered the shape distributions with respect to the transverse mass of the W boson and the transverse momentum of the charged-lepton, both at the Tevatron with cut on transverse momentum of the W boson. The distributions have been normalized on the ranges $[75, 85]$ GeV and $[35, 45]$ GeV respectively. The distributions have been generated with the hypothesis $m_{W,0} = 80.385\,\text{GeV}$ and bins with length of $0.5\,\text{GeV}$.

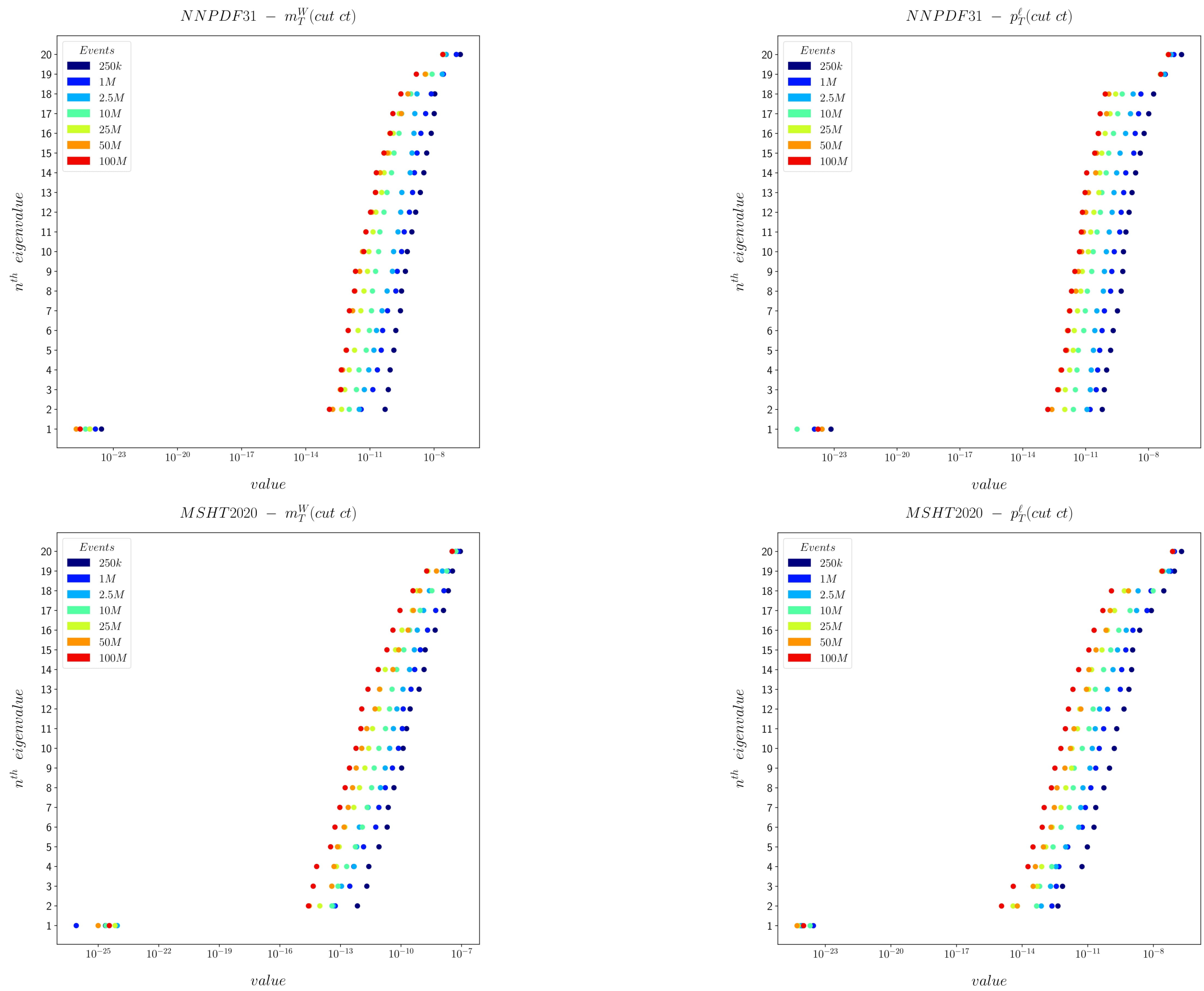

Figure 4: Evolution of the eigenvalues of the PDF covariance matrix with respect to the number of W boson production events. We considered the shape distributions with respect to the transverse mass of the W boson and the transverse momentum of the charged-lepton, both at the LHC with central cut. The distributions have been normalized on the ranges $[75, 85]$ GeV and $[35, 45]$ GeV respectively. The distributions have been generated with the hypothesis $m_{W,0} = 80.385\,\text{GeV}$ and bins with length of $0.5\,\text{GeV}$.

The goodness-of-fit that we defined in one of the previous sections depends strongly on the eigenvalues of the covariance matrix that we use. In particular, after diagonalizing the covariance matrix and considering that the goodness-of-fit is a quadratic form, it can be rewritten as a sum of terms of the form $\frac{(\tilde{B}-\tilde{T})^2}{\lambda}$, where λ is an eigenvalue of the PDF covariance matrix and $\tilde{B}$, $\tilde{T}$ are linear combinations of bins whose coefficients are the elements of the eigenvector associated to the eigenvalue λ; we use $\tilde{B}$ if the bins comes from the experimental distributions and $\tilde{T}$ if they come from the templates. It follows that if one of the eigenvalues is distinctively smaller than all the others, the width of the χ^2 will be strongly influenced by this particular eigenvalue making the distribution particularly narrow. We have studied the χ^2 distributions thoroughly and found out that it is common for the PDF uncertainty estimated via this method to be under 1 MeV. If this were true, then it would mean that the PDF uncertainty is always negligible with respect to the other sources of error, but such a strong statement needs to be properly justified. Is this behaviour physical or is it a byproduct of numerical errors? We then studied how the eigenvalues of the PDF covariance matrix change with respect to the number of events to understand which of them are physical and which are not. Those that are not should then be considered to be equal to 0: they are not 0 due to the finite precision of the algorithm one uses for computing eigenvalues. For an eigenvalue to be physical we expect it to converge as the number of events increases. In the figures above, most of the eigenvalues are not converging: this behaviour is quite intriguing and raised more questions without answering any of the previous. At this moment, we are still unable to conclude that those eigenvalues that are not converging are not physical. It may actually be due to other reasons, for example, 10^6 events might not be enough to let all the physical eigenvalues converge or maybe the PDF covariance matrix also depends on the number of replicas that constitutes the PDFs set. In this case, the `NNPDF31_nlo_as_0118` has 100 replicas, while `MSHT20_nlo_as_0118` has 65, but there exist PDFs set with 1000 replicas. We can only conclude that this aspect needs to be investigated in more details.

6.5 Dependence of the eigenvectors of the PDF covariance matrix on m_W

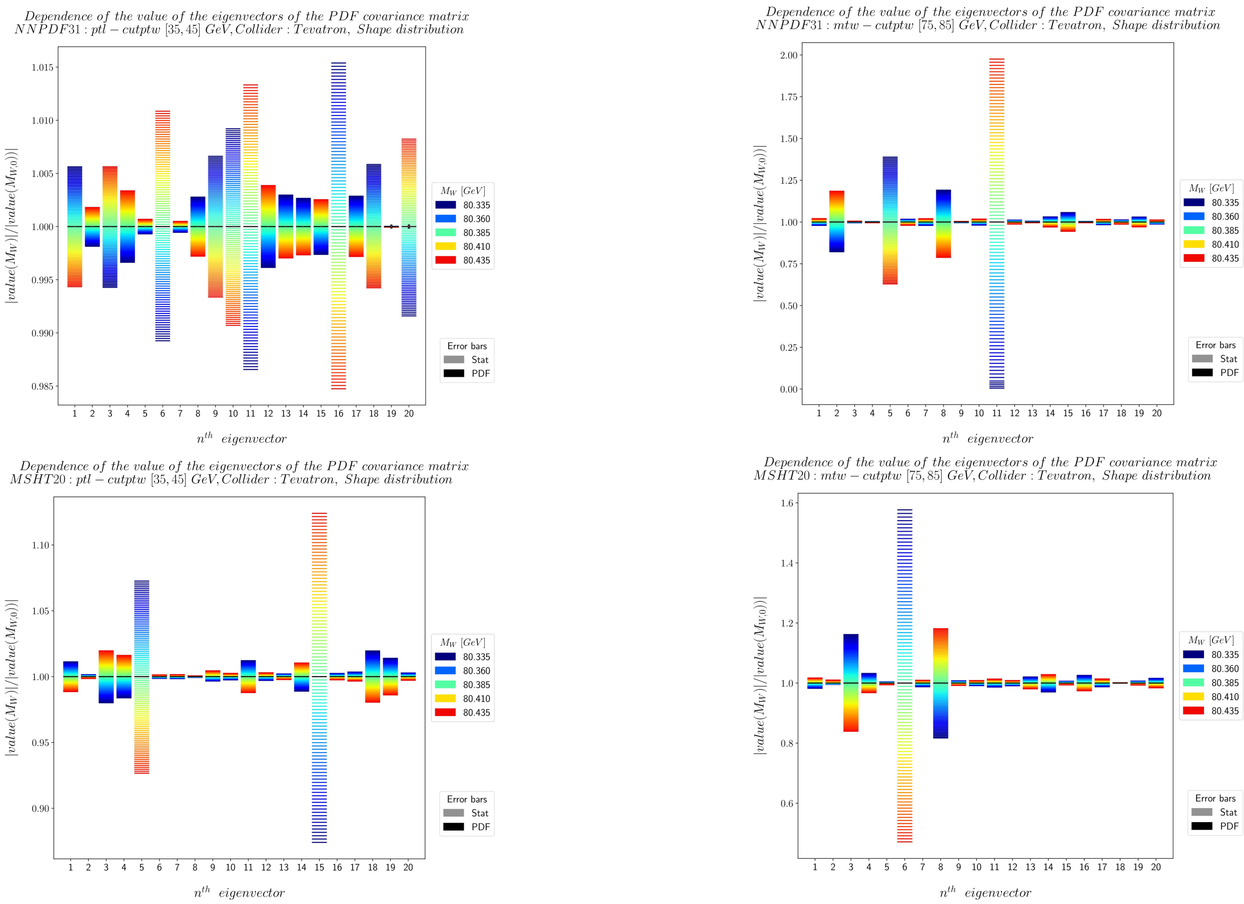

Figure 5: Dependence of the eigenvectors of the PDF covariance matrix on m_W. We considered the shape distributions with respect to the transverse mass of the W boson and the transverse momentum of the charged-lepton, both at the Tevatron with cut on transverse momentum of the W boson. The distributions have been normalized on the ranges $[75, 85]$ GeV and $[35, 45]$ GeV respectively. The distributions have been generated in the mass range $[80.335, 80.435]$ GeV and bins with length of $0.5\,$GeV.

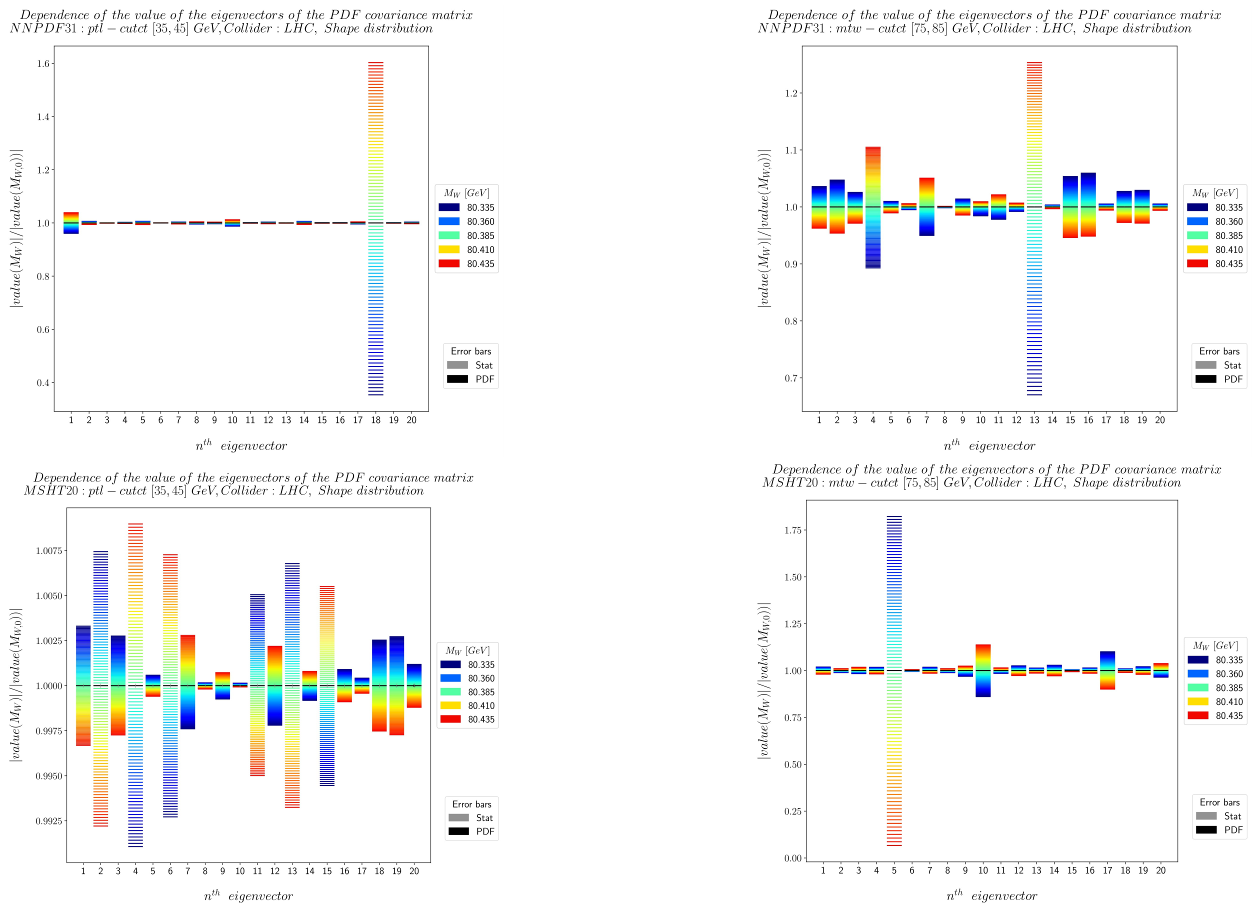

Figure 6: Dependence of the eigenvectors of the PDF covariance matrix on m_W. We considered the shape distributions with respect to the transverse mass of the W boson and the transverse momentum of the charged-lepton, both at the LHC with cut on transverse momentum of the W boson. The distributions have been normalized on the ranges $[75, 85]$ GeV and $[35, 45]$ GeV respectively. The distributions have been generated in the mass range $[80.335, 80.435]$ GeV and bins with length of $0.5\,$GeV.

In this section we further studied the PDF covariance matrix. In particular, we studied the dependence of its eigenvectors on m_W and compared their spread with their statistical and PDF errors. It is immediately clear that the spreading due to a change in m_W always wins over the statistical and PDF errors or it's of the same order of magnitude. The PDF and statistical errors never win over the dependence on m_W. Take the eigenvectors that are widely spread. A small change in m_W modifies the value of the eigenvectors significantly along with that of the χ^2 distribution. Consequently, such a change is penalized and the preferred value for m_W is brought near $m_{W,0}$. On the other hand, some eigenvectors may not change conspicuously even after a considerable change in m_W. These eigenvectors are dominated by the effect due to the statistical or PDF errors. We can therefore conclude that there are, without a doubt, some eigenvectors that can change significantly the χ^2 distribution only due to the dependence on the PDFs.

6.6 Result of the fits

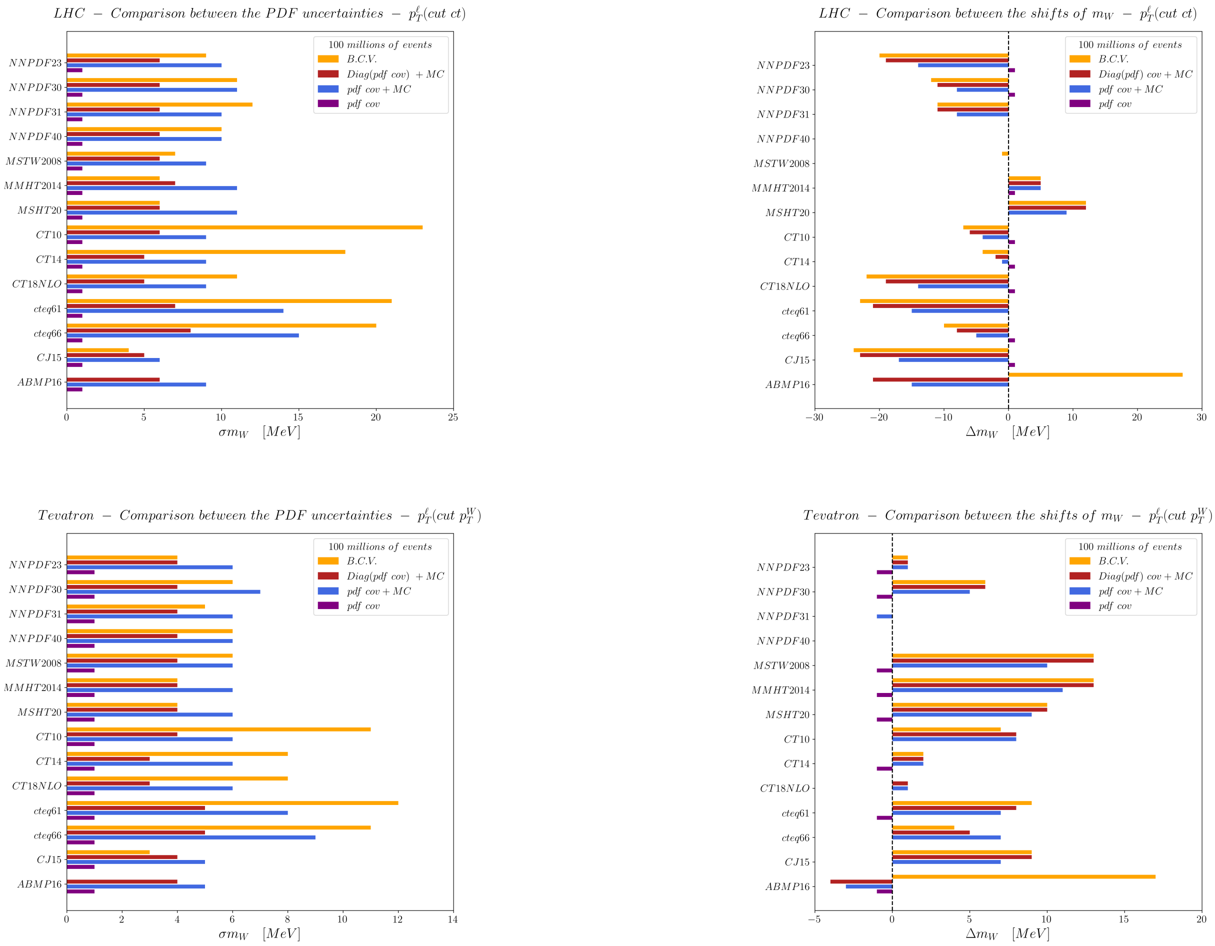

Figure 7: Comparison between the fitting techniques used to estimate the PDF uncertainty and shift on m_W. The templates were generated with NNPDF40_nlo_as_0118 and used to fit pseudodata generated with fourteen different PDFs sets in the $m_{W,0}$ hypotesis.

We applied the fitting technique we introduced in one of the previous sections using as covariance matrix the PDF covariance matrix. We have also employed the fitting technique of Ref.[6] to study how the PDF uncertainties and the shifts from $m_{W,0}$ vary with respect to the fitting method and the PDFs sets. We chose fourteen PDFs sets for the pseudodata, including NNPDF40_nlo_as_0118, and also chose this PDFs set for the templates. It follows that when the templates generated with a certain PDFs set are fitted on pseudodata generated with the same PDFs set, the shift on the best fit for m_W must be 0 as the distributions are the same and there can't be a better option to fit the pseudodata. It is interesting to notice that the fitting method of Ref.[6], which we denominated as *B.C.V.*, returns smaller PDF uncertainties, depending on the PDFs sets, when we use as covariance matrix the PDF covariance matrix plus the Monte Carlo error. This behaviour is rather curious and contributes to opening new questions around the fitting technique we employed. Why is the

$B.C.V.$ method better when it does not account for correlation? It also rather troubling to see that the PDF uncertainties are always 1 MeV when we only use the PDF covariance matrix: this undesired effect is most definitely due to some possible unphysical and really small eigenvalues of the PDF covariance matrix that makes the χ^2 distribution quite narrow. The results for the shifts of m_W are also very interesting. First of all, the shifts calculated via the $B.C.V.$ method are always bigger while those calculated with only the PDF covariance matrix are most of the time 0 or small, meaning that the possible unphysical eigenvalues make it so that any other distribution, apart from the one generated with the $m_{W,0}$ hypothesis, is extremely penalized. Such a strong statement can't be trusted before the problem has been thoroughly explored, but the results of the previous section are opening a way toward an answer for which more research is needed. We also notice that the for every PDFs setm the other shift are roughly the same. Moreover, there is no clear benefit in employing a fitting technique or the other.

7 Numerical results for the m_W covariance matrix

7.1 The correlation matrix

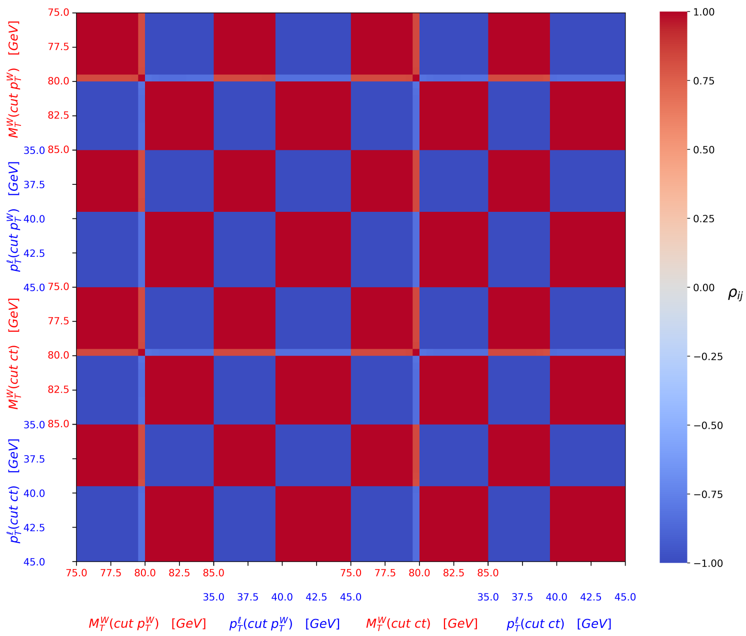

Figure 8: `NNPDF31_nlo_as_0118`

We show the correlation matrices with respect to m_W variations of the shape distributions. The observables that we considered are the transverse mass of the W boson and the charged-lepton transverse momentum with central cut at the LHC and the same observables at the Tevatron with cut on the transverse momentum of the W boson. The labels on the x and y axis indicate the observable and the bin and help identifying the sixteen matrices of which the joint correlation matrix is composed of.

This matrix exhibits an undeniably clear block structure with either perfect correlation or anti-correlation which are divided at the Jacobian Peak. This structure is shared by other observables and does not change with the PDFs set or the collider.

7.2 Evolution of the eigenvalues of the m_W covariance matrix

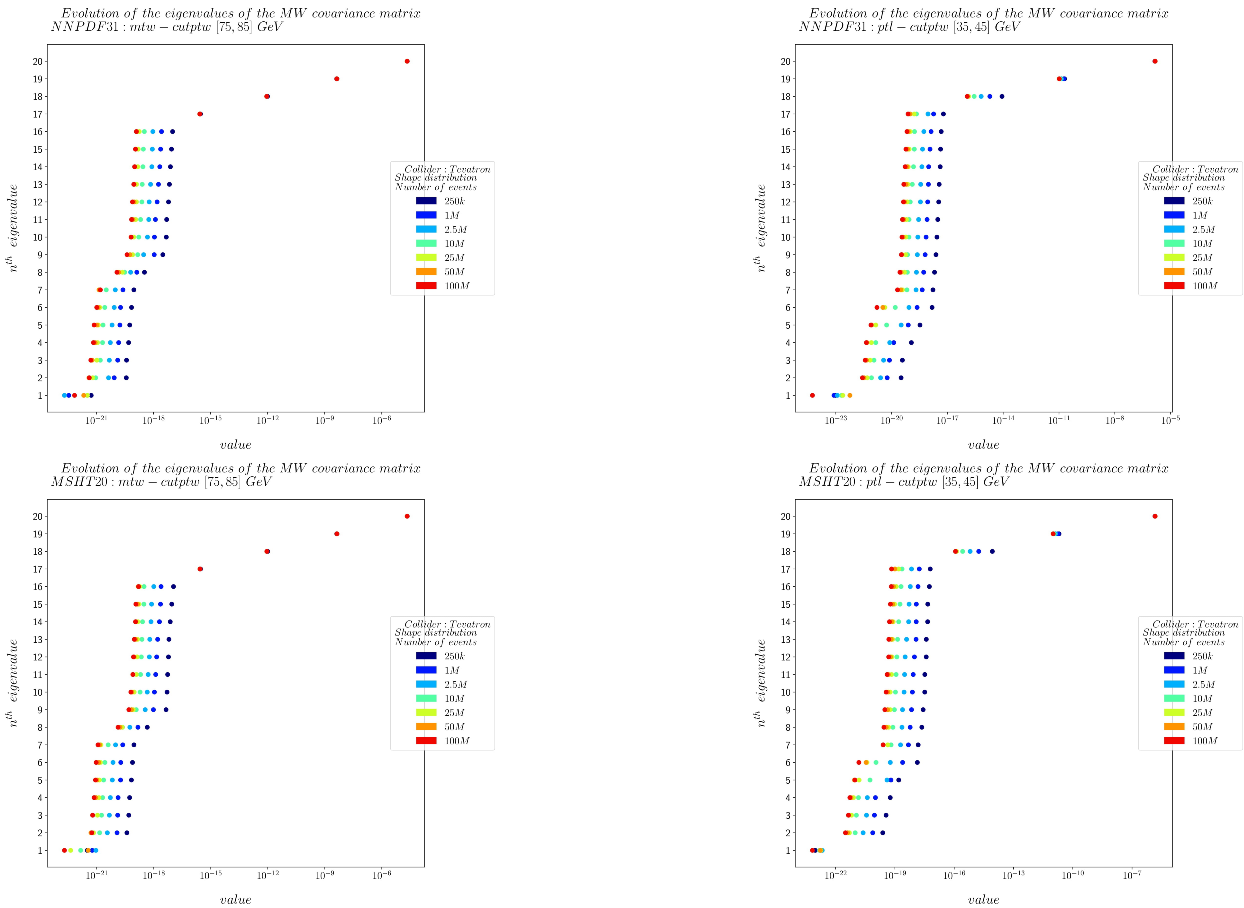

Figure 9: Evolution of the eigenvalues of the m_W covariance matrix with respect to the number of W boson production events. We considered the shape distributions with respect to the transverse mass of the W boson and the transverse momentum of the charged-lepton, both at the Tevatron with cut on transverse momentum of the W boson. The distributions have been normalized on the ranges [75, 85] GeV and [35, 45] GeV respectively. The distributions have been generated in the mass range [80.335, 80.435] GeV and bins with length of 0.5 GeV.

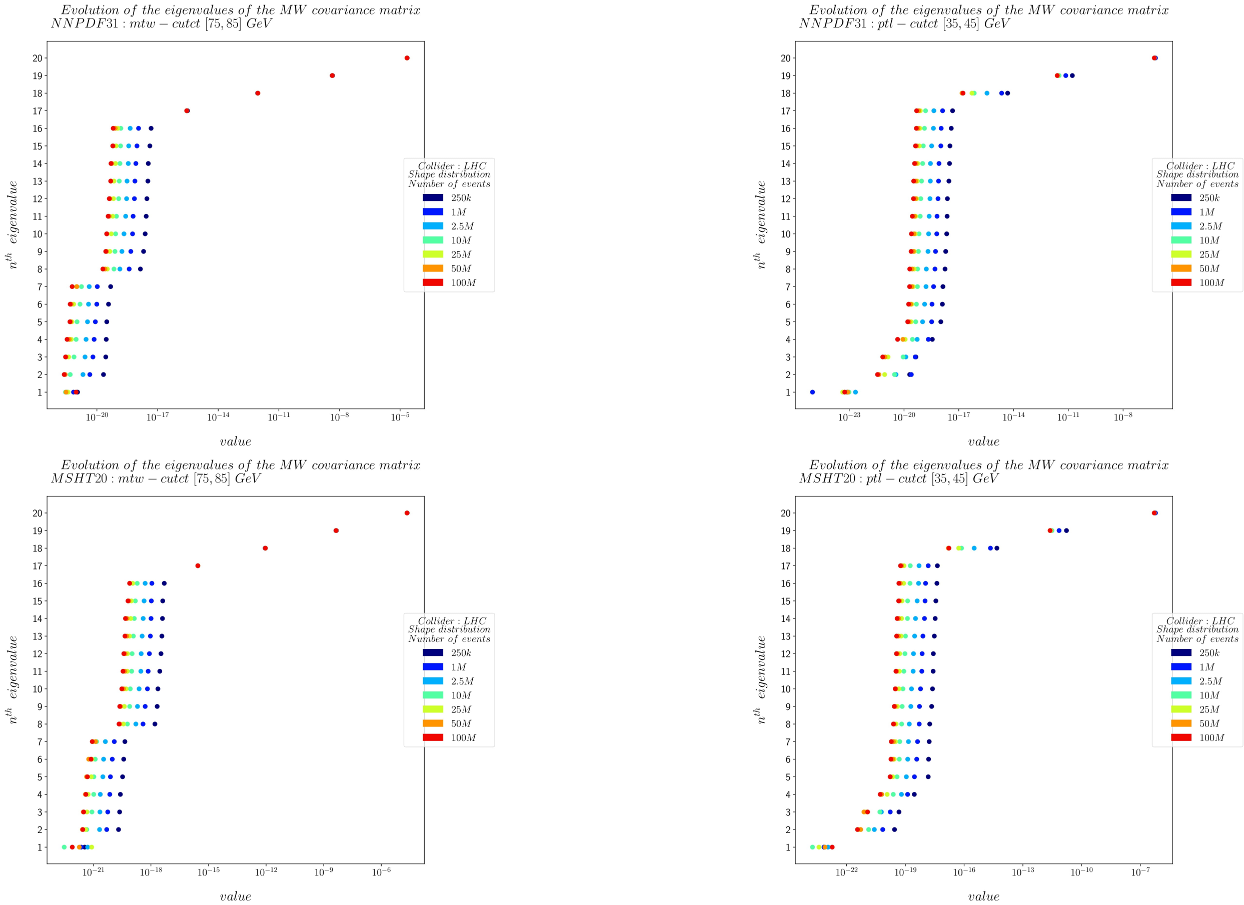

Figure 10: Evolution of the eigenvalues of the m_W covariance matrix with respect to the number of W boson production events. We considered the shape distributions with respect to the transverse mass of the W boson and the transverse momentum of the charged-lepton, both at the LHC with central cut. The distributions have been normalized on the ranges $[75, 85]$ GeV and $[35, 45]$ GeV respectively. The distributions have been generated in the mass range $[80.335, 80.435]$ GeV and bins with length of $0.5\,$GeV.

Most of the comments that we made in the section dedicated for the evolution of the eigenvalues of the PDF covariance matrix still hold true, but, in this case, there are more eigenvalues that are clearly converging or have converged. This kind of behaviour is, after all, something that we expect. The block-structure of the correlation is more defined and simpler, therefore we expect to need less eigenvalues to describes a simpler structure and for these eigenvalues to converge faster for the structure to emerge as clear as in the matrix of the previous section.

7.3 Dependence of the eigenvectors of the m_W covariance matrix on m_W

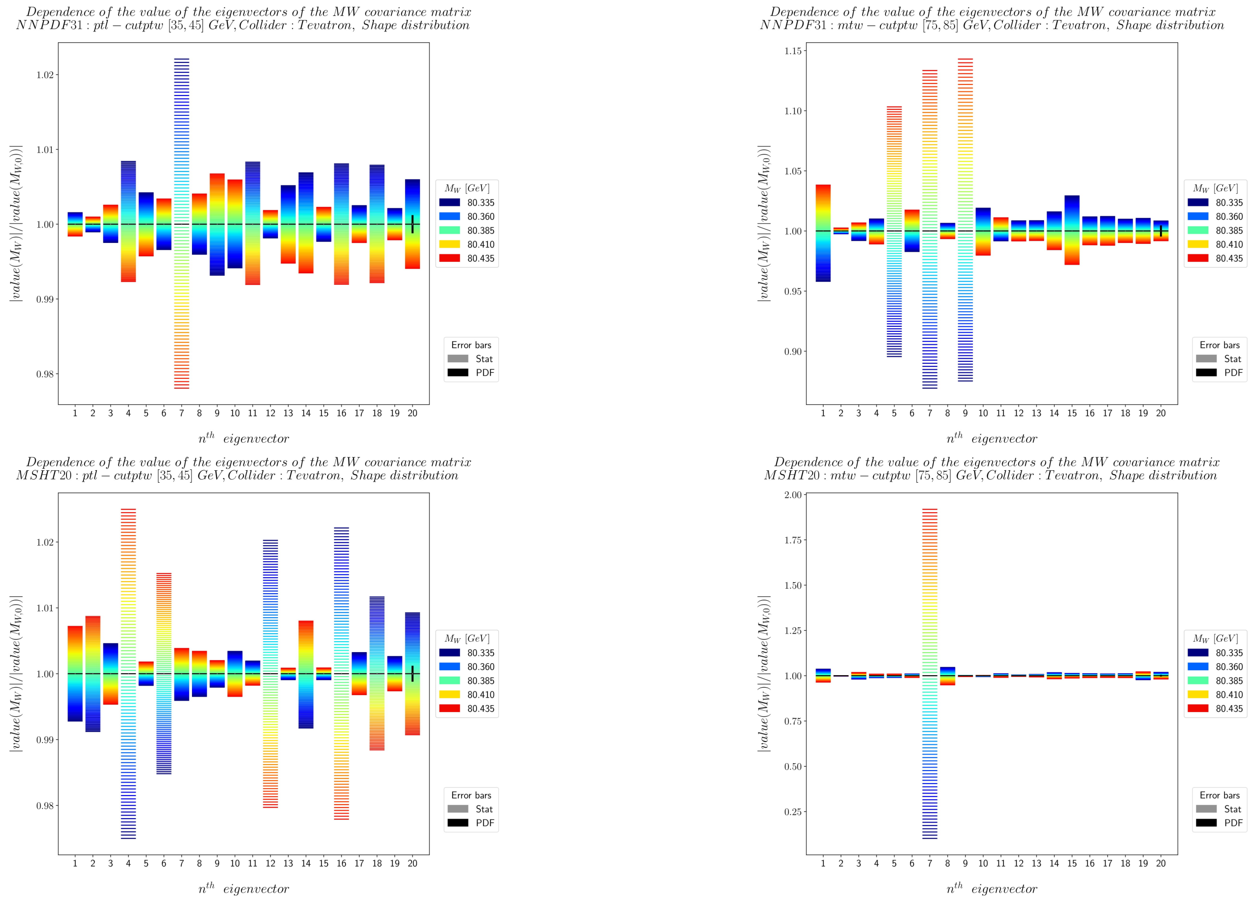

Figure 11: Dependence of the eigenvectors of the m_W covariance matrix on m_W. We considered the shape distributions with respect to the transverse mass of the W boson and the transverse momentum of the charged-lepton, both at the Tevatron with cut on transverse momentum of the W boson. The distributions have been normalized on the ranges $[75, 85]$ GeV and $[35, 45]$ GeV respectively. The distributions have been generated in the mass range $[80.335, 80.435]$ GeV and bins with length of 0.5 GeV.

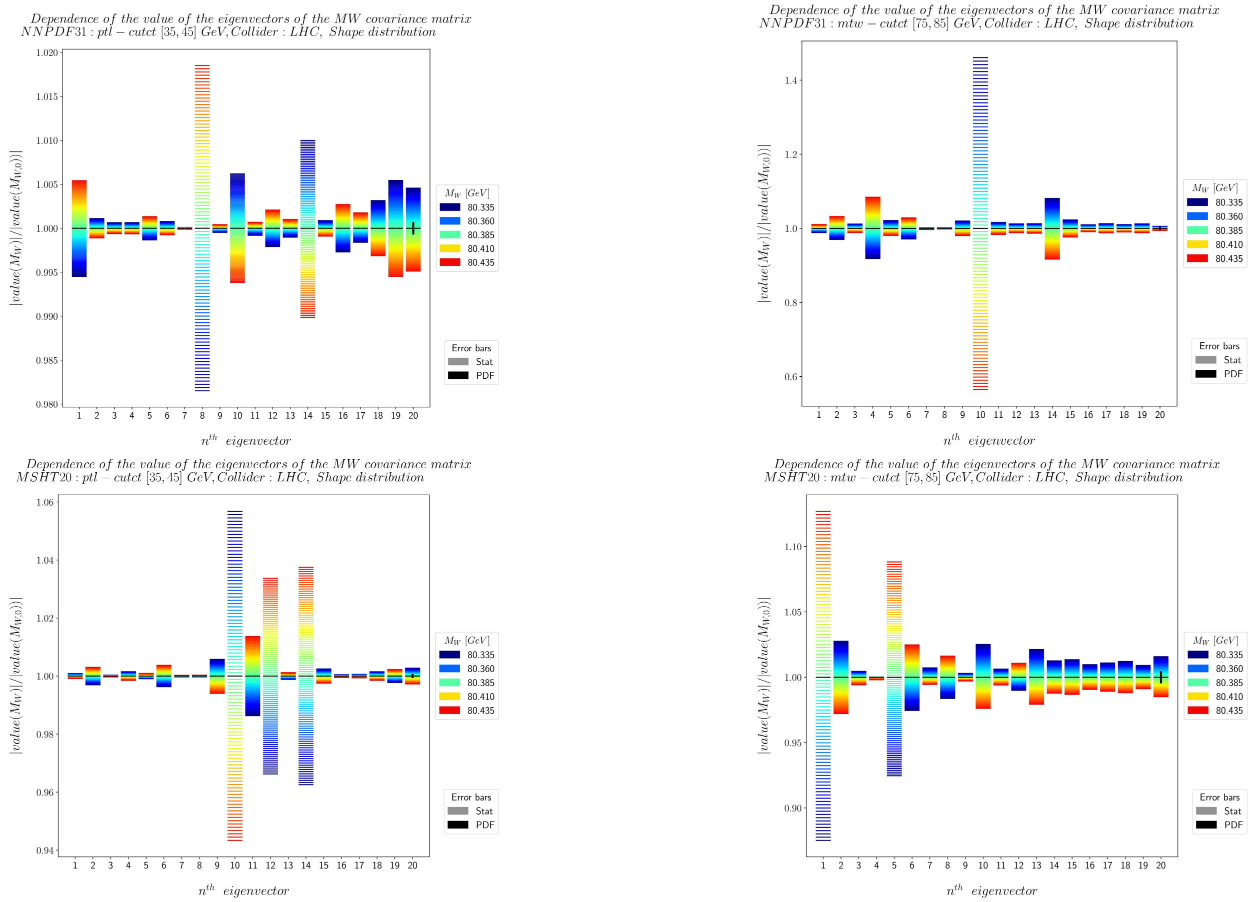

Figure 12: Dependence of the eigenvectors of the m_W covariance matrix on m_W. We considered the shape distributions with respect to the transverse mass of the W boson and the transverse momentum of the charged-lepton, both at the LHC with central cut. The distributions have been normalized on the ranges $[75, 85]$ GeV and $[35, 45]$ GeV respectively. The distributions have been generated in the mass range $[80.335, 80.435]$ GeV and bins with length of 0.5 GeV.

We also studied the eigenvectors of the m_W covariance matrix. In particular, we studied the dependence of its eigenvectors on m_W and compared their spread with its statistical error and its systematical error due to variations of m_W. The same comments that we did in the analogous section for the eigenvectors of the PDF covariance matrix still hold true, though in the present case, some error bands are more noticeable. On the whole, the PDF and m_W exhibits similar behaviours that we still struggle to fully understand.

8 Conclusions

All in all, our work raised more questions than it was able to answer. Nonetheless, we hope that we have at least paved a bit the way for a future study that may be a proper continuation of Ref.[3] and that will help to elucidate the physics of the W boson via validating or not the part of Ref.[10] that has to do with PDF uncertainties.

References

[1] M. Aaboud, G. Aad, B. Abbott, J. Abdallah, O. Abdinov, B. Abeloos, S. H. Abidi, O. S. AbouZeid, N. L. Abraham, and et al. Measurement of the w-boson mass in pp collisions at $\sqrt{s} = 7\ TeV$ with the atlas detector. *The European Physical Journal C*, 78(2), Feb 2018.

[2] T. Aoyama, N. Asmussen, M. Benayoun, J. Bijnens, T. Blum, M. Bruno, I. Caprini, C.M. Carloni Calame, M. Cè, G. Colangelo, and et al. The anomalous magnetic moment of the muon in the standard model. *Physics Reports*, 887:1–166, Dec 2020.

[3] Emanuele Bagnaschi and Alessandro Vicini. Parton density uncertainties and the determination of electroweak parameters at hadron colliders. *Phys. Rev. Lett.*, 126:041801, Jan 2021.

[4] Luca Barze, Guido Montagna, Paolo Nason, Oreste Nicrosini, and Fulvio Piccinini. Implementation of electroweak corrections in the POWHEG BOX: single W production. *JHEP*, 04:037, 2012.

[5] Sz. Borsanyi, Z. Fodor, J. N. Guenther, C. Hoelbling, S. D. Katz, L. Lellouch, T. Lippert, K. Miura, L. Parato, K. K. Szabo, F. Stokes, B. C. Toth, Cs. Torok, and L. Varnhorst. Leading hadronic contribution to the muon magnetic moment from lattice qcd. *Nature*, 593(7857):51–55, May 2021.

[6] G. Bozzi, L. Citelli, and A. Vicini. Parton density function uncertainties on the w boson mass measurement from the lepton transverse momentum distribution. *Phys. Rev. D*, 91:113005, Jun 2015.

[7] Luc Demortier. Equivalence of the best-fit and covariance-matrix methods for comparing binned data with a model in the presence of correlated systematic uncertainties, Apr 1999.

[8] Sidney D. Drell and Tung-Mow Yan. Massive lepton-pair production in hadron-hadron collisions at high energies. *Phys. Rev. Lett.*, 25:316–320, Aug 1970.

[9] Jens Erler and Matthias Schott. Electroweak precision tests of the standard model after the discovery of the higgs boson. *Progress in Particle and Nuclear Physics*, 106:68–119, May 2019.

[10] T. Aaltonen *et al.* High-precision measurement of the W boson mass with the cdf ii detector. *Science*, 376(6589):170–176, 2022.

[11] Richard P. Feynman. *The Behavior of Hadron Collisions at Extreme Energies*, pages 289–304. Springer Netherlands, Dordrecht, 1988.

[12] M.E. Peskin and D.V. Schroeder. *An Introduction To Quantum Field Theory*. Frontiers in Physics. Avalon Publishing, 1995.

[13] J. Pumplin, D. Stump, R. Brock, D. Casey, J. Huston, J. Kalk, H. L. Lai, and W. K. Tung. Uncertainties of predictions from parton distribution functions. ii. the hessian method. *Phys. Rev. D*, 65:014013, Dec 2001.

[14] Torbjörn Sjöstrand, Stefan Ask, Jesper R. Christiansen, Richard Corke, Nishita Desai, Philip Ilten, Stephen Mrenna, Stefan Prestel, Christine O. Rasmussen, and Peter Z. Skands. An introduction to pythia 8.2. *Computer Physics Communications*, 191:159–177, Jun 2015.

[15] D. Stump, J. Pumplin, R. Brock, D. Casey, J. Huston, J. Kalk, H. L. Lai, and W. K. Tung. Uncertainties of predictions from parton distribution functions. i. the lagrange multiplier method. *Phys. Rev. D*, 65:014012, Dec 2001.