Karlsruhe Institute of Technology

ETP-KA/2021-2

Impact of differential dijet cross section measurements on PDF fits at NNLO

Master thesis by

Jakob Stark

Handed in on 5 February 2021

Advisor: Priv. Doz. Dr. Klaus Rabbertz
Coadvisor: Prof. Dr. Günter Quast

Institute of Experimental Particle Physics (ETP)

Declaration of originality

I hereby declare that I have composed this paper by myself and without any assistance other than the sources given in my list of works cited. I have fully referenced any text passages directly or indirectly quoted from a source.

Karlsruhe, 05.02.2021, _____

Jakob Stark

Contents

1. Introduction

Scientist always searched for the most fundamental components of which matter is built. Ancient Greek philosophers and medieval alchemists founded their „fundamental" elements in theories, that we would consider being rather nonscientific today. Starting with the first prototype of the periodic table of elements in the 18th century and continuing to today's elementary particles in the standard model, chemists and physicists every once in a while have found smaller components of which the previous fundamental elements had been made. All ordinary matter consists of atoms, atoms consist of electrons, neutrons and protons, and protons in turn consist of quarks and gluons.

The internal proton (and neutron) structure is described by the theory of quantum chromodynamics. Unfortunately, some of the predictions of quantum chromodynamics are difficult to calculate and it happens that the proton structure is one of them. Because of that, the proton structure can up to today only be measured by experiments that is by colliding protons with other particles and analyzing the reaction products. The phenomenological description of the proton structure is encoded in the so-called parton distribution functions (PDFs). In the era of the Large Hadron Collider, especially in the upcoming high-luminosity run, it is important to have an accurate description of the proton structure, because the PDFs are needed to calculate any precise cross section prediction in proton colliders.

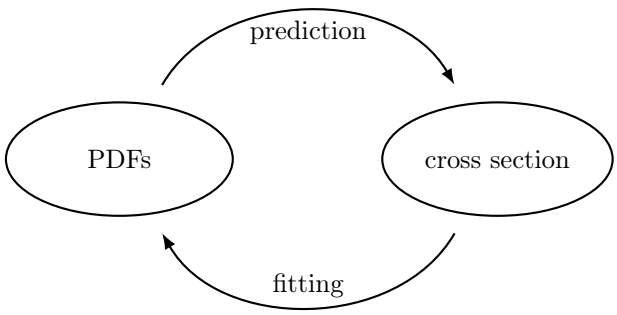


Figure 1.1: Schematic overview of PDF fitting

PDFs are determined from experimental data by basically a trial and error method. By varying the PDFs calculating cross section predictions and comparing these to the actually measured cross sections, one can try to improve the PDFs and ultimately find

the best-fitting PDFs. This procedure is called PDF fitting and may sound a bit like a chicken-and-egg problem at first glance, because unknown effects or systematic errors in the cross section measurement could misleadingly be included into the PDFs. However, by including a large number of cross section measurements from different experiments in so-called global PDF fits, proton PDFs can be extracted, that are probably close to the „true" proton structure.

In this work, the influence of a specific type of measurements – double- and triple-differential dijet cross sections – on fitted PDFs is studied. Dijet cross sections in proton-proton colliders quantify how often two directed, bundled streams of particles (called jets) emerge from a proton-proton collision. This statistically happens very often compared to other types of events and allows very precise cross section measurements. It is expected, that including jet data into PDF fits lowers the uncertainty on the gluon PDF.

Compared to previous PDF fits from the CMS collaboration with dijet data included, in this work, it is done for the first time with next-to-next-to leading order (NNLO) calculations that only became available in the last few years. NNLO calculations require a lot of time and are done on large scale computing clusters. As we will see later on in this work, NNLO cross section calculations lead to a great improvement of the predictions and decrease the theoretical uncertainties. As the cross section predictions are a key step in PDF fitting, smaller theoretical uncertainties also affect the fitted PDFs.

There is another parameter of perturbative QCD, namely the strong coupling constant α_s, that has to be determined from data. Because dijet production is dominated by QCD processes, fitting the strong coupling constant simultaneously to the PDFs is possible. In this work, such PDF plus α_s fits will also be analysed – with NNLO calculations for the first time as well.

Overall the goal of this work is not to extract a set of global PDFs to be released to the public. Rather, the impact of a specific type of cross section measurements – differential dijet data – is investigated. It will be demonstrated, that consistent fits are possible and yield much improved results, especially when taking the next-to-next-to leading order calculations into account.

2. Theoretical foundations

2.1 Quantum chromodynamics

The physical theory to predict most of the cross section measurements used here, is given by perturbative QCD. Quantum chromodynamics (QCD) is part of the standard model of particle physics, which again is the state-of-the-art theoretical framework in particle physics. QCD is a non-abelian gauge theory and describes the so-called strong interaction that is responsible for the formation and inner structure of hadrons. Introductions into QCD and quantum field theory in general can be found in standard textbooks like for example Ref. [46].

The mathematical framework to calculate QCD cross sections in high energy physics is perturbative QCD (pQCD). Any observable is expanded into powers of the renormalized strong coupling constant $\alpha_s(\mu_R^2)$. The renormalization scale μ_R is introduced by the renormalization of QCD. This absorbs divergences that occur in Feynman diagrams with loops, into measurable quantities, which again can be determined from experiment. The choice of μ_R is rather arbitrary and if the perturbative expansion was done up to infinite order, the pQCD predictions would not depend on μ_R anymore. In practice though the perturbative expansion is truncated after the first few terms and μ_R should be chosen close to the momentum transfer Q in a process to get a good approximation of the effective strength of the process [36].

The dependence of $\alpha_s(\mu_R^2)$ is given by the renormalization group equation:

$$\mu_R^2 \frac{\partial \alpha_s(\mu_R^2)}{\partial \mu_R^2} = \beta\left(\alpha_s(\mu_R^2)\right) \quad \text{with} \quad \beta(\alpha_s) = -\alpha_s^2\left(b_0 + b_1\alpha_s + b_2\alpha_s^2 + \dots\right) \quad , \tag{2.1}$$

where the coefficients b_i can be calculated again with perturbative QCD. If known at one specific scale $\mu_{R,1}^2$, the value of α_s can be „evolved" to any other scale $\mu_{R,2}^2$ by solving equation (2.1) with the desired accuracy. As the β-function in equation (2.1) is negatively signed, the value of $\alpha_s(\mu_R^2)$ decreases at higher scales μ_R and the effective coupling gets smaller. This feature, which is distinctive for QCD, is called *asymptotic freedom*. At lower scales on the other hand, the effective coupling increases until at some point the perturbative expansion does no longer converge and the whole framework of perturbative QCD breaks down. Lower scale predictions, like the formation and structure of hadrons, must be determined from experiments or calculated by some non-perturbative approach to QCD like for example lattice QCD.

The strong coupling constant α_s is determined from experiment and typically evolved to the scale of $\mu_R = M_z$, the mass of the Z-Boson. There have been many measurements of $\alpha_s(M_z)$ and an overview can be found in Ref. [36].

2.2 The proton structure

Protons are along with electrons and neutrons the main components of ordinary matter. The proton is a bound state of QCD and its properties and inner structure could in theory predicted by it. However, the energy scales, at which the proton exists, is in the region where perturbative QCD cannot be applied. Calculations involving non-perturbative approaches like for example lattice QCD are worked on [7], but up to now the proton structure is described best by the experimentally determined parton distributions functions.

The factorization theorem [20] makes it possible to calculate cross sections for processes with initial-state hadrons. Roughly speaking, the calculations are split in a way that effects (e.g. gluon emissions) below an arbitrary *factorization scale* μ_F are absorbed into the PDFs and effects happening at scales higher than μ_F are calculated in perturbative QCD. The PDFs are functions $f_i(x, \mu_F^2)$ that can be interpreted as the probability density of finding a parton i with a fraction x of the protons momentum at a scale of μ_F.

The PDFs satisfy their renormalization group equation, namely the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) equation, which relates (similar to the renormalization group equation of α_s, (2.1)) the PDFs at different factorization scales with each other. The DGLAP equation at leading order in α_s reads

$$\mu_F^2 \frac{\partial f_i(x, \mu_F^2)}{\partial \mu_F^2} = \sum_j \frac{\alpha_s(\mu_F^2)}{2\pi} \int_x^1 \frac{dz}{z} P_{ij}(z)\, f_j(x/z, \mu_F^2) \quad , \tag{2.2}$$

where the P_{ij} are the Altarelli-Parisi splitting functions that are again calculable in perturbative QCD. If given at a specific factorization scale $\mu_{F,1}^2$ the PDFs at another scale $\mu_{F,2}^2$ can be calculated by solving the DGLAP equations at the desired order in α_s.

PDFs are determined from a large variety of cross section measurements in global PDF fits, like in the MSTW [33], the CTEQ [24] or the NNPDF [22] PDF sets. Many of the public available PDFs are released through the `LHAPDF` interface [15], which provides a software library to store and evaluate PDFs as well as the uncertainties. Often they are defined at a specific starting scale μ_0 and then evolved to other factorization scales by solving the DGLAP equations (2.2) numerically.

3. The datasets

3.1 HERA I+II data

The HERA synchrotron was a proton electron (positron) collider at DESY. The data, which will later on be used in fits of PDFs and the strong coupling constant, are taken from a combined dataset of inclusive deep-inelastic scattering cross sections released by the H1 and ZEUS collaborations [4]. Most of the time HERA accelerated electrons (positrons) to a beam energy of $E_e \simeq 27.5\,\text{GeV}$ and protons to a beam energy of $E_p = 920\,\text{GeV}$. Additionally the dataset also includes cross section measurements taken with proton beam energies of $E_p = 820$, 575 and 460 GeV.

Deep-inelastic scattering (DIS) is a process, in which a lepton (here an electron or a positron) interacts with a proton at high energies. Figure 3.1 shows the two possible interactions that must be differentiated. The incoming lepton can emit either a neutral gauge boson (γ/Z) or a charged W boson. In the first kind of interaction (neutral current or short NC), the lepton is scattered but stays the same. If a W boson is exchanged (charged current or short CC), the outgoing particle is a neutrino, which cannot be seen in the detector. In both cases the exchanged gauge boson interacts with a part of the proton called a parton, that is scattered out of the proton, undergoing hadronization and leading to a jet.

In inclusive DIS we are only interested in a differential cross section depending on observables of the lepton or neutrino final state. As however a neutrino cannot be detected in general purpose particle detectors, in the charged current case the neutrino final state must be reconstructed using the hadronic remnants of the scattered proton.

3.1.1 DIS kinematics

DIS cross sections are typically measured differential in the two kinematic variables x and Q^2. Let k be the four momentum of the incoming lepton l in Figure 3.1, k' the one of the outgoing lepton or neutrino and $q = k - k'$ the one of the interacting gauge boson. The quantities x and Q^2 are then defined like

$$Q^2 = -q^2 \quad \text{and} \quad x = \frac{Q^2}{2q \cdot P} \tag{3.1}$$

where P is the four momentum of the initial proton. Q^2 is a measure for the energy scale and therefore also the spatial resolution at which the proton is probed. Depending

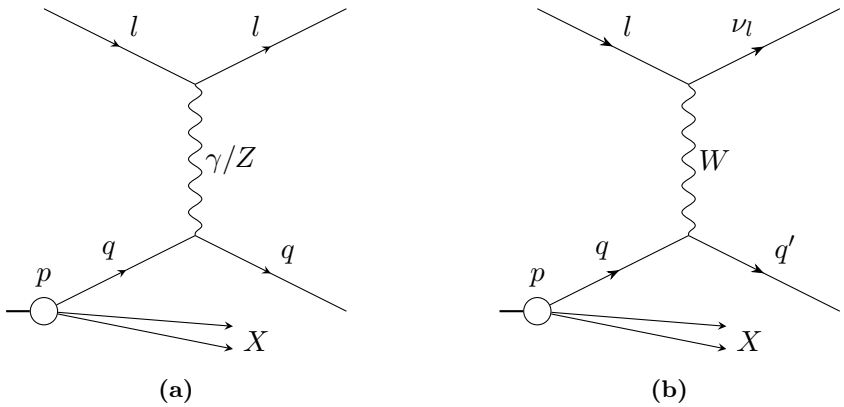


Figure 3.1: Deep-inelastic scattering with (a) neutral current interaction via a photon or Z-Boson and (b) charged current interaction via a W-Boson.

on the value of Q^2 one may „see" a different structure in the proton. This feature of the proton structure is theoretically described by the scale dependence of the structure functions (the structure functions depend on the proton PDFs, whose scale dependency is described by the DGLAP equations, see section 2.2). In a frame where the proton is fast-moving, the value x can be interpreted as the fraction of the nucleon's momentum that the interacting parton carries just before interacting with the probing gauge boson.

In case of neutral current interaction (see Figure 3.1a), the quantities Q^2 and x are approximately measured from the electron's (positron's) final state. Let θ_e be the scattering angle with respect to the proton beam, $E_e^{'}$ the scattered electron's energy and $p_{T,e}$ its transverse momentum. Then the kinematic variables are calculated with

$$Q^2 = \frac{p_{T,e}^2}{1-y_e} \quad \text{and} \quad x = \frac{Q^2}{s y_e}\,, \quad \text{where} \quad y_e = 1 - \frac{E_e^{'}(1-\cos\theta_e)}{2E_e} \quad . \tag{3.2}$$

As mentioned before, in case of charged current interaction, there is no electron or positron in the final state from which one could measure the event kinematics. Therefore the so-called „hadron method" was applied in charged current events:

$$Q^2 = \frac{p_{\mathrm{T},h}^2}{1-y_h} \quad \text{and} \quad x = \frac{Q^2}{s y_h}\,, \quad \text{where} \quad y_h = \frac{\sum_i (E_i - p_{Z,i})}{2E_e} \quad . \tag{3.3}$$

The sum is extending over the hadronic energy deposits E_i and their longitudinal momentum $p_{Z,i}$. A more detailed description of the exact measurement of the deep inelastic scattering dataset, that will be used later on, can be found in the corresponding publication [4].

3.1.2 Cross section predictions

The deep-inelastic cross section is theoretically described by a sum of so-called *structure functions* F_i. For the NC reduced cross section one gets

$$\sigma_{r,\mathrm{NC}}^{\pm} = \frac{d^2\sigma_{\mathrm{NC}}^{e^{\pm}p}}{dx\,dQ^2} \cdot \frac{Q^4 x}{2\pi\alpha^2 Y_+} = \tilde{F}_2 \mp \frac{Y_-}{Y_+} x\tilde{F}_3 - \frac{y^2}{Y_+}\tilde{F}_L \tag{3.4}$$

with the NC structure functions F_i and $Y_\pm = 1 \pm (1-y)^2$. For the charged current process an analogous expression yields

$$\sigma^{\pm}_{r,\mathrm{CC}} = \frac{d^2\sigma^{e^\pm p}_{\mathrm{CC}}}{dx\,dQ^2} \cdot \frac{2\pi x}{G_F^2}\left[\frac{M_W^2+Q^2}{M_W^2}\right]^2 = \frac{Y_+}{2}W_2^\pm \mp \frac{Y_-}{2}xW_3^\pm - \frac{y^2}{2}W_L^\pm \tag{3.5}$$

with the CC structure functions W_i.

The structure functions depend on x and on Q^2, where the latter was historically important, as deep inelastic scattering was first described with a parton model [10] that predicts scaling (Q^2-independent) structure functions. In modern QCD the structure functions are expressed in terms of the parton distribution functions:

$$F_a = \sum_i C_a^i \otimes f_a = \int_x^1 \frac{dy}{y} C_a^i(y) f(x/y, \mu_F) \tag{3.6}$$

where the coefficient functions C_a^i are calculable in perturbative QCD [36] and known up to the third order in α_s [21, 34, 43]. In the fits in Chapter 5 the `xFitter` framework is using the `QCDNUM` program to calculate cross section predictions for the HERA I+II inclusive DIS data accordingly [12].

3.2 CMS dijet data

There are several dijet datasets measured with CMS that will be used in PDF and α_s fitting in Chapter 5. The CMS detector is along with ATLAS one of the two large multi-purpose particle detectors mounted at the Large Hadron Collider (LHC) at CERN. The 8 TeV triple-differential dijet ($p+p \to \mathrm{jet}+\mathrm{jet}+x$) cross section measurement, that will mainly be used for the PDF fits including jet data in Chapter 5, is made from a data sample of 19.71 fb^{-1} of integrated luminosity at a center of mass energy of $\sqrt{s} = 8$ TeV collected by CMS in the 2012 proton-proton collision run period of the LHC [40].

The other dijet cross section measurement that will be used in Chapter 5 is a double-differential dijet cross section measurement at $\sqrt{s} = 7$ TeV from CMS during the 2011 run period of the LHC with an integrated luminosity of 5 fb^{-1} [18]. This publication also includes an inclusive jet cross section measurements, that has already been used in PDF and α_s fits [27]. In this work only the double-differential dijet measurement is used.

3.2.1 Jet reconstruction

A hadronic jet is typically understood to be a bundled stream of particles detected in the same region in the detector. However, to properly define a jet, a deterministic algorithm must be used to cluster single reconstructed particles into a jet. In order to be calculable in perturbative QCD this jet definition must satisfy the properties of infrared and collinear safety. An infrared- and collinear-safe jet definition has to be invariant under addition of a soft ($p \to 0$) particle and under splitting of a particle ($p \to p_1 + p_2$ with $p_1 \| p_2$, where p denotes the particle's 4-momentum). The de-facto

standard jet definition at CMS is the *anti-k_t* algorithm [16] out of the class of sequential recombination algorithms.

The anti-k_t algorithm operates on a set of input objects (e.g. particles reconstructed by the detector). A distance measure between two objects d_{ij} and a termination distance d_{iB} (also called „beam distance") are defined like

$$\begin{aligned} d_{ij} &= \min\left(p_{\mathrm{T},i}^{2k}, p_{\mathrm{T},j}^{2k}\right) \frac{\Delta R_{ij}^2}{R^2} \quad \text{and} \\ d_{iB} &= p_{\mathrm{T},i}^{2k}, \end{aligned} \tag{3.7}$$

where $p_{\mathrm{T},i}$ is the transverse momentum of the i-th object in the list and ΔR_{ij} is the distance in y (rapidity) – ϕ (azimuth) space. The algorithm repeatedly searches for the pair of objects with minimum distance d_{ij} and merges these two objects to a single one by adding their momentum four-vectors. If an object's termination distance d_{iB} is smaller than the minimum distance d_{ij} of any two objects, it is declared to be a jet and removed from the list of active objects. The algorithm terminates if ultimately all input objects have been merged into jets and are removed from the list.

The parameters k and R can be used to tune the algorithms and influence the shape and size of the jets. For the anti-k_t algorithm, k is set to $k = -1$ preferring merges with high-p_{T} objects, which results in rather circular jet shapes in the y-ϕ space. R influences the jets' sizes and typical values range from $R = 0.4$ to $R = 0.7$ or even $R = 1.0$.

Both the 7 TeV and 8 TeV dijet cross section measurements [18, 40] used in Chapter 5 use the anti-k_t algorithm with $R = 0.7$.

3.2.2 Dijet kinematics

An event recorded by the detector is considered to be a dijet event, if there are at least two jets satisfying some analysis-dependent requirements. The jets are ordered by their transverse momentum and the two jets with the highest p_{T} are called the leading and second leading jet. Dijet observables are then defined in terms of kinematic properties of these two jets.

The triple-differential cross section measured in Ref. [40], which will be used in the 8 TeV fits in Chapter 5, is defined as

$$\frac{d^3\sigma}{dp_{\mathrm{T,avg}}\, dy^*\, dy_b} \tag{3.8}$$

binned in the jet-related quantities

$$\begin{aligned} p_{\mathrm{T,avg}} &= \frac{1}{2}(p_{\mathrm{T},1} + p_{\mathrm{T},2}) \\ y^* &= \frac{1}{2}|y_1 - y_2| \quad \text{„ystar"} \\ y_b &= \frac{1}{2}|y_1 + y_2| \quad \text{„yboost"} \quad , \end{aligned} \tag{3.9}$$

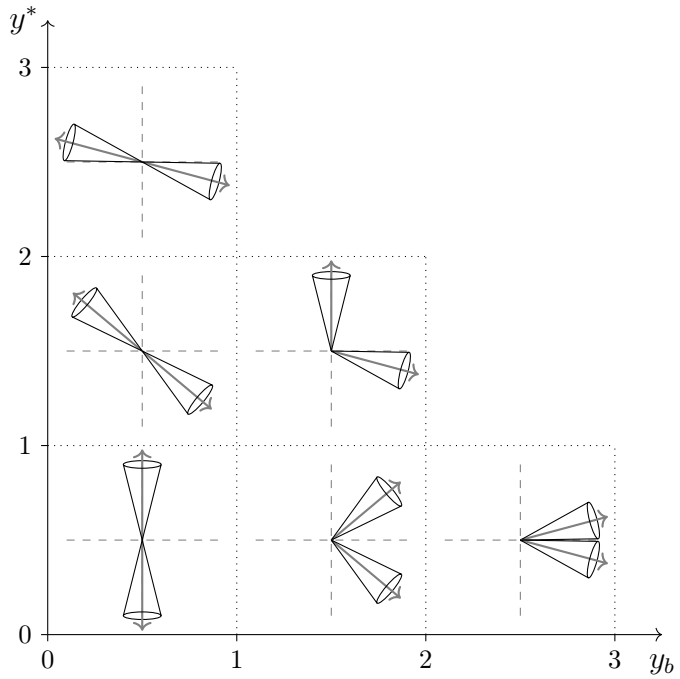


Figure 3.2: Typical dijet layouts in different rapidity bins. Larger values of y_b correspond to dijet events, where the parton-parton center-of-mass frame is more boosted along the beam line. The boost-invariant y^* corresponds to the jet rapidities in the parton-parton center-of-mass frame.

where $p_{\mathrm{T},1(2)}$ and $y_{1(2)}$ are the transverse momentums and rapidities of the leading and second leading jet respectively.

Figure 3.2 shows the six rapidity bins chosen in Ref. [40] and typical spatial configurations of the two leading jets in these bins. The y^* value of a dijet event is invariant under boosts in beam direction and can be interpreted as the absolute rapidity of the jets in the interacting partons' center-of-mass frame. The y_b value on the other side is a measure of how much the center-of-mass system of the initial partons is boosted with respect to the laboratory frame.

In the high-y_b region, events occur, where typically one of the interacting partons carries a larger momentum fraction than the other one. As in low x regions the gluon PDF dominates and in the high x regions the valence quarks dominate, in the high-y_b bins, most of the events originate from gluon-quark reaction. In the high-y^* region, the contribution of quark-quark reactions increases, especially in the high-p_{T} bins. This is because at high-y^* a significant momentum fraction of both the colliding partons is needed in order to produce high-p_{T} jets. Thus again, the valence quark distributions begin to dominate in this region. Details on this can be found in Ref. [40].

The fact, that these sub-processes contribute differently to the dijet cross sections dependent on the rapidity bin, makes the predictions sensitive to the PDFs. The ambition is, that correlations between fitted PDFs may be lowered with this type of rapidity

binning.

In the 7 TeV measurement in Ref. [18], the measured double-differential cross section is defined as

$$\frac{d^2\sigma}{dm_{12}\, d|y|_{\max}} \quad \text{with} \quad |y|_{\max} = \max(|y_1|, |y_2|) \tag{3.10}$$

with only a one-dimensional binning in rapidity. m_{12} is the invariant mass of the two leading jets and is called the dijet mass

$$m_{12} = \sqrt{(p_1^\mu + p_2^\mu)^2} \tag{3.11}$$

with the leading and second leading jets' four-momentums p_1^μ and p_2^μ.

More details on the dijet cross section measurements can be found in their corresponding publications. Because both of the measurements are done with CMS at LHC, there might be correlations in the systematic uncertainties of these datasets. Care must be taken, if PDF or α_s fits are performed with more than one of these datasets included at once.

3.2.3 Cross section predictions

The cross section predictions for dijet production in proton-proton colliders can be split into the three stages shown in Figure 3.3. The hard process (shown in the middle) is the only part that can be calculated in perturbative QCD. Two partons (one from each proton) are scattered and produce some other particles. This calculation can be performed at any order in α_s that can be handled with the available computing code and resources. The initial state, the two colliding protons, is not in the scale domain where perturbative QCD is applicable and the PDFs are used to describe it. The splitting of the initial state and the hard process is described by collinear factorization: The production cross section of any observable X ($p + p \to X$) can be written like

$$\sigma_{pp\to X} = \sum_{n=0}^{\infty} \alpha_s^n(\mu_R^2) \sum_{i,j} \int dx_1 dx_2 f_i(x_1, \mu_F^2) f_j(x_2, \mu_F^2) \hat{\sigma}^{(n)}_{ij\to X}(x_1, x_2, \mu_R^2, \mu_F^2) \quad , \tag{3.12}$$

that is, as a convolution of the proton PDFs $f_i(x, \mu_F^2)$ with the hard cross section matrix element $\hat{\sigma}_{ij\to X}$. As already pointed out, the PDFs are not calculable by perturbative QCD and must be determined from cross sections measurements.

In principle, the X in equation (3.12) can be any observable, as long as it satisfies the requirements of infrared and collinear safety. In the case of dijet (double as well as triple-differential) cross sections these requirements are satisfied by using the infrared and collinear safe anti-k_t jet algorithm (see above).

If the hard partons, emerging from the hard scattering process, are directly used as input to the jet clustering algorithms, the effects of showering and hadronization are neglected as well as effects coming from the underlying event. There are Monte Carlo event generators, like `PYTHIA` [41], `HERWIG` [9] or `SHERPA` [11] that offer a wide range of

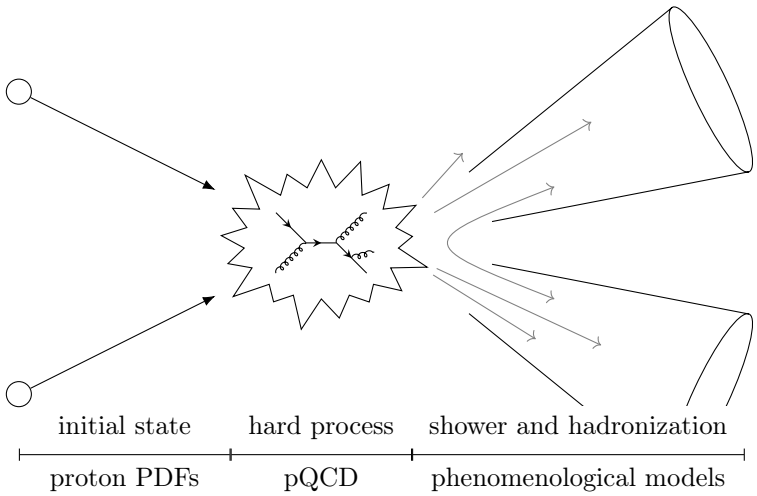


Figure 3.3: Dijet production process overview

models and algorithms to estimate these non-perturbative (NP) corrections to the final state. Typically, one of these generators is used to extract the NP effects as correction factors to each final state observable bin. These factors can then be applied to the cross sections calculated with equation (3.12).

Fitting PDFs to measured dijet cross section basically is done by varying the PDFs in equation (3.12) and finding those PDFs, that yield the best matching cross section prediction to the measurement. Dependent on the PDF parametrization, lots of variations are possible and it is crucial, that the cross section predictions can be calculated on a timescale of seconds. Unfortunately, the calculation in equation (3.12) needs a lot of time (especially at NLO or even NNLO) and there is no practical way to recalculate the cross sections for each variation of the PDFs fast enough to be useful in a PDF fit. To overcome this speed problem, the cross sections can be pre-calculated and stored in an interpolation grid. The calculation of the cross section in (3.12) then is simplified to just a summation over a number of interpolation nodes. Details on this procedure are pointed out in section 4.1.

The actual calculations of most of the cross section predictions for the fits in chapter 5 are done by the parton level Monte carlo Generator `NNLOJet` [23]. The interpolation tables are produced and evaluated with the `FastNLO` framework [14, 31].

4. Fitting procedure

There are several mathematical tools and algorithms used in proton PDF fitting. The following chapter gives an overview on the tools used in the fits in Chapter 5.

4.1 Fast interpolation grids

In order to fit the PDFs $f_i(x, \mu_F^2)$ in equation (3.12), a fast evaluation of the convolution is needed, as in each fitting iteration the PDFs may change. However, it is very challenging to calculate the $\hat{\sigma}_{ij\to X}$, especially in orders of α_s higher than the leading order, and a lot of computing time is needed to do so. To speed up the PDF fits, tools like `FastNLO` and `APPLgrid` were developed [17,31]. The basic idea is to store the $\hat{\sigma}$ function values in an interpolation grid which is computational expensive, but only a single time. To evaluate that interpolation grid, only a summation over the interpolation grid nodes is needed. There are small differences between `FastNLO` and `APPLgrid` and as `FastNLO` grids will be used later on, the `FastNLO` approach will be highlighted in the following.

The cross section in equation (3.12) is numerically calculated by generating random events that produce X and summing up their weights. The renormalization and factorization scales are typically chosen separately for each event, depending on event kinematic variables like Q^2 in inclusive deep-inelastic scattering or the transverse momentum p_T of the leading jet in jet production. Often there is no single „obvious" scale choice and different scale definitions could be used. In `FastNLO` the scales are thus considered to be functions of two internal quantities that we will call s_1 and s_2 in the following. So with $\mu_R = \mu_R(s_1, s_2)$ and $\mu_F = \mu_F(s_1, s_2)$ equation (3.12) yields

$$\begin{aligned}\sigma_{pp\to X} = \sum_{n,i,j} \int dx_1 dx_2 \; \alpha_s^n \left(\mu_R^2(s_1,s_2)\right) \; f_i\left(x_1, \mu_F^2(s_1,s_2)\right) \; f_j\left(x_2, \mu_F^2(s_1,s_2)\right) \\ \times \hat{\sigma}_{ij\to X}^{(n)}\left(x_1, x_2, \mu_R^2(s_1,s_2), \mu_F^2(s_1,s_2)\right) \quad .\end{aligned} \tag{4.1}$$

The $\hat{\sigma}_{ij\to X}^{(n)}$ can be expanded in powers r of $\log(\mu_F^2/\mu_c^2)$ and f of $\log(\mu_R^2/\mu_c^2)$ with some constant[1] μ_c^2 like

$$\hat{\sigma}_{ij\to X}^{(n)} = \sum_{r+f\leq n-n_\mathrm{LO}} \log^r(\mu_R^2/\mu_c^2) \log^f(\mu_F^2/\mu_c^2) \; \hat{\sigma}_{ij\to X}^{(n,r,f)} \quad , \tag{4.2}$$

[1] The standard choice in `FastNLO` is $\mu_c = 1\,\mathrm{GeV}$ in proton-proton collision tables

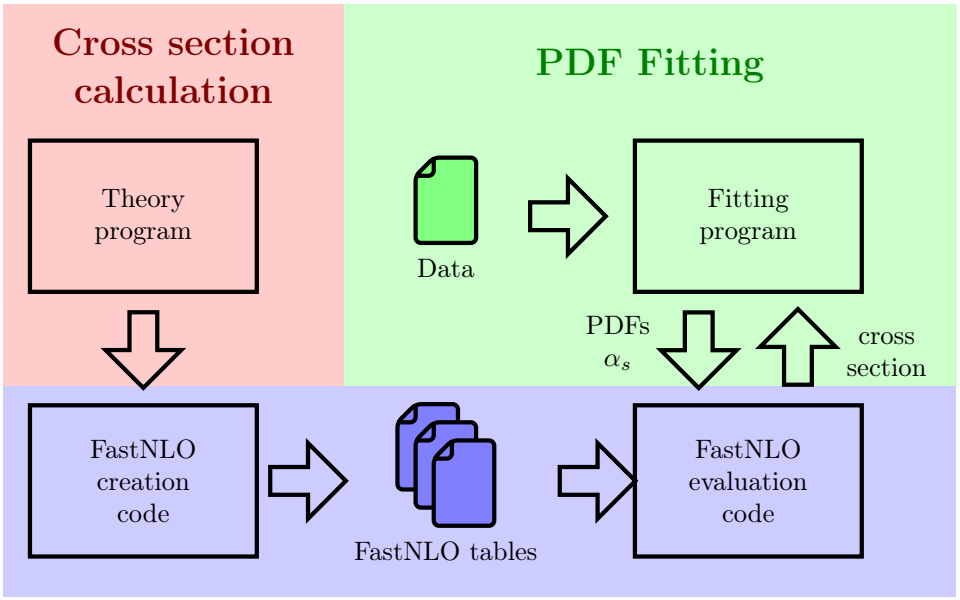


Figure 4.1: Schematic overview of a typical workflow involving `FastNLO`. The table creation needs a large amount of computing time and is done on powerful computing clusters, whereas the resulting tables can be stored and evaluated on a standard desktop PC. Fitting is done by a fitting program (e.g. `xFitter`) which is interfacing `FastNLO` to evaluate the tables and get cross section predictions for different PDF and α_s values.

which allows to factor the scale dependency out of the $\hat{\sigma}$ functions. In a next step the scale functions $\mu_{R/F}(s_1, s_2)$ and the PDFs $f_i(x)$ are replace by a set of interpolation functions. Possible functions could be for example triangular functions for linear interpolation or Lagrange polynomials for polynomial interpolation.

$$\begin{aligned} f_i(x_1, \mu_F) f_j(x_2, \mu_F) &= \sum_{\alpha,\beta} f_i(x_\alpha, \mu_F) f_j(x_\beta, \mu_F) E^\alpha(x_1) E^\beta(x_2) \\ \mu_{R/F}(s_1, s_2) &= \sum_{\delta,\varepsilon} \mu_{R/F}(s_\delta, s_\varepsilon) E^\delta(s_1) E^\varepsilon(s_2) \quad , \end{aligned} \tag{4.3}$$

where the $x_{\alpha/\beta}$ and the $s_{\delta/\varepsilon}$ are the interpolation nodes for the $x_{1/2}$ and $s_{1/2}$ values. This allows to separate the PDF and α_s evaluation from the folding like

$$\begin{aligned} \sigma_{pp\to X} = \sum_{n,i,j,r,f,\alpha,\beta,\delta,\varepsilon} & \alpha_s^n \left(\mu_R^2(s_\delta, s_\varepsilon)\right) f_i\left(x_\alpha, \mu_F^2(s_\delta, s_\varepsilon)\right) f_j\left(x_\beta, \mu_F^2(s_\delta, s_\varepsilon)\right) \\ & \log^r\left(\mu_R^2(s_\delta, s_\varepsilon)/\mu_c^2\right) \log^f\left(\mu_F^2(s_\delta, s_\varepsilon)/\mu_c^2\right) \times \tilde{\sigma}_{ij\to X}^{nrf\alpha\beta\delta\varepsilon} \quad , \end{aligned} \tag{4.4}$$

where

$$\tilde{\sigma}_{ij\to X}^{nrf\alpha\beta\delta\varepsilon} = \int dx_1 dx_2 \; E^\alpha(x_1) E^\beta(x_2) E^\delta(s_1) E^\varepsilon(s_2) \; \hat{\sigma}_{ij\to X}^{(n,r,f)}(x_1, x_2, s_1, s_2) \quad . \tag{4.5}$$

The computational expensive part (4.5) can be pre-calculated and then stored in a `FastNLO` interpolation table. To evaluate this table, only the contractions in equation

(4.4) have to be calculated, which allows fast switching of the PDFs and the α_s value, as needed for example in a PDF fit. Also the scale functions $\mu_{R/F}(s_1, s_2)$, as introduced previously, can be left undefined until the table is evaluated. The concept of factorizing out the scale dependency of the integral (4.5) by the expansion in (4.2) has been introduced in `FastNLO` version 2 allowing for so-called *flexible-scale* tables [14].

The `FastNLO` framework implements a toolkit to perform creation and evaluation of the interpolation tables. It interfaces to theory programs like for example `NLOJet++` or `NNLOJET` [23] that provide the code to integrate the integral in equation (4.5) for some observable X. This calculation is (dependent on the exact process and order n of α_s) computational very expensive and is performed on large computing clusters.

The evaluation of the grids can also be done with `FastNLO` which provides code to read the tables and compute cross sections given PDFs and α_s. To fit PDFs and/or the strong coupling constant α_s to data with `FastNLO`, one needs a pre-calculated interpolation grid matching the data scenario like observable binning and selection criteria. The fitting program than interfaces to `FastNLO`s evaluation code, provides PDFs and α_s at each fitting iteration, gets the cross section predictions and may compare them to the data to determine the next iteration step.

4.2 Minimization algorithms

Given a mathematical model for a physical process, it is a common task to fit free parameters in this model to measured data. We will later fit a model of the proton structure to measured cross section data. Such a fit is performed by first defining a cost function, which returns smaller values the better the model predictions match the data, and then minimizing this function with respect to the model parameters. The global minimization of an arbitrary function is a challenging problem and there are lots of different minimization algorithms for various problem classes.

The *MINUIT* minimization framework, implements several algorithms to minimize user defined functions. One of them is the MIGRAD algorithm. It performs a line search in direction of the gradient and updates the covariance matrix (inverse of the Hesse matrix) with each step. A detailed description of the algorithm can be found in [26]. As a byproduct to the found minimum it produces a quite accurate approximation of the covariance matrix at the minimum, which can be used for error analysis [25]. As with most minimization algorithms that perform some kind of gradient descent, MIGRAD may easily get stuck in a local minimum in parameter space. Also, it depends on an accurate estimation of the gradient of its target function. To address these shortcomings it is important to carefully analyze the minimization problem and verify the result by choosing different start values in the minimization process.

Another minimizer implemented in *MINUIT* is called SIMPLEX. It executes the minimization method of Nelder and Mead [35] that is based on an n-dimensional simplex. (A simplex is a geometric object defined by $n+1$ points in n dimensions, like e.g. a triangle in 2 dimensions). In each iteration the worst point of the simplex is replaced by a better one (in terms of function value), which is obtained by a scaled reflection at the

plane spanned by the other n points. Compared to MIGRAD the SIMPLEX algorithm does not need an estimate of the function gradient and therefore is - in some cases - more robust upon numerical errors. A downside is the very unreliable estimation of the covariance matrix. If parameter error estimations are needed, other methods must be used to study the detailed form of the function at its minimum.

In the fits in Chapter 5, unless noted otherwise, the MIGRAD minimizer is used. In some fits (especially those including α_s as free parameter) the gradient estimation of MIGRAD failed to accurately determine the derivation in direction of the α_s axis. This can lead to large step sizes in α_s direction and ultimately into a failure of the cross section calculation for too big or too small values of α_s. The best approach to solve this numerical problem was found to set limits to α_s. After that constrained fit converges the limits are released and the minimization is run again.

4.3 Goodness of fit estimator

The χ^2 function, which will be used as a goodness-of-fit estimator in the PDF fits later on, is a function that depends on the model parameters and the measured data points. It quantifies the agreement between model prediction and measurement, and will then be minimized with respect to the model parameters in the fit. If the χ^2 function is defined properly it can serve several other purposes in addition to fitting the best model parameters:

- The shape of the χ^2 function around its minimum provides information about the uncertainties and correlations of the fitted parameters.
- The value of the χ^2 function at its minimum may be used as a statistical test, to quantify how good the fitted model agrees with the measured data.

In the following, the goodness-of-fit estimator will be defined that is used in the fits later on.

Assuming normal distributed measurements $\boldsymbol{\mu}$, the joint probability distribution function reads

$$f(\boldsymbol{\mu};\boldsymbol{\theta}) = (2\pi)^{-\frac{n_\mu}{2}} \det(C)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\boldsymbol{\mu}-\boldsymbol{m}(\boldsymbol{\theta}))^T C^{-1}(\boldsymbol{\mu}-\boldsymbol{m}(\boldsymbol{\theta}))\right) \tag{4.6}$$

where $\boldsymbol{m}$ are the theory cross section predictions depending on the model parameters $\boldsymbol{\theta}$, n_μ is the number of measurements. C is the covariance matrix, that includes all the correlated and uncorrelated statistical uncertainties as well as uncorrelated systematical uncertainties (see definition below in (4.8)).

To include correlated systematical uncertainties, a nuisance parameter b_α for each such uncertainty source is introduced. The relative impacts γ_i^α of the correlated error source α on the i-th data point are stored in the datasets. To get the absolute impacts these values γ_i^α must be scaled with the actual value m_i to get the absolute correlated systematical errors Γ_i^α. The nuisance parameters are not known, but they are constructed in a way that ideally they are normally distributed with mean 0 and variance

1. To integrate this knowledge into (4.6) a Gaussian prior distribution is added

$$f(\boldsymbol{\mu};\boldsymbol{\theta},b_\alpha) = (2\pi)^{-\frac{n_\mu+n_b}{2}}\det(C)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\boldsymbol{\mu}-\boldsymbol{\Gamma}^\alpha b_\alpha-\boldsymbol{m}(\boldsymbol{\theta}))^T C^{-1}(\boldsymbol{\mu}-\boldsymbol{\Gamma}^\alpha b_\alpha-\boldsymbol{m}(\boldsymbol{\theta}))\right)\exp\left(-\frac{\sum_\alpha b_\alpha^2}{2}\right) \quad . \quad (4.7)$$

The covariance matrix C is given by

$$C_{ij} = \rho_{ij}\Delta_{i,\text{stat}}\Delta_{j,\text{stat}} + \delta_{ij}\Delta_{i,\text{syst}}\Delta_{j,\text{syst}} \quad (4.8)$$

with the Kronecker delta δ_{ij}, the statistical correlation matrix ρ and

$$\begin{aligned}\Delta_{i,\text{stat}} &= \delta_{i,\text{stat}}\sqrt{\mu_i m_i \exp(-\gamma_i^\alpha b_\alpha)}\\ \Delta_{i,\text{syst}} &= \delta_{i,\text{syst}} m_i\\ \Gamma_i^\alpha &= \gamma_i^\alpha m_i\end{aligned}$$

The correlation matrix ρ and the relative statistical $\boldsymbol{\delta}_{\text{stat}}$ and uncorrelated systematic uncertainties $\boldsymbol{\delta}_{\text{syst}}$ are provided with the datasets. As they are relative uncertainties, they must be scaled accordingly. The systematic uncertainties are scaled with the theory predicted value m_i, the statistical ones with the geometric mean of the measured data point and the predicted value. An additional correction factor $\exp(-\gamma_i^\alpha b_\alpha)$ accounts for the shift that the correlated systematic uncertainty sources impose on the predicted value m_i.

The reason why one does not just scale every uncertainty with the measured data can be understood by recalling the statistical relationship between the model and the measured data in the freqentist's point of view. While the prediction $\boldsymbol{m}(\boldsymbol{\theta})$ is a random variable, the measured data $\boldsymbol{\mu}$ is a sample drawn from this random variable. If the uncertainties are scaled with the measured data, and a data point happens to be an outlier, this can lead to a bias of the corresponding uncertainty. The uncertainty scaling settings used in this work have been used in the fits in Ref. [40] and are the `xFitter` standard settings except from the exponential correction factor.

A χ^2 is now constructed as the logarithm of the likelihood ratio between the null model $\boldsymbol{m} = \boldsymbol{m}(\boldsymbol{\theta})$ and the alternative model which is the best fitting model or *saturated model* ($\boldsymbol{m} = \boldsymbol{\mu}$ and $b_\alpha = 0$) if no constraints are imposed on the m_i parameters.

$$\begin{aligned}\chi^2 =& -2\log\frac{f(\boldsymbol{\mu};\boldsymbol{\theta},b_\alpha)}{f(\boldsymbol{\mu};\boldsymbol{\mu})}\\ =& (\boldsymbol{\mu}-\boldsymbol{\Gamma}^\alpha b_\alpha-\boldsymbol{m}(\boldsymbol{\theta}))^T C^{-1}(\boldsymbol{\mu}-\boldsymbol{\Gamma}^\alpha b_\alpha-\boldsymbol{m}(\boldsymbol{\theta})) + \sum_\alpha b_\alpha^2 + \log\frac{|C(\boldsymbol{m}(\boldsymbol{\theta}),b_\alpha)|}{|C(\boldsymbol{\mu},0)|}\\ \approx& (\boldsymbol{\mu}-\boldsymbol{\Gamma}^\alpha b_\alpha-\boldsymbol{m}(\boldsymbol{\theta}))^T C^{-1}(\boldsymbol{\mu}-\boldsymbol{\Gamma}^\alpha b_\alpha-\boldsymbol{m}(\boldsymbol{\theta}))\\ &+\sum_\alpha b_\alpha^2 + \sum_i \log\frac{\Delta_{i,\text{stat}}^2+\Delta_{i,\text{syst}}^2}{\mu_i^2(\delta_{i,\text{stat}}^2+\delta_{i,\text{syst}}^2)}\end{aligned} \quad (4.9)$$

neglecting the non-diagonal elements of C in the last step.

As mentioned before the $\chi^2(\theta)$ can be minimized (for example with one of the minimization algorithms from above) to find the best fitting model parameters. However as one could – in theory – fit an arbitrary model to the data, it is often desirable to quantify how well the fitted model describes the data in general. The following explains, how this information can be extracted from the χ^2 value at its minimum.

The alternative model has a maximum amount of freedom, and is parameterised by $\dim_{\text{alt}} = n_\mu + n_b$ parameters whereas the null model has dimension $\dim_0 = n_\theta + n_b$, the number of parameters θ plus the number of nuisance parameters. According to Wilks' theorem [44] the minimum of this likelihood ratio is χ^2 distributed with $\dim_{\text{alt}} - \dim_0 = n_\mu - n_\theta$ degrees of freedom in the large sample limit. Therefore the observed χ^2 value in the minimum can be used as a test statistic to quantify the agreement between the fitted model and the measured data. The expectation value of χ^2 in the minimum is $\text{ndf} = n_\mu - n_\theta$.

The shape of the χ^2 function around its minimum can be used to determine confidence intervals for the parameters $\boldsymbol{\theta}$. By varying one parameter θ_i around the minimum and each time minimizing the χ^2 function with respect to all the other θ parameters, one finds the limits at which this *profiled* χ^2 exceeds a difference of 1 to the minimum. However as this procedure is computational expensive, especially if the evaluation of χ^2 is non-trivial, the χ^2 function is often approximated by its second derivative (called the Hessian matrix) around its minimum, which then allows for an analytic approach as an alternative to the profiling method. Figure 4.2 shows an illustration of the profiling procedure in two dimensions and how the corresponding confidence intervals are constructed.

In PDF fits one is not so much interested in confidence intervals of the parameters θ_i, but rather in confidence intervals of the PDFs itself. Some methods to estimate those are described in Section 4.5.1 that treats experimental uncertainties on the fitted PDFs.

4.4 PDF parametrization

As PDFs are not (yet) calculable and must be determined from data, the problem of parametrization arises. A function that has to be determined from data must be dependent of some discrete values $\boldsymbol{\theta}$ that than can be fitted to data. There are different approaches here. The NNPDF group [22] for example uses a neural network with lots of free parameters to represent the PDFs. The trick is to find a parametrization, that is flexible enough to approximate the „real" unknown function, but on the other hand has not too much freedom. Too many parameters, especially if they are correlated, increase the difficulty to find an optimal set of parameters as the parameter space dimension increases with each parameter.

As the PDFs' dependency of the factorization scale μ_F^2 is given by the DGLAP equations, a parametrization at a fixed starting scale $\mu_{F_0}^2$ is sufficient to define the PDFs up to an arbitrary scale. Here a starting scale of $\mu_{F_0}^2 = 1.9\,\text{GeV}^2$ will be used which is just below the used charm quark mass. At this starting scale five independent

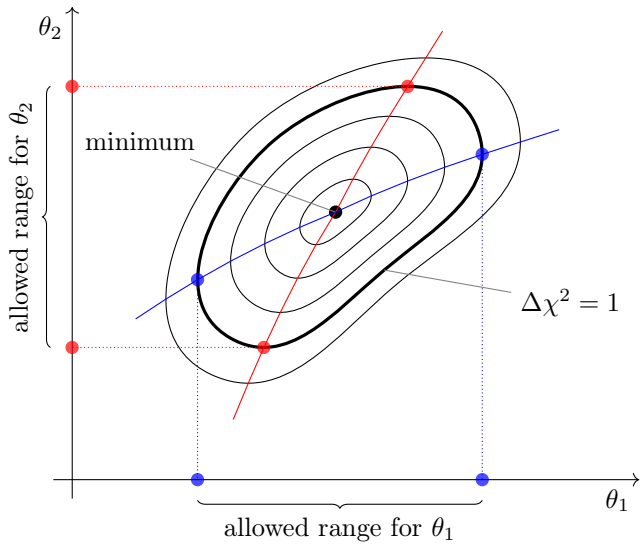


Figure 4.2: Profiling in 2D: The blue θ_1 profiling line is constructed, by minimizing χ^2 with respect to θ_2 for different fixed values of θ_1, and connecting the corresponding points in parameter space. The maximal allowed range (confidence interval) for θ_1 can be found by taking the θ_1 values at which $\Delta\chi^2$ exceeds 1. To get a confidence interval for the other parameter θ_2 one repeats the procedure with switched roles and gets the red interval.

linear combinations of PDFs will be parameterized, namely the valence up and down quark distributions xu_v and xd_v, the gluon PDF xg and the up and down type anti sea quark distributions $x\bar{U} = x\bar{u}$ and $x\bar{D} = (1 + f_s)x\bar{d}$. The strange quark PDF is modeled as f_s times the sea down quark PDF with a fixed value of $f_s = 0.4$. The charm and bottom quarks are dynamically produced at higher scales through solving the DGLAP equations with a general-mass variable-flavour-number scheme (GM-VFNS) [42].

4.4.1 HERAPDF-like parametrization

The parametrizations that will be used later on in the fits follow the HERAPDF style of parametrization [4]. For each of the PDF linear combinations listed above, there are three base parameters: A normalization factor A, the power B describing the PDF in the limit $x \to 0$ and the power C for the limit $x \to 1$. This yields

$$\begin{aligned} xg(x) &= A_g x^{B_g}(1-x)^{C_g} \cdot k_g(x) \\ xu_v(x) &= A_{u_v} x^{B_{u_v}}(1-x)^{C_{u_v}} \cdot k_{u_v}(x) \\ xd_v(x) &= A_{d_v} x^{B_{d_v}}(1-x)^{C_{d_v}} \cdot k_{d_v}(x) \\ x\bar{U}(x) &= A_{\bar{U}} x^{B_{\bar{U}}}(1-x)^{C_{\bar{U}}} \cdot k_{\bar{U}}(x) \\ x\bar{D}(x) &= A_{\bar{D}} x^{B_{\bar{D}}}(1-x)^{C_{\bar{D}}} \cdot k_{\bar{D}}(x) \quad . \end{aligned} \tag{4.10}$$

Not all of the parameters are independent though. The normalization parameters A_g, A_{u_v} and A_{d_v} are fixed by QCD number and momentum sum rules [33]. To ensure that

$x\bar{u} - x\bar{d} \to 0$ in the limit $x \to 0$, the further constraints $A_{\bar{U}} = A_{\bar{D}}(1 - f_s)$ and $B_{\bar{U}} = B_{\bar{D}}$ are imposed.

With $k_i(x) \equiv 1$ the parametrization (4.10) thus has $n_\theta = 10$ free parameters. The parametrization is, however, not flexible enough to describe the data and additional freedom has to be added by including free parameters into the k_i functions. The MSTW group for example used functions of the form

$$k_i(x) = 1 + \epsilon_i \sqrt{x} + \gamma_i x \tag{4.11}$$

with the free parameters ϵ_i and γ_i [33]. Here however, the HERAPDF approach will be used: The functions k_i are given by

$$k_i(x) = 1 + D_i x + E_i x^2 \tag{4.12}$$

with free parameters D_i and E_i. Thus there are additional 10 free parameters in the fit, which is generally too much. The HERAPDF group uses a procedure to determine how many of these new parameters are actually needed. Starting with a fit with $D_i = E_i = 0$, additional parameters are added until there is no longer an improvement in χ^2/ndf. First, the best 11-parameter fit is found by including one of the D_i or E_i parameters, then the best 12-parameter fit is determined in the same manner and so forth [1].

In the fits later on, no such procedure will be used, instead a fixed 13-parameter fit with free E_g, D_{u_v} and $D_{\bar{U}}$ parameters will be used. This has been found to be a good parametrization for HERA I+II DIS fits with CMS 8 TeV dijet data included at NLO [39]. The main goal will be, to examine the PDFs' changes between fits at NLO and NNLO as well as between fits at different scale definitions. Therefore it is suitable to use the same parametrization in each fit instead of finding the best set of parameters for each configuration.

4.4.2 Starting values

In addition to the theoretical problem of choosing an appropriate parameter set, another rather practical problem arises. Most minimization algorithms need some start values assigned to the parameters, that is a starting point in parameter space. If the χ^2 function is convex in the entire parameter space there is only one minimum, which is also the global one, and the choice of a starting point should in theory not affect the result. But often the χ^2 function does not have such nice properties and several different minima may exist. It is in general a good idea to choose the start values of the parameters to be as close to the final ones, as one can guess from physical arguments or already published PDF sets.

In the fits in Chapter 5 a well-defined procedure is used to choose the start values of the parameters. In the fits with only the HERA I+II inclusive DIS data, the starting values are taken from the HERAPDF 2.0 NNLO public PDF set [4] and $E_g = D_{u_v} = 0$ because these two parameters were not fitted in HERAPDF 2.0 NNLO. The final converged minima from these DIS only fits are then used as starting points for the fits with jet data included, whose final parameters are again used as input to the fits including the strong coupling constant. For the scale variation fits, the final parameters of the central fit are used as starting values.

4.5 Uncertainty estimation

Any physical measurement must include some kind of information of how reliable it is. Often this is done by publishing a confidence interval in addition to the actual value that was measured. In the case of PDF fitting these uncertainty intervals are especially important, because other physicists may use the PDF in their analyses where the PDF uncertainties propagate to the final result. There are several sources of uncertainties in the fits that are shown later on. The most obvious one is the datasets that include uncertainties on the measured cross section values. For example, this includes uncertainties on the calibrations and resolutions of the used detectors as well as on the luminosity provided by the particle accelerators.

In addition to these experimental uncertainties, there are uncertainties that originate from the theory calculations. These include model uncertainties, where fixed values for physical parameters (e.g. quark masses) are used without taking into account their uncertainties. Another uncertainty that can be counted towards the theory uncertainties and that is very difficult to determine, is the fact that we selected a specific PDF parametrization. The choice of parametrization already confines the outcome of the fits to a special set of functions and there is no obvious way to conclude, how much this biases the fitted PDFs and their estimated uncertainties. To estimate the parametrization uncertainties, the HERAPDF group used a procedure where additional free parameters where included into the fit after the central fit had been done. This allows to quantify, by how much this additional freedom leads to other fit results, and an error band on the PDFs can be constructed by taking the largest deviation from the central fit at each point x. Yet, one must be careful in interpreting these error bars, as they are not rooted in any exact statistical model and may still be biased by the choice of parametrization.

In this work only experimental uncertainties and scale uncertainties are analysed. They will be explained in the following.

4.5.1 Experimental uncertainties

There are several approaches to estimate the experimental uncertainties on the PDFs. Looking at the shape of the χ^2 function around its minimum, one can define an „allowed" range of parameters as the region in parameter space that is bordered by the $\Delta\chi^2 = 1$ contour. The maximal values for each of the parameters θ_i can be found by profiling them like it was shown in section 4.3. However, here, we are not interested in confidence intervals on the parameters, but rather on the PDFs' confidence bands. The exact approach would be to find the maximum and minimum values of each PDF at each point x in the entire $\Delta\chi^2 \leq 1$ region in parameter space. In the case that these extreme values lay on the $\Delta\chi^2 = 1$ edge and not in the interior region, the problem can be solved with the so-called Lagrange multiplier method [37], which is very slow computationally though.

Often an approximation is done to simplify the problem of finding the „allowed" range of the PDFs. The χ^2 function is expanded in second order around its minimum, where the first derivative is zero and second derivatives are given by the hesse matrix.

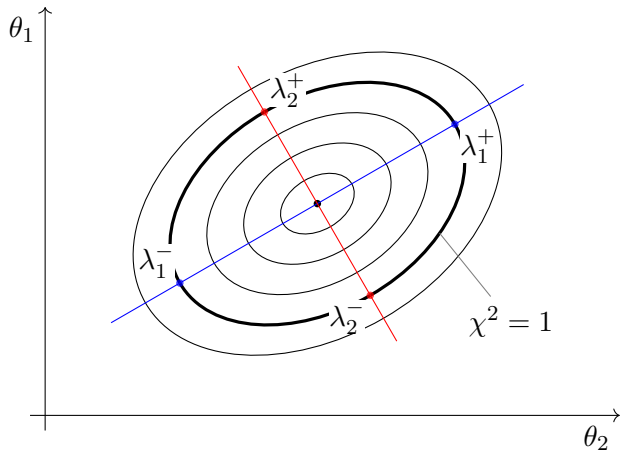


Figure 4.3: Hesse method of uncertainty estimation: The hesse matrix of the χ^2 is diagonalized and the PDFs are evaluated at the up and down variation in each eigenvector direction.

The PDFs dependency on the parameters θ_i is approximated to first order around the χ^2 minimum. As illustrated in Figure 4.3 the Hesse matrix is diagonalized in an orthonormal eigenbasis. The PDFs are evaluated at the up and down variation in each eigenvector direction ($\lambda_i^\pm$ in Figure 4.3) yielding $f(x|\lambda_i^\pm)$. The combined up and down variations of the central PDF $f(x)$ can then be calculated by the following rules:

$$\begin{aligned} \Delta f^+(x) &= \sqrt{\sum_i \left[\max\left(f(x|\lambda_i^+) - f(x), f(x|\lambda_i^-) - f(x), 0\right)\right]^2} \\ \Delta f^-(x) &= \sqrt{\sum_i \left[\min\left(f(x|\lambda_i^+) - f(x), f(x|\lambda_i^-) - f(x), 0\right)\right]^2} \end{aligned} \tag{4.13}$$

This method of estimating PDF uncertainties is called the *hesse* method. In general the PDFs $f(x|\lambda_i^\pm)$ are all stored along with the central PDF. This allows uncertainty propagation with a formula similar to equation (4.13) if the PDFs are used to calculate for example a cross section X:

$$\begin{aligned} \Delta X^+ &= \sqrt{\sum_i \left[\max\left(X_i^+ - X, X_i^- - X, 0\right)\right]^2} \\ \Delta X^- &= \sqrt{\sum_i \left[\min\left(X_i^+ - X, X_i^- - X, 0\right)\right]^2} \quad , \end{aligned} \tag{4.14}$$

where X is the central prediction and $X_i^\pm$ are the cross sections calculated with the $f(x|\lambda_i^\pm)$ PDFs.

4.5.2 Scale uncertainties

Truncating the expansion in α_s in perturbative QCD causes the calculation to depend on specific choices of the renormalization and factorization scale. Naturally, the question

arises, how this arbitrariness of scale choice can be quantified in a scale uncertainty. A commonly accepted estimation of the scale uncertainty on some quantity X is calculated to be the maximal deviations of X upon varying the renormalization and factorization scale independently by a factor two. This gives the six scale variations $(\mu_R, \mu_F) \in \{(1/2, 1/2), (1/2, 1), (1, 1/2), (1, 2), (2, 1), (2, 2)\}$, where the variations with large relative factors between μ_R and μ_F are avoided [38, pp. 38-40].

4.6 Software setup

The central piece of software that is used in the fits in chapter 5, is the PDF fitting program `xFitter`[2]. A detailed description of its predecessor `HERAFitter`, from which it inherits a lot of the code base, can be found in the corresponding publication [6]. `xFitter` interfaces the `MINUIT` minimization framework [26] by providing the χ^2 function that is to be minimized. Through a Minuit steering card, the type and order of several minimization algorithms supported by `MINUIT` can be selected, as well as starting values and limits for the parameters. Details on the steering cards can be found in the appendix chapter B. The χ^2 function is computed by `xFitter` according to the definition in section 4.3, which corresponds to the χ^2 settings that will be used in the fits in chapter 5, although other χ^2 definitions are possible in `xFitter` by changing the settings in the steering file.

Cross section predictions for the HERA I+II inclusive deep inelastic scattering dataset (see section 3.1.2) are calculated directly by `xFitter` with the help of the QCD evolution package `QCDNUM`[3] [12]. The cross section predictions for the CMS dijet datasets is calculated by evaluating pre-calculated interpolation tables with `FastNLO` that is also interfaced by `xFitter` for this purpose. The interpolation tables are filled with cross section predictions calculated by `NNLOJet` [23] and for the reproduction fits also with `NLOJet++`.

Instead of using a pre-installed version of `FastNLO`, the `xFitter` source code ships its own version of `FastNLO`, that was slightly modified to be able to be interfaced from `xFitter`. Unfortunately, the `FastNLO` code, that is distributed with the latest version of `xFitter` is outdated. This led to the necessity to patch some `FastNLO` code in the `xFitter` package, to enable the reading of the newest `FastNLO` table format. Exact instructions on how to do this are given in the appendix chapter A.

Detailed description on the `xFitter` settings and steering files that have been used to produce the fits in the following Chapter 5 may be found in the appendix Chapter B.

[2] in version 2.0.1
[3] in version 17-01-15

5. Fits and results

In the following, the results of the fits are presented. The first Section 5.1 shows the reproduction fits of results that were previously published. These fits are used to check that the upgraded software yields consistent results. In Section 5.2 new `FastNLO` tables are used to compare fits at NLO with fits at NNLO. In addition to that different central scale definitions in the theory calculation are compared.

Just like the PDFs, the strong coupling constant α_s must be determined from data in the framework of perturbative QCD. Fits, where the strong coupling constant is determined simultaneously along with the parton distributions, are presented in Section 5.3.

In Section 5.4, PDF fits including a double-differential dijet cross section measurement at 7 TeV are analyzed.

5.1 Reproduction of previous results

PDF fits with HERA I+II inclusive DIS data and CMS dijet data at $\sqrt{s} = 8\,\text{TeV}$ have been performed in the PhD thesis [39] as well as in the corresponding CMS publication [40]. The results presented in those two publications differ slightly:

1. Regarding the cross section measurement, there was an additional uncertainty added in the CMS publication, taking into account non-Gaussian tails in the jet energy resolution.
2. At the publication time of Ref. [39] there were no electroweak corrections available for the theory calculation. Therefore the fits in Ref. [39] do not include the corrections and only data bins with $p_{\text{T,avg}} < 1000\,\text{GeV}$ were fitted.
3. While Ref. [39] used a fixed 13 parameter parametrization for the PDF fit, this was changed to a fixed 16 parameter fit in the CMS publication [40]. This 16 parameter parametrization includes a negative term in the gluon distribution that was also used in the HERAPDF2.0 fit [4]:

$$xg(x) = A_g x^{B_g}(1-x)^{C_g} - A'_g x^{B'_g}(1-x)^{C'_g} \tag{5.1}$$

In the following, the PDF fitting results of Ref. [39] will be reproduced, one time without the electroweak corrections and the non-Gaussian tail uncertainty, as well as

on time with both included. The remaining settings for the fit are taken exactly as described in Ref. [39], which includes a cut on the inclusive DIS data of $Q^2 \geq 7.5\,\mathrm{GeV}^2$, a fixed value of $\alpha_s = 0.1180$ for the strong coupling constant and the 13 parameter PDF parametrization described in Section 4.4. The goal of these reproduction fits is mainly to check if the new version 2.0.1 of `xFitter` yields similar results in comparison to the original ones.

The `FastNLO` interpolation tables containing the theory calculations for the 8 TeV triple-differential dijet data that is used in Ref. [39] and Ref. [40], were calculated by the theory program `NLOJet++` and are available at hepdata[1]. Here, as well as in the original fits, the central scale choice $\mu = p_{\mathrm{T},1} e^{0.3y^*}$ is used. The new tables that will also be used later on in the NNLO fits are calculated by another theory program `NNLOJET`. Thus the second goal of the reproduction fits will be, to check if the NLO cross section predictions of `NLOJet++` and `NNLOJET` lead to the same results.

Figure 5.1 shows the gluon PDF from the fit in Ref. [39]. The most notable changes from including the dijet data into the fit were found to be in the gluon PDF. This is expected because the theory predictions of jet data are more sensitive to the gluon distribution than those of inclusive deep inelastic scattering data. In addition to the change of shape, a reduction of uncertainty is observed in the fit including dijet data [39].

The reproduction of the PDFs in Figure 5.1 is shown in Figure 5.2 as the blue and yellow bands. Ref. [39] did not publish their final fitted parameters. Thus a comparison is only possible by looking at the plotted PDFs and the final χ^2 values of the fits. Regarding the inclusive DIS fit without dijet data (blue band) there is almost no difference between the original fit in Figure 5.1 and the reproduction in Figure 5.2. The difference in χ^2 values is of the order of 2 per mille. Looking at the fits including 8 TeV dijet data (yellow band) one sees that the reproduction fit is also in good agreement to the original fit with a difference in final χ^2 values of about 3 per mille.

The other two fits in Figure 5.2 (green and red bands) show fits with the non-Gaussian tail uncertainty source and the electroweak corrections included and the $p_{\mathrm{T,avg}} < 1000\,\mathrm{GeV}$ cut removed. The green one is fitted with the same `FastNLO` table that was used in the original fits, and has been calculated by the theory program `NLOJet++`. The red band shows the same fit done with the new `FastNLO` tables that were calculated by `NNLOJET`. These new tables include theory calculations up to NNLO, but were used only up to NLO in order to compare the results to the old tables. Looking at the bands in Figure 5.2 one can see that the gluon PDFs agree very well. The χ^2 values are the same up to the per mille range. A detailed comparison including the fitted parameter values and partial χ^2 values can be found in the appendix Section C.1.

To conclude, it is safe to say, that the newest version of xFitter (v2.0.1) and the new `FastNLO` tables calculated by `NNLOJET` reproduce results that are consistent with the results of Ref. [39]. Also, it is notable that the inclusion of the electroweak corrections has almost no impact on the fitted PDFs. Indeed, the $p_{\mathrm{T,avg}}$ cut excludes some of the bins, where the electroweak corrections are expected to have the greatest impact,

[1] `https://www.hepdata.net/record/ins1598460`

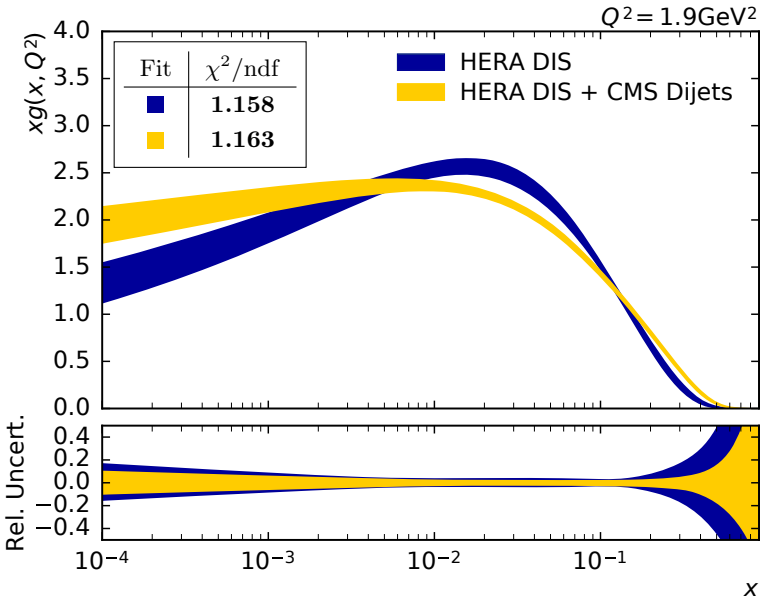


Figure 5.1: Fitted gluon PDF from Ref. [39]. Only the experimental uncertainty bands are shown and the colors were adjusted to match the ones in Figure 5.2.

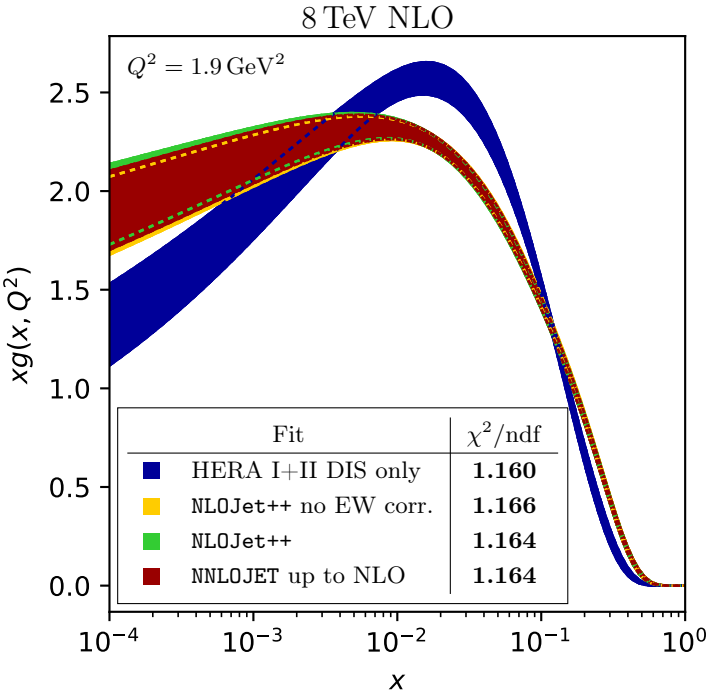


Figure 5.2: Gluon PDF of reproduction fits. Only the experimental uncertainty bands are shown. The prominent features of the original fits (■ and ■) could be reproduced; the difference between the `NLOJet++` (■) and the `NNLOJET` (■) calculations are negligible.

but a recent analysis of the NNPDF group does not find any significant impact of the electroweak corrections on PDF fit results for this specific dataset either [3].

5.2 Comparison between fit at NLO and NNLO

The new `FastNLO` tables, that are calculated by `NNLOJET` for cross section prediction of the CMS 8 TeV triple-differential dijet data, feature two main novelties. First, they include perturbative calculations of the dijet cross section up to next-to-next-to-leading order (NNLO) and second, they allow to choose the central scale for the evaluation from the two scale definitions

$$\mu_{R/F} = p_{\mathrm{T},1} e^{0.3y^*} \quad \text{„ptmax”} \quad \text{and} \quad \mu_{R/F} = m_{12} \quad \text{„m12”} \quad , \tag{5.2}$$

where the dijet mass m_{12} is defined as:

$$m_{12} = \sqrt{(p_1^\mu + p_2^\mu)^2} \tag{5.3}$$

with the leading and second leading jets' four-momentums p_1^μ and p_2^μ. These new tables can be used to study the impact of the new NNLO theory contribution on the PDF fit results. Like in the reproduction fits in the previous section, the PDFs are first fitted to the combined HERA I+II inclusive DIS cross section measurements. However in comparison to the fits in the previous section, the cut on Q^2 is increased from 7.5 to 10.0 GeV^2. On top of the inclusive DIS data, the 8 TeV triple-differential dijet cross sections are then included in additional fits at different orders of α_s and with different central scale definitions.

Detailed information on the results like fitted PDF parameters, shapes of all PDFs and partial χ^2 values are given in the appendix section C.2.

Figure 5.3 shows the gluon PDFs of all the fits. The most prominent feature in the PDFs coming out of the NLO fits, which are plotted on the left hand side, is the significant difference between the fits with different scale definitions. Looking at the χ^2 values, it can be seen that the „ptmax” ($\mu_{R/F} = p_{\mathrm{T},1} e^{0.3y^*}$) central scale definition yields cross section predictions that are consistent with the DIS only fit. Approximately a $\Delta\chi^2 = 1$ increase per additional data point is observed which is the expected value. With the dijet mass „m12” central scale definition, the resulting χ^2 value is significantly larger which indicates more tension between these cross section predictions and the measured data.

Interestingly, at NNLO all fits are much more in accordance to each other, which can be seen by comparing the final χ^2 values and by looking at the gluon PDF shapes shown in Figure 5.3. This is an indicator that at higher order perturbative calculations the scale dependency gets smaller. Theoretically such a behaviour is expected, because in a renormalizable theory, the dependence on the scales should get smaller if the perturbative series is expanded up to higher orders of α_s. The scale uncertainty, that is the theoretical uncertainty on the calculations due to specific scale choice, is typically estimated by varying the scales by factors 0.5 (down variation) and 2.0 (up variation)

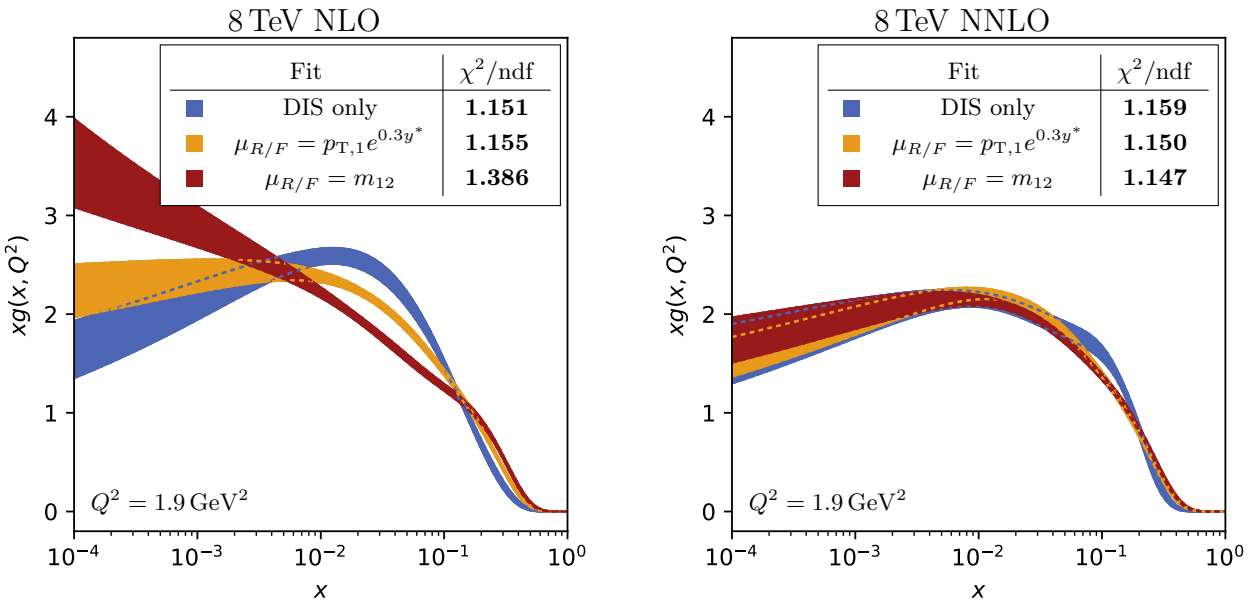


Figure 5.3: Gluon PDFs of the NLO (left) and the NNLO (right) fits. The blue band shows the fits with HERA I+II inclusive DIS data only. The yellow and orange bands both correspond to fits including CMS 8 TeV 3D dijet data but different central scale definitions in the theory calculations. Only the experimental uncertainties are shown.

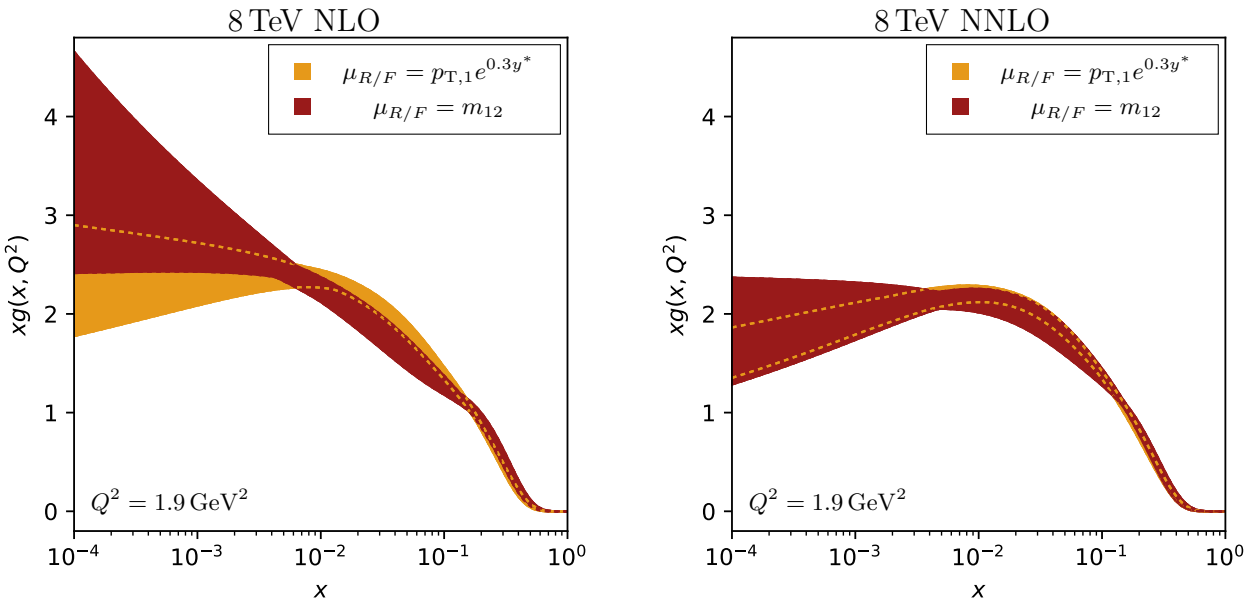


Figure 5.4: Scale uncertainties of the Gluon PDFs of the NLO (left) and the NNLO (right) fits. The scale uncertainties are constructed by taking the maximal envelope of all the scale variations for a given central scale including the experimental uncertainties.

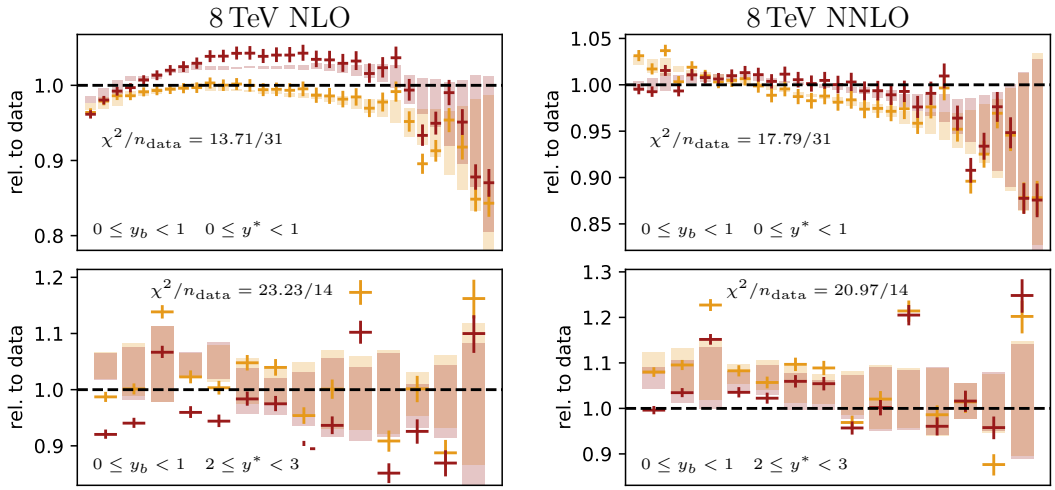


Figure 5.5: Theory (crosses) and shifted data (bands) relative to the central data for the $\mu_{R/F} = p_{\text{T},1} e^{0.3y^*}$ (orange ■) and $\mu_{R/F} = m_{12}$ (dark red ■) scale definitions for NLO (left) and NNLO (right) for the central rapidity bin (top) and the high-y^* bin. The partial χ^2 values for these bins are given for the „ptmax" central scale fit (orange).

and taking the maximal envelope of the results. Following a common practice, the renormalization μ_R and factorization scale μ_F used in the dijet cross section calculation are varied independently in six scale variations as described in Section 4.5. This leads to the uncertainty bands (including experimental uncertainties) shown in Figure 5.4, where the decrease of scale uncertainty is clearly visible, if one switches on the NNLO perturbative contribution.

It is not easy to judge if a fit describes some data well. While the overall χ^2 value is often a good quantity to measure the goodness of fit, it does not reveal which of the datapoints match the prediction well and which not. Further insights into the details of the fit's prediction quality can be found by comparing the prediction to the measurement or by looking at the partial χ^2 values.

A direct comparison between the cross section predictions in the central and the high-y^* rapidity bins is shown in Figure 5.5 (the remaining rapidity bins can be found in the appendix C.6). The theory predictions shown as crosses are calculated by evaluating the `FastNLO` tables with the fitted PDFs and the corresponding scale definition. The colored bands show the data points shifted with the corresponding fitted values of the correlated uncertainty sources (also called *shifts* in the following). First of all, one can see that in NLO there is a significant discrepancy between the predictions of the fit with $\mu = p_{\text{T},1} e^{0.3y^*}$ and the one with $\mu = m_{12}$. If the NNLO contribution is activated, the predictions get a lot more consistent. This is the same conclusion that has been made above by comparing the fitted gluon PDF shapes and the total χ^2 values of the fits. Another interesting feature can be seen in the central rapidity bin, where in NLO there is a „curvature", that vanishes for the most part in NNLO. This curvature is matched rather well in NLO by the shifted data except in the central $p_{\text{T,avg}}$ region for the $\mu = m_{12}$ fit. Yet, compensating the curvature by shifting the data points comes at

the cost of higher values for the fitted correlated uncertainty sources and this curvature may be responsible for some of the outliers in the fitted shifts, that are discussed below.

Tables 5.1 and 5.2 show the partial χ^2 values of the fits. While partial χ^2 values ignore some of the parts in the χ^2 definition, like systematic correlations between the different rapidity bins, they give a rough idea of which rapidity bins are the best matching, and which are the most problematic ones. According to the partial χ^2 values, the predictions match the measured cross sections better in the central rapidity bins with $y_b + y^* < 2$. Actually in NNLO, the partial χ^2 value in each of these bins is lower, than the number of datapoints, while it is higher in each of the outer bins ($y_b + y^* >= 2$).

The partial χ^2 tables also show the p-values of the final χ^2 values. This is the probability to find a minimum χ^2 value at least as extreme as the one actually found in the fit, assuming the goodness of fit estimator is χ^2-distributed and the physical model is true. Usually, this would mean that all of the fits presented above could be rejected with at least a 3σ significance (p-value 2.70×10^{-3}). There are, however, uncertainties that are not included into the goodness of fit estimator. In fact, none of the theoretical uncertainties, namely the scale-, model- and parametrization uncertainties are taken into account in the χ^2 calculation. Therefore the p-values in Table 5.1 and 5.2 can not be used to make statistical decisions, but can only be used to compare the different fits with each other.

Systematic correlated uncertainties are included in the PDF fits as nuisance parameters (see Section 4.3) and fitted in each iteration in the minimization algorithm. They are constructed in a way, that theoretically they would behave like normal distributed random variables with a mean 0 and a standard deviation of 1. There are 169 such systematic correlated uncertainty sources in the combined HERA I+II inclusive DIS dataset and 28 uncertainty sources in the triple-differential CMS 8 TeV dijet dataset. The fitted values of these correlated uncertainty sources can also be used to judge fit quality. The squared sum of n independent standard normal distributed random variables follows a χ^2 distribution with n degrees of freedom. The „correlated χ^2"-row in Tables 5.1 and 5.2 show this squared sum of shifts for all the correlated uncertainty sources included in the corresponding fit. The inclusive DIS only fit shows a value of 50.96 in NLO and 55.48 at NNLO, which is a lot smaller than the expected value of $E[\chi^2_{169}] = 169$ for $n = 169$ uncertainty sources. This may be a hint, that the correlated uncertainties are slightly overestimated in the combined HERA dataset.

Figure 5.6 shows the fitted shifts of the correlated uncertainty sources of the triple-differential dijet data set, as well as the sum of their squares for each fit. The uncertainties have different origins:

- **nperr** is the uncertainty of the NP corrections. It is the only theoretical uncertainty directly included in the fit. They tend to be larger in higher $p_{\mathrm{T,avg}}$ regions, except in the central rapidity bin, where they are larger in the low $p_{\mathrm{T,avg}}$ region.
- **jererr** and **nongaussiantails** have their origin in the jet energy resolution. These tend to be larger in the outer rapidity bins.
- **lumi** is a global scaling uncertainty taking into account possible systematics in

Table 5.1: Partial χ^2 values for the NLO fits in Section 5.2

		n_{data}	HERA I+II Dis only	with CMS dijets $\mu = p_{\text{T},1} e^{0.3y^*}$	with CMS dijets $\mu = m_{12}$
HERA I+II	combined	1016	1106.14	1128.45	1181.30
CMS 8 TeV dijets	yb0 ys0	31	–	13.71	35.82
	yb0 ys1	26	–	13.37	28.03
	yb0 ys2	14	–	23.23	76.82
	yb1 ys0	23	–	13.98	19.21
	yb1 ys1	17	–	20.33	27.45
	yb2 ys0	11	–	28.62	88.36
	combined	122	–	113.24	275.69
	correlated χ^2		50.96	69.88	115.15
	log penalty χ^2		-2.98	-11.96	-13.24
	combined		1154.12	1299.63	1558.90
	ndf		1003	1125	1125
	p-value		6.13×10^{-4}	2.15×10^{-4}	1.26×10^{-16}
	combined χ^2/ndf		**1.151**	**1.155**	**1.386**

Table 5.2: Partial χ^2 values for the NNLO fits in Section 5.2

		n_{data}	HERA I+II Dis only	with CMS dijets $\mu = p_{\text{T},1} e^{0.3y^*}$	with CMS dijets $\mu = m_{12}$
HERA I+II	combined	1016	1109.15	1116.89	1119.54
CMS 8 TeV dijets	yb0 ys0	31	–	17.79	16.70
	yb0 ys1	26	–	15.90	12.78
	yb0 ys2	14	–	20.97	23.38
	yb1 ys0	23	–	20.09	16.22
	yb1 ys1	17	–	25.70	20.40
	yb2 ys0	11	–	13.25	22.10
	combined	122	–	113.70	111.58
	correlated χ^2		55.48	63.96	63.17
	log penalty χ^2		-1.74	-1.04	-3.95
	combined		1162.89	1293.52	1290.33
	ndf		1003	1125	1125
	p-value		3.24×10^{-4}	3.33×10^{-4}	4.16×10^{-4}
	combined χ^2/ndf		**1.159**	**1.150**	**1.147**

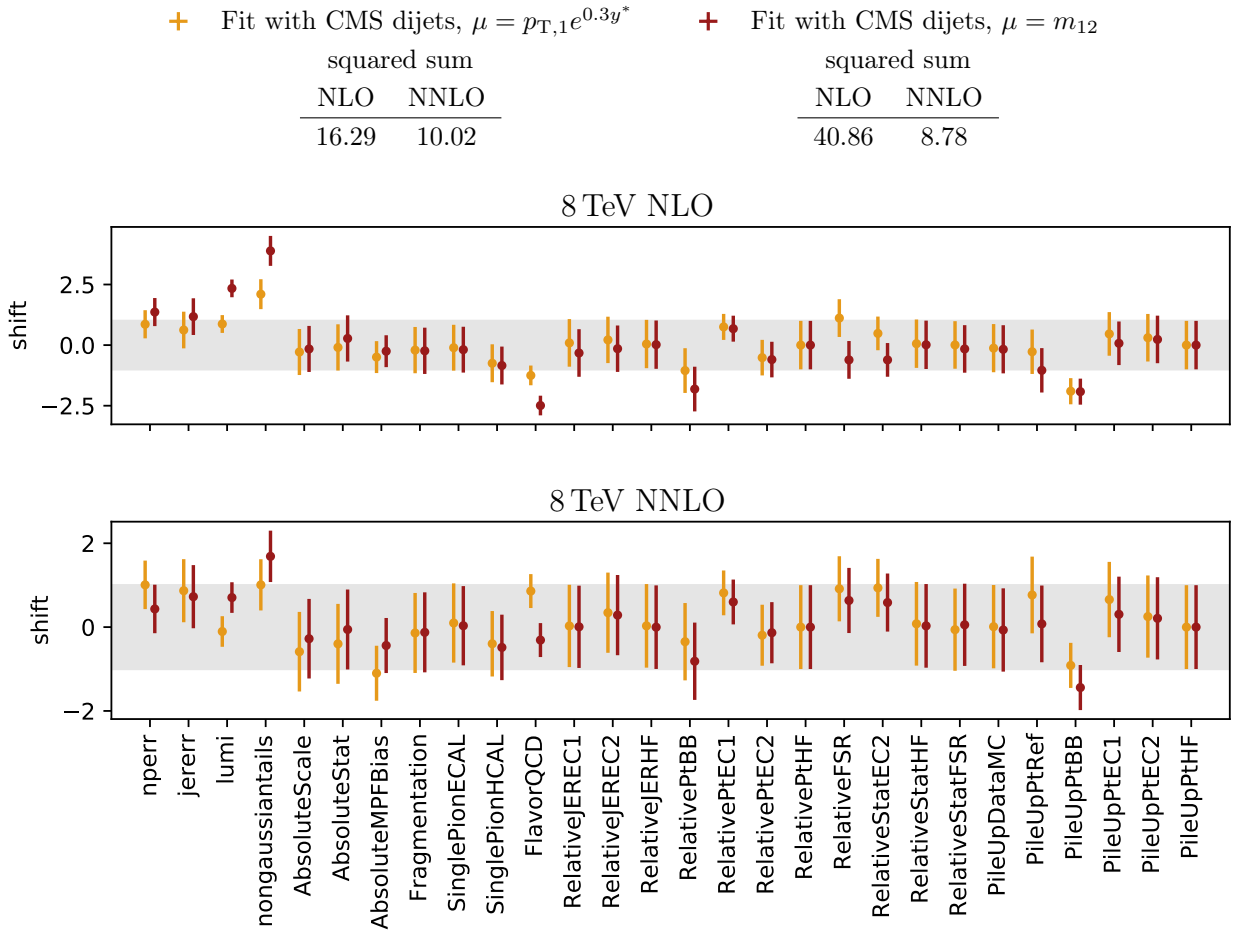


Figure 5.6: Fitted values on the correlated systematic uncertainty sources of the triple differential dijet dataset at 8 TeV for the NLO and the NNLO fits. The gray band marks the one sigma region. The tables on the top list the sums of the squared shift values.

the estimation of the LHCs luminosity. Nonzero values of the lumi uncertainty shift all datapoints up or down by the same relative factor.

- The remaining uncertainty sources originate in the jet energy calibration (JEC) and are described in detail in Ref. [29]. Their influence on the dijet cross section measurement, that is analyzed here, is documented in Ref. [39]. For the discussion here, it is interesting to look at the **FlavourQCD** and the **PileUpPtBB** sources. Both happen to have a large impact in one phase space region while having almost no impact in other phase-space regions. For example, the „PileUpPtBB" uncertainty is large in the central rapidity bins at large $p_{T,avg}$ and almost zero in the outer rapidity bins and at small $p_{T,avg}$ [39].

As already mentioned above and shown in Figure 5.5, the predictions at NLO exhibit a „curved" deviation from the measured data points in the central rapidity bin. To compensate for this, shifts are needed that push the data to smaller values at the

upper and lower end of the $p_{\mathrm{T,avg}}$ range while leaving them be in the middle. The „FlavourQCD" and „PileUpPtBB" uncertainties sources take rather large negative values in Figure 5.6 in the NLO case. This is due to the fact, that their non-uniform impact on the different data points permits a data shift to match the curved NLO prediction. Especially for the predictions with $\mu = m_{12}$ (red in Figure 5.5) also a global shift upwards is needed because the predictions are higher than the data in the middle of the $p_{\mathrm{T,avg}}$ range. The high positive value of the „lumi" shift is probably caused by this, as the luminosity uncertainty has an uniform impact on all data points. The extreme outlier values of the „nongaussiantails" uncertainty (especially in the NLO fit with $\mu = m_{12}$ could not be tracked down to a simple explanation. The non-Gaussian tails in the jet energy resolution have the most impact in the outer rapidity bins [40], where there is more discrepancy between measured data and the predictions. This may be a possible explanation for the high value of the „nongaussiantails" shift.

Figure 5.6 also shows the sum of squared shifts for each fit. Except for the NLO fit with the dijet mass as central scale definition, all these values also turn out to be below the expected value of 28, which like for the HERA data could be a hint for uncertainty overestimation.

Overall the essential finding here is the following: While the fits with NLO theory predictions work and yield reasonable results if looking separately at each of the fits, there is a large discrepancy between the two different central scale definitions, that are used in the theory calculation. If the NNLO contributions are added to the calculations, the differences in the results decrease drastically, while at the same time the scale uncertainties get smaller. It seems, that the NNLO contribution is crucial to get scale independent and consistent results in PDF fits to triple-differential dijet data.

5.3 Including the strong coupling constant

In the fits in Sections 5.1 and 5.2 a fixed value of $\alpha_s(M_z) = 0.1180$ was used. However the strong coupling constant, just like the PDFs, is also a quantity, that has to be determined from data. In fact, there is a strong correlation between the gluon PDF and the value of the strong coupling. Fitting PDFs with a fixed value of α_s puts an additional constraint on the fits. The `xFitter` framework allows to include $\alpha_s(M_z)$ as a free parameter into the PDF fits and the results of such fits will be studied in the following. Again, fits are performed at NLO and NNLO and with both available central scale definitions in the `FastNLO` tables.

Figure 5.7 shows the gluon PDF shapes, as well as the χ^2 and fitted $\alpha_s(M_z)$ values. The essential finding from Section 5.2, that the fits with different central scale definitions are more consistent in NNLO, can be found in the fits with free α_s as well. The best-fitting values of the strong coupling constant are found to be

$$\begin{aligned} \alpha_s(M_z) &= 0.1191 \pm 0.0015(\mathrm{exp})\, ^{+0.0028}_{-0.0016}(\mathrm{scale}) \qquad \text{at NLO and} \\ \alpha_s(M_z) &= 0.1155 \pm 0.0012(\mathrm{exp})\, ^{+0.0008}_{-0.0017}(\mathrm{scale}) \qquad \text{at NNLO} \end{aligned} \tag{5.4}$$

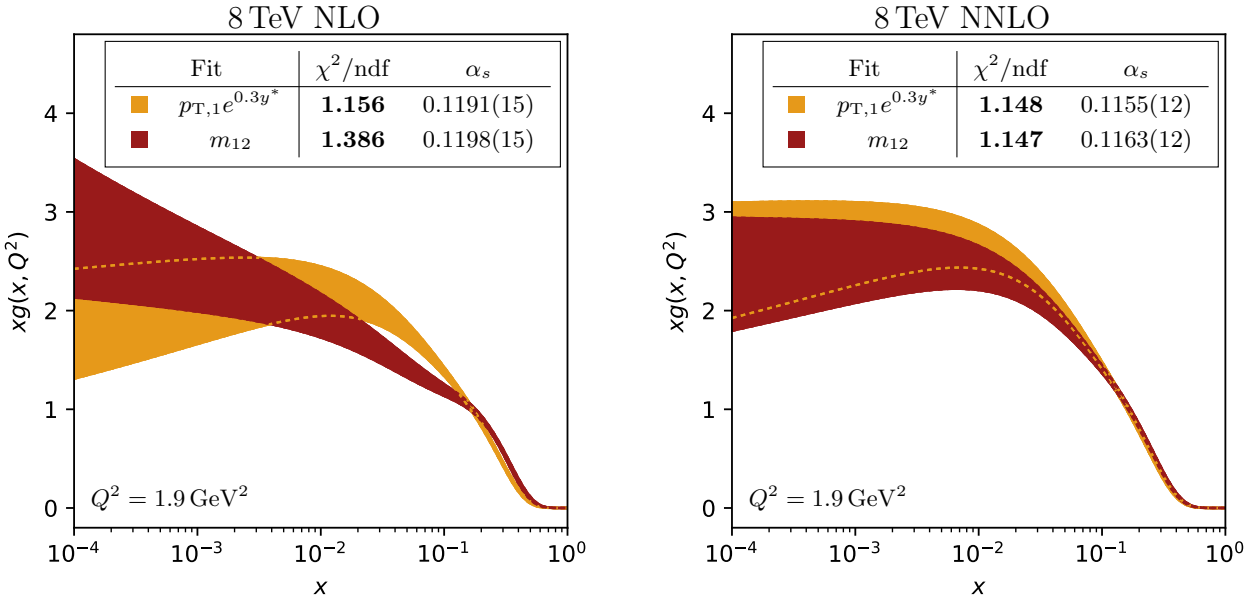


Figure 5.7: Gluon PDFs corresponding to the NLO (left) and NNLO (right) fits, with α_s included as free parameter. The bands show the increased experimental uncertainties in comparison to Figure 5.3.

with the central scale definition of $\mu = p_{T,1}e^{0.3y^*}$ („ptmax") and

$$\begin{aligned} \alpha_s(M_z) &= 0.1198 \pm 0.0015(\mathrm{exp})\, {}^{+0.0021}_{-0.0021}(\mathrm{scale}) \qquad \text{at NLO and} \\ \alpha_s(M_z) &= 0.1163 \pm 0.0013(\mathrm{exp})\, {}^{+0.0010}_{-0.0004}(\mathrm{scale}) \qquad \text{at NNLO} \end{aligned} \tag{5.5}$$

with the dijet mass as central scale („m12"). The experimental uncertainties are determined from the shape of the χ^2 function at its minimum in a similar way like the experimental uncertainties on the PDFs. The scale uncertainties are taken to be the maximal envelopes of the $\alpha_s(M_z)$-included fits upon the six scale variations described in section 4.5. The uncertainties on the α_s values presented here must be treated with caution, as only the experimental and scale uncertainties have been determined. Parametrization and model uncertainties, which have not been analyzed within the scope of this work, and remain to be investigated.

The NLO value in equation (5.4) is in good accordance to the value obtained by Ref. [39], where the same PDF parametrization and central scale definition were used:

$$\alpha_s(M_z) = 0.1194 \pm 0.0015(\mathrm{exp}) \pm 0.0002(\mathrm{mod})\, {}^{+0.0002}_{-0.0004}(\mathrm{par})\, {}^{+0.0031}_{-0.0019}(\mathrm{scale}) \quad . \tag{5.6}$$

Both the values for $\alpha_s(M_z)$ at NLO in (5.4) as well as in (5.5) agree well with other published values determined in NLO perturbative theory, like for example CMS publications of inclusive jet production ($\alpha_s(M_z) = 0.1185 \pm 0.0019(\mathrm{exp})^{+0.0060}_{-0.0037}(\mathrm{theo})$) [27] or differential 3-jet production ($\alpha_s(M_z) = 0.1171 \pm 0.0013(\mathrm{exp})^{+0.0073}_{-0.0047}(\mathrm{theo})$) [28]. An overview of different determinations of the strong coupling constant can be found

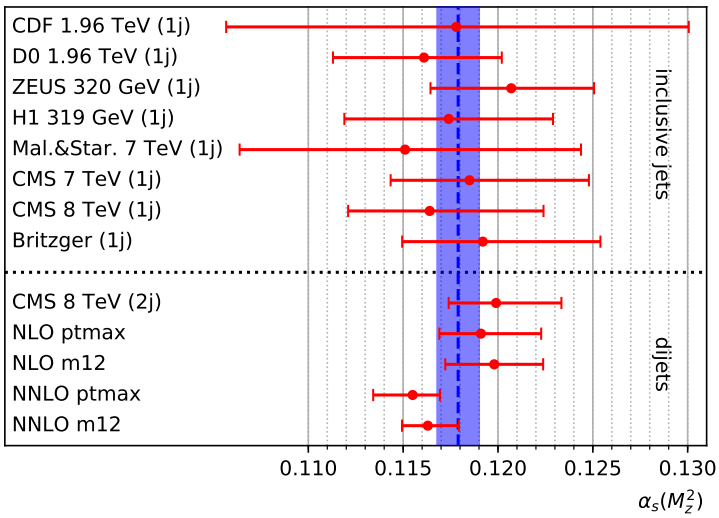


Figure 5.8: Overview of different $\alpha_s(M_z)$ determinations using jet cross sections. In the top section values fitted to inclusive jet cross section are shown. The bottom section shows the $\alpha_s(M_z)$ value obtained with triple-differential dijet cross sections at 8 TeV by CMS [40] as well as the values fitted with the same data in this work. With the exception of the two NNLO fits at the bottom, all $\alpha_s(M_z)$ determinations were done with theory calculations at NLO. The uncertainties are computed by a quadratic summation of all published uncertainties for each of the values. The uncertainties for the values that were obtained in this work are the quadratic sum of the experimental and the scale uncertainties presented in Equations (5.4) and (5.5). The blue band shows the world average value from Ref. [36].

in the recent „Review of particle physics" from the Particle Data Group [36]. They mention several $\alpha_s(M_z)$ determinations from inclusive jet cross section measurements. Figure 5.8 shows a comparison between these determinations (top section in the figure) [2, 5, 8, 13, 19, 27, 30, 32], the $\alpha_s(M_z)$ determination with triple-differential dijets at 8 TeV with the same data that is analyzed in this work [40] and the values presented above.

Figure 5.8 also shows the Particle Data Group's recent world average value of $\alpha_s(M_z) = 0.1179 \pm 0.0010$ [36] (blue band) which consists of a large variety of different α_s measurements as well as calculations. It in good agreement with the NLO results obtained in this work. The values determined using NNLO theory tend to be somewhat smaller as expected, but are also compatible with the world average.

The uncertainty bands in Figure 5.7 reveal an increased experimental uncertainty of the gluon distribution in NLO as well as in NNLO compared to the fixed-α_s fits in Section 5.2. This enlarged experimental uncertainty is caused by the correlation of the gluon PDF with the strong coupling. This means that the χ^2 function can not separate these quantities very well and if both of them are fitted without external constraints, it will result in larger uncertainties on the gluon PDF as well as on α_s. To further examine

the correlation a profiling of α_s is done. For a series of fixed values of $\alpha_s(M_z)$ in the interval $[0.114, 0.122]$, the PDFs are fitted and the minimum χ^2 values are plotted over $\alpha_s(M_z)$.

The results of these $\alpha_s(M_z)$ profile fits are shown in Figure 5.9 for the „ptmax" scale definition $\mu = p_{\mathrm{T},1} e^{0.3y^*}$ and in Figure 5.10 for the dijet mass scale definition $\mu = m_{12}$. On the right hand side, the minimum χ^2 values are drawn for each fixed $\alpha_s(M_z)$ value. On the left hand side, for each fit, the corresponding gluon PDF is plotted. In the range of $x < 0.1$ the correlation between the gluon PDF and the strong coupling constant is clearly visible. Larger fixed values of $\alpha_s(M_z)$ yield a smaller fitted gluon density in the low-x region. The Figures C.7, C.8 (for the $\mu = p_{\mathrm{T},1} e^{0.3y^*}$ fits), C.9 and C.10 (for the $\mu = m_{12}$ fits) in the appendix also show the corresponding valence and sea quark distributions. But there, the correlations with $\alpha_s(M_z)$ are by far not as strong as in the case of the gluon density.

From the α_s profile, the best fitting value of $\alpha_s(M_z)$ can be determined, which is a technique alternative to direct fitting like above and already employed in previous measurements of the strong coupling, like in Refs. [27, 28]. Through a quadratic fit (gray curves in Figures 5.9 and 5.9), the $\alpha_s(M_z)$ with the lowest possible χ^2 value can be found. Uncertainties on this value are constructed by finding the values of α_s where the quadratic fit ascends above a χ^2 difference of $\Delta\chi^2 = 1$ above the minimum. Since a quadratic function is fitted, this leads to symmetric uncertainties. The values found like this are

$$\begin{aligned} \alpha_s(M_z) &= 0.1192 \pm 0.0015(\mathrm{exp}) \qquad \text{at NLO and} \\ \alpha_s(M_z) &= 0.1155 \pm 0.0013(\mathrm{exp}) \qquad \text{at NNLO,} \end{aligned} \tag{5.7}$$

for a central scale of $\mu = p_{\mathrm{T},1} e^{0.3y^*}$ („ptmax") and

$$\begin{aligned} \alpha_s(M_z) &= 0.1198 \pm 0.0015(\mathrm{exp}) \qquad \text{at NLO and} \\ \alpha_s(M_z) &= 0.1163 \pm 0.0014(\mathrm{exp}) \qquad \text{at NNLO,} \end{aligned} \tag{5.8}$$

with the central scale definition of $\mu = m_{12}$ (dijet mass). These values almost exactly match the values obtained from directly including $\alpha_s(M_z)$ into the fits in equations (5.4) and (5.5).

5.4 Additional dijet data

The triple-differential measurement at 8 TeV in Ref. [40], which was used for the fits in the previous sections, is not the only differential dijet cross section measurement published by CMS. In the following, a double-differential measurement at 7 TeV in Ref. [18] will be analyzed. The fits are done with the exact same settings as the ones in Section 5.2. However, the `FastNLO` tables with the theory calculations for this measurement are only available with a single central scale definition, namely the dijet mass m_{12} which is also the observable of the measurement defined in equation (3.10).

Figure 5.11 shows the gluon distributions of the fits with the 7 TeV data included. For comparison, the gluon PDFs of the inclusive DIS data only fits and the fits with

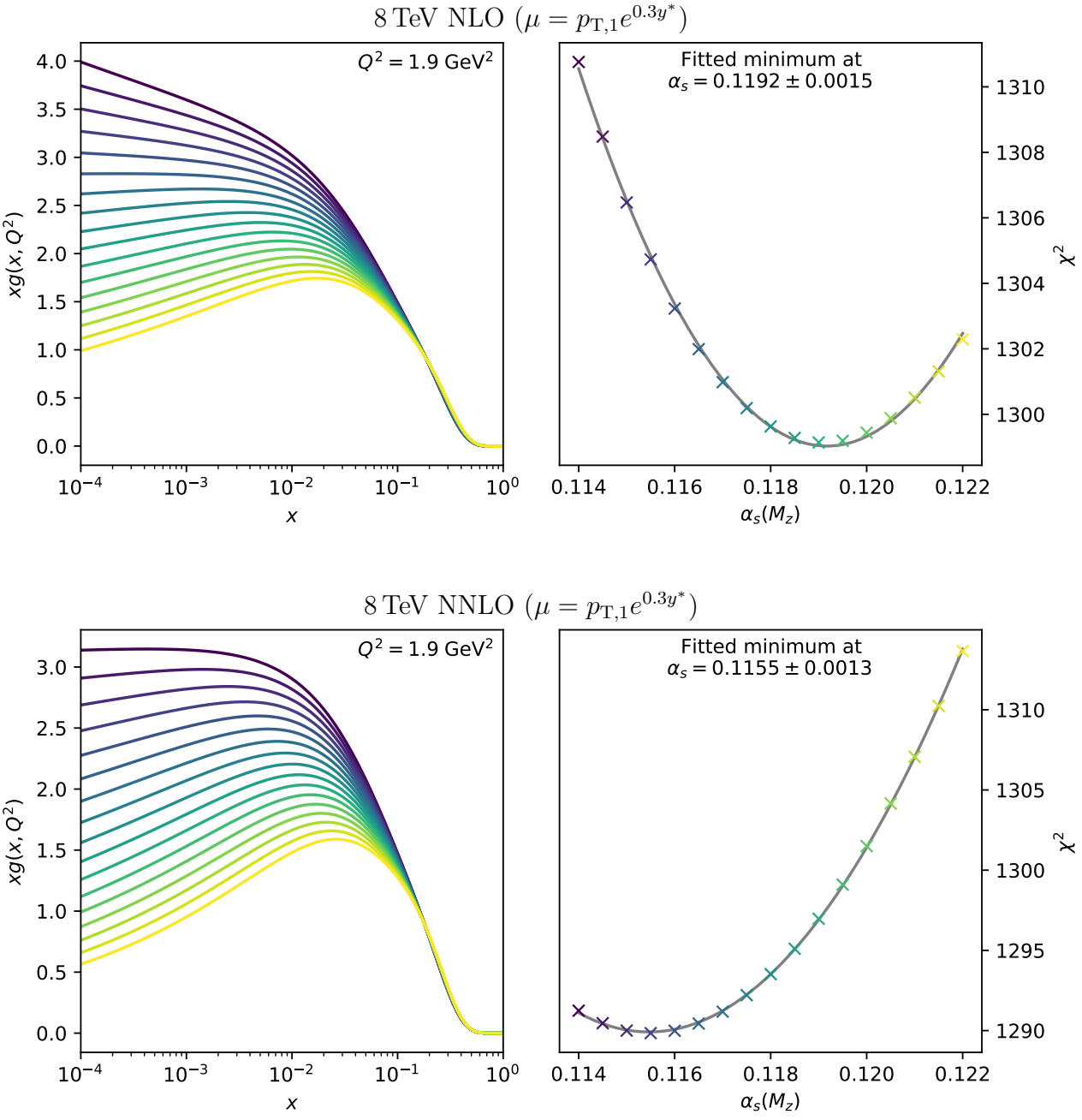


Figure 5.9: Profile of $\alpha_s(M_z)$ for NLO (top) and NNLO (bottom). Fits with the central scale $\mu = p_{T,1}e^{0.3y^*}$ are done for a series of fixed values of $\alpha_s(M_z)$ and the corresponding minimum χ^2 values plotted on the left panels. For each fit, the emerging gluon distribution is plotted on the right hand side in the corresponding color.

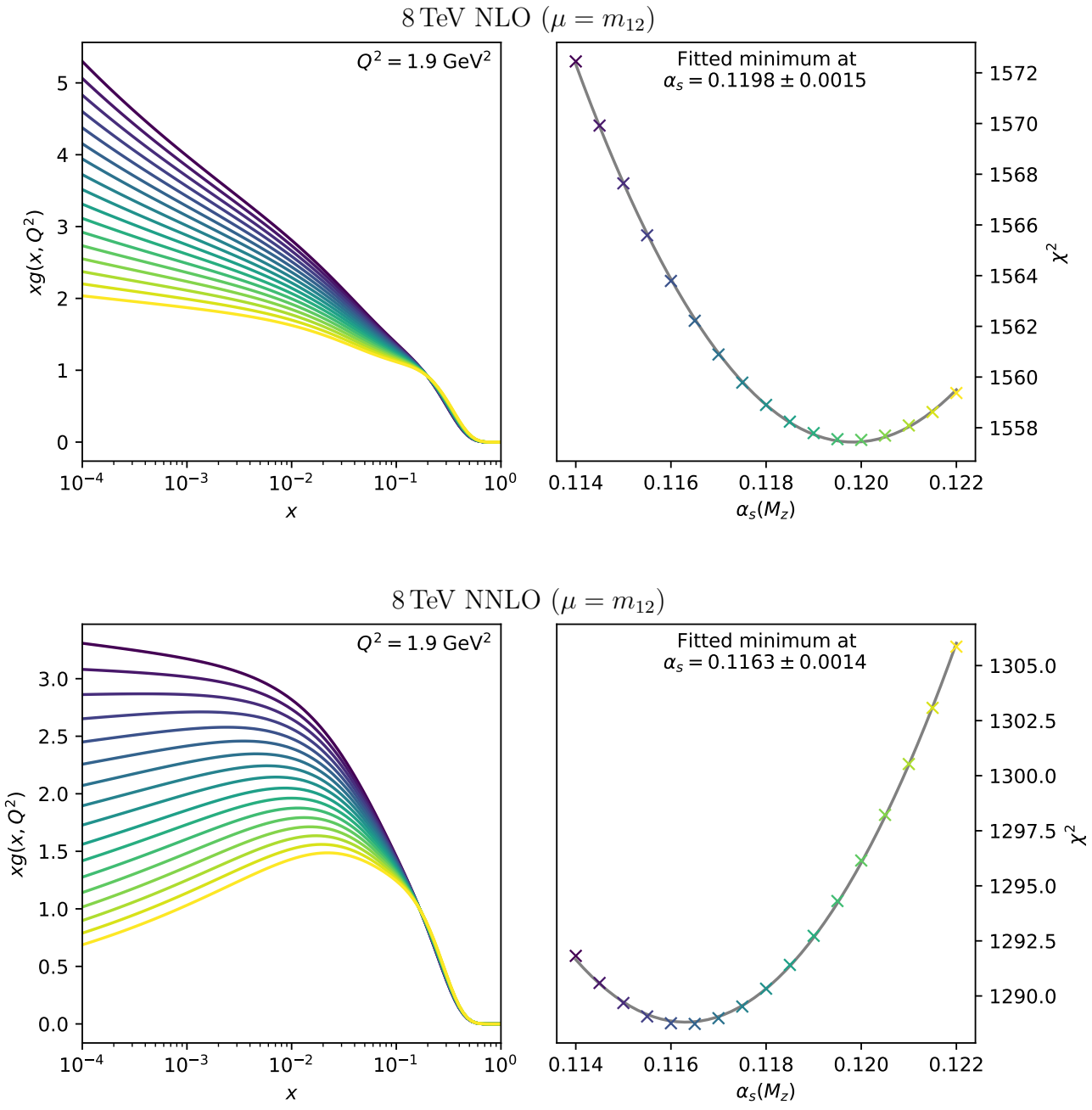


Figure 5.10: Profile of $\alpha_s(M_z)$ for NLO (top) and NNLO (bottom). Fits with the central scale $\mu = m_{12}$ are done for a series of fixed values of $\alpha_s(M_z)$ and the corresponding minimum χ^2 values plotted on the left panels. For each fit, the emerging gluon distribution is plotted on the right hand side in the corresponding color.

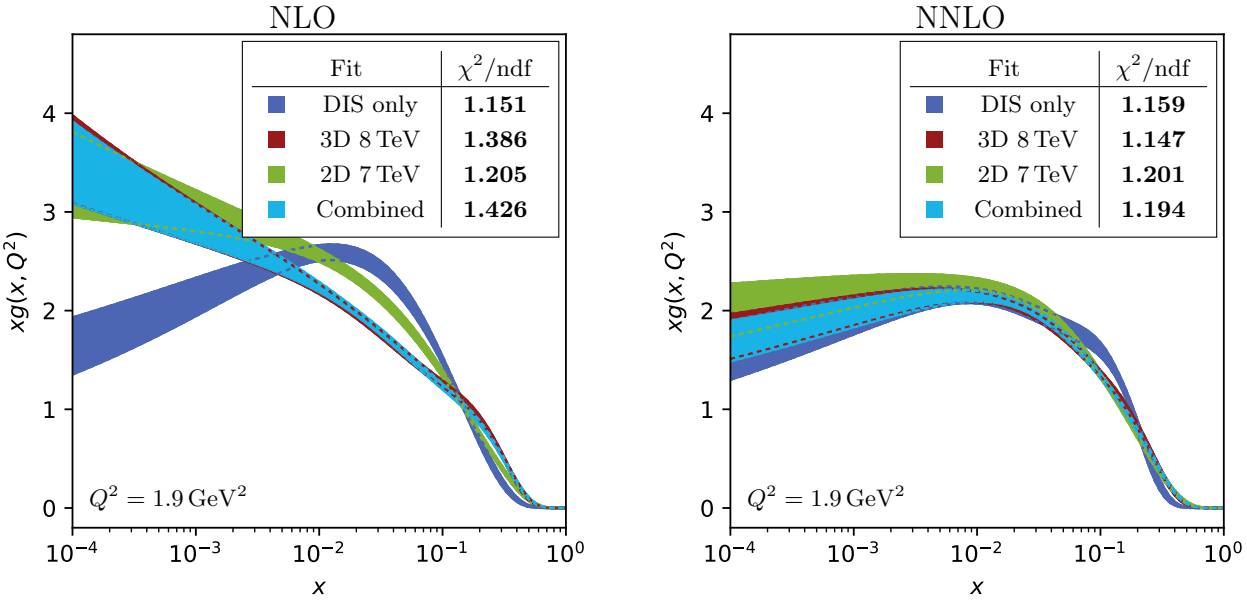


Figure 5.11: Gluon PDFs corresponding to the NLO (left) and NNLO (right) fits with different differential dijet cross section measurements included. Only experimental uncertainties are shown.

8 TeV data included from Section 5.2 are also drawn into Figure 5.11. At NLO both fits with jet data included result in notably higher χ^2/ndf values than the inclusive DIS only fit. However, at NLO, the dijet mass is not the preferred central scale choice for the theory calculations to describe either of the dijet datasets. Ref [39] used the scale definition of $\mu_R = \mu_F = p_{\mathrm{T},1} e^{0.3y^*}$ („ptmax") that better describes the 8 TeV triple-differential dijet cross sections at NLO. The theory calculations for the 7 TeV double-differential measurement, that were presented in the corresponding publication, use the average transverse momentum $\mu_R = \mu_F = p_{\mathrm{T,ave}}$ as central scale [18].

At NNLO, the shapes of all the fits' gluon PDFs agree much better. In comparison to the fit including 3D 8 TeV data from section 5.2, the one with 7 TeV 2D data yields only a small improvement in terms of χ^2/ndf. Apparently, the 7 TeV 2D theory calculation describes the measured data only somewhat better in NNLO than in NLO. A possible explanation for this may be that the 7 TeV double-differential data includes fewer data points (bins) and these have higher uncertainties than in the 8 TeV triple-differential measurement.

To get a better estimation of the PDFs, it is desirable to include more than one dijet dataset into the PDF fits. Therefore, in addition to the fits with inclusive DIS plus one single dijet data set, fits have been done with all three datasets included. Care must be taken because both datasets are measured by CMS at the LHC and there might be correlations between the measured cross sections in both datasets. On the other side, it can be argued, that the measurements have been made during different run periods

3D 8 TeV squared sum		2D 7 TeV squared sum		Combined squared sum	
NLO	NNLO	NLO	NNLO	NLO	NNLO
40.86	8.78	7.01	6.76	58.40	19.37

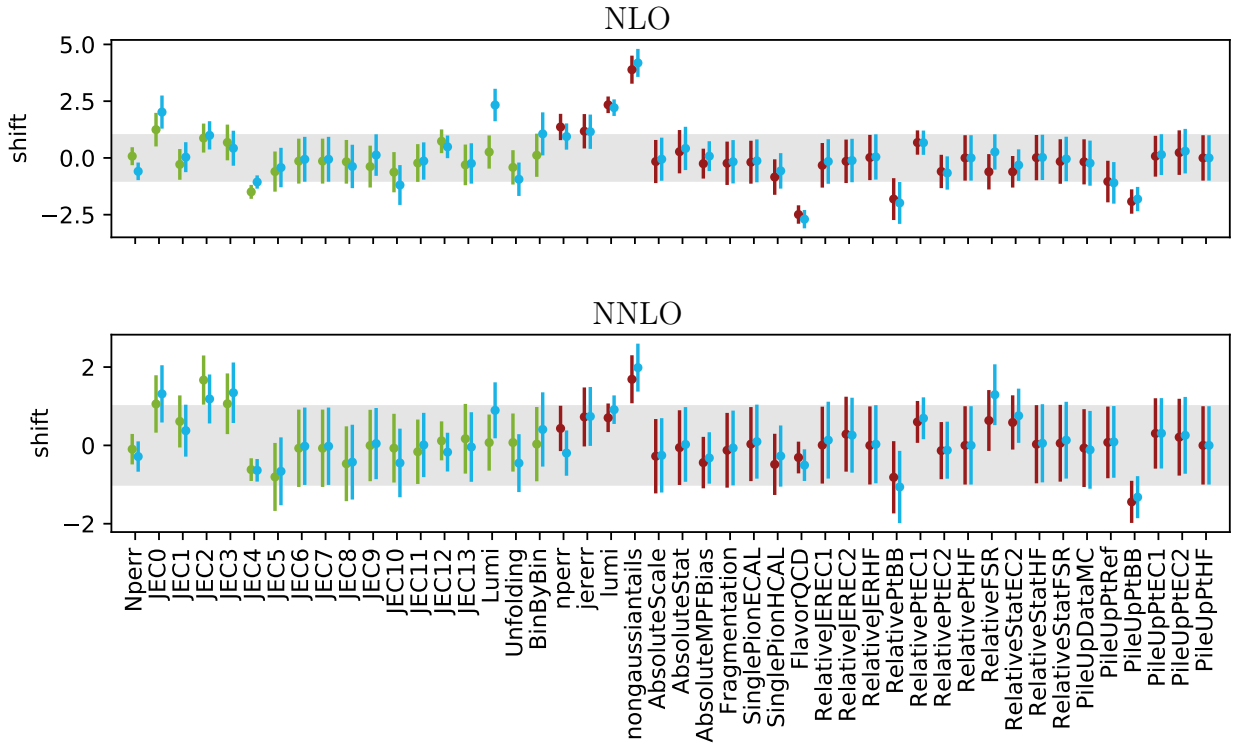


Figure 5.12: Fitted values on the correlated systematic uncertainty sources for the NLO and the NNLO fits with different differential dijet data included. The gray band marks the one sigma region. The tables on the top list the sums of the squared shift values.

of the LHC and should share only a small subset of the systematic uncertainties. In the combined fits presented here, such possible correlations are ignored without further investigation.

The results of these combined fits are also shown in Figure 5.11. It is interesting, that in the combined fit the triple-differential measurement at 8 TeV seems to dominate the fit, at least concerning the gluon distribution. This presumably originates from the smaller uncertainties in the 8 TeV measurement compared to the 7 TeV measurement [18, 40]. Details on these fits such as the final fitted parameters or the detailed partial χ^2 values can be found in the appendix Section C.4.

Figure 5.12 shows the individual shifts for each of the uncertainty sources in the 7 TeV double-differential and 8 TeV triple-differential dijet data sets. There are 18 such sources for the 7 TeV data which leads to an expected value of $E[\chi^2_{18}] = 18$ for their sum

Table 5.3: Partial χ^2 values for the different fits at NLO

	n_{data}	HERA I+II Dis only	CMS dijets 3D 8 TeV	CMS dijets 2D 7 TeV	Combined
HERA I+II	1016	1106.14	1181.30	1124.63	1178.64
CMS dijets 3D 8 TeV	122	–	275.69	–	283.74
CMS dijets 2D 7 TeV	54	–	–	94.24	97.36
correlated χ^2		50.96	115.15	60.71	132.39
log penalty χ^2		-2.98	-13.24	-6.07	-11.22
combined		1154.12	1558.90	1273.52	1680.89
ndf		1003	1125	1057	1179
p-value		6.13×10^{-4}	1.26×10^{-16}	1.38×10^{-6}	2.48×10^{-20}
combined χ^2/ndf		**1.151**	**1.386**	**1.205**	**1.426**

Table 5.4: Partial χ^2 values for the different fits at NNLO

	n_{data}	HERA I+II Dis only	CMS dijets 3D 8 TeV	CMS dijets 2D 7 TeV	Combined
HERA I+II	1016	1109.15	1119.54	1112.62	1120.46
CMS dijets 3D 8 TeV	122	–	111.58	–	111.93
CMS dijets 2D 7 TeV	54	–	–	93.31	103.04
correlated χ^2		55.48	63.17	63.93	73.81
log penalty χ^2		-1.74	-3.95	-0.49	-1.37
combined		1162.89	1290.33	1269.37	1407.88
ndf		1003	1125	1057	1179
p-value		3.24×10^{-4}	4.16×10^{-4}	6.63×10^{-6}	4.25×10^{-6}
combined χ^2/ndf		**1.159**	**1.147**	**1.201**	**1.194**

of squares. In the NLO as well as in the NNLO fit, the fitted values of 7.01 and 6.76 are much smaller though. For the combined fit in NLO the observed value of the sum of squared shifts, 58.40 is well around the expected value of the number of sources, 46, but is dominated by a few outliers, like for example the „nongaussiantails" uncertainty source. At NNLO again, the observed value is much smaller than the expected one. The behavior of these outliers also propagates itself to the overall sum of squared shifts, which can be found in the „correlated χ^2" row in Tables 5.3 and 5.4, where the values are almost twice as large in NLO as in NNLO for the 8 TeV only and the combined fit.

The outlier values of the shifts in the 8 TeV 3D data are shown on the right-hand side in Figure 5.12 and have already been discussed in detail in Section 5.2. They are related to the „curved" deviation of the NLO prediction from the measured data in the central rapidity bin (see Figure 5.13). If the double-differential data at 7 TeV is included in the combined fit, there is almost no change for the 3D data's shifts (red points compared

with cyan points in Figure 5.12). The shifts of the 7 TeV data are shown on the left-hand side. Their exact meaning can be found in the corresponding publications [27, 29]. For the fit including only the 7 TeV data apart from the DIS data (green points) there are no extreme outliers but if the 8 TeV triple-differential data is included in the combined fit (cyan points) there is a prominent outlier in NLO, namely the „Lumi" uncertainty source. A possible cause of this can be found by looking into Figure 5.13 where a direct comparison between the different fits' predictions and the corresponding shifted data is shown. The top part shows this comparison for two rapidity bins of the 7 TeV data's phase space. It was already mentioned, that the combined fit is probably dominated by the 8 TeV data. Apparently, the NLO predictions from this combined fit (cyan crosses) yield higher values than those of the fit with only 7 TeV data (green crosses). Such a global tendency to higher values can be compensated by a high value of the luminosity uncertainty source, which is exactly what can be seen in Figure 5.12. Interestingly, at next-to-next-to-leading order, this difference gets much smaller and the fitted „Lumi" correlated uncertainty drops into the 1-sigma region.

It is difficult to measure the concordance or consistency of two measurements with totally different observables, because they can not be compared directly, but rather must both be tested against a physical model. The physical model in this case depends on the PDF parametrization and fixed order pQCD and therefore includes some assumptions. The difficulty to quantify the goodness-of-fit and interpret the final χ^2 values has already been discussed in Section 5.2. The χ^2 values presented in Figure 5.11 and Table 5.3 can not be used in statistical tests like for example a χ^2 test (the corresponding p-values are calculated in Table 5.3) because important theoretical uncertainties, like the parametrization-, model- and scale uncertainties are not included in the χ^2 definition. In fact, the p-values shown in Tables 5.3 and 5.4 are all well beyond the 3σ bound and given the measured data points as well as their uncertainty estimations are correct, one could safely reject all the fits with at least a 3σ significance. Nevertheless, the χ^2 values or p-values can be used to compare the fit quality of different fits.

In the case of PDF fitting, the goal is not to find the „true" PDFs though, but rather to create a good empirical approximation to the proton structure. Therefore datasets may very well be called consistent if a fitted PDF can be found, that describes the combined data about as well as the best fitting PDFs for each of the datasets alone. As the partial χ^2 values in Tables 5.3 (NLO) and 5.4 (NNLO) reveal, this is the case in the fits done here. Comparing the second and the third column (3D 8 TeV fit and 2D 7 TeV fit) with the combined fit in the fourth column, one can see that in NLO as well as in NNLO the partial χ^2 values do not change dramatically if combining both the datasets. This means that the best fitting theory prediction for the combined data, namely the prediction of the combined fit, describes either of the datasets almost as well as the fits to one of them alone. If the datasets were inconsistent, then one would expect a significant increase in the combined fit's χ^2 values for at least one of the datasets compared to the corresponding single dataset fit.

To conclude, we have found, that in NLO there are some small incoherences between the 7 TeV and the 8 TeV data which can, for example, be seen in the „Lumi" shift outlier discussed above or in the fitted gluon PDFs' shapes in Figure 5.11. With the

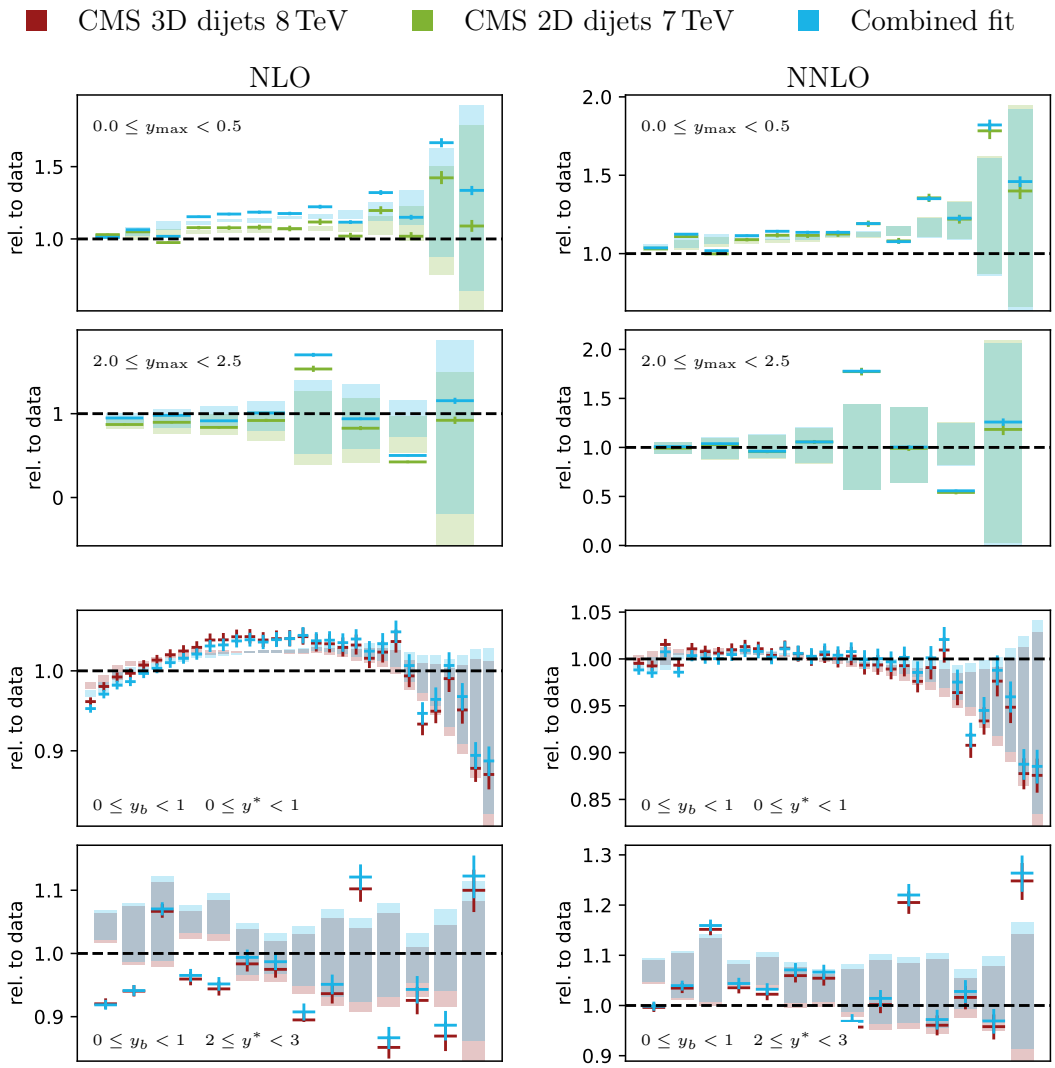


Figure 5.13: Theory (crosses) and shifted data (bands) relative to the central data for the inclusive DIS plus 8 TeV ($\mu = m_{12}$) fit (dark red ■), for the DIS plus 7 TeV fit (green ■) and the combined inclusive DIS 8 TeV and 7 TeV fit (cyan ■) for NLO (left) and NNLO (right). The theory predictions are calculated by evaluating the `FastNLO` table with the fitted PDFs. The shifted data bands are constructed by shifting the central data point with the fitted correlated uncertainty sources (shifts) from the corresponding fit. The top section shows two rapidity bins of the 7 TeV 2D dataset, whereas in the bottom section two rapidity bins of the 8 TeV 3D dataset are shown.

inclusion of the NNLO contribution, these discrepancies decrease drastically. The 8 TeV data apparently dominates the combined fit but nevertheless this combined fit yields reasonable predictions also for the 7 TeV data.

While the fits in this work are the first PDF fits to differential dijet data with full NNLO calculations, there have already PDF fits to dijet data been performed with K-factor based NNLO contributions by the NNPDF group. The corresponding publication presents fits to the same CMS 7 TeV double-differential and 8 TeV triple-differential dijet datasets, that are used in this work. Although they do not just use the HERA I+II combined inclusive DIS dataset as a base, but rather the complete NNPDF baseline data, the results are in good agreement with the key findings in this Section. They also observe combined fit dominated by the 8 TeV data as well as almost good consistency between both the datasets (even already at NLO) [3].

6. Outlook

In this thesis, PDF fits with and without the strong coupling constant included as parameter have been presented. First, previous PDF fits at NLO with triple-differential dijet data at 8 TeV have been reproduced. It was shown, that the newest version of `xFitter` yields the same results that have been found by Ref. [39]. The inclusion of electroweak corrections that had not yet been available to Ref. [39] does not have any significant impact.

There are new `FastNLO` tables for the 8 TeV triple-differential measurement, which are calculated using the `NNLOJET` program. In Section 5.1 a fit with the old `NLOJet++` tables, that were used by Ref. [39] and in the CMS publication for the 8 TeV dijet data [40], have been compared to a fit with these new `NNLOJET` tables (only evaluated at NLO). The difference between both results is negligible.

The main data of interest in this work have been the 8 TeV triple-differential dijet cross section measurement published Ref. [40]. In this publication, a PDF fit to the data, as well as a determination of $\alpha_s(M_z)$ have already been performed with NLO theory. The NNPDF global PDF fitting group recently fitted PDFs to this dataset using NLO theory supplemented with NNLO K-factors. They already found that the NNLO contribution is important for the compatibility of this dataset with the rest of their data [3]. In this work, fits with full next-to-next-to-leading order theory (calculated with the `NNLOJET` program) have been presented. In accordance to the findings of Ref . [3] we observe a drastic improvement in fit quality and consistency if the NNLO contribution is added. A comparison between fits with two different central scale definitions shows, that in NLO there are large differences in the results. At NNLO these discrepancies vanish for the most part and the results are compatible within the uncertainties even though not all the relevant uncertainties have been considered yet. In addition to more consistency between the different central scales, the results at NNLO show much decreased scale uncertainties. These insights provide strong evidence, that the NNLO contribution is crucial for consistent PDF fits including the 8 TeV triple-differential dijet measurement.

Differential dijet cross section calculations are in addition to the PDFs sensitive to the strong coupling constant. Determinations of $\alpha_s(M_z)$ from jet data have been done in numerous publications and are one of the standard use cases for jet related measurements. In this thesis, simultaneous fits of PDFs and $\alpha_s(M_z)$ have been performed. The best-fitting values are shown in Table 6.1. Notable are the slightly smaller

Table 6.1: Final $\alpha_s(M_z)$ values obtained from simultaneous PDF plus $\alpha_s(M_z)$ fits

	NLO	NNLO
$\mu = p_{\mathrm{T},1} e^{0.3y^*}$	$0.1191 \pm 0.0015(\mathrm{exp})^{+0.0028}_{-0.0016}(\mathrm{scale})$	$0.1155 \pm 0.0012(\mathrm{exp})^{+0.0008}_{-0.0017}(\mathrm{scale})$
$\mu = m_{12}$	$0.1198 \pm 0.0015(\mathrm{exp})^{+0.0021}_{-0.0021}(\mathrm{scale})$	$0.1163 \pm 0.0013(\mathrm{exp})^{+0.0010}_{-0.0004}(\mathrm{scale})$

values in NNLO compared to the NLO values, which is expected. Like the PDFs at NNLO, the $\alpha_s(M_z)$ values obtained from the NNLO fits show a significantly smaller scale uncertainty compared to the corresponding NLO values.

The uncertainties on the $\alpha_s(M_z)$ values are related to the PDF uncertainties, especially to the one of gluon PDF. A strong correlation between $\alpha_s(M_z)$ and the gluon PDF in the dijet cross section calculation has been observed in this work. Including $\alpha_s(M_z)$ as a free parameter in the simultaneous fits increases the experimental uncertainty of the gluon PDF compared to the fits with a fixed value for the strong coupling constant.

In addition to the 8 TeV triple-differential dijet measurement, there is a double-differential dijet cross section measurement at 7 TeV published by CMS [18]. Like for the 8 TeV measurement, full NNLO calculations for this dataset have now become available. In this work, PDF fits to either of the datasets alone, as well as to the combination of both are presented. Differences between the datasets can be seen in the fit results at NLO but upon inclusion of the NNLO contribution, these differences decrease drastically. Both the datasets are consistent and can be fitted together without complications in NNLO, although the 8 TeV data dominate the combined fit.

Within the scope of this work, there have not been made any estimations on PDF parametrization or model uncertainties, that remain to be investigated. This should be kept at the back of one's mind when looking at the results presented here. Especially the PDF parametrization uncertainties are difficult to estimate, as there is an infinite number of possible parametrizations.

The problem of determining the proton structure remains crucial for all kinds of precision measurements at hadron colliders. Jet related data like differential measurements allow fits of the PDFs as well as the strong coupling constant if they are supplemented by theory calculations with sufficient precision. Here, it is shown, that many of the problems arising in PDF fits with NLO theory can be fixed with NNLO calculations. Programs like `NNLOJET` and `FastNLO` together with large computing resources provide the basis for higher-order jet cross section calculations and will ultimately lead to a more detailed knowledge of the proton structure.

A. Patches to xFitter

The `xFitter` program uses `FastNLO` code to evaluate QCD interpolation tables. However, `FastNLO` is not dynamically linked into `xFitter` but rather a slightly modified version of `FastNLO` is distributed along with the `xFitter` source code and linked statically at compile time into the `xFitter` executable. The version of `FastNLO`, that is shipped with the latest stable release v2.0.1[1] of `xFitter` at the time of writing, is outdated and not capable of reading and evaluating the NNLO tables created using `NNLOJET`.

Basically the changes that were made to the official `xFitter` version 2.0.1 are the following:

1. Patch the `FastNLO` part in the `xFitter` source distribution by copying a recent version of `FastNLO` into the corresponding directory. This allows `xFitter` to read *flexible-scale* tables as introduced in `FastNLO` version 2, including NNLO contributions.
2. Enable linking to `zlib` in the `xFitter` build mechanism. This allows `xFitter` to read gzipped tables which saves a lot of disk space.
3. Add an additional setting in the `xFitter` steering file, which allows for selecting the desired central scale at which the flexible scale tables are evaluated. The options are `DataSetMurDef` and `DataSetMufDef` in the `Scales` namelist in the steering file (see appendix chapter B). The options will be ignored and should be set to -1 if the dataset is not linked to `FastNLO` tables for theory predictions. If the dataset uses predictions from `FastNLO`, the values select the central renormalization and factorization scale definitions. The information of which value corresponds to which scale definition can be found in the `FastNLO` table headers (use e.g. `FastNLO`'s `fnlo-tk-cppread` program for a nice printout).

The patched version of `xFitter` has been published online and can be downloaded from the CERN gitlab at

`https://gitlab.cern.ch/jastark/xfitter/-/tree/release_2.0.1` .

After downloading and compiling, this patched version of `xFitter` should be able to reproduce the results in Chapter 5, if provided with the correct input files.

[1] Recent xFitter releases are available for download at `https://www.xfitter.org/xFitter/`

B. Steering cards

In the following, the exact configuration for most of the fits in chapter 5 is listed. `xFitter` uses three configurations files, namely the main steering file `steering.txt`, the Minuit steering file `minuit.in.txt` and an additional configuration file for several physical constants `ewparams.txt`. All three of these files must be present in the current directory if `xFitter` is started.

B.1 steering.txt

The main steering file specifies most of the `xFitter` run options. It is in the Fortran Namelist format and most of the important options are documented in the `xFitter` manual [45]. There are however two options that are important to the fits in chapter 5, that are not documented and will be explained here.

The `QCDNUM` namelist specifies options for the PDF evolution grid, that will be used to evolve PDFs from one factorization scale to another. The important option here is the undocumented `QARR(2)` field, which specifies the upper limit of the evolution grid. Its default value is 2.05×10^8 (in GeV^2) which however is too small for the 13 TeV calculations and for some of the scale variations at 8 TeV. Therefore the value was increased to 1.0×10^9 which was sufficient for all the fits done in this work.

The `Scales` namelist specifies scale options for the theory calculations. The `DataSetMuR` and `DataSetMuF` fields specify the renormalization and factorization scale factors for the theory calculations for each of the input files, and are varied between 0.5, 1.0 and 2.0 if analyzing scale uncertainties. The `DataSetMurDef` and `DataSetMufDef` fields have been added with one of the patches described in appendix chapter A. These values specify the central scale definition for the `FastNLO` flexible-scale tables and are ignored for non-`FastNLO` theory calculations.

In the following listing a typical `steering.txt` configuration file as used in the fits in chapter 5 is presented:

steering.txt

```
1 &InFiles
2     NInputFiles = 13
3     InputFileNames(1)  = 'datafiles/1506.06042/HERA1+2_NCep_920.dat'
4     InputFileNames(2)  = 'datafiles/1506.06042/HERA1+2_NCep_820.dat'
```

```
5     InputFileNames(3)  = 'datafiles/1506.06042/HERA1+2_NCep_575.dat'
6     InputFileNames(4)  = 'datafiles/1506.06042/HERA1+2_NCep_460.dat'
7     InputFileNames(5)  = 'datafiles/1506.06042/HERA1+2_NCem.dat'
8     InputFileNames(6)  = 'datafiles/1506.06042/HERA1+2_CCep.dat'
9     InputFileNames(7)  = 'datafiles/1506.06042/HERA1+2_CCem.dat'
10    InputFileNames(8)  = 'datafiles/1705.02628/CMS_TD2Jet_yb0_ys0.dat'
11    InputFileNames(9)  = 'datafiles/1705.02628/CMS_TD2Jet_yb0_ys1.dat'
12    InputFileNames(10) = 'datafiles/1705.02628/CMS_TD2Jet_yb0_ys2.dat'
13    InputFileNames(11) = 'datafiles/1705.02628/CMS_TD2Jet_yb1_ys0.dat'
14    InputFileNames(12) = 'datafiles/1705.02628/CMS_TD2Jet_yb1_ys1.dat'
15    InputFileNames(13) = 'datafiles/1705.02628/CMS_TD2Jet_yb2_ys0.dat'
16 &End
17
18 &InCorr
19    NCorrFiles = 6
20    CorrFileNames(1)   = 'datafiles/1705.02628/CMS_TD2Jet_yb0_ys0.corr'
21    CorrFileNames(2)   = 'datafiles/1705.02628/CMS_TD2Jet_yb0_ys1.corr'
22    CorrFileNames(3)   = 'datafiles/1705.02628/CMS_TD2Jet_yb0_ys2.corr'
23    CorrFileNames(4)   = 'datafiles/1705.02628/CMS_TD2Jet_yb1_ys0.corr'
24    CorrFileNames(5)   = 'datafiles/1705.02628/CMS_TD2Jet_yb1_ys1.corr'
25    CorrFileNames(6)   = 'datafiles/1705.02628/CMS_TD2Jet_yb2_ys0.corr'
26 &End
27
28 &CovarToNuisance
29    LConvertCovToNui = .false.
30    Tolerance = 0.0
31    LSubtractStat = .false.
32 &End
33
34 &QCDNUM
35   Read_QCDNUM_Tables = .true.
36   QARR(2) = 1.0D9
37 &End
38
39 &OutDir
40    OutDirName = 'output'
41 &End
42
43 &Scales
44    DataSetMuR    = 7*1.0, 6*1.0
45    DataSetMuF    = 7*1.0, 6*1.0
46    DataSetMurDef = 7*0, 6*0
47    DataSetMufDef = 7*0, 6*0
48 &End
49
50 &xFitter
51    RunningMode = 'Fit'
52    TheoryType = 'DGLAP'
53    Order  = 'NLO'
54    Q02     = 1.9
55    HF_SCHEME = 'RT␣OPT'
```

```
56      PDFType = 'proton'
57      PDFStyle = 'HERAPDF'
58      CHI2SettingsName = 'StatScale', 'UncorSysScale', 'CorSysScale',
        'UncorChi2Type', 'CorChi2Type'
59      Chi2Settings     = 'Poisson'  , 'Linear',        'Linear'     ,
        'Diagonal'     , 'Hessian'
60      Chi2ExtraParam = 'ExtraSystRescale'
61      LUseAPPLgridCKM = True
62      LDEBUG     = False
63  &End
64
65  &ExtraMinimisationParameters
66      name  = 'alphas', 'fs', 'fcharm'
67      value = 0.1180  , 0.4 , 0.0
68      step  = 0.0     , 0.0 , 0.0
69  &End
70
71  &Output
72    DoBands = True
73    DoBandsSym = False
74    Q2VAL = 1.9, 3.0, 4.0, 5., 10., 100., 6464, 8317
75    OUTNX = 101
76    OUTXRANGE = 1E-4, 0.9999
77  &End
78
79  &Cuts
80     ProcessName(1)     = 'NC␣e+-p'
81     Variable(1)        = 'Q2'
82     CutValueMin(1)     = 10.0
83     CutValueMax(1)     = 1000000.0
84
85     ProcessName(2)     = 'NC␣e+-p'
86     Variable(2)        = 'x'
87     CutValueMin(2)     = 0.000001
88     CutValueMax(2)     = 1.0
89
90     ProcessName(3)     = 'CC␣e+-p'
91     Variable(3)        = 'Q2'
92     CutValueMin(3)     = 10.0
93     CutValueMax(3)     = 1000000.0
94
95     ProcessName(4)     = 'CC␣e+-p'
96     Variable(4)        = 'x'
97     CutValueMin(4)     = 0.000001
98     CutValueMax(4)     = 1.0
99  &End
```

B.2 minuit.in.txt

In this Minuit run card, parameters are specified and given start values (α_s is given in `steering.txt` though) and the minimization strategy is defined. The listing below shows a typical Minuit run card as used in the fits in chapter 5. First MIGRAD is run with a maximal iteration bound of 20000, than the hesse matrix is estimated. Note that `xFitter` may issue some additional MINUIT commands after `return` depending on the settings for uncertainty estimation. The MIGRAD algorithm, like specified in the listing below, works for most of the fits in chapter 5. For the fits including $\alpha_s(M_z)$ however it sometimes runs out of the allowed range for $\alpha_s(M_z)$ due to a bad estimation of the gradient (see Section 4.2). This can be fixed in most of the cases by limiting the $\alpha_s(M_z)$ parameter in a first call to MIGRAD and free it again in a second call, like shown in the commented lines 18-20 in the listing below.

minuit.in.txt

```
1  set title
2  new ETPPDF
3  parameters
4      2  'Bg'        0.121118     0.039061
5      3  'Cg'        9.947400     1.121916
6      5  'Eg'       13.288705     5.884189
7     12  'Buv'       0.426782     0.028425
8     13  'Cuv'       3.995199     0.049824
9     14  'Duv'      19.081485     2.830706
10    22  'Bdv'       1.024478     0.081755
11    23  'Cdv'       4.645417     0.366832
12    33  'CUbar'    15.097878     0.734351
13    34  'DUbar'    28.657550     3.685604
14    41  'ADbar'     0.219646     0.023236
15    42  'BDbar'    -0.127689     0.015826
16    43  'CDbar'     8.044305     1.591946
17
18 *set limits 101 0.100 0.130
19 *migrad 20000
20 *set limits 101
21 migrad 20000
22 hesse
23 save
24 return
```

B.3 ewparam.txt

The `ewparam.txt` sets several electroweak parameters as well as quark masses and other constants. In all the fits in this work, there were no changes to the standard `ewparam.txt` that is distributed with `xFitter` version 2.0.1. listed below.

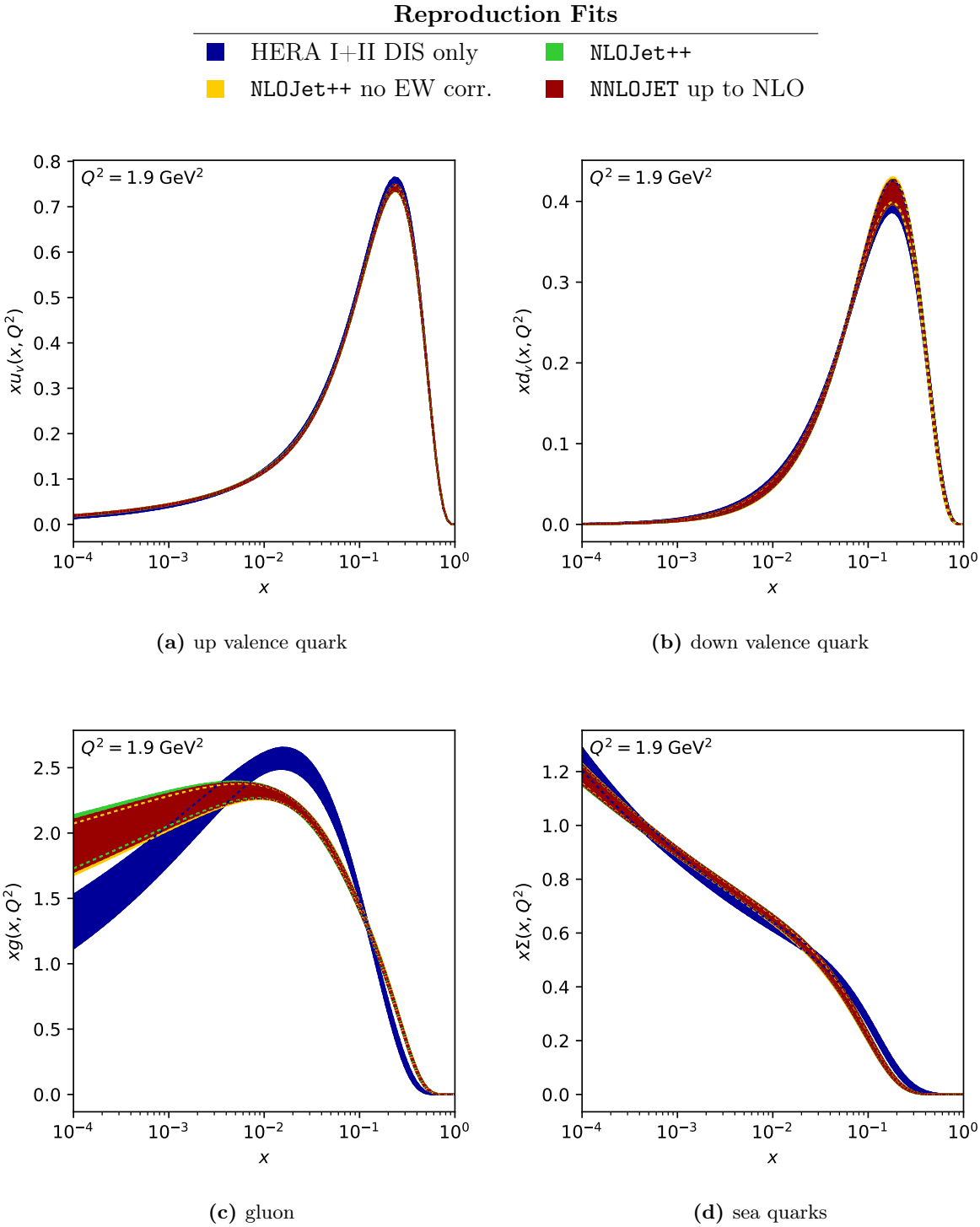
ewparam.txt

```
*
*  Electroweak parameters
*

* Note: DIS uses on-shell electroweak scheme,
* DY uses GFermi scheme

&EWPars
  ! Choice of EW scheme: 0 - alpha(0), 1 - G_mu, 2 - running alpha_EM
  ! EWSchemeFlag = 0

  ! 1/137.035999074(44) = 7.29735d-3
  alphaem  = 7.29735d-3
  gf       = 1.16638d-5
  sin2thw  = 0.23127d0
  ! alphas = 0.1176d0
  convfac  = 0.389379338d9

  ! boson masses
  mw     = 80.385d0
  mz     = 91.1876d0
  mh     = 125.9d0

  ! widths
  wz     = 2.4952d0
  ww     = 2.085d0
  wh     = 1d-3
  wtp    = 2.0d0

  ! charges
  ! euq  = 0.666666666666667d0
  ! edq  = -0.333333333333333d0

  ! CKM ( todo: add Vub & Vcb to DY)
  Vud    = 0.97427d0
  Vus    = 0.2254d0
  Vub    = 0.00358d0
  Vcd    = 0.22520d0
  Vcs    = 0.97344d0
  Vcb    = 0.04156d0
  Vtd    = 0.00872d0
  Vts    = 0.04076d0
  Vtb    = 0.999133d0

  !*** fermion masses

  ! lepton masses
  men    = 1d-10
  mel    = 0.510998928d-3
  mmn    = 1d-10
```

```
51    mmo     = 0.1056583715d0
52    mtn     = 1d-10
53    mta     = 1.77682d0
54
55   ! Light quark masses:
56    mup     = 0.06983d0
57    mdn     = 0.06983d0
58    mst     = 0.150d0
59
60   ! Heavy quark masses:
61    mch     = 1.43d0
62    mtp     = 173d0
63    mbt     = 4.5d0
64
65  &end
```

C. Fit details

C.1 Reproduction fits



(a) up valence quark

(b) down valence quark

(c) gluon

(d) sea quarks

Figure C.1: PDF overview for the reproduction fits from Section 5.1

Table C.1: Fitted parameters of the reproduction fits from Section 5.1

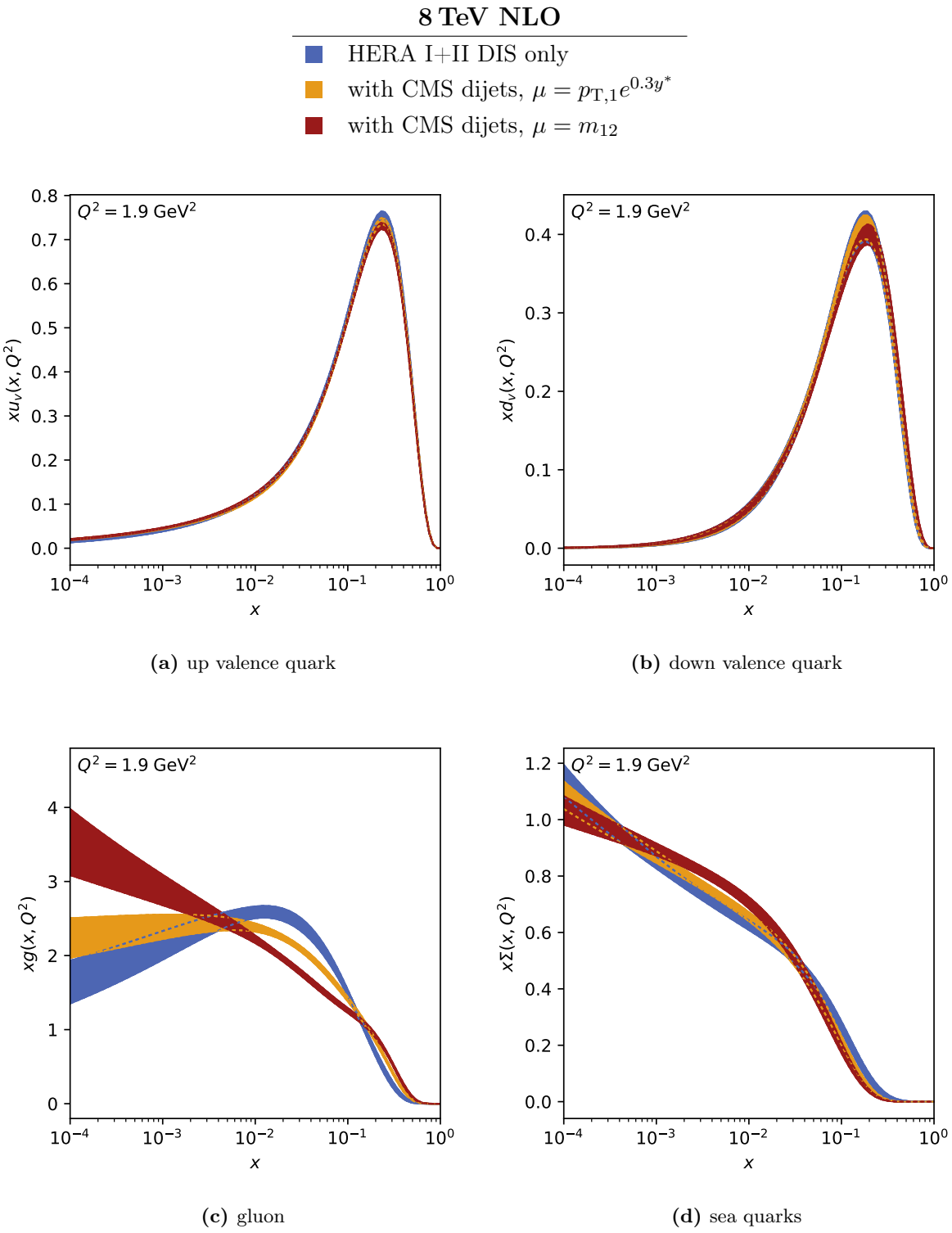
Par.	HERA I+II Dis only	NLOJet++ no EW corr.	NLOJET++	NNLOJET up to NLO
B_g	0.166 ± 0.035	0.069 ± 0.019	0.058 ± 0.020	0.061 ± 0.020
C_g	10.907 ± 1.167	8.693 ± 0.520	8.803 ± 0.531	8.754 ± 0.531
E_g	15.995 ± 6.742	23.365 ± 5.548	25.491 ± 5.967	24.510 ± 5.786
B_{u_v}	0.422 ± 0.028	0.359 ± 0.024	0.359 ± 0.023	0.350 ± 0.023
C_{u_v}	3.991 ± 0.050	3.924 ± 0.040	3.913 ± 0.044	3.904 ± 0.043
D_{u_v}	19.587 ± 2.926	25.731 ± 3.570	25.470 ± 3.527	26.967 ± 3.713
B_{d_v}	1.007 ± 0.082	1.081 ± 0.065	1.028 ± 0.065	1.030 ± 0.065
C_{d_v}	4.588 ± 0.363	4.907 ± 0.277	4.632 ± 0.280	4.656 ± 0.283
$C_{\bar{U}}$	14.896 ± 0.719	15.820 ± 0.612	15.613 ± 0.642	15.632 ± 0.642
$D_{\bar{U}}$	31.223 ± 3.706	25.068 ± 2.725	23.676 ± 2.564	23.915 ± 2.550
$A_{\bar{D}}$	0.193 ± 0.017	0.243 ± 0.014	0.251 ± 0.015	0.250 ± 0.015
$B_{\bar{D}}$	-0.152 ± 0.012	-0.123 ± 0.009	-0.118 ± 0.009	-0.119 ± 0.009
$C_{\bar{D}}$	6.937 ± 1.423	13.844 ± 2.434	12.909 ± 2.173	12.871 ± 2.147

Table C.2: Partial χ^2 values for the reproduction fits from Section 5.1

		n_{data}	HERA I+II Dis only	NLOJet++ no EW corr.	NLOJET++	NNLOJET up to NLO
HERA I+II	NCep 920	332	376.65	402.24	401.93	401.92
	NCep 820	63	61.66	62.07	61.64	61.73
	NCep 575	234	196.88	198.65	198.20	198.25
	NCep 460	187	205.18	205.72	205.96	205.91
	NCem	159	216.91	218.69	220.06	220.18
	CCep	39	37.56	38.35	37.44	37.36
	CCem	42	54.81	51.71	52.21	52.02
	combined	1059	1149.65	1177.43	1177.44	1177.37
CMS 8 TeV 3D Dijets		122[1]	–	132.86	108.10	111.42
	correlated χ^2		60.56	90.80	81.10	78.93
	log penalty χ^2		-0.54	-16.04	-10.91	-11.22
	combined	1178[1]	1209.68	1385.05	1355.72	1356.51
	combined χ^2/ndf		**1.160**	**1.200**	**1.164**	**1.164**

[1] Note that in the fit with no electroweak corrections (NLOJet++ no EW corr.) these number decrease to $n_{\text{data}} = 111$ and 1167 respectively due to the $p_{\text{T}} < 1000\,\text{GeV}$ cut

C.2 8 TeV fits at NLO and NNLO



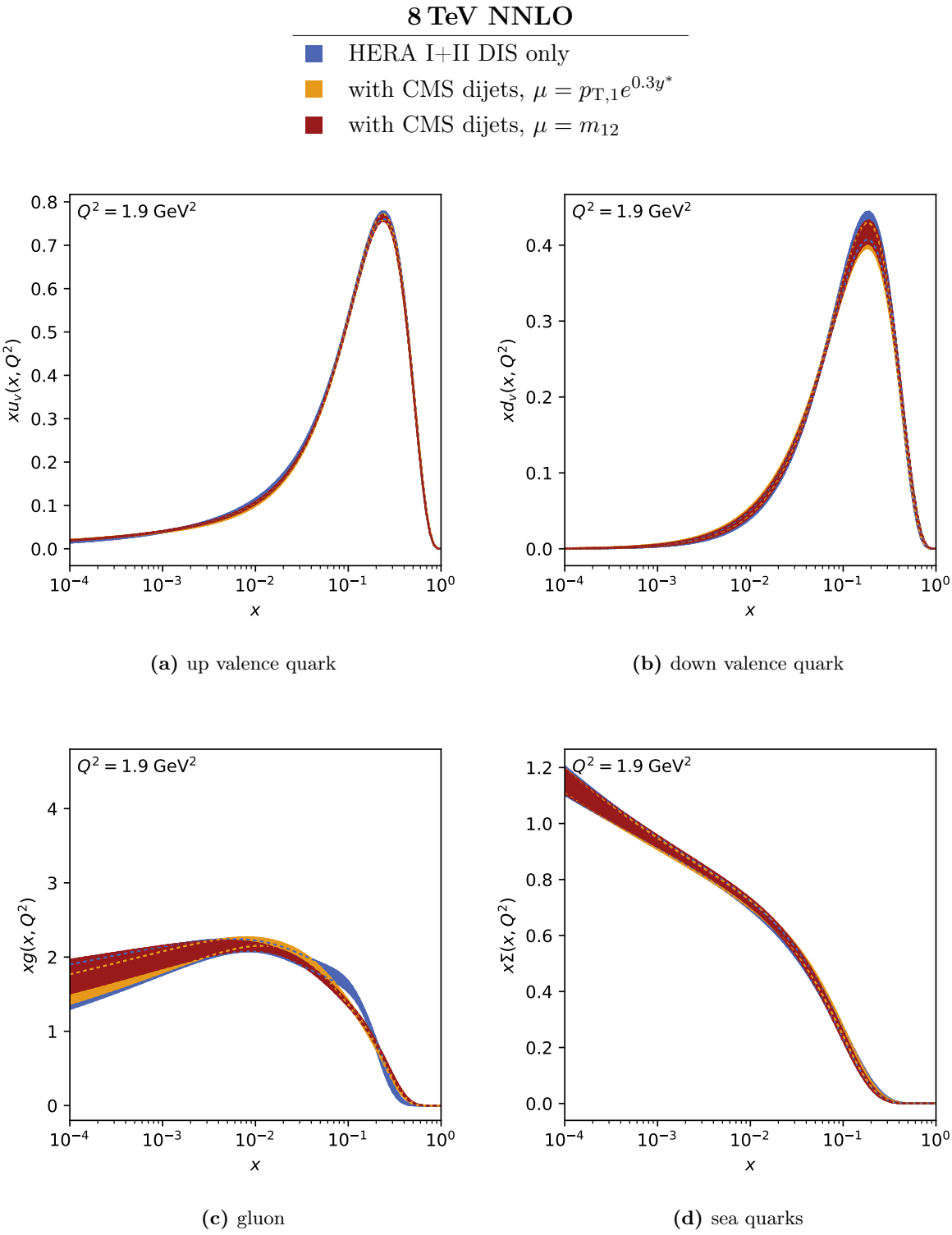
(a) up valence quark

(b) down valence quark

(c) gluon

(d) sea quarks

Figure C.2: PDF overview for the NLO fits in Section 5.2



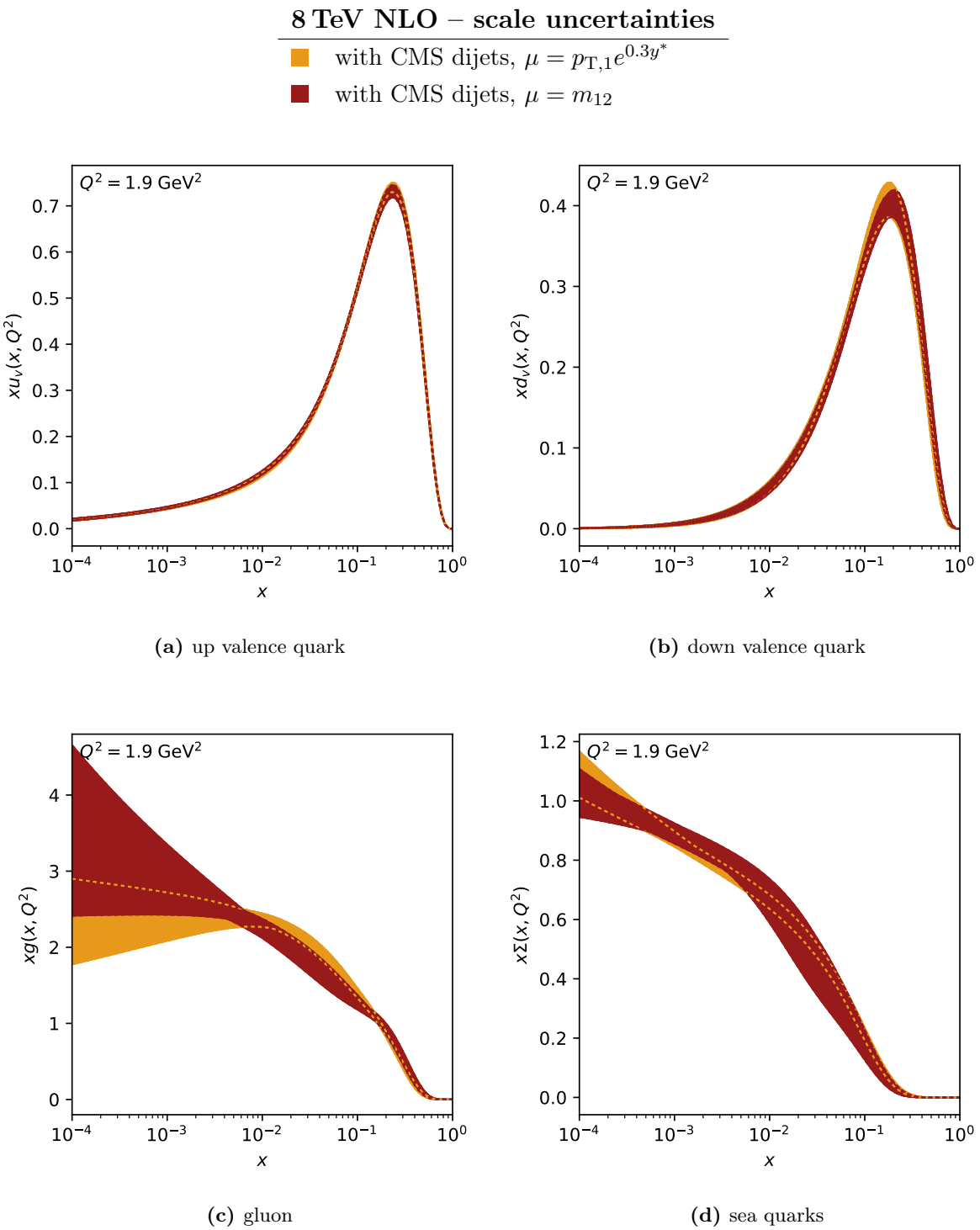
(a) up valence quark

(b) down valence quark

(c) gluon

(d) sea quarks

Figure C.3: PDF overview for the NNLO fits in Section 5.2



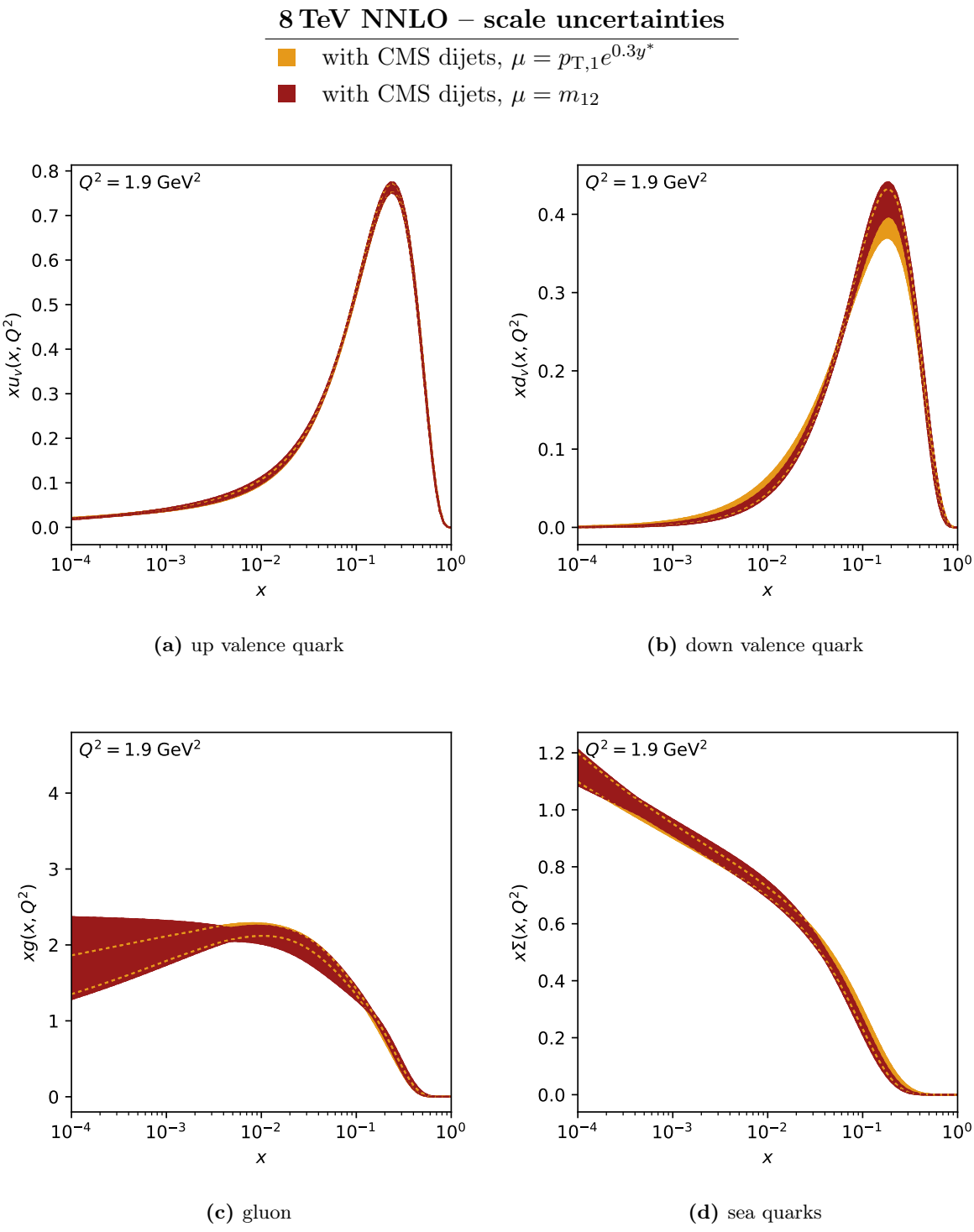
(a) up valence quark

(b) down valence quark

(c) gluon

(d) sea quarks

Figure C.4: PDF overview including scale variations for the NLO fits in Section 5.2



(a) up valence quark

(b) down valence quark

(c) gluon

(d) sea quarks

Figure C.5: PDF overview including scale variations for the NNLO fits in Section 5.2

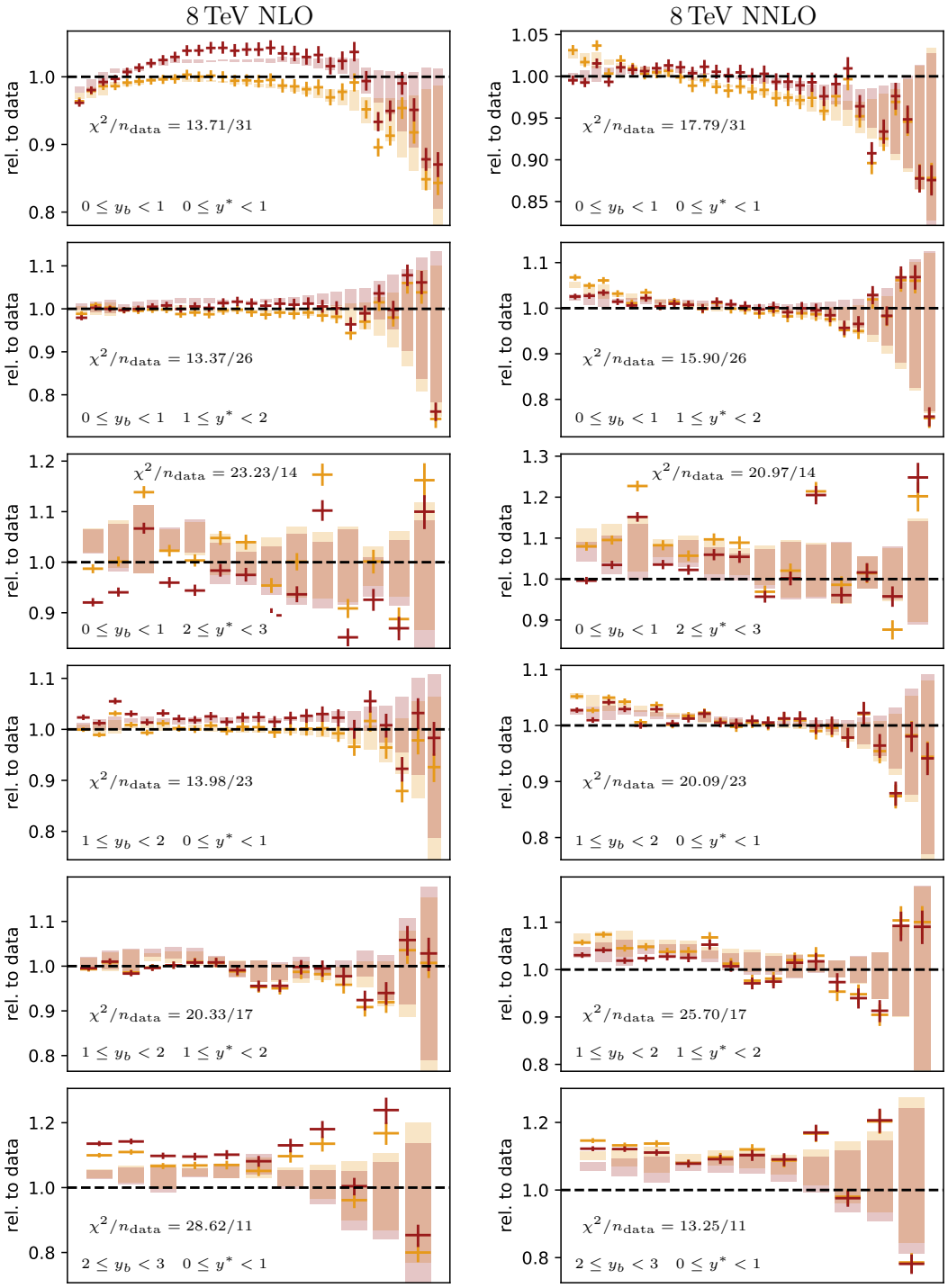


Figure C.6: Theory (crosses) and shifted data (bands) relative to the central data for the $\mu = p_{\text{T},1} e^{0.3y^*}$ (orange ■) and $\mu = m_{12}$ (dark red ■) scale definitions for NLO (left) and NNLO (right). The theory predictions are calculated by evaluating the `FastNLO` table with the fitted PDFs and the corresponding scale definition. The shifted data bands are constructed by shifting the central data point with the fitted correlated uncertainty sources (shifts) from the corresponding fit. The partial χ^2 values for each bin for the „ptmax" central scale fit are also given.

Table C.3: Fitted parameters of the NLO fits in Section 5.2. Only independent fitted parameters are shown. The uncertainties are the experimental ones, determined by diagonalizing the hesse matrix of the χ^2 function in the minimum (see Section 4.5)

Par.	HERA I+II Dis only	with CMS dijets, $\mu = p_{\mathrm{T},1} e^{0.3y^*}$	with CMS dijets, $\mu = m_{12}$
B_g	0.121 ± 0.039	0.032 ± 0.023	-0.084 ± 0.024
C_g	9.947 ± 1.122	8.854 ± 0.519	7.999 ± 0.415
E_g	13.289 ± 5.884	28.632 ± 6.456	48.224 ± 8.733
B_{u_v}	0.427 ± 0.028	0.351 ± 0.023	0.361 ± 0.026
C_{u_v}	3.995 ± 0.050	3.908 ± 0.044	3.922 ± 0.048
D_{u_v}	19.081 ± 2.831	26.741 ± 3.689	23.607 ± 3.518
B_{d_v}	1.024 ± 0.082	1.020 ± 0.065	0.967 ± 0.051
C_{d_v}	4.645 ± 0.367	4.607 ± 0.279	4.046 ± 0.199
$C_{\bar{U}}$	15.098 ± 0.734	15.635 ± 0.661	16.133 ± 0.682
$D_{\bar{U}}$	28.658 ± 3.686	21.644 ± 2.591	15.717 ± 2.585
$A_{\bar{D}}$	0.220 ± 0.023	0.282 ± 0.021	0.377 ± 0.031
$B_{\bar{D}}$	-0.128 ± 0.016	-0.096 ± 0.012	-0.059 ± 0.014
$C_{\bar{D}}$	8.044 ± 1.592	13.181 ± 2.076	17.528 ± 3.165

Table C.4: Fitted parameters of the NNLO fits in Section 5.2. Only independent fitted parameters are shown. The uncertainties are the experimental ones, determined by diagonalizing the hesse matrix of the χ^2 function in the minimum (see Section 4.5)

Par.	HERA I+II Dis only	with CMS dijets, $\mu = p_{\mathrm{T},1} e^{0.3y^*}$	with CMS dijets, $\mu = m_{12}$
B_g	0.094 ± 0.041	0.096 ± 0.024	0.067 ± 0.025
C_g	14.972 ± 1.407	10.149 ± 0.610	9.893 ± 0.558
E_g	154.716 ± 34.852	33.275 ± 8.023	41.327 ± 9.072
B_{u_v}	0.393 ± 0.027	0.271 ± 0.020	0.305 ± 0.021
C_{u_v}	3.969 ± 0.054	3.822 ± 0.046	3.877 ± 0.045
D_{u_v}	25.124 ± 3.638	51.540 ± 7.318	40.791 ± 5.710
B_{d_v}	1.088 ± 0.082	1.032 ± 0.072	1.049 ± 0.066
C_{d_v}	4.795 ± 0.369	4.586 ± 0.312	4.588 ± 0.276
$C_{\bar{U}}$	12.680 ± 1.172	14.790 ± 0.740	15.100 ± 0.731
$D_{\bar{U}}$	10.984 ± 2.726	17.302 ± 2.194	16.277 ± 2.083
$A_{\bar{D}}$	0.317 ± 0.027	0.309 ± 0.020	0.324 ± 0.021
$B_{\bar{D}}$	-0.089 ± 0.013	-0.092 ± 0.010	-0.087 ± 0.010
$C_{\bar{D}}$	11.069 ± 1.771	10.580 ± 1.513	12.032 ± 1.658

C.3 Fits including α_s

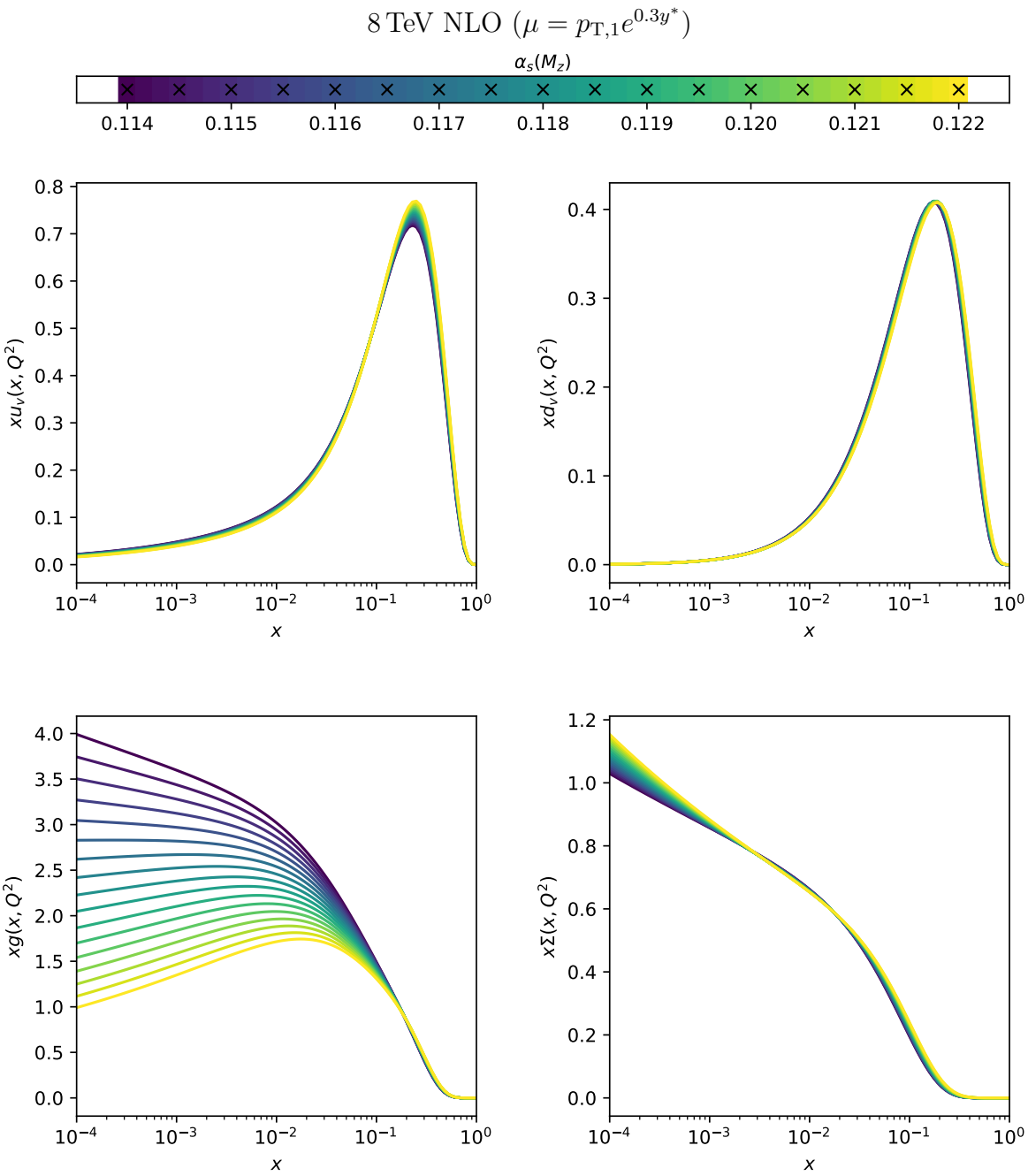


Figure C.7: Overview on the $\alpha_s(M_z)$ profiling with $\mu = p_{T,1}e^{0.3y^*}$ at NLO from Section 5.3. The extent to which a fitted PDF varies upon changing $\alpha_s(M_z)$ is a measure for its correlation with $\alpha_s(M_z)$. All PDFs are shown at $Q^2 = 1.9\,\mathrm{GeV}^2$.

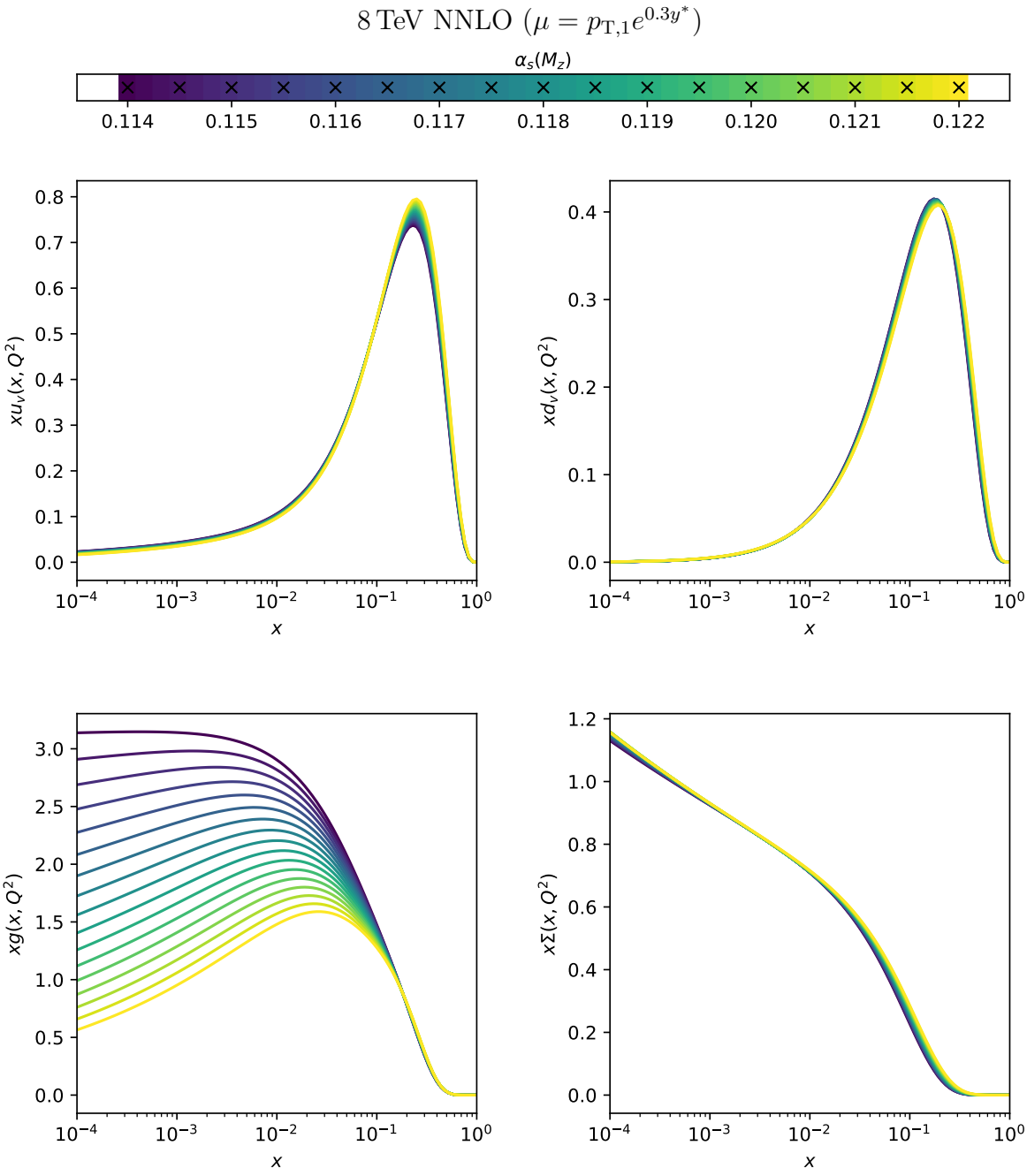


Figure C.8: Overview on the $\alpha_s(M_z)$ profiling with $\mu = p_{T,1} e^{0.3y^*}$ at NNLO from Section 5.3. The extent to which a fitted PDF varies upon changing $\alpha_s(M_z)$ is a measure for its correlation with $\alpha_s(M_z)$. All PDFs are shown at $Q^2 = 1.9\,\text{GeV}^2$.

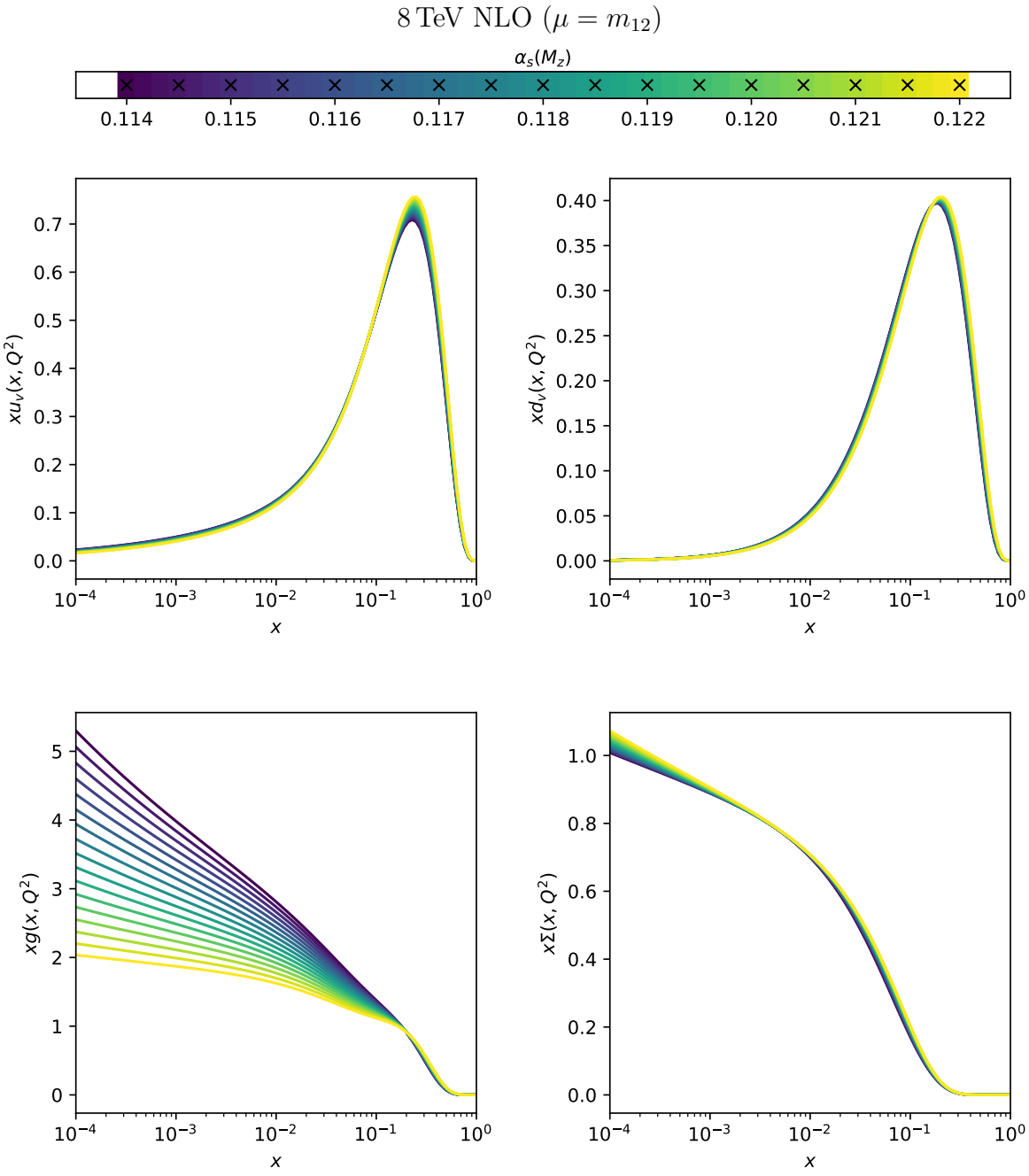


Figure C.9: Overview on the $\alpha_s(M_z)$ profiling with $\mu = m_{12}$ at NLO from Section 5.3. The extent to which a fitted PDF varies upon changing $\alpha_s(M_z)$ is a measure for its correlation with $\alpha_s(M_z)$. All PDFs are shown at $Q^2 = 1.9\,\mathrm{GeV}^2$.

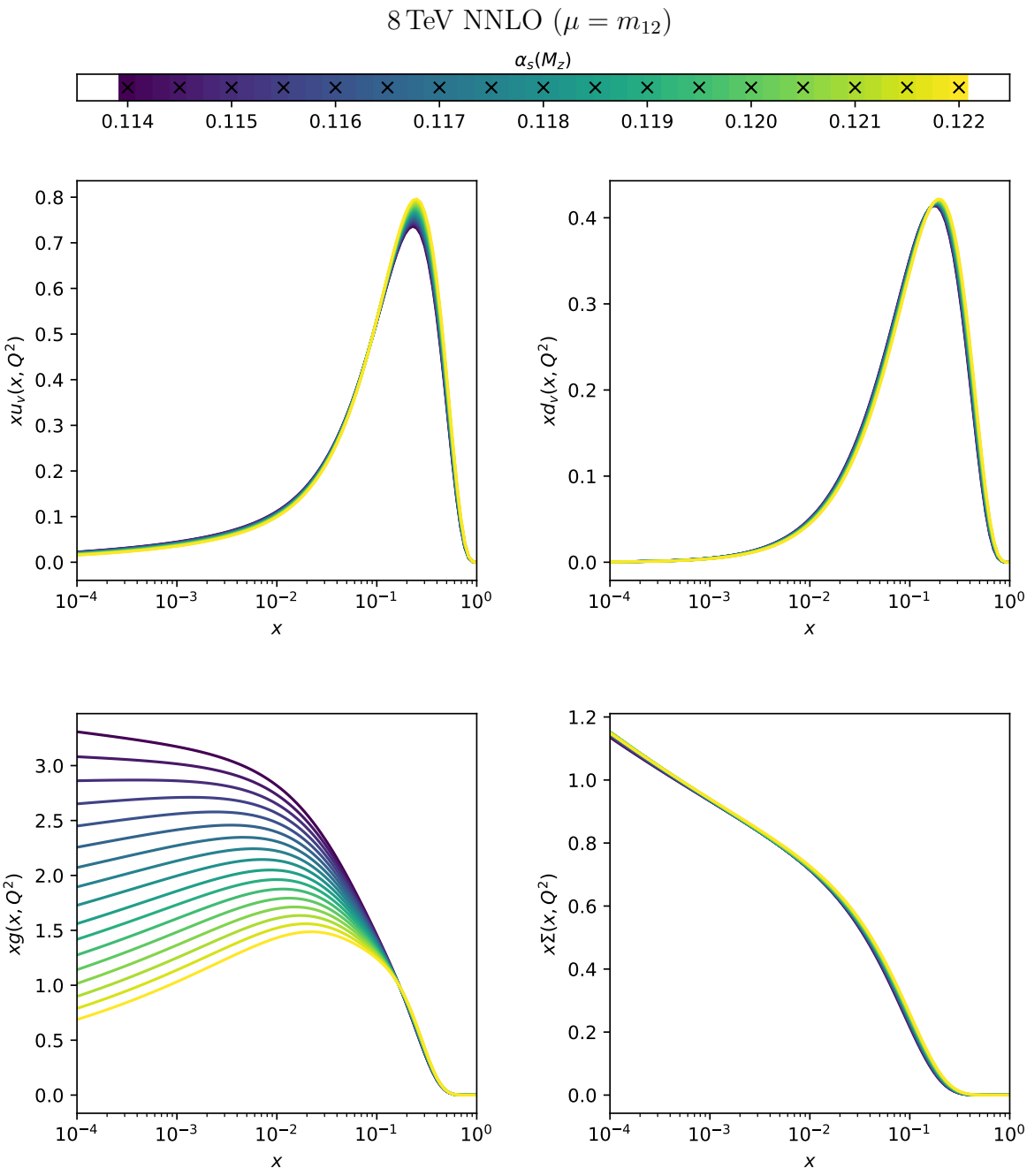
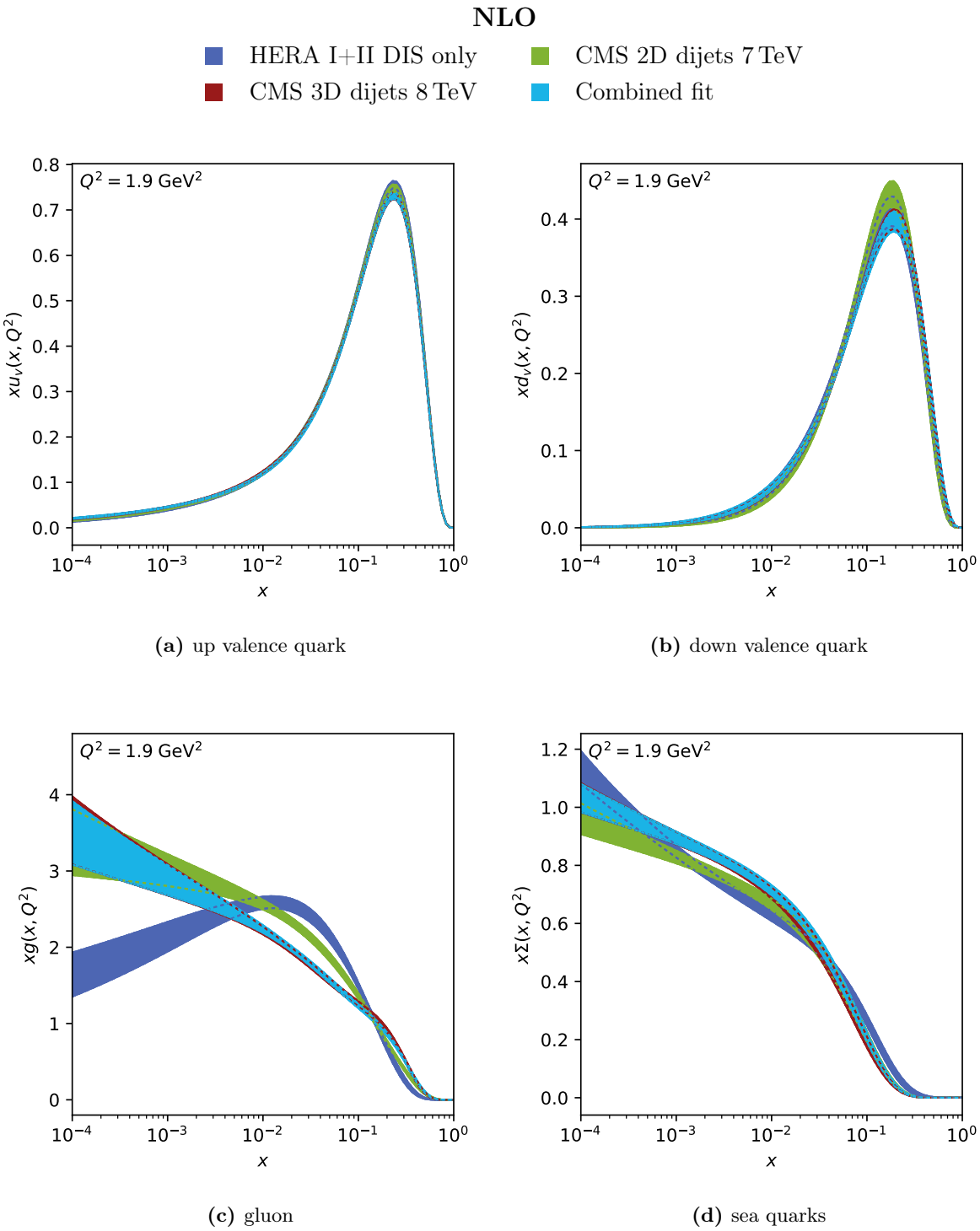


Figure C.10: Overview on the $\alpha_s(M_z)$ profiling with $\mu = m_{12}$ at NNLO from Section 5.3. The extent to which a fitted PDF varies upon changing $\alpha_s(M_z)$ is a measure for its correlation with $\alpha_s(M_z)$. All PDFs are shown at $Q^2 = 1.9\,\text{GeV}^2$.

C.4 Fits with 2D dijet data at 7 TeV included



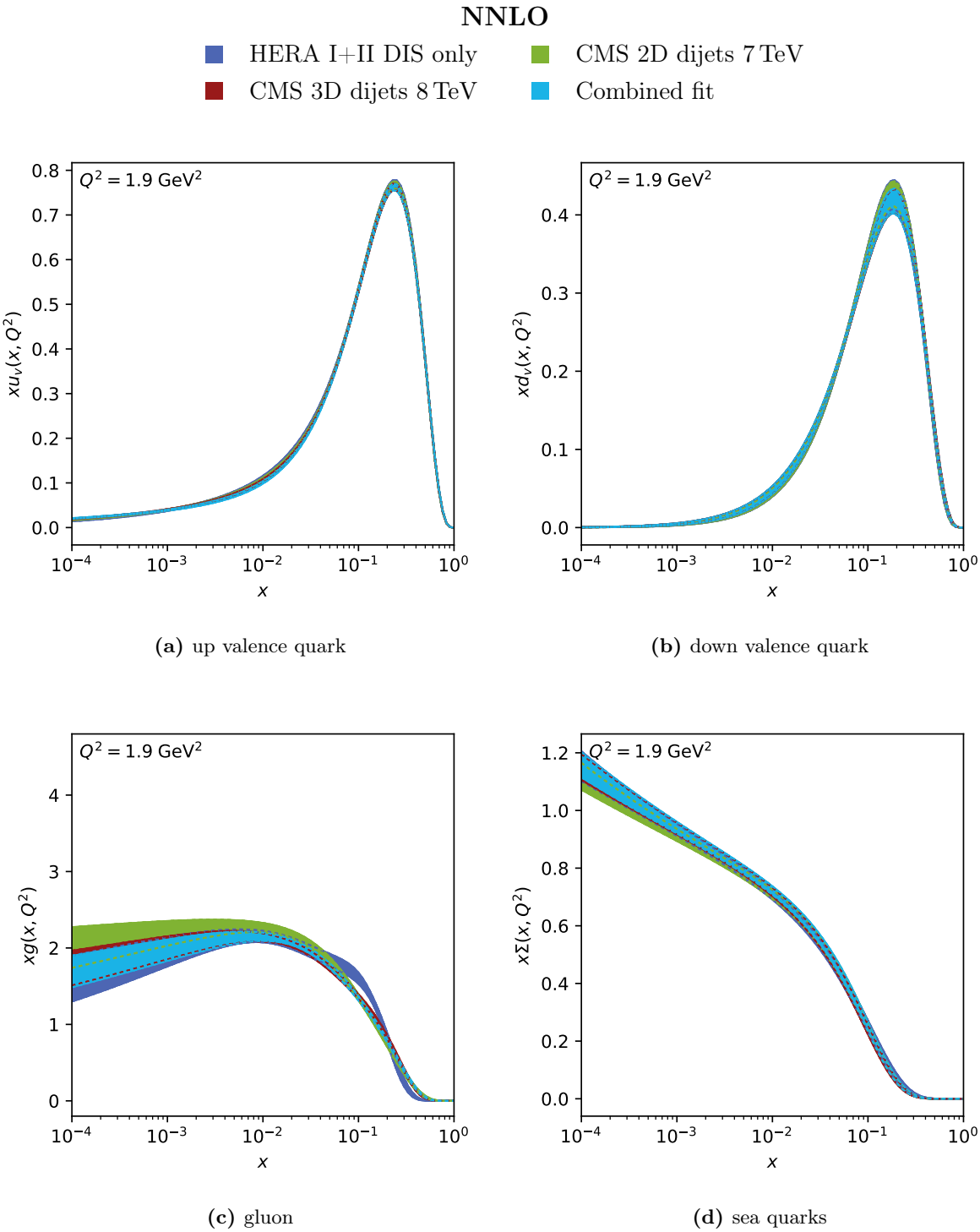
(a) up valence quark

(b) down valence quark

(c) gluon

(d) sea quarks

Figure C.11: PDF overview for the fits including 7 TeV double-differential dijet data in Section 5.4 at NLO



(a) up valence quark

(b) down valence quark

(c) gluon

(d) sea quarks

Figure C.12: PDF overview for the fits including 7 TeV double-differential dijet data in Section 5.4 at NNLO

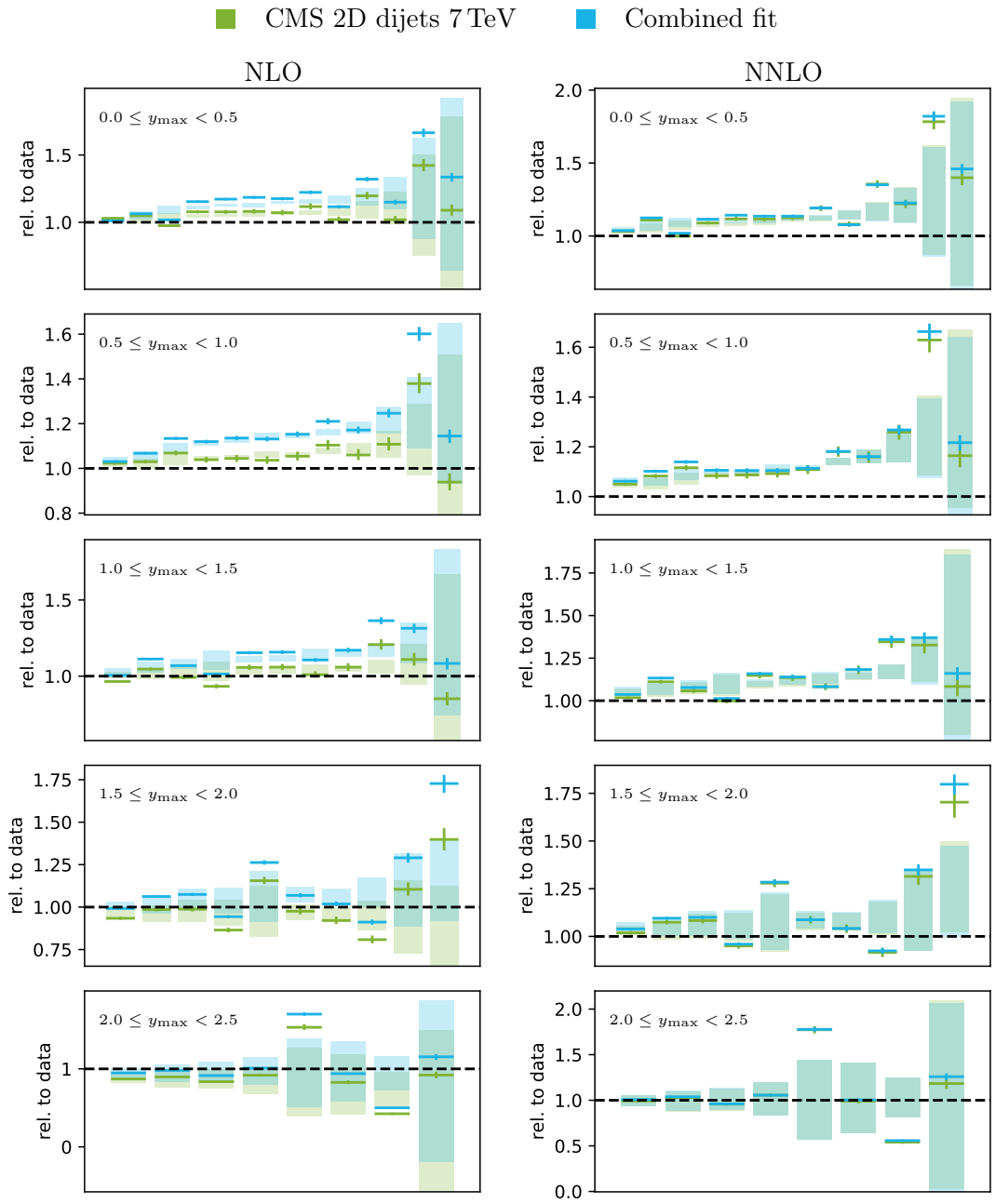


Figure C.13: Theory (crosses) and shifted data (bands) relative to the central data for the inclusive DIS plus 7 TeV fit (green ■) and the combined inclusive DIS 8 TeV and 7 TeV fit (cyan ■) for NLO (left) and NNLO (right). The theory predictions are calculated by evaluating the `FastNLO` table with the fitted PDFs. The shifted data bands are constructed by shifting the central data point with the fitted correlated uncertainty sources (shifts) from the corresponding fit.

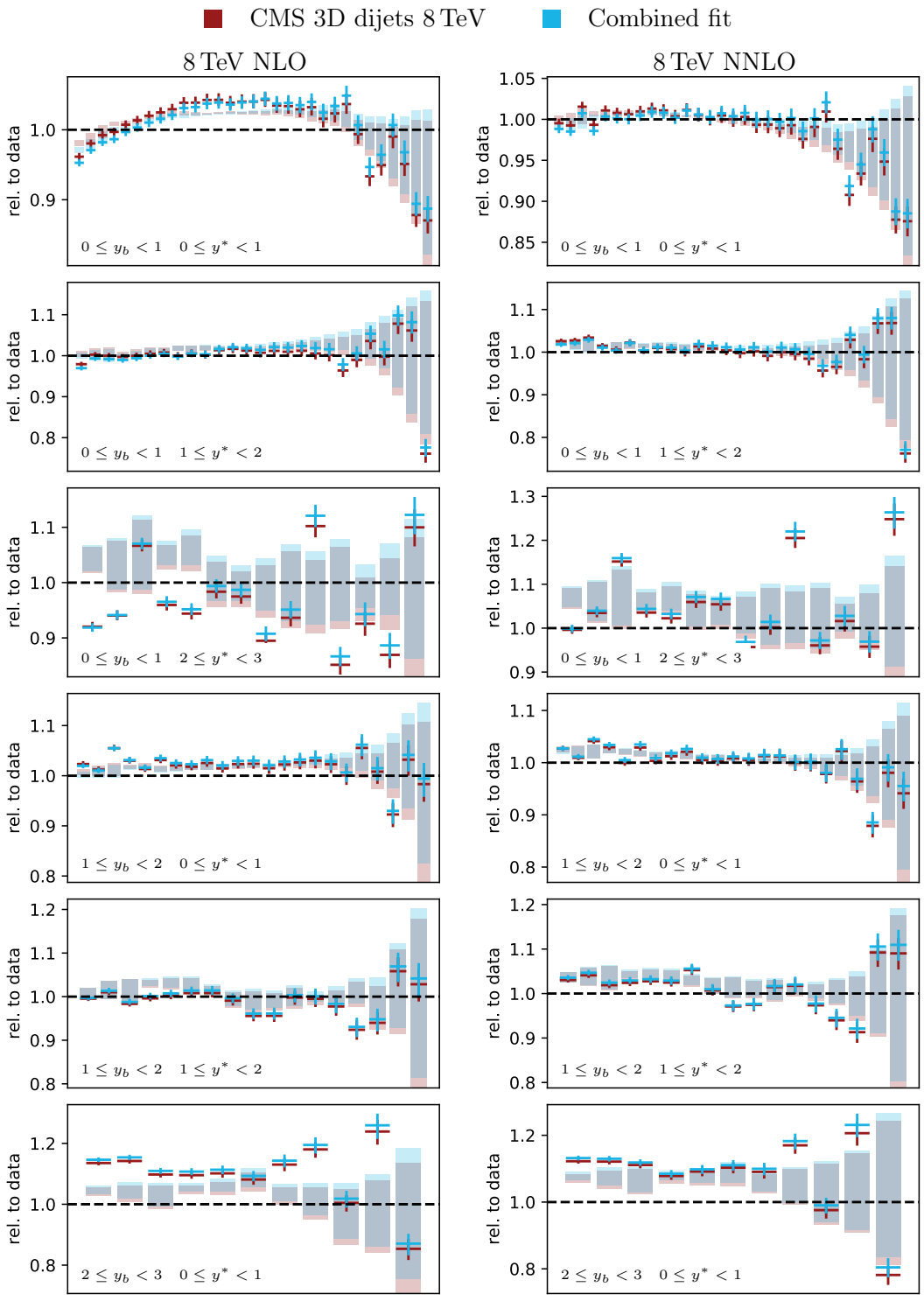


Figure C.14: Theory (crosses) and shifted data (bands) relative to the central data for the inclusive DIS plus 8 TeV ($\mu = m_{12}$) fit (dark red ■) and the combined inclusive DIS 8 TeV and 7 TeV fit (cyan ■) for NLO (left) and NNLO (right). The theory predictions are calculated by evaluating the `FastNLO` table with the fitted PDFs. The shifted data bands are constructed by shifting the central data point with the fitted correlated uncertainty sources (shifts) from the corresponding fit.

Table C.5: Fitted parameters of the NLO fits in Section 5.4. Only independent fitted parameters are shown. The uncertainties are the experimental ones, determined by diagonalizing the hesse matrix of the χ^2 function in the minimum (see Section 4.5)

Par.	HERA I+II Dis only	CMS 3D 8 TeV	CMS 2D 7 TeV	Combined
B_g	0.121 ± 0.039	-0.084 ± 0.024	-0.046 ± 0.026	-0.081 ± 0.023
C_g	9.947 ± 1.122	7.999 ± 0.415	6.219 ± 0.636	7.466 ± 0.398
E_g	13.289 ± 5.884	48.224 ± 8.733	6.996 ± 2.987	35.610 ± 6.617
B_{u_v}	0.427 ± 0.028	0.361 ± 0.026	0.402 ± 0.028	0.342 ± 0.023
C_{u_v}	3.995 ± 0.050	3.922 ± 0.048	3.979 ± 0.049	3.911 ± 0.050
D_{u_v}	19.081 ± 2.831	23.607 ± 3.518	20.983 ± 3.292	26.520 ± 3.607
B_{d_v}	1.024 ± 0.082	0.967 ± 0.051	1.121 ± 0.074	0.955 ± 0.053
C_{d_v}	4.645 ± 0.367	4.046 ± 0.199	5.033 ± 0.315	3.987 ± 0.207
$C_{\bar{U}}$	15.098 ± 0.734	16.133 ± 0.682	15.910 ± 0.651	16.215 ± 0.720
$D_{\bar{U}}$	28.658 ± 3.686	15.717 ± 2.585	21.211 ± 3.246	15.931 ± 2.347
$A_{\bar{D}}$	0.220 ± 0.023	0.377 ± 0.031	0.343 ± 0.032	0.377 ± 0.030
$B_{\bar{D}}$	-0.128 ± 0.016	-0.059 ± 0.014	-0.061 ± 0.015	-0.058 ± 0.013
$C_{\bar{D}}$	8.044 ± 1.592	17.528 ± 3.165	16.270 ± 2.624	15.218 ± 2.272

Table C.6: Partial χ^2 values for the NLO fits in Section 5.4

		n_{data}	HERA I+II Dis only	CMS dijets 3D 8 TeV	CMS dijets 2D 7 TeV	Combined
HERA I+II	combined	1016	1106.14	1181.30	1124.63	1178.64
CMS dijets 3D 8 TeV	yb0 ys0	31	–	35.82	–	30.92
	yb0 ys1	26	–	28.03	–	29.91
	yb0 ys2	14	–	76.82	–	81.68
	yb1 ys0	23	–	19.21	–	27.05
	yb1 ys1	17	–	27.45	–	26.34
	yb2 ys0	11	–	88.36	–	87.84
	combined	122	–	275.69	–	283.74
CMS dijets 2D 7 TeV	ym0	13	–	–	28.27	34.31
	ym1	12	–	–	18.71	19.29
	ym2	11	–	–	19.55	17.13
	ym3	10	–	–	16.61	15.20
	ym4	8	–	–	11.10	11.43
	combined	54	–	–	94.24	97.36
	correlated χ^2		50.96	115.15	60.71	132.39
	log penalty χ^2		-2.98	-13.24	-6.07	-11.22
	combined		1154.12	1558.90	1273.52	1680.89
	ndf		1003	1125	1057	1179
	p-value		6.13×10^{-4}	1.26×10^{-16}	1.38×10^{-6}	2.48×10^{-20}
	combined χ^2/ndf		**1.151**	**1.386**	**1.205**	**1.426**

Table C.7: Fitted parameters of the NNLO fits in Section 5.4. Only independent fitted parameters are shown. The uncertainties are the experimental ones, determined by diagonalizing the hesse matrix of the χ^2 function in the minimum (see Section 4.5)

Par.	HERA I+II Dis only	CMS 3D 8 TeV	CMS 2D 7 TeV	Combined
B_g	0.094 ± 0.041	0.067 ± 0.025	0.043 ± 0.027	0.071 ± 0.023
C_g	14.972 ± 1.407	9.893 ± 0.558	7.236 ± 0.737	8.797 ± 0.577
E_g	154.716 ± 34.852	41.327 ± 9.072	8.100 ± 3.710	23.508 ± 6.130
B_{u_v}	0.393 ± 0.027	0.305 ± 0.021	0.351 ± 0.025	0.266 ± 0.020
C_{u_v}	3.969 ± 0.054	3.877 ± 0.045	3.939 ± 0.050	3.830 ± 0.044
D_{u_v}	25.124 ± 3.638	40.791 ± 5.710	31.727 ± 4.541	52.942 ± 7.464
B_{d_v}	1.088 ± 0.082	1.049 ± 0.066	1.093 ± 0.071	1.051 ± 0.068
C_{d_v}	4.795 ± 0.369	4.588 ± 0.276	4.810 ± 0.310	4.624 ± 0.286
$C_{\bar{U}}$	12.680 ± 1.172	15.100 ± 0.731	15.032 ± 0.716	15.400 ± 0.735
$D_{\bar{U}}$	10.984 ± 2.726	16.277 ± 2.083	18.377 ± 2.484	17.335 ± 2.143
$A_{\bar{D}}$	0.317 ± 0.027	0.324 ± 0.021	0.320 ± 0.022	0.321 ± 0.019
$B_{\bar{D}}$	-0.089 ± 0.013	-0.087 ± 0.010	-0.085 ± 0.011	-0.088 ± 0.010
$C_{\bar{D}}$	11.069 ± 1.771	12.032 ± 1.658	11.810 ± 1.558	11.169 ± 1.493

Table C.8: Partial χ^2 values for the NNLO fits in Section 5.4

		n_{data}	HERA I+II Dis only	CMS dijets 3D 8 TeV	CMS dijets 2D 7 TeV	Combined
HERA I+II	combined	1016	1109.15	1119.54	1112.62	1120.46
CMS dijets 3D 8 TeV	yb0 ys0	31	–	16.70	–	17.25
	yb0 ys1	26	–	12.78	–	12.05
	yb0 ys2	14	–	23.38	–	23.53
	yb1 ys0	23	–	16.22	–	16.79
	yb1 ys1	17	–	20.40	–	20.17
	yb2 ys0	11	–	22.10	–	22.14
	combined	122	–	111.58	–	111.93
CMS dijets 2D 7 TeV	ym0	13	–	–	33.46	37.65
	ym1	12	–	–	18.72	22.45
	ym2	11	–	–	16.79	18.04
	ym3	10	–	–	12.45	14.17
	ym4	8	–	–	11.89	10.73
	combined	54	–	–	93.31	103.04
	correlated χ^2		55.48	63.17	63.93	73.81
	log penalty χ^2		-1.74	-3.95	-0.49	-1.37
	combined		1162.89	1290.33	1269.37	1407.88
	ndf		1003	1125	1057	1179
	p-value		3.24×10^{-4}	4.16×10^{-4}	6.63×10^{-6}	4.25×10^{-6}
	combined χ^2/ndf		**1.159**	**1.147**	**1.201**	**1.194**

Bibliography

[1] F. D. Aaron et al. Combined measurement and QCD analysis of the inclusive $e^{\pm}p$ scattering cross sections at HERA. *Journal of High Energy Physics*, 2010(1), January 2010.

[2] V. M. Abazov et al. Determination of the strong coupling constant from the inclusive jet cross section in $p\bar{p}$ collisions at sqrt(s)=1.96 TeV. *Phys. Rev. D*, 80:111107, 2009.

[3] Rabah Abdul Khalek et al. Phenomenology of NNLO jet production at the LHC and its impact on parton distributions. *Eur. Phys. J. C*, 80(8):797, 2020.

[4] H. Abramowicz et al. Combination of measurements of inclusive deep inelastic $e^{\pm}p$ scattering cross sections and QCD analysis of HERA data. *The European Physical Journal C*, 75(12), December 2015.

[5] T. Affolder et al. Measurement of the Strong Coupling Constant from Inclusive Jet Production at the Tevatron $\bar{p}p$ Collider. *Phys. Rev. Lett.*, 88:042001, 2002.

[6] S. Alekhin et al. HERAFitter, Open Source QCD Fit Project, 2015.

[7] Constantia Alexandrou et al. Lattice calculation of parton distributions. *Phys. Rev. D*, 92:014502, Jul 2015.

[8] V. Andreev et al. Measurement of multijet production in ep collisions at high Q^2 and determination of the strong coupling α_s. *Eur. Phys. J. C*, 75(2):65, 2015.

[9] Johannes Bellm et al. Herwig 7.0/Herwig++ 3.0 release note. *Eur. Phys. J. C*, 76(4):196, 2016.

[10] J.D. Bjorken and Emmanuel A. Paschos. Inelastic Electron Proton and gamma Proton Scattering, and the Structure of the Nucleon. *Phys. Rev.*, 185:1975–1982, 1969.

[11] Enrico Bothmann et al. Event Generation with Sherpa 2.2. *SciPost Phys.*, 7(3):034, 2019.

[12] M. Botje. Qcdnum: Fast qcd evolution and convolution. *Computer Physics Communications*, 182(2):490–532, Feb 2011.

[13] Daniel Britzger, Klaus Rabbertz, Daniel Savoiu, Georg Sieber, and Markus Wobisch. Determination of the strong coupling constant using inclusive jet cross section data from multiple experiments. *Eur. Phys. J. C*, 79(1):68, 2019.

[14] Daniel Britzger, Klaus Rabbertz, Fred Stober, and Markus Wobisch. New features in version 2 of the fastNLO project. In *20th International Workshop on Deep-Inelastic Scattering and Related Subjects*, pages 217–221, 2012.

[15] Andy Buckley et al. LHAPDF6: parton density access in the LHC precision era. *Eur. Phys. J. C*, 75:132, 2015.

[16] Matteo Cacciari, Gavin P Salam, and Gregory Soyez. The anti-k_t jet clustering algorithm. *Journal of High Energy Physics*, 2008(04):063–063, apr 2008.

[17] Tancredi Carli et al. A posteriori inclusion of parton density functions in NLO QCD final-state calculations at hadron colliders: the APPLGRID project. *The European Physical Journal C*, 66(3-4):503–524, Feb 2010.

[18] S. Chatrchyan et al. Measurements of differential jet cross sections in proton-proton collisions at $\sqrt{s} = 7\,\text{TeV}$ with the CMS detector. *Physical Review D*, 87(11), Jun 2013.

[19] S. Chekanov et al. Jet-radius dependence of inclusive-jet cross-sections in deep inelastic scattering at HERA. *Phys. Lett. B*, 649:12–24, 2007.

[20] John C. Collins, Davison E. Soper, and George F. Sterman. Factorization of Hard Processes in QCD. *Adv. Ser. Direct. High Energy Phys.*, 5:1–91, 1989.

[21] J. Davies, A. Vogt, S. Moch, and J.A.M. Vermaseren. Non-singlet coefficient functions for charged-current deep-inelastic scattering to the third order in QCD. *PoS*, DIS2016:059, 2016.

[22] Stefano Forte and Stefano Carrazza. Parton distribution functions, 2020.

[23] Thomas Gehrmann et al. Jet cross sections and transverse momentum distributions with NNLOJET. *PoS*, RADCOR2017:074, 2018.

[24] Tie-Jiun Hou et al. New CTEQ global analysis of quantum chromodynamics with high-precision data from the LHC, 12 2019.

[25] F. James. *MINUIT - Function Minimization and Error Analysis*. CERN Program Library, 94.1 edition, March 1994.

[26] F. James and M. Roos. Minuit - a system for function minimization and analysis of the parameter errors and correlations. *Computer Physics Communications*, 10(6):343–367, December 1975.

[27] V. Khachatryan et al. Constraints on parton distribution functions and extraction of the strong coupling constant from the inclusive jet cross section in pp collisions at $\sqrt{s} = 7\,\text{TeV}$. *The European Physical Journal C*, 75(6), Jun 2015.

[28] V. Khachatryan et al. Measurement of the inclusive 3-jet production differential cross section in proton–proton collisions at 7 TeV and determination of the strong

coupling constant in the TeV range. *The European Physical Journal C*, 75(5), May 2015.

[29] Vardan Khachatryan et al. Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV. *JINST*, 12(02):P02014, 2017.

[30] Vardan Khachatryan et al. Measurement and QCD analysis of double-differential inclusive jet cross sections in pp collisions at $\sqrt{s} = 8$ TeV and cross section ratios to 2.76 and 7 TeV. *JHEP*, 03:156, 2017.

[31] T. Kluge, K. Rabbertz, and M. Wobisch. Fast pQCD Calculations for PDF Fits. *Deep Inelastic Scattering DIS 2006*, Jan 2007.

[32] Bogdan Malaescu and Pavel Starovoitov. Evaluation of the Strong Coupling Constant α_s Using the ATLAS Inclusive Jet Cross-Section Data. *Eur. Phys. J. C*, 72:2041, 2012.

[33] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt. Parton distributions for the lhc. *The European Physical Journal C*, 63(2):189–285, Jul 2009.

[34] S. Moch, J.A.M. Vermaseren, and A. Vogt. Third-order qcd corrections to the charged-current structure function. *Nuclear Physics B*, 813(1-2):220–258, May 2009.

[35] J. A. Nelder and R. Mead. A simplex method for function minimization. *The Computer Journal*, 7(4):308–313, January 1965.

[36] Particle Data Group. Review of Particle Physics. *Progress of Theoretical and Experimental Physics*, 2020(8), 08 2020. 083C01.

[37] J. Pumplin, D. R. Stump, and W. K. Tung. Multivariate fitting and the error matrix in global analysis of data. *Physical Review D*, 65(1), Dec 2001.

[38] Klaus Rabbertz. *Jet Physics at the LHC*. Springer International Publishing, 2017.

[39] Georg Sieber. *Measurement of Triple-Differential Dijet Cross Sections with the CMS Detector at 8 TeV and PDF Constraints*. PhD thesis, Karlsruher Institut für Technologie (KIT), 2016.

[40] A. M. Sirunyan et al. Measurement of the triple-differential dijet cross section in proton-proton collisions at $\sqrt{s} = 8\,\text{TeV}$ and constraints on parton distribution functions. *The European Physical Journal C*, 77(11), Nov 2017.

[41] Torbjörn Sjöstrand et al. An introduction to PYTHIA 8.2. *Computer Physics Communications*, 191:159–177, Jun 2015.

[42] R. S. Thorne. Effect of changes of variable flavor number scheme on parton distribution functions and predicted cross sections. *Physical Review D*, 86(7), October 2012.

[43] J.A.M. Vermaseren, A. Vogt, and S. Moch. The third-order qcd corrections to deep-inelastic scattering by photon exchange. *Nuclear Physics B*, 724(1-2):3–182, Sep 2005.

[44] S. S. Wilks. The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics*, 9(1):60–62, March 1938.

[45] xFitter developers. *xFitter PDF Fitting package*, March 2017.

[46] Yndurain, Francisco J. *The Theory of Quark and Gluon Interactions*. Springer Berlin Heidelberg, 2006.