



PUBLISHED FOR SISSA BY SPRINGER

RECEIVED: May 5, 2014

REVISED: November 11, 2014

ACCEPTED: November 21, 2014

PUBLISHED: December 15, 2014

## PDF reweighting in the Hessian matrix approach

Hannu Paukkunen<sup>a,b</sup> and Pia Zurita<sup>c</sup>

<sup>a</sup>*Department of Physics, University of Jyväskylä,  
P.O. Box 35, FI-40014 University of Jyväskylä, Finland*

<sup>b</sup>*Helsinki Institute of Physics,  
P.O. Box 64, FI-00014 University of Helsinki, Finland*

<sup>c</sup>*Departamento de Física de Partículas and IGFAE, Universidade de Santiago de Compostela,  
E-15782 Galicia, Spain*

*E-mail:* [hannu.paukkunen@jyu.fi](mailto:hannu.paukkunen@jyu.fi), [pia.zurita@usc.es](mailto:pia.zurita@usc.es)

**ABSTRACT:** We introduce the Hessian reweighting of parton distribution functions (PDFs). Similarly to the better-known Bayesian methods, its purpose is to address the compatibility of new data and the quantitative modifications they induce within an existing set of PDFs. By construction, the method discussed here applies to the PDF fits that carried out a Hessian error analysis using a non-zero tolerance  $\Delta\chi^2$ . The principle is validated by considering a simple, transparent example. We are also able to establish an agreement with the Bayesian technique provided that the tolerance criterion is appropriately accounted for and that a purely exponential Bayesian likelihood is assumed. As a practical example, we discuss the inclusive jet production at the LHC.

**KEYWORDS:** Statistical Methods, Parton Model, QCD

**ARXIV EPRINT:** [1402.6623](https://arxiv.org/abs/1402.6623)

---

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>The Hessian method</b>	<b>2</b>
<b>3</b>	<b>The Hessian reweighting</b>	<b>4</b>
<b>4</b>	<b>Bayesian methods</b>	<b>6</b>
<b>5</b>	<b>Simple example</b>	<b>9</b>
<b>6</b>	<b>Equivalence of the Bayesian and Hessian reweighting</b>	<b>14</b>
<b>7</b>	<b>Inclusive jet production at the LHC</b>	<b>15</b>
<b>8</b>	<b>Conclusion</b>	<b>23</b>

---

## 1 Introduction

A large part of the present high-energy collider physics depends, in one way or another, on the knowledge of the parton distribution functions (PDFs). The use of PDFs leans on a cornerstone theorem of Quantum Chromo Dynamics (QCD), the collinear factorization [1, 2]. Although it can be formally proven only for the most simple cases, it is often assumed to work in general. Ultimately, it is the agreement with the experimental data that decides whether such an assumption is valid.

The PDFs are traditionally determined in global analyses [3–6] finding the parametrization that can optimally reproduce a variety of experimental data. This is a complex procedure requiring the ability to efficiently solve the parton evolution equations and to calculate higher-order QCD cross-sections. Nontrivial issues are also the actual way of finding the best fit and quantifying its uncertainties. Although various PDF parametrizations are publicly available for a general user, for a long period of time it was difficult for e.g. an experimental collaboration to understand what would be the implications of their measurements in the context of a global PDF fit. For example, although a given measurement would be known to be most sensitive to, say, the up quark distribution, some other data in a global fit may already provide the up quarks with more stringent constraints, and the real advantage of the measurement would be due to a sub-leading contribution, say, from the strange quarks.

The Bayesian reweighting technique, first introduced in [7] and later on elaborated by the Neural Network PDF (NNPDF) collaboration [8, 9], provides a way of addressing the consistency and quantitative effects of a new experimental evidence in terms of PDF fits. In

essence, the underlying probability distribution, represented in the NNPDF philosophy [6] by an ensemble ( $\sim 1000$ ) of PDF replicas, is updated by assigning each replica a certain weight based on the new data. This method has become an increasingly popular way to estimate the effects of e.g. new LHC measurements [10–16]. The drawback is that it has been proven to work only in conjunction with the NNPDF fits while the majority of the existing PDF fits use a rather different way of quantifying the PDFs and their uncertainties. Along with the best fit found by  $\chi^2$  minimization, they provide a collection ( $\sim 50$ ) of Hessian error sets [17] that quantify the neighborhood of the central fit within a certain confidence criterion  $\Delta\chi^2$ . An extension of the Bayesian reweighting technique to this particular case was suggested in [18], and has thereafter been used in some occasions [11, 19, 20]. However, a recent study [21] revealed clear deviations when comparing the results from reweighting to the ones obtained by a direct fit, indicating that the proposed generalization is not accurate.

Here we take a different strategy. Based on the ideas presented in ref. [22], our principal goal is to show how a general user can directly study the consistency and consequences of a new data set within an existing set of PDFs that comes with the Hessian error sets without having to lean on the Bayesian techniques. The method naturally incorporates the confidence criterion  $\Delta\chi^2$  defined in the original fit and, by considering a simple numerical example, we argue that it is perfectly compatible with a new fit. Our second objective is to understand how the procedure suggested here relates to the Bayesian reweighting and to the discrepancies found in [21]. We prove that the original Bayesian method, proposed in [7] and advocated recently in [23], is equivalent with the one introduced here once the  $\Delta\chi^2$  criterion is properly incorporated.

## 2 The Hessian method

The usual definition of an optimum correspondence between data and a set of PDFs  $f \equiv f(x, Q^2)$  that depends on certain fit parameters  $\{a\}$ , is the minimum of a  $\chi^2$ -function. In its most simple form, we can write it as

$$\chi^2\{a\} = \sum_k \left[ \frac{X_k^{\text{theory}}[f] - X_k^{\text{data}}}{\delta_k^{\text{data}}} \right]^2, \quad (2.1)$$

where  $X_k^{\text{theory}}[f]$  are the theory predictions depending on the PDFs. The corresponding experimental values are denoted by  $X_k^{\text{data}}$  and their uncertainty by  $\delta_k^{\text{data}}$ . Modifications to this definition are necessary if the experimental errors are correlated or some data sets are emphasized in the fit by assigning them an additional weight. In the Hessian approach to quantify the PDF errors [17], the behaviour of  $\chi^2$  around the best fit  $S_0$  is approximated by a second order polynomial in the space of fit parameters  $\{a\}$

$$\chi^2\{a\} \approx \chi_0^2 + \sum_{ij} \delta a_i H_{ij} \delta a_j, \quad (2.2)$$

where  $\delta a_j \equiv a_j - a_j^0$  are the excursions from the best-fit values and  $\chi_0^2$  is the minimum value of  $\chi^2$ . Being symmetric, the Hessian matrix  $H_{ij}$  has  $N_{\text{eig}}$  orthonormal eigenvectors

$\mathbf{v}^{(k)}$  and eigenvalues  $\epsilon_k$  satisfying

$$H_{ij}v_j^{(k)} = \epsilon_k v_i^{(k)}, \quad (2.3)$$

$$\sum_j v_j^{(k)} v_j^{(\ell)} = \sum_j v_k^{(j)} v_\ell^{(j)} = \delta_{k\ell}. \quad (2.4)$$

Defining a new set of variables as

$$z_k \equiv \sqrt{\epsilon_k} \sum_j v_j^{(k)} \delta a_j, \quad (2.5)$$

one easily finds that

$$\chi^2\{a\} \approx \chi_0^2 + \sum_i z_i^2. \quad (2.6)$$

That is, the transformation in eq. (2.5) diagonalizes the Hessian matrix. A criterion is needed to specify how much the term  $\sum_i z_i^2$  can grow while the corresponding PDFs still remain “acceptable”. Those PDF fits that employ the ideal choice  $\Delta\chi^2 = 1$  are usually limited a smaller set of data [24, 25], while the global fits prefer to take  $\Delta\chi^2 > 1$  [4, 5] to account for small inconsistencies among the data sets and to compensate for the parametrization bias [28]. It follows [17] that the corresponding uncertainty for a PDF-dependent quantity  $\mathcal{O} = \mathcal{O}[f]$  can be computed as

$$(\Delta\mathcal{O})^2 = \Delta\chi^2 \sum_k \left( \frac{\partial\mathcal{O}}{\partial z_k} \right)^2. \quad (2.7)$$

An essential feature of the Hessian approach is the introduction of the PDF error sets  $S_k^\pm$ , defined customarily (along with the best fit  $S_0$ ) in the  $z$ -space as

$$\begin{aligned} z(S_0) &= (0, 0, \dots, 0), \\ z(S_1^\pm) &= \pm\sqrt{\Delta\chi^2} (1, 0, \dots, 0), \\ z(S_2^\pm) &= \pm\sqrt{\Delta\chi^2} (0, 1, \dots, 0), \\ &\vdots \\ z(S_{N_{\text{eig}}}^\pm) &= \pm\sqrt{\Delta\chi^2} (0, 0, \dots, 1). \end{aligned} \quad (2.8)$$

Using these sets, one can evaluate the derivatives in eq. (2.7) by a linear approximation

$$\left( \frac{\partial\mathcal{O}}{\partial z_k} \right) \approx \frac{\mathcal{O}[S_k^+] - \mathcal{O}[S_k^-]}{2\sqrt{\Delta\chi^2}}, \quad (2.9)$$

such that

$$(\Delta\mathcal{O})^2 = \frac{1}{4} \sum_k (\mathcal{O}[S_k^+] - \mathcal{O}[S_k^-])^2. \quad (2.10)$$

This formula, or a generalization for asymmetric errors, provides an extremely simple and useful recipe for propagating the PDF-uncertainties to observables. Recently, it has become fashionable [4, 5] to abandon the fixed  $\Delta\chi^2$  tolerance and define the PDF error sets instead by a “dynamic tolerance”

$$z_i(S_k^\pm) \equiv \pm t_k^\pm \delta_{ik}, \quad (2.11)$$

which coincides with eq. (2.8) if the tolerance parameters  $t_k^\pm$  are equal to  $\sqrt{\Delta\chi^2}$ .

### 3 The Hessian reweighting

Let us now consider a new set of data  $\vec{y} = y_1, y_2, \dots, y_{N_{\text{data}}}$  with covariance matrix  $C$ . Our goal here is twofold: to find out whether these new data are consistent within the original set of PDFs and, if so, what would be the effect of incorporating them into the original analysis. In order to answer these questions, we consider a function  $\chi_{\text{new}}^2$  defined as

$$\chi_{\text{new}}^2 \equiv \chi_0^2 + \sum_k^{N_{\text{eig}}} z_k^2 + \sum_{i,j=1}^{N_{\text{data}}} (y_i[f] - y_i) C_{ij}^{-1} (y_j[f] - y_j), \quad (3.1)$$

where we have simply added the contribution of the new data on top of the “old”  $\chi^2$  in eq. (2.6). Using a similar linear approximation as earlier, we can estimate the theoretical values  $y_i[f]$  in arbitrary  $z$ -space coordinates by

$$y_i[f] \approx y_i[S_0] + \sum_{k=1}^{N_{\text{eig}}} \left. \frac{\partial y_i[S]}{\partial z_k} \right|_{S=S_0} z_k \approx y_i[S_0] + \sum_{k=1}^{N_{\text{eig}}} D_{ik} w_k, \quad (3.2)$$

where we have defined

$$D_{ik} \equiv \frac{y_i[S_k^+] - y_i[S_k^-]}{2} \quad (3.3)$$

$$w_k \equiv \frac{z_k}{\frac{1}{2}(t_k^+ + t_k^-)}. \quad (3.4)$$

Thus,  $\chi_{\text{new}}^2$  is a continuous, quadratic function of the parameters  $w_k$ , and its minimum is given simply by

$$\vec{w}^{\text{min}} = -\mathbf{B}^{-1} \vec{a}, \quad (3.5)$$

where the matrix  $\mathbf{B}$  and vector  $\vec{a}$  are

$$B_{kn} = \sum_{i,j} D_{ik} C_{ij}^{-1} D_{jn} + \left( \frac{t_k^+ + t_k^-}{2} \right)^2 \delta_{kn}, \quad (3.6)$$

$$a_k = \sum_{i,j} D_{ik} C_{ij}^{-1} (y_j[S_0] - y_j). \quad (3.7)$$

An important feature of the solution is the “penalty term”

$$P \equiv \sum_{k=1}^{N_{\text{eig}}} \left[ \left( \frac{t_k^+ + t_k^-}{2} \right) w_k^{\text{min}} \right]^2 \xrightarrow{t_k^\pm \rightarrow \sqrt{\Delta\chi^2}} \Delta\chi^2 \sum_{k=1}^{N_{\text{eig}}} (w_k^{\text{min}})^2, \quad (3.8)$$

which we can use to decide whether the new data set is consistent within the original PDFs. If an overall tolerance  $\Delta\chi^2$  was defined in the original fit, having  $P \ll \Delta\chi^2$  means that the new data could have been incorporated into the original fit without causing a conflict with the other data. On the other hand, if  $P \gtrsim \Delta\chi^2$  the new data appears to display significant tension with the considered set of PDFs. In the case that the original fit used a dynamic tolerance a simple interpretation like this is lost, and one can only check whether the new

$z$ -space coordinates remain within the tolerance parameters. However, even a large penalty term does not necessarily mean that the new data would be incompatible with the other data. A situation like this may arise if the new data probe unconstrained components of PDFs whose behaviour was fixed by hand. For example, some recent PDF fits [26] still assume  $s(x) \propto (u(x) + d(x))$  for the strange quark distribution and confronted with data sensitive to the strange quarks could lead to this kind of situation.

The components of the weight vector  $\vec{w}^{\min}$  also specify the set of PDFs  $f^{\text{new}}$  that corresponds to the new global minimum. They can be easily calculated by taking  $y_i = f(x, Q^2)$  in eq. (3.2). That is,

$$f^{\text{new}} \approx f_{S_0} + \sum_{k=1}^{N_{\text{eig}}} \left( \frac{f_{S_k^+} - f_{S_k^-}}{2} \right) w_k^{\min}. \quad (3.9)$$

The resulting new PDFs are linear combinations of the original ones — they have been “reweighted”. We note that the new PDFs constructed in this way still satisfy the necessary sum rules. For instance, as the original best fit  $S_0$  and the error sets  $S_k^\pm$  satisfy the momentum sum rule

$$\int_0^1 dx x \sum_f f_{S_0} = \int_0^1 dx x \sum_f f_{S_k^\pm} = 1, \quad (3.10)$$

then

$$\begin{aligned} \int_0^1 dx x \sum_f f^{\text{new}} &= \int_0^1 dx x \sum_f f_{S_0} + \sum_k \frac{w_k^{\min}}{2} \left[ \int_0^1 dx x \sum_f f_{S_k^+} - \int_0^1 dx x \sum_f f_{S_k^-} \right] \\ &= 1 + \sum_k \frac{w_k^{\min}}{2} [1 - 1] = 1. \end{aligned}$$

Due to the linearity of the parton evolution equations, also  $f^{\text{new}}$  satisfies them. Thus, the reweighted distributions comprise a proper set of PDFs which can be consistently utilized in perturbative QCD calculations. One can also construct the new PDF error sets. Indeed, eq. (3.1) can be rewritten as

$$\chi_{\text{new}}^2 = \chi_{\text{new}}^2 \Big|_{\vec{w}=\vec{w}^{\min}} + \sum_{ij} \delta w_i B_{ij} \delta w_j, \quad (3.11)$$

where  $\delta \vec{w} = \vec{w} - \vec{w}^{\min}$ , and the matrix  $\mathbf{B}$  takes the role of the Hessian matrix (compare to eq. (2.2)). This can be brought into a diagonal form by an analogue of the transformation in eq. (2.5), and the new error sets defined as

$$\delta w_i(\hat{S}_k^\pm) = \pm \hat{v}_i^{(k)} \sqrt{\frac{1}{\hat{\epsilon}_k}} \hat{t}_k^\pm, \quad (3.12)$$

where  $\hat{v}_i^{(k)}$  are the eigenvectors and  $\hat{\epsilon}_k$  the eigenvalues of the matrix  $\mathbf{B}$ . The original overall tolerance can be set easily by  $\hat{t}_k^\pm = \sqrt{\Delta \chi^2}$ .

**Non-linear extension of the Hessian reweighting.** The linear approximation of eq. (3.2) can be improved by including also quadratic terms of  $w_k$  as

$$y_i[S] = y_i[S_0] + \sum_{k=1}^{N_{\text{eig}}} \frac{1}{2} \left[ \frac{y_i[S_k^+] - y_i[S_0]}{t_k^+/t_k^-} - \frac{y_i[S_k^-] - y_i[S_0]}{t_k^-/t_k^+} \right] w_k \quad (3.13)$$

$$+ \sum_{k=1}^{N_{\text{eig}}} \frac{t_k^+ + t_k^-}{4} \left[ \frac{y_i[S_k^+] - y_i[S_0]}{t_k^+} + \frac{y_i[S_k^-] - y_i[S_0]}{t_k^-} \right] w_k^2$$

correcting for the possible non-linear behaviour. Using this in eq. (3.1),  $\chi_{\text{new}}^2$  becomes a quartic function of  $w_k$  whose minimum must be found by numerical methods. The corresponding PDFs can be computed by taking  $y_i = f(x, Q^2)$  in eq. (3.13). The matrix  $\mathbf{B}$  in eq. (3.11) gets replaced by

$$B_{kn}^{\text{non.lin}} = \sum_{i,j} \left( \frac{\partial y_i[f]}{\partial w_k} \right) C_{ij}^{-1} \left( \frac{\partial y_j[f]}{\partial w_n} \right) \quad (3.14)$$

$$+ \sum_{i,j} \left( \frac{\partial^2 y_i[f]}{\partial w_k \partial w_n} \right) C_{ij}^{-1} (y_j[f] - y_j) + \left( \frac{t_k^+ + t_k^-}{2} \right)^2 \delta_{kn},$$

where the partial derivatives read

$$\left( \frac{\partial y_i[f]}{\partial w_k} \right) = \frac{1}{2} \left[ \frac{y_i[S_k^+] - y_i[S_0]}{t_k^+/t_k^-} - \frac{y_i[S_k^-] - y_i[S_0]}{t_k^-/t_k^+} \right] \quad (3.15)$$

$$+ \frac{t_k^+ + t_k^-}{2} \left[ \frac{y_i[S_k^+] - y_i[S_0]}{t_k^+} + \frac{y_i[S_k^-] - y_i[S_0]}{t_k^-} \right] w_k$$

$$\left( \frac{\partial^2 y_i[f]}{\partial w_k \partial w_n} \right) = \frac{t_k^+ + t_k^-}{2} \left[ \frac{y_i[S_k^+] - y_i[S_0]}{t_k^+} + \frac{y_i[S_k^-] - y_i[S_0]}{t_k^-} \right] \delta_{kn}, \quad (3.16)$$

and are understood to be evaluated at the found minimum.

## 4 Bayesian methods

Given a large ensemble of PDFs  $f_k$ ,  $k = 1 \dots N_{\text{rep}}$ , such as those of the NNPDF collaboration [6], that represents the underlying probability distribution  $\mathcal{P}_{\text{old}}(f)$  of the PDFs, one can compute the expectation value  $\langle \mathcal{O} \rangle$  and variance  $\delta \langle \mathcal{O} \rangle$  for an observable  $\mathcal{O}$  as

$$\langle \mathcal{O} \rangle = \frac{1}{N_{\text{rep}}} \sum_{k=1}^{N_{\text{rep}}} \mathcal{O}[f_k], \quad (4.1)$$

$$\delta \langle \mathcal{O} \rangle = \sqrt{\frac{1}{N_{\text{rep}}} \sum_{k=1}^{N_{\text{rep}}} (\mathcal{O}[f_k] - \langle \mathcal{O} \rangle)^2}. \quad (4.2)$$

Using the laws of statistics, the initial probability distribution  $\mathcal{P}_{\text{old}}(f)$  can be updated to include also additional information contained in a new set of data  $\vec{y}$ , since, by the Bayes theorem,

$$\mathcal{P}_{\text{new}}(f) \propto \mathcal{P}(\vec{y}|f) \mathcal{P}_{\text{old}}(f), \quad (4.3)$$

where  $\mathcal{P}(\vec{y}|f)$  stands for the conditional probability (the likelihood function) for the new data, given a set of PDFs. It follows that the average value for any observable depending on the PDFs becomes a weighted average

$$\langle \mathcal{O} \rangle_{\text{new}} = \frac{1}{N_{\text{rep}}} \sum_{k=1}^{N_{\text{rep}}} \omega_k \mathcal{O}[f_k], \quad (4.4)$$

$$\delta \langle \mathcal{O} \rangle_{\text{new}} = \sqrt{\frac{1}{N_{\text{rep}}} \sum_{k=1}^{N_{\text{rep}}} \omega_k (\mathcal{O}[f_k] - \langle \mathcal{O} \rangle_{\text{new}})^2}, \quad (4.5)$$

where the weights  $\omega_k$  turn out to be proportional to the likelihood function  $\mathcal{P}(\vec{y}|f)$ . The question of how to choose the likelihood appropriately has been recently revisited in [23] but the conclusive answer, if it exists, remains to be given. Two options, corresponding to different choices of the likelihood, have been discussed in the literature. The one suggested originally by Giele and Keller (GK) [7] follows from taking  $\mathcal{P}(\vec{y}|f)d^n y$  as the probability to find the new data to be confined in a differential element  $d^n y$  around  $\vec{y}$  resulting in

$$\omega_k^{\text{GK}} = \frac{\exp[-\chi_k^2/2]}{(1/N_{\text{rep}}) \sum_{k=1}^{N_{\text{rep}}} \exp[-\chi_k^2/2]}, \quad (4.6)$$

where

$$\chi_k^2 = \sum_{i,j=1}^{N_{\text{data}}} (y_i[f_k] - y_i) C_{ij}^{-1} (y_j[f_k] - y_j). \quad (4.7)$$

The option advocated by the NNPDF collaboration derives from taking  $\mathcal{P}(\vec{y}|f)d\chi$  as the probability for the corresponding  $\chi \equiv \sqrt{\chi^2}$  to be confined in a differential volume  $d\chi$  around  $\chi$ , giving instead<sup>1</sup>

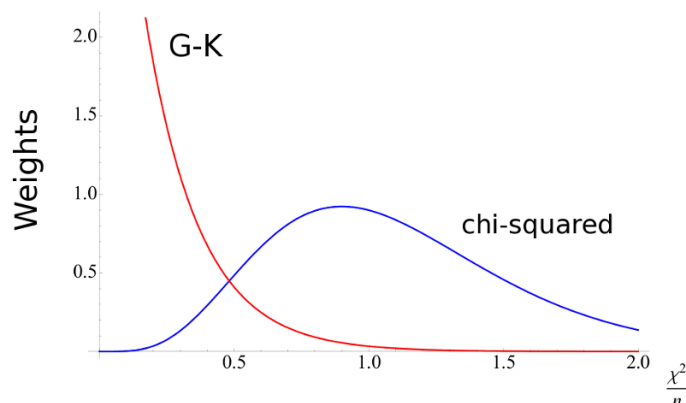
$$\omega_k^{\text{chi-squared}} = \frac{(\chi_k^2)^{(N_{\text{data}}-1)/2} \exp[-\chi_k^2/2]}{(1/N_{\text{rep}}) \sum_{k=1}^{N_{\text{rep}}} (\chi_k^2)^{(N_{\text{data}}-1)/2} \exp[-\chi_k^2/2]}, \quad (4.8)$$

which has been shown to be consistent with a direct fit in the NNPDF framework [8, 9]. It was pointed out in ref. [23] that the former weights contain more information on the new data than the latter ones, as a given data set uniquely determines the value of  $\chi^2$ , while a fixed  $\chi^2$  may correspond to various different data sets. The generic behaviour of these weights with respect to  $\chi^2$  per number of points for  $N_{\text{data}} = 10$  is shown in figure 1. While the GK weights are always higher for those replicas that give lower  $\chi^2$ , the NNPDF option obviously favors ones with  $\chi^2/N_{\text{data}} \approx 1$ . We note that the latter likelihood may lead to a following situation: if the value of  $\chi^2/N_{\text{data}}$  (computed with the expectation values of the observables or as an average of the individual  $\chi_k^2$ s) is less than unity before the reweighting, it can happen that the reweighting actually causes  $\chi^2$  for the new data to grow since the replicas with  $\chi^2/N_{\text{data}} \approx 1$  are favored. However, if new data are directly included into a PDF fit as in eq. (3.1), the value of  $\chi^2$  for the new data should only decrease.<sup>2</sup>

<sup>1</sup>We dub these weights  $\omega_k^{\text{chi-squared}}$  since their behaviour is very close to the usual  $\chi^2$  distribution.

<sup>2</sup>In PDF fits with extremely flexible fit functions, there is a danger that the PDFs that correspond to the minimum  $\chi^2$  unwantedly reproduce also random fluctuations of the data. Here, we assume that this is





**Figure 1.** Comparison of the likelihoods for the original derivation of the Bayesian re-weighting [7] (red) and the one proposed in [8, 9] (blue). In this plot the number of points is  $n = 10$ .

The large ensemble of PDFs required by the Bayesian approach can be constructed, in analogue to eq. (3.9), by

$$f_k \equiv f_{S_0} + \sum_i^{N_{\text{eig}}} \left( \frac{f_{S_i^+} - f_{S_i^-}}{2} \right) R_{ik}, \quad (4.9)$$

where the coefficients  $R_{ik}$  are random numbers drawn from a Gaussian distribution centered at zero and with variance one. An asymmetric version of eq. (4.9) to account for non-linearities was advocated in ref. [18]. Specifically, it was proposed that the replicas should be generated by

$$f_k^{\text{asym}} \equiv f_{S_0} + \sum_i^{N_{\text{eig}}} \left( f_{S_i^\pm} - f_{S_0} \right) |R_{ik}| \quad (4.10)$$

where  $f_{S_i^+}$  or  $f_{S_i^-}$  is chosen according to the sign of  $R_{ik}$ . However, in this case already before the reweighting the expectation values for the observables will not, in general, match those computed directly with the central set of the original fit. To accurately compare with the linear Hessian reweighting, we stick here to the symmetric prescription of eq. (4.9). As pointed out earlier, the replicas built in this way satisfy the PDF sum rules and the parton evolution equations. After computing the weights  $\omega_k$  for each replica, the reweighted PDFs can be written as

$$f_{\text{new}} = f_{S_0} + \sum_i^{N_{\text{eig}}} \left( \frac{f_{S_i^+} - f_{S_i^-}}{2} \right) \left( \frac{1}{N_{\text{rep}}} \sum_k \omega_k R_{ik} \right), \quad (4.11)$$

and, similarly to the Hessian case, one can calculate the “penalty” induced in the original fit by

$$P = \Delta\chi^2 \sum_i^{N_{\text{eig}}} \left( \frac{1}{N_{\text{rep}}} \sum_k \omega_k R_{ik} \right)^2. \quad (4.12)$$

---

not the case or that it has been taken into account e.g. by including suitable penalty terms in the original definition of  $\chi^2$  [28, 29].

We note that before reweighting  $w_k = 1$ , and the sums in the parenthesis above vanish since the mean of the random numbers  $R_{ik}$  is zero. Another useful indicator for the Bayesian methods is the effective number of replicas  $N_{\text{eff}}$ , defined as

$$N_{\text{eff}} \equiv \exp \left\{ \frac{1}{N_{\text{rep}}} \sum_{k=1}^{N_{\text{rep}}} \omega_k \log(N_{\text{rep}}/\omega_k) \right\}. \quad (4.13)$$

If a given replica  $f_k$  ends up having a small weight  $w_k \ll 1$ , it has a negligible effect in the new predictions computed by eqs. (4.4) and (4.5). The value of  $N_{\text{eff}}$  defined above serves as an estimate for such a “loss” of replicas. If  $N_{\text{eff}} \ll N_{\text{rep}}$ , the method becomes inefficient and is a sign that the new data contains too much new information or that it is incompatible with the previous data. Should this happen, also the penalty in eq. (4.12) is probably large.

## 5 Simple example

In this section, we will compare the different reweighting methods by invoking a rather simple, but illustrative example. We consider a function

$$g(x) = a_0 x^{a_1} (1-x)^{a_2} e^{x a_3} (1+x e^{a_4})^{a_5}, \quad (5.1)$$

which resembles a typical fit function used in PDF fits.<sup>3</sup> We proceed as follows:

- Construct a set of pseudodata (data set 1) for  $g(x)$ . The value of each data point  $y_k$  and its uncertainty  $\delta y_k$  are computed by

$$y_k = (1 + \alpha r_k) y_k^0, \quad \delta y_k = \alpha y_k^0$$

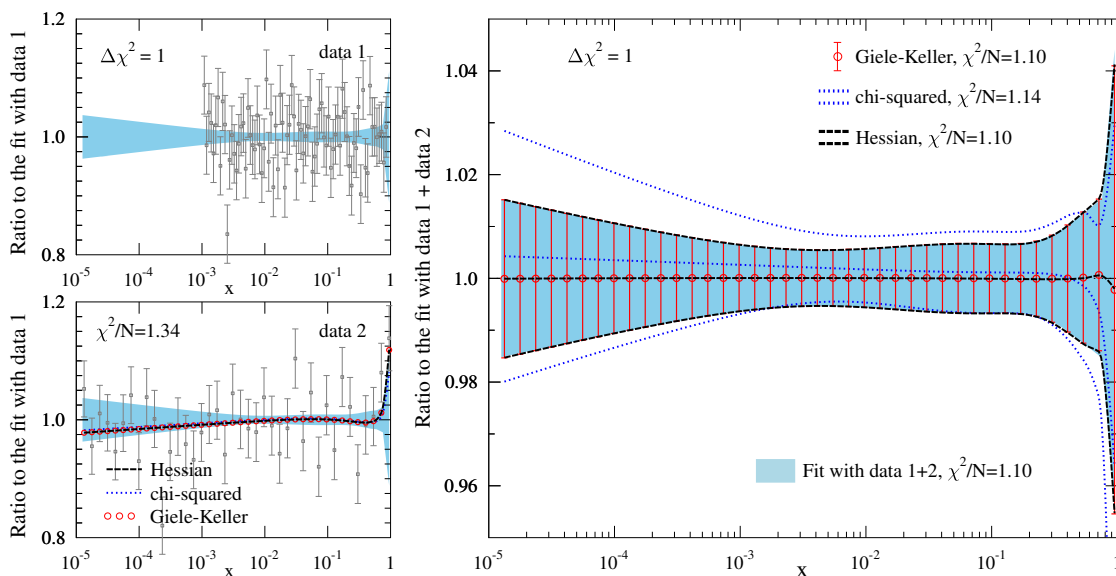
where  $y_k^0 = g(x)$  is evaluated with fixed parameters  $a_0^0 = 30$ ,  $a_1^0 = 0.5$ ,  $a_2^0 = 2.4$ ,  $a_3^0 = 4.3$ ,  $a_4 = 2.4$ , and  $a_5 = -3$ . The parameter  $\alpha = 0.05$  controls the size of the fluctuations generated by the Gaussian random numbers  $r_k$ .

- Perform a  $\chi^2$  fit with four free parameters  $a_0, a_1, a_2, a_3$  to these data, and construct the corresponding Hessian error sets using a certain  $\Delta\chi^2$  criteria.
- Construct a second set of pseudodata (data set 2) by the same procedure as in the case of data set 1 (using the same parameters), and apply the above-introduced reweighting techniques on these data.
- Perform a direct fit using both data sets and compare this “complete” result to the predictions given by the reweighting methods.

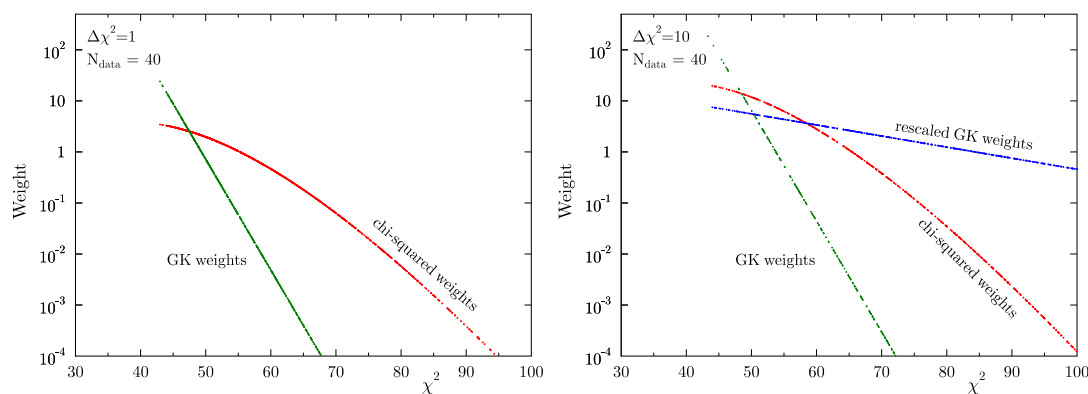
We begin by considering the ideal case  $\Delta\chi^2 = 1$  (that is, we take  $t_k^\pm = 1$ ), depicted in figure 2. We have chosen here an example in which the data set 2 (40 points) contains evidence from a region of  $x$  that the data set 1 (80 points) did not reach. In the case of

---

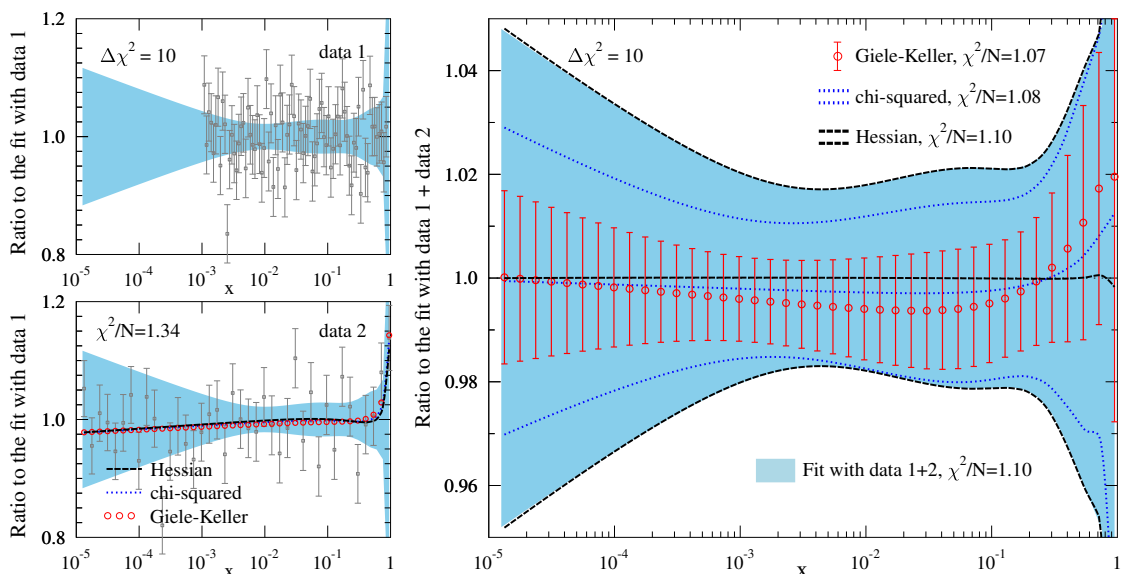
<sup>3</sup>In fact, the functional form and the parameter values we use here correspond to the gluons of CTEQ6 [27].



**Figure 2.** *Upper left-hand panel:* data set 1 normalized by the fit to these data. The light blue band shows the fit error defined by  $\Delta\chi^2 = 1$ . *Lower left-hand panel:* data set 2 normalized by the fit to the data set 1. The light blue band shows the original fit error and the black dashed line is the result using the Hessian reweighting on these data. Blue dotted line and the line marked by red circles are the corresponding results using the Bayesian reweighting with  $\omega_k^{\text{chi-squared}}$  and  $\omega_k^{\text{GK}}$ , respectively. The original value of  $\chi^2/N$  is indicated. *Right-hand panel:* the results of reweighting for the function  $g(x)$  normalized to the fit using data sets 1 and 2 (the light-blue band is the total  $\Delta\chi^2 = 1$  error band). The band enclosed by the black dashed lines corresponds the Hessian reweighting, and the one enclosed by the blue dotted lines to the Bayesian reweighting with  $\omega_k^{\text{chi-squared}}$ . The red circles with error bars are the results using Bayesian reweighting with  $\omega_k^{\text{GK}}$ . The resulting values of  $\chi^2/N$  for the data set 2 are indicated.



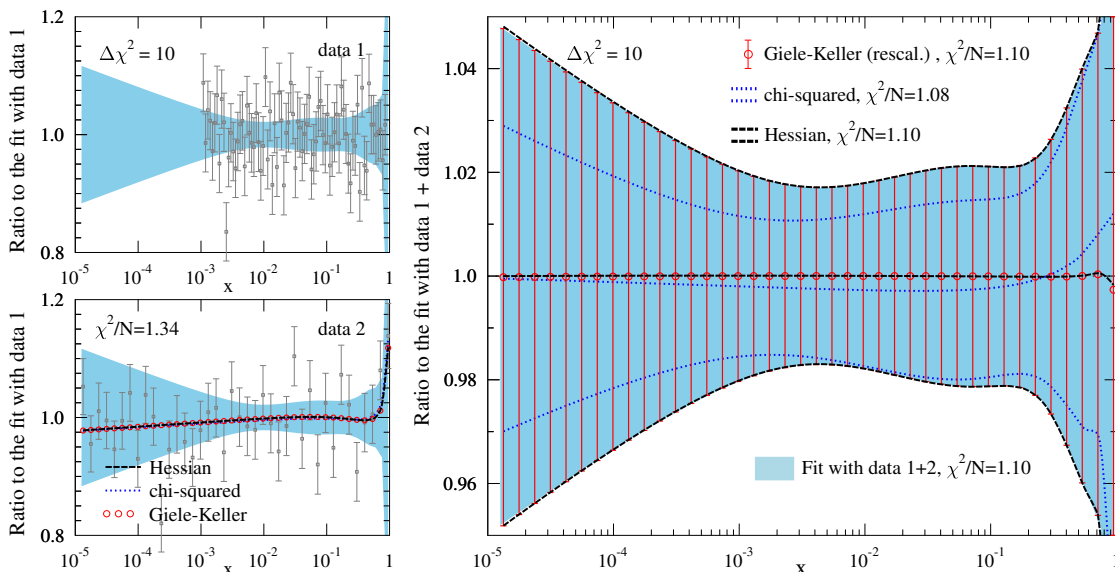
**Figure 3.** *Left-hand panel:* distribution of the Bayesian weights for  $\Delta\chi^2 = 1$  corresponding to figure 2. The line of green (red) points correspond to the GK (chi-squared) weights. *Right-hand panel:* as the left-hand panel but for  $\Delta\chi^2 = 10$  corresponding to figures 4 and 5. The additional line of blue dots corresponds to the rescaled GK weights.



**Figure 4.** As figure 2, but using  $\Delta\chi^2 = 10$ .

Bayesian methods, we have used a sufficiently large number ( $10^5$ ) of replicas to get rid of all numerical inaccuracies. To compare the methods as accurately as possible, the linear version of Hessian reweighting is used throughout this section. The results shown in figure 2 reveal that the Hessian reweighting and the Bayesian one with GK weights agree not only with each other, but also with the direct re-fit. The outcome with the chi-squared weights is similar but it does not fully coincide with the others. The reason for the similarity is that as the pseudodata we have used here are statistically consistent and we do not have too much freedom in the fit function, there are practically no replicas with  $\chi_k^2/N_{\text{data}} < 1$ , as can be appreciated from the left-hand panel of figure 3 where we plot the distribution of both Bayesian weights for this particular case. That is, the turnover for the chi-squared weights (see figure 1) does not play a role and it is very similar subset of replicas that mostly contributes in both cases.

Many global fits of PDFs use  $\Delta\chi^2$  clearly larger than unity. In our simple example here, a motivation for using  $\Delta\chi^2 > 1$  could be to compensate for the restricted functional form at small values of  $x$  where the data set 1 did not have constraints [28]. Thus, we repeat the exercise taking this time  $\Delta\chi^2 = 10$ . The results are shown in figure 4. While the Hessian reweighting can still accurately reproduce the re-fit, neither of the Bayesian methods coincides with them. In the case of GK reweighting the reason for the failure is that the likelihood function  $\mathcal{P}(\vec{y}|f)$  as such does not contain any information on  $\Delta\chi^2$ , although the distribution of replicas clearly depends on the value of  $\Delta\chi^2$ . As the spread among the replicas encoded by  $\Delta\chi^2 = 1$  is narrower than that covered by  $\Delta\chi^2 = 10$ , the new data appear more constraining than they actually are. This can also be verified from the distribution of Bayesian weights shown in the right-hand panel of figure 3. In comparison to the case with  $\Delta\chi^2 = 1$  the replicas with lowest  $\chi^2$  tend to get much higher



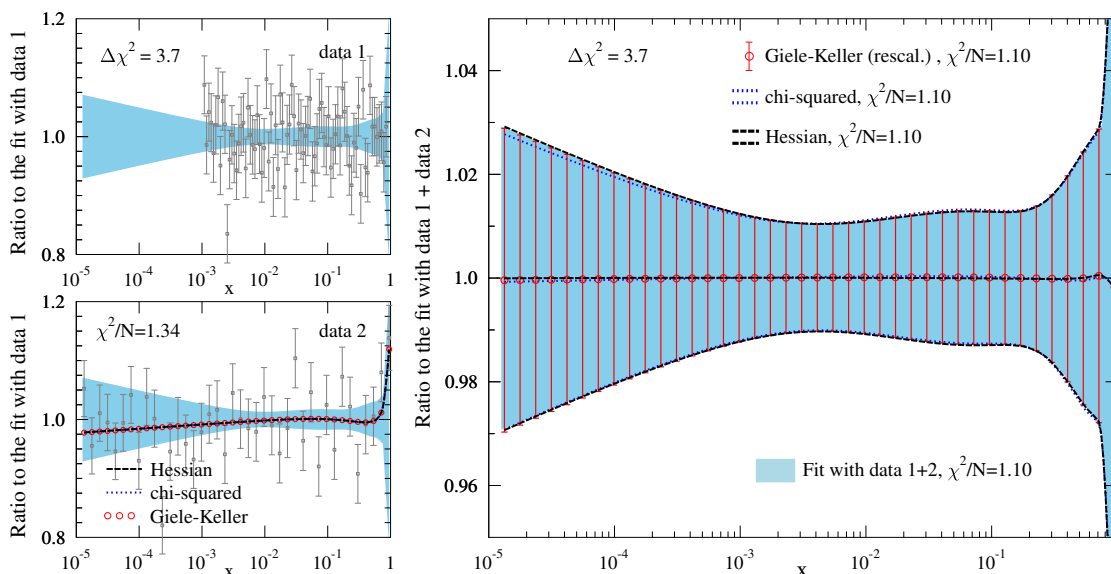
**Figure 5.** As figure 4, but rescaling the values of  $\chi^2$  by  $(\Delta\chi^2)^{-1}$  in the case of Bayesian reweighting with GK weights.

weight. However, the agreement encountered with  $\Delta\chi^2 = 1$  (figure 2) hints that it should be possible to generalize the Bayesian method with GK weights also to the case with  $\Delta\chi^2 > 1$ . The key point is to note that we could divide, for example, eq. (3.1) by  $\Delta\chi^2$ , and effectively use  $\Delta\chi^2 = 1$  thereafter. This observation instructs us to rescale the values of  $\chi_k^2$  in eqs. (4.6) as

$$\chi_k^2 \longrightarrow \frac{\chi_k^2}{\Delta\chi^2}, \tag{5.2}$$

when computing the weight for each replica. The corresponding results are shown in figure 5 which differs from figure 4 only in using the above mentioned rescaling when computing the Bayesian GK weights. Evidently, the agreement between the Bayesian method with GK weights, the Hessian reweighting, and the re-fit, is restored. The spectrum of rescaled GK weights is shown in the right-hand panel of figure 3 as well. The division of the individual  $\chi^2$  values by  $\Delta\chi^2$  makes the distribution of weights considerably flatter and it is actually a rather wide range of  $\chi^2$  values that still give a non-negligible contribution. We note that mathematically we would end up with the same reweighting result by narrowing the Gaussian distribution of random numbers instead of rescaling the individual  $\chi^2$  values. Indeed, replacing  $R_{ik} \rightarrow R_{ik}/\sqrt{\Delta\chi^2}$  in eq. (4.9) we would recover the same distribution of replicas that was obtained by using  $\Delta\chi^2 = 1$ . However, with  $\Delta\chi^2 = 10$  and rescaled GK weights the effective number of replicas is about twice as large as that with  $\Delta\chi^2 = 1$  ( $N_{\text{eff},\Delta\chi^2=1} \approx 19200$ ,  $N_{\text{eff},\Delta\chi^2=10} \approx 37500$ ). This makes the rescaling of  $\chi^2$  values a better option than narrowing the Gaussian distribution, although both procedures lead to the same result.

Our simple example here indicates that the Bayesian reweighting with chi-squared weights does not, in general, correspond to a direct re-fit although here they lead to a



**Figure 6.** As figure 5, but using  $\Delta\chi^2 = 3.7$ .

good approximation.<sup>4</sup> However, comparing figures 2 and 5 we notice that for  $\Delta\chi^2 = 1$  the results of chi-squared reweighting are a bit above the true result, but for  $\Delta\chi^2 = 10$  somewhat below. Similarly, for  $\Delta\chi^2 = 1$  the errorband is too wide, but for  $\Delta\chi^2 = 10$  too narrow. It therefore looks possible that by “tuning” the  $\Delta\chi^2$  appropriately the chi-squared reweighting could be made to coincide with the exact result. Since the rescaled GK weights appear to be the proper way to do the reweighting, the condition that the replicas that mostly contribute get the same weight in both cases is that the ratio

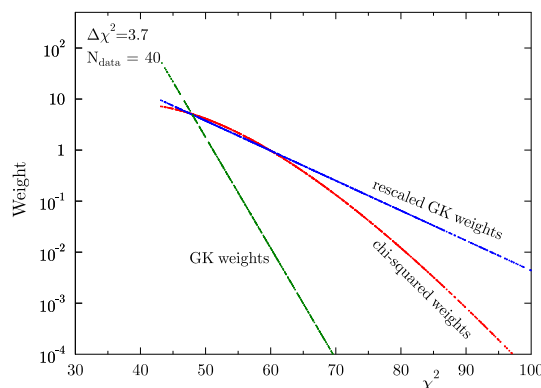
$$r \equiv \frac{(\chi^2)^{(N_{\text{data}}-1)/2} \exp[-\chi^2/2]}{\exp[-\chi^2/(2\Delta\chi^2)]} \quad (5.3)$$

is approximately constant. Requiring the derivative of this ratio to be zero, one finds

$$\Delta\chi^2 \Big|_{\frac{dr}{d\chi^2}=0} = \frac{\chi^2}{\chi^2 - (N_{\text{data}} - 1)}. \quad (5.4)$$

The individual replicas are always centered around the original fit and it thus makes sense to take the original value of  $\chi^2$  as a guideline in evaluating the value of  $\Delta\chi^2$  in the equation above. In the present example  $\chi^2/N \approx 1.34$  before the re-fit which corresponds to  $\Delta\chi^2 \approx 3.7$ . This particular value of  $\Delta\chi^2$  brings the chi-squared method to an excellent agreement with the direct re-fit as can be seen from figure 6. The corresponding distributions of the Bayesian weights are shown in figure 7 and, indeed, with  $\Delta\chi^2 \approx 3.7$  the chi-squared and rescaled GK weights go practically hand in hand in the region of  $\chi^2$  which mostly contributes to the final result. That is, if most of the replicas give  $\chi^2/N > 1$  and  $\Delta\chi^2$

<sup>4</sup>In a situation in which many replicas give  $\chi_k^2/N_{\text{data}} \ll 1$  the difference to the direct re-fit can become much larger.



**Figure 7.** As the right-hand panel of figure 3, but for  $\Delta\chi^2 = 3.7$ .

happens to be close to the value of eq. (5.4) evaluated with the original central  $\chi^2$ , both Bayesian reweightings may give approximately the same result.

In our simple example here, what mainly limits the accuracy of the reweighting is the precision of the original quadratic expansion of eq. (2.2). Indeed, the small mismatch between the results of reweighting and the real fit e.g. in figure 5 can be largely attributed to this approximation not being perfect. Although the same function  $g(x)$  was used in generating and fitting the pseudodata, we have checked that using a different fit function, e.g. a superposition of Chebyshev polynomials, would not alter our conclusions.

### 6 Equivalence of the Bayesian and Hessian reweighting

The close similarity of the results obtained using the (linear) Hessian reweighting and Bayesian one with the rescaled GK weights indicates that the two are actually one and the same. In this short section we will give a formal proof of this equivalence. From eq. (4.11) we see that the coordinates specifying the GK-reweighted PDFs in the eigenvector space are given by

$$w_k^{\text{GK}} = \frac{1}{N_{\text{rep}}} \sum_{\ell} \omega_{\ell} R_{k\ell} = \frac{1}{\mathcal{N}} \times \frac{1}{N_{\text{rep}}} \sum_{\ell} e^{-\frac{\chi_{\ell}^2}{2\Delta\chi^2}} R_{k\ell}, \tag{6.1}$$

where the denominator  $\mathcal{N}$  is

$$\mathcal{N} \equiv \frac{1}{N_{\text{rep}}} \sum_{\ell} e^{-\frac{\chi_{\ell}^2}{2\Delta\chi^2}}, \tag{6.2}$$

and we have applied the rescaling  $\chi_k^2 \rightarrow \chi_k^2/\Delta\chi^2$  as in eq. (5.2). Using the expression for  $\chi^2$  in eq. (4.7) and the linear approximation of eq. (3.2) with  $w_k = R_{k\ell}$ , we find

$$w_k^{\text{GK}} = \frac{1}{\mathcal{N}} \times \frac{1}{N_{\text{rep}}} \sum_{\ell} \exp \left[ -\frac{1}{2\Delta\chi^2} \sum_{n,m} R_{n\ell} \left( \sum_{i,j} D_{in} C_{ij}^{-1} D_{jm} \right) R_{m\ell} - \frac{1}{\Delta\chi^2} \sum_n a_n R_{n\ell} - \frac{1}{2\Delta\chi^2} \chi^2[f_{S_0}] \right] R_{k\ell}, \tag{6.3}$$

where  $\chi^2[f_{S_0}]$  is the value of  $\chi^2$  computed with the central set  $S_0$ , and the coefficients  $D_{ik}$  and  $a_k$  were defined in eqs. (3.3) and (3.7). In the limit of infinitely large  $N_{\text{rep}}$ , the sum over the replicas above can be replaced by an integral

$$\frac{1}{N_{\text{rep}}} \sum_{\ell=1}^{N_{\text{rep}}} \xrightarrow{N_{\text{rep}} \rightarrow \infty} (2\pi)^{-N_{\text{eig}}/2} \int_{-\infty}^{+\infty} d\vec{\mathbf{R}} \exp \left[ -\frac{1}{2} \vec{\mathbf{R}}^2 \right], \quad (6.4)$$

where the additional exponential stems from the probability distribution for the random numbers  $R_{m\ell}$  being Gaussian. Using this in eq. (6.3) above, we have

$$\begin{aligned} w_k^{\text{GK}} &= \frac{1}{\mathcal{N}} \times (2\pi)^{-N_{\text{eig}}/2} e^{-\frac{1}{2\Delta\chi^2} \chi^2[f_{S_0}]} \int_{-\infty}^{+\infty} d\vec{\mathbf{R}} e^{-\frac{1}{2\Delta\chi^2} \vec{\mathbf{R}}^T \mathbf{B} \vec{\mathbf{R}} - \frac{1}{\Delta\chi^2} \vec{\mathbf{a}}^T \vec{\mathbf{R}}} R_k \\ &= -\frac{\Delta\chi^2}{\mathcal{N}} \times (2\pi)^{-N_{\text{eig}}/2} e^{-\frac{1}{2\Delta\chi^2} \chi^2[f_{S_0}]} \frac{d}{da_k} \int_{-\infty}^{+\infty} d\vec{\mathbf{R}} e^{-\frac{1}{2\Delta\chi^2} \vec{\mathbf{R}}^T \mathbf{B} \vec{\mathbf{R}} - \frac{1}{\Delta\chi^2} \vec{\mathbf{a}}^T \vec{\mathbf{R}}} \\ &= -\frac{1}{\mathcal{N}} \times \sqrt{\frac{(\Delta\chi^2)^{N_{\text{eig}}}}{\det \mathbf{B}}} e^{-\frac{1}{2\Delta\chi^2} \chi^2[f_{S_0}] + \frac{1}{2\Delta\chi^2} \vec{\mathbf{a}}^T \mathbf{B}^{-1} \vec{\mathbf{a}}} (\mathbf{B}^{-1} \vec{\mathbf{a}})_k, \end{aligned}$$

where the elements of the matrix  $\mathbf{B}$  were given in eq. (3.6). The corresponding expression for the denominator  $\mathcal{N}$  is

$$\mathcal{N} = \sqrt{\frac{(\Delta\chi^2)^{N_{\text{eig}}}}{\det \mathbf{B}}} e^{-\frac{1}{2\Delta\chi^2} \chi^2[f_{S_0}] + \frac{1}{2\Delta\chi^2} \vec{\mathbf{a}}^T \mathbf{B}^{-1} \vec{\mathbf{a}}}. \quad (6.5)$$

Upon taking the ratio, the various prefactors cancel, and the coefficients  $w_k^{\text{GK}}$  reduce to

$$w_k^{\text{GK}} = -(\mathbf{B}^{-1} \vec{\mathbf{a}})_k, \quad (6.6)$$

which coincides with the eq. (3.5) specifying the coefficients for the linear Hessian reweighting. As shown in ref. [23], the weights  $w_k^{\text{chi-squared}}$  in eq. (4.8) emerge from  $w_k^{\text{GK}}$  by integrating over all possible data sets that give equal  $\chi^2$ . Being ‘‘contaminated’’ by such additional information readily explains their failure in the present context.

## 7 Inclusive jet production at the LHC

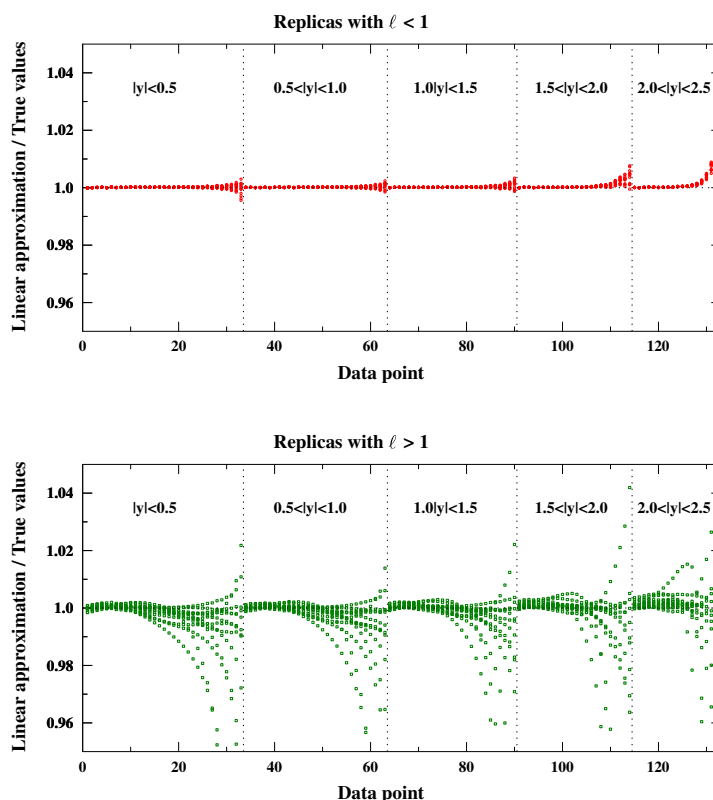
In this section, we apply the reweighting methods to the production of inclusive jets in proton+proton collisions at the LHC. In comparison to our example discussed earlier, a new source of non-linearity arises which could potentially decrease the accuracy of the linear reweighting. Namely, the quadratic PDF dependence of the proton+proton cross sections ( $\hat{\sigma}$  denotes the coefficient functions and jet definitions in general)

$$\sigma^{\text{PP}}[S] = f[S] \otimes \hat{\sigma} \otimes f[S], \quad (7.1)$$

may or may not be well approximated by

$$\sigma^{\text{PP}}[S_0 + \frac{1}{2} \sum_k (S_k^+ - S_k^-) w_k] \approx \sigma^{\text{PP}}[S_0] + \frac{1}{2} \sum_k (\sigma^{\text{PP}}[S_k^+] - \sigma^{\text{PP}}[S_k^-]) w_k, \quad (7.2)$$

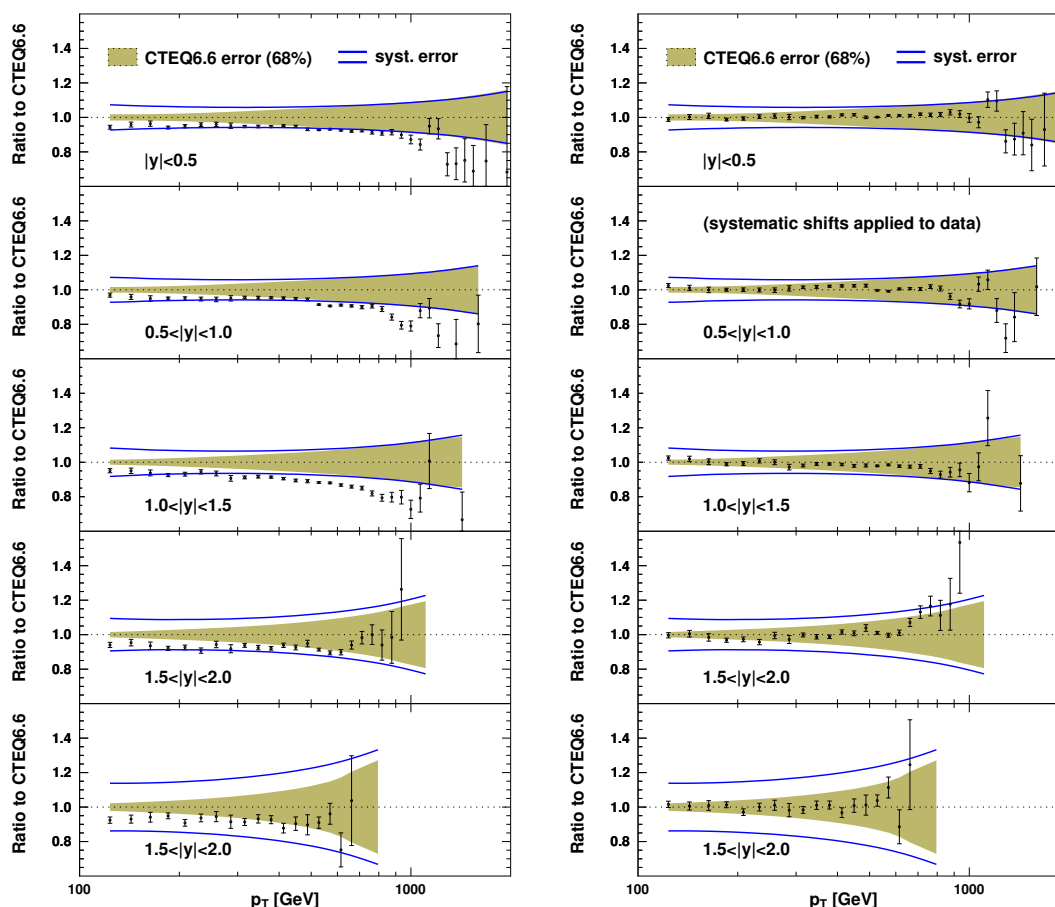




**Figure 8.** *Upper panel:* ratios between the jet cross sections computed by the linear approximation of eq. (7.2) divided by the “exact” value using a parametrization of the corresponding linear combination of PDFs inside FASTNLO. The results are plotted for 15 random replicas with the weight-vector length  $\ell < 1$ . The data points have been numbered with increasing transverse momentum. *Lower panel:* as the upper panel, but with  $\ell > 1$ .

which is the approximation one effectively makes when using eq. (3.2). In some other cases, such as deep-inelastic scattering (linear in PDFs), this would not be an issue whereas for the  $W$ -asymmetry, the non-linearities could be even more intricate. Here, we consider the recent CMS  $\sqrt{s} = 7$  TeV jet measurements [30] ( $N_{\text{data}} = 133$ ) for which a direct FASTNLO interface [31–33] is available to evaluate the cross sections at NLO accuracy. We use the CTEQ6.6 PDFs [34] whose error sets quantify the uncertainties within 90% confidence level the corresponding tolerance being  $\Delta\chi_{\text{CTEQ6.6}}^2(90\%) = 100$ . The renormalization scale  $\mu_r$  and factorization scale  $\mu_f$  were fixed to the jet transverse momentum as  $\mu_r = \mu_f = p_T/2$ , and the strong coupling was set to  $\alpha_s(M_Z) = 0.118$  at the Z boson pole.

Let us first discuss the adequacy of the approximation in eq. (7.2). To this end, we have prepared some random PDF replicas by eq. (4.9) separating the cases  $\ell = \sum_{i=1}^{N_{\text{eig}}} R_{ik}^2 < 1$  and  $\ell = \sum_{i=1}^{N_{\text{eig}}} R_{ik}^2 > 1$ . We compute the jet cross sections “exactly” (by constructing parametrizations corresponding to these PDF replicas and using them in FASTNLO computations) and, on the other hand, by the linear approximation of eq. (7.2). Typical results from such an exercise are shown in figure 8. First, if  $\ell < 1$ , the linear approximation proves



**Figure 9.** *Left-hand panels:* the CMS inclusive jet cross sections for the five rapidity intervals compared to the NLO calculation with CTEQ6.6 PDFs and taking  $\mu_f = \mu_r = p_T/2$ . The error bars in the data points show the statistical uncertainty, while the total systematic error is indicated by the blue lines. The colored bands show the CTEQ6.6 PDF uncertainty. *Right-hand panels:* as the left-hand panels, but after applying the systematic shifts.

rather accurate, the deviations being normally much less than couple of percents. In the latter case,  $\ell > 1$ , the linear approximation evidently breaks down. We conclude that if the reweighted PDFs end up sufficiently close to the original ones the linear approximation of eq. (7.2), and thereby the linear Hessian reweighting, should be rather accurate. In the case of Bayesian reweighting, replicas with  $\ell > 1$  often occur, and, from the lower panel of figure 8, we can consequently anticipate some differences whether the cross sections are evaluated using the replicas directly inside FASTNLO, or “on the fly” by the linear approximation of eq. (7.2).

We account for the correlated systematic errors by constructing a covariance matrix. To be specific, we compute the elements of the covariance matrix  $C$  by

$$C_{ij} = \delta_{ij} (\sigma_i^{\text{uncorr}})^2 + \sum_k \beta_i^k \beta_j^k, \quad (7.3)$$

where  $\sigma_i^{\text{uncorr}}$  is the uncorrelated error of data point  $i$ , and  $\beta_i^k$  denotes the absolute shift of this data point corresponding to 1-sigma deviation of the systematic parameter  $k$ . In addition to the luminosity, unfolding and jet energy scale uncertainties, we also treat the quoted uncertainty in the multiplicative non-perturbative corrections (underlying event, hadronization) as a correlated systematic error. The uncorrelated errors  $\sigma_i^{\text{uncorr}}$  include the statistical and 1% uncorrelated systematic uncertainty added in quadrature. Calculating the  $\chi^2$  using the covariance matrix  $C$  is equivalent to (see e.g. [35–37]) minimizing

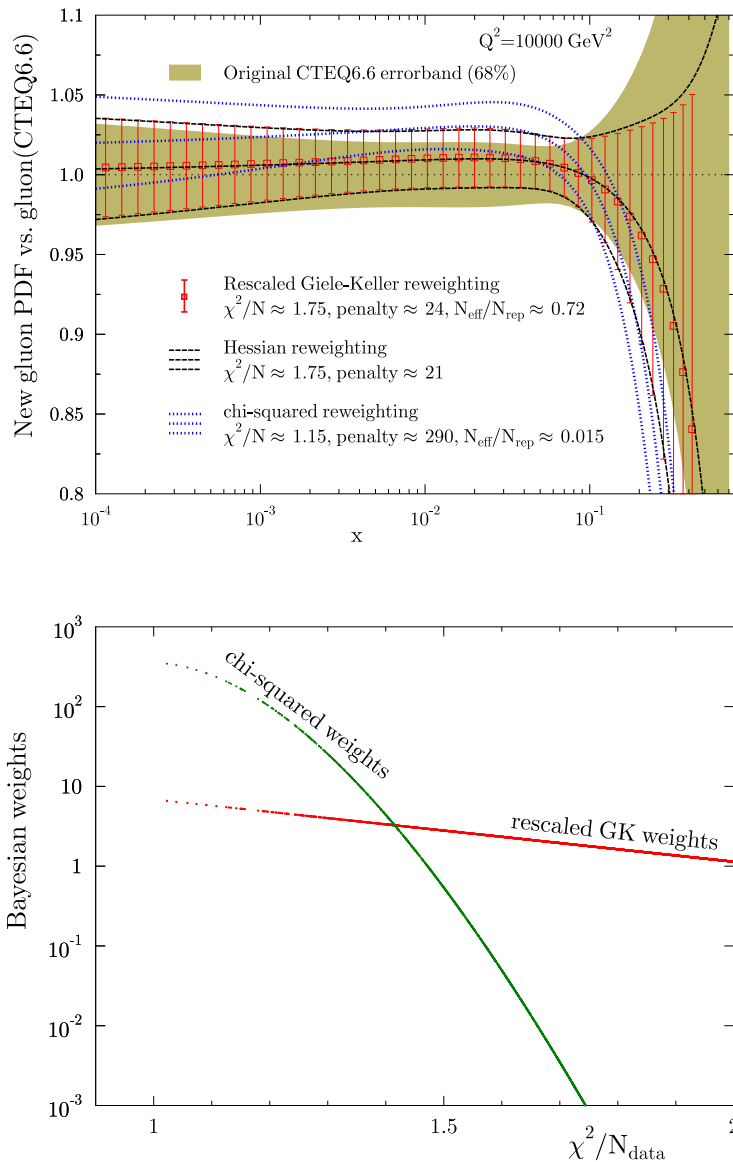
$$\chi^2 = \sum_i \left[ \frac{y_i^{\text{theory}} - y_i^{\text{data}} - \sum_k s_k \beta_i^k}{\sigma_i^{\text{uncorr}}} \right]^2 + \sum_k s_k^2, \quad (7.4)$$

with respect to the systematic parameters  $s_k$ . This occurs with the parameter values

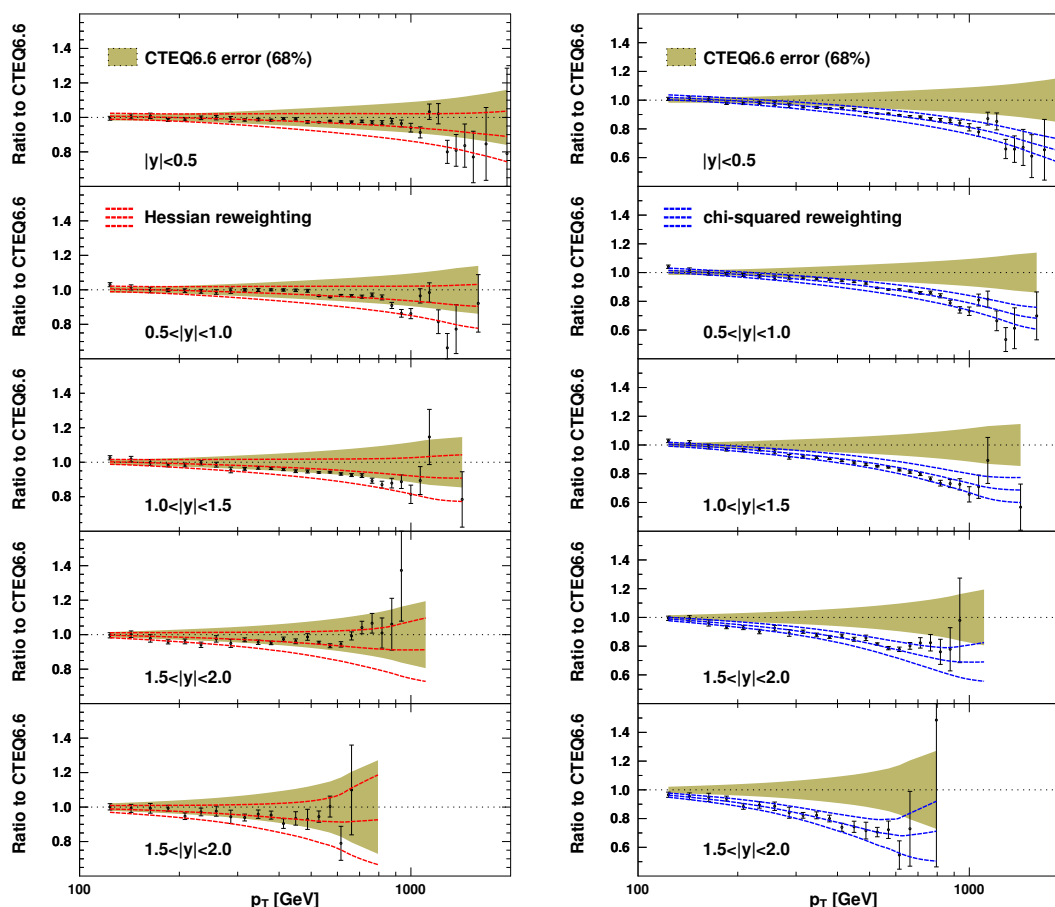
$$s_k^{\text{min}} = \sum_j \left[ \beta_j^k - \sum_{i,\ell,s} \beta_i^k C_{i\ell}^{-1} \beta_\ell^s \beta_j^s \right] \frac{y_j^{\text{theory}} - y_j^{\text{data}}}{(\sigma_j^{\text{uncorr}})^2}, \quad (7.5)$$

and  $-\sum_k s_k^{\text{min}} \beta_i^k$  is the net systematic shift for the data point  $y_i^{\text{data}}$ . Figure 9 presents a comparison between the CMS data and the NLO predictions, including the 68% PDF error bands (obtained downscaling the original 90% errors by 1/1.645 [11, 38]). The  $\chi^2$  value for the central CTEQ6.6 set is  $\chi^2/N_{\text{data}} \approx 2.1$  accounting for the data correlations as discussed above. The calculation appears to overpredict the experimental cross section by some 5% which, however, gets still easily hidden under the systematic shifts as demonstrated in figure 9.

It is a straightforward task to apply the reweighting methods on these jet data. For the needs of the Bayesian techniques we have generated  $10^4$  PDF replicas using eq. (4.9). As the data uncertainties correspond to one standard deviation, we have rescaled the random numbers in eq. (4.9) by  $R_{ik} \rightarrow R_{ik}/1.645$  which brings the PDF replicas to the 68% level as well. When computing the GK weights we divide the resulting values of  $\chi^2$  by  $\Delta\chi_{\text{CTEQ6.6}}^2(68\%) = 100/1.645^2 \approx 37$ , to appropriately modify the underlying likelihood. The chi-squared weights are always computed without rescaling the  $\chi^2$  values. We stress that the rescaling of CTEQ6.6 replicas from 90% to 68% confidence level affects only the outcome of chi-squared reweighting as, by construction, the central result of GK reweighting would remain the same by simply using the 90% replicas and rescaling the  $\chi^2$  values by  $\Delta\chi_{\text{CTEQ6.6}}^2(90\%)$ . The Hessian reweighting has been performed directly with the original 90% error sets rescaling the uncertainties by 1/1.645 only at the end. The resulting modifications in the gluon PDFs are presented in the upper panel of figure 10 (as the jet production is predominantly sensitive to the gluons, we find it reasonable to present the results here only in terms of gluon PDFs). From this plot we find that, as expected, the Hessian technique and the Bayesian method with rescaled GK weights are in good agreement also here. The small mismatch at large  $x$  originates mainly from eq. (7.2) not being exact. As the reweighting penalty  $P \approx 21$  is less than  $\Delta\chi_{\text{CTEQ6.6}}^2(68\%)$ , we conclude that these data could have been added to the CTEQ6.6 fit without causing a significant disagreement with the original data. In this sense these data are compatible with the



**Figure 10.** *Upper panel:* the gluon PDFs at  $Q^2 = 10000 \text{ GeV}^2$  after reweighting using the correlated errors. The red points with error bars correspond to the results using the Bayesian reweighting with rescaled GK weights, and the blue dotted lines mark the corresponding result with chi-squared weights. The result from Hessian reweighting is indicated by the black dashed lines and the colored band is the original CTEQ6.6 uncertainty. All results are normalized to the central set of CTEQ6.6. *Lower panel:* the distribution of Bayesian weights.



**Figure 11.** *Left-hand panels:* the CMS inclusive jet cross sections compared to the predictions after the Hessian reweighting and applying the systematic shifts. All values have been normalized to the central prediction of CTEQ6.6 and only the statistical data errors are shown. *Right-hand panels:* as the left-hand panels, but with the Bayesian reweighting with chi-squared weights.

CTEQ6.6 PDFs despite the largish  $\chi^2/N_{\text{data}} \approx 1.75$  which could hint, however, that some tension between the Tevatron Run-1 jet data [39, 40] (used in CTEQ6.6 fit to constrain large- $x$  gluons) and these new LHC measurements exist.<sup>5</sup> The line of red points in figure 10 shows the spectrum of rescaled GK weights. For dividing the values of  $\chi^2$  by the tolerance  $\Delta\chi^2_{\text{CTEQ6.6}}(68\%) \approx 37$ , this distribution is rather flat and the final result is affected by replicas with a wide range of  $\chi^2$  explaining also the rather large number of effective replicas,  $N_{\text{eff}}/N_{\text{rep}} \approx 0.72$ . The chi-squared weights give rise to effects which are qualitatively alike but quantitatively much larger. The agreement with the new jet data is admittedly better,  $\chi^2/N_{\text{data}} \approx 1.15$  (the green points in figure 10 indeed demonstrate that the chi-squared weights assign clearly larger weights to replicas with  $\chi^2/N_{\text{data}} < 1.4$  than the rescaled GK weights), but at the cost of increasing the original  $\chi^2$  by  $P \approx 290$  which is way beyond the

<sup>5</sup>Indeed, the latest CTEQ fit [35] has abandoned the Tevatron Run-1 jet data as they do not completely agree with the Run-2 data. See ref. [5] for a review and ref. [41] for more discussions.

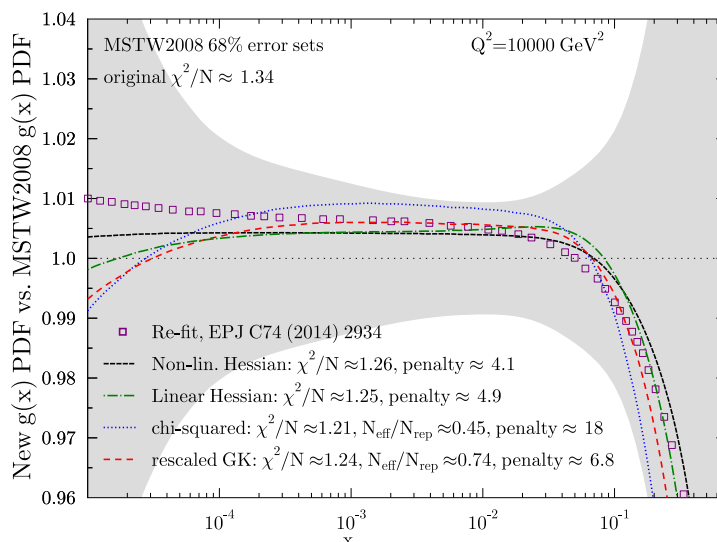
CTEQ6.6 tolerance (both 68% and 90%). The surviving number of replicas  $N_{\text{eff}}$  is also very low. That is, one could conclude that the jet data considered here were not in agreement with the CTEQ6.6 PDFs. However, as we have demonstrated in the previous sections, the result is misleading since the chi-squared weights do not appropriately account for the other data that were originally included in the global CTEQ6.6 fit. From eq. (5.4) we can estimate the value of  $\Delta\chi^2$  that would have brought the results of chi-squared reweighting close to the correct ones: using the original central value  $\chi^2/N_{\text{data}} \approx 2.1$  and  $N_{\text{data}} = 133$  we obtain  $\Delta\chi^2 \approx 1.9$ . This is much less than  $\Delta\chi^2_{\text{CTEQ6.6}}(68\%) = 37$  and concretely explains why the chi-squared weights lead to a result which is so far from the correct one.

The new predictions for the cross-section are shown in figure 11. The systematic shifts have been applied to the data and they depend significantly on which method of reweighting was used. The Hessian reweighting has caused a mild downward shift on the cross section which was to be expected given that the original central predictions somewhat overshoot the data (see figure 9). In the case of Bayesian reweighting with chi-squared weights the induced changes with respect to the original CTEQ6.6 predictions are way too large.

**Reweighting of MSTW2008.** In the previous example, we discussed PDF reweighting in the case of a fixed global tolerance  $\Delta\chi^2$  assuming the validity of the quadratic approximation in eq. (3.1). It is, however, well known that such quadratic profile is never perfect, and for the fit parameters that are not very well constrained severe deviations from the ideal quadratic behaviour may occur (see e.g. figures 5-6 in [5]). To account for such imperfections and improve the linear approximations in estimating the observables in the space of eigenvectors, the non-linear extension of the Hessian reweighting should provide some improvement. As an example, we consider the MSTW2008 PDFs [5] using the same CMS jet data as earlier. In fact, the effect of incorporating the  $\sqrt{s} = 7$  TeV CMS and ATLAS [42] inclusive jet measurements into the MSTW2008 framework was studied recently via a direct re-fit [21], but due to the large systematic uncertainties of the ATLAS data, the fit was mainly driven by the same CMS measurements we have discussed here. Thus, in what follows, we will contrast our results to what was obtained in that analysis.

The MSTW2008 package provides two separate sets of error PDFs, the 68% and 90% confidence-level sets, of which we will use the former one. As the numerical values for the determined tolerances  $t_k^\pm$  and the corresponding increases  $T_k^\pm$  of the global  $\chi^2$  are given, we can account also for the non-quadratic behaviour of the original  $\chi^2$  function in the space of eigenvectors. Specifically, in addition to using eq. (3.13) in evaluating the observables, we make the following substitution in eq. (3.1):

$$\begin{aligned} \sum_k^{N_{\text{eig}}} z_k^2 &\longrightarrow \sum_k^{N_{\text{eig}}} a_k z_k^2 + b_k z_k^3, \\ a_k &= \frac{t_k^-(T_k^+/t_k^+)^2 - t_k^+(T_k^-/t_k^-)^2}{t_k^+ + t_k^-}, \\ b_k &= \frac{(T_k^+/t_k^+)^2 - (T_k^-/t_k^-)^2}{t_k^+ + t_k^-}. \end{aligned} \tag{7.6}$$

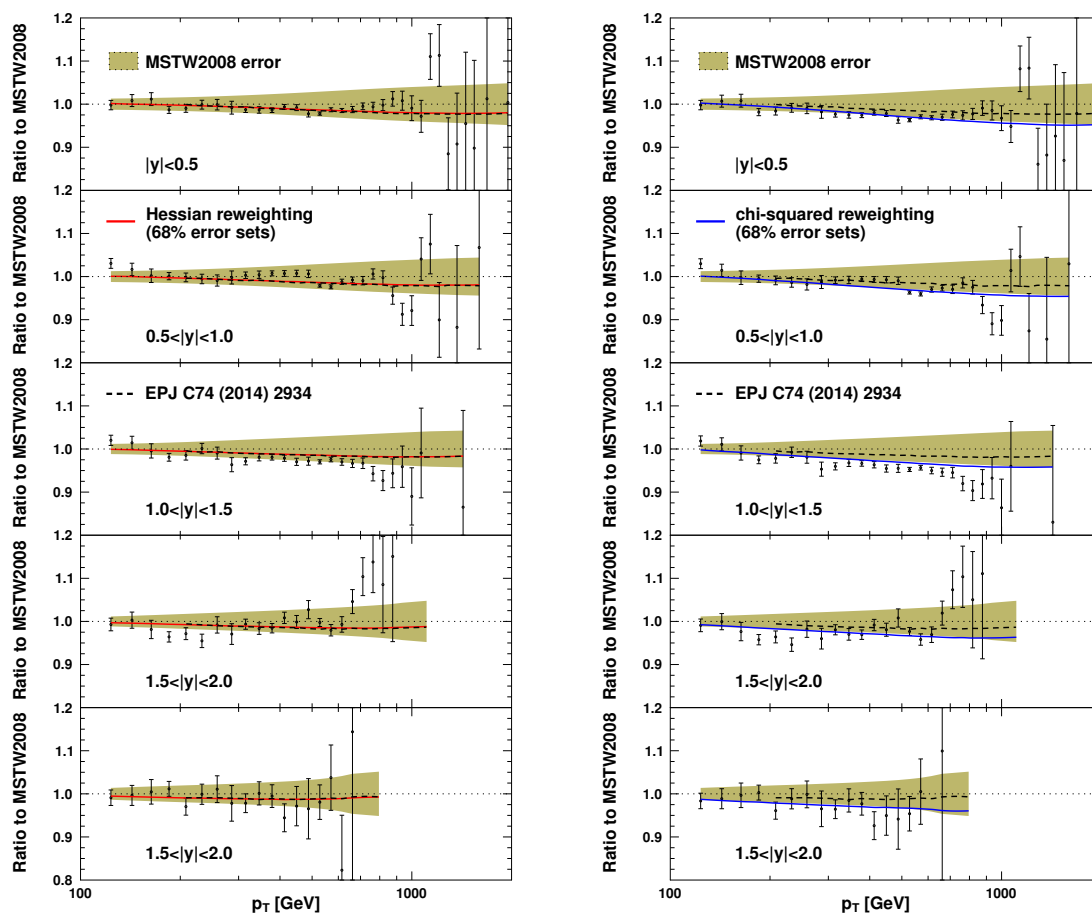


**Figure 12.** The reweighted gluon distributions at  $Q^2 = 10000 \text{ GeV}^2$  normalized by the central MSTW2008 gluons. The solid black line is the result from the non-linear Hessian reweighting, and the green dashed-dotted line corresponds to the linear Hessian reweighting. Blue dotted line is the prediction of Bayesian reweighting with chi-squared weights, and red dashed line results from using the rescaled GK weights. The shaded band is the original MSTW2008 uncertainty and the purple squares mark the result obtained in ref. [21].

We set  $\alpha_s(M_Z^2) = 0.1202$ , and use  $\mu_r = \mu_f = p_T$ . As it is not possible (with the available information) to consistently implement the dynamic tolerance after the reweighting, we discuss here only the resulting central values.

The gluon PDF after the non-linear Hessian reweighting is shown in figure 12. For comparison, the figure includes also the outcome from the linear Hessian reweighting, Bayesian reweighting with the chi-squared weights, Bayesian reweighting with GK weights rescaled by the average tolerance,  $\Delta\chi_{\text{MSTW2008}}^2(68\%) \approx 10$ , and the result obtained by a re-fit in ref. [21]. Apart from the Bayesian predictions with chi-squared weights the results of different prescriptions are more or less consistent with each other within  $10^{-4} < x < 10^{-1}$ , and also close to the “exact” result.<sup>6</sup> The differences between the linear and non-linear prescriptions are clearly more important at  $x < 10^{-4}$  and  $x > 10^{-1}$ , where the original PDF uncertainties become large. The chi-squared weights predict again modifications for the gluon PDFs which are similar to the correct ones but somewhat too pronounced (the estimated penalty stands out from what is obtained with the other methods). The reason can be understood by estimating from eq. (5.4) which value of  $\Delta\chi^2$  would have given consistent results: now  $\chi^2/N_{\text{data}} \approx 1.34$  which translates to  $\Delta\chi^2 \approx 3.9$ . While close, this is still less than  $\Delta\chi_{\text{MSTW2008}}^2(68\%) \approx 10$  and explains why the result is different from the direct re-fit.

<sup>6</sup>Note that in ref. [21] the open fit parameters were not restricted to those used in generating the 20 eigenvector sets. Thus, it is natural to expect deviations in comparison to the results of reweighting.



**Figure 13.** *Left-hand panels:* the jet cross sections after reweighting MSTW2008 PDFs. The results of the non-linear Hessian reweighting are shown by the red lines and the results of ref. [21] are the black dashed lines. All values have been normalized to the central predictions of MSTW2008. Systematic shifts have been applied to the data and only the statistical data errors are shown. *Right-hand panels:* as the left-hand panels, but with the Bayesian reweighting with chi-squared weights.

The predicted jet cross sections after the reweighting are shown in figure 13, normalized to the central MSTW2008 values. The results of Hessian reweighting clearly stay within the original uncertainties and appear to agree with the results of ref. [21]. The predictions of the Bayesian reweighting with chi-squared weights are, however, at the lower limit of the original error bands and disagree with the true re-fit.

## 8 Conclusion

We have discussed how to test the consistency and estimate the effects that a new set of experimental measurements have within an existing set of Hessian PDFs with non-zero tolerance  $\Delta\chi^2$ . To this end, the Hessian reweighting, was introduced as an alternative technique to the prevailing Bayesian methods. While the Hessian reweighting is straight-



forwardly derived considering a new set of data in a global  $\chi^2$  fit, the Bayesian methods are outwardly distinct, based on statistical inference. We compared the different approaches to a direct re-fit through a simple example verifying the adequacy of the Hessian method. In the case of the Bayesian procedure an agreement with a new fit was also established, but only after including the  $\Delta\chi^2$  criterion properly into the Bayesian likelihood function which — as we mathematically justified — must be a pure exponential as originally proposed by Giele and Keller. The conditions under which another commonly used (but in this case inadequate) likelihood function gives consistent results was also discussed. The inclusive jet production at the LHC was considered as an additional example. At first, our findings may appear to be in contrast to the works of NNPDF collaboration in which a different functional form for the likelihood is derived. However, as the NNPDF methodology for fitting PDFs is far more involved than the simple  $\chi^2$  minimization considered here, it is possible that a different functional form applies. For example, the NNPDF fits involve extremely flexible fit functions and to avoid fitting random fluctuations the data are divided into “training“ and “control” sets. While the actual  $\chi^2$  minimization is done only for the former subset of data, the latter one is used to decide when the minimization process should be stopped. The outcome of a direct  $\chi^2$  minimization with no stopping criteria would generally be different if the same functional form would be used. However, for the moment it is still unclear what exactly causes the need for using a different Bayesian likelihood.

The two types of methods discussed here have their pros and cons: while the Hessian procedure requires evaluating the observables only with the central and error sets (typically around 50 sets in total), in the Bayesian method one needs to deal with a much larger ensemble of PDFs (around  $10^3$  to find well-converging results). On the other hand, the Hessian reweighting is procedurally a bit more involved requiring e.g. numeric linear algebra, while the Bayesian technique is simpler. The reliability of the both methods depends basically on the accuracy of the quadratic approximation around the minimum  $\chi^2$  made in the original PDF fit and on the adequacy of the linear approximations that one makes. In the case of Hessian reweighting one can improve on these approximations by invoking non-linear corrections as we explained. However, when the PDF fits are updated details like the form of the fit function, data sets included, or value of  $\alpha_s(M_Z^2)$  are often altered. Such modifications cannot be easily accounted for by the reweighting procedures and in this sense the reweighting is always an approximation to a real fit. In any case, it is definitely useful in checking whether the new data appear consistent and which components of the PDFs are bound to undergo a change and how.

## Acknowledgments

We thank J. Rojo and S. Forte for useful discussions on the PDF reweighting and N. Armesto and C. Salgado for their comments regarding the manuscript. We also want to acknowledge the Referee of our paper whose suggestions led to significant improvements especially in section 5 concerning the discussion on the conditions that make the chi-squared reweighting to closely mimic a direct refit. During this work H.P. was supported by the Academy

of Finland, Project No. 133005. P.Z. was supported by European Research Council grant HotLHC ERC-2011-StG-279579; by Ministerio de Ciencia e Innovación of Spain under the Consolider-Ingenio 2010 Programme CPAN (CSD2007-00042); by Xunta de Galicia; and by FEDER.

**Open Access.** This article is distributed under the terms of the Creative Commons Attribution License ([CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/)), which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited.

## References

- [1] J.C. Collins, D.E. Soper and G.F. Sterman, *Factorization of Hard Processes in QCD*, *Adv. Ser. Direct. High Energy Phys.* **5** (1988) 1 [[hep-ph/0409313](#)] [[INSPIRE](#)].
- [2] J.C. Collins, D.E. Soper and G.F. Sterman, *Factorization for Short Distance Hadron - Hadron Scattering*, *Nucl. Phys.* **B 261** (1985) 104 [[INSPIRE](#)].
- [3] 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:1301.6754](#)] [[INSPIRE](#)].
- [4] H.-L. Lai, M. Guzzi, J. Huston, Z. Li, P.M. Nadolsky et al., *New parton distributions for collider physics*, *Phys. Rev.* **D 82** (2010) 074024 [[arXiv:1007.2241](#)] [[INSPIRE](#)].
- [5] A.D. Martin, W.J. Stirling, R.S. Thorne and G. Watt, *Parton distributions for the LHC*, *Eur. Phys. J.* **C 63** (2009) 189 [[arXiv:0901.0002](#)] [[INSPIRE](#)].
- [6] R.D. Ball, V. Bertone, S. Carrazza, C.S. Deans, L. Del Debbio et al., *Parton distributions with LHC data*, *Nucl. Phys.* **B 867** (2013) 244 [[arXiv:1207.1303](#)] [[INSPIRE](#)].
- [7] W.T. Giele and S. Keller, *Implications of hadron collider observables on parton distribution function uncertainties*, *Phys. Rev.* **D 58** (1998) 094023 [[hep-ph/9803393](#)] [[INSPIRE](#)].
- [8] NNPDF collaboration, R.D. Ball et al., *Reweighting NNPDFs: the W lepton asymmetry*, *Nucl. Phys.* **B 849** (2011) 112 [*Erratum ibid.* **B 854** (2012) 926] [*Erratum ibid.* **B 855** (2012) 927] [[arXiv:1012.0836](#)] [[INSPIRE](#)].
- [9] R.D. Ball, V. Bertone, F. Cerutti, L. Del Debbio, S. Forte et al., *Reweighting and Unweighting of Parton Distributions and the LHC W lepton asymmetry data*, *Nucl. Phys.* **B 855** (2012) 608 [[arXiv:1108.1758](#)] [[INSPIRE](#)].
- [10] CMS collaboration, *Measurement of the muon charge asymmetry in inclusive  $pp \rightarrow W + X$  production at  $\sqrt{s} = 7$  TeV and an improved determination of light parton distribution functions*, *Phys. Rev.* **D 90** (2014) 032004 [[arXiv:1312.6283](#)] [[INSPIRE](#)].
- [11] R. Gauld, *Feasibility of top quark measurements at LHCb and constraints on the large-x gluon PDF*, *JHEP* **02** (2014) 126 [[arXiv:1311.1810](#)] [[INSPIRE](#)].
- [12] M. Czakon, M.L. Mangano, A. Mitov and J. Rojo, *Constraints on the gluon PDF from top quark pair production at hadron colliders*, *JHEP* **07** (2013) 167 [[arXiv:1303.7215](#)] [[INSPIRE](#)].
- [13] L. Carminati, G. Costa, D. D'Enterria, I. Koletsou, G. Marchiori et al., *Sensitivity of the LHC isolated-gamma+jet data to the parton distribution functions of the proton*, *Europhys. Lett.* **101** (2013) 61002 [[arXiv:1212.5511](#)] [[INSPIRE](#)].
- [14] M. Beneke, P. Falgari, S. Klein, J. Piclum, C. Schwinn et al., *Inclusive Top-Pair Production Phenomenology with TOPIXs*, *JHEP* **07** (2012) 194 [[arXiv:1206.2454](#)] [[INSPIRE](#)].

- [15] D. d’Enterria and J. Rojo, *Quantitative constraints on the gluon distribution function in the proton from collider isolated-photon data*, *Nucl. Phys. B* **860** (2012) 311 [[arXiv:1202.1762](#)] [[INSPIRE](#)].
- [16] NNPDF collaboration, R.D. Ball et al., *Parton distributions with QED corrections*, *Nucl. Phys. B* **877** (2013) 290 [[arXiv:1308.0598](#)] [[INSPIRE](#)].
- [17] J. Pumplin, D. Stump, R. Brock, D. Casey, J. Huston et al., *Uncertainties of predictions from parton distribution functions. 2. The Hessian method*, *Phys. Rev. D* **65** (2001) 014013 [[hep-ph/0101032](#)] [[INSPIRE](#)].
- [18] G. Watt and R.S. Thorne, *Study of Monte Carlo approach to experimental uncertainty propagation with MSTW 2008 PDFs*, *JHEP* **08** (2012) 052 [[arXiv:1205.4024](#)] [[INSPIRE](#)].
- [19] N. Sato, *Threshold resummation in direct photon production and its implications on the large- $x$  gluon PDF*, [arXiv:1309.7995](#) [[INSPIRE](#)].
- [20] N. Armesto, J. Rojo, C.A. Salgado and P. Zurita, *Bayesian reweighting of nuclear PDFs and constraints from proton-lead collisions at the LHC*, *JHEP* **11** (2013) 015 [[arXiv:1309.5371](#)] [[INSPIRE](#)].
- [21] B.J.A. Watt, P. Motylinski and R.S. Thorne, *The Effect of LHC Jet Data on MSTW PDFs*, *Eur. Phys. J. C* **74** (2014) 2934 [[arXiv:1311.5703](#)] [[INSPIRE](#)].
- [22] H. Paukkunen and C.A. Salgado, *Agreement of Neutrino Deep Inelastic Scattering Data with Global Fits of Parton Distributions*, *Phys. Rev. Lett.* **110** (2013) 212301 [[arXiv:1302.2001](#)] [[INSPIRE](#)].
- [23] N. Sato, J.F. Owens and H. Prosper, *Bayesian Reweighting for Global Fits*, *Phys. Rev. D* **89** (2014) 114020 [[arXiv:1310.1089](#)] [[INSPIRE](#)].
- [24] ZEUS COLLABORATION, H1 collaboration, A.M. Cooper-Sarkar, *PDF Fits at HERA*, *PoS(EPS-HEP2011)320* [[arXiv:1112.2107](#)] [[INSPIRE](#)].
- [25] S. Alekhin, J. Blumlein and S. Moch, *Parton Distribution Functions and Benchmark Cross Sections at NNLO*, *Phys. Rev. D* **86** (2012) 054009 [[arXiv:1202.2281](#)] [[INSPIRE](#)].
- [26] J.F. Owens, A. Accardi and W. Melnitchouk, *Global parton distributions with nuclear and finite- $Q^2$  corrections*, *Phys. Rev. D* **87** (2013) 094012 [[arXiv:1212.1702](#)] [[INSPIRE](#)].
- [27] 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* **07** (2002) 012 [[hep-ph/0201195](#)] [[INSPIRE](#)].
- [28] J. Pumplin, *Parametrization dependence and Delta  $\chi^2$  in parton distribution fitting*, *Phys. Rev. D* **82** (2010) 114020 [[arXiv:0909.5176](#)] [[INSPIRE](#)].
- [29] A. Glazov, S. Moch and V. Radescu, *Parton Distribution Uncertainties using Smoothness Prior*, *Phys. Lett. B* **695** (2011) 238 [[arXiv:1009.6170](#)] [[INSPIRE](#)].
- [30] CMS collaboration, *Measurements of differential jet cross sections in proton-proton collisions at  $\sqrt{s} = 7$  TeV with the CMS detector*, *Phys. Rev. D* **87** (2013) 112002 [[arXiv:1212.6660](#)] [[INSPIRE](#)].
- [31] T. Kluge, K. Rabbertz and M. Wobisch, *FastNLO: Fast pQCD calculations for PDF fits*, [hep-ph/0609285](#) [[INSPIRE](#)].
- [32] FASTNLO collaboration, D. Britzger, K. Rabbertz, F. Stober and M. Wobisch, *New features in version 2 of the fastNLO project*, [arXiv:1208.3641](#) [[INSPIRE](#)].

- [33] FASTNLO collaboration, M. Wobisch, D. Britzger, T. Kluge, K. Rabbertz and F. Stober, *Theory-Data Comparisons for Jet Measurements in Hadron-Induced Processes*, [arXiv:1109.1310](#) [[INSPIRE](#)].
- [34] P.M. Nadolsky, H.-L. Lai, Q.-H. Cao, J. Huston, J. Pumplin et al., *Implications of CTEQ global analysis for collider observables*, *Phys. Rev. D* **78** (2008) 013004 [[arXiv:0802.0007](#)] [[INSPIRE](#)].
- [35] J. Gao, M. Guzzi, J. Huston, H.-L. Lai, Z. Li et al., *CT10 next-to-next-to-leading order global analysis of QCD*, *Phys. Rev. D* **89** (2014) 033009 [[arXiv:1302.6246](#)] [[INSPIRE](#)].
- [36] D. Stump, J. Pumplin, R. Brock, D. Casey, J. Huston et al., *Uncertainties of predictions from parton distribution functions. 1. The Lagrange multiplier method*, *Phys. Rev. D* **65** (2001) 014012 [[hep-ph/0101051](#)] [[INSPIRE](#)].
- [37] S. Albino, B.A. Kniehl and G. Kramer, *AKK Update: Improvements from New Theoretical Input and Experimental Data*, *Nucl. Phys. B* **803** (2008) 42 [[arXiv:0803.2768](#)] [[INSPIRE](#)].
- [38] S. Dulat, T.-J. Hou, J. Gao, J. Huston, P. Nadolsky et al., *Higgs Boson Cross Section from CTEQ-TEA Global Analysis*, *Phys. Rev. D* **89** (2014) 113002 [[arXiv:1310.7601](#)] [[INSPIRE](#)].
- [39] CDF collaboration, T. Affolder et al., *Measurement of the inclusive jet cross section in  $\bar{p}p$  collisions at  $\sqrt{s} = 1.8$  TeV*, *Phys. Rev. D* **64** (2001) 032001 [Erratum *ibid.* **D 65** (2002) 039903] [[hep-ph/0102074](#)] [[INSPIRE](#)].
- [40] D0 collaboration, B. Abbott et al., *High- $p_T$  jets in  $\bar{p}p$  collisions at  $\sqrt{s} = 630$  GeV and 1800 GeV*, *Phys. Rev. D* **64** (2001) 032003 [[hep-ex/0012046](#)] [[INSPIRE](#)].
- [41] J. Pumplin, J. Huston, H.L. Lai, P.M. Nadolsky, W.-K. Tung et al., *Collider Inclusive Jet Data and the Gluon Distribution*, *Phys. Rev. D* **80** (2009) 014019 [[arXiv:0904.2424](#)] [[INSPIRE](#)].
- [42] ATLAS collaboration, *Measurement of inclusive jet and dijet production in  $pp$  collisions at  $\sqrt{s} = 7$  TeV using the ATLAS detector*, *Phys. Rev. D* **86** (2012) 014022 [[arXiv:1112.6297](#)] [[INSPIRE](#)].