Theory Uncertainties in Parton Distribution Functions

Rosalyn Laura Pearson

Doctor of Philosophy
The University of Edinburgh
July 2021

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 involve the interaction of protons, which are composite particles. These interactions are described by combining perturbative Quantum Chromodynamics (QCD) 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 QCD. Uncertainties in PDFs arise from those in the experimental data and theoretical predictions, as well as from the fitting procedure itself. Those in the theory come from many sources. Here we consider two of the most important: the first are missing higher order uncertainties (MHOUs), arising due to truncating the predictions' perturbative expansion; the second are nuclear uncertainties, due to difficulty making predictions in a nuclear environment.

In this thesis we consider how to include theory uncertainties in PDF fits by constructing a theory covariance matrix and adding this to the experimental one. MHOUs are estimated and included as a proof of concept in next-to-leading order PDFs. We find that these capture many of the important features of the known PDFs at the next order above. We then investigate nuclear uncertainties, estimate their magnitude and assess their impact on the PDFs. Finally, we consider how to make predictions with theory uncertainties using PDFs which also include theoretical uncertainties. Here there are correlations between the PDFs and the predictions, which can lead to a shift in the predictions and their uncertainties, which can significantly improve their accuracy and precision.

Declaration

This thesis was composed by myself, and details work carried out as a member of the Neural Network Parton Distribution Function 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 the following places:

1. **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]

2. **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]

3. **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]

4. **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

5. **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]

6. **Deuteron Uncertainties in the Determination of Proton PDFs**, Richard D. Ball, Emanuele R. Nocera & Rosalyn L. Pearson (Oct 30, 2020) Published in: Eur.Phys.J.C 81 (2021) 1, 37, e-Print: 2011.00009 [hep-ph]

7. **Correlation of Theoretical Uncertainties in PDF Fits and Theoretical Uncertainties in Predictions**, Richard D. Ball & Rosalyn L. Pearson (July 22, 2021), e-Print: 2105.05114v2 [hep-ph]

8. **Next generation proton PDFs with deuteron and nuclear uncertainties**, Rosalyn L. Pearson, Richard D. Ball & Emanuele R. Nocera (Jun 23, 2021), e-Print: 2106.12349 [hep-ph]

In particular, Chapters 2 and 3 are based on 5, Chapter 4 draws on on 2, 6 and 7, and Chapter 5 is based on 8.

Unless stated otherwise, I have generated the figures myself, many of which were made using the `ReportEngine` software [1]. Some of these are also published in the above works.

(Rosalyn Laura Pearson, July 2021)

Acknowledgements

First and foremost I would like to thank my supervisor, Richard Ball, for his support and guidance in this project, and for steering me towards a fascinating and cohesive area of research. I have benefited very much from his insightful perspective, and meticulous understanding of the interplay of ingredients in PDF fits. Important thanks also go to Emanuele Nocera, who helped me find my feet in the early days of my PhD, and has been immensely helpful and supportive throughout, above and beyond what I could have hoped for. Without him I am not sure I would be writing this thesis! Also to Zahari Kassabov for his perpetual presence on GitHub, answering thousands of questions and helping me to develop good coding practices.

I have also met some lovely people in NNPDF, of which there are too many to list here. But in particular I would like to mention Mikey and Tommaso, the Edinburgh students in my year. They have helped me out more times than I can count with my research and with the code development, and I was very happy to share an office with them. I also really enjoyed collaborating closely with Cameron during the first couple of years. Thanks also go to Rabah and Shayan. Finally, I would like to thank Emma for helping me stick through the tougher moments, and providing mentorship from the year above.

Outwith my PhD I would like to thank my sister Julia, my partner Fred and my best friend Eleanor, all of whom I have also lived with. In particular, thanks to Eleanor for her help with the artwork for some of the figures.

Contents

List of Figures

List of Tables

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.
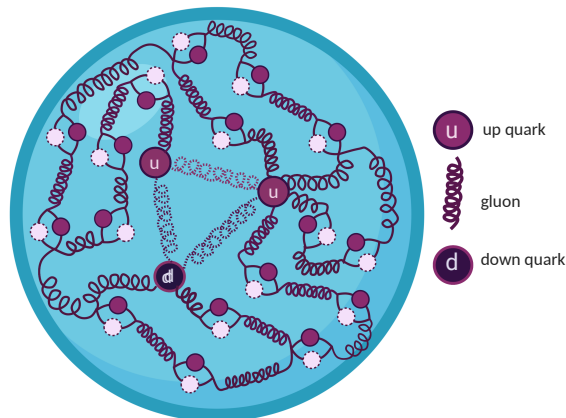
Figure 0.0.1   *A visualisation of the internal structure of the proton. Quarks are bound together by gluons. Artist: Eleanor Conole.*

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) [5] at CERN, and new

colliders are anticipated in coming years [6–8]. 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, which we describe using Quantum Chromodynamics (QCD). However, the proton (Fig. 0.0.1) is in reality highly complicated and the QCD calculations for it are intractable. This means protons must currently 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 are collected at the hadronic level, whereas theoretical predictions using QCD 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.

QCD can't give us the PDFs directly, but crucially they are process independent. This means that they can be determined in a global fit between multi-process experimental data and corresponding theoretical predictions. Fig. 0.0.2 shows the fitted functional form of the PDFs in NNPDF4.0 [9]. Once these have been determined, they can be used to make predictions for any observable involving protons.

Because the PDFs must be inferred from measurements and our current knowledge of QCD, they are not exactly known. The uncertainties in their form then propagate through to predictions made using them. There are three places these uncertainties can be introduced:

1. through uncertainties in the experimental measurements;

2. through uncertainties in the input theory;

3. through the fitting procedure.

Until recently, experimental uncertainties were the dominant source, meaning that theory uncertainties have been largely ignored in standard PDF fits. However, with the onset of increasingly high precision experiments and the corresponding drive of PDFs down to 1% accuracy [10, 11], a proper treatment of theory uncertainties is becoming pressing.
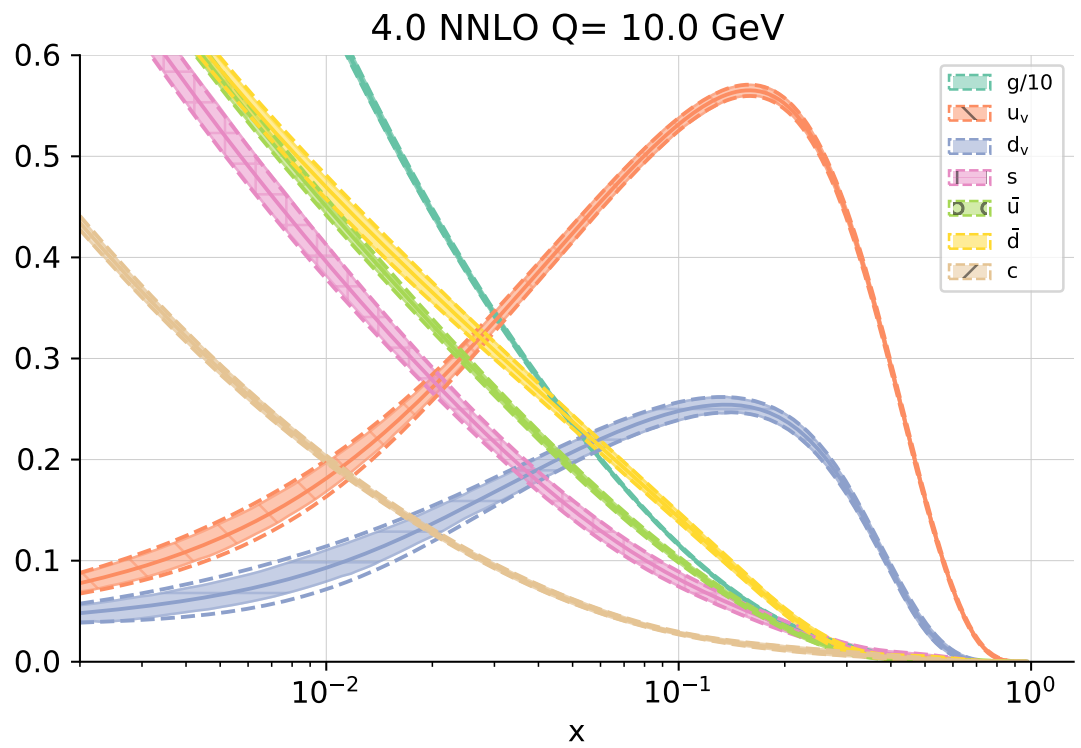
Figure 0.0.2 *The different PDF flavours determined in the latest NNPDF4.0 determination.*

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 neccesary 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, 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. We distinguish between the NNPDF3.1 methodology, on which the results in Chapters 3 and 5 are based, and the NNPDF4.0 methodology, on which the results in Chapter 4 are based. We note that the difference between these methodologies doesn't affect the main lines of argument in these chapters.

1.1 Physics background

1.1.1 Deep inelastic scattering

For a more in-depth analysis, see [12, 13]. In this section we rely heavily on [14, 15].

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-Marsden experiments from 1908-1913 and the subsequent uncovering of the atomic nucleus [16]. 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 are not point-like particles and so a series of important experiments were carried out in the 1960s at the Stanford 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. 1.1.1. 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.

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

$$|\mathcal{M}|^2 = \mathcal{N}_1 \frac{\alpha^2}{q^4} L_{\mu\nu} W^{\mu\nu}, \qquad (1.1.1)$$

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.
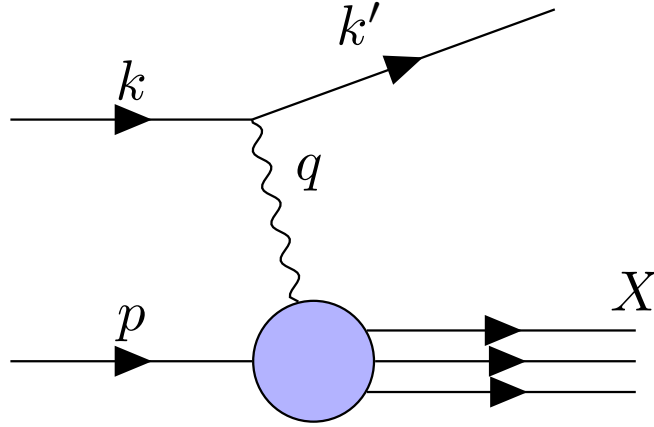
Figure 1.1.1 *Deep inelastic scattering.*

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

$$\begin{aligned}
L_{\mu\nu} &= \sum_{spins} \bar{u}(k')\gamma_\mu u(k)\bar{u}(k)\gamma_\nu u(k') \\
&= Tr(k\!\!\!/'\gamma_\mu k\!\!\!/\gamma_\nu) \\
&= \mathcal{N}_2\left(k_\mu k'_\nu + k_\nu k'_\mu - g_{\mu\nu}k\cdot k' \right) \\
&= \mathcal{N}_2\left(4k_\mu k_\nu - 2k_\mu q_\nu - 2k_\nu q_\mu + g_{\mu\nu}q^2 \right),
\end{aligned} \qquad (1.1.2)$$

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

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

(1.1.3)

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

$$\begin{aligned}
F_1(x, Q^2) &= W_1(p, q); \\
F_2(x, Q^2) &= \nu W_2(p, q); \\
F_L(x, Q^2) &= F_2(x, Q^2) - 2xF_1(x, Q^2),
\end{aligned}$$

(1.1.4)

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. 1.1.2 and 1.1.3 in Eqn. 1.1.1, 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:

$$|\mathcal{M}|^2 = \mathcal{N}_1 \mathcal{N}_2^2 \frac{\alpha^2}{q^4}\left\{(-2q^2)W_1(p, q) + \left(4(p \cdot k)^2 - 4(p \cdot q)(p \cdot k)\right)W_2(p, q)\right\}. \quad (1.1.5)$$

1.1.2 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 [17]. 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" [18–20].

Furthermore, $F_L(x)$ was measured to be 0, known as the Callan-Gross relation [17, 21], which suggests that the point-like particles could not absorb longitudinal photons. This fitted in nicely with the quark models developed shortly before

[22–25], 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. 1.1.2.
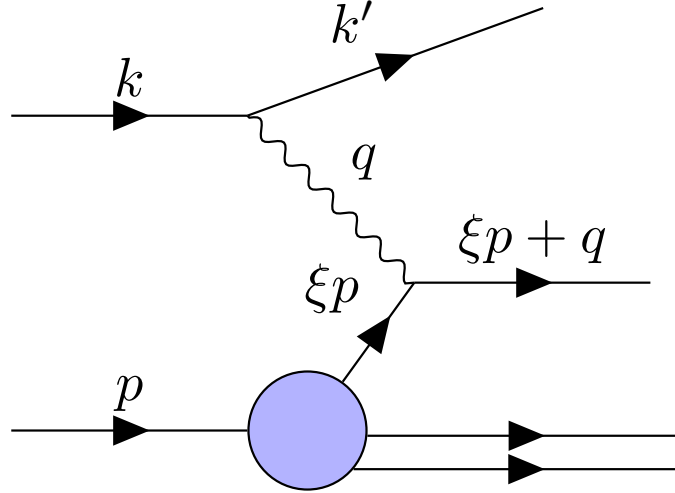
Figure 1.1.2 *DIS in the parton model. One parton with momentum p interacts with the virtual photon, and the other partons "spectate".*

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:

$$(\xi p + q)^2 = 0$$
$$\implies 2\xi p \cdot q + q^2 = 0$$
$$\implies 2\xi p \cdot q - Q^2 = 0 \tag{1.1.6}$$
$$\implies \xi = \frac{Q^2}{2p \cdot q} \equiv x.$$

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, $\hat{W}^q_{\mu\nu}$, as

$$W_{\mu\nu} = \sum_q f_q(x)\hat{W}^q_{\mu\nu}\delta(Q^2 - 2xp\cdot q) = \frac{1}{Q^2}\sum_q f_q(x)\hat{W}^q_{\mu\nu}\delta(1 - \frac{2xp\cdot q}{Q^2}), \quad (1.1.7)$$

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 $x \to x + dx$. 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 1.1.1, we can see that

$$|\mathcal{M}|^2 = \frac{1}{Q^2}\sum_q f_q(x)|\hat{\mathcal{M}}_q|^2. \quad (1.1.8)$$

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. 1.1.2, and the muon has the same, but with the substitutions $k \to p$ and $q \to -q$. Once again we can use $q_\mu L^{\mu\nu} = 0$ and the expression

$$|\mathcal{M}_{(e\mu)}|^2 = \mathcal{N}_1\frac{\alpha^2}{q^4}L^{(e)}_{\mu\nu}L^{\mu\nu}_{(\mu)} \quad (1.1.9)$$

to show (in the massless limit)

$$|\mathcal{M}_{(e\mu)}|^2 = \mathcal{N}_1\mathcal{N}_2^2\frac{\alpha^2}{q^4}\left(16(p\cdot k)^2 + 8q^2(p\cdot k) + 2q^4\right). \quad (1.1.10)$$

Using the symmetry of Fig. 1.1.2, we can see this is analogous to $|\hat{\mathcal{M}}_q|^2$ under the substitution $p \to xp$, provided we replace the charge of the electron, e, with that

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

$$
\begin{aligned}
|\hat{\mathcal{M}}_q|^2 &= \mathcal{N}_1 \mathcal{N}_2^2 \frac{e_q^2 \alpha^2}{q^4} \left\{ 4(2xp \cdot k)^2 + 4(2xp \cdot k)q^2 + 2q^4 \right\} \\
&= \mathcal{N}_1 \mathcal{N}_2^2 \frac{e_q^2 \alpha^2}{Q^4} \left\{ 4 \left(\frac{Q^2}{y} \right)^2 - 4 \left(\frac{Q^2}{y} \right) Q^2 + 2Q^4 \right\} \\
&= \mathcal{N}_1 \mathcal{N}_2^2 e_q^2 \alpha^2 \left\{ 2 + 4 \left(\frac{1-y}{y^2} \right) \right\}.
\end{aligned}
\tag{1.1.11}
$$

Now we can use this alongside Eqn. 1.1.5 in Eqn. 1.1.8, giving us

$$
\begin{aligned}
F_1 &\equiv W_1 = \sum_q f_q(x) e_q^2, \\
F_2 &\equiv \nu W_2 = 2x \sum_q f_q(x) e_q^2.
\end{aligned}
\tag{1.1.12}
$$

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 $Q^2 \to \infty$, 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).

1.1.3   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 [26].

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" [22–25] . The additional $SU(3)_c$ colour symmetry was put forwards in order that the quarks satisfied Fermi-Dirac statistics [27]. 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

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}^a F^{a\mu\nu} + \bar{q}^i(i\not{D}_i^j - m\delta_i^j)q_j, \tag{1.1.13}$$

where the covariant derivative is

$$\mathcal{D}_\mu \psi(x) = (\partial_\mu - i\sqrt{4\pi\alpha_s}T^a A_\mu^a)\psi(x), \tag{1.1.14}$$

and the field strength tensor is

$$F_{\mu\nu}^a = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a + \sqrt{4\pi\alpha_s}f^{abc}A_\mu^b A_\nu^c. \tag{1.1.15}$$

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 [12].

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

$$V(r) \sim \frac{\alpha}{r} + kr, \tag{1.1.16}$$

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. $\alpha_s \to \alpha_s(Q^2)$), which satisfies

$$Q^2 \frac{\partial \alpha_s}{\partial Q^2} = \beta(\alpha_s), \tag{1.1.17}$$

The beta function, $\beta(\alpha_s)$, can be expressed perturbatively as an expansion in α_s and is currently known to N^{3}LO.

At one-loop order the solution of this equation is

$$\alpha_s(Q^2) = \frac{\alpha_s(\mu_R^2)}{1 + \beta_0 \alpha_s(\mu_R^2) \ln\left(\frac{Q^2}{\mu_R^2}\right)}, \tag{1.1.18}$$

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 (Eqn. 1.1.17) 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 infrared 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. Unforunately, at the partonic level, structure functions and cross sections are not infrared safe.

1.1.4 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 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 infrared 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 infrared-safe observables the soft and collinear divergences exactly canel [28, 29], 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 [30], 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.
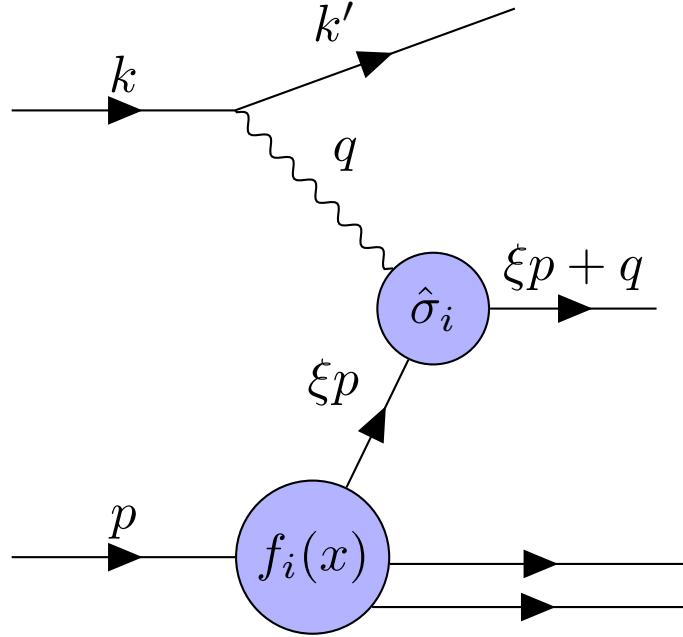
Figure 1.1.3 *Factorisation and the QCD improved parton model*

We can write a DIS cross section as

$$\sigma^{DIS} = \sum_i \int dx f_i(x, \mu_F^2) \hat{\sigma}_i \left(x, \frac{Q^2}{\mu_F^2} \right), \qquad (1.1.19)$$

corresponding to Fig 1.1.3.

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. 1.1.4. 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 $z = x/y$.
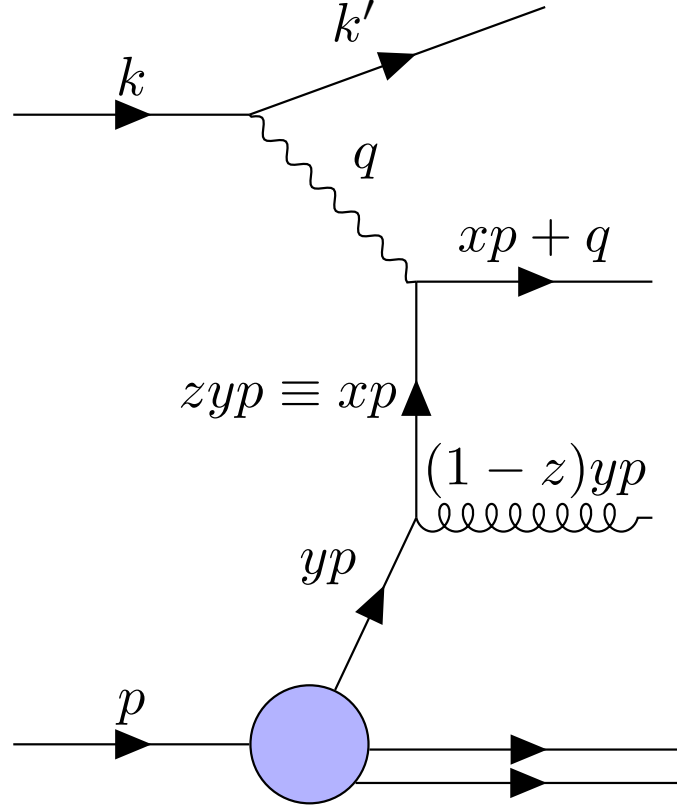
Figure 1.1.4   *A quark radiating a gluon before interacting.*

It transpires (see [13] for the derivation) that the structure function F_2 can be expressed as

$$\frac{F_2(x, Q^2)}{x} = \sum_i e_i^2 \int_x^1 \frac{dy}{y} f_i(y) \left[\delta\left(1 - \frac{x}{y}\right) + \frac{\alpha_s}{2\pi} \mathcal{P}_{qq}\left(\frac{x}{y}\right) \ln\left(\frac{Q^2}{m^2}\right) \right]. \quad (1.1.20)$$

m is a cutoff introduced to regularise the collinear divergence and you can see that as $m \to 0$ the structure function diverges. A divergence also occurs for $(1-z) \to 0$, 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

$$\mathcal{P}_{qq} = \frac{4}{3}\left(\frac{1+z^2}{1-z}\right). \qquad (1.1.21)$$

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

$$\mathcal{I}_{qq}^i(x) \equiv \frac{\alpha_s}{2\pi}\int_x^1 \frac{dy}{y} f_i(y)\mathcal{P}_{qq}\left(\frac{x}{y}\right), \qquad (1.1.22)$$

and separating 1.1.20 into a singular part and a calculable part, like

$$\frac{F_2(x,Q^2)}{x} = \sum_i e_i^2\left[f_i(x) + \mathcal{I}_{qq}^i(x)\ln\left(\frac{\mu_F^2}{m^2}\right) + \mathcal{I}_{qq}^i(x)\ln\left(\frac{Q^2}{\mu_F^2}\right)\right]. \qquad (1.1.23)$$

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

$$f_i(x,\mu_F^2) = f_i(x) + \mathcal{I}_{qq}^i(x)\ln\left(\frac{\mu_F^2}{m^2}\right), \qquad (1.1.24)$$

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. Noting that at leading order $f_i(y) = f_i(y,\mu_F^2)$, we are able to write

$$\frac{F_2(x,Q^2)}{x} = \sum_i e_i^2\left[f_i(x,\mu_F^2) + \frac{\alpha_s}{2\pi}\int_x^1\frac{dy}{y}f_i(y,\mu_F^2)\mathcal{P}_{qq}\left(\frac{x}{y}\right)\ln\left(\frac{Q^2}{\mu_F^2}\right)\right] + \mathcal{O}(\alpha_s^2). \qquad (1.1.25)$$

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

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

$$(1.1.26)$$

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

$$\frac{\partial f_i(x,\mu_F^2)}{\partial \ln \mu_F^2} = \frac{\alpha_s}{2\pi} \int_x^1 \frac{dy}{y} f_i(y,\mu_F^2)\mathcal{P}_{qq}\left(\frac{x}{y}\right). \tag{1.1.27}$$

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 [31–33]:

$$\frac{\partial f_i}{\partial \ln \mu_F^2} = \sum_i \frac{\alpha_s}{2\pi} \mathcal{P}_{ij} \otimes f_j, \tag{1.1.28}$$

where we have used the Mellin convolution, defined

$$\mathcal{P} \otimes f \equiv \int_x^1 \frac{dy}{y} \mathcal{P}\left(\frac{x}{y}\right) f(y,\mu_F^2), \tag{1.1.29}$$

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.

The DGLAP equations are commonly dealt with in Mellin space where the convolution is transformed into a product; the Mellin transform from x-space to Mellin space is defined as

$$M(n) \equiv \int_0^1 dx\ x^{n-1} M(x). \tag{1.1.30}$$

Considering the Mellin transform of the RHS of the DGLAP equation, and suppressing the parton indices, i, for clarity, we have

$$
\begin{aligned}
\int_0^1 dx\ x^{n-1} & \int_x^1 \frac{dy}{y} \frac{\alpha_s}{2\pi}\ \mathcal{P}\left(\frac{x}{y}\right) f(y,\mu_F^2) \\
&= \int_0^1 dx\ x^{n-1}\left[\int_0^1 dy \int_0^1 dz\ \frac{\alpha_s}{2\pi}\ \mathcal{P}(z)f(y)\delta(x-yz)\right] \\
&= \frac{\alpha_s}{2\pi} \int_0^1 dz\ z^{n-1}\ \mathcal{P}(z) \int_0^1 dy\ y^{n-1}f(y) \\
&= \frac{\alpha_s}{2\pi}\mathcal{P}(n)\ f(n) \\
&\equiv \gamma(n)\ f(n),
\end{aligned}
\tag{1.1.31}
$$

where $\gamma(n)_{ij}$ (with parton indices explicit) are known as the anomalous dimensions, and are calculable order by order in perturbation theory.

1.1.5 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. 1.1.5. Because two protons are involved the expression for the cross section is the natural extension of the DIS case (Eqn. 1.1.19):

$$\sigma = \sum_{i,j} \int dx_1 dx_2 f_i(x_1, \mu_F^2) f_j(x_2, \mu_F^2) \hat{\sigma}_{ij}\left(x_1, x_2, \frac{Q^2}{\mu_F^2}, ...\right). \qquad (1.1.32)$$
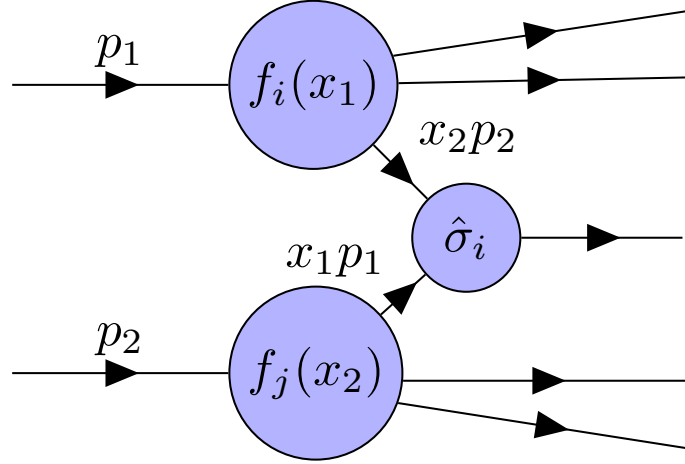
Figure 1.1.5 *Factorisation in hadron-hadron collisions.*

1.1.6 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" [12]. Intuitively, adding up all the momenta of the partons must equal the momentum of the proton. This enforces the condition

$$\int_0^1 dx \sum_i x f_i(x, Q^2) = 1. \qquad (1.1.33)$$

The other thing we know about the proton is that it is made up of two up and one down "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:

$$\int_0^1 dx(f_u - f_{\bar{u}}) = 2; \tag{1.1.34a}$$

$$\int_0^1 dx(f_d - f_{\bar{d}}) = 1; \tag{1.1.34b}$$

$$\int_0^1 dx(f_q - f_{\bar{q}}) = 0, \qquad q = s, c, t, b. \tag{1.1.34c}$$

Note that these conditions require that the PDFs are integrable.

1.2 Methodological background

In this section we review the necessary background for PDF determination within the NNPDF [34] framework. The results in this thesis are based on two different versions of this: NNPDF3.1 and NNPDF4.0. First we touch on the experimental and theoretical inputs to PDF fits, which are common to both versions. Then we outline the NNPDF3.1 framework, which was used to generate the results in Chapters 3 and 5. We summarise the NNPDF fitting strategy, and detail information on neural networks specific to this context. Finally, we explain the main differences between this and NNPDF4.0, which was used to generate the results in Chapter 4.

1.2.1 Experimental and theoretical input

NNPDF uses a variety of experimental data from a number of particle colliders, including those based at CERN [35] and Fermilab [36]. These are observables such as cross sections, differential cross sections and structure functions. Fig. 1.2.1 is a plot of the (x, Q^2) range spanned by the datasets in the upcoming NNPDF4.0 release, with those in NNPDF3.1 [9] shown without a black edge. Much 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
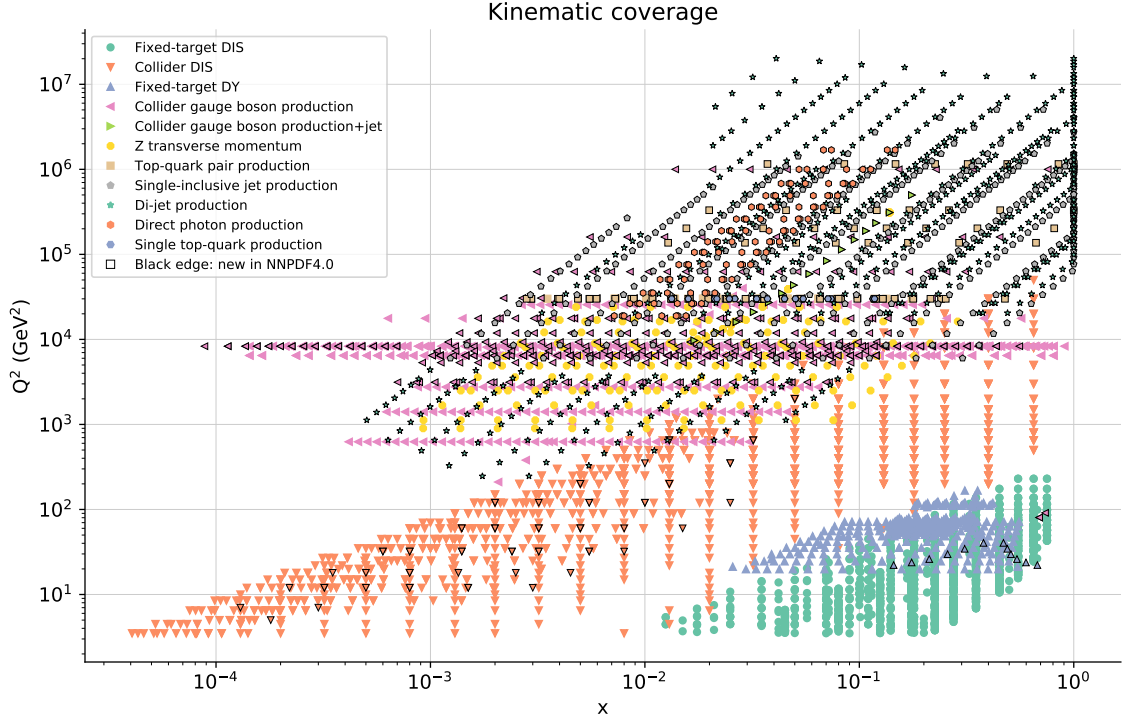
Figure 1.2.1 *Plot of the (x, Q^2) range spanned by data included in the upcoming NNPDF4.0 NLO fit. Datasets in NNPDF3.1 are those without the black edge.*

including $t\bar{t}$ production and high energy jets. For a full review of the data in both NNPDF3.1 and NNPDF4.0 see [37].

Theoretical predictions of the corresponding parton-level observables are computed using external codes [38–42]. These are converted to higher orders of perturbation theory as necessary using QCD and electroweak correction factors (known as "c" or "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. 1.1.28). This evolution is done using `APFEL` [43].

1.2.2 Experimental uncertainties

Experimental uncertainties are described using a covaraince matrix, C_{ij}, which gives the uncertainties and correlations between each of the data points $i, j = 1, ..., N_{dat}$. 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):

$$C_{ij} = \delta_{ij}\sigma_i^{uncorr}\sigma_j^{uncorr} + \sum_a \sigma_{i,a}\sigma_{j,a} + \left(\sum_m \sigma_{i,m}\sigma_{j,m}\right)D_iD_j, \qquad (1.2.1)$$

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 uncertaintes 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. 1.2.2 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,

$$\chi^2 = \frac{1}{N_{dat}}(D_i - T_i)C_{ij}^{-1}(D_j - T_j), \qquad (1.2.2)$$

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 [44] due to the presence of normalisation uncertainties. To avoid this, NNPDF employ the iterative $t0$ procedure [45] whereby D_i in Eqn. 1.2.1 are replaced initially with the predictions from a
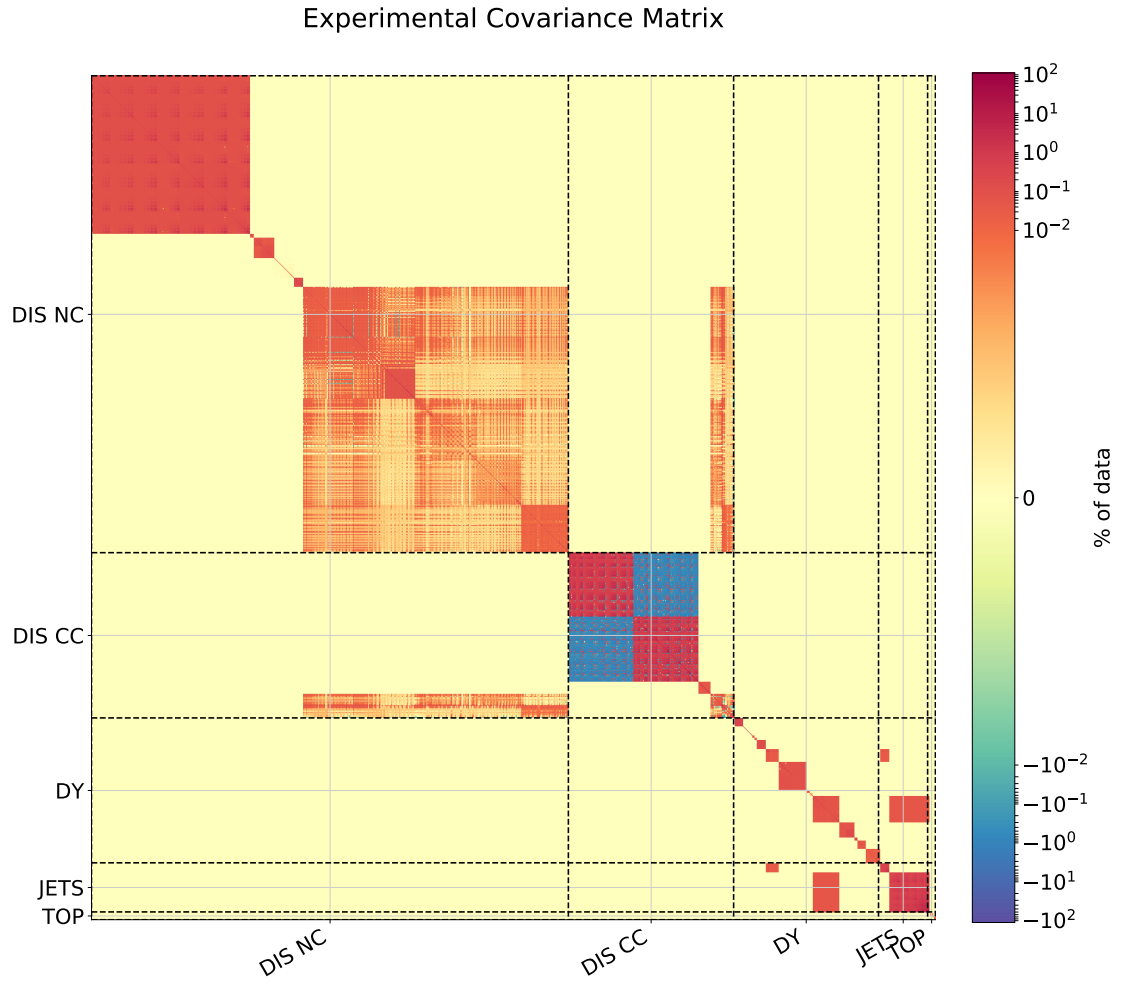
Figure 1.2.2 *An example of an experimental covariance matrix for data included in an NNPDF fit. The data are grouped according to what type of process the interaction belongs to (DIS charged current (CC) and neutral current (NC), Drell-Yan (DY), jets and top production).*

baseline fit, and the covariance matrix is iterated concurrently with preprocessing.

1.2.3 NNPDF fitting strategy

There are a number of groups currently active in carrying out proton PDF fits including MSHT [46], CTEQ [47], NNPDF [34], HERAPDF/xFitter [48] and ABM [49]. 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, which is summarised in Fig. 1.2.3. There are two main features which differ from other fitting collaborations' [50]. 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 [9, 51, 52].
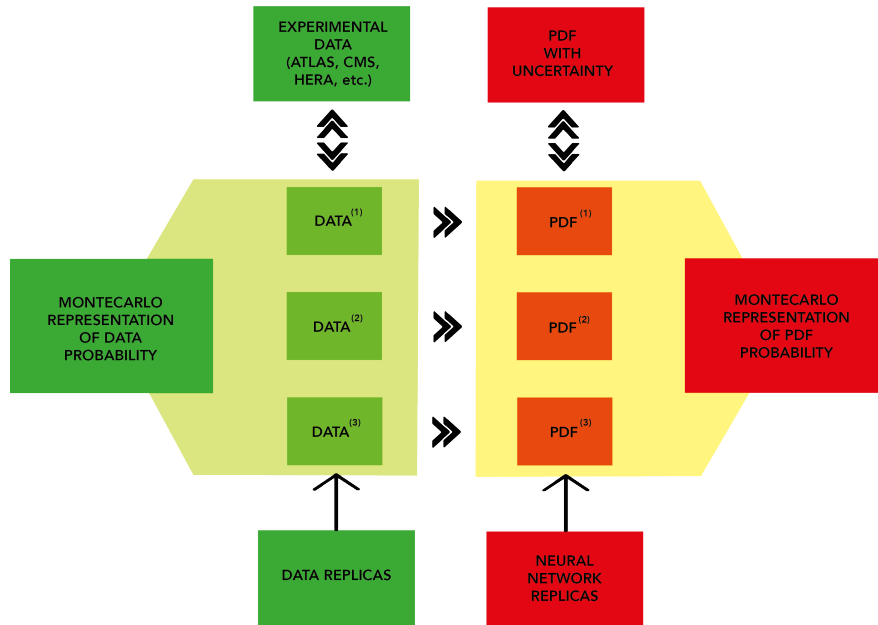
Figure 1.2.3 *NNPDF general strategy.*

1.2.4 Monte Carlo approach

The uncertainties in the functional form of PDFs come as a direct consequence of the uncertainties in the experimental and theoretical inputs. In order to
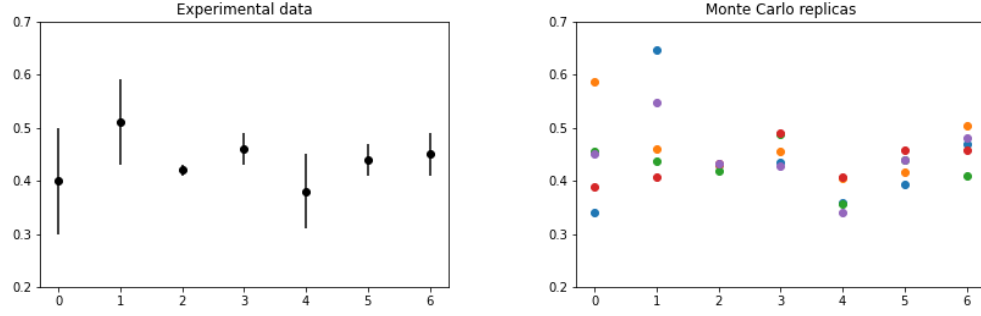
Figure 1.2.4 *Generation of Monte Carlo replicas of pseudodata from data with uncertainties. Left: experimental data; right: five Monte Carlo replicas.*

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. 1.2.4 is a schematic illustrating the generation of these "pseudodata", $D^{(k)}$, $k = 1, ..., N_{rep}$. They are generated using Gaussian random numbers $n_a^{(k)}$ and $\hat{n}_m^{(k)}$:

$$D^{(k)} = (D^0 + \sum_a n_a^{(k)} \sigma^a) \prod_m (1 + \hat{n}_m^{(k)} \sigma^p), \qquad (1.2.3)$$

where D^0 is the (symmetrised) experimental data value, and σ^a and σ^m are the additive and multiplicative uncertainties discussed in Sec. 1.2.2. Explicitly, the pseudodata replicas satisfy the relations:

$$\langle D_i^{(k)} \rangle = D_i^0; \qquad (\langle D_i^{(k)} \rangle - D_i^0)(\langle D_j^{(k)} \rangle - D_j^0) = C_{ij}, \qquad (1.2.4)$$

in the limit of $N_{rep} \to \infty$, where the notation $\langle \cdot \rangle$ denotes the mean over replicas. Fig. 1.2.5 shows the distribution of pseudodata for a single data point.

Once the pseudodata have been generated, each replica, $D^{(k)}$, is fitted separately to the theoretical predictions, $T_i[f_q^{(k)}]$, which depend on the PDF replicas, $f_q^{(k)}$ (where q runs over the fitted flavours: g, u, d, s, c, $\bar{u}$, $\bar{d}$, $\bar{s}$, $\bar{c}$). This is done by fitting the PDFs to minimise a target error function based on the χ^2:

$$\chi^{2\,(k)} = \frac{1}{N_{dat}} \sum_q \sum_{i,j=1}^{N_{dat}} (D_i^{(k)} - T_i[f_q^{(k)}]) C(t_0)_{ij}^{-1} (D_j^{(k)} - T_j[f_q^{(k)}]). \qquad (1.2.5)$$

20
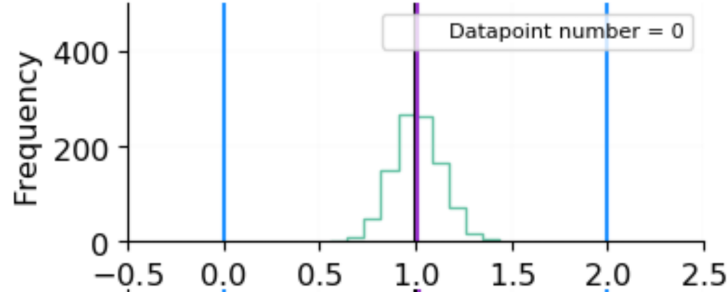
Figure 1.2.5 *Histogram of the distribution of 100 pseudodata replicas for a single data point, normalised to D^0. The purple line is the mean value $\langle D^{(k)} \rangle$, which is equal to D^0 (black line) to arbitrary precision.*

Here $C(t_0)$ is the t0 covariance matrix, which is Eqn. 1.2.1 with $D_i \rightarrow t_{0\ i}$, where t_0 are theory predictions from a similar previous fit. This is to remove d'Agostini bias [44, 45]. Doing this fit then results in a PDF set of each flavour. These act as a Monte Carlo parametrisation of the PDFs (for example, Fig. 1.2.6). 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 [53, 54].
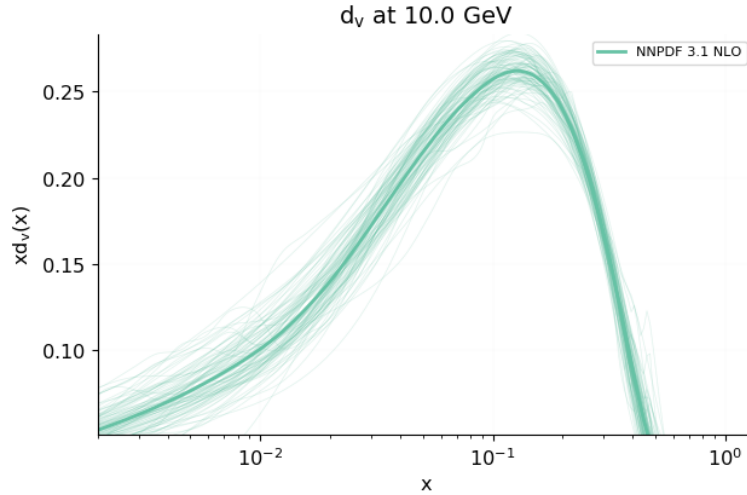
Figure 1.2.6 *Monte Carlo replicas for the down valence quark PDF NNPDF3.1 at NLO.*

1.2.5 Neural networks

Inspired by how the brain processes information, in machine learning neural networks are graphs of connected nodes. They are trained by example, so have

the capability to learn a PDF's functional form given a set of data. Using 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.

Neural networks in NNPDF3.1

Fig. 1.2.7 depicts the architecture used in NNPDF3.1. This is a "2-5-3-1" archiecture, where the numbers refer to the number of nodes in each layer. It
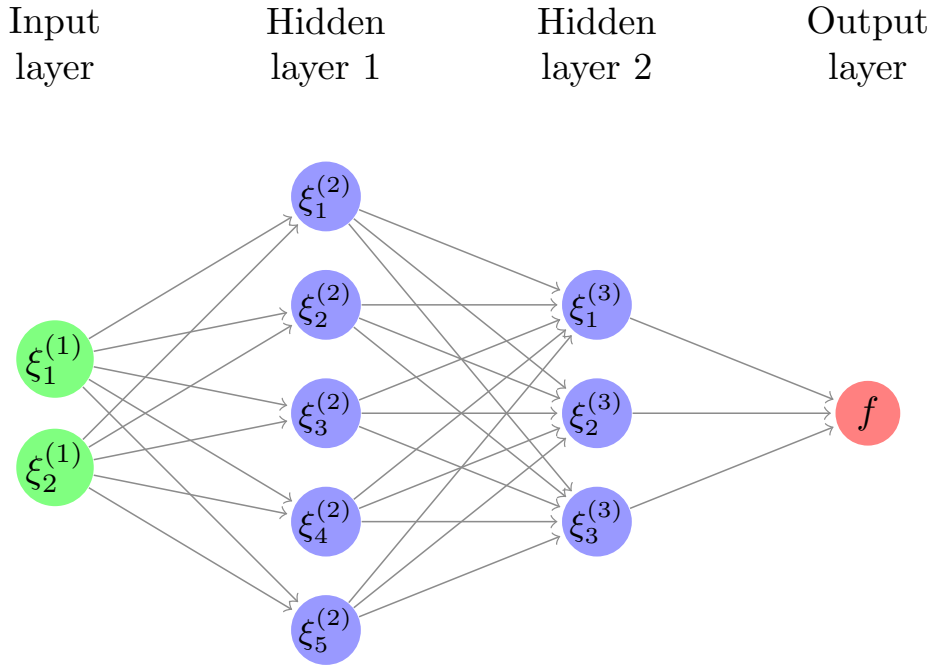
Figure 1.2.7 *Schematic depiction of the 2-5-3-1 architecture of an artificial neural network of NNPDF3.1. Here $\xi_1^{(1)}$ and $\xi_2^{(1)}$ are the variables x and $\ln(1/x)$ respectively.*

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 lth layer is given by

$$\xi_i^{(l)} = g\left(\sum_j^{inputs} \omega_{ij}^{(l)} \xi_j^{(l-1)} + \theta_i^{(l)} \right) \tag{1.2.6}$$

where the ωs and θs are "weights" and "thresholds"; parameters to be minimised with respect to. g is an "activation function" which is set to

$$g(z) = \begin{cases} \frac{1}{1+\exp(-z)} & \text{for hidden layers} \\ z & \text{for the output layer.} \end{cases} \tag{1.2.7}$$

The choice of this 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].

The training of the neural networks is implemented using a "genetic algorithm" [55] (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.

1.2.6 Parametrisation, preprocessing and postprocessing

A scale of $Q = 1.65$ 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. 1.1.28). The PDFs are fitted parametrised in a "fitting basis" of eight combinations of flavours, to help convergence [56], 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:

$$f_i(x) = A_i x^{-\alpha_i} (1 - x)^{\beta_i} N_i(x). \tag{1.2.8}$$

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.

1.2.7   Cross validation

Neural networks are effective at learning the functional form which underlies data.
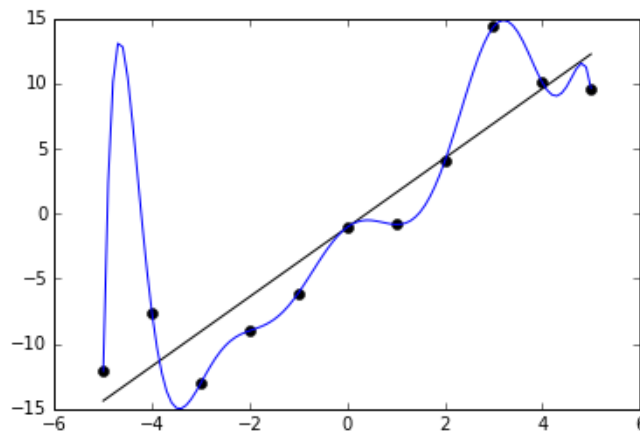
Figure 1.2.8   *Overlearning: the data points (black dots) fluctuate around the linear underlying law (black line), but the neural network continues to minimise the error function until it passes through every data point (blue curve), fitting the noise in the data.*

Sometimes, if there are more degrees of freedom in the PDF than in the data, they can be "too effective", picking up not just the underlying law but also the noise. This is known as "overlearning" (see Fig. 1.2.8 for an example).
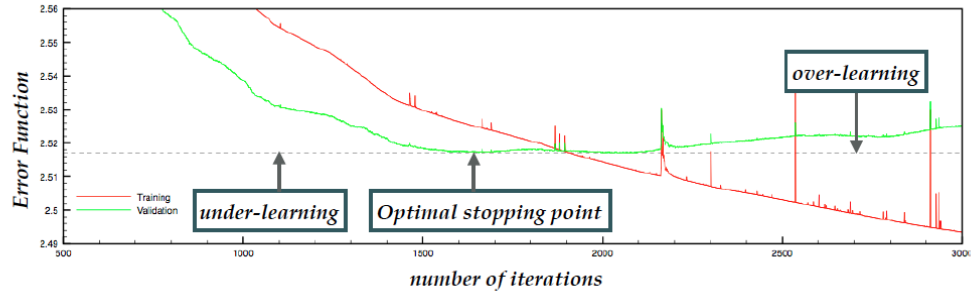
Figure 1.2.9   *Cross validation with the lookback method.*
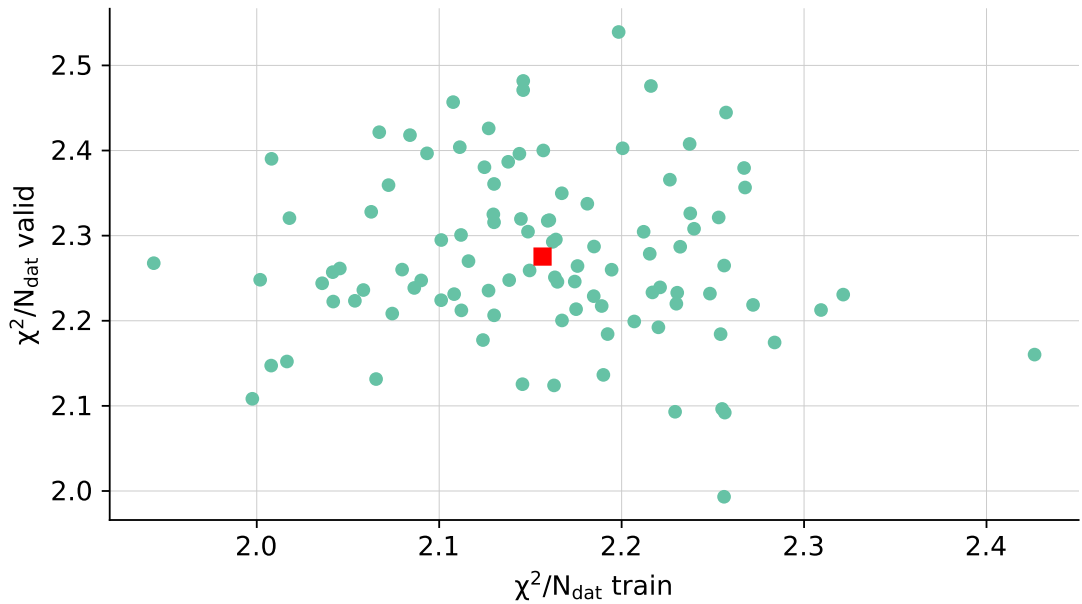
Figure 1.2.10   *Comparing the training and validation χ^2s for the 100 replicas (green circles) of a PDF fit. The red square gives the average.*

To circumvent this problem, the data are split into a training and a validation set. The training data are used to optimise the neural network, and the validation data are 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. In

NNPDF3.1 we determine the optimum fit using the "lookback" method, where after training the model corresponding to the minimum in the validation error function is selected.

1.3 NNPDF4.0

The earlier parts of this section describe the methodology for the NNPDF3.1 [9] release, on which Chapters 3 and 5 are based. However, at the time of writing this thesis a new release, NNPDF4.0, is being launched. The work in Chapter 4 is based on this methodology, and so we briefly explain the main developments between NNPDF3.1 and NNPDF4.0. For more information see [37, 57], and the future NNPDF4.0 paper.

1.3.1 Methodology

NNPDF4.0 heralds some significant methodological changes to the NNPDF procedure, the most important of which we will outline here. Perhaps most strikingly, the architecture of the neural network is changed [58] from that in Fig. 1.2.7, which is implemented separately for each flavour, to a single neural network with flexible architecture. The implementation is via `Keras` [59] and `Tensorflow` [60], and the optimisation is with an inbuilt gradient descent algorithm rather than the genetic algorithm in NNPDF3.1. The stopping criterion follows a patience method where the fit stops once the minimiser is stable for a set length of time. All of these parameters, including architecture are determined via hyperoptimisation with k-folding, which is a process to determine the best combination for stability and performance.

Additionally, positivity of PDF replicas is now strictly enforced rather than them being allowed to be negative within a threshold. This is following a recent study [61] which showed that $\overline{MS}$ PDFs are strictly positive.

1.3.2 Theory developments

There are two main theory developments between NNPDF3.1 and NNPDF4.0. One is the treatment of nuclear and deuteron data using an additional uncertainty.

This forms the basis of Chapter 4 so we will not discuss it further here.

The other is the inclusion of NLO electroweak (EW) corrections. This consists of QED corrections to the DGLAP evolution as well as NLO EW corrections for a variety of processes; wherever EW corrections are available they are added to increase precision. This is very important as EW corrections can be up to $\sim 20\%$ [62] in some regions. The corrections are provided as interpolation grids via the `PineAPPL` library [63].

1.3.3 Validation of PDFs

The effectiveness of the fitting methodology has traditionally been tested using closure tests [56], which use a separate PDF to create proxy known "true values". This procedure has been updated, whereby fits are carried out to many proxy PDF replicas, and the bias and variance of the results are compared. This is made possible by the significantly increased speed of the new fitting methodology [58].

1.3.4 New data

NNPDF4.0 also includes a large number of additional datasets, including many from the 13 TeV run at the LHC (see Fig. 1.2.1). What's more, many of the existing datasets' implementation has been improved [37].

Chapter 2

Theory uncertainties in PDFs

The concept of experimental uncertainties is one that is familiar to any scientist. Whenever we make a measurement it is accompanied by an uncertainty which quantifies our degree of confidence in its accuracy. The smaller this uncertainty is, the more useful the measurement is. Not providing an uncertainty arguably renders the measurement useless, as the implicit uncertainty could in principle be arbitrarily large.

But uncertainties also apply to theoretical predictions. In a broad sense, there is some uncertainty associated with our degree of confidence in a particular theoretical model, but even within the parameters of a model a prediction can often be uncertain. In the case of QCD there are many contributions to this uncertainty. One of the most obvious is the truncation of the perturbation series to fixed order, necessary because successively complex calculations are required for increasing orders. Other contributions include:

- non-perturbative effects such as higher-twist terms;

- treatment of heavy quarks and the impact of nuclear environments [64, 65];

- choice of model parameters such as α_s and particle masses, which have to be determined from experiment, and are subject to different theoretical definitions [66, 67].

PDFs are produced in fits which compare experimental measurements with theoretical predictions, so uncertainties in both these places should end up in the

PDFs. In the past we have got away with arguing that theory uncertainties are small compared to experimental uncertainties [65], and so the experimental ones with dominate the PDF uncertainty. In regions where this is not the case, such as the higher-twist region, the data are simply not included in the fit. But such an argument is increasingly becoming a stretch, due to ultra-precision measurements, the advent of the high-luminosity LHC [68], and anticipated future colliders [6–8].

To calculate theoretical uncertainties we must first consider what it is we are trying to calculate. Underlying the uncertainty is a "true" value, which we are unable to determine exactly. This suggests a Bayesian approach is most appropriate [69, 70], where the uncertainty quantifies our belief that the true value lies within a certain range. We will use such a Bayesian framework, and assume Gaussianity of the expected true value of the theory calculation in order to simplify the situation. This should be sufficient to capture the main features of the uncertainty.

In this section we will show that including general theory uncertainties in PDFs can be done by constructing a theory covariance matrix, S, to complement the experimental covariance matrix, C. Theory uncertainties can then be included in a fit by the replacement $C \to C + S$ [71]. The plus sign appears because the experimental and theoretical uncertainties are independent, so the uncertainties are combined in quadrature. They are also on an equal footing in terms of their effect on the PDFs. When there are many data sets, for example in a global fit, there can be very strong correlations in theory uncertainty, even outwith individual experiments. This is because the underlying theory connects different predictions, even when the corresponding data come from different experiments.

2.1 Fitting PDFs including theory uncertainties

Historically, experimental uncertainties have been the dominant source of uncertainty in PDF fits. In the NNPDF3.1 framework, both replica generation and computation of χ^2 are based entirely on these. We must now try to match the ongoing drive to increase experimental precision by including uncertainties introduced at the theoretical level. This is especially important given data sets new in NNPDF3.1 such as the Z boson transverse momentum distributions [72–74], which have very high experimental precision. Without the inclusion of theoretical uncertainties, this has led to tension with the other datasets.

In future NNPDF fits theoretical uncertainties will be included following a procedure outlined in Sec. 4.1 of [71]. This hinges on a result from Bayesian statistics which applies to Gaussian uncertainties. Namely, theory uncertainties can be included by directly adding a theoretical covariance matrix to the experimental covariance matrix prior to the fitting. We will now give a brief summary of this derivation.

When determining PDFs we incorporate information from experiments in the form of N_{dat} experimental data points D_i, $i = 1, ..., N_{dat}$. 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:

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

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. 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. In Chapter 5 we will take an in-depth look into these correlations.

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.

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

where S_{ij} is a "theory covariance matrix" encapsulating the magnitude and correlation of the various theory uncertainties. We will need to do some work to determine S_{ij} for the different sources of uncertainty, 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 it to $P(D|T)$ using Bayes' Theorem:

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

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 $\mathcal{T}$:

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

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

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

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

$$\Delta_i \equiv \mathcal{T}_i - T_i. \tag{2.1.6}$$

These shifts quantify the accuracy of the theoretical predictions, and can be thought of as nuisance parameters in the PDF fit. We can express Eqn. 2.1.5 in terms of the shifts, Δ_i, making use of the assumptions of Gaussianity in Eqns. 2.1.1 and 2.1.2:

$$
\begin{aligned}
P(D|T) \propto \int D^N\Delta \ \exp\bigg(& -\frac{1}{2}(D_i - T_i - \Delta_i) \\
& \times C_{ij}^{-1}(D_j - T_j - \Delta_j) - \frac{1}{2}\Delta_i S_{ij}^{-1}\Delta_j \bigg).
\end{aligned}
\tag{2.1.7}
$$

To evelute 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:

$$(D - T - \Delta)^T C^{-1}(D - T - \Delta) + \Delta^T S^{-1}\Delta$$
$$= D^T(C^{-1} + S^{-1})\Delta - \Delta^T C^{-1}(D - T) - (D - T)^T C^{-1}\Delta + (D - T)^T C^{-1}(D - T)$$
$$= (\Delta - (C^{-1} + S^{-1})^{-1}C^{-1}(D - T))^T (C^{-1} + S^{-1})$$
$$\times (\Delta - (C^{-1} + S^{-1})^{-1}C^{-1}(D - T))$$
$$- (D - T)^T C^{-1}(C^{-1} + S^{-1})^{-1}C^{-1}(D - T) + (D - T)^T C^{-1}(D - T).$$
$$(2.1.8)$$

Now, integrating Eqn. 2.1.7 over Δ leads to a constant from the Gaussian integrals, which we can absorb, and only the parts of the exponent without Δ remain:

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

We can further simplify this by noting that

$$(C^{-1} + S^{-1})^{-1} = (C^{-1}(C + S)S^{-1})^{-1} \qquad (2.1.10)$$
$$= S(C + S)^{-1}C,$$

which means we can rewrite

$$C^{-1} - C^{-1}(C^{-1} + S^{-1})^{-1}C^{-1} = C^{-1} - C^{-1}S(C + S)^{-1} \qquad (2.1.11)$$
$$= (C^{-1}(C + S) - C^{-1}S)(C + S)^{-1}$$
$$= (C + S)^{-1}.$$

Finally, with indices restored we are left with

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

Comparing this result to Eqn. 2.1.1, we can confirm that when we possess theoretical predictions, T_i, rather than true values, $\mathcal{T}_i$, we can account for this by adding a theoretical covariance matrix, S_{ij} to the experimental covariance matrix, C_{ij} [71]. This means the theory uncertainties are on an equal footing with experimental systematic uncertainties. Note that C_{ij} is positive definite

by construction and so $(C + S)_{ij}$ is always invertible, even if S_{ij} has negative eigenvalues.

Now all that remains is to construct a theory covariance matrix which parametrises each instance of theoretical uncertainty. This is a nebulous task, given that we are not privy to the true values, $\mathcal{T}$, and so are unable to simply apply the formal definition

$$S_{ij} = \langle (\mathcal{T}_i - T_i)(\mathcal{T}_j - T_j) \rangle, \tag{2.1.13}$$

where $\langle \cdot \rangle$ denotes an average over true values, $\mathcal{T}$. We need to find methods to calculate the various contributions, S_{ij}, (be they MHOUs, nuclear corrections, higher twist corrections etc.) which not only encapsulate the per-point theoretical uncertainties but also preserve the correlations between different data points. Unlike experimental uncertainties, these correlations can exist outwith individual experiments; in fact, all data in PDF fits depend themselves on PDFs, and this common link will lead to correlations between all datapoints, albeit of varying strength.

The rest of this thesis addresses two of the most important types of theoretical uncertainties: MHOUs (Chapter 3) and nuclear uncertainties (Chapter 4). For each type, we show how to construct a theoretical covariance matrix, and present and discuss the results of PDF fits including these covariance matrices. Then in Chapter 5 we consider how to use these PDFs with theory uncertainties to make new physics predictions which also include theory uncertainties.

Chapter 3

Missing higher order uncertainties

In this chapter we address the dominant source of theory uncertainty in current PDF fits: missing higher order uncertainties (MHOUs). In Sec. 3.1 we explain their origin, then in Sec. 3.2 we revise their standard method of estimation, through scale variation. We then show how to use this to construct a theory covariance matrix (Sec. 3.3), and test the validity of this at NLO against the known NNLO result (Sec. 3.5). Finally, we present the PDFs including MHOUs (Sec. 3.6).

3.1 Introduction

PDF fits rely on the comparison of experimental data with theoretical predictions at the partonic level. These predictions are carried out in the framework of perturbation theory, where results are expressed as an expansion in the strong coupling constant, α_s. The first non-zero contribution to the expansion is known as "leading order" (LO), the next is "next-to-leading order" (NLO), and so on (NNLO, N^{3}LO etc.). Because in the perturbative regime α_s is small $(0.11791 \pm 0.00009$ [75]), corrections from higher orders are increasingly small. Predictions must be directly calculated at each order by considering all the possible contributing Feynman diagrams, and this becomes exponentially more complicated with increasing orders; the cutting edge of calculations is currently at the N^{3}LO level. PDFs are fitted using predictions truncated at a given order, with NNLO PDFs being the modern standard.
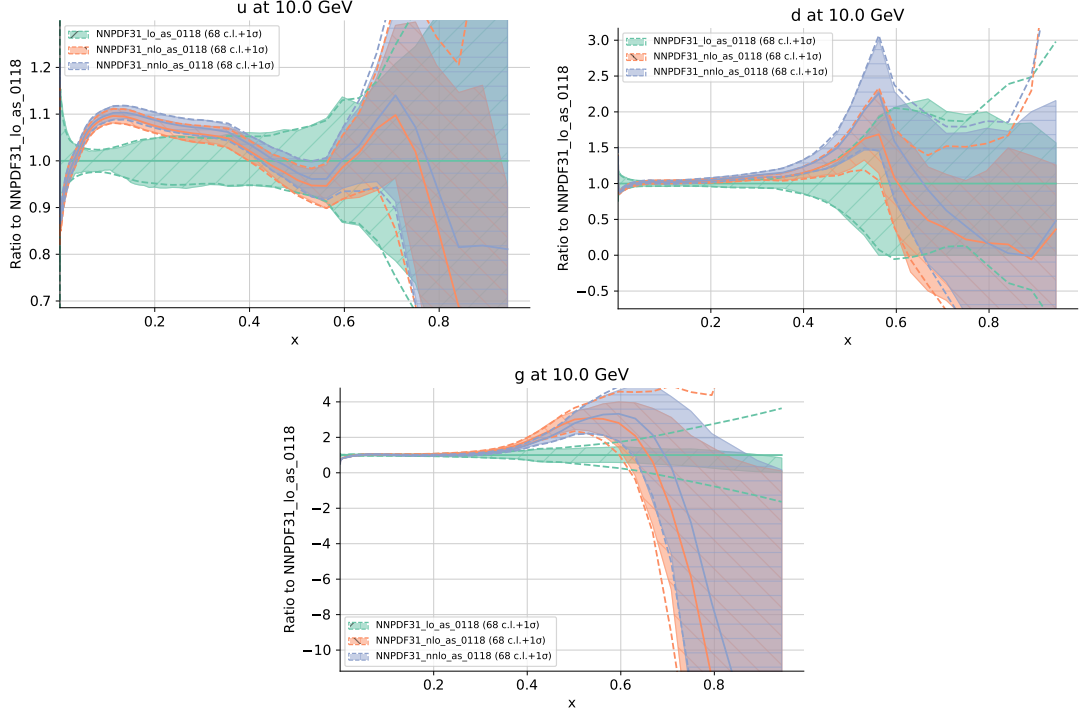
Figure 3.1.1 *Comparison of NNPDF3.1 PDFs at different perturbative orders: LO (green); NLO (orange); NNLO (blue). PDFs are normalised to the LO result, and displayed at scale Q = 10 GeV.*

These missing higher order terms in the expansion for theory predictions lead to MHOUs, which are currently the dominant source of uncertainty in PDF fits. We can see that going from LO to NLO to NNLO in Fig. 3.1.1 that the functional form of the PDF changes, and that the change from LO to NLO is greater than that from NLO to NNLO. MHOUs are currently not included in the PDF uncertainties, justified historically by the claim that they are small compared to experimental contributions to the PDF uncertainty, especially at NNLO. This justification, however, is now on shakier ground with PDF uncertainties dropping as low as 1% at the electroweak scale. QCD MHOU uncertainties themselves are typically $\mathcal{O}(1\%)$ [76] and, with the current push to N^{3}LO precision, will only become increasingly important as time goes on.

In addition to a missing source of per-point uncertainty on each data point, MHOUs can affect a PDF fit more insidiously by impacting on the desired weight of data sets relative to one another; regions of data with high MHOU are naturally to be trusted less when used to determine the PDFs, and so should carry less weight in the fitting process. If MHOUs are included, these data will be deweighted automatically because they will carry higher uncertainty, however in the absence of MHOUs they may impact on a fit to an undesirable degree.

Having established the importance of including MHOUs, in the next section we will go on to develop a formalism for estimating them, and constructing a MHOU covariance matrix.

3.2   Scale variation

The most popular method for estimating MHOUs is by "scale variation". This is based on making theoretical predictions at a range of values of the artificial renormalisation (μ_R) and factorisation (μ_F) scales introduced in Chapter 1. The renormalisation group equation (RGE; 1.1.17) and factorisation theorem only hold to all orders in perturbation theory, and in this case varying the scale values will have no effect on any results. However, when the perturbative expansion is truncated, there will be a residual μ_R and μ_F dependence which characterises the degree of MHOU. Varying these scales and observing the impact on the predictions can therefore provide an estimate of the MHOUs.

Although other approaches to estimating MHOUs, based on the current known orders, have been suggested [77–80], we adopted the method of scale variations not only because it is the most widely used, but also because it is the most easily implemented for our purposes. Firstly, the renormalisation group invariance is incorporated automatically, which ensures the MHOUs decrease as the perturbative order increases. Secondly, the scale dependece of $\alpha_s(\mu_R^2)$ and the PDFs is universal to all processes, which is important for PDF fits dealing with a range of interactions. Finally, correlations between data points are implicitly maintained because predictions for different scale values will be smooth functions of kinematics; this ensures that neighbouring regions of phase space will be strongly correlated.

There are, however, some disadvantages. Firstly, the definition of the two scales themselves has been historically approached in various ways, often differently for DIS and hadronic collisions, but also changing over time. Since PDFs use both DIS and hadronic data we need to settle on a consistent approach. Secondly, there is no cut and dry method for determining the range of varied scale choices, and in fact the choice of central scales are themselves to some degree arbitrary; for example, top production processes commonly have both central scales set to the top mass, m_t, and DIS processes have both set to Björken Q. Though there is physical motivation for these choices, we could equally well pick $2m_t$ rather

than m_t in the former case, for example. A standard approach is to take the 7-point envelope of the predictions obtained by varying (μ_F, μ_R) independently in $\{1/2, 1, 2\}$, excluding $(1/2, 1/2)$ and $(2, 2)$. However, for our purposes we do not want a per-point envelope but rather a covariance matrix which retains correlations between data points. We will address both of these drawbacks below.

Finally, scale variation techniques will not pick up any "new physics" at higher orders, be it additional colour configurations, singularities or mechanisms of interaction. This is harder to deal with, and requires resummation techniques among other methods. In this work we assume these effects to be less important, and do not address them for the time being.

In the remainder of this section we will review the technique of scale variation, and with it the definitions of μ_F and μ_R. We will converge on a general formalism that can be applied to both electroproduction and hadroproduction. We will show that there are two independent directions of scale variation and discuss how to combine them, both in single process and multi-process interactions. We will then go on to show how to use this to build a covariance matrix in Sec. 3.3.

3.2.1 Renormalisation group invariance

It is customary when making a theory prediction to pick a renormalisation scale, μ_R, that is indicative of the physical scale of the interaction, Q. We will denote this "central" theory prediction by $T(Q^2)$. In general, a theory prediction at scale μ_R can be written $\overline{T}(\alpha_s(\mu_R^2), \mu_R^2/Q^2)$, where we explicitly note that α_s itself depends on the renormalisation scale. From this we can see that

$$T(Q^2) \equiv \overline{T}(\alpha_s(Q^2), 1). \tag{3.2.1}$$

The strong coupling constant satisfies the RGE

$$\mu_R^2 \frac{d^2}{d\mu_R^2} \alpha_s(\mu_R^2) = \beta(\alpha_s(\mu_R^2)), \tag{3.2.2}$$

and we can expand the beta function perturbatively as

$$\beta(\alpha_s) = \beta_0 \alpha_s^2 + \beta_1 \alpha_s^3 + \beta_2 \alpha_s^4 + \dots . \tag{3.2.3}$$

As discussed in Chapter 1, renormalisation group invariance tells us that a

prediction of a physical quantity (such as $\overline{T}$) to all orders must be independent of μ_R, because this scale is unphysical. This means we can write

$$\mu_R^2 \frac{d}{d\mu_R^2} \overline{T}\left(\alpha_s(\mu_R^2), \mu_R^2/Q^2\right) = 0. \tag{3.2.4}$$

Before proceeding further, we introduce some variables to make the analysis clearer:

$$\mu_R^2 = kQ^2, \qquad t = \ln(Q^2/\Lambda^2), \qquad \kappa = \ln k = \ln \mu_R^2/Q^2, \tag{3.2.5}$$

where Λ is the QCD scale. This means $\alpha_s(\mu_R^2)$ is a function of $\ln \mu_R^2/\Lambda^2 = t + \kappa$.

Revisiting Eqn. 3.2.4, we can write this as

$$
\begin{aligned}
0 &= \frac{d}{d\kappa} \overline{T}(\alpha_s(t+\kappa), \kappa) \\
&= \frac{d}{d\kappa} \alpha_s(t+\kappa) \frac{\partial}{\partial \alpha_s} \overline{T}(\alpha_s(t+\kappa), \kappa)\bigg|_\kappa + \frac{\partial}{\partial \kappa} \overline{T}(\alpha_s(t+\kappa), \kappa)\bigg|_{\alpha_s},
\end{aligned}
\tag{3.2.6}
$$

assuming that $\overline{T}$ is analytic in α_s and κ. To simplify this we can use

$$\frac{d}{d\kappa} \alpha_s(t+\kappa) = \frac{d}{dt} \alpha_s(t+\kappa) = \frac{d}{d\ln \mu_R^2} \alpha_s(t+\kappa) = \beta(\alpha_s(t+\kappa)), \tag{3.2.7}$$

where we have used the definition of the beta function (Eqn. 3.2.2), and this means that

$$0 = \frac{\partial}{\partial t} \overline{T}(\alpha_s(t+\kappa), \kappa)\bigg|_\kappa + \frac{\partial}{\partial \kappa} \overline{T}(\alpha_s(t+\kappa), \kappa)\bigg|_{\alpha_s}. \tag{3.2.8}$$

We can now Taylor expand $\overline{T}(\alpha_s, \kappa)$ about the central scale $\mu_R^2 = Q^2 \implies k = 1 \implies \kappa = 0$ for fixed α_s:

$$
\begin{aligned}
\overline{T}(\alpha_s(t+\kappa), \kappa) &= \overline{T}(\alpha_s(t+\kappa), 0) \\
&\quad + \kappa \frac{\partial}{\partial \kappa} \overline{T}(\alpha_s(t+\kappa), 0)\bigg|_{\alpha_s} + \tfrac{1}{2}\kappa^2 \frac{\partial^2}{\partial \kappa^2} \overline{T}(\alpha_s(t+\kappa), 0)\bigg|_{\alpha_s} + \dots
\end{aligned}
\tag{3.2.9}
$$

.

Then, using Eqn. 3.2.6, we can replace $\frac{\partial}{\partial \kappa}$ with $-\frac{\partial}{\partial t}$, and write

$$
\begin{aligned}
\overline{T}(\alpha_s(t+\kappa), \kappa) &= \overline{T}(\alpha_s(t+\kappa), 0) - \kappa \frac{\partial}{\partial t}\overline{T}(\alpha_s(t+\kappa), 0)\Big|_\kappa \\
&\quad + \tfrac{1}{2}\kappa^2 \frac{\partial^2}{\partial t^2}\overline{T}(\alpha_s(t+\kappa), 0)\Big|_\kappa + \dots \\
&= T(t+\kappa) - \kappa \frac{d}{dt}T(t+\kappa) + \tfrac{1}{2}\kappa^2 \frac{d^2}{dt^2}T(t+\kappa) + \dots \quad (3.2.10)
\end{aligned}
$$

This tells us how to find a scale varied theoretical prediction, $\overline{T}$, in terms of the t dependence of the central prediction, T. Furthermore, we can express this t dependence as an α_s dependence using

$$
\frac{d}{dt}T(t) = \frac{d\alpha_s(t)}{dt}\frac{\partial}{\partial \alpha_s}\overline{T}(\alpha_s(t), 0) = \beta(\alpha_s(t))\frac{\partial}{\partial \alpha_s}\overline{T}(\alpha_s(t), 0). \tag{3.2.11}
$$

Noting that $\beta(\alpha_s) = \mathcal{O}(\alpha_s^2)$, we see that $\frac{1}{T}\frac{dT}{dt} = \mathcal{O}(\alpha_s)$ and $\frac{1}{T}\frac{d^2 T}{dt^2} = \mathcal{O}(\alpha_s^2)$ etc. The pattern follows that every time a derivative is taken with respect to t you pick up a power of α_s as a consequence of the chain rule in differentiating. Looking back at Eqn. 3.2.10 it is clear that each power of κ is associated with a power of α_s. Expressing the theory prediction perturbatively as

$$
T = \alpha_s T_{\mathrm{LO}} + \alpha_s^2 T_{\mathrm{NLO}} + \alpha_s^3 T_{\mathrm{NNLO}} + \dots , \tag{3.2.12}
$$

we can match powers of α_s in Eqn. 3.2.10 to obtain the expressions

$$
\begin{aligned}
\overline{T}_{\mathrm{LO}}(\alpha_s(t+\kappa), \kappa) &= T_{\mathrm{LO}}(t+\kappa), \\
\overline{T}_{\mathrm{NLO}}(\alpha_s(t+\kappa), \kappa) &= T_{\mathrm{NLO}}(t+\kappa) - \kappa \tfrac{d}{dt}T_{\mathrm{LO}}(t+\kappa), \\
\overline{T}_{\mathrm{NNLO}}(\alpha_s(t+\kappa), \kappa) &= T_{\mathrm{NNLO}}(t+\kappa) - \kappa \tfrac{d}{dt}T_{\mathrm{NLO}}(t+\kappa) \\
&\quad + \tfrac{1}{2}\kappa^2 \tfrac{d^2}{dt^2}T_{\mathrm{LO}}(t+\kappa).
\end{aligned} \tag{3.2.13}
$$

The difference between the scale varied prediction and the central scale prediction,

$$
\Delta(t, \kappa) = \overline{T}(\alpha_s(t+\kappa), \kappa) - T(t). \tag{3.2.14}
$$

can be used to estimate the MHOU. From Eqn. 3.2.13 we find the explicit

expressions for the theory uncertainties

$$\Delta_{\text{LO}}(t, \kappa) = T_{\text{LO}}(t + \kappa) - T_{\text{LO}}(t),$$
$$\Delta_{\text{NLO}}(t, \kappa) = (T_{\text{NLO}}(t + \kappa) - \kappa \tfrac{d}{dt} T_{\text{LO}}(t + \kappa)) - T_{\text{NLO}}(t),$$
$$\Delta_{\text{NNLO}}(t, \kappa) = (T_{\text{NNLO}}(t + \kappa) - \kappa \tfrac{d}{dt} T_{\text{NLO}}(t + \kappa)$$
$$+ \tfrac{1}{2} \kappa^2 \tfrac{d^2}{dt^2} T_{\text{LO}}(t + \kappa)) - T_{\text{NNLO}}(t) .$$

$$(3.2.15)$$

At LO we can see that the uncertainty results entirely from the choice of κ, in other words of μ_R in the α_s evaluation. At NLO we can see that the leading part of $T_{\text{NLO}}(t + \kappa)$ is subtracted off by the $\mathcal{O}(\kappa)$ term, meaning that the uncertainty is reduced with respect to LO. At NNLO, in addition, the $\mathcal{O}(\kappa^2)$ term subtracts off the subleading dependence of $T_{\text{NNLO}}(t + \kappa) - \kappa \tfrac{d}{dt} T_{\text{NLO}}(t + \kappa)$, and so the uncertainty is yet smaller. This pattern of decreased scale variation uncertainties with increased perturbative order reflects our general understanding of the behaviour of MHOUs.

It is also apparent that the size of MHOU depends on the value of κ, in other words on the size of scale variation. This introduces a degree of arbitrariness into MHOU estimation, with the historical empirical range of choice being $\kappa \in [-\ln 4, \ln 4]$. In practice, we must investigate the dependence of Δ on κ, using validation at lower orders against known higher orders to converge on a suitable prescription. This will be addressed in Sec. 3.3.

We will now go on to show how RG invariance can be applied to processes involving hadrons, where the partonic cross section is also convolved with a PDF. We will show that in this scenario there are two independent scales, and thus two independent sources of MHOU: one from the α_s dependence in the hard cross section; the other from the anomalous dimensions in the PDF evolution.

3.2.2   Scale variation in partonic cross sections

We will start with DIS, where there is only one hadron then move to the case of hadron-hadron collisions, such as those carried out at the LHC. In each case we will consider RG invariance to find an expression for μ_R variation in the partonic observable, for the case where the PDF is evaluated at the physical scale. Scale variation in PDF evolution, i.e. the μ_F variation, will be addressed in the next section.

Deep Inelastic Scattering

For DIS processes, theory predictions are of the structure functions discussed in Chapter 1. These can be expressed as a convolution of a parton level coefficient, C, with a PDF, f:

$$F(Q^2) = C(x, \alpha_s(Q^2)) \otimes f(x, Q^2), \qquad (3.2.16)$$

where $\otimes$ is a convolution in the momentum fraction, x, and there is an implicit sum over parton flavours. There will be a MHOU in F due to truncating the coefficient function, C, to fixed perturbative order. We can estimate this by keeping the PDF scale (or factorisation scale) fixed and varying the renormalisation scale in C. This will result in a scale-varied structure function,

$$\overline{F}(Q^2, \mu_R^2) = \overline{C}(\alpha_s(\mu_R^2), \mu_R^2/Q^2) \otimes f(Q^2), \qquad (3.2.17)$$

where we have made the x-dependence implicit and the scale dependence explicit in the coefficient function. We can use the quantities defined in Eqn. 3.2.5 to write this as

$$\overline{F}(t, \kappa) = \overline{C}(\alpha_s(t + \kappa), \kappa) \otimes f(t). \qquad (3.2.18)$$

We know that the structure function, an observable, is RG invariant, and, because we are keeping the factorisation scheme fixed, the PDF is independent of μ_R. This means that the coefficient functions must also obey RG invariance, and so in a parallel with Eqn. 3.2.10 we can write

$$\overline{C}(\alpha_s(t + \kappa), \kappa) = C(t + \kappa) - \kappa \tfrac{d}{dt} C(t + \kappa) + \tfrac{1}{2}\kappa^2 \tfrac{d^2}{dt^2} C(t + \kappa) + \ldots, \qquad (3.2.19)$$

where, like before, we denote the central scale quantities without a bar. In order to evaluate the derivatives, note that the coefficient function can be expressed as a perturbative expansion in α_s,

$$C(t) = c_0 + \alpha_s(t)c_1 + \alpha_s^2(t)c_2 + \alpha_s^3(t)c_3 + \ldots, \qquad (3.2.20)$$

and that $\tfrac{d}{dt}\alpha_s(t, \kappa) = \beta(\alpha_s(t, \kappa))$, where the beta function also admits the expansion in Eqn. 3.2.2. Explicitly, this leads us to

$$\begin{aligned}
\tfrac{d}{dt} C(t) &= \alpha_s^2(t)\beta_0 c_1 + \alpha_s^3(t)(\beta_1 c_1 + 2\beta_0 c_2) + \ldots, \\
\tfrac{d^2}{dt^2} C(t) &= 2\alpha_s^3(t)\beta_0^2 c_1 + \ldots,
\end{aligned} \qquad (3.2.21)$$

resulting in the perturbative expression for μ_R variation of C:

$$\overline{C}(\alpha_s(t+\kappa), \kappa) = c_0 + \alpha_s(t+\kappa)c_1 + \alpha_s^2(t+\kappa)(c_2 - \kappa\beta_0 c_1)$$
$$+ \alpha_s^3(t+\kappa)(c_3 - \kappa(\beta_1 c_1 + 2\beta_0 c_2) + \kappa^2\beta_0^2 c_1) + \dots .$$
(3.2.22)

Using Eqn. 3.2.16 we finally get an expression for the μ_R variation of F:

$$\overline{F}(t, \kappa) = c_0 \otimes f(t) + \alpha_s(t+\kappa)c_1 \otimes f(t) + \alpha_s^2(t+\kappa)\left(c_2 - \kappa\beta_0 c_1\right) \otimes f(t)$$
$$+ \alpha_s^3(t+\kappa)\left(c_3 - \kappa(\beta_1 c_1 + 2\beta_0 c_2) + \kappa^2\beta_0^2 c_1\right) \otimes f(t) + \dots .$$
(3.2.23)

In practice, when predicting scale varied observables, using these equations is relatively straightforward. Because the coefficients, c_i and β_i, are already known to some order, the workflow consists of some basic algebra to create the new, scale varied, coefficients at each order from the central-scale coefficients at the surrounding orders.

Hadron-hadron collisions

Hadron-hadron collisions can be considered in a similar way to DIS, the difference being that the observable cross section, Σ, depends on two PDFs, one for each of the hadrons:

$$\Sigma(t) = H(t) \otimes (f(t) \otimes f(t)),$$
(3.2.24)

where H is the parton level cross section and this time we have used $t = \ln(Q^2/\Lambda^2)$ from the outset. Once again, there is an implicit sum over parton flavours. As before, we can vary $\kappa = \ln(\mu^2/Q^2)$ in H whilst keeping f fixed, so that

$$\overline{\Sigma}(t, \kappa) = \overline{H}(\alpha_s(t+\kappa), \kappa) \otimes (f(t) \otimes f(t)),$$
(3.2.25)

where

$$\overline{H}(\alpha_s(t), \kappa) = H(t) - \kappa\frac{d}{dt}H(t) + \frac{1}{2}\kappa^2\frac{d^2}{dt^2}H(t) + \dots .$$
(3.2.26)

Because hadron-hadron collisions involve a range of processes, we consider a generic process starting at $O(\alpha_s^n)$ for $n \in \mathbb{Z}$, so that

$$H(t) = \alpha_s^n(t)h_0 + \alpha_s^{n+1}(t)h_1 + \alpha_s^{n+2}(t)h_2 + \dots .$$
(3.2.27)

Once again using $\frac{d}{dt}\alpha_s(t,\kappa) = \beta(\alpha_s(t,\kappa))$ and Eqn. 3.2.2 we arrive at

$$
\begin{aligned}
\frac{d}{dt}H(t) &= n\alpha_s^{n-1}(t)\beta(\alpha_s)h_0 + (n+1)\alpha_s^n(t)\beta(\alpha_s)h_1 + \dots \\
&= \alpha_s^{n+1}n\beta_0 h_0 + \alpha_s^{n+2}(n\beta_1 h_0 + (n+1)\beta_0 h_1) + \dots \\
\frac{d^2}{dt^2}H(t) &= \alpha_s^{n+2}n(n+1)\beta_0^2 h_0 + \dots.
\end{aligned}
\tag{3.2.28}
$$

Overall, to evaluate $\overline{\Sigma}$ we can therefore use Eqn. 3.2.25 along with

$$
\begin{aligned}
\overline{H}(\alpha_s,\kappa) &= \alpha_s^n h_0 + \alpha_s^{n+1}(h_1 - \kappa n\beta_0 h_0) \\
&+ \alpha_s^{n+2}(h_2 - \kappa(n\beta_1 h_0 + (n+1)\beta_0 h_1) \\
&+ \tfrac{1}{2}\kappa^2 n(n+1)\beta_0^2 h_1) + \dots.
\end{aligned}
\tag{3.2.29}
$$

Again, to evaluate the scale varied cross section, all that is needed is to modify the coefficients at each order using those at the central scale for the surrounding orders.

3.2.3   Scale variation in evolution of PDFs

We now turn to the effects of scale variation in the PDFs themselves. This is a crucial contribution to MHOUs because the PDFs are common to predictions for all processes, and therefore responsible for widespread correlations in uncertainty. MHOUs in the PDFs arise from the truncation of the perturbative expansion of the splitting functions (or, in Mellin space, the anomalous dimensions) in the DGLAP evolution equations (Eqn. 1.1.28) discussed in Chapter 1. The scale evolution of the PDFs can be encapsulated in Mellin space in the equation

$$
\mu_F^2 \frac{d}{d\mu_F^2} f(\mu_F^2) = \gamma(\alpha_s(\mu_F^2))f(\mu_F^2),
\tag{3.2.30}
$$

where the parton flavour indices are left implicit and the anomalous dimension, γ, can be expressed as an expansion in α_s as

$$
\gamma(t) = \alpha_s(t)\gamma_0 + \alpha_s^2(t)\gamma_1^2 + \alpha_s^3(t)\gamma_2^3 + \cdots.
\tag{3.2.31}
$$

Note that we refer to a separate factorisation scale, μ_F, distinct from the renormalisation scale, μ_R, in the previous section. This is because each scale is associated with a separate RGE and they are therefore independent; to explore the full space of scale variations they need to be considered separately.

To solve for the PDF, we can integrate Eqn. 3.2.30 to give

$$f(\mu_F^2) = \exp\left(\int_{\mu_0}^{\mu_F^2} \frac{d\mu^2}{\mu^2} \gamma(\alpha_s(\mu^2)) \right) f_0 \,, \qquad (3.2.32)$$

where f_0 is the PDF at the initial scale, μ_0. Note that the LHS is independent of the choice of μ_0. To consider the effect of scale variations on the PDF, we can proceed similarly to Sec. 3.2.2, defining

$$\mu_F^2 = kQ^2, \qquad t = \ln(Q^2/\Lambda^2), \qquad \kappa = \ln k = \ln \mu_F^2/Q^2, \qquad (3.2.33)$$

and finding the scale varied anomalous dimension

$$\overline{\gamma}(\alpha_s(t), \kappa) = \gamma(t) - \kappa \tfrac{d}{dt} \gamma(t) + \tfrac{1}{2} \kappa^2 \tfrac{d^2}{dt^2} \gamma(t) + \cdots. \qquad (3.2.34)$$

Once again we can use the beta function expansion (Eqn. 3.2.2) alongside Eqn. 3.2.31 to give

$$\begin{aligned}
\overline{\gamma}(\alpha_s(t+\kappa), \kappa) &= \alpha_s(t+\kappa)\gamma_0 + \alpha_s^2(t+\kappa)(\gamma_1 - \kappa\beta_0\gamma_0) \\
&+ \alpha_s^3(t+\kappa)(\gamma_2 - \kappa(\beta_1\gamma_0 + 2\beta_0\gamma_1) + \kappa^2\beta_0^2\gamma_0) \\
&+ \cdots,
\end{aligned} \qquad (3.2.35)$$

which has the same form as Eqn. 3.2.29 upon setting $n = 1$. We can use this equation to estimate MHOUs in PDFs, which can be done by refitting the PDF at each scale choice using different anomalous dimensions. This method has been applied in previous analyses [81–83], but the process of refitting the PDFs is computationally intensive and so we would like to avoid having to do this if possible. Instead, we can look directly at the PDF level and consider evaluating the PDFs themselves at different scales, as was done in [84].

We can define the scale varied PDF as that obtained by varying the scale in the anomalous dimension,

$$\overline{f}(\alpha_s(t+\kappa), \kappa) = \exp\left(\int_{t_0}^{t} dt' \overline{\gamma}(\alpha_s(t'+\kappa), \kappa) \right) f_0 \qquad (3.2.36)$$

Shifting integration variable $t' \to t' - \kappa$ whilst redefining the initial scale, we can

then apply Eqn. 3.2.34 and expand the exponential, *i.e.*

$$
\begin{aligned}
\overline{f}(\alpha_s(t+\kappa),\kappa) &= \exp\left(\int_{t_0}^{t+\kappa} dt' \overline{\gamma}(\alpha_s(t'),\kappa)\right) f_0 \\
&= \exp\left(\left[\int_{t_0}^{t+\kappa} dt' \gamma(t')\right] - \kappa\gamma(t+\kappa) + \tfrac{1}{2}\kappa^2 \frac{d}{dt}\gamma(t+\kappa) + \ldots\right) \\
&\quad \times \exp\left(\kappa\gamma(t_0) - \tfrac{1}{2}\kappa^2 \frac{d}{dt}\gamma(t_0) + \ldots\right) f_0 \qquad\qquad (3.2.37) \\
&= \left[1 - \kappa\gamma(t+\kappa) + \tfrac{1}{2}\kappa^2(\gamma^2(t+\kappa) + \frac{d}{dt}\gamma(t+\kappa)) + \ldots\right] \\
&\quad \times \exp\left(\int_{t_0}^{t+\kappa} dt' \gamma(t')\right) \exp\left(\kappa\gamma(t_0) - \tfrac{1}{2}\kappa^2 \frac{d}{dt}\gamma(t_0) + \ldots\right) f_0\,.
\end{aligned}
$$

We can absorb the factor resulting from variation of t_0 into the initial PDFs, f_0, so that $\exp\left(\kappa\gamma(t_0) - \tfrac{1}{2}\kappa^2 \frac{d}{dt}\gamma(t_0) + \ldots\right) f_0 \to f_0$. Then, noting also that $\exp\left(\int_{t_0}^{t+\kappa} dt' \gamma(t')\right) f_0 = f(t+\kappa)$, we end up with the result

$$
\overline{f}(\alpha_s(t+\kappa),\kappa) = \left[1 - \kappa\gamma(t+\kappa) + \tfrac{1}{2}\kappa^2(\gamma^2(t+\kappa) + \frac{d}{dt}\gamma(t+\kappa)) + \ldots\right] f(t+\kappa)\,.
$$

$$(3.2.38)$$

This result, which comes from varying the scale at which the PDF is evaluated, is equivalent to the result which comes from varying the scale in the anomalous dimension, Eqn. 3.2.35. This is because there is only one RGE and therefore only one scale which the PDF depends on. Furthermore, note that Eqn. 3.2.38 shows us that the scale dependence can be factorised out of the PDF. This means we are free to instead factor it into the parton level coefficient function, which will always appear convolved with the PDF. This is useful when making a scale varied prediction when only a central PDF is available, and has been used in the past to make predictions for new LHC processes (e.g. Higgs production [85]). However, in the case where we also want to consider μ_R variation in the coefficient function, the two scale variations will be mixed up, and this can lead to a complicated interplay, especially in the presence of heavy quark effects. In this work we adopt the method of scale variation for PDFs by using Eqn. 3.2.38.

3.2.4 Varying two scales together

We now consider the simultaneous variation of μ_R in the parton level observable and μ_F in the PDFs, in order to explore the full range of scale variation space. We will derive formulae for scale variation up to NNLO which are needed to

construct a MHOU covariance matrix.

For a DIS process we can write the double-scale-varied structure function as

$$\overline{F}(Q^2, \mu_F^2, \mu_R^2) = \overline{C}(\alpha_s(\mu_R^2), \mu_R^2/Q^2) \otimes \overline{f}(\alpha_s(\mu_F^2), \mu_F^2/Q^2). \tag{3.2.39}$$

Similarly to before, we can define the variables

$$\mu_{(F/R)}^2 = k_{(F/R)}Q^2, \qquad \kappa_{(F/R)} = \ln k_{(F/R)}, \qquad t_{(F/R)} = t + \kappa_{(F/R)} \tag{3.2.40}$$

and use them to write the structure function as

$$\overline{F}(t, \kappa_F, \kappa_R) = \overline{C}(t_R, \kappa_R)\overline{f}(t_F, \kappa_F). \tag{3.2.41}$$

We then need to apply the equations for the scale varied PDFs and coefficient functions from the previous section,

$$\begin{aligned}
\overline{f}(t_F, \kappa_F) &= f(t_F) - \kappa_F \tfrac{d}{dt} f(t_F) + \tfrac{1}{2}\kappa_F^2 \tfrac{d^2}{dt^2} f(t_F) + \dots, \\
\overline{C}(t_R, \kappa_R) &= C(t_R) - \kappa_R \tfrac{d}{dt} C(t_R) + \tfrac{1}{2}\kappa_R^2 \tfrac{d^2}{dt^2} C(t_R) + \dots,
\end{aligned} \tag{3.2.42}$$

and use the fact that $\frac{\partial}{\partial t} \sim \mathcal{O}(\alpha_s)$ to see that

$$\begin{aligned}
\overline{F}(t, \kappa_F, \kappa_R) = \ & C(t_R)f(t_F) - \left(\kappa_F C(t_R)\tfrac{d}{dt} f(t_F) + \kappa_R \tfrac{d}{dt} C(t_R)f(t_F)\right) \\
& + \tfrac{1}{2}\left(\kappa_F^2 C(t_R)\tfrac{d^2}{dt^2} f(t_F) + 2\kappa_R\kappa_F \tfrac{d}{dt} C(t_R)\tfrac{d}{dt} f(t_F) \right. \\
& \left. + \kappa_R^2 \tfrac{d^2}{dt^2} C(t_R)f(t_F)\right) + \mathcal{O}(\alpha_s^3).
\end{aligned} \tag{3.2.43}$$

Taking a closer look, and comparing to Eqn. 3.2.39, we can write this in terms of partial derivatives of F:

$$\begin{aligned}
\overline{F}(t, \kappa_F, \kappa_R) = \ & F(t_F, t_R) - \left(\kappa_F \left.\frac{\partial F}{\partial t_F}\right|_{t_R} + \kappa_R \left.\frac{\partial F}{\partial t_R}\right|_{t_F}\right) \\
& + \tfrac{1}{2}\left(\kappa_F^2 \left.\frac{\partial^2 F}{\partial t_F^2}\right|_{t_R} + 2\kappa_F\kappa_R \frac{\partial^2 F}{\partial t_F \partial t_R} \right. \\
& \left. + \kappa_R^2 \left.\frac{\partial^2 F}{\partial t_R^2}\right|_{t_F}\right) + \cdots.
\end{aligned} \tag{3.2.44}$$

Considering this expression, we can think of $\kappa_F \frac{\partial}{\partial t_F}$ and $\kappa_R \frac{\partial}{\partial t_R}$ as being the generators of the two types of scale variations.

46

For hadron-hadron processes, the double-scale-varied cross section is instead

$$\overline{\Sigma}(t_F, t_R, \kappa_F, \kappa_R) = \overline{H}(\alpha_s(t_R), \kappa_R) \otimes \left(\overline{f}(t_F, \kappa_F) \otimes \overline{f}(t_F, \kappa_F) \right) . \qquad (3.2.45)$$

and we can apply exactly the same approach as above, leading to

$$
\begin{aligned}
\overline{\Sigma}(t, \kappa_F, \kappa_R) \;=\; & \Sigma(t_F, t_R) - \left(2\kappa_F \left. \frac{\partial \Sigma}{\partial t_F} \right|_{t_R} + \kappa_R \left. \frac{\partial \Sigma}{\partial t_R} \right|_{t_F} \right) \\
& + \tfrac{1}{2} \left(2\kappa_F^2 \left. \frac{\partial^2 \Sigma}{\partial t_F^2} \right|_{t_R} + 4\kappa_F \kappa_R \frac{\partial^2 \Sigma}{\partial t_F \partial t_R} \right. \\
& \left. + \kappa_R^2 \left. \frac{\partial^2 \Sigma}{\partial t_R^2} \right|_{t_F} \right) + \cdots ,
\end{aligned}
\qquad (3.2.46)
$$

where this time we pick up a factor of 2 with each $\kappa_F \frac{\partial}{\partial t_F}$, due to the two PDFs.

3.2.5   Scale variation for multiple processes

We are now approaching a formalism which can be applied to carry out scale variations for the data included in PDF fits. But first we need to work out how to carry out simultaneous scale variations involving data from more than one process, for example DIS and Drell-Yan.

For the case of two separate processes, the parton level cross sections will be totally independent, so there will be two separate RGEs and therefore two separate renormalisation scales, and hence renormalisation scale variation should be uncorrelated. However, all the processes share a common factorisation scale, and so the factorisation scale variation must be correlated.

This correlation can be complex because the DGLAP equation is a matrix equation, and the anomalous dimension matrix has several independent eigenvalues (at NLO there are two singlet and one non-singlet, and more at higher orders). In order to fully preserve the correlations we ought to consider a separate factorisation scale for each of these components, and fully correlate this across all processes. In this current work, however, we attempt to reduce the complexity by retaining full correlation in the factorisation scale, only varying one scale. This approximation might be inaccurate when considering two processes with evolution dependent on different anomalous dimensions, in which case we would not be fully exploring the scale variation space. We draw attention to this limitation as an area of future study.

Sticking for the time being with correlated factorisation scale variation, for two processes we will have in general three scales: two renormalisation scales and one factorisation scale; μ_{R_1}, μ_{R_2} and μ_F. If we vary each scale independently by a factor of 2 about their central value we will have seven total scale choices to consider. Each time we add another process we will add another renormalisation scale, and in effect add another dimension to the scale variation. For p independent processes, labelled $a = 1, \ldots, p$, there will be $p + 1$ independent scale parameters and $3 + 2p$ total scale variations; one variation is the central scale, two variations up and down for the factorisation scale, and two variations up and down for each of the p renormalisation scales.

3.3   Building the theory covariance matrix

We now have all the components we need to set about constructing a theory covariance matrix; we can carry out double scale variation for both DIS and hadron-hadron processes, and correlate scale variation between multiple processes. All that remains is to formulate a prescription for estimating MHOUs given scale variations.

In Chapter 2 we formulated a method for including theory uncertainties in PDFs using a theory covariance matrix composed using a distribution of shifts between theory predictions at fixed order and the unknown all-order "true" theory. We know that scale variation can be used to provide an estimate of these shifts, but as discussed previously the exact combination of scales is arbitrary. To address this problem, we present a series of prescriptions for constructing the theory covariance matrix, which we will later go on to test using results at known orders and, in addition, by studying the impact on the PDFs; we can then select the best prescription.

We consider a set of data involving p different types of processes, each with data points $\{i_a\}$, $a = 1, \ldots, p$ and an associated renormalisation scale ratio $\kappa_{R_a} = \ln \mu_{R_a}^2 / Q^2$. The theory covariance matrix can be constructed by taking an average over the outer products of the shifts in scale varied theory relative to the central theory. For the a-th process, these shifts are

$$\Delta_{i_a}(\kappa_F, \kappa_{R_a}) \equiv T_{i_a}(\kappa_F, \kappa_{R_a}) - T_{i_a}(0, 0) \,. \tag{3.3.1}$$

For a given prescription, m, we then choose a set of points, V_m, in $p+1$-dimensional scale variation space, and construct the theory covariance matrix by summing over these points, normalised by a prescription-dependent factor N_m:

$$S_{ij} = N_m \sum_{V_m} \Delta_{i_a}(\kappa_f, \kappa_{R_a}) \Delta_{i_b}(\kappa_f, \kappa_{R_b}). \qquad (3.3.2)$$

Note that a and b can label either the same or different processes. Importantly, since the covariance matrix is assembled as a sum of outer products it will necessarily be positive semi-definite. However, given that the dimension of the data is $\mathcal{O}(1000)$ and the dimension of V_m will be in general considerably smaller, we expect S to be singular in most instances.

It now remains to develop a prescription, m. We must consider the full set of data, so there are two scenarios:

1. i and j belong to the same process;

2. i and j belong to different processes.

Because S is rank 2, we only need to consider a maximum of two different processes at any one time. Finally, we can choose how to correlate the scale variation; we can pick a "symmetric prescription", in which the scales are varied independently, or an "asymmetric prescription", where they are correlated. This second scenario amounts to setting $\mu_F = \mu_R$, which is like varying the scale in the physical cross section; because in a central prediction we typically set both scales to the physical scale of the process (e.g. Q for DIS), we can call this varying the scale of the process.

3.3.1 Symmetric prescriptions

In a symmetric prescription, the scales are varied in an uncorrelated way.

One process

For a single process ($p = 1$), there are two scales, κ_F and κ_R. Let us denote the number of independent scales as s, so here $s = 2$. We can write the theory
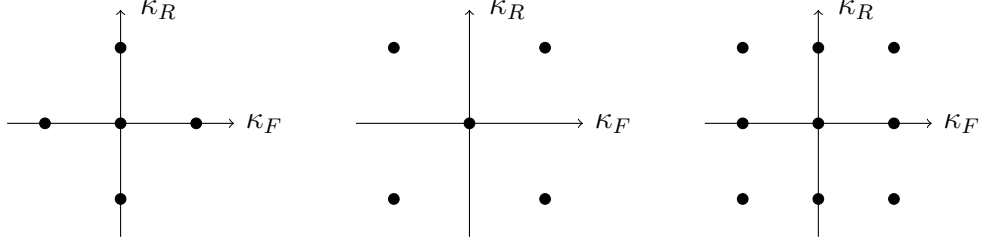
Figure 3.3.1 *Symmetric prescriptions for a single process, indicating the sampled values for the factorisation scale κ_F and renormalisation scale κ_R in each case. The origin of coordinates corresponds to the central scales $\kappa_F = \kappa_R = 0$. We show the three prescriptions 5-point (left), $\bar{5}$-point (middle) and 9-point (right).*

covariance matrix as

$$S_{ij} = n_m \sum_{v_m} \Delta_i(\kappa_F, \kappa_R)\Delta_j(\kappa_F, \kappa_R)\,, \qquad (3.3.3)$$

where v_m is the set of m scale-varied points and n_m is a normalisation factor, to be determined. Note that v_m excludes any points for which Δ_i vanishes, since these will not contribute to S. In practice, this means the central point ($\kappa_F = \kappa_R = 0$) is not included. Overall there is one central point and m variations about it, so we typically refer to a given prescription as an "($m + 1$)-point prescription". We can find n_m by summing over the number of independent scales, s, and averaging over the contributions from each scale, m. This means we can write

$$n_m = s/m. \qquad (3.3.4)$$

We will now outline three different prescriptions, depicted in Fig. 3.3.1. In each case we denote the values of the scales $(\kappa_F; \kappa_R)$, varying each by the fixed values $\kappa = \{0, \pm \ln 4\}$, which we denote $\{0, \pm\}$ respectively. We also adopt the notation $\Delta_i^{+0} = \Delta_i(+\ln 4, 0)$, $\Delta_i^{0-} = \Delta_i(0, -\ln 4)$, etc. for the shifts.

- **5-point**: $v_4 = \{(\pm; 0), (0; \pm)\}$ and $n_4 = 2/4 = 1/2$. This amounts to scale variation for each scale in turn, keeping the other fixed. We find the covariance matrix

$$S_{ij}^{(5\text{pt})} = \tfrac{1}{2}\{\Delta_i^{+0}\Delta_j^{+0} + \Delta_i^{-0}\Delta_j^{-0} + \Delta_i^{0+}\Delta_j^{0+} + \Delta_i^{0-}\Delta_j^{0-}\}\,. \qquad (3.3.5)$$

  We find that the variations of each scale are added together in quadrature, as one would expect for independent contributions to the MHOU.

- $\overline{5}$-**point**: $\overline{v}_4 = \{(\pm; \pm)\}$, where $(\pm; \pm)$ are assumed uncorrelated, and $\overline{n}_4 = 2/4 = 1/2$. This is the complement of 5-point, exploring a different region of scale variation space.

$$S_{ij}^{(\overline{5}\text{pt})} = \tfrac{1}{2}\{\Delta_i^{++}\Delta_j^{++} + \Delta_i^{--}\Delta_j^{--} + \Delta_i^{+-}\Delta_j^{+-} + \Delta_i^{-+}\Delta_j^{-+}\}. \qquad (3.3.6)$$

- **9-point**: $v_8 = v_4 \oplus \overline{v}_4$ (the union of 5-point and $\overline{5}$-point) and $n_8 = 2/8 = 1/4$. Here we include every combination, varying the scales totally independently.

$$
\begin{aligned}
S_{ij}^{(9\text{pt})} = \tfrac{1}{4}\{ & \Delta_i^{+0}\Delta_j^{+0} + \Delta_i^{-0}\Delta_j^{-0} + \Delta_i^{0+}\Delta_j^{0+} + \Delta_i^{0-}\Delta_j^{0-} \\
& +\Delta_i^{++}\Delta_j^{++} + \Delta_i^{+-}\Delta_j^{+-} + \Delta_i^{-+}\Delta_j^{-+} + \Delta_i^{--}\Delta_j^{--}\}.
\end{aligned}
\qquad (3.3.7)
$$

Two processes

In the case that $p = 2$ we can have either uncorrelated or partially correlated scale variations. We will have $p + 1 = 3$ independent scales to vary, and our set of scale variation points, V_m, will be much larger than for one process (v_m). If we label the processes $a = 1, b = 2$, we can view the two-process covariance matrix as

$$S_{ij} = \begin{pmatrix} S_{i_1 j_1} & S_{i_1 j_2} \\ S_{i_2 j_1} & S_{i_2 j_2} \end{pmatrix}, \qquad (3.3.8)$$

so the diagonal elements deal with data points in the same process, and the off-diagonals deal with data points in different processes. For the diagonal blocks, the form of S must be equivalent to the $p = 1$ case, and so

$$S_{i_1 j_1} = N_m \sum_{V_m} \Delta_{i_1}(\kappa_F, \kappa_{R_1})\Delta_{j_1}(\kappa_F, \kappa_{R_1}) = n_m \sum_{v_m} \Delta_{i_1}(\kappa_F, \kappa_{R_1})\Delta_{j_1}(\kappa_F, \kappa_{R_1}).$$
$$(3.3.9)$$

This means that v_m must be a subset of V_m, so that if we sum over V_m setting κ_{R_2} to 0, we should recover v_m up to a degeneracy factor, d_m, which is the number of copies of v_m in V_m. This means the overall normalisation factor is

$$N_m = n_m/d_m. \qquad (3.3.10)$$

We now go on to consider the 5-point, $\overline{5}$-point and 9-point prescriptions for the case of two processes, depicted in Fig. 3.3.2. We expand the notation to include three scales, that is $(\kappa_F; \kappa_{R_1}, \kappa_{R_2})$.
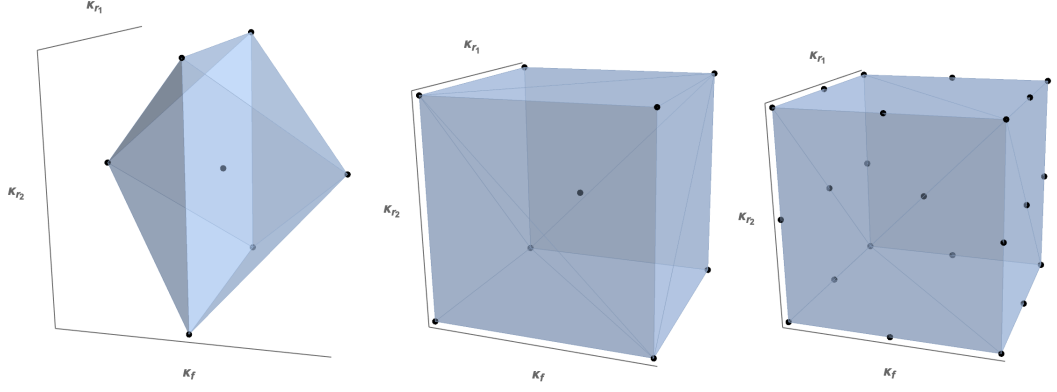
Figure 3.3.2 *Same as Fig. 3.3.1, now for the case of two different processes with a common factorisation scale, κ_F, and different renormalisation scales, κ_{R_1} and κ_{R_2}, so the diagrams are now in 3d. We again show the three prescriptions 5-point (left), $\bar{5}$-point (middle) and 9-point (right).*

- **5-point**: We vary the factorisation scale holding the renormalisation scales fixed, and the renormalisation scales holding the factorisation scale fixed, so that $V_4 = \{2(\pm; 0, 0), (0; \pm, \pm)\}$. This means V_4 has eight elements in total. The factor of two comes from considering the one-process case, where we can set $\kappa_{R_2} = 0$, and must return a multiple of v_4. Explicitly, we will get $\{2(\pm; 0, 0), 2(0; \pm, 0)\}$, picking up a factor of two on the second term because there are two terms implicitly here, for $\kappa_{R_2} = +$ and $-$. We need to include a factor of two from the outset on the first term so we end up with an overall factor at the end, in this case $d_4 = 2$. This means $N_4 = (1/2)/2 = 1/4$. The off-diagonal blocks of the covariance matrix are evaluated as

$$S_{i_1 j_2}^{(5\mathrm{pt})} = \tfrac{1}{4}\{2\Delta_{i_1}^{+0}\Delta_{j_2}^{+0} + 2\Delta_{i_1}^{-0}\Delta_{j_2}^{-0} + (\Delta_{i_1}^{0+} + \Delta_{i_1}^{0-})(\Delta_{j_2}^{0+} + \Delta_{j_2}^{0-})\}. \quad (3.3.11)$$

When generalising to 3 processes, we simply write $V_4 = \{4(\pm; 0, 0, 0), (0; \pm, \pm, \pm)\}$, and this time $d_4 = 4$ by the same arguments as above. However, we can still use Eqn. 3.3.11 to evalutate all of the off-diagonal blocks because we will only ever be considering two of the three processes at a time.

- **$\bar{5}$-point**: $\overline{V}_4 = \{(\pm; \pm, \pm)\}$, so we are essentially exploring the outer corners of the 3d scale variation space. $\overline{V}_4$ has eight elements, and there are two elements for each in $\overline{v}_4$, meaning that $\overline{N}_4 = (1/2)/2 = 1/4$. The off-diagonal

blocks of the covariance matrix are evaluated as

$$S_{i_1 j_2}^{(\overline{5}\text{pt})} = \tfrac{1}{4}\{(\Delta_{i_1}^{++} + \Delta_{i_1}^{+-})(\Delta_{j_2}^{++} + \Delta_{j_2}^{+-}) + (\Delta_{i_1}^{-+} + \Delta_{i_1}^{--})(\Delta_{j_2}^{-+} + \Delta_{j_2}^{--})\}\,.$$

$$(3.3.12)$$

For three processes the generalisation is simply $\overline{V}_4 = \{(\pm; \pm, \pm, \pm)\}$, with $\overline{N}_4 = 1/8$.

- **9-point**: Again, all three scales are varied completely independently and $V_8 = \{3(0; \pm, \pm), 2(\pm; \oplus, \oplus)\}$, where $\oplus$ means either $+, -$ or 0. Note the factors of 2 and 3, which arise from arguments similar to that for 5-point; namely, that V_8 must reduce to a multiple of v_8 when κ_{R_2} is set to 0. The overall $d_8 = 6$ and so $N_8 = 1/24$. The off-diagonal blocks of the covariance matrix are

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

$$(3.3.13)$$

For three processes, $V_8 = \{9(0; \pm, \pm, \pm), 4(\pm; \oplus, \oplus, \oplus)\}$ and $d_8 = 36$.

3.3.2 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". This is like setting $\kappa_F = \kappa_R$. In addition, it is also possible to consider varying the scale of the process on top of the variation of factorisation and renormalisation scales already considered. The logic behind this is that the three scales each estimate a different source of MHOU:

- Varying the scale of the process estimates the MHOU on the hard cross section which is proportional to collinear logarithms;

- Varying the renormalisation scale estimates the MHOU on the hard cross section which is proportional to the beta function;

- Varying the factorisation scale estimates the MHOU in the anomalous dimension.
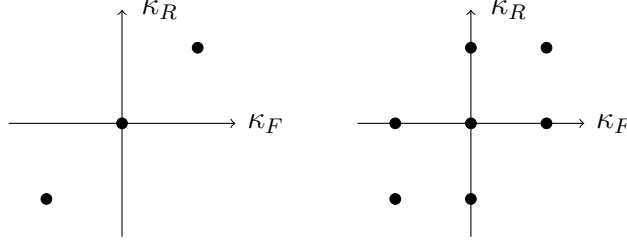
Figure 3.3.3 *Same as Fig. 3.3.1, now in the case of the asymmetric prescriptions for a single process with factorisation scale κ_F and renormalisation scale κ_R. We display the 3-point (left) and 7-point (right) prescriptions, defined in the text.*

However, both of these approaches will suppress correlations between uncertainties in PDF evolution across different processes, and may therefore overestimate the MHOU. Ultimately, the best scheme, be it a symmetric or an asymmetric prescription, must be established through a validation procedure such as the one outlined in later sections.

We will now formulate prescriptions for these two asymmetric prescriptions, being the 3-point and 7-point prescriptions, respectively. These are depicted for a single process in Fig. 3.3.3.

- **3-point**: We set $\kappa_F = \kappa_R$ and vary this scale. We have $v_2 = \{\pm\}$ and $s = 1$, $m = 2$ and $n_2 = 1/2$, so we are just averaging over the two scale-varied options. For a single process

$$S_{ij}^{(\mathrm{3pt})} = \tfrac{1}{2}\{\Delta_i^{++}\Delta_j^{++} + \Delta_i^{--}\Delta_j^{--}\}. \qquad (3.3.14)$$

For two different processes we have $V_2 = \{\pm, \pm\}$, and can see explicitly that we are ignoring the correlations in κ_F between the two processes. We have $d_2 = 2$ and so $N_2 = (1/2)/2 = 1/4$, and the off-diagonal blocks of the covariance matrix are evaluated as

$$S_{i_1 j_2}^{(\mathrm{3pt})} = \tfrac{1}{4}\{(\Delta_{i_1}^{++} + \Delta_{i_1}^{--})(\Delta_{j_2}^{++} + \Delta_{j_2}^{--})\}. \qquad (3.3.15)$$

- **7-point**: We combine varying the scale of the process with κ_R and κ_F variation. We will end up with essentially a combination of 3-point (scale of process) and 5-point (individual factorisation and renormalisation). For a single process, $v_6 = \{(\pm; 0), (0; \pm), (+; +), (-; -)\} = \{(\pm; 0), (0; \pm), (\mp; \pm)\}$, where $(\mp; \pm)$ simply means that the variation is fully correlated (so there are only 2 terms, not 4). Then $s = 2$ and $m = 6$

so $n_6 = 1/3$, and for a single process

$$S_{ij}^{(7\mathrm{pt})} = \tfrac{1}{3}\{\Delta_i^{+0}\Delta_j^{+0} + \Delta_i^{-0}\Delta_j^{-0} + \Delta_i^{0+}\Delta_j^{0+} + \Delta_i^{0-}\Delta_j^{0-} + \Delta_i^{++}\Delta_j^{++} + \Delta_i^{--}\Delta_j^{--}\}\,.$$

(3.3.16)

For more than one process, variations of the scale of the process are uncorrelated between processes so the μ_F variation enclosed in the scale of the process variation will be decorrelated. So overall for two processes we will be in a 4d scale variation space, $(\kappa_{F_1}, \kappa_{R_1}; \kappa_{F_2}, \kappa_{R_2})$. Then $V_6 = \{2(+, 0; +, 0), 2(-, 0; -, 0), (0, \pm; 0, \pm), (\overline{\pm, \pm; \pm, \pm})\}$, where $(\overline{\pm, \pm; \pm, \pm}) = \{(+, +; +, +), (+, +; -, -), (-, -; +, +), (-, -; -, -)\}$, and thus $d_6 = 2$, so $N_6 = 1/6$, and the off-diagonal theory covariance matrix reads

$$S_{i_1 j_2}^{(7\mathrm{pt})} = \tfrac{1}{6}\{2\Delta_{i_1}^{+0}\Delta_{j_2}^{+0} + 2\Delta_{i_1}^{-0}\Delta_{j_2}^{-0} + (\Delta_{i_1}^{0+} + \Delta_{i_1}^{0-})(\Delta_{j_2}^{0+} + \Delta_{j_2}^{0-})$$
$$+ (\Delta_{i_1}^{++} + \Delta_{i_1}^{--})(\Delta_{j_2}^{++} + \Delta_{j_2}^{--})\}\,.$$

(3.3.17)

3.4 Results for the theory covariance matrix

In this section we summarise the data used to calculate theory covariance matrices using the prescriptions in the previous section. We divide the data into processes, each with a distinct renormalisation scale. We present the theory covariance matrices for each prescription at NLO.

3.4.1 Input data and process categorisation

In order to use these prescriptions, we must first divide our data into distinct "processes". There is some degree of arbitrariness here, but the idea is to group together data which might have a similar structure of MHOUs under a common renormalisation scale. First we will review the data included in these fits, then we will outline the process categorisation. This is summarised in Table 3.4.1.

The data considered here are a mildly altered version of those in NNPDF3.1 [9]. More exactly, they include: fixed-target [86–93] and HERA [94] DIS structure functions; charm cross-sections from HERA [95]; gauge boson production from the Tevatron [96–99]; electroweak boson production, inclusive jet, Z p_T distributions, and $t\bar{t}$ total and differential cross-sections from ATLAS [74, 100–107], CMS [73, 108–115] and LHCb [116–119]. In total they make up 2819 data points.

Process Type	Dataset	Reference	N_dat	N_dat (process)
DIS NC	NMC	[86, 87]	134	
	SLAC	[88]	12	
	BCDMS	[89, 90]	530	1593
	HERA σ_{NC}^p	[94]	886	
	HERA σ_{NC}^c	[95]	31	
DIS CC	NuTeV dimuon	[91, 92]	41	
	CHORUS	[93]	430	552
	HERA σ_{CC}^p	[94]	81	
DY	ATLAS W, Z, 7 TeV 2010	[100]	30	
	ATLAS W, Z, 7 TeV 2011	[101]	34	
	ATLAS low-mass DY 2011	[102]	4	
	ATLAS high-mass DY 2011	[103]	5	
	ATLAS Z p_T 8 TeV (p_T^{ll}, M_{ll})	[74]	44	
	ATLAS Z p_T 8 TeV (p_T^{ll}, y_Z)	[74]	48	
	CMS Drell-Yan 2D 2011	[108]	88	
	CMS W asy 840 pb	[109]	11	
	CMS W asy 4.7 pb	[110]	11	
	CMS W rap 8 TeV	[111]	22	484
	CMS Z p_T 8 TeV (p_T^{ll}, M_{ll})	[73]	28	
	LHCb Z 940 pb	[116]	9	
	LHCb $Z \to ee$ 2 fb	[117]	17	
	LHCb $W, Z \to \mu$ 7 TeV	[118]	29	
	LHCb $W, Z \to \mu$ 8 TeV	[119]	30	
	CDF Z rap	[96]	29	
	D0 Z rap	[97]	28	
	D0 $W \to e\nu$ asy	[98]	8	
	D0 $W \to \mu\nu$ asy	[99]	9	
JET	ATLAS jets 2011 7 TeV	[104]	31	164
	CMS jets 7 TeV 2011	[112]	133	
TOP	ATLAS σ_{tt}^top	[105, 106]	3	
	ATLAS $t\bar{t}$ rap	[107]	10	26
	CMS σ_{tt}^top	[113, 114]	3	
	CMS $t\bar{t}$ rap	[115]	10	
Total			**2819**	**2819**

Table 3.4.1 *Input data and process categorisation. 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).*

We identify five categories to divide the data into: neutral current DIS (DIS NC); charged current DIS (DIS CC); Drell-Yan (DY); jet production (JET); and top production (TOP). These represent groups of data with similar expected MHO terms. For instance, predictions for DY processes can be differential in differing variables, but all are obtained by integrating the same underlying, fully differential, distribution. This means they should have a similar perturbative structure. On the other hand, these distributions constrain different PDFs and so complex correlations between the PDFs will be introduced. Note that DIS is split up into CC and NC due to the differing structure of interaction.

3.4.2 NLO theory covariance matrices

All calculations are performed using the same settings as in [9]. Theoretical predictions are provided by `APFEL` [43] for the DIS structure functions and by `APFELgrid` [120] combined with `APPLgrid` [121] for the hadronic cross-sections. They are all evaluated using the central NLO PDF obtained by performing a NLO fit to the same dataset, for consistency. The resulting predictions are then used to construct the theory covariance matrices using the `ReportEngine` software [1].

We now present the theory covariance matrices at NLO, and discuss their features. In Fig. 3.4.1 we compare the square root of the diagonal elements of the experiment (C) and theory (S) covariance matrices, normalised to experimental data values; these are equivalent to the fractional per-point uncertainties. In this and what follows, the data are grouped by process and, within each process, by dataset according to Table 3.4.1, wherein they are binned according to kinematics (the same as in the experimental papers). S is the 9-pt covariance matrix, being that with the largest number of independent scale variations. On the whole, the size of NLO MHOUs is comparable to the experimental uncertainties, and we can see that each will dominate in different regions. One particularly striking instance of this is seen in the HERA data, which can be found in the middle of DIS NC: for high (x, Q^2) there is low statistics and so the experimental uncertainties dominate; for low (x, Q^2) perturbation theory holds less strongly and the MHOUs dominate. We can also explore the structure of correlations between data points by viewing the covariance matrices as a whole.

In Fig. 3.4.2 we show C alongside S (9-point) as a % of experimental data value. It is immediately obvious that S contains a much richer and more vibrant structure than C; experimental correlations only exist within experiments as the experiments are isolated from one another. However, predictions for these experiments originate from a common theoretical framework, and therefore theory uncertainties can exist between any two data points, regardless of experimental origin. In particular, data points within the same process are assigned a common renormalisation scale, inducing correlations between them. Furthermore, all points are predicted using the same PDFs and, in the 9-point prescription, share a common factorisation scale. This introduces entries in S outwith individual processes. Along the block diagonal we note that correlations within an individual experiment are broadly positive, due to these points sharing a close kinematic region. In other regions we see a mixture of positive and negative correlations,

57
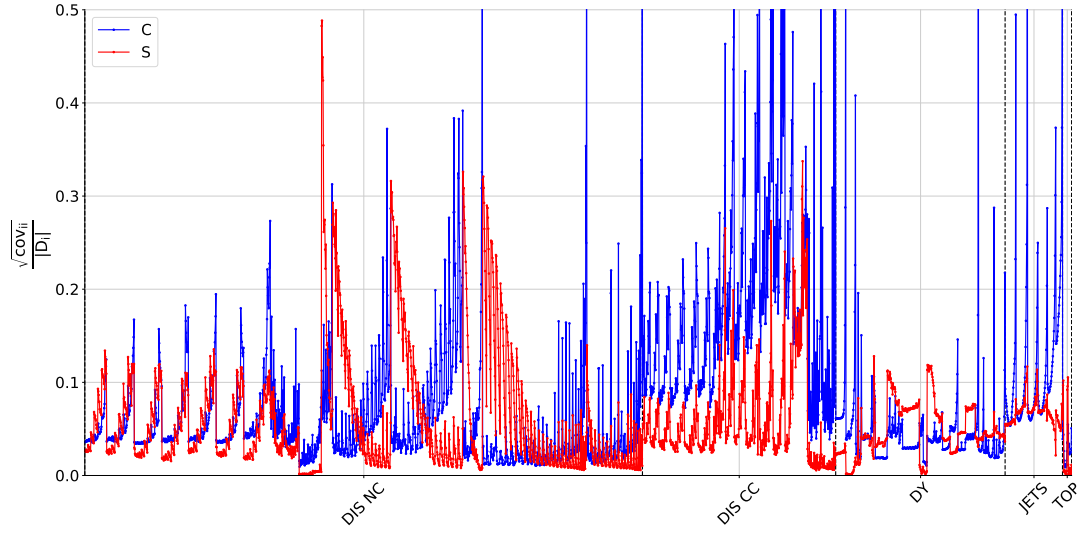
Figure 3.4.1 *Comparison of the per-point experimental (blue) and 9-point theoretical (red) uncertainties, normalised to data. In this and what follows, data are grouped by process and, within each process, by dataset, following Table 3.4.1*
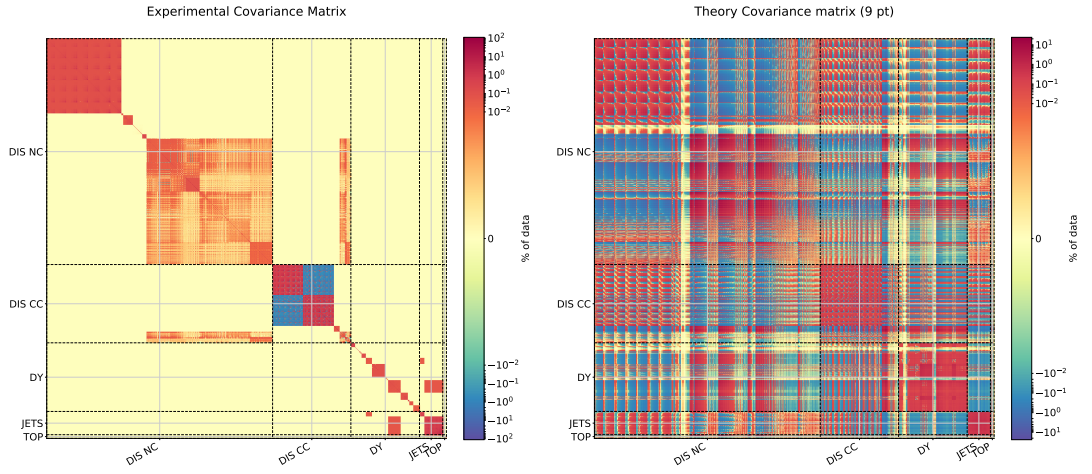
Figure 3.4.2 *Comparison of the experimental C_{ij} (left) and the 9-point theoretical S_{ij} (right) covariance matrices. Entries are displayed as a percentage of the experimental value.*

for example HERA is generally positively correlated with DY but negatively correlated with fixed target DIS.
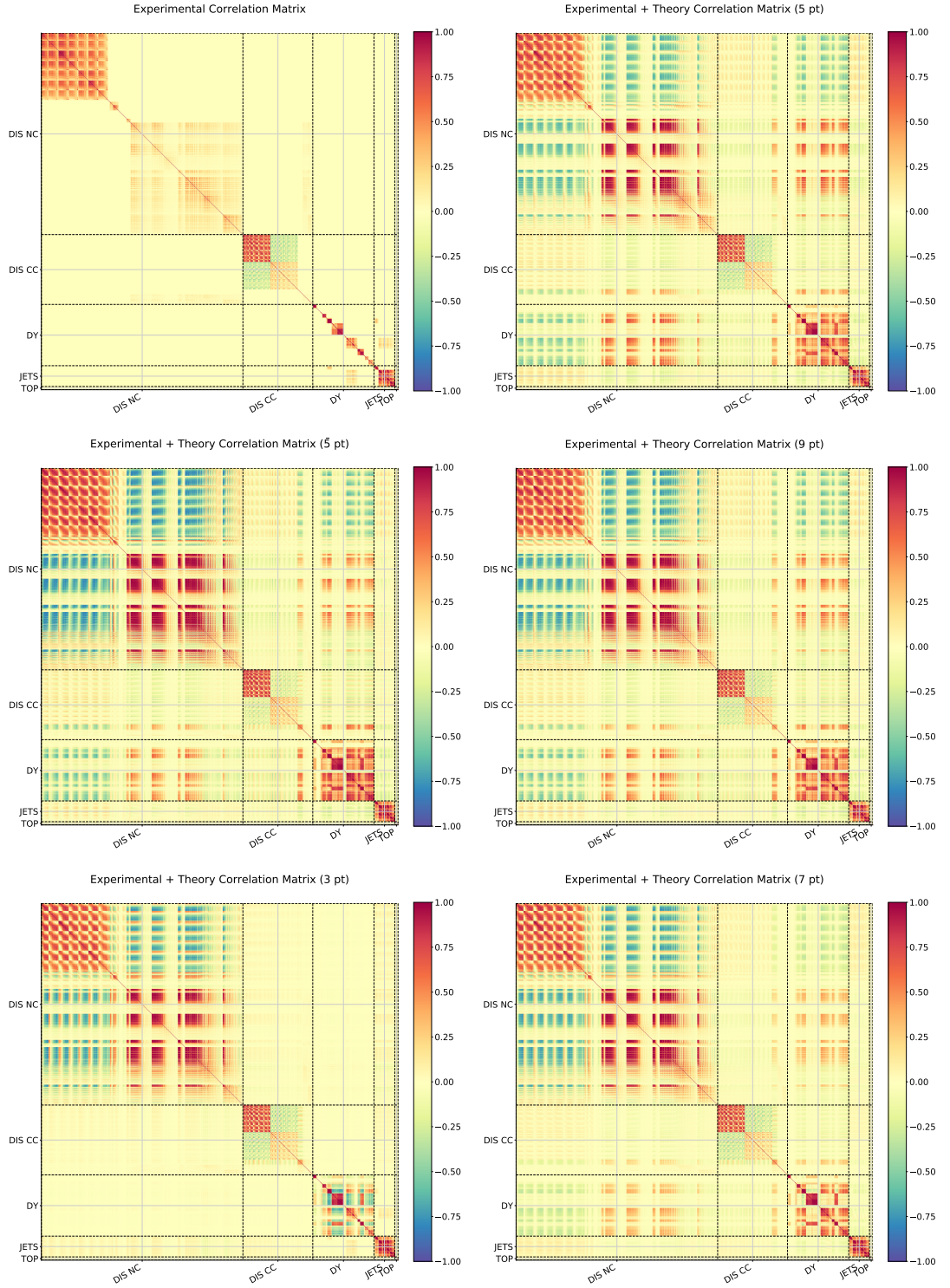
Figure 3.4.3 *Comparison of the experimental correlation matrix (top left) and the the combined experimental and theoretical correlation matrices computed using the prescriptions described in Sec. 3.3: the symmetric prescriptions (5-pt top right, $\overline{5}$-pt centre left, 9-pt centre right); and asymmetric prescriptions (3-pt bottom left, 7-pt bottom right).*

The exact structure of S is dependent on the chosen prescription in Sec. 3.3. In Fig. 3.4.3 we delve into these differences by comparing correlation matrices for each prescription, defined

$$\text{corr}_{ij}^A \equiv \frac{A_{ij}}{\sqrt{A_{ii}}\sqrt{A_{jj}}}. \qquad (3.4.1)$$

This removes the effect of the differing magnitude of entries, laying bare the underlying structure; a value of 1 corresponds to full correlation between two points, a value of 0 corresponds to no correlation and a value of -1 corresponds to full anticorrelation. This time we look at the impact of adding the theory covariance matrix to the experiement covariance matrix. In all cases a richer structure emerges, however we note that:

- For 3-point the correlations between processes are a lot weaker, and this is because both factorisation and renormalisation scale are uncorrelated between processes;

- For 7-point (the other asymmetric prescription), the correlations are weaker than in 5-point despite the fact that 7-point uses the same scale variation points as 5-point plus two additional ones. This is because the μ_F variation is combined with the uncorrelated "scale of process" variation;

- All three symmetric prescriptions show similar patterns of correlation.

From this, it seems that any of the symmetric prescriptions might be a suitable choice. However, in the next section we outline more quantitative tests to validate whether or not each prescription provides a reasonable estimate of MHOUs, and hence determine the best prescription.

3.5   Validating the theory covariance matrix

Whilst in general the structure of MHOs is unknown to us, due to the perturbative nature of calculations we do know that the MHOU is dominated by the next MHO; so at NLO we expect the MHOU to be dominated by the NNLO terms. A good test of the NLO theory covariance matrix, therefore, is that it reliably encapsulates the known NNLO predictions. In this section we describe a method

of validation based on this observation.We summarise the procedure here, before going into some considerable detail in Sec. 3.5.1.

We consider the space of experimental data, D, spanned by the vector of experimental data points, D_i, $i = 1, \ldots, N_{dat}$. The theory covmat, S_{ij} acts as a linear operator on this space. We know that S_{ij} is positive semi-definite and symmetric, and therefore has positive or zero eigenvalues only. As an uncertainty matrix, S_{ij} defines ellipsoids, E, of given confidence level. These lie in the subspace $S \in D$, and are centred on the NLO prediction, T_i^{NLO}. A test of the efficacy of S_{ij} is that the 1-σ ellipse broadly encapsulates the known shift to the next higher order, $\delta_i \sim T_i^{NNLO} - T_i^{NLO}$. Note that here T_i^{NNLO} must be evaluated using the same NLO PDF, to ensure that the shift is due only to perturbative differences and not to effects from refitting. This is a robust test, owing to the great difference between dim $D \sim 1000$ and dim $S \sim 10$; for a random matrix we would expect only 1% of δ_i to lie in S.

3.5.1 Details of validation procedure

Recall that the ellipse $E \in S \in D$, where dim $D = N_{dat}$ and dim $S = N_{sub}$. We define dimensionless quantities by normalising to the theory prediction at NLO:

$$\widehat{S}_{ij} = \frac{S_{ij}}{T_i^{NLO}T_j^{NLO}}; \qquad \delta_i = \frac{T_i^{NNLO} - T_i^{NLO}}{T_i^{NLO}}. \tag{3.5.1}$$

We expect the component of δ_i along each axis of the ellipse, E, to be the same order as the 1-σ ellipse. Physically, this means that the eigenvectors of S_{ij} correctly estimate all the independent directions of uncertainty in theory space. The size of the shift in each direction is given by the corresponding eigenvalue. The null subspace of E, $i.e.$, directions with vanishing eigenvalues, corresponds to directions in D where the theory uncertainty is very small and can be neglected.

We can find the eigenvectors and eigenvalues of $\widehat{S}_{ij}$ by diagonalising it. The non-zero eigenvalues are $\lambda^\alpha \equiv (s^\alpha)^2$; $\alpha = 1, \ldots, N_{sub}$ and there are $N_{dat} - N_{sub}$ additional zero eigenvalues - this is a large number. We choose the eigenvectors, e_i^α, to be orthonormal such that

$$\sum_i e_i^\alpha e_i^\beta = \delta^{\alpha\beta}. \tag{3.5.2}$$

This diagonalisation procedure is somewhat involved owing to the large number of zero eigenvalues. To make this easier, we can project $\widehat{S}$ onto the subspace, S, where all the eigenvalues are positive definite by construction. We can then perform the diagonalisation here. We can make this projection by noting that S is spanned by the vectors used to construct S_{ij}, that is $\{\Delta_i(\kappa_F, \kappa_{R_a}) : \kappa_F, \kappa_{R_a} \in V_m\}$. Correspondingly, $\widehat{S}_{ij}$ is spanned by $\widehat{\Delta}_i \equiv \delta_i/T_i^{NLO}$. The caveat is that not all of these vectors are linearly independent, and so we must find a linearly independent subset of these on a case by case basis, of which there will be N_{sub}. We now consider each of the prescriptions in Sec. 3.3 in turn.

3-point

Here the factorisation scale is always correlated with the renormalisation scale variation so we can consider $(\kappa_{R_1}, \kappa_{R_2}, \ldots, \kappa_{R_p})$ only. The table below summarises the possible permutations of scale variations under this scheme; each type of scale variation configuration is displayed alongisde the number of permutations of this type.

3-point	
No. of vectors	$(\kappa_{R_1}, \kappa_{R_2}, \ldots, \kappa_{R_p})$
1	$(+, +, +, \ldots)$
1	$(-, -, -, \ldots)$
p	$(-, +, +, \ldots)$ and cyclic
p	$(+, -, -, \ldots)$ and cyclic

So we naïvely have $2 + 2p$ vectors in this space. However, these are not all linearly independent. Explicitly, we have the following restrictions:

- If we sum all of the cyclic permutations in the lower two rows we get $p(\ (+, +, +, \ldots) + (-, -, -, \ldots)\ )$, which is just p times the sum of the first two rows. This is one restriction;

- Each pair of complimentary cyclic permutations, *e.g.*, $(-, +, +, \ldots)$ and $(+, -, -, \ldots)$, sum to the sum of the first two rows. These are another p restrictions.

Overall this means we have $2 + 2p - 1 - p = p + 1$ linearly independent contributions, and so $N_{sub}^{3pt} = p + 1$. This means we can choose rows 1 and 3 as our linearly independent vectors.

62

5-point

We can apply similar arguments here, this time also considering the factorisation scale.

5-point	
No. of vectors	$\left(\kappa_F; \kappa_{R_1}, \kappa_{R_2}, \ldots, \kappa_{R_p}\right)$
2	$(\pm; 0, 0, 0, \ldots)$
1	$(0; -, -, -, \ldots)$
1	$(0; +, +, +, \ldots)$
p	$(0; -, +, +, \ldots)$ and cyclic
p	$(0; +, -, -, \ldots)$ and cyclic

Again we the same $p+1$ restrictions as in 3-point and so $N_{sub} = 3 + 2p - (p+1) = p + 3$. We choose rows 1, 3 and 4 as our linearly independent vectors.

$\overline{5}$-point

$\overline{5}$-point	
No. of vectors	$\left(\kappa_F; \kappa_{R_1}, \kappa_{R_2}, \ldots, \kappa_{R_p}\right)$
2	$(\pm; +, +, +, \ldots)$
2	$(\pm; -, -, -, \ldots)$
$2p$	$(\pm; -, +, +, \ldots)$ and cyclic
$2p$	$(\pm; +, -, -, \ldots)$ and cyclic

This time since we have $\pm$ for every possibility, there are $2(p + 1)$ restrictions, and so $N_{sub} = 4 + 4p - 2(p+1) = 2p + 2$. We choose rows 1 and 3 as our linearly independent vectors.

7-point

This is just the sum of 3-point and 5-point, so $N_{sub} = p + 1 + p + 3 = 2p + 4$. Likewise we combine the vectors from 3-point and 5-point.

9-point

9-point	
No. of vectors	$(\kappa_F; \kappa_{R_1}, \kappa_{R_2}, \ldots, \kappa_{R_p})$
3	$(\oplus; +, +, +, \ldots)$
3	$(\oplus; -, -, -, \ldots)$
$3p$	$(\oplus; -, +, +, \ldots)$ and cyclic
$3p$	$(\oplus; +, -, -, \ldots)$ and cyclic
2	$(\pm; 0, 0, 0, \ldots)$
$2p$	$(\pm; 0, -, -, \ldots)$ and cyclic
$2p$	$(\pm; 0, +, +, \ldots)$ and cyclic

This time we have $3(p+1)$ restrictions from the top four rows and an additional $2(p+1)$ from the bottom three, so overall $N_{sub} = 6 + 6p + 2 + 4p - 3(p+1) - 2(p+1) = 5p + 3$. We choose rows 1, 3 and 7 as our linearly independent vectors.
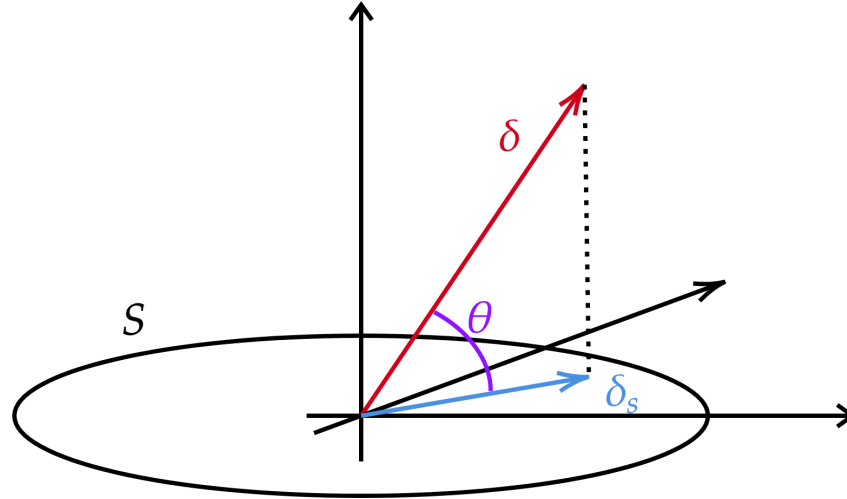
Figure 3.5.1 *Schematic representation of the geometric relation between the shift vector, $\delta \in D$, (here drawn as 3d), and the component, δ^S, of the shift vector which lies in the subspace, S (here drawn as 2d), containing the ellipse, E, defined by the theory covariance matrix. The angle θ between δ and δ^S is also shown; the dotted line shows the other side of the triangle, $\delta^{\mathrm{miss}} \in D/S$.*

Now that we have determined a suitable subspace, we can project the shift, δ_i, onto these eigenvectors:

$$\delta^\alpha = \sum_i \delta_i e_i^\alpha. \tag{3.5.3}$$

For a reasonable covariance matrix, δ^α should be a similar size to E in each dimension. We can then find the total component of the shift in S,

$$\delta_i^S = \sum_\alpha \delta^\alpha e_i^\alpha, \tag{3.5.4}$$

and the complementary component in the remaining space, D/S,

$$\delta_i^{miss} = \delta_i - \delta_i^S. \tag{3.5.5}$$

The validation will be a success if most of δ is in S, i.e. $|\delta_i^{miss}| \ll |\delta_i|$. Fig. 3.5.1 depicts the relationship between these objects. Note that δ, δ^S and δ^{miss} make up a right angled triangle, with some angle, θ, between δ and δ^S. For a successful validation, θ should be "small". Although there is no distinct cut-off of values here, note that given the dimension of D is 100 times larger than S, for a random covariance matrix we would expect θ very close to 90°.

3.5.2 Results of validation tests

We now apply the validation tests outlined thus far to the NLO theory covariance matrices for the various prescriptions.

A first check can be done by comparing the diagonal elements, σ_i where $\widehat{S}_{ii} = (\sigma_i)^2$, with the shifts, δ_i. This tests whether the per-point uncertainties encompass the NLO to NNLO shift. In Fig. 3.5.2, these comparisons are shown for all the prescriptions. Clearly, the shape of the MHOU is similar to the shape of the shift for the majority of the data and for each prescription. In fact, there is little difference between the prescriptions, except for the size of the uncertainty; 5-point is the least conservative and $\overline{5}$-point is the most conservative. From this we can see that the theory covmat is qualitatively descriptive of the observed higher order shift, in terms of size and correlation. We also note that the majority of the influence of prescription choice is on the off-diagonal elements. The range of scale variation looks to be broadly appropriate, therefore, with the caveat that for some points the MHOU is overestimated (see in particular NC DIS). This is a conservative treatment, however it could affect the weighting of data sets adversely.
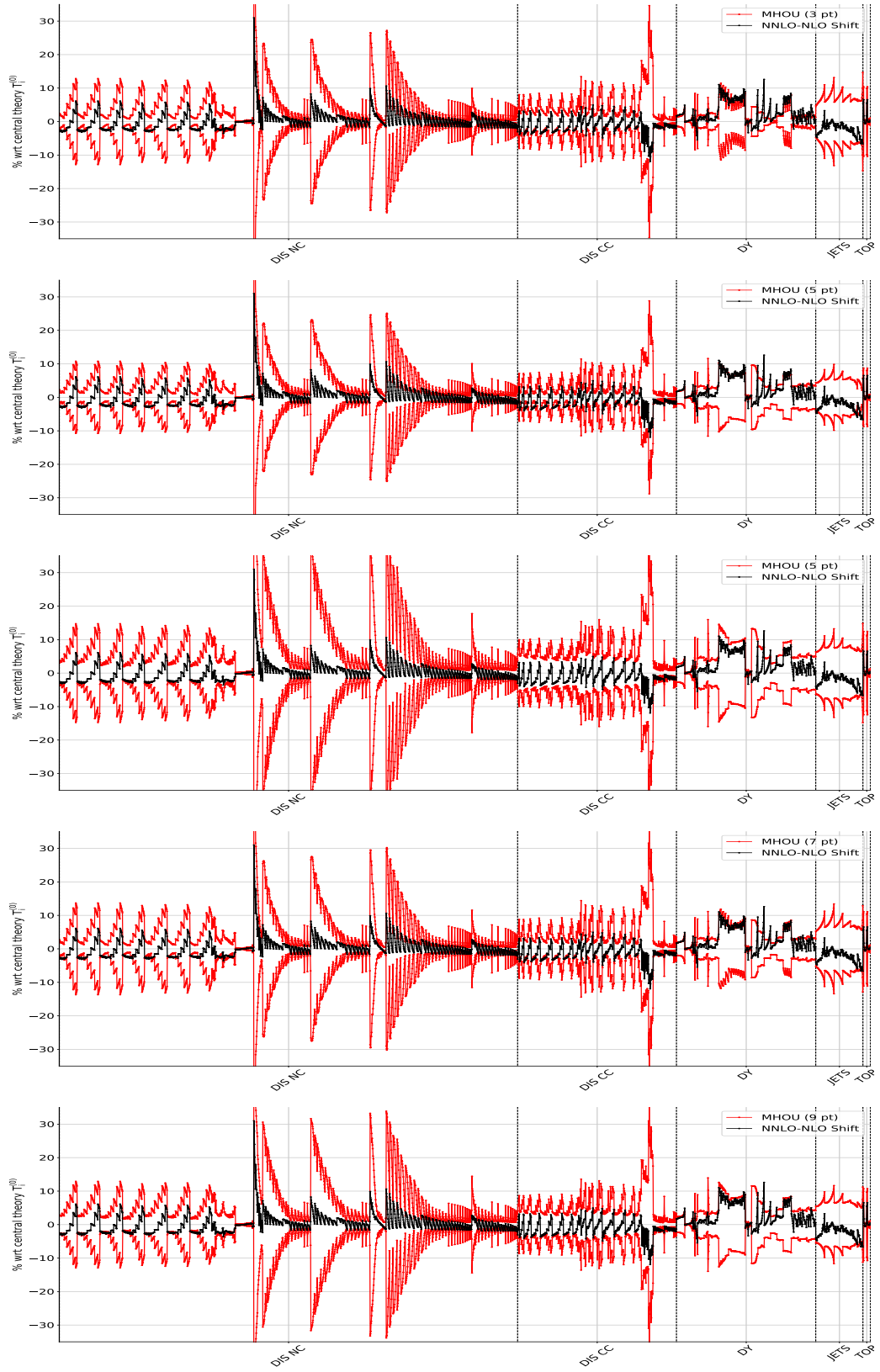
Figure 3.5.2 *The diagonal uncertainties σ_i (red) symmetrized about zero, compared to the shift δ_i for each datapoint (black), for the prescriptions. From top to bottom: 3-point, 5-point $\overline{5}$-point, 7-point and 9-point (bottom). All values are shown as percentage of the central theory prediction.*

To examine the efficacy of the correlations, we must turn to the methods discussed in the previous section. We first look at the angle, θ, between the shift and its component in the subspace, S, spanned by the theory covariance matrix.

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

Table 3.5.1 *The angle, θ, between the NNLO-NLO shift and its component, δ_i^S, lying within the subspace S (see Fig. 3.5.1) spanned by the theory covariance matrix for different prescriptions. The dimension of the subspace S in each case is also given.*

Tab. 3.5.1 displays these values for each of the prescriptions. All of these are reasonably small, given that $N_{sub} \ll N_{dat}$, but 9-point is the best, with $\theta = 26°$, due to the more comprehensive structure of scale variations. 3-point is unsurprisingly, the worst, suggesting that lack of correlation in the factorisation scale misses important correlations in the universal PDF evolution.

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

Table 3.5.2 *Same as Table 3.5.1 for each process of Table 3.4.1. The number of data points in each process is given directly below the name of the process.*

Tab. 3.5.2 shows the same analysis carried out individually for the various processes. The same hierarchy of prescriptions is evident within each process, with θ smallest for the processes with the least data (e.g. TOP). This is expected,

since larger collections of data span a greater kinematic range and include a richer structure, which is correspondingly harder to capture. DIS NC, the largest process, is the most poorly described. Note that the global value for θ is better than this, and so it is correlations within individual processes which are hardest to capture.

Having established what fraction of δ_i falls within S, we now go on to look at the complimentary component which falls outside, δ_i^{miss}. Fig. 3.5.3 shows this alongside δ_i. The missing element is non-zero for all processes, and with a shape following that of the shift. This suggests there could be a component of δ_i which is missing for most data points, pointing to a poor estimation of the MHOU in the PDFs, which is common to all data. A good candidate for this is that the factorisation scale variation, as mentioned before, is only approximate; it would be better to include a separate variation for each of the eigenvalues of evolution, e.g. to first order splitting up the singlet and non-singlet.
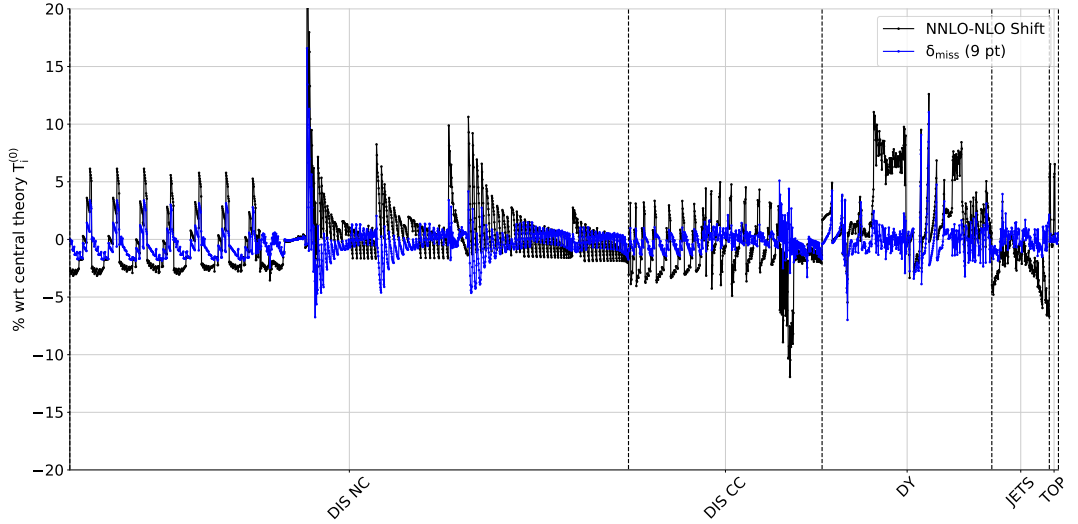
Figure 3.5.3   *The NNLO-NLO shift, δ_i (black), compared to its component, δ_i^{miss} (blue), which lies outside the subspace S, computed using the 9-point prescription.*

We know the component of the shift in the space of theory uncertainties, S, but we still need to see what fraction of this is encapsulated by the ellipse, E. To do this we look at the eigenvalues of the theory covariance matrix, $\lambda^\alpha = (s^\alpha)^2$. These are the lengths of the semi-axes of E, and there are as many as N_{sub}. We can compare them to δ^α, the projections of the shift onto each eigenvector.
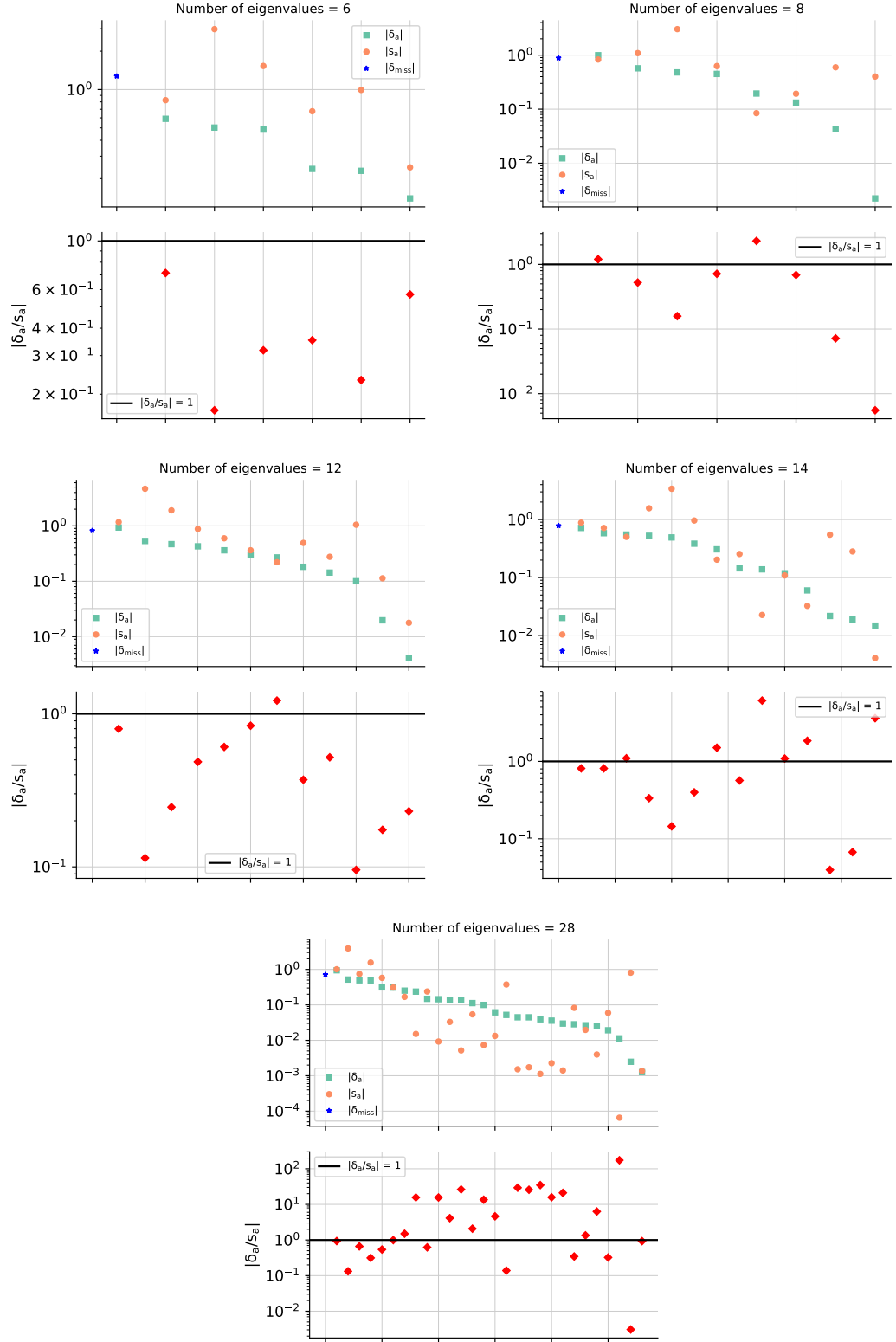
Figure 3.5.4 *The projection, δ^α, of the normalised shift vector, δ_i, along each eigenvector, e_i^α, of $\widehat{S}$, compared to the corresponding eigenvalue , s^α, ordered by the size of the projections (from largest to smallest). In each case results are shown as absolute (upper) and as ratios δ^α/s^α (lower). The magnitude of missing component, $|\delta_i^{\mathrm{miss}}|$ is also shown (blue star). Prescriptions: (top left) 3-point, 5-point $\overline{5}$-point, 7-point and 9-point (bottom right)*
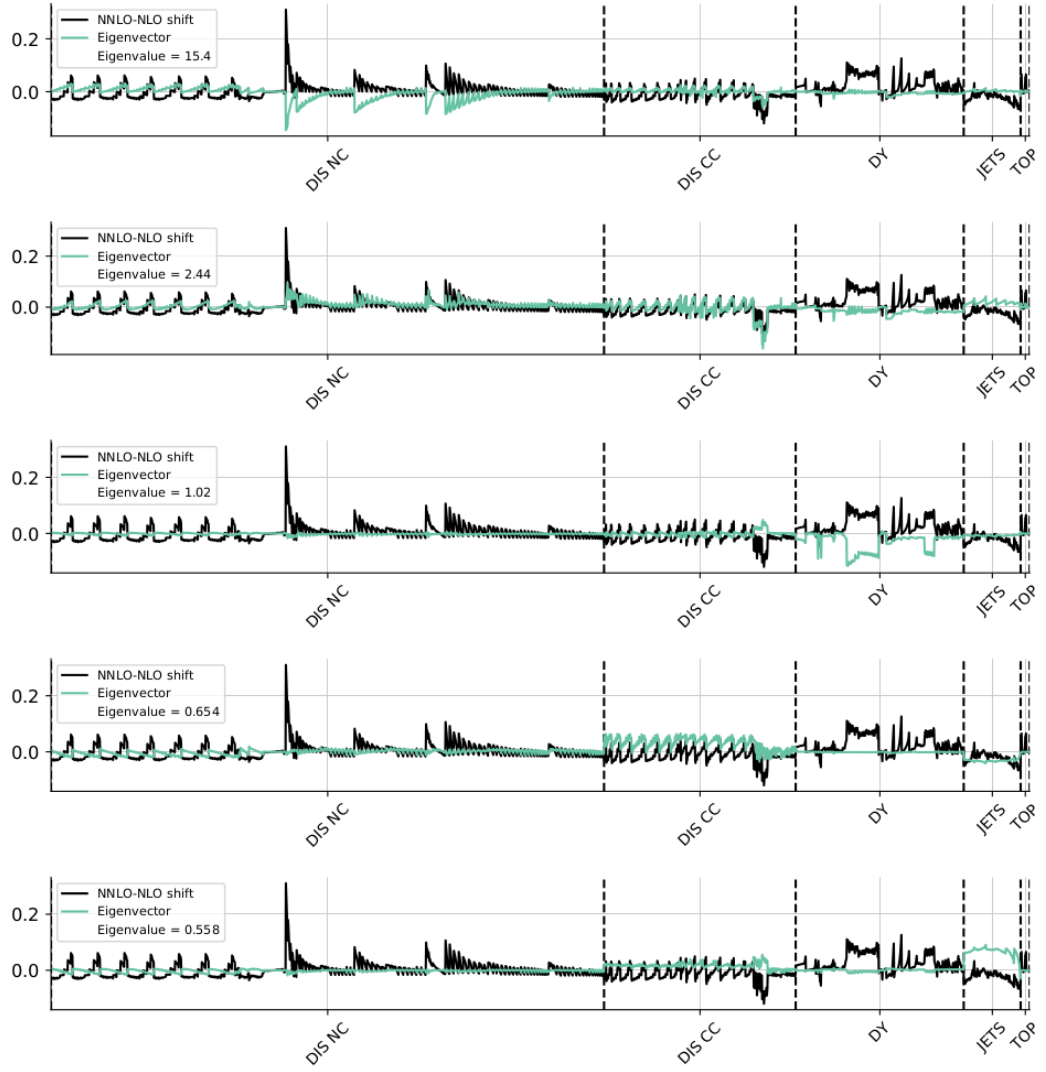
Figure 3.5.5 *The components, e_i^α (green), of the eigenvectors, corresponding to the five largest eigenvalues for the 9-point theory covariance matrix. The NNLO-NLO shift, δ_i (black), is shown for comparison.*

Fig. 3.5.4 shows these values for each prescription. All the prescriptions do a reasonable job, in that the largest eigenvalue is similar to the projection of the shift in that direction. Additionally, the size of the eigenvalues tends to decrease as the shift in that direction decreases. As expected, 3-point overestimates the uncertainty, with $\delta^\alpha < s^\alpha$ for all eigenvalues. This is also the case for most of 5- and $\bar{5}$-point. However, for the more complicated 7- and 9-point there is a better estimation for the larger eigenvalues, although a high scatter and poor prediction for the smaller ones; note that this is a result of only six scales being varied, so there is limited information available. Additionally, the missing component

70

is less than the largest projection for the symmetric prescriptions but is greater for the asymmetric prescriptions. This, once again, supports the adoption of a symmetric prescription.

Finally we investigate the components of the eigenvectors in the data space, D. The five largest of these for the 9-point are plotted in Fig. 3.5.5, alongside δ. Each of these can be identified with a component of the variation. From largest to smallest: 1 is largely DIS NC; 2 is a mixture of DIS NC and CC; 3 is DY; 4 is DIS CC; 5 is JETS. Unsurprisingly, the largest is dominated by the largest process, and so on, until the important TOP contribution appears at the (unshown) 9th eigenvector.

Overall, we have shown that all of the prescriptions capture most of the important features of the MHOs, and that 9-point does so the most accurately, given its more complex structure of scale variation. We therefore adopt 9-point as our chosen prescription, and proceed in the next section to include a 9-point theory covariance matrix in NLO PDF fits.

3.6 PDFs with missing higher order uncertainties

In this section we present the goal of this project: PDFs at NLO with the systematic inclusion of MHOUs. We compare NLO PDFs with and without MHOUs against the known NNLO PDFs, addressing the stability of the results to changes in prescription for the covariance matrix (9-pt vs 7-pt vs 3-pt). We break down the impact of MHOUs by including them separately in the Monte-Carlo sampling and the fitting.

To recap, there are only two places in which a theory covariance matrix changes the PDF determination:

Sampling

Recall from Chapter 1 that Monte Carlo pseudodata replicas $D_i^{(k)}$, $k = 1, \ldots N_{rep}$, are generated such that their covariance gives the total covariance matrix:

$$\lim_{N_{rep} \to \infty} \frac{1}{N_{rep}(N_{rep} - 1)} \sum_{k=1}^{N_{rep}} (D_i^{(k)} - \langle D_i \rangle)(D_j^{(k)} - \langle D_j \rangle) = C_{ij} + S_{ij}, \qquad (3.6.1)$$

where $\langle \dots \rangle$ denotes an average over replicas.

Fitting

During fitting the χ^2 is minimised, and this depends on S_{ij} through the equation:

$$\chi^2 = \frac{1}{N_{dat}} \sum_{i,j}^{N_{dat}} (D_i - T_i)(C + S)_{ij}^{-1}(D_j - T_j), \qquad (3.6.2)$$

where C_{ij} is the t0 covariance matrix to avoid d'Agostini bias [44, 45], as mentioned in Chapter 1.

We can assess the quality of the fit using this χ^2 estimator alongside the ϕ estimator, defined as

$$\phi = \sqrt{\langle \chi_{exp}^2[T_i]\rangle - \chi_{exp}^2[\langle T_i\rangle]}, \qquad (3.6.3)$$

where χ_{exp}^2 is evaluated using the (non-t0) experimental covariance matrix, C_{ij}. Following [56], this can be expressed as

$$\phi = \sqrt{\frac{1}{N_{dat}} \sum_{i,j=1}^{N_{dat}} (C_{ij} + S_{ij})^{-1} X_{ij}}, \qquad (3.6.4)$$

where $X_{ij} = \langle T_i T_j\rangle - \langle T_i\rangle\langle T_j\rangle$ is the covariance matrix of theoretical predictions. This is a measure of the consistency of the data; if they are consistent, then they should combine to reduce the uncertainty and so $\phi \ll 1$. The result should be a factor of r_ϕ greater than when S is not included, where

$$r_\phi = \sqrt{1 + \frac{1}{N_{dat}} \sum_{i,j=1}^{N_{dat}} C_{ij}^{-1} S_{ij}}. \qquad (3.6.5)$$

This means that if there were no other changes we would expect PDF uncertainties to increase by r_ϕ upon including MHOUs. The PDFs considered in this section are summarised in Table 3.6.2. NNLO PDFs with MHOUs are to be determined in a future work. The ϕ and χ^2 values for these fits are shown in Tables 3.6.1, 3.6.3, broken down by process type and, for χ^2 by dataset.

| Dataset | $N_{\rm dat}$ | $\chi^2/N_{\rm dat}$ in the NNPDF3.1 global fits | | | | | | NNLO |
| | | NLO | | | | | | |
		C	$C + S^{(9{\rm pt})}$	$C + S^{(7{\rm pt})}$	$C + S^{(3{\rm pt})}$	$C + S_{\rm fit}^{(9{\rm pt})}$	$C + S_{\rm samp}^{(9{\rm pt})}$	C
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 $\sigma_{\rm NC}^p$	886	1.086	1.044	1.046	1.079	1.044	1.190	1.098
HERA $\sigma_{\rm NC}^c$	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 $\sigma_{\rm CC}^p$	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 $Z \to ee$ 2 fb	17	1.305	1.303	1.305	1.313	1.334	1.250	1.203
LHCb $W, Z \to \mu$ 7 TeV	29	1.262	1.106	1.267	1.261	1.134	1.207	1.536
LHCb $W, Z \to \mu$ 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 $W \to e\nu$ asy	8	1.176	1.066	1.055	1.083	1.029	1.200	2.558
D0 $W \to \mu\nu$ asy	9	1.400	1.450	1.372	1.361	1.439	1.395	1.374
DY	**484**	**1.486**	**1.447**	**1.485**	**1.483**	**1.461**	**1.434**	**1.231**
ATLAS jets 2011 7 TeV	31	1.069	1.019	1.065	1.079	1.026	1.031	1.076
CMS jets 7 TeV 2011	133	0.869	0.786	0.790	0.830	0.795	0.883	0.921
JETS	**164**	**0.907**	**0.839**	**0.858**	**0.901**	**0.848**	**0.911**	**0.950**
ATLAS $\sigma_{tt}^{\rm top}$	3	2.577	0.787	0.853	0.982	0.770	2.442	0.903
ATLAS $t\bar{t}$ rap	10	1.258	0.955	0.867	0.910	0.935	1.355	1.424
CMS $\sigma_{tt}^{\rm top}$	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 3.6.1   *Breakdown of χ^2 values by dataset and process.*

Label	Order	Cov. Mat.	Comments
NNPDF31_nlo_as_0118_kF_1_kR_1	NLO	C	baseline Global NLO
NNPDF31_nlo_as_0118_scalecov_9pt	NLO	$C + S^{(9\text{pt})}$	
NNPDF31_nlo_as_0118_scalecov_7pt	NLO	$C + S^{(7\text{pt})}$	
NNPDF31_nlo_as_0118_scalecov_3pt	NLO	$C + S^{(3\text{pt})}$	
NNPDF31_nlo_as_0118_scalecov_9pt_fit	NLO	$C + S^{(9\text{pt})}$	S only in χ^2 definition
NNPDF31_nlo_as_0118_scalecov_9pt_sampl	NLO	$C + S^{(9\text{pt})}$	S only in sampling
NNPDF31_nnlo_as_0118_kF_1_kR_1	NNLO	C	baseline Global NNLO

Table 3.6.2 *Summary of the PDFs discussed in this section. The perturbative order and treatment of uncertainties for each are indicated.*

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

Table 3.6.3 *ϕ for fits with S compared to without. Results are shown for the 9-, 7-, and 3- point prescriptions. For 9-point the impact is broken down by inclusion of S in either fitting (S_{fit}) or sampling (S_{sampl}) only. The final column is a comparison to the NNLO C only fit.*

3.6.1 Fit quality

We can see that including MHOUs causes the χ^2 to decrease, both globally and for many individual datasets. This is an indication that the fit quality has improved, and is to be expected as we have added additional uncertainties. This varies slightly by prescription, with 9-point showing the greatest improvement, down 3%, which is comparable to the NNLO value. This suggests that the theory uncertainty is doing a reasonable job of accounting for the NNLO correction. Individual datasets follow a similar trend, with the caveat that at NNLO some datasets acquire additional uncertainties, which confuses the comparison (e.g. CMS Z p_T). ϕ increases by 30% for 9-point, less than the expected value of r_ϕ=1.69, suggesting a resolution of tension between datasets. This is relatively

stable across prescription choice, although a little higher for 9-point.

Let us now turn our attention to the fits were S is only included in the sampling or the fitting. For sampling only, we expect the uncertainty to increase, leading to broad fluctuations in the data replicas, but because this is not accounted for in the χ^2 the fit quality should be worse. This is indeed what we see, with the χ^2 increasing and the ϕ tripling in value. For fitting only, the MHOU should mostly affect the weighting between datasets, and therefore the central value of the PDFs. We can see that the χ^2 is close to that of the fit with S in fitting *and* sampling, and the ϕ increases less. But it is clear that it is the inclusion of S in the fitting which drives the best fit.

3.6.2 Form of PDFs

We now consider the form of the PDFs themselves. Fig 3.6.1 shows the NLO PDFs before and after adding (9-point) MHOUs. These are compared to the central value of the NNLO C-only PDF. Results are shown at the parametrisation scale, $Q = 1.6$ GeV, alongside a higher scale, $Q = 10$ GeV, for g, Σ, $\bar{d}$ and s. In the data region, the PDF uncertainties increase a small amount, but the biggest difference is in the significant shift of central value, up to 1-σ. This corresponds with what we saw earlier in the ϕ values, which increase only a small amount, suggesting that tension in the data region has been partially resolved. Outwith the data region, however, the uncertainties increase more, particularly in the poorly-understood extrapolation region at very low x.

We can see that the central values are compatible with the NNLO ones within uncertainties, but more strikingly it is especially evident at the parametrisation scale that adding MHOUs tends to shift the central value towards the NNLO one. Although this shift is scale dependent, given that NLO and NNLO PDFs evolve differently, this agreement is still somewhat apparent at the higher scale; the strange content shifts up and the gluon shifts down. Comparing the different prescriptions (Fig. 3.6.2), it is apparent that the results are fairly stable, with the asymmetric prescriptions (3-point and 7-point) showing closer similarity to one another but with 3-point having slightly larger uncertainties.
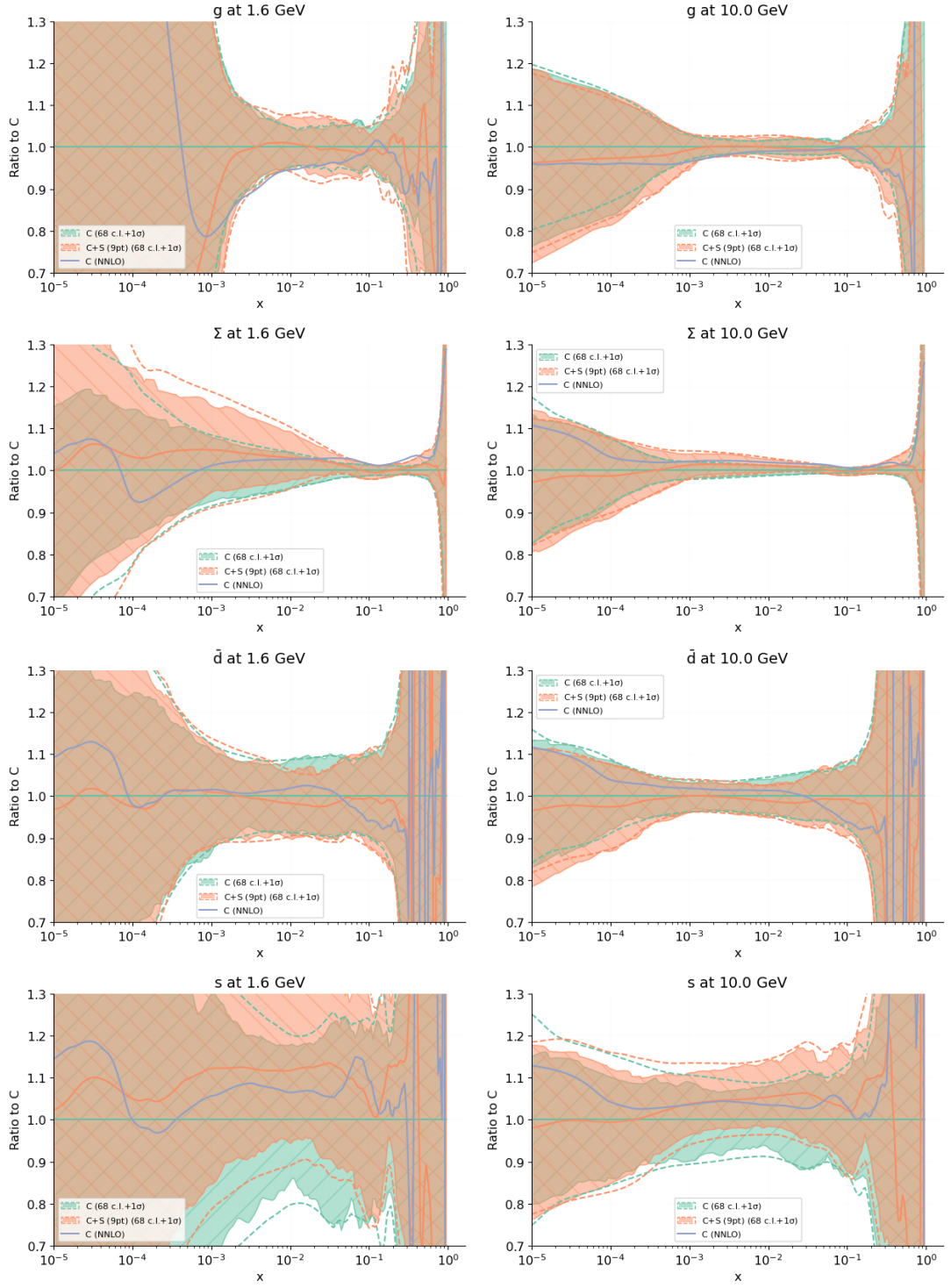
Figure 3.6.1 *NLO PDFs based on C (green) and $C+S^{(9\mathrm{pt})}$ (orange) normalised to the former, alongside the central value of the NNLO fit based on C (blue line). Results are shown at $Q = 1.6$ GeV (left column) and $Q = 10$ GeV (right column). From top to bottom: gluon; total quark singlet; anti-down quark; strange quark.*
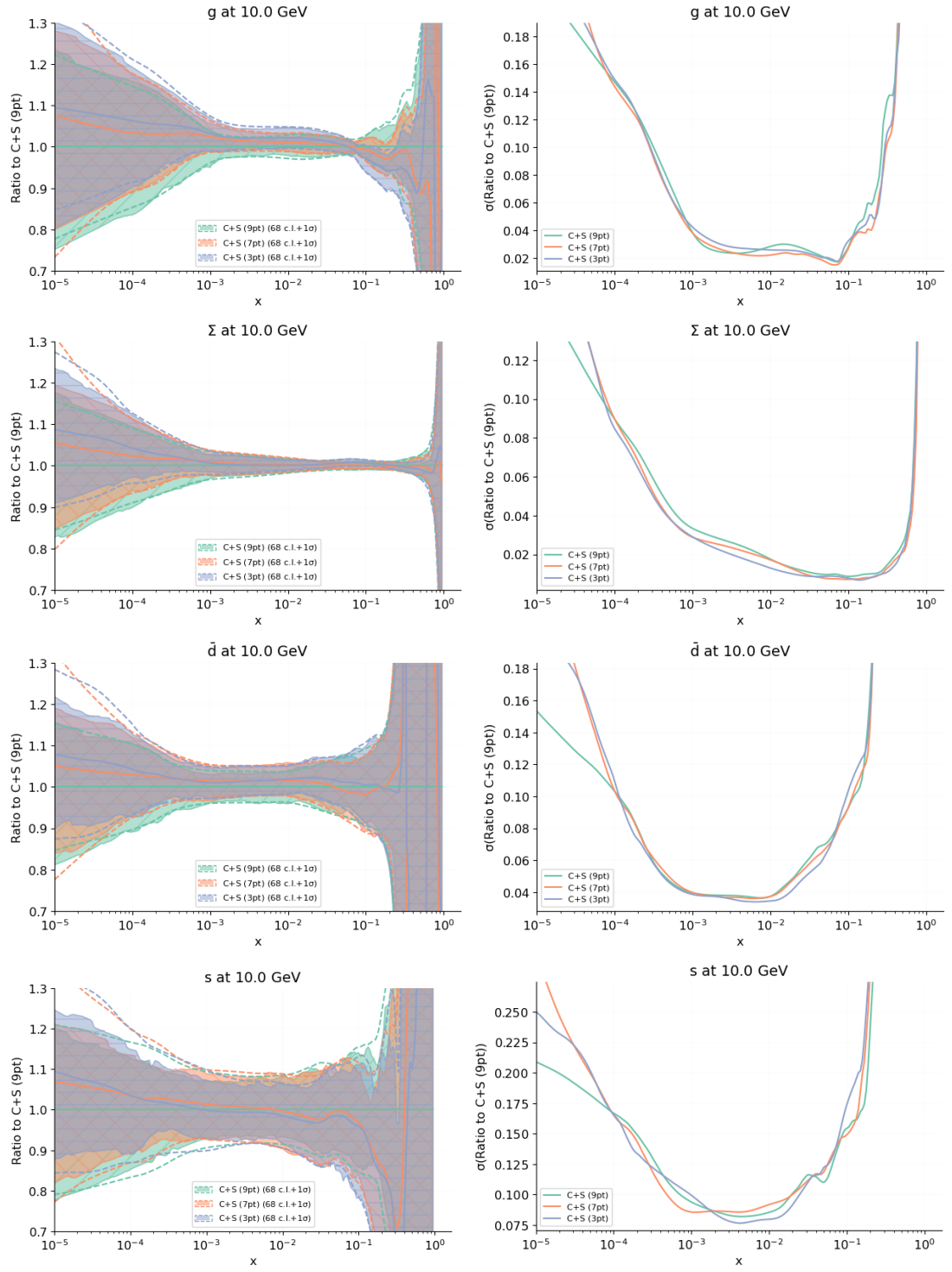
Figure 3.6.2 *Same as Fig. 3.6.1 but comparing the 3- (blue), 7- (orange), and 9-point (green) prescriptions, normalised to 9-point. The right hand panel shows the relative PDF uncertainties for clarity.*
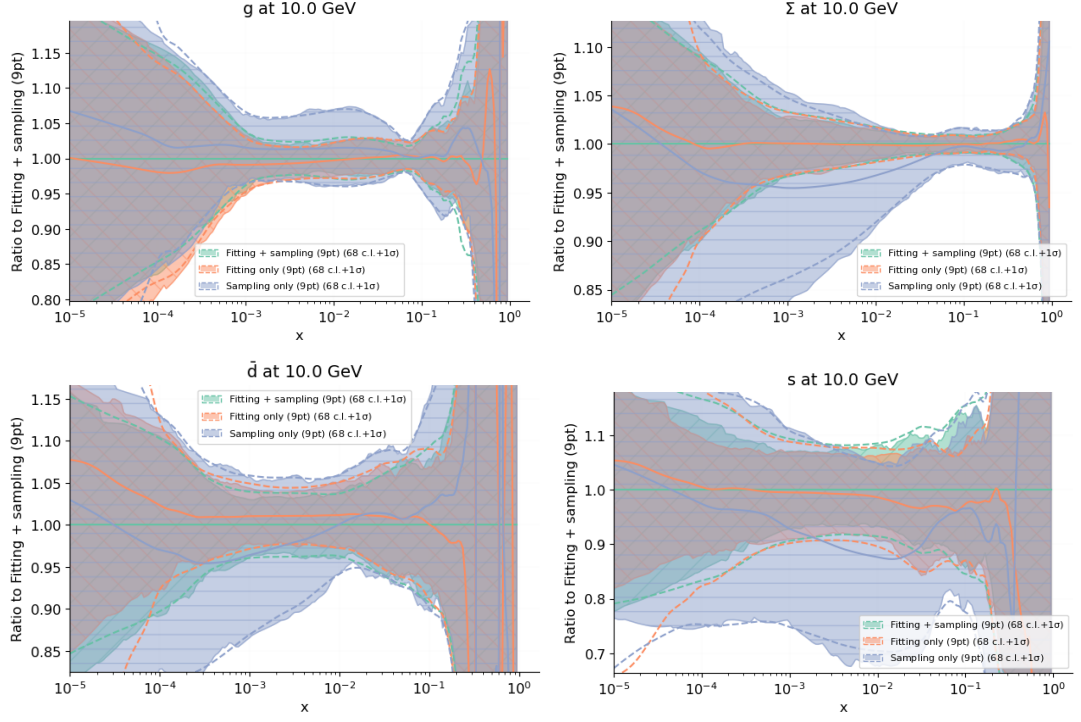
Figure 3.6.3 *Same as Fig. 3.6.1, now comparing $C + S^{(9\text{pt})}$ fit (green) with those in which the theory covariance matrix S is included either in the χ^2 definition (orange) or in the generation of Monte Carlo replicas (blue), but not in both.*

Finally we compare using S in only fitting or sampling (Fig 3.6.3). For sampling only, the PDF uncertainty increases dramatically, with poor fit quality, especially in the quark distributions. For fitting only, the central value is affected due to the change in relative weights of the datasets, such that it is similar to that for fitting + sampling. The uncertainty, however, shows only a very small change in the data region. This all arises because the inclusion of MHOUs in data generation cause the pseudodata broadness to increase dramatically, which is in turn balanced by a relaxation of tensions due to the inclusion of MHOUs in fitting. This has the effect of a sizeable shift in central values with only a small increase in uncertainties.

3.7   Summary

We have presented the first PDFs with MHOUs included in their uncertainties, paving the way for the routine inclusion of theory uncertainties in future PDFs. This chapter has been primarily focussed on the formalism necessary

for including MHOUs, and on the validation of the method. We find that scale variation appears to work well for this purpose, and that our prescriptions for combining scales are more successful and free from instabilities than the established "envelope" techniques. It is also clear that there is scope for more complex scale variation techniques, particularly for the factorisation scale, which as a first step could be split into singlet and non-singlet variation.

The PDFs detailed in this chapter, along with PDFs with varied scales, are available in LHAPDF format [54] from the NNPDF website:

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

It now remains to investigate the impact of including MHOUs in PDFs on phenomenology, that is, in using them to compute predictions for cross-sections. For this some thorough analysis is required to address the potential "double counting" of theory uncertainties, where they are included both in the PDF and in the hard cross-section. We will address this in considerable detail in Chapter 5. First, however, we will look at one other important form of theory uncertainties: those due to nuclear effects.

Chapter 4

Nuclear Uncertainties

The theory covariance formalism developed in Chapter 2 can be applied to any source of theory uncertainty in PDFs. One of the most important of these is nuclear uncertainties. A wide range of data is needed to pin down the form of PDFs, including that where the proton is not in a free state. More precisely, this encompasses DIS and DY fixed target measurements involving deuteron and heavy nuclear targets. In these cases the proton's interaction is altered due to the surrounding nuclear environment, and this difference propagates through to the fitted PDFs, leading to an unwanted shift in their central values and uncertainties. We cannot simply discard these data, as they play a crucial role in the strangeness content of the proton and also the light flavour separation at high x, a region important for searches for physics beyond the Standard Model. Instead, we must determine corrections to the PDF central value and additional uncertainties to account for the use of nuclear data.

Given their importance, there have been wide-ranging studies of deuteron and heavy nuclear corrections: deuteron corrections have been included in previous PDF determinations via nuclear smearing functions [65, 122–125] based on models of the deuteron wavefunction [126–130]; heavy nuclear corrections have been included following a selection of nuclear models [123, 131, 132] or fitting the data [124]. Using such models, however, can introduce a bias that is difficult to quantify precisely. In the past, NNPDF has opted to ignore nuclear effects on the assumption that they are small [9, 56, 65], however this is another source of uncertainty that is becoming increasingly important as PDF precision increases. Furthermore it is thought that the shape of PDFs can be affected, especially

at high x [122], and this was evidenced in previous NNPDF fits with deuteron corrections following Eq. (8) of [123], with parameter values from [133]; however, an increase in χ^2 here suggested that the nuclear uncertainty was not effectively determined.

In this chapter we show how to account for nuclear effects, both deuteron and heavy nuclear, in proton PDF fits. We do this using a theory covariance matrix of nuclear uncertainties, and propose two alternatives for their inclusion: one is to simply apply a nuclear uncertainty, effectively deweighting the affected datasets in the PDF fit proportionally (just like we did for MHOUs in Chapter 3); the other is to shift the PDF central values by applying a nuclear correction, including smaller nuclear uncertainties as a result. If the shift is estimated accurately, then for an uncertainty smaller than the shift the second method gives a more precise outcome.

We can determine nuclear corrections by comparing the theory predictions for nuclear observables using proton PDFs with those using the correct nuclear PDF (nPDF). This shift can be identified with Eqn. 2.1.6 in Chapter 2, i.e. quantifying the size of nuclear correction for that data point. The collective shifts can then be used to construct a theory covariance matrix based on Eqn. 2.1.13. In carrying out this work we looked first at heavy nuclear corrections (for Cu, Fe and Pb) and then at deuteron corrections, addressing them separately because deuterons, being only a proton and a neutron, are distinct from a heavy nuclear environment such as ^{56}Fe, with 26 protons and 30 neutrons bound together.

For the heavy nuclear PDFs we initially used [134] a combination of three available nPDF sets (DSSZ [135], nCTEQ15 [136], and EPPS16 [137]), but NNPDF subsequently released its own global nPDFs, nNNPDF2.0 [138], which is what we will consider in this Chapter. Given the enhanced difficulty of nPDF determination, all of these nPDF sets are only available at NLO. For the deuteron PDFs we developed a self-consistent iterative procedure to determine deuteron PDFs at NNLO within the NNPDF formalism, and used the output of this to determine deuteron corrections [139]. These deuteron PDFs have the advantage over those from nNNPDF2.0 that they are NNLO, but are based on less data so have larger uncertainties. This should at worst lead to a conservative uncertainty estimation, but we will discuss the comparison in Section 4.5.

This chapter is organised as follows. First we review the nuclear data in proton PDF fits (Sec. 4.1). Then we consider heavy nuclear uncertainties including the

resulting covariance matrix (Sec. 4.2). We then look at deuteron uncertainties in the same way (Sec. 4.3), before including combined nuclear uncertainties in NNPDF4.0 proton PDF fits in Sec 4.4. We summarise the results in Sec. 4.5.

4.1 Nuclear data in PDFs

We consider the NNPDF4.0 NNLO dataset which consists of ~ 4000 data points, of which $\sim 10\%$ are deuteron data and $\sim 20\%$ are heavy nuclear data. The table below summarises the datasets which make up the total nuclear data, giving the name of dataset, the observable it corresponds to, and the nuclear target involved.

Nuclear data			
Dataset	**Process**	N_{dat}	**Target**
DYE605 [140]	DY	85	$^{64}_{32}$Cu
NuTeV [141]	DIS CC	76	$^{56}_{26}$Fe
CHORUS [93]	DIS CC	832	$^{208}_{82}$Pb
SLAC [88]	DIS NC	67	^{2}H
BCDMS [142]	DIS NC	581	^{2}H
NMC [86]	DIS CC	204	^{2}H and p
DYE866/NuSea [143]	DY	15	^{2}H and p
DYE906/SeaQuest [144]	DY	6	^{2}H and p

Table 4.1.1   *The nuclear data in NNPDF4.0. The process (Deep inelastic scattering (DIS) charged current (CC), neutral current (NC) and Drell-Yan (DY) is displayed for each dataset, alongside the number of data (N_{dat}) and the target.*

4.2 Heavy nuclear uncertainties

We can include heavy nuclear uncertainties using the covariance matrix methodology previously developed in NNPDF [145]. To construct the covariance matrix, we can look directly at the source of uncertainty in our use of nuclear data: we currently calculate nuclear observables, T_i^N (where i labels the data point), using a proton PDF, f_p. Instead we should be using the corresponding nPDF, f_N. This means that each contribution to the covariance matrix can be determined

$$\Delta_i^{(k)} = T_i^N[f_N^{(k)}] - T_i^N[f_p], \tag{4.2.1}$$

82

where k indexes the nPDF replicas. Including N_{rep} contributions, one for each replica, means that the uncertainty in the nPDF is automatically incorporated into the nuclear uncertainty. A covariance matrix can then be constructed as

$$S_{ij} = \frac{1}{N_{rep}} \sum_k^{N_{rep}} \Delta_i^{(k)} \Delta_j^{(k)}. \tag{4.2.2}$$

We call this the "deweighted" approach, because this theory covariance matrix will deweight the nuclear datasets in the fit.

A more ambitious approach is to also try and correct the value of the nuclear observable we use, so that it is based on the nPDF rather than the proton one. This can be done by applying a shift,

$$\delta T_i^N = T_i^N[f_N] - T_i^N[f_p], \tag{4.2.3}$$

to the nuclear observables. In this case we must amend the contributions to the covariance matrix so that they are relative to the new central value

$$\Delta_i^{(k)} = T_i^N[f_N^{(k)}] - T_i^N[f_N]. \tag{4.2.4}$$

We call this the "shifted" approach.

We can then include the nuclear covariance matrix in a normal proton PDF fit, allowing uncertainties due to nuclear data to be automatically accounted for. Note that although we use nuclear data to determine uncertainties, we are not double counting the nuclear data; we use them once to determine the nuclear uncertainty and once to do a global fit to find the proton PDFs. Adding uncertainties actually makes the nuclear data count *less*.

In the above equations the nuclear observables, T_i^N, are calculated from the proton observables, T_i, by taking into account the non-isoscalarity of the target, i.e. by combining the proton and neutron observables in accordance with the atomic number, Z and mass number, A. Explicitly,

$$\begin{aligned} T_i^N[f_p] &= \frac{1}{A}\left(Z T_i[f_p] + (A - Z) T_i[f_n] \right), \\ T_i^N[f_N] &= \frac{1}{A}\left(Z T_i[f_{p/N}] + (A - Z) T_i[f_{n/N}] \right). \end{aligned} \tag{4.2.5}$$

The first line is what is done in standard NNPDF fits, and the second line is

the extension to the nPDF case. Here $f_{p/N}$ is the PDF for the proton bound in a nucleus, N, and $f_{n/N}$ is the same for the neutron. We assume that the two are related by swapping u and d quarks. We obtain these PDFs directly from nNNPDF2.0, but they relate to f_N via

$$f_N = \frac{1}{A}\left(Z f_{p/N} + (A - Z) f_{n/N}\right).$$

(4.2.6)
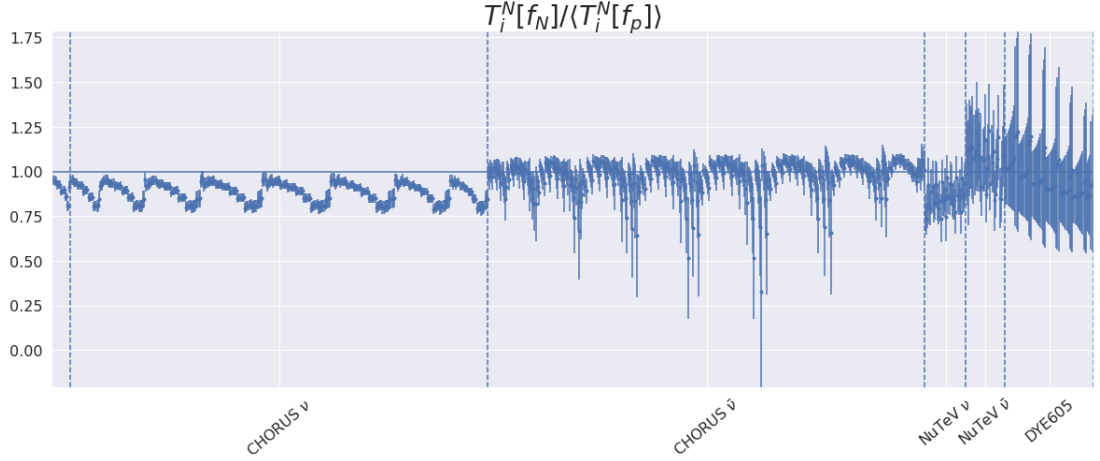
Figure 4.2.1 *Ratio between the nuclear observables computed with nPDFs, $T_i^N[f_N]$, and the central prediction computed with proton PDFs, $\langle T_i^N[f_p]\rangle$. The error is the standard deviation of the distribution of $T_i^N[f_N]$ replicas. Data are organised in bins of increasing (x, Q^2) within each dataset.*

Before proceeding to the covariance matrix itself, we can first investigate the change to the nuclear observables that arises from using nPDFs rather than proton ones. Fig. 4.2.1 shows the ratio of the observables calculated with nuclear PDFs to those with proton PDFs, for the heavy nuclear datasets. We can see in all datasets that there is a kinematic dependence, although this is especially evident in CHORUS. This is a result of the kinematic dependence of the ratio of proton and nuclear PDFs, which fits with the downwards turn at high x expected from nuclear shadowing models. CHORUS ν and NuTeV ν data in particular show a systematic shift downwards which is not comfortably within errors. This suggests that applying a shift as well as an uncertainty could be a sensible strategy.

4.2.1　The heavy nuclear covariance matrix

We now turn to the covariance matrix for heavy nuclear uncertainties. Fig. 4.2.2 shows the square roots of the diagonal elements of this covariance matrix, which are equivalent to the % per-point uncertainties. For the deweighted case, it
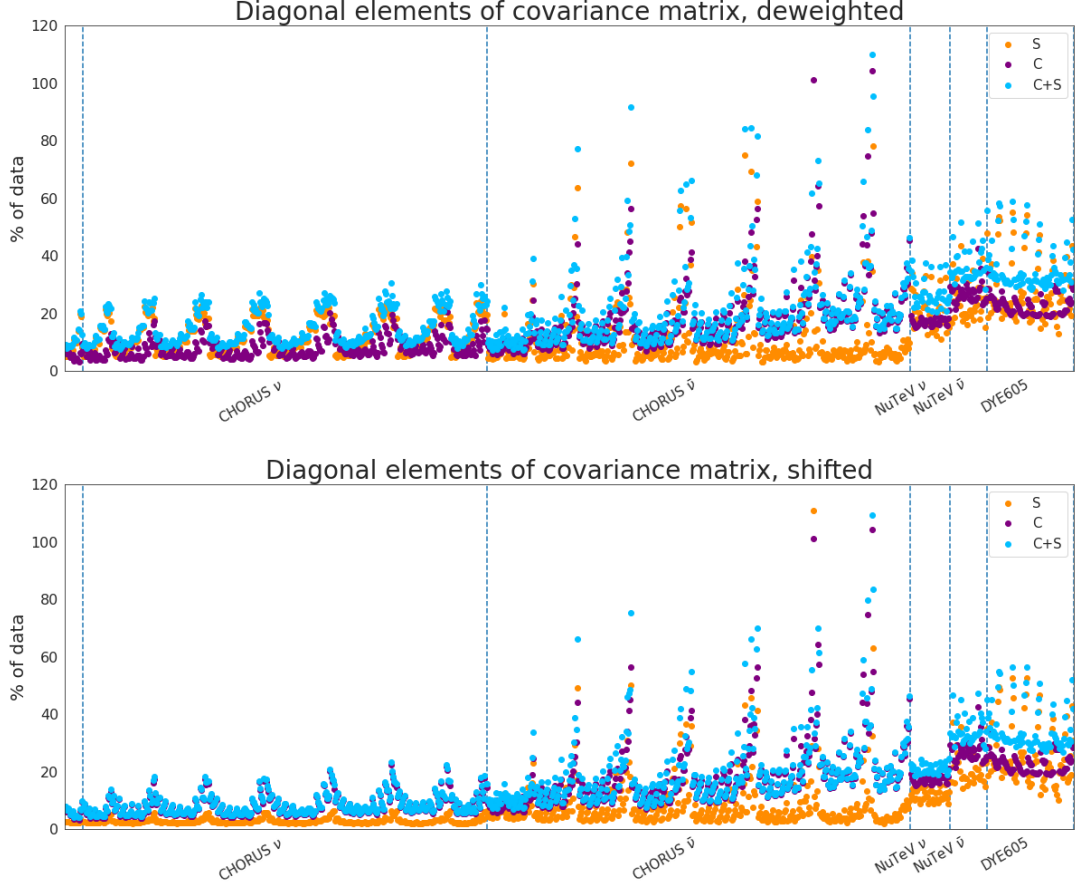
Figure 4.2.2　*Square root of diagonal elements of covariance matrices for C (purple), S (orange) and C + S (blue). All values are displayed as a % of data. Top: deweighted; bottom: shifted.*

is clear that the heavy nuclear uncertainties are comparable to the experimental uncertainties and are larger in most regions other than CHORUS $\bar{\nu}$. This suggests that all datasets apart from that will be significantly deweighted in the fit. We see that the plot has many features in common with Fig. 4.2.1, in particular the kinematic pattern, and this makes sense as the covariance matrix is composed using the difference in observables when using nPDFs versus proton ones. For the shifted case, we see a marked decrease in the diagonal of S, such that nuclear per-point uncertainties seem no longer significant for CHORUS.
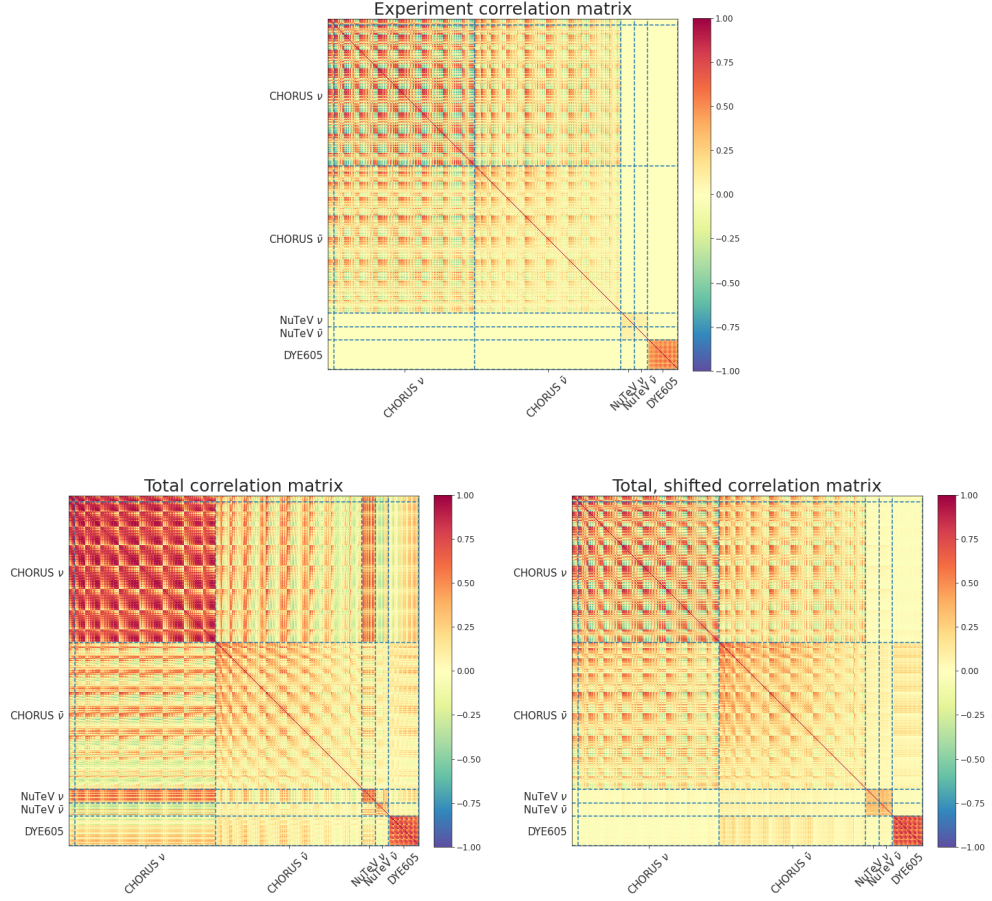
Figure 4.2.3 *Correlation matrices for heavy nuclear data. The experiment correlation matrix, C, is shown above the total correlation matrices, $C + S$, for both the deweighted (left) and the shifted (right) case.*

Fig. 4.2.3 investigates the pattern of correlations, displaying correlation matrices as defined in Sec. 3.4 of Chapter 3. Note that in [134] we considered covariance matrices experiment by experiment (a conservative approach), but here we compute the full heavy nuclear covariance matrix, including correlations between experiments. We display the correlation matrices for both the deweighted and shifted theory covariance matrices. Adding the theory covariance matrix introduces correlations on the off-block diagonals in both cases, but these are particularly strong in the deweighted case. CHORUS ν shows especially larger correlations upon introducing the deweighted S, but these are reduced almost to experiment level when the shifted S is used instead. This is a clear consequence of the systematic shift in Fig. 4.2.1, which is larger than the uncertainty from the nPDF. Once again, this is evidence that using the shifted formulism could be an appropriate choice.

4.3 Deuteron uncertainties

We now turn to uncertainties from deuteron data. The logic is the same as for heavy nuclear uncertainties, except we have a lot more deuteron data than heavy nuclear data from a particular element. This allows us to fit our own deuteron PDFs within the NNPDF methodology, resulting in NNLO PDFs to calculate uncertainties rather than NLO ones.

The whole procedure is outlined in Fig. 4.3.1. We split the global data into "proton" and "deuteron" data, where the proton data in fact include the heavy nuclear data considered in the previous section; this allows us to focus purely on deuteron uncertainties. The deuteron data are a combination of "pure" deuteron data, coming from deuteron-only targets (SLAC and BCDMS), and "mixed" deuteron data, which are ratios so depend also on proton target data (NMC, DYE866/NuSea and DYE906/SeaQuest). We can denote the pure data as $T_i^d[f_d]$ and the mixed data as $T_i^d[f_d, f_p]$, indicating the additional dependence of the latter on the proton PDF.

In normal proton PDF fits any deuteron observable is calculated using the isoscalar PDF,

$$f_s = \frac{1}{2}(f_p + f_n), \tag{4.3.1}$$

where f_n is the neutron PDF, obtained under the assumption of isospin invariance (by swapping u and d quarks in f_p). The procedure is as follows:

1. The proton data are used to fit pure proton PDFs (uncontaminated by deuteron data), $\{f_p^{(k)} : k = 1, \ldots, N_{rep}\}$ with central value $f_p^0 \equiv f_p$.

2. We cannot fit the pure deuteron PDFs without additional input because of the mixed ratio data ($T_i^d[f_d, f_p]$); these data require a proton PDF as input. We can use f_p^0 from the proton-only fit here, but since this is only the central value we must include a proton covariance matrix to account for the uncertainty due to the proton PDF. This is composed

$$\begin{aligned} S_{ij}^p &= \langle \Delta_i^{p,\,(k)}, \Delta_j^{p,\,(k)} \rangle \\ \Delta_i^{p,\,(k)} &= T_i^d[f_d^{(0)}, f_p^{(k)}] - T_i^d[f_d^{(0)}, f_p^{(0)}], \end{aligned} \tag{4.3.2}$$

where here i runs over the deuteron ratio data only. This encapsulates the correlations between the ratio datasets due to their common dependence on

87
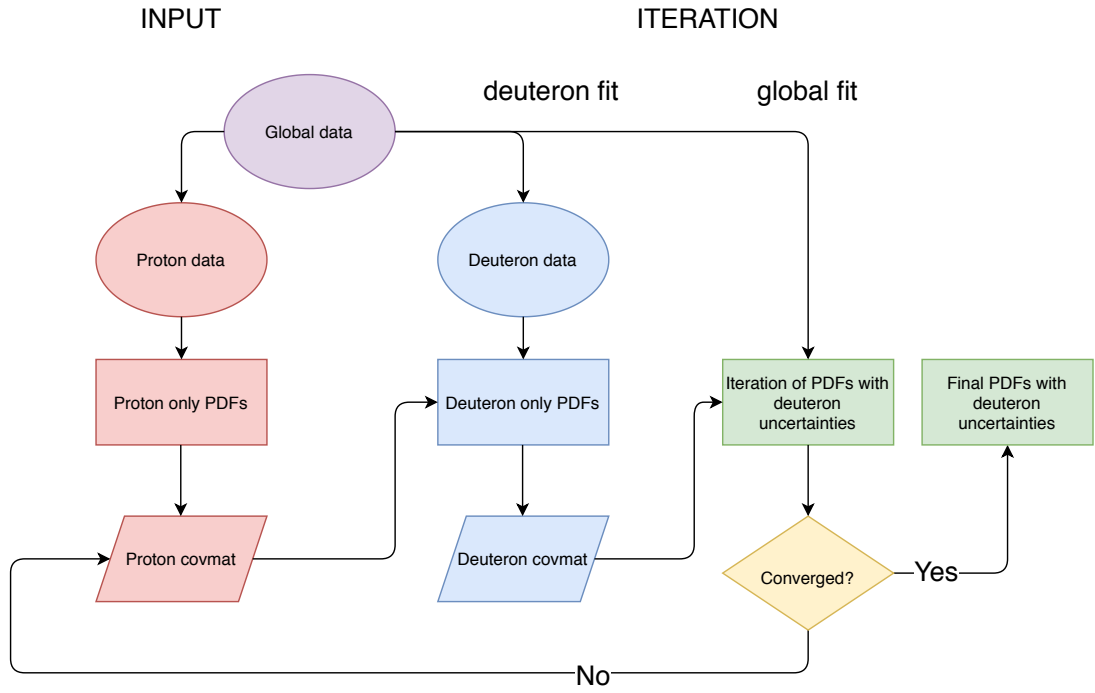
Figure 4.3.1 *Outline of the iterative procedure used to determine proton PDFs with deuteron uncertainties. The data are split into proton data and deuteron data. The proton data are used to find proton-only PDFs which are needed to fit deuteron observables in deuteron-only PDFs. These are used to construct a deuteron covariance matrix which is used in a global proton PDF fit. The whole process is iterated to consistency.*

the proton PDF.

3. Fit the deuteron PDFs using $C_{ij} + S_{ij}^p$ for the deuteron ratio data. However, in practice S^p depends on $f_d^{(0)}$ and we don't know this yet (that's what we're trying to find!). To first approximation we can replace $f_d^{(0)} \to f_s^{(0)}$, using the isoscalar PDF from Eqn. 4.3.1. This is a reasonable exchange, given that S_{ij}^p is a measure of uncertainty, and acts only to deweight data in the fit; using f_s should have only a very small effect. In principle, we could then iterate this to consistency, using the output f_d to determine a new S_{ij}^p and perform a new deuteron fit. However we are less interested in determining the deuteron PDFs themselves, more in their application in creating a covariance matrix for fitting proton PDFs. Any (already small) effect from using f_s in S^p will become smaller when finding f_d and then smaller again when finding f_p, where the only influence is via a covariance matrix which depends on f_d.

4. Use f_d to generate a deuteron covariance matrix:

$$S_{ij}^d = \langle \Delta_i^{d,\ (k)}, \Delta_j^{d,\ (k)} \rangle$$

$$\Delta_i^{d,\ (k)} = \begin{cases} T_i^d[f_d^{(k)}] - T_i^d[f_s^{(0)}] & i \in \text{pure} \\ T_i^d[f_d^{(k)}, f_p^{(0)}] - T_i^d[f_s^{(0)}, f_p^{(0)}] & i \in \text{mixed}, \end{cases} \qquad (4.3.3)$$

or, for the shifted case,

$$\Delta_i^{d,\ (k)} = \begin{cases} T_i^d[f_d^{(k)}] - T_i^d[f_d^{(0)}] & i \in \text{pure} \\ T_i^d[f_d^{(k)}, f_p^{(0)}] - T_i^d[f_d^{(0)}, f_p^{(0)}] & i \in \text{mixed}, \end{cases} \qquad (4.3.4)$$

$$\delta T_i^d = \begin{cases} T_i^d[f_d^{(0)}] - T_i^d[f_s^{(0)}] & i \in \text{pure} \\ T_i^d[f_d^{(0)}, f_p^{(0)}] - T_i^d[f_s^{(0)}, f_p^{(0)}] & i \in \text{mixed}. \end{cases} \qquad (4.3.5)$$

S^d incorporates correlations between the deuteron data due to their common dependence on the deuteron PDF and, for the ratio data, their consequential dependence on proton PDFs.

5. Perform a global proton PDF fit incorporating S^d for the deuteron data. These are PDFs with deuteron uncertainties included.

6. Use the resulting proton PDFs in place of the proton-only PDFs in Step 2, thus iterating the procedure. We expect this to converge rapidly for a few reasons: first, the influence of deuteron data in a proton fit is small; second, a small change in the proton PDF makes little difference to the deuteron uncertainty; third, the effect of deuteron uncertainties on the weight of data in the fit is anticipated to be small.

Note that once again in this procedure we are not double counting the deuteron data; we use them once to determine the deuteron uncertainty and once to do a global fit to find the proton PDFs. Adding uncertainties actually makes the deuteron data count *less*.

The fits performed are summarised in Tab. 4.3.1. The baseline fit, global-base, is the baseline fit for NNPDF4.0 [37] without nuclear uncertainties included (the uncertainties we determine in this Chapter will be included in the final NNPDF4.0 release). We determine a proton-only fit in what we term "Iteration 0", and then perform two iterations of determining the deuteron and global proton fits, termed

Iteration	Dataset	Fit ID	Description
Baseline	Proton and Deuteron	global-base	NNPDF4.0 fit without nuclear uncertainties
Iteration 0	Proton	proton-ite0	Same as baseline, but restricted to the proton dataset
Iteration 1	Deuteron	deuteron-ite1	Same as baseline, but restricted to the deuteron dataset and supplemented with a proton covariance matrix determined from the proton-ite0 fit according to Eqn. 4.3.2.
	Proton and Deuteron	global-ite1-dw	Same as baseline, but supplemented with a deuteron covariance matrix determined from the deuteron-ite1 fit according to Eqn. 4.3.3.
Iteration 2	Deuteron	deuteron-ite2	Same as deuteron-ite1, but with a proton covariance matrix determined from the global-ite1-dw fit.
	Proton and Deuteron	global-ite2-dw	Same as global-ite1-dw, but with a deuteron covariance matrix determined from the deuteron-ite2 fit.
	Proton and Deuteron	global-ite2-sh	Same as global-ite2-sh, but with a deuteron covariance matrix and shifts determined according to Eqn. refeq:deuteronshifted.

Table 4.3.1 *A summary of the fits performed in this study, see text for details.*

"Iteration 1" and "Iteration 2".

As in the case for heavy nuclear data, it is useful to first look at the effect on the deuteron observables of using the deuteron PDF rather than the proton one (Fig. 4.3.2). The uncertainties are quite large but the ratio of observables is consistent with 1 in most regions other than high x, where nuclear shadowing is expected to play a part leading to large negative corrections. This mirrors what was seen in the heavy nuclear case (Fig. 4.2.1). The observables for NuSea show a systematic offset outwith uncertainties, like what we saw for CHORUS and NuTeV ν.

4.3.1 The deuteron covariance matrix

We now go on to investigate the deuteron covariance matrix. The diagonal elements are displayed in Fig. 4.3.3. Again, we see a pattern that parallels the pattern in the deuteron observable ratios; the size of the per-point uncertainty depends on the kinematics.
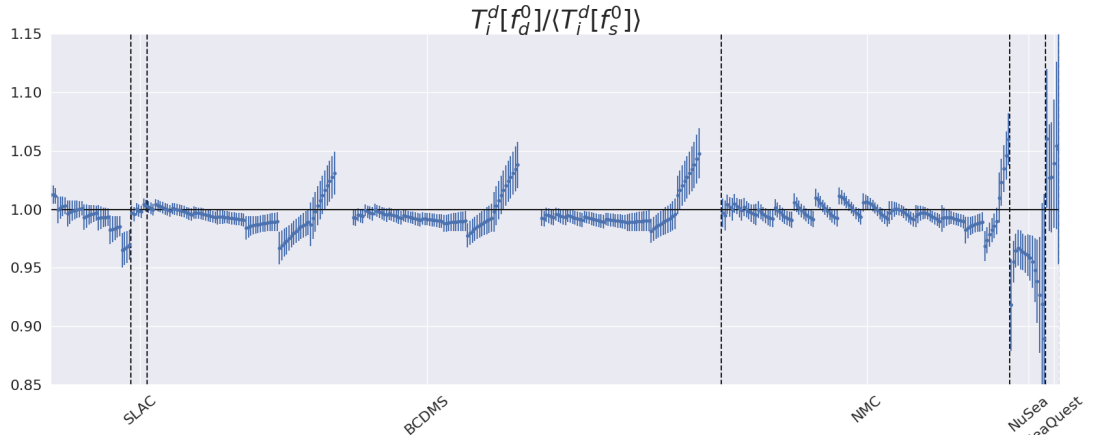
Figure 4.3.2 *Like Fig. 4.2.1 but for deuteron observables. Ratio between the deuteron observables computed with Iteration 1 deuteron PDFs, $T_i^d[f_d]$, and the central prediction computed with the isoscalar PDF, $\langle T_i^d[f_s] \rangle$. The error is the standard deviation of the distribution of $T_i^d[f_d]$ replicas. Data are organised in bins of increasing (x, Q^2) within each dataset.*
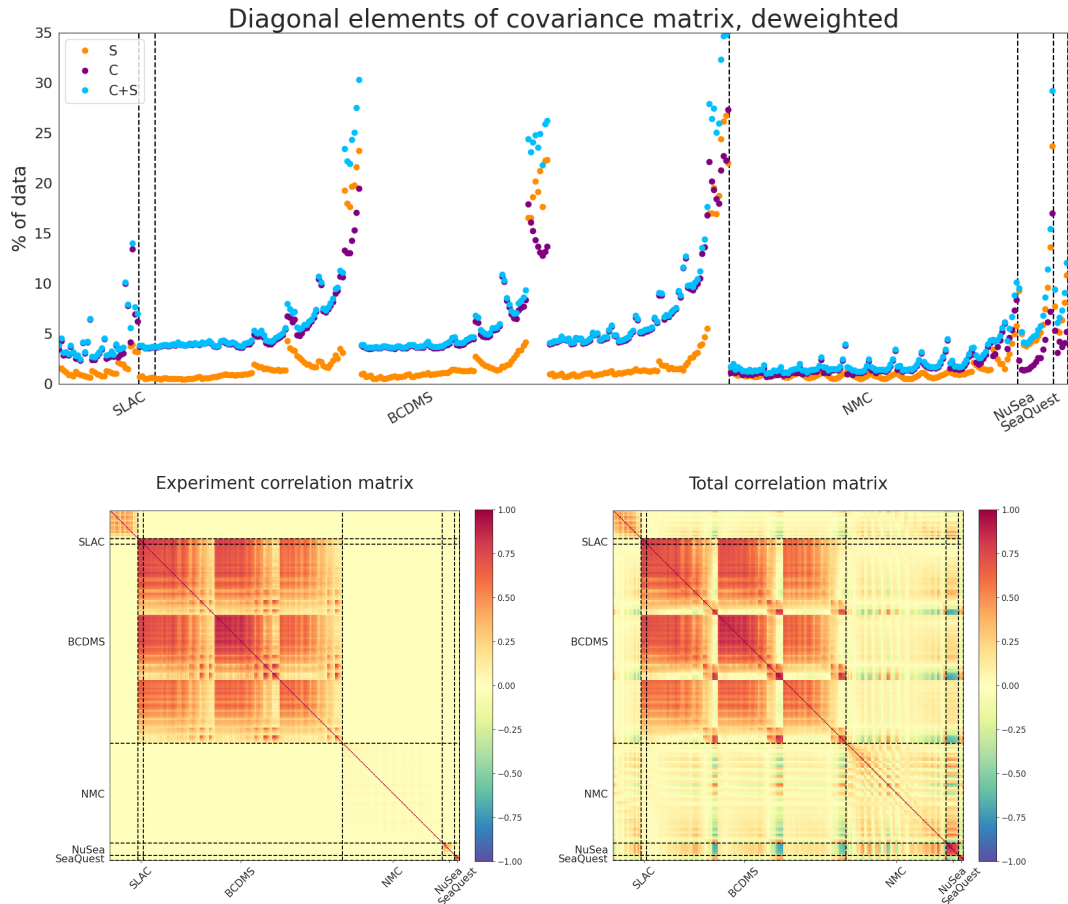
Figure 4.3.3 **Top panel:** *Square root of diagonal elements of covariance matrices for C (purple), S (deweighted; orange) and $C + S$ (deweighted; blue). All values are displayed as a % of data.* **Bottom panel:** *Correlation matrices. C (left) and $C + S$ (deweighted; right). The deweighted case only is displayed, but the qualitative features remain the same for the shifted case.*

91

The deuteron uncertainty is smaller than the experimental uncertainty for the pure deuteron datasets (SLAC and BCDMS), but is comparable for the mixed ratio data (NMC, NuSea and SeaQuest). This is because ratio data have smaller experimental uncertainties due to a significant cancellation of systematic errors. The bottom part of the figure shows the full matrix plots. As above, we see the most impact is on the ratio data. The figures show only the deweighted case, but this is qualitatively similar to the shifted case.

4.3.2 Deuteron correction factor

As an additional investigation, we can use the fitted deuteron PDFs to evaluate a correction to F_2 by computing the ratio F_2^d/F_2^p. This can then be compared to the result from nNNPDF2.0 and to the parametric correction used in MMHT2014 analyses [123], which is based on four fitted parameters.
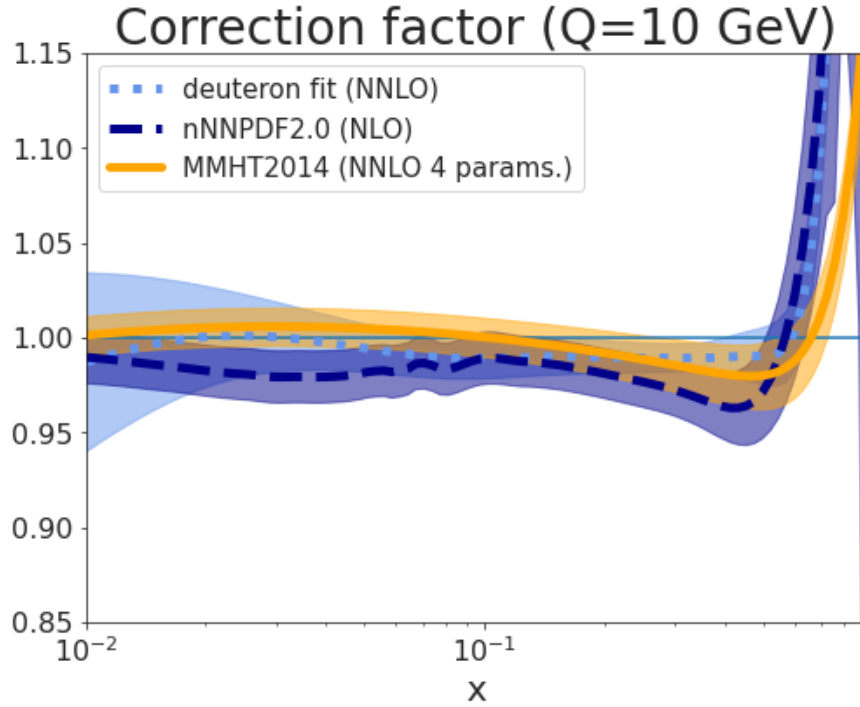
Figure 4.3.4 *F_2^d/F_2^p evaluated using deuteron PDFs from the present determination (deuteron-ite2), deuteron PDFs from nNNPDF2.0, and via the model correction used in MMHT2014 fits.*

This comparison can be seen in Fig. 4.3.4 at $Q = 10$ GeV. From this, deuteron correction to the structure function is clearly small, just a few percent across

the whole of x. The shape of the distribution also fits that expected from nuclear shadowing, with a dip at large x. All three determinations agree within uncertainties, giving us confidence in the robustness of our procedure. They also all have uncertainties of a similar size, despite the fact that nNNPDF2.0 is only at NLO. This is because it also contains heavy nuclear data through the constraint of continuity of A/Z. The uncertainty for the deuteron fit at NNLO is slightly larger at low x in the non-data region, reflecting the more conservative determination with no model dependence or continuity constraints imposed.

4.4 PDFs with nuclear uncertainties

Having constructed and studied deuteron and heavy nuclear uncertainties, we can then include them in fits. Table 4.4.1 summarises the various configurations of nuclear uncertainties and shifts. The NNPDF4.0 release includes (deweighted) deuteron and heavy nuclear uncertainties as a default, so it serves as the baseline for comparison.

Fit label	Description
NNPDF4.0	Baseline fit from NNPDF4.0
No nuclear unc.	Without nuclear uncertainties
Heavy nuclear unc.	Heavy nuclear uncertainties only
Heavy nuclear shifted	Heavy nuclear uncertainties with shifted central value
Deuteron unc.	Deuteron uncertainties only
Deuteron shifted	Deuteron uncertainties with shifted central value
Shifted	(All) nuclear uncertainties with shifted central value

Table 4.4.1 *A summary of the fits with nuclear uncertainties.*

	No nuc unc	Deut unc	Heavy nuc unc	**NNPDF4.0**	Deut shift	Heavy nuc shift	Shifted
χ^2	1.269	1.257	1.193	**1.162**	1.244	1.196	1.166
ϕ	0.160	0.158	0.160	**0.164**	0.158	0.170	0.169

Table 4.4.2 *Total χ^2 and ϕ values for nuclear data sets for the various fits.*

Table 4.4.2 gives the total χ^2 and ϕ values for these fits, where ϕ gives the ratio of uncertainties after fitting to the original data uncertainties, averaged over data points, and is defined in Eqn. 24 of [56]. The partial values per nuclear dataset can be seen in Fig. 4.4.1. Overall, including nuclear uncertainties causes the χ^2 to drop from 1.27 to 1.16, indicating a substantially better fit quality. This is to be expected, seeing as we are adding an uncertainty into the fit, accompanied by

an increase in ϕ. The NNPDF4.0 baseline gives the best fit, which is dominated by the inclusion of heavy nuclear uncertainties without a shift. The impact at the nuclear dataset level is striking, with a significant improvement in both χ^2 and ϕ for most of these datasets. The difference between the deweighted and shifted prescriptions, however, is minimal.
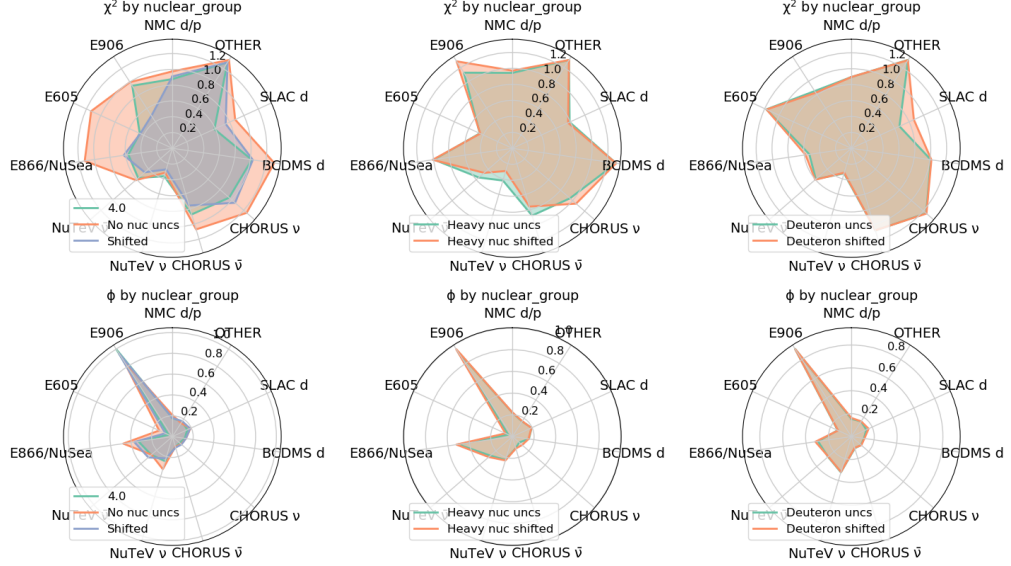
Figure 4.4.1 *Partial χ^2 (top row) and ϕ (bottom row) values broken down by nuclear dataset for the different configurations of uncertainties. All other datasets are collected under OTHER.*
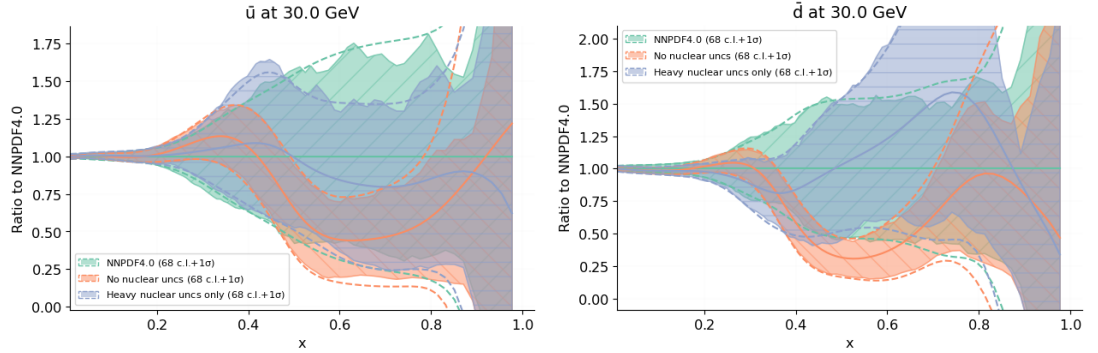
Figure 4.4.2 *Impact of including nuclear uncertainties in NNPDF4.0. The default (green) is to include them for all nuclear data. Fits with no nuclear uncertainties (orange) and with only heavy nuclear uncertainties (blue) are also shown.*

It's also helpful to look at the PDFs themselves. Nuclear uncertainties have an effect which is important in the large x region, where the nuclear data are. Firstly,

we would like to know what impact adding nuclear uncertainties has on the PDFs. In Fig. 4.4.2 we compare NNPDF4.0 to the fit without nuclear uncertainties. Including nuclear uncertainties causes a significant change to the shape of the PDFs in the large x region. This corresponds to the nuclear shadowing region, where nPDFs are lower compared to proton ones. Having no nuclear uncertainties causes the PDFs to be pulled downwards in this region, in the direction of the nPDFs.

Also displayed in Fig. 4.4.2 is a fit with only heavy nuclear uncertainties. It is clear that heavy nuclear uncertainties are responsible for the bulk of the impact, which is expected given the impact at the data level is more significant for thcese data (see Figs. 4.2.2 and 4.3.3).
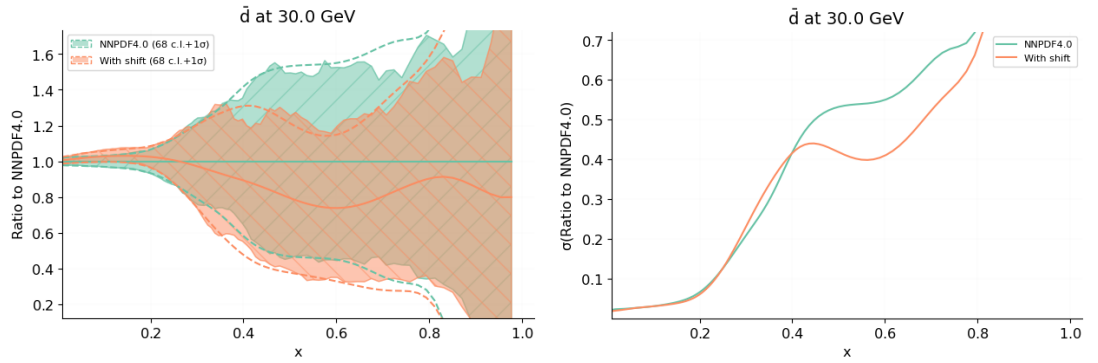
Figure 4.4.3   *The impact of shifting (orange) versus deweighting (green) for the $\bar{d}$ distribution. The right panel shows the uncertainties for clarity. The effects on the $\bar{u}$ distribution are qualitatively similar.*

Secondly, we would like to look at what happens when we also apply a shift to the nuclear data, following Eqn. 4.2.3. This can be seen in Fig. 4.4.3 for the $\bar{d}$ distribution. There is a clear reduction of PDF uncertainties when using the shifted prescription versus the deweighted prescription. However, as we saw before, there is little impact at the level of χ^2 and ϕ values. From this we can conclude that the deweighted prescription is the more conservative one. Given this, and that it leads to a slightly lower total χ^2 and ϕ, we opted to use the deweighted prescription in NNPDF4.0; including only an uncertainty for both deuteron and heavy nuclear data.

4.5 Summary

Nuclear data are important in PDF fits, but effects due to the nuclear environment are hard to quantify. To bring PDFs to 1% accuracy, we need to address these "small" but nevertheless important differencecs. We used the theory covariance matrix formalism outlined in Chapter 2 to include nuclear uncertainties in the next generation PDF fits. We adopted an empirical approach by recalculating predictions for nuclear observables with nuclear PDFs and using the shifts in predictions to construct a nuclear covariance matrix.

We analysed deuteron and heavy nuclear data separately, using nNNPDF2.0 for heavy nuclear PDFs, and fitting deuteron PDFs using an iterative proceduere in the NNPDF3.1 methodology. The resulting uncertainties are a crucial ingredient in the latest release, NNPDF4.0. They lead to a reduced global χ^2 and the relaxation of tension between the nuclear DIS and LHC DY data. Including the uncertainties causes a modest shift in the central value of PDFs and an increase in errors, in the high x nuclear data region. This difference is driven by the heavy nuclear data. We also investigated a procedure to shift the nuclear predictions, which was equivalent within errors to including only an uncertainty, but with a slightly higher global χ^2. Given this, we adopted using an uncertainty only as the conservative approach.

Chapter 5

Making predictions using PDFs with theoretical uncertainties

Earlier in this thesis we have discussed the importance of theoretical uncertainties, and how to include them in PDFs. We have also produced PDFs including MHOUs (see Chapter 3) and deuteron and nuclear uncertainties (see Chapter 4), which are freely available. In the future, PDFs with theory uncertainties will become the norm, and will be used widely to make theoretical predictions for observables by convoluting them with parton-level hard cross sections.

When making predictions for hadronic observables there are two sources of uncertainty: the hard cross section and the PDF. Typically, MHOUs for the former are estimated using scale variation, and these are added in quadrature to the latter. When the PDFs themselves also include MHOUs, we can think of the two scale evolutions considered being $Q_0 \to Q_{dat}$ and $Q_0 \to Q_{pred}$, where Q_0 is the PDF parametrisation scale, Q_{dat} is the scale of data in the PDF and Q_{pred} is the scale the prediction is made at. Then each of these sources has both a contribution from renormalisation scale variation and from factorisation scale variation. The PDFs themselves contain data from various processes, so when making a prediction for a process which is included in the PDFs there will invariably be correlations between the renormalisation scale variation in the PDF and that in the hard cross section. Even if the process is a new one, for example Higgs production, correlations due to factorisation scale variation will always be present. Simply combining the two sources of uncertainty in quadrature will miss these correlations and lead to an inflation in overall uncertainties. Hence we refer

to this method of combining uncertainties in quadrature as the "conservative prescription".

This issue was explored in detail in [146], noting that PDFs are a tool to express one observable in terms of others. This can be realised exactly in the simple case of fully correlated factorisation scale variation for non-singlet structure functions. Here the PDF can be eliminated entirely, and it is manifest that there is only one independent scale ($Q_{dat} \rightarrow Q_{pred}$), rather than two, with the MHOUs cancelling to a large degree. If MHOUs were used in both the PDF and the hard cross section in this case, it would amount to "double counting". It was also shown that correlations existed for renormalisation scale variation, albeit less strongly.

When using PDFs in predictions we include both sources of scale variation in an uncorrelated way, and so miss the MHOU cancellation and corresponding reduction in uncertainties. As noted in [4], this is a consequence of PDFs being universal. We cannot reconstruct the full data and MHOUs from the PDFs alone as information is lost in the fitting process; one set of PDFs could arise from many different data possibilities. However, if we want PDFs to be useful in a wide range of predictions, universality is required and so it would at first sight seem like we must live with this loss of correlation.

In [4] it was argued that we know the increase in PDF uncertainties due to MHOUs is small, and the effect is mostly realised in changes to the central value as the fit is rebalanced by changes to the weighting of different data. Indicative cases were explored in Chapter 7 of the study, where it was seen that the PDF uncertainty is consequently much smaller than the MHOU on the hard cross section. When combining these two in quadrature, the effect of missing correlations was therefore argued to be likely small, and so the overestimate of uncertainty would also be small. It was argued that a small overestimate of uncertainty is better than neglecting MHOUs altogether. It is therefore one of the aims of this study to see whether these claims are justified.

In this chapter we investigate correlations between PDF MHOUs and hard cross section MHOUs when making predictions. Although we focus on MHOUs, the analysis extends naturally to all sources of theory uncertainty. We develop a method for algebraically determining these correlations, and show how to include them when making a prediction. This is a complicated problem so we proceed incrementally:

- In Sec. 5.1 we show how a theoretical uncertainty can be reformulated in terms of a nuisance parameter, which holds the key to the propagation of uncertainties. We consider two extremal toy models: one where the theory is rigid with no unknown parameters ("pure theory"); the other where the theory is completely flexible and can fit the data exactly ("pure phenomenology").

- In Sec. 5.2 we move on to a model where the data are fitted using just one parameter. Here the correlations lead to a shift in the theoretical predictions. This is somewhere between the two scenarios in Sec. 5.1.

- In Sec. 5.3 we extend this analysis to a multi-parameter fit with multiple theory uncertainties, and then to a PDF fit, where the PDFs are continuous functions with a functional uncertainty.

- In Sec. 5.4 we present numerical results comparing this procedure with the naive approach of adding the uncorrelated contributions in quadrature, in the context of the NLO global fit with MHOUs discussed in Chapter 3. We make predictions including MHOUs for repetitions of all the experiments already in the fit (so-called "autopredictions"). We then investigate the scenario of a prediction for a process already in the fit (top production), and for a new process (Higgs production). We show that including these correlations leads to a shift in the central value of the prediction, which is within the uncertainties for the naive approach but takes the NLO predictions closer to the known NNLO result, reducing the χ^2 to experimental data. Furthermore, we find that for the autopredictions and top predictions there is a significant reduction in uncertainties due to the correlations, so the correlated predictions are both more precise and more accurate. For Higgs production, we find that the effect is much weaker, because there are only correlations through the factorisation scale and not through the renormalisation scale. We emphasise the power of these correlations as a way to improve theoretical predictions.

- In Sec. 5.5 we provide a summary.

5.1 Predictions with correlated theory uncertainties

We saw in Chapter 2 that to include theory uncertainties in a fit all you need to do is add a theory covariance matrix, S_{ij}, to the experimental covariance matrix C_{ij}. Recall that $i, j = 1, \ldots, N_{dat}$ run over data points. The only assumptions underlying this result are that all uncertainties are Gaussian, and that the theory uncertainties are independent of the experimental data. Since Gaussian experimental uncertainties are already assumed in NNPDF's framework, these assumptions are very reasonable. We can express the result as the conditional probability

$$P(T|D) \propto \exp\left(-\frac{1}{2}(T-D)^T (C+S)^{-1}(T-D) \right). \qquad (5.1.1)$$

Recall that both C and S are real and symmetric, that C is positive definite and that S is positive semidefinite and will generally possess many zero eigenvalues. In a fit we determine T from D by maximising $P(T|D)$, which amounts to minimising

$$\chi^2 = (T-D)^T (C+S)^{-1}(T-D) \qquad (5.1.2)$$

with respect to the free parameters which characterise the theory prediction.

In this section we start off by considering one single source of fully correlated theory uncertainty, so that

$$S = \beta\beta^T, \qquad (5.1.3)$$

where β are real and non-zero.

5.1.1 Nuisance parameters

We can model the theory uncertainty as a fully correlated shift in the theory prediction:

$$T \rightarrow T + \lambda\beta, \qquad (5.1.4)$$

where λ is a nuisance parameter characterising the scale of the shift. We will now show that this will lead us to Eqn. 5.1.1. Firstly, assuming Gaussian experimental

uncertainties, we can write

$$P(T|D\lambda) \propto \exp\left(-\frac{1}{2}(T + \lambda\beta - D)^T C^{-1}(T + \lambda\beta - D)\right). \tag{5.1.5}$$

Using Bayes' Theorem,

$$P(T|D\lambda)P(\lambda|D) = P(\lambda|TD)P(T|D). \tag{5.1.6}$$

We want to find $P(T|D)$, so we need an expression for the prior for λ, $P(\lambda|D) = P(\lambda)$, where we assume that the theory uncertainty is independent of the experimental data. We choose a unit-width Gaussian centred on zero,

$$P(\lambda) \propto \exp\left(-\frac{1}{2}\lambda^2\right). \tag{5.1.7}$$

Marginalising over λ, Eqn. 5.1.5 becomes

$$P(T|D) \propto \int d\lambda \exp\left(-\frac{1}{2}\left[(T + \lambda\beta - D)^T C^{-1}(T + \lambda\beta - D) + \lambda^2\right]\right). \tag{5.1.8}$$

We can evaluate the term in $[\cdot]$ by remembering $S = \beta\beta^T$, introducing the variable

$$Z \equiv (1 + \beta^T C^{-1}\beta)^{-1}, \tag{5.1.9}$$

and completing the square:

$$\begin{aligned}
[\cdot] &= (T - D)^T C^{-1}(T - D) + (T - D)^T C^{-1}\lambda\beta + \lambda\beta^T C^{-1}(T - D) \\
&\quad + \lambda\beta^T C^{-1}\lambda\beta + \lambda^2 \\
&= (T - D)^T C^{-1}(T - D) + (T - D)^T C^{-1}\lambda\beta + \lambda\beta^T C^{-1}(T - D) + \lambda^2 Z^{-1} \\
&= Z^{-1}(\lambda + Z\beta^T C^{-1}(T - D))^2 \\
&\quad - Z(\beta^T C^{-1}(T - D))^2 + (T - D)^T C^{-1}(T - D) \\
&= Z^{-1}(\lambda + Z\beta^T C^{-1}(T - D))^2 \\
&\quad + (T - D)^T(C^{-1} - ZC^{-1}SC^{-1})(T - D).
\end{aligned}$$
$$\tag{5.1.10}$$

Finally, we can use the Sherman-Morrison formula and Eqn. 5.1.2 to write this as

$$[\cdot] = Z^{-1}(\lambda + Z\beta^T C^{-1}(T-D))^2 + (T-D)^T(C+S)^{-1}(T-D)$$
$$= Z^{-1}(\lambda + Z\beta^T C^{-1}(T-D))^2 + \chi^2 \tag{5.1.11}$$
$$\equiv Z^{-1}(\lambda - \bar{\lambda}) + \chi^2,$$

where we have defined
$$\bar{\lambda} = Z\beta^T C^{-1}(D-T). \tag{5.1.12}$$

Plugging this back into Eqn. 5.1.8, we get

$$P(T|D) \propto \int d\lambda \; e^{-\frac{1}{2}\chi^2} \exp\left(-\frac{1}{2}Z^{-1}(\lambda - \bar{\lambda})^2\right)$$
$$\propto e^{-\frac{1}{2}\chi^2}, \tag{5.1.13}$$

which is Eqn. 5.1.1. The advantage of this approach is that we can also get the posterior distribution for λ (after fitting using D and T), by using Bayes' Theorem (Eqn. 5.1.6):

$$P(\lambda|TD) = \frac{P(T|D\lambda)P(\lambda)}{P(T|D)}$$
$$= \exp\left(-\frac{1}{2}\left[(T+\lambda\beta-D)^T C^{-1}(T+\lambda\beta-D)+\lambda^2-\chi^2\right]\right) \tag{5.1.14}$$
$$= \exp\left(-\frac{1}{2}Z^{-1}(\lambda-\bar{\lambda})\right),$$

where we recognised the similarity between the exponent here and in Eqn. 5.1.8. So the effect of the fit is to shift the centre of the distribution from $0 \to \bar{\lambda}$, and the width from $1 \to Z$. Note that from the definition of Z (Eqn. 5.1.9),

$$0 \leq Z \leq 1, \tag{5.1.15}$$

so the theory uncertainty is always reduced when information on D is added. Fig. 5.1.1 gives a sketch of this effect.
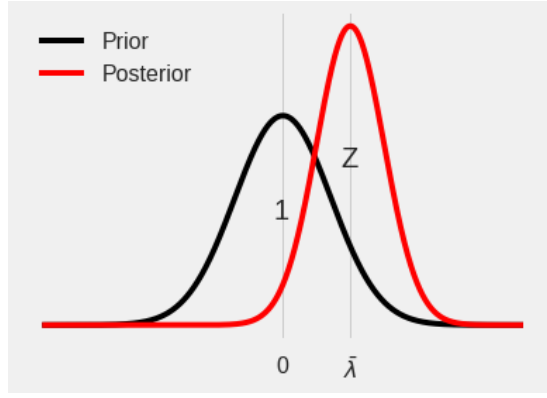
Figure 5.1.1 *Sketch of the prior (Eqn. 5.1.7) and posterior (Eqn. 5.1.14) distributions for λ. Adding information shifts the distribution and reduces the width.*

5.1.2   Predictions without fits

We will now test out this formalism for a toy model where we have "pure theory" values, T_0. These have no unknown parameters, so cannot be fitted. They do, however, have a theory uncertainty. Despite the fact we cannot fit them to the data, $T_0 \neq D$ in general, the data can still give us information on the predictions via the nuisance parameters in the theory uncertainties. The expectation value of λ can be evaluated

$$E[\lambda] = \mathcal{N}_\lambda \int d\lambda \; \lambda \; P(\lambda|T_0 D) = \bar{\lambda}(T_0, D). \qquad (5.1.16)$$

Note that $\mathcal{N}_\lambda$ is set such that $E[1] = 1$. We can then find the variance

$$Var[\lambda] \equiv E[(\lambda - E[\lambda])^2] = Z. \qquad (5.1.17)$$

Both of these can be seen straight away from the form of the posterior for λ, Eqn. 5.1.14.

In the nuisance parameter formalism, we can write

$$T(\lambda) = T_0 + \lambda\beta. \qquad (5.1.18)$$

Before comparing these theory predictions to the data, we could use the prior for

λ in this expression, which would give us

$$E[T(\lambda)] = T_0, \qquad \mathrm{Cov}[(T(\lambda)] = \beta\beta^T = S. \qquad (5.1.19)$$

But we could instead first compare T to D, and then use the posterior distribution. In that case we'd end up with

$$\begin{aligned}
E[T(\lambda)] &= T_0 + \bar{\lambda}(T, D)\beta \\
&= T_0 + \beta\beta^T (C + S)^{-1}(D - T_0) \qquad (5.1.20) \\
&= T_0 + S(C + S)^{-1}(D - T_0),
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{Cov}[(T(\lambda)] &= E[(T(\lambda) - E[T(\lambda)])^2] \\
&= \mathrm{Var}[\lambda]\beta\beta^T \qquad (5.1.21) \\
&= ZS.
\end{aligned}$$

We can think of these predictions as "autopredictions", i.e. we:

1. compare the data to the theory;

2. use the information from 1. to make new predictions for exact repetitions of the same experiments.

From Eqn. 5.1.20 we can see that there is a shift in the autopredictions of

$$\delta T = -S(C + S)^{-1}(T_0 - D), \qquad (5.1.22)$$

and that the uncertianties are reduced by a factor $\sqrt{Z}$, thanks to information provided by the data. Overall, the autoprediction covariance matrix is

$$\begin{aligned}
ZS &= S - S(C + S)^{-1}S \\
&= C(C + S)^{-1}S = S(C + S)^{-1}C. \qquad (5.1.23)
\end{aligned}$$

To see the impact of the shift in the predictions, we can compare the experimental χ^2 (i.e. using the experimental covariance matrix only) of the original predictions to the autopredictions. The original χ^2 is

$$\chi^2_{exp} = (T_0 - D)^T C^{-1}(T_0 - D), \qquad (5.1.24)$$

and, using Eqn. 5.1.22, the autoprediction χ^2 is

$$
\begin{aligned}
\chi^2_{auto} &= (T_0 + \delta T - D)^T C^{-1} (T_0 + \delta T - D) \\
&= (T_0 - D)(C + S)^{-1} C (C + S)^{-1} (T_0 - D).
\end{aligned}
\tag{5.1.25}
$$

From this we can see that $\chi^2_{auto} \leq \chi^2_{exp}$ because C and S are both positive semidefinite. In other words, the shifts always lead to improved quality of fit by exploiting the theory uncertainty to add information from the data.

We can investigate this more explicitly by using a simple model where the experimental covariance matrix is diagonal and the theory uncertainty is fully correlated. Writing $\beta = se$, where s is the size of correlated theory uncertainty and $e^T e = 1$, we have

$$
C = \sigma^2 \mathbb{I}, \qquad S = s^2 e e^T,
\tag{5.1.26}
$$

where σ is the per-point uncertainty. Then we can evaluate:

$$
\begin{aligned}
(C + S)^{-1} &= \frac{1}{\sigma^2} \left(1 - \frac{s^2}{\sigma^2 + s^2} e e^T \right); \\
Z &= (1 + s^2/\sigma^2)^{-1}.
\end{aligned}
\tag{5.1.27}
$$

so the reduction in the theory uncertainties depends on s^2/σ^2:

- for $s^2 \ll \sigma^2$, there is a small influence of the data on the theory uncertainty;

- for $\sigma^2 \ll s^2$, the theory uncertainty size is reduced from s to σ, because in the limit $s^2/\sigma^2 \to \infty$, $Z \to \sigma^2 e e^T$;

- for theory uncertainties the same size as experimental ones, $\sigma^2 \sim s^2/N$ and $Z \sim 1/(N + 1)$, so if there are a large number of independent data points then there is a large reduction in uncertainty; more data gives more information.

In this model the autoprediction shifts are

$$
\delta T = \frac{-s^2}{\sigma^2 + s^2} e^T (T_0 - D) e,
\tag{5.1.28}
$$

which are in the direction of the theory uncertinty, e, as we would expect. When $s^2/\sigma^2 \to \infty$, $e^T (T_0 + \delta T) \to e^T D$, so in this direction the autopredictions coincide with the data.

It can also be shown that the autoprediction χ^2 is

$$\chi^2_{auto} = (T_0 - D)^T \frac{1}{\sigma}\left(1 - \frac{s^2(s^2 + 2\sigma^2)}{(s^2 + \sigma^2)^2}ee^T\right)(T_0 - D), \qquad (5.1.29)$$

so the contributions to the χ^2 which are orthogonal to e are unchanged, and the contributions along e are reduced by Z^2. This means unless the theory uncertainties are very small, we will end up with a χ^2 for the autopredictions which is size $(N-1)$ rather than N, because the contribution along e will be substantially reduced.

We can also make genuine predictions, $\widetilde{T}_I$, $I = 1, \ldots \widetilde{N}$. In this scenario these also have no free parameters, but their theory uncertainty is correlated with that of T_i, $i = 1, \ldots, N$, for which we have experimental data, D_i. We can write the predictions in the nuisance parameter formalism as

$$\widetilde{T}(\widetilde{\lambda}) = \widetilde{T} + \widetilde{\lambda}\widetilde{\beta}, \qquad (5.1.30)$$

where the vector $\widetilde{\beta}$ gives the size and direction of the theory uncertainty in $\widetilde{T}$. If $\widetilde{\lambda}$ are independent of λ, then the theory uncertainties in $\widetilde{T}$ are uncorrelated with those in T and the covariance for the predictions is just

$$\widetilde{S} = \widetilde{\beta}\widetilde{\beta}^T. \qquad (5.1.31)$$

In the opposite scenario where they are fully correlated, $\widetilde{\lambda} = \lambda$ and we can use the data D for T to improve our prediction by using the posterior of λ, i.e. $P(\lambda|TD)$:

$$\begin{aligned} E[\widetilde{T}(\lambda)] &= \widetilde{T} + \overline{\lambda}(T, D)\widetilde{\beta} \\ &= \widetilde{T} + \widetilde{\beta}\beta^T(C + S)^{-1}(D - T), \end{aligned} \qquad (5.1.32)$$

so there is a shift in the predictions of

$$\delta\widetilde{T} = -\widehat{S}(C + S)^{-1}(T - D), \qquad (5.1.33)$$

where $\widehat{S} = \widetilde{\beta}\beta^T$ is the covariance matrix of cross correlations between the theory uncertainty in the theory for which there are data and that in the predictions.

The covariance matrix of the predictions can be calculated

$$
\begin{aligned}
\mathrm{Cov}[\widetilde{T}(\lambda)] &\equiv E[(\widetilde{T}(\lambda) - E[\widetilde{T}(\lambda)]^2) \\
&= E[((\widetilde{\lambda} - \overline{\lambda})\widetilde{\beta})((\widetilde{\lambda} - \overline{\lambda})\widetilde{\beta})^T] \\
&= \mathrm{Var}[\lambda]\widetilde{\beta}\widetilde{\beta}^T \\
&= Z\widetilde{S}.
\end{aligned}
\tag{5.1.34}
$$

So the covariance of the genuine predictions is reduced by the same factor, Z, as the autopredictions. This means the data can work via the correlations in theory uncertainties to produce more precise and (if the theory is correct) more accurate predictions for observables that aren't yet measured. This is accompanied by a shift which is proportional to the cross covariance between theory uncertainties.

We can understand a little more about the various theory covariance matrices by imagining we obtained some experimental measurements, $\widetilde{D}$, corresponding to predictions $\widetilde{T}$. Then we could add these to the fit, and would get a new fitting theory covariance matrix with dimensions $(N + \widetilde{N}) \times (N + \widetilde{N})$ and content

$$
\begin{pmatrix} S & \widehat{S} \\ \widehat{S}^T & \widetilde{S} \end{pmatrix} = \begin{pmatrix} \beta\beta^T & \beta\widetilde{\beta}^T \\ \widetilde{\beta}\beta^T & \widetilde{\beta}\widetilde{\beta}^T \end{pmatrix}.
\tag{5.1.35}
$$

Here the role of $\widehat{S}$ and $\widehat{S}^T$ as cross-correlation theory covariance matrices is clear. We can see that the theory uncertainties in the prediction are consistent with the theory uncertainties we would use when including observables in a fit, which makes sense in the comparison with the autopredictions case.

Ideally, the shifted predictions would give a better χ^2 to the new data $\widetilde{D}$, but this is not guaranteed because the shifts were induced from the old data, D, and there could be inconsistencies between D and $\widetilde{D}$.

5.1.3   Autopredictions in a perfect fit

We have just considered the scenario where T is unfitted to the data. Now consider the "opposite" situation of a perfect fit. Here T have a high level of flexibility, and can fit D exactly. $P(T|D)$ is always maximised when $T = D$, so $\chi^2 = 0$. We can extract the expectation value and covariance of T as

$$
E[T] = \mathcal{N}_T \int dT \; T \; P(T|D) = D,
\tag{5.1.36}
$$

where $\mathcal{N}_T$ is such that $E[1] = 1$, and

$$\text{Cov}[T] \equiv E[(T - E[T])^2] = C + S. \tag{5.1.37}$$

We can write the autopredictions again in the nuisance parameter formalism, so

$$T(\lambda) = T + \lambda\beta. \tag{5.1.38}$$

When calculating the expectation value of a function of T and λ, we must take some care. The data, D are always held fixed because they are set values from experiment. Then the expectation value can be calculated using conditional probabilities,

$$E[f(T, \lambda)] \equiv \mathcal{N}_T \int dT \; \mathcal{N}_\lambda \int d\lambda \; f(T, \lambda) P(\lambda | TD) P(T | D). \tag{5.1.39}$$

We must therefore integrate first over λ, because this is conditional on T, and then over T. So to get the expectation value of λ, we first take the expectation value over λ, with both T and D fixed, making use of Eqn. 5.1.12:

$$\begin{aligned} E[\lambda] &= E[\bar{\lambda}(T, D)] \\ &= \beta^T (C + S)^{-1} E[D - T]. \end{aligned} \tag{5.1.40}$$

Then we take the expectation value over T, keeping D fixed, which just gives us

$$E[\lambda] = \beta^T (C + S)^{-1} (D - D) = 0. \tag{5.1.41}$$

For the variance we have

$$\text{Var}[\lambda] = E[(\lambda - E[\lambda])^2] = E[\lambda^2]. \tag{5.1.42}$$

To evaluate this we can use the trick of adding and subtracting $\bar{\lambda}(T, D)$ to λ because we are aiming to put it in terms of $Z = E[(\lambda - \bar{\lambda}(T, D))^2]$. Making use

of Eqn. 5.1.9 and Eqn. 5.1.37,

$$
\begin{aligned}
\mathrm{Var}[\lambda] &= E[(\lambda - \overline{\lambda}(T,D) + \overline{\lambda}(T,D))^2] \\
&= E[(\lambda - \overline{\lambda}(T,D))^2] + E[\overline{\lambda}(T,D)^2] \\
&= Z + \beta^T (C+S)^{-1} E[(T-D)(T-D)^T](C+S)^{-1}\beta \\
&= Z + \beta^T (C+S)^{-1} \mathrm{Cov}[T](C+S)^{-1}\beta \\
&= 1 - \beta^T(C+S)^{-1}\beta + \beta^T(C+S)^{-1}\beta \\
&= 1.
\end{aligned}
\tag{5.1.43}
$$

So in a perfect fit, the posterior distribution of nuisance parameters is exactly the same as the prior. All the information from the data is absorbed into the fitted parameters, and so we are left with no update to the theoretical uncertainty. In the calculation of the variance you can see how the reduction by factor Z that we saw in the pure theory case is exactly cancelled by the factor due to the fluctuation of $\overline{\lambda}(T,D)$ due to the covariance of T.

We can now use the posterior of λ to calculate the autopredictions. First we calculate the expectation value:

$$
E[T(\lambda)] = E[T + \lambda\beta] = D,
\tag{5.1.44}
$$

which is a consistency check. Next we calculate the covariance, where we remember the expectation value is taken first over λ with T and D fixed, and then over T with D fixed:

$$
\begin{aligned}
\mathrm{Cov}[T(\lambda)] &= E[(T(\lambda) - E[T(\lambda)])^2] \\
&= E[(T - D + \lambda\beta)(T - D + \lambda\beta)^T] \\
&= E[(T-D)(T-D)^T] + E[\lambda\beta(T-D)^T] \\
&\quad + E[(T-D)\lambda\beta^T] + E[\lambda^2]\beta\beta^T.
\end{aligned}
\tag{5.1.45}
$$

The first term is just $\mathrm{Cov}[T]$ and the last term is $\mathrm{Var}[\lambda]S = S$. To calculate the middle terms, consider

$$
\begin{aligned}
E[\lambda\beta(T-D)^T] &= E[\beta\overline{\lambda}(T,D)(T-D)^T] \\
&= -S(C+S)^{-1}E[(T-D)(T-D)^T] \\
&= -S(C+S)^{-1}\mathrm{Cov}[T].
\end{aligned}
\tag{5.1.46}
$$

So overall

$$\begin{aligned}
\text{Cov}[T(\lambda)] &= \text{Cov}[T] - S(C+S)^{-1}\text{Cov}[T] - \text{Cov}[T](C+S)^{-1}S + \text{Var}[\lambda]S \\
&= (C+S) - S - S + S \\
&= C.
\end{aligned}$$

(5.1.47)

So in a perfect fit, the covariance of the autopredictions is equal to the covariance of the data. This happened because the covariance arising from the fit (the first term in Eqn. 5.1.45), and the covariance arising in the autopredictions (the last term in Eqn. 5.1.45) are each cancelled by the cross covariance between the fit and the prediction. This is the effect which was noted in [146]. In a perfect fit, there is no distinction between the autopredictions and the data, and so the theory uncertainty is irrelevant. In other words, the case of a perfect fit can be thought of as "pure phenomenology"; the only information we are left with is in the data. As a result we can't make genuine predictions for points we don't have experimental data for because there's no real underlying theory.

5.2   One-parameter fits

Previously, we looked at two unrealistic simple models;

1. Fixed theory which cannot be fitted to data (pure theory);

2. Over-flexible theory which is fitted perfectly to data (pure phenomenology).

These helped us to develop the nuisance parameter approach, but in reality we want somewhere between these two extremes; normally the theory has parameters which are constrained by the data and can be fitted, however the theory is rigid enough to be able to make new predictions, $\widetilde{T}$, where no data exist. We will see that in this more realistic case there are features from both the pure theory and the pure pheno cases, namely:

1. shifts in the central values;

2. reduction in uncertainty;

3. correlations of theory uncertainties.

First we will consider a theory with one fitted parameter. Then we will generalise this to many fitted parameters, which is a description of many modern PDF fits. Finally, we will consider an NNPDF fit, where the many parameters are encapsulated in a neural network.

Starting with the single parameter fit, $T(\theta)$ only depend on a single parameter, θ. The χ^2 (Eqn. 5.1.2) will be minimised for some $\theta = \theta_0$, with variance $\text{Var}[\theta]$. Once θ_0 has been determined we can then make some predictions, $\widetilde{T}(\theta_0)$, where the tilde denotes they are predictions for theories separate from the fit inputs. These predictions will have uncertainties proportional to $\text{Var}[\theta]$.

We have assumed that the uncertainties are Gaussian, and so they are differentiable. This means we can linearise $T(\theta)$ about $T(\theta_0) \equiv T_0$:

$$T(\theta) = T_0 + (\theta - \theta_0)\dot{T}_0 + \ldots. \tag{5.2.1}$$

We want to determine the uncertainty in the fitted θ, so we need to propagate the uncertainties in the data, D, and the theory, $T(\theta)$, into θ. We can do this using the standard NNPDF approach (see Chapter 1) of generating pseudodata replicas, $D^{(r)}$, which are Gaussianly distributed about the actual data, D, with covariance $C + S$. It is important to remember that these are just a device for propagating the uncertainty, and we must still hold D constant when taking expectation values. More explicitly, we can define the average over replicas for any function, F, of the replicas as

$$\langle F(D^{(r)}) \rangle = \lim_{N_{rep} \to \infty} \frac{1}{N_{\text{rep}}} \sum_{r=1}^{N_{\text{rep}}} F(D^{(r)}). \tag{5.2.2}$$

Then the replicas will satisfy

$$\langle D^{(r)} \rangle \equiv D, \qquad \langle (D^{(r)} - D)(D^{(r)} - D)^T \rangle = C + S. \tag{5.2.3}$$

in the limit of $N_{rep} \to \infty$.

The fit proceeds by fitting a parameter replica, $\theta^{(r)}$, for each pseudodata replica, $D^{(r)}$, by minimising

$$\chi_r^2[\theta] = (T(\theta) - D^{(r)})^T (C + S)^{-1} (T(\theta) - D^{(r)}), \tag{5.2.4}$$

with respect to θ. Using Eqn. 5.2.1, this leads to

$$\theta^{(r)} - \theta_0 = \frac{\dot{T}_0^T (C + S)^{-1} (D^{(r)} - T_0)}{\dot{T}_0^T (C + S)^{-1} \dot{T}_0}. \tag{5.2.5}$$

Now $\theta_0 = \langle \theta^{(r)} \rangle$, so using the replica averages in Eqn. 5.2.3 we find

$$\dot{T}_0^T (C + S)^{-1} (D - T_0) = 0, \tag{5.2.6}$$

so we can rewrite Eqn. 5.2.5 as

$$\theta^{(r)} - \theta_0 = \frac{\dot{T}_0^T (C + S)^{-1} (D^{(r)} - D)}{\dot{T}_0^T (C + S)^{-1} \dot{T}_0}. \tag{5.2.7}$$

Using the fact that C and S are symmetric,

$$
\begin{aligned}
\text{Var}[\theta] &= \langle (\theta^{(r)} - \theta_0)^2 \rangle \\
&= \frac{\dot{T}_0^T (C + S)^{-1} \langle (D^{(r)} - D)(D^{(r)} - D)^T \rangle (C + S)^{-1} \dot{T}_0}{(\dot{T}_0^T (C + S)^{-1} \dot{T}_0)^2} \\
&= (\dot{T}_0^T (C + S)^{-1} \dot{T}_0)^{-1}. \tag{5.2.8}
\end{aligned}
$$

Note that:

- data points with a large dependence on θ have large $\dot{T}_0$ and contribute more.

- directions with large uncertainty, $(C + S)$, contribute less.

Now we have the uncertainty in the fitted parameter, θ, we can find the fitting uncertainty. This is the covariance of $T(\theta)$ due to the experimental and theoretical uncertainties from fitting θ. We will call this covariance matrix X. Using the fact that $E[T] = \langle T(\theta^{(r)}) \rangle = T(\theta_0) = T_0$ and writing $T^{(r)} = T(\theta^{(r)})$,

$$
\begin{aligned}
X \equiv \text{Cov}[T(\theta)] &= \langle (T^{(r)} - T_0)(T^{(r)} - T_0)^T \rangle & (5.2.9) \\
&= \dot{T}_0 \langle (\theta^{(r)} - \theta_0)^2 \rangle \dot{T}_0^T & (5.2.10) \\
&= \dot{T}_0 (\dot{T}_0^T (C + S)^{-1} \dot{T}_0)^{-1} \dot{T}_0^T & (5.2.11) \\
&= n(n^T (C + S)^{-1} n)^{-1} n^T, & (5.2.12)
\end{aligned}
$$

where in the last line we define $\dot{T}_0 \equiv |\dot{T}_0| n$, i.e. n is a unit vector in the direction of $\dot{T}_0$. We can see that X depends only on the direction (n) of $\dot{T}_0$, not its magnitude.

Note that X is singular and also that

$$X = X(C + S)^{-1}X, \tag{5.2.13}$$

which will be useful later. Using Eqn. 5.2.7 in Eqn. 5.2.1, we can see that

$$T^{(r)} - T_0 = X(C + S)^{-1}(D^{(r)} - D^{(0)}), \tag{5.2.14}$$

so $X(C + S)^{-1}$ projects the data replicas onto the theory replicas.

Now let's revisit the model for covariance matrices, Eqn. 5.1.26. If we define the angle between the experimental uncertainties and the θ variation by $\cos \phi = n^T e$, we find that

$$n^T(C + S)^{-1}n = \frac{\sigma^2 + s^2 \sin^2 \phi}{\sigma^2(\sigma^2 + s^2)}, \tag{5.2.15}$$

$$n^T X n = \frac{\sigma^2(\sigma^2 + s^2)}{(\sigma^2 + s^2 \sin^2 \phi)}. \tag{5.2.16}$$

Note that using any vector other than n here gives 0. We can see that the effects from the theory uncertainty (s) depend on its degree of alignment with n, the direction of the parameter dependence.

- For complete alignment, $\phi = 0$ and the variance of T in this direction is $(\sigma^2 + s^2)$.

- When they are orthogonal, $\phi = \frac{\pi}{2}$ and the variance is σ^2, so the theory uncertainty doesn't factor into the fitting.

5.2.1 Autopredictions in single parameter fits

We can now get the expectation and covariance of the autopredictions in the one-parameter model. We write the autopredictions again as

$$T(\theta, \lambda) = T(\theta) + \lambda\beta. \tag{5.2.17}$$

As before, we compute the expectation values over λ using $P(\lambda|TD)$ and then over T using $P(T|D)$. We do the expectation over T by averaging over the theory replicas $T^{(r)} \equiv T(\theta^{(r)})$. All this time we must hold D fixed as the probabilities are conditional on the data. As stated before, $D^{(r)}$ are not physical, they are just

an artificial device we use to allow us to propagate the uncertainties. So we don't average over the data replicas when getting the expectation values. Explicitly,

$$E[f(T, \lambda)] = \langle (\mathcal{N}_\lambda \int d\lambda \, f(T^{(r)}, \lambda) P(\lambda | T^{(r)}, D)) \rangle, \qquad (5.2.18)$$

where we recall that $\langle \cdot \rangle$ denotes the replica average.

Expectation value

To get the expectation value of the autopredictions, we do the same steps as we did for the perfect fit, but now we have the theory replicas as well. So first we find the expectation value of λ,

$$\begin{aligned} E[\lambda] &= \langle \overline{\lambda}(T(\theta)^{(r)}), D) \rangle \\ &= \beta^T (C + S)^{-1} (D - T_0) \qquad (5.2.19) \\ &\equiv \overline{\lambda}_0. \end{aligned}$$

We can see the parallel here with the pure theory scenario, where the nuisance parameters can have non-zero expectation values. This is because the one parameter fit is not perfect. We can now calculate the expectation value of the autopredictions:

$$\begin{aligned} E[T(\theta, \lambda)] &= \langle T^{(r)} + \overline{\lambda}(T^{(r)}, D)\beta \rangle \\ &= T_0 + \overline{\lambda}_0 \beta \qquad (5.2.20) \\ &= T_0 + S(C + 1)^{-1}(D - T_0), \end{aligned}$$

and therefore the shift induced is

$$\delta T = -S(C + S)^{-1}(T_0 - D). \qquad (5.2.21)$$

Note that Eqn. 5.2.6 tells us that $n^T(C+S)^{-1}(T_0 - D) = 0$, so the shifts are only non-zero when n and e point in different directions. When they are parallel (i.e. $\phi = 0$), the theory uncertainty is absorbed by the fit, like in the perfect fit. We can use the same arguments as we did in the pure theory case to conclude that the shifts always improve the fit to experimental data.

114

Covariance

Now we can find the covariance of autopredictions. We start by computing the variance of λ, again using the trick of adding and subtracting $\overline{\lambda}(T^{(r)}, D)$. When taking the expectation value over T we use the replica average.

$$
\begin{aligned}
\mathrm{Var}[\lambda] &= E[(\lambda - E[\lambda])^2] \\
&= E[(\lambda - \overline{\lambda}(T^{(r)}, D) + \overline{\lambda}(T^{(r)}, D) - \overline{\lambda}_0)^2] \\
&= E[(\lambda - \overline{\lambda}(T, D))^2] + \langle(\overline{\lambda}(T^{(r)}, D) - \overline{\lambda}_0)^2\rangle \\
&= Z + \beta^T(C + S)^{-1}\langle(T^{(r)} - T_0)(T^{(r)} - T_0)^T\rangle(C + S)^{-1}\beta \\
&= 1 - \beta^T(C + S)^{-1}\beta + \beta^T(C + S)^{-1}X(C + S)^{-1}\beta \\
&\equiv \overline{Z}.
\end{aligned}
\tag{5.2.22}
$$

So unlike in a perfect fit, the last two terms don't cancel. So the information in the data can't just be totally absorbed into the fitted parameter, because there isn't enough flexibility for this to happen. As a result that information can have an impact on the nuisance parameters.

We can see that $\overline{Z} \geq Z$ because $(C + S)^{-1}X(C + S)^{-1}$ is positive semidefinite, and $\overline{Z} \leq 1$ because $X(C + S)^{-1}$ is projective (Eqn. 5.2.14), so its eigenvalues are either 0 or 1. Overall

$$
0 < Z \leq \overline{Z} \leq 1,
\tag{5.2.23}
$$

so the information from the data about the theory uncertainties is less in the single parameter fit than in the pure theory. This is due to the uncertainty in the fit. However, unlike in the perfect fit, the data do still constrain the theory uncertainties, provided that the fitted parameter and the theory uncertainties are in different directions.

In the simple model for uncertainties we introduced in Eqn. 5.1.26, it can be shown that

$$
\overline{Z} = \frac{\sigma^2}{\sigma^2 + s^2 \sin^2 \phi},
\tag{5.2.24}
$$

so

- $\overline{Z} = 1$ when $\phi = 0$, i.e. when $n = e$ and the parameter variation is aligned with the theory uncertainty;

- $\overline{Z} = Z$ only if $\phi = \pi/2$, i.e. $n \perp e$ and here the data have the greatest influence because there is no absorption of information into the fit.

115

Finally, we can use this information to calculate the covariance of the autopredictions. Again, we first take the expectation value over λ, holding T and D fixed, and then use the theory replicas to take the expectation value over T with D fixed.

$$
\begin{aligned}
\mathrm{Cov}[T(\theta, \lambda)] &= E[T(\theta, \lambda) - E[T(\theta, \lambda)])^2] \\
&= E[(T - T_0 + (\lambda - \overline{\lambda}_0)\beta)(T - T_0 + (\lambda - \overline{\lambda})\beta)^T] \\
&= \langle (T^{(r)} - T_0)(T^{(r)} - T_0)^T \rangle + E[(\lambda - \overline{\lambda}_0)\beta(T - T_0)^T] \\
&\quad + E[(T - T_0)(\lambda - \overline{\lambda}_0)\beta^T] + E[(\lambda - \overline{\lambda}_0)^2]\beta\beta^T.
\end{aligned}
\tag{5.2.25}
$$

The first term is $\mathrm{Cov}[T] = X$ and the last term is $\mathrm{Var}[\lambda]S$. The cross terms can be evaluated like

$$
\begin{aligned}
E[(\lambda - \overline{\lambda}_0)\beta(T - T_0)^T] &= \langle \beta(\overline{\lambda}(T^{(r)}, D) - \overline{\lambda}(T_0, D))(T^{(r)} - T_0)^T \rangle \\
&= -S(C + S)^{-1}\mathrm{Cov}[T] \\
&= -S(C + S)^{-1}X.
\end{aligned}
\tag{5.2.26}
$$

So overall

$$
\mathrm{Cov}[T(\theta, \lambda)] = X - S(C + S)^{-1}X - X(C + S)^{-1}S + \overline{Z}S.
\tag{5.2.27}
$$

- The first term is the fitting uncertainty. This includes contributions from both experiment and theory.

- The last term is the theory uncertainty in the prediction, reduced by a factor $\overline{Z}$ through exposure to the data.

- The middle two terms are correlations between these two uncertainty sources.

We can simplify this expression by noting that

$$
\overline{Z}S = S(C + S)^{-1}X(C + S)^{-1}S + ZS
\tag{5.2.28}
$$

and using the fact that

$$
X - S(C+S)^{-1}X - X(C+S)^{-1}S + S(C+S)^{-1}X(C+S)^{-1}S = C(C+S)^{-1}X(C+S)^{-1}C
\tag{5.2.29}
$$

to see that

$$
\mathrm{Cov}[T(\lambda)] = C(C + S)^{-1}X(C + S)^{-1}C + ZS.
\tag{5.2.30}
$$

Note that in a perfect fit, $X = C + S$ and so we just get C, which is what we ended up with before.

We didn't see the same cancellation as in the perfect fit section, however, because $\mathrm{Cov}[T]$ is X here, rather than $C + S$. So rather than getting the experimental covariance matrix, C, we end up with the sum in quadrature of the fitting uncertainty, X and the theory uncertainty, S, but each reduced due to the effects from correlation.

In the simple model of uncertainties (Eqn. 5.1.26) it can be shown that

$$\mathrm{Cov}[T(\lambda)] = \frac{\sigma^2(\sigma^2 + s^2)}{\sigma^2 + s^2 \sin^2 \phi} \left(nn^T - \frac{s^2}{\sigma^2 + s^2} \cos \phi (en^T + ne^T) + \frac{s^2}{\sigma^2 + s^2} ee^T \right).$$
$$(5.2.31)$$

Here the first term is X, the last term is $\overline{Z}S$ and the middle terms are the correlation.

- If $\phi = 0$, $n = e$ and the result is just $\sigma^2 nn^T$, which is the experimental uncertainty. This is the case of a perfect fit.

- If $\phi = \pi/2$, $n \perp e$ and the correlation terms vanish. Then you end up with $X + ZS$, i.e. the two contributions are added in quadrature, with the theory uncertainty reduced by Z. This is the pure theory case.

From this we can see that the one parameter fit interpolates smoothly between these two extremes.

Note that we can rewrite the covariance as

$$\mathrm{Cov}[T(\lambda)] = \frac{\sigma^2(\sigma^2 + s^2)}{\sigma^2 + s^2 \sin^2 \phi} \left(n - \frac{s^2 \cos \phi}{\sigma^2 + s^2} e \right) \left(n^T - \frac{s^2 \cos \phi}{\sigma^2 + s^2} e^T \right) + \frac{s^2 \sigma^2}{\sigma^2 + s^2} ee^T.$$
$$(5.2.32)$$

In this recasting, we end up with ZS as the last term. We can see that X is altered such that n gets an additional component in the direction of e, due to the correlation with the theory. This:

1. changes the direction;

2. reduces the magnitude of the fitting uncertainty by a factor $\sqrt{\sin^2 \phi + Z^2 \cos^2 \phi}$, although the resulting fitting uncertainty is still larger than it would have been had the theory uncertainty not been included in the fit.

5.2.2 Correlated predictions in single parameter fits

We are now in a position to consider genuine predictions in the case of a single parameter fit. These are denoted $\widetilde{T}_I(\theta)$, $I = 1, \ldots \widetilde{N}$. Note that the predictions, θ, depend on the same parameters as the fitted theory. There are two distinct sources of uncertainty in these predictions:

1. that in the determination of θ, which in turn comes from the experimental uncertainties in D_i and the theory uncertainties in T_i;

2. the theory uncertainties in $\widetilde{T}_I$.

1. are expressed via Eqns. 5.2.7 and 5.2.8. We can linearise the predictions, just like we did for the fitted theory in Eqn. 5.2.1:

$$\widetilde{T}(\theta) = \widetilde{T}_0 + (\theta - \theta_0)\dot{\widetilde{T}}_0 \equiv \widetilde{T}(\theta_0). \tag{5.2.33}$$

Then we can use the same approach as in Eqns. 5.2.9-5.2.12 to find the uncertainty in θ, which is equivalent to the fitting uncertainty, $\widetilde{X}$. Writing $\widetilde{T}^{(r)} \equiv \widetilde{T}(\theta^{(r)})$ and making use of Eqn. 5.2.7:

$$\begin{aligned}
\widetilde{X} &\equiv \mathrm{Cov}[\widetilde{T}(\theta)] \\
&= \langle (\widetilde{T}^{(r)} - \widetilde{T}_0)(\widetilde{T}^{(r)} - \widetilde{T}_0)^T \rangle \\
&= \langle ((\theta^{(r)} - \theta_0)\dot{\widetilde{T}}_0)((\theta^{(r)} - \theta_0)\dot{\widetilde{T}}_0)^T \rangle \\
&= \dot{\widetilde{T}}_0 \langle (\theta^{(r)} - \theta_0)^2 \rangle \dot{\widetilde{T}}_0^T \\
&= \dot{\widetilde{T}}_0 (\dot{T}_0^T (C + S)^{-1} \dot{T}_0)^{-1} \dot{\widetilde{T}}_0^T.
\end{aligned} \tag{5.2.34}$$

2. can be either uncorrelated or correlation with the theory uncertainty in T.

(a) **If it is uncorrelated**, e.g. if they are different types of theory uncertainty, then we can denote the nuisance parameter for the predictions as $\widetilde{\lambda}$. We choose the same prior of a unit Gaussian centred on zero. Then we can write

$$\widetilde{T}(\theta, \widetilde{\lambda}) = \widetilde{T}(\theta) + \widetilde{\lambda}\widetilde{\beta}, \tag{5.2.35}$$

where $\widetilde{\beta}$ gives the direction of the theory uncertainty in $\widetilde{T}$, and $\widetilde{\lambda}$ gives the size. Then we can calculate the expectation and covariance of the

predictions,

$$E[\widetilde{T}(\theta, \widetilde{\lambda})] = \widetilde{T}(\theta_0)$$
$$\mathrm{Cov}[\widetilde{T}(\theta, \widetilde{\lambda})] = \mathrm{Cov}[\widetilde{T}(\theta)] + \mathrm{Var}[\lambda^2]\widetilde{\beta}\widetilde{\beta}^T \qquad (5.2.36)$$
$$= \widetilde{X} + \widetilde{S}.$$

So if the uncertainties are uncorrelated, we add the theory uncertainty in $\widetilde{T}$ in quadrature with the uncertainty in θ from the fit.

(b) **If it is fully correlated**, e.g. factorisation scale variation, $\widetilde{\lambda} = \lambda$, which has non-zero expectation value and variance after the fit. Now

$$\widetilde{T}(\theta, \lambda) = \widetilde{T}_0 + \overline{\lambda}(T_0, D)\widetilde{\beta}, \qquad (5.2.37)$$

so we end up with a similar shift to that of the autopredictions, which is due to the correlation. Explicitly, the shift is

$$\delta\widetilde{T}(\theta_0) = \widetilde{\beta}\beta^T(C + S)^{-1}(D - T_0)$$
$$= -\widehat{S}(C + S)^{-1}(T_0 - D), \qquad (5.2.38)$$

where $\widehat{S} = \widetilde{\beta}\beta^T$ is the cross covariance matrix of the theory uncertainties in the prediction with those in the fitted theory. The covariance of the predictions is then

$$\mathrm{Cov}[\widetilde{T}(\theta, \lambda)] = E[(\widetilde{T}(\theta, \lambda) - E[\widetilde{T}(\theta, \lambda)])^2]$$
$$= E[\widetilde{T}(\theta) + \lambda\widetilde{\beta} - \widetilde{T}_0 - \overline{\lambda}(T_0, D)\widetilde{\beta})^2]. \qquad (5.2.39)$$

First take the expectation value over λ, then over T (using the theory replicas):

$$\mathrm{Cov}[\widetilde{T}(\theta, \lambda)] = E[(\widetilde{T}(\theta) - \widetilde{T}_0) + (\lambda - \overline{\lambda}_0)\widetilde{\beta})^2]$$
$$= \langle(\widetilde{T}^{(r)} - \widetilde{T}_0)(\widetilde{T}^{(r)} - \widetilde{T}_0)^T\rangle + E[(\lambda - \overline{\lambda}_0)\widetilde{\beta}(\widetilde{T} - \widetilde{T}_0)^T]$$
$$+ E[(\widetilde{T} - \widetilde{T}_0)\widetilde{\beta}^T(\lambda - \overline{\lambda}_0)] + E[(\lambda - \overline{\lambda}_0)^2]\widetilde{\beta}\widetilde{\beta}^T.$$
$$(5.2.40)$$

The first term is just $\widetilde{X}$ and the last term is $\mathrm{Var}[\lambda]\widetilde{S} = \overline{Z}\widetilde{S}$. We can evaluate

the middle terms like

$$E[(\lambda - \overline{\lambda}_0)\widetilde{\beta}(\widetilde{T} - \widetilde{T}_0)^T] = \langle \widetilde{\beta}(\overline{\lambda}(T^{(r)}, D) - \overline{\lambda}(T_0, D))(\widetilde{T}^{(r)} - \widetilde{T}_0)^T \rangle$$
$$= -\widehat{S}(C + S)^{-1}\langle (T^{(r)} - T_0)(\widetilde{T}^{(r)} - \widetilde{T}_0)^T \rangle \quad (5.2.41)$$
$$= -\widehat{S}(C + S)^{-1}\widehat{X}^T,$$

where

$$\widehat{X} = \langle (T^{(r)} - T_0)(\widetilde{T}^{(r)} - \widetilde{T}_0)^T \rangle$$
$$= \dot{\widetilde{T}}_0 \langle (\theta^{(r)} - \theta_0)^2 \rangle \dot{\widetilde{T}}_0^T \quad (5.2.42)$$
$$= \dot{\widetilde{T}}_0 (\dot{T}_0^T (C + S)^{-1} \dot{T}_0)^{-1} \dot{T}_0^T.$$

So overall

$$\text{Cov}[\widetilde{T}(\theta, \lambda)] = \widetilde{X} - \widehat{S}(C + S)^{-1}\widehat{X}^T - \widehat{X}(C + S)^{-1}\widehat{S}^T + \overline{Z}\widetilde{S}. \quad (5.2.43)$$

Note that we can write the last term as

$$\overline{Z}\widetilde{S} = Z\widetilde{S} + \widehat{S}(C + S)^{-1}\widehat{X}(C + S)^{-1}\widehat{S}^T$$
$$= \widetilde{S} - \widehat{S}(C + S)^{-1}\widehat{S}^T + \widehat{S}(C + S)^{-1}\widehat{X}(C + S)^{-1}\widehat{S}^T. \quad (5.2.44)$$

Here Z and $\overline{Z}$ are the same as for the autopredictions, and satisfy the same bounds found before (Eqn. 5.2.23). Note that because C is positive definite and S is positive semidefinite, then

$$0 \leq \widehat{S}(C + S)^{-1}\widehat{X}(C + S)^{-1}\widehat{S}^T \leq \widehat{S}(C + S)^{-1}\widehat{S}^T \leq \widetilde{S}, \quad (5.2.45)$$

so importantly the subtraction of $\widehat{S}(C + S)^{-1}\widehat{S}^T$ is never large enough to make the whole covariance matrix negative.

Summary

In summary, including correlations leads to three effects:

1. A shift in central value;

2. A reduction in theory uncertainties;

3. A reduction in fitting uncertainties.

During the fit, information that is implicit in the data about the theory is propoagated via the corrleations. This leads to more precise (and hopefully more accurate) predictions.

5.3 Correlated MHOUs in PDF fits

In this section we add another layer of complexity to the model we are building up. Now the theory values, $T_i[f]$, depend on PDFs, f, which are determined in a global fit to the N data points, D_i, with experimental covariance C_{ij}. The PDFs are then used to make $\widetilde{N}$ theory predictions, $\widetilde{T}_I[f]$.

In a PDF fit there are many potential sources of theory uncertainty, but here we will consider MHOUs, computed with scale variation using a prescription from Chapter 3. In this case S_{ij} and $\widetilde{S}_{IJ}$ have many non-zero eigenvalues. We can describe them using n nuisance parameters, λ_α, $\alpha = 1, \ldots, n$. Usually $n \ll N$, but we don't impose a limit on n here.

We will now find the expectation value and covariance of these nuisance parameters, and use those to find the shifts in predictions, and the change in their uncertainties.

5.3.1 Multiple nuisance parameters

Here we repeat the analysis of Sec. 5.1.1, but for multiple nuisance parameters rather than just one. Each nuisance parameter is associated with a shift in theory value $T_i[f] \to T_i[f] + \lambda_\alpha \beta_{i,\alpha}[f]$, using summation notation for α. Note that the $\beta_{i,\alpha}$ don't have to be mutually orthogonal. We pick a prior for each nuisance parameter the same as in Sec. 5.1.1, i.e.

$$P(\lambda|D) = P(\lambda) \propto \exp\left(-\tfrac{1}{2}\lambda_\alpha \lambda_\alpha\right). \tag{5.3.1}$$

Once again we assume Gaussianity, and now instead of Eqn. 5.1.5 we get

$$P(T|D\lambda) \propto \exp\left(-\tfrac{1}{2}(T[f] + \lambda_\alpha \beta_\alpha - D)^T C^{-1}(T[f] + \lambda_\alpha \beta_\alpha - D)\right). \tag{5.3.2}$$

We can marginalise over λ_α to get

$$P(T|D) \propto \int d^n\lambda \, \exp\left(-\tfrac{1}{2}[(T[f] + \lambda_\alpha\beta_\alpha - D)^T C^{-1}(T[f] + \lambda_\beta\beta_\beta - D) + \delta_{\alpha\beta}\lambda_\alpha\lambda_\beta]\right) \, . \tag{5.3.3}$$

The next step is to complete the square in the exponent. After some work, defining

$$Z_{\alpha\beta} = (\delta_{\alpha\beta} + \beta_\alpha^T C^{-1}\beta_\beta)^{-1}, \tag{5.3.4}$$

where the bracketed inverse is with respect to α, β, and

$$\overline{\lambda}_\alpha = Z_{\alpha\beta}\beta_\beta^T C^{-1}(D - T), \tag{5.3.5}$$

we end up with

$$P(T|D) \propto \int d^n\lambda \, \exp\left(-\tfrac{1}{2}(\lambda_\alpha - \overline{\lambda}_\alpha)Z_{\alpha\beta}^{-1}(\lambda_\beta - \overline{\lambda}_\beta) - \tfrac{1}{2}\chi^2\right) \propto \exp(-\tfrac{1}{2}\chi^2). \tag{5.3.6}$$

Note that χ^2 in this expression is given by Eqn. 5.1.2 but where $S = \beta_\alpha\beta_\alpha^T$. Note also the analogy between this and Eqn. 5.1.13.

We can then use Bayes' Theorem to get the posterior distribution,

$$P(\lambda|TD) \propto \exp\left(-\tfrac{1}{2}(\lambda_\alpha - \overline{\lambda}_\alpha)Z_{\alpha\beta}^{-1}(\lambda_\beta - \overline{\lambda}_\beta)\right), \tag{5.3.7}$$

and from this the expectation and covariance of λ_α are

$$E[\lambda_\alpha] = \overline{\lambda}_\alpha, \qquad E[(\lambda_\alpha - \overline{\lambda}_\alpha)(\lambda_\beta - \overline{\lambda}_\beta)] = Z_{\alpha\beta}. \tag{5.3.8}$$

If we write $\beta = e_\alpha\beta_\alpha$, such that e_α is a unit eigenvector of $Z_{\alpha\beta}$, then the corresponding eigenvalue is $z = (1 + \beta^T C^{-1}\beta)^{-1}$. This means $0 < z < 1$ and so $Z_{\alpha\beta}$ is positive definite, therefore invertible. What's more, if all $z < 1$ then $\delta_{\alpha\beta} - Z_{\alpha\beta}$ is also positive definite. We can view $Z_{\alpha\beta}$ as the matrix version of the coefficient from Sec. 5.1.1 as there are now many sources of uncertainty.

We can write, in analogy with before,

$$Z_{\alpha\beta} = \delta_{\alpha\beta} - \beta_\alpha^T(C + S)^{-1}\beta_\beta, \tag{5.3.9}$$

which allows us to rewrite $\overline{\lambda}_\alpha$, using $(1 - (C + S)^{-1}S)C^{-1} = (C + S)^{-1}$, as

$$\overline{\lambda}_\alpha = \beta_\alpha^T(C + S)^{-1}(D - T[f]). \tag{5.3.10}$$

5.3.2 Fitting PDFs with fixed parametrisation

Now we can use the previous section's results in the context of a PDF fit with MHOUs. In this section, we consider a fixed parametrisation of PDFs, like that adopted by, for example MSHT, CTEQ and ABM. Here the PDFs, $f(\theta)$, depend on m parameters, θ_p, $p = 1, \ldots, m$, where $m < N$ such that the data are able to determine all the parameters through χ^2 minimisation. We will move on to the somewhat different case of PDFs with neural networks (unsurprisingly relevant to NNPDF) in the next section.

We adopt the same approach as in Sec. 5.2, but fitting m parameters, θ_p, rather than a single one, θ. Writing the PDF that minimises the χ^2 as $f(\theta^0) \equiv f_0$, and using the notation $T_0 \equiv T(\theta^0)$, $T_p \equiv \partial T(\theta^0)/\partial \theta_p^0$ with summation convention for p, we can linearise $T(\theta)$ as

$$T(\theta) = T_0 + (\theta_p - \theta_p^0)T_p + \ldots. \tag{5.3.11}$$

Minimising this with respect to θ^p, we find

$$\theta_p^{(r)} - \theta_p^0 = (T_p^T(C+S)^{-1}T_q)^{-1}T_q^T(C+S)^{-1}(D^{(r)} - D), \tag{5.3.12}$$

where the inverse of the left section is with respect to p, q. So

$$\begin{aligned}
\text{Cov}_{pq}[\theta] &= \langle (\theta_p^{(r)} - \theta_p^0)(\theta_q^{(r)} - \theta_q^0) \rangle \\
&= (T_p^T(C+S)^{-1}T_q)^{-1}.
\end{aligned} \tag{5.3.13}$$

To find X, we can separate out the magnitude and direction of T_p by writing $T_p = |T_p|n_p$, where n_p are (not necessarily orthogonal) unit vectors. Then Eqn. 5.2.12 becomes

$$X \equiv \text{Cov}[T[f]] = n_p(n_p^T(C+S)^{-1}n_q)^{-1}n_q^T. \tag{5.3.14}$$

Note that $X(C+S)^{-1}$ still projects the data replicas onto the theory replicas, and the projective relation for X still holds.

Autopredictions

First consider the autopredictions, $T(f, \lambda) \equiv T[f] + \lambda_\alpha \beta_\alpha$. We can see that the results from Sec. 5.2 still hold. In particular, the central values of $\overline{\lambda}_\alpha$ are given

by

$$\mathrm{E}[\lambda_\alpha] = -\beta_\alpha^T (C+S)^{-1}(T[f_0] - D), \qquad (5.3.15)$$

and the shifts (Eqn. 5.2.21) are now

$$\delta T[f] = \beta_\alpha \beta_\alpha^T (C+S)^{-1}(D - T[f_0]) = -S(C+S)^{-1}(T[f_0] - D). \qquad (5.3.16)$$

These shifts will improve the χ^2 to experimental data, just like those in Sec. 5.1. The covariance of λ becomes an equation for a matrix rather than a coefficient:

$$\begin{aligned}
\mathrm{Cov}_{\alpha\beta}[\lambda] &= E[(\lambda_\alpha - E[\lambda_\alpha])(\lambda_\beta - E[\lambda_\beta])] \\
&= \delta_{\alpha\beta} - \beta_\alpha^T (C+S)^{-1}\beta_\beta - \beta_\alpha^T (C+S)^{-1} X (C+S)^{-1}\beta_\beta \equiv \overline{Z}_{\alpha\beta}.
\end{aligned}$$

$$(5.3.17)$$

Like before, both $\overline{Z}_{\alpha\beta}$ and $\delta_{\alpha\beta} - \overline{Z}_{\alpha\beta}$ are positive semidefinite, so

$$0 < Z_{\alpha\beta} \le \overline{Z}_{\alpha\beta} \le \delta_{\alpha\beta}. \qquad (5.3.18)$$

The covariance of the autopredictions then becomes

$$\begin{aligned}
\mathrm{Cov}[T(f,\lambda)] &= X - S(C+S)^{-1}X - X(C+S)^{-1}S + \beta_\alpha \overline{Z}_{\alpha\beta}\beta_\beta^T \\
&= C(C+S)^{-1}X(C+S)^{-1}C + S - S(C+S)^{-1}S.
\end{aligned}$$

$$(5.3.19)$$

Note that this result is identical to that in Sec. 5.2.

Predictions for new observables

Now consider true predictions for new observables. The shifts (Eqn. 5.1.33) can be written

$$\delta \widetilde{T}[f] = \widetilde{\beta}_\alpha \beta_\alpha^T (C+S)^{-1}(D - T[f_0]) = -\widehat{S}(C+S)^{-1}(T[f_0] - D), \qquad (5.3.20)$$

where we have defined $\widehat{S} = \widetilde{\beta}_\alpha \beta_\alpha^T$.

If the predictions are fully correlated, then $\widetilde{T}(f,\lambda) = \widetilde{T}[f] + \lambda_\alpha \widetilde{\beta}_\alpha$ and

$$\mathrm{Cov}[\widetilde{T}(f,\lambda)] = \widetilde{X} - \widehat{S}(C+S)^{-1}\widehat{X}^T - \widehat{X}(C+S)^{-1}\widehat{S}^T + \widetilde{\beta}_\alpha \overline{Z}_{\alpha\beta}\widetilde{\beta}_\beta^T, \qquad (5.3.21)$$

where $\widetilde{S} = \widetilde{\beta}_\alpha \widetilde{\beta}_\alpha^T$ and we have:

$$
\begin{aligned}
\widetilde{X} &= \widetilde{T}_p (T_p^T (C + S)^{-1} T_q)^{-1} \widetilde{T}_q^T; \\
\widehat{X} &= \widetilde{T}_p (T_p^T (C + S)^{-1} T_q)^{-1} T_q^T,
\end{aligned}
\tag{5.3.22}
$$

Using Eqn. 5.3.19, we can therefore write the last term as

$$
\begin{aligned}
\overline{Z} \widetilde{S} &= Z \widetilde{S} + \widehat{S} (C + S)^{-1} X (C + S)^{-1} \widehat{S}^T, \tag{5.3.23} \\
Z \widetilde{S} &= \widetilde{S} - \widehat{S} (C + S)^{-1} \widehat{S}^T. \tag{5.3.24}
\end{aligned}
$$

So we find that once again the result is exactly the same as in Sec. 5.3. In other words, once we have eliminated the nuisance parameters, the only difference in generalising one parameter to many is to amend the expressions for X, $\widetilde{X}$ and $\widehat{X}$.

5.3.3 Fitting NNPDFs

In NNPDF we don't use a fixed parametrisation, but instead have a neural network with a very large number of parameters, in general greater than the number of data points. Here the ability to overfit is a danger, so we adopt a cross-validation procedure when finding the optimal χ^2 (see Chapter 1). This means that when fitting each data replica, $D^{(r)}$, the χ^2 is not exactly minimised; there is random noise in the system which amounts to a "functional uncertainty" [56].

The fact that Eqn. 5.2.4 is not exactly minimised makes the analytical approach imprecise, and while in general the results in Sec. 5.3.1 are valid, the exact relations for the fitted parameters (e.g. Eqn. 5.3.12) and subsequent results do not hold. We are still able to use the PDF replicas compute, for example

$$
X \equiv \mathrm{Cov}[T[f]] = \langle (T^{(r)} - T^{(0)})(T^{(r)} - T^{(0)})^T \rangle, \tag{5.3.25}
$$

where $T^{(r)} \equiv T[f^{(r)}]$, and $T^{(0)} \equiv \langle T^{(r)} \rangle$. However, the projective relation for X is no longer satisfied, and now $X(C + S)^{-1}$ doesn't project the data replicas on to the theory replicas. We can confirm this for a given PDF by computing the matrix

$$
Y \equiv \mathrm{Cov}[T, D] = \langle (T^{(r)} - T^{(0)})(D^{(r)} - D)^T \rangle. \tag{5.3.26}
$$

For a fixed parametrisation, combining Eqn. 5.2.14 and Eqn. 5.2.3 shows that $Y = X = Y^T$. But explicit computation in an NNPDF fit shows us that Y is

generally considerably smaller than X, because the fluctuations of theory replicas are poorly correlated to those of the data replicas. This is despite the fluctuations of the data replicas being about an order of magnitude greater than that of the theory replicas. Although many $X(C+S)^{-1}$ eigenvalues will be zero (because $m < N$), a lot of the non-zero eigenvalues will be larger than one as a result of functional uncertainty. So although $\delta_{\alpha\beta} - Z_{\alpha\beta}$ is still positive definite, the upper bound on $\overline{Z}_{\alpha\beta}$, Eqn. 5.3.18, is no longer true; the covariance of the nuisance parameters can be greater than the prior if there is a large functional uncertainty.

Note that X is not invertible, but this is not a technical limitation. The mapping of the global dataset onto the PDFs is not invertible (excepting certain cases, for example the data from a single process at a single scale explored in [146]). This is because you can't recover the full data from the PDFs alone, which is in part because PDFs are only functions of x, but the data also depend on the scale; when PDFs are delivered, the data are all projected onto a common PDF scale. Additionally, the PDFs are universal and therefore process independent, so you can't say which processes were used to get the final PDFs.

Expectation and covariance of autopredictions

Once again we can consider autopredictions, now in a realistic NNPDF scenario. We find that the shifts are given by a similar expression to Eqn. 5.3.16,

$$\delta T[f] = -S(C+S)^{-1}(T^{(0)} - D), \tag{5.3.27}$$

and will once again reduce the experimental χ^2. The covariance of autopredictions is also still given by Eqn. 5.3.19:

$$P \equiv \mathrm{Cov}[T(f,\lambda)] = C(C+S)^{-1}X(C+S)^{-1}C + (S - S(C+S)^{-1}S). \tag{5.3.28}$$

If the theory uncertainty S is much smaller than the experimental uncertainty C, P approaches the result

$$P_{\mathrm{con}} = X + S; \tag{5.3.29}$$

the fitting uncertainty and theoretical uncertainty can be combined in quadrature, and the 'conservative' prescription recommended in [4] is a useful approximation.

More generally, we can think of the two contributions to P being the correlated PDF uncertainty and the correlated theory uncertainty. Because $C > 0$, and

$S \geq 0$, $X \geq 0$, both of these contributions are positive semidefinite. Additionally, the correlated theory uncertainty is bounded from above by the uncorrelated theory uncertainty:

$$0 \leq S - S(C + S)^{-1}S = C(C + S)^{-1}S \leq S. \tag{5.3.30}$$

At first glance it might appear that the correlated PDF uncertainty will also be bounded from above by the uncorrelated PDF uncertainty X. One might think this because since C is positive definite, and S positive semi-definite, $C \leq C + S$, so $C(C + S)^{-1} \leq 1$, and $C(C + S)^{-1}X(C + S)^{-1}C \leq X$. This argument is wrong, however, and the correlated PDF uncertainty can sometimes exceed the uncorrelated. Writing

$$
\begin{aligned}
C(C + S)^{-1}X(C + S)^{-1}C = {} & X - S(C + S)^{-1}X - X(C + S)^{-1}S \\
& + S(C + S)^{-1}X(C + S)^{-1}S,
\end{aligned}
\tag{5.3.31}
$$

in some circumstances the sum of the last three terms can be positive. For this reason it seems impossible to prove in general that $P \leq P_{\mathrm{con}}$, though in all practical applications we have tested so far this seems to be the case.

Genuine predictions

For genuine predictions, where the theory uncertainties in the predictions are correlated with those in the fit theory, the shifts are given by Eqn. 5.3.20,

$$\delta \widetilde{T}[f] = -\widehat{S}(C + S)^{-1}(T^{(0)} - D), \tag{5.3.32}$$

and the uncertainties are given by Eqn. 5.3.21, which are most usefully written as

$$
\begin{aligned}
\widetilde{P} \equiv {} & \mathrm{Cov}[\widetilde{T}(f, \lambda)] \\
= {} & \widetilde{X} - \widehat{X}(C + S)^{-1}\widehat{S}^T - \widehat{S}(C + S)^{-1}\widehat{X}^T + \widehat{S}(C + S)^{-1}X(C + S)^{-1}\widehat{S}^T \\
& + (\widetilde{S} - \widehat{S}(C + S)^{-1}\widehat{S}^T).
\end{aligned}
\tag{5.3.33}
$$

The second line are contributions to the correlated PDF uncertainty, and the third line are contributions to the correlated theory uncertainty. Note that we

also need to calculate

$$\widetilde{X} \equiv \mathrm{Cov}[\widetilde{T}[f, \lambda]] = \langle (\widetilde{T}^{(r)} - \widetilde{T}^{(0)})(\widetilde{T}^{(r)} - \widetilde{T}^{(0)})^T \rangle, \qquad (5.3.34)$$

$$\widehat{X} \equiv \mathrm{Cov}[\widetilde{T}[f, \lambda], T[f, \lambda]] = \langle (\widetilde{T}^{(r)} - \widetilde{T}^{(0)})(T^{(r)} - T^{(0)})^T \rangle. \quad (5.3.35)$$

When $\widehat{S}$ is very small, we end up with the conservative result. This will be the case for predictions of new processes, where the dominant MHOU is in the hard cross section. However, for existing processes, $\widetilde{S}$ and $\widehat{S}$ can be large, even if S is small.

5.4 Numerical results

In this section we will apply all the results we worked up to in Sec. 5.3. We have seen that in a realistic NNPDF fit, we just use the same analytic expressions (Eqns. 5.3.27, 5.3.28, 5.3.32, 5.3.33) as we would for a fit of a PDF with fixed parametrisation. This holds true despite the fact that PDFs possess a "functional uncertainty", due to the fact that PDF parameters are not uniquely fixed by the fit. All the additional information we need to find the correlated predictions and uncertainties are the matrices X, $\widetilde{X}$ and $\widehat{X}$, which we compute by taking the ensemble average over the PDF replicas from the fit. In this section we will compute these matrices for an NNPDF global fit with theory uncertainties, and use them to compute and predict both autopredictions and genuine predictions, which include the effect of correlated theory uncertainties. We use as a baseline the NNPDF3.1 NLO global fit with 9 point MHOUs which was generated in Chapter 3. This includes 2819 data points, split up into 5 processes. We show again the experimental and theory covariance matrices from this fit in Fig. 5.4.1 for reference.

5.4.1 Covariance of PDF uncertainties X

The first thing to do to calculate predictions is to compute X_{ij} (Eqn. 5.3.25), which is shown in Fig. 5.4.2 as a heat map alongside its corresponding correlation matrix. The off-diagonals of X are almost as large as the diagonals.
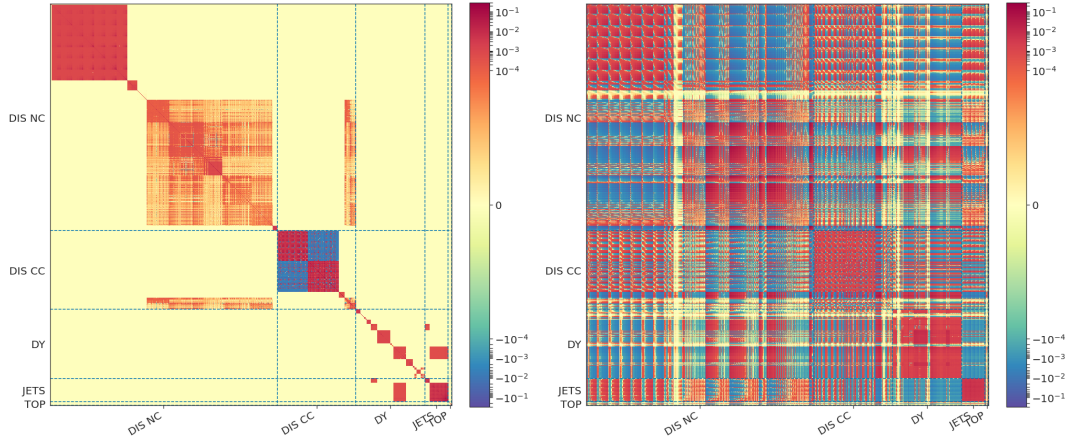
Figure 5.4.1 *The experimental covariance matrix, C_{ij}, normalised to the theoretical predictions $T_i^{(0)}$ (left), and the corresponding theory covariance matrix for MHOU, S_{ij} (right). The datasets are arranged in the order given in Fig. 5.4.7 below: so SLAC data are in the top left corner, and LHC top data in the lower right corner.*
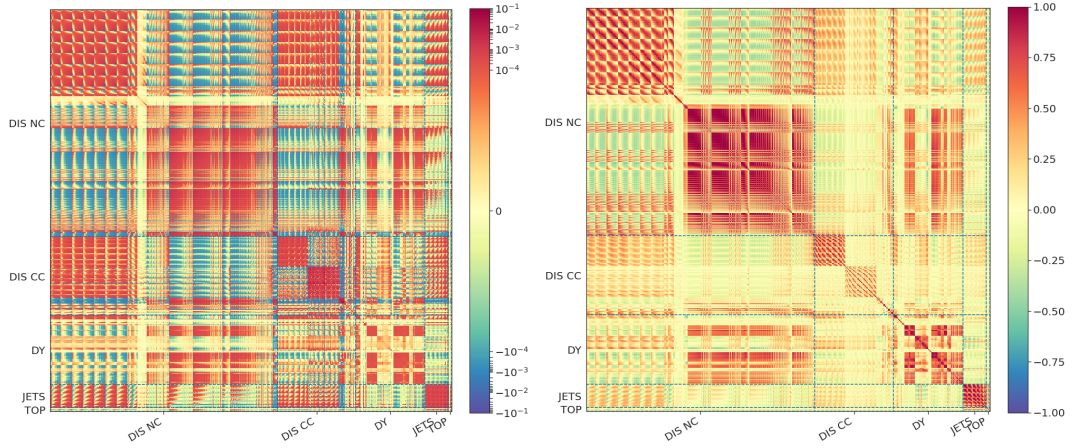
Figure 5.4.2 *The covariance matrix of PDF uncertainties, X_{ij}, normalised to the theoretical predictions $T_i^{(0)}$ (left), and the corresponding correlation matrix $X_{ij}/\sqrt{X_{ii}X_{jj}}$ (right). The datasets are arranged in the order given in Fig. 5.4.7 below: so SLAC data are in the top left corner, and LHC top data in the lower right corner.*

This is because theory predictions are often very strongly correlated, not only for nearby bins within the same experiment but also for different processes at nearby scales. This is primarily due to the smoothness of the underlying PDFs in (x, Q^2), but it is also a consequence of the highly correlated theory uncertainties included in the fit.

129
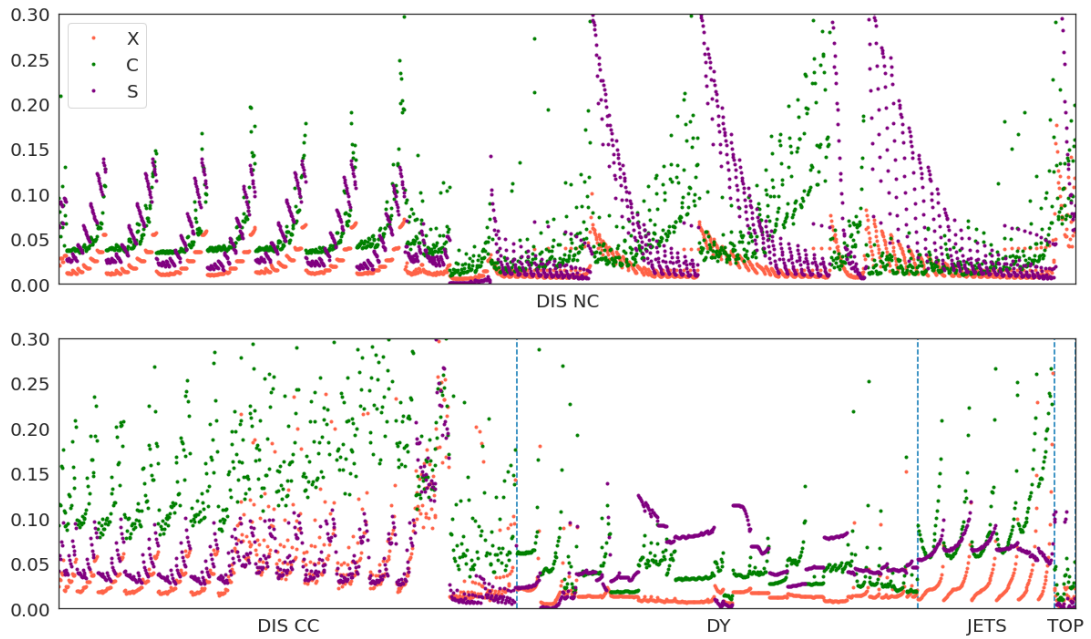
Figure 5.4.3 *The square root of the diagonal elements of the matrices X (in orange), C (in green) and S (in purple) normalised to the theoretical predictions $T_i^{(0)}$, with those for C and S the same as in Chapter 3. The datasets are arranged in the order given in Fig. 5.4.7 below.*

We compare PDF uncertainties to experimental and theory uncertainties by looking at the per-point uncertainty in Fig. 5.4.3. Recall from Eqn. 5.2.3 that $C + S$ is the covariance of the data replicas to which the PDFs are fitted. At NLO, the relative size of C_{ii} and S_{ii} can vary quite alot between datasets; for fixed target DIS, S_{ii} is generally greater than C_{ii}, except at large x. For HERA NC there is an interesting pattern whereby $S_{ii} \ll C_{ii}$ at large x, but the opposite is true at small x. The experimental uncertainty is also dominant for CHORUS, but for most of DY the theory uncertainty dominates.

The PDF uncertainties, X_{ii} are generally less than both the experimental and theoretical uncertainties. This makes sense because they are the product of a fit, and so the uncertainty on each point is influenced by all the other data points in the fit, which collectively conspire to reduce the uncertainty. We can clearly see this effect in DY and JETS. The exception to this are datasets with very small theory uncertainty, for example ratio datasets where systematic uncertainties cancel between the numerator and denominator (e.g. NMC d/p, asymmetry data and differential top). In these instances, X_{ii} is above S_{ii}, though still lower than C_{ii}.

130

5.4.2   Nuisance parameters

Now let's look at the nuisance parameters λ_α of the covariance matrix S. We showed in Chapter 3 that for five processes in the 9 point prescription there are 28 non-zero eigenvalues, and so we will have 28 nuisance parameters. Fig. 5.4.4 shows these eigenvalues in descending order in the top panel, with their nuisance parameters below. We computed the expectation value of the nuisance parameters using Eqn. 5.3.16, and their uncertainties using Eqn. 5.3.17. It helps to separate out the two contributions to the uncertainty on the nuisance parameters, namely:

1. the scale uncertainty (Eqn. 5.3.8);

2. the PDF uncertainty (last term in Eqn. 5.3.17).

These are shown as the lower two panels in Fig. 5.4.4.

Recall that the prior for the nuisance parameters was a unit Gaussian centred on zero (Eqn. 5.3.1). After fitting, we see that the total uncertainty in the nuisance parameters for the largest $\sim$9 eigenvalues has been substantially reduced, showing that the MHOU along these directions has been learnt in the fitting process. For the nuisance parameters corresponding to the smaller eigenvalues there is very little reduction. This tells us that the data don't constrain these directions very well. The central values for the largest three eigenvalues are very close to zero within uncertainties. This shows that the choice of prior was reasonable. The next three or so significantly deviate from zero, showing that the data must have significant information about the MHOUs in these directions. For the remaining smaller eigenvalues, the central values of the nuisance parameters are all compatible with zero. For the very small ones it seems the data have had no effect at all because the posterior distributions are the same as the prior. So only the largest eigenvalues are actually relevant for PDF determination, and the others are simply ignored by the fit.

Looking now at the split between scale uncertainty and PDF uncertainty, we see that the MHOU for the largest eigenvalues is reduced a lot, showing that it is learnt in the same way we saw in the simple models of Secs 5.1-5.3. However, there is very little information extracted about the smaller ones. The PDF uncertainty is also small for the largest and smallest eigenvalues, but dominates for the middle ones.
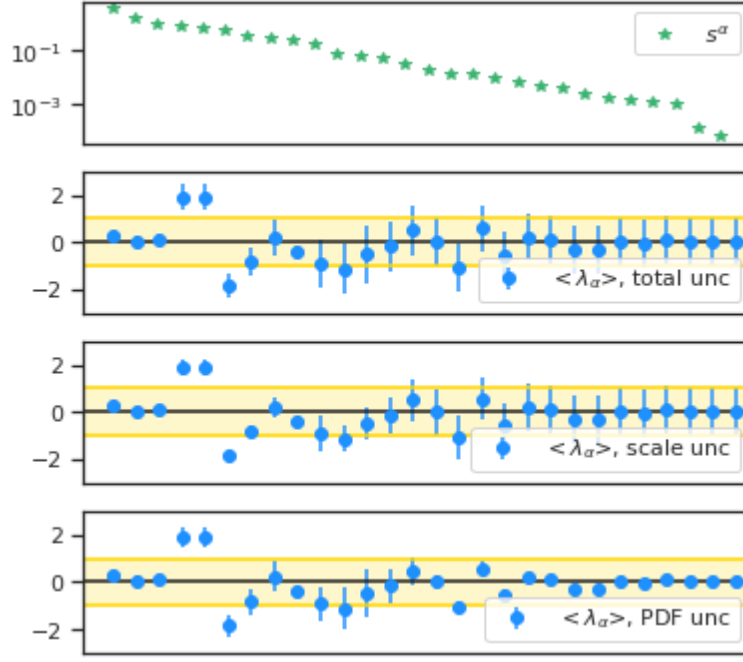
Figure 5.4.4 *The 28 positive eigenvalues s^α of the theory uncertainty matrix S_{ij} (above), shown in descending order, and 28 nuisance parameters λ_α corresponding to the 28 eigenvectors β_α (below), as given by Eqn. 5.3.15. The uncertainties in the nuisance parameters are shown in total (square roots of the diagonal entries of Eqn. 5.3.17, and broken down into the contribution from scale uncertainties alone (square roots of the diagonal entries of Eqn. 5.3.8 and from PDF uncertainties (square roots of the diagonal entries of the last term in Eqn. 5.3.17. The yellow bands highlight the region between ± 1.*
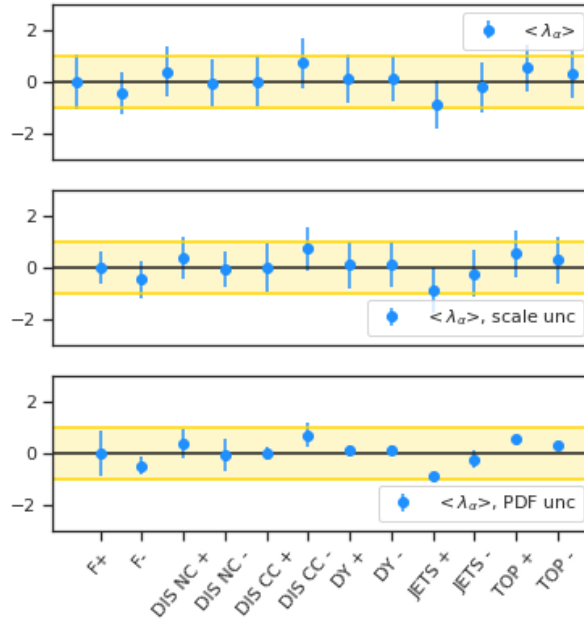
Figure 5.4.5 *Nuisance parameters λ for directions in the space of scale variations corresponding to up/down changes in factorization scale, and in renormalization scale for the five types of processes in the determination of the 9-pt theory covariance matrix for MHOU. The uncertainties in the nuisance parameters are shown in total, and broken down into the contribution from scale uncertainties alone and from PDF uncertainties, just as in Fig. 5.4.4. The yellow bands highlight the region between ± 1.*

Having seen that some directions of MHOU are more relevant in the fit, we may wonder whether these have a physical interpretation; in Fig. 3.5.5 of Chapter 3 we saw that the largest eigenvectors of S were driven by factorisation scale variation and then renormalisation scale variation for DIS NC. To investigate this, we can switch bases, choosing β_α to correspond to factorisation scale variations (up/down) and renormalisation scale variations (up/down for each process). Fig. 5.4.5 is a similar plot to Fig. 5.4.4, but for this "physical" basis.

We see that the central values fluctuate about zero, but stay in the ± 1 band, showing again that the impact of fitting the data on the nuisance parameters is not large. This is reassuring as it backs up the choice of central scales and the choice of range of scale variations (the latter being implicit in the prior for λ_α). Looking just at the scale uncertainty, it is apparent that the factorisation scale variation nuisance parameters have learnt the most information, which makes sense as factorisation scale variation is common to all data in the fit. NC DIS, being the largest process, is also learnt about to some extent. However, including the PDF uncertainty washes out these effects. In particular, the uncertainties in these directions can be slightly greater than one in total; in fact there is less learnt about the factorisation scale nuisance parameters that we had supposed in the prior.

Already we see that information from the data in the fit significantly updates the priors for the nuisance parameter distribution. From this it is likely that there will be an effect at the level of the autopredictions, which is the subject of the next section.

5.4.3 Autopredictions

As described in Sec. 5.1, autopredictions are where we fit a PDF and then use that PDF to make predictions for the data that went into the PDF. These are essentially postdictions, and are ideal for testing the extent of correlation between theory uncertainties in the PDF fit and in the (auto)predictions.

Although this situation is somewhat artificial because experiments are never redone in exactly the same way, the implications of this investigation will be general. This is because for a global fit of this size (2819 data points, 35 datasets and 5 processes), removing only one of the smaller datasets has negligible impact on the PDFs.
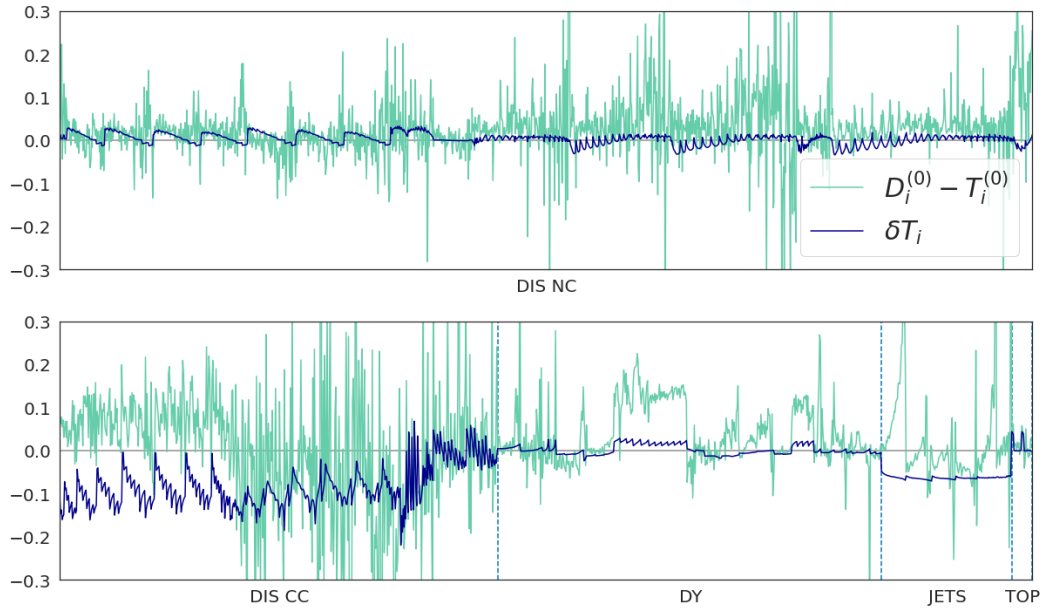
Figure 5.4.6   *The shifts δT_i, Eqn. 5.3.27 (in blue) compared to the differences between theory and data, $D_i - T_i^{(0)}$ (in green), both normalised to $T_i^{(0)}$.*

Removing even a large dataset will only increase PDF uncertainties without affecting the theory uncertainties for the remaining data. This means that if we did the fit with a certain dataset removed, and did the analysis with that fit instead, we would instead have a genuine prediction, and the correlations between MHOUs in the PDF and the prediction would be very close to what we have for the autopredictions.

To make the correlated autopredictions, we first compute δT_i (Eqn. 5.3.16). This is the shift in theory predictions arising due to theory correlations. We show this in Fig. 5.4.6, normalised to the orginal theoretical prediction $T_i^{(0)}$. We also show $D_i - T_i^{(0)}$ for comparison. The shifts tend to be small, however for some datasets (especially CHORUS and inclusive jets) there is a systematic overall shift of order $D_i - T_i^{(0)}$.

It's interesting to see whether the shifts improve the autopredictions. In Fig. 5.4.7 we show the χ^2 of the autopredictions, using the experimental covariance matrix only, for the following autoprediction central values:

- No theory uncertainty;

- Theory uncertainty in the fit;

- Shifted autopredictions.
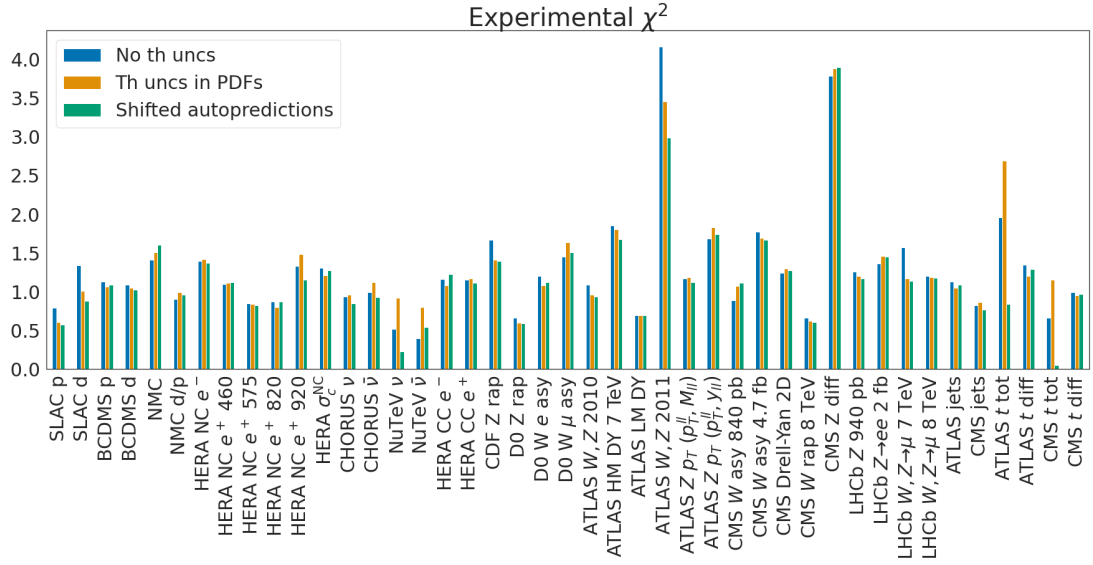
Figure 5.4.7 *The experimental χ^2 for each data set, comparing the original result of the NLO fit with no theory uncertainties to the fit with theory uncertainties, and then including the correlated shift in the autopredictions.*

We use only the experimental covariance matrix, in order to isolate the effects due to the changing central value from those due to adding additional uncertainties.

The results for all cases are very similar. Including the theory uncertainty in the fit has mixed results; some predictions get better at the expense of others getting worse. This is because the main effect of including theory uncertainties is to rebalance the fit. When the correlated shift is also included, the fit to most datasets improves, in some cases substantially. The exact values are broken down by process in Tab. 5.4.1. When including theory uncertainties, the χ^2 goes up

	DIS NC	DIS CC	DY	JETS	TOP	Total
No th uncs	1.13	0.98	1.56	0.88	1.20	1.17
Uncorr th uncs	1.15	1.06	1.53	0.90	1.27	1.19
Correlated th uncs	1.09	0.91	1.47	0.83	0.97	1.10

Table 5.4.1 *The experimental χ^2 per data point for each process, comparing the original result of the NLO fit with no theory uncertainties to the fit with theory uncertainties, and then including the shift in the autopredictions.*

slightly from 1.17 to 1.19. However, when the correlated shift is added, there is a significant improvement across all processes, with the total χ^2 dropping to 1.10.
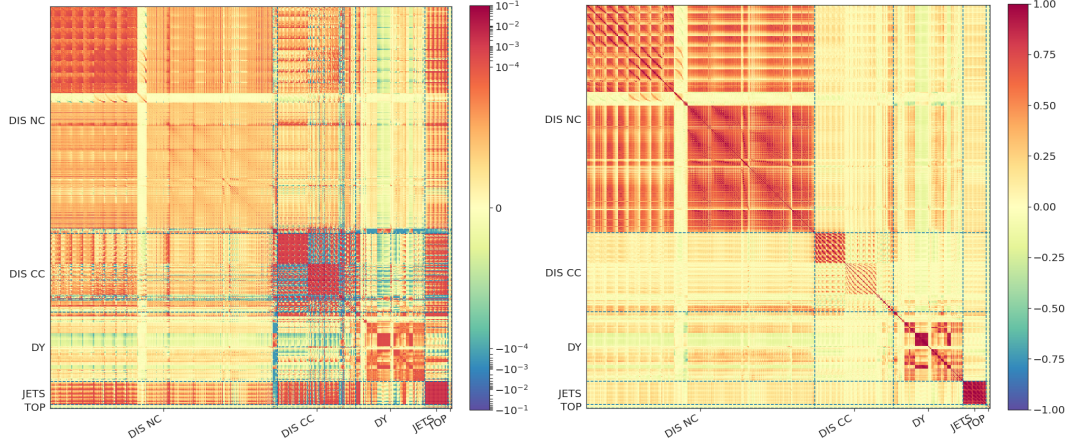
Figure 5.4.8   *The autoprediction covariance matrix P_{ij} Eqn. 5.3.28, normalised to the theoretical predictions $T_i^{(0)}$ (left), and the corresponding corrrelation matrix $P_{ij}/\sqrt{P_{ii}P_{jj}}$ (right).*
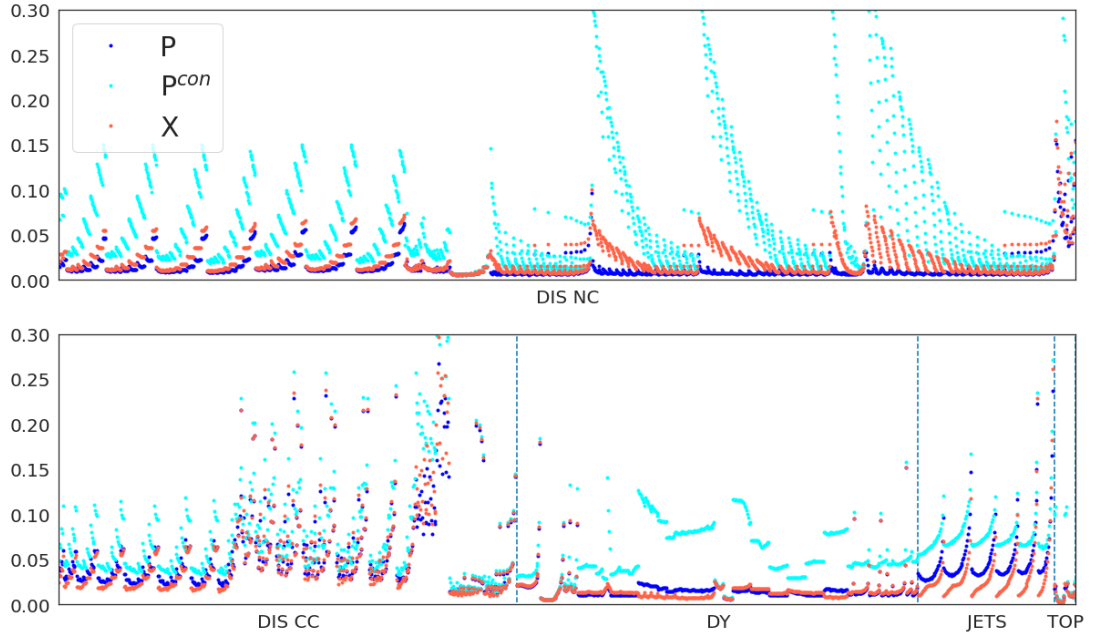
Figure 5.4.9   *The percentage uncertainties of the autopredictions $\sqrt{P_{ii}}$ Eqn. 5.3.28 (cyan) compared to the PDF uncertainty $\sqrt{X_{ii}}$ (orange), and the conservative result, $\sqrt{P_{ii}^{\mathrm{con}}}$ Eqn. 5.3.29 (dark blue), all normalised to the theoretical predictions $T_i^{(0)}$.*

Now let's see whether we also end up with an increase in precision, by considering the uncertainties in the autopredictions. In Fig. 5.4.8 we show the full covariance matrix of autopredictions normalised to theory predictions and also as a correlation matrix. Remember that P_{ij} is the sum of:

1. The PDF uncertainty, derived from a combination of the experimental and theory uncertainties in the fit;

2. The theory uncertainty in the autoprediction.

Each of these contributions is reduced due to the learning of the theory uncertainties and the correlation between the two sources of theory uncertainty. As might be expected, we can see that there are very large correlations in the autopredictions within datasets. These are due to:

1. Correlation of experimental uncertainties within datasets;

2. Smooth underlying PDFs;

3. Correlations of theory uncertainties.

Correlations are generally larger within each process than outwith them. This suggests that the factorisation scale correlation is small compared to the combination of the renormalisation scale correlation and the effects from PDF smoothness and experimental uncertainties.

Fig. 5.4.9 shows the percentage uncertainties of the autopredictions, $\sqrt{P_{ii}}/T_i$, compared to the original PDF uncertainties, $\sqrt{X_{ii}}/T_i$. It also shows the percentage uncertainties for the conservative prescription, $\sqrt{P_{ii}^{con}}/T_i$. The correlated autoprediction uncertainties are generally a similar size to the PDF uncertainties. They are larger for some of the DY datasets and JETS, and for most of DIS NC and some DY they are smaller. This is in stark contrast to the uncorrelated conservative prescription, which are greater than the PDF uncertainties across the board, sometimes by a lot, and typically a factor of two or more. This is because they don't take into account the correlation or the learning, which leads to an overestimate, especially where the theory uncertainties are a lot bigger than the PDF uncertainty. In fact, the conservative prescription only works well where the theory uncertainties are very small, for example the NMC d/p ratio data.

The upshot is that the correlated autopredictions are not only more accurate, they are also more precise. But we do need to be wary of this increase in precision because it depends implicitly on the assumptions made when modelling the prior MHOUs that we made in Chapter 3. In particular, it is dependent on the choice of independent scales, the size of variation, and the prescription for generating

137

S. For example, the aggressive reduction in small x uncertainty for HERA NC may well be due to the unseparated singlet and non-singlet factorisation scale variation; because of this the singlet evolution is overconstrained at small x [146]. We leave this as a matter for future work.
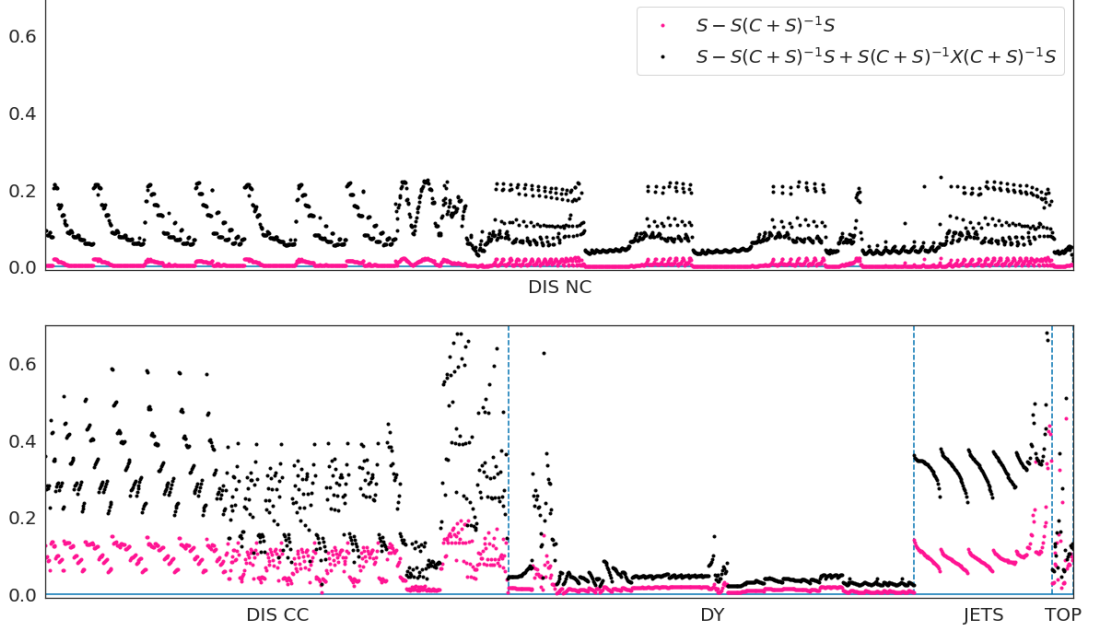
Figure 5.4.10 *The contributions to the diagonal elements of the correlated theory uncertainty normalised to diagonal elements of S: $(S - S(C + S)^{-1}S)_{ii}/S_{ii}$ (pink), and $(S - S(C + S)^{-1}S + S(C + S)^{-1}X(C + S)^{-1}S)_{ii}/S_{ii}$ (black).*

To understand better how these changes in uncertainty arise, we show in Fig. 5.4.10 a breakdown of the diagonal elements of the correlated theory uncertainty (the second term in Eqn. 5.3.28), normalised to the theory uncertainty prior, S_{ii}. Explicitly, the contributions are

$$S - S(C + S)^{-1}S + S(C + S)^{-1}X(C + S)^{-1}S. \tag{5.4.1}$$

The first term is the prior, the second term is due to learning, and the third term is due to PDF fluctuations. Note that the first two terms are ZS and the whole expression is $\overline{Z}S$ in the one parameter model.

We can see that the learning reduces the prior almost to zero for NC DIS and DY, and by an order of magnitude for the rest. It is likely that more flexibility is required in the prior. The PDF fluctuations then undo a lot of the learning, but the overall uncertainty is still less than the prior.

138

We can do a similar breakdown for the correlated PDF uncertainty diagonals, this time normalised to X_{ii}. This is the first term in Eqn. 5.3.28, but expanded out like in Eqn. 5.3.31. The correlation terms, $-S(C+S)^{-1}X - X(C+S)^{-1}S$, are very large as anticipated in [146]. This is especially true where there is a large theory uncertainty (small x HERA NC or JETS), and here they can overwhelm X and give a negative result. Despite that, the addition of PDF fluctuations in $S(C+S)^{-1}X(C+S)^{-1}S$ (rememeber the breakdown in Eqn. 5.3.31) always leaves the total positive, sometimes taking it higher than X. We can therefore see that adding orrelations generally reduces the uncertainties, but they can sometimes increase them. The learning, however, always reduces them.
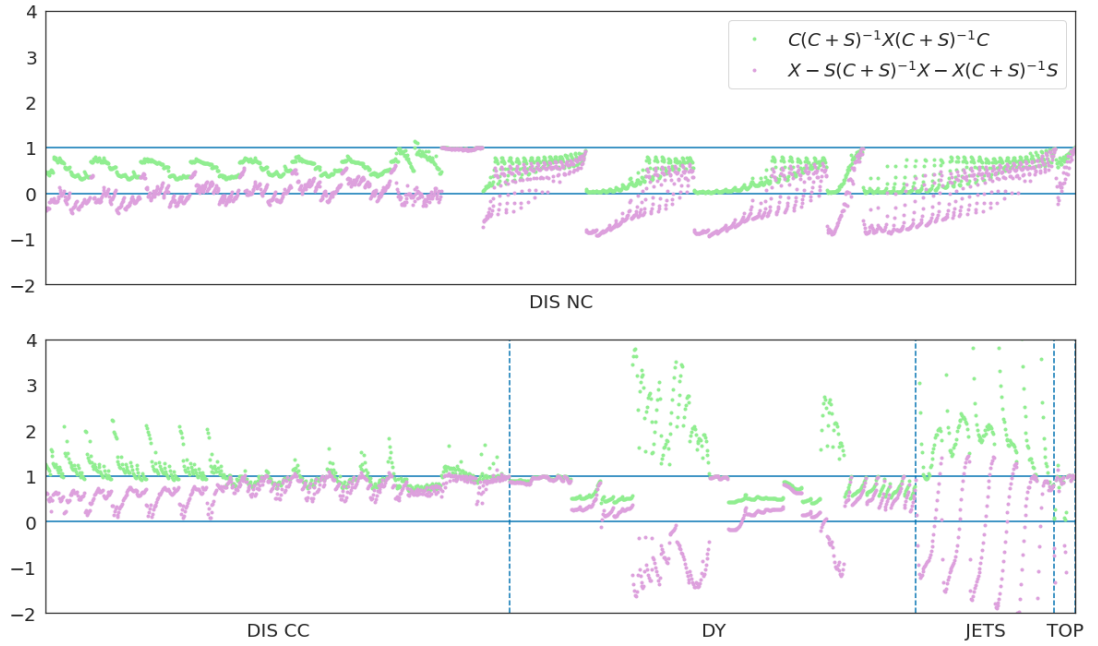
Figure 5.4.11 *The contributions to the diagonal elements of the correlated PDF uncertainty normalised to diagonal elements of X: $(X - S(C + S)^{-1}X - X(C + S)^{-1}S)_{ii}/X_{ii}$ (lilac), and $(C(C + S)^{-1}X(C + S)^{-1}C)_{ii}/X_{ii}$ (see Eqn. 5.3.31 (green).*

For the autopredictions, we expect high levels of learning and correlation, because we are making predictions for exact repetitions of experiments already in teh fit. However, as noted earlier, removing one of the smaller datasets will have little effect on the fit, which leads us to suspect that there will be similar results for genuine predictions if the process is already in the fit and, especially, if the kinematics are similar.

5.4.4 Predictions for an existing process: top production

Finally we can consider genuine predictions for experiments that weren't used in the PDF fits. These can either be for processes already in the fit, or for new ones. Here we consider the former, and in the next subsection we'll end with the latter.
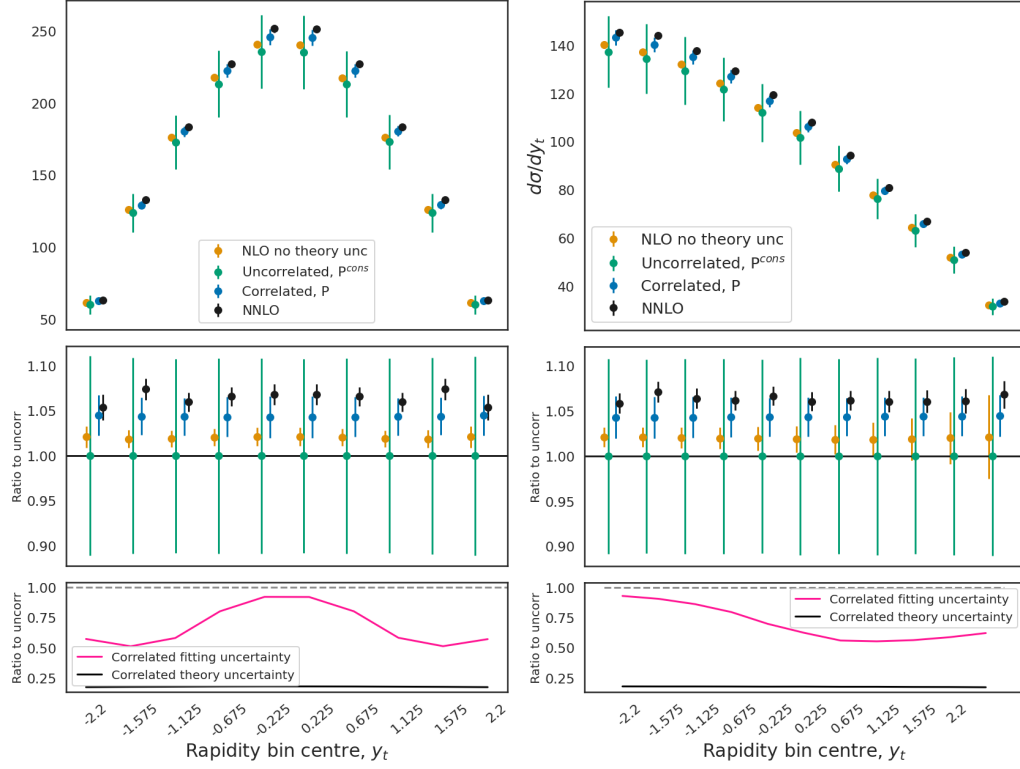
Figure 5.4.12 *The upper two panels show predictions for $t\bar{t}$ unnormalised rapidity distribution data taken at 13 TeV by CMS, the dilepton rapidity distribution [2] (left) and the lepton+jets distribution [3] (right). The four predictions show: the NLO fit with no MHOUs, PDF error only; the combined PDF and MHOU fit, ignoring correlations (thus $\sqrt{P_{II}^{\mathrm{con}}}$); the correlated result including the shift, and uncertainty computed using the simplified result (thus $\sqrt{P_{II}^{\mathrm{sim}}}$); the result with the same shift, but with the correlations included exactly (thus P_{II}), and the NNLO result with no MHOU. In the middle panels the same is shown, but normalised to the uncorrelated result. In the lower panels we show the fractional reduction in the PDF uncertainty and the theory uncertainty due to the inclusion of the correlations.*

We look at $t\bar{t}$ production rapidity distributions in two channels (dilepton and lepton + jets), measured by CMS at 13 TeV [2, 3]. There are a couple of reasons for this choice:

140

- The base fit contains $t\bar{t}$ total cross sections at 7, 8 and 13 TeV and normalised rapidity distributions at 8 TeV, all from both ATLAS and CMS.

- The NLO theory uncertainty for the 13 TeV data is ∼10%, considerably larger than the PDF uncertainty.

Both these things mean that we'd expect the correlation between the theory uncertainties in these data and the 13 TeV distributions to be high, and so we should see some of the largest effects currently possible with these PDFs. The CMS 13 TeV $t\bar{t}$ rapidity predictions were computed using the same procedure as the 8 TeV distributions in [9]: NLO theoretical predictions were generated with Sherpa [147], in a format compliant with APPLgrid [121], using the MCgrid code [148] and the Rivet [149] analysis package, with OpenLoops [150] for the NLO matrix elements. Renormalisation and factorisation scales have been chosen based on the recommendation of [151] as $H_T/4$.

The predictions are shown in Fig. 5.4.12. The correlated shift is sizeable, about 5%, but this is still comfortably within the ∼10% theory uncertainty. We'd anticipate this given that Fig. 5.4.5 tells us that the shift in nuisance parameters for the top renormalisation scale variation is also well within uncertainties. We also see that the shift is almost fully correlated across the whole distribution. This is because these are unnormalised distributions, so there is the restriction that they must sum to the total cross section. They are therefore strongly correlated with the measurements of the total cross section in the fit. We can confirm this by breaking down the contributions to the shift from the different fitted data points, seen in Table 5.4.2. The six total cross section measurements are responsibile for the vast majority of the shift, with the 8 TeV normalised rapidity distributions pushing the shift back down by about 25%. The rest of the data have almost no impact.

ATLAS				CMS				Other
tot			*diff*	*tot*			*diff*	
7 TeV	8 TeV	13 TeV	8 TeV	7 TeV	8 TeV	13 TeV	8 TeV	
0.37	0.11	0.24	-0.21	0.26	0.21	0.07	-0.04	-0.01

Table 5.4.2 *The fractional contributions of different data sets included in the fit to the shifts in the top rapidity distributions, averaged over all 21 data points.*

To see if the shift improves the predictions, we can compare it to the known NNLO-NLO shift, just like we did in Chapter 3. Therefore, in Fig. 5.4.12 we

also show the full NNLO result (without theory uncertainties). It's interesting that the shift due to correlations, which we saw is driven by the $t\bar{t}$ total cross section data, largely accounts for the NNLO correction; the data know that the NLO theory predictions are a bit low, and that knowledge is propagated into the predictions for the 13 TeV rapidity distributions.

In terms of the change in uncertainties, the middle panels show the same as the top panels but as a ratio to the uncorrelated case, making the uncertainties more visible. Comparing the difference between the uncorrelated and correlated is striking; the correlated uncertainties are far smaller than the uncorrelated ones. However, despite substantial reduction of the very large theory uncertainty, they are still larger than the pure PDF uncertainties. Even though the uncertainties have shrunk a lot, they are still compatible with the NNLO result thanks to a shift in the central values. While the conservative prescription is also compatible with the NNLO result, it is immediately obvious from the plot that it is inferior.

A breakdown of the reduction in uncertainties due to the correlations is shown in the bottom panels of Fig. 5.4.12. The correlated theory uncertainty is substantially reduced (uniformly across rapidity). This is due to the learning of the normalisation from the data already in the fit. The correlated PDF uncertainty is reduced a lot less, maximally a factor of two for where the cross section is small, but hardly at all where the cross section is large. From this it is clear that the dominant effect here is the learning of the theory uncertainty in the overall normalisation.

The theory uncertainties in the predictions are all highly correlated with one another, including between the two rapidity distributions. We can see this by looking at the correlation matrices for $\widetilde{X}$ and $\widetilde{P}$, shown in Fig. 5.4.13. The predictions are all to start with more than 50% correlated by the PDF. Then, when correlated theory uncertainties are included, the correlations bump up to > 70%. We saw before that this is due to the constraint that they must all sum to give the total cross section. The pattern of correlations nicely shows the symmetry within the dilepton distribution, and with the lepton + jets distribution; the greater the rapidity separation, the smaller the correlation.
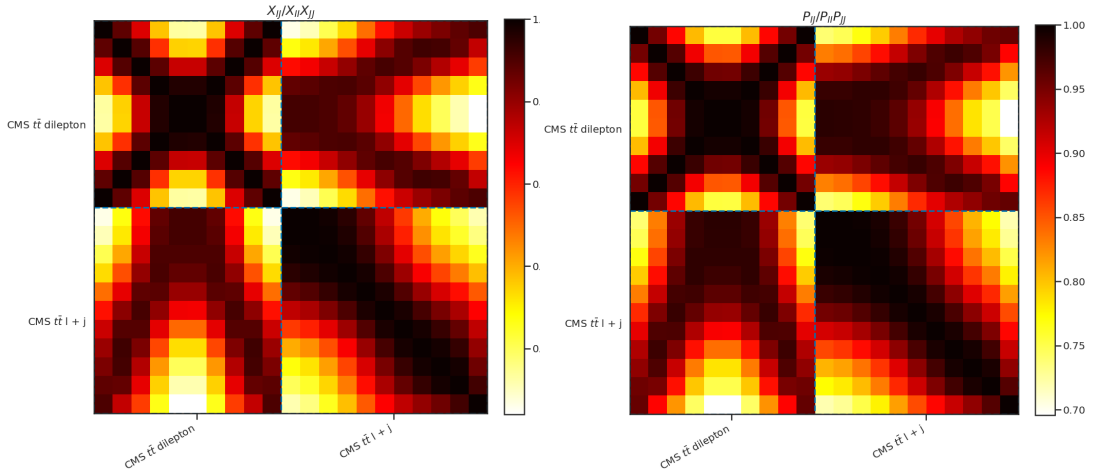
Figure 5.4.13 *The left hand plot shows the correlation matrix $\widetilde{X}_{IJ}/\sqrt{\widetilde{X}_{II}\widetilde{X}_{JJ}}$ of the contribution of the PDF uncertainties to the predictions for the 13 TeV rapidity distributions by CMS: the right hand plot shows the correlation matrix $\widetilde{P}_{IJ}/\sqrt{\widetilde{P}_{II}\widetilde{P}_{JJ}}$ of the total uncertainties including the correlated theoretical uncertainties. Note the expanded scales on the heat maps, different in each plot.*

5.4.5 Predicting a new process: Higgs production

At last we can make some predictions for a new process: one outwith the fit. For this we choose Higgs production via gluon fusion. We calculate the total cross section at 14 TeV using `ggHiggs` [152–154]. Renormalization and factorization scales are set to half the Higgs' mass, and the computation is performed using rescaled effective theory.

Our results are shown in in Fig. 5.4.14. All results use the baseline NLO PDFs with MHOUs, but the parton-level Higgs cross sections are computed at NLO, NNLO and N³LO. We break down the uncertainties into:

1. The PDF uncertainty, $\widetilde{X}$;

2. 3-point factorisation scale uncertainty;

3. 3-point renormalisation scale uncertainty;

4. Uncorrelated (conservative prescription), $\widetilde{P}_{cons}$;

5. Correlated, $\widetilde{P}$.

Note that 2. and 3. sum to give the 5-point prescription in Chapter 3.

143
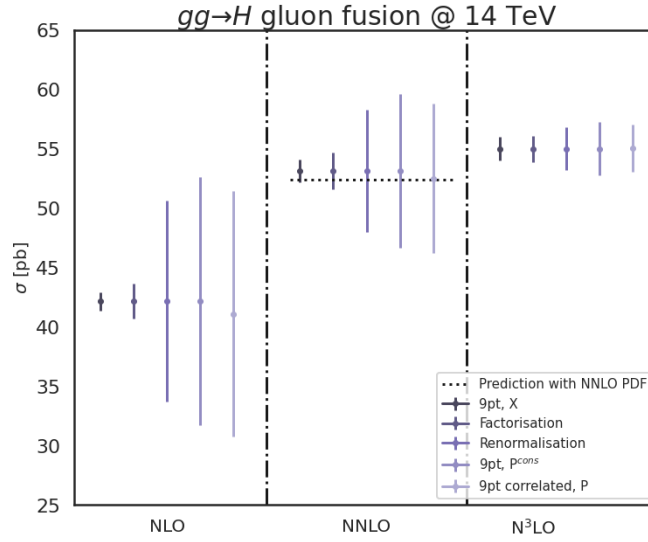
Figure 5.4.14 *Predictions for the Higgs total cross-section at 14 TeV, made using a variety of approximations. All results use NLO PDFs, while the Higgs total cross-section is computed at NLO (left panel), NNLO (centre panel) and N³LO (right panel). In each panel, we then have, from left to right: MHOU included only in the PDF determination in the 9pt scheme; the same but with the factorization scale uncertainty (MHOU in PDF evolution) included in quadrature; the same but with instead the renormalization scale uncertainty (MHOU in the Higgs cross-section); the total PDF uncertainty and 9pt MHOU combined in quadrature, as recommended in [4]; the total PDF plus 9pt MHOU, but now including also the shift and the correlation between theoretical uncertainties. In the centre panel we also show the NNLO prediction with NNLO PDFs (but no theoretical uncertainties), as a dashed line.*

At NLO the MHOU in the Higgs cross section (estimated by varying the Higgs renormalisation scale) completely dominates the other uncertainties, so the effect of correlations between the sources of MHOU is negligible. At higher orders, the renormalisation scale uncertainty shrinks dramatically until it is comparable to the other sources of uncertainty at N³LO. Notice also that the shift due to correlations is always very small compared to the overall uncertainty, and gets smaller order by order. Here, unlike for top production, the fit includes no information on Higgs production, so the renormalisation scale is totally uncorrelated. This means any information from the fit must propagate through factorisation scale uncertainties. We can see that at NNLO the small shift due to this pulls the NNLO prediction very close to the calculation using NNLO PDFs, although this is most likely coincidental. The effect of correlations on the total uncertainties is also not very large, and the difference between the simplified and

144

full calculations is small.

Note that if we used NNLO (or higher order) PDFs with MHOUs here (we can't - they don't exist!), the MHOU in the PDFs would have been smaller, and therefore the effects due to theory uncertainty correlations would be again smaller.

From these examples of autopredictions, and genuine predictions for top and Higgs, we have seen that the extent of the shift and correlation can vary quite significantly, depending on the type of prediction being made and what information is already contained in the PDFs. The conservative prescription recommended in [4] is certainly not appropriate in general, as the full inclusion of correlations can be quite substantially reduce uncertainties, as we saw both for the autopredictions and top predictions. However, when predicting a new process for which the PDF contains little information about correlated theoretical uncertainties, unsurprisingly the impact of correlations is small and the conservative prescription is sufficient.

5.5   Summary

Main conclusion: **When using PDFs which include MHOUs to make predictions, taking account of the correlations between the MHOUs in the PDF and in the predictions can provide significant improvements in accuracy and precision.**

We considered the scenario where PDFs with theory uncertainties are used to make predictions with theory uncertainties. We studied the correlations between these two sources of theory uncertainties. We did this by recasting the theory uncertainties as nuisance parameters for each PDF replica, which contain information about the experimental data's impact on the theory uncertainties. We built our way through increasingly complex and correspondingly realistic models of the fitting procedure, isolating three distinct but related effects, each of which has a significant impact on the final theoretical predictions.

1. **Shifts in central values:** We understand that we can use experimental data to determine PDFs. But we can also use them to find corrections to the theory which improve the agreement between data and theory. This is done via Bayesian learning through exposure of the fit theory to experimental data. The correlation between the theory uncertainty in the fit and in the predictions then propagates this knowledge through, leading to more accurate predictions. We identified this effect first in Sec. 5.2.

2. **Learning of theory uncertainties:** In the same mechanism as 1., information from the data is learnt by the fit theory and propagated via correlations to the predictions leading to a reduction in uncertainties. This was also first identified in Sec. 5.2.

3. **Correlations in theory uncertainties:** The correlations between the theory uncertainties in the fit and those in the prediction lead to a change in the PDF uncertainties in the prediction, even where there isn't a shift. If these correlations are unaccounted for, the theory uncertainty is "double counted". This was first noted in [146]. The effect is separate to Bayesian learning.

These three effects were found throughout in the simple models in Secs. 5.1, the one parameter fits in Sec. 5.2 and the multiparameter fits in Sec. 5.3. Using the

NNPDF3.1 NLO global fit with MHOUs, we saw explicitly in Sec. 5.4 that the shifts give reasonable estimates of NNLO corrections, and correspondingly reduce the χ^2 to the experimental data. We also showed that the uncertainty in the NLO predictions can still be thought of as a sum in quadrature of a theory uncertainty and a PDF uncertainty (which itself includes a theory uncertainty). However, these uncertiainties are reduced by a factor that depends on the relative size of the theory and experimental uncertainties, leading to significant shrinking in some cases. The upshot of this is that the conservative prescription is genuinely conservative, even if sometimes dramatically so. We expect these conclusions to also be true for global fits with fixed parametrisations and tolerance [46, 47], were MHOUs to be included.

We found that the degree of correlation is highly dependent on the type of prediction being made. For the autopredictions (predictions for new measurements of the same data points as those included in the fit), Sec. 5.4.3 where there is maximal correspondence between the data in the fit and the predictions being made, the correlation is very high, leading to shifts that improve the quality of the fit to the data, together with a significant reduction in uncertainties, in some cases down to a small fraction of the uncorrelated values. For genuine predictions for new measurements of processes already included in the PDF fit, such as the new measurements of differential top production discussed in Sec. 5.4.4, we observe that the shift takes the correlated NLO predictions very close to the NNLO prediction, with a significant reduction in uncertainties: the prediction is both more accurate and more precise. For Higgs production, discussed in Sec. 5.4.5, a process not included in the PDF fit, the level of correlation is much smaller, since the dominant uncertainty (the MHOU in the hard cross-section) is uncorrelated with the MHOU of the fitted processes. In this case the shift is well within uncertainties, and the reduction in uncertainty very modest, so here the use of the conservative prescription [4] is entirely appropriate. We expect this to be true of predictions for any new process with large theoretical uncertainties.

The main conclusion is that when using PDFs which include MHOUs to make predictions, taking account of the correlations between the MHOUs in the PDF and in the predictions can provide significant improvements in accuracy and precision. This is especially true if the predicted process is among those included in the fit.

However, we need to treat the correlated predictions with some care because their reliability is contingent on the appropriate prior being chosen for MHOUs.

147

If too many unjustified assumptions are made, we could end up with predictions that are too aggressive. Bearing this in mind, the conservative (uncorrelated) prescription could have its uses as an upper bound, especially for new processes where we expect the degree of correlation to be low.

In order to calculate fully correlated predictions and uncertainties, one requires besides the PDF replicas some additional information: the cross-correlations between the theoretical uncertainties in the prediction and those in the theoretical calculations used to determine the PDFs, $\widehat{S}_{Ij}$; and the cross-correlations between the PDF uncertainties in the prediction and all the calculations included in the fit, $\widehat{X}_{Ij}$. In the future, it may be possible to present this information in separate NNPDF deliverables to facilitate the calculation of the correlation effects.

Although we presented our numerical study of correlations in the context of MHOUs, we would expect similar results for other kinds of theoretical uncertainty, such as nuclear uncertainties, higher twist uncertainties, or indeed parametric uncertainties: once the theory covariance matrix has been computed, the linear algebra has no concern for the type of theoretical uncertainty it contains. This suggests a new technique for determining external parameters in PDF fits, such as quark masses or electroweak parameters, taking full account of all correlations with the PDFs and MHOU. We hope to explore this possibility in the near future.

Chapter 6

Conclusion

In this thesis we have considered uncertainties in the theoretical predictions that go into PDF fits, and how these theory uncertainties can impact the PDFs, both in changes to the PDFs' central values and in changes to their uncertainties. In Chapter 2 we showed how, under the assumptions that the theory uncertainties are Gaussian and independent of the experimental data, they can be included in PDF fits. This is by simply adding a theory covariance matrix to the existing experimental covariance matrix, so that uncertainties from theory and experiment stand on an equal footing. This theory covariance matrix is the covariance between the theoretical predictions and the unknown "true" values fron nature.

The complexity of this procedure lies primarily in constructing the theory covariance matrix, which cannot be determined exactly due to our lack of knowledge of the underlying truth. Instead we can consider a series of nuisance parameters which encapsulate the size of the shifts between predicted values and true values. This theory covariance matrix then acts as a prior when it is included in a PDF fit. We can in principle recalculate it using the new information obtained by the fit and then iterate to convergence. However, for a well determined prior this convergence should be fast.

We applied this procedure for including theory uncertainties in PDF fits to some of the dominant sources of uncertainties: missing higher order uncertainties (MHOUs, Chapter 3) and nuclear uncertainties (Chapter 4). MHOUs, which arise from the use of predictions to less than all orders in perturbation theory, were estimated using the established method of scale variation, where the artificial factorisation and renormalisation scales are varied to obtain a set of

predictions. We developed multiple prescriptions for combining these variations into a covariance matrix, and carried out PDF fits at NLO using the different prescriptions. We checked the efficacy of the prescriptions by comparing to the known results at NNLO. We adopted the "9 point" prescription, which performs the best, and encapsulates many features of the missing higher orders. We note that limitations in this prescription arise from the coarse categorisation of data into different processes, and the use of only one factorisation scale. For example, CHORUS data are categorised as charged current DIS, but in reality have a component of charged current and of neutral current. The factorisation scale variation could also be split up at the least into a singlet and a non-singlet component, to allow a better exploration of scale variation space. We saw that a large part of the missing higher orders that weren't encapsulated by the 9 point prescription was correlated globally, suggesting that this could be linked to the factorsation scale.

Nuclear uncertainties come from the use of data for deuteron and nuclear targets in fits for proton PDFs. The nuclear environment causes changes to the observables which are hard to quantify precisely, and these propagate through to the PDFs. To estimate the uncertainties we used an empirical approach using nuclear PDFs, which contain information about the nuclear environment. We constructed one nuclear covariance matrix for the deuteron data and one for the heavy nuclear data. We included these as default in the imminent NNPDF4.0 determination, both at NLO and NNLO, noting that they help to resolve tension between the nuclear and Drell-Yan data.

Finally, we considered the use of PDFs in making physics predictions, and how this is complicated by the presence of theory uncertainties. Theory uncertainties must be included in the PDF and in the prediction itself, but there exist correlations between these two which must be taken into account, otherwise the overall uncertainty will be inflated. We determined formulae for computing these correlations, and used them to make fully correlated predictions with theory uncertainties. We showed that when fitting a PDF, properly taking account of a combination of Bayesian learning of information from the experimental data during the fitting process, and correlations between the fit and the predictions, results in both more accurate and more precise predictions, and an improvement in χ^2. The improvement in accuracy and precision is dependent on the type of prediction being made, whereby predictions with a closer proximity to data in the fit (in terms of process type and kinematics) will be more significantly updated.

This work extends naturally to the systematic inclusion of theory uncertainties from other sources. These include: uncertainties due to chosen values of parameters such as the strong coupling constant and the quark masses; and uncertainties due to unknown higher twist contributions to the predictions. These are both the subject of current investigation, but will contribute smaller effects than MHOUs and nuclear uncertainties. Furthermore, the work on MHOUs will be extended to NNLO and become standard in future NNPDF releases. The extension from NLO to NNLO is conceptually trivial but requires technical hurdles to be overcome. Doing this upgrade would also be a good time to consider more complex renormalisation and factorisation scale splittings.

Bibliography

[1] Zahari Kassabov. Reportengine: A framework for declarative data analysis. https://doi.org/10.5281/zenodo.2571601, February 2019.

[2] Albert M Sirunyan et al. Measurements of $t\bar{t}$ differential cross sections in proton-proton collisions at $\sqrt{s} = 13$ TeV using events containing two leptons. *JHEP*, 02:149, 2019.

[3] Albert M Sirunyan et al. Measurement of differential cross sections for the production of top quark pairs and of additional jets in lepton+jets events from pp collisions at $\sqrt{s} = 13$ TeV. *Phys. Rev. D*, 97(11):112003, 2018.

[4] Rabah Abdul Khalek et al. Parton Distributions with Theory Uncertainties: General Formalism and First Phenomenological Studies. *Eur. Phys. J. C*, 79(11):931, 2019.

[5] Large hadron collider. https://home.cern/science/accelerators/large-hadron-collider. Accessed: 23.02.20.

[6] Future circular collider. https://home.cern/science/accelerators/future-circular-collider. Accessed: 23.02.20.

[7] Hiroaki Aihara et al. The International Linear Collider. A Global Project. 1 2019.

[8] Compact linear collider. https://home.cern/science/accelerators/compact-linear-collider. Accessed: 23.02.20.

[9] R. D. Ball et al. Parton distributions from high-precision collider data. *Eur. Phys. J.*, C77(10):663, 2017.

[10] M. Cepeda et al. Report from Working Group 2: Higgs Physics at the HL-LHC and HE-LHC. *CERN Yellow Rep. Monogr.*, 7:221–584, 2019.

[11] P. Azzi et al. Report from Working Group 1: Standard Model Physics at the HL-LHC and HE-LHC. *CERN Yellow Rep. Monogr.*, 7:1–220, 2019.

[12] R. K. (R. Keith) Ellis. *QCD and collider physics*. Cambridge monographs on particle physics, nuclear physics, and cosmology ; 8. Cambridge University Press, Cambridge ; New York, 1996.

[13] F. (Francis) Halzen. *Quarks and leptons : an introductory course in modern particle physics.* John Wiley, New York] ; [Chichester], 1984.

[14] Michiel Botje. Lecture notes particle physics ii: Quantum chromo dynamics, November 2013.

[15] Nathan Hartland. Proton structure at the LHC, 11 2014.

[16] Dr. H. Geiger 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):604–623, 1913.

[17] Jr. Callan, Curtis G. and David J. Gross. Bjorken scaling in quantum field theory. *Phys. Rev. D*, 8:4383–4394, 1973.

[18] R.P. Feynman. The behavior of hadron collisions at extreme energies. *Conf. Proc. C*, 690905:237–258, 1969.

[19] Richard P. Feynman. Very high-energy collisions of hadrons. *Phys. Rev. Lett.*, 23:1415–1417, 1969.

[20] R.P. Feynman. Photon-hadron interactions. 1973.

[21] C.G. Callan and David J. Gross. Crucial Test of a Theory of Currents. *Phys. Rev. Lett.*, 21:311–313, 1968.

[22] Murray Gell-Mann. Symmetries of baryons and mesons. *Phys. Rev.*, 125:1067–1084, 1962.

[23] Murray Gell-Mann. A Schematic Model of Baryons and Mesons. *Phys. Lett.*, 8:214–215, 1964.

[24] G. Zweig. *An SU(3) model for strong interaction symmetry and its breaking. Version 2.* 2 1964.

[25] Y. Dothan, Murray Gell-Mann, and Yuval Ne'eman. Series of Hadron Energy Levels as Representations of Noncompact Groups. *Phys. Lett.*, 17:148–151, 1965.

[26] B Grinstein. Introductory lectures on qcd. 2006.

[27] O.W. Greenberg. Spin and Unitary Spin Independence in a Paraquark Model of Baryons and Mesons. *Phys. Rev. Lett.*, 13:598–602, 1964.

[28] T. Kinoshita. Mass singularities of Feynman amplitudes. *J. Math. Phys.*, 3:650–677, 1962.

[29] T.D. Lee and M. Nauenberg. Degenerate Systems and Mass Singularities. *Phys. Rev.*, 133:B1549–B1562, 1964.

[30] John C. Collins, Davison E. Soper, and George F. Sterman. *Factorization of Hard Processes in QCD*, volume 5. 1989.

[31] G. Altarelli and G. Parisi. Asymptotic freedom in parton language. *Nucl. Phys. B*, 1977.

[32] Y. L. Dokshitzer. *Sov. Phys.*, 1977.

[33] L. N. Gribov and V. N. Lipatov. *Sov. J. Nucl. Phys*, 1972.

[34] Nnpdf website. https://nnpdf.mi.infn.it/. Accessed: 23.07.21.

[35] Cern website. https://home.cern/. Accessed: 23.07.21.

[36] Fermilab website. https://www.fnal.gov/. Accessed: 23.07.21.

[37] Emanuele R. Nocera. Towards NNPDF4.0: The structure of the proton to one-percent accuracy. "http://nnpdf.mi.infn.it/wp-content/uploads/2021/04/NOCERA_DIS2021.pdf", 2021.

[38] J.M. Campbell and R.K. Ellis. An update on vector boson pair production at hadron colliders. *Phys. Rev.*, 1999.

[39] J. Alwall, 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*, 07:079, 2014.

[40] S. Catani et al. Vector boson production at hadron colliders: a fully exclusive QCD calculation at NNLO. *Phys. Rev. Lett.*, 103:082001, 2009.

[41] Ryan Gavin, Ye Li, Frank Petriello, and Seth Quackenbush. FEWZ 2.0: A code for hadronic Z production at next-to-next-to-leading order. *Comput. Phys. Commun.*, 182:2388–2403, 2011.

[42] S. S. Catani and M. H. Seymour. A General algorithm for calculating jet cross-sections in NLO QCD. *Nucl. Phys.*, B485:291–419, 1997. [Erratum: Nucl. Phys.B510,503(1998)].

[43] V. Bertone et al. APFEL: A PDF Evolution Library with QED corrections. *Comput. Phys. Commun.*, 185:1647–1668, 2014.

[44] G. D'Agostini. On the use of the covariance matrix to fit correlated data. *Nucl. Instrum. Meth. A*, 346:306–311, 1994.

[45] Richard D. Ball, Luigi Del Debbio, Stefano Forte, Alberto Guffanti, Jose I. Latorre, Juan Rojo, and Maria Ubiali. Fitting Parton Distribution Data with Multiplicative Normalization Uncertainties. *JHEP*, 05:075, 2010.

[46] S. Bailey, T. Cridge, L. A. Harland-Lang, A. D. Martin, and R. S. Thorne. Parton distributions from LHC, HERA, Tevatron and fixed target data: MSHT20 PDFs. *Eur. Phys. J. C*, 81(4):341, 2021.

[47] Tie-Jiun Hou et al. New CTEQ global analysis of quantum chromodynamics with high-precision data from the LHC. *Phys. Rev. D*, 103(1):014013, 2021.

[48] A. M. Cooper-Sarkar. PDF Fits at HERA. *PoS*, EPS-HEP2011:320, 2011.

[49] S. Alekhin, J. Blümlein, and S. Moch. An update of the ABM16 PDF fit. In *26th International Workshop on Deep Inelastic Scattering and Related Subjects*. SISSA, 9 2019.

[50] S. Forte et al. Neural network parametrization of deep inelastic structure functions. *JHEP*, 05:062, 2002.

[51] Richard D. Ball, Luigi Del Debbio, Stefano Forte, Alberto Guffanti, Jose I. Latorre, Juan Rojo, and Maria Ubiali. A first unbiased global NLO determination of parton distributions and their uncertainties. *Nucl. Phys. B*, 838:136–206, 2010.

[52] Richard D. Ball et al. Parton distributions with LHC data. *Nucl. Phys. B*, 867:244–289, 2013.

[53] Lhapdf website. https://lhapdf.hepforge.org/. Accessed: 23.02.20.

[54] A. Buckley et al. LHAPDF6: parton density access in the LHC precision era. *Eur. Phys. J.*, C75:132, 2015.

[55] Luigi Del Debbio, Stefano Forte, Jose I. Latorre, Andrea Piccione, and Joan Rojo. Neural network determination of parton distributions: The Nonsinglet case. *JHEP*, 03:039, 2007.

[56] Richard D. Ball et al. Parton distributions for the LHC Run II. *JHEP*, 04:040, 2015.

[57] NNPDF collaboration. Theory and methodology: new developments. "http://nnpdf.mi.infn.it/wp-content/uploads/2021/04/pdf4lhc_all.pdf", 2021.

[58] Stefano Carrazza and Juan Cruz-Martinez. Towards a new generation of parton densities with deep learning models. *Eur. Phys. J. C*, 79(8):676, 2019.

[59] Franccois Chollet. keras. https://github.com/fchollet/keras, 2015.

[60] Abadi et al. Tensorflow: Large-scale machine learning on heterogeneous distributed systems, 2016.

[61] Alessandro Candido, Stefano Forte, and Felix Hekhorn. Can $\overline{\text{MS}}$ parton distributions be negative? *JHEP*, 11:129, 2020.

[62] W. Hollik et al. Electroweak physics. *Acta Phys. Polon. B*, 35:2533–2555, 2004.

[63] S. Carrazza, E. R. Nocera, C. Schwan, and M. Zaro. PineAPPL: combining EW and QCD corrections for fast evaluation of LHC processes. *JHEP*, 12:108, 2020.

[64] Stefano Forte, Eric Laenen, Paolo Nason, and Juan Rojo. Heavy quarks in deep-inelastic scattering. *Nucl. Phys. B*, 834:116–162, 2010.

[65] Richard D. Ball, Valerio Bertone, Luigi Del Debbio, Stefano Forte, Alberto Guffanti, Juan Rojo, and Maria Ubiali. Theoretical issues in PDF determination and associated uncertainties. *Phys. Lett. B*, 723:330–339, 2013.

[66] André H. Hoang. The Top Mass: Interpretation and Theoretical Uncertainties. In *7th International Workshop on Top Quark Physics*, 12 2014.

[67] Richard D. Ball, Stefano Carrazza, Luigi Del Debbio, Stefano Forte, Zahari Kassabov, Juan Rojo, Emma Slade, and Maria Ubiali. Precision determination of the strong coupling constant within a global PDF analysis. *Eur. Phys. J. C*, 78(5):408, 2018.

[68] High-luminosity lhc. https://home.cern/science/accelerators/high-luminosity-lhc. Accessed: 23.02.20.

[69] G. D'Agostini. Bayesian reasoning versus conventional statistics in high-energy physics. In *18th International Workshop on Maximum Entropy and Bayesian Methods (MaxEnt 98)*, 7 1998.

[70] G. D'Agostini. Bayesian inference in processing experimental data: Principles and basic applications. *Rept. Prog. Phys.*, 66:1383–1420, 2003.

[71] R. D. Ball and A. Deshpande. The Proton Spin, Semi-Inclusive processes, and a future Electron Ion Collider. 2018.

[72] G. Aad et al. Measurement of the Z/γ^* boson transverse momentum distribution in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector. *JHEP*, 09:145, 2014.

[73] V. Khachatryan et al. Measurement of the Z boson differential cross section in transverse momentum and rapidity in proton–proton collisions at 8 TeV. *Phys. Lett.*, B749:187–209, 2015.

[74] Georges et al. G. Aad. 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):291, 2016.

[75] Particle Data Group. Review of Particle Physics. *Progress of Theoretical and Experimental Physics*, 2020(8), 08 2020. 083C01.

[76] John Campbell, Joey Huston, and Frank Krauss. *The Black Book of Quantum Chromodynamics*. Oxford University Press, 2017.

[77] Matteo Cacciari and Nicolas Houdeau. Meaningful characterisation of perturbative theoretical uncertainties. *JHEP*, 1109:039, 2011.

[78] André David and Giampiero Passarino. How well can we guess theoretical uncertainties? *Phys. Lett.*, B726:266–272, 2013.

[79] Emanuele Bagnaschi, Matteo Cacciari, Alberto Guffanti, and Laura Jenniches. An extensive survey of the estimation of uncertainties from missing higher orders in perturbative calculations. *JHEP*, 02:133, 2015.

[80] Marco Bonvini. Probabilistic definition of the perturbative theoretical uncertainty from missing higher orders. *Eur. Phys. J. C*, 80(10):989, 2020.

[81] Alan D. Martin, R. G. Roberts, and W. James Stirling. Precision analysis of Λ (MS) and the gluon distribution and its implication for jet and top quark cross-sections. *Phys. Rev.*, D43:3648–3656, 1991.

[82] Marc Virchaux and Alain Milsztajn. A Measurement of alpha-s and higher twists from a QCD analysis of high statistics F-2 data on hydrogen and deuterium targets. *Phys. Lett.*, B274:221–229, 1992.

[83] G. Ridolfi and S. Forte. Renormalization and factorization scale dependence of observables in QCD. *J. Phys.*, G25:1555–1556, 1999.

[84] Guido Altarelli, Richard D. Ball, and Stefano Forte. Small x Resummation with Quarks: Deep-Inelastic Scattering. *Nucl. Phys.*, B799:199–240, 2008.

[85] D. de Florian et al. Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector. 2016.

[86] M. Arneodo et al. Accurate measurement of F_2^d/F_2^p and $R_d - R_p$. *Nucl. Phys.*, B487:3–26, 1997.

[87] M. Arneodo et al. Measurement of the proton and deuteron structure functions, F_2^p and F_2^d, and of the ratio σ_L/σ_T. *Nucl. Phys.*, B483:3–43, 1997.

[88] L. W. Whitlow, E. M. Riordan, S. Dasu, Stephen Rock, and A. Bodek. Precise measurements of the proton and deuteron structure functions from a global analysis of the SLAC deep inelastic electron scattering cross-sections. *Phys. Lett.*, B282:475–482, 1992.

[89] A. C. Benvenuti et al. A high statistics measurement of the proton structure functions $f_2(x, q^2)$ and r from deep inelastic muon scattering at high q^2. *Phys. Lett.*, B223:485, 1989.

[90] A. C. Benvenuti et al. A high statistics measurement of the deuteron structure functions $f_2(x, q^2)$ and r from deep inelastic muon scattering at high q^2. *Phys. Lett.*, B237:592, 1990.

[91] M. Goncharov et al. Precise measurement of dimuon production cross-sections in muon neutrino Fe and muon anti-neutrino Fe deep inelastic scattering at the Tevatron. *Phys. Rev.*, D64:112006, 2001.

[92] David Alexander Mason. Measurement of the strange - antistrange asymmetry at NLO in QCD from NuTeV dimuon data. FERMILAB-THESIS-2006-01.

[93] G. Onengut et al. Measurement of nucleon structure functions in neutrino scattering. *Phys. Lett.*, B632:65–75, 2006.

[94] H. Abramowicz et al. Combination of measurements of inclusive deep inelastic $e^{\pm}p$ scattering cross sections and QCD analysis of HERA data. *Eur. Phys. J.*, C75(12):580, 2015.

[95] H. Abramowicz et al. Combination and QCD Analysis of Charm Production Cross Section Measurements in Deep-Inelastic ep Scattering at HERA. *Eur.Phys.J.*, C73:2311, 2013.

[96] Timo Antero Aaltonen et al. Measurement of $d\sigma/dy$ of Drell-Yan e^+e^- pairs in the Z Mass Region from $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV. *Phys. Lett.*, B692:232–239, 2010.

[97] V. M. Abazov et al. Measurement of the shape of the boson rapidity distribution for $p\bar{p} \rightarrow Z/\gamma^* \rightarrow e^+e^- + X$ events produced at $\sqrt{s}$=1.96-TeV. *Phys. Rev.*, D76:012003, 2007.

[98] Victor Mukhamedovich Abazov et al. Measurement of the electron charge asymmetry in $\boldsymbol{p\bar{p}\rightarrow W+X\rightarrow e\nu+X}$ decays in $\boldsymbol{p\bar{p}}$ collisions at $\boldsymbol{\sqrt{s}=1.96}$ TeV. *Phys. Rev.*, D91(3):032007, 2015. [Erratum: Phys. Rev.D91,no.7,079901(2015)].

[99] Victor Mukhamedovich Abazov et al. Measurement of the muon charge asymmetry in $p\bar{p} \rightarrow$ W+X $\rightarrow \mu\nu$ + X events at $\sqrt{s}$=1.96 TeV. *Phys.Rev.*, D88:091102, 2013.

[100] Georges Aad et al. Measurement of the inclusive $W^{\pm}$ and Z/γ^* cross sections in the electron and muon decay channels in pp collisions at $\sqrt{s}$= 7 TeV with the ATLAS detector. *Phys.Rev.*, D85:072004, 2012.

[101] Morad Aaboud et al. Precision measurement and interpretation of inclusive W^+ , W^- and Z/γ^* production cross sections with the ATLAS detector. *Eur. Phys. J.*, C77(6):367, 2017.

[102] Georges Aad et al. Measurement of the low-mass Drell-Yan differential cross section at $\sqrt{s} = 7$ TeV using the ATLAS detector. *JHEP*, 06:112, 2014.

[103] Georges Aad et al. Measurement of the high-mass Drell–Yan differential cross-section in pp collisions at $\sqrt{s}$=7 TeV with the ATLAS detector. *Phys.Lett.*, B725:223, 2013.

[104] Georges Aad et al. Measurement of inclusive jet and dijet production in pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector. *Phys. Rev.*, D86:014022, 2012.

[105] Georges Aad et al. Measurement of the $t\bar{t}$ production cross-section using $e\mu$ events with b-tagged jets in pp collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS detector. *Eur. Phys. J.*, C74(10):3109, 2014. [Addendum: Eur. Phys. J.C76,no.11,642(2016)].

[106] Morad Aaboud et al. Measurement of the $t\bar{t}$ production cross-section using $e\mu$ events with b-tagged jets in pp collisions at $\sqrt{s}$=13 TeV with the ATLAS detector. *Phys. Lett.*, B761:136–157, 2016.

[107] Georges Aad et al. Measurements of top-quark pair differential cross-sections in the lepton+jets channel in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector. *Eur. Phys. J.*, C76(10):538, 2016.

[108] Sergei Chatrchyan et al. Measurement of the differential and double-differential Drell-Yan cross sections in proton-proton collisions at $\sqrt{s} = 7$ TeV. *JHEP*, 1312:030, 2013.

[109] Serguei Chatrchyan et al. Measurement of the electron charge asymmetry in inclusive W production in pp collisions at $\sqrt{s} = 7$ TeV. *Phys.Rev.Lett.*, 109:111806, 2012.

[110] Serguei Chatrchyan et al. Measurement of the muon charge asymmetry in inclusive pp to WX production at $\sqrt{s} = 7$ TeV and an improved determination of light parton distribution functions. *Phys.Rev.*, D90:032004, 2014.

[111] Vardan Khachatryan et al. Measurement of the differential cross section and charge asymmetry for inclusive pp $\rightarrow$ W$^\pm$ + X production at $\sqrt{s} = 8$ TeV. *Eur. Phys. J.*, C76(8):469, 2016.

[112] Serguei Chatrchyan et al. Measurements of differential jet cross sections in proton-proton collisions at $\sqrt{s} = 7$ TeV with the CMS detector. *Phys.Rev.*, D87:112002, 2013.

[113] Vardan Khachatryan et al. Measurement of the t-tbar production cross section in the e-mu channel in proton-proton collisions at sqrt(s) = 7 and 8 TeV. *JHEP*, 08:029, 2016.

[114] Vardan Khachatryan et al. Measurement of the top quark pair production cross section in proton-proton collisions at $\sqrt{(s)} = 13$ TeV. *Phys. Rev. Lett.*, 116(5):052002, 2016.

[115] Vardan Khachatryan et al. Measurement of the differential cross section for top quark pair production in pp collisions at $\sqrt{s} = 8$ TeV. *Eur. Phys. J.*, C75(11):542, 2015.

[116] R Aaij et al. Inclusive W and Z production in the forward region at $\sqrt{s} = 7$ TeV. *JHEP*, 1206:058, 2012.

[117] R Aaij et al. Measurement of the cross-section for $Z \to e^+e^-$ production in *pp* collisions at $\sqrt{s} = 7$ TeV. *JHEP*, 1302:106, 2013.

[118] Roel Aaij et al. Measurement of the forward Z boson production cross-section in *pp* collisions at $\sqrt{s} = 7$ TeV. *JHEP*, 08:039, 2015.

[119] Roel Aaij et al. Measurement of forward W and Z boson production in *pp* collisions at $\sqrt{s} = 8$ TeV. *JHEP*, 01:155, 2016.

[120] Valerio Bertone, Stefano Carrazza, and Nathan P. Hartland. APFELgrid: a high performance tool for parton density determinations. *Comput. Phys. Commun.*, 212:205–209, 2017.

[121] Tancredi Carli et al. A posteriori inclusion of parton density functions in NLO QCD final-state calculations at hadron colliders: The APPLGRID Project. *Eur.Phys.J.*, C66:503, 2010.

[122] J. F. Owens, A. Accardi, and W. Melnitchouk. Global parton distributions with nuclear and finite-Q^2 corrections. *Phys. Rev.*, D87(9):094012, 2013.

[123] L. A. Harland-Lang et al. Parton distributions in the LHC era: MMHT 2014 PDFs. *Eur. Phys. J.*, C75(5):204, 2015.

[124] A. Accardi et al. Constraints on large-x parton distributions from new weak boson production and deep-inelastic scattering data. *Phys. Rev.*, D93(11):114017, 2016.

[125] S. I. Alekhin, S. A. Kulagin, and R. Petti. Nuclear Effects in the Deuteron and Constraints on the d/u Ratio. *Phys. Rev. D*, 96(5):054005, 2017.

[126] Robert B. Wiringa, V.G.J. Stoks, and R. Schiavilla. An Accurate nucleon-nucleon potential with charge independence breaking. *Phys. Rev. C*, 51:38–51, 1995.

[127] W. Melnitchouk, Andreas W. Schreiber, and Anthony William Thomas. Relativistic deuteron structure function. *Phys. Lett. B*, 335:11–16, 1994.

[128] W. Melnitchouk, M. Sargsian, and M.I. Strikman. Probing the origin of the EMC effect via tagged structure functions of the deuteron. *Z. Phys. A*, 359:99–109, 1997.

[129] R. Machleidt. The High precision, charge dependent Bonn nucleon-nucleon potential (CD-Bonn). *Phys. Rev. C*, 63:024001, 2001.

[130] Franz Gross. Covariant spectator theory of np scattering: Deuteron magnetic moment. *Phys. Rev. C*, 89(6):064002, 2014. [Erratum: Phys.Rev.C 101, 029901 (2020)].

[131] S. Dulat et al. New parton distribution functions from a global analysis of quantum chromodynamics. *Phys. Rev.*, D93(3):033006, 2016.

[132] S. Alekhin, J. Blümlein, S. Moch, and R. Placakyte. Parton distribution functions, α_s, and heavy-quark masses for LHC Run II. *Phys. Rev.*, D96(1):014011, 2017.

[133] A.D. Martin, A.J. Th.M. Mathijssen, W.J. Stirling, R.S. Thorne, B.J.A. Watt, et al. Extended Parameterisations for MSTW PDFs and their effect on Lepton Charge Asymmetry from W Decays. *Eur.Phys.J.*, C73(2):2318, 2013.

[134] Richard D. Ball, Emanuele R. Nocera, and Rosalyn L. Pearson. Nuclear Uncertainties in the Determination of Proton PDFs. *Eur. Phys. J.*, C79(3):282, 2019.

[135] Daniel de Florian, Rodolfo Sassot, Pia Zurita, and Marco Stratmann. Global Analysis of Nuclear Parton Distributions. *Phys. Rev.*, D85:074028, 2012.

[136] K. Kovarik et al. nCTEQ15 - Global analysis of nuclear parton distributions with uncertainties in the CTEQ framework. *Phys. Rev.*, D93(8):085037, 2016.

[137] Kari J. Eskola, Petja Paakkinen, Hannu Paukkunen, and Carlos A. Salgado. EPPS16: Nuclear parton distributions with LHC data. *Eur. Phys. J.*, C77(3):163, 2017.

[138] Rabah Abdul Khalek, Jacob J. Ethier, Juan Rojo, and Gijs van Weelden. nNNPDF2.0: quark flavor separation in nuclei from LHC data. *JHEP*, 09:183, 2020.

[139] Richard D. Ball, Emanuele R. Nocera, and Rosalyn L. Pearson. Deuteron Uncertainties in the Determination of Proton PDFs. *Eur. Phys. J. C*, 81(1):37, 2021.

[140] J. G. Heinrich et al. Measurement of the Ratio of Sea to Valence Quarks in the Nucleon. *Phys. Rev. Lett.*, 63:356–359, 1989.

[141] M. Tzanov et al. Precise measurement of neutrino and anti-neutrino differential cross sections. *Phys. Rev.*, D74:012008, 2006.

[142] A. C. Benvenuti et al. A High Statistics Measurement of the Deuteron Structure Functions $F_2(x, Q^2)$ and R from Deep Inelastic Muon Scattering at High Q^2. *Phys. Lett.*, B237:592, 1990.

[143] R. S. Towell et al. Improved measurement of the anti-d/anti-u asymmetry in the nucleon sea. *Phys. Rev.*, D64:052002, 2001.

[144] J. Dove et al. The asymmetry of antimatter in the proton. *Nature*, 590(7847):561–565, 2021.

[145] Rabah Abdul Khalek et al. A First Determination of Parton Distributions with Theoretical Uncertainties. 2019.

[146] L. A. Harland-Lang and R. S. Thorne. On the Consistent Use of Scale Variations in PDF Fits and Predictions. *Eur. Phys. J.*, C79(3):225, 2019.

[147] T. Gleisberg et al. Event generation with SHERPA 1.1. *JHEP*, 02:007, 2009.

[148] Luigi Del Debbio, Nathan P. Hartland, and Steffen Schumann. MCgrid: projecting cross section calculations on grids. *Comput.Phys.Commun.*, 185:2115–2126, 2014.

[149] Andy Buckley, Jonathan Butterworth, Leif Lonnblad, David Grellscheid, Hendrik Hoeth, et al. Rivet user manual. *Comput.Phys.Commun.*, 184:2803–2819, 2013.

[150] Fabio Cascioli, Philipp Maierhofer, and Stefano Pozzorini. Scattering Amplitudes with Open Loops. *Phys. Rev. Lett.*, 108:111601, 2012.

[151] Michal Czakon, David Heymes, and Alexander Mitov. Dynamical scales for multi-TeV top-pair production at the LHC. *JHEP*, 04:071, 2017.

[152] Richard D. Ball, Marco Bonvini, Stefano Forte, Simone Marzani, and Giovanni Ridolfi. Higgs production in gluon fusion beyond NNLO. *Nucl.Phys.*, B874:746, 2013.

[153] Marco Bonvini, Richard D. Ball, Stefano Forte, Simone Marzani, and Giovanni Ridolfi. Updated Higgs cross section at approximate N^{3}LO. *J. Phys.*, G41:095002, 2014.

[154] Marco Bonvini, Simone Marzani, Claudio Muselli, and Luca Rottoli. On the Higgs cross section at N^{3}LO+N^{3}LL and its uncertainty. *JHEP*, 08:105, 2016.