Report on Summer Student Programme at DESY

Martin Bieker, TU Dortmund University

September 7, 2016

Abstract

This report is a summary of my summer student project on the fitting of parton density functions with top data. First a brief introduction to the theory of PDFs and the fitting process is given. In a theory study the impact of different parameters on the theory predictions are reviewed. The second part of the project focusses on the implementation of a chi^2 definition for use in fits with low precision data. This χ^2 function is compared to commonly used definitions. Using the newly implemented χ^2 function the effects of the inclusion of LHC top data taken at energies of 7 TeV is considered.

Contents

1 Introduction

1.1 Parton density functions

At the Large Hadron Collider (LHC) protons are collided in order to study the Standard Model of particle physics. Protons, or more general, hadrons are not elementary particles. They are compromised of quarks and gluons. So in order to make theory predictions for experiments at the LHC the inner structure of the proton must be known. The structure of hadrons is described by a set of Parton Density Functions $f_i(x, Q)$. These are probability densities that determine what fraction of the total momentum of the hadron each parton i carries.
Within the framework of perturbative QCD[1] it is not possible to directly determine the structure of hadrons. Therefore it is necessary to determine to the PDFs by some other means. This is done by fitting theory predictions based on said PDFs to experimental data. Once the PDF for a energy scale is known it can be evolved to higher scales by solving the so called DGLAP equations.
The goal of this project is to determine PDFs using these fits. The focus lies on the inclusion of $t\bar{t}$-data. The production of top quarks at the LHC is discussed in the next section.

1.2 $t\bar{t}$ production at the LHC

One objective of the project is to incorporate $t\bar{t}$-data of the Large Hadron Collider at a center of mass energy of 7 TeV into a fit to determine the parton density functions of the proton. At the LHC top quark pairs are produced in various reactions. The feynman graphs for the most import processes are shown in figure 1.

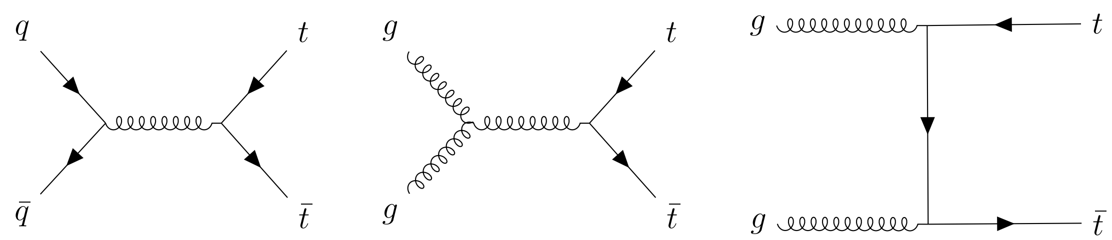


Figure 1: Feynman graphs of top pair production processes at the Large Hadron Collider

In addition to the hard scattering cross section, which can be calculated from the Feynman diagrams above, the $t\bar{t}$ production rate of proton proton collisions depends on the parton distribution functions. For this reason $t\bar{t}$ data can be used in PDF fits.

[1]Quantum Chromodynamics is the theory that describes the strong interactions between quarks and gluons.

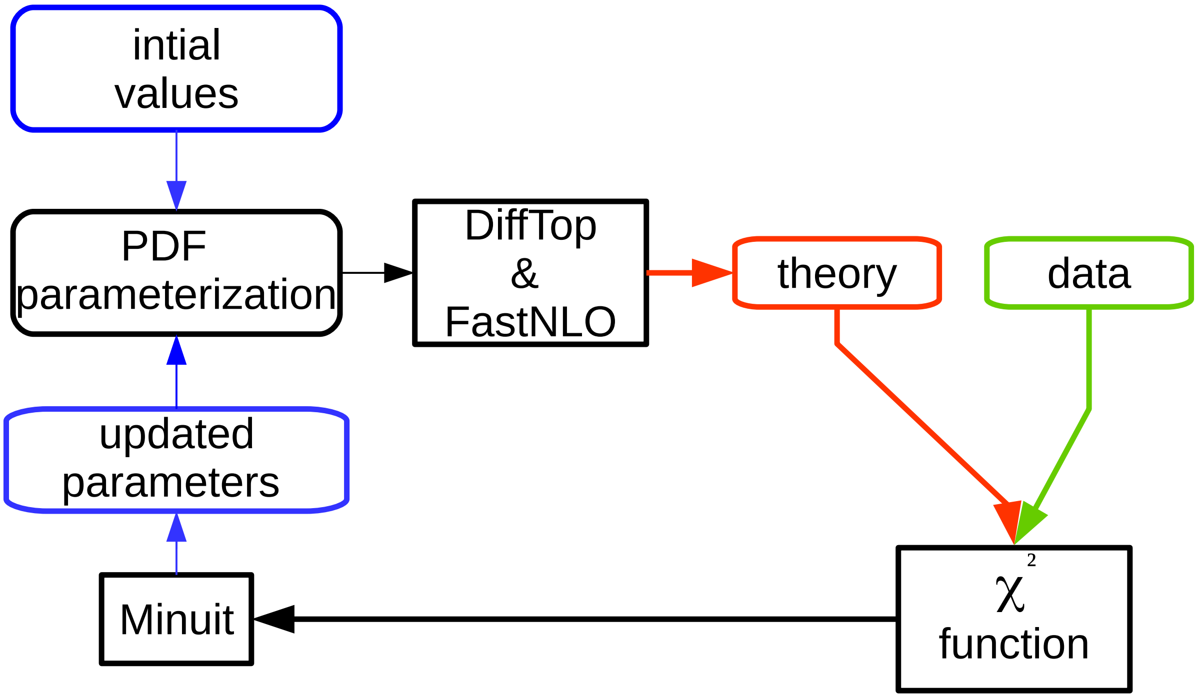


Figure 2: Overview of the process by which PDFs are extracted from data. The names of the software packets used within the Xfitter framework are also named.

1.3 PDF fitting process

The parton density functions are determined by fitting data to corresponding theory predictions. This is necessary, because they cannot be determined from (perturbative) theory calculations alone. The purpose of this section of the report is to describe this fitting process and introduce the software tools used.

Figure 2 provides a visual overview of the PDF fitting process. At first a parameterization for the parton density functions is chosen. Possible choices for this are discussed later. With these parameterizations and a set of initial values for the PDF parameters a theory predictions can be computed. These predictions are then compared to the measured data using a χ^2-function. χ^2 is a function of the data, it's uncertainties and the theory and provides a measure of agreement between data and theory predictions. The statistical reason for choosing a certain χ^2 definition are discussed in the next section of this report. In order to find set of PDFs that best fits the data, the χ^2-function is iteratively minimized.

There are different software packets which implement the process in order to extract the parton density functions from measured data. The theory predictions for the reactions are calculated with DiffTop [1] and fastNLO [2]. The former uses perturbative QCD in order to calculate the production for heavy quarks and the latter uses interpolation tables to allow fast evaluations of the cross sections with changing PDFs. This is done in order to prevent the time consuming invocation of DiffTop at each iteration of the fit. The χ^2-function then obtained with these theory predictions is then minimized using the

MINUIT package[3]. This is a program which finds the minimum of a specified function numerically by employing various minimization strategies.
These tools are all included in the XFITTER framework [4] which allows doing PDF fits for a wide range of different reactions.

1.4 The χ^2 -function

The χ^2 -function is used to quantify the agreement between the theory predictions and the measured data. It is a function of the parameters to be fitted to the data. In a data set which includes correlated uncertainties the following definition is usually used:

$$\chi^2(\vec{a}) = \sum_{ij} (d_i - t_i(\vec{a})) \left(V^{-1}\right)_{ij} (d_j - t_j(\vec{a})) \tag{1}$$

In this equation the d_i are the data points and the t_i are the corresponding theory predictions which depend on the fit parameters $\vec{a}$. The correlated uncertainties in the dataset are described by the inverse covariance matrix V^{-1}. The parameters $\vec{a}$ can be obtained by the minimization of the χ^2 -function.
This definition of the χ^2-function can be derived from the principle of maximum likelihood. For the derivation it is assumed that the data points follow a normal distribution around the theory value.

$$\vec{d} \sim \mathcal{N}(\vec{t}, V)$$

The likelihood function is then given by the probability density function of the multivariate normal distribution.

$$\mathcal{L} = \frac{1}{\sqrt{2\pi \det (V)}} \exp\left(-\frac{1}{2}\sum_{ij} (d_i - t_i(\vec{a})) \left(V^{-1}\right)_{ij} (d_j - t_j(\vec{a}))\right) \tag{2}$$

In order to estimate the parameters $\vec{a}$ equation 2 is maximized. Algebraically equivalent to this is to minimize to negative logarithm of the function. From this the χ^2-function can be obtained.

$$\chi^2(\vec{a}) = -2\ln\mathcal{L} = \sum_{ij} (d_i - t_i(\vec{a})) \left(V^{-1}\right)_{ij} (d_j - t_j(\vec{a})) + \text{const} \tag{3}$$

Other χ^2 definitions can be derived by assuming a different distribution of the data. It is important to note that the χ^2-function presented here is a function of the absolute uncertainties of the measurement. If the quoted uncertainty is relative this value has to be rescaled before entering the χ^2 calculation. This can be achieved by multiplying the relative error either by the data or the theory value. This choice of rescaling provides the possibility to introduce a statistical bias into the fit result.

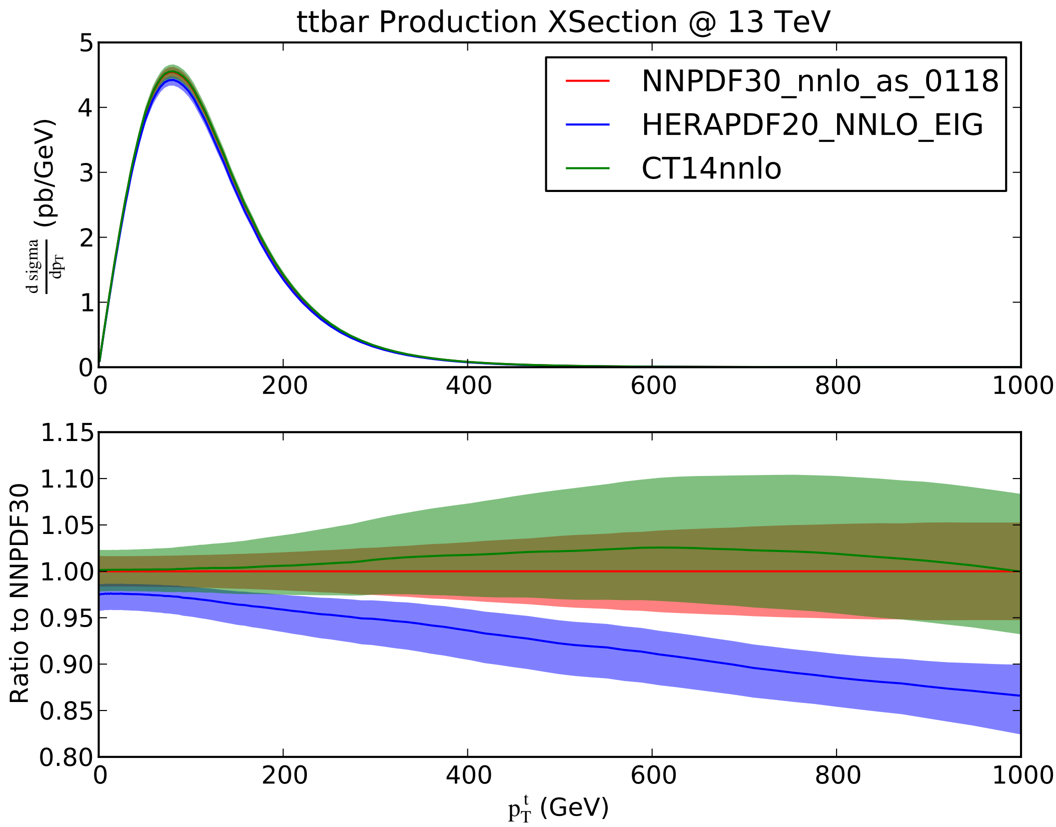
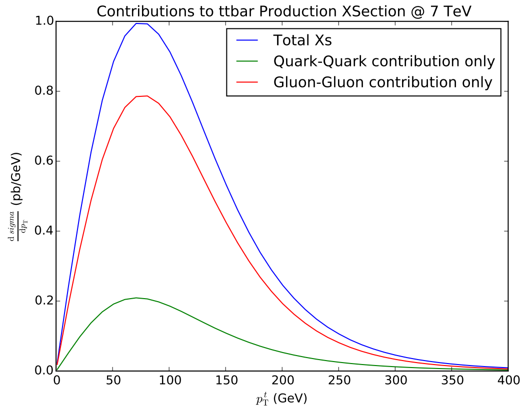


Figure 3: Differential $t\bar{t}$ production cross section as a function of transverse momentum p_{T}. The influence of three different PDF parameterizations is shown. Absolute cross sections are displayed in the upper plot. In the lower plot shows a ratio of the cross sections to the HERAPDF parameterization.

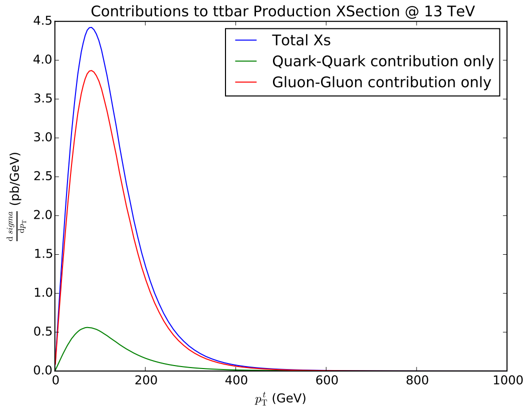
2 Project

2.1 Theory Studies

In this section the theory studies of my project will be discussed. This was done in order to see how different settings in FastNLO influences the predicted cross sections. Parameters which can be changed include the PDF parametrisation, center of mass energy, the factorization scale and the renormalization scale.
The influence of PDF parameterization is studied first. In order to do this the $t\bar{t}$ production cross section is calculated as a function of the transverse momentum p_T in fastNLO. This is done for a selection of different PDF parameterizations. The resulting distributions are then compared to each other to observe the impact of the different PDFs. The calculation was carried out at a center of mass energy of 13 TeV. Figure 3 shows this comparison for three different PDFs.
The plot shows that for low energies the differences between the PDFs is covered by



(a) $\sqrt{S} = 7\,\mathrm{TeV}$



(b) $\sqrt{S} = 13\,\mathrm{TeV}$

Figure 4: Different contributions to the $t\bar{t}$ production cross section as a function of p_T for center of mass energies of 7 TeV and 13 TeV.

the uncertainty bands. Only for higher energies a significant deviation is visible. For the relative errors this is caused by the low central values of the functions at high p_T. For this reason the choice of parameterization does not significantly influence the theory predictions. This means that one is relatively free in this choice. In the rest of this report the HERAPDF parameterization is used for all PDF fits. The reason for this is that the data for the comparison fits is done with data from the HERA experiments.
Another aspect of interest is the contributions of individual partons to the total production cross section. With this information it is possible to conclude which PDFs will be influenced mostly by the inclusion of the $t\bar{t}$ data. In the 4a and 4b the contribution to the total from quarks and gluons are displayed separately for 7 TeV and 13 TeV center of mass energy. An additional plot for 8 TeV center of mass energy (figure **??**) can be found in the appendix.
For higher energies the contribution of gluon introduced processes becomes greater. This suggests that $t\bar{t}$ data may be able to provide direct constraints on the gluon PDFs.
Lastly the influence of the factorization and rescaling scale will be considered. Reasonable choices for these scales μ_r and μ_f lie in the same order of magnitude as the scales of the observed processes. For the $t\bar{t}$ production sensible scales would be either the mass of the top quark m_t, the transverse momentum p_T or a combination of the former. The FASTNLO toolkit allows to study the influence of these scale choices.

- mass of the top quark $\mu^2 = m_\mathrm{t}^2$
- transverse momentum $\mu^2 = p_\mathrm{T}^2$
- quadratic mean $\mu^2 = \frac{m_\mathrm{t}^2 + p_\mathrm{T}^2}{2}$

The following plot shows the resulting cross section for three different scales. In this case both the factorization and the renormaliziation scale are taken to be the same value.

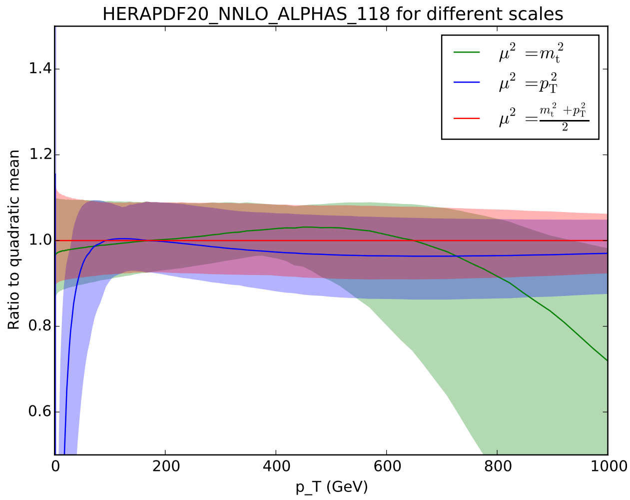


Figure 5: The differential $t\bar{t}$ production cross section as a function of p_{T} is shown for the different factorization and renormaliziation scales.

As expected the predictions for all three scales agree at $p_{\mathrm{T}} = m_{\mathrm{t}}$. For the most part the difference between the scale choices is covered by the uncertainty bands. Only for low values of p_{T} using this value as a scale leads to significant differences to the other scales. For this reason the transverse momentum is not an acceptable scale choice. For the fits in this project the scale is chosen to be the mass of the top quark from now on.

2.2 χ^2 - Implementation in Xfitter

The second objective of the project was to implement a new χ^2 function within the Xfitter framework. This was done, because the usual χ^2 definition presented in section 1.4 of this report has several disadvantages. The problem of rescaling the relative uncertainties at each step was already discussed. Rescaling causes the covariance matrix to change at each step of the iteration. For this reason the inversion of the matrix must be carried out at during each iteration. This numeric overhead would be reduced by using a χ^2 definition which does not include a changing covariance matrix. A χ^2 definition derived from the log normal probability distribution can solve this problem. This χ^2 function is given by

$$\chi^2 = \sum_{ij} (\ln d_i - \ln t_i(\vec{a})) \left(V_{\mathrm{rel}}^{-1}\right)_{ij} (\ln d_j - t \ln_j(\vec{a})) \tag{4}$$

and includes the relative uncertainties instead of absolute errors. With this χ^2 function there is no need for rescaling and therefore no changing covariance matrix. Basing the χ^2 on the log-normal probability distribution has further advantages. For data with low statistics and high uncertainties the log-normal probability distribution describes the data better than a Gaussian. An additional reason employing a χ^2 definition as

	reference fit	XFITTER default	log-normal χ^2
Total χ^2 / dof	1363 / 1131	1339 / 1131	1344 / 1131

Table 1: Values of the χ^2 function after minimization for three different χ^2 definitions.

described by equation 4 in a PDF fit is that the log-normal p.d.f. only allows for positive cross sections.
Large parts of the Xfitter framework are written in the FORTRAN programming language. The χ^2 function was implemented according to equation 4. In order to ensure the validity of the added code test-cases were developed. This was done using an already existing implementation of the log-normal χ^2 in an alternative fitting program named Alpos. Two tests were carried out: First the χ^2 values for a single iteration were compared. This test resulted in a exact agreement of the computed χ^2 value in both implementations. In order to catch errors that surface only after multiple iterations during the fitting process a second test was carried out. For this a full fit in both Alpos and Xfitter with an almost identical starting configuration was realized.
In order to compare the new χ^2 implementation to the default Xfitter a fit on HERA data[5, 6]. This data includes charged and neutral current deep inelastic scattering cross sections in ep collisions. The fits were done with three different χ^2 settings. These include the HERA default setting with linear error scaling and a bias correction term, the log normal χ^2 and a χ^2 definition that uses linear error scaling but does not contain the correction term.
The resulting χ^2 values after the minimization has converged are shown in table 1. These values are similar for all tested χ^2 definitions. The plots in figure 6 show the PDFs for the up valance quark, the gluon and the sea quarks. The results show that the differences between the different χ^2 functions are relatively small. In regions of x for which the PDF is not going towards zero the relative deviation is between the different χ^2 definition on the order of a few percent.

3 Inclusion of $7\,\mathrm{TeV}$ $t\bar{t}$ data

The main objective of the project was to study the impact of LHC $t\bar{t}$ data[7] on the parton distributions functions. This is done by including $t\bar{t}$ production cross sections taken at a center of mass energy of $7\,\mathrm{TeV}$ [7] into a fit. The results of this fit are compared to a fit using HERA data only. The log-normal χ^2 function introduced in the previous sections is used in these fits. A drawback of the $7\,\mathrm{TeV}$ $t\bar{t}$ data is the missing breakdown of uncertainties. In the published data set there is only a statistical error and an uncorrelated uncertainty included.
Some of the resulting PDFs is shown in figure 10 with a scale of $Q^2 = m_{\mathrm{t}}^2$. In this figure the PDFs for the up valence quark, the gluon and the sea quarks are displayed.
The inclusion of the $7\,\mathrm{TeV}$ LHC data reduces the uncertainty band on the gluon PDF especially at high x values. In general one can see a slight shift in the central value of

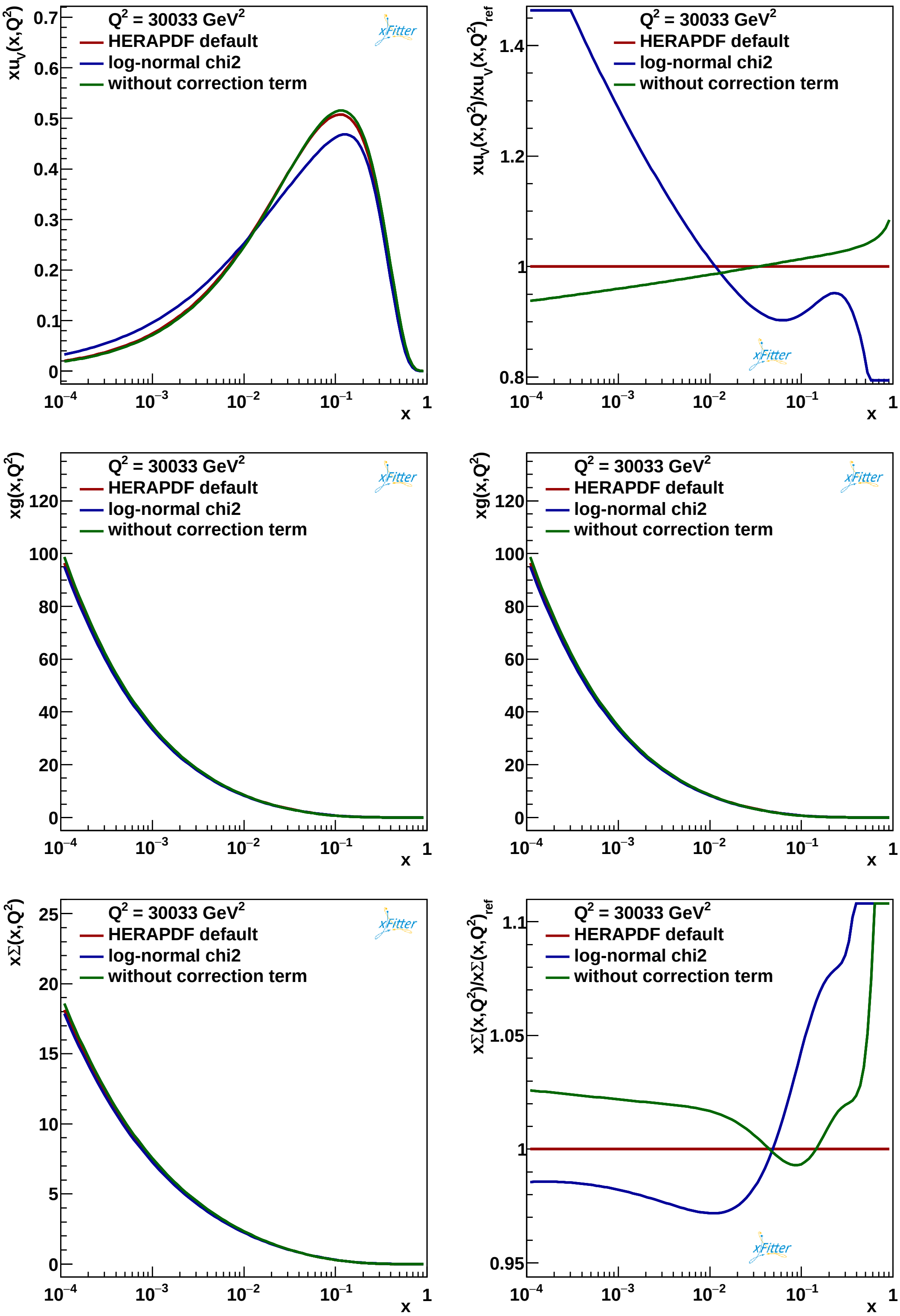


Figure 6: Overview of the fit results with HERA data comparing the log normal χ^2 definition to the HERAPDF and the Xfitter default settings.

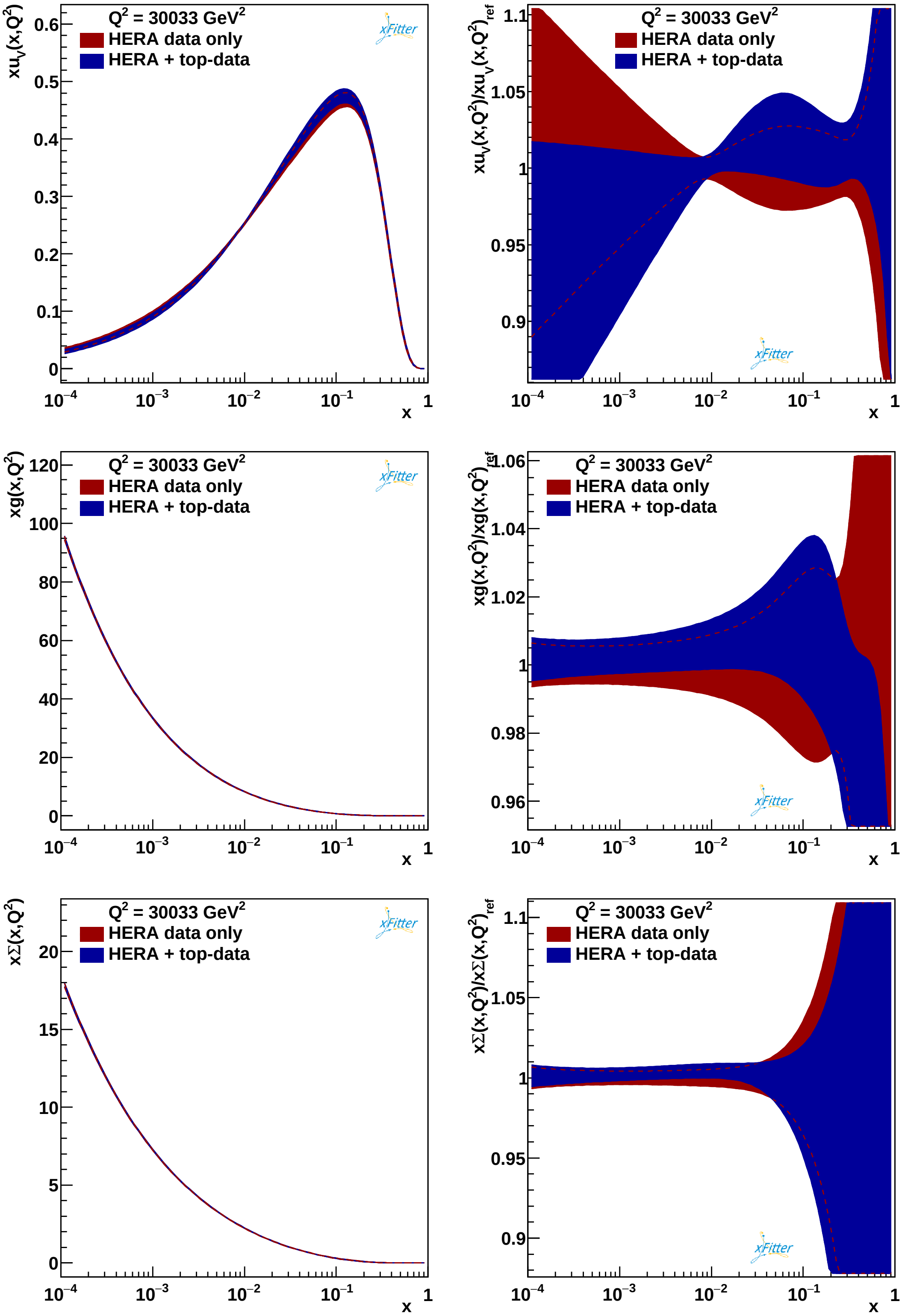


Figure 7: Overview of the fit results with 7 TeV $t\bar{t}$ data compared to the reference fit using HERA data only.

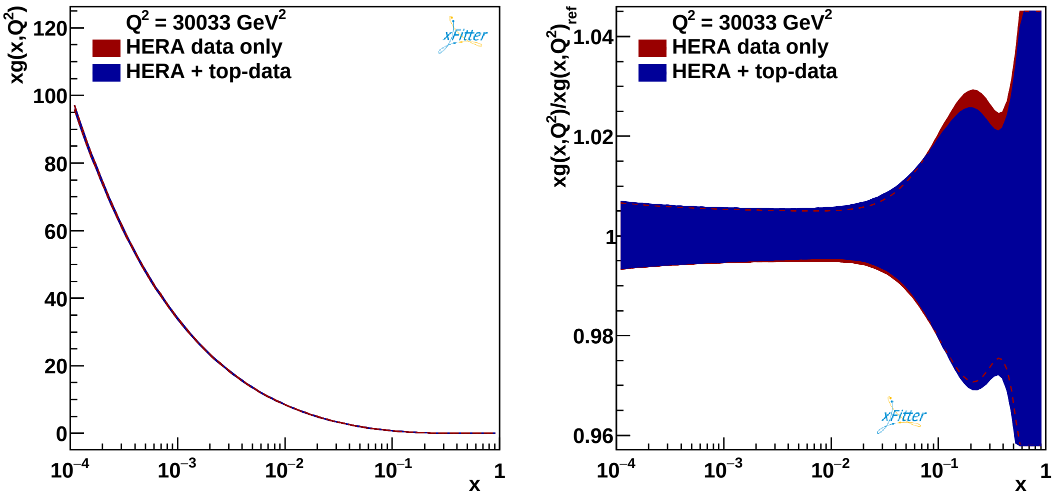


Figure 8: Gluon PDF fro a fit with 7 TeV $t\bar{t}$ data using the HERAPDF default χ^2 settings. Plots of the other PDFs can be found in the appendix.

the parton density functions. But for most partons there is no change in the size of the error bands. The gluon distribution is of particular interest, because as it was shown the largest contribution to the $t\bar{t}$ production cross section comes from gluon induced processes.
These results can be compared to a fit using the HERAPDF default χ^2 settings. The results of these fits can be found in figure 8. There is no impact of the 7 TeV data visible. This shows that the log normal χ^2 definition is more sensitve to the $t\bar{t}$ data with higher uncertainties.
In order to further reduce the PDF uncertainties it would be sensible to include higher energy respectively higher precision $t\bar{t}$ data from the Large Hadron Collider into the analysis. Alternatively the completed correlated systematic uncertainties of the 7 TeV data could be included into the fit, which would better model the fluctuations in the dataset than the use of a single uncorrelated value for the systematics.

4 Summary and conclusion

In the first part of the project several influences on the theory predictions were studied. It was shown how the center of mass energy, the PDF parameterization and factorization and renormalization scales impacted the $t\bar{t}$ production cross sections. Another important aspect was the composition of the cross section. Especially at higher energies the cross section is dominated by gluon induced processes.
The second objective of the project was to include a χ^2 function based on the log normal probability distribution function into the XFITTER fitting frame work. This definition of the χ^2 is more suited for fitting data with high uncertainties or low statistics. It also simplifies the handling of relative uncertainties in an iterative fit. The implementation

was crosschecked against an other PDF fitting program and also against other χ^2 definitions within the XFITTER framework. Some small differences between the different χ^2 functions were noticeable. But these deviations are of the some order of magnitude as other small influences in the fit like bias correction. It can be concluded that the log normal χ^2 definition can be used in PDF fitting.
This newly implemented feature was then used to do extend an existing fit of HERA data with $t\bar{t}$ production cross section taken at the LHC. The inclusion of this data only impacted the fit results noticeably when using the log normal χ^2 definition instead of the default HERAPDF settings. This result shows that the log normal χ^2 function is more sensitive to lower statistics data. Inclusion of the LHC $t\bar{t}$ data led to shifts in the central value of the PDFs and also lowered the uncertainty of the gluon PDF at high values of x. This confirms the predictions of the theory studies.
In order to further reduce the uncertainties further fits with higher precision data should be done. This would include 8 TeV and 13 TeV $t\bar{t}$ production cross section data from the LHC. These datasets were not available at the moment and the fits could not be done in the timescale of the summer student project.

5 Appendix

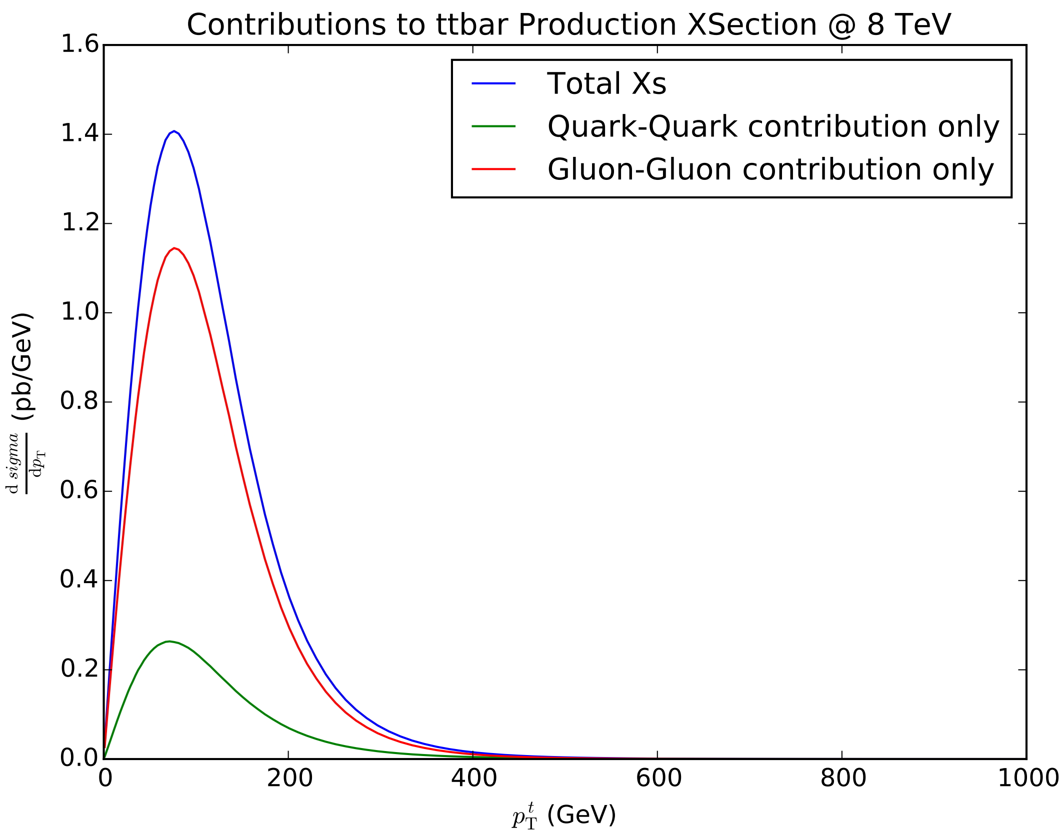


Figure 9: Different contributions to the $t\bar{t}$ production cross section as a function of p_{T} for a center of mass energy of 8 TeV.

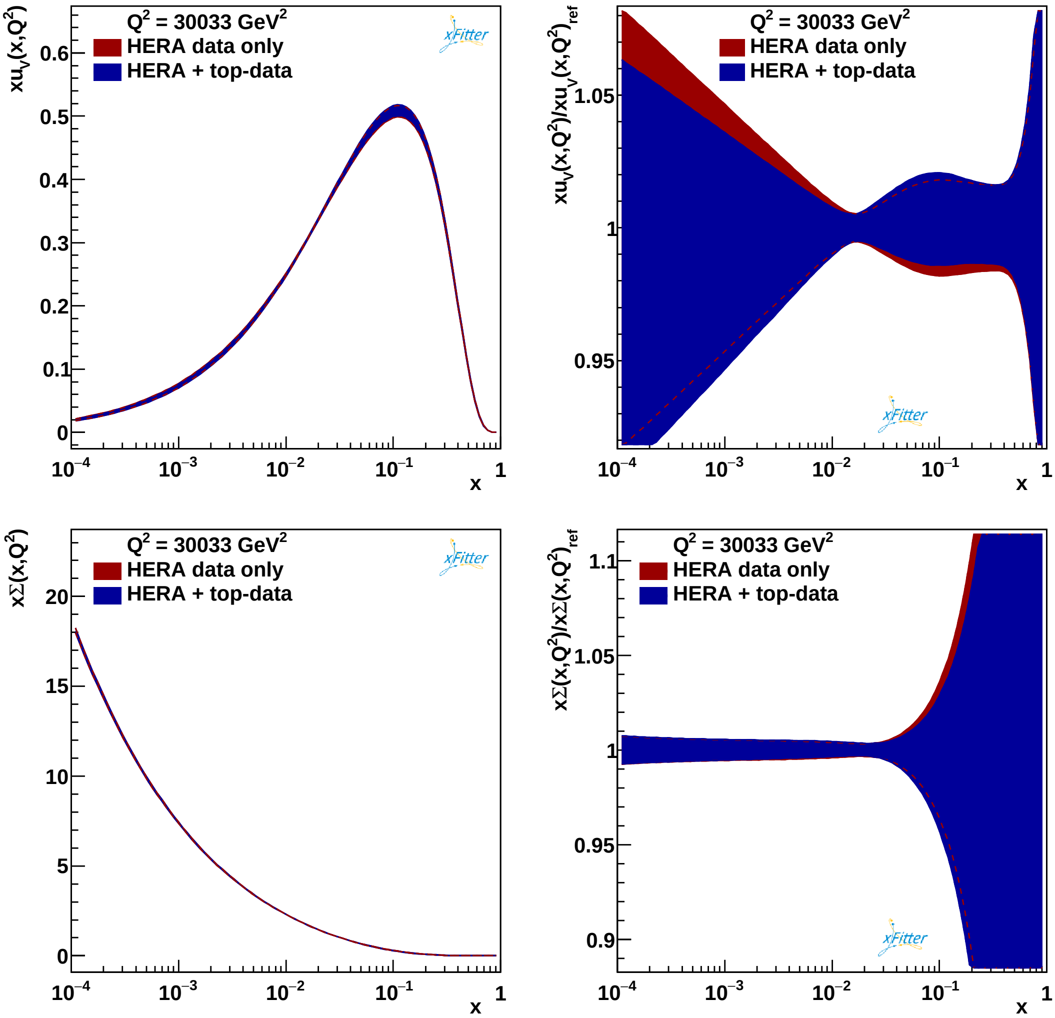


Figure 10: Overview of the fit results with 7 TeV $t\bar{t}$ data using the HERAPDF deafult χ^2 settings.

References

[1] M. Guzzi, K. Lipka, and S.-O. Moch, "Top-quark pair production at hadron colliders: differential cross section and phenomenological applications with DiffTop", JHEP **01**, 082 (2015), arXiv:1406.0386 [hep-ph].

[2] T. Kluge, K. Rabbertz, and M. Wobisch, "FastNLO: Fast pQCD calculations for PDF fits", in Deep inelastic scattering. Proceedings, 14th International Workshop, DIS 2006, Tsukuba, Japan, April 20-24, 2006 (2006), pp. 483–486, arXiv:hep-ph/0609285 [hep-ph].

[3] F. James, and M. Roos, "Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations", Comput. Phys. Commun. **10**, 343–367 (1975).

[4] S. Alekhin, et al., "HERAFitter", Eur. Phys. J. **C75**, 304 (2015), arXiv:1410.4412 [hep-ph].

[5] F. D. Aaron, et al., "Combined Measurement and QCD Analysis of the Inclusive e+- p Scattering Cross Sections at HERA", JHEP **01**, 109 (2010), arXiv:0911.0884 [hep-ex].

[6] H. Abramowicz, et al., "Combination of measurements of inclusive deep inelastic $e^{\pm}p$ scattering cross sections and QCD analysis of HERA data", Eur. Phys. J. **C75**, 580 (2015), arXiv:1506.06042 [hep-ex].

[7] G. Aad, et al., "Measurements of normalized differential cross sections for $t\bar{t}$ production in pp collisions at $\sqrt{s} = 7$TeV using the ATLAS detector", Phys. Rev. **D90**, 072004 (2014), arXiv:1407.0371 [hep-ex].