Uncertainties due to Nuclear Data in Proton PDF Fits

Rosalyn L. Pearson[*]

The Higgs Centre for Theoretical Physics,
University of Edinburgh, JCMB, KB, Mayfield Rd, Edinburgh EH9 3FD, Scotland.
E-mail: r.l.pearson@ed.ac.uk

Richard D. Ball

The Higgs Centre for Theoretical Physics,
University of Edinburgh, JCMB, KB, Mayfield Rd, Edinburgh EH9 3FD, Scotland.
E-mail: rdb@ph.ed.ac.uk

Emanuele R. Nocera

Nikhef Theory Group,
Science Park 105, 1098 XG Amsterdam, The Netherlands.
E-mail: enocera@nikhef.nl

We investigate the role of uncertainties due to the use of nuclear data when determining proton parton distribution functions (PDFs), and show how these uncertainties can be included in fits. In particular, we assess data with nuclear targets included in NNPDF3.1, namely from CHORUS, NuTeV and E605 with Pb, Fe and Cu targets respectively. We then analyse the resulting form of the PDFs, noting that the increase in uncertainty is small. Finally, we consider the impact on the $\bar{d}/\bar{u}$ ratio, total strangeness $s + \bar{s}$ and strange valence distribution $s - \bar{s}$. We find the effects to be negligible in all cases.

---

[*]Speaker.

Parton distribution functions (PDFs) are universal quantities encapsulating the internal structure of the proton, and are crucial for making predictions in particle physics [1]. To maximally constrain them, PDFs are determined by fitting a range of experimental data over a wide variety of processes and kinematic regimes. Some of this data consists of measurements on nuclear targets, rather than proton targets. In this case, the surrounding nuclear environment will have an effect on the measured observables, which in turn will influence the form of the fitted PDFs. The uncertainties associated with these effects are termed "nuclear uncertainties". Such uncertainties are small [2] [3] but becoming increasingly relevant with the advent of the Large Hadron Collider and the era of precision physics it has ushered in [1]. In these proceedings, we show how to use existing nuclear PDFs (nPDFs) to provide an estimate of nuclear uncertainties, and include them in future proton PDF fits within the Neural Network PDF (NNPDF) framework [4]. For a more detailed analysis, see [5].

There are three experiments with nuclear targets currently included in NNPDF analyses: charged current inclusive deep inelastic scattering (DIS) cross sections from CHORUS [6], on Pb; DIS dimuon cross sections from NuTeV [7] [8] on Fe; and Drell-Yan dimuon cross sections from E605 at Fermilab [9], on Cu. After cuts, nuclear data make up 993/4285 of the data points ($\sim 23\%$). For a complete summary of the data sets, see [10].

A study of the correlation between these measurements and the fitted PDFs reveals that the CHORUS data has most impact on the up- and down-valence distributions, NuTeV data has most impact on the strange quarks, and E605 data has most impact on the other light sea quarks: anti-up and anti-down. Therefore, we anticipate largest effects from nuclear uncertainties in these PDFs.

In a PDF fit we include an experimental covariance matrix, C_{ij} (where i, j run over the data points), describing the breakdown of statistical and systematic errors. Uncertainties due to nuclear data must be considered in addition to the experimental uncertainties, and in general they can be encapsulated in a theoretical covariance matrix, S_{ij}. In a PDF fit we simply add this to C_{ij} [11], so that the nuclear uncertainties act like additional experimental systematics.

We adopted an empirical approach to construct the nuclear uncertainties, using nPDFs rather than appealing to nuclear models, which rely on various assumptions [12]. We compared theoretical predictions for nuclear observables made with the correct corresponding nPDFs for an isotope "N", $T_i^N[f_N^{(n)}]$, to those with proton PDFs, $T_i^N[f_p]$. Here f_p is the central value for a proton PDF and $f_N^{(n)}$ is one Monte Carlo replica in an nPDF ensemble with central value f_N, where $n = 1, ..., N_{rep}$. To generate such an ensemble we combined a Monte Carlo version of three recent nPDF sets: DSSZ12 [13], nCTEQ15 [14] and EPPS16 [15]. Note that DSSZ12 does not provide a Cu PDF, so for the case of E605 we combined just two nPDF sets.

We considered two definitions of nuclear uncertainties:

1. **Def. 1,** (a conservative approach) where the only modification is to include nuclear uncertainties by considering the difference between nuclear and proton predictions,
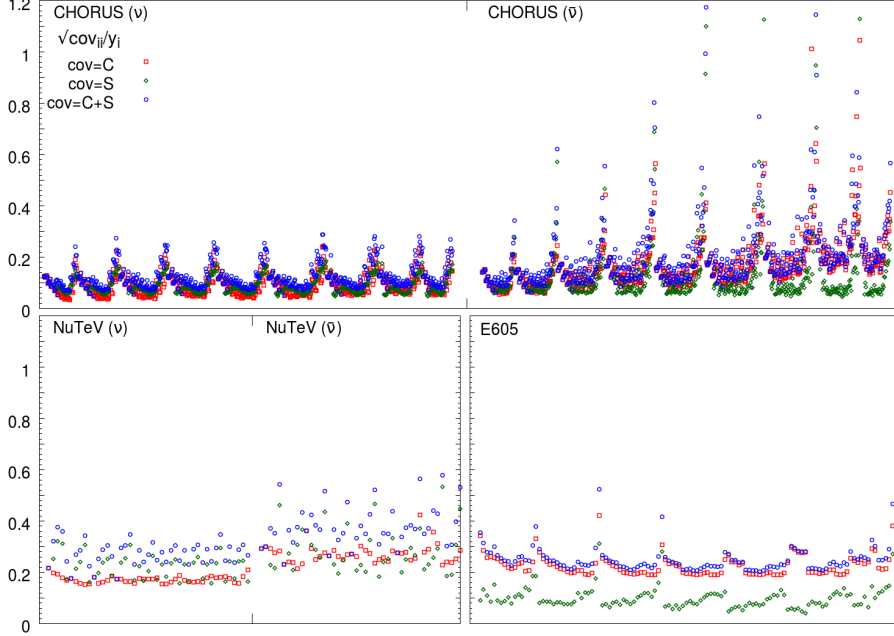
$$\Delta_i^{(n)} = T_i^N[f_N^{(n)}] - T_i^N[f_p];$$ (1)

2. **Def. 2,** (a more ambitious approach) where a shift,

$$\delta T_i^N = T_i^N[f_N] - T_i^N[f_p],$$ (2)

is also applied to the corresponding observable, meaning that the uncertainty should be defined relative to the shifted value,

$$\Delta_i^{(n)} = T_i^N[f_N^{(n)}] - T_i^N[f_N].$$

(3)

Figure 1: The square root of the diagonal elements of the covariance matrices, normalised to corresponding data. Experimental contributions are red, theory green and the total blue. Data from CHORUS and NuTeV are split into neutrino and anti-neutrino parts. Points are binned in (anti-)neutrino beam energy E: 25, 35, 45, 55, 70, 90, 110, 120, 170 GeV. In each bin x increases from left to right, $0.045 < x < 0.65$.

Whilst Def. 1 just deweights the nuclear data sets in a PDF fit, Def. 2 also attempts to directly apply a nuclear correction. In both cases we can construct a theoretical covariance matrix as

$$S_{ij} = \frac{1}{N_{rep}} \sum_{n=1}^{N_{rep}} \Delta_i^{(n)} \Delta_j^{(n)}.$$

(4)

We did this separately for each experiment, which is a conservative treatment.

Considering the square roots of the diagonal elements of the covariance matrices, normalised to data, y_i (Fig. 1), we see that the nuclear uncertainty has the largest impact on the NuTeV data, where the nuclear uncertainties dominate the data uncertainties. Given the high correlation of NuTeV observables with the s and $\bar{s}$ PDFs, the effect of including the uncertainties ought to be greatest for these PDFs.

To explicitly evaluate the impact of nuclear uncertainties, we compared four different PDF fits:

- **Baseline,** based on NNPDF3.1, with small improvements [10];
- **NoNuc,** Baseline with nuclear data removed;
- **NucUnc,** Baseline with nuclear uncertainties according to Def. 1;
- **NucCor,** Baseline with nuclear uncertainties and a nuclear correction according to Def. 2.

Table 1 shows the variation in χ^2 for selected data sets (for a full break-down see [10]). All of the fits show reduced χ^2 compared to Baseline, highlighting tension due to nuclear data. However,
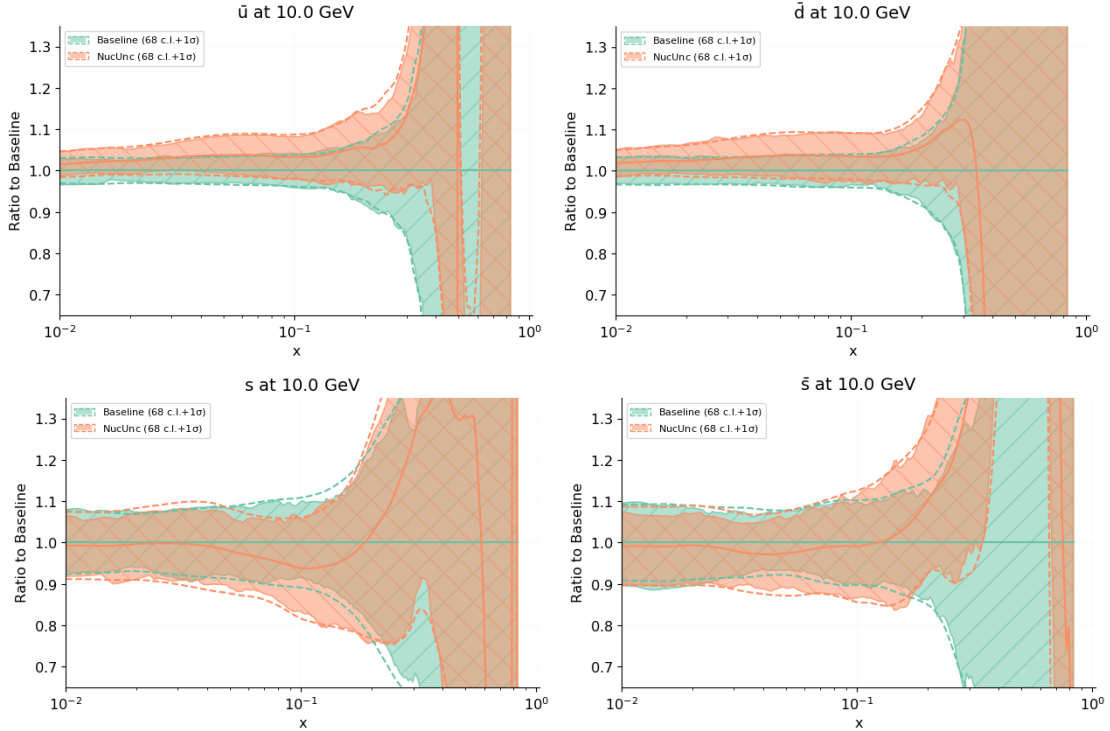
the strange-sensitive ATLAS W/Z at 7 TeV (2011) measurements [16] still have a poor χ^2, indicating that possible tensions with NuTeV were unlikely responsible for this; in any case, the data sets occupy different kinematic regions. The best fit is obtained for NucUnc, which has the largest uncertainties.

Fig. 2 shows the light sea quark PDFs for NucUnc compared to Baseline. These are the distributions with greatest impact, but there is little appreciable change other than a small shift in the central value and increase in uncertainties. NucCor behaves similarly. Overall, the nuclear uncertainties are small compared to the global experimental uncertainty.

Given the changes to the light sea quark PDFs, it is interesting to examine the impact on relevant phenomenological quantities, namely: the sea quark asymmetry, $\bar{u}/\bar{d}$; strangeness fraction, $R_s = (s + \bar{s})/(\bar{u} + \bar{d})$; and strange valence distribution, $xs^- = x(s - \bar{s})$ (Fig. 3). In all cases we found that removing the nuclear data has a significant effect, emphasising the need to retain this data in proton PDF fits. Adding nuclear uncertainties, however, makes very little difference. In

| Experiment	N_{dat}	Baseline	NoNuc	NucUnc	NucCor
CHORUS ν	416	1.29	–	0.97	1.04
CHORUS $\bar{\nu}$	416	1.20	–	0.78	0.83
NuTeV ν	39	0.41	–	0.31	0.40
NuTeV $\bar{\nu}$	37	0.90	–	0.62	0.83
E605 σ^p	85	1.18	–	0.85	0.89
ATLAS W/Z (2011)	34	1.97	1.78	1.87	1.94
ATLAS	360	1.08	1.04	1.04	1.05
CMS	409	1.07	1.07	1.07	1.07
LHCb	85	1.46	1.27	1.32	1.37
Total	4285	1.18	1.14	1.07	1.09

Table 1: χ^2 per data point for selected data sets. The final row shows results for the full fitted data.

particular, the known tension between ATLAS W/Z + HERA DIS data and NuTeV data, which is apparent in the strangeness fraction [17], is not relieved with the addition of nuclear uncertainties.

Figure 2: NucUnc fits with nuclear uncertainties (orange) compared to Baseline (green) for PDFs at 10 GeV. Clockwise from top left: $\bar{u}, \bar{d}, s$ and $\bar{s}$ PDFs. Error bands are 1 σ and the results are normalised to Baseline fit.

3

We found no appreciable difference between using NucUnc versus NucCor, so recommend to incorporate uncertainties using NucUnc (Def. 1) as this is the more conservative option.

We studied the role of nuclear data in proton PDF fits, and adopted an empirical approach to determine the nuclear uncertainties due to this data. We based our analysis on recent nPDF fits DSSZ12, nCTEQ15 and EPPS16. Using a theoretical covariance matrix, we included these uncertainties in proton PDF fits, and found that the fit quality was improved, with the largest effect on the light sea quark distributions. The PDF sets from this analysis are available upon request from the authors in LHAPDF format [18]. We found no significant impact on associated phenomenology.
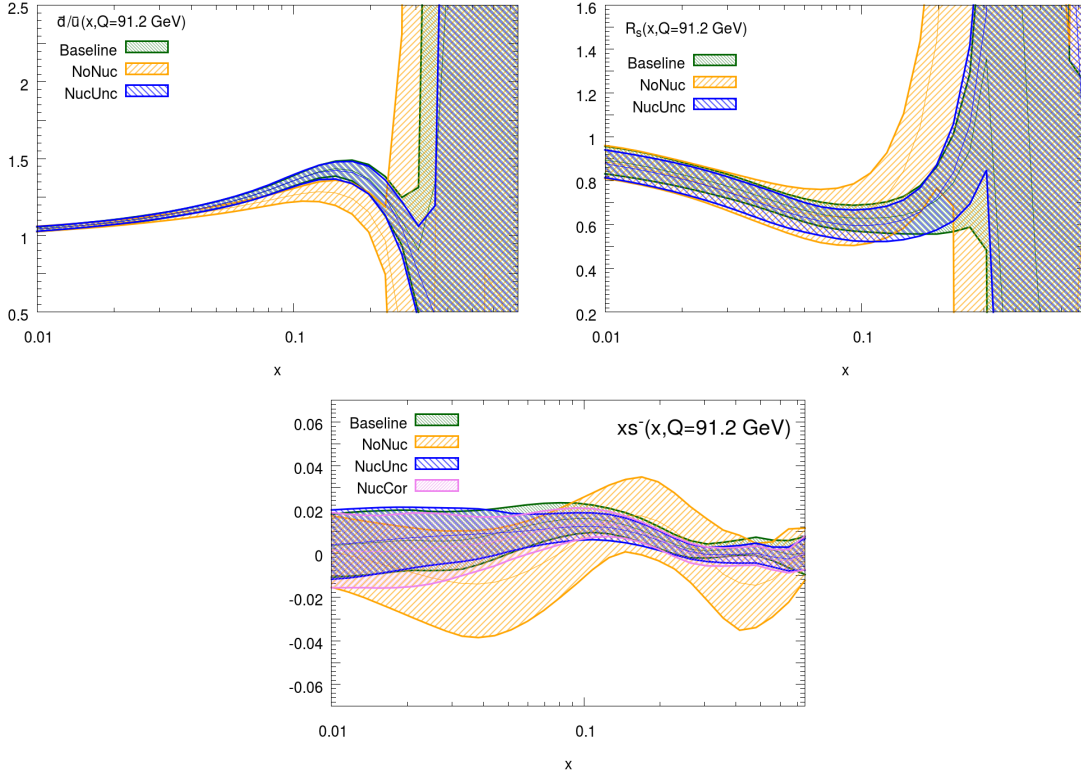
Figure 3: Effect of including nuclear uncertainties on phenomenology. Clockwise, from top left: sea quark asymmetry, strangeness fraction, strange valence distribution. Distributions correspond to the use of different PDF fits: Baseline (green), NoNuc (yellow), NucUnc (blue) and NucCor (pink). $Q = 91.2$ GeV. In the upper two plots, NucCor are indistinguishable from NucUnc so are omitted for readability.

We will extend this analysis to deuterium data, and in the future we will be able to use nuclear PDFs from NNPDF [19] to estimate uncertainties. These methods can be applied to other sources of theoretical uncertainties, such as higher twist effects, fragmentation functions, and missing higher order uncertainties [20].

References

[1] J. Gao, L. Harland-Lang, and J. Rojo, *The Structure of the Proton in the LHC Precision Era*, Phys. Rept. **742** (2018) 1–121, [arXiv:1709.04922].

[2] **NNPDF** Collaboration, R. D. Ball et al., *Precision determination of electroweak parameters and the strange content of the proton from neutrino deep-inelastic scattering*, Nucl. Phys. **B823** (2009) 195–233, [arXiv:0906.1958].

[3] **NNPDF** Collaboration, R. D. Ball et al., *Theoretical issues in PDF determination and associated uncertainties*, Phys. Lett. **B723** (2013) 330–339, [arXiv:1303.1189].

[4] **NNPDF** Collaboration, R. D. Ball et al., *A Determination of parton distributions with faithful uncertainty estimation*, Nucl. Phys. **B809** (2009) 1–63, [arXiv:0808.1231]. [Erratum: Nucl. Phys.B816,293(2009)].

[5] R. D. Ball, E. R. Nocera and R. L. Pearson, *Nuclear Uncertainties in the Determination of Proton PDFs*, Eur. Phys. J. **C79 (2019)**, [arXiv:1812.09074].

[6] **CHORUS** Collaboration, G. Onengut et al., *Measurement of nucleon structure functions in neutrino scattering*, Phys. Lett. **B632** (2006) 65–75.

[7] **NuTeV** Collaboration, M. Goncharov et al., *Precise measurement of dimuon production cross-sections in muon neutrino Fe and muon anti-neutrino Fe deep inelastic scattering at the Tevatron*, Phys. Rev. **D64** (2001) 112006, [hep-ex/0102049].

[8] **NuTeV** Collaboration, M. Tzanov et al., *Precise measurement of neutrino and anti-neutrino differential cross sections*, Phys. Rev. **D74** (2006) 012008, [hep-ex/0509010].

[9] **CDF** Collaboration, T. Aaltonen et al., *Measurement of the Inclusive Jet Cross Section at the Fermilab Tevatron p anti-p Collider Using a Cone-Based Jet Algorithm*, Phys. Rev. **D78** (2008) 052006, [arXiv:0807.2204]. [Erratum: Phys. Rev.D79,119902(2009)].

[10] **NNPDF** Collaboration, R. D. Ball et al., *Parton distributions from high-precision collider data*, Eur. Phys. J. **C77** (2017), no. 10 663, [arXiv:1706.00428].

[11] R. D. Ball and A. Deshpande, *The Proton Spin, Semi-Inclusive processes, and a future Electron Ion Collider*, 2018. [arXiv:1801.04842].

[12] M. Arneodo, *Nuclear effects in structure functions*, Phys. Rept. **240** (1994) 301–393.

[13] D. de Florian, R. Sassot, P. Zurita, and M. Stratmann, *Global Analysis of Nuclear Parton Distributions*, Phys. Rev. **D85** (2012) 074028, [arXiv:1112.6324].

[14] K. Kovarik et al., *nCTEQ15 - Global analysis of nuclear parton distributions with uncertainties in the CTEQ framework*, Phys. Rev. **D93** (2016), no. 8 085037, [arXiv:1509.00792].

[15] K. J. Eskola, et al., *EPPS16: Nuclear parton distributions with LHC data*, Eur. Phys. J. **C77** (2017), no. 3 163, [arXiv:1612.05741].

[16] **ATLAS** Collaboration, G. Aad et al., *Measurement of the inclusive $W^{\pm}$ and Z/gamma cross sections in the electron and muon decay channels in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, Phys. Rev. **D85** (2012) 072004, [arXiv:1109.5141].

[17] **ATLAS** Collaboration, G. Aad et al., *Determination of the strange quark density of the proton from ATLAS measurements of the $W \to \ell\nu$ and $Z \to \ell\ell$ cross sections*, Phys. Rev. Lett. **109** (2012) 012001, [arXiv:1203.4051].

[18] A. Buckley, et al., *LHAPDF6: parton density access in the LHC precision era*, Eur. Phys. J. **C75** (2015) 132, [arXiv:1412.7420].

[19] R. Abdul Khalek, J. J. Ethier and J. Rojo, *Nuclear Parton Distributions from Lepton-Nucleus Scattering and the Impact of an Electron-Ion Collider*, Eur. Phys. J. C **79** (2019) no.6, 471 doi:10.1140/epjc/s10052-019-6983-1 [arXiv:1904.00018].

[20] R. Abdul Khalek et al. [NNPDF Collaboration], *A First Determination of Parton Distributions with Theoretical Uncertainties*, Eur. Phys. J. **C75** (2015) 132, [arXiv:1905.04311].