NNPDF3.0

Next generation PDFs for the LHC Run II

Juan Rojo
STFC Rutherford Fellow, University of Oxford

Presented on behalf of the **NNPDF Collaboration:**
Richard D. Ball, Valerio Bertone, Stefano Carrazza, Christopher S. Deans, Luigi Del Debbio, Stefano Forte, Alberto Guffanti, Nathan P. Hartland, Jose I. Latorre, JR, and Maria Ubiali

ICHEP 2014
Valencia, 04/07/2014

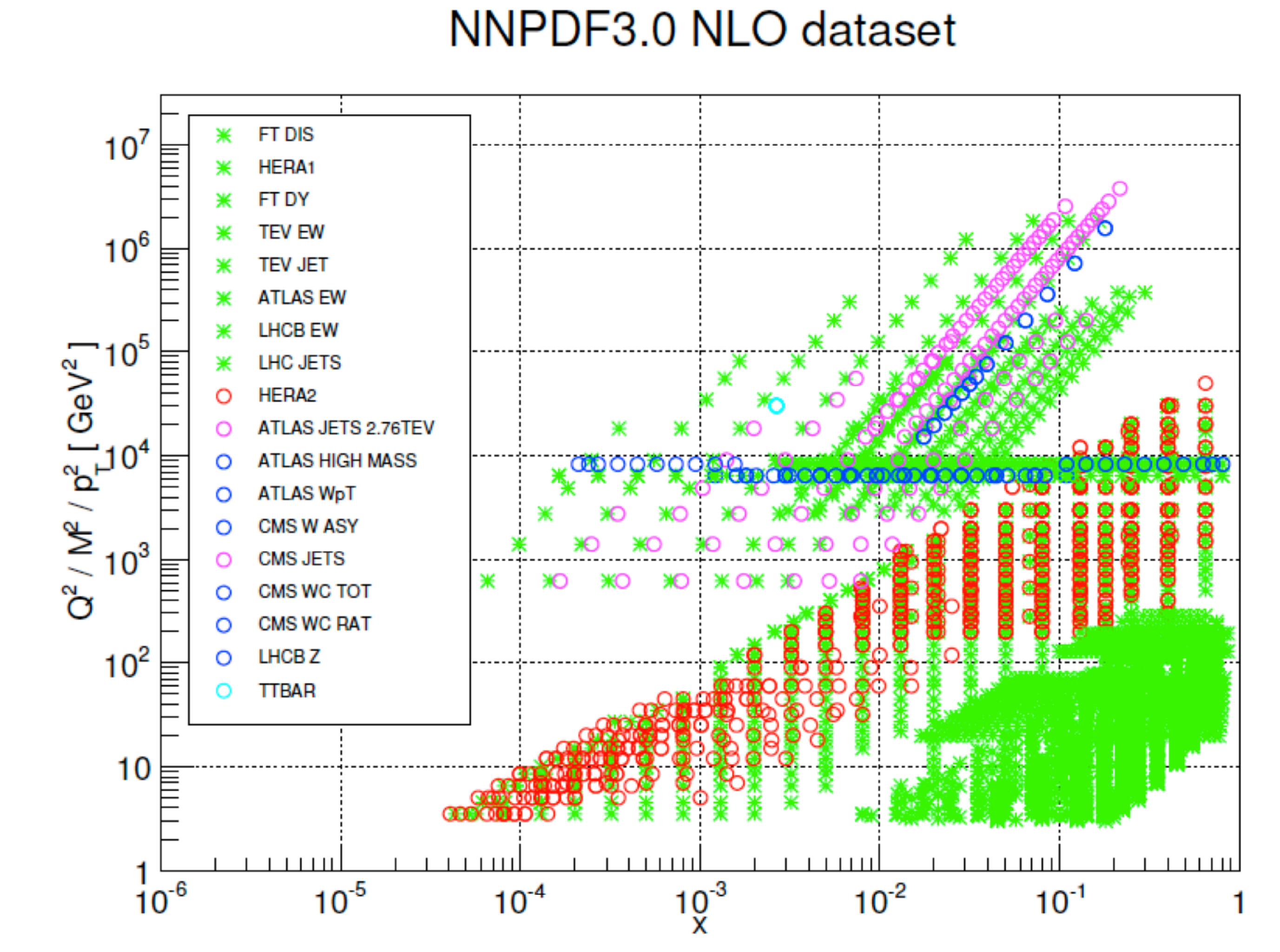
Parton Distributions for LHC Run II

- **PDFs** have been an essential ingredient for **Run I** phenomenology, and will be so even more at the upcoming **Run II at 13 TeV**
- Many crucial LHC analysis benefit from improved PDFs, from **precision Standard Model measurements**, like the W mass determination, and **Higgs boson characterization** to **BSM searches**
- **NNPDF3.0** is the new forthcoming PDF release from the NNPDF Collaboration, a major upgrade that accounts for recent progress in experimental constraints, theory calculations and methodological improvements. **To be released in the next few weeks**.
- In this talk I present **for the first time** the **final NNPDF3.0 sets,** including:
  - ☑ **New experimental data:** HERA-II structure functions, ATLAS and CMS jets, CMS W+charm, ATLAS and CMS Drell-Yan production, top quark production ....
  - ☑ **Improved theory calculations:** Approximate NNLO K-factors for jets, electroweak effects for Drell-Yan data, APPLgrid/FastNLO/aMCfast for all hadronic observables ...
  - ☑ **Fitting methodology:** C++ rewriting of the code, fitting strategy validated on closure tests, extended positivity, optimized Genetic Algorithms minimization ......
  - ☑ **Results** and **implications for LHC phenomenology**

New Experimental Data
LHC & HERA-II

New experimental data

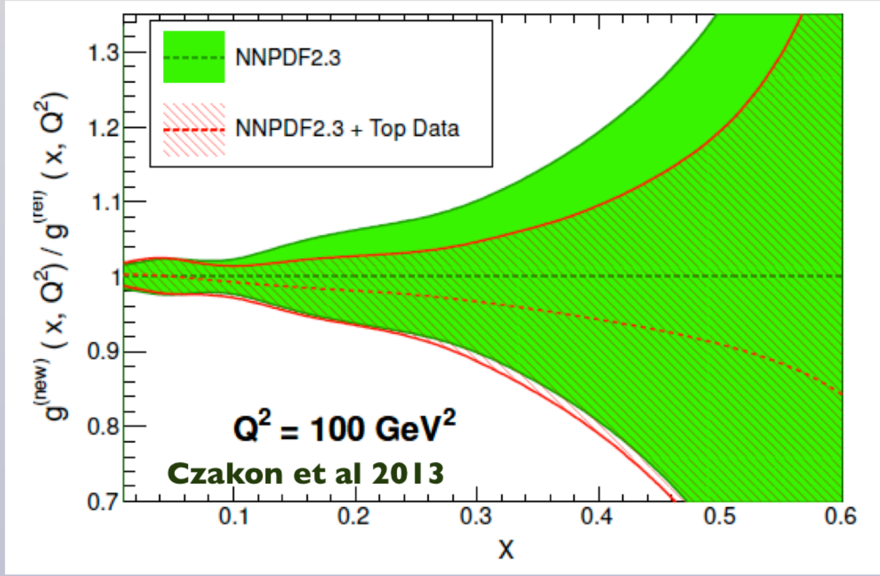
- More than **1000 new data points** from new HERA and LHC **data**
  - ☑ **HERA structure function data:** HERA-II structure functions from H1 and ZEUS, combined HERA F_{2c} cross-sections
  - ☑ **LHC jet data:** CMS **7 TeV inclusive jets** from 2011, ATLAS **2.76 TeV jets** including their correlation 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**

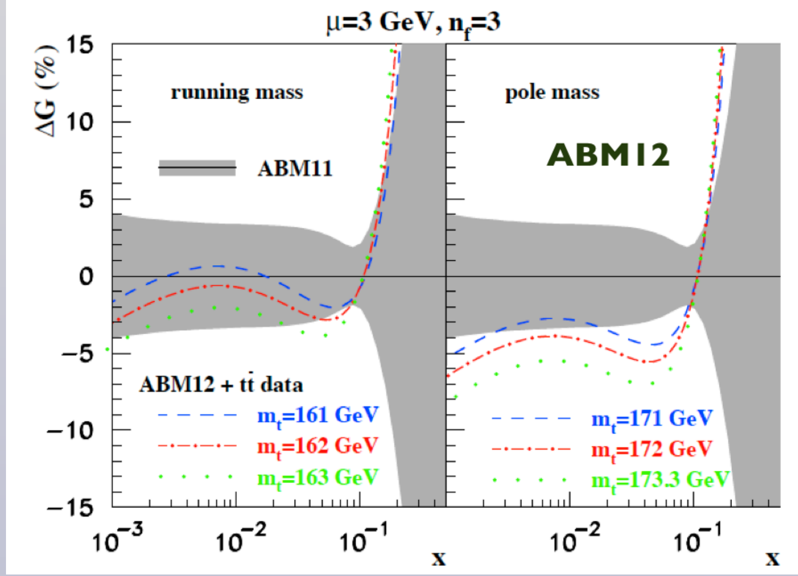


All these datasets **already reasonably well described by NNPDF2.3**

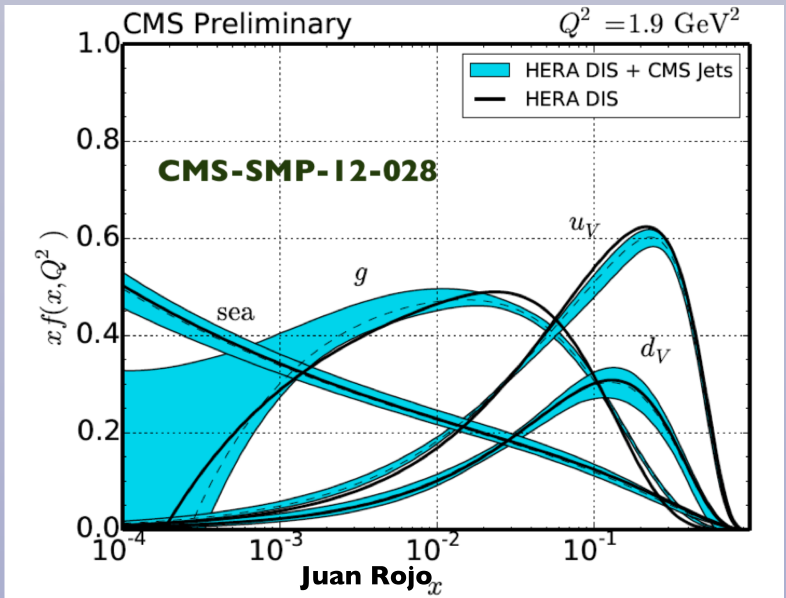
Motivation for new data

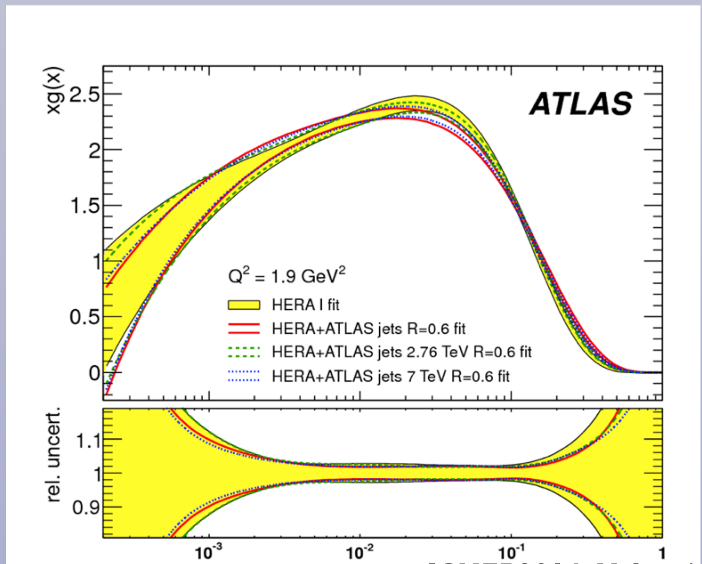
- **Top quark total cross-sections** allow to constrain the **large-x gluon PDF**





- **Jet cross-sections** pin down **medium and large-x gluon** and **large-x quarks**

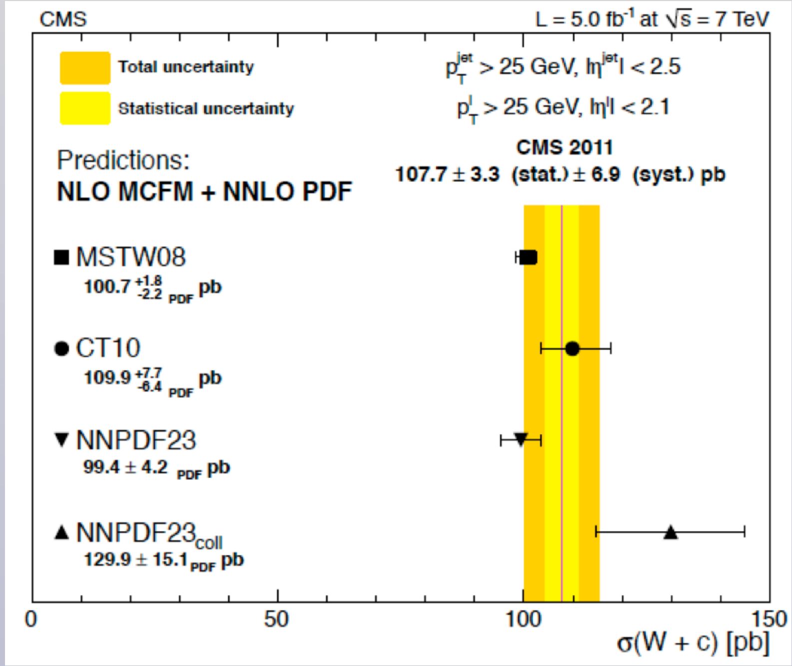


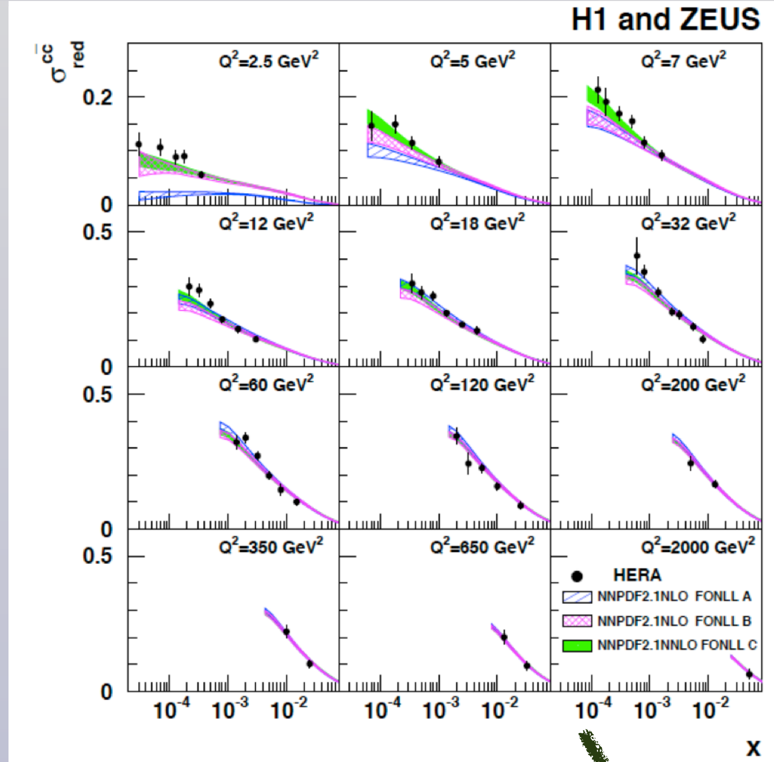


ATLAS data on the **ratio 7 TeV / 2.76 for jet production**, beautiful illustration of the PDF sensitivity of **cross-section ratios between different center-of-mass energies** (Mangano and JR 12)

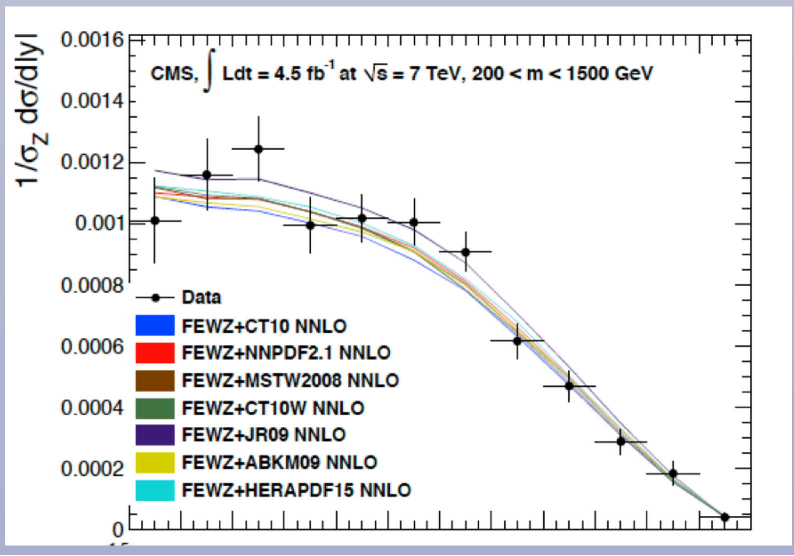
Motivation for new data

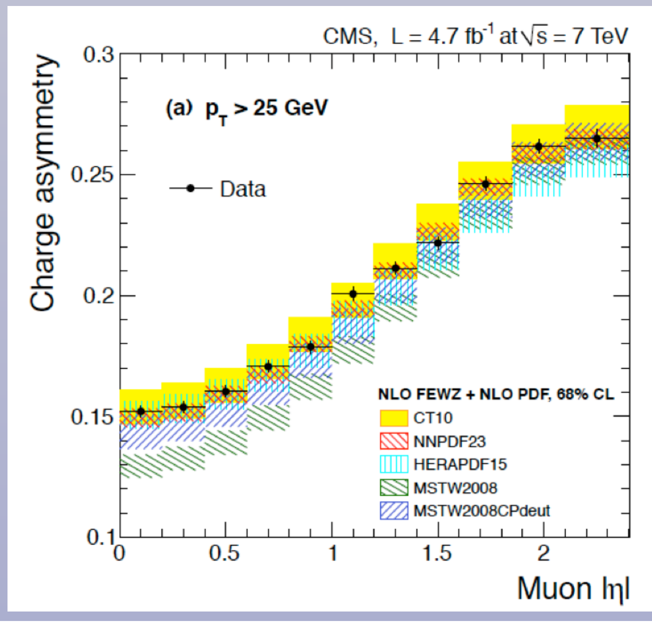
W+charm production data:
directly sensitive to the **strange PDF**





On-shell and off-shell **Drell-Yan data**:
Sensitive to **quark flavor separation**





HERA-II inclusive and charm data:
constrains on **medium and low-x quarks and gluons**

Theory & Tools

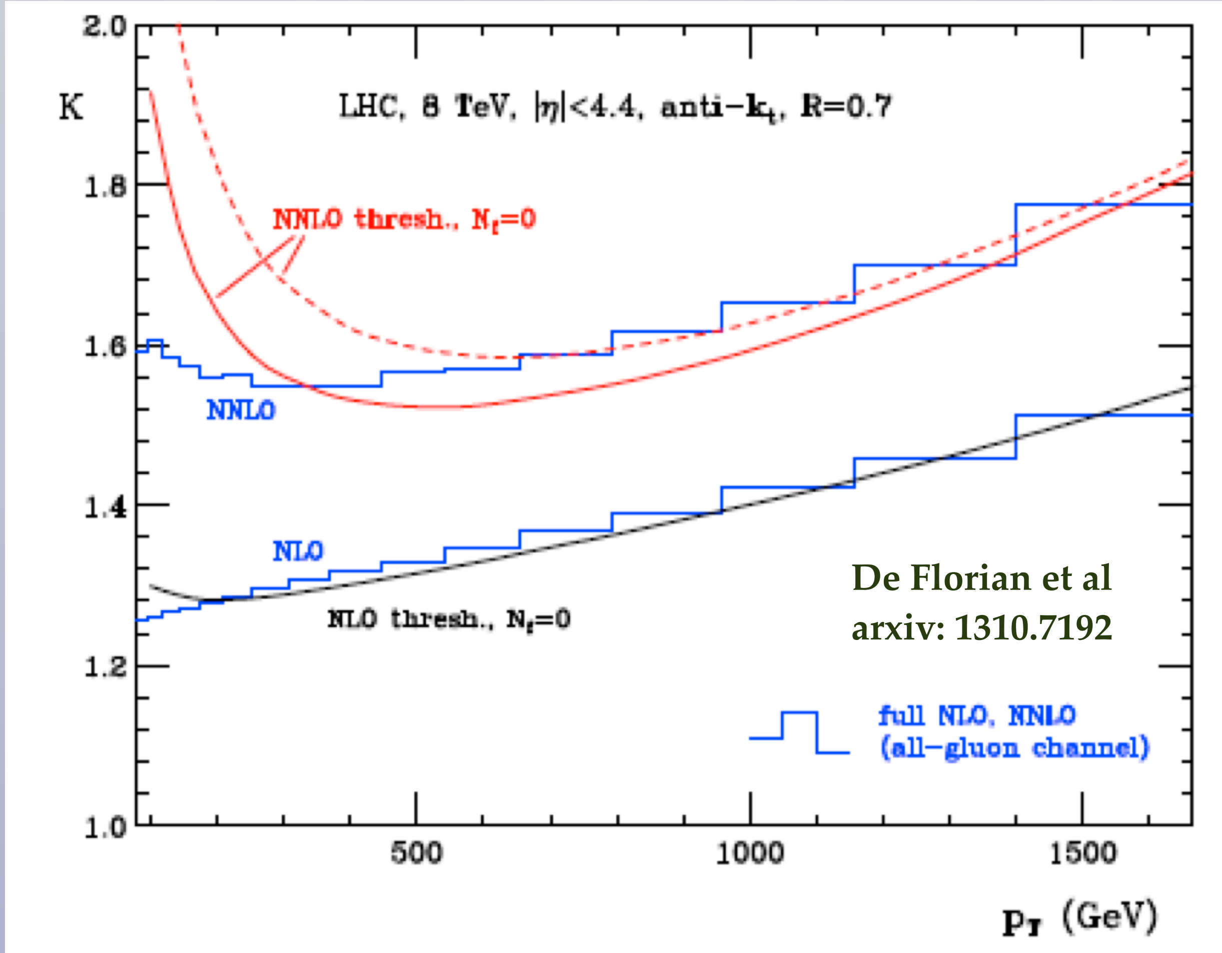
Fast interfaces for NLO calculations

- **(N)NLO QCD calculations** are too CPU-time intensive to be used directly into PDF analysis
- In the recent years, various approaches have been proposed to provide **fast interfaces to NLO calculations**, that can be used directly in PDF analysis, the main ones being:
  - **APPLgrid**: interface to **MCFM** and **NLOJet++** (arxiv:0911.2985)
  - **FastNLO**: interface to **NLOJet++** (arxiv:1109.1310)
  - **aMCfast:** interface to **Madgraph5_aMC@NLO** using the APPLgrid library (arxiv:1406.7693)
  - **FastKernel:** NNPDF internal (arxiv:1002.2312)
- In NNPDF3.0 we systematically use these **fast NLO calculations** for all collider data:
  - **NLOjet++/FastNLO:** CDF and CMS jet data
  - **NLOjet++/APPLgrid:** ATLAS jet data
  - **MCFM/APPLgrid:** ATLAS, CMS and LHCb electroweak gauge boson production, CMS W+charm production, ATLAS and CMS top quark data
  - **Madgraph5_aMC@NLO/aMCfast:** Higgs xsec in gluon fusion for positivity constraints
- For the **NNLO fits**, the NLO calculation are supplemented with **bin-by-bin C-factors** from the corresponding NNLO calculations: **top++** for top data, **FEWZ/DYNNLO** for Drell-Yan data,...

$$C_{\rm fact} = \frac{\hat{\sigma}^{\rm NNLO} \otimes \mathcal{L}^{\rm NNLO}}{\hat{\sigma}^{\rm NLO} \otimes \mathcal{L}^{\rm NNLO}}$$

Jets in NNLO global fits

- The recent calculation of the **gluon-gluon channel NNLO jet cross sections** (arxiv:1310.3993) is an important milestone for inclusion of jet data in NNLO fits: **O(20-25%) enhancements wrt NLO results**
- On the other hand, the **gg channel is small** at medium and large p_T at the LHC energies
- While full NNLO result becomes available, **approximate NNLO** results can be derived from the **improved threshold calculation**: reasonable approximation to exact at large p_T, **breaks down at small p_T**



Our strategy is the following:

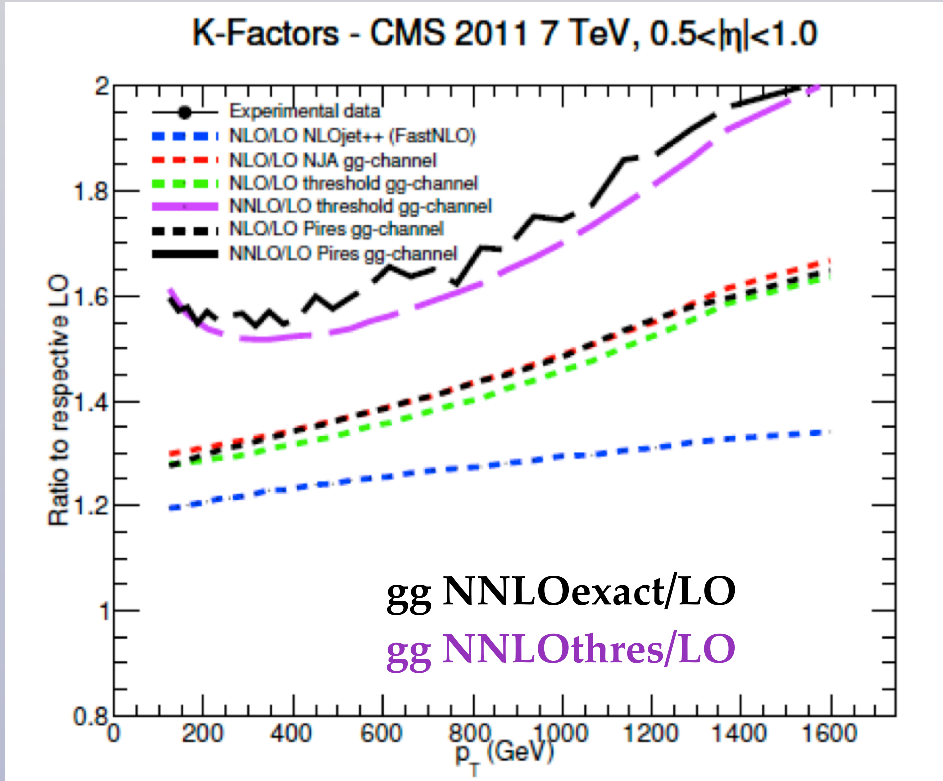
- Compute, for all jet data, the **NNLOexact** and **NNLOapprox** in the **gg channel**
- Use the exact calculation to determine the **range of validity** of the threshold calculation
- With this information, restrict the range of fitted jet data and use the **NNLOapprox calculation** accounting for all partonic channels
- To be **conservative**, we only include data points for which:

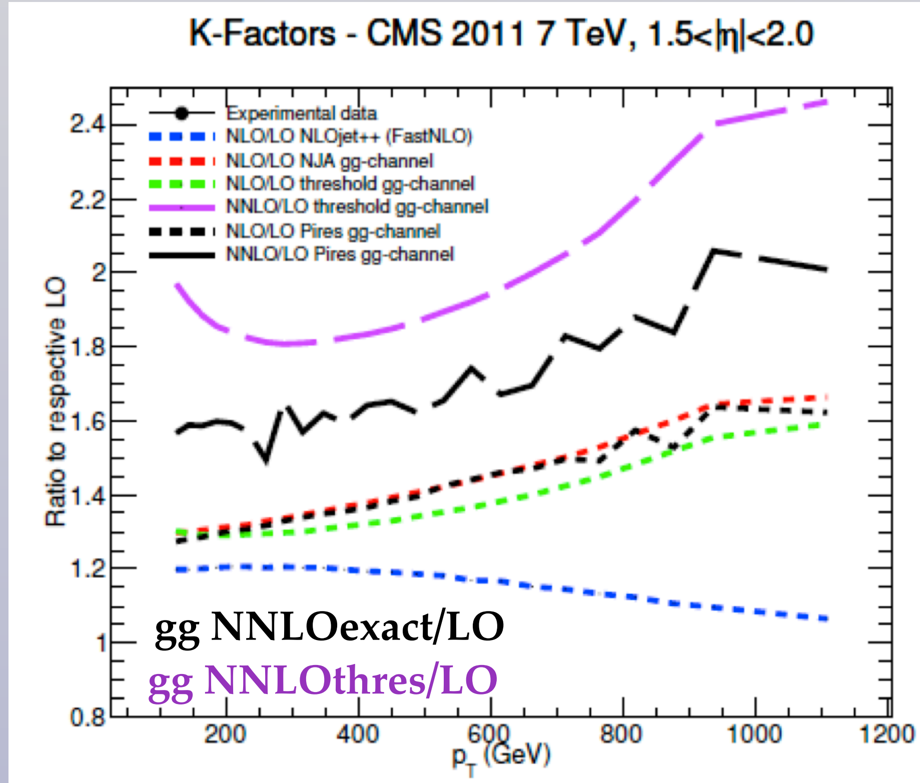
NNLOapprox < (**1.15** NLOexact)

Jets in NNLO global fits

- Therefore, **comparison with exact gg NNLO** can determine for which values of jet p_T and η the **NNLOthres** calculation can be trusted (assumption: NNLO K-factor similar in all channels)

S. Carrazza and J. Pires, in preparation





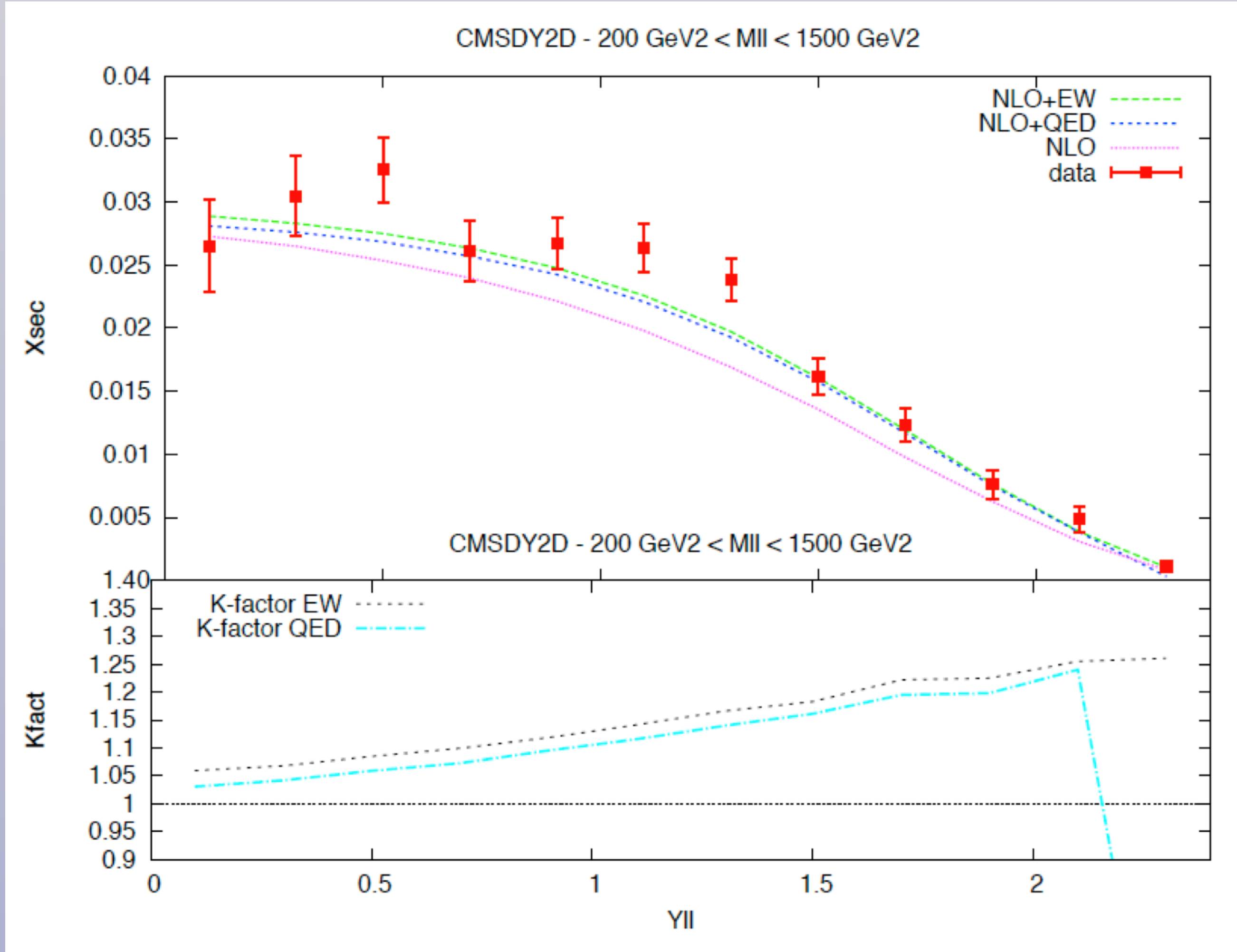
- We **discard jet data** in the kinematic range where the NNLO threshold calculation cannot be trusted
- This restriction will be superseded as soon as **full exact NNLO** results available
- This approximation allows to **keep a maximum of jet data in the NNLO fit**, without this data PDF uncertainties in the large-x gluon are much larger

Electroweak corrections

- Electroweak corrections are important for **W and Z** production specially at **large invariant masses and/or transverse momentum**

- In NNPDF3.0, full NLO electroweak corrections for all **neutral current Drell-Yan** datasets have been computed with **FEWZ3.1**

- As an illustration, **high-mass Drell-Yan** from CMS 7 TeV:



Electroweak corrections up to **25% at the highest dilepton masses**

QED corrections also large, but affected by the large uncertainties of **γ(x,Q)**

QED corrections not included in the NNPDF3.0 QCD-only fit (NNPDF3.0 QED to follow)

For 14 TeV, including EW corrections for most datasets will be mandatory

Methodology
&
Closure Testing

NNPDF3.0 is really NNPDF++

Completion of a major software development project: rewriting of most of the **NNPDF fitting framework** from **Fortran 77** to **C++** and **Python**

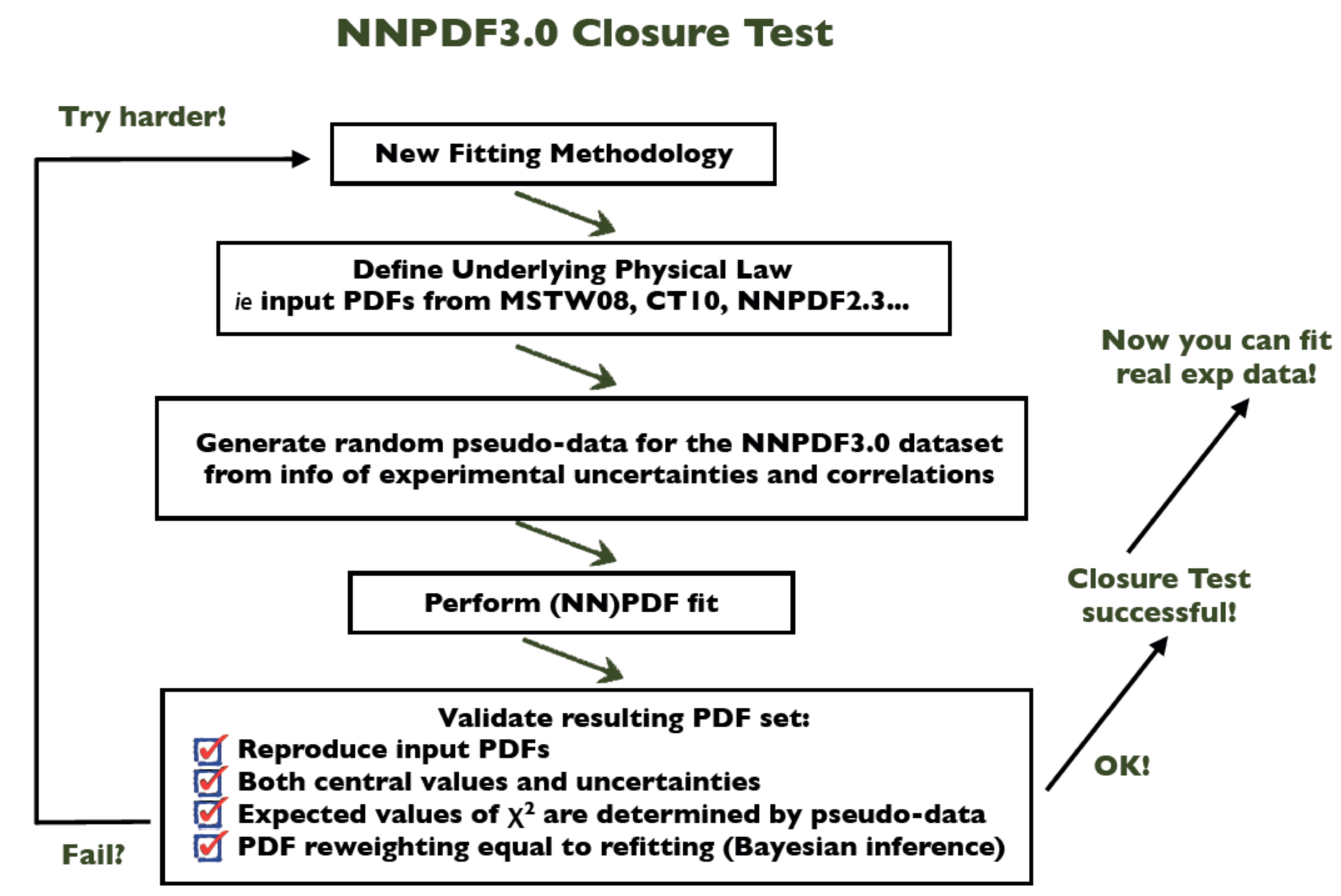
Lines of code by language in NNPDF3.0 fitting framework

Language	files	blank	comment	code
C++	106	6993	6048	26551
Fortran 77	113	115	10161	20872
C/C++ Header	134	1183	857	3920
make	34	792	447	1699
ASP.Net	1	511	0	1390
Bourne Shell	23	261	202	802
Python	8	187	168	565
Fortran 90	1	32	43	117
Bourne Again Shell	3	7	11	34
SUM:	423	10081	17937	55950

- ☑ **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**
- ☑ Fits can now be easily parallelized to run in **clusters** and in **Graphical Processing Units**
- ☑ Guarantees **robustness and stability for NNPDF development** in the medium and long term

Closure Testing

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



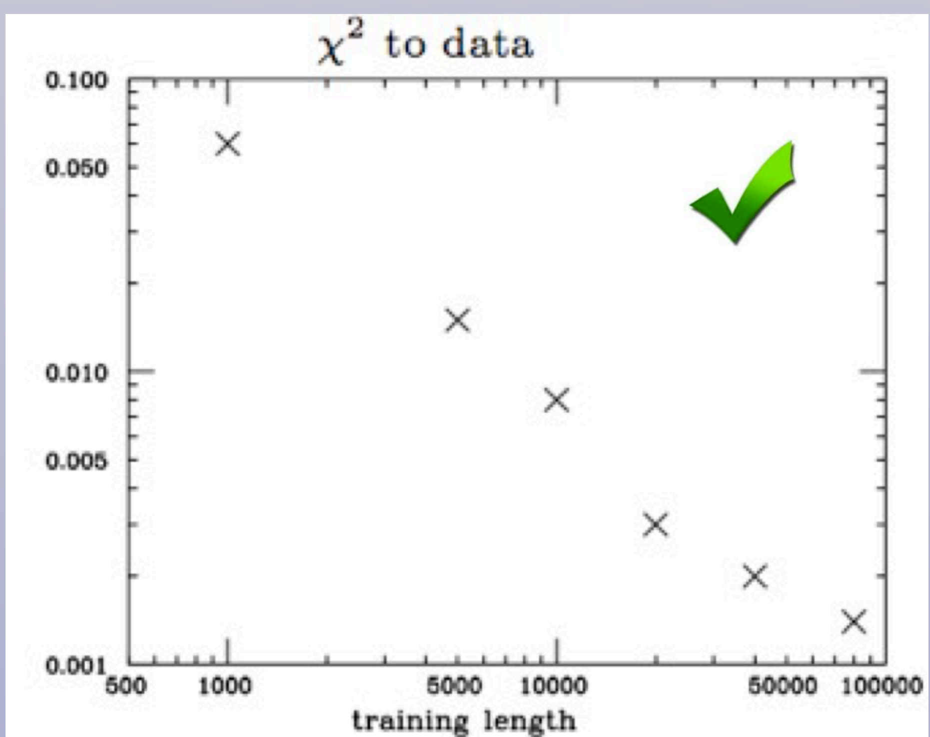
Closure Testing

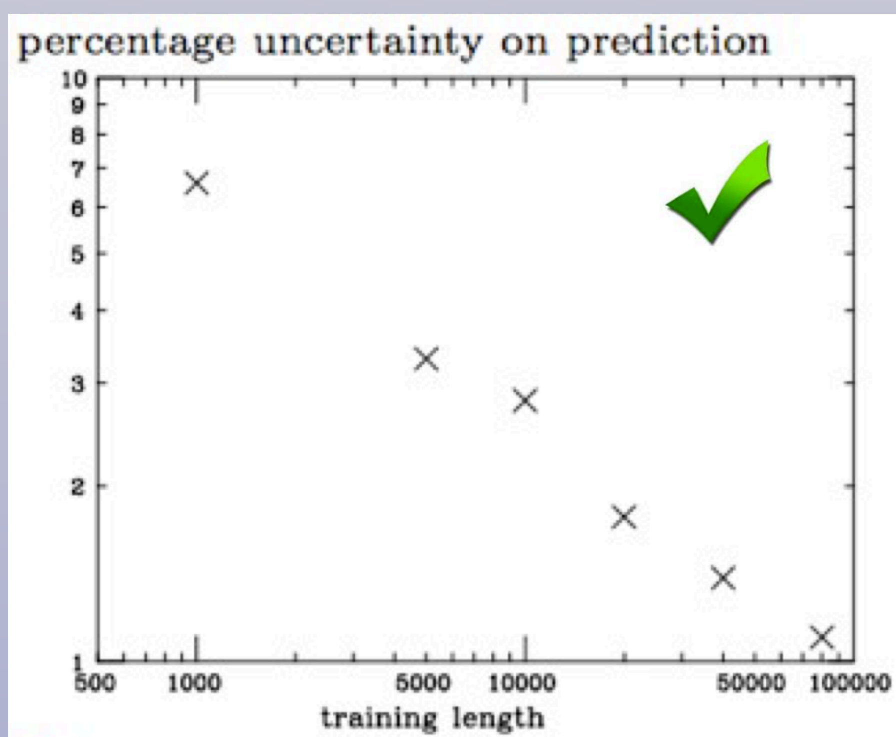
Validation and optimization of fitting strategy performed on **closure test** with **known** underlying PDFs

Three levels of closure tests:

- **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

Example: **Level 0 closure tests -** Fit results successfully **converge towards underlying law**: central **χ^2 to pseudo-data tends to zero**, same for PDF uncertainties on predictions (all replicas converge on same underlying law)

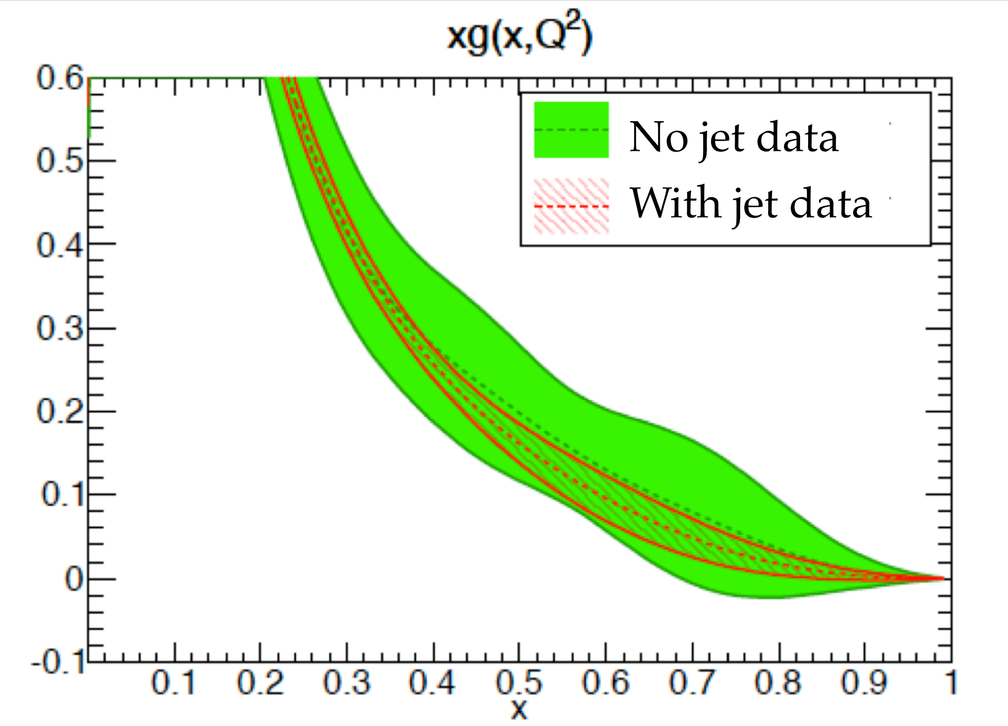




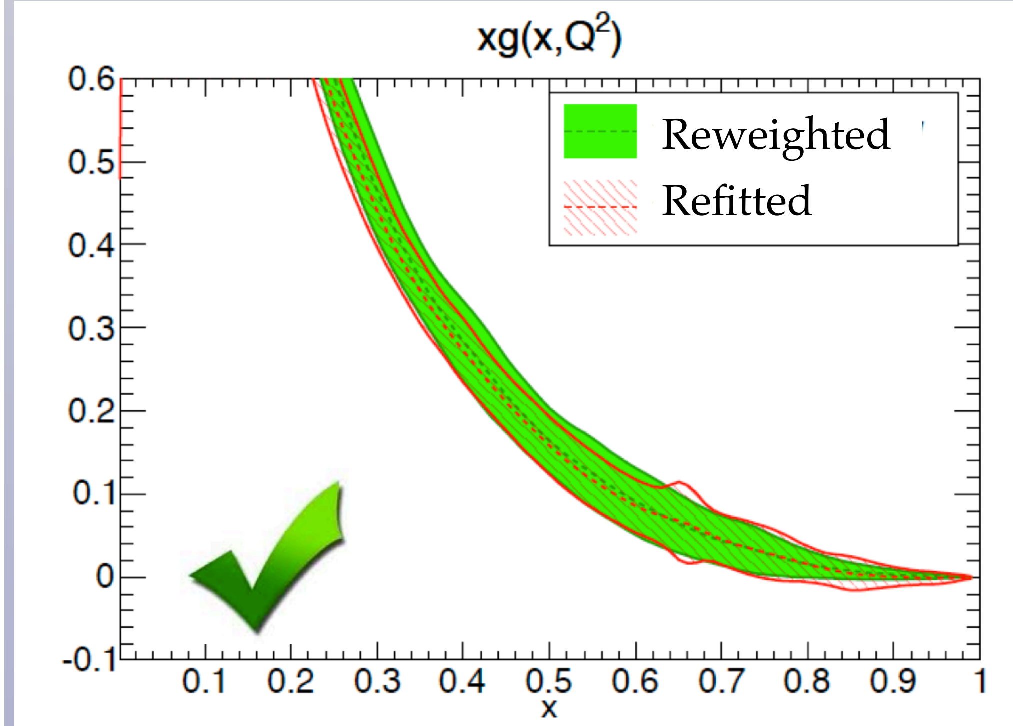
PDF reweighting

- **Statistically identical results** should be obtained when **refitting** or when **reweighting,** even more so in a **closure test**
- Thus **Bayesian inference** can be used as the **ultimate closure test**, sensitive to **all the moments** of the refitted / reweighting PDFs, not only central value and error
- As an illustration, compare effects of **CDF, D0 and ATLAS jet data** included by refitting and by reweighting. Compare both **NNPDF** and **Giele-Keller** prescriptions.

Fit wo jet data vs Fit with jet data



Fit with jet data vs RW with jet data



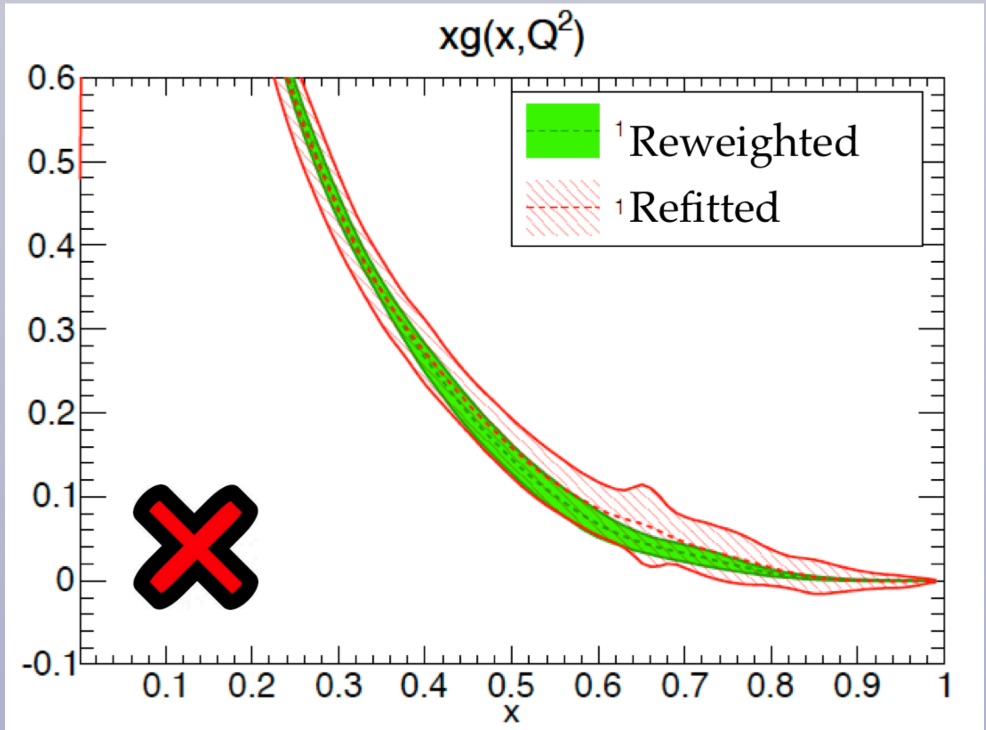
Note large uncertainties on gluon from missing jet data

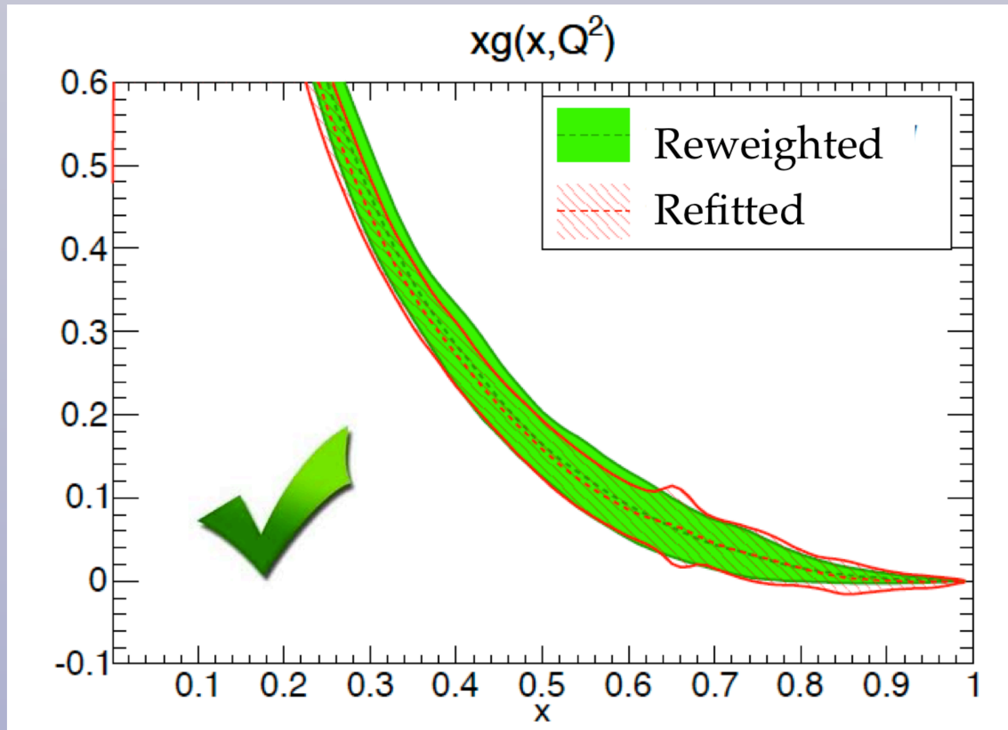
PDF reweighting

- **Statistically identical results** should be obtained when **refitting** or when **reweighting,** even more so in a **closure test**
- Thus **Bayesian inference** can be used as the **ultimate closure test**, sensitive to **all the moments** of the refitted/reweighting PDFs, not only central value and error
- As an illustration, compare effects of **CDF, D0 and ATLAS jet data** included by refitting and by reweighting. Compare both **NNPDF** and **Giele-Keller** prescriptions.

Giele-Keller $$w_k^G = \frac{e^{-\frac{1}{2}\chi_k^2}}{\frac{1}{N}\sum_{k=1}^N e^{-\frac{1}{2}\chi_k^2}}.$$

NNPDF $$w_k = \frac{(\chi_k^2)^{\frac{1}{2}(n-1)} e^{-\frac{1}{2}\chi_k^2}}{\frac{1}{N}\sum_{k=1}^N (\chi_k^2)^{\frac{1}{2}(n-1)} e^{-\frac{1}{2}\chi_k^2}}$$

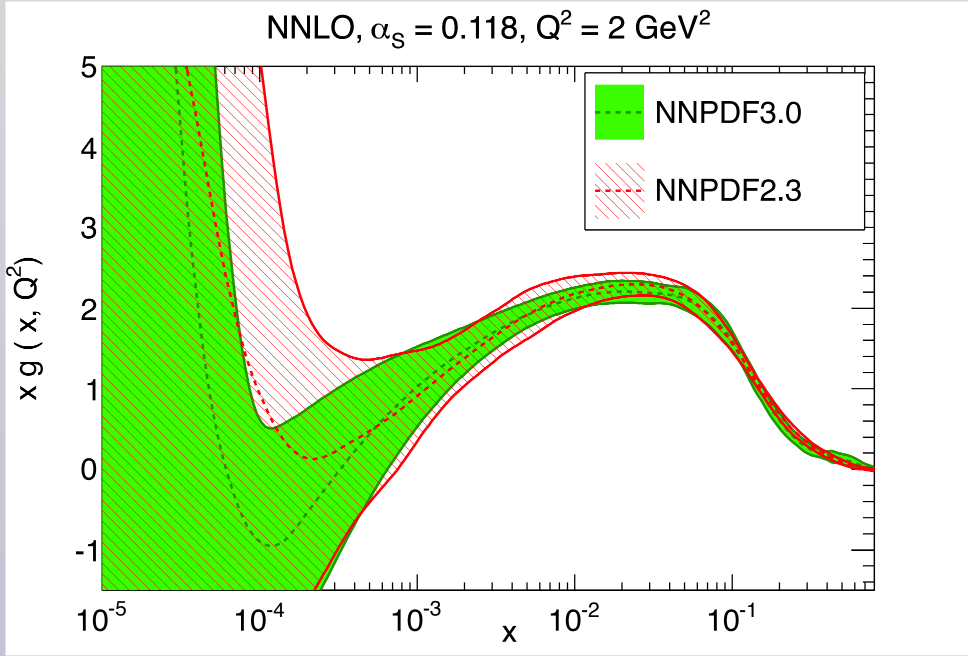




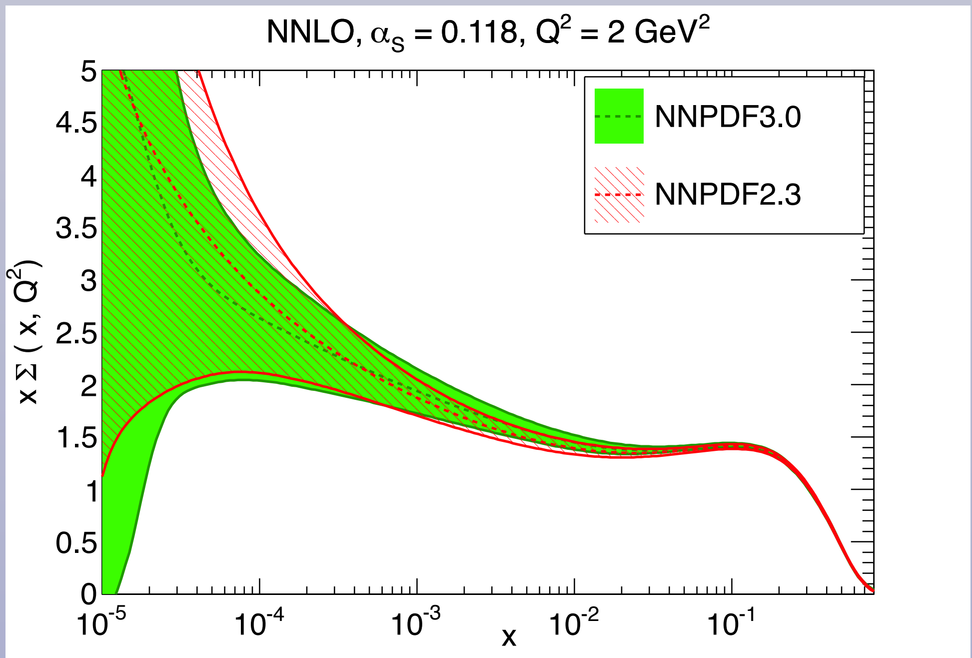
Refitted vs Reweighted

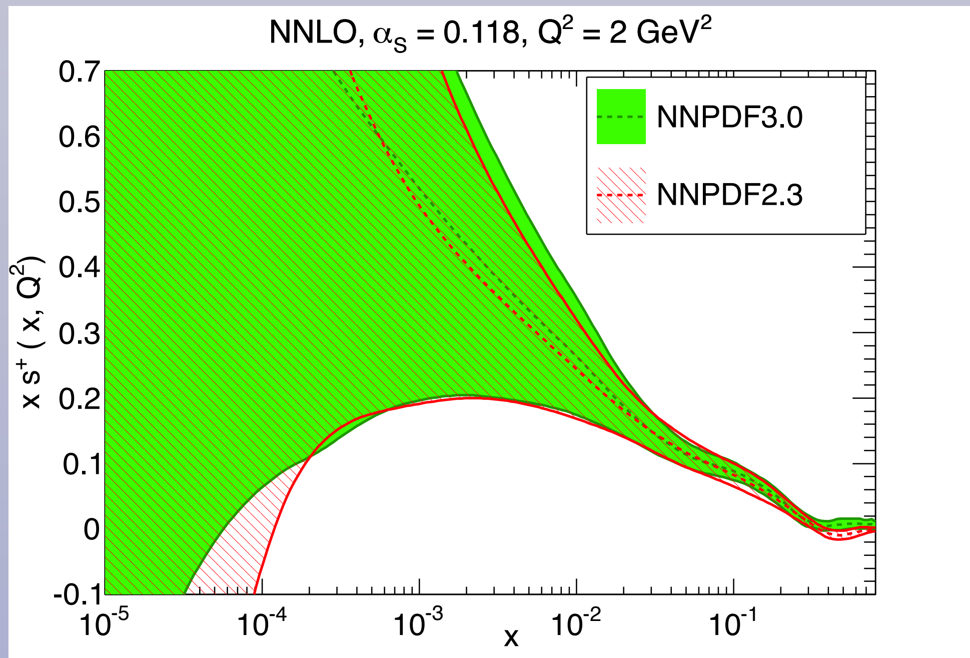
NNPDF3.0
Results &
Implications for the LHC

The NNPDF3.0 parton distributions

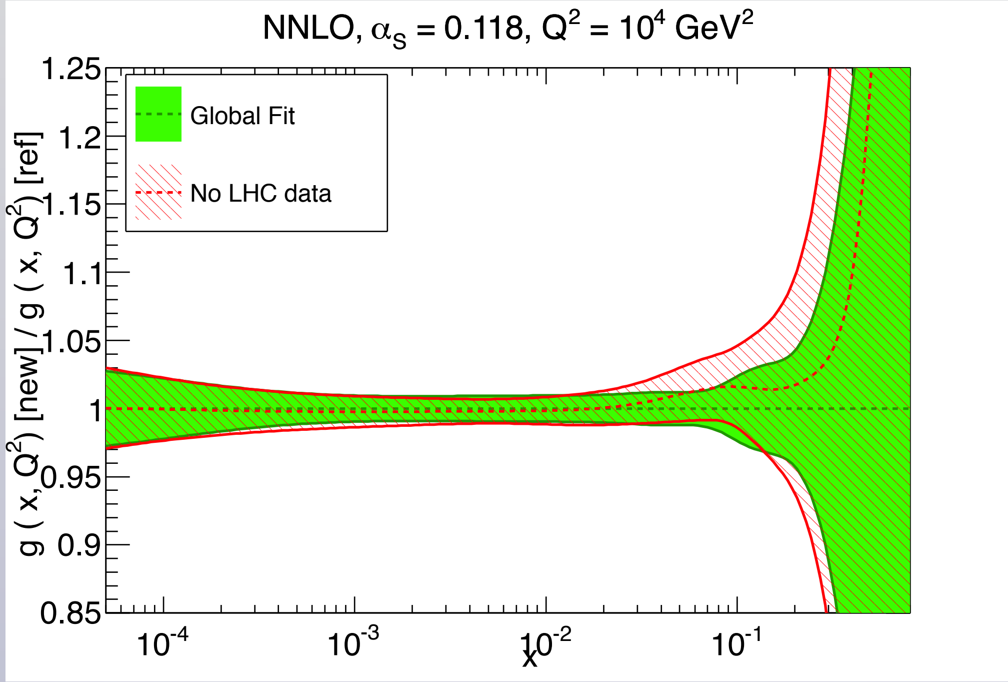


- Reasonable agreement between NNPDF2.3 and NNPDF3.0: as expected, since **all the new HERA and LHC data** already well described in NNPDF2.3
- Differences between PDFs at the 1-sigma level at most: impact of **new data** and of **updated theory and methodology**
- **PDF uncertainties are reduced** in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

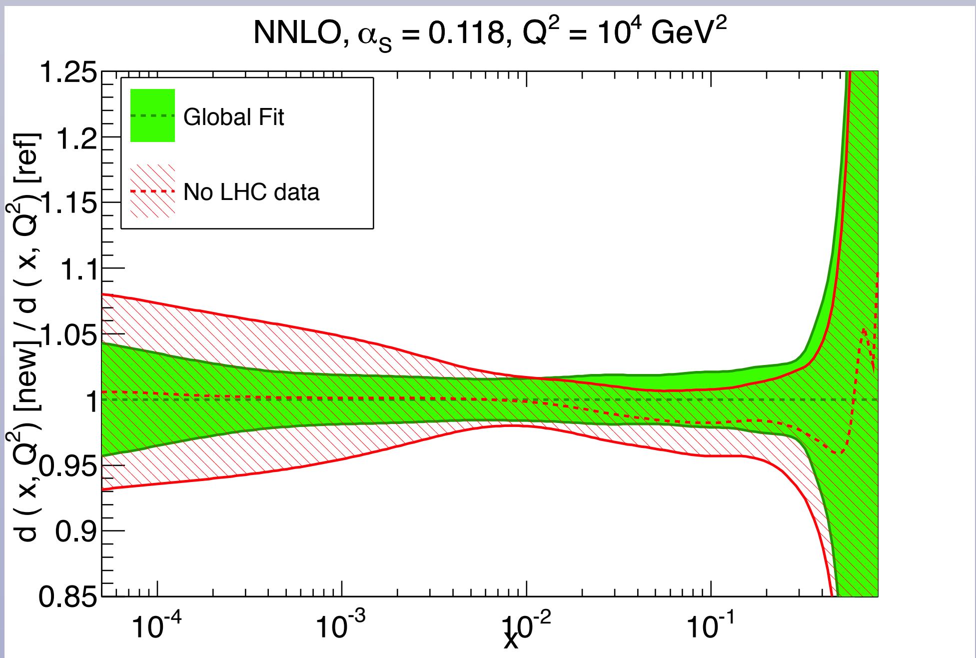


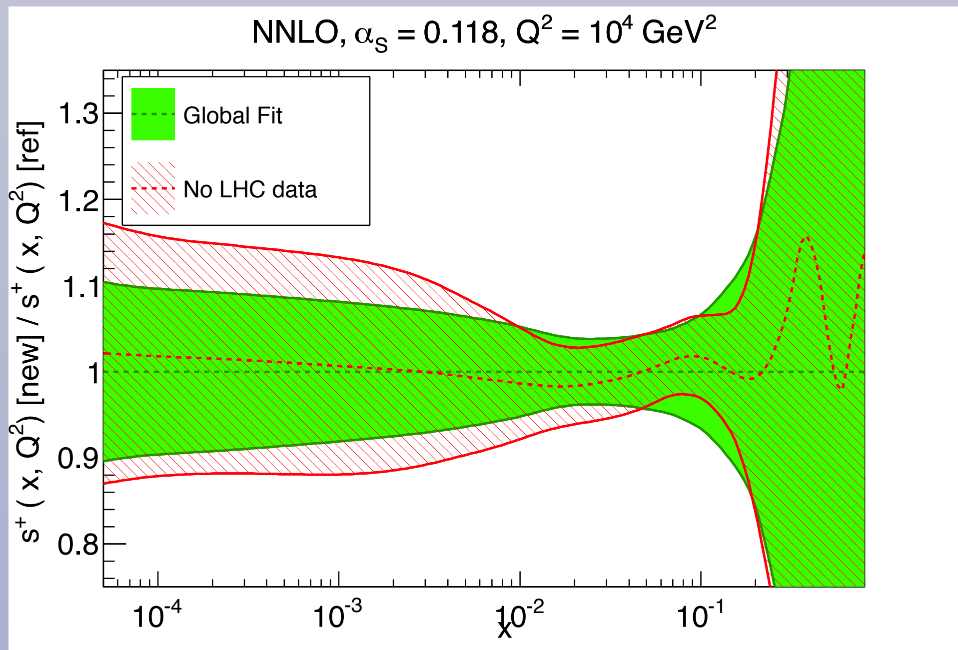


Impact of LHC data

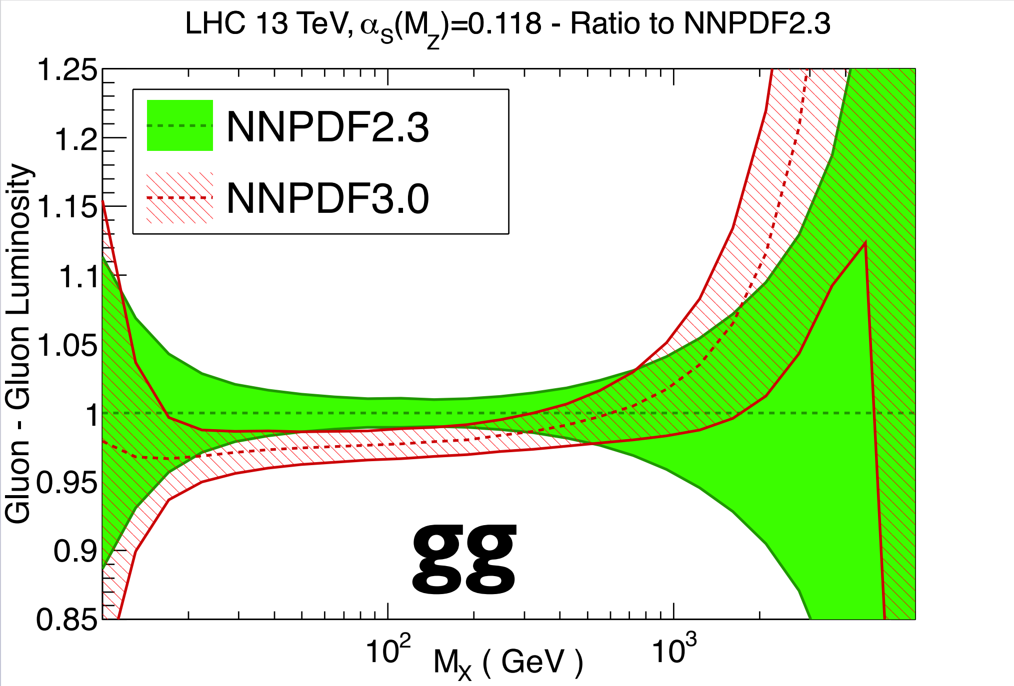


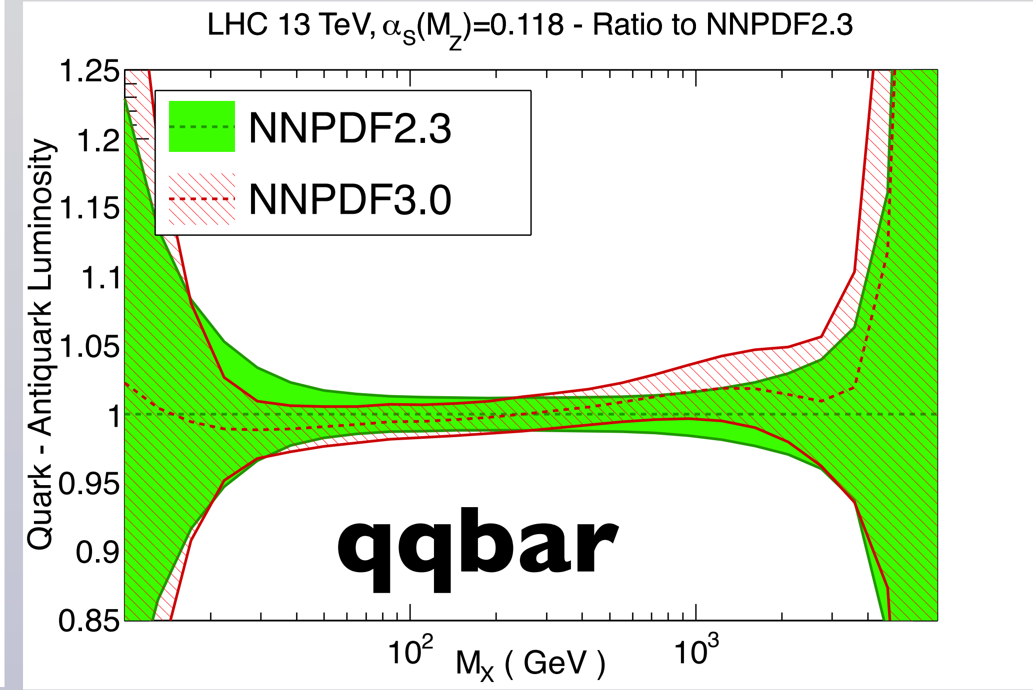
- Compare global NNPDF3.0 fit with a fit **without LHC data**
- PDF uncertainties on **large-x gluon** reduced due to **top quark and jet data**
- PDF uncertainties on **light quarks** reduced from the **Drell-Yan** and **W+charm data**
- The **description of all new LHC data**, already good in NNPDF2.3,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

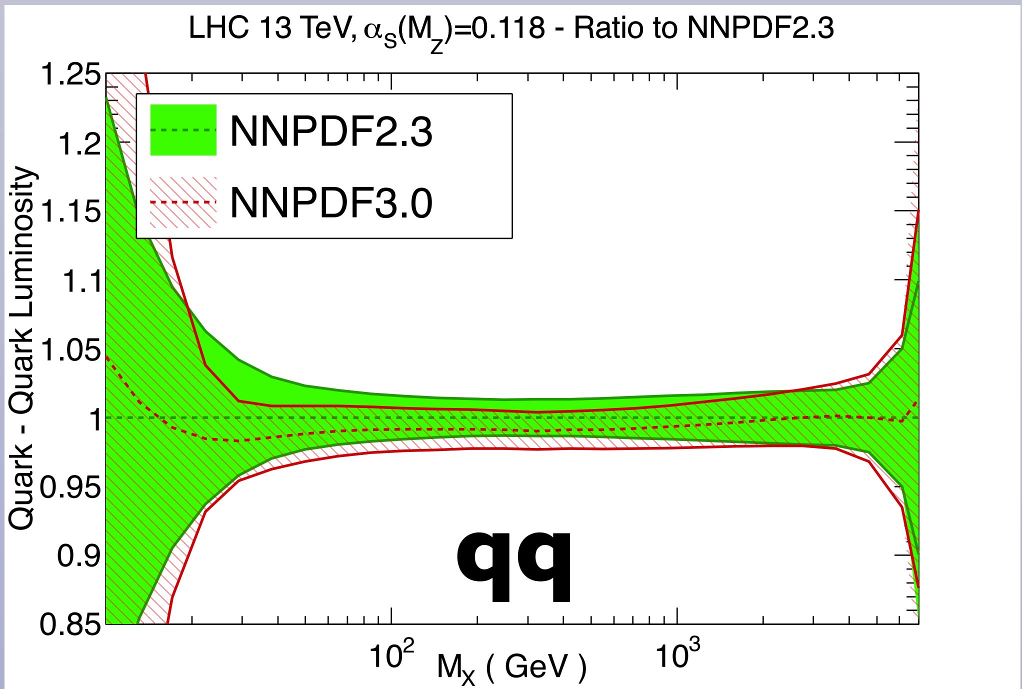




PDF luminosities

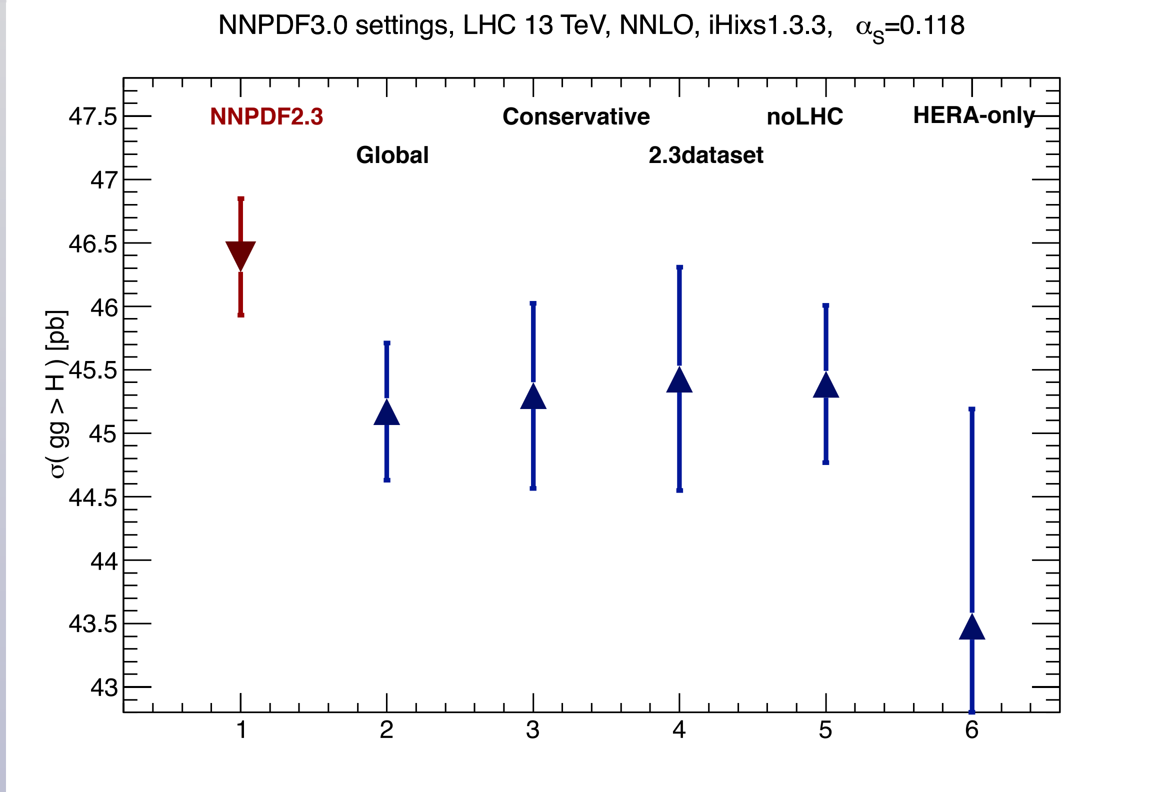






- **PDF luminosities** are useful to translate differences in PDFs into differences in LHC cross-sections
- For **quarks,** luminosities are similar between 2.3 and 3.0. **Antiquarks** in 3.0 at large masses are harder than in 2.3
- More differences for the **gluon-gluon luminosity**. NNPDF3.0 softer by about 1-sigma wrt NNPDF2.3 for M< 500 GeV: **implications for Higgs production**

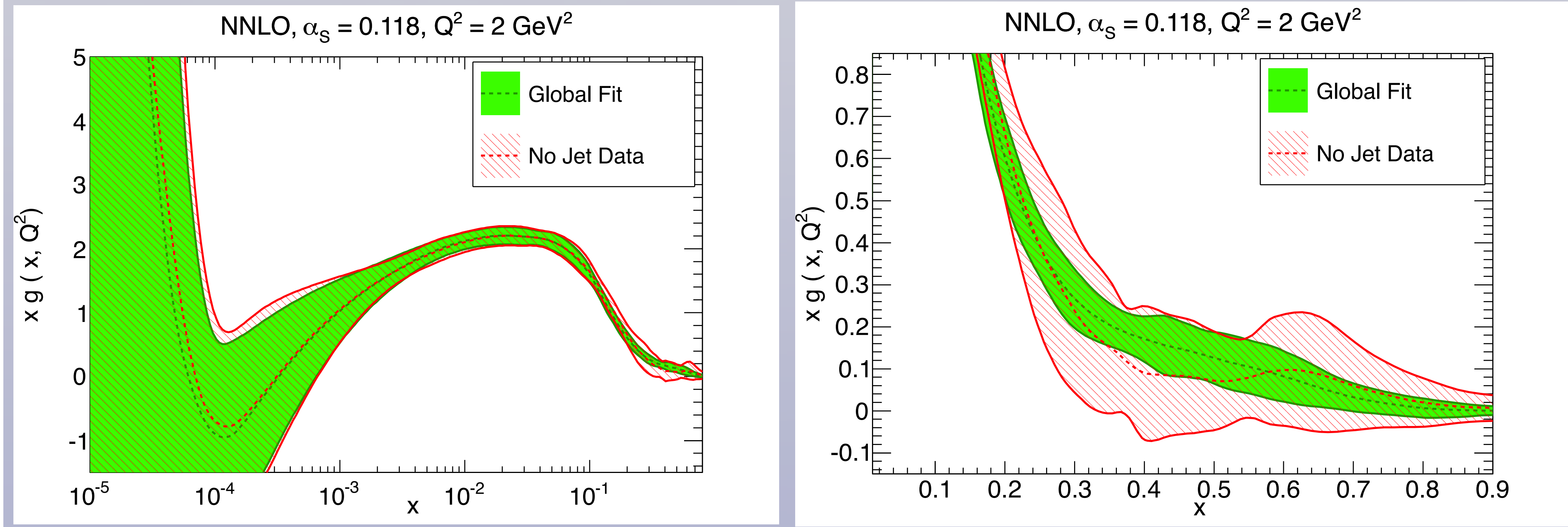
Higgs production in gluon fusion



- The softer gg luminosity in NNPDF3.0 leads to a **decrease in the ggH xsec** at the LHC 13 TeV
- This effect arises from the combination of the **improved methodology**, based on closure tests, and the **new HERA-II and LHC data**
- Remarkable **stability of the NNPDF3.0 predictions** with respect to the fitted dataset
- Interesting to compare with predictions of the **upcoming updates** from CT, HERAPDF and MSTW.
- In the pipeline: systematic comparison of NNPDF2.3 and NNPDF3.0 for a **wide range of LHC observables** using **MadGraph5_aMC@NLO**

Impact of jet data

- **Removing all jet data** from the NNLO global fit (until exact NNLO available) is not an option: substantial increase in **large-x gluon PDF uncertainties**
- Using the **NNLO threshold calculation**, benchmarked with the **exact NNLO results the gg channel**, allows to make the most of the **Tevatron and LHC jet data** until full NNLO result available
- Remarkably, in **NNPDF3.0 the central value for g(x) in jetless fit** is within 1-sigma of the global fit result: **consistency** between jet data and all other datasets



Summary & Outlook

Summary and outlook

NNPDF3.0 is the new upcoming release from the NNPDF collaboration.

It 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:** fitting strategy validated using **closure tests,** optimized Genetic Algorithms, extended positivity, fast Bayesian regularization, PDF fitting basis independence ....

LO, NLO and NNLO sets for a range of α_S **values** will become available in **LHAPDF6**

Also PDF sets based on **different datasets,** PDFs sets in **different VFN schemes,** ...

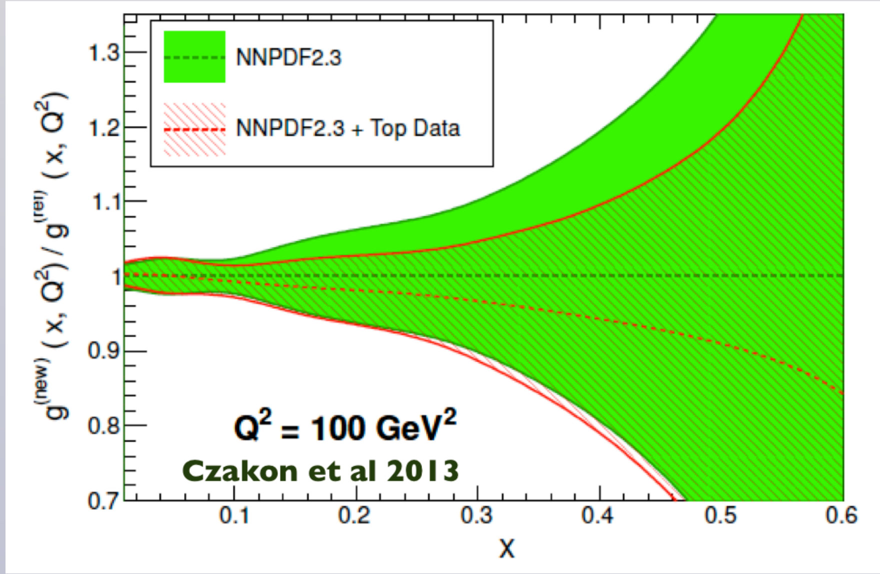
In the medium and long term, **NNPDF development plans**:

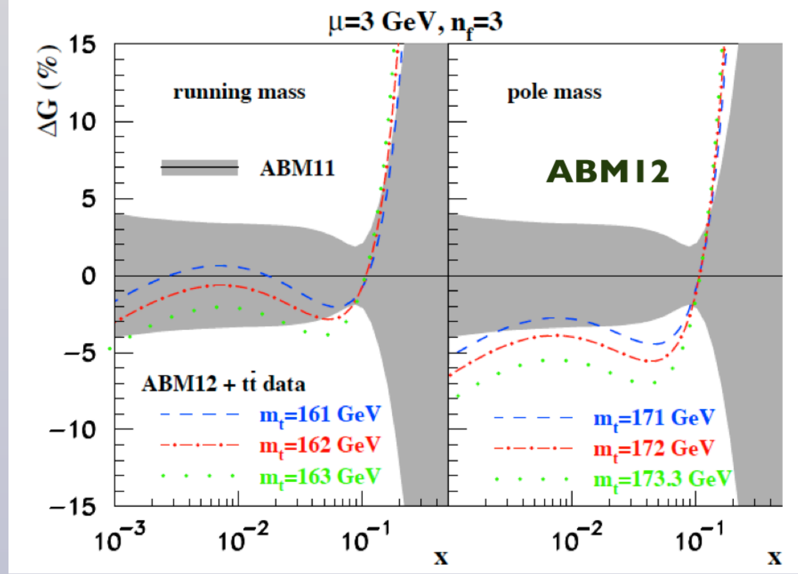
- Include all relevant **LHC Run I data**: Complete set of 8 TeV measurements, high pT Z+jets, direct photon production. Then from 2015 also add **LHC Run II data**
- **Upgrade theory calculations** as they become available: NNLO for top quark differential distributions, exact NNLO for jets and for Z+jets
- 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

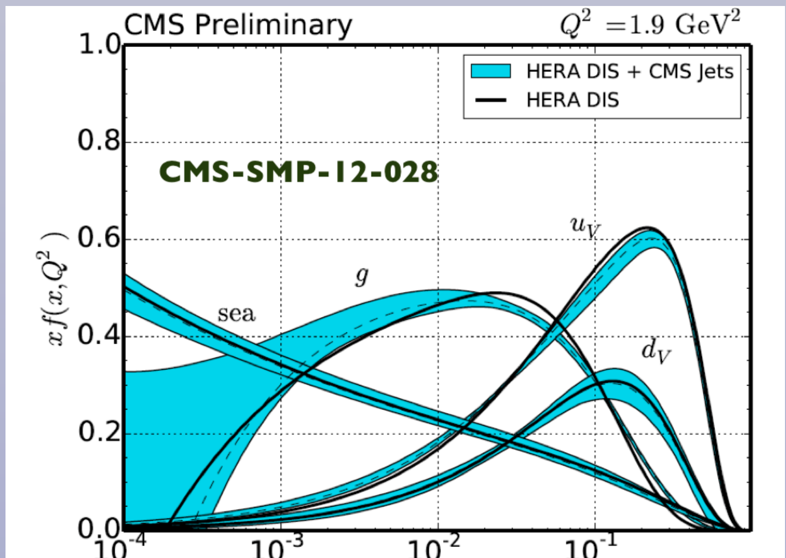
Motivation for new data in NNPDF3.0

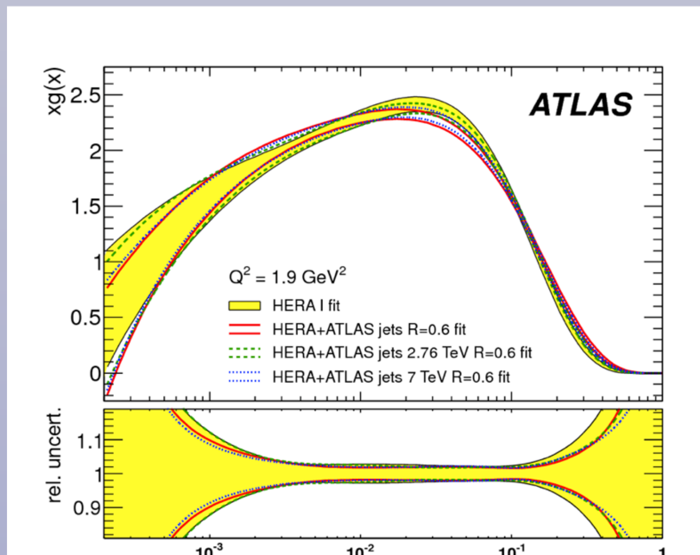
- **Top quark total cross-sections** allow to constrain the **large-x gluon PDF** (NNLO for differential distributions will be available soon, should be able to include as well differential top production measurements)





- **Jet cross-sections** pin down **medium and large-x gluon** and **large-x quarks** (important to include properly information on NNLO corrections, see later)



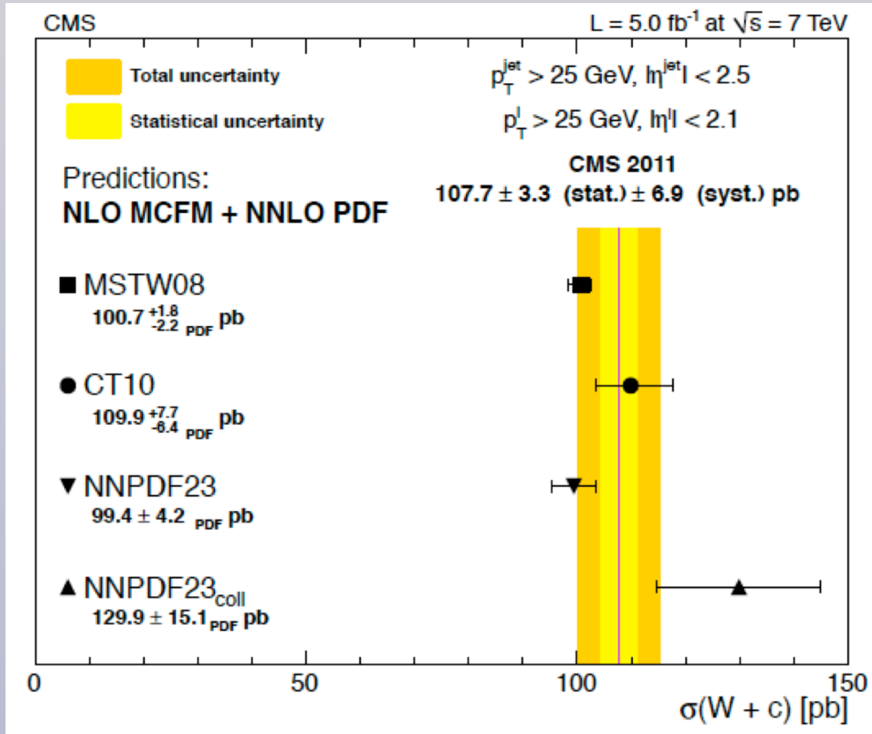


ATLAS data on the **ratio 7 TeV / 2.76 for jet production**, beautiful illustration of the PDF sensitivity of **cross-section ratios between different center-of-mass energies** (Mangano and JR 12)

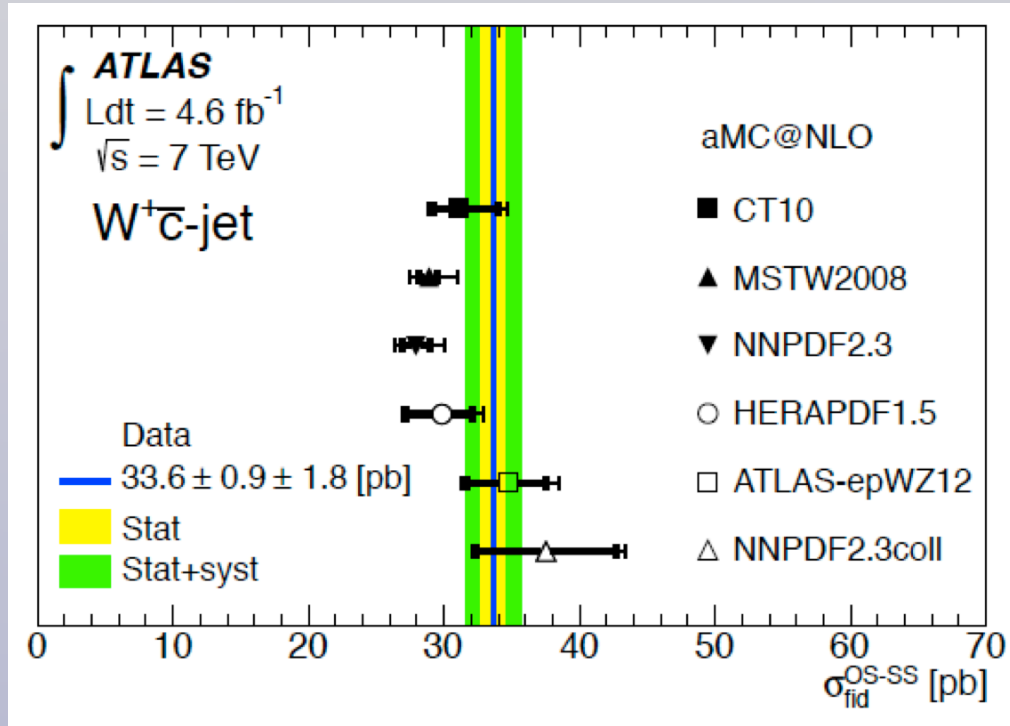
Motivation for new data in NNPDF3.0

- **W+charm** production data directly sensitive to the strange PDF
- Measured by ATLAS (arxiv:1402.6263) and CMS (arxiv:13101138) with somewhat opposite (?) conclusions

CMS: strange suppression in agreement with DIS data



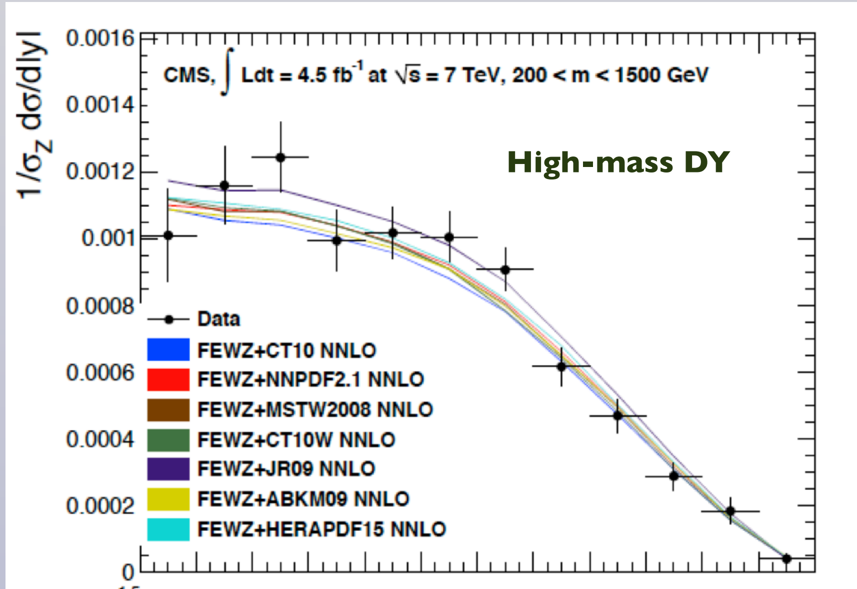
ATLAS: light quark sea symmetric preferred

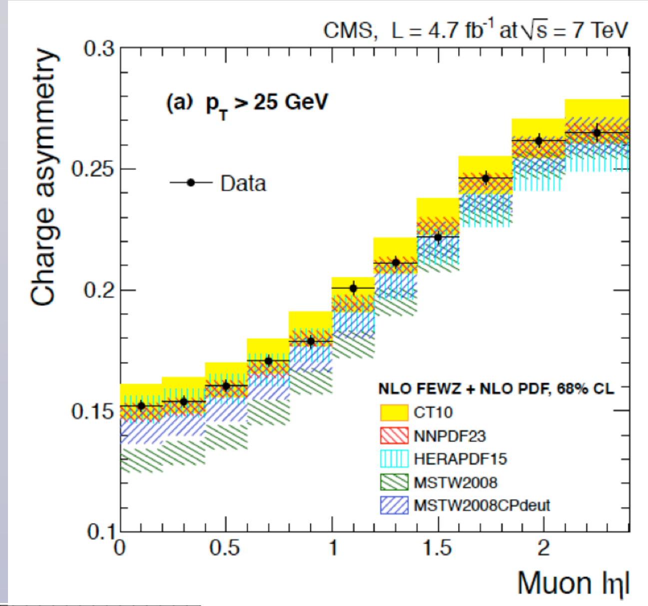


- However, only in the context of a **global fit** the **optimal** value for strangeness can be determined
- A recent analysis in the **ABM framework** (arxiv:1404.6469) suggest that **fits with symmetric strangeness** cannot describe properly fixed target DIS and Drell-Yan data (see also R. Thorne in past PDF4LHC) and that one can fit ATLAS data with still a suppressed strangeness (same as found in NNPDF2.3 for **incl W,Z**)
- The **NNPDF3.0** will perform a similar analysis, with the advantage of using a **completely flexible parametrization for s(x,Q)**, which in other analysis uses a very restrictive functional form

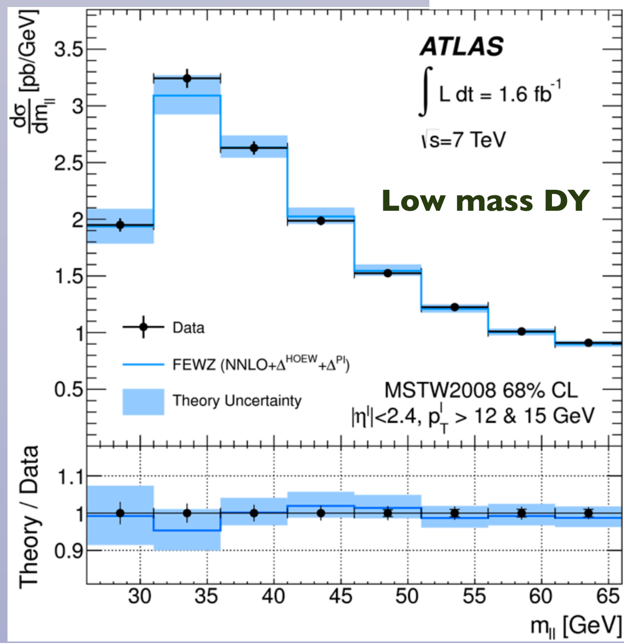
Motivation for new data in NNPDF3.0

- **Electroweak gauge boson production** is an essential measurement for **quark flavor separation**
- The relevance of LHC W,Z production is even greater in **collider-only fits**
- Data on the **Drell-Yan process** at **low and high masses** allow to extend the kinematical coverage in Bjorken-x





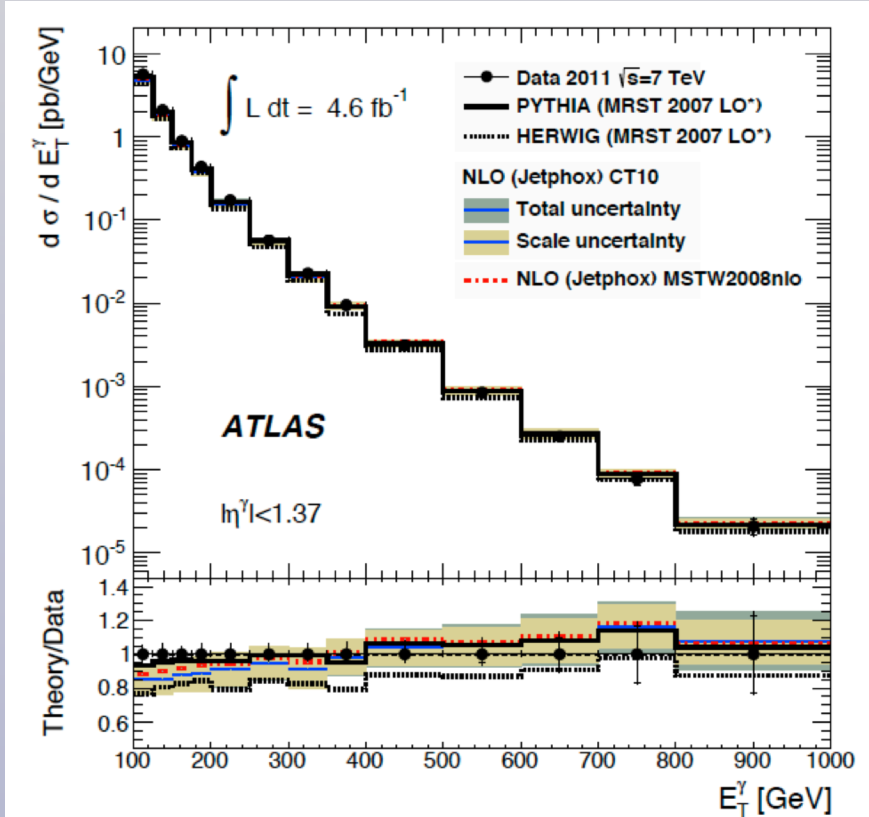
W asymmetry



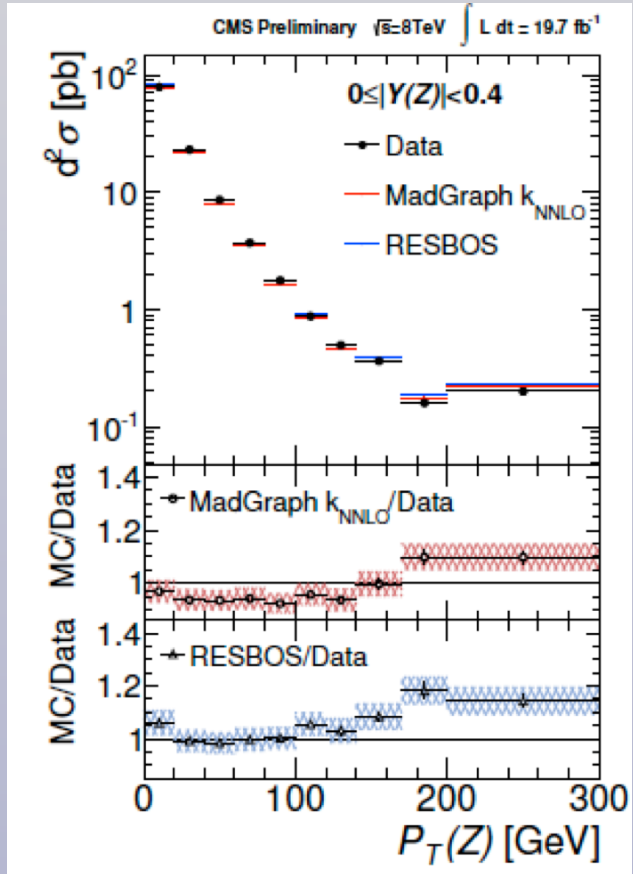
Given the data precision, NNLO corrections are essential, and also **electroweak corrections** at high masses

More data in the pipeline

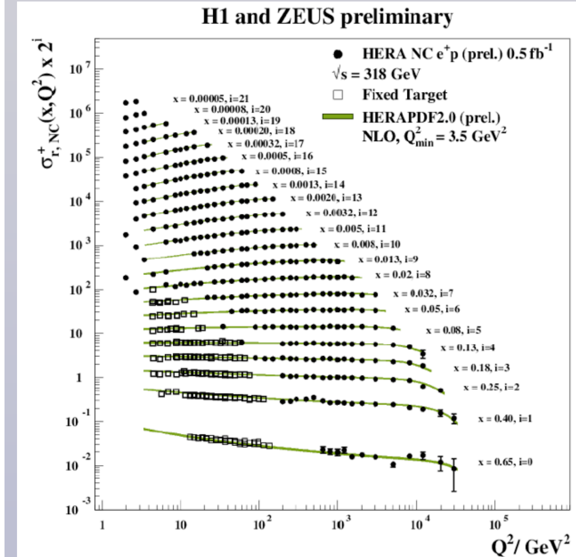
- **The plethora of new LHC data** that is becoming available for PDF fitting makes any PDF fit somewhat outdated shortly after it has been released
- At some point we need to put a **cut-off** about the data to include in NNPDF3.0
- Some **important measurements** that we might try to add in time for NNPDF3.0 include



ATLAS direct photon data from 2011 run
But need APPLgrids and K-factors, already available for the ATLAS analysis



CMS-SMP-13-013
Measurement of the Z pt spectrum at high pt
But data still preliminary
(see Markus's talk)



Combined HERA-II data
To supersede the separated H1 and ZEUS HERA-II data already included in 3.0

Independence of PDF fitting basis

- **Predictions for physical observables** should be independent of the specific choice of PDF fitting basis
- We have explored in **closure tests** that thanks to the **improved NNPDF3.0 methodology**, we achieve almost statistically equivalent fits using **two very different basis**

NNPDF2.3 basis

$$
\begin{aligned}
\Sigma(x,Q_0^2) &= \sum_{i=1}^{3}(q_i+\bar{q}_i)(x,Q_0^2) \\
T_3(x,Q_0^2) &= \left(u+\bar{u}-d-\bar{d}\right)(x,Q_0^2) \\
\Delta_S(x,Q_0^2) &= \left(\bar{d}-\bar{u}\right)(x,Q_0^2) \\
s^+(x,Q_0^2) &= (s+\bar{s})(x,Q_0^2) \\
V(x,Q_0^2) &= \sum_{i=1}^{3}(q_i-\bar{q}_i)(x,Q_0^2) \\
s^-(x,Q_0^2) &= (s-\bar{s})(x,Q_0^2)
\end{aligned}
$$

NNPDF3.0 basis

$$
\begin{aligned}
\Sigma(x,Q_0^2) &= \sum_{i=1}^{3}(q_i+\bar{q}_i)(x,Q_0^2) \\
T_3(x,Q_0^2) &= \left(u+\bar{u}-d-\bar{d}\right)(x,Q_0^2) \\
T_8(x,Q_0^2) &= \left(u+\bar{u}+d+\bar{d}-2s-2\bar{s}\right)(x,Q_0^2) \\
V(x,Q_0^2) &= \sum_{i=1}^{3}(q_i-\bar{q}_i)(x,Q_0^2) \\
V_3(x,Q_0^2) &= \left(u-\bar{u}-d+\bar{d}\right)(x,Q_0^2) \\
V_8(x,Q_0^2) &= \left(u-\bar{u}+d-\bar{d}-2s+2\bar{s}\right)(x,Q_0^2)
\end{aligned}
$$

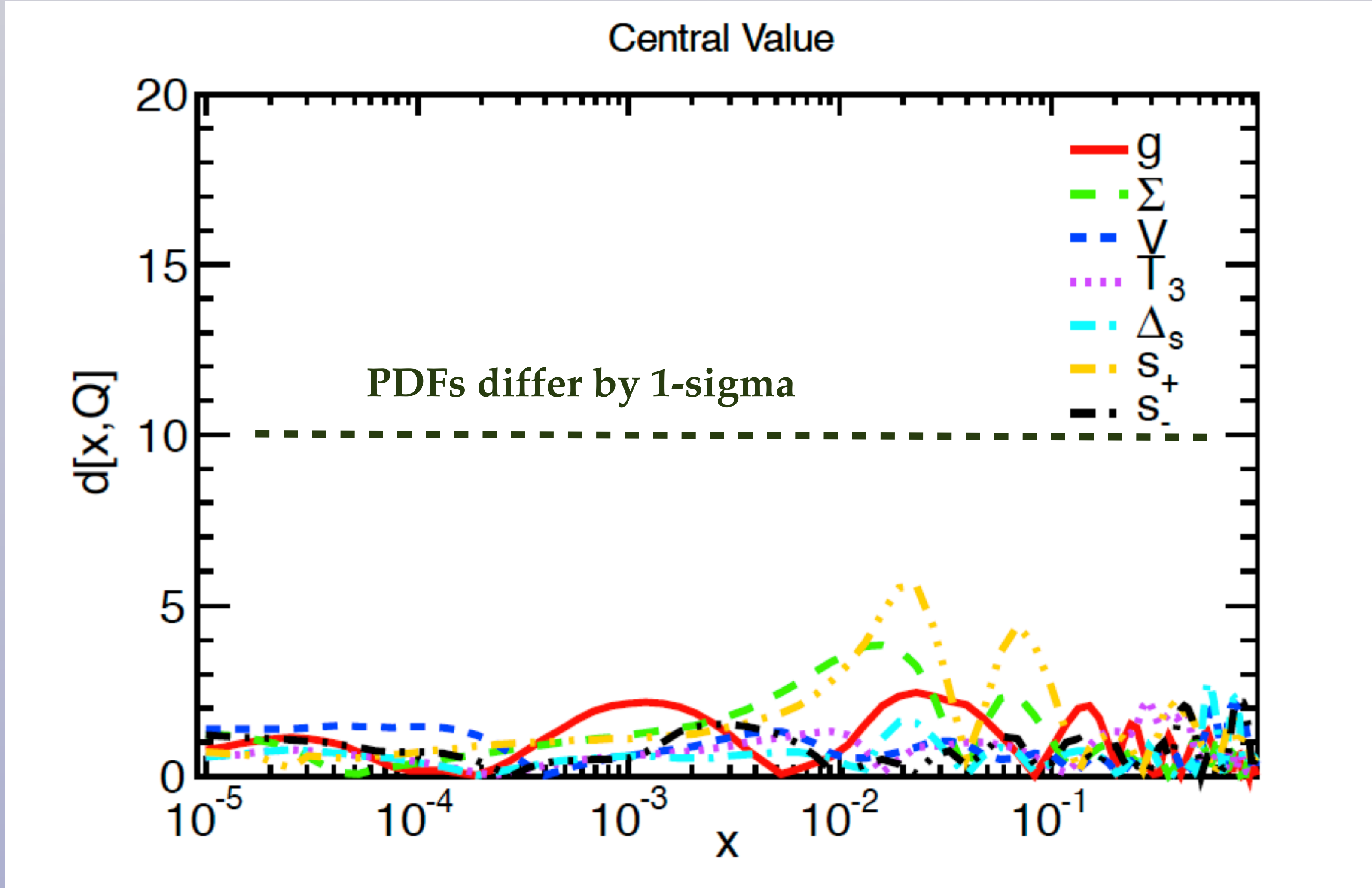
Natural basis from the point of view of physical observables

Natural basis from point of view of PDF evolution equations

Independence of PDF fitting basis

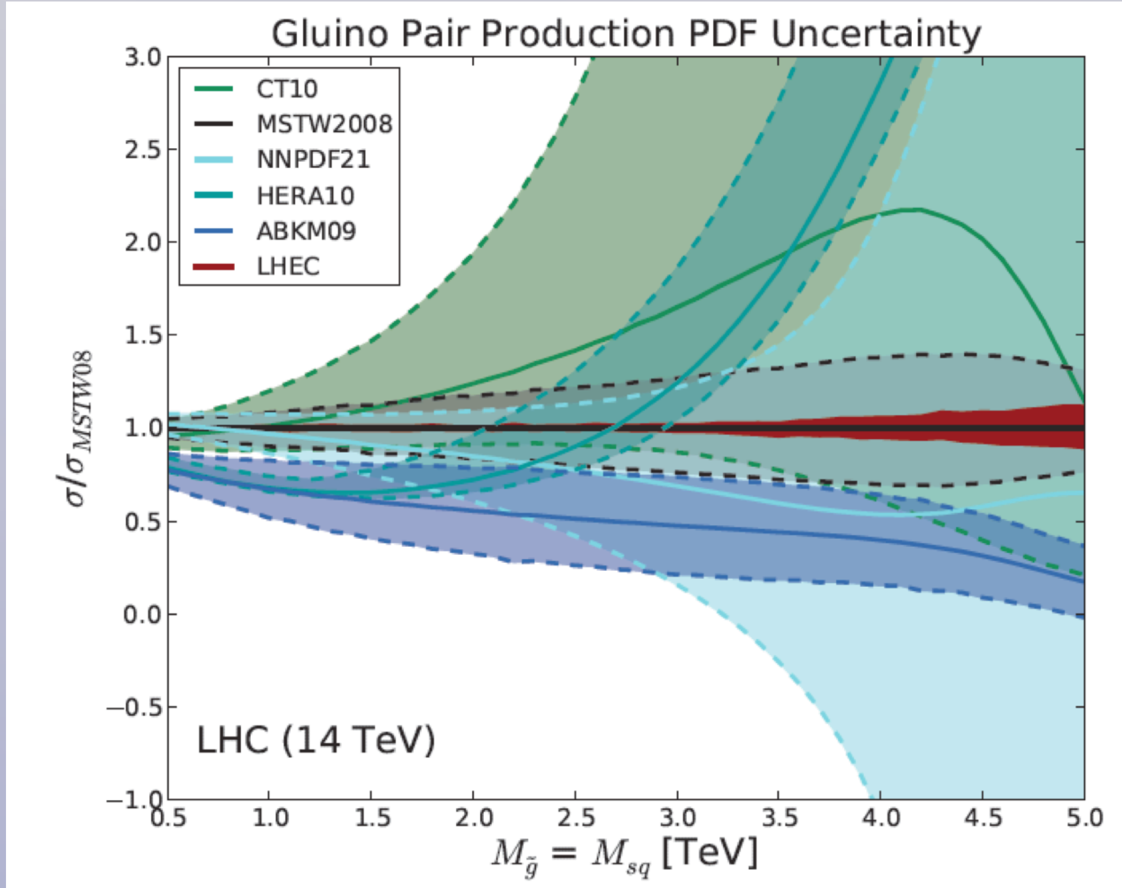
- **Predictions for physical observables** should be independent of the specific choice of PDF fitting basis
- We have explored in **closure tests** that thanks to the **improved NNPDF3.0 methodology**, we achieve almost statistically independent fits using **two very different basis**

Fit in NNPDF2.3 basis vs Fit in NNPDF3.0 basis



Positivity of physical cross-sections

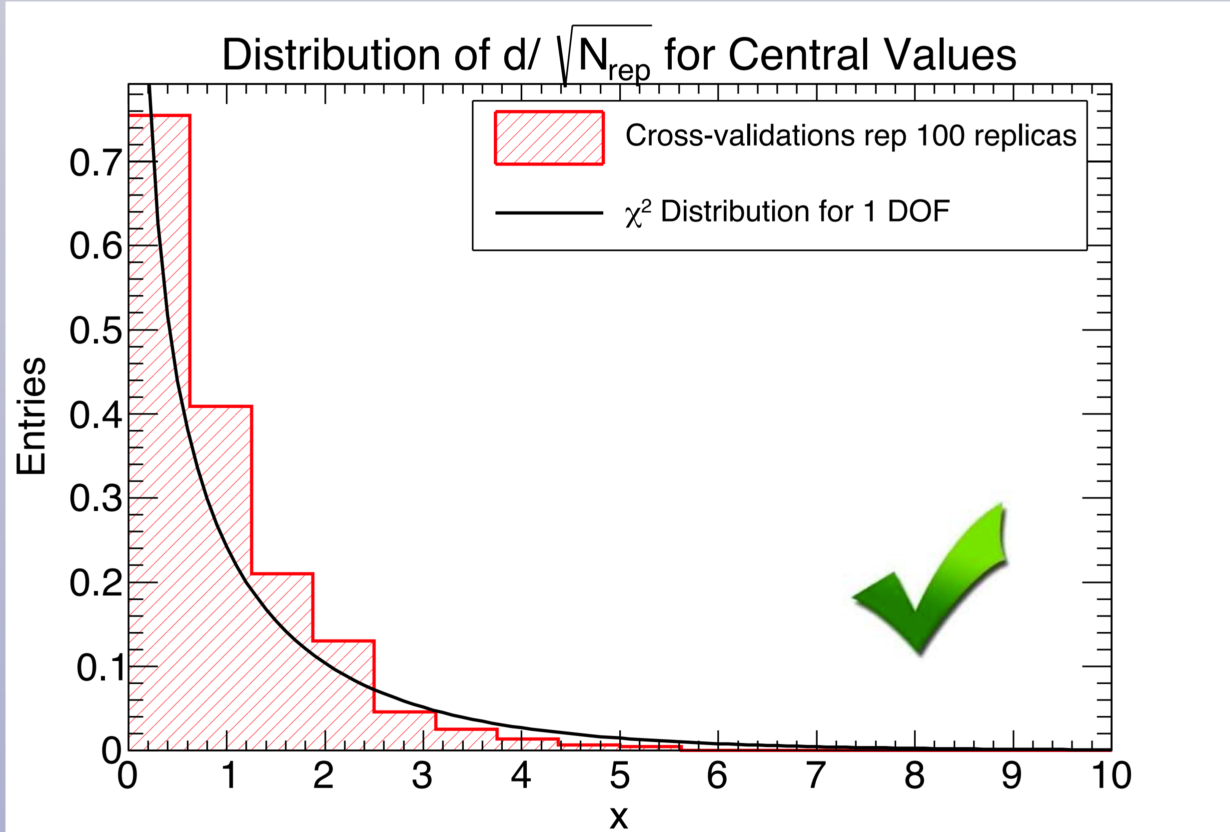
- While PDFs are not **positive definite beyond LO**, **physical cross-sections** should always be positive
- Implementing this condition, **without overconstraining PDFs with a too restrictive parametrization**, is essential for a reliable estimate of PDF uncertainties
- This is particularly crucial in the **large-x region,** production of BSM high-mass particles



PDF reweighting

- **Statistically identical results** should be obtained when **refitting** or when **reweighting,** even more so in a **closure test**
- Thus **Bayesian inference** can be used as the **ultimate closure test**, sensitive to all the higher moments of the refitted/reweighting PDFs
- As an illustration, compare effects of **CDF, D0 and ATLAS jet data** included by refitting and by reweighting. Compare also **NNPDF** and **Giele-Keller** prescriptions.

Distribution of distances between refit and RW



Positivity of physical cross-sections

- 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,** production of **BSM high-mass particles**

Quark positivity

Anti-Quark positivity

Gluon positivity

$$\begin{aligned}
F_2^u(x,Q^2) &\propto x\left(u(x,Q^2)+\bar{u}(x,Q^2)\right)+\mathcal{O}(\alpha_s)\\
F_2^d(x,Q^2) &\propto x\left(d(x,Q^2)+\bar{d}(x,Q^2)\right)+\mathcal{O}(\alpha_s)\\
F_2^s(x,Q^2) &\propto x\left(s(x,Q^2)+\bar{s}(x,Q^2)\right)+\mathcal{O}(\alpha_s)\\
\frac{d^2\sigma^{\mathrm{DY,u}}}{dM^2\,dy} &\propto u(x_1,Q^2)\bar{u}(x_3,Q^2)+\mathcal{O}(\alpha_s)\\
\frac{d^2\sigma^{\mathrm{DY,d}}}{dM^2\,dy} &\propto d(x_1,Q^2)\bar{d}(x_2,Q^2)+\mathcal{O}(\alpha_s)\\
\frac{d^2\sigma^{\mathrm{DY,s}}}{dM^2\,dy} &\propto s(x_1,Q^2)\bar{s}(x_2,Q^2)+\mathcal{O}(\alpha_s)\\
F_L(x,Q^2) &\propto C_g\otimes g(x,Q^2)+C_q\otimes q(x,Q^2)+\mathcal{O}(\alpha_s^2)\\
\frac{d\sigma^{\mathrm{higgs}}}{dy_H} &\propto g(x_1,M_H^2)g(x_2,M_H^2)+\mathcal{O}(\alpha_s^3)\qquad M_H\equiv Q_{\mathrm{pos}}
\end{aligned}$$

Positivity of physical cross sections imposed at a **low scale ~ 2 GeV**, then maintained by evolution

Preprocessing exponents

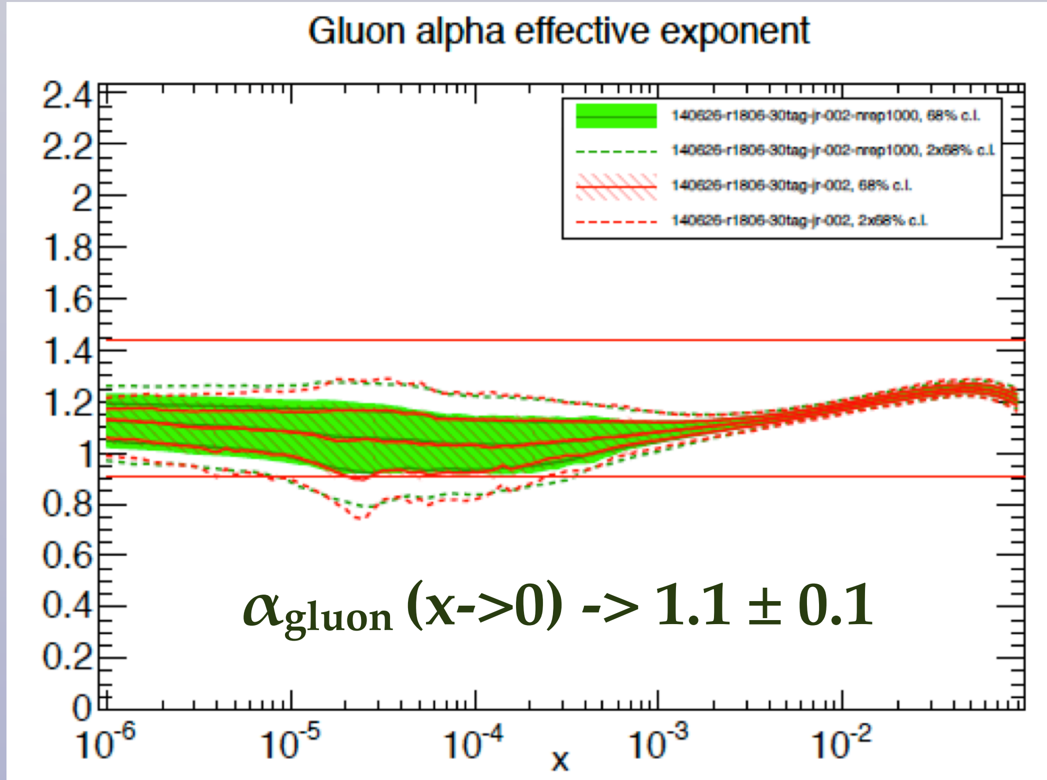
- In our PDF parametrization, the **neural networks** are complemented by **polynomial prefactors** whose goal is to s**peed-up the NN minimization**: preprocessing analysis, standard in **machine learning**

$$f_i(x,Q) = A_i\, x^{-\alpha_i}(1-x)^{\beta_i}\,\mathrm{NN}_i(x)$$

- These exponents are selected at **random for each replica**, in a wide range determined dynamically by iterating the determination of the **effective preprocessing exponents**

$$\alpha_{\mathrm{eff},i}(x) = -\frac{|f_i(x)|}{\log(x)}$$

$$\beta_{\mathrm{eff},i}(x) = \frac{|f_i(x)|}{\log(1-x)}$$



These effective exponents are also useful to validate different **models of non-perturbative QCD**