



UNIVERSITÀ DEGLI STUDI DI MILANO

Scuola di Dottorato in Fisica, Astrofisica e Fisica Applicata

Dipartimento di Fisica

Corso di Dottorato in Fisica, Astrofisica e Fisica Applicata

Ciclo XXXVII

Advancements in PDFs determination

Incorporation of QED effects and
new theoretical improvements in a
modern deep learning fitting framework

Settore Scientifico Disciplinare FIS/02

Supervisor: Prof. Stefano Forte

Coordinator: Prof. Aniello Mennella

Ph.D. Thesis of: Niccolò Laurenti

Anno Accademico 2023-2024

Commission of the final examination:

External Referees:

Emanuele Re
Ramon Winterhalder

External Members:

Fabio Maltoni
Katarzyna Wichmann

Internal Member:

Marco Zaro

Final examination:

November 8th, 2024

Università degli Studi di Milano, Dipartimento di Fisica, via Giovanni Celoria 16, Milano, Italy

MIUR subjects:

FIS/02 - Fisica Teorica, Modelli e Metodi Matematici

Cover illustration:

Generated through the public AI image generator Foocus¹ with the input "Draw the photon content of the proton"

¹Go to https://colab.research.google.com/github/l1lyasviel/Foocus/blob/main/foocus_colab.ipynb to play with the code.

To my Mother, source of inspiration

Abstract

The focus of this Ph.D. thesis is the improvement in the determination of parton distribution functions (PDFs) through the inclusion of new theoretical effects in the theoretical predictions that are compared to experimental data. Such effects are the inclusion of Quantum Electrodynamics (QED) corrections, the inclusion of the so-called missing higher orders uncertainties (MHOUs) and the inclusion of approximate next-to-next-to-next-to-leading order contributions (aN³LO). This work aims to overcome the NNPDF4.0 PDFs determination, that did not include such effects and whose methodology is at the basis of this analysis. We first present the methodology used in the NNPDF4.0 PDFs fit. Then, we explain how the QED effects have been added in the NNPDF4.0 fitting framework, describing the details of the implementation, the resulting PDFs determination and the impact on physical processes of such PDFs. Moreover, we explain how the MHOUs and aN³LO effects have been included in NNPDF4.0, presenting the results of the two effects both separately and combined. We also show how the QED effects have been combined to MHOUs and aN³LO and we will present the results of the PDFs determination that includes all the three effects. In conclusion, we will show an application to Monte Carlo event generator of the implementation of the QED effects and of the development of a more flexible and efficient theory pipeline for the computation of the theoretical predictions that has been interfaced to the NNPDF4.0 code and is at the basis of all the work described in this thesis.

Contents

Introduction	xiii
Part I : Prologue	3
1 Introduction on the theory of Strong Interactions	3
1.1 Basics of Quantum Chromodynamics	4
1.2 Renormalization	6
1.3 Running of the strong coupling	7
1.4 Parton model	9
1.5 Factorization of collinear divergences	12
1.6 DGLAP evolution equations	14
1.7 Factorization of large logarithms	22
1.8 Variable flavor number scheme	25
1.9 Matching and basis rotation	27
2 NNPDF4.0 fitting framework	29
2.1 Methodology	30
2.2 Dataset	37
2.3 New computational tools	38
Part II : A Tale of Light and Shadow	49
3 Including QED corrections	49
3.1 Photon PDF	50
3.2 DGLAP evolution	57
3.3 Benchmarking the solution	64
3.4 Results of the QED fit	67
3.5 Impact on LHC phenomenology	74
4 Combining QED with MHO and N³LO	83
4.1 MHO	83
4.2 N ³ LO	95
4.3 Combination	112

Part III : Epilogue	123
5 Applications to Monte Carlo event generators	123
5.1 Methodology	124
5.2 The NNPDF4.0MC PDFs	127
5.3 Impact on LHC physics	129
Summary and Conclusions	139
Appendices	139
A Attempts to construct a truncated solution of DGLAP equations in presence of two running couplings	141
B NNPDF4.0 truncated vs exact solution	145
C Dependence of the PDFs on the fitting scale	151
Bibliography	159
List of Publications	177
Acknowledgments	180

Introduction

*“Please allow me to introduce myself
I’m a man of wealth and taste”*

– The Rolling Stones, Sympathy For The Devil

Motivation

To explore physics beyond the Standard Model (BSM), it is essential to achieve a high understanding in the description of the Standard Model (SM) processes, both in experimental measurements and in theoretical predictions. Indeed, if one of the two did not have enough precision, in presence of data that slightly deviate from the SM theoretical prediction, we wouldn’t be able to understand whether this deviation is due to a BSM effect or to our poor knowledge, either from the experimental or the theoretical point of view, of the SM physics. For this reason, high-energy experimental physics is moving towards a phase of high precision [1–3]. In fact, during the High Luminosity LHC (HL-LHC) era, the goal will be to increase the LHC’s luminosity by a factor of 10 beyond its current level. This will significantly enhance the precision of experimental measurements, with the aim of achieving accuracy on the order of percent or better. Consequently, theoretical predictions will need to match this level of precision to enable meaningful comparisons between theory and experiments.

As we will discuss in the following chapters, in high-energy physics the hadronic cross sections are computed using the factorization theorem. This involves a convolution between partonic cross sections, where the external particles are the hadron’s elementary constituents (quarks and gluons) rather than the hadron itself, and the parton distribution functions (PDFs), which provide information about the hadron’s internal structure. The PDFs currently represent a significant source of uncertainty in theoretical predictions. Therefore, achieving a better accuracy in PDFs determination is essential in order to reach the percent accuracy in the theory predictions and therefore to be able to compare theory and experiments.

In the SM the observables are computed in perturbation theory as a power series in a certain coupling (there can be even more couplings at the same time). For example, in the case of the strong interaction, that is the one governing the internal structure of hadrons, a generic observable \mathcal{O} is computed as

$$\mathcal{O} = \mathcal{O}^{(0)} + \alpha_s \mathcal{O}^{(1)} + \alpha_s^2 \mathcal{O}^{(2)} + \alpha_s^3 \mathcal{O}^{(3)} + \dots, \quad (1)$$

where α_s is the strong coupling, that will be defined in Chap. 1. The first term of Eq. (1) is usually called leading order (LO), the second next-to-leading order (NLO), the third next-to-next-to-leading order (NNLO) and so on. As we will see in Chap. 1, truncating Eq. (1) up to a certain order and neglecting the higher orders is a good approximation only at scales much larger than the typical hadronization scale, i.e. 1 GeV. It means that the PDFs, that describe the internal structure of hadrons, cannot be computed in perturbation theory. Therefore, they must be extracted from data. Such procedure is called PDFs fit. In order to perform a PDFs fit three main ingredients are needed: a dataset, a methodology and certain theory settings.

The dataset is the set of all the experimental measurements that we want to use to compare experiments with theory predictions. They can be of two types: electron-proton scattering and proton-proton collisions. The first ones are very clean processes both from the experimental and the theoretical point of view, since they involve only one hadron in the initial state. For this reason they are easier to be studied theoretically and to be reconstructed experimentally. Moreover, the presence of only one hadron in the initial state results in the presence of only one PDF in the factorization theorem, and therefore provides a strong constraint in the fit. The proton-proton collisions instead involve two PDFs in the computation of the theory predictions, so finding the best configuration in the PDFs space is more complicated. However, they are very precise data that must be included in the fit in order to increase the accuracy of the PDFs determination. Moreover, all these experimental measurements are associated with an error that must be propagated into the fit, as it will be discussed in Chap 2.

The methodology is the way the PDFs are extracted from experimental data. In other words, it is the way the theory predictions are compared to the experimental measurements to find the best configuration for the PDFs that provide a faithful description of data. In 2021 the NNPDF collaboration released the NNPDF4.0 PDFs set [4], in which a lots of technical improvements in the fitting methodology were introduced with respect to its predecessor NNPDF3.1 [5]. Indeed, the new analysis is based on a new fitting algorithm based on stochastic gradient descent and it includes a systematic implementation of both positivity constraints and integrability of sum rules. Moreover, the new fitting algorithm is based on a semi-automatic selection of the features of the algorithm, called *hyperoptimization*. All these improvements aim to reduce the uncertainty associated to the fitting procedure itself. The NNPDF4.0 fitting framework is at the basis of the analyses described in this thesis. For this reason it will be briefly described in Chap. 2. For a more detailed description see Ref. [4].

The last ingredient of a PDFs fit is given by the theory settings used in the extraction. Indeed, the partonic cross sections are computed up to a given order in perturbation theory. Moreover, the dependence of the PDFs on the hard scale of the considered observable is given by an integro-differential equation that depends on the so-called splitting functions that are computed in perturbation theory. The perturbative order that we use in the theory predictions of the dataset used in the PDFs fit is one of the theory settings that define a fit. Other settings that we can choose are the inclusion of Quantum Electrodynamics (QED) correction, the inclusion of Electroweak (EW) effects, the use of theory uncertainties and so on. The NNPDF4.0 analysis is a NNLO PDFs fit, i.e. it uses the partonic cross section and the splitting functions at NNLO. Moreover, it uses pure Quantum Chromodynamics (QCD) predictions, i.e. it neglects QED effects, and it uses only experimental errors in the fitting procedure. However, this level of accuracy in the theoretical calculations is not enough if we want to reach the percent accuracy in the PDFs determination.

The subject of this thesis is the improvement of the accuracy of the theory predictions

in order to reach the percent accuracy needed to obtain a high precision PDFs determination that can match the accuracy of the experimental data in the HL-LHC era. Such improvement is carried out in three independent ways: through the inclusion of QED corrections in the PDFs fit, the inclusion of the so-called missing higher orders uncertainties (MHOU) in the theory predictions and the inclusion of approximate next-to-next-to-next-to-leading order (N^3 LO) corrections in electron-proton scattering observables and in the PDFs evolution equations. Each of these effects have been implemented independently and produced three different PDFs sets that have been compared against NNPDF4.0. Moreover, in the end they have been combined to produce another PDFs set that has been compared with the three of them and with NNPDF4.0. Let's briefly discuss each of them.

- When we include QED corrections in a PDFs fit we allow the quarks inside the protons to split into a quark-photon pair. For this reason we must consider a photon component of the proton, i.e. a photon PDF. Since this type of splitting is governed by the QED coupling, the photon PDF will be suppressed with respect to the other PDFs and therefore it will be a small correction. However, we will see that there are regions in which these effects are not negligible, i.e. their size is larger than the percent accuracy that we aim to reach. It means that we have to include such corrections to improve the precision of PDFs fits.
- When we compute an observable in perturbation theory using Eq. (1) we are introducing an error in the theory predictions that is related to the fact that we are truncating the series at a given order instead of using all the (infinite) terms. Obviously, as we go up with perturbative orders, this error becomes smaller and smaller but still of finite size. This error is what we call MHOU. Therefore, first we have to associate a MHOU to the theory predictions, and this will be done with the technique of scale variations, and then we have to include this additional uncertainty in the fitting algorithm. This last step will be done with the construction of a theory covariance matrix that will be combined with the experimental one.
- At NNLO, all the ingredients needed for a PDFs fit are currently known. Such ingredients are the theory predictions of electron-proton scattering, processes from proton-proton collision (that are known at NNLO through the so-called K -factor) and the splitting functions needed for the evolution equations. At N^3 LO instead, due to the complexity of the fixed order calculations, only partial information is currently available in the literature. For this reason we used all the terms that are known at N^3 LO and for the ones that are still missing we used approximations that are based on the information that are currently available.

In this thesis, these three fits will be referred to as NNPDF4.0QED [6], NNPDF4.0MHOU [7] and aN^3 LO NNPDF4.0 [8], respectively. The inclusion of these contributions will lead to an improvement in the precision of the theory predictions that propagates into an improvement in the accuracy of the PDFs determination.

Thesis overview

This thesis is organized as follows:

Chapter 1: Introduction on the theory of strong interactions.

In this chapter we will briefly overview the theory of strong interactions. We will start by writing down the Lagrangian of QCD, describing all the terms that compose

it. Then, we will give some details about renormalization and we will explain how it is related to the running of the strong coupling. Moreover, we will overview the parton model, that is the theory describing the collision between composite particles, i.e. the hadrons. We will also explain the factorization of collinear divergences and we will show how it causes the running of the PDFs with the hard scale of the process we are considering. This will lead us to the so-called DGLAP equations, whose solution will be also presented. Next, we will describe the insurgence of the large logarithms, how these terms spoil the perturbative convergence of the observables, and how this problem is solved. We will also show how this is related to the construction of a so-called variable flavor number scheme.

Chapter 2: Description of the NNPDF4.0 fitting framework.

In this chapter we will describe how the PDFs are extracted from experimental data in the NNPDF4.0 methodology. We will explain how the PDFs are parametrized at the fitting scale, and how the fit is performed, i.e. we will write down the χ^2 used in the fit and we will briefly describe the stopping algorithm. We will also give some details on the numerical techniques used to compare theory predictions with experimental data in an efficient way. Moreover, we will explain how the experimental uncertainties are propagated to the PDFs in the fitting procedure. Next, we will describe the experimental data used in the fit. In the end we will briefly present new computational tools used to automatize the computation of the theory predictions, and we will show that they give the same results with respect to the ones used for the NNPDF4.0 determination.

Chapter 3: Inclusion of QED corrections in PDFs fits.

In this chapter we will describe how the QED effects are included into a PDFs fit and how it has been interfaced with the NNPDF4.0 fitting framework. We will start with a brief overview of the LuxQED approach, that is used to determine the photon PDF, and we will describe how it has been used in order to perform a QED fit. We will describe how PDFs evolution equations are changed in presence of QED effects and how the solution of such equations is modified in this case. In this chapter we will also benchmark the solution of the QED DGLAP equations that we implemented, with the one implemented in the `APFEL` code, that uses a different solution method. We will also compare the solution in presence of QED corrections with the pure QCD one. Then we will show the results of the QED fit using the NNPDF4.0 methodology, and we will compare the results both with the pure QCD counterpart, i.e. NNPDF4.0, and with the most recent QED fits. In the end, we will show the impact of QED corrections on the physical processes of interests for the LHC studies.

Chapter 4: Inclusion of MHOU and N^3 LO corrections and their combination with QED effects.

In this chapter we will describe how the QED effects are combined with the inclusion of MHOU and N^3 LO corrections. We will start by reviewing how MHOU is included into a PDFs fit through the construction of a theory covariance matrix and how it is combined with the experimental one. Then, we will give the results of the pure QCD fit that includes the MHOU and we will compare it with the baseline fit, i.e. NNPDF4.0. Next, we will show how N^3 LO corrections have been included when they are known, and the approximation that has been used for the terms that are still missing. We will show the results of the fit, we will compare it with NNPDF4.0 and we will combine it with the inclusion of MHOU. In the end, we will show how the QED effects are combined both with MHOU, with N^3 LO and with the combination

of the two. We will compare the results with the pure QCD MHOU and N³LO fits and we will compare the photon PDFs obtained in the different cases. In the end we will briefly assess the impact on the phenomenology of the different fits.

Chapter 5: Applications to Monte Carlo event generators.

In this chapter we will apply the implementation of a new theory pipeline and the one of QED effects in the NNPDF4.0 code, to the construction of PDFs sets that are suitable for being used in Monte Carlo event generators. We will describe which are the features that such PDFs must satisfy, we will describe how the QED effects have been included, and we will present the results that we obtained, comparing them with the standard NNPDF4.0 fit. We will also briefly investigate the impact of such modified PDFs to the LHC phenomenology.

Part I
Prologue

Introduction on the theory of Strong Interactions

“We are only as strong as we are united, as weak as we are divided.”

– Albus Dumbledore, Harry Potter and the Goblet of Fire

The Standard Model (SM) is the theory describing the fundamental interactions between the elementary particles and it is based on the gauge symmetry group

$$SU(3)_C \otimes SU(2)_W \otimes U(1)_Y. \quad (1.1)$$

The part we will discuss in this chapter is $SU(3)_C$, that is the part that describes the so-called *strong interactions*. It is called Quantum Chromodynamics (QCD) and it is a quantum field theory based on the non Abelian symmetry group $SU(3)$ whose charge is called color.

The fundamental fields of QCD are the ones describing the quarks and the gluons: the quarks are spin 1/2 particles that belong to the fundamental representation of $SU(3)$ (i.e. the triplet) and therefore they appear with three possible colors. In addition to the quarks there are their anti-particles that are the anti-quarks: they are particles that have the same properties of their corresponding quark but opposite charge. They belong to the anti-fundamental representation of $SU(3)$ (i.e. the anti-triplet). The different kind of quarks and anti-quarks are the so-called *flavors*. Being the strong interaction independent on the flavor, the differences in the processes involving different quarks come only from the kinematics due to the different masses of the flavors. In the end there are the gluons, that are the gauge bosons of the strong interaction and therefore belong to the adjoint representation of $SU(3)$ (i.e. the octet). It means that they appear with eight possible colors.

Two fundamental properties of QCD that distinguishes it from QED and that will be further explained in this chapter are the *color confinement* and *asymptotic freedom*: color confinement is the property that quarks and gluons cannot be observed in the final state since only colorless particles, i.e. the hadrons, can be detected experimentally. This makes the computation of hadronic cross sections impossible in the context of perturbative QCD because of the the failure of the series expansion at scales of the order of the hadron masses. This problem is solved with the introduction of the *parton model* that will be introduced in Sec. 1.4. The asymptotic freedom is the property that the strong coupling decreases at high energies. Indeed, if we call Λ the typical scale of the hadronic physics, i.e. $\Lambda \sim 1$ GeV, then the strong coupling is small for energies much larger than Λ , while it is of $\mathcal{O}(1)$ at energies of the order of Λ . This means that we can obtain reasonable predictions only at high energies, since the perturbative expansion is well behaved,

being every term much smaller than the previous ones. Instead, at small energies perturbation theory breaks down since every term is comparable (or even larger) with the previous ones and therefore we should include an infinite number of terms in the series. It means that at small energies the theory is no longer predictive.

In this chapter we will describe the basics of the theory of the strong interactions. In Sec. 1.1 we will introduce the lagrangian of QCD and we will explain all the possible terms allowed by the symmetries, in Sec. 1.2 we will give some details about renormalization explaining its consequences and in Sec. 1.3 we will describe the running of the strong coupling. In Sec. 1.4 we will describe the parton model, that is needed in order to have predictions with composite initial states. In Sec. 1.5 we will explain how to deal with the collinear divergences coming from initial state emissions and in Sec. 1.6 we will describe how this procedure leads to evolution equations. In Sec. 1.7 we will describe how to treat large logarithms coming from the mass of the heavy quarks and in the end in Sec. 1.8 we will see how this applies when we deal with very different energy scales. In the end in Sec. 1.9 we will describe how the so-called matching conditions are applied to the PDFs, when a heavy quark threshold is crossed.

1.1 Basics of Quantum Chromodynamics

The strong interaction is described by the QCD lagrangian that reads

$$\mathcal{L}_{\text{QCD}} = \mathcal{L}_{\text{gauge}} + \mathcal{L}_{\text{CP}} + \mathcal{L}_{\text{gauge}}^{\text{fixing}} + \mathcal{L}_{\text{ghosts}}. \quad (1.2)$$

Let's see Eq. (1.2) term by term. The first contribution $\mathcal{L}_{\text{gauge}}$ contains the mass terms of the quarks, the kinetic terms of both quarks and gluons and the quark-gluon and the gluon-gluon interaction terms. It reads

$$\mathcal{L}_{\text{gauge}} = \sum_{f=\text{flavors}} \bar{\psi}_i^{(f)} \left(i \not{D}_{ij} - m^{(f)} \delta_{ij} \right) \psi_j^{(f)} - \frac{1}{4} F_{\mu\nu}^a F^{a,\mu\nu}, \quad (1.3)$$

where

$$D_{\mu,ij} = \delta_{ij} \partial_\mu - i g_s T_{ij}^a A_\mu^a, \quad (1.4)$$

$$F_{\mu\nu}^a = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a + g_s f^{abc} A_\mu^b A_\nu^c. \quad (1.5)$$

In Eqs. (1.3-1.5) $\psi_i^{(f)}$ and A_μ^a are respectively the quark and gluon fields, T_{ij}^a are the Gell-Mann matrices [9], i.e. the generators of the fundamental representation of $SU(3)$, and f^{abc} are the structure constant of $SU(3)$, i.e. they satisfy

$$[T^a, T^b] = i f^{abc} T^c, \quad a, b, c = 1, \dots, 8. \quad (1.6)$$

Moreover, the Gell-Mann matrices satisfy

$$\text{Tr} (T^a T^b) = T_R \delta^{ab}, \quad T_R = \frac{1}{2}, \quad (1.7)$$

$$T_{ik}^a T_{kj}^a = C_F \delta_{ij}, \quad C_F = \frac{4}{3}, \quad (1.8)$$

$$(T_{\text{adj}}^a)_{bc} (T_{\text{adj}}^a)_{cd} = C_A \delta_{bd}, \quad C_A = 3, \quad (1.9)$$

where

$$(T_{\text{adj}}^a)_{bc} = -i f^{abc}, \quad (1.10)$$

is the generator of the *adjoint representation*, i.e. the octet. We have that the indices are $i, j=1, 2, 3$ and $a, b, c, d=1, 2, \dots, 8$, i.e. they are the color indices respectively of the triplet and of the octet. The sum over repeated indices is understood. Inserting Eqs. (1.4-1.5) into Eq. (1.3) we find

$$\begin{aligned} \mathcal{L}_{\text{gauge}} = & \sum_f \left(i\bar{\psi}_i^{(f)} \not{\partial} \psi_i^{(f)} - m^{(f)} \bar{\psi}_i^{(f)} \psi_i^{(f)} + g_s \bar{\psi}_i^{(f)} \gamma^\mu T_{ij}^a A_\mu^a \psi_j^{(f)} \right) \\ & + \frac{1}{2} A^{a,\mu} (\eta_{\mu\nu} \partial^2 - \partial_\mu \partial_\nu) A^{a,\nu} - g_s f^{abc} A^{a,\mu} A^{b,\nu} \partial_\mu A_\nu^c \\ & - \frac{1}{4} g_s^2 f^{abe} f^{cde} A^{a,\mu} A^{b,\nu} A_\mu^c A_\nu^d. \end{aligned} \quad (1.11)$$

where γ^μ are the Dirac matrices [10].

The next term in Eq. (1.2) is the CP violation term, that reads

$$\mathcal{L}_{\text{CP}} = \theta \frac{\alpha_s}{4\pi} \tilde{F}_{\mu\nu}^a F^{a,\mu\nu}, \quad (1.12)$$

where θ is a dimensionless parameter, $\tilde{F}_{\mu\nu}^a$ is defined as

$$\tilde{F}_{\mu\nu}^a = \frac{1}{2} \epsilon_{\mu\nu\rho\sigma} F^{a,\rho\sigma}, \quad (1.13)$$

and we used that

$$\alpha_s = \frac{g_s^2}{4\pi}. \quad (1.14)$$

α_s is the coupling of QCD, also called strong coupling. This term violates CP because of the completely antisymmetric tensor $\epsilon_{\mu\nu\rho\sigma}$. However, CP violation in the strong sector has never been observed in experiments and this gives an experimental bound on θ such that $\theta \lesssim 10^{-10}$.

Then we have the gauge fixing term

$$\mathcal{L}_{\text{fixing}}^{\text{gauge}} = -\frac{1}{2\xi} \sum_{a=1}^8 (\partial_\mu A^{a,\mu})^2, \quad (1.15)$$

that is needed to avoid the infinities that arise in the functional integral formalism due to the gauge invariance of the lagrangian. Hence, this term is needed to correctly quantize the theory. The parameter ξ is what fixes the gauge. For example $\xi = 1$ is the Feynman gauge while $\xi = 0$ is the Landau gauge.

In the end, the last term of Eq. (1.2) is

$$\mathcal{L}_{\text{ghosts}} = (\partial^\mu \chi^a)^* D_\mu^{ab} \chi^b, \quad (1.16)$$

where D_μ^{ab} is the covariant derivative in the adjoint representation and reads

$$D_\mu^{ab} = \delta^{ab} \partial_\mu - g_s f^{abc} A_\mu^c, \quad (1.17)$$

and χ^b are the so-called Faddeev-Popov ghosts. They are anticommuting complex scalar fields (Grassmann variables). Also this term, like the previous one, is needed to correctly quantize the theory and it comes from the fact that QCD is a non-abelian gauge theory. Ghosts are not “real” particles in the sense that they cannot appear in the final states but only appear in virtual corrections and can be eliminated with a gauge transformation. They are also needed since in QCD when we sum over the gluon polarizations, the longitudinal polarizations do not cancel as it happens for the photon.

1.2 Renormalization

Whenever we compute a cross section beyond leading order in the SM, we observe the appearance of divergences. However, being the cross section a physical observable it must be finite. Therefore, these divergences are unphysical and they must be canceled. This procedure is called *renormalization*. In this section we will only discuss the divergences appearing in the virtual diagrams, i.e. diagrams that don't change the number of particles in the final state. In such diagrams we have integrals like

$$\int_0^\infty dk k^{D-1}, \quad (1.18)$$

where k is the modulus of the momentum flowing in the loop. This integral is divergent both for $k \rightarrow \infty$, called ultraviolet (UV) divergence, and for $k \rightarrow 0$, called infrared (IR) divergence. Let us concentrate for the moment only on the UV divergences (the IR ones will be discussed in Sec. 1.5). In order to cancel them we must “modify” the theory in a way such that all the divergences are regularized: it means that we must introduce a regulator that prevents all quantities from being divergent. In this way, the divergences are transformed into terms that diverge in the limit in which the regulator is removed. For example, a possible regulator is a cutoff on the integral in Eq. (1.18), i.e. we don't integrate up to infinity but up to a certain finite value Λ . In this case, removing the regulator means performing the limit $\Lambda \rightarrow \infty$. The problem of regularizing with a cutoff is that the regularized theory is no longer gauge invariant. Therefore, a much more convenient regularization is the so-called dimensional regularization: the theory is regularized performing the integration over the momenta in $d = 4 - 2\epsilon$ dimensions. In this way both the UV and IR divergences appear as poles in ϵ . Performing the renormalization means removing such poles. After regularizing the theory, we perform the computation of the observables in this modified theory assuring that the divergent quantities are canceled. In the end, we can remove the regulator recovering the “true” theory in which the observables are finite.

In order to cancel the divergences, we observe that the fields and the couplings of the divergent theory are non physical since they are the fields and the couplings of a divergent theory. We will call them *bare fields* and *bare couplings*. Bare quantities will be addressed with a subscript 0. Then, we rescale the bare quantities defining

$$\psi = \frac{\psi_0}{\sqrt{Z_2}}, \quad A_\mu^a = \frac{A_{0\mu}^a}{\sqrt{Z_3}}, \quad m = \frac{Z_2}{Z_m}, \quad g_s = \frac{Z_2 \sqrt{Z_3}}{Z_1} \mu^{-\epsilon} g_{s0}. \quad (1.19)$$

The objects Z_i are called renormalization constants. Observe that in the definition of g_s we had to introduce a scale μ that has the dimension of an energy. The reason is that, in dimensional regularization, the bare coupling g_{s0} is no more dimensionless but has dimension of an energy to the power of ϵ , i.e. $[g_{s0}] = \epsilon$. Therefore, the definition of g_s in Eq. (1.19) assures that it is dimensionless. μ is called renormalization scale and its choice is completely arbitrary. Now we can compute the renormalization constants order by order in perturbation theory so that the poles are removed in the matrix elements and we can safely remove the regulator, i.e. we can send $\epsilon \rightarrow 0$, and all the observables will

be finite. At $\mathcal{O}(\alpha_s)$ they read

$$Z_1 = 1 - (C_F + C_A) \frac{\alpha_s}{4\pi \epsilon} + \mathcal{O}(\alpha_s^2), \quad (1.20)$$

$$Z_2 = 1 - C_F \frac{\alpha_s}{4\pi \epsilon} + \mathcal{O}(\alpha_s^2), \quad (1.21)$$

$$Z_3 = 1 + \left(\frac{5}{3} C_A - \frac{4}{3} n_f T_F \right) \frac{\alpha_s}{4\pi \epsilon} + \mathcal{O}(\alpha_s^2). \quad (1.22)$$

In other words, we are reabsorbing the infinities into a redefinition of fields and couplings. This doesn't represent a problem since, as we already said, the bare fields and quantities are unphysical.

1.3 Running of the strong coupling

As a consequence of dimensional regularization, in order to cancel the divergences, we had to introduce an arbitrary scale μ with the dimension of an energy. Obviously, being μ arbitrary, we have to require that the physical observables are independent on its choice. Equivalently, we can observe that the bare strong coupling is independent on μ . Therefore, imposing that

$$\mu^2 \frac{d}{d\mu^2} \alpha_{s0} = 0, \quad (1.23)$$

we obtain

$$\mu^2 \frac{d}{d\mu^2} \alpha_s(\mu^2) = \beta(\alpha_s(\mu^2)), \quad (1.24)$$

that is called renormalization group equation (RGE) of QCD. In the end, we have found that the strong coupling acquires a dependence on the renormalization scale μ . $\beta(\alpha_s)$ is called beta function of QCD and it is computed in perturbation theory as

$$\beta(\alpha_s) = -\alpha_s^2 (\beta_0 + \beta_1 \alpha_s + \beta_2 \alpha_s^2 + \dots). \quad (1.25)$$

If we solve Eq. (1.24) at LO (i.e. keeping only the term β_0) with a certain initial condition $\alpha_s(\mu_0^2)$, we obtain

$$\alpha_s^{\text{LO}}(\mu^2) = \frac{\alpha_s(\mu_0^2)}{1 + \alpha_s(\mu_0^2) \beta_0 \log\left(\frac{\mu^2}{\mu_0^2}\right)} = \frac{1}{\beta_0 \log\left(\frac{\mu^2}{\Lambda^2}\right)}, \quad (1.26)$$

where we defined

$$\Lambda^2 = \mu_0^2 e^{-\frac{1}{\beta_0 \alpha_s(\mu_0^2)}}. \quad (1.27)$$

Λ is called Landau pole of QCD and it is the value such that Eq. (1.26) diverges. One can observe that, depending on the sign of β_0 , Eq. (1.26) can give two opposite behaviors:

- If $\beta_0 > 0$ the strong coupling goes to zero for $\mu^2 \gg \Lambda^2$. In this case the theory is called UV free.
- If $\beta_0 < 0$ the strong coupling goes to zero for $\mu^2 \ll \Lambda^2$. In this case the theory is called IR free.

The computation of β_0 in QCD gives

$$\beta_0 = \frac{11C_A - 4n_f T_F}{12\pi}. \quad (1.28)$$

where n_f is the number of quarks. Since in the SM we have $n_f = 6$ we are in the case $\beta_0 > 0$ and therefore QCD is UV free. This is the origin of asymptotic freedom in QCD. For computing the value of Λ we can use Eq. (1.27) with $\alpha_s(M_z^2) = 0.118$ [11] finding that $\Lambda \simeq 200$ MeV. It means that, at energies smaller than the typical hadronization scale, i.e. 1 GeV, the strong coupling is not small but is of $\mathcal{O}(1)$. Hence, at these scales perturbation theory breaks down: in order to have reliable predictions we should be able to include an infinite series of contributions, since every term in the perturbative expansion is of the same order of the previous ones.

1.3.1 Solution methods

We have seen that at LO the solution of the RGE is exact and is given by Eq. (1.26). Beyond LO instead, it is not possible to find an analytical solution. Therefore, approximate solutions are needed. A first possibility is to use Runge-Kutta solutions to solve iteratively Eq. (1.24), that with a change of variable can be rewritten as

$$\frac{d}{dt}\alpha_s(t) = \beta(\alpha_s(t)), \quad (1.29)$$

with $t = \log \mu^2$. In this way we find a numerical solution for $\alpha_s(\mu^2)$ starting from the initial condition $\alpha_s^0 = \alpha_s(\mu_0^2)$.

Another possibility is to obtain an expanded solution for the strong coupling. It means that we will find an expression that will be solution of Eq. (1.24) up to higher order corrections. Therefore, if we want to solve it at N^k LO we will obtain an expression that solves the N^k LO RGE up to terms of $\mathcal{O}(\alpha_s^{2+k})$. For example, at NLO we have that the expanded solution is

$$\alpha_s^{\text{NLO}}(\mu^2) = \alpha_s^{\text{LO}}(\mu^2) \left[1 + \frac{\beta_1}{\beta_0} \alpha_s^{\text{LO}}(\mu^2) \log \left(\frac{\alpha_s^{\text{LO}}(\mu^2)}{\alpha_s^0} \right) \right], \quad (1.30)$$

that, as it can be easily verified, satisfies

$$\mu^2 \frac{d}{d\mu^2} \alpha_s^{\text{NLO}}(\mu^2) = -(\alpha_s^{\text{NLO}}(\mu^2))^2 (\beta_0 + \alpha_s^{\text{NLO}}(\mu^2) \beta_1) + \mathcal{O}\left((\alpha_s^{\text{NLO}}(\mu^2))^3\right). \quad (1.31)$$

1.3.2 Running with different flavor numbers

From the expression of β_0 in Eq. (1.28) we can observe that the running of α_s of Eq. (1.26) depends on the number of flavors. Indeed, the computation of the beta function involves quark loops of all possible flavors. Therefore, as far as we know, we must use $n_f = 6$. However, we expect that if a particle is much heavier than the typical scale of the considered process, its contribution to the computation of an observable must be suppressed. For example, if there are other quark families that we don't know yet whose mass is higher than the energies that are currently accessible by experiments, it's reasonable to think that their contribution to observables at scales much smaller than their mass is completely negligible. This is the Appelquist-Carazzone decoupling theorem [12]. Indeed, using the so-called *decoupling renormalization scheme* (DS) we perform

the subtraction at zero momentum and so the propagators of the flavors with $\mu^2 \gg m^2$ go to zero. Therefore, the heavy flavor disappears from the computation of the physical observables. However, the beta function is not an observable and therefore Appellequist-Carazzone decoupling theorem does not apply. It means that the RGE of QCD must be written with $n_f = 6$ even if we use the DS.

This problem is avoided using effective field theories (EFT). Let us consider a theory with n_f massless quarks and one heavy quark with mass much larger than the scale of the process we are considering. This is the $n_f + 1$ flavor scheme. Then, we can consider a theory in which we have n_f light flavors and no heavy quark. This is the n_f flavor scheme. In practice, we have integrated out a degree of freedom of the theory (the heavy quark). In the n_f flavor scheme the RGE will be written in terms of n_f flavors while in the $n_f + 1$ scheme the RGE will be written in terms of $n_f + 1$ flavors. Obviously the two schemes must yield the same result for the physical observables. This is obtained using matching conditions between the strong couplings in the two theories, namely $\alpha_s^{[n_f+1]}$ and $\alpha_s^{[n_f]}$. Such matching conditions are, up to NNLO [13]

$$\frac{\alpha_s^{[n_f]}(\mu^2)}{\alpha_s^{[n_f+1]}(\mu^2)} = 1 - \frac{\alpha_s^{[n_f+1]}(\mu^2)}{6\pi} + \frac{\left(\alpha_s^{[n_f+1]}(\mu^2)\right)^2}{6\pi^2} \left(\frac{1}{6} \log^2 \frac{\mu^2}{m_h^2} - \frac{19}{4} \log \frac{\mu^2}{m_h^2} - \frac{7}{4} \right) + \mathcal{O} \left(\left(\alpha_s^{[n_f+1]} \right)^3 \right). \quad (1.32)$$

In this procedure, we have to introduce a matching scale μ_h such that if $\mu > \mu_h$ we use the $n_f + 1$ flavor scheme, while if $\mu < \mu_h$ we use the n_f flavor scheme. In order to avoid the appearance of large logarithms that would spoil the perturbative convergence we must use $\mu_h \sim \mathcal{O}(m_h)$.

In conclusion, we have constructed a scheme in which the number of flavors varies with the renormalization scale μ . This is a so-called variable flavor number (renormalization) scheme (VFNS).

1.4 Parton model

So far in this chapter we have discussed about the fundamental fields of QCD, i.e. quarks and gluons. However, due to the confinement property, these particles cannot be observed experimentally. Indeed, only color singlet can be observed as asymptotic states. Therefore, what are really observed in experiments are hadrons, that are composite states made of quark and gluons, that for this reason are called *partons*. However, a problem arises: because of the failure of perturbative QCD at the scales of the hadron confinement, computing cross sections involving hadrons in the external states is impossible. For this reason we must introduce the so called *parton model* to describe scattering processes involving hadrons. First of all, we will consider a simpler process in which there is only one hadron in the initial state. Such process is easier to be studied from the theoretical point of view and to be reconstructed experimentally. Then we will generalize to hadron-hadron collisions.

Let us consider a process in which an electron scatters off a proton via the exchange of a virtual photon. The differential cross section for this process is

$$\frac{d\sigma}{dx dy} = \frac{4\pi\alpha_{em}^2(S - m_p^2)}{Q^4} \left(y^2 x F_1(x, Q^2) + \left(1 - y - \frac{xy m_p^2}{S - m_p^2} \right) F_2(x, Q^2) \right), \quad (1.33)$$

where we defined

$$Q^2 = -q^2, \quad (1.34)$$

$$x = \frac{Q^2}{2P \cdot q}, \quad (1.35)$$

$$y = \frac{q \cdot P}{k \cdot P}, \quad (1.36)$$

where k and k' are the momenta of the electron before and after the scattering, q^2 is the virtuality of the emitted photon, i.e. $q^2 = (k - k')^2$, P and m_p are the momentum and the mass of the proton, S is the hadronic center of mass energy (squared), i.e. $S = (k + P)^2$, and F_a with $a = 1, 2$ are called the *structure functions* of the proton and parametrize the internal structure of the proton seen by the virtual photon. In this discussion we are assuming that the energy of the electron is low enough to completely neglect the possibility that the interaction is mediated by an electroweak boson.

If the electron has low energy then the scattering will be elastic and the proton will be still present in the final state. In this case it's easy to show that $x = 1$. Instead, if the electron has enough energy, it can break the internal structure of the proton and it will form new hadrons in the final state. In this case we will have that $x < 1$. This process is called *deep inelastic scattering* (DIS). Neglecting the proton mass, the cross section can be written as

$$\frac{d\sigma}{dx dQ^2} = \frac{2\pi\alpha_{em}^2}{xQ^4} ((1 + (1 - y)^2) F_2 - y^2 F_L), \quad (1.37)$$

where we defined $F_L = F_2 - 2xF_1$. In the parton model the hadronic cross section is computed as a incoherent sum of the cross sections for the scattering of the virtual photon against point-like particles that are the constituents of the proton, i.e. the partons. Let us suppose that the virtual photons interacts with a parton that carries a fraction z of the proton's momentum. We are neglecting the possibility that the parton has a momentum component in the transverse plane with respect to the one of the proton. Then, we have to introduce the *parton distribution function* (PDF) $f_i(z, Q^2)$ that is the probability density that the photon interacts with a parton of type i , where i can be quarks, anti-quarks or gluons, that carries a fraction z of the proton's momentum. Observe that the PDF depends also on Q^2 , since as we will see in the following, it depends also on the scale at which we probe the proton structure. Therefore, if we denote the partonic structure function, i.e. the one involving partons in the external states, with $\hat{F}_{i,a}$, we have that the hadronic structure functions, i.e. the ones involving hadrons in the final states, can be computed as

$$F_a(x, Q^2) = \sum_{i=q,\bar{q},g} \int_x^1 dz f_i(z, Q^2) \hat{F}_{i,a}\left(\frac{x}{z}, Q^2\right), \quad a = 2, L. \quad (1.38)$$

The label q means all possible type of quarks whose component in the proton is non negligible. We will further explain what we mean in the following. Eq. (1.38) is accurate as long as we are at energy scales such that the interactions between the hadron constituents can be neglected and the partons can be treated as free on-shell particles. For this reason, we are neglecting corrections of $\mathcal{O}(\Lambda^2/Q^2)$, called higher-twist corrections. Usually, the DIS structure functions are not written in terms of the partonic structure functions but in terms of the so-called *coefficient functions*, defined as

$$C_{a,i}(z, Q^2) = \frac{\hat{F}_{a,i}(z, Q^2)}{z}. \quad (1.39)$$

In this way Eq. (1.38) becomes

$$F_a(x, Q^2) = x \sum_{i=q, \bar{q}, g} \int_x^1 \frac{dz}{z} f_i(z, Q^2) C_{i,a} \left(\frac{x}{z}, Q^2 \right), \quad a = 2, L. \quad (1.40)$$

In the following of this thesis we will use the handy notation

$$(f \otimes g)(x) = \int_x^1 \frac{dz}{z} f(z) g \left(\frac{x}{z} \right) = \int_x^1 \frac{dz}{z} f \left(\frac{x}{z} \right) g(z), \quad (1.41)$$

so that Eq. (1.40) becomes

$$F_a(x, Q^2) = x \sum_{q, \bar{q}, g} C_{a,i}(Q^2) \otimes f_i(Q^2). \quad (1.42)$$

The coefficient functions are computed in perturbation theory as

$$C_{a,i} = C_{a,i}^{(0)} + \alpha_s C_{a,i}^{(1)} + \alpha_s^2 C_{a,i}^{(2)} + \dots \quad (1.43)$$

From the discussion we carried out so far, it is clear that the PDFs contain the information on the internal structure of the proton. For this reason they are non-perturbative objects: they are governed by a value of the strong coupling related to the typical hadronization scale, i.e. where it is of order $\mathcal{O}(1)$ and QCD is non-perturbative. For this reason, they cannot be computed in perturbation theory but they must be extracted from data. Now we may ask: which flavors do have a PDF? From what we said so far, the up, down and strange quarks (with their corresponding anti-quarks), having a mass that is much smaller than the hadronization scale, must be described by the non perturbative physics of the hadrons. Indeed, they are formed inside the hadron by quarks that split into a quark-gluon pair. This splitting happens at a scale that is clearly in the non perturbative regime and therefore they are non perturbative objects. For this reason, they do have a PDF. From now on, these three quarks will be called light quarks and their mass will be neglected since it gives a correction of the order $\mathcal{O}(m^2/Q^2)$ that is completely negligible with respect to the terms that Eq. (1.38) is neglecting, i.e. the higher-twist corrections. For the same reason, also the gluon content of the proton must be described by a PDF since it is generated at non perturbative scales.

Let us consider now the quarks whose mass is larger than the proton mass and that, for this reason, will be called heavy quarks. The bottom and the top quarks have a mass that is much larger than the proton mass and therefore they are generated by splittings happening at scales that are in the perturbative regime. It means that they will not be described by a non perturbative PDF, but rather by perturbative splittings included in the matrix elements. Instead, the charm quark has a mass slightly larger than the proton mass. Therefore, it cannot be considered light such that its mass is neglected, but we cannot even discard a non perturbative contribution. For this reason, we will consider also a charm PDF that is the so-called *intrinsic charm*.

The PDFs must also satisfy some sum rules coming from conservation arguments. First of all, requiring proton's momentum conservation, we have the *momentum sum rule* that reads

$$\int_0^1 dz \sum_{i=q, \bar{q}, g} z f_i(z, Q^2) = 1. \quad (1.44)$$

Then, requiring flavor conservation in the proton we have that

$$\int_0^1 dz (f_u(z, Q^2) - f_{\bar{u}}(z, Q^2)) = 2, \quad (1.45)$$

$$\int_0^1 dz (f_d(z, Q^2) - f_{\bar{d}}(z, Q^2)) = 1, \quad (1.46)$$

$$\int_0^1 dz (f_s(z, Q^2) - f_{\bar{s}}(z, Q^2)) = 0, \quad (1.47)$$

$$\int_0^1 dz (f_c(z, Q^2) - f_{\bar{c}}(z, Q^2)) = 0, \quad (1.48)$$

that are called *valence sum rules*.

So far we focused on the electron-proton scattering, a process in which we have only one hadron in the initial state and therefore only one PDF. The extension of the parton model to proton-proton collision is straightforward: the two protons interact through their point-like constituents that carry a fraction z_1 and z_2 of their momenta. The hadronic cross section will be given by

$$\sigma(Q^2) = \sum_{i,j=q,\bar{q},g} \int_0^1 dz_1 \int_0^1 dz_2 f_{H_1,i}(z_1, Q^2) f_{H_2,j}(z_2, Q^2) \hat{\sigma}_{ij}(z_1, z_2, Q^2), \quad (1.49)$$

where $f_{H_1,i}(z_1)$ is the probability of finding in the first hadron a parton of type i that carries a fraction z_1 of its momentum (and analogous considerations holds for $f_{H_2,j}(z_2)$) and $\hat{\sigma}_{ij}$ is the partonic cross section, i.e. the cross section involving partons as the external particles, and is computed in perturbation theory. With an abuse of notation with respect to the definition given in Eq. (1.41), Eq. (1.49) will be written in the compact form

$$\sigma(Q^2) = \sum_{i,j=q,\bar{q},g} \hat{\sigma}_{ij}(Q^2) \otimes f_{H_1,i}(Q^2) \otimes f_{H_2,j}(Q^2). \quad (1.50)$$

1.5 Factorization of collinear divergences

In Sec. 1.2 we observed that when we compute a cross section beyond the leading order in the SM, we have to deal with the insurgence of divergences. We already talked about the UV divergences, that are removed with renormalization. Now we will talk about the IR divergences.

In the computation of the cross sections IR divergences arise whenever we have to deal with massless particles, like gluons. Moreover, if we neglect the mass of the light quarks (as we usually do) we have the appearance of other IR divergences that we have to remove. IR divergences appear both in the virtual corrections and in the real ones. For the virtual corrections, they appear when the momentum of the massless particle flowing in the propagator goes to zero. The real corrections consist in the emission of real particles (where with real we mean on-shell) that change the number of particles in the final state. In this case IR divergences arise when the momenta of these particles go to zero or when the angle between the emitted particle and the emitting one go to zero. These two types of divergences are called respectively *soft* or *collinear*. Like the UV divergences, the IR ones are unphysical. Therefore, we have to find a way to remove them. Thanks to Bloch-Nordsieck theorem we know that when we sum real and virtual

diagrams the soft and collinear divergences coming from the final state emissions cancel. Moreover, the Kinoshita-Lee-Nauenberg theorem assures that the soft divergences coming from the initial state emissions also cancel. These cancellations happen at any given order in perturbation theory. We are left with the collinear divergences coming from the initial state emissions, that are removed with the so-called *factorization*

Also in the case of IR divergences we need to regularize them, in the same way we regularized the UV divergences. This can be done in various ways, like introducing a cutoff that prevents the momentum from going to zero or performing the dimensional regularization. In the following we will apply the latter. Let us focus on the DIS structure function F_2 . When we compute the $\mathcal{O}(\alpha_s)$ correction we have that [14]

$$F_2(x, Q^2) = \sum_{q, \bar{q}} e_q^2 x \int_x^1 \frac{dz}{z} q_0(z) \left[\delta \left(1 - \frac{x}{z} \right) + \alpha_s(\mu^2) \left(C_{2,q}^{(1)} \left(\frac{x}{z}, \frac{\mu^2}{Q^2} \right) - \frac{1}{\epsilon} P_{qq}^{(0)} \left(\frac{x}{z} \right) \right) \right] \\ + \sum_{q, \bar{q}} e_q^2 x \int_x^1 \frac{dz}{z} g_0(z) \alpha_s(\mu^2) \left[C_{2,g}^{(1)} \left(\frac{x}{z}, \frac{\mu^2}{Q^2} \right) - \frac{1}{\epsilon} P_{qg}^{(0)} \left(\frac{x}{z} \right) \right], \quad (1.51)$$

where $C_{2,q}^{(1)}$ and $C_{2,g}^{(1)}$ are the regular part of the first order correction to the coefficient functions in which we considered only light quarks and $P_{qq}^{(0)}$ and $P_{qg}^{(0)}$ are objects that describe respectively the splitting of a quark into a gluon-quark pair and of a gluon into a quark-anti-quark pair. They read

$$P_{qg}^{(0)}(z) = \frac{2T_F}{4\pi} (z^2 + (1-z)^2), \quad (1.52)$$

$$P_{qq}^{(0)}(z) = \frac{C_F}{2\pi} \left(\frac{1+z^2}{1-z} \right)_+, \quad (1.53)$$

where with the subscript $+$ we indicated the *plus distribution*. μ is the so-called *factorization scale*: like the renormalization scale it is an unphysical scale that has been introduced due to the dimensional regularization. It is in principle different from the renormalization scale, but a standard choice is to set $\mu_R = \mu_F = \mu$. Observe that the PDFs have been denoted with q_0 and g_0 since at this points they are the *bare* PDFs, i.e. the PDFs in the IR divergent theory. In order to cancel the poles in ϵ we reabsorb them into a redefinition of the PDFs:

$$q(x, \mu^2) = \int_x^1 \frac{dz}{z} q_0(z) \left[\delta \left(1 - \frac{x}{z} \right) - \frac{1}{\epsilon} \alpha_s(\mu^2) P_{qq}^{(0)} \left(\frac{x}{z} \right) \right] \\ + \int_x^1 \frac{dz}{z} g_0(z) \left[-\frac{1}{\epsilon} \alpha_s(\mu^2) P_{qg}^{(0)} \left(\frac{x}{z} \right) \right], \quad (1.54)$$

$$g(x, \mu^2) = g_0(x), \quad (1.55)$$

so that we get

$$F_2(x, Q^2) = \sum_{q, \bar{q}} e_q^2 x \int_x^1 \frac{dz}{z} q(z, \mu^2) \left[\delta \left(1 - \frac{x}{z} \right) + \alpha_s(\mu^2) C_{2,q}^{(1)} \left(\frac{x}{z}, \frac{\mu^2}{Q^2} \right) \right] \\ + \sum_{q, \bar{q}} e_q^2 x \int_x^1 \frac{dz}{z} g(z, \mu^2) \alpha_s(\mu^2) C_{2,g}^{(1)} \left(\frac{x}{z}, \frac{\mu^2}{Q^2} \right), \quad (1.56)$$

that is not divergent anymore.

So far we did a discussion at $\mathcal{O}(\alpha_s)$ in perturbation theory. However, at any order we can reabsorb the initial state collinear IR divergences into a redefinition of the PDFs. Indeed, starting from the hadronic structure functions in the regularized theory we have that

$$F_a = x \sum_{i=q,\bar{q},g} C_{0,a,i} \otimes f_{0,i}, \quad (1.57)$$

where $f_{0,i}$ are the bare PDFs and $C_{0,i}$ are the bare coefficient functions, i.e. the ones with the poles in ϵ . One can show that the bare coefficient functions always factorize as

$$C_{0,a,i} = \sum_{j=q,\bar{q},g} C_{a,j}(\mu^2) \otimes \Gamma_{ji}(\mu^2), \quad (1.58)$$

where $C_{a,i}$ is no longer singular and Γ_{ji} contain the collinear divergences and reads

$$\Gamma_{ij}(z, \mu^2) = \delta_{ij} \delta(1-z) - \frac{1}{\epsilon} \alpha_s(\mu^2) P_{ij}^{(0)}(z) + \mathcal{O}(\alpha_s^2). \quad (1.59)$$

If we plug Eq. (1.59) into Eq. (1.57) we find

$$F_a = x \sum_{i,j=q,\bar{q},g} C_{0,a,j} \otimes \Gamma_{ji}(\mu^2) \otimes f_{0,i}. \quad (1.60)$$

Therefore, if we redefine the PDFs as

$$f_j(\mu^2) = \sum_{i=q,\bar{q},g} \Gamma_{ji}(\mu^2) \otimes f_{0,i}, \quad (1.61)$$

we find

$$F_a = x \sum_{i=q,\bar{q},g} C_i(\mu^2) \otimes f_i(\mu^2). \quad (1.62)$$

that is written in terms of non-divergent quantities. Expanding Eqs. (1.61-1.62) at $\mathcal{O}(\alpha_s)$ we find exactly Eqs. (1.54-1.56). Like in the case of renormalization, we have reabsorbed the infinities into a redefinition of the PDFs. This is not a problem since the bare PDFs are unphysical objects, being the PDFs of the divergent theory.

1.6 DGLAP evolution equations

Requiring the independence of the bare PDFs on the factorization scale μ we can derive the so-called Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) equations that read

$$\mu^2 \frac{d}{d\mu^2} f_i(x, \mu^2) = \sum_{j=q,\bar{q},g} \int_x^1 \frac{dz}{z} P_{ij} \left(\frac{x}{z}, a_s(\mu^2) \right) f_j(z, \mu^2), \quad i = q, \bar{q}, g, \quad (1.63)$$

where we defined

$$a_s = \frac{\alpha_s}{4\pi}, \quad (1.64)$$

and P_{ij} are called *splitting functions* and are computed in perturbation theory as

$$P_{ij}(z) = a_s P_{ij}^{(0)} + a_s^2 P_{ij}^{(1)} + a_s^3 P_{ij}^{(2)} + \dots \quad (1.65)$$

DGLAP equations are a system of $2n_f + 1$ integro-differential equations, with n_f being the number of light quarks, whose solution gives the PDFs dependence on the factorization scale. It is the analogous for the PDFs of the RGE, Eq. (1.24), for the running of α_s . Solving Eq. (1.63) with a certain initial condition at a certain scale μ_0 , we obtain the PDFs at any arbitrary scale μ . It means that

$$f_i(\mu^2) = \sum_{j=q,\bar{q},g} E_{ij}(\mu^2 \leftarrow \mu_0^2) \otimes f_j(\mu_0^2), \quad i = q, \bar{q}, g. \quad (1.66)$$

The PDFs evolved using the splitting functions at N^kLO are usually called N^kLO PDFs. What about intrinsic heavy quarks? For these quarks we consider a PDF (i.e. we consider a non perturbative component) but the fact that the mass of the quark is not neglected prevents the insurgence of collinear divergences associated to the intrinsic quark. Since we have said that the factorization of collinear singularities is what produces the scale dependence of the PDFs, we conclude that the intrinsic heavy quarks PDF do not evolve with DGLAP. Therefore, we have that

$$f_i(\mu^2) = f_i(\mu_0^2), \quad i = c, \bar{c}. \quad (1.67)$$

Before explaining how to solve Eq. (1.63) we observe that DGLAP evolution is resumming to all orders in α_s large logarithms of the form $\log(\mu^2/\mu_0^2)$.

1.6.1 Maximally decoupling the system

In order to solve Eq. (1.63) we have to maximally decouple the system. First of all we rewrite Eq. (1.63) in the matrix form

$$\mu^2 \frac{d}{d\mu^2} \begin{pmatrix} q_i \\ g \\ \bar{q}_i \end{pmatrix} = \sum_{j=1}^{n_f} \begin{pmatrix} P_{q_i q_j} & P_{q_i g} & P_{q_i \bar{q}_j} \\ P_{g q_j} & P_{gg} & P_{g \bar{q}_j} \\ P_{\bar{q}_i q_j} & P_{\bar{q}_i g} & P_{\bar{q}_i \bar{q}_j} \end{pmatrix} \otimes \begin{pmatrix} q_j \\ g \\ \bar{q}_j \end{pmatrix}. \quad (1.68)$$

Then, we observe that, due to charge conjugation invariance and SU(n_f) symmetry, the splitting functions satisfy

$$\begin{aligned} P_{q_i q_j} &= P_{\bar{q}_i \bar{q}_j} = \delta_{ij} P_{qq}^V + P_{qq}^S, \\ P_{\bar{q}_i q_j} &= P_{q_i \bar{q}_j} = \delta_{ij} P_{q\bar{q}}^V + P_{q\bar{q}}^S, \\ P_{q_i g} &= P_{\bar{q}_i g} = P_{qg}, \\ P_{g q_i} &= P_{g \bar{q}_i} = P_{gq}. \end{aligned} \quad (1.69)$$

Plugging Eqs. (1.69) into Eq. (1.68) we obtain

$$\mu^2 \frac{d}{d\mu^2} \begin{pmatrix} q_i \\ g \\ \bar{q}_i \end{pmatrix} = \begin{pmatrix} P_{qq}^V & P_{qg} & P_{q\bar{q}}^V \\ P_{gq} & P_{gg} & P_{g\bar{q}} \\ P_{\bar{q}q}^V & P_{qg} & P_{q\bar{q}}^V \end{pmatrix} \otimes \begin{pmatrix} q_i \\ g \\ \bar{q}_i \end{pmatrix} + \begin{pmatrix} P_{qq}^S & 0 & P_{q\bar{q}}^S \\ 0 & 0 & 0 \\ P_{q\bar{q}}^S & 0 & P_{q\bar{q}}^S \end{pmatrix} \otimes \begin{pmatrix} \sum_{j=1}^{n_f} q_j \\ g \\ \sum_{j=1}^{n_f} \bar{q}_j \end{pmatrix}. \quad (1.70)$$

If now we define

$$q_i^\pm = q_i \pm \bar{q}_i, \quad (1.71)$$

we find that

$$\begin{aligned} \mu^2 \frac{d}{d\mu^2} \begin{pmatrix} q_i^+ \\ g \\ q_i^- \end{pmatrix} &= \begin{pmatrix} (P_{qq}^V + P_{q\bar{q}}^V) & 2P_{qg} & 0 \\ P_{gq} & P_{gg} & 0 \\ 0 & 0 & (P_{qq}^V - P_{q\bar{q}}^V) \end{pmatrix} \otimes \begin{pmatrix} q_i^+ \\ g \\ q_i^- \end{pmatrix} \\ &+ \begin{pmatrix} (P_{qq}^S + P_{q\bar{q}}^S) & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & (P_{qq}^S - P_{q\bar{q}}^S) \end{pmatrix} \otimes \begin{pmatrix} \sum_{i=1}^{n_f} q_i^+ \\ g \\ \sum_{i=1}^{n_f} q_i^- \end{pmatrix}. \end{aligned} \quad (1.72)$$

Defining the the singlet and valence distributions as

$$\Sigma = \sum_{i=1}^{n_f} q_i^+, \quad V = \sum_{i=1}^{n_f} q_i^-, \quad (1.73)$$

and the following splitting functions combinations

$$P^\pm = P_{qq}^V \pm P_{q\bar{q}}^V, \quad (1.74)$$

$$P_{qq} = P^+ + n_f(P_{qq}^S + P_{q\bar{q}}^S) = P^+ + P_{ps}, \quad (1.75)$$

$$P^V = P^- + n_f(P_{qq}^S - P_{q\bar{q}}^S) = P^- + P^S, \quad (1.76)$$

we find that the plus, minus and gluon distributions satisfy

$$\mu^2 \frac{d}{d\mu^2} g = P_{gg} \otimes g + P_{gq} \otimes \Sigma, \quad (1.77)$$

$$\mu^2 \frac{d}{d\mu^2} q_i^+ = P^+ \otimes q_i^+ + \frac{1}{n_f} (P_{qq} - P^+) \otimes \Sigma + 2P_{qg} \otimes g, \quad (1.78)$$

$$\mu^2 \frac{d}{d\mu^2} q_i^- = P^- \otimes q_i^- + \frac{1}{n_f} (P^V - P^-) \otimes V. \quad (1.79)$$

From Eqs. (1.78-1.79) we can observe that if we define linear combinations of q_i^+ (the same applies for q_i^-) that are differences between different flavors, we obtain a distribution whose DGLAP equation is decoupled from all the other distributions. The usually adopted combinations are the following

$$T_3 = u^+ - d^+, \quad (1.80)$$

$$T_8 = u^+ + d^+ - 2s^+, \quad (1.81)$$

$$T_{15} = u^+ + d^+ + s^+ - 3c^+, \quad (1.82)$$

$$T_{24} = u^+ + d^+ + s^+ + c^+ - 4b^+, \quad (1.83)$$

$$T_{35} = u^+ + d^+ + s^+ + c^+ + b^+ - 5t^+, \quad (1.84)$$

$$V_3 = u^- - d^-, \quad (1.85)$$

$$V_8 = u^- + d^- - 2s^-, \quad (1.86)$$

$$V_{15} = u^- + d^- + s^- - 3c^-, \quad (1.87)$$

$$V_{24} = u^- + d^- + s^- + c^- - 4b^-, \quad (1.88)$$

$$V_{35} = u^- + d^- + s^- + c^- + b^- - 5t^-, \quad (1.89)$$

$$(1.90)$$

whose evolution equation is given by

$$\mu^2 \frac{d}{d\mu^2} T_i = P^+ \otimes T_i, \quad (1.91)$$

$$\mu^2 \frac{d}{d\mu^2} V_i = P^- \otimes V_i. \quad (1.92)$$

Then, summing Eq. (1.79) over all flavours we find

$$\mu^2 \frac{d}{d\mu^2} V = P^V \otimes V, \quad (1.93)$$

that again evolves decoupled from all the other distributions. Instead, if we sum Eq. (1.78) over all flavours we cannot get a completely decoupled differential equation. Indeed, it is not possible to decouple the gluon from the singlet evolution. In the end, what we find is a two dimensional sector composed by the gluon and the singlet distributions that evolve coupled through the equation

$$\mu^2 \frac{d}{d\mu^2} \begin{pmatrix} \Sigma \\ g \end{pmatrix} = \begin{pmatrix} P_{qq} & 2n_f P_{qg} \\ P_{gq} & P_{gg} \end{pmatrix} \otimes \begin{pmatrix} \Sigma \\ g \end{pmatrix}. \quad (1.94)$$

The basis composed by the combinations defined in Eqs. (1.80-1.89), the singlet and the valence distributions defined in Eq. (1.73) and the gluon distribution is called *evolution basis*. On the other hand, the basis composed by quarks, anti-quarks and gluon is called *flavor basis*.

In presence of intrinsic flavors a further decomposition is needed. Let us consider for example the case $n_f = 3$. Σ and V in this case are composed only by up, down and strange quarks and anti-quarks. This is the same for T_3, T_8, V_3 and V_8 . However, the remaining non-singlet distributions are combinations of evolving and non evolving distributions. Therefore, in this case, instead of using $T_{15}, T_{24}, T_{35}, V_{15}, V_{24}$ and V_{35} , it is more convenient to use the heavy flavor in the flavor basis. It means that our basis will be given by

$$g, \Sigma, V, T_3, V_3, T_8, V_8, c, \bar{c}, b, \bar{b}, t, \bar{t}, \quad (1.95)$$

where the distributions of charm, bottom and top are non evolving. In the same way, if $n_f = 4$ the basis will be

$$g, \Sigma, V, T_3, V_3, T_8, V_8, T_{15}, V_{15}, b, \bar{b}, t, \bar{t}, \quad (1.96)$$

and so on. We will refer to this basis as *intrinsic evolution basis*.

1.6.2 Passing to Mellin space

A possible way to solve the differential equations in Eqs. (1.91-1.94) is performing the Mellin transform, defined as

$$\tilde{f}(N) \equiv \mathcal{M}[f](N) = \int_0^1 dz z^{N-1} f(z). \quad (1.97)$$

One can show that in N -space the convolutions, defined in Eq. (1.41), are transformed into products. Therefore, if we define

$$\gamma_{ij}(N) = -\mathcal{M}[P_{ij}](N), \quad (1.98)$$

that are called *anomalous dimensions*, Eqs. (1.80-1.89) become

$$\mu^2 \frac{d}{d\mu^2} \tilde{T}_i = -\gamma^+ \tilde{T}_i, \quad (1.99)$$

$$\mu^2 \frac{d}{d\mu^2} \tilde{V}_i = -\gamma^- \tilde{V}_i, \quad (1.100)$$

$$\mu^2 \frac{d}{d\mu^2} \tilde{V} = -\gamma^V \tilde{V}, \quad (1.101)$$

$$\mu^2 \frac{d}{d\mu^2} \begin{pmatrix} \tilde{\Sigma} \\ \tilde{g} \end{pmatrix} = - \begin{pmatrix} \gamma_{qq} & 2n_f \gamma_{qg} \\ \gamma_{gq} & \gamma_{gg} \end{pmatrix} \begin{pmatrix} \tilde{\Sigma} \\ \tilde{g} \end{pmatrix}, \quad (1.102)$$

where \tilde{T}_i , \tilde{V}_i , \tilde{V} , \tilde{g} and $\tilde{\Sigma}$ are the PDFs in N -space. Eqs. (1.99-1.102) are now easier to solve than their corresponding x -space equations. Indeed, finding the N -space solutions of the diagonal sectors is trivial. Regarding, the coupled sector, being the anomalous dimensions dependent on μ through the dependence on α_s , it is not possible to find an exact solution of Eq. (1.102)¹.

Once we have solved DGLAP equations in Mellin space, we have to perform the inverse Mellin transform to find the x -space solution. It is obtained as

$$f(x) = \mathcal{M}^{-1}[\tilde{f}](x) = \frac{1}{2\pi i} \int_{\mathcal{P}} x^{-N} \tilde{f}(N) dN, \quad (1.103)$$

where \mathcal{P} is an arbitrary path in the complex plane that passes to the right of the right-most singularity and such that its imaginary part starts from $-\infty$ and ends to ∞ . If the function $\tilde{f}(N)$ is holomorphic Eq. (1.103) can be written as

$$f(x) = \frac{1}{2\pi i} \int_0^1 x^{-\mathcal{P}(t)} \tilde{f}(\mathcal{P}(t)) \frac{d\mathcal{P}(t)}{dt} dt, \quad (1.104)$$

where $\mathcal{P}(t)$ is the parametrization of the integration path in the complex plane:

$$\mathcal{P}(t) : [0, 1] \rightarrow \mathbb{C} : t \rightarrow \mathcal{P}(t). \quad (1.105)$$

Since the imaginary part of the integration path starts at $-\infty$ and ends to ∞ , a convenient choice is to choose a path that is symmetric with respect to the real axis and that crosses it at $t = 1/2$. It means that

$$\mathcal{P}\left(\frac{1}{2} - t\right) = \mathcal{P}^*\left(\frac{1}{2} + t\right). \quad (1.106)$$

With this property, Eq. (1.104) becomes

$$f(x) = \frac{1}{\pi} \int_{\frac{1}{2}}^1 \operatorname{Re} \left(x^{-\mathcal{P}(t)} \tilde{f}(\mathcal{P}(t)) \frac{1}{i} \frac{d\mathcal{P}(t)}{dt} \right) dt, \quad (1.107)$$

Regarding the choice of the integration path $\mathcal{P}(t)$, even if in principle every paths are equivalent, in practice some choices provide a better numerical convergence than others. Indeed, being $x < 1$, if the real part of $\mathcal{P}(t)$ goes to $-\infty$ at the two ends of the integration

¹The solution of a linear differential equation of the form $\frac{d}{dt} \vec{y}(t) = \mathbf{A}(t) \cdot \vec{y}(t)$ can be computed analytically and is $\vec{y}(t) = e^{\int_{t_0}^t dt' \mathbf{A}(t')} \cdot \vec{y}(t_0)$ if and only if $[\mathbf{A}(t_1), \mathbf{A}(t_2)] = 0$ for all the choices of t_1 and t_2 .

path the convergence of the numerical integral is assured. A powerful choice for $\mathcal{P}(t)$ is to use the so-called *Talbot Path*, where it is parametrized as

$$\mathcal{P}(t) = o + r \left(\frac{\theta(t)}{\tan \theta(t)} + i\theta(t) \right), \quad (1.108)$$

with

$$\theta(t) = \pi (2t - 1), \quad (1.109)$$

and r and o are two free parameters to be fixed. Regarding r , in Ref. [15] it has been suggested $r = 0.4M/(-\log x)$ with M being the number of accurate digits, e.g. $M = 16$. However, this seems to yield unstable results for the Mellin inversion of the matching conditions in the large x region so we added an additional regularization, which makes the path less “edgy” there. The form that we used was

$$r = \frac{0.4 \cdot 16}{0.1 - \log x}. \quad (1.110)$$

Larger values of the additional regularization seem to produce again some instabilities in the large x region. Regarding the choice of o instead, since for the non-singlet anomalous dimensions the rightmost singularity is for $N = 0$, while in the singlet sector it is for $N = 1$, it has been chosen as

$$o = \begin{cases} 1, & \text{singlet sector} \\ 0, & \text{non-singlet sector} \end{cases}. \quad (1.111)$$

In this way, given that the path intersects the real axis for $N = o + r$ we are sure that the integration path passes to the right of all the singularities.

1.6.3 Solving the Mellin space equations

The last step we need to do to solve DGLAP equations is to solve the system in Mellin space, i.e. to solve Eqs. (1.99-1.102). For the diagonal parts the differential equation can be solved exactly. First of all we perform a change of variable from μ to a_s so that the differential equation becomes

$$\frac{d}{da_s} f_i = -\frac{\gamma_{\text{ns}}(a_s)}{\beta(a_s)} f_i, \quad (1.112)$$

where $f_i = \{\tilde{T}_i, \tilde{V}_i, \tilde{V}\}$ and $\gamma_{\text{ns}} = \{\gamma^+, \gamma^-, \gamma^V\}$. The solution of Eq. (1.112) is

$$f_i(a_s) = E_{\text{ns}}(a_s \leftarrow a_s^0) f_i(a_s^0), \quad (1.113)$$

with

$$\begin{aligned} E_{\text{ns}}(a_s \leftarrow a_s^0) &= \exp \left[- \int_{a_s^0}^{a_s} \frac{\gamma_{\text{ns}}(a'_s)}{\beta(a'_s)} da'_s \right] \\ &= \exp \left[\int_{a_s^0}^{a_s} \frac{a'_s \gamma_{\text{ns}}^{(0)} + a_s'^2 \gamma_{\text{ns}}^{(1)} + a_s'^3 \gamma_{\text{ns}}^{(2)}}{a_s'^2 \beta_0 + a_s'^3 \beta_1 + a_s'^4 \beta_2} da'_s \right] \\ &= \exp \left[\gamma_{\text{ns}}^{(0)} j^{(0,2)}(a_s, a_s^0) + \gamma_{\text{ns}}^{(1)} j^{(1,2)}(a_s, a_s^0) + \gamma_{\text{ns}}^{(2)} j^{(2,2)}(a_s, a_s^0) \right], \quad (1.114) \end{aligned}$$

where the functions $j^{(k,2)}$ are defined as

$$j^{(k,2)}(a_s, a_s^0) = \int_{a_s^0}^{a_s} \frac{a_s'^k}{a_s' \beta_0 + a_s'^2 \beta_1 + a_s'^3 \beta_2}, \quad (1.115)$$

and read

$$j^{(2,2)}(a_s, a_s^0) = \arctan \left[\frac{\beta_0(a_s - a_s^0)\Delta}{2\beta_0 + \beta_1(a_s + a_s^0) + 2a_s a_s^0 \beta_2} \right], \quad (1.116)$$

$$j^{(1,2)}(a_s, a_s^0) = \frac{2}{\beta_0 \Delta} \left[\arctan \left(\frac{\beta_1 + 2a_s \beta_2}{\beta_0 \Delta} \right) - \arctan \left(\frac{\beta_1 + 2a_s^0 \beta_2}{\beta_0 \Delta} \right) \right], \quad (1.117)$$

$$j^{(0,2)}(a_s, a_s^0) = \frac{\log \left(\frac{a_s}{a_s^0} \right)}{\beta_0} - \frac{\beta_1}{\beta_0} j^{(1,2)}(a_s, a_s^0) - \frac{\beta_2}{\beta_0} j^{(2,2)}(a_s, a_s^0), \quad (1.118)$$

where $\Delta = \sqrt{4\beta_0\beta_2 - \beta_1^2}/\beta_0$. From Eq. (1.114) is clear that the change of variable from μ^2 to a_s in Eq. (1.112) avoids us to compute the integral of $a_s(\mu^2)$, that is a complicate expression already at LO. In this way we only have to compute the initial and the final values of $a_s(\mu^2)$, i.e. $a_s = a_s(\mu^2)$ and $a_s^0 = a_s(\mu_0^2)$, solving the RGE, Eq. (1.24), with one of the methods described in Sec. 1.3 and then we can apply Eq. (1.114).

For the coupled sector in Eq. (1.102), as we already observed, it is not possible to find an exact solution, due to the dependence of the anomalous dimension on the factorization scale. However, approximate solutions are available in the literature. First of all, also in this case we can rewrite the differential equation in terms of a_s :

$$\frac{d}{da_s} \vec{f} = -\frac{\gamma_{\mathbf{S}}(a_s)}{\beta(a_s)} \cdot \vec{f}, \quad (1.119)$$

where

$$\vec{f} = \begin{pmatrix} \tilde{\Sigma} \\ \tilde{g} \end{pmatrix}, \quad \gamma_{\mathbf{S}} = \begin{pmatrix} \gamma_{qq} & 2n_f \gamma_{qg} \\ \gamma_{gq} & \gamma_{gg} \end{pmatrix}. \quad (1.120)$$

The solution of Eq. (1.119) is given by

$$\vec{f}(a_s) = \mathbf{E}_{\mathbf{S}}(a_s \leftarrow a_s^0) \cdot \vec{f}(a_s^0), \quad (1.121)$$

where

$$\mathbf{E}_{\mathbf{S}}(a_s \leftarrow a_s^0) = \mathcal{P} \exp \left[- \int_{a_s^0}^{a_s} \frac{\gamma_{\mathbf{S}}(a_s')}{\beta(a_s')} da_s' \right], \quad (1.122)$$

where \mathcal{P} is the path-ordering operator. In the remaining of this section we will review two possible approximate ways to compute Eq. (1.122).

Truncated solution

This solution is based on the observation that at LO Eq. (1.122) can be solved exactly. Indeed, we have that at LO it reads

$$\mathbf{E}_{\mathbf{S}}^{(0)}(a_s \leftarrow a_s^0) = \exp \left[\int_{a_s^0}^{a_s} \frac{\gamma_{\mathbf{S}}^{(0)}}{a_s' \beta_0} da_s' \right] = \exp \left[\gamma_{\mathbf{S}}^{(0)} \frac{\log \frac{a_s}{a_s^0}}{\beta_0} \right]. \quad (1.123)$$

Observe that at LO the path ordering operator can be removed and therefore the solution is just the exponential of a matrix. The solution beyond LO can be obtained expanding around the LO solution [16]. Indeed, we can rewrite Eq. (1.119) expanding the denominator so that it reads

$$\frac{d}{da_s} \vec{f} = \frac{\mathbf{R}(a_s)}{a_s} \cdot \vec{f}, \quad (1.124)$$

where $\mathbf{R}(a_s)$ satisfies a perturbative expansion, i.e. $\mathbf{R}(a_s) = \sum_{k=0}^{\infty} a_s^k \mathbf{R}^{(k)}$, and its coefficients read

$$\mathbf{R}^{(0)} = \frac{\gamma_{\mathbf{S}}^{(0)}}{\beta_0}, \quad \mathbf{R}^{(k)} = \frac{\gamma_{\mathbf{S}}^{(k)}}{\beta_0} - \sum_{i=1}^k b_i \mathbf{R}^{(k-i)}, \quad (1.125)$$

with $b_i = \beta_i/\beta_0$. In order to find the solution, we make the following ansatz:

$$\mathbf{E}_{\mathbf{S}}^{(k)}(a_s \leftarrow a_s^0) = \mathbf{U}(a_s) \mathbf{E}_{\mathbf{S}}^{(0)}(a_s \leftarrow a_s^0) \mathbf{U}^{-1}(a_s^0), \quad (1.126)$$

where

$$\mathbf{U}(a_s) = \mathbf{1} + \sum_{k=1}^{\infty} a_s \mathbf{U}^{(k)}, \quad (1.127)$$

with $\mathbf{U}^{(k)}$ to be determined. Inserting Eq. (1.127) into Eq. (1.124) and collecting the powers of a_s we derive a recursive series of commutation relations for the coefficients of the evolution operator \mathbf{U} , that read

$$[\mathbf{U}^{(1)}, \mathbf{R}^{(0)}] = \mathbf{R}^{(1)} - \mathbf{U}^{(1)}, \quad (1.128)$$

$$[\mathbf{U}^{(2)}, \mathbf{R}^{(0)}] = \mathbf{R}^{(2)} + \mathbf{R}^{(1)} \mathbf{U}^{(1)} - 2\mathbf{U}^{(2)}, \quad (1.129)$$

⋮

$$[\mathbf{U}^{(k)}, \mathbf{R}^{(0)}] = \mathbf{R}^{(k)} + \sum_{i=1}^{k-1} \mathbf{R}^{(k-i)} \mathbf{U}^{(i)} - k\mathbf{U}^{(k)} = \mathbf{R}'^{(k)} - k\mathbf{U}^{(k)}, \quad (1.130)$$

with

$$\mathbf{R}'^{(k)} = \mathbf{R}^{(k)} + \sum_{i=1}^{k-1} \mathbf{R}^{(k-i)} \mathbf{U}^{(i)}. \quad (1.131)$$

Then we decompose $\mathbf{R}^{(0)}$ as

$$\mathbf{R}^{(0)} = \mathbf{e}_+ \lambda_+ + \mathbf{e}_- \lambda_-, \quad (1.132)$$

where λ_{\pm} are the eigenvalues of $\mathbf{R}^{(0)}$ and \mathbf{e}_{\pm} are the projectors on their eigenspaces. They read

$$\lambda_{\pm} = \frac{1}{2\beta_0} \left[\gamma_{qq}^{(0)} + \gamma_{gg}^{(0)} \pm \sqrt{\left(\gamma_{qq}^{(0)} - \gamma_{gg}^{(0)} \right)^2 + 4\gamma_{qq}^{(0)} \gamma_{gg}^{(0)}} \right], \quad (1.133)$$

$$\mathbf{e}_{\pm} = \frac{1}{\lambda_{\pm} - \lambda_{\mp}} \left(\mathbf{R}^{(0)} - \lambda_{\mp} \mathbf{1} \right). \quad (1.134)$$

Obviously, the projectors satisfy the following relations

$$\mathbf{e}_{\pm} \cdot \mathbf{e}_{\pm} = \mathbf{e}_{\pm}, \quad \mathbf{e}_{\pm} \cdot \mathbf{e}_{\mp} = \mathbf{0}, \quad \mathbf{e}_+ + \mathbf{e}_- = \mathbf{1}. \quad (1.135)$$

Using Eq. (1.135) we can write the identity

$$\mathbf{U}^{(k)} = e_- \mathbf{U}^{(k)} e_- + e_- \mathbf{U}^{(k)} e_+ + e_+ \mathbf{U}^{(k)} e_- + e_+ \mathbf{U}^{(k)} e_+, \quad (1.136)$$

that, inserted into Eq. (1.130), yields

$$\mathbf{U}^{(k)} = \frac{e_- \mathbf{R}'^{(k)} e_- + e_+ \mathbf{R}'^{(k)} e_+}{k} + \frac{e_+ \mathbf{R}^{(k)} e_-}{\lambda_- - \lambda_+ + k} + \frac{e_- \mathbf{R}^{(k)} e_+}{\lambda_+ - \lambda_- + k}. \quad (1.137)$$

For consistency, when we use this solution, the values of a_s and a_s^0 are computed using the expanded solution.

Iterated solution

Another possible way to solve Eq. (1.122) is using a discretized path-ordering [17]. The evolution operator is computed as

$$\mathbf{E}_S (a_s \leftarrow a_s^0) = \prod_{k=n}^0 \mathbf{E}_S (a_s^{k+1} \leftarrow a_s^k), \quad (1.138)$$

with $a_s^{n+1} = a_s$ and

$$\mathbf{E}_S (a_s^{k+1} \leftarrow a_s^k) = \exp \left[- \frac{\gamma_S (a_s^{k+\frac{1}{2}})}{\beta (a_s^{k+\frac{1}{2}})} \Delta a_s \right], \quad (1.139)$$

where

$$a_s^k = a_0 + k \Delta a_s, \quad (1.140)$$

$$a_s^{k+\frac{1}{2}} = a_0 + \left(k + \frac{1}{2} \right) \Delta a_s, \quad (1.141)$$

$$\Delta a_s = \frac{a_s - a_s^0}{n+1}. \quad (1.142)$$

In Eq. (1.138) the order of the product is such that later operators are to the left. We have observed that choosing $n = 30$ is enough for the solution to converge. In this case, the values of a_s and a_s^0 are computed using the iterated solution. In the following of this thesis we will refer to this solution also as iterated-exact solution or just exact solution.

1.7 Factorization of large logarithms

In the past sections we have seen that when we treat some quark as massless, it gives rise to IR divergences (in addition to those coming from the fact that the gluon is massless). We have seen also that all such IR divergences cancel when we sum virtual diagrams with real emission diagrams, with the exception of the initial state collinear emissions. The way the remaining divergences are handled is by reabsorbing them into a redefinition of the PDFs and this gives rise to the running of the PDFs (of the massless particles) with the factorization scale. However, in addition to the light quarks there are also the heavy quarks whose mass is not neglected: when we keep the mass of a quark it acts as a IR regulator and the cross sections (in the following we will concentrate on the DIS

coefficient functions) will be finite. The IR divergence is replaced by a logarithm of the form $\log \frac{Q^2}{m^2}$, with Q and m being respectively the hard scale of the process and the mass of the heavy quark. However, also in this case some problems arise: indeed, the cross section will satisfy a perturbative expansion of the form

$$\begin{aligned}
\sigma &= a_s^0 c_0^{(0)} \\
&+ a_s \left(c_0^{(1)} \log \frac{Q^2}{m^2} + c_1^{(1)} \right) \\
&+ a_s^2 \left(c_0^{(2)} \log^2 \frac{Q^2}{m^2} + c_1^{(2)} \log \frac{Q^2}{m^2} + c_2^{(2)} \right) \\
&+ a_s^3 \left(c_0^{(3)} \log^3 \frac{Q^2}{m^2} + c_1^{(3)} \log^2 \frac{Q^2}{m^2} + c_2^{(3)} \log \frac{Q^2}{m^2} + c_3^{(3)} \right) \\
&+ \mathcal{O}(\alpha_s^4),
\end{aligned} \tag{1.143}$$

that clearly gives a large logarithms problem. In fact, at order k in a_s we have that the leading term is $a_s^k \log^k \frac{Q^2}{m^2}$. At scales such that $Q \sim m$ we have no problems since the logarithms are small and then we can safely truncate the perturbative series because every term is much smaller than the previous ones. At large scales instead, we have that $a_s \log \frac{Q^2}{m^2} \sim 1$ and therefore perturbation theory breaks down since every term in the series is of the same order or larger than the previous ones. It means that we can no longer truncate the series but we need to resum to all orders if we want to have reliable predictions.

The way this problem is handled, for example considering the charm as the heavy quark, is the following: the *three flavor scheme* (3FS) uses the three light flavors as massless and keeps the exact dependence on the mass of the charm quark. Therefore, only the light PDFs will run with DGLAP (they will be called active flavors) and both the QCD beta function, Eq. (1.28), and DGLAP equations, Eq. (1.91-1.94), will be written in terms of $n_f = 3$. From now on we will use the symbol q for the three light quarks and we will add c and \bar{c} to the sums to make the presence of intrinsic charm explicit. The DIS structure functions will be written as

$$F_a(x, Q^2) = x \sum_{i=q, \bar{q}, g, c, \bar{c}} C_{a,i}^{[3]} \left(\frac{m_c^2}{Q^2}, a_s^{[3]}(Q^2) \right) \otimes f_i^{[3]}(Q^2), \tag{1.144}$$

while the PDF running will be written as

$$f_i^{[3]}(Q^2) = \sum_{j=q, \bar{q}, g} U_{ij}^{[3]}(Q^2 \leftarrow Q_0^2) \otimes f_j^{[3]}(Q_0^2), \quad i = q, \bar{q}, g. \tag{1.145}$$

$$f_i^{[3]}(Q^2) = f_i^{[3]}(Q_0^2), \quad i = c, \bar{c}. \tag{1.146}$$

In Eq. (1.146) we used that, as we discussed in the past section, the intrinsic heavy flavor PDF does not evolve. For the discussion that we carried out so far, the 3FS is reliable when $Q \sim m_c$ since the mass logarithms will be small. In the 3FS, in the case of heavy quark production, the coefficient functions contain a factor $\theta(s - 4m_c^2)$ that comes from the momentum conservation. It means that, if $s < 4m_c^2$ the charm quark does not appear in the final state (and the same for the other heavy quarks). Moreover, the adoption of the DS scheme assures the decoupling of all the heavy quark loops. Therefore, if $Q^2 \ll m_c^2$ the dependence on the heavy quark mass completely disappears. For this reason if the

hard scale of the process is far below the mass of the charm, we can neglect the charm mass corrections. In the same way, if we are close to the charm mass, we can neglect the corrections given by the mass of the bottom.

When $Q^2 \gg m_c^2$ instead, the mass logarithms become large and must be resummed. This is achieved with the *four flavor scheme* (4FS): the large logarithms are reabsorbed into a redefinition of the PDFs and DGLAP evolution and a_s running are performed in terms of $n_f = 4$. In fact, in the massive coefficient functions computed in the 3FS the mass logarithms factorize as

$$C_{a,i}^{[3]} \left(\frac{m_c^2}{Q^2}, a_s^{[4]}(Q^2) \right) = \sum_{j=q,\bar{q},g,c,\bar{c}} C_{a,j}^{[4]} \left(\frac{m_c^2}{Q^2}, a_s^{[4]}(Q^2) \right) \otimes A_{ji} \left(\frac{m_c^2}{Q^2}, a_s^{[4]}(Q^2) \right), \quad (1.147)$$

where the functions A_{ji} are called *matching conditions* and read

$$A_{ij} \left(z, \frac{m_c^2}{Q^2}, a_s(Q^2) \right) = \delta(1-z)\delta_{ij} + \sum_{k=1}^{\infty} (a_s(Q^2))^k A_{ij}^{(k)} \left(z, \frac{m_c^2}{Q^2} \right). \quad (1.148)$$

The matching conditions exactly contain the large logarithms dependence of the coefficient functions. Therefore, the coefficient functions in the 4FS are free from large logarithms and we can safely take the massless limit, i.e.

$$C_{a,i}^{[4]} \left(\frac{m_c^2}{Q^2}, a_s^{[4]}(Q^2) \right) \xrightarrow{Q^2 \gg m_c^2} C_{a,i}^{[4]} \left(0, a_s^{[4]}(Q^2) \right) + \mathcal{O} \left(\frac{m_c^2}{Q^2} \right), \quad (1.149)$$

where $C_{a,i}^{[4]} \left(0, a_s^{[4]}(Q^2) \right)$ is the coefficient function computed neglecting the charm mass (also called massless coefficient function). Observe that, in Eq. (1.147) the coefficient functions in the 3FS are written in terms of $a_s^{[4]}$. This is easily done expanding $a_s^{[3]}$ in terms of $a_s^{[4]}$ using Eq. (1.32) and inserting it in the expression of the coefficient function in the 3FS.

Now we are ready to factorize the large logarithms into the PDFs: plugging Eq. (1.147) into Eq. (1.144) we have that

$$F_a(x, Q^2) = x \sum_{i,j=q,\bar{q},g,c,\bar{c}} C_{a,j}^{[4]} \left(\frac{m_c^2}{Q^2}, a_s^{[4]}(Q^2) \right) \otimes A_{ji} \left(\frac{m_c^2}{Q^2}, a_s^{[4]}(Q^2) \right) \otimes f_i^{[3]}(Q^2). \quad (1.150)$$

Then, defining the PDFs in the 4FS as

$$f_i^{[4]}(Q^2) = \sum_{j=q,\bar{q},g,c,\bar{c}} A_{ij} \left(\frac{m_c^2}{Q^2}, a_s^{[4]}(Q^2) \right) \otimes f_j^{[3]}(Q^2), \quad i = q, \bar{q}, g, c, \bar{c}, \quad (1.151)$$

we have that Eq. (1.150) can be rewritten as

$$F_a(x, Q^2) = x \sum_{i=q,\bar{q},g,c,\bar{c}} C_{a,i}^{[4]} \left(\frac{m_c^2}{Q^2}, a_s^{[4]}(Q^2) \right) \otimes f_i^{[4]}(Q^2). \quad (1.152)$$

In the 4FS the PDFs will run with $n_f = 4$:

$$f_i^{[4]}(Q^2) = \sum_{j=q,\bar{q},g,c,\bar{c}} U_{ij}^{[4]}(Q^2 \leftarrow Q_0^2) \otimes f_j^{[4]}(Q_0^2), \quad i = q, \bar{q}, g, c, \bar{c}. \quad (1.153)$$

Observe that in this scheme we have a component of the PDFs that is perturbatively generated. It is the result of reabsorbing into the PDFs the splittings involving the heavy quark, that in the 3FS were included in the coefficient functions. Moreover, in this scheme the heavy quark PDF evolves with DGLAP evolution together with the other PDFs. The evolution given by Eq. (1.153) is resumming the collinear logarithms to all orders in a_s .

In conclusion, we have that if $Q \sim m_c$ the 3FS gives reliable predictions since the perturbative expansion is well behaved, while if $Q^2 \gg m_c^2$ perturbation theory breaks down and we have to resum the large logarithms and use the 4FS. But the question is: how do we transition from the 3FS to the 4FS? In other words, how do we decide when we can truncate the series and when we have to resum? This is achieved by combining the fixed order result with the resummed one and subtracting the double counted terms. For example, observing Eq. (1.143), the fixed order result consists in the rows while the resummed one consists in the columns. Therefore, what we want to implement is

$$F_a(Q^2) = F_a^{\text{f.o.}}(Q^2) + F_a^{\text{res.}}(Q^2) - F_a^{\text{d.c.}}(Q^2). \quad (1.154)$$

In literature there are different prescriptions available that implement Eq. (1.154), like ACOT [18, 19], S-ACOT [20, 21], TR and TR' [22, 23] and FONLL [24–26]. In the FONLL scheme for example, the subtraction of Eq. (1.154) is implemented observing that the fixed order calculation is just the structure function in the 3FS, the resummed result is the structure function in the 4FS computed with the massless coefficient functions, i.e. in the limit $m_c^2/Q^2 \rightarrow 0$, and the double counting are the 3FS structure functions computed keeping the mass logarithms but sending the power terms to zero. In other words, the coefficient functions that we need to subtract are obtained as

$$C_{a,i}^{[3,0]} \left(\frac{m_c^2}{Q^2}, a_s^{[3]}(Q^2) \right) = \sum_{j=q,\bar{q},g,c,\bar{c}} C_{a,j}^{[4]}(0, a_s^{[3]}(Q^2)) \otimes A_{ji} \left(\frac{m_c^2}{Q^2}, a_s^{[3]}(Q^2) \right). \quad (1.155)$$

In conclusion, in the FONLL scheme Eq. (1.154) is given by

$$\begin{aligned} F_a^{\text{FONLL}}(Q^2, m_c^2) &= F_a^{(3)}(Q^2, m_c^2) + F_a^{(4)}(Q^2, 0) - F_a^{(3,0)}(Q^2, m_c^2) \\ &= \sum_{\substack{i= \\ q,\bar{q},g,c,\bar{c}}} \left(\sum_{\substack{j= \\ q,\bar{q},g,c,\bar{c}}} \left[C_{a,j}^{(3)} \left(\frac{m_c^2}{Q^2} \right) - C_{a,j}^{(3,0)} \left(\frac{m_c^2}{Q^2} \right) \right] \otimes A_{ji}^{-1} \left(\frac{m_c^2}{Q^2} \right) + C_{a,i}^{[4]}(0) \right) \otimes f_i^{[4]}(Q^2), \end{aligned} \quad (1.156)$$

where all the quantities are computed in terms of $a_s^{[4]}(Q^2)$. To conclude this section we remark that we considered the mass logarithms of the charm but the same considerations hold for the other heavy quarks, with the only difference that in these cases we don't have an intrinsic component. It means that the bottom PDF in the 5FS will be purely perturbative.

1.8 Variable flavor number scheme

We have seen that at scales much smaller than a certain heavy quark mass, the heavy quark completely decouples as it cannot be produced neither in loops nor in the final states. As we go up with the energy, the heavy quark starts being produced and we have unresummed mass logarithms (3FS). As we increase again the energy, these logarithms

become large and spoil the perturbative convergence of the theory. Therefore, we have to use the 4FS in which the logarithms are resummed. When we cross the threshold of the next heavy flavor, all these considerations apply in the same way. Therefore, we need to construct a scheme in which below a certain threshold the heavy flavor is completely decoupled and as we cross the threshold, that flavor contributes. When we approach the next heavy flavor the same situation applies: below its threshold it cannot be produced but as we cross it, it starts contributing. In practice, what we want is the following: we want to define thresholds for the production of each heavy quark, i.e. μ_c for the charm, μ_b for the bottom and μ_t for the top, and use an EFT such that below a certain threshold the corresponding heavy quark is completely decoupled. Then, in all the energy ranges we will implement Eq. (1.154) (that in the case of the FONLL scheme is Eq. (1.156)) with the corresponding value of n_f . When we will cross a given threshold, the new heavy quark will be switched on. For example, if $Q < \mu_c$ we will have $n_f = 3$ and no heavy quarks in the theory, if $\mu_c < Q < \mu_b$ we will have $n_f = 4$ and the only heavy quark present will be the charm, and so on. The equivalence between the physical observables in the different energy ranges is assured by the matching performed by Eq. (1.151).

So, if we want to evolve the PDFs from a scale μ_0^2 below μ_c to a scale μ^2 above it, we have to evolve first to μ_c with $n_f = 3$, then we have to perform the matching and in the end we can evolve to μ^2 with $n_f = 4$. It means that the evolution will read

$$\begin{aligned}
f_i^{[4]}(\mu^2) &= \sum_{j=q,\bar{q},g,c,\bar{c}} U_{ij}^{[4]}(\mu^2 \leftarrow \mu_c^2) \otimes f_j^{[4]}(\mu_c^2) \\
&= \sum_{\substack{j,k= \\ q,\bar{q},g,c,\bar{c}}} U_{ij}^{[4]}(\mu^2 \leftarrow \mu_c^2) \otimes A_{jk} \left(\frac{m_c^2}{\mu_c^2}, a_s^{[4]}(\mu_c^2) \right) \otimes f_k^{[3]}(\mu_c^2) \\
&= \sum_{\substack{j,k,h= \\ q,\bar{q},g,c,\bar{c}}} U_{ij}^{[4]}(\mu^2 \leftarrow \mu_c^2) \otimes A_{jk} \left(\frac{m_c^2}{\mu_c^2}, a_s^{[4]}(\mu_c^2) \right) \otimes U_{kh}^{[3]}(\mu_c^2 \leftarrow \mu_0^2) \otimes f_h^{[3]}(\mu_0^2) \\
&= \sum_{j=q,\bar{q},g,c,\bar{c}} T_{ij}(\mu^2 \leftarrow \mu_0^2, m_c^2) \otimes f_j^{[3]}(\mu_0^2), \tag{1.157}
\end{aligned}$$

with

$$T_{ij}(\mu^2 \leftarrow \mu_0^2, m_c^2) = \sum_{\substack{k,h= \\ q,\bar{q},g,c,\bar{c}}} U_{ik}^{[4]}(\mu^2 \leftarrow \mu_c^2) \otimes A_{kh} \left(\frac{m_c^2}{\mu_c^2}, a_s^{[4]}(\mu_c^2) \right) \otimes U_{hj}^{[3]}(\mu_c^2 \leftarrow \mu_0^2). \tag{1.158}$$

Observe that in Eq. (1.158) the dependence on the charm matching scale has to disappear. Indeed, being it an arbitrary scale, the physical quantities cannot depend on it. Therefore, since the coefficient functions in the 4FS do not depend on μ_c , if the PDFs in the 4FS depended on it, the left hand side of Eq. (1.152) would depend on it as well. This is true if we include all the orders in the series, but it is no longer true when we truncate at a given order in perturbation theory. Regarding the choice of the threshold scales, from Eq. (1.158) it is clear that we must choose $\mu_c \sim \mathcal{O}(m_c)$ otherwise we would have unresummed large logarithms coming from the matching conditions. Observe that Eq. (1.157) has been written in terms of the charm quark, but the very same procedure applies when we cross the threshold of the bottom quark. The only difference is that we are not considering an intrinsic bottom component and therefore the bottom PDF in the 5FS will be generated purely perturbatively by the matching conditions.

In conclusion, we have constructed a scheme in which the number of flavors that take part to DGLAP evolution (from now on called “active” flavors) varies with the factorization scale μ . This is called variable flavor number (factorization) scheme and it is analogous to what we did in Sec. 1.3 with the running of the strong coupling.

1.9 Matching and basis rotation

Let us now briefly describe how to perform the matching given in Eq. (1.158), for a generic value of n_f . Since in Mellin space convolutions are transformed into products, in this section we will drop the convolution sign. The matching between the n_f scheme and the $n_f + 1$ scheme reads

$$g^{[n_f+1]} = A_{gg}^S g^{[n_f]} + A_{gq}^S \Sigma_{(n_f)}^{[n_f]} + A_{gH}^S h^{[n_f]}, \quad (1.159)$$

$$q^{[n_f+1]} = \frac{1}{2n_f} A_{qq}^S g^{[n_f]} + \frac{1}{2n_f} A_{qq}^{\text{ps}} \Sigma_{(n_f)}^{[n_f]} + A_{qq}^{\text{ns}} q^{[n_f]}, \quad \text{and the same for } \bar{q}, \quad (1.160)$$

$$h^{[n_f+1]} = \frac{1}{2} A_{Hg}^S g^{[n_f]} + \frac{1}{2} A_{Hq}^{\text{ps}} \Sigma_{(n_f)}^{[n_f]} + A_{HH} h^{[n_f]}, \quad \text{and the same for } \bar{h}, \quad (1.161)$$

where $A_{qq}^S = A_{qq}^{\text{ns}} + A_{qq}^{\text{ps}}$ and in $\Sigma_{(n_f)}^{[n_f]}$ the label $[n_f]$ refers to the flavor scheme while the label (n_f) refers to the value in the sum in Eq. (1.73). From Eq. (1.160) we get that

$$q_+^{[n_f+1]} = \frac{1}{n_f} A_{qq}^S g^{[n_f]} + \frac{1}{n_f} A_{qq}^{\text{ps}} \Sigma_{(n_f)}^{[n_f]} + A_{qq}^{\text{ns}} q_+^{[n_f]}, \quad (1.162)$$

$$q_-^{[n_f+1]} = A_{qq}^{\text{ns}} q_-^{[n_f]}, \quad (1.163)$$

that implies

$$\Sigma_{(n_f)}^{[n_f+1]} = A_{qq}^S g^{[n_f]} + (A_{qq}^{\text{ns}} + A_{qq}^{\text{ps}}) \Sigma_{(n_f)}^{[n_f]}, \quad (1.164)$$

$$V_{(n_f)}^{[n_f+1]} = A_{qq}^{\text{ns}} V_{(n_f)}^{[n_f]}, \quad (1.165)$$

$$T_i^{[n_f+1]} = A_{qq}^{\text{ns}} T_i^{[n_f]}, \quad (1.166)$$

$$V_i^{[n_f+1]} = A_{qq}^{\text{ns}} V_i^{[n_f]}, \quad (1.167)$$

where $\Sigma_{(n_f)}^{[n_f+1]}$ and $V_{(n_f)}^{[n_f+1]}$ are the singlet and valence distributions composed by the first n_f flavors but expressed in the $n_f + 1$ flavor scheme. From Eq. (1.161) we get that

$$h_+^{[n_f+1]} = A_{Hg}^S g_{(n_f)}^{[n_f]} + A_{Hq}^{\text{ps}} \Sigma_{(n_f)}^{[n_f]} + A_{HH} h_+^{[n_f]}, \quad (1.168)$$

$$h_-^{[n_f+1]} = A_{HH} h_-^{[n_f]}. \quad (1.169)$$

In the end, in order to obtain the singlet and the valence in the $n_f + 1$ flavor scheme we have that

$$\Sigma_{(n_f+1)}^{[n_f+1]} = \Sigma_{(n_f)}^{[n_f+1]} + h_+^{[n_f+1]}, \quad (1.170)$$

$$V_{(n_f+1)}^{[n_f+1]} = V_{(n_f)}^{[n_f+1]} + h_-^{[n_f+1]}. \quad (1.171)$$

Moreover, in the new scheme we have to activate two new non-singlet combinations, i.e. $T_{(n_f+1)^2-1}$ and $V_{(n_f+1)^2-1}$. From the form of the non-singlet combinations, it is straightforward to prove that this is obtained simply by

$$T_j^{[n_f+1]} = \Sigma_{(n_f)}^{[n_f+1]} - n_f h_+^{[n_f+1]}, \quad (1.172)$$

$$V_j^{[n_f+1]} = V_{(n_f)}^{[n_f+1]} - n_f h_-^{[n_f+1]}, \quad (1.173)$$

with $j = (n_f + 1)^2 - 1$. Therefore, the matching between the n_f scheme and the $n_f + 1$ can be written in the matrix form

$$\begin{pmatrix} g \\ \Sigma_{(n_f+1)} \\ T_j \end{pmatrix}^{[n_f+1]} = \mathbf{R}_S^{(n_f)} \mathbf{A}_S^{(n_f)} \begin{pmatrix} g \\ \Sigma_{(n_f)} \\ h_+ \end{pmatrix}^{[n_f]}, \quad (1.174)$$

$$\begin{pmatrix} V_{(n_f+1)} \\ V_j \end{pmatrix}^{[n_f+1]} = \mathbf{R}_{\text{ns}}^{(n_f)} \mathbf{A}_{\text{ns}}^{(n_f)} \begin{pmatrix} V_{(n_f)} \\ h_- \end{pmatrix}^{[n_f]}, \quad (1.175)$$

$$T_i^{[n_f+1]} = A_{qq}^{\text{ns}} T_i^{[n_f]}, \quad (1.176)$$

$$V_i^{[n_f+1]} = A_{qq}^{\text{ns}} V_i^{[n_f]}, \quad (1.177)$$

where in Eqs. (1.176-1.177) we have that $i = 3, 8, \dots, n_f^2 - 1$. The matching and rotation matrices read

$$\mathbf{A}_S^{(n_f)} = \begin{pmatrix} A_{gg}^{\text{S}} & A_{gq}^{\text{S}} & A_{gH}^{\text{S}} \\ A_{qg}^{\text{S}} & A_{qq}^{\text{ns}} + A_{qq}^{\text{ps}} & 0 \\ A_{Hg}^{\text{S}} & A_{Hq}^{\text{S}} & A_{HH}^{\text{S}} \end{pmatrix}, \quad (1.178)$$

$$\mathbf{A}_{\text{ns}}^{(n_f)} = \begin{pmatrix} A_{qq}^{\text{ns}} & 0 \\ 0 & A_{HH}^{\text{ns}} \end{pmatrix}, \quad (1.179)$$

and

$$\mathbf{R}_S^{(n_f)} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & -n_f \end{pmatrix}, \quad (1.180)$$

$$\mathbf{R}_{\text{ns}}^{(n_f)} = \begin{pmatrix} 1 & 1 \\ 1 & -n_f \end{pmatrix}. \quad (1.181)$$

As a conclusion of this section we observe that the entry in the last column of the second line of Eq. (1.178) is null for the reason that is still unknown in the literature. It would correspond to a matching term A_{qH} .

“All I wanted to say,” bellowed the computer, “is that my circuits are now irrevocably committed to calculating the answer to the Ultimate Question of Life, the Universe, and Everything.” [...] “But the program will take me a little while to run.” Fook glanced impatiently at his watch. “How long?” he said. “Seven and a half million years.” said Deep Thought.

– The Hitchhiker’s Guide to the Galaxy

In Chap. 1 we saw that whenever we compute a cross section involving hadrons in the initial state, we have to use the parton model to split the long distance physics (hadronic physics) from the short distance physics (partonic physics) using the PDFs. Moreover, we said that the PDFs cannot be computed in QCD due to the failure of perturbation theory, but must be extracted from data. In the end we saw that DGLAP equations give the running of the PDFs with the factorization scale. Therefore, if we extract the PDFs at a certain fitting scale Q_0 we can evolve them at every scale using DGLAP. Given that the DGLAP evolution is known, the problem of PDFs fitting reduces into obtaining the PDFs at the fitting scale. In this chapter we will explain how the PDFs are extracted at the fitting scale Q_0 in the NNPDF4.0 fitting framework. NNPDF4.0 is the PDF set released by the NNPDF collaboration in 2021 that superseded the previous set NNPDF3.1. With respect to its predecessor, it included new data (mainly from LHC), but the main improvement consisted in a completely new fitting framework. Indeed, the fitting methodology has been redesigned in order to use stochastic gradient descent, that with respect to the previously adopted genetic algorithms provides better performances. Moreover, this new PDFs determination included a systematic application of positivity constraints and the enforcement of sum rule integrability.

This chapter is organized as follows: in Sec. 2.1 we will explain the fitting methodology describing the PDFs parametrization, the form of the χ^2 function (i.e. the function that has to be minimized in the fitting procedure, that in the context of deep learning is often called *loss function*), the way it is optimized and the treatment of the experimental uncertainties. In Sec. 2.2 we will describe the dataset used in the NNPDF4.0 analysis. In the end, in Sec. 2.3 we will describe a new theory pipeline for computing the theory predictions that has been developed after the release of NNPDF4.0 and we will compare the results of the new fit using this pipeline with the old one.

2.1 Methodology

As we said in Chap. 1, the observables in high energy physics are computed as

$$\begin{aligned} F_a(x, Q^2) &= x \sum_{i=q, \bar{q}, g} C_{a,i}(Q^2) \otimes f_i(Q^2) \\ &= x \sum_{i,j=q, \bar{q}, g} C_{a,i}(Q^2) \otimes U_{ij}(Q^2 \leftarrow Q_0^2) \otimes f_j(Q_0^2), \end{aligned} \quad (2.1)$$

for the DIS data, and as

$$\begin{aligned} \sigma(Q^2) &= \sum_{i,j=q, \bar{q}, g} \hat{\sigma}_{ij}(Q^2) \otimes f_i(Q^2) \otimes f_j(Q^2) \\ &= \sum_{i,j,k,l=q, \bar{q}, g} \hat{\sigma}_{ij}(Q^2) \otimes U_{ik}(Q^2 \leftarrow Q_0^2) \otimes U_{jl}(Q^2 \leftarrow Q_0^2) \otimes f_k(Q_0^2) \otimes f_l(Q_0^2), \end{aligned} \quad (2.2)$$

for the collider data. Given that using DGLAP evolution we can obtain the PDFs at every scales starting from the PDFs at a given scale, the goal of a PDFs fit is to extract the PDFs at the fitting scale, i.e. extracting $f_i(Q_0^2)$ appearing in Eqs. (2.1-2.2).

In this section we will review the methodology that NNPDF uses for the PDFs extraction. In particular we will describe the methodology used in the NNPDF4.0 analysis, that is at the basis of the work carried out in this thesis. First of all we will explain how the PDFs are parametrized at the fitting scale, how the theory predictions are obtained and then compared to data, we will explain how the experimental uncertainties are propagated in the PDFs and in the end how the PDFs are evolved towards all scales with DGLAP.

2.1.1 PDF parametrization

In order to extract the PDFs at fitting scale we have to choose a parametrization to parametrize $f_i(x, Q_0^2)$. This is a delicate issue since, from the fact that we have almost no information on the shape of the PDFs we are trying to fit, except for some sum rules or integrability constraints, we have to use a general enough parametrization in order not to introduce a bias on the fit. For example, this problem appears when the PDFs are parametrized using fixed functional forms. With this approach, called Hessian method, the PDFs are parametrized with functional forms constructed from polynomials in x and \sqrt{x} and the uncertainties are estimated performing a least square fit to the data using standard error propagation. With this approach it is observed that the PDFs obtained tend to underestimate the uncertainties of the theory predictions. For this reason a ‘‘tolerance’’ factor is introduced a posteriori to enlarge such uncertainties. Therefore, the use of fixed functional forms introduces a bias in the PDFs determination that propagates into a poor description of the uncertainties of theory predictions. For this reason, the NNPDF collaboration chooses a different way to parametrize the PDFs at the fitting scale, using the neural networks. The reason is that, in the limit of an infinite number of parameters, neural networks can approximate any given differentiable function. The parametrization that is adopted from NNPDF is

$$x f_i(x, Q_0^2; \boldsymbol{\theta}) = A_i x^{1-\alpha_i} (1-x)^{\beta_i} \text{NN}_i(x; \boldsymbol{\theta}), \quad i = 1, \dots, 8, \quad (2.3)$$

where NN_i is one of the outputs of the neural network, $\boldsymbol{\theta}$ is the set of all its parameters (the weights and the biases), A_i and the factor $x^{1-\alpha_i} (1-x)^{\beta_i}$ are prefactors included to assure the normalization given by the sum rules and to increase convergence.

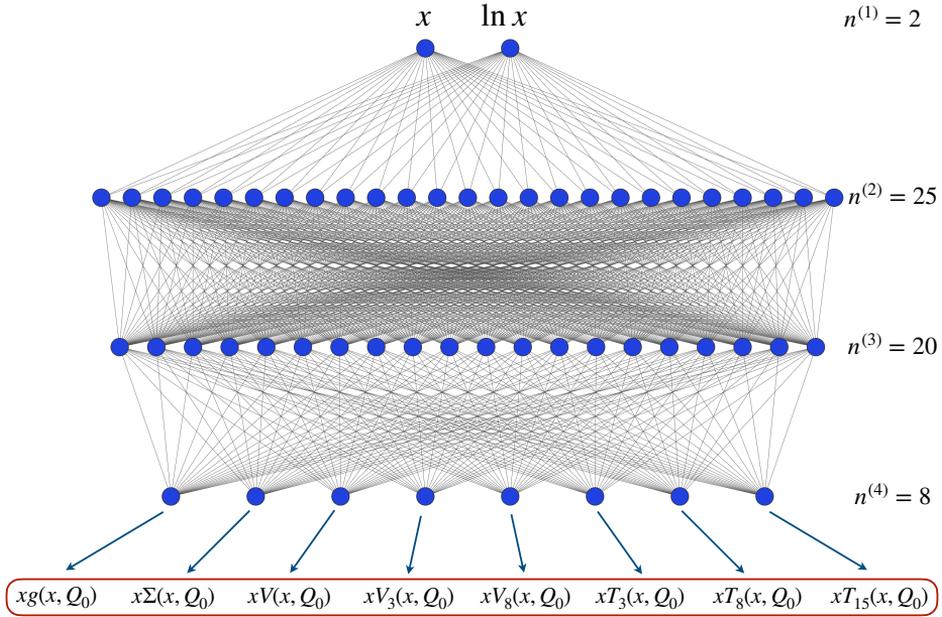


Figure 2.1: Architecture of the neural network used in the NNPDF4.0 PDFs determination.

Now we will briefly describe the architecture of the neural network used in the analysis of NNPDF4.0. First of all, we recall that a neural network provides a non linear mapping from a certain input space (that in our case is the space of the x values) to a certain output space (that in our case is the 8-dimensional space of the PDFs). It does it implementing a layer structure in which the output of every layer is the input of the next one. The input of the first layer and the output of the last one are respectively the input and the output of the neural network. Fig. 2.1 shows the architecture of the neural network used in the NNPDF4.0 methodology. The blue dots corresponds to the nodes of the network. The output of the i -th node of the l -th layer is given by

$$\xi_i^{(l)} = g \left(\sum_{j=1}^{n^{(l-1)}} w_{ij}^{(l)} \xi_j^{(l-1)} + b_i^{(l)} \right), \quad (2.4)$$

where $\xi_j^{(l-1)}$ is the output of the j -th node of the previous layer, $w_{ij}^{(l)}$ and $b_i^{(l)}$ are called weights and biases of the l -th layer and $g(x)$ is a non linear monotonic function called activation function. Different layers can have different activation functions. The set of the weights and biases of all the nodes are the free parameters of the network, i.e. what we have to vary to find the best configuration, and it is what we called θ in Eq. (2.3). Regarding the activation function, a common choice is to use the so-called sigmoid function that reads

$$g(x) = \frac{1}{1 + e^{-x}}. \quad (2.5)$$

This function gives an output that is bounded between 0 and 1 and therefore is the most suitable for approximating the PDFs.

Regarding the input of the neural network, we already said that the input space is the space of the x values. However, we can observe from Fig. 2.1 that the neural network

takes as input both x and $\log(x)$. Indeed, passing also $\log(x)$ to the network, assures that the PDFs at small- x have an adequate weight in the fitting algorithm, that otherwise would result in a poor description of the experimental data in that region.

In addition to its parameters, a neural network is characterized from the so-called hyperparameters, that are the settings of the network that we do not vary during the training. These are for example the number of layers, the number of nodes for each layer, the activation functions, the optimization algorithm and so on. Obviously, there are combinations of such hyperparameters that will perform better in training the network on the data, while others will perform worse. In the NNPDF4.0 methodology such hyperparameters are determined through a semi-automatic algorithm called hyperoptimization.

To avoid the polynomial prefactor in Eq. (2.3) from limiting the functional form and thus resulting in underestimated uncertainties, the exponents α_i and β_i are randomly chosen from a range that is determined in a self-consistent way [27, 28]. Specifically, when a change is made to the methodology or dataset, an initial fit is performed, with the effective exponents of each distribution calculated as

$$\alpha_{\text{eff},i} = \frac{\log f_i(x)}{\log 1/x}, \quad \beta_{\text{eff},i} = \frac{\log f_i(x)}{\log(1-x)}. \quad (2.6)$$

For the next fit, the sampling distribution for the α_i and β_i exponents is uniformly chosen within an interval defined by doubling the 68% confidence interval of the respective effective exponent. This procedure is repeated until the sampling range stabilizes.

As we can observe from Fig. 2.1, the fit is performed in the evolution basis defined in Eq. (1.73) and Eqs. (1.80-1.89). The fitting scale is

$$Q_0 = 1.65 \text{ GeV}. \quad (2.7)$$

Since the threshold used for the matching has been chosen equal to the heavy quark masses, the fit is performed at a scale in which the bottom and the top are not active. For this reason we don't need to fit these distributions since they are zero below their matching scale (we are not considering intrinsic bottom or intrinsic top) and they are perturbatively generated above such scale by the matching performed in Eq. (1.157). Therefore, among the combinations defined in Eqs. (1.80-1.89), we have to fit only the ones that do not include bottom or top. Regarding the charm, the fit is performed at a scale that is slightly above its threshold. Therefore, it will be an active flavor and it will be composed by a perturbative and a non perturbative component. Performing a backward matching to get the charm PDF below the charm mass we can extract the non perturbative component, i.e. the intrinsic charm [29]. In the NNPDF4.0 analysis it has been set for simplicity $c = \bar{c}$ at fitting scale. Indeed, thanks to flavor symmetry considerations, we expect the difference between charm and anti-charm to be very small (even if some asymmetry is generated by evolution). This condition can be relaxed. Indeed, in Ref. [30] another output has been added to the network to fit also \bar{c} . Requiring the symmetry between c and \bar{c} implies that we have to fit T_{15} but not V_{15} since in the latter the c^- component cancels. In App. C we performed a study in which the fitting scale has been varied both below the threshold of the charm and above the one of the bottom. In the first case we are fitting directly the non perturbative component of the charm PDF. The results show that the fit is stable when we perform the fit below or above a heavy quark threshold.

Regarding the base choice, it is obviously possible to perform the fit in the flavor basis, i.e. to fit $f_i = u, \bar{u}, d, \bar{d}, s, \bar{s}, c, g$. It has tested explicitly that performing a fit in

flavor basis doesn't affect the results since the fits performed in the two different basis agree within 1σ uncertainty band.

2.1.2 Performing the fit

Performing the fit means finding the best configuration of the parameters of the neural network $\theta = \{w_{ij}^{(l)}, b_i^{(l)}\}$, such that the theoretical predictions in Eqs. (2.1-2.2) agree with the experimental data. In order to do it, we have to provide as input to the fitting algorithm both the experimental data and a way to compute their theory prediction. The first one will be discussed in detail in Sec. 2.2. Regarding the second one instead, the situation is a little bit more delicate. Indeed, looking at Eqs. (2.1-2.2) we can observe that, in order to get the theory predictions (the right hand side) to be compared with the experimental data (the left hand side), we have to compute the DGLAP evolution from the PDFs at fitting scale to the scale Q^2 relative to the data, and then perform an integration (a double integration in the case of hadronic data) with the partonic cross section. These operations can be computationally heavy and slowing the fitting procedure. This problem is solved with the use of the so-called Fast-Kernel (FK) tables [28]. If we define an interpolation grid in x , i.e. $\{x_i\}$, we can define tables that encode all the information on the partonic cross sections and on the DGLAP evolution factors in a PDFs independent way, such that the theory predictions can be obtained as

$$F^{(n)} = \text{FK}_{i;k}^{(n)} f_i(x_k, Q_0^2), \quad (2.8)$$

for the DIS data, and

$$\sigma^{(n)} = \text{FK}_{ij;kl}^{(n)} f_i(x_k, Q_0^2) f_j(x_l, Q_0^2), \quad (2.9)$$

for collider data. The suffix n labels the FK tables for the different datasets. In this way, evaluating the theory predictions reduces to computing a matrix-vector product, that is much more efficient from the computational point of view.

Now that we can compute the theory predictions in an efficient way we have to compare them with experimental data. This is done computing the χ^2 , that reads

$$\chi^2 = \sum_{i,j=1}^{N_{\text{data}}} (D_i - P_i) (\text{cov}^{-1})_{ij} (D_j - P_j), \quad (2.10)$$

where D_i are the experimental data, P_i are the corresponding predictions and cov_{ij} is the covariance matrix. The experimental covariance matrix reads

$$(\text{cov}_{\text{exp}})_{ij} = \delta_{ij} \sigma_i^{(\text{uncorr})} \sigma_j^{(\text{uncorr})} + \left(\sum_{m=1}^{N_{\text{mult}}} \sigma_{i,m}^{(\text{norm})} \sigma_{j,m}^{(\text{norm})} + \sum_{l=1}^{N_{\text{corr}}} \sigma_{i,l}^{(\text{corr})} \sigma_{j,l}^{(\text{corr})} \right) D_i D_j, \quad (2.11)$$

where $\sigma_i^{(\text{uncorr})}$ are the uncorrelated uncertainties that are obtained by summing in quadrature the uncorrelated systematic uncertainties and the statistical uncertainties, $\sigma_{i,m}^{(\text{norm})}$ are the multiplicative normalization uncertainties and $\sigma_{i,m}^{(\text{corr})}$ are other correlated systematic uncertainties. The χ^2 computed from Eq. (2.10) gives an estimation of the agreement between experimental data and theory predictions. However, in order to avoid the D'Agostini bias [31], that arises in presence of multiplicative uncertainties when the

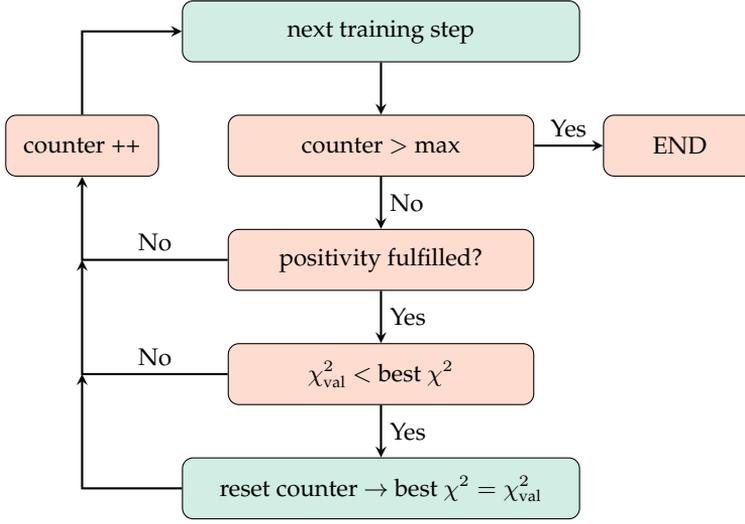


Figure 2.2: Flowchart illustrating the early stopping algorithm employed in NNPDF4.0 to identify the optimal stopping point for the fit.

experimental covariance matrix is used for minimization, the so-called t_0 prescription is applied. With this prescription, the χ^2 is minimized using the covariance matrix given by

$$(\text{cov}_{t_0})_{ij} = \delta_{ij} \sigma_i^{(\text{uncorr})} \sigma_j^{(\text{uncorr})} + \sum_{m=1}^{N_{\text{mult}}} \sigma_{i,m}^{(\text{norm})} \sigma_{j,m}^{(\text{norm})} P_i^{(0)} P_j^{(0)} + \sum_{l=1}^{N_{\text{corr}}} \sigma_{i,l}^{(\text{corr})} \sigma_{j,l}^{(\text{corr})} D_i D_j, \quad (2.12)$$

where $P_i^{(0)}$ is the central value of a theoretical prediction computed using PDFs from a previous fit. Therefore, the t_0 prescription introduces the need of an iterative procedure, in which at every step $P_i^{(0)}$ are computed using the PDFs obtained from the previous fit. The iteration is stopped only when the t_0 covariance matrix stabilizes. It is observed that two or three iterations are enough to achieve convergence.

Now the question is: when do we stop optimizing the neural network parameters? Indeed, a sufficiently large neural network has a flexibility such that the χ^2 can be reduced indefinitely and the network starts learning the noise present in the data. This is the so-called *overfitting* and regularization algorithms must be used to prevent it. In the NNPDF framework, this regularization method primarily depends on an early stopping algorithm that uses cross-validation, as illustrated by the flowchart in Fig. 2.2.

This stopping algorithm serves two main purposes: identifying the optimal set of neural network parameters found during training, and determining the point at which to stop further searching and end the training process. To determine the optimal instance of the neural network (defined as the one that generalizes best to new data) a cross-validation method is employed. In this method, the entire global NNPDF4.0 dataset is split into a validation dataset and a training dataset: each experimental dataset is randomly divided with 75% of its data points going into the training set and 25% into the validation set. Fig. 2.3 demonstrates how this division helps to identify the best instance of the neural network. Specifically, during the fitting process, the training set is used to establish a training error function χ_{tr}^2 which the optimizer targets and can

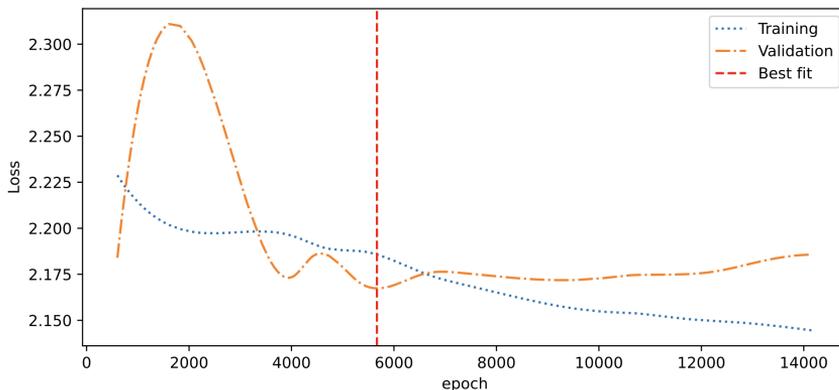


Figure 2.3: Idealized behavior of training loss (dotted, blue) and validation loss (dashed, orange) during a typical PDF fitting process. For clarity, the curves have been smoothed. While the optimization algorithm keeps reducing the training loss, monitoring the validation loss allows us to stop the training at the optimal point, avoiding the overfitting.

theoretically minimize indefinitely. This is represented by the blue curve. The validation set, however, is not accessible to the optimizer, but its error function χ_{val}^2 is evaluated at each training epoch and is shown by the orange line. As illustrated, after χ_{val}^2 reaches its minimum value just before 6000 epochs, it starts to increase again. This increase indicates overfitting, where the optimizer starts fitting the noise in the training data, thus failing to generalize to new data. The final fitting result corresponds to the instance where χ_{val}^2 is at its lowest value. The epoch representing the best neural network instance is marked by the vertical dashed line.

To determine when a neural network has completed its training, a counter is initiated once the validation loss, χ_{val}^2 , falls below a certain threshold. From this point, the counter monitors the number of epochs that pass, and the training stops if the validation loss does not improve for a specified number of epochs. If this occurs, the training stops, and the model reverts to the instance with the best validation loss. If the threshold value for the validation loss is never reached during training, the fit is considered inadequate and discarded. Additionally, to be deemed acceptable, the instance must meet certain positivity criteria [4] to ensure that the PDFs for up, down, and strange quarks and antiquarks, as well as the gluon PDF, are positive. These constraints are based on Ref. [32], which demonstrated that PDFs for individual quark flavors and the gluon, as defined in the $\overline{\text{MS}}$ factorization scheme, are non-negative. Lastly, there is a maximum epoch threshold for training the model. If the model is still improving upon reaching this threshold, the training will be terminated anyway.

After completing the training of the full set of replicas, first the initial scale PDFs are evolved through DGLAP equations to obtain the PDFs for a set of Q values logarithmically spaced between $Q = 1$ GeV and $Q = 100$ TeV. PDFs at scales that are between two consecutive Q values are obtained through interpolation. Then, specific post-fit criteria are assessed. Any replicas failing to meet these criteria are discarded. In particular, replicas with an arc-length or χ^2 value, calculated against the experimental data, that deviates more than 4σ from the central value of their distribution are removed. The post-fit

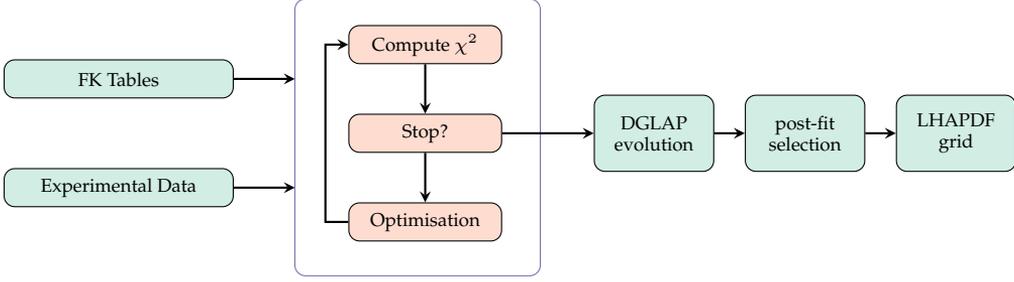


Figure 2.4: Diagrammatic representation of the NNPDF4.0 fitting framework.

check also ensures the integrability of the solutions by verifying that the inequality

$$\sum_k \left| x_k^{\text{int}} f_i(x_k^{\text{int}}, Q^2) \right| < \frac{1}{2}, \quad (2.13)$$

is satisfied for $f_i = V, V_3, V_8, T_3, T_8$ and $x_k^{\text{int}} = 10^{-9}, 10^{-8}, 10^{-7}$ with $Q^2 = 5 \text{ GeV}^2$. It is observed that roughly 1% of the replicas do not pass post-fit criteria.

In the end the output of the fitting procedure is converted into LHAPDF6 [33] format. In Fig. 2.4 the whole fitting procedure that we described is shown.

2.1.3 Treatment of experimental uncertainties

In order to provide a useful description of physics observables, it is necessary to provide an uncertainty band for the PDFs. The way this problem is handled in the NNPDF4.0 framework is with the use of N_{rep} Monte Carlo replicas (where usually $N_{\text{rep}} = 100$). It means that the uncertainties are estimated providing a set of N_{rep} different PDFs $f_i^{(r)}$, with $r = 1, \dots, N_{\text{rep}}$, that give an importance sampling of the PDFs probability distribution. The different replicas are obtained by constructing N_{rep} sets of artificial data, called “pseudodata”, that are generated by shifting the original data around their central values with a multigaussian distribution given by the covariance matrices associated to each dataset and then performing N_{rep} independent fits. Since all the replicas are equally probable, they are statistically uncorrelated. Therefore, a generic object $X[f]$ that depends on the PDFs, can be obtained by averaging over the PDFs set, i.e.

$$\langle X[f] \rangle = \frac{1}{N_{\text{rep}}} \sum_{r=1}^{N_{\text{rep}}} X[f^{(r)}], \quad (2.14)$$

where the object X can be a cross section, a DIS structure function or the PDF itself. In the same way, the variance is obtained as

$$\text{Var}[X[f]] = \frac{1}{N_{\text{rep}}} \sum_{r=1}^{N_{\text{rep}}} \left(X[f^{(r)}] - \langle X[f] \rangle \right)^2. \quad (2.15)$$

In Fig. 2.5 it is shown the set of replicas for the gluon PDF at fitting scale (on the left) and the uncertainty band constructed from these replicas. The continuous solid band corresponds to the 68% of the envelope of all replicas, while the dashed line is the Monte Carlo error obtained computing the variance of the replicas with Eq. (2.15). The agreement of the two kind of errors gives an estimate of the goodness of the gaussian approximation for the experimental uncertainties.

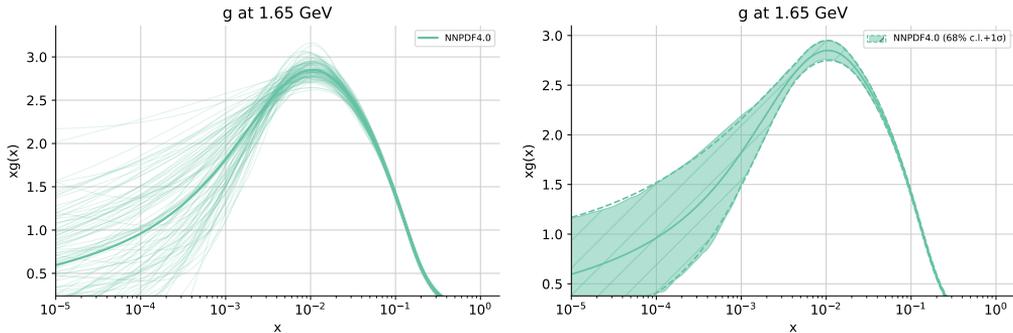


Figure 2.5: A distribution of 100 PDF replicas (left) and the corresponding 1σ interval at a 68% confidence level, calculated using Eqs. (2.14-2.15) with X as the identity operator (right). Both are displayed for the gluon distribution at 1.65 GeV.

Dataset	Ref.	N_{dat}	x	Q [GeV]
NMC F_2^d / F_2^p	[34]	260 (121/121)	[0.012, 0.680]	[2.1, 10.]
NMC $\sigma^{\text{NC},p}$	[35]	292 (204/204)	[0.012, 0.500]	[1.8, 7.9]
SLAC F_2^p	[36]	211 (33/33)	[0.140, 0.550]	[1.9, 4.4]
SLAC F_2^d	[36]	211 (34/34)	[0.140, 0.550]	[1.9, 4.4]
BCDMS F_2^p	[37]	351 (333/333)	[0.070, 0.750]	[2.7, 15.]
BCDMS F_2^d	[37]	254 (248/248)	[0.070, 0.750]	[2.7, 15.]
CHORUS σ_{CC}^{ν}	[38]	607 (416/416)	[0.045, 0.650]	[1.9, 9.8]
CHORUS $\sigma_{CC}^{\bar{\nu}}$	[38]	607 (416/416)	[0.045, 0.650]	[1.9, 9.8]
NuTeV σ_{CC}^{ν} (dimuon)	[39,40]	45 (39/39)	[0.020, 0.330]	[2.0, 11.]
NuTeV $\sigma_{CC}^{\bar{\nu}}$ (dimuon)	[39,40]	45 (36/37)	[0.020, 0.210]	[1.9, 8.3]
[NOMAD $\mathcal{R}_{\mu\mu}(E_{\nu})$] (*)	[41]	15 (—/15)	[0.030, 0.640]	[1.0, 28.]
[EMC F_2^c]	[42]	21 (—/16)	[0.014, 0.440]	[2.1, 8.8]
HERA I+II $\sigma_{\text{NC},\text{CC}}^p$	[43]	1306 (1011/1145)	$[4 \cdot 10^{-5}, 0.65]$	[1.87, 223]
HERA I+II σ_{NC}^c (*)	[44]	52 (—/37)	$[7 \cdot 10^{-5}, 0.05]$	[2.2, 45]
HERA I+II σ_{NC}^b (*)	[44]	27 (26/26)	$[2 \cdot 10^{-4}, 0.50]$	[2.2, 45]

Table 2.1: The DIS datasets examined for the NNPDF4.0 PDF determination are detailed here. For each dataset, we provide its name as used in this paper, the relevant reference, the number of data points included in the NLO/NNLO fits both before and after applying kinematic cuts (refer to Sec. 4 of Ref. [4]), and the kinematic ranges covered after the cuts. Datasets not previously included in NNPDF3.1 are marked with an asterisk, while those not part of the baseline determination are noted in square brackets. The Q coverage for NOMAD should be understood as an integration range.

2.2 Dataset

The data used in the NNPDF4.0 PDFs determination are outlined in Tabs. 2.1-2.5. These tables contain DIS, DIS jets, fixed-target Drell–Yan, inclusive gauge boson production at colliders, and other LHC processes, respectively. For each dataset, we provide the dataset name, the corresponding reference, the number of data points used in the NLO and NNLO fits both before and after applying kinematic cuts (i.e. some data points are discarded for being at too low scales, as detailed in Sec. 4 of Ref. [4]), and the kinematic ranges considered post-cuts. Datasets not previously featured in NNPDF3.1 are marked

Dataset	Ref.	N_{dat}	Q^2 [GeV ²]	p_T [GeV]
[ZEUS 820 (HQ) (1j)] (*)	[45]	30 (—/30)	[125,10000]	[8,100]
[ZEUS 920 (HQ) (1j)] (*)	[46]	30 (—/30)	[125,10000]	[8,100]
[H1 (LQ) (1j)] (*)	[47]	48 (—/48)	[5.5,80]	[4.5,50]
[H1 (HQ) (1j)] (*)	[48]	24 (—/24)	[150,15000]	[5,50]
[ZEUS 920 (HQ) (2j)] (*)	[49]	22 (—/22)	[125,20000]	[8,60]
[H1 (LQ) (2j)] (*)	[47]	48 (—/48)	[5.5,80]	[5,50]
[H1 (HQ) (2j)] (*)	[48]	24 (—/24)	[150,15000]	[7,50]

Table 2.2: As Tab. 2.1 for DIS jet data.

Dataset	Ref.	N_{dat}	$y_{\ell\ell}$	$m_{\ell\ell}$ [GeV]
E866 $\sigma^d/2\sigma^p$ (NuSea)	[50]	15 (15/15)	[0.07, 1.53]	[4.60, 12.9]
E866 σ^p (NuSea)	[51]	184 (89/89)	[0.00, 1.36]	[4.50, 8.50]
E605 σ^p	[52]	119 (85/85)	[-0.20, 0.40]	[7.10, 10.9]
E906 $\sigma^d/2\sigma^p$ (SeaQuest) (*)	[53]	6 (6/6)	[0.11, 0.77]	[4.71, 6.36]

Table 2.3: As Tab. 2.1 for fixed-target DY data.

with an asterisk, while those excluded from the baseline determination are shown in brackets.

The NNPDF4.0 analysis uses a total of 4426 data points at NLO and 4618 at NNLO (after the kinematic cuts). This is an increase compared to NNPDF3.1, which had 4295 data points at NLO and 4285 at NNLO. For a comparison of the datasets used in NNPDF4.0 with those in NNPDF3.1 and other recent PDF analyses (namely ABMP16 [54], CT18 [55], and MSHT20 [56]) refer to App. B of Ref. [4], see Tables B.1-B.6.

Fig. 2.6 shows the kinematic coverage in the (x, Q^2) plane for the NNPDF4.0 dataset used in the default NNLO fit. For hadronic data, the kinematic variables are calculated using leading-order (LO) kinematics. When an observable is averaged over rapidity, the midpoint of the integration range is used to determine the x values. Data points from datasets introduced in NNPDF4.0 are highlighted with a black border.

2.3 New computational tools

In Sec. 2.1 we described the fitting methodology of the NNPDF4.0 PDFs determination. We also mentioned how the theory predictions used in the fit to be compared with the experiments are obtained through the FKtables approach, i.e. Eqs. (2.8-2.9). These tables are numerically computed with a computing framework that relied on a combination of various codes, including APFEL [110] for PDF evolution and DIS structure function calculations, and APFELgrid [111] for creating interpolation grids of NLO partonic matrix elements. These grids utilized the APPLgrid [112] and FastNLO [113] interpolators. After the NNPDF4.0 analysis, the NNPDF collaboration started working on a new theory pipeline to replace the old one. The primary advantage of the new pipeline is its cohesive yet modular and adaptable structure. It is described in Ref. [114]. This new pipeline has been fully benchmarked with the old one for DGLAP evolution (both pure QCD and mixed QCD \otimes QED, as we will see in Chap. 3), for the computation of the structure functions, the interpolation of grids and the interface to mg5_aMC@NLO. With this new

Dataset	Ref.	N_{dat}	Kin ₁	Kin ₂ [GeV]
CDF Z differential	[57]	29 (29/29)	$0.0 \leq y_{\ell\ell} \leq 2.9$	$66 \leq m_{\ell\ell} \leq 116$
D0 Z differential	[58]	28 (28/28)	$0.0 \leq y_{\ell\ell} \leq 2.8$	$66 \leq m_{\ell\ell} \leq 116$
[D0 W electron asymmetry]	[59]	13 (13/8)	$0.0 \leq y_e \leq 2.9$	$Q = m_W$
D0 W muon asymmetry	[60]	10 (10/9)	$0.0 \leq y_\mu \leq 1.9$	$Q = m_W$
ATLAS low-mass DY 7 TeV	[61]	6 (4/6)	$ \eta_\ell \leq 2.1$	$14 \leq m_{\ell\ell} \leq 56$
ATLAS high-mass DY 7 TeV	[62]	13 (5/5)	$ \eta_\ell \leq 2.1$	$116 \leq m_{\ell\ell} \leq 1500$
ATLAS W, Z 7 TeV ($\mathcal{L} = 35 \text{ pb}^{-1}$)	[63]	30 (30/30)	$ \eta_\ell, y_Z \leq 3.2$	$Q = m_W, m_Z$
ATLAS W, Z 7 TeV ($\mathcal{L} = 4.6 \text{ fb}^{-1}$) (*)	[64]	61 (53/61)	$ \eta_\ell, y_Z \leq 2.5, 3.6$	$Q = m_W, m_Z$
CMS W electron asymmetry 7 TeV	[65]	11 (11/11)	$ \eta_e \leq 2.4$	$Q = m_W$
CMS W muon asymmetry 7 TeV	[66]	11 (11/11)	$ \eta_\mu \leq 2.4$	$Q = m_W$
CMS DY 2D 7 TeV	[67]	132 (88/110)	$ \eta_{\ell\ell} \leq 2.2$	$20.0 \leq m_{\ell\ell} \leq 200$
LHCb $Z \rightarrow ee$ 7 TeV	[68]	9 (9/9)	$2.0 \leq \eta_\ell \leq 4.5$	$Q = m_Z$
LHCb $W, Z \rightarrow \mu$ 7 TeV	[69]	33 (29/29)	$2.0 \leq \eta_\ell \leq 4.5$	$Q = m_W$
[ATLAS W 8 TeV] (*)	[70]	22 (—/22)	$ \eta_\ell < 2.4$	$Q = m_W$
ATLAS low-mass DY 2D 8 TeV (*)	[71]	84 (47/60)	$ y_{\ell\ell} < 2.4$	$46 \leq m_{\ell\ell} \leq 200$
ATLAS high-mass DY 2D 8 TeV (*)	[72]	48 (48/48)	$ y_{\ell\ell} < 2.4$	$116 \leq m_{\ell\ell} \leq 1500$
CMS W rapidity 8 TeV	[73]	22 (22/22)	$ \eta_\ell \leq 2.3$	$Q = m_W$
LHCb $Z \rightarrow ee$ 8 TeV	[74]	17 (17/17)	$2.00 < \eta_e < 4.25$	$Q = m_Z$
LHCb $W, Z \rightarrow \mu$ 8 TeV	[75]	34 (29/30)	$2.00 < \eta_\mu < 4.25$	$Q = m_Z$
[LHCb $W \rightarrow e$ 8 TeV] (*)	[76]	8 (—/8)	$2.00 < \eta_e < 4.25$	$Q = m_W$
ATLAS $\sigma_{W,Z}^{\text{tot}}$ 13 TeV (*)	[77]	3 (3/3)	—	$Q = m_W, m_Z$
LHCb $Z \rightarrow ee$ 13 TeV (*)	[78]	17 (15/15)	$2.00 < y_Z < 4.25$	$Q = m_Z$
LHCb $Z \rightarrow \mu\mu$ 13 TeV (*)	[78]	18 (16/16)	$2.00 < y_Z < 4.50$	$Q = m_Z$

Table 2.4: As Tab. 2.1 for collider (Tevatron, top, and LHC, bottom) inclusive gauge boson production data.

pipeline, interpolations grids for different processes have been recomputed:

- All fully inclusive DIS processes have been recomputed using the code `YADISM` [115, 116]. This code reimplements the FONLL prescription in a way that heavy quarks mass effects are treated in a better way [117].
- All fixed-target Drell–Yan processes have been recomputed using a modified version [114] of `VRAP` [118].
- The following datasets of Tabs. 2.4–2.5 have been recomputed using `mg5_aMC@NLO` (all grids, in the `PineAPPLE` format, are available at <https://github.com/NNPDF/pineapplgrids>):
 - ATLAS W, Z 7 TeV ($\mathcal{L} = 4.6 \text{ fb}^{-1}$)
 - ATLAS low-mass DY 7 TeV
 - ATLAS high-mass DY 7 TeV
 - ATLAS σ_{tt}^{tot} 7, 8
 - ATLAS σ_{tt}^{tot} 13 ($\mathcal{L} = 139 \text{ fb}^{-1}$)
 - ATLAS $t\bar{t}$ lepton+jets 8 TeV
 - ATLAS $t\bar{t}$ dilepton 8 TeV
 - CMS Drell–Yan 2D 7 TeV

Dataset	Ref.	N_{dat}	Kin ₁	Kin ₂ [GeV]
ATLAS $W^\pm + c$ 7 TeV (*)	[79]	22 (22/—)	$ \eta_\ell < 2.5$	$Q = m_W$
CMS $W^\pm + c$ 7 TeV	[80]	10 (10/—)	$ \eta_\ell < 2.1$	$Q = m_W$
CMS $W^\pm + c$ 13 TeV (*)	[81]	5 (5/—)	$ \eta_\ell < 2.4$	$Q = m_W$
ATLAS $W^\pm + \text{jet}$ 8 TeV (*)	[82]	32 (30/30)	$0 \leq p_T^W \leq 800$ GeV	$Q = m_W$
ATLAS Z p_T 8 TeV ($p_T, m_{\ell\ell}$)	[83]	64 (40/44)	$12 \leq m_{\ell\ell} \leq 150$ GeV	$30 \leq p_T^Z \leq 900$
ATLAS Z p_T 8 TeV (p_T, y_Z)	[83]	120 (18/48)	$ y_Z < 2.4$	$30 \leq p_T^Z \leq 150$
CMS Z p_T 8 TeV	[73]	50 (28/28)	$ y_Z < 1.6$	$30 \leq p_T^Z \leq 170$
CMS $\sigma_{t\bar{t}}^{\text{tot}}$ 5 TeV (*)	[84]	1 (1/1)	—	$Q = m_t$
ATLAS $\sigma_{t\bar{t}}^{\text{tot}}$ 7, 8 TeV	[85]	2 (2/2)	—	$Q = m_t$
CMS $\sigma_{t\bar{t}}^{\text{tot}}$ 7, 8 TeV	[86]	2 (2/2)	—	$Q = m_t$
ATLAS $\sigma_{t\bar{t}}^{\text{tot}}$ 13 TeV ($\mathcal{L}=139 \text{ fb}^{-1}$) (*)	[87]	1 (1/1)	—	$Q = m_t$
CMS $\sigma_{t\bar{t}}^{\text{tot}}$ 13 TeV	[88]	1 (1/1)	—	$Q = m_t$
[ATLAS $t\bar{t} \ell + \text{jets}$ 8 TeV ($1/\sigma d\sigma/dp_T^t$)]	[89]	8 (—/8)	$0 \leq p_T^t \leq 500$ GeV	$Q = m_t$
ATLAS $t\bar{t} \ell + \text{jets}$ 8 TeV ($1/\sigma d\sigma/dy_t$)	[89]	5 (4/4)	$ y_t < 2.5$	$Q = m_t$
ATLAS $t\bar{t} \ell + \text{jets}$ 8 TeV ($1/\sigma d\sigma/dy_{t\bar{t}}$)	[89]	5 (4/4)	$ y_{t\bar{t}} < 2.5$	$Q = m_t$
[ATLAS $t\bar{t} \ell + \text{jets}$ 8 TeV ($1/\sigma d\sigma/dm_{t\bar{t}}$)]	[89]	7 (—/7)	$345 \leq m_{t\bar{t}} \leq 1600$ GeV	$Q = m_t$
ATLAS $t\bar{t} 2\ell$ 8 TeV ($1/\sigma d\sigma/dy_{t\bar{t}}$) (*)	[90]	5 (5/5)	$ y_{t\bar{t}} < 2.8$	$Q = m_t$
CMS $t\bar{t} \ell + \text{jets}$ 8 TeV ($1/\sigma d\sigma/dy_{t\bar{t}}$)	[91]	10 (9/9)	$-2.5 < y_{t\bar{t}} < 2.5$	$Q = m_t$
CMS $t\bar{t} 2D 2\ell 8$ TeV ($1/\sigma d\sigma/dy_t dm_{t\bar{t}}$) (*)	[92]	16 (16/16)	$ y_t < 2.5$	$340 \leq m_t \leq 1500$
CMS $t\bar{t} \ell + \text{jet}$ 13 TeV ($d\sigma/dy_t$) (*)	[93]	10 (10/10)	$ y_t < 2.5$	$Q = m_t$
CMS $t\bar{t} 2\ell$ 13 TeV ($d\sigma/dy_t$) (*)	[94]	11 (11/11)	$ y_t < 2.5$	$Q = m_t$
[ATLAS incl. jets 7 TeV, R=0.6]	[95]	90 (—/90)	$ y^{\text{jet}} < 3.0$	$100 \leq p_T^{\text{jet}} \leq 1992$
[CMS incl. jets 7 TeV]	[96]	133 (—/133)	$ y^{\text{jet}} < 2.5$	$100 \leq p_T^{\text{jet}} \leq 2000$
ATLAS incl. jets 8 TeV, R=0.6 (*)	[97]	171 (171/171)	$ y^{\text{jet}} < 3.0$	$70 \leq p_T^{\text{jet}} \leq 2500$
CMS incl. jets 8 TeV (*)	[98]	185 (185/185)	$ y^{\text{jet}} < 3.0$	$74 \leq p_T^{\text{jet}} \leq 2500$
ATLAS dijets 7 TeV, R=0.6 (*)	[99]	90 (90/90)	$0.0 \leq y^* \leq 3.0$	$260 \leq m_{jj} \leq 4270$
CMS dijets 7 TeV (*)	[100]	54 (54/54)	$ y_{\text{max}} < 2.5$	$200 \leq m_{jj} \leq 5000$
[CMS 3D dijets 8 TeV] (*)	[101]	122 (122/122)	$0.0 < y_b, y^* < 3.0$	$133 \leq p_{T,\text{avg}} \leq 1780$
[ATLAS isolated γ prod. 8 TeV] (*)	[102]	49 (—/—)	$ \eta_\gamma < 2.37$	$E_T^\gamma < 1500$
ATLAS isolated γ prod. 13 TeV (*)	[103]	53 (53/53)	$ \eta_\gamma < 2.37$	$E_T^\gamma < 1500$
ATLAS single $t R_t$ 7 TeV (*)	[104]	1 (1/1)	—	$Q = m_t$
CMS single $t \sigma_t + \sigma_{\bar{t}}$ 7 TeV (*)	[105]	1 (1/1)	—	$Q = m_t$
ATLAS single $t R_t$ 8 TeV (*)	[106]	1 (1/1)	—	$Q = m_t$
CMS single $t R_t$ 8 TeV (*)	[107]	1 (1/1)	—	$Q = m_t$
ATLAS single $t R_t$ 13 TeV (*)	[108]	1 (1/1)	—	$Q = m_t$
CMS single $t R_t$ 13 TeV (*)	[109]	1 (1/1)	—	$Q = m_t$
ATLAS single t 7 TeV ($1/\sigma d\sigma/dy_t$) (*)	[104]	4 (3/3)	$ y_t < 3.0$	$Q = m_t$
ATLAS single t 7 TeV ($1/\sigma d\sigma/dy_{\bar{t}}$) (*)	[104]	4 (3/3)	$ y_{\bar{t}} < 3.0$	$Q = m_t$
ATLAS single t 8 TeV ($1/\sigma d\sigma/dy_t$) (*)	[106]	4 (3/3)	$ y_t < 2.2$	$Q = m_t$
ATLAS single t 8 TeV ($1/\sigma d\sigma/dy_{\bar{t}}$) (*)	[106]	4 (3/3)	$ y_{\bar{t}} < 2.2$	$Q = m_t$

Table 2.5: As Tab. 2.1, this table presents information for other LHC processes. The entries, listed from top to bottom, include: W -boson production associated with either a charm jet or a light quark jet; transverse momentum production of Z -bosons; both total and differential top quark pair production; single-inclusive and dijet production; inclusive production of isolated photons; and single top t -channel production, covering both total and differential aspects.

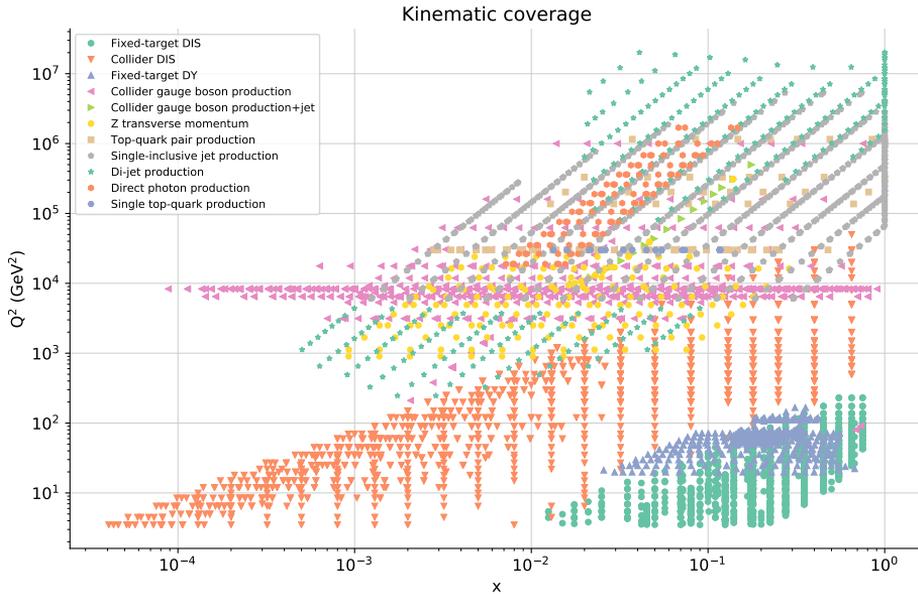


Figure 2.6: The kinematic coverage of the NNPDF4.0 dataset in the (x, Q^2) plane.

- CMS $t\bar{t}$ 2D dilepton 8 TeV
- CMS $t\bar{t}$ lepton+jets 13 TeV
- CMS $t\bar{t}$ dilepton 13 TeV
- CMS $\sigma_{t\bar{t}}^{\text{tot}}$ 5.02 TeV
- CMS $\sigma_{t\bar{t}}^{\text{tot}}$ 7, 8 TeV
- CMS $\sigma_{t\bar{t}}^{\text{tot}}$ 13 TeV
- LHCb Z 7 TeV ($\mathcal{L} = 940 \text{ pb}^{-1}$)
- LHCb $Z \rightarrow ee$ 8 TeV ($\mathcal{L} = 2 \text{ fb}^{-1}$)
- LHCb $W, Z \rightarrow \mu$ 7 TeV
- LHCb $W, Z \rightarrow \mu$ 8 TeV

Grids for all remaining datasets have been converted to the `PinEAPPL` format without recalculating them. Naturally, aside from the new FONLL implementation, none of these changes should affect the results, provided both the previous and new theoretical methods are numerically accurate. The new FONLL implementation introduces differences from the previous one due to sub-leading corrections, which result in NNLO variations in the NLO fit and N^3LO variations in the NNLO fit. These differences are limited to the charm and bottom mass corrections in DIS structure functions, impacting only a small number of data points at the NNLO level, and only by a fraction of a percent.

Finally, during the transition to the new pipeline, a few bugs were found in the implementation of some data points, including issues like incorrect normalization, incorrect scale assignment, or incorrect bin size. These corrections have a negligible impact in practice, as they affect only a handful of points out of more than 4500.

We have produced variants of the published NNPDF4.0 NLO and NNLO fits using the new theory pipeline. To demonstrate the equivalence between these fits and the

	NLO QCD		NNLO QCD	
	New	Published	New	Published
χ^2	1.26	1.24	1.17	1.16
$\langle E_{\text{tr}} \rangle$ rep	2.41 ± 0.06	2.43 ± 0.08	2.28 ± 0.05	2.27 ± 0.07
$\langle E_{\text{val}} \rangle$ rep	2.57 ± 0.10	2.62 ± 0.13	2.37 ± 0.11	2.35 ± 0.11
$\langle \chi^2 \rangle$ rep	1.29 ± 0.02	1.27 ± 0.02	1.20 ± 0.02	1.18 ± 0.02
$\langle \text{TL} \rangle$ rep	12900 ± 2000	13200 ± 2100	12400 ± 2600	13400 ± 2400
ϕ	0.156 ± 0.006	0.178 ± 0.007	0.153 ± 0.005	0.162 ± 0.005

Table 2.6: Comparison of statistical indicators for a set of 100 NNPDF4.0 NNLO PDF replicas generated with the new theory pipeline and the published NNPDF4.0 NNLO 100 replica set.

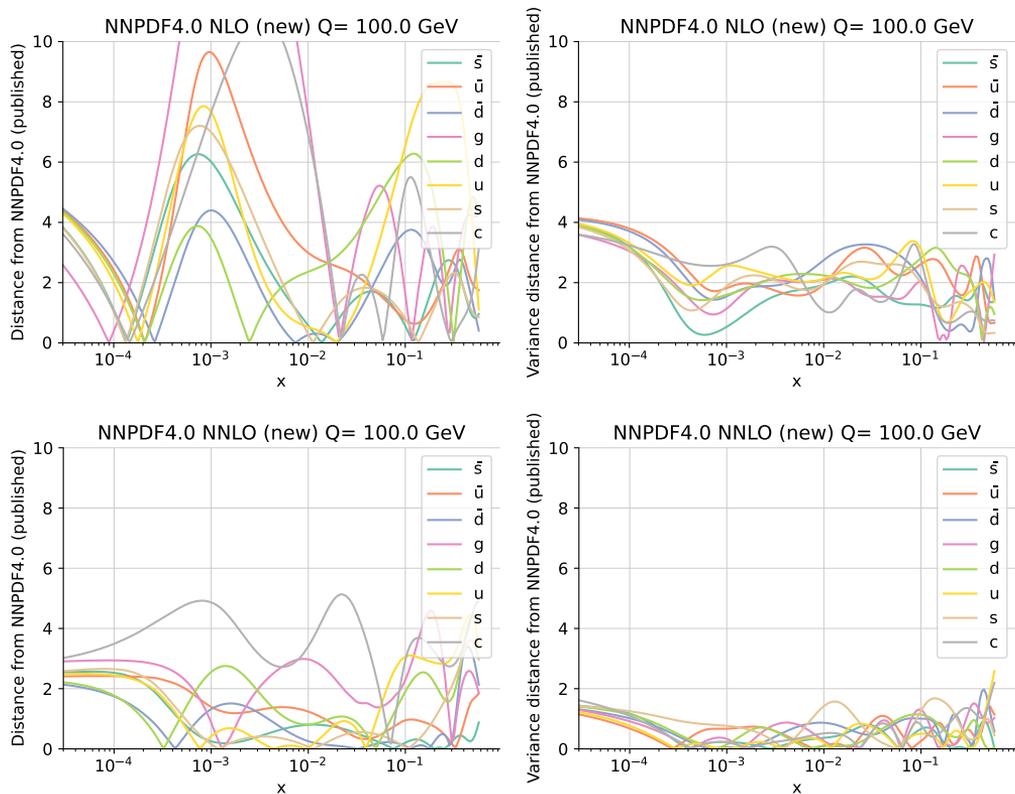


Figure 2.7: Distances at $Q = 100$ GeV between the central values (left) and uncertainties (right) of the 100 NNPDF4.0 PDF replicas at NLO (top) and NNLO (bottom) whose statistical indicators are compared in Tab. 2.6.

original NNPDF4.0 fits, which were generated using the previous theory pipeline, we compare each pair of fits. All fits consist of 100 replicas. The corresponding statistical estimators are presented in Tab. 2.6. Besides the fit quality estimators, we also include the ϕ estimator, defined in Equation (4.6) of Ref. [119]. This estimator measures the ratio of

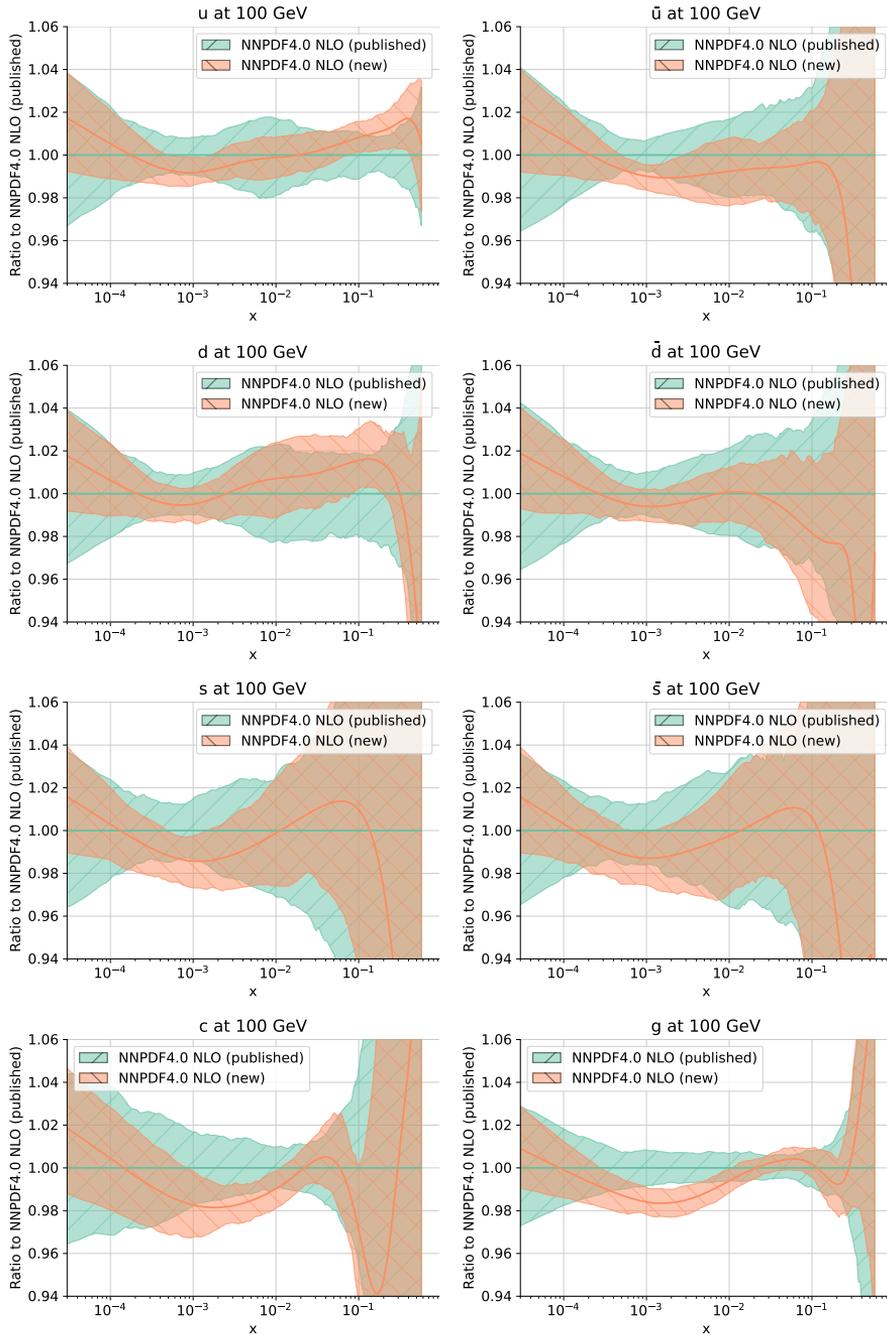


Figure 2.8: The NLO PDFs at $Q = 100$ GeV from the 100 replica sets compared in Tab. 2.6 and Fig. 2.7. Results are shown normalized to the central value of the published set. Bands correspond to 1σ uncertainties. From left to right and from top to bottom, we show the up, anti-up, down, anti-down, strange and gluon PDFs.

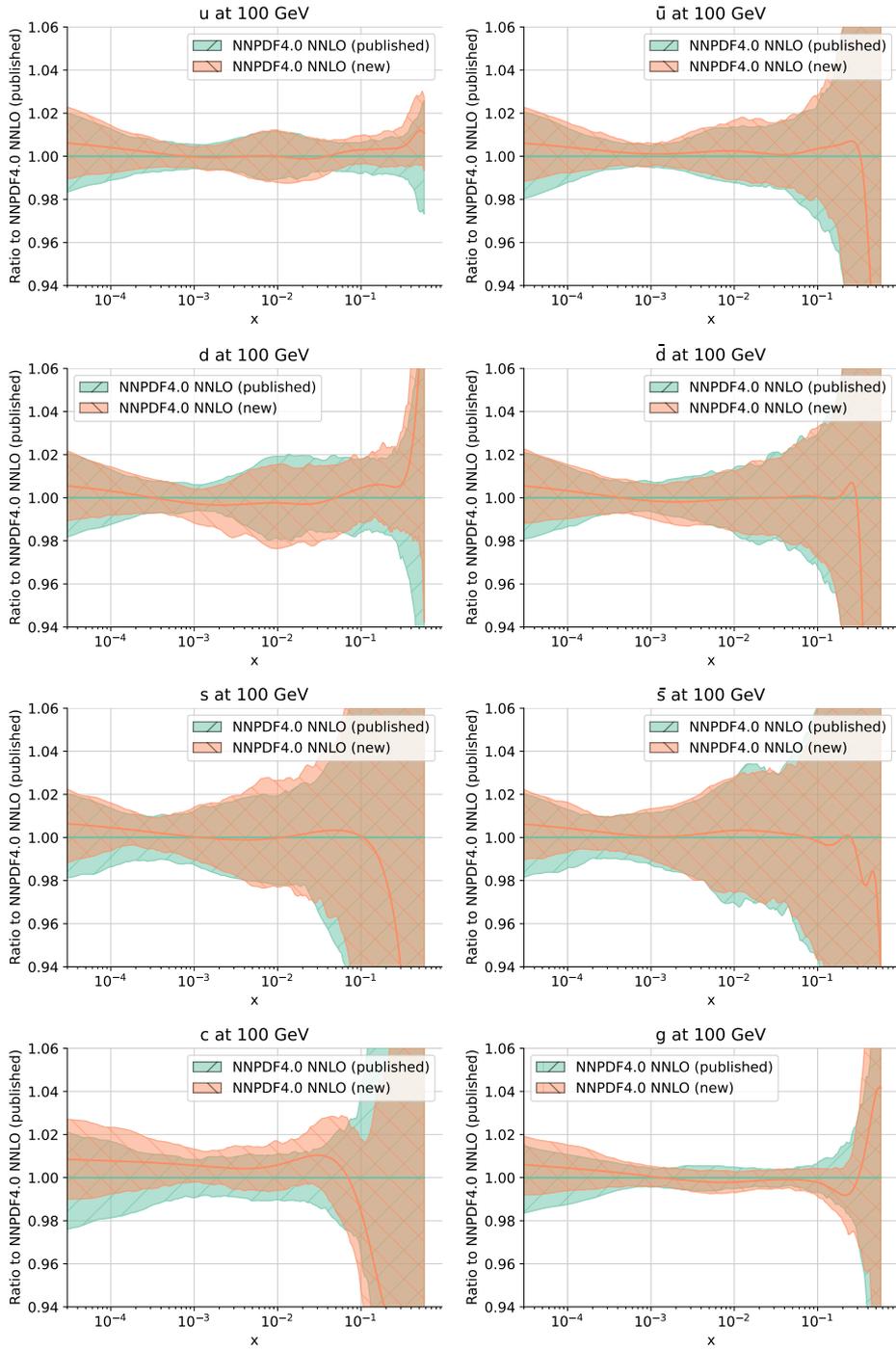


Figure 2.9: Same as Fig. 2.8 but at NNLO.

the average (correlated) PDF uncertainty on data points to the experimental uncertainty, making it a sensitive indicator of fit quality. All estimators match at NNLO (considering the statistical uncertainty on the χ^2 with the given number of data points is around 0.02), explicitly showing the equivalence between the two sets of replicas. At NLO, all estimators are also very close, with differences falling within the statistical uncertainty of the χ^2 .

PDFs distances are shown in Fig. 2.7, where the distances (as defined in Ref. [120]) between the central values and uncertainties are compared using two sets of replicas at $Q = 100$ GeV. When considering a set of 100 replicas, a distance of $d \sim 1$ indicates statistical similarity, while $d \sim 10$ signifies one-sigma deviations. At NNLO, distances are typically around 1, indicating that both replica sets are derived from the same underlying distribution. In contrast, at NLO, distances tend to be somewhat larger, approximately at the half-sigma level for most PDFs and occasionally reaching 1σ in specific instances, primarily due to differences in the FONLL implementation [117]. The actual PDFs themselves are displayed in Figs. 2.8-2.9, corresponding to NLO and NNLO, respectively. The comprehensive agreement observed between the two pairs of NNLO replica sets, alongside the minor discrepancies at NLO, is clearly evident.

From now on in this thesis, whenever we will refer to the NNPDF4.0 PDFs set, we will refer to the fit using the new theory baseline. Moreover, these PDFs sets are made available as a set of 100 Monte Carlo replicas both at NLO and at NNLO on the NNPDF collaboration website at the link

<https://nnpdf.mi.infn.it/nnpdf4-0-qcd/> ,

and are called

NNPDF40_nlo_as.01180_qcd ,
NNPDF40_nnlo_as.01180_qcd .

Part II

A Tale of Light and Shadow

“May it be a light to you in dark places, when all other lights go out.”

– Galadriel, *The Lord of the Rings: The Two Towers*

In Chap. 2 we discussed the fitting framework used in the NNPDF4.0 PDFs determination. As we already mentioned, this is a pure QCD NNLO fit. It means that QED effects are neglected and the QCD ones are treated at NNLO. Therefore, splitting functions, partonic cross sections and coefficient functions do not include corrections in the electromagnetic coupling α_{em} . Moreover, the photon content of the proton, i.e. the photon PDF, is neglected and for this reason we don’t include the photon initiated (PI) processes (processes in which the initial state parton is a photon) in the computation of hadronic observables. In this chapter we will describe how these kind of contributions are included in a PDFs fit and in particular how we included them in the NNPDF4.0 fitting framework.

Early efforts to create PDF sets that included QED effects used models for the photon PDF at the starting evolution scale [121, 122]. The first determination of QED PDFs driven by data, based on the NNPDF2.3 methodology [123, 124], produced a photon PDF with very large uncertainties. The fact that trying to determine a photon PDF from data results in large uncertainties was further confirmed in recent studies using the `xFitter` methodology [125], that used fixed functional forms to fit PDFs from data and analyzed high-mass ATLAS Drell-Yan distributions [72]. In 2016, a significant advancement in determining QED PDFs was made in Refs. [126, 127] (with related findings in Refs. [128, 129]). These studies demonstrated that the photon PDF could be calculated perturbatively in QED, provided the proton structure functions at all scales are given, ranging from the elastic ($Q^2 \rightarrow 0$) to the deep-inelastic ($Q^2 \rightarrow \infty$) regimes. Since then, this method, known as LuxQED, has become the foundation for all QED PDFs sets [130–134].

QED corrections enter a PDFs fit in different ways. First of all, given that we are allowing the quarks inside the proton to split into a quark-photon pair, we have to take into account also a photon PDF and therefore we have to consider PI contributions in the computation of theory predictions. Obviously, PI contributions will be suppressed with respect to the others since the photon PDF is much smaller than the quarks and gluon PDFs and since these partonic cross sections will be proportional to some power of α_{em} . The photon PDF affects the momentum sum rule, Eq. (1.44), subtracting part of the proton’s momentum from the sum of all the other PDFs. The second way QED corrections affect a PDFs fit is through DGLAP evolution. Indeed, in Eq. (1.65) we wrote the splitting functions as an expansion in powers of a_s . However, they admit corrections

also in α_{em} . For this reason, the photon mixes with the other PDFs through DGLAP evolution. This leads to a modification of the diagonalization of the DGLAP equation discussed in Sec. 1.6.

Even if the the inclusion of a photon PDF in a PDFs fits, weakly affects the quark and gluon PDFs, its impact is not always negligible with respect to the current size of the uncertainties. For example, the photon usually carries a fraction of the proton momentum that is roughly two orders of magnitude smaller than that of the gluon. Consequently, the resulting reduction in the gluon momentum fraction is relevant at the percent level. Hence, it is comparable to the present uncertainty in the gluon PDF. Moreover, we will show in this section that there are processes in which in some kinematical ranges (like the large rapidity region in Higgs production) the QED effects have a non negligible impact, resulting in a contribution of the order of 5%. Therefore, precise calculations of LHC processes that incorporate electroweak effects necessitate a consistent global QED PDFs determination.

In this chapter we will describe the way QED corrections are added to the NNPDF4.0 fitting framework. From now on a PDFs fit that includes QED corrections will be referred to as QED fit, while a fit that doesn't involve them will be called QCD fit. This chapter is organized as follows: in Sec. 3.1 we will briefly summarize the LuxQED approach, on which the photon PDF determination of this work is based, and we will describe how it was interfaced to the NNPDF4.0 fitting methodology; in Sec. 3.2 we will explain how the DGLAP equations, see Sec. 1.6, are modified by the inclusion of QED effects and how to solve them in this different case; in Sec. 3.3 we will describe the benchmarks we performed in order to check that our implementation of mixed QCD \otimes QED DGLAP evolution is correct; in Sec. 3.4 we will present the results of the QED fit, comparing it with NNPDF4.0 and with the most recent QED PDFs determinations; in the end, in Sec. 3.5 we will study the impact of including QED corrections to PDFs to phenomenological processes of interest for the LHC.

3.1 Photon PDF

The photon PDF is somehow special: indeed it is not fitted like the other PDFs, but it is obtained with the LuxQED approach in which it is linked to the other PDFs through a perturbative calculation. In this section we are going to briefly describe the LuxQED approach that is adopted also in our QED PDFs determination.

In the original LuxQED papers [126,127] the authors showed that the existing electron-proton scattering data contains all the necessary information to precisely determine the photon PDF. Typically, ep scattering is viewed as a process where a photon emitted by the electron probes the proton's structure. However, it can equally be considered as an electron probing the photon field produced by the proton itself. Consequently, the ep scattering cross section is intrinsically linked to the photon PDF. This is shown explicitly considering, instead of DIS, the fictitious process $l + p \rightarrow L + X$, where l and L are neutral leptons, with the first one being massless and the second one being massive. Then, they observe that there are two ways to estimate the heavy-lepton production cross section: the first one uses the standard proton structure functions, F_2 and F_L , while the second one utilizes the proton PDFs, with the dominant contribution coming from the photon PDF. By equating these two computations, one can determine f_γ . This leads to

the so-called LuxQED formula, that reads

$$x f_\gamma(x, \mu^2) = \frac{1}{2\pi\alpha_{em}(\mu^2)} \int_x^1 \frac{dz}{z} \left\{ \int_{\frac{x^2 m_p^2}{1-z}}^{\frac{\mu^2}{1-z}} \frac{dQ^2}{Q^2} \alpha_{em}^2(Q^2) \left[-z^2 F_L\left(\frac{x}{z}, Q^2\right) + \left(z P_{\gamma q}(z) + \frac{2x^2 m_p^2}{Q^2} \right) F_2\left(\frac{x}{z}, Q^2\right) \right] - \alpha_{em}^2(\mu^2) z^2 F_L\left(\frac{x}{z}, Q^2\right) \right\}, \quad (3.1)$$

where m_p is the mass of the proton and $P_{\gamma q}$ is the photon-quark splitting function. Eq. (1.40) holds in the $\overline{\text{MS}}$ scheme and neglects terms of order $\mathcal{O}(\alpha_s \alpha_{em})$ and $\mathcal{O}(\alpha_{em}^2)$. Being the DIS structure functions computed through the PDFs according to Eq. (1.40), Eq. (3.1) is linking the photon PDF to all the other PDFs. However, a remark is needed: Eq. (3.1) doesn't give a way to compute the photon PDF starting from quarks and gluon PDFs. Indeed, the PDFs that appear inside the computation of F_2 and F_L are implicitly QED dependent since they depend on the DGLAP evolution, that in presence of QED corrections evolve mixing with the photon, and they are obtained in a QED fit in which the effects of the photon PDF are taken into account. For this reasons, Eq. (3.1) must be seen as a constraint that the fitting algorithm must satisfy through all the steps of the PDFs determination, like the sum rules, rather than as a way of computing the photon PDF.

Now that we have introduced the LuxQED formula, we can explain how it is actually computed. We have seen that the LuxQED formula depends on DIS structure functions. First of all we observe that there are two kind of contributions, the elastic and the inelastic electron-proton scattering, and the total will be given by the sum of the two, i.e.

$$F_a = F_a^{(\text{inel})} + F_a^{(\text{el})}. \quad (3.2)$$

The elastic scattering is the one in which an electron scatters off a proton without enough energy to resolve its internal structure. Then, in this case we still have the proton in the final state and the structure functions will be proportional to $\delta(1-z)$. In this case their expression reads

$$F_2^{(\text{el})}(x, Q^2) = \frac{G_E^2(Q^2) + G_M^2(Q^2)\tau}{1 + \tau} \delta(1-x), \quad (3.3)$$

$$F_L^{(\text{el})}(x, Q^2) = \frac{G_E^2(Q^2)}{\tau} \delta(1-x), \quad (3.4)$$

where $\tau = Q^2/(4m_p^2)$ and G_E and G_M are called electric and magnetic Sachs form factors of the proton [135]. For G_E and G_M it is usually used the dipole approximation, in which they read

$$G_E(Q^2) = \frac{1}{\left(1 + \frac{Q^2}{m_{\text{dip}}^2}\right)^2}, \quad (3.5)$$

$$G_M(Q^2) = \mu_p G_E(Q^2), \quad (3.6)$$

with $m_{\text{dip}}^2 = 0.71 \text{ GeV}^2$ and $\mu_p \simeq 2.793$. Instead, the inelastic contribution is the one in which the electron has enough energy to break the internal structure of the proton and the structure functions are computed within the parton model with the formula in Eq. (1.40). However, observing Eq. (3.1) it is clear that the Q^2 integration extends down

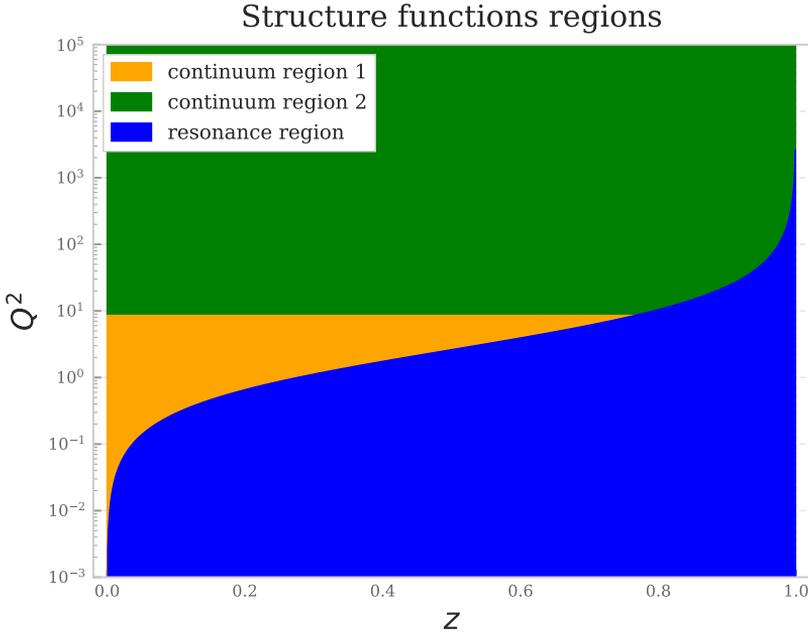


Figure 3.1: Division of the integration region of Eq. (3.1) into subregions in which a different way to compute the structure functions is used: in the resonance region (blue) they are computed from the fit of the CLAS collaboration; in the continuum region 1 (orange) they are computed from the GD11-P fit by HERMES collaboration; in the continuum region 2 they are computed in perturbation theory with Eq. (1.40).

to very small scales, where the perturbative computation of the DIS structure functions is no longer reliable. For this reason, at very low scale the DIS structure functions are fitted from experimental data. In this analysis, the LuxQED formula is computed with the public code `FiatLux` [130], in which the inelastic structure functions are computed from two separate integration regions that are divided by $W^2 = m_p^2 + Q^2(1-z)/z$:

- In the resonance region, defined as $W^2 \leq 3.5 \text{ GeV}^2$, the structure functions are computed from the fit of the CLAS collaboration [136], using also the parameterization of Ref. [137] in order to assess the uncertainty of it.
- The continuum region, defined as $W^2 \geq 3.5 \text{ GeV}^2$, is itself divided into two subregions according to Q^2 . For $Q^2 \leq 9 \text{ GeV}^2$, the structure functions are given by the GD11-P fit by HERMES collaboration [138, 139]. For $Q^2 \geq 9 \text{ GeV}^2$, the structure functions are computed in perturbation theory with Eq. (1.40).

Fig. 3.1 shows the division of the integration space into the different regions.

Now we are ready to see how the constraint of Eq. (3.1) is implemented. Indeed, one option would be to compute at every stage of the fit the photon PDF, for every point of the x grid, through the LuxQED formula. However, since it involves a double integration, that is computationally very heavy to perform, this approach would slow down the fit in a way that it would become impossible to perform in a reasonable amount of time. Therefore, according to the procedure developed for the NNPDF3.1QED [130]

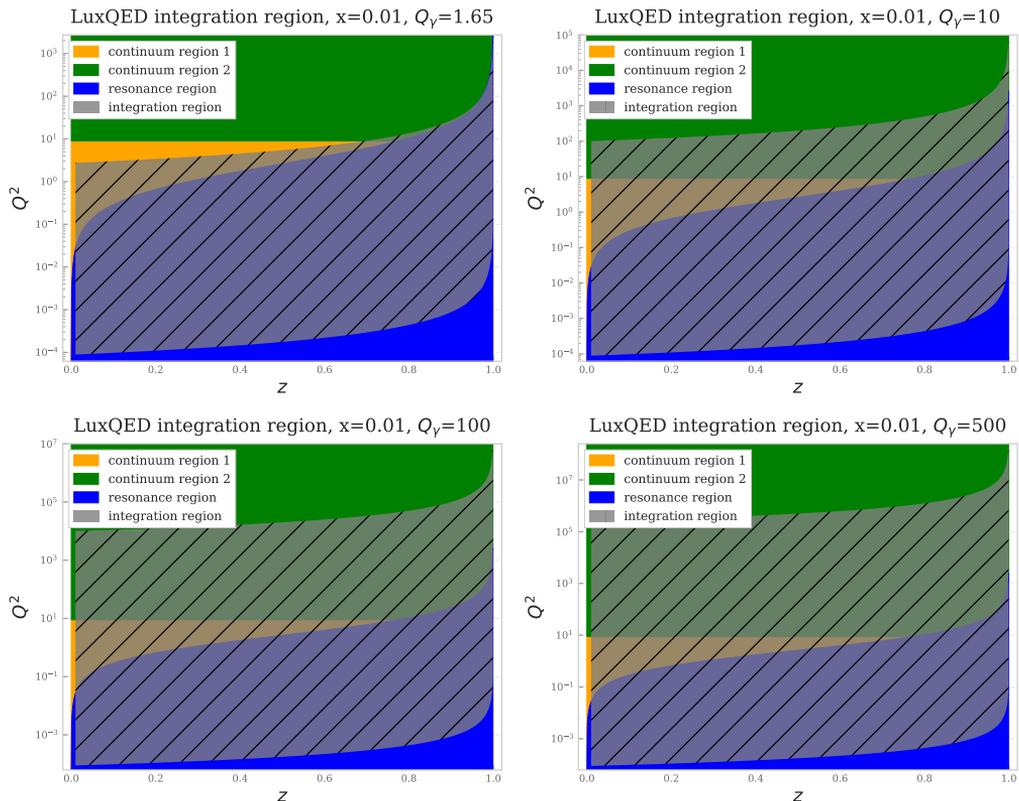


Figure 3.2: Integration region (gray) of the LuxQED formula, Eq. (3.1), for $x = 0.01$ and for different values of Q_γ , i.e. $Q_\gamma = 1.65$ GeV, 10 GeV, 100 GeV and 500 GeV.

PDFs determination, the constraint of Eq. (3.1) is implemented in an iterative way. The algorithm that is used to perform a QED fit is the following: first of all we start from a previous pure QCD fit, that in our case is NNPDF4.0. Then, we use such QCD PDFs to compute the structure functions needed to compute the photon PDF with Eq. (3.1). At the end of this step we have obtained $f_\gamma(Q_0^2)$. Once we have the photon PDF at fitting scale Q_0 we perform the QED fit in which the photon is kept constant, and the quark and gluon PDFs are fitted. The photon enters in the fit in three possible ways. First, it modifies the momentum sum rule of Eq. (1.44), that in presence of a photon PDF becomes

$$\int_0^1 dz (z\Sigma(z, Q_0^2) + zg(z, Q_0^2) + z\gamma(z, Q_0^2)) = 1. \quad (3.7)$$

It means that the photon is subtracting momentum from the sum of the other PDFs. Second, the photon modifies the DGLAP evolution of the PDFs. Indeed, in Sec. 3.2 we will see that it evolves coupled with quarks and gluons. Last, the presence of a photon PDF enables the PI partonic processes, i.e. diagrams in which we have a photon as initial particle. However, in this analysis we did not include PI contributions. The reason is that in the NNPDF4.0 PDFs determination, the dataset has been chosen in such a way that the PI contributions are negligible in the theory predictions for those data points. Moreover, defining the QED corrections for the DIS structure functions is not straightforward and

therefore also in that case the predictions include only diagrams that do not involve PI processes. Once that we have fitted the quark and gluon PDFs in presence of the photon that we have generated, we have a first QED fit. At this point the procedure is iterated: the QED fit that we have performed is used to recompute the DIS structure functions that are used in the LuxQED formula, Eq. (3.1). In this way a second photon is obtained and it is used to perform a second QED fit. This procedure is iterated until convergence is achieved: the QED fit is considered converged when the photon obtained at a certain step is compatible with the previous one within the uncertainties. It is observed that two iterations are enough to achieve convergence.

Now we can discuss about the generation of the photon PDF with Eq. (3.1), given a certain PDFs set. One possibility is to apply the LuxQED formula using $Q_\gamma = Q_0$, where we call Q_γ the scale at which the formula is applied. However, the integration region of Eq. (3.1) strongly depends on the choice of Q_γ . Indeed, in Fig. 3.2 (up-left) we can observe that with this choice of Q_γ the resulting photon is entirely determined by the structure functions fitted from data. This is not an ideal situation since we are much more confident in the structure functions obtained from the convolution of coefficient functions and PDFs than in the ones fitted from experimental data. Increasing the value of Q_γ to 10 GeV (up-right plot) we can observe that the integration region extends to the zone in which the structure functions are computed through PDFs, but the contribution coming from the experimental fits is still dominant. If we instead look at the lower plots of Fig. 3.2, where we used $Q_\gamma = 100$ GeV and 500 GeV, we can observe that, despite the fact that the integration on the regions in which the structure functions are fitted from data is still present, now the part of the integral in which the structure functions are computed through the PDFs is larger. In Fig. 3.3 it is shown the photon PDF obtained with the LuxQED formula, split in the different contributions, for the values of Q_γ that we discussed in Fig. 3.2. In particular, the photon PDF is obtained through the sum of three contributions: the elastic part (red), obtained computing the structure functions from Eqs. (3.3-3.4), the inelastic non perturbative part (blue), where the structure functions are fitted from data and the inelastic perturbative part (green), in which the structure functions are computed from PDFs. Observe that the blue contribution of Fig. 3.3 is the union of the blue and orange integration regions of Fig. 3.1. It is clear that when we compute the photon at $Q_\gamma = 100$ GeV or at $Q_\gamma = 500$ GeV, the inelastic perturbative part is the dominant contribution, being roughly one order of magnitude larger than the others. This is relevant since, as highlighted in Ref. [131], only the inelastic part of the photon PDF evolves with DGLAP equations coupled with the other quarks. The elastic component of the photon instead, evolves decoupled from all the other PDF with an evolution equation that reads

$$\mu^2 \frac{d}{d\mu^2} \gamma^{(\text{el})} = P_{\gamma\gamma} \otimes \gamma^{(\text{el})} + \delta\gamma, \quad (3.8)$$

where, see Ref. [131],

$$\delta\gamma(x, \mu^2) = \frac{\alpha_{em}(\mu^2)}{2\pi} \frac{1}{x} \left[\left(x P_{\gamma q}(x) + \frac{2x^2 m_p^2}{\mu^2} \right) \frac{G_E^2(\mu^2) + G_M^2(\mu^2)\tau}{1 + \tau} - x^2 \frac{G_E^2(\mu^2)}{\tau} \right]. \quad (3.9)$$

However, since when we compute the photon at large scale we have that the elastic part is smaller than the inelastic, except for very large x values where the photon PDF is small, we neglected the fact that the first one evolves with a different equation, and we

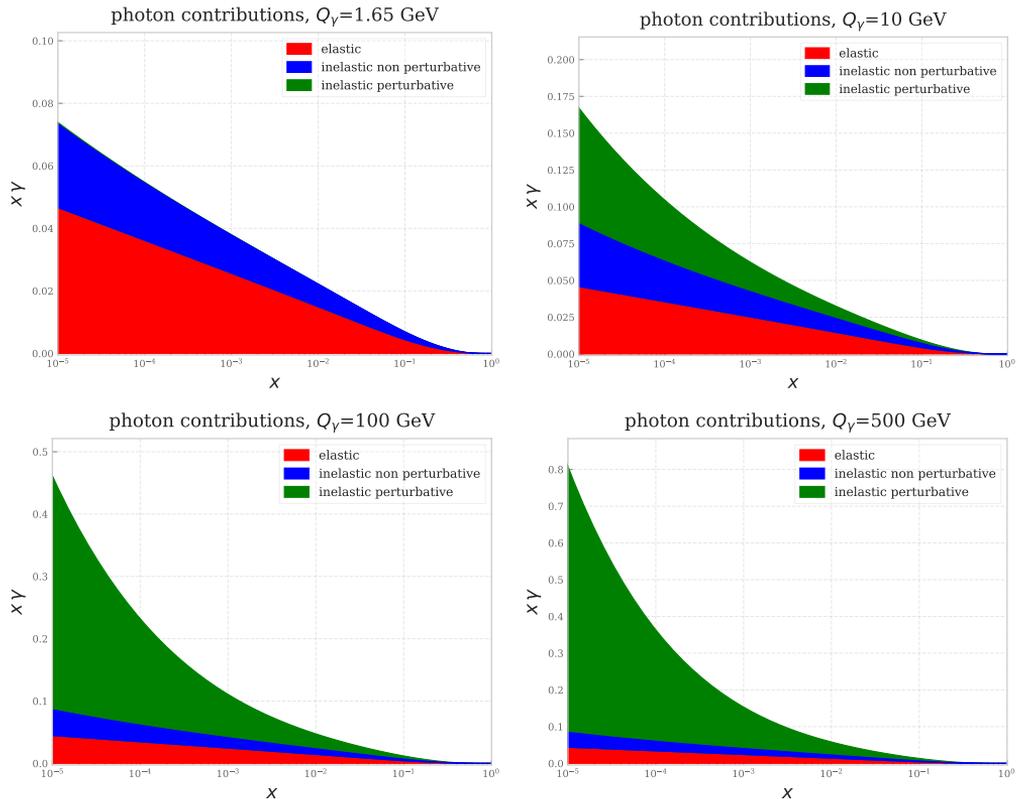


Figure 3.3: Breakdown of the photon PDF $\gamma(x, Q_\gamma^2)$ into the contributions originating from various components of the proton structure functions $F_i(x, Q^2)$ that determine it according to the LuxQED formula in Eq. (3.1). The results are shown as a function of x for four different scale choices of Q_γ : $Q_\gamma = 1.65$ GeV (top left), 10 GeV (top right), 100 GeV (bottom left), and 500 GeV (bottom right).

evolved the total PDF with DGLAP equations. In the end, Eq. (3.1) is neglecting higher twist corrections. Therefore, applying it at a high scale minimizes such unknown corrections. For all these reasons, in the NNPDF4.0QED fit the photon has been generated at $Q_\gamma = 100$ GeV, according to the suggestion made by the authors of the LuxQED approach [127], and then evolved back to the fitting scale Q_0 using backward DGLAP evolution. In this way we have the photon at fitting scale to be used in the fit. It necessary to observe that the fit is not directly performed at the scale at which the photon is generated since the fitting algorithm performs in a better way if the fitting scale is chosen to be small, see App. C for further details. Fig. 3.4 shows schematically what we discussed so far, i.e. how a QED fit is performed.

In order to conclude this section we will discuss the uncertainties of the photon PDF. We already discussed that the LuxQED formula depends on the proton structure functions, that are in turn determined from different contributions: the elastic part, the resonance component, the low Q continuum region and the high Q continuum region. Only the last one is computed from the PDFs. Therefore, the N_{rep} Monte Carlo replicas of the photon will account only for the uncertainty of the high Q continuum region, that comes from the Monte Carlo replicas of the PDFs set we used to compute the photon.

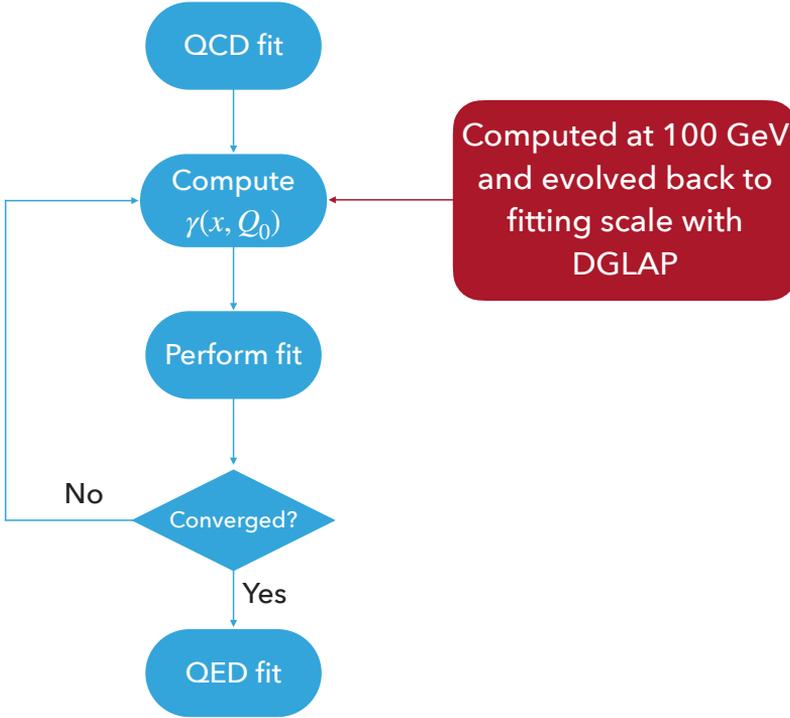


Figure 3.4: Flowchart illustrating the NNPDF4.0QED fitting strategy. In the final iteration, after the procedure has converged, the additional LUXqed17 contributions are incorporated into $\gamma(x, Q^2)$, as described in this section.

In order to correctly estimate the uncertainties of the photon PDF, we have to include other sources of error. The following sources of uncertainty are taken into account [127]: the uncertainty of the elastic contribution that comes from the A1 world proton form factor fits [140]; the uncertainty given by replacing the A1 world polarized fit just with the unpolarized data; the parametrization of the structure functions in the resonance region [136–138]; the parametrization of $R_{L/T}$ [139, 141, 142], which is the ratio between longitudinal and transverse structure functions; the scale Q_{match}^2 where we switch from low- Q^2 to high- Q^2 inelastic structure functions; a twist-4 modification of the structure function F_L [143, 144]; and in the end, an estimation of the missing higher-order corrections in the computation of the structure functions at high Q^2 . According to the methodology developed for NNPDF3.1QED, these additional uncertainties are introduced in the last iteration of the fitting procedure: when we compute the photon of the final QED fit, for each replica of the $n_{\text{sys}} = 7$ additional uncertainties of the LUXqed17 [127] fit are included as statistical fluctuations of the photon computed with the LuxQED formula at the scale Q_γ . It means that the final photon replicas at Q_γ will be given by

$$\tilde{\gamma}^{(k)}(x, Q_\gamma^2) = \gamma^{(k)}(x, Q_\gamma^2) + \sum_{j=1}^{n_{\text{sys}}} \delta\gamma_j^{(\text{luxe})}(x, Q_\gamma^2) \cdot \mathcal{N}(0, 1), \quad k = 1, \dots, N_{\text{rep}}, \quad (3.10)$$

where $\gamma^{(k)}(x, Q_\gamma^2)$ is the photon computed from the LuxQED formula, $\mathcal{N}(0, 1)$ is a uni-

variate Gaussian random number and $\delta\gamma_j^{(\text{lux})}$ is the normalized eigenvector for the j -th systematic uncertainty in the LUXqed17 PDFs set. In practice, $\delta\gamma_j^{(\text{lux})}(x, Q_\gamma^2)$ is obtained diagonalizing the covariance matrix for the extra LUXqed17 uncertainties, defined on the x grid of the fit, using a method that is similar to the one of Refs. [145,146]. This method is completely equivalent to using the Hessian eigenvectors of the PDFs set LUXqed17. The photon PDF $\tilde{\gamma}^{(k)}(x, Q_\gamma^2)$ is what is used in the last iteration of the fit.

3.2 DGLAP evolution

In this section we will describe how the QED corrections affect the DGLAP evolution equations. Indeed, in presence of QED effects, and therefore in presence of a photon PDF, DGLAP equations are written as

$$\mu^2 \frac{d}{d\mu^2} f_i(x, \mu^2) = \sum_{j=q, \bar{q}, g, \gamma} \int_x^1 \frac{dz}{z} P_{ij} \left(\frac{x}{z}, a_s(\mu^2), a_{em}(\mu^2) \right) f_j(z, \mu^2), \quad i = q, \bar{q}, g, \gamma, \quad (3.11)$$

where we defined a_{em} accordingly to a_s , Eq. (1.64), as

$$a_{em} = \frac{\alpha_{em}}{4\pi}. \quad (3.12)$$

Observe that in Eq. (3.11) the splitting functions have acquired a dependence on the electromagnetic coupling and we added the photon PDF that mixes with the quark and gluon PDFs. The splitting functions satisfy the double expansion in the two couplings, i.e.

$$P_{ij}(z, a_s, a_{em}) = \sum_{\substack{n,m=0 \\ (n,m) \neq (0,0)}}^{\infty} a_s^n a_{em}^m P_{ij}^{(n,m)}(z) \quad (3.13)$$

$$= a_s P_{ij}^{(1,0)}(z) + a_s^2 P_{ij}^{(2,0)}(z) + a_s^3 P_{ij}^{(3,0)}(z) \\ + a_{em} P_{ij}^{(0,1)}(z) + a_s a_{em} P_{ij}^{(1,1)}(z) + a_{em}^2 P_{ij}^{(0,2)}(z) + \dots \quad (3.14)$$

Observe that we used a different notation for the pure QCD splitting functions with respect to the one used in Chap. 1. In fact now we are specifying the absolute order in the couplings and not the relative one. It means that $P_{ij}^{(n,0)} \equiv P_{ij}^{(n-1)}$. The only known QED corrections to splitting functions are the ones of $\mathcal{O}(a_{em})$, $\mathcal{O}(a_s a_{em})$ and $\mathcal{O}(a_{em}^2)$ [147,148]. Regarding the pure QCD splitting functions instead, as we already discussed, they are exactly known up to NNLO in the strong coupling (that corresponds to $\mathcal{O}(a_s^3)$ since they start at $\mathcal{O}(a_s)$). In this work we used all and only the terms that are fully known. It means that we included the pure QCD splitting functions up to NNLO, i.e. the terms $P_{ij}^{(1,0)}$, $P_{ij}^{(2,0)}$ and $P_{ij}^{(3,0)}$, plus the QED corrections that are currently known, i.e. $P_{ij}^{(0,1)}$, $P_{ij}^{(1,1)}$ and $P_{ij}^{(0,2)}$.

When we add QED corrections to DGLAP equations, we have also to take into account QED corrections to the running of the couplings a_s and a_{em} . Indeed, they satisfy

coupled RGEs that read

$$\mu^2 \frac{d}{d\mu^2} a_s(\mu^2) = \beta_{\text{QCD}}(a_s, a_{em}) = -a_s^2 \left(\beta_{\text{QCD}}^{(2,0)} + a_s \beta_{\text{QCD}}^{(3,0)} + a_{em} \beta_{\text{QCD}}^{(2,1)} + a_s^2 \beta_{\text{QCD}}^{(4,0)} + \dots \right), \quad (3.15)$$

$$\mu^2 \frac{d}{d\mu^2} a_{em}(\mu^2) = \beta_{\text{QED}}(a_s, a_{em}) = -a_{em}^2 \left(\beta_{\text{QED}}^{(0,2)} + a_{em} \beta_{\text{QED}}^{(0,3)} + a_s \beta_{\text{QED}}^{(1,2)} + a_{em}^2 \beta_{\text{QED}}^{(0,4)} + \dots \right). \quad (3.16)$$

Observe that with respect to Eq. (1.24), Eqs. (3.15-3.16) are written in terms of a_s and not in terms of α_s . This choice will only affect the definition of the beta function coefficients, that will change by factors of 4π . Eqs. (3.15-3.16) form two coupled non linear differential equations, that must be solved simultaneously. In order to be consistent with the terms that we included in the splitting functions, in this analysis we used the QCD beta function up to NNLO in QCD, i.e. the terms $\beta_{\text{QCD}}^{(2,0)}$, $\beta_{\text{QCD}}^{(3,0)}$ and $\beta_{\text{QCD}}^{(4,0)}$, the QED beta functions up to NLO in QED, i.e. $\beta_{\text{QED}}^{(0,2)}$ and $\beta_{\text{QED}}^{(0,3)}$, and the two leading mixed corrections, i.e. $\beta_{\text{QCD}}^{(2,1)}$ and $\beta_{\text{QED}}^{(1,2)}$.

3.2.1 Evolution basis

Now we are ready to show how the DGLAP equations in presence of QED corrections are solved. First of all, we need to maximally decouple the system, in a way analogous to the way we maximally decoupled the pure QCD case in Sec. 1.6. However, now the flavor symmetry is broken by the different charges of the different kind of quarks. Therefore, we have to define the up-like and down-like singlet and valence distributions as

$$\Sigma_u = \sum_{k=1}^{n_u} u_k^+, \quad \Sigma_d = \sum_{k=1}^{n_d} d_k^+, \quad (3.17)$$

$$V_u = \sum_{k=1}^{n_u} u_k^-, \quad V_d = \sum_{k=1}^{n_d} d_k^-, \quad (3.18)$$

where n_u and n_d are respectively the numbers of active up-like and down-like quarks and $n_u + n_d = n_f$. From Eqs. (3.17-3.18) we construct the singlet and valence sums and differences:

$$\Sigma = \Sigma_u + \Sigma_d, \quad \Sigma_\Delta = \frac{n_d}{n_u} \Sigma_u - \Sigma_d, \quad (3.19)$$

$$V = V_u + V_d, \quad V_\Delta = \frac{n_d}{n_u} V_u - V_d, \quad (3.20)$$

Moreover, we construct the up-like and down-like non-singlet combinations as

$$T_3^u = u^+ - c^+, \quad T_3^d = d^+ - s^+, \quad (3.21)$$

$$V_3^u = u^- - c^-, \quad V_3^d = d^- - s^-, \quad (3.22)$$

$$T_8^u = u^+ + c^+ - 2t^+, \quad T_8^d = d^+ + s^+ - 2b^+, \quad (3.23)$$

$$V_8^u = u^- + c^- - 2t^-, \quad V_8^d = d^- + s^- - 2b^-. \quad (3.24)$$

Then, we split the splitting functions as

$$P_{ij}(x, a_s, a_{em}) = P_{ij}(x, a_s) + \tilde{P}_{ij}(x, a_s, a_{em}), \quad (3.25)$$

where P_{ij} contains only the pure QCD terms, while \tilde{P}_{ij} contains both the pure QED ones and the mixed QCD \otimes QED terms. With these definitions it is straightforward to write the form of the DGLAP equations in this basis. Indeed, the non-singlet distributions evolve according to

$$\mu^2 \frac{d}{d\mu^2} T_{3/8}^{u/d} = \left(P^+ + \tilde{P}_{u/d}^+ \right) \otimes T_{3/8}^{u/d}, \quad (3.26)$$

$$\mu^2 \frac{d}{d\mu^2} V_{3/8}^{u/d} = \left(P^- + \tilde{P}_{u/d}^- \right) \otimes V_{3/8}^{u/d}. \quad (3.27)$$

The valence sum and difference distributions satisfy the coupled equation

$$\mu^2 \frac{d}{d\mu^2} \begin{pmatrix} V \\ V_\Delta \end{pmatrix} = \begin{pmatrix} P^V + \langle \tilde{P}_q^- \rangle & \nu_u \tilde{P}_{\Delta q}^- \\ \nu_d \tilde{P}_{\Delta q}^- & P^- + \{ \tilde{P}_q^- \} \end{pmatrix} \otimes \begin{pmatrix} V \\ V_\Delta \end{pmatrix}, \quad (3.28)$$

where we defined $\nu_{u/d} = \frac{n_{u/d}}{n_f}$ and

$$\langle \tilde{P}_q^\pm \rangle = \nu_u \tilde{P}_u^\pm + \nu_d \tilde{P}_d^\pm, \quad \{ \tilde{P}_q^\pm \} = \nu_d \tilde{P}_u^\pm + \nu_u \tilde{P}_d^\pm, \quad \tilde{P}_{\Delta q}^\pm = \tilde{P}_u^\pm - \tilde{P}_d^\pm. \quad (3.29)$$

In the end we have a four dimensional system for the singlets sum and difference, the gluon and the photon distribution that read

$$\mu^2 \frac{d}{d\mu^2} \begin{pmatrix} g \\ \gamma \\ \Sigma \\ \Sigma_\Delta \end{pmatrix} = \mathbf{P} \otimes \begin{pmatrix} g \\ \gamma \\ \Sigma \\ \Sigma_\Delta \end{pmatrix}, \quad (3.30)$$

where

$$\mathbf{P} = \begin{pmatrix} P_{gg} + \tilde{P}_{gg} & \tilde{P}_{g\gamma} & P_{gq} + \langle \tilde{P}_{gq} \rangle & \nu_u \tilde{P}_{g\Delta q} \\ \tilde{P}_{\gamma g} & \tilde{P}_{\gamma\gamma} & \langle \tilde{P}_{\gamma q} \rangle & \nu_u \tilde{P}_{\gamma\Delta q} \\ 2n_f(P_{qq} + \langle \tilde{P}_{qq} \rangle) & 2n_f \langle \tilde{P}_{q\gamma} \rangle & P_{qq} + \langle \tilde{P}_q^+ \rangle + \langle e_q^2 \rangle^2 \tilde{P}_{ps} & \nu_u \tilde{P}_{\Delta q}^+ + \nu_u e_{\Delta q}^2 \langle e_q^2 \rangle \tilde{P}_{ps} \\ 2n_f \nu_d \tilde{P}_{\Delta qg} & 2n_f \nu_d \tilde{P}_{\Delta q\gamma} & \nu_d \tilde{P}_{\Delta q}^+ + \nu_d e_{\Delta q}^2 \langle e_q^2 \rangle \tilde{P}_{ps} & P^+ + \{ \tilde{P}_q^+ \} + \nu_u \nu_d \langle e_{\Delta q}^2 \rangle^2 \tilde{P}_{ps} \end{pmatrix},$$

where \tilde{P}_{ps} is defined from $\tilde{P}_{qq'}^{ps} = e_q^2 e_{q'}^2 \tilde{P}_{ps}$ [148], and the combinations $\langle \tilde{P}_{gq} \rangle$, $\langle \tilde{P}_{qq} \rangle$, $\tilde{P}_{g\Delta q}$, $\tilde{P}_{\Delta qq}$, $\langle \tilde{P}_{\gamma q} \rangle$, $\langle \tilde{P}_{q\gamma} \rangle$, $\tilde{P}_{\gamma\Delta q}$ and $\tilde{P}_{\Delta q\gamma}$ are constructed analogously to the ones defined in Eq. (3.29).

3.2.2 Matching and rotation at heavy quarks thresholds

Let us now see how the matching described in Sec. 1.9 applies to the new basis. For the g , Σ and V distributions the matching is identical to the one given in Eq. (1.159) and Eqs. (1.164-1.165). From Eqs. (1.162-1.163) it follows that all the non-singlet combinations defined in Eqs. (3.21-3.24) are matched diagonally, i.e.

$$\left(T_{3/8}^{u/d} \right)^{[n_f+1]} = A_{qq}^{ns} \left(T_{3/8}^{u/d} \right)^{[n_f]}, \quad (3.31)$$

$$\left(V_{3/8}^{u/d} \right)^{[n_f+1]} = A_{qq}^{ns} \left(V_{3/8}^{u/d} \right)^{[n_f]}. \quad (3.32)$$

Also the singlet and valence differences, i.e. Σ_Δ and V_Δ , are matched diagonally. For V_Δ it is a direct consequence of Eq. (1.163), while for Σ_Δ it is due to the fact that the off-diagonal terms exactly cancel, thanks to its definition. Therefore, we have that

$$\Sigma_\Delta^{[n_f+1]} = A_{qq}^{\text{ns}} \Sigma_\Delta^{[n_f]}, \quad (3.33)$$

$$V_\Delta^{[n_f+1]} = A_{qq}^{\text{ns}} V_\Delta^{[n_f]}. \quad (3.34)$$

We remark that in this work we did not consider QED corrections to the matching conditions since they are not currently known and they are a very small correction.

In this basis, when we cross an heavy quark threshold, the basis rotation we have to perform is more complicated than the one in the pure QCD case. For the singlet and valence distributions, being identical in the two cases, the basis rotation is performed in the same way of Eqs. (1.170-1.171). For the singlet and valence differences, the situation is more complicated. Let's see the singlet since for the valence the rotation is identical. We want to write $\Sigma_{\Delta(n_f+1)}$ as a combination of $\Sigma_{(n_f)}$, $\Sigma_{\Delta(n_f)}$ and h_+ . First of all, from the definition of Σ_Δ in Eq. (3.19) we have that

$$\Sigma_{\Delta(n_f+1)} = \frac{n_d(n_f+1)}{n_u(n_f+1)} \Sigma_{u(n_f)} - \Sigma_{d(n_f)} + r_{23}(n_f) h_+. \quad (3.35)$$

with

$$r_{23}(n_f) = \begin{cases} \frac{n_d(n_f+1)}{n_u(n_f+1)} & \text{if } h \text{ is up-like} \\ -1 & \text{if } h \text{ is down-like} \end{cases}. \quad (3.36)$$

From Eq. (3.19) it's easy to obtain Σ_u and Σ_d as a function of Σ and Σ_Δ , i.e.

$$\Sigma_u = \frac{n_u}{n_f} (\Sigma + \Sigma_\Delta), \quad (3.37)$$

$$\Sigma_d = \frac{n_d}{n_f} \Sigma - \frac{n_u}{n_f} \Sigma_\Delta. \quad (3.38)$$

Inserting Eqs. (3.37-3.38) into Eq. (3.35) we obtain

$$\Sigma_{\Delta(n_f+1)} = r_{21}(n_f) \Sigma_{(n_f)} + r_{22}(n_f) \Sigma_{\Delta(n_f)} + r_{23}(n_f) h_+. \quad (3.39)$$

with

$$r_{21}(n_f) = \frac{n_d(n_f+1)}{n_u(n_f+1)} n_u(n_f) - n_d(n_f), \quad (3.40)$$

$$r_{22}(n_f) = \frac{n_f+1}{n_u(n_f+1)} \frac{n_u(n_f)}{n_f}. \quad (3.41)$$

The rotation to be applied to V_Δ is identical and reads

$$V_{\Delta(n_f+1)} = r_{21}(n_f) V_{(n_f)} + r_{22}(n_f) V_{\Delta(n_f)} + r_{23}(n_f) h_-. \quad (3.42)$$

In the end, we have to find the rotation for the non-singlet components. Being

$$T_3^u = u^+ - c^+ = \Sigma_{u(n_f)} - h_+, \quad \text{for } n_f = 3, \quad (3.43)$$

$$T_8^d = d^+ + s^+ - 2b^+ = \Sigma_{d(n_f)} - 2h_+, \quad \text{for } n_f = 4, \quad (3.44)$$

$$T_8^u = u^+ + c^+ - 2t^+ = \Sigma_{u(n_f)} - 2h_+, \quad \text{for } n_f = 5, \quad (3.45)$$

and using Eqs. (3.37-3.38) to express Σ_u and Σ_d as a function of Σ and Σ_Δ , we can write T_j , i.e. the new non-singlet combination that is switched on in the $n_f + 1$ scheme, as a function of $\Sigma_{(n_f)}$, $\Sigma_{\Delta(n_f)}$ and h_+

$$T_j = r_{31}(n_f)\Sigma_{(n_f)} + r_{32}(n_f)\Sigma_{\Delta(n_f)} + r_{33}(n_f)h_+, \quad (3.46)$$

with

$$r_{31}(n_f) = \begin{cases} \frac{n_u(n_f)}{n_f}, & \text{if } h \text{ is up-like } (n_f=3,5) \\ \frac{n_d(n_f)}{n_f}, & \text{if } h \text{ is down-like } (n_f=4) \end{cases}, \quad (3.47)$$

$$r_{32}(n_f) = \begin{cases} \frac{n_u(n_f)}{n_f}, & \text{if } h \text{ is up-like } (n_f=3,5) \\ -\frac{n_u(n_f)}{n_f}, & \text{if } h \text{ is down-like } (n_f=4) \end{cases}, \quad (3.48)$$

$$r_{33}(n_f) = \begin{cases} -1, & \text{if } h \text{ is } c \text{ } (n_f=3) \\ -2, & \text{if } h \text{ is } b, t \text{ } (n_f=4,5) \end{cases}. \quad (3.49)$$

For the V_j components the same relations hold and we get

$$V_j = r_{31}(n_f)V_{(n_f)} + r_{32}(n_f)V_{\Delta(n_f)} + r_{33}(n_f)h_-, \quad (3.50)$$

In conclusion, we have found that the basis rotation is performed as

$$\begin{pmatrix} \Sigma_{(n_f+1)} \\ \Sigma_{\Delta(n_f+1)} \\ T_j \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 \\ r_{21}(n_f) & r_{22}(n_f) & r_{23}(n_f) \\ r_{31}(n_f) & r_{32}(n_f) & r_{33}(n_f) \end{pmatrix} \begin{pmatrix} \Sigma_{(n_f)} \\ \Sigma_{\Delta(n_f)} \\ h_+ \end{pmatrix}, \quad (3.51)$$

$$\begin{pmatrix} V_{(n_f+1)} \\ V_{\Delta(n_f+1)} \\ V_j \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 \\ r_{21}(n_f) & r_{22}(n_f) & r_{23}(n_f) \\ r_{31}(n_f) & r_{32}(n_f) & r_{33}(n_f) \end{pmatrix} \begin{pmatrix} V_{(n_f)} \\ V_{\Delta(n_f)} \\ h_- \end{pmatrix}, \quad (3.52)$$

3.2.3 Solving the system

Now we are ready to describe the solution to Eqs. (3.26-3.28) and Eq. (3.30). First of all, we pass into N -space, so that the equations become

$$\mu^2 \frac{d}{d\mu^2} \tilde{T}_{3/8}^{u/d} = -(\gamma^+ + \tilde{\gamma}_{u/d}^+) \tilde{T}_{3/8}^{u/d}, \quad (3.53)$$

$$\mu^2 \frac{d}{d\mu^2} \tilde{V}_{3/8}^{u/d} = -(\gamma^- + \tilde{\gamma}_{u/d}^-) \tilde{V}_{3/8}^{u/d}, \quad (3.54)$$

for the non-singlet distributions, and

$$\mu^2 \frac{d}{d\mu^2} \vec{f}_V = -\gamma_V \cdot \vec{f}_V, \quad (3.55)$$

$$\mu^2 \frac{d}{d\mu^2} \vec{f}_S = -\gamma_S \cdot \vec{f}_S, \quad (3.56)$$

for the valence and singlet sectors. Let us start from the diagonal terms. In this case, rewriting Eqs. (3.53-3.56) in terms of a_s would be counterproductive since we should

solve $a_{em}(a_s) = a_{em}(\mu^2(a_s))$, that is possible (even if only numerically) but not convenient. Therefore, in this case the differential equation is solved in terms of μ^2 and reads

$$\frac{d}{d \log \mu^2} f_i = -\gamma_{\text{ns}}(a_s(\mu^2), a_{em}(\mu^2)) f_i, \quad (3.57)$$

where $f_i = \{\tilde{T}_3^u, \tilde{T}_3^d, \tilde{T}_8^u, \tilde{T}_8^d, \tilde{V}_3^u, \tilde{V}_3^d, \tilde{V}_8^u, \tilde{V}_8^d\}$ and $\gamma_{\text{ns}} = \{\gamma^+ + \tilde{\gamma}_u^+, \gamma^+ + \tilde{\gamma}_d^+, \gamma^- + \tilde{\gamma}_u^-, \gamma^- + \tilde{\gamma}_d^-\}$. The solution of this equation is obtained exponentiating the integral of the non-singlet anomalous dimension. However, due to the presence of two running couplings, solving such integral analytically is impossible. In order to overcome this obstacle, we observe that if we neglect the running of a_{em} , the solution of Eq. (3.57) can be computed exactly as

$$f_i(\mu^2) = E_{\text{ns}}(\mu^2 \leftarrow \mu_0^2) f_i(\mu_0^2), \quad (3.58)$$

with

$$\begin{aligned} E_{\text{ns}}(\mu^2 \leftarrow \mu_0^2) &= \exp \left[- \int_{\log \mu_0^2}^{\log \mu^2} \gamma_{\text{ns}}(a_s(\mu'^2), a_{em}) d \log \mu'^2 \right] \\ &= \exp \left[- \int_{\log \mu_0^2}^{\log \mu^2} \left(a_{em} \gamma_{\text{ns}}^{(0,1)} + a_{em}^2 \gamma_{\text{ns}}^{(0,2)} + a_s(\mu'^2) \gamma_{\text{ns}}^{(1,0)} + a_s(\mu'^2) a_{em} \gamma_{\text{ns}}^{(1,1)} \right. \right. \\ &\quad \left. \left. + a_s^2(\mu'^2) \gamma_{\text{ns}}^{(2,0)} + a_s^3(\mu'^2) \gamma_{\text{ns}}^{(3,0)} \right) d \log \mu'^2 \right] \\ &= \exp \left[- \int_{\log \mu_0^2}^{\log \mu^2} \gamma_{\text{ns}}^{\text{QED}} d \log \mu'^2 + \int_{a_s^0}^{a_s} \frac{a'_s \gamma_{\text{ns}}^{\prime(1,0)} + a_s'^2 \gamma_{\text{ns}}^{(2,0)} + a_s'^3 \gamma_{\text{ns}}^{(3,0)}}{a_s'^2 \beta_{\text{QCD}}^{\prime(2,0)} + a_s'^3 \beta_{\text{QCD}}^{(3,0)} + a_s'^4 \beta_{\text{QCD}}^{(4,0)}} da'_s \right]. \end{aligned} \quad (3.59)$$

where we defined

$$\gamma_{\text{ns}}^{\text{QED}} = a_{em} \gamma_{\text{ns}}^{(0,1)} + a_{em}^2 \gamma_{\text{ns}}^{(0,2)}, \quad (3.60)$$

$$\gamma_{\text{ns}}^{\prime(1,0)} = \gamma_{\text{ns}}^{(1,0)} + a_{em} \gamma_{\text{ns}}^{(1,1)}, \quad (3.61)$$

$$\beta_{\text{QCD}}^{\prime(2,0)} = \beta_{\text{QCD}}^{(2,0)} + a_{em} \beta_{\text{QCD}}^{(2,1)}. \quad (3.62)$$

Therefore, if a_{em} does not run, we have that

$$\begin{aligned} E_{\text{ns}}(\mu^2 \leftarrow \mu_0^2) &= \\ \exp \left[\gamma_{\text{ns}}^{\text{QED}} \log \frac{\mu_0^2}{\mu^2} \right] &\exp \left[\gamma_{\text{ns}}^{\prime(1,0)} j^{(0,2)}(a_s, a_s^0) + \gamma_{\text{ns}}^{(2,0)} j^{(1,2)}(a_s, a_s^0) + \gamma_{\text{ns}}^{(3,0)} j^{(2,2)}(a_s, a_s^0) \right], \end{aligned} \quad (3.63)$$

Observe that the functions $j^{(k,2)}$ are the ones defined in Eqs. (1.116-1.118), with the replacement $\beta_0 \rightarrow \beta_{\text{QCD}}^{(2,0)}$.

Eq. (3.63) has been derived in the case in which a_{em} is fixed. However, even if the running of a_{em} gives just a small correction, it cannot be totally neglected. Let us see how we included the dependence on μ^2 of a_{em} . In order to do it, we split the μ^2 integration range in intervals logarithmically spaced and in each of them we keep a_{em} fixed, so that in each interval we can apply Eq. (3.63). In practice, we are approximating $a_{em}(\mu^2)$ with a step function. Being the running of $a_{em}(\mu^2)$ very moderate with respect to the one of

the strong coupling, the error associated to this approximation is completely negligible. With this approximation, the solution of the non-singlet sector is

$$E_{\text{ns}}(\mu^2 \leftarrow \mu_0^2) = \prod_{k=n}^0 E_{\text{ns}}(\mu_{k+1}^2 \leftarrow \mu_k^2), \quad (3.64)$$

where $\mu_{n+1}^2 = \mu^2$ and the values of the couplings used to compute $E_{\text{ns}}(\mu_{k+1}^2 \leftarrow \mu_k^2)$ from Eq. (3.63) are given by the substitution

$$a_s \rightarrow a_s^{k+1} = a_s(\mu_{k+1}^2), \quad (3.65)$$

$$a_s^0 \rightarrow a_s^k = a_s(\mu_k^2), \quad (3.66)$$

$$a_{em} \rightarrow a_{em}^{k+\frac{1}{2}} = a_{em} \left(\frac{\mu_{k+1}^2 + \mu_k^2}{2} \right). \quad (3.67)$$

Now we can discuss how we solved the two coupled systems of Eqs. (3.55-3.56). The equation we want to solve is

$$\frac{d}{d \log \mu^2} \vec{f} = -\gamma(a_s(\mu^2), a_{em}(\mu^2)) \cdot \vec{f}, \quad (3.68)$$

with $\vec{f} = \{\vec{f}_V, \vec{f}_S\}$ and $\gamma = \{\gamma_V, \gamma_S\}$. The solution of Eq. (3.68) is given by

$$\vec{f}(\mu^2) = \mathbf{E}(\mu^2 \leftarrow \mu_0^2) \cdot \vec{f}(\mu_0^2), \quad (3.69)$$

with

$$\mathbf{E}(\mu^2 \leftarrow \mu_0^2) = \mathcal{P} \exp \left[- \int_{\log \mu_0^2}^{\log \mu^2} \gamma(a_s(\mu'^2), a_{em}(\mu'^2)) d \log \mu'^2 \right], \quad (3.70)$$

In this case, due to the presence of two running couplings, constructing a fully analytical expanded solution is impossible, see App. A for more details. A numerical approach can be used to construct a truncated solution, beginning with the exact path-ordered solution given in Eqs. (3.69-3.70) and then performing a numerical expansion. However, this method requires approximating higher-order derivatives using finite differences [110], which can result in numerical instabilities. For these reasons, we opted for the exact solution described in Sec. 1.6. Also in this case the μ^2 interval has been divided into sub-intervals that are logarithmically spaced so that using a discretized path-ordering as we did in Sec. 1.6, we obtain

$$\mathbf{E}(\mu^2 \leftarrow \mu_0^2) = \prod_{k=n}^0 \mathbf{E}(\mu_{k+1}^2 \leftarrow \mu_k^2), \quad (3.71)$$

where $\mu_{n+1}^2 = \mu^2$ and

$$\mathbf{E}(\mu_{k+1}^2 \leftarrow \mu_k^2) = \exp \left[-\gamma \left(a_s^{k+\frac{1}{2}}, a_{em}^{k+\frac{1}{2}} \right) \Delta \log \mu_k^2 \right], \quad (3.72)$$

with

$$a_s^{k+\frac{1}{2}} = a_s \left(\frac{\mu_k^2 + \mu_{k+1}^2}{2} \right), \quad (3.73)$$

$$\Delta \log \mu_k^2 = \log \mu_{k+1}^2 - \log \mu_k^2. \quad (3.74)$$

Regarding the solution of the two coupled RGEs for the couplings, Eqs. (3.15-3.16), due to the presence of two running couplings, it is not possible to find an expanded solution like it happened in the pure QCD case. For this reason, the only possible solution is the numerical one also for solving the running of the couplings.

3.3 Benchmarking the solution

3.3.1 EKO vs APFEL

The solution described in Sec. 3.2 has been implemented in the public code `EKO` [149] and it has been benchmarked with the QED evolution implemented in the `APFEL` code [110]. The latter, is the code used in the `NNPDF3.1QED` PDFs determination. It uses a different approach to solve DGLAP equations: it solves Eq. (3.11) directly in the x space and therefore it doesn't need to pass into Mellin space and to compute the inverse Mellin transform. The comparisons between the two solutions has been performed by taking a starting PDFs set, that in our case was `NNPDF3.1QED`, that has been evolved from $\mu_0 = 1.65$ GeV up to $\mu_0 = 100$ GeV with the two codes. Observe that the choice of the starting PDFs set is influential for the sake of the comparison. Indeed, we want to assess that the two codes evolve a given initial condition to the same result, no matter what the initial condition is. Fig. 3.5 shows the comparison in percentage distance between the result of `APFEL` QED exact solution and three different curves. First of all there is the comparison with `APFEL` truncated solution (green dot-dashed line). Then, there is the comparison with `EKO` using the same settings for the a_{em} running of `APFEL` (red dashed line): indeed, in `APFEL` the running of the electromagnetic coupling is pure LO since it neglects both the pure QED NLO term $\beta_{\text{QED}}^{((0,3))}$ and the first QCD correction of the QED RGE, i.e. $\beta_{\text{QED}}^{((1,2))}$. Moreover, in `APFEL` also the first QED correction to the QCD RGE, $\beta_{\text{QCD}}^{((2,1))}$, is not considered. Therefore, in this case the two RGE can be solved independently. In the end, there is the comparison with `EKO` in which these terms have been included (blue line), using the setting that are referred as `NNPDF4.0QED` settings. We can observe that the distance between the results produced by the two codes with the same solution method is around $10^{-2}\%$ for all PDFs in a wide range of x value, with the exception of those at very large x , where the PDFs are very small so in that case we are comparing numbers that are very close to zero. When comparing exact and truncated solution the differences increase but remain at the permille level, with the exception of very small x values, in which it reaches the percent level. Also in the case in which `EKO` and `APFEL` use their own settings for the running of a_{em} , the distances are still very small, with the exception of the photon PDF. This is a consequence of the fact that the quarks and gluon PDFs are not very sensible to the NLO corrections in the running of a_{em} while for the photon it is a non negligible effect. Moreover, it has been observed that the quarks and gluon PDFs are much more sensible to the inclusion or not of the QED correction to the QCD beta function, than to the running of a_{em} . This was expected since the running of those PDFs is not very affected by QED effects [150], but they are sensible to the running of the strong coupling. This benchmark validates the correctness of the implementation of $\text{QCD} \otimes \text{QED}$ evolution in `EKO` since it yields results that are identical, up to numerical precision and methodological differences, to the ones given by `APFEL` that uses a completely independent implementation based on a different solution strategy. Observe that between μ_0 and μ there is the bottom threshold. So this check also assures that the implementation of the matching conditions applied to our evolution base is correct.

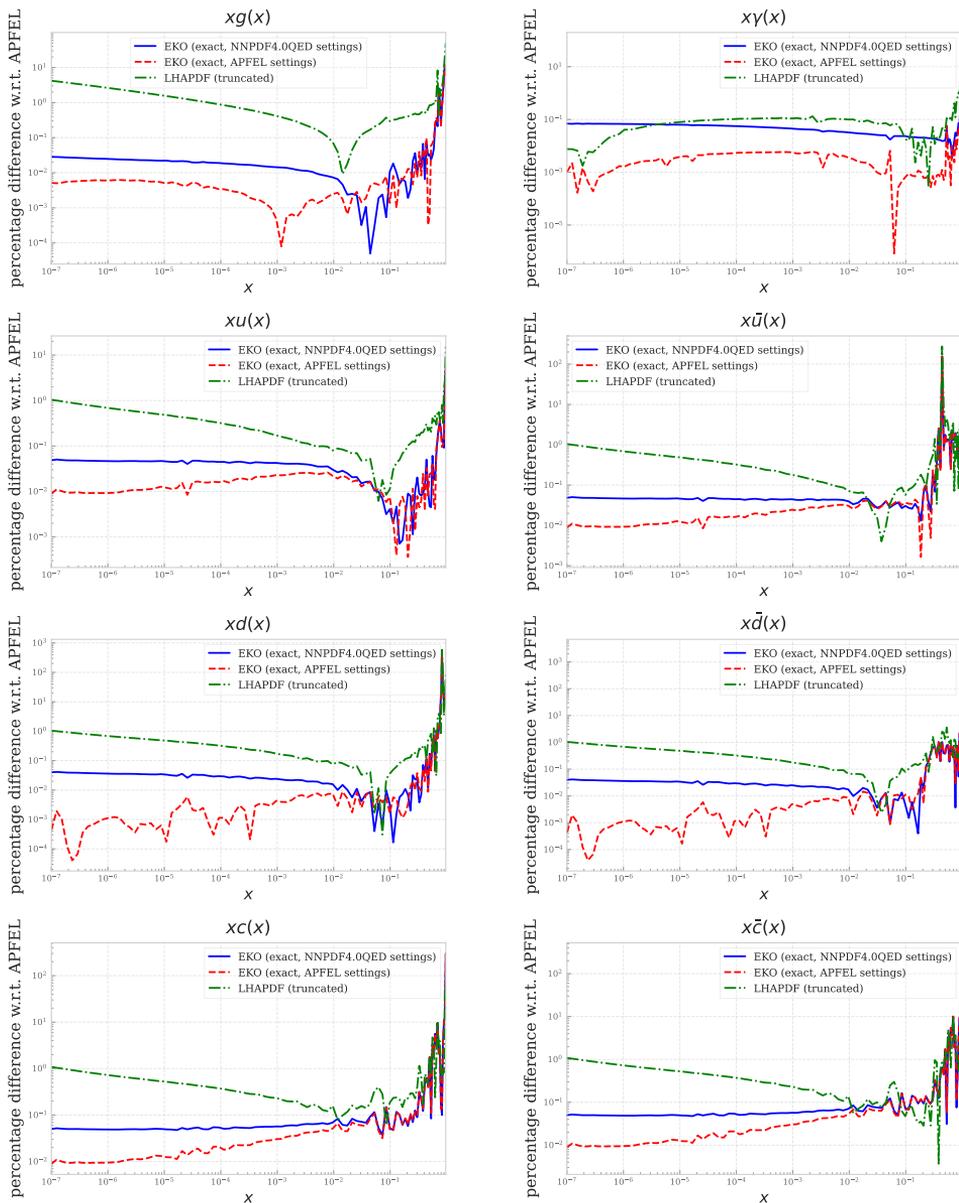


Figure 3.5: Percentage difference between pairs of PDFs at $Q = 100$ GeV, resulting from evolving the NNPDF3.1QED PDFs from $Q_0 = 1.65$ GeV using various QED \otimes QCD evolution implementations. The comparisons are shown from top to bottom for the gluon, up, down, and charm PDFs (left), and for the photon, anti-up, anti-down, and anti-charm PDFs (right). The three curves compare: APFEL exact vs. truncated evolution (green, dot-dashed); APFEL vs. EKO exact, using default settings for coupling evolution in each case (blue, solid); and APFEL vs. EKO exact, both with APFEL settings for coupling evolution (red, dashed). Observe that the y -axis uses a logarithmic scale, and the y -axis range for the gluon plot differs from that of the other PDFs.

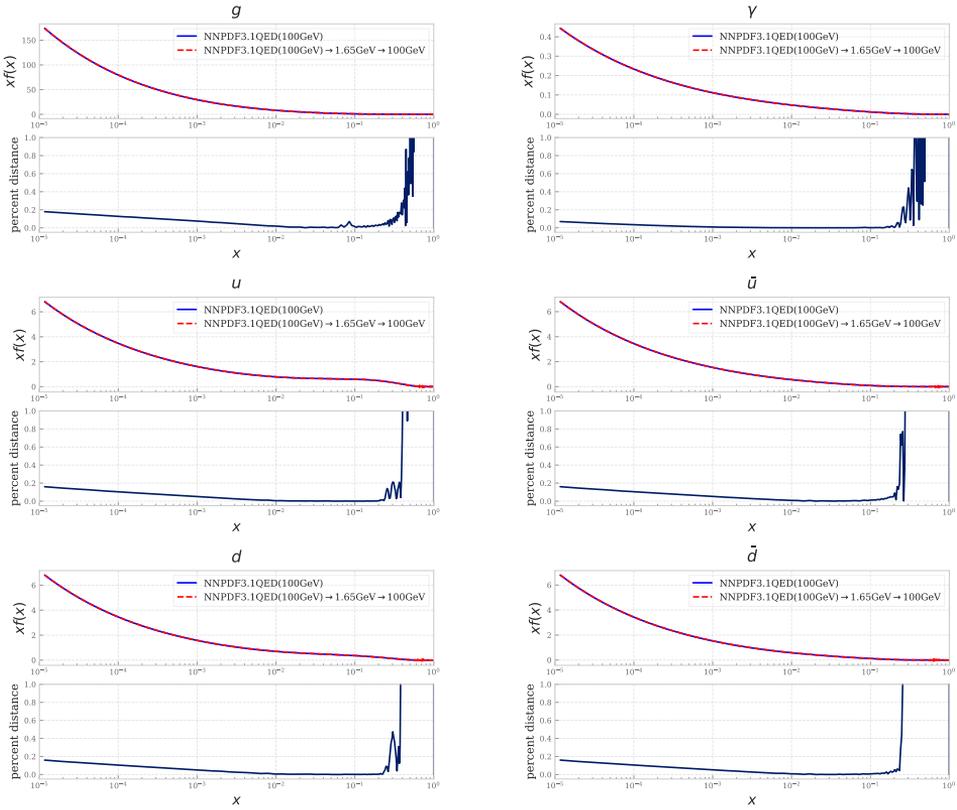


Figure 3.6: Comparison of the NNPDF3.1QED PDFs at $Q_a = 100$ GeV with the results obtained by evolving these PDFs from Q_a to $Q_b = 1.65$ GeV and then back to Q_a . The figures display, from top to bottom, the gluon, up and down quarks (left), and the photon, anti-up, and anti-down quarks (right). Each set includes both the pairs of PDFs and their percentage relative difference.

3.3.2 Invertibility of exact solution

As a further benchmark of the implementation of the $\text{QCD} \otimes \text{QED}$ evolution in EKO we checked the invertibility of the exact solution. Indeed, the evolution operator \mathbf{E} must satisfy

$$\mathbf{E}(\mu^2 \leftarrow \mu_0^2) \mathbf{E}(\mu_0^2 \leftarrow \mu^2) = \mathbf{1}. \quad (3.75)$$

This is not completely true for the truncated solution since in that case Eq. (3.75) is true only up to higher orders (with respect to the ones used in the splitting functions). Therefore, deviations from Eq. (3.75) decrease as we go up with the perturbative orders of the splitting functions. This check has been performed by taking a PDFs set, that was again NNPDF3.1QED even if also this check is not dependent on the choice of the initial condition, that has been evolved from $\mu = 100$ GeV to $\mu_0 = 1.65$ GeV and then back again to μ . Observe that also in this case the threshold of the bottom is crossed. Since this time it is crossed also in the backward evolution, with this check we are benchmarking the implementation of the inverse matching. Indeed, it can be done exactly only in Mellin space since it is the inverse of a matrix, while in x space it can be done only in an expanded way, being it the kernel of a convolution. Fig. 3.6 shows the results of the com-

parison. We can observe that the difference between the starting and the final PDFs is at the level of few permille, with the exception of the large x region where also in this case the PDFs are very small. Hence, we conclude that the $\text{QCD}\otimes\text{QED}$ backward evolution implemented in EKO has a permille accuracy.

3.3.3 QED evolution in the theory predictions

The last check that we performed for testing the $\text{QCD}\otimes\text{QED}$ evolution implemented in EKO , has been the comparisons of the theory predictions for the different datasets used in the fit with and without QED corrections in the evolution. In these predictions the QED effects entered only in the evolutions of the PDFs. It means that we didn't consider PI processes. The reason was that at this stage we only wanted to test the QED effects in the evolution factors. Moreover, as we already mentioned, in this analysis we didn't include PI processes in the PDFs determination. The predictions have been obtained using Eqs. (2.1-2.2) by convoluting the partonic processes, the evolution factor and the PDFs at the fitting scale. The numerical results have been computed through FKTables using Eqs. (2.8-2.9). In this test the PDFs at the fitting scale were NNPDF3.1QED (however also in this case the choice of the initial condition was influential since we are benchmarking the solution method that does not depend on the initial condition).

Fig. 3.7 shows the percent distances between the theory predictions of some representative datasets, among the ones used in the fit and listed in Tabs.2.1-2.5, with and without QED effects in the evolution of the PDFs. The first (red plus) is the distance between pure QCD evolution using the truncated solution (that is the one used in the NNPDF4.0 analysis) and the mixed $\text{QCD}\otimes\text{QED}$ evolution (that therefore is using the exact solution). The second (blue cross) is the same comparison but in the pure QCD evolution it is used the exact solution. What we can observe is that the distance between pure QCD evolution and $\text{QCD}\otimes\text{QED}$ evolution is at the level of permille for most of the data points, with some exceptions in which the difference becomes of the order of percent. Moreover, we can see that, as we expected, the $\text{QCD}\otimes\text{QED}$ exact solution provides results that are closer to the pure QCD exact solution than to the truncated one. The other datasets that are used in the fit but not in this plot show a similar behavior. In the end we conclude that the QED effects give a small correction to the PDFs evolution, and as a consequence it gives a small correction to the hadronic observables.

3.4 Results of the QED fit

Now we are ready to show the results of the NNPDF4.0QED fit. It has been obtained implementing the LuxQED methodology, described in Sec. 3.1, into the NNPDF4.0 fitting framework, described in Chap. 2. All the settings of the fit (the hyperparameters of the neural network, the theoretical constants etc.) have been chosen identical to the ones used in NNPDF4.0. As we already discussed, the only difference is the method used to solve the DGLAP equations, that is the exact one for the QED fit and the truncated one for NNPDF4.0. As we did in the NNPDF4.0 analysis we performed both the NLO fit and the NNLO one, where NLO and NNLO are referred to the strong coupling perturbative order, while for the QED coupling we always included all the splitting kernels that are currently known, i.e. the ones described in Sec. 3.2.

Tab. 3.1 compares the statistical estimators of the NLO and NNLO QED fits with the corresponding QCD fits, i.e. NNPDF4.0 at NLO and NNLO respectively. We can see that all values, and in particular the χ^2 are almost identical in the two cases. Some differences are visible at NLO but they are related to the differences between exact and truncated

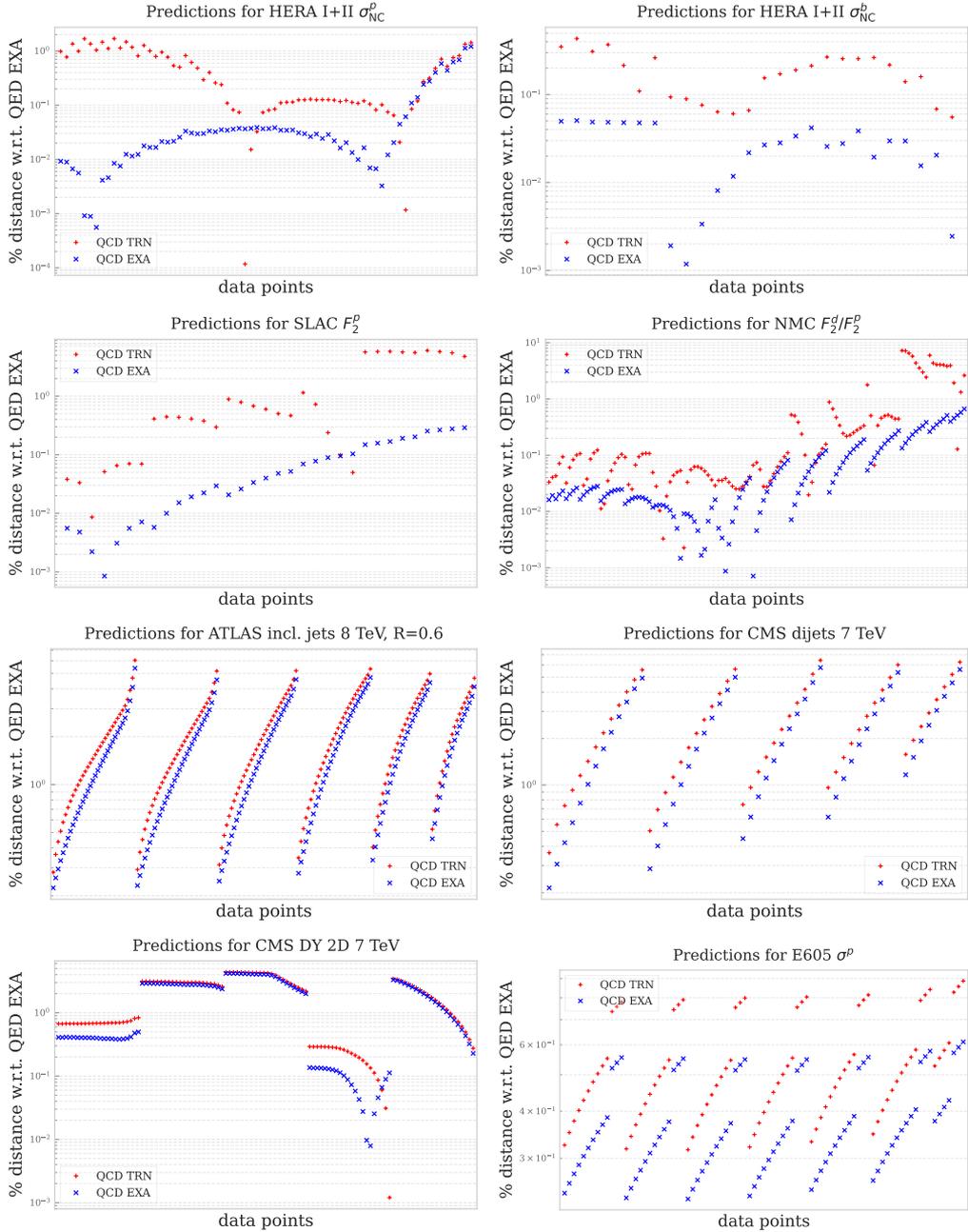


Figure 3.7: Comparison between the theory predictions obtained using QCD⊗QED DGLAP evolution and pure QCD evolution. For the latter, both the exact (blue crosses) and the truncated (red pluses) methods are displayed. The results are shown as a percent difference with respect to the QCD⊗QED evolution. All the data points for some representative data sets that are used in the fit are displayed in the plots. The datasets are among the ones listed in Tabs.2.1-2.5.

	NNPDF4.0 NLO		NNPDF4.0 NNLO	
	QCD×QED	QCD	QCD×QED	QCD
χ^2	1.31	1.26	1.17	1.17
$\langle E_{\text{tr}} \rangle_{\text{rep}}$	2.47±0.07	2.41±0.06	2.27±0.06	2.28±0.05
$\langle E_{\text{val}} \rangle_{\text{rep}}$	2.66±0.11	2.57±0.10	2.39±0.10	2.37±0.11
$\langle \chi^2 \rangle_{\text{rep}}$	1.337±0.016	1.286±0.017	1.192±0.014	1.195±0.015

Table 3.1: Statistical estimators for NNPDF4.0QED at NLO and NNLO are compared to those for NNPDF4.0 in pure QCD. The table, from top to bottom, includes: the total χ^2 , the average training and validation figures of merit $\langle E_{\text{tr}} \rangle_{\text{rep}}$ and $\langle E_{\text{val}} \rangle_{\text{rep}}$ across replicas, the average χ^2 over replicas $\langle \chi^2 \rangle_{\text{rep}}$.

solutions for the DGLAP equations. A discussion on the differences introduced by the change of solution of DGLAP equation is presented in App. B. Therefore, we conclude that the inclusion of QED effects does not affect the quality of the fit (as we expected, being such effects small).

Figs. 3.8-3.9 show the comparisons between NNPDF4.0 and NNPDF4.0QED for the quarks and gluon PDFs both at fitting scale and at 100 GeV. The comparison are shown both for NNPDF4.0 (that uses the truncated solution for DGLAP equations) and for a variant of that determination that uses the exact solution, as the NNPDF4.0QED fit does. We can observe that at fitting scale the QED PDFs are compatible within the uncertainty band with the pure QCD ones. Indeed, since the photon affects the fit only through DGLAP evolution and changing the momentum sum rules, we expect the quark and gluon PDFs to be almost unchanged, being the photon PDF very small. Instead, at 100 GeV the differences between NNPDF4.0 and NNPDF4.0QED increase due to evolution effects that are moderate but non negligible. The PDF that is most affected by QED evolution is the gluon in which the differences can reach up to 1σ level. That difference is also related to the different type of solution used in the two fits. Indeed, in the comparison between NNPDF4.0QED and the version of NNPDF4.0 in which the exact solution is used, the difference between the two gluons is reduced, especially at small x .

Regarding the photon PDF, Fig. 3.10 shows the comparison between NNPDF4.0QED and the most recent QED fits, i.e. NNPDF3.1QED [130], MSHT20QED [132] and CT18QED [133]. We can observe that the photons from the different PDFs sets agree at the percent level. This is a consequence of the fact that, even though these sets exhibits large differences in the quarks and gluon PDFs, the photon PDF is determined with the LuxQED formula in all of these QED fits, that in turn depends on the proton structure functions that are well described by all of these PDFs sets, since they are constrained by experimental data both at large and at low scales. Regarding the comparison with NNPDF3.1QED, the two results agree beyond the percent level. The main differences between the photons of the two analyses comes from the differences between the two types of solutions used in the two fits, i.e. truncated and exact. It means that the differences between the two photons are due to higher order corrections. The residual difference in the two photons are due to the change in the quark and gluon PDFs of the two sets. The comparison with MSHT20QED and CT18QED shows larger differences. This is mainly due to the fact that the methodologies are slightly different between the different sets (for example in MSHT20QED the photon is generated at fitting scale) and that the quark and gluon PDFs used to compute the LuxQED formula show some differences between the different sets. The uncertainties of the photons from the different PDFs sets are also compatible, since

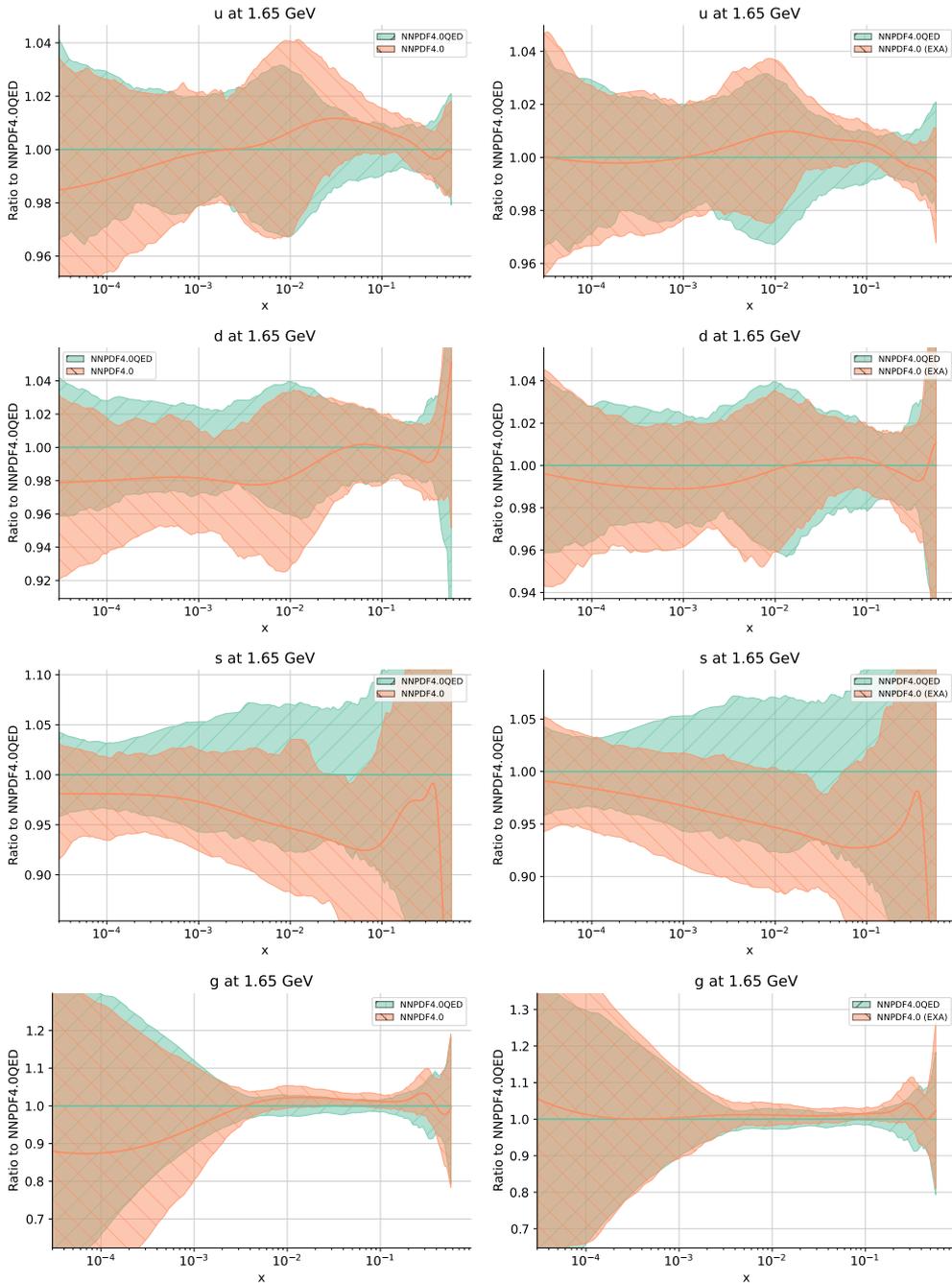


Figure 3.8: Comparison between NNPDF4.0 and NNPDF4.0QED (left) and comparison between a variation of NNPDF4.0 that uses the exact solution and NNPDF4.0QED (right). The PDFs for the gluon and for the quarks up, down and strange are shown as a ratio to NNPDF4.0QED. Results are displayed at fitting scale.

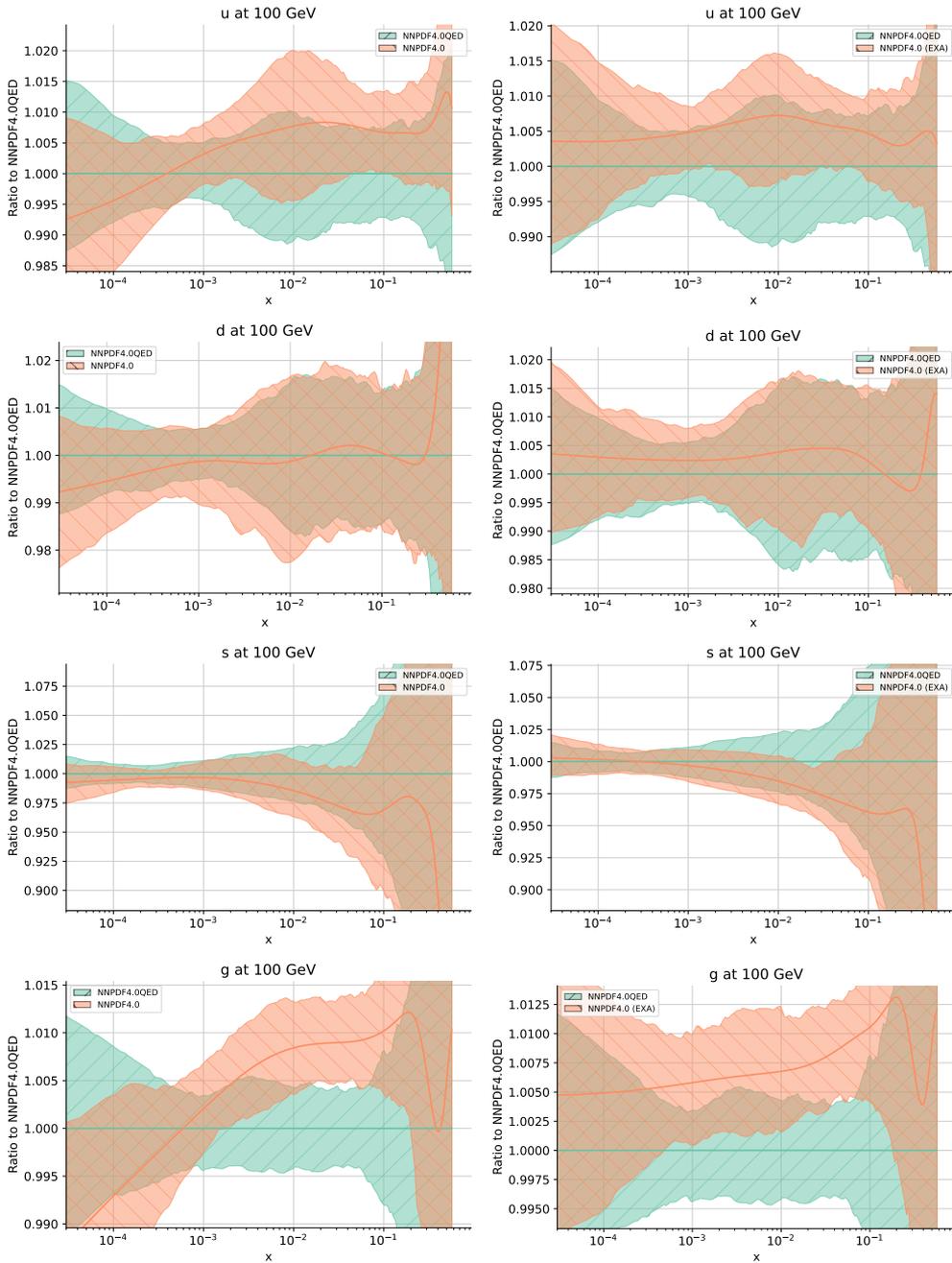


Figure 3.9: Comparison between NNPDF4.0 and NNPDF4.0QED (left) and comparison between a variation of NNPDF4.0 that uses the exact solution and NNPDF4.0QED (right). The PDFs for the gluon and for the quarks up, down and strange are shown as a ratio to NNPDF4.0QED. Results are displayed at 100 GeV.

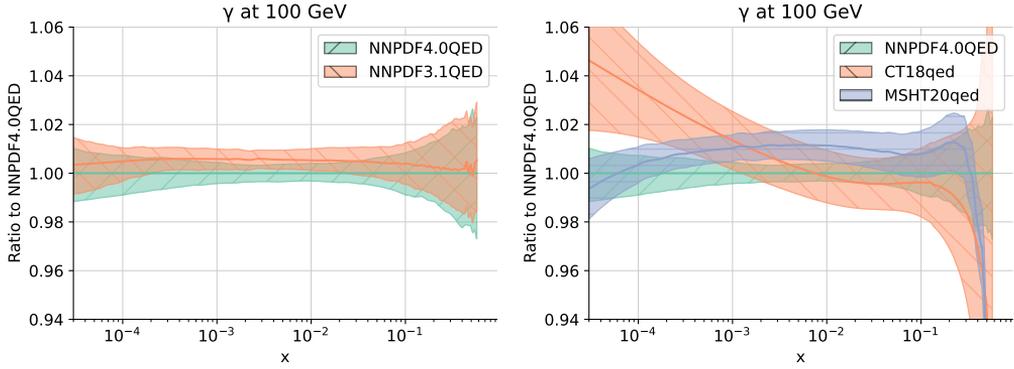


Figure 3.10: Left: Comparison between the photons of the NNPDF4.0QED and NNPDF3.1QED determinations as a ratio plot to the central value of NNPDF4.0QED at the scale $Q_\gamma = 100$ GeV. Right: same as left but comparing with MSHT20QED and CT18QED.

they are completely dominated by the theoretical uncertainties of the LuxQED approach, which have been discussed in Sec. 3.1. In our QED fit, the uncertainty on the photon are estimated in the way presented in Sec. 3.1, that follows Ref. [127]. Also the MSHT20QED determination follows this reference, with extra higher-twist corrections coming from the the choice of the low scale for Q_γ . Indeed, it shows very similar uncertainties to the ones of NNPDF4.0QED. CT18QED instead, adopts a different determination of the elastic contribution, resulting in larger uncertainties.

In Fig. 3.11, the central values of the photon PDFs from various sets are compared at different scales. Uncertainties are omitted to keep the plot clear. It's important to observe that differences in the different photons at the various scales shown in the figure are also due to different methods used to solve the QCD \otimes QED evolution equations. Specifically, the NNPDF3.1QED set uses a numerical implementation of the truncated solution (refer to App. B). Despite these differences in the way DGLAP equations are solved, these differences are primarily caused by the interaction with the quark and gluon PDFs. The increase in differences at lower scales supports this assumption. Despite this, all photon PDFs show an agreement within approximately 3% for all $x \gtrsim 10^{-3}$, even at the lowest scale of $Q = 1.65$ GeV. In Fig. 3.11, we also present the central photon PDF obtained of a variant of the the NNPDF4.0QED determination, obtained with the photon PDF computed at a scale of $Q_\gamma = 10$ GeV, instead of the default $Q_\gamma = 100$ GeV. When a lower Q_γ is chosen, the upper limit of integration in Q^2 of Eq. (3.1) is correspondingly reduced. As explained in Sec. 3.1, this leads to a significant contribution to the LuxQED formula from the low-scale region, where the structure function is derived from a fit to the data, making $\mathcal{O}(m_p/Q)$ corrections to the LuxQED formula potentially relevant. The observed shift in the central photon PDF when using this lower scale, compared to the default, is at most of the order of the uncertainty of the photon PDF. This indicates that the uncertainty in the LuxQED procedure is accurately estimated.

As we already discussed, the photon's impact in the fit is in the DGLAP evolution and in the modification of the momentum sum rule, i.e. Eq. (3.7). We asses the size of the latter by computing the momentum fraction carried by a given PDF f_i as

$$M[f_i(Q^2)] = \int_0^1 dz z f_i(z, Q^2). \quad (3.76)$$

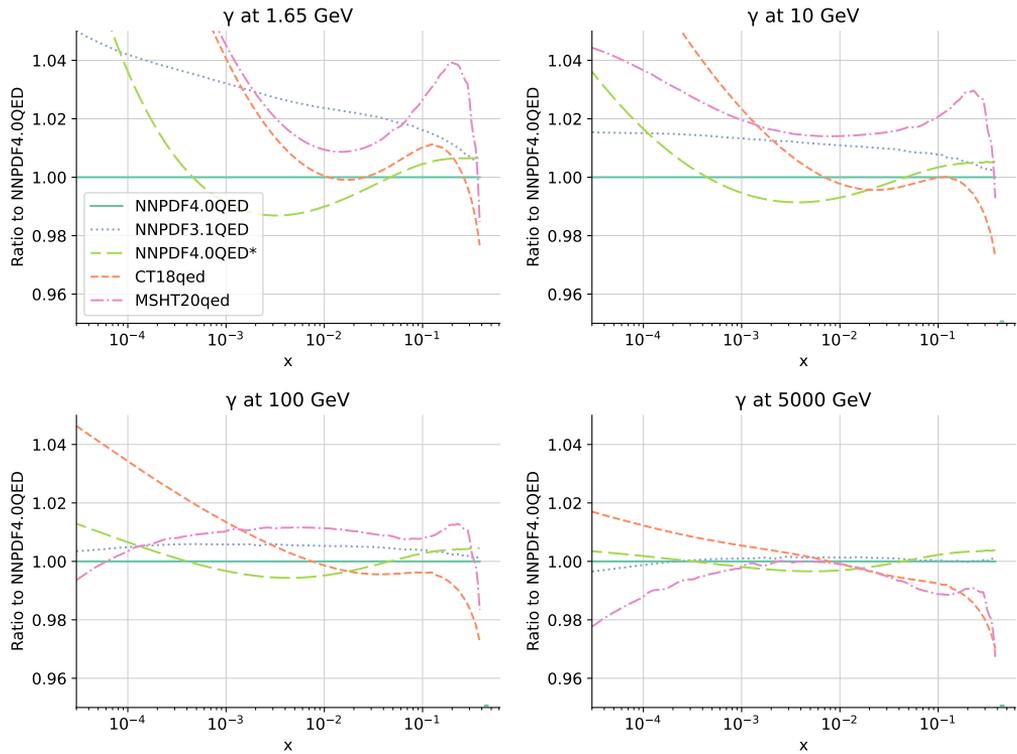


Figure 3.11: Variation with scale of the central photon PDF in the NNPDF4.0QED, NNPDF3.1QED, MSHT20QED and CT18QED PDF sets, for the scales $Q = 1.65, 10, 100, 5000$ GeV. It is also shown a variation of the NNPDF4.0QED determination in which the photon PDF is computed at $Q_\gamma = 10$ GeV (marked as NNPDF4.0QED*). The plots are shown as ratio plots with respect to NNPDF4.0QED.

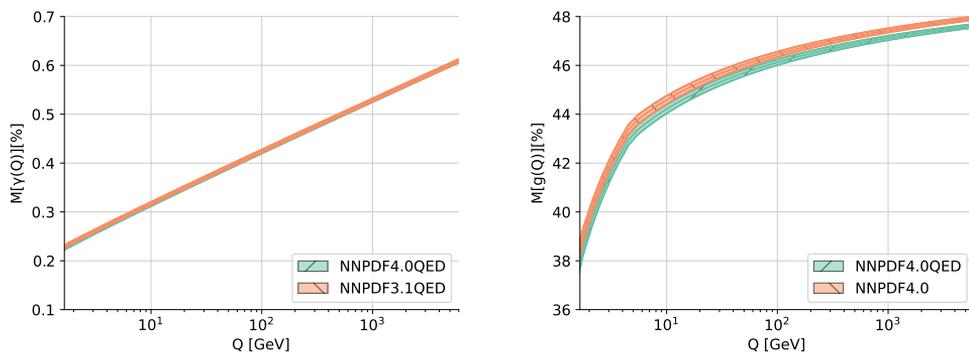


Figure 3.12: Left: The percentage of momentum fraction carried by the photon PDF $\gamma(x, Q^2)$ in NNPDF4.0QED and NNPDF3.1QED as a function of the scale Q , with bands representing 1σ uncertainties. Right: The same for the momentum fraction carried by the gluon PDF in NNPDF4.0QED and NNPDF4.0 (pure QCD).

In Fig. 3.12, the momentum fractions carried by the photon and gluon PDFs in the NNPDF4.0QED set are displayed as a function of scale, shown in percentages. The photon momentum fraction is compared with that of the NNPDF3.1QED set, while the gluon momentum fraction is compared with that of the NNPDF4.0 (pure QCD) set. The photon momentum fraction in both the NNPDF3.1QED and NNPDF4.0QED sets is nearly identical: the photon accounts for approximately 0.2% of the proton's momentum at a low scale (around $Q \sim 1$ GeV), and this fraction increases logarithmically with Q , reaching about 0.6% at the multi-TeV scale. The inclusion of the photon reduces the momentum fraction carried by the gluon by a similar amount, as it can be observed in the right part of the figure.

3.5 Impact on LHC phenomenology

In this section we will study the phenomenological impact of the NNPDF4.0QED PDFs set. Initially, we will compare parton luminosities with those calculated using other QED PDF sets. Next, we will evaluate the influence of QED corrections on specific processes by comparing calculations that incorporate PI contributions to those that only consider the pure QCD ones, and by directly comparing the results obtained from NNPDF4.0QED and NNPDF4.0 (pure QCD) PDFs.

3.5.1 Luminosities

The impact of including a photon PDF on parton luminosities, defined as

$$\mathcal{L}_{ij}(m_X, \sqrt{s}) = \frac{1}{s} \int_{\tau}^1 \frac{dz}{z} f_i(z, m_X^2) f_j\left(\frac{\tau}{z}, m_X^2\right), \quad (3.77)$$

with $\tau = m_X^2/s$ and \sqrt{s} is the center of mass energy, is both direct, due to PI partonic channels, and indirect, as the presence of the photon affects other PDFs, primarily through the reduction of the gluon PDF to maintain the momentum sum rule, as described in Sec. 3.1. Fig. 3.13 compares the PI channels for NNPDF4.0QED NNLO with those from NNPDF3.1QED (top) and MSHT20QED and CT18QED (bottom). There is a high level of agreement, which aligns with the similarities observed between the respective photon PDFs in Fig. 3.10.

Fig. 3.14 displays the luminosities for all other parton channels, comparing the set NNPDF4.0QED with both the previous QED set, NNPDF3.1QED, and the pure QCD NNPDF4.0. As discussed in this chapter, the inclusion of the photon has a moderate impact on the other PDFs. Thus, the comparison between the two QED sets resembles the comparison between the pure QCD NNPDF4.0 and NNPDF3.1, shown in Fig. 9.1 of Ref. [4], where it was demonstrated that NNPDF4.0 is backward compatible with NNPDF3.1, agreeing within uncertainties. The QED corrections mainly affect the gluon-gluon channel, causing a suppression of a few percent around the $m_X \sim 100$ GeV region due to momentum transfer to the photon. Fig. 3.15 shows the comparison with MSHT20QED and CT18QED: when comparing to other PDF sets, differences in quark and gluon PDFs, which are more significant than differences in the photon PDF, dominate the comparison, making it similar to the pure QCD luminosity comparison in Figure 9.3 of Ref. [4].

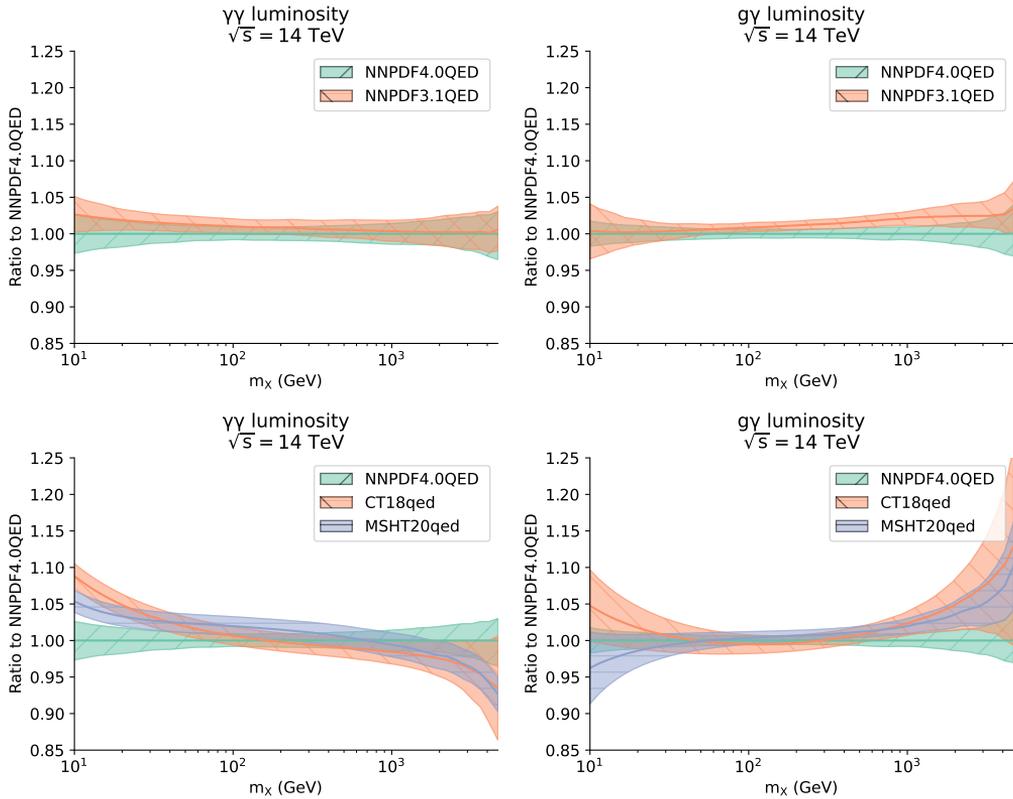


Figure 3.13: The photon-induced contributions to the luminosity at the LHC with a center-of-mass energy of $\sqrt{s} = 14$ TeV are shown as a function of the invariant mass m_X . These are compared for NNPDF4.0QED NNLO against its NNPDF3.1QED version (top), and also against MSHT20QED and CT18QED, with all comparisons presented as ratios relative to NNPDF4.0QED.

3.5.2 Physics processes

We now examine the impact of the photon PDF on several relevant processes for the physics at the LHC: Drell–Yan production (both neutral and charged-current), Higgs production via gluon-gluon fusion, vector boson fusion, and associated production with weak bosons, diboson production, and top-quark pair production, all at the LHC with a center-of-mass energy of $\sqrt{s} = 14$ TeV. Theory predictions for these processes have been calculated using the `PineAPPLE` interface to the automated QCD and EW calculations provided by `mg5_aMC@NLO`. As we already discussed, the `PineAPPLE` program generates interpolation grids, accurate to NLO in both strong and electroweak couplings, that are independent of PDFs, facilitating the variation of input PDF sets with the same grid used for all considered PDFs. Since we aim to compare differences between PDF sets rather than conduct precision phenomenology, we exclude NNLO QCD corrections and the uncertainty band presented here is only due to PDFs uncertainties. For a detailed list of parameters and cuts used in the grid calculations, refer to Sec. 9 of Ref. [4].

For all processes, we present predictions in the figures below using a standardized format. The top panels display the absolute distributions, while the bottom panels show the ratio to the central value obtained using NNPDF4.0QED PDFs, including PI chan-

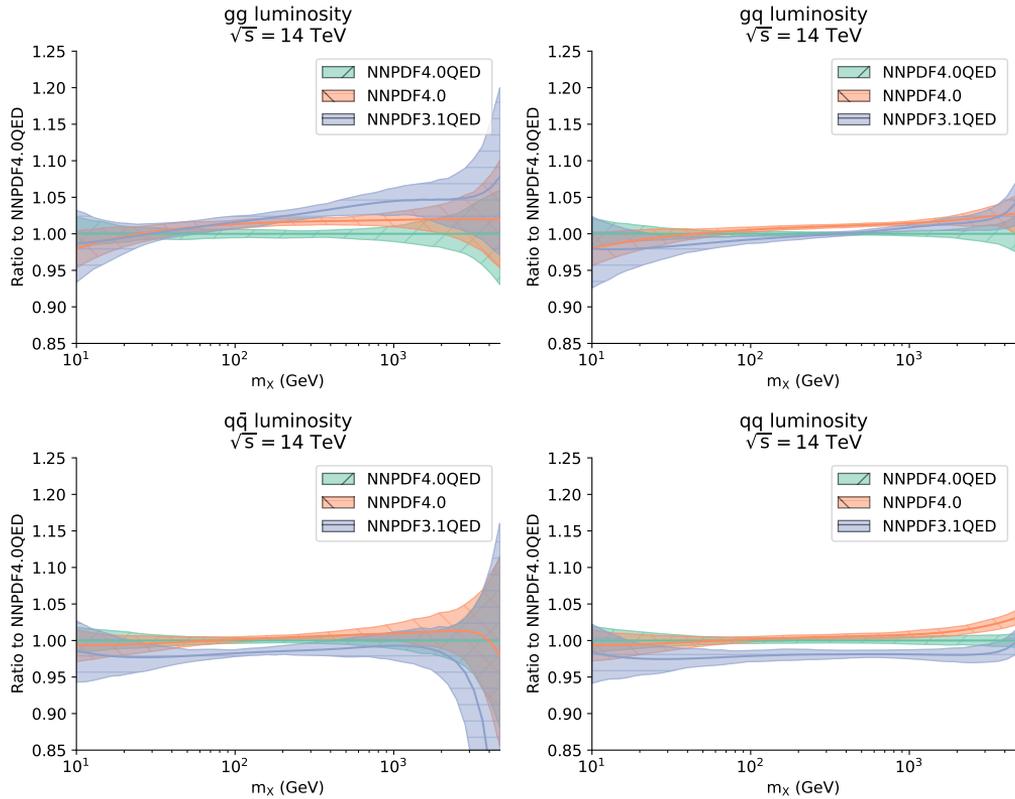


Figure 3.14: Comparison of parton luminosities in the NNPDF4.0QED and the NNPDF4.0 (pure QCD) sets is presented as a ratio to the former for the LHC with $\sqrt{s} = 14$ TeV. From left to right and from top to bottom, The gluon-gluon, gluon-quark, quark-anti-quark, quark-quark luminosities are displayed.

nels, which we refer to as NNPDF4.0QED. The left plots compare NNPDF4.0QED with NNPDF4.0QED without PI channels, NNPDF4.0 pure QCD, and NNPDF3.1QED. The right plots compare NNPDF4.0QED with MSHT20QED and CT18QED. The left plots help assess the overall magnitude of the QED corrections by comparing QED and pure QCD results, i.e. NNPDF4.0 and NNPDF4.0QED, determine the size of the PI contributions by comparing predictions with and without PI channels, and evaluate the impact of changes in the quark and gluon PDFs due to QED effects by comparing NNPDF4.0 with NNPDF4.0QED with PI channels turned off.

In Fig. 3.16, we present the results for inclusive Drell-Yan production in both the neutral-current and charged-current channels. For the neutral current, we provide the invariant mass distribution of the dilepton pair, while for the charged current, we display the rapidity distribution of the lepton. Fig. 3.17 contains the predictions for the Higgs rapidity distribution for gluon fusion, associated production with a W^+ boson, and vector boson fusion. Lastly, Fig. 3.18 illustrates predictions for weak-boson pair production (W^+W^- and W^+Z) as a function of the dilepton transverse momentum, as well as for top-quark pair production based on the invariant mass of the top quark pair.

For charged-current Drell-Yan processes, the QED corrections have virtually no im-

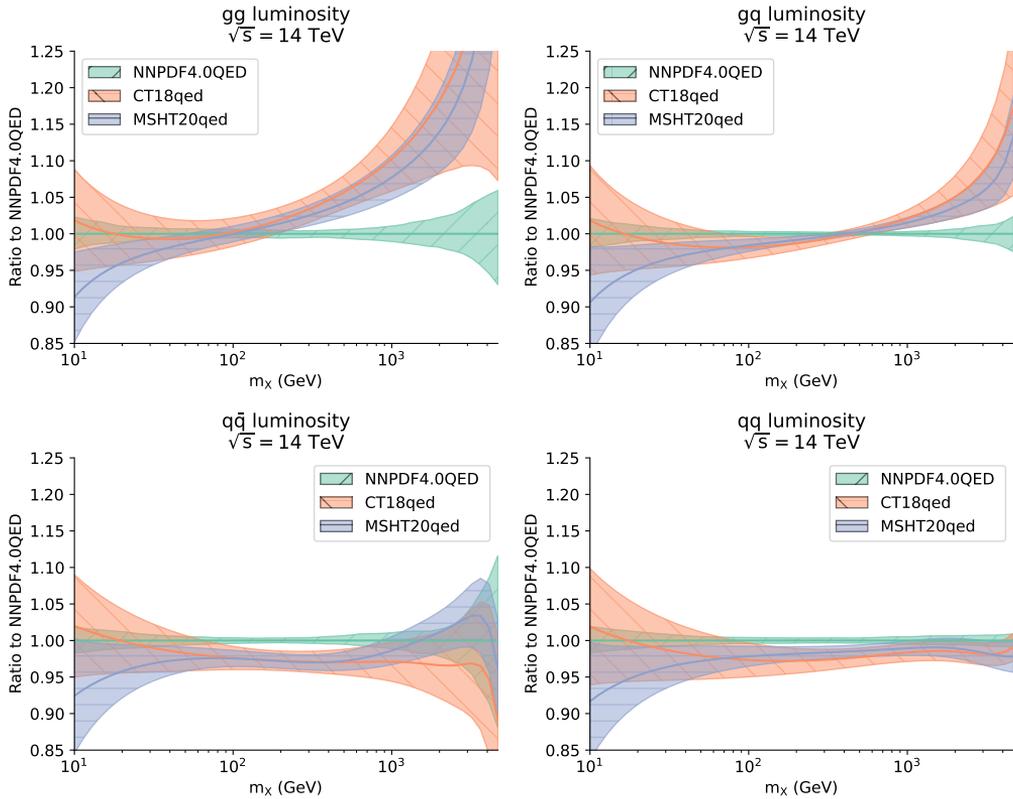


Figure 3.15: As Fig. 3.14 but comparing with MSHT20 [132] and CT18qed [134].

pact. In all other scenarios we consider, the QED correction effects fall into one of two categories. Either the inclusion of QED effects leads to an enhancement of the cross-section, which only occurs when the PI contribution is included, while the NNPDF4.0QED result without the PI contribution remains very close to the pure QCD result. Alternatively, there is a suppression of the cross-section, with the NNPDF4.0QED results being nearly identical whether the PI contribution is included or not. The former situation can be attributed to a significant positive PI contribution, whereas the latter can be explained by the reduction in gluon luminosity due to the transfer of momentum fraction from the gluon to the photon, as illustrated in Fig. 3.14.

The first scenario (where enhancement occurs due to the PI contribution) is seen in neutral-current dilepton production. In this case, the enhancement increases with invariant mass, reaching up to 5% at the TeV scale. It's important to note that the PI contribution to the dilepton final state can also occur through leading-order t - and u -channel diagrams. The absence of QED enhancement at the Z peak suggests that this non-resonant contribution is the main component of the PI contribution. A similar pattern is observed in W^+W^- and ZW^+ production, where the enhancement grows with p_T , reaching 5% for W^+W^- and 2% for ZW^+ at TeV scale for transverse momenta. Additionally, enhancement is seen in associated Higgs production with W^+ and in vector boson fusion. For associated Higgs production, the enhancement is most significant at forward rapidity, peaking at 4%, and decreases to 2% at the highest rapidity of $y_H = 2.5$.

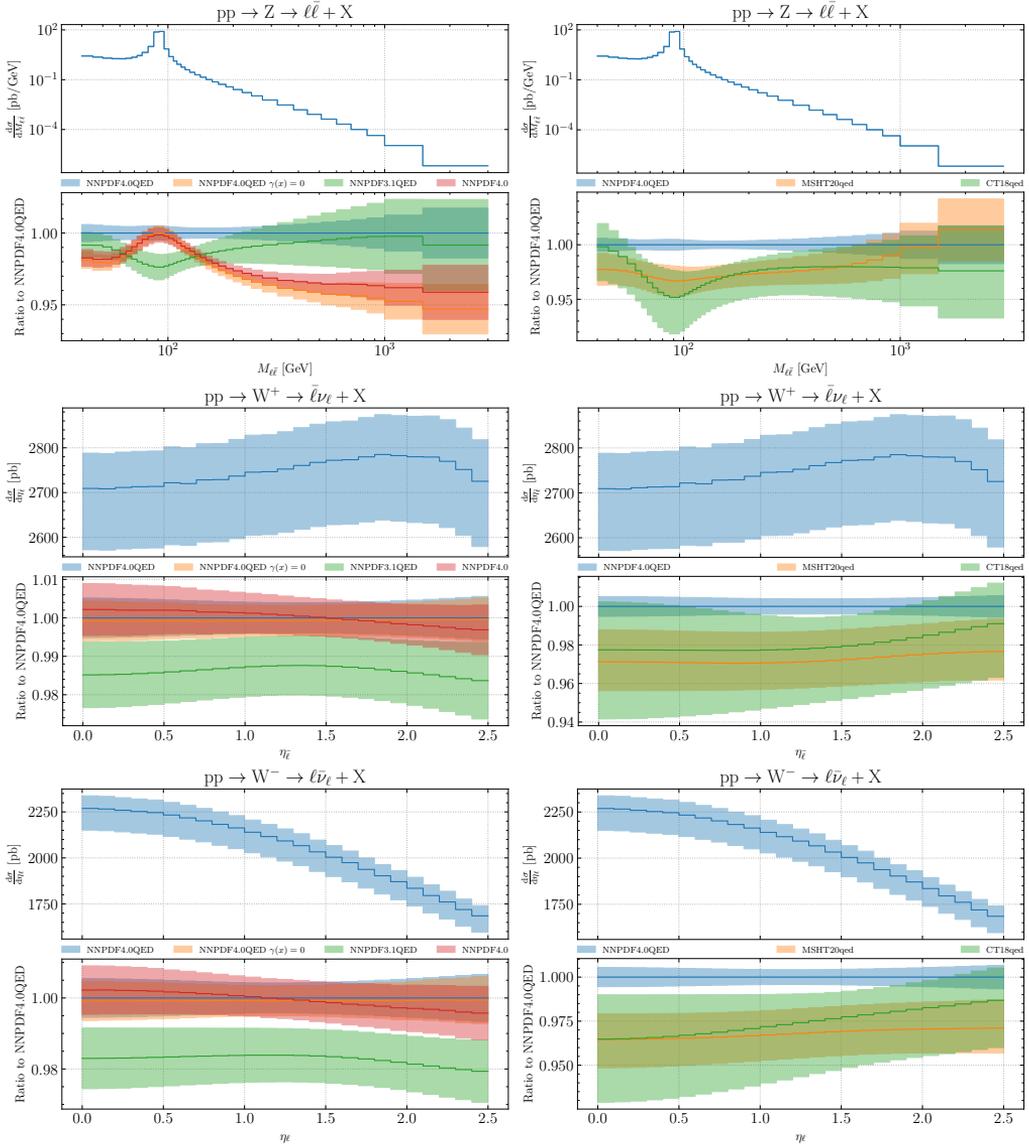


Figure 3.16: Predictions for inclusive Drell–Yan production at the LHC with a center-of-mass energy of $\sqrt{s} = 14$ TeV computed at NLO accuracy in both QCD and electroweak couplings. The uncertainties shown are exclusively PDFs uncertainties. From top to bottom, the figures display: dilepton production mediated by a γ^*/Z boson as a function of the dilepton invariant mass $m_{\ell\bar{\ell}}$; W^+ production as a function of the antilepton pseudo-rapidity $\eta_{\bar{\ell}}$; W^- production as a function of the antilepton pseudo-rapidity $\eta_{\bar{\ell}}$. For each process, the absolute distributions are shown in the top panels, and the ratio to the central value obtained with NNPDF4.0QED PDFs, including PI channels, is shown in the bottom panels. The left panels compare the full NNPDF4.0QED results with NNPDF4.0 (QCD only), NNPDF3.1QED, and NNPDF4.0QED without PI channels. The right panels compare NNPDF4.0QED with MSHT20QED [132] and CT18QED [134].

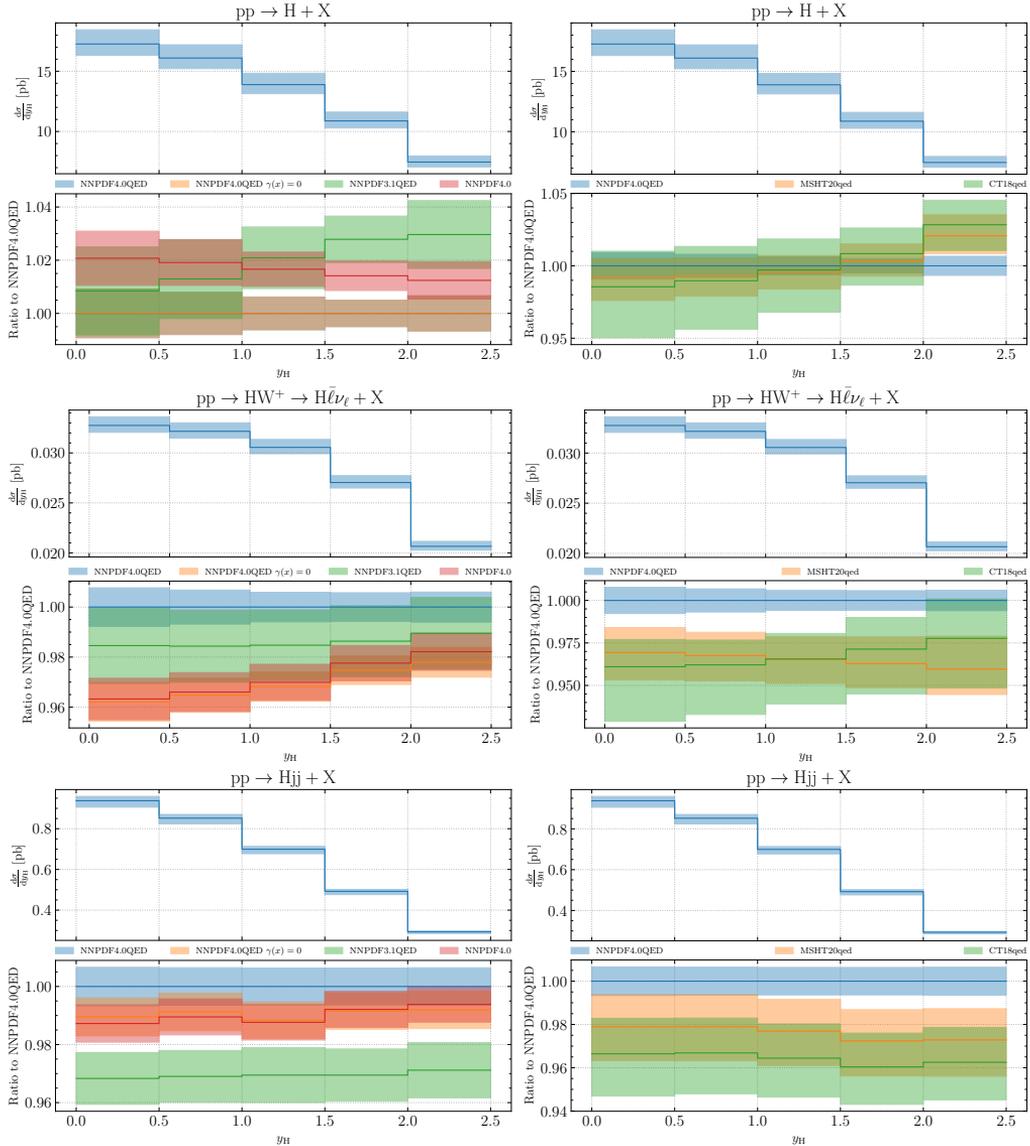


Figure 3.17: Same as Fig. 3.16 but it shows the rapidity distribution of the Higgs produced in gluon-gluon fusion (top panel), in association with a W^+ boson (middle panel), and through vector-boson fusion (bottom panel).

In vector boson fusion, the enhancement is modest, around 1%, and remains nearly constant with a slight decrease as rapidity increases.

The second scenario (suppression that is independent of the PI contribution) occurs in processes driven by gluon fusion. This effect is evident in Higgs production via gluon fusion, where suppression slightly depends on rapidity, ranging from 2% at central rapidity to about 1% at the highest rapidity. A similar but more moderate effect is observed in top pair production, with a suppression of around 1%, which remains largely independent of the invariant mass of the top pair.

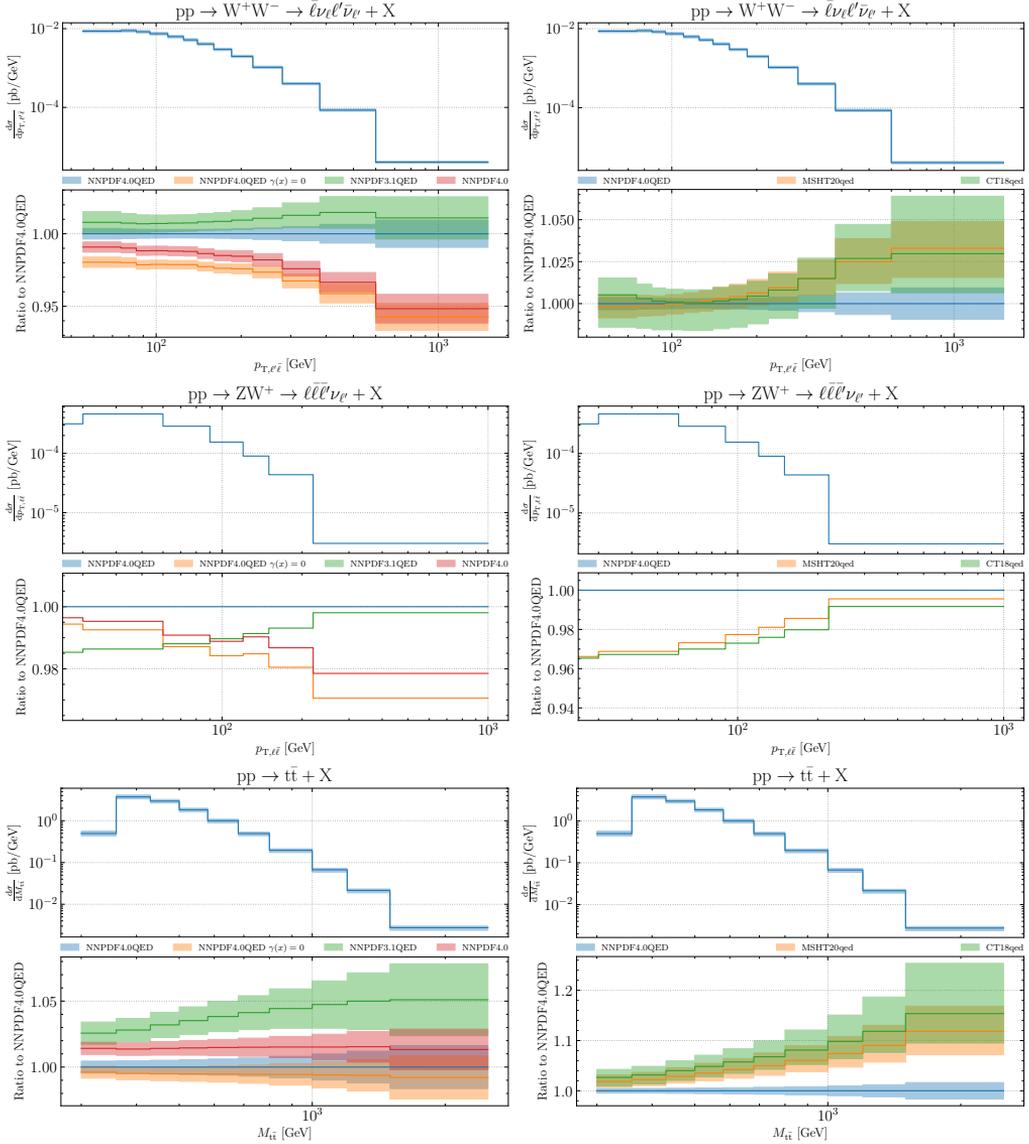


Figure 3.18: Same as Fig. 3.16 but it shows the transverse momentum distribution of dileptons from weak boson pair production (W^+W^- and W^+Z) and the invariant mass distribution for top-quark pair production. No acceptance cuts have been applied to the decay products of the W , Z bosons, or the top quark.

When comparing results obtained using different QED PDFs sets, i.e. NNPDF3.1QED, MSHT20QED, or CT18QED, the observed differences are likely due to variations in the quark and gluon PDFs, considering the similarity of the photon PDFs across these sets. In fact, these comparisons closely resemble those in Section 9.3 of Ref. [4], where the same processes were analyzed under pure QCD conditions.

In conclusion, we determined that incorporating QED corrections is crucial for achieving percent-level precision in phenomenological studies, even when PI contributions are

minimal, such as in Higgs production via gluon fusion. Ignoring the indirect effects of including the photon PDF leads to an overestimation of the peak cross-section (and consequently the total cross-section) by approximately 2%. This introduces a bias in the prediction comparable to the PDF uncertainty, which is not accounted for in the uncertainty estimation.

Combining QED with MHOUs and N³LO

“Avengers...Assemble!”

– Captain America, Avengers: Endgame

In Chap. 3 we described how the QED effects are included in a PDFs fit. In this chapter we will describe how they are combined with the inclusion of MHOUs [7], and with the inclusion of approximate N³LO corrections [8]. In order to do it, we will first describe how these effects are accounted for in a pure QCD PDFs fit, and then how they are combined with the QED effects.

This chapter is organized as follows: in Sec. 4.1 we will describe how MHOUs are included in a PDFs fit and we will briefly present the NNPDF4.0MHOUs PDFs determination; in Sec. 4.2 we will show how we included at N³LO the contributions that are known and how we estimated the ones that are still missing and then we will present the aN³LO NNPDF4.0 PDFs; in the end, in Sec. 4.3, we will describe how the MHOUs and the aN³LO effects are combined with the QED corrections in a PDFs fit, we will present the resulting PDFs and we briefly assess the impact on the phenomenology.

4.1 MHOUs

As we already mentioned, PDFs uncertainties, as the ones in standard PDF sets such as NNPDF4.0 [4], CT18 [55], MSHT20 [56], or ABMP16 [54], generally do not account for theoretical uncertainties. These are uncertainties that impact the theoretical predictions that are compared to the data during the process of determining PDFs from experimental measurements. The only exceptions are the parametric uncertainty associated with the value of the strong coupling constant α_s , which has been routinely included since the early days of LHC physics [151], and nuclear uncertainties affecting, for example, DIS data on nuclear targets (such as neutrino DIS data), which are included in the NNPDF4.0 PDF determination [152, 153].

Theoretical uncertainties can arise from various sources, including both parametric uncertainties (such as the values of heavy quark masses) and non-parametric uncertainties (such as the nuclear corrections that mentioned earlier). In particular, theory uncertainties related to missing higher-order terms in QCD computations are important because they impact all predictions. The standard perturbative accuracy for QCD computations is currently at NNLO, with N³LO corrections known only in a few cases [154]. At NNLO, MHOUs are typically on the order of a few percent or more. For precision observables at the LHC used in PDF determination, such as gauge boson or top-pair production, these uncertainties can be comparable to experimental systematic uncertainties

and often exceed the experimental statistical uncertainties. Given that both experimental measurement uncertainties and theoretical prediction uncertainties are equally significant in the figure of merit used for PDF determination [155], it is important to include both types of uncertainties when they are of similar magnitude.

The aim of this section is to describe how to incorporate MHOUs into the NNPDF4.0 NLO and NNLO parton distribution sets using the methods outlined in Refs. [156, 157]. As a result, we will present the versions of the NNPDF4.0 global PDFs determination that now include the impact of MHOUs in the PDFs determination process. Refs. [156, 157] demonstrated that incorporating MHOUs into the PDF determination affects central values and enhances perturbative convergence. Therefore, parton distributions that include MHOUs in the PDFs uncertainty should now be considered the standard approach.

The inclusion of MHOUs into a PDFs fit is achieved by supplementing theoretical predictions with a covariance matrix that reflects their expected correlated variations when higher-order corrections are included. This estimation is conducted through scale variation. A comprehensive explanation of this methodology can be found in Refs. [156, 157]. In this section we will review the primary elements of the procedure, particularly focusing on its implementation in the `EKO` code [149], which is an integral part of the new NNPDF theory pipeline [114] employed in this work, and that we discussed in Sec. 2.3. We start by outlining the definition of perturbative expansions for various quantities. We then review the estimation of MHOUs on hard cross-sections and anomalous dimensions through scale variation. Finally, we summarize the construction of the MHOU covariance matrix.

4.1.1 Perturbative expansion and factorization

Theoretical predictions at hadron colliders rely on two quantities calculated perturbatively: the partonic cross-sections (or coefficient functions in the case of DIS data), and the splitting functions (or equivalently the anomalous dimensions in Mellin space) which govern the scale dependence of the PDF. Both quantities can be represented as series in the strong coupling $a_s(Q^2)$, which itself is determined perturbatively through the solution to Eq. (1.24), based on the strong coupling value at a reference scale, typically $a_s(M_Z^2)$. The MHOU in these predictions arises from truncating these perturbative expansions at a certain order.

In a VFNS, as discussed in Sec. 1.8, another MHOU arises from truncating the perturbative expansion of the matching conditions that connect PDFs across different schemes with a certain number of active flavors. These uncertainties are particularly significant when dealing with PDFs below the charm threshold, such as when determining the intrinsic charm PDF [29, 30]. However, for precision LHC phenomenology, predictions are typically made in a $n_f = 5$ scheme. Moreover, PDFs are derived by comparing predictions to data, most of which are also computed in the $n_f = 5$ scheme. As a result, matching uncertainties mainly impact the small amount of data below the bottom threshold (no data below the charm threshold is used), and this influence is minimal due to the small MHOU at the bottom threshold. Consequently, the MHOU associated with the matching conditions are minor and will be neglected in this discussion.

We therefore concentrate on MHOUs affecting the hard cross-sections and anomalous dimensions. To estimate these MHOUs through scale variation, we generate different expressions for any given perturbative result at a specified accuracy. These expressions vary by the sub-leading terms produced when the evaluation scale of the strong coupling is changed. Beginning with the DIS coefficient functions, we construct a scale-

varied N^kLO coefficient function. In this section we will adopt a relative counting for the perturbative orders of the coefficient functions, instead of the absolute one used in this thesis, see Eq. (1.43). With this notation the coefficient function's expansion takes the form

$$C_{a,i}(a_s(Q^2)) = a_s^m(Q^2) \sum_{n=0}^k (a_s(Q^2))^n C_{a,i}^{(n)}, \quad (4.1)$$

where we used that the coefficient function for a given process can start at $\mathcal{O}(a_s^m)$. For example, for the massive coefficient function of the gluon we have $m = 1$ while for the one of the quark $m = 2$.

The scale varied coefficient function is constructed as

$$\bar{C}_{a,i}(a_s(\mu^2), \rho_r) = a_s^m(\mu^2) \sum_{n=0}^k (a_s(\mu^2))^n \bar{C}_{a,i}^{(n)}(\rho_r), \quad (4.2)$$

with $\rho_r = \mu_r^2/Q^2$, by requiring that

$$\bar{C}_{a,i}(a_s(\rho_r Q^2), \rho_r) = C_{a,i}(a_s(Q^2)) [1 + \mathcal{O}(a_s)], \quad (4.3)$$

which fixes $\bar{C}_{a,i}^{(n)}(\rho_r)$ in terms of $C_{a,i}^{(n)}$. The explicit expressions of the scale varied coefficient functions are given by

$$\bar{C}_{a,i}^{(0)}(\rho) = C_{a,i}^{(0)}, \quad (4.4)$$

$$\bar{C}_{a,i}^{(1)}(\rho) = C_{a,i}^{(1)} + m C_{a,i}^{(0)} \beta_0 \ln \rho, \quad (4.5)$$

$$\bar{C}_{a,i}^{(2)}(\rho) = C_{a,i}^{(2)} + \frac{m(m+1)}{2} C_{a,i}^{(0)} (\beta_0)^2 \ln^2 \rho + \left((m+1) C_{a,i}^{(1)} \beta_0 + m C_{a,i}^{(0)} \beta_1 \right) \ln \rho, \quad (4.6)$$

$$\begin{aligned} \bar{C}_{a,i}^{(3)}(\rho) = & C_{a,i}^{(3)} + \frac{m(m+1)(m+2)}{6} C_{a,i}^{(0)} (\beta_0)^3 \ln^3 \rho \\ & + \left(\frac{(m+1)(m+2)}{2} C_{a,i}^{(1)} (\beta_0)^2 + \frac{m(2m+3)}{2} C_{a,i}^{(0)} \beta_0 \beta_1 \right) \ln^2 \rho \\ & + \left((m+2) C_{a,i}^{(2)} \beta_0 + (m+1) C_{a,i}^{(1)} \beta_1 + m C_{a,i}^{(0)} \beta_2 \right) \ln \rho. \end{aligned} \quad (4.7)$$

At any given order, $C_{a,i}$ and $\bar{C}_{a,i}$ differ by sub-leading terms. This difference serves as an estimate of the missing higher orders and can be used to construct an MHOU covariance matrix, as it will be discussed in the following of this section. This method of estimating MHOU on partonic cross-sections is referred to as renormalization scale variation.

With an analogous procedure we can obtain an estimation of the MHOU of the anomalous dimensions. Indeed, we construct a scale-varied anomalous dimension as

$$\bar{\gamma}_{ij}(a_s(\mu^2), \rho_f) = a_s(\mu^2) \sum_{n=0}^k (a_s(\mu^2))^n \bar{\gamma}_{ij}^{(n)}(\rho_f), \quad (4.8)$$

with $\rho_f = \mu_f^2/Q^2$, by requiring that

$$\bar{\gamma}_{ij}(a_s(\rho_f Q^2), \rho_f) = \gamma_{ij}(a_s(Q^2)) [1 + \mathcal{O}(a_s)], \quad (4.9)$$

that fixes $\bar{\gamma}_{ij}^{(n)}(\rho_f)$ in terms of $\gamma_{ij}^{(n)}$. The explicit expressions are those obtained from Eqs. (4.4-4.7) setting $m = 1$. Similarly, the sub-leading difference between $\gamma_{ij}^{(n)}$ and $\bar{\gamma}_{ij}^{(n)}$

can be used to estimate the MHOU on anomalous dimensions. This uncertainty subsequently translates into an MHOU on the PDF $f_i(Q^2)$ when it is expressed in terms of the PDFs at the parametrization scale as

$$f_i(Q^2) = \mathcal{P} \exp \left(- \int_{Q_0^2}^{Q^2} \frac{d\mu^2}{\mu^2} \gamma(a_s(\mu^2)) \right)_{ij} f_j(Q_0^2) = E_{ij}(Q^2 \leftarrow Q_0^2) f_j(Q_0^2), \quad (4.10)$$

where the product is a standard product in N space while it is a convolution in x space. This method of estimating the MHOU on the scale dependence of the PDF is referred to as factorization scale variation.

By substituting the scale-varied anomalous dimension $\bar{\gamma}_{ij}(\alpha(\mu^2), \rho_f)$ into the expression for the PDF in Eqs. (4.10), it can be demonstrated [157] that factorization scale variation can also be performed directly at the PDFs level. This is done by defining a scale-varied PDF $\bar{f}_i(Q^2, \rho_f)$, where its scale dependence is determined by a scale-varied evolution operator $\bar{E}_{ij}(Q^2 \leftarrow \mu_0^2, \rho_f)$:

$$\bar{f}_i(Q^2, \rho_f) = \bar{E}_{ij}(Q^2 \leftarrow \mu_0^2, \rho_f) f_j(\mu_0^2), \quad (4.11)$$

and the scale-varied evolution operator \bar{E} differs by sub-leading terms from the original one:

$$\bar{E}_{ij}(Q^2 \leftarrow \mu_0^2, \rho_f) = E_{ij}(Q^2 \leftarrow \mu_0^2) [1 + O(a_s)]. \quad (4.12)$$

Then, the scale-varied evolution operator can be constructed as

$$\bar{E}_{ij}(Q^2 \leftarrow \mu_0^2, \rho_f) = K_{il}(a_s(\rho_f Q^2), \rho_f) E_{lj}(\rho_f Q^2 \leftarrow \mu_0^2), \quad (4.13)$$

where at $N^k\text{LL}$ (i.e. with the anomalous dimension computed at $N^k\text{LO}$) the additional evolution kernel $K_{ij}(a_s(\rho_f Q^2), \rho_f)$ reads

$$K_{ij}(a_s(\rho_f Q^2), \rho_f) = \sum_{n=0}^k (a_s(\rho_f Q^2))^n K_{ij}^{(n)}(\rho_f). \quad (4.14)$$

Substituting this expansion into Eq. (4.12) determines all the coefficients $K_{ij}^{(j)}(\rho_f)$ in terms of $\gamma_{ij}^{(j)}$. Their specific expressions up to $N^3\text{LO}$ read

$$K_{ij}^{(0)}(\rho) = \delta_{ij}, \quad (4.15)$$

$$K_{ij}^{(1)}(\rho) = \gamma_{ij}^{(0)} \ln \rho, \quad (4.16)$$

$$K_{ij}^{(2)}(\rho) = \frac{1}{2} \gamma_{il}^{(0)} \left(\beta_0 \delta_{lj} + \gamma_{lj}^{(0)} \right) \ln^2 \rho + \gamma_{ij}^{(1)} \ln \rho, \quad (4.17)$$

$$K_{ij}^{(3)}(\rho) = \frac{1}{6} \gamma_{il}^{(0)} \left(2(\beta_0)^2 \delta_{lj} + 3\beta_0 \gamma_{lj}^{(0)} + \gamma_{lm}^{(0)} \gamma_{mj}^{(0)} \right) \ln^3 \rho \\ + \frac{1}{2} \left(\beta_1 \gamma_{ij}^{(0)} + 2\beta_0 \gamma_{ij}^{(1)} + \gamma_{il}^{(0)} \gamma_{lj}^{(1)} + \gamma_{il}^{(1)} \gamma_{lj}^{(0)} \right) \ln^2 \rho + \gamma_{ij}^{(2)} \ln \rho. \quad (4.18)$$

Eqs. (4.12-4.13) imply that the scale-varied evolution kernel in Eq. (4.12) evolves from μ_0^2 to $\rho_f Q^2$, and then from $\rho_f Q^2$ back to Q^2 , with the latter evolution expanded to fixed $N^k\text{LO}$.

The two approaches for performing factorization scale variation, either on anomalous dimensions as in Eqs. (4.8-4.9) or on PDFs as Eqs. (4.11-4.12) are equivalent. When

applied at N^kLO, both methods produce the same sub-leading N^{k+1}LO terms, although the higher-order terms may differ, namely

$$\begin{aligned} \bar{f}_i(Q^2, \rho_f) - f_i(Q^2) = \\ \mathcal{P} \exp \left(- \int_{\mu_0^2}^{\mu^2} \frac{d\mu'^2}{\mu'^2} \left[\bar{\gamma} \left(a_s(\rho_f \mu'^2), \rho_f \right) - \gamma \left(a_s(\mu'^2) \right) \right] \right) f_j(\mu_0^2) [1 + \mathcal{O}(a_s)] , \end{aligned} \quad (4.19)$$

These two methods of performing factorization scale variation by adjusting the scale of the anomalous dimension or by altering the scale of the PDFs are referred to as scheme A and scheme B, respectively, in Ref. [157]. There is a third method, called scheme C in the same reference, which involves using the scale-varied PDF from Eqs. (4.11-4.13), specifically

$$\bar{f}_i(Q^2, \rho_f) = K_{il}(a_s(\rho_f Q^2), \rho_f) E_{lh}(\rho_f Q^2 \leftarrow \mu_0^2) f_h(\mu_0^2) , \quad (4.20)$$

in the factorized expression, Eq. (1.42), but incorporating $K(a_s(\rho_f Q^2), \rho_f)$ into the coefficient function instead of the PDF. This effectively evaluates the PDF at a different scale, using a modified coefficient function. Indeed, inserting Eq. (4.20) into Eq. (1.42) the expression for the hadronic observable after factorization scale variation is

$$\begin{aligned} F_a(Q^2) &= C_{a,i}(Q^2) \bar{f}_i(Q^2, \rho_f) \\ &= C_{a,i}(Q^2) K_{il}(a_s(\rho_f Q^2), \rho_f) E_{lh}(\rho_f Q^2 \leftarrow \mu_0^2) f_h(\mu_0^2) \end{aligned} \quad (4.21)$$

$$= \bar{C}_{a,i}(Q^2, \rho_f) f_i(\rho_f Q^2) [1 + \mathcal{O}(a_s)] , \quad (4.22)$$

where we defined

$$\bar{C}_{a,i}(Q^2, \rho_f) = C_{a,j}(Q^2) K'_{ji}(a_s(Q^2), \rho_f) = a_s^m(Q^2) \sum_{n=0}^k (a_s(Q^2))^n \bar{C}_{a,i}^{(j)}(\rho_f) , \quad (4.23)$$

and K'_{ij} is in turn computed by re-expressing $K_{ij}(\rho_f Q^2, \rho_f)$ as a series in $a_s(Q^2)$, i.e.

$$K'_{ij}(a_s(Q^2), \rho_f) = \sum_{n=0}^k (a_s(Q^2))^n K_{ij}^{(n)}(\rho_f) , \quad (4.24)$$

with the requirement

$$K_{ij}(a_s(\rho_f Q^2), \rho_f) = K'_{ij}(a_s(Q^2), \rho_f) [1 + \mathcal{O}(a_s)] . \quad (4.25)$$

In the end we get

$$K_{ij}^{(0)}(\rho) = \delta_{ij} , \quad (4.26)$$

$$K_{ij}^{(1)}(\rho) = \gamma_{ij}^{(0)} \ln \rho , \quad (4.27)$$

$$K_{ij}^{(2)}(\rho) = \frac{1}{2} \gamma_{il}^{(0)} \left(-\beta_0 \delta_{lj} + \gamma_{lj}^{(0)} \right) \ln^2 \rho + \gamma_{ij}^{(1)} \ln \rho , \quad (4.28)$$

$$\begin{aligned} K_{ij}^{(3)}(\rho) &= \frac{1}{6} \gamma_{il}^{(0)} \left(2(\beta_0)^2 \delta_{lj} - 3\beta_0 \gamma_{lj}^{(0)} + \gamma_{lm}^{(0)} \gamma_{mj}^{(0)} \right) \ln^3 \rho \\ &+ \frac{1}{2} \left(-\beta_1 \gamma_{ij}^{(0)} - 2\beta_0 \gamma_{ij}^{(1)} + \gamma_{il}^{(0)} \gamma_{lj}^{(1)} + \gamma_{il}^{(1)} \gamma_{lj}^{(0)} \right) \ln^2 \rho + \gamma_{ij}^{(2)} \ln \rho , \end{aligned} \quad (4.29)$$

which, substituted in Eq. (4.23) leads to

$$\bar{C}_{a,i}^{(0)}(\rho) = C_{a,i}^{(0)}, \quad (4.30)$$

$$\bar{C}_{a,i}^{(1)}(\rho) = C_{a,i}^{(1)} + \gamma_{ij}^{(0)} C_{a,j}^{(0)} \ln \rho, \quad (4.31)$$

$$\bar{C}_{a,i}^{(2)}(\rho) = C_{a,i}^{(2)} + \frac{1}{2} \gamma_{il}^{(0)} \left(-\beta_0 \delta_{lj} + \gamma_{lj}^{(0)} \right) C_{a,j}^{(0)} \ln^2 \rho + \left(\gamma_{ij}^{(0)} C_{a,j}^{(1)} + \gamma_{ij}^{(1)} C_{a,j}^{(0)} \right) \ln \rho, \quad (4.32)$$

$$\begin{aligned} \bar{C}_{a,i}^{(3)}(\rho) = & C_{a,i}^{(3)} + \frac{1}{6} \gamma_{il}^{(0)} \left(2(\beta_0)^2 \delta_{lj} - 3\beta_0 \gamma_{lj}^{(0)} + \gamma_{lm}^{(0)} \gamma_{mj}^{(0)} \right) C_{a,j}^{(0)} \ln^3 \rho \\ & + \frac{1}{2} \left(-\beta_1 \gamma_{ij}^{(0)} - 2\beta_0 \gamma_{ij}^{(1)} + \gamma_{il}^{(0)} \gamma_{lj}^{(1)} + \gamma_{il}^{(1)} \gamma_{lj}^{(0)} \right) C_{a,j}^{(0)} \ln^2 \rho \\ & + \frac{1}{2} \gamma_{il}^{(0)} \left(-\beta_0 \delta_{lj} + \gamma_{lj}^{(0)} \right) C_{a,j}^{(1)} \ln^2 \rho \\ & + \left(\gamma_{ij}^{(0)} C_{a,j}^{(2)} + \gamma_{ij}^{(1)} C_{a,j}^{(1)} + \gamma_{ij}^{(2)} C_{a,j}^{(0)} \right) \ln \rho. \end{aligned} \quad (4.33)$$

In standard practice, factorization scale variation is typically done using scheme C, as this method does not require altering the PDFs, which are generally obtained from an external source. However, within the context of determining PDFs, factorization scale variation using scheme B, as defined in Eq. (4.11), is simpler because it only involves modifying the evolution operator used for PDF evolution. This is the approach we will use for factorization scale variation in this section.

4.1.2 Construction of the covariance matrix

The MHOU arising from the perturbative truncation of partonic cross-sections and the scale dependence of PDFs are estimated using renormalization scale variation, Eqs. (4.2-4.3), and factorization scale variation using scheme B from Ref. [157], Eq. (4.11) and Eq. (4.14). These uncertainties are incorporated via an MHOU covariance matrix, which is constructed as follows [156, 157].

First, we define the shift in the theoretical prediction for the i -th data point due to renormalization and factorization scale variation. It reads

$$\Delta_i(\rho_f, \rho_r) \equiv T_i(\rho_f, \rho_r) - T_i(0, 0), \quad (4.34)$$

where $T_i(\rho_f, \rho_r)$ is the prediction for the i -th data point with the renormalization and factorization scales varied by factors ρ_r and ρ_f , respectively.

Next, we select a correlation pattern for scale variation as follows [156, 157]:

- factorization scale variation is correlated across all data points because the scale dependence of PDFs is universal.
- renormalization scale variation is correlated for data points within the same category, such as the same observable (e.g., fully inclusive DIS cross-sections) or different observables within the same process (e.g., Z transverse momentum and rapidity distributions).

This approach necessitates categorizing processes. For example, we treat charged-current and neutral-current DIS as distinct processes.

These choices are based on the assumptions that factorization and renormalization scale variations adequately capture the MHOU for anomalous dimensions and partonic cross-sections, respectively, and that missing higher-order terms are similar in nature

and magnitude across all processes within a given category. Other assumptions are possible; for instance, one might decorrelate the renormalization scale variation for different partonic sub-channels within the same process or introduce an additional variation in the process scale on top of the renormalization and factorization scale variations discussed above (see Section 4.3 of Ref. [157] for a more detailed discussion).

Then, we define an MHOUs covariance matrix, with the matrix element between two data points i and j given by:

$$S_{ij} = n_m \sum_{V_m} \Delta_i(\rho_f, \rho_{r_i}) \Delta_j(\rho_f, \rho_{r_j}), \quad (4.35)$$

where the sum runs over the space V_m of the m included scale variations. The factorization scale ρ_f is always varied in a correlated manner, while the renormalization scales ρ_{r_i} and ρ_{r_j} are varied in a correlated way ($\rho_{r_i} = \rho_{r_j}$) if data points i and j belong to the same category, but are varied independently if they belong to different categories. The term n_m is a normalization factor. Calculating this normalization factor is complex because it must account for the mismatch in the dimension of the scale variation space when two data points are in the same category (resulting in one correlated set of renormalization scale variations) versus when they are not (resulting in two independent sets of variations). These normalization factors were computed for various choices of the scale variation space V_m and for different values of m in Ref. [157], which provides detailed explanations.

As in Refs. [156, 157], we consider scale variations by a factor of 2, so:

$$\kappa_f = \ln \rho_f = \pm \ln 4, \quad \kappa_r = \ln \rho_r = \pm \ln 4. \quad (4.36)$$

In Ref. [157], various choices for the space of allowed variations were considered. These included the 9-point prescription, where κ_r and κ_f can take any of the values $(0, \pm \ln 4)$, resulting in $m = 8$ (eight variations around the central value); and the commonly used 7-point prescription, with $m = 6$, derived from the former by excluding the two outermost variations ($\kappa_r = +\ln 4, \kappa_f = -\ln 4$ and $\kappa_r = -\ln 4, \kappa_f = +\ln 4$). It can be shown [7] that both the 7-point and 9-point prescriptions exhibit similar behavior, consistent with the findings of Ref. [157]. Other prescriptions, with a more limited set of independent scale variations, were shown in Ref. [157] to perform less effectively, and we will not consider them further. The explicit expressions for the MHOUs covariance matrix using the 7-point and 9-point prescriptions are provided in Eqs. (4.18-4.19) and Eq. (4.15) of Ref. [157], respectively.

The assumptions regarding the correlation patterns of renormalization and factorization scale variations, the process categorization, the scale variation range, and the specific choice of variation points are somewhat arbitrary. This is unavoidable because MHOUs is an estimate of the probability distribution for the size of an unknown quantity that has a single true value, making it intrinsically Bayesian. The only way to validate this type of estimate is by comparing its performance to cases where the true value is known, as it has been done in Ref. [7].

4.1.3 Results

We now focus on the primary results of this section: the NNPDF4.0 NLO and NNLO PDFs sets with MHOUs. These sets are obtained by repeating the NNPDF4.0 PDF determinations, but this time including an MHOUs covariance matrix determined using a 7-point prescription, as discussed previously in this section. The dataset used is identical

Dataset	N_{dat}	NLO		NNLO	
		$C + S^{(\text{nucl})}$	$C + S^{(\text{nucl})} + S^{(7\text{pt})}$	$C + S^{(\text{nucl})}$	$C + S^{(\text{nucl})} + S^{(7\text{pt})}$
DIS NC	2100	1.30	1.22	1.23	1.20
DIS CC	989	0.92	0.87	0.90	0.90
DY NC	736	2.01	1.71	1.20	1.15
DY CC	157	1.48	1.42	1.48	1.37
Top pairs	64	2.08	1.24	1.21	1.43
Single-inclusive jets	356	0.84	0.82	0.96	0.81
Dijets	144	1.52	1.84	2.04	1.71
Prompt photons	53	0.59	0.49	0.75	0.67
Single top	17	0.36	0.35	0.36	0.38
Total	4616	1.34	1.23	1.17	1.13

Table 4.1: The number of data points and the χ^2 per data point for the NLO and NNLO NNPDF4.0 PDF sets without and with MHOUs.

to that for the NNPDF4.0 NNLO PDFs [4]. As mentioned in Ref. [7], we use the same dataset for both NLO and NNLO, whereas in Ref. [4] a slightly different dataset was used for NLO. Therefore, we compare four PDF sets here: NLO and NNLO, with and without MHOUs, all based on the same underlying data.

In Tab. 4.1 we present the number of data points and the χ^2 per data point for the NLO and NNLO NNPDF4.0 PDF determinations, both before and after incorporating MHOUs. Without MHOUs, the covariance matrix is defined as in Ref. [4], consisting of the experimental covariance matrix C and a theory covariance matrix $S^{(\text{nucl})}$, which accounts for missing nuclear corrections as detailed in Refs. [152, 153], with its effects discussed in Sec. 8.6 of Ref. [4]. When MHOUs are included, the covariance matrix also incorporates the contribution described in Eq. (4.35) and discussed in previously, which we refer to as $S^{(7\text{pt})}$.

Note that the MHOUs contribution is either excluded or included in both the definition of the χ^2 used by the NNPDF algorithm (i.e. for pseudodata generation and in the training and validation loss functions) and in the covariance matrix used to calculate the values provided in Tab. 4.1. Additionally, the experimental covariance matrix used to compute the values in Tab. 4.1 differs from the one used in the NNPDF algorithm. The latter applies the t_0 method [158] to handle multiplicative uncertainties and avoid the d’Agostini bias, while the former is simply the published experimental covariance matrix.

Tab. 4.1 shows that including the MHOUs covariance matrix results in a reduction of the total χ^2 for both NLO and NNLO fits, with a more significant decrease observed at NLO. Despite this inclusion, the NLO χ^2 remains slightly higher than the NNLO χ^2 . We have confirmed that this is caused by a small number of highly precise data points (excluded by the NLO cuts in Ref. [4]), where NNLO corrections are significantly underestimated by scale variation. However, for most data points and process categories, the MHOUs covariance matrix adequately accounts for the discrepancy between data and theoretical predictions at NLO due to missing NNLO terms.

In Fig. 4.1 it is shown the comparison between the individual PDFs at NLO and NNLO, with and without MHOUs, at $Q = 100$ GeV. The displayed distributions include the gluon, singlet and non-singlet ($V, V_3, V_8, T_3, T_8, T_{15}$) PDFs, defined in Eq. (1.73) and Eqs. (1.80-1.89), all normalized to the NNLO PDFs with MHOUs. The associated one-

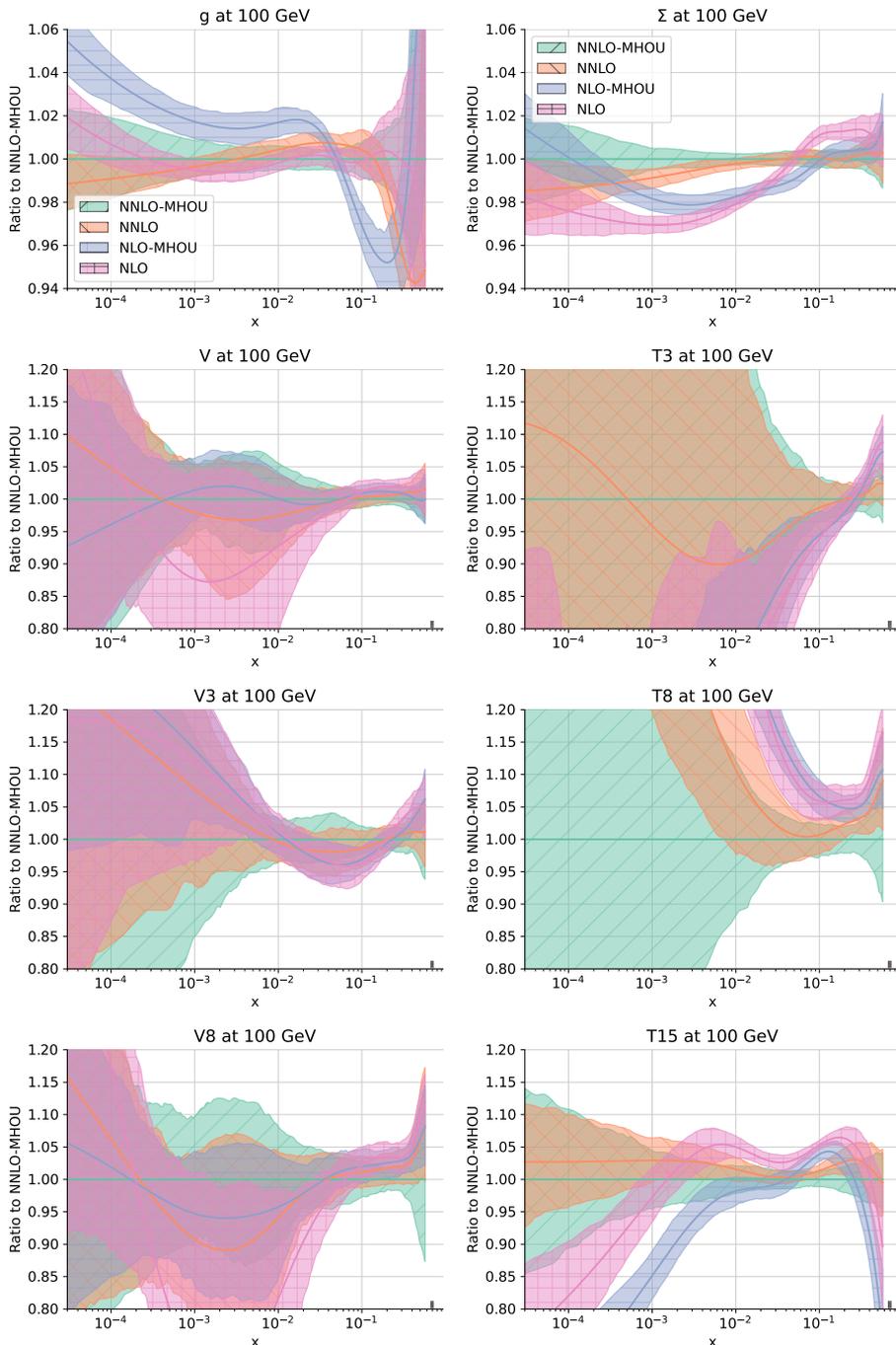


Figure 4.1: The NLO and NNLO PDFs, both with and without MHOUs, at $Q = 100$ GeV. The gluon, singlet, valence (V , V_3 , V_8), and triplet (T_3 , T_8 , T_{15}) PDFs are displayed. All curves are normalized to the NNLO PDFs with MHOUs. The bands represent the 1σ uncertainty.

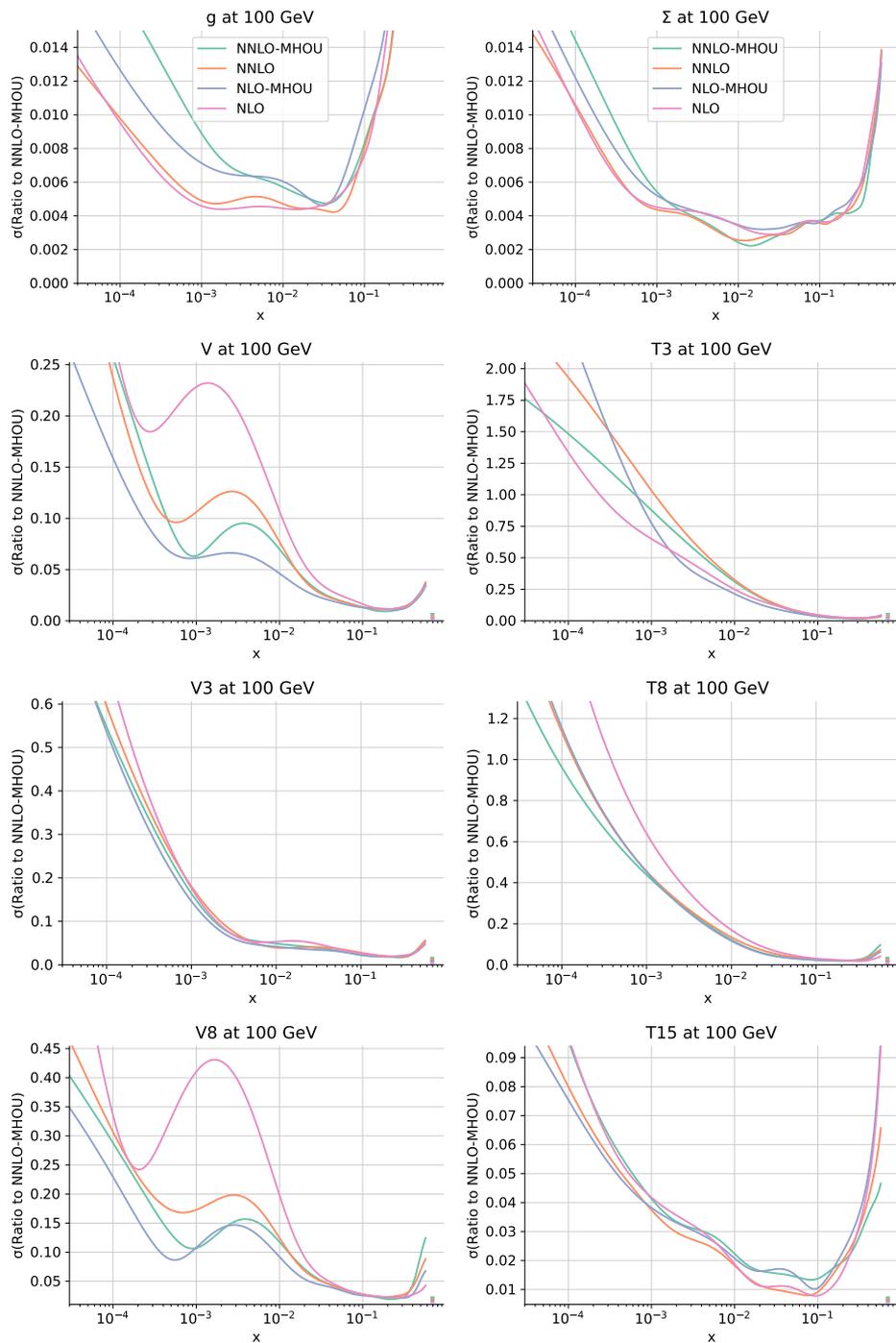


Figure 4.2: Relative 1σ uncertainties for the PDFs shown in Fig. 4.1. All uncertainties are normalized to the corresponding central NNLO PDFs with MHOUs.

sigma uncertainties are illustrated in Fig. 4.2. The inclusion of MHOUs generally causes only a moderate shift in the central values at NNLO; however, at NLO, the shift is notable for the gluon and singlet PDFs, though relatively modest for other PDF combinations.

Observing Fig. 4.2, it is evident that, on average, the PDFs uncertainty at NNLO remains largely unaffected by the inclusion of MHOUs in the data region, i.e. for $x \gtrsim 10^{-4}$ (see Fig. 2.6). However, in the singlet sector, the uncertainty increases at small x , notably for the gluon, where the increase is evident up to $x \sim 0.1$. At NLO, the uncertainty generally decreases in the non-singlet sector, whereas it rises across all x values in the singlet sector, with the gluon showing a particularly significant increase. This is in line with the fact that at NLO, the MHOUs from scale variation does not fully explain the shift from NLO to NNLO for some datasets. The somewhat unexpected observation that the PDF uncertainty does not necessarily increase, and may even decrease, when adding an additional uncertainty source to the χ^2 was previously noted in Refs. [152, 153]. This result shows the improved data compatibility achieved by incorporating the MHOUs.

The impact of incorporating MHOUs on PDFs uncertainties varies with both x and the specific PDF, so a comprehensive quantitative assessment requires evaluating the PDFs uncertainty on physics predictions. This can be achieved using the ϕ estimator, introduced in Ref. [119], defined by:

$$\phi_{\chi^2} = \sqrt{\langle \chi^2 \rangle - \chi^2}, \quad (4.37)$$

where $\langle \chi^2 \rangle$ represents the average χ^2 (per data point) computed for each individual replica and then averaged over all replicas, while χ^2 is the value listed in Tab. 4.1 and is obtained using the best-fit PDF, which is the average over replicas. For a single data point, ϕ is simply the ratio of the PDFs uncertainty to the data uncertainty. For multiple uncorrelated data points, ϕ is the square root of the average ratio of PDF variance to data variance. For correlated data points, it is this ratio calculated using the eigenvectors of the experimental covariance matrix, averaging the ratios of the diagonal elements of the theory covariance matrix to the eigenvalues of the experimental covariance matrix. Thus, ϕ measures the PDFs uncertainty relative to the experimental uncertainties, providing an estimate of the data consistency. A ϕ value less than 1 indicates that, on average, the uncertainties in the predictions are smaller than those of the original data, suggesting successful combination of consistent data by the underlying theory (refer to Sec. 6 of Ref. [157] for further discussion).

Tab. 4.2 presents the ϕ values both before and after incorporating MHOUs at NLO and NNLO. It is evident that the inclusion of MHOUs either leaves ϕ unchanged or leads to a decrease. The reduction is notably more pronounced for processes that involve non-singlet combinations, such as charged-current Drell-Yan, aligning with the patterns observed in the PDFs uncertainties shown in Fig. 4.2. Overall, the decrease is more pronounced at NNLO compared to NLO.

The decrease in ϕ indicates that incorporating MHOUs generally reduces the PDFs uncertainties for physical predictions within the data region. However, it's notable that this reduction, while evident at NLO for some PDF combinations, is not always reflected at NNLO in the PDF plots of Figs. 4.1-4.2, where an increase in uncertainty, particularly in the singlet sector, is observed. It is important to note that the uncertainties shown in Figs. 4.1-4.2 represent diagonal uncertainties, which combine both correlated and uncorrelated uncertainties in quadrature. In contrast, the calculation of physical observables involves integrating over different values of the momentum fraction, which introduces correlations between these values. These correlations are accounted for in the ϕ indicator. In fact, MHOUs tend to be highly correlated because they generally have a smooth

Dataset	NLO		NNLO	
	$C + S^{(\text{nucl})}$	$C + S^{(\text{nucl})} + S^{(7\text{pt})}$	$C + S^{(\text{nucl})}$	$C + S^{(\text{nucl})} + S^{(7\text{pt})}$
DIS NC	0.14	0.13	0.15	0.13
DIS CC	0.12	0.12	0.12	0.12
DY NC	0.19	0.17	0.18	0.17
DY CC	0.37	0.30	0.35	0.32
Top pairs	0.19	0.16	0.17	0.17
Single-inclusive jets	0.13	0.12	0.13	0.13
Dijets	0.10	0.09	0.11	0.10
Prompt photon	0.06	0.06	0.06	0.06
Single top	0.04	0.04	0.04	0.04
Total	0.16	0.15	0.17	0.15

Table 4.2: The ϕ estimator of Eq. (4.37) for the PDFs at NLO and NNLO with and without MHOU's..

impact, affecting kinematically close data points similarly. Correlated uncertainties are always smaller than or equal to uncorrelated uncertainties. Therefore, the different behaviors of diagonal PDF uncertainties and ϕ suggest that the correlation of MHOU's contributes to their impact on PDFs being highly correlated. It should also be noted that the ϕ indicator does not provide information on uncertainty behavior in extrapolation regions. Consequently, when making predictions outside the current dataset's kinematic range, PDF uncertainties may increase despite the inclusion of MHOU's.

It is interesting to compare the behavior of the ϕ indicator shown in Tab. 4.2 with that discussed in Ref. [157] (at NLO only), specifically in Section 6 (Tab. 6 and Tab. 8). Notably, the ϕ value before the inclusion of MHOU in that reference was more than twice as high as it is here. This difference is due to the uncertainties in the NNPDF4.0 PDFs set discussed here being significantly smaller than those in the previously available NNPDF3.1 PDF set. This reduction in uncertainties is attributed to methodological improvements, even with the same underlying dataset, as extensively discussed in Section 8.2 (see Figure 46) of Ref. [4]. For NNPDF4.0 NNLO, $\phi = 0.16$ (Tab. 31 of Ref. [4]), while for NNPDF3.1 NNLO, $\phi = 0.36$ (Tab. 8 of Ref. [157]).

Furthermore, Ref. [157] noted that adding a MHOU term to the covariance matrix used in the fit would theoretically increase the uncertainty of the result. For a single data point, this increase would be the quadrature sum of the MHOU term and the experimental uncertainty. For multiple correlated measurements, this can be expressed as an expected increase in the ϕ indicator based on the experimental and MHOU covariance matrices (see Eq.(6.5) of Ref. [157]). It was observed that while the ϕ value did increase with the inclusion of MHOU's, the increase was less than expected, suggesting that the data became more compatible when MHOU's were included.

Here, the inclusion of MHOU's actually reduces the value of ϕ , despite the lower initial value of ϕ with the NNPDF4.0 methodology. This indicates a more significant improvement in data compatibility. This improvement is partly due to the shift in emphasis from DIS data in NNPDF3.1 to hadron collider data in NNPDF4.0 (see Sections 7.2.4 and 7.2.5 of Ref. [4]). Since higher-order corrections have a greater impact on hadronic processes than on DIS, the influence of MHOU's is correspondingly greater. Additionally, with the NNPDF4.0 methodology, uncertainties are smaller for the same data uncertainties, which amplifies the effect of tension between data from different processes due to MHO corrections and enhances the impact of improved compatibility upon inclusion of

MHOUs. Thus, the evidence suggests that data compatibility is significantly increased by including PDF uncertainties, particularly at NNLO.

A priori, PDF sets with and without MHOUs are not necessarily expected to be compatible within uncertainties, since the latter do not account for an existing source of uncertainty. In practice, they agree in the non-singlet sector, where MHOUs are at most comparable to the PDF uncertainty before including MHOUs. However, in the singlet sector, where NNLO corrections can be very large, they generally do not agree.

Including MHOUs generally shifts the NLO PDFs closer to the NNLO PDFs, thereby enhancing perturbative convergence, with the exception of the gluon. For the singlet sector, while the NLO PDFs move towards the NNLO PDFs with the inclusion of MHOUs, the NNLO result remains outside the NLO uncertainty band, especially at small x . This indicates that there are substantial NNLO corrections to the NLO result in the singlet sector that MHOUs determined by scale variation underestimate. At small x , this can be attributed to unresummed small- x logarithms [159], which increase with perturbative order and are not accounted for by scale variation.

The overall conclusion is that including MHOUs estimated through scale variation enhances data compatibility. This inclusion causes a moderate shift in PDF central values and reduces PDF uncertainties on physics predictions in the data region, as evidenced by a lower ϕ indicator. At NNLO, diagonal PDF uncertainties remain mostly unchanged, but at NLO, they are reduced in the non-singlet sector and generally increase for the gluon at both NLO and NNLO. At NLO, including MHOUs accounts for the effect of MHO terms on fit quality while having a moderate impact on PDF uncertainties and central values in the non-singlet sector. In the singlet sector, it has a more significant impact on central values and increases uncertainties, though it does not fully account for the largest missing NNLO corrections. In the small- x extrapolation region, PDF uncertainties generally increase with the inclusion of MHOUs at both NLO and NNLO.

A more in-depth examination of perturbative stability and the impact of including MHOUs will be covered in the next section, which extends the PDF determination and the results presented here to N³LO.

4.2 N³LO

Calculations of hard-scattering cross-sections at N³LO have been available for massless DIS for a long time [160–163]. More recently, N³LO calculations have become available for a large set of hadron collider processes. These include inclusive Higgs production via gluon-fusion [164, 165], bottom-fusion [166], in association with vector bosons [154], and via vector-boson-fusion [167], as well as Higgs pair production [168], inclusive Drell-Yan production [169, 170], differential Higgs production [171–175], and differential Drell-Yan distributions [176, 177]. For a comprehensive overview, see Ref. [178].

To achieve predictions for hadronic observables with this level of accuracy, it is essential to pair these partonic cross-sections with PDFs determined at the same perturbative order. These PDFs must be derived by comparing theoretical predictions calculated to the same precision with experimental data. The primary challenge in this process is the absence of complete expressions for the N³LO splitting functions, which dictate the scale dependence of the PDFs. Currently, only partial information is available on these functions [179–189]. This information includes a set of integer N -Mellin moments, terms proportional to n_f^k with $k \geq 1$, and limits at large and small x . By including these partial results, it is possible to make an approximate determination of the N³LO splitting functions [188, 190], similar to the successful approach used for NNLO [191].

Currently, a global PDF determination at N³LO must rely on partial information: approximate knowledge of splitting functions and complete partonic cross-sections only for a limited range of processes. A preliminary effort to address this was recently undertaken in Ref. [190]. In this work, the lacking theoretical details for N³LO calculations were modeled using a set of nuisance parameters, which were then determined alongside the PDFs through a fit to experimental data.

In our analysis, we employ a somewhat different strategy. Specifically, we use a theory covariance matrix formalism to address the absence of certain perturbative information. Ref. [152] shows that nuclear uncertainties can be incorporated via a theory covariance matrix. In Sec. 4.1 we have described the construction of such a matrix to account for MHOUs, which are estimated through variations in renormalization and factorization scales. In this section, we apply the same formalism to create a theory covariance matrix for incomplete higher-order uncertainties (IHOUs). These IHOUs are related to the incomplete knowledge of N³LO theory, particularly concerning the splitting functions and the massive DIS coefficient functions. With these theory covariance matrices in hand, we can determine PDFs at an “approximate N³LO” level (aN³LO). This determination includes contributions from both the incomplete knowledge of N³LO splitting functions and massive coefficient functions (IHOUs), as well as missing N³LO corrections to the partonic cross-sections for hadronic processes (MHOUs).

We will present the aN³LO NNPDF4.0 PDF determination, which will be added to the existing LO, NLO, and NNLO sets [4], as well as the more recent NNPDF4.0MHOUs PDFs [7] that include MHOUs in the PDF uncertainty. Unlike the methodology used in the MSHT20 study [190], our approach incorporates newer exact results [186–189] that enhance the stability of the N³LO splitting function parametrization. Our construction is carried out within the open-source NNPDF framework [192]. Specifically, our aN³LO evolution is integrated into EKO [149], while the N³LO DIS coefficient functions are implemented in the `Adani` code [193, 194] and interfaced to `YADISM` [115] for the construction of the FONLL general-mass scheme. By determining PDFs from a unified global dataset and using the same methodology across four consecutive perturbative orders, we can now evaluate perturbative stability more precisely and provide a more accurate estimation of uncertainties.

4.2.1 Approximate N³LO evolution

We will start with the development and application of aN³LO evolution. First, we outline our approach to approximating the N³LO evolution equations, explaining how this approximation is used to derive aN³LO anomalous dimensions and splitting functions, and how we assess the uncertainty in this approximation. Next, we describe our implementation of aN³LO evolution and analyze how this affects the perturbative evolution of PDFs showing the results obtained evolving a given initial condition with and without the aN³LO corrections.

As we have seen in Eqs. (1.99-1.102) the maximally decoupled DGLAP equations in Mellin space depend on seven independent anomalous dimensions, which rule the evolution of three PDFs sectors that are decoupled from each other. Such sectors are:

- The non-singlet combinations, Eqs. (1.80-1.89), depending on γ^\pm .
- The valence distribution, Eq. (1.73), depending on $\gamma^V = \gamma^- + \gamma^S$.
- The singlet sector, composed by gluon and singlet distribution defined in Eq. (1.73), depending on $\gamma_{gg}, \gamma_{qg}, \gamma_{gq}$ and $\gamma_{qq} = \gamma^+ + \gamma_{ps}$.

Each of the seven independent anomalous dimensions satisfy a perturbative expansion in a_s . Our aim is to find approximate expressions at N³LO for the seven anomalous dimensions that we need.

The information that can be utilized to achieve this goal comes from three different sources:

1. full analytic knowledge of contributions to the anomalous dimensions proportional to the highest powers of the number of flavors n_f ;
2. large- x and small- x resummations provide all-order information on terms that are logarithmically enhanced by powers of $\ln(1-x)$ and $\ln x$, respectively;
3. analytic knowledge of a finite set of integer moments.

To construct an approximation based on this information, we first separate the analytically known terms (1-2). Next, we expand the remainder using a set of basis functions and use the known moments to determine the expansion coefficients. Finally, we vary the set of basis functions to estimate the uncertainties.

Schematically, we proceed in the following way:

1. We include all terms in the expansion

$$\gamma_{ij}^{(3)}(N) = \gamma_{ij}^{(3,0)}(N) + n_f \gamma_{ij}^{(3,1)}(N) + n_f^2 \gamma_{ij}^{(3,2)}(N) + n_f^3 \gamma_{ij}^{(3,3)}(N), \quad (4.38)$$

of the anomalous dimension in powers of n_f that are fully or partially known analytically. We call such terms $\gamma_{ij,n_f}^{(3)}(N)$.

2. We incorporate all terms from large- x and small- x resummation up to the highest known logarithmic accuracy, including all known sub-leading power corrections in both limits. These terms are denoted as $\gamma_{ij,N \rightarrow \infty}^{(3)}(N)$ and $\gamma_{ij,N \rightarrow 0}^{(3)}(N)$, as well as $\gamma_{ij,N \rightarrow 1}^{(3)}(N)$. Any potential double counting arising from the overlap of these terms with $\gamma_{ij,n_f}^{(3)}(N)$ is eliminated.

3. We write the approximate anomalous dimension matrix element $\gamma_{ij}^{(3)}(N)$ as the sum of the exactly known terms and a leftover $\tilde{\gamma}_{ij}^{(3)}(N)$ according to

$$\gamma_{ij}^{(3)}(N) = \gamma_{ij,n_f}^{(3)}(N) + \gamma_{ij,N \rightarrow \infty}^{(3)}(N) + \gamma_{ij,N \rightarrow 0}^{(3)}(N) + \gamma_{ij,N \rightarrow 1}^{(3)}(N) + \tilde{\gamma}_{ij}^{(3)}(N). \quad (4.39)$$

We write $\tilde{\gamma}_{ij}^{(3)}(N)$ as a linear combination of a set of n^{ij} interpolating functions $G_\ell^{ij}(N)$ (kept fixed, see Ref. [8] for the explicit form that has been chosen) and $H_\ell^{ij}(N)$ (to be varied)

$$\tilde{\gamma}_{ij}^{(3)}(N) = \sum_{\ell=1}^{n^{ij}-n_H} b_\ell^{ij} G_\ell^{ij}(N) + \sum_{\ell=1}^{n_H} b_{n^{ij}-2+\ell}^{ij} H_\ell^{ij}(N), \quad (4.40)$$

with n^{ij} equal to the number of exactly known Mellin moments of $\gamma_{ij}^{(3)}(N)$. We fix the coefficients b_ℓ^{ij} by equating the values of $\gamma_{ij}^{(3)}(N)$ to the known moments of the given splitting functions.

4. In the singlet sector, we set $n_H = 2$ and make \tilde{N}_{ij} different selections for the two functions $H_1^{ij}(N)$ and $H_2^{ij}(N)$ from a list of distinct basis functions. This results in \tilde{N}_{ij} expressions for the remainder $\tilde{\gamma}_{ij}^{(3)}(N)$ and, consequently, for the N³LO anomalous dimension matrix element $\gamma_{ij}^{(3)}(N)$ through Eq. (4.39). These expressions are used to construct the approximate anomalous dimension matrix and its associated uncertainty. The functions $H_\ell^{ij}(N)$ have been chosen in order to reproduce the singularity structure of the anomalous dimensions in N space. Instead, the number of varying H_{ij} (therefore, the value of \tilde{N}_{ij}) is chosen to be larger when less exact information is known. See Ref. [8] for a more detailed discussion.
5. In the non-singlet sector we set $n_H = 0$, meaning that we use a single approximation and disregard its uncertainty. The reason of this choice is that its uncertainty is very small with respect to the one of the singlet sector [8].

4.2.2 The approximate anomalous dimensions and their uncertainty

The procedure outlined so far yields an ensemble of \tilde{N}_{ij} different approximations to the N³LO anomalous dimension, represented as $\gamma_{ij}^{(3),(k)}(N)$ for $k = 1, \dots, \tilde{N}_{ij}$. Our best estimate for the approximate anomalous dimension is then the average of these approximations

$$\gamma_{ij}^{(3)}(N) = \frac{1}{\tilde{N}_{ij}} \sum_{k=1}^{\tilde{N}_{ij}} \gamma_{ij}^{(3),(k)}(N). \quad (4.41)$$

We account for the uncertainty in the approximation, and the resulting uncertainty in N³LO theory predictions, using the general formalism for theory uncertainties developed in Refs. [152, 156, 157]. Specifically, we treat the uncertainty in each anomalous dimension matrix element due to its incomplete knowledge as a source of theoretical prediction uncertainty, independent of other uncertainty sources, and we disregard potential correlations between the incomplete knowledge of individual anomalous dimensions $\gamma_{ij}^{(3)}$. This IHOU for N³LO terms is managed similarly to the MHOu.

Specifically, we determine the shift in the theory prediction for the m -th data point when the central anomalous dimension matrix element $\gamma_{ij}^{(3)}(N)$, given by Eq. (4.41), is replaced by each of the $\gamma_{ij}^{(3),(k)}(N)$ instances, treating each instance as an independent nuisance parameter:

$$\Delta_m(ij, k) = T_m(ij, k) - \bar{T}_m, \quad (4.42)$$

where \bar{T}_m is the prediction for the m -th data point using the best estimate from Eq. (4.41) for the full anomalous dimension matrix, and $T_m(ij, k)$ is the prediction obtained when the ij matrix element of the best estimate is replaced with the k -th instance $\gamma_{ij}^{(3),(k)}(N)$.

We then construct the covariance matrix for theory predictions for individual data points due to the IHOU on the ij N³LO matrix element as the covariance of the shifts $\Delta_m(ij, k)$ across all \tilde{N}_{ij} instances:

$$\text{cov}_{mn}^{(ij)} = \frac{1}{\tilde{N}_{ij} - 1} \sum_{k=1}^{\tilde{N}_{ij}} \Delta_m(ij, k) \Delta_n(ij, k). \quad (4.43)$$

We note that we do not assign an IHOU to the non-singlet anomalous dimensions and assume, conservatively, that there is no correlation between the different singlet anoma-

lous dimension matrix elements. Thus, the total contribution to the theory covariance matrix due to IHOU can be written as:

$$\text{cov}_{mn}^{\text{IHOU}} = \text{cov}_{mn}^{(gg)} + \text{cov}_{mn}^{(gq)} + \text{cov}_{mn}^{(qg)} + \text{cov}_{mn}^{(qq)}. \quad (4.44)$$

The mean square uncertainty on the anomalous dimension matrix element is then determined by treating it as a pseudo-observable, calculated as the variance

$$(\sigma_{ij}(N))^2 = \frac{1}{\tilde{N}_{ij} - 1} \sum_{k=1}^{\tilde{N}_{ij}} \left(\gamma_{ij}^{(3),(k)}(N) - \gamma_{ij}^{(3)}(N) \right)^2. \quad (4.45)$$

For all the details of the construction of the non-singlet and singlet anomalous dimensions through the procedure outlined in this section see Ref. [8], where the explicit form of all the ingredients is written down.

4.2.3 Results: aN³LO evolution

The aN³LO anomalous dimensions discussed in the previous sections have been integrated into the Mellin-space open-source evolution code `EKO` [149]. This code is part of the new pipeline [114] adopted by NNPDF for producing theoretical predictions used in PDF determination. The parameterization is formulated using a basis of Mellin space functions that are efficient to evaluate numerically. To achieve full aN³LO accuracy, in addition to the anomalous dimensions, the four-loop running of the strong coupling constant $\alpha_s(Q)$ and the N³LO matching conditions, which govern the transitions between schemes with different numbers of active quark flavors, have also been implemented.

The N³LO matching conditions were initially presented in Ref. [195] and later computed analytically in Refs. [196–205]. However, the $a_{Hg}^{(3)}$ entry of the matching condition matrix remains unknown¹. This unknown entry is currently parameterized using the first five known moments [195] and the LLx contribution, as described in Ref. [207]. These matching conditions have also been implemented in `EKO`, allowing for an assessment of the impact of including aN³LO terms on perturbative evolution.

In Fig. 4.3, we compare the evolution of a fixed set of PDFs from $Q_0 = 1.65$ GeV to $Q = 100$ GeV at NLO, NNLO, and aN³LO. Using the NNPDF4.0NNLO PDF set as input, the results are normalized to the aN³LO evolution. The comparisons are made for all combinations that evolve differently, as discussed at the beginning of this section, specifically the singlet, gluon, total valence, and non-singlet \pm combinations. The x axis is on a logarithmic scale for the singlet sector and on a linear scale for the valence and non-singlet combinations. The relative uncertainty for the gluon and singlet is shown in Fig. 4.4, with MHOU and IHOU separately displayed at N³LO.

In all cases, the perturbative expansion seems to converge uniformly, showing minimal differences between NNLO and aN³LO, except at very small $x \lesssim 10^{-3}$. Here, the singlet evolution is less pronounced at aN³LO compared to NNLO. This dip in gluon-driven small- x evolution affects the behavior of all quark and gluon PDFs in this region. Notably, this behavior is an all-order effect that remains consistent even after small- x resummation. In fact, the total theory uncertainty at aN³LO is at the sub-percent level for all $x \gtrsim 10^{-3}$. Therefore, not only has the MHOU become negligible, but the impact of IHOU on PDF evolution is also only significant at very small x .

¹The terms recently computed in Ref. [206] are not yet included and will be addressed in future updates.

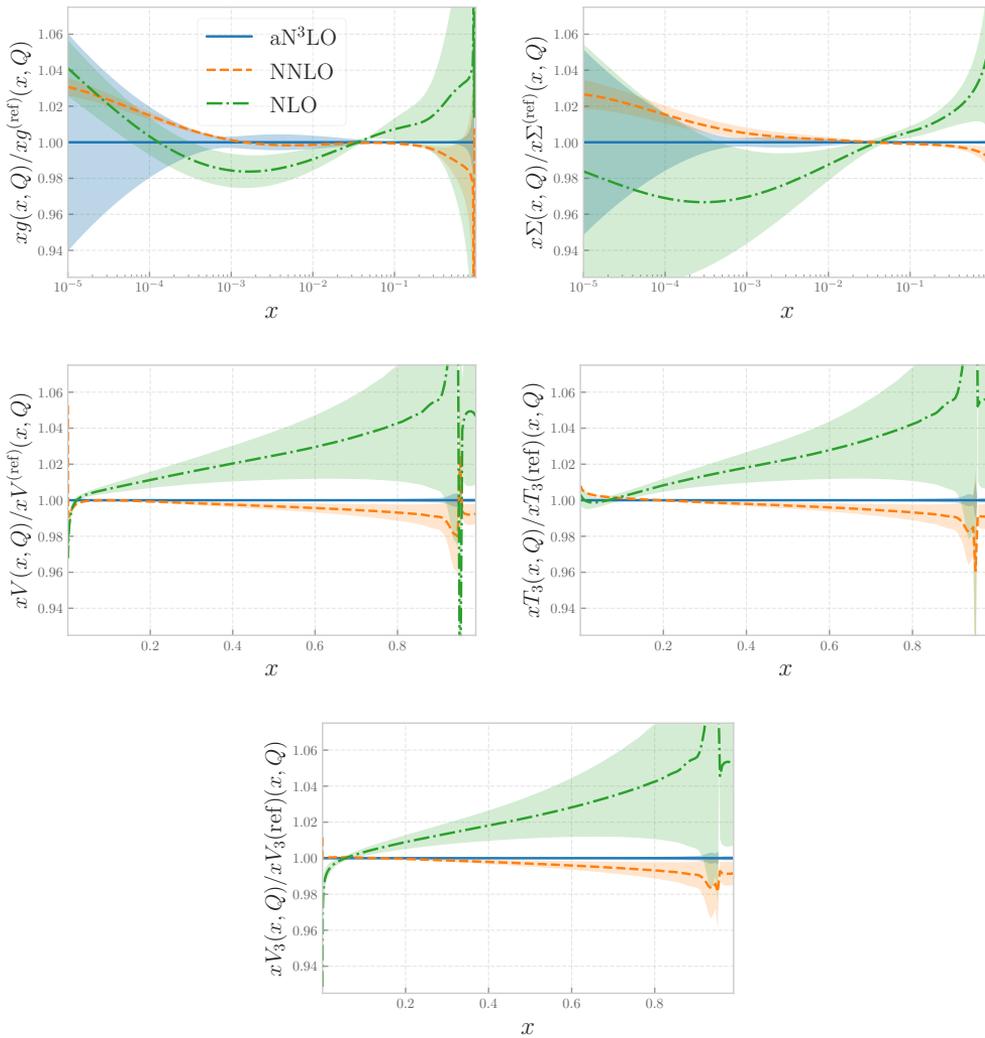


Figure 4.3: Comparison obtained evolving from $Q_0 = 1.65$ GeV to $Q = 100$ GeV at NLO, NNLO, and aN^3LO , using NNPDF4.0 NNLO as the fixed starting PDF. The results are presented as ratios to the aN^3LO (arranged from left to right and from top to bottom) for the gluon, singlet Σ , and the quark eigenstates V , V_3 , and T_3 in perturbative evolution. The total theory uncertainty is displayed in all cases, including the MHOÜ at NLO and NNLO, and the combined uncertainty of MHOÜ and IHOÜ at aN^3LO .

4.2.4 N^3LO partonic cross sections

Determining PDFs at N^3LO not only requires the splitting functions but also hard cross-sections at the same perturbative order. Exact N^3LO massless DIS coefficient functions have been known for several years [160–163, 208, 209]. However, for massive coefficient functions, only various approximations are currently available [193, 194, 207]. For hadronic processes, N^3LO results are available for inclusive Drell-Yan production, including both the total cross-section [154, 169, 170] and distributions for rapidity [176] and

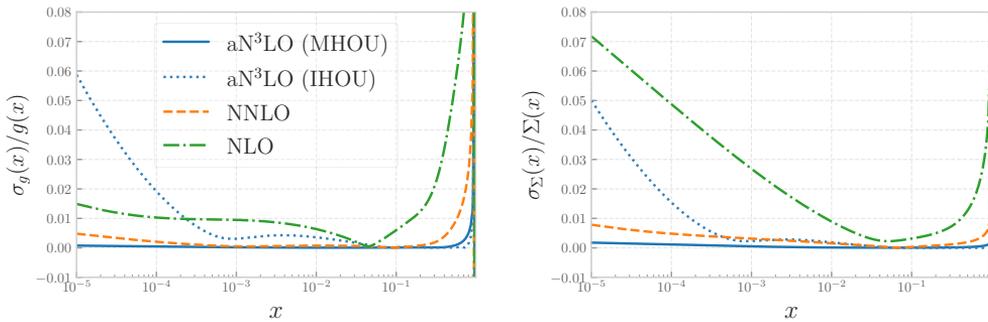


Figure 4.4: Relative uncertainty of the gluon and singlet PDFs (shown in Fig. 4.3). The MHO is shown in all cases, and at aN³LO the IHO is also shown.

transverse momentum [177], though these results have not been made publicly available.

We will now explain the implementation of these corrections. First, we review the available results on DIS coefficient functions and outline the key features of the approximations used for massive coefficient functions [193, 194]. Then, we address the N³LO corrections for hadronic processes and explore various approaches for incorporating these corrections into PDF determination.

N³LO corrections to massive DIS structure functions

The coefficient functions, assuming all quarks are massless, were computed at N³LO in [160, 161] for neutral-current charged-lepton scattering, with independent verification conducted recently in [209]. The analogous results for charged-current scattering can be found in [162, 163, 208]. The massive coefficient functions have been computed exactly up to NNLO for photon [210, 211], Z [212, 213], and W [214] exchange (for transitions from massless to massive states only). At N³LO, however, only partial results are available [194, 207, 215, 216], or in the limit where $Q^2 \gg m_h^2$ [195, 196, 198, 199, 201, 217].

We employ an approximation for the N³LO term $C_{i,k}^{(3)}(x, \alpha_s, m_h^2/Q^2)$ in the massive coefficient functions for photon-induced DIS, disregarding the axial-vector coupling of the Z boson and treating heavy quarks in the massless approximation for W boson exchange. This approximation, which is based on partial results, was initially introduced in Ref. [207] and revisited more recently in Ref. [193]. These references use the same exact results but differ in how they combine and interpolate them. Here, we follow the approach of Ref. [193], and also refer to Ref. [194] for additional details. Exact results are derived from threshold resummation and high-energy resummation, which are then combined with the asymptotic large- Q^2 limit. This ensures that the approximate massive coefficient function matches the exact massless result as $Q^2/m_h^2 \rightarrow \infty$. In the methodology of Refs. [193, 194], the massive coefficient functions are expressed as

$$C_{i,k}^{(3)}(x, m_h^2/Q^2) = C_{i,k}^{(3),\text{thr}}(x, m_h^2/Q^2)f_1(x) + C_{i,k}^{(3),\text{asy}}(x, m_h^2/Q^2)f_2(x), \quad (4.46)$$

where $C_{i,k}^{(3),\text{thr}}$ and $C_{i,k}^{(3),\text{asy}}$ represent the contributions from different resummation methods, and $f_1(x)$ and $f_2(x)$ are appropriate matching functions.

For massive quarks, the threshold limit is defined as $x \rightarrow x_{\text{max}}$, where $x_{\text{max}} = \frac{Q^2}{4m_h^2 + Q^2}$, or equivalently, $\beta \rightarrow 0$, with $\beta \equiv \sqrt{1 - \frac{4m_h^2}{s}}$ and $s = Q^2 \frac{1-x}{x}$ representing the

center-of-mass energy of the partonic cross-section. In this limit, the coefficient function includes logarithmically enhanced terms of the form $\alpha_s^n \ln^m \beta$, where $m \leq 2n$, due to soft gluon emissions, as predicted by threshold resummation [218]. Additionally, there are contributions of the form $\alpha_s^n \beta^{-m} \ln^l \beta$, with $m \leq n$, resulting from Coulomb exchange between the heavy quark and anti-quark, which can also be resummed using non-relativistic QCD methods [219]. At N³LO, all these contributions are known and can be extracted from existing resummed results [207]; they are incorporated into $C_{i,k}^{(3),\text{thr}}$.

In the high-energy limit, the coefficient function includes logarithmically enhanced terms of the form $\alpha_s^n \ln^m x$ where $m \leq n - 2$. These terms can be determined at all orders through small- x resummation at the leading logarithmic (LL) level [216], allowing for the extraction of the N³LO expansion [207]. This result can be further refined [193, 194] by incorporating a specific class of next-to-leading logarithmic (NLL) terms related to NLL perturbative evolution and the running of the coupling. In the approach described in Refs. [193, 194], the high-energy contributions are combined into $C_{i,k}^{(3),\text{asy}}$ along with the asymptotic $Q^2 \gg m_h^2$ limit of the coefficient function within the decoupling scheme [195, 196, 198, 199, 201, 217], while subtracting overlapping terms. This ensures that in the $Q^2 \gg m_h^2$ limit, the structure function computed from $C_{i,k}^{(3),\text{asy}}$, when combined with decoupling-scheme PDFs, matches the structure function obtained when the heavy quark mass is neglected and the heavy quark is treated as a massless parton. However, the asymptotic limit can only be determined approximately, as the term $a_{Qg}^{(3,0)}$ in the $A_{Qg}^{(3)}$ matching conditions was still unknown at the time the aN³LO NNPDF4.0 PDFs set was released. For this reason, we used an approximate result [207].

The interpolating functions used to combine the two contributions in Eq. (4.46) are designed to meet the following criteria:

$$\begin{aligned} f_1(x) &\xrightarrow{x \rightarrow 0} 0, & f_1(x) &\xrightarrow{x \rightarrow x_{\text{max}}} 1, \\ f_2(x) &\xrightarrow{x \rightarrow 0} 1, & f_1(x) &\xrightarrow{x \rightarrow x_{\text{max}}} 0, \end{aligned} \tag{4.47}$$

These conditions ensure that the threshold contribution disappears in the small- x limit, and vice versa. This setup guarantees that the approximation in Eq. (4.46) remains reliable across a broad kinematic range in the (x, Q^2) plane: $C_{i,k}^{(3),\text{asy}}$ accurately represents the massless limit for large Q^2 values and across all x , including the small- x limit, while $C_{i,k}^{(3),\text{thr}}$ captures the threshold limit, with x approaching x_{max} . The uncertainty of the approximate coefficient function can be assessed by varying the functional forms of the interpolating functions and the components that are not fully known. This includes the NLL small- x resummation and the matching functions involved in the asymptotic high Q^2 limit. This uncertainty diminishes as $x \rightarrow x_{\text{max}}$, where the exact known limit is recovered (with a fixed choice for the unknown constant β -independent terms), and increases in the intermediate η region. The interpolating functions and their uncertainties are optimized using the same approach at NNLO, where the full result is known. For a detailed discussion of this construction, refer to Ref. [193].

In order to obtain the best configuration for the parameters used in the $f_{1/2}$ functions, the approximation has been developed at NNLO, where the exact result is known. In this way we have found the parameters that best reproduced the exact result. After obtaining an approximation that described in an exhaustive way the known NNLO coefficient functions, Eq. (4.46) has been applied to the N³LO, with no modifications for the interpolating functions $f_{1/2}$.

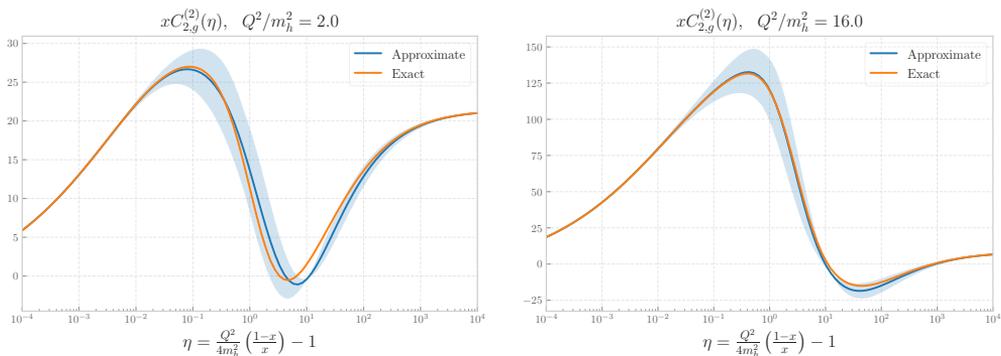


Figure 4.5: A comparison is made between the exact NNLO massive gluon-initiated coefficient function $xC_{2,g}^{(2)}(\eta)$ and the approximation given by Eq. (4.46). This comparison is plotted as a function of η , defined by Eq. (4.48), for fixed values of Q^2 . Results are presented for two different Q^2 values: one near the threshold at $Q^2 = 2m_h^2$ (left) and another at higher scales with $Q^2 = 16m_h^2$ (right). The uncertainty in the approximate result is derived from variations in the interpolating functions $f_1(x)$ and $f_2(x)$ used in Eq. (4.46).

This approximation at NNLO is illustrated in Fig. 4.5, where it is compared to the exact result for the massive gluon-initiated coefficient function $xC_{2,g}^{(2)}(\eta)$, expressed using the variable

$$\eta = \frac{Q^2(1-x)}{4m_h^2 x} - 1. \quad (4.48)$$

The comparison is presented for two different Q^2/m_h^2 ratios, one near the threshold and another at higher energy scales. It's important to note that as $\eta \rightarrow 0$, this corresponds to $x \rightarrow x_{\max}$ (the threshold limit), while $\eta \rightarrow \infty$ corresponds either to $Q^2/m_h^2 \rightarrow \infty$ for a fixed x (the asymptotic limit) or to $x \rightarrow 0$ for a fixed Q^2 (the high-energy limit). In this comparison, the uncertainty band is generated by varying only the interpolating functions.

The results obtained for the gluon and quark singlet coefficient functions at N³LO are shown in Fig. 4.6, where they are compared with the approximation from Ref. [207], each accompanied by their respective uncertainty estimates. There is a strong agreement between the different approximations, particularly for the dominant gluon coefficient function. The approximations align well in the asymptotic limits as $\eta \rightarrow 0$ and $\eta \rightarrow \infty$, and across most of the η range. However, they show some differences in the asymptotic large η region at fixed Q^2 , which corresponds to the small- x limit at fixed Q^2 . These discrepancies can be attributed to the inclusion of a specific class of NLL terms in the procedure used by Ref. [193, 194].

The uncertainty associated with the approximation can be incorporated as an additional IHOU alongside the one of the splitting functions, by adding a new component to the theory covariance matrix. Specifically, we define

$$\text{cov}_{mn}^C = \frac{1}{2} (\Delta_m(+)\Delta_n(+) + \Delta_m(-)\Delta_n(-)). \quad (4.49)$$

Here, $\Delta_m(\pm)$ represents the change in the prediction for the m -th DIS data point when substituting the central approximation of the massive coefficient function with the upper or lower bounds of the uncertainty range provided in Ref. [193] and illustrated as an

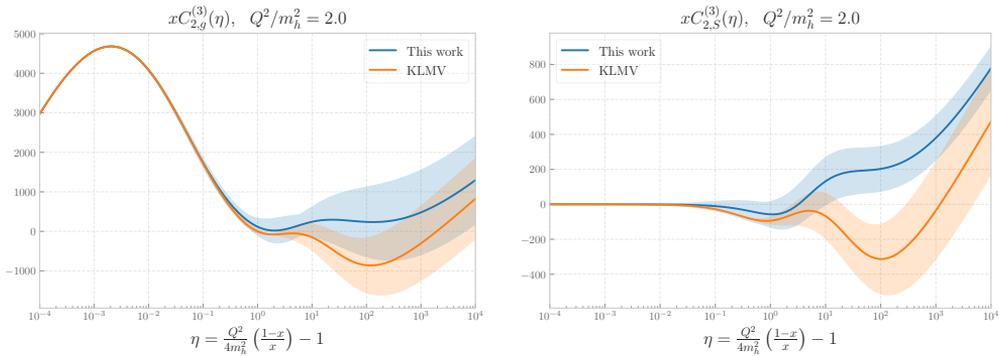


Figure 4.6: The approximate N³LO massive gluon (on the left) and quark singlet (on the right) coefficient functions are plotted as a function of η for a fixed $Q^2 = 2m_h^2$. Our results, based on the approximation from Ref. [193], are compared with those from the approximation provided by Ref. [207] (KLMV).

uncertainty band in Fig. 4.6. Unlike in Eq. (4.44), we divide by the number of independent variations without subtracting one, because the central value is not an average of the variations and is thus considered independent. This additional contribution, given by Eq. (4.49), is then included in the IHOU covariance matrix as an extra term on the right-hand side of Eq. (4.44).

The effect of this additional component on the IHOU is evaluated in Fig. 4.7, which displays the square root of the diagonal elements of the covariance matrix for all DIS data points in our dataset. This comparison shows the IHOU both before and after incorporating to Eq. (4.44) the extra term from Eq. (4.49) related to the IHOU of the massive coefficient function. It is evident that the influence of IHOUs due to perturbative evolution is generally minimal, consistent with the findings discussed in this section and depicted in Fig. 4.3. IHOUs related to splitting functions are only significant at small x , but since the available small- x data are at relatively low scales with short evolution lengths, their impact is limited. On the other hand, IHOUs affecting massive coefficient functions are more notable for data involving tagged bottom and charm structure functions. However, for other structure function data, the impact is moderate and mainly significant near the heavy quark production thresholds.

4.2.5 N³LO corrections to hadronic processes

N³LO corrections to the total cross-section for inclusive neutral-current (NC) and charged-current (CC) Drell-Yan production [169,170] are provided in the `n3loxS` public code [154]. These corrections are available both for on-shell W and Z bosons and as a function of the dilepton invariant mass $m_{\ell\ell}$. Differential distributions for leptonic observables in these processes have also been calculated [176,177], but they are not publicly accessible. Currently, no N³LO calculations are available for other processes covered in the NNPDF4.0 dataset.

The ratio of the NC total cross-section calculated at successive perturbative orders is displayed in Fig. 4.8. This ratio is evaluated using a fixed set of PDFs: NNPDF4.0 NNLO for comparing NNLO to NLO results, and NNPDF4.0 aN³LO for comparing N³LO to NNLO results. The results are presented in the high-mass region as a function of $m_{\ell\ell}$, with binning matching that of the ATLAS 7 TeV measurement [62]. The perturbative

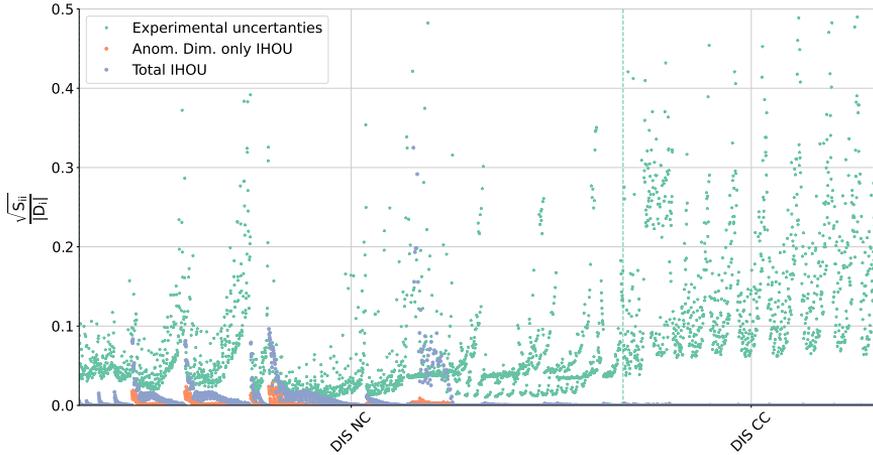


Figure 4.7: The square root of the diagonal elements of the IHOU covariance matrix for the DIS datasets, normalized to the experimental central value D_i , is presented. The IHOU values are shown both before and after incorporating the additional component from Eq. (4.44), which accounts for uncertainty related to the massive coefficient function, into the covariance matrix described by Eq. (4.49). For comparison, the experimental uncertainty is also displayed.

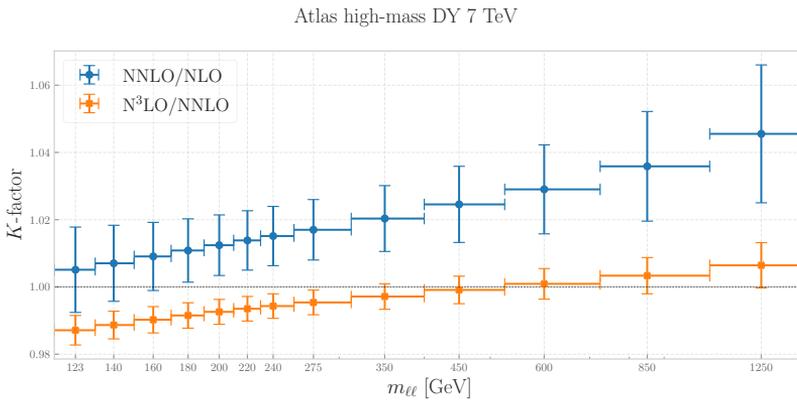


Figure 4.8: The ratio of the total neutral-current (NC) Drell-Yan cross-section is plotted as a function of the NNLO/NLO and N³LO/NNLO calculations for the inclusive NC Drell-Yan cross-section, binned by the invariant mass of the dilepton pair, $m_{\ell\ell}$. These ratios are computed using the `n3lox`s code and are integrated over all other kinematic variables. The $m_{\ell\ell}$ binning matches that of the ATLAS 7 TeV high-mass Drell-Yan measurement [62]. The same PDF set, NNPDF4.0 NNLO (for NNLO/NLO) and aN³LO (for N³LO/NNLO), is used for both the numerator and denominator. The vertical bands indicate the MHOU on the K -factors, derived from scale variations.

convergence is evident, with the N³LO/NNLO ratio being closer to unity and exhibiting less variation compared to the NNLO/NLO ratio. While NNLO corrections vary between +0.5% and +4%, the N³LO corrections are reduced to between -1.2% and +0.5%.

Total cross-section data are derived by extrapolating from measurements taken within a specific fiducial region. For NC Drell-Yan production in the central rapidity region and for dilepton invariant masses near the Z -peak, the N³LO/NNLO cross-section ratio

Dataset	Ref.	n_{dat}	Kin ₁	Kin ₂ [GeV]	C -factor N ³ LO/NNLO
ATLAS high-mass DY 7 TeV	[62]	13	$ \eta_\ell \leq 2.1$	$116 \leq m_{\ell\ell} \leq 1500$	$\frac{d\sigma}{dm_{\ell\ell}}$
ATLAS Z 7 TeV ($\mathcal{L} = 35 \text{ pb}^{-1}$)	[63]	8	$ \eta_\ell, y_Z \leq 3.2$	$Q = m_Z$	$\frac{d\sigma}{dm_{\ell\ell}}$ ($66 < m_{\ell\ell} < 150$)
ATLAS Z 7 TeV ($\mathcal{L} = 4.6 \text{ fb}^{-1}$) CC	[64]	24	$ \eta_\ell, y_Z \leq 2.5, 3.6$	$Q = m_Z$	$\frac{d\sigma}{dm_{\ell\ell}}$ ($46 < m_{\ell\ell} < 116$)
ATLAS $\sigma_{W,Z}^{\text{tot}}$ 13 TeV	[77]	3	—	$Q = m_W, m_Z$	σ

Table 4.3: The LHC NC DY production datasets in the NNPDF4.0 framework for which an N³LO K -factor has been incorporated in a variant of the standard aN³LO PDF determination are listed here. For each dataset, we provide the published references, the number of data points, and the relevant kinematic variables.

shows only a minor dependence on the dilepton rapidity $y_{\ell\ell}$ [176, 177]. However, it is uncertain whether this holds true for off-peak or extremely high or low rapidities. As a result, the inclusion of N³LO corrections for hadronic processes is not fully reliable at this time. Therefore, these corrections are not incorporated into our primary analysis but are considered in a specialized variant to evaluate their impact.

The datasets for which N³LO corrections have been applied in this variant are detailed in Tab. 4.3. This includes the high-mass NC cross-section, the Z rapidity distribution in the central region for on-shell Z -production, and the total W and Z cross-sections. For each of these processes, the N³LO cross-section is calculated by multiplying the NNLO result by a K -factor derived using a fixed underlying PDF set, specifically the aN³LO NNPDF4.0 PDF set, which will be discussed later. For the rapidity distribution, we use the same fixed K -factor as applied to the total cross-section. We exclude off-shell or double-differential rapidity distributions (such as those from CMS), off-forward rapidity distributions (such as those from LHCb), and low-mass total cross-sections, as the assumption of a constant K -factor or simple fiducial extrapolation is less reliable for these cases. The datasets are identified as in Tab. 2.4 of Ref. [4]².

Although we are currently unable to reliably calculate N³LO corrections for the existing LHC measurements, we aim to incorporate the complete NNPDF4.0 dataset into our aN³LO PDF determination. To achieve this, we apply an additional uncertainty to all data where N³LO corrections are not included. This uncertainty is similar to the MHOU used in the NNPDF4.0MHOU determination described in Sec. 4.1.

Thus, when N³LO corrections to the hard cross-section are not included, the theoretical prediction is obtained by combining aN³LO evolution with NNLO cross-sections. This prediction is then enhanced with a theory covariance matrix, calculated by varying the renormalization scale μ_R using a three-point method as outlined in Refs. [156, 157]:

$$\text{cov}_{mn}^{\text{NNLO}} = \frac{1}{2} (\Delta_m(+)\Delta_n(+) + \Delta_m(-)\Delta_n(-)) , \quad (4.50)$$

This is similar to Eq. (4.49), but here, $\Delta_m(\pm)$ represents the change in the prediction for the m -th data point when the coefficient functions are modified by varying the renor-

²Note that the number of data points for rapidity distributions differs here because we only consider Z distributions.

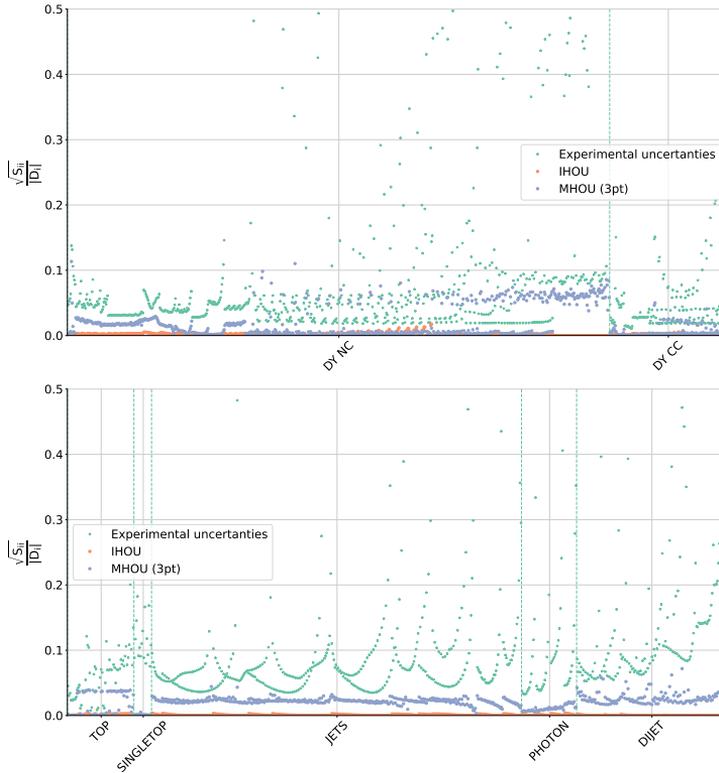


Figure 4.9: Similar to Fig. 4.7, this figure compares the IHOUs from Eq. (4.44) with the MHOUs from Eq. (4.50), which accounts for the missing N³LO correction to the matrix element. The results are presented for all hadronic datasets in the NNPDF4.0 collection: specifically, Drell-Yan (top), as well as top pair, single top, single-inclusive jet, prompt photon, and dijet production (bottom).

malization scale up or down, as described in Ref. [157] and implemented in Ref. [7], Eq. (2.9). The resulting MHOUs covariance matrix is then added to the IHOUs covariance matrix as an additional term of the theory covariance matrix. Therefore, putting together Eq. (4.44), Eq. (4.49) and Eq. (4.50), we find

$$\text{cov}_{mn}^{\text{IHOUs}} = \text{cov}_{mn}^{(gg)} + \text{cov}_{mn}^{(gq)} + \text{cov}_{mn}^{(qg)} + \text{cov}_{mn}^{(qq)} + \text{cov}_{mn}^{\text{C}} + \text{cov}_{mn}^{\text{NNLO}}. \quad (4.51)$$

The effect of this hadronic uncertainty is illustrated in Fig. 4.9, where we present, for all hadronic datasets, the square root of the diagonal entries of the MHOUs covariance matrix in Eq. (4.50). This is compared with the IHOUs covariance matrix Eq. (4.44) and the experimental uncertainties, all normalized to the central theory prediction. The MHOUs are typically greater than the IHOUs, suggesting that the impact of missing N³LO terms in the hard cross-sections is more significant than the IHOUs uncertainty related to N³LO perturbative evolution. Experimental uncertainties are generally even larger.

In addition to the NNPDF4.0 aN³LO baseline PDF set derived in this manner, we will also create an NNPDF4.0 MHOUs aN³LO set, following the approach used for the NLO and NNLO MHOUs sets recently described in Sec. 4.1, see also Ref. [7]. This set will incorporate MHOUs for both perturbative evolution and hard matrix elements, using the methods outlined in Refs. [156, 157]. We will determine the theory covariance matrix by performing combined correlated variations of renormalization and factorization

scales with a 7-point prescription, as detailed in Ref. [7]. Specifically, scale variations will be conducted at the perturbative order at which the expressions are computed, namely, aN^3LO for anomalous dimensions and DIS coefficient functions, and NNLO for hadronic processes. This approach inherently produces larger scale variations and appropriately reduces the weight of processes lacking N^3LO corrections. The ability to include processes with different perturbative orders in a single PDF determination highlights the benefit of incorporating MHOUs, as noted in Refs. [220, 221].

4.2.6 NNPDF4.0 at aN^3LO

We now introduce the aN^3LO NNPDF4.0 PDF sets. These have been developed using the dataset and methodology outlined in Ref. [4], which was employed for the LO, NLO, and NNLO NNPDF4.0 sets presented in that work, but now extended to aN^3LO . The aN^3LO results are derived using the approximate N^3LO contributions that we have discussed in this section, i.e. the splitting functions, the exact massless and approximate massive N^3LO coefficient functions and NNLO hadronic cross-sections with an added uncertainty. Moreover, similarly to the NNPDF4.0QED and the NNPDF4.0MHOU PDFs determinations, this work used the new theory pipeline described in Sec. 2.3 and more in detail in Ref. [114].

Alongside the standard NNPDF4.0 aN^3LO PDF set, we also introduce a variant that incorporates MHOUs for all theoretical predictions involved in the PDF determination. This variant is produced using the same approach described in Sec. 4.1 that is used to develop the NNPDF4.0MHOU PDFs set [7]. To evaluate perturbative convergence and the effects of MHOUs, we will also provide an NNPDF4.0MHOU NLO PDF set, constructed with the same methodology and dataset as the default NNPDF4.0 NLO PDF set (distinct from the NNPDF4.0 NNLO dataset).

Tab. 4.4 present the number of data points and the χ^2 per data point for the NLO, NNLO, and aN^3LO NNPDF4.0 fits, both with and without MHOUs. It organizes the datasets based on the process classification outlined in Ref. [7]. Additionally, Fig. 4.10 illustrates the total χ^2 per data point as a function of perturbative order.

The NLO and NNLO results without MHOUs are derived from the NLO and NNLO NNPDF4.0 PDF sets. For the NNLO results with MHOUs, the NNPDF4.0MHOU NNLO set from Ref. [7] is used. The NNPDF4.0MHOU NLO set, introduced here for the first time, applies the same methodology as the NNPDF4.0MHOU NNLO set [7], but uses the dataset from the NNPDF4.0 NLO [4]. Therefore, while the datasets with and without MHOUs remain consistent, the NLO and NNLO datasets differ according to Ref. [4]. The N^3LO dataset is the same of the NNLO dataset. In all instances, the theoretical predictions used for computing the χ^2 are generated with the updated theory pipeline. The covariance matrix is calculated as outlined in Sec. 4.1. The N^3LO predictions are based on the aN^3LO PDF sets, which utilize the same datasets and kinematic cuts as the NNPDF4.0 NNLO PDF sets, and incorporate theoretical predictions discussed in this section. These predictions are further enhanced with an IHOU covariance matrix as described in detail, and a MHOU for hadronic processes where N^3LO hard cross-sections are not available, as mentioned previously.

Tab. 4.4 and Fig. 4.10 reveal that the fit quality improves as the perturbative order increases, even when MHOUs are not included. This trend continues from NNLO to N^3LO , even though N^3LO corrections are only partially implemented, with hadronic matrix elements still computed at NNLO. This indicates that the inclusion of N^3LO corrections to evolution and DIS coefficient functions has a significant effect on fit quality, aligning with the anticipated improvement associated with adding an additional per-

Dataset	N_{dat}	NLO		N_{dat}	NNLO		N_{dat}	aN ³ LO	
		no MHOUs	MHOUs		no MHOUs	MHOUs		no MHOUs	MHOUs
DIS NC	1980	1.30	1.22	2100	1.22	1.20	2100	1.22	1.20
DIS CC	988	0.92	0.87	989	0.90	0.90	989	0.91	0.92
DY NC	667	1.49	1.32	736	1.20	1.15	736	1.17	1.16
DY CC	193	1.31	1.27	157	1.45	1.37	157	1.37	1.36
Top pairs	64	1.90	1.24	64	1.27	1.43	64	1.23	1.41
Single-inclusive jets	356	0.86	0.82	356	0.94	0.81	356	0.84	0.83
Dijets	144	1.55	1.81	144	2.01	1.71	144	1.78	1.67
Prompt photons	53	0.58	0.47	53	0.76	0.67	53	0.72	0.68
Single top	17	0.35	0.34	17	0.36	0.38	17	0.35	0.36
Total	4462	1.24	1.16	4616	1.17	1.13	4616	1.15	1.14

Table 4.4: The tables show the count of data points and the χ^2 per data point for the NLO, NNLO, and aN³LO NNPDF4.0 fits, both with and without MHOUs. The datasets are organized according to the process classification method employed in Ref. [7].

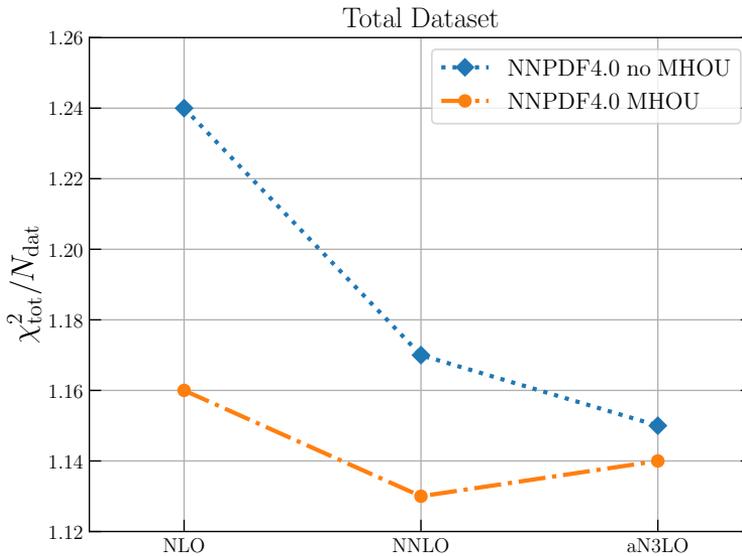


Figure 4.10: The total χ^2 per data point for the NNPDF4.0 fits at NLO, NNLO, and aN³LO, both with and without MHOUs.

turbative order, similar to the enhancement observed when progressing from NLO to NNLO.

On the other hand, when MHOUs are included, the fit quality becomes independent of the perturbative order within uncertainties (noting that, with $N_{\text{dat}} = 4462$, $\sigma_{\chi^2} = 0.03$). This indicates that the MHOUs covariance matrix estimated through scale variation is accurately capturing the observed shifts between perturbative orders, reflecting the true MHOUs. If this is correct, it also implies that at aN³LO, the missing

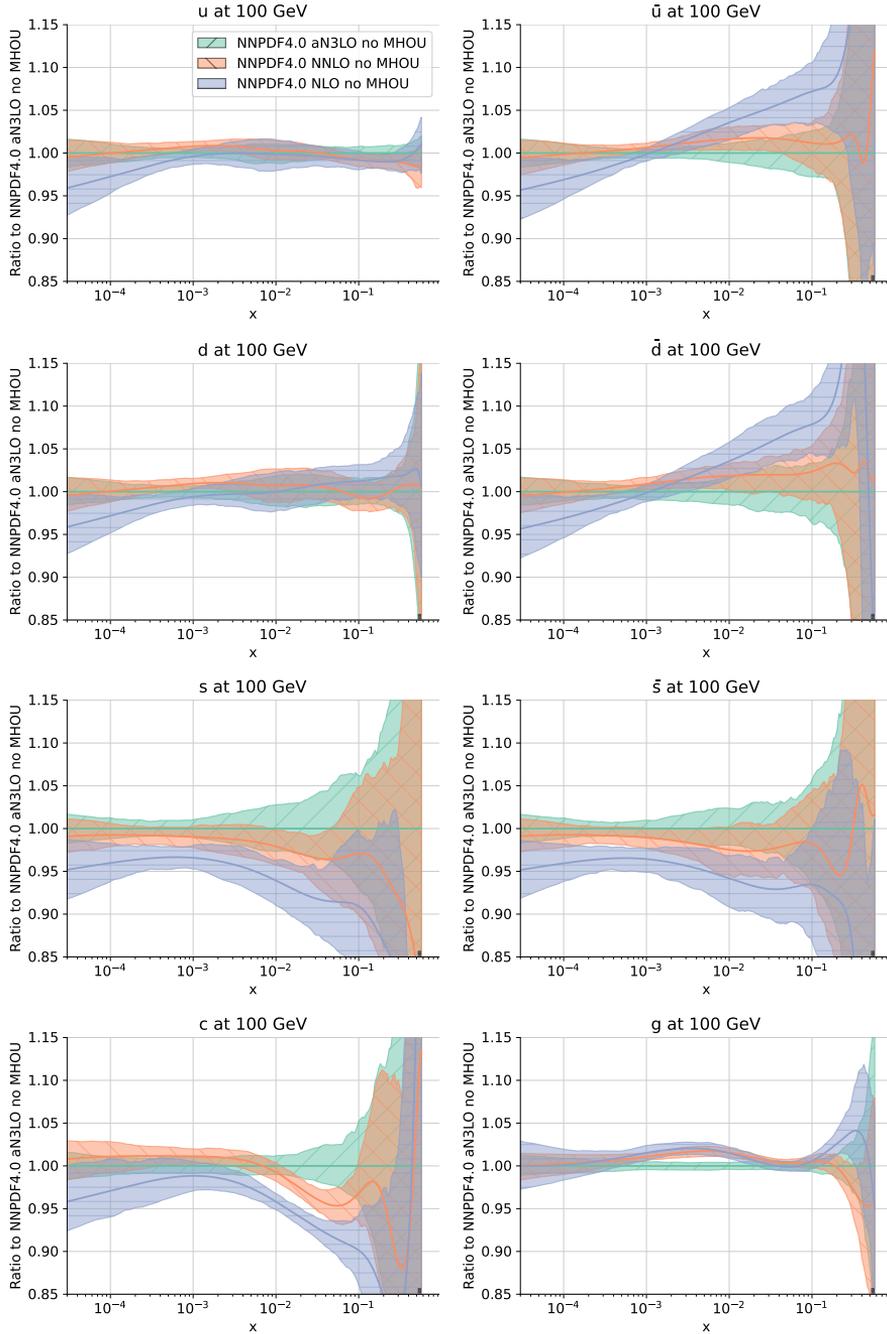


Figure 4.11: The NLO, NNLO, and aN^3LO NNPDF4.0 PDFs at $Q = 100$ GeV are shown. We present the up, anti-up, down, anti-down, strange, anti-strange, charm, and gluon PDFs, all normalized to the aN^3LO result. The error bands represent one sigma PDF uncertainties and do not include MHOUs on the theoretical predictions used in the fit.

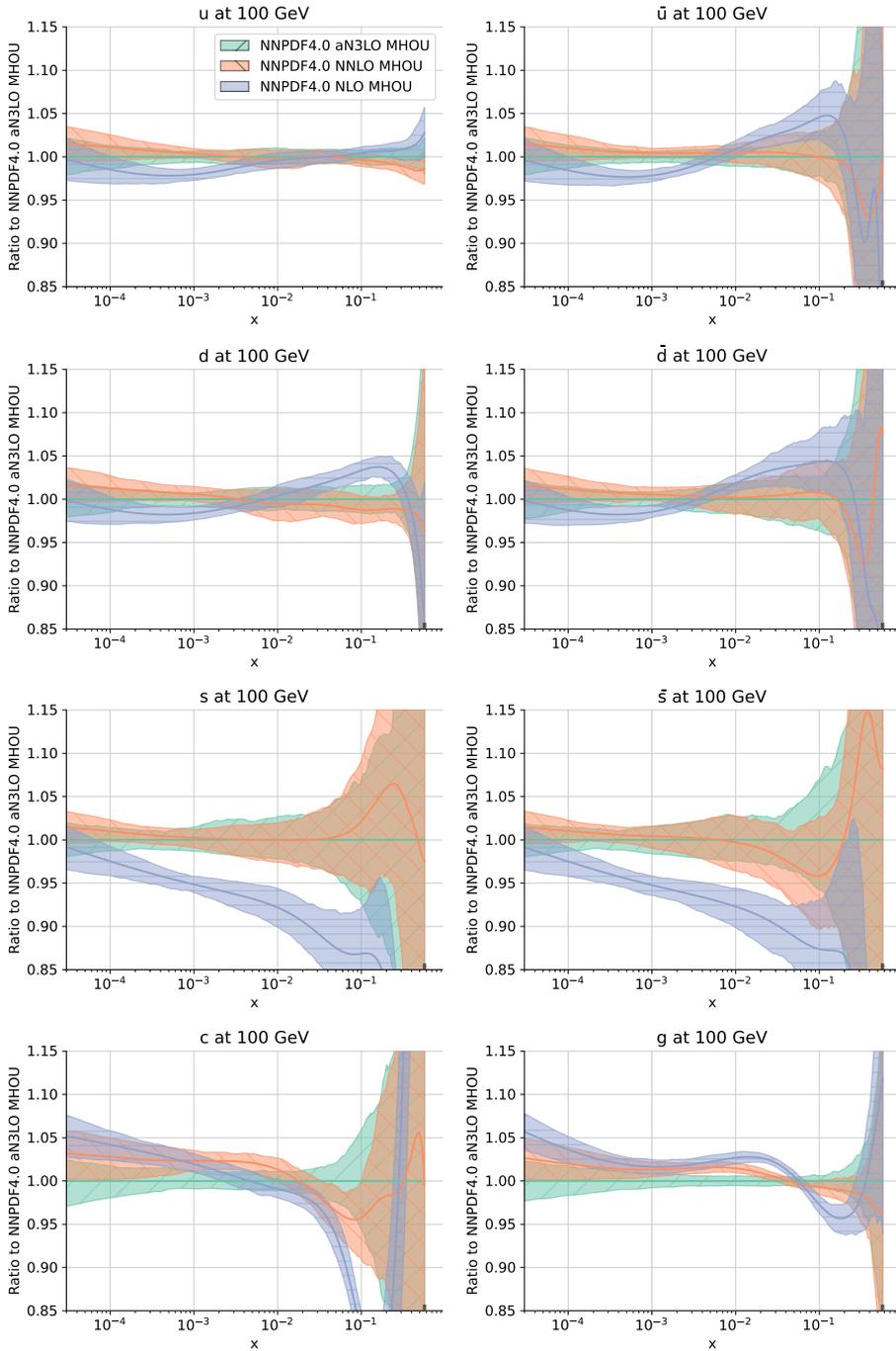


Figure 4.12: This figure is similar to Fig. 4.11, but it depicts the NNPDF4.0MHOUs PDF sets. The error bands represent one sigma PDF uncertainties, including MHOUs on the theoretical predictions used in the fit.

N^3 LO corrections to hadronic processes are effectively accounted for by the corresponding MHOU, which is consistently included. Additionally, at aN^3 LO, the fit quality remains consistent within uncertainties regardless of whether MHOUs are included. This strongly suggests that the inclusion of higher-order terms in perturbative evolution and DIS coefficient functions would not yield further improvements, meaning that with current experimental uncertainties, methodology, and dataset, the perturbative expansion has likely converged.

Let's now analyze the NNPDF4.0 aN^3 LO parton distribution functions (PDFs). The NLO, NNLO, and aN^3 LO NNPDF4.0 PDFs, both with and without MHOUs, are compared in Figs. 4.11-4.12, respectively. These figures display the up, anti-up, down, anti-down, strange, anti-strange, charm, and gluon PDFs at $Q = 100\text{GeV}$, normalized to the aN^3 LO results, as a function of x on both logarithmic and linear scales. The error bands represent one sigma PDF uncertainties, which include MHOUs for the MHOU sets and exclude them for the non-MHOU sets. The same PDF sets, with and without MHOUs, were used to compute the χ^2 values presented in Tab. 4.4.

The strong perturbative convergence observed in the fit quality is also evident in the PDFs. Notably, the NNLO PDFs are either very similar to or nearly indistinguishable from their aN^3 LO counterparts. Including MHOUs enhances the alignment between NNLO and aN^3 LO PDFs, bringing them almost into perfect agreement. This indicates that the NNLO PDFs become more precise with the inclusion of MHOUs, and the aN^3 LO PDFs have effectively converged, as previously discussed. However, there are some exceptions to this stability. The charm and gluon PDFs show significant changes due to aN^3 LO corrections. For the charm PDF, these corrections increase the central value by approximately 4% around $x \sim 0.05$, while for the gluon PDF, there is a reduction of about 2-3% around $x \sim 0.005$. Additionally, the inclusion of MHOUs results in an increase in PDF uncertainties by about 1-2% for both the charm and gluon PDFs. This adjustment makes the NNLO and aN^3 LO charm PDFs with MHOUs consistent within uncertainties, and the NNLO and aN^3 LO gluon PDFs with MHOUs nearly consistent.

Overall, these findings indicate that aN^3 LO corrections are typically minimal, with the notable exception of the gluon PDF. At the aN^3 LO level, the perturbative expansion has nearly reached convergence, as evidenced by the close agreement between NNLO and aN^3 LO PDFs, particularly when MHOUs are included. Additionally, MHOUs generally enhance the precision of PDFs, although their effect at aN^3 LO is quite limited. The implications of these results for phenomenology discussed in detail in Ref. [8].

4.3 Combination

In the past two sections, we discussed two theoretical developments to PDFs determination: including MHOU and approximate N^3 LO corrections. Moreover, we discussed the combination of the two effects. In this section we will discuss the combination of these two effects with the inclusion of QED corrections, that we described in detail in Chap. 3. Indeed, considering that the three theoretical advancements mentioned earlier contribute independently to improving the accuracy of the extraction of the PDFs, it makes sense to integrate them into a unified global PDF determination, following the approach recently undertaken by the MSHT group [190,222].

Taking advantage of the adaptability offered by the new NNPDF theory pipeline, we introduce for the first time variants of the NNPDF4.0 global analysis that integrate these three theoretical improvements in different combinations. Specifically, we have developed a version of NNPDF4.0QED NNLO incorporating MHOUs, as well as two versions of NNPDF4.0 at approximate N^3 LO with QED corrections, one with MHOUs and one

without. It's important to note that, as discussed in Sec. 4.2, our aN³LO fits consistently apply 3-point MHOU to the hard-scattering cross-sections of hadronic processes to account for the absent N³LO contributions.

First of all we will assess the impact of the QED effects on the fit including MHOU. Then we will discuss inclusion of QED corrections to the N³LO fit. In this case we will study both the case in which no MHOU is included in the fit, but also the case in which a MHOU is considered. In the end we will compare the photon obtained in these three QED fits with the NNPDF4.0QED fit, to assess how much these new effects impact the photon PDF obtained with the LuxQED method.

4.3.1 MHOU+QED

In order to produce the NNPDF4.0QED+MHOU fit it has been used the methodology adopted for the QED fit described in Chap. 3, using the MHOU theory covariance matrix described in Sec. 4.1. However, a remark is now needed. In order to perform such fit, it has been used the theory covariance matrix used in the pure QCD PDFs determination with MHOU, i.e. the one used for the NNPDF4.0MHOU fit. This leads to two small inconsistencies. The first one is that such theory covariance matrix has been computed using a pure QCD theory. However, as we already mentioned, the NNPDF4.0 dataset has been chosen in such a way that QED corrections to the theory predictions of the data points are completely negligible. Therefore, the error we are introducing in this way is very small. The second inconsistency is related to the fact that the theory covariance matrix used in the NNPDF4.0MHOU PDFs extraction uses the truncated solution of the DGLAP equations. Instead, in the combination of QED and MHOU, we use the exact solution, for the reasons that we are explained in Chap. 3. However, also in this case we have seen that the differences between the two solutions are mostly at the permille level at NNLO, with some data points that can reach at most the percent differences.

Fig. 4.13 shows the comparison for the quarks and gluon PDFs between the pure QCD fit that includes MHOU and its QED variant. The PDFs are shown at the scale $Q = 100$ GeV. We can observe that adding QED corrections to the MHOU PDFs fit gives a very small correction to the quark PDFs. The gluon shows a larger difference at large x values, but this can be explained with the fact that we are comparing two fits that use two different solution method for the DGLAP equations, as it happened in the comparison between NNPDF4.0 and NNPDF4.0QED. Moreover, we can observe that the difference between truncated and exact solutions (that is mostly visible in the gluon) are smaller with respect to Fig. 3.9. This is because of the fact that the inclusion of MHOU improves the perturbative convergence of the series.

We conclude that adding QED effects to the PDFs fit that include MHOU gives a small correction to the quarks and to the gluon since the photon PDF is very small compared to the other distributions. This is exactly what we found in Sec. 3.4: adding the QED corrections to a pure QCD fit doesn't affect much the quarks and gluon PDFs because of the moderate size of the photon PDF.

4.3.2 aN³LO+QED

Now we will consider the combination of the approximate N³LO and QED effects and its impact to the NNPDF4.0 methodology. As we already mentioned, in this case we will consider two cases. In the first one we will not include MHOU in the fit. It means that we will include a theory covariance matrix that is given only by the MHOU and by the 3-point MHOU of the hadronic predictions whose N³LO correction is still missing, as explained

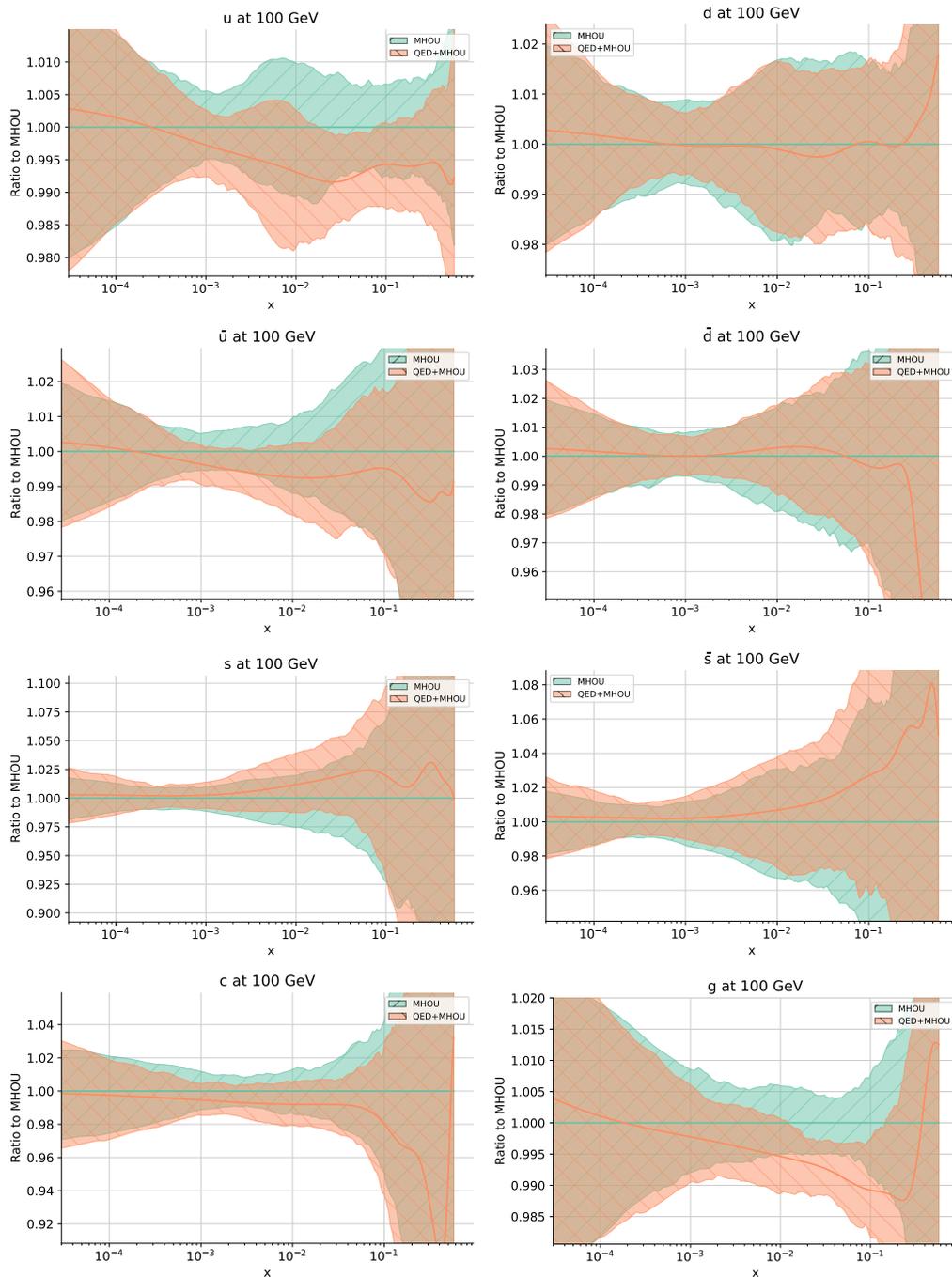


Figure 4.13: Comparison for quarks and gluon PDFs between the MHOUs fits with and without QED effects, as a ratio plot with respect to the second. The uncertainty band corresponds to the 1σ confidence level. The PDFs are plotted at the scale $Q = 100$ GeV.

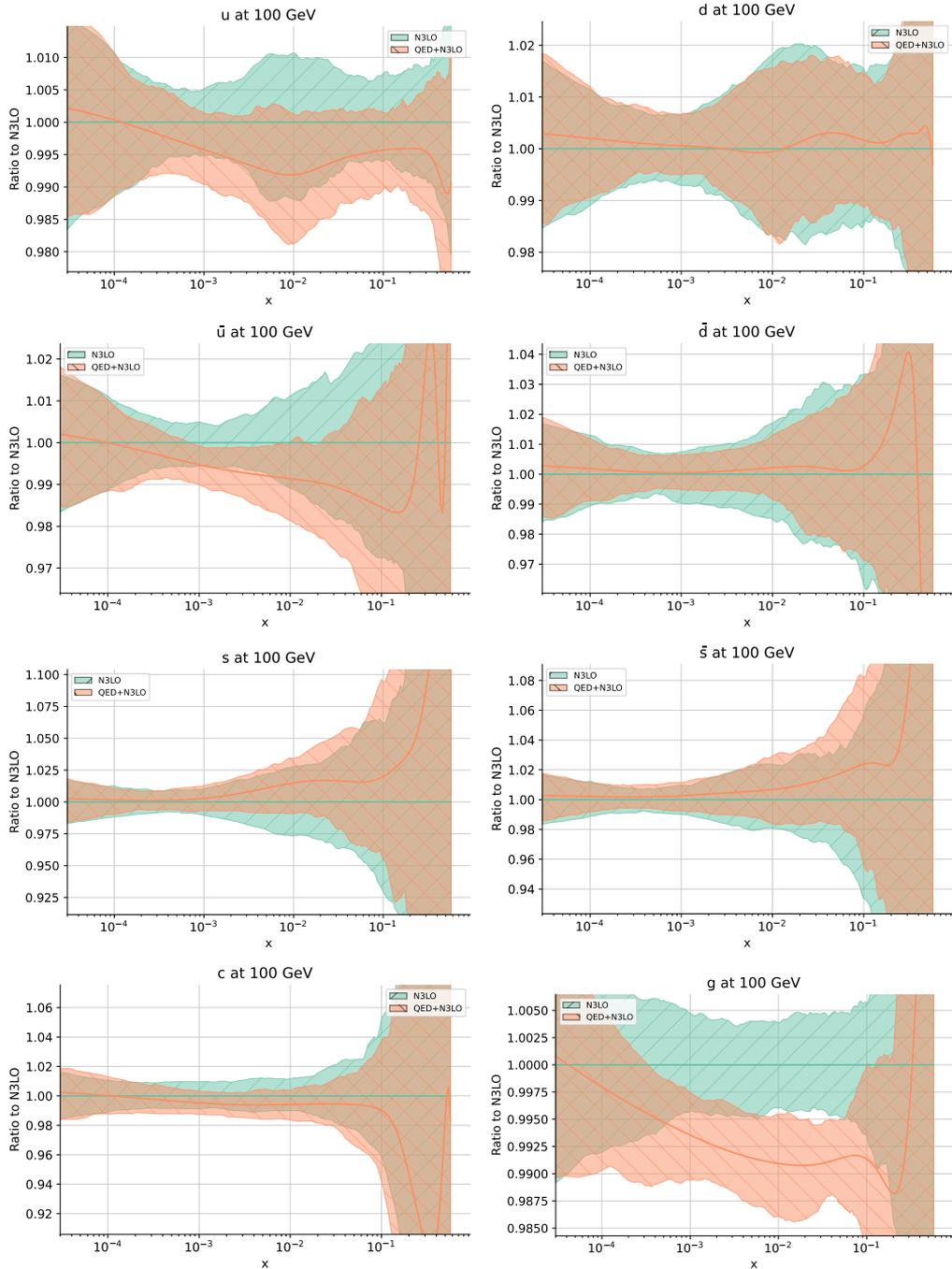


Figure 4.14: Comparison for quarks and gluon PDFs between the N³LO fits with and without QED effects, as a ratio plot with respect to the second. The uncertainty band corresponds to the 1σ confidence level. The PDFs are plotted at the scale $Q = 100$ GeV.

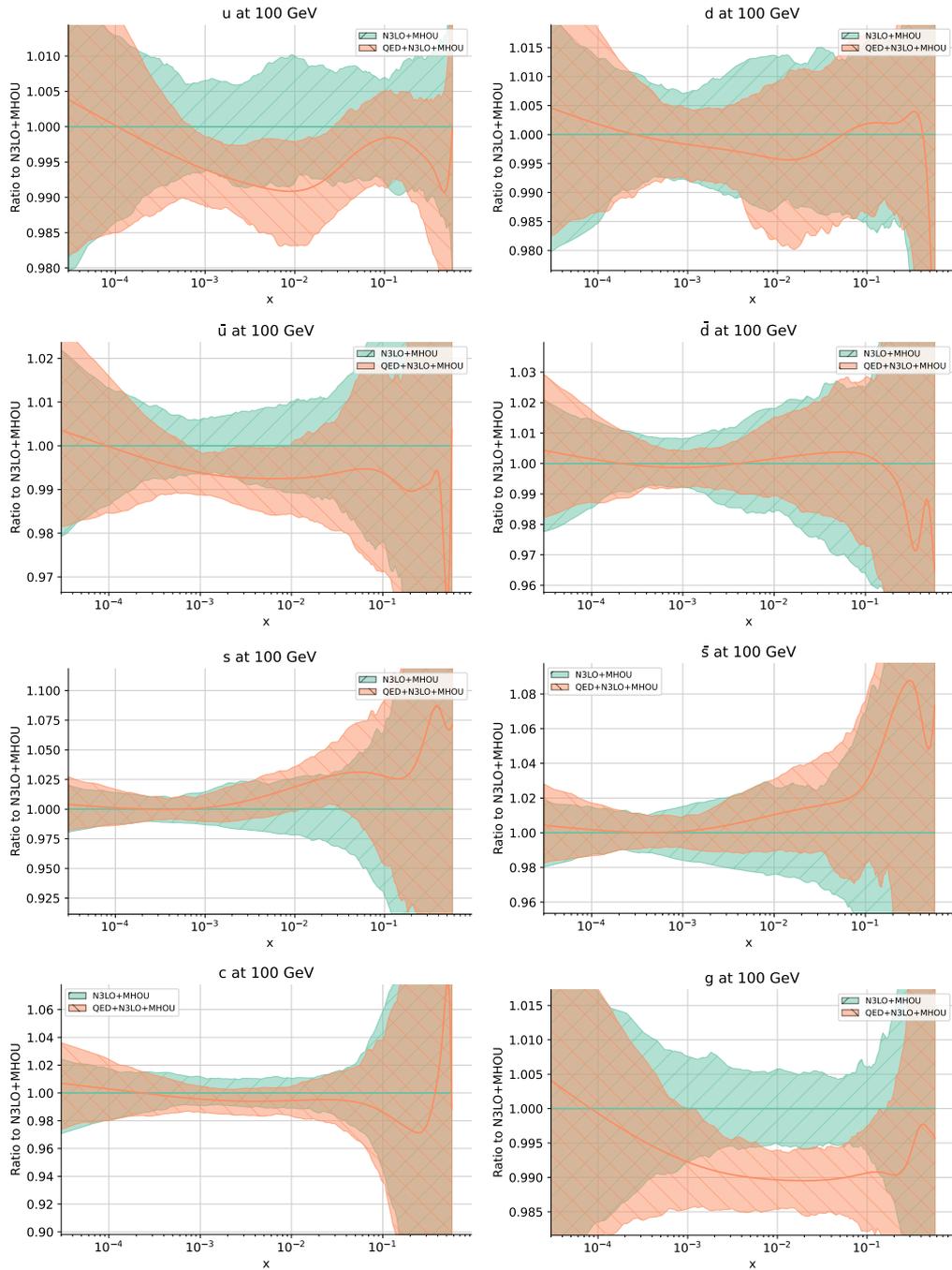


Figure 4.15: Comparison for quarks and gluon PDFs between the N³LO fits including MHOu with and without QED effects, as a ratio plot with respect to the second. The uncertainty band corresponds to the 1σ confidence level. The PDFs are plotted at the scale $Q = 100$ GeV.

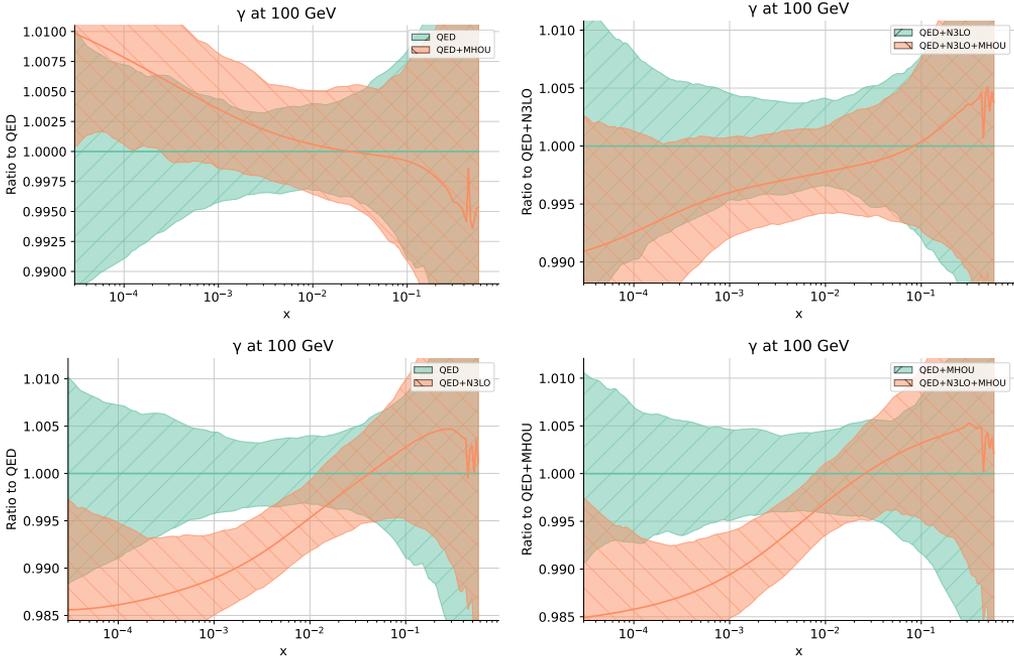


Figure 4.16: Comparisons of the photon PDFs obtained in the different cases. QED compared with QED+MHOU at NNLO (up left) and at N³LO (up right), QED at NNLO compared with QED at N³LO without MHOU (down left) and with MHOU (down right). The plots are shown as ratio plots and the PDFs are plotted at 100 GeV. The uncertainty band corresponds to the 1 σ confidence level.

in Sec. 4.2. This fit will be addressed as the QED+N³LO fit. In the second case instead, we will include also the 7-point MHOU in the N³LO fit. It means that the theory covariance matrix will include also a contribution given by the MHOU estimated from the 7-point scale variations of all the theory predictions. As in the previous case, both the IHOU and the MHOU covariance matrices, will be the ones used in the aN³LO NNPDF4.0 fit, that is a pure QCD fit that uses the truncated solution for the DGLAP equations. Also in this case, this leads to an inconsistency whose effect is completely negligible since QED effects on theory predictions are very small and the difference between truncated and exact solution at N³LO is even smaller than the one at NNLO.

Figs. 4.14-4.15 show the comparison between the N³LO fits with and without QED effects, first in the case in which MHOU are not considered and then in the case in which they are considered. The quarks and gluon PDFs are shown at the scale $Q = 100$ GeV. We can observe good agreement between the pure QCD fits and their QED version. The only exception, also in this case, is given by the gluon PDF at x values $10^{-2} \lesssim x \lesssim 10^{-1}$. Again, this is explained with the fact that we are comparing two fits that use two different solution methods for DGLAP evolution. However, if we compare the NNLO results with the N³LO one without MHOU, i.e. we compare Fig. 3.9 with Fig. 4.14, we can see that the difference between exact and truncated solution is smaller at N³LO than at NNLO, as we expected since we said that the two solutions are equivalent up to higher orders corrections and therefore it decreases as we go up with the perturbative expansion. Also in this case we can observe that in the case in which MHOU is considered, the

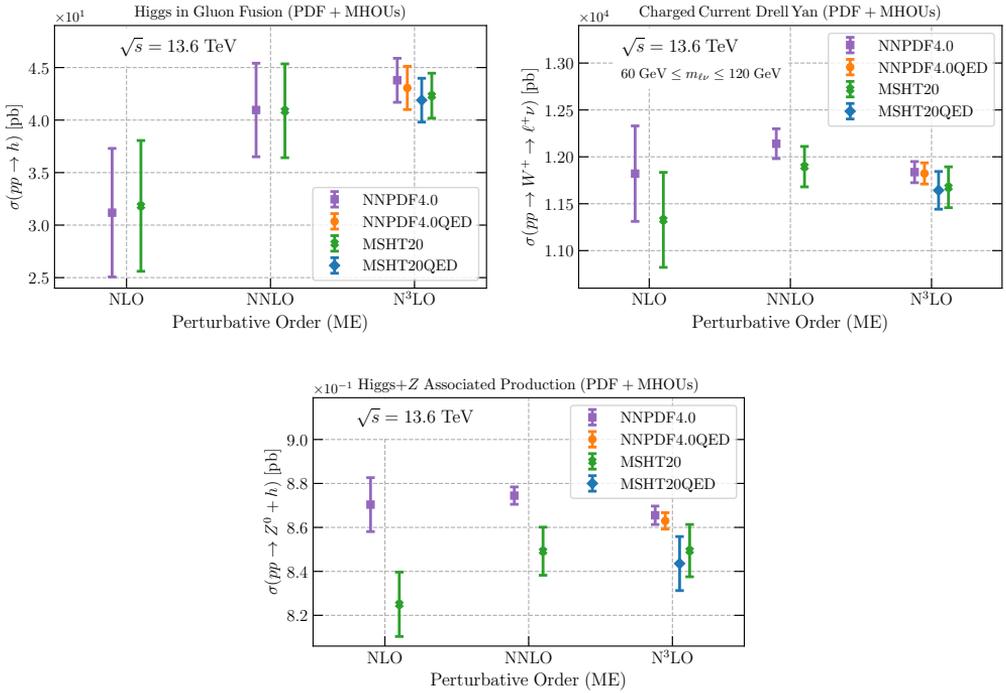


Figure 4.17: The perturbative convergence of Higgs production through gluon fusion and in association with Z bosons, as well as W production at the LHC with a center-of-mass energy of $\sqrt{s} = 13.6$ TeV, is analyzed. For the $N^3\text{LO}$ calculations, we also present predictions using the QED variants of the NNPDF4.0 and MSHT20 $aN^3\text{LO}$ sets.

differences between exact and truncated solutions decrease since the inclusion of MHOUs improves the perturbative convergence. We conclude that, also in the case of the $aN^3\text{LO}$ PDFs determination, the QED effects give a small correction to the pure QCD fit, because of the small impact of the photon PDF.

We also compare the photons obtained in the four different cases, i.e. in the QED, QED+MHOUs, QED+ $N^3\text{LO}$ and QED+ $N^3\text{LO}$ +MHOUs fits. In the two upper plots of Fig. 4.16 we can assess the impact of the MHOUs to the photon PDF both at NNLO (left) and at $N^3\text{LO}$ (right). We can observe that it gives a very small correction that is at per-mille level for almost all values of x with the exception of $x \simeq 10^{-5}$ where the difference can reach the percent level both at NNLO and at $N^3\text{LO}$. This difference is mostly given by the change in the shape of the quark and gluon PDFs when we include MHOUs, that is propagated into the photon PDF through the LuxQED formula, Eq. (3.1). In the two lower plots of Fig. 4.16 we can assess the impact of the inclusion of approximate $N^3\text{LO}$ corrections. In this case the differences are larger with respect to the previous case. Indeed, for $x \simeq 10^{-5}$ the photons have a difference of about 1.5% both with and without MHOUs. The reason is twofold: first of all the quarks and gluon PDFs used in the construction of the photon PDF at $N^3\text{LO}$ are slightly different with respect to the ones at NNLO; then the DIS coefficient functions at $N^3\text{LO}$ are modified by the presence of the approximate $N^3\text{LO}$ terms that we discussed in Sec. 4.2. So, the overall effect is the one shown in the lower plots of Fig. 4.16. Despite this difference at small x , the two comparisons show that for $x \gtrsim 10^{-2}$ the photon PDF at $N^3\text{LO}$ agrees with the one at NNLO

within the uncertainty band, both with and without MHOU.

In conclusion, we can assess the impact of the combination of aN³LO and QED corrections on the LHC phenomenology. Fig. 4.17 shows the perturbative convergence of the inclusive cross-section of three representative processes of the LHC physics at $\sqrt{s} = 13.6$ TeV. We present Higgs production through gluon fusion, associated production with Z bosons, and W^+ production at NLO, NLO, and N³LO accuracy levels. For the N³LO computations, we also provide predictions using the QED variants of the NNPDF4.0 and MSHT20 aN³LO sets [222]. This comparison reveals that incorporating QED corrections into the aN³LO sets results in a similar qualitative effect in both determinations, specifically, a moderate decrease in the central cross-section values, which remains comfortably within the theoretical uncertainty range.

Part III
Epilogue

Applications to Monte Carlo event generators

“Yeah man, they call gambling a disease, but it’s the only disease where you can win a bunch of money.”

– Norm McDonald

In this chapter we will describe how the implementation of the new theory pipeline, described in Sec. 2.3 and the implementation of QED corrections in the NNPDF4.0 fitting framework, described in Sec. 3, were applied for the construction of a PDFs set suitable for the use of Monte Carlo (MC) event generators. Indeed, MC event generators require PDFs to satisfy additional constraints than standard PDFs sets. Such constraints include PDFs positivity down to a very low scale as $Q \sim 1$ GeV, smooth extrapolation in the small x and large x regions, and numerically stable results also in extreme regions of the phase space for all PDFs.

MC event generators [223–226] give a complete description of the final state in particle collisions, and therefore are an essential ingredient for the interpretation of particle physics experiments. The most used event generators for LHC physics are `Pythia8` [227, 228], `HERWIG7` [229, 230], `SHERPA` [231, 232], `POWHEG` [233], `mg5_aMC@NLO` [234], and more recently `PanScales` [235–238].

In an MC event generator, PDFs [239, 240] are used not only for calculating hadronic cross-sections through their convolution with partonic matrix elements, but also for initiating backward parton showers in the initial state and for inputting into models of non-perturbative phenomena [241], such as the underlying event (UE), multiple parton interactions (MPI), and associated soft QCD processes. For these latter applications, PDFs must respect some additional requirements beyond those of standard PDFs. Specifically, their use in initial-state showers necessitates that they remain non-negative down to the perturbative cutoff at $Q \simeq 1$ GeV. Additionally, their implementation in modeling the UE, MPI, and other low-energy QCD phenomena requires a very smooth extrapolation to very small x and very low Q^2 values, with a gluon PDF that increases rapidly in the small x region. To avoid numerical issues during Monte Carlo integration and sampling, PDFs must be numerically stable even in extreme regions of phase space that might be irrelevant to phenomenology. Lastly, to align with standard parton showers, the charm PDF must be generated perturbatively (i.e., without an intrinsic component). Moreover, for accommodating electroweak corrections, it should be possible to include a photon PDF $\gamma(x, Q^2)$ and allow for QED splittings in perturbative evolution.

Several groups [242–246] have developed variations of their LO PDF sets specifically designed for use in MC event generators. For example, the NNPDF2.3QED LO PDFs introduced in [123, 124, 247] were integrated into `Pythia8` and served as one of the inputs for its widely-used `Monash` tune [248], which is utilized in modeling non-perturbative

QCD physics. Beyond leading order, BFKL-resummed versions of the NNPDF3.1 PDF set, which incorporate constraints on the small- x gluon based on D -meson production data from LHCb as discussed in [130, 159, 249], also meet the aforementioned criteria and are available in `Pythia8` as a separate PDF set.

In this chapter, we introduce variants of NNPDF4.0 [4, 192] at LO, and for the first time, at NLO and NNLO, specifically designed for use in modern MC event generators. The primary aim of these NNPDF4.0MC sets is to meet the previously discussed requirements while also offering the most accurate representation of the NNPDF4.0 dataset, especially at NLO and NNLO.

5.1 Methodology

Unless otherwise stated, we use the same experimental data, theoretical calculations, and methodology as in the recent MHO, QED, and aN³LO variants of NNPDF4.0 [6–8]. Specifically, we utilize the new NNPDF theory pipeline [114], that we already discussed in this thesis. The same input values for the Standard Model parameters are applied, particularly $\alpha_s(M_Z^2) = 0.118$ for the LO, NLO, and NNLO fits. We provide only a central PDF, rather than a set of PDF replicas to quantify uncertainty, because with additional constraints, uncertainties might become unreliable, and they are not essential for applications in MC event generators.

5.1.1 Positivity and perturbative charm

The positivity of MC PDFs is crucial for both initial-state showering and the modeling of soft QCD phenomena. At LO, PDFs correspond directly to physical cross-sections and are thus positive-definite. However, this is not necessarily the case at NLO and higher orders, where PDFs become scheme-dependent and can be negative in certain regions of phase space. While PDFs in the commonly used $\overline{\text{MS}}$ scheme generally remain positive at NLO and beyond, this holds true primarily in the perturbative region, i.e., at sufficiently high scales [32, 250, 251]. As a result, PDF positivity might not be maintained when extrapolating to low Q values.

In the baseline NNPDF4.0 analysis, PDF positivity is enforced at the initial parametrization scale ($Q_0 = 1.65$ GeV) for LO, and at a higher scale, $Q_{\text{pos}}^2 = 5$ GeV², for NLO and beyond, in accordance with the approach outlined in [32, 251]. Additionally, the positivity of certain physical observables is ensured at Q_{pos}^2 . Consequently, within the NNPDF4.0 framework, NLO and NNLO PDFs may be negative at low Q^2 values, provided they evolve to become positive at $Q^2 \geq Q_{\text{pos}}^2$. Although this may occur in phase space regions lacking direct experimental constraints or where the fixed-order leading-twist approximation is not valid, positivity remains a necessary condition for MC generators.

We begin with the perturbative charm variant of NNPDF4.0, where charm is determined by perturbative matching conditions at the matching scale $\mu_c = m_c$. We then enforce the positivity of $g(x, Q_0)$ and $\Sigma(x, Q_0)$ at $Q_0 = 1$ GeV by squaring the outputs of the corresponding neural networks. This ensures that the gluon and quark singlet PDFs are positive at $Q_0 = 1$ GeV, and consequently remain positive for $Q > Q_0$, due to their growth at small x driven by perturbative QCD evolution as the scale increases. Positivity for individual quark and anti-quark PDFs is imposed at $Q_{\text{pos}}^2 = 5$ GeV², as in the default approach. This is sufficient to ensure positivity down to Q_0 , both at large x , where perturbative evolution is moderate even at low scales, and at small x , where non-singlet PDFs vanish. This strategy results in positive-definite PDFs across the entire (x, Q^2) range relevant to MC generators at LO and NLO.

At NNLO, the perturbative matching conditions result in a charm PDF that is negative at small $x \lesssim 10^{-2}$ when $Q = m_c$, although it becomes positive across all x values once $Q^2 \gtrsim 5 \text{ GeV}^2$. Therefore, at NNLO, it is not possible to meet both the requirement for charm to be positive and for it to be determined by perturbative matching at $\mu_c = m_c$. As we will discuss in Sect. 5.2, enforcing the positivity of the gluon at small x at low scales is not favored by the data, and doing so leads to a slight decline in the fit quality.

5.1.2 Extrapolation in x and Q^2

General-purpose MC event generators must deliver reliable results across the widest possible range of phase space. This demands input PDFs that exhibit smooth behavior across a broad Q range, from around $Q \simeq 1 \text{ GeV}$ (for initial-state showers and non-perturbative QCD modeling) up to $Q \sim 100 \text{ TeV}$ (important for future particle colliders and astroparticle physics applications), and from $x \simeq 10^{-9}$ (relevant for forward particle production) to large x values near the elastic limit at $x = 1$ (critical for high-mass new physics searches). Since these regions often fall outside the scope of available data, a reliable extrapolation procedure is essential.

While the extrapolation of PDFs in Q^2 is determined by perturbative QCD evolution, extrapolation in x relies on certain assumptions. In the NNPDF4.0 approach, the extrapolation to small and large x regions is guided by the output of a preprocessed neural network, which is influenced by both the neural network itself and the preprocessing function. However, extrapolation to low Q^2 and large x values can lead to numerical instabilities, both inherent and due to their representation as LHAPDF grids. Specifically, the behavior at low Q^2 is driven by evolution from higher scales, which can amplify small differences in the initial conditions because of the increasing value of $\alpha_s(Q^2)$. At large x , PDFs become very small, making them particularly susceptible to numerical instabilities. These issues are connected, as even minor numerical differences on the order of $\mathcal{O}(10^{-5})$ in solving the evolution equations can significantly distort PDFs in the large x region where they nearly vanish. Although these instabilities are generally harmless for phenomenological applications, they can cause numerical problems when PDFs are used in MC generators.

To mitigate these instabilities and ensure that the MC PDFs remain smooth and well-behaved across the entire range, the NNPDF4.0MC PDFs are provided as an LHAPDF grid with a denser coverage in x specifically for the range $x \in [0.7, 0.95]$. For $x \gtrsim 0.95$, where PDFs become negligible, any residual fluctuations can be safely set to zero. Moreover, a dedicated Gaussian filter is used to average out instabilities that are within the accuracy limits of the LHAPDF interpolation. To avoid issues with backward evolution, PDFs are parametrized at $Q_0 = 1 \text{ GeV}$, eliminating the need for backward evolution. As a result, we provide LHAPDF grids that offer interpolated values for all $x \in [10^{-9}, 1]$ and $Q \in [1, 10^6] \text{ GeV}$.

5.1.3 QED evolution and the photon PDF

As demonstrated in [6, 126, 130, 132, 252] and related searches, incorporating a photon PDF along with quark and gluon PDFs has a modest impact, primarily reducing the gluon momentum fraction by up to about 0.5% in favor of the photon. For our analysis, we use the photon PDF $\gamma(x, Q^2)$ at $Q = 1 \text{ GeV}$ from the NNPDF4.0QED NNLO PDF set [6]. This photon PDF is included as a boundary condition in the $\text{QCD} \otimes \text{QED}$ evolution of the LO, NLO, and NNLO NNPDF4.0MC PDFs, and we enforce a momentum sum

ID	Ref.	evolution (Q_0)	Positivity (Q_{pos})	Charm
NNPDF23_lo_as_0130_qed	[247]	$\text{QCD}_{\text{LO}} \otimes \text{QED}_{\text{LO}}$ TRN (1 GeV)	$g, q_i, \bar{q}_i > 0$ (1 GeV)	pert.
NNPDF40_lo_as_01180	[4]	QCD_{LO} TRN (1.65 GeV)	$g, q_i, \bar{q}_i > 0$ (1.65 GeV)	fitted
NNPDF40_lo_pch_as_01180	[4]	QCD_{LO} TRN (1.65 GeV)	$g, q_i, \bar{q}_i > 0$ (1 GeV)	pert.
NNPDF40MC_lo_as_01180	t.w.	QCD_{LO} TRN (1.0 GeV)	$g, q_i, \bar{q}_i > 0$ (1 GeV)	pert.
NNPDF40MC_lo_as_01180_qed	t.w.	$\text{QCD}_{\text{LO}} \otimes \text{QED}_{\text{LO}}$ EXA (1.0 GeV)	$g, q_i, \bar{q}_i > 0$ (1 GeV)	pert.
NNPDF40_nlo_as_01180	[4]	QCD_{NLO} TRN (1.65 GeV)	$g, q_i, \bar{q}_i > 0$ ($\sqrt{5}$ GeV)	fitted
NNPDF40_nlo_pch_as_01180	[4]	QCD_{NLO} TRN (1 GeV)	$g, q_i, \bar{q}_i > 0$ ($\sqrt{5}$ GeV)	pert.
NNPDF40MC_nlo_as_01180	t.w.	QCD_{NLO} TRN (1 GeV)	$g, \Sigma > 0$ (1 GeV) $q_i, \bar{q}_i > 0$ ($\sqrt{5}$ GeV)	pert.
NNPDF40_nlo_as_01180_qed	[6]	$\text{QCD}_{\text{NLO}} \otimes \text{QED}_{\text{NLO}}$ EXA (1.65 GeV)	$g, q_i, \bar{q}_i > 0$ ($\sqrt{5}$ GeV)	fitted
NNPDF40MC_nlo_as_01180_qed	t.w.	$\text{QCD}_{\text{NLO}} \otimes \text{QED}_{\text{NLO}}$ EXA (1 GeV)	$g, \Sigma > 0$ (1 GeV) $q_i, \bar{q}_i > 0$ ($\sqrt{5}$ GeV)	pert.
NNPDF40_nnlo_as_01180	[4]	QCD_{NNLO} TRN (1.65 GeV)	$g, q_i, \bar{q}_i > 0$ ($\sqrt{5}$ GeV)	fitted
NNPDF40_nnlo_pch_as_01180	[4]	QCD_{NNLO} TRN (1 GeV)	$g, q_i, \bar{q}_i > 0$ ($\sqrt{5}$ GeV)	pert.
NNPDF40MC_nnlo_as_01180	t.w.	QCD_{NNLO} TRN (1 GeV)	$g, \Sigma > 0$ (1 GeV) $q_i, \bar{q}_i > 0$ ($\sqrt{5}$ GeV)	pert.
NNPDF40_nnlo_as_01180_qed	[6]	$\text{QCD}_{\text{NNLO}} \otimes \text{QED}_{\text{NLO}}$ EXA (1.65 GeV)	$g, q_i, \bar{q}_i > 0$ ($\sqrt{5}$ GeV)	fitted
NNPDF40MC_nnlo_as_01180_qed	t.w.	$\text{QCD}_{\text{NNLO}} \otimes \text{QED}_{\text{NLO}}$ EXA (1 GeV)	$g, \Sigma > 0$ (1 GeV) $q_i, \bar{q}_i > 0$ ($\sqrt{5}$ GeV)	pert.

Table 5.1: The NNPDF4.0MC PDFs presented in this work (t.w.) and their baseline counterparts.

rule that now also accounts for the photon contribution. We employ the exact-iterated solution of the $\text{QCD} \otimes \text{QED}$ evolution equations, as implemented in EKO [149] and described in Chap. 3.2, which provides further details. For pure QCD evolution, we use the truncated solution, following Ref. [4], ensuring that each PDF set presented here is based on the same form of the evolution equations solution as their default versions.

5.1.4 NNPDF4.0MC overview

Tab. 5.1 provides a summary of the settings used for the NNPDF4.0MC PDFs compared to their baseline counterparts. It includes details such as the LHAPDF naming ID, publication reference, PDF parametrization scale, the method used for solving the evolution equations, the positivity scale and the treatment of the charm PDF (whether data-driven or determined by perturbative matching). In this table, q_i and \bar{q}_i represent the PDFs for light quarks (up, down, and strange) and anti-quarks, respectively, noting that, as per [32, 251], the positivity of the charm PDF is not enforced.

Dataset by process	NLO					NNLO				
	n_{dat}	QCD		QCD \otimes QED		n_{dat}	QCD		QCD \otimes QED	
		BL	MC	BL	MC		BL	MC	BL	MC
DIS NC	1953	1.35	1.37	1.38	1.54	2110	1.22	1.30	1.22	1.29
DIS CC	988	0.91	0.92	0.94	0.95	989	0.90	0.89	0.90	0.89
DY NC	669	1.58	1.84	1.67	2.04	736	1.20	1.30	1.22	1.33
DY CC	197	1.38	1.56	1.40	1.61	157	1.45	1.55	1.47	1.57
Top pairs	66	2.40	2.14	2.51	2.47	64	1.27	1.16	1.31	1.27
Single-inclusive jets	356	0.82	0.88	0.83	0.93	356	0.94	1.01	0.93	1.00
Dijets	144	1.51	1.55	1.56	1.62	144	2.01	2.01	1.94	1.93
Photon	53	0.57	0.60	0.64	0.74	53	0.76	0.67	0.74	0.68
Single top	17	0.36	0.36	0.38	0.36	17	0.37	0.38	0.39	0.40
Total	4443	1.28	1.30	1.30	1.44	4626	1.16	1.22	1.17	1.22

Table 5.2: The number of data points and the χ^2 per data point for the NLO and NNLO baseline NNP4.0 fits (BL), compared to their NNP4.0MC counterparts (MC), with the same process categorisation as in Ref. [7]. The χ^2 values are provided for the QCD-only (NNPDF40(MC)_<order>.as.01180) and for the QCD \otimes QED (NNPDF40(MC)_<order>.as.01180_qed) fits of Tab. 5.1.

5.2 The NNP4.0MC PDFs

In this section we will compare the NNP4.0MC PDFs set to its baseline NNP4.0 and to the NNP2.3QED LO PDFs used for the *Monash* tune [248] of *Pythia8*. Here we only show some representative results; an extensive set of comparisons can be found online in the NNP website.¹

Tab. 5.2 summarizes the fit quality for the NLO and NNLO PDF sets detailed in Tab. 5.1, presenting the number of data points and the χ^2 per data point. LO χ^2 values are not included, as the fit quality at LO is generally poor and the specific χ^2 value is not very meaningful. For pure QCD PDFs, we observe that at NLO (NNLO), the χ^2 per data point for the baseline fit increases from 1.28 (1.16) to 1.30 (1.22), corresponding to about 1σ (3σ) in terms of the statistical variance of the χ^2 distribution with $n_{\text{dat}} = 4443$ (4626) data points. Thus, the imposition of MC PDF conditions has no statistically significant effect at NLO, but at NNLO, it leads to a slight degradation in fit quality. This is because the rapid increase of the gluon at small x with increasing scale can result in a negative gluon at scales $Q^2 \lesssim \text{few GeV}^2$ [251], making it challenging to ensure a positive gluon at low scales. For the QCD \otimes QED sets, a similar pattern is observed at NNLO. However, at NLO, there is a more noticeable decline in fit quality. This is due to the fact that the sub-leading terms included in the exact solution of the evolution equations result in a faster perturbative evolution compared to the truncated solution, particularly when the anomalous dimension is large, making more evident the issue of maintaining positive gluons at low scales. Consequently, the difference in fit quality between the pure QCD and QCD \otimes QED cases at NLO reflects the impact of missing NNLO corrections.

Fig. 5.1 compares the MC and baseline LO and NLO PDFs by displaying the gluon,

¹https://data.nnpdf.science/vp-public/NNPDF40MC_comparisons/

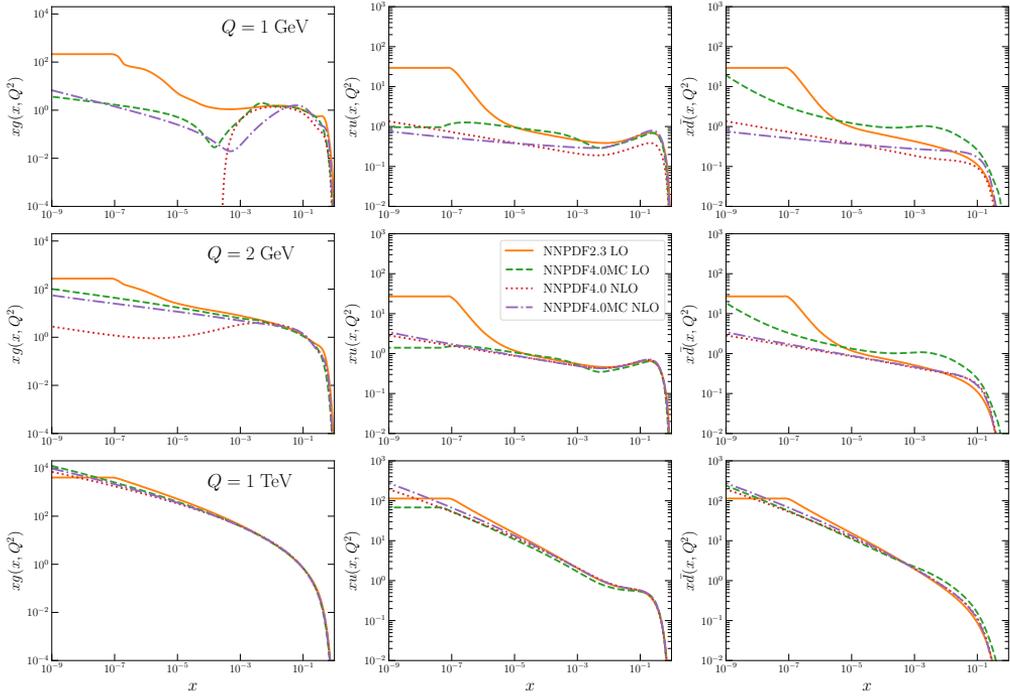


Figure 5.1: The NNPDF4.0MC LO and NLO PDFs for the gluon, up quark, and anti-down quark (from left to right) are compared to the NNPDF2.3LO and NNPDF4.0 NLO PDFs across three different scales: $Q = 1$ GeV, 2 GeV, and 1 TeV (from top to bottom). Only the central values are displayed in the region covered by LHAPDF.

up quark, and anti-down quark PDFs at $Q = 1$ GeV, 2 GeV, and 1 TeV. It is important to note that the small- x behavior of all quark and anti-quark PDFs is similar and primarily influenced by the singlet quark distribution. The figure shows the entire x range covered by the NNPDF4.0MC PDFs via the LHAPDF interpolation, which spans from 10^{-9} to 1. The NNPDF2.3LO set, in contrast, was only available for $x \geq 10^{-7}$, with PDFs being held constant at $x = 10^{-7}$ for smaller values. Beyond this difference, a key distinction between the NNPDF2.3 and NNPDF4.0 LO sets is that the NNPDF4.0MC set exhibits a qualitatively similar rise in the small- x gluon at both LO and NLO, which helps in tuning soft QCD models within MC event generators. The primary difference between the MC and default NLO PDFs is related to the small- x positivity of the gluon at low scales. As the scale Q increases, the differences between the various PDF sets become less pronounced due to perturbative evolution.

To illustrate the smoothness of the NNPDF4.0MC sets in the large- x extrapolation region, Fig. 5.2 shows the NLO and NNLO NNPDF4.0MC PDFs for $x = 0.85$ as a function of scale, alongside the central values of their baseline counterparts. The Q range displayed encompasses the entire interpolation range provided in the LHAPDF grids. All the PDFs presented demonstrate a satisfactory degree of smoothness.

To thoroughly evaluate the differences between the MC sets and their baseline counterparts, Fig. 5.3 shows the ratio of the NNPDF4.0MC NLO PDFs to the baseline PDFs, including the 68% confidence level uncertainties for the latter. For comparison, the

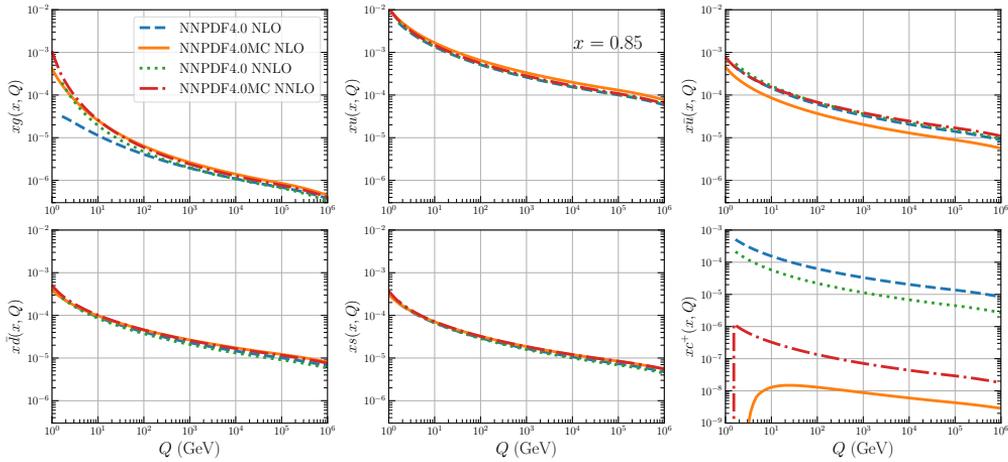


Figure 5.2: The NNPDF4.0MC NLO and NNLO PDFs for the gluon, up quark, anti-up quark, anti-down quark, strange quark, and total charm (arranged from left to right and from top to bottom) are compared to their baseline counterparts as a function of scale for a fixed large $x = 0.85$. The Q range displayed, from 1 to 10^6 GeV, represents the entire interpolation range available in the LHAPDF grids that we provide.

NNPDF4.0 NLO set with perturbative charm, as detailed in Tab. 5.1, is also presented. In the region where $x \gtrsim 10^{-3}$, which encompasses most of the experimental data, the quark MC PDFs generally fall within the uncertainty band of the baseline PDFs. However, larger discrepancies, attributable to the requirement for low-scale positivity, are observed in the gluon PDF, particularly at small $x \lesssim 10^{-2}$. These differences extend to the other PDFs at small x , resulting in a more pronounced small- x rise in the extrapolation region $x \lesssim 10^{-3}$ compared to the baseline.

The findings in Fig. 5.3 suggest that the additional model assumptions used in the MC PDFs do not significantly alter the baseline PDFs within the 1σ uncertainty range in most of the data region. This indicates that when the cross-sections are calculated with the NNPDF4.0MC sets, they should generally align with those derived from the baseline PDFs. Specifically, Fig. 5.3 shows that for many PDFs, particularly sea quark PDFs, a substantial portion of the difference between the MC and default PDFs can be attributed to the inclusion of perturbative charm.

5.3 Impact on LHC physics

In this section we will briefly assess the phenomenological impact of the NNPDF4.0MC PDFs set and we will study the similarities and the differences between the MC PDFs and their baseline counterparts shown in Figs. 5.1–5.3.

First, 5.4 presents the gluon-gluon, quark-anti-quark, and quark-quark parton luminosities at the LHC with a center-of-mass energy of $\sqrt{s} = 13.6$ TeV, as a function of the invariant mass of the final state, m_X . These luminosities are calculated using the same PDFs shown in Fig. 5.3 and are compared as ratios to the NNPDF4.0 baseline. The luminosities are integrated across the full rapidity range, making them primarily influenced by the PDF behavior in the central rapidity region, where $x_1 \sim x_2 \sim m_X/\sqrt{s}$. For $50 \text{ GeV} \lesssim m_X \lesssim 1 \text{ TeV}$, which corresponds to a medium-small x range, differences

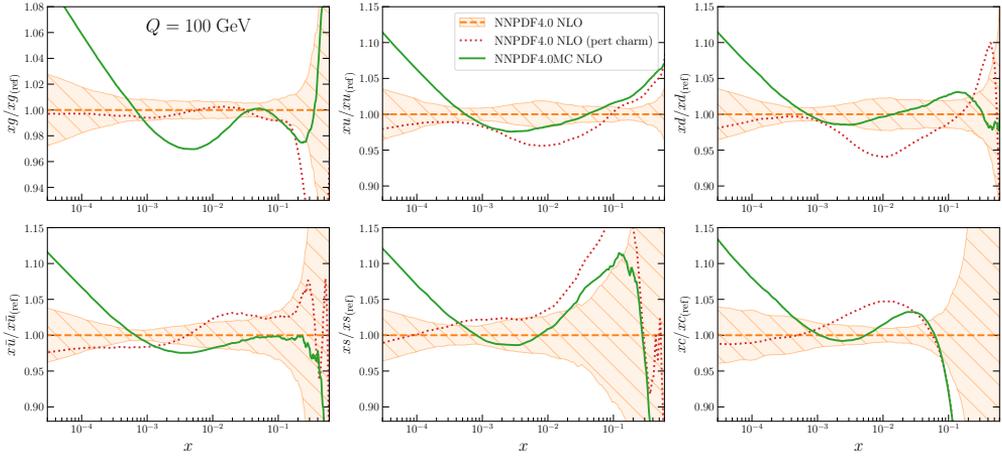


Figure 5.3: The NLO NNPDF4.0MC PDFs for the gluon, up quark, down quark, anti-up quark, strange quark, and charm (arranged from left to right and top to bottom) at $Q = 100$ GeV are presented as ratios to their baseline counterparts. The displayed uncertainty represents the 68% confidence level for the baseline. Additionally, the baseline variant with perturbative charm is included for comparison.

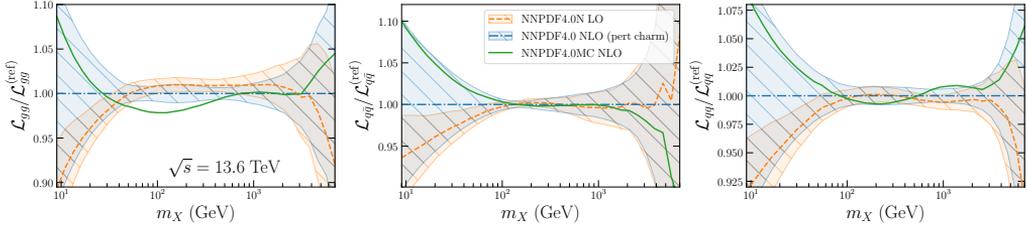


Figure 5.4: The gluon-gluon, quark-anti-quark, and quark-quark parton luminosities at the LHC with a center-of-mass energy of $\sqrt{s} = 13.6$ TeV are plotted as a function of the invariant mass m_X . These luminosities are computed using the same PDFs shown in Fig. 5.3 and are displayed as ratios relative to the NNPDF4.0 baseline.

between the MC PDFs and the baseline are generally modest and are mainly noticeable for the gluon. Specifically, the gluon-gluon luminosity shows a reduction of about 2% with the MC PDFs compared to the baseline for $100 \text{ GeV} \lesssim m_X \lesssim 3 \text{ TeV}$. Otherwise, variations between the NNPDF4.0 NLO PDFs and their MC counterparts are around 1% and become more pronounced, though still within uncertainties, for $m_X \lesssim 100 \text{ GeV}$ due to the more pronounced small x rise in the MC PDFs.

We then examine representative inclusive hard cross-sections, specifically Higgs and gauge boson production at the LHC with $\sqrt{s} = 13.6 \text{ TeV}$. These calculations are performed using the `ggHiggs` [253], `n3l0xs` [154], and `proVBFH` [167, 254] codes. In Fig. 5.5, we compare NLO and NNLO results (both for the PDFs and the matrix elements) obtained with the MC sets against their baseline counterparts. For the baseline sets, we also include `aN3LO` results using the settings from Ref. [8]. The uncertainty shown for the MC sets reflects only the missing higher-order contributions in the matrix element, assessed using the standard 7-point scale variation. For the baseline sets, the

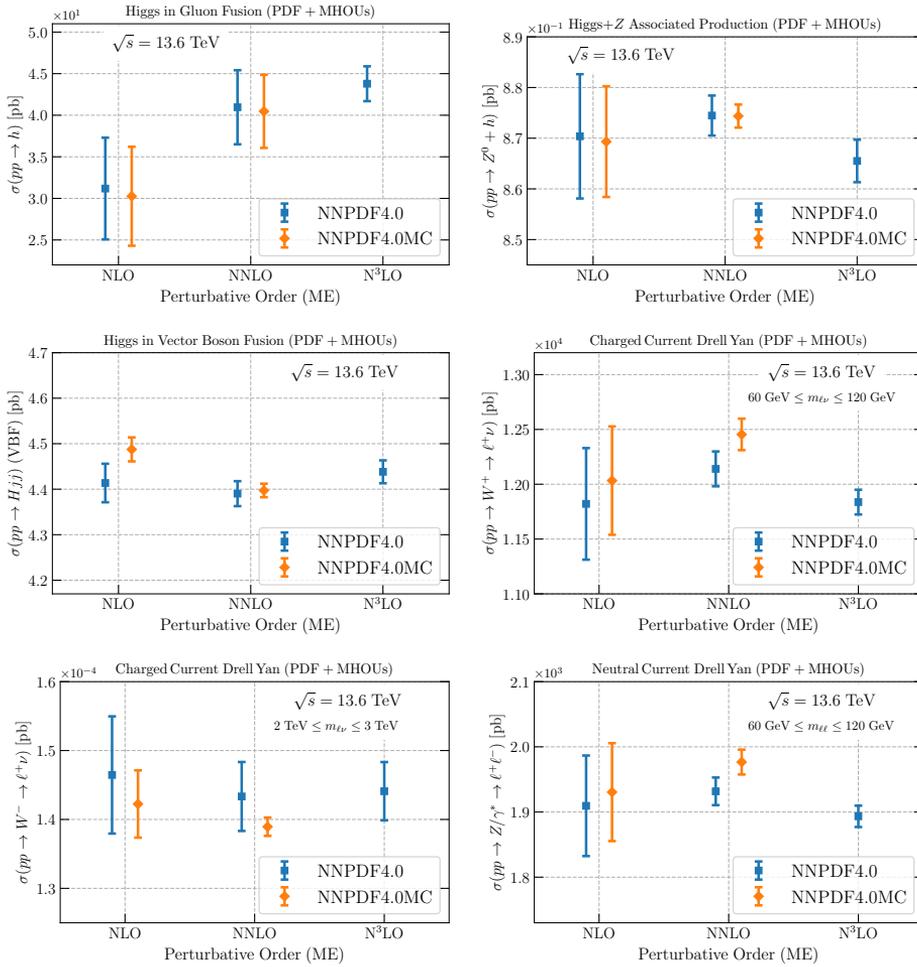


Figure 5.5: The inclusive NLO and NNLO cross-sections for Higgs production via gluon fusion, in association with a Z boson, and through vector boson fusion (top), as well as for on-shell and high-mass W production and on-shell Z production at the LHC with $\sqrt{s} = 13.6$ TeV (bottom), are compared between the NNPDF4.0MC PDFs and the baseline. Additionally, the aN³LO result is shown for the baseline NNPDF4.0. The uncertainty displayed reflects scale variation using the 7-point prescription for the MC PDFs, and for the baseline sets, it also includes PDF uncertainty, combined in quadrature.

uncertainty also includes the PDF uncertainty, combined in quadrature. The moderate differences observed align with those seen in the parton luminosities.

Next, we examine processes sensitive to soft physics, focusing on LHC differential distributions at leading order, generated using `Pythia8` and analyzed with the `Rivet` toolkit [255]. PDF uncertainties are not considered, and only central values obtained with NNPDF2.3LO, NNPDF4.0 NLO, and NNPDF4.0MC NLO PDFs are presented. We first look at the normalized transverse momentum distribution of the Z boson, reconstructed from bare dilepton events (electrons or muons), which is influ-

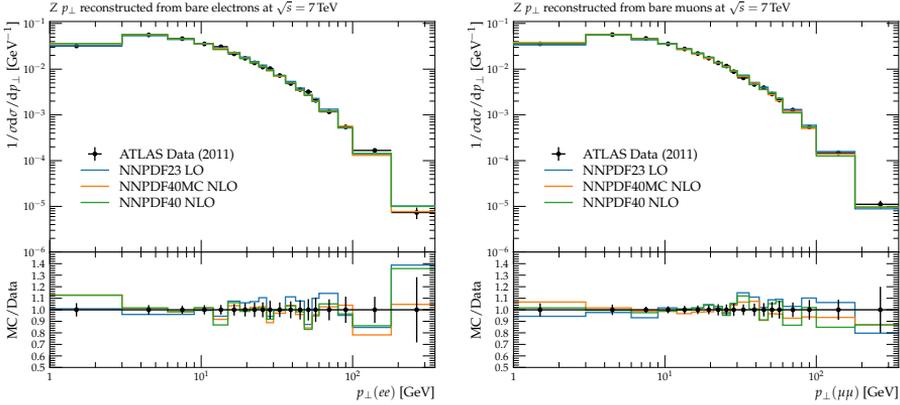


Figure 5.6: The normalized Z boson transverse momentum (p_T) distribution at LO, calculated using `Pythia8` and `Rivet`, with NNPDF2.3LO, NNPDF4.0 NLO, and NNPDF4.0MC NLO PDFs. The predictions are compared to ATLAS [256] data at $\sqrt{s} = 7$ TeV, using bare electron (left) or muon (right) pairs. Error bars on the data reflect both statistical and systematic uncertainties. The figure displays the absolute distribution (top) as well as the ratio of the theoretical prediction to the data (bottom).

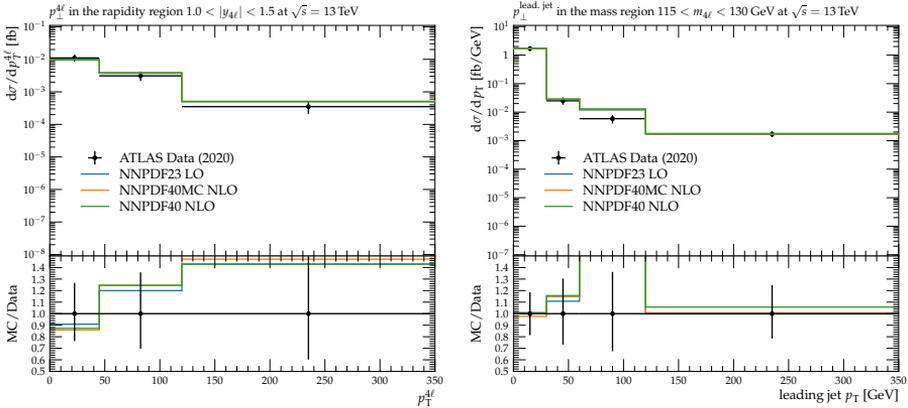


Figure 5.7: Same as Fig. 5.6, this figure presents the fiducial cross-section for Higgs production at $\sqrt{s} = 13$ TeV in the $H \rightarrow ZZ^* \rightarrow 4\ell$ ($\ell = e, \mu$) decay channel. It shows the four-lepton p_T distribution for $1.0 < |y_{4\ell}| < 1.5$ (left) and the p_T of the leading jet in events with at least one jet in the invariant mass range $115 < m_{4\ell} < 130$ GeV (right), compared with ATLAS data [257].

enced by both soft and hard QCD effects. In Fig. 5.6, the `Pythia8` LO predictions for $1\text{GeV} \leq p_{\perp}(\ell\ell) \leq 300$ GeV are compared with 7 TeV ATLAS data from Ref. [256]. The low and high p_{\perp} regions probe soft and hard QCD radiation, respectively. For these normalized distributions, higher-order QCD corrections partially cancel out. The differences between PDF sets are minimal, and all sets show good agreement with the data, except at very low p_{\perp} in the electron channel.

Next, we examine the fiducial cross-sections for Higgs production in the $H \rightarrow ZZ^* \rightarrow 4\ell$ ($\ell = e, \mu$) decay channel. In Fig. 5.7, we compare predictions with ATLAS data collected at $\sqrt{s} = 13$ TeV with an integrated luminosity of $\mathcal{L} = 139$ fb $^{-1}$ [257]. The results cover the transverse momentum distribution of the four hardest leptons in the event, $p_T^{4\ell}$,

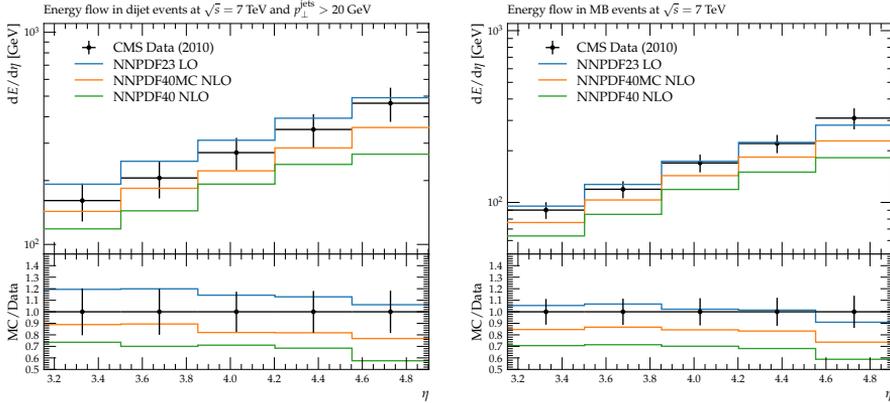


Figure 5.8: Same as Fig. 5.6, but for the energy flow in dijet (left) and minimum-bias events with $\sqrt{s} = 7$ TeV and $3.2 \leq \eta \leq 4.9$ compared to CMS data [258].

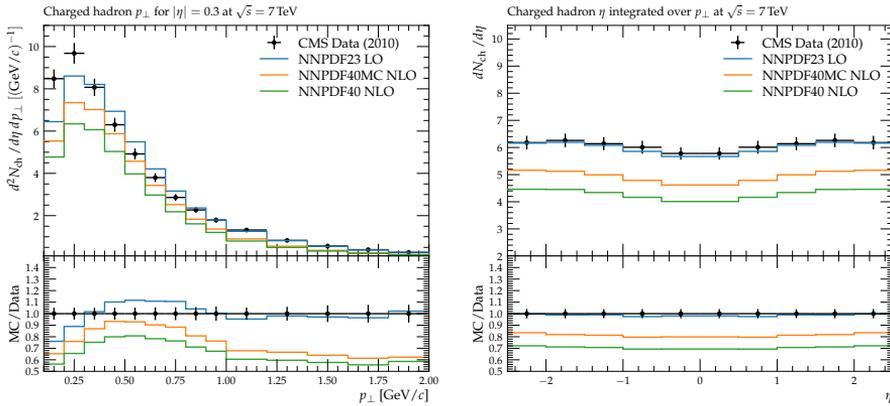


Figure 5.9: Similar to Fig. 5.8, this figure displays the charged-hadron transverse momentum (left) and pseudorapidity (right) distributions in proton-proton collisions at $\sqrt{s} = 7$ TeV, compared to the CMS data from [259].

within the rapidity range $1.0 < |y_{4\ell}| < 1.5$, and the transverse momentum of the leading jet in the invariant mass range $115 < m_{4\ell} < 130$ GeV. Differences between the various PDF sets are minimal, and the predictions show good agreement with the data.

We then examine the energy flow, defined as

$$\frac{dE}{d\eta} = \frac{1}{|\eta_{\max} - \eta_{\min}|} \left(\frac{1}{N_{\text{inel}}} \sum_{i=1}^{n_{\text{part}}} E_i \theta(\eta_i > \eta_{\min}) \theta(\eta_i < \eta_{\max}) \right), \quad (5.1)$$

where η represents the midpoint of the rapidity interval $[\eta_{\min}, \eta_{\max}]$, N_{inel} is the number of inelastic pp collisions, and n_{part} is the number of stable particles in the event with energy E_i . The energy flow in dijet events and minimum-bias events at $\sqrt{s} = 7$ TeV in the forward region $3.2 \leq \eta \leq 4.9$ is shown in Fig. 5.8, compared with the CMS data from [258]. For the dijet sample, a $p_{\perp}^{\text{jet}} > 20$ GeV cut is applied. Simulations based on NNPDF2.3LO show good agreement with the data for both dijet and minimum-bias

events, whereas simulations using NNPDF4.0 NLO sets (both MC and baseline) tend to underestimate the experimental measurements, indicating a possible need for a dedicated tune of soft QCD physics.

Finally, in Fig. 5.9, we present the charged-hadron multiplicity distribution, differentiated by pseudorapidity and transverse momentum, $d^2N_{\text{ch}}/d\eta dp_{\perp}$. This is shown as a function of p_{\perp} at a fixed rapidity of $|\eta|=0.3$ and as a function of η integrated over the entire p_{\perp} range. The predictions are compared to CMS measurements from [259], focusing on events with $p_{\perp} \leq 2$ GeV and $|\eta| < 2.5$ to emphasize sensitivity to non-perturbative QCD dynamics. As with the energy flow results, the NNPDF2.3LO set offers the best match to the experimental data, while the NNPDF4.0 sets tend to underestimate the CMS measurements. Both the energy flow in Fig. 5.8 and the charged-hadron differential distributions in Fig. 5.9 are sensitive to non-perturbative QCD processes, highlighting the need for a dedicated tune of soft QCD to achieve accurate descriptions. The Monash 2013 tune of `Pythia8` used in these simulations is based on NNPDF2.3LO, which explains the good agreement observed with this set.

Summary and Conclusions

“There is no real ending. It’s just the place where you stop the story.”

– Frank Herbert

In this thesis we have presented different studies whose main focus was the inclusion of new theoretical effects in the NNPDF4.0 fitting framework, that is the newest release of PDFs from the NNPDF collaboration. Such studies regarded the inclusion of QED effects, MHOU and N³LO corrections in a PDFs fit. The aim was a more precise extraction of the PDFs from experimental data, in order to have a better accuracy in the computation of high energy physics theoretical predictions.

In Chap. 1 we briefly presented the theoretical underground of this work, i.e. the theory describing strong interactions, we defined the PDFs, we showed how the DGLAP evolution equations are derived and how they are solved and in the end we showed how to deal with massive quarks.

In Chap. 2 we described the NNPDF4.0 fitting methodology, on which the work presented in this thesis is based. In particular, we described how the PDFs are parametrized at the initial scale, how the fit is performed, which constraints are imposed on the PDFs, how the experimental uncertainties are included in the fit and we briefly described the dataset used in the analysis.

In Chap. 3 we described how the QED corrections are included into a PDFs fit and how they have been implemented in the NNPDF code. We have presented a novel determination of QED PDFs using the NNPDF4.0 set of parton distributions, building on the approach previously employed for generating the NNPDF3.1QED PDFs [130]. This approach utilizes the LuxQED method [126, 127] to calculate the photon PDF. Our findings align with earlier searches, showing that while QED effects have a modest but non-negligible influence primarily on the gluon PDF, photon-initiated contributions are especially significant in high-mass processes like neutral-current Drell–Yan production. This determination of PDFs was achieved using a new NNPDF pipeline designed for generating theoretical predictions [114]. Due to the integration of QED evolution into this pipeline, particularly through the use of the `EKO` evolution code [149], the generation of QED variants of NNPDF determinations is now largely automated and will become standard in future releases. Indeed, there is generally no justification for excluding the photon PDF, even if its impact is minor, nor for ignoring its effect on the momentum fraction carried by the gluon.

The NNPDF4.0QED PDFs sets have been released through the `LHAPDF6` [33] interface:

<http://lhapdf.hepforge.org/> .

These sets are provided as collections of $N_{\text{rep}} = 100$ Monte Carlo replicas. In particular, we publish the NLO and NNLO global fits, constructed with the configurations outlined in Secs. 3.1-3.2, and labeled as:

NNPDF40_nlo_as.01180_qed
 NNPDF40_nnlo_as.01180_qed

These sets are also available on the NNPDF collaboration website:

<https://nnpdf.mi.infn.it/nnpdf4-0-qed/> .

They should be viewed as the QED PDFs version of the previously published NNPDF4.0 QCD-only PDF sets [4].

An important progression for QED PDFs is the full integration of electroweak corrections in the theoretical predictions used for PDF determination. This allows for an expansion of both the processes and the kinematic ranges that can be utilized in PDF determination, which are currently limited by the requirement to keep electroweak corrections minimal. This is particularly significant as electroweak effects become more influential in regions of phase space sensitive to large- x PDFs, which are crucial for searches for new physics [260, 261]. The availability of the `PineAPPLE` interface to automated Monte Carlo generators, such as `mg5_aMC@NLO`, within the new pipeline [114], will greatly facilitate this development. These advancements will contribute to achieving PDF determination with percent-level or even sub-percent-level accuracy.

In Chap. 4 we presented both the inclusion of MHOU and $N^3\text{LO}$ corrections into a PDFs fit and then we described how they have been combined with the QED corrections. We have introduced NLO and NNLO parton distribution sets where the PDF uncertainty included not only the uncertainties from the data and the analysis methodology used to derive the PDFs from the data, but also the uncertainty due to the perturbative truncation of the computations that generate the theoretical predictions compared to the data, i.e. the MHOU. We adopted the approach developed in Refs. [156, 157], which was used to construct the first NLO PDF sets incorporating MHOU. This methodology can now be extended to determine MHOU up to $N^3\text{LO}$, thanks to the availability of the `EKO` evolution code [149], and the integration of the NNPDF code [192] with both `EKO` and flexible tools for computing physical processes (such as the `YADISM` module [115] for DIS) through the new theory pipeline. The MHOU on PDFs discussed in this paper should be regarded as an additional component of the overall PDF uncertainty. They represent the uncertainty in the theoretical predictions used for PDF determination and are therefore comparable to the experimental uncertainty in the data itself. These uncertainties are independent of the separate MHOU associated with the hard cross-section of the processes being predicted. Thus, to determine the total uncertainty on predictions, one must combine the PDF uncertainty, now including the MHOU component, with the MHOU uncertainty on the hard cross-section calculation. This hard cross-section uncertainty is typically estimated as the envelope of a 7-point scale variation (see, for example, Ref. [262]). Alternatively, MHOU on the hard cross-section can be incorporated by constructing a theory covariance matrix for the cross-section. This approach has the advantage of accounting for the correlation between the theory uncertainty in the process used for PDF determination and that in the hard cross-section, which may become important if the experimental uncertainties are small and the process being predicted was also used in PDF determination [263, 264]. This can be achieved by determining the cross-correlation of MHO and PDF uncertainties between the predicted process and

those used for PDF determination [265]. In this context, it is worth noting that an alternative approach for incorporating MHOU's into predictions (one that accounts for MHOU's on PDFs and their full correlation with MHOU's on the hard cross-sections) is to include scale variation directly in the Monte Carlo sampling [264]. A comparative analysis of this method with the one used in this thesis, as well as with other approaches to scale variation (such as different methods for constructing the theory covariance matrix discussed in Sec. 4.1), will be reserved for future research.

The NNPDF4.0MHOU NNLO PDFs set is made publicly available via the LHAPDF6 interface,

<http://lhpdf.hepforge.org/> .

It is published as a set of $N_{\text{rep}} = 100$ Monte Carlo replicas, and it is called

NNPDF40_nnlo_as_01180_mhou

It must be considered a more precise version of the already published NNPDF4.0 NNLO PDF set [4]. This PDF set is also made available via the NNPDF collaboration website

<https://nnpdf.mi.infn.it/nnpdf4-0-mhou/> .

Moreover, we have introduced the first aN^3LO PDF sets within the NNPDF framework by developing a complete set of approximate N^3LO splitting functions, utilizing available partial results and known limits, along with approximate massive DIS coefficient functions. We summarized the newly released PDF sets, highlight the key findings regarding their characteristics, and outline our plans for future developments. The NNPDF4.0 aN^3LO PDF sets are now accessible through the LHAPDF6 interface:

<http://lhpdf.hepforge.org/> .

Specifically, we provide an aN^3LO NNPDF4.0 set:

NNPDF40_an3lo_as_01180

This set complements the LO, NLO, and NNLO sets from Ref. [4]. We are also releasing NLO and aN^3LO NNPDF4.0MHOU sets:

NNPDF40_nlo_as_01180_mhou

NNPDF40_an3lo_as_01180_mhou

These sets expand upon the NNLO NNPDF4.0MHOU PDF set that we discussed before. They incorporate the MHOU on the processes used for PDF determination into the PDF uncertainty, while otherwise remaining consistent with the default sets, allowing for direct comparison. All these sets are also available on the NNPDF Collaboration website,

<https://nnpdf.mi.infn.it/nnpdf4-0-n3lo/> .

We have conducted an initial evaluation of these PDF sets by comparing them with their NLO and NNLO counterparts, both with and without MHOU's. Our key findings are as follows:

- All PDFs show good perturbative convergence, with differences decreasing as the perturbative order increases. The aN^3LO results consistently align with the NNLO within the uncertainties.

- For quark PDFs, the difference between NNLO and aN³LO results is minimal, indicating that, given the current data and methodology, the influence of higher-order corrections is negligible.
- For the gluon PDF, a more noticeable shift is observed between NNLO and aN³LO, highlighting the importance of including aN³LO for precision phenomenology.
- Including MHOU's enhances perturbative convergence, primarily by shifting central values at each order closer to the higher-order result, with this adjustment diminishing as the perturbative order increases.
- With the inclusion of MHOU's, the fit quality becomes nearly independent of the perturbative order, and PDF uncertainties generally decrease (or remain stable) due to improved data compatibility.
- The effect of MHOU's at aN³LO is negligible for quarks but more significant for the gluon PDF.

Overall, these findings emphasize the importance of incorporating aN³LO corrections and MHOU's for achieving sub-percent accuracy in precision phenomenology.

In the end we have shown how to combine MHOU and N³LO effects with QED corrections. The PDFs sets are available through the `LHAPDF6` interface and are labeled as

```
NNPDF40_nnlo_as_01180_qed_mhou
NNPDF40_an3lo_as_01180_qed_mhou
```

The differences between the PDFs sets with and without QED effects are minimal, as we found comparing NNPDF4.0QED with NNPDF4.0, confirming that overall, QED effects are a percent correction on a QED fit. Moreover, the different photons agree beyond percent accuracy for $x \gtrsim 10^{-3}$, while at very small x may differ by few percent.

In Chap. 5 we showed how the implementation of the new theory pipeline and of QED corrections has been applied for the construction of Monte Carlo PDFs, i.e. PDFs sets used for Monte Carlo event generators. The NNPDF4.0MC PDFs introduced in this work meet the requirements of event generators not only at LO but also at NLO and NNLO accuracy. The NLO and NNLO sets offer a robust description of the global dataset while minimizing differences compared to baseline sets, ensuring their reliability for calculating hard cross-sections at the LHC and other facilities. This allows the combination of the precision and accuracy of global PDF sets at NLO and NNLO with the usability of these PDFs in generators for initial-state radiation and modeling of soft QCD processes. To ensure agreement with data for non-perturbative processes like the underlying event, pileup, and low- p_T radiation, the soft QCD models specific to each event generator will need to be tuned using these new NNPDF4.0MC PDFs, especially considering their behavior in low- x physics, which is a key component of tuning models. Dedicated tunes will be necessary for the NNPDF4.0MC PDF sets to be instrumental in the development of a new generation of Monte Carlo codes with higher perturbative accuracy. To this end, we plan to collaborate with event generator developers to integrate NNPDF4.0MC into their frameworks and produce dedicated tunes for soft QCD physics, enabling the comprehensive simulation of LHC processes, from soft to perturbative regions, within a single physics model. The NNPDF4.0 MC sets are made available through the `LHAPDF6` interface, and the different fits are labeled as presented in Tab. 5.1, and the NNPDF Collaboration website.

<https://nnpdf.mi.infn.it/nnpdf4-0-mc/>.

Regarding the future developments of the NNPDF methodology, given that there is no reason to neglect QED, MHOU and N³LO effects, they will be included by default in the NNPDF fits. The advancements made in various aspects of the NNPDF analysis framework, as discussed in this thesis, will be included in NNPDF4.1, an updated global fit scheduled for release in 2025. This new release will incorporate QED effects, MHOU_s, and aN³LO corrections, and will feature a substantially larger dataset, including high-precision LHC measurements from the entire Run II luminosity. By extending the `PineAPPLE` grids to NNLO fixed-order calculations, NNPDF4.1 will provide a global PDF determination without relying on the QCD K -factor approximation at NNLO. Additionally, the update will benefit from methodological enhancements, including advanced hyperparameter optimization techniques based on estimators derived from the complete PDF probability distributions. The implementation of NNPDF4.1 on GPU architectures will also significantly reduce computational overhead, facilitating more efficient and innovative PDF studies.

Attempts to construct a truncated solution of DGLAP equations in presence of two running couplings

In this appendix we will describe the attempts to construct a truncated solution for mixed QED \otimes QCD DGLAP equations and the reason why this is not possible in an analytical way. In fact, as we already mentioned, a truncated solution is still possible adopting numerical techniques, like the so-called “epsilon trick”, as it has been done for NNPDF3.1QED [130]. However, such approach exhibits numerical instability and for this reason we chose not to adopt it.

As we discussed in Sec. 1.6, perturbative solutions of equations of the form

$$\mu^2 \frac{d}{d\mu^2} f_i = - \left(a_s \gamma_{ij}^{(1,0)} + a_s^2 \gamma_{ij}^{(2,0)} + \dots \right) f_j, \quad (\text{A.1})$$

are based on the observation that at LO it can be solved exactly, with the solution being the one given in Eq. (1.123). Observe that we are using the convention of Eq. (3.13) for the anomalous dimensions. In order to find an expanded solution at NLO we can write the ansatz

$$f_i = f_{\text{LO},i} + a_s f_{\text{NLO},i}, \quad (\text{A.2})$$

and impose that it is solution of the equation

$$\frac{d}{dt} f_i = - \left(a_s \gamma_{ij}^{(1,0)} + a_s^2 \gamma_{ij}^{(2,0)} \right) f_j. \quad (\text{A.3})$$

Inserting Eq. (A.2) into Eq. (A.3) we obtain

$$\frac{d}{dt} f_{\text{NLO},i} = -a_s \left(\delta_{ij} \beta_{\text{QCD}}^{(2,0)} - \gamma_{ij}^{(1,0)} \right) f_{\text{NLO},j} - a_s \gamma_{ij}^{(2,0)} f_{\text{LO},j} + \mathcal{O}(a_s^2), \quad (\text{A.4})$$

with $t = \log \mu^2$. Considering the matrix \mathbf{R} that diagonalizes $\gamma_{ij}^{(1,0)}$, i.e.

$$\gamma_D^{(1,0)} = \mathbf{R} \gamma^{(1,0)} \mathbf{R}^{-1}, \quad (\text{A.5})$$

and defining

$$f'_{\text{NLO},i} = R_{ij} f_{\text{NLO},j}, \quad (\text{A.6})$$

$$\gamma'^{(2,0)} = \mathbf{R} \gamma^{(2,0)} \mathbf{R}^{-1}, \quad (\text{A.7})$$

$$f'_{\text{LO},i} = R_{ij} f_{\text{LO},j} = R_{ij} \left(e^{\gamma^{(1,0)} L} \right)_{jk} R_{kl}^{-1} R_{lm} f_{0,m} = \left(e^{\gamma_D^{(1,0)} L} \right)_{ij} f'_{0,j}, \quad (\text{A.8})$$

with $L = \log\left(\frac{a_s}{a_s^0}\right) / \beta_{\text{QCD}}^{(2,0)}$, we can write that

$$\frac{d}{dt} f'_{\text{NLO},i} = -a_s \left(\delta_{ij} \beta_{\text{QCD}}^{(2,0)} - \gamma_{D,ij}^{(1,0)} \right) f'_{\text{NLO},j} - a_s \gamma_{ij}^{\prime(2,0)} f_{\text{LO},j}^D + \mathcal{O}(a_s^2). \quad (\text{A.9})$$

It means that with a basis rotation we have decoupled the components of f'_{NLO} , so that Eq. (A.9) can be solved. Observe that this approach works since the rotation matrix \mathbf{R} does not depend on t , i.e. on μ , but only on the Mellin space variable N . This solution is equivalent to the one presented in Sec. 1.6.

In the case in which we consider QED corrections to the anomalous dimensions, Eq. (A.1) reads

$$\mu^2 \frac{d}{d\mu^2} f_i = - \left(a_s \gamma_{ij}^{\prime(1,0)} + \gamma_{ij}^{\text{QED}} + a_s^2 \gamma_{ij}^{\prime(2,0)} + \dots \right) f_j, \quad (\text{A.10})$$

with γ^{QED} and $\gamma_{ij}^{\prime(1,0)}$ defined as in Eqs. (3.60-3.61). For simplicity we are neglecting the running of a_{em} . If now we consider the LO solution as the one given in Eq. (1.123), i.e. we treat the QED corrections as NLO corrections, and we make the same ansatz of Eq. (A.2) we obtain

$$\frac{d}{dt} f_{\text{NLO},i} = -a_s \left(\delta_{ij} \beta_{\text{QCD}}^{\prime(2,0)} - \gamma_{ij}^{\prime(1,0)} \right) f_{\text{NLO},j} - a_s \gamma_{ij}^{\prime(2,0)} f_{\text{LO},j} - \gamma_{ij}^{\text{QED}} f_{\text{NLO},j} + \mathcal{O}(a_s^2), \quad (\text{A.11})$$

with $\beta_{\text{QCD}}^{\prime(2,0)}$ defined according to Eq. (3.62). It is clear now that performing the rotation in Eq. (A.6) doesn't decouple the components of f_{NLO} because of the non commutativity of $\gamma_{ij}^{\prime(1,0)}$ and γ_{ij}^{QED} . This argument shows that it is not possible to apply the standard truncated solution to the mixed QCD \otimes QED evolution because of the presence of two couplings. In the case in which we consider the running of a_{em} the situation is even more complicated. In order to solve this problem we may try to include the QED corrections into the LO expression and to treat the two couplings independently trying an expansion in both a_s and a_{em} . However, in this case we would have that the LO term is given by

$$f_{\text{LO}}(Q^2) = \mathcal{P} \exp \left(- \int_{Q_0^2}^{Q^2} \frac{d\mu^2}{\mu^2} \left(a_s(\mu^2) \gamma^{(1,0)} + a_{em}(\mu^2) \gamma^{(0,1)} \right) \right) f(Q_0^2), \quad (\text{A.12})$$

that does not admit an analytic solution because of the non commutativity of $\gamma^{(1,0)}$ and $\gamma^{(0,1)}$.

The failure of the standard truncated solution in the QCD \otimes QED case can be observed when we try to apply the method described in Sec. 1.6 in presence of QED corrections. First of all we will consider the case of fixed a_{em} . In this case Eq. (1.124) becomes

$$\frac{d}{da_s} \vec{f} = \left(\frac{\mathbf{R}^{\text{QED}}}{a_s^2} + \frac{\mathbf{R}^{(0)}}{a_s} + \mathbf{R}^{(1)} + \mathcal{O}(a_s) \right) \cdot \vec{f}, \quad (\text{A.13})$$

with

$$\mathbf{R}^{\text{QED}} = \frac{\gamma^{\text{QED}}}{\beta_{\text{QCD}}^{\prime(2,0)}}, \quad (\text{A.14})$$

$$\mathbf{R}^{(0)} = \frac{\gamma^{(0)}}{\beta_{\text{QCD}}^{\prime(2,0)}} - b_1 \mathbf{R}^{\text{QED}}, \quad (\text{A.15})$$

$$\mathbf{R}^{(1)} = \frac{\mathbf{R}^{(1)}}{\beta_{\text{QCD}}^{\prime(2,0)}} + (b_1^2 - b_2) \mathbf{R}^{\text{QED}}, \quad (\text{A.16})$$

and similarly for the higher orders. The presence of the term proportional to $\mathcal{O}(a_s^{-2})$ is what breaks the ansatz in Eq. (1.127): indeed, when we insert that ansatz in the DGLAP equation, it does not generate a term of $\mathcal{O}(a_s^{-2})$ to be compared with the one in Eq. (A.13). For this reason the truncated solution described in Sec. 1.6 fails when applied to the QCD \otimes QED case. Also in this case if we consider the running of a_{em} the situation is even more complicated since we have to insert the LO QED term inside the definition of LO solution, resulting in Eq. (A.12), that cannot be solved in an analytical way.

NNPDF4.0 truncated vs exact solution

In this appendix we will quantify the impact of the choice of solution of the DGLAP equations on the NNPDF4.0 PDFs determination. Indeed, in Sec. 1.6 we have seen that there are two possible kind of solutions to PDFs evolution equations, i.e. truncated and exact. Both of them are approximate solution and therefore they will not provide results that are numerically identical. However, the two solutions provide results that are the same up to higher orders. It means that the differences between the two solutions decrease as we increase the perturbative order. Hence, we expect that the two solutions provide results that at NNLO are much more in agreement with respect to NLO. To this goal we have performed a pure QCD PDFs fit both at NLO and at NNLO using the exact solution in DGLAP evolution, instead of the truncated one that is used for NNPDF4.0. For the computation of the theory predictions using the two different evolutions we have used the new theory pipeline described in Sec. 2.3.

The statistical indicators comparing the fits using respectively truncated and exact solution are compared in Tab. B.1 both at NLO and at NNLO. We can observe that at NLO the exact solution slightly deteriorates all the parameters of the fit, while at NNLO they are almost unchanged. Indeed, as we already discussed, the differences between the two solutions become smaller as we increase the perturbative accuracy of the splitting functions.

In Figs. B.1-B.4 there are the comparisons between the different PDFs sets, both at NLO and at NNLO and both at fitting scale and at 100 GeV. It is evident that the PDFs

	NLO QCD		NNLO QCD	
	Truncated solution	Exact solution	Truncated solution	Exact solution
χ^2	1.26	1.30	1.17	1.17
$\langle E_{\text{tr}} \rangle_{\text{rep}}$	2.41 ± 0.06	2.46 ± 0.07	2.28 ± 0.05	2.26 ± 0.06
$\langle E_{\text{val}} \rangle_{\text{rep}}$	2.57 ± 0.10	2.62 ± 0.11	2.37 ± 0.11	2.34 ± 0.10
$\langle \chi^2 \rangle_{\text{rep}}$	1.29 ± 0.02	1.34 ± 0.02	1.20 ± 0.02	1.19 ± 0.01
$\langle \text{TL} \rangle_{\text{rep}}$	12900 ± 2000	12600 ± 2000	12400 ± 2600	12100 ± 2500
ϕ	0.156 ± 0.006	0.164 ± 0.005	0.153 ± 0.005	0.147 ± 0.005

Table B.1: Comparison of statistical indicators for a set of 100 NNPDF4.0 NNLO PDF replicas generated with the new theory pipeline and the published NNPDF4.0 NNLO 100 replica set.

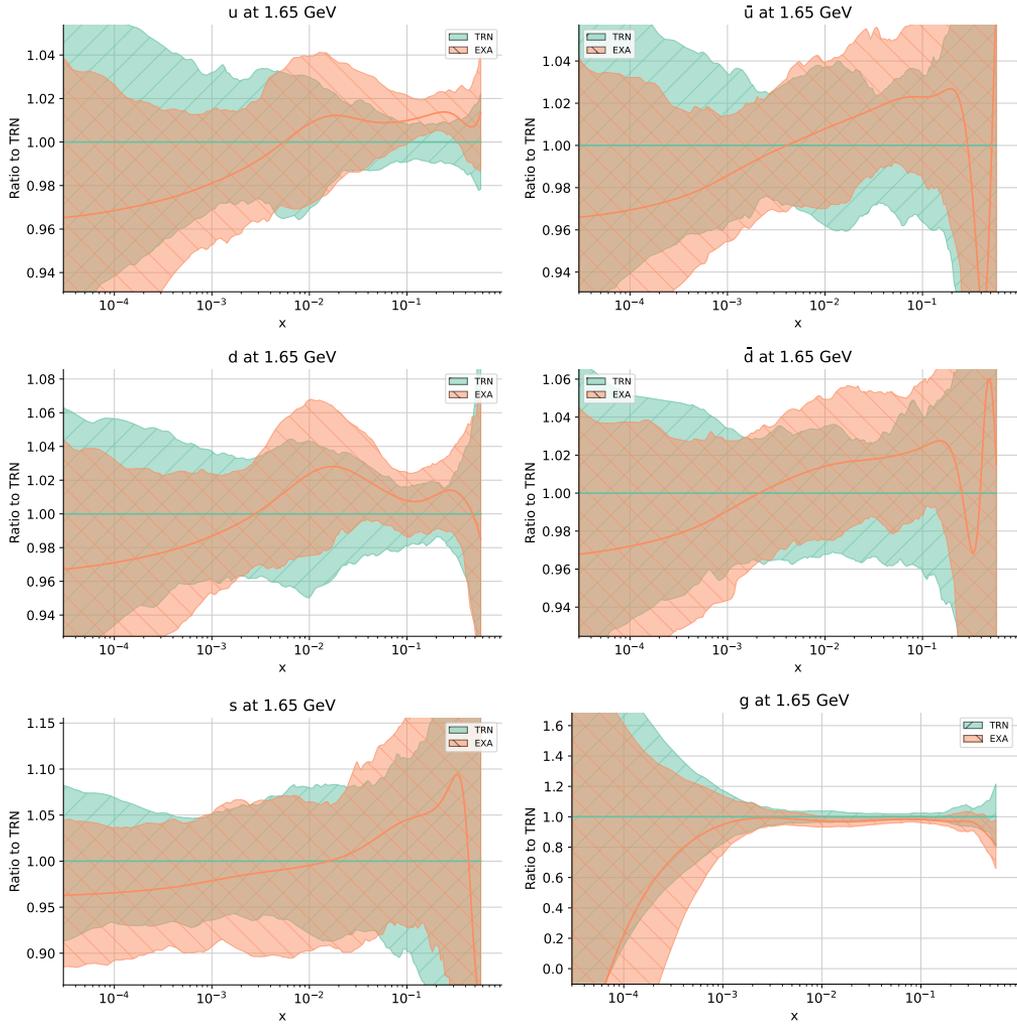


Figure B.1: Comparison between the NNPDF4.0 determinations at NLO using respectively the truncated (default) and the exact solutions to DGLAP equations. Gluon and up, anti-up, down, anti-down and strange quarks are shown at fitting scale as a ratio to the central value of the truncated solution.

are statistically indistinguishable at fitting scale, with slightly larger differences at NLO, as we expected. At 100 GeV instead, the NLO fits start to differ because of the two different solution methods, while the NNLO fits are almost identical for $x \geq 10^{-3}$. At smaller x values, the two NNLO fits begin to differ at the half-sigma level, with the maximum differences reaching the one-sigma level for the gluon.

This aligns with the fact that the exact and truncated solutions differ due to higher-order perturbative corrections, which exceed the NNLO precision of the calculation. It is well established that at small x , perturbative convergence deteriorates because high-energy logarithms require resummation for accurate PDF determination [159]. The qualitative behavior of the PDFs depicted in Fig. B.4 supports this explanation: the exact

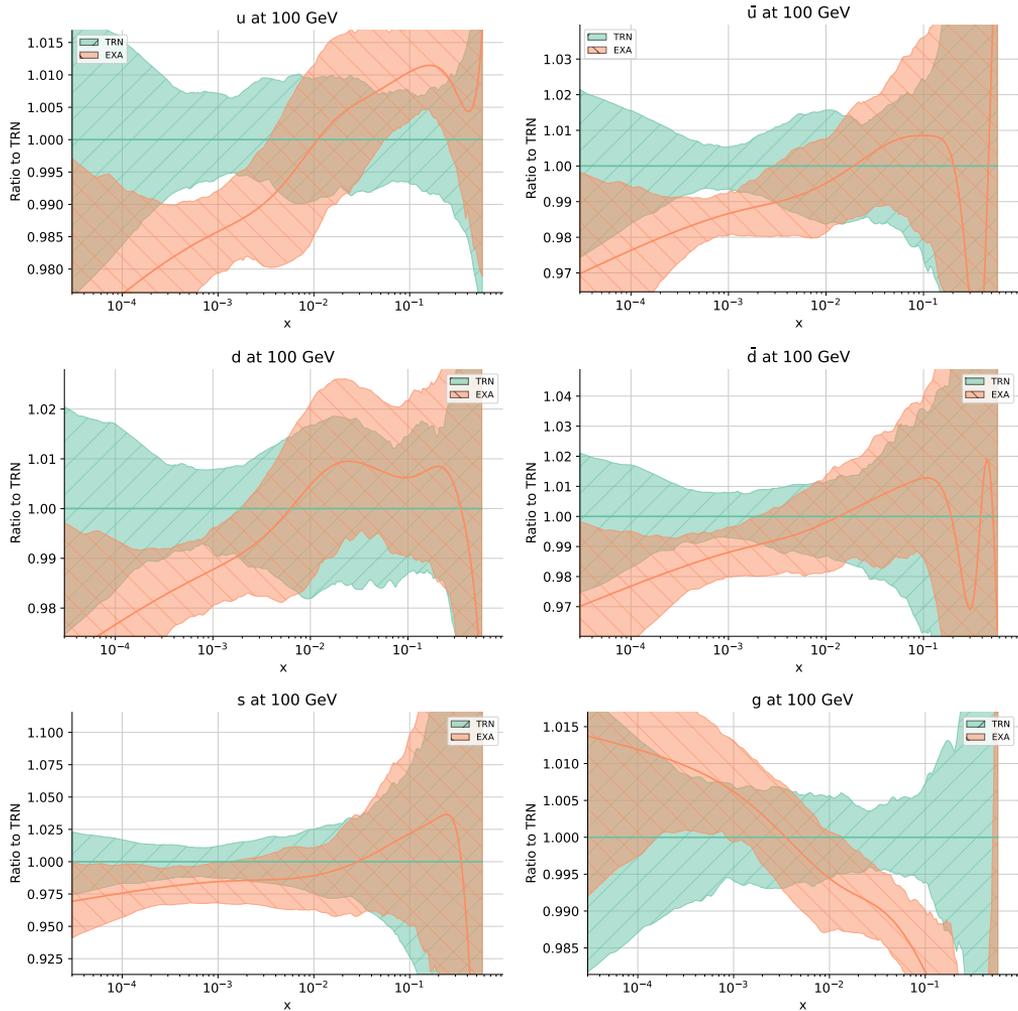


Figure B.2: Comparison between the NNPDF4.0 determinations at NLO using respectively the truncated (default) and the exact solutions to DGLAP equations. Gluon and up, anti-up, down, anti-down and strange quarks are shown at 100 GeV as a ratio to the central value of the truncated solution.

solution exponentiates a set of sub-leading small- x logarithms, while the truncated solution linearizes them. Consequently, in the small- x region, where data is lacking, these contributions cause a more pronounced increase in the gluon, subsequently affecting the quark-anti-quark sea.

In the end, we conclude that, at least at NNLO, the PDFs fits do not strongly depend on the method for solving the DGLAP equations in the most of x range, with some differences that arise due to evolution at small values of x . At NLO instead there is still some discrepancy between the two evolution strategies that are visible at fitting scale but mostly at larger scales because of the evolution.

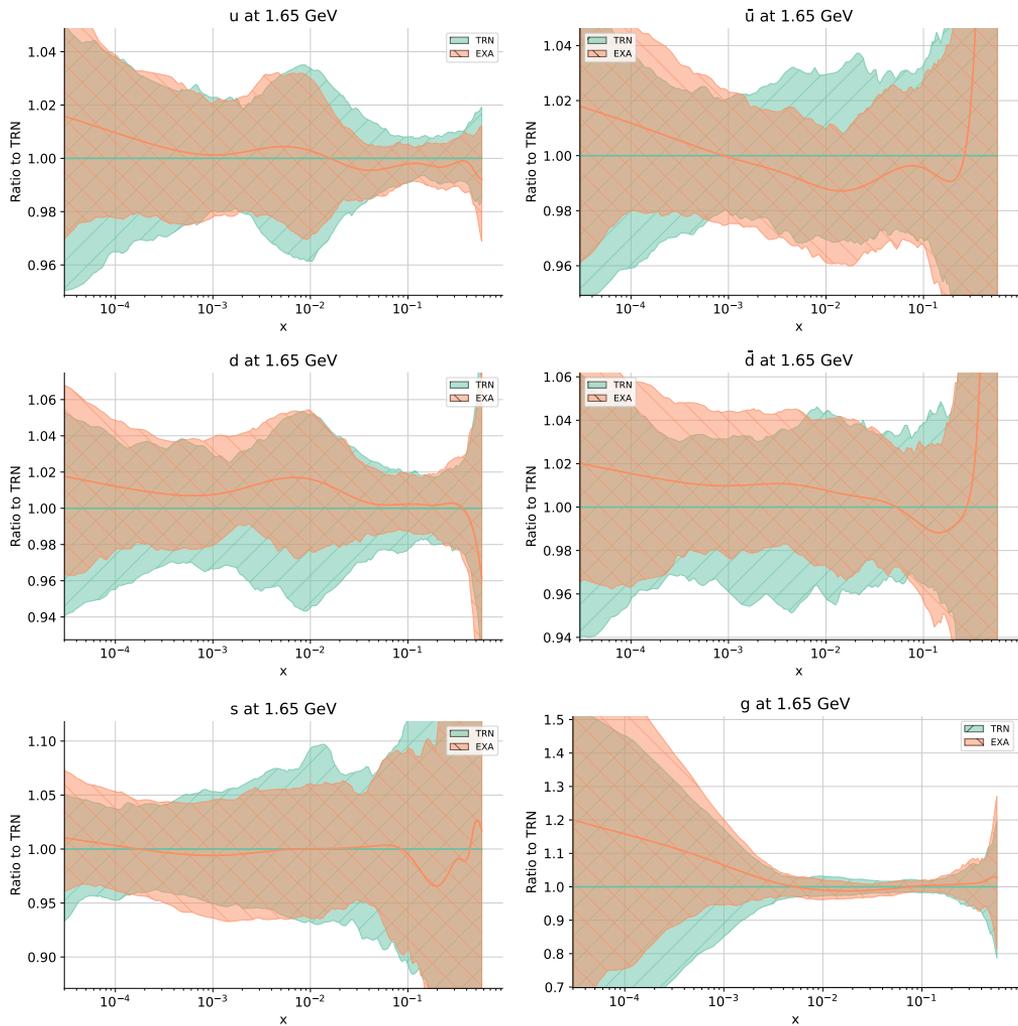


Figure B.3: Comparison between the NNPDF4.0 determinations at NNLO using respectively the truncated (default) and the exact solutions to DGLAP equations. Gluon and up, anti-up, down, anti-down and strange quarks are shown at fitting scale as a ratio to the central value of the truncated solution.

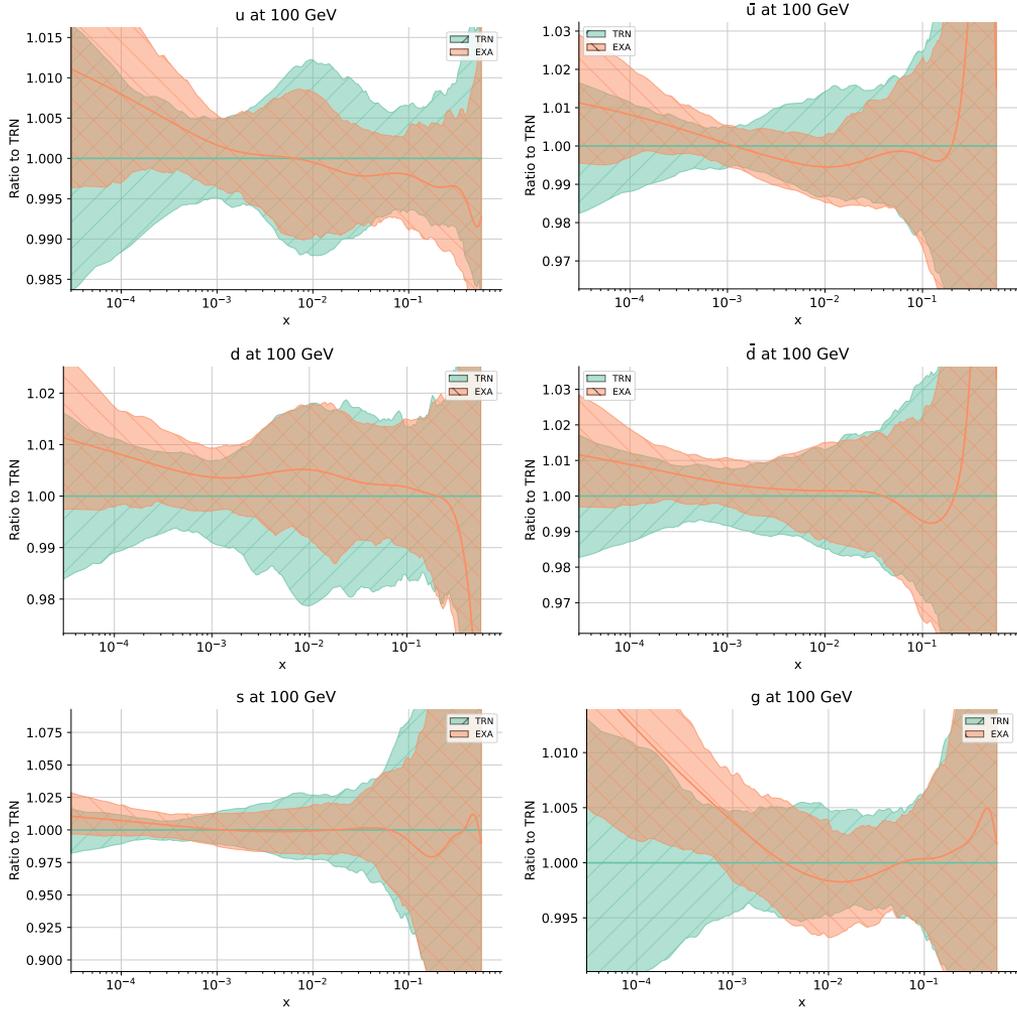


Figure B.4: Comparison between the NNPDF4.0 determinations at NNLO using respectively the truncated (default) and the exact solutions to DGLAP equations. Gluon and up, anti-up, down, anti-down and strange quarks are shown at 100 GeV as a ratio to the central value of the truncated solution.

Dependence of the PDFs on the fitting scale

In Sec. 2.1 we have shown that in order to extract the PDFs from experimental data, we need to parametrize them at an initial scale Q_0 , see Eqs. (2.1-2.2) and Eq. (2.3), since the PDFs at higher (or lower) scales can be obtained through DGLAP evolution. The choice of the fitting scale Q_0 is somehow arbitrary and PDFs determinations done with different values of Q_0 should be compatible within the uncertainties. However, some choices are better than others. First of all, we observe that the differences in PDFs extracted at different scales are given by higher orders terms. Therefore, as we vary Q_0 a change in the values of the PDFs will be observed, and it will increase for values of Q_0 that are very far apart. Then, we know that DGLAP evolution shrinks the PDFs, in the sense that if we evolve initial conditions that are very distant, the percent difference of the evolved PDFs will decrease. It means that if we perform a fit at a high scale, our fitting algorithm will have to choose the best configuration between curves that are very close to each other. This will result in a worse performing fitting algorithm that may take more time to converge (or not converge at all). Last, in Chap. 2 we have said that the hyperparameters of the NNPDF4.0 fitting algorithm have been selected through a procedure called hyperoptimization. Such optimization has been performed setting Q_0 to its default value of $Q_0 = 1.65$ GeV. Therefore, varying Q_0 we may need to re-hyperoptimize since the shapes of the PDFs will change due to DGLAP evolution. Moreover, if we move Q_0 below the mass of the charm or above the mass of the bottom the shapes of the PDFs will also be modified by the matching conditions, see Sec. 1.8, and we may expect that the hyperoptimization performed on the default NNPDF4.0 does not apply on the new PDFs shape. For these reasons we expect that as we vary Q_0 the PDFs will start differing but still being compatible within the uncertainties, while at a certain point the algorithm will break down giving results that are not compatible anymore with the default setting of Q_0 .

As a first step we started varying Q_0 between $m_c = 1.51$ GeV and $m_b = 4.92$ GeV in order to check that the results are stable within a given flavor number scheme (in this case $n_f = 4$). Figs.C.1-C.5 show the comparison between the fit performed with the default value of the fitting scale, i.e. $Q_0 = 1.65$, and fits performed with $Q_0 = 1.52, 2, 3, 4, 4.9$ GeV. We can observe that for the values of Q_0 that are closer to 1.65 GeV, i.e. $Q_0 = 1.52, 2$ GeV, the results are totally compatible with the default fit, with the differences below the percent level. As we increase the fitting scale, the differences with respect to the default fit increase, especially at small x . This is related to what we discussed earlier: different choices of the fitting scale should yield the same result up to higher orders differences and, being the PDFs band narrower at larger scales due to the evolution, the fitting code has more difficulties in finding the best configuration for the

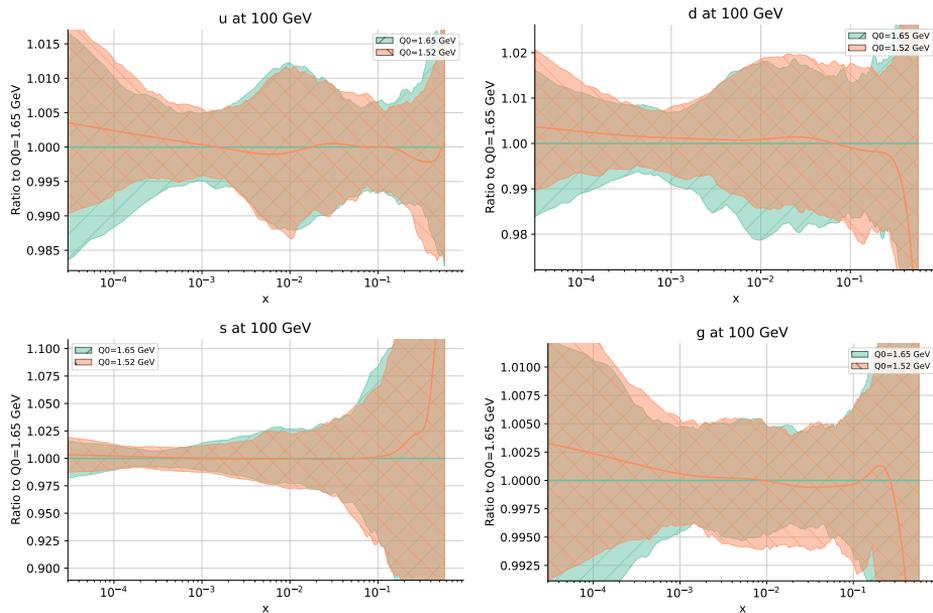


Figure C.1: Comparison between the up, down, strange and gluon PDFs fitted at $Q_0 = 1.52$ GeV (orange) and the default $Q_0 = 1.65$ GeV (green). The PDFs are shown at $Q = 100$ GeV and the plots are given as ratio plots with respect to the default value of Q_0 , i.e. $Q_0 = 1.65$. The uncertainty band corresponds to the 1σ uncertainties.

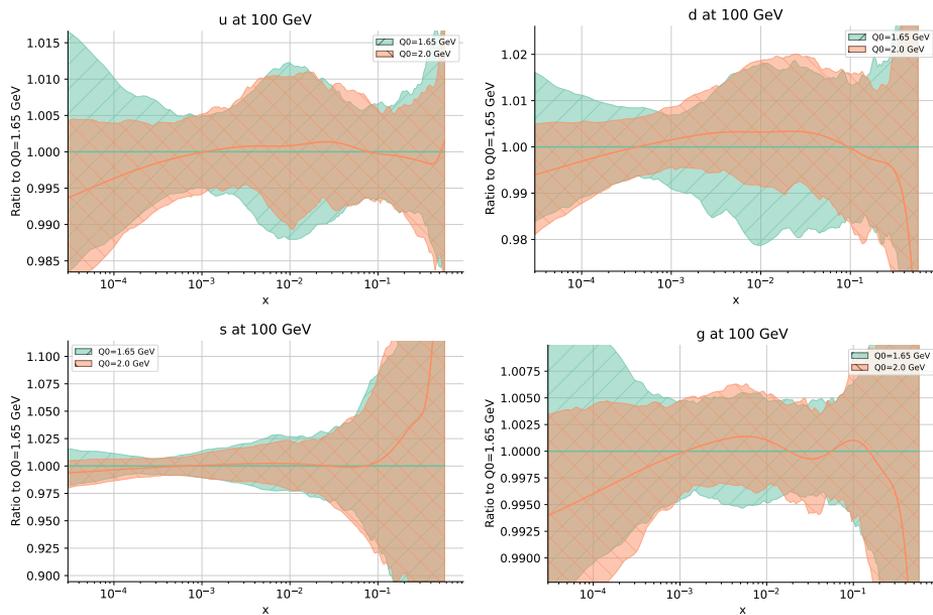


Figure C.2: As Fig. C.1 but for $Q_0 = 2.0$ GeV.

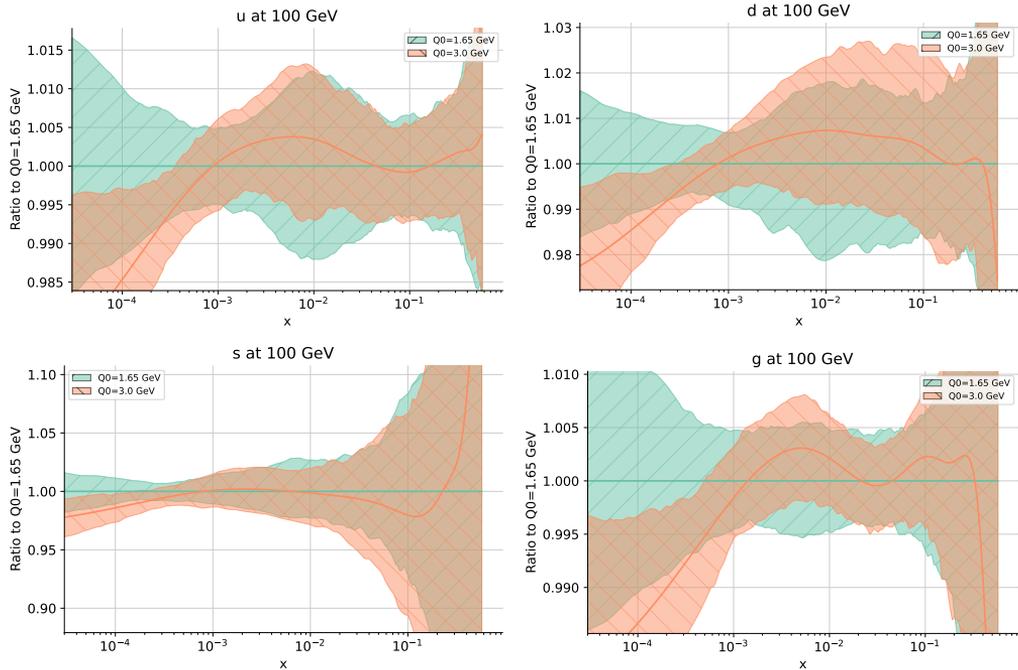


Figure C.3: As Fig. C.1 but for $Q_0 = 3$ GeV.

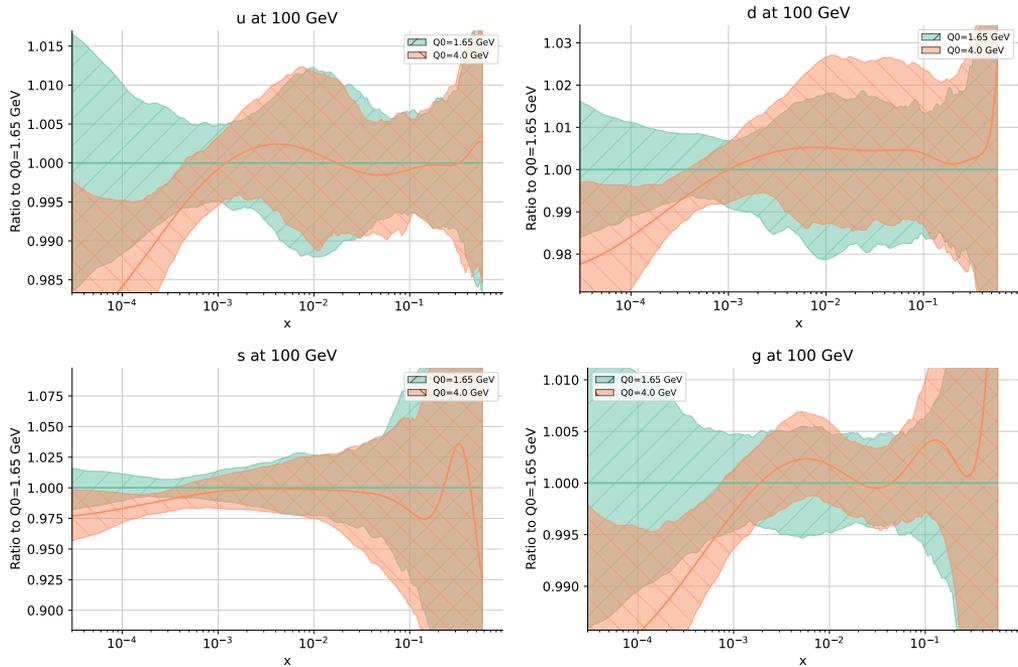


Figure C.4: As Fig. C.1 but for $Q_0 = 4$ GeV.

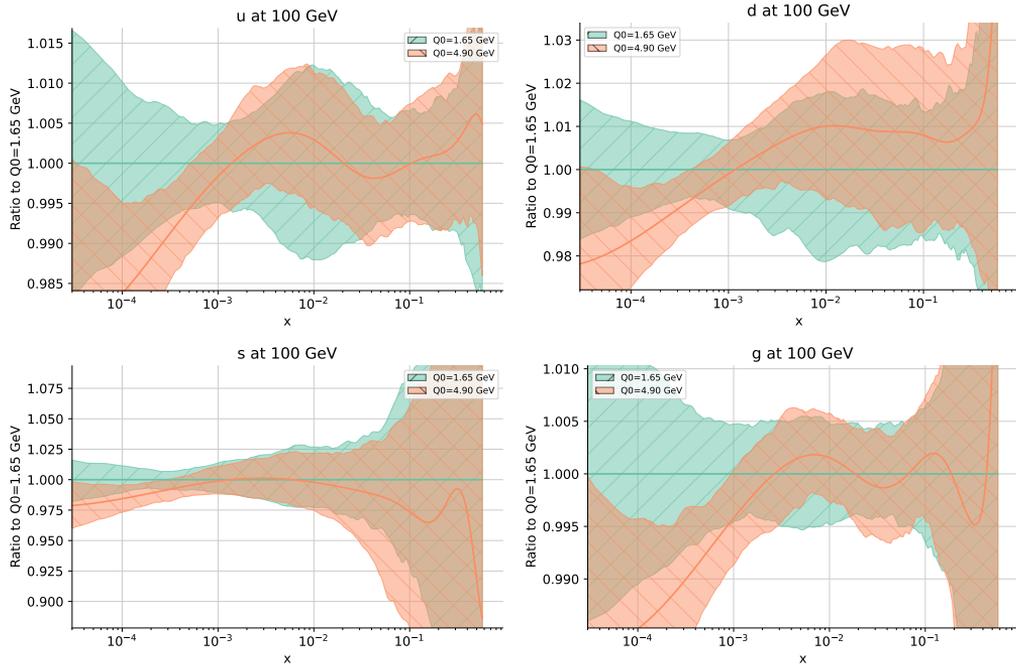


Figure C.5: As Fig. C.1 but for $Q_0 = 4.90$ GeV.

Q_0	χ^2	$\langle E_{\text{tr}} \rangle_{\text{rep}}$	$\langle E_{\text{val}} \rangle_{\text{rep}}$	$\langle \chi^2 \rangle_{\text{rep}}$	$\langle \text{TL} \rangle_{\text{rep}}$	ϕ
1.65 GeV (default)	1.171	2.28	2.37	1.20	12400	0.153
1.52 GeV	1.171	2.29	2.39	1.20	12300	0.155
2.0 GeV	1.170	2.26	2.35	1.19	12300	0.152
3.0 GeV	1.168	2.26	2.35	1.19	12800	0.150
4.0 GeV	1.170	2.26	2.35	1.19	12900	0.148
4.9 GeV	1.174	2.26	2.35	1.20	12600	0.148

Table C.1: Comparison of statistical indicators for a set of 100 NNPDF4.0 NNLO PDFs replicas generated with the default choice of Q_0 and with $Q_0 = 1.52, 2, 3, 4, 4.9$ GeV.

PDFs to agree with experimental data. Moreover, the larger differences are observed in the small x region, where there are no data. Therefore, there is no experimental value to constrain the PDFs in that region.

If we study the fit quality of the different fits in Tab. C.1 we can observe that the statistical indicators are almost unchanged for the different fits. The χ^2 shows a very small worsening as we increase Q_0 . This is related to the fact that, as we already discussed, at large Q_0 the fitting algorithm has more difficulties in finding the best configuration in the PDFs space. This behavior is observed also in the training lengths: at large Q_0 the fit takes slightly longer to converge. However, these differences with the default fit are minimal despite a larger differences in the PDFs. This is due to the fact that the PDFs change mainly in regions in which there are no data to constrain them. In the end, we conclude that varying Q_0 between the mass of the charm and the mass of the bottom,

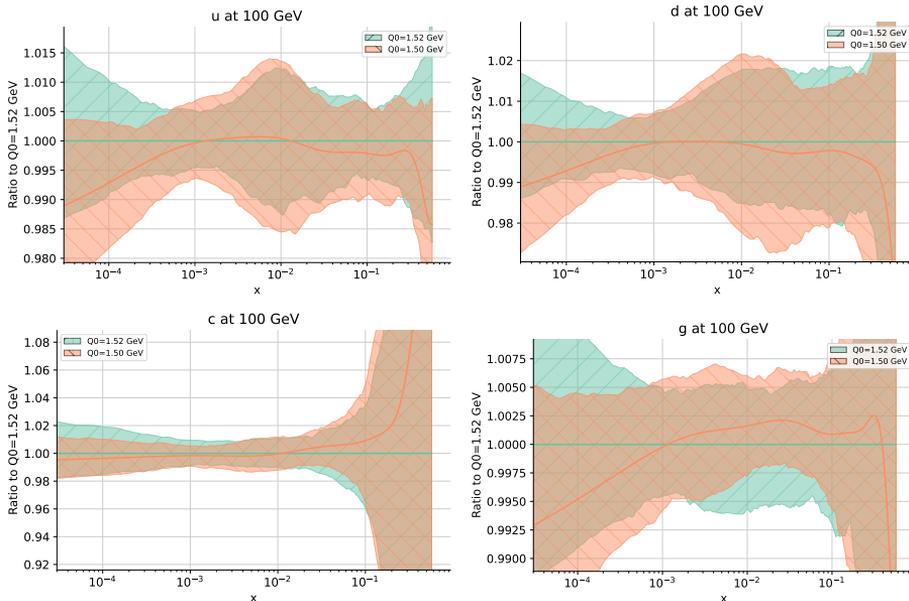


Figure C.6: Comparison between the fits obtained above ($Q_0 = 1.52$ GeV, green) and below ($Q_0 = 1.50$ GeV, orange) the charm threshold. The PDFs are plotted at 100 GeV and the plots are displayed as ratio plots with respect to the fit with $Q_0 = 1.52$ GeV.

Q_0	χ^2	$\langle E_{\text{tr}} \rangle_{\text{rep}}$	$\langle E_{\text{val}} \rangle_{\text{rep}}$	$\langle \chi^2 \rangle_{\text{rep}}$	$\langle \text{TL} \rangle_{\text{rep}}$	ϕ
1.50 GeV	1.166	2.26	2.34	1.19	13300	0.158
1.52 GeV	1.171	2.29	2.39	1.20	12300	0.155
4.90 GeV	1.174	2.26	2.35	1.20	12600	0.148
4.93 GeV	1.173	2.27	2.35	1.20	12900	0.150

Table C.2: Comparison of statistical indicators for a set of 100 NNPDF4.0 NNLO PDFs replicas generated with the fitting scale below and above the charm and bottom thresholds.

yields results that are fully compatible with the default choice of NNPDF4.0, with some differences that arise mainly in region where there are no data to constrain the PDFs.

The next thing we want to check is if crossing the charm or bottom threshold gives further problems to the fitting algorithm. Indeed, we expect that applying the matching conditions slightly changes the shape of the PDFs below and above the thresholds. This may result in problems for the fit since, as we said in Chap. 2, the fitting algorithm has been optimized on the PDFs at $Q_0 = 1.65$ GeV. Therefore, if applying the matching conditions, modifies such shapes in a considerable way, we may need to hyperoptimize again the hyperparameters of the neural network, to be able to have the same efficiency in the fitting procedure.

Figs. C.6-C.7 show the comparisons of the fits performed with the fitting scale below and above the charm and bottom thresholds, respectively. The values of Q_0 that have been adopted are $Q_0 = 1.50$ GeV and $Q_0 = 4.93$ GeV (we remind that $m_c = 1.51$ GeV and $m_b = 4.92$ GeV). In the case of the charm threshold we can observe very good

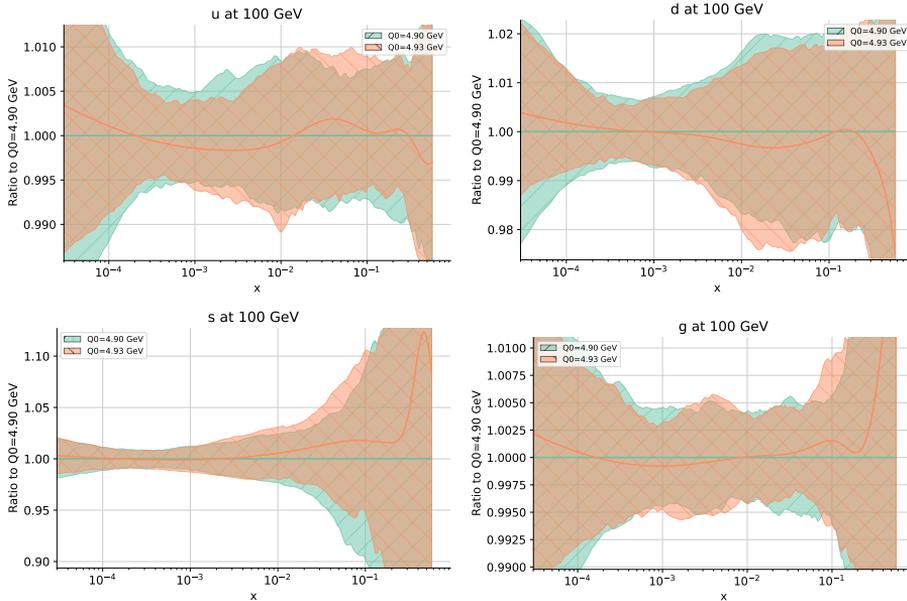


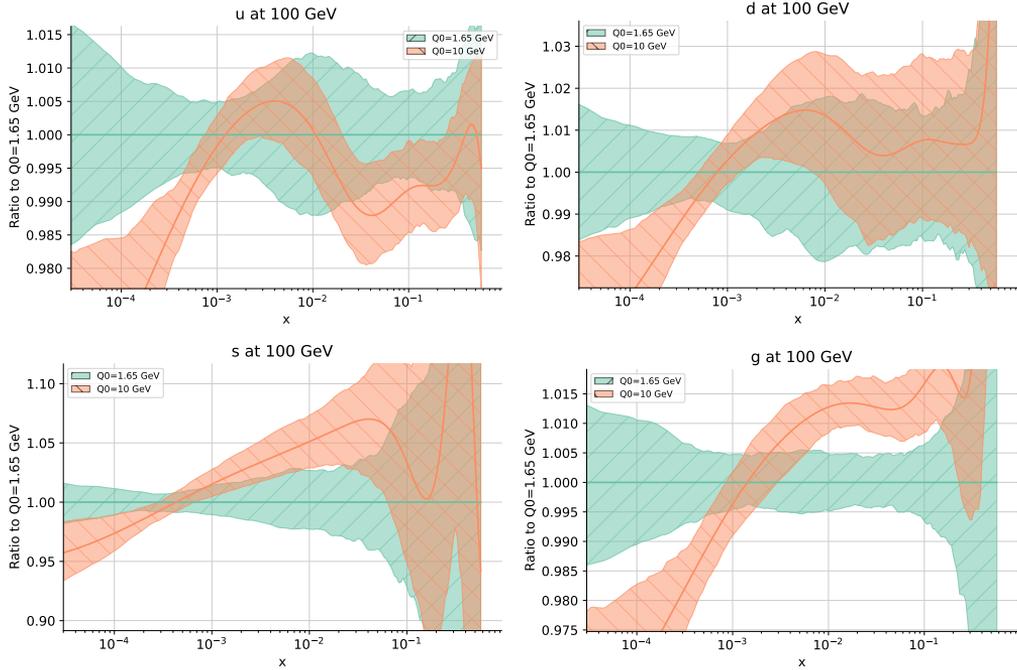
Figure C.7: Comparison between the fits obtained above ($Q_0 = 4.93$ GeV, orange) and below ($Q_0 = 4.90$ GeV, green) the bottom threshold. The PDFs are plotted at 100 GeV and the plots are displayed as ratio plots with respect to the fit with $Q_0 = 4.90$ GeV.

agreement in the PDFs, with the differences that are at the sub-percent level in all the x range, with the exception of the very small x values, where they reach the percent level. Observe that in Fig. C.6 we showed the charm PDF instead of the strange, as we were doing in the other plots of this appendix, in order to show that the charm PDF is not affected by performing the fit below m_c . Indeed, the differences in the two charms are below the percent level. Since, in this case, what we are fitting is exactly the intrinsic charm component, and not the sum of intrinsic and perturbative parts, as it happens performing the fit above m_c , we conclude that the fitting algorithm is able to correctly fit the charm in the two configurations. For the bottom threshold, the agreement is even better since it can be observed that the PDFs bands are completely overlapped. The slightly larger discrepancies observed in the case of the charm can be explained with the fact that at those scales, the value of the strong coupling isn't completely in a perturbative regime and therefore it results in larger uncertainties.

Tab. C.2 compares the statistical indicators of these fits with the ones immediately above (or below) the heavy quark thresholds. We can observe that all the values are compatible with the fits done in the $n_f = 4$ scheme. It is interesting to observe that the χ^2 of the fit done below the charm threshold is decreased of a considerable amount (but still being compatible with the one of the fit at $Q_0 = 1.52$ GeV).

We conclude that performing the PDFs fits for values of Q_0 that are in a different flavor number scheme with respect to the default one does not affect the quality of the fit and does not change the shapes of the PDFs. Therefore, applying the matching conditions to the PDFs doesn't change their shape in a considerable way, such that the fitting algorithm has to be re-hyperoptimized in order to perform well and fit the PDFs.

After concluding that applying the matching conditions does not worsen the quality of the fit, and therefore that the main difference comes from the choice of the value of


Figure C.8: As Fig. C.1 but for $Q_0 = 10$ GeV.

Q_0	χ^2	$\langle E_{\text{tr}} \rangle_{\text{rep}}$	$\langle E_{\text{val}} \rangle_{\text{rep}}$	$\langle \chi^2 \rangle_{\text{rep}}$	$\langle \text{TL} \rangle_{\text{rep}}$	ϕ
1.65 GeV (default)	1.171	2.28	2.37	1.20	12400	0.153
10 GeV	1.972	2.85	2.97	2.0	13100	0.151

Table C.3: Comparison of statistical indicators for a set of 100 NNPDF4.0 NNLO PDFs replicas generated with the default choice of Q_0 and with $Q_0 = 10$ GeV.

Q_0 (and not from the n_f scheme in which it is), we can perform a last check increasing again Q_0 and see if the fitting algorithm breaks down. For this reason we performed a last fit with $Q_0 = 10$ GeV. Fig. C.8 shows the comparison between the PDFs obtained in this case with the default choice of Q_0 . We can observe larger differences in all the x range for almost all the PDFs. As we explained previously, this is both because of differences coming from higher orders and for the difficulties that the fitting code has to fit the optimal configuration at a high scale. Tab. C.3 shows the statistical parameters of the fit at $Q_0 = 10$ GeV and compares them with the ones at $Q_0 = 1.65$ GeV. We can observe a large worsening of all the parameters but in particular of the χ^2 . We conclude that the fitting algorithm is not able anymore to efficiently perform the fit at such a high scale. Therefore, the optimal choice of the fitting scale for the PDFs is choosing Q_0 at small scales, close to the charm threshold (both above or below).

With this study we conclude that, investigating the PDFs dependence on the choice of the parametrization scale Q_0 , we have found that keeping it at low values is the most optimal choice since the fit is performed efficiently, with a small value of the χ^2 . Crossing heavy quarks thresholds does not worsen the quality of the fit since fit performed

slightly above or below a certain threshold provide results that are completely compatible within the uncertainties. However, increasing the value of Q_0 too much breaks the algorithm since the fit cannot be performed efficiently any longer and provides results that do not describe experimental data as good as fits done at a lower scale.

Bibliography

- [1] X. Cid Vidal et al., *Report from Working Group 3: Beyond the Standard Model physics at the HL-LHC and HE-LHC*, CERN Yellow Rep. Monogr. **7** (2019) 585–865, [arXiv:1812.07831].
- [2] **HL-LHC, HE-LHC Working Group** Collaboration, P. Azzi et al., *Standard Model Physics at the HL-LHC and HE-LHC*, arXiv:1902.04070.
- [3] **Physics of the HL-LHC Working Group** Collaboration, M. Cepeda et al., *Higgs Physics at the HL-LHC and HE-LHC*, arXiv:1902.00134.
- [4] **NNPDF** Collaboration, R. D. Ball et al., *The path to proton structure at 1% accuracy*, *Eur. Phys. J. C* **82** (2022), no. 5 428, [arXiv:2109.02653].
- [5] **NNPDF** Collaboration, R. D. Ball et al., *Parton distributions from high-precision collider data*, *Eur. Phys. J. C* **77** (2017), no. 10 663, [arXiv:1706.00428].
- [6] **NNPDF** Collaboration, R. D. Ball et al., *Photons in the proton: implications for the LHC*, *Eur. Phys. J. C* **84** (2024), no. 5 540, [arXiv:2401.08749].
- [7] **NNPDF** Collaboration, R. D. Ball et al., *Determination of the theory uncertainties from missing higher orders on NNLO parton distributions with percent accuracy*, *Eur. Phys. J. C* **84** (2024), no. 5 517, [arXiv:2401.10319].
- [8] **NNPDF** Collaboration, R. D. Ball et al., *The path to N^3 LO parton distributions*, *Eur. Phys. J. C* **84** (2024), no. 7 659, [arXiv:2402.18635].
- [9] H. Georgi, *Lie Algebras In Particle Physics: from Isospin To Unified Theories*, vol. 54. CRC Press, 1982.
- [10] M. E. Peskin and D. V. Schroeder, *An Introduction to quantum field theory*. Addison-Wesley, Reading, USA, 1995.
- [11] **Particle Data Group** Collaboration, R. L. Workman et al., *Review of Particle Physics*, *PTEP* **2022** (2022) 083C01.
- [12] T. Appelquist and J. Carazzone, *Infrared Singularities and Massive Fields*, *Phys. Rev. D* **11** (1975) 2856.

- [13] K. G. Chetyrkin, B. A. Kniehl, and M. Steinhauser, *Strong coupling constant with flavor thresholds at four loops in the modified minimal-subtraction scheme*, *Physical Review Letters* **79** (Sep, 1997) 2184–2187, [hep-ph/9706430].
- [14] R. K. Ellis, W. J. Stirling, and B. R. Webber, *QCD and collider physics*. Cambridge University Press, 1996.
- [15] J. Abate and P. Valko, *Multi-precision laplace transform inversion*, *International Journal for Numerical Methods in Engineering* **60** (2003) 979–993.
- [16] A. Vogt, *Efficient evolution of unpolarized and polarized parton distributions with qcd-pegasus*, *Comput. Phys. Commun.* **170** (2005) 65–92, [hep-ph/0408244].
- [17] M. Bonvini, *Resummation of soft and hard gluon radiation in perturbative QCD*. PhD thesis, Genoa U., 2012. arXiv:1212.0480.
- [18] M. A. G. Aivazis, F. I. Olness, and W.-K. Tung, *Leptoproduction of heavy quarks. I. General formalism and kinematics of charged current and neutral current production processes*, *Physical Review D* **50** (Sep, 1994) 3085–3101, [hep-ph/9312318].
- [19] M. A. G. Aivazis, J. C. Collins, F. I. Olness, and W.-K. Tung, *Leptoproduction of heavy quarks. II. A unified QCD formulation of charged and neutral current processes from fixed-target to collider energies*, *Physical Review D* **50** (Sep, 1994) 3102–3118, [hep-ph/9312319].
- [20] J. C. Collins, *Proof of factorization for diffractive hard scattering*, *Phys. Rev. D* **57** (Mar, 1998) 3051–3056, [hep-ph/9709499].
- [21] M. Krämer, F. I. Olness, and D. E. Soper, *Treatment of heavy quarks in deeply inelastic scattering*, *Physical Review D* **62** (Oct, 2000) [hep-ph/0003035].
- [22] R. S. Thorne and R. G. Roberts, *Ordered analysis of heavy flavor production in deep-inelastic scattering*, *Physical Review D* **57** (Jun, 1998) 6871–6898, [hep-ph/9709442].
- [23] R. S. Thorne, *Variable-flavor number scheme for next-to-next-to-leading order*, *Physical Review D* **73** (Mar, 2006) [hep-ph/0601245].
- [24] M. Buza, Y. Matiounine, J. Smith, and W. L. van Neerven, *Charm electroproduction viewed in the variable-flavour number scheme versus fixed-order perturbation theory*, *The European Physical Journal C* **1** (Mar, 1998) 301–320, [hep-ph/9612398].
- [25] M. Cacciari, M. Greco, and P. Nason, *The $p(T)$ spectrum in heavy-flavour hadroproduction*, *JHEP* **05** (1998) 007, [hep-ph/9803400].
- [26] S. Forte, E. Laenen, P. Nason, and J. Rojo, *Heavy quarks in deep-inelastic scattering*, *Nucl. Phys.* **B834** (2010) 116–162, [arXiv:1001.2312].
- [27] **The NNPDF Collaboration**, R. D. Ball et al., *Precision determination of electroweak parameters and the strange content of the proton from neutrino deep-inelastic scattering*, *Nucl. Phys.* **B823** (2009) 195–233, [arXiv:0906.1958].
- [28] **NNPDF Collaboration**, R. D. Ball et al., *Parton distributions for the LHC Run II*, *JHEP* **04** (2015) 040, [arXiv:1410.8849].

- [29] NNPDF Collaboration, R. D. Ball, A. Candido, J. Cruz-Martinez, S. Forte, T. Giani, F. Hekhorn, K. Kudashkin, G. Magni, and J. Rojo, *Evidence for intrinsic charm quarks in the proton*, *Nature* **608** (2022), no. 7923 483–487, [arXiv:2208.08372].
- [30] NNPDF Collaboration, R. D. Ball, A. Candido, J. Cruz-Martinez, S. Forte, T. Giani, F. Hekhorn, G. Magni, E. R. Nocera, J. Rojo, and R. Stegeman, *Intrinsic charm quark valence distribution of the proton*, *Phys. Rev. D* **109** (2024), no. 9 L091501, [arXiv:2311.00743].
- [31] G. D’Agostini, *Bayesian reasoning in data analysis: A critical introduction*. World Scientific, 2003.
- [32] A. Candido, S. Forte, and F. Hekhorn, *Can \overline{MS} parton distributions be negative?*, *JHEP* **11** (2020) 129, [arXiv:2006.07377].
- [33] A. Buckley, J. Ferrando, S. Lloyd, K. Nordström, B. Page, et al., *LHAPDF6: parton density access in the LHC precision era*, *Eur.Phys.J.* **C75** (2015) 132, [arXiv:1412.7420].
- [34] New Muon Collaboration, M. Arneodo et al., *Accurate measurement of F_2^d/F_2^p and $R_d - R_p$* , *Nucl. Phys.* **B487** (1997) 3–26, [hep-ex/9611022].
- [35] New Muon Collaboration, M. Arneodo et al., *Measurement of the proton and deuteron structure functions, F_2^p and F_2^d , and of the ratio σ_L/σ_T* , *Nucl. Phys.* **B483** (1997) 3–43, [hep-ph/9610231].
- [36] L. W. Whitlow, E. M. Riordan, S. Dasu, S. Rock, and A. Bodek, *Precise measurements of the proton and deuteron structure functions from a global analysis of the SLAC deep inelastic electron scattering cross-sections*, *Phys. Lett.* **B282** (1992) 475–482.
- [37] BCDMS Collaboration, A. C. Benvenuti et al., *A High Statistics Measurement of the Proton Structure Functions $F_2(x, Q^2)$ and R from Deep Inelastic Muon Scattering at High Q^2* , *Phys. Lett.* **B223** (1989) 485.
- [38] CHORUS Collaboration, G. Onengut et al., *Measurement of nucleon structure functions in neutrino scattering*, *Phys. Lett.* **B632** (2006) 65–75.
- [39] NuTeV Collaboration, M. Goncharov et al., *Precise measurement of dimuon production cross-sections in $\nu_\mu Fe$ and $\bar{\nu}_\mu Fe$ deep inelastic scattering at the Tevatron*, *Phys. Rev.* **D64** (2001) 112006, [hep-ex/0102049].
- [40] D. A. Mason, *Measurement of the strange - antistrange asymmetry at NLO in QCD from NuTeV dimuon data*, . FERMILAB-THESIS-2006-01.
- [41] NOMAD Collaboration, O. Samoylov et al., *A Precision Measurement of Charm Dimuon Production in Neutrino Interactions from the NOMAD Experiment*, *Nucl.Phys.* **B876** (2013) 339, [arXiv:1308.4750].
- [42] European Muon Collaboration, J. J. Aubert et al., *Production of charmed particles in 250-GeV μ^+ - iron interactions*, *Nucl. Phys.* **B213** (1983) 31–64.
- [43] ZEUS, H1 Collaboration, H. Abramowicz et al., *Combination of measurements of inclusive deep inelastic $e^\pm p$ scattering cross sections and QCD analysis of HERA data*, *Eur. Phys. J.* **C75** (2015), no. 12 580, [arXiv:1506.06042].

- [44] **H1, ZEUS** Collaboration, H. Abramowicz et al., *Combination and QCD analysis of charm and beauty production cross-section measurements in deep inelastic ep scattering at HERA*, *Eur. Phys. J.* **C78** (2018), no. 6 473, [arXiv:1804.01019].
- [45] **ZEUS** Collaboration, S. Chekanov et al., *Inclusive jet cross-sections in the Breit frame in neutral current deep inelastic scattering at HERA and determination of α_s* , *Phys. Lett. B* **547** (2002) 164–180, [hep-ex/0208037].
- [46] **ZEUS** Collaboration, S. Chekanov et al., *Inclusive-jet and dijet cross-sections in deep inelastic scattering at HERA*, *Nucl. Phys. B* **765** (2007) 1–30, [hep-ex/0608048].
- [47] **H1** Collaboration, V. Andreev et al., *Measurement of Jet Production Cross Sections in Deep-inelastic ep Scattering at HERA*, *Eur. Phys. J. C* **77** (2017), no. 4 215, [arXiv:1611.03421].
- [48] **H1** Collaboration, V. Andreev et al., *Measurement of multijet production in ep collisions at high Q^2 and determination of the strong coupling α_s* , *Eur. Phys. J. C* **75** (2015), no. 2 65, [arXiv:1406.4709].
- [49] **ZEUS** Collaboration, H. Abramowicz et al., *Inclusive dijet cross sections in neutral current deep inelastic scattering at HERA*, *Eur. Phys. J. C* **70** (2010) 965–982, [arXiv:1010.6167].
- [50] **FNAL E866/NuSea** Collaboration, R. S. Towell et al., *Improved measurement of the anti-d/anti-u asymmetry in the nucleon sea*, *Phys. Rev.* **D64** (2001) 052002, [hep-ex/0103030].
- [51] **NuSea** Collaboration, J. C. Webb et al., *Absolute Drell-Yan dimuon cross sections in 800-GeV/c p p and p d collisions*, hep-ex/0302019.
- [52] G. Moreno et al., *Dimuon production in proton - copper collisions at $\sqrt{s} = 38.8$ -GeV*, *Phys. Rev.* **D43** (1991) 2815–2836.
- [53] **SeaQuest** Collaboration, J. Dove et al., *The asymmetry of antimatter in the proton*, *Nature* **590** (2021), no. 7847 561–565, [arXiv:2103.04024].
- [54] S. Alekhin, J. Blümlein, S. Moch, and R. Placakyte, *Parton distribution functions, α_s , and heavy-quark masses for LHC Run II*, *Phys. Rev.* **D96** (2017), no. 1 014011, [arXiv:1701.05838].
- [55] T.-J. Hou et al., *New CTEQ global analysis of quantum chromodynamics with high-precision data from the LHC*, *Phys. Rev. D* **103** (2021), no. 1 014013, [arXiv:1912.10053].
- [56] S. Bailey, T. Cridge, L. A. Harland-Lang, A. D. Martin, and R. S. Thorne, *Parton distributions from LHC, HERA, Tevatron and fixed target data: MSHT20 PDFs*, *Eur. Phys. J. C* **81** (2021), no. 4 341, [arXiv:2012.04684].
- [57] **CDF** Collaboration, T. A. Aaltonen et al., *Measurement of $d\sigma/dy$ of Drell-Yan e^+e^- pairs in the Z Mass Region from $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV*, *Phys. Lett.* **B692** (2010) 232–239, [arXiv:0908.3914].
- [58] **D0** Collaboration, V. M. Abazov et al., *Measurement of the shape of the boson rapidity distribution for $p\bar{p} \rightarrow Z/\gamma^* \rightarrow e^+e^- + X$ events produced at $\sqrt{s}=1.96$ -TeV*, *Phys. Rev.* **D76** (2007) 012003, [hep-ex/0702025].

- [59] **D0** Collaboration, V. M. Abazov et al., *Measurement of the muon charge asymmetry in $p\bar{p} \rightarrow W+X \rightarrow \mu\nu + X$ events at $\sqrt{s}=1.96$ TeV*, *Phys.Rev.* **D88** (2013) 091102, [arXiv:1309.2591].
- [60] **D0** Collaboration, V. M. Abazov et al., *Measurement of the electron charge asymmetry in $p\bar{p} \rightarrow W + X \rightarrow e\nu + X$ decays in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV*, *Phys. Rev.* **D91** (2015), no. 3 032007, [arXiv:1412.2862]. [Erratum: *Phys. Rev.* **D91**,no.7,079901(2015)].
- [61] **ATLAS** Collaboration, G. Aad et al., *Measurement of the low-mass Drell-Yan differential cross section at $\sqrt{s} = 7$ TeV using the ATLAS detector*, *JHEP* **06** (2014) 112, [arXiv:1404.1212].
- [62] **ATLAS** Collaboration, G. Aad et al., *Measurement of the high-mass Drell-Yan differential cross-section in pp collisions at $\sqrt{s}=7$ TeV with the ATLAS detector*, *Phys.Lett.* **B725** (2013) 223, [arXiv:1305.4192].
- [63] **ATLAS** Collaboration, G. Aad et al., *Measurement of the inclusive W^\pm and Z/γ^* cross sections in the electron and muon decay channels in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, *Phys.Rev.* **D85** (2012) 072004, [arXiv:1109.5141].
- [64] **ATLAS** Collaboration, M. Aaboud et al., *Precision measurement and interpretation of inclusive W^+ , W^- and Z/γ^* production cross sections with the ATLAS detector*, *Eur. Phys. J.* **C77** (2017), no. 6 367, [arXiv:1612.03016].
- [65] **CMS** Collaboration, S. Chatrchyan et al., *Measurement of the electron charge asymmetry in inclusive W production in pp collisions at $\sqrt{s} = 7$ TeV*, *Phys.Rev.Lett.* **109** (2012) 111806, [arXiv:1206.2598].
- [66] **CMS** Collaboration, S. Chatrchyan et al., *Measurement of the muon charge asymmetry in inclusive pp to WX production at $\sqrt{s} = 7$ TeV and an improved determination of light parton distribution functions*, *Phys.Rev.* **D90** (2014) 032004, [arXiv:1312.6283].
- [67] **CMS** Collaboration, S. Chatrchyan et al., *Measurement of the differential and double-differential Drell-Yan cross sections in proton-proton collisions at $\sqrt{s} = 7$ TeV*, *JHEP* **1312** (2013) 030, [arXiv:1310.7291].
- [68] **LHCb** Collaboration, R. Aaij et al., *Measurement of the cross-section for $Z \rightarrow e^+e^-$ production in pp collisions at $\sqrt{s} = 7$ TeV*, *JHEP* **1302** (2013) 106, [arXiv:1212.4620].
- [69] **LHCb** Collaboration, R. Aaij et al., *Measurement of the forward Z boson production cross-section in pp collisions at $\sqrt{s} = 7$ TeV*, *JHEP* **08** (2015) 039, [arXiv:1505.07024].
- [70] **ATLAS** Collaboration, G. Aad et al., *Measurement of the cross-section and charge asymmetry of W bosons produced in proton-proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, *Eur. Phys. J. C* **79** (2019), no. 9 760, [arXiv:1904.05631].
- [71] **ATLAS** Collaboration, M. Aaboud et al., *Measurement of the Drell-Yan triple-differential cross section in pp collisions at $\sqrt{s} = 8$ TeV*, *JHEP* **12** (2017) 059, [arXiv:1710.05167].

- [72] **ATLAS** Collaboration, G. Aad et al., *Measurement of the double-differential high-mass Drell-Yan cross section in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, *JHEP* **08** (2016) 009, [arXiv:1606.01736].
- [73] **CMS** Collaboration, V. Khachatryan et al., *Measurement of the Z boson differential cross section in transverse momentum and rapidity in proton-proton collisions at 8 TeV*, *Phys. Lett.* **B749** (2015) 187–209, [arXiv:1504.03511].
- [74] **LHCb** Collaboration, R. Aaij et al., *Measurement of forward $Z \rightarrow e^+e^-$ production at $\sqrt{s} = 8$ TeV*, *JHEP* **05** (2015) 109, [arXiv:1503.00963].
- [75] **LHCb** Collaboration, R. Aaij et al., *Measurement of forward W and Z boson production in pp collisions at $\sqrt{s} = 8$ TeV*, *JHEP* **01** (2016) 155, [arXiv:1511.08039].
- [76] **LHCb** Collaboration, R. Aaij et al., *Measurement of forward $W \rightarrow e\nu$ production in pp collisions at $\sqrt{s} = 8$ TeV*, *JHEP* **10** (2016) 030, [arXiv:1608.01484].
- [77] **ATLAS** Collaboration, G. Aad et al., *Measurement of W^\pm and Z-boson production cross sections in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *Phys. Lett.* **B759** (2016) 601–621, [arXiv:1603.09222].
- [78] **LHCb** Collaboration, R. Aaij et al., *Measurement of the forward Z boson production cross-section in pp collisions at $\sqrt{s} = 13$ TeV*, *JHEP* **09** (2016) 136, [arXiv:1607.06495].
- [79] **ATLAS** Collaboration, G. Aad et al., *Measurement of the production of a W boson in association with a charm quark in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, *JHEP* **1405** (2014) 068, [arXiv:1402.6263].
- [80] **CMS** Collaboration, S. Chatrchyan et al., *Measurement of associated W + charm production in pp collisions at $\sqrt{s} = 7$ TeV*, *JHEP* **02** (2014) 013, [arXiv:1310.1138].
- [81] **CMS** Collaboration, A. M. Sirunyan et al., *Measurement of associated production of a W boson and a charm quark in proton-proton collisions at $\sqrt{s} = 13$ TeV*, *Eur. Phys. J. C* **79** (2019), no. 3 269, [arXiv:1811.10021].
- [82] **ATLAS** Collaboration, M. Aaboud et al., *Measurement of differential cross sections and W^+/W^- cross-section ratios for W boson production in association with jets at $\sqrt{s} = 8$ TeV with the ATLAS detector*, *JHEP* **05** (2018) 077, [arXiv:1711.03296]. [Erratum: *JHEP* 10, 048 (2020)].
- [83] **ATLAS** Collaboration, G. Aad et al., *Measurement of the transverse momentum and ϕ_η^* distributions of Drell-Yan lepton pairs in proton-proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, *Eur. Phys. J.* **C76** (2016), no. 5 291, [arXiv:1512.02192].
- [84] **CMS** Collaboration, A. M. Sirunyan et al., *Measurement of the inclusive $t\bar{t}$ cross section in pp collisions at $\sqrt{s} = 5.02$ TeV using final states with at least one charged lepton*, *JHEP* **03** (2018) 115, [arXiv:1711.03143].
- [85] **ATLAS** Collaboration, G. Aad et al., *Measurement of the $t\bar{t}$ production cross-section using $e\mu$ events with b-tagged jets in pp collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS detector*, *Eur. Phys. J.* **C74** (2014), no. 10 3109, [arXiv:1406.5375]. [Addendum: *Eur. Phys. J.* C76, no. 11, 642 (2016)].

- [86] S. Spannagel, *Top quark mass measurements with the CMS experiment at the LHC*, PoS **DIS2016** (2016) 150, [arXiv:1607.04972].
- [87] **ATLAS** Collaboration, G. Aad et al., *Measurement of the $t\bar{t}$ production cross-section in the lepton+jets channel at $\sqrt{s} = 13$ TeV with the ATLAS experiment*, *Phys. Lett. B* **810** (2020) 135797, [arXiv:2006.13076].
- [88] **CMS** Collaboration, V. Khachatryan et al., *Measurement of the top quark pair production cross section in proton-proton collisions at $\sqrt{s} = 13$ TeV*, *Phys. Rev. Lett.* **116** (2016), no. 5 052002, [arXiv:1510.05302].
- [89] **ATLAS** Collaboration, G. Aad et al., *Measurements of top-quark pair differential cross-sections in the lepton+jets channel in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector*, *Eur. Phys. J.* **C76** (2016), no. 10 538, [arXiv:1511.04716].
- [90] **ATLAS** Collaboration, M. Aaboud et al., *Measurement of top quark pair differential cross-sections in the dilepton channel in pp collisions at $\sqrt{s} = 7$ and 8 TeV with ATLAS*, *Phys. Rev. D* **94** (2016), no. 9 092003, [arXiv:1607.07281]. [Addendum: *Phys.Rev.D* 101, 119901 (2020)].
- [91] **CMS** Collaboration, V. Khachatryan et al., *Measurement of the differential cross section for top quark pair production in pp collisions at $\sqrt{s} = 8$ TeV*, *Eur. Phys. J.* **C75** (2015), no. 11 542, [arXiv:1505.04480].
- [92] **CMS** Collaboration, A. M. Sirunyan et al., *Measurement of double-differential cross sections for top quark pair production in pp collisions at $\sqrt{s} = 8$ TeV and impact on parton distribution functions*, *Eur. Phys. J.* **C77** (2017), no. 7 459, [arXiv:1703.01630].
- [93] **CMS** Collaboration, A. M. Sirunyan et al., *Measurement of differential cross sections for the production of top quark pairs and of additional jets in lepton+jets events from pp collisions at $\sqrt{s} = 13$ TeV*, *Phys. Rev.* **D97** (2018), no. 11 112003, [arXiv:1803.08856].
- [94] **CMS** Collaboration, A. M. Sirunyan et al., *Measurements of $t\bar{t}$ differential cross sections in proton-proton collisions at $\sqrt{s} = 13$ TeV using events containing two leptons*, *JHEP* **02** (2019) 149, [arXiv:1811.06625].
- [95] **ATLAS** Collaboration, G. Aad et al., *Measurement of the inclusive jet cross-section in proton-proton collisions at $\sqrt{s} = 7$ TeV using 4.5 fb^{-1} of data with the ATLAS detector*, *JHEP* **02** (2015) 153, [arXiv:1410.8857]. [Erratum: *JHEP*09,141(2015)].
- [96] **CMS** Collaboration, S. Chatrchyan et al., *Measurement of the Ratio of Inclusive Jet Cross Sections using the Anti- k_T Algorithm with Radius Parameters $R=0.5$ and 0.7 in pp Collisions at $\sqrt{s} = 7$ TeV*, *Phys. Rev. D* **90** (2014), no. 7 072006, [arXiv:1406.0324].
- [97] **ATLAS** Collaboration, M. Aaboud et al., *Measurement of the inclusive jet cross-sections in proton-proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, *JHEP* **09** (2017) 020, [arXiv:1706.03192].
- [98] **CMS** Collaboration, V. Khachatryan et al., *Measurement and QCD analysis of double-differential inclusive jet cross sections in pp collisions at $\sqrt{s} = 8$ TeV and cross section ratios to 2.76 and 7 TeV*, *JHEP* **03** (2017) 156, [arXiv:1609.05331].

- [99] **ATLAS Collaboration** Collaboration, G. Aad et al., *Measurement of dijet cross sections in pp collisions at 7 TeV centre-of-mass energy using the ATLAS detector*, *JHEP* **1405** (2014) 059, [arXiv:1312.3524].
- [100] **CMS Collaboration**, S. Chatrchyan et al., *Measurements of differential jet cross sections in proton-proton collisions at $\sqrt{s} = 7$ TeV with the CMS detector*, *Phys.Rev.* **D87** (2013) 112002, [arXiv:1212.6660].
- [101] **CMS Collaboration**, A. M. Sirunyan et al., *Measurement of the triple-differential dijet cross section in proton-proton collisions at $\sqrt{s} = 8$ TeV and constraints on parton distribution functions*, *Eur. Phys. J. C* **77** (2017), no. 11 746, [arXiv:1705.02628].
- [102] **ATLAS Collaboration**, G. Aad et al., *Measurement of the inclusive isolated prompt photon cross section in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, *JHEP* **08** (2016) 005, [arXiv:1605.03495].
- [103] **ATLAS Collaboration**, M. Aaboud et al., *Measurement of the cross section for inclusive isolated-photon production in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector*, *Phys. Lett. B* **770** (2017) 473–493, [arXiv:1701.06882].
- [104] **ATLAS Collaboration**, G. Aad et al., *Comprehensive measurements of t -channel single top-quark production cross sections at $\sqrt{s} = 7$ TeV with the ATLAS detector*, *Phys. Rev. D* **90** (2014), no. 11 112006, [arXiv:1406.7844].
- [105] **CMS Collaboration**, S. Chatrchyan et al., *Measurement of the Single-Top-Quark t -Channel Cross Section in pp Collisions at $\sqrt{s} = 7$ TeV*, *JHEP* **12** (2012) 035, [arXiv:1209.4533].
- [106] **ATLAS Collaboration**, M. Aaboud et al., *Fiducial, total and differential cross-section measurements of t -channel single top-quark production in pp collisions at 8 TeV using data collected by the ATLAS detector*, *Eur. Phys. J. C* **77** (2017), no. 8 531, [arXiv:1702.02859].
- [107] **CMS Collaboration**, V. Khachatryan et al., *Measurement of the t -channel single-top-quark production cross section and of the $|V_{tb}|$ CKM matrix element in pp collisions at $\sqrt{s} = 8$ TeV*, *JHEP* **06** (2014) 090, [arXiv:1403.7366].
- [108] **ATLAS Collaboration**, M. Aaboud et al., *Measurement of the inclusive cross-sections of single top-quark and top-antiquark t -channel production in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *JHEP* **04** (2017) 086, [arXiv:1609.03920].
- [109] **CMS Collaboration**, A. M. Sirunyan et al., *Cross section measurement of t -channel single top quark production in pp collisions at $\sqrt{s} = 13$ TeV*, *Phys. Lett. B* **772** (2017) 752–776, [arXiv:1610.00678].
- [110] V. Bertone, S. Carrazza, and J. Rojo, *APFEL: A PDF Evolution Library with QED corrections*, *Comput.Phys.Commun.* **185** (2014) 1647, [arXiv:1310.1394].
- [111] V. Bertone, S. Carrazza, and N. P. Hartland, *APFELgrid: a high performance tool for parton density determinations*, *Comput. Phys. Commun.* **212** (2017) 205–209, [arXiv:1605.02070].
- [112] T. Carli et al., *A posteriori inclusion of parton density functions in NLO QCD final-state calculations at hadron colliders: The APPLGRID Project*, *Eur.Phys.J.* **C66** (2010) 503, [arXiv:0911.2985].

- [113] **fastNLO** Collaboration, M. Wobisch, D. Britzger, T. Kluge, K. Rabbertz, and F. Stober, *Theory-Data Comparisons for Jet Measurements in Hadron-Induced Processes*, arXiv:1109.1310.
- [114] A. Barontini, A. Candido, J. M. Cruz-Martinez, F. Hekhorn, and C. Schwan, *Pineline: Industrialization of high-energy theory predictions*, *Comput. Phys. Commun.* **297** (2024) 109061, [arXiv:2302.12124].
- [115] A. Candido, F. Hekhorn, G. Magni, T. R. Rabemananjara, and R. Stegeman, *Yadism: Yet Another Deep-Inelastic Scattering Module*, arXiv:2401.15187.
- [116] A. Candido, A. Garcia, G. Magni, T. Rabemananjara, J. Rojo, and R. Stegeman, *Neutrino Structure Functions from GeV to EeV Energies*, *JHEP* **05** (2023) 149, [arXiv:2302.08527].
- [117] A. Barontini, A. Candido, F. Hekhorn, G. Magni, and R. Stegeman, *An FONLL prescription with coexisting flavor number PDFs*, arXiv:2408.07383.
- [118] C. Anastasiou, L. J. Dixon, K. Melnikov, and F. Petriello, *High precision QCD at hadron colliders: Electroweak gauge boson rapidity distributions at NNLO*, *Phys. Rev.* **D69** (2004) 094008, [hep-ph/0312266].
- [119] **NNPDF** Collaboration, R. D. Ball et al., *Parton distributions for the LHC Run II*, *JHEP* **04** (2015) 040, [arXiv:1410.8849].
- [120] **The NNPDF** Collaboration, R. D. Ball et al., *A first unbiased global NLO determination of parton distributions and their uncertainties*, *Nucl. Phys.* **B838** (2010) 136, [arXiv:1002.4407].
- [121] A. D. Martin, R. G. Roberts, W. J. Stirling, and R. S. Thorne, *Parton distributions incorporating QED contributions*, *Eur. Phys. J.* **C39** (2005) 155, [hep-ph/0411040].
- [122] C. Schmidt, J. Pumplin, D. Stump, and C. P. Yuan, *CT14QED parton distribution functions from isolated photon production in deep inelastic scattering*, *Phys. Rev.* **D93** (2016), no. 11 114015, [arXiv:1509.02905].
- [123] R. D. Ball et al., *Parton distributions with LHC data*, *Nucl.Phys.* **B867** (2013) 244, [arXiv:1207.1303].
- [124] **NNPDF** Collaboration, R. D. Ball et al., *Parton distributions with QED corrections*, *Nucl.Phys.* **B877** (2013) 290–320, [arXiv:1308.0598].
- [125] **xFitter Developers’ Team** Collaboration, F. Giuli et al., *The photon PDF from high-mass Drell–Yan data at the LHC*, *Eur. Phys. J.* **C77** (2017), no. 6 400, [arXiv:1701.08553].
- [126] A. Manohar, P. Nason, G. P. Salam, and G. Zanderighi, *How bright is the proton? A precise determination of the photon parton distribution function*, *Phys. Rev. Lett.* **117** (2016), no. 24 242002, [arXiv:1607.04266].
- [127] A. V. Manohar, P. Nason, G. P. Salam, and G. Zanderighi, *The Photon Content of the Proton*, *JHEP* **12** (2017) 046, [arXiv:1708.01256].
- [128] L. A. Harland-Lang, V. A. Khoze, and M. G. Ryskin, *Photon-initiated processes at high mass*, *Phys. Rev. D* **94** (2016), no. 7 074008, [arXiv:1607.04635].

- [129] L. A. Harland-Lang, V. A. Khoze, and M. G. Ryskin, *Sudakov effects in photon-initiated processes*, *Phys. Lett.* **B761** (2016) 20–24, [arXiv:1605.04935].
- [130] **NNPDF** Collaboration, V. Bertone, S. Carrazza, N. P. Hartland, and J. Rojo, *Illuminating the photon content of the proton within a global PDF analysis*, *SciPost Phys.* **5** (2018), no. 1 008, [arXiv:1712.07053].
- [131] L. A. Harland-Lang, A. D. Martin, R. Nathvani, and R. S. Thorne, *Ad Lucem: QED Parton Distribution Functions in the MMHT Framework*, *Eur. Phys. J. C* **79** (2019), no. 10 811, [arXiv:1907.02750].
- [132] T. Cridge, L. A. Harland-Lang, A. D. Martin, and R. S. Thorne, *QED parton distribution functions in the MSHT20 fit*, *Eur. Phys. J. C* **82** (2022), no. 1 90, [arXiv:2111.05357].
- [133] K. Xie, T. Hobbs, T.-J. Hou, C. Schmidt, M. Yan, and C.-P. Yuan, *The photon content of the proton in the CT18 global analysis*, *SciPost Phys. Proc.* **8** (2022) 074, [arXiv:2107.13580].
- [134] K. Xie, B. Zhou, and T. J. Hobbs, *The Photon Content of the Neutron*, arXiv:2305.10497.
- [135] G. Ricco, S. Simula, and M. Battaglieri, *Power corrections in the longitudinal and transverse structure functions of proton and deuteron*, *Nuclear Physics B* **555** (Aug., 1999) 306–334.
- [136] **CLAS** Collaboration, M. Osipenko et al., *A Kinematically complete measurement of the proton structure function F_2 in the resonance region and evaluation of its moments*, *Phys. Rev.* **D67** (2003) 092001, [hep-ph/0301204].
- [137] M. E. Christy and P. E. Bosted, *Empirical fit to precision inclusive electron-proton cross-sections in the resonance region*, *Phys. Rev.* **C81** (2010) 055213, [arXiv:0712.3731].
- [138] **HERMES** Collaboration, A. Airapetian et al., *Inclusive Measurements of Inelastic Electron and Positron Scattering from Unpolarized Hydrogen and Deuterium Targets*, *JHEP* **05** (2011) 126, [arXiv:1103.5704].
- [139] H. Abramowicz, E. M. Levin, A. Levy, and U. Maor, *A Parametrization of $\sigma(\gamma^* p)$ above the resonance region $Q^2 \geq 0$* , *Phys. Lett.* **B269** (1991) 465–476.
- [140] **A1** Collaboration, J. C. Bernauer et al., *Electric and magnetic form factors of the proton*, *Phys. Rev.* **C90** (2014), no. 1 015206, [arXiv:1307.6227].
- [141] **Jefferson Lab Hall C E94-110** Collaboration, Y. Liang et al., *Measurement of $R = \sigma(L) / \sigma(T)$ and the separated longitudinal and transverse structure functions in the nucleon resonance region*, nucl-ex/0410027.
- [142] **E143** Collaboration, K. Abe et al., *Measurements of $R = \sigma(L) / \sigma(T)$ for $0.03 \leq x \leq 0.1$ and fit to world data*, *Phys. Lett.* **B452** (1999) 194–200, [hep-ex/9808028].
- [143] A. M. Cooper-Sarkar, I. Abt, B. Foster, M. Wing, V. Myronenko, and K. Wichmann, *Study of HERA data at Low Q^2 and Low x* , *PoS DIS2016* (2016) 013, [arXiv:1605.08577].

- [144] L. A. Harland-Lang, A. D. Martin, P. Motylinski, and R. S. Thorne, *The impact of the final HERA combined data on PDFs obtained from a global fit*, *Eur. Phys. J. C* **76** (2016), no. 4 186, [arXiv:1601.03413].
- [145] S. Carrazza, S. Forte, Z. Kassabov, J. I. Latorre, and J. Rojo, *An Unbiased Hessian Representation for Monte Carlo PDFs*, *Eur. Phys. J.* **C75** (2015), no. 8 369, [arXiv:1505.06736].
- [146] S. Carrazza, S. Forte, Z. Kassabov, and J. Rojo, *Specialized minimal PDFs for optimized LHC calculations*, *Eur. Phys. J.* **C76** (2016), no. 4 205, [arXiv:1602.00005].
- [147] D. de Florian, G. F. R. Sborlini, and G. Rodrigo, *QED corrections to the Altarelli–Parisi splitting functions*, *Eur. Phys. J.* **C76** (2016), no. 5 282, [arXiv:1512.00612].
- [148] D. de Florian, G. F. R. Sborlini, and G. Rodrigo, *Two-loop QED corrections to the Altarelli–Parisi splitting functions*, *JHEP* **10** (2016) 056, [arXiv:1606.02887].
- [149] A. Candido, F. Hekhorn, and G. Magni, *EKO: evolution kernel operators*, *Eur. Phys. J. C* **82** (2022), no. 10 976, [arXiv:2202.02338].
- [150] S. Carrazza, *Parton distribution functions with qed corrections*, 2015.
- [151] F. Demartin, S. Forte, E. Mariani, J. Rojo, and A. Vicini, *The impact of PDF and α_s uncertainties on Higgs Production in gluon fusion at hadron colliders*, *Phys. Rev.* **D82** (2010) 014002, [arXiv:1004.0962].
- [152] **NNPDF Collaboration**, R. D. Ball, E. R. Nocera, and R. L. Pearson, *Nuclear Uncertainties in the Determination of Proton PDFs*, *Eur. Phys. J.* **C79** (2019), no. 3 282, [arXiv:1812.09074].
- [153] R. D. Ball, E. R. Nocera, and R. L. Pearson, *Deuteron Uncertainties in the Determination of Proton PDFs*, *Eur. Phys. J. C* **81** (2021), no. 1 37, [arXiv:2011.00009].
- [154] J. Baglio, C. Duhr, B. Mistlberger, and R. Szafron, *Inclusive production cross sections at N^3LO* , *JHEP* **12** (2022) 066, [arXiv:2209.06138].
- [155] L. Del Debbio, T. Giani, and M. Wilson, *Bayesian approach to inverse problems: an application to NNPDF closure testing*, *Eur. Phys. J. C* **82** (2022), no. 4 330, [arXiv:2111.05787].
- [156] **NNPDF Collaboration**, R. Abdul Khalek et al., *A first determination of parton distributions with theoretical uncertainties*, *Eur. Phys. J. C* (2019) 79:838, [arXiv:1905.04311].
- [157] **NNPDF Collaboration**, R. Abdul Khalek et al., *Parton Distributions with Theory Uncertainties: General Formalism and First Phenomenological Studies*, *Eur. Phys. J. C* **79** (2019), no. 11 931, [arXiv:1906.10698].
- [158] **The NNPDF Collaboration**, R. D. Ball et al., *Fitting Parton Distribution Data with Multiplicative Normalization Uncertainties*, *JHEP* **05** (2010) 075, [arXiv:0912.2276].

- [159] R. D. Ball, V. Bertone, M. Bonvini, S. Marzani, J. Rojo, and L. Rottoli, *Parton distributions with small- x resummation: evidence for BFKL dynamics in HERA data*, *Eur. Phys. J.* **C78** (2018), no. 4 321, [arXiv:1710.05935].
- [160] J. A. M. Vermaseren, A. Vogt, and S. Moch, *The third-order QCD corrections to deep-inelastic scattering by photon exchange*, *Nucl. Phys.* **B724** (2005) 3, [hep-ph/0504242].
- [161] S. Moch, J. A. M. Vermaseren, and A. Vogt, *The longitudinal structure function at the third order*, *Phys. Lett.* **B606** (2005) 123, [hep-ph/0411112].
- [162] S. Moch, M. Rogal, and A. Vogt, *Differences between charged-current coefficient functions*, *Nucl. Phys.* **B790** (2008) 317–335, [arXiv:0708.3731].
- [163] S. Moch, J. A. M. Vermaseren, and A. Vogt, *Third-order QCD corrections to the charged-current structure function $F(3)$* , *Nucl. Phys. B* **813** (2009) 220–258, [arXiv:0812.4168].
- [164] C. Anastasiou, C. Duhr, F. Dulat, F. Herzog, and B. Mistlberger, *Higgs Boson Gluon-Fusion Production in QCD at Three Loops*, *Phys. Rev. Lett.* **114** (2015), no. 21 212001, [arXiv:1503.06056].
- [165] B. Mistlberger, *Higgs boson production at hadron colliders at N^3 LO in QCD*, *JHEP* **05** (2018) 028, [arXiv:1802.00833].
- [166] C. Duhr, F. Dulat, and B. Mistlberger, *Higgs Boson Production in Bottom-Quark Fusion to Third Order in the Strong Coupling*, *Phys. Rev. Lett.* **125** (2020), no. 5 051804, [arXiv:1904.09990].
- [167] F. A. Dreyer and A. Karlberg, *Vector-Boson Fusion Higgs Production at Three Loops in QCD*, *Phys. Rev. Lett.* **117** (2016), no. 7 072001, [arXiv:1606.00840].
- [168] L.-B. Chen, H. T. Li, H.-S. Shao, and J. Wang, *Higgs boson pair production via gluon fusion at N^3 LO in QCD*, *Phys. Lett. B* **803** (2020) 135292, [arXiv:1909.06808].
- [169] C. Duhr, F. Dulat, and B. Mistlberger, *Charged current Drell-Yan production at N^3 LO*, *JHEP* **11** (2020) 143, [arXiv:2007.13313].
- [170] C. Duhr and B. Mistlberger, *Lepton-pair production at hadron colliders at N^3 LO in QCD*, *JHEP* **03** (2022) 116, [arXiv:2111.10379].
- [171] F. Dulat, B. Mistlberger, and A. Pelloni, *Differential Higgs production at N^3 LO beyond threshold*, *JHEP* **01** (2018) 145, [arXiv:1710.03016].
- [172] F. Dulat, B. Mistlberger, and A. Pelloni, *Precision predictions at N^3 LO for the Higgs boson rapidity distribution at the LHC*, *Phys. Rev. D* **99** (2019), no. 3 034004, [arXiv:1810.09462].
- [173] X. Chen, T. Gehrmann, E. W. N. Glover, A. Huss, B. Mistlberger, and A. Pelloni, *Fully Differential Higgs Boson Production to Third Order in QCD*, *Phys. Rev. Lett.* **127** (2021), no. 7 072002, [arXiv:2102.07607].
- [174] G. Billis, B. Dehnadi, M. A. Ebert, J. K. L. Michel, and F. J. Tackmann, *Higgs p_T Spectrum and Total Cross Section with Fiducial Cuts at Third Resummed and Fixed Order in QCD*, *Phys. Rev. Lett.* **127** (2021), no. 7 072001, [arXiv:2102.08039].

- [175] S. Camarda, L. Cieri, and G. Ferrera, *Drell–Yan lepton-pair production: q_T resummation at N³LL accuracy and fiducial cross sections at N³LO*, *Phys. Rev. D* **104** (2021), no. 11 L111503, [arXiv:2103.04974].
- [176] X. Chen, T. Gehrmann, N. Glover, A. Huss, T.-Z. Yang, and H. X. Zhu, *Dilepton Rapidity Distribution in Drell-Yan Production to Third Order in QCD*, *Phys. Rev. Lett.* **128** (2022), no. 5 052001, [arXiv:2107.09085].
- [177] X. Chen, T. Gehrmann, N. Glover, A. Huss, T.-Z. Yang, and H. X. Zhu, *Transverse mass distribution and charge asymmetry in W boson production to third order in QCD*, *Phys. Lett. B* **840** (2023) 137876, [arXiv:2205.11426].
- [178] F. Caola, W. Chen, C. Duhr, X. Liu, B. Mistlberger, F. Petriello, G. Vita, and S. Weinzierl, *The Path forward to N³LO*, in *Snowmass 2021*, 3, 2022. arXiv:2203.06730.
- [179] J. Davies, A. Vogt, B. Ruijl, T. Ueda, and J. A. M. Vermaseren, *Large- n_f contributions to the four-loop splitting functions in QCD*, *Nucl. Phys. B* **915** (2017) 335–362, [arXiv:1610.07477].
- [180] S. Moch, B. Ruijl, T. Ueda, J. A. M. Vermaseren, and A. Vogt, *Four-Loop Non-Singlet Splitting Functions in the Planar Limit and Beyond*, *JHEP* **10** (2017) 041, [arXiv:1707.08315].
- [181] J. Davies, C. H. Kom, S. Moch, and A. Vogt, *Resummation of small- x double logarithms in QCD: inclusive deep-inelastic scattering*, *JHEP* **08** (2022) 135, [arXiv:2202.10362].
- [182] J. M. Henn, G. P. Korchemsky, and B. Mistlberger, *The full four-loop cusp anomalous dimension in $\mathcal{N} = 4$ super Yang-Mills and QCD*, *JHEP* **04** (2020) 018, [arXiv:1911.10174].
- [183] C. Duhr, B. Mistlberger, and G. Vita, *Soft integrals and soft anomalous dimensions at N³LO and beyond*, *JHEP* **09** (2022) 155, [arXiv:2205.04493].
- [184] S. Moch, B. Ruijl, T. Ueda, J. A. M. Vermaseren, and A. Vogt, *Low moments of the four-loop splitting functions in QCD*, *Phys. Lett. B* **825** (2022) 136853, [arXiv:2111.15561].
- [185] G. Soar, S. Moch, J. A. M. Vermaseren, and A. Vogt, *On Higgs-exchange DIS, physical evolution kernels and fourth-order splitting functions at large x* , *Nucl. Phys. B* **832** (2010) 152–227, [arXiv:0912.0369].
- [186] G. Falcioni, F. Herzog, S. Moch, and A. Vogt, *Four-loop splitting functions in QCD – The quark-quark case*, arXiv:2302.07593.
- [187] G. Falcioni, F. Herzog, S. Moch, and A. Vogt, *Four-loop splitting functions in QCD – The gluon-to-quark case*, *Phys. Lett. B* **846** (2023) 138215, [arXiv:2307.04158].
- [188] S. Moch, B. Ruijl, T. Ueda, J. Vermaseren, and A. Vogt, *Additional moments and x -space approximations of four-loop splitting functions in QCD*, *Phys. Lett. B* **849** (2024) 138468, [arXiv:2310.05744].

- [189] G. Falcioni, F. Herzog, S. Moch, J. Vermaseren, and A. Vogt, *The double fermionic contribution to the four-loop quark-to-gluon splitting function*, *Phys. Lett. B* **848** (2024) 138351, [arXiv:2310.01245].
- [190] J. McGowan, T. Cridge, L. A. Harland-Lang, and R. S. Thorne, *Approximate N^3 LO parton distribution functions with theoretical uncertainties: MSHT20aN³LO PDFs*, *Eur. Phys. J. C* **83** (2023), no. 3 185, [arXiv:2207.04739].
- [191] W. L. van Neerven and A. Vogt, *Improved approximations for the three loop splitting functions in QCD*, *Phys. Lett. B* **490** (2000) 111–118, [hep-ph/0007362].
- [192] NNPDF Collaboration, R. D. Ball et al., *An open-source machine learning framework for global analyses of parton distributions*, *Eur. Phys. J. C* **81** (2021), no. 10 958, [arXiv:2109.02671].
- [193] N. Laurenti, *Construction of a next-to-next-to-next-to-leading order approximation for heavy flavour production in deep inelastic scattering with quark masses*, Master's thesis, Rome Sapienza U., 2021.
- [194] A. Barontini, M. Bonvini, and N. Laurenti, *Implementation of DIS at N³LO for PDF determinations, in preparation* (2024).
- [195] I. Bierenbaum, J. Blümlein, and S. Klein, *Mellin Moments of the $O(\alpha_s^3)$ Heavy Flavor Contributions to unpolarized Deep-Inelastic Scattering at $Q^2 \gg m^2$ and Anomalous Dimensions*, *Nucl. Phys. B* **820** (2009) 417–482, [arXiv:0904.3563].
- [196] I. Bierenbaum, J. Blumlein, S. Klein, and C. Schneider, *Two-Loop Massive Operator Matrix Elements for Unpolarized Heavy Flavor Production to $O(\epsilon)$* , *Nucl. Phys. B* **803** (2008) 1–41, [arXiv:0803.0273].
- [197] I. Bierenbaum, J. Blümlein, and S. Klein, *The Gluonic Operator Matrix Elements at $O(\alpha_s^2)$ for DIS Heavy Flavor Production*, *Phys. Lett. B* **672** (2009) 401–406, [arXiv:0901.0669].
- [198] J. Ablinger, J. Blümlein, S. Klein, C. Schneider, and F. Wissbrock, *The $O(\alpha_s^3)$ Massive Operator Matrix Elements of $O(n_f)$ for the Structure Function $F_2(x, Q^2)$ and Transversity*, *Nucl. Phys. B* **844** (2011) 26–54, [arXiv:1008.3347].
- [199] J. Ablinger, A. Behring, J. Blümlein, A. De Freitas, A. Hasselhuhn, A. von Manteuffel, M. Round, C. Schneider, and F. Wißbrock, *The 3-Loop Non-Singlet Heavy Flavor Contributions and Anomalous Dimensions for the Structure Function $F_2(x, Q^2)$ and Transversity*, *Nucl. Phys. B* **886** (2014) 733–823, [arXiv:1406.4654].
- [200] J. Ablinger, J. Blümlein, A. De Freitas, A. Hasselhuhn, A. von Manteuffel, M. Round, and C. Schneider, *The $O(\alpha_s^3 T_F^2)$ Contributions to the Gluonic Operator Matrix Element*, *Nucl. Phys. B* **885** (2014) 280–317, [arXiv:1405.4259].
- [201] A. Behring, I. Bierenbaum, J. Blümlein, A. De Freitas, S. Klein, and F. Wißbrock, *The logarithmic contributions to the $O(\alpha_s^3)$ asymptotic massive Wilson coefficients and operator matrix elements in deeply inelastic scattering*, *Eur. Phys. J. C* **74** (2014), no. 9 3033, [arXiv:1403.6356].
- [202] J. Blümlein, J. Ablinger, A. Behring, A. De Freitas, A. von Manteuffel, C. Schneider, and C. Schneider, *Heavy Flavor Wilson Coefficients in Deep-Inelastic Scattering: Recent Results*, *PoS QCDEV2017* (2017) 031, [arXiv:1711.07957].

- [203] J. Ablinger, J. Blümlein, A. De Freitas, A. Hasselhuhn, A. von Manteuffel, M. Round, C. Schneider, and F. Wißbrock, *The transition matrix element $a_{qq}(n)$ of the variable flavor number scheme at $o(\alpha_s^3)$* , *Nuclear Physics B* **882** (May, 2014) 263–288.
- [204] J. Ablinger, A. Behring, J. Blümlein, A. De Freitas, A. von Manteuffel, and C. Schneider, *The 3-loop pure singlet heavy flavor contributions to the structure function $f_2(x, q^2)$ and the anomalous dimension*, *Nuclear Physics B* **890** (Jan, 2015) 48–151.
- [205] J. Ablinger, A. Behring, J. Blümlein, A. De Freitas, A. Goedicke, A. von Manteuffel, C. Schneider, and K. Schönwald, *The unpolarized and polarized single-mass three-loop heavy flavor operator matrix elements $A_{gg, Q}$ and $\Delta A_{gg, Q}$* , *JHEP* **12** (2022) 134, [arXiv:2211.05462].
- [206] J. Ablinger, A. Behring, J. Blümlein, A. De Freitas, A. von Manteuffel, C. Schneider, and K. Schönwald, *The first-order factorizable contributions to the three-loop massive operator matrix elements $A_{Qg(3)}$ and $\Delta A_{Qg(3)}$* , *Nucl. Phys. B* **999** (2024) 116427, [arXiv:2311.00644].
- [207] H. Kawamura, N. A. Lo Presti, S. Moch, and A. Vogt, *On the next-to-next-to-leading order QCD corrections to heavy-quark production in deep-inelastic scattering*, *Nucl. Phys. B* **864** (2012) 399–468, [arXiv:1205.5727].
- [208] J. Davies, A. Vogt, S. Moch, and J. A. M. Vermaseren, *Non-singlet coefficient functions for charged-current deep-inelastic scattering to the third order in QCD*, *PoS DIS2016* (2016) 059, [arXiv:1606.08907].
- [209] J. Blümlein, P. Marquard, C. Schneider, and K. Schönwald, *The massless three-loop Wilson coefficients for the deep-inelastic structure functions F_2 , F_L , xF_3 and g_1* , *JHEP* **11** (2022) 156, [arXiv:2208.14325].
- [210] E. Laenen, S. Riemersma, J. Smith, and W. L. van Neerven, *Complete $O(\alpha_s)$ corrections to heavy flavor structure functions in electroproduction*, *Nucl. Phys.* **B392** (1993) 162–228.
- [211] S. I. Alekhin and J. Blümlein, *Mellin representation for the heavy flavor contributions to deep inelastic structure functions*, *Phys. Lett.* **B594** (2004) 299–307, [hep-ph/0404034].
- [212] F. Hekhorn and M. Stratmann, *Next-to-Leading Order QCD Corrections to Inclusive Heavy-Flavor Production in Polarized Deep-Inelastic Scattering*, *Phys. Rev. D* **98** (2018), no. 1 014018, [arXiv:1805.09026].
- [213] F. Hekhorn, *Next-to-Leading Order QCD Corrections to Heavy-Flavour Production in Neutral Current DIS*. PhD thesis, Tübingen U., Math. Inst., 2019. arXiv:1910.01536.
- [214] J. Gao, *Massive charged-current coefficient functions in deep-inelastic scattering at NNLO and impact on strange-quark distributions*, *JHEP* **02** (2018) 026, [arXiv:1710.04258].
- [215] E. Laenen and S.-O. Moch, *Soft gluon resummation for heavy quark electroproduction*, *Phys. Rev. D* **59** (1999) 034027, [hep-ph/9809550].
- [216] S. Catani, M. Ciafaloni, and F. Hautmann, *High-energy factorization and small x heavy flavor production*, *Nucl. Phys. B* **366** (1991) 135–188.

- [217] J. Ablinger, A. Behring, J. Blümlein, A. De Freitas, A. von Manteuffel, et al., *The 3-Loop Pure Singlet Heavy Flavor Contributions to the Structure Function $F_2(x, Q^2)$ and the Anomalous Dimension*, arXiv:1409.1135.
- [218] R. Bonciani, S. Catani, M. L. Mangano, and P. Nason, *NLL resummation of the heavy quark hadroproduction cross-section*, *Nucl. Phys. B* **529** (1998) 424–450, [hep-ph/9801375]. [Erratum: *Nucl.Phys.B* 803, 234 (2008)].
- [219] A. Pineda and A. Signer, *Heavy Quark Pair Production near Threshold with Potential Non-Relativistic QCD*, *Nucl. Phys. B* **762** (2007) 67–94, [hep-ph/0607239].
- [220] F. Faura, S. Iranipour, E. R. Nocera, J. Rojo, and M. Ubiali, *The Strangest Proton?*, *Eur. Phys. J. C* **80** (2020), no. 12 1168, [arXiv:2009.00014].
- [221] S. Amoroso et al., *Snowmass 2021 Whitepaper: Proton Structure at the Precision Frontier*, *Acta Phys. Polon. B* **53** (2022), no. 12 12–A1.63–67, [arXiv:2203.13923].
- [222] T. Cridge, L. A. Harland-Lang, and R. S. Thorne, *Combining QED and Approximate N^3 LO QCD Corrections in a Global PDF Fit: MSHT20qed_an3lo PDFs*, arXiv:2312.07665.
- [223] A. Buckley et al., *General-purpose event generators for LHC physics*, *Phys. Rept.* **504** (2011) 145–233, [arXiv:1101.2599].
- [224] P. Nason and B. Webber, *Next-to-Leading-Order Event Generators*, *Ann. Rev. Nucl. Part. Sci.* **62** (2012) 187–213, [arXiv:1202.1251].
- [225] S. Höche, *Introduction to parton-shower event generators*, in *Theoretical Advanced Study Institute in Elementary Particle Physics: Journeys Through the Precision Frontier: Amplitudes for Colliders*, pp. 235–295, 2015. arXiv:1411.4085.
- [226] J. M. Campbell et al., *Event Generators for High-Energy Physics Experiments*, in *Snowmass 2021*, 3, 2022. arXiv:2203.11110.
- [227] T. Sjostrand, S. Mrenna, and P. Z. Skands, *A Brief Introduction to PYTHIA 8.1*, *Comput. Phys. Commun.* **178** (2008) 852–867, [arXiv:0710.3820].
- [228] C. Bierlich et al., *A comprehensive guide to the physics and usage of PYTHIA 8.3*, *SciPost Phys. Codeb.* **2022** (2022) 8, [arXiv:2203.11601].
- [229] J. Bellm et al., *Herwig 7.0/Herwig++ 3.0 release note*, *Eur. Phys. J. C* **76** (2016), no. 4 196, [arXiv:1512.01178].
- [230] J. Bellm et al., *Herwig 7.2 release note*, *Eur. Phys. J. C* **80** (2020), no. 5 452, [arXiv:1912.06509].
- [231] T. Gleisberg, S. Hoeche, F. Krauss, M. Schonherr, S. Schumann, F. Siegert, and J. Winter, *Event generation with SHERPA 1.1*, *JHEP* **02** (2009) 007, [arXiv:0811.4622].
- [232] **Sherpa** Collaboration, E. Bothmann et al., *Event Generation with Sherpa 2.2*, *SciPost Phys.* **7** (2019), no. 3 034, [arXiv:1905.09127].
- [233] S. Alioli, P. Nason, C. Oleari, and E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, *JHEP* **1006** (2010) 043, [arXiv:1002.2581].

- [234] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, *JHEP* **1407** (2014) 079, [arXiv:1405.0301].
- [235] M. van Beekveld et al., *Introduction to the PanScales framework, version 0.1*, arXiv:2312.13275.
- [236] M. van Beekveld, S. Ferrario Ravasio, K. Hamilton, G. P. Salam, A. Soto-Ontoso, G. Soyez, and R. Verheyen, *PanScales showers for hadron collisions: all-order validation*, *JHEP* **11** (2022) 020, [arXiv:2207.09467].
- [237] M. van Beekveld, S. Ferrario Ravasio, G. P. Salam, A. Soto-Ontoso, G. Soyez, and R. Verheyen, *PanScales parton showers for hadron collisions: formulation and fixed-order studies*, *JHEP* **11** (2022) 019, [arXiv:2205.02237].
- [238] M. van Beekveld et al., *A new standard for the logarithmic accuracy of parton showers*, arXiv:2406.02661.
- [239] J. Gao, L. Harland-Lang, and J. Rojo, *The Structure of the Proton in the LHC Precision Era*, *Phys. Rept.* **742** (2018) 1–121, [arXiv:1709.04922].
- [240] K. Kovařík, P. M. Nadolsky, and D. E. Soper, *Hadronic structure in high-energy collisions*, *Rev. Mod. Phys.* **92** (2020), no. 4 045003, [arXiv:1905.06957].
- [241] P. Z. Skands, *Tuning Monte Carlo Generators: The Perugia Tunes*, *Phys. Rev. D* **82** (2010) 074018, [arXiv:1005.3457].
- [242] H.-L. Lai et al., *Parton Distributions for Event Generators*, *JHEP* **04** (2010) 035, [arXiv:0910.4183].
- [243] M. Yan, T.-J. Hou, P. Nadolsky, and C. P. Yuan, *CT18 global PDF fit at leading order in QCD*, *Phys. Rev. D* **107** (2023), no. 11 116001, [arXiv:2205.00137].
- [244] **The NNPDF Collaboration**, R. D. Ball et al., *Unbiased global determination of parton distributions and their uncertainties at NNLO and at LO*, *Nucl.Phys.* **B855** (2012) 153, [arXiv:1107.2652].
- [245] A. Sherstnev and R. S. Thorne, *Different PDF approximations useful for LO Monte Carlo generators*, in *16th International Workshop on Deep Inelastic Scattering and Related Subjects*, p. 149, 7, 2008. arXiv:0807.2132.
- [246] A. Sherstnev and R. S. Thorne, *Parton Distributions for LO Generators*, *Eur. Phys. J.* **C55** (2008) 553–575, [arXiv:0711.2473].
- [247] S. Carrazza, S. Forte, and J. Rojo, *Parton Distributions and Event Generators*, in *43rd International Symposium on Multiparticle Dynamics*, pp. 89–96, 2013. arXiv:1311.5887.
- [248] P. Skands, S. Carrazza, and J. Rojo, *Tuning PYTHIA 8.1: the Monash 2013 Tune*, *European Physical Journal* **74** (2014) 3024, [arXiv:1404.5630].
- [249] V. Bertone, R. Gauld, and J. Rojo, *Neutrino Telescopes as QCD Microscopes*, *JHEP* **01** (2019) 217, [arXiv:1808.02034].
- [250] J. Collins, T. C. Rogers, and N. Sato, *Positivity and renormalization of parton densities*, *Phys. Rev. D* **105** (2022), no. 7 076010, [arXiv:2111.01170].

- [251] A. Candido, S. Forte, T. Giani, and F. Hekhorn, *On the positivity of \overline{MS} parton distributions*, *Eur. Phys. J. C* **84** (2024), no. 3 335, [arXiv:2308.00025].
- [252] **CTEQ-TEA Collaboration**, K. Xie, T. J. Hobbs, T.-J. Hou, C. Schmidt, M. Yan, and C. P. Yuan, *Photon PDF within the CT18 global analysis*, *Phys. Rev. D* **105** (2022), no. 5 054006, [arXiv:2106.10299].
- [253] M. Bonvini, R. D. Ball, S. Forte, S. Marzani, and G. Ridolfi, *Updated Higgs cross section at approximate N^3LO* , *J. Phys.* **G41** (2014) 095002, [arXiv:1404.3204].
- [254] M. Cacciari, F. A. Dreyer, A. Karlberg, G. P. Salam, and G. Zanderighi, *Fully Differential Vector-Boson-Fusion Higgs Production at Next-to-Next-to-Leading Order*, *Phys. Rev. Lett.* **115** (2015), no. 8 082002, [arXiv:1506.02660].
- [255] C. Bierlich et al., *Robust Independent Validation of Experiment and Theory: Rivet version 3*, *SciPost Phys.* **8** (2020) 026, [arXiv:1912.05451].
- [256] **ATLAS Collaboration**, G. Aad et al., *Measurement of the transverse momentum distribution of Z/γ^* bosons in proton–proton collisions at $\sqrt{s}=7$ TeV with the ATLAS detector*, *Phys. Lett. B* **705** (2011) 415–434, [arXiv:1107.2381].
- [257] **ATLAS Collaboration**, G. Aad et al., *Measurements of the Higgs boson inclusive and differential fiducial cross sections in the 4ℓ decay channel at $\sqrt{s} = 13$ TeV*, *Eur. Phys. J. C* **80** (2020), no. 10 942, [arXiv:2004.03969].
- [258] **CMS Collaboration**, S. Chatrchyan et al., *Measurement of energy flow at large pseudorapidities in pp collisions at $\sqrt{s} = 0.9$ and 7 TeV*, *JHEP* **11** (2011) 148, [arXiv:1110.0211]. [Erratum: *JHEP* 02, 055 (2012)].
- [259] **CMS Collaboration**, V. Khachatryan et al., *Transverse-momentum and pseudorapidity distributions of charged hadrons in pp collisions at $\sqrt{s} = 7$ TeV*, *Phys. Rev. Lett.* **105** (2010) 022002, [arXiv:1005.3299].
- [260] R. D. Ball, A. Candido, S. Forte, F. Hekhorn, E. R. Nocera, J. Rojo, and C. Schwan, *Parton distributions and new physics searches: the Drell–Yan forward–backward asymmetry as a case study*, *Eur. Phys. J. C* **82** (2022), no. 12 1160, [arXiv:2209.08115].
- [261] E. Hammou, Z. Kassabov, M. Madigan, M. L. Mangano, L. Mantani, J. Moore, M. M. Alvarado, and M. Ubiali, *Hide and seek: how PDFs can conceal new physics*, *JHEP* **11** (2023) 090, [arXiv:2307.10370].
- [262] **LHC Higgs Cross Section Working Group Collaboration**, D. de Florian et al., *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, arXiv:1610.07922.
- [263] L. A. Harland-Lang and R. S. Thorne, *On the Consistent Use of Scale Variations in PDF Fits and Predictions*, *Eur. Phys. J.* **C79** (2019), no. 3 225, [arXiv:1811.08434].
- [264] Z. Kassabov, M. Ubiali, and C. Voisey, *Parton distributions with scale uncertainties: a Monte Carlo sampling approach*, *JHEP* **03** (2023) 148, [arXiv:2207.07616].
- [265] R. L. Pearson, R. D. Ball, and E. R. Nocera, *Next generation proton PDFs with deuteron and nuclear uncertainties*, *SciPost Phys. Proc.* **8** (2022) 026, [arXiv:2106.12349].

List of Publications

- A. Barontini, M. Bonvini, N. Laurenti, *Implementation of DIS at N^3 LO for PDF determinations*, in preparation.
- J. Cruz-Martinez, S. Forte, N. Laurenti, T. R. Rabemananjara, and J. Rojo, *LO, NLO, and NNLO Parton Distributions for LHC Event Generators*, *JHEP* **09** (2024) 088, arXiv:2406.12961.
- A. Barontini, N. Laurenti, and J. Rojo, *NNPDF4.0 at N^3 LO PDFs with QED corrections*, in *31st International Workshop on Deep-Inelastic Scattering and Related Subjects*, 6, 2024, arXiv:2406.01779.
- **NNPDF** collaboration, R. D. Ball et al., *The path to N^3 LO parton distributions*, *Eur. Phys. J. C* **84** (2024), no. 7 659, arXiv:2402.18635.
- **NNPDF** collaboration, R. D. Ball et al., *Determination of the theory uncertainties from missing higher orders on NNLO parton distributions with percent accuracy*, *Eur. Phys. J. C* **84** (2024), no. 5 517, arXiv:2401.10319.
- **NNPDF** collaboration, R. D. Ball et al., *Photons in the proton: implications for the LHC*, *Eur. Phys. J. C* **84** (2024), no. 5 540, arXiv:2401.08749.
- **NNPDF** collaboration, N. Laurenti, *Inclusion of QED corrections in PDFs fits*, *Nucl. Part. Phys. Proc.*, **343** (2024) 23-27.
- N. Laurenti, T. R. Rabemananjara, and R. Stegeman, *Approximating missing higher-orders in transverse momentum distributions using resummations*, in *29th International Workshop on Deep-Inelastic Scattering and Related Subjects*, 7, 2022, arXiv:2207.12265.

Acknowledgments

With the end of my Ph.D. there are many people that I have met me during this journey and that I want to thank. First of all, I would like to thank my Supervisor Prof. Stefano Forte, who gave me the opportunity to work in this interesting field, helped me during these three years with useful explanations and discussions and for sure played a fundamental role in my education.

I would like also to thank all the members of the NNPDF collaboration and in particular the ones of the N³PDF group, Roy Stegeman, Juan Cruz-Martinez, Felix Hekhorn, Alessandro Candido and Kirill Kudashkin who welcomed me in the Milan physics department and helped me during the first period of my Ph.D. In particular, I would like to thank Alessandro and Felix for supervising me in the first steps of my journey. A special thank goes to Roy, as big Davide once said, the guy with the perfect body, for participating to all the coffee breaks and for making the office a nicer place.

Among all of them, a special thank goes to Andrea Barontini, that is a friend before being a colleague, but in particular was my flatmate (and someone says also lover, but we will keep the mystery unsolved) for three years. I want to thank him for making this period more fun, for the useful discussions about anime and for helping each other during our Ph.D. path.

Then, I would like to thank my desk neighbor Davide Maria Tagliabue (he would kill me if I forgot the "Maria"), a.k.a. big Davide, for teaching me the basics about religion and for enlightening me with very useful discussions about super expensive watches, sport cars and bodybuilding competitions, that have been fundamental for my academic education (but for my mental sanity I will never write with him a Latex document anymore). Moreover, I would like to thank Davide Morgante, a.k.a. small Davide, for all the laughs that we made together and for creating lots of awkward situations asking to all the people in the department the infamous "twenty years question".

I would also like to thank Jacopo D'Alberto, the guy who "si abbenda" and that convinced me not once but twice to support him in some low level disco club because he had to flirt with a girl, and Michele Caresana, a.k.a. the Bee-Keeper, the only person that has already done all the coolest things in the world. Thank you for our skiing trips which were helpful to learn Jacopo's dialect.

Then, I would like to thank Davide Abriola, a.k.a. the average size Davide, Andrea Afify, Sara Cesare and Laura Nasella for all the lunches together and for being part of my Ph.D. journey and making it more pleasant and fun. In particular, I thank Laura for never introducing the actress she is friend with to Jacopo.

Then, I would like to thank my mother, for calling me every week (more than one a week is too much, sorry mum), asking me if I have had lunch and dinner, taking and bringing me to the train station every time I was back home and for helping me whenever I needed it.

I also want to thank my brothers Jacopo (that had for three years all the room for himself) and Simone.

In the end, I would like to thank my all friends, but in particular Matteo, for being my friend my whole life, and Rosanna for always giving a good reason to come back home.