NNPDF 2.3

Parton Densities with LHC data

Valerio Bertone
Albert-Ludwigs-Universität Freiburg,
Physikalisches Institut

DIFFRACTION 2012

September 11 - 15, 2012, Lanzarote, Canarie Islands

SUMMARY

- The **NNPDF** methodology
- The **NNPDF 2.3** set:
  - Dataset
  - Results
  - Phenomenology
- Conclusions

The NNPDF Methodology

Main ingredients

- **Monte Carlo** by importance sampling:
    - construct a set of data replicas which reproduces the statistical features of the original dataset.
- **Neural networks** as interpolants:
    - unbiased basis parametrized by a very large set of parameters.
- **Genetic Algorithm** for neural network training:
    - suitable exploration of the space of parameters avoiding to fall into local minima.
- Determination of the best fit by **cross-validation**:
    - proper fitting avoiding overlearning.

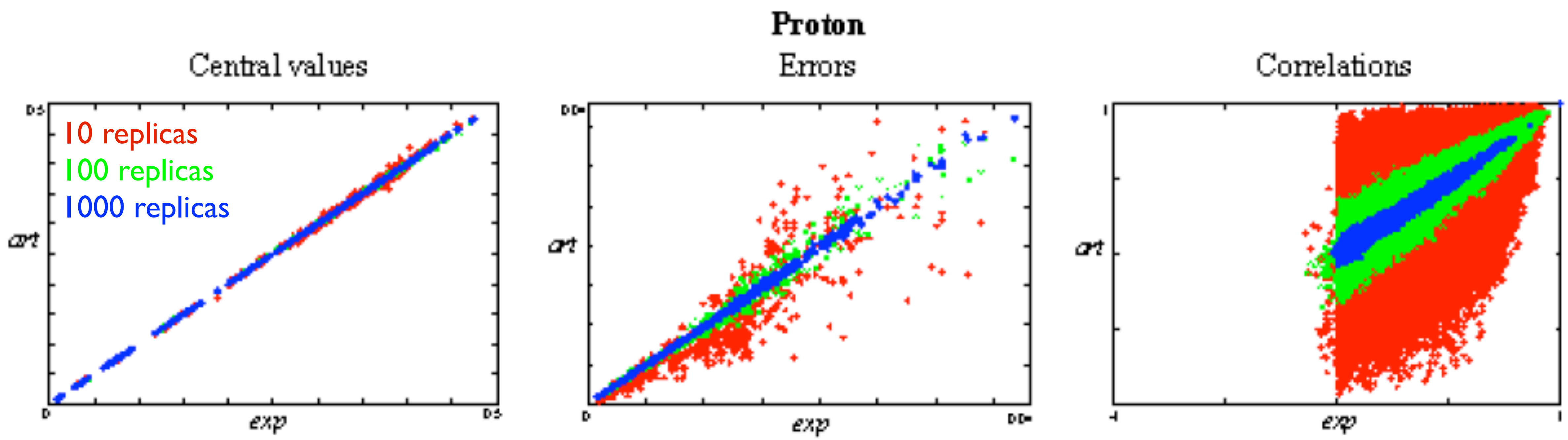
The NNPDF Methodology
Monte Carlo Sampling

- Generation of **artificial MC data** according to the distribution:

$$\mathcal{O}_i^{(art)(k)} = (1 + r_{norm}^{(k)}\sigma_{norm})\left[\mathcal{O}_i^{(exp)} + r_{stat}^{(k)}\sigma_{stat} + \sum_{p=1}^{N_{sys}} r_{sys,p}^{(k)}\sigma_{sys,p}\right]$$

where $r_j^{(k)}$ are univariate (normally distributed) random numbers.

- **Validation** of the MC replicas against data:



- ***O*(1000) replicas** to reproduce correlations at the percent accuracy.

NNPDF

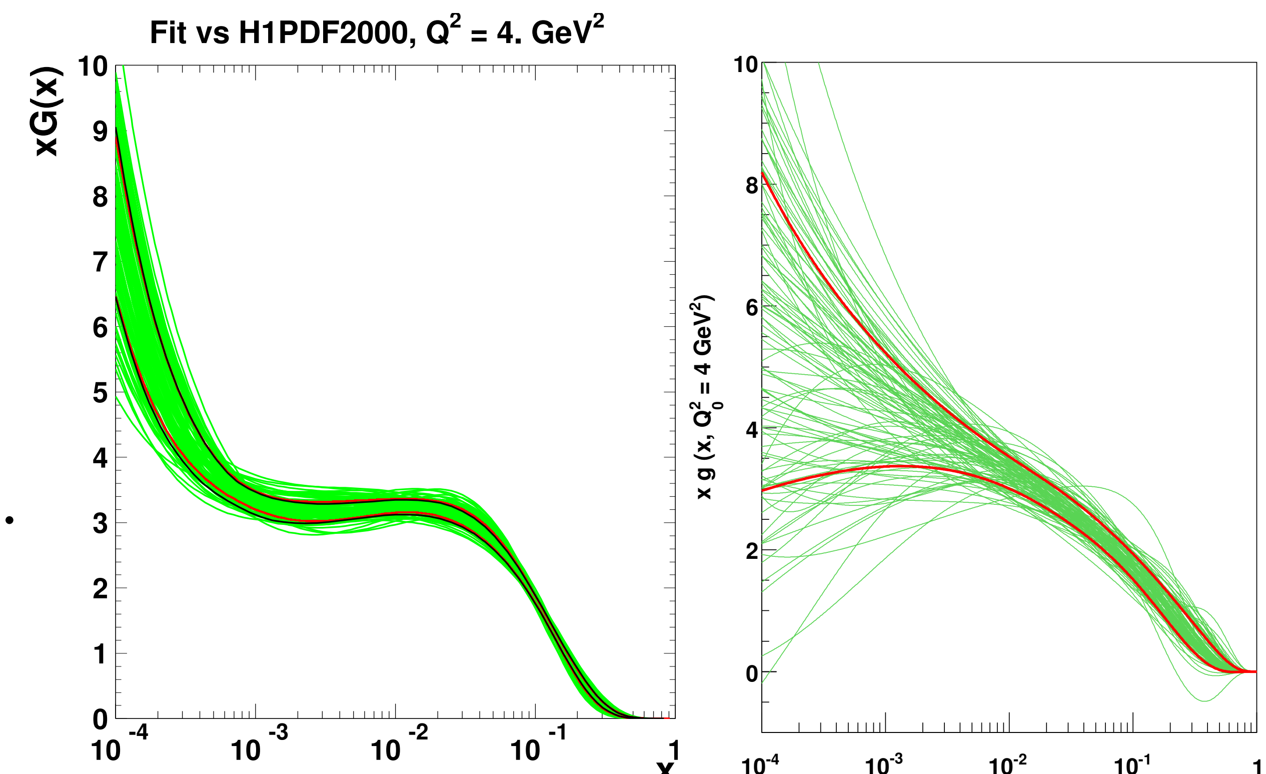
The NNPDF Methodology
Neural Networks

- **7 independent PDFs** at the initial scale $Q_0=\sqrt{2}$ GeV:

Gluon: $g(x)$
Singlet: $\Sigma(x) = \sum_q [q(x) + \overline{q}(x)]$
Valence: $V(x) = \sum_q [q(x) - \overline{q}(x)]$
Triplet: $T_3(x) = [u(x) + \overline{u}(x)] - [d(x) + \overline{d}(x)]$
Sea asymmetry: $\Delta_s(x) = \overline{d}(x) - \overline{u}(x)$
Total strangeness: $s^+(x) = s(x) + \overline{s}(x)$
Strange valence: $s^-(x) = s(x) - \overline{s}(x)$

- Each PDF parametrized by a **Neural Network** having architecture 2-5-3-1 (37 free parameters).

- **Redundant** parametrization **to avoid biases** from the choice of the PDF functional form:
  - Polinomial form vs. Neural Network,
  - stable under the change of architecture.



The NNPDF Methodology

Dynamical Stopping

Drawback:

A redundant parametrization might fit not only to the physical behaviour but also to the **random statistical fluctuations** of data.

NNPDF

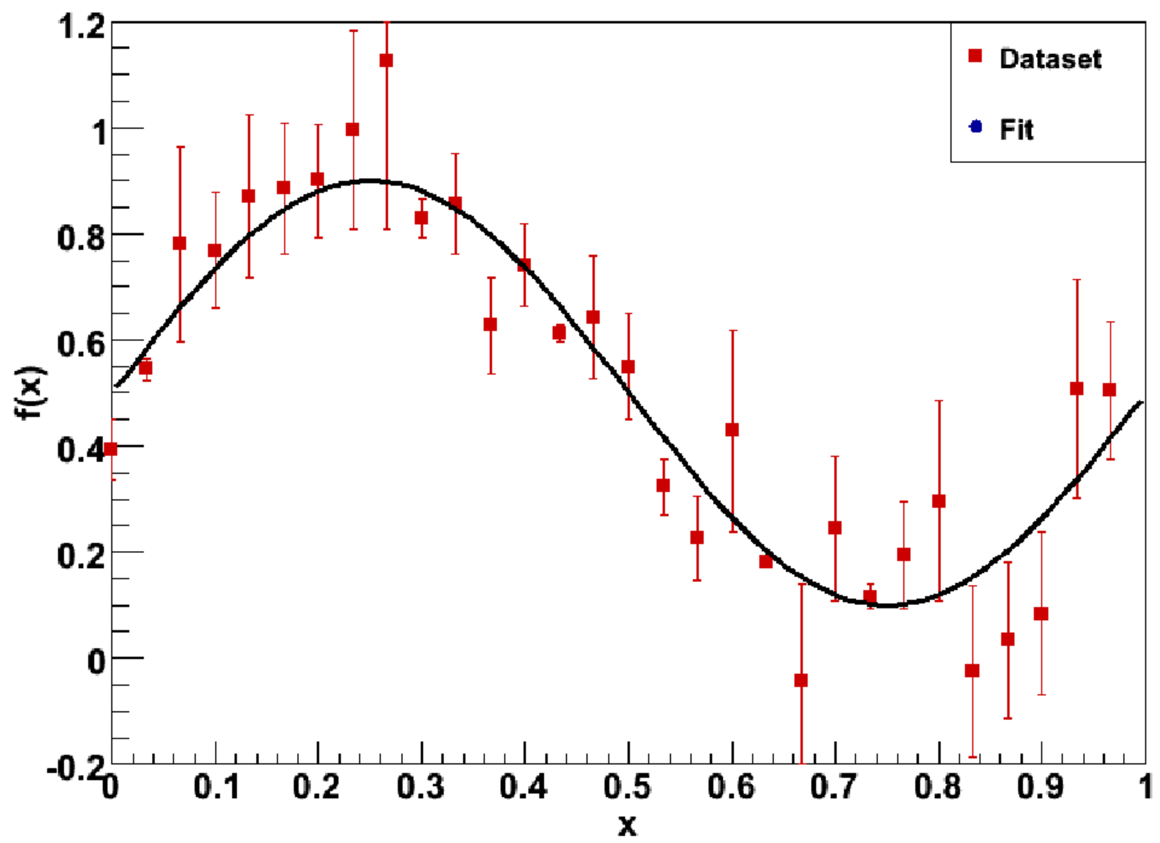
The NNPDF Methodology

Dynamical Stopping

Drawback:

A redundant parametrization might fit not only to the physical behaviour but also to the **random statistical fluctuations** of data.

Underlying physical law



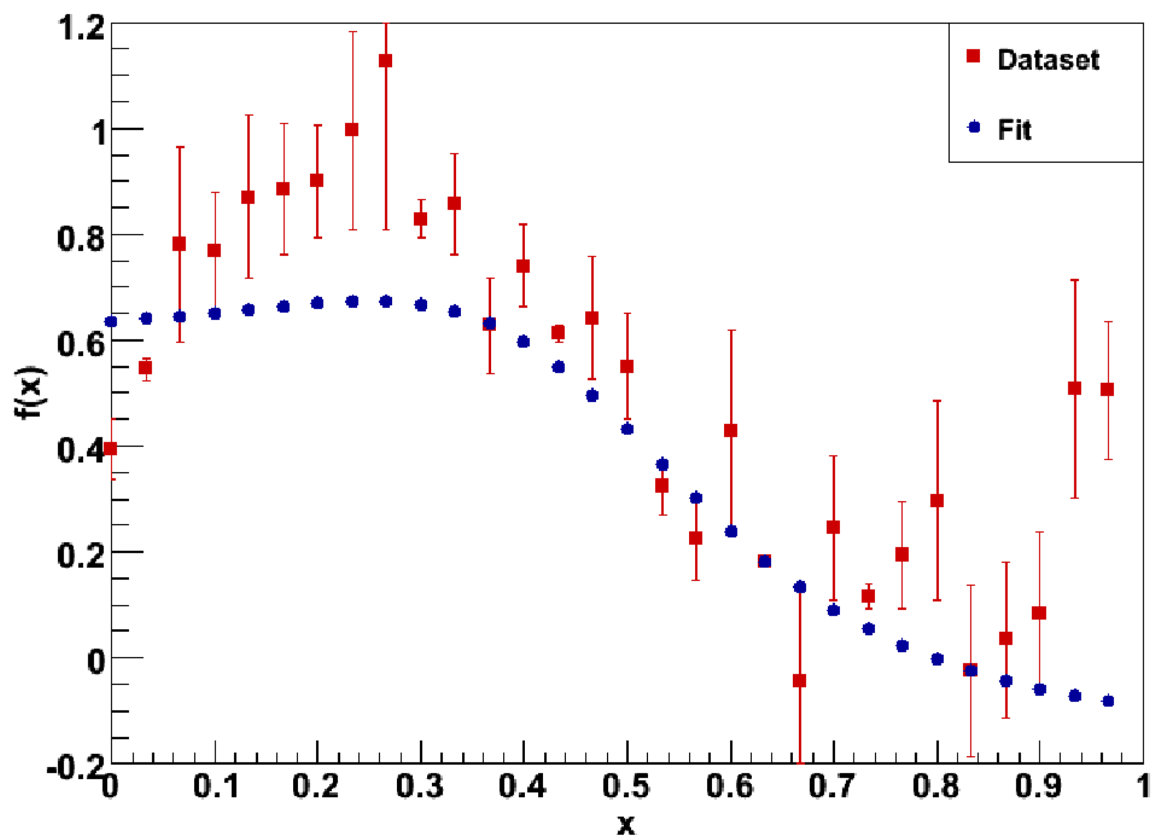
NNPDF

The NNPDF Methodology
Dynamical Stopping

Drawback:

A redundant parametrization might fit not only to the physical behaviour but also to the **random statistical fluctuations** of data.

Underlearning



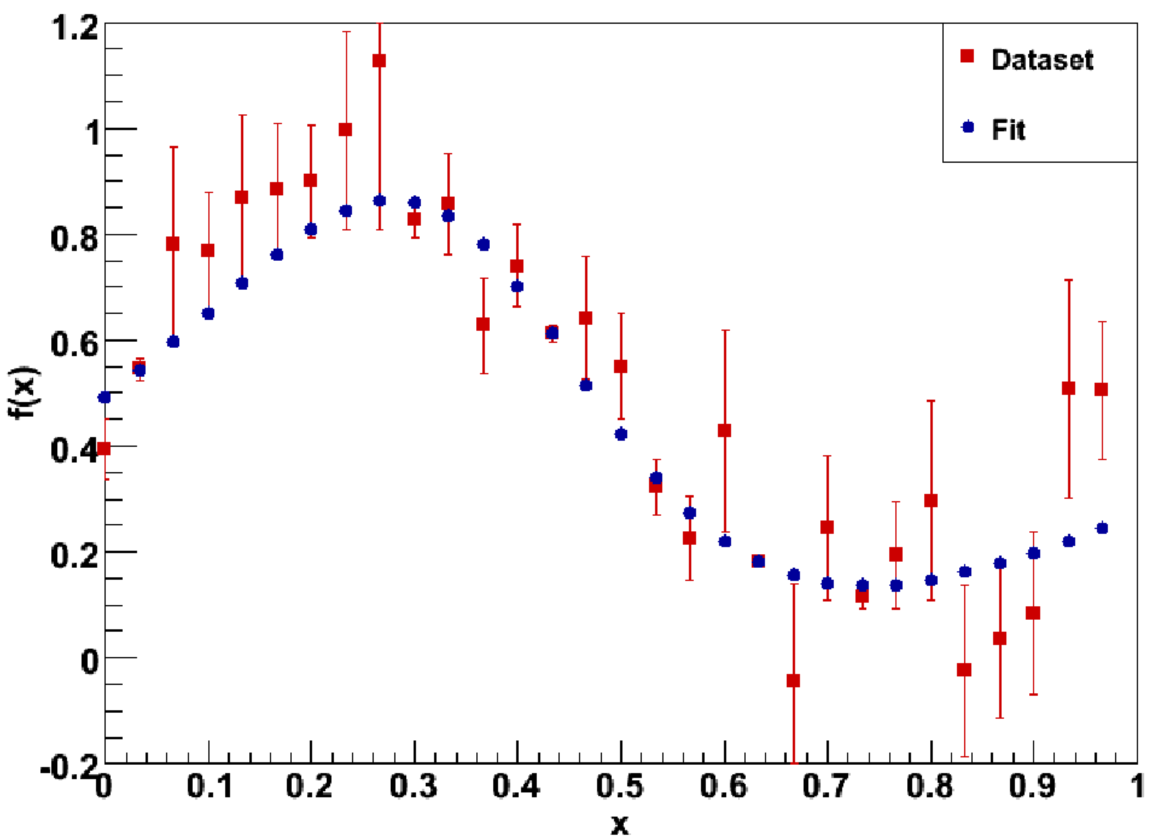
NNPDF

The NNPDF Methodology
Dynamical Stopping

Drawback:

A redundant parametrization might fit not only to the physical behaviour but also to the **random statistical fluctuations** of data.

Proper learning



NNPDF

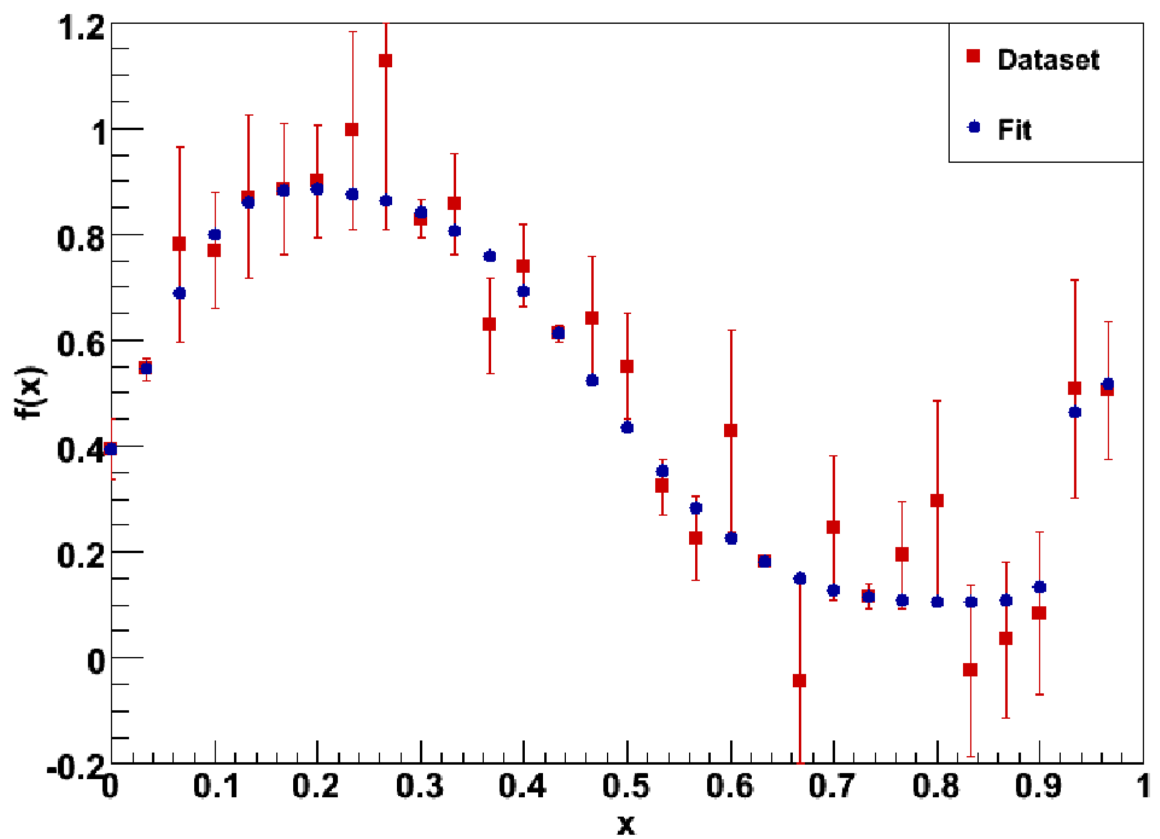
The NNPDF Methodology

Dynamical Stopping

Drawback:

A redundant parametrization might fit not only to the physical behaviour but also to the **random statistical fluctuations** of data.

Overlearning



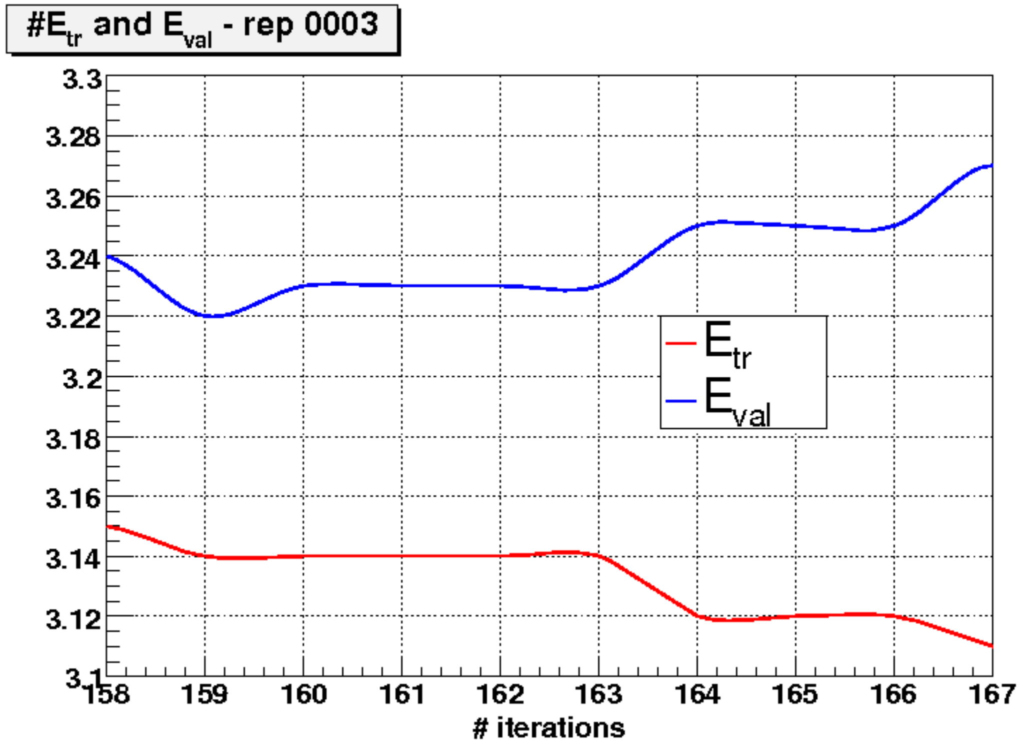
The NNPDF Methodology

Dynamical Stopping

Need for a **dynamical stopping** criterion!

Cross-validation method:

- Divide data in two sets: **training** and **validation** (for each experiment).
- **Random division** for each replica (tipically $f_t = f_v = 0.5$).
- **Minimization** performed only on the training set. Meantime, the validation χ^2 is monitored.
- When the training χ^2 still decreases while the validation χ^2 stops decreasing, **STOP the fit.**



The NNPDF Methodology

... to summarize

- Generate **Monte Carlo replicas** of the experimental data taking into account all experimental correlations.

- Fit to each replica a set of PDFs, parametrized at the initial scale Q_0 with **Neural Networks** using the **Genetic Algorithm** to minimize the χ^2 and the **Cross-Validation** method to stop the fit.

- **Expectation values** for observables are then given by:

$$\langle \mathcal{O}\{f\} \rangle = \frac{1}{N} \sum_{k=1}^{N} \mathcal{O}[f_k] \quad \text{and} \quad \delta \langle \mathcal{O}\{f\} \rangle = \sqrt{\langle \mathcal{O}^2\{f\} \rangle - \langle \mathcal{O}\{f\} \rangle^2}$$

... and corresponding formulae used to compute correlations.

NNPDF

NNPDF timeline

... how we got here

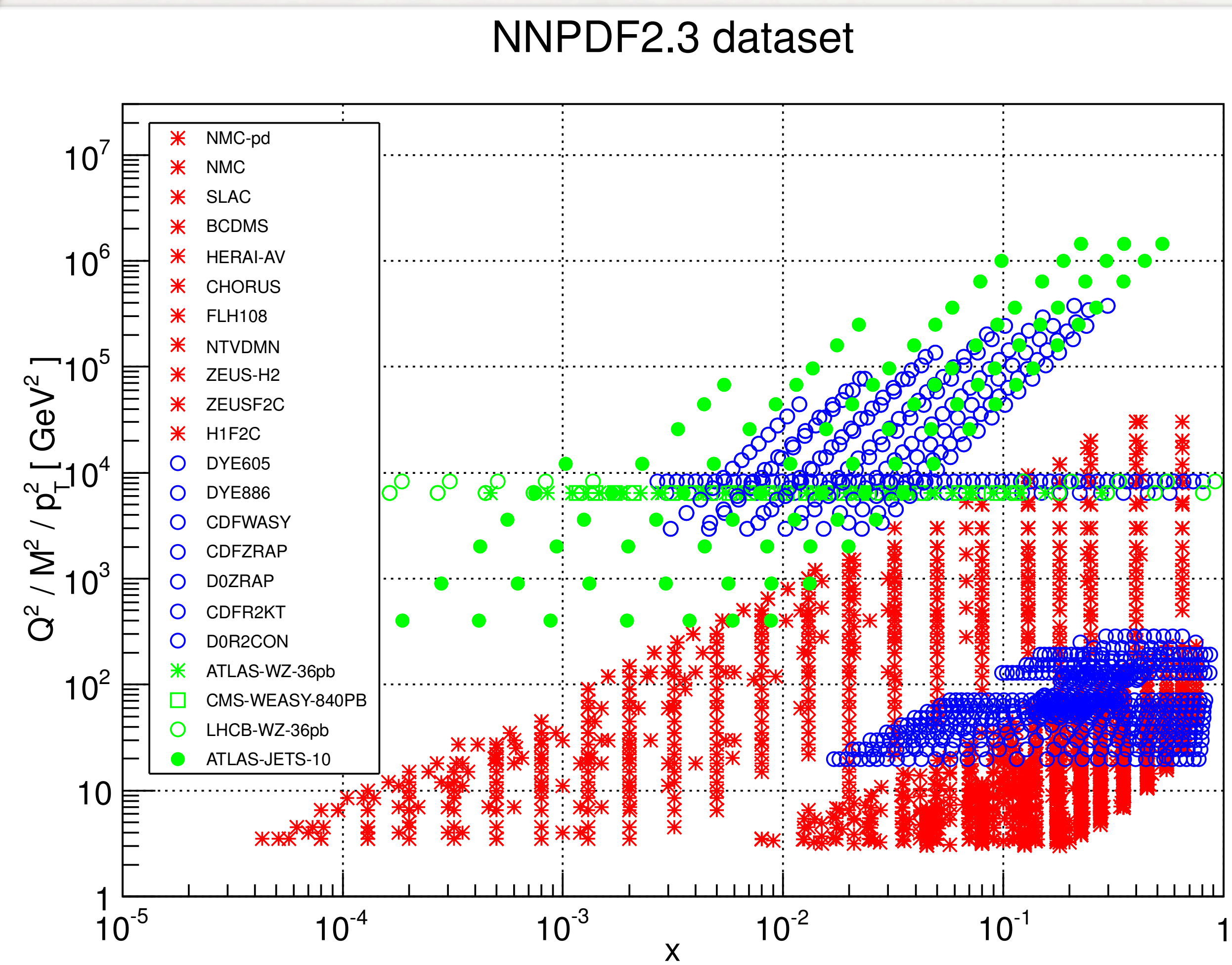
	NNPDF 1.0	NNPDF 1.2	NNPDF 2.0	NNPDF 2.1 NLO	NNPDF 2.1 LO & NNLO	NNPDF 2.3
DIS	✔	✔	✔	✔	✔	✔
Drell-Yan	✘	✘	✔	✔	✔	✔
Jets	✘	✘	✔	✔	✔	✔
LHC	✘	✘	✘	✘	✘	✔
Antistrange indep. param	✘	✔	✔	✔	✔	✔
Heavy quark masses	✘	✘	✘	✔	✔	✔
NNLO	✘	✘	✘	✘	✔	✔

NNPDF

NNPDF 2.3

The Dataset

The **NNPDF 2.3** is the first PDF fit set at NLO and NNLO which **includes LHC data**.



3501 data points (NNLO)

Experiment	Data
Fixed Target DIS	1952
Fixed Target DY	318
HERA DIS	834
Tevatron W/Z	70
Tevatron Jets	186
ATLAS incl. Jets	**90**
ATLAS W/Z lept. rap.	**30**
CMS W letp. asym.	**11**
LHCb W rap.	**10**

Only the relevant LHC data for which the **full covariance matrix** is available are included.

NNPDF

NNPDF 2.3
Methodological Improvements

High order corrections **computationally intensive**

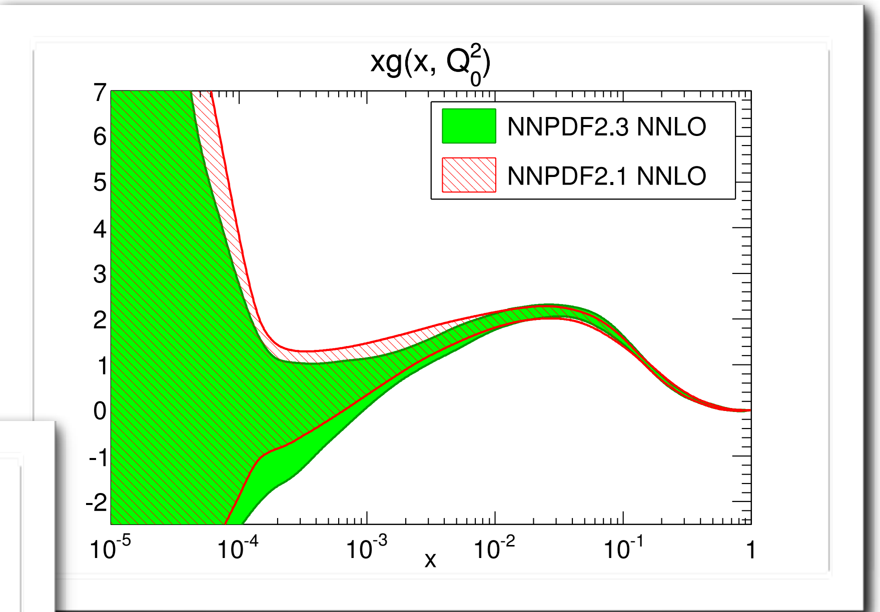
- Combinantion of **FastKernel** with **FastNLO/APPLgrid**:
  - substatial **speed-up** in computation of observables during the fit.
- More **advanced minimization** procedure to fit PDFs:
  - allowed by fast computation of observables,
  - **retuning of the minimization** algorithm (genetic algorithm),
  - retraining of replicas having χ^2 “too away” from the average (**outliers**).

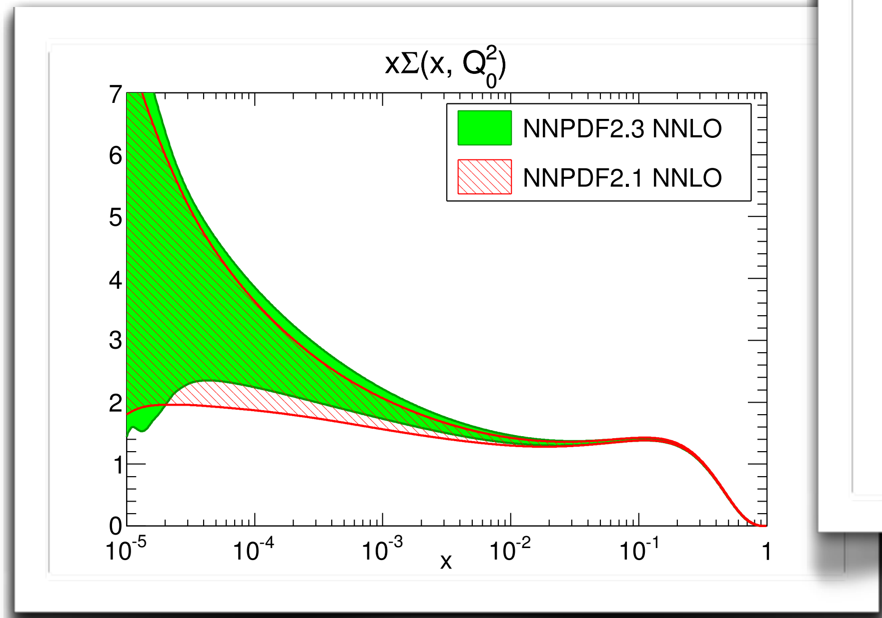
NNPDF

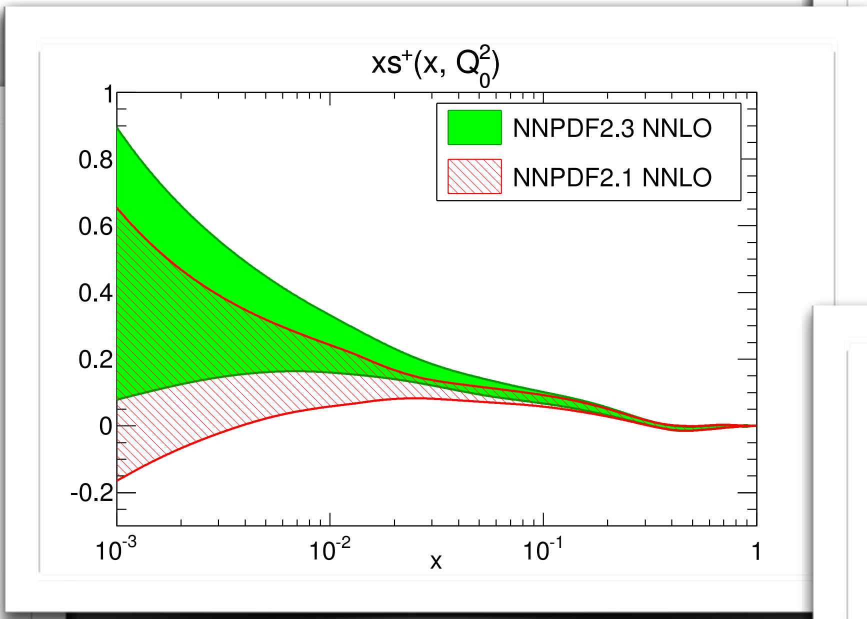
NNPDF 2.3

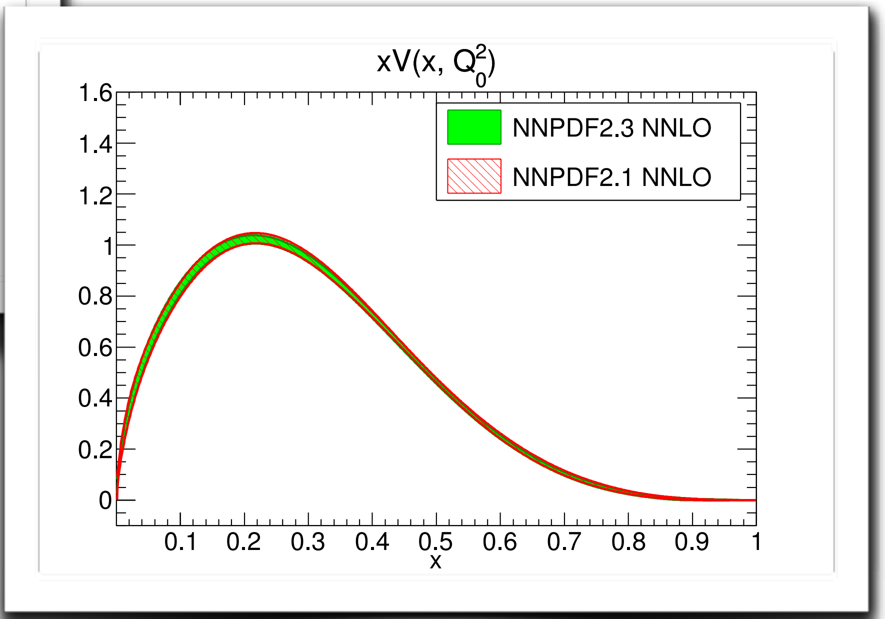
Parton Distributions: 2.3 vs. 2.1

- Addition of the **LHC data**,
- improved **minimization**,
- corrected error in the **di-muon cross-section**.









- Moderate but significant reduction in uncertainty,
- change in **strangeness.**

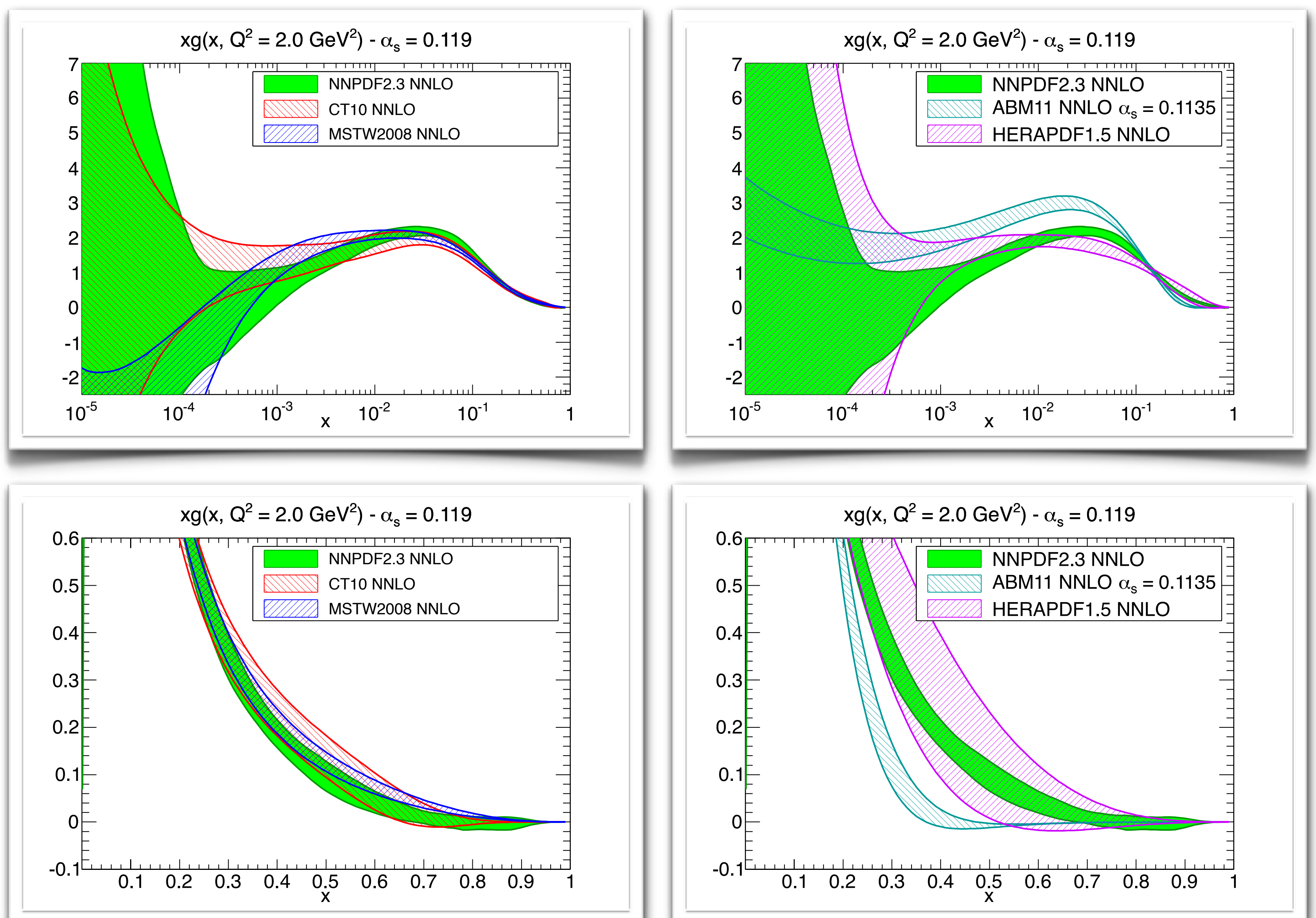
NNPDF

NNPDF 2.3

Parton Distributions: NNPDF vs. other collaborations

- Other NNLO PDF sets present on the market:
  - **MSTW2008**: **global fit** (DIS+DY+Jets+EW)
  - **CT10**: **global fit** (DIS+DY+Jets+EW)
  - **ABM11**: **red. dataset** (DIS+DY)
  - **HERAPDF1.5**: **red. dataset** (HERA only)

- good **agreement between global fits** (**NNPDF**, **MSTW** and **CT**) all over the *x* range,
- larger **differences** between **NNPDF** and **ABM** at both large and small-*x* (different α_s?),
- **ABM**: **unnatually small uncertainty** at small-*x*,
- **HERAPDF** larger uncertainty at large-*x* (**no collider data**) but better agreement at small-*x*. **Underestimation** of the error in the **quark sector**.



NNPDF 2.3

Parton Distributions: NNPDF vs. other collaborations

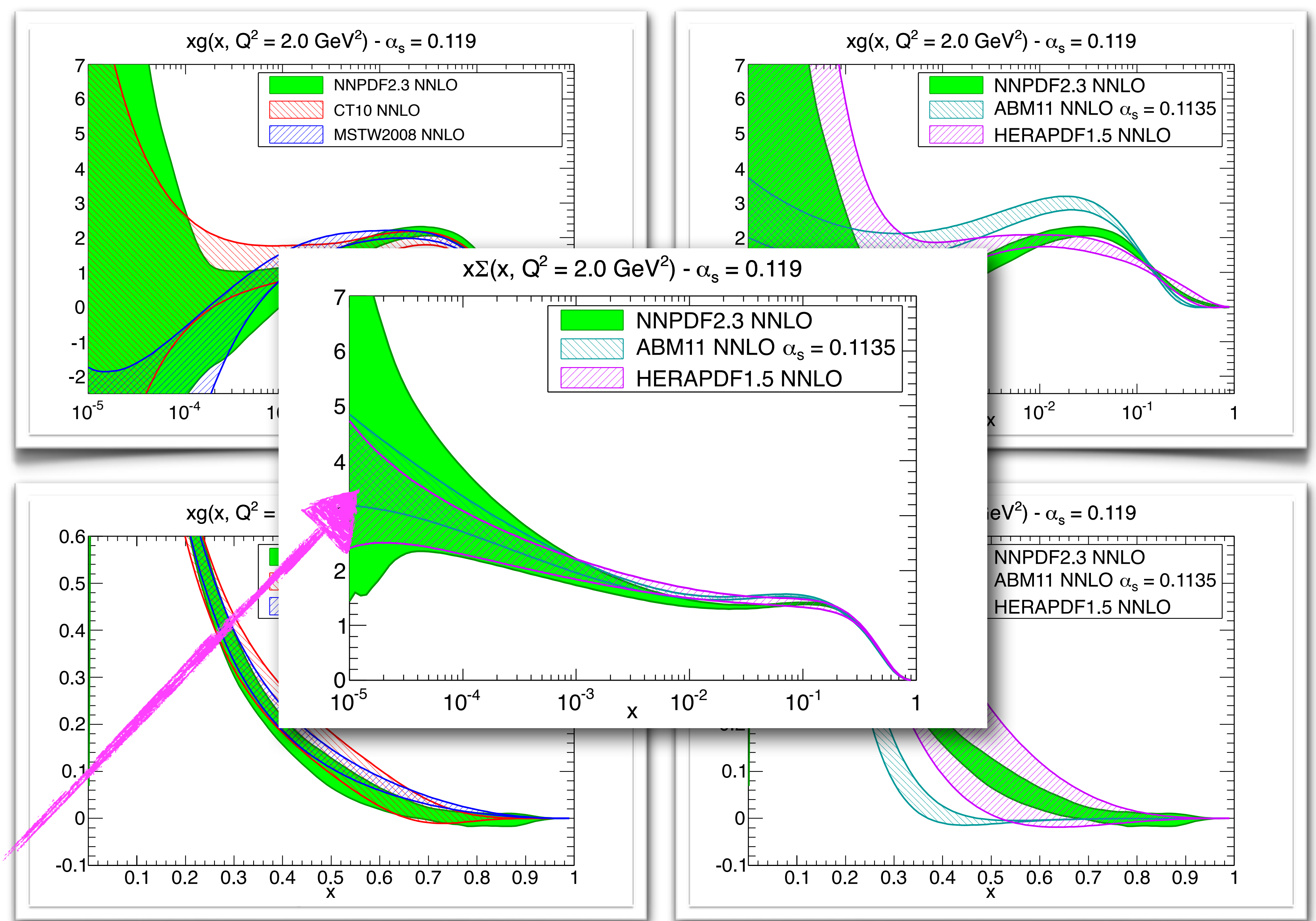
- Other NNLO PDF sets present on the market:
  - **MSTW2008**: **global fit** (DIS+DY+Jets+EW)
  - **CT10**: **global fit** (DIS+DY+Jets+EW)
  - **ABM11**: **red. dataset** (DIS+DY)
  - **HERAPDF1.5**: **red. dataset** (HERA only)

- good **agreement between global fits** (**NNPDF**, **MSTW** and **CT**) all over the *x* range,

- larger **differences** between **NNPDF** and **ABM** at both large and small-*x* (different α_s?),

- **ABM**: **unnatually small uncertainty** at small-*x*,

- **HERAPDF** larger uncertainty at large-*x* (**no collider data**) but better agreement at small-*x*. **Underestimation** of the error in the **quark sector**.

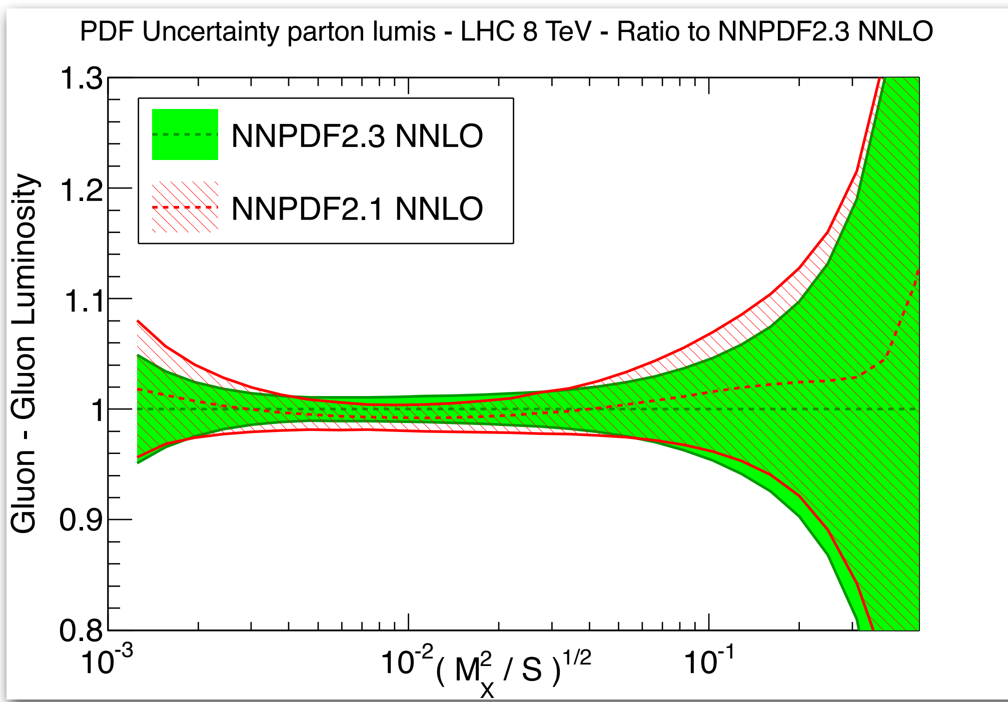


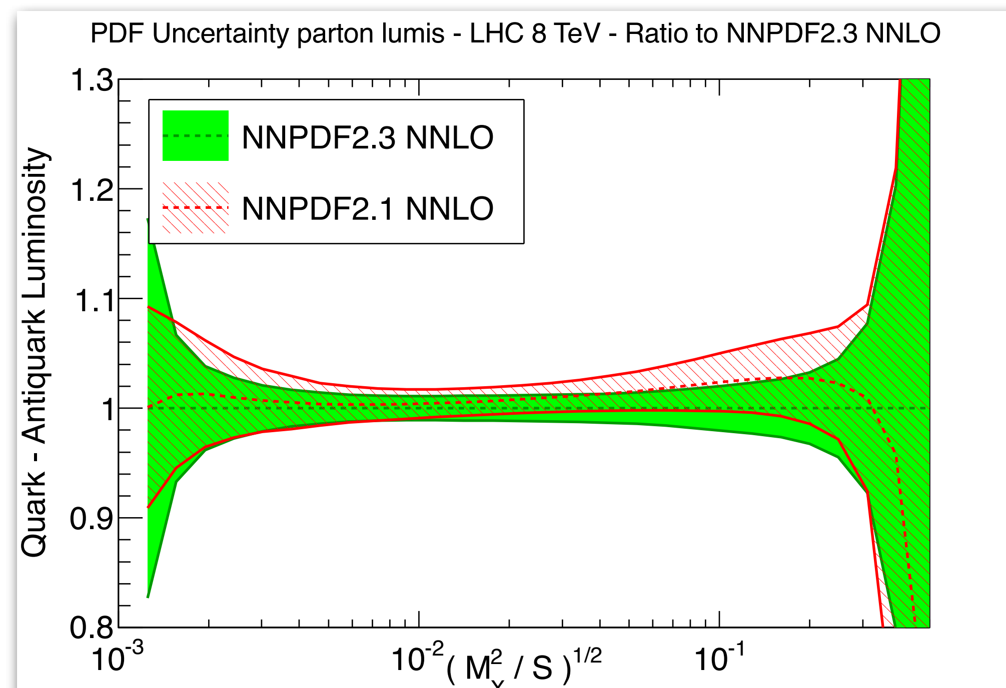
NNPDF 2.3

Parton Luminosity: 2.3 vs. 2.1

At hadron colliders observables depend on **parton luminosities**:

$$\Phi_{ij}(M_X^2) = \frac{1}{s}\int_\tau^1 \frac{dx_1}{x_1} f_i(x_1, M_X^2) f_j(\tau/x_1, M_X^2)$$

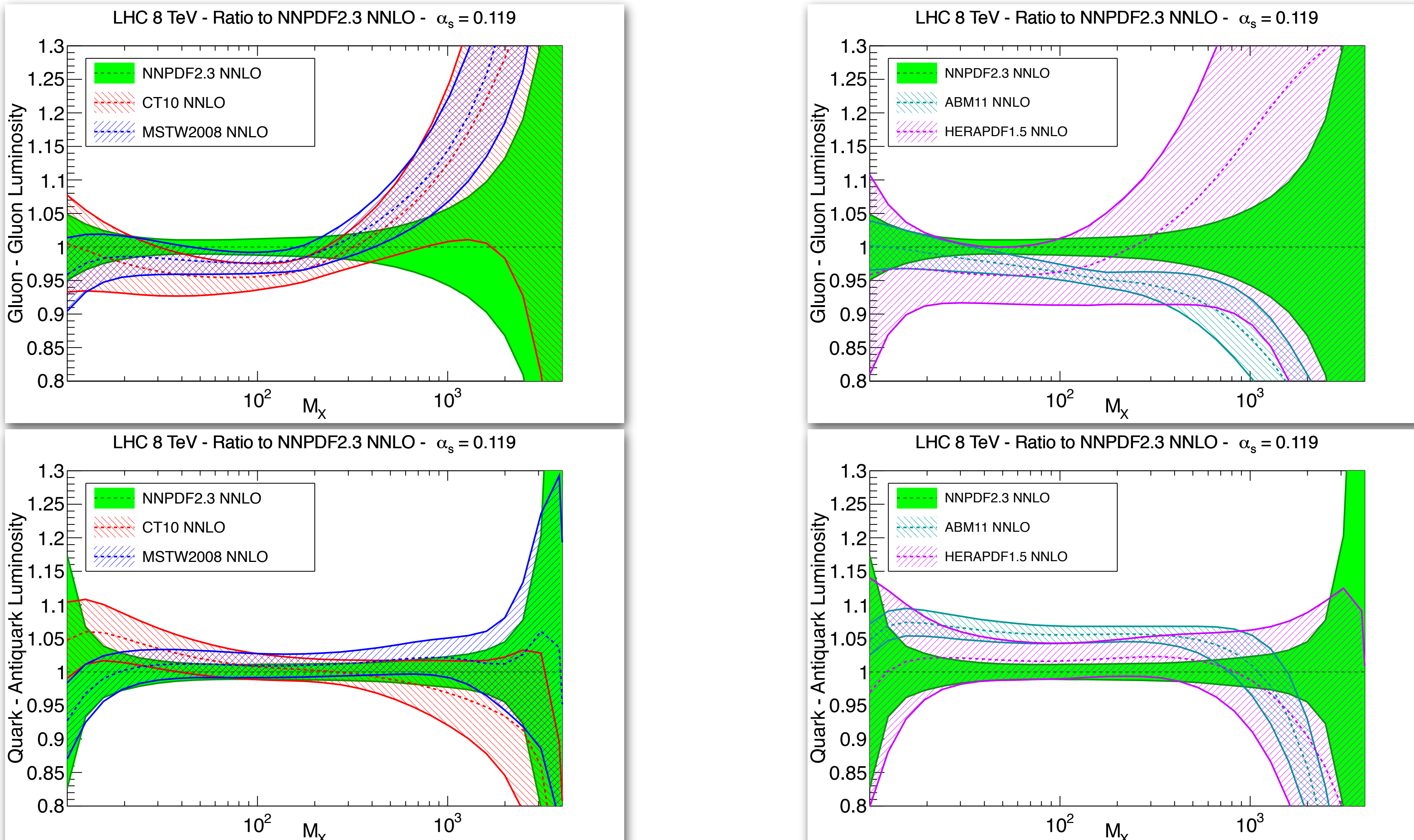




- When going from NNPDF 2.1 to NNPDF 2.3, the **gluon-gluon** luminosity:
  - stable in the Standard Model Higgs/Top region [$M_X/\sqrt{s} \simeq 1.5 - 2 \times 10^{-2}$]
  - reduction of the uncertainty for larger final state invariant masses.

NNPDF 2.3

Parton Luminosity: NNPDF vs. other collaborations

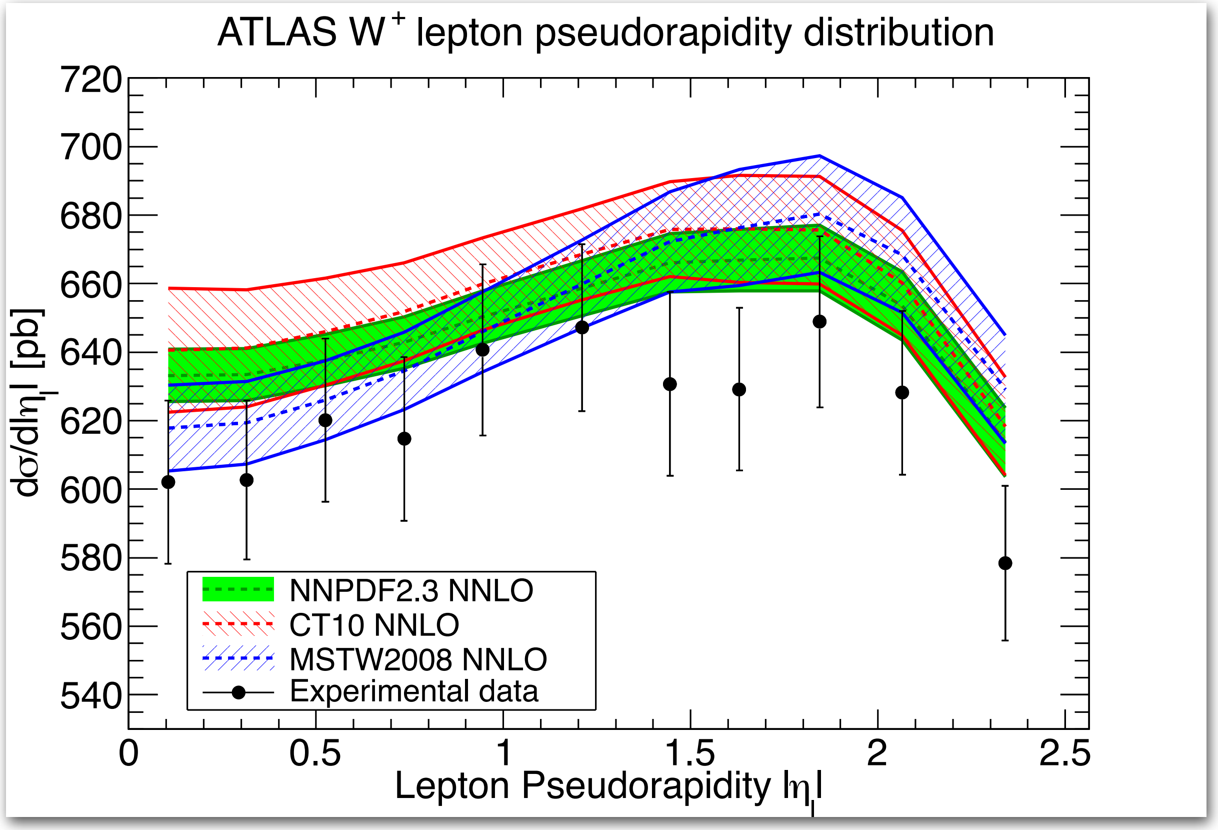


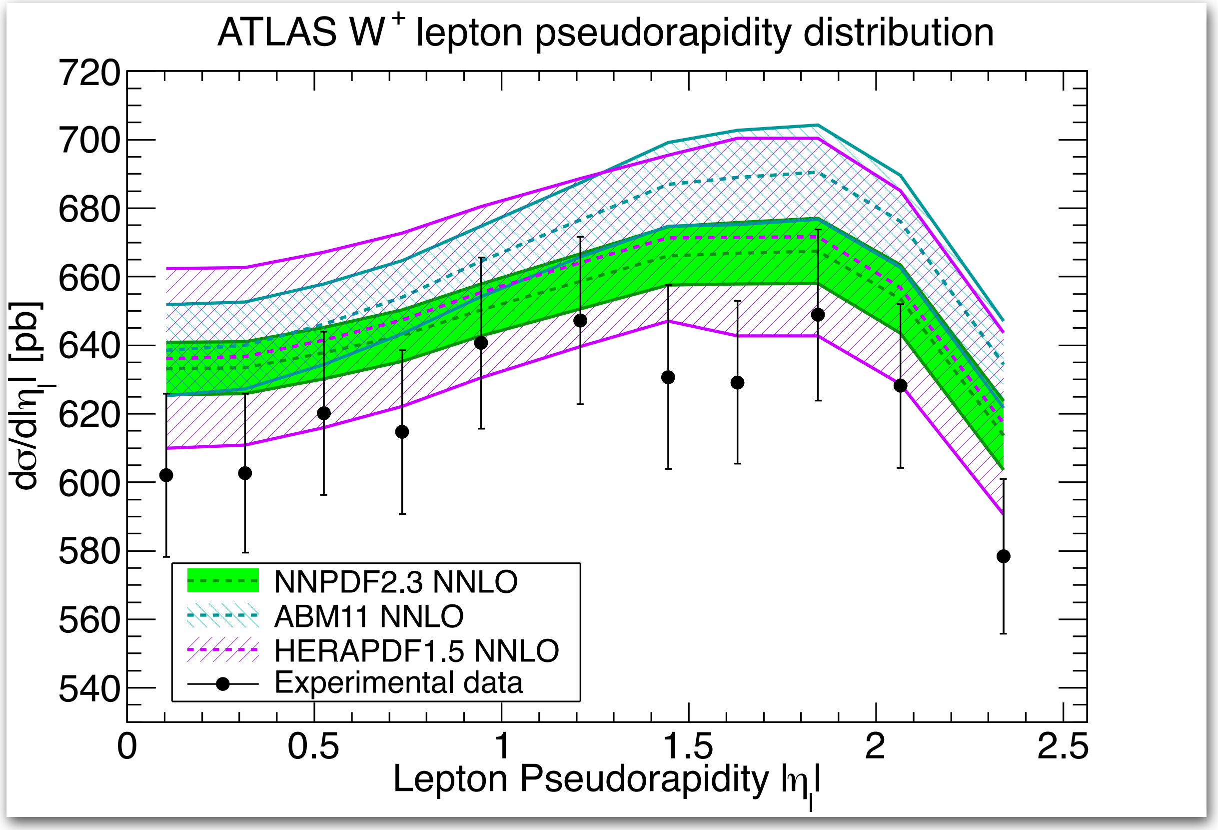
- **Global fits in good agreement** all over the relevant range of invariant mass,
- **larger differences** between **NNPDF** and **ABM** even using the same value for α_s,
- **HERAPDF** in **general agreement** with **NNPDF** but larger uncertanty.

NNPDF

NNPDF 2.3

Phenomenology: Differential Distributions



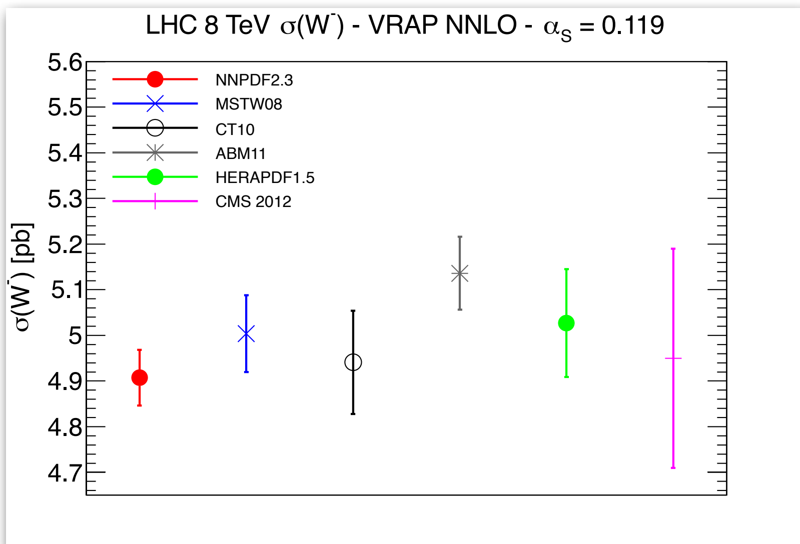


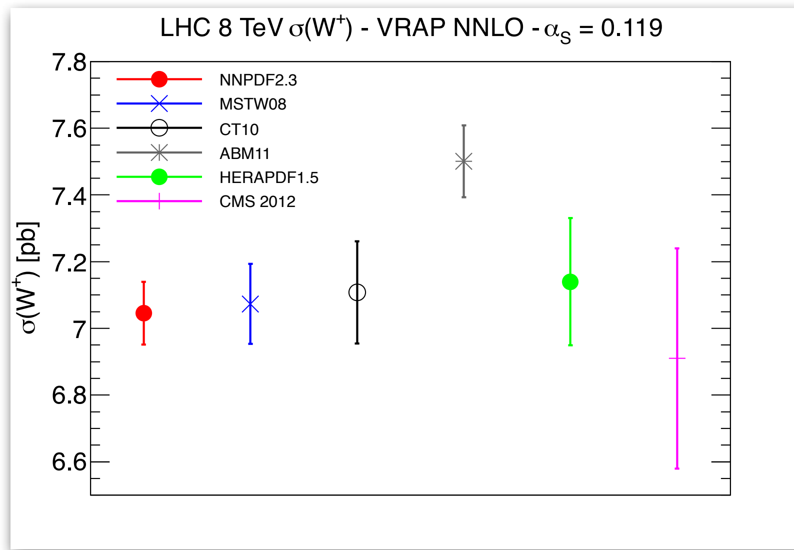
Dataset	NNPDF2.3	MSTW08	CT10	ABM11	HERAPDF1.5
ATLAS W, Z	1.46	3.20	1.16	2.06	1.87
CMS W el asy	0.84	3.86	1.77	1.61	0.81
LHCb W, Z	0.89	1.09	0.98	2.04	0.80
ATLAS jets	1.41	1.47	1.54	1.52	1.61

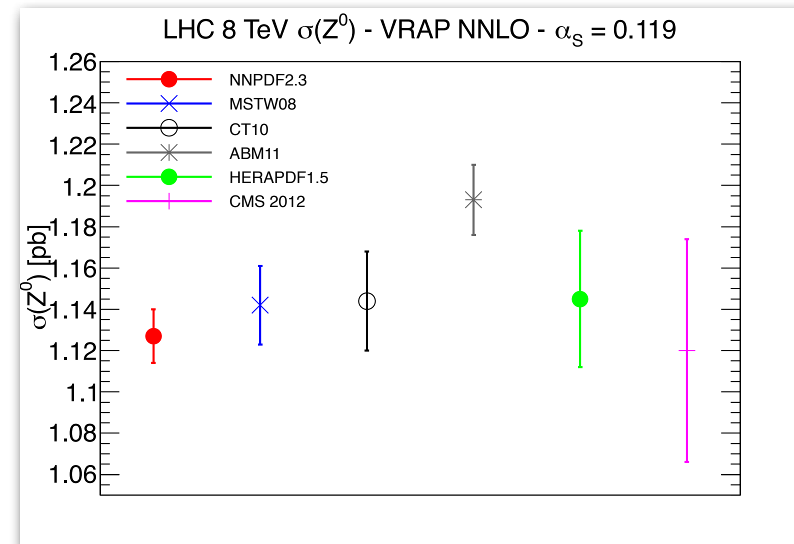
NNPDF 2.3

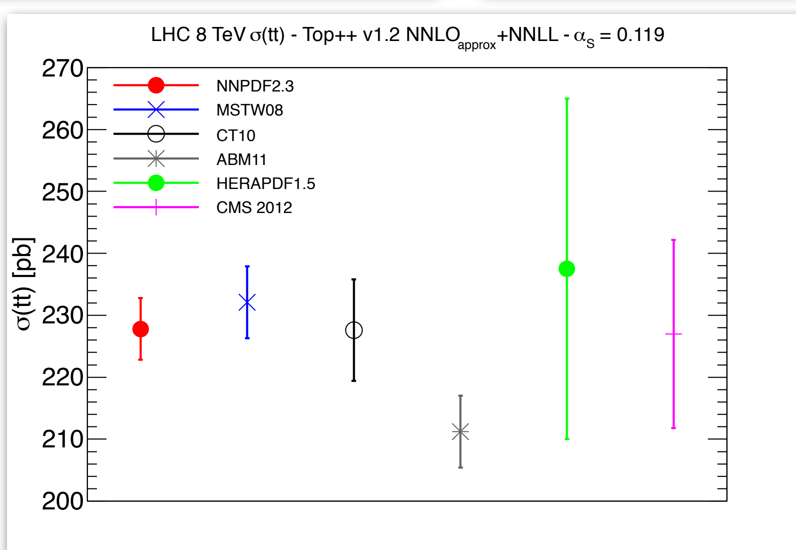
Phenomenology: Inclusive Cross-sections at 8 TeV

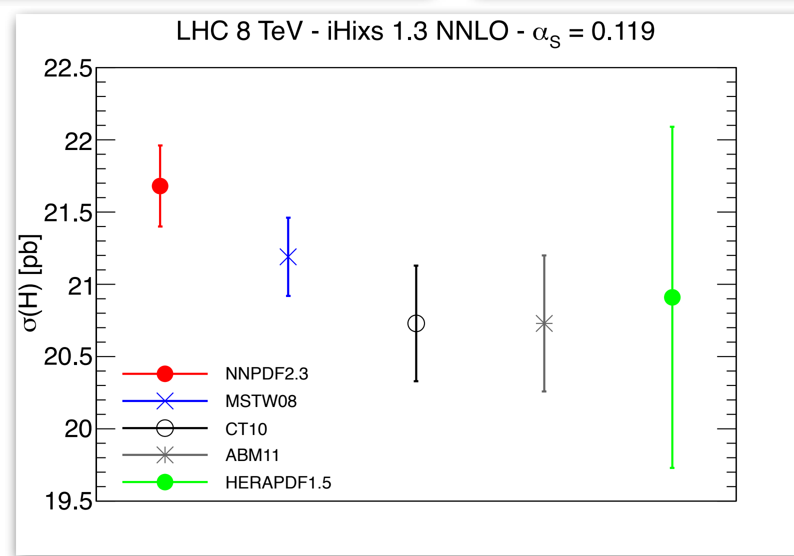
- Theory predictions vs. first LHC measurements at 8 TeV:











General **good agreement** between all the **global** PDF sets and with data.

Conclusions and Outlook

- The **NNPDF 2.3** is the **first PDF fit** including LHC data,
- **improvement** in the **fitting methodology** allowed by a faster observable computation,
- impact of the LHC data small but **non-negligible**.
- Inclusion of more LHC data and new processes:
  - W + *charm* production for strangeness determination,
  - *top pair* production for the large-x gluon determination.

All the NNPDF 2.3 sets are available from the LHAPDF interface and from the web site:

http://nnpdf.hepforge.org/

where also a Mathematica interface is provided.

NNPDF

Backup Slides

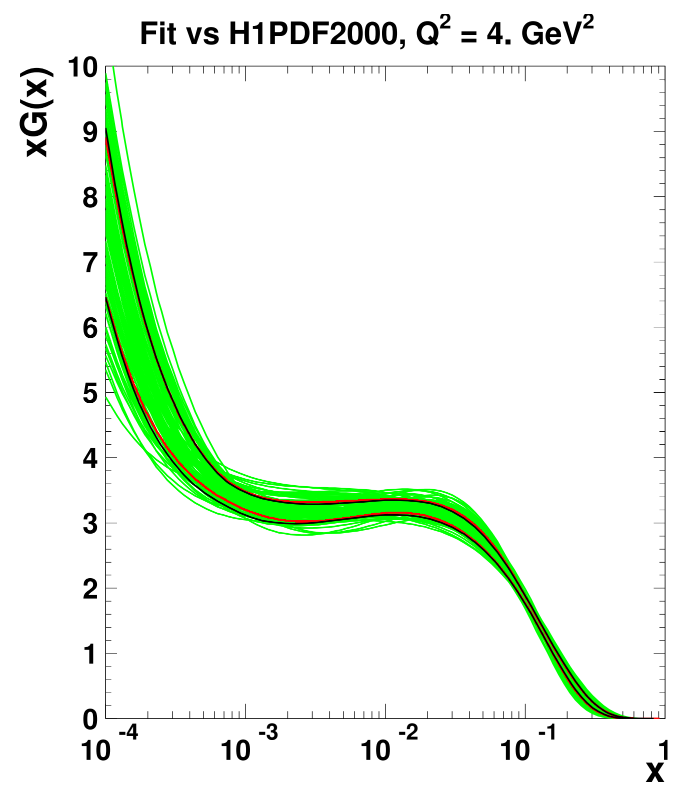
The NNPDF Methodology
Main ingredients

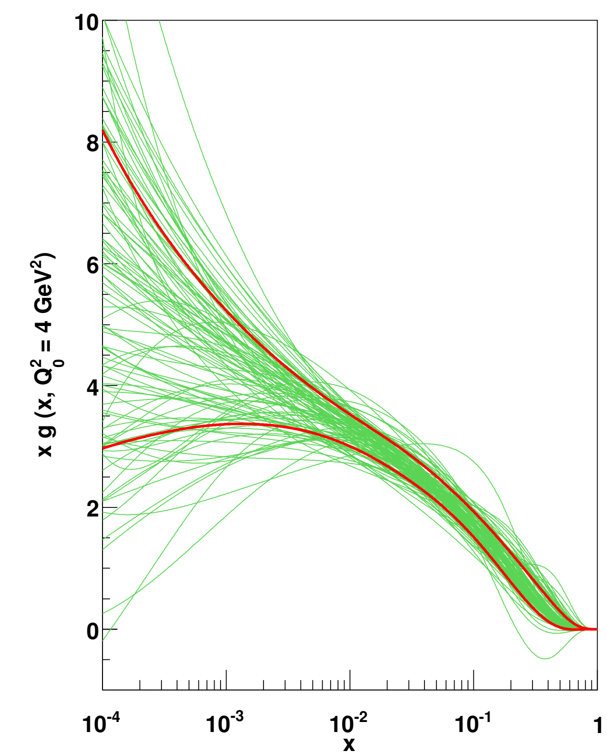
- **Monte Carlo** determination of **uncertainties**:
  - **no** need to rely on **linear propagation** of errors,
  - possibility to test the impact of **non-gaussian** uncertainties,
  - possibility to test for **non-gaussian behaviour** of fitted PDFs.
- Parametrization of PDFs using **Neural Networks**:
  - provide an **unbiased parametrization**.
- Determine the **best fit** PDFs using **Cross-Validation**:
  - ensures **proper fitting** avoiding overlearning.

The NNPDF Methodology

Neural Networks

- **7 independent PDFs** at the initial scale $Q_0=\sqrt{2}$ GeV.
- Each PDF parametrized by a **Neural Network** having architecture 2-5-3-1 (37 free parameters).
- **Redundant** parametrization **to avoid biases** from the choice of the PDF functional form:
    - Polinomial form vs. Neural Network:





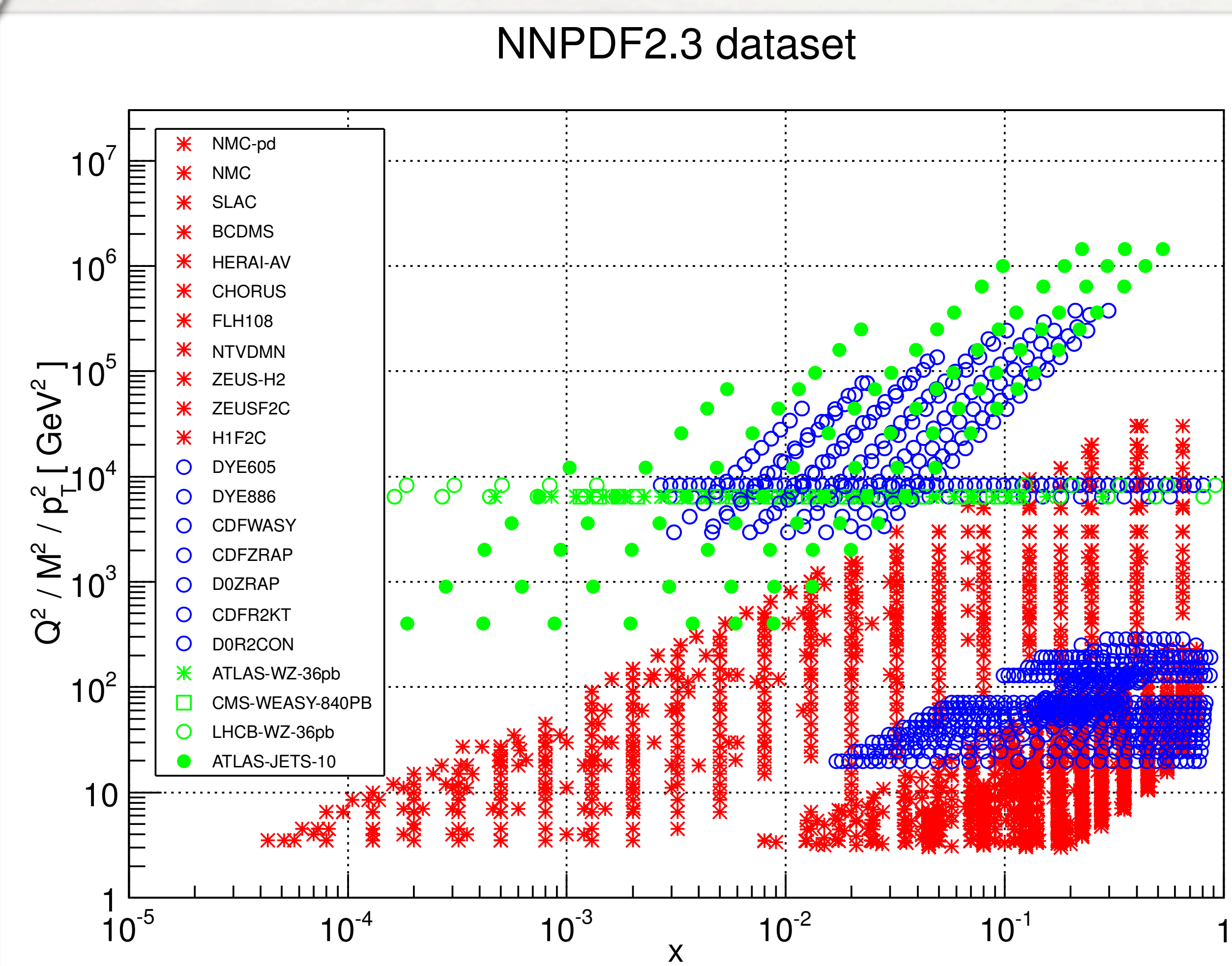
NNPDF 2.3
The Dataset: The LHC data

- The **NNPDF 2.3** is the first PDF fit set at NLO and NNLO which **includes LHC data**.
- The dataset includes the relevant LHC data for which the **full covariance matrix** is available:
  - **ATLAS** Inclusive Jets, 36 pb^{-1} [arXiv:1112.6297]
  - **ATLAS** W/Z lepton rapidity distributions, 36 pb^{-1} [arXiv:1109.5141]
  - **CMS** W lepton asymmetry, 840 pb^{-1} [arXiv:1206.2598]
  - **LHCb** W rapidity distributions, 36 pb^{-1} [arXiv:1204.1620]

NNPDF

NNPDF 2.3

The Dataset: no LHC fit



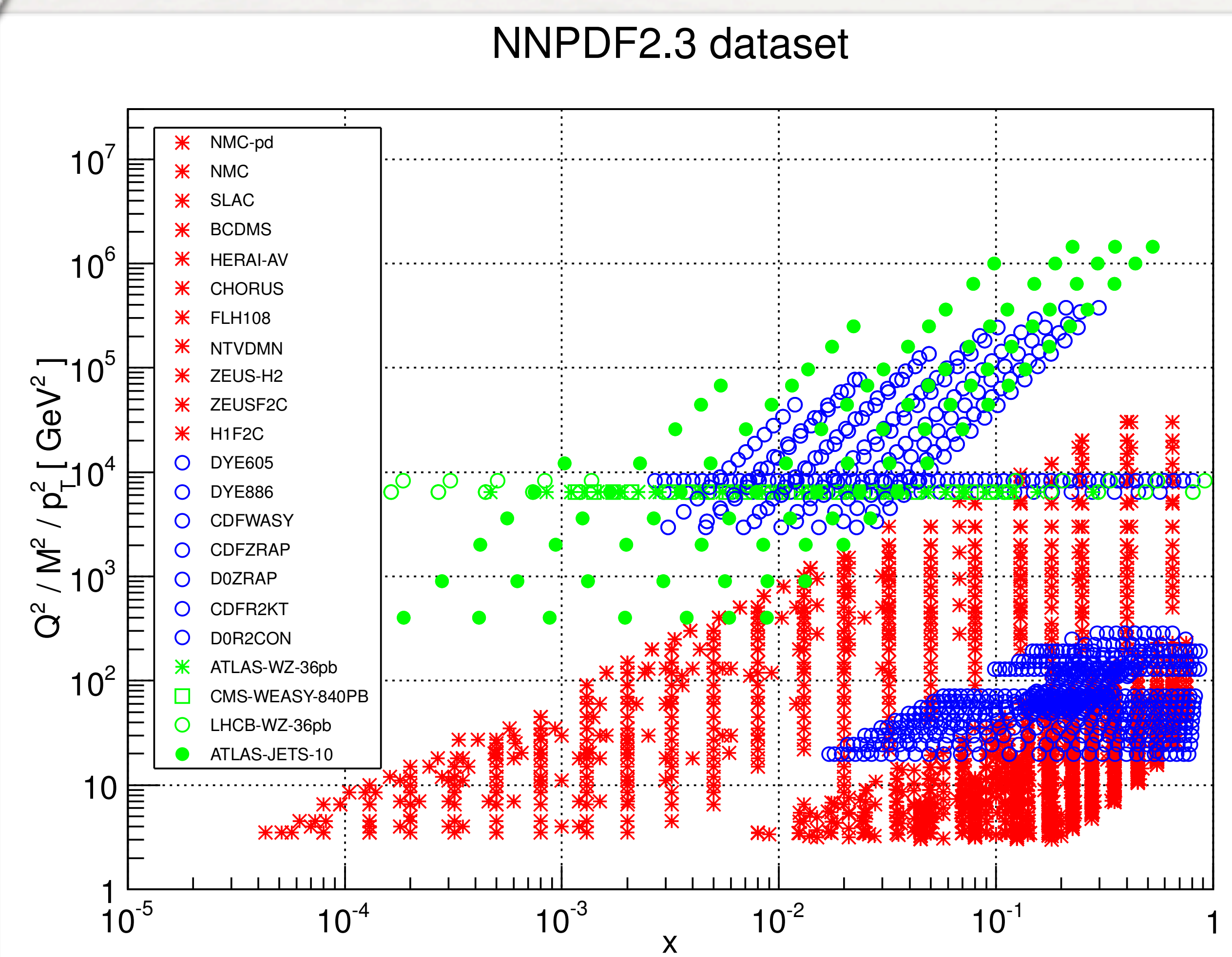
	Experiment	Data
1	Fixed Target DIS	1952
	Fixed Target DY	318
2	HERA DIS	834
	Tevatron W/Z	70
	Tevatron Jets	186
3	LHC W/Z	51
	LHC Jets	90

3360 data points (NNLO)

... to assess the impact of the LHC data

NNPDF 2.3

The Dataset: collider only fit



	Experiment	Data
1	Fixed Target DIS	1952
	Fixed Target DY	318
2	HERA DIS	834
	Tevatron W/Z	70
	Tevatron Jets	186
3	LHC W/Z	51
	LHC Jets	90

1231 data points (NNLO)

... to study the compatibility of the low energy data

NNPDF

NNPDF 2.3
LHC data: Impact and consistency (NLO)

Compare the quality of the fit to LHC data **before** and **after** the **inclusion**

- Including LHC data in the fit improves the quality of their description **without deteriorating** the quality of the fit to other datasets.
- Moderate impact of the LHC data, supporting the **consistency** of the global fit framework.
- Fit quality is comparable at NLO and NNLO, though the former is marginally better.

NLO	NNPDF2.3noLHC	NNPDF2.3
Total	1.101	1.121
NMCpd	0,93	0,93
NMC	1,59	1,61
SLAC	1,28	1,26
BCDMS	1,20	1,19
HERA-I	1,01	1,00
CHORUS	1,09	1,10
NuTeV	0,42	0,45
DYE605	0,85	0,88
DYE866	1,24	1,28
CDFWASY	1,45	1,54
CDFZRAP	1,77	1,79
D0ZRAP	0,57	0,57
ATLAS-WZ	1,37	1,27
CMS-WEASY	1,50	1,04
LHCb-WZ	1,24	1,21
CDFR2KT	0,60	0,61
D0R2CON	0,84	0,84
ATLAS-JETS-2010	1,57	1,55

NNPDF

NNPDF 2.3
LHC data: Impact and consistency (NNLO)

Compare the quality of the fit to LHC data **before** and **after** the **inclusion**

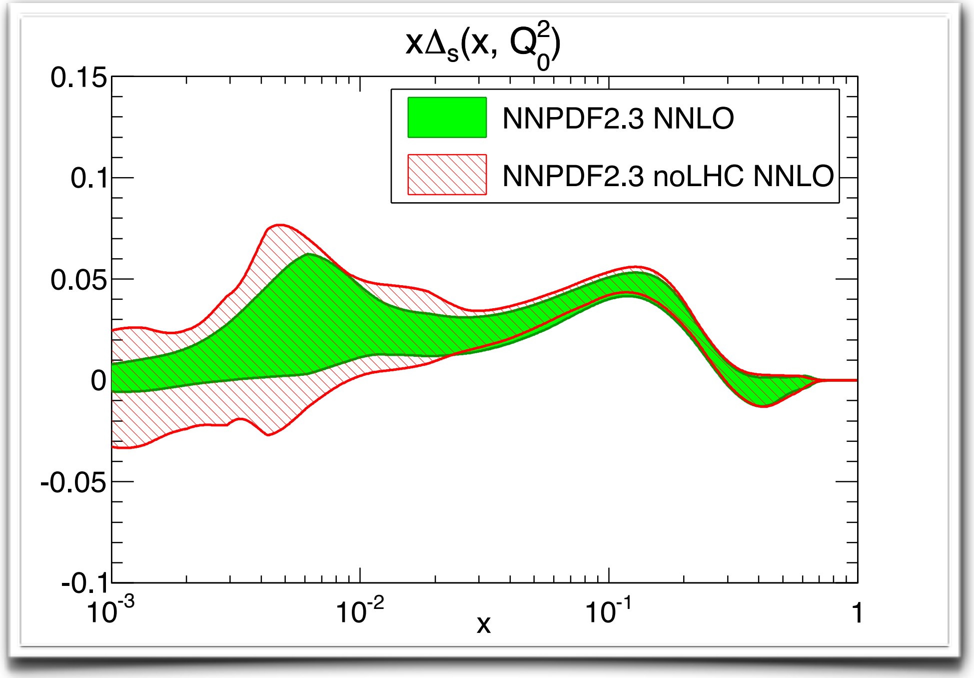
- Including LHC data in the fit improves the quality of their description **without deteriorating** the quality of the fit to other datasets.
- Moderate impact of the LHC data, supporting the **consistency** of the global fit framework.
- Fit quality is comparable at NLO and NNLO, though the former is marginally better.

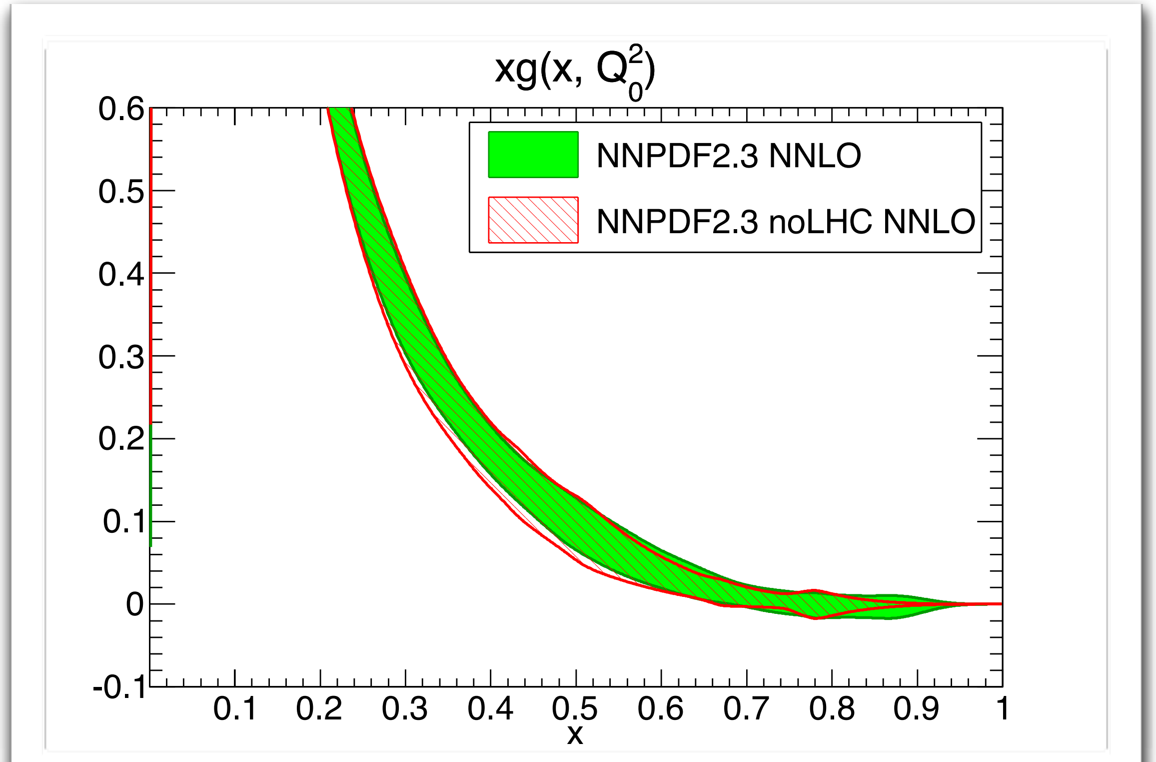
NNLO	NNPDF2.3noLHC	NNPDF2.3
Total	1.147	1.153
NMCpd	0,94	0,94
NMC	1,56	1,57
SLAC	1,04	1,02
BCDMS	1,28	1,29
HERA-I	1,03	1,01
CHORUS	1,07	1,06
NuTeV	0,48	0,55
DYE605	1,07	1,02
DYE866	1,61	1,62
CDFWASY	1,66	1,70
CDFZRAP	2,15	2,12
D0ZRAP	0,64	0,63
ATLAS-WZ	1,94	1,46
CMS-WEASY	1,37	0,96
LHCb-WZ	1,33	1,22
CDFR2KT	0,67	0,67
D0R2CON	0,94	0,93
ATLAS-JETS-2010	1,45	1,42

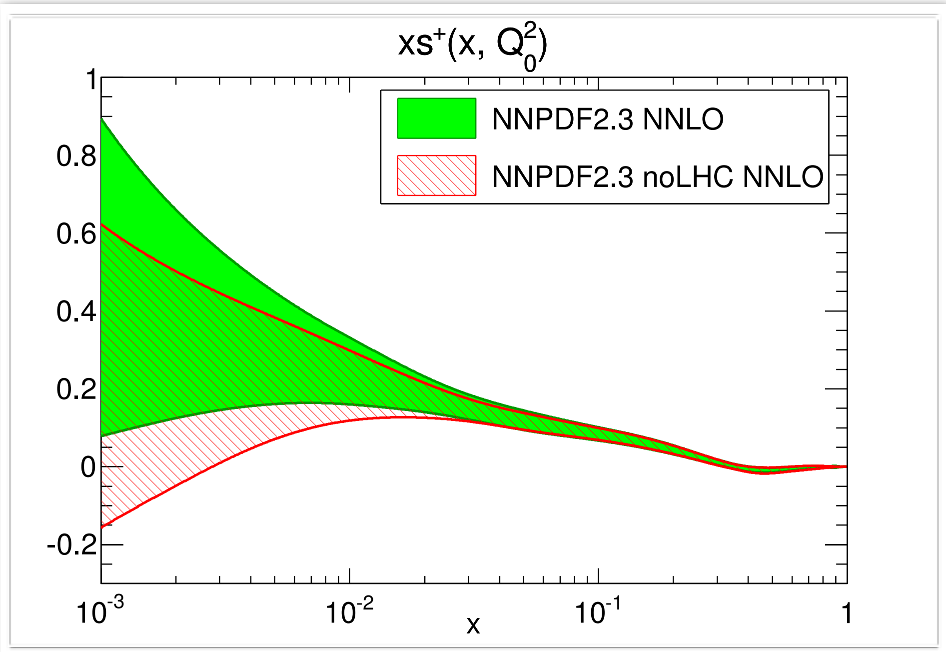
NNPDF

NNPDF 2.3
LHC data: Parton distributions

- More accurate light quark **flavor decomposition**:
  - due to the **LHC electroweak boson production** data.
- Reduction of the uncertainty in the **large *x* gluon**:
  - due to the **ATLAS jet** data.







NNPDF

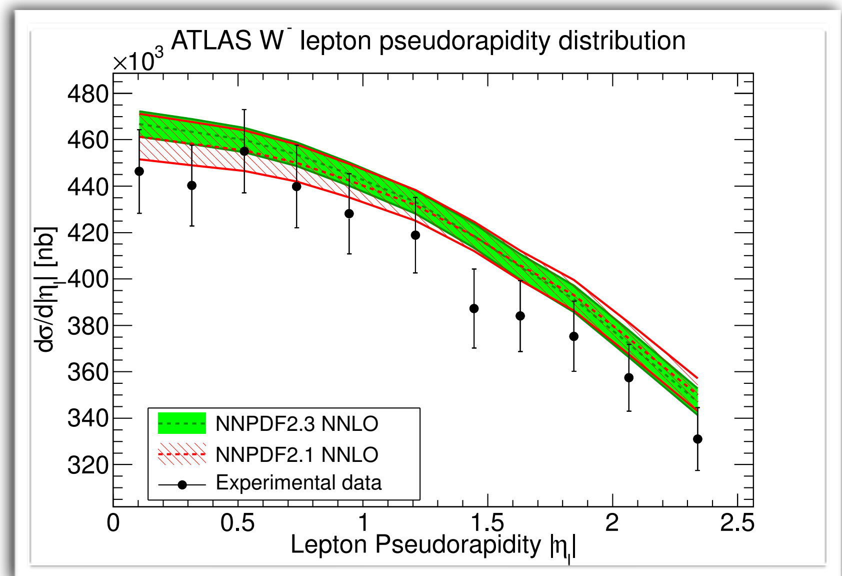
NNPDF 2.3

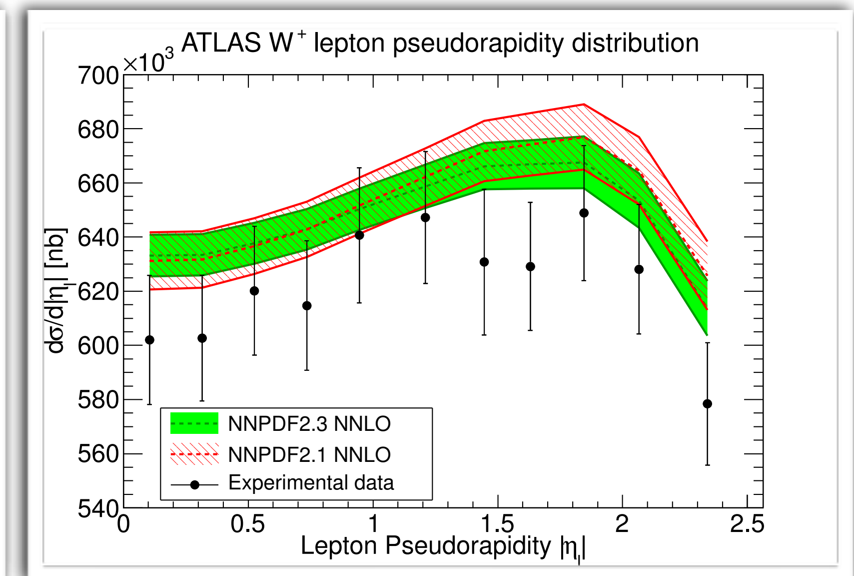
LHC data: Observables

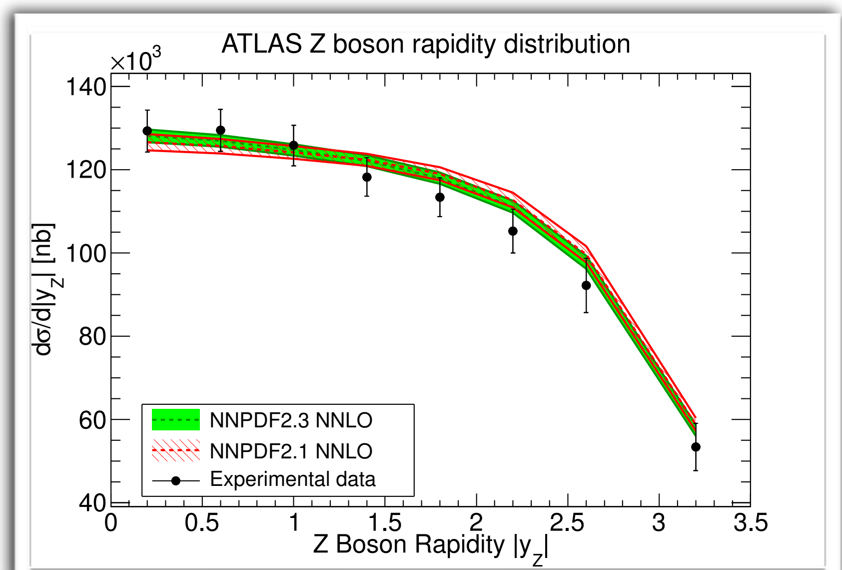
- Predictions for the LHC data already **acceptable before** including them in the fit.

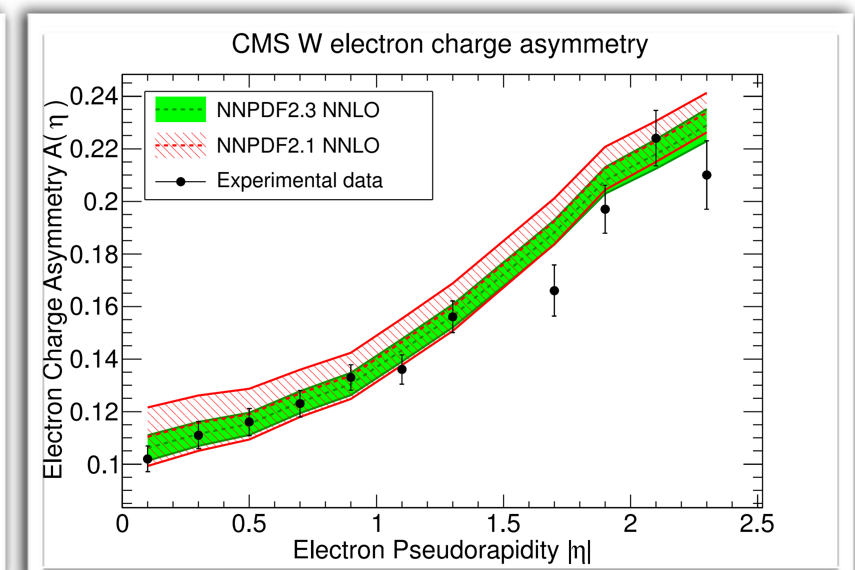
- **Quite good** description of the LHC data **after** the inclusion in the fit.

- Substantial **reduction of the uncertainty** on the observable predictions.





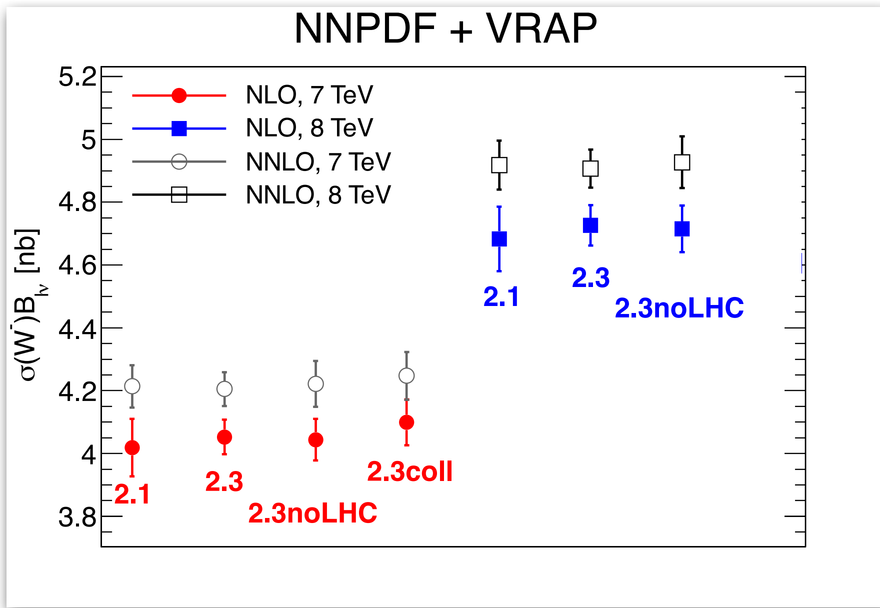


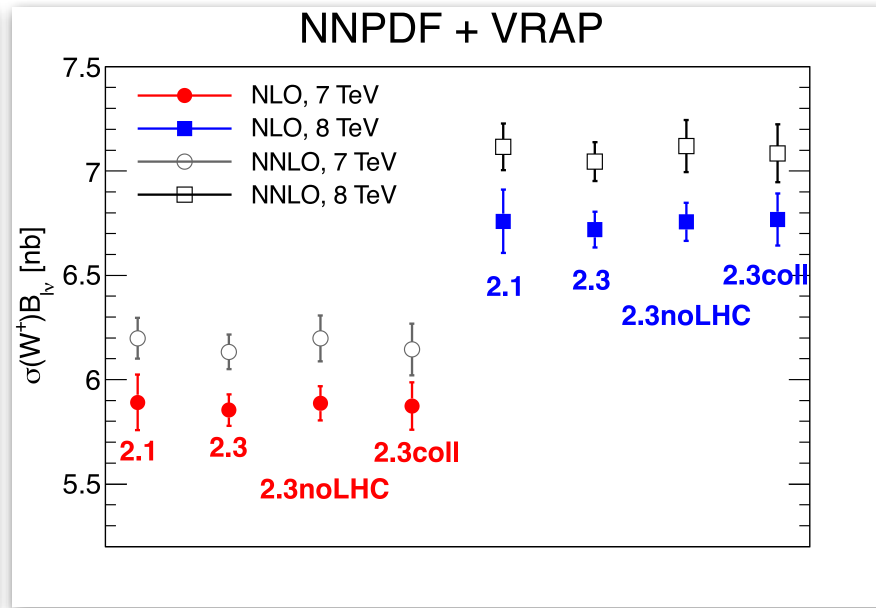


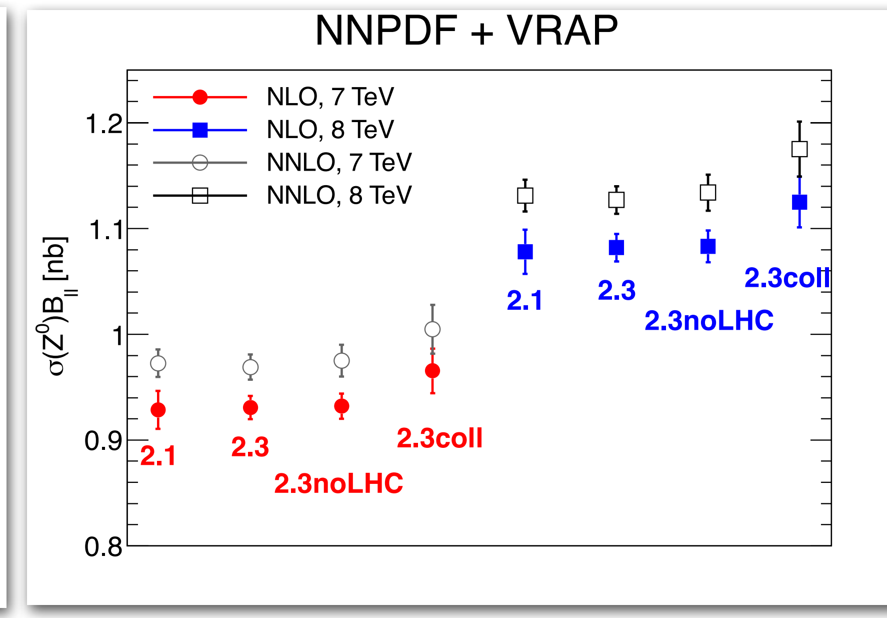
NNPDF

NNPDF 2.3

Phenomenology: W/Z production



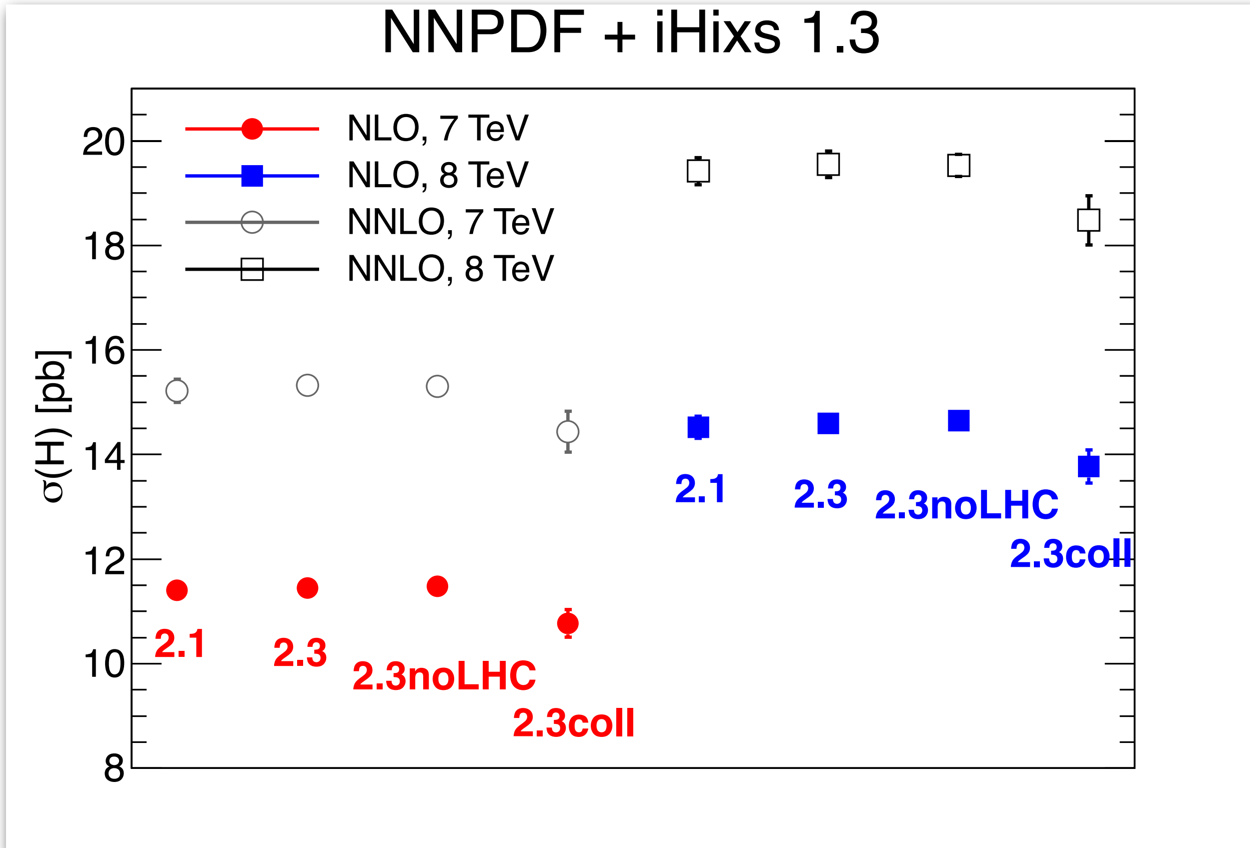


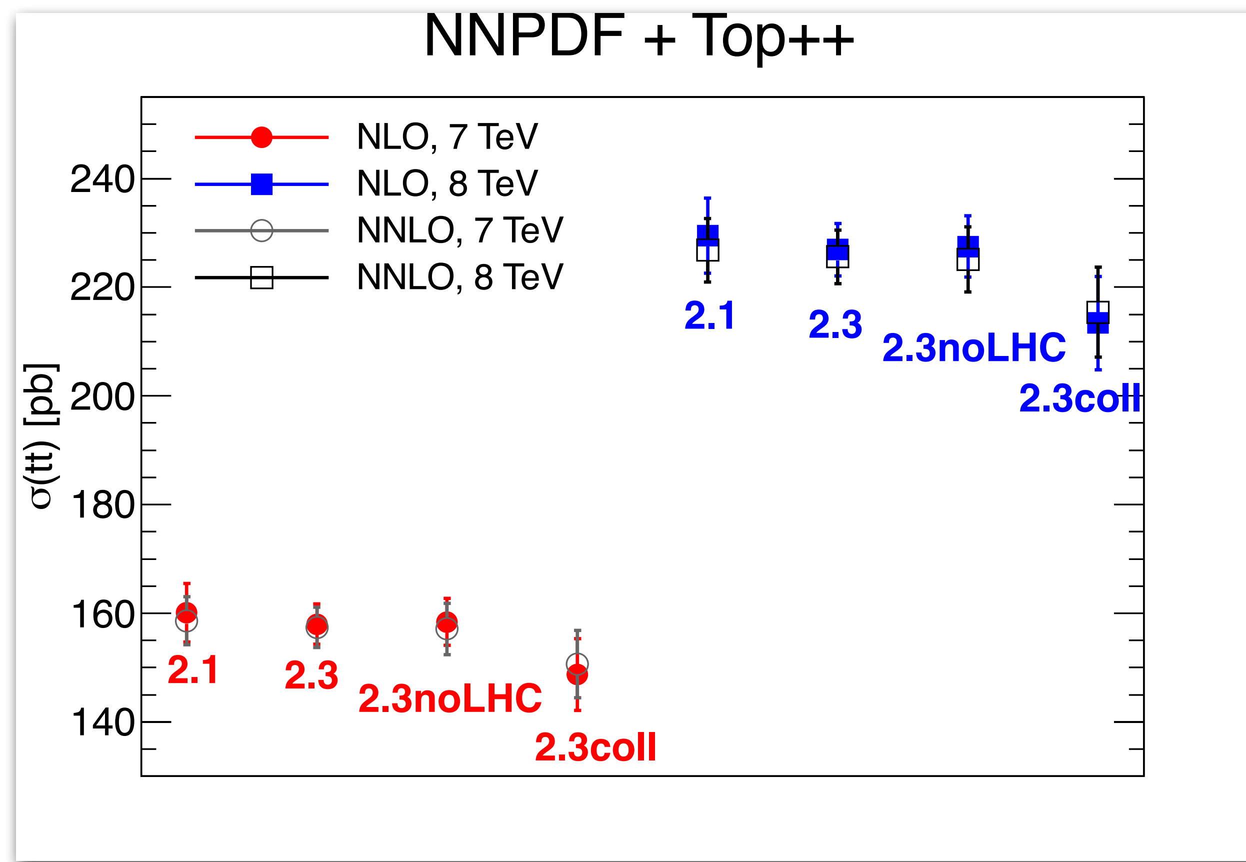


- Only PDF uncertainties shown,
- *W* and *Z* production mostly sensitive to the **quark-antiquark** luminosity,
- predictions with NNPDF 2.1 and NNPDF 2.3 **always compatible**,
- the **accuracy increases** when going from NNPDF 2.1 to 2.3:
  - partly due to the improved methodology (NNPDF 2.1 vs. NNPDF 2.3 noLHC),
  - partly due to the inclusion of the LHC data (NNPDF 2.3 vs. NNPDF 2.3 noLHC).

NNPDF 2.3

Phenomenology: Top & Higgs production



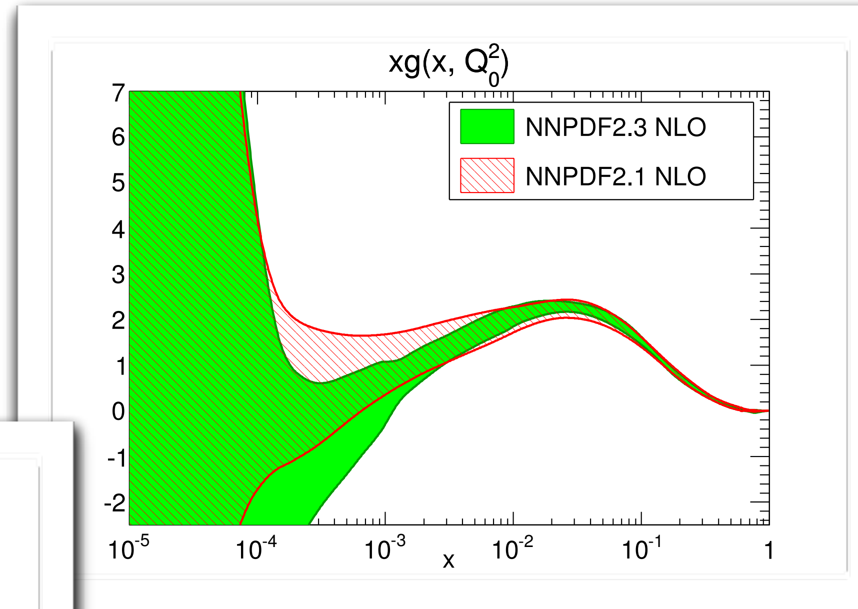


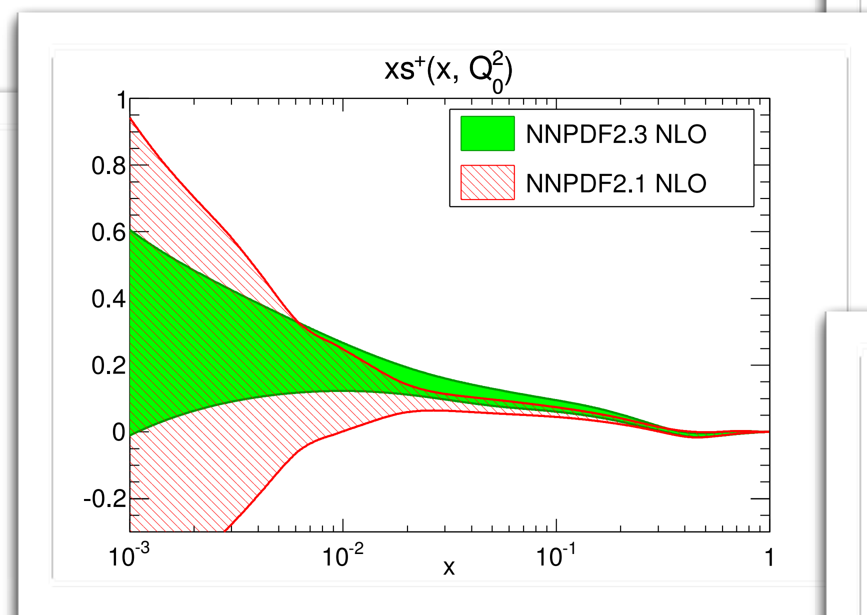
- Only PDF uncertainties shown,
- *Top* and *Higgs* production mostly sensitive to the **gluon-gluon** luminosity,
- predictions with NNPDF 2.1 and NNPDF 2.3 **always compatible**.

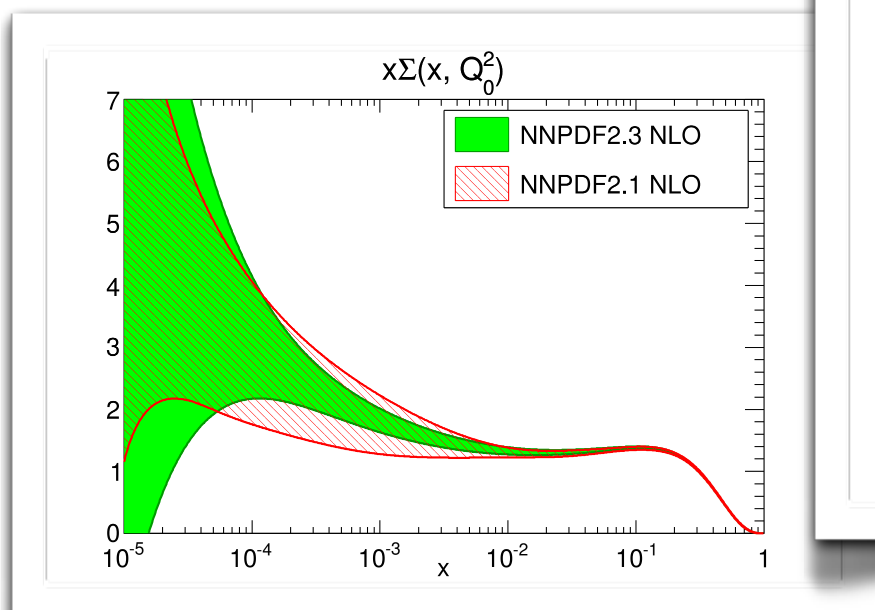
NNPDF 2.3

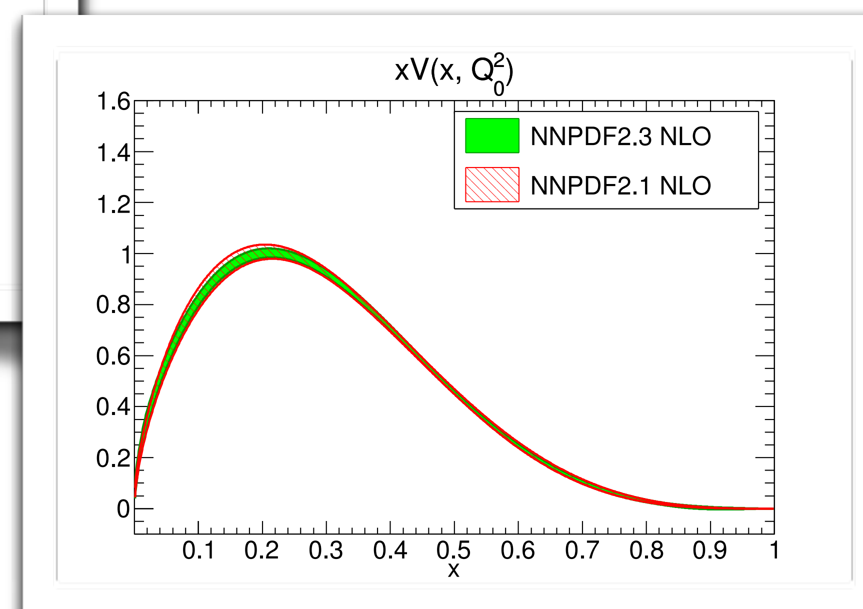
Parton Distributions: 2.3 vs. 2.1

- Addition of the **LHC data**,
- improved **minimization**,
- corrected error in the **di-muon cross-section**.









- Marked improvement of the fit quality at NLO,
- changes at the **half sigma** level for all PDFs,
- consequence of the **improved minimization**.

NNPDF