nnpdfrw:
an NNPDF reweighting tool

Alberto Guffanti
Niels Bohr International Academy & Discovery Center
Niels Bohr Institute - Copenhagen

HERAfitter Meeting
CPPM (Marseille)
February 13-14, 2012

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.

The outcome is a set fo $N_{\rm rep}$ **equiprobable** replicas

NNPDF Methodology

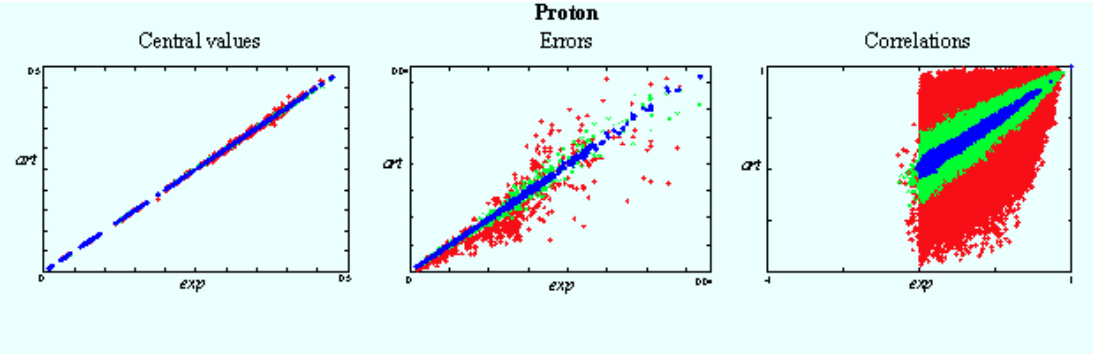
Monte Carlo replicas generation

- **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 (gaussianly distributed) 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

Reweighting (NN)PDFs

Assessing the impact of new data on PDF fits

[R. D. Ball et al., arXiv:1012.0836]
[R. D. Ball et al., arXiv:1108.1758]

- The $N_{\rm rep}$ **replicas** of a NNPDF fit give the **probability density** in the space of PDFs
- **Expectation values** for observables computed as

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

(... 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 without refitting.

Reweighting (NN)PDFs

Assessing the impact of new data on PDF fits

[R. D. Ball et al., arXiv:1012.0836]
[R. D. Ball et al., arXiv:1108.1758]

- Using **Bayes Theorem** we obtain

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

- **Averages over the sample** 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 (NN)PDFs

Assessing the impact of new data on PDF fits

[R. D. Ball et al., arXiv:1012.0836]
[R. D. Ball et al., arXiv:1108.1758]

- The quantity of information included in the reweighted set is quantified by the Shannon Entropy

$$N_{\rm eff} = \exp\left\{\frac{1}{N}\sum_{n=1}^{N} w_k \ln\left(\frac{N}{w_k}\right)\right\}$$

  i.e. the information contained in the reweighted set correspond to the one of a standard fit with $N_{\rm eff}$ replicas

- $N_{\rm eff} << N_{\rm rep}$ indicates that the new data are either very constraining or incompatible with the ones included in the global fit

Reweighting (NN)PDFs

Assessing the impact of new data on PDF fits

$$\ln(N_{\rm eff}) = \frac{1}{N}\sum_{k=1}^{N} w_k \ln\left(\frac{N}{w_k}\right)$$

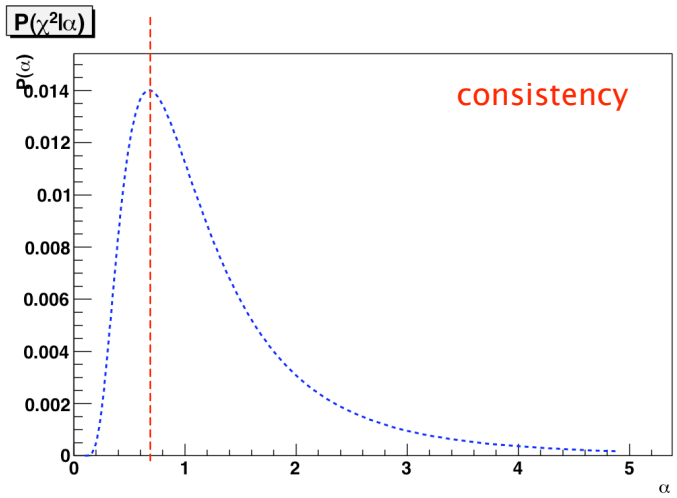
If N_{eff} small, two possibilities:

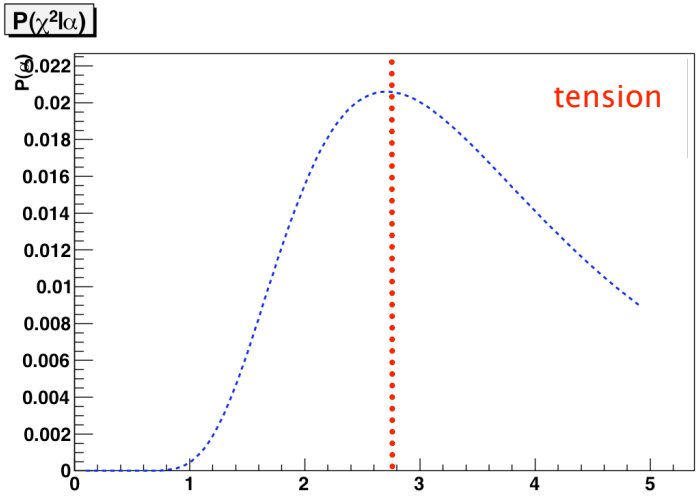
a) new data are very constraining
b) new data are inconsistent

How to distinguish?

$$\begin{aligned} \chi^2_\alpha &= \chi^2/\alpha \\ w_k(\alpha) &= \left(\chi^2_\alpha\right)^{n/2-1} \mathrm{e}^{-\chi^2_\alpha/2} \\ P(\alpha) &= \frac{\mathcal{N}}{\alpha}\sum_{k=1}^{N} w_k(\alpha) \end{aligned}$$

* Rescale covariance matrix by a factor α
* Compute probability for the rescaling parameter α
* w_α proportional to the probability of f_k given the new data with rescaled errors.

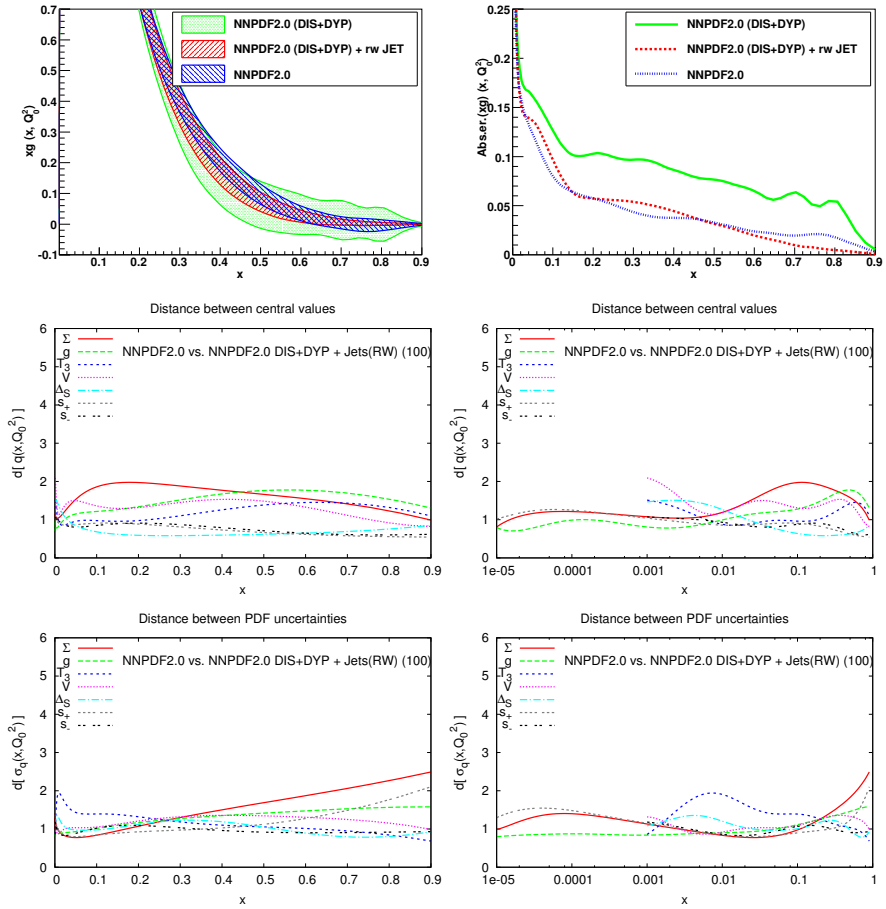




Reweighting (NN)PDFs

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

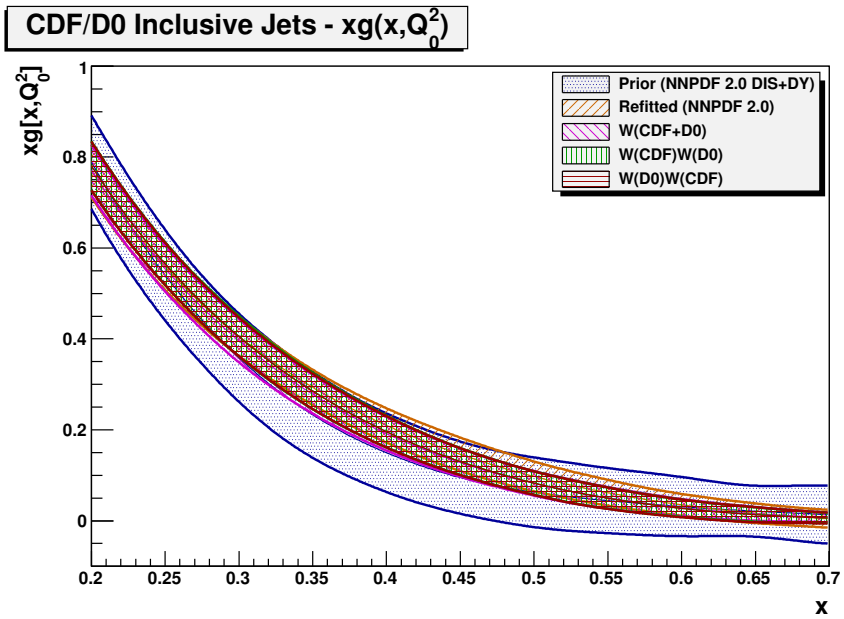
- Use **DIS+DY-fit** as **prior** probability distribution
- Add Tevatron Inclusive Jet data through refitting and through reweighting
- **Reweighting** and **refitting** yield **statistically equivalent** results

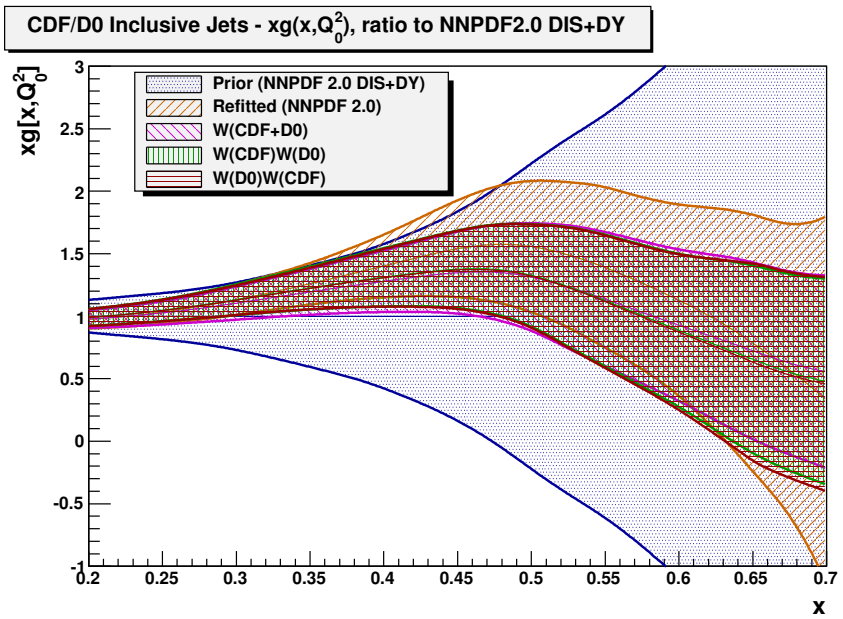


Reweighting NNPDFs

Testing the reweighting procedure - Adding multiple datasets

- What happens if we add more than one (independent) set by reweighting?
  - Is successive reweighting equivalent to reweighting with combined dataset?
  - Do successive reweightings commute?



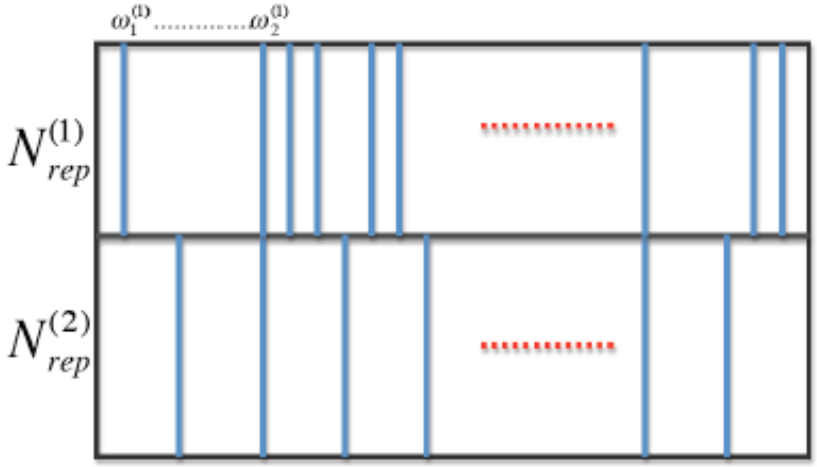


Reweighting NNPDFs

Unweighting reweighted PDFs

From reweighted ensemble to unweighted one for usual LHAPDF interface

- From reweighted set construct unweighted "normal" set
- Reweighted ensemble and uniform ensemble obtained by unweighting are statistically equivalent if N_{eff} is large enough
- In this way NNPDF2.2 was obtained, the first PDF fit including LHC data



$$W_k \equiv W_{k-1} + \frac{w_k}{N_{\rm rep}^{(1)}} = \sum_{j=0}^{k} \frac{w_j}{N_{\rm rep}^{(1)}}$$

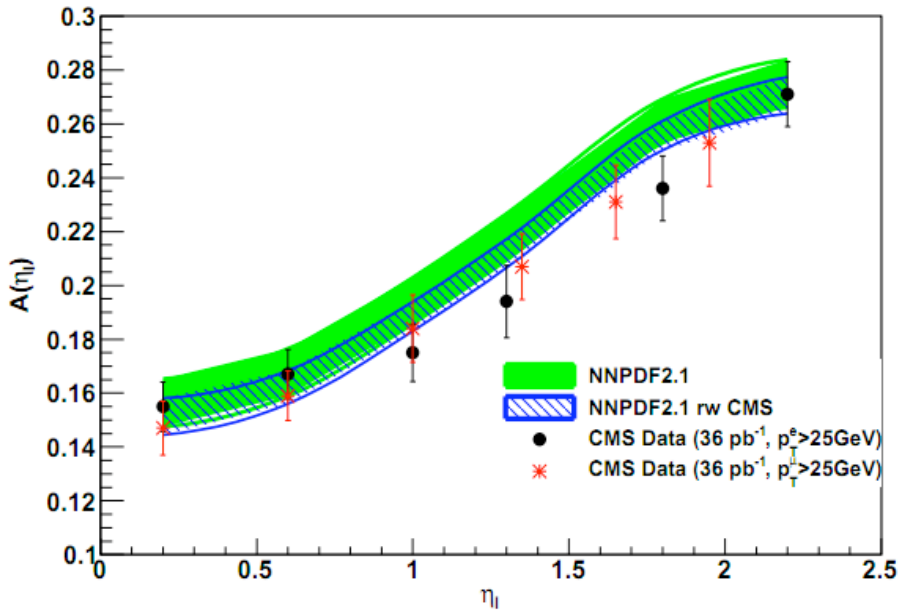
$$\tilde{w}_k = \sum_{j=1}^{N_{\rm rep}^{(2)}} \Theta\left(\frac{j}{N_{\rm rep}^{(2)}} - W_{k-1}\right) \Theta\left(W_k - \frac{j}{N_{\rm rep}^{(2)}}\right)$$

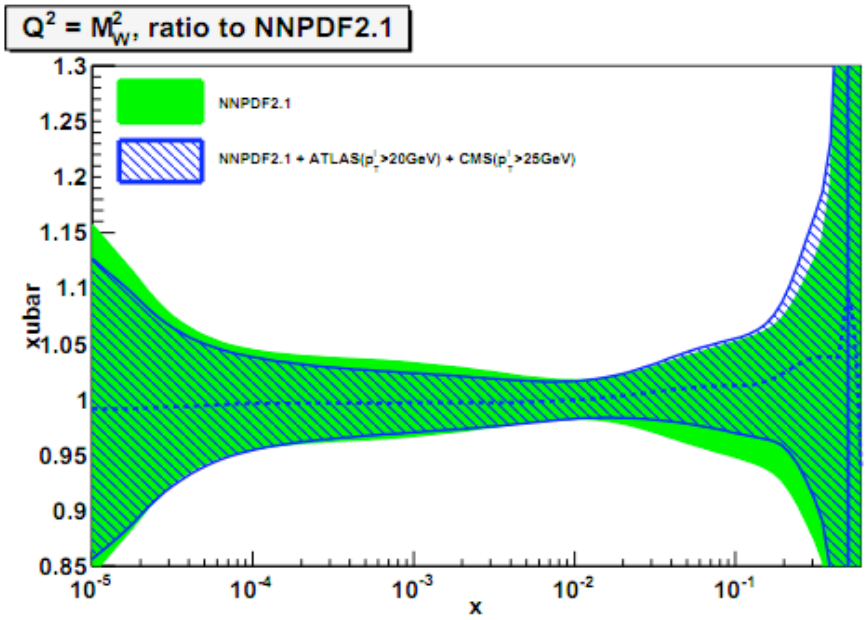
$\sum_k \tilde{w}_k = N_{\rm rep}^{(2)}$ number of effective replicas

- Each w_k can be either 0,1,..., $N^{(2)}_{rep}$
- Take w_1 copies of rep 1, w_2 copies of rep 2...
- End up with uniform sampling of $N^{(2)}_{rep}$

Reweighting NNPDFs

Reweighting at work - The NNPDF2.2 parton set

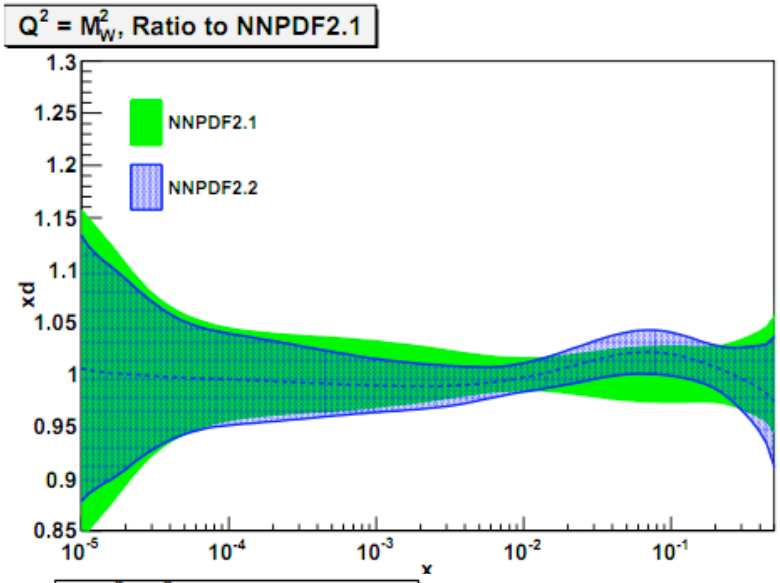


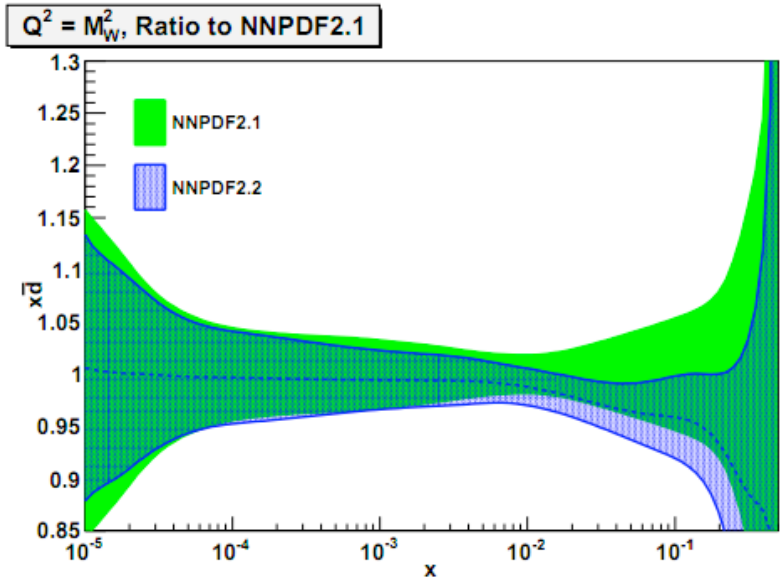


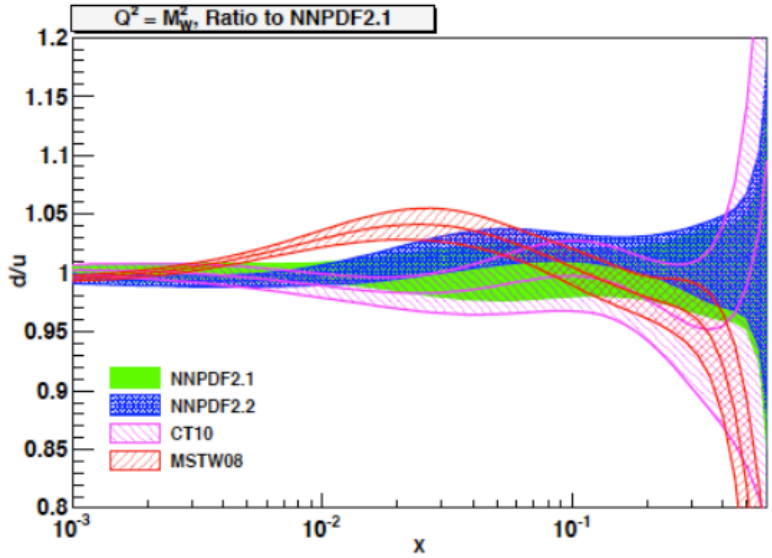
- Data compatible with data included in global PDF analysis
- After reweighting prediction gets closer to experimental data and uncertainty is reduced
- Non-negligible reduction of uncertainty at medium-small x for light (anti)quark

Reweighting NNPDFs

Reweighting at work - The NNPDF2.2 parton set





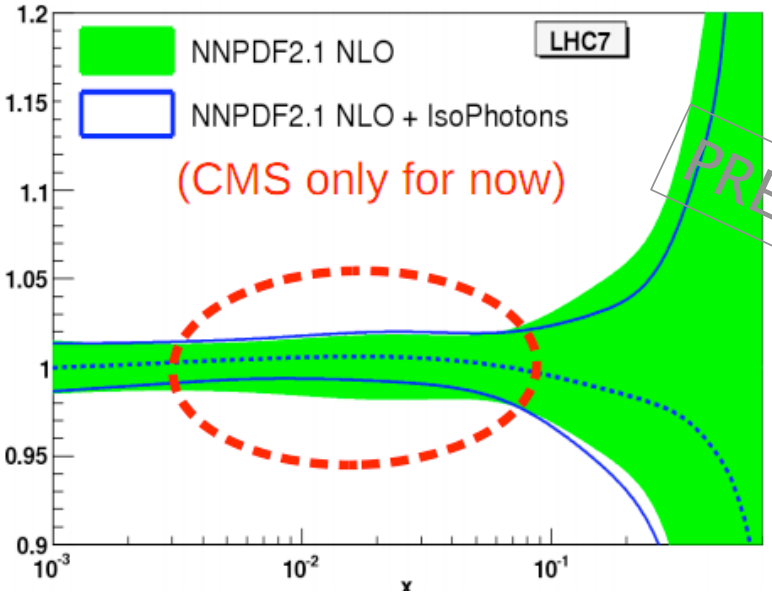


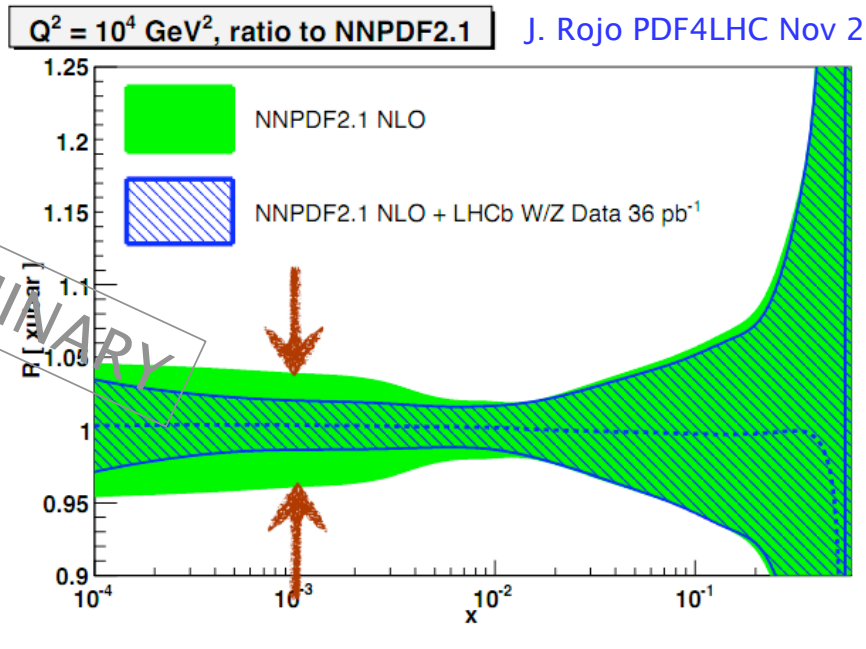
- NNPDF2.2: Added ATLAS and CMS W lepton asymmetry data and Tevatron D0 electron and muon asymmetry data at the same time.
- Their inclusion reduces uncertainty and moves central values
- Total uncertainty reduction is significant!

Reweighting NNPDFs

Reweighting at work - Other examples

D'Enterria, PDF4LHC Nov 2011





J. Rojo PDF4LHC Nov 2011

PRELIMINARY

- W lepton asymmetry data from ATLAS and CMS ➔ medium, small-x region light quarks/antiquarks
- Direct photon ➔ medium-x gluon
- LHCb high rapidity data (still preliminary) ➔ small-x region

LHC data already have a non-negligible impact!!

Reweighting NNPDFs

`nnpdfrw`: an NNPDF reweighting tool

[L. Del Debbio, N. P. Hartland and AG]

- Publicly (almost) available C++ code
- **Input**
  - experimental data and covariance matrix (format to be decided)
  - predictions for the observables for the $N_{\rm rep}$ replicas in the prior set
- **Output**
  - Estimators of reweighting quality (Shannon Entropy, χ^2 and $\mathcal{P}(\alpha)$ plots)
  - LHAPDF .LHgrid file for **unweighted** set including the new data

Conclusions & Outlook

- Reweighting provides a fast and effective alternative to study the impact of new data on PDF fits without need for refitting
- Can be used both for Monte Carlo and Hessian (see R. Mc Nulty's talks at PDF4LHC - Aug. 2011) PDF sets
- `nnpdfrw` is a minimal tool for performing reweighting anlyses with NNPDF that will be available soon (and Kristin promised to implement it in the HERAfitter framework)