Unpolarised Parton Distribution Functions today: needs, achievements and challenges

Thirteenth Conference
on the Intersections of Particle and Nuclear Physics

Emanuele R. Nocera

School of Physics and Astronomy - University of Edinburgh

Indian Wells – 1st June 2018
Outline

1 Needs
   ▶ Accuracy and precision

2 Achievements
   ▶ Data: impact of latest LHC measurements
   ▶ Theory: NNLO QCD corrections, fitting charm, the photon PDF, resummed PDFs

3 Challenges
   ▶ Theory: including missing higher order uncertainties in a fit
   ▶ Methodology: tools for compression, visualisation and minimisation

4 Conclusions

DISCLAIMER

I will focus on collinear, unpolarised parton distribution functions

Emphasis on recent achievements and on topics which I’ve worked on recently

Apologies in advance for not discussing your favourite subject

For an extensive review of topics not addressed in this talk, please see

[Phys.Rept. 742 (2018) 1; WG1 summary talk at DIS2018]
1. Needs
Factorisation of physical observables

\[ \mathcal{O}_I = \sum_{f=q,\bar{q},g} C_I f(x, \alpha_s(\mu^2)) \otimes f(x, \mu^2) + \text{p.s. corrections} \]

\[ f \otimes g = \int_x^1 \frac{dy}{y} f \left( \frac{x}{y} \right) g(y) \]

<table>
<thead>
<tr>
<th>Process</th>
<th>Reaction</th>
<th>Subprocess</th>
<th>PDFs probed</th>
<th>x</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\ell^{\pm}{p,n} \to \ell^{\pm} + X)</td>
<td>(\gamma^* q \to q)</td>
<td>(q, \bar{q}, g)</td>
<td>(x \gg 0.01)</td>
<td></td>
</tr>
<tr>
<td>(\ell^{\pm} n/p \to \ell^{\pm} + X)</td>
<td>(\gamma^* d/u \to d/u)</td>
<td>(d/u)</td>
<td>(x \gg 0.01)</td>
<td></td>
</tr>
<tr>
<td>(\nu(\bar{\nu})N \to \mu^- (\mu^+) + X)</td>
<td>(W^* q \to q')</td>
<td>(q, \bar{q})</td>
<td>(0.01 \lesssim x \lesssim 0.5)</td>
<td></td>
</tr>
<tr>
<td>(\nu N \to \mu^- \mu^+ + X)</td>
<td>(W^* s \to c)</td>
<td>(s)</td>
<td>(0.01 \lesssim x \lesssim 0.2)</td>
<td></td>
</tr>
<tr>
<td>(\bar{\nu} N \to \mu^+ \mu^- + X)</td>
<td>(W^* \bar{s} \to \bar{c})</td>
<td>(\bar{s})</td>
<td>(0.01 \lesssim x \lesssim 0.2)</td>
<td></td>
</tr>
<tr>
<td>(e^\pm p \to e^\pm + X)</td>
<td>(\gamma^* q \to q)</td>
<td>(g, q, \bar{q})</td>
<td>(0.0001 \lesssim x \lesssim 0.1)</td>
<td></td>
</tr>
<tr>
<td>(e^+ p \to \bar{\nu} + X)</td>
<td>(W^+ {d, s} \to {u, c})</td>
<td>(d, s)</td>
<td>(x \gg 0.01)</td>
<td></td>
</tr>
<tr>
<td>(e^\pm p \to e^\pm c\bar{c} + X)</td>
<td>(\gamma^* c \to c, \gamma^* g \to c\bar{c})</td>
<td>(c, g)</td>
<td>(0.0001 \lesssim x \lesssim 0.1)</td>
<td></td>
</tr>
<tr>
<td>(e^\pm p \to jet(s) + X)</td>
<td>(\gamma^* g \to q\bar{q})</td>
<td>(g)</td>
<td>(0.01 \lesssim x \lesssim 0.1)</td>
<td></td>
</tr>
<tr>
<td>(pp \to \mu^+ \mu^- + X)</td>
<td>(u\bar{u}, d\bar{d} \to \gamma^*)</td>
<td>(\bar{q})</td>
<td>(0.015 \lesssim x \lesssim 0.35)</td>
<td></td>
</tr>
<tr>
<td>(pn/pp \to \mu^+ \mu^- + X)</td>
<td>((ud)/(u\bar{u}) \to \gamma^*)</td>
<td>(\bar{d}/\bar{u})</td>
<td>(0.015 \lesssim x \lesssim 0.35)</td>
<td></td>
</tr>
<tr>
<td>(p\bar{p}(pp) \to jet(s) + X)</td>
<td>(gg, qg, qq \to 2\text{jets})</td>
<td>(g, q)</td>
<td>(0.005 \lesssim x \lesssim 0.5)</td>
<td></td>
</tr>
<tr>
<td>(p\bar{p} \to (W^\pm \to \ell^\pm \nu^\pm) + X)</td>
<td>(ud \to W^+, \bar{u}\bar{d} \to W^-)</td>
<td>(u, d, \bar{u}, \bar{d})</td>
<td>(x \gg 0.05)</td>
<td></td>
</tr>
<tr>
<td>(pp \to (W^\pm \to \ell^\pm \nu^\pm) + X)</td>
<td>(u\bar{d} \to W^+, u\bar{d} \to W^-)</td>
<td>(u, d, \bar{u}, \bar{d}, (g))</td>
<td>(x \gg 0.01)</td>
<td></td>
</tr>
<tr>
<td>(p\bar{p}(pp) \to (Z \to \ell^+ \ell^-) + X)</td>
<td>(uu, dd(u\bar{u}, dd) \to Z)</td>
<td>(u, d(g))</td>
<td>(x \gg 0.01)</td>
<td></td>
</tr>
<tr>
<td>(pp \to (W + c) + X)</td>
<td>(gs \to W^- c, g\bar{s} \to W^+ \bar{c})</td>
<td>(s, \bar{s})</td>
<td>(x \sim 0.01)</td>
<td></td>
</tr>
<tr>
<td>(pp \to t\bar{t} + X)</td>
<td>(gg \to t\bar{t})</td>
<td>(g)</td>
<td>(x \sim 0.01)</td>
<td></td>
</tr>
</tbody>
</table>
A global determination of parton distribution functions
A mathematically ill-posed problem: determine a set of functions from a finite set of data

METHODOLOGY

1. Parametrisation: general, smooth, flexible at an initial scale $Q_0^2$

$$x f_i(x, Q_0^2) = A_{f_i} x^{a_{f_i}} (1 - x)^{b_{f_i}} \mathcal{F}(x, \{c_{f_i}\})$$

small $x$

$$x f_i(x, Q^2) \xrightarrow{x \to 0} x^{a_{f_i}}$$

large $x$

$$x f_i(x, Q^2) \xrightarrow{x \to 1} (1 - x)^{b_{f_i}}$$

smooth interpolation in between

2. A prescription to determine/compute expectation values and uncertainties

$$E[\mathcal{O}] = \int \mathcal{D} \Delta f \mathcal{P}(\Delta f|\text{data}) \mathcal{O}(\Delta f) \quad V[\mathcal{O}] = \int \mathcal{D} \Delta f \mathcal{P}(\Delta f|\text{data}) [\mathcal{O}(\Delta f) - E[\mathcal{O}]]^2$$

Monte Carlo: $\mathcal{P}(\Delta f|\text{data}) \longrightarrow \{\Delta f_k\}$

Maximum likelihood: $\mathcal{P}(\Delta f|\text{data}) \longrightarrow \Delta f_0$

$$E[\mathcal{O}] \approx \frac{1}{N} \sum_k \mathcal{O}(\Delta f_k) \quad E[\mathcal{O}] \approx \mathcal{O}(\Delta f_0)$$

$$V[\mathcal{O}] \approx \frac{1}{N} \sum_k [\mathcal{O}(\Delta f_k) - E[\mathcal{O}]]^2 \quad V[\mathcal{O}] \approx \text{Hessian, } \Delta \chi^2 \text{ envelope, …}$$

3. A self-validating procedure (closure test, dynamic tolerance)

COMBINED WITH THEORY AND DATA TO FIND BEST-FIT PDFs

theory: NNLO QCD, GM-VFNS, charm, photon, …

data set: as global as possible
Example: the gluon PDF

**circa 2012**

- incompatible results from different groups
- benchmarking exercise largely inconclusive
- recommendation: ignore individual group uncertainties
- take the envelope of individual determinations

**circa 2015**

- compatible results from different groups
- PDF uncertainties become meaningful
- recommendation (PDF4LHC): combine individual group uncertainties into a statistically meaningful set

Agreement keeps improving

- residual differences among groups can be explained in terms of differences in the data set, details of the QCD analysis and methodology
  
  [PRD 86 (2012) 074017]
Example: the gluon PDF

Agreement keeps improving.

residual differences among groups can be explained in terms of differences in the data set, details of the QCD analysis and methodology [PRD 86 (2012) 074017]

[Tie-Jiun Hou, DIS 2018]
### Overview of recent PDF determinations

<table>
<thead>
<tr>
<th></th>
<th>NNPDF3.1</th>
<th>MMHT2014</th>
<th>CT14</th>
<th>HERAPDF2.0</th>
<th>CJ15</th>
<th>ABMP16</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed target DIS</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>JLAB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>HERA I+II</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>HERA jets</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Fixed target DY</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Tevatron $W, Z$</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Tevatron jets</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>LHC jets</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>LHC vector boson</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>LHC top (incl.)</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>LHC (diff.)</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>statistical treatment</th>
<th>Monte Carlo</th>
<th>Hessian $\Delta \chi^2$ dynamical</th>
<th>Hessian $\Delta \chi^2$ dynamical</th>
<th>Hessian $\Delta \chi^2 = 1$</th>
<th>Hessian $\Delta \chi^2 = 1.645$</th>
<th>Hessian $\Delta \chi^2 = 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>parametrisation</td>
<td>Neural Network (259 pars)</td>
<td>Chebyshev pol. (37 pars)</td>
<td>Bernstein pol. (30-35 pars)</td>
<td>polynomial (14 pars)</td>
<td>polynomial (24 pars)</td>
<td>polynomial (15 pars)</td>
</tr>
<tr>
<td>HQ scheme</td>
<td>FONLL</td>
<td>TR′</td>
<td>ACOT-$\chi$</td>
<td>TR′</td>
<td>ACOT-$\chi$</td>
<td>FFN</td>
</tr>
</tbody>
</table>

See also recommendations for PDF usage in computations of (LHC) high-energy processes [JPG 43 (2016) 023001, EPJC 76 (2016) 471]
The role of PDF uncertainties

1. Higgs boson characterisation
   PDF uncertainty often dominant contribution to theory uncertainty

2. Determination of SM parameters
   PDF uncertainty largest theoretical uncertainty in $M_W$ determination

3. BSM gluino production
   the larger the mass of the final state the larger the PDF uncertainty

---

[PRD 91 (2015) 113005]

- $\sigma_{13 \text{ TeV}}$
  - $ggF (N3LO)$: 48.5 pb
  - $VBF (N2LO)$: 3.78 pb
  - $WH (N2LO)$: 1.37 pb
  - $ZH (N2LO)$: 0.88 pb
  - $ttH (N1LO)$: 0.51 pb

---

[EPJ C76 (2016) 53]

- $\hat{K}^{\text{NLO}+\text{NLL}}(pp \to \bar{g}g + X)$
  - $\sqrt{S} = 13 \text{ TeV}$
  - $\mu_F = \mu_R = m$

- $m_{\bar{q}} = m_{\bar{g}} = m$
2. Achievements
A plethora of new data

1. **GLUON**
   - inclusive jets and dijets (medium/large $x$)
   - isolated photon and $\gamma$+jets (medium/large $x$)
   - top pair production (large $x$)
   - high $p_T$ $V$ production (small/medium $x$)

2. **QUARKS**
   - high $p_T W (+$ jets) ratios (medium/large $x$)
   - $W$ and $Z$ production (medium $x$)
   - low and high mass DY (small and large $x$)
   - $W + c$ (strange at medium $x$)

3. **PHOTON**
   - low and high mass DY
   - $WW$ production

4. Great progress also in interface NLO (NNLO) codes to PDF fitting codes
   - FASTNLO [Kluge et al., 2010]
   - aMCfast [JHEP 1408 (2014) 166]
   - MCgrid [CPC 185 (2014) 2115]
   - APFELgrid [CPC 212 (2017) 205]
A wealth of new NNLO calculations

explosion of calculations in past 24 months

[Slide: courtesy of G. Salam, updated April 2017]
The gluon PDF at large $x$: $t\bar{t}$ differential distributions

NNLO, global fits, LHC 13 TeV

ATLAS and CMS rapidity distributions at $\sqrt{s} = 8$ TeV

Significant reduction of $gg$ luminosity uncertainties at $M_X \geq O(1)$ TeV

e.g., at $M_X \sim 2$ TeV, uncertainties decrease from 13% to 5%

Impact of $t\bar{t}$ differential data similar to that of jet data
though jet data analysed neglecting NNLO QCD corrections in the matrix element

A precision determination of the gluon PDF at large $x$ is now possible at NNLO
the situation should only improve thanks to the recent NNLO jet calculation

$t\bar{t}$ differential distributions are included in the NNPDF3.1 PDF release

[see JHEP 1704 (2017) 044 and EPJ C77 (2017) 663 for details]
The gluon PDF at medium $x$: the $Z$-boson $p_T$ distribution

ATLAS and CMS $p_T$ distributions at $\sqrt{s} = 8$ TeV in various rapidity bins in the $Z$-peak region

NNLO/NLO $K$-factors 5%-10% depending on the rapidity/invariant mass region challenge: measurements have sub-percent experimental errors

Complementary information on the gluon PDF e.g., at $M_X \sim 2$ TeV, uncertainties decrease from 13% to 8%

$Z$ $p_T$ distributions are included in the NNPDF3.1 PDF release

[see JHEP 1707 (2017) 130 and EPJ C77 (2017) 663 for details]
The gluon PDF at small $x$: forward charm production

$D$ meson production from LHCb at different center-of-mass energies

$$N_{ij}^{X} = \frac{d^{2}\sigma(X \text{ TeV})}{dy_{i}^{D} dp_{T}^{D}} \frac{d^{2}\sigma(X \text{ TeV})}{dy_{j}^{D} ref dp_{T}^{D}}$$

$$R_{13}^{ij}/X = \frac{d^{2}\sigma(13 \text{ TeV})}{dy_{i}^{D} dp_{T}^{D}} \frac{d^{2}\sigma(X \text{ TeV})}{dy_{j}^{D} dp_{T}^{D}}$$

Gluon PDF errors are reduced by up to a factor 10 below $x \sim 10^{-5}$ robust w.r.t theoretical uncertainties (charm mass, scale variations, alternative reference bins)

Combine result with future LHeC measurements of $F_{L}$

test for BFKL resummations and non-linear QCD dynamics

Application: ultra high-energy (UHE) neutrino-nucleus cross-sections

NLO QCD provides a prediction accurate to $\lesssim 10\%$ at $E_{\nu} \sim 10^{12}$ GeV

[see PRL 118 (2017) 072001 for details]
Quark flavour separation from LHC data

High-precision $W$ and $Z$ production data from ATLAS, CMS and LHCb handle on quark/antiquark flavour separation

Largest impact on light quarks at large $x$ provided by LHCb data

error reduction by a factor 2 in NNPDF3.1 at $x \sim 0.1$

Combined effect of (LHC) CMS, LHCb and (Tevatron) D0 $W$, $Z$ data

improved determination of $x(u_V - d_V)$

[see R. Thorne’s talk at DIS2017 and EPJ C77 (2017) 663 for details]
In most PDF fits the strange PDF is suppressed w.r.t up and down sea quark PDFs effect mostly driven by neutrino dimuon data

A symmetric strange sea PDF is preferred by collider data in particular by ATLAS $W, Z$ rapidity distributions (2011) [EPJ C77 (2017) 367]

$$R_s(x, Q^2) = \frac{s(x, Q^2) + \bar{s}(x, Q^2)}{\bar{u}(x, Q^2) + \bar{d}(x, Q^2)} \begin{cases} \sim 0.5 \text{ from neutrino and CMS } W + c \text{ data} \\ \sim 1.0 \text{ from ATLAS } W, Z \end{cases}$$

The ATLAS data can be accommodated in the global fit increased strangeness, though not as much as in a collider-only fit some tension remains between collider and neutrino data Suppressed strangeness confirmed by recent $W + c$ CMS analysis [CMS PAS SMP-17-014]
Parametrise the $c^+(x, Q^2_0)$, quark and gluon PDFs on the same footing
stabilise the dependence of LHC processes upon variations of $m_c$
quantify the nonperturbative charm component in the proton (BHPS? sea-like?)
take into account massive charm-initiated contribution to the DIS structure functions

Fitted charm found to differ from perturbative charm at scales $Q \sim m_c$ in NNPDF3.1
preference for a BHPS-like shape
shape driven by LHCb $W,Z$ data + EMC data

At $Q = 1.65$ GeV charm carry $0.26 \pm 0.42$ % of the proton momentum
but it is affected by large uncertainties, especially if no EMC data are included
The photon PDF: how bright is the proton?

The photon PDF $\gamma(x, Q)$ in LUXqed

$$xf_{\gamma/p}(x, \mu^2) = \frac{1}{2\pi\alpha(\mu^2)} \int_x^1 \frac{dz}{z} \left\{ \int_{x\frac{\mu^2}{Q^2}}^{x\frac{\mu^2}{1-z\mu^2}} \frac{dQ^2}{Q^2} \alpha^2(Q^2) \right\}$$

$$\left( zp_{\gamma q}(z) + \frac{2x^2m_p^2}{Q^2} \right) F_2(x/z, Q^2) - z^2 F_L \left( \frac{x}{z}, Q^2 \right)$$

$$- \alpha^2(\mu^2) z^2 F_2 \left( \frac{x}{z}, \mu^2 \right)$$

Good agreement with NNPDF3.0QED, model-independent fit from LHC DY data

NNPDF3.0QED: model-independent determination of $\gamma(x, Q)$ from LHC $W, Z$ data affected by large uncertainties, $\mathcal{O}(100\%)$ due to limited experimental information

LUXQED: compute $\gamma(x, Q)$ in terms of inclusive structure functions $F_2$ and $F_L$ significant improvement in the PDF uncertainty implications for high-mass processes for BSM searches, e.g. DY production at the TeV scale

NNPDF3.1LUXQED: consistent NNPDF fit with LUXQED constraint good agreement, but smaller uncertainties sizable impact on precision physics: e.g. associated Higgs production with $W$

Beyond fixed-order accuracy

small $x$: \( \frac{1}{x} \ln^k x \)
high-energy gluon emission: single logs

Large logs $\alpha_s \ln \sim 1$ spoil the convergence of the perturbative series

large $x$: \( \left( \frac{\ln^k (1-x)}{(1-x)} \right)^+$
soft gluon emission: double logs

PDFs with threshold resummation [JHEP 1509 (2015) 191] (only DIS, DY $Z/\gamma$, total $t\bar{t}$ + evol.)
suppression in PDFs partially or totally compensates enhancements in partonic cross-sections
accuracy of the resummed fit competitive with the fixed-order fit, except for the large-$x$ gluon
large uncertainties for MSSM particle resummed cross-sections [EPJ C76 (2016) 53]

PDFs with high-energy resummation [EPJ C78 (2018) 321] (only DIS + evol.)
Resummed PDFs enhanced at small $x$, uncertainties reduced
Large effects for future colliders, or $b$ production at LHC
The correlated replica method and $\alpha_s$  

How can we take into account PDF/$\alpha_s$ correlations in a Monte Carlo way?

for each data sample (replica),
perform a scan in $\alpha_s$

each replica has a preferred value of the $\alpha_s$ (the minimum of each parabola)
these preferred values form a Monte Carlo distribution

$$\alpha_s^{\text{NNLO}}(M_Z) = 0.1185 \pm 0.0005_{\text{exp}} \pm 0.0001_{\text{meth}} \pm 0.0011_{\text{th}} = 0.1185 \pm 0.0012(1\%)$$
3. Challenges
Towards 1% PDF uncertainties

Typical PDF uncertainty in data region of order 1%
Can we believe in 1% PDF uncertainties? What are the consequences?
Higher data precision, more fit challenges

Example 1: ATLAS 7 TeV jets [EPJ C78 (2018) 248]
Each rapidity bin can be fitted with $\chi^2/d.o.f. \sim 1$, best-fit PDFs indistinguishable
If all bins are fitted simultaneously, $\chi^2/d.o.f. \sim 3$
⇒ misestimated correlations?

from 2011 to 2012, uncorrelated uncertainties down to sub-permille
2011: $\chi^2/d.o.f. \sim 1$; 2012: impossible to fit better than $\chi^2/d.o.f. \sim 3$
⇒ pathological behaviour of covariance matrix, what is the uncertainty on it?

Example 3: The ATLAS 7 TeV $p_T$ distribution [EPJ C77 (2017) 663]
uncorrelated statistical uncertainties at permille level
large NNLO corrections $\sim 10\%$, but nominal $K$-factor uncertainties very small
⇒ fit only possible with estimate of theory uncertainties
Including the theory covariance matrix in a fit

Very preliminary

\[ \mathcal{O}_i(\mu_R, \mu_F), \ i = 1, N_{\text{dat}} \]

\[ \Delta^+_i = \mathcal{O}_i(\mu_R, \mu_F) - \mathcal{O}_i(2\mu_R, 2\mu_F) \]

\[ \Delta^-_i = \mathcal{O}_i(\mu_R, \mu_F) - \mathcal{O}_i(\frac{1}{2}\mu_R, \frac{1}{2}\mu_F) \]

\[
\text{Cov}_{\text{th}}[\mathcal{O}_i, \mathcal{O}_j] = \Delta^+_i \Delta^+_j - \Delta^-_i \Delta^-_j
\]

\[ \text{Cov}_{\text{tot}} = \text{Cov}_{\text{exp}} + \text{Cov}_{\text{th}} \]
Computational efficiency

**Issue 1: PDF fits are computationally expensive**
Can modern optimisation tools (evolutionary strategies, analytical gradients) help?
Assess the impact of the data without refitting
Bayesian reweighting \[\text{NPB 855 (2012) 608}\] and Hessian profiling \[\text{JHEP 12 (2014) 100}\]

**Issue 2: Monte Carlo sets are delivered in terms of a large number of replicas**
Option 1: compression \[\text{EPJ C75 (2015) 474}\]
select a subset of replicas whose statistical features are as close as possible to those of the prior
Option 2: Monte Carlo to Hessian conversion \[\text{EPJ C75 (2015) 369}\]
sample the replicas on a discrete grid, select the eigenvectors of the ensuing covariance matrix

**Issue 3: PDF sets are not optimised for specific processes**
Tools for visualising sensitivity of PDFs to (hadronic) data
SMPDF \[\text{EPJ C76 (2016) 205}\] and PDFSense \[\text{arXiv:1803.02777}\]
select subset of the covariance matrix correlated to a given set of processes
perform single value decomposition on the covariance matrix and select dominant eigenvector
project out orthogonal subspace and iterate until desired accuracy reached
Input from Lattice QCD

Moments

Unpolarized moments
\( \langle x \rangle_{u^+ - d^+} \)
\( \langle x \rangle_{u^+} \)
\( \langle x \rangle_{d^+} \)
\( \langle x \rangle_{s^+} \)
\( \langle x \rangle_g \)
\( \mu^2 = Q^2 = 4 \text{ GeV}^2 \)

Quasi-PDFs

Various lattice QCD methods to determine PDF-related quantities

Need for a rigorous characterisation of the systematic uncertainties

Promising results, but still not competitive with global QCD analyses
4. Conclusions
Summary and outlook

1. The impact of the data
   - LHC data have now the dominant impact on PDFs (gluon and flavour separation) although collider-only fits are still not competitive
   - Methodology and theory must adapt accordingly

2. The (limits of the) methodology
   - statistical analysis tools necessary to cope with data accuracy
   - PDF uncertainties are faithful, but not optimised

3. The theory frontier
   - with sub-percent data uncertainties, theory uncertainties become dominant
   - resummation advantageous, electroweak corrections mandatory

4. Beyond the frontier
   - NNPDF http://nnpdf.mi.infn.it/
   - N3PDF http://n3pdf.mi.infn.it/
Summary and outlook

1. The impact of the data
   ▶ LHC data have now the dominant impact on PDFs (gluon and flavour separation) although collider-only fits are still not competitive
   ▶ Methodology and theory must adapt accordingly

2. The (limits of the) methodology
   ▶ statistical analysis tools necessary to cope with data accuracy
   ▶ PDF uncertainties are faithful, but not optimised

3. The theory frontier
   ▶ with sub-percent data uncertainties, theory uncertainties become dominant
   ▶ resummation advantageous, electroweak corrections mandatory

4. Beyond the frontier
   ▶ NNPDF http://nnpdf.mi.infn.it/
   ▶ N3PDF http://n3pdf.mi.infn.it/

Thank you