UNIVERSITÀ DEGLI STUDI DI MILANO

FACOLTÀ DI SCIENZE E TECNOLOGIE

DIPARTIMENTO DI FISICA

ON THE SCALE DEPENDENCE OF PARTON PARAMETRIZATIONS

Ivan Andrea Godino
988088

Relatore
Prof. Stefano Forte

Anno accademico 2022/2023

Summary

This work has the purpose of analysing the dependence of parton distribution function on the scale at which they are parameterized. In particular, we analyse the effects of passing through the threshold for production of heavy quarks.

The first chapter presents a review of the basic concepts of Quantum ChromoDynamics, in more detail the phenomena of confinement and asymptotic freedom. Then, we give an overview of PDFs and their role in describing scattering processes.

The second chapter presents the NNPDF methodology for fitting PDFs. Namely, it introduces the elements of the fitting process and it contains a brief explanation of the ideas behind a neural network.

The third chapter presents the results of the study. In particular the results are divided into different regimes at which the PDFs are parameterized, specifically discussing the possible dependence on heavy quark threshold.

INDEX

1 Theoretical review of QCD and PDFs 7
- 1.1 QCD formulation . . . 7
- 1.2 Experimental evidence & PDFs . . . 10
- 1.3 Factorization . . . 11

2 PDF determination with NNs 14
- 2.1 The NNPDF approach . . . 14
- 2.2 Brief review of Neural Network principles . . . 16
- 2.3 NNPDF use of Neural Networks . . . 18
- 2.4 On the scale dependence of parton parameterizations . . . 20

3 Results 22
- 3.1 Results before iteration . . . 22
- 3.2 Need of a second iteration . . . 22
- 3.3 Results after iteration . . . 27
- 3.4 On the gluon behavior at small x . . . 28

1 Theoretical review of QCD and PDFs

1.1 QCD formulation

Our current understanding of strong interaction is based on *Quantum ChromoDynamics* (QCD), a theory firstly proposed in 1973 by Fritzsch, Leutwyler and Gell-Mann. According to this theory hadrons are not fundamental particles themselves, but are constituted by quarks, charged fermions with spin 1/2, which are fundamental particles. The force that ties quarks together in a hadron is the strong force. This type of force is among the four fundamental interactions and its source is the *colour charge*. Each quark, in fact, in addition to electric charge, carries a colour charge, which conventionally can be red, green, blue, anti-red, anti-blue, anti-green. This colour should be understood just as a name to this additional degree of freedom carried by quarks, but there is no relationship with the common concept of colours for objects. Even if each quark carries a colour charge, they are confined in hadrons, which carry a neutral colour charge (they are *white*). In order to better understand this feature of hadrons, it is useful to study in greater detail the formulation of strong interaction.

QCD is a gauge theory, based on the $SU(3)$ symmetry of colour. The basic idea behind the formulation of this gauge theory is that if we locally redefine the convention for colour charges and we accordingly modify the gauge field, we find again the same physical laws. This can be schematically represented, with the lagrangian $\mathcal{L}$, as:

$$G^{SU(3)_c}(x)\mathcal{L}(\psi, A) \rightarrow \mathcal{L}(\psi^*, A^*),$$

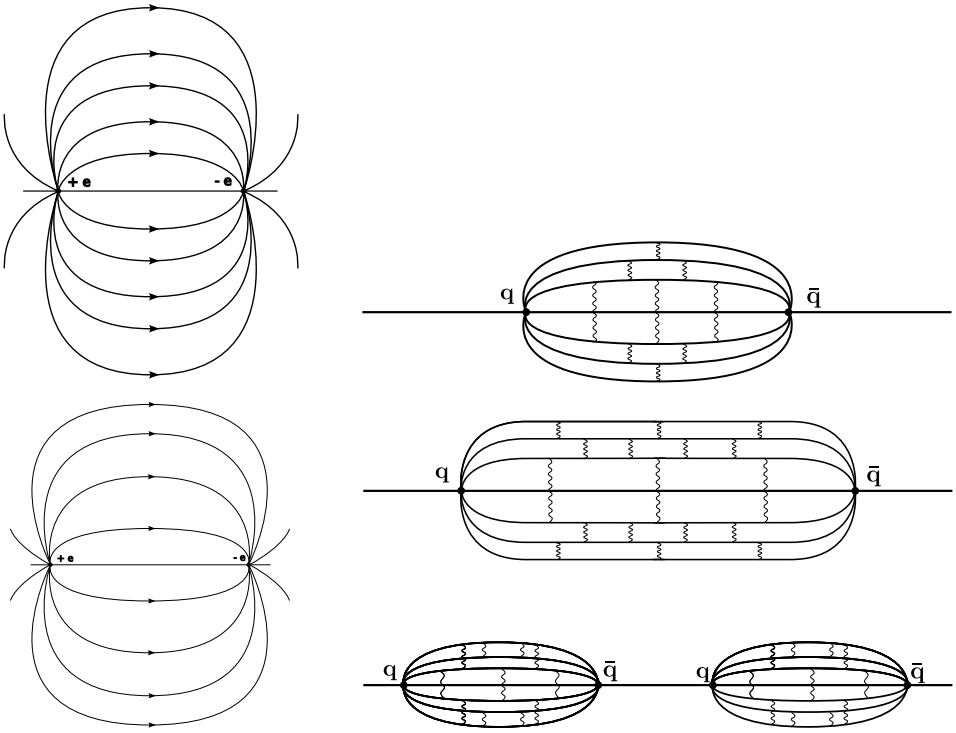
where ψ is the field, A is the gauge field and $G^{SU(3)_c}(x)$ represents the local symmetry operation applied. What the above expression means, in other words, is that if we change everywhere the convention for colour, in a different way for each point, it is still possible to have a theory which describes the same physics, but we have to change consequently the field, adding a local phase, and change the gauge field. As in QED (Quantum ElectroDynamics), quanta of the gauge field are massless spin-1 gauge particles and, within QCD, they are called *gluons*. Despite the similarities, gluons carry colour charge themselves and are not neutral (like the photon). The reason of this peculiar behavior has its roots in the fact that $SU(3)$ is a non-abelian group (while $U(1)$, the group behind QED, is abelian). This different aspect gives rise to a non-abelian field theory with a self-interacting field A. Ac-

cordingly, gluons, quanta of this field, carry a colour charge and are self-interacting.

1.1.1 Confinement and asymptotic freedom

We present now *confinement* and *asymptotic freedom*. While the latter one is explained with QCD, for the first one we only have an intuitive explanation, because no prediction in this sense can be calculated from first principles in QCD.

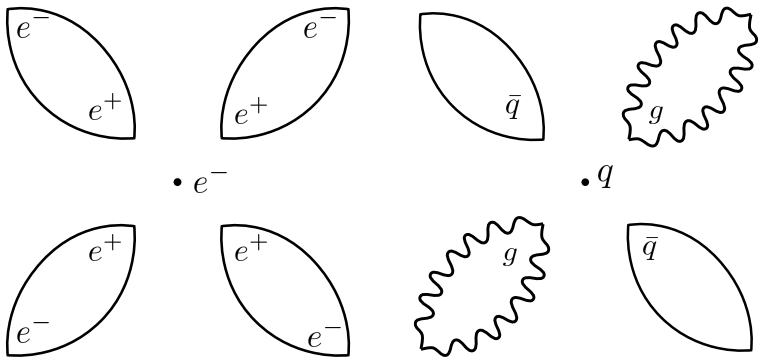
Confinement, the phenomenon responsible for the fact that quarks were never observed outside hadrons, can be understood thanks to the self-interacting gauge field we introduced above. It is useful to make a comparison with electrodynamics. As shown in the figure 1a,



(a) QED field lines **(b)** QCD field lines

Figure 1: The different behavior of field lines in the case of electrodynamics and in the case of strong interaction: for electrodynamics field lines, the density becomes lower as charges are drawn apart. For strong-interaction field lines, thanks to a self-interacting gauge field, the interaction does not become weaker as quarks are separated.

when two charges are drawn apart the field lines become less dense: this means that the force experienced by the two charges is weaker. As shown in the other figure, 1b, when we apply the same procedure to a couple of quark and anti-quark, the effect is the opposite: the field lines are always intense, due to self interaction. This means that



(a) QED **(b)** QCD

Figure 2: In the picture is depicted the different behavior of the vacuum sea polarization, in the cases of QED and QCD.

pulling away two quarks from one another, does not become easier if they are far away. So, if we try to separate two quarks, we should put a lot of work to increase their distance up to a point where the system has enough energy to create a couple of quark & anti-quark from the vacuum, giving rise to a new meson. This argument, not rigorous, can easily give an idea of why quarks are always confined and they have never been seen alone.

Another important effect theorized by QCD is the effect that goes under the name of *asymptotic freedom*. It is, in a certain way, the opposite effect of confinement. When two quarks are moved close to one another the force between them becomes increasingly small. To explain this effect, we have to deal with renormalization. We are tackling this topic from an intuitive, not rigorous, point of view, making a comparison with QED. In particular, in QED, a charge in vacuum polarizes the vacuum sea with couples of electrons and positrons. An intuitive representation of this phenomenon is depicted in figure 2a. When distances are large, the net effect of this polarization is a shielding of the bare electric charge. As a result, the usual fundamental charge, that we know in classical physics, is just the effect of the bare electric charge, shielded by the polarization of the vacuum sea with couples of electrons and positrons.

A similar effect, with some differences, appears also in QCD (figure 2b). In QCD, around a quark, we have a sea of quark-antiquarks and gluons. The feature that the shielding is also constituted by gluons provide a greater, opposite effect in comparison with QED: in QCD we have an increase in the effective colour charge. We found the sought effect of *asymptotic freedom*: as the distance decreases, the colour charge of quarks decreases.

1.2 Experimental evidence & PDFs

Even if their existence outside hadrons was never observed, we can say that we "saw" quarks for the first time in 1968, with deep inelastic scattering experiments (DIS), at SLAC accelerator. In these experiments an energetic lepton (in the specific case an electron) was scattered off a proton. In particular, it is useful to analyse the cross section:

$$\frac{d^2\sigma}{dq^2 d\nu} = \frac{4\pi\alpha^2}{q^4}\frac{E_f}{E_i M_p}\left[\frac{M_p}{\nu}F_2(q^2,\nu)\cos^2\frac{\theta}{2} + 2F_1(q^2,\nu)\sin^2\frac{\theta}{2}\right],$$

where $\nu = E_i - E_f$ is the difference between the initial energy and the final energy of the electron, M_p is the proton mass, θ is the angle of scattering for the electron, α is the fine-structure constant, q is the momentum transfer. F_1 and F_2 are *structure functions* and correspond to the scattering of the 2 possible polarization states of the virtual photon exchanged. Analysing the experimental data from the experiments, it became evident that F_1 and F_2 didn't depend on q^2, but only on a dimensionless quantity (Bjorken variable):

$$x = \frac{Q^2}{2p\cdot q} = \frac{q^2}{2M_n\nu},$$

where $Q^2 = -q^2$, p is the momentum of the initial hadron.

This feature led to the so-called *scaling hypothesis*. The key observation here was that it was possible to interpret the results of deep inelastic scattering (DIS) as the results of an elastic scattering process between a free point-like electron and a point like constituent of the hadron (parton). In this way, x becomes the fraction of the hadron longitudinal momentum carried by the parton. To sum up, in electron-proton scattering, when the momentum transfer is really small (with respect to $\hbar r_p^{-1}$) we can see the scattering experiment as an elastic scattering between an electron and a point-like proton. On the opposite way, when the momentum transfer is really high, we can interpret the process as the scattering with point-like particles confined into the nucleon.

This is coherent with *asymptotic freedom*: if we have a high momentum transfer, we are able to resolve shorter distances and see the quarks as "free". The subsequent formulation of QCD achieved immediately a great success, because, as discussed above, the theory was able to explain asymptotic freedom, i.e. the fact that at high Q^2, the quarks could be approximate as free. Of course, quarks are not actually free and the dependence on Q^2 cannot, in general, be neglected. The Bjorken variable can still be interpreted as above, but this approximation is valid only at leading order in QCD.

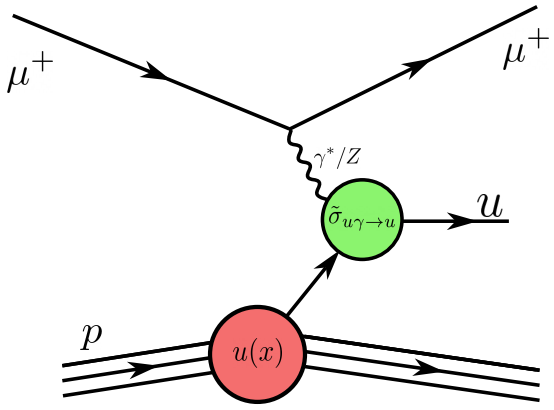


Figure 3: Schematic representation of the scattering process between an hadron (proton) and a lepton (muon). We assume that the interaction is between the muon and an up quark.

1.3 Factorization

In QCD, for quark-lepton scatterings (figure 3), we can exploit an useful property: *factorization*. Thanks to it, the total cross section (assuming a muon interacting with an up quark) can be written as:

$$\sigma_{\mu p\to\mu X}(Q^2) = \tilde{\sigma}_{\gamma^* u\to u} \otimes f_u(x, Q^2).$$

Here, the index i stands for the type of parton and can assume 6 different values for the different flavours, 6 other values for the antiquarks and an additional one for the gluon. The term $\tilde{\sigma}_{\gamma^* u\to u}$ is the so-called "hard" cross section for photon-quark scattering, computable in QCD, and $\otimes$ indicates the convolution product.

As we see there is a quantity, $f_i(x, Q^2)$, naturally raising in this treatment. These functions (mathematically *distributions*) are called *parton distribution functions* (PDFs) and are distinctive for each hadron. At leading order in QCD can be interpreted as probability densities: $f(x, Q^2)dx$ represent the probability of finding a parton i with a fraction of the momentum of the hadron between x and $x + dx$. In an analogous way, we can write the total cross section for a scattering process between two quarks (one up and one down) confined in two protons as:

$$\sigma_{pp\to W} = \tilde{\sigma}_{ud\to W} \otimes f_u(x, Q^2) \otimes f_d(x, Q^2).$$

The importance of PDFs stands in their *universality*: their expression is the same for every type of collision and is, therefore, a property of the hadron. Thanks to their universality, it is possible to determine PDFs from one process and use them for another. As an example, PDFs were among the input information used in 2012, when the Higgs boson was discovered in ATLAS and CMS experiments.

1.3.1 PDF properties

It is important to notice that PDFs are subjected to some constrains. In particular, in each hadron the flavour number is a quantity well-defined. For example, a proton has, in total, 2 quarks up, 1 quark down, no other quarks. However, it is possible to encounter scattering experiments which comprehend also, for example, strange quarks. This happens because the scattering processes that comprehend a strange quark and those which comprehend an anti-strange quark happen with the same probability. An analogous argument stands for the other flavours. This can be written in formula as:

$$\int_0^1 dx \left(f_u(x,Q^2) - f_{\bar{u}}(x,Q^2)\right) = 2 \quad \int_0^1 dx \left(f_d(x,Q^2) - f_{\bar{d}}(x,Q^2)\right) = 1$$

$$\int_0^1 dx \left(f_j(x,Q^2) - f_{\bar{j}}(x,Q^2)\right) = 0 \quad \text{for } j \neq u, d.$$

Another constrain is given by energy conservation, that imposes the momentum sum rule:

$$\int_0^1 dx\, x \left(\sum_{i=1}^{n_f} \left(f_{q_i}(x,Q^2) + f_{\bar{q}_i}(x,Q^2)\right) + f_g(x,Q^2) \right) = 1.$$

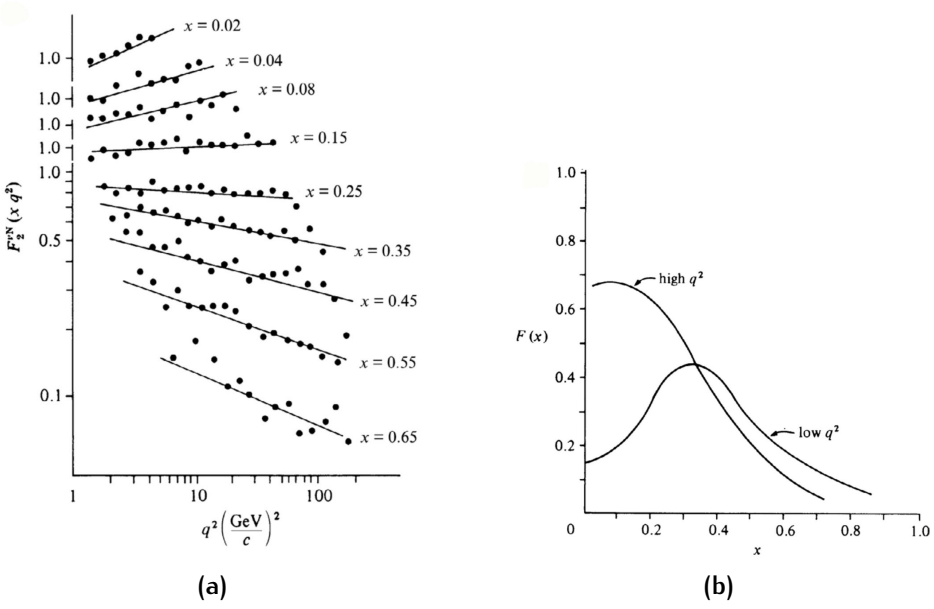
A further aspect emerging from this treatment is that it is possible to write F_1, F_2 as:

$$\begin{cases} F_1(x,Q^2) = \sum_i f_i(x,Q^2) c_i^2 \\ F_2(x,Q^2) = x \sum_i f_i(x,Q^2) c_i^2 \end{cases} \tag{1}$$

Here, the coefficients c_i indicate the charge of the quark. If quarks were free, the structure function and the PDFs would depend only on x. In the first experiments, a limited range of x and Q^2 was experimentally accessible and structure function seemed not to depend on q. When more data became available, as it is shown in figure 4, it became evident that structure functions depend on q^2 in a well-defined way. Accordingly, PDFs, linked to structure functions through 1, depend on q^2. The dependence on q^2 is perturbatively computable in quantum chromodynamics: it is governed by a set of integro-differential equations, *evolution equations*, which take the form:

$$Q^2 \frac{\partial^2}{\partial Q^2} f_i(x,Q^2) = \sum_{j=1}^{n_f} P_{ij}(x, \alpha_s(Q^2)) \otimes f_j(x,Q^2).$$

Here the index i and j are referred to the flavour of the quark (anti-quark, gluon ...), α_s is the strong coupling constant and P_{ij} are perturbative kernels (calculated perturbatively). Theoretically, also the x-dependence is fixed by first principles in QCD, but the actual physical law cannot be computed, since we are in the non-perturbative regime of QCD, i.e. it is not possible to derive physical results using perturbation theory.



(a) (b)

Figure 4: These two figures represent the experimental evidence for the violation of the so-called scaling hypothesis. On the left, it is possible to see that the structure function does actually depend on q^2, in a well-defined way. The scaling hypothesis was due to the fact that the first experiments were carried out only for a small range of q^2 and x. On the right, two different structure functions for different values of q^2. The images come from [1].

2 PDF DETERMINATION WITH NNS

2.1 THE NNPDF APPROACH

In this section we investigate in detail the NNPDF approach for PDF determination. The distinctive characteristics of this approach are the use of neural networks for PDF parameterization and the use of Monte Carlo representation of uncertainties. The fitting process can be schematically represented as in figure 7. At the end of this section all the steps of the fitting methodology in that picture should be clear. This section is mainly inspired by [3], [5].

2.1.1 PDF determination as an AI problem

The basic idea of using artificial intelligence to infer PDFs is to see the problem of PDF fitting as a specific type of pattern recognition problem: we want to know a set of unknown underlying functions from data instances, with almost no knowledge on their functional form.

Historically, the first type of parameterization used to fit PDFs (at a certain scale $Q_0 = \sqrt{Q^2}$) was:

$$f_i = x^{\alpha_i}(1-x)^{\beta_i}.$$

Clearly, in this way, we are assuming a specific form for the parameterization of PDFs, based on the assumption of having a power-like behavior as $x \to 0$ and as $x \to 1$. However, this assumption cannot be verified since only the region $10^{-4} \lesssim x \lesssim 0.5$ is experimentally accessible. Another problem is the fact that even if such a functional form is true at a certain scale Q_0, using evolution equations it is possible to see that this functional form is not preserved at any scale.

More sophisticated forms of parameterization were developed over the years, to overcome these problems. The underlying issue, common to this new ways of parameterizations, was the evaluation of uncertainties: uncertainties on fit parameters, determined with standard error propagation, resulted too small when new data were available. Furthermore, the addition of new data often led to a more complex parameterization, which led to an increase of uncertainties. These evidences suggest that each type of parameterization proposed was affected by bias. In this framework, the purpose of using artificial intelligence is to provide a bias-free parameterization.

However, in contrast to normal pattern recognition problems, the functions provide a continuous output and data depend only indirectly on the functions to be determined (the dependence is through the convolution integrals). This feature is common to more complex pattern recognition problems, but there are two peculiarities typical of this specific problem.

The first one is the intrinsic non-deterministic nature of quantum mechanics: PDFs are probabilities densities of observables, meaning that even if PDFs were known exactly, the cross section is just an expression of the probability of the observation of an event. In PDFs determination, we are determining the probability distribution of PDFs, hence a probability distribution of probabilities distributions, in other words, we are determining a probability functional.

The second one is the problem of estimation of uncertainties: currently PDF uncertainties are a dominant source of uncertainties in predictions for high-energy physics measurements. The additional complexity in this case, which differenziate this problem with respect to other similar AI problems, is that uncertainties among two different PDFs are correlated. This type of correlation has to be taken into account through a covariance matrix of uncertainties in the space of probabilities distributions. In particular, this matrix includes both experimental and theoretical uncertainties. Including also theoretical errors has the effect of connecting different measurements, which are experimentally uncorrelated.

2.1.2 The method in detail

In order to break down the problem of determining a probability in a space of functions to a set of problems in which a unique best-fit set of functions is determined, we use a Monte Carlo representation. In practice, this means that we start from the probability distribution in the space of data, whose mean and covariance are coherent with experimental values and correlated uncertainties. The Monte Carlo representation is obtained extracting a set of replica instances from the probability distribution, in such a way that, with a sufficiently large number of replicas, mean and covariance of the set of replica reproduce the ones of the underlying distribution.

The number of replicas is actually determined a posteriori and, for the work here proposed, the number of replicas is fixed to 100. Afterwards, PDFs are fitted through neural networks, minimizing a suitable figure of merit, i.e. an appropriate metric of performance. Finally, we end up with a set of PDF replicas, one for each data replica: we have now a representation of probability densities in the space of PDFs. By usual tool of statistics, it is possible to determine central values, uncertainties and correlations.

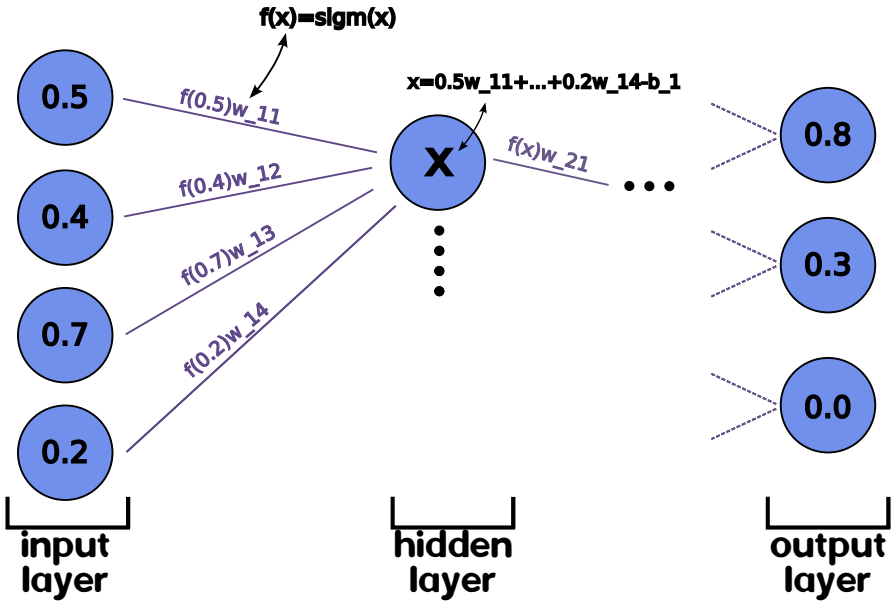


Figure 5: Here a schematic representation of a NN is depicted. Each node has as input a weighted sum of all the outputs of the nodes belonging to the previous layer, minus the threshold. The input of a node is the input of the activation function, the output of the function is the output of the node, which will enter as an input for the next layer.

As mentioned above, data are compared to theoretical predictions through convolution integrals (using the property of factorization and perturbatively solving evolution equations). In practice, the convolutions are turned into multiplications of pre-computed tables (Fast-Kernel tables).

2.2 Brief review of Neural Network principles

This section briefly explains the basic ideas behind the mechanism of Neural Networks (NNs), only scratching the surface of this field: we present the simplest kind of NN possible. NNs are classic tools used in machine learning, originally inspired by the structure of the brain. Their basic purpose is to find the underlying function that, given a certain input, provides a certain output.

Recalling the names of brain components, the structure consists in a set of nodes, called *artificial neurons*, connected by edges. Neurons are organized in *layers*, as shown in figure 5. When the NN receives an input, it is processed from the input layer to the output one, crossing the so-called hidden layers. For example, the NNPDF layer structure is 2-25-20-8, as in figure 6.

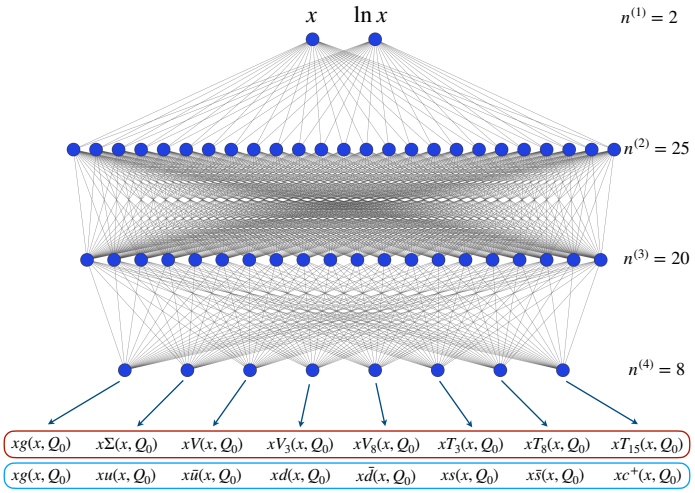


Figure 6: Architecture of the neural network used in NNPDF code. This image is taken from [2]

Starting from the left, each input value is taken as an argument of a so-called *activation function*. A common choice for this type of functions are, for example, sigmoids or hyperbolic tangents. When the chosen function is a step function, the node is called *perceptron*. The introduction of this function is fundamental, because it allows to take into account nonlinear behaviors. For each node, the output of the activation function is then multiplied by a weight (a free parameter of the network) and finally all these terms are summed over. Often to the sum is subtracted another parameter, called bias (or activation threshold), which is another free parameter of the network.

This final sum is stored in the corresponding node belonging to the next layer and it is taken again as input of an activation function. Repeating this process, we cross all the layers, up to the output.

With such a structure it is easy to see that by introducing just few nodes, as in NNPDF case, we give rise to a huge number (hundreds in our case) of free parameters.

Now, in order to find the optimal value for each of this parameters, we have to make the NN undergo a process of *training*, which has the purpose of determining the best values for the free parameters. This, in the context of machine learning problems, is usually done with input-output pairs, but, as discussed above, this doesn't work for our specific case. However, the general strategy remains the same: we want that the difference between the output computed by the neural network and the expected output to be the minimum possible, varying all the parameters. Moreover, we want to minimize a suitable chi-squared figure of merit, based on the difference between

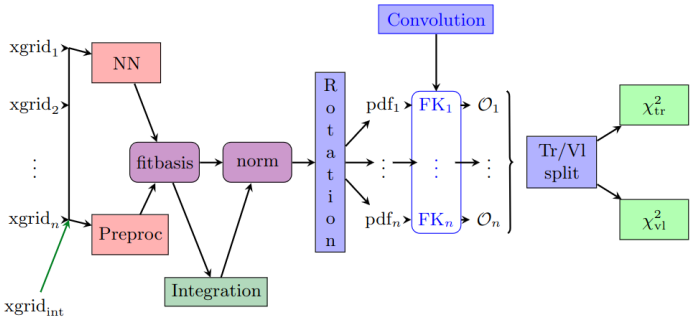


Figure 7: Global strategy for PDFs determination. This image is taken from [3].

output and experimental results, varying the parameters. Specifically, the figure of merit minimized is the following:

$$\chi^2 = \sum_{i,j}^{N_{dat}} (D - P)_i \sigma_{ij} (D - P)_j, \tag{2}$$

where D_i is the i-th datapoint, P_i is the prediction of the corresponding datapoint and σ_{ij} is the covariance between datapoints i and j. This minimization is carried out thanks to an algorithm of gradient descent (GD), applied to the space of parameters on which the NN depend.

For completeness, we also quote here a theorem that proves the actual convergence for neural networks fits. The so-called *universal approximation theorem* states that any continuous function $f : [0,1]^n \to [0,1]$ can be approximated arbitrarily well by a multi-layer perceptron with at least 1 hidden layer and a finite number of hidden units.

2.3 NNPDF use of Neural Networks

In our case, PDFs are parametrized as:

$$xf_i(x, Q_0) = A_i x^{-\alpha_i+1}(1-x)^{\beta_i} NN_i(x). \tag{3}$$

Here, A_i is just an overall normalization constant, and $x^{-\alpha_i}(1-x)^{\beta_i}$ is a prefactor mainly used in order to accelerate the convergence speed. The parameters of α_i and β_i are uniformly randomly chosen across a wide range (a posteriori determined). The structure of the neural network can be seen in figure 6.

The input is provided both in the form x, $\log x$ in order to take into account the two different regimes experimentally accessible: $10^{-4} \lesssim x \lesssim 0.03$ (in logarithmic regime) and $0.03 \lesssim x \lesssim 0.5$ (in linear regime).

At this point of the discussion, it should be clear that the choice of NN as fitting methodology was determined by the necessity of an

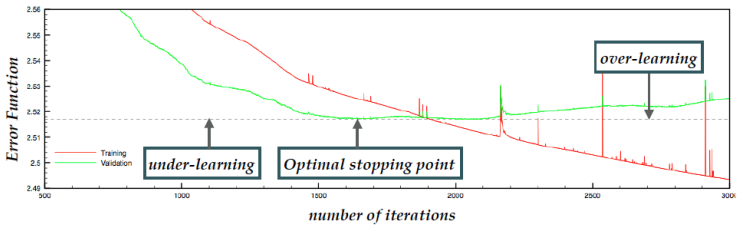


Figure 8: The picture shows the different behavior of the chi squared of the 2 different sets (validation and training), as the NN fits. At the beginning we have a similar behavior for the two sets. After passing the optimal stopping point, the two behaviors are opposite: the χ^2_{val} starts rising, while the χ^2_{tr} is continuing to decrease. This image is taken from the NNPDF website.

unbiased way of fitting. In other words, NNs provide the opportunity of fitting without a specific fixed functional form as in the past. This technique, however, carries some drawbacks with it. One of this is overfitting. In fact, if we leave the NN training for an arbitrary long time, the NN will fit not only the sought PDF, but also the underlying statistical noise. In order to overcome this inconvenience, a specific stopping criterion is adopted. For each replica, the data are split into two sets: validation and training. On the training set we minimize the chi-squared with all the procedure mentioned above. At the same time, the chi-squared of the validation set is constantly monitored.

Assuming that the physical law we want to fit is predominant above the noise, initially the two chi-squared decrease. At some point the algorithm will start to fit more noise than the physical law. At this point the chi-squared of the validation set attains its minimum. After this point, the more noise is fitted, the more the chi-squared of the validation set increases. Therefore, in order not to fit noise, the stopping point is chosen looking at the minimum for the chi-squared validation set. A schematic representation of the overall process is illustrated in figure 8.

To sum up, in figure 7 the global strategy is depicted. We start calculating a fit with the neural network at a certain reference value Q_0 (to do this, we calculate the different contributions $x^{\alpha_i}(1-x)^{\beta_i}$, $NN(x)$ and we calculate by integration the normalization pre-factor A_i). In order to evolve the fitted PDFs to different scales, it is useful to evolve linear combinations of PDFs, which decouple many of the evolution equations. Namely, through a rotation, we go from the physical basis, to the evolution basis, to calculate the evolution.

In practice, the evolution and the theoretical calculations are stored in precalculated Fast-Kernel tables. After the different evolutions at different values of Q (corresponding to the different values at which the experiments are carried out) are calculated, we compare the re-

sults predicted with the calculations to the experimental data. In particular, the chi-squared is calculated separately for the two sets of data validation and training. The fitting procedure is then iterated minimizing the training chi-squared. The best-fit is chosen looking at the minimum for the validation chi-squared.

2.4 On the scale dependence of parton parameterizations

The purpose of this work is to check the validity of this fitting method, verifying the correctness of some underlying assumptions. Namely, our results should not depend on the value of Q_0 chosen to fit the PDFs, since the Q dependence is fixed by the evolution equations. In particular, the current PDFs are calculated for a value of $Q_0 = 1.65\,\text{GeV}$.

In practice, we computed the different fits for different values of Q_0 and we compared the results with the reference fit at scale $Q_0 = 1.65\,\text{GeV}$. The additional subtle point has to deal with the different form of the perturbative solution of evolution equations within different values of Q_0. In fact, there are different regimes of Q^2, where the effects of heavy quarks on PDFs are different (different renormalization and factorization schemes are adopted).

We will consider as representative the case of the charm quark. In the cases where Q^2 is much bigger than m_c^2, the natural choice to treat this quark is to consider it massless (all terms of order $\frac{m_c^2}{Q^2}$ are neglected). On the other hand, below m_c^2, a decoupling renormalization scheme is usually used, which effectively removes heavy quarks' contributions. This means that, below the mass of the charm, we consider $n_f = 3$, while above we consider $n_f = 4$. In terms of evolution equation, this means that the charm PDF will start evolving only after passing the value of the charm mass. Clearly, the two different regimes, high and low Q^2, should be in some way matched near the value of m_c^2, where terms of $\frac{m_c^2}{Q^2}$ cannot be neglected anymore. Without going too deep in the discussion (for a deeper detail, see [6], [7]), another scheme is used to consider this contributions, just above the threshold of the charm mass. Finally, the matching is given by perturbative computable matching conditions. Still, changing the value of Q_0 near the threshold remains a delicate aspect, and this will be an important point in the next section.

The other important feature of evolution equation to keep in mind is the stability if we raise or lower Q_0. If we assume to consider the limiting case where $Q \to \infty$, then, as explained above, we reach the limit of *asymptotic freedom*: quarks are not experiencing anymore

strong interaction. This means, in other words, that different PDFs at finite Q_0 converge to the same limiting case, as Q_0 goes to infinity. What we are saying is that, if we start from a low Q_0, the evolution to a higher Q is, in general, stable, since we are converging to an unique limit. On the other hand, if we start from a PDF calculated at high Q_0, we have the opposite problem: a small change in the function at Q_0 causes a large change in the function at small Q.

In this sense, in principle, we expect, keeping Q_0 low, results to be more stable: the fitting procedure uses the evolution equation in the stable sense (from low Q_0 to high Q) and, after the fitting is complete, it is possible to evolve the PDFs at higher scales without any problem. In practice we will verify this in the next section.

3 Results

The purpose of this section is to present the results of the calculations carried out with different values of Q_0. Specifically, the values of Q_0 are varied firstly within the usual thresholds for heavy quark production, and then verifying what happens when passing them. Currently, $Q_0 = 1.65\,\text{GeV}$, while $m_c = 1.51\,\text{GeV}$ and $m_b = 4.92\,\text{GeV}$. In particular, many fits are iterated more than one time, as explained in the specific section and results are mainly presented after iteration.

Since many different plots are available for each calculation, few representative plots are chosen to be presented in the text: the up quark (u) and the gluon (g). In addiction, we present the *distances* plots: in these plots we present the absolute value of the difference between the fits calculated at different scales, divided by the square root of the sum of uncertainties squared. These plots should oscillate around the value of one, since the difference between the two plots should be due to statistical uncertainty. PDF plots are presented in logarithmic scale, distances both in linear and logarithmic scale. Further plots are present in the appendix. All the fits calculated within this work are based on NNPDF4.0 code, available at the NNPDF webpage.

3.1 Results before iteration

3.1.1 $Q_0 = 2.00\,\text{GeV}$ & $Q_0 = 1.52\,\text{GeV}$

In this section we present the results for a small change with respect to $Q_0 = 1.65\,\text{GeV}$: for $Q_0 = 2.00\,\text{GeV}$ (figure 9) and $Q_0 = 1.52\,\text{GeV}$ (figure 10). As a consequence, results are compatible within the given uncertainties and the χ^2, table 1, remains optimal.

3.2 Need of a second iteration

Calculating fits as in the above section, we noticed that the further we move from the reference scale $Q_0 = 1.65\,\text{GeV}$, the larger is the change in the results. This trend can be appreciated looking at the chi-squared in table 1. These differences are solved if we iterate fits, as

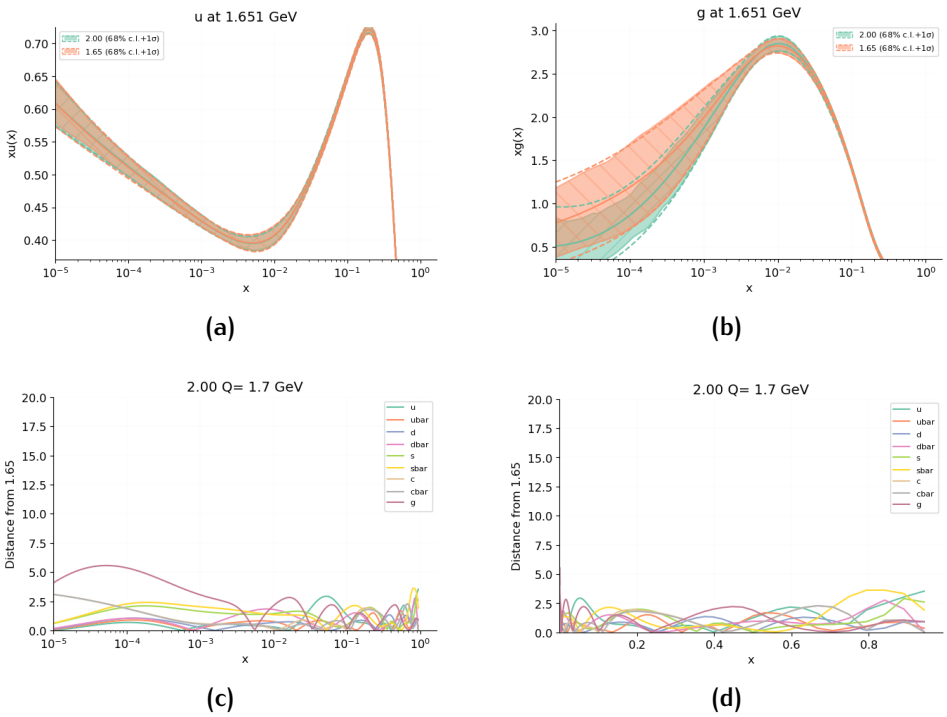


Figure 9: Plots for $Q_0 = 2.00\,\text{GeV}$

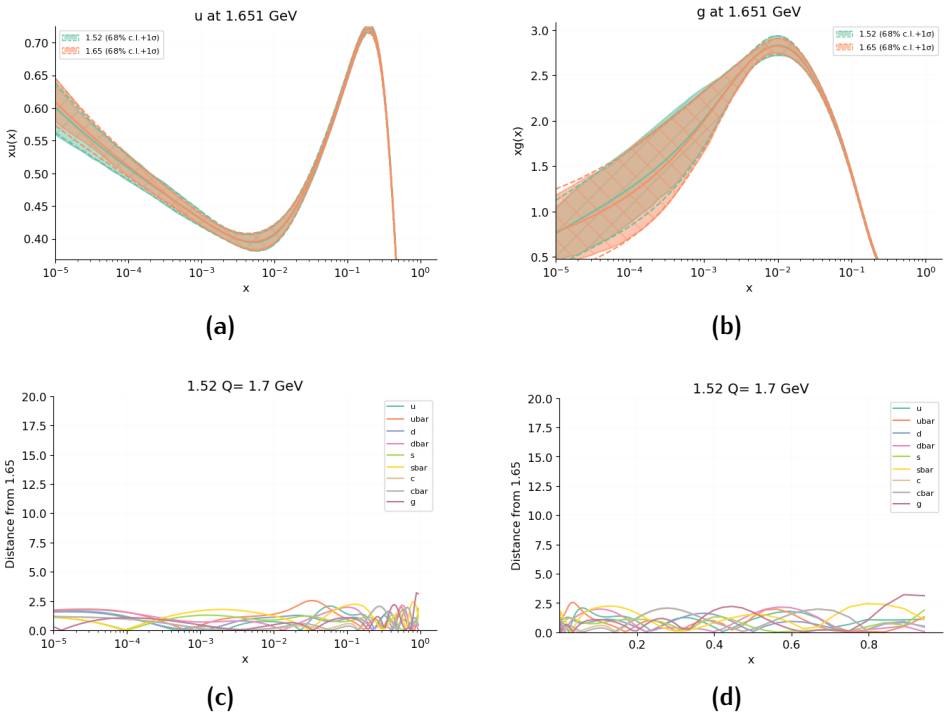


Figure 10: Plots for $Q_0 = 1.52\,\text{GeV}$

provided for in NNPDF methodology, and thus they can be explained thanks to two different aspects.

Table 1: Different values of χ^2 for different fits, not iterated.

Q_0[GeV]	χ^2
1.65	1.171
2.00	1.170
1.52	1.171
3.00	1.192
4.00	1.211
4.90	1.240
1.50	1.168
4.93	1.240

The first is the value of the pre-processing factor present in equation 3. This factor has the purpose of increasing the speed of convergence for the neural network. The coefficients α_i and β_i are randomly and uniformly chosen in a wide interval, a-posteriori determined, in order to be significantly wider with respect to the confidence interval of the effective values of α_i and β_i fitted. When changing the value of Q_0, also the effective values of these factors are changing. In other words, if we leave the exponents in the form they are at a different scale, we are no more helping the NN with the fitting, but we are externally imposing an incorrect functional form that the NN has to correct. With iteration, the exponents' interval for the new fit is calculated from the effective exponents of the previous one.

The second effect we are facing is due to the way used to estimate the covariance matrix. In particular, we have to separately consider the cases of multiplicative errors, such as normalization errors, and addictive ones. When considering multiplicative errors, with different correlation, there's an effect, the so-called *D'Agostini* bias, affecting our results. An uncertainty is multiplicative when its value σ is proportional to the quantity d that is being measured: $\sigma = \sigma_{rel} d$, where σ_{rel} is fixed. The bias arises if the value used for d is also the prediction which is being fitted: $d = t$. The solution consists of using instead for d the value of t_0 obtained from some previous existing determined fit. The best fit value of t is used for a new fit, and the procedure is iterated with convergence.

3.2.1 $Q_0 = 4.90\,\text{GeV}$

To give an example of why fits need to be iterated, we present here the case of $Q_0 = 4.90\,\text{GeV}$, just below the bottom mass. For this value of Q_0, results are no longer satisfactory: the chi-squared has incresed (table 1) and the shape of PDFs plot (figure 11) is different from the reference one. A more detailed look at in these first results, analysing

training lengths, reveals that the neural network is not learning well. Training lengths represent the number of steps of our algorithm of gradient descent, until it reaches the minimum.

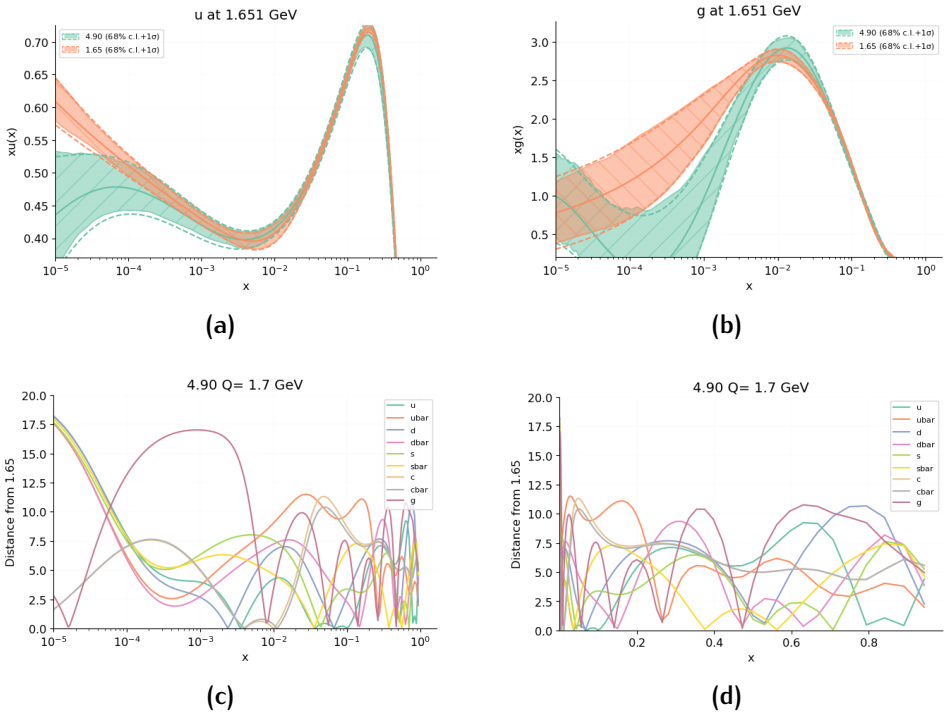


Figure 11: Plots for $Q_0 = 4.90\,\text{GeV}$, before iteration

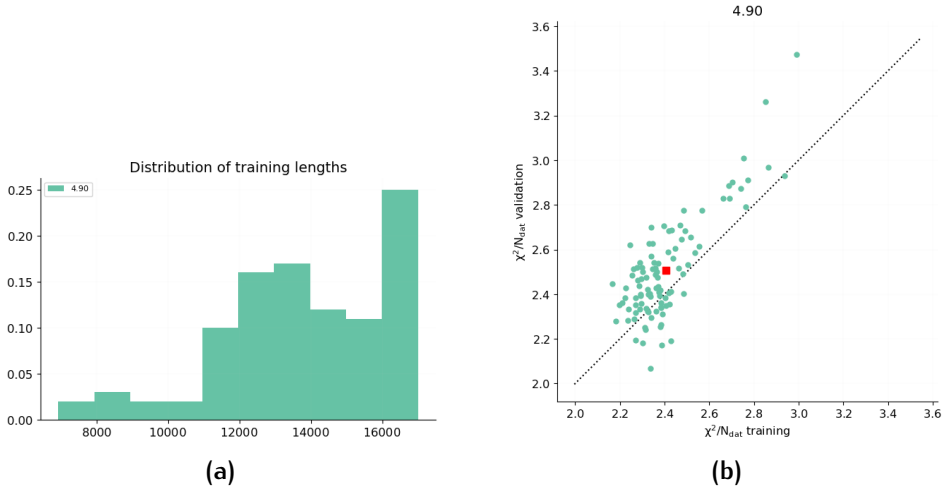


Figure 12: Distribution of training lengths (a) and χ^2_{tr}-χ^2_{val} plot for $Q_0 = 4.90\,\text{GeV}$ (b), before iteration. In the appendix we present, as a reference, the same graphs for $Q_0 = 1.65\,\text{GeV}$.

The number has a maximum value possible, in order not to have the algorithm work for an arbitrary long amount of time. The graph

in figure 12a shows the distribution of training lengths among replicas, i.e. on the *y-axes* we have the number of replicas which stopped with a certain value of training length. This maximum value for training lengths is set to be 17000 by default, so a peak at the right part of the graph indicates that many replicas are not reaching the minimum sought by the GD algorithm. Furthermore, another indication of this problem can be seen in the χ^2_{tr}-χ^2_{val} plot, figure 12b. We can see that there is a certain correlation between the values of the two chi-squared. This is an indicator that the algorithm is not fitting well: if the algorithm is stopped before reaching the minimum, we expect to have a correlation between the two chi-squares. Furthermore, this is not a representation of all the chi-squares. In fact, within the replica process, replicas with a chi-squared too large (bigger than 3.5) are cut away. This procedure has the purpose of cutting away outliers of the distribution and, in general, few replicas are excluded (less than 10%). In this case it was necessary to exclude more than 25% of replicas produced. In other words, we didn't have anymore a faithful representation of probability distribution of PDFs in the space of functions.

We therefore iterated the results and we got a significant improvement both in the chi-squared and in the plots (figure 13). Also the

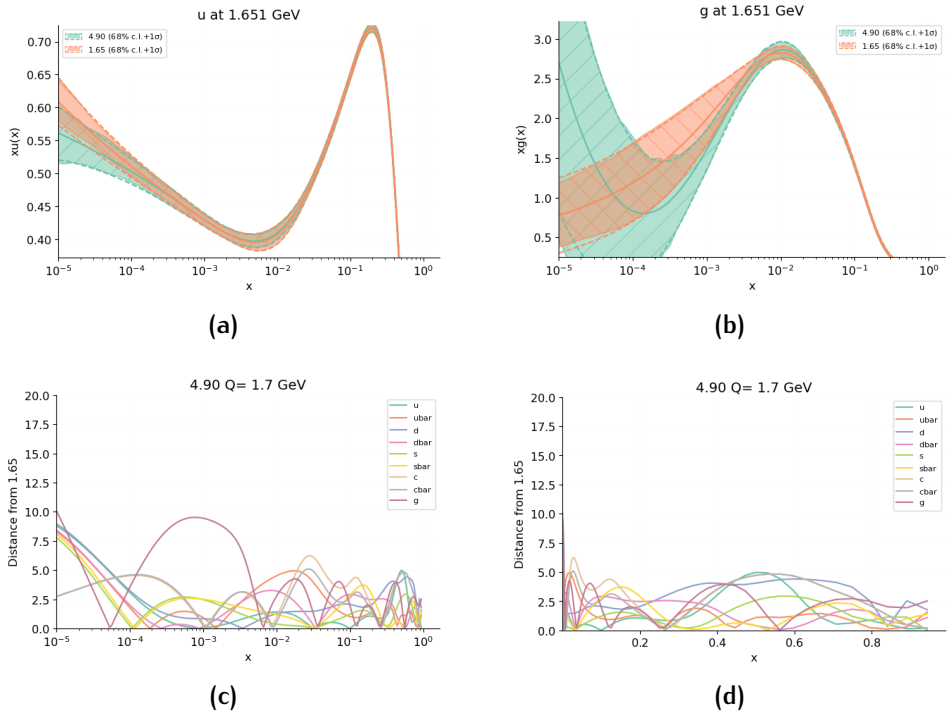
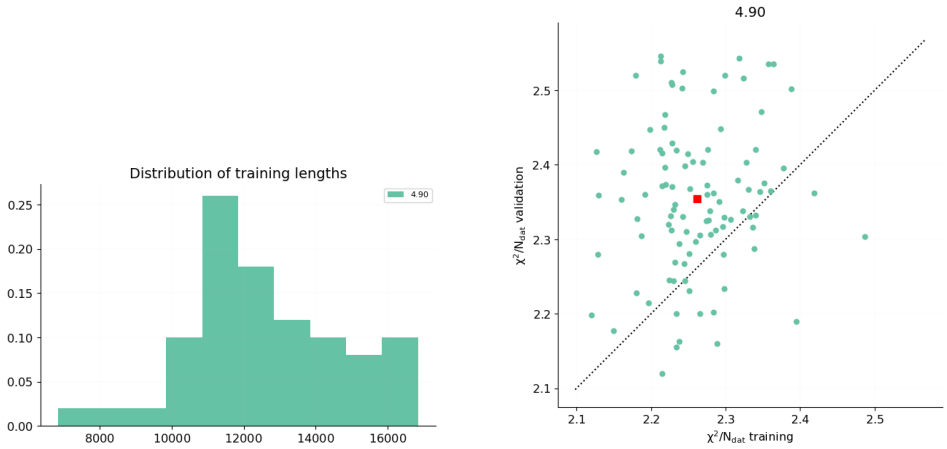


Figure 13: Plots for $Q_0 = 4.90\,\text{GeV}$, after iteration

distribution of training lengths and the χ^2_{val}-χ^2_{tr} plot is consistent (figure 14). On the other hand, the behavior at small x remains slightly different, expecially for the gluon.



(a) (b)

Figure 14: Distribution of training lengths (a) and χ^2_{tr}-χ^2_{val} plot for $Q_0 = 4.90\,GeV$ (b), after iteration. In the appendix we present, as a reference, the same graphs for $Q_0 = 1.65\,GeV$.

3.3 Results after iteration

3.3.1 $Q_0 = 3.00\,GeV$ & $Q_0 = 4.00\,GeV$

As we can see in figure 15 ($Q_0 = 3.00\,GeV$) and 16 ($Q_0 = 4.00\,GeV$), the plots are compatible with the reference plot, especially for quarks. Again the chi-squared is better. However, it is still present a different behavior for the gluon at small x.

3.3.2 $Q_0 = 4.93\,GeV$

As in the case of $Q_0 = 4.90\,GeV$, the results (figure 17) are consistent both in term of chi-squared and plots. However, the gluon has always a different behavior at small x. We notice that the fact of having passed the threshold of $m_b = 4.91\,GeV$ does not affect the plots in any way.

3.3.3 $Q_0 = 1.50\,GeV$

The results (figure 18) are stable, expecially in terms of chi-squared (table 2). We therefore notice that also passing the threshold of m_c does not affect our results, and even the charm PDF is consistent. Also the distribution of training lengths and the χ^2_{val}-χ^2_{tr} plot (figure 19b) is consistent. However, we found that the distribution of training lengths shows a peak at 17k (figure 19a). Therefore, we tried to increase a little bit the number of epochs, up to 20k, in order to see if the results were better. After this change plots are stable (figure 20),

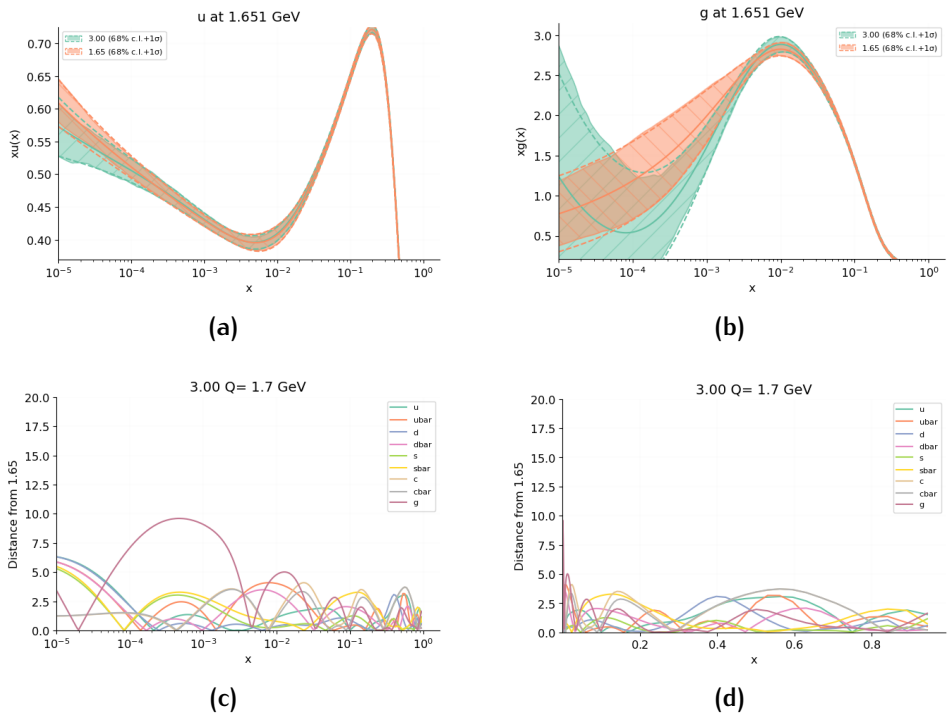


Figure 15: Plots for $Q_0 = 3.00\,\text{GeV}$, after iteration

Table 2: Different values of χ^2 for different fits, after iteration. The term −20k indicates the fit done setting the number of epochs to 20k.

Q_0[GeV]	χ^2
3.00	1.168
4.00	1.170
4.90	1.174
4.93	1.173
1.50	1.166
1.50 − 20k	1.164

the chi-squared decreases and the distribution of training lengths (figure 21a) has a more natural shape.

3.4 On the gluon behavior at small x

In this section, we try to investigate why the gluon has an unexpected behavior at small x, as shown in the previous sections. The small x behavior is not driven by the data, because data stop around $x \approx 10^{-3} - 10^{-4}$. The value and uncertainty at small x are known to a large extent based on guessing a reasonable extrapolation. Preprocessing at last input drives its extrapolation. The singlet and gluon PDFs at small x display a rise driven by perturbative evolution stronger than

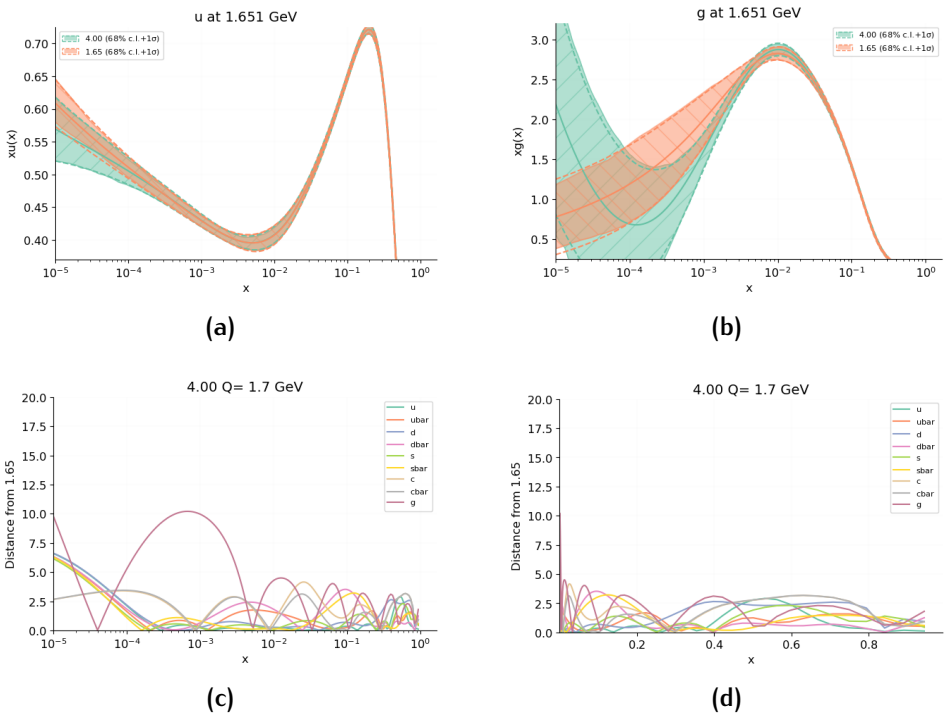


Figure 16: Plots for $Q_0 = 4.00\,\text{GeV}$, after iteration

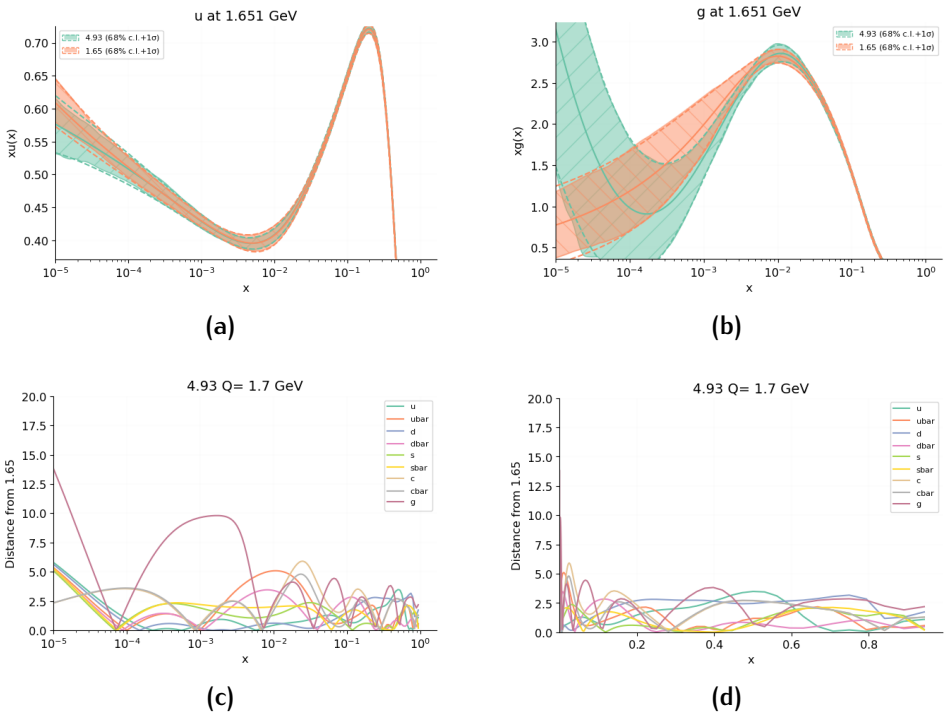


Figure 17: Plots for $Q_0 = 4.93\,\text{GeV}$, after iteration

any power of $\ln x$, but weacker than any power of x, and stronger as Q^2 increase. Therefore, if preprocessing exponents are computed at a

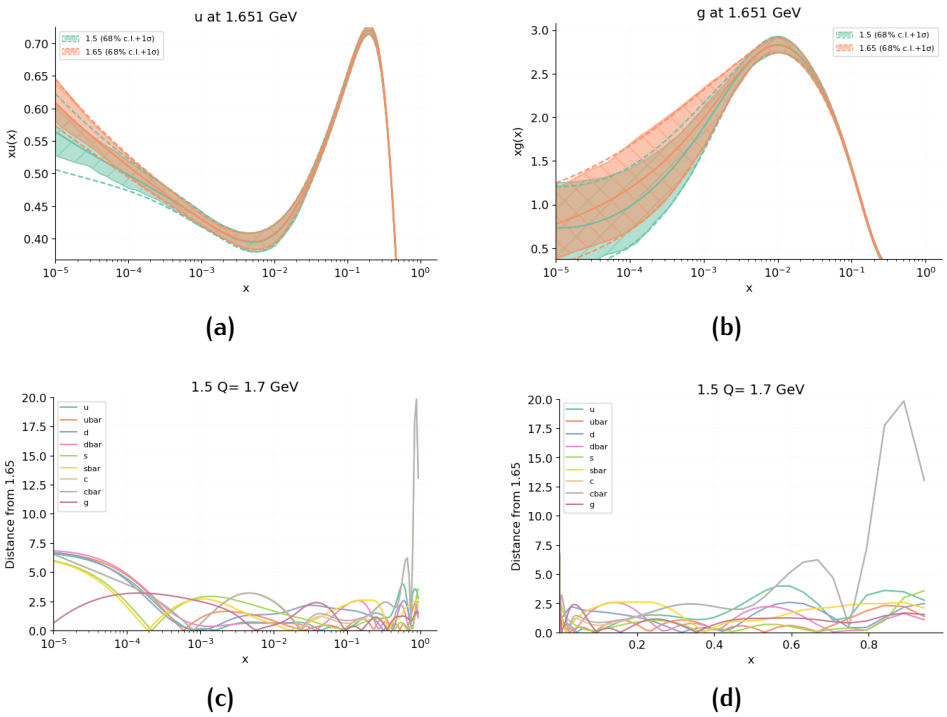
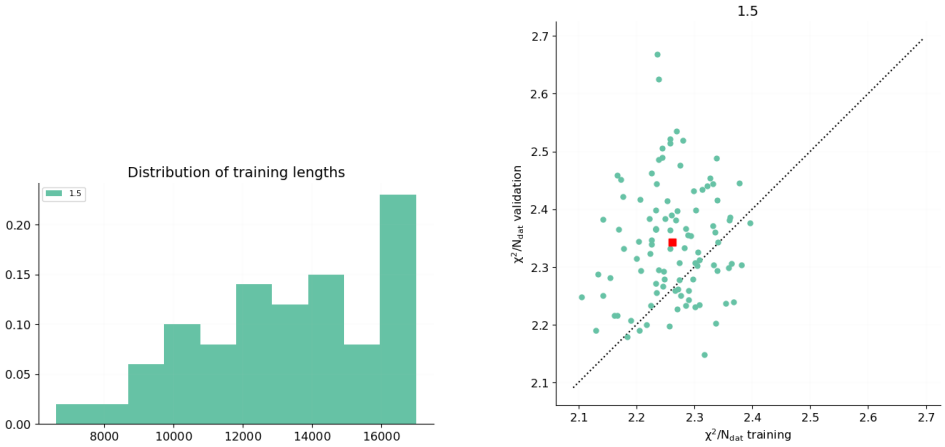


Figure 18: Plots for $Q_0 = 1.50\,\text{GeV}$, after iteration

different scale, the value of x at which they are determined must be changed in order for the results to remain the same.

By default the preprocessing exponents are computed at $x = 10^{-3}$ and $x = 10^{-6}$. We have verified that if computed at 10^{-3} only and then iterated (as in figure 22), the small x behavior becomes essentially the same as that of the default for quarks and more similar to that of the default for gluons. A dedicated study should be required in order to quantify the choice of x values as the scale is changed in a way that preserves the small x behavior.



(a) (b)

Figure 19: Distribution of training lengths (a) and χ^2_{tr}-χ^2_{val} plot for $Q_0 = 1.50\,GeV$ (b), after iteration. In the appendix we present, as a reference, the same graphs for $Q_0 = 1.65\,GeV$.

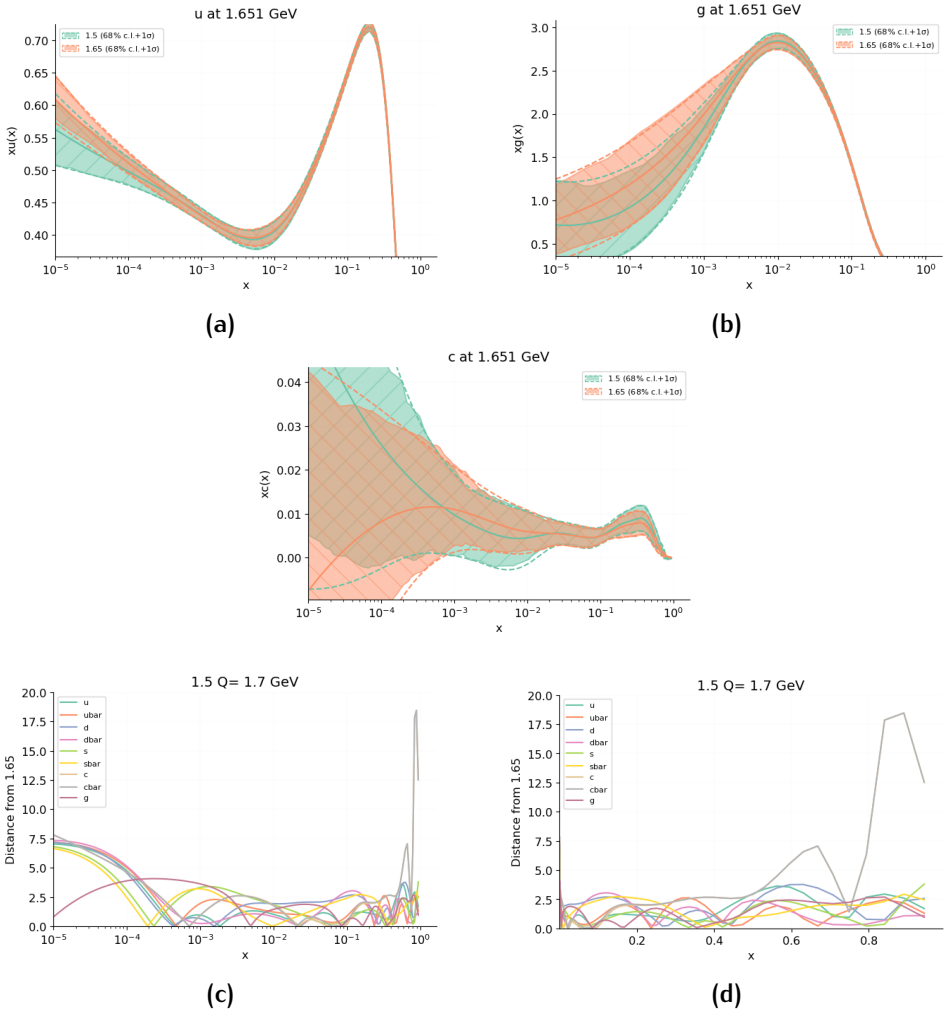
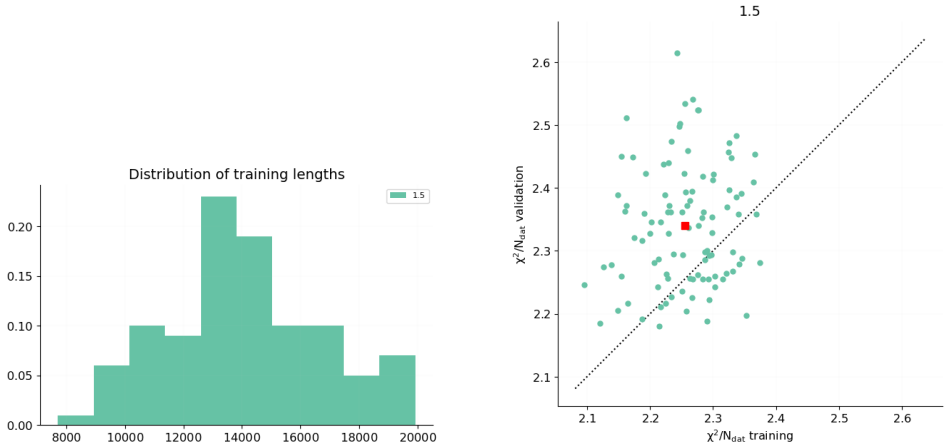
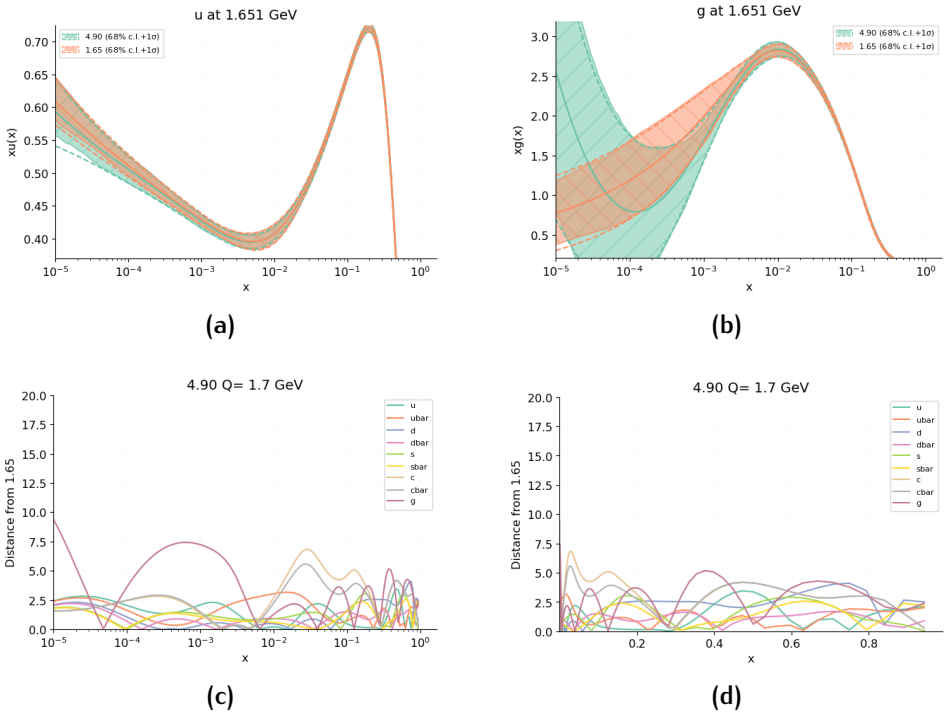


Figure 20: Plots for $Q_0 = 1.50\,\text{GeV}$, after iteration and with the number of epochs set to 20k. We decided to include also the charm PDF in order to show that it's behavior doesnt change passing through the threshold at 1.50 GeV.



(a) (b)

Figure 21: Distribution of training lengths (a) and χ^2_{tr}-χ^2_{val} plot for $Q_0 = 1.50\,\text{GeV}$ (b), after iteration and with the number of epochs set to 20k. In the appendix we present, as a reference, the same graphs for $Q_0 = 1.65\,\text{GeV}$.



(a) (b)

(c) (d)

Figure 22: Plots for $Q_0 = 4.90\,\text{GeV}$, iterated and with the pre-processing exponents calculated at $x = 10^{-3}$.

Conclusions

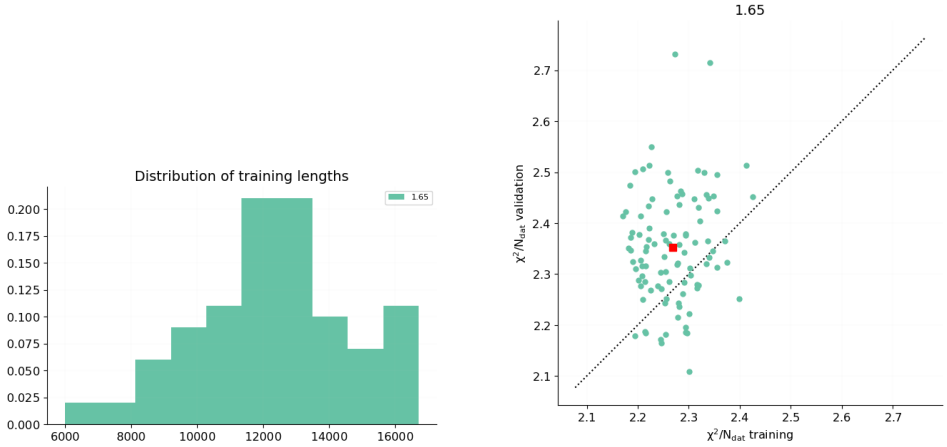
Here we sum up the results of this work. The purpose of the thesis was to analyze the dependence of PDFs on the scale at which they are parametrized. The results of this work confirm the fact that, in the data region, results are stable as the parameterization scale is varied. Results are unchanged when passing a heavy quark threshold. However, results are generally more stable, and the fit quality improves as the scale is lowered.

This last point can be understood thanks to the different stability of evolution equations, as explained in section 2.4. In particular, if PDFs are parametrized at a small value of Q_0, for the NN it is easier to infer the PDFs, since a big change in the shape at low Q_0 causes a small change at big Q.

We also found that PDFs, especially the gluon, have a slightly different behavior at small x. It is important to emphasize that in that region we don't have experimental data, therefore the NN is extrapolating. In this sense, uncertainties are big, and therefore the result found is still good. However, we tried to investigate this problem in greater detail as discussed in section 3.4. We found a strong indication that the value of x at which they are computed should be optimized as a function of the scale in order to obtain stable results.

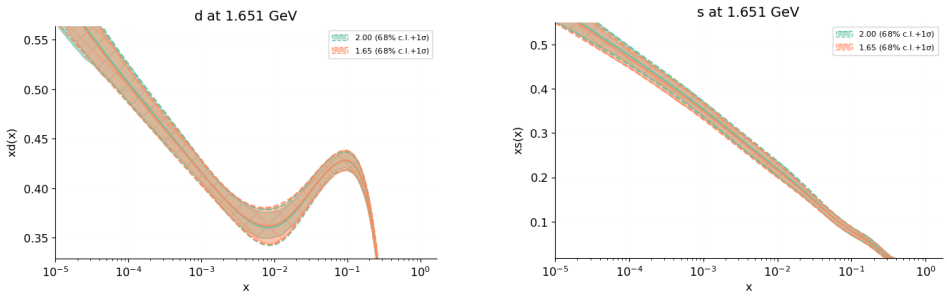
Appendix

We present here some additional plots, useful for the interested reader who wants to have a more complete view on PDFs plots. We choose to show plots before iteration and, for iterated fits, additional plots for down quark (d) and strange quark (s). In addition, on top of the appendix, we present the training lengths distribution and the χ^2_{val}-χ^2_{tr} plot for the 1.65 GeV, that can be compared to the same graphs at different scales.



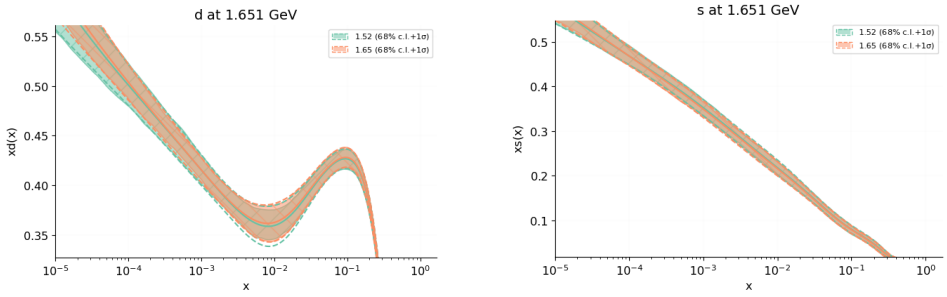
(a) (b)

Figure 23: Distribution of training lengths (a) and χ^2_{tr}-χ^2_{val} plot for $Q_0 = 1.65\,\text{GeV}$ (b).



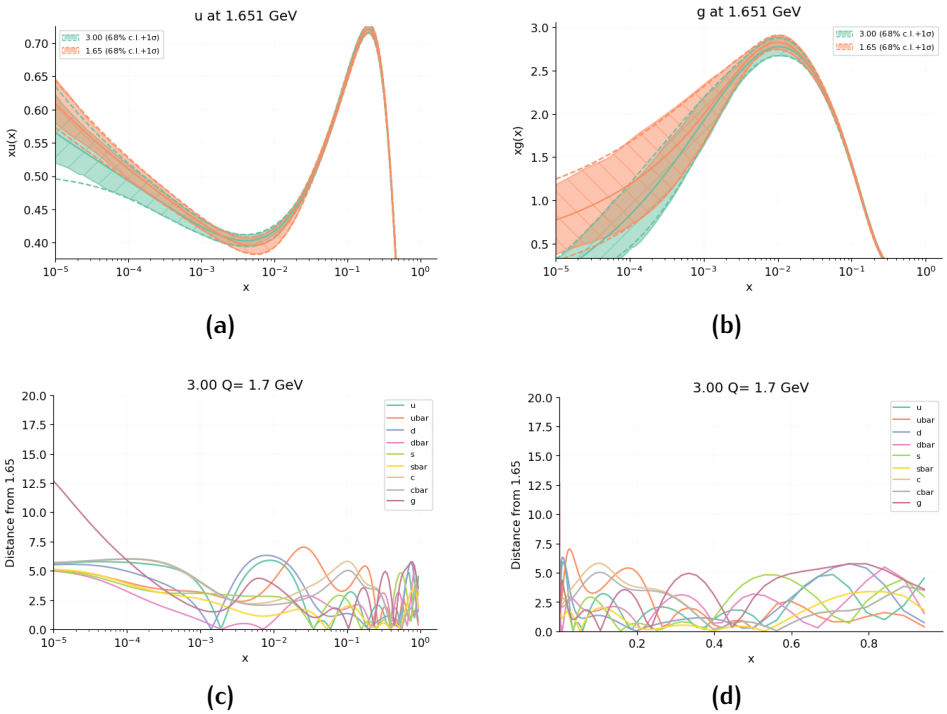
(a) (b)

Figure 24: Additional plots for $Q_0 = 2.00\,\text{GeV}$, before iteration



(a) (b)

Figure 25: Additional plots for $Q_0 = 1.52\,\text{GeV}$, before iteration



(a) (b)

(c) (d)

Figure 26: Plots for $Q_0 = 3.00\,\text{GeV}$, before iteration

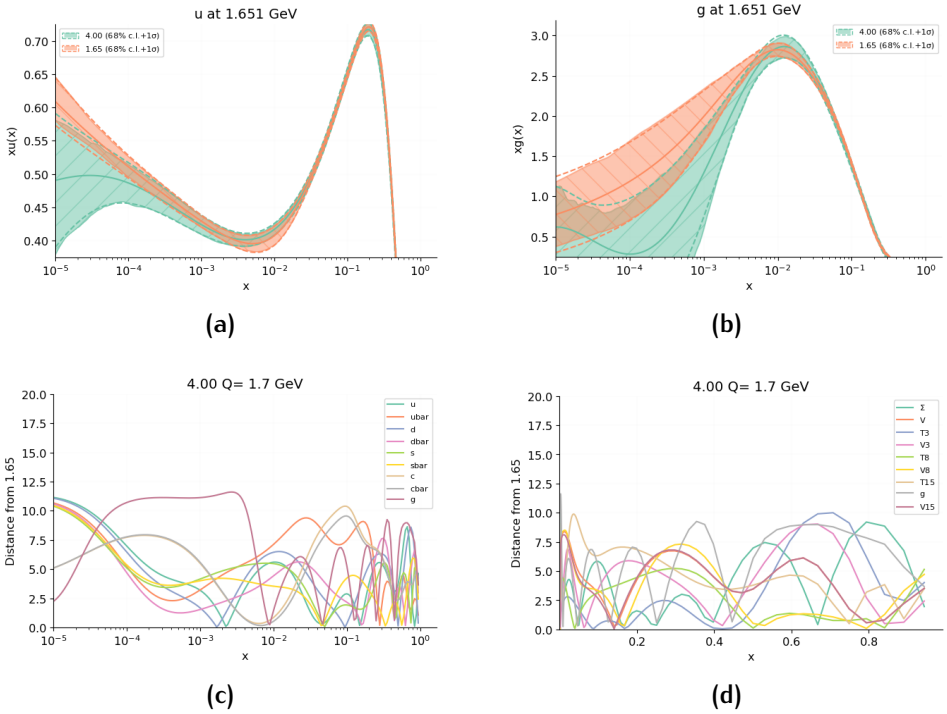


Figure 27: Plots for $Q_0 = 4.00\,\text{GeV}$, before iteration

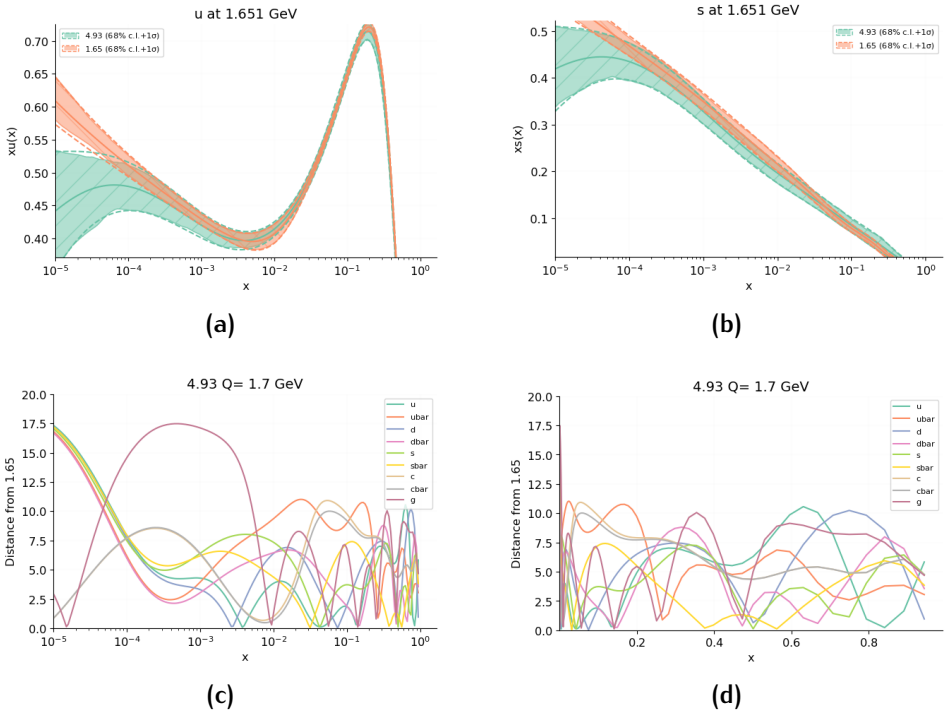
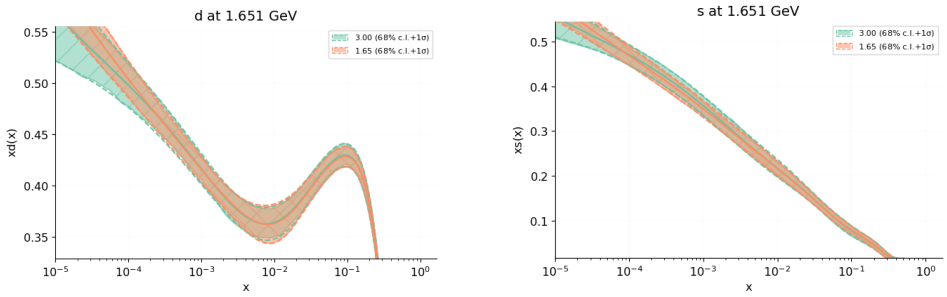
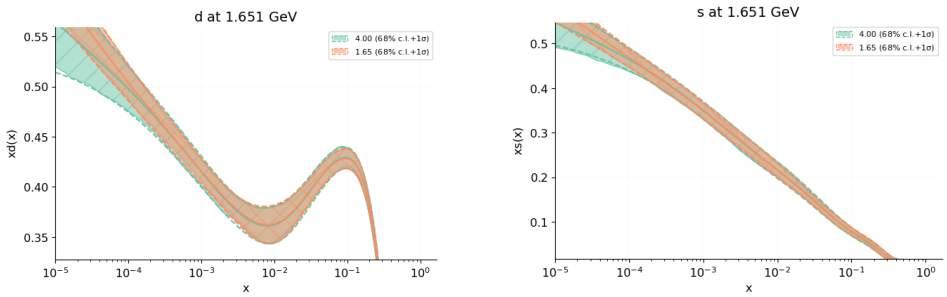


Figure 28: Plots for $Q_0 = 4.93\,\text{GeV}$, before iteration



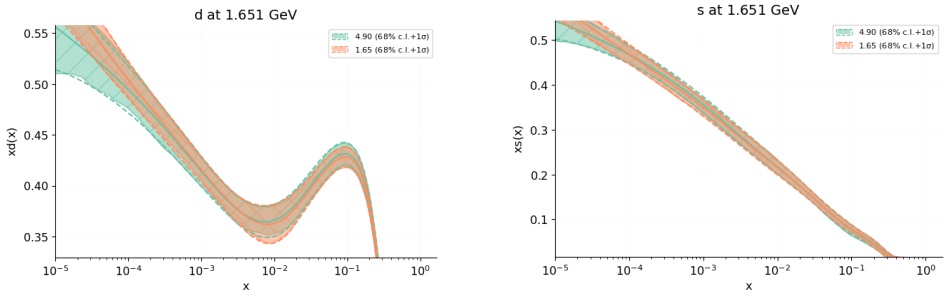
(a) (b)

Figure 29: Additional plots for $Q_0 = 3.00\,\text{GeV}$, after iteration



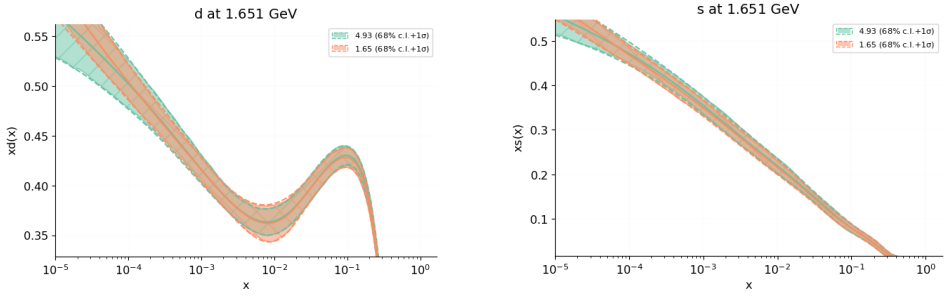
(a) (b)

Figure 30: Additional plots for $Q_0 = 4.00\,\text{GeV}$, after iteration



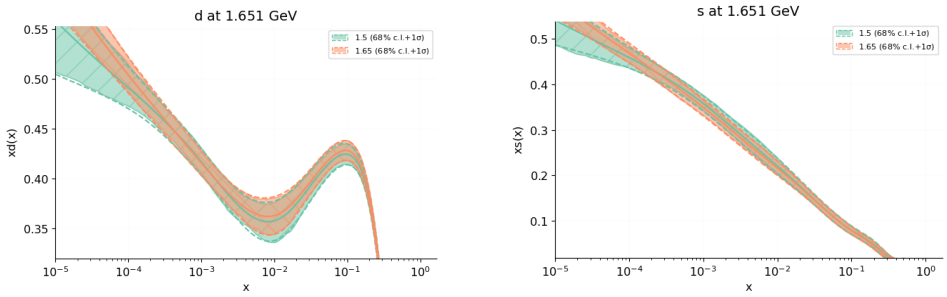
(a) (b)

Figure 31: Additional plots for $Q_0 = 4.90\,\text{GeV}$, after iteration



(a) (b)

Figure 32: Additional plots for $Q_0 = 4.93\,\text{GeV}$, after iteration



(a) (b)

Figure 33: Additional plots for $Q_0 = 1.50\,\text{GeV}$, after iteration and after having increased the number of epochs.

Bibliography

[1] J. E. Dodd and B. Gripaios. *The Ideas of Particle Physics*. Cambridge University Press, 3nd edition, 2006

[2] R. D. Ball, S. Carrazza, J .Cruz-Martinez, L. Del Debbio, S. Forte, T. Giani, S. Iranipour, Z. Kassabov, J. I. Latorre, E. R. Nocera, R. L. Pearson, J. Rojo, R. Stegeman, C. Schwan, M. Ubiali, C. Voisey, M. Wilson. *The Path to Proton Structure at One-Percent Accuracy*. arXiv:2109.02653, May 2022

[3] P. Calafiura, D. Rousseau and K. Terao. *Artificial Intelligence for High Energy Physics* (chapter 19: *Parton Distribution Functions* by S. Forte and S. Carrazza). World Scientific, Gen 2022

[4] J. Rojo. *The Partonic Content of Nucleons and Nuclei*. arXiv:1910.03408, Feb 2020.

[5] S. Carrazza and J. Cruz-Martinez, *Towards a new generation of parton densities with deep learning models*. arXiv:1907.05075, Jul 2019

[6] S. Forte, E. Laenen, P. Nason and J. Rojo. *Heavy quarks in deep-inelastic scattering*. Nucl. Phys. B **834** (2010), 116-162, doi:10.1016/j.nuclphysb.2010.03.014, arXiv:1001.2312

[7] S. Forte, *ACTA PHYSICA POLONICA B, Parton Distributions at the dawn of the LHC*. Nov 2010

[8] The NNPDF Collaboration: R. D. Ball, L. Del Debbio, S. Forte, A. Guffanti, J. I. Latorre, J. Rojo, M. Ubiali, *Fitting Parton Distribution Data with Multiplicative Normalization Uncertainties*. arXiv:0912.2276, Jun 2010

[9] G. D'Agostini, *Bayesian reasoning in data analysis - A critical introduction*, World Scientific Publishing, 2003

[10] A. Das, T. Ferbel *Introduction to Nuclear and Particle Physics*. World Scientific Publishing. Dec 2003

[11] A. Buckley, J. Ferrando, S. Lloyd, K. Nordstrom, B. Page, M. Ruefenacht, M. Schoenherr, G. Watt *LHAPDF6: parton density access in the LHC precision era*. arXiv:1412.7420, Dec 2014

Acknowledgements

First of all, I want to express my gratitude for professor Stefano Forte and his PhD student Niccolò Laurenti. Professor Forte firstly proposed this engaging work to me few months ago and I immediately embraced his proposal. During these months, he was always available and clarified every doubt or lack of understanding with passion and exactitude. Niccolò Laurenti helped me constantly, practically every day, with NNPDF code. He demonstrated great patience and kindness helping me to understand the functioning of NNPDF complex machine.

A heartfelt thanks go to my family, who supported me across these years both economically and giving me advice on the decisions of my life. Besides my parents, who are the reason why I am here, I want to remember my grandparents, especially my grandmother, who transmitted me her great passion for study and knowledge.

Secondly, I cannot forget my physics companion and my friends: Alice, Gabriele, Giulia, Ludovica, Marco, Matteo, Michele and in particular Emma and Simone, who supported me especially in these final months.

Finally, I am grateful to the all the professors I encountered throughout my studies at the University, underlying the fact that I never found just lecturers, but always persons keen on their subject and interested in the personal, besides academic, development of their students.