



Parton distributions with theory uncertainties: general formalism and first phenomenological studies

NNPDF Collaboration

Rabah Abdul Khalek^{1,2}, Richard D. Ball³, Stefano Carrazza⁴, Stefano Forte^{4,a}, Tommaso Giani³, Zahari Kassabov⁵, Rosalyn L. Pearson³, Emanuele R. Nocera², Juan Rojo^{1,2}, Luca Rottoli^{6,7}, Maria Ubiali⁸, Cameron Voisey⁵, Michael Wilson³

¹ Department of Physics and Astronomy, VU University, 1081 HV, Amsterdam, The Netherlands

² Nikhef Theory Group, Science Park 105, 1098 XG Amsterdam, The Netherlands

³ The Higgs Centre for Theoretical Physics, University of Edinburgh, JCMB, KB, Mayfield Rd, Edinburgh, Scotland EH9 3JZ, UK

⁴ Tif Lab, Dipartimento di Fisica, Università di Milano and INFN, Sezione di Milano, Via Celoria 16, 20133 Milan, Italy

⁵ Cavendish Laboratory, University of Cambridge, Cambridge CB3 0HE, UK

⁶ Dipartimento di Fisica G. Occhialini, U2, Università degli Studi di Milano-Bicocca, Piazza della Scienza, 3, 20126 Milan, Italy

⁷ INFN, Sezione di Milano-Bicocca, 20126 Milan, Italy

⁸ DAMTP, University of Cambridge, Wilberforce Road, Cambridge CB3 0WA, UK

Received: 1 July 2019 / Accepted: 16 October 2019 / Published online: 15 November 2019

© The Author(s) 2019

Abstract We formulate a general approach to the inclusion of theoretical uncertainties, specifically those related to the missing higher order uncertainty (MHOU), in the determination of parton distribution functions (PDFs). We demonstrate how, under quite generic assumptions, theory uncertainties can be included as an extra contribution to the covariance matrix when determining PDFs from data. We then review, clarify, and systematize the use of renormalization and factorization scale variations as a means to estimate MHOUs consistently in deep inelastic and hadronic processes. We define a set of prescriptions for constructing a theory covariance matrix using scale variations, which can be used in global fits of data from a wide range of different processes, based on choosing a set of independent scale variations suitably correlated within and across processes. We set up an algebraic framework for the choice and validation of an optimal prescription by comparing the estimate of MHOU encoded in the next-to-leading order (NLO) theory covariance matrix to the observed shifts between NLO and NNLO predictions. We perform a NLO PDF determination which includes the MHOU, assess the impact of the inclusion of MHOUs on the PDF central values and uncertainties, and validate the results by comparison to the known shift between NLO and NNLO PDFs. We finally study the impact of the inclusion of MHOUs in a global PDF determination on LHC cross-sections, and provide guidelines for their use in precision phenomenology. In addition, we also compare the results based on the theory

covariance matrix formalism to those obtained by performing PDF determinations based on different scale choices.

Contents

1	Introduction	2
2	A theoretical covariance matrix	4
3	MHOUs from scale variations	6
3.1	Renormalization group invariance	7
3.2	Scale variation for partonic cross-sections	8
3.2.1	Electroproduction	9
3.2.2	Hadronic processes	10
3.3	Scale variation for PDF evolution	10
3.4	Double scale variations	13
3.4.1	Electroproduction	13
3.4.2	Hadronic processes	15
3.5	Multiple scale variations	16
4	Scale variation prescriptions for the theory covariance matrix	16
4.1	Symmetric prescriptions for individual processes	17
4.2	Symmetric prescriptions for multiple processes	18
4.3	Asymmetric prescriptions	20
5	Validation of the theory covariance matrix	21
5.1	Input data and process categorization	21
5.2	The theory covariance matrices at NLO	22
5.3	Construction of validation tests	25
5.4	Results of validation tests	27

^ae-mail: stefano.forte@mi.infn.it

6	PDFs with missing higher order uncertainties	32
6.1	DIS-only PDFs	35
6.2	Global PDFs	36
7	Implications for phenomenology	37
7.1	Higgs production	40
7.2	Top quark pair production	44
7.3	Z and W gauge boson production	45
8	Usage and delivery	47
8.1	Combining MHOUs in PDFs and hard matrix elements	47
8.2	Computation of the total uncertainty	48
8.3	Delivery	49
9	Summary and outlook	49
A	Diagonalisation of the theory covariance matrix	50
B	PDF sets with different scale choices	52
B.1	DIS-only PDFs	53
B.2	Global PDFs	55
B.3	The envelope prescription for MHOUs	56
	References	59

1 Introduction

An accurate estimate of the uncertainty in Standard Model (SM) predictions is a crucial ingredient for precision phenomenology at the Large Hadron Collider (LHC). Now, and for several years to come [1,2], theoretical uncertainties for hadron collider processes are dominated by the missing higher order uncertainty (MHOUs) in perturbative QCD calculations, usually estimated by scale variation, and by parton distribution function (PDF) uncertainties. Of course, PDFs summarize the information on the nucleon structure extracted from other SM processes [3]: effectively, PDFs provide a way of obtaining a prediction for a given process in terms of other processes. This way of thinking about PDFs immediately shows that MHOUs are present not only in the perturbative prediction for a particular process, but also in the underlying processes used for the PDF determination.

Current PDF uncertainties essentially only include the propagated uncertainty arising from statistical and systematic uncertainties in the experimental data used in their determination. Methodological uncertainties related for example to the choice of functional form for the PDFs, or the fitting methodology employed, can be kept under control using closure tests [4], and with care can be made negligible in the data region. Parametric uncertainties, such as those related to the value of the strong coupling $\alpha_s(m_Z)$ or the charm mass m_c can be included by performing fits for a range of parameters. However up until now MHOUs have never been included in a PDF fit: what is usually called the “PDF uncertainty” does not include the MHOUs in the theoretical calculations used for PDF determination, and, more generally, does not typically include any source of theory uncertainty.

Historically, this is related to the fact that MHOUs have always been considered as likely to be small in comparison to other PDF uncertainties, especially since NNLO PDFs have become the default standard. However, it is clear that as PDF uncertainties become smaller and smaller, at some point MHOUs will become significant. In the most recent NNPDF set, NNPDF3.1 [5], PDF uncertainties at the electroweak scale can be as low as 1%. Given that the typical size of MHOUs on NNLO QCD processes is at the percent level (see e.g. [6]) their neglect seems difficult to justify a priori.

Besides contributing to the overall size of PDF uncertainty, more subtly the MHOUs might affect the relative weights of different datasets included in the fit: a dataset which is accurately described by NNLO theory because it has small MHOUs should in principle carry more weight than one which is poorly described because it has large MHOUs. The neglect of MHOUs might thus be biasing current global PDF fits.

It is the purpose of this paper to set up a general formalism for the inclusion of theoretical uncertainties, specifically MHOUs, in PDF determinations, and then to perform a first exploration of their impact on LHC phenomenology. The development of this treatment of MHOUs will involve three main ingredients. The first is the formulation of a general theory for the inclusion in PDF fits of generic theoretical uncertainties, of which MHOUs are a particular case. The second is the choice of a specific method for estimating the MHOUs in each of the cross-sections that enter the PDF fit. The third is the construction of a set of tools for the validation of this methodology, to check that the MHOUs are being correctly estimated.

The first ingredient in our approach is common to any kind of theory uncertainty: theory uncertainties include not only MHOUs, but also any other aspect in which the theory used to obtain predictions for the physical processes that enter the PDF fit is incompletely known. These include higher twists (see Refs. [7,8] and Ref. therein) and other power-suppressed corrections, nuclear corrections when nuclear targets are involved (see Refs. [9,10] and Ref. therein), final state corrections for non-inclusive processes, and so forth. All of these uncertainties are only meaningful in a Bayesian sense: there is only one correct value of the next-order perturbative correction, not a distribution of values. They thus necessarily involve a process of informed estimation or guesswork: the only way to actually know the size of, say, a missing higher order correction, is to calculate it.

We will show by adopting a Bayesian point of view, and assigning a Gaussian probability distribution to the expected true value of the theory calculation, that the impact of any missing theoretical contribution can be encoded as an additive contribution to the experimental covariance matrix used in the PDF fit [11]. The combination is additive because experimental and theoretical uncertainties are by their nature independent, and are thus combined in quadrature. In a global

fit, theoretical uncertainties can be strongly correlated not only across data points within a given experiment, but also between different experiments, and even different processes, so we need a theoretical covariance matrix which includes all these correlations across all the datasets included in the fit.

This then immediately raises the issue of choosing a meaningful way to estimate the MHOUs, which in particular incorporates these correlations. The standard way of estimating MHOUs in perturbative QCD calculations is to perform a variation of the renormalization and factorization scales, denoted as μ_r and μ_f respectively, with various choices for the range and combination of variations existing. While the shortcomings of this method are well known, and various alternatives have been discussed [12–14], this remains the default and most widely used option. In the present context, its main advantage is its universality (it can be applied in the same way to any of the processes used in the fit), and the way in which it implicitly incorporates correlations (for example predictions for data points in the same process which are kinematically close will be automatically correlated), even across different processes (through the PDFs, which are the same in every process). Thus while in principle our covariance matrix formalism allows for the inclusion of any method for estimating MHOUs in a PDF determination, here we will specifically use scale variation.

In order to do this, we need to examine systematically the underpinnings of scale variation as a means to estimate theory uncertainties, since different definitions of scale variation have been used in different contexts. Indeed, the standard definitions of renormalization and factorization scale typically used for deep-inelastic scattering and hadronic collisions are not the same. Because PDF fits include both types of processes, it is important to understand in detail how these definitions relate to each other, in order to be able to correlate the scale variations in a meaningful way. Specifically, we will show that one may estimate the MHOUs for any process by combining two independent scale variations: one to estimate the MHOUs in the perturbative evolution of the PDFs (missing higher orders in the DGLAP splitting functions), and the other to estimate the MHOUs in the perturbative calculation of the partonic cross-sections (missing higher orders in the hard-scattering matrix elements).

Once the scales to be varied are understood, the remaining task is to choose a particular prescription to be used to construct the theoretical covariance matrix. In estimating MHOUs for a given process, the most commonly adopted option is the so-called seven-point envelope prescription, in which μ_r and μ_f are independently varied by a factor of two about the central choice while ensuring that $1/2 \leq \mu_r/\mu_f \leq 2$, and the MHOUs is then taken as the envelope of the results. For our purposes this is insufficient: rather than taking an envelope, we wish to construct a covariance matrix out of the

scale variations. In particular, because theoretical uncertainties are correlated across processes (through the evolution of the PDFs), we need a prescription for determining the entries of the covariance matrix both within a single process and across pairs of processes.

We will discuss in detail a variety of options to achieve this, based on a general “ n -point prescription”. These options will differ from each other in the choice of the number of independent variations, the directions of such variations in the (μ_r, μ_f) plane, and the way the variations are correlated (or not) across different processes.

The validation of these point prescriptions, and the choice of the optimal one to be used for PDF determinations is a nontrivial problem, which however admits an elegant solution. The validation can be performed at NLO, by comparing the estimate of the MHOUs encoded in the theory covariance matrix to the known next (NNLO) order correction. The problem is then to compare the probability distribution of expected higher-order results to the unique answer given by the NNLO calculation. The solution to this problem is to view the set of shifts between the NLO and NNLO computations for all the processes under consideration as a vector, with one component for each of the data points. The theory covariance matrix corresponding to each prescription then defines a one-sigma ellipsoid in a subspace of this space. The validation is performed by projecting the shift vector into the ellipsoid: if the theory covariance matrix gives a sensible estimate of the MHOUs at NLO, the shift vector will lie almost entirely within the ellipsoid. Using this strategy, we will validate a variety of scale variation prescriptions on a similar dataset to that of the global NNPDF3.1 analysis. Since the dimension of the space of datapoints is typically two orders of magnitude higher than the dimension of the subspace of the ellipsoid, this is a highly nontrivial test.

Once a prescription has been selected and used to construct the theory covariance matrix, it is possible to perform a PDF fit based on it. Within the NNPDF methodology, an ensemble of PDF replicas is fitted to data replicas. Data replicas are generated in a way which reflects the uncertainties and correlations of the underlying data, as encoded in their covariance matrix. The best-fit PDF replica for each data replica is then determined by minimizing a figure of merit (χ^2) which is computed using the covariance matrix. As mentioned, and as we shall show in Sect. 2, the theory contribution appears as an independent contribution to the total covariance matrix, uncorrelated with the experimental one and simply added to it. Therefore, once the covariance matrix is supplemented by an extra theory contribution coming from MHOUs, this should be treated on the same footing as any other contribution, and it will thus affect both the data replica generation, and the fitting of PDF replicas to data replicas.

Qualitatively, one may expect the inclusion of the MHOUs in the data replica generation to increase the spread of the data

replicas, and thus lead in itself to an increase in overall PDF uncertainties. On the other hand the inclusion of the MHOU in the fitting might also reduce tensions within the fit due to the imperfection of the theory and, since these are highly correlated, result in significant shifts in central values, and overall a better fit with reduced uncertainties. The combined effect of including the MHOU in both the data generation and the fitting is thus not at all obvious.

We will investigate these effects by performing PDF determinations in which MHOUs are included in either, or both, the replica generation and the PDF replica fitting. Once again, results can be validated at NLO by comparing NLO PDFs determined with the theory covariance matrix to NNLO PDFs. A successful validation should show that the best-fit NLO PDF moves towards the central NNLO result upon inclusion of the theory covariance matrix in both replica generation and fitting, due to a relaxation of tensions in the NLO fit, and that the NNLO PDF differs from the NLO PDF by an amount which is correctly estimated by the NLO uncertainty band. As we shall see, this is indeed the case, and in fact it will turn out that often the uncertainty band does not increase or even decreases upon inclusion of the theory covariance matrix.

Having determined PDFs which now account for the MHOU associated to the processes that enter the fit, the natural questions which then arise are what is their impact, and more generally how they should be used for precision LHC phenomenology. In order to address the first question, we will compute predictions with MHOUs for typical LHC standard candle processes, both with and without including the MHOU in the PDF, and provide a first phenomenological exploration and assessment of the impact of these uncertainties.

The second question is not entirely trivial and we will address it in detail. Indeed, scale variation is routinely performed in order to estimate the MHOU in theoretical predictions for hadron collider processes. Clearly, when obtaining a prediction, we should avoid double counting a MHOU which has already been included in the PDF. Instances in which this might happen include not only the trivial situation in which a prediction is obtained for a process which has already been used for PDF determination, but also the somewhat more subtle situation in which the MHOU in the PDF and the observable which is being predicted are correlated through perturbative evolution [15]. We will discuss this situation, and provide guidelines for the usage of PDFs with MHOUs.

This paper is broadly divided into two main parts. In the first part, we construct a general formalism for the inclusion of theory uncertainties and specifically MHOUs in PDF determination, and show how to construct and validate a theory covariance matrix. In the second part, we perform a first investigation of the phenomenological implications of these theory uncertainties. The structure of the paper is the follow-

ing: in Sect. 2 we show, using a Bayesian approach, that under certain assumptions any type of theory uncertainty can be included as a contribution to the covariance matrix. In Sect. 3 we summarize the theory of scale variation and use it to review, compare and systematize different definitions which have been used in the literature. In Sect. 4 we then formulate a number of “point prescriptions” for the theory covariance matrix, both for a single process, and also to account for correlations between a pair of processes. In Sect. 5 we compute the theory covariance matrix for a variety of prescriptions, we test them against known higher order corrections, and use this comparison to select an optimal prescription.

We then move to the second, more phenomenological, part of the paper. The centerpiece of this section is the determination of NLO PDF sets with MHOU, presented in Sect. 6. We first only include deep-inelastic scattering data (DIS-only fit), and then adopt a global data set, which is compared to PDFs without MHOU, and validated against NNLO PDFs. In Sect. 7 we present initial studies of the phenomenological impact of the inclusion of MHOUs in PDFs for representative LHC processes. Finally in Sect. 8 we provide guidelines for the usage of PDFs with MHOU, in particular concerning the combination of the PDF uncertainties with the MHOU on the hard matrix element, and present the delivery of the PDF sets produced in this work.

Two appendices contain further studies and technical details. In Appendix A we provide additional details concerning the procedure adopted to diagonalise the theory covariance matrix. Then in Appendix B we study another possible validation of the results of Sect. 6, by comparing PDFs with MHOUs to the PDFs obtained by adopting different choices of renormalization and factorization scales in the PDF determination. Families of fits which only differ in choices of scale have never been carried out before and will be presented here for the first time. Whereas they do not necessarily give a fair estimate of the MHOU on PDFs, they surely do provide an indication of the expected impact of scale variation on PDFs, and the pattern of MHOU correlations.

A concise discussion of the main results of this work was presented in Ref. [16], of which this paper represents the extended companion.

2 A theoretical covariance matrix

Parton distribution functions are determined from a set of N_{dat} experimental data points, which we represent by an N_{dat} -dimensional vector D_i , $i = 1, \dots, N_{\text{dat}}$. These data points have experimental uncertainties that may be correlated with each other, and this information is encoded in an experimental covariance matrix C_{ij} . This covariance matrix may be block-diagonal if some sets of data are uncorrelated. Each experimental data point has associated with it a “true”

value \mathcal{T}_i – the value given by Nature – whose determination is the goal of the experiment. Since the experimental measurements are imperfect, they cannot determine \mathcal{T} exactly, but they can be used to estimate the Bayesian probability of a given hypothesis for \mathcal{T} . Assuming that the experimental results are Gaussianly distributed about this hypothetical true value, the conditional probability for the true values \mathcal{T} given the measured cross-sections D is

$$P(\mathcal{T}|D) = P(D|\mathcal{T}) \propto \exp\left(-\frac{1}{2}(\mathcal{T}_i - D_i)C_{ij}^{-1}(\mathcal{T}_j - D_j)\right), \tag{2.1}$$

up to an overall normalization constant. Note that this tacitly assumes equal priors for both D and \mathcal{T} .

Of course the true values \mathcal{T}_i are unknown. However we can calculate theoretical predictions for each data point D_i , which we denote by T_i . These predictions are computed using a theory framework which is generally incomplete: for example because it is based on the fixed-order truncation of a perturbative expansion, or because it excludes higher-twist effects, or nuclear effects, or some other effect that is difficult to calculate precisely. Furthermore, these theory predictions T_i depend on PDFs, evolved to a suitable scale also using incomplete theory. While the theory predictions may correspond to a variety of different observables and processes, they all depend on the same underlying (universal) PDFs.

We now assume, in the same spirit as when estimating experimental systematics, that the true values \mathcal{T}_i are centered on the theory predictions T_i , and Gaussianly distributed about the theory predictions, with which they would coincide if the theory were exact and the PDFs were known with certainty. The conditional probability for the true values \mathcal{T} given theoretical predictions T is then

$$P(\mathcal{T}|T) = P(T|\mathcal{T}) \propto \exp\left(-\frac{1}{2}(\mathcal{T}_i - T_i)S_{ij}^{-1}(\mathcal{T}_j - T_j)\right), \tag{2.2}$$

again up to a normalization constant, where S_{ij} is a “theory covariance matrix”, to be estimated in due course.

PDFs are determined by maximizing the probability of the theory given the data $P(T|D)$, marginalised over the true values \mathcal{T} which of course remain unknown. Now using Bayes’ theorem

$$P(\mathcal{T}|DT)P(D|\mathcal{T}) = P(D|\mathcal{T}T)P(\mathcal{T}|T). \tag{2.3}$$

Moreover, since the experimental data do not depend on the theorists’ calculations T , but only on the ‘truth’ \mathcal{T} ,

$$P(D|\mathcal{T}T) = P(D|\mathcal{T}). \tag{2.4}$$

Then because by construction $\int D^N \mathcal{T} P(\mathcal{T}|D) = 1$,

$$P(D|T) = \int D^N \mathcal{T} P(\mathcal{T}|D)P(\mathcal{T}|T), \tag{2.5}$$

where the N -dimensional integral is over all of the possible values of \mathcal{T}_i . The probability of the experimental data D is now conditional on the theory T because we have marginalised over the underlying ‘truth’ \mathcal{T} , which is common to both.

Writing the difference between the true \mathcal{T}_i and the actual T_i values of the theory prediction as

$$\Delta_i \equiv \mathcal{T}_i - T_i, \tag{2.6}$$

we can change variables of integration to convert the integral over \mathcal{T}_i into an integral over the shifts Δ_i : using the Gaussian hypotheses Eqs. (2.1) and (2.2), Eq. (2.5) becomes that

$$P(D|T) \propto \int D^N \Delta \exp\left(-\frac{1}{2}(D_i - T_i - \Delta_i) \times C_{ij}^{-1}(D_j - T_j - \Delta_j) - \frac{1}{2}\Delta_i S_{ij}^{-1}\Delta_j\right). \tag{2.7}$$

The Gaussian integrals can now be performed explicitly. Adopting a vector notation in order to make the algebra more transparent, we rewrite the exponent as

$$\begin{aligned} & (D - T - \Delta)^T C^{-1}(D - T - \Delta) + \Delta^T S^{-1}\Delta \\ &= \Delta^T (C^{-1} + S^{-1})\Delta - \Delta^T C^{-1}(D - T) \\ & \quad - (D - T)^T C^{-1}\Delta + (D - T)^T C^{-1}(D - T) \\ &= (\Delta - (C^{-1} + S^{-1})^{-1}C^{-1}(D - T))^T (C^{-1} + S^{-1}) \\ & \quad \times (\Delta - (C^{-1} + S^{-1})^{-1}C^{-1}(D - T)) \\ & \quad - (D - T)^T C^{-1}(C^{-1} + S^{-1})^{-1}C^{-1}(D - T) \\ & \quad + (D - T)^T C^{-1}(D - T), \end{aligned} \tag{2.8}$$

where we used the fact that both C and S are symmetric matrices, and in the last line we completed the square. Integrating over Δ , ignoring the normalization, Eq. (2.7) then becomes

$$P(T|D) = P(D|T) \propto \exp\left(-\frac{1}{2}(D - T)^T \times (C^{-1} - C^{-1}(C^{-1} + S^{-1})^{-1}C^{-1})(D - T)\right) \tag{2.9}$$

However

$$(C^{-1} + S^{-1})^{-1} = (C^{-1}(C + S)S^{-1})^{-1} = S(C + S)^{-1}C, \tag{2.10}$$

so that

$$\begin{aligned} C^{-1} - C^{-1}(C^{-1} + S^{-1})^{-1}C^{-1} &= C^{-1} - C^{-1}S(C + S)^{-1} \\ &= (C^{-1}(C + S) - C^{-1}S)(C + S)^{-1} = (C + S)^{-1}. \end{aligned} \quad (2.11)$$

Restoring the indices, we thus find the simple result

$$P(\mathcal{T}|D) \propto \exp\left(-\frac{1}{2}(D_i - T_i)(C + S)_{ij}^{-1}(D_j - T_j)\right). \quad (2.12)$$

Comparison of Eq. (2.12) with Eq. (2.1) indicates that when replacing the true \mathcal{T}_i by the theoretical predictions T_i in the expression of the χ^2 of the data, the theoretical covariance matrix S_{ij} should simply be added to the experimental covariance matrix C_{ij} [11]. In effect this implies that, at least within this Gaussian approximation, when determining PDFs theoretical uncertainties can be treated simply as another form of experimental systematic: it is an additional uncertainty to be taken into account when trying to find the truth from the data on the basis of a specific theoretical prediction. The experimental and theoretical uncertainties are added in quadrature because they are in principle uncorrelated.

In the case for which theoretical uncertainties can be neglected, i.e. if $S_{ij} \rightarrow 0$, then $P(\mathcal{T}|T)$ in Eq. (2.2) becomes proportional to $\delta^N(\mathcal{T}_i - T_i)$. As a result, in this case Eq. (2.12) reduces to Eq. (2.1) with \mathcal{T}_i replaced by the predictions T_i . This shows that Eq. (2.12) remains true even if S_{ij} has zero eigenvalues and is thus not invertible. Note however that by construction C_{ij} is positive definite, since any experimental measurement always has uncorrelated statistical uncertainties due to the finite number of events, so $(C + S)_{ij}$ will always be invertible.

The question remains of how to estimate the theory covariance matrix, S_{ij} . The Gaussian hypothesis Eq. (2.2) implies that

$$S_{ij} = \langle (\mathcal{T}_i - T_i)(\mathcal{T}_j - T_j) \rangle = \langle \Delta_i \Delta_j \rangle, \quad (2.13)$$

where the average is taken over the true theory values \mathcal{T} using the probability distribution $P(\mathcal{T}|T)$, and $\langle \Delta_i \rangle = 0$ consistent with the assumption that the probability distribution of the truth \mathcal{T} is centred on the theoretical calculation T . In practice however the formal definition Eq. (2.13) is not very helpful: we need some way to estimate the shifts Δ_i – ‘nuisance parameters’, in the language of systematic error determination – in a way that takes into account the theoretical correlations between different kinematic points within the same dataset, between different datasets measuring the same physical process, and between datasets corresponding to different processes (with initial state hadrons). Note that theory

correlations will always be present even for entirely different processes, through the universal parton distributions: the only processes with truly independent theoretical uncertainties are those with only leptons in the initial state, which are of course irrelevant for PDF determination.

The most commonly used method of estimating the theory corrections due to MHOUs, which can naturally incorporate all these theoretical correlations, is scale variation. This method is reviewed in Sect. 3 in general terms and then used in Sect. 4 in order to formulate specific prescriptions for constructing the theory covariance matrix S_{ij} . Other approaches which have been discussed in the literature involve estimating MHOUs based on the behaviour of the known perturbative orders [12–14]; however, at least at present, these do not appear to provide a formalism which is sufficiently well-established, and of appropriately general applicability. We emphasize however that the formalism presented in this section is independent of the specific method adopted to estimate the correlated theory shifts Δ_i that enter Eq. (2.13).

3 MHOUs from scale variations

The variation of the renormalization and factorization scales is the most popular approach for estimating missing higher order uncertainties (MHOUs) in QCD perturbative calculations. It has a number of advantages: it naturally incorporates renormalization group (RG) invariance, thereby ensuring that as the perturbative order increases, estimates of MHOUs decrease; the same procedure can be used for any perturbative process, since the scale dependence of the strong coupling $\alpha_s(\mu^2)$ and of PDFs is universal; the estimates of MHOUs it produces are smooth functions of the kinematics, and thereby correctly incorporate the strong correlations in nearby regions of phase space; and correlations between different processes due to universal ingredients such as PDFs can be easily incorporated. Its drawbacks are also well known: there is no unique principle to determine the specific range of the scale variation (nor even the precise central scale to be adopted); and it misses uncertainties associated with new singularities or color structures present at higher orders but missing at lower orders. The former problem may be dealt with, at least qualitatively, by validating a given range in situations where the next order corrections are known. We will attempt such a validation in this paper. The latter problem is more challenging, requiring resummation in the case of unresummed logarithms, or other methods of estimating new types of corrections, and it is unclear whether or not it admits a general solution.

While scale variation has been discussed many times in a variety of contexts, there is no standard, commonly accepted formulation of it, and specifically none that can be applied to both electroproduction and hadroproduction processes, as

we need to do if we wish to use scale variation in the context of global PDF analyses. In fact, it turns out that the most commonly adopted approaches to scale variation differ, typically according to the nature of the process which is being considered, though also as a function of time, with different prescriptions being favored in the past than those in common use at the present. Moreover, even the terminology is not uniform: it has evolved over time, resulting in the same names being used for what are essentially different scale variations.

To formulate prescriptions for the general use of scale variation for MHO estimation which can be applied to any process included in present or future PDF determinations, it is thus necessary to first review the underpinnings of scale variation, and to then use them in order to set up a generally applicable formalism. This will be done in the current section, by specifically discussing the cases of electroproduction and hadroproduction. In particular, we will show that for factorized processes MHOs on the partonic cross-sections and on perturbative evolution are independent and can be estimated through independent scale variations. We will then discuss how they can be combined, first with a single process and then for several processes, both correlated and uncorrelated.

3.1 Renormalization group invariance

The basic principle of scale variation is based on the observation that scale-dependent contributions to a perturbative prediction are fixed by RG invariance, and therefore scale variation can be used to generate higher order contributions, which are then taken as a proxy for the whole missing higher orders.

More explicitly, consider a generic theoretical prediction (typically a perturbative cross-section) of the form $\bar{T}(\alpha_s(\mu^2), \mu^2/Q^2)$, where μ^2 is the renormalization scale and Q^2 is some physical scale in the process. Thus \bar{T} indicates the theory prediction T when it is evaluated at some renormalization scale μ^2 instead of being evaluated at the physical scale Q^2 : if we instead set $\mu^2 = Q^2$, then

$$T(Q^2) \equiv \bar{T}(\alpha_s(Q^2), 1). \tag{3.1}$$

The QCD running coupling $\alpha_s(\mu^2)$ satisfies the RG equation

$$\mu^2 \frac{d}{d\mu^2} \alpha_s(\mu^2) = \beta(\alpha_s(\mu^2)), \tag{3.2}$$

where the QCD beta function has the following perturbative expansion:

$$\beta(\alpha_s) = \beta_0 \alpha_s^2 + \beta_1 \alpha_s^3 + \beta_2 \alpha_s^4 + \dots \tag{3.3}$$

RG invariance is the statement that the all-order prediction is independent of the renormalization scale:

$$\mu^2 \frac{d}{d\mu^2} \bar{T}(\alpha_s(\mu^2), \mu^2/Q^2) = 0. \tag{3.4}$$

It will be useful in what follows to define the variables

$$\mu^2 = kQ^2, \quad t = \ln(Q^2/\Lambda^2), \quad \kappa = \ln k = \ln \mu^2/Q^2, \tag{3.5}$$

so $\alpha_s(\mu^2)$ is a function of $\ln \mu^2/\Lambda^2 = t + \kappa$. We can then write the RG equation (3.4) as

$$\begin{aligned} 0 &= \frac{d}{d\kappa} \bar{T}(\alpha_s(t + \kappa), \kappa) \\ &= \frac{d}{d\kappa} \alpha_s(t + \kappa) \frac{\partial}{\partial \alpha_s} \bar{T}(\alpha_s(t + \kappa), \kappa) \Big|_{\kappa} \\ &\quad + \frac{\partial}{\partial \kappa} \bar{T}(\alpha_s(t + \kappa), \kappa) \Big|_{\alpha_s} \\ &= \frac{\partial}{\partial t} \bar{T}(\alpha_s(t + \kappa), \kappa) \Big|_{\kappa} + \frac{\partial}{\partial \kappa} \bar{T}(\alpha_s(t + \kappa), \kappa) \Big|_{\alpha_s}, \end{aligned} \tag{3.6}$$

where in the second line we assume that \bar{T} is analytic in α_s and κ , and in the third we use

$$\frac{d}{d\kappa} \alpha_s(t + \kappa) = \frac{d}{dt} \alpha_s(t + \kappa) = \beta(\alpha_s(t + \kappa)). \tag{3.7}$$

Taylor expanding $\bar{T}(\alpha_s, \kappa)$ in κ about $\kappa = 0$ (i.e. $k = 1$, $\mu^2 = Q^2$) at fixed coupling α_s ,

$$\begin{aligned} \bar{T}(\alpha_s(t + \kappa), \kappa) &= \bar{T}(\alpha_s(t + \kappa), 0) \\ &\quad + \kappa \frac{\partial}{\partial \kappa} \bar{T}(\alpha_s(t + \kappa), 0) \Big|_{\alpha_s} \\ &\quad + \frac{1}{2} \kappa^2 \frac{\partial^2}{\partial \kappa^2} \bar{T}(\alpha_s(t + \kappa), 0) \Big|_{\alpha_s} + \dots \\ &= \bar{T}(\alpha_s(t + \kappa), 0) - \kappa \frac{\partial}{\partial t} \bar{T}(\alpha_s(t + \kappa), 0) \Big|_{\kappa} \\ &\quad + \frac{1}{2} \kappa^2 \frac{\partial^2}{\partial t^2} \bar{T}(\alpha_s(t + \kappa), 0) \Big|_{\kappa} + \dots, \end{aligned} \tag{3.8}$$

where in the second line we use the RG invariance condition, Eq. (3.6), to replace $\frac{\partial}{\partial \kappa}$ with $-\frac{\partial}{\partial t}$. We can thus determine the κ dependence of $\bar{T}(\alpha_s, \kappa)$ using the dependence of $T(t) = \bar{T}(\alpha_s(t), 0)$ on t :

$$\begin{aligned} \bar{T}(\alpha_s(t + \kappa), \kappa) &= T(t + \kappa) - \kappa \frac{d}{dt} T(t + \kappa) \\ &\quad + \frac{1}{2} \kappa^2 \frac{d^2}{dt^2} T(t + \kappa) + \dots \end{aligned} \tag{3.9}$$

Now since

$$\frac{d}{dt}T(t) = \frac{d\alpha_s(t)}{dt} \frac{\partial}{\partial\alpha_s} \bar{T}(\alpha_s(t), 0) = \beta(\alpha_s(t)) \frac{\partial}{\partial\alpha_s} \bar{T}(\alpha_s(t), 0), \quad (3.10)$$

and $\beta(\alpha_s) = \mathcal{O}(\alpha_s^2)$, we see that $\frac{1}{T} \frac{dT}{dt} = \mathcal{O}(\alpha_s)$, while $\frac{1}{T} \frac{d^2T}{dt^2} = \mathcal{O}(\alpha_s^2)$ etc.: derivatives with respect to t always add one power of α_s . It follows that in Eq. (3.9), the term $\mathcal{O}(\kappa)$ is $\mathcal{O}(\alpha_s)$ with respect to the leading term, and the term $\mathcal{O}(\kappa^2)$ is $\mathcal{O}(\alpha_s^2)$ with respect to the leading term, and so on. We thus see explicitly that the scale-dependent terms (those that depend on κ), at a given order in perturbation theory, are determined by derivatives of the cross-section lower down the perturbation series.

This implies that if we know the cross-section $T(t)$ as a function of the central scale Q^2 to a given order in perturbation theory, we can then use Eq. (3.9) to determine the scale-dependent κ terms directly from $T(t)$ at any given order, by differentiating terms lower down the perturbative expansion. For instance, truncating at LO, NLO, or NNLO, one has

$$\begin{aligned} \bar{T}_{\text{LO}}(\alpha_s(t+\kappa), \kappa) &= T_{\text{LO}}(t+\kappa), \\ \bar{T}_{\text{NLO}}(\alpha_s(t+\kappa), \kappa) &= T_{\text{NLO}}(t+\kappa) - \kappa \frac{d}{dt} T_{\text{LO}}(t+\kappa), \\ \bar{T}_{\text{NNLO}}(\alpha_s(t+\kappa), \kappa) &= T_{\text{NNLO}}(t+\kappa) - \kappa \frac{d}{dt} T_{\text{NLO}}(t+\kappa) \\ &\quad + \frac{1}{2} \kappa^2 \frac{d^2}{dt^2} T_{\text{LO}}(t+\kappa). \end{aligned} \quad (3.11)$$

The differentiation may be performed analytically, which is trivial for a fixed order expansion, or numerically, which can be useful in a resummed expression where the dependence on $\alpha_s(t)$ can be nontrivial [17]. Note that when the renormalization scale coincides with the physical scale of the process, $\mu^2 = Q^2$, then $\kappa = 0$ and $\bar{T} = T$ at every order in the perturbative expansion.

The MHO can now be estimated as the difference between the scale varied cross-section and the cross-section evaluated at the central scale, namely

$$\Delta(t, \kappa) = \bar{T}(\alpha_s(t+\kappa), \kappa) - T(t). \quad (3.12)$$

Thus at LO, NLO and NNLO we have, using Eq. (3.11), that the theory nuisance parameters are given by

$$\begin{aligned} \Delta_{\text{LO}}(t, \kappa) &= T_{\text{LO}}(t+\kappa) - T_{\text{LO}}(t), \\ \Delta_{\text{NLO}}(t, \kappa) &= \left(T_{\text{NLO}}(t+\kappa) - \kappa \frac{d}{dt} T_{\text{LO}}(t+\kappa) \right) - T_{\text{NLO}}(t), \\ \Delta_{\text{NNLO}}(t, \kappa) &= \left(T_{\text{NNLO}}(t+\kappa) - \kappa \frac{d}{dt} T_{\text{NLO}}(t+\kappa) \right. \\ &\quad \left. + \frac{1}{2} \kappa^2 \frac{d^2}{dt^2} T_{\text{LO}}(t+\kappa) \right) - T_{\text{NNLO}}(t). \end{aligned} \quad (3.13)$$

One finds that while at LO the theory uncertainty is entirely due to the scale chosen for α_s , at NLO the dependence on scale is milder since the leading dependence is subtracted off by the $\mathcal{O}(\kappa)$ term. At NNLO it is milder still, since the $\mathcal{O}(\kappa)$ term subtracts the leading dependence in the first term, and the $\mathcal{O}(\kappa^2)$ removes the subleading dependence in the first two terms. RG invariance then guarantees that the terms generated by scale variation are always subleading, so if the perturbation series is well behaved, the theory shifts Δ become smaller and smaller as the order of the expansion is increased.

Clearly the size of the MHO, estimated in this way, will depend on the size of the scale variation, and thus on the value chosen for κ . Typically one varies the renormalization scale by a factor of two in each direction, i.e. $\kappa \in [-\ln 4, \ln 4]$, since this range is empirically found to yield sensible results for many processes. However, in principle, one should treat κ as a free parameter, whose magnitude needs to be validated whenever possible by comparing to known higher order results.

In the present work, we are specifically interested in the application of this method to processes with one or more hadrons in the initial state, i.e. to cross-sections factorized into a hard cross-section convoluted with a PDF or a parton luminosity. There are then two independent sources of MHO: the perturbative expansion of the hard partonic cross-section, and the perturbative expansion of the anomalous dimensions that determine the perturbative evolution of the parton distributions. It is convenient to obtain each of these from an independent scale variation, and this can be done by writing separate RG equations for the hard cross-section and for the PDF, as we will demonstrate below. This approach is completely equivalent to the perhaps more familiar point of view in which MHOs on perturbative evolution are instead obtained by varying the scale at which the PDF is evaluated in the factorized expression, as we will also show.

We will begin by considering the MHO in the hard-scattering partonic cross-sections; we will then turn to a discussion of MHOs in the PDF evolution, and show that the latter can be obtained by several equivalent procedures. We will then discuss how both scale variations can be obtained from double scale variation of the hard cross-section, and how this fact also offers the possibility of performing scale variation in alternative ways whereby these two sources of MHO are mixed. We will discuss these for completeness, since in the past scale variations were often performed in this way. Finally, we will address scale variations and their correlations when several processes are considered at once.

3.2 Scale variation for partonic cross-sections

We start by considering scale variation in hard-scattering partonic cross-sections, first in the case of electroproduction (that is, for lepton-proton deep-inelastic scattering, DIS),

and then for the case of hadroproduction (proton-proton or proton-antiproton collisions).

3.2.1 Electroproduction

Consider first an electroproduction process, such as DIS, with an associated structure function given by

$$F(Q^2) = C(\alpha_s(Q^2)) \otimes f(Q^2), \tag{3.14}$$

where \otimes is the convolution in the momentum fraction x between the perturbative coefficient function $C(x, \alpha_s)$ and the PDF $f(x, Q^2)$, and where the sum over parton flavors is left implicit. In Eq. (3.14) both α_s and the PDF are evaluated at the physical scale of the process, so nothing depends on unphysical renormalization or factorization scales. We can determine the MHOU associated with the structure function F due to the truncation of the perturbative expansion of the coefficient function by fixing the factorization scheme and keeping fixed the scale at which the PDF is evaluated (usually referred to as factorization scale), but varying the renormalization scale used in the computation of the coefficient function itself.

The scale-dependent structure function \bar{F} will then be given by

$$\bar{F}(Q^2, \mu^2) = \bar{C}(\alpha_s(\mu^2), \mu^2/Q^2) \otimes f(Q^2), \tag{3.15}$$

where μ^2 is the renormalization scale used in the computation of the coefficient function, or equivalently by

$$\bar{F}(t, \kappa) = \bar{C}(\alpha_s(t + \kappa), \kappa) \otimes f(t), \tag{3.16}$$

where as in Eq. (3.5) we are using the notation $t = \ln Q^2/\Lambda^2$ and $\kappa = \ln \mu^2/Q^2$. Note that in Eq. (3.15) the structure function is written as a function of μ^2 in the sense of the RG equation (3.4): the dependence on μ^2 cancels order by order, and the residual dependence can be used to estimate the MHOU.

In phenomenological applications, it is more customary to write $F(Q^2)$, i.e. *not* to write the dependence of F on μ^2 , thereby emphasizing the renormalization scale independence of the physical observable, and just to indicate the scale dependence of the hard coefficient function $\bar{C}(\alpha_s(\mu^2), \mu^2/Q^2)$. Here and in the sequel we will stick to the notation used in RG equations since we wish to emphasize that, as the scale is varied, we are dealing with a one-parameter family of theory predictions for the physical (RG invariant) observable, which all coincide to the accuracy at which they are calculated but which differ by higher order terms.

Now, the RG invariance of physical cross-sections, and therefore of the structure function F , requires RG invariance

of the coefficient function. This is because we are not varying the factorization scheme, so the PDF is independent of the renormalization scale μ . It follows that, as in Eq. (3.11),

$$\begin{aligned} \bar{C}(\alpha_s(t + \kappa), \kappa) &= C(t + \kappa) - \kappa \frac{d}{dt} C(t + \kappa) \\ &+ \frac{1}{2} \kappa^2 \frac{d^2}{dt^2} C(t + \kappa) + \dots, \end{aligned} \tag{3.17}$$

where $C(t) = \bar{C}(\alpha_s(t), 0)$ is the coefficient function evaluated at $\mu^2 = Q^2$, and thus $\kappa = 0$. Then, given the perturbative expansion of the coefficient function,

$$C(t) = c_0 + \alpha_s(t)c_1 + \alpha_s^2(t)c_2 + \alpha_s^3(t)c_3 + \dots, \tag{3.18}$$

its derivatives can be easily evaluated using the beta function expansion Eq. (3.3),

$$\begin{aligned} \frac{d}{dt} C(t) &= \alpha_s^2(t)\beta_0 c_1 + \alpha_s^3(t)(\beta_1 c_1 + 2\beta_0 c_2) + \dots, \\ \frac{d^2}{dt^2} C(t) &= 2\alpha_s^3(t)\beta_0^2 c_1 + \dots, \end{aligned} \tag{3.19}$$

and we find that the renormalization scale variation of the coefficient function is

$$\begin{aligned} \bar{C}(\alpha_s(t + \kappa), \kappa) &= c_0 \\ &+ \alpha_s(t + \kappa)c_1 + \alpha_s^2(t + \kappa)(c_2 - \kappa\beta_0 c_1) \\ &+ \alpha_s^3(t + \kappa)(c_3 - \kappa(\beta_1 c_1 + 2\beta_0 c_2) \\ &+ \kappa^2\beta_0^2 c_1) + \dots. \end{aligned} \tag{3.20}$$

Again, note that in the case where $\mu^2 = Q^2$, and so $\kappa = 0$, one recovers the standard perturbative expansion Eq. (3.18). We can now find the scale-dependent structure function,

$$\begin{aligned} \bar{F}(t, \kappa) &= c_0 \otimes f(t) + \alpha_s(t + \kappa)c_1 \otimes f(t) \\ &+ \alpha_s^2(t + \kappa)(c_2 - \kappa\beta_0 c_1) \otimes f(t) \\ &+ \alpha_s^3(t + \kappa)(c_3 - \kappa(\beta_1 c_1 + 2\beta_0 c_2) + \kappa^2\beta_0^2 c_1) \\ &\otimes f(t) + \dots. \end{aligned} \tag{3.21}$$

Note that evaluating these expressions is numerically very straightforward, in that the scale-varied expression Eq. (3.21) has the same form, involving the same convolutions of c_i with f , as the convolution with the PDFs to the given order at the central scale Eqs. (3.14) and (3.18), only with rescaled coefficients. This means there is no need to recompute NNLO corrections, K -factors, etc.: all that is necessary is to change the coefficients in the perturbative expansion at the central scale according to Eq. (3.21).

3.2.2 Hadronic processes

MHOUs in the partonic hard cross-sections of hadronic processes can be computed in the same way as for DIS. The only additional complication is that the physical observable – typically, a cross-section Σ – now depends on the convolution of two PDFs:

$$\Sigma(t) = H(t) \otimes (f(t) \otimes f(t)), \tag{3.22}$$

where again the physical scale is $t = \ln(Q^2/\Lambda^2)$, $H(t)$ is the partonic hard-scattering cross-section, the PDFs are convoluted together into a parton luminosity $\mathcal{L} = f \otimes f$, and the sum over parton flavors is left implicit. Then, varying the renormalization scale $\kappa = \ln \mu^2/Q^2$ in the hard cross-section, we have

$$\overline{\Sigma}(t, \kappa) = \overline{H}(\alpha_s(t + \kappa), \kappa) \otimes (f(t) \otimes f(t)). \tag{3.23}$$

where, just as for electroproduction, for PDFs evaluated at a fixed scale T , the RG invariance tells us that $\overline{H}(\alpha_s(t), \kappa)$ is given in terms of $H(t)$ by Eq. (3.9):

$$\overline{H}(\alpha_s(t), \kappa) = H(t) - \kappa \frac{d}{dt} H(t) + \frac{1}{2} \kappa^2 \frac{d^2}{dt^2} H(t) + \dots \tag{3.24}$$

If the partonic process begins at $O(\alpha_s^n)$, with $n = 0, 1, 2, \dots$, then one can expand the hard cross-section as follows

$$H(t) = \alpha_s^n(t)h_0 + \alpha_s^{n+1}(t)h_1 + \alpha_s^{n+2}(t)h_2 + \dots \tag{3.25}$$

Then, as in the case of electroproduction, using Eq. (3.3) we can readily evaluate these derivatives,

$$\begin{aligned} \frac{d}{dt} H(t) &= n\alpha_s^{n-1}(t)\beta(\alpha_s)h_0 + (n+1)\alpha_s^n(t)\beta(\alpha_s)h_1 + \dots \\ &= \alpha_s^{n+1}n\beta_0h_0 + \alpha_s^{n+2}(n\beta_1h_0 + (n+1)\beta_0h_1) + \dots \\ \frac{d^2}{dt^2} H(t) &= \alpha_s^{n+2}n(n+1)\beta_0^2h_0 + \dots \end{aligned} \tag{3.26}$$

so that, putting everything together, the expression for the scale-varied partonic cross-section to be used to evaluate the scale-varied hadronic cross-section $\overline{\Sigma}$, Eq. (3.23), will be given by

$$\begin{aligned} \overline{H}(\alpha_s, \kappa) &= \alpha_s^n h_0 + \alpha_s^{n+1} (h_1 - \kappa n \beta_0 h_0) \\ &\quad + \alpha_s^{n+2} (h_2 - \kappa (n \beta_1 h_0 + (n+1) \beta_0 h_1)) \\ &\quad + \frac{1}{2} \kappa^2 n(n+1) \beta_0^2 h_1 + \dots \end{aligned} \tag{3.27}$$

This is rather more involved than Eq. (3.21), but shares the same advantages: the convolutions to be evaluated in Eq. (3.23) have the same structure as those in Eq. (3.22), so all that is required to vary the renormalization scale is to modify their coefficients.

3.3 Scale variation for PDF evolution

The renormalization scale variation described in the previous section can be used to estimate the MHOU in any partonic cross-section of an electroproduction or hadroproduction process evaluated to a fixed order in perturbation theory. However, when computing factorized observables of the form Eqs. (3.14, 3.22), an entirely independent source of MHOU arises from the truncation of the perturbative expansion of the splitting functions (or anomalous dimensions in Mellin space) that govern the PDF evolution equations. We now show that this MHOU can again be estimated by scale variation; we will also show that this scale variation can be performed in different ways: either at the level of the anomalous dimension; or at the level of the PDFs themselves; or finally at the level of the hard-scattering partonic coefficient functions, by exploiting the fact that physical results cannot depend on the scale at which the PDF is evaluated, and so one may trade the effect of scale variation between the PDF and the hard coefficient function.

Consider a PDF $f(\mu^2)$, where μ is the scale at which the PDF is evaluated. For simplicity, in this section all the argument is presented implicitly assuming a Mellin space formalism, so that convolutions are replaced by ordinary products. Also, indices labeling different PDFs are left implicit, so our argument applies directly to the nonsinglet case but can be straightforwardly generalized to the singlet evolution and to other flavor combinations.

The scale dependence of $f(\mu^2)$ is fixed by the evolution equation

$$\mu^2 \frac{d}{d\mu^2} f(\mu^2) = \gamma(\alpha_s(\mu^2)) f(\mu^2), \tag{3.28}$$

which applies also to the general singlet case assuming that a sum over parton flavors is left implicit. The anomalous dimension admits a perturbative expansion of the form

$$\gamma(t) = \alpha_s(t)\gamma_0 + \alpha_s^2(t)\gamma_1^2 + \alpha_s^3(t)\gamma_2^3 + \dots \tag{3.29}$$

Eq. (3.28) can be integrated to give

$$f(\mu^2) = \exp\left(\int^{\mu^2} \frac{d\mu'^2}{\mu'^2} \gamma(\alpha_s(\mu'^2))\right) f_0, \tag{3.30}$$

where f_0 indicates the PDF at the initial scale μ_0 . Of course, the left-hand side of the equation is independent of this initial scale μ_0 , so the dependence can be left implicit also on

the right-hand side, by not specifying the lower limit on the integral. In practice, if the PDF f_0 were extracted from data, any change in this scale would be entirely reabsorbed by the fitting procedure.

We now observe the well-known fact that the anomalous dimension in Eq. (3.28) is a RG invariant quantity, and therefore the scale on which it depends is physical. However, this physical scale can in general be different from the renormalization scale used to determine the anomalous dimension itself (e.g. if it were determined through the renormalization of a twist-two operator). We let $\mu^2 = kQ^2$, where as in the general argument of Sect. 3.1, μ^2 is an arbitrary renormalization scale and Q^2 is a physical scale. We can make γ independent of the renormalization scale order by order in perturbation theory if we define its scale-varied counterpart in the same way as before

$$\bar{\gamma}(\alpha_s(t), \kappa) = \gamma(t) - \kappa \frac{d}{dt} \gamma(t) + \frac{1}{2} \kappa^2 \frac{d^2}{dt^2} \gamma(t) + \dots, \quad (3.31)$$

with κ given by Eq. (3.5) and $\gamma(t) = \bar{\gamma}(\alpha_s(t), 0)$, so that given the perturbative expansion Eq. (3.29) one has that

$$\begin{aligned} \bar{\gamma}(\alpha_s(t + \kappa), \kappa) &= \alpha_s(t + \kappa)\gamma_0 + \alpha_s^2(t + \kappa)(\gamma_1 - \kappa\beta_0\gamma_0) \\ &\quad + \alpha_s^3(t + \kappa)(\gamma_2 - \kappa(\beta_1\gamma_0 + 2\beta_0\gamma_1) \\ &\quad + \kappa^2\beta_0^2\gamma_0) + \dots \end{aligned} \quad (3.32)$$

is independent of κ up to higher orders terms, order by order. Note that Eq. (3.32) has the same form as Eqs. (3.25–3.27) (with $n = 1$).

We have shown that variation of the scale on which the anomalous dimension depends can be used, in the usual way, to generate higher order terms which estimate MHOUs in the expansion of the anomalous dimension itself. We now show how the same result can be obtained by scale variation at the PDF level. Inserting the result Eq. (3.32) in the solution of the evolution equations for the PDFs, Eq. (3.30), one finds that the evolution factor can be expressed as

$$\begin{aligned} &\exp\left(\int^t dt' \bar{\gamma}(\alpha_s(t' + \kappa), \kappa)\right) \\ &= \exp\left(\int^{t+\kappa} dt' \bar{\gamma}(\alpha_s(t'), \kappa)\right) \\ &= \exp\left(\left[\int^{t+\kappa} dt' \gamma(t')\right] \right. \\ &\quad \left. - \kappa \gamma(t + \kappa) + \frac{1}{2} \kappa^2 \frac{d}{dt} \gamma(t + \kappa) + \dots\right) \\ &= \left[1 - \kappa \gamma(t + \kappa) + \frac{1}{2} \kappa^2 (\gamma^2(t + \kappa) \right. \\ &\quad \left. + \frac{d}{dt} \gamma(t + \kappa)) + \dots\right] \exp\left(\int^{t+\kappa} dt' \gamma(t')\right), \end{aligned} \quad (3.33)$$

where in the first line we changed integration variable (ignoring any change in the lower limit of integration), in the second we used Eq. (3.31), and in the third we expanded the exponential perturbatively. We can now use this result to determine renormalization scale variation in the evolution directly from the scale dependence of the PDF, as in Ref. [17]. Defining a scale-varied PDF as

$$\bar{f}(\alpha_s(t + \kappa), \kappa) = \exp\left(\int^t dt' \bar{\gamma}(\alpha_s(t' + \kappa), \kappa)\right) f_0, \quad (3.34)$$

that is, as the PDF obtained by varying the renormalization scale in the anomalous dimension, then $f(t) = \bar{f}(\alpha_s(t), 0)$, and using Eq. (3.33) we find that

$$\begin{aligned} \bar{f}(\alpha_s(t + \kappa), \kappa) &= \left[1 - \kappa \gamma(t + \kappa) + \frac{1}{2} \kappa^2 (\gamma^2(t + \kappa) \right. \\ &\quad \left. + \frac{d}{dt} \gamma(t + \kappa)) + \dots\right] f(t + \kappa), \end{aligned} \quad (3.35)$$

provided only that any variation of the initial scale μ_0 due to changes in κ has been reabsorbed into the initial PDF f_0 .

Equation (3.35) is the same as the result obtained from varying the scale μ^2 at which the PDF is evaluated about the physical scale Q^2 : just as in the derivation of Eq. (3.24), this gives

$$\begin{aligned} \bar{f}(\alpha_s(t + \kappa), \kappa) &= f(t + \kappa) - \kappa \frac{d}{dt} f(t + \kappa) \\ &\quad + \frac{1}{2} \kappa^2 \frac{d^2}{dt^2} f(t + \kappa) + \dots \\ &= f(t + \kappa) - \kappa \gamma f(t + \kappa) \\ &\quad + \frac{1}{2} \kappa^2 (\gamma^2 + \frac{d}{dt} \gamma) f(t + \kappa) + \dots, \end{aligned} \quad (3.36)$$

where in the second line we used the PDF evolution equation, Eq. (3.28). Thus there is little point in varying the renormalization scale of the anomalous dimension and the scale at which the PDF is evaluated independently: provided we absorb changes in the initial scale in the initial PDF, and use the linearised solution of the evolution equation, the result (Eq. (3.35) or Eq. (3.36)) is precisely the same. This is essentially because the PDF $f(t)$ depends on only a single scale.

Equation (3.35) indicates that the κ dependence can be factorized out of the PDF. We can use this property to factor it into the hard-scattering coefficient function. Consider for example electroproduction, whose factorized structure function is given by Eq. (3.14):

$$\begin{aligned} \widehat{F}(t, \kappa) &= C(t) \bar{f}(\alpha_s(t + \kappa), \kappa) \\ &= C(t) \left[1 - \kappa \gamma(t + \kappa) + \frac{1}{2} \kappa^2 (\gamma^2(t + \kappa) \right. \end{aligned}$$

$$\begin{aligned} & + \frac{d}{dt} \gamma(t + \kappa) + \dots \Big] f(t + \kappa) \\ \equiv & \widehat{C}(t, \kappa) f(t + \kappa), \end{aligned} \quad (3.37)$$

where in the second line we used the expansion Eq. (3.35), and the third line should be viewed as the definition of the scale-varied coefficient function $\widehat{C}(t + \kappa, \kappa)$. Moreover, given the relation

$$\frac{d}{dt} \gamma(\alpha_s) = \beta(\alpha_s) \frac{d\gamma}{d\alpha_s}, \quad (3.38)$$

and then using the perturbative expansions of the beta function β , the anomalous dimension γ , and the coefficient function C , Eqs. (3.3), (3.29), and (3.18), respectively, one finds

$$\begin{aligned} \widehat{C}(t, \kappa) = & c_0 + \alpha_s(t)(c_1 - \kappa\gamma_0) + \alpha_s^2(t) \\ & \times \left(c_2 - \kappa(\gamma_0 c_1 + \gamma_1 c_0) + \frac{1}{2} \kappa^2 \gamma_0(\gamma_0 + \beta_0) c_0 \right) \\ & + \dots \end{aligned} \quad (3.39)$$

Note that this result for $\widehat{C}(t, \kappa)$ is not the same as $\overline{C}(t + \kappa, \kappa)$, Eq. (3.20). The reason is that $\overline{C}(t + \kappa, \kappa)$ is obtained from the variation of the renormalization scale of the hard coefficient function, and can be used to estimate the MHO in the perturbative expansion of the coefficient function, while $\widehat{C}(t, \kappa)$ is obtained from the variation of the renormalization scale of the anomalous dimension, and can be used to estimate the MHO in the perturbative evolution of the PDF. We have obtained the former from RG invariance of the hard cross-section, and the latter from RG invariance of the anomalous dimension. However, Eq. (3.37) can be equivalently viewed as expressing the fact that the physically observable structure function cannot depend on the scale at which the PDF is evaluated in the factorized expression, usually referred to as factorization scale: provided we absorb changes in the initial scale in the initial PDF, varying the scale of the anomalous dimension is identical to varying the scale of the PDF.

It is customary to refer to the scale variation which estimates MHO in the coefficient function as renormalization scale variation: this corresponds to evaluating $\overline{C}(t + \kappa, \kappa)$ in Eq. (3.20). The scale variation which estimates MHO in the anomalous dimension, and corresponds to $\widehat{C}(t + \kappa, \kappa)$ in Eq. (3.39), is usually called instead factorization scale variation. This terminology is used for example by the Higgs Cross-Section working group [18] and more generally within the context of LHC physics; in the older DIS literature the same terminology has a somewhat different meaning, as we shall discuss in Sect. 3.4 below.

The previous discussion entails that in practice there are (at least) three different ways of estimating the MHO associated to the PDF evolution in terms of the anomalous dimen-

sion at fixed order in perturbation theory by means of scale variations:

- (A) The renormalization scale of the anomalous dimension can be varied directly, using Eq. (3.32). This approach works well provided that the initial PDF f_0 is refitted, but if it is held fixed care must be taken to absorb scale variations of the initial scale into the initial PDF. This method was used for DIS renormalization scale variations in many older papers, see e.g. Refs. [19–21]). It has the disadvantage that it requires refitting the PDF as the scale is varied, which is cumbersome for most applications.
- (B) The scale at which the PDF is evaluated can be varied, either analytically or numerically, using Eq. (3.36). This is in many ways the simplest method, as the initial PDF remains unchanged, while only the PDF is involved so the result is manifestly universal. Furthermore it is easily adapted to a variable flavor number scheme (VFNS), since the MHOs in the PDFs with different numbers of active flavors can each be estimated separately. The numerical method was employed in [17], in the context of small x resummation. It has the disadvantage that if one wishes to estimate the impact on a given physical observable one needs to first generate the scale-varied PDF, before combining it with the hard coefficient function.
- (C) The scale at which the PDF is evaluated is varied, but the compensating scale-dependent terms are factorized into the coefficient function using for example Eq. (3.39). This factorization scale variation is most commonly used when evaluating a new process using an established PDF set, e.g. in studies of LHC processes (as in Ref. [18]) since it has the advantage that it can be implemented directly using an external interpolated PDF set (such as provided by LHAPDF [22]). It has the conceptual disadvantage that the universality of the variation is obscured, since the scale dependent terms are mixed in the expansion of the coefficient function (this is particularly complicated in a VFNS, where the coefficient functions also depend on heavy quark masses), and the practical disadvantage that it requires the evaluation of new contributions to the coefficient function involving additional convolutions. Also, it can be impractical in situations where higher order corrections are difficult to evaluate precisely due to numerical issues.

Note that whereas these methods are in principle completely equivalent, they can differ by subleading terms according to the convention used to truncate the perturbation expansion. Indeed, in method (A) the expansion of the anomalous dimension is truncated, but higher order terms in the exponentiation may be retained depending on the form

of the solution to the evolution equations adopted; in method (B) the exponential has been expanded (see Eq. (3.33)) so the result is the same as would be obtained with a linearized solution of the evolution equation; while in method (C) cross-terms between the expansion of linearized evolution and coefficient function expansion have also been dropped (compare Eq. (3.37) with Eq. (3.39)). However, since the differences always involve higher order terms, each method can be regarded as giving an equally valid estimate of the MHOUs in the perturbative evolution: differences between methods should be viewed as the uncertainty on the MHOUs itself when estimated by scale variation.

3.4 Double scale variations

We now discuss the combination of the two independent scale variations of Sects. 3.2 and 3.3, respectively estimating MHOUs in the hard cross-section and in perturbative evolution, thereby deriving master formulae for scale variation up to NNLO which will then be used in the subsequent sections. For completeness, we will also discuss different options for scale variation which have been considered in the literature, and clarify some terminological mismatches, especially between the older studies of DIS and the more recent applications to LHC processes.

3.4.1 Electroproduction

Consider first the more general factorization of an electroproduction cross-section, such as a DIS structure function:

$$\overline{F}(Q^2, \mu_f^2, \mu_r^2) = \overline{C}(\alpha_s(\mu_r^2), \mu_r^2/Q^2) \otimes \overline{f}(\alpha_s(\mu_f^2), \mu_f^2/Q^2), \tag{3.40}$$

where here and in the following we adopt the (standard) terminology that we introduced in Sect. 3.3, and the viewpoint which corresponds to option (B) of that section: μ_r denotes the renormalization scale, whose dependence is entirely contained in the hard coefficient function \overline{C} (as in Eq. (3.15)), and whose variation estimates MHOUs in its expansion; while μ_f denotes the factorization scale, whose dependence is entirely contained in the PDF (as in Eq. (3.34)), and whose variation estimates MHOUs in the expansion of the anomalous dimension (or equivalently the splitting functions). In the following, as in Sect. 3.3, we will omit the convolution as well as the parton indices.

Note that again, as in Eq. (3.15), and then in Eqs. (3.23), (3.31), and (3.36), the dependence on the scales μ_f and μ_r should be understood in the sense of the RG equation: the structure function does not depend on them, but as the scales are varied there remains a subleading dependence which estimates the MHOUs. As already mentioned, this notation, while

standard in the context of RG equations, is somewhat unusual in the context of factorization, where instead it is more customary to omit the scale dependence of the physical observable.

Given that the structure function $\overline{F}(Q^2, \mu_f^2, \mu_r^2)$ factorizes into the hard coefficient function and the PDF, the factorization and renormalization scales μ_f and μ_r can be chosen completely independently; the scale dependence will also factorize. Explicitly, we define

$$\begin{aligned} \mu_f^2 &= k_f Q^2, \quad \mu_r^2 = k_r Q^2, \quad \text{with } t_f = t + \kappa_f, \\ t_r &= t + \kappa_r, \end{aligned} \tag{3.41}$$

and then $\kappa_f = \ln k_f, \kappa_r = \ln k_r$. In terms of these variables, the factorized structure function will be given by

$$\overline{F}(t, \kappa_f, \kappa_r) = \overline{C}(t_r, \kappa_r) \overline{f}(t_f, \kappa_f), \tag{3.42}$$

where, as in Sects. 3.2 and 3.3, the scale-varied PDF and coefficient functions are

$$\begin{aligned} \overline{f}(t_f, \kappa_f) &= f(t_f) - \kappa_f \frac{d}{dt} f(t_f) + \frac{1}{2} \kappa_f^2 \frac{d^2}{dt^2} f(t_f) + \dots, \\ \overline{C}(t_r, \kappa_r) &= C(t_r) - \kappa_r \frac{d}{dt} C(t_r) + \frac{1}{2} \kappa_r^2 \frac{d^2}{dt^2} C(t_r) + \dots, \end{aligned} \tag{3.43}$$

where $f(t_f) \equiv \overline{f}(t_f, 0)$ and $C(t_r) \equiv \overline{C}(t_r, 0)$ stand for the PDF and the coefficient function evaluated at the central scale, $\mu_f^2 = Q^2$ and $\mu_r^2 = Q^2$, respectively. Recalling that $\frac{\partial}{\partial t} \sim \mathcal{O}(\alpha_s)$, the structure function is therefore given by

$$\begin{aligned} \overline{F}(t, \kappa_f, \kappa_r) &= C(t_r) f(t_f) \\ &\quad - \left(\kappa_r \frac{d}{dt} C(t_r) f(t_f) + \kappa_f C(t_r) \frac{d}{dt} f(t_f) \right) \\ &\quad + \frac{1}{2} \left(\kappa_r^2 \frac{d^2}{dt^2} C(t_r) f(t_f) \right. \\ &\quad \left. + 2\kappa_r \kappa_f \frac{d}{dt} C(t_r) \frac{d}{dt} f(t_f) + \kappa_f^2 C(t_r) \frac{d^2}{dt^2} f(t_f) \right) \\ &\quad + \mathcal{O}(\alpha_s^3). \end{aligned} \tag{3.44}$$

From this expression, it follows that scale variations with respect to κ_f can be determined by taking derivatives with respect to t_f while holding t_r fixed and vice-versa, so one has

$$\begin{aligned} \overline{F}(t, \kappa_f, \kappa_r) &= F(t_f, t_r) - \left(\kappa_f \frac{\partial F}{\partial t_f} \Big|_{t_r} + \kappa_r \frac{\partial F}{\partial t_r} \Big|_{t_f} \right) \\ &\quad + \frac{1}{2} \left(\kappa_f^2 \frac{\partial^2 F}{\partial t_f^2} \Big|_{t_r} \right. \\ &\quad \left. + 2\kappa_f \kappa_r \frac{\partial^2 F}{\partial t_f \partial t_r} + \kappa_r^2 \frac{\partial^2 F}{\partial t_r^2} \Big|_{t_f} \right) + \dots. \end{aligned} \tag{3.45}$$

In other words, we can think of the two variations as being generated by $\kappa_f \frac{\partial}{\partial t_f}$ and $\kappa_r \frac{\partial}{\partial t_r}$ respectively.

We can equivalently treat the factorization scale variation using method (C) of the previous section, and thus factorize both scale variations into the coefficient function, as done in Eq. (3.39). In the case of electroproduction, inserting the expansions of Eq. (3.18) in Eq. (3.44) one obtains

$$\overline{F}(t, \kappa_f, \kappa_r) = \widehat{C}(\alpha_s(t_r), \kappa_f, \kappa_r) f(t_f), \tag{3.46}$$

with now all dependence on κ_r and κ_f encoded into a redefined coefficient function:

$$\begin{aligned} \widehat{C}(\alpha_s(t_r), \kappa_f, \kappa_r) &\equiv c_0 + \alpha_s(t_r)c_1 - \alpha_s(t_f)\kappa_f c_0\gamma_0 \\ &+ \alpha_s(t_r)^2(c_2 - \kappa_r \beta_0 c_1) - \alpha_s(t_r)\alpha_s(t_f)\kappa_f c_1\gamma_0 \\ &+ \alpha_s^2(t_f)(-\kappa_f c_0\gamma_1 + \frac{1}{2}\kappa_f^2 c_0\gamma_0(\beta_0 + \gamma_0)) + \dots \\ &= c_0 + \alpha_s(t_r)(c_1 - \kappa_f c_0\gamma_0) + \alpha_s^2(t_r) \\ &\times (c_2 - \kappa_r \beta_0 c_1 - \kappa_f (c_1\gamma_0 + c_0\gamma_1) \\ &+ \frac{1}{2}\kappa_f^2 c_0\gamma_0(\gamma_0 - \beta_0) + \kappa_f \kappa_r \beta_0 c_0\gamma_0) + \dots \end{aligned} \tag{3.47}$$

up to terms of $\mathcal{O}(\alpha_s^3(t_r))$, given that one can change the scale that enters the coupling using

$$\alpha_s(t_f) = \alpha_s(t_r) + (\kappa_f - \kappa_r)\beta_0\alpha_s^2(t_r) + \dots \tag{3.48}$$

Note that in the expression for \widehat{C} the coupling constant is always evaluated at the renormalization scale μ_r , and that for $\kappa_r = \kappa_f = 0$ one gets back the original perturbative expansion Eq. (3.18).

However, especially in the context of PDF determinations, as opposed to the situation in which a pre-computed PDF set is being used, it is rather more convenient to use either of methods (A) or (B) from Sect. 3.3 when estimating the MHOU in the scale dependence of the PDF, since this can be done without reference to any particular process. We can then determine the universal μ_f variation by varying the scale in the PDF evolution, as done for instance in Eq. (3.32) or Eq. (3.36), while instead the process-dependent μ_r variation is estimated by varying the renormalization scale in the coefficient function, as done in Eq. (3.20), or Eq. (3.27) in the case of hadronic processes.

Note that since all scale-varied terms ultimately derive from the scale dependence of the universal QCD coupling $\alpha_s(\mu^2)$, it is reasonable to treat the independent scale variations of μ_f and μ_r symmetrically, e.g. by varying in the range $|\kappa_f|, |\kappa_r| \leq \ln 4$. Indeed, this symmetry is an advantage of the method: we use the same variation for estimating all MHOU's. Since μ_f and μ_r can each be varied independently, a simple option is to perform the double scale variations by considering the five scale choices $(\kappa_f, \kappa_r) = (0, 0), (\pm \ln 4, 0), (0, \pm \ln 4)$. We will refer to

this as 5-point scale variation; alternative schemes will be considered in the next section.

Note finally that if we set the renormalization and factorization scales in Eq. (3.40) to be equal to each other, $\mu_f^2 = \mu_r^2 = \tilde{\mu}^2$, we have the factorization

$$\widetilde{F}(Q^2, \tilde{\mu}^2) = \widetilde{C}(\alpha_s(\tilde{\mu}^2), \tilde{\mu}^2/Q^2) f(\tilde{\mu}^2). \tag{3.49}$$

In most of the earlier papers, mainly concerned with DIS structure functions, e.g. [19,20,23–25], the scale $\tilde{\mu}^2$ was termed the factorization scale: this originates in the earliest papers on the OPE. However, in our current terminology it corresponds to both renormalization and factorization scales taken equal to each other. Likewise, in the earlier papers what here we call the factorization scale μ_f was referred to as the renormalization scale. Here, to avoid confusion, we will call $\tilde{\mu}^2$ in Eq. (3.49) the scale of the process.

For clarity the different nomenclatures for the various scales used in the earlier papers, and in more modern work (and in this paper), are summarized in Table 1.

Consider now the effect on the structure function of varying the scale of the process. As before, we define $\tilde{\kappa} = \ln \tilde{\mu}^2/Q^2$ and write

$$\widetilde{F}(t + \tilde{\kappa}, \tilde{\kappa}) = \widetilde{C}(\alpha_s(t + \tilde{\kappa}), \tilde{\kappa}) f(t + \tilde{\kappa}). \tag{3.50}$$

Now the renormalization group invariance of the cross-section [i.e. Eq. (3.4)] requires a cancellation between scale variations in the coefficient function and the PDF: with $F(t) \equiv \widetilde{F}(t, 0)$,

$$\begin{aligned} \widetilde{F}(t + \tilde{\kappa}, \tilde{\kappa}) &= F(t + \tilde{\kappa}) - \tilde{\kappa} \frac{d}{dt} F(t + \tilde{\kappa}) \\ &+ \frac{1}{2}\tilde{\kappa}^2 \frac{d^2}{dt^2} F(t + \tilde{\kappa}) + \dots \\ &= C f - \tilde{\kappa} \left(\frac{d}{dt} C + \gamma C \right) f \\ &+ \frac{1}{2}\tilde{\kappa}^2 \left(\frac{d^2}{dt^2} C + 2\gamma \frac{d}{dt} C + C \frac{d}{dt} \gamma + C \gamma^2 \right) f \\ &+ \dots \end{aligned} \tag{3.51}$$

where the first line is the same as Eq. (5.8) in Ref. [17] while in the second line we used Eq. (3.36) for scale variation of the PDF. Then, expanding in the usual way, we find that

$$\begin{aligned} \overline{C}(t + \tilde{\kappa}, \kappa) &= c_0 + \alpha_s(t + \tilde{\kappa})(c_1 - \tilde{\kappa} c_0\gamma_0) \\ &+ \alpha_s^2(t + \tilde{\kappa})(c_2 - \tilde{\kappa}(\beta_0 c_1 + c_1\gamma_0 + c_0\gamma_1) \\ &+ \frac{1}{2}\tilde{\kappa}^2 c_0\gamma_0(\beta_0 + \gamma_0)) + \dots \end{aligned} \tag{3.52}$$

which indeed coincides with the expression for what is referred to as factorization scale variation in this earlier lit-

Table 1 Nomenclatures for the different scale variations used in some of the earlier papers (mainly in the context of DIS), and in more recent work (mainly in the context of hadronic processes), as discussed in detail in the text. The ‘modern’ terminology is adopted throughout this paper

Scale	MHOU	‘Traditional’ name [19,20,23–25]	‘Modern’ name [26], [PDG]
μ_r	In hard xsec	–	Renormalization scale
μ_f	In PDF evolution	Renormalization scale	Factorization scale
$\tilde{\mu}$	In physical xsec	Factorization scale	Scale of the process

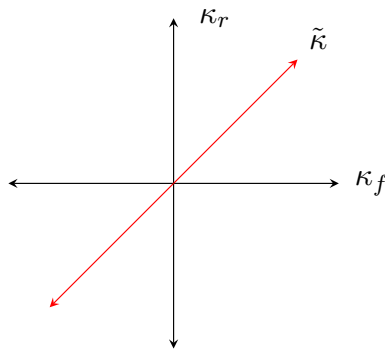


Fig. 1 The two-dimensional space of scale variations for a single process: κ_r is the renormalization scale (giving the MHOU in the hard cross-section), κ_f is the factorization scale (giving the MHOU in the evolution of the PDF) and $\tilde{\kappa}$ is the variation of the scale of the process (called factorization scale variation in the earlier literature), obtained by setting $\kappa_f = \kappa_r$

erature: see e.g. Ref. [26], Eq. (2.17). Therefore, varying the scale of the process mixes together the scale dependence in the coefficient function and the scale dependence in the PDF: indeed, if in Eq. (3.47) we set $\kappa_f = \kappa_r = \tilde{\kappa}$, it reduces to Eq. (3.52).

Clearly, variations of $\tilde{\mu}^2$ are not independent of the variations of μ_f^2 or μ_r^2 : rather they are generated by $\tilde{\kappa} (\frac{\partial}{\partial t_f} + \frac{\partial}{\partial t_r})$, so they correspond to directions along the diagonal in the space of κ_f and κ_r , see Fig. 1. In the earlier literature, MHOUs were estimated by combining renormalization scale variation with this latter variation, namely by varying $\tilde{\mu}^2$ and μ_f^2 : see e.g. Refs. [19,20]. This however has the disadvantage of generating large scale ratios: performing variations of $\tilde{\mu}^2$ and μ_f^2 sequentially we can obtain $\kappa_f = 2 \ln 4$, because

$$\tilde{\kappa} \left(\frac{\partial}{\partial t_f} + \frac{\partial}{\partial t_r} \right) + \kappa_f \frac{\partial}{\partial t_f} = (\tilde{\kappa} + \kappa_f) \frac{\partial}{\partial t_f} + \tilde{\kappa} \frac{\partial}{\partial t_r}. \quad (3.53)$$

A way of avoiding these large ratios was constructed in Ref. [26]: first do the scale variation of Eq. (3.52), but then substitute

$$c_2 \rightarrow c_2 - (\kappa_r - \kappa_f) \beta_f c_1 = c_2 - \left(\ln \mu_f^2 / \mu_r^2 \right) \beta_0 c_1, \quad (3.54)$$

where care must be taken to use the correct argument of α_s in each term. Indeed, this procedure then agrees with Eq. (3.46)

given that

$$\kappa_f \frac{\partial}{\partial t_f} + \kappa_r \frac{\partial}{\partial t_r} = \kappa_f \left(\frac{\partial}{\partial t_f} + \frac{\partial}{\partial t_r} \right) + (\kappa_r - \kappa_f) \frac{\partial}{\partial t_r}. \quad (3.55)$$

3.4.2 Hadronic processes

Consider now the case of hadronic process as in Eq. (3.22). For these processes, the factorization has the general form

$$\overline{\Sigma}(t_f, t_r, \kappa_f, \kappa_r) = \overline{H}(\alpha_s(t_r), \kappa_r) \otimes (\overline{f}(t_f, \kappa_f) \otimes \overline{f}(t_f, \kappa_f)). \quad (3.56)$$

The hard coefficient function will have the same expansion as Eq. (3.27). Just as for electroproduction, it is possible to factorize variations of κ_f into the hard coefficient functions: then

$$\overline{\Sigma}(t_f, t_r, \kappa_f, \kappa_r) = \widehat{H}(\alpha_s(t_r), \kappa_r, \kappa_f) \otimes (f(t_f) \otimes f(t_f)), \quad (3.57)$$

where (using as above Mellin space, to avoid the convolutions), one finds

$$\begin{aligned} \widehat{H} = & \alpha_s^n(t_r) h_0 + \alpha_s^{n+1}(t_r) (h_1 - \kappa_r \beta_0 h_0) \\ & - 2\alpha_s^n(t_r) \alpha_s(t_f) \kappa_0 h_0 \gamma_0 \\ & + \alpha_s^{n+2}(t_r) (h_2 - \kappa_2 (n\beta_1 h_0 + (n+1)\beta_0 h_1) \\ & + \frac{1}{2} \kappa_2^2 n(n+1) \beta_0^2 h_1) \\ & - \alpha_s^{n+1}(t_r) \alpha_s(t_f) (\kappa_0 (h_1 - \kappa_2 \beta_0 h_0) 2\gamma_0) \\ & + \alpha_s^n(t_r) \alpha_s^2(t_f) \left(-\kappa_0 h_0 2\gamma_1 + \frac{1}{2} \kappa_0^2 h_0 2\gamma_0 (\beta_0 + 2\gamma_0) \right) \\ & + \dots \end{aligned} \quad (3.58)$$

However these expressions are even more cumbersome than in the case of electroproduction, thereby demonstrating the greater clarity of methods (A) or (B) in determining the dependence on the scale μ_f . By adopting one of these two methods, we can determine the MHOU in a hadronic process through independent variations of the factorization scale μ_f and the renormalization scale μ_r in just the same way as we estimated the MHOU in the deep inelastic structure function in the previous section.

3.5 Multiple scale variations

We finally consider simultaneous scale variation in a pair of processes: for instance the electroproduction process of Sect. 3.4.1 and a hadronic process as in Sect. 3.4.2. Clearly, the PDF is universal, but the coefficient functions are process-dependent. It follows that while the scale variations of κ_r in the two coefficient functions will be totally independent, the scale variation κ_f of the PDF will be correlated between these two processes.

The degree of this correlation is somewhat subtle: indeed, κ_f generates MHO terms in anomalous dimensions, but the anomalous dimension matrix has several independent eigenvalues (two singlet and one nonsinglet which at NLO and beyond further splits into C-even and C-odd). Hence in principle one should introduce an independent factorization scale variation for each of these components, which is then fully correlated across all processes. For the time being, we will perform fully correlated variations of the factorization scale. This is an approximation, which may not be accurate particularly for processes which depend on PDFs whose evolution is controlled by different anomalous dimensions (such as, say, the singlet and the isospin triplet). We will comment further on this approximation in the sequel.

Now, considering both processes together, we have three independent scales to vary, μ_f , μ_{r_1} , and μ_{r_2} , where μ_{r_1} is the renormalization scale for the deep inelastic process, and μ_{r_2} is the renormalization scale for the hadronic process. The relation of the factorization scale μ_f to the physical scale of each process (whatever that is) is the same for both processes, since the PDFs are universal. Thus if we vary all scales independently by a factor two about their central value we end up with seven scale choices. We can think of the additional renormalization scale as an extra dimension in the space of possible scale variations.

By trivial generalization for p independent processes π_a , $a = 1, \dots, p$, we will have $p + 1$ independent scale parameters $\mu_f, \mu_{r_1}, \dots, \mu_{r_p}$ corresponding to a total of $3+2p$ scale variations. Writing $\kappa_{r_a} = \ln \mu_{r_a}^2 / Q^2$ with $a = 1, \dots, p$, the traditional range of variation of $\kappa_f, \kappa_{r_1}, \dots, \kappa_{r_p}$ would then be defined by

$$|\kappa_f| \leq \ln 4, \quad |\kappa_{r_a}| \leq \ln 4, \quad a = 1, \dots, p.$$

Clearly all prescriptions constructed in this way will be symmetrical in the different scales.

We now see why, for the determination of MHOUs in PDFs, it is advantageous to work with the independent scales κ_f, κ_{r_a} , $a = 1, \dots, p$ rather than with the traditional factorization scales $\tilde{\kappa}$ used in the older treatments of scale variation: while the scale κ_f used to estimate MHOUs in the PDF evolution is universal, the scales κ_{r_a} used to estimate MHOUs in the hard cross-sections are instead process-dependent. We

can therefore only define process scales $\tilde{\kappa}$ by either introducing artificial correlations between the scales of the hard cross-sections for different processes (which would result in underestimated MHOUs in the hard cross-sections), or else by sacrificing universality of the PDFs, with uncorrelated evolution uncertainties for different processes (which would result in overestimated MHOUs from PDF evolution). Neither of these options is very satisfactory, though we consider the latter briefly in Sect. 4.3 below, where it gives rise to asymmetric scale-variation prescriptions.

4 Scale variation prescriptions for the theory covariance matrix

Having set out a general formalism for the inclusion of MHOUs through a theory covariance matrix, based on assuming a distribution of shifts between a theory calculation at finite perturbative order and the true all-order value (Sect. 2), and having discussed how scale variation can be used to produce estimates for such shifts (Sect. 3), we now provide an explicit prescription for the construction of a theory covariance matrix from scale variation. Because of the intrinsic arbitrariness involved in the procedure, we actually propose several alternative prescriptions, which will be then validated in the next section by studying cases in which the next perturbative order is in fact known. We will also assess the impact at the PDF fit level of varying the prescription used for constructing the theory covariance matrix.

We consider a situation in which we have p different types of processes $\pi_a = \{i_a\}$, where i_a labels the data points belonging to the a -th process and $a = 1, \dots, p$. Each of the p processes is characterized by a factorization scale μ_f (associated with the PDFs) and a renormalization scale μ_{r_a} (associated with the hard coefficient functions), to be understood in the sense of the ‘modern’ terminology in Table 1. We will perform scale variation of both scales following Sect. 3.4, by taking them as independent, as discussed in that section. When considering a pair of different processes, as explained in Sect. 3.5, we assume the variations of μ_{r_a} to be uncorrelated among them, while those of μ_f are taken to be fully correlated.

The theory covariance matrix is then constructed by averaging outer products of the shifts with respect to the central scales, given for the a -th process as

$$\Delta_{i_a}(\kappa_f, \kappa_{r_a}) \equiv T_{i_a}(\kappa_f, \kappa_{r_a}) - T_{i_a}(0, 0), \quad (4.1)$$

over points in the space of scales. Here, as before, we have defined $\kappa_{r_a} = \ln k_{r_a} = \ln \mu_{r_a}^2 / Q^2$ and $\kappa_f = \ln k_f = \ln \mu_f^2 / Q^2$. In Eq. (4.1), $T_{i_a}(\kappa_f, \kappa_{r_a})$ indicates the theoretical prediction evaluated at these scales with $T_{i_a}(0, 0)$ being

the central theory prediction, and the index i_a running over all data points corresponding to process a .

We assume here that all scale variations correspond to the same range

$$|\kappa_f| \leq w, \quad |\kappa_{r_a}| \leq w, \quad a = 1, \dots, p,$$

for some w (typically $w = \ln 4$, as in Eq. (3.5)). In practice, in each prescription the three points $\kappa = 0, \pm w$ are sampled for each scale. The theory covariance matrix is then

$$S_{ij} = N_m \sum_{V_m} \Delta_{i_a}(\kappa_f, \kappa_r) \Delta_{i_b}(\kappa_f, \kappa_r) \quad (4.2)$$

where $i_a \in \pi_a$ and $i_b \in \pi_b$ indicate two data points, possibly corresponding to different processes π_a and π_b , m labels the prescription, V_m is the set of scale points to be summed over in the given prescription, and N_m is a normalization factor, both to be determined. Different prescriptions to construct the theory covariance matrix S_{ij} vary in the set of combination of scales which are summed over in Eq. (4.2), as we will discuss below.

Because Eq. (4.2) is a sum of outer products of shifts, the theory covariance matrix S_{ij} is positive semi-definite by construction. To demonstrate this, consider a real vector v_i : then it follows that

$$\sum_{ij} v_i S_{ij} v_j = N_m \sum_{V_m} \left(\sum_i v_i \Delta_i \right)^2 \geq 0. \quad (4.3)$$

Note however that because the number of elements of V_m is finite, S_{ij} will generally be singular, since for any vector z_j which is orthogonal to the space S spanned by the set of vectors $\{\Delta_{i_a}(\kappa_f, \kappa_{r_a}) : \kappa_f, \kappa_{r_a} \in V_m\}$, $S_{ij} z_j = 0$. This property will be important when we come to validate the covariance matrix in the following section, by constructing the set of orthonormal eigenvectors e_i^α which span the space S .

It is interesting to note that the diagonalization of \widehat{S}_{ij} can be rephrased in terms of nuisance parameters of the systematic uncertainties associated with the MHO. For example, following the notation of Appendix A.2 of Ref. [27], the absolute correlated uncertainties $\beta_{i,\alpha}$ may be expressed in terms of the eigenvalues and eigenvectors of the normalised covariance matrix $\widehat{S}_{ij} = \sum_{\alpha=1}^{N_{\text{sub}}} (s^\alpha)^2 e_i^\alpha e_j^\alpha$ as

$$\beta_{i,\alpha} = T_i^{\text{NLO}} s^\alpha e_i^\alpha, \quad (4.4)$$

for $\alpha = 1, \dots, N_{\text{sub}}$. An algorithm for constructing the eigenvectors e_i^α from the shifts induced by scale variation is given in Appendix A. This way of looking at the theory covariance matrix might be useful in that the nuisance parameters can be interpreted in terms of missing higher order contributions. For instance, the values of the nuisance parameters

which optimize the agreement between data and theory are the most likely guess for MHO terms which is favored by the data, everything else being equal.

We now consider various prescriptions. Because S_{ij} will in general span the full set of data points, we must consider both the case in which points i, j in Eq. (4.2) belong to the same process (“single process prescription”) and the case in which they belong to two different processes (“multiple process prescription”). We first discuss the case of symmetric scale variation, in which the two scales are varied independently, and then the case in which the two scales are varied in a correlated way, the latter scenario being equivalent to varying the “scale of the process” (in the sense of Table 1), thereby leading to asymmetric prescriptions as already mentioned in Sect. 3.5.

4.1 Symmetric prescriptions for individual processes

We consider first the prescriptions for when there is just a single process, that is, $p = 1$. In this case, there are at most two independent scales, the factorization and renormalization scales κ_f and κ_r . The theory covariance matrix is then constructed as

$$S_{ij} = n_m \sum_{v_m} \Delta_i(\kappa_f, \kappa_r) \Delta_j(\kappa_f, \kappa_r), \quad (4.5)$$

where again v_m represents the set of points to be summed over in the given prescription, limited here to points in the space of the two scales κ_f and κ_r , and n_m is the normalization factor. Let s be the number of independent scales being varied (so $s = 1$ or $s = 2$), and m be the number of points in the variation (so m is the number of elements of v_m): a given scheme is then usually described as an ‘ $(m + 1)$ -point scheme’. Note that we do not include in v_m trivial points for which Δ_i vanishes (which in practice means the single point $\kappa_f = \kappa_r = 0$), since these do not contribute to the sum.

The normalization factor n_m in Eq. (4.5) is determined by averaging over the number of points associated with the variation of each scale, and adding the contributions from variation of independent scales. This means that

$$n_m = s/m. \quad (4.6)$$

We consider three different prescriptions, represented schematically in Fig. 2.

- **5-point** We vary κ_f keeping $\kappa_r = 0$ and vice versa, so $v_4 = \{(\pm; 0), (0; \pm)\}$, where the pairs denote the values of the two independent scales $(\kappa_f; \kappa_r)$. Then $s = 2, m = 4$, and the normalisation is $n_4 = 1/2$. This definition implies that we can average over the two nontrivial values

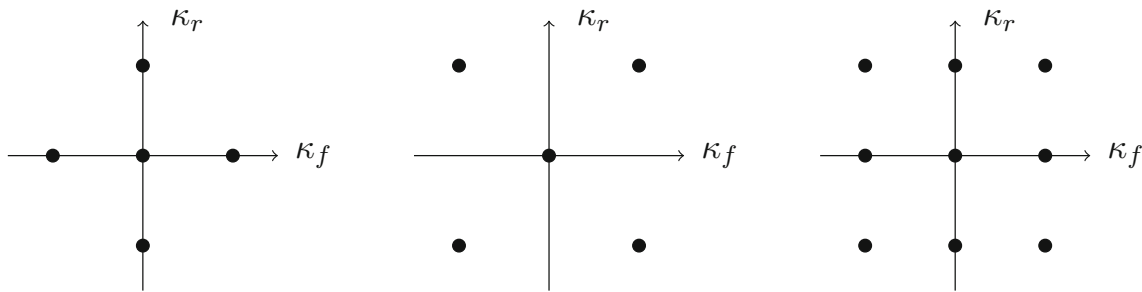


Fig. 2 Symmetric prescriptions for a single process, indicating the sampled values for the factorization scale κ_f and renormalization scale κ_r in each case. The origin of coordinates corresponds to the central

scales $\kappa_f = \kappa_r = 0$. We show the three prescriptions 5-point (left), $\bar{5}$ -point (center) and 9-point (right)

of the each scale in turn, and add the results:

$$S_{ij}^{(5pt)} = \frac{1}{2} \left\{ \Delta_i^{+0} \Delta_j^{+0} + \Delta_i^{-0} \Delta_j^{-0} + \Delta_i^{0+} \Delta_j^{0+} + \Delta_i^{0-} \Delta_j^{0-} \right\}, \tag{4.7}$$

where we have adopted the abbreviated notation $\Delta_i^{+0} = \Delta_i(+w, 0)$, $\Delta_i^{0-} = \Delta_i(0, -w)$, etc. for the shifts.

Note that the variations of κ_f and κ_r are added in quadrature since they are independent: this is why it is important to make sure that the variations are indeed independent, as is the case for renormalization and factorization scales, as discussed in Sect. 3.4.

- **$\bar{5}$ -point** This is an alternative 5-point prescription, which is basically the complement of 5-point: $\bar{v}_4 = \{(\pm; \pm)\}$, where $(\pm; \pm)$ are assumed uncorrelated, i.e. 4 independent points.

The counting is the same as for 5-point: $s = 2$, $m = 4$ and again $\bar{n}_4 = 1/2$:

$$S_{ij}^{(\bar{5}pt)} = \frac{1}{2} \left\{ \Delta_i^{++} \Delta_j^{++} + \Delta_i^{--} \Delta_j^{--} + \Delta_i^{+-} \Delta_j^{+-} + \Delta_i^{-+} \Delta_j^{-+} \right\}. \tag{4.8}$$

As before, the two scales are varied in a manifestly independent way.

- **9-point** Here we vary κ_f and κ_r completely independently, giving the union of the 5-point and $\bar{5}$ -point prescriptions: $v_8 = v_4 \oplus \bar{v}_4$. Now $s = 2$, $m = 8$ and $n_8 = 1/4$, and the theory covariance matrix is given by

$$S_{ij}^{(9pt)} = \frac{1}{4} \left\{ \Delta_i^{+0} \Delta_j^{+0} + \Delta_i^{-0} \Delta_j^{-0} + \Delta_i^{0+} \Delta_j^{0+} + \Delta_i^{0-} \Delta_j^{0-} + \Delta_i^{++} \Delta_j^{++} + \Delta_i^{+-} \Delta_j^{+-} + \Delta_i^{-+} \Delta_j^{-+} + \Delta_i^{--} \Delta_j^{--} \right\}. \tag{4.9}$$

4.2 Symmetric prescriptions for multiple processes

Now we consider multiple processes, i.e. $p > 1$, with scale variations either uncorrelated or partially correlated. In Eq. (4.2), the set V_m now involves possible variations of the $p+1$ scales $\kappa_f, \kappa_{r_1}, \dots, \kappa_{r_p}$, where κ_{r_a} indicates the renormalization scale for process $a = 1, \dots, p$. This implies that now V_m is a much bigger set than v_m . However any given element of S_{ij} in Eq. (4.2) can involve at most two different processes, π_a and π_b , so to compute this element we can simply ignore the other processes. Consequently, it is sufficient to consider $p = 2$, since generalization to $p > 2$ will then be straightforward.

For a given pair of processes, say π_1 and π_2 , the covariance matrix has diagonal elements $S_{i_1 j_1}, S_{i_2 j_2}$ and off-diagonals $S_{i_1 j_2} = S_{j_2 i_1}$, where as above the extra subscript indicates the process: $i_1, j_1 \in \pi_1, i_2, j_2 \in \pi_2$. Thus one can write

$$S_{ij} = \begin{pmatrix} S_{i_1 j_1} & S_{i_1 j_2} \\ S_{i_2 j_1} & S_{i_2 j_2} \end{pmatrix}. \tag{4.10}$$

Consider first the diagonal blocks $S_{i_1 j_1}$ and $S_{i_2 j_2}$. Adding process π_2 cannot change the theoretical uncertainty in process π_1 , although the two uncertainties may be correlated. Consequently $S_{i_1 j_1}$ and $S_{i_2 j_2}$ are each given by the same expression as in the single-process case, Eq. (4.5), so we must have

$$S_{i_1 j_1} = N_m \sum_{V_m} \Delta_{i_1}(\kappa_f, \kappa_{r_1}) \Delta_{j_1}(\kappa_f, \kappa_{r_1}) = n_m \sum_{v_m} \Delta_{i_1}(\kappa_f, \kappa_{r_1}) \Delta_{j_1}(\kappa_f, \kappa_{r_1}). \tag{4.11}$$

This can only be true if the set of points v_m in Eq. (4.5) is a subset of the set V_m in Eq. (4.2): so when for example computing $S_{i_1 j_1}$, Δ_{i_1} and Δ_{j_1} depend only on the scales κ_f and κ_{r_1} associated with π_1 , and are independent of the scale κ_{r_2} associated with π_2 . Consequently, when we sum over V_m in Eq.(4.2), performing the trivial sum over κ_{r_2} must reduce V_m to its subset v_m , up to a degeneracy factor d_m which

counts the number of copies of elements of v_m contained in V_m . This fixes the overall normalization factor N_m :

$$N_m = n_m/d_m. \tag{4.12}$$

It remains to determine V_m for a given $(m + 1)$ -point prescription. It is easy to see that in each case we obtain a unique result, which is in a sense a direct product of p copies of v_m , taking into account the common scale κ_f . The points in the $(\kappa_f, \kappa_{r_1}, \kappa_{r_2})$ space that are being sampled in each prescription when there are two processes are shown in Fig. 3 (corresponding to the single-process prescriptions shown in Fig. 2).

To show how this works, we consider each prescription in turn, starting with the $\bar{5}$ -point prescription which is easier to construct than 5-point.

- $\bar{5}$ -point** For two processes, π_1 and π_2 say, we now have three scales, $\kappa_f, \kappa_{r_1}, \kappa_{r_2}$ which can each be varied independently. For the $\bar{5}$ -point prescription we only consider variations in which none of the scales is at the central value: $\bar{v}_4 = \{(\pm; \pm)\}$, where the \pm variations are performed independently. It follows that $\bar{V}_4 = \{(\pm; \pm, \pm)\}$, where the triples denote the three independent scales, $(\kappa_f; \kappa_{r_1}, \kappa_{r_2})$, varied independently. The set \bar{V}_4 thus has eight points in total. For each element of \bar{v}_4 , there are two elements of \bar{V}_4 , so $\bar{d}_4 = 2$, and since $\bar{n}_4 = 1/2, \bar{N}_4 = 1/4$. The result for the off-diagonal blocks of the theory covariance matrix in this prescription is thus given by

$$S_{i_1 j_2}^{(\bar{5}\text{pt})} = \frac{1}{4} \left\{ (\Delta_{i_1}^{++} + \Delta_{i_1}^{+-}) (\Delta_{j_2}^{++} + \Delta_{j_2}^{+-}) \right.$$

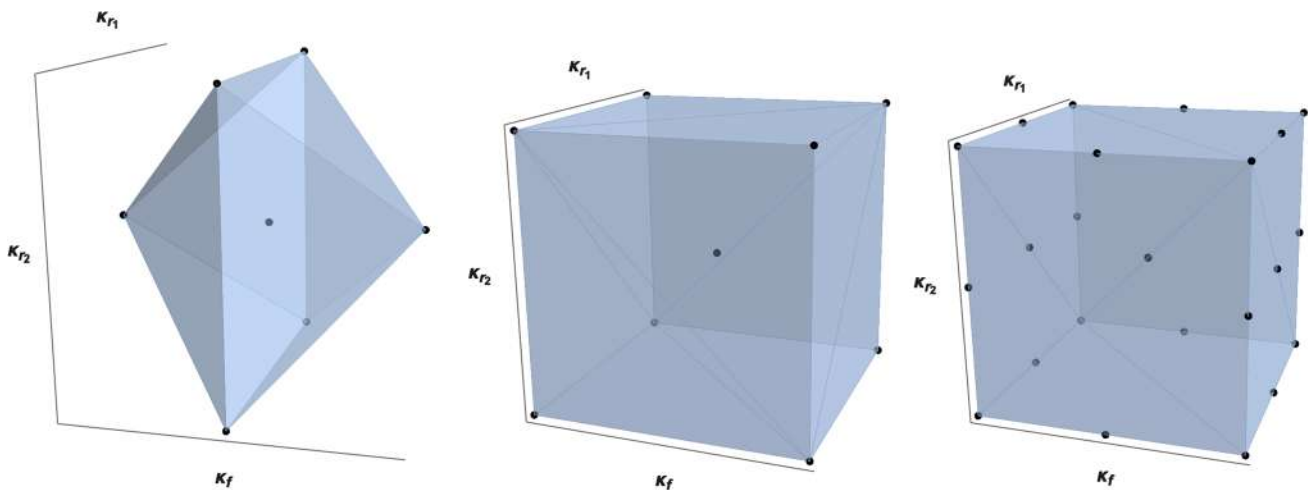


Fig. 3 Same as Fig. 2, now for the case of two different processes π_1 and π_2 with a common factorization scale κ_f and different renormalization scales κ_{r_1} and κ_{r_2} , so the diagrams are now in three dimen-

$$+ (\Delta_{i_1}^{--} + \Delta_{i_1}^{+-}) (\Delta_{j_2}^{--} + \Delta_{j_2}^{+-}) \}. \tag{4.13}$$

From this expression it is clear that while the scale κ_f is varied coherently between the two processes, the scales κ_{r_1} and κ_{r_2} are varied incoherently, as required.

It is straightforward to generalize this procedure to three processes: then $\bar{V}_4 = \{(\pm; \pm, \pm, \pm)\}$, so $\bar{d}_4 = 4$, and since $\bar{n}_4 = 1/2, \bar{N}_4 = 1/8$. However Eq. (4.13) remains unchanged, in the sense that it can be used to evaluate all three off-diagonal blocks $s_{i_1 j_2}, s_{i_2 j_3}, s_{i_3 j_1}$: this must always be the case, since each term in the sum Eq. (4.2) involves at most three scales. For p processes, it is easy to see that the number of distinct elements of V_4 is 2^{p+1} .

- 5-point** Again for two processes we have three scales, but now one varies each holding the other fixed to its central value: $v_4 = \{(\pm; 0), (0; \pm)\}$, so $V_4 = \{2(\pm; 0, 0), (0; \pm, \pm)\}$, where the two in front of the first element indicates that there are two copies of it, so V_4 has eight elements in total. Then for each element of v_4 , there are precisely two elements of V_4 , so $d_4 = 2$, and since $n_4 = 1/2, N_4 = 1/4$.

The result for the off-diagonal entries of the theory covariance matrix in this prescription is thus given by

$$S_{i_1 j_2}^{(5\text{pt})} = \frac{1}{4} \left\{ 2\Delta_{i_1}^{+0} \Delta_{j_2}^{+0} + 2\Delta_{i_1}^{-0} \Delta_{j_2}^{-0} + (\Delta_{i_1}^{0+} + \Delta_{i_1}^{0-}) (\Delta_{j_2}^{0+} + \Delta_{j_2}^{0-}) \right\}. \tag{4.14}$$

Note that also in this expression the variations of κ_f are manifestly correlated between the two processes, whereas the variations of $\kappa_{r_1}, \kappa_{r_2}$ are not.

When there are three processes, it is easy to see that $V_4 = \{4(\pm; 0, 0, 0), (0; \pm, \pm, \pm)\}$, i.e. it has 16 ele-

sions. The origin of coordinates is associated to the central scale, $\kappa_f = \kappa_{r_1} = \kappa_{r_2} = 0$. We again show the three prescriptions 5-point (left), $\bar{5}$ -point (center) and 9-point (right)

ments, though only 10 are distinct: the other six are simply copies, necessary to obtain the correct coefficients in Eq. (4.7) and Eq. (4.14). There are now four elements of V_4 for each element of v_4 , so now $d_4 = 4$, and $N_4 = 1/8$. Again Eq. (4.14) can be used to calculate all three off-diagonal blocks. For p processes, it is easy to see that V_4 has 2^{p+1} elements, but that only $2 + 2^p$ of these are actually distinct.

- 9-point** Here we vary the three scales completely independently: $v_8 = v_4 \oplus \bar{v}_4$. Constructing V_8 is now somewhat more involved, since while terms with $\kappa_f = 0$ have degeneracy 2, terms where $\kappa_f \neq 0$ is varied have degeneracy 3, so we need three copies of the former and two of the latter to take the overall degeneracy to 6. The solution is thus $V_8 = \{3(0; \pm, \pm), 2(\pm; \oplus, \oplus)\}$, where \oplus means either +, - or 0. Thus V_8 has 48 elements, of which only 22 are actually distinct. Since the first term of V_8 has a degeneracy of 2, while the last has a degeneracy of 3, the overall degeneracy is $d_8 = 6$, and since $n_8 = 1/4$, $N_8 = 1/24$. It follows that the off-diagonal blocks of the theory covariance matrix in this prescription are

$$S_{i_1 j_2}^{(9pt)} = \frac{1}{24} \left\{ 2 \left(\Delta_{i_1}^{+0} + \Delta_{i_1}^{++} + \Delta_{i_1}^{+-} \right) \times \left(\Delta_{j_2}^{+0} + \Delta_{j_2}^{++} + \Delta_{j_2}^{+-} \right) + 2 \left(\Delta_{i_1}^{-0} + \Delta_{i_1}^{-+} + \Delta_{i_1}^{--} \right) \left(\Delta_{j_2}^{-0} + \Delta_{j_2}^{-+} + \Delta_{j_2}^{--} \right) + 3 \left(\Delta_{i_1}^{0+} + \Delta_{i_1}^{0-} \right) \left(\Delta_{j_2}^{0+} + \Delta_{j_2}^{0-} \right) \right\}. \tag{4.15}$$

The pattern of correlations in the variation of the three scales in this expression should be clear from the way it is written.

When there are three processes, $V_8 = \{9(0; \pm, \pm, \pm), 4(\pm; \oplus, \oplus, \oplus)\}$, whence $d_8 = 36$, and since $n_8 = 1/4$, $N_8 = 1/144$. Again, Eq. (4.15) can be used to calculate all three off-diagonal blocks. V_8 now has 288 elements, of which 62 are distinct. For p processes, there are $2^p + 2 \cdot 3^p$ distinct elements.

4.3 Asymmetric prescriptions

It is sometimes argued that since only the cross-section is actually physical, a single process has only one scale, namely the scale of the process in the sense of Table 1 and Eq. (3.49). Therefore, in order to estimate the MHOUs, only this single scale should be varied. Alternatively, one may consider the variation of the scale of the process *on top* of the variation of the renormalization and factorization scales considered previously.

The logic of the first alternative (variation of the scale of the process only) is that after all there is only one scale in the factorised expressions, for example those given by the Wilson

expansion applied to DIS. The logic of the second alternative (variation of the scale of the process, the renormalization scale, and the factorization scale) is that each of these estimates a different source of MHOUs: varying the scale of the process generates terms related to missing higher order contributions to the hard cross-section which are proportional to collinear logarithms, the renormalization scales to missing higher order contributions to the hard cross-section which are proportional to the beta function, and finally the factorization scale to missing higher order contributions to the anomalous dimension.

On the other hand, both alternatives might be criticized on the grounds that they suppress correlations between the uncertainties in PDF evolution across different processes, and thus seriously overestimate uncertainties (the first worse than the second). Ultimately, however, they can be considered as possible options to be tested in a situation in which the true answer is known. Such a validation will be performed in the next section.

We now consider these two options in turn, both for the single-process case, which is represented schematically in Fig. 4, and for multiple-processes.

- 3-point** For a single process, we set $\kappa_f = \kappa_r$ and only vary the single resulting scale. Then $v_2 = \{\pm\}$ in an obvious notation, and $s = 1$, $m = 2$ and $n_2 = 1/2$, i.e. we simply average over the two nontrivial values of the single scale. For a single process we thus find that

$$S_{ij}^{(3pt)} = \frac{1}{2} \left\{ \Delta_i^{++} \Delta_j^{++} + \Delta_i^{--} \Delta_j^{--} \right\}, \tag{4.16}$$

whenever $i, j \in \pi$.

Likewise, for two different processes π_1 and π_2 , we set $\kappa_f = \kappa_{r_1}$ for π_1 , set $\kappa_f = \kappa_{r_2}$ for π_2 , and then vary κ_{r_1} and κ_{r_2} independently. This procedure necessarily ignores the correlations in the variation of κ_f between π_1 and π_2 . Since $v_2 = \{\pm\}$, $V_2 = v_2 \otimes v_2 = \{\pm, \pm\}$, where the ordered pairs denote the two independent

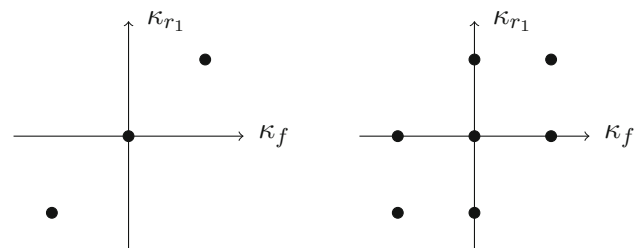


Fig. 4 Same as Fig. 2, now in the case of the asymmetric prescriptions for a single process with factorization scale κ_f and renormalization scale κ_r . We display the 3-point (left) and 7-point (right) prescriptions, defined in the text

scales $(\kappa_{r1}, \kappa_{r2})$. Clearly, for each element of v_2 there are two elements of V_2 , so $d_2 = 2$, Eq. (4.12) gives $N_2 = 1/4$, and the off-diagonal entries of the theory covariance matrix are

$$S_{i_1 j_2}^{(3pt)} = \frac{1}{4} \left\{ \left(\Delta_{i_1}^{++} + \Delta_{i_1}^{--} \right) \left(\Delta_{j_2}^{++} + \Delta_{j_2}^{--} \right) \right\}. \quad (4.17)$$

It can be seen from this factorised expression that the variations for each process are entirely uncorrelated. Generalization to more than two processes is straightforward: for p processes V_2 has 2^p elements, all of them distinct. Because in this prescription we ignore correlations in the PDF evolution uncertainties, we expect this prescription to significantly overestimate the MHOUs. Note that a fully correlated 3-point prescription in which we set $\kappa_f = \kappa_{r1} = \kappa_{r2}$ would instead significantly underestimate the MHOUs, which is why we do not consider it.

- 7-point** We now combine the variation of the scale of the process to the variation of renormalization and factorization scales. As we saw in Sect. 3.4, a change in the scale of the process is generated by $\tilde{\kappa}(\partial_{t_r} + \partial_{t_f})$, so it moves diagonally in the (κ_f, κ_r) plane. Thus for a single process, varying the scale of the process just corresponds to a new point-prescription, symmetric only about the line $\kappa_f = \kappa_r$, but asymmetric about the κ_f and κ_r axes. However, because variations of the scale of the process are assumed uncorrelated across different processes, while μ_f variations are correlated, such a scheme can give reduced correlations when there several processes.

For a single process, variation of the scale of the process just gives two extra points $(+; +)$, $(-; -)$ (in the same notation as before, i.e. variations in the $\kappa_f = \kappa_r$ plane), so $v_4 = \{(\pm; 0), (0; \pm)\}$ becomes $v_6 = \{(\pm; 0), (0; \pm), (+; +), (-; -)\} = \{(\pm; 0), (0; \pm), (\pm; \pm)\}$, where $(\pm; \pm)$ simply means that the variation is fully correlated (so there are only 2 terms, not 4).

We then have $v_6 = \{(\pm; 0), (0; \pm), (\pm; \pm)\}$, $s = 2$ (note we still have only two independent scales), $m = 6$ and $n_6 = 1/3$, and thus for a single process

$$S_{ij}^{(7pt)} = \frac{1}{3} \left\{ \Delta_i^{+0} \Delta_j^{+0} + \Delta_i^{-0} \Delta_j^{-0} + \Delta_i^{0+} \Delta_j^{0+} + \Delta_i^{0-} \Delta_j^{0-} + \Delta_i^{++} \Delta_j^{++} + \Delta_i^{--} \Delta_j^{--} \right\}. \quad (4.18)$$

When there is more than one process, we have to remember that variations of the scale of the process are uncorrelated between different processes, so they can decorrelate the allowed variations of μ_f . This means the allowed variations for two processes are in a space of four dimensions rather than three: call these say $(\kappa_{f1}, \kappa_{r1}; \kappa_{f2}, \kappa_{r2})$. The extension of v_6 is then $V_6 =$

$\{2(+, 0; +, 0), 2(-, 0; -, 0), (0, \pm; 0, \pm), (\pm, \pm; \pm, \pm)\}$, where $(\pm, \pm; \pm, \pm) = \{(+, +; +, +), (+, +; -, -), (-, -; +, +), (-, -; -, -)\}$, and thus $d_6 = 2$, so $N_6 = 1/6$, and the off-diagonal theory covariance matrix reads

$$S_{i_1 j_2}^{(7pt)} = \frac{1}{6} \left\{ 2\Delta_{i_1}^{+0} \Delta_{j_2}^{+0} + 2\Delta_{i_1}^{-0} \Delta_{j_2}^{-0} + \left(\Delta_{i_1}^{0+} + \Delta_{i_1}^{0-} \right) \left(\Delta_{j_2}^{0+} + \Delta_{j_2}^{0-} \right) + \left(\Delta_{i_1}^{++} + \Delta_{i_1}^{--} \right) \left(\Delta_{j_2}^{++} + \Delta_{j_2}^{--} \right) \right\}. \quad (4.19)$$

This prescription gives smaller correlations than the symmetric prescriptions since the variation of the two factorization scales μ_{f1} and μ_{f2} is now entirely uncorrelated. Generalization to p processes is again straightforward: since the variations of the scale of the process are in effect independent of the separate variations of μ_f and μ_r , $V_6 = V_4 \oplus V_2$ for any number of processes, so there are in total $2 + 2^{p+1}$ distinct elements.

5 Validation of the theory covariance matrix

In this section we determine the theory covariance matrix S_{ij} at NLO using the different prescriptions formulated in Sect. 4, we introduce a method for the validation of the theory covariance matrix when the next-order result is known, and we use it to validate the theory covariance matrices that we computed against the known NNLO results. This validation is performed on a global dataset based on the same processes as those used in the NNPDF3.1 PDF determination. This dataset will then be used to produce fits incorporating MHOUs using the theory covariance matrix (Sect. 6), and also, for comparison, fits using scale-varied theories (Appendix B).

5.1 Input data and process categorization

The validation of the theory covariance matrix and the PDF determination to be discussed in the next section are performed using a set of theory predictions for a dataset which is very similar to that used in the NNPDF3.1 PDF determination [5], but differs from it in some details, as we now discuss.

The input dataset used here includes fixed-target [28–35] and HERA [36] deep-inelastic inclusive structure functions; charm cross-sections from HERA [37]; gauge boson production from the Tevatron [38–41]; and electroweak boson production, inclusive jet, $Z p_T$ distributions, and $t\bar{t}$ total and differential cross-sections from ATLAS [42–50], CMS [51–59] and LHCb [60–63] at $\sqrt{s} = 7$ and 8 TeV (two data points for the ATLAS and CMS total $t\bar{t}$ cross-sections are at 13 TeV).

This input dataset differs in many small respects from that used in the NNPDF3.1 baseline. Firstly, the fixed-target Drell–Yan (DY) cross-sections [64–67] are excluded from the fit since APFEL currently does not allow the calculation of scale-varied fixed-target DY cross-sections. Secondly, the value of the lower kinematic cut has been increased from $Q_{\min}^2 = 2.69 \text{ GeV}^2$ to 13.96 GeV^2 in order to ensure the validity of the perturbative QCD expansion when scales are varied downwards. Thirdly, we include only jet data for which the exact NNLO calculations are available, as discussed in [68], namely the ATLAS and CMS inclusive jet cross-sections at 7 TeV from the 2011 dataset. Finally, we exclude the bottom structure function F_2^b measurements, for which the implementation of scale variations is complicated by the crossing of the heavy quark thresholds.

Also, in original NNPDF3.1 determination somewhat different cuts were applied to data at NLO and NNLO (essentially in order to remove from the NLO fit data which are subject to large NNLO corrections). Here we wish to have exactly the same dataset at NLO and NNLO, in order to make sure that the differences between NLO and NNLO are due purely to differences in the theoretical calculations, and not in the input datasets. Therefore, the baseline kinematic cuts of NNPDF3.1 have been slightly modified so that the data points excluded at NLO are also excluded at NNLO and vice-versa.

Taking into account all these modifications, in total the input dataset includes $N_{\text{dat}} = 2819$ datapoints. The fact that the dataset differs somewhat from that of Ref. [5] must be kept in mind when assessing the impact of theory uncertainties, and indeed to this purpose in Sect. 6.2 we will construct a new baseline PDF set which only differs from that of Ref. [5] in that it is based on the dataset we present here. Specifically, the loss of Drell–Yan data will lead to an increased uncertainty in the $\bar{u} - \bar{d}$ combination, and the higher Q^2 cutoff to somewhat larger uncertainties in the small- x region where the low Q^2 data are concentrated. Here our main goal is to assess the impact of theory uncertainties, not to construct the most competitive, state-of-the-art PDF set, which will be the subject of future work.

Because the prescriptions in Sect. 4 assume that renormalization scale variation is fully correlated within a given process, but uncorrelated between different processes, it is necessary to define what it is meant by “process”, i.e., to classify datasets into processes. This requires an educated guess as to which theory computations share the same higher order corrections. For example, it is necessary to decide whether charged-current (CC) and neutral-current (NC) DIS are the same process or not, and whether the transverse momentum and rapidity distributions for one observable (such as, say, Z production) should be grouped together. Our categorization is summarized in Table 2.

Specifically, we group the data into five distinct categories: DIS NC, DIS CC, Drell–Yan (DY), inclusive jet production (JET), and top quark pair production (TOP). More refined categorizations will be considered elsewhere, but we consider this to be sufficient for a first study. The logic underlying this choice is that we group together processes that are likely to share the same MHO terms. Thus for instance the predictions for all DY processes are obtained by integrating the same underlying fully differential distributions, and thus have a similar perturbative structure. Because different distributions impact different PDF combinations – so e.g. the Z p_T distribution mostly impacts the gluon, while the W rapidity distributions mostly impact flavor separation – this will induce nontrivial correlations in the PDF fitting.

All calculations are performed using the same settings as in [5]: PDF evolution and the calculation of DIS structure functions up to NNLO are carried out using the APFEL [69] program; heavy quark mass effects are included by means of the FONLL general-mass variable flavor number scheme [70–72]; the charm PDF is fitted alongside the light quark PDFs [73], rather than being generated from perturbative evolution of light partons; the charm quark pole mass is taken to be $m_c = 1.51 \text{ GeV}$, and the strong coupling constant is fixed to be $\alpha_s(m_Z) = 0.118$, consistent with the latest PDG average [74].

In order to evaluate the theory covariance matrix S_{ij} , it is necessary to be able to evaluate both DIS structure functions and hadronic cross-sections for a range of values of the factorization and renormalization scales, i.e., in the notation of Eq. (3.41), for $\kappa_f \neq 0$ and $\kappa_r \neq 0$. In this case, the entries of the NLO theory covariance matrix have been constructed by means of the ReportEngine software [75] taking the scale-varied NLO theory cross-sections $T_i(k_f, k_r)$ as input. These are provided by APFEL [69] for the DIS structure functions and by APFELgrid [76] combined with APPLgrid [77] for the hadronic cross-sections. The evaluation of these scale-varied cross-sections has been validated by means of independent programs, in particular with HOPPET [78] and OpenQCDrad [79] for the DIS structure functions, and with the built-in scale variation functionalities of APPLgrid. All these NLO cross-sections are evaluated using the central NLO PDF obtained by performing a NLO fit to the same dataset, for consistency.

5.2 The theory covariance matrices at NLO

We now present results for the theory covariance matrices, constructed using NLO calculations and evaluated according to the prescriptions introduced in Sect. 4, and discuss some of their qualitative features.

In Fig. 5 we show the diagonal elements of the experimental and theory covariance matrices, or more specifically the experimental uncertainty normalized to the data,

Table 2 The categorization of the input datasets into different processes adopted in this work. Each dataset is assigned to one of five categories: neutral-current DIS (DIS NC), charged-current DIS (DIS CC), Drell–Yan (DY), jet production (JET) and top quark pair production (TOP). For each dataset, we also provide the corresponding publication reference and the number of data points after cuts. We also show the total number of points in each of the five categories of process

Process type	Dataset	References	N_{dat}	N_{dat} (total)
DIS NC	NMC	[28,29]	134	1593
	SLAC	[30]	12	
	BCDMS	[31,32]	530	
	HERA σ_{NC}^p	[36]	886	
DIS CC	HERA σ_{NC}^c	[37]	31	552
	NuTeV dimuon	[33,34]	41	
	CHORUS	[35]	430	
DY	HERA σ_{CC}^p	[36]	81	484
	ATLAS $W, Z, 7 \text{ TeV } 2010$	[42]	30	
	ATLAS $W, Z, 7 \text{ TeV } 2011$	[43]	34	
	ATLAS low-mass DY 2011	[44]	4	
	ATLAS high-mass DY 2011	[45]	5	
	ATLAS $Z p_T 8 \text{ TeV } (p_T^H, M_{ll})$	[46]	44	
	ATLAS $Z p_T 8 \text{ TeV } (p_T^H, y_Z)$	[46]	48	
	CMS Drell–Yan 2D 2011	[51]	88	
	CMS $W \text{ asy } 840 \text{ pb}$	[52]	11	
	CMS $W \text{ asy } 4.7 \text{ pb}$	[53]	11	
	CMS $W \text{ rap } 8 \text{ TeV}$	[54]	22	
	CMS $Z p_T 8 \text{ TeV } (p_T^H, M_{ll})$	[55]	28	
	LHCb $Z 940 \text{ pb}$	[60]	9	
	LHCb $Z \rightarrow ee 2 \text{ fb}$	[61]	17	
	LHCb $W, Z \rightarrow \mu 7 \text{ TeV}$	[62]	29	
	LHCb $W, Z \rightarrow \mu 8 \text{ TeV}$	[63]	30	
	CDF $Z \text{ rap}$	[38]	29	
	D0 $Z \text{ rap}$	[39]	28	
	D0 $W \rightarrow ev \text{ asy}$	[40]	8	
	D0 $W \rightarrow \mu\nu \text{ asy}$	[41]	9	
JET	ATLAS jets 2011 7 TeV	[47]	31	164
	CMS jets 7 TeV 2011	[56]	133	
TOP	ATLAS σ_{tt}^{top}	[48,49]	3	26
	ATLAS $t\bar{t} \text{ rap}$	[50]	10	
	CMS σ_{tt}^{top}	[57,58]	3	
	CMS $t\bar{t} \text{ rap}$	[59]	10	
Total			2819	2819

$(C_{ii})^{1/2}/D_i$, and the MHO normalized to the data, $(S_{ii})^{1/2}/D_i$, for $i = 1, \dots, N_{\text{dat}}$, where D_i is the i -th datapoint. Here and henceforth, the experimental covariance matrix C_{ij} includes all uncorrelated statistical uncertainties as well as correlated systematic uncertainties, as published by the respective experiments, and used to assess fit quality as e.g. in Sect. 3.2 of Ref. [68]. Note that this differs from the covariance matrix C_{ij} used for PDF minimization in the treatment of multiplicative uncertainties (such as normalization or luminosity uncertainties) in that the latter must be treated using the so-called t_0 method of Ref. [80] in order to avoid bias. As in all previous NNPDF determinations, the t_0 covariance matrix is used for PDF minimization while the experi-

mental covariance matrix is used in order to assess fit quality, in order to ensure reproducibility of results.

The datapoints are grouped by process and, within a process by experiment, following Table 2. The theory covariance matrix S_{ij} is computed using the 9-point prescription (the one with the largest number of independent variations; recall Sect. 4). Broadly speaking, the estimated NLO MHO is roughly comparable to experimental uncertainties, as expected. However for some datapoints the experimental uncertainty is dominant (and thus the theory uncertainty will have only a small effect), while for others the MHO is dominant. These latter datapoints will carry less weight in a PDF fit with MHO included, depending also on the underlying

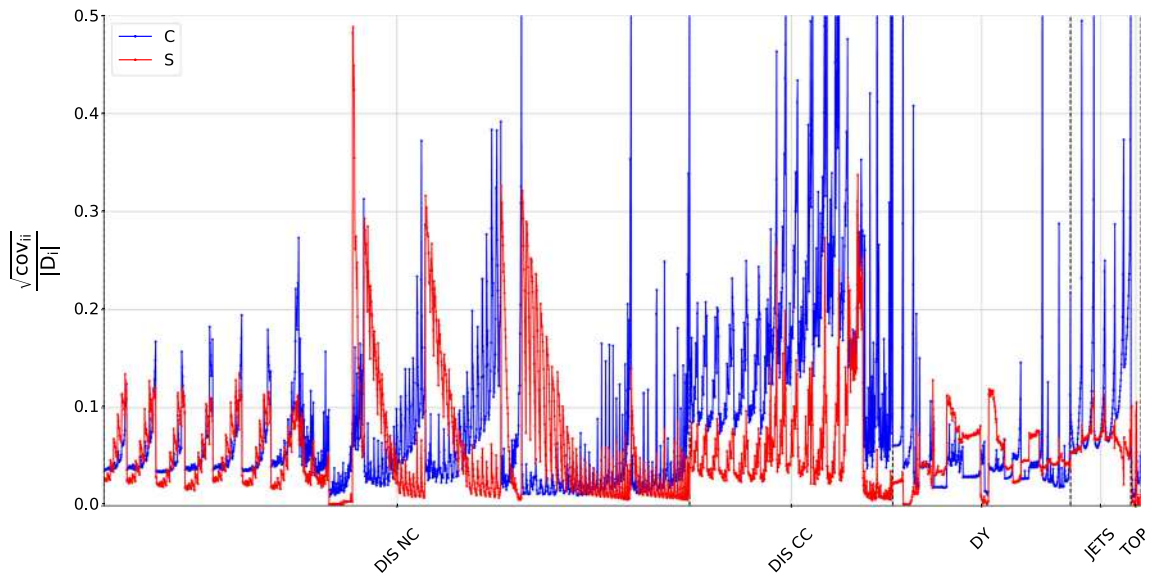


Fig. 5 Comparison of the diagonal experimental uncertainties (blue) and the diagonal theoretical uncertainties evaluated using the 9-point prescription (red), all normalized to the central experimental value. The

correlation pattern. Some datasets have datapoints in both these categories: the HERA NC DIS are particularly striking, since at high Q^2 (where statistics are low) the dominant uncertainty is experimental, while at low Q^2 (and thus small x , where perturbation theory is less reliable) the dominant uncertainty is due to MHO.

In Fig. 6 we compare the complete experimental covariance matrix C_{ij} to the theory covariance matrix S_{ij} , again computed using the 9-point prescription. Both covariance matrices are displayed as heat maps, with each entry expressed as a fraction with respect to the corresponding experimental central value; i.e. $C_{ij}/D_i D_j$ and $S_{ij}/D_i D_j$. It is clear from Fig. 6 that the theory covariance matrix has, as expected, a richer structure of correlations than its experimental counterpart: for example data from the same process (such as DIS) are correlated even when the corresponding experimental measurements are completely uncorrelated (such as HERA and fixed target). Furthermore, correlation of the factorization scale variation between disparate processes, such as DIS processes and hadronic processes, results in nonzero entries in the theory covariance matrix even in these regions.

The precise structure of these theory-induced correlations depends on the choice of prescription adopted. To illustrate this, Fig. 7 compares the experimental correlation matrix, given by

$$\rho_{ij}^{(C)} = \frac{C_{ij}}{\sqrt{C_{ii}}\sqrt{C_{jj}}}, \tag{5.1}$$

with the corresponding combined experimental and theoretical correlation matrices, defined by

data are grouped by process and, within a process, by experiment, following Table 2

$$\rho_{ij}^{(C+S)} = \frac{(C+S)_{ij}}{\sqrt{(C+S)_{ii}}\sqrt{(C+S)_{jj}}}, \tag{5.2}$$

for all the prescriptions defined in Sect. 4. Specifically, from top left to bottom right we have the experimental correlations $\rho^{(C)}$ followed by $\rho^{(C+S)}$ for the symmetric 5, $\bar{5}$, 9 point prescriptions, and the asymmetric 3 and 7 point prescriptions. As in Fig. 6, the cross-sections are grouped by process type and, within that, by experiment.

Some qualitative features of the theory-induced correlations are apparent. There are clearly large positive correlations within individual experiments along the diagonal blocks, this being particularly evident for DIS NC and DY data, which have large numbers of points which are relatively close kinematically. Off the diagonal, but still within the same process, there are large correlations between experiments for the DY, jets and top data points, and large anticorrelations for the DIS NC data points (these mostly between fixed target and HERA). Correlations and anticorrelations between different processes are also often present but are somewhat weaker: for example the DY data points (from LHC) are quite correlated with the HERA NC DIS data points, but anticorrelated with fixed target NC DIS data points.

When comparing different prescriptions, it is clear that the 3-point prescription leads to especially small correlations between processes, which is expected because with this prescription the factorization scale and renormalization scale variations are uncorrelated between processes. The correlations between processes are also weaker in 7-point than in 5-point, due to the fact that (as discussed in Sect. 4.3) the correlated variation of the factorization scale is combined with the uncorrelated variations of the scale of the process for the

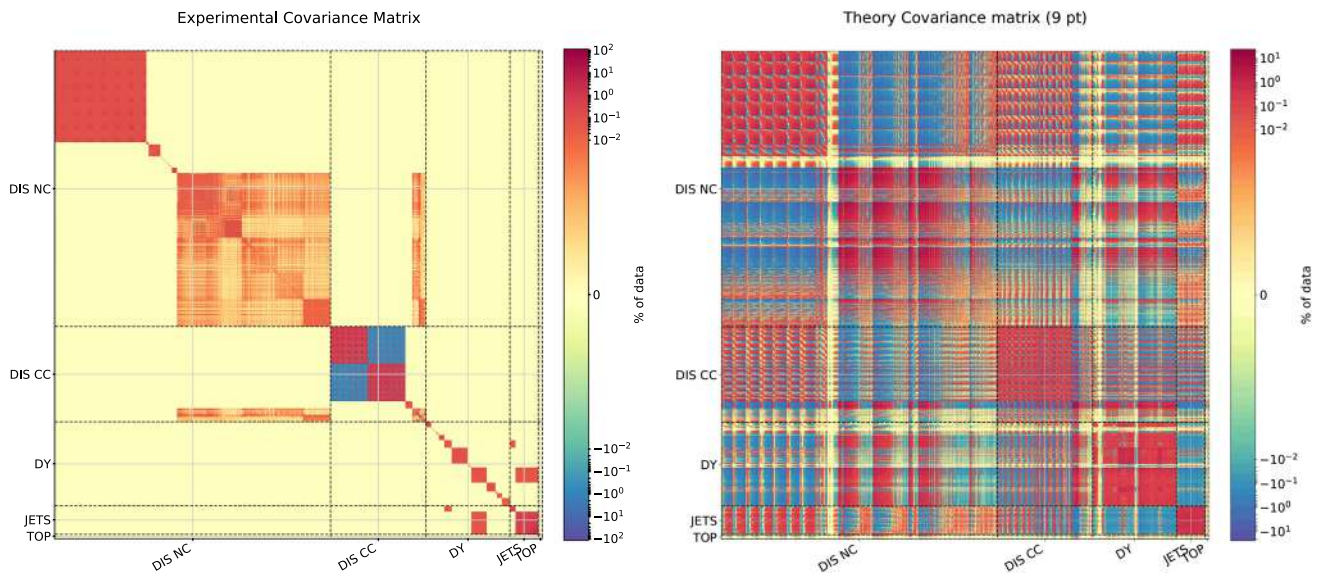


Fig. 6 Comparison of the experimental C_{ij} (left) and the theoretical S_{ij} (right) covariance matrices, the latter evaluated using the 9-point prescription. All entries are normalized to the central experimental value.

The data are grouped by process and, within a process, by experiment, following Table 2

pair of processes involved. It is worth noting, however, that the pattern of correlations is similar for all the symmetric prescriptions.

In order to decide which prescriptions are best, and more generally whether or not they produce a reliable estimate of MHOUs, we must proceed to their validation.

5.3 Construction of validation tests

We wish to construct a validation test for the NLO theory covariance matrix, by comparing it to the known NNLO theoretical result. We do so by viewing the set of experimental data as a vector with components D_i , where $i = 1, \dots, N_{\text{dat}}$. The vector lives in an N_{dat} -dimensional “data” space D , on which the theory covariance matrix S_{ij} acts as a linear operator. The matrix S_{ij} is symmetric and positive semi-definite, meaning that all of its non-zero eigenvalues are positive. In a PDF fit, S_{ij} always enters as an additive contribution to the experimental covariance matrix C_{ij} , and thus their sum is always invertible, owing to the non-zero statistical uncertainties on the data, which bound the eigenvalues from below.

The matrix S_{ij} defines ellipsoids E corresponding to a given confidence level in the data space, centered on the NLO theoretical prediction, $T_i^{\text{NLO}} \equiv T_i^{\text{NLO}}(0, 0)$ evaluated using the central scale choice. In the context of MHOUs, we can take T_i^{NLO} to be the predictions at NLO, with the one-sigma ellipsoid $E_{1\sigma}$ estimating a 68% confidence level for the MHO correction. We can validate whether S_{ij} correctly predicts both the size and the correlation pattern of the MHO terms by testing the extent to which the shift vector

$\delta_i \sim T_i^{\text{NNLO}} - T_i^{\text{NLO}}$, i.e. the difference between the NNLO and NLO predictions for T_i , falls within a given ellipsoid E . These predictions should be taken with a fixed underlying PDF (which could indeed be any standard reference PDF): it is the change in prediction due to the change of perturbative evolution and hard matrix element which are relevant here. Note that the dimensionality of the subspace spanned by the ellipsoid E is much smaller than that of the data space D : in a global fit the data space has dimension $\mathcal{O}(3000)$ (Table 2), while even the most complex prescriptions in Sect. 4 have $\mathcal{O}(30)$ independent variations, not all of which correspond to independent eigenvectors, as we will see shortly. So E actually lives in a subspace S of dimension N_{sub} of the full space D : $E \in S \in D$. For a single process we expect N_{sub} to be of order a dozen or so at most. In fact, even for a single process (see Table 2) we always have $N_{\text{sub}} \ll N_{\text{dat}}$. Hence, a nontrivial validation of the theory covariance matrix is if the component of the shift vector δ_i lying outside E is small, i.e. if the angle between δ_i and the projection of δ_i onto S is small.

Furthermore we expect the component of δ_i along each axis of the ellipsoid E to be of the same order as the typical one-sigma variation. The physical interpretation of such a successful validation is that the eigenvectors of S_{ij} correctly estimate the independent directions of uncertainty in theory space, with the size of the shift estimated by the corresponding eigenvalue. The null subspace of E , i.e. the directions of vanishing eigenvalues, would then correspond to directions in D for which the theory uncertainty is so small that it cannot be reliably estimated and so can be safely neglected.

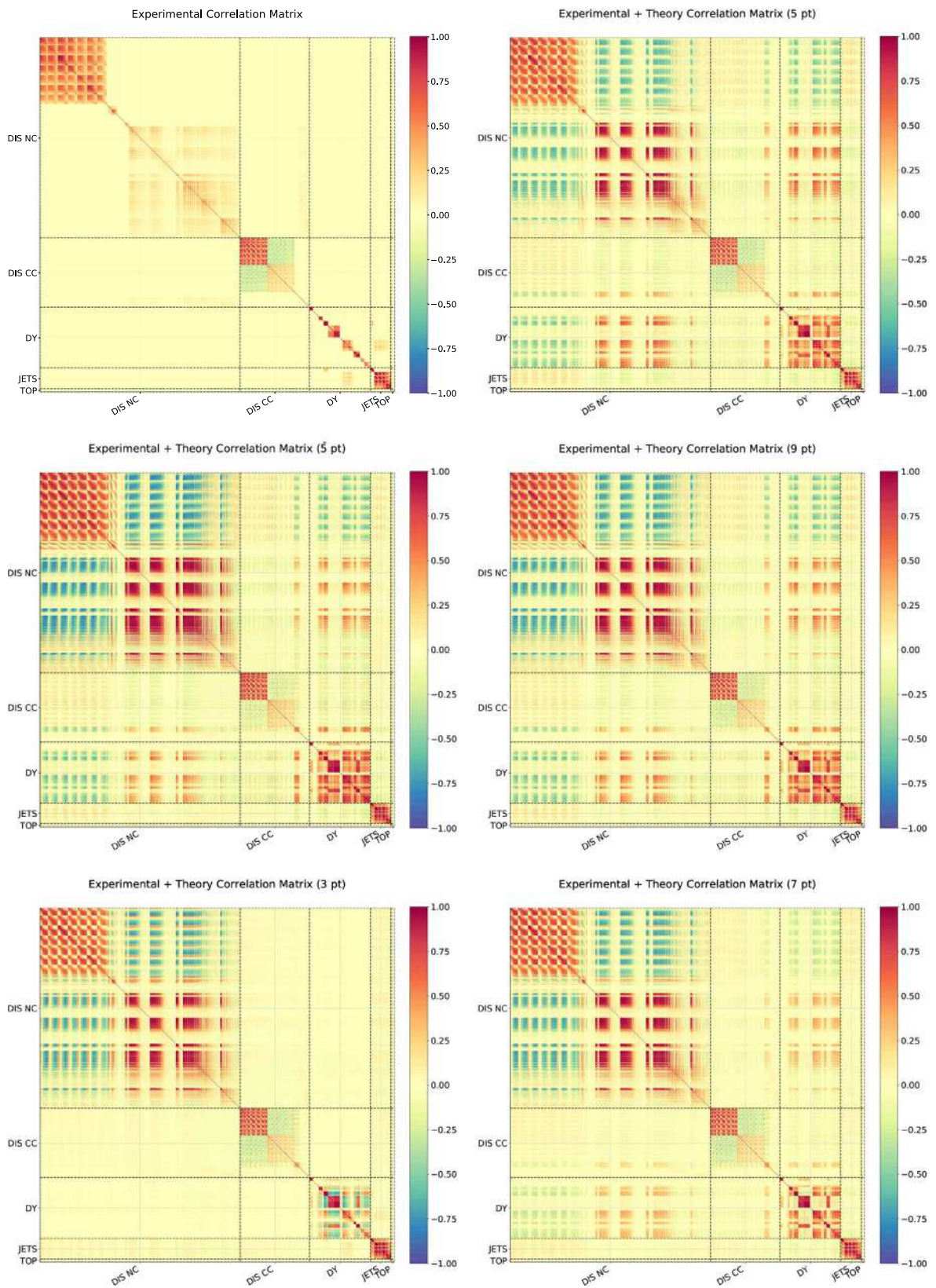


Fig. 7 Comparison of the experimental correlation matrix Eq. (5.1) (top left) and the the combined experimental and theoretical correlation matrices Eq. (5.2) computed using the prescriptions described in Sect. 4: the symmetric prescriptions (5-pt top right, 5-pt center left, 9-pt

center right), and asymmetric prescriptions (3-pt bottom left, 7-pt bottom right). The data are grouped by process and within a process by experiment, as in Fig. 6

These are highly nontrivial tests, given the huge discrepancy between the dimensionality of the space D , and the dimensionality of S .

Let us now see how this works in detail. First, we need to identify the spaces E and S . To do this, we normalize the NLO theory covariance matrix S_{ij} to the central NLO theory prediction T_i , so that all its elements are dimensionless, allowing a meaningful comparison: we define

$$\widehat{S}_{ij} = S_{ij} / (T_i^{\text{NLO}} T_j^{\text{NLO}}). \tag{5.3}$$

Likewise, we define a normalized shift vector with components

$$\delta_i = (T_i^{\text{NNLO}} - T_i^{\text{NLO}}) / T_i^{\text{NLO}}. \tag{5.4}$$

The NNLO prediction T_i^{NNLO} is computed using NNLO matrix elements and parton evolution, but with the same NLO PDF set used in the computation of T_i^{NLO} and S_{ij} . In this way the shift δ_i only takes account of the perturbative effects due to NNLO corrections, which are estimated by S_{ij} , and not the additional effect of refitting.

We then diagonalize \widehat{S}_{ij} , to give eigenvectors, e_i^α (chosen to be orthonormal, i.e. $\sum_i e_i^\alpha e_i^\beta = \delta^{\alpha\beta}$), with corresponding non-zero eigenvalues, $\lambda^\alpha = (s^\alpha)^2$; $\alpha = 1, \dots, N_{\text{sub}}$. All these eigenvalues are real and positive, see Eq. (4.3). The eigenvectors span the subspace S . There are also $N_{\text{dat}} - N_{\text{sub}}$ zero eigenvalues. These are degenerate, and their eigenvectors span the space D/S . Because of the zero eigenvalues, the diagonalization of \widehat{S}_{ij} is in practice rather difficult: the procedure we use to identify the subspace S and its dimensionality N_{sub} , and then diagonalize the projection of \widehat{S}_{ij} into S , is described in some detail in Appendix A.

Next we project the shift vector δ_i onto the eigenvectors,

$$\delta^\alpha = \sum_{i=1}^{N_{\text{dat}}} \delta_i e_i^\alpha. \tag{5.5}$$

These projections δ^α should be of the same order as the size of the ellipse in this direction, i.e. the s^α : more specifically in an ideal world 68% of the δ^α / s^α would be less than one. This is all the meaningful statistical information that is contained in \widehat{S}_{ij} .

Finally, we can now resolve the shift vector δ_i into its component lying within S

$$\delta_i^S = \sum_{\alpha=1, \dots, N_{\text{sub}}} \delta^\alpha e_i^\alpha, \tag{5.6}$$

and the complementary component within the remaining space D/S , $\delta_i^{\text{miss}} = \delta_i - \delta_i^S$. For a successful test, we expect most of δ to lie within S , so $|\delta_i^{\text{miss}}| \ll |\delta_i|$, or equivalently

$|\delta_i^S| \approx |\delta_i|$. By construction δ_i^S and δ_i^{miss} are orthogonal (since the subspaces S and D/S are orthogonal spaces), thus the three vectors δ_i^S , δ_i^{miss} and δ_i form a right-angled triangle, with δ_i being its hypotenuse. The geometrical relation between the shift vector δ_i , and the component of the shift vector which lies in the subspace S , δ_i^S is illustrated in Fig. 8.

With these definitions, the theory covariance matrix S_{ij} provide a reasonable estimate of the MHOU if the angle

$$\theta = \arccos \left(\frac{|\delta_i^S|}{|\delta_i|} \right) = \arcsin \left(\frac{|\delta_i^{\text{miss}}|}{|\delta_i|} \right) \tag{5.7}$$

between the shift δ_i and its component in the subspace S , δ_i^S is reasonably small. As mentioned above, for a global PDF fit the typical situation that one encounters is that $N_{\text{dat}} \gg N_{\text{sub}}$ (in the present case $N_{\text{dat}} \sim \mathcal{O}(3000)$, while $N_{\text{sub}} \sim \mathcal{O}(30)$). So this validation test is highly nontrivial, since finding the relatively small subspace S in the huge space D is rather hard: for a random symmetric matrix S_{ij} , components of δ_i in D/S will generally be as large as those in S , and thus $|\delta_i^S| \ll |\delta_i|$, and θ will be very close to a right angle.

5.4 Results of validation tests

We now explicitly perform the validation tests discussed in Sect. 5.3, with the NLO theory covariance matrices \widehat{S}_{ij} (normalized to NLO theory, as in Eq. (5.3)) constructed from scale variations for all data points in Table 2, and for each prescription of Sect. 4. These are then validated using the shift vector δ_i constructed as the difference of NNLO and NLO theory, normalized to the latter, as in Eq. (5.4).

A very first comparison can be done at the level of diagonal elements σ_i , where $\widehat{S}_{ii} = (\sigma_i)^2$, by comparing them directly to the normalized shifts δ_i Eq. (5.4). This already

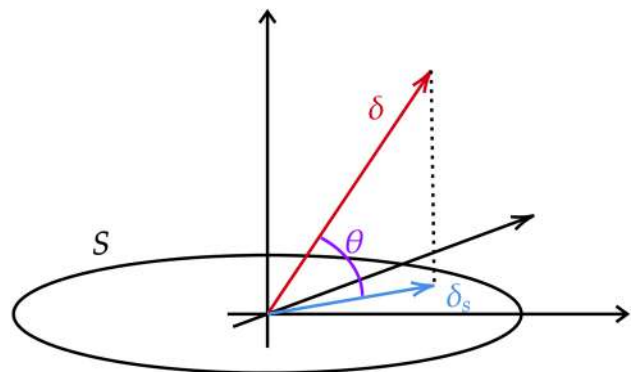


Fig. 8 Schematic representation of the geometric relation between the shift vector $\delta \in D$ (here drawn as a three dimensional space), and the component δ^S of the shift vector which lies in the subspace S (here drawn as a two dimensional space, containing the ellipse E defined by the theory covariance matrix). The angle θ between δ and δ^S is also shown: the dotted line shows the other side of the triangle, $\delta^{\text{miss}} \in D/S$

tells us whether the overall size of the scale variation is of the right order of magnitude: one expects the shifts δ_i and the uncertainties σ_i to be of roughly the same order.

These comparisons are shown in Figs. 9 and 10. In each plot the data points are presented sequentially on the horizontal axis, organized by process as in Table 2. The shape of the estimated MHOUs imitates the shape of the true shift rather faithfully, for each of the five processes, and for each prescription. This shows that the theory covariance matrix gives a qualitatively reliable estimate of the true MHOUs, in the sense that the estimate is small when the MHOUs is small,

large when it is large, and moreover correctly incorporates the correlations in the MHOUs between nearby kinematic regions, responsible for the shape. There is little discernible difference between all the various point prescriptions, except in the overall size of the estimates: for example comparing the symmetric prescriptions, we see that 5-point is the least conservative and $\bar{5}$ -point is the most conservative, whilst 9-point lies somewhere between the two. This is particularly noticeable in the DY data.

It is clear from these plots that the overall size of the estimated uncertainties, given by varying renormalization and

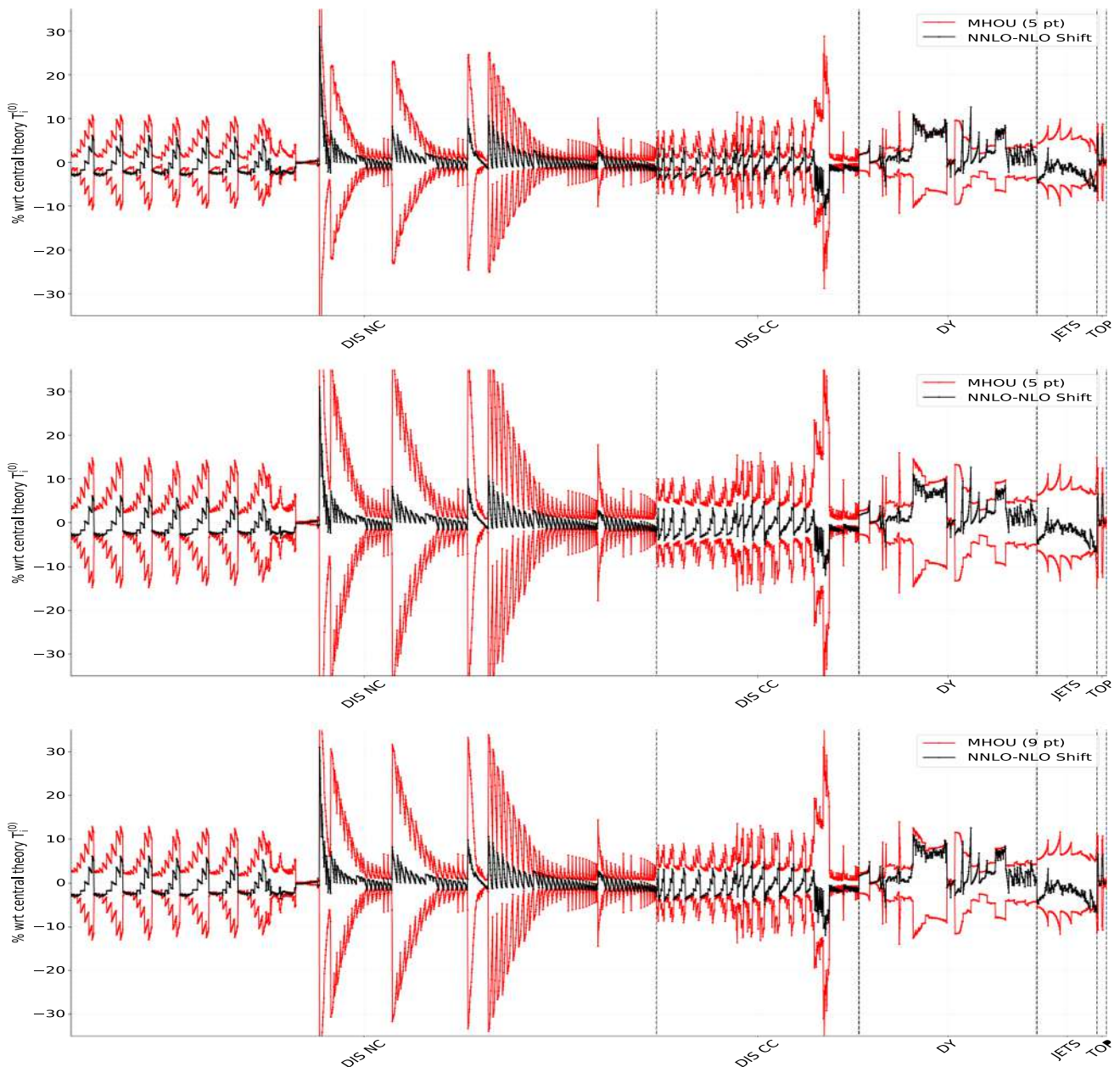


Fig. 9 The diagonal uncertainties σ_i (red) symmetrized about zero, compared to the shift δ_i for each datapoint (black), for the symmetric prescriptions: 5-point (top), $\bar{5}$ -point (middle), and 9-point (bottom). All values are shown as percentage of the central theory prediction

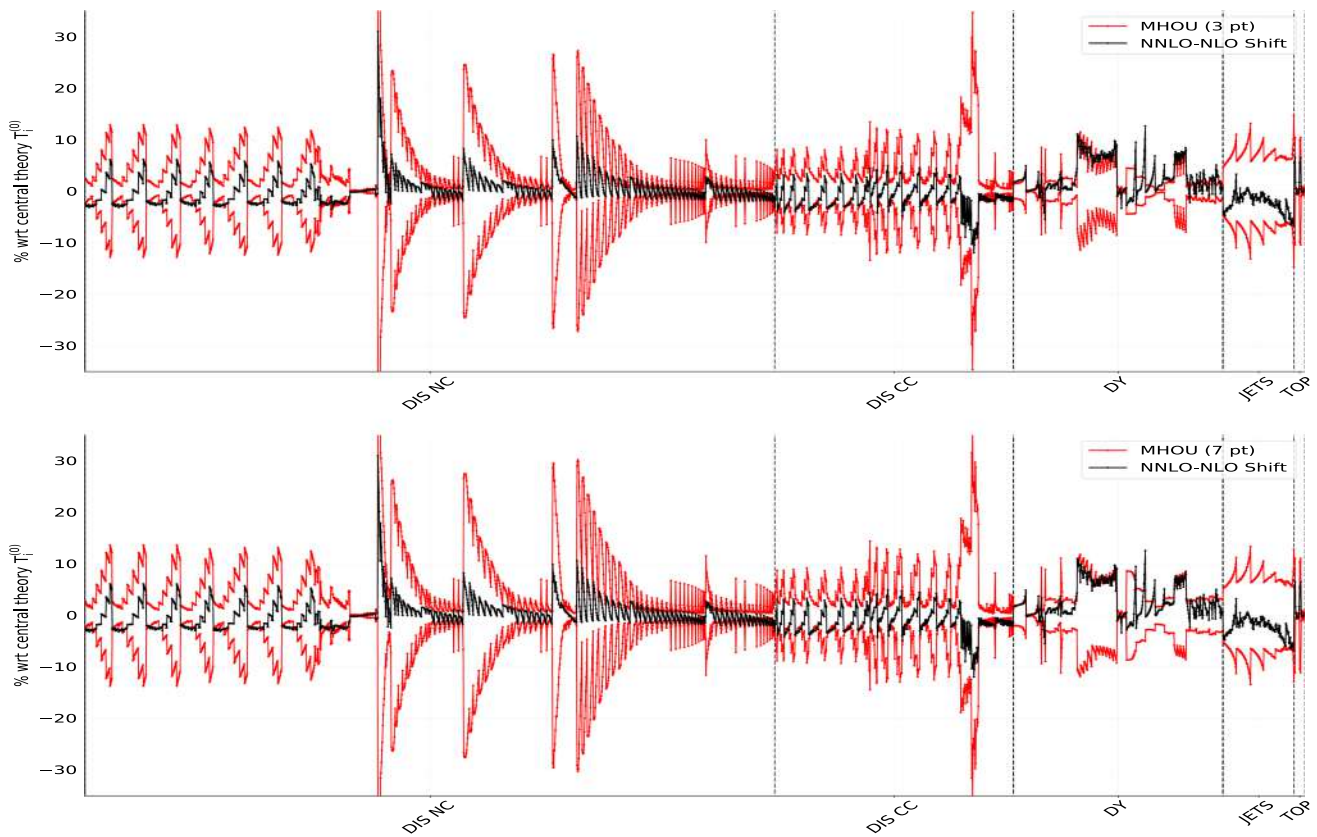


Fig. 10 Same as Fig. 9 but for the asymmetric prescriptions: 3-point (top) and 7-point (bottom)

factorization scales by a factor of two in either direction (i.e. as in Eq. (4) with $w = \ln 4$) is, by and large, roughly correct: if the range were significantly smaller, some of the uncertainties would have been underestimated, whereas if it were larger all uncertainties would have been overestimated. This said, for several data points the MHOUs at NLO is clearly overestimated by scale variation: this is particularly true of the small- x NC DIS data from HERA in the center of the plot.

Overall, these plots demonstrate that since there are only small differences in the diagonal elements of each prescription, it is in the detailed correlations between data points where the differences in performance between the prescriptions lies. To expose this, we need to diagonalize the theory covariance matrix (using the procedure in Appendix A), so that we can see in detail which components of the shift vector are correctly estimated, and which are missed, as explained in Sect. 5.3.

As discussed in Sect. 5.3, once we have the eigenvectors corresponding to the nonzero eigenvalues of the theory covariance matrix, the first validation test consists of checking how much of the shift vector δ_i lies within the space spanned by these eigenvectors, S , and has thus been included in the estimation of MHOUs provided by the theory covariance matrix. The results of this test for the global dataset,

described in Sect. 5.1, are shown in Table 3: for each prescription we give the dimension N_{sub} of S , i.e. the number of linearly independent eigenvectors e_i^α of S_{ij} , and then the

Table 3 The angle θ Eq. (5.7) between this shift and its component δ_i^S lying within the subspace S (see Fig. 8) spanned by the theory covariance matrix for different prescriptions. The dimension of the subspace S in each case is also given

Prescription	N_{sub}	θ
5-pt	8	33°
$\bar{5}$ -pt	12	31°
9-pt	28	26°
3-pt	6	52°
7-pt	14	29°

Table 4 Same as Table 3 for each process of Table 2. The number of data points in each process is given directly below the name of the process

Presc.	N_{sub}	DIS NC	DIS CC	DY	JET	TOP
		1593	552	484	164	26
5-pt	4	39°	21°	25°	17°	11°
$\bar{5}$ -pt	4	38°	17°	23°	22°	10°
9-pt	8	32°	16°	22°	14°	3°
3-pt	2	54°	36°	39°	24°	12°
7-pt	6	35°	17°	22°	16°	3°

value of the angle θ , defined in Eq. (5.7), between the shift δ_i and its component δ_i^S , defined in Eq. (5.6), lying within the subspace S spanned by e_i^α . We note that all the angles are reasonably small, despite the fact that N_{sub} is so much smaller than the dimension 2819 of the data space.

The 9-point prescription performs best, with an angle of $\theta = 26^\circ$ between the shift δ_i and its projection δ_i^S in the subspace S : clearly the more complicated pattern of scale variations (compared to the other two symmetric prescriptions) improves the estimation of the MHO. The 3-point prescription performs worst, suggesting that lack of correlation in the factorization scale between processes in this prescription means that much of the correlation in the MHO due to universal PDF evolution has been missed. The 7-point prescription is however only a little worse than 9-point, presumably due to the dilution of the correlation in factorization scale variation which is a feature of this prescription. Note that since these results for θ are geometrical, they are largely independent of the range of the scale variation Eq. (4).

It is interesting to ask whether all processes are equally well described, and whether there are significant differences in correlations between processes or within a process. To this purpose, in Table 4 we list the angle θ computed for each individual process using the various prescriptions. Three conclusions emerge from inspection of this table. First, when each process is taken individually, the results seen in Table 3 for the relative merits of each prescription are replicated process by process: again 3-point is worst, and 9-point is best. Secondly, processes with large numbers of data points are

much harder to describe than those with only a few data points (i.e. θ is smallest for smaller datasets): this is hardly surprising, since the larger datasets cover a wider kinematic range and thus have more structure to predict. Finally, the quality of the description of the global dataset for each prescription is in each case dominated by the process (DIS NC) which is described worst, however the global dataset is actually described a little better (for each prescription) than the dataset for this process, particularly for 9-point, less so for 3-point. This suggests that correlations across processes are actually described reasonably well, and are anyway less critical than correlations within processes.

We next look in more detail at the part of δ_i which falls outside the subspace S , $\delta_i^{\text{miss}} = \delta_i - \delta_i^S$. This is shown for the 9-point prescription in Fig. 11. While this is generally uniformly small, of order a few percent, across the full range of processes, it also has nonzero components in all datasets, and all processes. Furthermore, for most processes the shape of δ_i^{miss} closely follows that of the shift δ_i . This may suggest that a significant fraction of δ_i^{miss} might be due to the fact that there is a component of δ_i which is systematically missing for most or all processes. This in turn suggests that a sizable part of δ_i^{miss} might be due to poor estimation of the MHO in PDF evolution, rather than poor estimation of MHO in hard cross-sections which can vary substantially between different processes (and indeed different kinematics). Indeed, as already mentioned in Sect. 3.5, our current treatment of factorization scale variation is only approximate, and a more

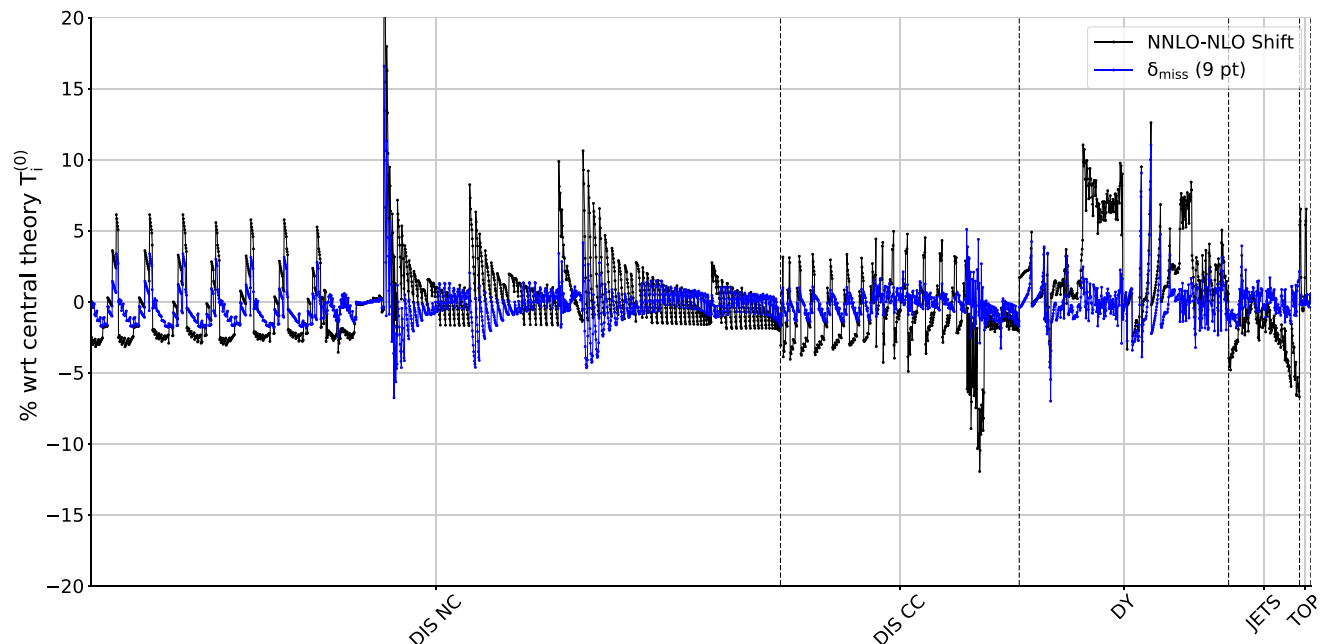


Fig. 11 The NNLO-NLO shift δ_i (black) compared to its component δ_{miss} (blue) which lies outside the subspace S , computed using the 9-point prescription

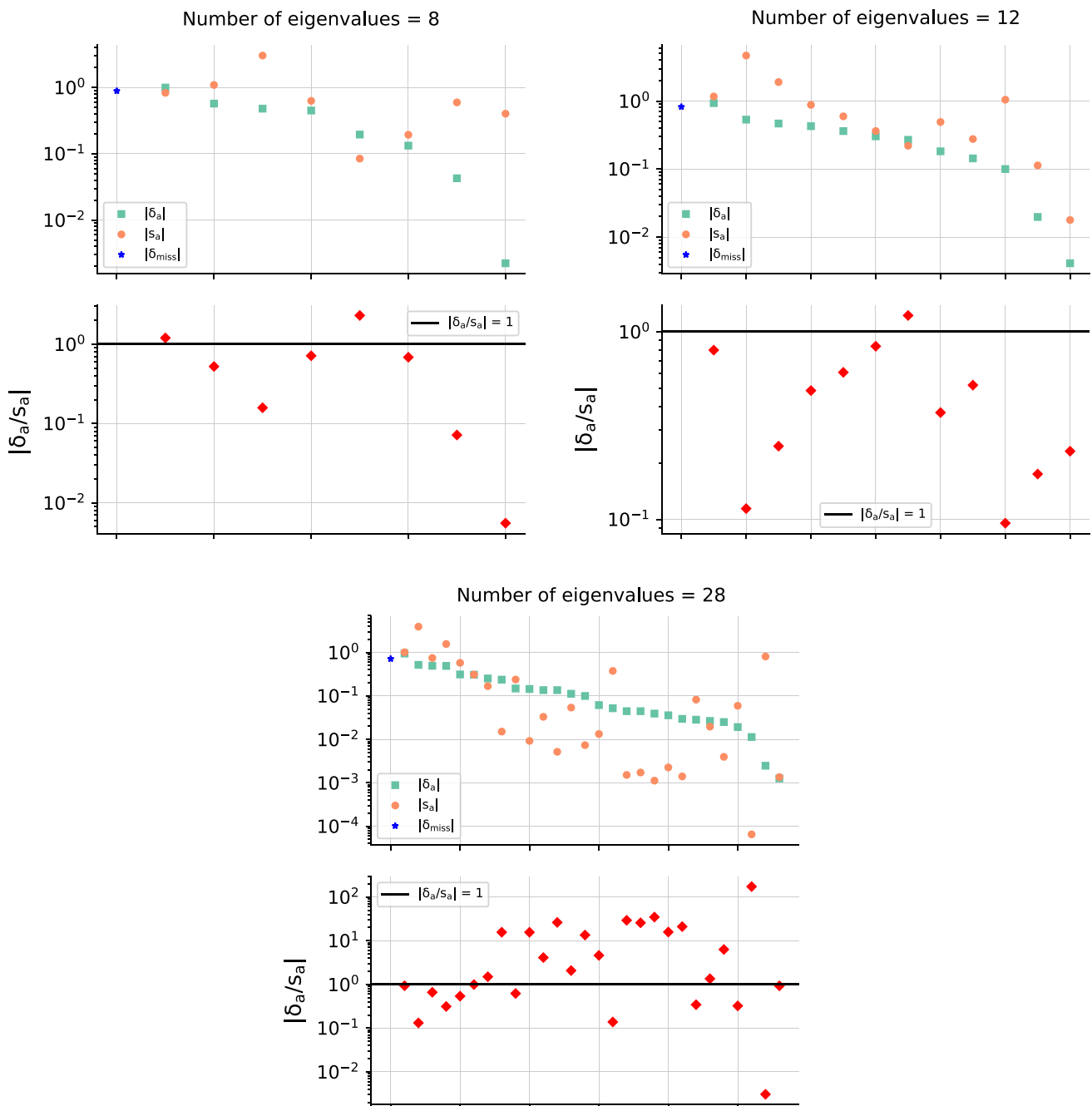


Fig. 12 The projection δ^α Eq. (5.5) of the normalized shift vector δ_i Eq. (5.4) along each eigenvector e_i^α of the normalized theory covariance matrix Eq. (5.3), compared to the corresponding eigenvalue s^α , ordered by the size of the projections (from largest to smallest). In each case results are shown as absolute (upper) and as ratios δ^α/s^α (lower),

the horizontal line indicating when this ratio is one. The length of the component of δ_i that is not captured at all by the theory covariance matrix, $|\delta_i^{\text{miss}}|$ is also shown (blue star). Results are shown for the symmetric prescriptions: 5-point (top left), $\bar{5}$ -point (top right), and 9-point (bottom)

sophisticated treatment would involve performing separate scale variation for each eigenvalue of perturbative evolution.

Having established that most of the NNLO-NLO shift δ_i lies within S , we now proceed to examine what fraction of δ_i^S lies with the error ellipse E specified by the theory covariance

matrix. To that end, the eigenvalues $\lambda^\alpha = (s^\alpha)^2$ of the theory covariance matrix of the global dataset are shown in Fig. 12 for symmetric prescriptions, and in Fig. 13 for the asymmetric ones: these define the length of the semi-axes of E . Since there are five distinct processes, there are 8, 12 and 28

positive eigenvalues for the symmetric 5-point, $\bar{5}$ -point and 9-point prescriptions respectively, and 6, 14 positive eigenvalues for the asymmetric 3-point and 7-point prescriptions, as explained in Appendix A. Also shown are the projections δ^α of the normalized shift vector δ Eq. (5.4) along each corresponding eigenvector e_i^α , Eq. (5.5).

Inspection of these plots confirms that all the prescriptions seem to perform reasonably well. The largest eigenvalue is always very similar in size to the shift, and the size of the eigenvalues generally falls as the projected shifts get smaller. As expected, the 3-point prescription clearly overestimates uncertainties, since $\delta^\alpha < s^\alpha$ for all the eigenvalues. The same is true, but to a lesser extent, for both 5-point and $\bar{5}$ -point. For the more complicated 7-point and 9-point prescriptions the largest projections (corresponding to the first seven or eight eigenvalues) are estimated rather well, though still perhaps a little conservatively, but for the smaller projections the scatter increases significantly, with some projected shifts hardly predicted at all. This is perhaps not surprising: when varying just six independent scales, we can only expect to obtain a limited amount of information on the MHO terms. However the correct estimation of the largest projected shifts shows that the theory covariance matrix is giving a reasonable estimation of the MHO, especially when implemented through the more complicated prescriptions.

On each of these plots, we also show the length of the component δ_i^{miss} that is orthogonal to S , and thus completely outside E . For the symmetric prescriptions, $|\delta_i^{\text{miss}}|$ is always less than the largest component of δ in S , while for the asymmetric prescriptions it is greater, very significantly so for the 3-point prescription. This is another indication that the symmetric prescriptions give a better account of the correlations in theoretical uncertainties.

A more detailed understanding of the physical meaning of each eigenvector can be acquired by inspecting its components e_i^α in the data space. These are shown in Fig. 14 for the eigenvectors corresponding to the five largest eigenvalues in the 9-point prescription: the shift vector δ_i is also shown for comparison. It is clear that there is a close correspondence between eigenvectors and MHO contributions to individual processes. For instance the first eigenvector contributes mostly to DIS NC, the second to both DIS NC and DIS CC, the third to DY, the fourth mainly to DIS CC, and the fifth mainly to JETS. Clearly the ordering of these larger eigenvalues is related to the number of data points for the respective processes: the more datapoints, the larger the eigenvalue of the (correlated) uncertainty estimate. Even relatively small eigenvalues can give an important contribution, though to processes with fewer datapoints: for example the ninth eigenvector (not shown) clearly dominates TOP.

In summary, from these validation tests it is apparent that the 9-point prescription gives a reasonable estimate of most of the MHO, both for individual processes and for the global

dataset, with the 7-point being just slightly worse. Based on this, we will therefore adopt 9-point as a default prescription for the theory covariance matrix in the PDF determination to be discussed in the next section.

6 PDFs with missing higher order uncertainties

We can now present the main results of this work: the first determination of the parton distributions of the proton which systematically accounts for the MHOUs affecting the theory calculations of the input processes for the fit. First we present the results for PDFs obtained by fitting only DIS data. This provides us with an initial test case, which we will study by comparing PDFs obtained including the combined experimental and theoretical covariance matrix to the corresponding baseline fit in which only experimental uncertainties are included.

We then turn to the global PDF determination, which offers a nontrivial validation of our methodology, specifically by comparing NLO PDFs, with and without MHOUs, to NNLO PDFs. For global fits, we also study the stability of the results to changes in the prescription used for the computation of the theory covariance matrix: specifically, we compare PDFs obtained with the 9-point prescription (which is our default) to those based on the 7- and 3-point ones. We also study PDFs determined by only partially including the theory covariance matrix, either only in the data generation or only in the fitting. As discussed in the introduction, this provides us with a way of disentangling the impact of the theory covariance matrix on the central value of the PDFs or on the PDF uncertainty.

As discussed in Sect. 2, the theory uncertainties are included by simply replacing the experimental covariance matrix C_{ij} with the sum $(C+S)_{ij}$ of the experimental and theory covariance matrices in the expression for the likelihood of the true value given the data. The NNPDF methodology, as used specifically in the determination of the most recent NNPDF3.1 PDF set [5], is otherwise unchanged. Within this methodology, the covariance matrix is used to generate N_{rep} pseudodata replicas $D_i^{(k)}$ for each datapoint i , with $k = 1, \dots, N_{\text{rep}}$, whose distribution must reproduce the covariance of any two data points. This means that with theory uncertainties included,

$$\lim_{N_{\text{rep}} \rightarrow \infty} \frac{1}{(N_{\text{rep}} - 1)} \sum_{k=1}^{N_{\text{rep}}} \left(D_i^{(k)} - \langle D_i \rangle \right) \left(D_j^{(k)} - \langle D_j \rangle \right) = C_{ij} + S_{ij}, \quad (6.1)$$

with $\langle D_i \rangle = \frac{1}{N_{\text{rep}}} \sum_{k=1}^{N_{\text{rep}}} D_i^{(k)}$ denoting the average over Monte Carlo replicas.

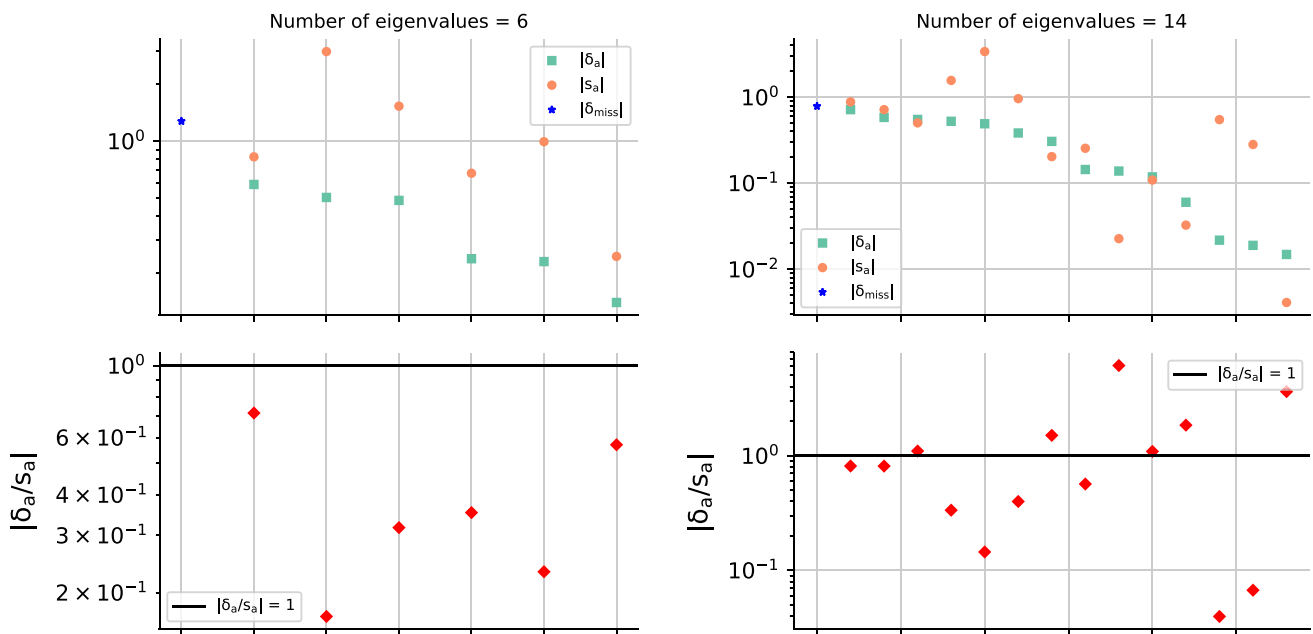


Fig. 13 Same as Fig. 12 but for the asymmetric prescriptions: 3-point (left) and 7-point (right)

A PDF replica is then fitted to each pseudodata replica $D_i^{(k)}$ by minimizing a figure of merit, which in the presence of theory uncertainties becomes

$$\chi^2 = \frac{1}{N_{\text{dat}}} \sum_{i,j=1}^{N_{\text{dat}}} (D_i - T_i) (C + S)_{ij}^{-1} (D_j - T_j), \quad (6.2)$$

where T_i is the theory prediction evaluated with the central scale choice, and the theory covariance matrix S_{ij} is computed using one of the prescriptions presented in Sect. 4.

It is thus clear that the inclusion of a theory-induced contribution in the covariance matrix affects only two steps of the procedure: the pseudodata generation, and the minimization. Everything else is unchanged, and is identical to the default NNPDF methodology. Note that in particular the experimental covariance matrix C used in the fitting is determined, as in NNPDF3.1 and previous NNPDF releases using the so-called t_0 method for the treatment of multiplicative uncertainties, in order to avoid d’Agostini bias (see Refs. [27, 80] for a detailed discussion). As in previous NNPDF releases, minimization is thus performed using the t_0 definition of the χ^2 , but all χ^2 values shown are computed using the covariance matrix as published by the respective experiments.

In the sequel, in order to assess fit quality we will provide χ^2 values, and also, we will study the estimator, defined in Ref. [4]

$$\phi = \sqrt{\langle \chi_{\text{exp}}^2 [T_i] \rangle - \chi_{\text{exp}}^2 [\langle T_i \rangle]}, \quad (6.3)$$

where by $\chi_{\text{exp}}^2 [T_i]$ we denote the value of the χ^2 computed using the i -th PDF replica, and only including the experimental covariance matrix (thus Eq. (6.2), but with S_{ij} set to zero). The average χ^2 values which enter Eq. (6.3) are then $\langle \chi^2 [T_i] \rangle$, the mean value of this χ^2 averaged over replicas, and $\chi^2 [\langle T_i \rangle]$, the value of the χ^2 computed using the “central” PDF set which is found by averaging over replicas.

It was shown in Ref. [4] that ϕ then gives the average over all datapoints of the ratio of the uncertainties of the predictions to the uncertainties of the original experimental data, taking account of correlations:

$$\phi = \left(\frac{1}{N_{\text{dat}}} \sum_{i,j=1}^{N_{\text{dat}}} (C)_{ij}^{-1} T_{ij} \right)^{1/2}, \quad (6.4)$$

where $T_{ij} = \langle T_i T_j \rangle - \langle T_i \rangle \langle T_j \rangle$ is the covariance matrix of the theoretical predictions. For an uncorrelated covariance matrix, this is just the ratio of the uncertainty in the prediction using the output PDF to that of the original data. Hence, the value of ϕ provides an estimate of the mutual theoretical consistency of the data which are being fitted: consistent data are combined by the underlying theory and lead to an uncertainty in the prediction which is significantly smaller than that of the original data. Note that ϕ is always defined so that the uncertainty in the prediction is normalized to the original experimental uncertainty (rather than combined experimental and theory uncertainties). In particular, when considering PDFs determined including a theory covariance matrix, this means that PDFs are determined minimizing the χ^2 Eq. (6.2), but χ_{exp}^2 is instead used in the computation of ϕ Eq. (6.3).

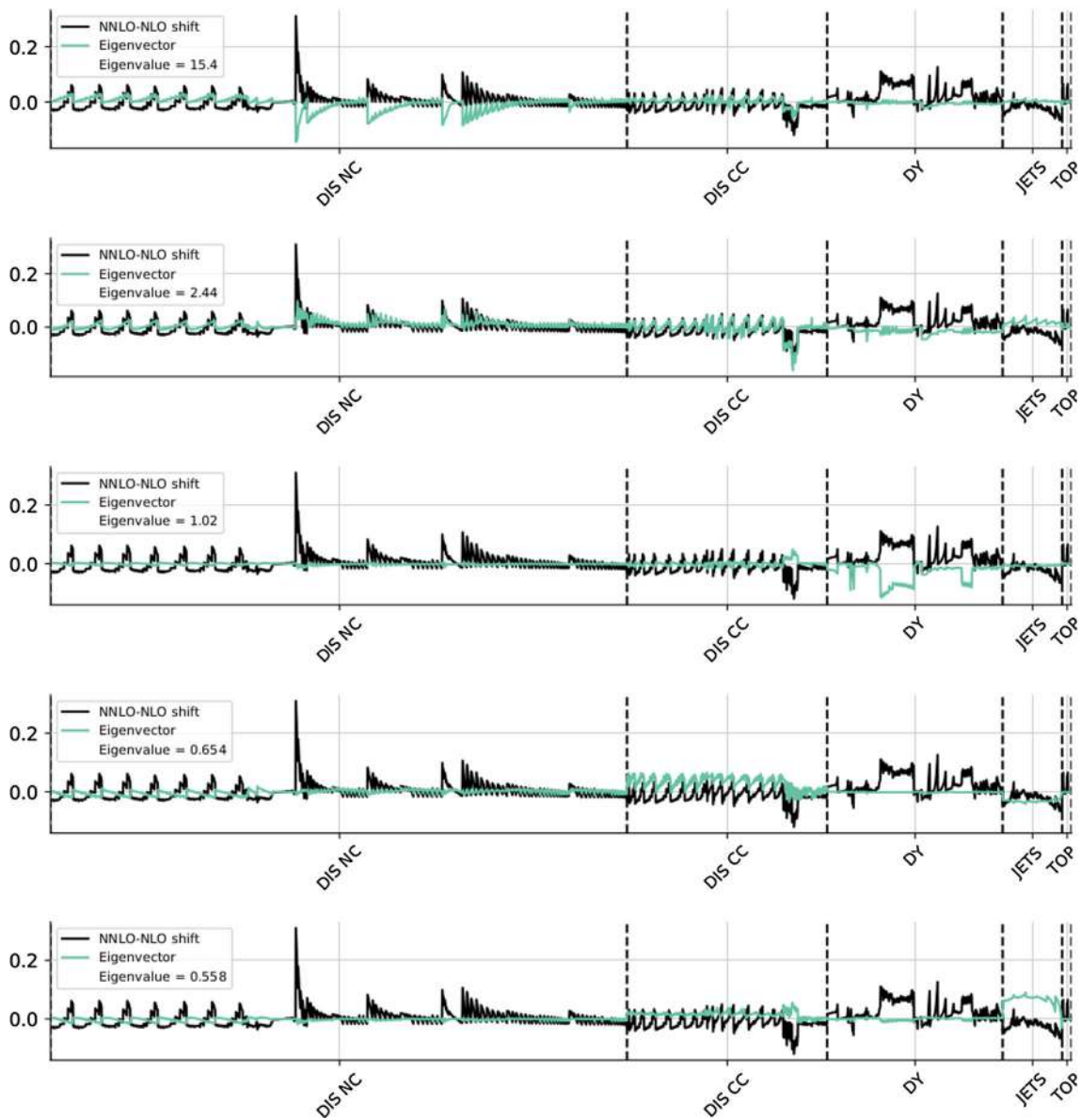


Fig. 14 The components e_i^α (green) of the eigenvectors, corresponding to the five largest eigenvalues for the 9-point theory covariance matrix, shown in the same format as Fig. 9. The NNLO-NLO shift, δ_i (black), is shown for comparison

When changing the covariance matrix from C to $C' = C + S$ the fluctuations of the replicas will change, according to Eq. (6.1), and if theoretical uncertainties change in the same proportion one would expect the value of ϕ to become $\phi' = r_\phi \phi$, with

$$r_\phi = \left(1 + \frac{1}{N_{\text{dat}}} \sum_{i,j=1}^{N_{\text{dat}}} (C)_{ij}^{-1} S_{ij} \right)^{1/2}. \tag{6.5}$$

Thus, when including MHO, all else being equal, we would expect PDF uncertainties to increase by a factor r_ϕ . This will

provide us with a baseline to which we can compare the change in uncertainty which is actually observed.

All the PDF sets which have been produced and which will be discussed in this section are listed in Table 5. For each of the fits, we indicate its label, the input dataset, the perturbative order and the covariance matrix used.

For the fits that include a theory covariance matrix, we also indicate the prescription with which it has been constructed. In the remainder of this section we discuss the main features of these PDF sets.

Table 5 Summary of the PDF sets discussed in this section. The dataset, perturbative order and nature of the treatment of uncertainties for each set are indicated

Label	Dataset	Order	Cov. mat.	Comments
NNPDF31_nlo_as_0118_dis_kF_1_kR_1	DIS	NLO	C	Baseline DIS-only NLO
NNPDF31_nlo_as_0118_dis_scalecov_9pt	DIS	NLO	$C + S^{(9pt)}$	
NNPDF31_nnlo_as_0118_dis_kF_1_kR_1	DIS	NNLO	C	Baseline DIS-only NNLO
NNPDF31_nlo_as_0118_kF_1_kR_1	Global	NLO	C	Baseline Global NLO
NNPDF31_nlo_as_0118_scalecov_9pt	Global	NLO	$C + S^{(9pt)}$	
NNPDF31_nlo_as_0118_scalecov_7pt	Global	NLO	$C + S^{(7pt)}$	
NNPDF31_nlo_as_0118_scalecov_3pt	Global	NLO	$C + S^{(3pt)}$	
NNPDF31_nlo_as_0118_scalecov_9pt_fit	Global	NLO	$C + S^{(9pt)}$	S only in χ^2 definition
NNPDF31_nlo_as_0118_scalecov_9pt_sampl	Global	NLO	$C + S^{(9pt)}$	S only in sampling
NNPDF31_nnlo_as_0118_kF_1_kR_1	Global	NNLO	C	Baseline Global NNLO

6.1 DIS-only PDFs

We first discuss PDF sets based on DIS data only. Fit quality indicators are collected in Table 6. The theory covariance matrix is always constructed using the 9-point prescription. We show the value of χ^2/N_{dat} and of the ϕ estimator defined in Eqs. (6.2, 6.3) respectively. Results are shown for both the total dataset and for the individual DIS experiments of Table 2. Note that the total χ^2 is no longer just the weighted sum of the individual χ^2 s, because it now also includes correlations between experiments.

It is apparent from Table 6 that in all cases the χ^2 improves when including the theory covariance matrix, both for individual experiments and for the total dataset. Specifically, the χ^2 decreases by about 2-3% when including theory a covariance matrix $S^{(9pt)}$ evaluated with the 9-point prescription.

The value of ϕ increases very substantially, suggesting a significant increase in the PDF uncertainty. The expected increase according to Eq. (6.5) is $r_\phi = 2.07$: NLO MHOUs in DIS are much larger than experimental uncertainties. The observed increase, by a factor of 2.17, is in good agreement

with this expectation. It is interesting to observe that the NNLO value of ϕ is actually also rather larger than the NLO value, though not quite so much larger, suggesting that at NNLO the MHOUs in DIS might still be quite large.

Next we compare PDFs: in Fig. 15 we compare the gluon and the total quark singlet PDF at $Q = 10$ GeV with and without MHOUs in the covariance matrix, determined using the 9-point prescription.

The NLO results are also compared with the central value of the NNLO fit based on the experimental covariance matrix only. Note that in these comparison plots the PDF uncertainty band is always computed using standard NNPDF methodology, i.e., as the standard deviation over the PDF replica sample. Therefore, this uncertainty band has a different meaning dependent on whether or not the theory covariance matrix is included: when it is not included, the band represents the conventional ‘‘PDF uncertainty’’, reflecting the uncertainties from the data (and methodology), while when it is included, the band provides the combined ‘‘PDF’’ and MHO uncertainty.

Table 6 The values of the χ^2/N_{dat} and of the ϕ estimator in the NNPDF3.1 DIS-only fits with the theory covariance matrix $S^{(9pt)}$, compared to the results based on including only the experimental covariance matrix C

NNPDF3.1 DIS-only fits							
Dataset	n_{dat}	χ^2/N_{dat}			ϕ		
		NLO		NNLO	NLO		NNLO
		C	$C + S^{(9pt)}$	C	C	$C + S^{(9pt)}$	C
NMC	134	1.259	1.236	1.239	0.388	0.464	0.444
SLAC	12	0.908	0.543	0.791	0.247	0.590	0.343
BCDMS	530	1.046	1.017	1.047	0.339	0.505	0.389
CHORUS	430	0.982	0.856	1.124	0.409	0.418	0.512
NuTeV	41	0.628	0.491	0.872	0.940	0.994	1.35
HERA incl	967	1.097	1.066	1.104	0.280	1.013	0.335
HERA F_2^c	31	1.047	0.997	1.033	0.526	1.097	0.631
Total	2145	1.061	1.032	1.095	0.358	0.780	0.441

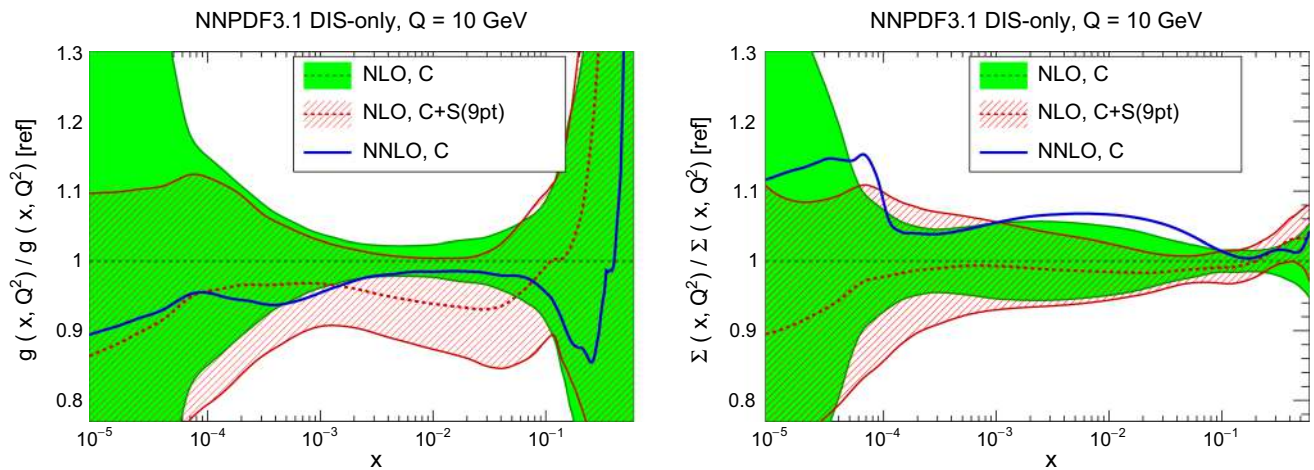


Fig. 15 Comparison of DIS-only PDFs determined with and without MHOUs in the covariance matrix. The gluon (left) and quark singlet (right) are shown at $Q = 10$ GeV. The theory covariance matrix S has been constructed using the 9-point prescription. The central value of the NNLO determined without MHOUs is also shown. All results are shown

The comparison shows that for PDFs which are strongly constrained by data, such as the quark singlet PDF for $x \gtrsim 10^{-3}$, the uncertainty does not increase much upon inclusion of the theory covariance matrix, and sometimes it even decreases. However, for several PDFs, including the gluon PDF, which is only loosely constrained by the DIS data, the uncertainty increases substantially with MHOUs. This is of course consistent with the fact that, in the absence of stringent experimental constraints, an extra contribution to the covariance matrix will lead to increased uncertainties in the best fit.

6.2 Global PDFs

We now discuss PDFs determined from the global dataset presented in Sect. 5.1. Only NLO PDFs will be discussed here, with global NNLO PDFs left for future work. The χ^2 values and ϕ values are shown in Tables 7 and 8 respectively, both for the total dataset and for the individual processes of Table 2. In comparison to the DIS-only case of Table 6 we now also show results obtained using the 7-point and 3-point prescriptions, and also for the default 9-point prescription but where the results were obtained by including the theory covariance matrix either only in the χ^2 definition Eq. (6.2), or only in the data generation Eq. (6.1), in order to understand better the two distinct effects. The baseline NLO and NNLO PDF sets (without theory covariance matrix) are identical in all respects to the NNPDF3.1 PDF sets [68], except for the somewhat different dataset as discussed in Sect. 5.1.

As in the case of the DIS-only fit, upon adding the MHOUs we find a reduction of χ^2 both for the global fit and for individual datasets. Specifically, the χ^2 for the NLO global fit

as a ratio to the central value of the set with theory covariance matrix not included. Note that the uncertainty band has a different meaning according to whether the theory covariance matrix is included or not: if not it is the standard PDF uncertainty coming from data, while if it is included, then it is the total uncertainty including the MHOUs

with theory covariance matrix computed with the 9-point prescription decreases by about 3%, and almost coincides with the NNLO χ^2 , suggesting that indeed the theory uncertainty is correctly accounting for the missing NNLO correction. The pattern at the level of individual datasets is more complex, due to a variety of reasons. In particular, consider the CMS $Z p_T$ distribution, where a very significant decrease in χ^2 is observed when going from NLO to NNLO, but not when adding the theory uncertainty to the NLO. This turns out to be due to a sizable uncorrelated uncertainty which must be added to the NNLO theory prediction in order to account for numerical instabilities (see the discussion of Fig. 6 in Ref. [68]).

On the other hand, the value of ϕ now increases much less than expected: with our favorite 9-point prescription the increase is by about 30%, while the expected $r_\phi = 1.69$. This is an indication that by accounting for the missing NNLO terms, the inclusion of MHOUs resolves some of the tensions in the fit with only the experimental uncertainties, thus reducing the overall effect of the MHOUs. The NNLO fit also shows an increase, to 0.36, but the fact that this is already quite close to 0.41 perhaps suggests that the effect of adding MHOUs to the NNLO global fit will be relatively modest.

Comparing the different prescriptions, results are reasonably stable, even when comparing to the 3-point prescription which, as discussed in Sects. 4.3–5.3, spans a much smaller subspace of theory variations. However, the 9-point prescription appears to perform best in terms of χ^2 quality with very little difference in ϕ , in agreement with the results of Sect. 5.4.

We finally turn to fits in which the theory covariance matrix is included either in the χ^2 definition Eq. (6.2) but

not in the data generation Eq. (6.1), or in the data generation Eq. (6.1) but not in the χ^2 definition Eq. (6.2). In the former case, we expect the MHOUs to affect mostly the central value (since the relative weighting of different data points is altered during the fitting according to the relative size of their MHOUs), and to a lesser extent the uncertainties (since the data replicas only fluctuate according to the experimental uncertainties). The results show that indeed including the MHOUs in the χ^2 definition alone leads to a χ^2 value which is very close to that found when the MHOUs is fully included, consistent with the expectation that it is the inclusion of the theory covariance matrix in the χ^2 which mostly drives the best fit, while the ϕ value increases somewhat less. In the latter case, we expect to obtain increased uncertainties but a worse fit, since the data replica fluctuations are wider due to the MHOUs, and this is not accounted for in the χ^2 . The results indeed show a significant deterioration of fit quality, as expected for an inconsistent fit: the χ^2 goes up, and also the ϕ value goes up dramatically, showing the increase in uncertainty due to the inclusion of MHOUs in the sampling, now uncompensated by a rebalancing of the datasets through the inclusion of MHOUs in the fit.

We now move on to discuss the corresponding results at the PDF level, in analogy with the comparisons presented for the DIS-only fits in Fig. 15. Specifically, in Fig. 16 we show the results of the NLO fits based on C and $C + S^{(9pt)}$, as well as the central value of the NNLO fit based on C , for the gluon, the total quark singlet, the anti-down quark, and the total strangeness PDFs, all at $Q = 10$ GeV. We also show in Fig. 16 the same PDFs but at the scale $Q = 1.6$ GeV at which PDFs are parametrized.

We find that in the data region the PDF uncertainty is only very moderately increased by the inclusion of the theory covariance matrix, while central values can shift significantly, by up to one sigma. This is consistent with the observation that the ϕ values in Table 8 increase by only a moderate amount upon inclusion of the theory covariance matrix. This provides evidence that in the data region the inclusion of the theory covariance matrix resolves tensions which are otherwise present in the global dataset. In contrast, in regions where PDFs which are only loosely constrained by the data, and in particular in the extrapolation regions, the PDF uncertainty increases significantly.

When comparing PDFs at the parametrization scale in Fig. 17, an especially interesting comparison is with respect to the central NNLO value: not only is this quite compatible with the uncertainty band, but there is now clear evidence that upon inclusion of the NLO MHOUs the central best fit moves towards the correct NNLO result. Of course, this improved agreement of the best-fit NLO and NNLO PDFs is scale-dependent, since PDFs at NLO and NNLO evolve in different ways, and the scale at which NLO and NNLO become closest will depend on the scale of the data which dominate the deter-

mination of each PDF combination. However, the agreement is seen in Fig. 16 to persist by and large also at high scale. This is further evidence that indeed the theory covariance matrix has resolved tensions due to MHOUs. This improved agreement of the central value of the NLO $C + S^{(9pt)}$ with the NNLO C fits is non-trivial: for instance, inclusion of the theory covariance matrix leads to a suppression of the gluon at large x and an enhancement of strangeness, both of which are indeed also observed at NNLO.

Next, in Fig. 18 we compare PDFs obtained using different prescriptions. The corresponding relative PDF uncertainties are compared in Fig. 19. In agreement with what we saw for the χ^2 and ϕ values in Tables 7, 8 results are quite stable with respect to the choice of prescription, though in the most extreme case of the 3-point prescription, where factorization scale variations are entirely uncorrelated between different processes, we observe somewhat smaller uncertainties, and a central value which is closer to that when the MHOUs is not included.

Finally, in Fig. 20 we compare PDFs obtained including the theory covariance matrix only in the χ^2 definition Eq. (6.2) but not in the data generation Eq. (6.1) and conversely. We see that when the theory covariance matrix is included in the replica generation but not in the χ^2 , uncertainties increase very significantly. This result is in agreement with the observation from Table 7 that in this case the fit quality significantly deteriorates, which is because the fit becomes inconsistent due to the χ^2 not matching the wider fluctuations in the data. The effect is particularly visible for the quark distributions. On the other hand, including the theory covariance matrix only in the χ^2 singles out the effect of the theory covariance matrix on central values, due to rebalancing of datapoints in the fit according to their relative MHOUs. Indeed in this case the central value is very close to that obtained when including the MHOUs is both data generation and fit. We also see that the change in uncertainties in the data region is now very small, consistent with Table 7. These results confirm our expectation that in the full fit, while the MHOUs results in a substantial increase in the fluctuations of data replicas, this is compensated by a relaxation of tensions due to the inclusion of MHOUs the fit, with the net result that while central values shift, overall uncertainties do not increase much.

7 Implications for phenomenology

Whereas a full assessment of the impact of the inclusion of MHOUs in PDFs will be possible only once we have global NNLO sets with MHOUs, it is worth performing a first phenomenological investigation, by computing reference LHC standard candles with the NLO PDF sets which include MHOUs presented in Sect. 6, and comparing to results with

Table 7 The values of the χ^2/N_{dat} in NLO global fits with the theory covariance matrix S , compared to the results based on including only the experimental covariance matrix C . Results are shown for the 9-, 7-, and 3-point prescriptions. For the 9-point prescription we also show results obtained including the theory covariance matrix in the χ^2 defi-

inition Eq. (6.2) but not in the data generation Eq. (6.1) (marked $S_{\text{fit}}^{9\text{pt}}$) and then in the data generation Eq. (6.1) but not in the χ^2 definition Eq. (6.2) (marked $S_{\text{samp}}^{9\text{pt}}$). Values corresponding to the NNLO fit with experimental covariance matrix C only are also shown

Dataset	n_{dat}	χ^2/N_{dat} in the NNPDF3.1 global fits						NNLO C
		NLO						
		C	$C + S^{(9\text{pt})}$	$C + S^{(7\text{pt})}$	$C + S^{(3\text{pt})}$	$C + S_{\text{fit}}^{(9\text{pt})}$	$C + S_{\text{samp}}^{(9\text{pt})}$	
NMC	134	1.241	1.239	1.264	1.253	1.235	1.246	1.222
SLAC	12	0.868	0.503	0.485	0.509	0.493	0.738	0.693
BCDMS	530	1.040	1.029	1.046	1.062	1.033	1.042	1.062
HERA σ_{NC}^p	886	1.086	1.044	1.046	1.079	1.044	1.190	1.098
HERA σ_{NC}^c	31	1.395	1.037	1.082	1.172	1.055	1.563	1.163
DIS NC	1593	1.088	1.079	1.086	1.095	1.081	1.227	1.084
NuTeV dimuon	41	0.474	0.388	0.355	0.359	0.421	0.406	0.470
CHORUS	430	1.037	0.891	0.896	0.900	0.898	1.081	1.124
HERA σ_{CC}^p	81	1.154	1.070	1.067	1.106	1.062	1.103	1.126
DIS CC	552	1.012	0.928	0.933	0.960	0.929	1.036	1.079
ATLAS W, Z 7 TeV 2010	30	0.999	0.880	0.916	0.975	0.892	0.984	0.935
ATLAS W, Z 7 TeV 2011	34	3.306	2.224	2.282	2.389	2.205	3.107	1.807
ATLAS low-mass DY 7 TeV	4	0.684	0.654	0.668	0.690	0.660	0.733	1.024
ATLAS high-mass DY 7 TeV	5	1.677	1.736	1.700	1.660	1.667	1.577	1.498
ATLAS Z p_T 8 TeV (p_T^l, M_{ll})	44	1.171	1.067	1.070	1.067	1.062	1.183	0.907
ATLAS Z p_T 8 TeV (p_T^l, y_{ll})	48	1.666	1.583	1.614	1.688	1.638	1.641	0.865
CMS Drell–Yan 2D 2011	88	1.220	1.067	1.098	1.169	1.062	1.132	1.319
CMS W asy 840 pb	11	0.965	1.022	0.966	0.987	1.045	1.034	0.863
CMS W asy 4.7 fb	11	1.662	1.670	1.704	1.713	1.659	1.657	1.750
CMS W rap 8 TeV	22	0.955	0.611	0.609	0.587	0.627	0.665	0.826
CMS Z p_T 8 TeV (p_T^l, M_{ll})	28	3.895	3.745	3.712	3.836	3.706	3.905	1.339
LHCb Z 940 pb	9	1.238	1.191	1.162	1.179	1.165	1.281	1.437
LHCb $Z \rightarrow ee$ 2 fb	17	1.305	1.303	1.305	1.313	1.334	1.250	1.203
LHCb $W, Z \rightarrow \mu$ 7 TeV	29	1.262	1.106	1.267	1.261	1.134	1.207	1.536
LHCb $W, Z \rightarrow \mu$ 8 TeV	30	1.194	1.027	1.125	1.154	1.054	1.152	1.438
CDF Z rap	29	1.554	1.313	1.433	1.505	1.311	1.418	1.510
D0 Z rap	28	0.649	0.601	0.626	0.640	0.597	0.618	0.604
D0 $W \rightarrow ev$ asy	8	1.176	1.066	1.055	1.083	1.029	1.200	2.558
D0 $W \rightarrow \mu\nu$ asy	9	1.400	1.450	1.372	1.361	1.439	1.395	1.374
DY	484	1.486	1.447	1.485	1.483	1.461	1.434	1.231
ATLAS jets 2011 7 TeV	31	1.069	1.019	1.065	1.079	1.026	1.031	1.076
CMS jets 7 TeV 2011	133	0.869	0.786	0.790	0.830	0.795	0.883	0.921
JETS	164	0.907	0.839	0.858	0.901	0.848	0.911	0.950
ATLAS $\sigma_{t\bar{t}}^{\text{top}}$	3	2.577	0.787	0.853	0.982	0.770	2.442	0.903
ATLAS $t\bar{t}$ rap	10	1.258	0.955	0.867	0.910	0.935	1.355	1.424
CMS $\sigma_{t\bar{t}}^{\text{top}}$	3	0.984	0.170	0.234	0.333	0.158	0.859	0.140
CMS $t\bar{t}$ rap	10	0.950	0.910	0.923	0.933	0.916	0.942	1.039
TOP	26	1.260	1.012	1.016	1.077	1.001	1.264	1.068
Total	2819	1.139	1.109	1.129	1.139	1.113	1.220	1.105

Table 8 Same as Table 7, but for the values of the ϕ estimator

Process	ϕ in the NNPDF3.1 global fits						NNLO
	NLO						
	C	$C + S^{(9pt)}$	$C + S^{(7pt)}$	$C + S^{(3pt)}$	$C + S_{fit}^{(9pt)}$	$C + S_{saml}^{(9pt)}$	
DIS NC	0.266	0.412	0.393	0.384	0.414	1.137	0.305
DIS CC	0.389	0.408	0.427	0.442	0.388	0.502	0.471
DY	0.361	0.377	0.369	0.379	0.378	0.603	0.380
JETS	0.295	0.359	0.327	0.333	0.336	0.461	0.392
TOP	0.375	0.443	0.387	0.405	0.382	0.612	0.363
Total	0.314	0.405	0.394	0.394	0.400	0.932	0.362

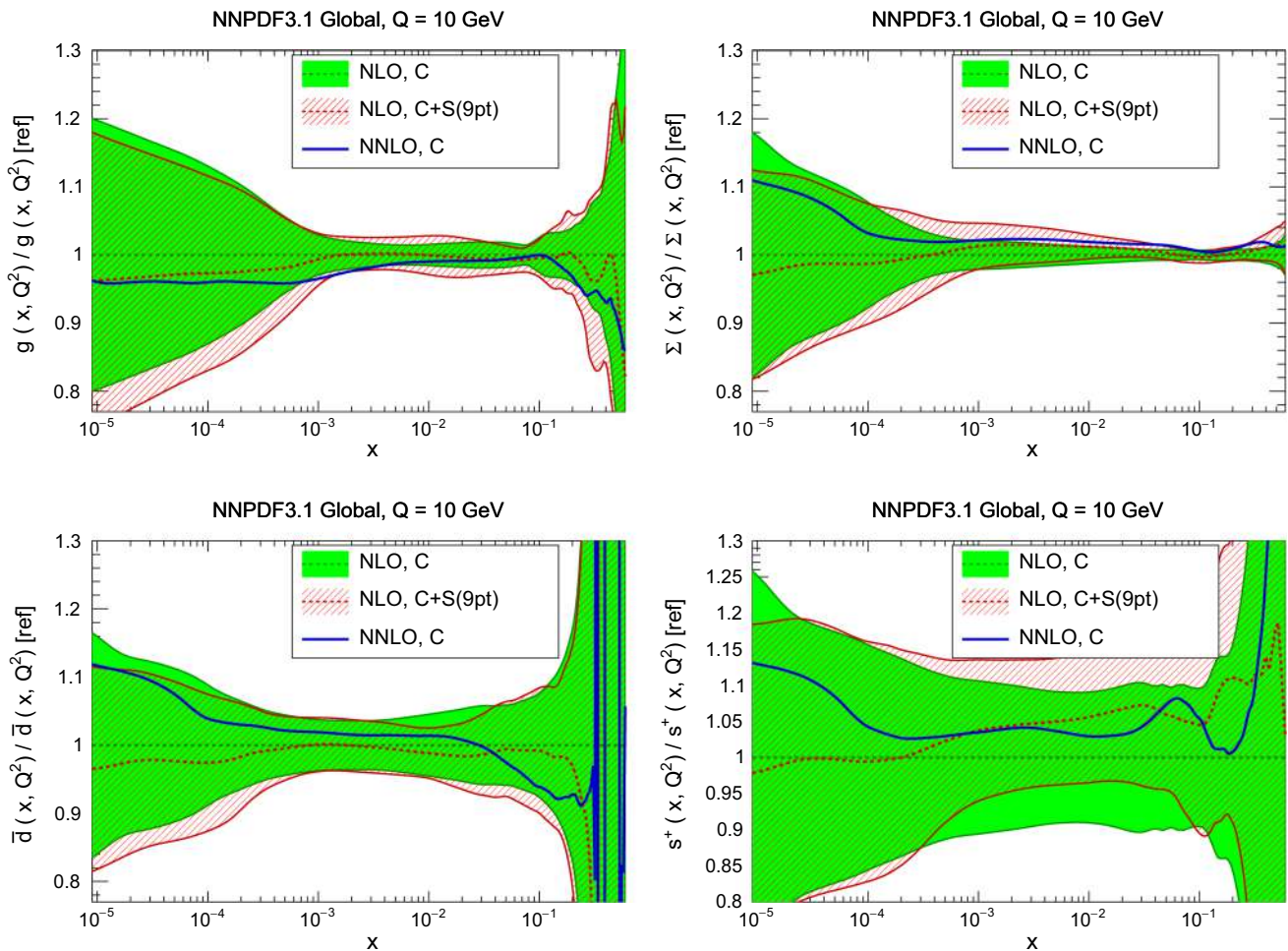


Fig. 16 Same as Fig. 15 now for the NNPDF3.1 global fits. We show the results of the NLO fits based on C and $C + S^{(9pt)}$ normalized to the former, as well as the central value of the NNLO fit based on C , for the gluon, the total quark singlet, the anti-down quark, and the total strangeness PDFs, all at $Q = 10$ GeV

the corresponding NLO PDF sets in which no MHOU is included.

In this section we will specifically consider Higgs boson production in gluon-fusion and in vector-boson fusion, top quark pair and Z and W electroweak gauge boson production. Note that the latter processes are among those which have been used for PDF determination, see Table 2. This raises the

issue of possible double counting of uncertainties between the MHOU in the PDF and in the hard matrix element. This will be addressed in Sect. 8.1 below.

As discussed in Sect. 6, once the MHOU is included in the covariance matrix, the standard NNPDF methodology can be used, but with the PDF uncertainties now also including a theory-induced contribution. Specifically, PDF uncertainties

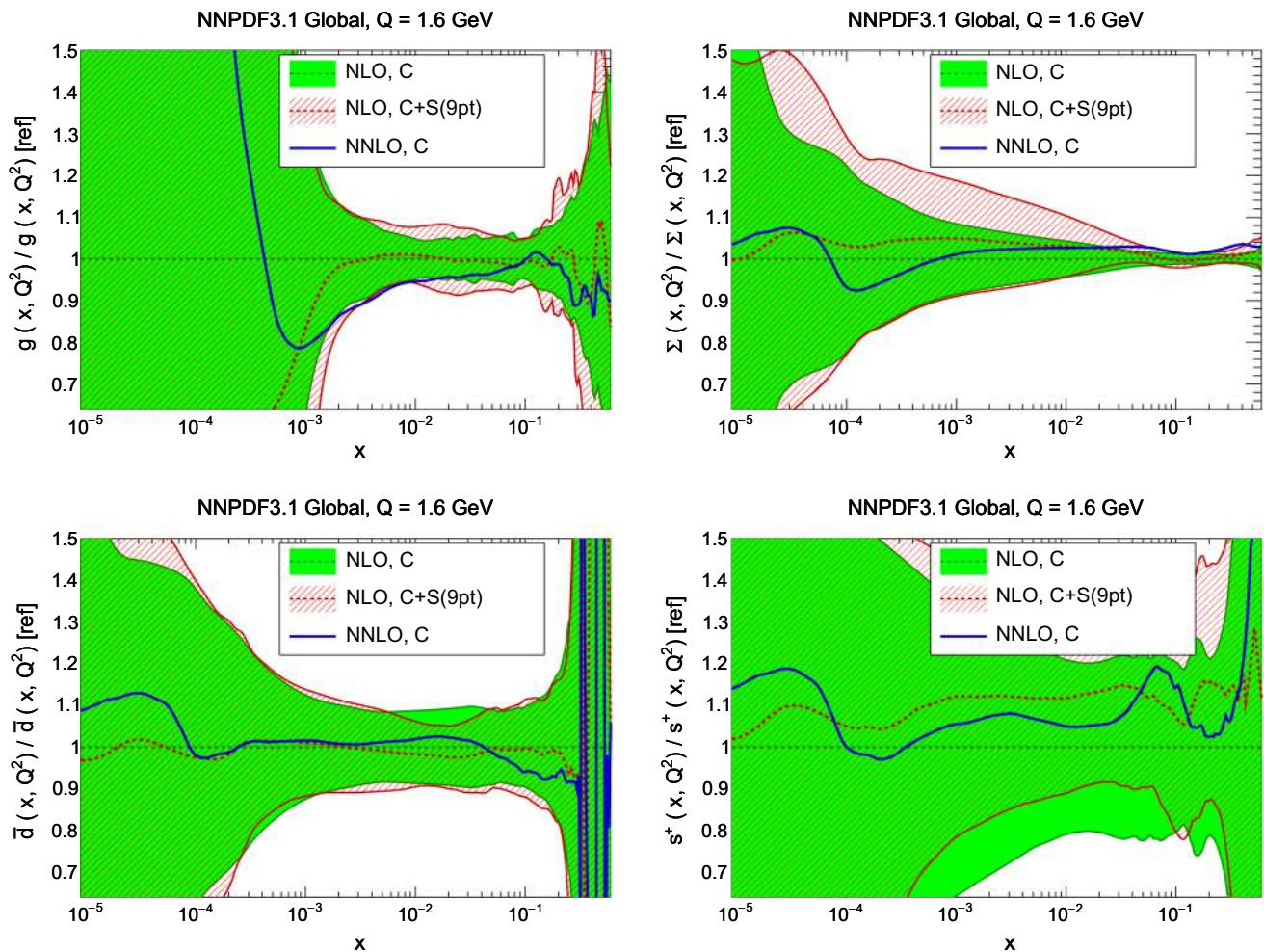


Fig. 17 Same as Fig. 16 but now with results shown at the scale $Q = 1.6$ GeV at which PDFs are parametrized

(which now include the MHOU uncertainty) are obtained as standard deviations over the replica sample. The total uncertainty on a physical prediction is then obtained by combining this uncertainty with that on the hard cross-section for the given process. The latter is conventionally obtained as the envelope of a 7-point scale variation, see e.g. Ref. [18]. Of course, an alternative possibility is to compute the theory uncertainty on the hard cross-sections in exactly the same way as we compute it when performing PDF determination, i.e. using the theory covariance matrix. In this case, the MHOU on any measurement is found as the diagonal element of the covariance matrix, evaluated for the given measurement. Here we will compute the theory uncertainty both using the theory covariance matrix (with the 9-point prescription, given in Eq. (4.9)), and as a 7-point envelope. The MHOU uncertainty on the hard cross-section can then be combined with the total uncertainty on the PDF (which includes both MHOU and data uncertainties) in quadrature. A more detailed discussion of prescriptions for the computation

of the total uncertainty on a physical observable, including explicit formulae, will be given in Sect. 8.1 below.

The current state of the art for precision phenomenology is NNLO, and thus NNLO PDFs would be needed for accurate predictions. However, as discussed in Sect. 6, at present only NLO global PDFs with MHOU are available. In principle, NNLO PDFs from a DIS only fit are also available. However, also as discussed in Sect. 6, some of these PDFs (specifically the gluon) are affected by large uncertainties due to the lack of experimental constraints. The comparison of PDFs with and without MHOU for such sets would thus be rather misleading. Therefore, in this section we will focus on NLO PDFs. It should of course be kept in mind that NNLO PDFs with MHOU are likely to have smaller uncertainties.

7.1 Higgs production

We first discuss Higgs production in gluon fusion (ggF) and in vector boson fusion (VBF). These two processes are of direct relevance for the characterization of the Higgs sector and

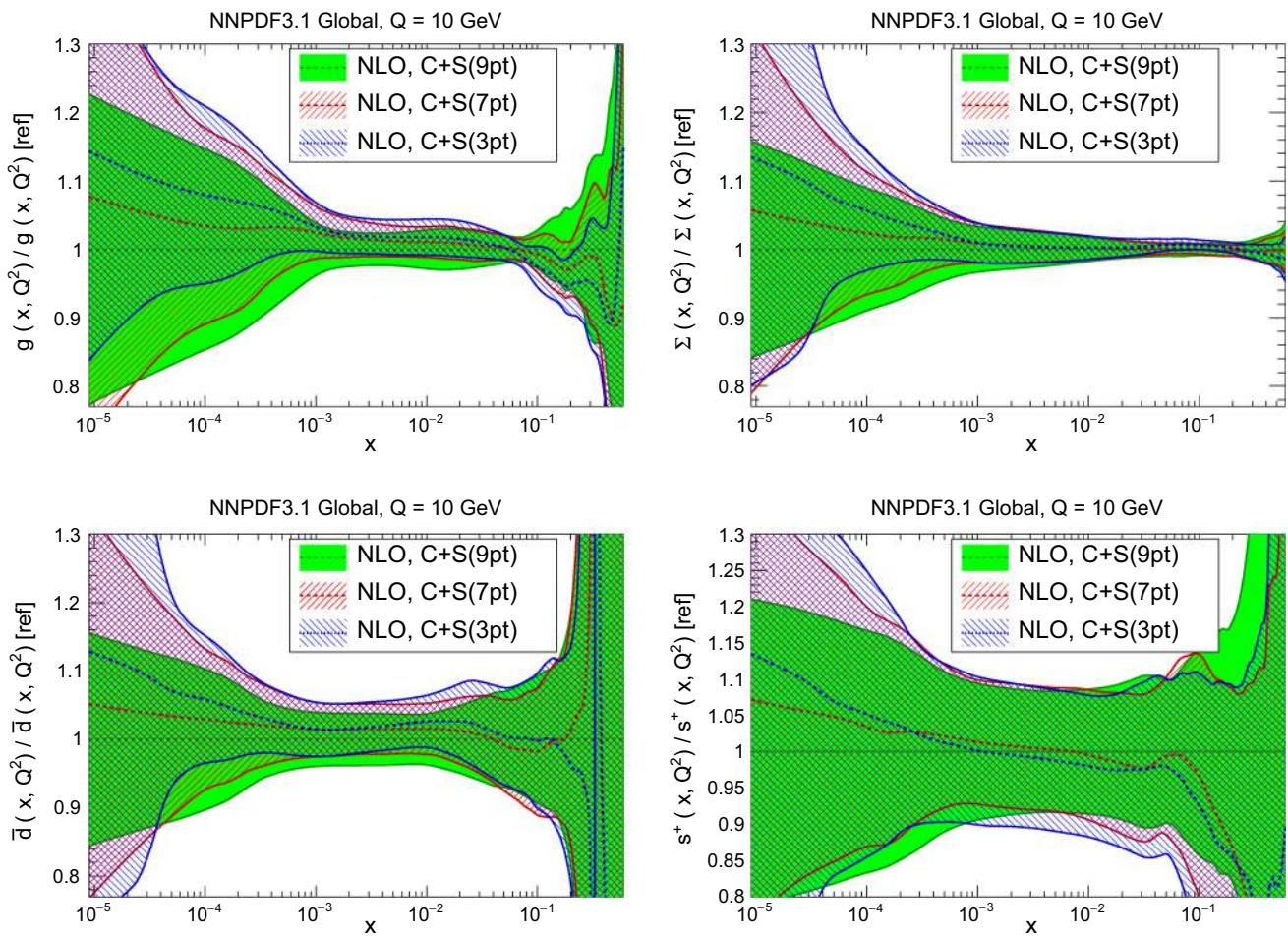


Fig. 18 Same as Fig. 16 now comparing the results of the NNPDF3.1 global fits with the theory covariance matrix constructed accordingly to the 3-, 7-, and 9-point prescriptions, normalized to the central value of the latter

are both currently known at $N^3\text{LO}$ accuracy [81–84]. Note that the perturbative behavior and leading partonic channels for these processes are quite different. Higgs production in gluon fusion is driven by the gluon-gluon luminosity and its perturbative expansion converges slowly, with manifest convergence reached only at $N^3\text{LO}$. Vector boson fusion is driven by the quark-antiquark luminosity and it exhibits fast perturbative convergence.

In Table 9 we present predictions for Higgs production in gluon fusion at the LHC for $\sqrt{s} = 13$ TeV. We perform the calculation at NLO, NNLO and $N^3\text{LO}$ in the rescaled effective theory approximation using `ggHiggs` [85–90] with $\mu_f = \mu_r = m_H/2$ as central scale, with the NLO global sets obtained in this paper, with and without MHOU, as input PDFs at all orders. The results are displayed graphically in Fig. 21, where, for the NNLO computation, we also show the central value found using NNLO PDFs.

We find that for all perturbative orders the central values obtained with PDFs with and without MHOU are very similar, while the PDF uncertainty is about 50% larger when

MHOU are included in the PDF fit. This can be understood by noticing that for the intermediate values of the momentum fraction, $x \simeq 10^{-2}$, relevant for Higgs production in gluon fusion, the PDF uncertainty of the gluon is increased in the $C + S^{(9\text{pt})}$ fit as compared to the C -only fits, see Fig. 16. Comparison to the result obtained using NNLO PDFs (for the NNLO computation) shows that upon inclusion of the MHOU the PDF uncertainty band of the result with NLO PDFs now includes the NNLO PDF result, while it would not in the absence of MHOU, both because of the (small) shift in central value and of the widening of the uncertainty band.

From Table 9 one can also observe that the MHOU on the hard matrix element uncertainty $\sigma_{\mathcal{F}}^{\text{th}}$ evaluated using the 9-point theory covariance matrix, Eq. (4.9), is compatible with the canonical 7-point envelope if the latter is symmetrized by taking the maximum value between the lower and upper uncertainties. In particular, the theory covariance matrix estimate is slightly larger than the envelope prescription at NLO and at NNLO, while it becomes a little smaller at $N^3\text{LO}$. Even

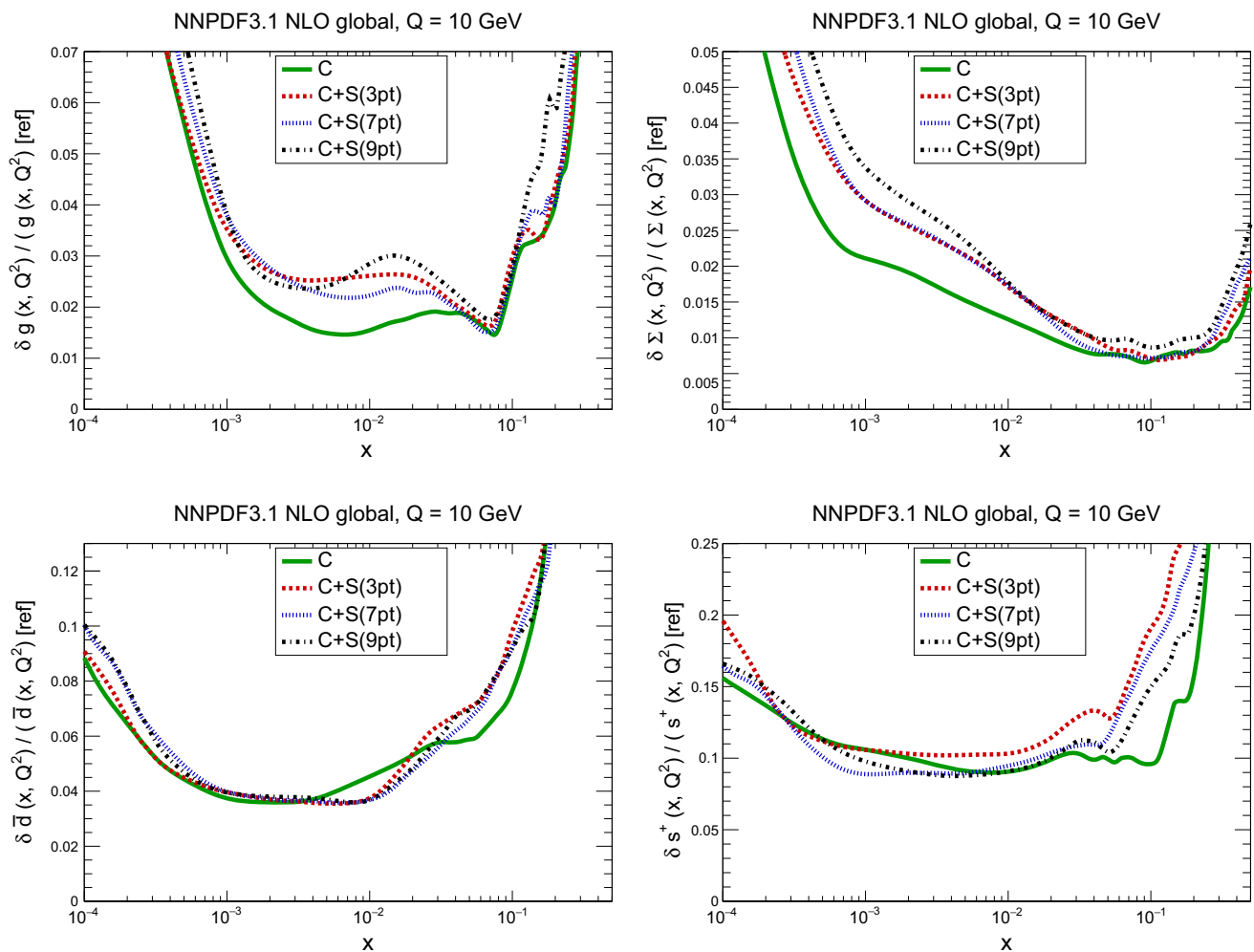


Fig. 19 Same as Fig. 18, now showing relative PDF uncertainties, normalized to the central value of the baseline set. Note that the y-axes ranges are different for each PDF combination

so, the NLO uncertainty band does not contain the NNLO central value, which lies just above the edge of the band.

We conclude that using NLO PDFs in the N^3 LO calculation, the inclusion of MHOU in the PDFs translates into a few per-mille increase of the PDF uncertainty at the cross-section level. In Ref. [82] NNLO PDFs were used with the N^3 LO calculation in order to provide a state-of-the-art result, and a MHOU uncertainty on the NNLO PDF was estimated based on the difference between results obtained using NLO and NNLO PDFs. Once NNLO PDFs with MHOUs determined within our approach are available it will be interesting to compare our results with this estimate.

We now turn to Higgs production in vector boson fusion. We perform the calculation at N^3 LO accuracy using `proVBFH-inclusive` [84,91] with central factorization and renormalization scales set equal to the squared four-momentum of the vector boson. Results are collected in Table 10 and shown in Fig. 21. The MHOU corrections to the PDFs are very small, so PDF uncertainties with or without

theory covariance matrix are very similar. Also in this case, like for gluon fusion, the uncertainty on the hard matrix element computed with the 9-point theory covariance matrix is similar to the one obtained by symmetrizing the 7-point envelope.

The smallness of the MHOU in the PDF follows from the fact that VBF Higgs production is driven by the quark-antiquark luminosity, which in turn is dominated by the quark PDF in the data region, whose uncertainties, as we have seen in Sect. 6.2, are almost unaffected by the inclusion of MHOU. Comparison to the result obtained using NNLO PDFs (for the NNLO computation) shows that the NNLO PDF result is at the edge of the PDF uncertainty band of the result with NLO PDFs if MHOU are included, while it is off by almost two σ if they are not. This is essentially due to the significant shift in central value, in agreement with the observation made in Sect. 6.2, where we noticed that MHOUs have the effect of moving the central value of the PDFs in the data region towards the NNLO result. The shift in the central value of

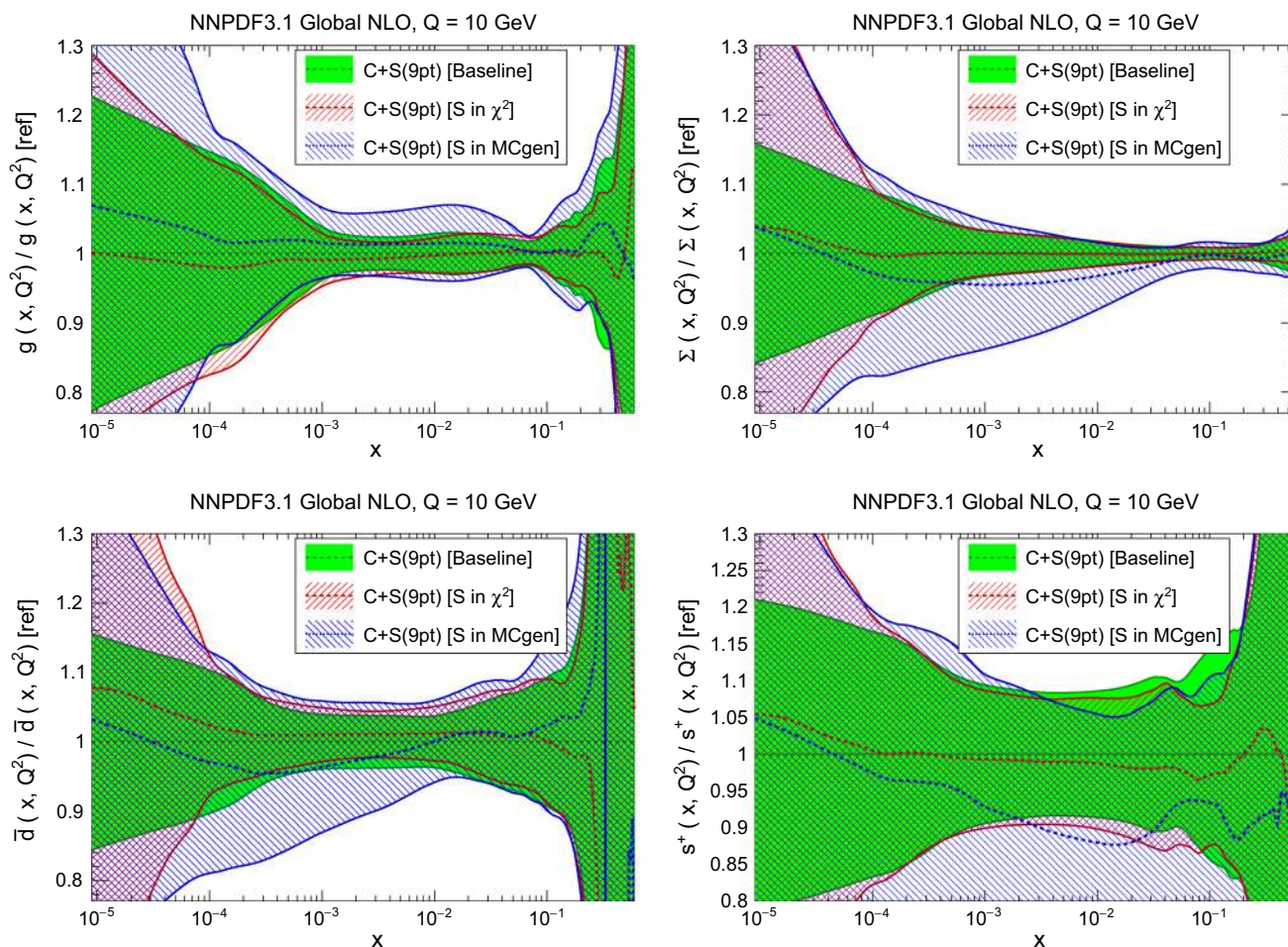


Fig. 20 Same as Fig. 16, now comparing the results of the baseline $C + S^{(9pt)}$ fit with those in which the theory covariance matrix S is included either in the χ^2 definition or in the generation of Monte Carlo replicas, but not in both

Table 9 The total cross-sections for Higgs production in gluon fusion (in pb) obtained by using NLO global PDFs based on either C or $C + S^{(9pt)}$, see Table 5. We quote the central prediction, the total PDF uncertainty (first) and the MHOU uncertainty on the hard cross-section (second) expressed as a percentage of the central value. The latter is

evaluated both using the theory covariance matrix (9-point prescription) or, in parenthesis, a (symmetrized) envelope of the 7-point scale variations (see Sect. 8.1), obtained by taking the maximum value between the lower and upper uncertainties

Higgs production in gluon fusion at 13 TeV		
	C	$C + S^{(9pt)}$
NLO	$37.63 \pm 1.14\% \pm 24.67$ (22.69) %	$37.45 \pm 1.69\% \pm 24.67$ (22.69) %
NNLO	$47.38 \pm 1.12\% \pm 11.82$ (10.09) %	$47.16 \pm 1.65\% \pm 11.83$ (10.09) %
N ³ LO	$49.04 \pm 1.12\% \pm 3.35$ (3.85) %	$48.81 \pm 1.65\% \pm 3.35$ (3.85) %

Table 10 Same as Table 9, now for Higgs production in vector boson fusion

Higgs production in VBF at 13 TeV		
	C	$C + S^{(9pt)}$
NLO	$3.966 \pm 1.12\% \pm 0.66$ (0.66) %	$3.995 \pm 1.17\% \pm 0.66$ (0.65) %
NNLO	$3.928 \pm 1.12\% \pm 0.37$ (0.42) %	$3.956 \pm 1.17\% \pm 0.37$ (0.41) %
N ³ LO	$3.925 \pm 1.12\% \pm 0.11$ (0.15) %	$3.953 \pm 1.17\% \pm 0.11$ (0.14) %

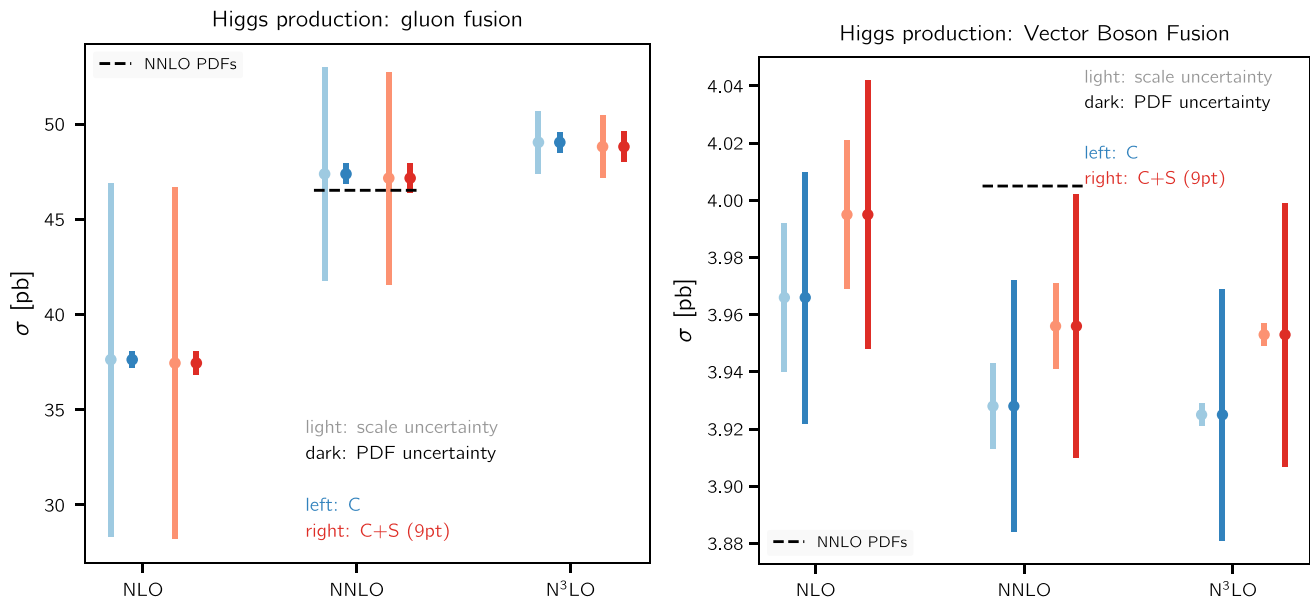


Fig. 21 Graphical representation of the results of Tables 9 and 10. At each perturbative order the pair of uncertainty bands on the left (blue) is computed with PDFs based on the experimental covariance matrix C , while the pair of uncertainty bands on the right (red) with PDFs based on the combined experimental and theoretical covariance matrix $C + S$ (9-

point prescription). The light-shaded bands represent the uncertainty on the hard cross-section (“scale uncertainty”) evaluated using the theory covariance matrix (see text) the dark bands represent the PDF uncertainty. For the NNLO result, we also show the central value obtained using NNLO PDFs as a dashed horizontal line

the VBF cross-section due to the MHOUs is in fact quite significant: its size is comparable to the MHOUs $\sigma_{\mathcal{F}^{\text{th}}}$ on the NLO matrix element, and indeed the shift when going from NLO to NNLO matrix elements, and thus much larger than the corresponding N³LO correction.

We conclude that for VBF the main effect of including the MHOUs in the PDF is a significant shift in the central value of the prediction. Also in this case estimates of the MHOUs on the NNLO PDF were presented in Ref. [84], and it will be interesting to compare them to our approach once NNLO PDFs with MHOUs determined within our approach are available.

A common feature of gluon fusion and vector-boson fusion is that it is only upon inclusion of the MHOUs that the result found using NNLO PDFs is within or at the edge of the PDF uncertainty band of the result found with NLO PDFs.

7.2 Top quark pair production

We now study the impact of the PDF-related MHOUs on the total top-quark pair production cross-section at the LHC for different center-of-mass energies. In Table 11 we collect, using the same format as Table 9, the predictions for the top-quark pair-production cross-sections at $\sqrt{s} = 7, 8$ and 13 TeV obtained using the $\tau_{\text{OP++}}$ code [92] and setting the central scales to $\mu_f = \mu_r = m_t = 172.5$ TeV. The results in the case of 8 and 13 TeV are also displayed in Fig. 22, where

again at NNLO we also show the result obtained using NNLO PDFs.

Just as in the case of Higgs production via gluon-gluon fusion, we find that for top-quark pair production the central values obtained with PDFs with and without MHOUs are rather similar, and well within the one- σ PDF uncertainty. We also observe that the PDF uncertainty at $\sqrt{s} = 7$ and 8 TeV (13 TeV) is about 50% (20%) larger once MHOUs are included in the determination of the PDFs. This is again compatible with the corresponding behavior of the gluon PDF shown in Fig. 16, where it can be observed that, for $x \simeq 0.1$, relevant for top pair production at $\sqrt{s} = 7$ and 8 TeV, the PDF uncertainty is increased in the $C + S^{(9\text{pt})}$ fit compared to the C -only fit, while this increase is less marked for $x \sim 0.3$, relevant for top pair production at $\sqrt{s} = 13$ TeV. Also in this case, the NNLO prediction using NLO PDFs is in better agreement with the that using NNLO PDFs once MHOUs are included, and in fact only in this case the latter is within the PDF error band of the former.

In addition, we note once again that the uncertainty on the hard cross-section $\sigma_{\mathcal{F}^{\text{th}}}$ evaluated using the 9-point covariance matrix is rather similar to that obtained from the symmetrized 7-point envelope. In particular, the 9-point result is slightly larger (smaller) than the 7-point envelope at NNLO (NLO). Finally, from Fig. 22 we notice that for this process the MHOUs on the hard cross-section dominates the PDF uncertainty (with or without MHOUs included), even with NLO PDFs.

Table 11 Same as Table 9, now for top-quark pair-production at $\sqrt{s} = 7, 8$ and 13 TeV

	C	$C + S^{(9pt)}$
$t\bar{t}$ production at 7 TeV		
NLO	$155.42 \pm 1.57\% \pm 12.2$ (13.0) %	$153.94 \pm 2.45\% \pm 12.2$ (13.0) %
NNLO	$174.48 \pm 1.55\% \pm 5.52$ (6.46) %	$172.81 \pm 2.42\% \pm 5.52$ (6.45) %
$t\bar{t}$ production at 8 TeV		
NLO	$222.45 \pm 1.44\% \pm 12.3$ (12.8) %	$220.42 \pm 2.17\% \pm 12.3$ (12.8) %
NNLO	$249.41 \pm 1.43\% \pm 5.43$ (6.28) %	$247.14 \pm 2.14\% \pm 5.43$ (6.27) %
$t\bar{t}$ production at 13 TeV		
NLO	$734.21 \pm 1.11\% \pm 12.4$ (11.8) %	$728.57 \pm 1.38\% \pm 12.3$ (11.8) %
NNLO	$819.43 \pm 1.11\% \pm 5.16$ (5.64) %	$813.17 \pm 1.35\% \pm 5.16$ (5.64) %

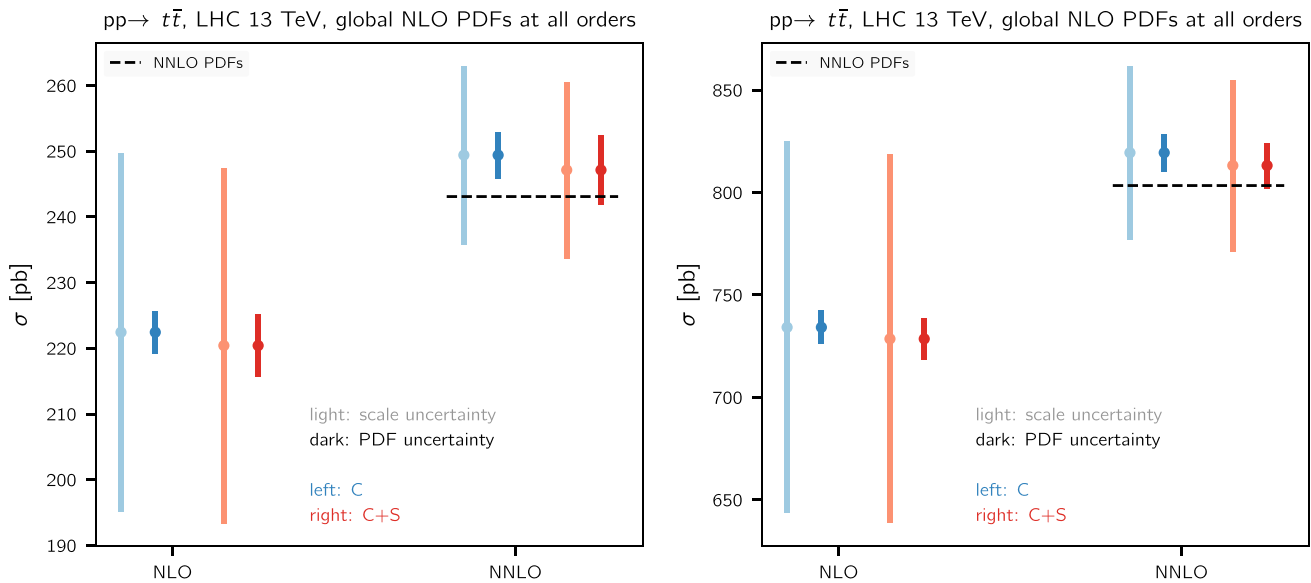


Fig. 22 Same as Fig. 21 for top-quark pair production at 8 and 13 TeV, see also Table 11

7.3 Z and W gauge boson production

We finally turn to gauge boson production, for which we obtain predictions using the computational framework Matrix [93]. In this formalism, all tree-level and one-loop amplitudes are obtained from OpenLoops [94–96]. For these theoretical predictions for inclusive W and Z production cross sections at $\sqrt{s} = 13$ TeV, we adopt realistic kinematic cuts similar to those applied by ATLAS and CMS. The fiducial phase space for the W^\pm cross-section is defined by requiring $p_{l,T} \geq 25$ GeV and $|\eta_l| \leq 2.5$ for the charged lepton transverse momentum and pseudo-rapidity and a missing energy from the neutrino of $p_{\nu,T} \geq 25$ GeV. In the case of Z production, we require $p_{l,T} \geq 25$ GeV and $|\eta_l| \leq 2.5$ for the charged leptons transverse momentum and rapidity and $66 \leq m_{ll} \leq 116$ GeV for the di-lepton invariant mass.

In Table 12 we display a similar comparison as in Table 9 now for W and Z gauge boson production at $\sqrt{s} = 13$ TeV. The corresponding graphical representation of the results is

provided in Fig. 23, again using the same conventions as in Fig. 21 and again also showing the NNLO result with NNLO PDFs.

We find that when including the MHOU the PDF uncertainty is increased by $\simeq 70\%$, 30% and 75% for Z, W^+ , and W^- production respectively. Given that W and Z production at ATLAS and CMS at $\sqrt{s} = 13$ TeV is sensitive to the light sea quarks down to $x \simeq 10^{-3}$, this increase in the PDF uncertainty once MHOU are accounted for is consistent with the corresponding increase reported in the case of the singlet PDF in Fig. 19.

Similarly to Higgs production in vector-boson-fusion, we find that the inclusion of MHOU in the PDF shifts the central value of the prediction, by an amount which is comparable to or larger than the data-driven PDF uncertainty. Yet again, the agreement of the NNLO prediction with NLO PDFs with that which is obtained when NNLO PDFs are used is significantly improved: for Z production within the PDF error band and for W production just barely outside it. We conclude that for

Table 12 Same as Table 9, now for W and Z gauge boson production at $\sqrt{s} = 13$ TeV. The cross-section is given in nb

	C	$C + S^{(9pt)}$
Z production at 13 TeV		
NLO	$0.759 \pm 0.96\% \pm 4.18$ (4.18) %	$0.767 \pm 1.63\% \pm 4.16$ (4.15) %
NNLO	$0.749 \pm 0.97\% \pm 0.94$ (0.63) %	$0.760 \pm 1.64\% \pm 0.93$ (0.66) %
W^- production at 13 TeV		
NLO	$3.534 \pm 0.92\% \pm 4.28$ (4.34) %	$3.560 \pm 1.58\% \pm 4.28$ (4.34) %
NNLO	$3.474 \pm 0.92\% \pm 1.03$ (0.64) %	$3.511 \pm 1.59\% \pm 0.99$ (0.63) %
W^+ production at 13 TeV		
NLO	$4.614 \pm 1.00\% \pm 4.09$ (4.15) %	$4.643 \pm 1.73\% \pm 4.08$ (4.14) %
NNLO	$4.582 \pm 0.99\% \pm 0.88$ (0.58) %	$4.631 \pm 1.72\% \pm 0.87$ (0.62) %

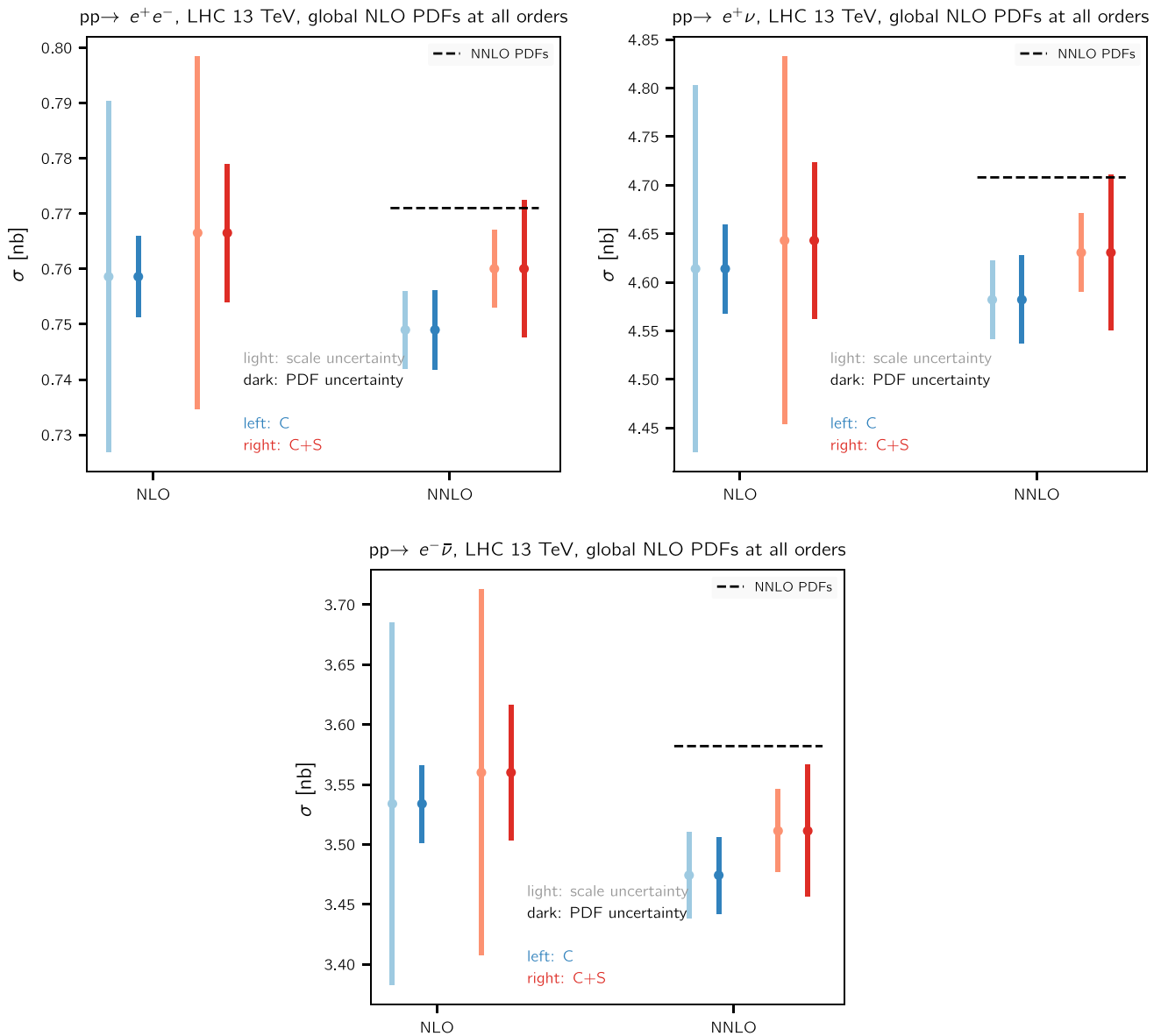


Fig. 23 Same as Fig. 21 for W^\pm and Z gauge boson production at $\sqrt{s} = 13$ TeV, see also Table 12

weak gauge boson production at the LHC the impact of the MHOU associated to the PDFs is twofold: on the one hand an overall increase in the PDF uncertainties that ranges between 30 and 70% depending on the process, and on the other hand a shift in the central values which is comparable to that of the PDF uncertainties of the fit without MHOU.

8 Usage and delivery

As mentioned previously, the PDF sets with MHOU presented in Sect. 6 can be used in essentially the same way as the standard NNPDF sets. In this section we discuss how MHOU included in PDF sets should be combined with those in hard matrix elements, specifically addressing some conceptual issues, and we then provide detailed instructions for their use. We then discuss the delivery of the PDF sets presented in this work, and provide a list of the sets which are being made publicly available by means of the LHAPDF interface.

8.1 Combining MHOU in PDFs and hard matrix elements

As discussed in the introduction, the MHOU on PDFs discussed in this paper arises due to the fact that PDFs are determined using perturbative computations performed at a finite order in the perturbative expansion, and it manifests itself in the fact that PDFs change when varying the order at which they are determined: NLO and NNLO PDFs differ. We have further seen in Sect. 3 that there exist two distinct sources of MHOU in the PDF: that related to MHOU in the computation of the hard cross-sections for those processes used for PDF determination, and that coming from MHOU in the anomalous dimensions. These two sources of MHOU in the PDFs are respectively associated with renormalization and factorization scale variation and can be treated as independent of each other, at least with the definition given here and summarized in Table 1.

On top of this MHOU on the PDF, when computing a factorized prediction for a PDF-dependent hard process not used in the determination of the PDFs, but rather predicted using a given PDF set, there is then the usual MHOU on the hard process itself. This, in turn, just like the MHOU on the PDF, comes from two separate sources: the MHOU on the hard cross-section for the given process, and the MHOU on the evolution of the PDF from the initial scale to the scale of the process. This has been seen explicitly in the phenomenological results presented in Sect. 7, Tables 9, 10, 11, 12 and Figs. 21, 22, 23. So each prediction carries two uncertainties, a PDF uncertainty, which includes the MHOU in the determination of the PDFs (shown as a dark band in the plots, and given as the first uncertainty in the tables), and a “scale” uncertainty in the prediction (shown as a light band in the

plots, and given as the second uncertainty in the tables). Note that in all these plots and tables the PDF uncertainty (when including the theory covariance matrix) includes both the MHOU, and the standard PDF uncertainty due to the uncertainties in the experimental data, while the “scale” uncertainty is just the usual MHOU in the prediction.

In summary, a factorized prediction is affected by two different sources of MHOU: the MHOU in the PDF determination, included in the PDF uncertainty, and then the MHOU in the calculation of the prediction itself. Each in turn receives contributions from both renormalization and factorization scale variation. This immediately raises the question as to whether some of these uncertainties are correlated, and – if this is the case – whether this correlation can be easily accounted for.

A first obvious source of correlation arises when producing a prediction for a process which is among those included for the PDF determination. Examples of this category of processes are top quark pair and gauge boson production, discussed in Sect. 7. They are already included among the processes of Table 2. The MHOU coming from renormalization scale variation is then correlated. Indeed, we know from Fig. 7 that any two predictions for the same physical process are highly correlated, particularly at points which are kinematically close. One might choose to ignore this problem, on the grounds that the main purpose of PDF determinations is to predict new processes, such as Higgs production, or BSM processes: after all, if there is new data for an existing process, it can be included in the PDF fit, and then all correlations would be retained. However this (partial) solution is not available for factorization scale variations, which are used to estimate the MHOU in the evolution between different scales: since the PDFs are universal, these MHOU are correlated across all processes, both within the fit and also in any predictions made subsequently using the PDFs.

The existence of correlations between MHOU in the fitted process and MHOU in the predicted process can be demonstrated rather clearly [15] by noting that PDFs are merely a tool to express a physical observable in terms of other physical observables. In particular QCD predicts the cross-section for one observable in terms of measurements of cross-sections for the same or other observables. Normally to do this one first extracts the PDF from the cross-section data at a range of scales, and then computes cross-sections at some other scale using the extracted PDFs. However in the case of nonsinglet structure functions (discussed in Ref. [15] as a simple paradigm), where the relation between structure function and PDF is straightforward and linear, one can eliminate the nonsinglet PDF altogether: given the structure function at one scale, QCD then predicts the structure function at a different scale, with no reference to any PDF.

Now, it is clear that when expressing one process in terms of another process directly, without any PDFs, there is a sig-

nificant cancellation of MHOUs, specifically that related to perturbative evolution, estimated by means of factorization scale variation. In the example of the nonsinglet structure function, if the structure function at one scale is predicted from its value at some different scale, the factorization scale uncertainty will only depend on the evolution between the two scales involved. Hence there is only one source of MHOUs in the prediction. On the other hand, when using a PDF, there are, as explained above, two sources of MHOUs estimated through factorization scale variation: that from evolving the initial PDF up to the scale of the data used in fit, and that from evolving the initial PDF up to the scale of the prediction. Hence, one has in effect two sources of MHOUs, and if these are assumed to be uncorrelated, and thus added in quadrature, any cancellations are lost and the result will inevitably be an over-estimate of the uncertainty.

If PDFs are to be delivered in the usual way as a universal (*i.e.* process independent) PDF set, much of the detailed information about the specific data, their uncertainties, and the theoretical calculations, and in particular their MHOUs that have gone into determining the PDFs is lost: all that remains are the process independent PDFs. Given only the PDFs, it is clearly impossible to reconstruct the original data, or the MHOUs specific to calculations at each data point, since many different data sets, from different processes, can yield the same PDFs. Consequently, when using PDFs to make a prediction, the correlation between the MHOUs in the prediction and that in the calculations used to determine the PDFs cannot be computed, even in principle: with only the universal PDFs as input, the correlation is no longer available. The loss of this correlation is the inevitable price to pay for PDF universality.

Having understood that neglecting such correlations is inevitable, at least without extending the range of deliverables, one may ask how serious the issue is. The total MHOUs in the determination of the PDF arises from the combination of the MHOUs of theoretical predictions made for a large number of datapoints. The correlations between all these are automatically kept into account by the fitting procedure. Inevitably the fit adjusts to take the MHOUs into account: datapoints associated with large MHOUs (compared to the experimental uncertainty) will be downweighted in the fit, while the effect of data with small MHOUs (compared to their experimental uncertainty) will be relatively unchanged. This rebalancing of the fit is one of the main consequences of including the MHOUs.

Hence, as we saw in our global fit results, the MHOUs have only a relatively small impact on the overall PDF uncertainty: rather by resolving tensions in the fit due to MHOUs in the theoretical predictions, they lead to significant shifts in the central value. However when making a prediction, the uncertainty due to MHOUs in the hard process can be large: in fact in many cases as large or even larger than the total

PDF uncertainty (including its MHOUs). Neglecting the correlation between the MHOUs in the prediction (which might be large) and the MHOUs in the PDFs (which is relatively small) by adding them in quadrature is then likely to be a small effect. Note that this does not mean that the MHOUs on the PDF was negligible in the first place: and indeed as we have seen it may significantly affect the central value of the prediction. Rather, it is its effect on the overall PDF uncertainty which, at least in the data region that we are discussing here, is relatively small. Furthermore, because what is being neglected is a correlation which would lead to a cancellation of uncertainties, it can at worst lead to a small overestimate of uncertainties.

We conclude that while there is clearly a correlation between the MHOUs in the determination of the PDFs and the MHOUs of the hard matrix element of the predicted process, ignoring this correlation, and thus adding the two sources of MHOUs in quadrature, will give a result which is at worst a little conservative. Given all the well known uncertainties intrinsic to the estimate of MHOUs through scale variation, we consider such an approach both pragmatic and realistic.

8.2 Computation of the total uncertainty

Having concluded that uncorrelated combination of the MHOUs on the PDF and on the hard matrix element is justified, we summarize our procedure for computing uncertainties in practice.

To begin with, the PDF uncertainty $\sigma_{\mathcal{F}}^{\text{PDF}}$ associated with a given cross-section \mathcal{F} is evaluated as usual in the NNPDF methodology as the standard deviation over the replica set:

$$\sigma_{\mathcal{F}}^{\text{PDF}} = \left(\frac{1}{N_{\text{rep}} - 1} \sum_{k=1}^{N_{\text{rep}}} (\mathcal{F}[\{q^{(k)}\}] - \langle \mathcal{F}[\{q\}] \rangle)^2 \right)^{1/2}. \quad (8.1)$$

If this prescription is applied to a PDF set with “standard” PDF uncertainty (such as the published NNPDF3.1 [5]) set, the resulting uncertainty only includes the correlated statistical and systematic uncertainties from the data, and the methodological uncertainty intrinsic to any PDF fit. If the PDF sets including MHOUs presented in Sect. 5 of this paper are used instead, the resulting uncertainty obtained from Eq. (8.1) accounts for both the data-driven and MHOUs on the PDF, with all correlations taken into account.

Because the MHOUs on the hard matrix element is treated as uncorrelated to the PDF uncertainty, it can in principle be computed with any prescription preferred by the end-user. A commonly used prescription is 7-point scale variation [18]. Our preferred prescription is instead to use the same methodology as used for the computation of the theory covariance

matrix. In this case, the uncertainty on the cross-section \mathcal{F} is then simply the corresponding diagonal entry of the covariance matrix element, namely

$$\sigma_{\mathcal{F}}^{\text{th}} = \left[S_{\mathcal{F}\mathcal{F}}^{(9\text{pt})} \right]^{1/2}, \tag{8.2}$$

where $S_{\mathcal{F}\mathcal{F}}^{(9\text{pt})}$ is evaluated using our default 9-point prescription defined by Eq. (4.9), with Δ_{ij} computed for $i = j = \mathcal{F}$, i.e. the theory prediction for the given observable. We showed in Sect. 7 that for various standard candles our 9-point theory covariance matrix prescription and the 7-point envelope prescription give very similar results, provided the envelope prescription is symmetrized.

The PDF uncertainty Eq. (8.1) and the uncertainty on the hard matrix element Eq. (8.2) can then be treated as uncorrelated uncertainties. It is then appropriate to combine them in quadrature, so the total uncertainty on the cross-section \mathcal{F} is simply

$$\sigma_{\mathcal{F}}^{\text{tot}} = \left(\left(\sigma_{\mathcal{F}}^{\text{th}} \right)^2 + \left(\sigma_{\mathcal{F}}^{\text{PDF}} \right)^2 \right)^{1/2}. \tag{8.3}$$

We believe that this prescription provides a conservative estimate of the combined MHOU on the predicted cross-section.

Note that when using a χ^2 to assess the quality of the agreement between experimental data and the associated theory predictions for a PDF set which includes MHOU, the MHOU must be always be included in the definition of the χ^2 estimator, ideally (though not necessarily) by means of the theory covariance matrix. This is because, as seen in Sect. 6.2, the inclusion of MHOU modifies the best-fit central value, and thus if the MHOU were not included in the χ^2 , these PDFs would not provide the best fit, and the results might be misleading. Because the theory covariance matrix has been included in the fitting (based on the argument of Sect. 2) as uncorrelated to the experimental covariance matrix, when assessing fit quality it should be regarded as an additional systematic uncertainty, specific to the determination of PDFs from the data, to be added in quadrature to the usual experimental systematics.

8.3 Delivery

The variants of the NNPDF3.1 NLO global sets presented in this work are publicly available in the LHAPDF format [22] from the NNPDF website:

<http://nnpdf.mi.infn.it/nnpdf3-1th/>

In the following, we list the PDF sets that are made available. The NLO sets based on the theory covariance matrix are:

NNPDF31_nlo_as_0118_scalecov_9pt
 NNPDF31_nlo_as_0118_scalecov_7pt

NNPDF31_nlo_as_0118_scalecov_3pt

which correspond to the fits based on Eq. (6.2) in the cases in which the theory covariance matrix S_{ij} has been evaluated with the 9-, 7-, and 3-point prescriptions, respectively.

We have also constructed NLO PDF sets based on scale-varied theories, to be discussed in Appendix B below. These are determined using Eq. (B.1), and they are

NNPDF31_nlo_as_0118_kF_1_kR_1
 NNPDF31_nlo_as_0118_kF_2_kR_2
 NNPDF31_nlo_as_0118_kF_0p5_kR_0p5
 NNPDF31_nlo_as_0118_kF_2_kR_1
 NNPDF31_nlo_as_0118_kF_1_kR_2
 NNPDF31_nlo_as_0118_kF_0p5_kR_1
 NNPDF31_nlo_as_0118_kF_1_kR_0p5

where the naming convention indicates the values of the scale ratios k_f and k_r . Note that the NNPDF31_nlo_as_0118_kF_1_kR_1 set is also the baseline (central scales and experimental covariance matrix only) to be used in the comparisons with the fits based on the theory covariance matrix listed above. Finally, we also provide the set

NNPDF31_nnlo_as_0118_kF_1_kR_1

which corresponds to the NNLO fit with central scales and experimental covariance matrix only, that has been produced for validation purposes.

It is important to bear in mind that the variants of the NNPDF3.1 fits presented in this work are based on a somewhat different dataset to that used in the default NNPDF3.1 analysis. Therefore, when using these sets it is important to be consistent: for example by comparing fits with and without MHOU that are based on a common input dataset.

In addition to the sets listed above, the other PDF sets presented in this paper, such as the DIS-only fits based on scale-varied calculations and on the theory covariance matrix, are available from the authors upon request.

9 Summary and outlook

In this work we have presented the first PDF determination that includes MHOU as part of the PDF uncertainty. This is in principle required for consistency, given that MHOU are routinely part of the theoretical predictions for hadron collider processes, and likely to become a requirement for precision collider phenomenology as other sources of uncertainties decrease.

The bulk of our work amounted to establishing a general language and formalism for the inclusion of MHOU when multiple processes are considered at once in the global PDF fit, constructing prescriptions for estimating these MHOU by means of scale variation, and for validating them in cases in

which the higher order corrections are known. The formalism presented here is sufficiently flexible that it can also be applied to different sources of theoretical uncertainty, such as nuclear corrections or higher twists, and could also be used in conjunction with alternative ways of estimating MHO, such as for example the Cacciari–Houdeau method.

The validation studies presented here suggest however that the conventional scale variation method to estimate the MHO works remarkably well. Indeed, when coupled to the theory covariance matrix formalism that we introduced, this method turns out to be free of the instabilities that plague envelope techniques, and it leads to results which appear to be reasonably stable and thus insensitive to the arbitrary choices that are inherent to its implementation. The reason for these properties is essentially that, within a covariance matrix approach, possible directions which do not correspond to actual MHO have no impact on the fitting.

Our results however also suggest that even more realistic estimates of MHO might be obtained through more complex patterns of scale variation than those considered here. Specifically, a more refined treatment of factorization scale variation is likely to be advantageous, in which independent variation is performed for each eigenvalue of the anomalous dimension matrix. Also, it might be advantageous to vary independently the renormalization scales in different partonic sub-channels. Indeed, we have observed from the validation of our estimate of MHO, while always reasonably successful for the datasets considered here, deteriorates as the size of the dataset increases, which suggests that more complex structures might be required. Here we have performed a first investigation, and the exploration of these more complex patterns of scale variation will be left for future work.

On the phenomenological side, our results show that at least at NLO the main effect of the inclusion of MHO in PDF determination is to improve the accuracy of the result, while not significantly reducing its precision. Indeed, whenever experimental information is abundant, in particular for a global dataset, we have found that the total PDF uncertainty is only moderately affected by the inclusion of MHO – in fact, for the datapoints included in PDF determination it even decreases – but the central value moves closer to the true result. Moreover, the fit quality improves, thereby showing that the main effect of the inclusion of MHO is in reducing tensions between datasets due to imperfections in their theoretical description.

The most interesting future phenomenological development will be of course the extension of our methodology to the determination of MHO in a state-of-the-art global NNLO PDF set. It will be interesting to assess to what extent the behaviour observed at NLO persists there. More generally, the inclusion of MHO at NNLO is expected to lead to the most precise and accurate PDF sets that can be deter-

mined with currently available theoretical and experimental information.

Acknowledgements We are grateful to V. Bertone and N. P. Hartland for collaboration in the early stages of this work. We would also like to acknowledge useful discussions with J. Bendavid, M. Bonvini, F. Caola, M. Duehrssen, L. Harland-Lang, P.F. Monni, G. Salam, and R. Thorne on the topic of theory uncertainties and PDFs. R.D.B. is supported by the UK Science and Technology Facility Council through grant ST/P000630/1. S.F. is supported by the European Research Council under the European Union’s Horizon 2020 research and innovation Programme (grant agreement n.740006). T.G. is supported by The Scottish Funding Council, grant H14027. Z.K. is supported by the European Research Council Consolidator Grant “NNLOforLHC2”. E.R.N. is supported by the European Commission through the Marie Skłodowska-Curie Action ParDHonS FFs.TMDs (grant number 752748). R.L.P. and M.W. are supported by the STFC grant ST/R504737/1. J.R. is supported by the European Research Council Starting Grant “PDF4BSM” and by the Netherlands Organization for Scientific Research (NWO). L.R. is supported by the European Research Council Starting Grant “REINVENT” (grant number 714788). M.U. is partially supported by the STFC grant ST/L000385/1 and funded by the Royal Society grants DH150088 and RGF/EA/180148. C.V. is supported by the STFC grant ST/R504671/1.

Data Availability Statement This manuscript has associated data in a data repository. [Authors’ comment: Parton distributions with theory uncertainties discussed here are publicly available in LHAPDF format from the NNPDF website: <http://nnpdf.mi.infn.it/nnpdf3-1th/>, as discussed in Sect. 8.3 of the paper.]

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. Funded by SCOAP³.

A Diagonalisation of the theory covariance matrix

To carry out the validation described in Sect. 5.3 and thus compute the angles θ defined Eq. (5.7), we must first diagonalise \widehat{S}_{ij} for the various prescriptions. In this appendix we provide details concerning this diagonalization process.

The diagonalisation of the theory covariance matrix \widehat{S}_{ij} is difficult due to the very large number of zero eigenvalues. To get around this problem we first project \widehat{S}_{ij} onto S , and then perform the diagonalization in this subspace (in which all eigenvalues are positive, by construction). The projection is easily achieved, since S is spanned by the vectors $\{\Delta_i(\kappa_f, \kappa_{r_a}) : \kappa_f, \kappa_{r_a} \in V_m\}$ defined in Eq. (4.1), used to construct S_{ij} in Eq. (4.2). Similarly, \widehat{S}_{ij} is constructed from normalized vectors $\{\widehat{\Delta}_i(\kappa_f, \kappa_{r_a}) : \kappa_f, \kappa_{r_a} \in V_m\}$, where $\widehat{\Delta}_i = \Delta_i/T_i^{\text{NLO}}$. However these vectors are not all linearly independent, and a linearly independent set is best constructed on a case by case basis. This construction also gives us N_{sub} , the dimension of S , for each of the prescriptions.

- 5-point** When there are p processes, V_4 has $2 + 2^p$ distinct elements $\{(\pm; 0, 0, 0, \dots), (0; \pm, \pm, \pm, \dots)\}$, so there are $2 + 2^p$ different vectors, $\widehat{\Delta}_i^{\pm 0}$ and $\widehat{\Delta}_i^{0\pm}$: if $i_n \in \pi_n$, with $n = 1, \dots, p$ labeling the different processes, then

$$\widehat{\Delta}_i^{+0} \equiv (\widehat{\Delta}_{i_1}^{+0}, \widehat{\Delta}_{i_2}^{+0}, \dots, \widehat{\Delta}_{i_p}^{+0}), \tag{A.1}$$

and similarly for $\widehat{\Delta}_i^{-0}$, while

$$\widehat{\Delta}_i^{0\pm} \equiv (\widehat{\Delta}_{i_1}^{0\pm}, \widehat{\Delta}_{i_2}^{0\pm}, \dots, \widehat{\Delta}_{i_p}^{0\pm}), \tag{A.2}$$

where for each process the renormalization scale is varied independently. Not all of the second class of vectors, $\widehat{\Delta}_i^{0\pm}$, are linearly independent: for example, when $p = 2$, there exists the linear relation

$$(\widehat{\Delta}_{i_1}^{0+}, \widehat{\Delta}_{i_2}^{0+}) - (\widehat{\Delta}_{i_1}^{0-}, \widehat{\Delta}_{i_2}^{0-}) = (\widehat{\Delta}_{i_1}^{0+}, \widehat{\Delta}_{i_2}^{0-}) - (\widehat{\Delta}_{i_1}^{0-}, \widehat{\Delta}_{i_2}^{0+}), \tag{A.3}$$

so there are five rather than six linearly independent vectors. The number of linear relations for general p can be deduced inductively: if we have in total n_p linearly independent vectors v_i^a for p processes, where $a = 1, \dots, p$, $i = (i_1, \dots, i_p)$, then when there are $p + 1$ processes, we have $2n_p$ distinct vectors $(v_i^a, \widehat{\Delta}_{i_{p+1}}^{0\pm})$, but with linear relations

$$(v_i^a, \widehat{\Delta}_{i_{p+1}}^{0+}) - (v_i^{a+1}, \widehat{\Delta}_{i_{p+1}}^{0+}) = (v_i^a, \widehat{\Delta}_{i_{p+1}}^{0-}) - (v_i^{a+1}, \widehat{\Delta}_{i_{p+1}}^{0-}). \tag{A.4}$$

There are $n_p(n_p - 1)$ of these relations, but of these only $n_p - 1$ are linearly independent. So $n_{p+1} = 2n_p - (n_p - 1) = n_p + 1$. For $p = 2$, $n_p = 5$, so in general we must have $n_p = p + 3$, i.e. the dimension of the subspace S is $N_{\text{sub}} = p + 3$ for the 5-point prescription.

- 5-point** When there are p processes, \overline{V}_4 has 2^{p+1} distinct elements $\{(\pm; \pm, \pm, \pm, \dots)\}$, so there are 2^{p+1} different vectors $\widehat{\Delta}_i^{+\pm}$ and $\widehat{\Delta}_i^{-\pm}$, where

$$\widehat{\Delta}_i^{+\pm} \equiv (\widehat{\Delta}_{i_1}^{+\pm}, \widehat{\Delta}_{i_2}^{+\pm}, \dots, \widehat{\Delta}_{i_p}^{+\pm}), \tag{A.5}$$

and similarly for $\widehat{\Delta}_i^{-\pm}$: the scale variation of $\ln k_f$ is fully correlated across all processes, but all the renormalization scales for the processes π_r are varied independently from it and the others. Again not all of these 2^{p+1} vectors are linearly independent, but it can be shown using a similar inductive argument as for 5-point that the number of independent vectors is $N_{\text{sub}} = 2p + 2$.

- 9-point** When there are p processes, V_8 has $2^p + 2 \cdot 3^p$ distinct elements $V_8 = \{(0; \pm, \pm, \pm, \dots), (\pm; \oplus, \oplus, \oplus, \dots)\}$, so the corresponding vectors are $\widehat{\Delta}_i^{0\pm}$, $\widehat{\Delta}_i^{+\pm}$, $\widehat{\Delta}_i^{+0}$, $\widehat{\Delta}_i^{-\pm}$ and $\widehat{\Delta}_i^{-0}$ where

$$\widehat{\Delta}_i^{0\pm} \equiv (\widehat{\Delta}_{i_1}^{0\pm}, \widehat{\Delta}_{i_2}^{0\pm}, \dots, \widehat{\Delta}_{i_p}^{0\pm}), \tag{A.6}$$

while

$$\begin{aligned} \widehat{\Delta}_i^{+\pm} &\equiv (\widehat{\Delta}_{i_1}^{+\pm}, \widehat{\Delta}_{i_2}^{+\pm}, \dots, \widehat{\Delta}_{i_p}^{+\pm}), \\ \widehat{\Delta}_i^{+0} &\equiv (\widehat{\Delta}_{i_1}^{+0}, \widehat{\Delta}_{i_2}^{+0}, \dots, \widehat{\Delta}_{i_p}^{+0}), \end{aligned} \tag{A.7}$$

and similarly for $\widehat{\Delta}_i^{-\pm}$ and $\widehat{\Delta}_i^{-0}$. Again there are many linear relations between these vectors: if for p processes there are n_p^0 independent vectors of class $\widehat{\Delta}_i^{0\pm}$, and n_p^\pm independent vectors of classes $\widehat{\Delta}_i^{+\pm}$, $\widehat{\Delta}_i^{+0}$ and $\widehat{\Delta}_i^{-\pm}$, $\widehat{\Delta}_i^{-0}$ then while for $p + 1$ processes $n_{p+1}^0 = 2n_p^0 - (n_p^0 - 1)$ (i.e. $n_p^0 - 1$ linearly independent linear relations), $n_{p+1}^\pm = 3n_p^\pm - 2(n_p^\pm - 1)$ (i.e. $2(n_p^\pm - 1)$ linearly independent linear relations). So we now find $n_p^0 = p + 1$, $n_p^\pm = 2p + 1$, and $N_{\text{sub}} = n_p^0 + n_p^+ + n_p^- = 5p + 3$.

- 3-point** When there are p processes, V_2 has 2^p distinct elements $V_2 = \{(\pm, \pm, \pm, \dots)\}$, so the independent vectors are $\widehat{\Delta}_i^{++}$ and $\widehat{\Delta}_i^{--}$, where

$$\widehat{\Delta}_i^{++} \equiv (\widehat{\Delta}_{i_1}^{++}, \widehat{\Delta}_{i_2}^{++}, \dots, \widehat{\Delta}_{i_p}^{++}), \tag{A.8}$$

and similarly for $\widehat{\Delta}_i^{--}$. The number of linearly independent vectors is $N_{\text{sub}} = p + 1$.

- 7-point** Since $V_6 = V_4 \oplus V_2$, the $2 + 2^{p+1}$ independent vectors are simply those for 5-point and those for 3-point together, i.e. $\widehat{\Delta}_i^{\pm 0}$, $\widehat{\Delta}_i^{0\pm}$, $\widehat{\Delta}_i^{++}$ and $\widehat{\Delta}_i^{--}$. The number of linearly independent vectors is thus $N_{\text{sub}} = 2p + 4$.

Once we have a set of linearly independent vectors spanning the space, we can use them to construct an orthonormal basis v_i^a , such that $\sum_i v_i^a v_i^b = \delta^{ab}$. Then the projection of \widehat{S}_{ij} into the subspace S will be given by

$$\widehat{S}^{ab} = \sum_{i,j} v_i^a v_j^b \widehat{S}_{ij}, \tag{A.9}$$

and the diagonalization of \widehat{S}^{ab} gives the positive eigenvalues $\lambda_\alpha = (s^\alpha)^2$. The eigenvectors e_i^α can then be constructed from the basis vectors v_i^a : if e_a^α is the eigenvector of s^{ab} corresponding to eigenvalue λ_α , then $e_i^\alpha = \sum_a v_i^a e_a^\alpha$.

Table 13 List of PDF sets with different choices of the renormalization μ_f and factorization μ_f . For each set, we indicate its label, the input dataset, the perturbative order, and the ratios $k_r = \mu_r/Q$ and $k_f = \mu_f/Q$ to the central scale Q

Label	Dataset	Order	$k_f = \mu_r/Q$	$k_r = \mu_f/Q$
NNPDF31_dis_nlo_as_0118_kF_1_kR_1	DIS	NLO	1	1
NNPDF31_dis_nlo_as_0118_kF_2_kR_2	DIS	NLO	2	2
NNPDF31_dis_nlo_as_0118_kF_0p5_kR_0p5	DIS	NLO	$\frac{1}{2}$	$\frac{1}{2}$
NNPDF31_dis_nlo_as_0118_kF_2_kR_1	DIS	NLO	2	1
NNPDF31_dis_nlo_as_0118_kF_1_kR_2	DIS	NLO	1	2
NNPDF31_dis_nlo_as_0118_kF_0p5_kR_1	DIS	NLO	$\frac{1}{2}$	1
NNPDF31_dis_nlo_as_0118_kF_1_kR_0p5	DIS	NLO	1	$\frac{1}{2}$
NNPDF31_dis_nlo_as_0118_kF_2_kR_0p5	DIS	NLO	2	$\frac{1}{2}$
NNPDF31_dis_nlo_as_0118_kF_0p5_kR_2	DIS	NLO	$\frac{1}{2}$	2
NNPDF31_nlo_as_0118_kF_1_kR_1	Global	NLO	1	1
NNPDF31_nlo_as_0118_kF_2_kR_2	Global	NLO	2	2
NNPDF31_nlo_as_0118_kF_0p5_kR_0p5	Global	NLO	$\frac{1}{2}$	$\frac{1}{2}$
NNPDF31_nlo_as_0118_kF_2_kR_1	Global	NLO	2	1
NNPDF31_nlo_as_0118_kF_1_kR_2	Global	NLO	1	2
NNPDF31_nlo_as_0118_kF_0p5_kR_1	Global	NLO	$\frac{1}{2}$	1
NNPDF31_nlo_as_0118_kF_1_kR_0p5	Global	NLO	1	$\frac{1}{2}$
NNPDF31_nlo_as_0118_kF_2_kR_0p5	Global	NLO	2	$\frac{1}{2}$
NNPDF31_nlo_as_0118_kF_0p5_kR_2	Global	NLO	$\frac{1}{2}$	2
NNPDF31_dis_nnlo_as_0118_kF_1_kR_1	DIS	NNLO	1	1
NNPDF31_nnlo_as_0118_kF_1_kR_1	Global	NNLO	1	1

Table 14 The values of χ^2/N_{dat} for the DIS-only PDF sets based on scale-varied theories. We display the values of the χ^2/N_{dat} for the nine combination of scale variations, (k_f, k_r) , listed in Table 13. For each dataset, we also indicate the number of data points after cuts, see also Table 2

Dataset	N_{dat}	χ^2/N_{dat} NNP3.1 DIS-only NLO									NNLO
		(1, 1)	(2, 2)	$(\frac{1}{2}, \frac{1}{2})$	(2, 1)	(1, 2)	$(1, \frac{1}{2})$	$(\frac{1}{2}, 1)$	$(2, \frac{1}{2})$	$(\frac{1}{2}, 2)$	
NMC	134	1.259	1.274	1.272	1.263	1.268	1.247	1.286	1.277	1.301	1.244
SLAC	12	0.908	1.144	0.941	0.881	0.922	0.709	0.7651	0.668	0.745	0.794
BCDMS	530	1.046	1.047	1.057	1.046	1.040	1.050	1.059	1.093	1.053	1.046
CHORUS	430	0.982	1.024	1.069	1.031	1.018	1.030	1.024	1.055	1.038	1.093
NuTeV	41	0.628	0.564	0.712	0.711	0.642	0.693	0.634	0.725	0.736	0.892
HERA incl	967	1.097	1.126	1.136	1.167	1.091	1.152	1.131	1.357	1.122	1.103
HERA F_2^c	31	1.047	0.983	1.153	1.058	1.012	1.257	2.122	1.868	2.137	1.055
Total	2145	1.061	1.083	1.103	1.104	1.064	1.098	1.104	1.218	1.104	1.089

B PDF sets with different scale choices

The approach that we have pursued in this work for the determination of MHOUs in PDFs is based on the idea of utilising scale variation of the theory prediction to produce an estimate of the MHOU, and then using this information to construct a theory covariance matrix to be used in PDF fitting. Results from this approach have been presented in Sect. 6. An alternative, and perhaps more naive, option would be that of simply performing PDF fits in which different choices are made for the factorization and renormalization scales used in the fit.

One may then take the envelope of the resulting fits, for some set of scale choices, as an estimate of the MHOU.

In this appendix, we will construct PDF sets based on varying the renormalization and factorization scales in the PDF fit. These PDFs are obtained from the minimization of the usual figure of merit

$$\chi^{2(s)} = \frac{1}{N_{\text{dat}}} \sum_{i,j=1}^{N_{\text{dat}}} (D_i - T_i^{(s)})(C_0^{-1})_{ij}(D_j - T_j^{(s)}), \quad (\text{B.1})$$

where $T_i^{(s)} = T_i(\kappa_f^{(s)}, \kappa_r^{(s)})$, s labels the scale choices used for the determination of each PDF set, and C_0 is the

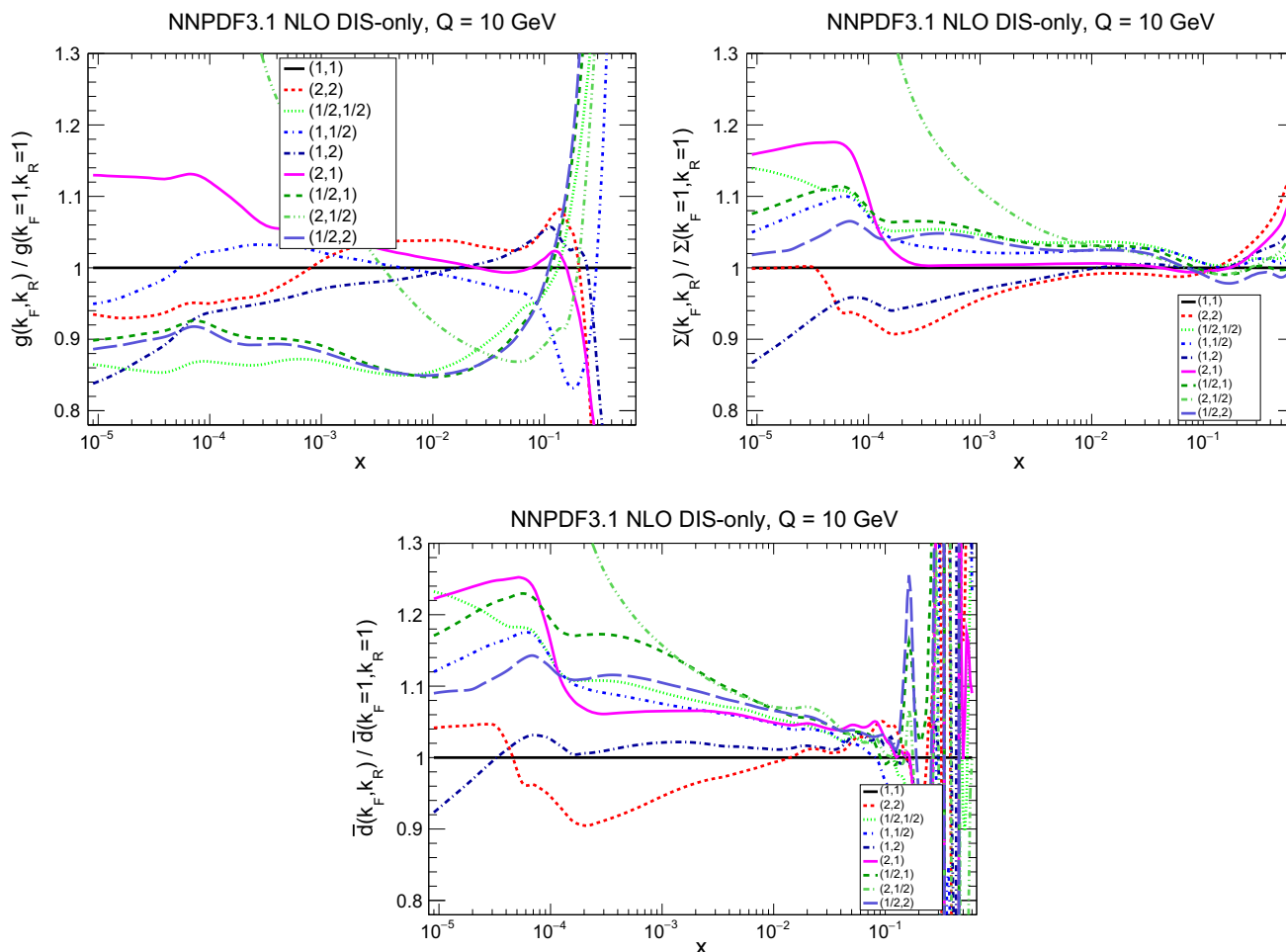


Fig. 24 Comparison of the central values of the DIS-only PDF with different values of (k_f, k_r) . All results are normalized to the baseline $(k_f, k_r) = (1, 1)$. The gluon, the total quark singlet and down antiquark (top to bottom) are shown at $Q = 10$ GeV

experimental covariance matrix evaluated using the usual t_0 -prescription. We will then study the resulting PDFs.

As we shall see, whereas this approach provides an independent way of assessing the dependence of PDFs on scale choice, it does not provide a stable way of estimating MHOUs. Although these PDF sets do not appear to be advantageous for MHOU, we present them here because they are nevertheless interesting for their own sake. This is especially true in view of the fact that PDF sets based on systematic scale variation of the underlying theory have never been presented before.

Based on the experimental and theoretical settings described in Sect. 5.1, we have produced a number of PDF sets with different choices for k_r and k_f , input dataset, and perturbative order, which are summarized in Table 13. The PDF sets corresponding to the central scale choices are the same as discussed in Sect. 6. In the same way as in Sect. 6, we determine PDFs at NLO both from a DIS-only dataset and

a global dataset, with NNLO PDFs determined with central scale choices as a reference.

In all of these PDF determinations, the factorization and renormalization scale are varied in a fully correlated way between all datasets. So for example, if $k_r = 2$, then the renormalization scale is taken to be twice its default value for all processes. This immediately exposes a defect in this method: in principle, the MHOU in the hard cross-sections of different processes are uncorrelated. However, uncorrelated variations of k_r across the five processes that we consider would require 3^6 fits of each type, or $\mathcal{O}(70,000)$ replicas, which is of course impractical.

B.1 DIS-only PDFs

In Table 14 we collect the values of χ^2/N_{dat} for the PDFs determined from a DIS-only dataset with various choices of renormalization and factorization scale. We note that the central scale choice leads to the lowest value of χ^2 .

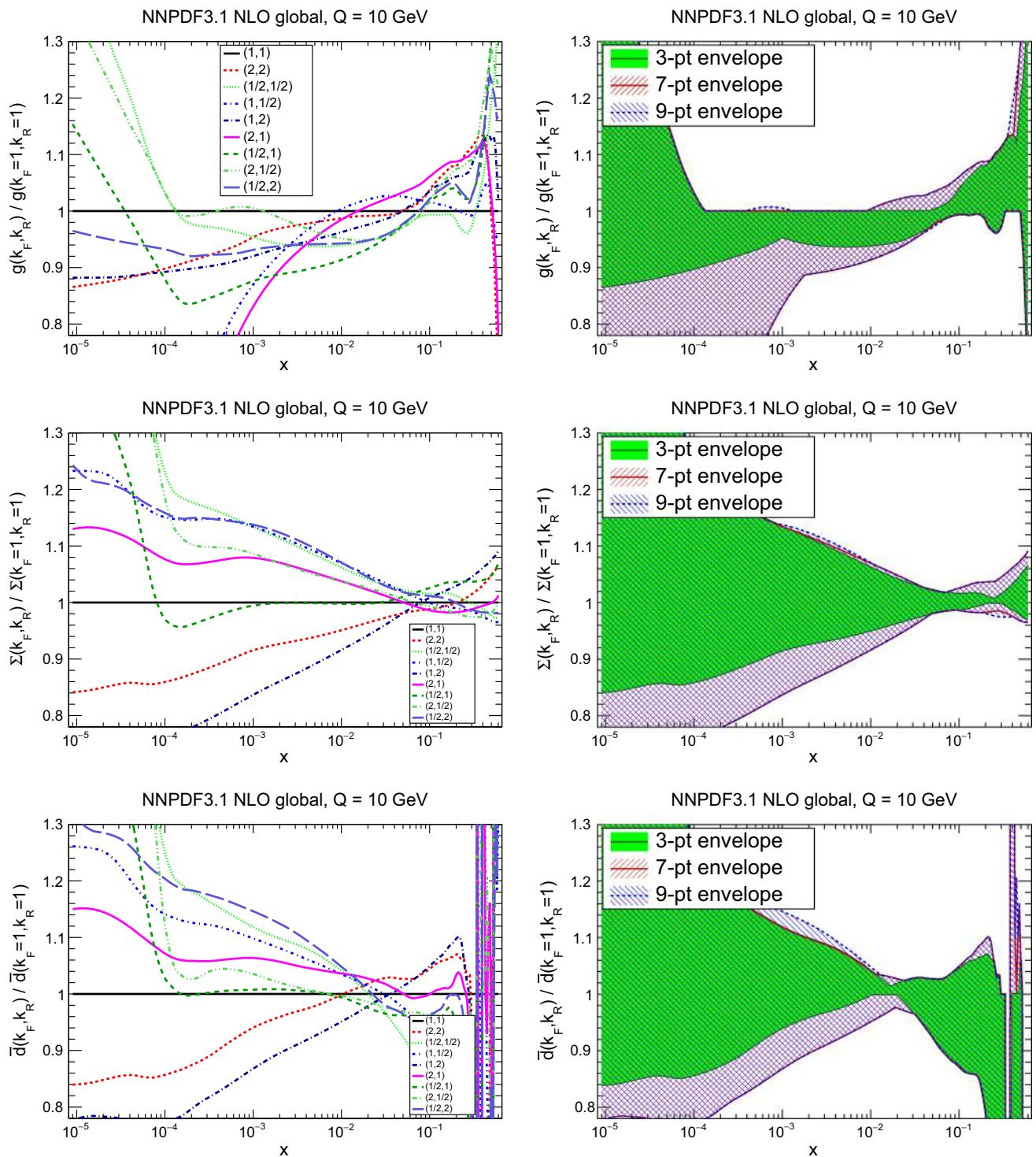


Fig. 25 Left: same as Fig. 24 (left) for global PDF sets. Right: Envelope of scale variations computed using various prescriptions (see text)

The scale choice $(2, \frac{1}{2})$ leads to a much larger χ^2 than any other choice. This choice, which involves a large scale ratio, is typically omitted when estimating scale uncertainties from an envelope prescription. Note however that the reciprocal choice $(\frac{1}{2}, 2)$, which is also usually discarded for the same reason, leads to a χ^2 which is not particularly large.

Variation of χ^2 values with the scales is more marked for HERA experiments than for fixed target, consistent with the observation that scale variation is larger in the small x region covered by the HERA data.

We next assess the impact of scale variation on the PDFs. The impact on PDF uncertainties turns out to be moderate,

Table 15 Same as Table 14 for global PDFs

Process	N_{dat}	χ^2/N_{dat} NNPDF3.1 global NLO									NNLO
		(1, 1)	(2, 2)	$(\frac{1}{2}, \frac{1}{2})$	(2, 1)	(1, 2)	$(1, \frac{1}{2})$	$(\frac{1}{2}, 1)$	$(2, \frac{1}{2})$	$(\frac{1}{2}, 2)$	
DIS NC	1593	1.088	1.182	1.209	1.191	1.103	1.144	1.188	1.394	1.200	1.084
DIS CC	552	1.012	1.014	1.045	1.018	1.020	1.042	1.089	1.065	1.079	1.079
DY	484	1.486	1.500	1.437	1.439	1.461	1.347	1.441	1.772	1.664	1.231
JETS	164	0.907	0.875	0.947	0.911	0.874	0.914	0.938	1.023	0.945	0.950
TOP	26	1.260	2.542	1.390	1.143	2.352	1.277	1.121	1.493	1.756	1.068
Total	2819	1.139	1.200	1.256	1.214	1.153	1.190	1.240	1.405	1.253	1.105

and thus we concentrate on central values. In Fig. 24 we compare the central values of the DIS-only PDFs obtained with the values of (k_f, k_r) listed in Table 13, normalized to the $(k_f, k_r) = (1, 1)$ baseline. The gluon, total quark singlet and down antiquark are shown at $Q = 10$ GeV. The behaviour for other quark flavors is similar.

Scale variation for the gluon turns out to be reasonably asymmetric, with all scale choices leading to a gluon which is below the central scale choice for $x \lesssim 10^{-2}$. The scale choice $(k_f, k_r) = (2, \frac{1}{2})$ which, as already noted, leads to a much worse χ^2 value appears to lead to a rather unstable PDF. In the singlet case the spread of scale variations about the central choice is more symmetric. Whereas for the gluon the spread is considerable for all x values, for the quark singlet the spread becomes quite small at large $x \gtrsim 0.1$. The behavior of the down antiquark is similar to that of the singlet.

In general, the sensitivity of results to scale variation appears to be directly linked to the data-driven PDF uncertainty, with a much wider spread observed whenever the information coming from data is reduced and PDF uncertainties are large.

B.2 Global PDFs

We now turn to PDF fits based on the global dataset. As we will show, the use of a global dataset reduces not only the PDF uncertainties but also the relative impact of varying the scales, compared to the DIS-only fits. This is consistent with previous results [97] showing that the perturbative stability of PDFs improves as the size of the dataset used for their determination grows.

In Table 14 we collect the values of χ^2/N_{dat} for the NNPDF3.1 NLO global PDF sets for all scale choices, with the NNLO value for the central scale choice also shown. As in the DIS-only case, the best fit is found for the central scale choice, with all others leading to a worse χ^2 value, and the scale choice $(2, \frac{1}{2})$ leading to much worse fit quality. All scale choices at NLO give a worse fit than the NNLO fit, due to the fact that NNLO corrections are needed for a good description of several high-precision LHC data [5].

The corresponding PDFs are shown in Fig. 25, where we also display various envelopes of scale variations which will be discussed in Sect. B.3 below. The general features are similar to those of the DIS-only fits shown in Fig. 24. In the case of the gluon, varying the scales with respect to the central choice leads in general to a suppression for $x \lesssim 10^{-2}$ and an enhancement for larger values. For the singlet and down quark PDFs, one can observe a large spread as the scales are varied at small- x , which is approximately symmetric, a point of minimum sensitivity around $x \simeq 0.1$, and then a further increase of the spread of central values at large- x especially for the poorly constrained down antiquark. The scale combination $(k_f, k_r) = (2, \frac{1}{2})$ also leads to a large distortion of the PDF central values in this case, as in the DIS-only fits.

Table 16 The values of $\Delta_{\chi^2}^{\pm}$ Eq. (B.2) for DIS-only and global PDFs. The maximum and minimum of the χ^2 are evaluated for the PDF sets listed in Table (13)

Dataset	N_{dat}	$[\Delta_{\chi^2}^-, \Delta_{\chi^2}^+]$ DIS-only NLO
NMC	134	[-2, +6]
SLAC	12	[-3, +3]
BCDMS	530	[-3, +25]
CHORUS	430	[0, +38]
NuTeV	41	[-3, +4]
HERA incl	967	[-6, +251]
HERA F_2^c	31	[-3, +34]
Total	2145	[0, +337]
Process	N_{dat}	$[\Delta_{\chi^2}^-, \Delta_{\chi^2}^+]$ global NLO
DIS NC	1593	[0, +487]
DIS CC	552	[0, +42]
DY	484	[-67, +138]
JETS	164	[-5, +19]
TOP	26	[-4, +33]
Total	2819	[0, +750]

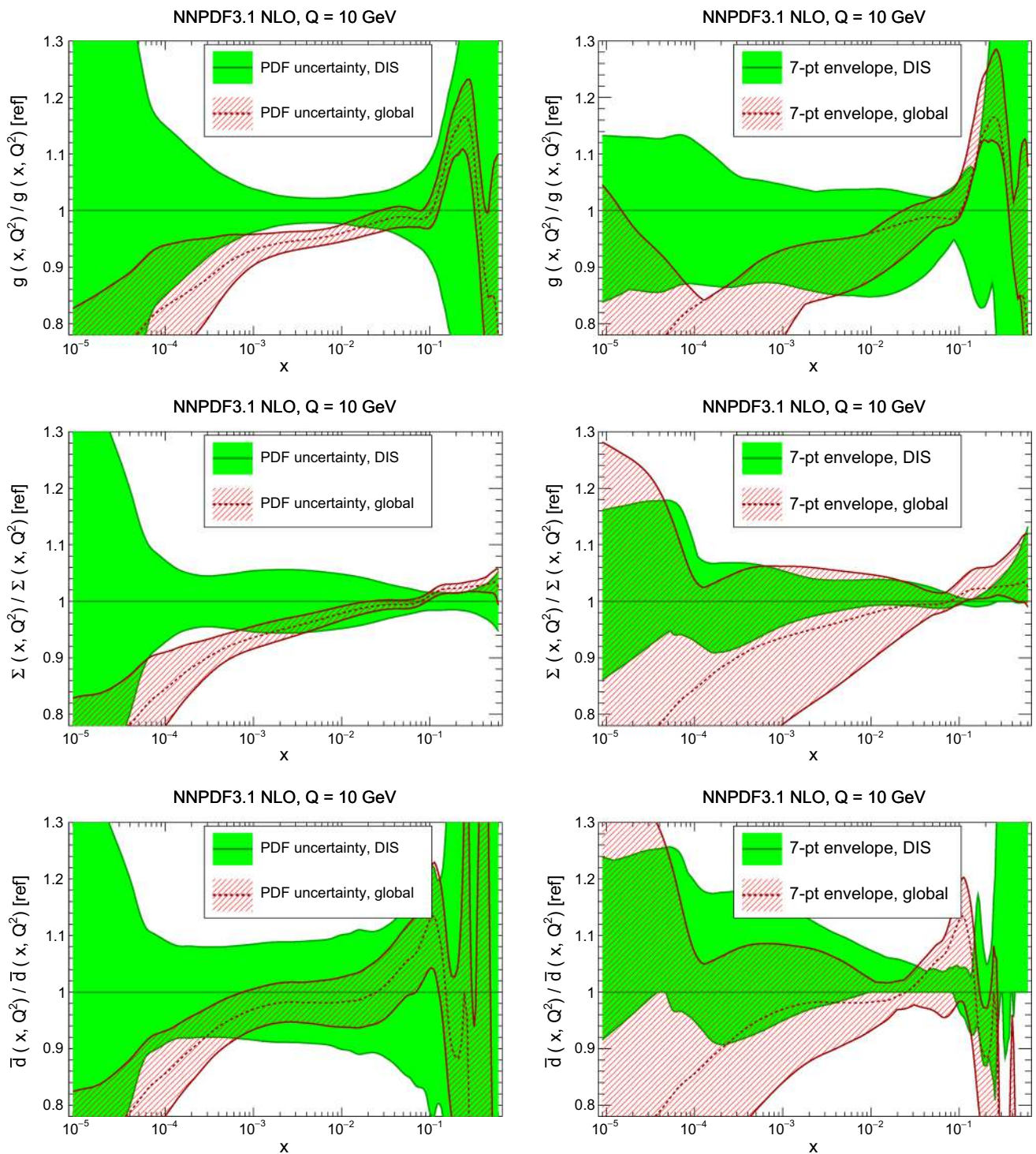


Fig. 26 Comparison between PDF uncertainties (left) and 7-point envelopes of scale variations (right) for the DIS and global PDFs. The gluon, quark singlet and antidown at $Q = 10$ GeV are shown

B.3 The envelope prescription for MHOU

Now we would like to assess the possibility of using PDF sets obtained with different choices of renormalization and factorization scales as a means to estimate MHOU. We first

systematically compare fit quality as a function of the scale choice. In order to facilitate this comparison we evaluate

$$\Delta_{\chi^2}^+ \equiv (\chi_{\max}^2 - \chi_{\text{central}}^2); \quad \Delta_{\chi^2}^- \equiv (\chi_{\min}^2 - \chi_{\text{central}}^2), \quad (\text{B.2})$$

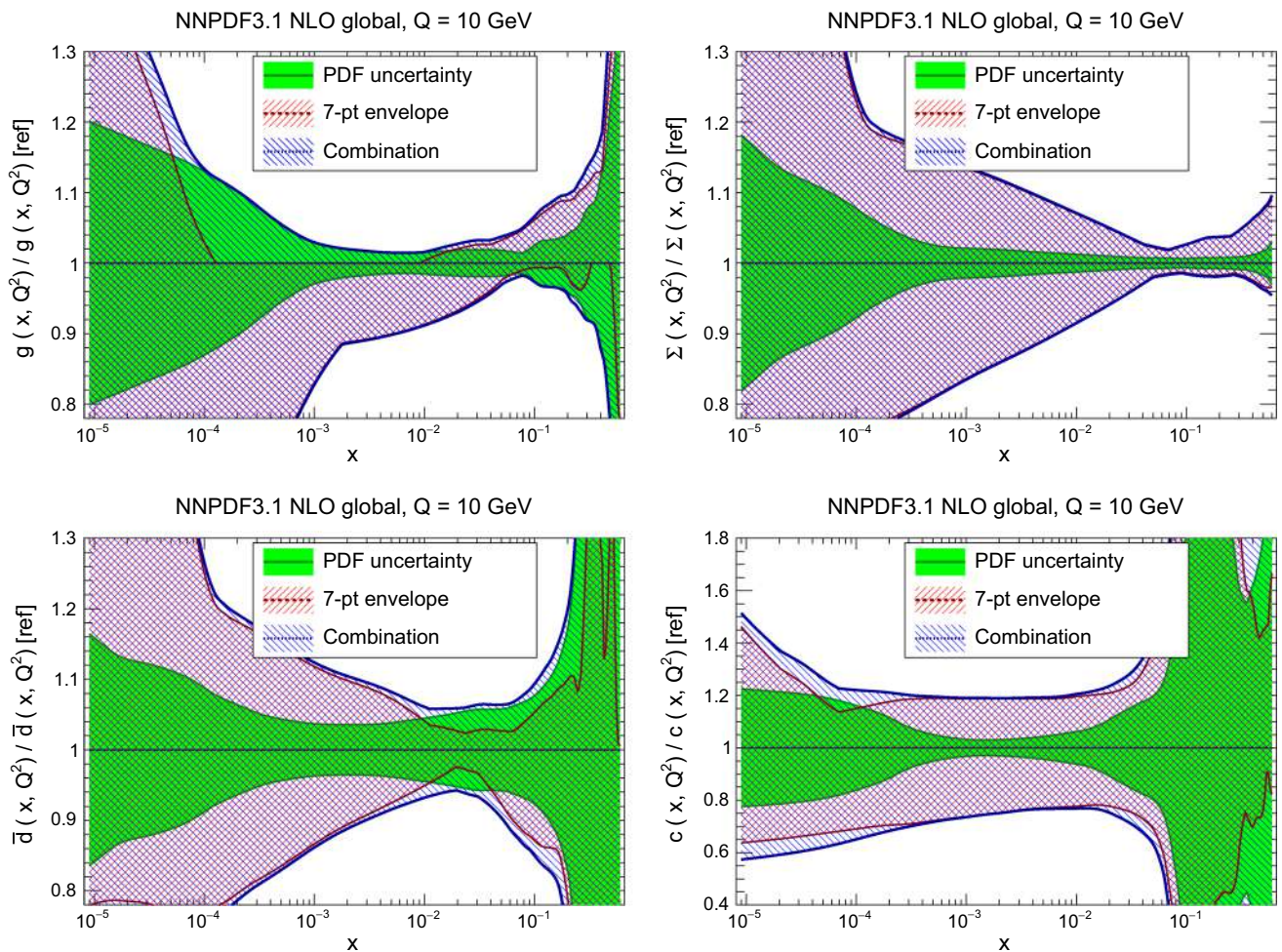


Fig. 27 The PDF uncertainties Eq. (B.3): the PDF uncertainty σ^{PDF} computed using standard NNPDF methodology, the 7-point envelope estimate of the MHO uncertainty $\sigma^{\text{mho},\pm}$ and the total combined uncer-

tainty $\sigma^{\text{tot},\pm}$. Results are shown at $Q = 10$ GeV for the gluon (top left), quark singlet (top right), down antiquark (bottom left), and charm PDFs (bottom right) normalized to the central value

where χ^2_{max} (χ^2_{min}) denotes the largest (smallest) value of the χ^2 , for either the total dataset or for individual experiments, among all the NLO fits based on scale-varied theories listed in Tables 14 and 15, and χ^2_{central} indicates the values from the baseline fit with central scales.

The values of these quantities for both the DIS-only fits and the global fits are collected in Table 16, both for individual experiments and for the total dataset. From this comparison, see that for DIS-only PDFs the HERA data drive the differences in χ^2 values, with rather smaller contributions from the fixed-target experiments.

The global fits behave in a similar way, with the HERA data still dominating χ^2 differences. However, a marked χ^2 spread is now also seen for the LHC experiments. This shows that the precise collider (HERA and LHC) data are most sensitive to higher order corrections. As in the case of DIS-only fits, for almost all experiments the central scale choice $(k_f, k_r) = (1, 1)$ provides the best overall description of the

various datasets. A notable exception is DY, for which the scale choice $(1, \frac{1}{2})$ leads to an improved fit (see Table 15), in agreement with the argument (often used for Higgs production in gluon fusion) that the natural renormalization scale for inclusive production of a colorless object is half its mass.

The fact that, with the exception of the combination $(k_f, k_r) = (2, \frac{1}{2})$ all scale choices lead to PDFs in reasonable agreement with the data, suggests that an estimate of MHO might be obtained by taking an envelope of PDFs determined with different scale choices. We consider in particular: the 9-point envelope, in which all combinations of scales of Table. (13) are included in the envelope; the 7-point envelope, in which the two choices $(k_f, k_r) = (2, \frac{1}{2})$ and $(k_f, k_r) = (\frac{1}{2}, 2)$ are removed from the 9-point envelope; and the 3-point envelope, in which only the two choices $(k_f, k_r) = (2, 2)$ and $(k_f, k_r) = (\frac{1}{2}, \frac{1}{2})$ are considered together with the central scale choices. These envelopes are shown in Fig. 25. It is clear that the size of the envelope is

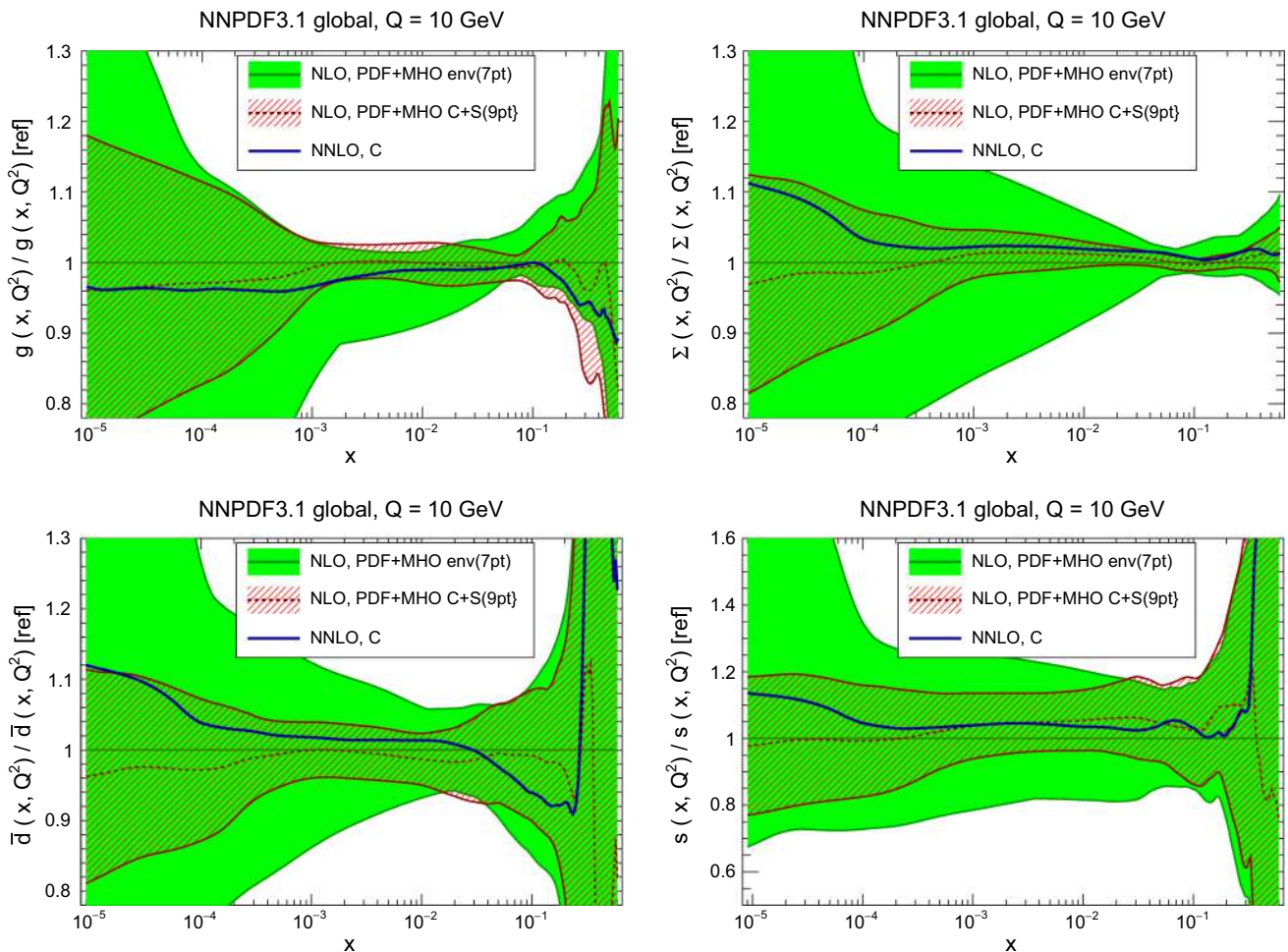


Fig. 28 Same as Fig. 27 comparing the total uncertainty obtained from the 7-point estimate of the MHO using Eq. (B.3) to the total uncertainty obtained from the theory covariance matrix [same as Fig. 16] computed using the 9-point prescription. The central NNLO value

obtained using the experimental covariance matrix is also shown. All results are normalized to the central NLO value with experimental covariance matrix

extremely sensitive to the choice of scales to be included. Of course, by construction, an envelope including more scale choices always leads to a wider band than envelopes with fewer choice, and indeed the 7-point envelope leads to significantly larger uncertainties than the 3-point envelope, though the 7-point and 9-point envelopes essentially coincide.

Further insight on the envelope method can be obtained by comparing the envelope of scale variations, taken as a candidate MHO, for DIS-only and global PDF sets. This is done in Fig. 26, where the standard, data-driven PDF uncertainties are also shown for reference. Whereas the PDF uncertainties always decrease when from DIS-only to global PDFs (and so did the theory uncertainties when determined using the theory covariance matrix in Sect. 6) the theory uncertainties estimated from the envelope prescription behave more erratically, with the envelope for the global fit leading to a wider band in the case of the singlet distribution. Quite in general, the envelope estimates of MHOs appear to be rather large in

comparison to PDF uncertainties, and unstable upon changes in dataset.

Finally, we can ask how MHOs estimated from an envelope could be combined with the data-driven PDF uncertainties. In the case of the covariance matrix approach discussed in Sect. 6, results are found using default NNPDF methodology including an extra contribution to the covariance matrix. Here, however, we need a prescription for the combination of MHOs (obtained from the envelope of scale variations) with the data-driven PDF uncertainties (obtained using default NNPDF methodology with a purely experimental covariance matrix).

A possible prescription would be to calculate the total uncertainty on PDFs as a sum in quadrature of the envelope MHO and the standard PDF uncertainty. Taking into account that these envelopes are asymmetric the prescription is then

$$\sigma^{\text{tot},\pm} = \left((\sigma^{\text{mho},\pm})^2 + (\sigma^{\text{PDF}})^2 \right)^{1/2}, \quad (\text{B.3})$$

where $\sigma_q^{\text{mho},+(-)}$ indicates the upper (lower) limit of the envelope, and σ_q^{PDF} is the standard PDF uncertainty.

In Fig. 27 we show the uncertainties $\sigma^{\text{tot},\pm}$, $\sigma^{\text{mho},\pm}$ and σ^{PDF} Eq. (B.3) using the 7-point envelope for the gluon, the quark singlet, the down antiquark, and the charm PDFs, all normalized to the central value. In Fig. 28 we further compare, for the same PDF combinations, the total uncertainties obtained with the envelope method (shown in Fig. 27) with the total uncertainties obtained with our theory covariance matrix methodology (shown in Fig. 16), all normalized to the central curve of the envelope method. The NNLO central curve (with experimental covariance matrix only) is also shown. Results are obtained using the baseline settings: the 7-point prescription for the envelope method and the 9-point prescription for the theory covariance matrix.

It is clear that some qualitative features are common to both uncertainty estimates; in particular, the asymmetry of the envelope prescription favors variations which go towards the direction of the true NNLO result. However, it is also clear that the envelope prescription has a number of shortcomings: it leads to discontinuous and asymmetric uncertainties, which are difficult to accommodate in a Gaussian framework; it is very unstable and strongly dependent on arbitrary choices for the set of the scale variations over which the envelope should be taken; it is quite cumbersome and again arbitrary in requiring one to postulate a specific way of combining MHO and data-induced PDF uncertainties; it leads to very large MHOs which appear to be overestimated in comparison to the known shift to the NNLO result if a 7-point prescription is used.

The reason for the much greater stability of MHO estimated using the covariance matrix prescription should be clear: when using an envelope prescription, any large deviation in a given direction leads to large uncertainties in that direction, regardless of whether indeed there are large MHO or not. In a covariance matrix approach, a large eigenvalue in any given direction will allow the fit to move in that direction. This, however, at least in the presence of abundant experimental information, will actually happen only if the data pull in that direction due to MHO, and otherwise it will have little effect. Note also that, by construction, in an envelope approach the best fit will be the same as that in which MHO are not included. So it is possible to have a more conservative estimate of the overall uncertainty, but not a more accurate result.

We conclude that, whereas results for MHO based on scale varied fits and an envelope prescription are by and large consistent with those obtained with a covariance matrix approach, they are less stable, less reliable, and less accurate.

References

1. Physics of the HL-LHC Working Group Collaboration, M. Cepeda et al., Higgs Physics at the HL-LHC and HE-LHC. [arXiv:1902.00134](#)
2. HL-LHC, HE-LHC Working Group Collaboration, P. Azzi et al., Standard model physics at the HL-LHC and HE-LHC. [arXiv:1902.04070](#)
3. J. Gao, L. Harland-Lang, J. Rojo, The Structure of the Proton in the LHC Precision Era. *Phys. Rept.* **742**, 1–121 (2018). [arXiv:1709.04922](#)
4. NNPDF Collaboration, R.D. Ball et al., Parton distributions for the LHC Run II. *JHEP* **04**, 040 (2015). [arXiv:1410.8849](#)
5. NNPDF Collaboration, R.D. Ball et al., Parton distributions from high-precision collider data. *Eur. Phys. J. C* **77**(10), 663 (2017). [arXiv:1706.00428](#)
6. J. Campbell, J. Huston, F. Krauss, *The Black Book of Quantum Chromodynamics* (Oxford University Press, Oxford, 2017)
7. S.I. Alekhin, Global fit to the charged leptons DIS data: alpha(s) parton distributions, and high twists. *Phys. Rev. D* **63**, 094022 (2001). [arXiv:hep-ph/0011002](#)
8. The NNPDF Collaboration, R.D. Ball et al., Theoretical issues in PDF determination and associated uncertainties. *Phys. Lett. B* **723**, 330 (2013). [arXiv:1303.1189](#)
9. H. Paukkunen, Nuclear PDFs Today. *PoS HardProbes2018*, 014 (2018). [arXiv:1811.01976](#)
10. NNPDF Collaboration, R.D. Ball, E.R. Nocera, R.L. Pearson, nuclear uncertainties in the determination of proton PDFs. *Eur. Phys. J. C* **79**(3), 282 (2019). [arXiv:1812.09074](#)
11. R.D. Ball, A. Deshpande, The Proton Spin, semi-Inclusive processes, and a future Electron Ion Collider (2018). [arXiv:1801.04842](#)
12. M. Cacciari, N. Houdeau, Meaningful characterisation of perturbative theoretical uncertainties. *JHEP* **1109**, 039 (2011). [arXiv:1105.5152](#)
13. A. David, G. Passarino, How well can we guess theoretical uncertainties? *Phys. Lett. B* **726**, 266–272 (2013). [arXiv:1307.1843](#)
14. E. Bagnaschi, M. Cacciari, A. Guffanti, L. Jenniches, An extensive survey of the estimation of uncertainties from missing higher orders in perturbative calculations. *JHEP* **02**, 133 (2015). [arXiv:1409.5036](#)
15. L.A. Harland-Lang, R.S. Thorne, On the Consistent Use of Scale Variations in PDF Fits and Predictions. *Eur. Phys. J. C* **79**(3), 225 (2019). [arXiv:1811.08434](#)
16. NNPDF Collaboration, R. Abdul Khalek et al., A first determination of parton distributions with theoretical uncertainties. [arXiv:1905.04311](#)
17. G. Altarelli, R.D. Ball, S. Forte, Small x resummation with quarks: Deep-inelastic scattering. *Nucl. Phys. B* **799**, 199–240 (2008). [arXiv:0802.0032](#)
18. LHC Higgs Cross Section Working Group Collaboration, D. de Florian et al., Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector. [arXiv:1610.07922](#)
19. A.D. Martin, R.G. Roberts, W.J. Stirling, Precision analysis of Λ (MS) and the gluon distribution and its implication for jet and top quark cross-sections. *Phys. Rev. D* **43**, 3648–3656 (1991)
20. M. Virchaux, A. Milsztajn, A measurement of alpha-s and higher twists from a QCD analysis of high statistics F-2 data on hydrogen and deuterium targets. *Phys. Lett. B* **274**, 221–229 (1992)
21. G. Ridolfi, S. Forte, Renormalization and factorization scale dependence of observables in QCD. *J. Phys. G* **25**, 1555–1556 (1999)
22. A. Buckley, J. Ferrando, S. Lloyd, K. Nordström, B. Page et al., LHAPDF6: parton density access in the LHC precision era. *Eur. Phys. J. C* **75**, 132 (2015). [arXiv:1412.7420](#)

23. G. Altarelli, M. Diemoz, G. Martinelli, P. Nason, Total cross-sections for heavy flavor production in hadronic collisions and QCD. Nucl. Phys. B **308**, 724–752 (1988)
24. P. Nason, S. Dawson, R.K. Ellis, The total cross-section for the production of heavy quarks in hadronic collisions. Nucl. Phys. B **303**, 607–633 (1988)
25. F.E. Close, R.G. Roberts, G.G. Ross, Factorization scale independence, the connection between alternative explanations of the EMC Effect and QCD predictions for nuclear properties. Nucl. Phys. B **296**, 582–610 (1988)
26. W.L. van Neerven, A. Vogt, NNLO evolution of deep inelastic structure functions: The Singlet case. Nucl. Phys. B **588**, 345–373 (2000). [arXiv:hep-ph/0006154](#)
27. R.D. Ball, S. Carrazza, L. Del Debbio, S. Forte, J. Gao et al., Parton distribution benchmarking with LHC data. JHEP **1304**, 125 (2013). [arXiv:1211.5142](#)
28. New Muon Collaboration, M. Arneodo et al., Accurate measurement of F_2^d/F_2^p and $R_d - R_p$. Nucl. Phys. B **487**, 3–26 (1997). [arXiv:hep-ex/9611022](#)
29. 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 σ_L/σ_T . Nucl. Phys. B **483**, 3–43 (1997). [arXiv:hep-ph/9610231](#)
30. L.W. Whitlow, E.M. Riordan, S. Dasu, S. Rock, 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. B **282**, 475–482 (1992)
31. 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. B **223**, 485 (1989)
32. 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. B **237**, 592 (1990)
33. NuTeV Collaboration, M. Goncharov et al., Precise measurement of dimuon production cross-sections in $\nu_\mu\text{Fe}$ and $\bar{\nu}_\mu\text{Fe}$ deep inelastic scattering at the Tevatron. Phys. Rev. D **64**, 112006 (2001). [arXiv:hep-ex/0102049](#)
34. D.A. Mason, Measurement of the strange—antistrange asymmetry at NLO in QCD from NuTeV dimuon data. Ph D Thesis, University of Oregon. FERMILAB-THESIS-2006-01. <https://lss.fnal.gov/archive/thesis/2000/fermilab-thesis-2006-01.pdf> (2019)
35. CHORUS Collaboration, G. Onengut et al., Measurement of nucleon structure functions in neutrino scattering. Phys. Lett. B **632**, 65–75 (2006)
36. ZEUS, H1 Collaboration, H. Abramowicz et al., Combination of measurements of inclusive deep inelastic $e^\pm p$ scattering cross sections and QCD analysis of HERA data. Eur. Phys. J. C **75**(12), 580 (2015). [arXiv:1506.06042](#)
37. H1, ZEUS Collaboration, H. Abramowicz et al., Combination and QCD analysis of charm production cross section measurements in deep-inelastic ep scattering at HERA. Eur. Phys. J. C **73**, 2311 (2013). [arXiv:1211.1182](#)
38. CDF Collaboration, T.A. Aaltonen et al., Measurement of $d\sigma/dy$ of Drell-Yan e^+e^- pairs in the Z Mass Region from $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV. Phys. Lett. B **692**, 232–239 (2010). [arXiv:0908.3914](#)
39. D0 Collaboration, V. M. Abazov et al., Measurement of the shape of the boson rapidity distribution for $p\bar{p} \rightarrow Z/\gamma^* \rightarrow e^+e^- + X$ events produced at $\sqrt{s}=1.96$ -TeV. Phys. Rev. D **76**, 012003 (2007). [arXiv:hep-ex/0702025](#)
40. D0 Collaboration, V. M. Abazov et al., Measurement of the electron charge asymmetry in $p\bar{p} \rightarrow W + X \rightarrow e\nu + X$ decays in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV. Phys. Rev. D **91**(3) 032007 (2015). [arXiv:1412.2862](#) [Erratum: Phys. Rev. D **91**, no.7, 079901 (2015)]
41. D0 Collaboration, V.M. Abazov et al., Measurement of the muon charge asymmetry in $p\bar{p} \rightarrow W + X \rightarrow \mu\nu + X$ events at $\sqrt{s} = 1.96$ TeV. Phys. Rev. D **88**, 091102 (2013). [arXiv:1309.2591](#)
42. ATLAS Collaboration, G. Aad et al., Measurement of the inclusive W^\pm and Z/γ^* cross sections in the electron and muon decay channels in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector. Phys. Rev. D **85**, 072004 (2012). [arXiv:1109.5141](#)
43. ATLAS Collaboration, M. Aaboud et al., Precision measurement and interpretation of inclusive W^+ , W^- and Z/γ^* production cross sections with the ATLAS detector. Eur. Phys. J. C **77**(6), 367 (2017). [arXiv:1612.03016](#)
44. ATLAS Collaboration, G. Aad et al., Measurement of the low-mass Drell–Yan differential cross section at $\sqrt{s} = 7$ TeV using the ATLAS detector. JHEP **06**, 112 (2014). [arXiv:1404.1212](#)
45. ATLAS Collaboration, G. Aad et al., Measurement of the high-mass Drell–Yan differential cross-section in pp collisions at $\sqrt{s}=7$ TeV with the ATLAS detector. Phys. Lett. B **725**, 223 (2013). [arXiv:1305.4192](#)
46. ATLAS Collaboration, G. Aad et al., Measurement of the transverse momentum and ϕ_η^* distributions of Drell–Yan lepton pairs in proton–proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. Eur. Phys. J. C **76**(5), 291 (2016). [arXiv:1512.02192](#)
47. ATLAS Collaboration, G. Aad et al., Measurement of inclusive jet and dijet production in pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector. Phys. Rev. D **86**, 014022 (2012). [arXiv:1112.6297](#)
48. ATLAS Collaboration, G. Aad et al., Measurement of the $t\bar{t}$ production cross-section using $e\mu$ events with b-tagged jets in pp collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS detector. Eur. Phys. J. C **74**(10), 3109 (2014). [arXiv:1406.5375](#). [Addendum: Eur. Phys. J. C **76**(11), 642 (2016)]
49. ATLAS Collaboration, M. Aaboud et al., Measurement of the $t\bar{t}$ production cross-section using $e\mu$ events with b-tagged jets in pp collisions at $\sqrt{s}=13$ TeV with the ATLAS detector. Phys. Lett. B **761**, 136–157 (2016). [arXiv:1606.02699](#)
50. ATLAS Collaboration, G. Aad et al., Measurements of top-quark pair differential cross-sections in the lepton+jets channel in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector. Eur. Phys. J. C **76**(10), 538 (2016). [arXiv:1511.04716](#)
51. CMS Collaboration, S. Chatrchyan et al., Measurement of the differential and double-differential Drell–Yan cross sections in proton–proton collisions at $\sqrt{s} = 7$ TeV. JHEP **1312**, 030 (2013). [arXiv:1310.7291](#)
52. CMS Collaboration, S. Chatrchyan et al., Measurement of the electron charge asymmetry in inclusive W production in pp collisions at $\sqrt{s} = 7$ TeV. Phys. Rev. Lett. **109**, 111806 (2012). [arXiv:1206.2598](#)
53. CMS Collaboration, S. Chatrchyan et al., Measurement of the muon charge asymmetry in inclusive pp to WX production at $\sqrt{s} = 7$ TeV and an improved determination of light parton distribution functions. Phys. Rev. D **90**, 032004 (2014). [arXiv:1312.6283](#)
54. CMS Collaboration, V. Khachatryan et al., Measurement of the differential cross section and charge asymmetry for inclusive $pp \rightarrow W^\pm + X$ production at $\sqrt{s} = 8$ TeV. Eur. Phys. J. C **76**(8), 469 (2016). [arXiv:1603.01803](#)
55. CMS Collaboration, V. Khachatryan et al., Measurement of the Z boson differential cross section in transverse momentum and rapidity in protonproton collisions at 8 TeV. Phys. Lett. B **749**, 187–209 (2015). [arXiv:1504.03511](#)
56. CMS Collaboration, S. Chatrchyan et al., Measurements of differential jet cross sections in proton–proton collisions at $\sqrt{s} = 7$ TeV with the CMS detector. Phys. Rev. D **87**, 112002 (2013). [arXiv:1212.6660](#)
57. CMS Collaboration, V. Khachatryan et al., Measurement of the t-bar production cross section in the e-mu channel in proton–proton collisions at $\sqrt{s} = 7$ and 8 TeV. JHEP **08**, 029 (2016). [arXiv:1603.02303](#)

58. CMS Collaboration, V. Khachatryan et al., Measurement of the top quark pair production cross section in proton–proton collisions at $\sqrt{s} = 13$ TeV. *Phys. Rev. Lett.* **116**(5), 052002 (2016). [arXiv:1510.05302](#)
59. CMS Collaboration, V. Khachatryan et al., Measurement of the differential cross section for top quark pair production in pp collisions at $\sqrt{s} = 8$ TeV. *Eur. Phys. J. C* **75**(11), 542 (2015). [arXiv:1505.04480](#)
60. LHCb Collaboration, R. Aaij et al., Inclusive W and Z production in the forward region at $\sqrt{s} = 7$ TeV. *JHEP* **1206**, 058 (2012). [arXiv:1204.1620](#)
61. LHCb Collaboration, R. Aaij et al., Measurement of the cross-section for $Z \rightarrow e^+e^-$ production in pp collisions at $\sqrt{s} = 7$ TeV. *JHEP* **1302**, 106 (2013). [arXiv:1212.4620](#)
62. LHCb Collaboration, R. Aaij et al., Measurement of the forward Z boson production cross-section in pp collisions at $\sqrt{s} = 7$ TeV. *JHEP* **08**, 039 (2015). [arXiv:1505.07024](#)
63. LHCb Collaboration, R. Aaij et al., Measurement of forward W and Z boson production in pp collisions at $\sqrt{s} = 8$ TeV. *JHEP* **01**, 155 (2016). [arXiv:1511.08039](#)
64. NuSea Collaboration, J. C. Webb et al., Absolute Drell–Yan dimuon cross sections in 800-GeV/c p p and p d collisions. [arXiv:hep-ex/0302019](#)
65. J.C. Webb, Measurement of continuum dimuon production in 800-GeV/c proton nucleon collisions. [arXiv:hep-ex/0301031](#)
66. FNAL E866/NuSea Collaboration, R.S. Towell et al., Improved measurement of the anti-d/anti-u asymmetry in the nucleon sea. *Phys. Rev. D* **64**, 052002 (2001). [arXiv:hep-ex/0103030](#)
67. G. Moreno et al., Dimuon production in proton–copper collisions at $\sqrt{s} = 38.8$ -GeV. *Phys. Rev. D* **43**, 2815–2836 (1991)
68. NNPDF Collaboration, R.D. Ball, S. Carrazza, L. Del Debbio, S. Forte, Z. Kassabov, J. Rojo, E. Slade, M. Ubiali, Precision determination of the strong coupling constant within a global PDF analysis. *Eur. Phys. J. C* **78**(5), 408 (2018). [arXiv:1802.03398](#)
69. V. Bertone, S. Carrazza, J. Rojo, APFEL: A PDF evolution library with QED corrections. *Comput. Phys. Commun.* **185**, 1647 (2014). [arXiv:1310.1394](#)
70. S. Forte, E. Laenen, P. Nason, J. Rojo, Heavy quarks in deep-inelastic scattering. *Nucl. Phys. B* **834**, 116–162 (2010). [arXiv:1001.2312](#)
71. R.D. Ball, M. Bonvini, L. Rottoli, Charm in deep-inelastic scattering. *JHEP* **11**, 122 (2015). [arXiv:1510.02491](#)
72. R.D. Ball, V. Bertone, M. Bonvini, S. Forte, P.G. Merrild, J. Rojo, L. Rottoli, Intrinsic Charm in a matched general-mass scheme. *Phys. Lett. B* **754**, 49–58 (2016). [arXiv:1510.00009](#)
73. NNPDF Collaboration, R.D. Ball, V. Bertone, M. Bonvini, S. Carrazza, S. Forte, A. Guffanti, N.P. Hartland, J. Rojo, L. Rottoli, A determination of the charm content of the proton. *Eur. Phys. J. C* **76**(11), 647 (2016). [arXiv:1605.06515](#)
74. Particle Data Group Collaboration, C. Patrignani et al., Review of Particle Physics. *Chin. Phys. C* **40**(10), 100001 (2016)
75. Z. Kassabov, Reportengine: A framework for declarative data analysis (2019). <https://doi.org/10.5281/zenodo.2571601>
76. V. Bertone, S. Carrazza, N.P. Hartland, APFELgrid: a high performance tool for parton density determinations. *Comput. Phys. Commun.* **212**, 205–209 (2017). [arXiv:1605.02070](#)
77. T. Carli et al., A posteriori inclusion of parton density functions in NLO QCD final-state calculations at hadron colliders: The APPLGRID Project. *Eur. Phys. J. C* **66**, 503 (2010). [arXiv:0911.2985](#)
78. G.P. Salam, J. Rojo, A higher order perturbative parton evolution toolkit (HOPPET). *Comput. Phys. Commun.* **180**, 120–156 (2009). [arXiv:0804.3755](#)
79. S. Alekhin, J. Blümlein, S. Moch, Parton distribution functions and benchmark cross sections at NNLO. *Phys. Rev. D* **86**, 054009 (2012). [arXiv:1202.2281](#)
80. The NNPDF Collaboration, R.D. Ball et al., Fitting parton distribution data with multiplicative normalization uncertainties. *JHEP* **05**, 075 (2010). [arXiv:0912.2276](#)
81. C. Anastasiou, C. Duhr, F. Dulat, F. Herzog, B. Mistlberger, Higgs Boson Gluon-Fusion Production in QCD at Three Loops. *Phys. Rev. Lett.* **114**(21), 212001 (2015). [arXiv:1503.06056](#)
82. C. Anastasiou, C. Duhr, F. Dulat, E. Furlan, T. Gehrmann, F. Herzog, A. Lazopoulos, B. Mistlberger, High precision determination of the gluon fusion Higgs boson cross-section at the LHC. *JHEP* **05**, 058 (2016). [arXiv:1602.00695](#)
83. B. Mistlberger, Higgs boson production at hadron colliders at N^3 LO in QCD. *JHEP* **05**, 028 (2018). [arXiv:1802.00833](#)
84. F.A. Dreyer, A. Karlberg, Vector-Boson Fusion Higgs Production at Three Loops in QCD. *Phys. Rev. Lett.* **117**(7), 072001 (2016). [arXiv:1606.00840](#)
85. R.D. Ball, M. Bonvini, S. Forte, S. Marzani, G. Ridolfi, Higgs production in gluon fusion beyond NNLO. *Nucl. Phys. B* **874**, 746 (2013). [arXiv:1303.3590](#)
86. M. Bonvini, R.D. Ball, S. Forte, S. Marzani, G. Ridolfi, Updated Higgs cross section at approximate N^3 LO. *J. Phys. G* **41**, 095002 (2014). [arXiv:1404.3204](#)
87. M. Bonvini, S. Marzani, C. Muselli, L. Rottoli, On the Higgs cross section at N^3 LO+ N^3 LL and its uncertainty. *JHEP* **08**, 105 (2016). [arXiv:1603.08000](#)
88. T. Ahmed, M. Bonvini, M.C. Kumar, P. Mathews, N. Rana, V. Ravindran, L. Rottoli, Pseudo-scalar Higgs boson production at N^3 LO_A + N^3 LL'. *Eur. Phys. J. C* **76**(12), 663 (2016). [arXiv:1606.00837](#)
89. M. Bonvini, S. Marzani, Double resummation for Higgs production. *Phys. Rev. Lett.* **120**(20), 202003 (2018). [arXiv:1802.07758](#)
90. M. Bonvini, Small- x phenomenology at the LHC and beyond: HELL 3.0 and the case of the Higgs cross section. *Eur. Phys. J. C* **78**(10), 834 (2018). [arXiv:1805.08785](#)
91. M. Cacciari, F.A. Dreyer, A. Karlberg, G.P. Salam, G. Zanderighi, Fully differential vector-boson-fusion higgs production at next-to-next-to-leading order. *Phys. Rev. Lett.* **115**(8), 082002 (2015). [arXiv:1506.02660](#)
92. M. Czakon, A. Mitov, Top++: A program for the calculation of the top-pair cross-section at hadron colliders. *Comput. Phys. Commun.* **185**, 2930 (2014). [arXiv:1112.5675](#)
93. M. Grazzini, S. Kallweit, M. Wiesemann, Fully differential NNLO computations with MATRIX. *Eur. Phys. J. C* **78**(7), 537 (2018). [arXiv:1711.06631](#)
94. F. Cascioli, P. Maierhofer, S. Pozzorini, Scattering amplitudes with open loops. *Phys. Rev. Lett.* **108**, 111601 (2012). [arXiv:1111.5206](#)
95. T. Matsuura, S.C. van der Marck, W.L. van Neerven, The calculation of the second order soft and virtual contributions to the Drell–Yan cross-section. *Nucl. Phys. B* **319**, 570–622 (1989)
96. A. Denner, S. Dittmaier, L. Hofer, Collier: a fortran-based Complex one-loop library in extended regularizations. *Comput. Phys. Commun.* **212**, 220–238 (2017). [arXiv:1604.06792](#)
97. S. Forte, Z. Kassabov, J. Rojo, L. Rottoli, Theoretical uncertainties and dataset dependence of parton distributions. In *Proceedings of the 2017 Les Houches workshop 'Physics at TeV colliders*. [arXiv:1803.07977](#)