PARTON DISTRIBUTION FUNCTIONS

- Global analysis
- Practical applications

Pavel Nadolsky

Department of Physics
Southern Methodist University (Dallas, TX)

Lecture 3
June 2013

Stages of the PDF analysis

1. Select experimental data
2. Assemble all relevant theoretical cross sections and verify their mutual consistency
3. Choose the functional form for PDF parametrizations
4. Perform a fit
5. Make the new PDFs and their uncertainties available to end users

1. Selection of experimental data

- **Neutral-current ep DIS data from HERA are most extensive and precise** among all data sets
  - In addition, their systematic errors were reduced recently by cross calibration of H1 and ZEUS detectors
- However, by their nature they constrain only a limited number of PDF parameters
- Thus, two complementary approaches to the selection of the data are possible

Two strategies for selection of experimental data

DIS-based analyses ⇒ focus on the most precise (HERA DIS) data

- NC DIS, CC DIS, NC DIS jet, c and b production *(ABM, HERAPDF, JR)*

Global analyses (*CT10*, *MSTW'2008, NNPDF*)
⇒ focus on completeness, reliable flavor decomposition

- all HERA data + fixed-target DIS data
  - notably, CCFR and NuTeV νN DIS constraining $s(x,Q)$
- low-Q Drell-Yan (E605, E866), W lepton asymmetry, Z rapidity *(CT10, MSTW'08, NNPDF2)*
- Tevatron Run-2 and LHC jet production, $t\bar{t}$ production

2. Theoretical cross sections

Process	Number of QCD loops	Mass scheme*	
Neutral current	2	ZM	*Moch, Vermaseren, Vogt*
DIS	2	GM	*Riemersma, Harris, Smith, van Neerven*
			Buza, Matiounine, Smith, van Neerven
Charged current	2	ZM	*Moch, Vermaseren, Vogt*
DIS	1	GM	
$pN \xrightarrow{\gamma^*, W, Z} \ell\bar{\ell}^{(\prime)} X$	2	ZM	*Anastasiou, Dixon, Melnikov, Petriello*
$p\bar{p} \to jX$	1; 2 expected	ZM	
$ep \to jjX$	2	ZM	

*ZM/GM: zero-mass/general-mass approximation for c, b contributions

Although "NNLO" PDF fits include most of the NNLO hard cross sections, more work is needed to bring them to true NNLO accuracy

3. Requirements for PDF parametrizations

A. A valid set of $f_{a/p}(x,Q)$ must satisfy QCD sum rules

Valence sum rule

$$\int_0^1 \left[u(x,Q)-\bar{u}(x,Q)\right]dx = 2 \qquad \int_0^1 \left[d(x,Q)-\bar{d}(x,Q)\right]dx = 1$$

$$\int_0^1 \left[s(x,Q)-\bar{s}(x,Q)\right]dx = 0$$

A proton has net quantum numbers of 2 u quarks + 1 d quark

Momentum sum rule

$$[\text{proton}] \equiv \sum_{a=g,q,\bar{q}} \int_0^1 x\, f_{a/p}(x,Q)\, dx = 1$$

momenta of all partons must add up to the proton's momentum

Through this rule, normalization of $g(x,Q)$ is tied to the first moments of quark PDFs

3. Requirements for PDF parametrizations

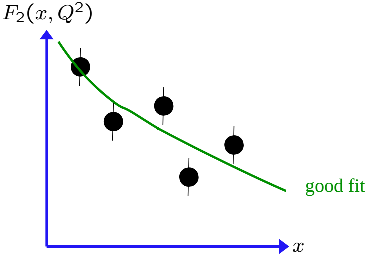
B. A valid PDF set must **not** produce unphysical predictions for observable quantities

Example

- Any concievable hadronic cross section σ must be non-negative: $\sigma \geq 0$
  - this is typically realized by requiring $f_{a/p}(x, Q) > 0$
- Any cross section asymmetry A must lie in the range $-1 \leq A \leq 1$
  - this constrains the range of allowed PDF pararametrizations

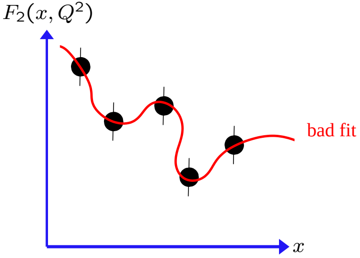
3. Requirements for PDF parametrizations

C. PDF parametrizations for $f_{a/p}(x,Q)$ must be "flexible just enough" to reach agreement with the data, without reproducing random fluctuations



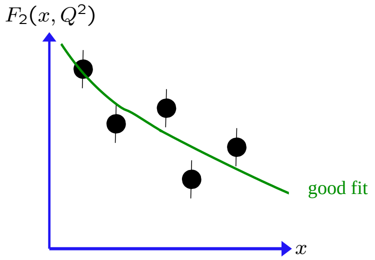
3. Requirements for PDF parametrizations

C. PDF parametrizations for $f_{a/p}(x,Q)$ must be "flexible just enough" to reach agreement with the data, without reproducing random fluctuations



3. Requirements for PDF parametrizations

C. PDF parametrizations for $f_{a/p}(x,Q)$ must be "flexible just enough" to reach agreement with the data, without reproducing random fluctuations



Traditional solution

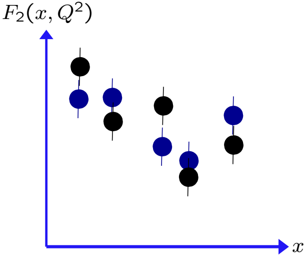
"Theoretically motivated" functions with a few parameters

$$f_{i/p}(x,Q_0) = a_0 x^{a_1}(1-x)^{a_2} \times F(x;\, a_3, a_4, \ldots)$$

- $x \to 0$: $f \propto x^{a_1}$ – Regge-like behavior
- $x \to 1$: $f \propto (1-x)^{a_2}$ – quark counting rules
- $F(a_3, a_4, \ldots)$ affects intermediate x; just a convenient functional form

3. Requirements for PDF parametrizations

C. PDF parametrizations for $f_{a/p}(x,Q)$ must be "flexible just enough" to reach agreement with the data, without reproducing random fluctuations



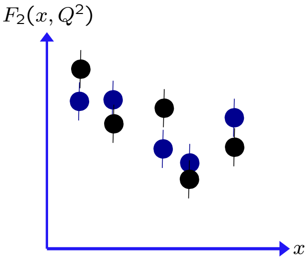
Radical solution

Neural Network PDF collaboration

- Generate N replicas of the experimental data, randomly scattered around the original data in accordance with the probability suggested by the experimental errors
- Divide the replicas into a fitting sample and control sample

3. Requirements for PDF parametrizations

C. PDF parametrizations for $f_{a/p}(x,Q)$ must be "flexible just enough" to reach agreement with the data, without reproducing random fluctuations



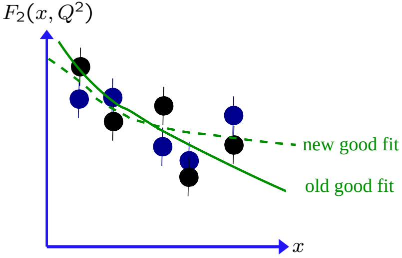
Radical solution

Neural Network PDF collaboration

- Parametrize $f_{a/p}(x,Q)$ by ultra-flexible functions — neural networks
- A statistical theorem states that any function can be approximated by a neural network with a sufficient number of nodes (in practice, of order 10)

3. Requirements for PDF parametrizations

C. PDF parametrizations for $f_{a/p}(x,Q)$ must be "flexible just enough" to reach agreement with the data, without reproducing random fluctuations



Radical solution

Neural Network PDF collaboration

- Fit the neural nets to the fitting sample, while demanding good agreement with the control sample
- Smoothness of $f_{a/p}(x,Q)$ is preserved, despite its nominal flexibility

4. Statistical aspects

J. Pumplin et al., JHEP 0207, 012 (2002), and references therein; J. Collins & J. Pumplin, hep-ph/0105207

Suppose there are N PDF parameters $\{a_i\}$, N_{exp} experiments, M_k data points and N_k correlated systematic errors in each experiment

Each systematic error is associated with a random parameter r_n, assumed to be distributed according to a Gaussian distribution with unit dispersion

The best external estimate of syst. errors corresponds to $\{r_n = 0\}$; but we must allow for $r_n \neq 0$

The most likely combination of $\{a\}$ and $\{r\}$ is found by minimizing

$$\chi^2 = \sum_{k=1}^{N_{exp}} w_k \chi_k^2$$

$w_k > 0$ are weights applied to emphasize or de-emphasize contributions from individual experiments (default: $w_k = 1$)

4. Statistical aspects

J. Pumplin et al., JHEP 0207, 012 (2002), and references therein; J. Collins & J. Pumplin, hep-ph/0105207

χ^2 for one experiment is

$$\chi_k^2 = \sum_{i=1}^{M_k} \frac{1}{\sigma_i^2} \left(D_i - T_i(\{a\}) - \sum_{n=1}^{R_k} r_n \beta_{ni} \right)^2 + \sum_{n=1}^{R_k} r_n^2$$

D_i and T_i are **data** and **theory** values at each point

$\sigma_i = \sqrt{\sigma_{stat}^2 + \sigma_{syst,uncor}^2}$ is the total statistical + systematical **uncorrelated** error

$\sum_n \beta_{ni} r_k$ are **correlated** systematic shifts

β_{ni} is the **correlation** matrix; is provided with the data or theoretical cross sections before the fit

$\sum_n r_n^2$ is the penalty for deviations of r_n from their expected values, $r_n = 0$

4. Statistical aspects

J. Pumplin et al., JHEP 0207, 012 (2002), and references therein; J. Collins & J. Pumplin, hep-ph/0105207

Each χ_k can be **analytically** minimized with respect to **the Gaussian** r_n, with the result

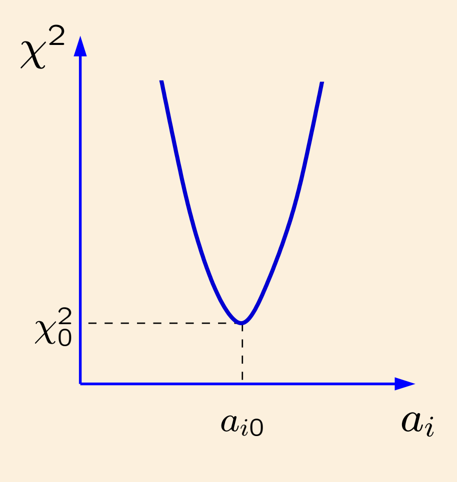
$$r_n\left(\{a\}\right) = \sum_{n'=1}^{R_k} (A^{-1})_{nn'} B_{n'}(\{a\})$$

$$A_{nn'} = \delta_{nn'} + \sum_{i=1}^{M_k} \frac{\beta_{ni}\beta_{n'i}}{\sigma_i^2}; \qquad B_n(\{a\}) = \sum_{i=1}^{M_k} \frac{\beta_{ni}(D_i - T_i)}{\sigma_i^2}$$

$$\chi_k^2 = \sum_{i=1}^{M_k} \frac{1}{\sigma_i^2}\left(D_{ki} - T_{ki}(\{a\})\right)^2 + \sum_{n,n'=1}^{R_k} B_n (A^{-1})_{nn'} B_{n'} \qquad (1)$$

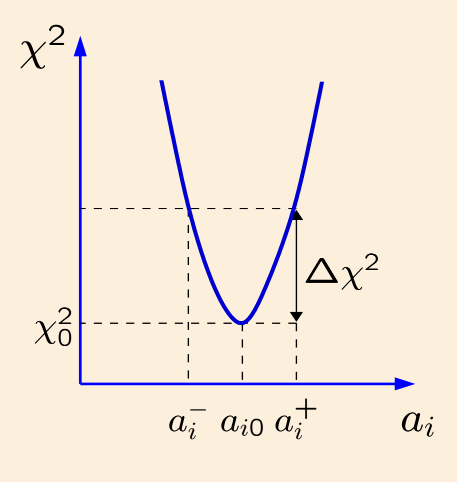
Numerical minimization of $\sum_k w_k \chi_k^2(a, r(a))$ (with χ_k from Eq. (1)) then establishes the region of acceptable $\{a\}$, which includes the largest possible variations of $\{a\}$ that are allowed by the systematic effects

Multi-dimensional error analysis



- Minimization of a likelihood function (χ^2) with respect to ~ 30 theoretical (mostly PDF) parameters $\{a_i\}$ and > 100 experimental systematical parameters

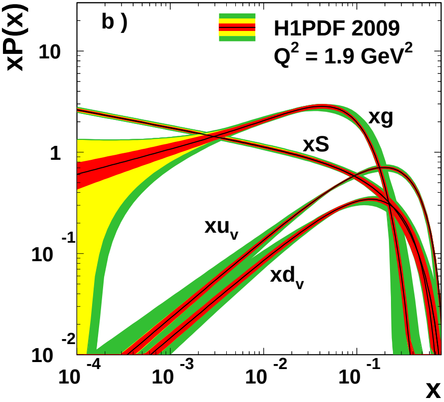
Multi-dimensional error analysis



- Establish a confidence region for $\{a_i\}$ for a given tolerated increase in χ^2
- In the ideal case of perfectly compatible Gaussian errors, 68% c.l. on a physical observable X corresponds to $\Delta\chi^2 = 1$ independently of the number N of PDF parameters

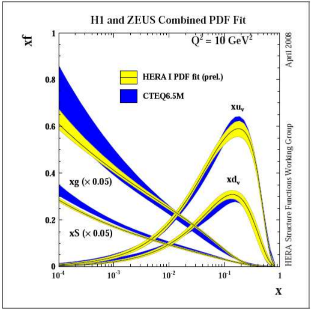
See, e.g., P. Bevington, K. Robinson, Data analysis and error reduction for the physical sciences

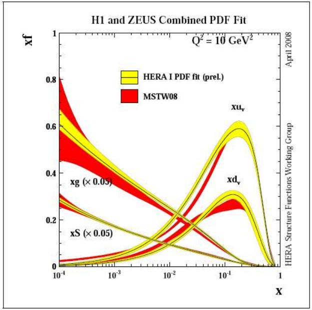
H1-2009 fit (arXiv:0904.3513)



- HERA-based fits are the closest to reproducing this ideal situation
- Example: the H1-2009 fit to the complete DIS data from HERA-1
- Color bands: experimental ($\Delta\chi^2 = 1$), theoretical, total uncertainty
- Heavy-flavor effects evaluated in GM-VFN scheme

HERAPDF0.1 set based on the combined H1+ZEUS data



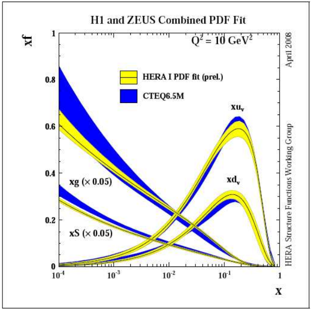


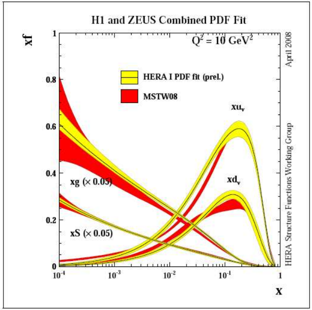
Updated HERAPDF0.2 fit was released this spring

The combined H1+ZEUS sample has a much smaller systematical uncertainty than the H1 and ZEUS samples individually

Nominally, very small uncertainty compared to CTEQ-MSTW-NNPDF!

HERAPDF0.1 set based on the combined H1+ZEUS data



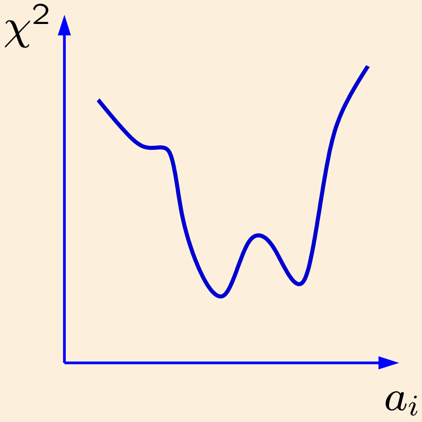


However:

- insufficient PDF flavor separation [neutral-current DIS probes only $4/9\,(u+\bar{u}+c+\bar{c})+1/9\,(d+\bar{d}+s+\bar{s})$]
- too rigid PDF parametrizations $\Rightarrow$ less flexibility to probe all allowed PDF behavior, notably at small x
- typical gluon forms, e.g., $g(x,Q_0)=Ax^B(1-x)^C\,(1+Dx)$, are ruled out by the Tevatron jet data *(Pumplin et al., arXiv:0904.2424)*

But if we combine the HERA data with the other experiments:

Multi-dimensional error analysis

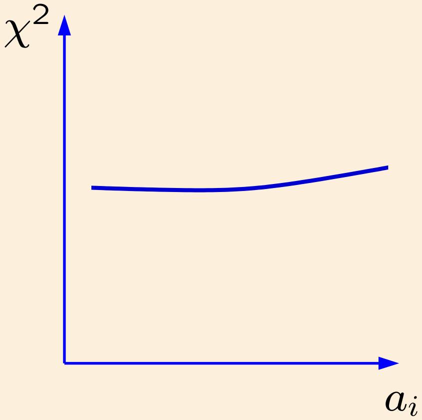


Pitfalls to avoid

- "Landscape"
  - disagreements between the experiments

In the worst situation, significant disagreements between M experimental data sets can produce up to $N \sim M!$ possible solutions for PDF's, with $N \sim 10^{500}$ reached for "only" about 200 data sets

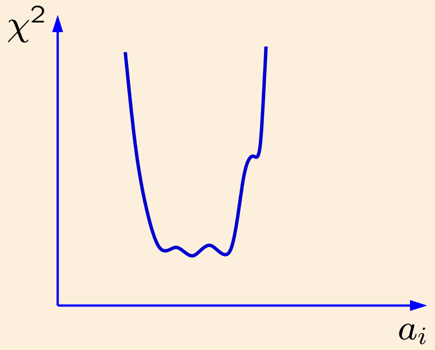
Multi-dimensional error analysis



Pitfalls to avoid

- Flat directions
  - unconstrained combinations of PDF parameters
  - dependence on free theoretical parameters, especially in the PDF parametrization
  - impossible to derive reliable PDF error sets

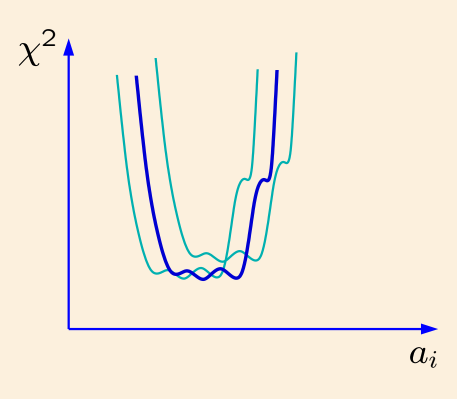
Multi-dimensional error analysis



The actual χ^2 function shows

- a well pronounced global minimum χ_0^2
- weak tensions between data sets in the vicinity of χ_0^2 (mini-landscape)
- some dependence on assumptions about flat directions

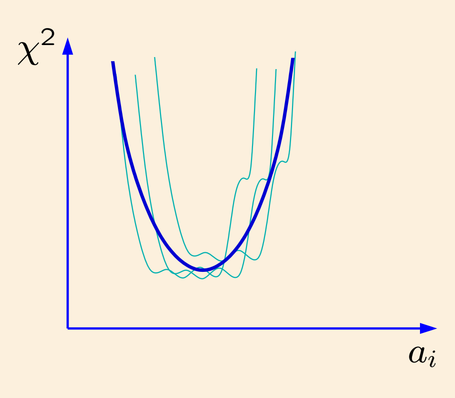
Multi-dimensional error analysis



The actual χ^2 function shows

- a well pronounced global minimum χ_0^2
- weak tensions between data sets in the vicinity of χ_0^2 (mini-landscape)
- some dependence on assumptions about flat directions

Multi-dimensional error analysis

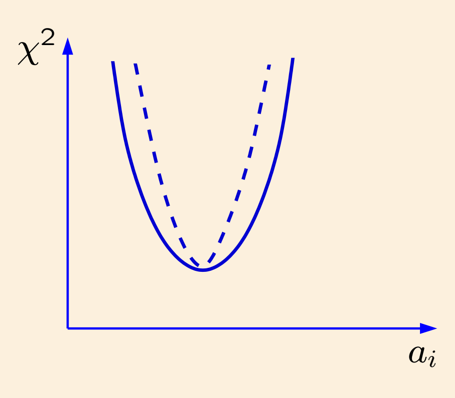


The actual χ^2 function shows

- a well pronounced global minimum χ^2_0
- weak tensions between data sets in the vicinity of χ^2_0 (mini-landscape)
- some dependence on assumptions about flat directions

The likelihood is approximately described by a quadratic χ^2 with a revised tolerance condition $\Delta\chi^2 \leq T^2$

Multi-dimensional error analysis

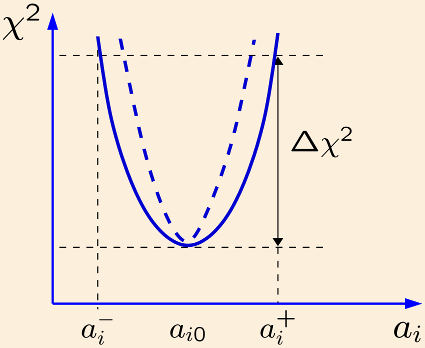


The actual χ^2 function shows

- a well pronounced global minimum χ_0^2
- weak tensions between data sets in the vicinity of χ_0^2 (mini-landscape)
- some dependence on assumptions about flat directions

The likelihood is approximately described by a quadratic χ^2 with a revised tolerance condition $\Delta\chi^2 \leq T^2$

Multi-dimensional error analysis



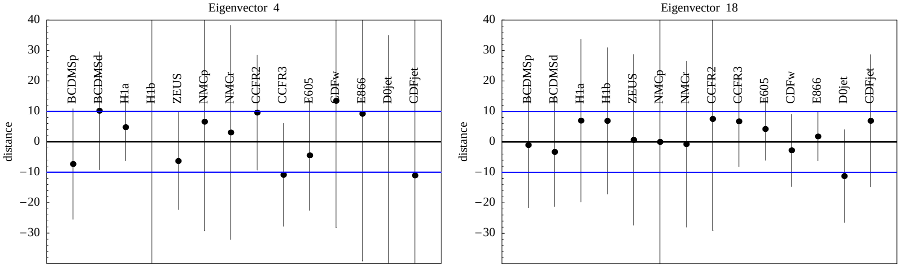
The actual χ^2 function shows

- a well pronounced global minimum χ_0^2
- weak tensions between data sets in the vicinity of χ_0^2 (mini-landscape)
- some dependence on assumptions about flat directions

The likelihood is approximately described by a quadratic χ^2 with a revised tolerance condition $\Delta\chi^2 \leq T^2$

CTEQ6 tolerance criterion (2001)

Acceptable values of PDF parameters must agree at ≈90% c.l. with all experiments included in the fit, *for a plausible range of assumptions about the PDF parametrization, scale dependence, experimental systematics, ...*

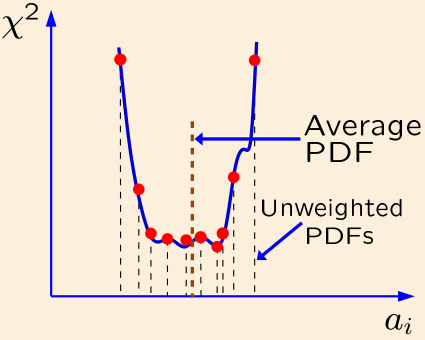


Can be crudely approximated (but does not have to) by assuming $T \approx 10$ for all PDF parameters

Refined variants of this criterion are applied in the MSTW'08 and CT10 analyses

Confidence intervals in global PDF analyses

Monte-Carlo sampling of the PDF parameter space



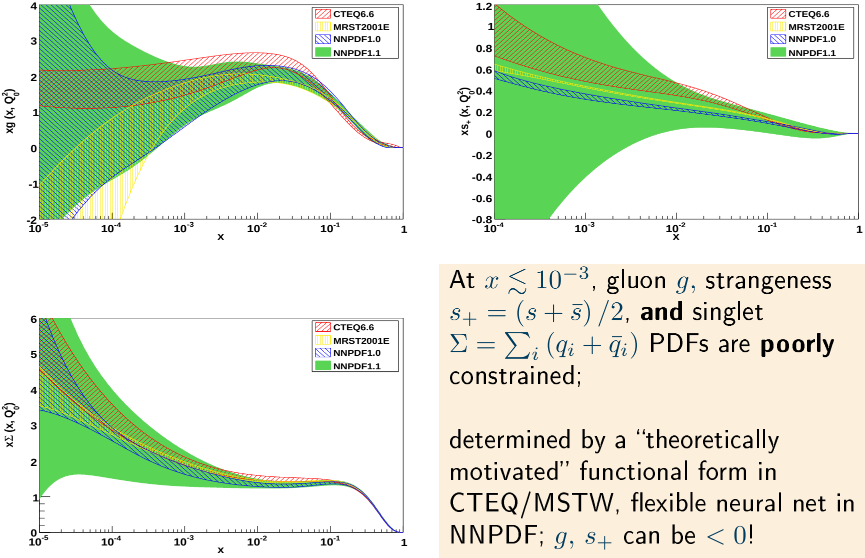
A very general approach that

- realizes stochastic sampling of the probability distribution

(Alekhin; Giele, Keller, Kosower; NNPDF)

- can parametrize PDF's by flexible neural networks (NNPDF)
- does not rely on smoothness of χ^2 or Gaussian approximations

NNPDF1.1 vs. other PDFs at $Q^2 = 2$ GeV^2 (arXiv:0811.2288)



At $x \lesssim 10^{-3}$, gluon g, strangeness $s_+ = (s+\bar{s})/2$, **and** singlet $\Sigma = \sum_i (q_i + \bar{q}_i)$ PDFs are **poorly** constrained;

determined by a "theoretically motivated" functional form in CTEQ/MSTW, flexible neural net in NNPDF; g, s_+ can be < 0!

PDF analysis
continues to be
an active field
open to young researchers!

Origin of differences between PDF sets

1. Corrections of wrong or outdated assumptions

lead to significant differences between new (≈post-2010) and old (≈pre-2010) PDF sets

- inclusion of NNLO QCD, heavy-quark hard scattering contributions
  - CT10 and MSTW'2008 PDFs implement complete heavy-quark treatment; previous PDFs are obsolete without it
- relaxation of ad hoc constraints on PDF parametrizations
- improved numerical approximations

Origin of differences between PDF sets

2. PDF uncertainty

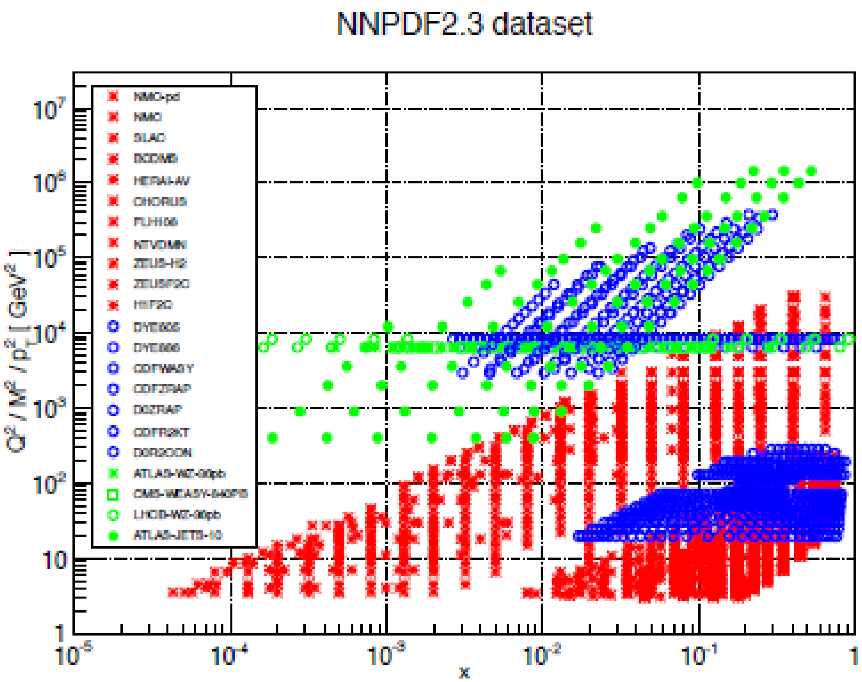
a range of allowed PDF shapes for plausible input assumptions, **partly** reflected by the PDF error band

is associated with

- the choice of fitted experiments
- experimental errors propagated into PDF's
- handling of inconsistencies between experiments
- choice of factorization scales, parametrizations for PDF's, higher-twist terms, nuclear effects,...

leads to non-negligible differences between the newest PDF sets

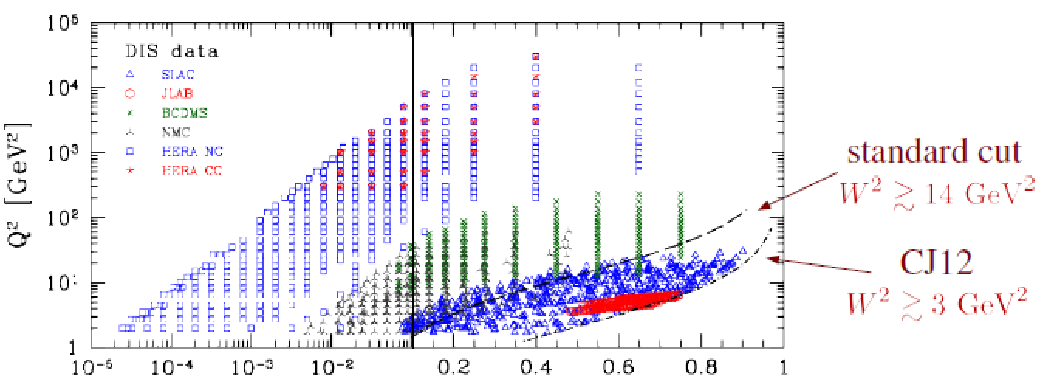
Selection of experiments, 2013



NNLO PDF analyses by ABM, CT, HERA, JR, MSTW, NNPDF groups

Precise PDF parametrizations for a variety of QCD applications

ABM: arXiv:1302.1516
CT10 NNLO: arXiv:1302.6246
NNPDF2.3: arXiv:1207.1303
MSTW'12: arXiv:1211.1215

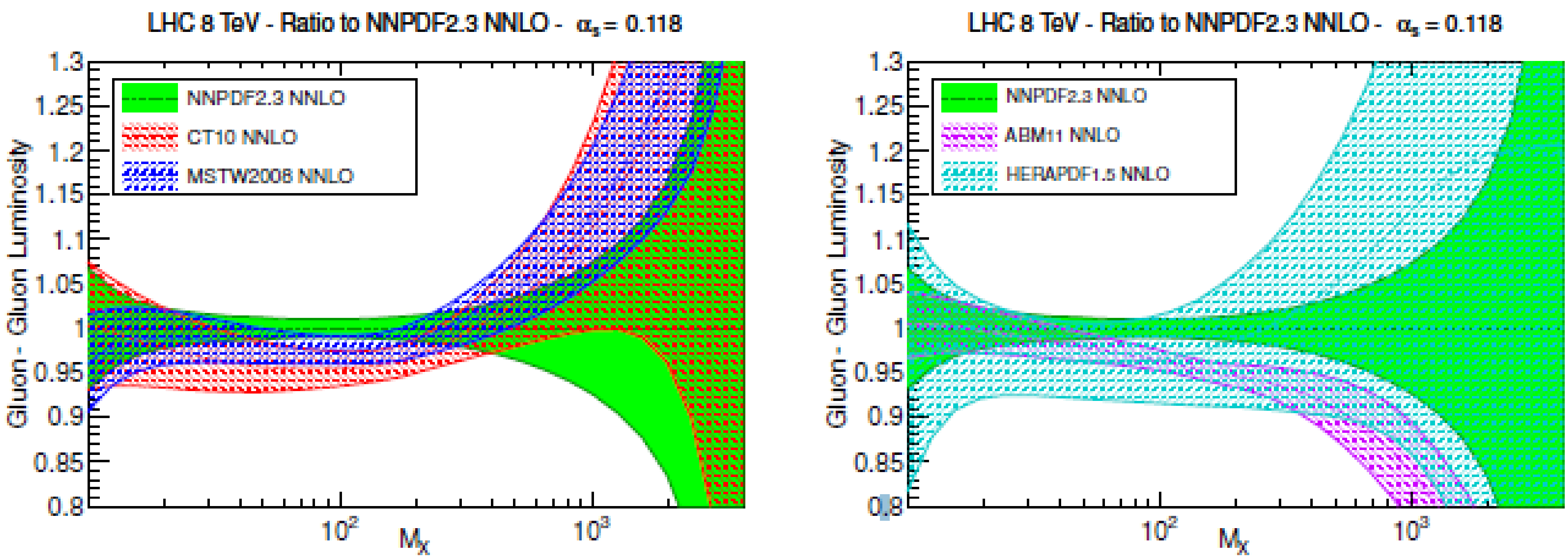


CTEQ-Jlab (CJ) (Accardi): NLO global QCD analysis of large-x, small-Q DIS region

important for fixed-target experiments, collider searches for TeV

2012 Comparison of unpolarized PDFs

[arXiv:1211.5142] - Benchmark study of different PDF determinations.
Detailed comparison at common α_S of the most up to date NNLO fits from the ABM, CT, HERAPDF, MSTW and NNPDF collaborations.

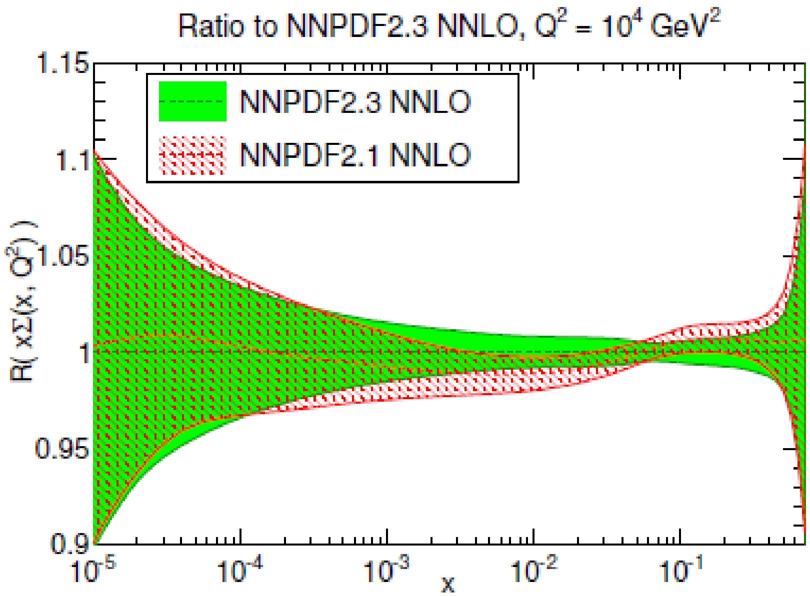


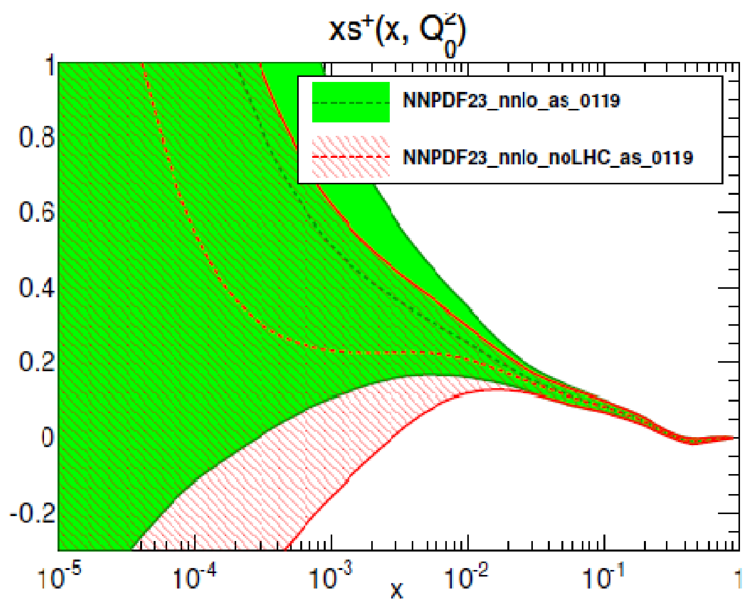
Reasonable agreement was found between CT, MSTW, NNPDF.

ABM softer large-x gluon and harder quarks.

Central values of HERAPDF1.5 NNLO agree with global fits, larger uncertainties due to reduced dataset.

LHC data⇒ new PDFs





NNPDF2.3: the first published PDF set that includes LHC 7 TeV data sets:

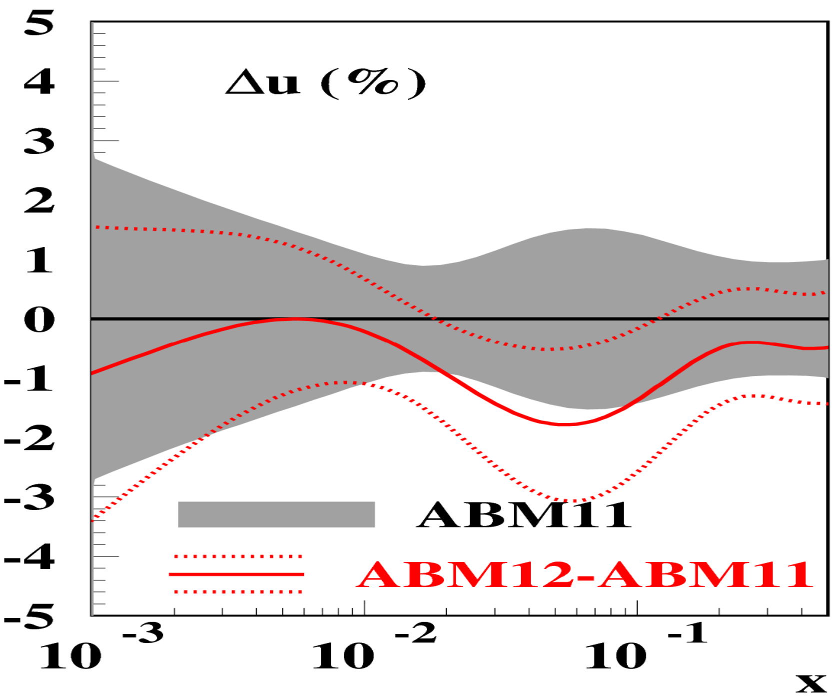
ATLAS inc. jets and $W^{\pm}/Z$ rapidity distributions,
LHCb $W^{\pm}$ rapidity distributions,
CMS W asymmetry

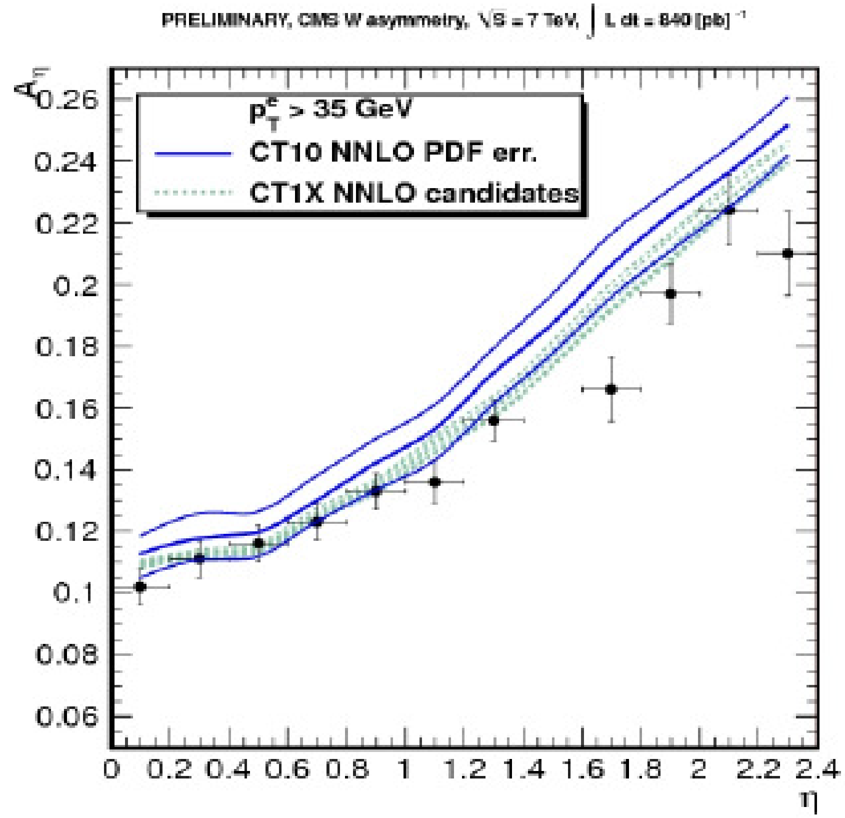
Some reduction in the PDF uncertainty compared to pre-LHC PDFs

Reduced error on strangeness PDFs

Large constraint for “collider only PDFs”

LHC data⇒ new PDFs





ABM: inclusion of ATLAS W/Z data modifies u and d PDFs

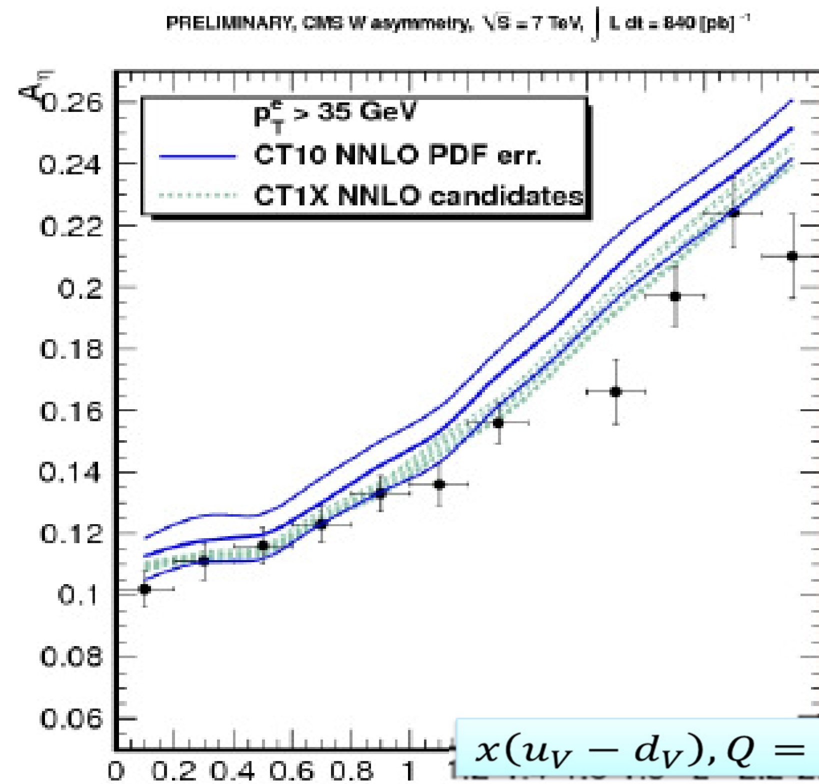
Preliminary fits CT1X and MMSTWW with LHC data

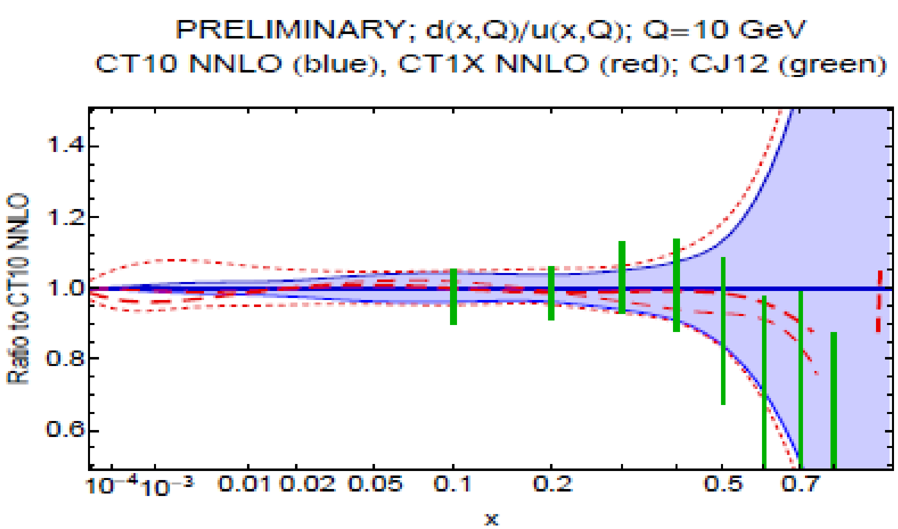
The CMS W asymmetry modifies separation between $u, \bar{u}, d, \bar{d}$ PDFs at $x \sim 0.01$ and d/u at $x > 0.1$

LHC data⇒ new PDFs

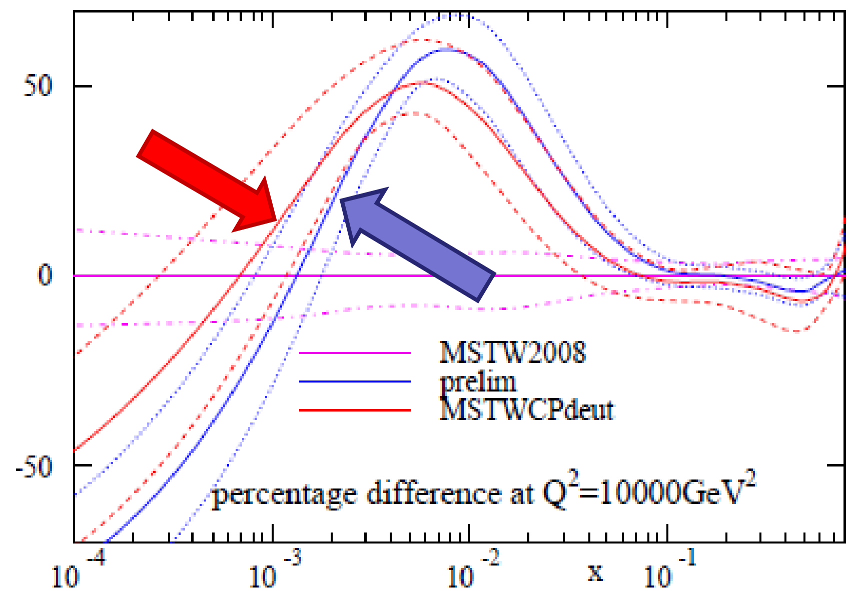
CT1X: modified d/u at x>0.1, increased uncertainties on d/u and $\bar{d}/\bar{u}$ at $x \to 0$

MSTW'2012: $d(x,Q)$ is modified across all x, now in agreement with CMS W asy data





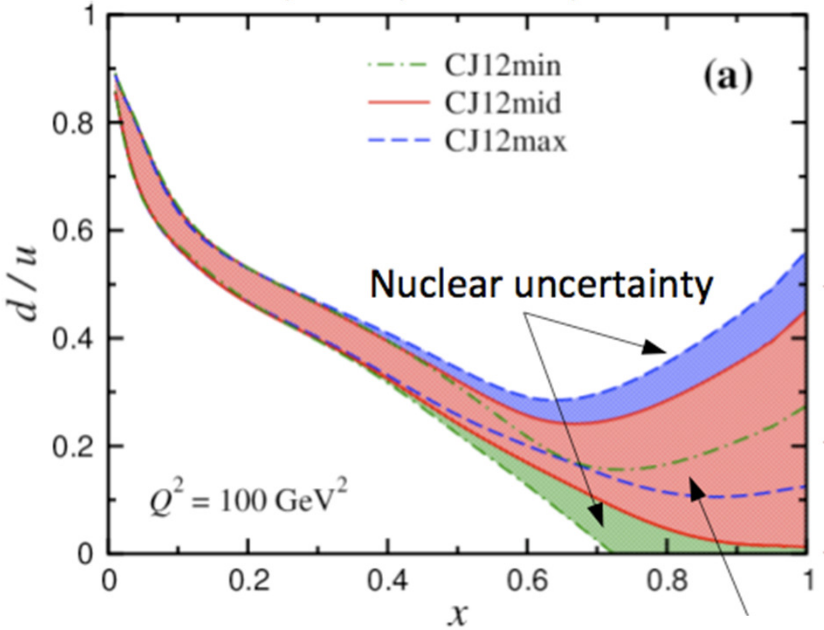
$x(u_V - d_V), Q = 100\ GeV$



Nuclear PDFs

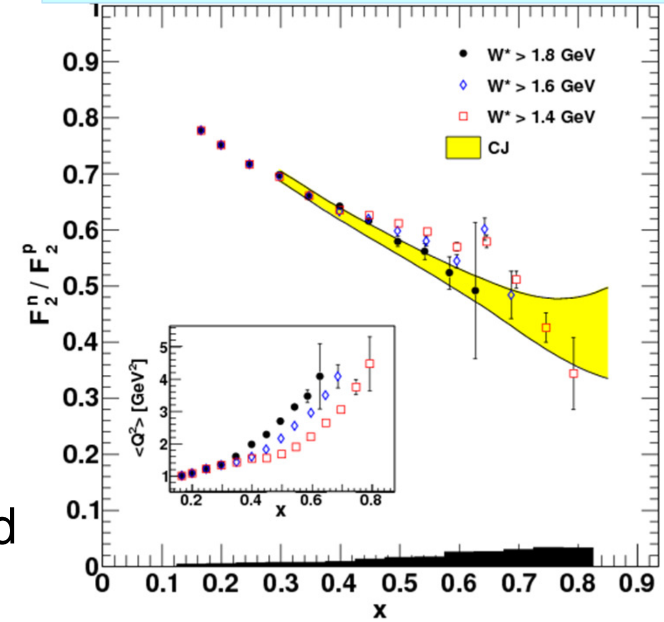
CJ 12 (Accardi), d(x)/u(x) at large x

Owens, Accardi, Melnitchouk, arXiv:1212.1702



- Systematic evaluation of deuteron/neutron corrections improves large x pdf precision
- Leverages extended data (JLab) data set

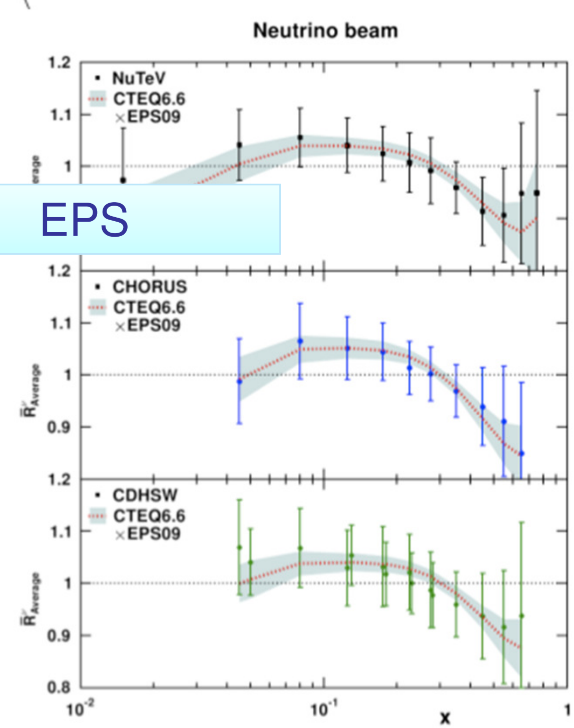
BONUS Experiment at Jlab



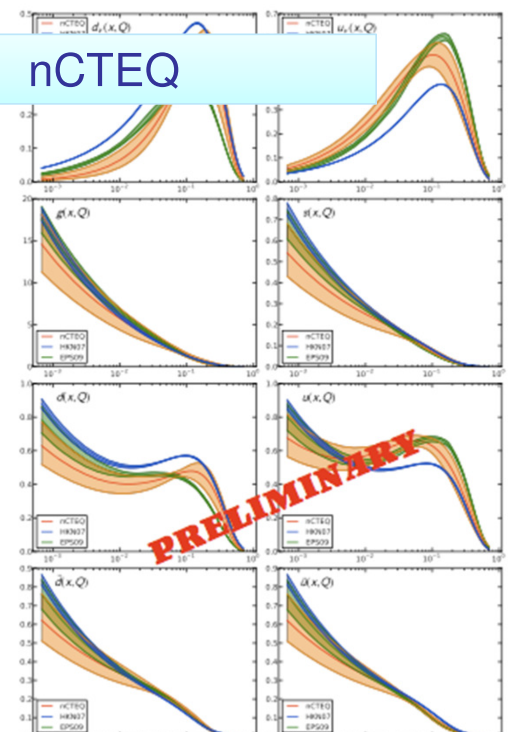
NUCLEAR pdfs

- Needed for ATLAS, RHIC, neutrino oscillations,…..
- Some controversy over including (or not!) neutrino data
- EIC, LHeC,… will help A LOT

EPS



nCTEQ

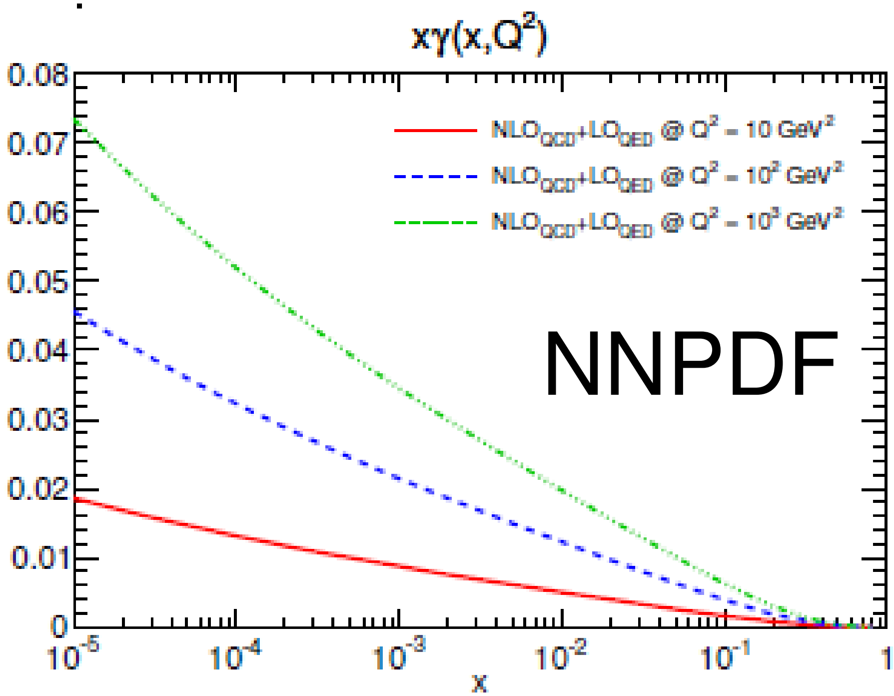


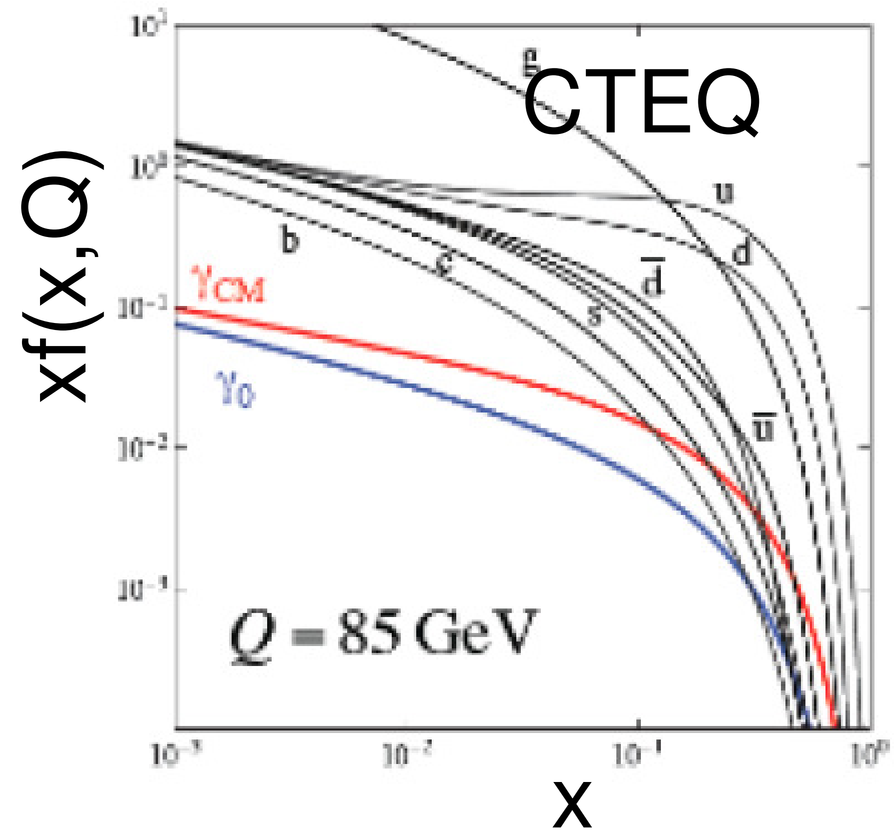
Photon PDFs: include γ as a new parton

Important for EW precision physics (W mass measurements), require deep revisions in the PDF analysis

The only existing QCD+QED PDF set is MRST'2004 QED, not updated for detailed studies

Preliminary NNLO QCD+LO QED PDFs presented by CTEQ and NNPDF groups, undergo validation





PDF whishlist at the LHC

Traditional

- Inclusive jets and dijets, central and forward: large-x quarks and gluons
- Inclusive W and Z production and asymmetries: quark flavor separation, strangeness

Relevant for M_W

New @ LHC

- Isolated photons, photons+jets: medium-x gluons
- W production with charm quarks: direct handle on strangeness
- W and Z production at high p_T: medium and small-x gluon
- Off resonance Drell-Yan and W production at high mass: quarks at large-x
- Low mass Drell-Yan production: small-x gluon
- Top quark cross-sections and differential distributions: large-x gluon

Relevant for M_W

Relevant for M_W

Relevant for M_W

Speculative

- Z+charm: intrinsic charm PDF
- Single top production: gluon and bottom PDFs
- Charmonium production: small-x gluon
- Open heavy quark production: gluon and intrinsic heavy flavor

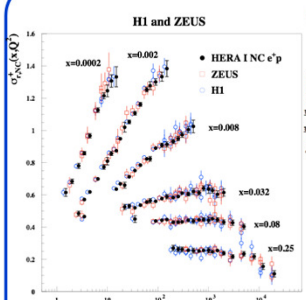
Relevant for M_W

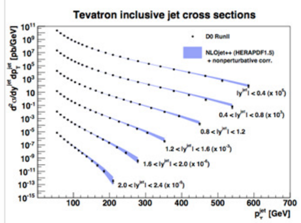
Benchmarking tools: HERAFitter

Developed at HERA, extended to LHC and theory groups

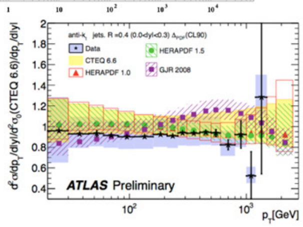
Study the impact of different data on PDFs and test different theory approaches

experimental input





experiments:
HERA, Tevatron, LHC, fixed target



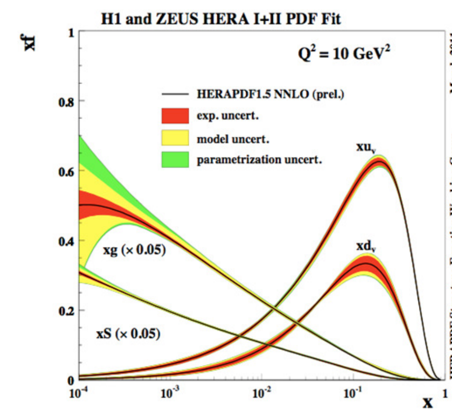
processes:
NC, CC DIS, jets, diffraction, heavy quarks (c,b,t)
Drell-Yan, W production

theoretical calculations/tools

Heavy quark schemes:	MSTW, CTEQ, ABM
Jets, W, Z production:	fastNLO, Applgrid
Top production	NNLO (Hathor)
QCD Evolution	DGLAP (QCDNUM)
	k_T factorisation
Alternative tools	NNPDF reweighting
Other models	Dipole model

+ Different error treatment models
+ Tools for data combination (HERAaverager)

HERAFitter



PDF or DPDF,...

$\alpha_S(M_Z), m_c, m_b, m_t, f_s,..$

Theory predictions

Benchmarking

Comparison of schemes

Open source code, available at https://www.herafitter.org/HERAFitter

Version 0.3.0 released in March 2013.