CAVENDISH-HEP-17-06
CERN-TH-2017-077
Edinburgh 2017/08
Nikhef/2017-006
OUTP-17-04P
TIF-UNIMI-2017-3

Parton distributions from high-precision collider data

The NNPDF Collaboration:
Richard D. Ball,[1] Valerio Bertone,[2] Stefano Carrazza,[3] Luigi Del Debbio,[1] Stefano Forte,[4] Patrick Groth-Merrild,[1] Alberto Guffanti,[5] Nathan P. Hartland,[2] Zahari Kassabov,[4,5] José I. Latorre,[6,7] Emanuele R. Nocera,[8] Juan Rojo,[2] Luca Rottoli,[8] Emma Slade,[8] and Maria Ubiali[9]

[1] *The Higgs Centre for Theoretical Physics, University of Edinburgh, JCMB, KB, Mayfield Rd, Edinburgh EH9 3JZ, Scotland*
[2] *Department of Physics and Astronomy, VU University, NL-1081 HV Amsterdam, and Nikhef Theory Group, Science Park 105, 1098 XG Amsterdam, The Netherlands*
[3] *Theoretical Physics Department, CERN, CH-1211 Geneva, Switzerland*
[4] *Tif Lab, Dipartimento di Fisica, Università di Milano and INFN, Sezione di Milano, Via Celoria 16, I-20133 Milano, Italy*
[5] *Dipartimento di Fisica, Università di Torino and INFN, Sezione di Torino, Via P. Giuria 1, I-10125, Turin, Italy*
[6] *Departament de Física Quàntica i Astrofísica, Universitat de Barcelona, Diagonal 645, 08028 Barcelona, Spain*
[7] *Center for Quantum Technologies, National University of Singapore*
[8] *Rudolf Peierls Centre for Theoretical Physics, 1 Keble Road, University of Oxford, OX1 3NP Oxford, United Kingdom*
[9] *Cavendish Laboratory, HEP group, University of Cambridge, J.J. Thomson Avenue, Cambridge CB3 0HE, United Kingdom*

SLAC [32]; the combined HERA-I inclusive structure function dataset [33] and HERA-II inclusive measurements from H1 and ZEUS [34–37]; the HERA combined measurements of the charm production cross-section $\sigma_c^{\rm NC}$ [38]; CHORUS inclusive neutrino DIS [39], and NuTeV dimuon production data [40, 41]. From the Tevatron, CDF [42] and D0 [43] Z rapidity distributions; and CDF [44] Run-II one-jet inclusive cross-sections were used. Constraints from fixed-target Drell-Yan came from the E605 [45] and E866 [46–48] experiments. LHC measurements included electroweak boson production data from ATLAS [49–51], CMS [52–54] and LHCb [55, 56]; one-jet inclusive cross-sections from ATLAS [57, 58] and CMS [59]; the differential distributions for W production in association with charm quarks from CMS [60]; and total cross-section measurements for top quark pair production data from ATLAS and CMS at 7 and 8 TeV [61–66].

For NNPDF3.1 we have made a number of improvements to the NNPDF3.0 dataset. Firstly we have included the final datasets for several experiments which have now concluded, replacing superseded data in the NNPDF3.0 analysis. The HERA-I data and the H1 and ZEUS HERA-II inclusive structure functions have been replaced by the final HERA combination [9]. The HERA dataset has also been enlarged by the inclusion of H1 and ZEUS measurements of the bottom structure function $F_2^b(x, Q^2)$ [67,68], which may prove useful in specific applications such as in the determination of the bottom quark mass m_b. In order to perform dedicated studies of the charm content of the proton, we have constructed a PDF set also including the EMC measurements of charm structure functions at large-x [69], which will be discussed in Sect. 5.3. However, these measurements are not included in the standard dataset. The legacy W lepton asymmetries from D0 using the complete Tevatron luminosity, both in the electron [14] and in the muon [13] channels have been added. These precise weak gauge boson production measurements provide important information on the quark flavor separation at large-x, as demonstrated in [70].

Aside from the updated legacy datsets, in NNPDF3.1 a large number of recent measurements from ATLAS, CMS and LHCb are included. For ATLAS, we now include the Z boson (p_T^Z, y_Z) and (p_T^Z, M_{ll}) double differential distributions measured at 8 TeV [71]; the inclusive W^+, W^- and Z rapidity distributions at 7 TeV from the 2011 dataset [72], the top-quark pair production normalized y_t distribution at 8 TeV [73]; total cross-sections for top quark pair production at 7, 8 and 13 TeV [74,75]; inclusive jet cross-sections at 7 TeV from the 2011 dataset [76]; and finally low mass Drell-Yan M_{ll} distributions at 7 TeV from the 2010 run [77]. The transverse momentum spectrum at 7 TeV (2011 dataset) [78] will be studied in Sec. 4.2 but it is not included in the default set. The total top cross-section is the only data point at 13 TeV which is included. For CMS, NNPDF3.1 includes the W^+ and W^- rapidity distributions at 8 TeV [79], together with their cross-correlations; the inclusive jet production cross-sections at 2.76 TeV [80]; top-quark pair production normalized $y_{t\bar{t}}$ distributions at 8 TeV [81], total inclusive $t\bar{t}$ cross-sections at 7, 8 and 13 TeV [82]; the distribution of the Z boson double differentially in (p_T, y_Z) at 8 TeV [83]. The double-differential distributions (y_{ll}, M_{ll}) in Drell-Yan production at 8 TeV [84] will be studied in Sect. 4.8 below, but it is not included in the default PDF determination. For LHCb, NNPDF3.1 includes the complete 7 and 8 TeV measurements of inclusive W and Z production in the muon channel [85,86], which supersedes all previous measurements in the same final state.

An overview of the data included in NNPDF3.1 is presented in Tables 2.1, 2.2, and 2.3, for the DIS structure function data, the fixed target and Tevatron Drell-Yan experiments, and the LHC datasets, respectively. For each dataset we indicate the corresponding published reference, the number of data points in the NLO/NNLO PDF determinations before and after (in parenthesis) kinematic cuts, the kinematic range covered in the relevant variables after cuts, and the code used to compute the NLO and NNLO results. Datasets included for the first time in NNPDF3.1 are flagged with an asterisk. The datasets not used for the default determination are in brackets. The total number of data points for the default PDF determination is 4175/4295/4285 at LO/NLO/NNLO.

In Fig. 2.1 we show the kinematic coverage of the NNPDF3.1 dataset in the $\left(x, Q^2\right)$ plane.

Experiment	Obs.	Ref.	$N_{\mathbf{dat}}$	x range	Q range (GeV)	Theory
NMC	F_2^d/F_2^p	[28]	260 (121/121)	$0.012 \leq x \leq 0.68$	$2.1 \leq Q \leq 10$	APFEL
	$\sigma^{\rm NC,p}$	[29]	292 (204/204)	$0.012 \leq x \leq 0.50$	$1.8 \leq Q \leq 7.9$	
SLAC	F_2^p	[32]	211 (33/33)	$0.14 \leq x \leq 0.55$	$1.9 \leq Q \leq 4.4$	APFEL
	F_2^d	[32]	211 (34/34)	$0.14 \leq x \leq 0.55$	$1.9 \leq Q \leq 4.4$	
BCDMS	F_2^p	[30]	351 (333/333)	$0.07 \leq x \leq 0.75$	$2.7 \leq Q \leq 15.1$	APFEL
	F_2^d	[31]	254 (248/248)	$0.07 \leq x \leq 0.75$	$3.0 \leq Q \leq 15.1$	
CHORUS	$\sigma^{\rm CC,\nu}$	[39]	607 (416/416)	$0.045 \leq x \leq 0.65$	$1.9 \leq Q \leq 9.8$	APFEL
	$\sigma^{\rm CC,\bar{\nu}}$	[39]	607 (416/416)	$0.045 \leq x \leq 0.65$	$1.9 \leq Q \leq 9.8$	
NuTeV	σ_ν^{cc}	[40,41]	45 (39/39)	$0.02 \leq x \leq 0.33$	$2.0 \leq Q \leq 10.8$	APFEL
	$\sigma_{\bar{\nu}}^{cc}$	[40,41]	45 (37/37)	$0.02 \leq x \leq 0.21$	$1.9 \leq Q \leq 8.3$	
HERA	$\sigma^p_{\rm NC,CC}$ (*)	[9]	1306 (1145/1145)	$4 \cdot 10^{-5} \leq x \leq 0.65$	$1.87 \leq Q \leq 223$	APFEL
	$\sigma^c_{\rm NC}$	[38]	52 (47/37)	$7 \cdot 10^{-5} \leq x \leq 0.05$	$2.2 \leq Q \leq 45$	
	F_2^b (*)	[67,68]	29 (29/29)	$2 \cdot 10^{-4} \leq x \leq 0.5$	$2.2 \leq Q \leq 45$	
EMC	[F_2^c] (*)	[69]	21 (16/16)	$0.014 \leq x \leq 0.44$	$2.1 \leq Q \leq 8.8$	APFEL

Table 2.1: Deep-inelastic scattering data included in NNPDF3.1. The EMC F_2^c data are in brackets because they are only included in a dedicated set but not in the default dataset. New datasets, not included in NNPDF3.0, are denoted (*). The kinematic range covered in each variable is given after cuts are applied. The total number of DIS data points after cuts is 3102/3092 for the NLO/NNLO PDF determinations (not including the EMC F_2^c data).

Exp.	Obs.	Ref.	$N_{\mathbf{dat}}$	Kin_1	Kin_2 (GeV)	Theory
E866	$\sigma^d_{\rm DY}/\sigma^p_{\rm DY}$	[48]	15 (15/15)	$0.07 \leq y_{ll} \leq 1.53$	$4.6 \leq M_{ll} \leq 12.9$	APFEL+Vrap
	$\sigma^p_{\rm DY}$	[46,47]	184 (89/89)	$0 \leq y_{ll} \leq 1.36$	$4.5 \leq M_{ll} \leq 8.5$	APFEL+Vrap
E605	$\sigma^p_{\rm DY}$	[45]	119 (85/85)	$-0.2 \leq y_{ll} \leq 0.4$	$7.1 \leq M_{ll} \leq 10.9$	APFEL+Vrap
CDF	$d\sigma_Z/dy_Z$	[42]	29 (29/29)	$0 \leq y_{ll} \leq 2.9$	$66 \leq M_{ll} \leq 116$	Sherpa+Vrap
	k_t incl jets	[87]	76 (76/76)	$0 \leq y_{\rm jet} \leq 1.9$	$58 \leq p_T^{\rm jet} \leq 613$	NLOjet++
D0	$d\sigma_Z/dy_Z$	[43]	28 (28/28)	$0 \leq y_{ll} \leq 2.8$	$66 \leq M_{ll} \leq 116$	Sherpa+Vrap
	W electron asy (*)	[14]	13 (13/8)	$0 \leq y_e \leq 2.9$	$Q = M_W$	MCFM+FEWZ
	W muon asy (*)	[13]	10 (10/9)	$0 \leq y_\mu \leq 1.9$	$Q = M_W$	MCFM+FEWZ

Table 2.2: Same as Table 2.1 for the Tevatron fixed-target Drell-Yan and W, Z and jet collider data. The total number of Tevatron data points after cuts is 345/339 for NLO/NNLO fits.

Exp.	Obs.	Ref.	$N_{\mathbf{dat}}$	Kin_1	Kin_2 (GeV)	Theory
ATLAS	W, Z 2010	[49]	30 (30/30)	$0 \leq \lvert\eta_l\rvert \leq 3.2$	$Q = M_W, M_Z$	MCFM+FEWZ
	W, Z 2011 **(*)**	[72]	34 (34/34)	$0 \leq \lvert\eta_l\rvert \leq 2.3$	$Q = M_W, M_Z$	MCFM+FEWZ
	high-mass DY 2011	[50]	11 (5/5)	$0 \leq \lvert\eta_l\rvert \leq 2.1$	$116 \leq M_{ll} \leq 1500$	MCFM+FEWZ
	low-mass DY 2011 **(*)**	[77]	6 (4/6)	$0 \leq \lvert\eta_l\rvert \leq 2.1$	$14 \leq M_{ll} \leq 56$	MCFM+FEWZ
	$[Z\ p_T$ 7 TeV $(p_T^Z, y_Z)]$ **(*)**	[78]	64 (39/39)	$0 \leq \lvert y_Z\rvert \leq 2.5$	$30 \leq p_T^Z \leq 300$	MCFM+NNLO
	$Z\ p_T$ 8 TeV (p_T^Z, M_{ll}) **(*)**	[71]	64 (44/44)	$12 \leq M_{ll} \leq 150$ GeV	$30 \leq p_T^Z \leq 900$	MCFM+NNLO
	$Z\ p_T$ 8 TeV (p_T^Z, y_Z) **(*)**	[71]	120 (48/48)	$0.0 \leq \lvert y_Z\rvert \leq 2.4$	$30 \leq p_T^Z \leq 150$	MCFM+NNLO
	7 TeV jets 2010	[57]	90 (90/90)	$0 \leq \lvert y^{\text{jet}}\rvert \leq 4.4$	$25 \leq p_T^{\text{jet}} \leq 1350$	NLOjet++
	2.76 TeV jets	[58]	59 (59/59)	$0 \leq \lvert y^{\text{jet}}\rvert \leq 4.4$	$20 \leq p_T^{\text{jet}} \leq 200$	NLOjet++
	7 TeV jets 2011 **(*)**	[76]	140 (31/31)	$0 \leq \lvert y^{\text{jet}}\rvert \leq 0.5$	$108 \leq p_T^{\text{jet}} \leq 1760$	NLOjet++
	$\sigma_{\text{tot}}(t\bar{t})$	[74, 75]	3 (3/3)	-	$Q = m_t$	top++
	$(1/\sigma_{t\bar{t}})d\sigma(t\bar{t})/y_t$ **(*)**	[73]	10 (10/10)	$0 < \lvert y_t\rvert < 2.5$	$Q = m_t$	Sherpa+NNLO
CMS	W electron asy	[52]	11 (11/11)	$0 \leq \lvert\eta_e\rvert \leq 2.4$	$Q = M_W$	MCFM+FEWZ
	W muon asy	[53]	11 (11/11)	$0 \leq \lvert\eta_\mu\rvert \leq 2.4$	$Q = M_W$	MCFM+FEWZ
	$W + c$ total	[60]	5 (5/0)	$0 \leq \lvert\eta_l\rvert \leq 2.1$	$Q = M_W$	MCFM
	$W + c$ ratio	[60]	5 (5/0)	$0 \leq \lvert\eta_l\rvert \leq 2.1$	$Q = M_W$	MCFM
	2D DY 2011 7 TeV	[54]	124 (88/110)	$0 \leq \lvert\eta_{ll}\rvert \leq 2.2$	$20 \leq M_{ll} \leq 200$	MCFM+FEWZ
	[2D DY 2012 8 TeV]	[84]	124 (108/108)	$0 \leq \lvert\eta_{ll}\rvert \leq 2.4$	$20 \leq M_{ll} \leq 1200$	MCFM+FEWZ
	$W^{\pm}$ rap 8 TeV **(*)**	[79]	22 (22/22)	$0 \leq \lvert\eta_l\rvert \leq 2.3$	$Q = M_W$	MCFM+FEWZ
	$Z\ p_T$ 8 TeV **(*)**	[83]	50 (28/28)	$0.0 \leq \lvert y_Z\rvert \leq 1.6$	$30 \leq p_T^Z \leq 170$	MCFM+NNLO
	7 TeV jets 2011	[59]	133 (133/133)	$0 \leq \lvert y^{\text{jet}}\rvert \leq 2.5$	$114 \leq p_T^{\text{jet}} \leq 2116$	NLOjet++
	2.76 TeV jets **(*)**	[80]	81 (81/81)	$0 \leq \lvert y_{\text{jet}}\rvert \leq 2.8$	$80 \leq p_T^{\text{jet}} \leq 570$	NLOjet++
	$\sigma_{\text{tot}}(t\bar{t})$	[82, 88]	3 (3/3)	-	$Q = m_t$	top++
	$(1/\sigma_{t\bar{t}})d\sigma(t\bar{t})/y_{t\bar{t}}$ **(*)**	[81]	10 (10/10)	$-2.1 < y_{t\bar{t}} < 2.1$	$Q = m_t$	Sherpa+NNLO
LHCb	Z rapidity 940 pb	[55]	9 (9/9)	$2.0 \leq \eta_l \leq 4.5$	$Q = M_Z$	MCFM+FEWZ
	$Z \to ee$ rapidity 2 fb	[56]	17 (17/17)	$2.0 \leq \eta_l \leq 4.5$	$Q = M_Z$	MCFM+FEWZ
	$W, Z \to \mu$ 7 TeV **(*)**	[85]	33 (33/29)	$2.0 \leq \eta_l \leq 4.5$	$Q = M_W, M_Z$	MCFM+FEWZ
	$W, Z \to \mu$ 8 TeV **(*)**	[86]	34 (34/30)	$2.0 \leq \eta_l \leq 4.5$	$Q = M_W, M_Z$	MCFM+FEWZ

Table 2.3: Same as Table 2.1, for ATLAS, CMS and LHCb data from the LHC Run I at $\sqrt{s} = 2.76$ TeV, $\sqrt{s} = 7$ TeV and $\sqrt{s} = 8$ TeV. The ATLAS 7 TeV $Z\ p_T$ and CMS 2D DY 2012 are in brackets because they are only included in a dedicated study but not in the default PDF set. The total number of LHC data points after cuts is 848/854 for NLO/NNLO fits (not including ATLAS 7 TeV $Z\ p_T$ and CMS 2D DY 2012).

For hadronic data, leading-order kinematics have been assumed for illustrative purposes, with central rapidity used when rapidity is integrated over and the plotted value of Q^2 set equal to the factorization scale. It is clear that the new data added in NNPDF3.1 are distributed in a wide range of scales and x, considerably extending the kinematic reach and coverage of the dataset.

In Table 2.4 we present a summary of the kinematic cuts applied to the various processes included in NNPDF3.1 at NLO and NNLO. These cuts ensure that only data where theoretical calculations are reliable are included. Specifically, we always remove from the NLO dataset points for which the NNLO corrections exceed the statistical uncertainty. The further cuts collected in Table 2.4, specific to individual datasets, will be described when discussing each dataset in turn. All computations are performed up to NNLO in QCD, not including electroweak corrections. We have checked that with the cuts described in Table 2.4, electroweak corrections never exceed experimental uncertainties.

The codes used to perform NLO computations will be discussed in each subsection below.

Dataset	NLO	NNLO
DIS structure functions	$W^2 \geq 12.5$ GeV^2 $Q^2 \geq 3.5$ GeV^2	$W^2 \geq 12.5$ GeV^2 $Q^2 \geq 3.5$ GeV^2
HERA $\sigma_c^{\rm NC}$ (in addition)	-	$Q^2 \geq 8$ GeV^2 (fitted charm)
ATLAS 7 TeV inclusive jets 2011	$\lvert y_{\rm jet}\rvert \leq 0.4$	$\lvert y_{\rm jet}\rvert \leq 0.4$
Drell-Yan E605 and E866	$\tau \leq 0.080$ $\lvert y/y_{\rm max}\rvert \leq 0.663$	$\tau \leq 0.080$ $\lvert y/y_{\rm max}\rvert \leq 0.663$
D0 $W \to l\nu$ asymmetries	-	$\lvert A_l\rvert \geq 0.03$
CMS Drell-Yan 2D 7 TeV	30 GeV $\leq M_{ll} \leq 200$ GeV $\lvert y_Z\rvert \leq 2.2$	$M_{ll} \leq 200$ GeV $\lvert y_Z\rvert \leq 2.2$
[CMS Drell-Yan 2D 8 TeV]	$M_{ll} \geq 30$ GeV	$M_{ll} \geq 30$ GeV
LHCb 7 TeV and 8 TeV $W, Z \to \mu$	-	$\lvert y_l\rvert \geq 2.25$
[ATLAS Z p_T 7 TeV]	30 GeV $\leq p_T^Z \leq 500$ GeV	30 GeV $\leq p_T^Z \leq 500$ GeV
ATLAS Z p_T 8 TeV (p_T, M_{ll})	$p_T^Z \geq 30$ GeV	$p_T^Z \geq 30$ GeV
ATLAS Z p_T 8 TeV (p_T, y_Z)	30 GeV $\leq p_T^Z \leq 150$ GeV	30 GeV $\leq p_T^Z \leq 150$ GeV
CMS Z p_T 8 TeV (p_T, y_Z)	30 GeV $\leq p_T^Z \leq 170$ GeV $\lvert y_Z\rvert \leq 1.6$	30 GeV $\leq p_T^Z \leq 170$ GeV $\lvert y_Z\rvert \leq 1.6$

Table 2.4: Full set of kinematical cuts applied to the processes used for NNPDF3.1 PDF determination at NLO and at NNLO. Only data satisfying the constraints in the table are retained. The experiments in brackets are not part of the global dataset and only used for dedicated studies. The cut on the HERA charm structure function data at NNLO is applied only when charm is fitted, and it is applied in addition to the other DIS kinematical cuts.

With the exception of deep-inelastic scattering, NNLO corrections are implemented by computing at the hadron level the bin-by-bin ratio of the NNLO to NLO prediction with a pre-defined PDF set, and applying the correction to the NLO computation (see Sect. 2.3 of Ref. [5]). For all new data included in NNPDF3.1, the PDF set used for the computation of these correction factors (often refereed to as K-factors, and in Ref. [5] as C-factors) is NNPDF3.0, except for the CMS W rap 8 TeV and ATLAS W/Z 2011 entries of Tab. 2.3 for which published `xFitter` results have been used and the CMS 2D DY 2012 data for which MMHT PDFs have been used [89] (see Sect. 2.5 below); the PDF dependence of the correction factors is much smaller than all other relevant uncertainties as we will demonstrate explicitly in Sect. 2.7 below.

2.2 Deep-inelastic structure functions

The main difference between the NNPDF 3.0 and 3.1 DIS structure function datasets is the replacement of the separate HERA-I and ZEUS/H1 HERA-II inclusive structure function measurements by the final legacy HERA combination [9]. The impact of the HERA-II data on a global fit which includes HERA-I data is known [5, 90–92] to be moderate to begin with; the further impact of replacing the separate HERA-I and HERA-II data used in NNPDF3.0 with their combination has been studied in [93] and found to be completely negligible.

Additionally, the NNPDF3.1 dataset includes the H1 and ZEUS measurements of the bottom structure function $F_2^b(x, Q^2)$ [67,68]. While the F_2^b dataset is known to have a very limited pull, the inclusion of this dataset is useful for applications, such as the determination of the bottom mass [94].

While it is not included in the default NNPDF3.1 dataset, the EMC data on charm structure functions [69] will also be used for specific studies of the charm content of the proton in Sect. 5.3. As discussed in Refs. [23,95], the EMC dataset has been corrected by updating the $\mathrm{BR}(D \to \mu)$ branching ratio: the value used in the original analysis [69] is replaced with the latest PDG value [96]. A conservative uncertainty on this branching ratio of $\pm 15\%$ is also included.

The cuts applied to DIS data are as follows. As in NNPDF3.0, for all structure function datasets we exclude data with $Q^2 < 3.5$ GeV^2 and $W^2 < 12.5$ GeV^2, i.e. the region where higher twist corrections might become relevant and the perturbative expansion may become unreliable. At NNLO we also remove F_2^c data with $Q^2 < 8$ GeV^2 in order to minimize the possible impact of unknown NNLO terms related to initial-state charm (see below).

The computation of structure functions has changed in comparison to previous NNPDF releases. Indeed, in NNPDF3.0 the solution of the DGLAP evolution equations and the structure functions were computed with the internal NNPDF code `FKgenerator` [97, 98], based on the Mellin space formalism. In NNPDF3.1, as was already the case in the charm study of Ref. [23], PDF evolution and DIS structure functions are computed using the `APFEL` public code [99], based instead on the x-space formalism. The two codes have been extensively benchmarked against each other, see App. A. DIS structure functions are computed at NLO in the FONLL-B general-mass variable flavor number scheme, and at NNLO in the FONLL-C scheme [100]. All computations include target mass corrections.

In NNPDF3.1 we now parametrize charm independently, and thus the FONLL GM-VFN has been extended in order to include initial-state heavy quarks. This is accomplished using the formalism of Refs. [21, 22]. Within this formalism, a massive correction to the charm-initiated contribution is included alongside the contribution of fitted charm as a non-vanishing boundary condition to PDF evolution. At NNLO this correction requires knowledge of massive charm-initiated contributions to the DIS coefficient functions up to $\mathcal{O}\left(\alpha_S^2\right)$, which are currently only known to $\mathcal{O}\left(\alpha_S\right)$ [101]. Therefore, in the NNLO PDF determination, the NLO expression for this correction is used: this corresponds to setting the unknown $\mathcal{O}\left(\alpha_S^2\right)$ contribution to the massive charm-initiated term to zero. Such an approximation was used Ref. [23], where it was shown that it is justified by the fact that even setting to zero the full correction (i.e. using the LO expression for the massive correction) has an effect which at the PDF level is much smaller than PDF uncertainties (see in particular Fig. 10 of Ref. [23]).

Finally, as in previous NNPDF studies, no nuclear corrections are applied to the deuteron structure function and neutrino charged-current cross-section data taken on heavy nuclei, in particular NuTeV and CHORUS. We will return to this issue in Sect. 4.11.

2.3 Fixed-target Drell-Yan production

In NNPDF3.1 we have included the same fixed-target Drell-Yan (DY) data as in NNPDF3.0, namely the Fermilab E605 and E866 datasets; in the latter case both the proton-proton data

and the ratio of cross-sections between deuteron and proton targets, $\sigma^d_{\rm DY}/\sigma^p_{\rm DY}$ are included. However, the kinematic cuts applied to these two experiments differ from those in NNPDF3.0, based on the study of [102], which showed that theoretical predictions for data points too close to the production threshold become unstable. Requiring reliability of the fixed-order perturbative approximation leads to the cuts

$$\tau \leq 0.08 \quad \text{and} \quad |y/y_{\rm max}| \leq 0.663\,, \tag{2.1}$$

where $\tau = M_{ll}^2/s$ and $y_{\rm max} = -\frac{1}{2}\ln\tau$, with M_{ll} the dilepton invariant mass distribution and $\sqrt{s}$ the center of mass energy of the collision.

As in the case of DIS, NLO fixed-target Drell-Yan cross-sections were computed in NNPDF3.0 using the Mellin-space `FKgenerator` code, while in NNPDF3.1 they are obtained using `APFEL`. The two computations are benchmarked in App. A. NNLO corrections are determined using `Vrap` [103]. Once more, as in previous NNPDF studies, no nuclear corrections are applied; again we will return to this issue in Sect. 4.11 below.

2.4 Single-inclusive jets

Four single-inclusive jet cross-section measurements were part of the NNPDF3.0 dataset: CDF Run II k_T [44], CMS 2011 [59], ATLAS 7 TeV 2010 and ATLAS 2.76 TeV, including correlations to the 7 TeV data [57, 58]. On top of these, in NNPDF3.1 we also include the ATLAS 7 TeV 2011 [76] and CMS 2.76 TeV [80] data. Some of these measurements are available for different values of the jet R parameter; the values used in NNPDF3.1 are listed in Table 2.5.

Dataset	Ref.	Jet Radius
CDF Run II k_t incl jets	[87]	$R = 0.7$
ATLAS 7 TeV jets 2010	[57]	$R = 0.4$
ATLAS 2.76 TeV jets	[58]	$R = 0.4$
ATLAS 7 TeV jets 2011	[76]	$R = 0.6$
CMS 7 TeV jets 2011	[59]	$R = 0.7$
CMS 2.76 TeV jets	[80]	$R = 0.7$

Table 2.5: Values of the jet R parameter used for the jet production datasets included in NNPDF3.1.

No cuts are applied to any of jet datasets included in NNPDF3.1, except for the ATLAS 2011 7 TeV data, for which achieving a good description turns out to be impossible if all five rapidity bins are included simultaneously. We can obtain a good agreement between data and theory when using only the central rapidity bin, $|\eta^{\rm jet}| < 0.4$. The origin of this state of affairs is not understood: we have verified that a reasonable description can be obtained if some of the systematic uncertainties are decorrelated, but we have no justification for such a procedure. We have therefore chosen to only include in NNPDF3.1 data from the central rapidity bin, $|\eta^{\rm jet}| < 0.4$ for this set. This is also the rapidity bin with the largest PDF sensitivity [104].

In NNPDF3.1, all NLO jet cross-sections are computed using `NLOjet++` [105] interfaced to `APPLgrid` [106]. The jet p_T is used as the central factorization and renormalization scale in all cases, as this choice exhibits improved perturbative convergence compared to other scale choices such as the leading jet p_T^1 [107, 108].

While the NNLO calculation of inclusive jet production has been recently published [20, 108], results are not yet available for all datasets included in NNPDF3.1. Therefore, jet data are included as default in the NNPDF3.1 NNLO determination using NNLO PDF evolution but NLO matrix elements, while adding to the covariance matrix an additional fully correlated theoretical

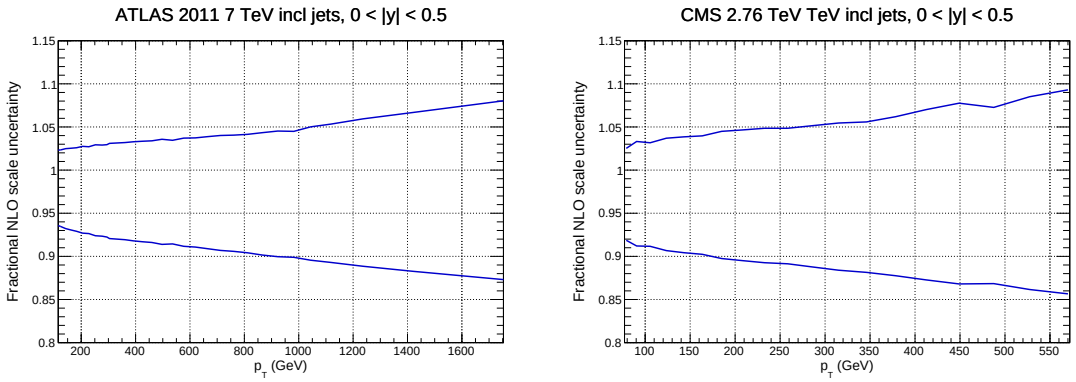

Figure 2.2: The fractional scale uncertainty on NLO single-inclusive jet production, as a function of the jet p_T for the central rapidity bins of ATLAS 7 TeV 2011 (left) and the CMS 2.76 TeV (right).

systematic uncertainty estimated from scale variation of the NLO calculation. The NLO scale variations are performed using `APPLgrid` interfaced to `HOPPET` [109]. We take the associated uncertainty as the the envelope of the result of seven-point scale variation $\mu_F \in [p_T/2, 2p_T]$ and $\mu_R \in [p_T/2, 2p_T]$ with $1/2 \leq \mu_F/\mu_R \leq 2$.The NNLO corrections are generally well within this scale variation band when the jet p_T is chosen as a central scale [108]. This scale uncertainty is shown in Fig. 2.2 for ATLAS 7 TeV 2011 and CMS 2.76 TeV as a function of the jet p_T for the central rapidity bin. It is seen to range between a few percent at low p_T up to around 10% at the largest p_T. A similar behaviour is observed in other rapidity bins, with a more asymmetric band at forward rapidity.

In order to gauge the reliability of our approximate treatment of the jet data, we have produced a PDF determination in which all data for which NNLO corrections are known, namely the 7 TeV ATLAS and CMS datasets, are included using exact NNLO theory. This will be discussed in Sect. 4.4. Representative NNLO corrections are shown in Fig. 2.3, where we show the NNLO/NLO ratio for the central rapidity bin ($0 \leq |y_{\rm jet}| \leq 0.5$) of the ATLAS and CMS 7 TeV 2011 datasets, plotted as a function of p_T [110]: note (see Table 2.5) that the values of R are different, thereby explaining the different size of the correction, which for CMS is $\sim -2\%$ for $p_T \sim 100$ GeV, increasing up to $\sim 5\%$ for $p_T \sim 2$ TeV, and for ATLAS it ranges from $\sim -4\%$ increasing up to $\sim 9\%$ as a function of p_T. Unlike in the case of the Z transverse momentum distribution, to be discussed in Sect. 2.6, the lack of smoothness of the corrections seen in Fig. 2.3 is not problematic as the fluctuations are rather smaller than typical uncorrelated uncertainties on these data.

2.5 Drell-Yan production at hadron colliders

The NNPDF3.0 determination already included a wide set of collider Drell-Yan data, both at the W and Z peak and off-shell. This dataset has been further expanded in NNPDF3.1. We discuss here invariant mass and rapidity distributions; transverse momentum distributions will be discussed in Sect. 2.6.

In NNPDF3.1 we include for the first time D0 legacy W asymmetry measurements based on the complete dataset in the electron [14] and muon [13] channels. The only cut applied to this dataset is at NNLO, where we remove data with $\mathcal{A}_l(y_l) \leq 0.03$ in both the electron and muon channel data. This is due to the fact that when the asymmetry is very close to zero, even with high absolute accuracy on the NNLO theoretical calculation, it is difficult to achieve high percentage accuracy, thereby making the NNLO correction to the asymmetry unreliable. The NLO computation is performed using `APPLgrids` from the HERAfitter study of [70], which we have cross-checked using `Sherpa` [111] interfaced to `MCgrid` [112]. NNLO corrections are

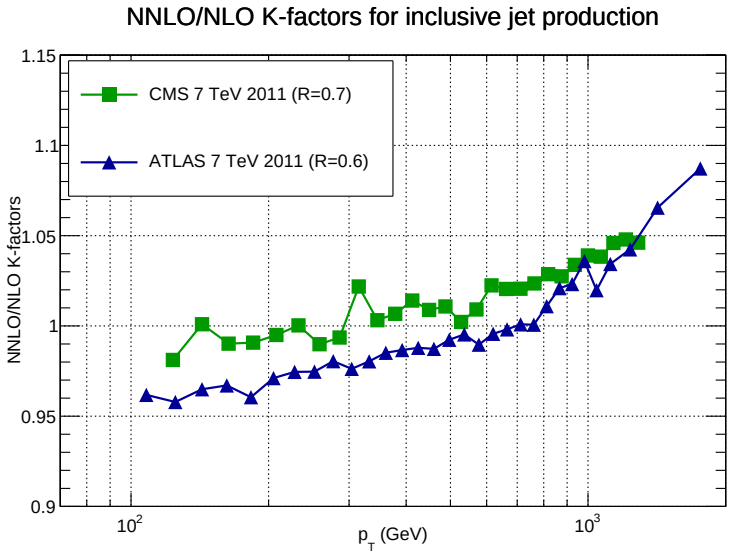

Figure 2.3: The NNLO/NLO cross-section ratio [110] for the central rapidity bin ($0 \leq |y_{\rm jet}| \leq 0.5$) of the ATLAS and CMS 7 TeV 2011 jet data, with the values of R of Tab. 2.5, plotted vs. p_T.

computed using `FEWZ` [113–115].

New results are included for ATLAS, CMS and LHCb. For ATLAS, NNPDF3.0 included 2010 W and Z 7 TeV rapidity distributions and their cross-correlations [49]. A recent update of the same measurement [72], based on the entire 7 TeV integrated luminosity of 4.6 fb^{-1} is included in NNPDF3.1, albeit partially. This measurement provides differential distributions in lepton pseudo-rapidity $|\eta_l|$ in the range $0 \leq |\eta_l| \leq 2.5$ for on-shell W^+ and W^- production. For Z/γ^* production results are provided either with both leptons measured in the range $0 \leq |\eta_l| \leq 2.5$, or with one lepton with $0 \leq |\eta_l| \leq 2.5$ and the other with $2.5 \leq |\eta_l| \leq 4.9$. The central rapidity data sre given for three bins in the dilepton invariant mass $46 < m_{ll} < 66$, $66 < m_{ll} < 116$ and $116 < m_{ll} < 150$ GeV, and the forward rapidity data in the last two mass bins (on-peak and high-mass). We only include the on-shell, $0 \leq |\eta_l| \leq 2.5$ data, thereby neglecting the two low- and high-mass Z production bins in the central rapidity region, and the on-peak and high-mass Z production bins at forward rapidity. The full dataset will be included in future NNPDF releases. No other cuts are applied to the dataset. Theoretical predictions are obtained using NLO `APPLgrids` [106] generated using `MCFM` [116], while the NNLO corrections are taken from the `xFitter` analysis of Ref [72].

Also new to NNPDF3.1 is the ATLAS low-mass Drell-Yan data from Ref. [77]. We use only the low-mass DY cross-sections in the muon channel measured from 35 pb^{-1} 2010 dataset, which extends down to $M_{ll} = 12$ GeV. The 2011 7 TeV data with invariant masses between 26 GeV and 66 GeV are not included because they are affected by large electroweak corrections and are therefore excluded by our cuts. Furthermore, two datapoints are removed from the NLO datasets because NNLO corrections exceed experimental uncertainties. Theoretical predictions are obtained at NLO using `APPLgrids` [106] constructed using `MCFM`, and at NNLO corrections are computed using `FEWZ`.

For CMS, NNPDF3.1 includes 8 TeV W^+ and W^- rapidity distributions, including information on their correlation [79]. No cuts have been applied to this dataset. Theoretical predictions are obtained using the NLO `APPLgrids` generated with `MCFM` and the NNLO correction factors computed using `FEWZ` in the context of the `xFitter` [117] analysis presented in Ref. [79]. Double differential rapidity y_{ll} and invariant mass M_{ll} distributions for Z/γ^* production from the

	NNPDF3.1			NNPDF3.0	
Dataset	NNLO	NLO	LO	NNLO	NLO
NMC	1.30	1.35	3.25	1.29	1.36
SLAC	0.75	1.17	3.35	0.66	1.08
BCDMS	1.21	1.17	2.20	1.31	1.21
CHORUS	1.11	1.06	1.16	1.11	1.14
NuTeV dimuon	0.82	0.87	4.75	0.69	0.61
HERA I+II inclusive	1.16	1.14	1.77	1.25	1.20
HERA $\sigma_c^{\rm NC}$	1.45	1.15 ()	1.21	[1.61]	[2.57]
HERA F_2^b	1.11	1.08	11.2	[1.13]	[1.12]
DY E866 $\sigma^d_{\rm DY}/\sigma^p_{\rm DY}$	0.41	0.40	1.06	0.47	0.53
DY E886 σ^p	1.43	1.05	0.81	1.69	1.17
DY E605 σ^p	1.21	0.97	0.66	1.09	0.87
CDF Z rap	1.48	1.619	1.54	1.55	1.28
CDF Run II k_t jets	0.87	0.84	1.07	0.82	0.95
D0 Z rap	0.60	0.67	0.65	0.61	0.59
D0 $W \to e\nu$ asy	2.70	1.59	1.75	[2.68]	[4.58]
D0 $W \to \mu\nu$ asy	1.56	1.52	2.16	[2.02]	[1.43]
ATLAS total	**1.09**	**1.36**	**5.34**	**1.92**	**1.98**
ATLAS W,Z 7 TeV 2010	0.96	1.04	2.38	1.42	1.39
ATLAS high-mass DY 7 TeV	1.54	1.88	4.05	1.60	2.17
ATLAS low-mass DY 2011	0.90	0.69	2.86	[0.94]	[0.81]
ATLAS W,Z 7 TeV 2011	2.14	3.70	27.2	[8.44]	[7.6]
ATLAS jets 2010 7 TeV	0.94	0.92	1.22	1.12	1.07
ATLAS jets 2.76 TeV	1.03	1.03	1.50	1.31	1.32
ATLAS jets 2011 7 TeV	1.07	1.12	1.59	[1.03]	[1.12]
ATLAS Z p_T 8 TeV (p_T^{ll}, M_{ll})	0.93	1.17	-	[1.05]	[1.28]
ATLAS Z p_T 8 TeV (p_T^{ll}, y_{ll})	0.94	1.77	-	[1.19]	[2.49]
ATLAS $\sigma_{tt}^{\rm tot}$	0.86	1.92	53.2	0.67	1.07
ATLAS $t\bar{t}$ rap	1.45	1.31	1.99	[3.32]	[1.50]
CMS total	**1.06**	**1.20**	**2.13**	**1.19**	**1.33**
CMS W asy 840 pb	0.78	0.86	1.55	0.73	0.85
CMS W asy 4.7 fb	1.75	1.77	3.16	1.75	1.82
CMS $W+c$ tot	-	0.54	16.5	-	0.93
CMS $W+c$ ratio	-	1.91	3.21	-	2.09
CMS Drell-Yan 2D 2011	1.27	1.23	2.15	1.20	1.19
CMS W rap 8 TeV	1.01	0.70	4.32	[1.24]	[0.96]
CMS jets 7 TeV 2011	0.84	0.84	0.93	1.06	0.98
CMS jets 2.76 TeV	1.03	1.01	1.09	[1.22]	[1.18]
CMS Z p_T 8 TeV (p_T^{ll}, M_{ll})	1.32	3.65	-	[1.59]	[3.86]
CMS $\sigma_{tt}^{\rm tot}$	0.20	0.59	53.4	0.56	0.10
CMS $t\bar{t}$ rap	0.94	0.96	1.32	[1.15]	[1.01]
LHCb total	**1.47**	**1.62**	**5.16**	**2.11**	**2.67**
LHCb Z 940 pb	1.49	1.27	2.51	1.29	0.91
LHCb $Z \to ee$ 2 fb	1.14	1.33	6.34	1.21	2.31
LHCb $W,Z \to \mu$ 7 TeV	1.76	1.60	4.70	[2.59]	[2.36]
LHCb $W,Z \to \mu$ 8 TeV	1.37	1.88	7.41	[2.40]	[3.74]
Total dataset	**1.148**	**1.168**	**2.238**	**1.284**	**1.307**

Table 3.1: The values of $\chi^2/N_{\rm dat}$ for the global fit and for all the datasets included in the NNPDF3.1 LO, NLO and NNLO PDF determinations. Values obtained using the NNPDF3.0 NLO and NNLO PDFs are also shown: numbers in brackets correspond to data not fitted in NNPDF3.0. Note that NNPDF3.0 values are produced using NNPDF3.1 theory settings, and are thus somewhat worse than those quoted in Ref. [5].

	NNPDF3.1 pert. charm		NNPDF3.1	
Dataset	NNLO	NLO	NNLO	NLO
NMC	1.38	1.38	1.30	1.35
SLAC	0.70	1.22	0.75	1.17
BCDMS	1.27	1.24	1.21	1.17
CHORUS	1.10	1.07	1.11	1.06
NuTeV dimuon	1.27	1.01	0.82	0.87
HERA I+II inclusive	1.21	1.15	1.16	1.14
HERA $\sigma_c^{\rm NC}$	1.20 (1.42)	1.21 (1.35)	1.45	1.15 (1.35)
HERA F_2^b	1.16	1.12	1.11	1.08
DYE866 $\sigma^d_{\rm DY}/\sigma^p_{\rm DY}$	0.46	0.48	0.41	0.40
DYE886 σ^p	1.38	1.09	1.43	1.05
DYE605 σ^p	1.05	0.83	1.21	0.97
CDF Z rap	1.44	1.46	1.48	1.62
CDF Run II k_t jets	0.86	0.86	0.87	0.84
D0 Z rap	0.60	0.64	0.60	0.67
D0 $W \to e\nu$ asy	2.71	1.63	2.70	1.59
D0 $W \to \mu\nu$ asy	1.42	1.38	1.56	1.52
ATLAS total	**1.17**	**1.45**	**1.09**	**1.3**
ATLAS W,Z 7 TeV 2010	1.04	1.08	0.96	1.04
ATLAS high-mass DY 7 TeV	1.66	2.08	1.54	1.88
ATLAS low-mass DY 2011	0.83	0.70	0.90	0.69
ATLAS W,Z 7 TeV 2011	2.74	4.29	2.14	3.70
ATLAS jets 2010 7 TeV	0.96	0.95	0.94	0.92
ATLAS jets 2.76 TeV	1.06	1.13	1.03	1.03
ATLAS jets 2011 7 TeV	1.11	1.14	1.07	1.12
ATLAS Z p_T 8 TeV (p_T^{ll}, M_{ll})	0.94	1.19	0.93	1.17
ATLAS Z p_T 8 TeV (p_T^{ll}, y_{ll})	0.96	1.84	0.94	1.77
ATLAS $\sigma_{tt}^{\rm tot}$	0.80	2.03	0.86	1.92
ATLAS $t\bar{t}$ rap	1.39	1.18	1.45	1.31
CMS total	**1.09**	**1.2**	**1.06**	**1.20**
CMS W asy 840 pb	0.69	0.80	0.78	0.86
CMS W asy 4.7 fb	1.75	1.76	1.75	1.77
CMS $W+c$ tot	-	0.49	-	0.54
CMS $W+c$ ratio	-	1.92	-	1.91
CMS Drell-Yan 2D 2011	1.33	1.27	1.27	1.23
CMS W rap 8 TeV	0.90	0.65	1.01	0.70
CMS jets 7 TeV 2011	0.87	0.86	0.84	0.84
CMS jets 2.76 TeV	1.06	1.05	1.03	1.01
CMS Z p_T 8 TeV (p_T^{ll}, y_{ll})	1.29	3.50	1.32	3.65
CMS $\sigma_{tt}^{\rm tot}$	0.21	0.67	0.20	0.59
CMS $t\bar{t}$ rap	0.96	0.96	0.94	0.96
LHCb total	**1.48**	**1.77**	**1.47**	**1.62**
LHCb Z 940 pb	1.31	1.08	1.49	1.27
LHCb $Z \to ee$ 2 fb	1.47	1.66	1.14	1.33
LHCb $W,Z \to \mu$ 7 TeV	1.54	1.51	1.76	1.60
LHCb $W,Z \to \mu$ 8 TeV	1.51	2.28	1.37	1.88
Total dataset	**1.187**	**1.197**	**1.148**	**1.168**

Table 3.2: Same as Tab. 3.1, but now comparing the default NNPDF3.1 NNLO and NNLO sets to the variant in which charm is perturbatively generated. For HERA $\sigma_c^{\rm NC}$ the number in parenthesis refer to the subset of data to which the NNLO FC cut of Table. 2.4 is applied.

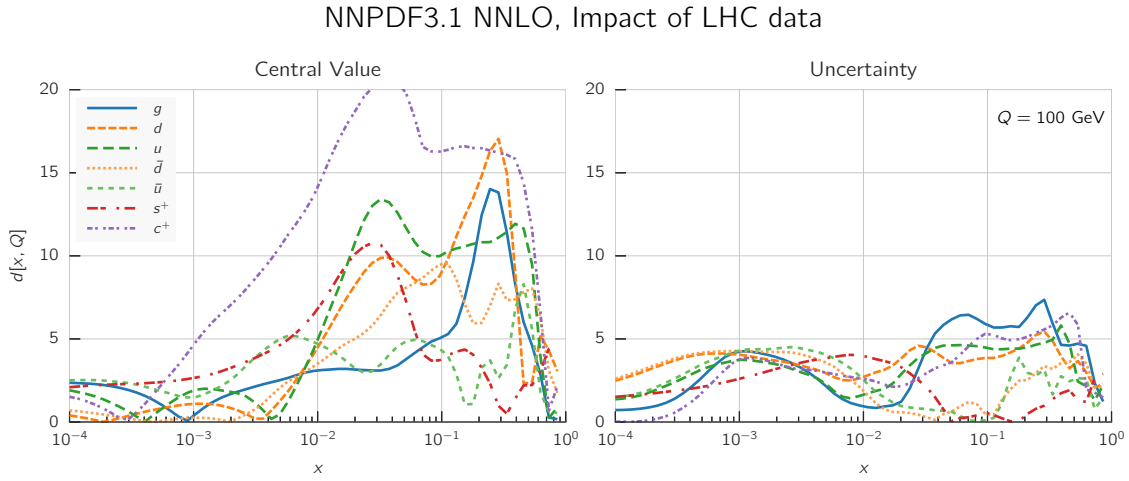

Figure 4.26: Same as Fig. 4.3 but now excluding all LHC data.

level of deterioration is not so great as to make searches for new physics altogether impossible.

4.11 Nuclear targets and nuclear corrections.

The NNPDF3.1 dataset includes several measurements taken upon nuclear targets. DIS data from the SLAC, BCDMS and NMC experiments along with the E886 fixed-target Drell-Yan data involve measurements of deuterium. All neutrino data and the fixed-target E605 Drell-Yan data, are obtained with heavy nuclear targets. All of these data were already included in previous PDF determinations, including NNPDF3.0. The impact of nuclear corrections was studied in Ref. [5] and found to be under control. However, the much wider dataset might now permit the removal of these data from the global dataset: whereas removing data inevitably entails some loss of precision, this might be more than compensated by the increase in accuracy due to the complete elimination of any dependence on uncertain nuclear corrections.

In order to assess this, we performed two additional PDF determinations with the NNPDF3.1 methodology. Firstly, by removing all heavy nuclear target data but keeping deuterium data, and secondly removing all nuclear data and only keeping proton data. The distances between the default and these two PDF sets are shown in Fig. 4.28. At large x the impact of nuclear target data is significant, at the one to two sigma level, mostly on the flavor separation of the sea. The deuterium data also have a significant impact, particularly in the intermediate x range.

A direct comparison of PDFs, in Fig. 4.29, and their uncertainties, in Fig. 4.32, shows that indeed PDFs determined with no heavy nuclear target data are reasonably compatible with the global set, though with rather larger uncertainties, especially for strangeness. Indeed, best-fit results without heavy nuclear targets, or even without deuterium data, are all compatible within their respective uncertainties, which is consistent with the previous conclusion that the absence of nuclear corrections for these data does not lead to significant bias at the level of current PDF uncertainties. On the other hand, PDFs determined with only proton data while compatible to within one sigma with the global set within their larger uncertainties, show a substantial loss of precision. This is particularly notable for down quarks, due to the importance of deuterium data in pinning down the isospin triplet PDF combinations.

Because deuterium data have a significant impact on the fit, one may worry that nuclear corrections to the deuterium data are now no longer negligible, at the accuracy of the present PDF determination. In order to investigate this issue in greater detail, we have performed a

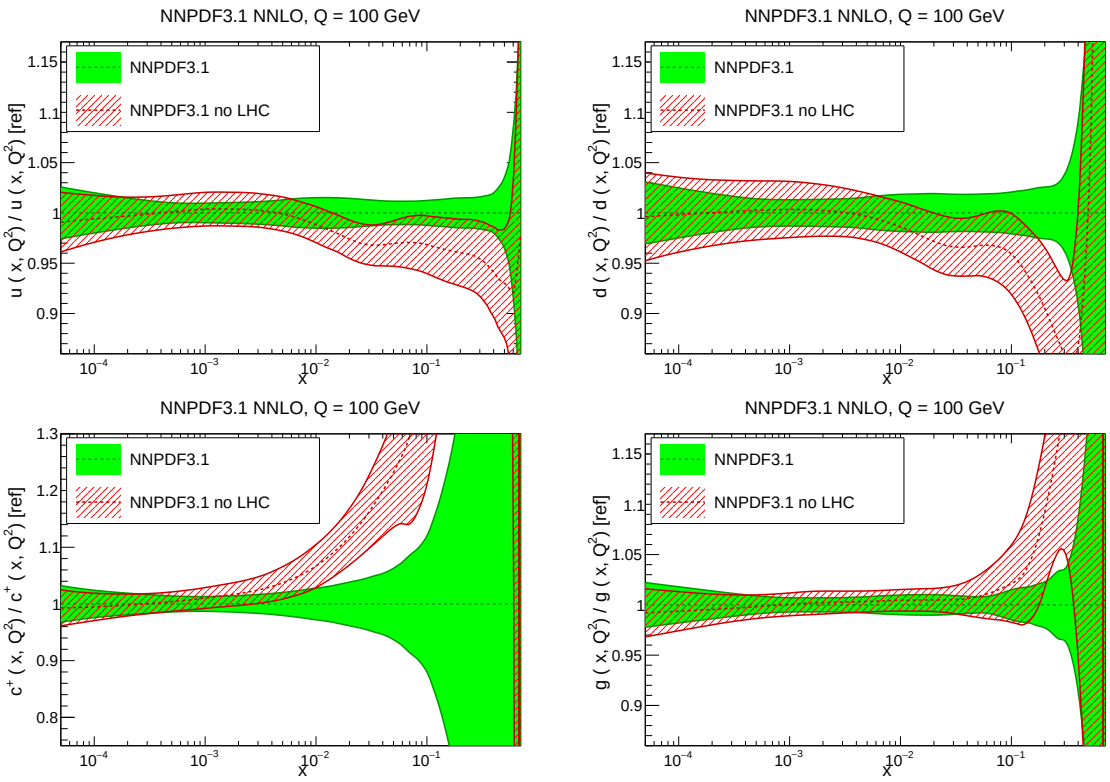

Figure 4.27: Same as Fig. 4.4 but now excluding all LHC data. Results are shown for the up (top left), down (top right), charm (bottom left) and gluon (bottom right) PDFs.

variant of the NNPDF3.1 NNLO default PDF determination in which all deuterium data are corrected using the same nuclear corrections as used by MMHT14 (specifically, Eqs. (9,10) of Ref. [7]).

In terms of fit quality we find that the inclusion of nuclear corrections leads to a slight deterioration in the quality of the fit, with a value of $\chi^2/N_{\rm dat} = 1.156$, to be compared to the defaut $\chi^2/N_{\rm dat} = 1.148$ (see Table 3.1). In particular we find that for the NMC, SLAC, and BCDMS data the values of $\chi^2/N_{\rm dat}$ with (without) nuclear corrections are respectively 0.94(0.95), 0.71(0.70), and 1.11(1.11). Therefore, the addition of deuterium corrections has no significant impact on the fit quality to these data.

The distances between PDFs determined including deuterium corrections and the default are shown in Fig. 4.30. They are seen to be moderate and always below the half-sigma level, and confined mostly to the up and down PDFs, as expected. These PDFs are shown in Fig. 4.31, which confirms the moderate effect of the deuterium correction. It should be noticed that the PDF uncertainty, also shown in Fig. 4.31, is somewhat increased when the deuterium corrections are included. The relative shift for other PDFs are yet smaller since they are affected by larger uncertainties, which are also somewhat increased by the inclusion of the nuclear corrections.

In view of the theoretical uncertainty involved in estimating nuclear corrections, and bearing in mind that we see no evidence of an improvement in fit quality while we note a slight increase in PDF uncertainties when including deuterium corrections using the model of Ref. [7], we conclude that the impact of deuterium corrections on the NNPDF3.1 results is sufficiently small that they may be safely ignored even within the current high precision of PDF determination. Nevertheless, more detailed dedicated studies of nuclear corrections, also in relation to the construction of nuclear PDF sets, may well be worth pursuing in future studies.

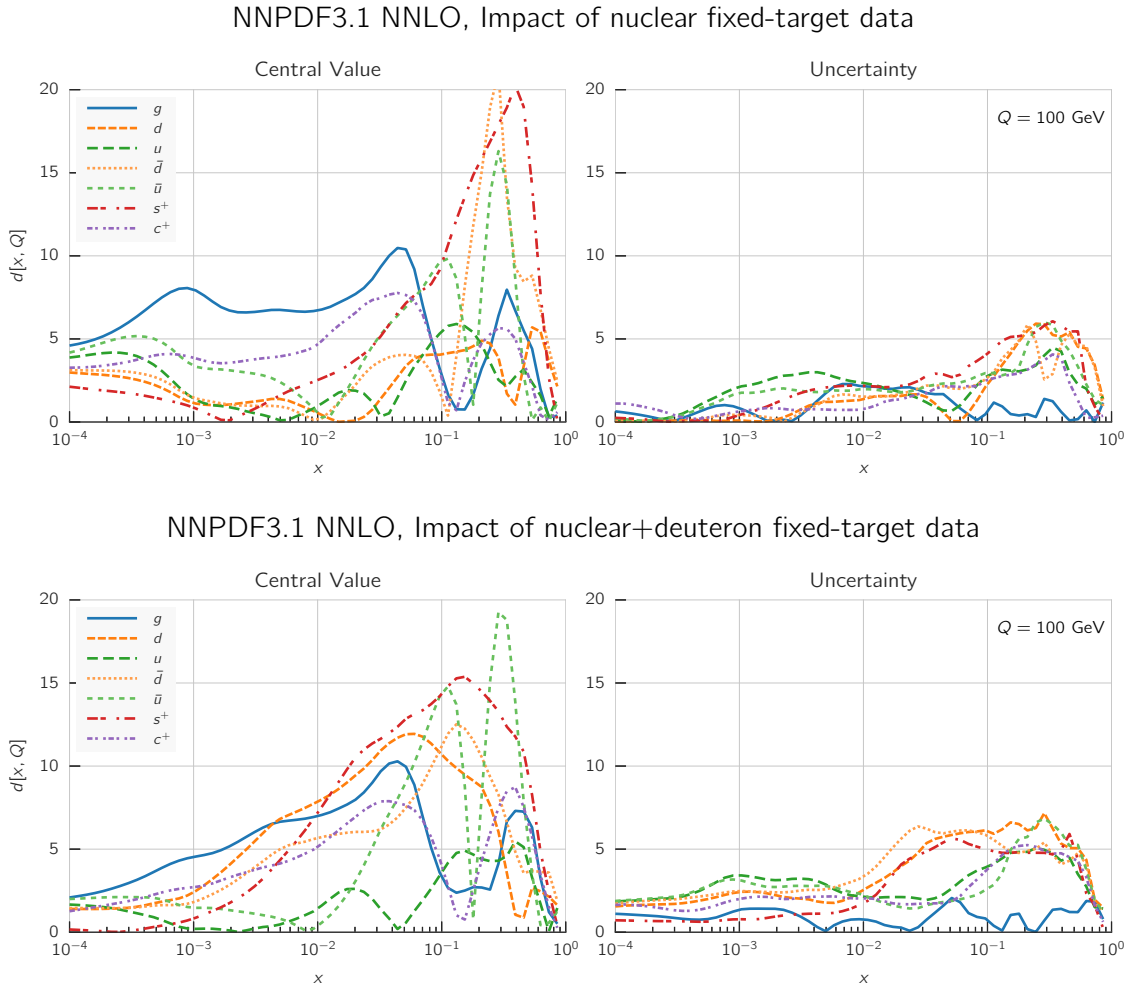

Figure 4.28: Same as Fig. 4.3 but now excluding all data with heavy nuclear targets, but keeping deuterium data (top) or excluding all data with any nuclear target and only keeping proton data (bottom)

In conclusion, for the time being it is still appears advantageous to retain nuclear target data in the global dataset for general-purpose PDF determination. However, if very high accuracy is required (such as, for instance, in the determination of standard model parameters) it might be preferable to use PDF sets from which all data with nuclear targets have been omitted.

4.12 Collider-only parton distributions

A yet more conservative option to that discussed in the previous Section is to retain only collider data from HERA, the Tevatron and the LHC. The motivation for this suggestion, first presented in the NNPDF2.3 study [133], is that this excludes data taken at low scales, which may be subject to potentially large perturbative and non-perturbative corrections. Furthermore, data taken on nuclear targets, and all of the older datasets are eliminated, thereby leading to a more reliable set of PDFs. However, previous collider-only PDFs had very large uncertainties, due to the limited collider dataset then available.

In order to re-assess the situation with the current, much wider LHC dataset, we have repeated a collider-only PDF determination. This amounts to repeating the proton only PDF

and the corresponding ratio of momentum fractions

$$K_s(Q^2) = \frac{\int_0^1 dx\, x\left(s(x,Q^2)+\bar{s}(x,Q^2)\right)}{\int_0^1 dx\, x\left(\bar{u}(x,Q^2)+\bar{d}(x,Q^2)\right)}\,, \tag{5.2}$$

have been traditionally assumed to be significantly smaller than one, and in PDF sets produced before the strange PDF could be extracted from the data, such as e.g. NNPDF1.0 [97], it was often assumed that $R_s \sim \frac{1}{2}$, for all x, and thus also $K_s \sim \frac{1}{2}$. This level of strangeness suppression is indeed found in many recent global PDF sets, in which the strongest handle on the strange PDF is provided by deep-inelastic neutrino inclusive F_2 and charm F_2^c ("dimuon") data.

This was challenged in Ref. [135] where, on the basis of ATLAS W and Z production data, combined with HERA DIS data, it was argued instead that, in the measured region, the strange fraction R_s is of order one. In Refs. [3,133], respectively based on the NNPDF2.3, and NNPDF3.0 global analyses, both of which includes the data of Ref. [135], it was concluded that whereas the ATLAS data do favor a larger strange PDF, they have a moderate impact on the global PDF determination due to large uncertainties, and also, that if the strange PDF is only determined from HERA and ATLAS data, the central value is consistent with the conclusion of Ref. [135], but the uncertainty is large enough to lead to agreement with the suppressed strangeness of the global PDF sets to within one sigma. In Ref. [3] it was also shown that the CMS $W+c$ production data [60], which were included there for the first time and which are also included in NNPDF3.1, though only in the NLO determination because of lack of knowledge of the NNLO corrections, have a negligible impact due to their large uncertainties.

As we discussed in Section 4.7, ATLAS W and Z production data have been supplemented by the rather more accurate dataset of Ref. [72], also claimed to favor enhanced strangeness. Indeed, we have seen in Section 4.7 that strangeness is significantly enhanced by the inclusion of these data, and also, in Section 3.4, that this enhancement can be accommodated in the global PDF determination thanks to the independently parametrized charm PDF, which is a new feature to NNPDF3.1. It is thus interesting to re-asses strangeness in NNPDF3.1, by comparing theoretically motivated choices of dataset: we will thus compare to the previous NNPDF3.0 results for strangeness obtained using the default NNPDF3.1, the collider-only PDF set of Section 4.12, which can be considered to be theoretically more reliable, and a PDF set which we have constructed by using NNPDF3.1 methodology, but only including all HERA inclusive structure function data from Tab. 2.1 and the ATLAS data of Ref. [72]. Because inclusive DIS data alone cannot determine separately strangeness [1] this is then a determination of strangeness which fully relies on the ATLAS data.

In Table 5.1 we show NNLO results, obtained using these different PDF sets, for $R_s(x,Q)$ Eq. 5.1 at $Q = 1.38$ GeV (thus below charm threshold) and $Q = m_Z$ and $x = 0.023$, an x value chosen by ATLAS in order to maximize sensitivity. Results are also compared to that of Ref. [72]. A graphical representation of the table is in Fig. 5.2.

First, comparison of the NNPDF3.1 HERA+ATLAS W,Z result with that of Ref. [72], based on the same data, shows agreement at the one-sigma level, with a similar central value and a greatly increased uncertainty, about four times larger, most likely because of the more flexible parametrization and because of independently parametrizing charm. Second, strangeness in NNPDF3.1 is rather larger than in NNPDF3.0: as we have shown in Sects. 3.4,4.7 this is largely due to the effect of the ATLAS W,Z 2011 data, combined with determining charm from the data: indeed, it is clear from Fig. 4.2 that the new data and new methodology both lead to strange enhancement, with the former effect dominant but the latter not negligilbe. This enhancement is more marked in the collider-only PDF set, which leads to a value which is very close to that coming from the ATLAS data. This suggests some tension between strangeness preferred by collider data and the rest of the dataset, i.e., most likely, neutrino data.

PDF set	$R_s(0.023, 1.38\ \text{GeV})$	$R_s(0.023, M_Z)$
NNPDF3.0	0.45±0.09	0.71±0.04
NNPDF3.1	0.59±0.12	0.77±0.05
NNPDF3.1 collider-only	0.82±0.18	0.92±0.09
NNPDF3.1 HERA + ATLAS W, Z	1.03±0.38	1.05±0.240
`xFitter` HERA + ATLAS W, Z (Ref. [72])	$1.13\,^{+0.11}_{-0.11}$	-

Table 5.1: The strangeness fraction $R_s(x, Q)$ Eq. (5.1) at $x = 0.023$, at a low scale and a high scale. We show results obtained using NNPDF3.0, and NNPDF3.1 baseline, collider-only and HERA+ATLAS W, Z sets, compared to the `xFitter` ATLAS value Ref. [72].

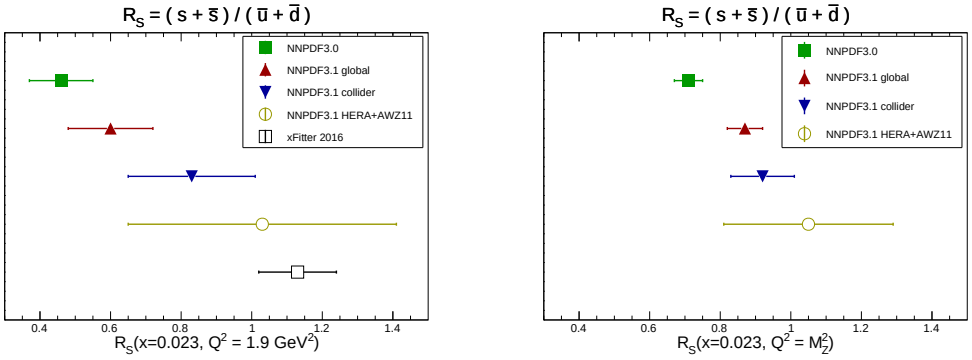

Figure 5.2: Graphical representation of the results of Table 5.1.

It is interesting to repeat this analysis for the full x range. This is done in Fig. 5.3, where $R_s(x, Q)$ Eq. (5.1) is plotted as a function of x again at low and high scales, now only including NNPDF3.0, and the default and collider-only versions of NNPDF3.1. It is clear that in the collider-only PDF set strangeness is largely unconstrained at large x, whereas the global fit is constrained by neutrino data to have a suppressed value $R_s \sim 0.5$. At lower x we see the tension between this and the constraint from the collider data, which prefer a larger value.

In Fig. 5.3 we also compare the strangeness ratio $R_s(x, Q)$ of NNPDF3.1 with that of CT14 and MMHT14. We find that there is good consistency in the entire range of x, while the PDF errors in NNPDF3.1 are typically smaller than those of the other two sets, especially at large scales. It is also interesting to note how in NNPDF3.1 the PDF uncertainties in the ratio R_s blow up at very large x, reflecting the lack of direct information on strangeness in that kinematic region.

We now turn to the strange momentum fraction $K_s(Q^2)$ Eq. (5.2); values for the same PDF sets and scales are shown in Table 5.2. Results are quite similar to those found from the analysis of Table 5.1. For the NNPDF3.1 collider-only and especially the HERA + ATLAS W, Z fits, the central value of K_s is unphysical, with a huge uncertainty; essentially, all one can say is that the strange momentum fraction K_s is completely uncertain. This shows rather dramatically that the relatively precise values in Table 5.1 only hold in a rather narrow x range. It will be interesting to see whether more LHC data, possibly leading to a competitive collider-only fit, will confirm strangeness enhancement and allow for an accurate determination of strangeness in a wider range of x.