NNPDF3.0

Next generation PDFs for the LHC Run II

Juan Rojo
STFC Rutherford Fellow
Rudolf Peierls Center for Theoretical Physics
University of Oxford

PDF4LHC Meeting
CERN, 03/11/2014

arXiv:1410.8849

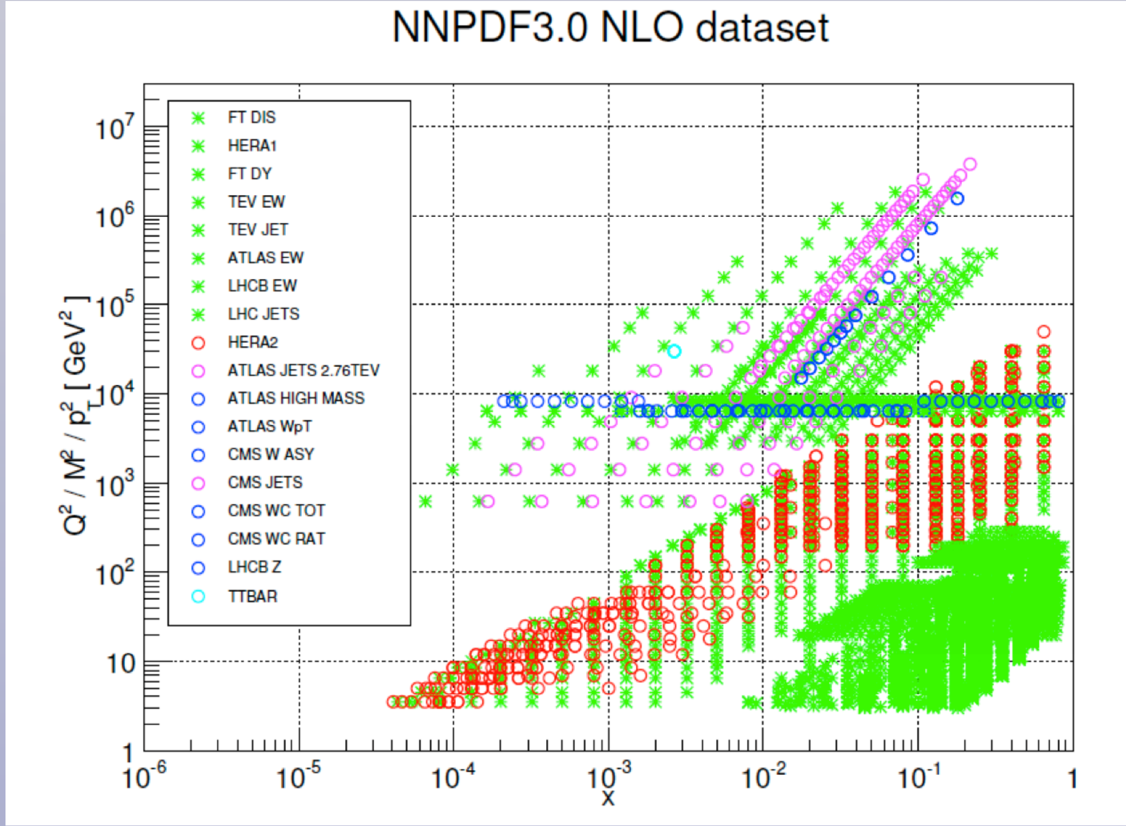
Parton distributions for the LHC Run II

The NNPDF Collaboration:
Richard D. Ball[1,2], Valerio Bertone[2], Stefano Carrazza[4,2], Christopher S. Deans[1], Luigi Del Debbio[1], Stefano Forte[4], Alberto Guffanti[5], Nathan P. Hartland[1], José I. Latorre[3], Juan Rojo[6] and Maria Ubiali[7]

Abstract

We present NNPDF3.0, the first set of parton distribution functions (PDFs) determined with a methodology validated by a closure test. NNPDF3.0 uses a global dataset including HERA-II deep-inelastic inclusive cross-sections, the combined HERA charm data, jet production from ATLAS and CMS, vector boson rapidity and transverse momentum distributions from ATLAS, CMS and LHCb, $W+c$ data from CMS and top quark pair production total cross sections from ATLAS and CMS. Results are based on LO, NLO and NNLO QCD theory and also include electroweak corrections. To validate our methodology, we show that PDFs determined from pseudo-data generated from a known underlying law correctly reproduce the statistical distributions expected on the basis of the assumed experimental uncertainties. This closure test ensures that our methodological uncertainties are negligible in comparison to the generic theoretical and experimental uncertainties of PDF determination. This enables us to determine with confidence PDFs at different perturbative orders and using a variety of experimental datasets ranging from HERA-only up to a global set including the latest LHC results, all using precisely the same validated methodology. We explore some of the phenomenological implications of our results for the upcoming 13 TeV Run of the LHC, in particular for Higgs production cross-sections.

- ☑ **HERA structure function data: HERA-II structure functions from H1 and ZEUS**, combined HERA F_{2c} cross-sections
- ☑ **LHC jet data:** CMS **7 TeV inclusive jets** from 2011, ATLAS **2.76 TeV jets** including their correlation with the 7 TeV jet data
- ☑ **LHC electroweak data:** CMS **muon asymmetries** from 2011, LHCb **Z rapidity distributions** from 2011, CMS **W+charm** production data, ATLAS and CMS **Drell-Yan production**, ATLAS **W p_T** distributions
- ☑ **ATLAS and CMS top quark pair production data**

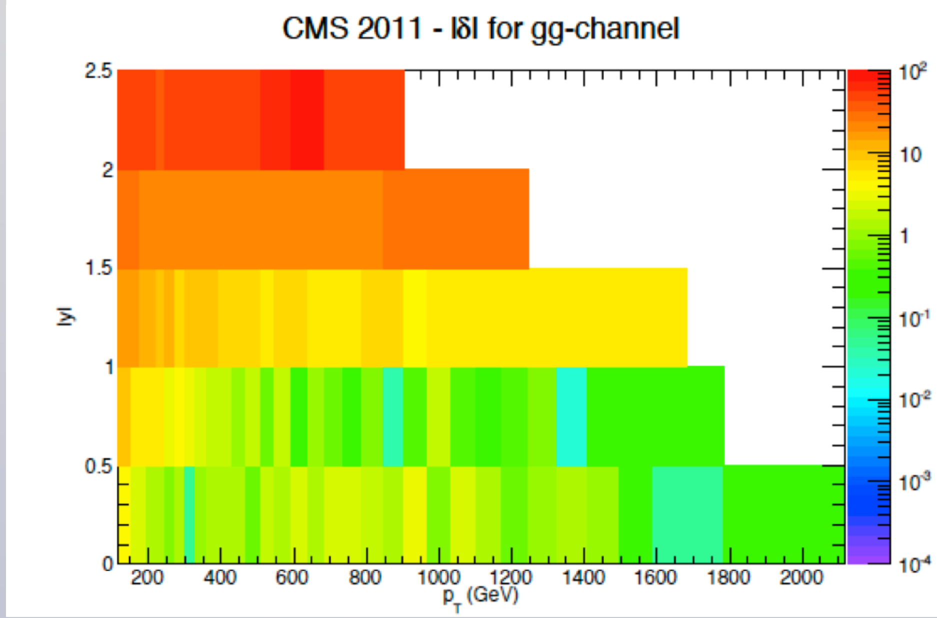


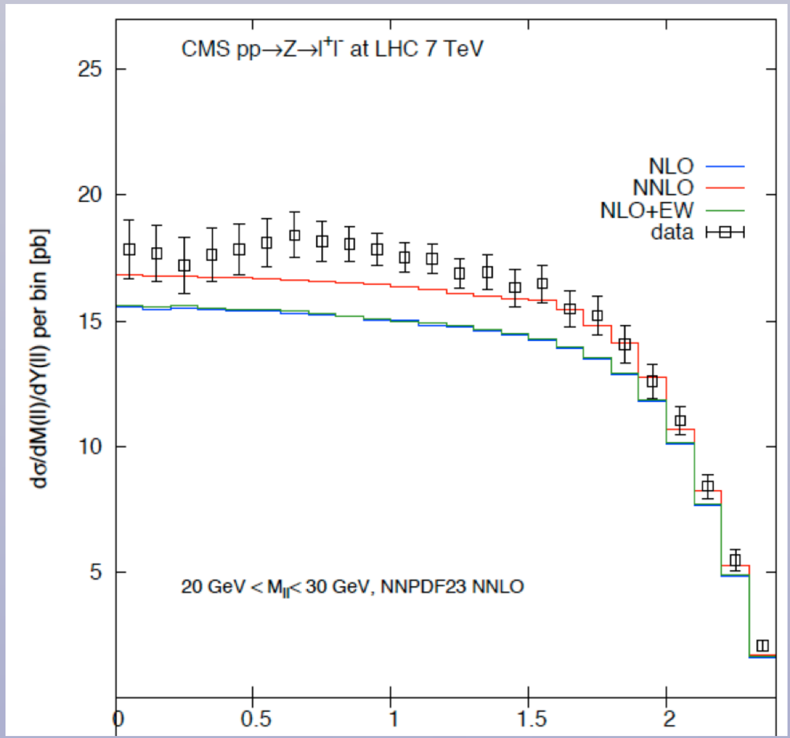
Theoretical treatment

- All NLO calculations for collider processes computed with **fast interfaces:** APPLgrid, FastNLO and aMCfast
- NLO theory supplemented by **NNLO K-factors** for Drell-Yan, top quark production and inclusive jet production

$$C^{\rm nnlo} \equiv \frac{\hat{\sigma}^{\rm nnlo} \otimes \mathcal{L}^{\rm nnlo}}{\hat{\sigma}^{\rm nlo} \otimes \mathcal{L}^{\rm nnlo}}$$

- **Electroweak corrections** also included for all neutral current Drell-Yan datasets
- Jet data at NNLO using **threshold calculation**, validated by **bin-by-bin comparison with the exact NNLO calculation** in the gg channel





Exclusion regions of jet data in NNLO fit

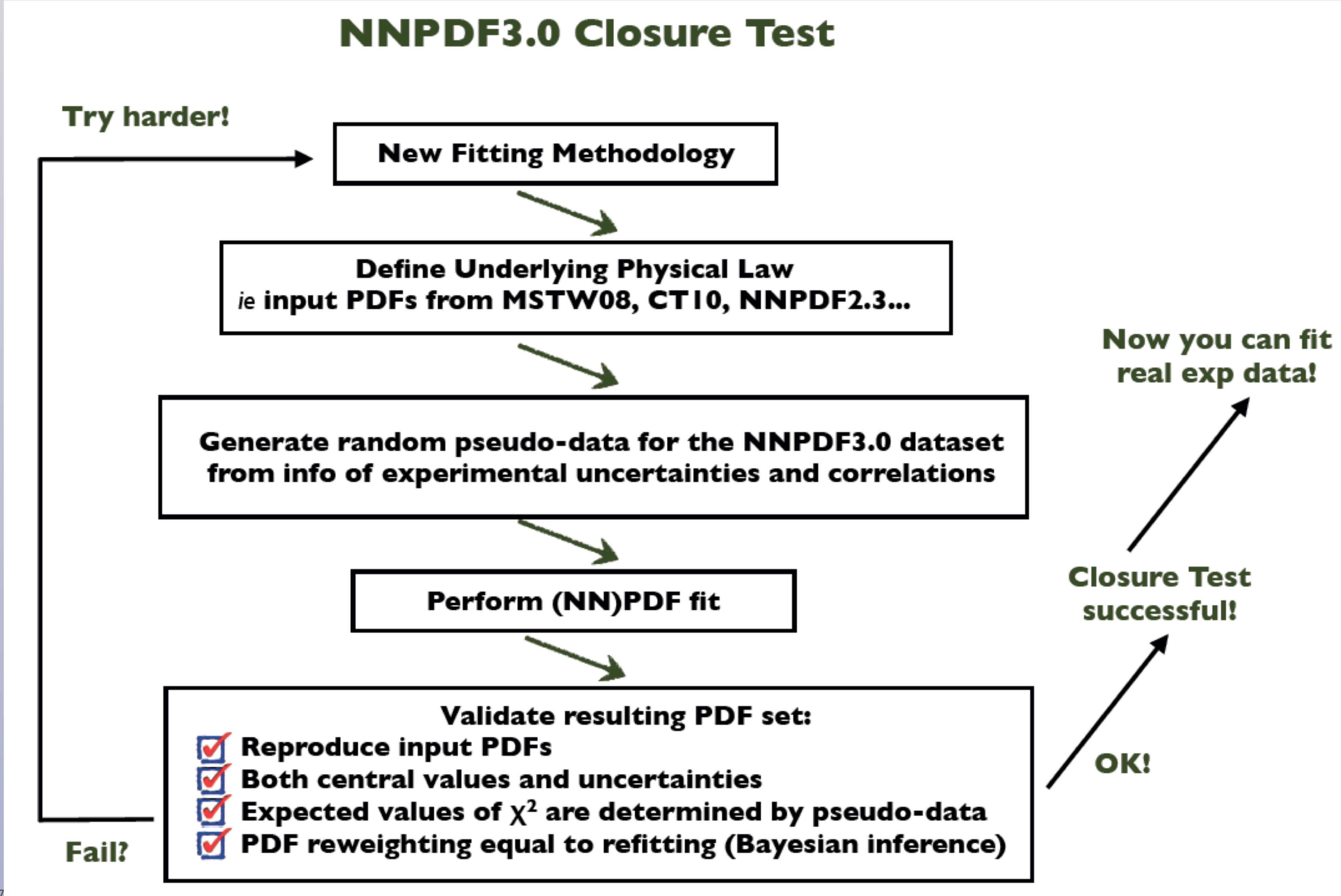
Experiment	$N_{\rm dat}$	Exclusion regions in the (y, p_T) plane	
CDF Run-II k_t jets [83]	52	$1.1 < \|y\| < 1.6$	$p_T < 224$ GeV, $p_T > 298$ GeV
		$\|y\| > 1.6$	all p_T bins
ATLAS 2.76 TeV jets [63]	3	$0.0 < \|y\| < 0.3$	$p_T < 260$ GeV
		$\|y\| > 0.3$	all p_T bins
ATLAS 7 TeV jets 2010 [50]	9	$0.0 < \|y\| < 0.3$	$p_T < 400$ GeV
		$0.3 < \|y\| < 0.8$	$p_T < 800$ GeV
		$\|y\| > 0.8$	all p_T bins
CMS jets 2011 [62]	83	$1.0 < \|y\| < 1.5$	$p_T < 272$ GeV
		$\|y\| > 1.5$	all p_T bins

- ☑ Completely rewritten Fortran NNPDF fitting code into **C++** and **Python**
- ☑ **Modular structure**: each dataset is an individual object, with the associated theory encapsulated in individual **FK tables:** easy to include new measurements and to upgrade theory for existing ones
- ☑ **Greatly improved fitting efficiency**: main bottleneck for PDF fits is convolution between input PDFs and theory, performed here with **assembly-like structure**
- ☑ **Generalized PDF parametrization**: fits can now be performed in any **arbitrary input PDF basis**, with new self-consistent method to determine **preprocessing exponent ranges**
- ☑ Optimization of the **generalized positivity of PDFs**: crucial for robust estimates of PDF errors in extrapolation regions
- ☑ **These technical and conceptual improvements** guarantee **robustness and stability for NNPDF development** in the medium and long term

Closure Testing

Closure Testing

Validation and optimization of fitting strategy performed on **closure test with known** underlying PDF set



Closure Testing

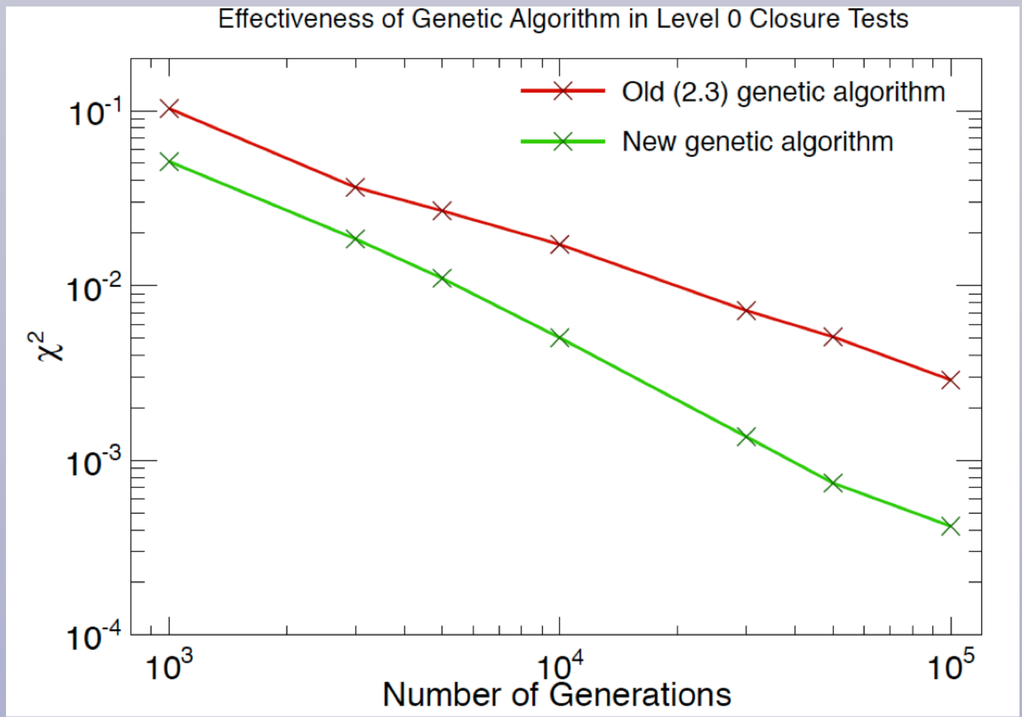
- **Level 0: no fluctuations** on pseudo-data, **no Monte Carlo** replica generation
- **Level 1**: **with fluctuations** on pseudo-data, **no Monte Carlo** replica generation
- **Level 2**: **with fluctuations** on pseudo-data, **with Monte Carlo** replica generation

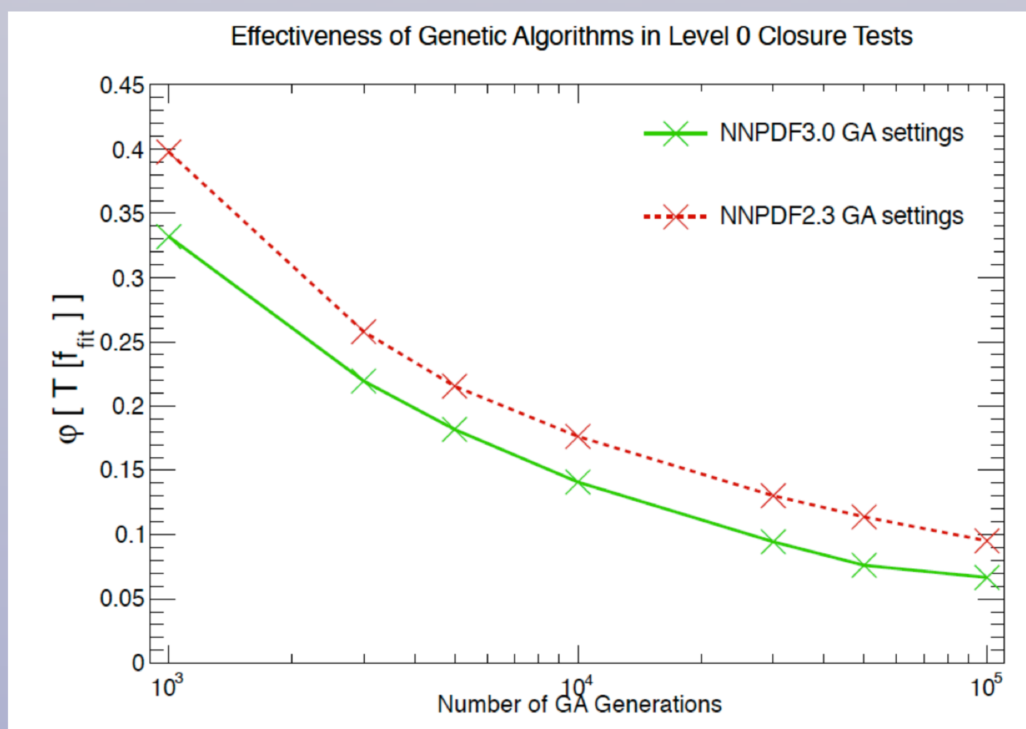
Level 0 Closure Tests:

- ☑ **Central values of input PDF** reproduced with arbitrary accuracy
- ☑ **PDF uncertainties on the fitted data points** can become arbitrarily small

Genetic Algorithms minimization efficiency substantially improved wrt NNPDF2.3

$$\varphi_{\chi^2} \equiv \sqrt{\langle \chi^2[\mathcal{T}[f_{\rm fit}], \mathcal{D}_0] \rangle - \chi^2[\langle \mathcal{T}[f_{\rm fit}] \rangle, \mathcal{D}_0]} \,.$$





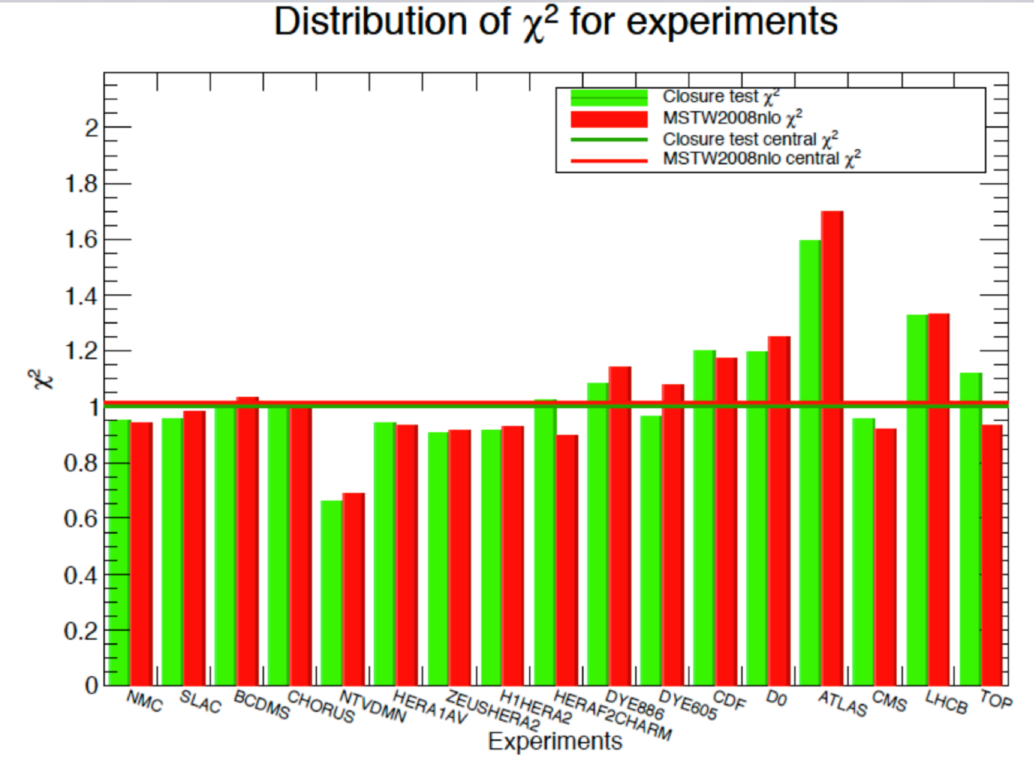
Measure of PDF uncertainties in the fitted cross-sections

Closure Testing

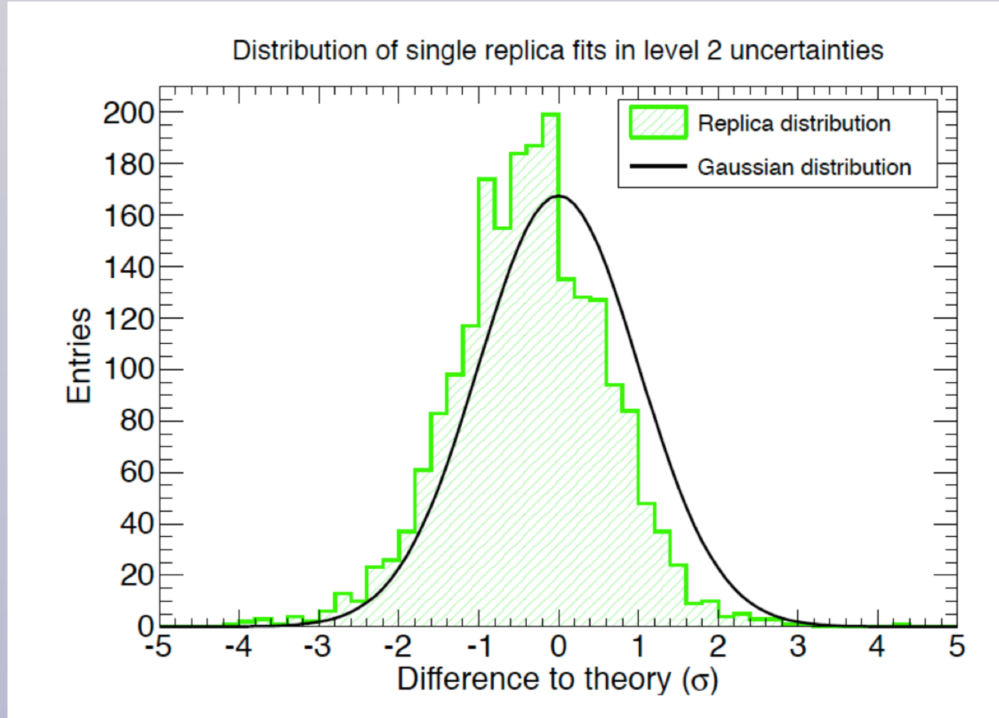
Level 2 Closure Tests:

☑ Reproduced **χ^2 of input PDF** - both total and individual experiments

☑ **Fitted PDF central values** fluctuate around input values by the same amount as expected from the **size of the PDF uncertainties**



Difference between fit and input PDFs central values in units of the PDF uncertainties



$$\xi_{n\sigma} = \frac{1}{N_{\rm PDF}} \frac{1}{N_x} \frac{1}{N_{\rm fits}} \sum_{i=1}^{N_{\rm PDF}} \sum_{j=1}^{N_x} \sum_{l=1}^{N_{\rm fits}} I_{[-n\sigma^{i(l)}_{\rm fit}(x_j), n\sigma^{i(l)}_{\rm fit}(x_j)]} \left(\langle f^{i(l)}_{\rm fit}(x_j) \rangle - f^i_{\rm in}(x_j) \right)$$

$$\xi^{(12)}_{1\sigma} = 0.699\,, \qquad \xi^{(12)}_{2\sigma} = 0.948\,,$$

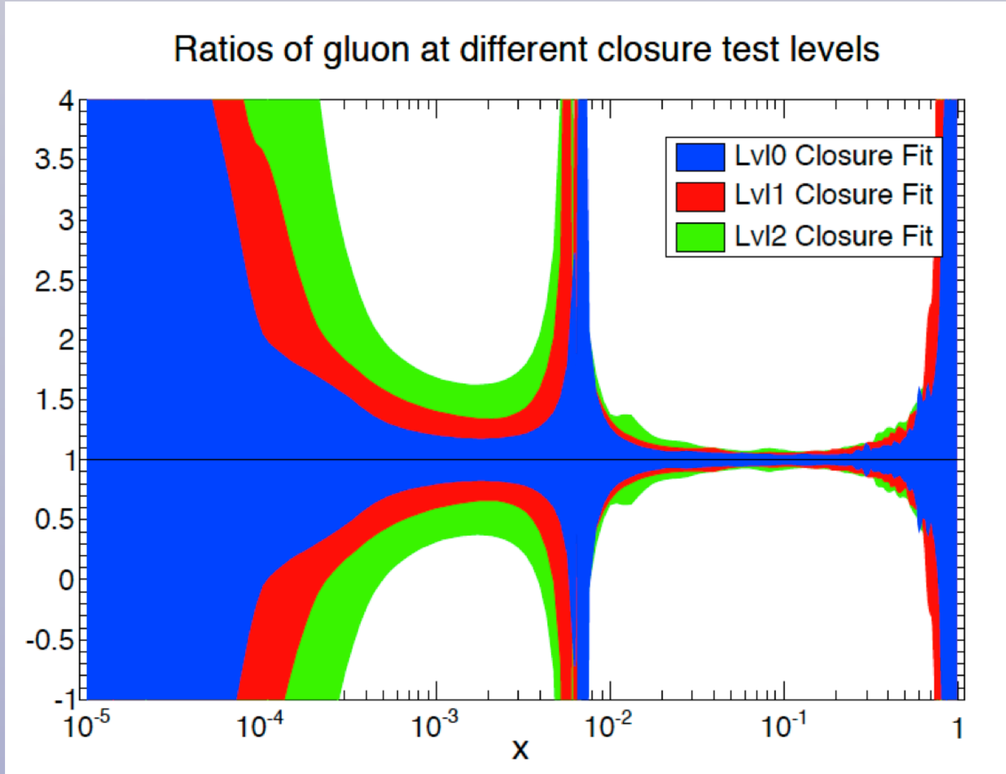
☑ The **central values of the fitted PDFs** fall in the **one(two)-sigma intervals** around **68%(95%)** of the times (average over x values and flavors)

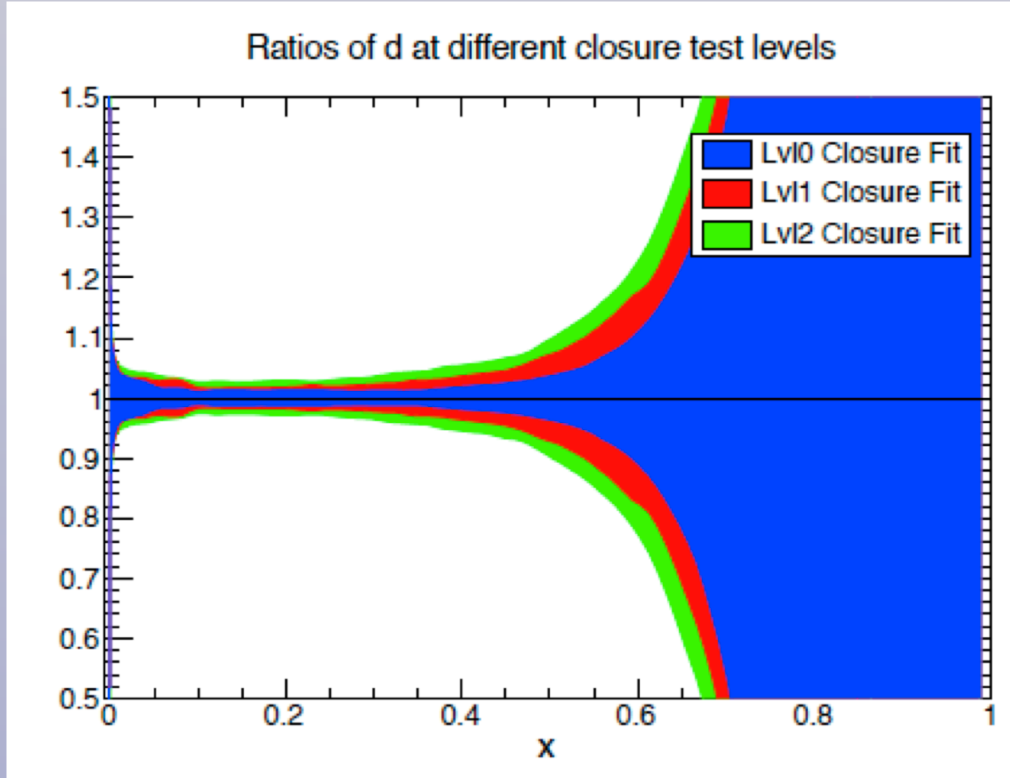
Closure Testing

Comparing **Level 0, 1 and 2 closure tests** provide a quantitative determination of the **different components of the total PDF uncertainty:**

- **Level 0:** ***Extrapolation*** **uncertainty,** due to limited kinematical coverage of experimental data. Here best-fit PDF uniquely defined - the input PDF set
- **Level 1:** ***Functional*** **uncertainty,** where a large number of different functional forms can provide an equally good fit to the pseudo-data (which now includes fluctuations)
- **Level 2:** ***Experimental*** **data uncertainty,** due to the genuine fluctuations in the experimental measurements around their true value

In regions with experimental data, the **three components are of similar size**: mandatory to include all of them for a robust estimate of PDF uncertainties





PDF Parametrization basis independence

NNPDF3.0 basis

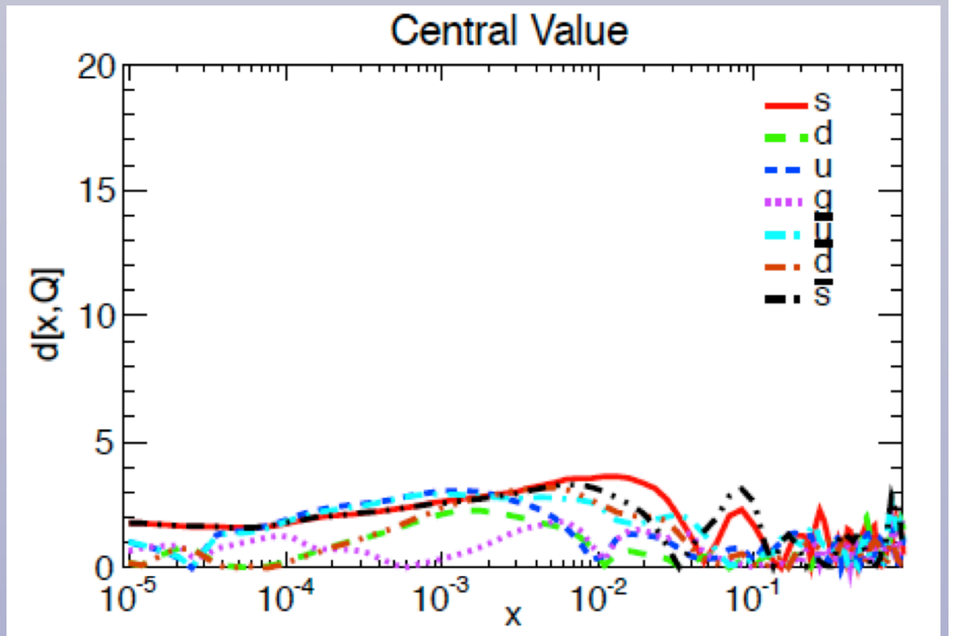
$$
\begin{aligned}
\Sigma(x,Q_0^2) &= (u+\bar u+d+\bar d+s+\bar s)(x,Q_0^2)\\
T_3(x,Q_0^2) &= (u+\bar u-d-\bar d)(x,Q_0^2)\\
T_8(x,Q_0^2) &= (u+\bar u+d+\bar d-2s-2\bar s)(x,Q_0^2)\\
V(x,Q_0^2) &= (u-\bar u+d-\bar d+s-\bar s)(x,Q_0^2)\\
V_3(x,Q_0^2) &= (u-\bar u-d+\bar d)(x,Q_0^2)\\
V_8(x,Q_0^2) &= (u-\bar u+d-\bar d-2s+2\bar s)(x,Q_0^2),
\end{aligned}
$$

NNPDF2.3 basis

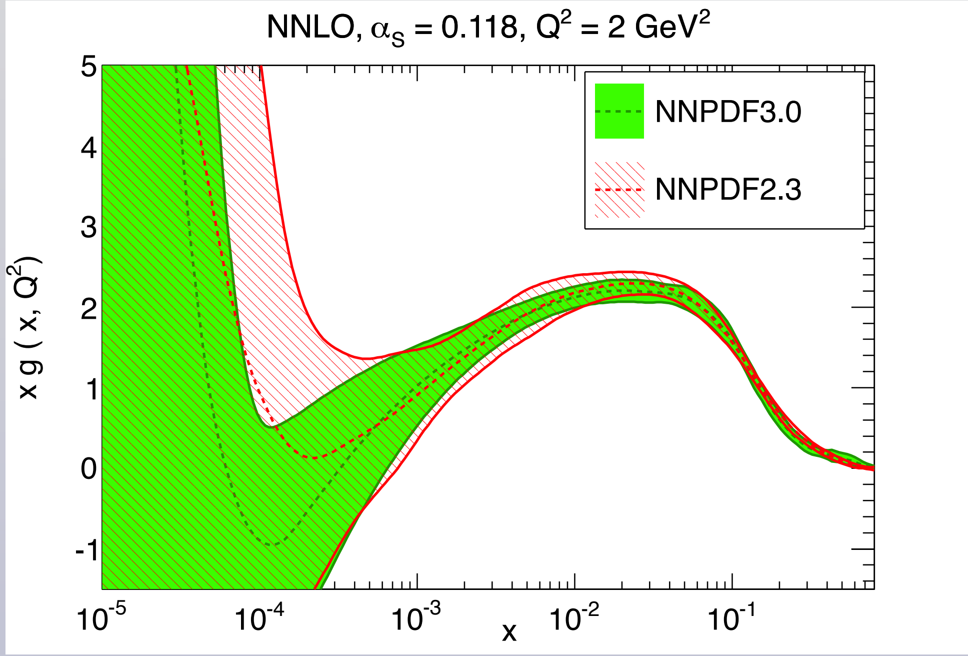
$$
\begin{aligned}
\Sigma(x,Q_0^2) &= (u+\bar u+d+\bar d+s+\bar s)(x,Q_0^2)\\
T_3(x,Q_0^2) &= (u+\bar u-d-\bar d)(x,Q_0^2)\\
V(x,Q_0^2) &= (u-\bar u+d-\bar d+s-\bar s)(x,Q_0^2)\\
\Delta_S(x,Q_0^2) &= (\bar d-\bar u)(x,Q_0^2)\\
s^+(x,Q_0^2) &= (s+\bar s)(x,Q_0^2)\\
s^-(x,Q_0^2) &= (s-\bar s)(x,Q_0^2)\,,
\end{aligned}
$$

In NNPDF3.0 a **new input PDF parametrization basis** is used, directly related to the eigenvectors of DGLAP evolution

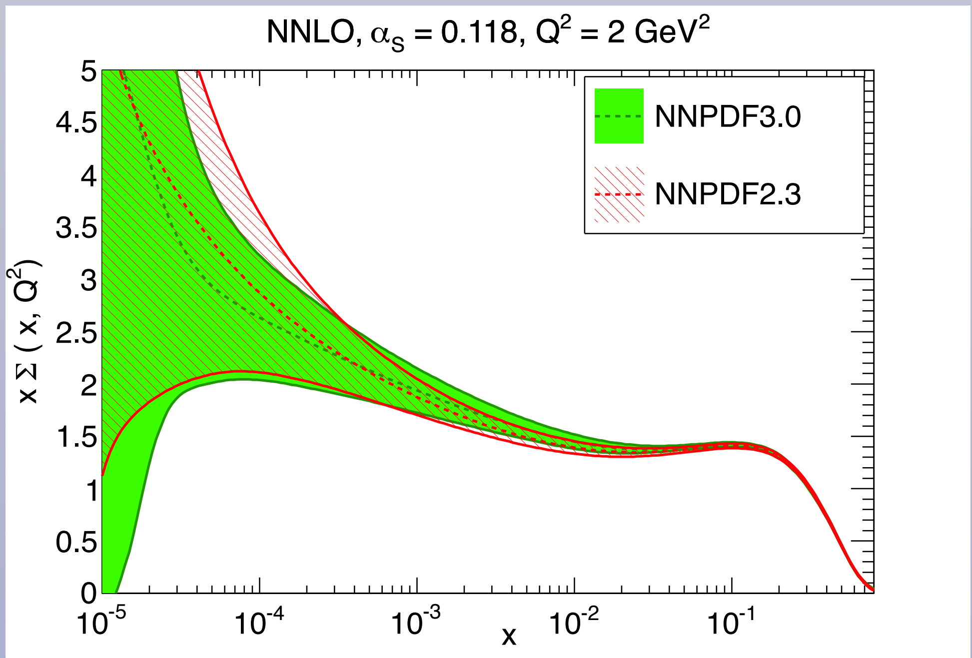
- **Checked robustness of the results** comparing fits in the ``NNPDF2.3" and ``NNPDF3.0" basis
- Verified that shapes of poorly known PDFs (like dbar-ubar or strange asymmetry) are **genuine results of the fit** and not artificial byproducts of the choice of basis

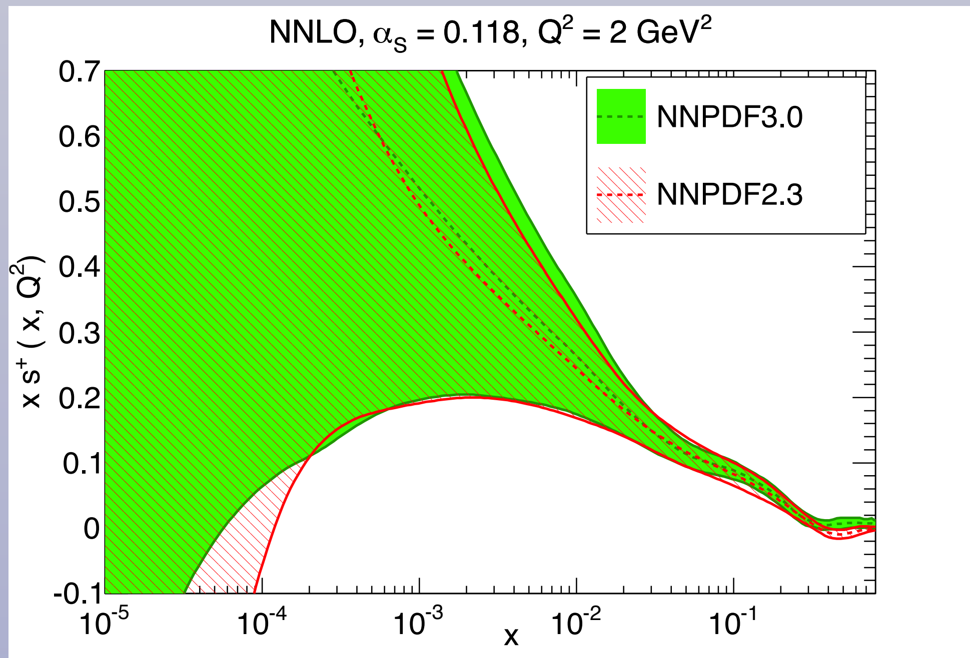


The NNPDF3.0 parton distributions

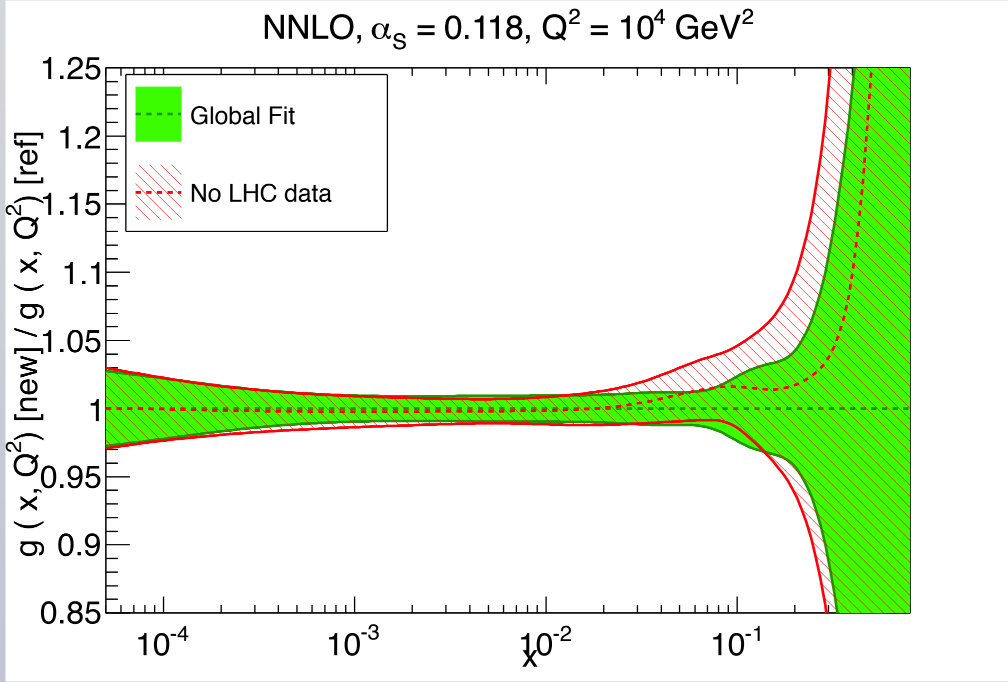


- Reasonable agreement between NNPDF2.3 and NNPDF3.0: as expected, since **all the new HERA and LHC data** already well described in NNPDF2.3
- Differences between PDFs at the 1-sigma level at most: impact of **new data** and of **updated theory and methodology**
- **PDF uncertainties are reduced** in many cases: small and large-x gluon, down quarks, strangeness...

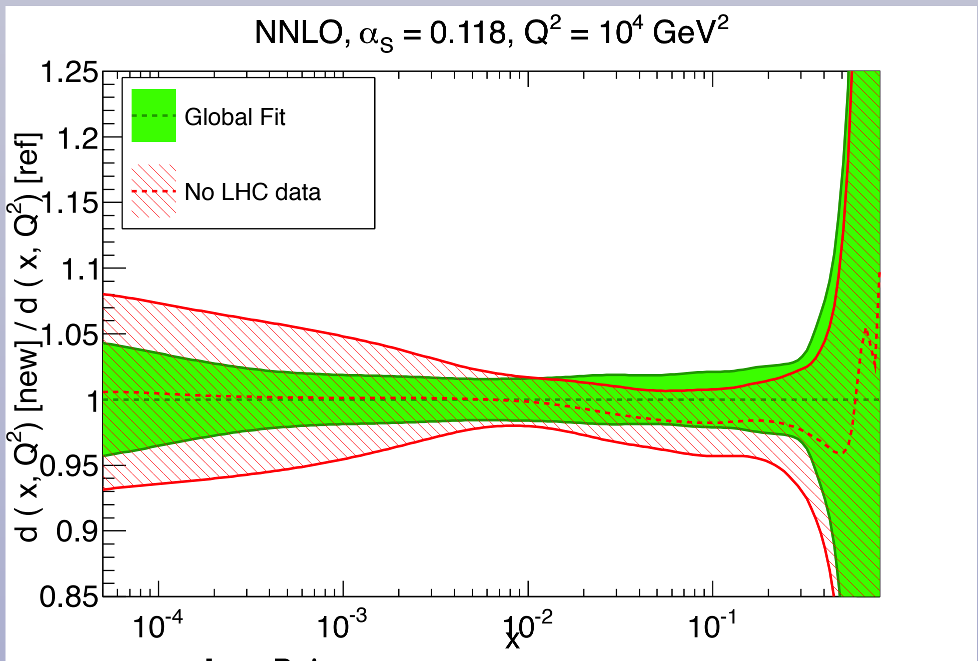


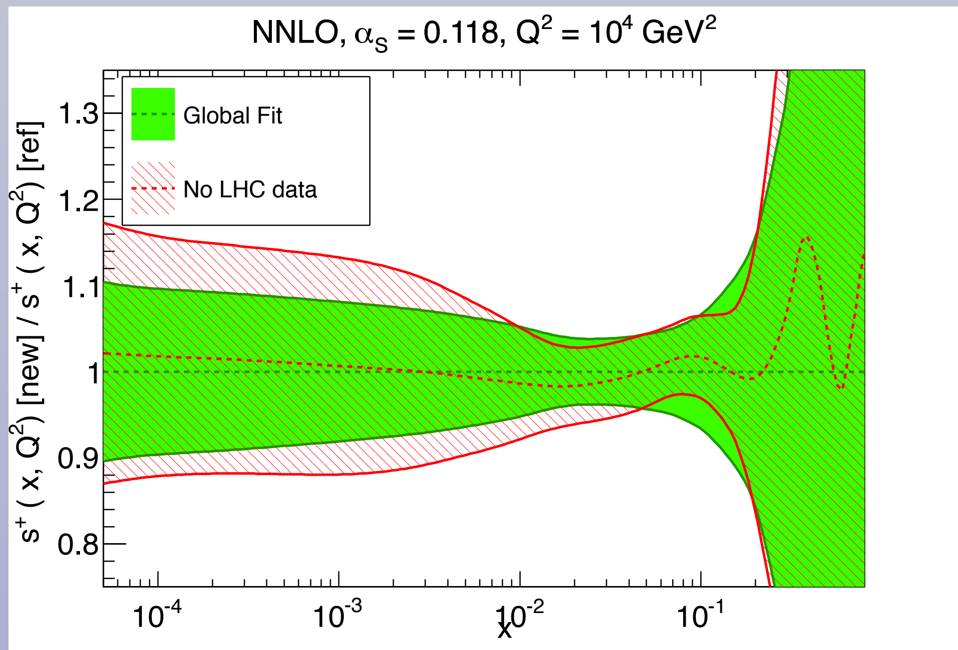


Impact of LHC data



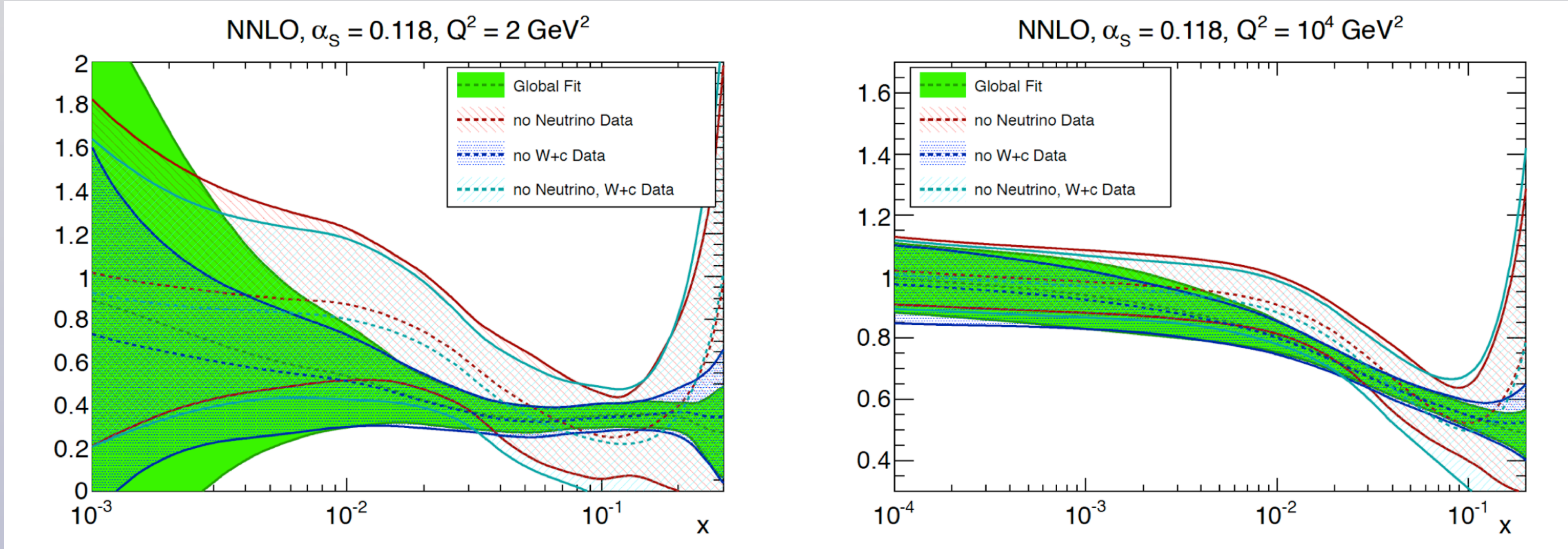
- Compare global NNPDF3.0 fit with a fit **without LHC data**
- PDF uncertainties on **large-x gluon** reduced due to **top quark and jet data**
- PDF uncertainties on **light quarks** reduced from the **Drell-Yan** and **W+charm data**
- The **description of all new LHC data**, already good in NNPDF2.3, is further improved in NNPDF3.0





$$r_s(x,Q^2) = \frac{s(x,Q^2)+\bar{s}(x,Q^2)}{\bar{d}(x,Q^2)+\bar{u}(x,Q^2)}$$

Strangeness



- The CMS **W+charm** and ATLAS **inclusive W,Z** data provide a handle on the strangeness content of the proton
- A number of fits have been performed **adding/removing experiments** with sensitivity to the strangeness: CMS W+charm, ATLAS inclusive W,Z, inclusive and dimuon neutrino data. In all cases, **good agreement within PDF uncertainties is found**, **no evidence for tension** between any data in terms of their implications for strangeness
- A fit with ATLAS WZ, but no neutrino data, leads to a fair description of W+c, but a **poor χ^2 for neutrinos**

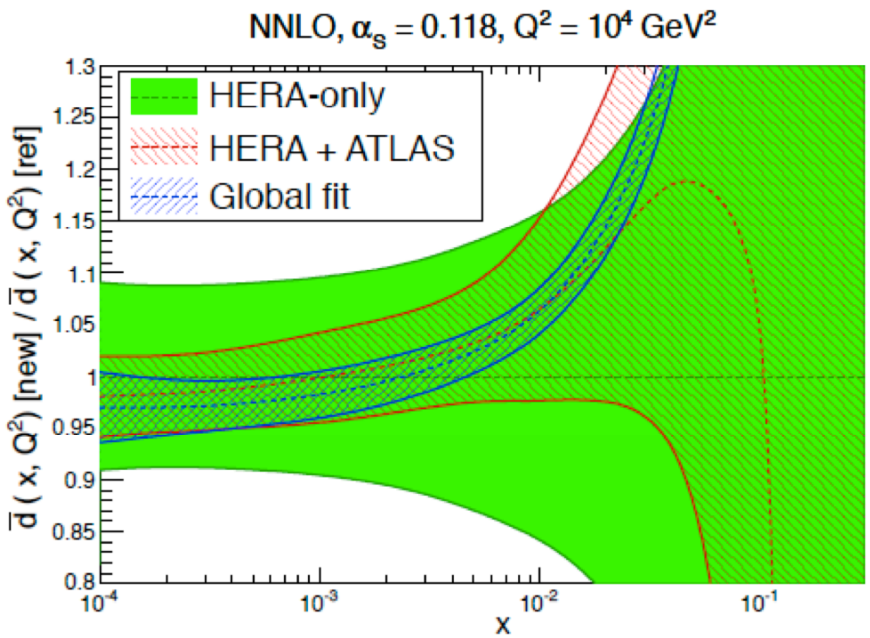
	$\chi^2_{\rm exp}$			
	Global	No neutrino	No $W+c$	No neutrino/$W+c$
CHORUS	1.13	*3.87*	1.09	*3.45*
NuTeV	0.62	*4.31*	0.66	*6.45*
ATLAS W,Z 2010	1.21	1.05	1.24	1.08
CMS $W+c$ 2011	0.86	0.50	*0.90*	*0.61*

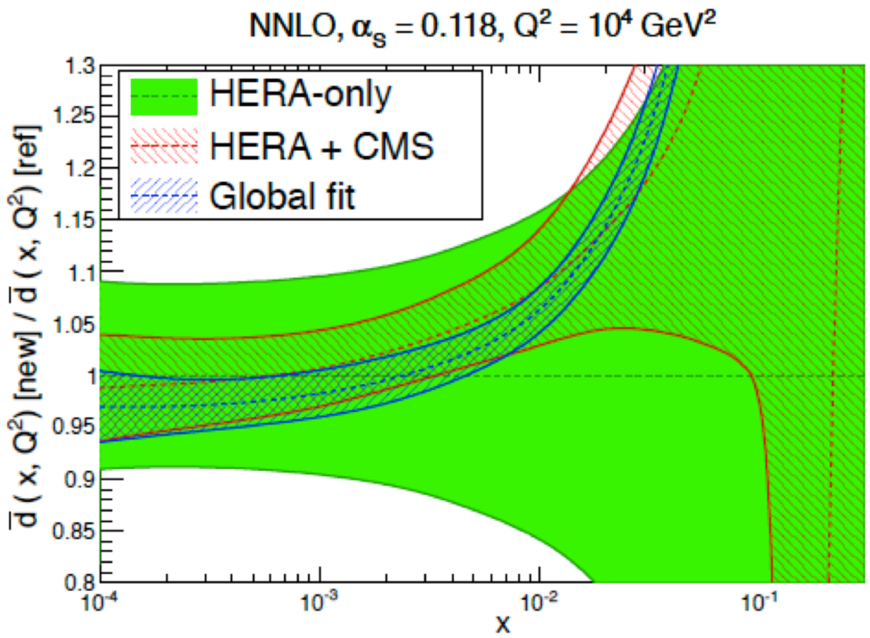
PDF sets based on reduced datasets

- The stability of NNPDF3.0 as a function of the dataset used in the fit has been extensively studied by providing **PDF fits based on reduced datasets**
- These include: HERA-only, HERA-I-only, HERA +CMS, HERA+ATLAS, fits wo jet data, fits based on a conservative dataset, fits wo LHC data, ...
- The fits with **HERA+ATLAS** and **HERA+CMS** data represent a **very substantial improvement over HERA-only fits**, but still not competitive with global PDF fits

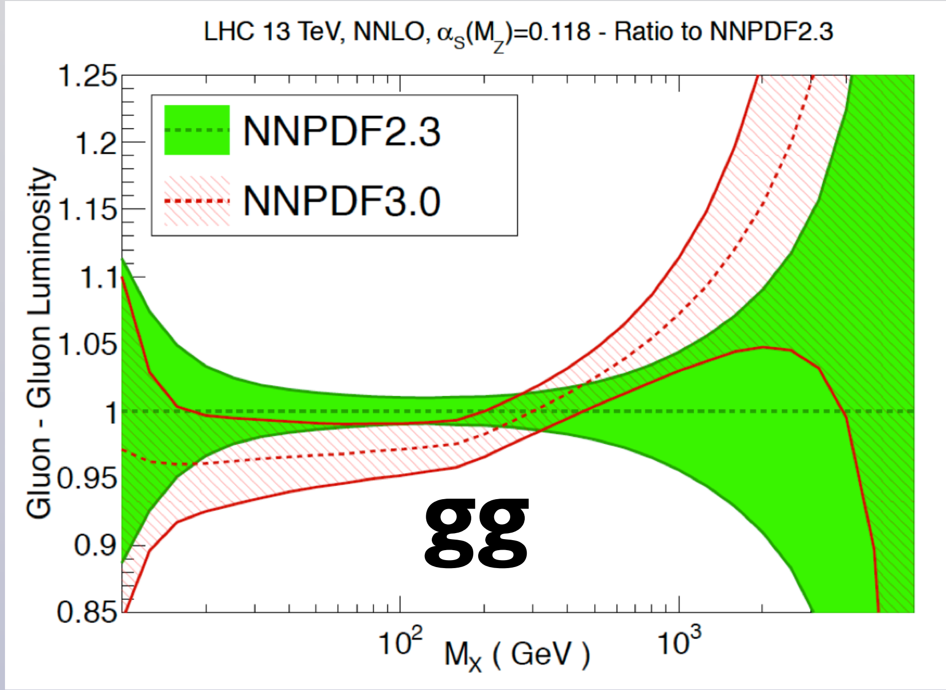
Measure of PDF uncertainties in the fitted cross-sections

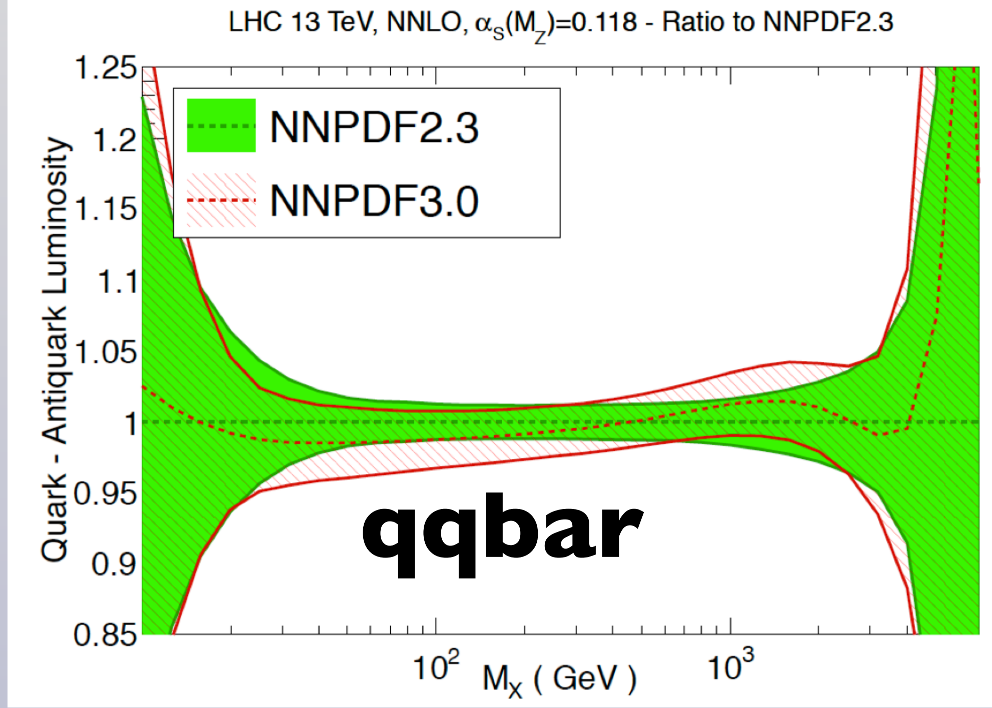
Dataset	φ_{χ^2} NLO	φ_{χ^2} NNLO
Global	0.291	0.302
HERA-I	0.453	0.439
HERA all	0.375	0.343
HERA+ATLAS	-	0.330
HERA+CMS	-	0.315
Conservative	0.422	0.478
no LHC	0.312	0.316

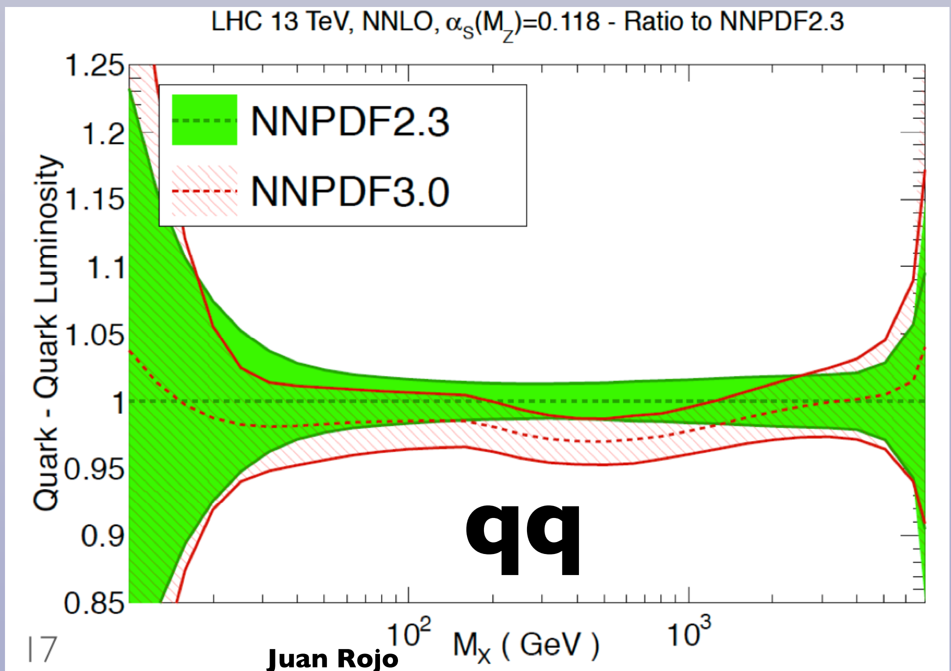




PDF luminosities

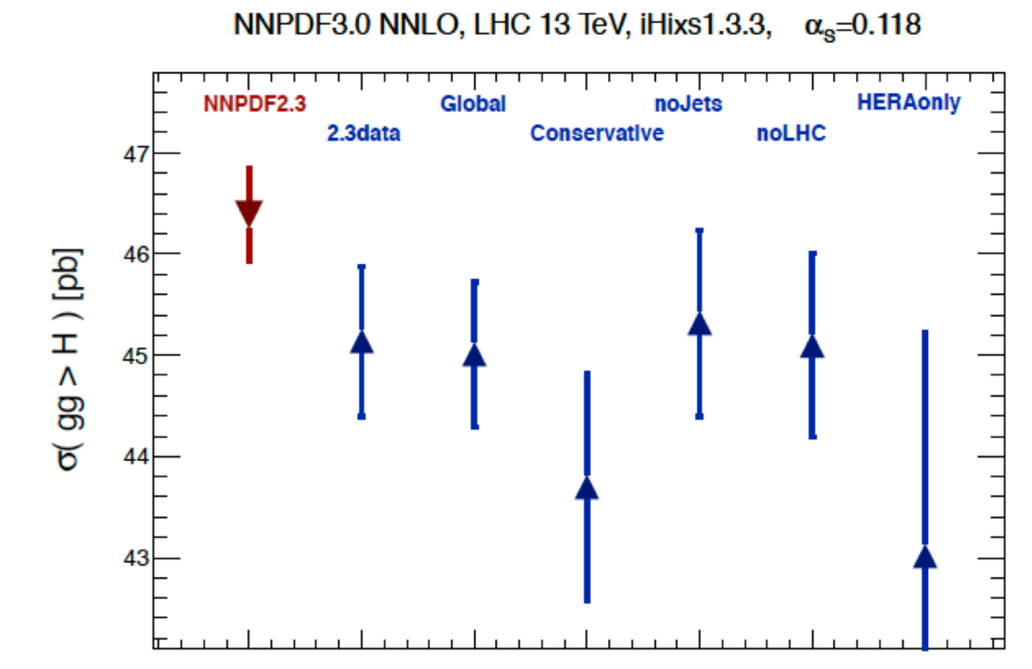


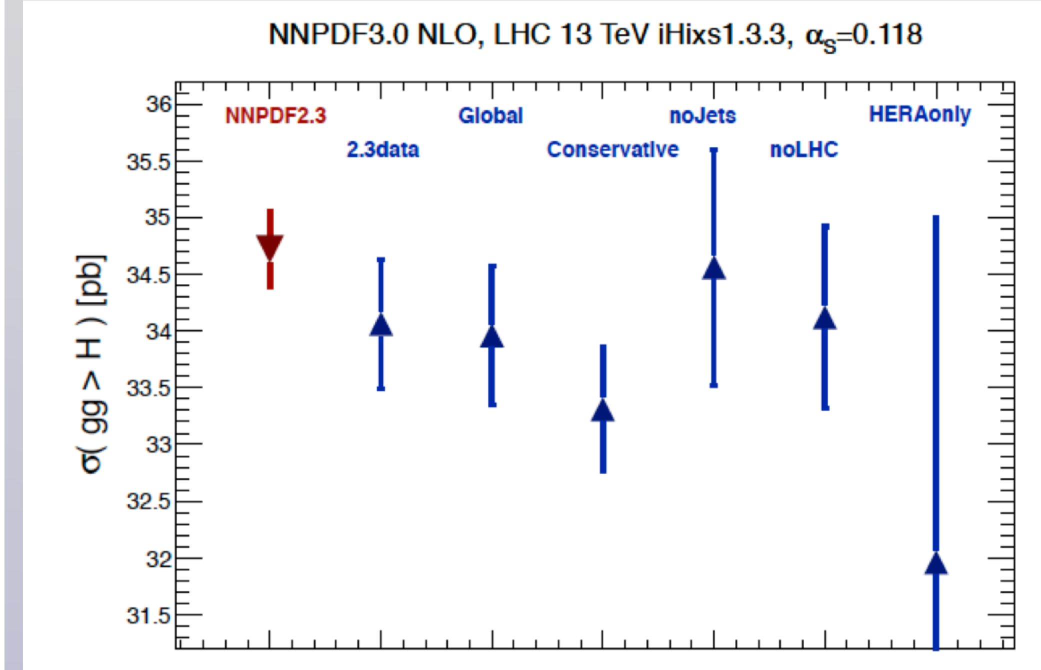




- **PDF luminosities** are useful to translate differences in PDFs into differences in LHC cross-sections
- For **qq, 3.0** softer than 2.3 for 300 GeV < M < 1 TeV, very good agreement at small and large invariant masses
- For **qqbar**, good agreement in the whole range of invariant masses
- For the **gg lumi,** NNPDF3.0 softer by about 1-sigma wrt NNPDF2.3 for **M< 200 GeV: implications for Higgs production in ggF**

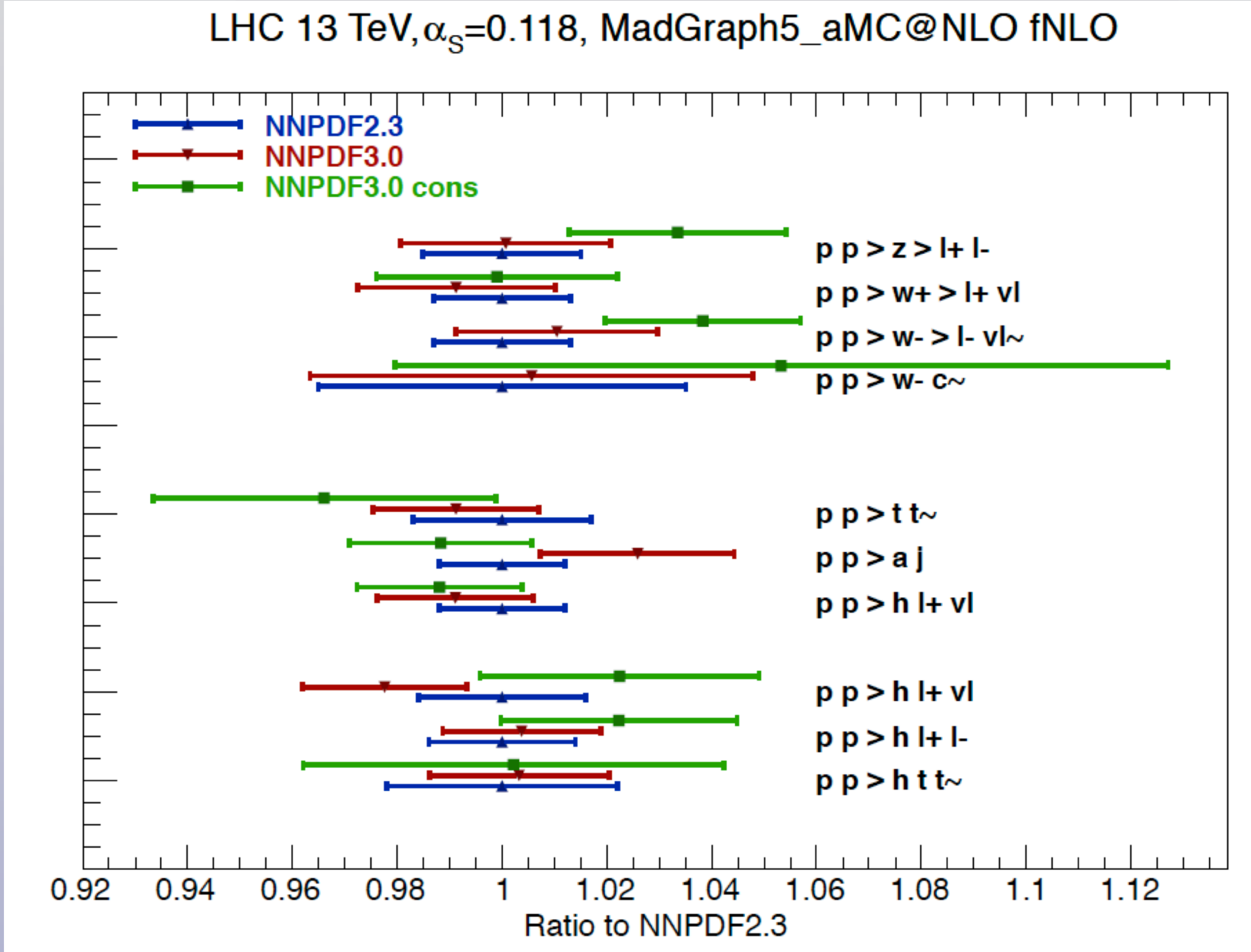
Higgs production in gluon fusion





- The softer **gg luminosity** in NNPDF3.0 leads to a **decrease in the ggH xsec** at the LHC 13 TeV
- The effect is more marked at NLO and NNLO, though even in the latter the pull is only **P ~1.5**
- The ggH process is different from many other LHC xsecs because there are **no direct experimental constraints on the gluon at x ~ 0.01**, and thus predictions for **ggH** are more sensitive to modifications in the methodology or in the choice of dataset (that indirectly affects g(x) for **x ~ 0.01**)
- In NNPDF3.0, the changes in the **ggH xsec arise mostly from** the **improved fitting methodology**, validated on the closure tests

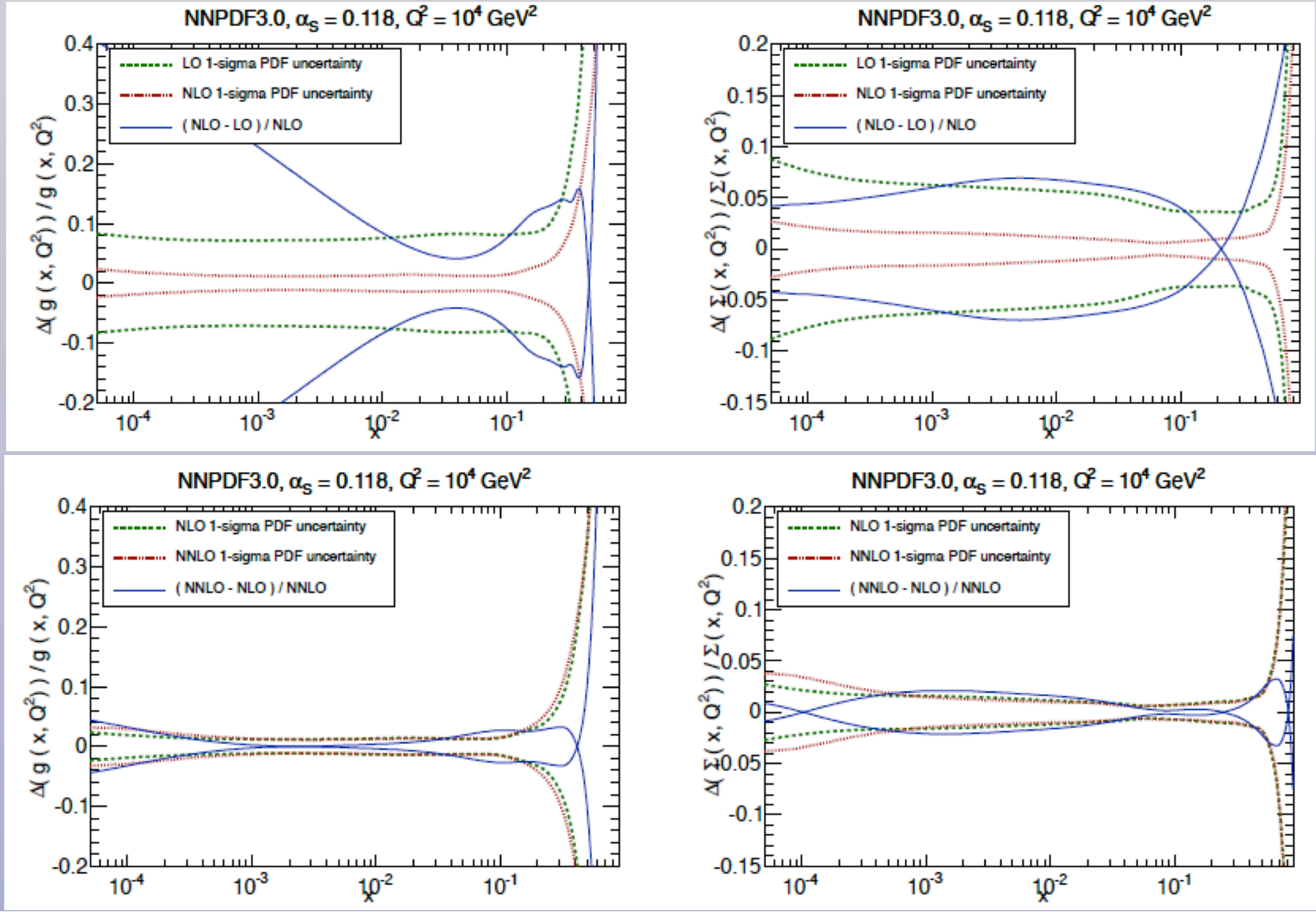
LHC phenomenology



- Reasonable stability between **NNPDF2.3 and NNPDF3.0** for a wide range of LHC cross-sections
- *Conservative* **PDFs** in good agreement with global fit within larger uncertainties

Theoretical uncertainties

- Given the wealth of available **experimental data**, and the refinements in the PDF fitting methodology, a careful assessment of theoretical uncertainties on PDFs is becoming more pressing than ever (talks by Jun and Alberto)
- Rough estimate: compare **difference between subsequent perturbative orders** with **intrinsic PDF uncertainties** to see where theory errors on PDFs could be relevant phenomenologically. At NNLO, **gluon at large x**, **quark sea at medium x, ...**



Summary and outlook

NNPDF3.0 represents a **substantial improvement over NNPDF2.3** both in terms of data, theory and methodology:

- ☑ **Data:** all available **H1 and ZEUS HERA-II data included**, and many new **LHC** measurements from **ATLAS, CMS, LHCb** including W asymmetry, W+charm, inclusive jets, high and low mass Drell-Yan, top quark production, ...
- ☑ **Theory:** Improved **approximate NNLO K-factors for jet data** based on the partial exact NNLO results, **electroweak corrections** included for all relevant data, FONLL-B for NLO sets
- ☑ **Methodology: new C++ code,** fitting strategy validated using **closure tests,** optimized Genetic Algorithms, generalized positivity, improved stopping, PDF fitting basis independence ....

NNPDF3.0 NLO and NNLO for α_S =0.118 in **LHAPDF6** since August

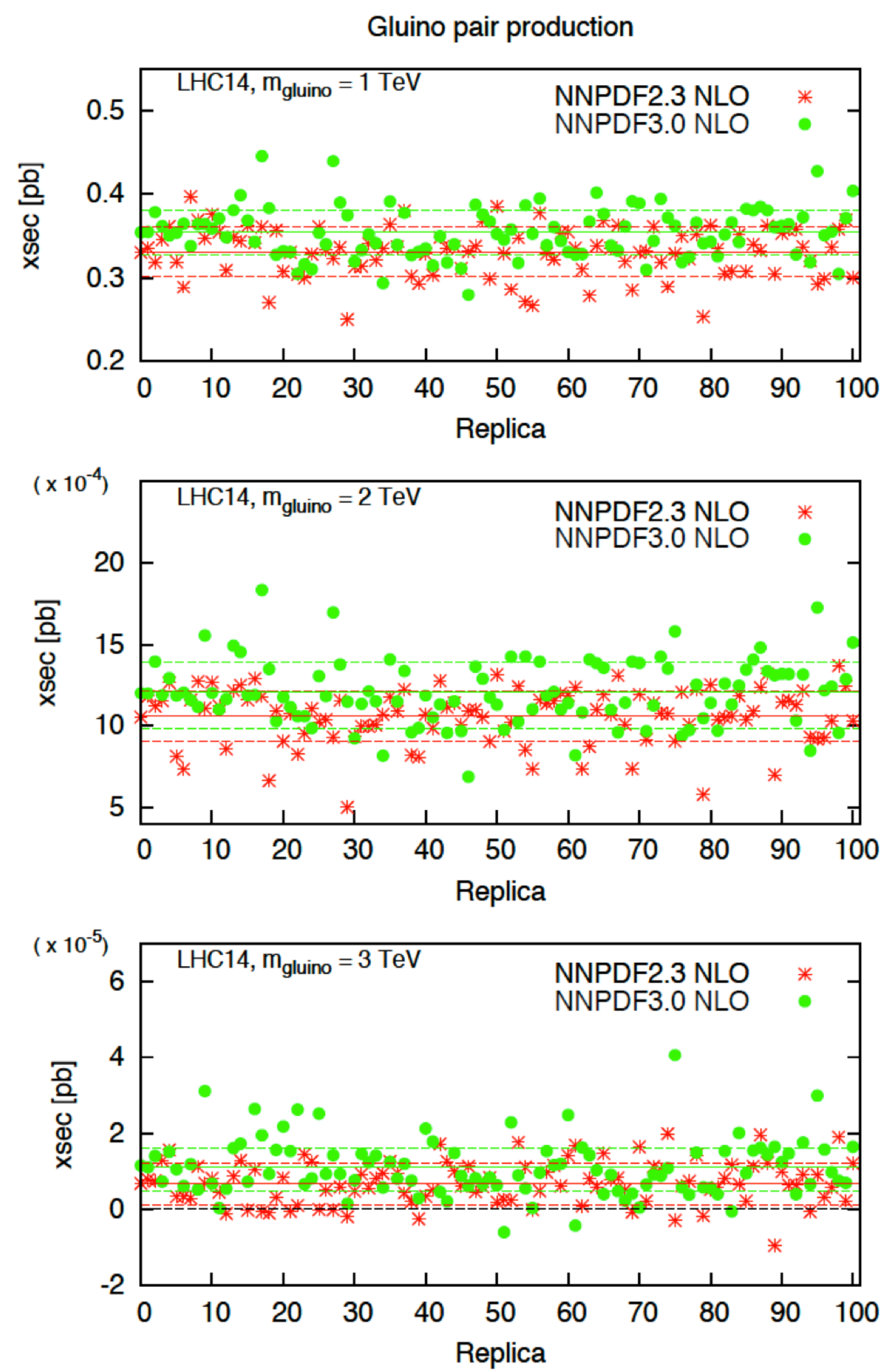
LO PDFs, PDF sets in a **wide range of α_S value**s, and **PDFs based on different datasets,** available in LHAPDF6 by the end of this week

In the medium and long term, **NNPDF development plans**:

- Include all relevant **LHC Run I data**: Complete set of 8 TeV measurements, high pT Z+jets, direct photon production. Then from 2015 also add **LHC Run II data**
- Careful estimate of theoretical uncertainties in global PDF fits
- **Upgrade theory calculations** as they become available: NNLO for top quark differential distributions, exact NNLO for jets and for Z+jets
- Produce NNPDF3.0 sets with **QED corrections**, **intrinsic charm**, threshold and high-energy **resummation,** as well as PDF sets specific for NLO **Monte Carlo event generators**

Extra Material

PDF uncertainties and BSM physics



- While PDFs are not **positive definite beyond LO**, **physical cross-sections** should always be positive
- Implementing this condition, **without overconstraining PDFs with a too restrictive parametrization**, is essential for a reliable estimate of PDF uncertainties
- This is particularly crucial in the **large-x region,** required for the production of **BSM high-mass particles**
- NNPDF3.0 also improves NNPDF2.3 in this respect: the new generalized positivity leads to positive 68% CL intervals for cross-sections in the few TeV scale
- Explicitly verified with **SUSY Prospino calculations**, for LHC 14 TeV, and for the production for squark pairs, gluino pairs and squark-antisquark pairs, up to **M_X =6 TeV**
- Similar checks performed for **very high-mass Drell-Yan production**