How PDF fits became Precision Physics

LightCone 2018, Jefferson Laboratory

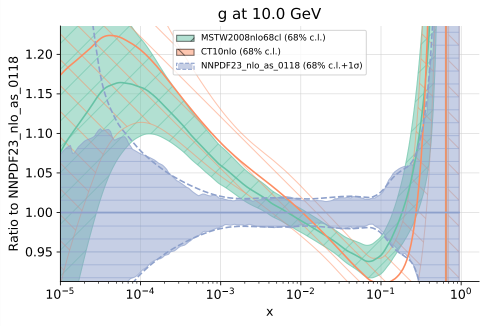
Zahari Kassabov

May 16th 2018

Cavendish Laboratory, University of Cambridge

PDF fits at the time of the Higgs discovery

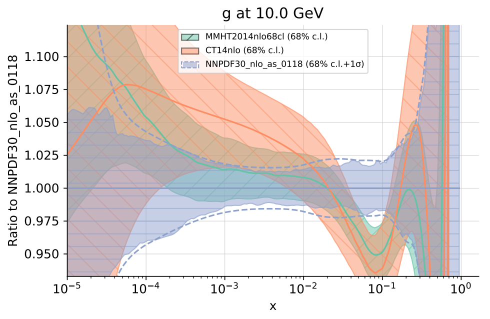
- Several groups obtained incompatible results.



- Benchmarking exercises to understand the differences largely inconclusive.
- Recommendation: For the most part, ignore individual group uncertainties. Take “envelope” of individual determinations.

PDFs including LHC data

- Situation much improved: Several groups produced compatible results.



- PDF4LHC 2015 recommendation [arXiv:1510.03865] recognized most import features in PDF fits.
- PDF uncertainties become meaningful: Propagate experimental uncertainties to best fit of unbiased interpolants.

Some features of contemporary PDF fits

- Data driven approach: Avoid making assumptions on the non perturbative behaviour. Use flexible parametrization:
  - CTEQ and MMHT: Bernstein and Chebichev polynomials
  - NNPDF: Neural networks
- Self validating procedures:
  - CTEQ and MMHT: dynamic tolerance.
  - NNPDF: closure tests.
- Use as much data as possible
- Use the best available perturbative QCD theory:
  - NNLO thoery with a General Mass Variable Flavour Number Scheme for heavy quark emission
  - Photon PDF
  - Fitted charm

Differences with other groups can be explained by differences in these assumptions (see e.g. Thorne 1201.6180).

PDF timeline [Global sets, published only]

	2008		2009		2010		2011	2012		2013		2014		2015	2017	
SET	CTEQ6.6	NNPDF1.0	MSTW	ABKM09	NNPDF2.0	CT10 (NLO)	NNPDF2.1 (NNLO)	ABM11	NNPDF2.3	CT10 (NNLO)	ABM12	NNPDF3.0	MMHT	CT14	ABMP16	NNPDF3.1
MONTH	(02)	(08)	(01)	(08)	(02)	(07)	(07)	(02)	(07)	(02)	(10)	(10)	(12)	(06)	(01)	(06)
F. T. DIS	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
ZEUS+H1-HI	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
COMB. HI	✗	✗	✗	✗	✔	✗	✔	✗	✔	✗	✔	✔	✗	✗	✔	✔
ZEUS+H1-HII	✗	✗	✗	✗	✗	✗	some	✗	✗	some	✗	✔	✗	✗	✔	✔
HERA JETS	✗	✗	✔	✗	✗	✗	✗	✗	✗	✗	✗	✗	✔	✗	✗	✗
F. T. DY	✔	✗	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
TEV W+Z	✔	✗	✔	✗	✔	✔	✔	✗	✔	✔	✗	✔	✔	✔	✗	✔
LHC W+Z	✗	✗	✗	✗	✗	✗	✗	✗	✔	✗	some	✔	✔	✔	some	✔
TEV JETS	✔	✗	✔	✗	✔	✔	✗	✔	✔	✔	✗	✔	✔	✔	✗	✔
LHC JETS	✗	✗	✗	✗	✗	✗	✗	✗	✔	✗	✗	✔	✔	✔	✗	✔
TOP TOTAL	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✔	✔	✗	✗	✔	✔
SINGLE TOP TOTAL	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✔	✗
TOP DIFFERENTIAL	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✔
W p_T	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✔	✗	✗	✗	✗
W+C	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✔	✗	✗	✗	✗
Z p_T	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✔

Figure 1: S.Forte 2018

Agreement keeps improving

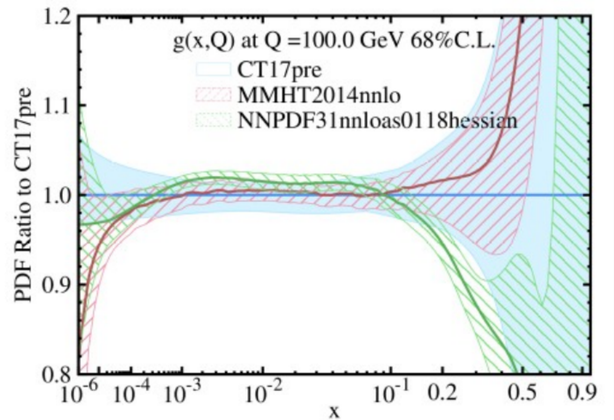
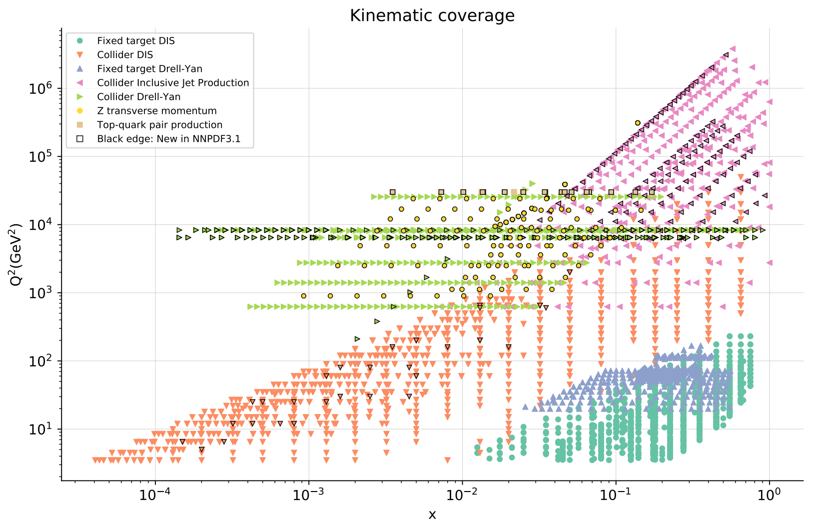


Figure 2: Tie-Jiun Hou, DIS 2018 (Unpublished)

NNPDF 3.1

- Updates NNPDF 3.0 with:
    - New data (LHC + Updated HERA).
    - Data driven description of the charm PDF.
- Results largely compatible but:
    - Smaller uncertainties
    - Improved accuracy

Dataset in NNPDF 3.1



Difficulties with high precision data

- Increased precision in experiment data and theoretical calculation imposes challenges in the PDF fitting toolchain (e.g. interpolation of the DGLAP solution can make a big difference)
- Parton level calculation can be problematic. E.g. large numerical instabilities in $Z\,p_T$ calculation (Boughezal et al, arxiv:1705.00343).
- Experimental input can be problematic:
    - Improved statistics $\to$ Correlated systematics dominant $\to$ Covariance matrices (used for χ^2 goodness of fit) near singular $\to$ χ^2 given by not so well controlled systematics.

Example: ATLAS 7 TeV Jets

- Fit quality in the MMHT framework increases very substantially if one relaxes some correlations across rapidity bins:

	Full	21	62	21,62
$\chi^2/N_{\text{pts.}}$	2.85	1.58	2.36	1.27

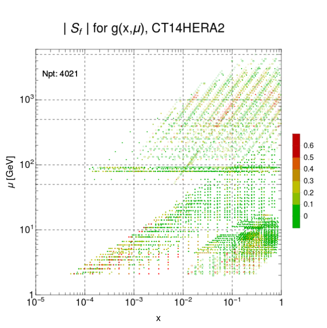
Table 1: χ^2 per number of data points ($N_{\text{pts}} = 140$) for fit to ATLAS jets data [23], with the default systematic error treatment ('full') and with certain errors, defined in the text, decorrelated between jet rapidity bins.

Figure 3: (Harland-Lang, Martin, Thorne arxiv:1711.05757)

- Highly non trivial analyses are required to understand problems with data. Better tools necessary.

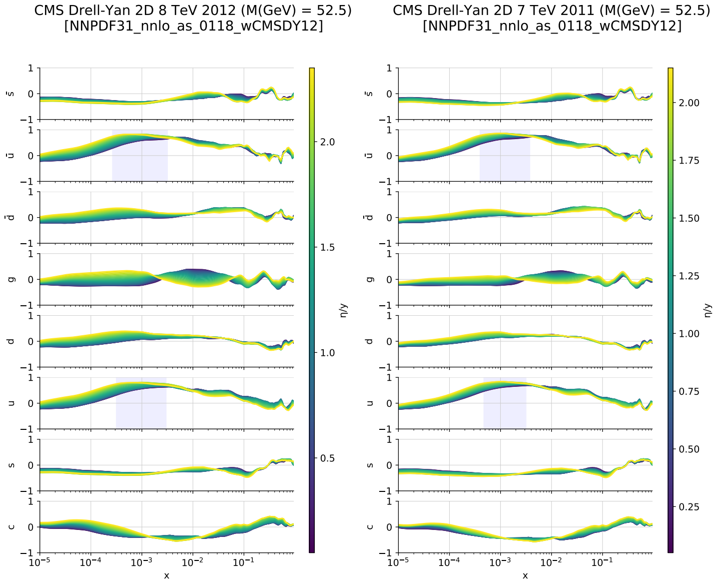
Tools for compression, visualization and data impact

- PDFSense (Wang et al, arxiv:1803.02777). Define a *sensitivity* based on correlations of data and parameters, and on the residuals to the best fit.

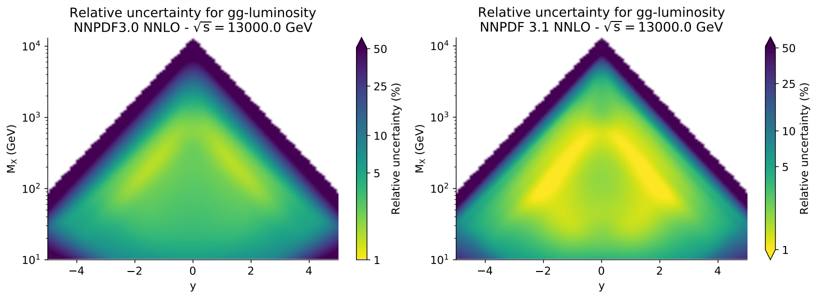


- SMPDF (Carrazza, Forte, Z.K., Rojo, arxiv:1602.00005) Use correlations directly in *x* space to construct an optimized representation or study the PDF dependence.

Example: Studying data dependence with SMPDF

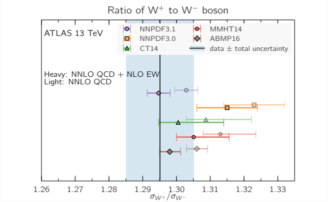


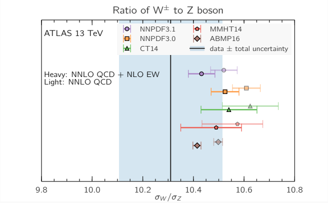
Some results for NNPDF 3.1

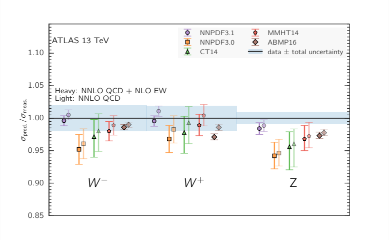


- PDF Uncertainties in data region can be below 1%!
  - How much can these be trusted?

Impact on the W and Z cross sections







Note:

- These are *predictions* for standard candles.
- Improvements likely driven by fitted charm PDF.

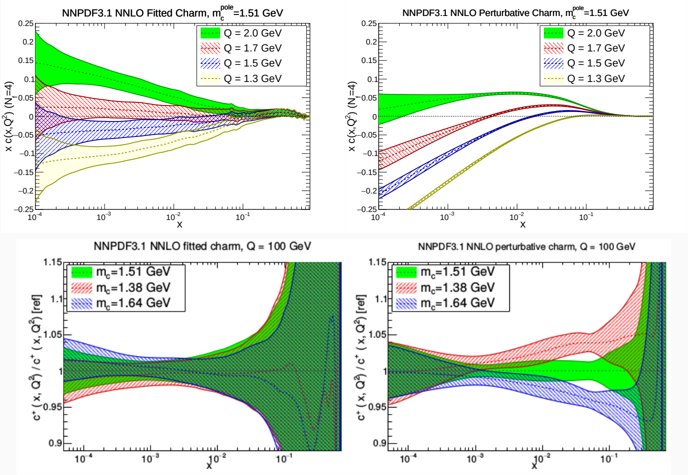
Determining charm PDF from the data

- Usually, charm generated only perturbatively (or parametrized with models e.g. CT14 arxiv:1707.00657).
  - Induces large dependence on the charm mass.
  - Possible non perturbative effects missed.
- Motto: Replace assumptions on non perturbative physics and theoretical ambiguities by data driven approach
  - $\Longrightarrow$Fit the charm PDF.
- Teach the GM-VNFS to account for massive quarks: Ball et al, arXiv:1510.02491.

Effects of fitting charm

Compared to a purely perturbative charm PDF,

- Reduced scale dependence + charm vanishing at $Q^2 \gtrsim m_c$
  $\rightarrow$Increased strangeness at high Q^2 $\rightarrow$ Better fit to the LHC data.



The photon PDF from LUXQED

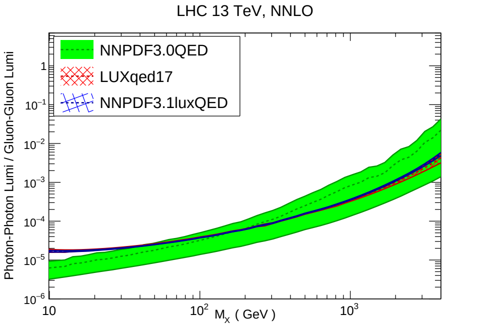
- Important uncertainty for many analyses (e.g. Higgs+W). Most requested feature since NNPDF3.1.
- Previously determined using a restricted model (MRST2004) or constrained from data with large uncertainties (NNPDF 2.3 QED).
- Breakthrough in the understanding (Manohar, Nason, Salam, Zanderighi, arxiv:1607.04266):
  - Photon PDF can be written in terms of the structure functions:

$$x f_{\gamma/p}(x,\mu^2) = \frac{1}{2\pi\alpha(\mu^2)} \int_x^1 \frac{dz}{z} \left\{ \int_{\frac{x^2 m_p^2}{1-z}}^{\frac{\mu^2}{1-z}} \frac{dQ^2}{Q^2} \alpha^2(Q^2) \left[\left(z p_{\gamma q}(z) + \frac{2x^2 m_p^2}{Q^2} \right) F_2(x/z, Q^2) - z^2 F_L\left(\frac{x}{z}, Q^2\right) \right] - \alpha^2(\mu^2) z^2 F_2\left(\frac{x}{z}, \mu^2\right) \right\},$$

Photon PDF in NNPDF

Matched the LUXQED ansatz with the NNPDF formalism e.g. to conserve momentum fraction.

- In agreement with the previous data driven fit, but with much reduced uncertainties.



Using precise PDFs: A determination of the string coupling constant

Adapted the NNPDF methodology to work as if PDFs and α_S were fitted simultaneously. Apply to the NNPDF 3.1 dataset [arxiv: 1802.03398].

$$\alpha_S^{\mathrm{NNLO}}(M_Z) = 0.11845 \pm 0.00052(0.4\%)^{(\mathrm{exp})} \pm 0.0011^{(\mathrm{MHOU})}$$

$$\alpha_S^{\mathrm{NLO}}(M_Z) = 0.12067 \pm 0.00064(0.4\%)^{(\mathrm{exp})}$$

- Negligible systematic uncertainties induced in the process, studied extensively.
- Experimental uncertainty comparable to most precise determinations in the PDG.
- Only way to find the best fit to *all* data (i.e. also in PDFs) in fits of α_s to collider data [ZK, arxiv:1802.05236].

Missinh higher order uncertainties in α_S

$$\alpha_S^{\text{NNLO}}(M_Z) = 0.11845 \pm 0.00052(0.4\%)^{(\text{exp})} \pm 0.0011^{(\text{MHOU})}$$

- No good way of estimating Missing Higher Order Uncertainties. Conservatively, take half the difference between the result at NLO and NNLO.
- There are some arguments that this is overestimated due to the bad fit [Carrazza, Forte, ZK, Rojo, Rottoli, arxiv:1803.07977].

Missing higher order uncertainties in PDFs (Very preliminary!)

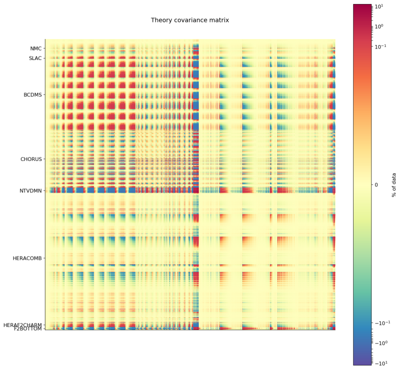
Include a theory covariance matrix as well as an experimental covariance matrix in the PDF fit.

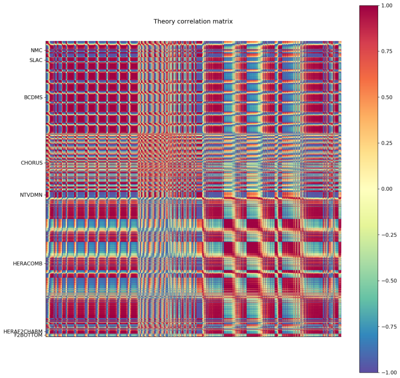
- Compute e.g. using scale variations

$$s_{ij} = \frac{1}{2}\left(\left(\sigma_i(2\mu) - \sigma_i(\mu)\right)\left(\sigma_j(2\mu) - \sigma_j(\mu)\right) + \left(\sigma_i(\tfrac{1}{2}\mu) - \sigma_i(\mu)\right)\left(\sigma_j(\tfrac{1}{2}\mu) - \sigma_j(\mu)\right)\right)$$

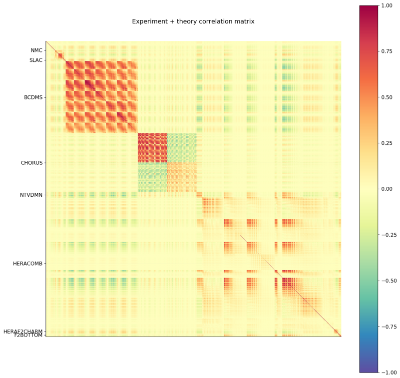
- s_{ij} element of the theoretical covariance between data points i and j.
- $\sigma_i(\mu)$ Theoretical prediction for point i evaluated with scales $\mu_r = \mu_f = \mu$.

Theoretical covariance and correlation





Theory + Experiment covariance (Very preliminary)



- Propagates *both* experimental and theory uncertainties to the fit.
- Weights down points with large perturbative corrections.
- Allows various correlation models to be studied.

Thank you!