News from NNPDF

Alberto Guffanti

Niels Bohr International Academy & Discovery Center
Niels Bohr Institute, Copenhagen

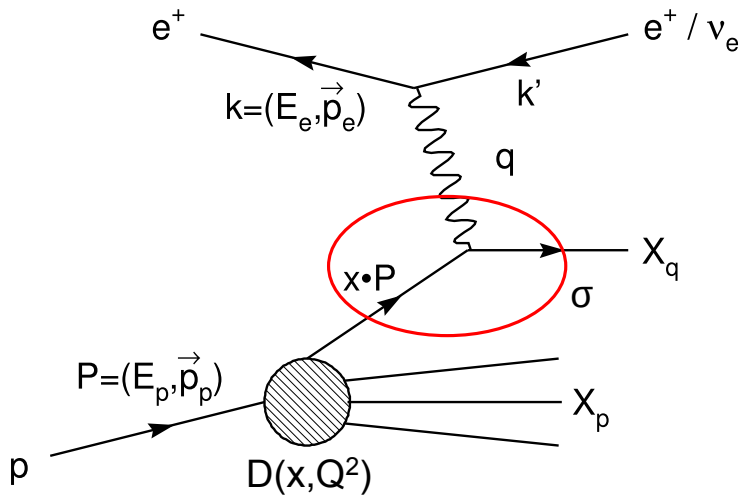
On behalf of

The NNPDF Collaboration
R. D. Ball, L. Del Debbio (Edinburgh), F. Cerutti, J. I. Latorre (Barcelona),
S. Forte, J. Rojo (Milano), V. Bertone(Freiburg), M. Ubiali(Aachen)

PANIC 2011
MIT (Cambridge, MA), July 24-29, 2011

What are Parton Distribution Functions?

- Consider a process with one hadron in the initial state



- According to the **Factorization Theorem** we can write the cross section as

$$d\sigma = \sum_a \int_0^1 \frac{d\xi}{\xi} D_a(\xi, \mu^2) d\hat{\sigma}_a \left(\frac{x}{\xi}, \frac{\hat{s}}{\mu^2}, \alpha_s(\mu^2) \right) + \mathcal{O}\left(\frac{1}{Q^p} \right)$$

What are Parton Distribution Functions?

- The **absolute value** of PDFs at a given x and Q^2 **cannot be computed** in QCD Perturbation Theory
  (Lattice? In principle yes, but ...)
- ... but their **scale dependence** is governed by **DGLAP** evolution equations

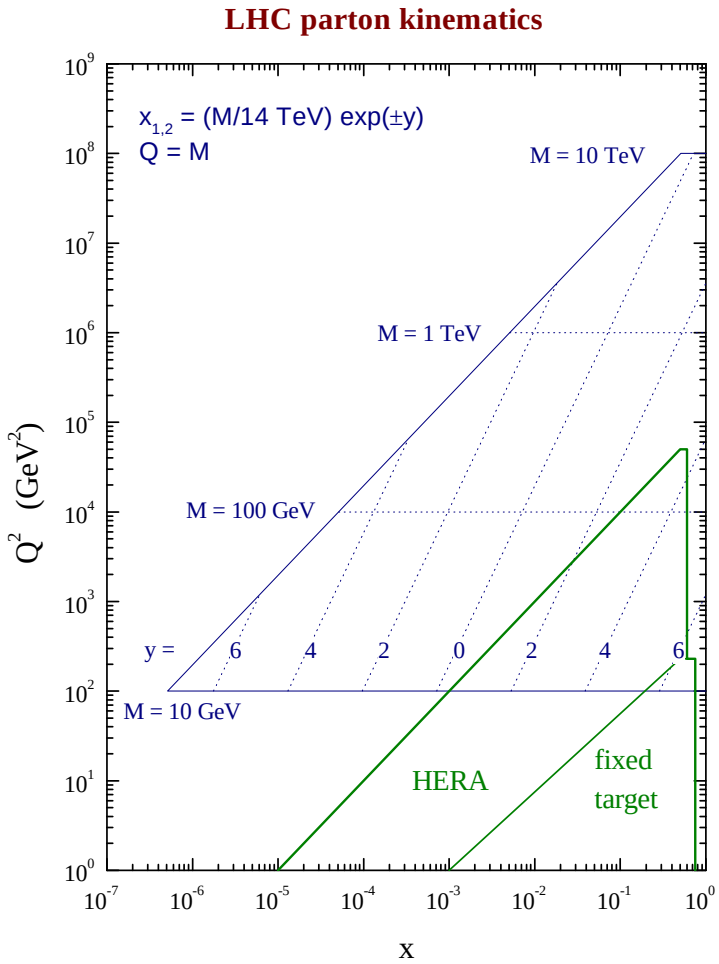
$$\frac{\partial}{\ln Q^2} q^{NS}(\xi, Q^2) = P^{NS}(\xi, \alpha_s) \otimes q^{NS}(\xi, Q^2)$$

$$\frac{\partial}{\ln Q^2} \begin{pmatrix} \Sigma \\ g \end{pmatrix} (\xi, Q^2) = \begin{pmatrix} P_{qq} & P_{qg} \\ P_{gq} & P_{gg} \end{pmatrix} (\xi, \alpha_s) \otimes \begin{pmatrix} \Sigma \\ g \end{pmatrix} (\xi, Q^2)$$

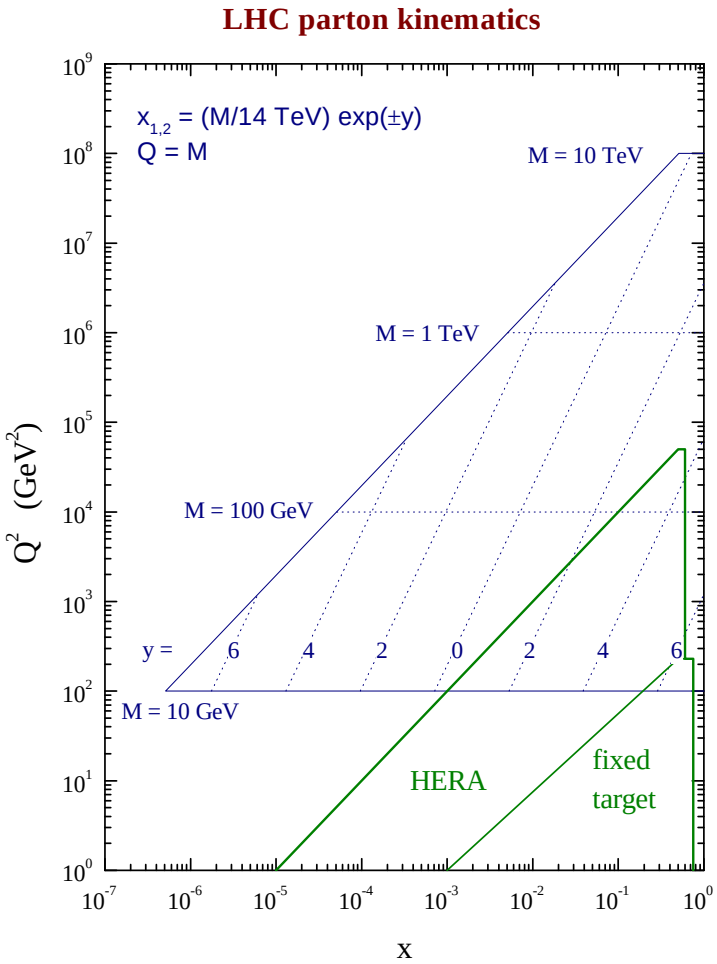
- ... where the **splitting functions** can be computed in PT and are known up to **NNLO**

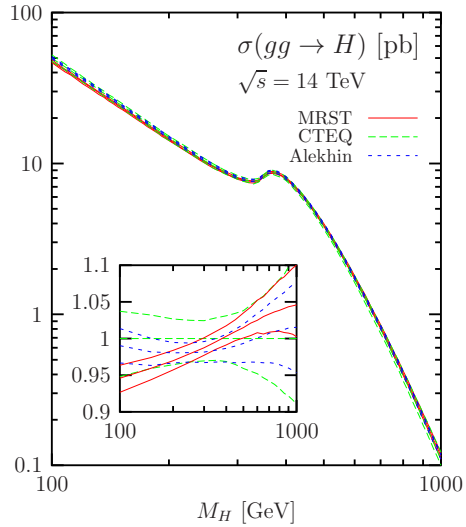
(**LO** - Dokshitzer; Gribov, Lipatov; Altarelli, Parisi; 1977)
(**NLO** - Floratos, Ross, Sachrajda; Gonzalez-Arroyo, Lopez, Yndurain; Curci, Furmanski, Petronzio, 1981)
(**NNLO** - Moch, Vermaseren, Vogt; 2004)

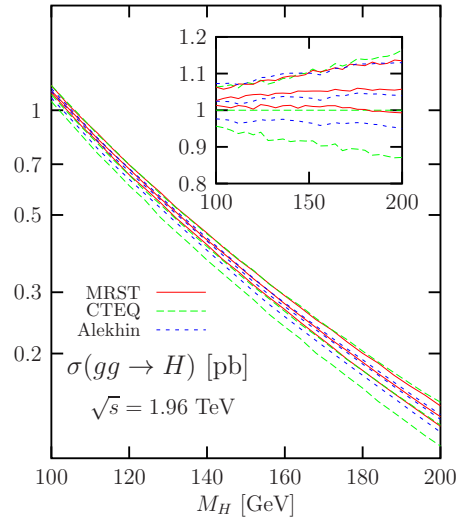
Why care about PDFs (and their uncertainties)?



Why care about PDFs (and their uncertainties)?

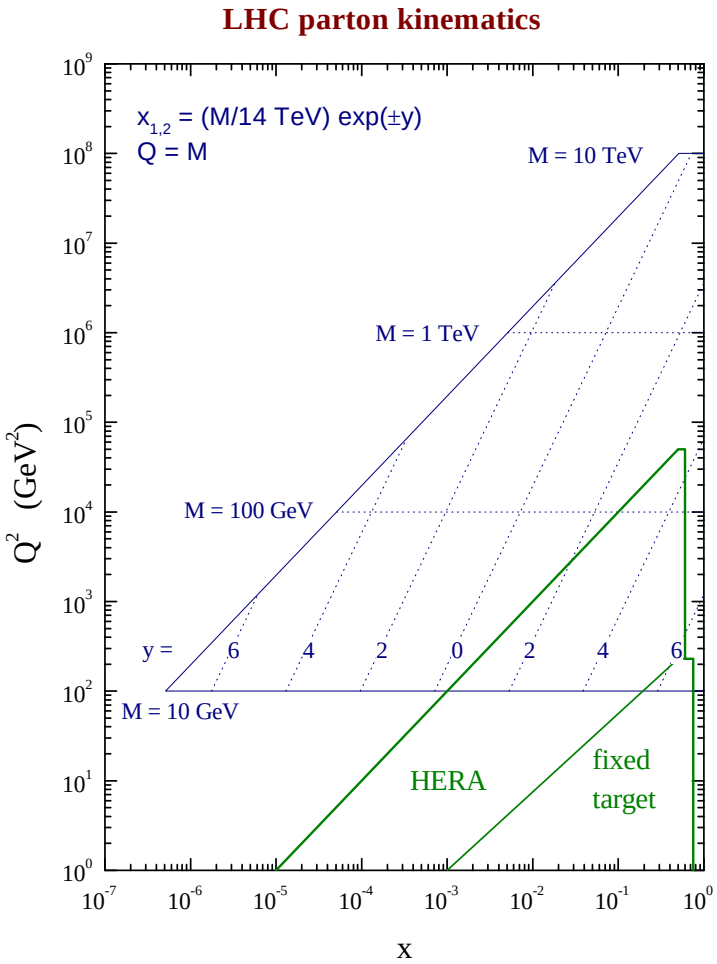


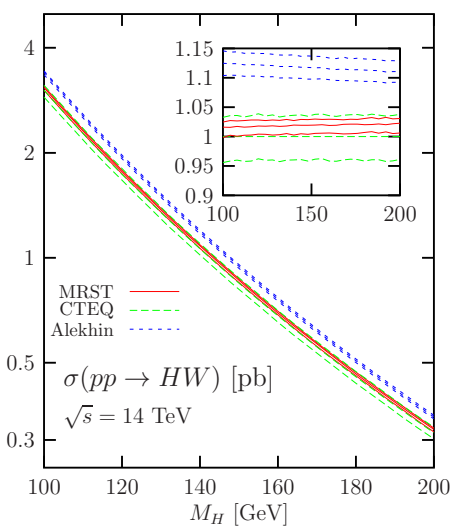


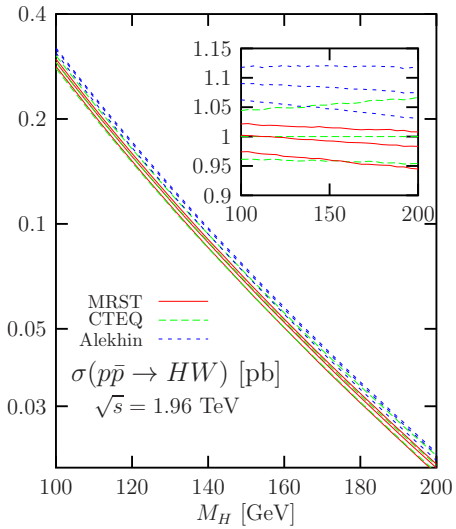


[A. Djouadi and S. Ferrag, hep-ph/0310209]

Why care about PDFs (and their uncertainties)?







[A. Djouadi and S. Ferrag, hep-ph/0310209]

Why care about PDFs (and their uncertainties)?

- Errors on PDFs are in some cases the **dominating theoretical error** on precision observables

Ex. $\sigma(Z^0)$ at the LHC: $\delta_{PDF} \sim 3\%$, $\delta_{NNLO} \sim 2\%$

[J. Campbell, J. Huston and J. Stirling, (2007)]

Why care about PDFs (and their uncertainties)?

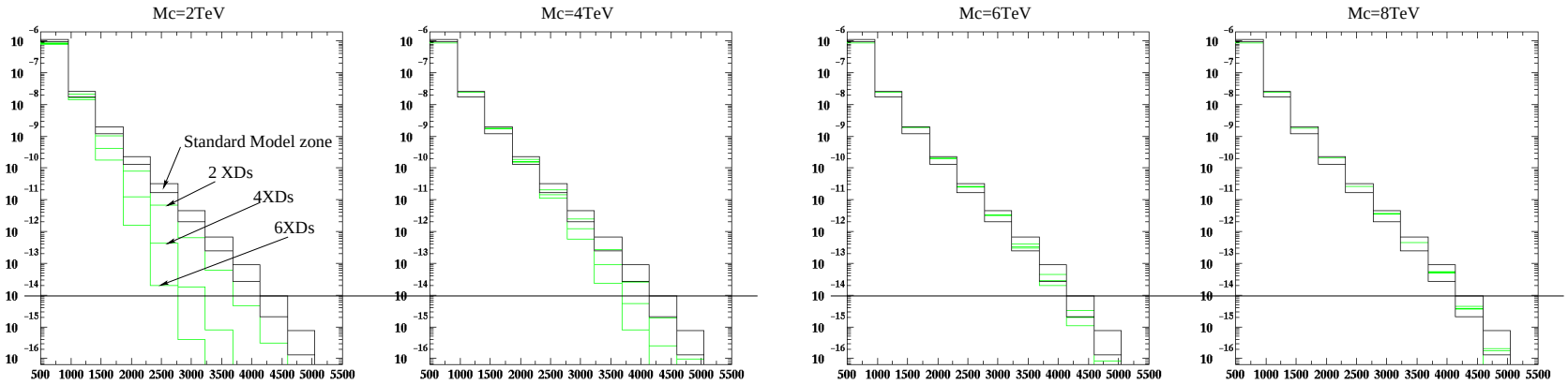
- Errors on PDFs are in some cases the **dominating theoretical error** on precision observables

Ex. $\sigma(Z^0)$ at the LHC: $\delta_{PDF} \sim 3\%$, $\delta_{NNLO} \sim 2\%$

[J. Campbell, J. Huston and J. Stirling, (2007)]

- Errors on PDFs might **reduce sensitivity to New Physics**

Ex. Extra Dimensions discovery in dijet cross section at the LHC:



[S. Ferrag (ATLAS), hep-ph/0407303]

Problem

Faithful estimation of errors on PDFs

- Single observable: **1-σ interval**
- Correlated observables: **1-σ contours**
- Function: need an **"error band" in the space of functions**
  (*i.e.* the probability density $\mathcal{P}[f]$ in the space of functions $f(x)$)

Expectation values are Functional integrals

$$\langle \mathcal{F}[f(x)] \rangle = \int \mathcal{D}f \mathcal{F}[f(x)] \mathcal{P}[f(x)]$$

Problem

Faithful estimation of errors on PDFs

- Single observable: **1-σ interval**
- Correlated observables: **1-σ contours**
- Function: need an **"error band" in the space of functions**
  (*i.e.* the probability density $\mathcal{P}[f]$ in the space of functions $f(x)$)

Expectation values are Functional integrals

$$\langle \mathcal{F}[f(x)] \rangle = \int \mathcal{D}f \mathcal{F}[f(x)] \mathcal{P}[f(x)]$$

Determine a function from a finite set of data points

Solution

Standard Approach

- Introduce a simple functional form with enough free parameters

$$q(x, Q^2) = x^{\alpha}(1-x)^{\beta} P(x; \lambda_1, ..., \lambda_n).$$

- Fit parameters minimizing χ^2.

Solution

Standard Approach

- Introduce a simple functional form with enough free parameters

$$q(x, Q^2) = x^{\alpha}(1-x)^{\beta} P(x; \lambda_1, ..., \lambda_n).$$

- Fit parameters minimizing χ^2.

Open problems:

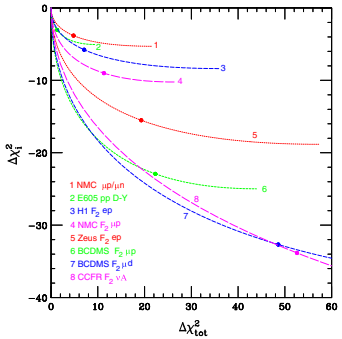
- **Error propagation** from data to parameters and from parameters to observables is **not trivial**.
- **Theoretical bias** due to the chosen **parametrization** is difficult to assess.

Shortcomings of the Standard approach

What is the meaning of a one-σ uncertainty?

- Standard $\Delta\chi^2 = 1$ criterion is **too restrictive** to account for large discrepancies among experiments.

[Collins & Pumplin, 2001]

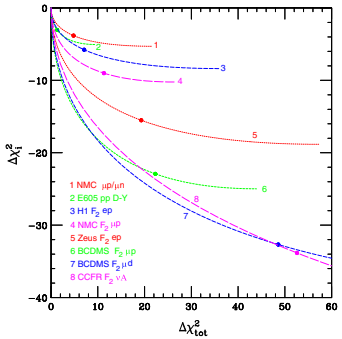


Shortcomings of the Standard approach

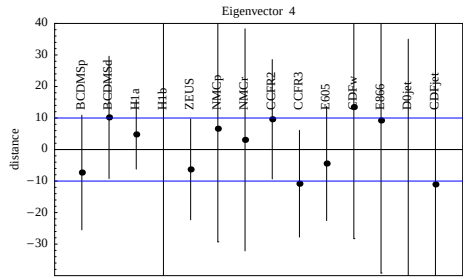
What is the meaning of a one-σ uncertainty?

- Standard $\Delta\chi^2 = 1$ criterion is **too restrictive** to account for large discrepancies among experiments.

[Collins & Pumplin, 2001]



- Introduce a **TOLERANCE** criterion, i.e. take the envelope of uncertainties of experiments to determine the $\Delta\chi^2$ to use for the global fit (CTEQ).

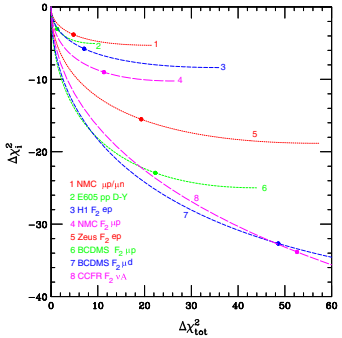


Shortcomings of the Standard approach

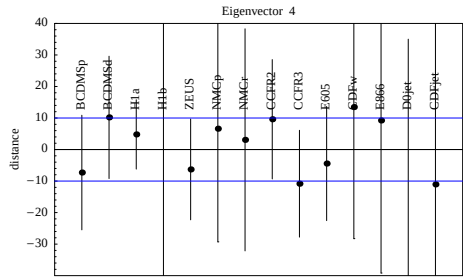
What is the meaning of a one-σ uncertainty?

- Standard $\Delta\chi^2 = 1$ criterion is **too restrictive** to account for large discrepancies among experiments.

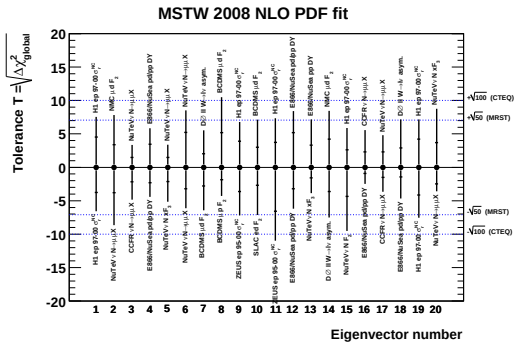
  [Collins & Pumplin, 2001]



- Introduce a **TOLERANCE** criterion, i.e. take the envelope of uncertainties of experiments to determine the $\Delta\chi^2$ to use for the global fit (CTEQ).



- Make it **DYNAMICAL**, i.e. determine $\Delta\chi^2$ separately for each hessian eigenvector (MSTW).

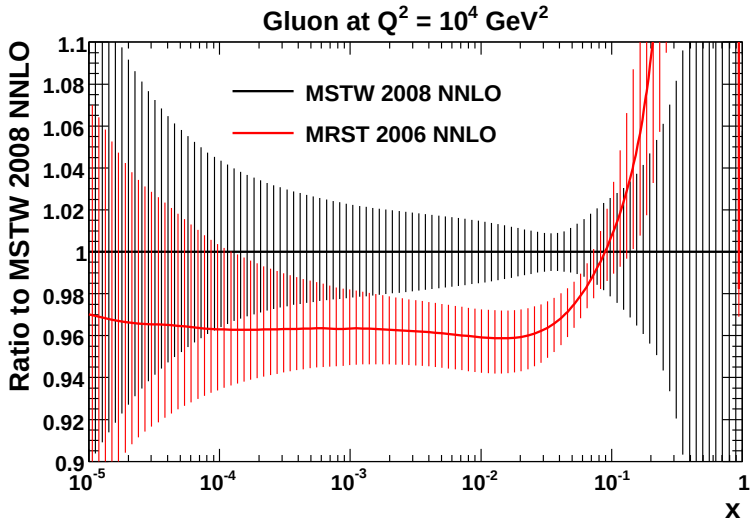


Shortcomings of the standard approach

What determines PDF uncertainties?

- Uncertainties in standard fits often increase when adding new data to the fit.
- Related to the need of extending the parametriztion in order to accomodate the new data

Smaller high-x gluon (and slightly smaller α_S) results in larger small-x gluon – now shown at NNLO.



Larger small-x uncertainty due to extrat free parameter.

[R. Thorne, PDF4LHC]

NNPDF Methodology

Main Ingredients

- **Monte Carlo** determination of errors
  - No need to rely on linear propagation of errors
  - Possibility to test for the impact of non gaussianly distributed errors
  - Possibility to test for non-gaussian behaviour in fitted PDFs ($1-\sigma$ vs. 68% CL)
- **Neural Networks**
  - Provide an **unbiased** parametrization
- **Stopping based on Cross-Validation**
  - Ensures proper fitting avoiding overlearning

NNPDF Methodology

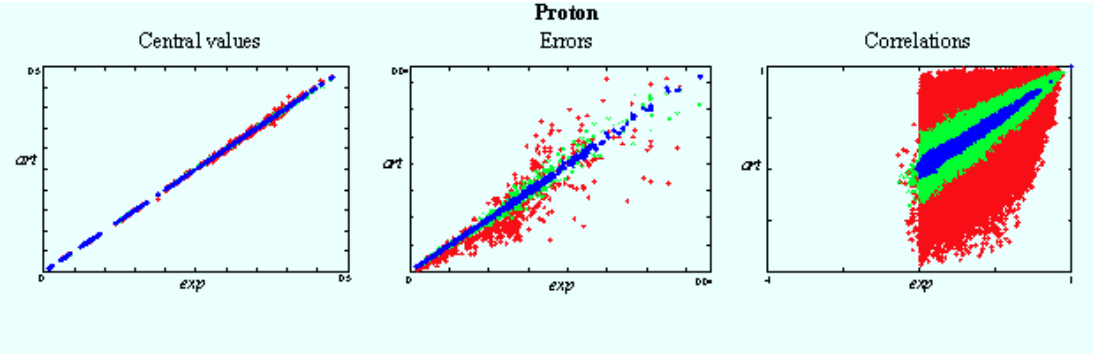
Monte Carlo replicas generation

- Generate artificial data according to distribution

$$O_i^{(art)\,(k)} = (1 + r_N^{(k)}\,\sigma_N)\left[O_i^{(exp)} + \sum_{p=1}^{N_{sys}} r_p^{(k)}\,\sigma_{i,p} + r_{i,s}^{(k)}\,\sigma_s^i\right]$$

where r_i are univariate gaussian random numbers

- Validate Monte Carlo replicas against experimental data
(statistical estimators, faithful representation of errors, convergence rate increasing N_{rep})



- $\mathcal{O}(1000)$ replicas needed to reproduce correlations to percent accuracy

NNPDF Methodology

Neural Networks ... a suitable basis of functions

- We use **Neural Networks** as **functions** to represent **PDFs at the starting scale**
- We employ **Multilayer Feed-Forward** Neural Networks trained using a **Genetic Algorithm**
- Activation determined by weights and thresholds

$$\xi_i = g\left(\sum_j \omega_{ij}\xi_j - \theta_i\right), \qquad g(x) = \frac{1}{1+e^{-\beta x}}$$

NNPDF Methodology

Neural Networks ... a suitable basis of functions

- We use **Neural Networks** as **functions** to represent **PDFs at the starting scale**
- We employ **Multilayer Feed-Forward** Neural Networks trained using a **Genetic Algorithm**
- Activation determined by weights and thresholds

$$\xi_i = g\left(\sum_j \omega_{ij}\xi_j - \theta_i\right), \qquad g(x) = \frac{1}{1+e^{-\beta x}}$$

Ex.: 1-2-1 NN:

$$\xi_1^{(3)}(\xi_1^{(1)}) = \frac{1}{1+e^{\theta_1^{(3)} - \frac{\omega_{11}^{(2)}}{1+e^{\theta_1^{(2)} - \xi_1^{(1)}\omega_{11}^{(1)}}} - \frac{\omega_{12}^{(2)}}{1+e^{\theta_2^{(2)} - \xi_1^{(1)}\omega_{21}^{(1)}}}}}$$

NNPDF Methodology

Neural Networks ... a suitable basis of functions

- We use **Neural Networks** as **functions** to represent **PDFs at the starting scale**
- We employ **Multilayer Feed-Forward** Neural Networks trained using a **Genetic Algorithm**
- Activation determined by weights and thresholds

$$\xi_i = g\left(\sum_j \omega_{ij}\xi_j - \theta_i\right), \qquad g(x) = \frac{1}{1+e^{-\beta x}}$$

Ex.: 1-2-1 NN:

$$\xi_1^{(3)}(\xi_1^{(1)}) = \frac{1}{1+e^{\theta_1^{(3)} - \frac{\omega_{11}^{(2)}}{1+e^{\theta_1^{(2)} - \xi_1^{(1)}\omega_{11}^{(1)}}} - \frac{\omega_{12}^{(2)}}{1+e^{\theta_2^{(2)} - \xi_1^{(1)}\omega_{21}^{(1)}}}}}$$

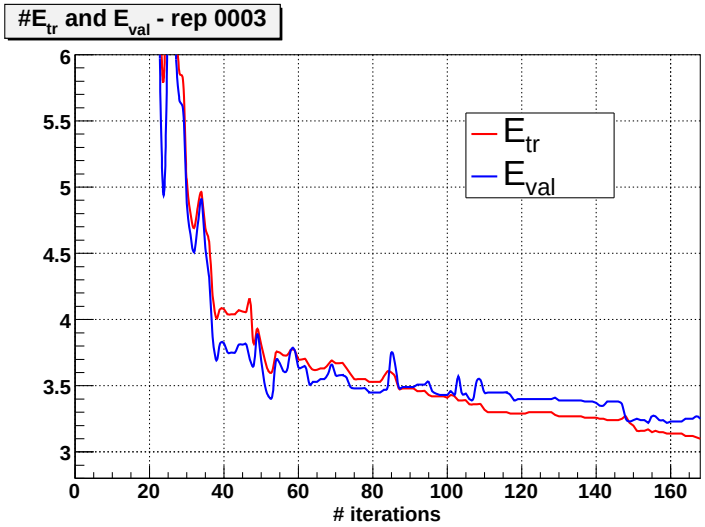
- They provide a parametrization which is **redundant** and **robust** against variations

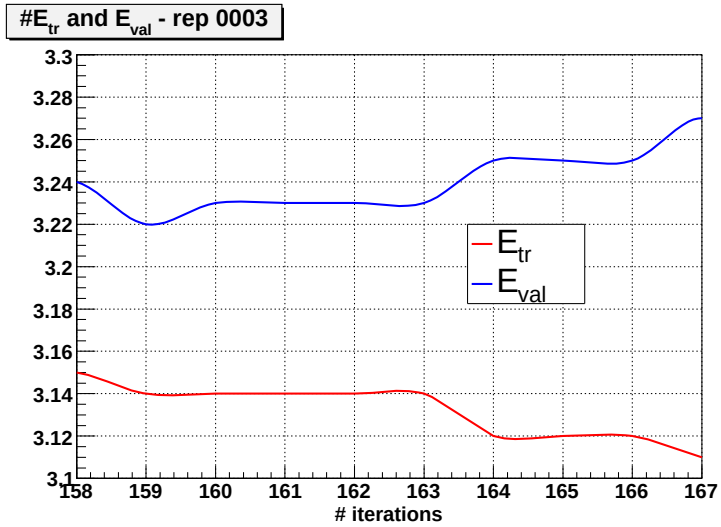
Stopping criterion based on Training-Validation separation

- Divide the data in two sets: Training and Validation
- Minimize the χ^2 of the data in the Training set
- Compute the χ^2 for the data in the Validation set
- When Validation χ^2 stops decreasing, **STOP** the fit

Stopping criterion based on Training-Validation separation

- Divide the data in two sets: Training and Validation
- Minimize the χ^2 of the data in the Training set
- Compute the χ^2 for the data in the Validation set
- When Validation χ^2 stops decreasing, **STOP** the fit





NNPDF Methodology

... in a Nutshell

- Generate N_{rep} **Monte-Carlo replicas** of the experimental data
  (sampling of the probability density in the space of data)
- Fit a set of Parton Distribution Functions on each replica
  (sampling of the probability density in the space of PDFs)
- **Expectation values** for observables are **Monte Carlo integrals**

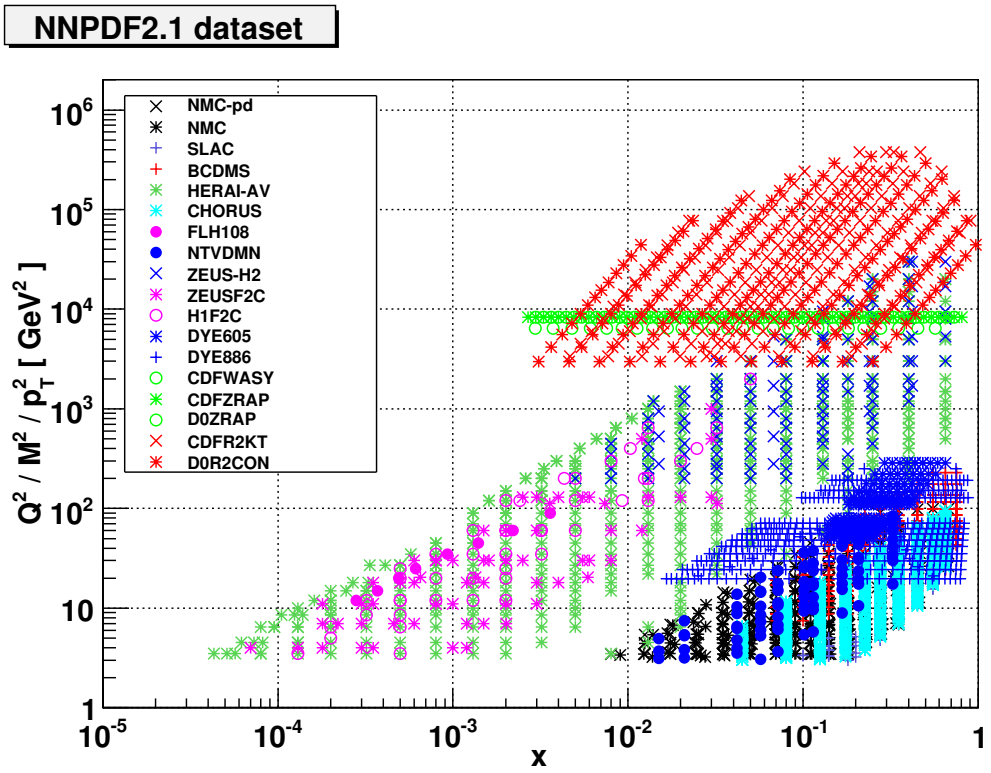
$$\langle \mathcal{F}[f_i(x,Q^2)] \rangle = \frac{1}{N_{rep}} \sum_{k=1}^{N_{rep}} \mathcal{F}\left(f_i^{(net)(k)}(x,Q^2) \right)$$

... the same is true for errors, correlations, etc.

NNPDF 2.1

Dataset

[R. D. Ball et. al, arXiv:1101.1300] - **NLO**
[R. D. Ball et. al, arXiv:1107.2652] - **LO/NNLO**



3415 data points

(for comparison MSTW08 includes 2699 data points)

OBS	*Data set*
Deep Inelastic Scattering	
F_2^d / F_2^p	NMC-pd
F_2^p	NMC, SLAC, BCDMS
F_2^d	SLAC, BCDMS
$\sigma_{NC}^{\pm}$	HERA-I, ZEUS (HERA-II)
$\sigma_{CC}^{\pm}$	HERA-I, ZEUS (HERA-II)
F_L	H1
σ_ν , $\sigma_{\bar{\nu}}$	CHORUS
dimuon prod.	NuTeV
F_2^c	ZEUS, H1
Drell-Yan & Vector Boson prod.	
$d\sigma^{\rm DY} / dM^2 dy$	E605
$d\sigma^{\rm DY} / dM^2 dx_F$	E866
W asymm.	CDF
Z rap. distr.	D0/CDF
Inclusive jet prod.	
Incl. $\sigma^{(\rm jet)}$	CDF (k_T) - Run II
Incl. $\sigma^{(\rm jet)}$	D0 (cone) - Run II

NNPDF 2.1

Heavy Flavour treatment - FONLL

- We adopt the **FONLL** General Mass-Variable Flavour Number Scheme

  [M. Cacciari, M. Greco and P. Nason, (1998)]

  [S. Forte, P. Nason E. Laenen and J. Rojo, (2010)]

- FONLL gives a prescription to **combine FFN** (Massive) and **ZM-VFN** (Massless) computations, at any given order, **avoiding double counting**.
- With results available three implementations of FONLL are possibile:
  - **FONLL-A**: $\mathcal{O}(\alpha_s)$ Massless + $\mathcal{O}(\alpha_s)$ Massive - (NLO fit)
  - **FONLL-B**: $\mathcal{O}(\alpha_s)$ Massless + $\mathcal{O}(\alpha_s^2)$ Massive
  - **FONLL-C**: $\mathcal{O}(\alpha_s^2)$ Massless + $\mathcal{O}(\alpha_s^2)$ Massive - (NNLO fit)
- **Fixed Flavour Number Scheme** (3-, 4-, 5-) fits **available**.

NNPDF 2.1

Parametrization

Parton Distributions Combination		NN architechture
Singlet ($\Sigma(x)$)	$\Longrightarrow$	2-5-3-1 (37 pars)
Gluon ($g(x)$)	$\Longrightarrow$	2-5-3-1 (37 pars)
Total valence ($V(x) \equiv u_V(x) + d_V(x)$)	$\Longrightarrow$	2-5-3-1 (37 pars)
Non-singlet triplet ($T_3(x)$)	$\Longrightarrow$	2-5-3-1 (37 pars)
Sea asymmetry ($\Delta_S(x) \equiv \bar{d}(x) - \bar{u}(x)$)	$\Longrightarrow$	2-5-3-1 (37 pars)
Total Strangeness ($s^+(x) \equiv (s(x) + \bar{s}(x))/2$)	$\Longrightarrow$	2-5-3-1 (37 pars)
Strange valence ($s^-(x) \equiv (s(x) - \bar{s}(x))/2$)	$\Longrightarrow$	2-5-3-1 (37 pars)

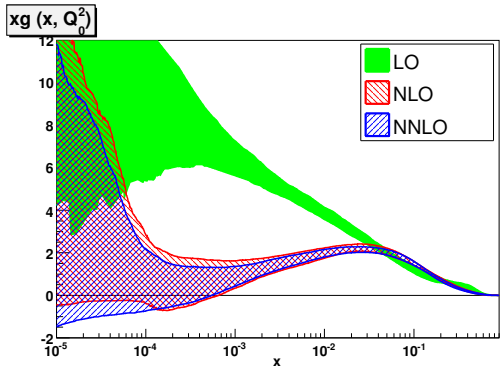
259 parameters

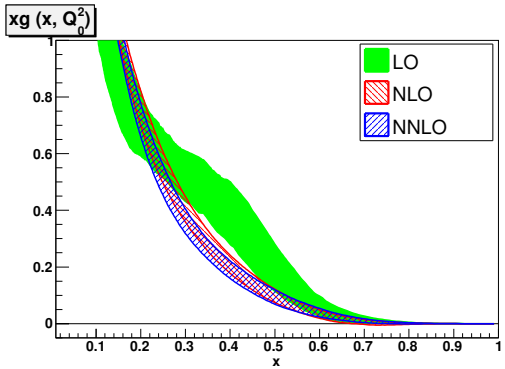
Standard fits have $\sim$ 25 parameters in total

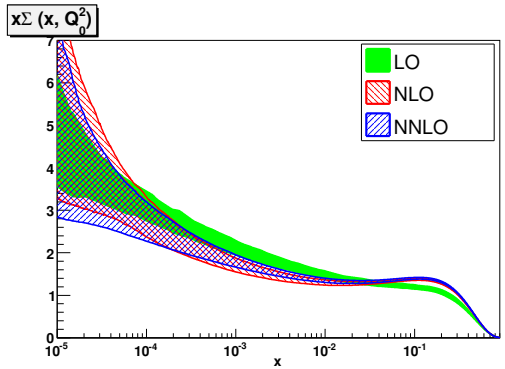
No change in the parametrization since NNPDF1.2 ... despite substantial **enlargement of the dataset**.

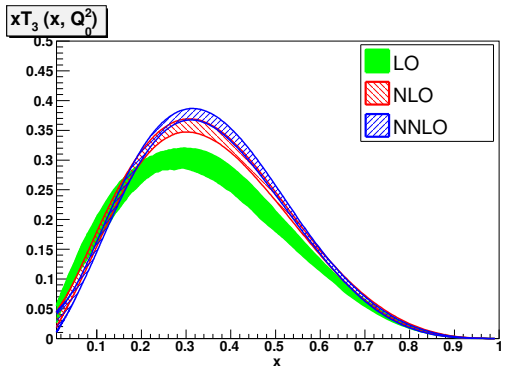
NNPDF2.1@NNLO

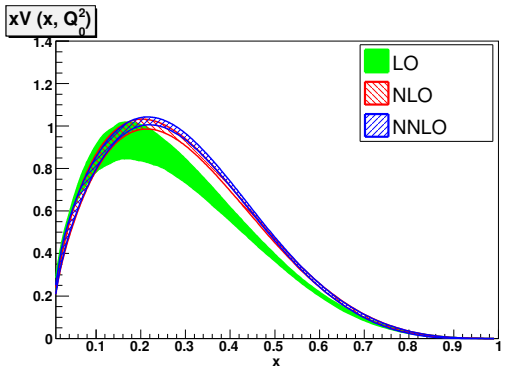
Partons - Comparison to NLO

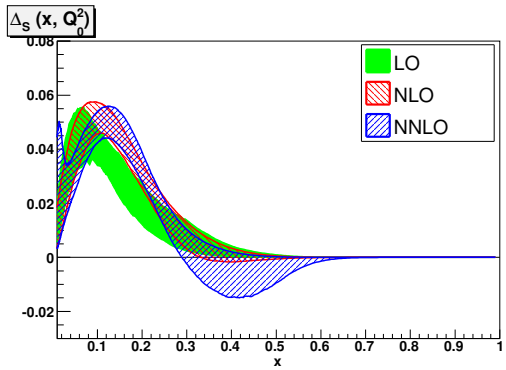


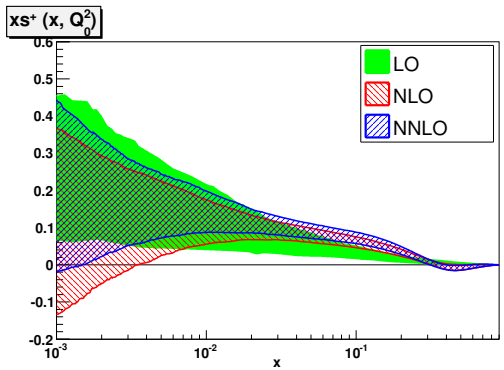


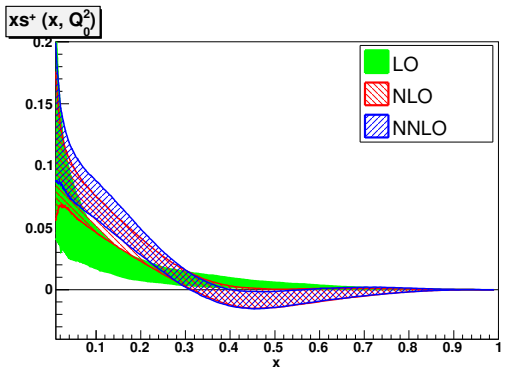


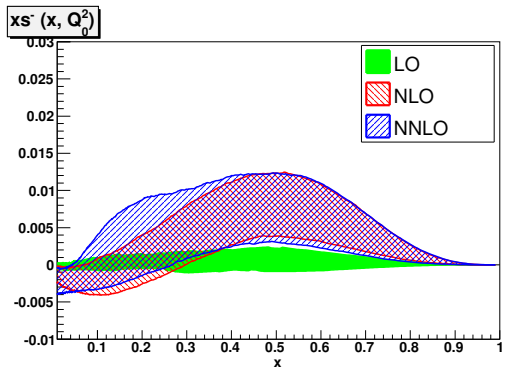






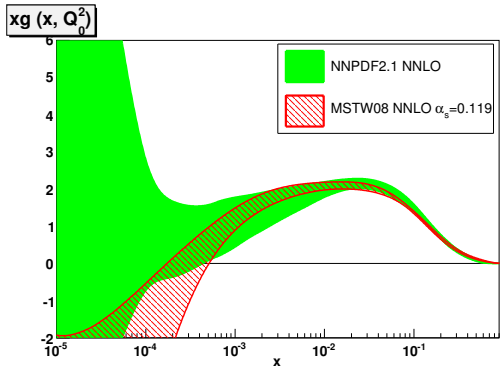


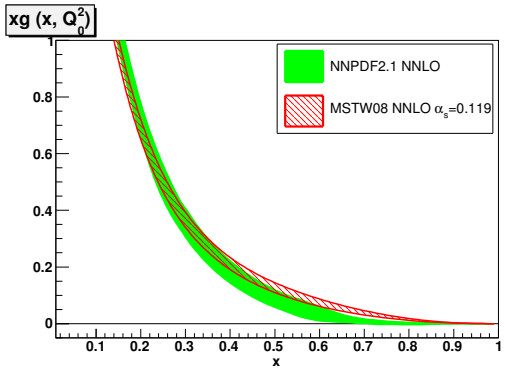


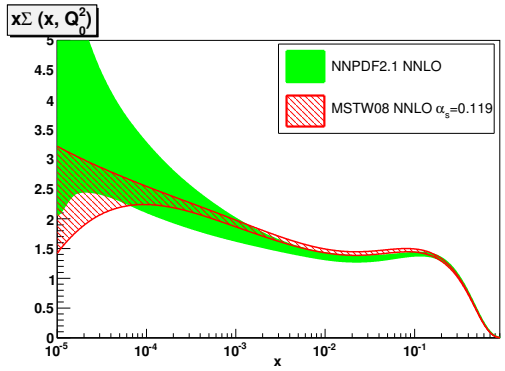


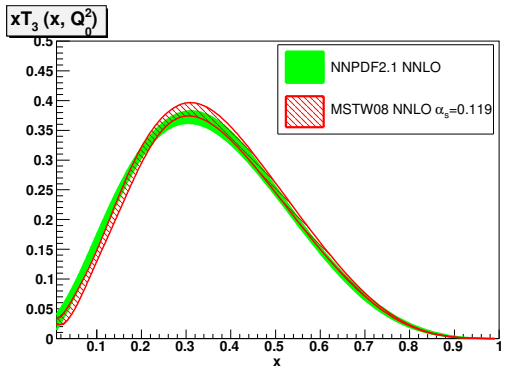
NNPDF2.1@NNLO

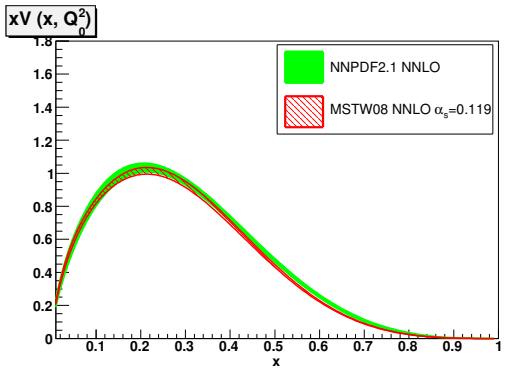
Partons - Comparison to MSTW08

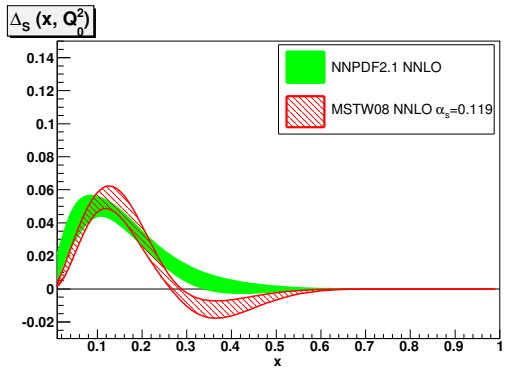


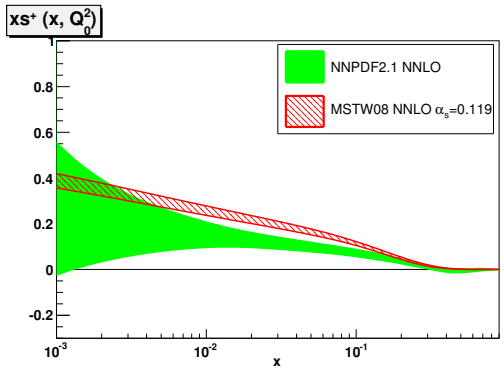


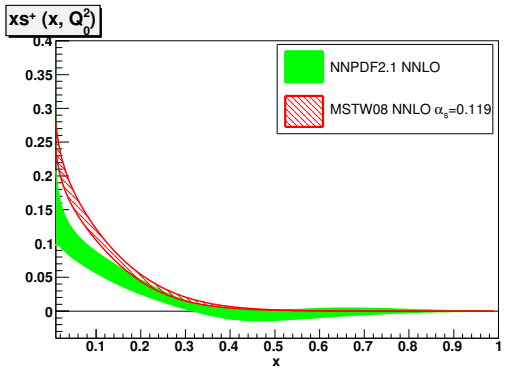


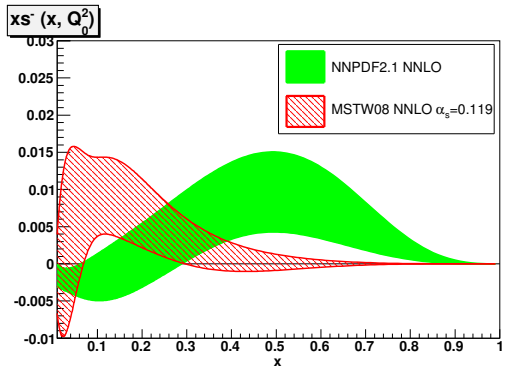








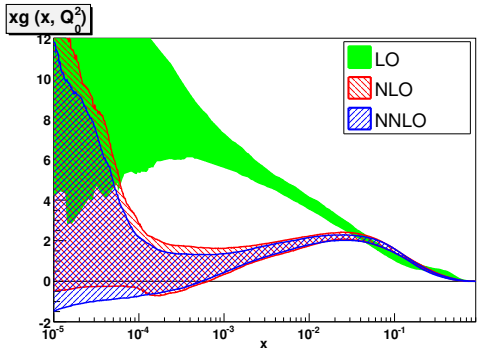




NNPDF2.1@NNLO

Partons - A couple of upshots

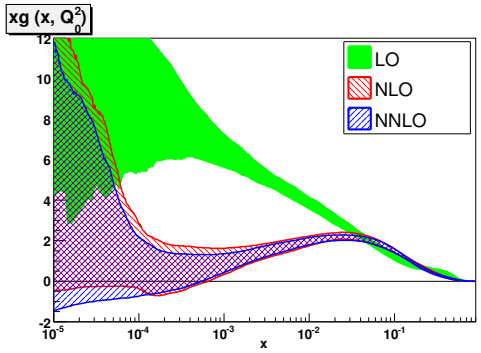
- Stability of parton determined when increasing the perturbative order of the analysis.



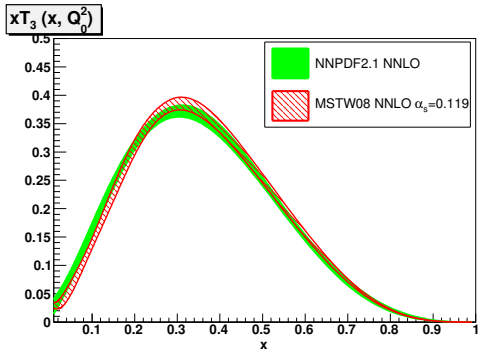
NNPDF2.1@NNLO

Partons - A couple of upshots

- Stability of parton determined when increasing the perturbative order of the analysis.



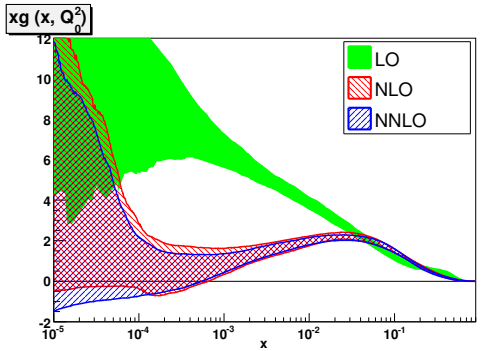
- **Uncertainties** on PDFs have **size comparable** to those obtained by other groups in kinematic regions where there are significant **contraints from data** ...



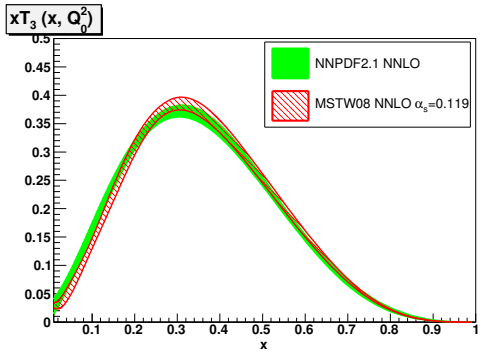
NNPDF2.1@NNLO

Partons - A couple of upshots

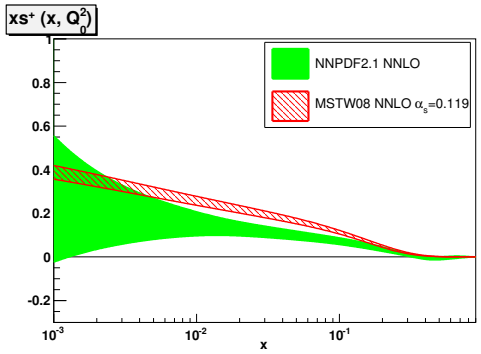
- Stability of parton determined when increasing the perturbative order of the analysis.



- **Uncertainties** on PDFs have **size comparable** to those obtained by other groups in kinematic regions where there are significant **contraints from data** ...



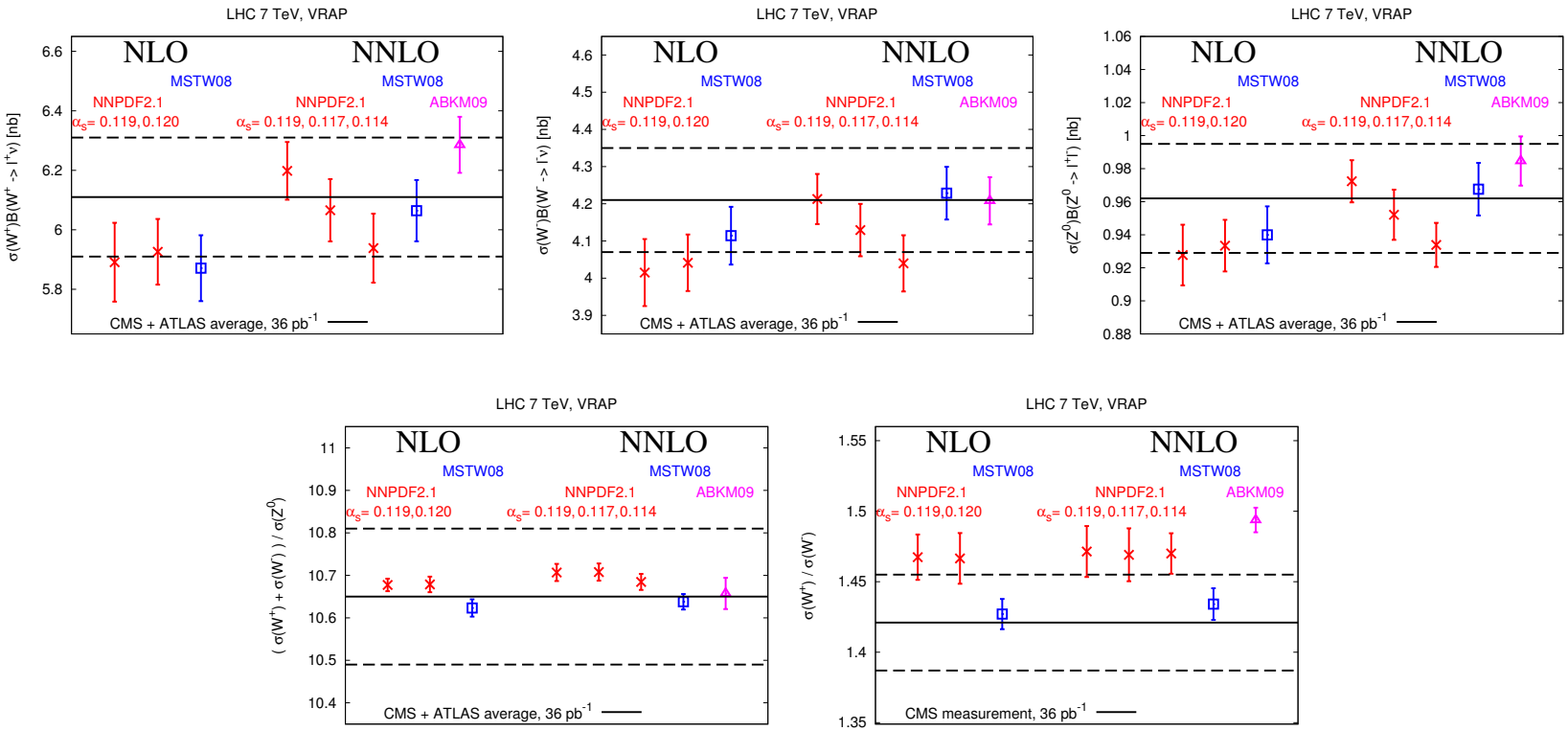
- ... but still retain **unbiasedness** in kinematic regions where there are little or **no experimental constraints**.



NNPDF2.1@NNLO

Phenomenology - LHC Standard Candles

- Predictions for **LHC Standard Candles** compared to other **theory predictions** & **LHC data**



Reweighting NNPDFs

Assessing the impact of new data on PDF fits

[R. D. Ball et al., arXiv:1012.0836]

- Originally inspired by an idea of Giele and Keller [hep-ph/9803393]
- The $N_{\rm rep}$ **replicas** of a NNPDF fit give the **probability density** in the space of PDFs
- **Expectation values** for observables are **Monte Carlo integrals**

$$\langle \mathcal{F}[f_i(x,Q^2)] \rangle = \frac{1}{N_{rep}} \sum_{k=1}^{N_{rep}} \mathcal{F}\left(f_i^{(net)(k)}(x,Q^2) \right)$$

(... the same is true for errors, correlations, etc.)

- We can **assess the impact** of including **new data** in the fit updating the probability density distribution.

Reweighting NNPDFs

Assessing the impact of new data on PDF fits

- According to **Bayes Theorem** we have

$$\mathcal{P}_{\rm new}(\{f\}) = \mathcal{N}_\chi \mathcal{P}(\chi^2|\{f\})\mathcal{P}_{\rm init}(\{f\}), \quad \mathcal{P}(\chi^2|\{f\}) = [\chi^2(y,\{f\})]^{\frac{n_{dat}-1}{2}} e^{-\frac{\chi^2(y,\{f\})}{2}}$$

- **Monte Carlo integrals** are now **weighted sums**

$$\langle \mathcal{F}[f_i(x,Q^2)]\rangle = \sum_{k=1}^{N_{rep}} w_k \mathcal{F}\left(f_i^{(net)(k)}(x,Q^2)\right)$$

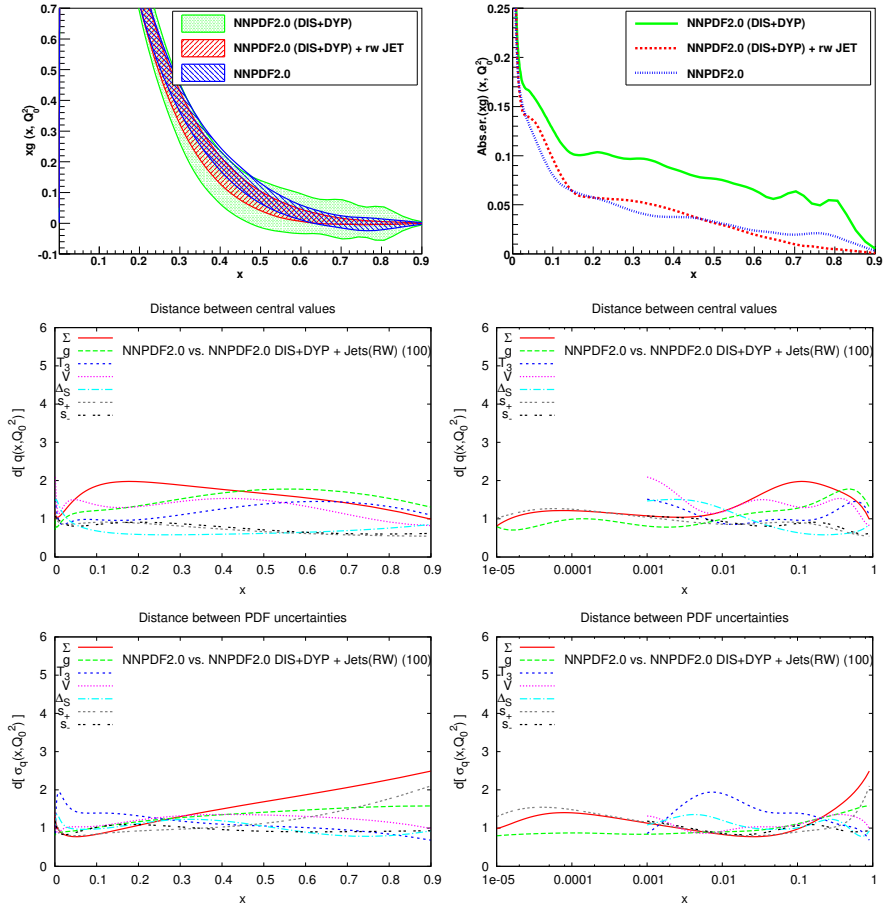
where the **weights** are

$$w_k = \frac{[\chi^2(y,f_k)]^{\frac{n_{dat}-1}{2}} e^{-\frac{\chi^2(y,f_k)}{2}}}{\sum_{i=1}^{N_{\rm rep}} [\chi^2(y,f_i)]^{\frac{n_{dat}-1}{2}} e^{-\frac{\chi^2(y,f_i)}{2}}}$$

Reweighting NNPDFs

Proof-of-concept: Inclusive Jet data, reweighting vs. refitting

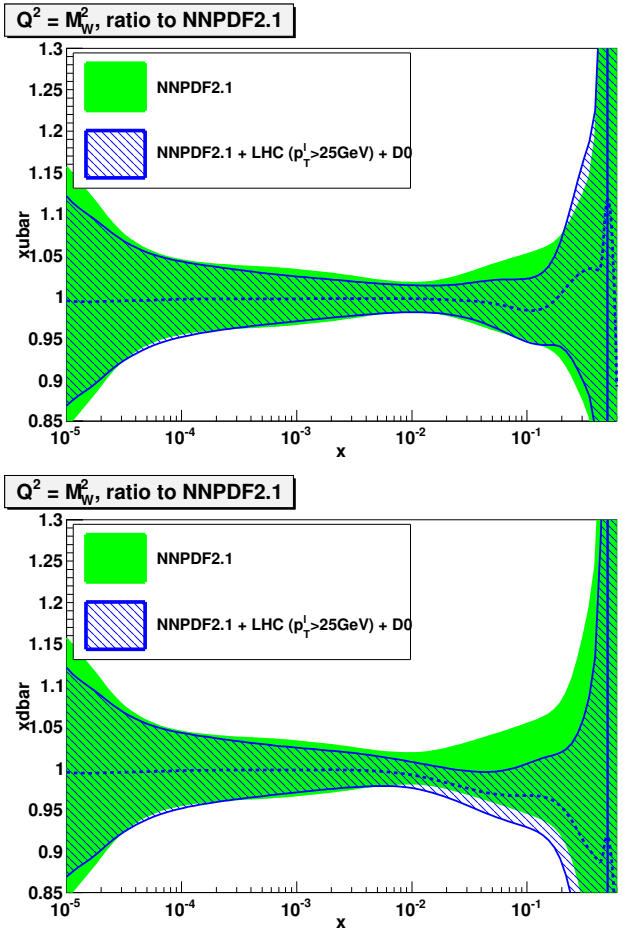
- Use **DIS+DY-fit** as **prior** probability distribution
- Add Tevatron Inclusive Jet data through refitting and through reweighting
- **Reweighting** and **refitting** yield **statistically equivalent** results



Reweighting NNPDFs

Real data: Tevatron/LHC W lepton asymmetry

- **D0 and first ATLAS and CMS** (36 pb^{-1}) W lepton asymmetry data have the potential to constrain PDFs.
- **Included** on top of NNPDF2.1 NLO **using reweighting techniques**.
- **No** need of **refitting**.
- Main impact on medium-/large-x quark distributions.
- Looking forward to 1 fb^{-1} data!



Conclusions

- A **reliable** estimation of **PDF uncertainties** is crucial in order to exploit the full physics potential of the LHC experiments.
- The **NNPDF2.1** family fulfills the requirement of an ideal PDF set for precision phenomenology at the LHC
  - it is based on a comprehensive, up-to-date, dataset,
  - it is free of parametrization bias,
  - it is provided with a reliable, statistically meaningful estimation of uncertainties,
  - it includes NLO corrections without resorting to K-factor approximations (local K-factors are used for NNLO corrections to hadronic observables),
  - it includes a consistent treatment of heavy quark effects,
  - it is available for a variety of values of α_s and quark masses.
- **Reweighting** techniques provide an easy way to **assess impact of/incorporate new data** in PDF fits without need for refitting.
- NNPDF sets are **available** within the **LHAPDF** interface.

BACKUP SLIDES

PDF Uncertainties and Correlations

A practitioner's guide to NNPDF predictions

Central Value

$$\langle \mathcal{F} \rangle = \frac{1}{N_{\rm set}} \sum_{k=1}^{N_{\rm set}} \mathcal{F}[q^{(k)}]$$

Standard Deviation

$$\sigma_{\mathcal{F}} = \left(\frac{1}{N_{\rm N_{set}}} \sum_{k=1}^{N_{\rm set}} \left(\mathcal{F}[\{q^{(k)}\}] - \langle \mathcal{F}[\{q\}] \rangle \right)^2 \right)^{1/2}$$

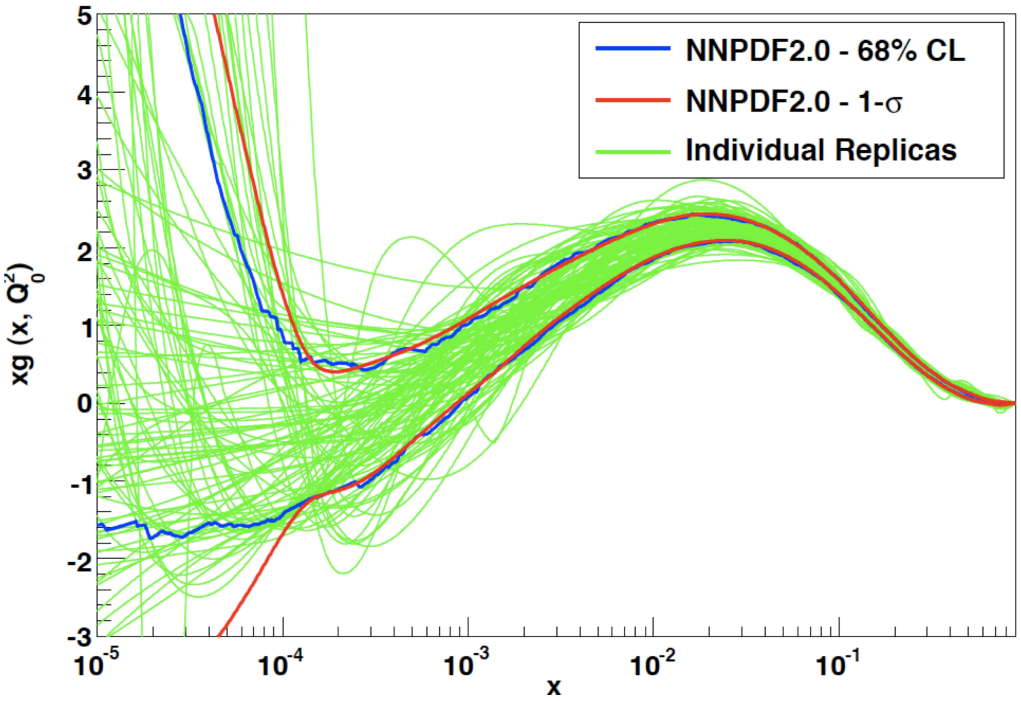
Correlation

$$\rho \equiv \cos\varphi(\mathcal{F},\mathcal{G}) = \frac{< \mathcal{F}\,\mathcal{G} >_{\rm rep} - < \mathcal{F} >_{\rm rep} < \mathcal{G} >_{\rm rep}}{\sqrt{< \mathcal{F}^2 >_{\rm rep} - < \mathcal{F} >^2_{\rm rep}}\sqrt{< \mathcal{G}^2 >_{\rm rep} - < \mathcal{G} >^2_{\rm rep}}}$$

Confidence Level Intervals

Testing for non gaussian distribution of fitted PDFs

- **Confidence Level intervals** can be computed directly from the replicas distribution
- Comparison of 68% C.L. and symmetric 1σ especially in extrapolation regions where theory constraints dominate on experimental information



- A **generic DIS observable** in the FONLL scheme is written as:

$$F^{FONLL}(x,Q^2) = \mathcal{D}(Q^2)F^{(d)}(x,Q^2) + F^{(n_l)}(x,Q^2)$$

where the **threshold damping factor** is given by

$$\mathcal{D}(Q^2) = \theta(Q^2 - m^2)\left(1 - \frac{m^2}{Q^2}\right)$$

and the **subtraction term** is

$$F^{(d)} = \left[F^{(n_l+1)}(x,Q^2) - F^{(n_l,0)}(x,Q^2)\right]$$

with the massless limit of the massive contributions being

$$F^{n_l,0}(x,Q^2) = x\int_x^1 \frac{dy}{y}\sum_{i=q,\bar{q},g} B_i^{(0)}\left(\frac{x}{y},\frac{Q^2}{m^2},\alpha_S^{(n_l+1)}(Q^2)\right) f_i^{(n_l+1)}(y,Q^2)$$

with

$$\lim_{m\to 0}\left[B_i\left(x,\frac{Q^2}{m^2}\right) - B_i^{(0)}\left(x,\frac{Q^2}{m^2}\right)\right] = 0$$