

Theoretical Uncertainties in Parton Distribution Functions

Rosalyn Laura Pearson



Doctor of Philosophy
The University of Edinburgh
May 2020

Abstract

We are now in the era of high precision particle physics, spurred on by a wealth of new data from the Large Hadron Collider (LHC). In order to match the precision set by modern experiments and test the limits of the Standard Model, we must increase the sophistication of our theoretical predictions. Much of the data available involves the interaction of protons, which are composite particles. Their interactions are described by combining perturbative Quantum Field Theory (QFT) with parton distribution functions (PDFs), which encapsulate the non-perturbative behaviour. Increasing accuracy and precision of these PDFs is therefore of great value to modern particle physics.

PDFs are determined by multi-dimensional fits of experimental data to theoretical predictions from QFT. Uncertainties in PDFs arise from those in the experimental data and theoretical predictions, as well as from the methodology of the fit. At the current levels of precision theoretical uncertainties are increasingly significant, but have so far not been included in PDF fits. Such uncertainties arise from many sources, an important one being the truncation of the perturbative expansion for the theoretical predictions to a fixed order, resulting in missing higher order uncertainties (MHOUs).

In this thesis we consider how to include theory uncertainties in future PDF fits, and address several sources of uncertainties. MHOUs are estimated and included as a proof of concept in next-to-leading order (NLO) PDFs. We find that these capture many of the important features of the known PDFs at the next order above (NNLO). We then go on to investigate uncertainties from previously ignored heavy nuclear effects and higher twist effects, estimate their magnitude and assess the impact of their inclusion on the PDFs.

Declaration

I declare that this thesis was composed by myself, and details work carried out as a member of the NNPDF collaboration, and alongside my supervisor Richard Ball. Unless explicitly stated in the text, all results are mine or come from collaborative projects to which I have made a significant contribution. This work has not been submitted for any other degree or professional qualification.

Parts of this work have been published in [].

(Rosalyn Laura Pearson, May 2020)

Acknowledgements

Insert people you want to thank here.

Contents

Abstract	i
Declaration	ii
Acknowledgements	iii
Contents	v
List of Figures	vi
List of Tables	vii
0.1 List of Publications	vii
Introduction	vii
1 Background	1
1.0.1 Deep inelastic scattering.....	2
1.0.2 The parton model.....	4
1.0.3 Quantum Chromodynamics (QCD).....	7
1.0.4 The QCD improved parton model and factorisation.....	9
1.0.5 Hadroproduction	13
1.0.6 Sum rules	14

1.1	Determining PDFs	15
1.1.1	Experimental and theoretical input.....	15
1.1.2	Experimental uncertainties	15
1.1.3	NNPDF fitting strategy	17
1.1.4	Monte Carlo approach	19
1.1.5	Neural Networks.....	21
1.1.6	Parametrisation, preprocessing and postprocessing	23
1.1.7	Cross validation	24
2	Theory uncertainties in PDFs - 10%	26
2.1	Introduction	26
2.2	A theoretical covariance matrix	33
2.3	MHOUs from scale variations.....	37
2.3.1	Renormalization group invariance.....	38
2.3.2	Scale variation for partonic cross-sections	41
2.3.3	Scale variation for PDF evolution	45
2.3.4	Double scale variations.....	50
2.3.5	Multiple scale variations.....	56
2.4	Scale variation prescriptions for the theory covariance matrix.....	58
2.4.1	Symmetric prescriptions for individual processes	60
2.4.2	Symmetric prescriptions for multiple processes	61
2.4.3	Asymmetric prescriptions	65
2.5	Validation of the theory covariance matrix	68
2.5.1	Input data and process categorization	68

2.5.2	The theory covariance matrices at NLO	71
2.5.3	Construction of validation tests.....	74
2.5.4	Results of validation tests.....	77
2.6	PDFs with missing higher order uncertainties	82
2.6.1	DIS-only PDFs	83
2.6.2	Global PDFs	86
2.7	Implications for phenomenology	102
2.7.1	Higgs production	103
2.7.2	Top quark pair production	105
2.7.3	Z and W gauge boson production	107
2.8	Usage and delivery	108
2.8.1	Combining MHOUs in PDFs and hard matrix elements.....	112
2.8.2	Computation of the total uncertainty	114
2.8.3	Delivery	116
2.9	Summary and outlook.....	117
3	Missing higher order uncertainties 0%	119
3.1	Prescriptions to generate the theory covariance matrix	120
3.2	Validating the theory covariance matrix	120
3.3	PDFs with missing higher order uncertainties	120
3.4	Impact on phenomenology.....	120
3.5	Usage and delivery	120
4	Nuclear Uncertainties - 90%	121
4.1	Introduction	121

4.2	Nuclear Data	122
4.3	Determining Nuclear Uncertainties.....	122
4.4	The Impact on Global PDFs.....	124
4.5	Phenomenology	126
4.6	Conclusions	128
5	Deuteron Uncertainties - 0%	130
6	Higher Twist Uncertainties - 0%	131
6.1	The role of higher twist data in PDFs.....	131
6.2	Ansatz for a higher twist correction.....	131
6.3	Using a neural network to model higher twist	131
6.3.1	Model architecture.....	131
6.3.2	Training and validating the neural network.....	131
6.4	Form of the higher twist correction	131
6.5	The higher twist covariance matrix	131
6.6	PDFs with higher twist uncertainties	131
7	Conclusion - 0%	132
A	Diagonalisation of the theory covariance matrix	133
B	PDF sets with different scale choices	134
	Bibliography	135

Contents

List of Figures

(0.1.1) A visualisation of the internal structure of the proton. Quarks are bound together by gluons.	viii
(0.1.2) The different PDF flavours determined in the latest NNPDF3.1 release [16].	ix
(1.0.1) Deep inelastic scattering	2
(1.0.2) DIS in the parton model. One parton with momentum p interacts with the virtual photon, and the other partons “spectate”.	5
(1.0.3) Factorisation and the QCD improved parton model	10
(1.0.4) A quark radiating a gluon before interacting.	11
(1.0.5) Factorisation in hadron-hadron collisions.	14
(1.1.1) Plot of the (x, Q^2) range spanned by data included in the latest NNPDF3.1 NLO fit.	16
(1.1.2) An example of an experimental covariance matrix for data included in an NNPDF fit. The data are grouped according to what type of process the interaction belongs to (DIS charged current (CC) and neutral current (NC), Drell-Yan (DY), jets and top production).	18
(1.1.3) Schematic of the generation of Monte Carlo replicas of pseudodata from data with uncertainties.	19
(1.1.4) Histogram of distribution of 100 pseudodata replicas for a single data point, normalised to D^0 . The purple line is the mean value $\langle D^{(k)} \rangle$, which is equal to D^0 to arbitrary precision.	20
(1.1.5) NNPDF general strategy.	20
(1.1.6) Monte Carlo replicas for the down valence quark PDF NNPDF3.1 at NLO.	21

(1.1.7) Schematic depiction of the 2-5-3-1 architecture of an artificial neural network currently used by NNPDF. In the NNPDF methodology $\xi_1^{(1)}$ and $\xi_2^{(1)}$ are the variables x and $\log x$ respectively.	22
(1.1.8) Overlearning: the data points (black dots) fluctuate around the linear underlying law (black line), but the neural network continues to minimise the error function until it passes through every data point (blue curve), fitting the noise in the data.	24
(1.1.9) Cross validation with the lookback method.	25
(1.1.10) Comparing the training and validation χ^2 s for the 100 replicas (green circles) of a PDF fit. The red square gives the average.	25
(2.3.1) 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 $\kappa_f = \kappa_r$	55
(2.4.1) 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), 5-point (center) and 9-point (right).	60
(2.4.2) Same as Fig. 2.4.1, 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_{r1} and κ_{r2} , so the diagrams are now in three dimensions. The origin of coordinates is associated to the central scale, $\kappa_f = \kappa_{r1} = \kappa_{r2} = 0$. We again show the three prescriptions 5-point (left), 5-point (center) and 9-point (right).	63
(2.4.3) Same as Fig. 2.4.1, 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.	66
(2.5.1) Comparison of the diagonal experimental uncertainties (blue) and the diagonal theoretical uncertainties (red), each normalised to the data, with the experimental uncertainties 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 2.5.1	73

(2.5.2) 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 2.5.1	73
(2.5.3) Comparison of the experimental correlation matrix Eq. (2.5.1) (top left) and the combined experimental and theoretical correlation matrices Eq. (2.5.2) computed using the prescriptions described in Sect. 2.4: the symmetric prescriptions (5-pt top right, 5-pt centre left, 9-pt centre 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. 2.5.2	91
(2.5.4) 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, $\delta^{\text{miss}} \in D/S$	92
(2.5.5) 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), 5-point (middle), and 9-point (bottom). All values are shown as percentage of the central theory prediction.	93
(2.5.6) Same as Fig. 2.5.5 but for the asymmetric prescriptions: 3-point (top) and 7-point (bottom).	93
(2.5.7) 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.	94
(2.5.8) The projection δ^α Eq. (2.5.5) of the normalized shift vector δ_i Eq. (2.5.4) along each eigenvector e_i^α of the normalized theory covariance matrix Eq. (2.5.3), 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), 5-point (top right), and 9-point (bottom).	95
(2.5.9) Same as Fig. 2.5.8 but for the asymmetric prescriptions: 3-point (left) and 7-point (right).	96

(2.5.1) 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. 2.5.5. The NNLO-NLO shift, δ_i (black), is shown for comparison.	97
(2.6.1) Comparison of DIS-only PDFs determined with and without MHOUs in the covariance matrix. The gluon (top) and quark singlet (bottom) at $Q = 10$ GeV at NLO (left) and NNLO (right) are shown. The theory covariance matrix S has been constructed using the 9-point prescription. In the NLO plots the central value of the NNLO determined without MHOU is also shown. All results are shown 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 MHOU.	98
(2.6.2) Same as Fig. 2.6.1 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 . Results are shown at $Q = 10$ GeV for the gluon, the total quark singlet, the anti-down quark, and the total strangeness PDFs.	99
(2.6.3) Same as Fig. 2.6.2 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.	100
(2.6.4) Same as Fig. 2.6.3, 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.	101
(2.6.5) Same as Fig. 2.6.2, 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 on both.	109
(2.7.1) Graphical representation of the results of Tables 2.7.1 and 2.7.2. 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.	110

(2.7.2) same as Fig. 2.7.1 for top-quark pair production at 8 and 13 TeV, see also Table 2.7.3.	110
(2.7.3) same as Fig. 2.7.1 for W^\pm and Z gauge boson production at $\sqrt{s} = 13$ TeV, see also Table 2.7.4.	111
(4.3.1) The square root of the diagonal elements of the covariance matrices, normalised to corresponding data. Experimental contributions are red, theory green and the total blue. Data from CHORUS and NuTeV are split into neutrino and anti-neutrino parts. Points are binned in (anti-)neutrino beam energy E : 25, 35, 45, 55, 70, 90, 110, 120, 170 GeV. In each bin x increases from left to right, $0.045 < x < 0.65$	123
(4.4.1) Correlation matrices, $\rho_{ij}^{cov} = \frac{cov_{ij}}{\sqrt{cov_{ii}cov_{jj}}}$, before (left) and after (right) including nuclear uncertainties. Data are binned the same as in Fig. 4.4.2. The top row corresponds to CHORUS, middle row to NuTeV and bottom row to E605. Results are displayed for Def. 1 but are qualitatively similar for Def. 2.	125
(4.4.2) The square root of the diagonal elements of the covariance matrices, normalised to corresponding data. Experimental contributions are red, theory green and the total blue. Data from CHORUS and NuTeV are split into neutrino and anti-neutrino parts. Points are binned in (anti-)neutrino beam energy E : 25, 35, 45, 55, 70, 90, 110, 120, 170 GeV. In each bin x increases from left to right, $0.045 < x < 0.65$	127
(4.5.1) NucUnc fits with nuclear uncertainties (orange) compared to Baseline (green) for PDFs at 10 GeV. Clockwise from top left: \bar{u} , \bar{d} , s and \bar{s} PDFs. Error bands are 1σ ; results are normalised to Baseline fit.	128
(4.6.1) Effect of including nuclear uncertainties on phenomenology. From left to right: sea quark asymmetry, strangeness fraction, strange valence distribution. Distributions correspond to the use of different PDF fits: Baseline (green), NoNuc (yellow), NucUnc (blue) and NucCor(pink). $Q = 91.2$ GeV. In the left two plots, NucCor are indistinguishable from NucUnc so are omitted for readability.	129

List of Tables

(2.3.1) 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.	54
(2.5.1) 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.	70
(2.5.2) The angle θ Eq. (2.5.7) between this shift and its component δ_i^S lying within the subspace S (see Fig. 2.5.4) spanned by the theory covariance matrix for different prescriptions. The dimension of the subspace S in each case is also given.	78
(2.5.3) Same as Table 2.5.2 for each process of Table 2.5.1. The number of data points in each process is given directly below the name of the process.	79
(2.6.1) 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.	84
(2.6.2) 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 $S^{(9\text{pt})}$, compared to the results based on including only the experimental covariance matrix C . Results are provided both for the NLO and for the NNLO fits.	85

(2.6.3) 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 definition Eq. (2.6.2) but not in the data generation Eq. (2.6.1) (marked S_{fit}^{9pt}) and then in the data generation Eq. (2.6.1) but not in the χ^2 definition Eq. (2.6.2) (marked S_{sampler}^{9pt}). Values corresponding to the NNLO fit with experimental covariance matrix C only are also shown.	88
(2.6.4) Same as Table 2.6.3, but for the values of the ϕ estimator.	89
(2.7.1) 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 2.6.1. 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 (symmetrised) envelope of the 7-point scale variations (see Sect. 2.8.1), obtained by taking the maximum value between the lower and upper uncertainties.	104
(2.7.2) Same as Table 2.7.1, now for Higgs production in vector boson fusion.	105
(2.7.3) Same as Table 2.7.1, now for top-quark pair-production at $\sqrt{s} = 7, 8$ and 13 TeV.	106
(2.7.4) Same as Table 2.7.1, now for W and Z gauge boson production at $\sqrt{s} = 13$ TeV. The cross-section is given in nb.	108
(4.4.1) χ^2 per data point for selected data sets. The final row shows results for the full fitted data.	126

0.1 List of Publications

- **Towards parton distribution functions with theoretical uncertainties**, Pearson, R. L. and Voisey, C. Nuclear and Particle Physics Proceedings, Volumes 300-302, July-September 2018, Pages 24-29, e-Print: 1810.01996 [hep-ph]
- **Nuclear Uncertainties in the Determination of Proton PDFs**, NNPDF Collaboration: Richard D. Ball et al. (Dec 21, 2018), Published in: Eur.Phys.J.C 79 (2019) 3, 282, e-Print: 1812.09074 [hep-ph]
- **A first determination of parton distributions with theoretical**

uncertainties, NNPDF Collaboration: Rabah Abdul Khalek et al. (May 10, 2019), Published in: Eur.Phys.J. C (2019) 79:838, e-Print: 1905.04311 [hep-ph]

- **Uncertainties due to Nuclear Data in Proton PDF Fits**, Rosalyn Pearson, Richard Ball, Emanuele Roberto Nocera (Jun 14, 2019), Published in: PoS DIS2019 (2019) 027, Contribution to: DIS 2019, 027
- **Parton Distributions with Theory Uncertainties: General Formalism and First Phenomenological Studies**, NNPDF Collaboration: Rabah Abdul Khalek et al. (Jun 25, 2019) Published in: Eur.Phys.J.C 79 (2019) 11, 931, e-Print: 1906.10698 [hep-ph]

Introduction

Over the past 100 years, following the discovery of the atomic nucleus by Rutherford in 1911, great strides have been made towards understanding subatomic structure. We now know that atoms are made up of hadrons (such as protons and neutrons) and leptons (such as the electron). Probing hadrons with high energy photons shows that they are composed of quarks and gluons.

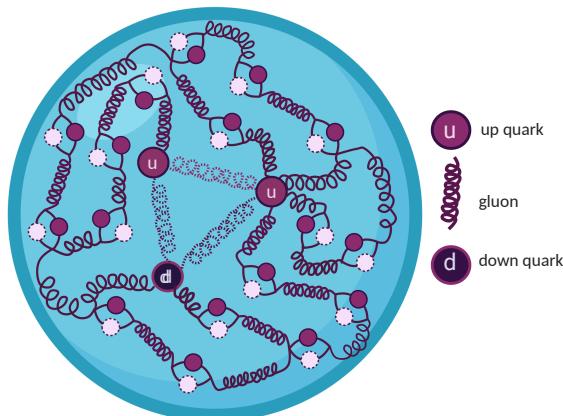


Figure 0.1.1 A visualisation of the internal structure of the proton. Quarks are bound together by gluons.

The Standard Model of particle physics has proven thus far to be an extremely accurate model of nature at the subatomic scale, and the current focus is on providing ever more precise experimental and theoretical results to test it and search for new physics which it cannot explain.

Cutting edge high energy physics experiments are currently being carried out at colliders such as the Large Hadron Collider (LHC) at CERN, and there are

many plans for new colliders such as the FCC, ILC and CLIC. Many of these experiments involve the collision of protons. At a basic level we can think of a proton as being composed of two up quarks and one down quark (uud), bound together by the strong interaction. However, the proton (Fig. ??) is in reality highly complicated and inaccessible to the normal perturbative calculations of Quantum Field Theory (QFT), and can only be treated using probabilistic methods.

When two protons collide we do not know which constituents, or “partons” are interacting, or what individual properties they have, such as their momentum and spin. We need some way of relating the known properties of the proton to the unknown properties of the partons. One way of doing this is using parton distribution functions (PDFs), which to first approximation give the probability of picking out a certain type of parton with certain properties.

Confinement of the quarks means experimental data is collected at the hadronic level, whereas theoretical predictions using QFT are made at the partonic level. The parton model provides a link between the two; in this framework partonic predictions are convolved with corresponding PDFs, summing over all possible partonic interactions. This produces PDF-dependent hadronic predictions. For useful theoretical predictions we therefore need as precise and accurate a handle on the PDFs as possible.

PDFs are unknowns in perturbative Quantum Chromodynamics (QCD), the theory of the strong interaction. Crucially, they are process independent. This means that they can be determined in a global fit between a wealth of experimental data and theoretical predictions. Once constrained, they can then be applied to any process. Fig. 4.5.1 shows the fitted functional form of the PDFs.

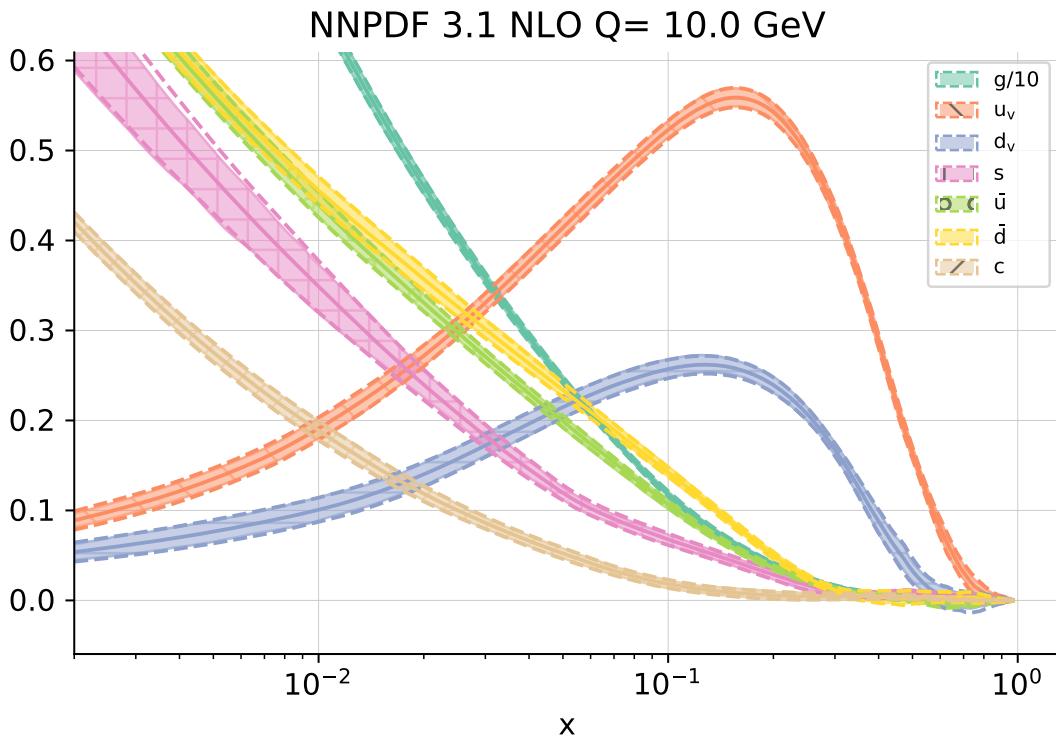


Figure 0.1.2 The different PDF flavours determined in the latest NNPDF3.1 release [16].

Any uncertainties in PDF determination will propagate through to future predictions. There are three places these uncertainties can be introduced:

1. experimental uncertainties;
2. theory uncertainties;
3. methodological uncertainties (errors from the fitting procedure).

Until recently, experimental uncertainties were the dominant source of error, meaning that theory uncertainties have been largely ignored in standard PDF fits. However, with the onset of increasingly high precision experiments, a proper treatment of theoretical uncertainties is becoming pressing.

Chapter 1

Background

Parton distribution functions (PDFs) bridge the gap between short and long range physics, allowing perturbative Quantum Chromodynamics (QCD) to be applied at the hadronic scale. They embody the incalculable strongly coupled dynamics, and are determined by a comparison of perturbative theory with experiment. Once determined, their form is process-independent and so they can be re-deployed in future calculations.

This section provides some background to PDFs necessary for understanding the remainder of this thesis. It is divided in to two main parts, being the necessary physics and the necessary methodology of PDF determination.

To review the physics, we begin by looking at the process of deep inelastic scattering (DIS), and how the naïve parton model was developed to explain these experimental observations. Next we look at this in the context of QCD, see how PDFs fit into the picture, and how they evolve with the scale of the physics. Finally we briefly touch on hadron-hadron collisions, which along with DIS constitute the bulk of the processes in modern PDF fits.

To review the methodology we consider the NNPDF fitting strategy, explaining how theory and experiment are used together with neural networks to determine PDFs.

1.0.1 Deep inelastic scattering

For a more in-depth analysis, see Refs. [36, 50]. In the following background sections we rely heavily on Ref. [28].

The notion of bombarding matter to uncover its structure has led to many important discoveries in the last hundred or so years, starting with the Geiger-Marsen experiments from 1908-1913 and the subsequent discovery of the atomic nucleus [44]. In the decades following the discovery of the neutron in 1932, nuclei were probed at higher energies, leading to them being understood in terms of “form factors” which parametrised their electric and magnetic distributions. At this stage it was clear that they were not point-like particles and so a series of important experiments were carried out in the 1960s at the Standford Linear Accelerator (SLAC), involving a high energy beam of charged leptons scattering off a stationary hadronic target. This process is known as Deep Inelastic Scattering (DIS).

In this section we will consider the example of electrons incident on protons, as shown in Fig. ???. In the deep inelastic regime, there is a large momentum transfer, $q = k - k'$, mediated by a virtual photon. The proton, P , with initial momentum p , fragments into some hadronic state X , and the electron starts with energy E and momentum k and ends with energy E' and momentum k' . The momentum transfer is large enough that the masses of the proton and electron can be neglected.

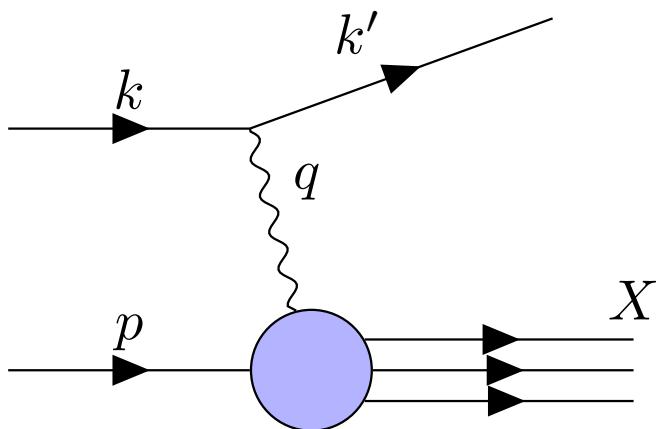


Figure 1.0.1 Deep inelastic scattering

It is customary to define some useful variables for help in the analysis, listed in the table below.

Variable	Definition	Interpretation
Q^2	$-q^2 = -(k - k')^2$	momentum transfer
ν	$p \cdot q = M(E' - E)$	energy transfer
x	$\frac{Q^2}{2\nu}$	scaling parameter
y	$\frac{q \cdot p}{k \cdot p} = 1 - \frac{E'}{E}$	inelasticity $\in [0, 1]$

The interaction is made up of a leptonic current (that of the electron) and a hadronic current (the fragmentation of the proton from P to X). This means we can express the squared matrix element, $|\mathcal{M}|^2$, as

$$|\mathcal{M}|^2 = \mathcal{N}_1 \frac{\alpha^2}{q^4} L_{\mu\nu} W^{\mu\nu}, \quad (1.0.1)$$

where $L_{\mu\nu}$ is the leptonic part, determined from perturbative Quantum Electrodynamics (QED), and $W^{\mu\nu}$ is the hadronic part, containing the incalculable strongly coupled dynamics. α is the QED coupling constant and \mathcal{N}_1 is a normalisation constant.

From QED, for an unpolarised photon beam in the DIS regime we can use the Feynman rules to write

$$\begin{aligned} L_{\mu\nu} &= \sum_{spins} \bar{u}(k') \gamma_\mu u(k) \bar{u}(k) \gamma_\nu u(k') \\ &= Tr(k' \gamma_\mu k \gamma_\nu) \\ &= \mathcal{N}_2 \left(k_\mu k'_\nu + k_\nu k'_\mu - g_{\mu\nu} k \cdot k' \right) \\ &= \mathcal{N}_2 \left(4k_\mu k_\nu - 2k_\mu q_\nu - 2k_\nu q_\mu + g_{\mu\nu} q^2 \right), \end{aligned} \quad (1.0.2)$$

where in the last line we used the fact that the electron is massless so $0 = k'^2 = q^2 + k^2 - 2q \cdot k \implies q^2 = 2q \cdot k$. We have also introduced another constant, \mathcal{N}_2 .

Finding the hadronic tensor is more difficult because we lack knowledge of the hadronic states P and X , so our only constraints are that $W^{\mu\nu}$ is Lorentz-invariant and that the electromagnetic current must be conserved, so $q \cdot W = 0$. This allows

us to write its general form as

$$W^{\mu\nu}(p, q) = -\left(g^{\mu\nu} - \frac{q^\mu q^\nu}{q^2}\right)W_1(p, q) + \left(p^\mu - q^\mu \frac{p \cdot q}{q^2}\right)\left(p^\nu - q^\nu \frac{p \cdot q}{q^2}\right)W_2(p, q), \quad (1.0.3)$$

where W_1 and W_2 are scalar functions which encapsulate the strong dynamics. These scalar functions are often written as:

$$\begin{aligned} F_1(x, Q^2) &= W_1(p, q); \\ F_2(x, Q^2) &= \nu W_2(p, q); \\ F_L(x, Q^2) &= F_2(x, Q^2) - 2x F_1(x, Q^2), \end{aligned} \quad (1.0.4)$$

and are known as the “structure functions”. Often the hadronic tensor is parametrised in terms of F_2 and F_L , the latter of which is the longitudinal structure function and encapsulates the longitudinal component.

We can now combine Eqns. 1.0.2 and 1.0.3 in Eqn. 1.0.1, making use of the fact that due to current conservation $q^\mu L_{\mu\nu} = 0$ to help simplify things. This leads us to the result:

$$|\mathcal{M}|^2 = \mathcal{N}_1 \mathcal{N}_2 \frac{\alpha^2}{q^4} \left\{ (-2q^2)W_1(p, q) + \left(4(p \cdot k)^2 - 4(p \cdot q)(p \cdot k)\right)W_2(p, q) \right\}. \quad (1.0.5)$$

1.0.2 The parton model

Carrying out DIS experiments allows us to measure the structure functions for different values of x and Q^2 . It transpired that no clear Q^2 dependence was observed, and this is known as Björken scaling [30]. Because Q^2 is the photon’s squared momentum, it corresponds to the energy at which the hadron is being probed. The fact that the structure functions are not dependent on this suggests that the interaction is point-like. This led to the formulation of the “parton model”, which described the proton as a composite state made up of point-like particles termed “partons” [38–40].

Furthermore, $F_L(x)$ was measured to be 0, known as the Callan-Gross relation ??, which suggests that the point-like particles could not absorb longitudinal photons. This fitted in nicely with the quark models developed shortly before [35, 45, 46, 58], which described hadrons in terms of spin-1/2 quarks; spin-

$1/2$ particles cannot interact with longitudinal photons. This was the first experimental evidence for the existence of quarks.

In the DIS regime, Q^2 is large and so the virtual photon probes at the short timescale $1/Q$, meaning that the interaction will be effectively instantaneous when compared with the inner proton dynamics which operate at the QCD scale $1/\lambda_{QCD} \sim 1$ fm. In the parton model we make the assumption that the partons have only a small momentum transverse to the proton's, and that they are effectively on shell for the interaction ($k^2 \approx 0$). In addition, we consider the process in the infinite momentum frame of the proton, in which it is Lorentz contracted by M/P (a small number), so we can assume the photon will only interact with one parton because it will only traverse a narrow cross-section of the proton. The updated picture is shown in Fig. ??.

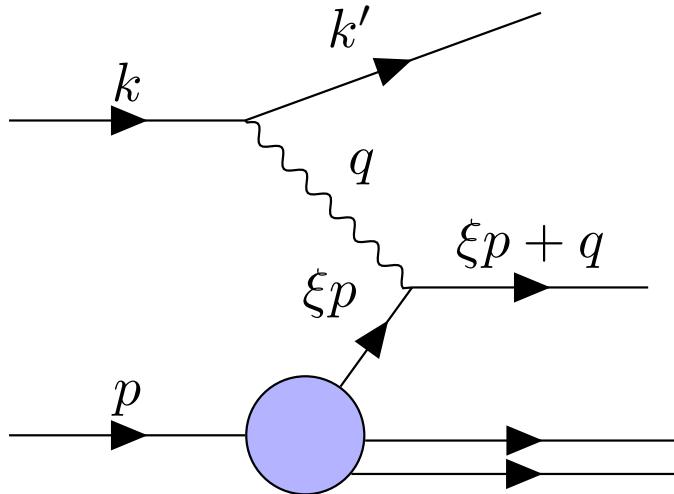


Figure 1.0.2 *DIS in the parton model. One parton with momentum p interacts with the virtual photon, and the other partons “spectate”.*

We parametrise the momentum of the interacting parton as ξp , $\xi \in [0, 1]$. The parton in the final state has negligible mass so its momentum squared is zero:

$$\begin{aligned}
 (\xi p + q)^2 &= 0 \\
 \implies 2\xi p \cdot q + q^2 &= 0 \\
 \implies 2\xi p \cdot q - Q^2 &= 0 \\
 \implies \xi &= \frac{Q^2}{2p \cdot q} \equiv x.
 \end{aligned} \tag{1.0.6}$$

This allows us to identify the parton's momentum fraction in this frame with the Björken x variable.

We can think of the total collection of interactions in terms of a weighted sum over the interactions between the photon and the individual point-like partons, and so can write the proton-level hadronic tensor, $W_{\mu\nu}$ in terms of the parton-level ones, $\hat{W}_{\mu\nu}^q$, as

$$W_{\mu\nu} = \sum_q f_q(x) \hat{W}_{\mu\nu}^q \delta(Q^2 - 2xp \cdot q) = \frac{1}{Q^2} \sum_q f_q(x) \hat{W}_{\mu\nu}^q \delta(1 - \frac{2xp \cdot q}{Q^2}), \quad (1.0.7)$$

where q runs over the possible quark flavours and f_q are distributions, with $f_q(x)dx$ giving the probability that in an interaction a parton of flavour q will be found in the momentum range $x \rightarrow x + dx$. We call these functions “parton distribution functions” (PDFs). The delta function appears due to integration over the final phase space of X , and enforces conservation of momentum. Using Eqn 1.0.1, we can see that

$$|\mathcal{M}|^2 = \frac{1}{Q^2} \sum_q f_q(x) |\hat{\mathcal{M}}_q|^2. \quad (1.0.8)$$

This means that the total amplitude can be expressed in terms of the partonic amplitudes and the PDFs. If we assume that the partons are massless Dirac particles, we can infer the partonic amplitudes directly from that of electron-muon scattering. In this scenario the electron has a current like Eqn. 1.0.2, and the muon has the same, but with the substitutions $k \rightarrow p$ and $q \rightarrow -q$. Once again we can use $q_\mu L^{\mu\nu} = 0$ and the expression

$$|\mathcal{M}_{(e\mu)}|^2 = \mathcal{N}_1 \frac{\alpha^2}{q^4} L_{\mu\nu}^{(e)} L_{(\mu)}^{\mu\nu} \quad (1.0.9)$$

to show (in the massless limit)

$$|\mathcal{M}_{(e\mu)}|^2 = \mathcal{N}_1 \mathcal{N}_2 \frac{\alpha^2}{q^4} \left(16(p \cdot k)^2 + 8q^2(p \cdot k) + 2q^4 \right). \quad (1.0.10)$$

Using the symmetry of Fig. ??, we can see this is analogous to $|\hat{\mathcal{M}}_q|^2$ under the substitution $p \rightarrow xp$, provided we replace the charge of the electron, e , with that

of the parton, e_q , so that $\alpha \rightarrow e_q\alpha$. Making use of the expression $p \cdot k = Q^2/2xy$,

$$\begin{aligned} |\hat{\mathcal{M}}_q|^2 &= \mathcal{N}_1 \mathcal{N}_2^2 \frac{e_q^2 \alpha^2}{q^4} \left\{ 4(2xp \cdot k)^2 + 4(2xp \cdot k)q^2 + 2q^4 \right\} \\ &= \mathcal{N}_1 \mathcal{N}_2^2 \frac{e_q^2 \alpha^2}{Q^4} \left\{ 4 \left(\frac{Q^2}{y} \right)^2 - 4 \left(\frac{Q^2}{y} \right) Q^2 + 2Q^4 \right\} \\ &= \mathcal{N}_1 \mathcal{N}_2^2 e_q^2 \alpha^2 \left\{ 2 + 4 \left(\frac{1-y}{y^2} \right) \right\}. \end{aligned} \quad (1.0.11)$$

Now we can use this alongside Eqn. 1.0.5 in Eqn. 1.0.8, giving us

$$\begin{aligned} F_1 &\equiv W_1 = \sum_q f_q(x) e_q^2, \\ F_2 &\equiv \nu W_2 = 2x \sum_q f_q(x) e_q^2. \end{aligned} \quad (1.0.12)$$

We see immediately that the Callan-Gross relation, $F_L(x) \equiv F_2(x) - 2xF_1(x) = 0$, is satisfied, as was observed experimentally.

However, it was soon observed that this relation only held in the limit $Q^2 \rightarrow \infty$, and that at smaller scales there were so-called “scaling violations”. In order to understand this behaviour it is necessary to revisit the parton model in the light of Quantum Chromodynamics (QCD).

1.0.3 Quantum Chromodynamics (QCD)

QCD is the theory of the strong force. This is responsible for binding together hadrons, and explains the short-range interactions which occur within them. It is a gauge theory where the quark fields are realised as fundamental representations of the $SU(3)$ symmetry group and interactions between them are carried out via gauge bosons termed “gluons”, which are expressed in the adjoint representation [49].

Quark models showed that the structure of observed hadrons can be explained using the $SU(3)_f$ group alongside the association of quarks with different “flavours” [35, 45, 46, 58]. The additional $SU(3)_c$ colour symmetry was put forwards in order that the quarks satisfied Fermi-Dirac statistics [47]. Each quark is assigned an additional colour ((anti-)red, green or blue) in such a way that the composite hadrons are colourless. The additional local symmetry is accompanied by eight gauge bosons, the gluons. Colour is the charge of QCD, just as electric

charge is for QED. An important difference is that, unlike chargeless photons in QED, the gluons themselves also have colour and this leads to complex self-interactions.

QCD can be expressed through the Lagrangian

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}^a F^{a\mu\nu} + \bar{q}^i(i\cancel{D}_i^j - m\delta_i^j)q_j, \quad (1.0.13)$$

where the covariant derivative is

$$\mathcal{D}_\mu \psi(x) = (\partial_\mu - i\sqrt{4\pi\alpha_s}T^a A_\mu^a)\psi(x), \quad (1.0.14)$$

and the field strength tensor is

$$F_{\mu\nu}^a = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a + \sqrt{4\pi\alpha_s}f^{abc}A_\mu^b A_\nu^c. \quad (1.0.15)$$

The indices μ, ν are spacetime indices, i, j are quark colour indices and a, b, c are gluon colour indices. The first term in the Lagrangian arises from the self-interacting gluons, A , and the second term from the quarks, q , which obey the Dirac equation. α_s is the strong coupling constant, which dictates the strength of the interaction, and T^a are the eight $SU(3)$ generators. f^{abc} are the $SU(3)$ structure constants. For simplicity we have assumed all quarks have the same mass, m . Note that gauge fixing and ghost terms are omitted. For more information see Ref. [36].

Colour self-interactions give rise to the important properties of “confinement” and “asymptotic freedom”. The QCD potential is of the form

$$V(r) \sim \frac{\alpha}{r} + kr, \quad (1.0.16)$$

where the first term drops off with distance like QED, but the second term comes from the self-interactions and means that separating two quarks takes infinite energy. This explains why we have not observed free quarks (“confinement”). Additionally, the QCD colour charge decreases with shorter distances. This means that at very short distances or high energies the quarks become “free”, which is known as “asymptotic freedom”. This crucial fact allows us to apply the tool of perturbation theory in such regimes.

QCD is subject to divergences in the ultra-violet (high energies) and infra-red (low energies). The former are regulated by renormalisation, which introduces a

“renormalisation scale”, μ_R . This is non-physical, and so observables cannot depend on it. This observation leads to a “renormalisation group equation” (RGE), which can be solved by the introduction of a running coupling, dependent on the scale Q^2 (i.e. $\alpha_s \rightarrow \alpha_s(Q^2)$), which satisfies

$$Q^2 \frac{\partial \alpha_s}{\partial Q^2} = \beta(\alpha_s), \quad (1.0.17)$$

The beta function, $\beta(\alpha_s)$, can be expressed perturbatively as an expansion in α_s and is currently known to N³LO.

At one-loop order the solution of this equation is

$$\alpha_s(Q^2) = \frac{\alpha_s(\mu_R^2)}{1 + \beta_0 \alpha_s(\mu_R^2) \ln(\frac{Q^2}{\mu_R^2})}, \quad (1.0.18)$$

where β_0 is the first coefficient of the β expansion. From this solution we can explicitly see asymptotic freedom because α_s decreases as the energy scale increases. We also see the role of the renormalisation scale in specifying a particular reference value for α_s . This solution is not exact because the RGE 1.0.17 only holds to all orders. Any residual μ_R dependence characterises the accuracy of our calculation, because going to higher and higher orders should reduce this dependence, eventually to 0.

Quantities are infra-red safe if they do not depend on long-distance physics. This means we can apply perturbation theory because α_s is small enough in the short-distance regime. Unfortunately at the partonic level structure functions and cross sections are not infra-red safe.

1.0.4 The QCD improved parton model and factorisation

In the naïve parton model, we did not include any interactions involving gluons; their incorporation leads to the QCD improved parton model. The addition of gluons leads to significant complications, owing to the fact that the interacting quarks are free to emit gluons at some stage before detection (remember the detector is at a long-distance so we cannot ignore the long-distance physics). When these gluons are “soft” (low energy) or collinear to one of the partons we run into IR divergences. This situation is equivalent to the internal propagator quark going on-shell, or in other words there is a large time separation between the

partonic interaction and the gluon emission. The observed violation of Björken scaling has its origins in interactions with gluons. In IR-safe observables the soft and collinear divergences exactly cancel [52, 53], but for other cases we need a way of dealing with the disparate short- and long- scale physics.

This is done using the factorisation theorem ??, which allows us to factorise the incalculable long-distance physics into the PDFs, meaning we are able to use perturbative QCD as a predictive theory. The PDFs are then non-perturbative, meaning we must obtain them from experiments, but they are universal quantities and so once determined can be applied everywhere, much like the coupling constants. This process introduces the artificial ‘‘factorisation scale’’, μ_F , in addition to the renormalisation scale. The factorisation scale separates the short- and long- distance physics; loosely, if a parton’s transverse momentum is less than μ_F it is considered part of the hadron and is factored into the PDFs, otherwise it is seen as taking part in the hard scattering process, and will appear in the partonic cross section.

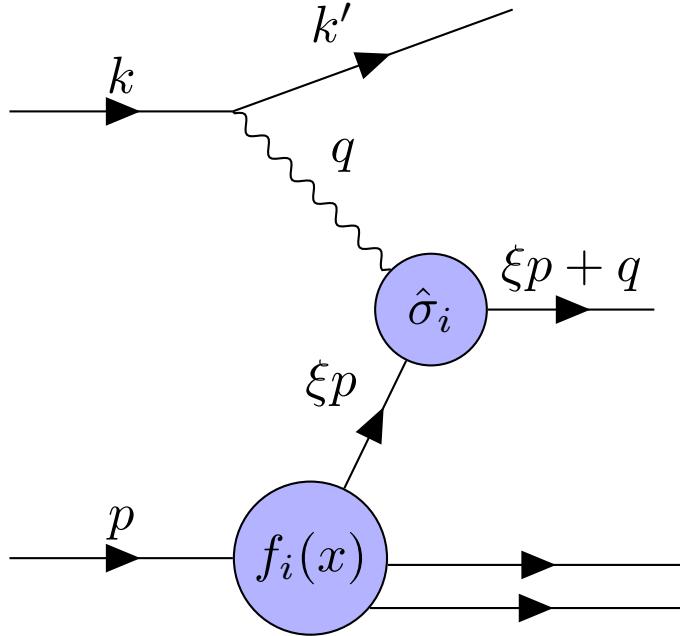


Figure 1.0.3 Factorisation and the QCD improved parton model

We can write a DIS cross section as

$$\sigma^{DIS} = \sum_i \int dx f_i(x, \mu_F^2) \hat{\sigma}_i \left(x, \frac{Q^2}{\mu_F^2} \right), \quad (1.0.19)$$

corresponding to Fig ??.

We can see how this works in practice by considering the case where a quark emits a gluon before interaction with the photon, such as in Fig. ?? . Here the parent parton, with fraction y of the proton's momentum, emits a gluon giving rise to a daughter parton with a fraction z of the parent hadron's momentum. We can see that $z = x/y$.

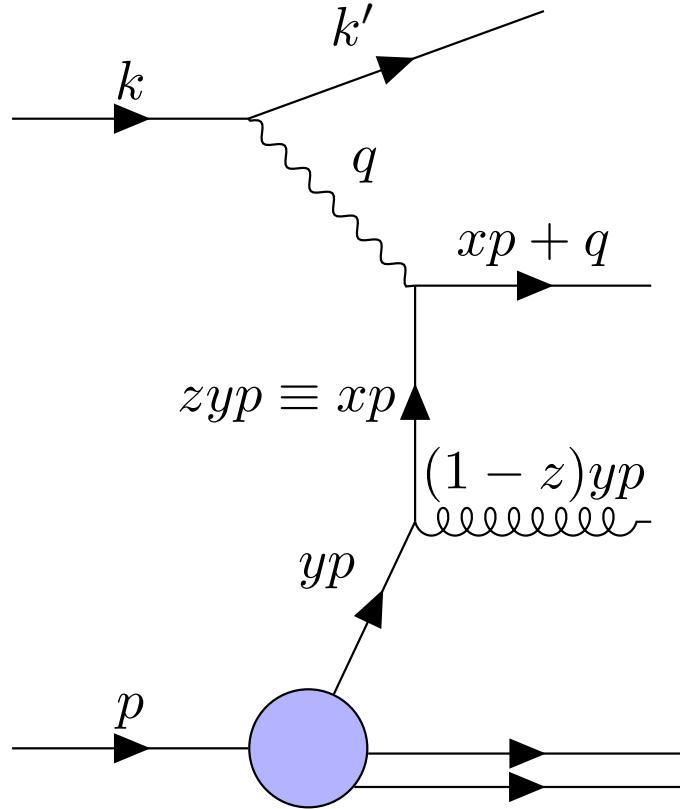


Figure 1.0.4 A quark radiating a gluon before interacting.

It transpires (see Ref. [50] for the derivation) that the structure function F_2 can be expressed as

$$\frac{F_2(x, Q^2)}{x} = \sum_i e_i^2 \int_x^1 \frac{dy}{y} f_i(y) \left[\delta\left(1 - \frac{x}{y}\right) + \frac{\alpha_s}{2\pi} \mathcal{P}_{qq} \left(\frac{x}{y}\right) \ln \left(\frac{Q^2}{m^2}\right) \right]. \quad (1.0.20)$$

m is a cutoff introduced to regularise the collinear divergence and you can see that as $m \rightarrow 0$ the structure function diverges. A divergence also occurs for $(1-z) \rightarrow 0$, and this is a soft divergence because it corresponds to the gluon being emitted

with zero momentum. The quantity \mathcal{P}_{qq} is the quark-quark ‘‘splitting function’’, detailing the probability that a quark emits a gluon leaving a daughter quark with fraction z of the parent’s momentum. In the \overline{MS} renormalisation scheme this has the form

$$\mathcal{P}_{qq} = \frac{4}{3} \left(\frac{1+z^2}{1-z} \right). \quad (1.0.21)$$

We want an expression which is free from the soft and collinear divergences. We can proceed by defining

$$\mathcal{I}_{qq}^i(x) \equiv \frac{\alpha_s}{2\pi} \int_x^1 \frac{dy}{y} f_i(y) \mathcal{P}_{qq} \left(\frac{x}{y} \right), \quad (1.0.22)$$

and separating ?? into a singular part and a calculable part, like

$$\frac{F_2(x, Q^2)}{x} = \sum_i e_i^2 \left[f_i(x) + \mathcal{I}_{qq}^i(x) \ln \left(\frac{\mu_F^2}{m^2} \right) + \mathcal{I}_{qq}^i(x) \ln \left(\frac{Q^2}{\mu_F^2} \right) \right]. \quad (1.0.23)$$

Notice we introduced the artificial factorisation scale, μ_F , to do this. Grouping the singular terms together as

$$f_i(x, \mu_F^2) = f_i(x) + \mathcal{I}_{qq}^i(x) \ln \left(\frac{\mu_F^2}{m^2} \right), \quad (1.0.24)$$

we have factorised the divergences into the PDF $f_i(x)$, giving a new PDF, $f_i(x, \mu_F^2)$, which also depends on μ_F . and noting that at leading order $f_i(y) = f_i(y, \mu_F^2)$, we are able to write

$$\frac{F_2(x, Q^2)}{x} = \sum_i e_i^2 \left[f_i(x, \mu_F^2) + \frac{\alpha_s}{2\pi} \int_x^1 \frac{dy}{y} f_i(y, \mu_F^2) \mathcal{P}_{qq} \left(\frac{x}{y} \right) \ln \left(\frac{Q^2}{\mu_F^2} \right) \right] + \mathcal{O}(\alpha_s^2). \quad (1.0.25)$$

We know that F_2 is an observable quantity and thus should be independent of μ_F , leading to a RGE:

$$\begin{aligned} \frac{1}{e_i^2 x} \frac{\partial F_2(x, Q^2)}{\partial \ln \mu_F^2} &= \frac{\partial f_i(x, \mu_F^2)}{\partial \ln \mu_F^2} \\ &+ \frac{\alpha_s}{2\pi} \int_x^1 \frac{dy}{y} \left(\frac{\partial f_i(y, \mu_F^2)}{\partial \ln \mu_F^2} \ln \left(\frac{Q^2}{\mu_F^2} \right) - f_i(y, \mu_F^2) \right) \mathcal{P}_{qq} \left(\frac{x}{y} \right) \\ &= 0. \end{aligned} \quad (1.0.26)$$

This can be further simplified by noting that $\frac{\partial f_i(y, \mu_F^2)}{\partial \ln \mu_F^2}$ is of $\mathcal{O}(\alpha_s^2)$, and so

$$\frac{\partial f_i(x, \mu_F^2)}{\partial \ln \mu_F^2} = \frac{\alpha_s}{2\pi} \int_x^1 \frac{dy}{y} f_i(y, \mu_F^2) \mathcal{P}_{qq} \left(\frac{x}{y} \right). \quad (1.0.27)$$

This equation describes the evolution of the newly defined PDFs with scale, a product of the factorisation of the divergences into them. In practice this equation is solved numerically.

When we also include the gluon as a parton, we open ourselves up to more splitting possibilities (e.g. gluon \rightarrow quark and gluon \rightarrow gluon), and this result generalises to a set of coupled differential equations known as the DGLAP equations [21, 34, 48]:

$$\frac{\partial f_i}{\partial \ln \mu_F^2} = \sum_i \frac{\alpha_s}{2\pi} \mathcal{P}_{ij} \otimes f_j, \quad (1.0.28)$$

where we have used the Mellin convolution, defined

$$\mathcal{P} \otimes f \equiv \int_x^1 \frac{dy}{y} \mathcal{P} \left(\frac{x}{y} \right) f(y, \mu_F^2), \quad (1.0.29)$$

and the index i runs from $-n_f$ to n_f (where n_f is the number of flavours), with the negative indices referring to the antiquarks, 0 to the gluon and the positive ones to the quarks.

1.0.5 Hadroproduction

At the LHC most processes involve the interaction of two protons. Hadron-hadron collisions can be approached in much the same way as DIS, but instead the process is like in Fig. ???. Because two protons are involved the expression for the cross section is the natural extension of the DIS case (??):

$$\sigma = \sum_{i,j} \int dx_1 dx_2 f_i(x_1, \mu_F^2) f_j(x_2, \mu_F^2) \hat{\sigma}_{ij} \left(x_1, x_2, \frac{Q^2}{\mu_F^2}, \dots \right). \quad (1.0.30)$$

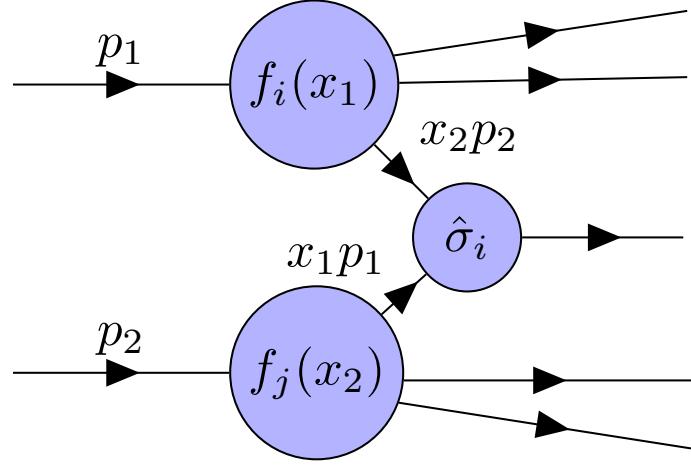


Figure 1.0.5 Factorisation in hadron-hadron collisions.

Write about higher order corrections and factorisation!

1.0.6 Sum rules

Although PDFs may seem at first sight to be totally unknown there are some theoretical observations which we can use to constrain their form. These are known as the “sum rules” [?]. Intuitively, adding up all the momenta of the partons must equal the momentum of the proton. This enforces the condition

$$\int_0^1 dx \sum_i x f_i(x, Q^2) = 1. \quad (1.0.31)$$

The other thing we know about the proton is that it is made up of two up and one down (and no strange) “valence” quarks. Any other quarks must be pair-produced from the sea, and therefore come with an antiquark of the same flavour. So we can normalise the PDFs using the expressions:

$$\int_0^1 dx (f_u - f_{\bar{u}}) = 2; \quad (1.0.32a)$$

$$\int_0^1 dx (f_d - f_{\bar{d}}) = 1; \quad (1.0.32b)$$

$$\int_0^1 dx(f_q - f_{\bar{q}}) = 0, \quad q = s, c, t, b. \quad (1.0.32c)$$

Note that these conditions require that the PDFs are integrable.

1.1 Determining PDFs

In this section we review the necessary background for PDF determination within the NNPDF [1] framework. First we touch on the experimental and theoretical inputs to PDF fits, then we summarise the NNPDF fitting strategy, and finally we detail information on neural networks specific to this context.

1.1.1 Experimental and theoretical input

NNPDF uses a variety of experimental data from a number of particle colliders, including those based at CERN [2] and Fermilab [3]. These are observables such as cross sections, differential cross sections and structure functions. Fig. 1.1.1 is a plot of the (x, Q^2) range spanned by the datasets in the latest NNPDF3.1 [16] release. The majority of the data are from DIS processes, which are crucial in determining PDF functional form, but in recent years increasingly more LHC collider data has been added including $t\bar{t}$ production, high energy jets and single top production [54]. For a full review of the data, see Ref. ??.

Theoretical predictions of the corresponding parton-level observables are computed using external codes such as MCFM [31], aMC@NLO [22], DYNNLO [17], FEWZ [43] and NLOjet++ [56]. These are converted to higher orders of perturbation theory as necessary using QCD and electroweak correction (“ k ”) factors. They are then combined with DGLAP evolution kernels, which evolve PDFs from an initial reference energy scale to the energy scale of each experiment using the DGLAP equations (Eqn. 1.0.28).

1.1.2 Experimental uncertainties

Experimental uncertainties are described using a covariance matrix, C_{ij} , which gives the uncertainties and correlations between each of the data points $i, j =$

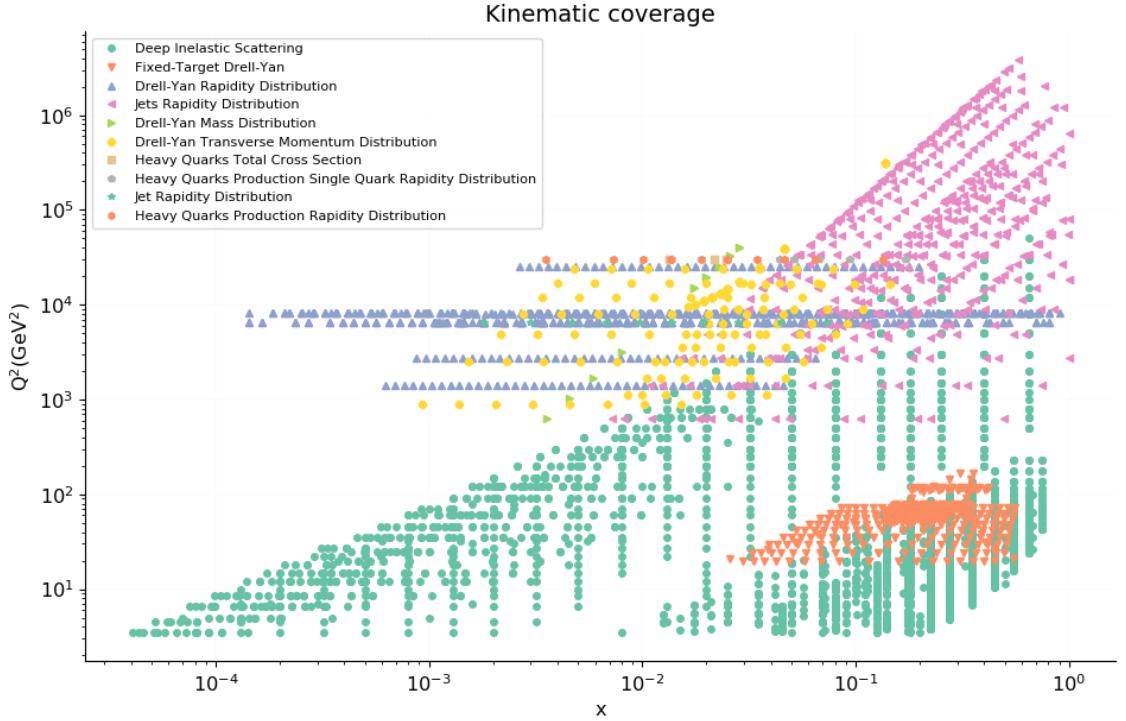


Figure 1.1.1 Plot of the (x, Q^2) range spanned by data included in the latest NNPDF3.1 NLO fit.

$1, \dots, N_{dat}$. It encapsulates the total breakdown of errors, σ , and can be constructed using uncorrelated errors (σ_i^{uncorr}), and additive ($\sigma_{i,a}$) and multiplicative ($\sigma_{i,m}$) correlated systematic errors (more on these below):

$$C_{ij} = \delta_{ij} \sigma_i^{uncorr} \sigma_j^{uncorr} + \sum_a^{add.} \sigma_{i,a} \sigma_{j,a} + \left(\sum_m^{mult.} \sigma_{i,m} \sigma_{j,m} \right) D_i D_j, \quad (1.1.1)$$

where D_i are the experimental data values.

Structurally, the uncorrelated statistical uncertainties appear down the diagonal and these are what we would recognise intuitively as the statistical error “on a data point”. However, correlated systematic uncertainties can also appear on the off-diagonals. Correlated uncertainties include those which link multiple data points, for example systematic uncertainties from a particular detector which will affect all of its data in a similar way.

Systematic uncertainties further divide into two types, “additive” and “multiplicative”. Additive systematics are perhaps a more familiar type of error, and are independent of the datapoint values themselves. On the other hand, multiplicative systematics depend on the measured values. In the context of

particle physics experiments, a common example is total detector luminosity. This is because recorded cross sections are dependent on the luminosity of the detector; a higher luminosity means more collisions will take place so the measured cross section will be greater.

Fig. 1.1.2 is an example of an experimental covariance matrix for data included in an NNPDF fit. The data are grouped according to what type of process the interaction belongs to (DIS charged current (CC) and neutral current (NC), Drell-Yan (DY), jets and top production). Systematic correlations within experiments are responsible for off-diagonal contributions, and these are mostly positive correlations but there is some anticorrelated behaviour in DIS CC, as a result of data in different kinematic regimes.

The covariance matrix can be used to define the χ^2 figure of merit,

$$\chi^2 = \frac{1}{N_{dat}}(D_i - T_i)C_{ij}^{-1}(D_j - T_j), \quad (1.1.2)$$

which measures the goodness of fit between the experimental data D_i with associated error breakdown C_{ij} , and theory predictions T_i . In practice, this definition is subject to d'Agostini bias [33] due to the presence of normalisation uncertainties. To avoid this, NNPDF employ the iterative *t0* procedure [27] whereby D_i in Eqn. ?? are replaced initially with the predictions from a baseline fit, and the covariance matrix is iterated concurrently with preprocessing.

1.1.3 NNPDF fitting strategy

There are a number of groups currently active in carrying out PDF fits including MSTW [8], CTEQ [18], NNPDF [1], HERAPDF/xFitter [32] and ABM [55].

The work in this thesis has been carried out in the framework developed by the NNPDF collaboration, so we will concentrate on this fitting strategy. There are two main features which differ from other fitting collaborations' [19]. These are:

1. The use of Monte Carlo approach to error analysis;
2. Fitting using artificial neural networks.

In the following sections we will provide an overview of these aspects, which can be found in more detail in Refs. [16? ?].

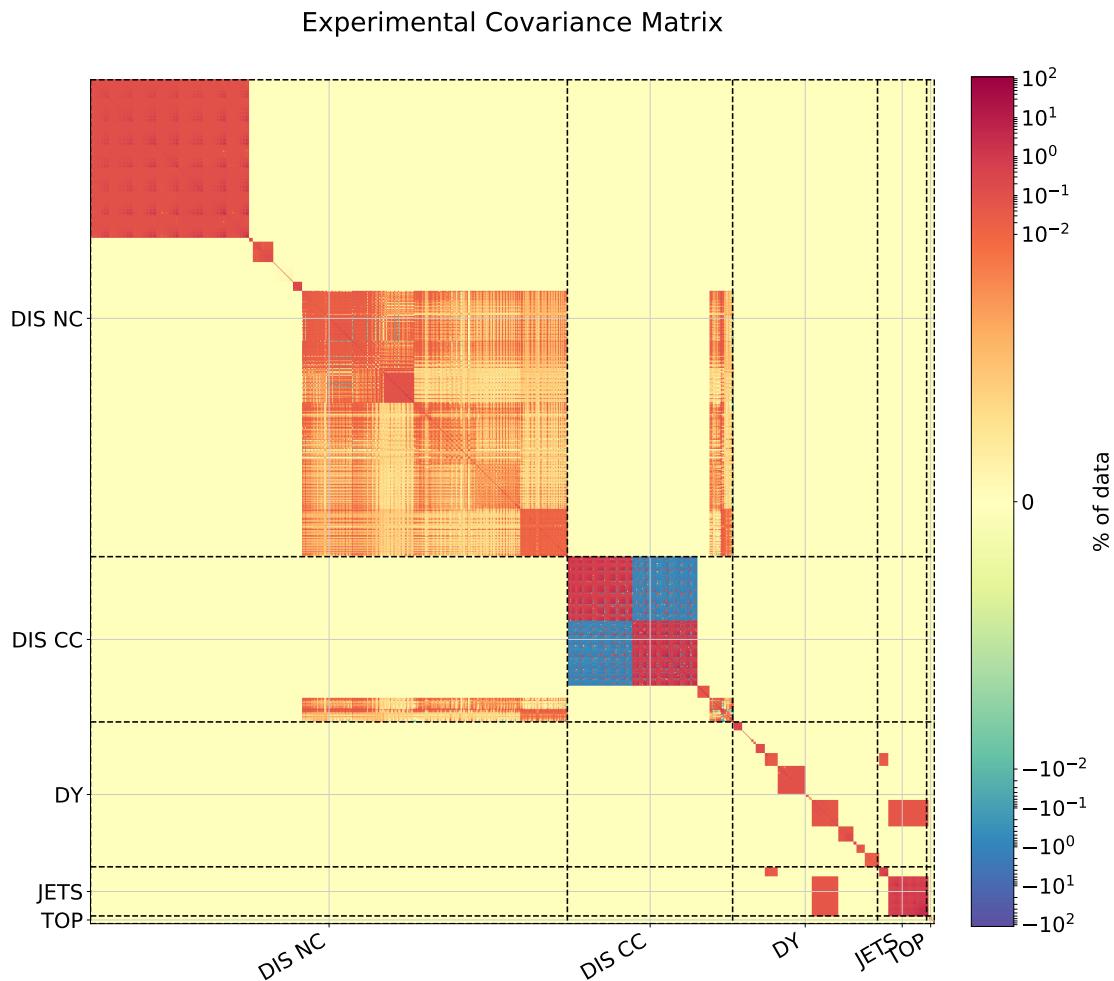


Figure 1.1.2 An example of an experimental covariance matrix for data included in an NNPDF fit. The data are grouped according to what type of process the interaction belongs to (DIS charged current (CC) and neutral current (NC), Drell-Yan (DY), jets and top production).

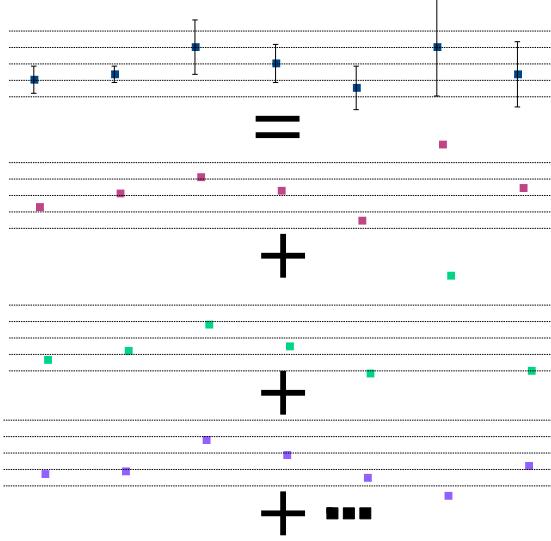


Figure 1.1.3 Schematic of the generation of Monte Carlo replicas of pseudodata from data with uncertainties.

1.1.4 Monte Carlo approach

The uncertainties in the functional form of PDFs come as a direct consequence of the uncertainties in the experimental and theoretical input. In order to propagate experimental uncertainties through to the PDFs, NNPDF represent the experimental data (central values and uncertainty distribution) as a Monte Carlo ensemble. This is a set of N_{rep} Monte Carlo “replicas” which, given high enough replica number, have a mean value equal to the data central value and covariance equal to the experimental covariance. Fig. 1.1.3 is a schematic illustrating the generation of these “pseudodata”, $D^{(k)}$, $k = 1, \dots, N_{rep}$. They are generated using Gaussian random numbers $n_a^{(k)}$ and $\hat{n}_p^{(k)}$:

$$D^{(k)} = (D^0 + \sum_a n_a^{(k)} \sigma^a) \prod_m (1 + \hat{n}_m^{(k)} \sigma^p), \quad (1.1.3)$$

where D_0 is the (symmetrised) experimental data value, and σ^a and σ^p are the additive and multiplicative uncertainties discussed in Sec. 1.1.2. Explicitly, the pseudodata replicas satisfy the relations:

$$\langle D_i^{(k)} \rangle = D_i^0; \quad (\langle D_i^{(k)} \rangle - D_i^0)(\langle D_j^{(k)} \rangle - D_j^0) = C_{ij}, \quad (1.1.4)$$

where the notation $\langle \cdot \rangle$ denotes the mean over replicas. Fig. 1.1.4 shows the distribution of pseudodata for a single data point.

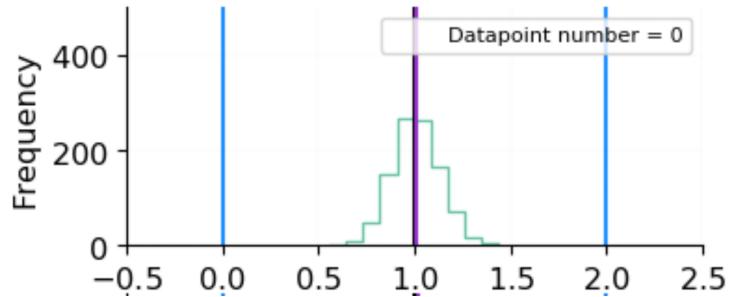


Figure 1.1.4 Histogram of distribution of 100 pseudodata replicas for a single data point, normalised to D^0 . The purple line is the mean value $\langle D^{(k)} \rangle$, which is equal to D^0 to arbitrary precision.

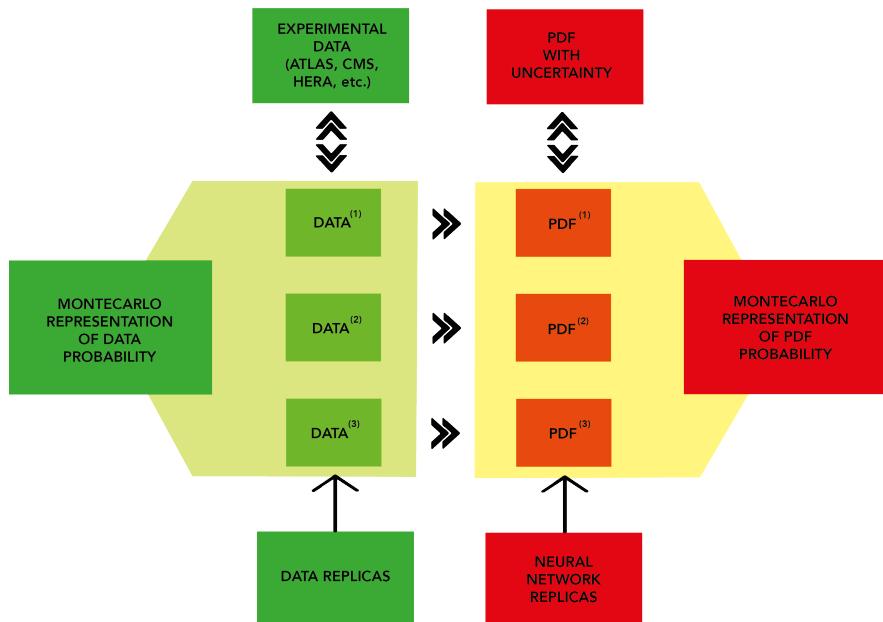


Figure 1.1.5 NNPDF general strategy.

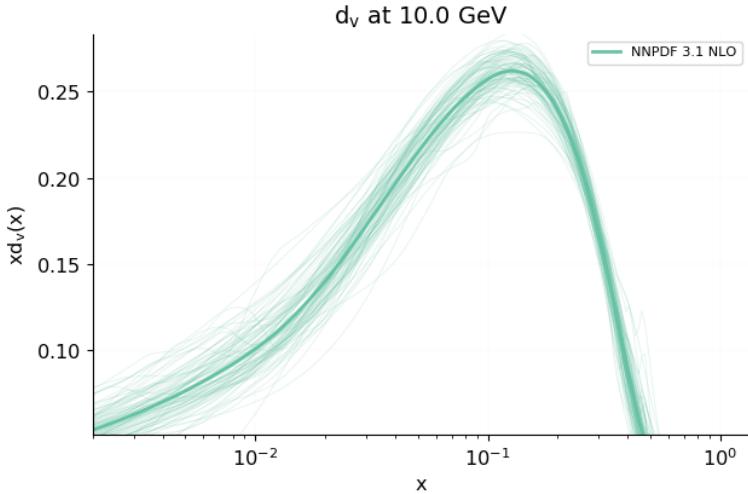


Figure 1.1.6 *Monte Carlo replicas for the down valence quark PDF NNPDF3.1 at NLO.*

Once the pseudodata have been generated, each of these ($D^{(k)}$) is fitted separately to the theoretical predictions by minimising a target error function based on the χ^2 (Eqn. 1.1.2), resulting in a PDF set of each flavour, $f_q^{(k)}$ (where q runs over the fitted flavours: $g, u, d, s, c, \bar{u}, \bar{d}, \bar{s}, \bar{c}$). These act as a Monte Carlo parametrisation of the PDFs (for example, Fig. 1.1.6). This means that the PDFs and their errors can be extracted by taking the means and standard deviations over the ensemble. The final PDFs are made publicly available as downloadable files on the LHAPDF website [7?].

1.1.5 Neural Networks

Inspired by how the brain processes information, in machine learning neural networks are a graph of connected nodes. They are trained by example, so have the capability to learn a PDF’s functional form given a set of data. The use of neural networks rather than specific functional forms allows us to avoid the theoretical bias which goes into selecting such a functional form. The layout, or “architecture”, consists of input layers, hidden layers and output layers. Nodes can be either input nodes or activation nodes, the latter of which have an associated activation function which is applied to their output. Fig. ?? depicts the architecture currently used by NNPDF. This is a “2-5-3-1” architecture, where the numbers refer to the number of nodes in each layer. It is a “multilayer perceptron”, meaning the graph is fully connected, and it is a feed-forward; information can only be passed in one direction through the layers (from input

to output). The two inputs are x and $\ln(1/x)$, and the output, f , is the PDF at the parametrisation scale, Q_0 . In this network the output of a node in the l^{th} layer is given by

$$\xi_i^{(l)} = g \left(\sum_j^{inputs} \omega_{ij}^{(l)} \xi_j^{(l-1)} + \theta_i^{(l)} \right) \quad (1.1.5)$$

where the ω s and θ s are “weights” and “thresholds”; parameters to be minimised with respect to. g is an “activation function” which is set to

$$g(z) = \begin{cases} \frac{1}{1+e^{-z}} & \text{for hidden layers} \\ a & \text{for output layer.} \end{cases} \quad (1.1.6)$$

The choice of sigmoid activation function for the hidden layers allows sufficient non-linear freedom in the functional form, and the linear activation function for the output layer ensures the range of the PDFs is not restricted to [0,1].

Input layer	Hidden layer 1	Hidden layer 2	Output layer
----------------	-------------------	-------------------	-----------------

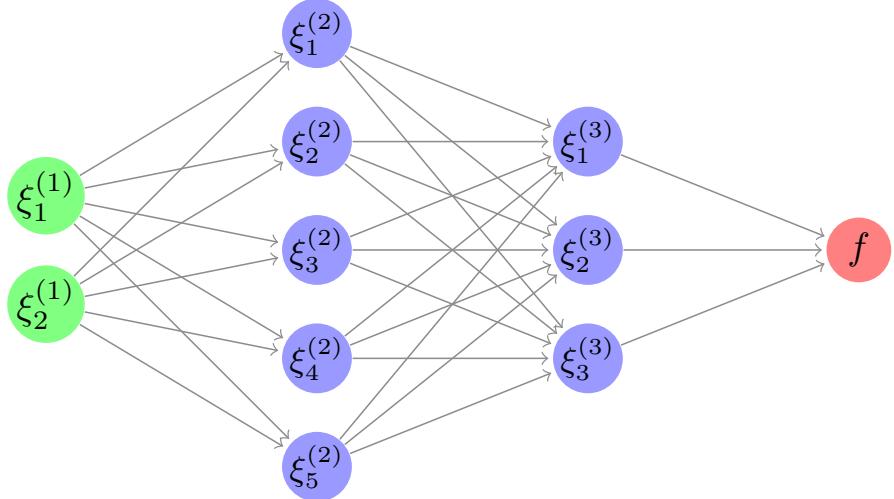


Figure 1.1.7 Schematic depiction of the 2-5-3-1 architecture of an artificial neural network currently used by NNPDF. In the NNPDF methodology $\xi_1^{(1)}$ and $\xi_2^{(1)}$ are the variables x and $\log x$ respectively.

The training of the neural networks is implemented using a “genetic algorithm” (CMA-ES), so-called because of the introduction of mutation to the fitting parameters. This additional degree of randomness helps to avoid getting stuck in local minima. In practice, this involves “mutating” some chosen fraction of the

thresholds, θ , by perturbing them at random.

1.1.6 Parametrisation, preprocessing and postprocessing

A scale of $Q = 1.65 \text{ GeV}$ is chosen to parametrise the PDFs at, and then they can be determined at any other scale by evolution using the DGLAP equations (Eqn. 1.0.28). The PDFs are fitted parametrised in a “fitting basis”, to help convergence [?], defined:

- g ;
- $\Sigma \equiv \sum_{u,d,s} q_i + \bar{q}_i$;
- $T_3 \equiv u - d$;
- $T_8 \equiv u + d - 2s$;
- $V \equiv \sum_{u,d,s} q_i - \bar{q}_i$;
- $V_3 \equiv \bar{u} - \bar{d}$;
- $V_8 \equiv \bar{u} - \bar{d} - 2\bar{s}$;
- c .

Since the form of the neural networks ($N_i(x)$) is determined by training on experimental data, the output is not meaningful outwith the data region. The functional form of the PDFs in this so-called “extrapolation region” is in practice fixed through enforcement of the known high and low x behaviour via “preprocessing”; the PDFs are parametrised as:

$$f_i(x) = A_i x^{-\alpha_i} (1 - x)^{\beta_i} N_i(x). \quad (1.1.7)$$

A_i are normalisation coefficients set by the sum rules and fixed at each iteration of the fit. The powers α_i and β_i are fitted parameters determined by iteration from one fit to the next. This preprocessing has the effect that the PDFs approach 0 at large x , and generally grow at small x . This is because the probability of the existence of a parton is generally small at high x and larger with decreasing x outwith the data region.

Postprocessing is also applied to the PDF replicas to remove those which don't satisfy certain quality conditions. That is, where the target error function or arc-length of the replica is more than four standard deviations outwith the mean, or where the positivity of the resulting cross-sections is not satisfactorily maintained.

1.1.7 Cross validation

Neural networks are effective at learning the functional form underlying data. Sometimes they can be “too effective”, picking up not just the underlying law but also the noise. This is known as “overlearning” (see Fig. ?? for an example).

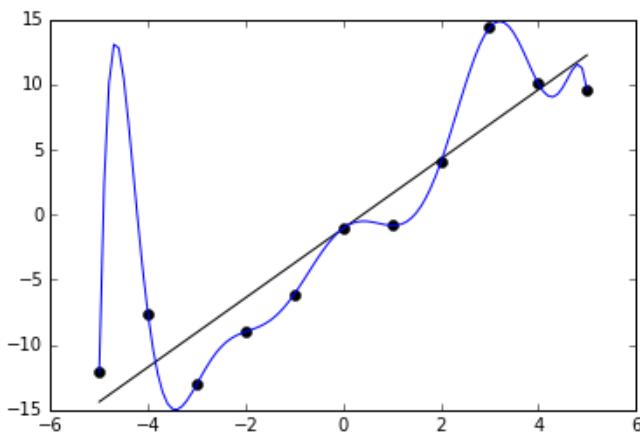


Figure 1.1.8 Overlearning: the data points (black dots) fluctuate around the linear underlying law (black line), but the neural network continues to minimise the error function until it passes through every data point (blue curve), fitting the noise in the data.

To circumvent this problem, the data is split into a training and a validation set. The training data is used to optimise the neural network, and the validation data is used to test the network output, in a process known as “cross validation”. As training epochs elapse, the target error function compared to both the training and validation data should decrease as the network learns the underlying law. At some point, however, the network will begin to learn the noise in the training data, at which point the training error function will continue to decrease, but the validation error function will stop decreasing and start to increase again. The optimum fit is determined using the “lookback” method, where after training the model corresponding to the minimum in the validation error function is selected.

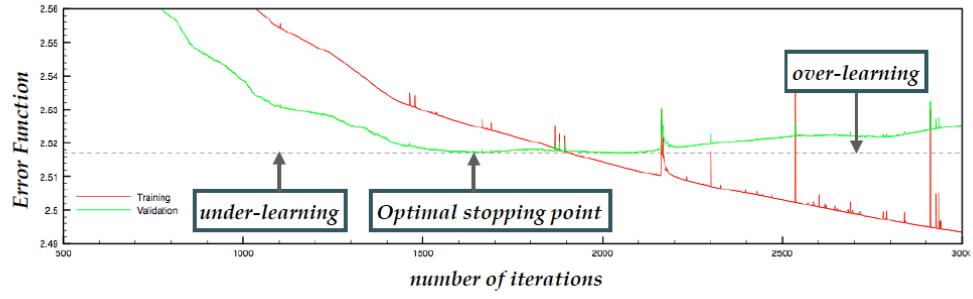


Figure 1.1.9 Cross validation with the lookback method.

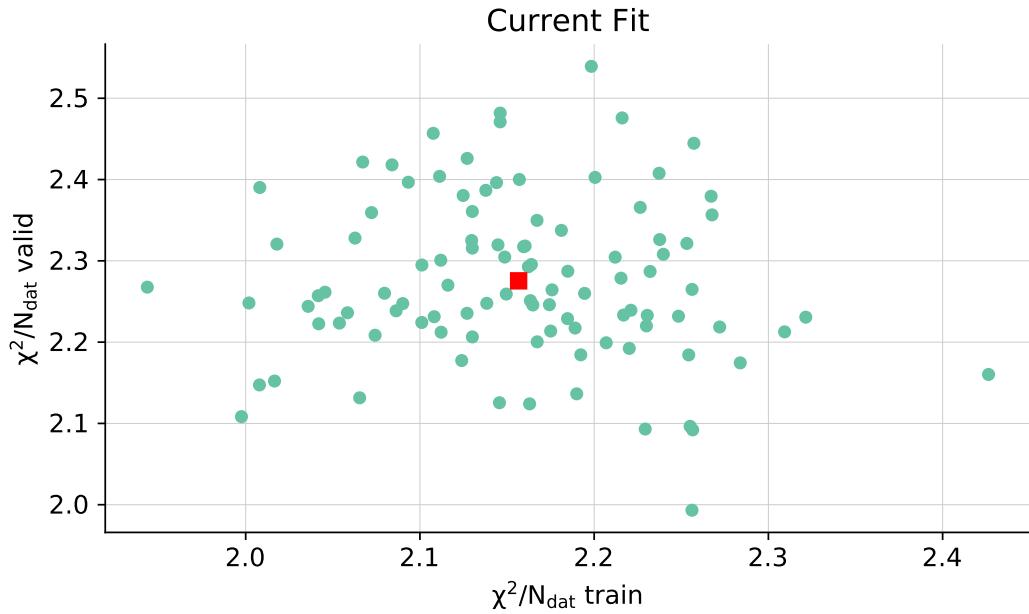


Figure 1.1.10 Comparing the training and validation χ^2 s for the 100 replicas (green circles) of a PDF fit. The red square gives the average.

Chapter 2

Theory uncertainties in PDFs - 10%

- What are theory uncertainties?
- Why are they now important?
- Types of unc - see below
- Bayesian interpretation of these uncertainties - there is one "true" value e.g. for the higher order value, so need to estimate uncertainty bc will never know the size of e.g. MHOU unless you go ahead and calc - maybe ref d'Agostini paper on Bayesian interpretation
- Will use Bayesian framework and assume Gaussianity of the expected true value of theory calc
- Show C+S in fit - plus sign because exp and th unc are independent so combine errors in quadrature. They are also on an equal footing in terms of their effect on the PDFs
- When many datasets/global fit, can have v strong theory correlations even across different experiments, because the underlying theory connects them

2.1 Fitting PDFs including theory uncertainties

Historically, experimental uncertainties have been the dominant source of error in PDF fits. In the NNPDF framework both replica generation and computation of

χ^2 are currently based entirely on these. We must now try to match the ongoing drive to increase experimental precision by including errors introduced at the theoretical level. This is especially important given recent data sets such as the Z boson transverse momentum distributions [10] [20] [42], which have very high experimental precision. Without the inclusion of theoretical errors, this has led to tension with the other datasets.

In future NNPDF fits theoretical uncertainties will be included following a procedure outlined by Ball & Desphande [24]. This hinges on a result from Bayesian statistics which applies to Gaussian errors. Namely, theory uncertainties can be included by directly adding a theoretical covariance matrix to the experimental covariance matrix prior to the fitting. A brief summary of the derivation is given below.

When determining PDFs we incorporate information from experiments in the form of N_{dat} experimental data points D_i , $i = 1, \dots, N_{dat}$. The associated uncertainties and their correlations are encapsulated in an experimental covariance matrix C_{ij} . Parts of the matrix which associate two independent experiments will be populated by zeros. However we would expect there to be correlations between data points from the same detector, for example.

Each data point is a measurement of some fundamental “true” value, \mathcal{T}_i , dictated by the underlying physics. In order to make use of the data in a Bayesian framework, we assume that the experimental values follow a Gaussian distribution about the unknown \mathcal{T} . Then, assuming the same prior for D and \mathcal{T} , we can write an expression for the conditional probability of \mathcal{T} given the known data D :

$$P(\mathcal{T}|D) = P(D|\mathcal{T}) \propto \exp\left(-\frac{1}{2}(\mathcal{T}_i - D_i)C_{ij}^{-1}(\mathcal{T}_j - D_j)\right). \quad (2.1.1)$$

However, in a PDF fit we cannot fit to the unknown true values \mathcal{T} , and must make do with predictions based on current theory T_i . This is the origin of theory uncertainties in PDF fits; where our theory is incomplete, fails to describe the physics well enough, or where approximations are made, we will introduce all kinds of subtle biases into the PDF fit. The theory predictions themselves also depend on PDFs, so uncertainties already present in the PDFs are propagated through. This, in particular, leads to a high level of correlation because the PDFs are universal, and shared between all the theory predictions.

We can take a similar approach when writing an expression for the conditional

probability of the true values \mathcal{T} given the available theory predictions T , by assuming that the true values are Gaussianly distributed about the theory predictions.

$$P(\mathcal{T}|T) = P(T|\mathcal{T}) \propto \exp\left(-\frac{1}{2}(\mathcal{T}_i - T_i)S_{ij}^{-1}(\mathcal{T}_j - T_j)\right), \quad (2.1.2)$$

where S_{ij} is a “theory covariance matrix” encapsulating the magnitude and correlation of the various theory errors. We will need to do some work to determine S_{ij} for the different sources of error, and this will be outlined in detail in the following chapters.

When we fit PDFs we aim to maximise the probability that a PDF-dependent theory is true given the experimental data available. This amounts to maximising $P(T|D)$, marginalised over the unknown true values \mathcal{T} . To make this more useful for fitting purposes, we can relate this to $P(D|T)$ using Bayes’ Theorem:

$$P(D|T)P(\mathcal{T}|DT) = P(\mathcal{T}|T)P(D|\mathcal{T}T), \quad (2.1.3)$$

where we note that the experimental data D do not depend on our modelled values T , so $P(D|\mathcal{T}T) = P(D|\mathcal{T})$. So we can integrate Bayes’ Theorem over the possible values of the N -dimensional true values \mathcal{T} :

$$\int D^N \mathcal{T} P(D|T) P(\mathcal{T}|DT) = \int D^N \mathcal{T} P(\mathcal{T}|T) P(D|\mathcal{T}), \quad (2.1.4)$$

and, because $\int D^N \mathcal{T} P(\mathcal{T}|TD) = 1$ as all possible probabilities for the true values must sum to one,

$$P(D|T) = \int D^N \mathcal{T} P(\mathcal{T}|T) P(D|\mathcal{T}). \quad (2.1.5)$$

We can always write the theory predictions T in terms of their shifts Δ relative the true values \mathcal{T} :

$$\Delta_i \equiv \mathcal{T}_i - T_i. \quad (2.1.6)$$

These shifts quantify the accuracy of the theoretical predictions, and can be thought of as nuisance parameters in the PDF fit. We can express the above integral in terms of the shifts Δ_i , making use of the assumptions of Gaussianity

in Eqns. ?? and ??:

$$P(D|T) \propto \int D^N \Delta \exp \left(-\frac{1}{2}(D_i - T_i - \Delta_i)^2 - \frac{1}{2}C_{ij}^{-1}(D_j - T_j - \Delta_j)^2 - \frac{1}{2}\Delta_i S_{ij}^{-1}\Delta_j \right). \quad (2.1.7)$$

To evaluate the Gaussian integrals, consider the exponent: switching to a vector notation for the time being, we can expand this out and then complete the square, making use of the symmetry of S and C :

$$\begin{aligned} & (D - T - \Delta)^T C^{-1} (D - T - \Delta) + \Delta^T S^{-1} \Delta \\ &= D^T (C^{-1} + S^{-1}) \Delta - \Delta^T C^{-1} (D - T) - (D - T)^T C^{-1} \Delta + (D - T)^T C^{-1} (D - T) \\ &= (\Delta - (C^{-1} + S^{-1})^{-1} C^{-1} (D - T))^T (C^{-1} + S^{-1}) \\ &\quad \times (\Delta - (C^{-1} + S^{-1})^{-1} C^{-1} (D - T)) \\ &\quad - (D - T)^T C^{-1} (C^{-1} + S^{-1})^{-1} C^{-1} (D - T) + (D - T)^T C^{-1} (D - T). \end{aligned} \quad (2.1.8)$$

Now, integrating Eqn. ?? over Δ leads to a constant from the Gaussian integrals, which we can absorb, and only the parts of the exponent without Δ remain:

$$P(T|D) = P(D|T) \propto \exp \left(-\frac{1}{2}(D - T)^T (C^{-1} - C^{-1} (C^{-1} + S^{-1})^{-1} C^{-1}) (D - T) \right). \quad (2.1.9)$$

We can further simplify this by noting that

$$\begin{aligned} (C^{-1} + S^{-1})^{-1} &= (C^{-1} (C + S) S^{-1})^{-1} \\ &= S (C + S)^{-1} C, \end{aligned} \quad (2.1.10)$$

which means we can rewrite

$$\begin{aligned} C^{-1} - C^{-1} (C^{-1} + S^{-1})^{-1} C^{-1} &= C^{-1} - C^{-1} S (C + S)^{-1} \\ &= (C^{-1} (C + S) - C^{-1} S) (C + S)^{-1} \\ &= (C + S)^{-1}. \end{aligned} \quad (2.1.11)$$

Finally, with indices restored we are left with

$$P(T|D) \propto \exp \left(-\frac{1}{2}(D_i - T_i)(C + S)_{ij}^{-1}(D_j - T_j) \right). \quad (2.1.12)$$

Use the expression for conditional probability $P(X \cap Y) = P(X|Y)P(Y)$ and integrate over all possible values of the true theory T (as this is an unknown):

$$\begin{aligned} \int dT P(T|y \cap f)P(y|f) &= \int dT P(y|T \cap f)P(T|f) \\ P(y|f) &= \int dT P(y|T \cap f)P(T|f). \end{aligned} \quad (2.1.13)$$

Now assume Gaussian uncertainties for data and theory, of the form $\exp(-\frac{1}{2}\chi^2)$

$$P(y|Tf) \propto \exp\left(-\frac{1}{2}(y - T)^T \sigma^{-1}(y - T)\right) \quad (2.1.14)$$

$$P(T|f) \propto \exp\left(-\frac{1}{2}(T - T[f])^T s^{-1}(T - T[f])\right) \quad (2.1.15)$$

and substitute these into Eq. ?? to get

$$P(y|f) \propto \int dT \exp\left(-\frac{1}{2}\left[(y - T)^T \sigma^{-1}(y - T) + (T - T[f])^T s^{-1}(T - T[f])\right]\right). \quad (2.1.16)$$

Note that the difference between the full theory T and the theory predictions $T[f]$ defines the total correction, C . Therefore the substitution $T = T[f] + C \Rightarrow dT = dT[f] + dC$ can be made, noting that $dT[f] = 0$ because $T[f]$ is the fixed output of NNPDF analysis. The overall expression then becomes

$$P(y|f) \propto \int dC \exp\left(-\frac{1}{2}\left[(y - T[f] - C)^T \sigma^{-1}(y - T[f] - C) + C^T s^{-1}C\right]\right) \quad (2.1.17)$$

which can be evaluated by Gaussian integration over shifted variables, leading to

$$P(y|f) \propto \exp\left(-\frac{1}{2}(y - T[f])^T (\sigma + s)^{-1}(y - T[f])\right). \quad (2.1.18)$$

The final result is that you can treat theoretical errors in exactly the same way as you treat experimental errors.

2.2 Sources of Theoretical Uncertainties

The next step is to estimate the theory covariance matrix, s . This can include a number of different theoretical uncertainties that may appear in PDF fits, such as:

1. **Statistical uncertainties** such as from Monte Carlo generators. These provide diagonal entries to s .
2. **Systematic uncertainties.** These are trickier, and can be estimated by varying some fit parameter ξ from its value at the central prediction, ξ_0 , and applying

$$s_{ij} = \langle (T[f; \xi] - T[f; \xi_0])_i (T[f; \xi] - T[f; \xi_0])_j \rangle \quad (2.2.1)$$

where the angled brackets denote the averaging over a given range of ξ according to some prescription.

Including systematic uncertainties will pose the biggest challenge. We need to identify the places they are being introduced and then make a suitable choice of ξ . Examples of systematic uncertainties we have begun to address are:

- **Missing higher order uncertainties (MHOUs).** These are a result of calculations being done only up to a certain perturbative order in the expansion of α_s . As discussed in more detail below, we can get a handle on these by varying the artificial renormalisation (μ_R) and factorisation (μ_F) scales introduced in the calculation. Here ξ can be thought of as a vector (μ_R, μ_F) .
- **Nuclear and deuteron corrections.** Here data are taken from nuclear targets but the theoretical treatment does not account for this. We can re-calculate the observables under different nuclear models or parametrisations. In this case ξ indexes the model or parametrisation used. The use of multiple models helps to remove any systematic bias introduced by each individual one.

2.3 Renormalisation group invariance

2.4 Scale variation in partonic cross-sections

2.5 Scale variation in PDF evolution

2.6 Double scale variations

2.7 Multiple scale variations

2.8 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 [? ?], theoretical uncertainties for hadron collider processes are dominated by the missing higher order uncertainty (MHOU) 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 [12]: 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 [?], 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 MHOU 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 [16], PDF uncertainties at the electroweak scale can be as low as 1%. Given that the typical size of MHOU on NNLO QCD processes is at the percent level (see e.g. [?]) their neglect seems difficult to justify a priori.

Besides contributing to the overall size of PDF uncertainty, more subtly the MHOU 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 MHOU 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 phenomenological 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 MHOU 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 MHOU is 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 in order to obtain predictions for the physical processes that enter the PDF fit is incompletely known. These include higher twists [25] and other power-suppressed corrections, nuclear corrections when nuclear targets are involved [?], 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 [24]. 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 MHOU, 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 [? ? ?], 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 MHOU for any process by combining two independent scale variations: one to estimate the MHOU in the perturbative evolution of the PDFs (missing higher orders in the DGLAP splitting functions), and the other to estimate the MHOU 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 MHOU 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 MHOU 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 MHOU 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. 2.2, 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 MHOU 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. On 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 [?]. 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 following: in Sect. 2.2 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. 2.3 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. 2.4 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. 2.5 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. 2.6. 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. For the DIS-only PDF determination, we also determine a NNLO PDF set including MHOU. In Sect. 4.5 we present initial studies of the phenomenological impact of the inclusion of MHOUs in PDFs for representative LHC processes. Finally in Sect. 2.8 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 ?? we provide additional details concerning the procedure adopted to diagonalise the theory covariance matrix. Then in Appendix ?? we study another possible validation of the results of Sect. 2.6, 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. [?], of which this paper represents the extended companion.

2.9 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

$$P(\mathcal{T}|D) = P(D|\mathcal{T}) \propto \exp(-\frac{1}{2}(\mathcal{T}_i - D_i)C_{ij}^{-1}(\mathcal{T}_j - D_j)), \quad (2.9.1)$$

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

$$P(\mathcal{T}|T) = P(T|\mathcal{T}) \propto \exp\left(-\frac{1}{2}(\mathcal{T}_i - T_i)S_{ij}^{-1}(\mathcal{T}_j - T_j)\right), \quad (2.9.2)$$

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

$$P(\mathcal{T}|DT)P(D|T) = P(D|\mathcal{T}T)P(\mathcal{T}|T). \quad (2.9.3)$$

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

$$P(D|\mathcal{T}T) = P(D|\mathcal{T}). \quad (2.9.4)$$

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

$$P(D|T) = \int D^N \mathcal{T} P(\mathcal{T}|D)P(\mathcal{T}|T), \quad (2.9.5)$$

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

$$\Delta_i \equiv \mathcal{T}_i - T_i, \quad (2.9.6)$$

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 Eqns. (2.2.1) and (2.2.2), Eq. (2.2.5) becomes that

$$P(D|T) \propto \int D^N \Delta \exp\left(-\frac{1}{2}(D_i - T_i - \Delta_i)C_{ij}^{-1}(D_j - T_j - \Delta_j) - \frac{1}{2}\Delta_i S_{ij}^{-1}\Delta_j\right). \quad (2.9.7)$$

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

$$\begin{aligned}
& (D - T - \Delta)^T C^{-1} (D - T - \Delta) + \Delta^T S^{-1} \Delta \\
= & \Delta^T (C^{-1} + S^{-1}) \Delta - \Delta^T C^{-1} (D - T) - (D - T)^T C^{-1} \Delta + (D - T)^T C^{-1} (D - T) \\
= & (\Delta - (C^{-1} + S^{-1})^{-1} C^{-1} (D - T))^T (C^{-1} + S^{-1}) (\Delta - (C^{-1} + S^{-1})^{-1} C^{-1} (D - T)) \\
- & (D - T)^T C^{-1} (C^{-1} + S^{-1})^{-1} C^{-1} (D - T) + (D - T)^T C^{-1} (D - T),
\end{aligned} \tag{2.9.8}$$

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. (2.2.7) then becomes

$$P(T|D) = P(D|T) \propto \exp \left(-\frac{1}{2} (D - T)^T (C^{-1} - C^{-1} (C^{-1} + S^{-1})^{-1} C^{-1}) (D - T) \right). \tag{2.9.9}$$

However

$$(C^{-1} + S^{-1})^{-1} = (C^{-1} (C + S) S^{-1})^{-1} = S (C + S)^{-1} C, \tag{2.9.10}$$

so that

$$\begin{aligned}
C^{-1} - C^{-1} (C^{-1} + S^{-1})^{-1} C^{-1} &= C^{-1} - C^{-1} S (C + S)^{-1} \\
&= (C^{-1} (C + S) - C^{-1} S) (C + S)^{-1} = (C + S)^{-1}.
\end{aligned} \tag{2.9.11}$$

Restoring the indices, we thus find the simple result

$$P(T|D) \propto \exp \left(-\frac{1}{2} (D_i - T_i) (C + S)_{ij}^{-1} (D_j - T_j) \right). \tag{2.9.12}$$

Comparison of Eq. (2.2.12) with Eq. (2.2.1) indicates that when replacing the true 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} [24]. 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. (2.2.2) becomes proportional to $\delta^N(\mathcal{T}_i - T_i)$. As a result, in this case Eq. (2.2.12) reduces to Eq. (2.2.1) with \mathcal{T}_i replaced by the predictions T_i . This shows that Eq. (2.2.12) 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. (2.2.2) implies that

$$S_{ij} = \langle (\mathcal{T}_i - T_i)(\mathcal{T}_j - T_j) \rangle = \langle \Delta_i \Delta_j \rangle, \quad (2.9.13)$$

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. (2.2.13) 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. 2.3 in general terms and then used in Sect. 2.4 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 [? ? ?]; 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. (2.2.13).

2.10 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 MHOU 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 MHOU 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 MHOU 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 MHOUs 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.

2.10.1 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

$$T(Q^2) \equiv \bar{T}(\alpha_s(Q^2), 1). \quad (2.10.1)$$

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

$$\mu^2 \frac{d}{d\mu^2} \alpha_s(\mu^2) = \beta(\alpha_s(\mu^2)), \quad (2.10.2)$$

where the QCD beta function has the following perturbative expansion:

$$\beta(\alpha_s) = \beta_0 \alpha_s^2 + \beta_1 \alpha_s^3 + \beta_2 \alpha_s^4 + \dots \quad (2.10.3)$$

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

$$\mu^2 \frac{d}{d\mu^2} \bar{T}(\alpha_s(\mu^2), \mu^2/Q^2) = 0. \quad (2.10.4)$$

It will be useful in what follows to define the variables

$$\mu^2 = kQ^2, \quad t = \ln(Q^2/\Lambda^2), \quad \kappa = \ln k = \ln \mu^2/Q^2, \quad (2.10.5)$$

so $\alpha_s(\mu^2)$ is a function of $\ln \mu^2/\Lambda^2 = t + \kappa$. We can then write the RG equation (2.3.4) as

$$\begin{aligned} 0 &= \frac{d}{d\kappa} \bar{T}(\alpha_s(t + \kappa), \kappa) \\ &= \frac{d}{d\kappa} \alpha_s(t + \kappa) \frac{\partial}{\partial \alpha_s} \bar{T}(\alpha_s(t + \kappa), \kappa) \Big|_{\kappa} + \frac{\partial}{\partial \kappa} \bar{T}(\alpha_s(t + \kappa), \kappa) \Big|_{\alpha_s} \\ &= \frac{\partial}{\partial t} \bar{T}(\alpha_s(t + \kappa), \kappa) \Big|_{\kappa} + \frac{\partial}{\partial \kappa} \bar{T}(\alpha_s(t + \kappa), \kappa) \Big|_{\alpha_s}, \end{aligned} \quad (2.10.6)$$

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

$$\frac{d}{d\kappa} \alpha_s(t + \kappa) = \frac{d}{dt} \alpha_s(t + \kappa) = \beta(\alpha_s(t + \kappa)). \quad (2.10.7)$$

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 ,

$$\begin{aligned} \bar{T}(\alpha_s(t + \kappa), \kappa) &= \bar{T}(\alpha_s(t + \kappa), 0) \\ &\quad + \kappa \frac{\partial}{\partial \kappa} \bar{T}(\alpha_s(t + \kappa), 0) \Big|_{\alpha_s} + \frac{1}{2} \kappa^2 \frac{\partial^2}{\partial \kappa^2} \bar{T}(\alpha_s(t + \kappa), 0) \Big|_{\alpha_s} + \dots \\ &= \bar{T}(\alpha_s(t + \kappa), 0) - \kappa \frac{\partial}{\partial t} \bar{T}(\alpha_s(t + \kappa), 0) \Big|_{\kappa} + \frac{1}{2} \kappa^2 \frac{\partial^2}{\partial t^2} \bar{T}(\alpha_s(t + \kappa), 0) \Big|_{\kappa} \end{aligned} \quad (2.10.8)$$

where in the second line we use the RG invariance condition, Eq. (2.3.6), 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 :

$$\bar{T}(\alpha_s(t + \kappa), \kappa) = T(t + \kappa) - \kappa \frac{d}{dt} T(t + \kappa) + \frac{1}{2} \kappa^2 \frac{d^2}{dt^2} T(t + \kappa) + \dots. \quad (2.10.9)$$

Now since

$$\frac{d}{dt} T(t) = \frac{d\alpha_s(t)}{dt} \frac{\partial}{\partial \alpha_s} \bar{T}(\alpha_s(t), 0) = \beta(\alpha_s(t)) \frac{\partial}{\partial \alpha_s} \bar{T}(\alpha_s(t), 0), \quad (2.10.10)$$

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. (2.3.9), 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. (2.3.9) 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

$$\begin{aligned}\bar{T}_{\text{LO}}(\alpha_s(t + \kappa), \kappa) &= T_{\text{LO}}(t + \kappa), \\ \bar{T}_{\text{NLO}}(\alpha_s(t + \kappa), \kappa) &= T_{\text{NLO}}(t + \kappa) - \kappa \frac{d}{dt} T_{\text{LO}}(t + \kappa), \\ \bar{T}_{\text{NNLO}}(\alpha_s(t + \kappa), \kappa) &= T_{\text{NNLO}}(t + \kappa) - \kappa \frac{d}{dt} T_{\text{NLO}}(t + \kappa) + \frac{1}{2} \kappa^2 \frac{d^2}{dt^2} T_{\text{LO}}(t + \kappa).\end{aligned}\tag{2.10.11}$$

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 [?]. 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 MHOU can now be estimated as the difference between the scale varied cross-section and the cross-section evaluated at the central scale, namely

$$\Delta(t, \kappa) = \bar{T}(\alpha_s(t + \kappa), \kappa) - T(t).\tag{2.10.12}$$

Thus at LO, NLO and NNLO we have, using Eq. (2.3.11), that the theory nuisance parameters are given by

$$\begin{aligned}\Delta_{\text{LO}}(t, \kappa) &= T_{\text{LO}}(t + \kappa) - T_{\text{LO}}(t), \\ \Delta_{\text{NLO}}(t, \kappa) &= (T_{\text{NLO}}(t + \kappa) - \kappa \frac{d}{dt} T_{\text{LO}}(t + \kappa)) - T_{\text{NLO}}(t), \\ \Delta_{\text{NNLO}}(t, \kappa) &= (T_{\text{NNLO}}(t + \kappa) - \kappa \frac{d}{dt} T_{\text{NLO}}(t + \kappa) + \frac{1}{2} \kappa^2 \frac{d^2}{dt^2} T_{\text{LO}}(t + \kappa)) - T_{\text{NNLO}}(t).\end{aligned}\tag{2.10.13}$$

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 $O(\kappa)$ term. At NNLO it is milder still, since the $O(\kappa)$ term subtracts the leading dependence in the first term, and the $O(\kappa^2)$ 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 becomes smaller and smaller as the order of the expansion is increased.

Clearly the size of the MHOU, 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 MHOU: 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 MHOUs 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 MHOU in the hard-scattering partonic cross-sections; we will then turn to a discussion of MHOUs 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 MHOU 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.

2.10.2 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).

Electroproduction

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

$$F(Q^2) = C(\alpha_s(Q^2)) \otimes f(Q^2), \quad (2.10.14)$$

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. (2.3.14) 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

$$\bar{F}(Q^2, \mu^2) = \bar{C}(\alpha_s(\mu^2), \mu^2/Q^2) \otimes f(Q^2), \quad (2.10.15)$$

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

$$\bar{F}(t, \kappa) = \bar{C}(\alpha_s(t + \kappa), \kappa) \otimes f(t), \quad (2.10.16)$$

where as in Eq. (2.3.5) we are using the notation $t = \ln Q^2/\Lambda^2$ and $\kappa = \ln \mu^2/Q^2$. Note that in Eq. (2.3.15) the structure function is written as a function of μ^2 in the sense of the RG equation (2.3.4): 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. (2.3.11),

$$\bar{C}(\alpha_s(t + \kappa), \kappa) = C(t + \kappa) - \kappa \frac{d}{dt} C(t + \kappa) + \frac{1}{2} \kappa^2 \frac{d^2}{dt^2} C(t + \kappa) + \dots, \quad (2.10.17)$$

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,

$$C(t) = c_0 + \alpha_s(t)c_1 + \alpha_s^2(t)c_2 + \alpha_s^3(t)c_3 + \dots, \quad (2.10.18)$$

its derivatives can be easily evaluated using the beta function expansion Eq. (2.3.3),

$$\begin{aligned} \frac{d}{dt} C(t) &= \alpha_s^2(t)\beta_0 c_1 + \alpha_s^3(t)(\beta_1 c_1 + 2\beta_0 c_2) + \dots, \\ \frac{d^2}{dt^2} C(t) &= 2\alpha_s^3(t)\beta_0^2 c_1 + \dots, \end{aligned} \quad (2.10.19)$$

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

$$\begin{aligned} \bar{C}(\alpha_s(t + \kappa), \kappa) &= c_0 + \alpha_s(t + \kappa)c_1 + \alpha_s^2(t + \kappa)(c_2 - \kappa\beta_0 c_1) \\ &\quad + \alpha_s^3(t + \kappa)(c_3 - \kappa(\beta_1 c_1 + 2\beta_0 c_2) + \kappa^2\beta_0^2 c_1) + \dots. \end{aligned} \quad (2.10.20)$$

Again, note that in the case where $\mu^2 = Q^2$, and so $\kappa = 0$, one recovers the standard perturbative expansion Eq. (2.3.18). We can now find the scale-dependent structure function,

$$\begin{aligned} \bar{F}(t, \kappa) &= c_0 \otimes f(t) + \alpha_s(t + \kappa)c_1 \otimes f(t) + \alpha_s^2(t + \kappa)(c_2 - \kappa\beta_0 c_1) \otimes f(t) \\ &\quad + \alpha_s^3(t + \kappa)(c_3 - \kappa(\beta_1 c_1 + 2\beta_0 c_2) + \kappa^2\beta_0^2 c_1) \otimes f(t) + \dots. \end{aligned} \quad (2.10.21)$$

Note that evaluating these expressions is numerically very straightforward, in that the scale-varied expression Eq. (2.3.21) 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. (2.3.14) and (2.3.18), 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. (2.3.21).

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:

$$\Sigma(t) = H(t) \otimes (f(t) \otimes f(t)), \quad (2.10.22)$$

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

$$\bar{\Sigma}(t, \kappa) = \bar{H}(\alpha_s(t + \kappa), \kappa) \otimes (f(t) \otimes f(t)). \quad (2.10.23)$$

where, just as for electroproduction, for PDFs evaluated at a fixed scale T , the RG invariance tells us that $\bar{H}(\alpha_s(t), \kappa)$ is given in terms of $H(t)$ by Eq. (2.3.9):

$$\bar{H}(\alpha_s(t), \kappa) = H(t) - \kappa \frac{d}{dt} H(t) + \frac{1}{2} \kappa^2 \frac{d^2}{dt^2} H(t) + \dots . \quad (2.10.24)$$

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

$$H(t) = \alpha_s^n(t) h_0 + \alpha_s^{n+1}(t) h_1 + \alpha_s^{n+2}(t) h_2 + \dots . \quad (2.10.25)$$

Then, as in the case of electroproduction, using Eq. (2.3.3) we can readily evaluate these derivatives,

$$\begin{aligned} \frac{d}{dt} H(t) &= n \alpha_s^{n-1}(t) \beta(\alpha_s) h_0 + (n+1) \alpha_s^n(t) \beta(\alpha_s) h_1 + \dots \\ &= \alpha_s^{n+1} n \beta_0 h_0 + \alpha_s^{n+2} (n \beta_1 h_0 + (n+1) \beta_0 h_1) + \dots \\ \frac{d^2}{dt^2} H(t) &= \alpha_s^{n+2} n(n+1) \beta_0^2 h_0 + \dots \end{aligned} \quad (2.10.26)$$

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 $\bar{\Sigma}$,

Eq. (2.3.23), will be given by

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

This is rather more involved than Eq. (2.3.21), but shares the same advantages: the convolutions to be evaluated in Eq. (2.3.23) have the same structure as those in Eq. (2.3.22), so all that is required to vary the renormalization scale is to modify their coefficients.

2.10.3 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. (2.3.14, 2.3.22), 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

$$\mu^2 \frac{d}{d\mu^2} f(\mu^2) = \gamma(\alpha_s(\mu^2)) f(\mu^2), \quad (2.10.28)$$

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

$$\gamma(t) = \alpha_s(t)\gamma_0 + \alpha_s^2(t)\gamma_1^2 + \alpha_s^3(t)\gamma_2^3 + \dots \quad (2.10.29)$$

Eq. (2.3.28) can be integrated to give

$$f(\mu^2) = \exp\left(\int^{\mu^2} \frac{d\mu'^2}{\mu'^2} \gamma(\alpha_s(\mu'^2))\right) f_0, \quad (2.10.30)$$

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. (2.3.28) 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. 2.3.1, 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

$$\bar{\gamma}(\alpha_s(t), \kappa) = \gamma(t) - \kappa \frac{d}{dt} \gamma(t) + \frac{1}{2} \kappa^2 \frac{d^2}{dt^2} \gamma(t) + \dots, \quad (2.10.31)$$

with κ given by Eq. (2.3.5) and $\gamma(t) = \bar{\gamma}(\alpha_s(t), 0)$, so that given the perturbative expansion Eq. (2.3.29) one has that

$$\begin{aligned} \bar{\gamma}(\alpha_s(t + \kappa), \kappa) &= \alpha_s(t + \kappa)\gamma_0 + \alpha_s^2(t + \kappa)(\gamma_1 - \kappa\beta_0\gamma_0) \\ &\quad + \alpha_s^3(t + \kappa)(\gamma_2 - \kappa(\beta_1\gamma_0 + 2\beta_0\gamma_1) + \kappa^2\beta_0^2\gamma_0) \end{aligned} \quad (2.10.32)$$

is independent of κ up to higher orders terms, order by order. Note that Eq. (2.3.32) has the same form as Eqs. (2.3.25-2.3.27) (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. (2.3.32) in the solution of the evolution equations for the PDFs, Eq. (2.3.30), one finds that the evolution factor can be expressed as

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

where in the first line we changed integration variable (ignoring any change in the lower limit of integration), in the second we used Eq. (2.3.31), 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. [?]. Defining a scale-varied PDF as

$$\bar{f}(\alpha_s(t + \kappa), \kappa) = \exp\left(\int^t dt' \bar{\gamma}(\alpha_s(t' + \kappa), \kappa)\right) f_0, \quad (2.10.34)$$

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. (2.3.33) we find that

$$\bar{f}(\alpha_s(t + \kappa), \kappa) = \left[1 - \kappa \gamma(t + \kappa) + \frac{1}{2} \kappa^2 (\gamma^2(t + \kappa) + \frac{d}{dt} \gamma(t + \kappa)) + \dots\right] f(t + \kappa), \quad (2.10.35)$$

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 .

Eq. (2.3.35) 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. (2.3.24), this gives

$$\begin{aligned} \bar{f}(\alpha_s(t + \kappa), \kappa) &= f(t + \kappa) - \kappa \frac{d}{dt} f(t + \kappa) + \frac{1}{2} \kappa^2 \frac{d^2}{dt^2} f(t + \kappa) + \dots \\ &= f(t + \kappa) - \kappa \gamma f(t + \kappa) + \frac{1}{2} \kappa^2 (\gamma^2 + \frac{d}{dt} \gamma) f(t + \kappa) + \dots, \end{aligned} \quad (2.10.36)$$

where in the second line we used the PDF evolution equation, Eq. (2.3.28). 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. (2.3.35) or Eq. (2.3.36)) is precisely the same. This is essentially because the PDF $f(t)$ depends on only a single scale.

Equation (2.3.35) 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. (2.3.14):

$$\begin{aligned}\widehat{F}(t, \kappa) &= C(t) \overline{f}(\alpha_s(t + \kappa), \kappa) \\ &= C(t) [1 - \kappa\gamma(t + \kappa) + \frac{1}{2}\kappa^2(\gamma^2(t + \kappa) + \frac{d}{dt}\gamma(t + \kappa)) + \dots] f(t + \kappa) \\ &\equiv \widehat{C}(t, \kappa) f(t + \kappa),\end{aligned}\tag{2.10.37}$$

where in the second line we used the expansion Eq. (2.3.35), 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

$$\frac{d}{dt}\gamma(\alpha_s) = \beta(\alpha_s) \frac{d\gamma}{d\alpha_s},\tag{2.10.38}$$

and then using the perturbative expansions of the beta function β , the anomalous dimension γ , and the coefficient function C , Eqs. (2.3.3), (2.3.29), and (2.3.18), respectively, one finds

$$\widehat{C}(t, \kappa) = c_0 - \alpha_s(t)\kappa(c_1 + \gamma_0) + \alpha_s^2(t) (c_2 - \kappa\gamma_1 + \frac{1}{2}\kappa^2(\gamma_0^2 + \beta_0\gamma_0)) + \dots\tag{2.10.39}$$

Note that this result for $\widehat{C}(t, \kappa)$ is not the same as $\overline{C}(t + \kappa, \kappa)$, Eq. (2.3.20). 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 MHOU 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 MHOU 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. (2.3.37) 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 MHOU in the coefficient function as renormalization scale variation: this corresponds to evaluating $\bar{C}(t+\kappa, \kappa)$ in Eq. (2.3.20). The scale variation which estimates MHOU in the anomalous dimension, and corresponds to $\hat{C}(t + \kappa, \kappa)$ in Eq. (2.3.39), is usually called instead factorization scale variation. This terminology is used for example by the Higgs Cross-Section working group [?] 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. 2.3.4 below.

The previous discussion entails that in practice there are (at least) three different ways of estimating the MHOU associated to the PDF evolution in terms of the anomalous dimension 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. (2.3.32). 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. [? ? ?]. 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. (2.3.36). 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 MHOUs in the PDFs with different numbers of active flavors can each be estimated separately. The numerical method was employed in [?], 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. (2.3.39). 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. [?]) since it has the advantage that it can be implemented directly using an external interpolated PDF set (such as provided by LHAPDF [7]). 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. (2.3.33)) 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. (2.3.37) with Eq. (2.3.39)). However, since the differences always involve higher order terms, each method can be regarded as giving an equally valid estimate of the MHOU in the perturbative evolution: differences between methods should be viewed as the uncertainty on the MHOU itself when estimated by scale variation.

2.10.4 Double scale variations

We now discuss the combination of the two independent scale variations of Sects. 2.3.2 and 2.3.3, 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.

Electroproduction

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

$$\overline{F}(Q^2, \mu_f^2, \mu_r^2) = \overline{C}(\alpha_s(\mu_r^2), \mu_r^2/Q^2) \otimes \overline{f}(\alpha_s(\mu_f^2), \mu_f^2/Q^2), \quad (2.10.40)$$

where here and in the following we adopt the (standard) terminology that we introduced in Sect. 2.3.3, 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. (2.3.15)), 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. (2.3.34)), and whose variation estimates MHOUs in the expansion of the anomalous dimension (or equivalently the splitting functions). In the following, as in Sect. 2.3.3, we will omit the convolution as well as the parton indices.

Note that again, as in Eq. (2.3.15), and then in Eqs. (2.3.23), (2.3.31), and (2.3.36), 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 MHOU. 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

$$\mu_f^2 = k_f Q^2, \quad \mu_r^2 = k_r Q^2, \quad \text{with} \quad t_f = t + \kappa_f, \quad t_r = t + \kappa_r, \quad (2.10.41)$$

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

$$\overline{F}(t, \kappa_f, \kappa_r) = \overline{C}(t_r, \kappa_r) \overline{f}(t_f, \kappa_f), \quad (2.10.42)$$

where, as in Sects. 2.3.2 and 2.3.3, the scale-varied PDF and coefficient functions

are

$$\begin{aligned}\bar{f}(t_f, \kappa_f) &= f(t_f) - \kappa_f \frac{d}{dt} f(t_f) + \frac{1}{2} \kappa_f^2 \frac{d^2}{dt^2} f(t_f) + \dots, \\ \bar{C}(t_r, \kappa_r) &= C(t_r) - \kappa_r \frac{d}{dt} C(t_r) + \frac{1}{2} \kappa_r^2 \frac{d^2}{dt^2} C(t_r) + \dots,\end{aligned}\quad (2.10.43)$$

where $f(t_f) \equiv \bar{f}(t_f, 0)$ and $C(t_r) \equiv \bar{C}(t_r, 0)$ 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

$$\begin{aligned}\bar{F}(t, \kappa_f, \kappa_r) &= C(t_r) f(t_f) - \left(\kappa_r \frac{d}{dt} C(t_r) f(t_f) + \kappa_f C(t_r) \frac{d}{dt} f(t_f) \right) + \frac{1}{2} \left(\kappa_r^2 \frac{d^2}{dt^2} C(t_r) f(t_f) \right. \\ &\quad \left. + 2\kappa_r \kappa_f \frac{d}{dt} C(t_r) \frac{d}{dt} f(t_f) + \kappa_f^2 C(t_r) \frac{d^2}{dt^2} f(t_f) \right) + \mathcal{O}(\alpha_s^3).\end{aligned}\quad (2.10.44)$$

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

$$\begin{aligned}\bar{F}(t, \kappa_f, \kappa_r) &= F(t_f, t_r) - \left(\kappa_f \frac{\partial F}{\partial t_f} \Big|_{t_r} + \kappa_r \frac{\partial F}{\partial t_r} \Big|_{t_f} \right) \\ &\quad + \frac{1}{2} \left(\kappa_f^2 \frac{\partial^2 F}{\partial t_f^2} \Big|_{t_r} + 2\kappa_f \kappa_r \frac{\partial^2 F}{\partial t_f \partial t_r} + \kappa_r^2 \frac{\partial^2 F}{\partial t_r^2} \Big|_{t_f} \right) + (2.10.45)\end{aligned}$$

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 subsection, and thus factorize both scale variations into the coefficient function, as done in Eq. (2.3.39). In the case of electroproduction, inserting the expansions of Eq. (2.3.18) in Eq. (2.3.44) one obtains

$$\bar{F}(t, \kappa_f, \kappa_r) = \hat{\bar{C}}(\alpha_s(t_r), \kappa_f, \kappa_r) f(t_f), \quad (2.10.46)$$

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

$$\begin{aligned}\hat{\bar{C}}(\alpha_s(t_r), \kappa_f, \kappa_r) &\equiv c_0 + \alpha_s(t_r) c_1 - \alpha_s(t_f) \kappa_f c_0 \gamma_0 \\ &\quad + \alpha_s(t_r)^2 (c_2 - \kappa_r \beta_0 c_1) - \alpha_s(t_r) \alpha_s(t_f) \kappa_f c_1 \gamma_0 \\ &\quad + \alpha_s^2(t_f) (-\kappa_f c_0 \gamma_1 + \frac{1}{2} \kappa_f^2 c_0 \gamma_0 (\beta_0 + \gamma_0)) + \dots \\ &= c_0 + \alpha_s(t_r) (c_1 - \kappa_f c_0 \gamma_0) + \alpha_s^2(t_r) (c_2 - \kappa_r \beta_0 c_1 - \kappa_f (c_1 \gamma_0 + c_0 \gamma_1) \\ &\quad + \frac{1}{2} \kappa_f^2 c_0 \gamma_0 (\gamma_0 - \beta_0) + \kappa_f \kappa_r \beta_0 c_0 \gamma_0) + \dots\end{aligned}\quad (2.10.47)$$

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

$$\alpha_s(t_f) = \alpha_s(t_r) + (\kappa_f - \kappa_r)\beta_0\alpha_s^2(t_r) + \dots \quad (2.10.48)$$

Note that in the expression for $\widehat{\bar{C}}$ 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. (2.3.18).

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. 2.3.3 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. (2.3.32) or Eq. (2.3.36), while instead the process-dependent μ_r variation is estimated by varying the renormalization scale in the coefficient function, as done in Eq. (2.3.20), or Eq. (2.3.27) 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 MHOUs. 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. (2.3.40) to be equal to each other, $\mu_f^2 = \mu_r^2 = \tilde{\mu}^2$, we have the factorization

$$\widetilde{F}(Q^2, \tilde{\mu}^2) = \widetilde{C}(\alpha_s(\tilde{\mu}^2), \tilde{\mu}^2/Q^2) f(\tilde{\mu}^2). \quad (2.10.49)$$

In most of the earlier papers, mainly concerned with DIS structure functions, e.g. [? ? ? ? ?], 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

Scale	MHOU	‘Traditional’ name [? ? ? ? ?]	‘Modern’ name [?],[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

Table 2.10.1 *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.*

the renormalization scale. Here, to avoid confusion, we will call $\tilde{\mu}^2$ in Eq. (2.3.49) 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 2.3.1.

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

$$\tilde{F}(t + \tilde{\kappa}, \tilde{\kappa}) = \tilde{C}(\alpha_s(t + \tilde{\kappa}), \tilde{\kappa}) f(t + \tilde{\kappa}). \quad (2.10.50)$$

Now the renormalization group invariance of the cross-section [i.e. Eq. (2.3.4)] requires a cancellation between scale variations in the coefficient function and the PDF: with $F(t) \equiv \tilde{F}(t, 0)$,

$$\begin{aligned} \tilde{F}(t + \tilde{\kappa}, \tilde{\kappa}) &= F(t + \tilde{\kappa}) - \tilde{\kappa} \frac{d}{dt} F(t + \tilde{\kappa}) + \frac{1}{2} \tilde{\kappa}^2 \frac{d^2}{dt^2} F(t + \tilde{\kappa}) + \dots \\ &= C f - \tilde{\kappa} \left(\frac{d}{dt} C + \gamma C \right) f + \frac{1}{2} \tilde{\kappa}^2 \left(\frac{d^2}{dt^2} C + 2\gamma \frac{d}{dt} C + C \frac{d}{dt} \gamma + C \gamma^2 \right) f + \dots \end{aligned} \quad (2.10.51)$$

where the first line is the same as Eq. (5.8) in Ref. [?] while in the second line we used Eq. (2.3.36) for scale variation of the PDF. Then, expanding in the usual way, we find that

$$\begin{aligned} \overline{C}(t + \tilde{\kappa}, \kappa) &= c_0 + \alpha_s(t + \tilde{\kappa})(c_1 - \tilde{\kappa} c_0 \gamma_0) \\ &\quad + \alpha_s^2(t + \tilde{\kappa})(c_2 - \tilde{\kappa}(\beta_0 c_1 + c_1 \gamma_0 + c_0 \gamma_1) + \frac{1}{2} \tilde{\kappa}^2 c_0 \gamma_0 (\beta_0 + \gamma_0)) + \dots \end{aligned} \quad (2.10.52)$$

which indeed coincides with the expression for what is referred to as factorization scale variation in this earlier literature: see e.g. Ref. [?], 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. (2.3.47) we set $\kappa_f = \kappa_r = \tilde{\kappa}$, it reduces to Eq. (2.3.52).

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. 2.3.1. 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. [? ?]. 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

$$\tilde{\kappa} \left(\frac{\partial}{\partial t_f} + \frac{\partial}{\partial t_r} \right) + \kappa_f \frac{\partial}{\partial t_f} = (\tilde{\kappa} + \kappa_f) \frac{\partial}{\partial t_f} + \tilde{\kappa} \frac{\partial}{\partial t_r}. \quad (2.10.53)$$

A way of avoiding these large ratios was constructed in Ref. [?]: first do the scale variation of Eq. (2.3.52), but then substitute

$$c_2 \rightarrow c_2 - (\kappa_r - \kappa_f) \beta_f c_1 = c_2 - (\ln \mu_f^2 / \mu_r^2) \beta_0 c_1, \quad (2.10.54)$$

where care must be taken to use the correct argument of α_s in each term. Indeed,

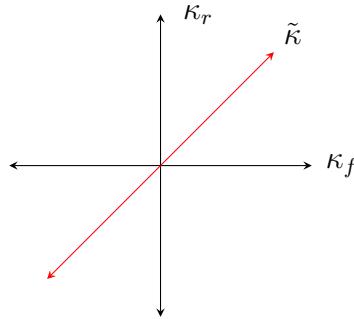


Figure 2.10.1 The two-dimensional space of scale variations for a single process: κ_r is the renormalization scale (giving the MHOU in the hard cross-section), κ_f is the factorization scale (giving the MHOU in the evolution of the PDF) and $\tilde{\kappa}$ is the variation of the scale of the process (called factorization scale variation in the earlier literature), obtained by setting $\kappa_f = \kappa_r$.

this procedure then agrees with Eq. (2.3.46) given that

$$\kappa_f \frac{\partial}{\partial t_f} + \kappa_r \frac{\partial}{\partial t_r} = \kappa_f \left(\frac{\partial}{\partial t_f} + \frac{\partial}{\partial t_r} \right) + (\kappa_r - \kappa_f) \frac{\partial}{\partial t_r}. \quad (2.10.55)$$

Hadronic processes

Consider now the case of hadronic process as in Eq. (2.3.22). For these processes, the factorization has the general form

$$\bar{\Sigma}(t_f, t_r, \kappa_f, \kappa_r) = \bar{H}(\alpha_s(t_r), \kappa_r) \otimes (\bar{f}(t_f, \kappa_f) \otimes \bar{f}(t_f, \kappa_f)). \quad (2.10.56)$$

The hard coefficient function will have the same expansion as Eq. (2.3.27). Just as for electroproduction, it is possible to factorize variations of κ_f into the hard coefficient functions: then

$$\bar{\Sigma}(t_f, t_r, \kappa_f, \kappa_r) = \widehat{H}(\alpha_s(t_r), \kappa_r, \kappa_f) \otimes (f(t_f) \otimes f(t_f)), \quad (2.10.57)$$

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

$$\begin{aligned} \widehat{H} = & \alpha_s^n(t_r) h_0 + \alpha_s^{n+1}(t_r)(h_1 - \kappa_r \beta_0 h_0) - 2\alpha_s^n(t_r)\alpha_s(t_f)\kappa_0 h_0 \gamma_0 \\ & + \alpha_s^{n+2}(t_r)(h_2 - \kappa_2(n\beta_1 h_0 + (n+1)\beta_0 h_1) + \frac{1}{2}\kappa_2^2 n(n+1)\beta_0^2 h_1) \\ & - \alpha_s^{n+1}(t_r)\alpha_s(t_f)(\kappa_0(h_1 - \kappa_2\beta_0 h_0)2\gamma_0) \\ & + \alpha_s^n(t_r)\alpha_s^2(t_f)(-\kappa_0 h_0 2\gamma_1 + \frac{1}{2}\kappa_0^2 h_0 2\gamma_0(\beta_0 + 2\gamma_0)) + \dots \end{aligned} \quad (2.10.58)$$

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.

2.10.5 Multiple scale variations

We finally consider simultaneous scale variation in a pair of processes: for instance the electroproduction process of Sect. 2.3.4 and a hadronic process as

in Sect. 2.3.4. Clearly, the PDF is universal, but the coefficient functions are process-dependent. It follows that scale variation κ_f of the PDF will be totally correlated between these two processes, while the scale variations of κ_r in the two coefficient functions will be totally independent.

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 MHOU 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 MHOU from PDF evolution). Neither of these options is very satisfactory, though we consider the latter briefly in Sect. 2.4.3 below, where it gives rise to asymmetric scale-variation prescriptions.

2.11 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. 2.2), and having discussed how scale variation can be used to produce estimates for such shifts (Sect. 2.3), 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 2.3.1. We will perform scale variation of both scales following Sect. 2.3.4, by taking them as independent, as discussed in that section. When considering a pair of different processes, as explained in Sect. 2.3.5, 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

$$\Delta_{i_a}(\kappa_f, \kappa_{r_a}) \equiv T_{i_a}(\kappa_f, \kappa_{r_a}) - T_{i_a}(0, 0), \quad (2.11.1)$$

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. (2.4.1), $T_{i_a}(\kappa_f, \kappa_{r_a})$ 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. (2.3.5)). In practice, in each prescription the three points $\kappa = 0, \pm w$ are sampled for each scale. The theory covariance matrix is then

$$S_{ij} = n_m \sum_{V_m} \Delta_{i_a}(\kappa_f, \kappa_r) \Delta_{i_b}(\kappa_f, \kappa_s) \quad (2.11.2)$$

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. (2.4.2), as we will discuss below.

Because Eq. (2.4.2) 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 e_i : then it follows that

$$\sum_{ij} e_i S_{ij} e_j = N_m \sum_{V_m} \left(\sum_i e_i \Delta_i \right)^2 \geq 0. \quad (2.11.3)$$

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.

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. (2.4.2) 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 2.3.1), thereby leading to asymmetric prescriptions as already mentioned in Sect. 2.3.5.

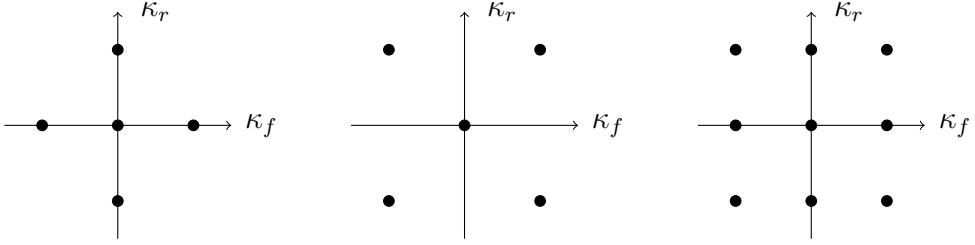


Figure 2.11.1 Symmetric prescriptions for a single process, indicating the sampled values for the factorization scale κ_f and renormalization scale κ_r in each case. The origin of coordinates corresponds to the central scales $\kappa_f = \kappa_r = 0$. We show the three prescriptions 5-point (left), 5-point (center) and 9-point (right).

2.11.1 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

$$S_{ij} = n_m \sum_{v_m} \Delta_i(\kappa_f, \kappa_r) \Delta_j(\kappa_f, \kappa_r), \quad (2.11.4)$$

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. (2.4.4) 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

$$n_m = s/m. \quad (2.11.5)$$

We consider three different prescriptions, represented schematically in Fig. 2.4.1.

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

$$S_{ij}^{(5\text{pt})} = \frac{1}{2}\{\Delta_i^{+0}\Delta_j^{+0} + \Delta_i^{-0}\Delta_j^{-0} + \Delta_i^{0+}\Delta_j^{0+} + \Delta_i^{0-}\Delta_j^{0-}\}, \quad (2.11.6)$$

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. 2.3.4.

- **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 $\bar{n}_4 = 1/2$:

$$S_{ij}^{(\bar{5}\text{pt})} = \frac{1}{2}\{\Delta_i^{++}\Delta_j^{++} + \Delta_i^{--}\Delta_j^{--} + \Delta_i^{+-}\Delta_j^{+-} + \Delta_i^{-+}\Delta_j^{-+}\}. \quad (2.11.7)$$

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

$$S_{ij}^{(9\text{pt})} = \frac{1}{4}\{\Delta_i^{+0}\Delta_j^{+0} + \Delta_i^{-0}\Delta_j^{-0} + \Delta_i^{0+}\Delta_j^{0+} + \Delta_i^{0-}\Delta_j^{0-} + \Delta_i^{++}\Delta_j^{++} + \Delta_i^{+-}\Delta_j^{+-} + \Delta_i^{-+}\Delta_j^{-+} + \Delta_i^{--}\Delta_j^{--}\}. \quad (2.11.8)$$

2.11.2 Symmetric prescriptions for multiple processes

Now we consider multiple processes, i.e. $p > 1$, with scale variations either uncorrelated or partially correlated. In Eq. (2.4.2), 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. (2.4.2) 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

$$S_{ij} = \begin{pmatrix} S_{i_1 j_1} & S_{i_1 j_2} \\ S_{i_2 j_1} & S_{i_2 j_2} \end{pmatrix}. \quad (2.11.9)$$

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. (2.4.4), so we must have

$$S_{i_1 j_1} = N_m \sum_{V_m} \Delta_{i_1}(\kappa_f, \kappa_{r_1}) \Delta_{j_1}(\kappa_f, \kappa_{r_1}) = n_m \sum_{v_m} \Delta_{i_1}(\kappa_f, \kappa_{r_1}) \Delta_{j_1}(\kappa_f, \kappa_{r_1}). \quad (2.11.10)$$

This can only be true if the set of points v_m in Eq. (2.4.4) is a subset of the set V_m in Eq. (2.4.2): 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.(2.4.2), 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 :

$$N_m = n_m / d_m. \quad (2.11.11)$$

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. 2.4.2 (corresponding to the single-process prescriptions shown in Fig. 2.4.1).

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

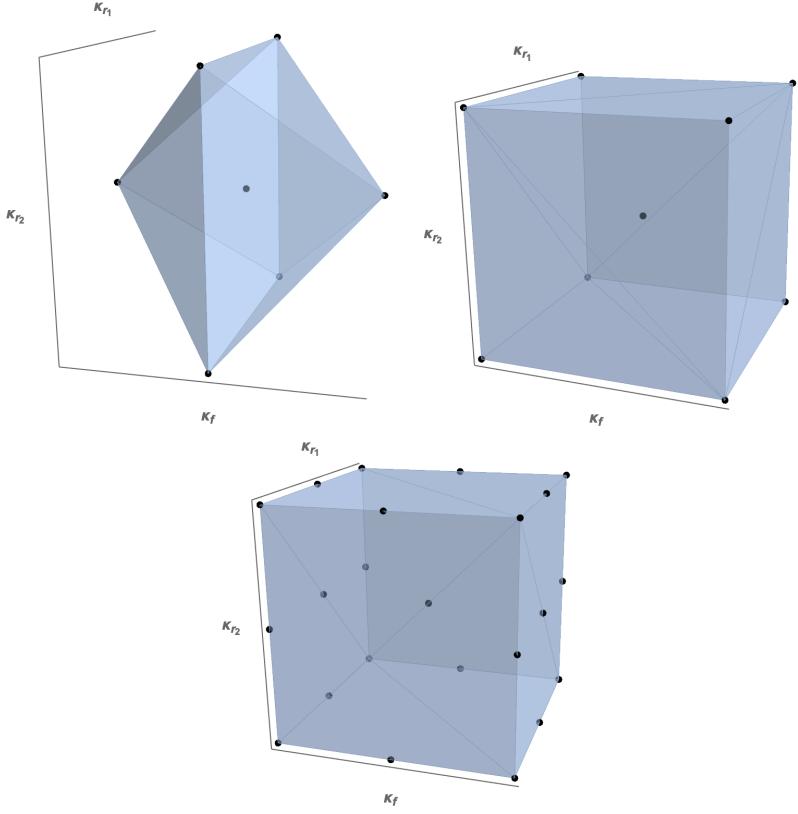


Figure 2.11.2 Same as Fig. 2.4.1, now for the case of two different processes π_1 and π_2 with a common factorization scale κ_f and different renormalization scales κ_{r1} and κ_{r2} , so the diagrams are now in three dimensions. The origin of coordinates is associated to the central scale, $\kappa_f = \kappa_{r1} = \kappa_{r2} = 0$. We again show the three prescriptions 5-point (left), $\bar{5}$ -point (center) and 9-point (right).

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, $(\kappa_f; \kappa_{r1}, \kappa_{r2})$, 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

$$S_{i_1 j_2}^{(\bar{5}\text{pt})} = \frac{1}{4} \{ (\Delta_{i_1}^{++} + \Delta_{i_1}^{+-})(\Delta_{j_2}^{++} + \Delta_{j_2}^{+-}) + (\Delta_{i_1}^{-+} + \Delta_{i_1}^{--})(\Delta_{j_2}^{-+} + \Delta_{j_2}^{--}) \}. \quad (2.11.12)$$

From this expression it is clear that while the scale κ_f is varied coherently between the two processes, the scales κ_{r1} and κ_{r2} 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. (2.4.12) 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. (2.4.2) 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

$$S_{i_1 j_2}^{(5\text{pt})} = \frac{1}{4} \{2\Delta_{i_1}^{+0}\Delta_{j_2}^{+0} + 2\Delta_{i_1}^{-0}\Delta_{j_2}^{-0} + (\Delta_{i_1}^{0+} + \Delta_{i_1}^{0-})(\Delta_{j_2}^{0+} + \Delta_{j_2}^{0-})\}. \quad (2.11.13)$$

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 elements, though only 10 are distinct: the other six are simply copies, necessary to obtain the correct coefficients in Eq. (2.4.6) and Eq. (2.4.13). 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. (2.4.13) 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 $\kappa_f = 0$ 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

$$\begin{aligned} S_{i_1 j_2}^{(9\text{pt})} = & \frac{1}{24} \{ 2(\Delta_{i_1}^{+0} + \Delta_{i_1}^{++} + \Delta_{i_1}^{+-})(\Delta_{j_2}^{+0} + \Delta_{j_2}^{++} + \Delta_{j_2}^{+-}) \\ & + 2(\Delta_{i_1}^{-0} + \Delta_{i_1}^{-+} + \Delta_{i_1}^{--})(\Delta_{j_2}^{-0} + \Delta_{j_2}^{-+} + \Delta_{j_2}^{--}) \} \\ & + 3(\Delta_{i_1}^{0+} + \Delta_{i_1}^{0-})(\Delta_{j_2}^{0+} + \Delta_{j_2}^{0-}) \}. \end{aligned} \quad (2.11.14)$$

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. (2.4.14) 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.

2.11.3 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 2.3.1 and Eq. (2.3.49). 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 MHOU: 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

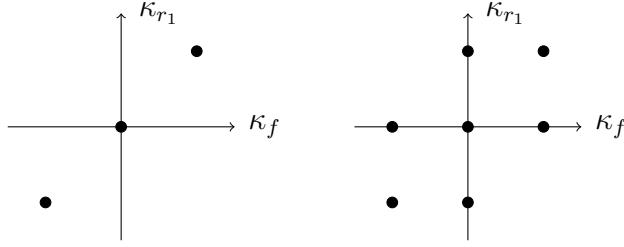


Figure 2.11.3 Same as Fig. 2.4.1, now in the case of the asymmetric prescriptions for a single process with factorization scale κ_f and renormalization scale κ_r . We display the 3-point (left) and 7-point (right) prescriptions, defined in the text.

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. 2.4.3, 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

$$S_{ij}^{(3pt)} = \frac{1}{2}\{\Delta_i^{++}\Delta_j^{++} + \Delta_i^{--}\Delta_j^{--}\}, \quad (2.11.15)$$

whenever $i, j \in \pi$.

Likewise, for two different processes π_1 and π_2 , we set $\kappa_f = \kappa_{r1}$ for π_1 , set $\kappa_f = \kappa_{r2}$ for π_2 , and then vary κ_{r1} and κ_{r2} 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 scales $(\kappa_{r1}, \kappa_{r2})$. Clearly, for each element of v_2 there are two elements of V_2 , so $d_2 = 2$, Eq. (2.4.11) gives $N_2 = 1/4$, and the off-diagonal entries of the theory covariance matrix are

$$S_{i_1 j_2}^{(3pt)} = \frac{1}{4}\{(\Delta_{i_1}^{++} + \Delta_{i_1}^{--})(\Delta_{j_2}^{++} + \Delta_{j_2}^{--})\}. \quad (2.11.16)$$

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 $\kappa_f = \kappa_{r_1} = \kappa_{r_2}$ 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. 2.3.4, 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 are 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 $v_6 = \{(\pm; 0), (0; \pm), (+; +), (-; -)\} = \{(\pm; 0), (0; \pm), (\mp; \mp)\}$, where $(\mp; \mp)$ simply means that the variation is fully correlated (so there are only 2 terms, not 4).

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

$$S_{ij}^{(7pt)} = \frac{1}{3} \{ \Delta_i^{+0} \Delta_j^{+0} + \Delta_i^{-0} \Delta_j^{-0} + \Delta_i^{0+} \Delta_j^{0+} + \Delta_i^{0-} \Delta_j^{0-} + \Delta_i^{++} \Delta_j^{++} + \Delta_i^{--} \Delta_j^{--} \}. \quad (2.11.17)$$

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 $(\kappa_{f_1}, \kappa_{r_1}; \kappa_{f_2}, \kappa_{r_2})$. The extension of v_6 is then $V_6 = \{2(+, 0; +, 0), 2(-, 0; -, 0), (0, \pm; 0, \pm), (\mp, \mp; \mp, \mp)\}$, where $(\mp, \mp; \mp, \mp) = \{(+, +; +, +), (+, +; -, -), (-, -; +, +), (-, -; -, -)\}$, and thus $d_6 = 2$, so

$N_6 = 1/6$, and the off-diagonal theory covariance matrix reads

$$S_{i_1 j_2}^{(7\text{pt})} = \frac{1}{6} \{ 2\Delta_{i_1}^{+0} \Delta_{j_2}^{+0} + 2\Delta_{i_1}^{-0} \Delta_{j_2}^{-0} + (\Delta_{i_1}^{0+} + \Delta_{i_1}^{0-})(\Delta_{j_2}^{0+} + \Delta_{j_2}^{0-}) \\ + (\Delta_{i_1}^{++} + \Delta_{i_1}^{--})(\Delta_{j_2}^{++} + \Delta_{j_2}^{--}) \}. \quad (2.11.18)$$

This prescription gives smaller correlations than the symmetric prescriptions since the variation of the two factorisation scales μ_{f_1} and μ_{f_2} 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.

2.12 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. 2.4, 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. 2.6), and also, for comparison, fits using scale-varied theories (Appendix ??).

2.12.1 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 [16], but differs from it in some details, as we now discuss.

The input dataset used here includes fixed-target [11, 14? ? ? ? ? ?] and HERA [?] deep-inelastic inclusive structure functions; charm cross-sections from HERA [?]; gauge boson production from the Tevatron [? ? ? ?]; and electroweak boson production, inclusive jet, $Z p_T$ distributions, and $t\bar{t}$ total and

differential cross-sections from ATLAS [5, 42? ? ? ? ? ? ?], CMS [20? ? ? ? ? ? ?] and LHCb [? ? ? ?] 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 [? ? ? ?] 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 $Q_{\min}^2 = 2.69$ 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 [?], 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.

Because the prescriptions in Sect. 2.4 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 2.5.1. 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).

Process Type	Dataset	Reference	N_{dat}	$N_{\text{dat}} \text{ (total)}$
DIS NC	NMC	[? ?]	134	
	SLAC	[?]	12	
	BCDMS	[? ?]	530	1593
	HERA σ_{NC}^p	[?]	886	
	HERA σ_{NC}^c	[?]	31	
DIS CC	NuTeV dimuon	[14?]	41	
	CHORUS	[11]	430	552
	HERA σ_{CC}^p	[?]	81	
DY	ATLAS W, Z , 7 TeV 2010	[5]	30	
	ATLAS W, Z , 7 TeV 2011	[?]	34	
	ATLAS low-mass DY 2011	[?]	4	
	ATLAS high-mass DY 2011	[?]	5	
	ATLAS Z p_T 8 TeV (p_T^{ll}, M_{ll})	[42]	44	
	ATLAS Z p_T 8 TeV (p_T^{ll}, y_Z)	[42]	48	
	CMS Drell-Yan 2D 2011	[?]	88	
	CMS W asy 840 pb	[?]	11	
	CMS W asy 4.7 pb	[?]	11	
	CMS W rap 8 TeV	[?]	22	484
	CMS Z p_T 8 TeV (p_T^{ll}, M_{ll})	[20]	28	
	LHCb Z 940 pb	[?]	9	
	LHCb $Z \rightarrow ee$ 2 fb	[?]	17	
	LHCb $W, Z \rightarrow \mu$ 7 TeV	[?]	29	
	LHCb $W, Z \rightarrow \mu$ 8 TeV	[?]	30	
	CDF Z rap	[?]	29	
	D0 Z rap	[?]	28	
	D0 $W \rightarrow e\nu$ asy	[?]	8	
	D0 $W \rightarrow \mu\nu$ asy	[?]	9	
JET	ATLAS jets 2011 7 TeV	[?]	31	
	CMS jets 7 TeV 2011	[?]	133	164
TOP	ATLAS σ_{tt}^{top}	[? ?]	3	
	ATLAS $t\bar{t}$ rap	[?]	10	
	CMS σ_{tt}^{top}	[? ?]	3	26
	CMS $t\bar{t}$ rap	[?]	10	76
Total			2819	2819

Table 2.12.1 The categorization of the input datasets into different processes adopted in this work. Each dataset is assigned to one of five

More refined categorizations will be considered elsewhere, but we consider this to be sufficient for a first study.

All calculations are performed using the same settings as in [16]: PDF evolution and the calculation of DIS structure functions up to NNLO are carried out using the `APFEL` [?] program; heavy quark mass effects are included by means of the `FONLL` general-mass variable flavor number scheme [? ? ?]; the charm PDF is fitted alongside the light quark PDFs [?], rather than being generated from perturbative evolution of light partons; the charm quark pole mass is taken to be $m_c = 1.51$ GeV, and the strong coupling constant is fixed to be $\alpha_s(m_Z) = 0.118$, consistent with the latest PDG average [?].

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. (2.3.41), 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 [?] taking the scale-varied NLO theory cross-sections $T_i(k_f, k_r)$ as input. These are provided by `APFEL` [?] for the DIS structure functions and by `APFELgrid` [?] combined with `APPLgrid` [?] 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` [?] and `OpenQCDrad` [?] 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.

2.12.2 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. 2.4, and discuss some of their qualitative features.

In Fig. 2.5.1 we show the diagonal elements of the experimental and theory covariance matrices, or more specifically the experimental uncertainty normalised to the data, $(C_{ii})^{1/2}/D_i$, and the MHOU normalised to the data, $(S_{ii})^{1/2}/D_i$, for $i = 1, \dots, N_{rmdat}$, where D_i is the i -th datapoint. The datapoints are grouped by process and, within a process by experiment, following Table 2.5.1. The theory covariance matrix S_{ij} is computed using the 9-point prescription (the one with the

largest number of independent variations; recall Sect. 2.4). Broadly speaking, the estimated NLO MHOU 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 MHOU is dominant. These latter datapoints will have relatively little weight in a PDF fit with MHOU included, since the large theory uncertainty makes them less useful for the extraction of PDFs. 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. 2.5.2 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. 2.5.2 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. 4.4.1 compares the experimental correlation matrix, given by

$$\rho_{ij}^{(C)} = \frac{C_{ij}}{\sqrt{C_{ii}}\sqrt{C_{jj}}}, \quad (2.12.1)$$

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

$$\rho_{ij}^{(C+S)} = \frac{(C + S)_{ij}}{\sqrt{(C + S)_{ii}}\sqrt{(C + S)_{jj}}}, \quad (2.12.2)$$

for all the prescriptions defined in Sect. 2.4. 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

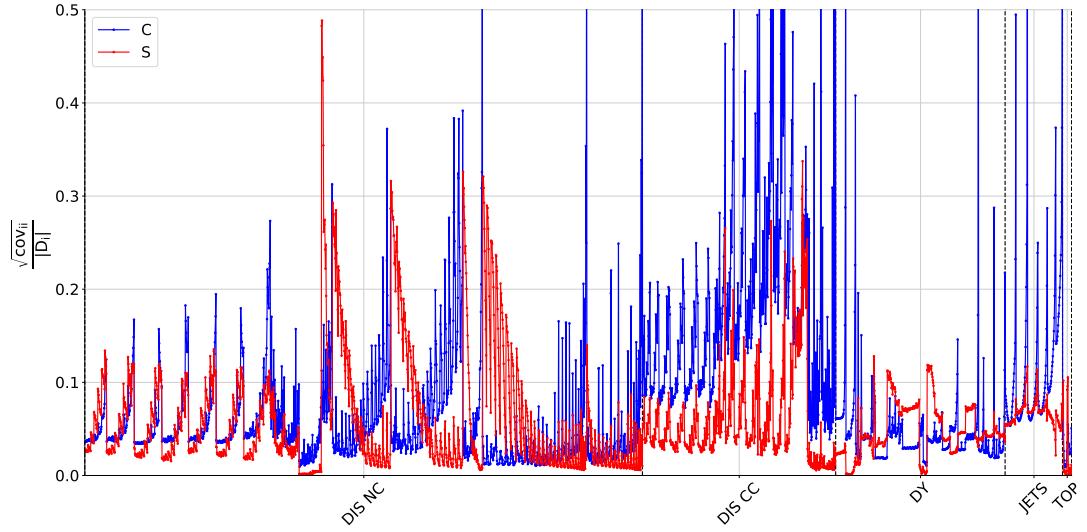


Figure 2.12.1 Comparison of the diagonal experimental uncertainties (blue) and the diagonal theoretical uncertainties (red), each normalised to the data, with the experimental uncertainties evaluated using the 9-point prescription. All entries are normalized to the central experimental value. The data are grouped by process and, within a process, by experiment, following Table 2.5.1

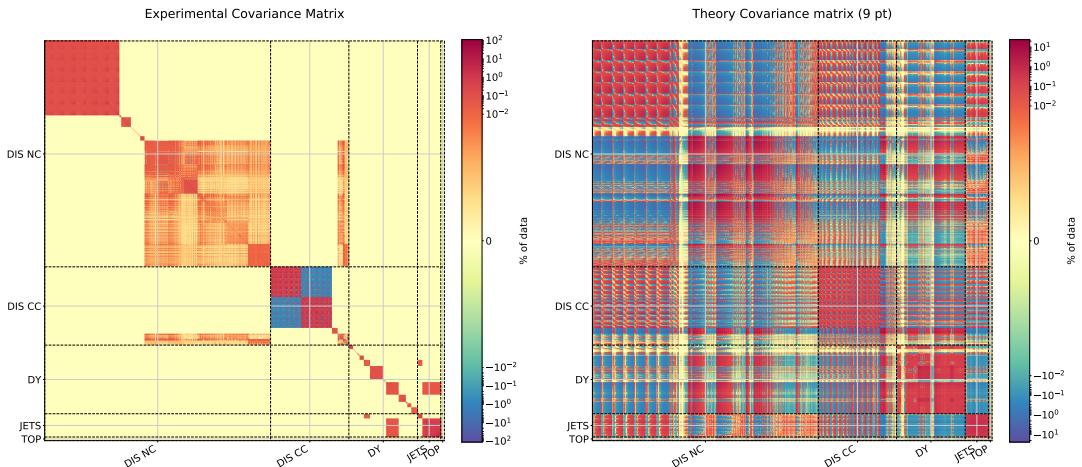


Figure 2.12.2 Comparison of the experimental C_{ij} (left) and the theoretical S_{ij} (right) covariance matrices, the latter evaluated using the 9-point prescription. All entries are normalized to the central experimental value. The data are grouped by process and, within a process, by experiment, following Table 2.5.1

prescriptions. As in Fig. 2.5.2, 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. 2.4.3) the correlated variation of the factorization scale is combined with the uncorrelated variations of the scale of the process for the 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.

2.12.3 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. 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.

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 . 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 2.5.1), while even the most complex prescriptions in Sect. 2.4 have at $\mathcal{O}(30)$ 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 : $E \in S \subset D$. 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 2.5.1) 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. 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

$$\widehat{S}_{ij} = S_{ij}/(T_i^{\text{NLO}} T_j^{\text{NLO}}). \quad (2.12.3)$$

Likewise, we define a normalized shift vector with components

$$\delta_i = (T_i^{\text{NNLO}} - T_i^{\text{NLO}})/T_i^{\text{NLO}}. \quad (2.12.4)$$

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. (2.4.3). 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 ??.

Next we project the shift vector δ_i onto the eigenvectors,

$$\delta^\alpha = \sum_{i=1}^{N_{\text{dat}}} \delta_i e_i^\alpha. \quad (2.12.5)$$

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

$$\delta_i^S = \sum_{\alpha=1, \dots, N_{\text{sub}}} \delta^\alpha e_i^\alpha, \quad (2.12.6)$$

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. 2.5.4.

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

$$\theta = \arccos\left(\frac{|\delta_i^S|}{|\delta_i|}\right) = \arcsin\left(\frac{|\delta_i^{\text{miss}}|}{|\delta_i|}\right) \quad (2.12.7)$$

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.

2.12.4 Results of validation tests

We now explicitly perform the validation tests discussed in Sect. 2.5.3, with the NLO theory covariance matrices \widehat{S}_{ij} (normalised to NLO theory, as in Eq. (2.5.3)) constructed from scale variations for all data points in Table 2.5.1, and for each prescription of Sect. 2.4. These are then validated using the shift vector δ_i constructed as the difference of NNLO and NLO theory, normalised to the latter, as in Eq. (2.5.4).

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 normalised shifts δ_i Eq. (2.5.4). This already 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. 2.5.5-2.5.6. In each plot the data points are presented sequentially on the horizontal axis, organised by process as in Table 2.5.1. The shape of the estimated MHOU 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 MHOU, in the sense that the estimate is small when the MHOU is small, large when it is large, and moreover correctly incorporates the correlations in the

HOU between nearby kinematic regions, responsible for the shape. There is little discernible difference between all the various point prescriptions, except is 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 factorization scales by a factor of two in either direction (i.e. as in Eq. (2.4) with $w = \ln 4$) is 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 some data points the MHOU at NLO is clearly overestimated by scale variation: this is particularly true of the small- x NC DIS data from HERA in the centre 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 diagonalise the theory covariance matrix (using the procedure in Appendix ??), so that we can see in detail which components of the shift vector are correctly estimated, and which are missed, as explained in Sect. 2.5.3.

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 2.12.2 The angle θ Eq. (2.5.7) between this shift and its component δ_i^S lying within the subspace S (see Fig. 2.5.4) spanned by the theory covariance matrix for different prescriptions. The dimension of the subspace S in each case is also given.

As discussed in Sect. 2.5.3, once we have the eigenvectors corresponding to the nonzero eigenvalues of the theory covariance matrix, the first validation test

Presc.	N_{sub}	DIS NC	DIS CC	DY	JET	TOP
		1593	552	484	164	26
5-pt	4	39°	21°	25°	17°	11°
5-bar-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°

Table 2.12.3 Same as Table 2.5.2 for each process of Table 2.5.1. The number of data points in each process is given directly below the name of the process.

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 MHOU provided by the theory covariance matrix. The results of this test for the global dataset, described in Sect. 2.5.1, are shown in Table 2.5.2: 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 value of the angle θ , defined in Eq. (2.5.7), between the shift δ_i and its component δ_i^S , defined in Eq. (2.5.6), 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 MHOU. 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 MHOU 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. (2.4).

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 2.5.3 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 2.5.2 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. 2.5.7. 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. This may suggest that much of it is due to poor estimation of the MHOU in PDF evolution, rather than poor estimation of MHOU in hard cross-sections which can vary substantially between different processes (and indeed different kinematics). A more sophisticated treatment of the factorization scale variation may help improve this.

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 within 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. 2.5.8 for symmetric prescriptions, and in Fig. 2.5.9 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, 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 ???. Also shown are the projections δ^α of the normalized shift vector δ Eq. (2.5.4) along each corresponding eigenvector e_i^α , Eq. (2.5.5).

Inspection of these plots confirms that all the prescriptions seem to perform reasonably well, with the eigenvalues of comparable size to the shifts, the size of the eigenvalues generally falling 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 only 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 MHOU, 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. 2.5.10 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 MHOU, 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.

2.13 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 both at NLO and NNLO 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. 2.2, 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 [16], 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,

$$\lim_{N_{\text{rep}} \rightarrow \infty} \frac{1}{N_{\text{rep}}(N_{\text{rep}} - 1)} \sum_{k=1}^{N_{\text{rep}}} \left(D_i^{(k)} - \langle D_i \rangle \right) \left(D_j^{(k)} - \langle D_j \rangle \right) = C_{ij} + S_{ij}, \quad (2.13.1)$$

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.

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

$$\chi^2 = \frac{1}{N_{\text{dat}}} \sum_{i,j=1}^{N_{\text{dat}}} \left(D_i - T_i^{(0)} \right) (C + S)^{-1}_{ij} \left(D_j - T_j^{(0)} \right), \quad (2.13.2)$$

where $T_i \equiv T_i(0, 0)$ is the theory prediction evaluated with the central scale choice $\kappa_f = \kappa_r = 0$, and with the theory covariance matrix S_{ij} computed using one of the prescriptions presented in Sect. 2.4.

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?] 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.

All the PDF sets which have been produced and which will be discussed in this section are listed in Table 2.6.1. 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.

2.13.1 DIS-only PDFs

We first discuss PDF sets based on DIS data only. Fit quality indicators are collected in Table 2.6.2. The theory covariance matrix is always constructed using the 9-point prescription. We show the value of χ^2/N_{dat} and the ϕ estimator, defined in Ref. [?]. This estimator is a measure of the size of the uncertainty on the prediction: for an uncorrelated covariance matrix, it reduces to the ratio of the uncertainty in the predictions using the output PDFs to that in the original data, which is then generalized to the correlated case. Results are shown for both the total dataset and for the individual DIS experiments of Table 2.5.1. Note that

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	$C + S^{(9\text{pt})}$	
NNPDF31_nnlo_as_0118_dis_kF_1_kR_1	DIS	NNLO	C	baseline DIS-only NNLO
NNPDF31_nnlo_as_0118_dis_scalecov_9pt	DIS	NNLO	$C + S^{(9\text{pt})}$	
NNPDF31_nlo_as_0118_kF_1_kR_1	Global	NLO	C	baseline Global NLO
NNPDF31_nlo_as_0118_scalecov_9pt	Global	NLO	$C + S^{(9\text{pt})}$	
NNPDF31_nlo_as_0118_scalecov_7pt	Global	NLO	$C + S^{(7\text{pt})}$	
NNPDF31_nlo_as_0118_scalecov_3pt	Global	NLO	$C + S^{(3\text{pt})}$	
NNPDF31_nlo_as_0118_scalecov_9pt_fit	Global	NLO	$C + S^{(9\text{pt})}$	S only in χ^2 definition
NNPDF31_nlo_as_0118_scalecov_9pt_sampl	Global	NLO	$C + S^{(9\text{pt})}$	S only in sampling
NNPDF31_nnlo_as_0118_kF_1_kR_1	Global	NNLO	C	baseline Global NNLO

Table 2.13.1 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.

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 2.6.2 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% for the NLO and NNLO fits when including theory a covariance matrix $S^{(9\text{pt})}$ evaluated with the 9-point prescription. Furthermore the value of ϕ also slightly decreases upon inclusion of the theory covariance matrix, both for the total dataset and experiment by experiment.

The decrease in the χ^2 value when including an extra contribution to the covariance matrix is to be expected, so this is essentially a consistency check. However, if the χ^2 reduction was simply a consequence of having increased uncertainties, one would also expect a corresponding increase in uncertainty in the theory predictions, i.e. an increase in ϕ . The fact that both χ^2 and ϕ decrease suggests that the inclusion of MHOUs resolves some tensions in the fit which are present when only the experimental covariance matrix is included. It is interesting that for DIS data these theoretical tensions are greater in the NNLO fit than at NLO.

Dataset	n_{dat}	NNPDF3.1 DIS-only fits							
		χ^2/N_{dat}				ϕ			
		NLO		NNLO		NLO		NNLO	
		C	$C + S^{(9\text{pt})}$	C	$C + S^{(9\text{pt})}$	C	$C + S^{(9\text{pt})}$	C	$C + S^{(9\text{pt})}$
NMC	134	1.259	1.236	1.239	1.163	0.388	0.378	0.444	0.411
SLAC	12	0.908	0.543	0.791	0.589	0.247	0.222	0.343	0.242
BCDMS	530	1.046	1.017	1.047	1.039	0.339	0.319	0.389	0.352
CHORUS	430	0.982	0.856	1.124	0.936	0.409	0.389	0.512	0.435
NuTeV	41	0.628	0.491	0.872	0.540	0.940	0.847	1.35	0.989
HERA incl	967	1.097	1.066	1.104	1.091	0.280	0.272	0.335	0.306
HERA F_2^c	31	1.047	0.997	1.033	1.033	0.526	0.516	0.631	0.474
Total	2145	1.061	1.032	1.095	1.051	0.358	0.355	0.441	0.389

Table 2.13.2 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 $S^{(9\text{pt})}$, compared to the results based on including only the experimental covariance matrix C . Results are provided both for the NLO and for the NNLO fits.

Next we compare PDFs: in Fig. 2.6.1 we compare the gluon and the total quark singlet PDF at $Q = 10$ GeV at NLO and NNLO 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.

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 upon inclusion of the theory covariance matrix, in fact it even decreases. This

is consistent with the previous observation that the uncertainty on the theory predictions themselves decreases somewhat. In the case of the gluon PDF, which is only loosely constrained by the DIS data, the uncertainty increases substantially with MHOUs, particularly in the NNLO fit. 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. So this indirectly supports the conclusion that in the regions where data are abundant the decrease in uncertainty is due to the fact that the theory contribution to the covariance matrix helps to resolve tensions in the fit.

Comparison to the NNLO best fit shows that its value is clearly compatible with the total uncertainty band.

2.13.2 Global PDFs

We now discuss PDFs determined from the global dataset presented in Sect. 2.5.1. 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 2.6.3 and 2.6.4 respectively, both for the total dataset and for the individual processes of Table 2.5.1. In comparison to the DIS-only case of Table 2.6.2 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. (2.6.2), or only in the data generation Eq. (2.6.1), in order to understand better the two distinct effects.

As in the case of the DIS-only fit, on adding the MHOU we find a reduction of χ^2 both for the global fit and for individual datasets, while the value of ϕ is roughly unchanged. We again conclude that the inclusion of the theory covariance matrix improves fit quality by removing some tension which would be otherwise present between datasets. Indeed the total χ^2 for the NLO fit with MHOU (in 9-point prescription) is now comparable to that of the NNLO fit. Comparing different prescriptions, results are reasonably stable, even when comparing to the 3-point prescription which, as discussed in Sects. 2.4.3-2.5.3, 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. 2.5.4.

We finally turn to fits in which the theory covariance matrix is included either in

the χ^2 definition Eq. (2.6.2) but not in the data generation Eq. (2.6.1), or in the data generation Eq. (2.6.1) but not in the χ^2 definition Eq. (2.6.2). 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 MHOU in the χ^2 definition alone leads to a χ^2 value which is very close to that found when the MHOU 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. On the other hand, the ϕ value is somewhat reduced, due to the relaxation of some of the tensions in the fit, now uncompensated by the great fluctuation of the replicas. In the latter case, we expect to obtain increased uncertainties but a worse fit, since the data replica fluctuations are wider due to the MHOU, 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, showing the increase in uncertainty due to the inclusion of MHOU in the sampling, now uncompensated by a rebalancing of the datasets through the inclusion of MHOU 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. 2.6.1. Specifically, in Fig. 2.6.2. we show the results of the NLO fits based on C and $C + S^{(9\text{pt})}$, 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.

We find that in the data region the PDF uncertainty is not substantially 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 2.6.4 do not increase 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 can increase significantly.

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 MHOU the central best fit moves

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^{ll}, 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^{ll}, 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^{ll}, 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 $W \rightarrow e\nu$ 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 σ_{tt}^{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 σ_{tt}^{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 2.13.3 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 definition Eq. (2.6.2) but not in the data generation Eq. (2.6.1) (marked $S_{\text{fit}}^{9\text{pt}}$) and then in the data generation Eq. (2.6.1) but not in the χ^2 definition Eq. (2.6.2) (marked $S_{\text{samp}}^{9\text{pt}}$). Values corresponding to the NNLO fit with experimental covariance matrix C only are also shown.

Process	ϕ in the NNPDF3.1 global fits						
	NLO						NNLO
	C	$C + S^{(9\text{pt})}$	$C + S^{(7\text{pt})}$	$C + S^{(3\text{pt})}$	$C + S_{\text{fit}}^{(9\text{pt})}$	$C + S_{\text{sampl}}^{(9\text{pt})}$	
DIS NC	0.266	0.268	0.262	0.261	0.261	1.137	0.305
DIS CC	0.389	0.376	0.367	0.391	0.369	0.502	0.471
DY	0.361	0.343	0.340	0.358	0.349	0.603	0.380
JETS	0.295	0.312	0.279	0.291	0.298	0.461	0.392
TOP	0.375	0.352	0.318	0.331	0.319	0.612	0.363
Total	0.314	0.315	0.304	0.313	0.309	0.932	0.362

Table 2.13.4 Same as Table 2.6.3, but for the values of the ϕ estimator.

towards the correct NNLO answer. This is further evidence that indeed the theory covariance matrix has resolved tensions due to MHOs. This improved agreement of the central value of the NLO $C + S^{(9\text{pt})}$ 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. 2.6.3 we compare PDFs obtained using different prescriptions. The corresponding relative PDF uncertainties are compared in Fig. 2.6.4. In agreement with what we saw for the χ^2 and ϕ values in Tables 2.6.3, 2.6.4 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 MHOU is not included.

Finally, in Fig. 2.6.5 we compare PDFs obtained including the theory covariance matrix only in the χ^2 definition Eq. (2.6.2) but not in the data generation Eq. (2.6.1) 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 2.6.3 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 MHOU. Indeed in this case the central value is very close to that obtained when including the MHOU 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 2.6.3. These results confirm our expectation that in the full fit, while the MHOU results in a substantial increase in the fluctuations of data replicas, this is compensated by a relaxation of tensions due to the inclusion of MHOU the fit, with the net result that while central values shift, overall uncertainties remain relatively unchanged.

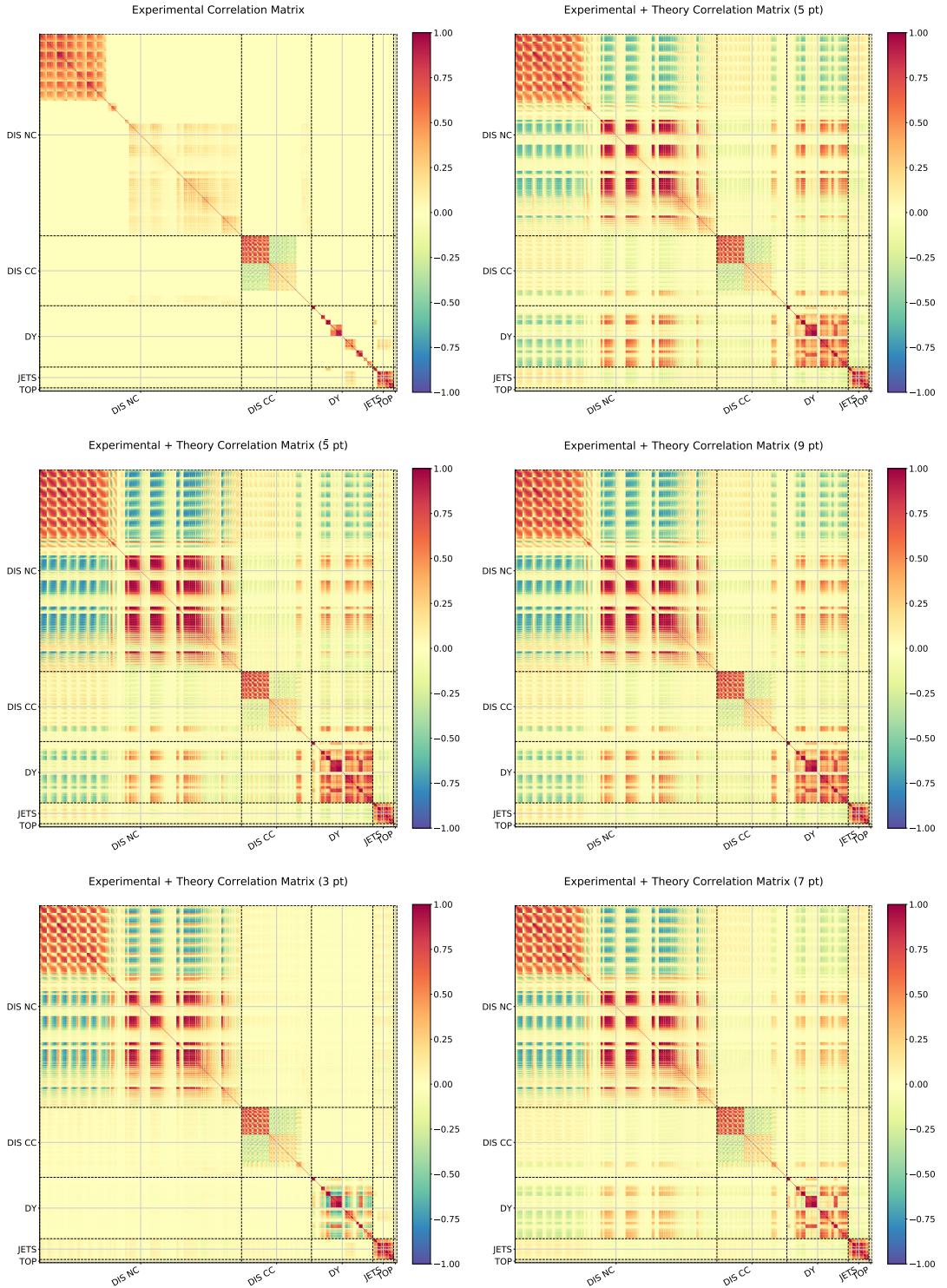


Figure 2.12.3 Comparison of the experimental correlation matrix Eq. (2.5.1) (top left) and the the combined experimental and theoretical correlation matrices Eq. (2.5.2) computed using the prescriptions described in Sect. 2.4: the symmetric prescriptions (5-pt top right, 5-pt centre left, 9-pt centre 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. 2.5.2.

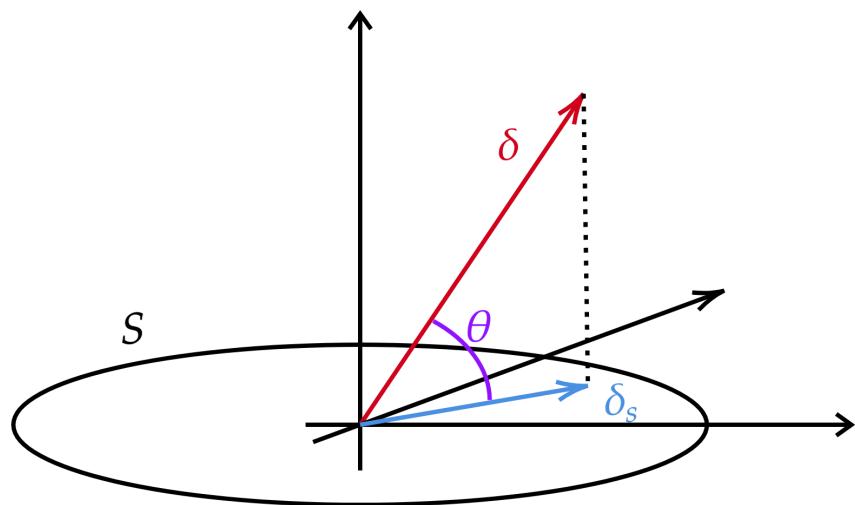


Figure 2.12.4 Schematic representation of the geometric relation between the shift vector $\delta \in D$ (here drawn as a three dimensional space), and the component δ^S of the shift vector which lies in the subspace S (here drawn as a two dimensional space, containing the ellipse E defined by the theory covariance matrix). The angle θ between δ and δ^S is also shown: the dotted line shows the other side of the triangle, $\delta^{\text{miss}} \in D/S$.

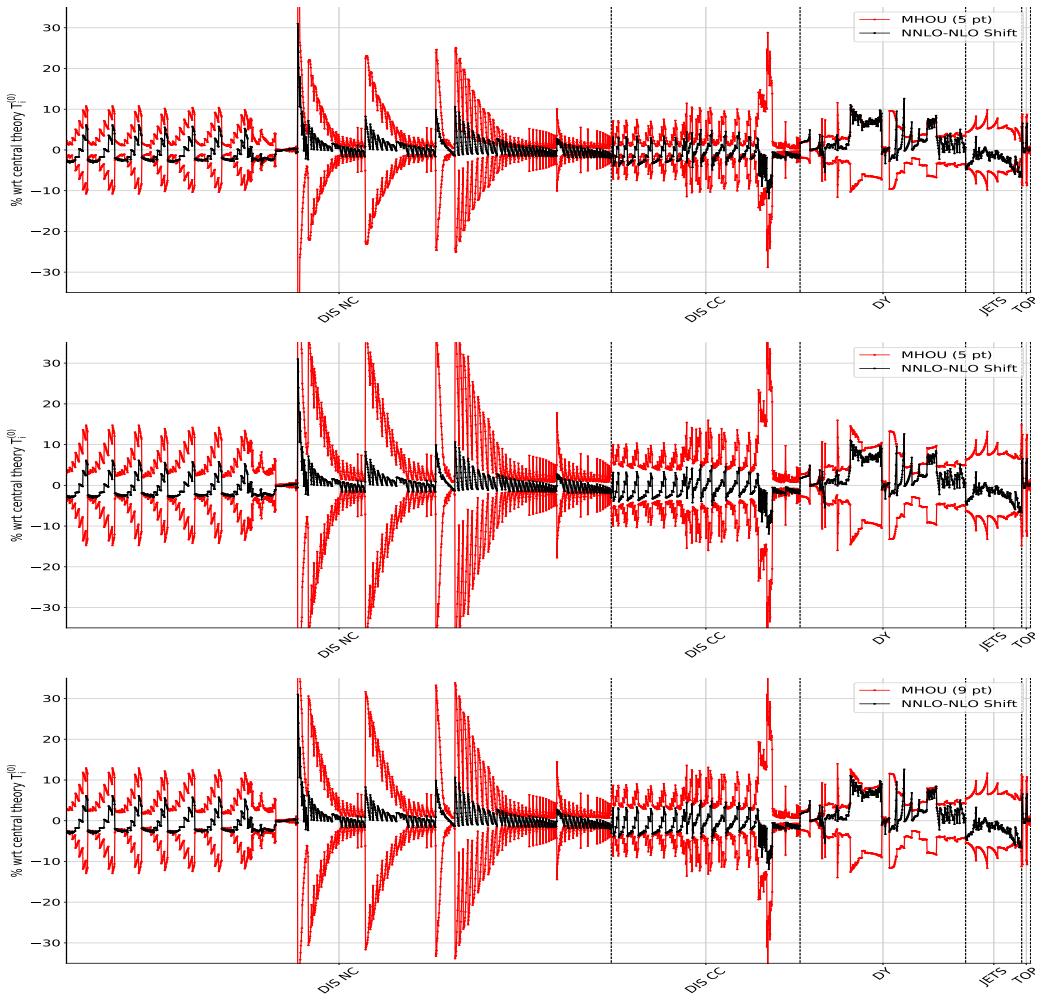


Figure 2.12.5 The diagonal uncertainties σ_i (red) symmetrized about zero, compared to the shift δ_i for each datapoint (black), for the symmetric prescriptions: 5-point (top), 5-point (middle), and 9-point (bottom). All values are shown as percentage of the central theory prediction.

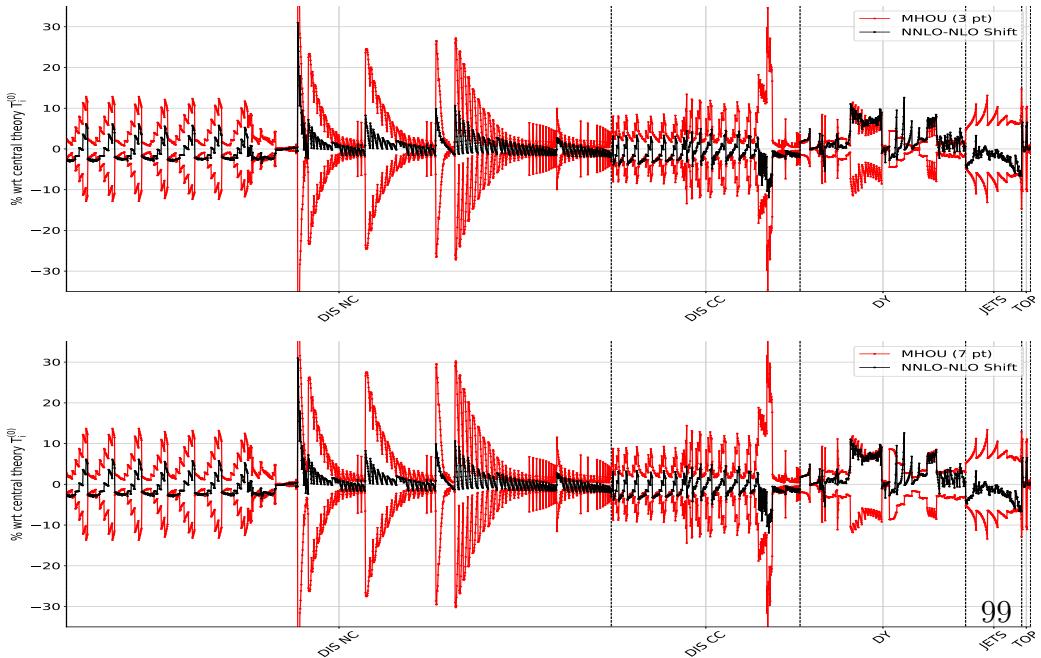


Figure 2.12.6 Same as Fig. 2.5.5 but for the asymmetric prescriptions: 3-point (top) and 7-point (bottom).

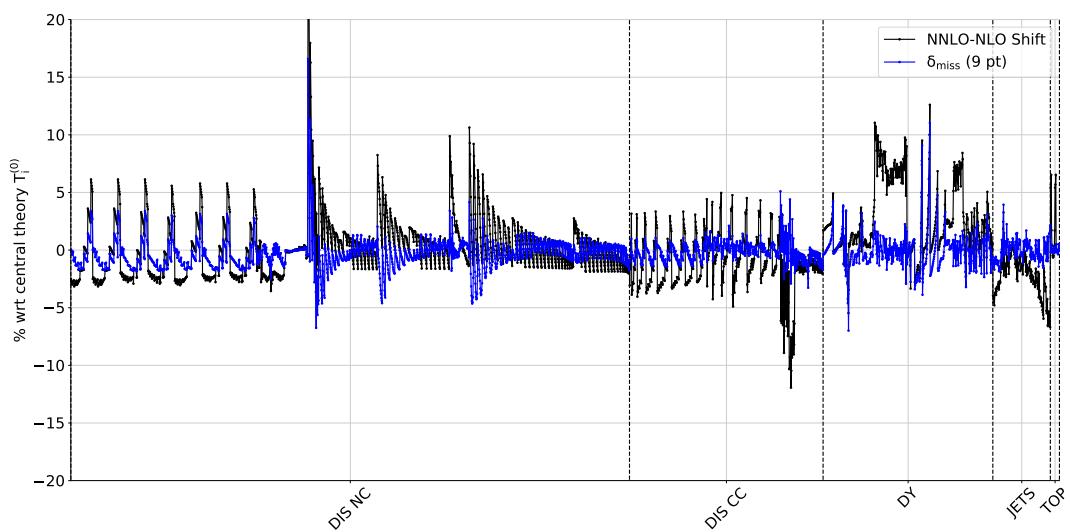


Figure 2.12.7 The NNLO-NLO shift δ_i (black) compared to its component δ_{miss} (blue) which lies outside the subspace S , computed using the 9-point prescription.

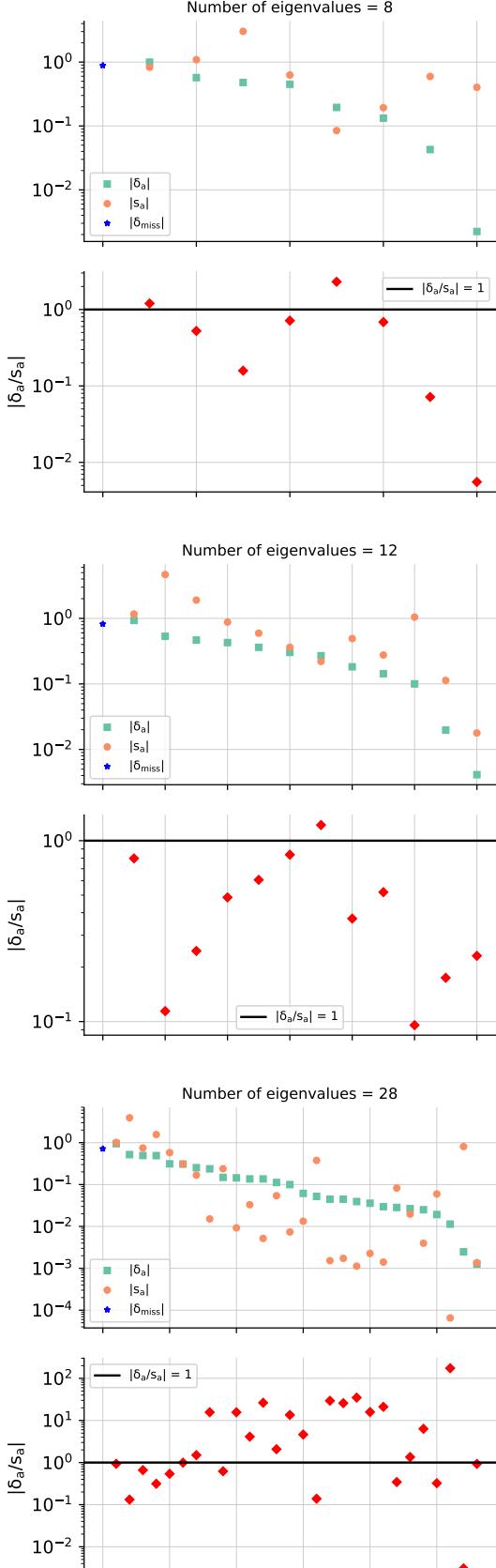


Figure 2.12.8 The projection δ^α Eq. (2.5.5) of the normalized shift vector δ_i Eq. (2.5.4) along each eigenvector e_i^α of the normalized theory covariance matrix Eq. (2.5.3), compared to the corresponding eigenvalue s^α , ordered by the size of the projections (from largest to smallest). In each case 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 matrix $\tilde{\Sigma}_a$ (left), $\tilde{\Sigma}_{\text{cov}}$ (middle), and $\tilde{\Sigma}_{\text{miss}}$ (right).

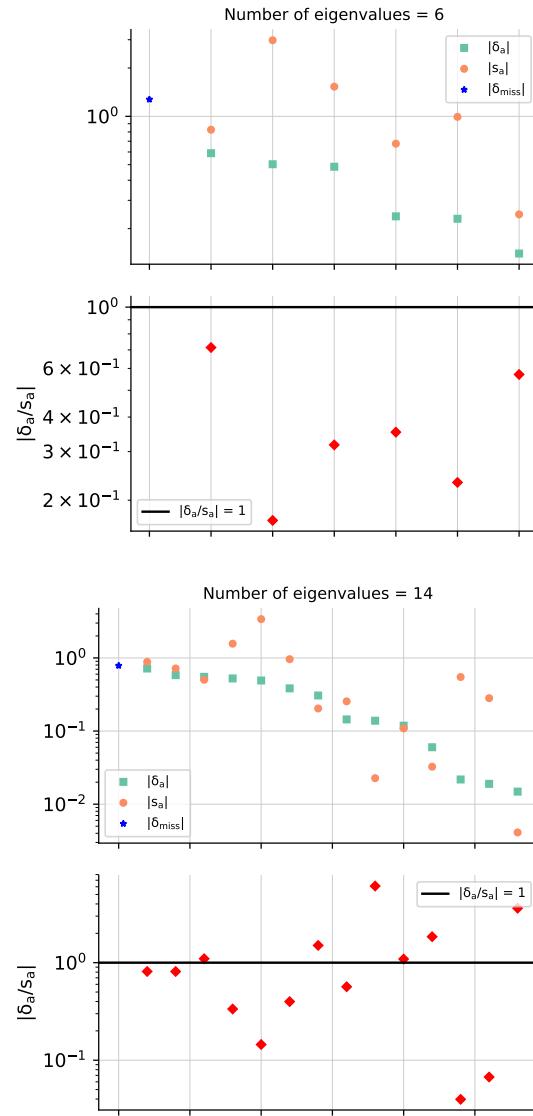


Figure 2.12.9 Same as Fig. 2.5.8 but for the asymmetric prescriptions: 3-point (left) and 7-point (right).

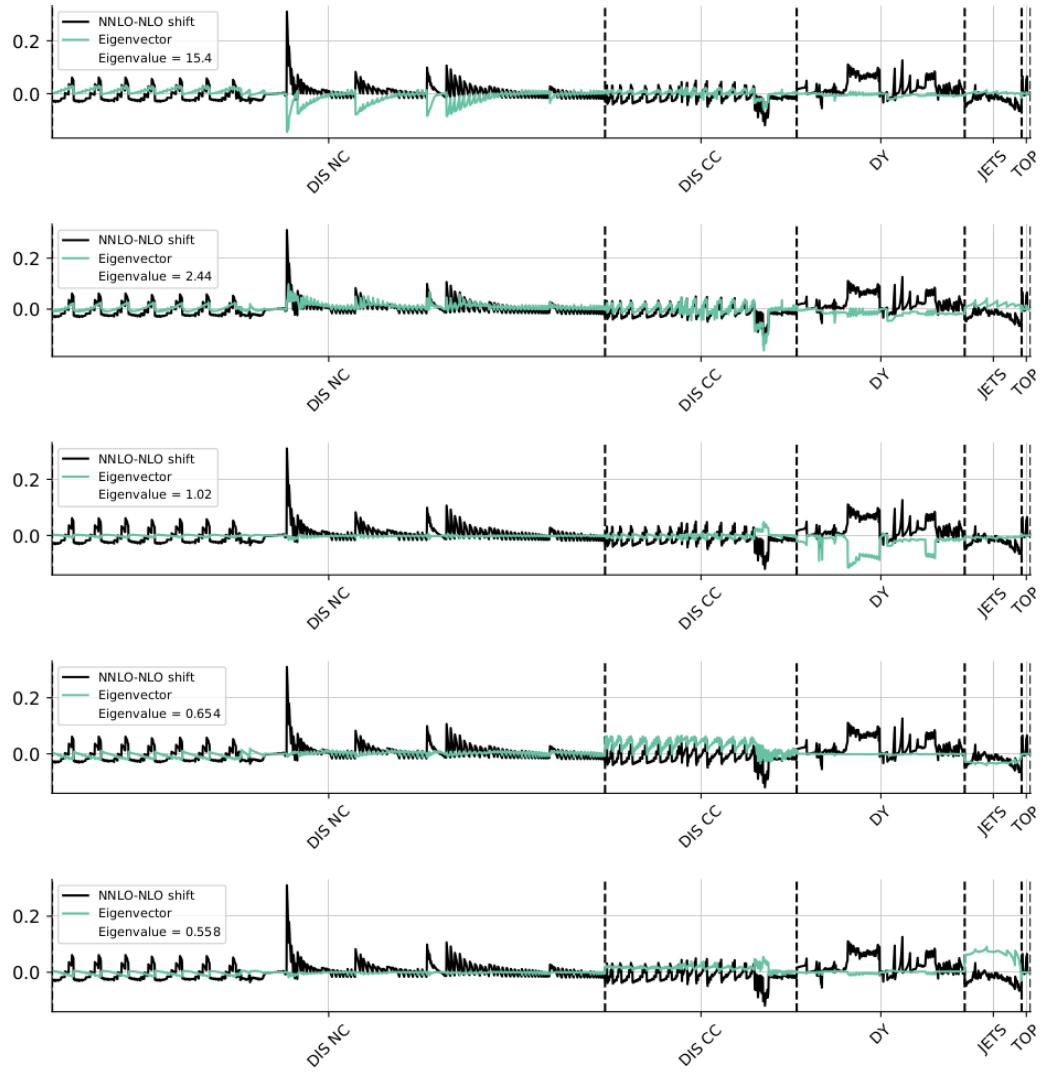


Figure 2.12.10 The components e_i^α (green) of the eigenvectors, corresponding to the five largest eigenvalues for the 9-point theory covariance matrix, shown in the same format as Fig. 2.5.5. The NNLO-NLO shift, δ_i (black), is shown for comparison.

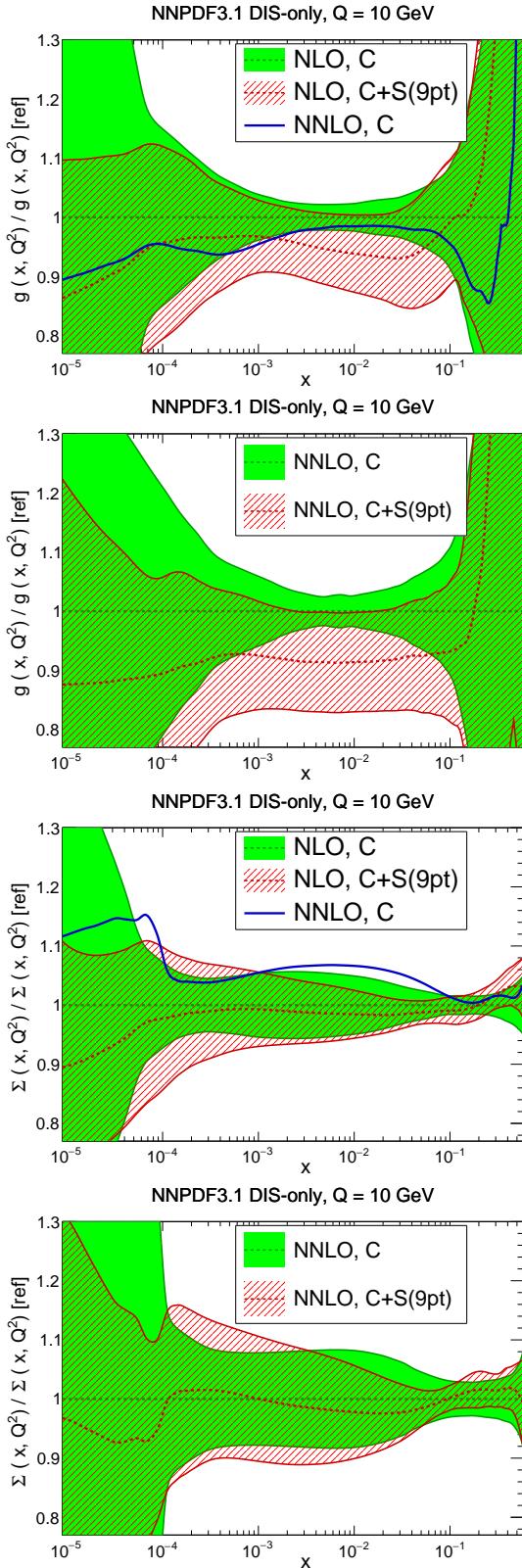


Figure 2.13.1 Comparison of DIS-only PDFs determined with and without MHOUs in the covariance matrix. The gluon (top) and quark singlet (bottom) at $Q = 10 \text{ GeV}$ at NLO (left) and NNLO (right) are shown. The theory covariance matrix S has been constructed using the 9-point prescription. In the NLO plots the central value of the NNLO determined without MHOU is also shown. All results are shown as a ratio to the central value of the set 104 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 MHOU.

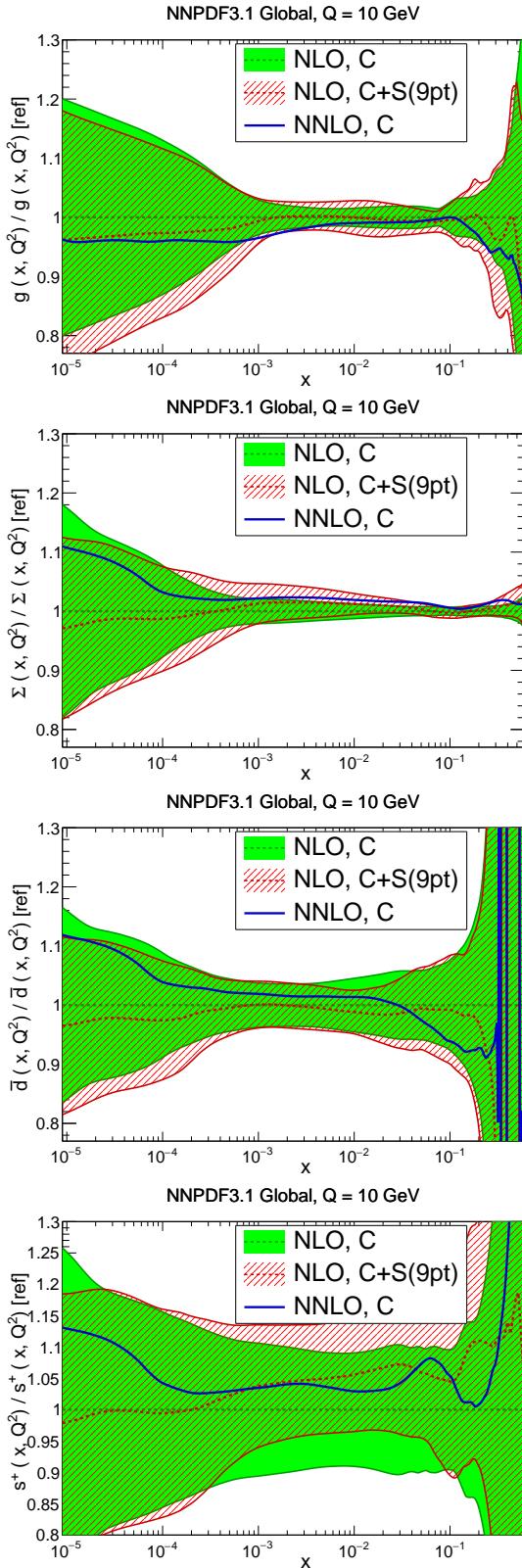


Figure 2.13.2 Same as Fig. 2.6.1 now for the NNPDF3.1 global fits. We show the results of the NLO fits based on C and $C+S^{(9pt)}$ normalized to the former, as well as the central value of the NNLO fit based on C . Results are shown at $Q = 10 \text{ GeV}$ for the gluon, the total quark singlet, the anti-down quark, and the total strangeness PDFs.

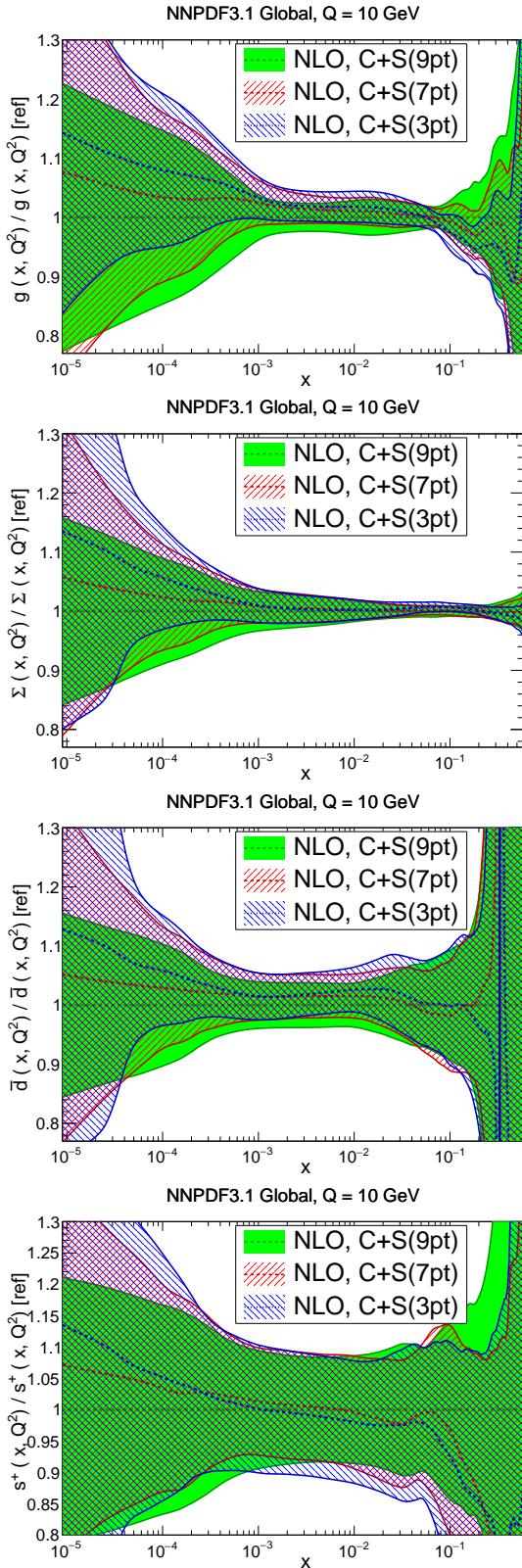


Figure 2.13.3 Same as Fig. 2.6.2 now comparing the results of the NNPDF3.1 global fits with the theory covariance matrix constructed accordingly to the 3-, 7-, and 9-point prescriptions, normalized to the central value of the latter.

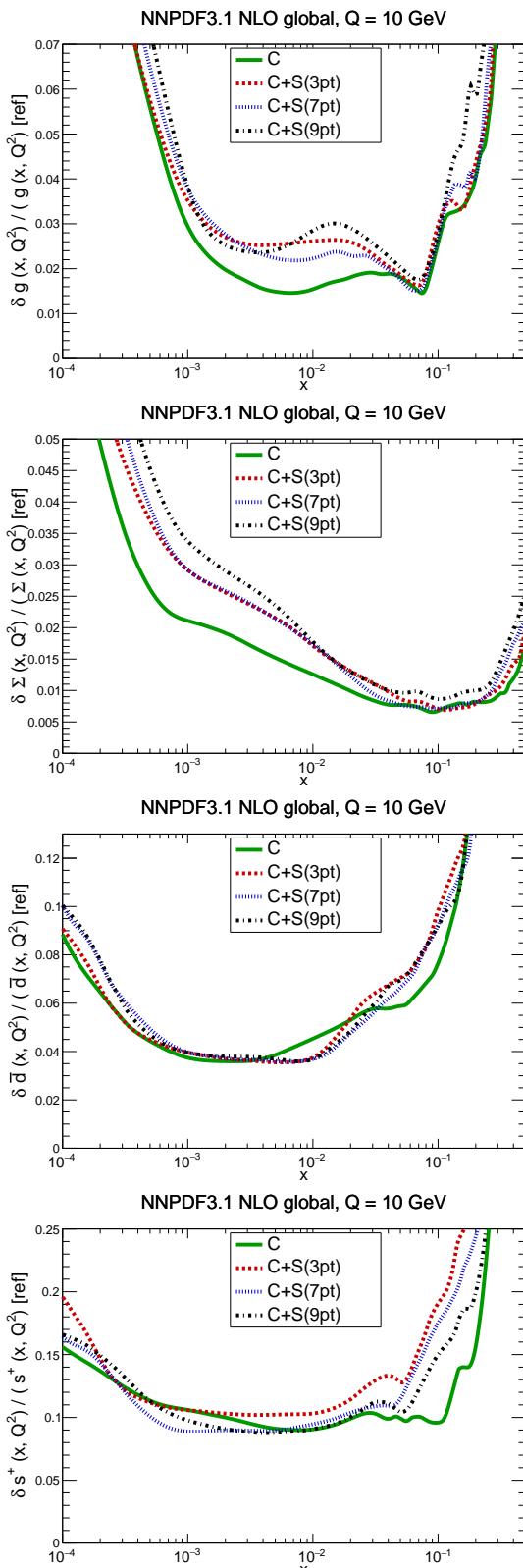


Figure 2.13.4 Same as Fig. 2.6.3, now showing relative PDF uncertainties, normalized to the central value of the baseline set. Note that the y-axes ranges are different for each PDF combination.

2.14 Implications for phenomenology

Whereas a full assessment of the impact of the inclusion of MHOU in PDFs will be possible only once we have global NNLO sets with MHOU, 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. 2.6, and comparing to results with 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 Tab. 2.5.1. 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. 2.8.1 below.

As discussed in Sect. 2.6, 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 (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. [?]. 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. (2.4.8)), 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. 2.8.1 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. 2.6, 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. 2.6, 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.

2.14.1 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 are both currently known at $N^3\text{LO}$ accuracy [? ? ? ?]. 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 2.7.1 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` [? ? ? ? ? ?] 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. 2.7.1.

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. 2.6.2.

From Table 2.7.1 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. (2.4.8), is compatible with the canonical 7-point envelope if the latter is symmetrised by taking the maximum value between the lower and upper uncertainties. In particular, the theory covariance matrix estimate is slightly larger than the

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

Table 2.14.1 *The total cross-sections for Higgs production in gluon fusion (in pb) obtained by using NLO global PDFs based on either C or $C + S^{(9\text{pt})}$, see Table 2.6.1. 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 (symmetrised) envelope of the 7-point scale variations (see Sect. 2.8.1), obtained by taking the maximum value between the lower and upper uncertainties.*

envelope prescription at NLO and at NNLO, while it becomes a little smaller at $N^3\text{LO}$. Even 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 $N^3\text{LO}$ 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. [?] NNLO PDFs were used with the $N^3\text{LO}$ 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 MHOU 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 $N^3\text{LO}$ accuracy using `proVBFH-inclusive` [? ?]. with central factorization and renormalization scales set equal to the squared four-momentum of the vector boson. Results are collected in Table 2.7.2 and shown in Fig. 2.7.1. 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

Higgs production in VBF at 13 TeV

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

Table 2.14.2 Same as Table 2.7.1, now for Higgs production in vector boson fusion.

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. 2.6.2, are almost unaffected by the inclusion of MHOU. On the other hand, in Sect. 2.6.2 we have also seen that MHOU have the effect of moving the central value of the PDFs in the data region towards the NNLO result, and indeed, the shift in the central value of the VBF cross-section due to the MHOU turns out to be significant, by an amount which is comparable to the MHOU σ_F^{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 $N^3\text{LO}$ correction.

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

2.14.2 Top quark pair production

We now study the impact of the PDF-related MHOU on the total top-quark pair production cross-section at the LHC for different center-of-mass energies. In Table 2.7.3 we collect, using the same format as Table 2.7.1, the predictions for the top-quark pair-production cross-sections at $\sqrt{s} = 7, 8$ and 13 TeV obtained using the `top++` code [?] and setting the central scales to $\mu_f = \mu_r = m_t = 172.5$

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

Table 2.14.3 Same as Table 2.7.1, now for top-quark pair-production at $\sqrt{s} = 7, 8$ and 13 TeV.

TeV. The results in the case of 8 and 13 TeV are also displayed in Fig. 2.7.2.

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 MHOU 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 MHOU are included in the determination of the PDFs. This is again compatible with the corresponding behavior of the gluon PDF shown in Fig. 2.6.2, 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 as 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. 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 symmetrised 7-point envelope. In particular, the 9-point result is slightly larger (smaller) than the 7-point envelope at NNLO (NLO). Finally, from Fig. 2.7.2 we notice that for this process the MHOU on the hard cross-section dominates the PDF uncertainty

(with or without MHOU included), even with NLO PDFs.

2.14.3 Z and W gauge boson production

We finally turn to gauge boson production, for which we obtain predictions using the computational framework `Matrix` [?]. In this formalism, all tree-level and one-loop amplitudes are obtained from `OpenLoops` [? ? ?]. 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 2.7.4 we display a similar comparison as in Table 2.7.1 now for W and Z gauge boson production at $\sqrt{s} = 13$ TeV. The corresponding graphical representation of the results is provided in Fig. 2.7.3, again using the same conventions as in Fig. 2.7.1.

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. 2.6.4.

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. We conclude that for 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.

Z production at 13 TeV			
	C	$C + S^{(9\text{pt})}$	
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			
	C	$C + S^{(9\text{pt})}$	
NLO	$3.534 \pm 2.33\% \pm 4.28$ (4.34) %	$3.560 \pm 2.97\% \pm 4.28$ (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			
	C	$C + S^{(9\text{pt})}$	
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) %	

Table 2.14.4 Same as Table 2.7.1, now for W and Z gauge boson production at $\sqrt{s} = 13$ TeV. The cross-section is given in nb.

2.15 Usage and delivery

As mentioned previously, the PDF sets with MHOU presented in Sect. 2.6 can be used in essentially the same way as the standard NNPDF sets. In this section we discuss how MHous 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.

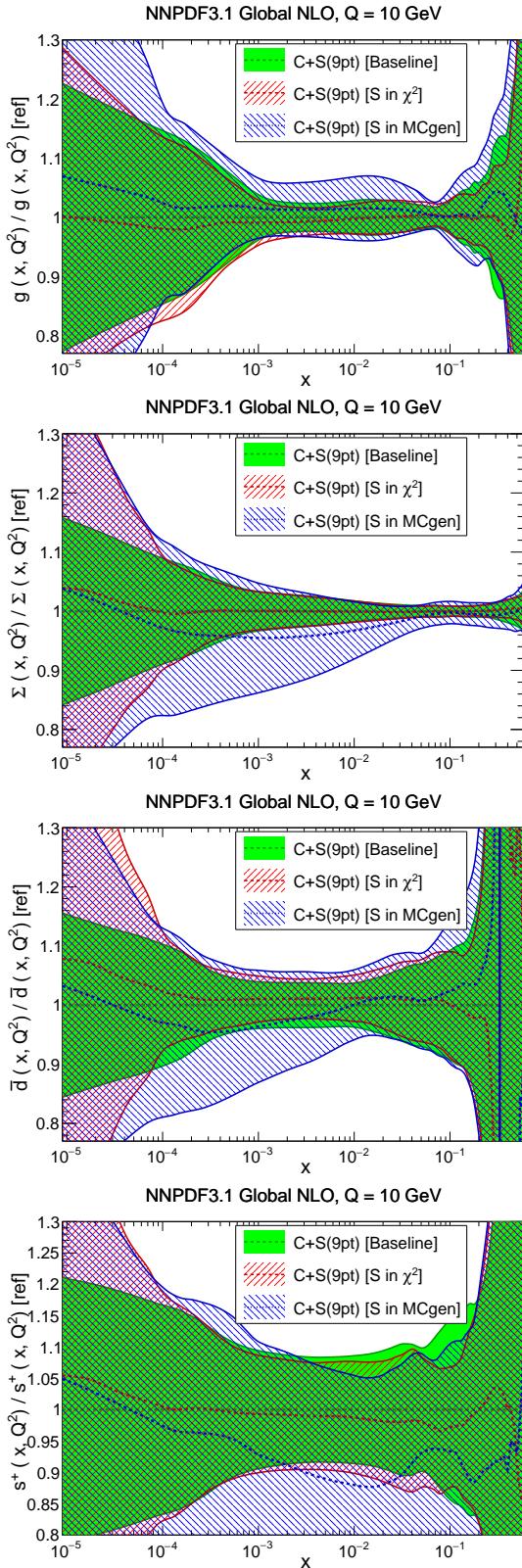


Figure 2.13.5 Same as Fig. 2.6.2, now comparing the results of the baseline $C + S^{(9pt)}$ fit with those in which the theory covariance matrix S is included either in the χ^2 definition or in the generation of Monte Carlo replicas, but not on both.

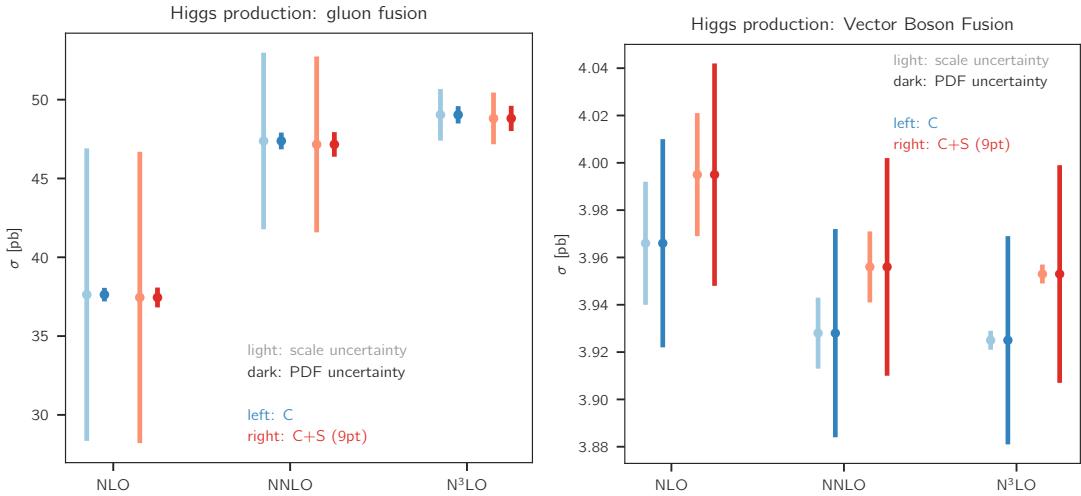


Figure 2.14.1 Graphical representation of the results of Tables 2.7.1 and 2.7.2. At each perturbative order the pair of uncertainty bands on the left (blue) is computed with PDFs based on the experimental covariance matrix C , while the pair of uncertainty bands on the right (red) with PDFs based on 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.

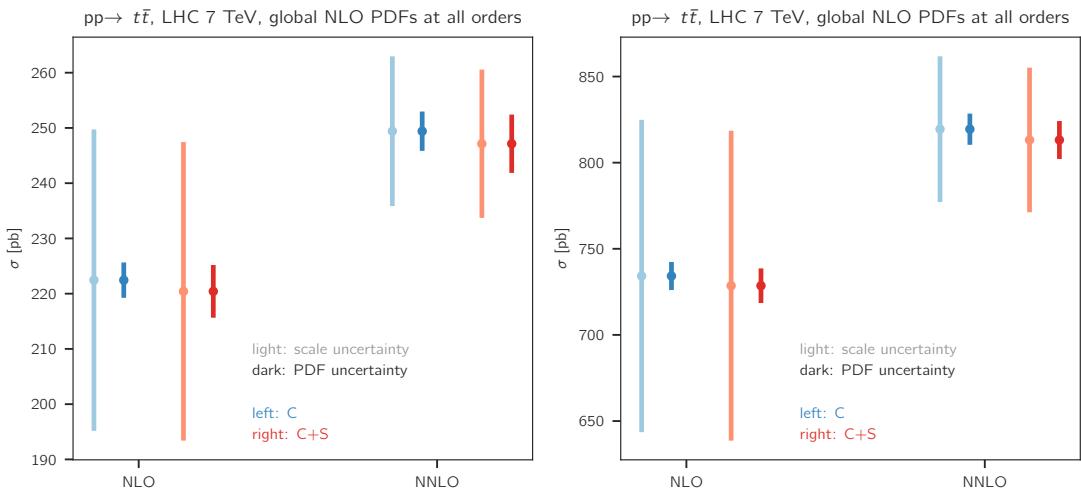


Figure 2.14.2 Same as Fig. 2.7.1 for top-quark pair production at 8 and 13 TeV, see also Table 2.7.3.

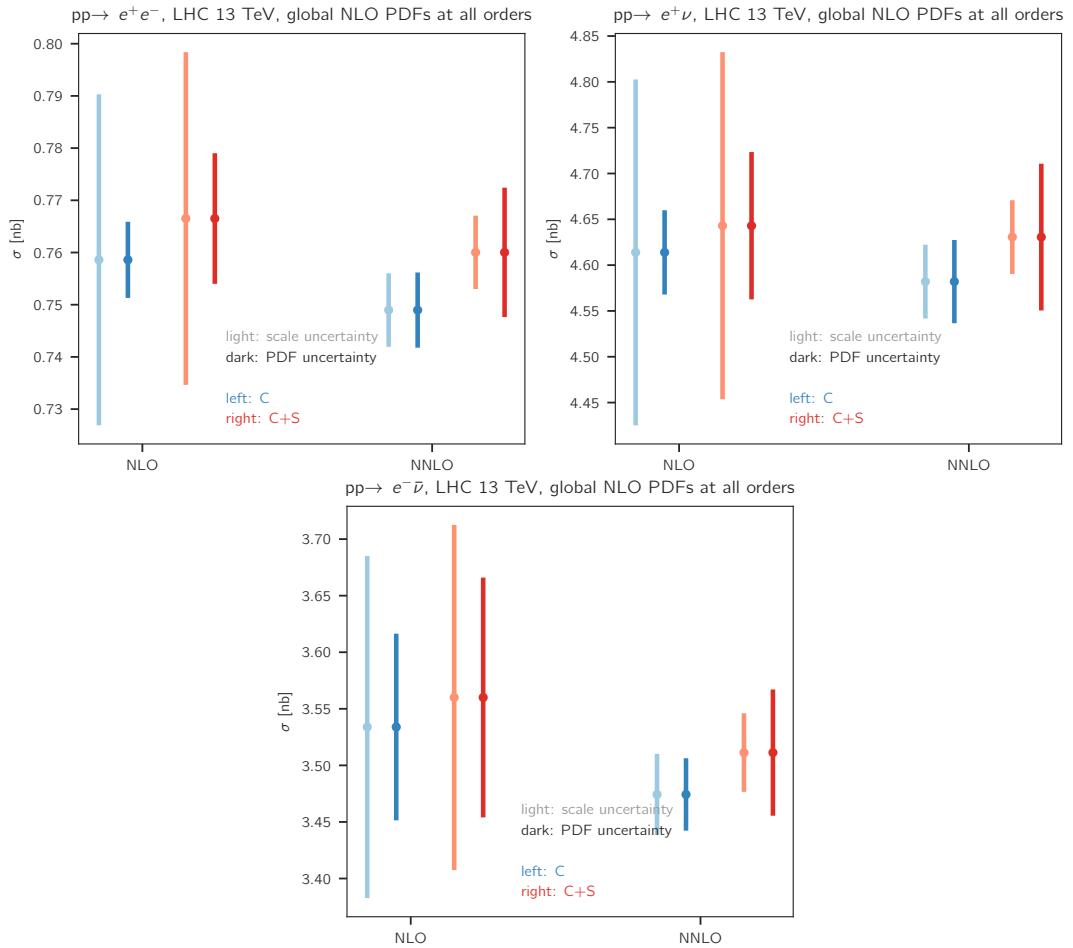


Figure 2.14.3 Same as Fig. 2.7.1 for W^\pm and Z gauge boson production at $\sqrt{s} = 13$ TeV, see also Table 2.7.4.

2.15.1 Combining MHOUs in PDFs and hard matrix elements

As discussed in the introduction, the MHOU on PDFs discussed in this paper arise due to the fact that PDFs are determined using perturbative computations performed at 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: in other words, NLO and NNLO PDFs differ. We have further seen in Sect. 2.3 that there exist two distinct sources of MHOU in the PDF: that related to MHOs in the computation of the hard cross-sections for those processes used for PDF determination, and that coming from MHOs 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 2.3.1.

Thus when considering a factorised prediction for a PDF-dependent hard process not used in the determination of the PDFs, but rather predicted using a given PDF set, there are two independent sources of MHOUs: those specific to the hard process itself (in the hard cross-section for the process and in the evolution of the PDF to the scale of the process), and those in the PDF (due the MHOUs in the hard cross-sections and evolution kernels used to determine the PDF). Clearly, each of these two sources 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 how and whether this correlation should be kept into account.

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. 4.5 and which are already included among the processes of Table 2.5.1. The MHOU coming from renormalization scale variation is then in principle correlated. Indeed, we know from Fig. 4.4.1 that any two predictions for the same physical process are highly correlated. This correlation is not taken into account if the MHOU in the PDF and in the hard matrix element are simply combined in quadrature, as will be recommended in Sect. 2.8.2. This correlation is mostly positive, e.g. between predictions for the same process for points which are kinematically close, so neglecting it provides a conservative estimate of

uncertainties, but it may also be negative, for points in different regions of phase space, or perhaps for different observables related to the same process, such as e.g. a rapidity and a transverse momentum distribution.

Given that we have demonstrated that MHOU on PDFs are generally small, and that most individual processes give a relatively small contribution in a global fit based on a wide dataset (see e.g. Ref. [16]), this lack of correlation is likely to be a very small or even negligible effect. However, if it is felt that a particular process should be predicted in a way that is free of this problem, the only completely safe way to proceed would be to remove that process from the dataset used for PDF determination. The situation here is akin to what happens when one wishes to avoid the prediction for a given process to be biased by the inclusion of that same process in the PDF determination. Indeed, several PDF sets in which specific datasets are removed were presented and discussed in Ref. [16], and additional custom-made sets excluding particular datasets have been provided for specific applications.

A potentially more serious problem was recently raised in Ref. [?], where it was pointed out that the factorization scale dependence of the PDF and the hard matrix element are strongly correlated, and in fact fully correlated for a process (such as nonsinglet deep-inelastic structure function) which depends on one PDF only. This can be easily understood recalling that factorization scale variation estimates MHOU in the anomalous dimension, and thus in perturbative evolution. Consider then a PDF set in which all PDFs are parametrised at a certain scale Q_0 . Assume for the sake of argument that one wishes to predict a hard process whose scale Q is this same scale Q_0 . It is clear that then there should be no evolution uncertainty, and thus one should not vary the factorization scale in the hard matrix element. This is the limiting case of the cases discussed in Ref. [?]: the MHOU related to factorization scale, being related to perturbative evolution, is in principle correlated between the process used for PDF determination and the process which is being predicted in a way that only depends on the evolution length.

However, further reflection shows that this correlation is always lost for realistic PDF determinations, because PDF sets are parametrised at a (typically low) scale Q_0 which does not coincide with the scale of typical hard processes, and that is actually in most cases below the minimum kinematic cut Q_{\min} . This implies that the information on the correlation of the evolution uncertainty when evolving back to the scale of hard processes is lost when constructing the PDF set at

Q_0 . In principle, one could produce various PDF sets, each parametrised at the scale of an individual process, such as “Higgs PDFs” with $Q_0 = m_H/2$ (the scale of Higgs production), “Drell-Yan PDFs” with $Q_0 = m_Z$, and so on. Predictions obtained with these PDFs for the corresponding processes would then have all factorization scale uncertainties (due to evolving from the scales of data used in the PDF fit to Q_0) already incorporated into the PDF uncertainty, so the only additional uncertainty in the prediction would be that from renormalization scale variation of the hard cross-section. However the price to pay would be loss of PDF universality: the “Higgs PDFs” and “Drell-Yan PDFs” would be different, in the sense that one could no longer be obtained from the other entirely through fixed order evolution.

Note that the correlation of factorization scale between all processes used for PDF determination is instead fully accounted for by our formalism, and is included in the theory covariance matrix when adopting any prescription (such as our default 9-point prescription) in which factorization scale variation is fully correlated. Therefore, in practice, by neglecting this correlation in factorization scale between the evolution to Q_0 in the PDF determination and the evolution from Q_0 when making our prediction, we are at worst producing an overestimate of the MHOU. Given the moderate effect of MHOUs on PDFs, this is likely to be a small effect.

We conclude that the uncorrelated combination of the MHOU on the PDF with the MHOU of the hard matrix element of the predicted process is both pragmatic and realistic, especially given the well known uncertainties intrinsic to the estimation of MHOUs.

2.15.2 Computation of the total uncertainty

Having concluded that uncorrelated combination of the MHOU 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:

$$\sigma_{\mathcal{F}}^{\text{PDF}} = \left(\frac{1}{N_{\text{rep}} - 1} \sum_{k=1}^{N_{\text{rep}}} (\mathcal{F}[\{q^{(k)}\}] - \langle \mathcal{F}[\{q\}] \rangle)^2 \right)^{1/2}. \quad (2.15.1)$$

If this prescription is applied to a PDF set with “standard” PDF uncertainty (such as the published NNPDF3.1 [16]) 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 MHOU presented in Sect. 2.5 of this paper are used instead, the resulting uncertainty obtained from Eq. (2.8.1) accounts for both the data-driven and MHOU on the PDF, with all correlations taken into account.

Because the MHOU 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 [?]. 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

$$\sigma_{\mathcal{F}}^{\text{th}} = \left[S_{\mathcal{F}\mathcal{F}}^{(9\text{pt})} \right]^{1/2}, \quad (2.15.2)$$

where $S_{\mathcal{F}\mathcal{F}}^{(9\text{pt})}$ is evaluated using our default 9-point prescription defined by Eq. (2.4.8), with Δ_{ij} computed for $i = j = \mathcal{F}$, i.e. the theory prediction for the given observable. We showed in Sect. 4.5 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. (2.8.1) and the uncertainty on the hard matrix element Eq. (2.8.2) can then be combined as uncorrelated uncertainties. For instance, one can combine them in quadrature and thus the total uncertainty on the cross-section \mathcal{F} is simply

$$\sigma_{\mathcal{F}}^{\text{tot}} = \left((\sigma_{\mathcal{F}}^{\text{th}})^2 + (\sigma_{\mathcal{F}}^{\text{PDF}})^2 \right)^{1/2}. \quad (2.15.3)$$

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 MHOUs, 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. 2.6.2, 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. In this sense, the theory covariance matrix 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.

2.15.3 Delivery

The variants of the NNPDF3.1 NLO global sets presented in this work are publicly available in the LHAPDF format [7] 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. (2.6.2) 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 ?? below. These are determined using Eq. (??), 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

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.

2.16 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 MHOU, such as for example the Cacciari-Houdeau method.

The validation studies presented here suggest however that the conventional scale variation method to estimate the MHOU 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 MHOU might be obtained through more complex patterns of scale variation than those considered here. For example, it might be advantageous to vary independently the renormalization scales in different partonic sub-channels, or the factorization scales for singlet, nonsinglet and valence partons that evolve independently. Indeed, we have observed from the validation of our estimate of MHOU, 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 advantageous. 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 MHOU 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 MHOU — in fact, for the datapoints included in PDF determination it even decreases — but the central value moves close to the true result. Moreover, the fit quality improves, thereby showing that the main effect of the inclusion of MHOU 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 MHOU 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 MHOU at NNLO is expected to lead to the most precise and accurate PDF sets that can be determined with currently available theoretical and experimental information.

Chapter 3

Nuclear Uncertainties - 90%

3.1 Introduction

Parton distribution functions (PDFs) are universal quantities encapsulating the internal structure of the proton, and are crucial for making predictions in particle physics [29]. To maximally constrain them, PDFs are determined by fitting a range of experimental data over a wide variety of processes and kinematic regimes. Some of this data consists of measurements on nuclear targets, rather than proton targets. In this case, the surrounding nuclear environment will have an effect on the measured observables, which in turn will influence the form of the fitted PDFs. The uncertainties associated with these effects are termed "nuclear uncertainties". Such uncertainties are small [15][25] but becoming increasingly relevant with the advent of the Large Hadron Collider and the era of precision physics it has ushered in [12].

In these proceedings, we show how to use existing nuclear PDFs (nPDFs) to provide an estimate of nuclear uncertainties, and include them in future proton PDF fits within the Neural Network PDF (NNPDF) framework [26]¹. We first review the nuclear data (Sec. 4.2), then outline the construction and form of nuclear uncertainties (Sec. 4.3). Finally, we assess the impact on the PDFs and associated phenomenology (Secs. 4.4 and 4.5).

¹For a more detailed analysis, see [?].

3.2 Nuclear Data

There are three experiments with nuclear targets currently included in NNPDF analyses: charged current inclusive deep inelastic scattering (DIS) cross sections from CHORUS [11], on Pb; DIS dimuon cross sections from NuTeV [14][57] on Fe; and Drell-Yan dimuon cross sections from E605 at Fermilab [6], on Cu. After cuts, nuclear data make up 993/4285 of the data points ($\sim 23\%$). For a complete summary of the data sets, see [16].

A study of the correlation between these measurements and the fitted PDFs reveals that the CHORUS data has most impact on the up- and down-valence distributions, NuTeV data has most impact on the strange, and E605 data has most impact on the other light sea quarks: anti-up and anti-down. Therefore, we anticipate largest effects from nuclear uncertainties in these PDFs.

3.3 Determining Nuclear Uncertainties

In a PDF fit we include an experimental covariance matrix, C_{ij} , describing the breakdown of statistical and systematic errors, where i, j run over the data points. Uncertainties due to nuclear data must be considered in addition to the experimental uncertainties, and in general they can be encapsulated in a theoretical covariance matrix, S_{ij} . In a PDF fit we simply add this to C_{ij} [24], so that the nuclear uncertainties act like experimental systematics.

We adopted an empirical approach to construct the nuclear uncertainties, using nPDFs rather than appealing to nuclear models, which rely on various assumptions [23]. We compared theoretical predictions for nuclear observables made with the correct corresponding nPDFs for an isotope “ N ”, $T_i^N[f_N^{(n)}]$, to those with proton PDFs, $T_i^N[f_p]$. Here f_p is the central value for a proton PDF and $f_N^{(n)}$ is one Monte Carlo replica in an nPDF ensemble, where $n = 1, \dots, N_{rep}$ [29]. To generate such an ensemble we combined three recent nPDF sets: DSSZ12 [41], nCTEQ15 [13] and EPPS16 [37]. Note that DSSZ12 does not provide a Cu PDF, so for the case of E605 we combined just two nPDF sets.

We considered two definitions of nuclear uncertainties:

1. **Def. 1,** (a conservative approach) where the only modification is to include

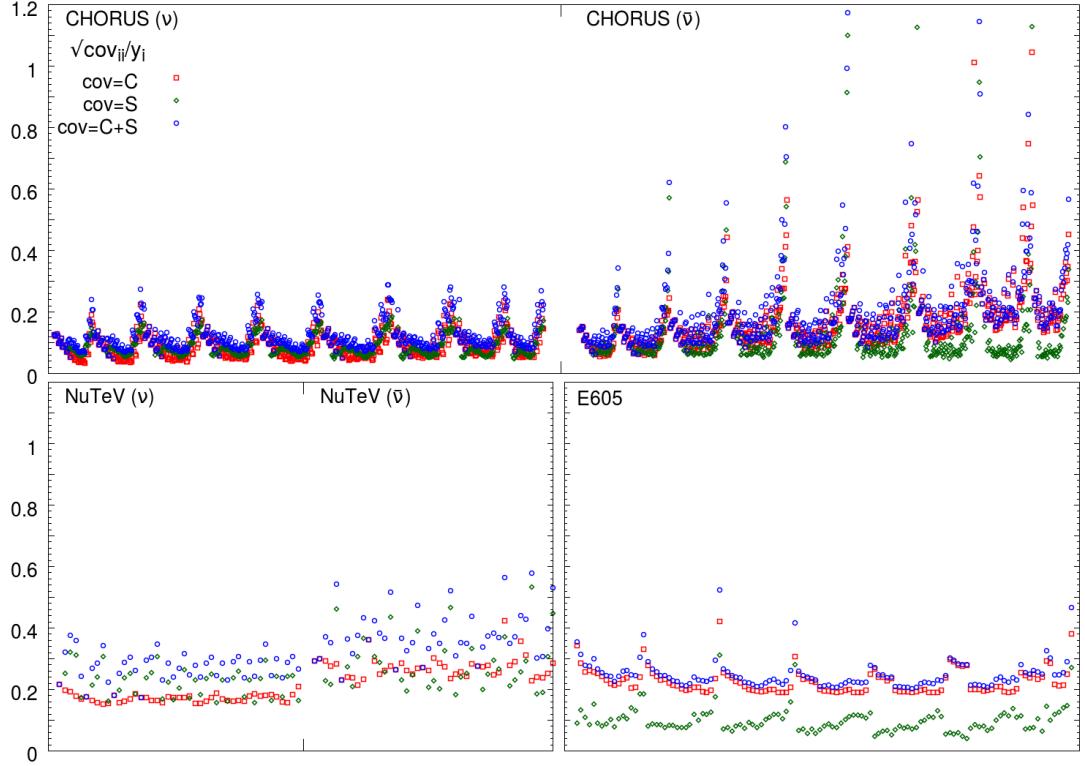


Figure 3.3.1 The square root of the diagonal elements of the covariance matrices, normalised to corresponding data. Experimental contributions are red, theory green and the total blue. Data from CHORUS and NuTeV are split into neutrino and anti-neutrino parts. Points are binned in (anti-)neutrino beam energy E : 25, 35, 45, 55, 70, 90, 110, 120, 170 GeV. In each bin x increases from left to right, $0.045 < x < 0.65$.

nuclear uncertainties, with

$$\Delta_i^{(n)} = T_i^N[f_N^{(n)}] - T_i^N[f_p]; \quad (3.3.1)$$

2. **Def. 2,** (a more ambitious approach) where a shift,

$$\delta T_i^N = T_i^N[f_N] - T_i^N[f_p], \quad (3.3.2)$$

is also applied to the corresponding observable, meaning that the uncertainty should be defined relative to the shifted value,

$$\Delta_i^{(n)} = T_i^N[f_N^{(n)}] - T_i^N[f_N]. \quad (3.3.3)$$

Whilst Def. 1 just deweights the nuclear data sets in a PDF fit, Def. 2 also attempts to directly apply a nuclear correction. In both cases we can construct a theoretical covariance matrix as

$$S_{ij} = \frac{1}{N_{rep}} \sum_{n=1}^{N_{rep}} \Delta_i^{(n)} \Delta_j^{(n)}. \quad (3.3.4)$$

We did this separately for each experiment, which is a conservative treatment.

Considering the diagonal elements of the covariance matrices (Fig. 4.4.2), we see that the nuclear uncertainty has the largest impact on the NuTeV data, where the nuclear uncertainties dominate the data uncertainties. This is mirrored in the off-diagonal elements (Fig. 4.4.1). Given the high correlation of NuTeV observables with the s and \bar{s} PDFs, the effect of including the uncertainties ought to be greatest for these PDFs.

3.4 The Impact on Global PDFs

We compared four different PDF fits:

- **Baseline**, based on NNPDF3.1, with small improvements [24];
- **NoNuc**, Baseline with nuclear data removed;
- **NucUnc**, Baseline with nuclear uncertainties according to Def. 1.
- **NucCor**, Baseline with nuclear uncertainties and a nuclear correction according to Def. 2.

Table 4.4.1 shows the variation in χ^2 for selected data sets ². All of the fits show reduced χ^2 compared to Baseline, highlighting tension due to nuclear data. However, the strange-sensitive ATLAS W/Z at 7 TeV (2011) measurements [5] still have a poor χ^2 , indicating that possible tensions with NuTeV were unlikely responsible for this; in any case, the data sets occupy different kinematic regions. The best fit is obtained for NucUnc, which has the largest uncertainties.

Fig. 4.5.1 shows the light sea quark PDFs for NucUnc compared to Baseline. These are the distributions with greatest impact, but there is little appreciable

²For a full break-down see [24].

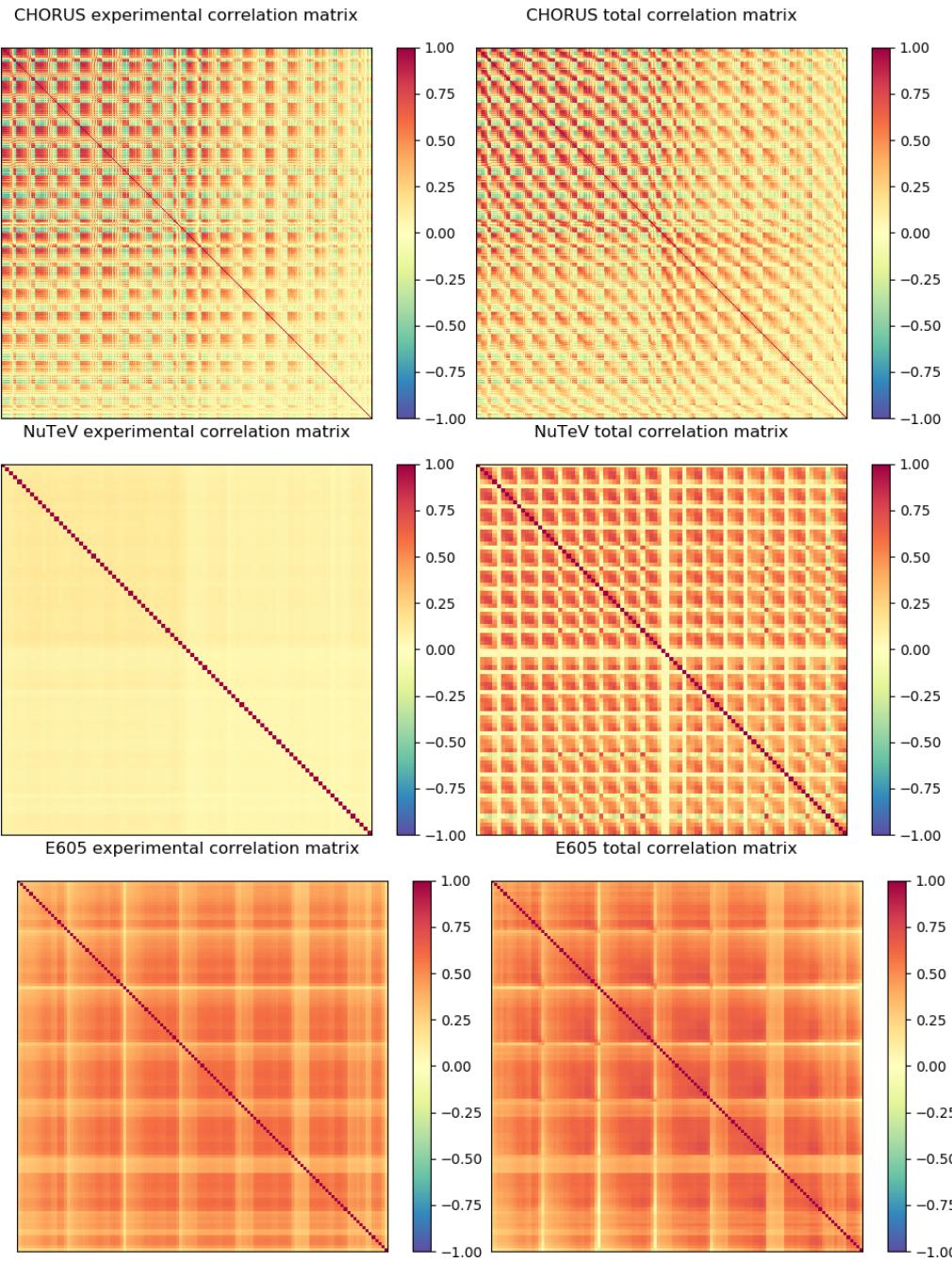


Figure 3.4.1 Correlation matrices, $\rho_{ij}^{cov} = \frac{cov_{ij}}{\sqrt{cov_{ii}cov_{jj}}}$, before (left) and after (right) including nuclear uncertainties. Data are binned the same as in Fig. 4.4.2. The top row corresponds to CHORUS, middle row to NuTeV and bottom row to E605. Results are displayed for Def. 1 but are qualitatively similar for Def. 2.

Experiment	N_{dat}	Baseline	NoNuc	NucUnc	NucCor
CHORUS ν	416	1.29	–	0.97	1.04
CHORUS $\bar{\nu}$	416	1.20	–	0.78	0.83
NuTeV ν	39	0.41	–	0.31	0.40
NuTeV $\bar{\nu}$	37	0.90	–	0.62	0.83
E605 σ^p	85	1.18	–	0.85	0.89
ATLAS W/Z (2011)	34	1.97	1.78	1.87	1.94
ATLAS	360	1.08	1.04	1.04	1.05
CMS	409	1.07	1.07	1.07	1.07
LHCb	85	1.46	1.27	1.32	1.37
Total	4285	1.18	1.14	1.07	1.09

Table 3.4.1 χ^2 per data point for selected data sets. The final row shows results for the full fitted data.

change other than a small shift in the central value and increase in uncertainties. NucCor behaves similarly. Overall, the nuclear uncertainties are small compared to the global experimental uncertainty.

3.5 Phenomenology

Given the changes to the light sea quark PDFs, it is interesting to examine the impact on relevant phenomenological quantities, namely: the sea quark asymmetry, \bar{u}/\bar{d} ; strangeness fraction, $R_s = (s + \bar{s})/(\bar{u} + \bar{d})$; and strange valence distribution, $xs^- = x(s - \bar{s})$ (Fig. 4.6.1). In all cases it is clear that removing the nuclear data has a significant effect, emphasising the need to retain this data in proton PDF fits. Adding nuclear uncertainties, however, makes very little difference. In particular, the known tension between ATLAS W/Z + HERA DIS data and NuTeV data, which is apparent in the strangeness fraction [9], is not relieved with the addition of nuclear uncertainties.

We found no appreciable difference between using NucUnc versus NucCor, so opt to incorporate uncertainties using NucUnc (Def. 1) as this is the more conservative option.

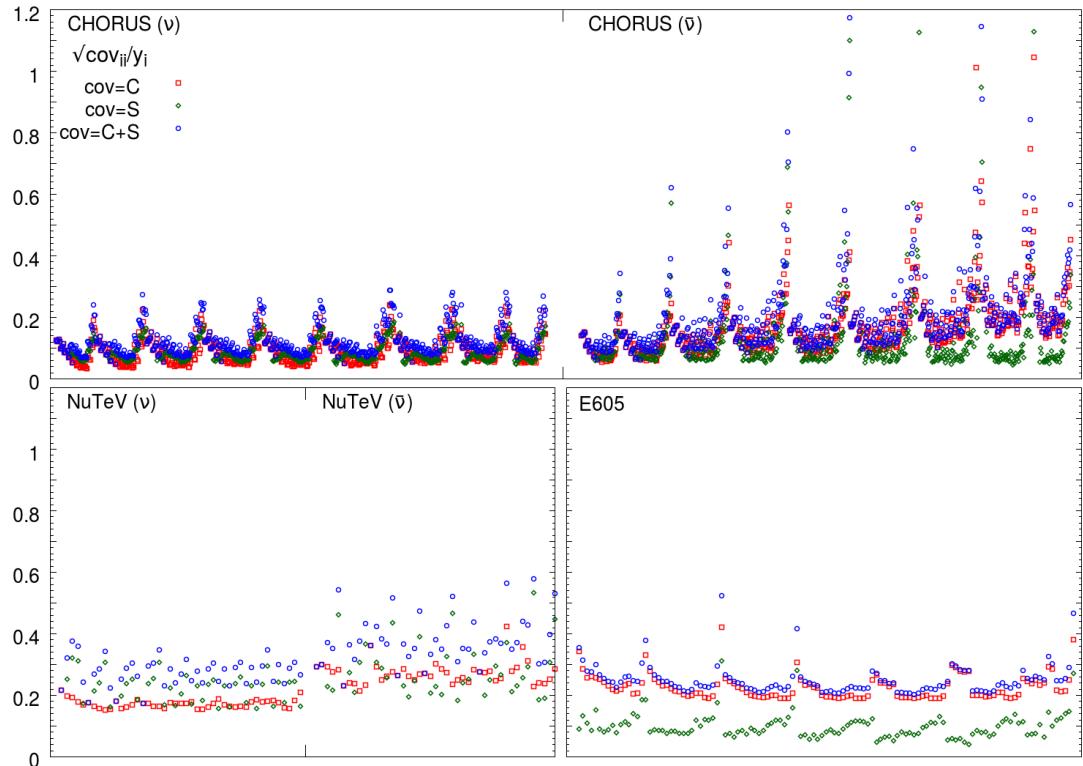


Figure 3.4.2 The square root of the diagonal elements of the covariance matrices, normalised to corresponding data. Experimental contributions are red, theory green and the total blue. Data from CHORUS and NuTeV are split into neutrino and anti-neutrino parts. Points are binned in (anti-)neutrino beam energy E : 25, 35, 45, 55, 70, 90, 110, 120, 170 GeV. In each bin x increases from left to right, $0.045 < x < 0.65$.

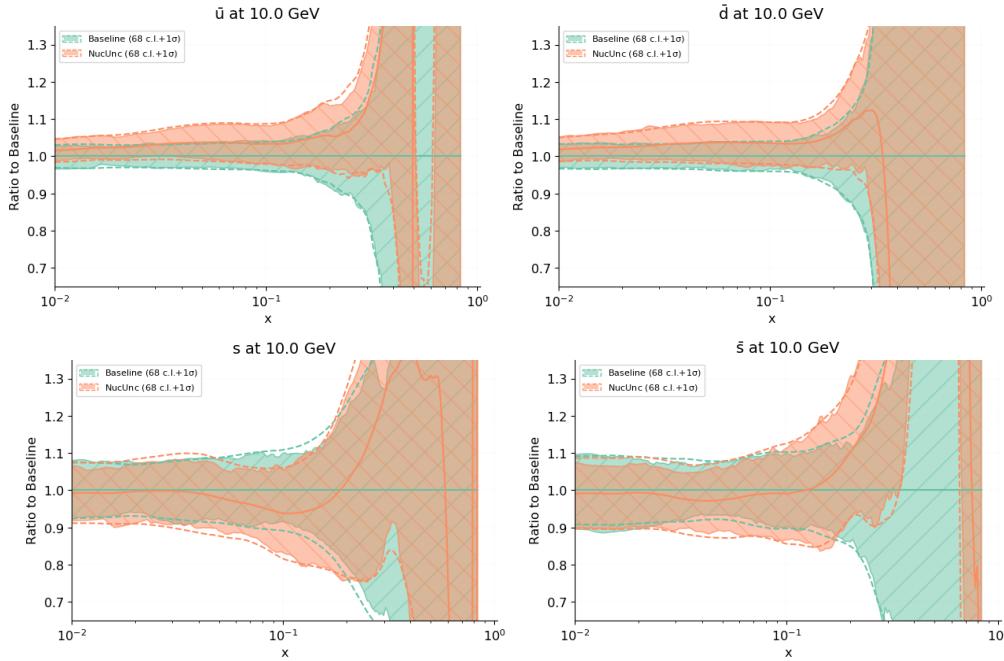


Figure 3.5.1 *NucUnc fits with nuclear uncertainties (orange) compared to Baseline (green) for PDFs at 10 GeV. Clockwise from top left: \bar{u} , \bar{d} , s and \bar{s} PDFs. Error bands are 1σ ; results are normalised to Baseline fit.*

3.6 Conclusions

We studied the role of nuclear data in proton PDF fits, and adopted an empirical approach to determine the nuclear uncertainties due to this data. We based our analysis on recent nPDF fits DSSZ12, nCTEQ15 and EPPS16. Using a theoretical covariance matrix, we included these uncertainties in proton PDF fits, and found that the fit quality was improved, with the largest effect on the light sea quark distributions³. We found no significant impact on associated phenomenology.

We will extend this analysis to deuterium data, and in the future we will be able to use nuclear PDFs from NNPDF [51] to estimate uncertainties. Furthermore, these methods can be applied to other sources of theoretical uncertainties, such as higher twist effects, fragmentation functions, and missing higher order uncertainties [?].

³The PDF sets from this analysis are available upon request from the authors in LHAPDF format [7].

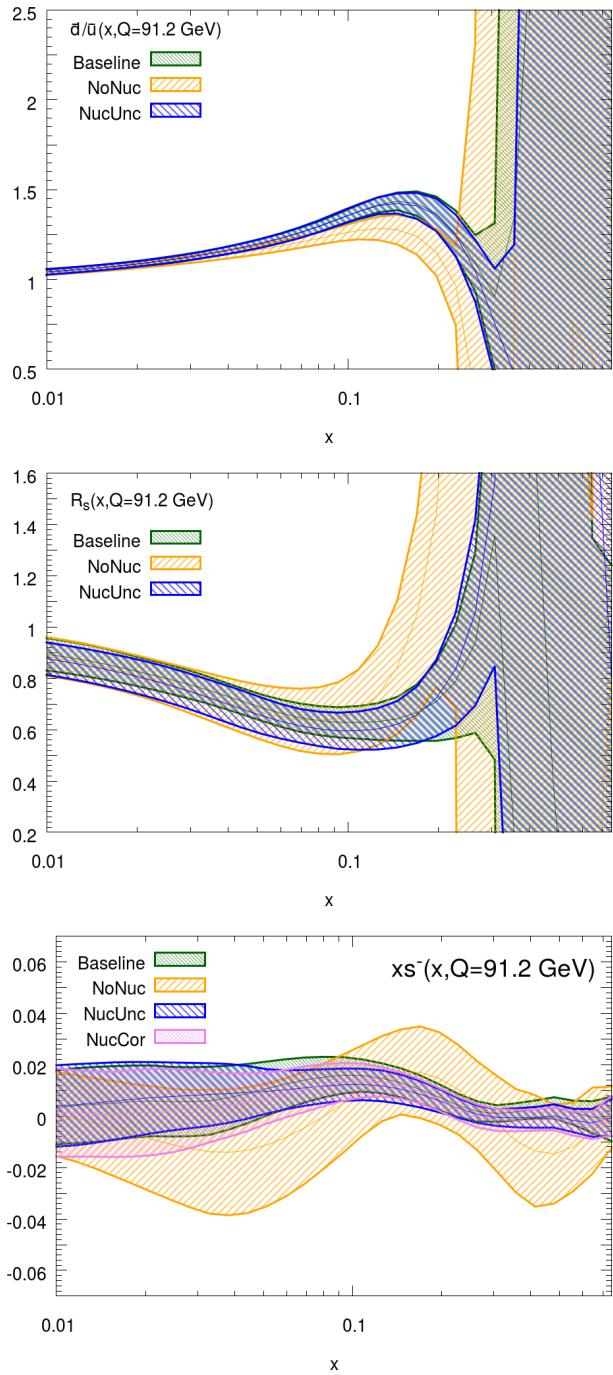


Figure 3.6.1 Effect of including nuclear uncertainties on phenomenology. From left to right: sea quark asymmetry, strangeness fraction, strange valence distribution. Distributions correspond to the use of different PDF fits: Baseline (green), NoNuc (yellow), NucUnc (blue) and NucCor(pink). $Q = 91.2 \text{ GeV}$. In the left two plots, NucCor are indistinguishable from NucUnc so are omitted for readability.

Chapter 4

Deuteron Uncertainties - 0%

Chapter 5

Higher Twist Uncertainties - 0%

5.1 The role of higher twist data in PDFs

5.2 Ansatz for a higher twist correction

5.3 Using a neural network to model higher twist

5.3.1 Model architecture

5.3.2 Training and validating the neural network

5.4 Form of the higher twist correction

5.5 The higher twist covariance matrix

5.6 PDFs with higher twist uncertainties

Chapter 6

Conclusion - 0%

Appendix A

Diagonalisation of the theory covariance matrix

Appendix B

PDF sets with different scale choices

Bibliography

- [1] , . <https://nnpdf.mi.infn.it/>.
- [2] , . <https://home.cern/>.
- [3] , . <https://www.fnal.gov/>.
- [4] https://www.desy.de/news/news_search/index_eng.html?openDirectAnchor=829.
- [5] Aad, G., et al. “Measurement of the inclusive W^\pm and Z/gamma cross sections in the electron and muon decay channels in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector.” *Phys. Rev.* D85: (2012) 072,004.
- [6] Aaltonen, T., et al. “Measurement of the Inclusive Jet Cross Section at the Fermilab Tevatron p anti-p Collider Using a Cone-Based Jet Algorithm.” *Phys. Rev.* D78: (2008) 052,006. [Erratum: Phys. Rev.D79,119902(2009)].
- [7] et al., A. B. “LHAPDF6: parton density access in the LHC precision era.” *Eur. Phys. J.* C75: (2015) 132.
- [8] et al., A. D. M. “Uncertainties on alpha(S) in global PDF analyses and implications for predicted hadronic cross sections.” *Eur. Phys. J.* C64: (2009) 653–680.
- [9] et al., G. A. “Determination of the strange quark density of the proton from ATLAS measurements of the $W \rightarrow \ell\nu$ and $Z \rightarrow \ell\ell$ cross sections.” *Phys. Rev. Lett.* 109: (2012) 012,001.
- [10] ———. “Measurement of the Z/γ^* boson transverse momentum distribution in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector.” *JHEP* 09: (2014) 145.
- [11] et al., G. O. “Measurement of nucleon structure functions in neutrino scattering.” *Phys. Lett.* B632: (2006) 65–75.
- [12] et al., J. G. “The Structure of the Proton in the LHC Precision Era.” .
- [13] et al., K. K. “nCTEQ15 - Global analysis of nuclear parton distributions with uncertainties in the CTEQ framework.” *Phys. Rev.* D93, 8: (2016) 085,037.

- [14] et al., M. G. “Precise measurement of dimuon production cross-sections in muon neutrino Fe and muon anti-neutrino Fe deep inelastic scattering at the Tevatron.” *Phys. Rev.* D64: (2001) 112,006.
- [15] et al., R. D. B. “Precision determination of electroweak parameters and the strange content of the proton from neutrino deep-inelastic scattering.” *Nucl. Phys.* B823: (2009) 195–233.
- [16] ———. “Parton distributions from high-precision collider data.” *Eur. Phys. J.* C77, 10: (2017) 663.
- [17] et al., S. C. “Vector boson production at hadron colliders: a fully exclusive QCD calculation at NNLO.” *Phys. Rev. Lett.* 103: (2009) 082,001.
- [18] et al., S. D. “New parton distribution functions from a global analysis of quantum chromodynamics.” *Phys. Rev.* D93, 3: (2016) 033,006.
- [19] et al., S. F. “Neural network parametrization of deep inelastic structure functions.” *JHEP* 05: (2002) 062.
- [20] et al., V. K. “Measurement of the Z boson differential cross section in transverse momentum and rapidity in proton–proton collisions at 8 TeV.” *Phys. Lett.* B749: (2015) 187–209.
- [21] Altarelli, G., and G. Parisi. “Asymptotic Freedom in Parton Language.” *Nucl. Phys. B* .
- [22] Alwall, J., R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. “The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations.” *JHEP* 07: (2014) 079.
- [23] Arneodo, M. “Nuclear effects in structure functions.” *Phys. Rept.* 240: (1994) 301–393.
- [24] Ball, R. D., and A. Deshpande. “The Proton Spin, Semi-Inclusive processes, and a future Electron Ion Collider.” <https://inspirehep.net/record/1648159/files/arXiv:1801.04842.pdf>.
- [25] Ball, R. D., V. Bertone, L. Del Debbio, S. Forte, A. Guffanti, J. Rojo, and M. Ubiali. “Theoretical issues in PDF determination and associated uncertainties.” *Phys. Lett.* B723: (2013) 330–339.
- [26] Ball, R. D., L. Del Debbio, S. Forte, A. Guffanti, J. I. Latorre, A. Piccione, J. Rojo, and M. Ubiali. “A Determination of parton distributions with faithful uncertainty estimation.” *Nucl. Phys.* B809: (2009) 1–63. [Erratum: Nucl. Phys.B816,293(2009)].
- [27] Ball, R. D., L. Del Debbio, S. Forte, A. Guffanti, J. I. Latorre, J. Rojo, and M. Ubiali. “Fitting Parton Distribution Data with Multiplicative Normalization Uncertainties.” *JHEP* 05: (2010) 075.

- [28] Botje, M. “Lecture notes Particle Physics II: Quantum Chromo Dynamics.”, 2013. <https://www.nikhef.nl/~h24/qcdcourse/section-all.pdf>.
- [29] Butterworth, J., et al. “PDF4LHC recommendations for LHC Run II.” *J. Phys.* G43: (2016) 023,001.
- [30] Callan, J., Curtis G., and D. J. Gross. “Bjorken scaling in quantum field theory.” *Phys. Rev. D* 8: (1973) 4383–4394.
- [31] Campbell, J., and R. Ellis. “An update on vector boson pair production at hadron colliders.” *Phys. Rev.* .
- [32] Cooper-Sarkar, A. M. “PDF Fits at HERA.” *PoS* EPS-HEP2011: (2011) 320.
- [33] D’Agostini, G. “On the use of the covariance matrix to fit correlated data.” *Nucl. Instrum. Meth. A* 346: (1994) 306–311.
- [34] Dokshitzer, Y. L. *Sov. Phys.* .
- [35] Dothan, Y., M. Gell-Mann, and Y. Ne’eman. “Series of Hadron Energy Levels as Representations of Noncompact Groups.” *Phys. Lett.* 17: (1965) 148–151.
- [36] Ellis, R. K. R. K. *QCD and collider physics*. Cambridge monographs on particle physics, nuclear physics, and cosmology ; 8. Cambridge ; New York: Cambridge University Press, 1996.
- [37] Eskola, K. J., P. Paakkinen, H. Paukkunen, and C. A. Salgado. “EPPS16: Nuclear parton distributions with LHC data.” *Eur. Phys. J.* C77, 3: (2017) 163.
- [38] Feynman, R. P. “Very high-energy collisions of hadrons.” *Phys. Rev. Lett.* 23: (1969) 1415–1417.
- [39] Feynman, R. “The behavior of hadron collisions at extreme energies.” *Conf. Proc. C* 690905: (1969) 237–258.
- [40] ———. “Photon-hadron interactions.” .
- [41] de Florian, D., R. Sassot, P. Zurita, and M. Stratmann. “Global Analysis of Nuclear Parton Distributions.” *Phys. Rev.* D85: (2012) 074,028.
- [42] G. Aad, G. e. a. “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.* C76, 5: (2016) 291.
- [43] Gavin, R., Y. Li, F. Petriello, and S. Quackenbush. “FEWZ 2.0: A code for hadronic Z production at next-to-next-to-leading order.” *Comput. Phys. Commun.* 182: (2011) 2388–2403.

- [44] Geiger, D. H., and E. Marsden. “LXI. The laws of deflexion of a particles through large angles.” *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 25, 148: (1913) 604–623. <https://doi.org/10.1080/14786440408634197>.
- [45] Gell-Mann, M. “Symmetries of baryons and mesons.” *Phys. Rev.* 125: (1962) 1067–1084.
- [46] ———. “A Schematic Model of Baryons and Mesons.” *Phys. Lett.* 8: (1964) 214–215.
- [47] Greenberg, O. “Spin and Unitary Spin Independence in a Paraquark Model of Baryons and Mesons.” *Phys. Rev. Lett.* 13: (1964) 598–602.
- [48] Gribov, L. N., and V. N. Lipatov. *Sov. J. Nucl. Phys.* .
- [49] Grinstein, B. “Introductory lectures on QCD.” .
- [50] Halzen, F. F. *Quarks and leptons : an introductory course in modern particle physics.* New York] ; [Chichester]: John Wiley, 1984.
- [51] Khalek, R. A., J. J. Ethier, and J. Rojo. “Nuclear Parton Distributions from Neural Networks.” In *Diffraction and Low-x 2018 (Diffflowx2018) Reggio Calabria, Italy, August 26-September 1, 2018.* 2018.
- [52] Kinoshita, T. “Mass singularities of Feynman amplitudes.” *J. Math. Phys.* 3: (1962) 650–677.
- [53] Lee, T., and M. Nauenberg. “Degenerate Systems and Mass Singularities.” *Phys. Rev.* 133: (1964) B1549–B1562.
- [54] Nocera, E. R., M. Ubiali, and C. Voisey. “Single Top Production in PDF fits.” *JHEP* 05: (2020) 067.
- [55] S. Alekhin, J. B., and S. Moch. “The ABM parton distributions tuned to LHC data.” *Phys. Rev.* D89, 5: (2014) 054,028.
- [56] S. Catani, S., and M. H. Seymour. “A General algorithm for calculating jet cross-sections in NLO QCD.” *Nucl. Phys.* B485: (1997) 291–419. [Erratum: *Nucl. Phys.*B510,503(1998)].
- [57] Tzanov, M., et al. “Precise measurement of neutrino and anti-neutrino differential cross sections.” *Phys. Rev.* D74: (2006) 012,008.
- [58] Zweig, G. *An SU(3) model for strong interaction symmetry and its breaking. Version 2*, 1964, 22–101.