



**UNIVERSITÀ DEGLI STUDI DI MILANO**

---

**FACOLTÀ DI SCIENZE E TECNOLOGIE**

Corso di Laurea Triennale in Fisica

**IMPACT OF SYSTEMATIC  
UNCERTAINTIES ON PDF  
DETERMINATION**

Relatore:

**Stefano Forte**

Correlatore:

**Stefano Carrazza**

Candidato:

**Francesco Saverio Dambrosio**

Matricola n. 829003

Codice PACS 12.38.-t

---

**Anno Accademico 2015-2016**

# Summary

The purpose of this thesis is to study the dependence of Parton Distribution Functions (PDFs) determination on the systematic uncertainties of experimental data. In fact, PDFs cannot be determined from first principles, but need a fitting procedure of experimental data in order to be computed.

In Chapter 1 we explain what is a PDF, how it can be used for QCD computations, such as cross sections, and how PDFs at different scales can be related. Then we briefly recall the recent history of PDFs determination.

In Chapter 2 we present the main procedures for the determination of a PDF set and the relative uncertainties, with particular attention to the Hessian method [9] and the Monte Carlo method [1]. While the first is often used in fitting procedures based on standard parametrization of the PDFs, the second is preferred when one has to deal with a large number of parameters in the fit. For example, in this work we use a neural network parametrization, for which a Monte Carlo method to determine PDFs uncertainties is mandatory. We also provide a series of criteria in order to evaluate the goodness of a fitting procedure.

In Chapter 3 we present the results we got in the developed fit simulations. The basic idea of the work is the following: we produce a reference fit with experimental datasets taken from charged lepton beam experiments and neutrino beam experiments; then, we introduce an inconsistency into these datasets, by rescaling the systematic uncertainties of the data by hand, and we fit again the datasets, comparing the new results with the reference by means of suitable criteria.

# Contents

<b>1</b>	<b>Parton Distribution Functions</b>	<b>3</b>
1.1	The main characters of Quantum Chromodynamics . . . . .	3
1.2	Factorization in QCD . . . . .	4
1.3	Definition of a Parton Distribution Function . . . . .	5
1.4	Progress in PDF determination . . . . .	7
<b>2</b>	<b>PDF fitting methodologies</b>	<b>9</b>
2.1	Hessian approach . . . . .	10
2.1.1	Uncertainties with Hessian approach . . . . .	10
2.1.2	The tolerance problem . . . . .	11
2.2	Monte Carlo approach . . . . .	12
2.2.1	Neural network parametrization . . . . .	12
2.2.2	Uncertainties with Monte Carlo approach . . . . .	13
2.3	Closure testing . . . . .	15
2.3.1	Level 0 closure test . . . . .	16
2.3.2	Level 1 closure test . . . . .	16
2.3.3	Level 2 closure test . . . . .	17
2.3.4	PDF uncertainties in closure tests . . . . .	17
2.3.5	Quantitative criteria for closure test validation . . . . .	18
<b>3</b>	<b>Simulations and results</b>	<b>21</b>
3.1	Methodology . . . . .	21
3.2	Input PDF set and datasets used in closure tests . . . . .	23
3.3	Reference closure test . . . . .	25
3.4	Rescaling of additive systematic uncertainties in closure test fluctuations . . . . .	27
3.5	Rescaling of multiplicative systematic uncertainties in closure test fluctuations . . . . .	30
3.6	Rescaling of additive systematic uncertainties in covariance matrix . . . . .	33
3.7	Final conclusions . . . . .	35
	<b>Bibliography</b>	<b>37</b>

# Chapter 1

## Parton Distribution Functions

Parton Distribution Functions (PDFs) are a powerful ingredient to study and understand the physics of processes at hadrons colliders, such as LHC. For instance, a PDF set allows one to calculate the cross section of scattering processes between hadrons, and make other predictions about strong interactions physics. In this chapter we briefly describe what is a PDF, what it can be used for, and how a PDF set can be computed, which is still an open issue.

### 1.1 The main characters of Quantum Chromodynamics

The substructure of hadrons, and in particular of nucleons, can be expressed in terms of the basic fields, or degrees of freedom of the theory that describes the strong interactions between them, i.e. quantum chromodynamics (QCD). These fields are classified into 6 different flavour quarks (each of them coming with its anti-flavour partner), and a gluon, the gauge boson which carries out the interaction between quarks; quarks and gluon represent the fundamental constituents of the nucleon. Quarks feature an internal degree of freedom, the colour, which is the analogous of the electric charge but, differently from this (which can be only positive or negative), the colour charge can have three different states, identified with three colours blue, red and green. In this picture, all known particles have neutral colour charge: for example, protons and neutrons (and in general baryons) have three quarks of different colours (blue, red and green), while mesons have two quarks, one of one colour and the other of the relative anti-colour. The gluon carries out the interaction between coloured quarks, so it shows two coloured components, one colour and one anti-colour; in this way, there exist 8 independent gluons, which can be combined to give different basis. The next tables report the family of the quarks and a possible basis for the gluons.

Flavour	Symbol	Mass (MeV/c <sup>2</sup> )	Anti-flavour	Symbol
Up	$u$	1.7 – 3.3	Antiup	$\bar{u}$
Down	$d$	4.1 – 5.8	Antidown	$\bar{d}$
Strange	$s$	$101^{+29}_{-21}$	Antistrange	$\bar{s}$
Charm	$c$	$1270^{+70}_{-90}$	Anticharm	$\bar{c}$
Bottom	$b$	$4190^{+180}_{-60}$	Antibottom	$\bar{b}$
Top	$t$	$(172 \pm 1.3) \cdot 10^3$	Antitop	$\bar{t}$

Gluon basis	$r\bar{g}$	$r\bar{b}$	$g\bar{b}$	$g\bar{r}$	$b\bar{r}$	$b\bar{g}$	$\frac{r\bar{r}-g\bar{g}}{\sqrt{2}}$	$\frac{r\bar{r}-g\bar{g}-2b\bar{b}}{\sqrt{6}}$
-------------	------------	------------	------------	------------	------------	------------	--------------------------------------	--

## 1.2 Factorization in QCD

In this section we present a basic property holding in QCD, i.e. the factorization, which will lead us to the definition of a PDF.

A fundamental property of QCD is factorization, which separates the cross section of a process into two parts: a process-dependent parton cross section, and a set of universal functions which are convolved with parton cross sections in order to obtain the cross section of a process. For instance, factorization for hadroproduction processes allows to express the cross section of a process at a scale  $M_X^2$  as

$$\begin{aligned}\sigma_X(s, M_X^2) &= \sum_{a,b} \int_{x_{min}}^1 dx_1 dx_2 f_{a/h_1}(x_1, M_X^2) f_{b/h_2}(x_2, M_X^2) \hat{\sigma}_{ab \rightarrow X}(x_1 x_2 s, M_X^2) \\ &= \sigma_{ab}^0 \sum_{a,b} \int_{\tau}^1 \frac{dx_1}{x_1} \int_{\tau/x_1}^1 \frac{dx_2}{x_2} f_{a/h_1}(x_1, M_X^2) f_{b/h_2}(x_2, M_X^2) C\left(\frac{\tau}{x_1 x_2}, \alpha_S(M_X^2)\right),\end{aligned}\tag{1.2.1}$$

where  $f_{a/h_i}(x_i)$  is the distribution of partons of type  $a$  in the  $i$ -th coming hadron,  $\hat{\sigma}_{q_b q_b \rightarrow X}$  is the parton-level cross section for the production of the final state  $X$ , and  $x_{min} = \tau := \frac{M_X^2}{s}$ . In the next step, the partonic cross section is replaced by  $\sigma_0 C(\tau, \alpha_S(M_X^2))$  where the hard coefficient function  $C(\tau, \alpha_S(M_X^2))$  is defined by viewing  $\hat{\sigma}_{q_b q_b \rightarrow X}$  as a function of the hard scale  $M_X^2$  and the dimensionless ratio of this scale to the centre-of-mass energy  $\hat{s}$  of the partonic subprocess in terms of the scaling variable  $\frac{M_X^2}{\hat{s}} = \frac{\tau}{x_1 x_2}$ .

Another example of factorization the standard factorization for the deep-inelastic structure functions  $F_i(x, Q^2)$ :

$$F_i(x, Q^2) = x \sum_a \int_x^1 \frac{dz}{z} C_{i,a}\left(\frac{x}{z}, \alpha_S(Q^2)\right) f_a(z, Q^2),\tag{1.2.2}$$

where  $x = \frac{Q^2}{2p \cdot q}$  is the standard Bjorken variable,  $C_{i,a}$  is the structure function computed with an incoming parton, and  $f_a(z, Q^2)$  is the distribution of the parton  $a$  in the only incoming hadron. Structure functions are then used in factorization for electroproduction to parametrize the inclusive deep-inelastic scattering cross section:

$$\frac{d^2 \sigma^{NC, \ell^\pm}}{dx dQ^2}(x, y, Q^2) = \frac{2\pi \alpha^2}{x Q^4} [Y_+ F_2^{NC}(x, Q^2) \mp Y_- x F_3^{NC}(x, Q^2) - y^2 F_L^{NC}(x, Q^2)],\tag{1.2.3}$$

for neutral-current charged lepton  $\ell^\pm$  DIS, where the longitudinal structure function is defined as

$$F_L(x, Q^2) := F_2(x, Q^2) - 2x F_1(x, Q^2),\tag{1.2.4}$$

and

$$Y_\pm := 1 \pm (1 - y)^2 \quad y := \frac{p \cdot q}{p \cdot k} = \frac{Q^2}{xs},\tag{1.2.5}$$

where  $p$  and  $k$  are respectively the incoming proton and lepton momenta,  $q$  is the virtual photon momentum ( $q^2 = Q^2$ ) and  $s$  is the centre-of-mass energy of the lepton-proton collision (last step holds neglecting the proton mass). Similar expressions hold for charged-current charged and neutral lepton scattering.

So in all these expressions we observe the presence of parton distribution functions, which actually have big importance in this framework: we will describe them in detail in the next section.

### 1.3 Definition of a Parton Distribution Function

In this section we define a parton distribution function, we explain how to relate PDFs at different energy scales, and we give some constraints which PDFs have to satisfy.

At the leading order of QCD theory, we can say that the Parton Distribution Function  $f_i(x, Q^2)$  is the probability density of finding in the proton a parton flavour  $i$  (quarks or gluon) carrying a fraction  $x$  of the proton momentum, where  $Q$  is the energy scale of the hard interaction. This definition is not well-posed at every order of QCD theory, in the sense that a PDF is not a probability distribution as much as, say, the square module of the wave function of a state; in fact, a PDF can also be negative, while a probability distribution must be positive defined. However, the definition makes sense when we consider PDFs not insulated, but convolved with partonic cross sections in factorized expressions such as eq.(1.2.1) and eq.(1.2.2). By this definition, it's clear that a PDF set includes 13 functions (6 quarks, 6 anti-quarks and a gluon), each of them defined between  $x = 0$  and  $x = 1$ . Fig. 1.1 shows a typical PDF set, including some parton flavours at two different scale energies.

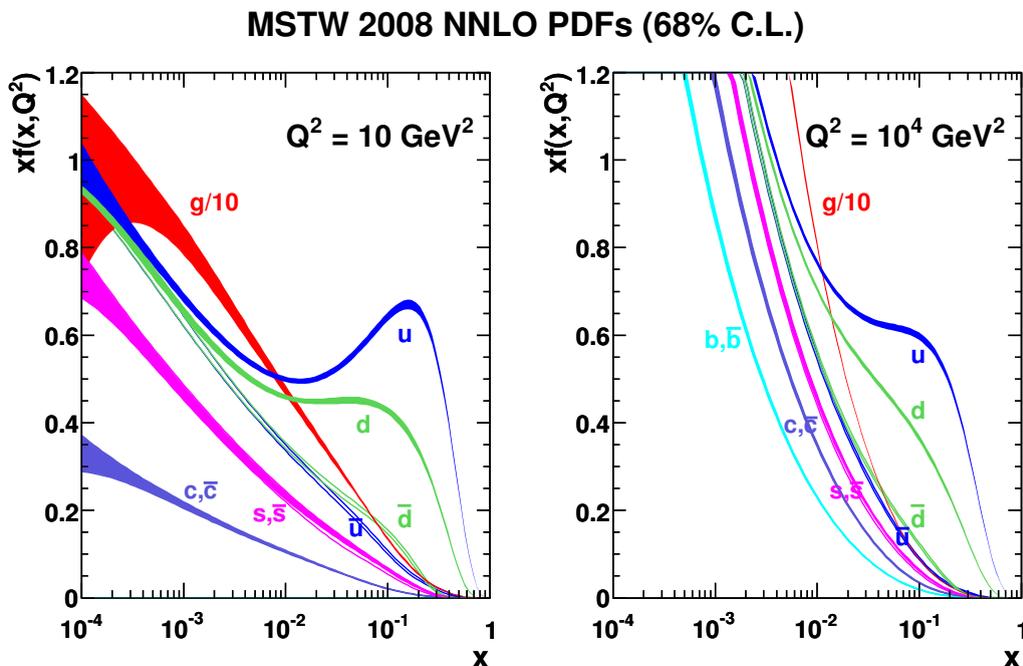


Fig. 1.1: MSTW 2008 NLO PDF set (68% C.L.) at  $Q^2 = 10 \text{ GeV}^2$  and  $Q^2 = 10 \text{ GeV}^4$ .

An important feature of PDFs is that they are universal, in the sense that they are process-independent; so, if we have a PDF set, we can calculate expressions such as eq.(1.2.1) for every process, provided we know the parton cross section for that specific process (partonic cross

sections are computed in perturbation theory using the quark and gluon degrees of freedom of the QCD Lagrangian).

An interesting fact of PDFs is that, computing the fraction of momentum carried by quarks, one finds

$$\frac{p_{quark}}{p_{proton}} = \int_0^1 dx \sum_{i \in quark} x f_i(x, Q^2) \approx \frac{1}{2}, \quad (1.3.1)$$

so quarks carry only half of the total momentum of the proton, with the rest being carried by the gluon. This expectation is confirmed by experimental observations in deep inelastic scattering (DIS) experiments with hadrons, and it is a good evidence of the existence of gluon.

The dependence of PDFs on the scale of the process can be evaluated with a set of perturbative evolution equations, which in fact are integro-differential equations, known as the DGLAP equations; a way to write them is the following [2]:

$$\frac{\partial}{\partial \ln Q^2} \begin{pmatrix} \Sigma(x, Q^2) \\ g(x, Q^2) \end{pmatrix} = \int_x^1 \frac{dy}{y} \begin{pmatrix} P_{qq}^S \left( \frac{x}{y}, \alpha_S(Q^2) \right) & 2n_f P_{qg}^S \left( \frac{x}{y}, \alpha_S(Q^2) \right) \\ P_{gq}^S \left( \frac{x}{y}, \alpha_S(Q^2) \right) & P_{gg}^S \left( \frac{x}{y}, \alpha_S(Q^2) \right) \end{pmatrix} \begin{pmatrix} \Sigma(y, Q^2) \\ g(y, Q^2) \end{pmatrix}, \quad (1.3.2)$$

$$\frac{\partial}{\partial \ln Q^2} q_{ij}^{NS}(x, Q^2) = \int_x^1 \frac{dy}{y} P_{ij}^{NS} \left( \frac{x}{y}, \alpha_S(Q^2) \right) q_{ij}^{NS}(y, Q^2), \quad (1.3.3)$$

where  $g$  is the gluon and  $\Sigma$  denotes the singlet quark distribution defined as

$$\Sigma(x, Q^2) := \sum_{i=1}^{n_f} [q_i(x, Q^2) + \bar{q}_i(x, Q^2)], \quad (1.3.4)$$

while the nonsinglet quark distributions are defined as any linearly independent set of  $2n_f - 1$  differences of quark and antiquark distribution

$$q_{ij}^{NS}(x, Q^2) := q_i(x, Q^2) - \bar{q}_j(x, Q^2). \quad (1.3.5)$$

The splitting functions  $P_{ab}$  describe the  $b \rightarrow a$  parton splitting probability, and they are perturbative series in the strong coupling constant  $\alpha_S$ , starting at order  $\alpha_S$  at leading-order computations:

$$P_{ab}(\alpha_S, z) = P_{ab}^{LO}(z) + \alpha_S P_{ab}^{NLO}(z) + \alpha_S^2 P_{ab}^{NNLO}(z) + \dots \quad (1.3.6)$$

In this way, all PDF sets are parametrized at a fixed reference scale  $Q_0^2$ , and the solution to the evolution equations provides tables of PDFs as a function of  $x$  and  $Q^2$ .

Finally, it must be said that typical conservation laws give some constraints on the distribution functions: for instance, the conservation of baryon number ( $n_u = 2$ ,  $n_d = 1$ ,  $n_{s,c,b,t} = 0$ ) implies

$$\int_0^1 dx (q_i(x, Q^2) - \bar{q}_i(x, Q^2)) = n_i, \quad (1.3.7)$$

while the conservation of total energy-momentum implies

$$\int_0^1 dx x \left[ \sum_{i=1}^{n_f} (q_i(x, Q^2) + \bar{q}_i(x, Q^2)) + g(x, Q^2) \right] = 1. \quad (1.3.8)$$

## 1.4 Progress in PDF determination

In this section we briefly recall the recent history of the PDF determination.

At the current stage of knowledge of strong interactions, PDFs cannot be computed from first principles (this would be equivalent to knowing the whole wave function of the proton), so they are determined by comparing the PDF-dependent predictions for one or more physical processes, such as cross sections, with their actual measured value, in a quite complex fitting procedure. Obviously, this procedure is non trivial because one has to fit a finite discrete ensemble of experimental data, and extract from that a set of functions, which are defined everywhere in  $[0, 1]$  (that is, in principle, a mathematically ill-posed problem). This procedure to determine a PDF set involves that every function comes out with an uncertainty band that arises from three sources: the underlying data to fit (which are affected by statistical and systematic errors), the theory which is used to describe them (which is typically based on the truncation of a perturbative expansion) and the procedure which is used to extract the PDFs from the data; in next chapters we will focus on the first and the last of them.

The determination of PDFs has gone through various stages, depending both on theoretical and phenomenological understanding of QCD, and on experiments in DIS and “hard” proton-proton (or proton-antiproton) high-energy collisions, with ever increasing precision. After some early attempts to determine a semi-quantitative set of PDFs [4], based on model assumptions and first results on DIS, the determination of the gluon distribution in 1982 [5] allowed to produce the first PDF set [6], based on the data-theory comparison for a set of different lepton-hadron and hadron-hadron scattering processes. These analyses were all performed at leading order (LO), i.e. using, in QCD calculations, the lowest perturbative order of the expansion in the strong coupling constant  $\alpha_S(Q^2)$ , which was accurate enough for those sets of data.

In the next decade high-precision deep-inelastic scattering and hadron collider data became available and, thanks in particular to DIS data from HERA collider, determination of PDF sets including NLO theory, which at the time was well understood, was possible; so, combination of more accurate experimental data and full knowledge of theory led to global parton sets showing agreement with NLO QCD computations, such as CTEQ5 [7] and MRST2001 [8]. Despite some differences in technical details, these PDF sets are based on similar determination approaches: a functional form for PDFs is assumed, parametrized by a relatively small number of parameters, that are determined by optimizing the fit of the computed observables to the experimental data.

Soon, as precision in experimental data and theoretical QCD computations increased, an estimate of the uncertainty in PDF sets became needed. A first attempt to obtain a PDF set with uncertainties was made in the late 90’s fitting to a subset of DIS experiments but retaining all the information on the correlated uncertainties in the underlying data and propagating it through the fitting procedure. However, the need of a more systematic and consistent approach to determine PDF uncertainties was recognized.

As we will explain in more details in the next chapter, in the 2000s two main ways of determining PDF uncertainties developed. The first approach is based on a conventional methodology which uses a least squares style method in order to determine “error” PDF sets along with the central best-fit, with the determination of a  $1-\sigma$  contour in parameter space about the best fit. This method is used by two groups that have produced widely used PDF sets, MSTW [9] and CTEQ [10].

---

A second approach was proposed in [11] and produced, by the NNDFP collaboration, a first PDF set based on DIS data [12] and then a PDF set from a global fit [13]. This approach differs in two main aspects from the previous one: first, PDFs are represented as a Monte Carlo sample, from which the central value and uncertainty can be computed respectively as a mean and standard deviation, rather than from a best-fit and error sets; second, the functional form used for the PDF parametrization, based on neural networks, has a very large number of parameters (more than 250, to be compared to about 30 for sets based on a standard parametrization). More details, also regarding the determination of uncertainties, will be given later.

Finally, in the late 2000s the increasing availability of calculations to next-to-next-to-leading order (NNLO) in QCD led the various groups to produce PDF sets using NNLO QCD theory in their determination, and also to include heavy-quark mass effects, with all these sets available for a variety of values of the strong coupling. All the sets cited above, MSTW, CTEQ and NNPDF, include these features.

## Chapter 2

# PDF fitting methodologies

In this chapter we describe two PDF fitting methodologies and how they deal with the problem of determining PDF uncertainties; then, we'll illustrate a method for the validation of a fitting procedure.

As we said in previous chapter, PDFs are not computable from first principles because that would involve the knowledge of the proton wave function. So, in order to determine a PDF set, a fitting procedure must be adopted, and it consists of comparing factorized expressions such as eqs.(1.2.1)–(1.2.2) with experimental data, while minimizing a suitable measure of goodness-of-fit.

Also, we said that this is not a well-posed problem, because we have to solve an infinite-dimensional problem (determining a set of functions<sup>1</sup>) with a finite set of experimental data. The task is simplified by the observation that parton distributions are expected to be smooth functions of the variable  $x \in [0, 1]$ , so they can be represented in terms of a finite basis of functions or, which is equivalent, with a finite set of parameters. The problem is then reduced to the choice of an optimal parametrization that, for given accuracy, minimizes the number of parameters without introducing a bias; given a parametrization of PDFs, fitting procedure involves computing a number of physical processes with them, and minimizing a suitable figure of merit, such as a  $\chi^2$  or likelihood function in the parameters' space in order to determine a best-fit set of PDFs. Obviously, this procedure can be developed at a single reference scale  $Q_0^2$ , because PDFs at other scales are related to these ones by the evolution equations (1.3.2)–(1.3.3).

There exist various strategies for parametrization and fitting of a PDF set, but typical choices are a standard parametrization (inspired by various QCD arguments), with a relatively small number of parameters, or a parametrization based on a large number of parameters, such as a neural network parametrization. Having chosen a parametrization, different approaches are possible in order to determine PDF uncertainties: typically, with standard parametrization a “Hessian” approach is adopted, in which the best-fit result is given in the form of an optimal set of parameters and an error matrix centered on this optimal fit to compute uncertainties; with neural network parametrization a Monte Carlo approach is preferred, in which the best fit is determined from the Monte Carlo sample by averaging and uncertainties are obtained

---

<sup>1</sup>In principle thirteen functions: six quarks, six antiquarks and a gluon. However, in practice, charm and heavier quark PDFs are not independently determined in all current PDF sets, and are instead assumed only to be generated by QCD radiation, so the standard for current precision studies is to have a set of seven independent PDFs.

as variances of the sample. However, these choices are not exclusive, for example one could use a parametrization based on neural networks and a Hessian approach for determination of uncertainties.

## 2.1 Hessian approach

In this section we describe a fitting methodology based on standard parametrization, in which the form of PDFs is assumed to be polynomial. Then, we will present the Hessian approach for the determination of PDF uncertainties, which is often combined to standard parametrizations.

Once a suitable basis of PDFs has been chosen (we can choose a flavour basis, or the singlet combination eq.(1.3.4)), we must adopt a parametrization of PDFs at a reference scale  $Q_0^2$ ; a standard choice is the following:

$$f_i(x, Q_0^2) = x^{\alpha_i} (1-x)^{\beta_i} g_i(x), \quad (2.1.1)$$

where  $g_i(x)$  tends to a constant for both  $x \rightarrow 0$  and  $x \rightarrow 1$  and usually is a polynomial or the exponential of a polynomial in  $x$  or  $\sqrt{x}$ . This choice is motivated by the expectation that PDFs behave as a power of  $x$  as  $x \rightarrow 0$  due to Regge theory, and as a power of  $(1-x)$  as  $x \rightarrow 1$  due to quark counting rules (see [14] for details). Typical contemporary PDF sets based on this choice of functional form are parametrized by about 20–30 parameters.

Once adopted a parametrization of PDFs, the task becomes determining best fit values and uncertainty ranges for the parameters and this is done, in Hessian approach, by minimizing a figure of merit such as

$$\chi^2(\vec{a}) = \frac{1}{N_{\text{dat}}} \sum_{i,j=1}^{N_{\text{dat}}} (d_i - \bar{d}_i(\vec{a})) \text{cov}_{ij} (d_j - \bar{d}_j(\vec{a})), \quad (2.1.2)$$

where  $d_i$  are experimental data with experimental covariance matrix  $\text{cov}_{ij}$  (including all correlated and uncorrelated statistical and systematic uncertainties);  $\bar{d}_i(\vec{a})$  are theoretical predictions, obtained by evolving the starting PDFs at any scale  $Q^2$  with the evolution equations (1.3.2)–(1.3.3) and folding the result with known partonic cross sections using factorized expressions eqs.(1.2.1)–(1.2.2); the vector  $\vec{a}$  denotes the full set of parameters on which the PDFs at scale  $Q_0^2$  depend, and it lies in a  $k$ -dimensional space, where  $k$  is the number of parameters. So  $\chi^2$  is a function of the parameters  $\vec{a}$  in which the dependence is expressed by the predictions  $\bar{d}_i(\vec{a})$ . Hence, the best fit set of parameters can be identified with the absolute minimum of the  $\chi^2$  in parameter space.

### 2.1.1 Uncertainties with Hessian approach

Once the best-fit has been determined, a confidence level (C.L.) about it is determined by expanding the  $\chi^2$  in parameter space about its minimum to lowest nontrivial order. The desired confidence level is obtained as the volume in parameter space about the minimum that corresponds to a fixed increase of the  $\chi^2$ . For Gaussian uncertainties, the 68%, i.e. 1- $\sigma$ , confidence level corresponds to the volume enclosed by the  $\chi^2 = \chi_{\text{min}}^2 + 1$  surface. Obviously, the first order expansion of  $\chi^2$  is zero (we expand about a minimum), so the lowest nontrivial order is the second, and hence the confidence level is determined by the covariance matrix in parameter

space, which actually is the inverse of the matrix of second derivatives (from which Hessian name) of  $\chi^2$  with respect to the parameters, evaluated at the minimum, i.e.

$$\sigma_{ij}^{-1} = \partial_i \partial_j \chi^2 |_{\min}. \quad (2.1.3)$$

An advantage of the Hessian method is that it provides a simple tool for computing PDF uncertainties, which consists in finding eigenvectors of the Hessian matrix and rescale them by their respective eigenvalues; in other words, uncertainties on PDFs, or any observables that depend on them, are found by adding in quadrature the variation along the direction of each eigenvector. So in a Hessian approach one delivers a central set of PDFs  $S_0$ , and  $N_{\text{par}}$  1- $\sigma$  error sets  $S_i$ , corresponding to the variation of each eigenvector in turn. The best-fit value of any quantity  $F(S)$  which depends on the PDF set (such as a cross section, or a PDF itself), and its 1- $\sigma$  uncertainty, are respectively:

$$F_0 = F(S_0), \quad \sigma_F = \sqrt{\sum_{i=1}^{N_{\text{par}}} [F(S_i) - F(S_0)]^2}. \quad (2.1.4)$$

The disadvantage of this procedure is that the computation of the Hessian matrix and its diagonalization rapidly become unmanageable if the number of parameters is too large.

### 2.1.2 The tolerance problem

In practice, in actual PDF fits involving large numbers of experimental data points from different experiments, it turns out that the criterion of varying  $\chi^2$  by  $\Delta\chi^2 = 1$  in order to determine the 1- $\sigma$  contour leads to unrealistic results. For example, in [15] it was shown, by comparing the parameter values that provide the best fit to each set of experimental data, that these best-fit values fluctuate much more than one would expect if  $\Delta\chi^2 = 1$  did actually provide a 68% confidence level in parameter space. This behaviour emerges from the fact that if the fit involves a large number of experiments, the best-fit is not simultaneously a best-fit for individual datasets. The idea to solve this problem is to determine the desired confidence level for the global fit, say 1- $\sigma$  contour, while simultaneously checking the degree of agreement of individual experiments with it: this is performed by introducing the concept of ‘‘tolerance’’.

Tolerance is computed as follows. First, the Hessian matrix is diagonalized. Then, each eigenvector is moved away from the minimum of the global fit in either direction, and a  $\chi^2$  is computed for each experiment. Then, for each experiment one determines both the position of the minimum of the  $\chi^2$  and the 1- $\sigma$  interval<sup>2</sup> (or another confidence level) about it. Finally, one takes the envelope of the error bands for individual experiments at the the 1- $\sigma$  interval. For example, at the 68% confidence level one determines the range of variation in parameter space along this eigenvector about the minimum such that the 1- $\sigma$  interval of each experiment overlaps with this range. This gives a tolerance interval for the given eigenvector, whose width can be measured in units of the variation of the  $\chi^2$  of the global fit. The tolerance, defined as

$$T^2 = \Delta\chi^2, \quad (2.1.5)$$

is the width of the envelope of the error bands for individual experiments at the desired confidence level, and it gives the uncertainty band (for instance 1- $\sigma$  for 68% confidence level) for each

<sup>2</sup>Corresponding to the  $\Delta\chi^2 = 1$  variation about the minimum.

eigenvector, i.e for each parameter. This procedure makes the uncertainty determination much more reliable, in the sense that if new experiments show a behaviour similar to experiments already included in the fit, then they are expected to fall within the  $1\text{-}\sigma$  band in 68% of cases; by comparing results of the different strategies, it has been found that this happen if the  $1\text{-}\sigma$  band is defined in this way, while if the  $1\text{-}\sigma$  band was defined on the basis of standard statistics then probability of measurements falling outside the band would be much higher. The problem with tolerance is that typical values of  $T$  are 5-10, and it seems quite unlikely that every source of experimental uncertainty is underestimated by such a large factor, so the real origin of tolerance is still unclear.

## 2.2 Monte Carlo approach

In this section we present a fitting procedure based on a neural network parametrization, which uses a large number of parameters. When a neural network parametrization is adopted, using a Hessian approach for uncertainty determination can be unmanageable: we will present an advantageous alternative to Hessian approach, the Monte Carlo approach.

The main difference between Hessian and Monte Carlo approach is in the way the uncertainties are propagated from parameters to observables. In a few words, in this approach the probability distribution of PDFs is given by assigning a Monte Carlo sample of a large number of PDF replicas. The advantage of Monte Carlo approach is that it makes the computation of uncertainties much easier when a “non standard” parametrization is adopted, for instance a parametrization less manageable with respect to eq.(2.1.1), or a parametrization with a very large number of parameters. As Monte Carlo approach has often been used together with neural networks as a parton parametrization (in particular in this work), we describe this choice of parametrization, and then come back to the problem of uncertainty determination.

### 2.2.1 Neural network parametrization

Neural networks provide another functional form with a very large number of parameters, which can fit any continuous function in the limit of infinite number of parameters; a finite-dimensional truncation of the neural network parametrization is adequate to fit a very wide class of functions (for instance, both periodic and non periodic) without the need to adjust the form of the parametrization to the desired problem. A neural network can be constructed by iterating recursively a response function on nodes arranged in layers which feed forward to the next layer, with the first (last) layer containing the input (output) variables. For exemple, if we choose  $g(x) = \frac{1}{1+e^{\theta-\beta x}}$  as response function, and adopt a 1-2-1 iterative architecture (one iteration at the first node, two at the second, one at the third), we get the following function:

$$f(x) = \frac{1}{1 + e^{\theta_1^{(3)} - \frac{\omega_{11}^{(2)}}{1+e^{\theta_1^{(2)} - x\omega_{11}^{(1)}}} - \frac{\omega_{12}^{(2)}}{1+e^{\theta_2^{(2)} - x\omega_{21}^{(1)}}}}}, \quad (2.2.1)$$

where  $\theta_n^{(i)}$  and  $\omega_{nm}^{(i)}$  are free parameters.

NNPDF group uses PDFs parametrized with 2-5-3-1 neural networks, with 37 free parameters, for six flavours, six antiflavours and the gluon, hence in total 259 free parameters, much more than in standard parametrization. An obvious advantage in using neural networks is that, as

they have a large number of parameters, they provide a much more flexible parametrization, that can handle different behaviours of the experimental data at different regions of the fit. But the large number of parameters is not the only advantage of neural networks: in fact, in principle we could use a standard parametrization with a very high degree polynomial, reaching in this way an equal number of parameters as in neural network parametrization; however, if we fit a dataset with, say, a 259<sup>th</sup> degree polynomial, typically we get a very oscillating function, which is not very sensible. Conversely, neural network parametrizations start with a very simple, smooth functional form, which gets progressively complicated as the fitting procedure goes on, without showing rapidly oscillating behaviour as high degree polynomials do.

The price to pay for having a large number of parameters is that if the best fit is determined as the absolute minimum of the  $\chi^2$  one may end up fitting data fluctuations, which is clearly not desirable. This problem is solved by using a cross-validation method, which works as follows. The data are randomly divided into two sets (“training” and “validation”); the  $\chi^2$  is then computed for both sets separately, but only the  $\chi^2$  of the training set is minimized. Initially both the training and validation  $\chi^2$  decrease, but at some point the training  $\chi^2$  keeps decreasing while the validation  $\chi^2$  starts increasing. The point at which this happens defines the best fit. Picking a different partition of the data into training and validation sets for each replica ensures that there is no information loss, though of course this is only true in the limit of a large number of replicas. This procedure is shown in Fig. 2.1, in which a cross-validated fit is compared with a non-cross-validated one: while in the former the minimization algorithm stops as the validation  $\chi^2$  starts increasing, determining the best fit, in the latter it goes on, reaching an “overlearning” point as data fluctuations are being fitted.

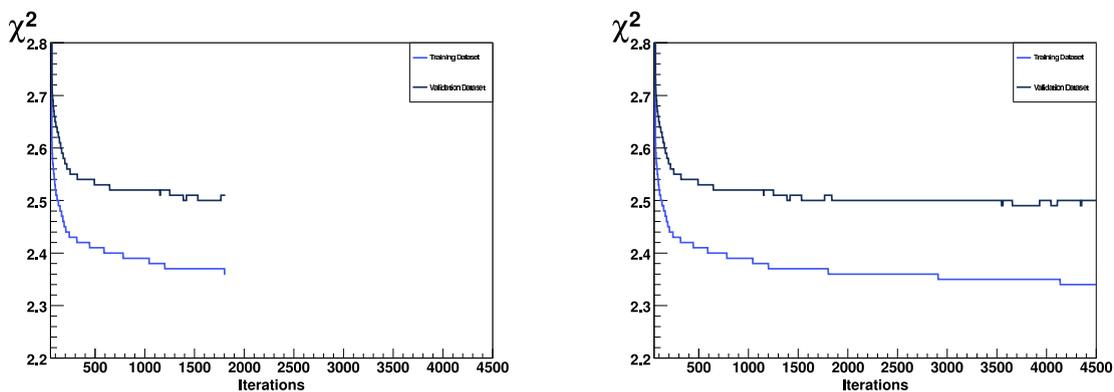


Fig. 2.1: Cross-validation method: on the left it is applied, on the right it is not. The  $\chi^2$  is shown as a function of the number of iterations of the minimization algorithm.

### 2.2.2 Uncertainties with Monte Carlo approach

Whereas in a Hessian approach parameters are assumed to be gaussianly distributed with covariance matrix  $\sigma_{ij}$  given by eq.(2.1.3), in Monte Carlo approach the probability distribution of PDFs is given by assigning a Monte Carlo sample of PDF replicas, namely  $N_{\text{rep}}$  PDF sets  $S^k$ . So, the best-fit value of any quantity  $F(S)$  which depends on the PDF set (such as a cross section, or a PDF itself) and its 1- $\sigma$  interval are now determined as the mean value and the

standard deviation over the replicas, respectively:

$$F_0 = \frac{1}{N_{\text{rep}}} \sum_{k=1}^{N_{\text{rep}}} F(S^k), \quad \sigma_F = \sqrt{\frac{1}{N_{\text{rep}} - 1} \sum_{k=1}^{N_{\text{rep}}} [F(S^k) - F_0]^2}. \quad (2.2.2)$$

The obvious advantage of the Monte Carlo method is that it does not require assumptions to be made on the form of the probability distribution in parameter space, and also that it provides a direct representation of the probability distribution.

A way of constructing a Monte Carlo sampling of parameter space is to construct a Monte Carlo representation of the original data. In this way, instead of having experimental data  $D_i$  with their covariance matrix, one constructs a set of  $N_{\text{rep}}$  data replicas  $D_i^k$  with  $i = 1, \dots, N_{\text{dat}}$  and  $k = 1, \dots, N_{\text{rep}}$ , that reproduce the probability distribution of the data, such as

$$\langle D_i \rangle := \frac{1}{N_{\text{rep}}} \sum_{k=1}^{N_{\text{rep}}} D_i^k, \quad (2.2.3)$$

$$\text{cov}_{ij} := \frac{1}{N_{\text{rep}} - 1} \sum_{i=1}^{N_{\text{rep}}} \sum_{j=1}^{N_{\text{rep}}} (D_i^k - \langle D_i \rangle) (D_j^k - \langle D_j \rangle) \quad (2.2.4)$$

tend respectively to the experimental data points and to the experimental covariance matrix in the limit of  $N_{\text{rep}}$  large enough. The Monte Carlo sample of PDFs is then determined by fitting a PDF set  $S^k$  to each data replica, which can be done by minimizing a suitable figure of merit; in this way the set of data replicas is mapped onto a set of PDF replicas. This procedure is very advantageous when the PDF parametrization uses a very large number of parameters, as in the neural network methodology, because in this case determination and diagonalization of the Hessian matrix may be impractical or impossible. On the other hand, when a large number of parameters is adopted, a cross-validation procedure is also implemented as above.

For a typical set of data used in a fit the number of replicas required turns out to be surprisingly small: Fig. 2.2 shows a scatter plot of the averages vs. central values and variances vs. standard deviations for the set of  $N_{\text{dat}} = 3372$  data points included in the NNPDF1.2 [16] parton fit, computed using  $N_{\text{rep}} = 10, 100, 1000$  Monte Carlo replicas. The deviations from experimental data are a few percent already for  $N_{\text{rep}} = 10$  as far as central values is concerned; however, increasing the number of replicas is useful for a more accurate determination of uncertainties.

Another way to deliver a Monte Carlo sampling is to start from a Hessian fit and construct the PDF replicas  $S^k$  by generating a multi-Gaussian distribution of parameter values, centered at the best fit and with width provided by the Hessian matrix itself:

$$F(S^k) = F(S_0) + \sum_{j=1}^{N_{\text{par}}} [F(S_j) - F(S_0)] R_j^k \quad (2.2.5)$$

where  $R_j^k$  is a random number taken from a Gaussian with mean zero and variance one, and  $S_0$  and  $S_j$  are the usual best-fit and eigenvector PDF sets. In this case, it is possible to verify a posteriori that constructing 50–100 Monte Carlo PDF replicas is enough to reproduce the original central value and Hessian covariance matrix.

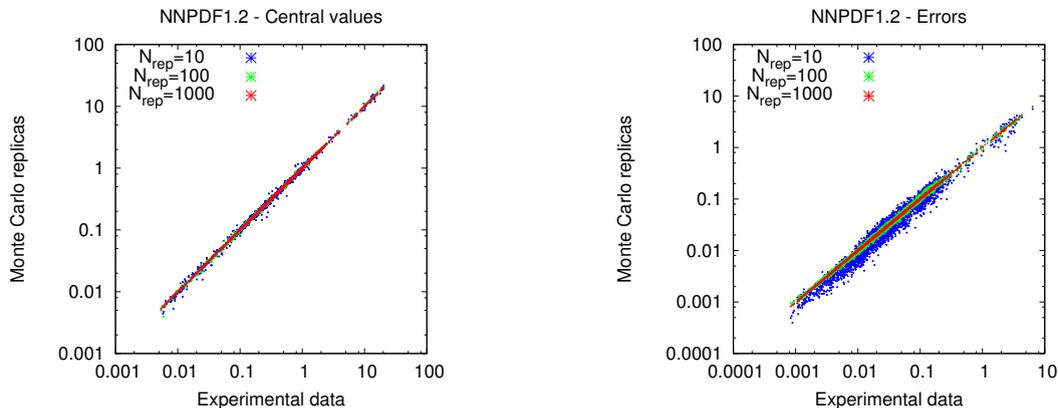


Fig. 2.2: Scatter plot of central values (left) and uncertainties (right) of a Monte Carlo sample compared to the data for the dataset of the NNPFD1.2 parton fit [16].

We can conclude this section with two reasons why it is useful to construct a Monte Carlo representation of a PDF set, even if it has been determined in a Hessian approach. The first reason is that if a Monte Carlo representation of a PDF set is available, new data can be added without performing a new fit, through Bayesian reweighting; in this case, a weight is associated to the original Monte Carlo replicas in order to deal with the effects of the new data. The second reason is that with a Monte Carlo representation of a PDF set it is possible to combine results obtained by different groups, especially if they arrive at independent PDF determinations using the same (or almost the same) data and theory, but different methodologies; an effective way of combining results, while keeping into account the possibility of methodological differences, and thus arriving at a more reliable result, is to simply produce a Monte Carlo set in which an equal (or conveniently weighted) fraction of replicas comes from each of the various groups. Furthermore, a Monte Carlo set can be transformed into a Hessian set, and viceversa: an advantage in having a Hessian set is that if one can diagonalize the Hessian matrix, every eigenvector represents an independent source of uncertainties, and this can be useful in experimental data analysis.

## 2.3 Closure testing

In this section we describe a procedure that is used in order to validate a fitting methodology, the closure testing, and we will consider it when applied to the Monte Carlo methodology described above. As we said in Sect.1.4, one of the sources of uncertainties in PDF determination lies in the methodology chosen to fit experimental data: the aim of the closure test is right to eliminate methodological uncertainty, so that the only uncertainties in PDFs are experimental and theoretical. The idea of the closure test is the following: we take a given assumed form for the PDFs (for example MSTW08 [9]), a given theoretical model (for example NLO perturbative QCD), and with them generate a set of global pseudo-data with known but realistic statistical properties determined with the covariance matrices of the real datasets. So these pseudo data have known statistical properties, no internal inconsistencies, and full agreement with the theoretical model used to produce them, therefore if we fit them we should get a set of PDFs that reproduce the assumed underlying PDFs, within the correct uncertainties, and criteria such as the quality of the matching between the input ( $f_{\text{in}}$ ) and the fitted ( $f_{\text{fit}}$ ) PDFs, or the agree-

ment between respective uncertainties, give a measure of how valid is our fitting methodology (including uncertainty determination).

As already explained, the fitting procedure is performed by minimizing a figure of merit, such as

$$\chi^2[\mathcal{T}[f], \mathcal{D}] = \frac{1}{N_{\mathcal{D}}} \sum_{I,J} (T_I[f] - D_I) C_{IJ}^{-1} (T_J[f] - D_J), \quad (2.3.1)$$

where  $\mathcal{D} = \{D_I\}$  is the set of pseudo-data,  $\mathcal{T}[f] = \{T_I[f]\}$  is a set of theoretical predictions delivered for the PDF  $f$  being fitted,  $C_{IJ}$  is the covariance matrix of the data and  $N_{\mathcal{D}}$  is the total number of data points of the dataset. It's important to note that the procedure used in fitting the pseudo-data is exactly the same as in fitting the real data: same figure of merit eq.(2.3.1) (evaluated with the real data replaced by pseudo-data), same covariance matrix, same fitting methodology. If we change by hand the level of uncertainties incorporated within the pseudo-data, we can use this test to study different sources of uncertainties related to the fitting procedure, or inconsistencies affecting the dataset. In particular, we can add different amounts of stochastic noise to the pseudo-data generated with the input PDF set; in order to have fluctuations and correlations of the pseudo-data reproducing precisely those of the real experimental data, the stochastic noise is generated using the experimental covariance matrix. So we define three different levels of closure test, Level 0, Level 1 and Level 2, each of them with an increasing amount of stochastic noise added to the pseudo-data.

### 2.3.1 Level 0 closure test

In Level 0 closure test no stochastic noise is added to pseudo-data  $\mathcal{D}_0 = \{D_I^0\}$  generated with the input PDF set. Then we perform  $N_{\text{rep}}$  fits with the same set of pseudo-data, minimizing the error function  $\chi^2[\mathcal{T}[f], \mathcal{D}_0]$  (computed with the covariance matrix of the data) for each fit, and for each fit we use a different seed to initialize random numbers used in minimization procedure. This produces an ensemble of PDF replicas  $\{f_{\text{fit}}^k\}$ , where  $k = 1, \dots, N_{\text{rep}}$ .

As pseudo-data have no inconsistencies by construction, and no stochastic noise is added, it's clear that in a Level 0 closure test the fit quality can be arbitrarily good. So, if the functional form used for parametrizing the fitted PDFs is flexible enough, the value of the  $\chi^2$  should decrease monotonically towards zero for each replica as the fit proceeds, as well as the best-fit  $\chi^2$ , i.e. the  $\chi^2$  evaluated for the average of all replicas. In this way, the Level 0 closure test is a highly non-trivial test of the efficiency of the minimization.

### 2.3.2 Level 1 closure test

In Level 1 closure test we use the pseudo-data generate in Level 0 closure test, and add to them stochastic fluctuations as follows:

$$D_I^1 = (1 + r_I^{\text{nor}} \sigma_I^{\text{nor}}) \left( D_I^0 + \sum_{p=1}^{N_{\text{sys}}} r_{I,p}^{\text{sys}} \sigma_{I,p}^{\text{sys}} + r_I^{\text{stat}} \sigma_I^{\text{stat}} \right), \quad (2.3.2)$$

where  $\sigma_I^{\text{stat}}$ ,  $\sigma_{I,p}^{\text{sys}}$  and  $\sigma_I^{\text{nor}}$  are the statistic, systematic and normalization uncertainties for each dataset, and the random numbers  $r_I^{\text{nor}}$ ,  $r_{I,p}^{\text{sys}}$  and  $r_I^{\text{stat}}$  are generated with the appropriate distribution to reproduce the experimental covariance matrix. These pseudo-data are used to fit all the  $N_{\text{rep}}$  replicas but, as in Level 0 closure test, we use a different seed to initialize the minimization

procedure for each replica. No other stochastic fluctuations are added to these pseudo-data, so experimental uncertainties are not propagated into the uncertainties of the fitted PDFs, and hence the ensemble of PDF replicas  $\{f_{\text{fit}}^k\}$  resulting from the Level 1 fits is expected to underestimate the PDF uncertainties.

If we make the pseudo-data fluctuate around the Level 0 by one standard deviation, we expect that the best-fit  $\chi^2$  will be around one. Moreover, there exist some “perfect” fits for which the  $\chi^2$  of the fitted PDFs coincides exactly with the  $\chi^2$  computed with the input PDFs used for the generation of the pseudo-data, i.e. such that  $\chi^2[\mathcal{T}[f_{\text{fit}}], \mathcal{D}_1] = \chi^2[\mathcal{T}[f_{\text{in}}], \mathcal{D}_1]$ . Actually, not all fits will feature this property, so, while in Level 0 the distribution of the fitted PDFs was close to a delta, here the distribution will spread a bit, but still remain too narrow.

### 2.3.3 Level 2 closure test

In Level 2 closure test we start from the pseudo-data generated in Level 1 closure test and construct  $N_{\text{rep}}$  Monte Carlo replicas producing, for each replica  $k$ , a set of pseudo-data  $\mathcal{D}_2^k = \{D_I^{2,k}\}$  as follows:

$$D_I^{2,k} = \left(1 + r_I^{\text{nor},k} \sigma_I^{\text{nor}}\right) \left(D_I^1 + \sum_{p=1}^{N_{\text{sys}}} r_{I,p}^{\text{sys},k} \sigma_{I,p}^{\text{sys}} + r_I^{\text{stat},k} \sigma_I^{\text{stat}}\right), \quad (2.3.3)$$

where the set of random numbers is different for each replica. Then, the pseudo-data are fitted in same way of real data, as already explained above. So in Level 2 fits, each Monte Carlo replica represents a fluctuation around the Level 1 pseudo-data, and the fluctuations of pseudo-data, due to experimental statistical, systematic and normalization uncertainties, are correctly propagated into the fitted PDFs, hence the ensemble of PDF replicas  $\{f_{\text{fit}}^k\}$  contains all the information on PDF uncertainties and correlations.

In this way, each replica contains two fluctuations, one from Level 1 pseudo-data generation, and one from Level 2, so the  $\chi^2[\mathcal{T}[f_{\text{fit}}^k], \mathcal{D}_2^k]$  computed for Level 2 is expected to be close to two. Finally, we expect the input PDFs  $f_{\text{in}}$  to lie within the  $1\text{-}\sigma$  band of the fitted PDFs  $f_{\text{fit}}$  with a probability of around 68%.

### 2.3.4 PDF uncertainties in closure tests

The different way to construct the pseudo-data in Level 0, Level 1 and Level 2 closure tests provide a different, increasing size of uncertainty on the fitted PDFs. In Level 0 there is only an interpolation and extrapolation uncertainty; in Level 1, the fluctuations added to the pseudo-data, the same for each replica, imply that different functional forms can produce equally good fits, yielding a source of uncertainty; in Level 2, as the fluctuations added to the pseudo-data are different for each replica, there is a real uncertainty on the data.

The uncertainty from Level 0 closure test is due to the fact that experimental data used in the fit has finite kinematical coverage, so, even if the fit can be arbitrarily good, between one point and another (i.e. in the interpolation region) and outside the data coverage (i.e. in the extrapolation region) the uncertainty of the fit will never tend to zero. In particular, in regions where  $x \rightarrow 0$  and  $x \rightarrow 1$  there is no experimental information, so here the uncertainties remain very large. This source of uncertainty is referred to as extrapolation uncertainty.

In Level 1 closure test the central values of the data fluctuate around the theoretical prediction, so, if the PDF parametrization is flexible enough, the minimization algorithm should find a large number of different functional forms that yield an equally good  $\chi^2[\mathcal{T}[f_{\text{fit}}], \mathcal{D}_1] \approx 1$ , adding to extrapolation uncertainty a new component, the functional uncertainty.

In Level 2 closure test, starting from Level 1 pseudo-data, we generate a set of  $N_{\text{rep}}$  Monte Carlo replicas that reflects the statistical and systematic errors given by the experimental measurements. As we said, in a Level 2 closure test we expect  $\chi^2[\mathcal{T}[f_{\text{fit}}], \mathcal{D}_2^k] \approx 2$ . The fit will give us an ensemble of fitted PDFs  $\{f_{\text{fit}}^k\}$  whose statistical properties are a faithful propagation of the fluctuations in the underlying dataset. The increase in the uncertainty from Level 1 to Level 2 fits represents a new component in the amount of uncertainties, which we call the data uncertainty.

If we compare Level 0, Level 1 and Level 2 closure tests with the same initial conditions, we can estimate the size of the three sources of uncertainties, extrapolation from Level 0, functional from Level 1 and data from Level 2. This is precisely what is done in Fig. 2.3, which shows the ratios of the uncertainty of the fitted PDFs to the respective central values in each case, with the MSTW08 NLO set used as input PDF  $f_{\text{in}}$ , at the input parametrization scale of  $Q^2 = 1 \text{ GeV}^2$ ; for the three fits, the PDF uncertainty bands are defined as the 68% confidence interval from the sample of  $N_{\text{rep}} = 100$  fitted replicas; PDFs for quark up, quark down and gluon are shown (other flavours feature similar behaviour). We can see that Level 0 uncertainties are smaller than Level 1 and in turn these are smaller than those at Level 2, confirming that raising up the level of the closure test increases the amount of uncertainties. We also observe that in the small- $x$  and large- $x$  regions the extrapolation uncertainty dominates, and hence Level 1 and Level 2 don't add a significative and new contribution to the uncertainty. An interesting thing is that in regions where we have a good coverage from available data, the three components are roughly of similar size, so the extrapolation, functional and data components are all significative for a correct determination of uncertainties. In general, we can conclude that a fit which misses out one of these three components will underestimate the overall PDF uncertainty.

Finally, the overall error of PDFs is computed as standard deviation over replicas, with the central value given by the average over replicas:

$$\langle f_{\text{fit}} \rangle = \frac{1}{N_{\text{rep}}} \sum_{k=1}^{N_{\text{rep}}} f_{\text{fit}}^k, \quad \sigma_{\text{fit}}^2 = \langle (f_{\text{fit}}^2 - \langle f_{\text{fit}} \rangle)^2 \rangle = \frac{1}{N_{\text{rep}}} \sum_{k=1}^{N_{\text{rep}}} (f_{\text{fit}}^k - \langle f_{\text{fit}} \rangle)^2. \quad (2.3.4)$$

### 2.3.5 Quantitative criteria for closure test validation

We now want to provide some quantitative criteria in order to validate a closure test and give an estimate of the goodness and stability of the fit; some of them will be used for the purpose of this work.

As in Level 1 and Level 2 closure test we expect the central  $\chi^2$ , computed averaging the fitted PDFs, to reproduce the  $\chi^2$  of the input PDFs, we can define the following estimator:

$$\Delta_{\chi^2} := \frac{\chi^2[\langle \mathcal{T}[f] \rangle, \mathcal{D}_i] - \chi^2[\mathcal{T}[f_{\text{in}}], \mathcal{D}_i]}{\chi^2[\mathcal{T}[f_{\text{in}}], \mathcal{D}_i]}, \quad (2.3.5)$$

which quantify the difference between the central  $\chi^2$  of the closure test fit, computed with an average over replica PDFs, and the  $\chi^2$  of the input PDF set, both computed with respect to the

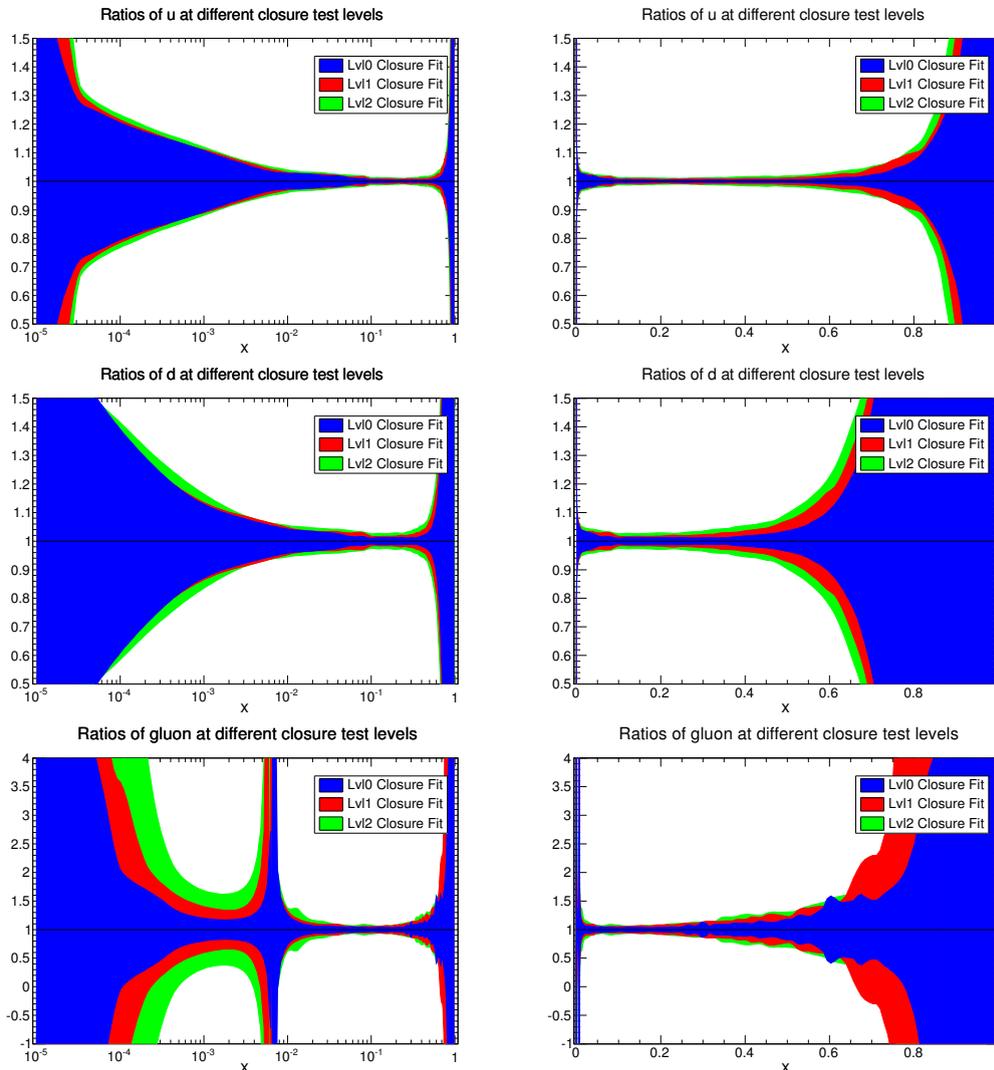


Fig. 2.3: Comparison of relative PDF uncertainties obtained from Level 0, Level 1 and Level 2 closure test fits with MSTW2008 NLO as input. The plots show the 68% confidence level PDF uncertainty band for each of the fits, normalized to the corresponding central value of each fit.

same closure test dataset; hence this estimator measures how close the closure test fit reproduces the theoretical predictions of the input PDFs, giving a quantitative measure of the success of the test. In particular,  $\Delta_{\chi^2} > 0$  arises from the situation in which the optimal  $\chi^2$  has not been reached yet (underlearning), while  $\Delta_{\chi^2} < 0$  indicates that the fit is following the fluctuations of the data (overlearning); obviously,  $\Delta_{\chi^2} = 0$  corresponds to perfect learning of the underlying law.

We now define an estimator to evaluate the accuracy with which PDF uncertainties are reproduced. If we assume gaussianity, the  $n$ - $\sigma$  interval about a prediction can be interpreted as confidence levels for the true value, in the sense that the true value will fall within 1- $\sigma$  band in 68.3% of cases, within 2- $\sigma$  band in 95.5% of cases, within 3- $\sigma$  band in 99.7% of cases, and so on.

Hence we define the following estimator:

$$\xi_{n\sigma} := \frac{1}{N_{\text{PDF}}} \frac{1}{N_x} \frac{1}{N_{\text{fits}}} \sum_{i=1}^{N_{\text{PDF}}} \sum_{j=1}^{N_x} \sum_{l=1}^{N_{\text{fits}}} I_{[-n\sigma_{\text{fit}}^{i(l)}(x_j), n\sigma_{\text{fit}}^{i(l)}(x_j)]} \left( \langle f_{\text{fit}}^{i(l)}(x_j) \rangle - f_{\text{in}}^i(x_j) \right), \quad (2.3.6)$$

where  $n$  is a positive integer,  $N_{\text{PDF}}$ ,  $N_x$  and  $N_{\text{fits}}$  are the number of PDF flavours,  $x$  values and fits respectively, over which averages are performed,  $I_A(x)$  is the indicator function of the interval  $A$  (which is one, if its argument lies in the interval  $A$ , and zero for all other values of its argument),  $\langle f_{\text{fit}}^{i(l)} \rangle$  and  $\sigma_{\text{fit}}^{i(l)}$  are the average PDFs and the corresponding standard deviation of the  $i$  PDF flavour for fit  $l$ , computed over the sample of  $N_{\text{rep}} = 100$  replicas of fit  $l$ . The PDFs can be sampled at various values of  $x$ . The estimators  $\xi_{1\sigma}$ ,  $\xi_{2\sigma}$ ,  $\dots$  provide the fraction of those fits for which the input PDF falls within one sigma, two sigma, etc of the central PDF  $\bar{f}_{\text{fit}}^{i(l)}$ , averaged over PDF flavours and values of  $x$ . In a successful closure test we should thus find that  $\xi_{1\sigma} \approx 0.68$ ,  $\xi_{2\sigma} \approx 0.95$ , etc. In principle, in order to calculate  $\xi_{1\sigma}$  we would need to, for instance, generate 100 closure test fits each one with  $N_{\text{rep}} = 100$  replicas. However performing this very large number of fits is very computationally expensive, and we would instead like to obtain an estimate of  $\xi_{1\sigma}$  which involves fewer fits. To achieve this, we can approximate the mean of each fit,  $\langle f_{\text{fit}}^{i(l)} \rangle$ , by fitting a single replica to each set of closure test data at Level 1, i.e. without additional replica fluctuations. We can also replace the individual values of  $\sigma^{i(l)}$  in with the corresponding values taken from a single 100 replica fit, making use of the fact that the variation in the PDF uncertainties between different closure test fits is small.

Another useful estimator is the distance between the fitted PDFs and the input PDFs, measured in units of the standard deviation of the fitted PDFs:

$$d_{\sigma} [f_{i,\text{fit}}, f_{i,\text{in}}] (x, Q) := \sqrt{\frac{(\bar{f}_{i,\text{fit}}(x, Q) - f_{i,\text{in}}(x, Q))^2}{\sigma^2 [f_{i,\text{fit}}] (x, Q)}}, \quad (2.3.7)$$

where  $i$  stands for the PDF flavour.

# Chapter 3

## Simulations and results

The purpose of this work is to study and analyze how the fitting procedure responds to the introduction of an inconsistency into the data; this inconsistency will be introduced by biasing the systematic uncertainties of the datasets, and we will discuss how the fitted PDFs and their uncertainties get modified with respect to a reference fit with no inconsistencies. In the next sections we explain the procedure we adopted to do this, the data used in the simulations, and the results we got.

### 3.1 Methodology

In this section we present the methodology we adopted in order to study how the fitting procedure responds to the introduction of an inconsistency into the data, i.e. which fitting procedure and datasets we used, how we introduced an inconsistency into the data, which criteria we used in comparing the results.

First of all, for all our simulations we use NNPDF group Monte Carlo fitting procedure [3], with a 2-5-3-1 neural network parametrization, and perform a series of Level 2 closure tests, as explained in Sect.2.3. For parametrization, we don't use the flavour basis, but the basis which diagonalize the DGLAP eqs.(1.3.2)–(1.3.3), because it directly relates physical observables to PDFs; this basis is the following:

$$\Sigma(x, Q_0^2) = (u + \bar{u} + d + \bar{d} + s + \bar{s})(x, Q_0^2) \quad (3.1.1)$$

$$T_3(x, Q_0^2) = (u + \bar{u} - d - \bar{d})(x, Q_0^2) \quad (3.1.2)$$

$$T_8(x, Q_0^2) = (u + \bar{u} + d + \bar{d} - 2s - 2\bar{s})(x, Q_0^2) \quad (3.1.3)$$

$$V(x, Q_0^2) = (u - \bar{u} + d - \bar{d} + s - \bar{s})(x, Q_0^2) \quad (3.1.4)$$

$$V_3(x, Q_0^2) = (u - \bar{u} - d + \bar{d})(x, Q_0^2) \quad (3.1.5)$$

$$V_8(x, Q_0^2) = (u - \bar{u} + d - \bar{d} - 2s + 2\bar{s})(x, Q_0^2), \quad (3.1.6)$$

and then the gluon PDF  $g(x, Q^2)$ . Each basis PDF at the reference scale is parametrized in terms of a neural network times a preprocessing factor:

$$f_i(x, Q_0) = A_i x^{-\alpha_i} (1-x)^{\beta_i} \text{NN}_i(x) \quad (3.1.7)$$

where  $A_i$  is an overall normalization constant, and the preprocessing term  $x^{-\alpha_i}(1-x)^{\beta_i}$  is introduced to speed up the minimization, without biasing the fit; finally,  $\text{NN}_i(x)$  is obviously the neural network.

Chosen an ensemble of experiments, i.e. of datasets, we adopt an existing PDF set as our “true theory” for producing pseudo-data in the closure test framework, as explained in Sec.2.3; details on which datasets and PDF set are used will be given in the next section. So, we perform a first level 2 closure test with the pure unmodified data, and this test will be our reference for all the other simulations. We do this because we want to have a reference with which we can compare the results of the fits with biased datasets.

Of this reference fit, we compute some estimators, in particular: a  $\chi_{\text{TD}}^2$  between the true PDF set and the generated pseudo-data, a  $\chi_{\text{FD}}^2$  between the fitted PDFs and the generated pseudo-data, and a  $\chi_{\text{FT}}^2$  between fitted and input PDFs. Moreover, we compute the criterion eq.(2.3.6) in order to evaluate the fraction of fits for which the theory falls within 1- $\sigma$ , 2- $\sigma$ , 3- $\sigma$  band of the central fitted PDF (this implies that for this reference level 2 closure test, and in general for every level 2 closure test, we perform also 100 level 1 closure tests with a single replica, as explained in Sec.2.3.5). The fits are evaluated on the PDFs of gluon, up, down, strange, anti-up, anti-down and anti-strange, at  $x = 0.05$ ,  $x = 0.1$  and  $x = 0.2$ .

Then, we introduce an inconsistency into the data generated for the closure test (obviously using the same theory as before): we take the level 2 pseudo-data eq.(2.3.3) and we rescale the systematic uncertainties by a suitable factor (2, 5 and 10), leaving unmodified the covariance matrix of the data used for the fit; in this way, we are increasing the fluctuations of the pseudo-data, but the uncertainties for the fit are remaining equal, and hence they will be underestimated. So, we perform the level 2 closure test with these biased data and evaluate the criteria explained above: the  $\chi_{\text{TD}}^2$  will tell us how much the pseudo-data have moved from the theory, while  $\chi_{\text{FD}}^2$  and  $\chi_{\text{FT}}^2$  will show if the fit has followed the data or the theory; the criterion eq.(2.3.6) will tell us how the distribution of the fits has changed with respect to the reference fit.

The systematic uncertainties of the experimental data are of two types, additive and multiplicative. While the former represents a shift to the data and do not depend on them, the latter depend on the data, in the sense that the bigger is the value of the data and the bigger is the associated multiplicative systematic uncertainty. In this work we have studied both of them, introducing the bias separately into each one.

Another way of introducing an inconsistency into the data is to conversely rescale the covariance matrix, i.e. dividing the matrix entries by the same factor, but leaving unmodified the fluctuations of the pseudo-data in eq.(2.3.3); in this way we are directly underestimating the uncertainties for the fit; this second way has been exploited only for additive systematic uncertainties.

## 3.2 Input PDF set and datasets used in closure tests

In this section we present the datasets we used in all our tests, and which PDF set we adopted in the closure test procedure.

For all our closure tests, we use MSTW08 NLO PDF set [9] as input PDFs  $f_{\text{in}}$ . As far as datasets is concerned, we use fixed-target deep inelastic scattering (DIS) data from experiments with charged lepton beams (NMC, SLAC, BCDMS) and neutrino beams (CHORUS, NuTeV); we use this small ensemble of datasets because we are not interested in producing an accurate determination of a PDF set, but we want to study how the fitting procedure responds to the introduction of an inconsistency into the datasets, and for this purpose few datasets are enough. Table 3.1 shows in details the datasets of each experiment, with reference and number of available data.

Experiment	Dataset	Reference	$N_{\text{dat}}$
NMC	NMC $d/p$	[17]	121
	NMC $\sigma^{\text{NC,p}}$	[18]	204
SLAC	SLAC $p$	[19]	33
	SLAC $d$	[19]	34
BCDMS	BCDMS $p$	[20]	333
	BCDMS $d$	[21]	248
CHORUS	CHORUS $\nu$	[22]	416
	CHORUS $\bar{\nu}$	[22]	416
NuTeV	NuTeV $\nu$	[23, 24]	39
	NuTeV $\bar{\nu}$	[23, 24]	37

Table 3.1: Experiments and datasets used in this work.

The physical processes involved in these experiments are listed in the Table 3.2.

Process	Subprocess	Partons	$x$ range	Experiment
$\ell^\pm \{p, n\} \rightarrow \ell^\pm X$	$\gamma^* q \rightarrow q$	$q, \bar{q}, g$	$x \geq 0.01$	NMC, SLAC, BCDMS
$\ell^\pm n/p \rightarrow \ell^\pm X$	$\gamma^* d/u \rightarrow d/u$	$d/u$	$x \geq 0.01$	NMC, SLAC, BCDMS
$\nu(\bar{\nu}) N \rightarrow \mu^-(\mu^+) X$	$W^* q \rightarrow q'$	$q, \bar{q}$	$0.01 \leq x \leq 0.5$	CHORUS, NUTEV
$\nu N \rightarrow \mu^- \mu^+ X$	$W^* s \rightarrow c$	$s$	$0.01 \leq x \leq 0.2$	CHORUS, NUTEV
$\bar{\nu} N \rightarrow \mu^+ \mu^- X$	$W^* \bar{s} \rightarrow \bar{c}$	$\bar{s}$	$0.01 \leq x \leq 0.2$	CHORUS, NUTEV

Table 3.2: Processes involved in experiments of Table 3.1.

In Table 3.3 we report, for each dataset, the measured observable, the most significant systematic uncertainty and its size with respect to the observable.

Dataset	Observable	Most Significant Systematic Uncertainty	Size
NMC $d/p$	$F_d^2/F_p^2$	radiative corrections and functional form of the ratio $F_d^2/F_p^2$ parametrizations	1%
NMC $\sigma^{\text{NC},p}$	$\frac{d^2\sigma_p}{dx dQ^2}, \frac{d^2\sigma_d}{dx dQ^2}$	acceptance	1–2%
SLAC $p$	$F_p^2$	overall normalization of the combined SLAC data	2%
SLAC $d$	$F_d^2$	overall normalization of the combined SLAC data	2%
BCDMS $p$	$F_p^2$	normalization uncertainty correlated for all data points	3%
BCDMS $d$	$F_d^2$	normalization uncertainty correlated for all data points	3%
CHORUS $\nu$	$\frac{d\sigma(\nu_\mu N \rightarrow \mu^- \mu^+ X)}{dx dy}$	correction factor to obtain differential cross-sections corrected for QED radiation effects	1–10%
CHORUS $\bar{\nu}$	$\frac{d\sigma(\bar{\nu}_\mu N \rightarrow \mu^+ \mu^- X)}{dx dy}$	correction factor to obtain differential cross-sections corrected for QED radiation effects – 1-10%	
NuTeV $\nu$	$\frac{d\sigma(\nu_\mu N \rightarrow \mu^- \mu^+ X)}{dx dy}$	overall sys. unc. obtained from different contributions	5–10%
NuTeV $\bar{\nu}$	$\frac{d\sigma(\bar{\nu}_\mu N \rightarrow \mu^+ \mu^- X)}{dx dy}$	overall sys. unc. obtained from different contributions	5–10%

Table 3.3: Most significant systematic uncertainties for experiments of Table 3.1.

All the systematic uncertainties of Table 3.3 are treated as additive systematics, while the only multiplicative systematic uncertainties in fixed-target DIS experiments are normalizations.

### 3.3 Reference closure test

Here we report the results of the reference closure test which, as said before, is performed with datasets of Table 3.1 and all systematic uncertainties unmodified. We perform a reference closure test fit in order to have a “perfect”, unbiased fit that we can compare with the fits in which we will introduce an inconsistency.

Fig. 3.1 shows the three  $\chi^2_{\text{TD}}$ ,  $\chi^2_{\text{FD}}$  and  $\chi^2_{\text{FT}}$  for the 5 fitted datasets, with global  $\chi^2$  reported in the title.

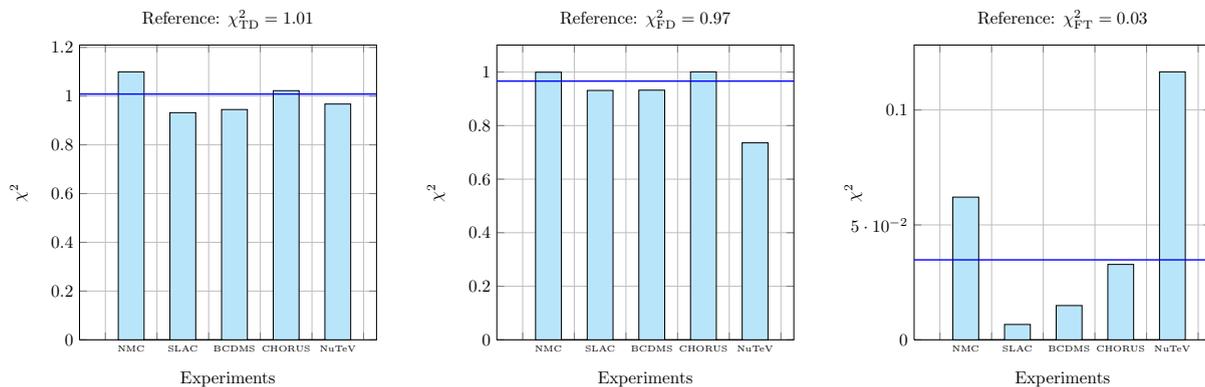


Fig. 3.1:  $\chi^2_{\text{TD}}$ ,  $\chi^2_{\text{FD}}$  and  $\chi^2_{\text{FT}}$  for the reference closure test, with global  $\chi^2$  in blue line.

We see that with no inconsistencies in the datasets, the pseudo-data fluctuate around the theory by  $1\text{-}\sigma$ , as shown by  $\chi^2_{\text{TD}}$ ; the fit follows more the theory than the data, as  $\chi^2_{\text{FT}}$  is almost vanishing, while  $\chi^2_{\text{FD}}$  is quite similar to  $\chi^2_{\text{TD}}$ ; so, as we can expect, the fit fluctuate less than the data. Finally, we can use the criterion eq.(2.3.6) and construct the distribution of the deviations of the fitted PDFs from the input PDFs, which is shown in Fig. 3.2 together with a suitably scaled gaussian distribution with standard deviation 1.

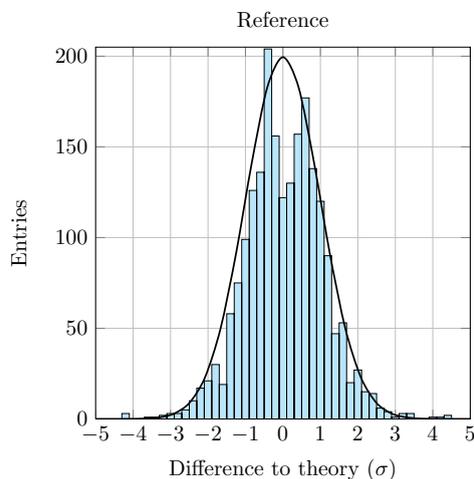


Fig. 3.2: Histograms for the difference between the input PDFs and the single replica fits in units of the fitted PDFs standard deviation.

We can see that the resulting distribution is very close to a gaussian one, but in order to give a more quantitative evaluation, we can compute the fractions of fits which fall within  $1\text{-}\sigma$ ,  $2\text{-}\sigma$  or  $3\text{-}\sigma$  band from the input PDFs, i.e. the estimators  $\xi_{n\sigma}$  of eq.(2.3.5). Results are shown in Table 3.4, from which we see that the agreement with theory is good also from a more quantitatively point of view.

Band	Theory (%)	Reference (%)
$1\text{-}\sigma$	68.3	69.8
$2\text{-}\sigma$	95.5	94.5
$3\text{-}\sigma$	99.7	99.0

Table 3.4: Fractions of fits falling within  $n\text{-}\sigma$  band from input PDFs.

### 3.4 Rescaling of additive systematic uncertainties in closure test fluctuations

In this section we present the results of the closure tests in which we have rescaled the systematic uncertainties of Table 3.3 in the fluctuations of the pseudo-data eq.(2.3.3). As said, the rescaling consists in multiplying the uncertainties by factors 2, 5 and 10, while keeping unmodified the covariance matrix.

First of all, Fig. 3.3 shows the distributions of  $\chi_{\text{TD}}^2$  (between pseudo-data and true theory) computed for the three simulations (scaling factor 2, 5 and 10) and compared to the reference closure test; global  $\chi^2$  is reported in the title.

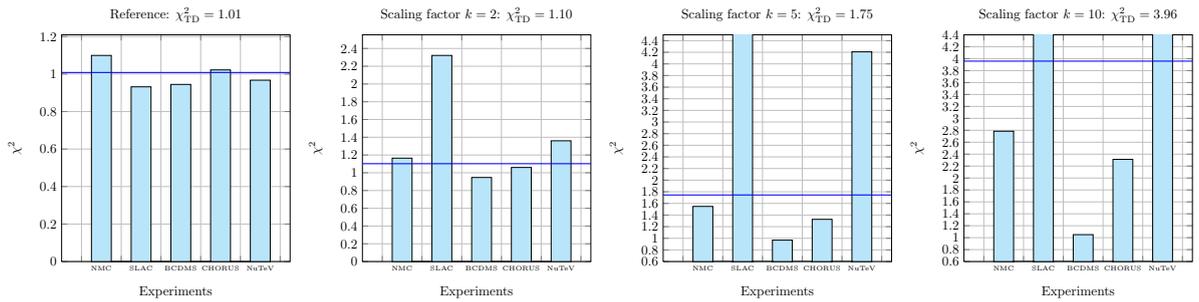


Fig. 3.3:  $\chi_{\text{TD}}^2$  for the closure tests after having rescaled by  $k$  the additive systematic uncertainties of all datasets entering the fluctuations in pseudo-data; global  $\chi_{\text{TD}}^2$  in blue line.

We can see that the  $\chi_{\text{TD}}^2$  increases as the fluctuations in the pseudo-data are progressively increased, and this is telling us that pseudo-data are moving away from the theory; if we rescale the fluctuation by a factor 10, pseudo-data move away from the theory by almost  $4\sigma$ . Moreover, we can see that  $\chi_{\text{TD}}^2$  doesn't increase equally for every dataset: for example, the increase for SLAC dataset is much greater than for BCDMS dataset; this is due to the fact that the fluctuations added to the pseudo-data are not equally selected for every dataset, but they are weighted by a random number, as explained in eq.(2.3.3); hence, if the random number for a given dataset is small, then the dataset will not be sensitive to an increasing of the fluctuations, and this is the case of BCDMS.

Now we can compute  $\chi_{\text{FD}}^2$  (between fit and data) and  $\chi_{\text{FT}}^2$  (between fit and true theory) in order to see how the fit responds to the introduction of the inconsistency; the two  $\chi^2$  distributions are shown in Fig. 3.4 and Fig. 3.5.

We see that the distribution of  $\chi_{\text{FD}}^2$  is quite similar to  $\chi_{\text{TD}}^2$  for each scaled fit, while  $\chi_{\text{FT}}^2$  of each scaled fit remains similar to the reference one, meaning that as we introduce an inconsistency into the datasets, the pseudo-data move away from the theory, but the fit remains close to it.

Now, in order to understand what happens with the uncertainties on the final fit, we construct the distributions of the deviations of the fitted PDFs from the input PDFs for each closure test, which is shown in Fig. 3.6 together with the reference closure test. The fractions of fits which fall within  $1\text{-}\sigma$ ,  $2\text{-}\sigma$  or  $3\text{-}\sigma$  band from the input PDFs are listed in Table 3.5.

From a qualitative point of view, we can see that the distributions get more peaked towards the central bins as the fluctuations in the pseudo-data are increased. This fact is confirmed more

### 3.4. Rescaling of additive systematic uncertainties in closure test fluctuations 28

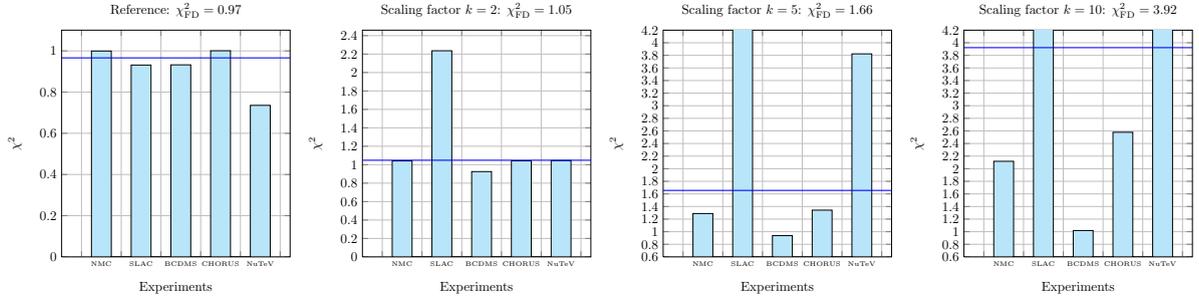


Fig. 3.4:  $\chi_{\text{FD}}^2$  for the closure tests after having rescaled by  $k$  the additive systematic uncertainties of all datasets entering the fluctuations in pseudo-data; global  $\chi_{\text{FD}}^2$  in blue line.

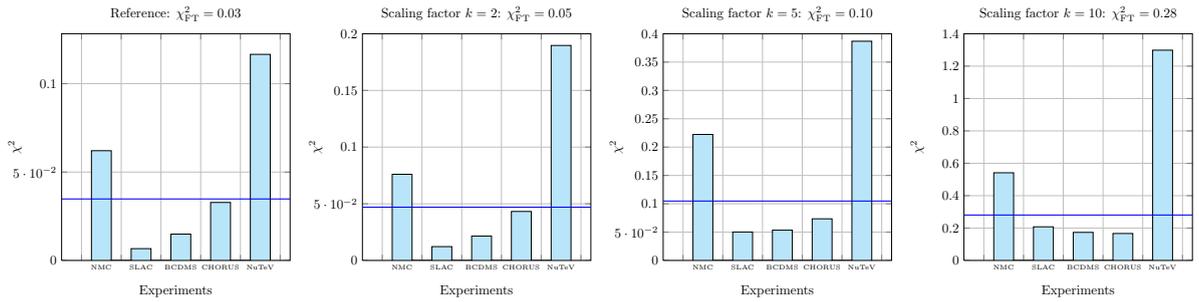


Fig. 3.5:  $\chi_{\text{FT}}^2$  for the closure tests after having rescaled by  $k$  the additive systematic uncertainties of all datasets entering the fluctuations in pseudo-data; global  $\chi_{\text{FT}}^2$  in blue line.

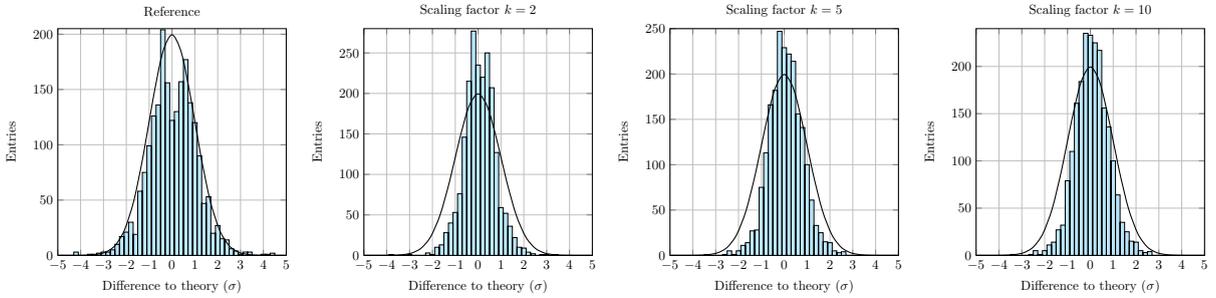


Fig. 3.6: Histograms for the difference between the input PDFs and the single replica fits in units of the fitted PDFs standard deviation, after having rescaled by  $k$  the additive systematic uncertainties of all datasets entering the fluctuations in pseudo-data.

Band	Theory (%)	Reference (%)	$k = 2$ (%)	$k = 5$ (%)	$k = 10$ (%)
1- $\sigma$	68.3	69.8	86.3	84.4	83.7
2- $\sigma$	95.5	94.5	99.3	98.8	98.8
3- $\sigma$	99.7	99.0	99.6	99.9	100

Table 3.5: Fractions of fits falling within  $n$ - $\sigma$  band from input PDFs.

quantitatively by the fractions of fits falling in 1- $\sigma$  band and 2- $\sigma$  band, which increase with respect to the reference fit. The explanation to this fact is that as we introduce an inconsistency by increasing the fluctuations in the pseudo-data, the central fit remains close to the true theory,

as confirmed by the  $\chi_{\text{FT}}^2$  distribution, however the final uncertainties on the fit get larger, so the  $n$ - $\sigma$  bands of the fitted PDFs are wider and they can accommodate the input PDFs in more cases than in reference fit, resulting in increasing 1- $\sigma$  fraction and 2- $\sigma$  fraction. We can see this in Fig. 3.7 which shows the PDFs of the gluon and the nonsinglet quark distribution  $T_3$  at high  $x$ -value, for the  $k = 10$  scaled fit in comparison to reference fit: the error bands of the scaled fit are larger than those of reference fit by a factor up to 2-3. So if we rescale the fluctuations by a factor 10, the final uncertainties on the fit increase, but not by the same factor, rather by a factor 2-3; and even less the central fitted PDFs move from the theory.

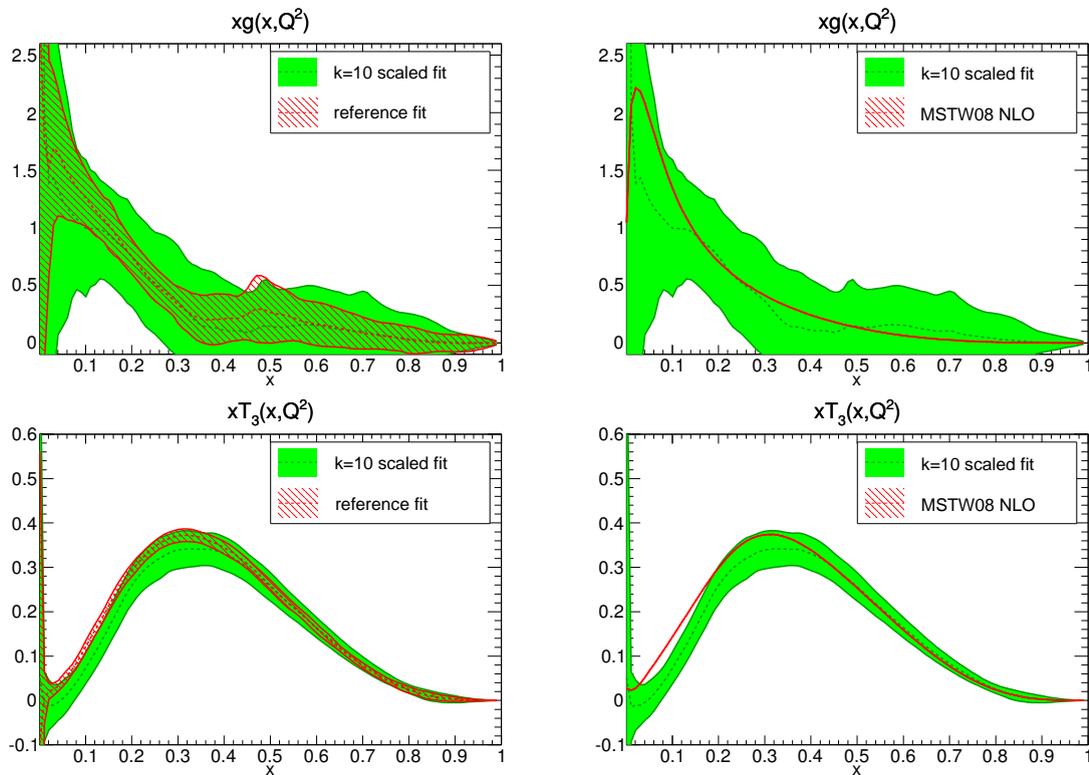


Fig. 3.7: Parton distribution functions for the gluon (top) and the nonsinglet quark distribution  $T_3$  (bottom): reference fit vs.  $k = 10$  scaled fit on the left,  $k = 10$  scaled fit vs. MSTW08 NLO on the right.

### 3.5 Rescaling of multiplicative systematic uncertainties in closure test fluctuations

In this section we present the results of the closure tests in which we have rescaled the multiplicative systematic uncertainties of all datasets in the fluctuations of the pseudo-data eq.(2.3.3). Again, the rescaling consists in multiplying the uncertainties by factors 2, 5 and 10, while keeping unmodified the covariance matrix.

Fig. 3.8 shows the distributions of  $\chi_{TD}^2$  (between pseudo-data and true theory) computed for the three simulations (scaling factor 2, 5 and 10) and compared to the reference closure test; global  $\chi^2$  is reported in the title.

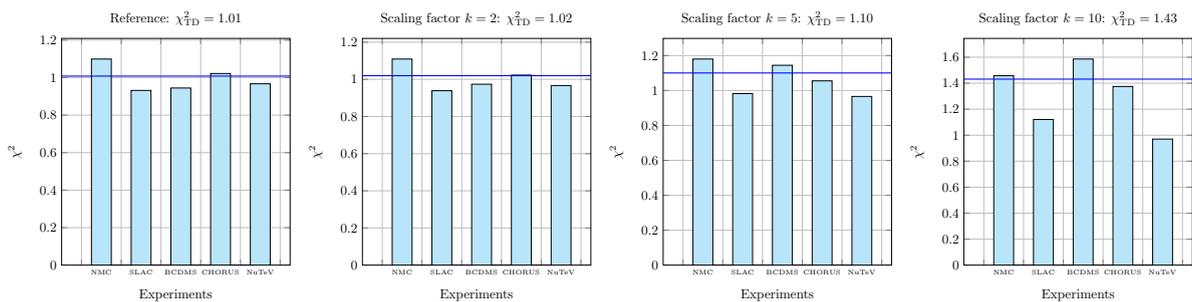


Fig. 3.8:  $\chi_{TD}^2$  for the closure tests after having rescaled by  $k$  the multiplicative systematic uncertainties of all datasets entering the fluctuations in pseudo-data; global  $\chi_{TD}^2$  in blue line.

We can see that the increase in the  $\chi_{TD}^2$ , as the fluctuations are progressively increased, is very small, compared to the case of additive systematics. This happens because multiplicative systematic uncertainties are proportional to the data, so if we make pseudo-data fluctuate more, their multiplicative uncertainties will fluctuate as well, they will get larger and hence the  $\chi_{TD}^2$  between data and theory will not change very much. In other words, we are always moving the data, but the farther we move the data and the bigger will be the uncertainties (keeping  $\chi_{TD}^2$  stable).

This behaviour is confirmed also by  $\chi_{FD}^2$  (between fit and data) and  $\chi_{FT}^2$  (between fit and true theory), whose distributions are shown in Fig. 3.9 and Fig. 3.10.

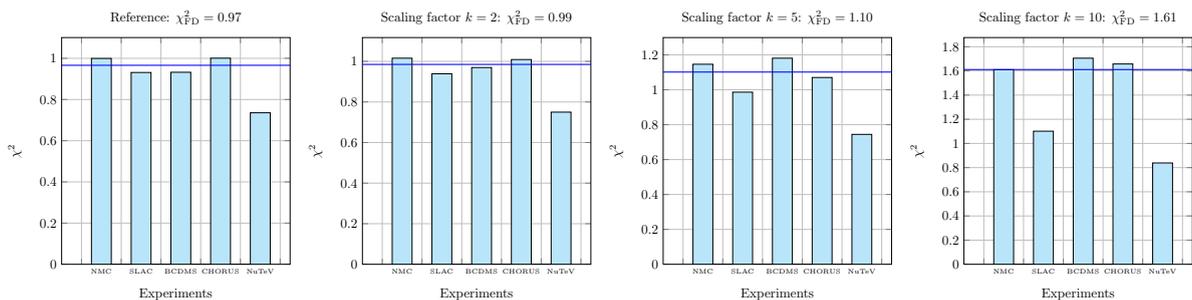


Fig. 3.9:  $\chi_{FD}^2$  for the closure tests after having rescaled by  $k$  the multiplicative systematic uncertainties of all datasets entering the fluctuations in pseudo-data; global  $\chi_{FD}^2$  in blue line.

We see that the fit follows the true theory, as the  $\chi_{FT}^2$  remains close to zero, but also the  $\chi_{FD}^2$  doesn't increase very much: also in this case, as the data are moved away from the theory

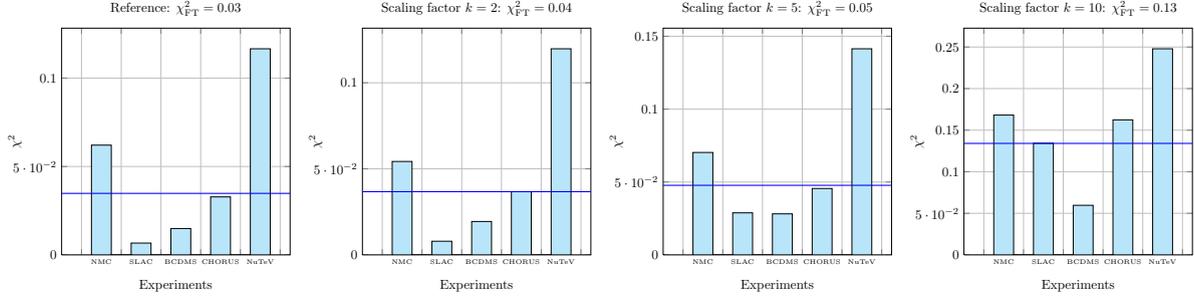


Fig. 3.10:  $\chi_{\text{FT}}^2$  for the closure tests after having rescaled by  $k$  the multiplicative systematic uncertainties of all datasets entering the fluctuations in pseudo-data; global  $\chi_{\text{FT}}^2$  in blue line.

by increasing the fluctuations coming from multiplicative systematics, the fit distances from the data, but, as multiplicative uncertainties are proportional to the fluctuated data, the  $\chi_{\text{FD}}^2$  between fit and data does not increase very much.

Finally, we construct the distributions of the deviations of the fitted PDFs from the input PDFs for the  $k = 2$  scaled closure test, which is shown in Fig. 3.11 together with the reference closure test. The fractions of fits which fall within  $1\text{-}\sigma$ ,  $2\text{-}\sigma$  or  $3\text{-}\sigma$  band from the input PDFs are listed in Table 3.6.

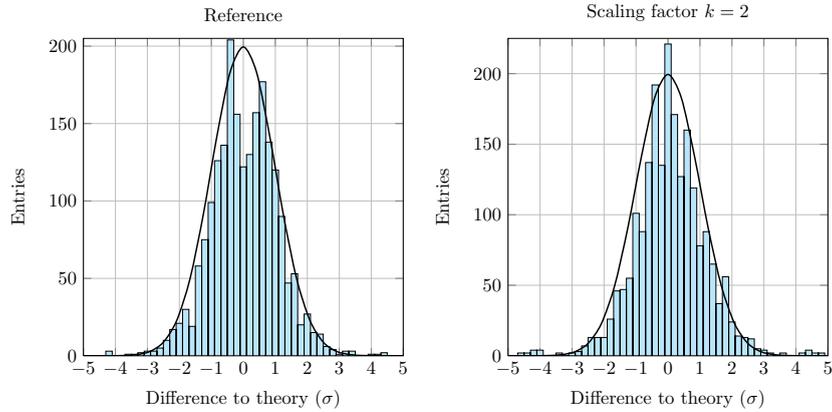


Fig. 3.11: Histograms for the difference between the input PDFs and the single replica fits in units of the fitted PDFs standard deviation, after having rescaled by  $k$  the multiplicative systematic uncertainties of all datasets entering the fluctuations in pseudo-data.

Band	Theory (%)	Reference (%)	$k = 2$ (%)
$1\text{-}\sigma$	68.3	69.8	68.0
$2\text{-}\sigma$	95.5	94.5	93.9
$3\text{-}\sigma$	99.7	99.0	98.6

Table 3.6: Fractions of fits falling within  $n\text{-}\sigma$  band from input PDFs.

Also with this criterion we see no big changes compared to reference: the shape of the distribution remains quite gaussian, and the fractions of fits falling within  $1\text{-}\sigma$ ,  $2\text{-}\sigma$  or  $3\text{-}\sigma$  band from the

input PDFs are similar to the reference ones. So we find that if we introduce an inconsistency into the data by rescaling the multiplicative uncertainties, the fit doesn't experience big changes, because multiplicative uncertainties are proportional to data, and hence a bigger fluctuations in the data will involve bigger uncertainties, so everything is rescaled in the same way. As a matter of fact, the final uncertainties on the PDFs are similar to the reference ones (while in the additive systematics case they were much larger), as we can see from the PDFs of the gluon and the nonsinglet quark distribution  $T_3$  at high  $x$ -value, for the  $k = 2$  scaled fit in comparison to reference fit, shown in Fig. 3.12.

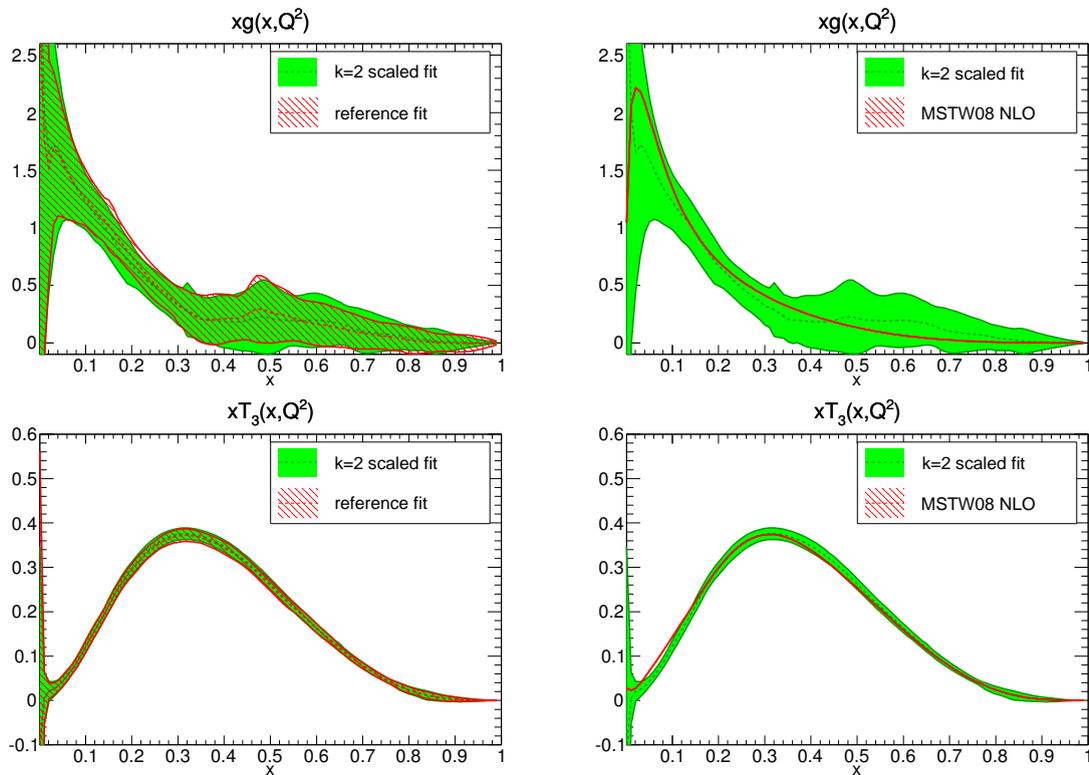


Fig. 3.12: Parton distribution functions for the gluon (top) and the nonsinglet quark distribution  $T_3$  (bottom): reference fit vs.  $k = 2$  scaled fit on the left,  $k = 2$  scaled fit vs. MSTW08 NLO on the right.

### 3.6 Rescaling of additive systematic uncertainties in covariance matrix

In this section we present the results of the closure tests in which the inconsistency has been introduced not by rescaling the fluctuations of the pseudo-data, but directly reducing the systematic uncertainties entering the covariance matrix used for the fit. In this way we are underestimating the additive systematic uncertainties, and hence we expect results similar to those obtained by rescaling the additive systematic uncertainties entering the fluctuations.

Fig. 3.13 shows the distributions of  $\chi_{\text{TD}}^2$  (between pseudo-data and true theory) computed for the three simulations (scaling factor 2, 5 and 10) and compared to the reference closure test; global  $\chi^2$  is reported in the title.

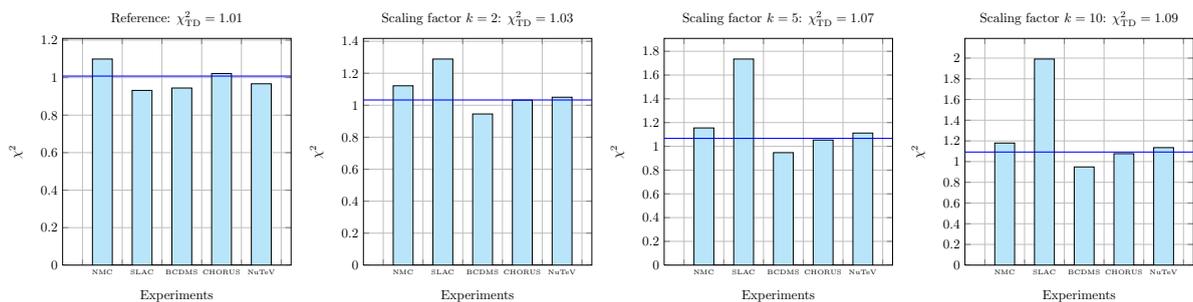


Fig. 3.13:  $\chi_{\text{TD}}^2$  for the closure tests after having divided by  $k$  the additive systematic uncertainties of all datasets entering the covariance matrix; global  $\chi_{\text{TD}}^2$  in blue line.

We see that the  $\chi_{\text{TD}}^2$  remains stable as the inconsistency is introduced: this is not surprising, as we are not rescaling the fluctuations, i.e. we are not moving the data from the theory, but we are reducing the uncertainties in the covariance matrix.

Fig. 3.14 and Fig. 3.15 show the distributions of  $\chi_{\text{FD}}^2$  (between fit and data) and  $\chi_{\text{FT}}^2$  (between fit and true theory).

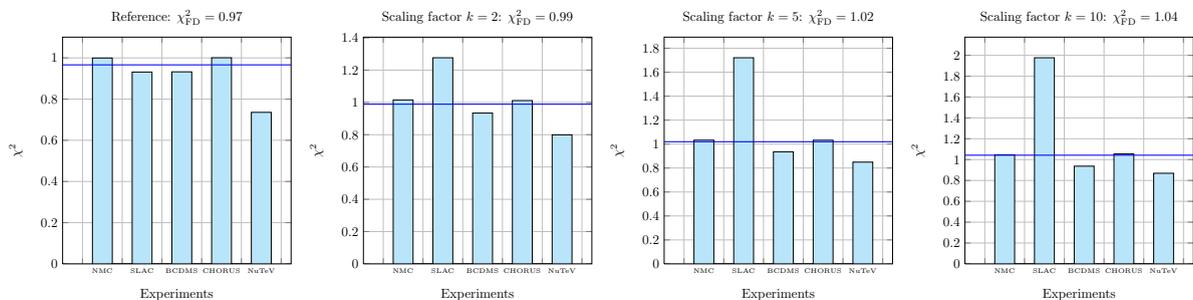


Fig. 3.14:  $\chi_{\text{FD}}^2$  for the closure tests after having divided by  $k$  the additive systematic uncertainties of all datasets entering the covariance matrix; global  $\chi_{\text{FD}}^2$  in blue line.

As expected, the  $\chi_{\text{FD}}^2$  doesn't change significantly, and is quite similar to the  $\chi_{\text{TD}}^2$ ; as a matter of fact, the fit remains close the theory, with  $\chi_{\text{FT}}^2$  being stable and vanishing.

Finally, we construct the distributions of the deviations of the fitted PDFs from the input PDFs for each closure test, which is shown in Fig. 3.16 together with the reference closure test. The

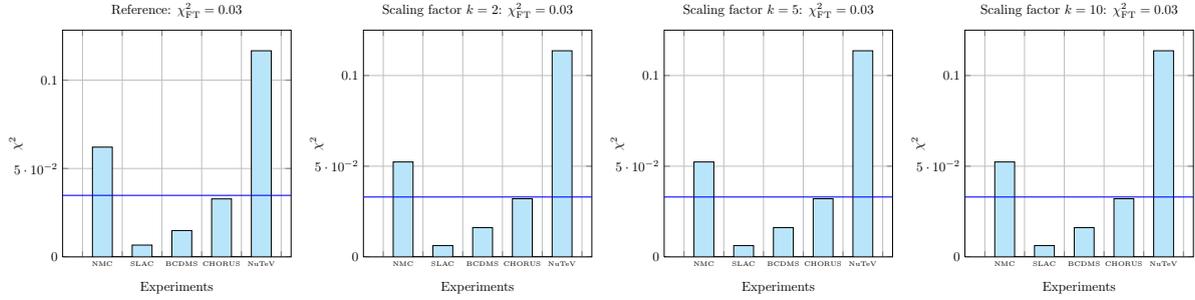


Fig. 3.15:  $\chi_{\text{FT}}^2$  for the closure tests after having divided by  $k$  the additive systematic uncertainties of all datasets entering the covariance matrix; global  $\chi_{\text{FT}}^2$  in blue line.

fractions of fits which fall within  $1\text{-}\sigma$ ,  $2\text{-}\sigma$  or  $3\text{-}\sigma$  band from the input PDFs are listed in Table 3.7.

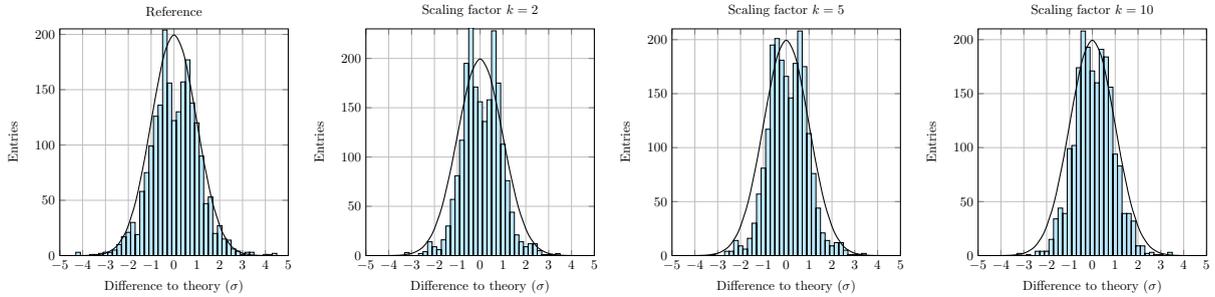


Fig. 3.16: Histograms for the difference between the input PDFs and the single replica fits in units of the fitted PDFs standard deviation, after having divided by  $k$  the additive systematic uncertainties of all datasets entering the covariance matrix.

Band	Theory (%)	Reference (%)	$k = 2$ (%)	$k = 5$ (%)	$k = 10$ (%)
$1\text{-}\sigma$	68.3	69.8	80.0	80.0	77.7
$2\text{-}\sigma$	95.5	94.5	96.9	96.9	98.4
$3\text{-}\sigma$	99.7	99.0	99.7	99.8	99.8

Table 3.7: Fractions of fits falling within  $n\text{-}\sigma$  band from input PDFs.

From a qualitative point of view, we see that the tails of the distributions get lower and the shape is more peaked, although not as much as we found by rescaling the fluctuations. This behaviour is confirmed more quantitatively by the fractions of fits falling within  $1\text{-}\sigma$  and  $2\text{-}\sigma$  band from the input PDFs, which are greater than the reference ones, as found by rescaling the fluctuations (and also in this case the increase with respect to the reference is smaller than in the “fluctuation case”). So we get results quite similar to those obtained by rescaling the additive systematic uncertainties in the fluctuations, but this time the changes with respect to the reference are a bit smaller. The reason why in this case the effects are a bit smaller when we increase the inconsistency is that in the covariance matrix the uncertainties are summed in quadrature, so if we halve the most significant uncertainty, this will not be significant anymore (because there is not a big difference in size between the different sources of uncertainty), and if

we keep dividing it by greater factors, we are doing nothing but reducing a source of uncertainty which is already negligible with respect to the others, hence there will be no effects in the fit.

### 3.7 Final conclusions

To sum up the work, our task was to study how the NNPDF fitting procedure, featuring neural network parametrization and Monte Carlo approach for uncertainties, responds to the introduction of an inconsistency into the datasets.

We have used a Level 2 closure test procedure, in which we have introduced an inconsistency into the datasets, performed a closure test fit and compared it to a reference closure test fit with no bias. In all tests we have used only fixed-target deep inelastic scattering data from experiments with charged lepton beams (NMC, SLAC, BCDMS) and neutrino beams (CHORUS, NuTeV).

Firstly, we have introduced the inconsistency by increasing the most significant additive systematic uncertainty of each dataset, widening in this way the fluctuations added to the pseudo-data in the closure test, and keeping fixed the covariance matrix of the datasets (in this way, we are underestimating the uncertainties on the datasets); fluctuations have been increased by factor 2, 5 and 10. In this way, we have obtained that pseudo-data generated by the level 2 closure test procedure move from the true theory (with a  $\chi_{\text{TD}}^2$  between data and theory passing from 1.01 to 3.96), and the fit remains far from the data by the same quantity ( $\chi_{\text{FD}}^2$  between fit and data passing from 0.97 to 3.92), and very close to the true theory ( $\chi_{\text{FT}}^2$  between fit and theory equal to 0.03 for the reference and 0.28 for the  $k = 10$  scaled fit). Moreover, the fraction of fits falling within  $1\text{-}\sigma$  band from the input PDF set increase when an inconsistency is introduced, passing from the value of 69.8% for the reference fit to  $\sim 85\%$  for the scaled fits. This suggests that if we increase the fluctuations coming from the additive systematics, the fit doesn't move away from the theory, but the final uncertainties on the fit will be larger, and the  $1\text{-}\sigma$  band gets wider. So if we introduce an inconsistency into the datasets, via additive systematics, the fit will respond to the bias increasing the final uncertainties with no need to increase the covariance matrix of the datasets; this suggests that the NNPDF methodology doesn't need a tolerance, as the Hessian approach, in order to account for inconsistencies introduced into the datasets, but it responds to them by itself.

Second, we have done the same work with multiplicative systematic uncertainties, which are proportional to the data. In this case, when we introduce the inconsistency by increasing the fluctuations coming from multiplicative systematics, the pseudo-data generated by the level 2 closure test procedure remain quite close to the true theory ( $\chi_{\text{TD}}^2$  passing from 1.01 to 1.43), and so does the fit with respect to data ( $\chi_{\text{FD}}^2$  passing from 0.97 to 1.61), and to theory ( $\chi_{\text{FT}}^2$  passing from 0.03 to 0.13). Also the fraction of fits falling within  $1\text{-}\sigma$  band from the input PDF set remains stable (69.8% for the reference fit and 68% for the  $k = 2$  scaled fit). In this case, as multiplicative systematics are proportional to the data, if we make the data fluctuate more, the uncertainties will increase as well, so even if the data move away from the theory, the  $\chi_{\text{TD}}^2$  will not be very sensitive to this (because uncertainties are larger). We also find that final uncertainties on the fit are quite similar to the reference ones, so if we introduce an inconsistency into the datasets via multiplicative systematics, the fit doesn't experience significant changes.

Finally, we have introduced the inconsistency in a different way, i.e. reducing the additive systematic uncertainties entering the covariance matrix used for the fitting procedure, and keeping

---

unmodified the fluctuations for the closure test. In this framework, obviously the pseudo-data don't move away from the true theory ( $\chi_{\text{TD}}^2$  passing from 1.01 to 1.09), and the fit remains very close to the true theory ( $\chi_{\text{TD}}^2 = 0.03$  as for the reference fit). We find a change in the fraction of fits falling within  $1\text{-}\sigma$  band from the input PDF set, which increases from the reference 69.8% to  $\sim 80\%$  of the scaled fits, with a behaviour quite similar to the first case, although with smaller effects. Also in this case we are underestimating the uncertainties on the datasets, and we get a final fit with wider bands of uncertainties, so the  $1\text{-}\sigma$  fraction increases. The difference with the first case is that in covariance matrix uncertainties are summed in quadrature, so if we reduce only the most significant one, the others will produce the same effects, and hence the global changes will be smaller.

# Bibliography

- [1] S. Forte, *Parton distributions at the dawn of the LHC*, Acta Phys.Polon. **B41** (2010) 2859–2920, [arXiv:1011.5247v2](#)
- [2] S. Forte and G. Watt, *Progress in the Determination of the Partonic Structure of the Proton* Ann. Rev. Nucl. Part. Sci. **63** (2013) 291, [arXiv:1301.6754v1](#)
- [3] **NNPDF** Collaboration, R.D. Ball, S. Carrazza, S. Forte, et al., *Parton distributions for the LHC Run II*, JHEP **04** (2015) 040, [arXiv:1410.8849v4](#)
- [4] A.J. Buras and K.J.F. Gaemers, *Simple Parametrizations of Parton Distributions with  $q^2$  Dependence Given by Asymptotic Freedom*, Nucl. Phys. **B132** (1978) 249.
- [5] M. Glück, E. Hoffmann, and E. Reya, *Scaling Violations and the Gluon Distribution of the Nucleon*, Zeit. Phys. **C13** (1982) 119.
- [6] D.W. Duke and J.F. Owens,  *$Q^2$  Dependent Parametrizations of Parton Distribution Functions*, Phys. Rev. **D30** (1984) 49–54.
- [7] **CTEQ** Collaboration, H.-L. Lai et al., *Global QCD analysis of parton structure of the nucleon: CTEQ5 parton distributions*, Eur. Phys. J. **C12** (2000) 375–392, [arXiv:hep-ph/9903282](#)
- [8] A.D. Martin, R.G. Roberts, W.J. Stirling, and R.S. Thorne, *MRST2001: Partons and  $\alpha_S$  from precise deep inelastic scattering and Tevatron jet data*, Eur. Phys. J. **C23** (2002) 7387, [arXiv:hep-ph/0110215](#)
- [9] A.D. Martin, W.J. Stirling, R.S. Thorne, and G. Watt, *Parton distributions for the LHC*, Eur.Phys.J. **C63** (2009) 189–285, [arXiv:0901.0002](#)
- [10] H.-L. Lai, M. Guzzi, J. Huston, Z. Li, P. M. Nadolsky, et al., *New parton distributions for collider physics*, Phys.Rev. **D82** (2010) 074024, [arXiv:1007.2241](#)
- [11] S. Forte, L. Garrido, J.I. Latorre, and A. Piccione, *Neural network parametrization of deep inelastic structure functions*, JHEP **0205** (2002) 062, [arXiv:hep-ph/0204232](#)
- [12] **NNPDF** Collaboration, R.D. Ball et al., *A Determination of parton distributions with faithful uncertainty estimation*, Nucl.Phys. **B809** (2009) 1–63, [arXiv:0808.1231](#)
- [13] **NNPDF** Collaboration, R.D. Ball, L. Del Debbio, S. Forte, A. Guffanti, J.I. Latorre, et al., *A first unbiased global NLO determination of parton distributions and their uncertainties*, Nucl.Phys. **B838** (2010) 136–206, [arXiv:1002.4407](#)

- [14] R.G. Roberts, *The structure of the proton*, Cambridge Univ. Press, Cambridge, UK, 1990.
- [15] J. Pumplin, D.R. Stump, J. Huston, H.-L. Lai, P.M. Nadolsky, et al., *New generation of parton distributions with uncertainties from global QCD analysis*, JHEP **0207** (2002) 012, [arXiv:hep-ph/0201195](#)
- [16] **NNPDF** Collaboration, R.D. Ball, L. Del Debbio, S. Forte, et al., *Precision determination of electroweak parameters and the strange content of the proton from neutrino deep-inelastic scattering*, Nucl.Phys. **B823** (2009) 195, [arXiv:0906.1958](#)
- [17] **New Muon** Collaboration, M. Arneodo et al., *Accurate measurement of  $F_2^d/F_2^p$  and  $R_d-R_p$* , Nucl. Phys. **B487** (1997) 326, [arXiv:hep-ex/9611022](#)
- [18] **New Muon** Collaboration, M. Arneodo et al., *Measurement of the proton and deuteron structure functions,  $F_2^p$  and  $F_2^d$ , and of the ratio  $\sigma_L/\sigma_T$* , Nucl. Phys. **B483** (1997) 343, [arXiv:hep-ph/9610231](#)
- [19] L. W. Whitlow, E. M. Riordan, S. Dasu, S. Rock, and A. Bodek, *Precise measurements of the proton and deuteron structure functions from a global analysis of the SLAC deep inelastic electron scattering cross-sections*, Phys. Lett. **B282** (1992) 475482.
- [20] **BCDMS** Collaboration, A. C. Benvenuti et al., *A high statistics measurement of the proton structure functions  $f_2(x, q^2)$  and  $r$  from deep inelastic muon scattering at high  $q^2$* , Phys. Lett. **B223** (1989) 485.
- [21] **BCDMS** Collaboration, A. C. Benvenuti et al., *A high statistics measurement of the deuteron structure functions  $f_2(x, q^2)$  and  $r$  from deep inelastic muon scattering at high  $q^2$* , Phys. Lett. **B237** (1990) 592.
- [22] **CHORUS** Collaboration, G. Onengut et al., *Measurement of nucleon structure functions in neutrino scattering*, Phys. Lett. **B632** (2006) 6575.
- [23] **NuTeV** Collaboration, M. Goncharov et al., *Precise measurement of dimuon production cross-sections in  $\nu_\mu Fe$  and  $\bar{\nu}_\mu Fe$  deep inelastic scattering at the Tevatron*, Phys. Rev. **D64** (2001) 112006, [arXiv:hep-ex/0102049](#)
- [24] D. A. Mason, *Measurement of the strange - antistrange asymmetry at NLO in QCD from NuTeV dimuon data*, FERMILAB-THESIS-2006-01.