Theoretical Uncertainties in Parton Distribution Functions

Rosalyn Laura Pearson

Abstract

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

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

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

Declaration

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

Parts of this work have been published in [].

(Rosalyn Laura Pearson, May 2020)

Acknowledgements

Insert people you want to thank here.

Contents

Contents

List of Figures

List of Tables

0.1 List of Publications

- **Towards parton distribution functions with theoretical uncertainties**, Pearson, R. L. and Voisey, C. Nuclear and Particle Physics Proceedings, Volumes 300-302, July-September 2018, Pages 24-29, e-Print: 1810.01996 [hep-ph]

- **Nuclear Uncertainties in the Determination of Proton PDFs**, NNPDF Collaboration: Richard D. Ball et al. (Dec 21, 2018), Published in: Eur.Phys.J.C 79 (2019) 3, 282, e-Print: 1812.09074 [hep-ph]

- **A first determination of parton distributions with theoretical uncertainties**, NNPDF Collaboration: Rabah Abdul Khalek et al. (May 10, 2019), Published in: Eur.Phys.J. C (2019) 79:838, e-Print: 1905.04311 [hep-ph]

- **Uncertainties due to Nuclear Data in Proton PDF Fits**, Rosalyn Pearson, Richard Ball, Emanuele Roberto Nocera (Jun 14, 2019), Published in: PoS DIS2019 (2019) 027, Contribution to: DIS 2019, 027

- **Parton Distributions with Theory Uncertainties: General Formalism and First Phenomenological Studies**, NNPDF Collaboration: Rabah Abdul Khalek et al. (Jun 25, 2019) Published in: Eur.Phys.J.C 79 (2019) 11, 931, e-Print: 1906.10698 [hep-ph]

Introduction

Over the past 100 years, following the discovery of the atomic nucleus by Rutherford in 1911, great strides have been made towards understanding

subatomic structure. We now know that atoms are made up of hadrons (such as protons and neutrons) and leptons (such as the electron). Probing hadrons with high energy photons shows that they are composed of quarks and gluons.
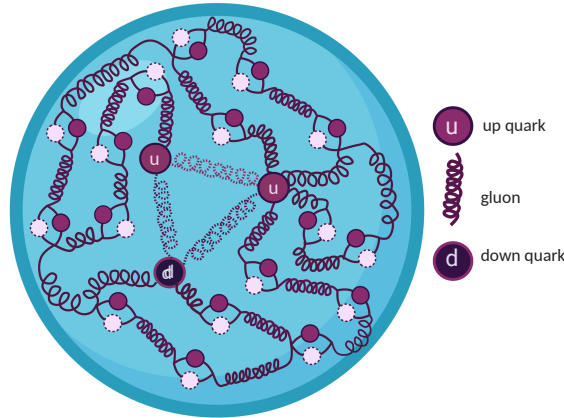
Figure 0.1.1 *A visualisation of the internal structure of the proton. Quarks are bound together by gluons.*

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

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

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

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

useful theoretical predictions we therefore need as precise and accurate a handle on the PDFs as possible.

PDFs are unknowns in perturbative Quantum Chromodynamics (QCD), the theory of the strong interaction. Crucially, they are process independent. This means that they can be determined in a global fit between a wealth of experimental data and theoretical predictions. Once constrained, they can then be applied to any process. Fig. 4.5.1 shows the fitted functional form of the PDFs.
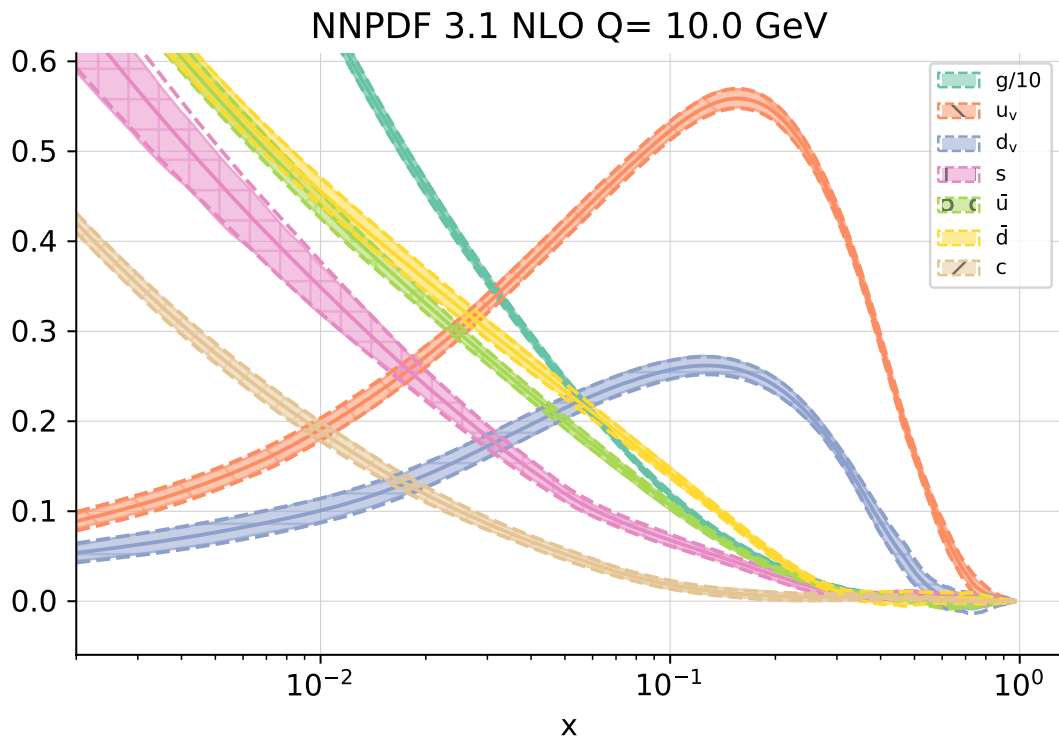
Figure 0.1.2 *The different PDF flavours determined in the latest NNPDF3.1 release [16].*

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

1. experimental uncertainties;

2. theory uncertainties;

3. methodological uncertainties (errors from the fitting procedure).

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

Chapter 1

Background

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

This section provides some background to PDFs 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 (DIS), and how the naïve parton model was developed to explain these experimental observations. Next we look at this in the context of QCD, see how PDFs fit into the picture, and how they evolve with the scale of the physics. Finally we briefly touch on hadron-hadron collisions, which along with DIS constitute the bulk of the processes in modern PDF fits.

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

1.0.1 Deep inelastic scattering

For a more in-depth analysis, see Refs. [39, 54]. In the following background sections we rely heavily on Ref. [31].

The notion of bombarding matter to uncover its structure has led to many important discoveries in the last hundred or so years, starting with the Geiger-Marsen experiments from 1908-1913 and the subsequent discovery of the atomic nucleus [47]. In the decades following the discovery of the neutron in 1932, nuclei were probed at higher energies, leading to them being understood in terms of "form factors" which parametrised their electric and magnetic distributions. At this stage it was clear that they were not point-like particles and so a series of important experiments were carried out in the 1960s at the Standford Linear Accelerator (SLAC), involving a high energy beam of charged leptons scattering off a stationary hadronic target. This process is known as Deep Inelastic Scattering (DIS).

In this section we will consider the example of electrons incident on protons, as shown in Fig. **??**. In the deep inelastic regime, there is a large momentum transfer, $q = k - k'$, mediated by a virtual photon. The proton, P, with initial momentum p, fragments into some hadronic state X, and the electron starts with energy E and momentum k and ends with energy E' and momentum k'. The momentum transfer is large enough that the masses of the proton and electron can be neglected.
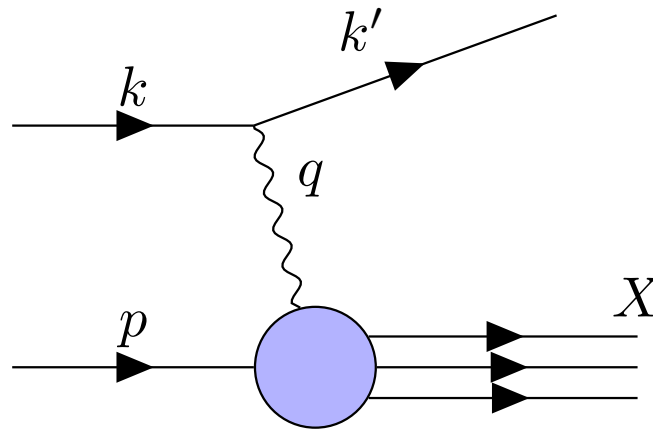
Figure 1.0.1 *Deep inelastic scattering*

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

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

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

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

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

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

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

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

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

us to write its general form as

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

(1.0.3)

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

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

(1.0.4)

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

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

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

(1.0.5)

1.0.2 The parton model

Carrying out DIS experiments allows us to measure the structure functions for different values of x and Q^2. It transpired that no clear Q^2 dependence was observed, and this is known as Björken scaling [33]. 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"[41–43].

Furthermore, $F_L(x)$ was measured to be 0, known as the Callan-Gross relation ??, which suggests that the point-like particles could not absorb longitudinal photons. This fitted in nicely with the quark models developed shortly before [38, 48, 49, 62], which described hadrons in terms of spin-1/2 quarks; spin-

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

In the DIS regime, Q^2 is large and so the virtual photon probes at the short timescale $1/Q$, meaning that the interaction will be effectively instantaneous when compared with the inner proton dynamics which operate at the QCD scale $1/\lambda_{QCD} \sim 1$ fm. In the parton model we make the assumption that the partons have only a small momentum transverse to the proton's, and that they are effectively on shell for the interaction ($k^2 \approx 0$). In addition, we consider the process in the infinite momentum frame of the proton, in which it is Lorentz contracted by M/P (a small number), so we can assume the photon will only interact with one parton because it will only traverse a narrow cross-section of the proton. The updated picture is shown in Fig. **??**.
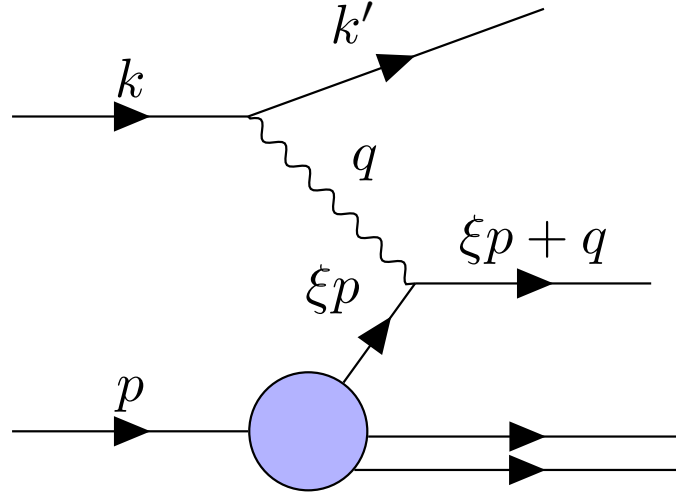
Figure 1.0.2 *DIS in the parton model. One parton with momentum p interacts with the virtual photon, and the other partons "spectate".*

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

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

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, $\hat{W}_{\mu\nu}^q$, as

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

where q runs over the possible quark flavours and f_q are distributions, with $f_q(x)dx$ giving the probability that in an interaction a parton of flavour q will be found in the momentum range $x \to x + dx$. 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 1.0.1, we can see that

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

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

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

to show (in the massless limit)

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

Using the symmetry of Fig. **??**, 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 $p \cdot k = Q^2/2xy$,

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

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

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

We see immediately that the Callan-Gross relation, $F_L(x) \equiv F_2(x) - 2x F_1(x) = 0$, 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).

1.0.3   Quantum Chromodynamics (QCD)

QCD is the theory of the strong force. This is responsible for binding together hadrons, and explains the short-range interactions which occur within them. It is a gauge theory where the quark fields are realised as fundamental representations of the $SU(3)$ symmetry group and interactions between them are carried out via gauge bosons termed "gluons", which are expressed in the adjoint representation [52].

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" [38, 48, 49, 62] . The additional $SU(3)_c$ colour symmetry was put forwards in order that the quarks satisfied Fermi-Dirac statistics [50]. 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

$$\mathcal{L} = -\frac{1}{4}F^a_{\mu\nu}F^{a\mu\nu} + \bar{q}^i(i\not{D}^j_i - m\delta^j_i)q_j, \tag{1.0.13}$$

where the covariant derivative is

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

and the field strength tensor is

$$F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + \sqrt{4\pi\alpha_s}f^{abc}A^b_\mu A^c_\nu. \tag{1.0.15}$$

The indices μ, ν are spacetime indices, i, j are quark colour indices and a, b, c are gluon colour indices. The first term in the Lagrangian arises from the self-interacting gluons, A, and the second term from the quarks, q, which obey the Dirac equation. α_s is the strong coupling constant, which dictates the strength of the interaction, and T^a are the eight $SU(3)$ generators. f^{abc} are the SU(3) structure constants. For simplicity we have assumed all quarks have the same mass, m. Note that gauge fixing and ghost terms are omitted. For more information see Ref. [39].

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

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

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

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

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

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

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

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

1.0.4   The QCD improved parton model and factorisation

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

partonic interaction and the gluon emission. The observed violation of Björken scaling has its origins in interactions with gluons. In IR-safe observables the soft and collinear divergences exactly canel [56, 57], but for other cases we need a way of dealing with the disparate short- and long- scale physics.

This is done using the factorisation theorem **??**, which allows us to factorise the incalculable long-distance physics into the PDFs, meaning we are able to use perturbative QCD as a predictive theory. The PDFs are then non-perturbative, meaning we must obtain them from experiments, but they are universal quantities and so once determined can be applied everywhere, much like the coupling constants. This process introduces the artificial "factorisation scale", μ_F, in addition to the renormalisation scale. The factorisation scale separates the short- and long- distance physics; loosely, if a parton's transverse momentum is less than μ_F it is considered part of the hadron and is factored into the PDFs, otherwise it is seen as taking part in the hard scattering process, and will appear in the partonic cross section.
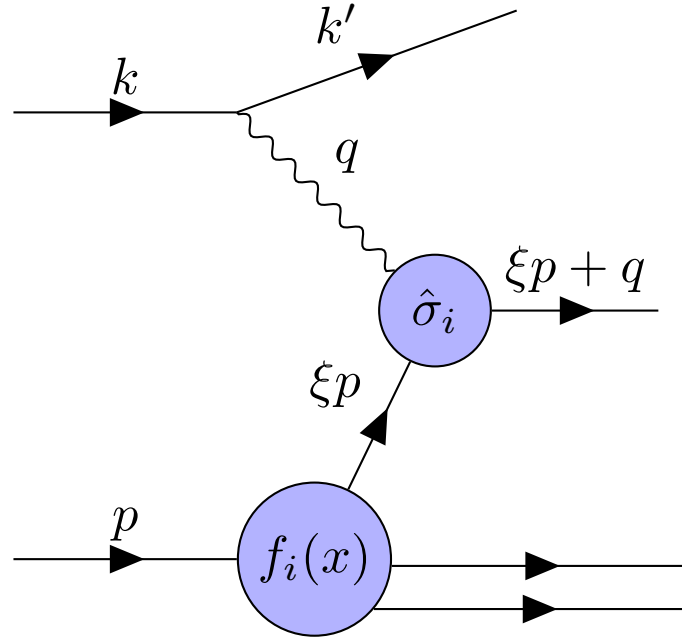
Figure 1.0.3 *Factorisation and the QCD improved parton model*

We can write a DIS cross section as

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

corresponding to Fig **??**.

We can see how this works in practice by considering the case where a quark emits a gluon before interaction with the photon, such as in Fig. **??**. Here the parent parton, with fraction y of the proton's momentum, emits a gluon giving rise to a daughter parton with a fraction z of the parent hadron's momentum. We can see that $z = x/y$.
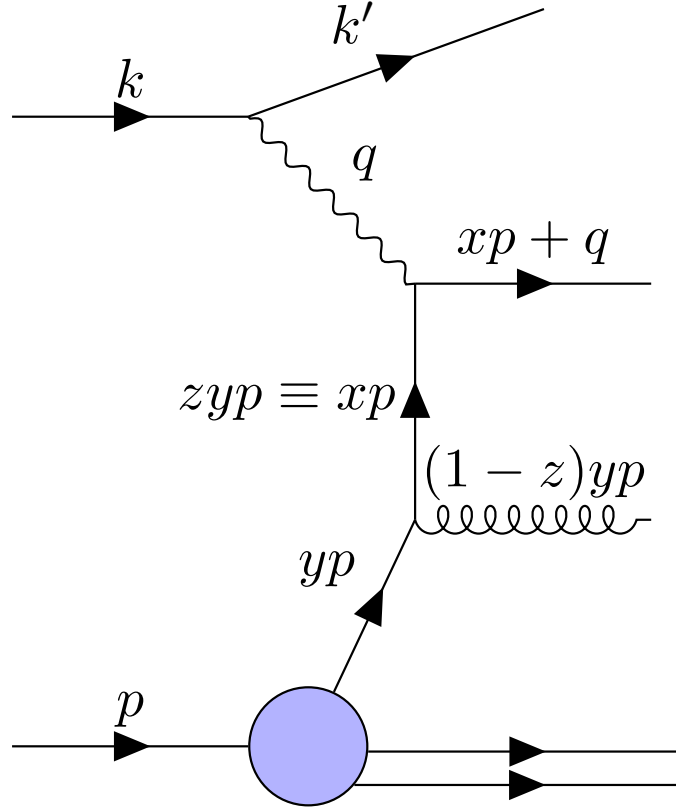
Figure 1.0.4 *A quark radiating a gluon before interacting.*

It transpires (see Ref. [54] for the derivation) that the structure function F_2 can be expressed as

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

m is a cutoff introduced to regularise the collinear divergence and you can see that as $m \to 0$ 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

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

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

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

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

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

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

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

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

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

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

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

$$(1.0.26)$$

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

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

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

When we also include the gluon as a parton, we open ourselves up to more splitting possibilities (e.g. gluon $\rightarrow$ quark and gluon $\rightarrow$ gluon), and this result generalises to a set of coupled differential equations known as the DGLAP equations [21, 37, 51]:

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

where we have used the Mellin convolution, defined

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

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

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

$$M(n) \equiv \int_0^1 dx \; x^{n-1} M(x). \tag{1.0.30}$$

Considering the Mellin transform of the RHS of the DGLAP equation, and

suppressing the parton indices, i, for clarity, we have

$$
\begin{aligned}
\int_0^1 dx\ x^{n-1} \quad & \int_x^1 \frac{dy}{y}\frac{\alpha_s}{2\pi}\ \mathcal{P}\left(\frac{x}{y}\right) f(y,\mu_F^2)\\
& = \int_0^1 dx\ x^{n-1}\left[\int_0^1 dy \int_0^1 dz\ \frac{\alpha_s}{2\pi}\ \mathcal{P}(z)f(y)\delta(x-yz)\right]\\
& = \frac{\alpha_s}{2\pi}\int_0^1 dz\ z^{n-1}\ \mathcal{P}(z)\int_0^1 dy\ y^{n-1}f(y)\\
& = \frac{\alpha_s}{2\pi}\mathcal{P}(n)\ f(n)\\
& \equiv \gamma(n)\ f(n),
\end{aligned}
\tag{1.0.31}
$$

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

1.0.5   Hadroproduction

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

$$
\sigma = \sum_{i,j}\int dx_1 dx_2 f_i(x_1,\mu_F^2)f_j(x_2,\mu_F^2)\hat{\sigma}_{ij}\left(x_1,x_2,\frac{Q^2}{\mu_F^2},...\right).
\tag{1.0.32}
$$
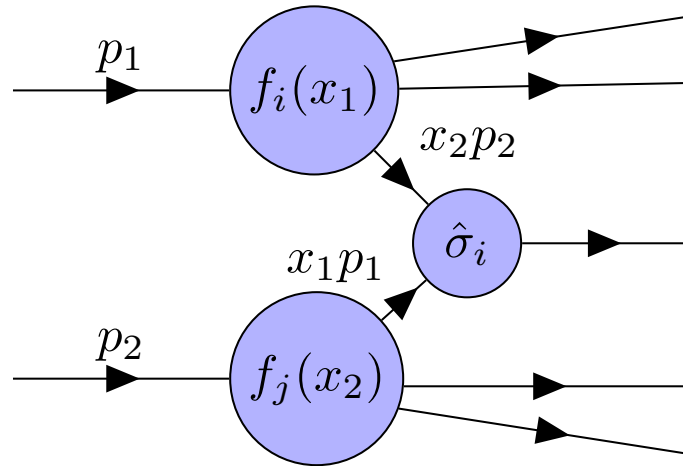
Figure 1.0.5   *Factorisation in hadron-hadron collisions.*

Write about higher order corrections and factorisation!

1.0.6 Sum rules

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

$$\int_0^1 dx \sum_i x f_i(x, Q^2) = 1. \tag{1.0.33}$$

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

$$\int_0^1 dx(f_u - f_{\bar{u}}) = 2; \tag{1.0.34a}$$

$$\int_0^1 dx(f_d - f_{\bar{d}}) = 1; \tag{1.0.34b}$$

$$\int_0^1 dx(f_q - f_{\bar{q}}) = 0, \qquad q = s, c, t, b. \tag{1.0.34c}$$

Note that these conditions require that the PDFs are integrable.

1.1 Determining PDFs

In this section we review the necessary background for PDF determination within the NNPDF [1] framework. First we touch on the experimental and theoretical inputs to PDF fits, then we summarise the NNPDF fitting strategy, and finally we detail information on neural networks specific to this context.
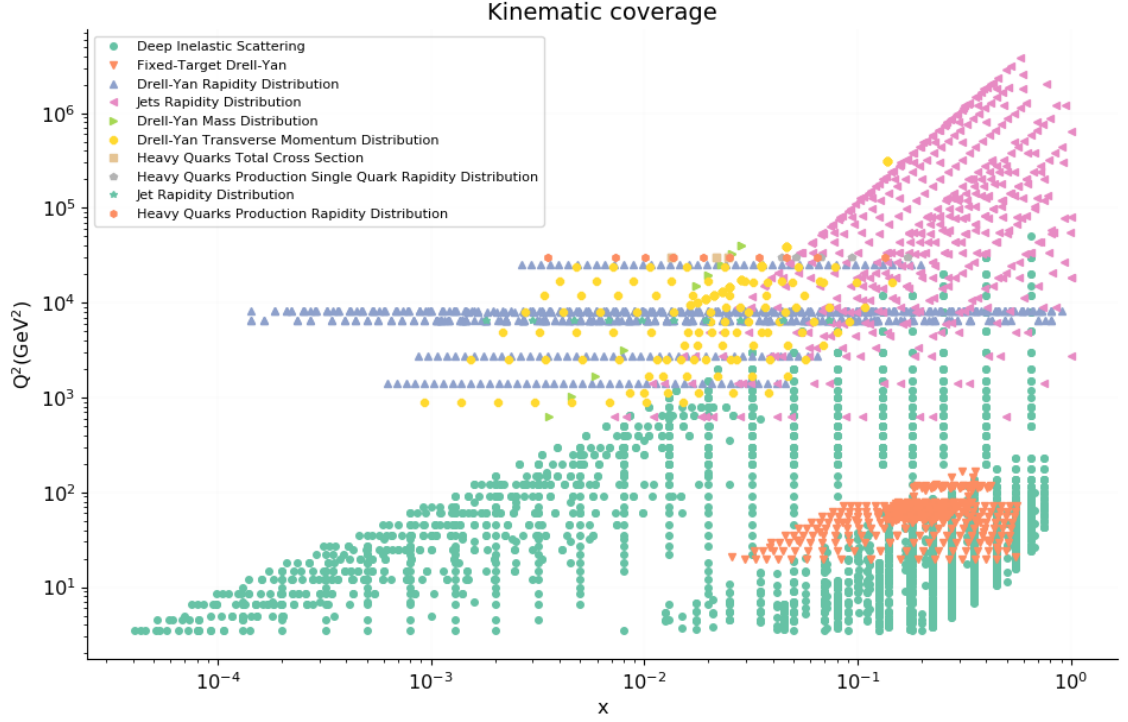
Figure 1.1.1 *Plot of the (x, Q^2) range spanned by data included in the latest NNPDF3.1 NLO fit.*

1.1.1 Experimental and theoretical input

NNPDF uses a variety of experimental data from a number of particle colliders, including those based at CERN [2] and Fermilab [3]. These are observables such as cross sections, differential cross sections and structure functions. Fig. 1.1.1 is a plot of the (x, Q^2) range spanned by the datasets in the latest NNPDF3.1 [16] release. The majority of the data are from DIS processes, which are crucial in determining PDF functional form, but in recent years increasingly more LHC collider data has been added including $t\bar{t}$ production, high energy jets and single top production [58]. For a full review of the data, see Ref. **??**.

Theoretical predictions of the corresponding parton-level observables are computed using external codes such as `MCFM` [34], `aMC@NLO` [22], `DYNNLO` [17], `FEWZ` [46] and `NLOjet++` [60]. These are converted to higher orders of perturbation theory as necessary using QCD and electroweak correction ("k") factors. They are then combined with DGLAP evolution kernels, which evolve PDFs from an initial reference energy scale to the energy scale of each experiment using the DGLAP equations (Eqn. 1.0.28).

1.1.2 Experimental uncertainties

Experimental uncertainties are described using a 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):

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

where D_i are the experimental data values.

Structurally, the uncorrelated statistical uncertainties appear down the diagonal and these are what we would recognise intuitively as the statistical error "on a data point". However, correlated systematic uncertainties can also appear on the off-diagonals. Correlated 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. 1.1.2 is an example of an experimental covariance matrix for data included in an NNPDF fit. The data are grouped according to what type of process the interaction belongs to (DIS charged current (CC) and neutral current (NC), Drell-Yan (DY), jets and top production). Systematic correlations within experiments are responsible for off-diagonal contributions, and these are mostly positive correlations but there is some anticorrelated behaviour in DIS CC, as a result of data in different kinematic regimes.
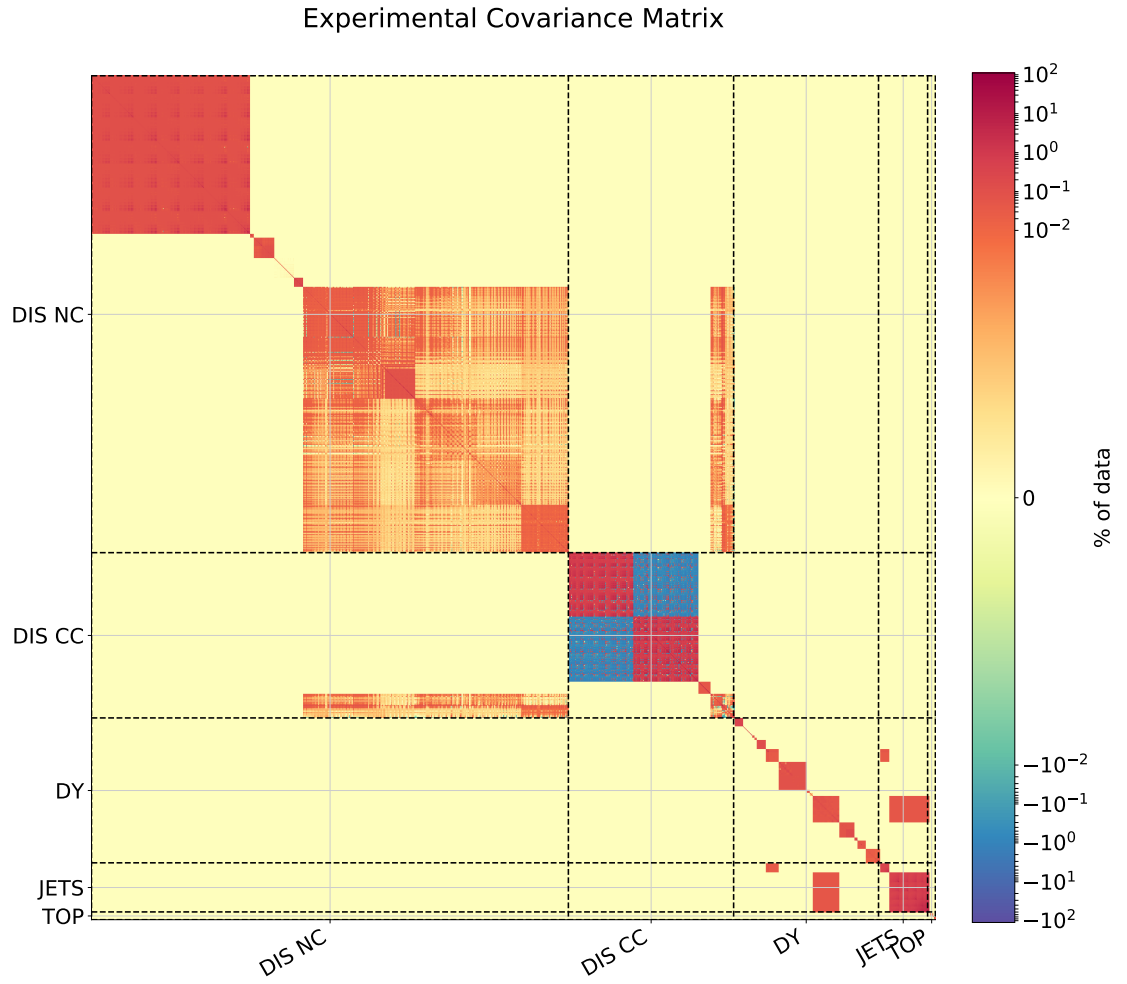
Figure 1.1.2 *An example of an experimental covariance matrix for data included in an NNPDF fit. The data are grouped according to what type of process the interaction belongs to (DIS charged current (CC) and neutral current (NC), Drell-Yan (DY), jets and top production).*

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

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

which measures the goodness of fit between the experimental data D_i with associated error breakdown C_{ij}, and theory predictions T_i. In practice, this definition is subject to d'Agostini bias [36] due to the presence of normalisation uncertainties. To avoid this, NNPDF employ the iterative $t0$ procedure [28] whereby D_i in Eqn. **??** are replaced initially with the predictions from a baseline fit, and the covariance matrix is iterated concurrently with preprocessing.

1.1.3   NNPDF fitting strategy

There are a number of groups currently active in carrying out PDF fits including MSTW [8], CTEQ [18], NNPDF [1], HERAPDF/xFitter [35] and ABM [59].

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

1. The use of Monte Carlo approach to error analysis;

2. Fitting using artificial neural networks.

In the following sections we will provide an overview of these aspects, which can be found in more detail in Refs. [16, 27, 29].

1.1.4   Monte Carlo approach

The uncertainties in the functional form of PDFs come as a direct consequence of the uncertainties in the experimental and theoretical input. In order to propagate experimental uncertainties through to the PDFs, NNPDF represent the experimental data (central values and uncertainty distribution) as a Monte Carlo ensemble. This is a set of N_{rep} Monte Carlo "replicas" which, given high enough replica number, have a mean value equal to the data central value and covariance equal to the experimental covariance. Fig. 1.1.3 is a schematic illustrating the generation of these "pseudodata", $D^{(k)}$, $k = 1, ..., N_{rep}$. They are generated using
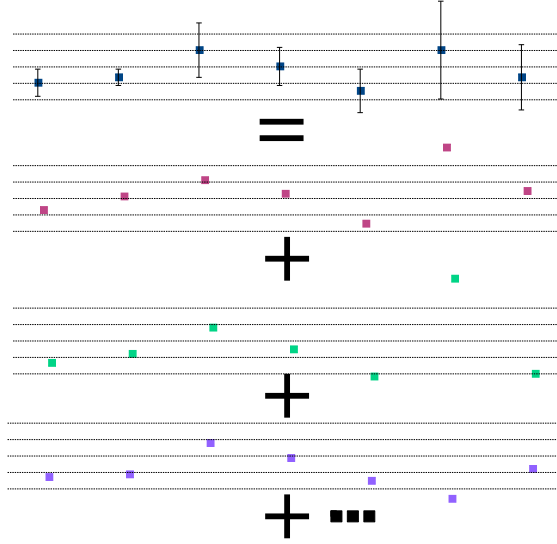
Figure 1.1.3 *Schematic of the generation of Monte Carlo replicas of pseudodata from data with uncertainties.*
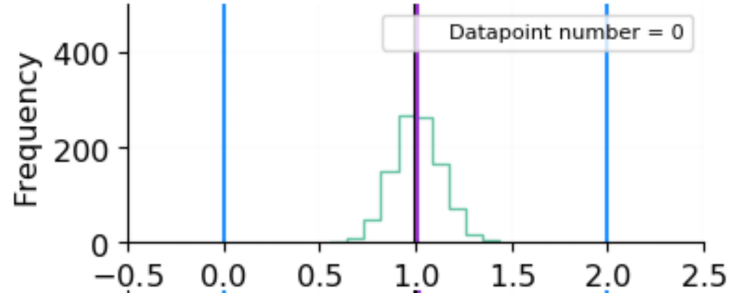
Figure 1.1.4 *Histogram of distribution of 100 pseudodata replicas for a single data point, normalised to D^0. The purple line is the mean value $\langle D^{(k)} \rangle$, which is equal to D^0 to arbitrary precision.*

Gaussian random numbers $n_a^{(k)}$ and $\hat{n}_p^{(k)}$:

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

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

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

where the notation $\langle \cdot \rangle$ denotes the mean over replicas. Fig. 1.1.4 shows the distribution of pseudodata for a single data point.
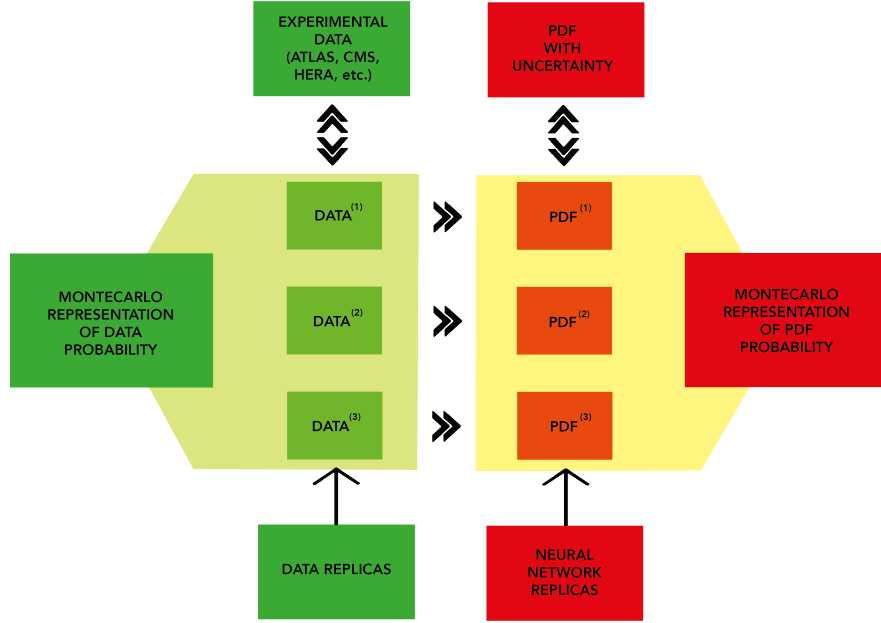
Figure 1.1.5 *NNPDF general strategy.*

Once the pseudodata have been generated, each of these $(D^{(k)})$ is fitted separately to the theoretical predictions by minimising a target error function based on the χ^2 (Eqn. 1.1.2), resulting in a PDF set of each flavour, $f_q^{(k)}$ (where q runs over the fitted flavours: g, u, d, s, c, $\bar{u}$, $\bar{d}$, $\bar{s}$, $\bar{c}$). These act as a Monte Carlo parametrisation of the PDFs (for example, Fig. 1.1.6). This means that the PDFs and their errors can be extracted by taking the means and standard deviations over the ensemble. The final PDFs are made publicly available as downloadable files on the LHAPDF website [4, 7].

1.1.5 Neural Networks

Inspired by how the brain processes information, in machine learning neural networks are a graph of connected nodes. They are trained by example, so have the capability to learn a PDF's functional form given a set of data. The use of neural networks rather than specific functional forms allows us to avoid the theoretical bias which goes into selecting such a functional form. The layout, or "architecture", consists of input layers, hidden layers and output layers. Nodes can be either input nodes or activation nodes, the latter of which have an associated activation function which is applied to their output. Fig. **??** depicts the architecture currently used by NNPDF. This is a "2-5-3-1" archiecture, where the numbers refer to the number of nodes in each layer. It is a "multilayer
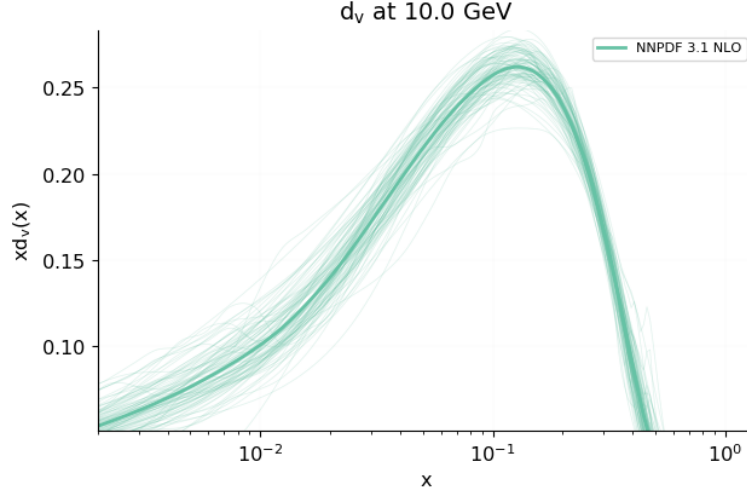
Figure 1.1.6 *Monte Carlo replicas for the down valence quark PDF NNPDF3.1 at NLO.*

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

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

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

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

The choice of sigmoid activation function for the hidden layers allows sufficient non-linear freedom in the functional form, and the linear activation function for the output layer ensures the range of the PDFs is not restricted to [0,1].
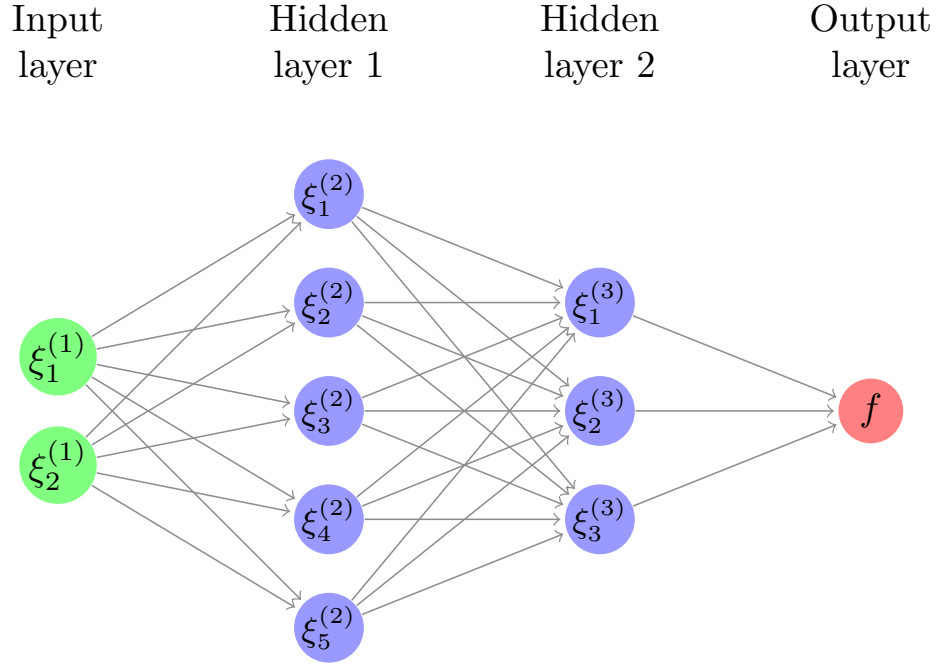
Figure 1.1.7 *Schematic depiction of the 2-5-3-1 architecture of an artificial neural network currently used by NNPDF. In the NNPDF methodology $\xi_1^{(1)}$ and $\xi_2^{(1)}$ are the variables x and $\log x$ respectively.*

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

1.1.6 Parametrisation, preprocessing and postprocessing

A scale of $Q = 1.65 \ GeV$ is chosen to parametrise the PDFs at, and then they can be determined at any other scale by evolution using the DGLAP equations (Eqn. 1.0.28). The PDFs are fitted parametrised in a "fitting basis", to help convergence [30], 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:

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

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

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

1.1.7 Cross validation

Neural networks are effective at learning the functional form underlying data. Sometimes they can be "too effective", picking up not just the underlying law but also the noise. This is known as "overlearning" (see Fig. ?? for an example).
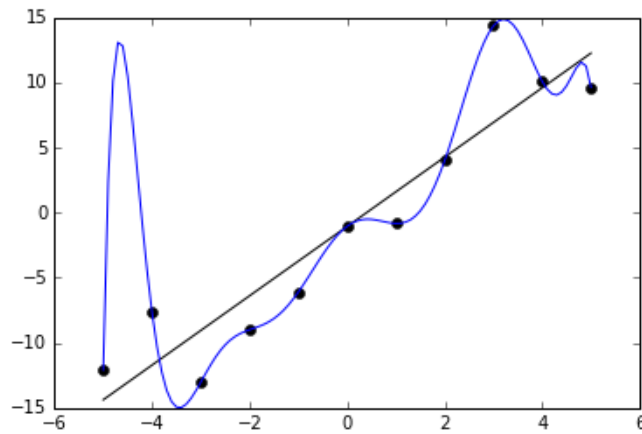
Figure 1.1.8 *Overlearning: the data points (black dots) fluctuate around the linear underlying law (black line), but the neural network continues to minimise the error function until it passes through every data point (blue curve), fitting the noise in the data.*

To circumvent this problem, the data is split into a training and a validation set. The training data is used to optimise the neural network, and the validation data is used to test the network output, in a process known as "cross validation". As training epochs elapse, the target error function compared to both the training and validation data should decrease as the network learns the underlying law. At some point, however, the network will begin to learn the noise in the training data, at which point the training error function will continue to decrease, but the validation error function will stop decreasing and start to increase again. The optimum fit is determined using the "lookback" method, where after training the model corresponding to the minimum in the validation error function is selected.
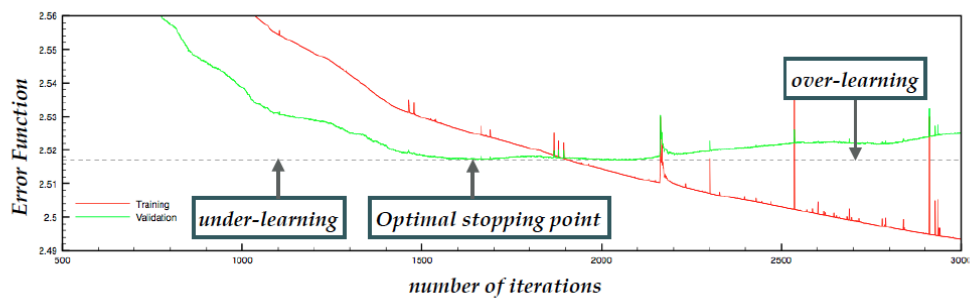
Figure 1.1.9 *Cross validation with the lookback method.*
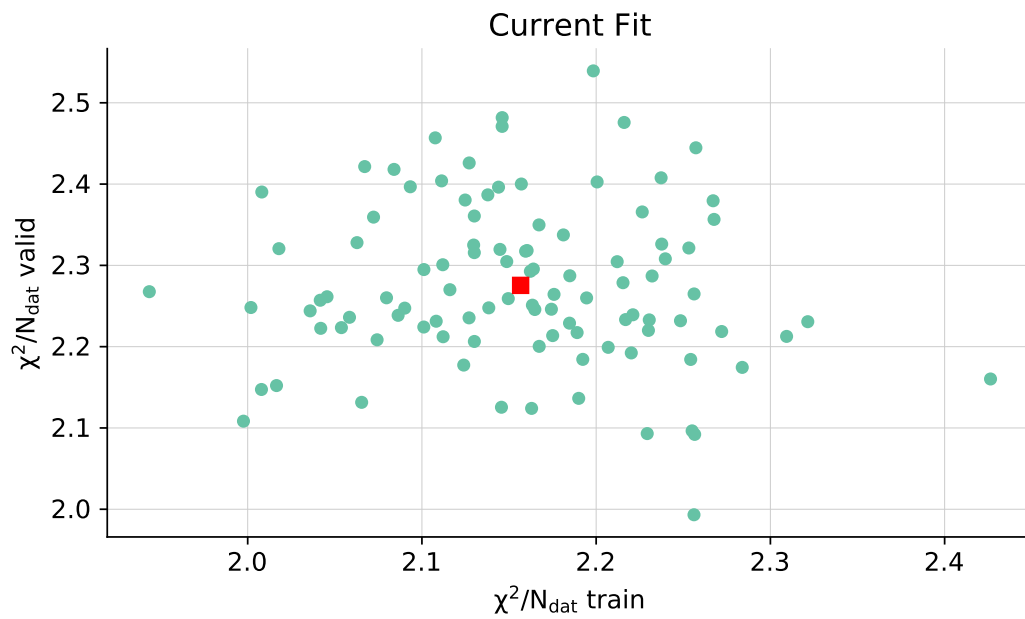
Figure 1.1.10 *Comparing the training and validation χ^2s for the 100 replicas (green circles) of a PDF fit. The red square gives the average.*

Chapter 2

Theory uncertainties in PDFs

- What are theory uncertainties?

- Why are they now important?

- Types of unc - see below

- Bayesian interpretation of these uncertainties - there is one "true" value
  e.g. for the higher order value, so need to estimate uncertainty bc will
  never know the size of e.g. MHOU unless you go ahead and calc - maybe
  ref d'Agostini paper on Bayesian interpretation

- Will use Bayesian framework and assume Gaussianity of the expected true
  value of theory calc

- Show C+S in fit - plus sign because exp and th unc are independent so
  combine errors in quadrature. They are also on an equal footing in terms
  of their effect on the PDFs

- When many datasets/global fit, can have v strong theory correlations even
  across different experiments, because the underlying theory connects them

2.1 Fitting PDFs including theory uncertainties

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

χ^2 are currently based entirely on these. We must now try to match the ongoing drive to increase experimental precision by including errors introduced at the theoretical level. This is especially important given recent data sets such as the Z boson transverse momentum distributions [10, 20, 45], which have very high experimental precision. Without the inclusion of theoretical errors, this has led to tension with the other datasets.

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

When determining PDFs we incorporate information from experiments in the form of N_{dat} experimental data points D_i, $i = 1, ..., N_{dat}$. 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_i. This is the origin of theory uncertainties in PDF fits; where our theory is incomplete, fails to describe the physics well enough, or where approximations are made, we will introduce all kinds of subtle biases into the PDF fit. The theory predictions themselves also depend on PDFs, so uncertainties already present in the PDFs are propagated through. This, in particular, leads to a high level of correlation because the PDFs are universal, and shared between all the theory predictions.

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

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

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

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

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

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

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 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 the above integral in terms of the shifts Δ_i, making use of the assumptions of Gaussianity

in Eqns. 2.1.1 and 2.1.2:

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

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

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

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:

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

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} [24]. 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 them MHOUs, nuclear corrections, higher twist corrections etc.) which not only encapsulate the per-point theoretical errors 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 following chapters address several important types of theoretical uncertainties: MHOUs; nuclear uncertainties; deuteron uncertainties. 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.

Chapter 3

Missing higher order uncertainties

In this chapter we address the dominant source of theoretical 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. 3.4). Finally, we present the PDFs including MHOUs (Sec. 3.5) and assess the impact on relevant phenomenology (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 (0.118 [53]), 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.
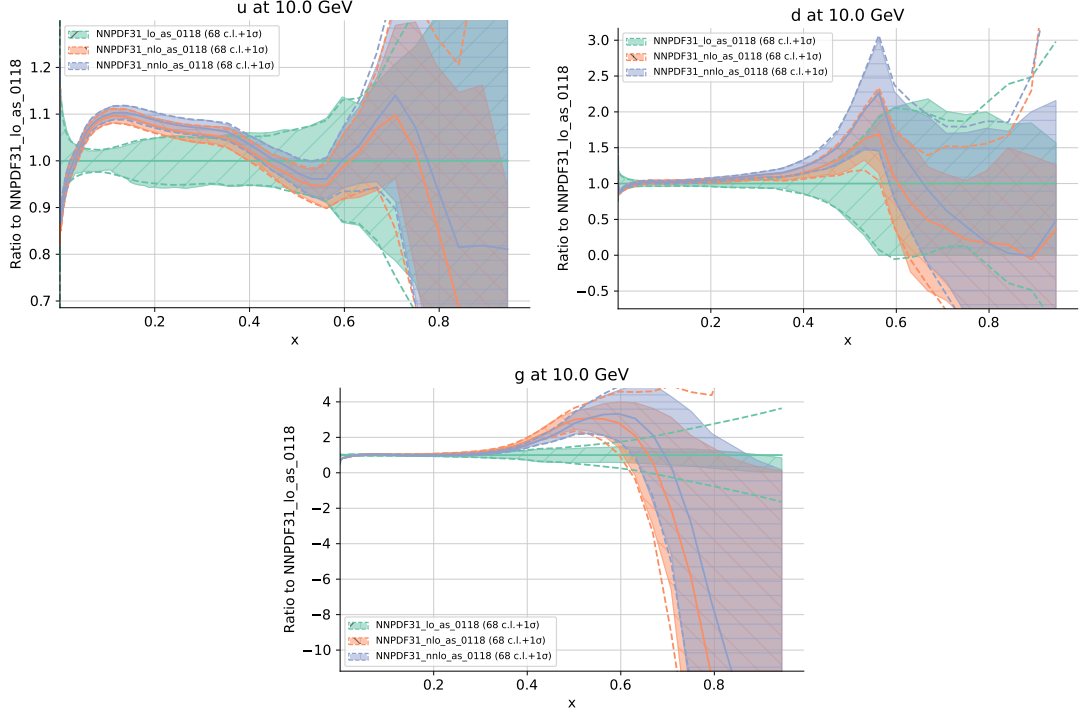
Figure 3.1.1 *Comparison of NNPDF3.1 PDFs at different perturbative orders: LO (green); NLO (orange); NNLO (blue). PDFs are normalised to the LO result, and displayed at scale Q = 10 GeV.*

These missing higher order terms in the expansion for theory predictions lead to missing higher order uncertainties (MHOUs), which are currently the dominant source of error 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 errors, justified historically by the claim that they are small compared to experimental contributions to the PDF error, 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 errors themselves are typically $\mathcal{O}(1\%)$ [?] and, with the current push to N³LO 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; 1.0.17) 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 [**? ? ?**], 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 have 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 genral 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

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

Noting that $\beta(\alpha_s) = \mathcal{O}(\alpha_s^2)$, we see that $\frac{1}{T}\frac{dT}{dt} = \mathcal{O}(\alpha_s)$ and $\frac{1}{T}\frac{d^2T}{dt^2} = \mathcal{O}(\alpha_s^2)$ 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

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

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

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

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

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

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

expressions for the theory uncertainties

$$
\begin{aligned}
\Delta_{\mathrm{LO}}(t,\kappa) &= T_{\mathrm{LO}}(t+\kappa) - T_{\mathrm{LO}}(t), \\
\Delta_{\mathrm{NLO}}(t,\kappa) &= (T_{\mathrm{NLO}}(t+\kappa) - \kappa \tfrac{d}{dt} T_{\mathrm{LO}}(t+\kappa)) - T_{\mathrm{NLO}}(t), \\
\Delta_{\mathrm{NNLO}}(t,\kappa) &= (T_{\mathrm{NNLO}}(t+\kappa) - \kappa \tfrac{d}{dt} T_{\mathrm{NLO}}(t+\kappa) \\
&\quad + \tfrac{1}{2}\kappa^2 \tfrac{d^2}{dt^2} T_{\mathrm{LO}}(t+\kappa)) - T_{\mathrm{NNLO}}(t) \,.
\end{aligned}
\tag{3.2.15}
$$

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 $T_{\mathrm{NLO}}(t+\kappa)$ 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 $T_{\mathrm{NNLO}}(t+\kappa) - \kappa \tfrac{d}{dt} T_{\mathrm{NLO}}(t+\kappa)$, 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

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

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,

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

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

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

resulting in the perturbative expression for μ_R variation of C:

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

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

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

. 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:

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

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

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

where

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

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

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

40

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. 1.0.28) 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

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

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

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

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 [? ? ?], 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 Ref. [?].

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.*

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

We can absorb the factor resulting from variation of t_0 into the initial PDFs, f_0, so that $\exp\left(\kappa\gamma(t_0) - \tfrac{1}{2}\kappa^2\frac{d}{dt}\gamma(t_0) + \dots\right) f_0 \to f_0$. 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

$$
\overline{f}(\alpha_s(t+\kappa),\kappa) = \left[1 - \kappa\gamma(t+\kappa) + \tfrac{1}{2}\kappa^2(\gamma^2(t+\kappa) + \frac{d}{dt}\gamma(t+\kappa)) + \dots\right] f(t+\kappa)\,.
$$
$$(3.2.38)$$

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 [**?**]). 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.

For hadronic-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
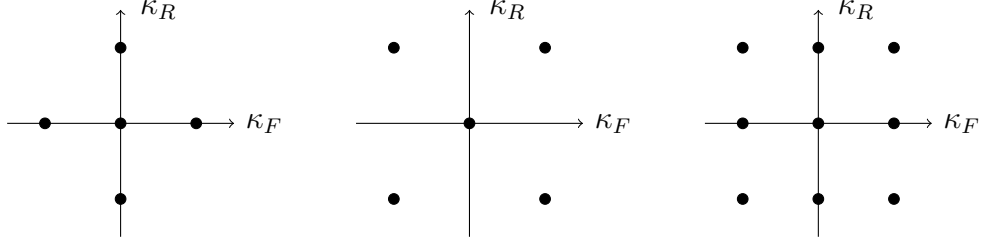
Figure 3.3.1 *Symmetric prescriptions for a single process, indicating the sampled values for the factorisation scale κ_F and renormalisation scale κ_R in each case. The origin of coordinates corresponds to the central scales $\kappa_F = \kappa_R = 0$. We show the three prescriptions 5-point (left), $\bar{5}$-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, 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

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

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

- **5̄-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 5̄-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

$$S_{i_1 j_1} = N_m \sum_{V_m} \Delta_{i_1}(\kappa_f, \kappa_{r_1})\Delta_{j_1}(\kappa_f, \kappa_{r_1}) = n_m \sum_{v_m} \Delta_{i_1}(\kappa_f, \kappa_{r_1})\Delta_{j_1}(\kappa_f, \kappa_{r_1}). \qquad (3.3.9)$$

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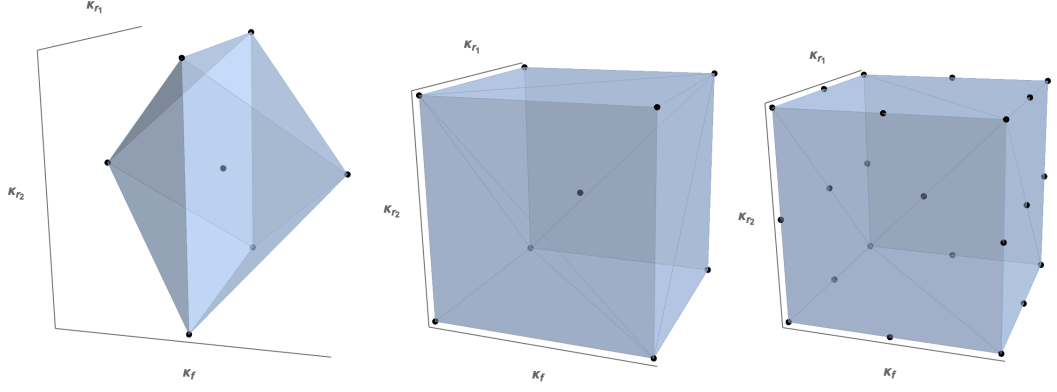
Figure 3.3.2 *Same as Fig. 3.3.1, now for the case of two different processes with a common factorisation scale, κ_F, and different renormalisation scales, κ_{R_1} and κ_{R_2}, so the diagrams are now in 3d. We again show the three prescriptions 5-point (left), $\bar{5}$-point (middle) and 9-point (right).*

- **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

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

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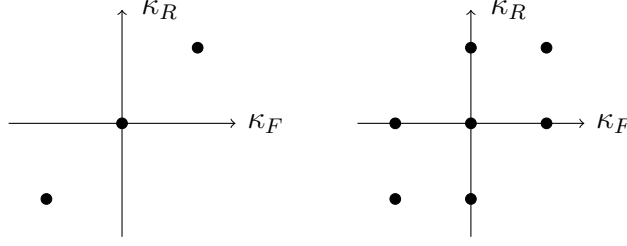
Figure 3.3.3 *Same as Fig. 3.3.1, now in the case of the asymmetric prescriptions for a single process with factorisation scale κ_F and renormalisation scale κ_R. We display the 3-point (left) and 7-point (right) prescriptions, defined in the text.*

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

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

  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

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

- **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, $v_6 = \{(\pm; 0), (0; \pm), (+; +), (-; -)\} = \{(\pm; 0), (0; \pm), (\overline{\pm; \pm})\}$, where $(\overline{\pm; \pm})$ 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

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

(3.3.17)

One process

Two processes

3.3.3 Categorising processes

3.3.4 NLO theory covariance matrices

3.4 Validating the theory covariance matrix

3.5 PDFs with missing higher order uncertainties

3.6 Impact on phenomenology

Chapter 4

Nuclear Uncertainties - 90%

4.1 Introduction

Parton distribution functions (PDFs) are universal quantities encapsulating the internal structure of the proton, and are crucial for making predictions in particle physics [32]. To maximally constrain them, PDFs are determined by fitting a range of experimental data over a wide variety of processes and kinematic regimes. Some of this data consists of measurements on nuclear targets, rather than proton targets. In this case, the surrounding nuclear environment will have an effect on the measured observables, which in turn will influence the form of the fitted PDFs. The uncertainties associated with these effects are termed "nuclear uncertainties". Such uncertainties are small [15][25] but becoming increasingly relevant with the advent of the Large Hadron Collider and the era of precision physics it has ushered in [12].

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

---

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

4.2 Nuclear Data

There are three experiments with nuclear targets currently included in NNPDF analyses: charged current inclusive deep inelastic scattering (DIS) cross sections from CHORUS [11], on Pb; DIS dimuon cross sections from NuTeV [14][61] on Fe; and Drell-Yan dimuon cross sections from E605 at Fermilab [6], on Cu. After cuts, nuclear data make up 993/4285 of the data points ($\sim 23\%$). For a complete summary of the data sets, see [16].

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

4.3 Determining Nuclear Uncertainties

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

We adopted an empirical approach to construct the nuclear uncertainties, using nPDFs rather than appealing to nuclear models, which rely on various assumptions [23]. We compared theoretical predictions for nuclear observables made with the correct corresponding nPDFs for an isotope "N", $T_i^N[f_N^{(n)}]$, to those with proton PDFs, $T_i^N[f_p]$. Here f_p is the central value for a proton PDF and $f_N^{(n)}$ is one Monte Carlo replica in an nPDF ensemble, where $n = 1, ..., N_{rep}$ [32]. To generate such an ensemble we combined three recent nPDF sets: DSSZ12 [44], nCTEQ15 [13] and EPPS16 [40]. Note that DSSZ12 does not provide a Cu PDF, so for the case of E605 we combined just two nPDF sets.

We considered two definitions of nuclear uncertainties:

1. **Def. 1,** (a conservative approach) where the only modification is to include
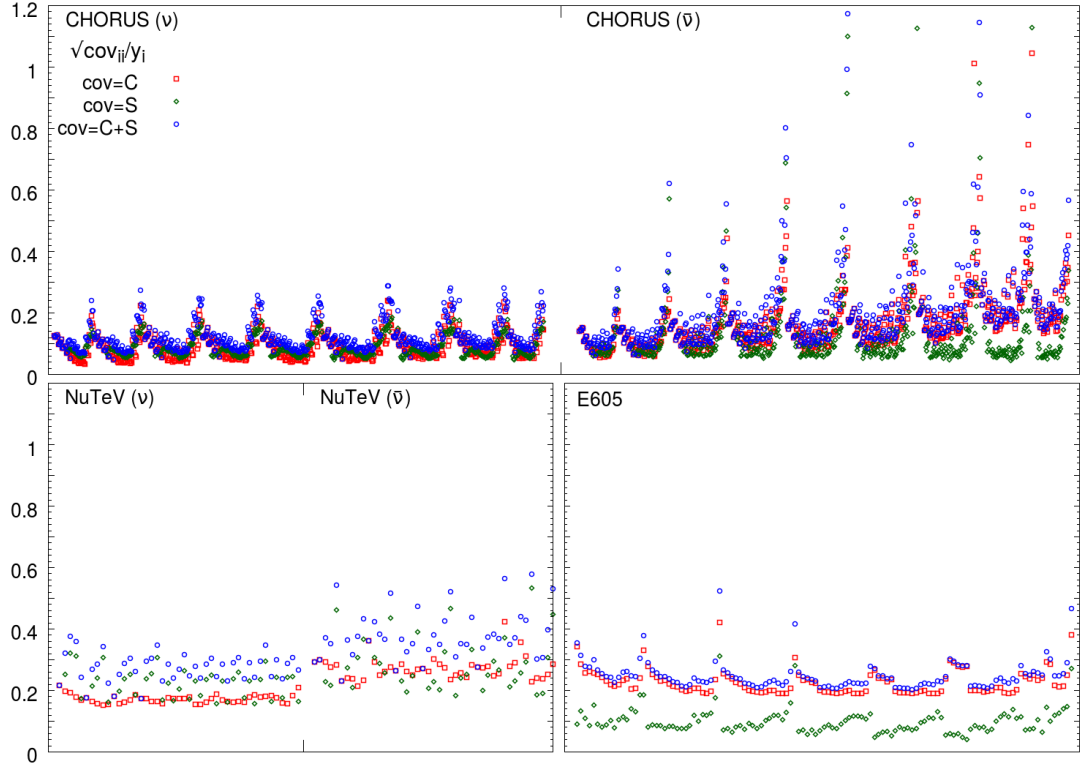
Figure 4.3.1 *The square root of the diagonal elements of the covariance matrices, normalised to corresponding data. Experimental contributions are red, theory green and the total blue. Data from CHORUS and NuTeV are split into neutrino and anti-neutrino parts. Points are binned in (anti-)neutrino beam energy E: 25, 35, 45, 55, 70, 90, 110, 120, 170 GeV. In each bin x increases from left to right, 0.045 < x < 0.65.*

nuclear uncertianties, with

$$\Delta_i^{(n)} = T_i^N[f_N^{(n)}] - T_i^N[f_p]; \tag{4.3.1}$$

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

$$\delta T_i^N = T_i^N[f_N] - T_i^N[f_p], \tag{4.3.2}$$

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

$$\Delta_i^{(n)} = T_i^N[f_N^{(n)}] - T_i^N[f_N]. \tag{4.3.3}$$

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

$$S_{ij} = \frac{1}{N_{rep}} \sum_{n=1}^{N_{rep}} \Delta_i^{(n)} \Delta_j^{(n)}. \tag{4.3.4}$$

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

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

4.4 The Impact on Global PDFs

We compared four different PDF fits:

- **Baseline,** based on NNPDF3.1, with small improvements [24];

- **NoNuc,** Baseline with nuclear data removed;

- **NucUnc,** Baseline with nuclear uncertainties according to Def. 1.

- **NucCor,** Baseline with nuclear uncertainties and a nuclear correction according to Def. 2.

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

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

---

[2] For a full break-down see [24].
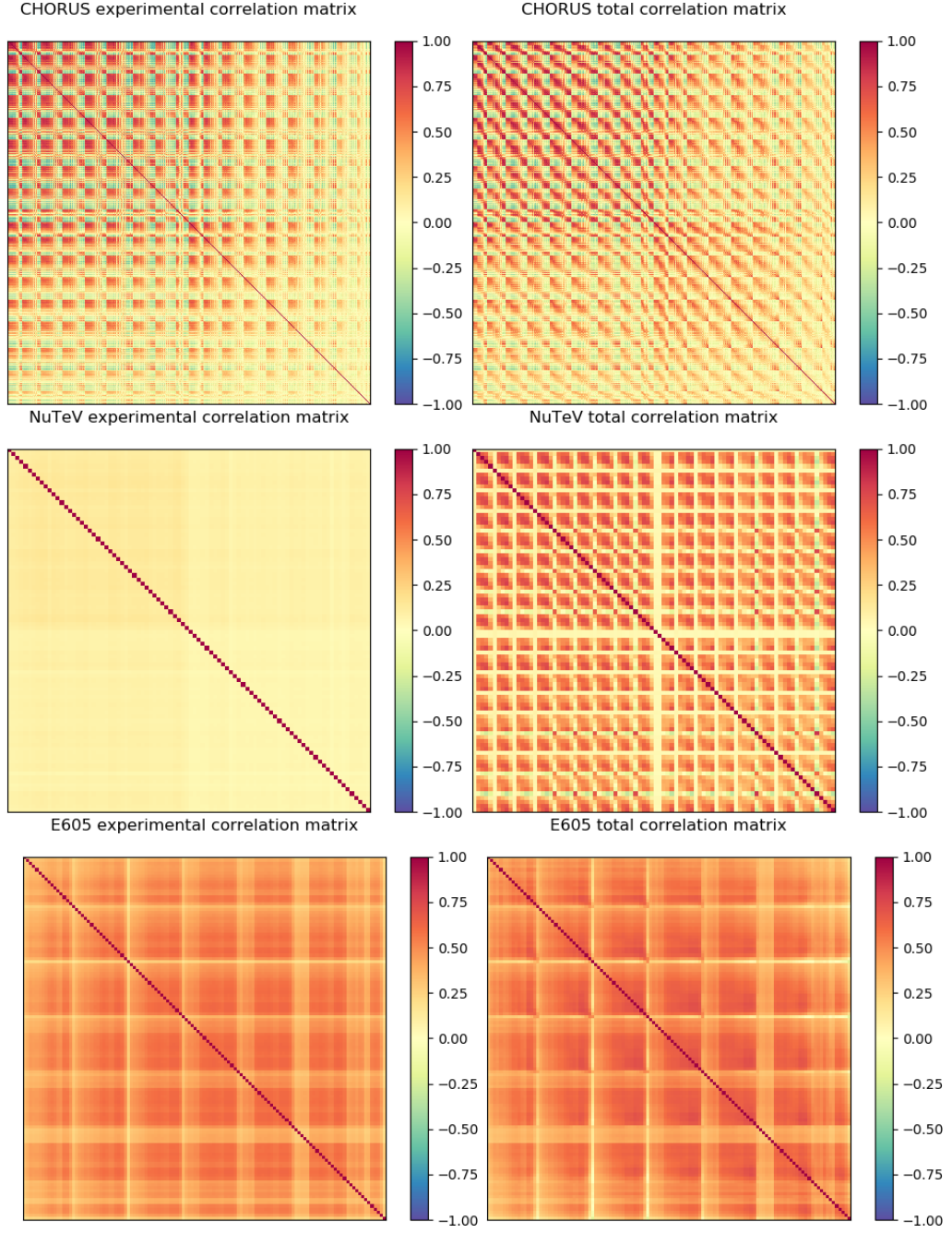
Figure 4.4.1 *Correlation matrices, $\rho_{ij}^{cov} = \frac{cov_{ij}}{\sqrt{cov_{ii}cov_{jj}}}$, before (left) and after (right) including nuclear uncertainties. Data are binned the same as in Fig. 4.4.2. The top row corresponds to CHORUS, middle row to NuTeV and bottom row to E605. Results are displayed for Def. 1 but are qualitatively similar for Def. 2.*

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

Table 4.4.1 *χ^2 per data point for selected data sets. The final row shows results for the full fitted data.*

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

4.5 Phenomenology

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

We found no appreciable difference between using NucUnc versus NucCor, so opt to incorporate uncertainties using NucUnc (Def. 1) as this is the more conservative option.
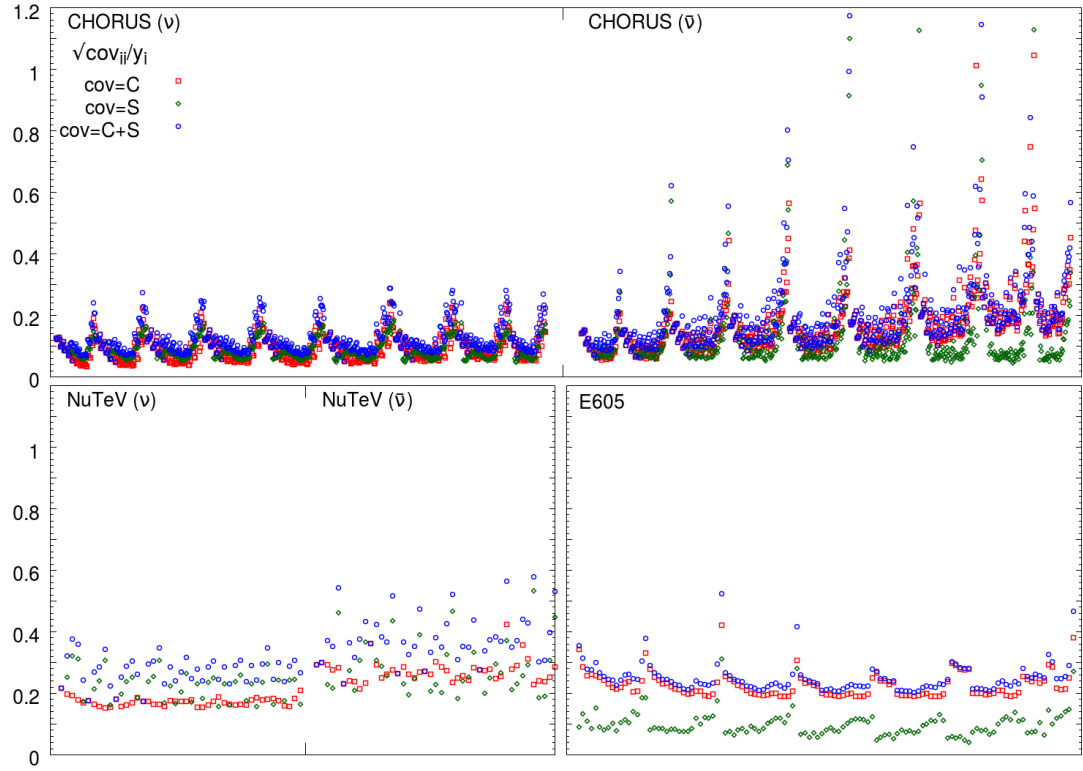
Figure 4.4.2 *The square root of the diagonal elements of the covariance matrices, normalised to corresponding data. Experimental contributions are red, theory green and the total blue. Data from CHORUS and NuTeV are split into neutrino and anti-neutrino parts. Points are binned in (anti-)neutrino beam energy E: 25, 35, 45, 55, 70, 90, 110, 120, 170 GeV. In each bin x increases from left to right, 0.045 < x < 0.65.*
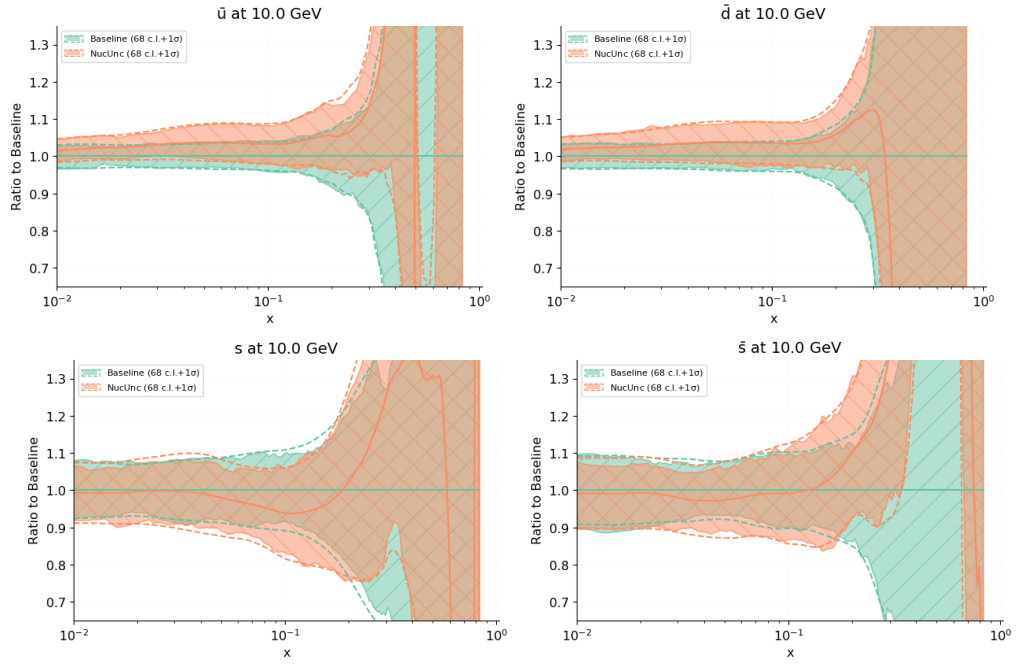
Figure 4.5.1 *NucUnc fits with nuclear uncertainties (orange) compared to Baseline (green) for PDFs at 10 GeV. Clockwise from top left: $\bar{u}$, $\bar{d}$, s and $\bar{s}$ PDFs. Error bands are 1 σ; results are normalised to Baseline fit.*

4.6   Conclusions

We studied the role of nuclear data in proton PDF fits, and adopted an empirical approach to determine the nuclear uncertainties due to this data. We based our analysis on recent nPDF fits DSSZ12, nCTEQ15 and EPPS16. Using a theoretical covariance matrix, we included these uncertainties in proton PDF fits, and found that the fit quality was improved, with the largest effect on the light sea quark distributions [3]. We found no significant impact on associated phenomenology.

We will extend this analysis to deterium data, and in the future we will be able to use nuclear PDFs from NNPDF [55] to estimate uncertainties. Furthermore, these methods can be applied to other sources of theoretical uncertainties, such as higher twist effects, fragmentation functions, and missing higher order uncertainties [?].

---

[3]The PDF sets from this analysis are available upon request from the authors in LHAPDF format [7].
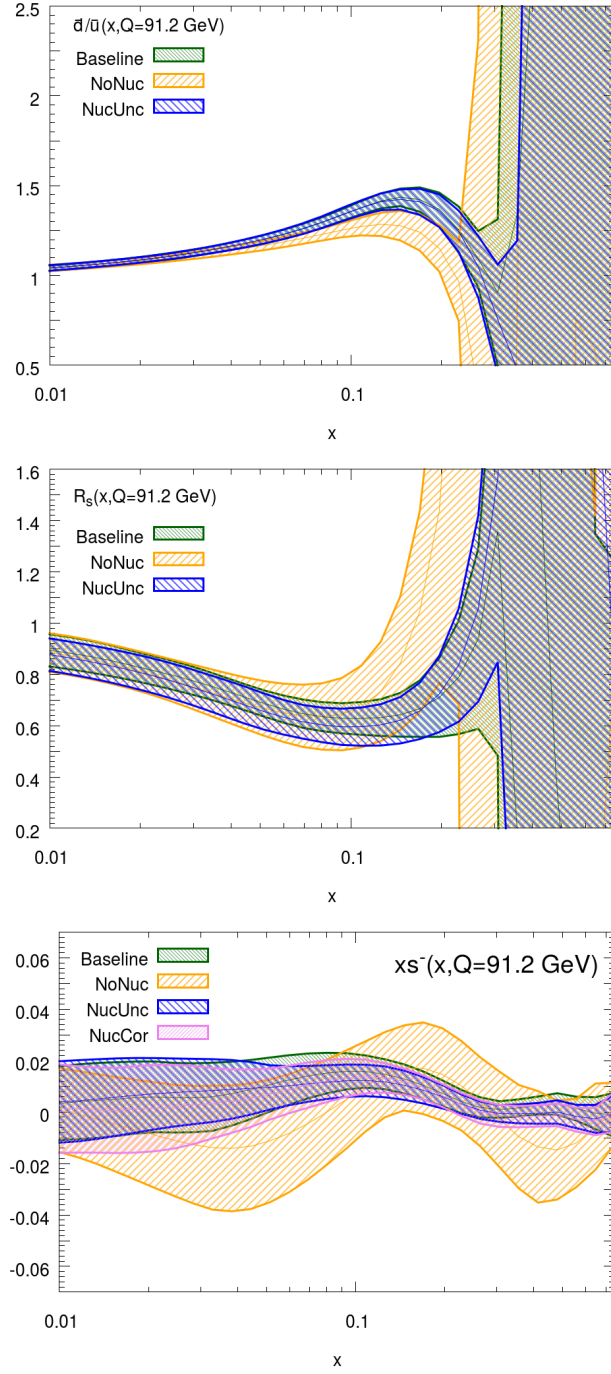
Figure 4.6.1 *Effect of including nuclear uncertainties on phenomenology. From left to right: sea quark asymmetry, strangeness fraction, strange valence distribution. Distributions correspond to the use of different PDF fits: Baseline (green), NoNuc (yellow), NucUnc (blue) and NucCor(pink). Q = 91.2 GeV. In the left two plots, NucCor are indistinguishable from NucUnc so are omitted for readability.*

Chapter 5

Deuteron Uncertainties - 0%

Chapter 6

Higher Twist Uncertainties - 0%

6.1   The role of higher twist data in PDFs

6.2   Ansatz for a higher twist correction

6.3   Using a neural network to model higher twist

6.3.1   Model architecture

6.3.2   Training and validating the neural network

6.4   Form of the higher twist correction

6.5   The higher twist covariance matrix

6.6   PDFs with higher twist uncertainties

Chapter 7

Conclusion - 0%

Appendix A

Diagonalisation of the theory covariance matrix

Appendix B

PDF sets with different scale choices

Bibliography

[1] , . https://nnpdf.mi.infn.it/.

[2] , . https://home.cern/.

[3] , . https://www.fnal.gov/.

[4] , . https://lhapdf.hepforge.org/.

[5] Aad, G., et al. "Measurement of the inclusive $W^{\pm}$ and Z/gamma cross sections in the electron and muon decay channels in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector." *Phys. Rev.* D85: (2012) 072,004.

[6] Aaltonen, T., et al. "Measurement of the Inclusive Jet Cross Section at the Fermilab Tevatron p anti-p Collider Using a Cone-Based Jet Algorithm." *Phys. Rev.* D78: (2008) 052,006. [Erratum: Phys. Rev.D79,119902(2009)].

[7] et al., A. B. "LHAPDF6: parton density access in the LHC precision era." *Eur. Phys. J.* C75: (2015) 132.

[8] et al., A. D. M. "Uncertainties on alpha(S) in global PDF analyses and implications for predicted hadronic cross sections." *Eur. Phys. J.* C64: (2009) 653–680.

[9] et al., G. A. "Determination of the strange quark density of the proton from ATLAS measurements of the $W \to \ell\nu$ and $Z \to \ell\ell$ cross sections." *Phys. Rev. Lett.* 109: (2012) 012,001.

[10] ———. "Measurement of the Z/γ^* boson transverse momentum distribution in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector." *JHEP* 09: (2014) 145.

[11] et al., G. O. "Measurement of nucleon structure functions in neutrino scattering." *Phys. Lett.* B632: (2006) 65–75.

[12] et al., J. G. "The Structure of the Proton in the LHC Precision Era." .

[13] et al., K. K. "nCTEQ15 - Global analysis of nuclear parton distributions with uncertainties in the CTEQ framework." *Phys. Rev.* D93, 8: (2016) 085,037.

[14] et al., M. G. "Precise measurement of dimuon production cross-sections in muon neutrino Fe and muon anti-neutrino Fe deep inelastic scattering at the Tevatron." *Phys. Rev.* D64: (2001) 112,006.

[15] et al., R. D. B. "Precision determination of electroweak parameters and the strange content of the proton from neutrino deep-inelastic scattering." *Nucl. Phys.* B823: (2009) 195–233.

[16] ———. "Parton distributions from high-precision collider data." *Eur. Phys. J.* C77, 10: (2017) 663.

[17] et al., S. C. "Vector boson production at hadron colliders: a fully exclusive QCD calculation at NNLO." *Phys. Rev. Lett.* 103: (2009) 082,001.

[18] et al., S. D. "New parton distribution functions from a global analysis of quantum chromodynamics." *Phys. Rev.* D93, 3: (2016) 033,006.

[19] et al., S. F. "Neural network parametrization of deep inelastic structure functions." *JHEP* 05: (2002) 062.

[20] et al., V. K. "Measurement of the Z boson differential cross section in transverse momentum and rapidity in proton–proton collisions at 8 TeV." *Phys. Lett.* B749: (2015) 187–209.

[21] Altarelli, G., and G. Parisi. "Asymptotic Freedom in Parton Language." *Nucl. Phys. B* .

[22] Alwall, J., R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations." *JHEP* 07: (2014) 079.

[23] Arneodo, M. "Nuclear effects in structure functions." *Phys. Rept.* 240: (1994) 301–393.

[24] Ball, R. D., and A. Deshpande. "The Proton Spin, Semi-Inclusive processes, and a future Electron Ion Collider." https://inspirehep.net/record/1648159/files/arXiv:1801.04842.pdf.

[25] Ball, R. D., V. Bertone, L. Del Debbio, S. Forte, A. Guffanti, J. Rojo, and M. Ubiali. "Theoretical issues in PDF determination and associated uncertainties." *Phys. Lett.* B723: (2013) 330–339.

[26] Ball, R. D., L. Del Debbio, S. Forte, A. Guffanti, J. I. Latorre, A. Piccione, J. Rojo, and M. Ubiali. "A Determination of parton distributions with faithful uncertainty estimation." *Nucl. Phys.* B809: (2009) 1–63. [Erratum: Nucl. Phys.B816,293(2009)].

[27] Ball, R. D., L. Del Debbio, S. Forte, A. Guffanti, J. I. Latorre, J. Rojo, and M. Ubiali. "A first unbiased global NLO determination of parton distributions and their uncertainties." *Nucl. Phys. B* 838: (2010) 136–206.

[28] ———. "Fitting Parton Distribution Data with Multiplicative Normalization Uncertainties." *JHEP* 05: (2010) 075.

[29] Ball, R. D., et al. "Parton distributions with LHC data." *Nucl. Phys. B* 867: (2013) 244–289.

[30] ———. "Parton distributions for the LHC Run II." *JHEP* 04: (2015) 040.

[31] Botje, M. "Lecture notes Particle Physics II: Quantum Chromo Dynamics.", 2013. https://www.nikhef.nl/~h24/qcdcourse/section-all.pdf.

[32] Butterworth, J., et al. "PDF4LHC recommendations for LHC Run II." *J. Phys.* G43: (2016) 023,001.

[33] Callan, J., Curtis G., and D. J. Gross. "Bjorken scaling in quantum field theory." *Phys. Rev. D* 8: (1973) 4383–4394.

[34] Campbell, J., and R. Ellis. "An update on vector boson pair production at hadron colliders." *Phys. Rev.* .

[35] Cooper-Sarkar, A. M. "PDF Fits at HERA." *PoS* EPS-HEP2011: (2011) 320.

[36] D'Agostini, G. "On the use of the covariance matrix to fit correlated data." *Nucl. Instrum. Meth. A* 346: (1994) 306–311.

[37] Dokshitzer, Y. L. *Sov. Phys.* .

[38] Dothan, Y., M. Gell-Mann, and Y. Ne'eman. "Series of Hadron Energy Levels as Representations of Noncompact Groups." *Phys. Lett.* 17: (1965) 148–151.

[39] Ellis, R. K. R. K. *QCD and collider physics.* Cambridge monographs on particle physics, nuclear physics, and cosmology ; 8. Cambridge ; New York: Cambridge University Press, 1996.

[40] Eskola, K. J., P. Paakkinen, H. Paukkunen, and C. A. Salgado. "EPPS16: Nuclear parton distributions with LHC data." *Eur. Phys. J.* C77, 3: (2017) 163.

[41] Feynman, R. P. "Very high-energy collisions of hadrons." *Phys. Rev. Lett.* 23: (1969) 1415–1417.

[42] Feynman, R. "The behavior of hadron collisions at extreme energies." *Conf. Proc. C* 690905: (1969) 237–258.

[43] ———. "Photon-hadron interactions." .

[44] de Florian, D., R. Sassot, P. Zurita, and M. Stratmann. "Global Analysis of Nuclear Parton Distributions." *Phys. Rev.* D85: (2012) 074,028.

[45] G. Aad, G. e. a. "Measurement of the transverse momentum and ϕ_η^* distributions of Drell–Yan lepton pairs in proton–proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector." *Eur. Phys. J.* C76, 5: (2016) 291.

[46] Gavin, R., Y. Li, F. Petriello, and S. Quackenbush. "FEWZ 2.0: A code for hadronic Z production at next-to-next-to-leading order." *Comput. Phys. Commun.* 182: (2011) 2388–2403.

[47] Geiger, D. H., and E. Marsden. "LXI. The laws of deflexion of a particles through large angles." *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 25, 148: (1913) 604–623. https://doi.org/10.1080/14786440408634197.

[48] Gell-Mann, M. "Symmetries of baryons and mesons." *Phys. Rev.* 125: (1962) 1067–1084.

[49] ———. "A Schematic Model of Baryons and Mesons." *Phys. Lett.* 8: (1964) 214–215.

[50] Greenberg, O. "Spin and Unitary Spin Independence in a Paraquark Model of Baryons and Mesons." *Phys. Rev. Lett.* 13: (1964) 598–602.

[51] Gribov, L. N., and V. N. Lipatov. *Sov. J. Nucl. Phys* .

[52] Grinstein, B. "Introductory lectures on QCD." .

[53] Group, P. D. "Review of Particle Physics." *Progress of Theoretical and Experimental Physics* 2020, 8. https://doi.org/10.1093/ptep/ptaa104. 083C01.

[54] Halzen, F. F. *Quarks and leptons : an introductory course in modern particle physics.* New York] ; [Chichester]: John Wiley, 1984.

[55] Khalek, R. A., J. J. Ethier, and J. Rojo. "Nuclear Parton Distributions from Neural Networks." In *Diffraction and Low-x 2018 (Difflowx2018) Reggio Calabria, Italy, August 26-September 1, 2018.* 2018.

[56] Kinoshita, T. "Mass singularities of Feynman amplitudes." *J. Math. Phys.* 3: (1962) 650–677.

[57] Lee, T., and M. Nauenberg. "Degenerate Systems and Mass Singularities." *Phys. Rev.* 133: (1964) B1549–B1562.

[58] Nocera, E. R., M. Ubiali, and C. Voisey. "Single Top Production in PDF fits." *JHEP* 05: (2020) 067.

[59] S. Alekhin, J. B., and S. Moch. "The ABM parton distributions tuned to LHC data." *Phys. Rev.* D89, 5: (2014) 054,028.

[60] S. Catani, S., and M. H. Seymour. "A General algorithm for calculating jet cross-sections in NLO QCD." *Nucl. Phys.* B485: (1997) 291–419. [Erratum: Nucl. Phys.B510,503(1998)].

[61] Tzanov, M., et al. "Precise measurement of neutrino and anti-neutrino differential cross sections." *Phys. Rev.* D74: (2006) 012,008.

[62] Zweig, G. *An SU(3) model for strong interaction symmetry and its breaking. Version 2*, 1964, 22–101.