Parton Distribution Functions and their applications

Pavel Nadolsky

Department of Physics
Southern Methodist University (Dallas, TX)

Lecture 2
July 22, 2017

Recap, lecture 1

In the first lecture, we discussed the basic behavior of parton distribution functions, motivation for precision studies of PDFs, and key experiments from which the PDFs are determined

Today we will review practical issues associated with the determination and usage of PDFs: differences between the PDF sets, computation of PDF uncertainties, implementation of correlated errors, and combination of PDF sets.

Classes of PDFs

General-purpose

For (N)NLO calculations with
$N_f \leq 5$ active quark flavors

From several groups:
ABMP'16
HERA2.0
CT14 (→ 17p)
MMHT'14 (→ 16)
NNPDF3.0 (→3.1)
…

Specialized

For instance, for CT14:
CT14 LO
CT14 $N_f = 3, 4, 6$
CT14 HERA2
CT14 Intrinsic charm
CT14 QCD+QED
CT14 MC

ATLAS & CMS exploratory

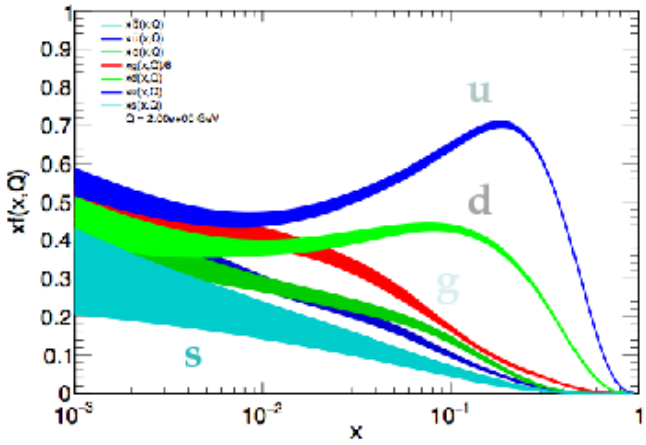
Combined

PDF4LHC'15=CT14+MMHT'14+NNPDF3.0

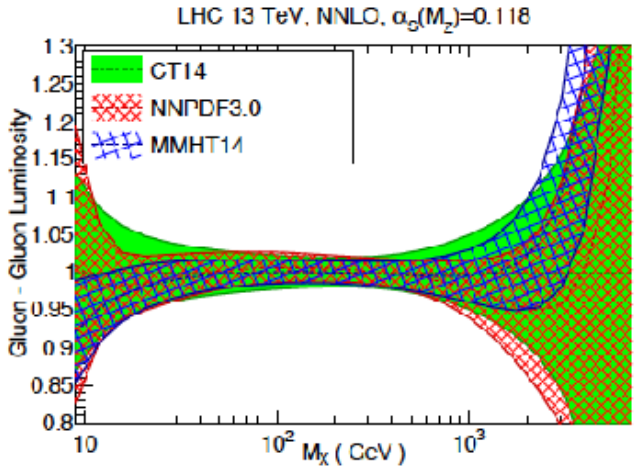
CT14 Parton distributions [1506.07443]

- Last major release on general-purpose PDFs, CT14 NNLO/NLO sets including alternative α_s series and n_f=3, 4, 6

CT14 NNLO PDFs



gluon-gluon luminosity

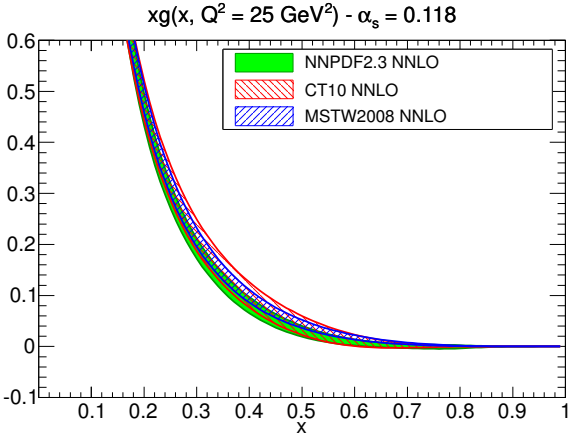


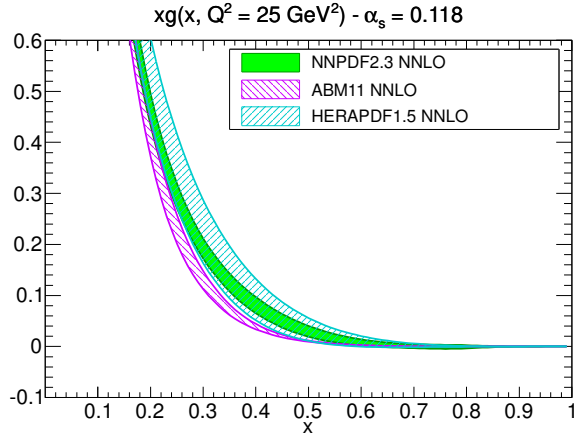
- old D0 W-electron asymmetry data superseded by the new one with full luminosity; combined HERA charm production, H1 FL data in NC DIS
- early LHC Run I data on W/Z charged lepton rapidity and asymmetry data; inclusive jet production from ATLAS and CMS
- more flexible parametrization for gluon, d/u at large-x, both d/u and dbar/ubar at small-x, 28 eigenvectors comparing to 25 for CT10

CT14 remains as our official sets for general purpose use

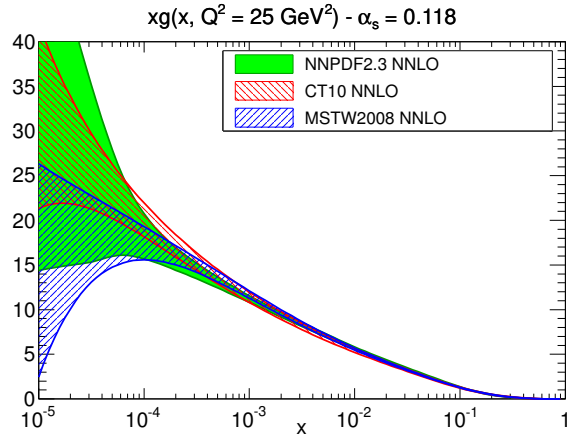
NNLO gluon PDF $xg(x, Q)$ from 5 groups

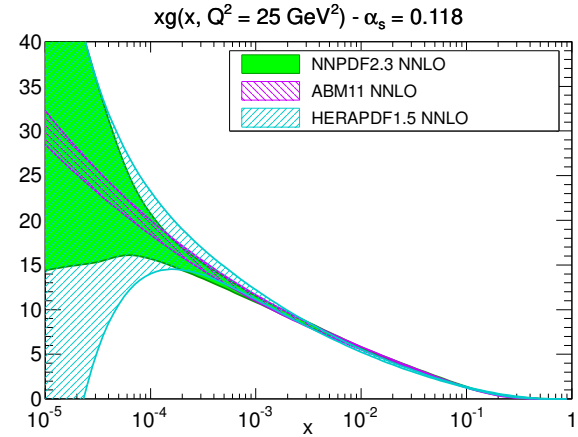
Linear x scale





Logarithmic x scale





R. Ball, et al., 1211.5142

CT10, MSTW'08, NNPDF2.3 PDFs are in good general agreement.
We see some differences with HERAPDF1.5 and ABM'12 sets.
Where do these differences come from?

Origin of differences between PDF sets

1. Corrections of wrong or outdated assumptions

lead to significant differences between new (≈post-2014) and old (≈pre-2014) PDF sets

- inclusion of NNLO QCD, heavy-quark hard scattering contributions
- 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

Stages of the PDF analysis

1. Select valid experimental data
2. Assemble most precise theoretical cross sections and verify their mutual consistency
3. Choose the functional form for PDF parametrizations
4. Implement a procedure to handle nuisance parameters (>200 sources of correlated experimental errors)
5. Perform a fit
6. Make the new PDFs and their uncertainties available to end users

1. Selection of experimental data

- **Neutral-current ep DIS data from HERA are STILL the 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

Global analyses *(~ABMP'16, CT14, MMHT'14, NNPDF3.1)* $\Rightarrow$ 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 and p_T
- Tevatron Run-2 and LHC jet production, $t\bar{t}$ production

Restricted analyses $\Rightarrow$ focus on the most precise data, esp. HERA DIS

- NC DIS, CC DIS, NC DIS jet, c and b production *(ABM'12, HERAPDF1.0,1.5, 2.0; ATLAS; CMS; JR)*

2. Available 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	2	GM	*Berger et al.*
$pN \xrightarrow{\gamma^*,W,Z} \ell\ell^{(\prime)}X$	2,3	ZM	*Anastasiou et al.; Boughezal et al; Gehrmann*
$p\bar{p} \to jX$	2	ZM	*Gehrmann et al.*
$ep \to jjX$	2	ZM	
$pp \to t\bar{t}X$	2	ZM	*Csakon, Mitov et al.*

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

Breath-taking progress in (N)NNLO calculations!

3. Requirements for PDF parametrizations

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

Valence sum rule

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

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

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

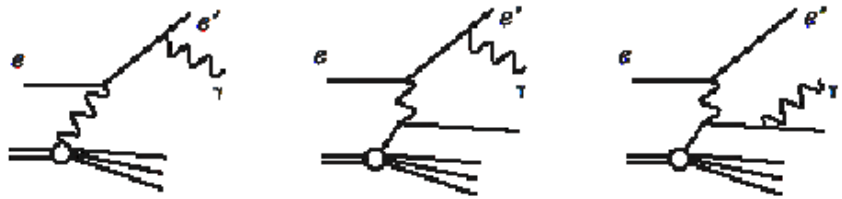
- Any concievable hadronic cross section σ must be non-negative: $\sigma \geq 0$
  - this is typically realized by requiring $f_{a/p}(x, Q) > 0$
- Any cross section asymmetry A must lie in the range $-1 \leq A \leq 1$
  - this constrains the range of allowed PDF pararametrizations
- Independent parametrizations for $s - \bar{s}$,γ, c may be difficult to constrain or separate from other factors

CT14 QED PDFs [1509.02905]

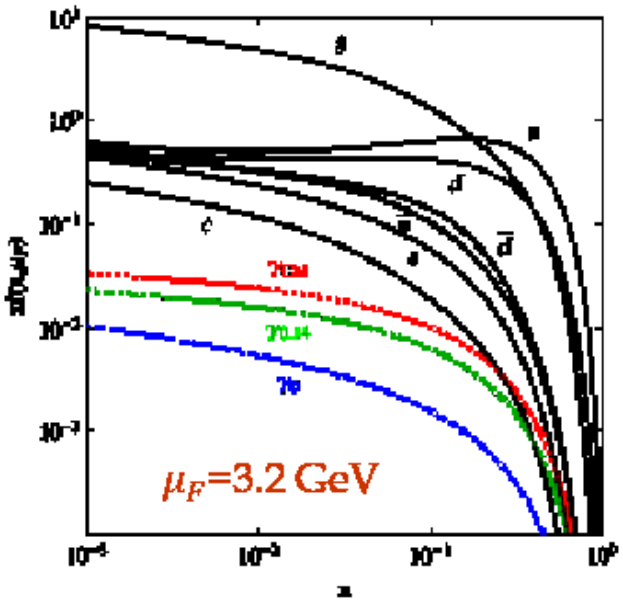
- CT14 set including photon PDF (NLO QCD+LO QED) based on radiative ansatz and with constraints from photon production in DIS

  - elastic part: equivalent photon approach, momentum frac. ~0.15%
  - inelastic part: radiative ansatz with one free parameter (momentum frac.), similar to MRST QED

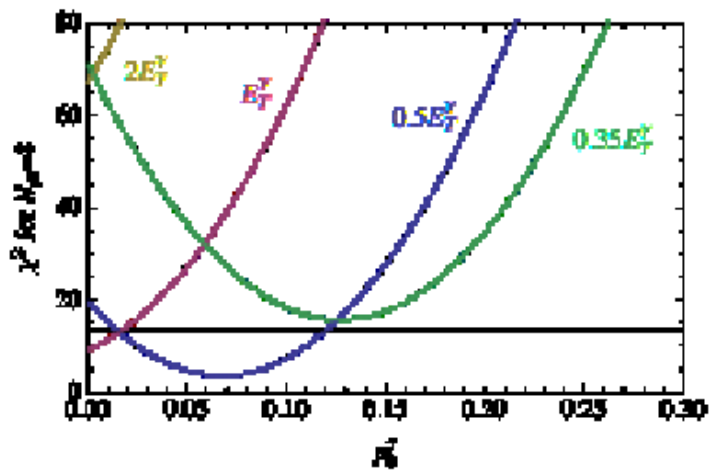
momentum frac. of inelastic part constrained from ZEUS isolated photon data



PDFs with photon (inelastic)



90% C.L. limit on momentum frac. ~0.14%



PDFs with fitted charm or intrinsic charm (IC)

Several studies conclude that IC may carry no more than 1% of the proton's momentum

Constraints depend on data selection (e.g., on whether the EMC F_2^c data are included) and methodology (CTEQ vs. NNPDF)

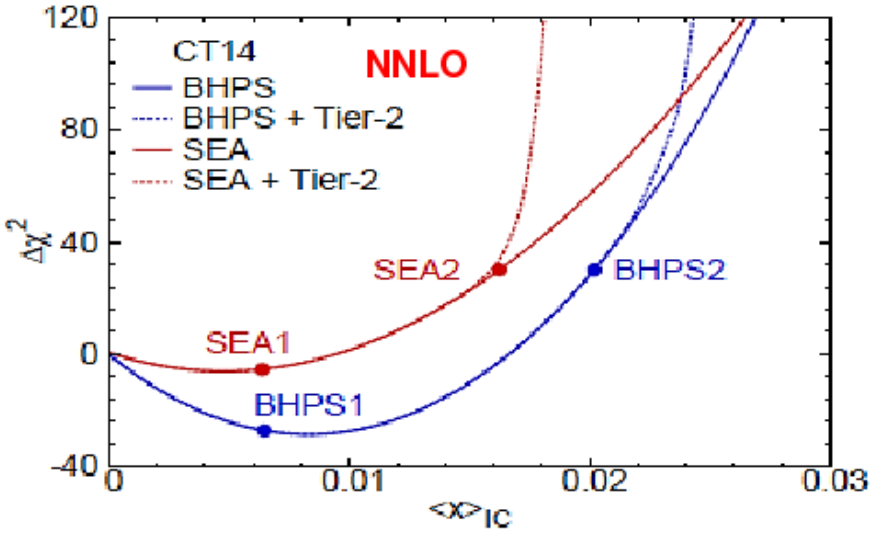
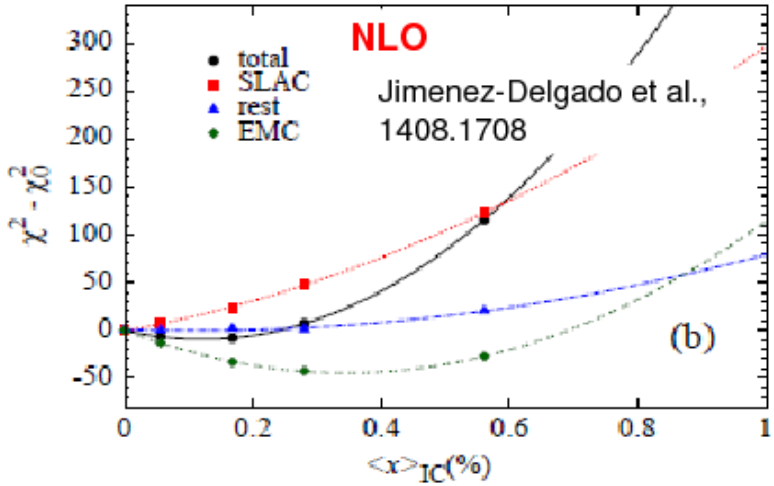
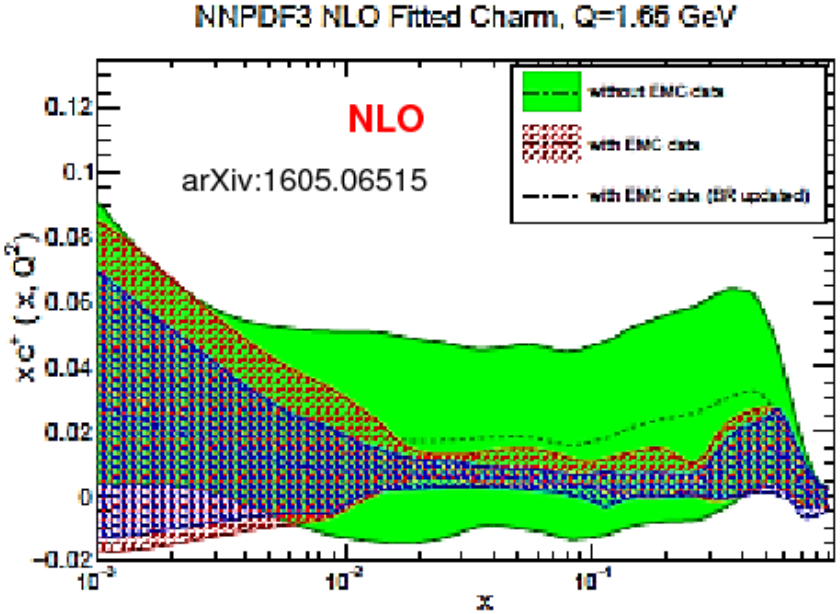


Figure 1: The $\Delta\chi^2$ versus the momentum fraction of charm $\langle x \rangle_{IC}$. **PoS DIS2015 (2015) 166**





PDF fits may include a ``fitted charm'' PDF

``Fitted charm'' = ``nonperturbative charm''
+ other (possibly not universal)
higher O(α_s) / higher power terms

QCD factorization theorem for DIS structure function $F(x,Q)$ [Collins, 1998]:

All α_s orders:
$$F(x,Q) = \sum_{a=0}^{N_f} \int_x^1 \frac{d\xi}{\xi} C_a\left(\frac{x}{\xi}, \frac{Q}{\mu}, \frac{m_c}{\mu}; \alpha(\mu)\right) f_{a/p}(\xi,\mu) + \mathcal{O}(\Lambda^2/m_c^2, \Lambda^2/Q^2).$$

The PDF fits implement this formula up to (N)NLO ($N_{ord} = 1$ or 2):

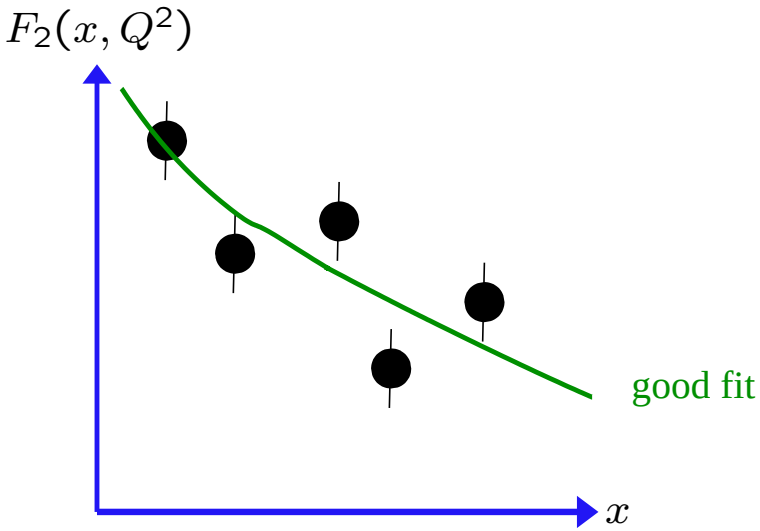
PDF fits:
$$F(x,Q) = \sum_{a=0}^{N_f} \int_x^1 \frac{d\xi}{\xi} C_a^{(N_{ord})}\left(\frac{x}{\xi}, \frac{Q}{\mu}, \frac{m_c}{\mu}; \alpha(\mu)\right) f_{a/p}^{(N_{ord})}(\xi,\mu).$$

The perturbative charm PDF component cancels at $Q \approx m_c$ up to a higher order

The 'fitted charm component' may approximate for missing terms of orders α_s^p with $p > N_{ord}$, or Λ^2/m_c^2, or Λ^2/Q^2

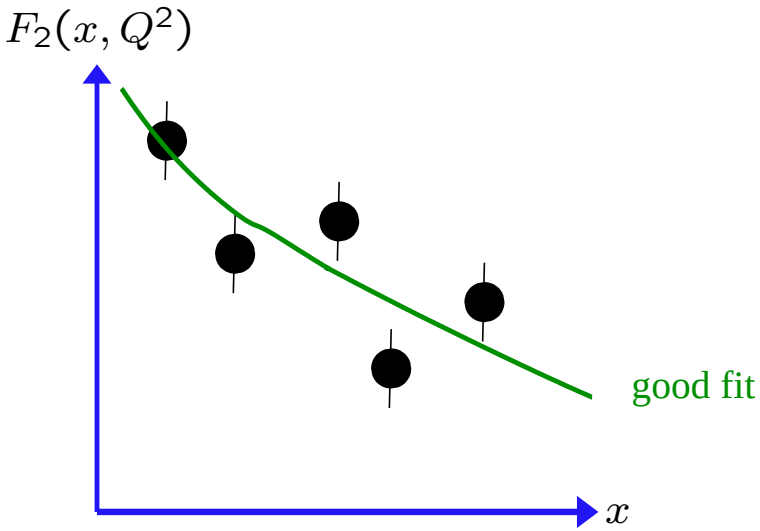
Requirements for PDF parametrizations

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



Requirements for PDF parametrizations

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



Traditional solution

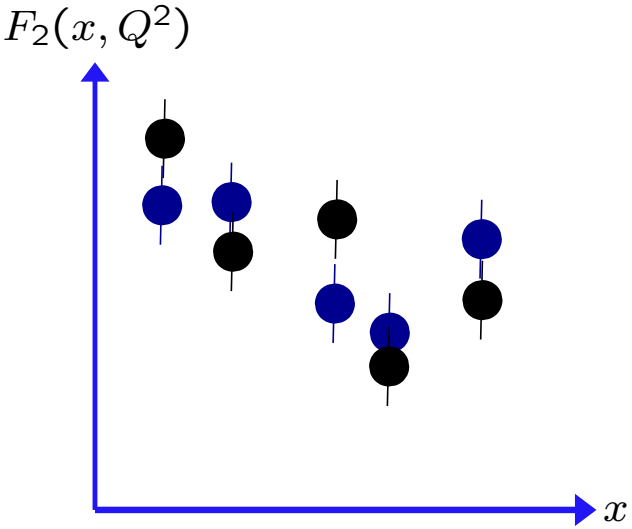
"Theoretically motivated" functions with a few parameters

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

- $x \to 0$: $f \propto x^{a_1}$ – Regge-like behavior
- $x \to 1$: $f \propto (1-x)^{a_2}$ – quark counting rules

Requirements for PDF parametrizations

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



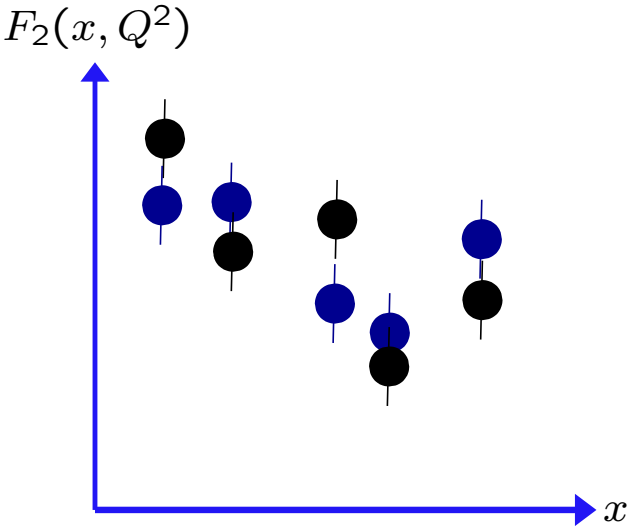
Radical solution

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

Requirements for PDF parametrizations

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



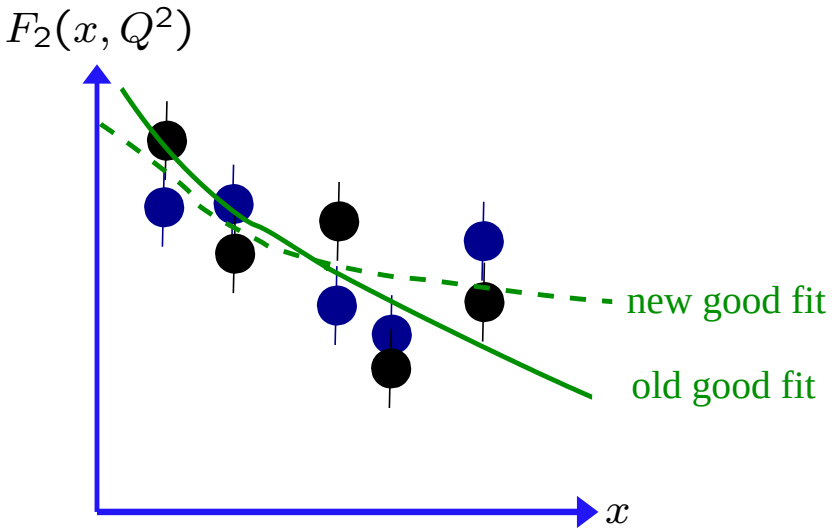
Radical solution

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)

Requirements for PDF parametrizations

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

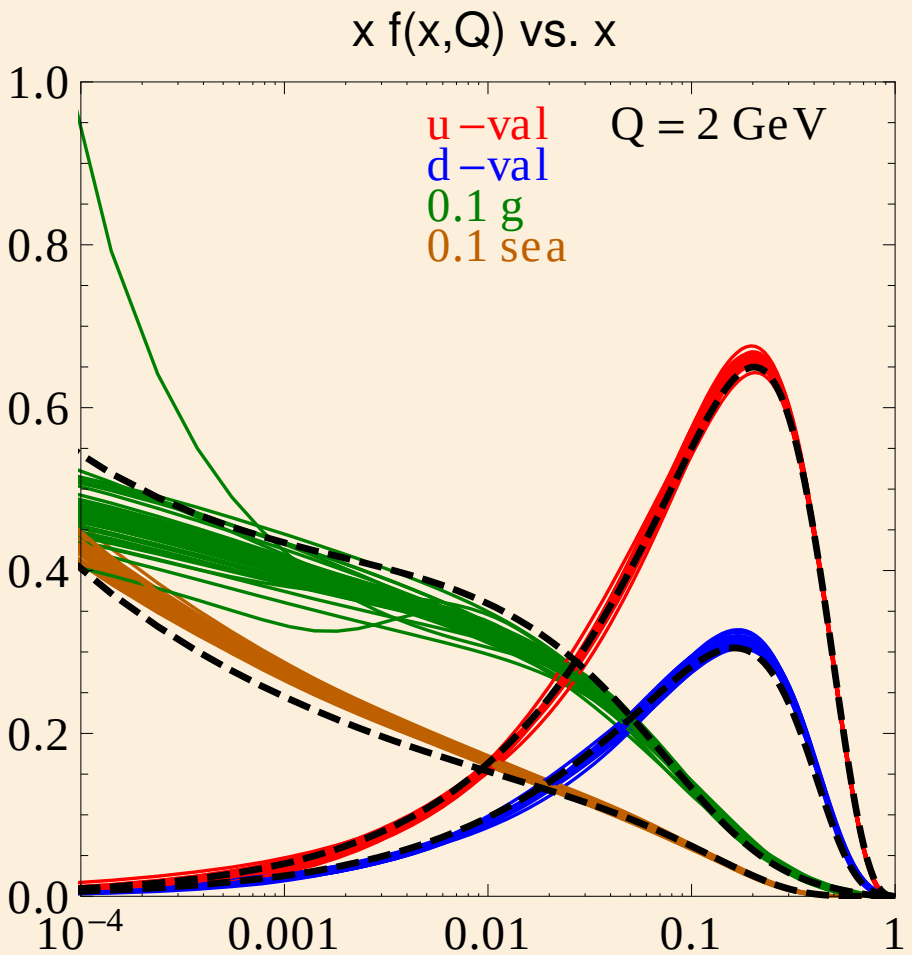


Radical solution

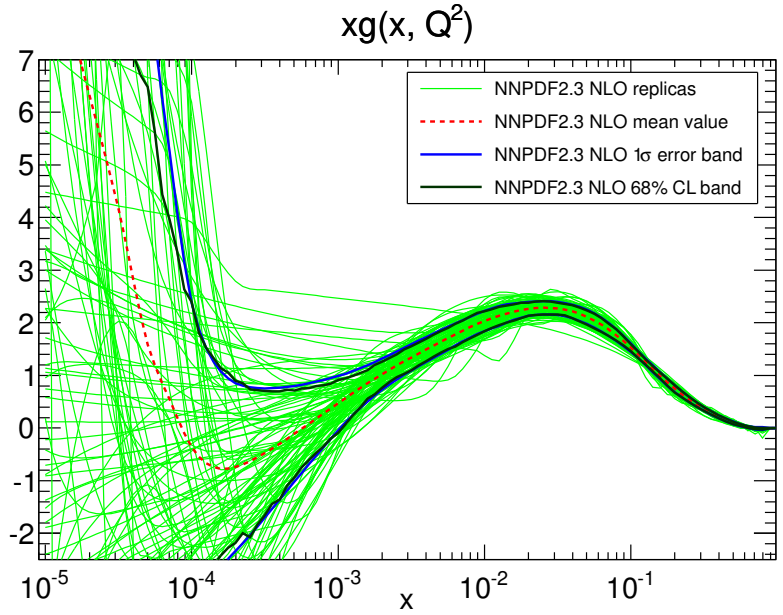
Neural Network PDF collaboration

- 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

Hessian error PDFs (CT10)



Neural network PDFs

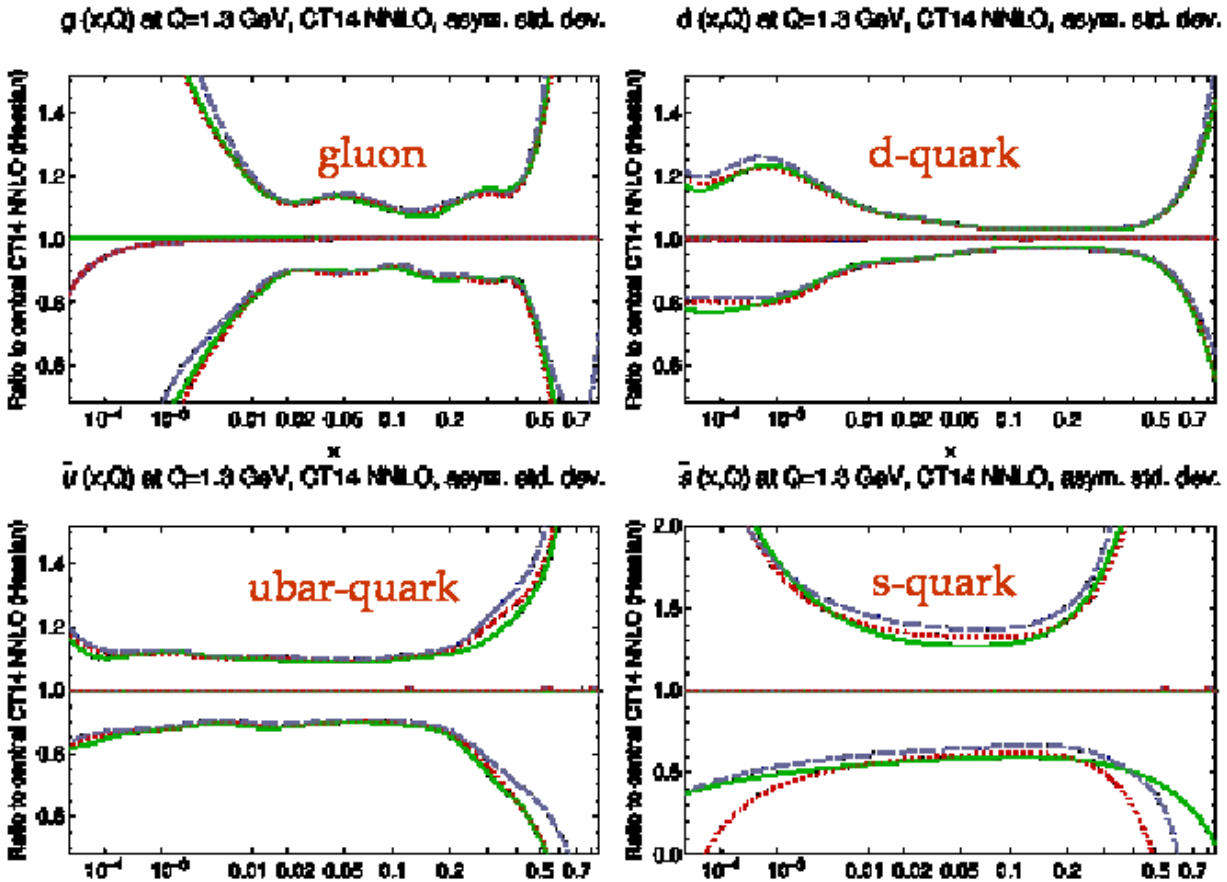


Several methods exist for converting MC error sets into Hessian error sets, and back

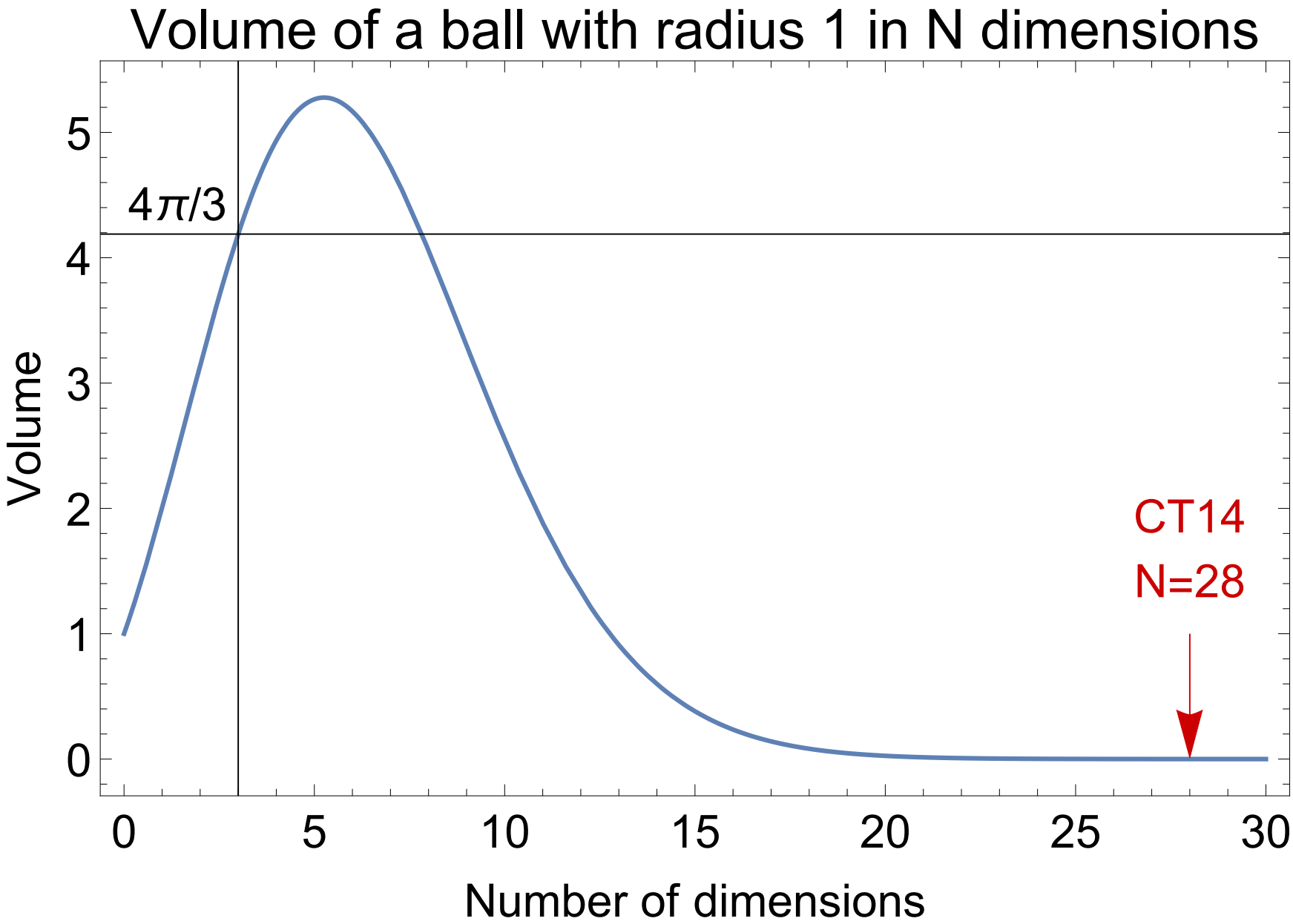
CT14 Monte-Carlo replicas [1607.06066]

- Two ensembles of CT14 MC replicas, Linear sampling(MC1), Log sampling(MC2), both with 1000 replicas

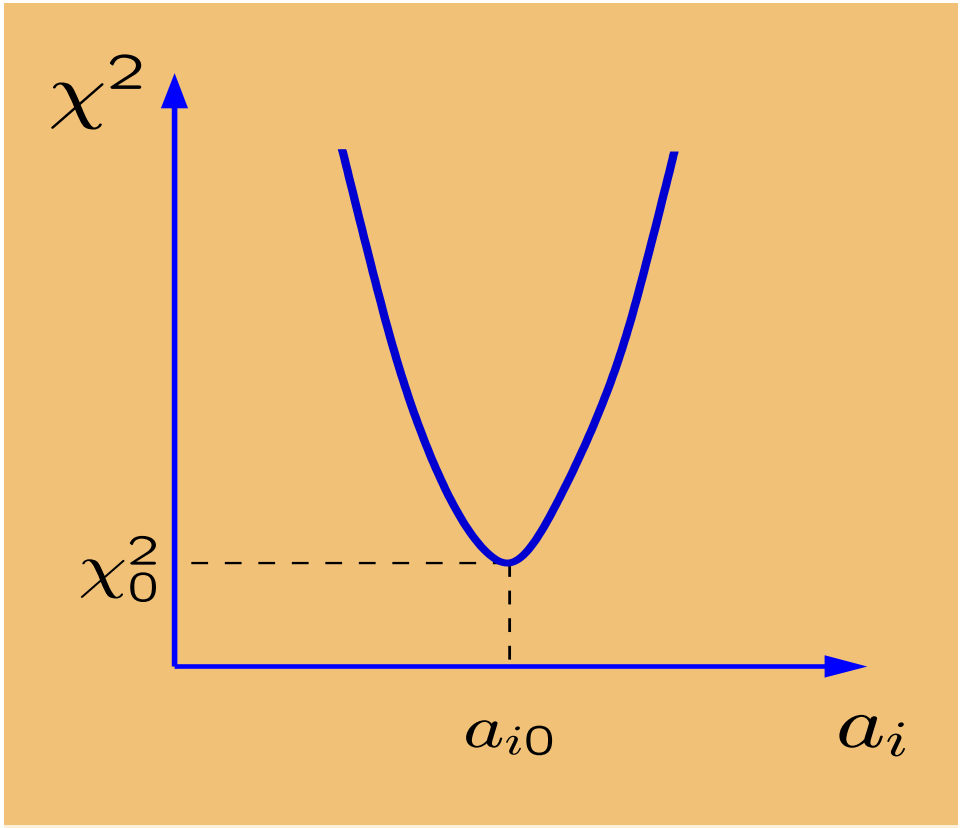
Hessian, MC1, MC2: solid, short-dashed, long-dashed



- reproducing statistical measures given by Hessian sets with small numbers of replicas; maintain positivity conditions as imposed in CT14

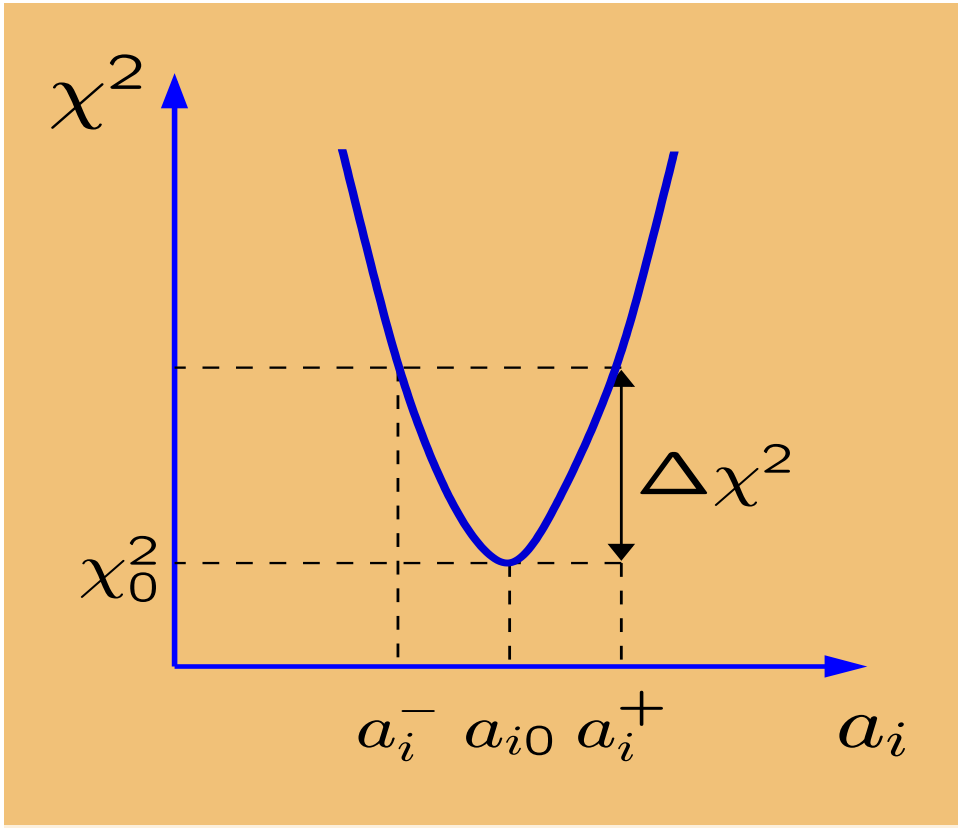

Volume of a ball with radius 1 in N dimensions
Volume
Number of dimensions
4π/3
CT14
N=28
0
1
2
3
4
5
0
5
10
15
20
25
30

Multi-dimensional error analysis



- Minimization of a likelihood function (χ^2) with respect to ~ 40 theoretical (mostly PDF) parameters $\{a_i\}$ and > 300 experimental systematical parameters

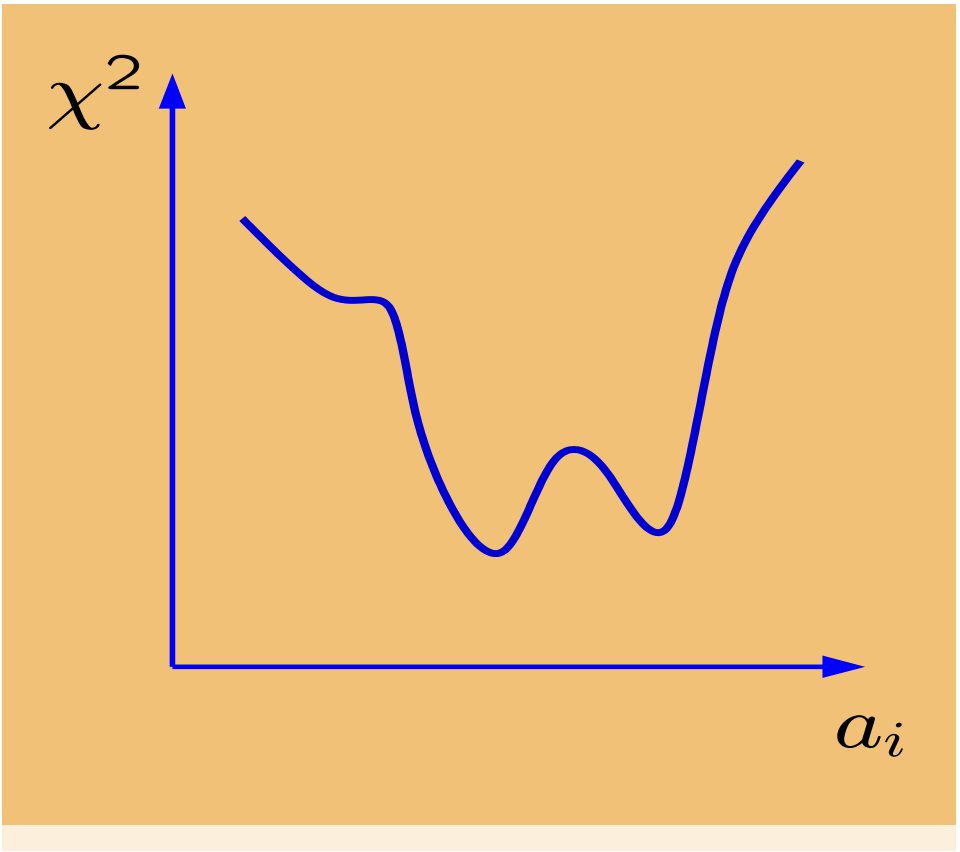
Multi-dimensional error analysis



- 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

Multi-dimensional error analysis

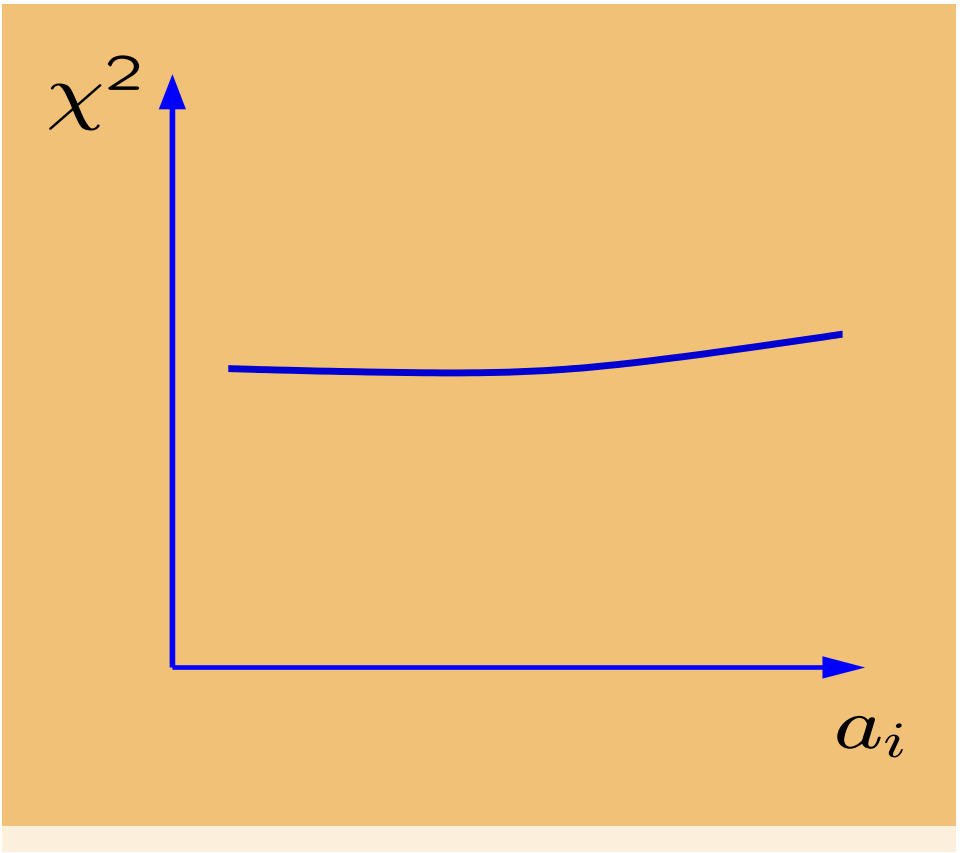


Pitfalls to avoid

- "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

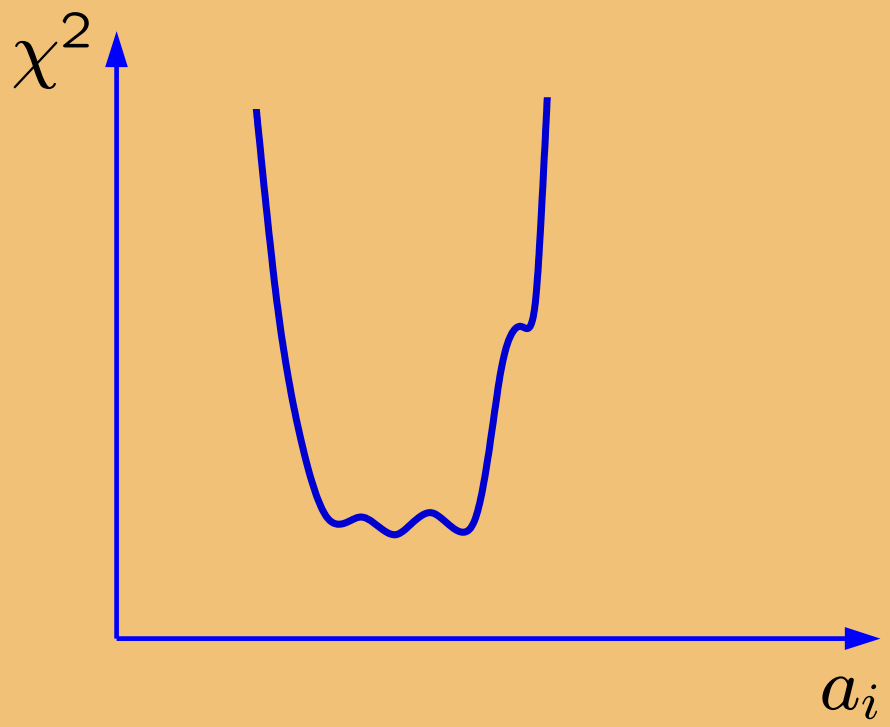
Multi-dimensional error analysis



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

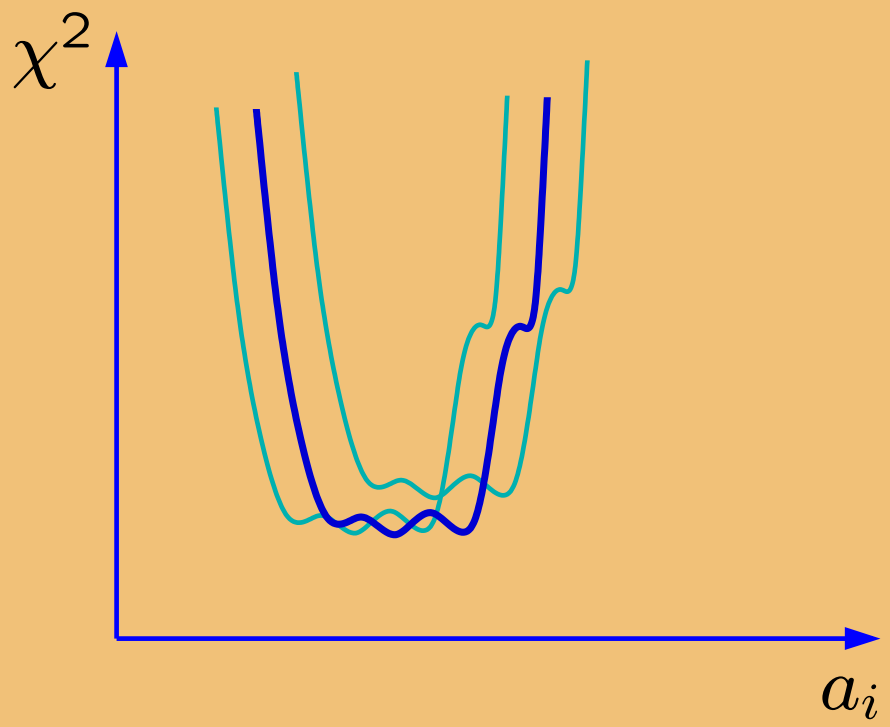
Multi-dimensional error analysis



The actual χ^2 function shows

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

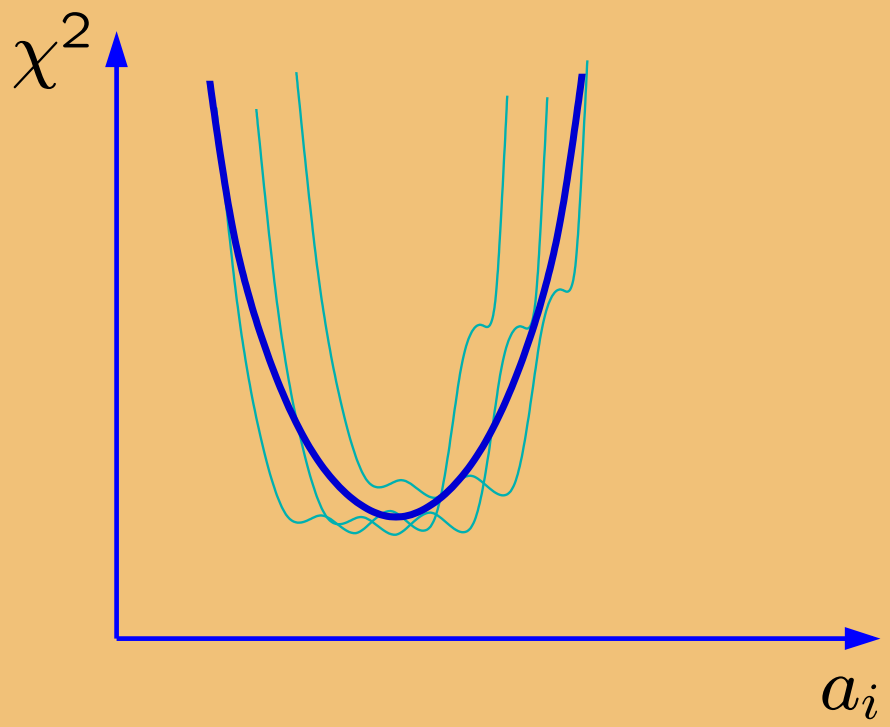
Multi-dimensional error analysis



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

Multi-dimensional error analysis

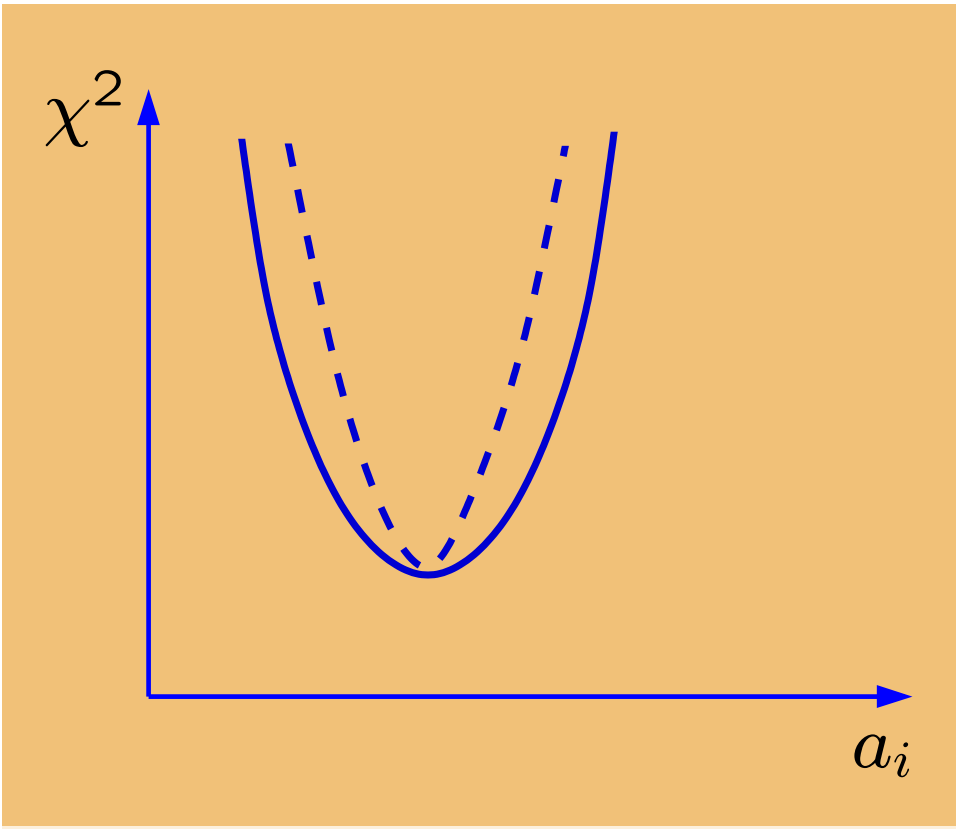


The actual χ^2 function shows

- 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

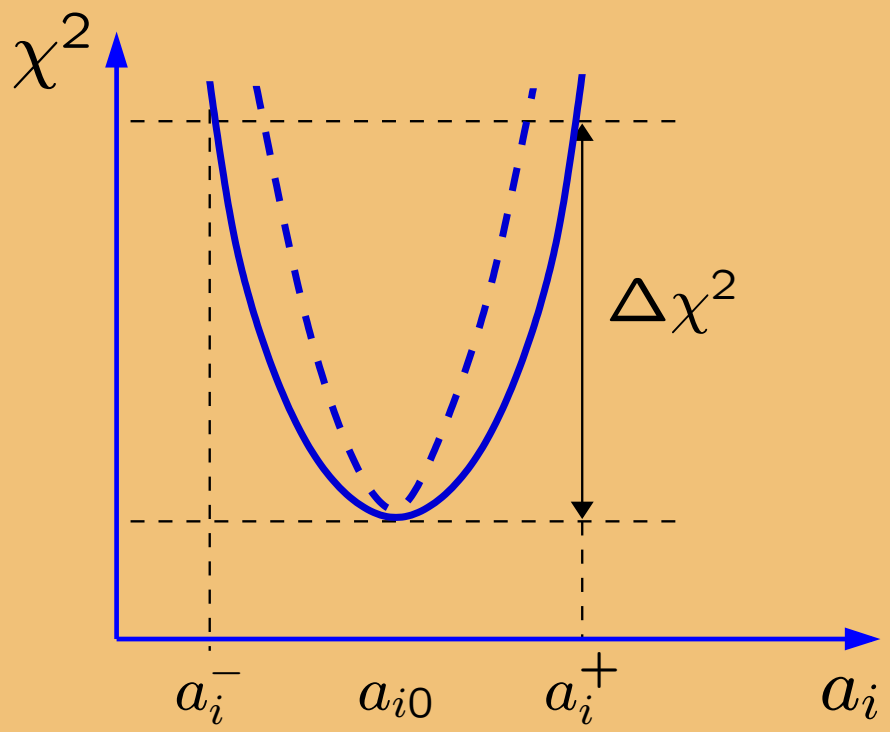


The actual χ^2 function shows

- a well pronounced global minimum χ^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



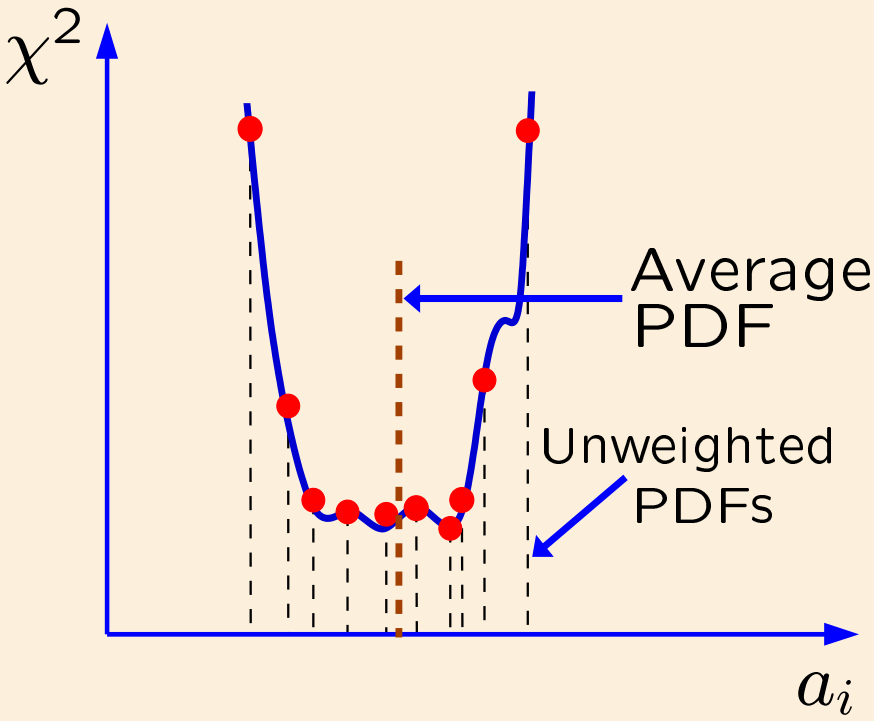
The actual χ^2 function shows

- a well pronounced global minimum χ^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 $\Delta\chi^2 \leq T^2$

Confidence intervals in a global PDF analysis

Monte-Carlo sampling of the PDF parameter space



A very general approach that

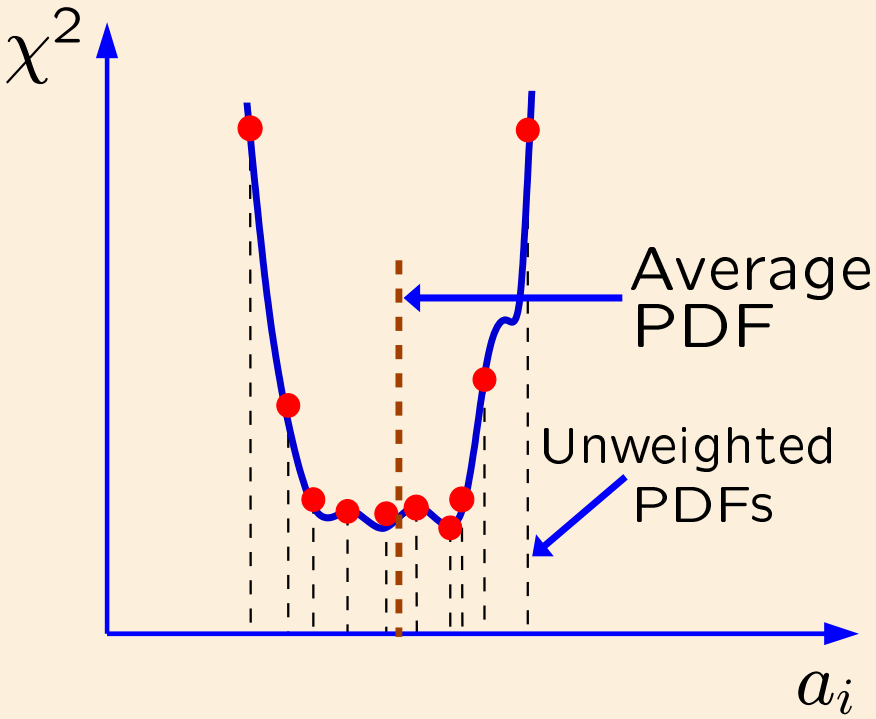
- realizes stochastic sampling of the probability distribution

(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

Confidence intervals in a global PDF analysis

Monte-Carlo sampling of the PDF parameter space

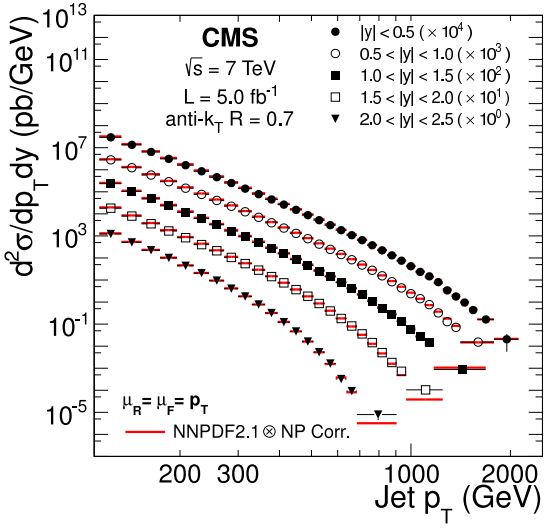


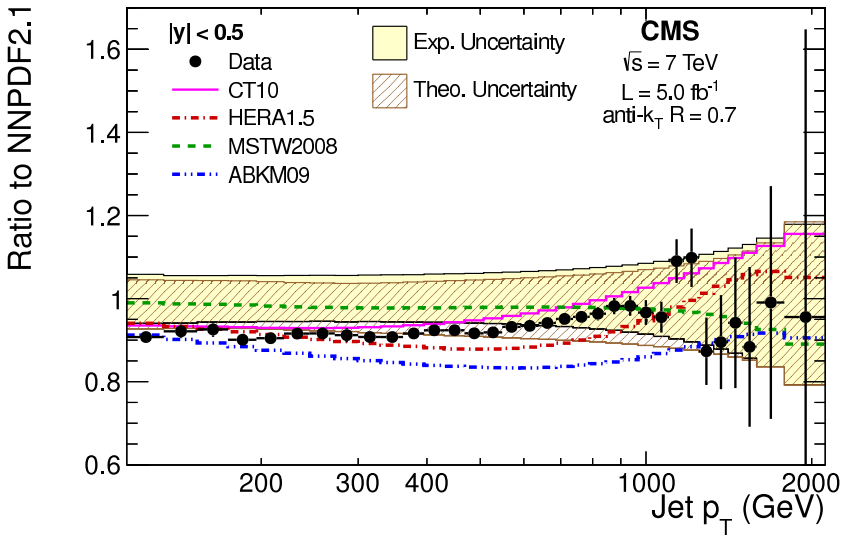
- Most individual replicas are **very bad** fits with $\Delta\chi^2 > 100$ *(1607.06066)*

  Recall that the $\Delta\chi^2 < 100$ ball has almost no volume in $N > 30$ dimensions

- A good fit is obtained by averaging over the replicas

Correlated systematic errors in the PDF analysis





At high luminosities, statistical errors in the collider data, such as jet production, are much smaller than correlated systematic errors.

The latest data sets are provided with many ($\sim 10 - 200$) sources of correlated errors.

Special methods are developed to include the effect of correlated systematic errors into the PDF uncertainties.

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 as 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 the 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

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

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

$\sigma_i = \sqrt{\sigma^2_{stat} + \sigma^2_{syst,uncor}}$ 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

Key formulas of the Hessian method

The Hessian method is used by CTXX PDFs to propagate the PDF uncertainty into QCD predictions.

It provides several simple formulas for computing the PDF uncertainties and PDF-induced correlations.

Tolerance hypersphere in the PDF space

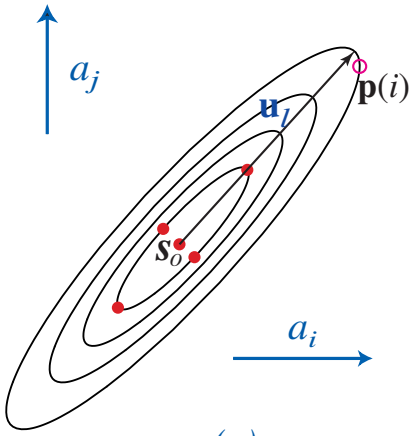
2-dim (i,j) rendition of N-dim (22) PDF parameter space

contours of constant χ^2_{global}

$\mathbf{u}_l$: *eigenvector in the l-direction*

$\mathbf{p}(i)$: *point of largest* a_i *with tolerance* T

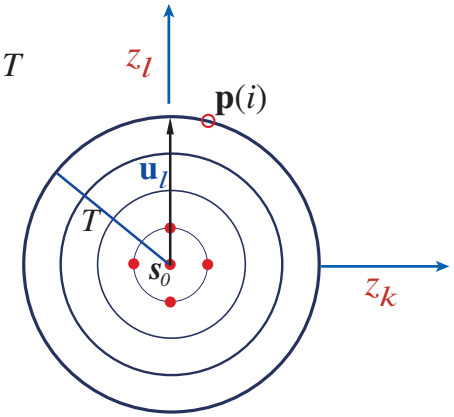
$\boldsymbol{s}_0$: *global minimum*



(a)
Original parameter basis

diagonalization and rescaling by the iterative method

- *Hessian eigenvector basis sets*

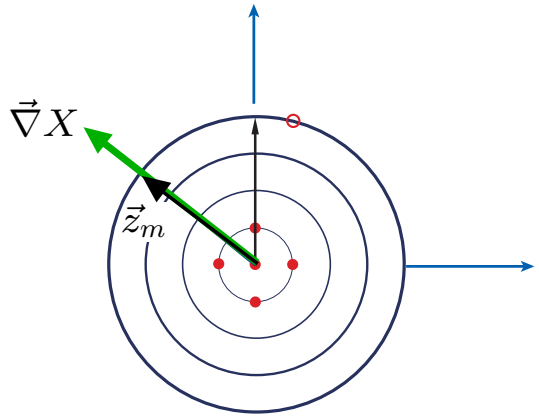


(b)
Orthonormal eigenvector basis

A hyperellipse $\Delta\chi^2 \leq T^2$ in space of N physical PDF parameters $\{a_i\}$ is mapped onto a filled hypersphere of radius T in space of N orthonormal PDF parameters $\{z_i\}$

Tolerance hypersphere in the PDF space

2-dim (i,j) rendition of N-dim (22) PDF parameter space



(b)
Orthonormal eigenvector basis

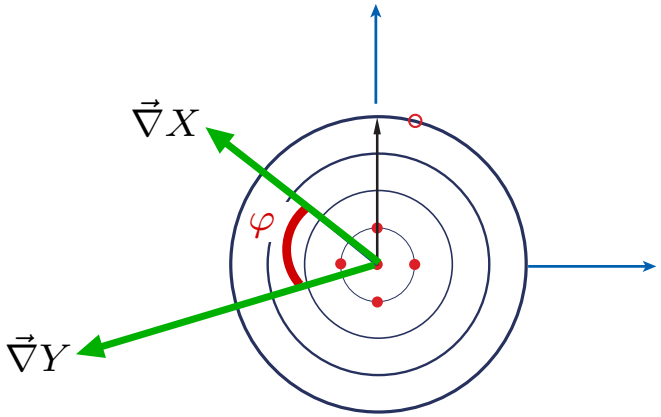
A symmetric PDF error for a physical observable X is given by

$$\Delta X = \vec{\nabla} X \cdot \vec{z}_m = \left|\vec{\nabla} X\right| = \frac{1}{2}\sqrt{\sum_{i=1}^{N}\left(X_i^{(+)} - X_i^{(-)}\right)^2}$$

Asymmetric PDF errors can be also computed.

Tolerance hypersphere in the PDF space

2-dim (i,j) rendition of N-dim (22) PDF parameter space



(b)
Orthonormal eigenvector basis

Correlation cosine for observables X and Y:

$$\cos\varphi = \frac{\vec{\nabla}X \cdot \vec{\nabla}Y}{\Delta X \Delta Y} = \frac{1}{4\Delta X\, \Delta Y} \sum_{i=1}^{N} \left(X_i^{(+)} - X_i^{(-)}\right)\left(Y_i^{(+)} - Y_i^{(-)}\right)$$

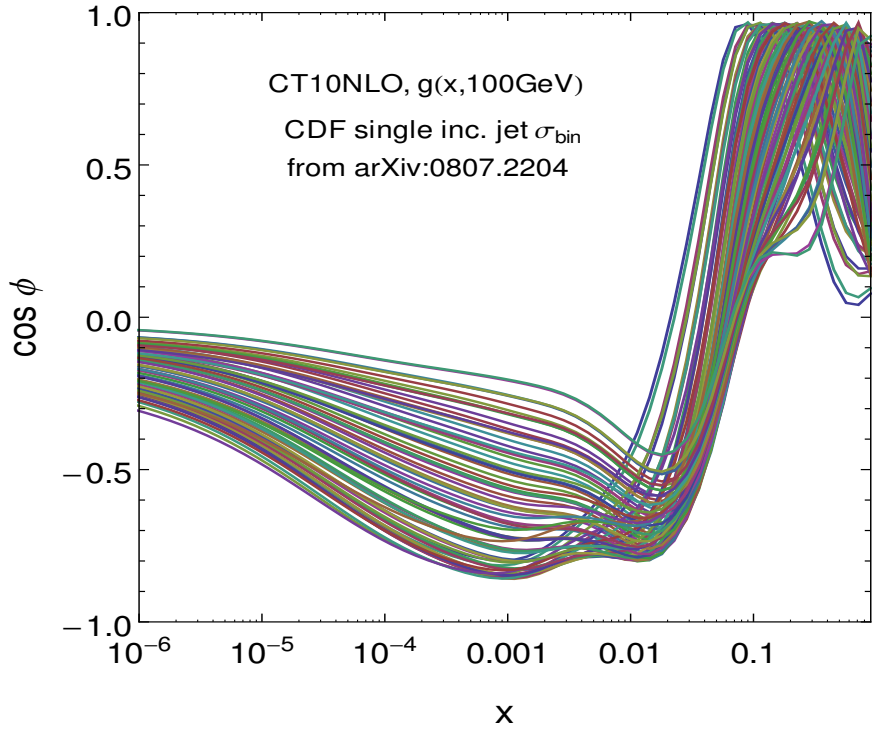
Sensitivity of PDFs to input cross sections

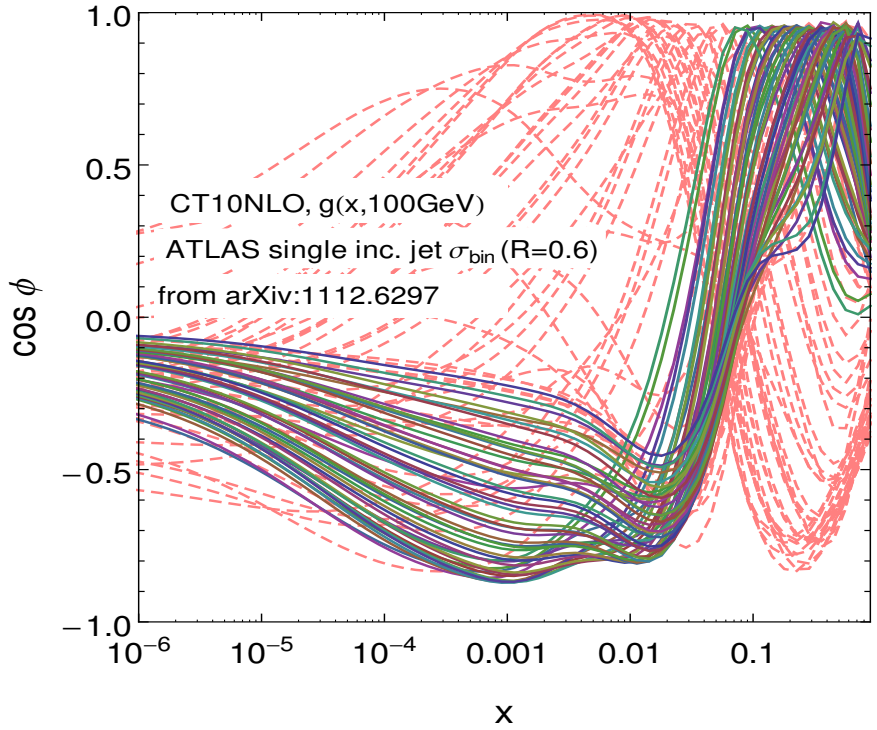
It is often interesting to know which specific data sets included in the PDF analysis constrain a given $f_{a/p}(x, Q)$; or which PDF flavors drive the PDF uncertainty in a given QCD observable X

This question can be addressed in several ways:

- by computing the correlation cosine $\cos\phi$ in the Hessian method *(Nadolsky et al., 0802.0007)*
- by a Lagrange multiplier scan of X *(Stump et al., hep-ph/0101051)*
- or by introducing effective Gaussian variables S_n dependent on X *(Dulat et al., 1309.0025, 1310.7601; Lai et al., 1007.2241)*

Sensitivity of $p\overset{(-)}{p} \to \text{jet} + X$ to gluon PDF

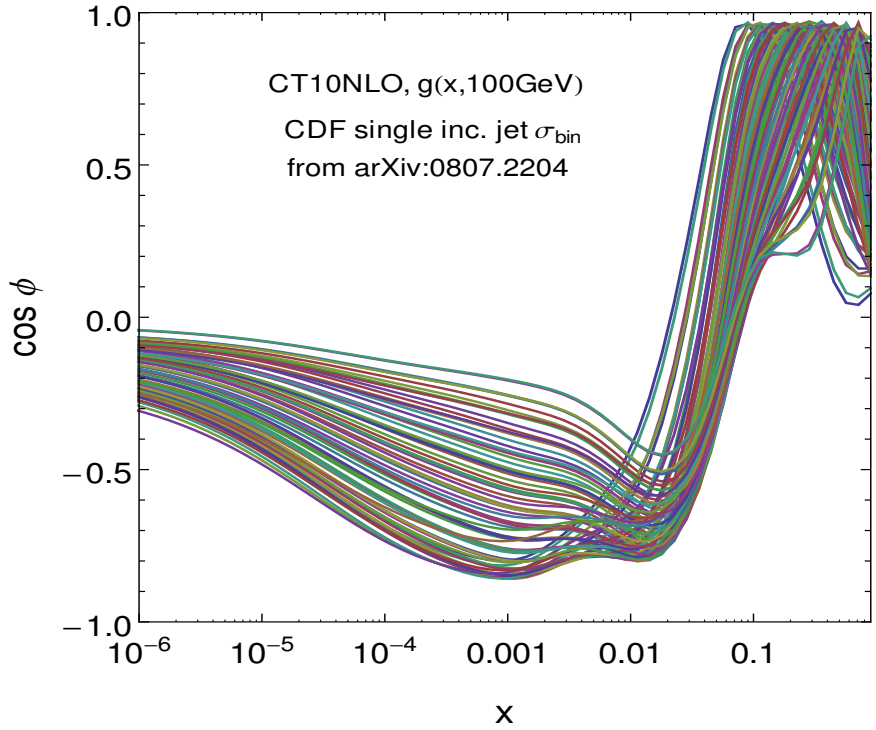


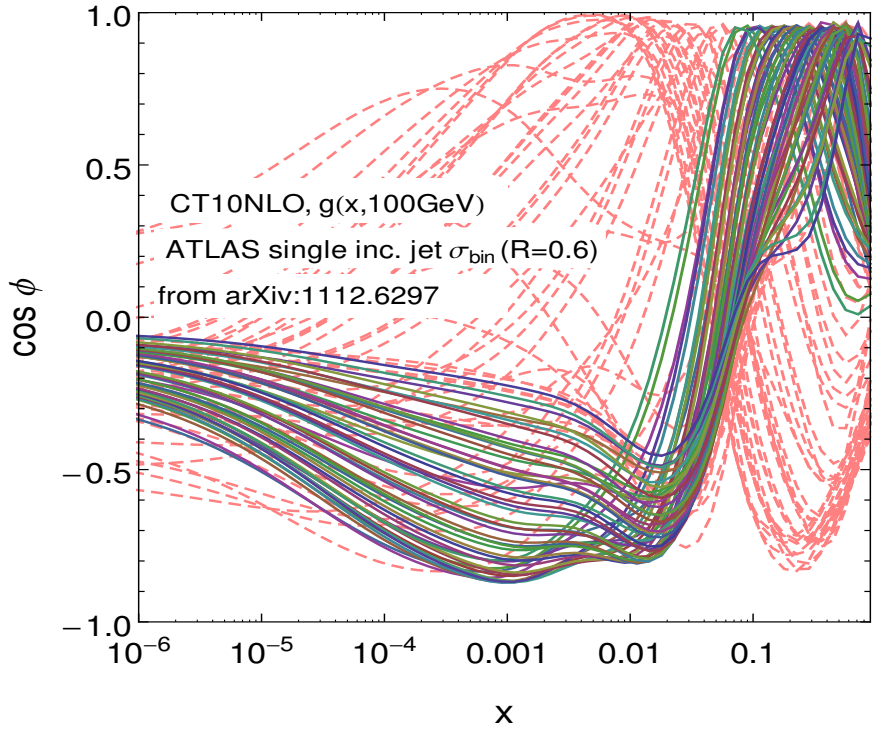


$\cos\phi(x)$ measures the correlation between the PDF uncertainty of $d^2\sigma/(dp_T dy)$ in inclusive jet production, and $g(x,Q)$ at a given x.

The curves show $\cos\varphi$ between the NLO theory cross sections in experimental $\{p_{Tj}, y_j\}$ bins and $g(x,Q)$, for various x values in $g(x,Q)$.

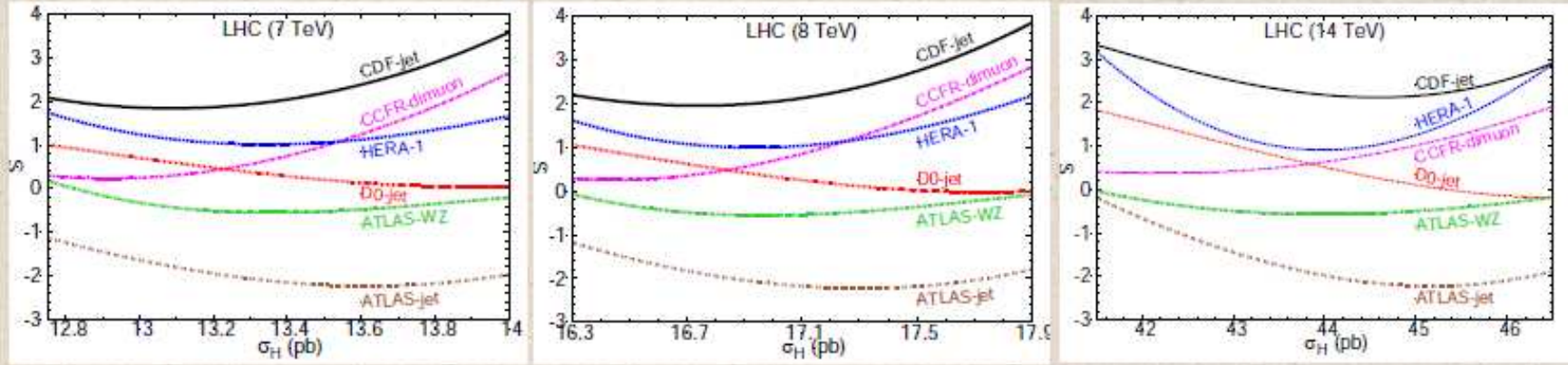
Sensitivity of $p\overset{(-)}{p} \to \text{jet} + X$ to gluon PDF





The CDF (left) and ATLAS (right) jet data are correlated with $g(x, Q)$ when $\cos\phi \gtrsim 0.7$ or $\cos\phi \lesssim -0.7$. The CDF (ATLAS) jet measurement are directly sensitive to $g(x, Q)$ with $x \gtrsim 0.1\ (0.01)$, and indirectly at $x \sim 0.001$ via the momentum sum rule. In the ATLAS jet data, the bins with $p_T^j < 200\ \text{GeV}$ (pink dashed curves) probe $g(x, Q)$ in a wider range than at CDF.

Sensitivity to Data Sets



- How sensitive are included data sets to value of σ_H?
- "Effective Gaussian Variable" S maps cumulative χ^2 distribution for N_{pt} onto cumulative Gaussian distribution
  - +1,+2,+3,… equivalent to that many sigma deviations
  - Negative values correspond to anomalously well-fit data
- Most strongly correlated data: high p_T jet, inclusive HERA, CCFR-dimuon
  - HERA more strongly correlated with 14 TeV—smaller x
  - CCFR dimuon correlation due to gluon-strange interdependence

$\alpha_s(M_Z)$ and heavy-quark masses in the PDF fits

- The QCD coupling, $\alpha_s(M_Z)$, and $\overline{MS}$ heavy-quark masses, $m_c(m_c)$, $m_b(m_b)$, and $m_t(m_t)$, can be also varied in the PDF fits.

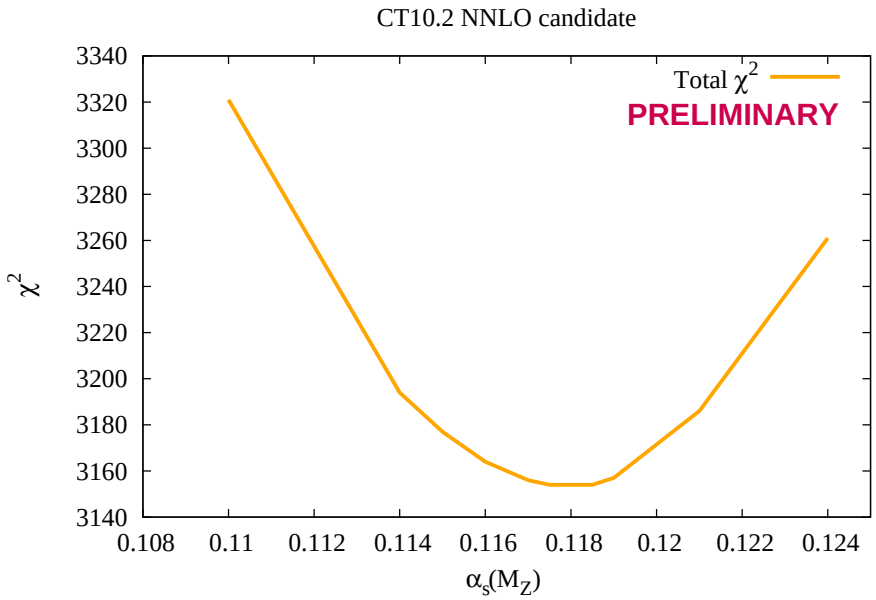
  However, the resulting constraints on these parameters are much weaker than from their dedicated measurements.

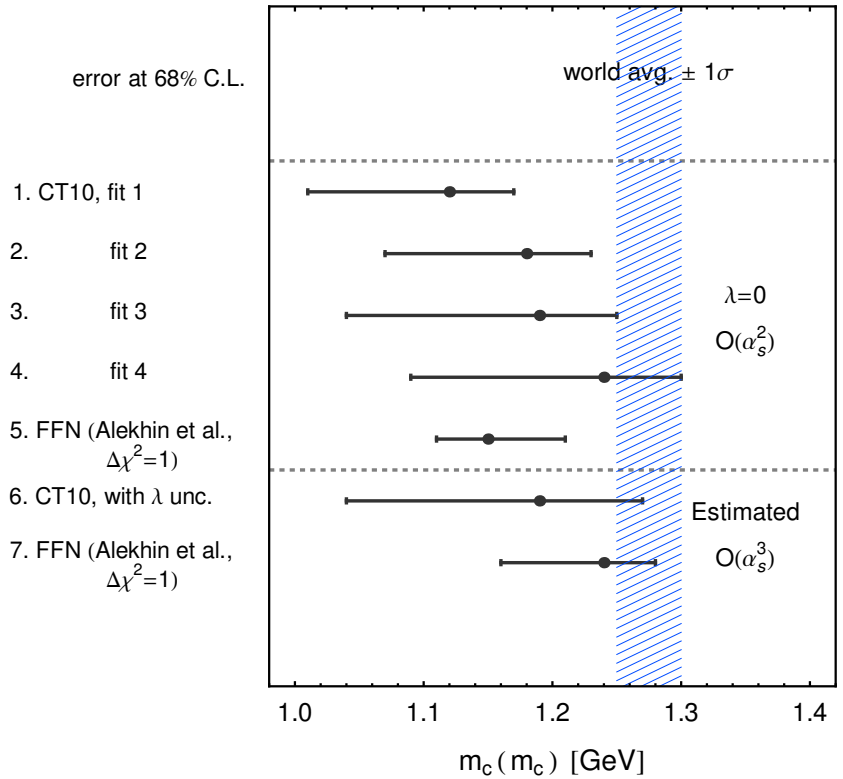
- CT10 NNLO assumes a fixed $\alpha_s(M_Z) = 0.118 \pm 0.002$ at 90%c.l., which is equal to the world average $\alpha_s(M_Z)$.

  For any X, the α_s uncertainty $\Delta_{\alpha_s} X$ is correlated with the PDF uncertainty $\Delta_{PDF} X$, the two must be combined in the full prediction.

  CT10 provides PDF error sets to compute the PDF+α_s uncertainty with full correlation.

Constraints on $\alpha_s(M_Z)$ and the $\overline{MS}$ charm mass $m_c(m_c)$ in the CT10 NNLO fit





- NLO: $\alpha_s(M_Z) = 0.11964 \pm 0.0064$ at 90% c.l.
- NNLO: $\alpha_s(M_Z) = 0.118 \pm 0.005$

Gao, Guzzi, Nadolsky, 1304.0494
$m_c(m_c) = 1.19^{+0.08}_{-0.15}$ GeV at 68% c.l.

$\alpha_s(M_Z)$ and heavy-quark masses in the PDF fits

- The QCD coupling, $\alpha_s(M_Z)$, and $\overline{MS}$ heavy-quark masses, $m_c(m_c)$, $m_b(m_b)$, and $m_t(m_t)$, can be also varied in the PDF fits.

  However, the resulting constraints on these parameters are much weaker than from their dedicated measurements.

- CT10 NNLO assumes a fixed $\alpha_s(M_Z) = 0.118 \pm 0.002$ at 90%c.l., which is equal to the world average $\alpha_s(M_Z)$.

  For any X, the α_s uncertainty $\Delta_{\alpha_s} X$ is correlated with the PDF uncertainty $\Delta_{PDF} X$, the two must be combined in the full prediction.

  CT10 provides PDF error sets to compute the PDF+α_s uncertainty with full correlation.

CT10 estimate of the PDF+α_s uncertainty

CT10 NNLO set provides

- $2N + 1 = 51$ PDF error sets at $\alpha_s(M_Z) = 0.118$, to compute

$$\Delta_{PDF} X = \frac{1}{2}\sqrt{\sum_{i=1}^{25}\left(X_i^{(+)} - X_i^{(-)}\right)^2} \quad \text{(the PDF uncertainty);}$$

- 2 best-fit PDFs for for $\alpha_s = 0.116$ and 0.120, to compute

$$\Delta_{\alpha_s} X = \frac{X(\alpha_s = 0.120) - X(\alpha_s = 0.116)}{2} \quad \text{(the } \alpha_s \text{ uncertainty).}$$

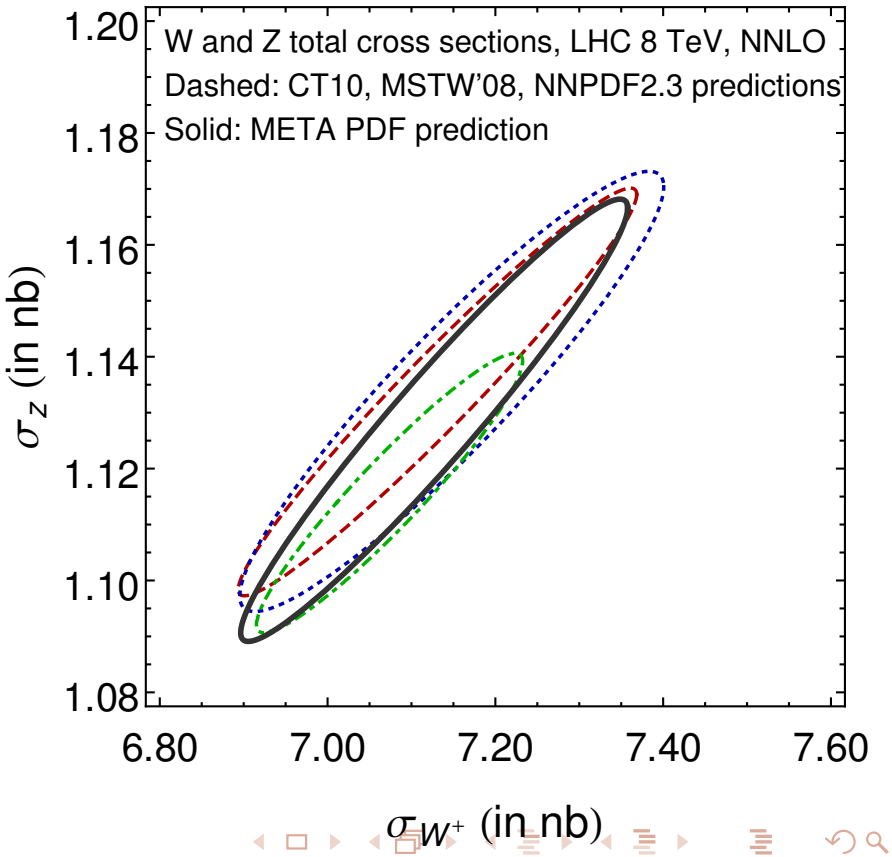
The formula for the combined PDF+α_s uncertainty with **with full correlation** is *(Lai et al., 1004.4624)*

$$\Delta_{PDF+\alpha_s} X = \sqrt{(\Delta_{PDF} X)^2 + (\Delta_{\alpha_s} X)^2}$$

Combination of PDF uncertainties from several groups

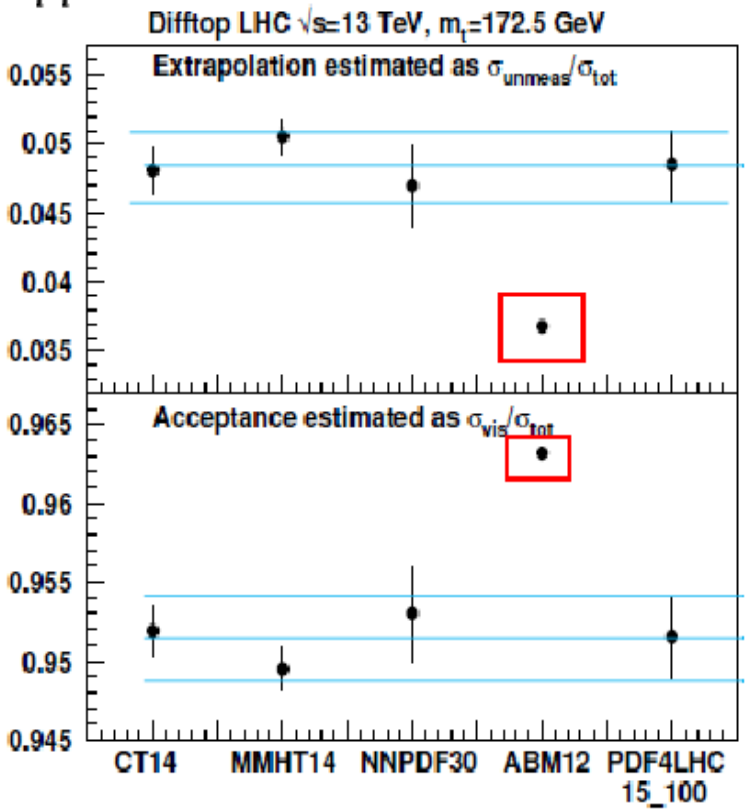
As the final step, it may be necessary to combine QCD cross sections X_i from several PDF ensembles and find the ultimate PDF uncertainty. The 2015 PDF4LHC recommendation provides guidelines for combination and three PDF4LHC'15 PDF sets constructed from CT14, MMHT'14, and NNPDF3.0 PDF sets

The META PDF method first averages the input PDFs in the PDF parameter space, before computing X_i *(J. Gao, P. Nadolsky, 1401.0013).*



Choosing the estimator for the PDF+α_s uncertainty

$pp \to t\bar{t}X$



⇒**PDF4LHC recommendations for LHC Run-II** (arXIV:1510.03865)

See also
A Critical Appraisal and Evaluation of Modern PDFs (arXIV:1603.08906)

Given numerous PDF sets, what is the PDF uncertainty in my analysis?

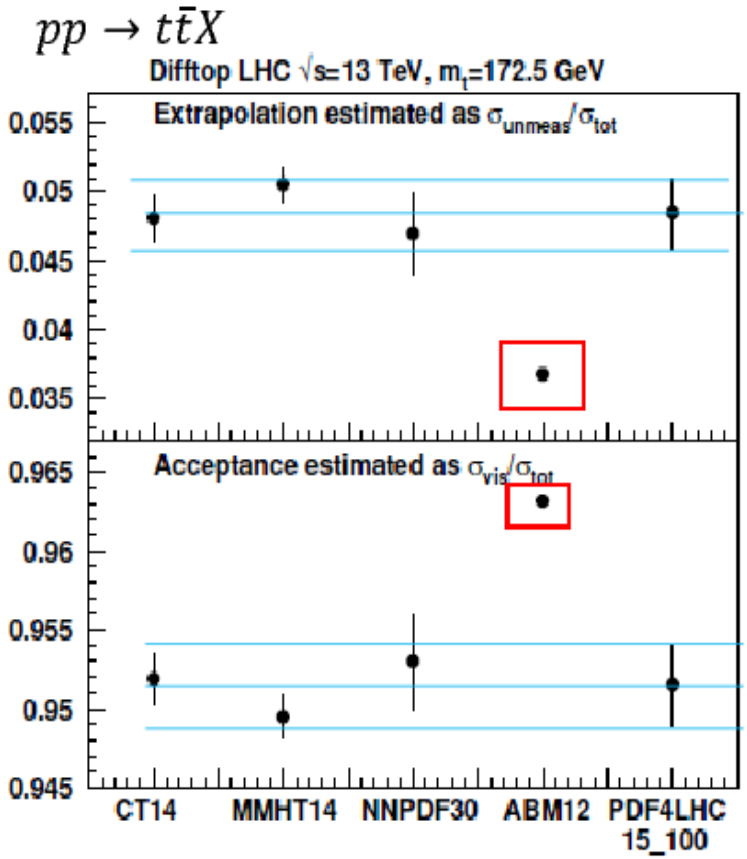


Figure: K. Lipka
1603.08906

The procedure for computing the PDF uncertainty must vary depending on the goals. The options may include

a) Using one individual set out of several similar ones (e.g., CT, MMHT, or NNPDF)

b) Using an envelope of all sets, including the outlier sets

c) 2015 recommendation by the PDF4LHC working group (arXiv:1510.03865)**:**

1. Several procedures spelled out for computation of PDF uncertainties, depending on the context
2. Estimation of PDF uncertainties is streamlined in many cases by using combined PDF4LHC15 sets based on CT14, MMHT14, and NNPDF3.0

Why PDF4LHC recommendation is necessary

Estimates of PDF uncertainties may vary drastically depending on the method.

An overly conservative estimate greatly reduces sensitivity to BSM physics.

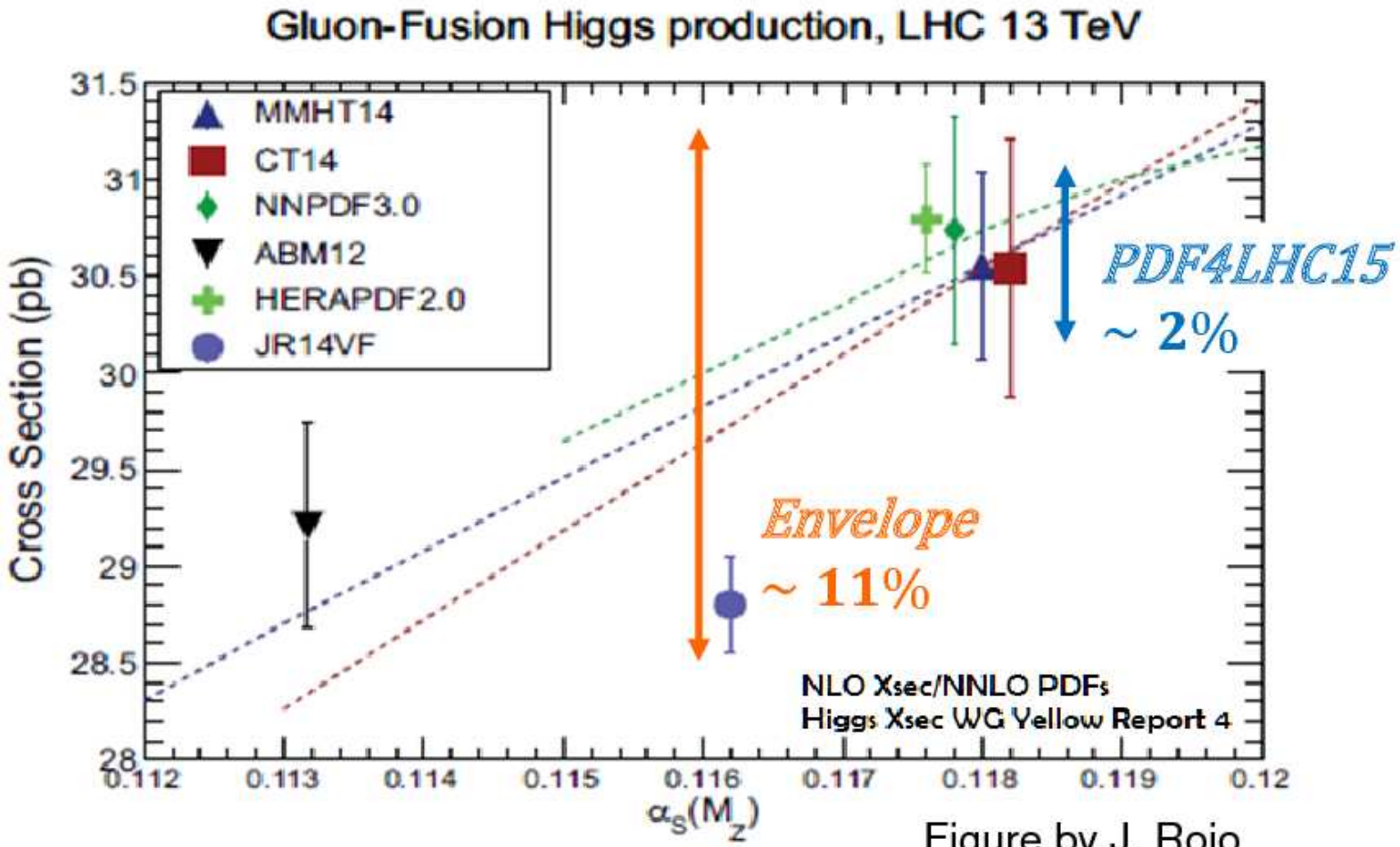


Figure by J. Rojo

Why PDF4LHC recommendation is needed

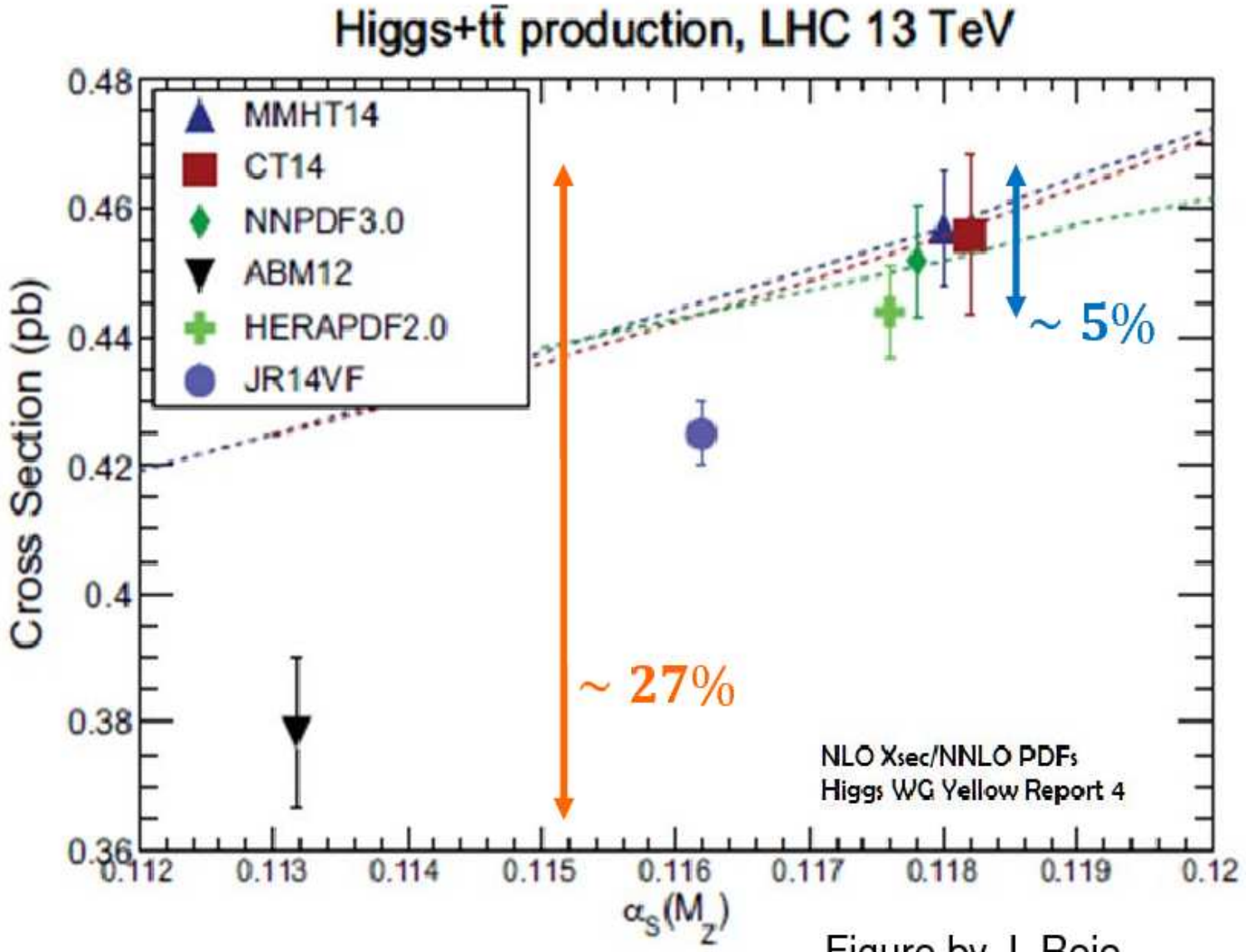


Figure by J. Rojo

OUTP-15-17P
SMU-HEP-15-12
TIF-UNIMI-2015-14
LCTS/2015-27

PDF4LHC recommendations for LHC Run II

Jon Butterworth[1], Stefano Carrazza[2], Amanda Cooper-Sarkar[3], Albert De Roeck[4,5], Joël Feltesse[6], Stefano Forte[2], Jun Gao[7], Sasha Glazov[8], Joey Huston[9], Zahari Kassabov[2,10], Ronan McNulty[11], Andreas Morsch[4], Pavel Nadolsky[12], Voica Radescu[13], Juan Rojo[14] and Robert Thorne[1].

[1] *Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK.*
[2] *TIF Lab, Dipartimento di Fisica, Università di Milano and INFN, Sezione di Milano, Via Celoria 16, I-20133 Milano, Italy*
[3] *Particle Physics, Department of Physics, University of Oxford, 1 Keble Road, Oxford OX1 3NP, UK.*
[4] *PH Department, CERN, CH-1211 Geneva 23, Switzerland*
[5] *Antwerp University, B2610 Wilrijk, Belgium*
[6] *CEA, DSM/IRFU, CE-Saclay, Gif-sur-Yvette, France*
[7] *High Energy Physics Division, Argonne National Laboratory, Argonne, Illinois 60439, U.S.A.*
[8] *Deutsches Elektronen-Synchrotron (DESY), Notkestrasse 85, D-22607 Hambury, Germany.*
[9] *Department of Physics and Astronomy, Michigan State University, East Lansing, MI 48824 U.S.A.*
[10] *Dipartimento di Fisica, Università di Torino and INFN, Sezione [illegible] Via Pietro Giuria 1, I-10125 Torino, Italy*
[11] *School of Physics, University College Dublin Science Centre . UCD Belfeld, Dublin 4, Ireland*
[12] *Department of Physics, Southern Methodist University, Dallas, TX 752[illegible]*
[13] *Physikalisches Institut, Universität Heidelberg, Heidelberg, Ge[illegible]*
[14] *Rudolf Peierls Centre for Theoretical Physics, 1 Keble Ro[illegible] University of Oxford, OX1 3NP Oxford, UK*

A major revision of the previous PDF4LHC recommendation in arxiv:1101.0538, arXiv:1211.5142

Before the PDF4LHC recommendations

Everyone proceeds at their own volition

After the PDF4LHC'15 recommendation

- Most traffic follows "the standard direction"
- Contraflow is tolerated in a few circumstances specified in the recommendation report

PDF4LHC publication, topics

1. **Review of updates on PDFs from various groups**

NNLO Global PDF sets: CT14, MMHT'14, NNPDF3

PDFs using other methodologies: ABM'12, CJ15, HERAPDF2.0

2. Average PDF sets by PDF4LHC group: PDF4LHC15_30, _100, _MC

Criteria for combination

$\alpha_s(M_Z) = 0.1180 \pm 0.0015$ at 68% c.l.

3. Recommendation on selecting PDF sets for various LHC applications

A. New physics searches

B. Precision tests of SM and PDFs

C. Monte-Carlo simulations

D. Acceptance estimates

Average PDF sets **can** be used **for bulk of applications** in A, C, D

Conclusion: the role of the PDF analysis now and in the future

It is a payoff decade for the PDF analysis efforts!

- A windfall of new data to compare with (LHC, HERA, Tevatron, JLab, RHIC,...)
- Tests of QCD factorization at new $\sqrt{s}$, targeting 1% precision
- New tools (HERA Fitter, APPLGRID, FastNLO,...) and methods (MC sampling, PDF reweighting) revolutionize the PDF analysis
- Understanding of the PDFs at the (N)NNLO level will be crucial for the success of the LHC physics program

Backup slides

CT14: parametrization forms

- Candidate CT14 fits have 30-35 free parameters
- In general, $f_a(x, Q_0) = A x^{a_1} (1-x)^{a_2} P_a(x)$
- **CT10** assumed $P_a(x) = \exp\left(a_0 + a_3\sqrt{x} + a_4 x + a_5 x^2\right)$
  - exponential form conveniently enforces positive definite behavior
  - but power law behaviors from a_1 and a_2 may not dominate
- In **CT14**, $P_a(x) = G_a(x) F_a(z)$; $G_a(x)$ is a smooth factor
  - $z = 1 - 1\left(1-\sqrt{x}\right)^{a_3}$ preserves desired Regge-like behavior at low x and high x (with $a_3 > 0$)
- Express $F_a(z)$ as a linear combination of Bernstein polynomials:

$$z^4, 4z^3(1-z), 6z^2(1-z)^2, 4z(1-z)^3, (1-z)^4$$

  - each basis polynomial has a single peak, with peaks at different values of z; reduces correlations among parameters

(N+1)-dim. perspective eliminates wrong N-dim. solutions

What about $s(x,Q) \neq \bar{s}(x,Q)$?

This can be tested in subprocesses

$$W^+ s \to c \text{ and } W^- \bar{s} \to \bar{c}$$

In the experiment, charm quarks can be identified by their semileptonic decays,

$$c \to s\mu^+\nu \text{ and } \bar{c} \to \bar{s}\mu^-\bar{\nu}$$

So one sees

$$\nu N \to \mu^- c X \to \mu^-\mu^+ X$$
$$\bar{\nu} N \to \mu^+ c X \to \mu^+\mu^- X$$

— SIDIS muon pair production (NuTeV)

Total strangeness and strangeness asymmetry

Denote

$$[q_i](Q) \equiv \int_0^1 x\, q_i(x,Q)\, dx$$ —net moment fraction carried by q_i

and introduce $s^{\pm}(x) = s(x) \pm \bar{s}(x)$ (total strangeness and its asymmetry). It is possible that

$$\int_0^1 s^-(x)dx = 0$$

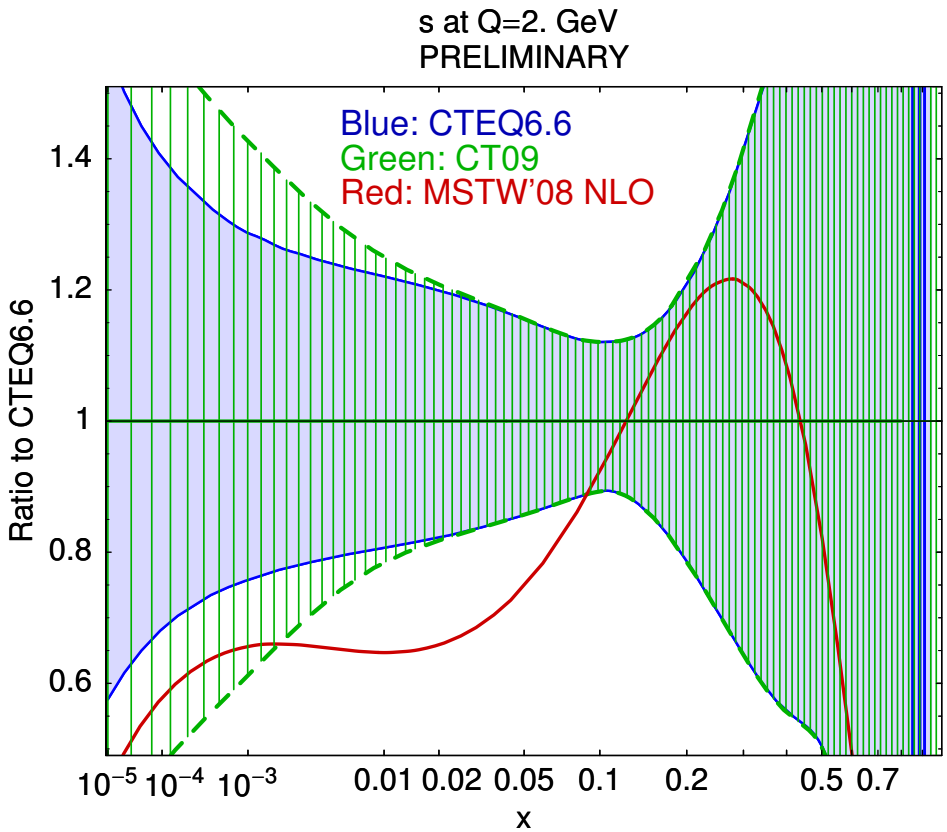
(a proton has no net strangeness), but

$$[S^-] \equiv \int_0^{-1} x s^-(x)dx \neq 0$$

(s and $\bar{s}$ have different x distributions)

A large non-vanishing $[S^-]$ might explain "the NuTeV weak angle anomaly"

CCFR (inclusive νN DIS) and NuTeV (SIDIS dimuon production): constraints on strangeness



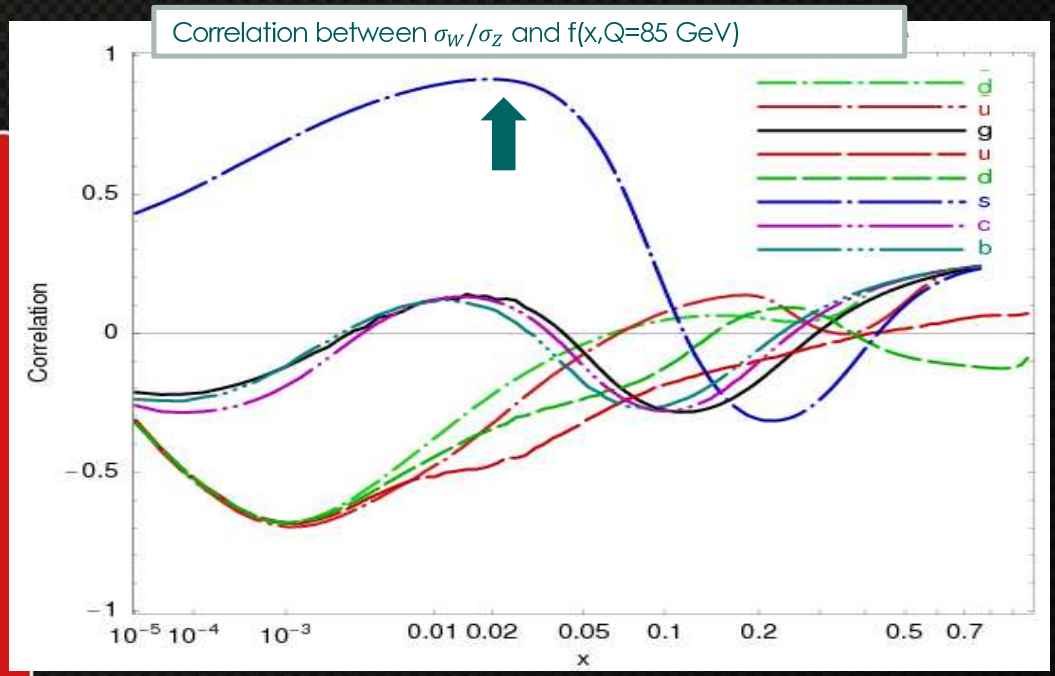
$s^+(x,Q)$ is reasonably well constrained at $x > 0.01$; practically unknown at $x < 0.01$

2009 NNPDF estimate (least biased by the parametrization of $s^-(x,Q)$):

$$[S^-](Q^2 = 20\ \text{GeV}^2) = 0 \pm 0.009$$

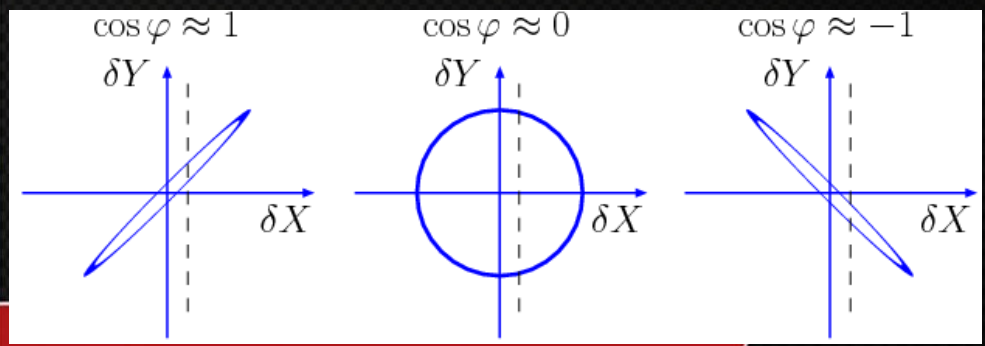
No statistically significant $[S^-]$; but the PDF error is large enough to eliminate the NuTeV anomaly (!)

Constraining strangeness PDF by LHC W and Z cross sections



2008, CTEQ6.6 (arXiv:0802.0007): the ratio σ_W/σ_Z at LHC must be sensitive to the strange PDF $s(x,Q)$

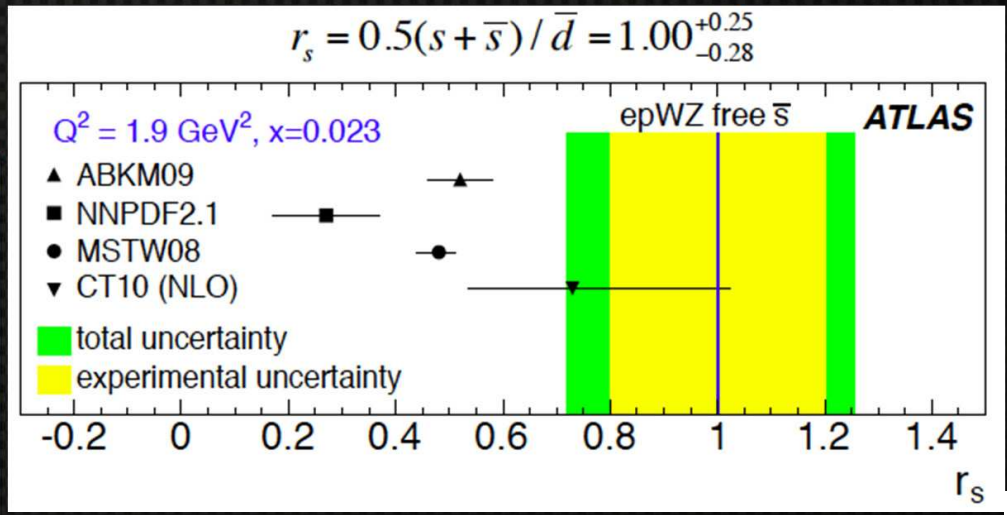
The uncertainty on $s(x,Q)$ limits the accuracy of the W boson mass measurement at the LHC



Correlation cosine $\cos\varphi \approx \pm 1$:

⇔ Measurement of X imposes tight constraints on Y

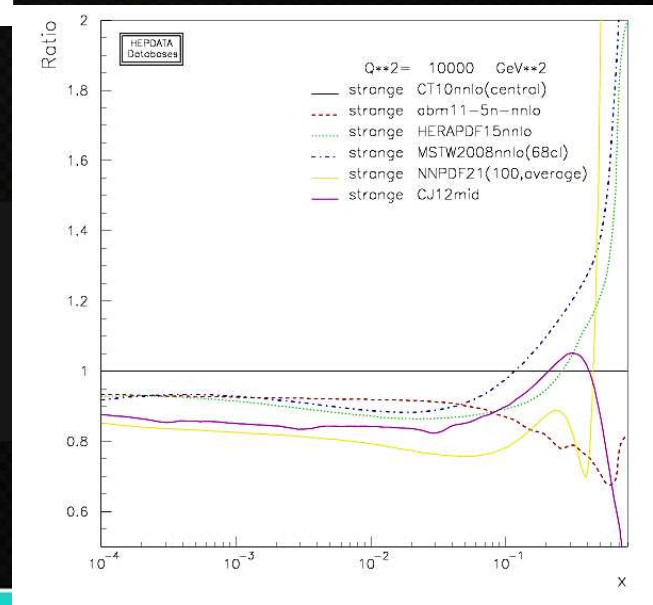
Constraining strangeness PDF by LHC W and Z cross sections



2012: the ATLAS analysis (arXiv:1203.4051) of W and Z production suggests

$\bar{s}(x,Q)/\bar{d}(x,Q) = 1.00^{+0.25}_{-0.28}$

at $x = 0.023$ and $Q^2 = 1.9$ GeV^2

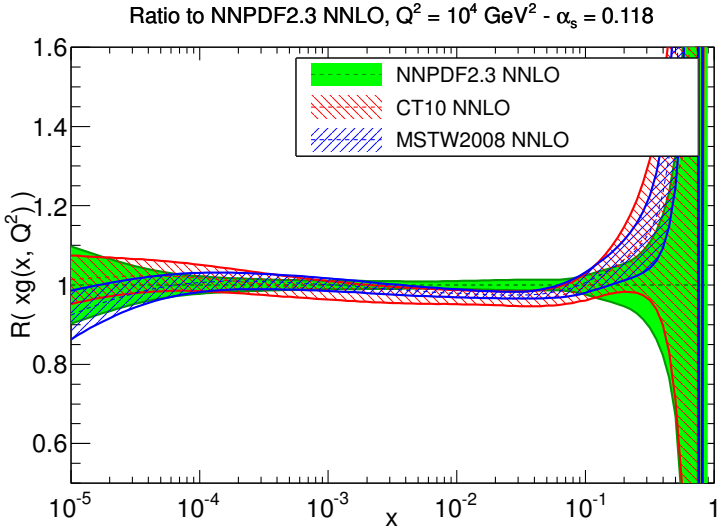


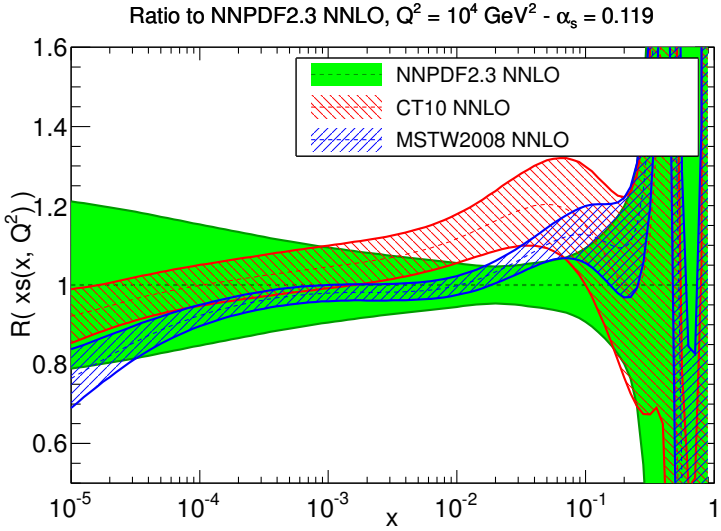
Similarly, $\sigma(W^+c)/\sigma(W^-c)$ cross section ratios show preference for $\frac{\bar{s}(x,Q)}{\bar{d}(x,Q)} \sim 1$, larger than in most pre-LHC PDF sets

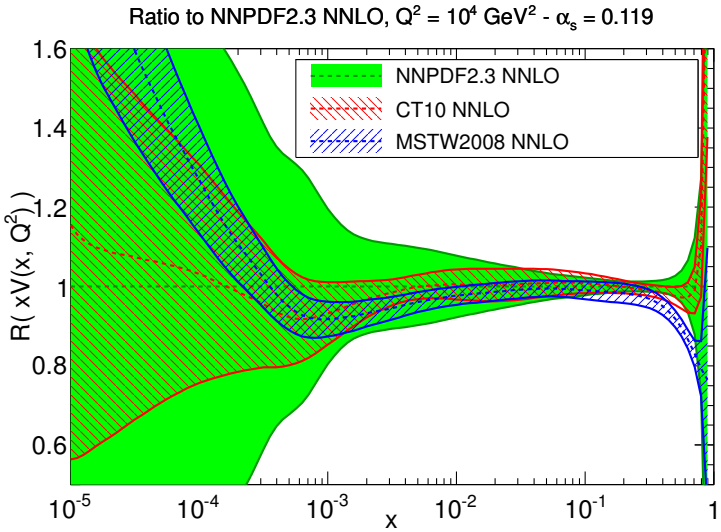
f

Ratios of CT10 and MSTW NNLO PDFs to NNPDF2.3 at $Q = 100$ GeV

(arXiv: 1211.5142)







At $x \lesssim 10^{-3}$, gluon g, strangeness s, **and** valence $V = u - d$ PDFs are **poorly** constrained;

determined by a "theoretically motivated" functional form in CTEQ/MSTW, flexible neural net in NNPDF