NNPDF: Neural Networks, Monte Carlo techniques and Parton Distribution Functions

Alberto Guffanti

Albert-Ludwigs-Universität Freiburg

on behalf of the **NNPDF Collaboration**:
R. D. Ball, L. Del Debbio (Edinburgh), S. Forte, J. Rojo(Milano),
J. I. Latorre, F. Cerutti (Barcelona), M. Ubiali (Aachen), V. Bertone and AG (Freiburg)

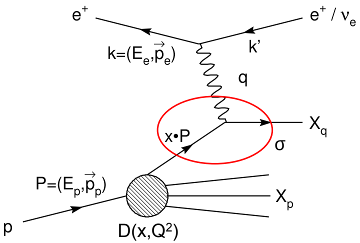
HiX2010

3rd International Workshop on Nucleon Structure at Large Bjorken *x*

Jefferson Lab, Newport News, Virginia
October 13 - 15, 2010

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 the **scale dependence** is governed by **DGLAP** evolution equations

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

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

- ... and the **splitting functions** P 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)

Problem

Faithful estimation of errors on PDFs

- Single quantity: 1-σ error
- Multiple quantities: 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 quantity: 1-σ error
- Multiple quantities: 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

$$q(x, Q_0^2) = x^\alpha (1-x)^\beta P(x; \lambda_1, ..., \lambda_n).$$

- Fit parameters minimizing χ^2.

Solution

Standard Approach

- Introduce a simple functional form with enough free parameters

$$q(x, Q_0^2) = x^\alpha (1-x)^\beta P(x; \lambda_1, ..., \lambda_n).$$

- Fit parameters minimizing χ^2.

Open problems:

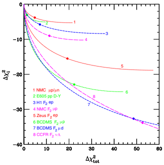
- **Error propagation** from data to parameters and from parameters to observables is **not trivial**.
- **Theoretical bias** due to the chosen **parametrization** is difficult to assess.

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]

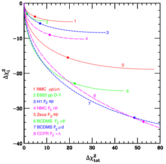


Shortcomings of the Standard approach

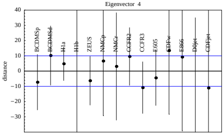
What is the meaning of a one-σ uncertainty?

- 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).

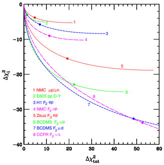


Shortcomings of the Standard approach

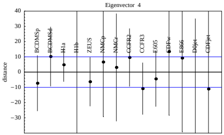
What is the meaning of a one-σ uncertainty?

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

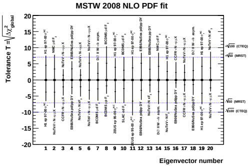
  [Collins & Pumplin, 2001]



- 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).

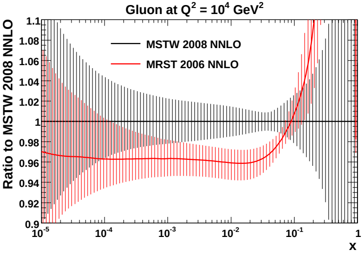


Shortcomings of the Standard approach

What determines PDF uncertainties?

- Uncertainties in standard fits often increase when adding new data to the fit.
- 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.



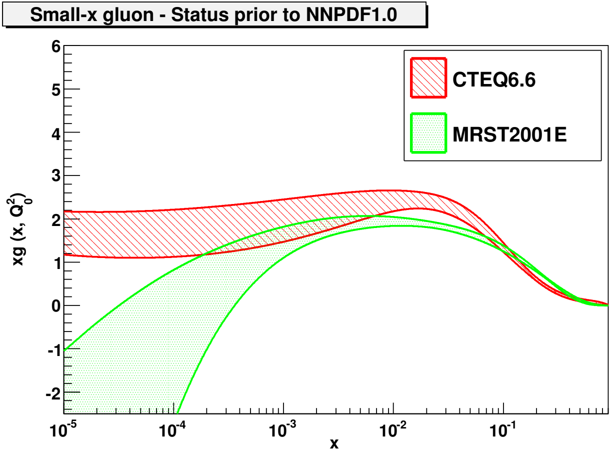
Larger small-x uncertainty due to extrat free parameter.

[R. Thorne, PDF4LHC]

PDF4LHCMSTW 24

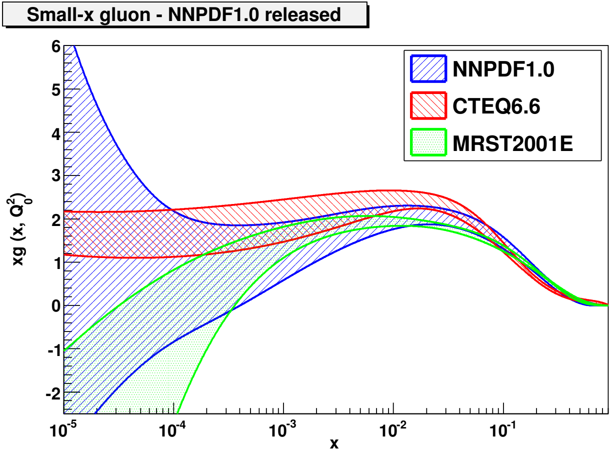
Shortcomings of the standard approace

What determines PDF uncertainties? Parmetrization bias?



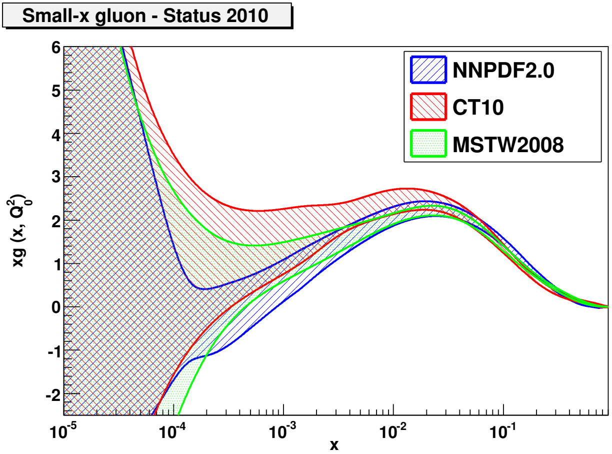
Shortcomings of the standard approace

What determines PDF uncertainties? Parmetrization bias?



Shortcomings of the standard approace

What determines PDF uncertainties? Parmetrization bias?



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

... 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 Methodology

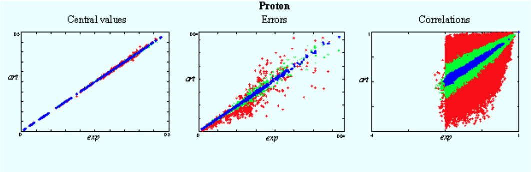
Monte Carlo replicas generation

- 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

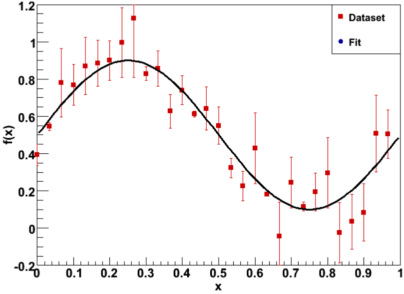
- Validate Monte Carlo replicas against experimental data (statistical estimators, faithful representation of errors, convergence rate increasing N_{rep})



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

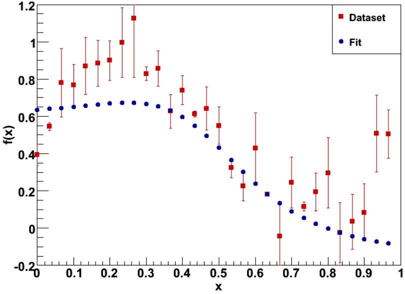
Proper Fitting avoiding Overlearning

- Let's see how proper fitting works in a toy model



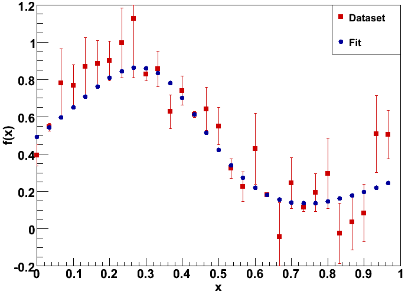
Proper Fitting avoiding Overlearning

- Let's see how proper fitting works in a toy model



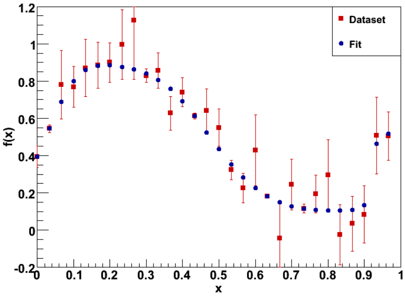
Proper Fitting avoiding Overlearning

- Let's see how proper fitting works in a toy model



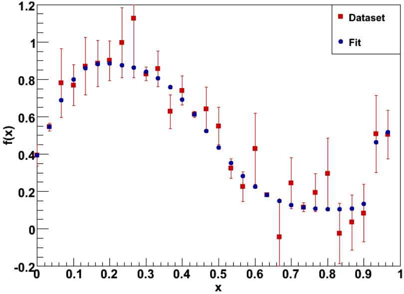
Proper Fitting avoiding Overlearning

- Let's see how proper fitting works in a toy model



Proper Fitting avoiding Overlearning

- Let's see how proper fitting works in a toy model



- Need a **redundant parametrization** to avoid parametrization bias.
- Need a way of **stopping the fit before overlearning** sets in to avoid fitting statistical noise.

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

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

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:

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

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

Neural Networks

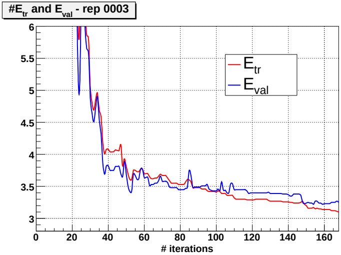
Stopping criterion

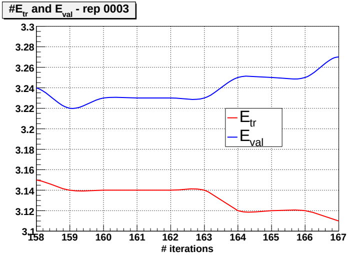
Stopping criterion based on Training-Validation separation

- 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

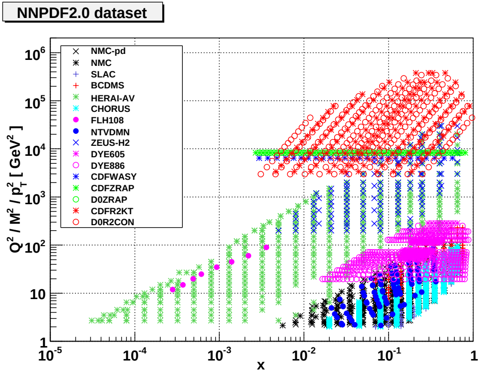
- Divide the data in two sets: Training and Validation
- Minimize the χ^2 of the data in the Training set
- Compute the χ^2 for the data in the Validation set
- When Validation χ^2 stops decreasing, **STOP** the fit





NNPDF 2.0

Dataset



3415 data points

(for comparison MSTW08 includes 2699 data points)

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 comb. ZEUS (HERA-II)
$\sigma_{CC}^{\pm}$	HERA-I comb. ZEUS (HERA-II)
F_L	H1
$\sigma_\nu, \sigma_{\bar\nu}$	CHORUS
dimuon prod.	NuTeV
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 2.0

Technical improvements

- The fit is carried at **NLO** QCD, in the **Zero-Mass Variable Flavour Number Scheme**
- Fast DGLAP evolution based on higher-order interpolating polynomials
- Improved treatment of normalization errors (t_0 method)
  - For details see [R. D. Ball et al., arXiv:0912.2276]
- Improvements in training/stopping
  - Target Weighted Training
  - Improved stopping for avoiding under-/over-learning
- All details given in [R. D. Ball et al., arXiv:1002.4407]

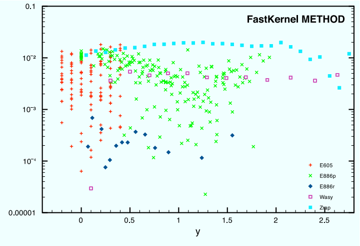
NNPDF 2.0

FastKernel

- NLO computation of hadronic observables too slow for parton global fits.
- MSTW08 and CTEQ include Drell-Yan NLO as (local) K factors rescaling the LO cross section
- K-factor depends on PDFs and it is not always a good approximation.

* NNPDF2.0 includes full NLO calculation of hadronic observables.
* Use available fastNLO interface for jet inclusive cross-sections.[hep-ph/0609285]
* Built up our own **FastKernel** computation of DY observables.

$$\int_{x_{0,1}}^{1} dx_1 \int_{x_{0,2}}^{1} dx_2\, f_a(x_1) f_b(x_2) C^{ab}(x_1, x_2) \rightarrow \sum_{\alpha,\beta=1}^{N_x} f_a(x_{1,\alpha}) f_b(x_{2,\beta}) \int_{x_{0,1}}^{1} dx_1 \int_{x_{0,2}}^{1} dx_2\, \mathcal{I}^{(\alpha,\beta)}(x_1, x_2) C^{ab}(x_1, x_2)$$



- DGLAP evol. and double convolution improved
  - Use high-orders polynomial interpolation
  - Precompute all Green Functions

A truly NLO analysis

NNPDF 2.0

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)

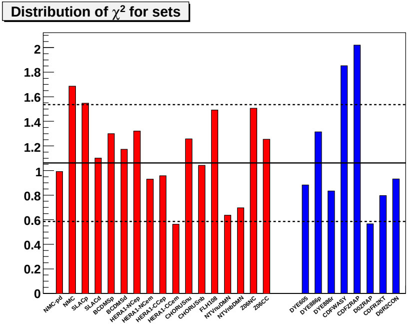
259 parameters
Standard fits have $\sim$ 25 parameters in total

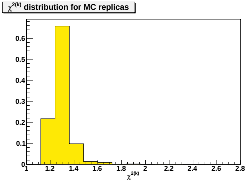
No change in the parametrization from NNPDF1.2 ... despite substantial **enlargement of the dataset**

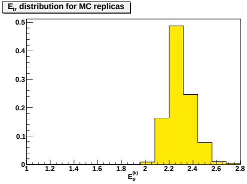
NNPDF 2.0

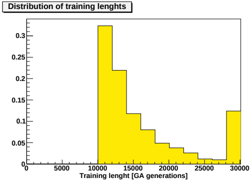
General features of the fit

$\chi^2_{\rm tot}$	1.21
$\langle E \rangle \pm \sigma_E$	2.32 ± 0.10
$\langle E_{\rm tr} \rangle \pm \sigma_{E_{\rm tr}}$	2.29 ± 0.11
$\langle E_{\rm val} \rangle \pm \sigma_{E_{\rm val}}$	2.35 ± 0.12
$\langle {\rm TL} \rangle \pm \sigma_{\rm TL}$	16175 ± 6257
$\langle \chi^{2(k)} \rangle \pm \sigma_{\chi^2}$	1.29 ± 0.09



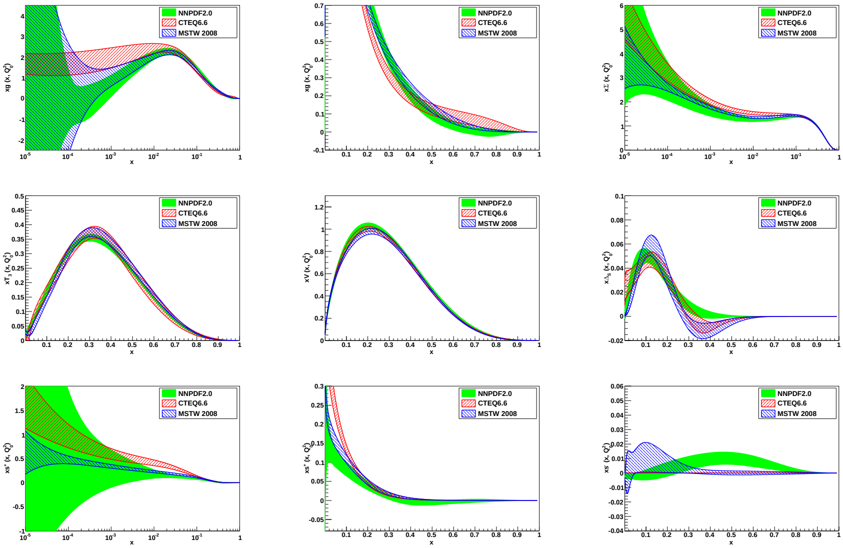






NNPDF 2.0

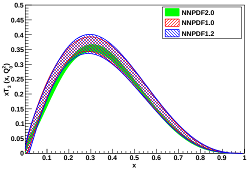
Partons - Comparison to other global fits



NNPDF2.0

Results - Partons - A couple of upshots

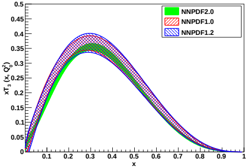
- **Reduction of uncertainties** with respect to older NNPDF sets due to **inclusion of new data**



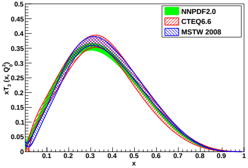
NNPDF2.0

Results - Partons - A couple of upshots

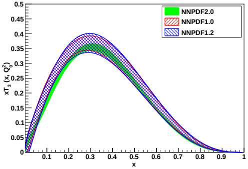
- **Reduction of uncertainties** with respect to older NNPDF sets due to **inclusion of new data**



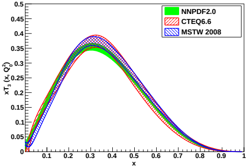
- **Uncertainties** on PDFs **competitive** with results from other groups ...



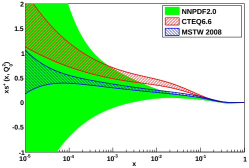
- **Reduction of uncertainties** with respect to older NNPDF sets due to **inclusion of new data**



- **Uncertainties** on PDFs **competitive** with results from other groups ...



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



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

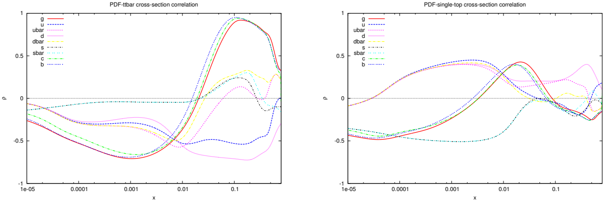
$$\rho \equiv \cos[\varphi(\mathcal{F},\mathcal{G})] = \frac{< \mathcal{F}\,\mathcal{G} >_{\rm rep} - < \mathcal{F} >_{\rm rep} < \mathcal{G} >_{\rm rep}}{\sqrt{< \mathcal{F}^2 >_{\rm rep} - < \mathcal{F} >^2_{\rm rep}}\sqrt{< \mathcal{G}^2 >_{\rm rep} - < \mathcal{G} >^2_{\rm rep}}}$$

PDF induced correlations

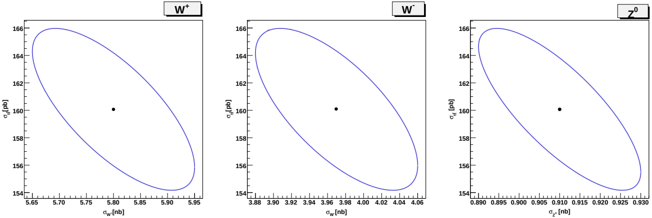
Ex.: Top-quark studies within the NNPDF framework

[J. Rojo and AG, arXiv:1008.4671]

- It is easy to compute **correlations** among **PDFs and observables**



- ... or **pairs of observbles**



Reweighting PDFs

Assessing the impact of new data on PDF fits

- Inspired by Giele and Keller [hep-ph/9803393]
- The N_{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 PDFs

Assessing the impact of new data on PDF fits

- According to **Bayes Theorem** we have

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

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

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

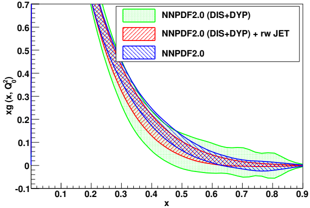
where the **weights** are

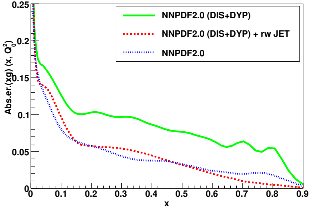
$$w_k = \frac{[\chi^2(y,f_k)]^{\frac{n_{dat}}{2}-1} e^{-\frac{\chi^2(y,f_k)}{2}}}{\sum_{i=1}^{N_{\rm rep}} [\chi^2(y,f_i)]^{\frac{n_{dat}}{2}-1} e^{-\frac{\chi^2(y,f_i)}{2}}}$$

Reweighting 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

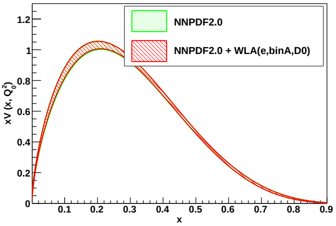


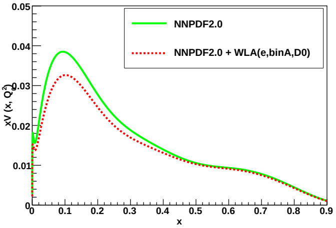


Reweighting PDFs

Reweighting real data: W lepton asymmetry

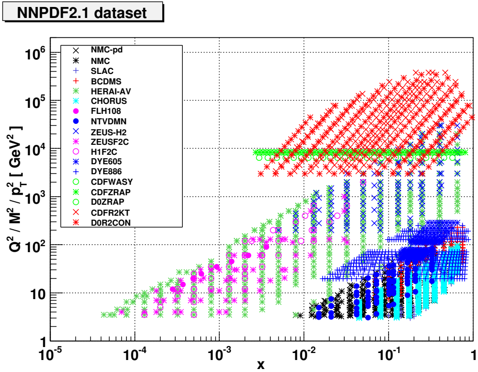
- In the NNPDF2.0 fit we only included CDF W asymmetry data
- We evaluated W electron asymmetry with NNPDF20 1000 replicas set using **DYNNLO** [Catani et al., arXiv:0903.2120].
- .. and included D0 W electron asymmetry data points through reweighting.
- Main impact on reduction of middle-x Valence uncertainty.
- **No** need of **refitting**





NNPDF 2.1

Dataset



3554 data points

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
$\sigma_\nu, \sigma_{\bar{\nu}}$	CHORUS
dimuon prod.	NuTeV
F_2^c	**ZEUS (99,03,08,09)**
F_2^c	**H1 (01,09,10)**

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

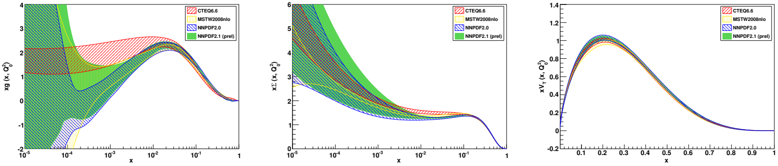
Inclusive jet prod.	
Incl. $\sigma^{\rm (jet)}$	CDF (k_T) - Run II
Incl. $\sigma^{\rm (jet)}$	D0 (cone) - Run II

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

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

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

- **Preliminary** results for partons



- Small- and medium-x **gluon** and **Singlet** are most sensitive, **Non-Singlet** combinations **mostly unaffected**
- Effect of improved heavy flavour treatment is within **one-σ**, both on PDFs and LHC Standard Candles
- **Fixed-**(3- and 4-)**Flavour Number Scheme** PDFs will also be released

Conclusions

- The **NNPDF methodology** based on using **Monte Carlo** techniques and **Neural Networks** is well suited to address problems of standard fits.
- **NNPDF2.0** is the first **global NNPDF fit**
  - Exact inclusion of NLO corrections
  - **No** sign of **strong tension** among different datasets
- NNPDF sets are **available** within the **LHAPDF** interface.
- **Reweighting** technique allows to study impact of **new data without refitting**
- Next steps:
  - Improved treatment of Heavy Flavour contributions (FONLL)
  - Inclusion of higher order contributions (NNLO QCD/EW effects)
  - Study the impact of theoretical uncertainties (α_S, quark masses ...)
  - Inclusion of resummation (small-/large-x) effects