



**UNIVERSITÀ DEGLI STUDI DI MILANO**  
**FACOLTÀ DI SCIENZE E TECNOLOGIE**

Laurea Triennale in Fisica

**Correlation between statistical and physical  
properties of ensembles of Parton Distribution  
Functions**

Relatore:  
**Prof. Stefano Forte**

Correlatore:  
**Dott. Juan Cruz Martinez**

**Elisa Stabilini**  
Matricola n° 942529  
A.A. 2021/2022

*To all the women  
who have studied and will study  
physics*

# Summary

The purpose of this thesis is to study and verify the presence of correlations between the statistical properties of a PDFs-set ensemble, and their physical properties such as arclength or kinetic energy. In particular, this work is focused on the meaning of the different  $\chi^2$  values associated with each PDF replica obtained with the NNPDF code and its correlation with other replicas' properties such as their functional form.

In Chapter 1 we introduce the theoretical framework of this work, starting with a description of the standard model of particle physics (SM), and Quantum Chromodynamics (QCD). Then we define what a parton distribution function (PDF) is, why it is important in perturbative QCD and how PDFs can be in principle derived from experimental data.

Chapter 2 contains a brief introduction to the main concepts of Machine Learning, the classification of ML algorithms, based on the different strategies on which the "learning" relies, and the different purposes for which can be used. Then there's a focus on artificial neural networks and their properties starting from the simplest example of a neural network, the single-layer perceptron. In particular, we give the definition of *weight* and *activation function* in the context of ML, then we describe the processes of training and generalization of a neural network algorithm.

In Chapter 3 is described the general NNPDF strategy, explaining how the NNPDF approach allows researchers to avoid theoretical bias when inferring the PDFs' functional forms from experimental data, namely by using neural networks as unbiased interpolants. Then we proceed to describe the role of Monte Carlo replicas of the data in the NNPDF approach, and how the replicas are used for both training and validation.

Considering the importance of the  $\chi^2$  figure of merit for this work we focus on its definition and on the difference between the central-data- $\chi^2$  and the replica- $\chi^2$ . Moreover, in the last part of the chapter, while presenting the cross-validation strategy adopted by the NNPDF research group, we define the role of the training and validation  $\chi^2$ .

Chapter 4 contains the description of the data analysis conducted. For the analysis two sets of replicas were used: the main 1000-replicas set and a 150-replicas subset of the first one. Indeed working directly on the 1000 replicas set resulted computationally inefficient, for this reason, the 150-replicas subset was used for testing purposes. We developed the analysis with the following scheme: first of all, we plotted a distribution of the central-data- $\chi^2$ , then a fitting evaluation

---

was run in order to control the possible over-fitting of the network and the possible correlation between the "fitting-level" of the NN and the central-data- $\chi^2$  for each replica.

From this evaluation emerged the absence of correlation between the central-data- $\chi^2$  and the learning of the NN. For this reason, we proceeded to plot a number of replicas with the highest and lowest central-data- $\chi^2$  in order to verify the presence of any particular feature. Then to highlight the main differences in the PDF replicas properties with respect to the central-data- $\chi^2$  we plotted the arclength distribution, the PDFs derivatives, and the kinetic energy (for the highest and lowest central-data- $\chi^2$ ).

The conclusion summarizes the main results of the analysis, in particular the evidence of a systematic change in the PDF functional form with respect to the central-data- $\chi^2$ .

# Contents

<b>1</b>	<b>Parton Distribution Functions</b>	<b>7</b>
1.1	Quantum Chromodynamics . . . . .	7
1.2	Definition of a PDF . . . . .	8
1.3	Parameterization of a PDF . . . . .	8
1.4	Factorization . . . . .	9
1.5	Sum rules . . . . .	10
<b>2</b>	<b>Machine learning</b>	<b>11</b>
2.1	Classification . . . . .	11
2.2	Neural Networks . . . . .	11
2.3	Multi-layered Neural Networks . . . . .	12
2.4	Learning and generalization . . . . .	13
<b>3</b>	<b>NNPDF strategy</b>	<b>15</b>
3.1	Fitting strategy . . . . .	15
3.2	MC approach and MC replicas . . . . .	16
3.3	Fitting methodology . . . . .	16
3.3.1	Fit constraints . . . . .	17
3.4	Neural Network architecture . . . . .	18
3.5	$\chi^2$ Figure of Merit . . . . .	19
3.6	Cross Validation . . . . .	20
<b>4</b>	<b>Results</b>	<b>23</b>
4.1	Analysis overview . . . . .	23
4.2	$\chi^2$ distribution . . . . .	24
4.3	Fitting evaluation . . . . .	25
4.4	PDF replicas plot . . . . .	26
4.5	Arc-length . . . . .	29
4.6	Derivative . . . . .	29
4.7	Kinetic energy . . . . .	34
<b>5</b>	<b>Conclusions</b>	<b>37</b>
	<b>Bibliography</b>	<b>46</b>



# Chapter 1

## Parton Distribution Functions

The Standard Model is the current theory that describes the fundamental particles and their interactions, in particular, it describes three of the four fundamental forces: electromagnetic, weak, and strong interactions, excluding gravity.

The branch of physics that studies fundamental particles and their interaction is high energy physics (HEP). Scattering experiments are an important tool in high-energy physics, they are used not only to study the internal structure of particle systems, such as atomic nuclei or protons but also to study the interactions between different types of particles.

The main reason to study Parton Distribution Functions or PDFs is that they play a fundamental role in the study of the strong interaction, especially regarding theoretical predictions in a wide range of scattering experiments.

Moreover, considering possible future experiments, a precise determination of the PDFs, would help physicists in making predictions that could open scenarios of new physics beyond the Standard Model (for reference see [14] and for example, [4]).

### 1.1 Quantum Chromodynamics

At first, the quark model was proposed as a mathematical tool to organize the different properties, such as symmetry, of the increasing number of particles that were observed at the time. Then, the first pieces of evidence of fundamental constituents of hadrons came from deep inelastic scattering experiments (DIS) only later, during the first years of the 1970s.

This evidence came together with the indication of an internal degree of freedom for quarks, the **color**, and the presence of a particle mediating the interaction between quarks that is sensitive to this color charge, the gluon, the gauge boson of the strong force. There are three possible values for this color charge, arbitrarily labeled blue, green, and red.

Quantum Chromodynamics (QCD) is indeed the theory of the strong interaction between quarks, in particular, describes how quarks and gluons bind together to form hadrons. Nowadays QCD is considered by scientists one of the cornerstones of the Standard Model [10].

**Asymptotic freedom and color confinement** As mentioned before, the first step toward the systematical studies on strong interaction, was the interpretation of quarks as physical entities, this could indeed explain the data obtained from DIS. This interpretation opened the way to the study of the internal structure of hadrons which, as opposed to elementary particles such as leptons, are bound states composed of quarks and gluons, collectively known as partons.

Today the most important sources of data in high-energy physics are collision experiments because particle colliders provide the highest available center of mass energies,  $s$ . Having the highest possible value of  $s$  is of key importance for probing the structure of matter at the shortest distances available.

This last fact is explained by QCD with the concept of **asymptotic freedom**, namely the tendency of quarks to be free at small distances (or equivalently, high energy scale) when the force between them becomes weaker. On the opposite side of the energy scale, at great distances, quarks are seen as bound together for **confinement**: since the force between two color charges remains constant with the distance, when two color charges are separated the energy grows until a quark-antiquark pair is spontaneously produced turning the initial hadron state into a pair-of-hadrons state instead of an isolated color charge [3].

## 1.2 Definition of a PDF

Nowadays we know that inside nucleons and hadrons there are quarks and gluons, although we do not know their precise form because we can not resolve QCD with a non-perturbative approach. Indeed the internal structure of nucleons is not well defined for the confinement, because their components can not be free. For this reason, the internal structure of nucleons is described by PDFs, Parton Distribution Functions, which encode information about the structure of strong-interacting particles and give a parameterization of the nucleon's internal components[9].

For example, at the leading order of perturbative QCD, a PDF  $f_i(x, Q^2)$  represents the probability density of finding the parton flavor  $i$  carrying a fraction  $x$  of the nucleon momentum at scale  $Q^2$ , where  $Q$  is the energy scale of the interaction.

With this definition, it is clear that a complete PDF set includes 13 functions (6 quarks, 6 anti-quarks, and a gluon) defined for  $x$  between 0 and 1. A PDF set is usually determined through the comparison between the PDF-dependent prediction for a physical process and its actual experimental value.

Fig. 1.1 shows a PDF set for the proton including some parton flavors at two different energy scales.

## 1.3 Parameterization of a PDF

At a reference scale,  $Q_0^2$  the standard parametrization that is used for a PDF reads

$$f_i(x, Q_0^2) = x^{\alpha_i} (1-x)^{\beta_i} g_i(x), \quad (1.1)$$

where  $g_i(x)$  tends to a constant for both  $x \rightarrow 0$  and  $x \rightarrow 1$ . Usually  $g_i(x)$  is a polynomial function or the exponential of a polynomial in  $x$  or  $\sqrt{x}$ . This choice is guided by the expectation that PDFs behave as a power of  $x$  for  $x \rightarrow 0$ , due to Regge theory, and as a power of  $(1-x)$  as  $x \rightarrow 1$  because of quarks counting rules.

As will be further explained in the following, the aim of the NNPDF group is to infer the shape of parton distribution functions *without* assumptions on their functional forms.

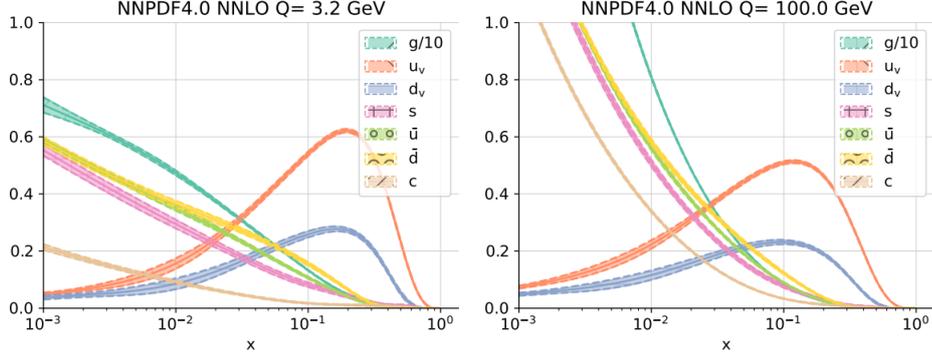


Figure 1.1: Examples of parton distribution functions obtained from NNPDF4.0

It is important to note that, to define a PDF is necessary to choose a reference scale in fact, the PDF form depends on the scale of the process. This dependence can be evaluated with DGLAP (Dokshitzer–Gribov–Lipatov–Altarelli–Parisi) evolution equations: a set of perturbative QCD equations which describe the variation of parton distribution function with the energy scale [15].

## 1.4 Factorization

PDFs have an important role in the study of "hard" processes, which are high-energy processes that admit a perturbative description. Indeed, the fundamental property of QCD which enables the perturbative computation of cross sections for processes with hadrons in the initial state is the **factorization**[11]. Factorization allows separating the cross-section of a process into two parts: a process-dependent partonic cross-section, and a set of universal parton distribution functions that characterize the hadronic bound states.

For example, if we consider a process with two hadrons in the initial states, the cross-section for this process can be written as follow:

$$\sigma = \sum_{a,b} \hat{\sigma}_{a,b} \otimes f_{a/h_1} \otimes f_{b/h_2}, \quad (1.2)$$

where  $\sigma$  is the experimentally measured cross section,  $\hat{\sigma}_{a,b}$  is the partonic cross section, computed with quarks and gluons in the initial state,  $f_{a/h_i}(x_1, Q^2)$  is the distribution of partons of type  $a$  in the  $i$ -th incoming hadron, and the convolution product of the  $a$ -th parton with the  $b_i$  hadron PDFs represents an integration over the relevant initial-state kinematic variables.

If we consider instead a process with a single hadron in the initial state the cross-section can be written as follow:

$$F_i(x, Q^2) = C_{i,a} \otimes f_a(x, Q^2), \quad (1.3)$$

where  $F_i(x, Q^2)$  is the deep-inelastic structure-function,  $C_{i,a}$  is the structure-function computed with an incoming parton,  $x := \frac{Q^2}{2p \cdot q}$  is the standard Bjorken variable, and  $f_a$  is the PDF for the proton  $a$ . By making explicit the integral computation, we can rewrite the equation as

$$F_i(x, Q^2) = x \sum_a \int_x^1 \frac{dz}{z} C_{i,a} \left( \frac{x}{z}, \alpha_S(Q^2) \right) f_a(z, Q^2), \quad (1.4)$$

## 1.5 Sum rules

Typical conservation laws put some constraints on the distribution functions, for example, the total momentum-energy conservation and the conservation of baryon number, with the PDF formalism can, respectively, be expressed as follow:

$$\int_0^1 dx x \left[ \sum_{i=1}^{n_f} (q_i(x, Q^2) + \bar{q}_i(x, Q^2)) + g(x, Q^2) \right] = 1, \quad (1.5)$$

$$\int_0^1 dx (q_i(x, Q^2) - \bar{q}_i(x, Q^2)) = n_i, \quad (1.6)$$

In particular, the baryon number conservation laws for the proton read as follows:

$$\int_0^1 dx (u(x, Q^2) - \bar{u}(x, Q^2)) = 2, \quad (1.7)$$

$$\int_0^1 dx (d(x, Q^2) - \bar{d}(x, Q^2)) = 1, \quad (1.8)$$

$$\int_0^1 dx (s(x, Q^2) - \bar{s}(x, Q^2)) = 0, \quad (1.9)$$

## Chapter 2

# Machine learning

Machine learning (ML) is a field of computer science that has seen great expansion in the last decades, mainly thanks to the increase in the computational power of modern computers.

### 2.1 Classification

ML algorithms can be distinguished on the basis of the type of training they undergo. In particular, three main classes of ML algorithms can be identified.

**Supervised learning** Supervised learning concerns learning from *labeled data*, this kind of algorithm is mainly used to handle regression and classification problems. In the NNPDF code supervised learning is used to determine parton distribution functions [5] and fragmentation functions [1]

**Unsupervised learning** Unsupervised learning is used to find patterns and structure in unlabeled data, these kinds of algorithms are particularly useful for the analysis of big data sets, especially when looking for clusters.

**Reinforcement learning** Differently from the previous cases, in a reinforcement learning algorithm, the algorithm, often called (*agent*), learns by interacting with the environment. A reinforcement learning algorithm works on a reward system: the agent learns by making, step by step, decisions that maximize its reward and minimize its penalty.

### 2.2 Neural Networks

With the term Artificial Neural Networks (ANNs) or simply Neural Networks (NNs) we define a large class of computing systems that mimic the biological neural networks constituting animal brains. A NN is constituted by a collection of fundamental units, *nodes*, called *artificial neurons*. These nodes are connected through *edges*. Edges carry a *weight* that adjusts as learning proceeds.

The great power of NNs is determined by two key properties, that allow the developer to modulate the learning parameters of the network to make them efficient for the specific problem they are applied to. Firstly there is the possibility to aggregate the neurons in *layers*, each of which can perform a different transformation

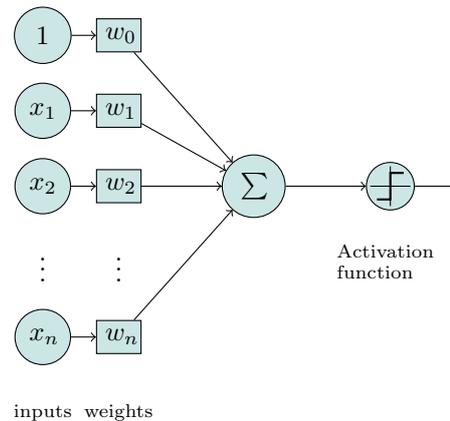


Figure 2.1: Example of single layer perceptron, from left: the different inputs, to each of them, is attached a weight ( $w_i$ ), the perceptron  $\Sigma$  is considered active if the weighted sum is greater than a certain threshold value.

on its inputs, then there's the possibility to define a *threshold of activation* for the neuron, under this threshold the neuron is considered inactive; indeed the function that defines the threshold of activation is called *activation function*.

**Single-layer perceptron** The simplest example of NN is a *single-layer* perceptron, even if today it is more common to use different models of artificial neurons. Fig. 2.1 shows an example of this simplest neural network, several inputs are fed to the perceptron through the edges, and each edge carries a weight, which determines the importance of the respective input for the node  $\Sigma$ . In a perceptron, the output is a simple binary output, 0 or 1, and it is determined by whether the weighted sum of the inputs is greater or not than a certain threshold value.

## 2.3 Multi-layered Neural Networks

Today's artificial neurons have a structure similar to the perceptron's, the main difference is the fact that the output of the node is not strictly binary. The most common neuron is the sigmoid neuron, where the output function is much smoother than the step function. In the sigmoid neuron, a small change in the input only causes a small change in the output as opposed to the stepped output of the perceptron.

As mentioned before, neurons in a NN can form layers. If the NN is constituted by more than one layer it's called a *multi-layered neural network*. Every multi-layered NN has an *input layer* and an *output layer*, respectively the first and last layer, these layers can be separated by a series of *hidden layers*.

An example of multi-layer neural networks are *deep neural networks* (DNNs), constituted by multiple layers of neurons that progressively extract information from raw data.

In multi-layer NNs and in DNNs hidden layers can be connected in different ways, the simpler NNs are the Feed-forward Neural Networks (FNNs), which are neural networks where the connections between different nodes *do not* form a cycle [16].

However, the greater the number of layers, the greater the number of *optimization parameters*, and the best is the approximation we get from the NN. Indeed the so-called *universal approximation theorem* states that any continuous function  $f : [0, 1]^n \rightarrow [0, 1]$  can be approximated arbitrarily well by a multi-layer perceptron with at least 1 hidden layer and a finite number of hidden units. Moreover, in 2017 proved a generalization of this theorem [13] which states that: width  $n + 4$  networks with ReLU activation functions can approximate any Lebesgue integrable function on  $n$ -dimensional input space if the depth of the network is allowed to grow.

## 2.4 Learning and generalization

When working with machine learning algorithms, the first step usually consists of a *training phase*. During this phase, the algorithm "learns" from data, for example, recognizes patterns or fits the input data. The training phase usually requires the minimization of a given *cost function*, which allows for an evaluation of the NN performance.

The aim of the training is to develop the algorithm's ability to generalize successfully its model, i.e. make predictions on data. Specifically, the term 'generalization' refers to the model's capability to adapt and react properly to previously unseen, new data, which has been drawn from the same distribution as the one used to build the model. Indeed, after the training, the NN undergoes the *validation phase* in which the network performance is measured by evaluating the same cost function.

**Correct generalization problem** Ideally, after the training the ML algorithm should be able to make correct predictions on data, avoiding either under- or over-learning. While under-learning happens when the model is not trained with adequate data, and subsequently the algorithm fails to make accurate predictions even with the training data, in case of over-learning the model will not be able to generalize, meaning that would make inadequate predictions on new data. For this reason, the validation phase of an ML algorithm is particularly important as shows how good or how bad the predictions of the NN are.

The learning evaluation is the domain of the statistical learning theory (for reference see [12]), but a few basic ideas can be summarized as follows. Firstly, it is clear that minimizing the training, or *in-sample* error is not enough because the validation or *out-of-sample* error can still be large, meaning that the NN cannot make a good prediction on the model underlying the data. Indeed, a large difference between the training and validation error indicates an *over-fit* of the data; the NN's parameters adapted too much to the training dataset causing a loss of generality for the model.

This fact is related to a second observation, which is that the validation error is a function of the *model complexity*, often related to the number of parameters used by the NN to approximate the result. The validation error is generally minimized for models with intermediate complexity, in fact, many-parameters models can easily produce over-learning, while NN models with not enough parameters can never reach an optimal learning level.



## Chapter 3

# NNPDF strategy

Machine Learning techniques have different applications in High Energy Physics, here specifically are used for the parameterization and optimization of PDFs [7]. In particular, while in other fitting approaches, the PDF shape is parameterized by relatively simple functional forms (see sect.1.2), inspired by QCD models, here NNs are used as unbiased interpolants, in order to avoid any theoretical bias. In this chapter, we present the NNPDF fitting strategy, with a focus on some specific features of the NNPDF fitting algorithm.

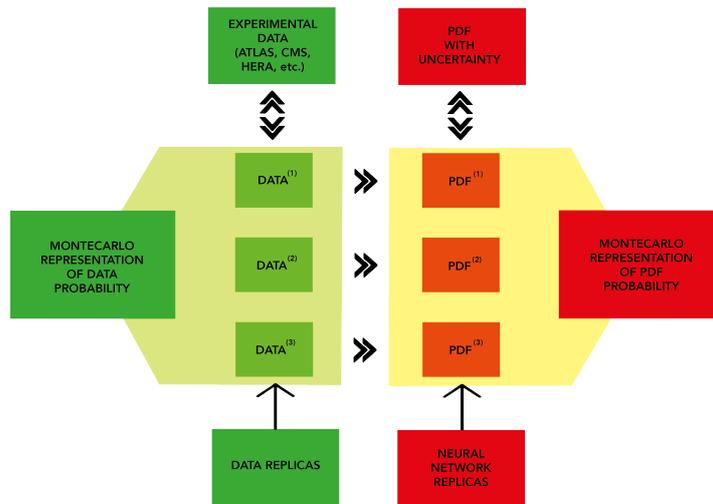


Figure 3.1: Schematic representation of the NNPDF general strategy

### 3.1 Fitting strategy

PDFs are not computable from first principles because that would involve the knowledge of the proton wave function. So, in order to determine a PDF set, a fitting procedure must be adopted which consists in comparing factorized expressions, such as Eq. 1.4, with experimental data while minimizing a specific cost function used to measure the goodness of the fit.

In this context, a determination of PDFs with uncertainties involves determining a probability distribution in a space of several independent functions. A

possible method assumes a specific functional form for parton distributions, projecting the infinite-dimensional problem onto a finite-dimensional parameter space so that a representation of PDFs is possible in terms of a finite number of parameters and there is an optimal parametrization.

Two different methodologies are currently used: the Hessian approach and the Monte Carlo approach, while in the first case the parametrization is standard and inspired by QCD arguments (see for example Eq. 1.1 in section PDF parameterization); in a Monte Carlo approach, the best fit is determined from Monte Carlo sample by an average of the data, and uncertainties are obtained as variances of the sample. In the following, we will present the Monte Carlo approach.

## 3.2 MC approach and MC replicas

In this Monte Carlo (MC) approach, in order to use a statistical framework, a number  $N_{dat}$  of experimental data (**data-point**) are used to build an ensemble of  $N_{rep}$  artificial Monte Carlo pseudo-data, hereinafter **replicas**

$$F_i^{(art)(k)} = (1 + r_N^{(k)} \sigma_N) \left( F_i^{(exp)} + \sum_{p=1}^{N_{sys}} r_p^{(k)} \sigma_{i,p} + r_i^{(k)} \sigma_{i,s} \right), \quad (3.1)$$

with  $k = 1, \dots, N_{rep}$ ,  $i = 1, \dots, N_{dat}$ ; where  $F_i$  represents one single data point,  $\sigma_N$  is the total normalization uncertainty,  $\sigma_{i,p}$  are the  $N_{sys}$  correlated systematic errors,  $\sigma_{i,c} = \sum_{p=1}^{N_{sys}} \sigma_{i,p}$  is the sum of all the correlated systematic errors, and  $r^{(k)}$  are independent univariate Gaussian random numbers.

We see, from Eq. 3.1, that replicas are randomly generated in accordance with a multi-Gaussian distribution, centered around each-data point, with variance given by the experimental uncertainty. Each replica contains the same number of data points as the original experimental measurement. So, with enough replicas, the set contains complete experimental information and the experimental value can be retrieved by taking the mean and the experimental variance (calculated over different replicas).

The experimental (or central) data used in the NNPDF strategy, come from a number of experiments, including fixed target, colliders, deep inelastic scattering, and Drell-Yan experiments. The experimental data are in form of observables such as cross-sections and structure formations and are provided in different formats from different research groups. In order to produce replicas and fit them, the data are converted to an NNPDF common-format (*commondata*) by the `buildmaster` code [2].

## 3.3 Fitting methodology

The next step consists in the interpolation between data points with neural networks, which means training  $N_{rep}$  neural networks, each of them based on the data in one single replica. In the end, we have  $N_{rep}$  PDFs and from these values, we can determine the mean value of the parton distribution for each  $x$  as the average over all the replicas, while the uncertainty is the variance of the values. Examples of PDF replicas and final PDF plots can be found in the Appendix, Fig. 5.1 - Fig. 5.4.

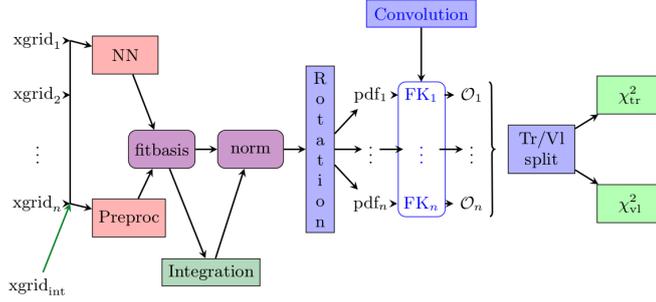


Figure 3.2: Schematic of the NNPDF algorithm workflow

### 3.3.1 Fit constraints

In the NNPDF approach, the parton distribution functions (or the fragmentation functions) are parameterized at a low scale, around the boundary between the perturbative and non-perturbative regimes of QCD, namely  $Q_0 \simeq 1$  GeV (the proton mass). QCD provides little guidance about the behavior of PDFs at the input parameterization scale  $Q_0$ , such as integrability conditions and the momentum and valence sum rules, and does not provide any further information on their  $x$  dependence at low scales. The NNPDF4.0 algorithm implements in the code these restraints together with a positivity constraint.

**Positivity constraints** As PDFs below leading order do not represent probabilities, they may be negative. Nevertheless, since was demonstrated (see [6]) that PDFs for individual quark flavors and for the gluon in the minimal subtraction scheme are non-negative, a positivity condition is imposed along with the positivity of physical cross-sections. Note that the positivity of PDFs in the minimal subtraction scheme is neither necessary nor sufficient in order to ensure cross-section positivity: they are independent (though of course related) constraints that limit the space of acceptable PDFs.

**Integrability constraints** At small- $x$  the PDFs behavior is constrained by integrability requirements. The gluon and singlet PDFs must satisfy the momentum sum rule, Eq. 1.5, which implies

$$\lim_{x \rightarrow 0} x^2 f_k(x, Q) = 0, \quad \forall Q, \quad f_k = g, \Sigma, \quad (3.2)$$

Similarly the valence sum rule, Eq. 1.6, force the small- $x$  behaviour of the valence distributions:

$$\lim_{x \rightarrow 0} x f_k(x, Q) = 0, \quad \forall Q, \quad f_k = V, V_3, V_8 \quad (3.3)$$

Moreover, standard Regge theory arguments indirectly imply that

$$\lim_{x \rightarrow 0} x f_k(x, Q), \quad \forall Q, \quad f_k = T_3, T_8; \quad (3.4)$$

where  $\Sigma$  is the singlet quark,  $g$  the gluon,  $V_i$  the valence and  $T_i$  the non-singlet sea combinations (listed in the following).

$$\Sigma(x, Q_0^2) = (u + \bar{u} + d + \bar{d} + s + \bar{s} + 2c)(x, Q_0^2), \quad (3.5)$$

$$T_3(x, Q_0^2) = (u + \bar{u} - d - \bar{d})(x, Q_0^2), \quad (3.6)$$

$$T_8(x, Q_0^2) = (u + \bar{u} + d + \bar{d} - 2s - 2\bar{s})(x, Q_0^2), \quad (3.7)$$

$$V(x, Q_0^2) = (u - \bar{u} + d - \bar{d} + s - \bar{s})(x, Q_0^2), \quad (3.8)$$

$$V_3(x, Q_0^2) = (u - \bar{u} - d + \bar{d})(x, Q_0^2), \quad (3.9)$$

$$V_8(x, Q_0^2) = (u - \bar{u} + d + \bar{d} - 2s + 2\bar{s})(x, Q_0^2), \quad (3.10)$$

These constraints that are imposed, by means of the Lagrange multipliers, affect the cost function. In particular, for each constraint that's imposed, an additional contribution is added to the total cost function. This fact is important for considerations that will follow as a higher or lower value of the cost function makes respectively, increase or decrease the probability of having a specific output.

### 3.4 Neural Network architecture

The NNPDF fitting code supports multi-layered, feed-forward NNs, called sequential dense NNs in the ML code framework (for more details on the NNPDF code see [2]).

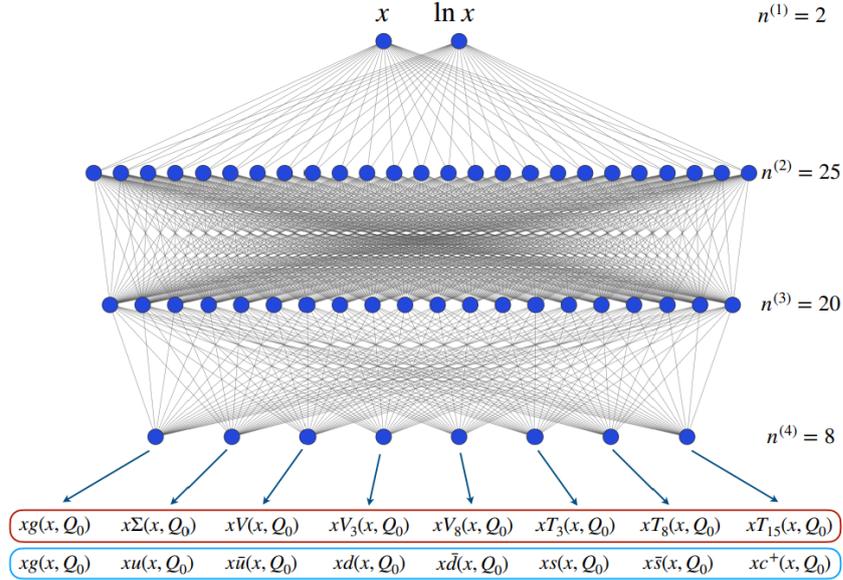


Figure 3.3: Example NNPDF neural network architecture, this architecture corresponds to the optimal choice in the evolution basis as results from hyper-optimization studies.

Image 3.3 shows an example of NNPDF neural network architecture. The neural network has two inputs ( $x$  and  $\ln x$ ) and eight output neurons, one for each parton flavor. In particular, image 3.3 shows a 2-25-20-8 NN, this particular architecture corresponds to the optimal choice of *hyper-parameter* in the evolution basis. Here the term hyper-parameter refers to the parameters of the NN architecture that can be decided by the developer; the number of layers, of neurons per layer, the activation function associated to each neuron, and the optimization

function are examples of hyper-parameters. The NNPDF4.0 code implements an automatic hyper-parameters study, the hyper-optimization which aims to find the best choice for the architecture's parameters of the NN.

As in other NNs, the neurons of this neural network are characterized by a series of parameters. In particular,  $\xi_i^{(l)}$  denotes the activation state of each neuron, where  $l$  represents the layer and  $i$  is the neuron index within the layer. These values of activation,  $\xi_i^{(l)}$ , are determined in terms of the previous layer,  $(l-1)$ .  $\omega_{ij}^{(l)}$  denotes the weight associated to the edge connecting the  $i$ -node of the  $l$  layer and the  $j$ -node of the  $(l-1)$  layer; and  $\theta_i^{(l)}$  the activation thresholds of each neuron.

For example, for a sigmoid neuron, its activation state can be computed explicitly as follow:

$$\xi_j^{(l)} = g(h_j^{(l)}), \quad (3.11)$$

where  $g(x)$  is the sigmoid function

$$g(x) = \frac{1}{1 + e^{-x}}, \quad (3.12)$$

and  $h_j^{(l)}$  is computed as

$$h_j^{(l)} = \sum_{i=1}^{n_{l-1}} \omega_{ji}^{(l)} \xi_i^{(l-1)} - \theta_j, \quad (3.13)$$

### 3.5 $\chi^2$ Figure of Merit

The goodness of a PDF set is measured by means of a suitable figure of merit, subsequently, the optimal fit is obtained by minimizing this figure of merit (or cost function). In the NNPDF approach, this figure of merit is the  $\chi^2$ ,

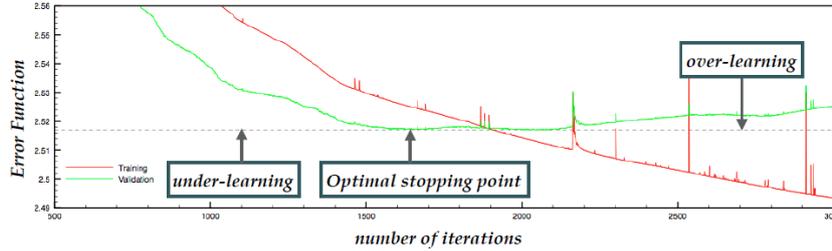
$$\chi^2 = \sum_{i,j}^{N_{dat}} (D_i - P_i) C_{ij}^{-1} (D_j - P_j), \quad (3.14)$$

where  $D_i$  is the  $i$ -th data point,  $P_i$  is the prediction of the corresponding data point, and  $C_{ij}$

$$C_{ij} = \delta_{ij} (\sigma_i^{uncorr})^2 + \sum_{k=1}^{N_{corr}} \sigma_{k,i}^{corr} \sigma_{k,j}^{corr} \quad (3.15)$$

is the experimental covariance matrix element between data-points  $i$  and  $j$ ; where every  $i$ -th data point ( $i = 1, \dots, N_{dat}$ ) is affected by uncorrelated uncertainty  $\sigma_i^{uncorr}$  and correlated systematic uncertainty  $\sigma_{k,i}^{corr}$  for  $k = 1, \dots, N_{corr}$ .

**Central-data and replica  $\chi^2$**  Depending on which data point  $D_i$  is considered when calculating the  $\chi^2$ , we can compute different quantities. For each replica, we can compute the replica- $\chi^2$  or the central-data- $\chi^2$ , in the first case the  $\chi^2$  is computed to the replicas, meaning that data point  $D_i$  used in the calculation are *replicas* of the original data. The central-data- $\chi^2$  is calculated with respect to the original experimental data, in this case,  $D_i$  indicates the experimental data point.

Figure 3.4: Training and validation  $\chi^2$  at different epochs

### 3.6 Cross Validation

As explained in section 2.2, weights and activation thresholds are free parameters of the network which adjust as learning proceeds. This means that it is necessary to optimize these parameter values in order to have a good fit. When having a large number of parameters, determining the best fit could be non-trivial, due to false minima and fluctuations. One possible way to avoid this problem is the *cross-validation* method [8]. This method consists in randomly dividing the data into two sets, training, and validation, and computing their  $\chi^2$  separately, but only the training ( $\chi_{tr}^2$ ) is minimized.

**$\chi^2$  minimization** Currently, the minimization strategy used in the NNPFD algorithm involves gradient descent algorithms, which require the evaluation of the  $\chi^2$  gradient with respect to the weights and the activation threshold of the neuron, namely  $\frac{\partial \chi^2}{\partial w_{ij}^{(l)}}$  and  $\frac{\partial \chi^2}{\partial \theta_i^{(l)}}$ . When having a large number of parameters a deterministic gradient descent may not be efficient as the algorithm would easily get stuck in local minima. To avoid this problem a stochastic gradient descent algorithm (SGD) is useful. In particular, the NNPFD algorithm has been tested with the Adam, AdaGrad, and RMS stochastic gradient descent function.

The cross-validation approach is thus important to determine the optimal length of the fit. While in the beginning, both  $\chi_{tr}^2$  and  $\chi_{val}^2$  decrease, at some point, the training error continues decreasing, while the validation error starts increasing determining a global minimum (see Fig. 3.4). This global minimum of the  $\chi_{val}^2$  is the optimal stopping point, a too-short fit would cause *under-learning*, meaning the NN is not yet able to fit data correctly. On the contrary, a longer fit would lead to *over-learning*, where the NN ends up fitting statistical fluctuations. In particular, over-learning is characterized by the increase in  $\chi_{val}^2$  as the number of iterations increases, indicating that what is being learned in the training sample is not present in the validation one (namely the fluctuations).

**Patience algorithm** The optimal stopping point can be determined with different strategies: in a *look-back strategy* the optimal stopping point is defined as the global minimum of  $\chi_{val}^2$ , computed over a large, fixed number of iterations. Instead, in a so-called *patience algorithm* the fit is stopped after  $\chi_{val}^2$  no longer improves for a defined number of iterations.

NNPDF4.0 algorithm implements a patience algorithm that, together with positivity constraints, defines when a fit is allowed to stop. Fig. 3.5 shows the

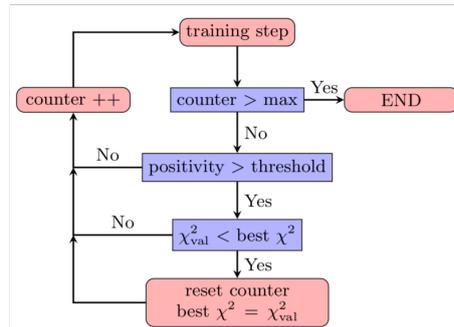


Figure 3.5: Flowchart describing the algorithm implemented in NNPDF4.0 to determine the optimal length of the fit based on the look-back cross-validation stopping method

workflow of the fit algorithm after the patience algorithm is enabled.



# Chapter 4

## Results

The aim of this work is to verify the presence of possible correlations between the statistical and physical properties of PDFs ensembles obtained with the NNPDF collaboration's code, and eventually identify the main features of the better-fitting PDFs. As the NNPDF fits produce a distribution of results, to have a better understanding of this result distribution, it's interesting to compare the results with properties of the experimental data. The central-data- $\chi^2$  is useful in this task because shows how well a PDF replica describes experimental data.

### 4.1 Analysis overview

The analysis that will be described in the following involves two different sets of replicas: the main set of 1000 replicas and a subset of 150 replicas. The 150 replicas subset, was used mainly for testing purposes as working directly on the 1000 replicas set resulted computationally inefficient. In particular, all the analyses that will be presented were, in the first moment tested on the 150 replicas subset, and then reproduced on the 1000 replicas main set. After studying the main statistical features of the two ensembles, we studied the statistical and physical properties of the highest and lowest central-data- $\chi^2$  replicas. In particular, to investigate the systematic behavior of the PDF replicas in relation to the central-data- $\chi^2$ , from both the main set and its subset the 2, 5, 10, 15, and 20 replicas with the highest and lowest central-data- $\chi^2$  were analyzed.

In order to summarize the results of these analyses, in the following we report the plots of the analyzed statistical and physical quantities

1. of the 5 replicas with the highest and lowest central-data- $\chi^2$ , plotted respectively in green and orange.
2. of all the replicas with a color scale associated to the central-data- $\chi^2$  of the replicas. In particular, in the 150-replicas plot, the color scale goes from green for the lowest-central-data- $\chi^2$ -replicas, to orange for the highest-central-data- $\chi^2$ -replicas. In the 1000-replicas this color scale wasn't appropriate as did not allow for a clear understanding of the systematic feature we wanted to enhance, indeed we used a color scale that goes from blue, for the lowest-central-data- $\chi^2$  replicas, to yellow, for the highest-central-data- $\chi^2$ -replicas.

The plots for the 5 replicas with the highest and lowest central-data- $\chi^2$  represent clearly the feature studied while the color-scale plots are useful to represent

the systematic behavior of PDFs replica with respect to the central-data- $\chi^2$ .

## 4.2 $\chi^2$ distribution

The analysis starts from the central-data- $\chi^2$  distribution plots (Fig. 4.1 and Fig. 4.2).

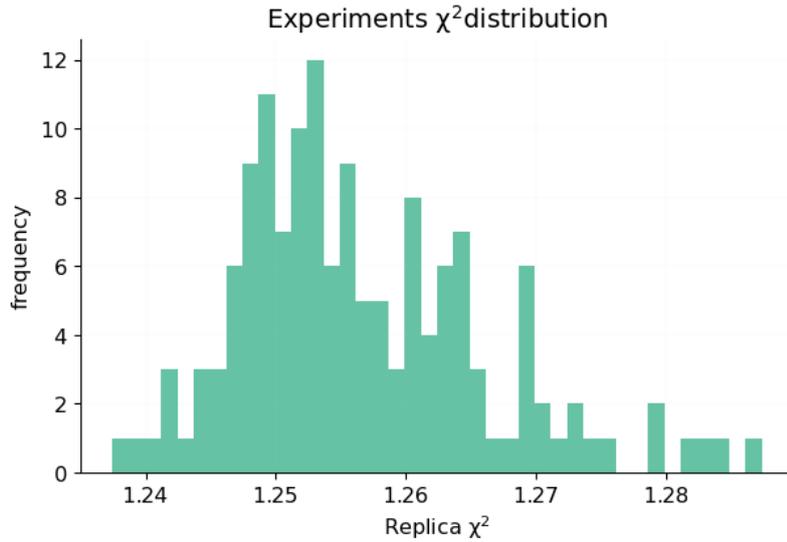


Figure 4.1: Chi2 distribution for 150 replica

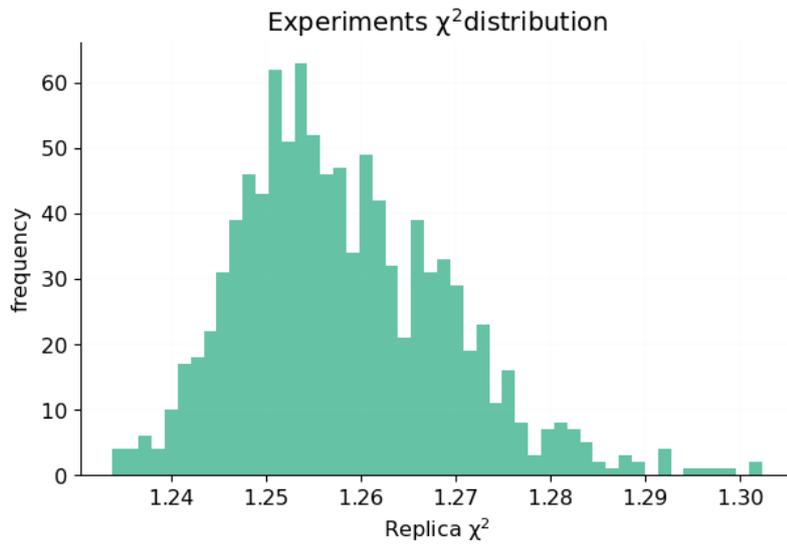


Figure 4.2: Chi2 distribution for 1000 replica

As we would expect, the two  $\chi^2$  distributions do not show particular differences. Both plots follow a standard  $\chi^2$  probability distribution and have only a

few replicas with a central-data- $\chi^2$  smaller than 1.24, and a few replicas with a central-data- $\chi^2$  value, greater than 1.28 or 1.29; meaning the two sets are statistically equivalent.

At first one could think that the central-data- $\chi^2$  value of a replica was related to the learning level of the NN algorithm while producing that replica. We tested this hypothesis by carrying out a fitting evaluation, in order to verify the possible presence of a correlation between the over-fitting of the NN associated with the replica and the low central-data- $\chi^2$  value.

### 4.3 Fitting evaluation

The learning (or fitting) level of each replica was studied by means of a scatter-plot in which, each point's  $(x, y)$  coordinates are defined as  $(\chi_{tr}^2, \chi_{val}^2)$ . In this way, the  $y = x$  line on the plane represents the case of "optimal learning" in which the NN is able to fit the valuation data set as well as the training data set. On the other hand, a point situated above this line represents a PDF replica affected by over-learning, while a point situated under this line represents the under-learning situation.

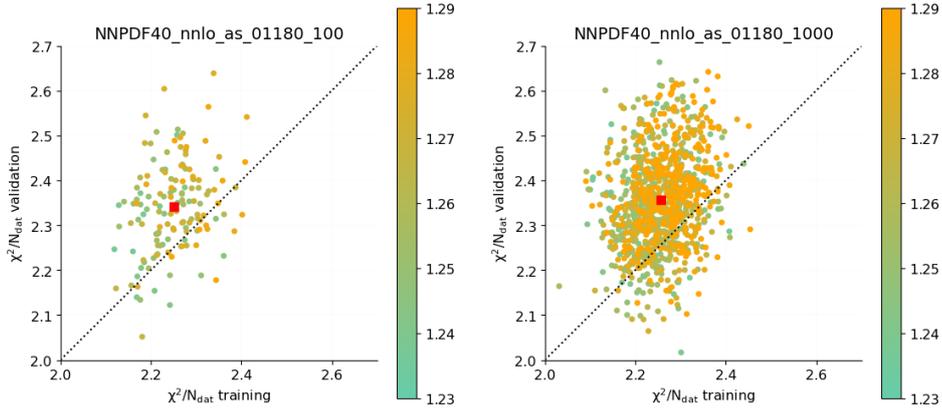


Figure 4.3: Color scale plot of the 150 replicas position in the  $\chi_{tr}^2, \chi_{val}^2$  plane. The color indicates the central- $\chi^2$  value

Figure 4.4: Color scale plot of the 1000 replicas position in the  $\chi_{tr}^2, \chi_{val}^2$  plane. The color indicates the central- $\chi^2$  value

In Fig. 4.3 and Fig. 4.4, each data point is plotted in a different color, associated with its central-data- $\chi^2$ , going from green for the lowest central-data- $\chi^2$  to orange for the highest central-data- $\chi^2$ . The red square represents a sort of "mean learning level", that quantifies the mean learning of the algorithm on its own replica data set. Its position over the bisector line indicates a small over-learning.

The scatter plots show that there is not any correlation between the central- $\chi^2$  value and the learning of the NN, if a correlation was present we would expect a majority of green points over the bisector line and a majority of orange points under the bisector line, meaning that central-data- $\chi^2$  would be related to over-learning.

Fig.4.6 and 4.5 show the same plots for the five replicas with the highest and lowest central-data- $\chi^2$  for each set of replicas. The points associated with the low central-data- $\chi^2$  replicas are plotted in green, while the ones associated with high central-data- $\chi^2$  replicas are plotted in orange. From these plots, it is even more

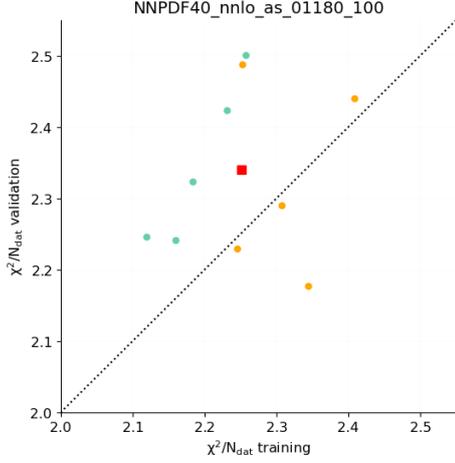


Figure 4.5: Plot of the position of the selected replicas (from 150 replicas fit) on the  $\chi_{tr}^2, \chi_{val}^2$  plane

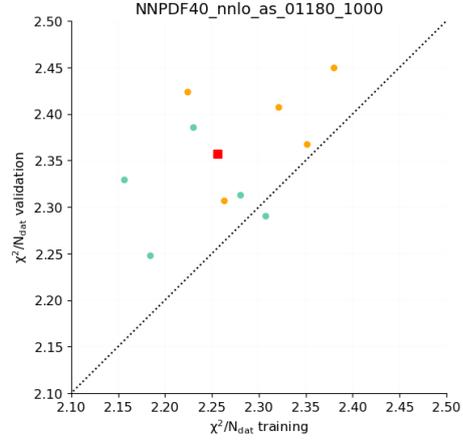


Figure 4.6: Plot of the position of the selected replicas (from 1000-replicas-fit) on the  $\chi_{tr}^2, \chi_{val}^2$  plane

evident the difference in the central-data- $\chi^2$  value, is not related to a difference in the learning on the NN, as both replicas with the highest and lowest central-data- $\chi^2$  show a small over-learning.

#### 4.4 PDF replicas plot

In order to understand whether the high or low central-data- $\chi^2$  could be related to a physical property of a PDF set, the first step is looking at the plots of the five highest and lowest central-data- $\chi^2$  replicas. Fig. 4.7 - 4.8 shows PDF plots for the 5 lowest (green) and highest (orange) central-data- $\chi^2$  replicas.

From the analyses presented so far, we know that the 1000 replicas set and its 150 subsets are statistically equivalent but we can not say anything about the physical properties of the two sets. For this reason, the 5 replicas with the highest and lowest central-data- $\chi^2$  have been selected from both replicas set.

These plots (Fig. 4.8 and Fig. 4.7) shows that the lowest central-data- $\chi^2$  replicas have a more complex functional structure. This is especially evident for the gluon (4.8 (d) and 4.7 (d)).

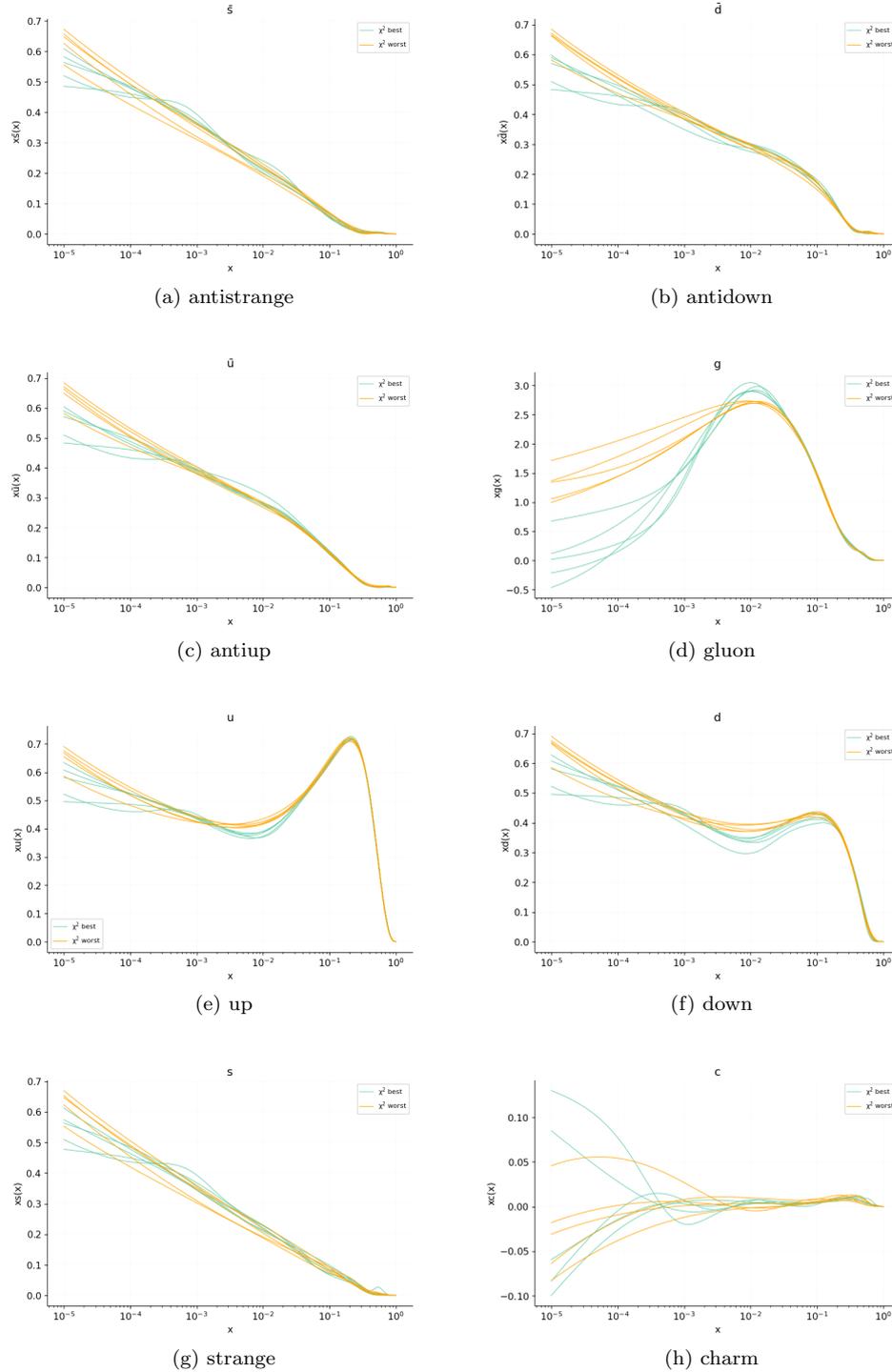


Figure 4.7: Plot of the 5 highest and lowest central-data- $\chi^2$  replicas from the 150 replicas set

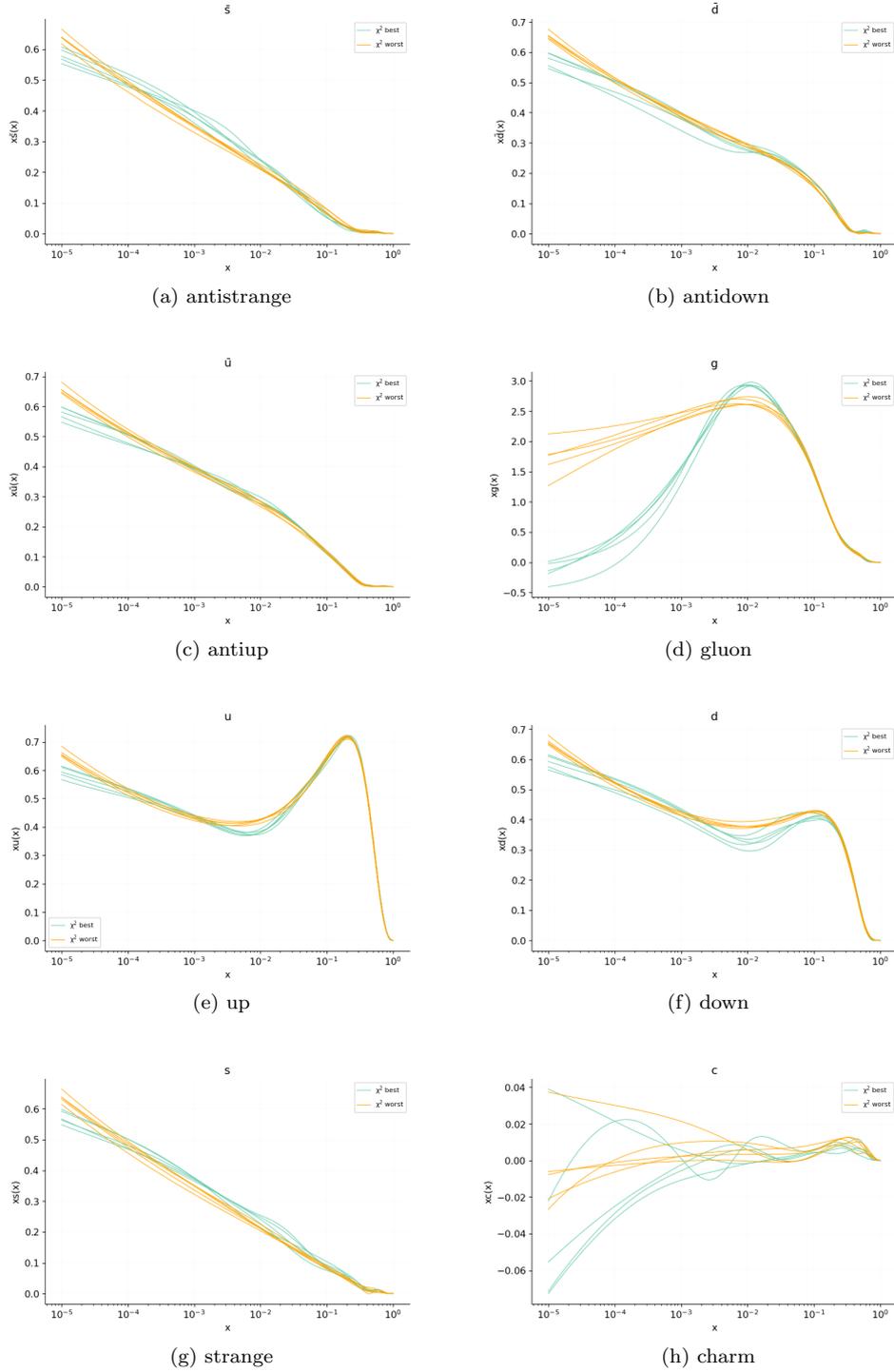


Figure 4.8: Plot of the 5 highest and lowest central-data- $\chi^2$  replicas from the 1000 replicas set

## 4.5 Arc-length

Intuitively one can imagine that the difference in the functional structure should result in different PDFs arclength,

$$\int_0^1 \sqrt{1 + \left(\frac{d}{d \log x} x f_i(x)\right)^2} dx. \quad (4.1)$$

In particular, we would expect to see an arc-length distribution centered around a higher value for the lowest central-data- $\chi^2$  replicas (more oscillating PDFs) and a distribution centered around a lower value for the high central-data- $\chi^2$  replicas (more smooth PDFs).

Figures Fig. 4.9 and Fig. 4.10, show the arc length distributions for the 5 lowest (green) and highest (orange) central-data- $\chi^2$  for both the 150-replicas subset and the 1000-replicas main set.

The same figures do not show any significant difference between distributions: for each of the eight flavors, the arc lengths for both the 5 highest and lowest central-data- $\chi^2$  replicas are distributed around the same value. For example, considering the gluon PDF (Fig. 4.9 if we consider the set of 150 replicas (d), and Fig. 4.10 (d) if we instead consider the 1000 replicas set), the arclength for both high- $\chi^2$  and low- $\chi^2$  vary around a central value of 1.28, despite the considerable difference in the PDF shape.

Anyway from this arclength analysis, and together with the PDFs plots in the previous section we can conclude that the main 1000-replicas set and the 150-replicas subset are statistically also with respect to the "physical" properties of the PDFs. For this reason, from this point on the analysis results and plots are reported only for the main replicas set.

## 4.6 Derivative

In order to have a better understanding of the trend of the curve, it is interesting to study the PDFs' derivative, especially considering that it is very similar to the integrand function in equation Eq. 4.1. For this reason, for the 5 replicas with the highest and lowest central-data- $\chi^2$ , we plotted the derivative (Fig. 4.11). As in previous plots, the derivative of the low- $\chi^2$  replicas is represented by green functions, and orange was used to plot the high- $\chi^2$  replicas derivative. In order to have more detailed information on the PDFs trend we computed the log derivative instead of the standard derivative. The use of log-scale indeed is useful to have a look at both the limits:  $x \rightarrow 0$  and  $x \rightarrow 1$ .

Looking at the derivative's plots (4.11) it is clear the reason behind the similarity of the arclength distributions: the arc-length value is dominated by the peak that all the PDFs present for  $x \rightarrow 1$ . Moreover, it is clear that replicas with the highest central-data- $\chi^2$  have a more fluctuating derivative, while the lowest central-data- $\chi^2$  replicas are related to a more smooth function and subsequently a less oscillating derivative. Figure 4.12 shows how this is a systematic behavior.

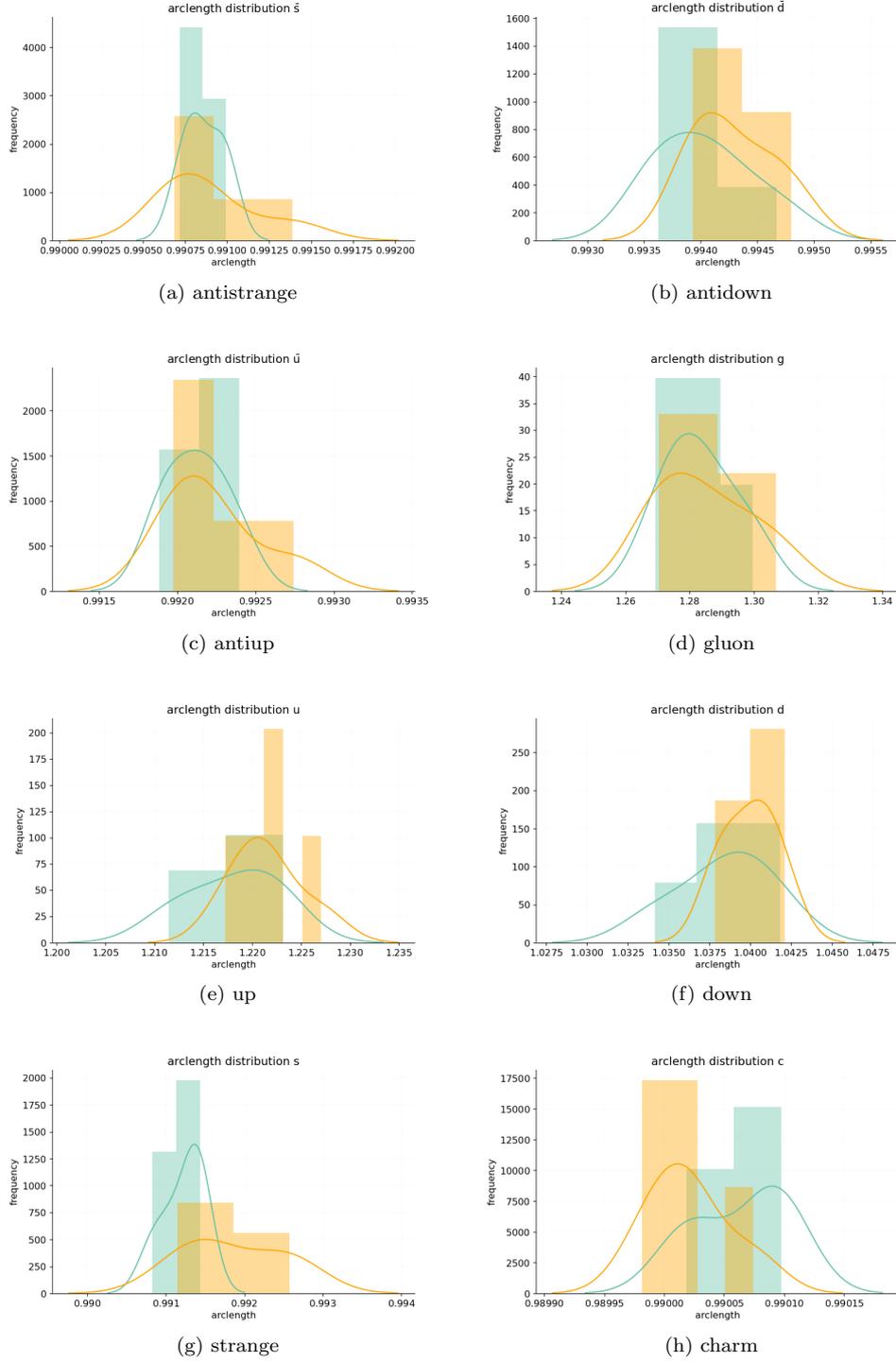


Figure 4.9: Plot of the arclength distributions of different parton flavors divided in respect with the  $\chi^2$  value - 150 replicas set

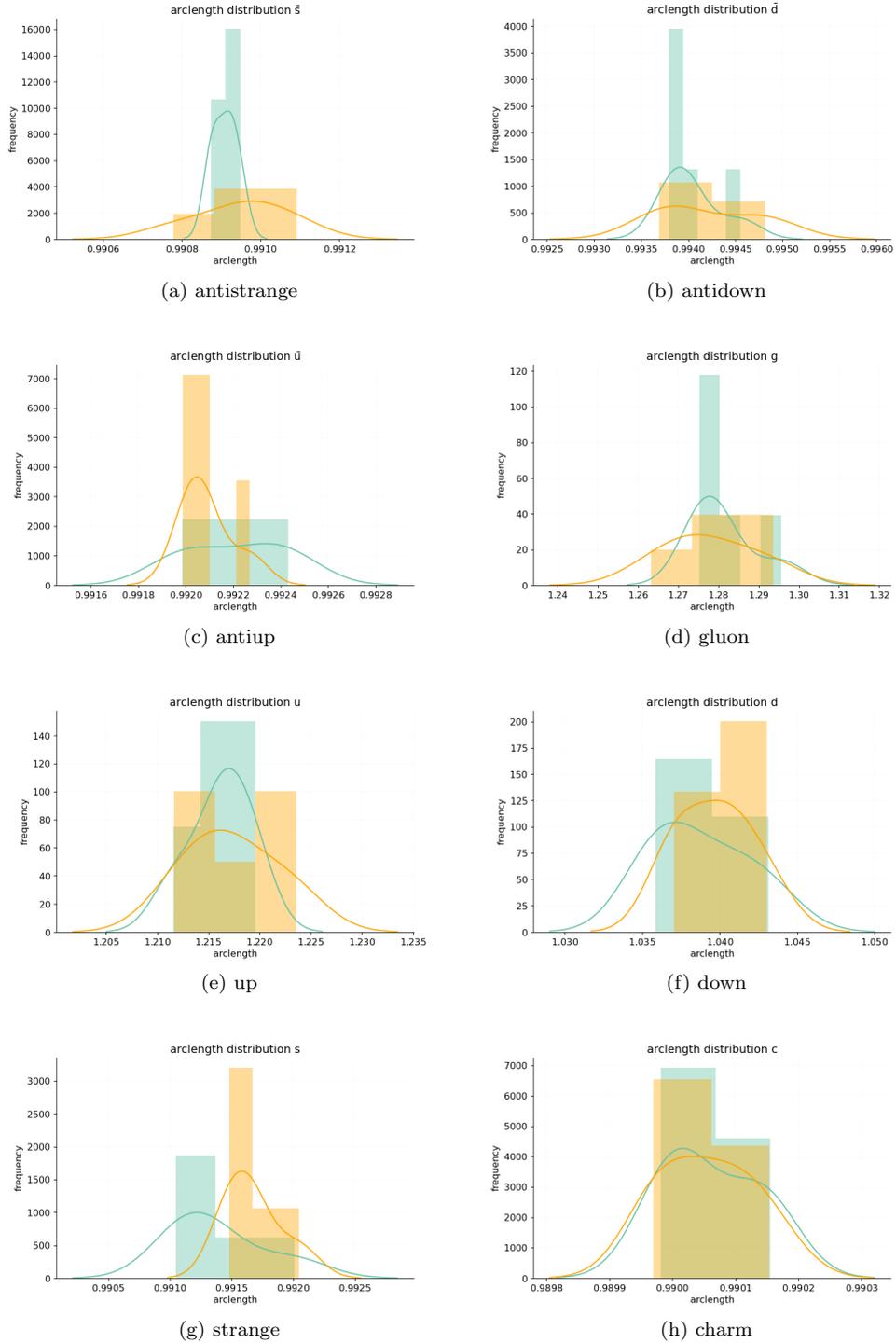


Figure 4.10: Plot of the arclength distributions of different parton flavors divided in respect with the  $\chi^2$  value - 1000 replicas set

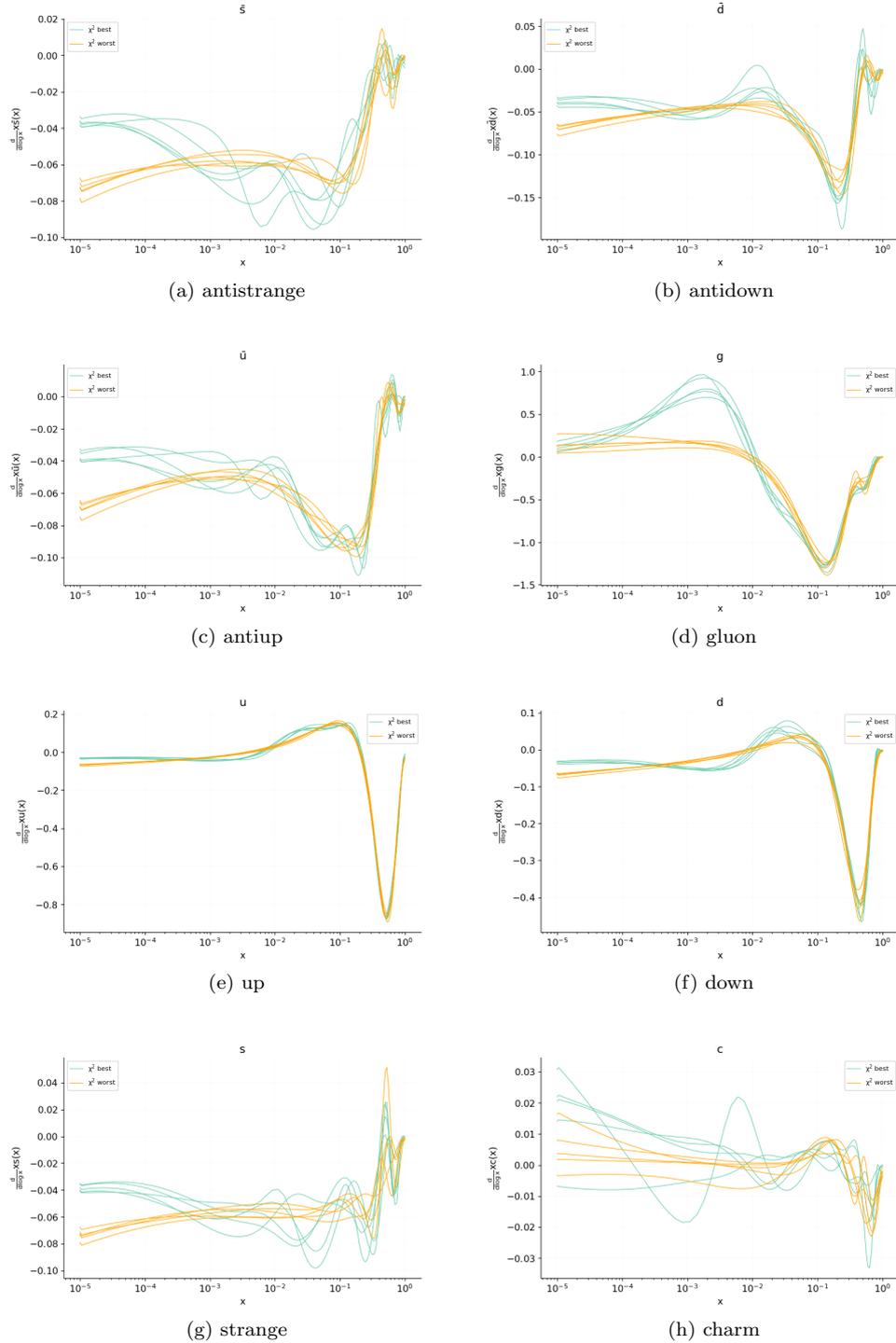


Figure 4.11: Plot of the PDFs' log-derivatives of different parton flavors separated according to the  $\chi^2$  value - 1000 replicas set

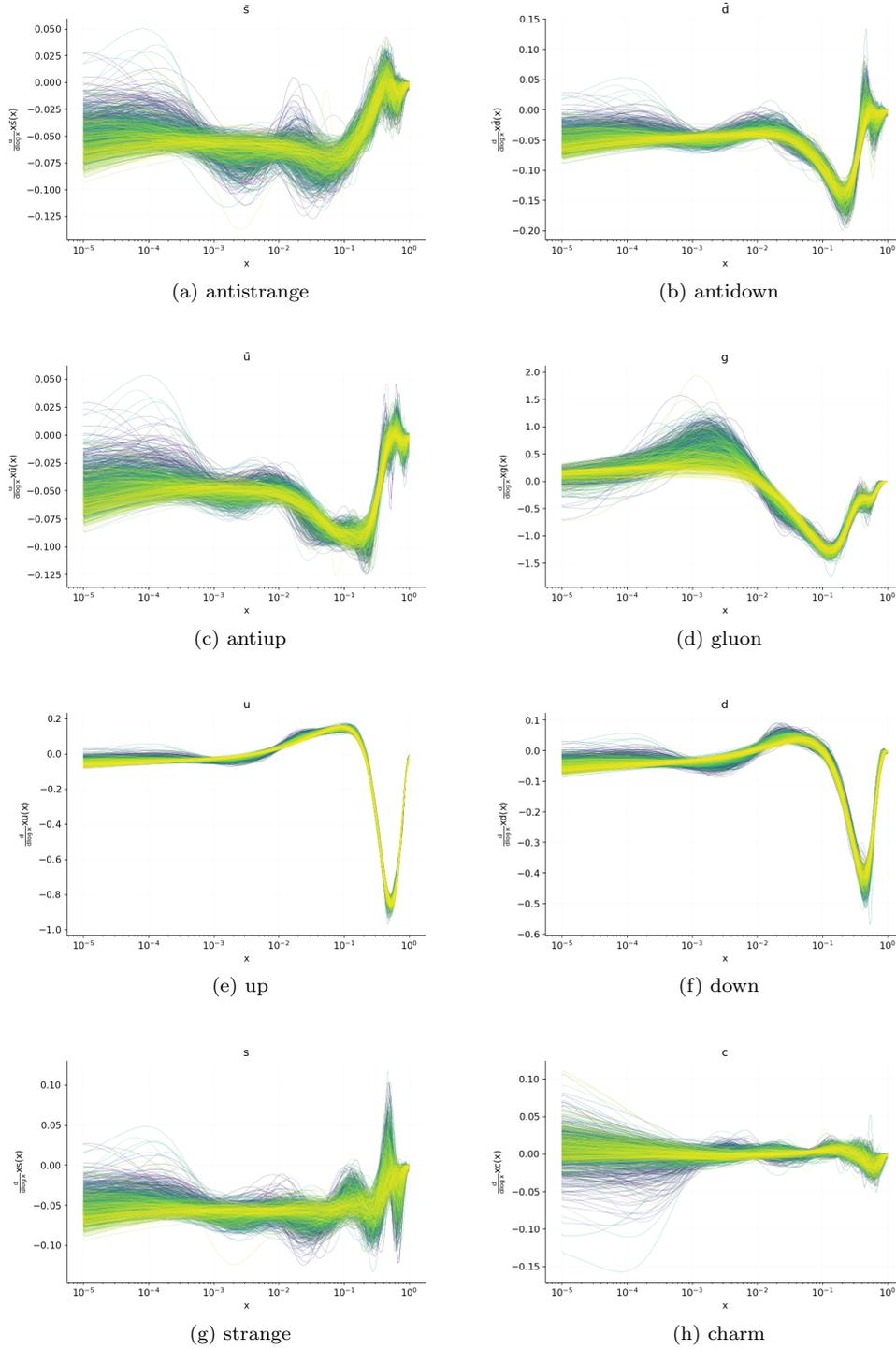


Figure 4.12: Plot of the PDFs' log-derivatives of different replicas in a color scale related to the central- $\chi^2$  value - 1000 replicas.

## 4.7 Kinetic energy

To conclude the analysis, the last quantity that we consider is kinetic energy. The term kinetic energy is used to identify the integrand in eq. 4.1 because it coincides with the Lagrangian of a quantum free particle, indeed its kinetic energy. The PDF kinetic energy is defined as

$$\sqrt{1 + \left(\frac{d}{d \log x} x f_i(x)\right)^2}. \quad (4.2)$$

Eq. 4.2 shows that the kinetic energy is ultimately the module of the derivative the kinetic energies, for this reason, the kinetic energies plots enhance the difference in the PDF replicas that were already identified in the derivative's plots. Fig 4.13 shows the kinetic energy plots for the 5 replicas with the lowest (green) and highest (orange) central-data- $\chi^2$ .

From this final analysis of the kinetic energy, it is clear that a low-central-data- $\chi^2$  can be related to a particular shape of the PDFs function for different flavors. In particular, it can be related to higher-kinetic energy and subsequently to more oscillating functions. This is especially true for the gluon PDF (4.13, (d)) where the high central-data- $\chi^2$  replicas form almost straight lines while the low central-data- $\chi^2$  replicas have a well-defined peak for  $x = 10^{-2}$ . The same plot for the complete replicas set (Fig. 4.14) confirms the observation, indeed, in the PDF gluon's plot (Fig. 4.13 (d)) it's possible to see a color gradient from blue to yellow in accordance with the results seen in (4.13, (d)).

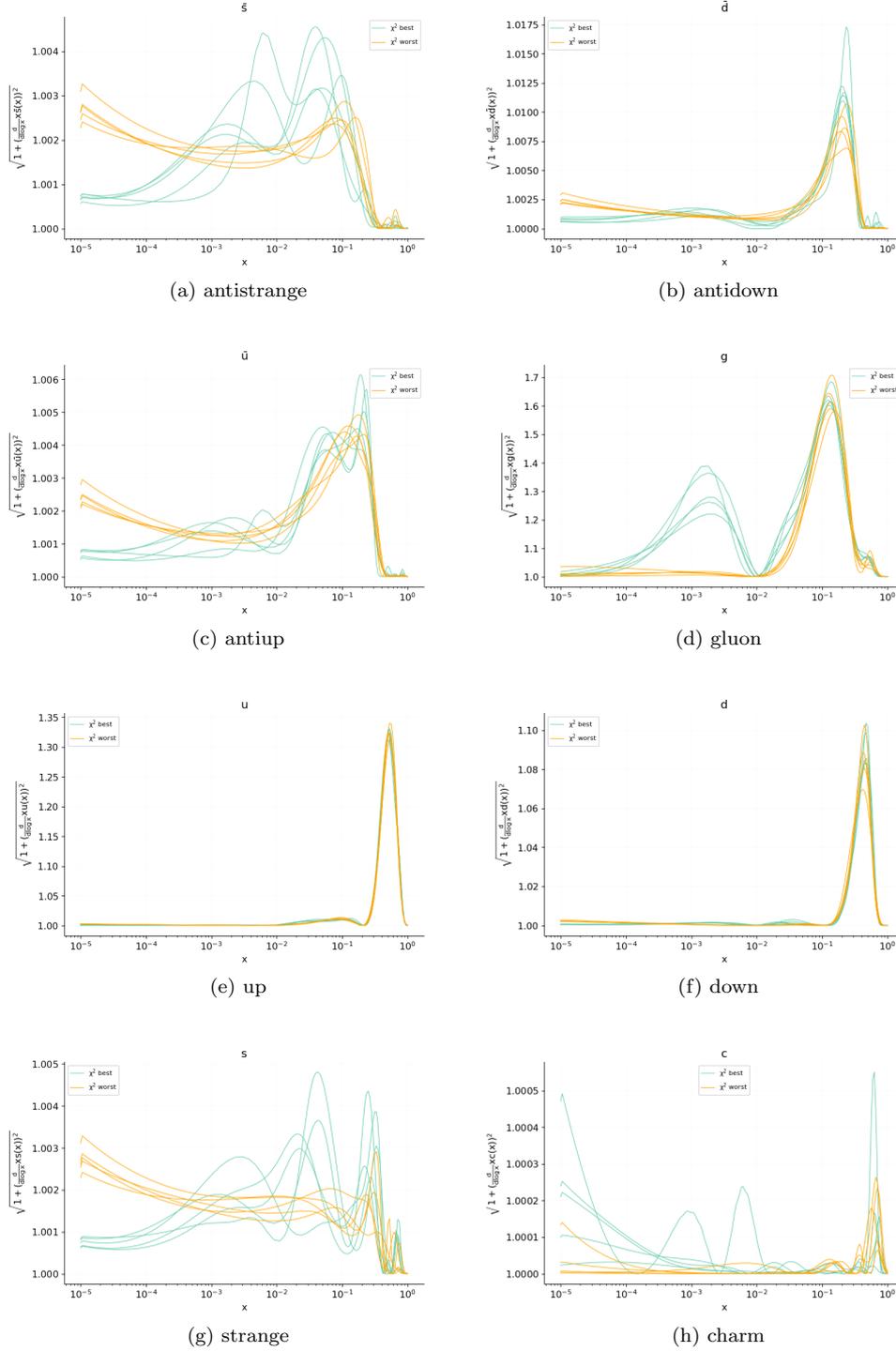


Figure 4.13: Plot of the PDFs' log-derivatives of different parton flavors separated according to the  $\chi^2$  value - 1000 replicas set

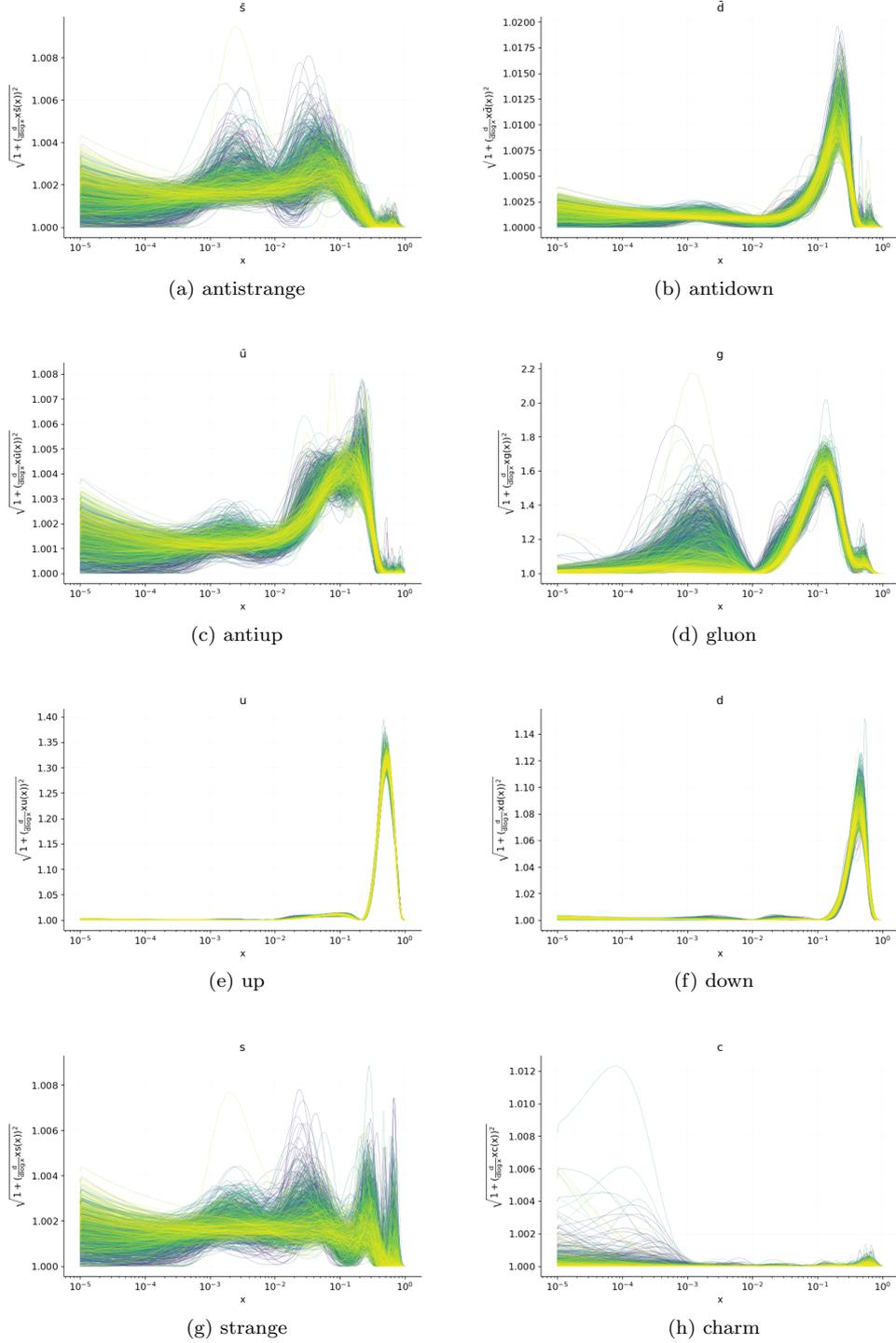


Figure 4.14: Plot of the PDFs' kinetic energies with a colors scale indicating how well they fit the central data according to the central- $\chi^2$  value - 1000 replicas set

## Chapter 5

# Conclusions

The analysis presented shows a few interesting results. First of all, we can state that a low-central-data- $\chi^2$  is related to a specific functional form for each PDF's flavor. We can also state that the reason why these replicas have a low-central-data- $\chi^2$ , is simply that they reflect, the low- $\chi^2$  value is not related to better learning of the NN.

The analysis also shows that these functional forms are outliers of the central- $\chi^2$  distribution. Only a few data replicas produce PDFs set with this feature. This can, in turn, be due to different reasons; it is possible that central data reflects a certain distribution made unlikely by the fitting methodology, for example, highly-fluctuating functions are penalized by the algorithm. This first hypothesis is supported by the lack of correlation between the central-data- $\chi^2$  and the fitting performance measured by the ratio between  $\chi_{tr}^2$  and  $\chi_{val}^2$ .

On the other side, it is possible that the fitting methodology is correct and that the fact that these functional forms are unlikely, is related to experimental data features: for example, the fact that low-central-data- $\chi^2$  shows high kinetic energy may be due to the experimental results obtained until this point and the fact of having only a few replicas with this feature is that experimentally it's difficult to obtain high-enough kinetic energy.

It is also possible that the particular functional form of low-central-data- $\chi^2$  replicas may represent only a statistical fluctuation of the experimental results and not a particularly "good" parton distribution functions and that, with further experiments, thus having more and different central-data- $\chi^2$  these same functional forms result in a high central-data- $\chi^2$ .

The other important observation emerging from this study, is the absence of correlation between the low-central-data- $\chi^2$  and the possible over-fitting of the data by the NN. Indeed, if a correlation was present, one would expect to see a clear separation between the low-central-data- $\chi^2$  and the high-central-data- $\chi^2$  on the over-fitting plane, to be precise I'd expect to see a higher number of low-central-data- $\chi^2$ -points (green) in the upper part of the plane, and a higher number of high-central-data- $\chi^2$ -points (orange) points in the lower part, under the bisector line, but that's not the case (see Fig. 4.3 and Fig. 4.4).

Moreover, if a correlation between the central- $\chi^2$  and the fitting performance of the NN was present, one would also expect to see an overall progressive shift from high-central-data- $\chi^2$  to low-central-data- $\chi^2$  (from orange to green), moving upward in the plot, which instead does not happen.

Considering these observations, in order to have a better understanding of the relation between the central- $\chi^2$ , the available data, and PDFs functional forms, it may be interesting to reproduce the scheme of analysis on a set of PDFs produced directly fitting the central data instead of the replicas. This could help understand whether the small number of low- $\chi^2$  replicas is due to the fitting methodology or if it reflects features of the data that need to be further investigated.

# Appendix

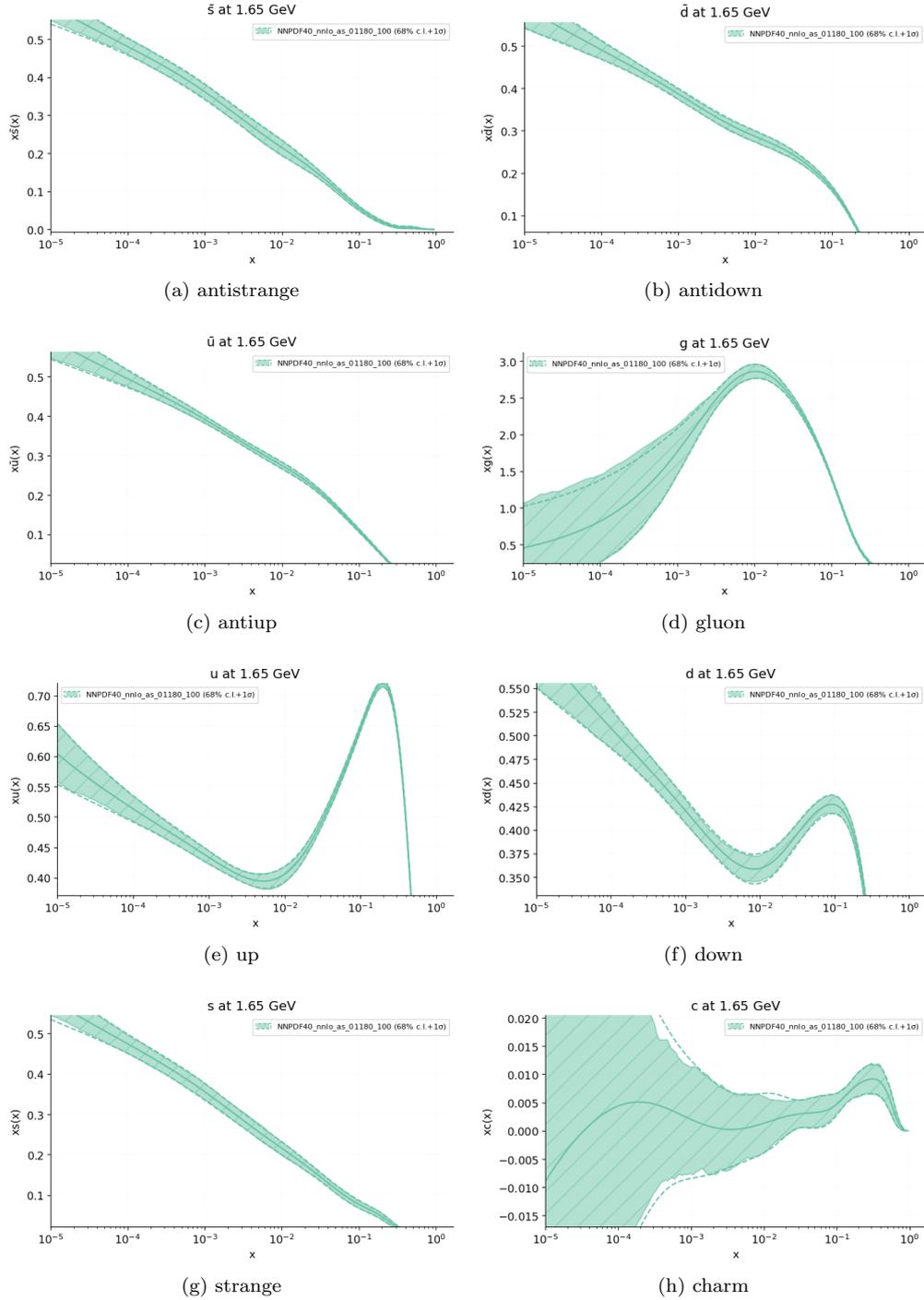


Figure 5.1: Best PDF set obtained from the 150-replicas sett

## 5. Conclusions

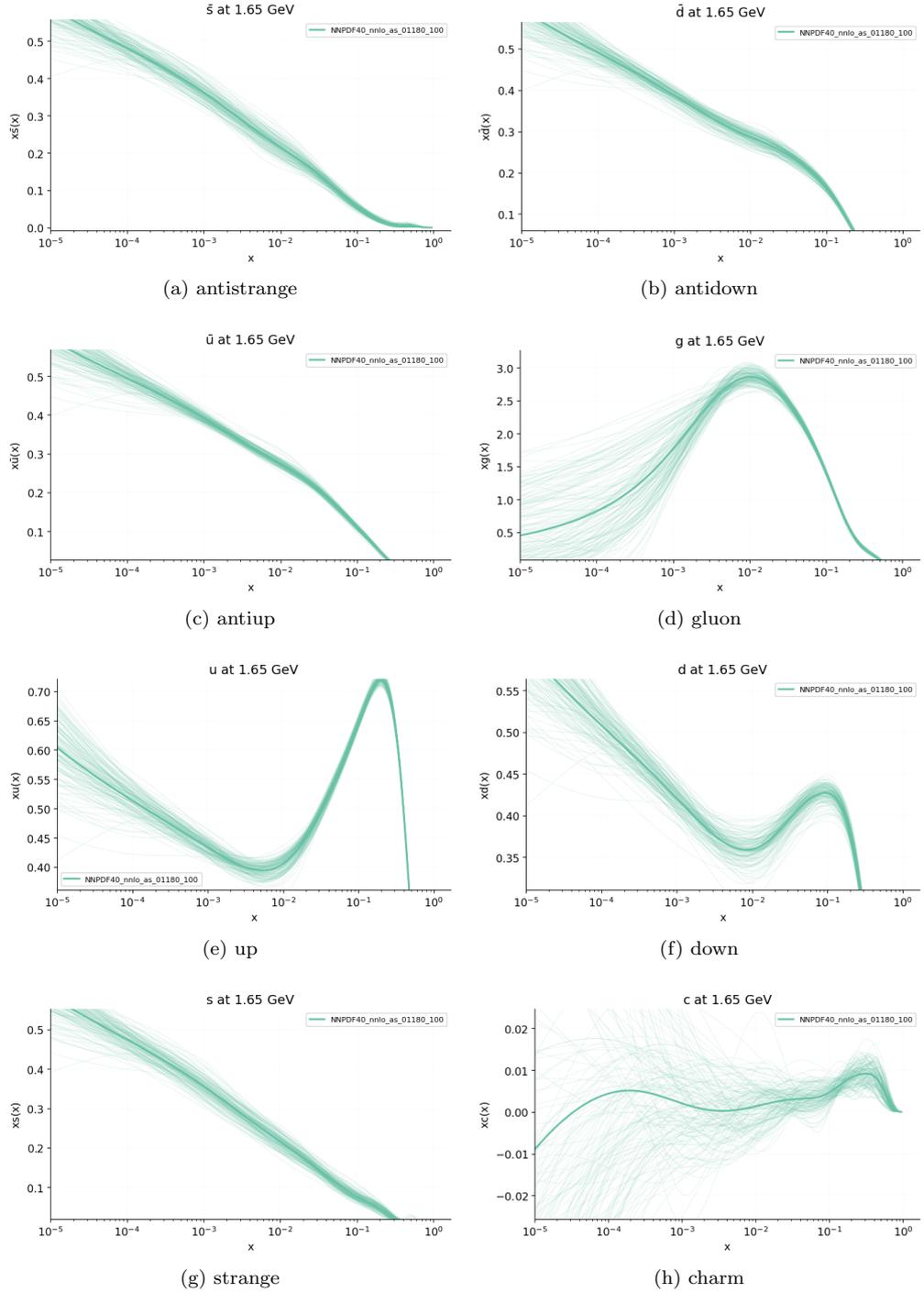


Figure 5.2: Plot of the 150 replicas PDF-set

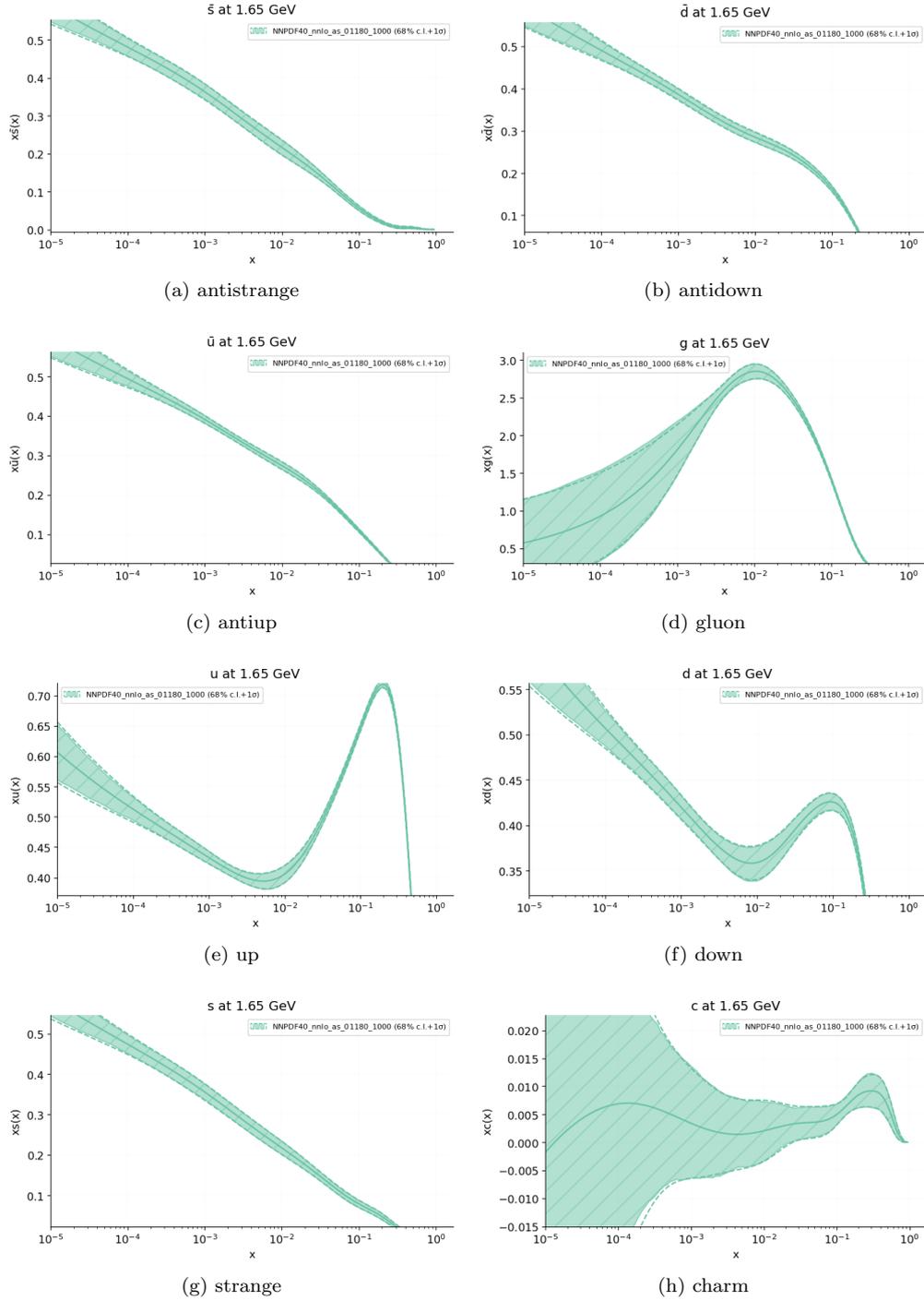


Figure 5.3: PDF set obtained from the 1000-replicas set

## 5. Conclusions

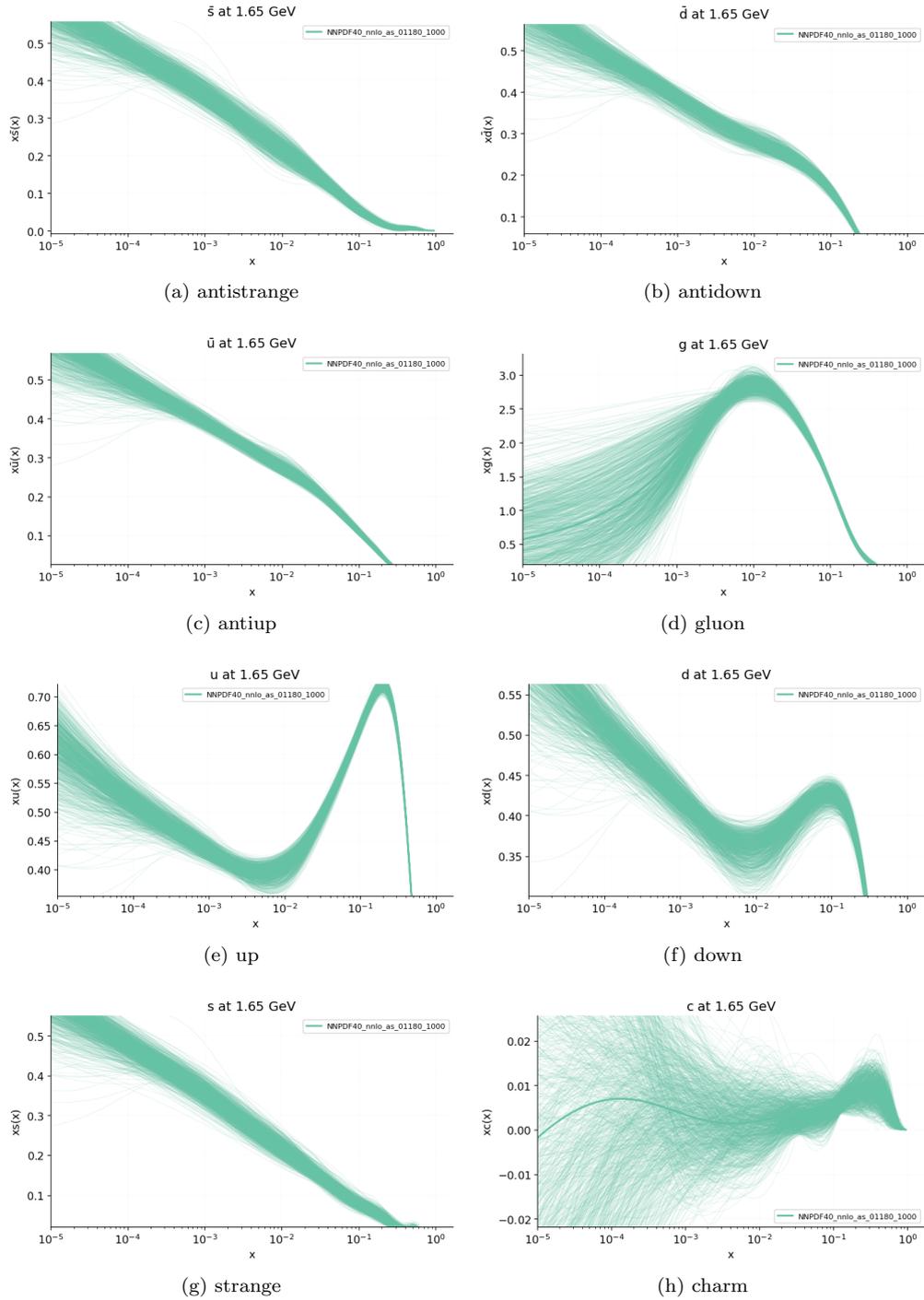


Figure 5.4: Plot of the 1000 replicas PDF-set

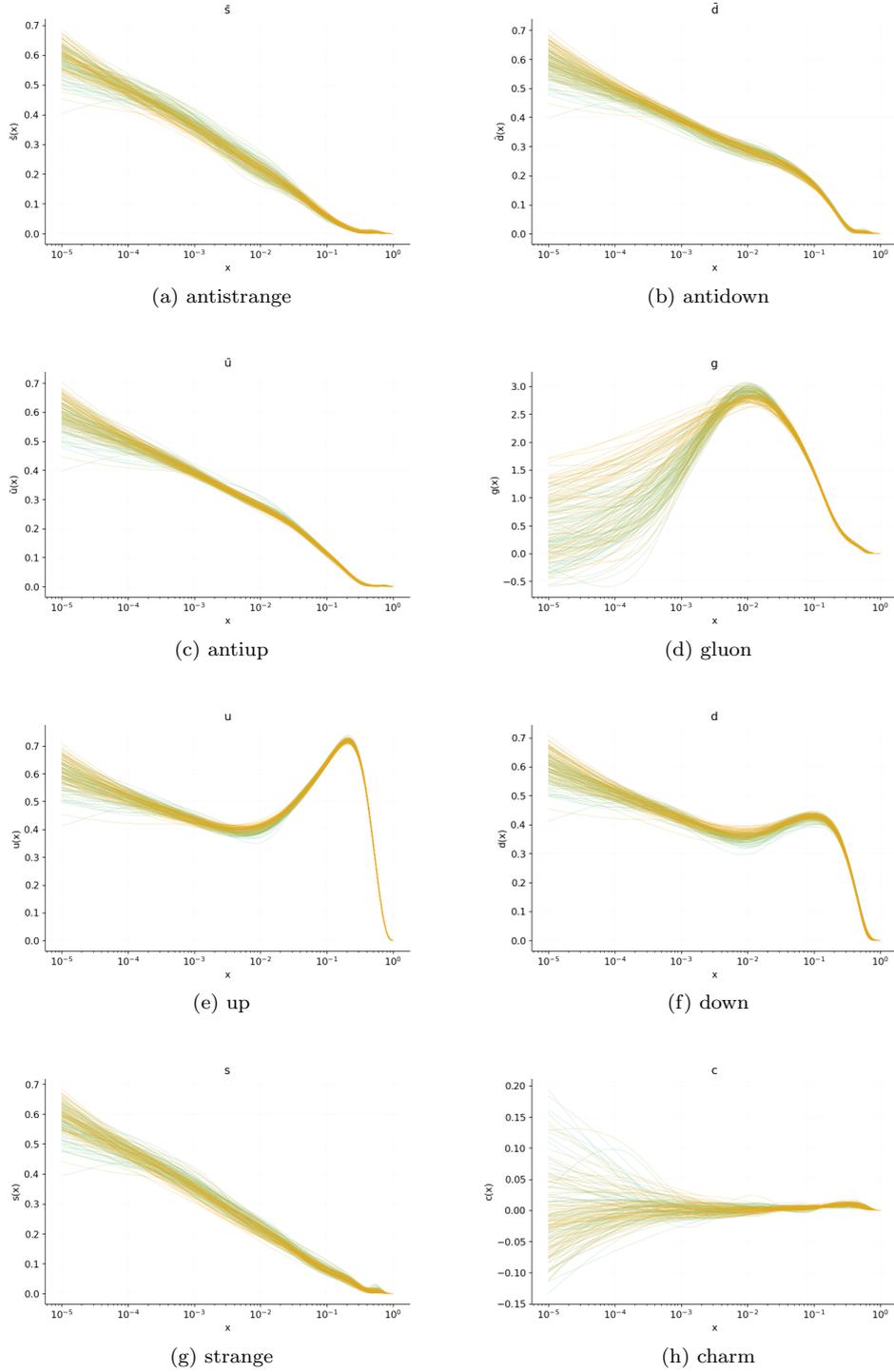


Figure 5.5: Plot of the 15 PDFs replicas data-set with a color scale representing how good they fit the central data

5. Conclusions

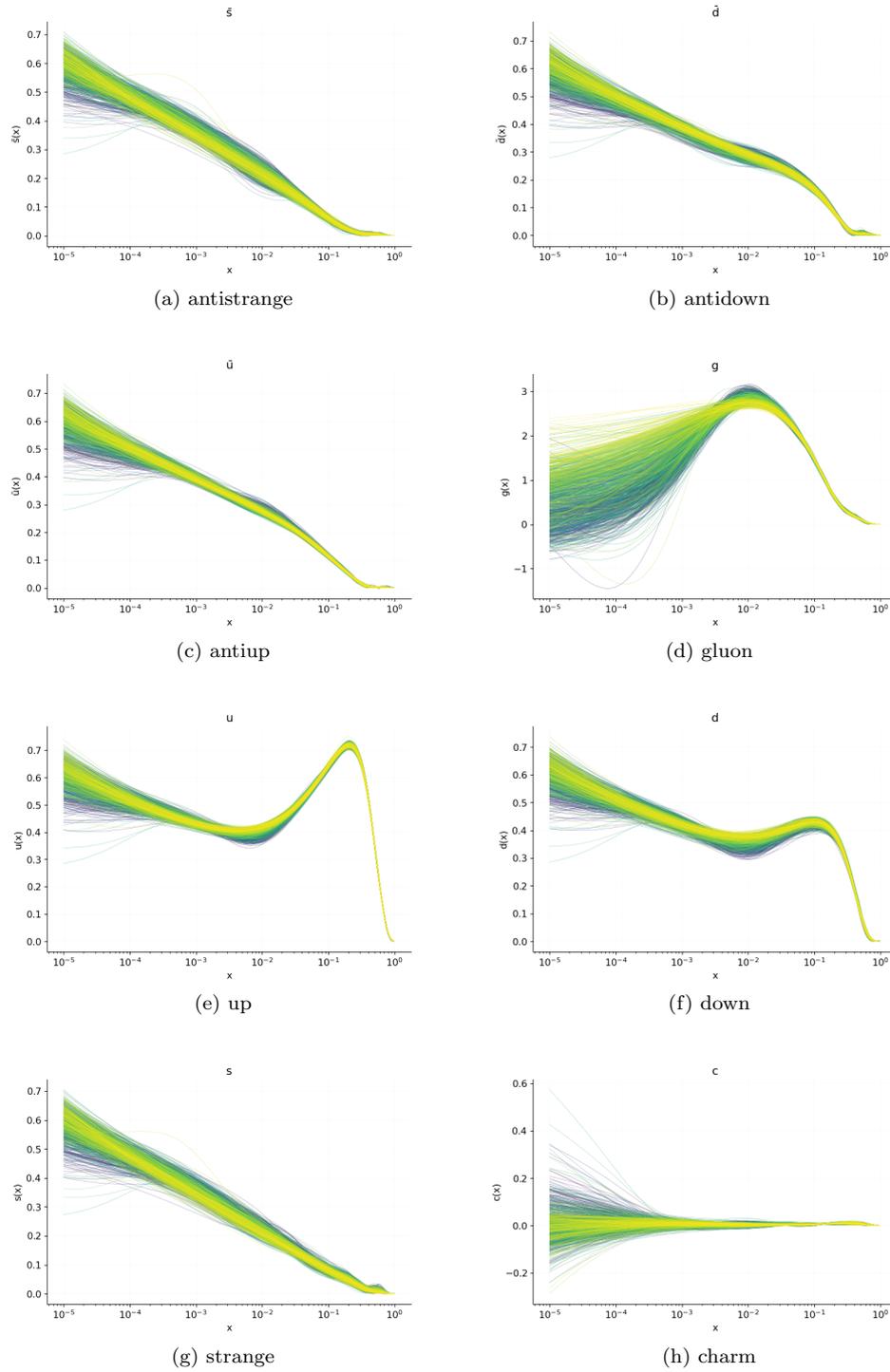


Figure 5.6: Plot of the 1000 PDFs replicas data-set with a color scale representing how well they fit the central data

## Bibliography

- [1] Richard D. Ball et al. Parton distributions from high-precision collider data. *Eur. Phys. J. C*, 77(10):663, 2017.
- [2] Richard D. Ball et al. An open-source machine learning framework for global analyses of parton distributions. *Eur. Phys. J. C*, 81(10):958, 2021.
- [3] Barger, Vernon D and Phillips, Roger J N. *Collider Physics*. Westview Press, Philadelphia, PA, December 1996.
- [4] Wim Beenakker, Christoph Borschensky, Michael Krämer, Anna Kulesza, Eric Laenen, Simone Marzani, and Juan Rojo. NLO+NLL squark and gluino production cross-sections with threshold-improved parton distributions. *Eur. Phys. J. C*, 76(2):53, 2016.
- [5] Valerio Bertone, Stefano Carrazza, Nathan P. Hartland, Emanuele R. Nocera, and Juan Rojo. A determination of the fragmentation functions of pions, kaons, and protons with faithful uncertainties. *Eur. Phys. J. C*, 77(8):516, 2017.
- [6] Alessandro Candido, Stefano Forte, and Felix Hekhorn. Can  $\overline{\text{MS}}$  parton distributions be negative? *JHEP*, 11:129, 2020.
- [7] Stefano Carrazza. Machine learning challenges in theoretical HEP. *J. Phys. Conf. Ser.*, 1085(2):022003, 2018.
- [8] Luigi Del Debbio, Stefano Forte, Jose I. Latorre, Andrea Piccione, and Joan Rojo. Neural network determination of parton distributions: The Nonsinglet case. *JHEP*, 03:039, 2007.
- [9] James E. Dodd and Ben Gripaios. *The Ideas of Particle Physics*. Cambridge University Press, 4th edition, 9 2020.
- [10] R. Keith Ellis, W. James Stirling, and B. R. Webber. *QCD and collider physics*, volume 8. Cambridge University Press, 2 2011.
- [11] Stefano Forte and Graeme Watt. Progress in the Determination of the Partonic Structure of the Proton. *Ann. Rev. Nucl. Part. Sci.*, 63:291–328, 2013.
- [12] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An introduction to statistical learning an introduction to statistical learning*. Springer texts in statistics. Springer, New York, NY, 2 edition, July 2021.
- [13] Zhou Lu, Hongming Pu, Feicheng Wang, Zhiqiang Hu, and Liwei Wang. The expressive power of neural networks: A view from the width. *CoRR*, abs/1709.02540, 2017.
- [14] Joseph D. Lykken. Beyond the Standard Model. 5 2010.
- [15] Mark Thomson. *Modern particle physics*. Cambridge University Press, New York, 2013.
- [16] Andreas Zell. *Simulation of Neural Networks*. Addison-Wesley Professional, 1 edition, 1994.

# Ringraziamenti

Al termine di questo percorso, finora durato tre anni ma destinato a continuare, sono contenta di poter ringraziare tuttø coloro che mi hanno sostenuto.

Ci tengo in primo luogo a ringraziare il professor Forte, per avermi dato la possibilità di svolgere questo lavoro in un ambito di ricerca interessantissimo. Lo ringrazio soprattutto per aver corretto con pazienza i miei errori e risposto a tutte le mie domande.

Ringrazio il dott. Cruz-Martinez, per essere sempre stato estremamente disponibile e puntuale nel rispondere e dirimere tutti i miei dubbi, anche i più banali.

Ringrazio i miei genitori, Davide e Raffaella, per avermi consentito di affrontare serenamente questo percorso di studio e per avermi sostenuto quando la serenità mancava. Ringrazio in modo speciale mia sorella Anna, che mi ha sempre ascoltato nei momenti difficili, mi ha confortato con un abbraccio e una carezza quando ne avevo bisogno, mi è stata vicina e non mi ha mai sminuito, se sono qui oggi è grazie a te; già che ci sono insieme a te ringrazio anche la nostra Wendolina. Ringrazio la nonna Palmira per essersi sempre divertita ad ascoltare le mie divagazioni e i miei racconti sugli aspetti più strani della fisica e per avermi sempre lasciato lo spazio di crescere nella direzione che preferivo. Ringrazio tutta la mia famiglia, siete compagni e compagne di viaggio fantastici.

Ringrazio le mie meravigliose compagne di laboratorio Dani e Rosa con cui ho pianto sullo studio degli esperimenti e l'analisi dei dati, con cui ho condiviso le serate passate a scrivere relazioni e con cui ho condiviso anche la gioia (incommensurabile) dell'aver terminato gli esami di laboratorio. Ringrazio Ale, con cui, dal primo parziale di meccanica ho condiviso l'ansia per gli esami e la gioia e la frustrazione al loro termine. Ringrazio tutti i compagni di studio e di pranzi, Max, Ale, Carlo, Tommi, Andre e tutti gli altri per aver reso questo percorso un po' meno faticoso e sicuramente meno solitario. Ringrazio Leo per avermi spronato a fare del mio meglio e aver creduto in me anche quando non ci credevo io, grazie.

Ringrazio la mia prof di Matematica e Fisica del liceo, Camilla Cervi, che mi ha fatto scoprire per la prima volta un mondo incredibilmente affascinante, fatto di particelle e onde e governato da equazioni, per avermi mostrato un modo di pensare fuori dagli schemi che mi ha spinto a scegliere questo percorso.

Ringrazio me stessa, per essermi ascoltata, per essermi messa in gioco nel fare qualcosa che mi sembrava impossibile ed essermi fatta forza in ogni momento. Ringrazio me stessa anche per essere stata in grado di chiedere aiuto e di fermarmi a riposare quando andare avanti sarebbe stato troppo difficile. Ringrazio per

questo anche tutte le specialiste che mi sono state vicine durante questo ultimo anno, la dott.ssa Rovaris, psicologa, che per prima mi ha accolto e indirizzato in questo percorso, la dott.ssa Pinto, la mia psichiatra, che mi ha aiutato a prendere confidenza con me stessa, il mio corpo e miei pensieri, la dott.ssa Ferrari, mia terapeuta che mi ha mostrato il mondo che si nascondeva dentro di me e che mi ha aiutata a guardarlo con comprensione. Ringrazio anche tuttø coloro che non ho nominato ma che mi sono statø accanto in modi diversi, grazie di esserci statø.