UNIVERSITÀ DEGLI STUDI DI MILANO

FACOLTÀ DI SCIENZE E TECNOLOGIE

Corso di Laurea triennale in Fisica

DETERMINATION OF THE CHARM QUARK MASS IN A GLOBAL PARTON FIT

Relatore:
Prof. Stefano Forte
Correlatore:
Dott. Stefano Carrazza

Tesi di Laurea di:
Elisa Radaelli
PACS: 12.38-t
Matricola: 866750

Anno Accademico 2017/2018

The first principle is that you must not fool yourself — and you are the easiest person to fool

-Richard Feynman

Contents

Chapter 1

Parton Distribution Functions

1.1 Quantum Chromodynamics

Quantum Chromodynamics (QCD) is the theory of the strong interaction, one of the four fundamental forces in nature. It describes the interactions between quarks and gluons, and in particular how they bind together to form the class of particles called hadrons.

QCD emerged as a mathematically consistent theory in the 1970s, and nowadays is considered by scientists as one of the cornerstones of the "Standard Model" of the elementary particles and their interactions.

One of the most used methods in high-energy physics research is collider physics: particle colliders provide the highest available centre-of-mass energies and this permit us to probe the structure of matter at the shortest distances availables [1].

This last fact can be explained with the **asymptotic freedom**, the tendency of quark to be free at small distances (high energy scale), when the forces between them become weaker. On the opposite side of the energy scale (at long distances) quarks are seen as bind together for the **confinement**, the tendency of components inside nucleons to be considered as highly interacting states.

This behaviour is explained also by the values of the strong coupling constant α_s: it decrease at short distances and grows ad higher ones.

The quark model derived from the **parton model**, described in 1969 by Feynman, in order to give a simple explanation about the nucleons and their smaller components, labeled at the time as **partons**. Now we know that inside the nucleons and hadrons there are quarks and gluons, although they do not have a well defined form. The internal structure of nucleons

is described by Parton Distribution Functions (PDFs), which are literally distribution functions that are able to provide a good explanation of the features of particles inside hadrons and their behaviour [2].

1.2 PDFs

Parton Distribution Functions (PDFs) encode informations about the structure of strong interacting particles (like hadrons), and in particular of nucleons, which are made of partons. Nowadays the internal structure of nucleons is not well defined for the confinement, because their components can't be free. For this reason to describe their internal structure are used this function distributions, which give a parameterization of the nucleons internal components. In this framework nucleons are composed of quarks and gluons, representing the elementary degrees of freedom for Quantum Chromodynamics (QDC).

Physically, PDFs have an important role in scattering processes. In particular, measuring the cross section σ for any process involving hadrons in the initial state, that is determined by folding PDFs with the perturbatively computable cross section that describes the interaction between partons (equation (1.1))

$$\sigma = \sum_{a,b} \hat{\sigma}_{a,b} \otimes f_{a/b_1} \otimes f_{b/b_2}, \tag{1.1}$$

where σ is experimentally measured and $\hat{\sigma}_{a,b}$ is the partonic cross section computed with quarks and gluons in the initial state. The convolution product of the a^{th} parton with the the b_i hadron PDFs in (1.1) represents an integration over the relevant initial-state kinematic variables.

Parton distributions are a set of functions determined through the comparison between the PDF-dependent prediction for physical processes and its actual experimental value.

1.2.1 Theoretical framework

To determine PDFs it is necessary to start from a theoretical prediction of various processes, which is compared to the experimental data.

The basic property that enables the perturbative computation of the cross sections for processes with hadrons in the initial state is their factorization into a partonic cross section (computed in perturbation theory) and parton

distribution (which characterizes the hadronic bound states and are universal).
The factorization is different for the hadroproduction and for electroproduction: in the first case the scaling variable is a parameter τ, which depend on the mass of the hadron in the final state and on the center-of-mass energy of the hadronic collision; the second case uses structure functions $F(x, Q^2)$, where x is the Bjorken scaling variable and Q^2 the momentum of the virtual photon produced.

Theory prediction for collider cross section can be written also as Equation (1.2)

$$\sigma^{(th)}\{\omega, \theta\} = \hat{\sigma}_{ij}(Q^2) \otimes \Gamma_{ij,kl}(Q^2, Q_0^2) \otimes q_k Q_0\{\omega, \theta\} \otimes q_l Q_0\{\omega, \theta\}, \quad (1.2)$$

where the $\otimes$ is the convolution product on x, $\hat{\sigma}ij$ (the hard-scattering cross section) and $\Gamma_{ij,kl}$ (the kernel of the DGLAP evolution), using computational parameters (for further details see Section 2.3).

In order to describe in a much compact way cross sections it is possible to use an other approach, where the datasets from experiments are divided into grids, represented as rows of a table, created using the DGLAP evolution operator, which is the responsible of the evolution in perturbative framework of the various data in energy.

In this way, the cross sections can be expressed in a much compact way, as in Equation (1.3)

$$\sigma^{(th)}\{\omega, \theta\} = \sum_{i,j=1}^{n_f} \sum_{a,b=1}^{n_x} FK_{ijab} q_i x_a, Q_0, \{\omega, \theta\} q_j x_b, Q_0, \{\omega, \theta\}, \quad (1.3)$$

where I labeled as "FK" the tables created from the various rows of data and in Equation (1.3) represents tables where a and b run over a grid in x.

The PDFs are also defined using luminosities, defined as the number of events in units of time and solid angles, written as Equation (1.4) [1] and (1.5), considering the relevant parton-parton luminosities, rather than the PDFs themselves:

$$L_{ab}(x, M_X^2) \equiv \int_x^1 \frac{dz}{x} f_{a/b_1}(z, M_X^2) f_{b/b_2}\left(\frac{x}{z}, M_X^2\right) = \int_x^1 \frac{dz}{z} f_{a/b_1}\left(\frac{x}{z}, M_X^2\right) f_{b/b_2}(z, M_X^2), \quad (1.4)$$

[1] where X is the final states and the a and b pedix label the incoming hadrons.

$$\sum_{q=u,d,s,c,b} (L_{q\bar{q}} + L_{\bar{q}q}). \tag{1.5}$$

The evolution of PDF determination has gone through various stages that mirror the evolution of the theoretical and phenomenological understanding of QCD. There was an improvement in methodology, given by the growing accuracy of experimental data and an improving confidence in perturbative QCD. Analyses are done using tools not only at the leading order (LO), but especially at high perturbative orders (NLO, next to leading order and NNLO, next-to-next leading order).

An important development came from the deep-inelastic data from the HERA collider, which gave a great increase in both accuracy and kinematic coverage, and yielded the global parton sets. For PDFs was assumed a functional form, which was then parameterized by a relatively small number of constants. The global parton sets were eventually determined by optimizing the fit of the computer observables to the experimental data.

Figure 1.1 [3] shows a set of PDFs at two different energy scales: to understand properties of hadronic cross sections and the impact of PDFs on them, one should consider the relevant parton-parton luminosities, defined as Equation (1.4).

1.3 The Determination of PDFs

1.3.1 Fitting methodology

Parton distributions are determined through comparison between the factorized expressions of them and the experimental data. To obtain the best fit, one determines a "confidence interval" in the PDFs space. By introducing a PDF parameterization it is possible to reduce an infinite-dimensional problem of representing a space of functions in a finite-dimensional form (which is mathematically well defined).

1.3.2 Uncertainties in the PDF

At first the only way to estimate uncertainties in the PDFs was to compare results obtained with several parton sets, although historically physicist obtained the set of PDFs with uncertainties by fitting only a restricted data sets, typically from deep inelastic scattering (DIS) experiments, and propagating all the correlated uncertainties through the fitting procedure.

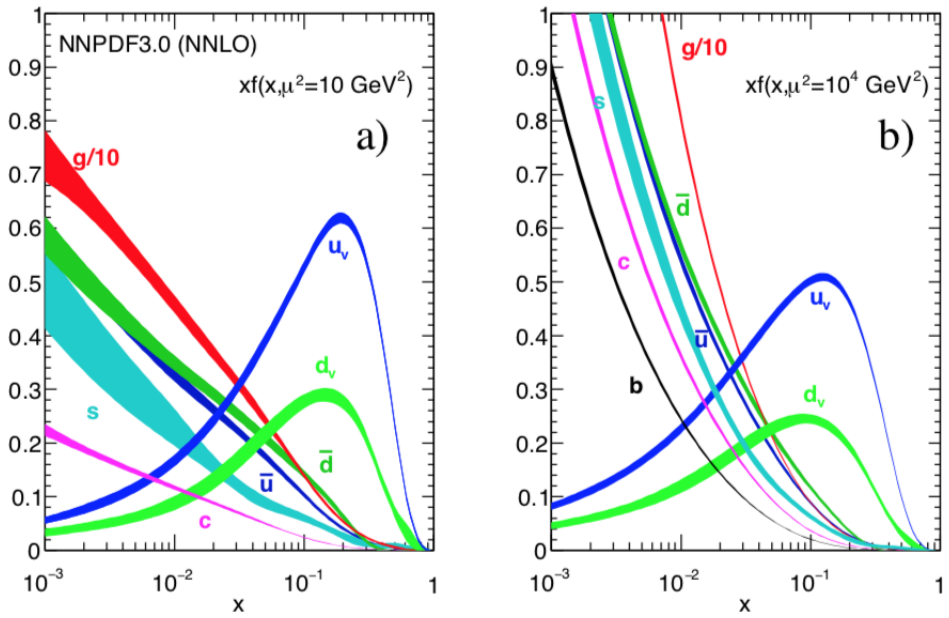


Figure 1.1: Parton distribution obtained in NNLO NNPDF global analysis at scales (a) $\mu^2 = 10$ GeV^2 and (b) $\mu^2 = 10^4$ GeV^2 with $\alpha_S = 0.118$.

It was however necessary to use a systematic approach: to obtain statistically meaningful results it is necessary to use a *tolerance* rescaling, defined as a dynamic tolerance procedure for the determination of $\Delta\chi^2$, considered as > 1, introduced to take account of tension between data sets in the fit (see [4] and [5]). Thereafter one can determine "error" PDF sets along with the central best fit, allowing for the determination of a 1-σ contour in the parameter space of the best fit. Later more refined versions of the tolerance method were developed and have been used in subsequent global fits (see [6] and [7]).

There is an alternative approach to PDF fitting, proposed by the NNPDF collaboration, used in this work and better explained in Chapter 2.

1.3.3 Parton parameterization

A set of PDFs is a set of functions, one for each parton entering in the equation used for the two different productions named in Subsection 1.2.1. In principle there could be 13 independent PDFs in a given hadron (six quarks, six antiquarks and the gluon). At first studies there were only two PDFs, one for the quarks and one for the gluons, because the PDFs for all

the quarks are assumed as equal. After a lot of studies, and considering that valence quarks' PDFs are different from sea quarks' ones, it is possible now to parameterize eight PDFs for each hadron (because the PDFs of charm and anti-charm are considered as the same).

Once a suitable basis set of PDFs has been chosen, all existing PDF determinations are based on the choice of a parameterization of PDFs at the reference scale:

$$f_i(x, Q^2) = x^{\alpha_i}(1-x)^{\beta_i} g_i(x), \tag{1.6}$$

where $g_i(x)$ tends to a constant for both $x \to 0$ and $x \to 1$.
A common choice for $g_i(x)$ is a polynomial function or the exponential of a polinomial in x or $\sqrt{x}$. Moreover, typical contemporary PDF sets based on this choices of functional forms are parametrized by $\sim$ 20-30 parameters.

There are also two ways to parameterize PDFs without using any theoretical prejudice: the Hessian Method or the MonteCarlo approach.

When unbiased PDF parameterizations are adopted and especially when the number of free parameters is very large, the absolute minimum of the figure of merit is not necessarily the best fit, because it may consider random fluctuations in the data or display oscillations [8].

1.3.4 Global fits

For each process considered and for each datasets a great amount of information is used to determine PDFs; this can become computationally intensive. Current global fits use various processes to control different aspects of PDFs.

Different data constrains different aspects of PDFs and are for example:

- Information on the overall shape of quarks and gluons at medium x comes from fixed-target DIS data on proton and deuterium targets;
- The determination of the behavior of the gluon and quark is done at small x and of the individual light flavors at medium x;
- Information on the flavor separation at small x comes from Tevatron Drell-Yan data (in particular from the W asymmetry);
- The flavor separation at medium x is mostly controlled by Tevatron Drell-Yan data on fixed proton and nucleus target;
- The total valence component is constrained by the neutrino inclusive DIS data;

- Strangeness is constrained by neutrino dimuon data and the interplay of W and Z production data with lower scale DIS and Drell-Yan data;
- The large x gluon, already determined by DIS scaling violations is constrained by Tevatron jet data.

Other global fits may differ in some detail (such as the specific choice of experiments or using different data sets), but all of them are based on the datasets constructed on a similar logic.

Now there are newer data and best methods to make a global fits, considering processes with less uncertainties.

Chapter 2

Machine Learning and Neural Networks

2.1 General strategy

In order to reach the final goal of this work (see Chapter 3) is necessary to determine the PDFs of the proton in the global QCD analysis framework. All data used for the determination derive from proton-proton and hadron-hadron collisions.

For a more detailed description of the tools here described see for example [9] and [10] and references therein.

2.1.1 Elements of the NNPDF global analyses

Global QCD fit is based on three inputs:

- experimental data;
- higher order perturbative calculations;
- statistical framework based on PDFs parameterization, estimating and propagating their uncertainties.

Combining together these three elements it is possible to make a fit. One defines the minimization of the figure of merit, the χ^2, which includes all the relevant sources of uncertainties and leads to define the shape of each PDF.

Fits are validated using complementary diagnosis tools and are translated into LHAPDF standard interface suitable for its public delivery and its integration into other HEP codes and into the analysis framework of LHC experiments. The general strategy used is summarized in Figure 2.1 [11].

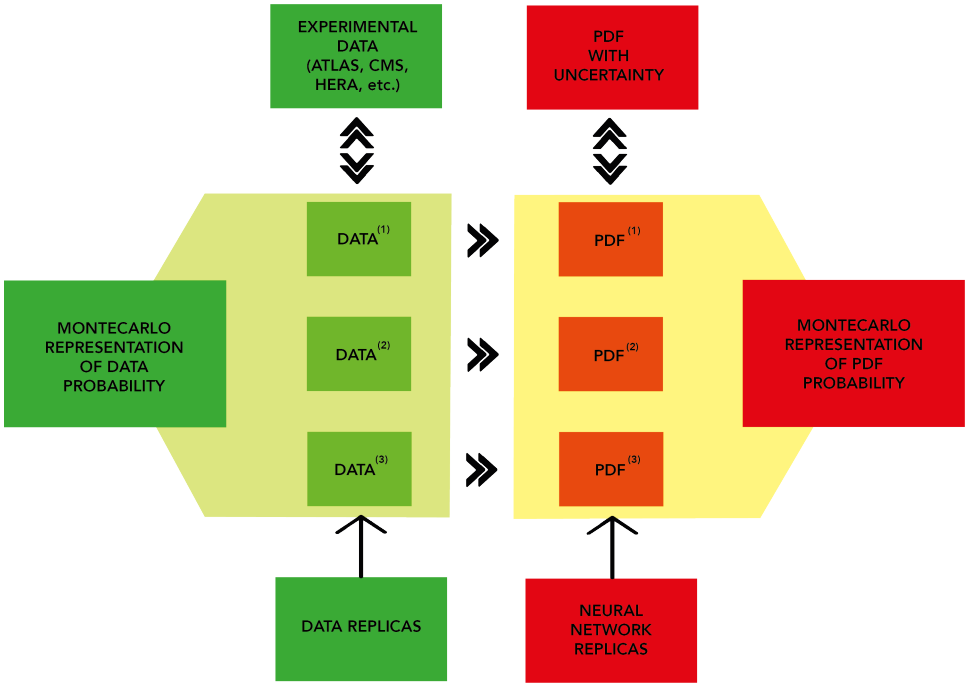


Figure 2.1: General strategy used in the NNPDF approach.

2.1.2 Machine learning

Machine learning techniques are useful in high-energy physics and are peculiar in the PDFs parameterization and optimization. It is possible in this way to define the best fit parameters and the subsequent validation, by means of closure testing.

Multi-layer Neural Networks are used to parameterize a PDF independent model and to do fragmentation, genetic algorithms are used for training and optimization and then a final test is useful in order to have a sistematic convalid of fitting methodology.

2.2 The Monte Carlo approach

PDFs at all scales can be obtained by solving evolution equations. A determination of PDFs with uncertainties involves determining a probability distribution in a space of several independent functions. Various methods can be used to analyse experimental data. In particular, one possible choice is that of the time-honored method, which assumes a specific functional form for parton distributions, projecting the infinite dimensional problem onto a finite-dimensional parameter space, and so that a representation of PDFs

must be possible in terms of a finite number of parameters and there is an optimal parameterization.

Two methodologies are complete and actually used: the Hessian approach, in which the best fit result is given in the form of an optimal set of parameters and an error matrix centered on this optimal fit to compute uncertainties, and a Monte Carlo approach, in which 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 first case the parametrization is standard and inspired by QCD arguments; in the second method it is possible to use artificial Neural Networks as interpolating functions in an attempt to reduce the bias related to the choice of functional form.

2.2.1 The role of replicas in the Monte Carlo approach

In a Monte Carlo approach the probability distribution in the parameter space is given by assigning a Monte Carlo sample of replicas of the total parameter set.

In order to use a statistical framework, experimental data are converted into an ensemble of N artificial MC replicas, which are named *pseudodata*: these are randomly generated in accordance with multi-Gaussian distributions centered around each data point, with variance given by experimental uncertainty.

Each replica contains the same number of data points as the original experimental measurements. With enough replicas the set contains complete experimental informations; experimental central value can be retrieved by taking the mean and the experimental variance (which is calculated over different replicas).

All possible approach includes data from various experiments of fixed target colliders, deep inelastic scattering and Drell-Yan experiments; the results obtained are in form of observables such as cross sections and structure formation. Experimental data are converted from the row format provided by a single dataset to a common format used in order to fit them.

2.2.2 Monte Carlo treatment of uncertainties

Given an observable X I will have N_{rep} replicas, each time using a different parameter replica: the central value for X is the average of these N_{rep} results and the standard deviation is the variance. The procedure is as follows: at first one starts with experimental data (denoted as F_I) from which are generated N replicas. Subsequently, it is found a set of PDFs (denoted as $q_0(i)$) for each replica. The PDFs can be parametrized at some reference

scale and they are fitted to the data replicas, evolving them to the scale of the data and using them to compute observables and minimizing the χ^2 of the comparison to the data with respect to the parameters.

However, the problem of constructing an adeguate sampling of parameters space has been reduced to that of constructing an adeguate Monte Carlo representation of the original data. A given set of replicas provides an accurate enough representation of the data that may be checked explicitly for a given sample by comparing means, variances and covariances from the sample with the desired features of the data.

2.3 Neural Networks

Neural Networks are a set of computing systems, used to fit some data without giving a precise function, because the networks adapt themselves to the experimental results.

The parton distribution functions are parameterized at low scale, around boundary between perturbative and non-perturbative regimes of QCD, around the proton mass of $Q_0 \approx 1$ GeV. The PDF shape, instead, is not parameterized in terms of simple function forms (inspired in QCD models), but are used artificial Neural Networks as unbiased interpolants (NN do not use a specifical model of functional form).

QCD itself provides some limits about the behaviour of PDFs at the input parameterization scale Q_0, such as the integrability conditions and the momentum and valence sum rules.

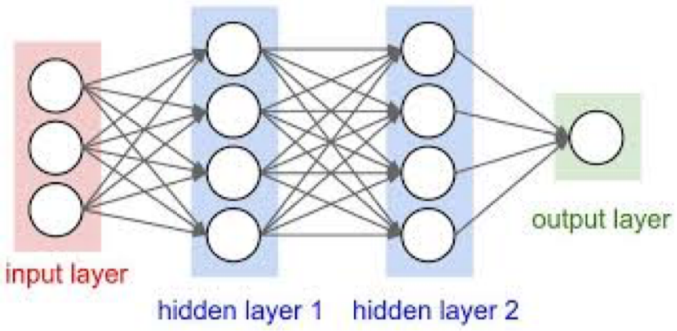


Figure 2.2: General Architecture of a Neural Network structure.

Neural Networks are developed in a multi-layer structure, with a given number of parameters put in the input layer, others in the hidden layers and at the end there is the output, as schematically represented in figure 2.2 [12].

The easiest form of them, with a 1-2-1 architecture, can be written as in Equation (2.1)

$$f(x) = \frac{1}{1 + e^{\theta_1^{(3)}} - \frac{\omega_{11}^{(2)}}{1+e^{\theta_1^{(2)} - x\omega_{11}^{(1)}}} - \frac{\omega_{12}^{(2)}}{1+e^{\theta_2^{(2)} - x\omega_{21}^{(1)}}}}, \tag{2.1}$$

where the ω_{ij} are the weights and the θ_i are the theresholds.
The NNPDF fit use a multi-layer feed-forward artificial NN with a 2-5-3-1 architecture for each PDF, with two inputs and one output neuron (directly related to the value of PDF at input parameterization scale Q_0). There are 296 parameters in the PDF determination in this framework.

The activation threshold of each neuron is denoted by ξ_i^l, with (l) as label of the layer and i is the specific neuron within each layer.

The values of each activation states of neurons in the layer l are determined in terms of those of the previous layer $(l-1)$ and weights $\{\omega_{ij}^{(l)}\}$, connecting them as the activation thresholds of each neuron $\{\theta_i^{(l)}\}$. The training of Neural Network consists in determining values of weights and thresholds that fulfill the constraints of a given optimization problem.

2.3.1 Neural Network methodology

The PDFs are functional forms, and they may be fitted in the limit of infinite number of parameters. They are nonlinear and "unbiased", meaning that a finite-dimensional truncation of the neural network parametrization is adeguate to fit a very wide class of functions without the need to adjust the form of the parametrization to the desired problem.

The best fit is instead determined using a cross-validation method. The data are randomly divided into two samples: the training sample and the validation one. The χ^2 is computed for both the samples of data. The validation stops when the χ^2 for the two groups are in the same point, over which the data is considered to be background noise [13].

2.4 Minimization and cross validation methods

2.4.1 Minimization methodology

Once PDFs has been parameterized, the optimal fit is obtained varying parameters of neural network in a way where the figure of merit is minimized.

The training (or learning) phase of Neural Network is possible to make using two different strategies: the deterministic one, with the gradient-descent and the back-propagation method, and the stochastic, used in this work, which include a random selection of the data. The determination of fit parameters requires evaluation of gradients of χ^2

$$\frac{\partial \chi^2}{\partial \omega_{ij}^{(l)}} \qquad and \qquad \frac{\partial \chi^2}{\partial \theta_i^{(l)}}, \tag{2.2}$$

and computing these gradients in the NNPDF case give a non linear relation between experimental data and input PDFs, preceeds through convolutions with DGLAP evolution kernels and hard-scattering partonic cross section (encoded into APFELgrid fast interpolation strategy).

2.4.2 Genetic Algorithm

In proton NNPDF global analysis, the NN training carried out by the Genetic Algorithms is based on a combination of deterministic and stochastic ingredients. In this way is possible to explore complex parameters spaced without getting stuck in the local minima. In this way is not required the knowledge of χ gradients, but only local values.

Starting from a random population of solutions far from the minimum, the spread (variance) of the population increases while at the same time the average (center) solution moves closer to the minimum. As the number of generations increases, the average solution remains close to the minimum but now the variance has been reduced significantly, indicating that the algorithm has converged.

2.4.3 Cross-validation

It is important to say that, although all said before, fit has not an absolute minimum of the figure of merit, because in this way he would fit also the noise. In order to avoid this is possible to use the cross-validation method to determine the optimal fit.

The cross validation stopping strategy can be explained in four steps:

1. All the input experimental measurements or all the data replicas are divided in a random way in two categories: the training sample and the validation one, both with the same probability. In the fit is used only the former, while the lattice is used as a control sample, to monitor and validate the training process.

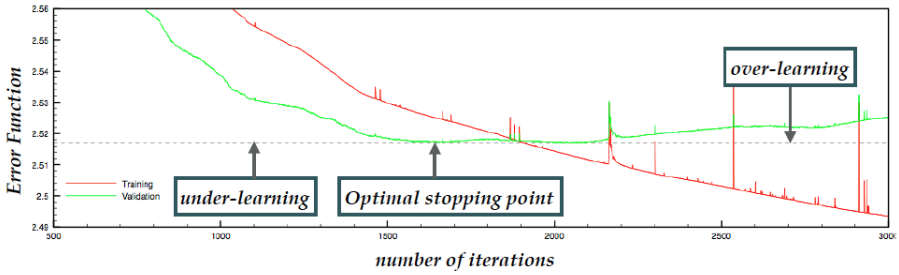


Figure 2.3: Trading and validation sample fits.

2. In the figure in 2.3 [14] is labeled the optimal stopping point: one defines the global minimum of the χ^2 of the validation sample, computed over a large fixed number of iterations (and from here the name"look-back").

3. As shown in 2.3 the shorter fit is in an under-learning part, and NN are not learning well the law. On the contrary, longer fits leads to the over-learning, where NN ends up fitting statistical fluctuations.

4. The tell-tale sign of the latter is the addition of the validation of χ^2, that grows as the number of interactions increases. What is learned in the trading sample is not in the validation (because here there are the fluctuations).

Chapter 3

Results

The goal of this thesis is the determination of the charm quark mass, the lightest of the so-called heavy quarks, within the context of a global fit of parton distributions, a set of PDFs (obtained from experimental data).

In this last chapter I will explain my work in by describing step by step the procedure I followed. I start talking about the theories and what could be varied in each one of them. Then I explain the treatment of grids and experimental data, how Neural Networks are used and the computational features used in order to obtain the final result. At the end I say how I created a code to plot the results of the fit (actually the χ^2 from the comparison of two fits) and find the minimum. This value is the charm quark mass that I find in this framework.

In some parts of this chapter I give some informations that I will not discuss further, because it is beyond the goal of this thesis.

3.1 Theories

By "Theory" I mean the set of the predictions to be compared to the data. These depends on a large number of parameters.
In the NNPDF repository all the theories are stored in a file named "*theory.db*", a database file with now 198 theories.

All of them depend on diffferent parameters, necessary to determine all possible PDFs corresponding to each theory. These includes for example:

- the perturbation order (which could be LO, NLO or NNLO),

- the flavour number scheme,

- the intrinsic charm switch,
- the DGLAP solution mode
- the FK table initial scale Q_0,
- the value of the strong coupling constant α_S,
- values that express the dependence on QED,
- Q_{ref}, the reference scale for α_S in GeV,
- the charm quark mass and the "c" production threshold ratio,
- the same values for the bottom and the top quark (the two heaviest quarks),
- CKM matrix elements,
- Z boson and W boson mass (in GeV),
- target mass corrections,
- the proton mass M_P,

and other parameters that are expressed in the QCD Lagrangian [15].

I only varied the charm quark mass and its reference scale (both in GeV), choosing as reference theory the one with a charm mass of 1.51 GeV (at NLO perturbation order). This value is what is labeled in Equation I.1.19 in [16], where the pole charm mass is $m_c = 1.51 \pm 0.13$ GeV.

I changed this value from 1.11 GeV from 1.81 GeV, spaced by 0.1 each time, as summed in table 3.1.

3.2 APFELcomb and FK tables

3.2.1 Experimental data

The APFELcomb project consists of a set of tools for the generation of FK tables, which provide the mechanism for computing predictions from the theories in the NNPDF framework. This is achieved by taking DGLAP evolution kernels from the APFEL code [17] and combining them with interpolated parton-level observable kernels of various forms.

Theory ID	Charm Mass (GeV)
190	1.11
191	1.21
192	1.31
193	1.41
194	1.51
195	1.61
196	1.71
197	1.81

Table 3.1: Summary of the new theories with the charm mass changed.

For this reason I considered the experimental data, divided into various datasets. In a technical language, *Dataset* refers to the result of a specific measurement, typically associated with a single experimental paper; instead *Experiment* refers to a collection of *Datasets*, which have correlated uncertainty.

Data, made available by experimental collaborations, come in a variety of formats. For use in a fitting code, this data must be converted into a common format that contains all the required information called as **CommonData**.

Each datapoint has an associated process type string: for each datapoint three kinematic values are given: the first is the principal differential quantity used in the measurement, the second defines the scale of the process and the third represent the centre-of-mass energy of the process, or its inelasticity.

There are three basic process types:

- **DIS**: Deep Inelastic scattering measurements,
- **DYP**: Fixed-Target Drell-Yan measurements,
- **APP**: Hadronic measurements,

each including a large number of experiments, summarised in Table 3.2, where the numbers from 1 to 8 indicate DIS experiments, 9 and 10 the DYP and the remaining the APP ones.

All the data from the various datasets are computed in a list of grids and subgrids (using evolution operator), which are merged in a single **FK table** for each dataset.

1. NMC.
2. SLAC,
3. BCDMS,
4. CHORUS,
5. NTVDMN,
6. HERACOMB,
7. HERAF2CHARM,
8. F2BOTTOM,
9. DYE886,
10. DYE605,
11. CDF,
12. D0,
13. ATLAS,
14. CMS,
15. LHC

Table 3.2: Experiments from which are derived all datasets used.

3.2.2 FK tables

In the NNPDF project, FK tables (or grids) are used to provide information required to compute QCD cross sections in a compact way. With the FK method a typical hadronic observable datapoint O is computed as

$$O_d = \sum_{\alpha,\beta}^{N_x} \sum_{i,j}^{N_{pdf}} \sigma_{\alpha\beta ij}^{(N_x)(d)} N_i^0(x_\alpha) N_j^0(x_\beta), \tag{3.1}$$

where $\sigma_{\alpha\beta ij}^{(N_x)(d)}$ is the FK table, a five index object with two indices in flavour (i and j), two indices in x (α and β)) and a datapoint index (d). $N_i^0(x_\alpha)$ is the i^{th} initial scale PDF in the evolution basis at x-grid point $x = x_\alpha$. Each FK table has an internally specified x-grid upon which the PDFs are interpolated.

For each FK table the following information is provided:

- GridDesc: a short description of the table,
- VersionInfo: a list that specify the versions of the various pieces of code used in the generation of the table,

- GridInfo: a list that specify the architectural points of the FK tables,
- TheoryInfo: a list of theory parameters used in the generation of the table,
- FlavourMap: a segment (for DIS) or a matrix (for APP) that describes the flavour structure of the grid,
- xGrid: a segment that defines the x grid upon which the FK grid is defined.

All the FK tables are stored as *FastKernel*, which is different for all the processes considered. In addition there are the *Compound* files and the *CFactor* ones.

Additional multiplicative factors to be applied at the output of the FK convolution may be introduced with CFactor files, while Compound files contains information on how to build the observable from FK tables. These last files implement operations between the final observable prediction for one definite point in the Dataset and the observable prediction for that point arised from the FK table calculation.

3.3 Neural networks fits

The NNPDF approach uses a MonteCarlo strategy to perform fit of PDFs as described in Section 2.2.

This includes:

- The MonteCarlo treatment for experimental data,
- The parameterization of PDFs with Artificial Neural Networks,
- The minimization strategy based on Genetic Algorithm,

all explained in Chapter 2.

The procedure is simple and I did it for each theory above defined (in my case from theory 190 to theory 197).

As first step it was necessary for me to prepare the fit: this command take in input an "*.yml*" file, one for each theory, where all the datasets that I wish to fit are defined in the FK tables form.

After that I launched the **nnfit** program, using 200 replicas and taking as input the folder of the theory considered, called *runcard folder*, created in the *setupfit* step: this number of replicas is necessary in order to be sure that 100 replicas survive after post selection.

The post-fit program was launched with 100 replicas and has in input the runcard folder. This allowed me to finalize the PDF set, applying the post-selection criteria, which discard e.g. replicas from which minimization has made convergence [18].

As final step I produced a set of comparison plots between each fit and the reference (with $m_c = 1.51$ GeV).

In Figure 3.1 - 3.2 these plots are shown for all PDFs for theory 190, corresponding to $m_c = 1.11$ GeV.

Figure 3.1: Comparison between PDFs with $m_c = 1.11$ GeV and $m_c = 1.51$ GeV.

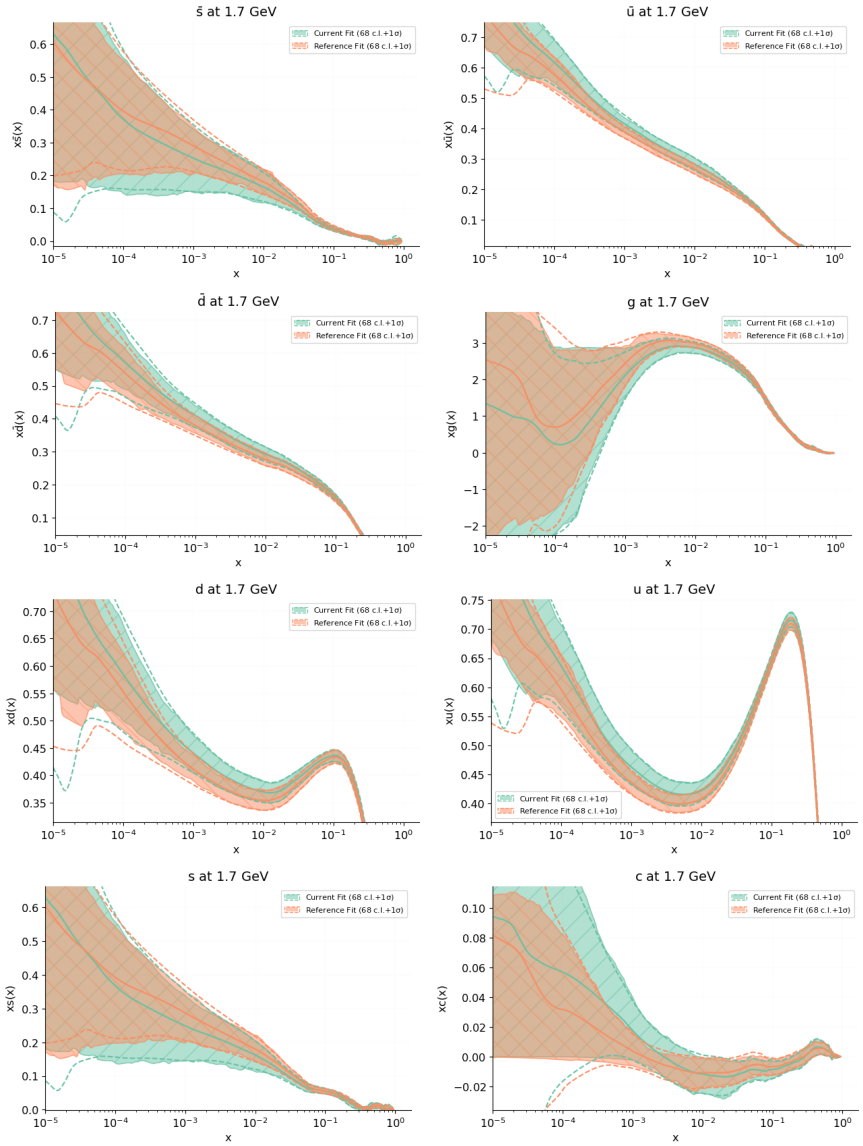
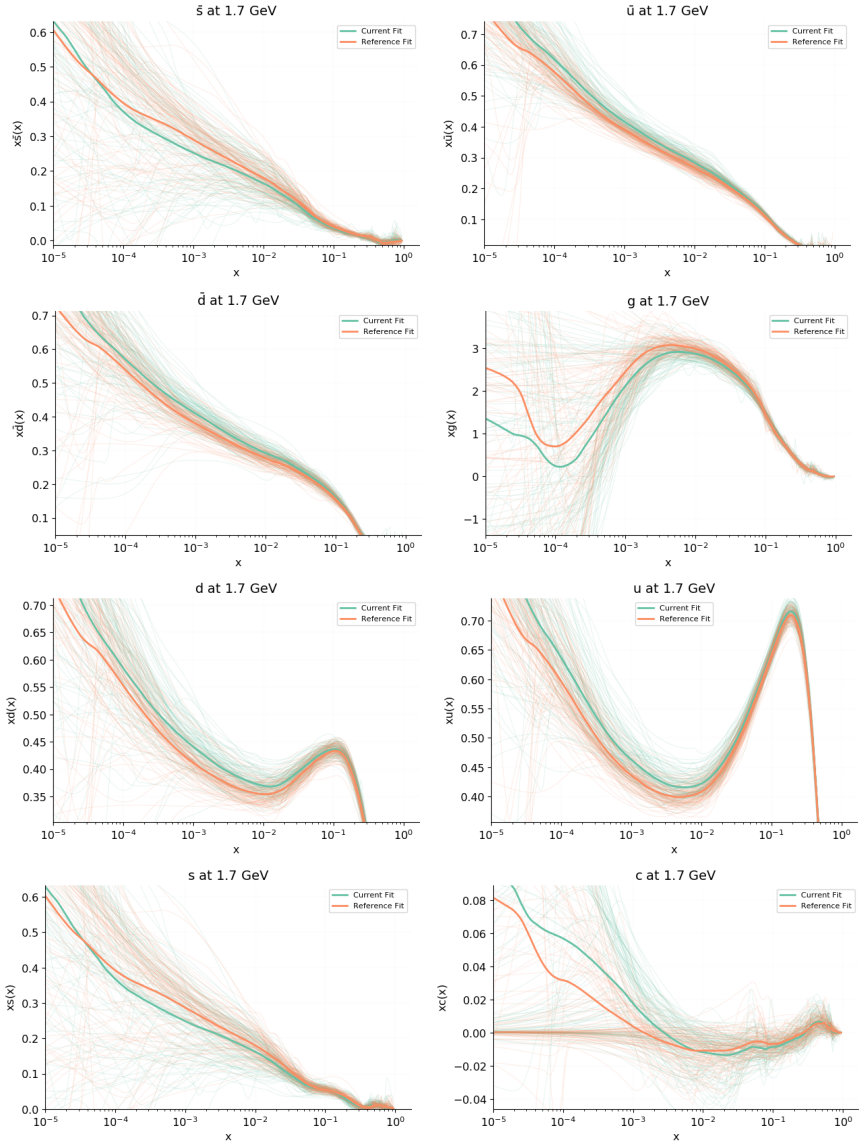


Figure 3.2: Same as 3.1 showing individual replicas.



It was not possible to achieve a fit for theories 196 and 197, corresponding to $m_c = 1.71$ GeV and 1.81 GeV, because the charm mass was above the starting scale of evolution. A way to correct this problem is under investigation.

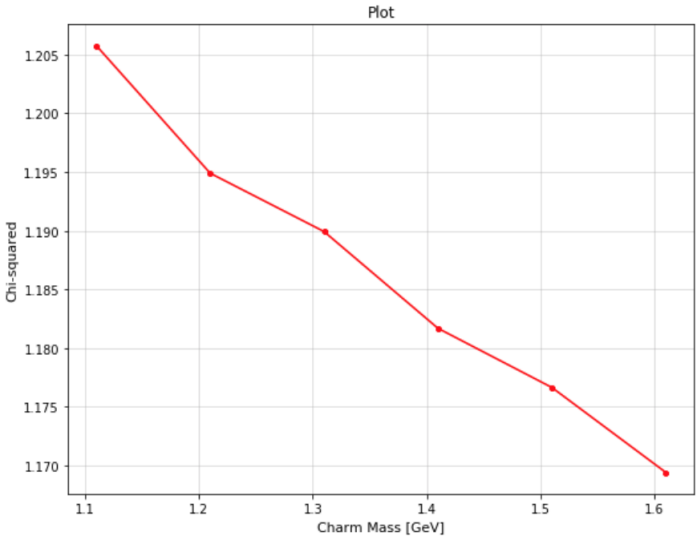
3.4 Final result

The χ^2 measure of fit quality for all the fits I performed are collected in Table 3.3. They are shown in Figure 3.3, while in Figure 3.4 a quadratic fit is shown.

Theory	Mass (Gev)	χ^2
190	1.11	1.20580
191	1.21	1.19490
192	1.31	1.18993
193	1.41	1.18169
194	1.51	1.17665
195	1.61	1.16942

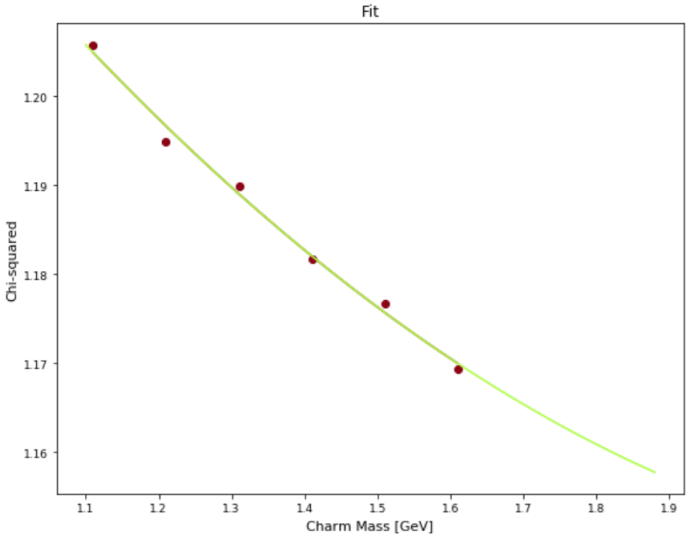
Table 3.3: Results obtained from the compare of two fits.

Figure 3.3: Plot of the resulted values.



It is clear that no minimum is found in this range. This is could have two possible explanations.

Figure 3.4: Quadratic fit to the datapoints.



It is possible that the charm mass is heavier than those supposed in [16] as pole mass. In this way the real mass value is out of the range considered in this work. In order to reach the exact value it is so necessary to consider heavier masses, in order to find the correct minimum of the χ^2 of the fits.

An other possible reason is that the datasets used are not sufficient to pin down a well defined value of the mass. A possible solution is consider new datasets, more sensitive to the charm mass than those used so far.

Ringraziamenti

Ringrazio sentitamente innanzitutto il Professor Stefano Forte e il Dottor Stefano Carrazza, per la costante disponibilità dimostratami in questi mesi e per l'infinita pazienza, e per avermi permesso di approfondire un argomento dalle grandi applicazioni pratiche, lavorando con strumenti innovativi e all'avanguardia.

Ringrazio mio fratello Gabriele, che con la sua spontaneità è sempre riuscito a tirarmi su di morale, nei momenti più bui di questi 3 anni. Ringrazio i miei genitori e i miei nonni, per avermi sostenuto e incoraggiato, ciascuno in modo diverso. Nonostante i loro tentennamenti iniziali, sono stati tenaci nel loro aiuto, tanto quanto me.

Ringrazio i miei colleghi dell'università, soprattutto per i momenti più leggeri, e per quelli che servivano a dimenticare tutti gli altri problemi. Un gigantesco grazie va a Vale e Sofia, per aver sostenuto la mia autostima e avermi supportato anche quando credevo di non farcela, standomi sempre vicino.

Ringrazio di cuore i miei compagni dell'AISF, che, nell'ultimo periodo, mentre scrivevo questa tesi, mi hanno dato un supporto inimmaginabile. E grazie per avermi fatto sentire parte di qualcosa di grande ed incredibilmente bello.

Ringrazio i miei amici, della mia o di altre bande, per aver preso un posto nel mio cuore e per farmi divertire.

Ringrazio Dio, per avermi donato la forza di arrivare fino a qui, senza arrendermi.

Bibliography

[1] R.K. Ellis, W.J. Stirling, and Webber B.R. *QCD and Collider Physics.* Cambridge: Cambridge University Press, 1996.

[2] Guy D. Coughlan, James E. Dodd, and Gripaios Ben M. *The ideas of Particle Physics. And Introduction for Scientists.* Cambridge University Press, 1984.

[3] B. Foster, R.S. Thorne, and M.G. Vincter. "Structure Functions". In: *Particle Data Group* (2018).

[4] J Pumplin et al. "New Generation of Parton Distributions with Uncertainties from Global QCD Analysis". In: *J. High Energy Phys.* 0207.012 (2002). DOI: 10.1088/1126-6708/2002/07/012. arXiv: 0201195 [hep-ph].

[5] AD Martin et al. "Uncertainties of predictions from parton distributions, 1: Experimental errors". In: *Eur. Phys. J.* C28.455-473 (2003). DOI: 10.1140/epjc/s2003-01196-2. arXiv: 0211080 [hep-ph].

[6] AD Martin et al. "Uncertainties on α S in global PDF analyses and implications for predicted hadronic cross sections". In: *Eur. Phys. J.* C63.189 (2009). DOI: 10.1140/epjc/s10052-009-1164-2. arXiv: 0905.3531 [hep-ph].

[7] H.L. Lai et al. "New parton distributions for collider physics". In: *Phys. Rev.* D82.074024 (2010). DOI: 10.1103/PhysRevD.82.074024. arXiv: 1007.2241 [hep-ph].

[8] S. Forte and G. Watt. "Progress in the Determination of the Partonic Structure of the Proton". In: *Annu. Rev. Nucl. Part. Sci.* 63.291-328 (2013). DOI: 10.1146/annurev-nucl-102212-170607. arXiv: 1301.6754 [hep-ph].

[9] J. Rojo. "Machine Learning tools for global PDF fits". In: *Conference: C18-07-31* (2018). arXiv: 1809.0439.

[10] R. D. Ball and others [NNPDF Collaboration]. "Parton distributions from high-precision collider data". In: *Eur. Phys. J.* C77.10,663 (2017). DOI: 10.1140/epjc/s10052-017-5199-5. arXiv: 1706.00428.

[11] *The role of replicas in the NNPDF approach.* URL: http://nnpdf.mi.infn.it/research/general-strategy/.

[12] *Neural Network architecture.* URL: http://n3pdf.mi.infn.it/wp-content/uploads/2019/03/SForte_Milano_032019.pdf.

[13] S. Forte. "Parton distributions at the dawn of the LHC". In: *Acta Phys.Polon.* B41.2859-2920 (2010). arXiv: 1011.5247 [hep-ph].

[14] *The cross-validation method.* URL: http://nnpdf.mi.infn.it/research/cross-validation/.

[15] Nathan Hartland. *nnpf++ data layout.*

[16] D. de Florian and others [LHC Higgs Cross Section Working Group]. "Deciphering the Nature of the Higgs Sector". In: *Handbook of LHC Higgs Cross Sections* 4 (2016), p. 849. DOI: 10.23731/CYRM-2017-002. arXiv: 1610.07922 [hep-ph].

[17] Valerio Bertone, Stefano Carrazza, and Juan Rojo. "A PDF Evolution Library with QED corrections". In: *Comput. Phys. Commun.* 185.1647-1668 (2014). DOI: 10.1016/j.cpc.2014.03.007. arXiv: 1310.1394 [hep-ph].

[18] A. Buckley et al. "LHAPDF6: parton density access in the LHC precision era". In: *Eur. Phys. J.* C75.3,132 (2015). URL: http://arxiv.org/abs/1412.7420.