

# **Extraction of parton distribution functions using recent LHCb and ALICE heavy-flavour measurements**

PROSA Collaboration

## **Abstract**

The impact of recent measurements of heavy-flavour production in deep inelastic ep scattering and in pp collisions on parton distribution functions is studied in a QCD analysis at next-to-leading order. Recent combined results of inclusive and heavy-flavour production cross sections in deep inelastic scattering at HERA are investigated together with heavy-flavour production measurements at the LHC. Differential cross sections of charm- and beauty-hadron production measured by the LHCb collaboration at the center of mass energies of 5, 7 and 13 TeV as well as the measurements of the ALICE experiment at the center of mass energies of 5 and 7 TeV are explored. These data impose additional constraints on the gluon and the sea-quark distributions at low partonic fractions  $x$  of the proton momentum, down to  $x \approx 10^{-6}$ , which is the kinematic region of relevance for high-energy neutrino production. The impact of the resulting uncertainties in the nucleon composition, affecting predictions for the atmospheric prompt neutrino fluxes is presented.

# 1 Introduction

The fundamental structure of the nucleon is described by the theory of strong interaction, quantum chromodynamics (QCD). In the collinear factorisation, the nucleon structure is expressed in terms of parton distribution functions (PDFs), defined as probability densities for parton to carry  $x$  fraction of the nucleon momentum at a factorisation scale  $\mu_f$ . While the scale evolution of the PDFs is calculated in perturbative QCD (pQCD) and presented by the DGLAP equations [?], the  $x$ -dependence must be constrained using the experimental measurements. The constraining power of experimental data on particular parton distribution is to large extent defined by the acceptance of the experiment. Measurements of neutral current (NC) and charged current (CC) cross sections in deep inelastic scattering (DIS) at HERA probe the  $x$ -range of  $10^{-4} < x < 10^{-1}$  and impose most significant constraints on the light quark PDFs and probe the gluon distribution via scaling violations. Additional constraints on flavour separation of the quark sea and on the gluon distribution at low and high  $x$  are obtained by using the measurements at fixed target experiments and in proton-(anti)proton collisions. The measurements of heavy-flavour production in proton-proton collisions at the LHC is dominated by the gluon-gluon fusion and probe the gluon distribution directly. The measurements of forward charm [1] and beauty [2] production by the LHCb experiment at the center-of-mass energy  $\sqrt{s}=7$  TeV was used for the first time by the PROSA collaboration [3] to improve constraints on the gluon distribution at  $10^{-4} < x < 5 \times 10^{-6}$ , in the region not covered by any other data to that date. The resulting PDFs were further used to investigate the uncertainties of the prompt neutrino fluxes in the atmosphere [4].

Recent improvements in precision of HERA measurements, new experimental data on heavy flavour production at the LHC at different  $\sqrt{s}$ , together with new developments on the theory and improvements in the phenomenological tools, offer possibilities for stronger constraints of the gluon distribution at low  $x$ . These improvements in experimental measurements and the theory are explored in the QCD analysis presented in this paper, which updates the earlier result [3]. The results are used in the predictions for the prompt atmospheric neutrino fluxes.

## 2 Input data sets and used theory predictions

The main objective of the present QCD analysis is to demonstrate the constraining power of the updated measurements of heavy-flavour production in DIS and proton-proton collisions for the determination of the PDFs of the proton. The QCD analysis is performed at next-to-leading order (NLO) using the xFitter framework [5]. The updated combinations of the inclusive DIS cross sections [6] and of charm and beauty production cross sections [7] are used together with the measurements of charm and beauty hadroproduction in proton-proton scattering at the LHC. The latter include the measurements of charm hadroproduction by the LHCb collaboration at 5 TeV [8], 7 TeV [1] and 13 TeV [9], and by ALICE at 5 TeV [10] and 7 TeV [11]. The measurements of beauty hadroproduction by the LHCb at 7 TeV [2] are also used.

The cross sections measured by the LHCb and ALICE in each  $p_T$  range are normalized in rapidity  $y$ ,  $\frac{d^2\sigma}{dydp_T}/(\frac{d^2\sigma}{dydp_T})_0$ . Here,  $(\frac{d^2\sigma}{dydp_T})_0$  is the cross section in the central LHCb rapidity bin,  $3 < y < 3.5$ . Note that ALICE measurements at  $|y| < 0.5$  are normalised to the LHCb cross section

measurement in  $3 < y < 3.5$ . The advantage of using the normalised cross section, demonstrated in the earlier PROSA analysis [3], is a significant reduction of the scale dependence in the theoretical prediction, while retaining the sensitivity to the PDFs.

In the presented QCD analysis, bin-to-bin correlations in the input measurements are taken into account as described in the following. The treatment of correlated experimental uncertainties for the HERA data follows that of the original publications [6, 7].

The correlated uncertainties in the ALICE and LHCb measurements reported in original publications [1, 2, 8–11] as single numbers in  $p_T$  and  $y$  bins are treated as fully correlated, and the uncorrelated uncertainties are obtained by subtracting the correlated ones from the total uncertainties, in quadrature. The systematic uncertainties, reported as error intervals, see e.g. Table (2) of [8], are assumed uncorrelated, since no details about their size in individual  $p_T$  and  $y$  bins are provided. For different final state measurement within one experiment, the tracking and luminosity uncertainties are treated as correlated. All experimental uncertainties are treated as uncorrelated among measurements at different center-of-mass energies. The uncorrelated uncertainties in the normalized cross sections  $\frac{d\sigma}{dy_0}$  are propagated as correlated uncertainties to the respective complementary rapidity bins.

In the presented QCD analysis, the scale evolution of partons is calculated through DGLAP equations at NLO, as implemented in the QCDNUM programme [12].

The theoretical FFNS predictions for the HERA data are obtained using OPENQCDRAD [12] code in the fixed-flavour-number scheme (FFNS) with the three active flavours in the proton, following the Ref. [7]. Similar to the earlier PROSA analysis [3], the theoretical predictions for the fully differential heavy-quark hadroproduction in proton-proton collisions, available at NLO in FFNS, are used. These are calculated using the MNR code, with the single-particle inclusive distributions computed using the pole mass scheme for the heavy quarks, and translated into the  $\overline{\text{MS}}$  scheme expressions following the Ref. [13]. The  $\overline{\text{MS}}$  mass scheme is then consistently used in the calculations for all used processes.

The factorisation and renormalisation scales are chosen to be  $Q$  for the inclusive DIS, and  $\mu_r = \mu_f = \sqrt{Q^2 + 4m_Q(m_Q)^2}$  for the heavy quark production in DIS, respectively, with  $m_Q(m_Q)$  representing the heavy-quark mass in the  $\overline{\text{MS}}$  scheme. For heavy quark production in proton-proton collision,  $\mu_r = \mu_f = \sqrt{4m_Q(m_Q)^2 + p_T^2}$  is assumed.

The calculations for the heavy quark hadroproduction are supplemented with phenomenological non-perturbative fragmentation functions to describe the transition of heavy quarks into hadrons. The fragmentation of charm quarks into D-mesons is described by the Kartvelishvili function with  $\alpha_K = 4.4 \pm 1.7$  as measured at HERA [14, 15], and for the fragmentation of beauty quarks  $\alpha_K = 11 \pm 3$  is used as measured at LEP [16], following the previous PROSA analysis [3].

The main QCD analysis is performed in the FFNS and the sensitivity of the heavy quark measurements to the PDFs and to the masses of the charm and beauty quarks is fully explored by treating  $m_c(m_c)$  and  $m_b(m_b)$  as free parameters in the fit. Alternatively, the fit is performed in the variable flavour number scheme (VFNS) to allow for possible applications of the results in the LHC analyses, e.g. tuning of the underlying event in the Monte Carlo simulations.

### 3 PDF parametrisation

The PDFs are parametrised at the starting evolution scale of  $\mu_{f0}^2 = 1.9 \text{ GeV}^2$ , similar as in Ref. [6] and Ref. [17], as follows:

$$\begin{aligned}
xg(x) &= A_g x^{B_g} (1-x)^{C_g} (1 + F_g \log x), \\
u_v(x) &= A_{u_v} x^{B_{u_v}} (1-x)^{C_{u_v}} (1 + D_{u_v} x), \\
d_v(x) &= A_{d_v} x^{B_{d_v}} (1-x)^{C_{d_v}}, \\
x\bar{U}(x) &= A_{\bar{U}} x^{B_{\bar{U}}} (1-x)^{C_{\bar{U}}} (1 + D_{\bar{U}} x), \\
x\bar{D}(x) &= A_{\bar{D}} x^{B_{\bar{D}}} (1-x)^{C_{\bar{D}}}.
\end{aligned} \tag{1}$$

Here,  $xg(x)$ ,  $xu_v(x)$  and  $xd_v(x)$  represent the gluon, up and down valence quark distributions, respectively. The sea quark distribution is defined as  $x\Sigma(x) = x\bar{u}(x) + x\bar{d}(x) + x\bar{s}(x)$ , with  $x\bar{u}(x)$ ,  $x\bar{d}(x)$ , and  $x\bar{s}(x)$  denoting the up, down, and strange antiquark distributions, respectively. For the up- and down-type antiquark distributions,  $x\bar{U}(x)$  and  $x\bar{D}(x)$ , relations  $x\bar{U}(x) = x\bar{u}(x)$  and  $x\bar{D}(x) = x\bar{d}(x) + x\bar{s}(x)$  are assumed. The normalisation parameters  $A_{u_v}$ ,  $A_{d_v}$ , and  $A_g$  are determined by the QCD sum rules. The strangeness fraction  $f_s = x\bar{s}/(x\bar{d} + x\bar{s})$  is fixed to  $f_s = 0.4$  as in the HERAPDF2.0 analysis [6]. Additional constraints  $B_{\bar{U}} = B_{\bar{D}}$  and  $A_{\bar{U}} = A_{\bar{D}}(1 - f_s)$  are imposed to ensure the same normalisation for the  $x\bar{u}$  and  $x\bar{d}$  distributions as  $x \rightarrow 0$ . The term  $F_g \log x$  was proposed in [17] to provide a more flexible functional form at low  $x$ .

For all measured data, the predicted and measured cross sections together with their corresponding uncertainties are used to build a global  $\chi^2$ , minimized to determine the initial PDF parameters. The  $\chi^2$  definition follows that of Eq. (32) in Ref. [6]. In the minimisation, performed using the MINUIT package [18], the experimental uncertainties in the heavy-quark normalised cross sections are treated as additive, and the treatment of the experimental uncertainties for the HERA DIS data follows the prescription given in Ref. [6].

The parameters in Eqs.(1) are selected by first fitting with all D, E and F parameters set to zero. The other parameters are then included in the fit one at a time. The improvement in  $\chi^2$  of the fits is monitored and the procedure is stopped when no further improvement is observed. The inclusion of the  $F_g$  parameter does not lead to sizable change in  $\chi^2$ , in particular, its fitted value is consistent with 0 within uncertainty, however the variation of  $F_g$  significantly affects the fit uncertainties.

To ensure that the gluon PDF at low  $x$  is not overconstrained in the fit, different functional forms in the parametrisation are tested, as used in the ABMP16 [19], CT14 [20], HERAPDF2.0 [6], MMHT2014 [21] and Bonvini-Giuli (BG) [17] PDF fits:

$$\begin{aligned}
\text{ABMP16:} \quad & xg(x) = A(1-x)^b x^{a(1+\gamma_1 x)}, \\
\text{CT14:} \quad & xg(x) = Ax^{a_1}(1-x)^{a_2}(e_0(1-y)^2 + e_1(2y(1-y)) + y^2), y = 2\sqrt{x} - x, \\
\text{MMHT2014:} \quad & xg(x) = Ax^B(1-x)^C(1 + a_1y + a_2(2y^2 - 1)) + A'_g x^{B'_g}(1-x)^{25}, y = 1 - 2\sqrt{x}, \\
\text{HERAPDF2.0:} \quad & xg(x) = A_g x^{B_g}(1-x)^{C_g} + A'_g x^{B'_g}(1-x)^{25}, \\
\text{HERAPDF2.0 no flex. } g: \quad & xg(x) = A_g x^{B_g}(1-x)^{C_g}, \\
\text{BG:} \quad & xg(x) = A_g x^{B_g}(1-x)^{C_g}(1 + F_g \log x + G_g \log^2 x),
\end{aligned} \tag{2}$$

These functional forms are characterised by 3 (HERAPDF2.0 no flex.  $g$ ), 4 (ABMP16), 5 (CT14, HERAPDF2.0, BG) or 7 (MMHT2014) parameters controlling the gluon PDF (c.f. 4 parameters in the presented nominal parametrisation of Eq. 1). The resulting gluon distributions are presented in Fig. 4. The parametrisations of ABMP16, HERAPDF2.0 without the flexible gluon, and BG provide very similar results to that of the nominal parametrisation in Eq. 1. Note that also the HERAPDF2.0 analysis considered the parametrisation without the flexible gluon, stamped it as ‘alternative’ gluon parameterisation ?? and provided primarily for predictions of cross sections at very low  $x$ , such as very high-energy neutrino cross sections.

The fit using the HERAPDF2.0 and CT14 parametrisations yielded a gluon distribution with a sharp turnover to negative values at  $x \sim 10^{-6}$ , i.e. at the edge of the kinematic reach of the used measurements. Using such PDFs would lead to a negative prediction for the total charm hadroproduction cross sections at  $\sqrt{s} \gtrsim 20$  TeV, similar to the observation of Ref. [22]. Therefore these parametrisations are discarded.

The MMHT2014 parametrisation is similar to the one used in HERAPDF2.0 at low  $x$ , but has more flexibility at high  $x$ . In the presented analysis, it was not possible to achieve convergence using the MMHT2014 parametrisation, because of not essential sensitivity of the used data sets to the gluon distribution at high  $x$ . However discarding parameters controlling the high- $x$  behaviour would result in the HERAPDF2.0 parametrisation.

Should we include description of HERAPDF2.0 without the “flexible gluon” here?

## 4 PDF uncertainties

The PDF uncertainties are investigated according to the general approach of HERAPDF2.0 analysis [6], with the fit, model, and parametrisation uncertainties taken into account.

The fit uncertainties arising from the uncertainties in the measurements are estimated by using the Hessian method [23], adopting the tolerance criterion of  $\Delta\chi^2 = 1$  and correspond to 68% confidence level.

To investigate the impact of model assumptions on the resulting PDFs, alternative fits are performed and the differences to the central result are considered as model uncertainties. The strangeness fraction is varied as  $0.3 \leq f_s \leq 0.5$  and the value of  $Q_{\min}^2$  imposed on the HERA data as  $2.5 \leq Q_{\min}^2 \leq 5.0 \text{ GeV}^2$ . The FFNS strong coupling constant is assumed as  $0.105 < \alpha_s^{n_f=3}(M_Z) < 0.107$  (corresponding to the VFNS values of  $0.117 < \alpha_s^{n_f=5}(M_Z) < 0.119$  [24]). The variation

of the fragmentation parameters  $\alpha_K = 4.4 \pm 1.7$  for charm hadrons and  $\alpha_K = 11 \pm 3$  for beauty hadrons is performed. The scales  $\mu_f$  and  $\mu_r$  for heavy quark production are varied independently and simultaneously up and down by a factor of two. Note that for the normalized cross section predictions, the simultaneous variation of the  $\mu_f$  and  $\mu_r$  scales results in the largest deviation in the resulting PDFs and is considered as one PDF uncertainty eigenvector.

The parametrisation uncertainty is estimated by extending the functional form in Eq. (1) with additional parameters, which are added or removed one at a time and do not impact the  $\chi^2$ . In particular, the shape of the gluon PDF is extended by adding the  $+G_g \log^2 x$  term [17]. This modification does not result in an improvement in  $\chi^2$ , however leads to the fit instability, once additional parameters are introduced during assessment of the parametrisation uncertainties. Therefore, this term is not considered in the nominal parametrisation. The variation of the starting scale,  $1.6 < \mu_{f0}^2 < 2.2 \text{ GeV}^2$ , is also taken into account as contribution to the parametrisation uncertainty. The parametrisation uncertainty is constructed as an envelope built from the maximal differences between the PDFs resulting from all the parameterization variations and the central fit at each  $x$  value.

The total PDF uncertainty is obtained by adding experimental, model, and parameterization uncertainties in quadrature.

## 5 Results

The quality of the overall fit can be judged based on the global  $\chi^2$  divided by the number of degrees of freedom,  $n_{dof}$ . For each data set included in the fit, a partial  $\chi^2$  divided by the number of measurements (data points),  $n_{dp}$ , is provided. The correlated part of  $\chi^2$  quantifies the influence of the correlated systematic uncertainties in the fit. The global and partial  $\chi^2$  values for each data set are listed in Table 1, illustrating a general agreement among all the data sets. The central values and the uncertainties of the fitted parameters are given in Table 2.

The fitted PDFs with their total uncertainties at the scale  $\mu_f^2 = 10 \text{ GeV}^2$  are shown in Fig. 1. These are compared to the result of the PROSA 2015 fit [3]. In Fig. 2 (left), the gluon distribution normalised to the one from the PROSA 2015 fit is shown. The two results are in a very good agreement and a significant improvement in the precision of the gluon PDF is achieved at  $x < 10^{-4}$ , as compared to the PROSA 2015 fit. The valence and sea quark PDFs are in good agreement with the result of HERAPDF2.0 analysis [6] and the observed differences in these distributions to the PROSA 2015 analysis are attributed to the update of the DIS measurements [25] used in Ref. [3] to the final combination [6] of the HERA data.

The relative total, fit, model and parametrisation uncertainties for the gluon PDF are shown in Fig. 2 (right). The total uncertainties are dominated by the model uncertainties, with the largest contributions arising from the scale variations in predictions for heavy-quark hadroproduction. Reduction of these uncertainties would require theoretical calculations at higher order.

The resulting PDFs are available in the LHAPDF format at [...] in both FFNS and VFNS. The performance of the new PDFs presented here was tested by computing predictions for the inclusive and multi-jet production in DIS [26–30] and jet [31] and top quark-antiquark production [32, 33] in proton-proton collisions. The VFNS variant of the presented PDFs is used. The resulting pre-

Dataset	$\chi^2/\text{ndp}$
HERA1+2 CCep	62 / 39
HERA1+2 CCem	49 / 42
HERA1+2 NCem	227 / 159
HERA1+2 NCep 820	68 / 70
HERA1+2 NCep 920	440 / 377
HERA1+2 NCep 460	223 / 204
HERA1+2 NCep 575	223 / 254
HERA c	49 / 52
HERA b	18 / 27
LHCb 7TeV [1302.2864] dzero nor. in y	15 / 30
LHCb 7TeV [1302.2864] dch nor. in y	19 / 29
LHCb 7TeV [1302.2864] ds nor. in y	14 / 20
LHCb 7TeV [1302.2864] dstar nor. in y	16 / 22
LHCb 7TeV Bzero pT-y cross section	52 / 76
LHCb 7TeV Bch pT-y cross section	129 / 108
LHCb 7TeV Bs pT-y cross section	37 / 60
LHCb 5TeV [1610.02230] dzero nor. in y	60 / 35
LHCb 5TeV [1610.02230] dch nor. in y	25 / 35
LHCb 5TeV [1610.02230] ds nor. in y	30 / 29
LHCb 5TeV [1610.02230] dstar nor. in y	35 / 30
LHCb 13TeV [1510.01707] dzero nor. in y	111 / 60
LHCb 13TeV [1510.01707] dch nor. in y	72 / 64
LHCb 13TeV [1510.01707] ds nor. in y	69 / 55
LHCb 13TeV [1510.01707] dstar nor. in y	82 / 54
ALICE 7TeV [1702.00766] dzero nor. in y	5.1 / 8
ALICE 7TeV [1702.00766] dch nor. in y	0.75 / 7
ALICE 7TeV [1702.00766] dstar nor. in y	2.3 / 6
ALICE 5TeV [1901.07979] dzero nor. in y	6.3 / 10
ALICE 5TeV [1901.07979] dch nor. in y	5.8 / 9
ALICE 5TeV [1901.07979] ds nor. in y	2.5 / 4
ALICE 5TeV [1901.07979] dstar nor. in y	1.7 / 9
Correlated $\chi^2$	282
Log penalty $\chi^2$	-32
Total $\chi^2$ / dof	2401 / 1969

Table 1: The global and partial  $\chi^2$  values for each data set together with the corresponding number of data points (ndp). The correlated  $\chi^2$  and the log penalty  $\chi^2$  entries refer to the  $\chi^2$  contributions from the correlated uncertainties and from the logarithmic term, respectively, as described in Ref. [6].

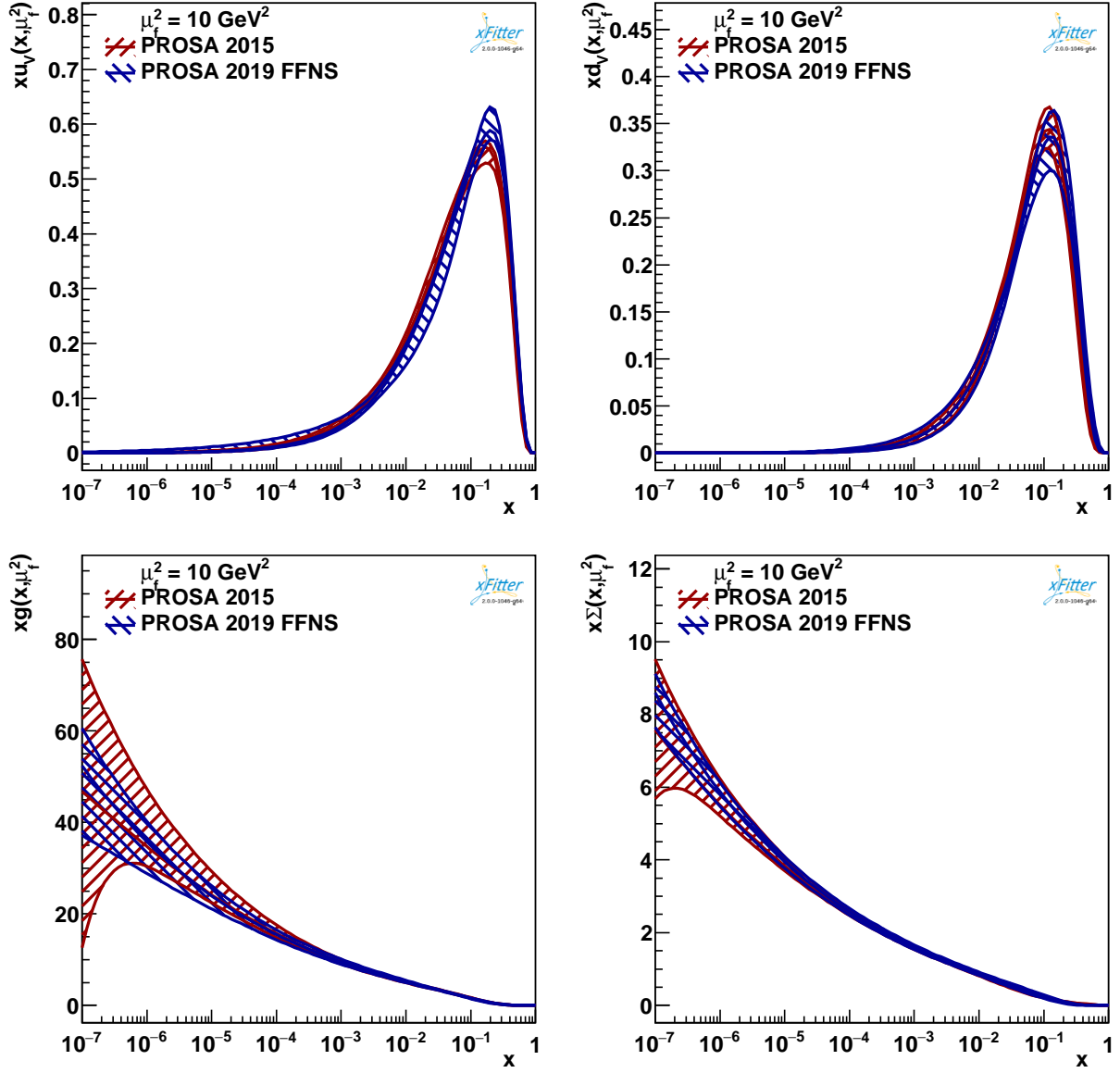


Figure 1: The fitted PDFs with their total uncertainties at the scale  $\mu_f^2 = 10 \text{ GeV}^2$  compared with the distributions from the PROSA 2015 fit.



Parameter	Value
$B_g$	$0.004 \pm 0.053$
$C_g$	$6.25 \pm 0.29$
$Fg$	$0.068 \pm 0.024$
$B_{u_v}$	$0.644 \pm 0.030$
$C_{u_v}$	$4.862 \pm 0.076$
$E_{u_v}$	$15.8 \pm 2.2$
$B_{d_v}$	$0.873 \pm 0.076$
$C_{d_v}$	$4.61 \pm 0.35$
$C_{\bar{U}}$	$7.36 \pm 0.77$
$D_{\bar{U}}$	$10.1 \pm 2.4$
$A_{\bar{D}}$	$0.1061 \pm 0.0058$
$B_{\bar{D}}$	$-0.1661 \pm 0.0062$
$C_{\bar{D}}$	$12.7 \pm 3.0$
$m_c(m_c)$	$1.230 \pm 0.031 \text{ GeV}$
$m_b(m_b)$	$3.977 \pm 0.100 \text{ GeV}$

Table 2: The resulting parameters for the PDFs with their fit uncertainties.

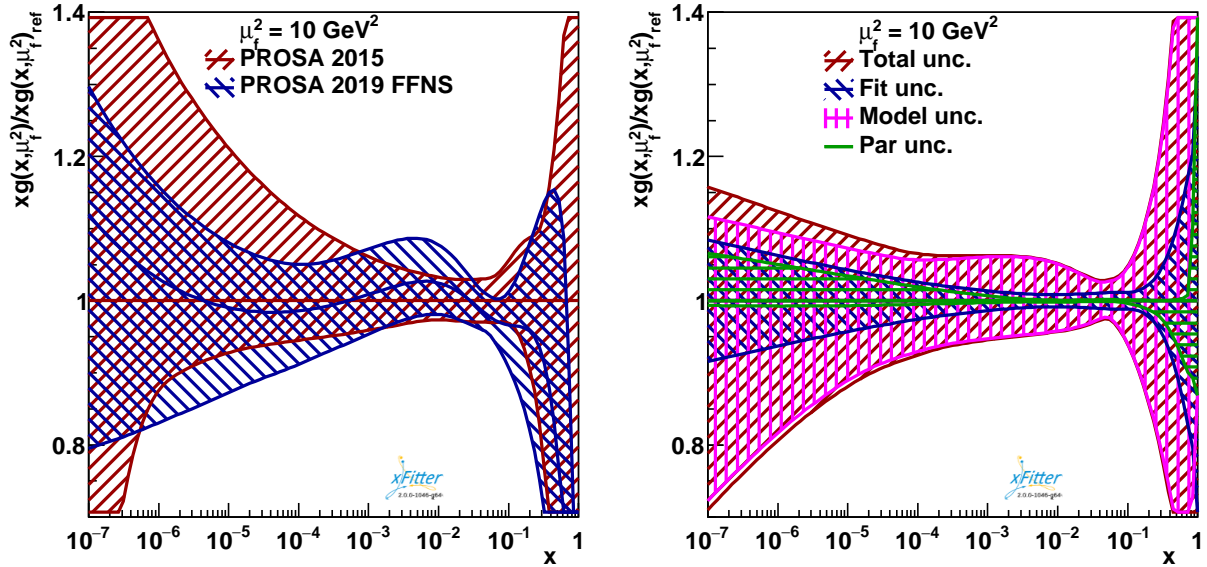


Figure 2: (left) The gluon PDF with their total uncertainties at the scale  $\mu_f^2 = 10 \text{ GeV}^2$  divided by the gluon PDF from the PROSA 2015 fit. (right) The relative total, fit, model and parametrisation uncertainties for the gluon PDF at the scale  $\mu_f^2 = 10 \text{ GeV}^2$ .

distributions are compared to the measurements by H1 [REF](#), ZEUS [REF](#) and CMS [REF](#) experiments in [would be great to have 2 representative examples in a Figure](#). The results are found similar to those using HERAPDF2.0 PDF.

## 6 Predictions for prompt atmospheric neutrino fluxes

## 7 Summary

## Acknowledgements

The work of O. Z. has been supported by Bundesministerium für Bildung und Forschung (contract 05H18GUCC1).

## References

- [1] R Aaij et al. Prompt charm production in pp collisions at  $\sqrt{s}=7$  TeV. *Nucl. Phys.*, B871:1–20, 2013.
- [2] R Aaij et al. Measurement of B meson production cross-sections in proton-proton collisions at  $\sqrt{s} = 7$  TeV. *JHEP*, 08:117, 2013.
- [3] O. Zenaiev et al. Impact of heavy-flavour production cross sections measured by the LHCb experiment on parton distribution functions at low x. *Eur. Phys. J.*, C75(8):396, 2015.
- [4] M. V. Garzelli, S. Moch, O. Zenaiev, A. Cooper-Sarkar, A. Geiser, K. Lipka, R. Placakyte, and G. Sigl. Prompt neutrino fluxes in the atmosphere with PROSA parton distribution functions. *JHEP*, 05:004, 2017.
- [5] S. Alekhin et al. HERAFitter. *Eur. Phys. J.*, C75(7):304, 2015. See [xfitter.org](http://xfitter.org).
- [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(12):580, 2015.
- [7] H. Abramowicz et al. Combination and QCD analysis of charm and beauty production cross-section measurements in deep inelastic  $ep$  scattering at HERA. *Eur. Phys. J.*, C78(6):473, 2018.
- [8] Roel Aaij et al. Measurements of prompt charm production cross-sections in pp collisions at  $\sqrt{s} = 5$  TeV. *JHEP*, 06:147, 2017.
- [9] Roel Aaij et al. Measurements of prompt charm production cross-sections in  $pp$  collisions at  $\sqrt{s} = 13$  TeV. *JHEP*, 03:159, 2016. [Erratum: *JHEP*05,074(2017)].

- [10] Shreyasi Acharya et al. Measurement of  $D^0$ ,  $D^+$ ,  $D^{*+}$  and  $D_s^+$  production in pp collisions at  $\sqrt{s} = 5.02$  TeV with ALICE. *Eur. Phys. J.*, C79(5):388, 2019.
- [11] Shreyasi Acharya et al. Measurement of D-meson production at mid-rapidity in pp collisions at  $\sqrt{s} = 7$  TeV. *Eur. Phys. J.*, C77(8):550, 2017.
- [12] S. Alekhin. “OPENQCDRAD”. See <http://www-zeuthen.desy.de/~alekhin/OPENQCDRAD/>.
- [13] Matthew Dowling and Sven-Olaf Moch. Differential distributions for top-quark hadro-production with a running mass. *Eur. Phys. J.*, C74(11):3167, 2014.
- [14] F. D. Aaron et al. Study of Charm Fragmentation into  $D^{*+}$  Mesons in Deep-Inelastic Scattering at HERA. *Eur. Phys. J.*, C59:589–606, 2009.
- [15] S. Chekanov et al. Measurement of the charm fragmentation function in  $D^*$  photoproduction at HERA. *JHEP*, 04:082, 2009.
- [16] Paolo Nason and Carlo Oleari. A Phenomenological study of heavy quark fragmentation functions in  $e^+ e^-$  annihilation. *Nucl. Phys.*, B565:245–266, 2000.
- [17] Marco Bonvini and Francesco Giuli. A new simple PDF parametrization: improved description of the HERA data. 2019.
- [18] 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.
- [19] S. Alekhin, J. Blümlein, S. Moch, and R. Placakyte. Parton distribution functions,  $\alpha_s$ , and heavy-quark masses for LHC Run II. *Phys. Rev.*, D96(1):014011, 2017.
- [20] Sayipjamal Dulat, Tie-Jiun Hou, Jun Gao, Marco Guzzi, Joey Huston, Pavel Nadolsky, Jon Pumplin, Carl Schmidt, Daniel Stump, and C. P. Yuan. New parton distribution functions from a global analysis of quantum chromodynamics. *Phys. Rev.*, D93(3):033006, 2016.
- [21] L. A. Harland-Lang, A. D. Martin, P. Motylinski, and R. S. Thorne. Parton distributions in the LHC era: MMHT 2014 PDFs. *Eur. Phys. J.*, C75(5):204, 2015.
- [22] A. Accardi et al. A Critical Appraisal and Evaluation of Modern PDFs. *Eur. Phys. J.*, C76(8):471, 2016.
- [23] J. Pumplin, D. Stump, R. Brock, D. Casey, J. Huston, J. Kalk, H. L. Lai, and W. K. Tung. Uncertainties of predictions from parton distribution functions. II. The Hessian method. *Phys. Rev. D*, 65:014013, 2001.
- [24] M. Tanabashi et al. Review of Particle Physics. *Phys. Rev.*, D98(3):030001, 2018.
- [25] H1 and ZEUS Collaborations. Combined measurement and QCD analysis of the inclusive  $e^\pm p$  scattering cross sections at HERA. *JHEP*, 01:109, 2010.

- [26] S. Chekanov et al. Inclusive jet cross-sections in the Breit frame in neutral current deep inelastic scattering at HERA and determination of  $\alpha(s)$ . *Phys. Lett.*, B547:164–180, 2002.
- [27] S. Chekanov et al. Inclusive-jet and dijet cross-sections in deep inelastic scattering at HERA. *Nucl. Phys.*, B765:1–30, 2007.
- [28] H. Abramowicz et al. Inclusive dijet cross sections in neutral current deep inelastic scattering at HERA. *Eur. Phys. J.*, C70:965–982, 2010.
- [29] A. Aktas et al. Measurement of inclusive jet production in deep-inelastic scattering at high  $Q^{*2}$  and determination of the strong coupling. *Phys. Lett.*, B653:134–144, 2007.
- [30] F. D. Aaron et al. Jet Production in ep Collisions at Low  $Q^{*2}$  and Determination of  $\alpha(s)$ . *Eur. Phys. J.*, C67:1–24, 2010.
- [31] Serguei Chatrchyan et al. Measurements of Differential Jet Cross Sections in Proton-Proton Collisions at  $\sqrt{s} = 7$  TeV with the CMS Detector. *Phys. Rev.*, D87(11):112002, 2013. [Erratum: *Phys. Rev.* D87, no. 11, 119902(2013)].
- [32] Albert M Sirunyan et al. Measurement of double-differential cross sections for top quark pair production in pp collisions at  $\sqrt{s} = 8$  TeV and impact on parton distribution functions. *Eur. Phys. J.*, C77(7):459, 2017.
- [33] Albert M Sirunyan et al. Measurement of  $t\bar{t}$  normalised multi-differential cross sections in pp collisions at  $\sqrt{s} = 13$  TeV, and simultaneous determination of the strong coupling strength, top quark pole mass, and parton distribution functions. *Submitted to: Eur. Phys. J.*, 2019.
- [34] Valerio Bertone, Stefano Carrazza, and Juan Rojo. APFEL: A PDF Evolution Library with QED corrections. *Comput. Phys. Commun.*, 185:1647–1668, 2014.
- [35] Stefano Forte, Eric Laenen, Paolo Nason, and Juan Rojo. Heavy quarks in deep-inelastic scattering. *Nucl. Phys.*, B834:116–162, 2010.
- [36] V. Bertone et al. Impact of the heavy quark matching scales in PDF fits. *Eur. Phys. J.*, C77(12):837, 2017.

## A Fit in VFNS

The fit in the VFNS is performed using the APFEL library [34] interfaced in xFitter. The theoretical predictions for the HERA data are computed using the FONLL-B scheme [35] with the pole charm and beauty quark masses set to  $m_c^{\text{pole}} = 1.4$  GeV and  $m_b^{\text{pole}} = 4.5$  GeV respectively. However, no VFNS calculation for heavy-quark hadroproduction is interfaced to public tools and can be used for PDF fitting. To keep using the MNR calculations with the VFNS, we exploit the feature of the APFEL library to choose arbitrary heavy-quark matching thresholds [36]. These thresholds are set to values:

$$\begin{aligned}\mu_c &= 4.5m_c^{\text{pole}} = 6.3 \text{ GeV}, \\ \mu_b &= 4.5m_b^{\text{pole}} = 20.25 \text{ GeV}.\end{aligned}\tag{3}$$

We imposed kinematic cuts  $p_T < 5$  GeV and  $p_T < 16$  GeV on the LHC charm and beauty data, respectively, to ensure that we are working with not more than 3 (4) light flavours when calculating predictions for charm (beauty) data. The strong coupling strength is set to  $\alpha_s^{n_f=5}(M_Z) = 0.118$  [24], while all other settings are the same as in the FFNS fit. The choice of the matching thresholds is arbitrary and amounts to a renormalisation scheme choice [36], therefore we have verified that the fit results are stable under variations within  $3.1 \leq \mu_Q/m_Q^{\text{pole}} \leq 6$ , with the  $p_T$  cuts on the charm and beauty LHC data modified accordingly.

The resulting PDFs are available in the LHAPDF format at [...]. No PDF uncertainties are provided with this set.

## B Additional plots

(These plots provided on request of M.V.: do we want them in the paper?)

The fitted PDFs with their total uncertainties at the starting scale of PDF evolution  $\mu_f^2 = 1.9 \text{ GeV}^2$  are shown in Fig. 3, superimposed with the PDFs from the PROSA 2015 fit [3]. The gluon distribution obtained in the fit using different parametrisations (see Eq. 2) are shown on Fig. 4.

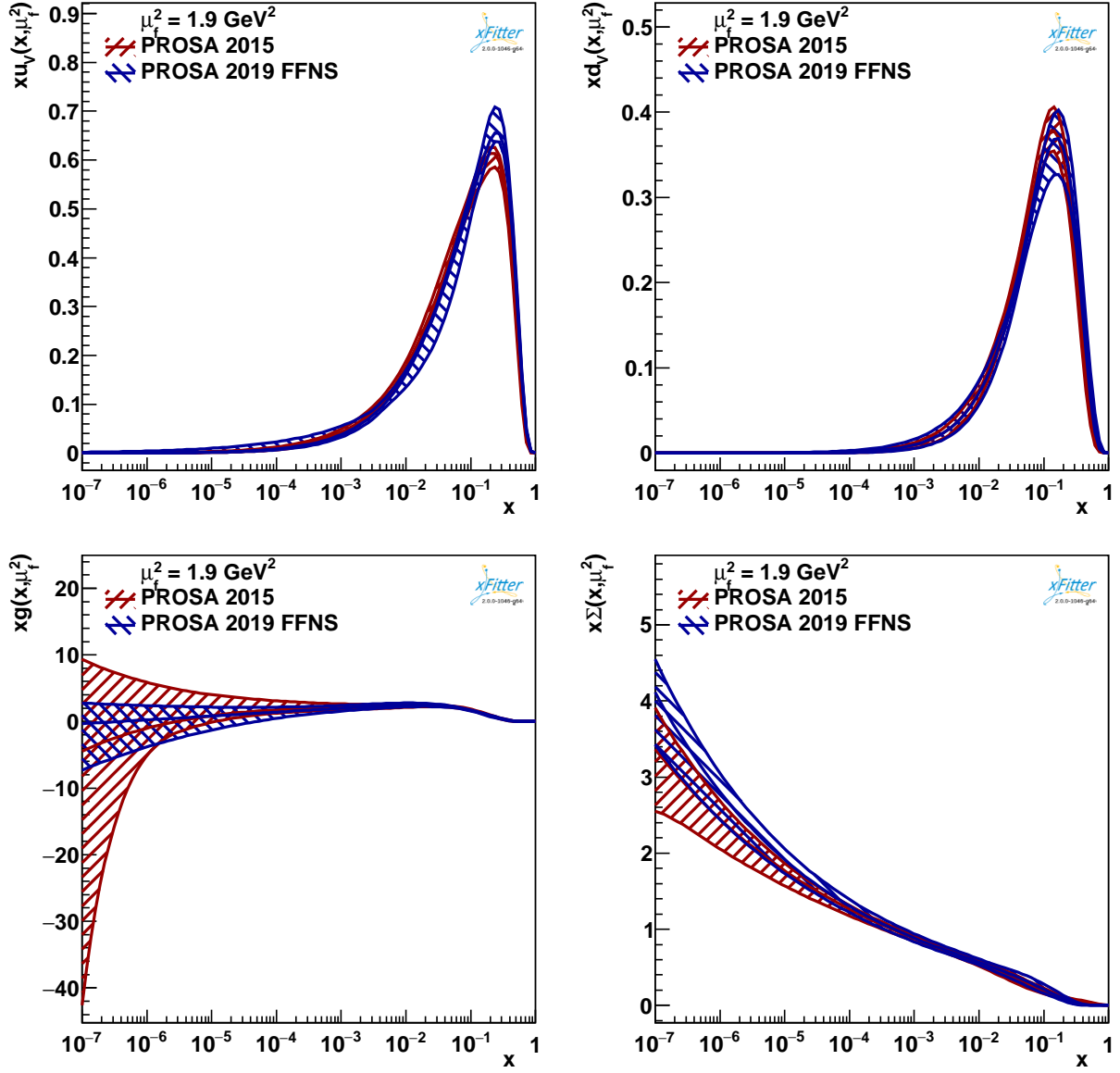


Figure 3: The fitted PDFs with their total uncertainties at the scale  $\mu_f^2 = 1.9 \text{ GeV}^2$  compared with the distributions from the PROSA 2015 fit.

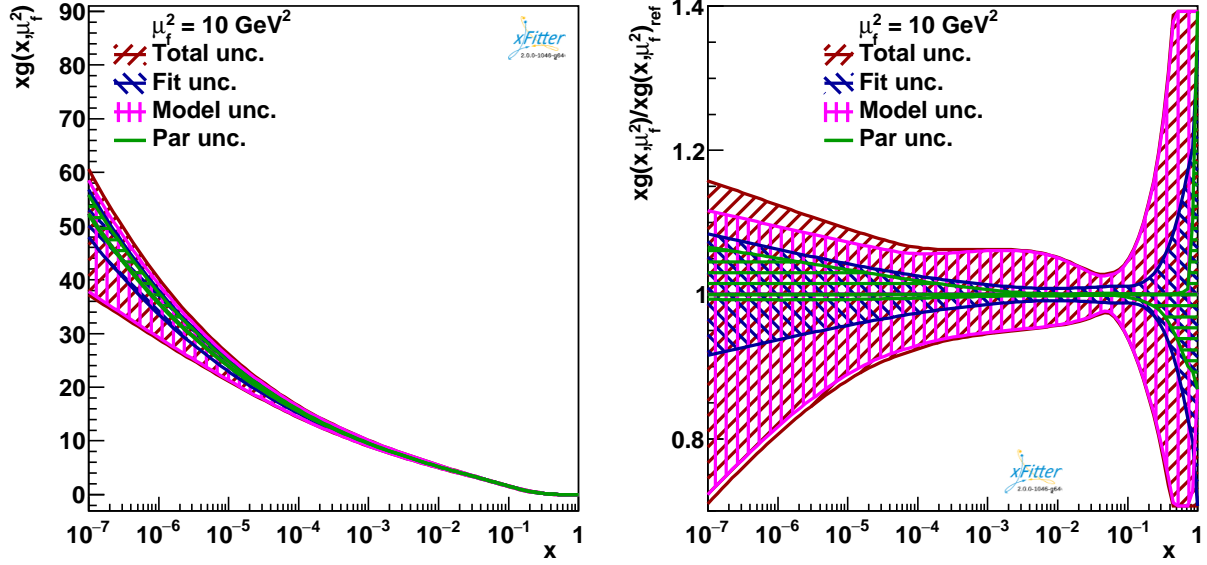


Figure 4: (left) The gluon PDF with their total uncertainties at the scale  $\mu_f^2 = 10 \text{ GeV}^2$  obtained using different gluon parametrisations (see Eq. 2). (right) The same PDFs normalised to the distribution obtained using the nominal parametrisation.