First set of Neural Partons for the LHC

Maria Ubiali

School of Physics, University of Edinburgh

University of Oxford
7 November 2008

"A determination of parton distributions with faithful uncertainty estimation,"
arXiv:0808.1231 [hep-ph], Nucl. Phys. B (in press)

The NNPDF Collaboration
R.D.Ball[1], L.Del Debbio[1], S.Forte[2], A.Guffanti[3], J.I.Latorre[4], A. Piccione[2], J. Rojo[2], M.U.[1]

[1] PPT Group, School of Physics, University of Edinburgh
[2] Dipartimento di Fisica, Università di Milano
[3] Physikalisches Institut, Albert-Ludwigs-Universität Freiburg
[4] Departament d'Estructura i Constituents de la Matèria, Universitat de Barcelona

Outline

1. Introduction
2. NNPDF approach: the main ingredients
   - Monte Carlo Determination of Errors
   - Neural Network as unbiased and redundant parametrization
   - Dynamical Stopping Criterion
3. Results
   - The NNPDF1.0 partons
   - Independence on parametrization
   - Dependence on data sets
   - Phenomenology
4. Conclusions and outlook

Outline

1. Introduction
2. NNPDF approach: the main ingredients
   - Monte Carlo Determination of Errors
   - Neural Network as unbiased and redundant parametrization
   - Dynamical Stopping Criterion
3. Results
   - The NNPDF1.0 partons
   - Independence on parametrization
   - Dependence on data sets
   - Phenomenology
4. Conclusions and outlook

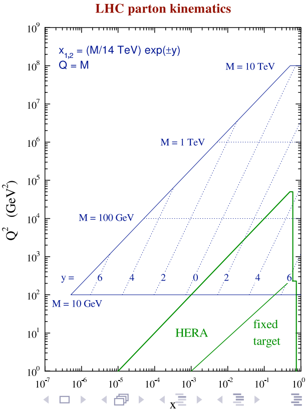
Parton Distribution Functions

- Factorization Theorem ($Q^2 \gg \Lambda^2_{\rm QCD}$):

$$\frac{d\sigma_H}{dX} = \sum_{a,b} \int dx_1 dx_2 f_a(x_1,\mu_f) f_b(x_2,\mu_f) \otimes \frac{d\hat{\sigma}}{dX}(\alpha_s(\mu_r),\mu_r,\mu_f,x_1,x_2,Q^2)$$

- DGLAP equations:

$$\frac{d}{dt}\begin{pmatrix} q \\ g \end{pmatrix} = \frac{\alpha_s}{2\pi}\begin{pmatrix} P_{qq} & P_{qg} \\ P_{gq} & P_{gg} \end{pmatrix} \otimes \begin{pmatrix} q \\ g \end{pmatrix} + O(\alpha_s^2)$$



The accurate computation of physical observables at the LHC requires good knowledge of PDFs and of their error.

The name of the game

- Given a set of data points we must determine a set of functions with error.
- We need an error band in the space of functions, i.e. a probability density $\mathcal{P}[f(x)]$ in the space of PDFs, f(x). For an observable $\mathcal{F}$ depending on PDFs :

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

Standard approach

- Choose a basis of functions and project PDFs on it.
- Determine best-fit values of parameters.
- Determine error by propagation of error in the space of parameters (Hessian method).

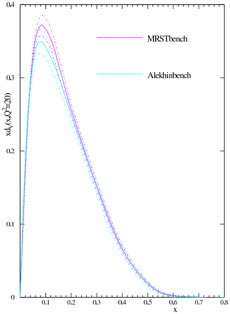
Issues

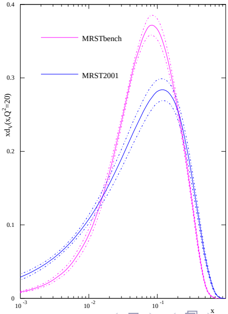
- Non trivial propagation of errors: non-gaussian errors and incompatible data.
- The error associated to the choice of parametrization is difficult to assess.

What is the problem?

PDF4LHC, February '08

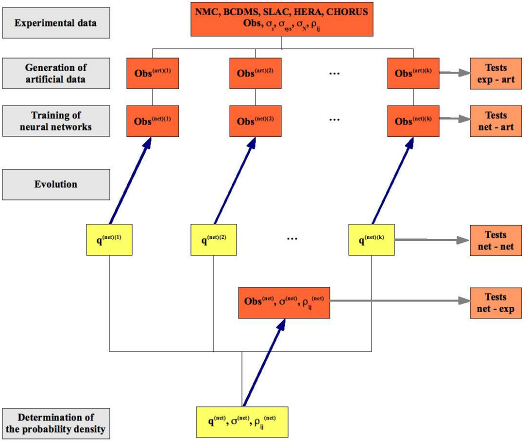
- Benchmark partons on reduced set of experiments do not agree with global fits within errors: incompatible experiments, not enough flexibility in the parametrization or both?
- Tolerance criterion $\Delta\chi^2 > 2.7$ means that error on experimental measurements is blown up by a factor $S = \sqrt{\Delta\chi^2/2.7}$ (B. Cousins).
- $S_{\rm CTEQ} \sim 6$; $S_{\rm MSTW} \sim 4.5$: is that factor mandatory?
- In Alekhin DIS+DY fit $\Delta\chi^2$=1





Outline

NNPDF approach



Monte Carlo errors

Non-gaussian errors and non trivial error propagation.

Neural Networks

Reduce bias from a restrictive fixed functional form.

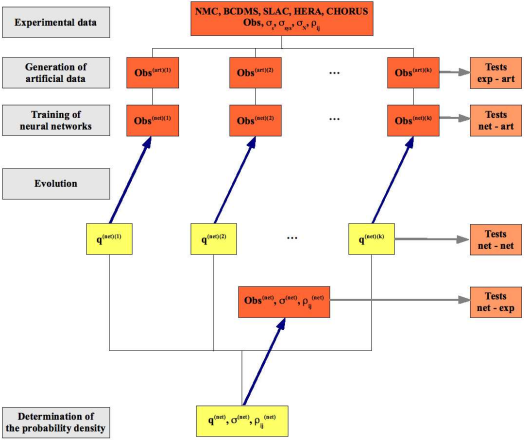
Dynamical Stopping

No looking for absolute minimum but learning from data.

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

$$\langle \mathcal{F}[f(x)] \rangle = \frac{1}{N_{\mathrm{rep}}} \sum_{k=1}^{N_{\mathrm{rep}}} \mathcal{F}[f^{(k)(\mathrm{net})}(x)]$$

Step 1: Monte Carlo Errors

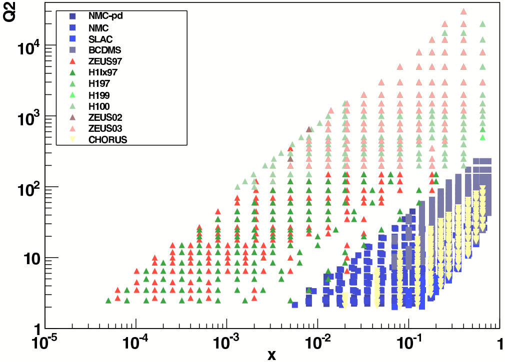


Monte Carlo errors

Non-gaussian errors and non trivial error propagation.

Experimental data

$$F_I^{(exp)}(x_p, Q_p^2) \qquad i = 1, \ldots, N_{\mathrm{data}}$$



OBS	Data set	OBS	Data set
F_2^p	NMC	σ^-_{NC}	ZEUS
	SLAC		H1
	BCDMS	σ^+_{CC}	ZEUS
F_2^d	SLAC		H1
	BCDMS	σ^-_{CC}	ZEUS
σ^+_{NC}	ZEUS		H1
	H1	$\sigma_\nu\, ,\, \sigma_{\bar{\nu}}$	CHORUS
$F_2^d\,/\,F_2^p$	NMC-pd	F_L	H1

- Kinematical cuts:
  $Q^2 > 2\ \mathrm{GeV}^2$
  $W^2 = Q^2(1-x)/x > 12.5\ \mathrm{GeV}^2$
- $\sim$ **3000** points.

Monte Carlo sample

Generate a $N_{\rm rep}$ Monte Carlo sets of artificial data, or "pseudo-data" of the original N_{data} data points

$$F_i^{(art)(k)}(x_p, Q_p^2) \equiv F_{i,p}^{(art)(k)} \qquad \begin{aligned} i &= 1, \ldots, N_{\rm data} \\ k &= 1, \ldots, N_{\rm rep} \end{aligned}$$

Multi-gaussian distribution centered on each data point:

$$F_{i,p}^{(art)(k)} = S_{p,N}^{(k)} F_{i,p}^{\rm exp} \left(1 + r_p^{(K)} \sigma_p^{\rm stat} + \sum_{j=1}^{N_{\rm sys}} r_{p,j}^{(k)} \sigma_{p,j}^{\rm sys} \right)$$

If two points have correlated systematic uncertainties

$$r_{p,j}^{(k)} = r_{p',j}^{(k)}$$

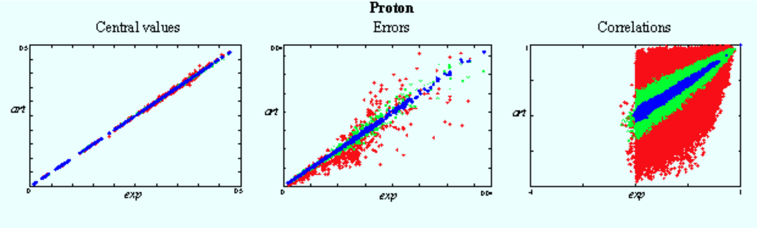
Correlations are properly taken into account.

Validation of the MC sample

Experiment	ZEUS	CHORUS	Total
$\langle PE\left[\langle F^{(\mathrm{art})}\rangle_{\mathrm{rep}}\right]\rangle_{\mathrm{dat}}$	$8.5\cdot 10^{-4}$	$1.8\cdot 10^{-3}$	$7.1\cdot 10^{-5}$
$r\left[F^{(\mathrm{art})}\right]$	1.000	1.000	0.980
$\langle PE\left[\langle \sigma^{(\mathrm{art})}\rangle_{\mathrm{rep}}\right]\rangle_{\mathrm{dat}}$	$9.6\cdot 10^{-3}$	$1.8\cdot 10^{-2}$	$3.0\cdot 10^{-3}$
$\langle \sigma^{(\mathrm{exp})}\rangle_{\mathrm{dat}}$	0.0607	0.1088	0.0556
$\langle \sigma^{(\mathrm{art})}\rangle_{\mathrm{dat}}$	0.0603	0.1109	0.0562
$r\left[\sigma^{(\mathrm{art})}\right]$	1.000	0.998	0.980
$\langle \rho^{(\mathrm{exp})}\rangle_{\mathrm{dat}}$	0.079	0.650	0.145
$\langle \rho^{(\mathrm{art})}\rangle_{\mathrm{dat}}$	0.082	0.657	0.146
$r\left[\rho^{(\mathrm{art})}\right]$	0.982	0.996	0.996
$\langle \mathrm{cov}^{(\mathrm{exp})}\rangle_{\mathrm{dat}}$	$1.53\cdot 10^{-4}$	$2.03\cdot 10^{-2}$	$1.07\cdot 10^{-3}$
$\langle \mathrm{cov}^{(\mathrm{art})}\rangle_{\mathrm{dat}}$	$1.57\cdot 10^{-4}$	$2.11\cdot 10^{-2}$	$1.01\cdot 10^{-3}$
$r\left[\mathrm{cov}^{(\mathrm{art})}\right]$	0.996	0.998	0.997

A MC sample with $\mathcal{O}(1000)$ replicas reproduces mean values, variances, correlations of experimental data within 1% accuracy.

Convergence rate increases with N_{rep}.



Monte Carlo Errors

For each replica $^{(k)}$ of the experimental data we fit a set of independent PDFs

Ensemble of fitted replicas of PDFs: representation of the probability distribution in the space of PDFs

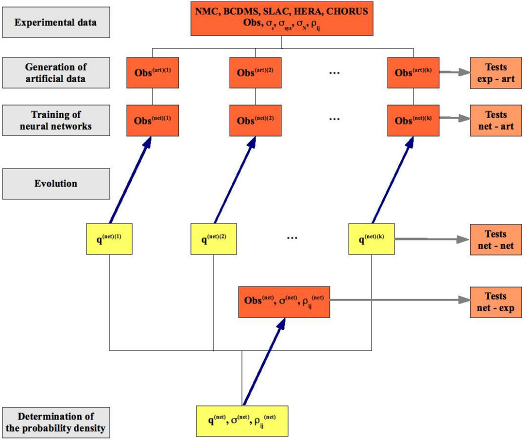
Uncertainties, central values and any other statistical property (e. g. correlations) of the PDFs (or any function of them) can be evaluated using standard statistical methods.

$$\langle \mathcal{F}[f(x)] \rangle = \frac{1}{N_{\rm rep}} \sum_{k=1}^{N_{\rm rep}} \mathcal{F}[f^{(k)({\rm net})}(x)]$$

$$\sigma_{\mathcal{F}[f(x)]} = \sqrt{\langle \mathcal{F}[f(x)]^2 \rangle - \langle \mathcal{F}[f(x)] \rangle^2}$$

$$\rho[f_a(x_1, Q_1^2), f_b(x_2, Q_2^2)] = \frac{\langle f_a(x_1, Q_1^2) f_b(x_2, Q_2^2) \rangle - \langle f_a(x_1, Q_1^2) \rangle \langle f_b(x_2, Q_2^2) \rangle}{\sigma_a(x_1, Q_1^2) \sigma_b(x_2, Q_2^2)}$$

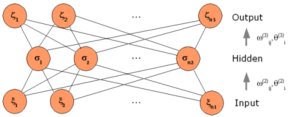
Step 2: Neural Network as unbiased and redundant parametrization



Neural Networks

Reduce bias from a restrictive fixed functional form.

What are neural networks?



* Each neuron receives input from neurons in preceding layer.
* Activation determined by weights and thresholds according to a non linear function:

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

In a simple case (1-2-1) we have,

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

7 parameters

...Just a convenient functional form which provides a redundant and flexible parametrization.

We want the best fit to be independent of any assumption made on the parametrization.

Basis set

- Each independent PDF at the initial scale $Q_0^2 = 2\mathrm{GeV}^2$ is parameterized by an individual NN.
- Little constraint on strange → Flavor Assumptions:
  - Symmetric strange sea $s(x) = \bar{s}(x)$
  - Strange sea proportional to non-strange sea $\bar{s}(x) = \frac{C}{2}(\bar{u}(x) + \bar{d}(x))$ (C = 0.5)
  - Intrinsic heavy quarks contributions neglected.
- Parametrization of (4+1) combinations of PDFs at $Q_0^2 = 2\ \mathrm{GeV}^2$:

Singlet : $\Sigma(x)$	$\longmapsto \mathrm{NN}_\Sigma(x)$	2-5-3-1 37 pars
Gluon : $g(x)$	$\longmapsto \mathrm{NN}_g(x)$	2-5-3-1 37 pars
Total valence : $V(x) \equiv u_V(x) + d_V(x)$	$\longmapsto \mathrm{NN}_V(x)$	2-5-3-1 37 pars
Non-singlet triplet : $T_3(x)$	$\longmapsto \mathrm{NN}_{T3}(x)$	2-5-3-1 37 pars
Sea asymmetry : $\Delta_S(x) \equiv \bar{d}(x) - \bar{u}(x)$	$\longmapsto \mathrm{NN}_\Delta(x)$	2-5-3-1 37 pars

185 parameters

Normalization and sum rules

$$
\begin{aligned}
\Sigma(x, Q_0^2) &= (1-x)^{m_\Sigma} x^{-n_\Sigma} \mathrm{NN}_\Sigma(x) , \\
V(x, Q_0^2) &= A_V (1-x)^{m_V} x^{-n_V} \mathrm{NN}_V(x) , \\
T_3(x, Q_0^2) &= (1-x)^{m_{T_3}} x^{-n_{T_3}} \mathrm{NN}_{T_3}(x) , \\
\Delta_S(x, Q_0^2) &= A_{\Delta_S} (1-x)^{m_{\Delta_S}} x^{-n_{\Delta_S}} \mathrm{NN}_{\Delta_S}(x) , \\
g(x, Q_0^2) &= A_g (1-x)^{m_g} x^{-n_g} \mathrm{NN}_g(x) .
\end{aligned}
$$

- Polynomial Preprocessing → Training Efficiency
  Need to verify the independence
- Normalization → Fixed by valence and momentum sum rules
  Theoretical constraint

$$
\begin{aligned}
\int_0^1 dx \; x \left(\Sigma(x) + g(x)\right) &= 1 \\
\int_0^1 dx \; \left(u(x) - \bar{u}(x)\right) &= 2 \\
\int_0^1 dx \; \left(d(x) - \bar{d}(x)\right) &= 1 .
\end{aligned}
$$

From the starting scale to the data: the evolution code

* To train NN we need to evolve from Q_0^2 to the experimental scales.

$$f_i(x, Q^2) = \sum_j \Gamma_{ij}(x, \alpha_s, \alpha_s^0) \otimes f_j(x, Q_0^2)$$

* Observables are a convolution over x of PDFs and Coefficient Functions.

$$F_I(x, Q^2) = \sum_j C_{Ij}(x, \alpha_s) \otimes f_j(x, Q^2) = \sum_{j,k} C_{Ij}(x, \alpha_s) \otimes \Gamma_{jk}(x, \alpha_s, \alpha_s^0) \otimes f_k(x, Q_0^2)$$

We want: Mellin space evolution

$$K_{Ik}(N, \alpha_s, \alpha_s^0) = \sum_j C_{Ij}(N, \alpha_s) \Gamma_{jk}(N, \alpha_s, \alpha_s^0)$$

We do not want: Complex Neural Networks

$$K_{Ik}(y, \alpha_s, \alpha_s^0) = \frac{1}{2\pi i} \int_{\mathcal{C}} dN\, y^{-N} K_{Ik}(N, \alpha_s, \alpha_s^0)$$

$$F_I(x, Q^2) = \sum_k \int_x^1 \frac{dy}{y} K_{Ik}(y, \alpha_s, \alpha_s^0)\, f_k(\frac{x}{y}, Q_0^2)$$

Kernels for a physical observable

F_2 proton structure function

$$F_2^p = x\{\frac{5}{18}C_{2,q}^s \otimes \Sigma + \frac{1}{6}C_{2,q} \otimes (T_3 + \frac{1}{3}(T_8 - T_{15}) + \frac{1}{5}(T_{24} - T_{35})) + \langle e_q^2 \rangle C_{2,g} \otimes g\}$$

$$F_2^p = x\{K_{\mathrm{F2},\Sigma} \otimes \Sigma_0 + K_{\mathrm{F2},g} \otimes g_0 + K_{\mathrm{F2},+} \otimes \left(T_{3,0} + \frac{1}{3}(T_{8,0} - T_{15,0})\right)\}$$

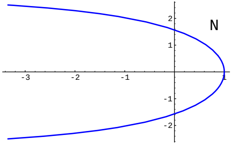
In Mellin space

$$K_{\mathrm{F2},\Sigma} = \frac{5}{18}C_{2,q}^s\Gamma_{\mathrm{S}}^{qq} + \frac{1}{30}C_{2,q}(\Gamma_{\mathrm{S}}^{24,q} - \Gamma_{\mathrm{S}}^{35,q}) + \langle e_q^2 \rangle C_{2,g}\Gamma_{\mathrm{S}}^{gq}$$

$$K_{\mathrm{F2},g} = \frac{5}{18}C_{2,q}^s\Gamma_{\mathrm{S}}^{qg} + \frac{1}{30}C_{2,q}(\Gamma_{\mathrm{S}}^{24,g} - \Gamma_{\mathrm{S}}^{35,g}) + \langle e_q^2 \rangle C_{2,g}\Gamma_{\mathrm{S}}^{gg}$$

$$K_{\mathrm{F2},+} = \frac{1}{6}C_{2,q}\Gamma_{\mathrm{NS}}^{+}$$

LH evolution benchmark



$$f_i(x,Q^2) = \sum_j \gamma_{ij} f_j(x,Q_0^2) + \int_x^1 \frac{dy}{y}\,\Gamma_{ij}(y,\alpha_s,\alpha_s^0)\left[f_j\left(\frac{x}{y},Q_0^2\right) - y f_j(x,Q_0^2)\right]$$

$$\gamma_{ij} = \int_C \frac{dN}{2\pi i}\,\frac{\Gamma_{ij}(N,\alpha_s,\alpha_s^0)}{1-N} - \int_0^x dy\,\Gamma_{ij}(y,\alpha_s,\alpha_s^0).$$

x	$\epsilon_{\rm rel}(u_v)$	$\epsilon_{\rm rel}(d_v)$	$\epsilon_{\rm rel}(\Sigma)$	$\epsilon_{\rm rel}(\bar d+\bar u)$	$\epsilon_{\rm rel}(s+\bar s)$	$\epsilon_{\rm rel}(g)$
$N_{\rm iter}=6$						
10^{-7}	$2.2\ 10^{-5}$	$8.1\ 10^{-6}$	$4.9\ 10^{-6}$	$1.5\ 10^{-5}$	$1.2\ 10^{-6}$	$2.2\ 10^{-5}$
10^{-6}	$6.3\ 10^{-6}$	$3.2\ 10^{-6}$	$9.8\ 10^{-6}$	$1.1\ 10^{-5}$	$5.4\ 10^{-6}$	$3.0\ 10^{-6}$
10^{-5}	$1.8\ 10^{-5}$	$1.4\ 10^{-5}$	$8.3\ 10^{-6}$	$3.0\ 10^{-6}$	$3.6\ 10^{-6}$	$1.4\ 10^{-6}$
10^{-4}	$3.1\ 10^{-5}$	$1.6\ 10^{-5}$	$3.6\ 10^{-5}$	$4.3\ 10^{-5}$	$3.3\ 10^{-5}$	$3.2\ 10^{-5}$
10^{-3}	$1.8\ 10^{-6}$	$1.2\ 10^{-5}$	$5.9\ 10^{-6}$	$5.8\ 10^{-6}$	$8.9\ 10^{-6}$	$3.6\ 10^{-6}$
10^{-2}	$2.8\ 10^{-5}$	$1.5\ 10^{-5}$	$4.7\ 10^{-5}$	$4.3\ 10^{-5}$	$4.6\ 10^{-5}$	$8.2\ 10^{-5}$
0.1	$3.2\ 10^{-6}$	$1.3\ 10^{-5}$	$3.0\ 10^{-6}$	$9.4\ 10^{-6}$	$2.1\ 10^{-5}$	$5.1\ 10^{-7}$
0.3	$1.9\ 10^{-6}$	$2.4\ 10^{-5}$	$6.5\ 10^{-6}$	$1.0\ 10^{-5}$	$3.2\ 10^{-6}$	$2.6\ 10^{-6}$
0.5	$1.70\ 10^{-5}$	$1.3\ 10^{-5}$	$1.5\ 10^{-5}$	$1.3\ 10^{-5}$	$3.0\ 10^{-6}$	$3.5\ 10^{-6}$
0.7	$7.0\ 10^{-5}$	$8.0\ 10^{-6}$	$5.9\ 10^{-5}$	$8.9\ 10^{-6}$	$2.4\ 10^{-5}$	$9.9\ 10^{-6}$
0.9	$1.4\ 10^{-5}$	$6.2\ 10^{-6}$	$1.3\ 10^{-5}$	$7.4\ 10^{-4}$	$1.8\ 10^{-3}$	$5.1\ 10^{-5}$

Benchmark evolution tables (hep-ph/0204316) reproduced with $\mathcal{O}(10^{-5})$ accuracy.

Theoretical errors

* Higher perturbative orders → NLO fit
* Heavy quark treatment → Zero Mass Variable Flavor Number scheme.
  Ignore intrinsic heavy quarks contributions, quarks are radiatively generated at thresholds.(Thorne,Tung, arXiv:0809.0714)
* Target Mass Corrections included and factorized into the hard kernels.

$$\widetilde{F}_2(\xi, Q^2) = \frac{x^2}{\tau^{3/2}} \frac{F_2(\xi, Q^2)}{\xi^2} + 6 \frac{M_N^2}{Q^2} \frac{x^3}{\tau^2} I_2(\xi, Q^2)$$

$$\tau = 1 + \frac{4M_N^2 x^2}{Q^2}$$

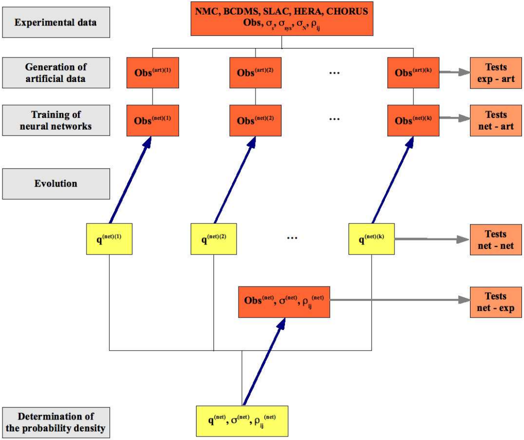
$$\xi = \frac{2x}{1+\sqrt{\tau}}$$

$$I_2(\xi, Q^2) = \int_\xi^1 \frac{dz}{z^2} F_2(z, Q^2).$$

Taking Mellin transforms with respect to ξ, define a new target mass corrected coefficient function

$$\widetilde{C}_{2,j}(N, \alpha_s, \tau) = \frac{(1+\tau^{1/2})^2}{4\tau^{3/2}} \left(1 + \frac{3\left(1 - \tau^{-1/2}\right)}{N+1}\right) C_{2,j}(N, \alpha_s)$$

Step 3: Training and dynamical stopping



Dynamical Stopping

No looking for absolute minimum but learning from data.

Fitting Strategy

Our fitting strategy is very different from that of normally used: instead of a set of basis functions with a small number of parameters, we have an unbiased basis of functions parameterized by a very large and redundant set of parameters.

CTEQ,MSTW,AL

$\mathcal{O}(20)$ parm

NNPDF

$\mathcal{O}(200)$ parm

Not trivial because ...

1. A redundant parametrization might accommodate also random fluctuations of statistical data.
2. Very large space of parameters

Ingredients of fitting procedure

1. Flexible and redundant parametrization
2. Genetic Algortihm minimization
3. Dynamical stopping criterion

Genetic Algorithm

- Set neural network parameters randomly.
- Make clones of the parameter vector and mutate them.
- Evaluate the figure of merit for each clone:

Error function

$$E^{2(k)}[\omega] = \sum_{i,j}^{N_{\rm dat}} (F_i^{(\rm art)(k)} - F_i^{(\rm net)(k)}) \left(\left(\overline{\rm cov}^{(k)} \right)^{-1} \right)_{ij} (F_j^{(\rm art)(k)} - F_j^{(\rm net)(k)})$$

$\overline{\rm cov}^{(k)}$ defined from an experimental covariance matrix which does not include normalization errors. (G. D'Agostini,2003)

$$\overline{\rm cov}_{ij}^{(k)} = \overline{\rm cov}_{ij}^{(exp)} S_{i,N}^{(k)} S_{j,N}^{(k)}$$

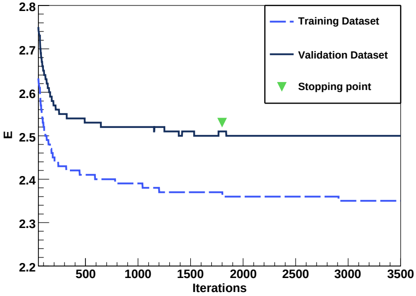
- Select the best ones and iterate the procedure until a stability is reached.

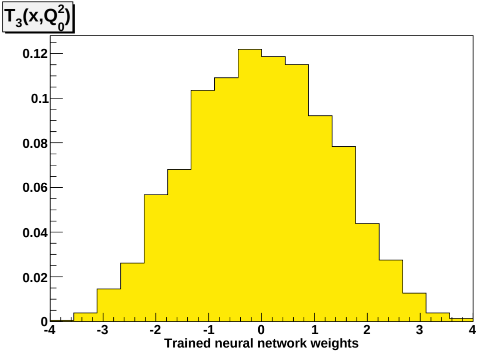
Dynamical Stopping Criterion

* GA is monotonically decreasing by construction.
* We do not want to reach the absolute minimum.
* The best fit is given by an optimal training beyond which the figure of merit improves only because we are fitting statistical noise of the data.

Cross-validation method

* Divide data in two sets: training and validation.
* Random division for each replica ($f_t = f_v = 0.5$).
* Minimisation is performed only on the training set. The validation χ^2 for the set is computed.
* When the training χ^2 still decreases while the validation χ^2 stops decreasing $\rightarrow$ STOP.

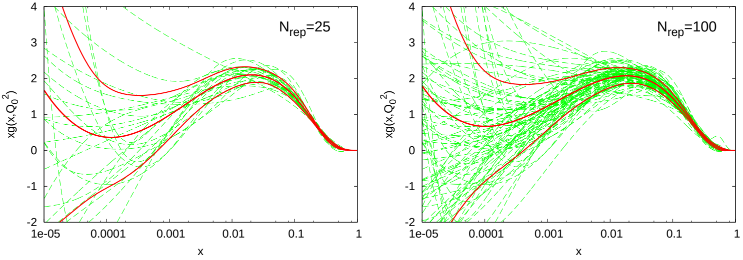




- Set of neural networks at stopping provides our best-fit.
- No physical interpretation for parameters: most unconstrained or zero.

Outline

Individual replicas vs Average quantities



Even though individual replicas may fluctuate significantly, average quantities such as central values and error bands are smooth inasmuch as stability is reached due to the number of replicas increasing.

How PDFs uncertainties must be evaluated

- Monte Carlo prescription (NNPDF)

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

- HEPDATA prescription (CTEQ and MRST/MSTW)

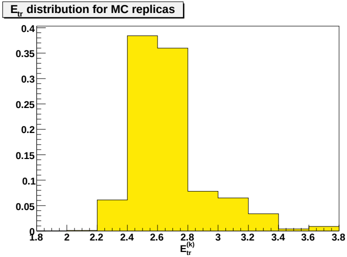
$$\sigma_{\mathcal{F}} = \frac{1}{2C_{90}} \left(\sum_{k=1}^{N_{\rm set}/2} \left(\mathcal{F}[\{f^{(2k-1)}\}] - \mathcal{F}[\{f^{(2k)}\}] \right)^2 \right)^{1/2} , \quad C_{90} = 1.64485$$

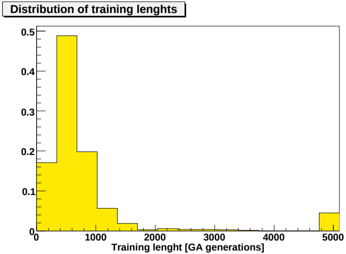
C_{90} accounts for the fact that the upper and lower parton sets correspond to 90% confidence levels rather than to one-σ uncertainties.

- HEPDATA* prescription (Alekhin)

$$\sigma_{\mathcal{F}} = \left(\sum_{k=1}^{N_{\rm set}} \left(\mathcal{F}[\{f^{(k)}\}] - \mathcal{F}[\{f^{(0)}\}] \right)^2 \right)^{1/2} .$$

NNPDF1.0: Statistical features



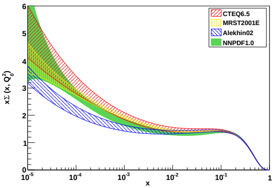


$\chi^2_{\rm tot}$	1.34
$\langle E \rangle$	2.71
$\langle E_{\rm tr} \rangle$	2.68
$\langle E_{\rm val} \rangle$	2.72
$\langle {\rm TL} \rangle$	824
$N_{\rm rep}$	1000

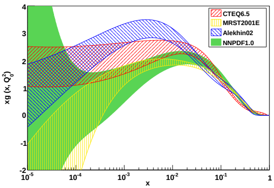
- Avg over replicas $\sim$ 1.
- Avg of replicas $\sim$ 2 for best fits.
- Poissonian distribution for Training Lenghts.

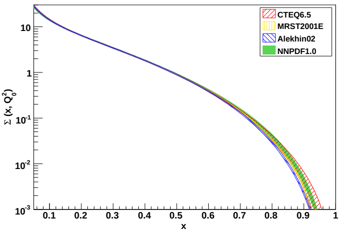
Experiment	$\chi^2_{\rm tot}$
SLAC	1.27
BCDMS	1.59
NMC	1.70
NMC-pd	1.53
ZEUS	1.11
H1	1.03
CHORUS	1.40
FLH108	1.62

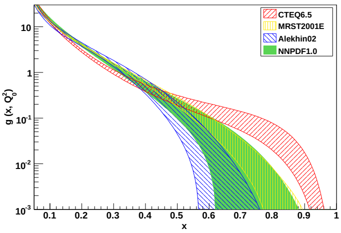
The NNPDF1.0 parton set



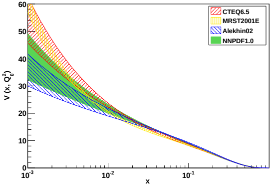
$$Q^2 = Q_0^2 = 2\ \mathrm{GeV}^2$$



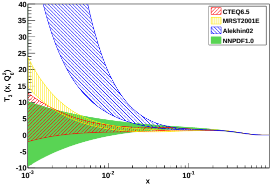


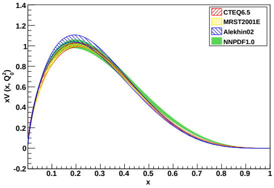


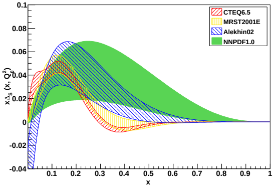
The NNPDF1.0 parton set

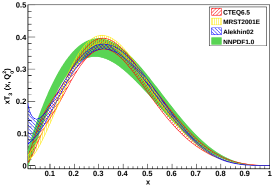


- More reliable estimation of errors.
- Compatible with other fits.
- Larger error in the extrapolation region.
- Some regions unconstrained in DIS fit.

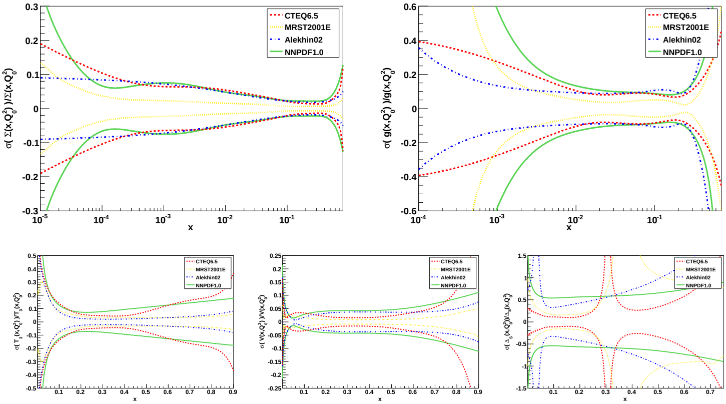








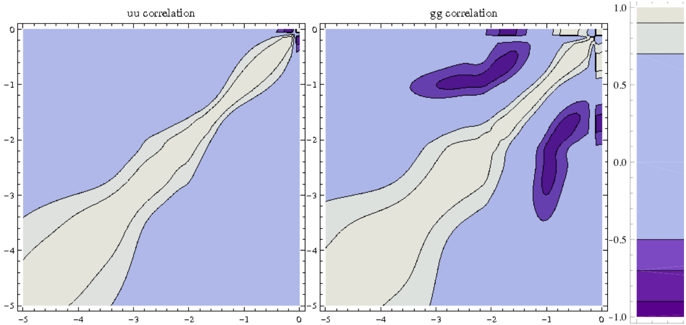
NNPDF1.0: Relative uncertainties



PDFs correlations

Correlations between $u - u$ and $g - g$ (Q=85GeV)

P.Nadolsky arXiv:0802.0007



$$\rho\left[f_a(x_1,Q_1^2)f_b(x_2,Q_2^2)\right] = \frac{\langle f_a(x_1,Q_1^2)f_b(x_2,Q_2^2)\rangle_{\rm rep} - \langle f_a(x_1,Q_1^2)\rangle_{\rm rep}\langle f_b(x_2,Q_2^2)\rangle_{\rm rep}}{\sigma_a(x_1,Q_1^2)\sigma_b(x_2,Q_2^2)}.$$

Statistical estimator: distance between MC ensembles.

* All features of the NNPDF parton set can be assessed by using standard statistical tools.
* Distances between two probability distributions:
  Quark $\left\{ f_{ik}^{(1)} = f_k^{(1)}(x_i, Q_0^2) \right\}$

$$\langle d[f] \rangle = \sqrt{\left\langle \frac{\left(\langle f_i \rangle_{(1)} - \langle f_i \rangle_{(2)} \right)^2}{\sigma^2[f_i^{(1)}] + \sigma^2[f_i^{(2)}]} \right\rangle_{\rm pts}}$$

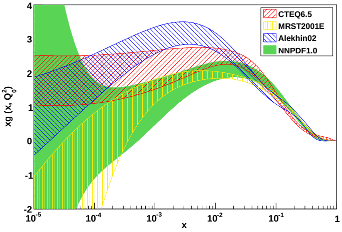
* With:

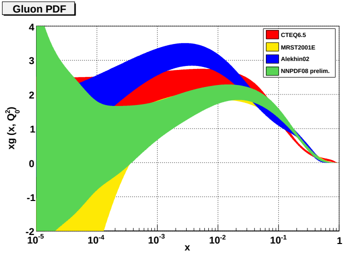
$$\langle f_i \rangle_{(1)} \equiv \frac{1}{N_{\rm rep}^{(1)}} \sum_{k=1}^{N_{\rm rep}^{(1)}} f_{ik}^{(1)} \ ,$$

$$\sigma^2[f_i^{(1)}] \equiv \frac{1}{N_{\rm rep}^{(1)}(N_{\rm rep}^{(1)} - 1)} \sum_{k=1}^{N_{\rm rep}^{(1)}} \left(f_{ik}^{(1)} - \langle f_i \rangle_{(1)} \right)^2$$

* For statistically equivalent PDF sets: $\langle d[f] \rangle \sim \langle d[\sigma_f] \rangle \sim 1$

Stability under variation of the parametrization

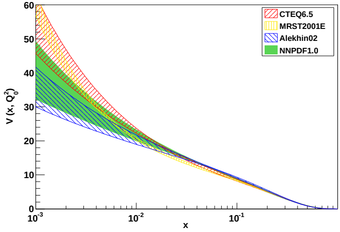


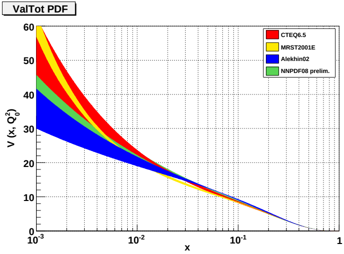


	Data	Extrapolation
$\Sigma(x, Q_0^2)$	$5\ 10^{-4} \leq x \leq 0.1$	$10^{-5} \leq x \leq 10^{-4}$
$\langle d[f]\rangle$	0.98	1.25
$\langle d[\sigma]\rangle$	1.14	1.34
$g(x, Q_0^2)$	$5\ 10^{-4} \leq x \leq 0.1$	$10^{-5} \leq x \leq 10^{-4}$
$\langle d[f]\rangle$	1.52	1.15
$\langle d[\sigma]\rangle$	1.16	1.07
$T_3(x, Q_0^2)$	$0.05 \leq x \leq 0.75$	$10^{-3} \leq x \leq 10^{-2}$
$\langle d[f]\rangle$	1.00	1.11
$\langle d[\sigma]\rangle$	1.76	2.27
$V(x, Q_0^2)$	$0.1 \leq x \leq 0.6$	$3\ 10^{-3} \leq x \leq 3\ 10^{-2}$
$\langle d[f]\rangle$	1.30	0.90
$\langle d[\sigma]\rangle$	1.10	0.98
$\Delta_S(x, Q_0^2)$	$0.1 \leq x \leq 0.6$	$3\ 10^{-3} \leq x \leq 3\ 10^{-2}$
$\langle d[f]\rangle$	1.04	1.91
$\langle d[\sigma]\rangle$	1.44	1.80

* Stability under change of architecture of the nets: **37 pars** $\longrightarrow$ **31 pars**
* Independence on the parametrization!

Stability under variation of the parametrization





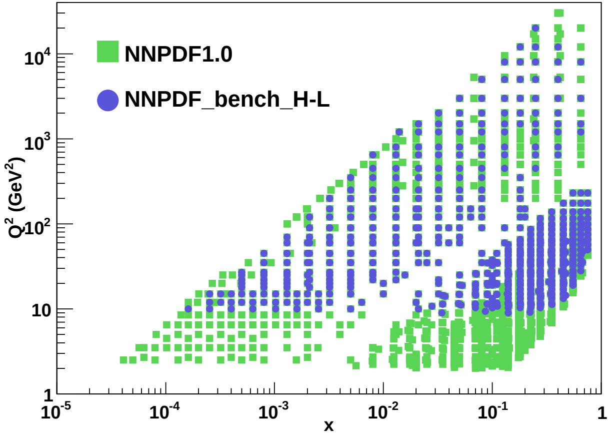
	Data	Extrapolation
$\Sigma(x, Q_0^2)$	$5\ 10^{-4} \leq x \leq 0.1$	$10^{-5} \leq x \leq 10^{-4}$
$\langle d[f] \rangle$	0.98	1.25
$\langle d[\sigma] \rangle$	1.14	1.34
$g(x, Q_0^2)$	$5\ 10^{-4} \leq x \leq 0.1$	$10^{-5} \leq x \leq 10^{-4}$
$\langle d[f] \rangle$	1.52	1.15
$\langle d[\sigma] \rangle$	1.16	1.07
$T_3(x, Q_0^2)$	$0.05 \leq x \leq 0.75$	$10^{-3} \leq x \leq 10^{-2}$
$\langle d[f] \rangle$	1.00	1.11
$\langle d[\sigma] \rangle$	1.76	2.27
$V(x, Q_0^2)$	$0.1 \leq x \leq 0.6$	$3\ 10^{-3} \leq x \leq 3\ 10^{-2}$
$\langle d[f] \rangle$	1.30	0.90
$\langle d[\sigma] \rangle$	1.10	0.98
$\Delta_S(x, Q_0^2)$	$0.1 \leq x \leq 0.6$	$3\ 10^{-3} \leq x \leq 3\ 10^{-2}$
$\langle d[f] \rangle$	1.04	1.91
$\langle d[\sigma] \rangle$	1.44	1.80

* Stability under change of architecture of the nets: **37 pars** $\longrightarrow$ **31 pars**
* Independence on the parametrization!

Dependence on data sets

HERA-LHC benchmark

Benchmark PDF fit to a reduced consistent set of DIS data.(hep-ph/0511119)



Dependence on data sets

HERA-LHC benchmark

Benchmark PDF fit to a reduced consistent set of DIS data.(hep-ph/0511119)

Set	$N_{\rm dat}$	$x_{\rm min}$	$x_{\rm max}$	$Q^2_{\rm min}$	$Q^2_{\rm max}$
SLACp	211 (47)	.07000	.85000	0.6	29.
SLACd	211 (47)	.07000	.85000	0.6	29.
BCDMSp	351 (333)	.07000	.75000	7.5	230.
BCDMSd	254 (248)	.07000	.75000	8.8	230.
NMC	288 (245)	.00350	.47450	0.8	61.
NMC-pd	260 (153)	.00150	.67500	0.2	99.
Z97lowQ2	80	.00006	.03200	2.7	27.
Z97NC	160	.00080	.65000	35.0	20000.
Z97CC	29	.01500	.42000	280.0	17000.
Z02NC	92	.00500	.65000	200.0	30000.
Z02CC	26	.01500	.42000	280.0	30000.
Z03NC	90	.00500	.65000	200.0	30000.
Z03CC	30	.00800	.42000	280.0	17000.
H197mb	67 (55)	.00003	.02000	1.5	12.
H197lowQ2	80	.00016	.20000	12.0	150.
H197NC	130	.00320	.65000	150.0	30000.
H197CC	25	.01300	.40000	300.0	15000.
H199NC	126	.00320	.65000	150.0	30000.
H199CC	28	.01300	.40000	300.0	15000.
H199NChy	13	.00130	.01050	100.0	800.
H100NC	147	.00131	.65000	100.0	30000.
H100CC	28	.01300	.40000	300.0	15000.
CHORUSν	607 (471)	.02000	.65000	0.3	95.
CHORUS$\bar{\nu}$	607 (471)	.02000	.65000	0.3	95.
FLH108	8	.00028	.00360	12.0	95.

$$Q^2 > 2\,\mathrm{GeV}^2$$
$$W^2 > 12.5\,\mathrm{GeV}^2$$

Dependence on data sets

HERA-LHC benchmark

Benchmark PDF fit to a reduced consistent set of DIS data.(hep-ph/0511119)

Set	N_{dat}	x_{min}	x_{max}	Q^2_{min}	Q^2_{max}
BCDMSp	322	$7\ 10^{-2}$	0.75	10.3	230
NMC	95	0.028	0.48	9	6
NMC-pd	73	0.035	0.67	11.4	99
Z97NC	206	$1.6\ 10^{-4}$	0.65	10	$2\ 10^4$
H197lowQ^2	77	$3.2\ 10^{-4}$	0.2	12	150

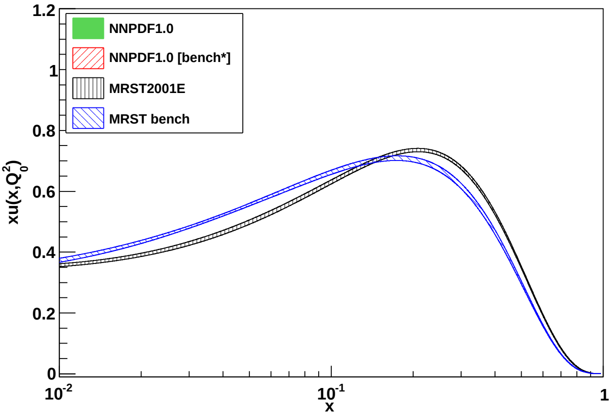
$$Q^2 > 9\,\mathrm{GeV}^2$$
$$W^2 > 15\,\mathrm{GeV}^2$$

3163 data $\longrightarrow$ **773** data

Dependence on data sets

HERA-LHC benchmark

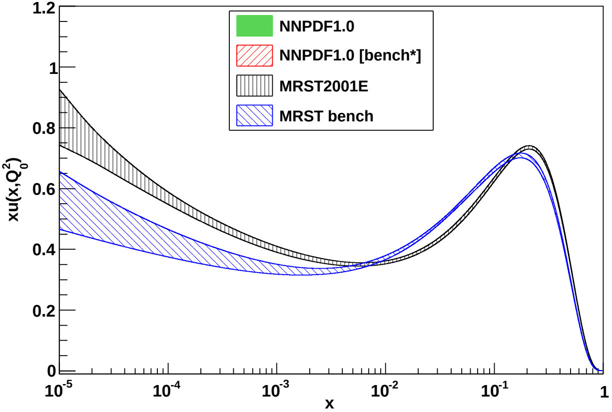
Comparison between collaborations and between benchmark/global partons.
$u(x, Q^2 = 2\mathrm{GeV}^2)$: MRST data region



Dependence on data sets

HERA-LHC benchmark

Comparison between collaborations and between benchmark/global partons.
$u(x, Q^2 = 2\text{GeV}^2)$: MRST extrapolation region



Dependence on data sets

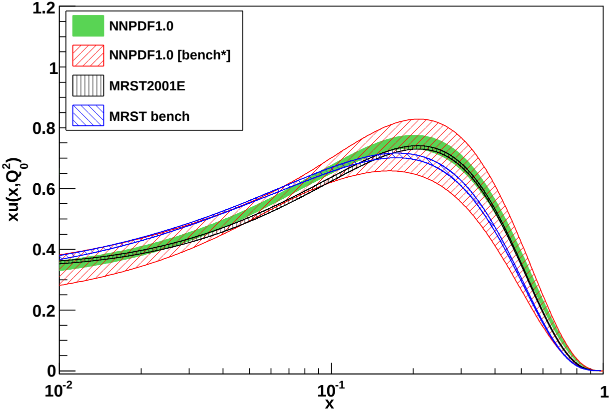
HERA-LHC benchmark

- Benchmark partons and global partons do not agree within error!
- Note that PDFs input parametrization, flavor assumptions and statistical treatment ($\Delta\chi^2_{\rm global} = 50$, $\Delta\chi^2_{\rm bench} = 1$) are tuned to data.
- This is not satisfactory especially to predict the behaviour of PDFs in the extrapolation region (LHC)

Dependence on data sets

HERA-LHC benchmark

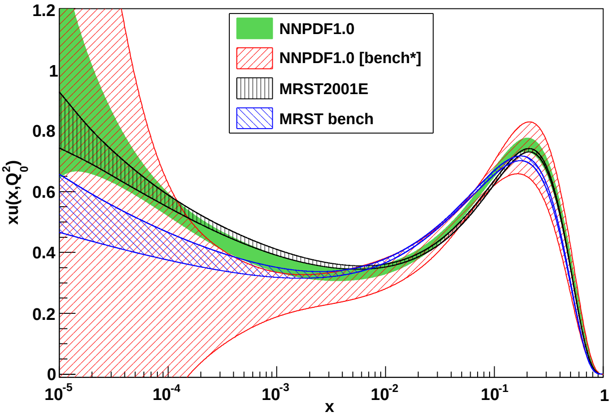
Comparison between collaborations and between benchmark/global partons.
$u(x, Q^2 = 2\mathrm{GeV}^2)$: Data region



Dependence on data sets

HERA-LHC benchmark

Comparison between collaborations and between benchmark/global partons.
$u(x, Q^2 = 2\mathrm{GeV}^2)$: Extrapolation Region

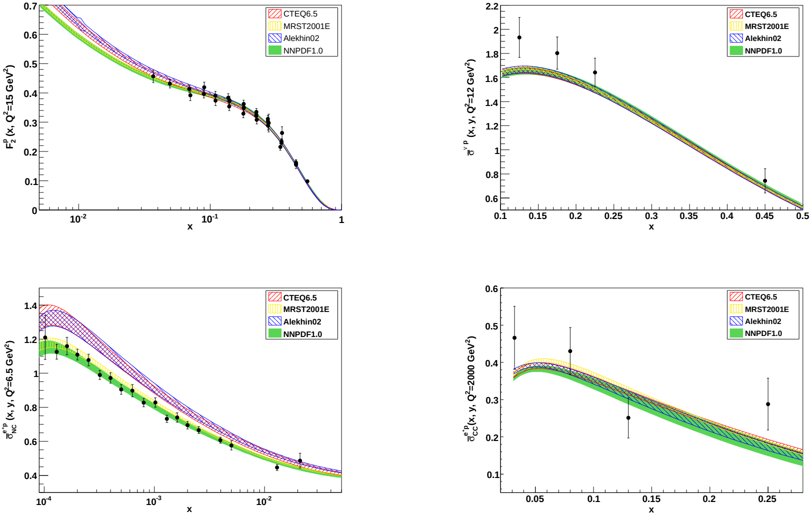


Dependence on data sets

HERA-LHC benchmark

- NNPDF1.0 is consistent with MRST global fit.
- NNPDFbench is consistent with NNPDF1.0 and MRST.
- Same parametrization and flavour assumption.
- Same statistical treatment.
- Underestimation of the error in the standard approach.

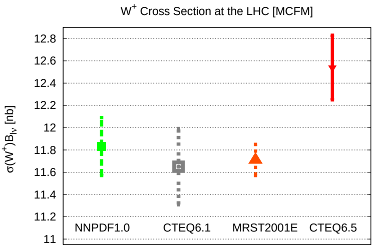
Comparison with present experimental data

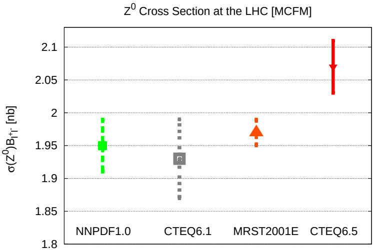


Prediction on LHC standard candle processes

- Gauge boson production at the LHC.
- All quantities have been computed at NLO with MCFM (http://mcfm.fnal.gov)
- Quoted uncertainties are the one-σ bands due to the PDF uncertainty only.

	$\sigma_{W^+}\mathcal{B}_{l^+\nu_l}$	$\Delta\sigma_{W^+}/\sigma_{W^+}$	$\sigma_Z\mathcal{B}_{l^+l^-}$	$\Delta\sigma_Z/\sigma_Z$
NNPDF1.0	11.83 ± 0.26	2.2%	1.95 ± 0.04	2.1%
CTEQ6.1	11.65 ± 0.34	2.9%	1.93 ± 0.06	3.1%
MRST01	11.71 ± 0.14	1.2%	1.97 ± 0.02	1.0%
CTEQ6.5	12.54 ± 0.29	2.3%	2.07 ± 0.04	1.9%





Outline

Conclusions

- Standard approaches with fixed parametrization tend to underestimate uncertainties unless experimental errors are inflated by essentially arbitrary amount.
- Monte Carlo ensemble
  - Any statistical property of PDFs can be calculated using standard statistical methods.
  - No need of any tolerance criterion.
- The Neural Network parametrization
  - Small uncertainties come from an underlying physical law, not from parametrization bias.
  - Inconsistent data or underestimated uncertainties do not require a separate treatment and are automatically signalled by a larger value of the χ^2.
- The first NNPDF parton set [arXiv:0808.1231] is available on the common LHAPDF interface (http://projects.hepforge.org/lhapdf).

Outlook

- Inclusion of hadronic data to
  - improve the accuracy of gluon at large x (jets)
  - determine the light antiquark sea asymmetry (Drell-Yan)
  - allow for a direct determination of the strange distribution (Dimuon data)
- More accurate treatment of Heavy Quark thresholds by including terms proportional to the heavy quark mass.
- LO parton set in view of its use in Monte Carlo generators.
- More sophisticated theoretical treatment: NNLO parton distributions, large and small x resummation corrections should also be considered.
- Study of the impact of PDFs uncertainties on LHC phenomenology (optimized small set of parton distribution)

EXTRA MATERIAL

Stability under variation of preprocessing exponents

- Polynomial preprocessing functions are introduced in order to speed up the training but should not affect final results.
- Checked the stability of result upon variation of the preprocessing exponents away from their default values.

Valence sector			Singlet sector		
	χ^2	$\langle \mathrm{TL} \rangle$		χ^2	$\langle \mathrm{TL} \rangle$
$n_{T_3} = n_V = 0.1$	1.38	771	$n_\Sigma = n_g = 0.8$	1.39	1002
$n_{T_3} = n_V = 0.5$	1.34	1629	$n_\Sigma = n_g = 1.6$	1.52	2287
$m_{T_3} = m_V = 2$	1.55	1186	$m_\Sigma = m_g - 1 = 2$	1.37	647
$m_{T_3} = m_V = 4$	1.28	1311	$m_\Sigma = m_g - 1 = 4$	1.41	1306

- An increase of $N_{\mathrm{gen}}^{\mathrm{max}}$ or a more efficient minimization algorithm would be required to obtain a satisfactory fit. Thus, in these cases we expect reduced stability of results: this applies to the case of increase of the small x preprocessing exponent n for the singlet, and to a lesser extent to the decrease of the large x preprocessing exponent m for the nonsinglet and valence.
- Randomized preprocessing?

Statistical Estimators I: observables

- Central value of the i-th experimental point

$$\left\langle F_i^{(\mathrm{art})} \right\rangle_{\mathrm{rep}} = \frac{1}{N_{\mathrm{rep}}} \sum_{k=1}^{N_{\mathrm{rep}}} F_i^{(\mathrm{art})(k)} \; .$$

- Variance of the i-th experimental point

$$\sigma_i^{(\mathrm{art})} = \sqrt{\left\langle \left(F_i^{(\mathrm{art})} \right)^2 \right\rangle_{\mathrm{rep}} - \left\langle F_i^{(\mathrm{art})} \right\rangle_{\mathrm{rep}}^2} \; .$$

- Associated covariance:

$$\rho_{ij}^{(\mathrm{art})} = \frac{\left\langle F_i^{(\mathrm{art})} F_j^{(\mathrm{art})} \right\rangle_{\mathrm{rep}} - \left\langle F_i^{(\mathrm{art})} \right\rangle_{\mathrm{rep}} \left\langle F_j^{(\mathrm{art})} \right\rangle_{\mathrm{rep}}}{\sigma_i^{(\mathrm{art})} \sigma_j^{(\mathrm{art})}} \; .$$

$$\mathrm{cov}_{ij}^{(\mathrm{art})} = \rho_{ij}^{(\mathrm{art})} \sigma_i^{(\mathrm{art})} \sigma_j^{(\mathrm{art})} .$$

Statistical Estimators II: replicas vs data

- Mean variance and percentage error on central values over the $N_{\rm dat}$ data points.

$$\left\langle V\left[\left\langle F^{(\rm art)}\right\rangle_{\rm rep}\right]\right\rangle_{\rm dat} = \frac{1}{N_{\rm dat}}\sum_{i=1}^{N_{\rm dat}}\left(\left\langle F_i^{(\rm art)}\right\rangle_{\rm rep} - F_i^{(\rm exp)}\right)^2 ,$$

$$\left\langle PE\left[\left\langle F^{(\rm art)}\right\rangle_{\rm rep}\right]\right\rangle_{\rm dat} = \frac{1}{N_{\rm dat}}\sum_{i=1}^{N_{\rm dat}}\left[\frac{\left\langle F_i^{(\rm art)}\right\rangle_{\rm rep} - F_i^{(\rm exp)}}{F_i^{(\rm exp)}}\right] .$$

- $\left\langle V\left[\left\langle \sigma^{(\rm art)}\right\rangle_{\rm rep}\right]\right\rangle_{\rm dat}$, $\left\langle V\left[\left\langle \rho^{(\rm art)}\right\rangle_{\rm rep}\right]\right\rangle_{\rm dat}$, $\left\langle V\left[\left\langle {\rm cov}^{(\rm art)}\right\rangle_{\rm rep}\right]\right\rangle_{\rm dat}$

  $\left\langle PE\left[\left\langle \sigma^{(\rm art)}\right\rangle_{\rm rep}\right]\right\rangle_{\rm dat}$, $\left\langle PE\left[\left\langle \rho^{(\rm art)}\right\rangle_{\rm rep}\right]\right\rangle_{\rm dat}$, $\left\langle PE\left[\left\langle {\rm cov}^{(\rm art)}\right\rangle_{\rm rep}\right]\right\rangle_{\rm dat}$

  relative to errors, correlations and covariances are defined in the same way.
- These estimators indicate how close are the averages over generated data and the experimental values.

Stability estimators III: replicas vs data

- Scatter correlation:

$$r\left[F^{(\mathrm{art})}\right]=\frac{\left\langle F^{(\mathrm{exp})}\left\langle F^{(\mathrm{art})}\right\rangle_{\mathrm{rep}}\right\rangle_{\mathrm{dat}}-\left\langle F^{(\mathrm{exp})}\right\rangle_{\mathrm{dat}}\left\langle\left\langle F^{(\mathrm{art})}\right\rangle_{\mathrm{rep}}\right\rangle_{\mathrm{dat}}}{\sigma_s^{(\mathrm{exp})}\sigma_s^{(\mathrm{art})}},$$

  where the scatter variances are defined as

$$\sigma_s^{(\mathrm{exp})}=\sqrt{\left\langle\left(F^{(\mathrm{exp})}\right)^2\right\rangle_{\mathrm{dat}}-\left(\left\langle F^{(\mathrm{exp})}\right\rangle_{\mathrm{dat}}\right)^2},$$

$$\sigma_s^{(\mathrm{art})}=\sqrt{\left\langle\left(\left\langle F^{(\mathrm{art})}\right\rangle_{\mathrm{rep}}\right)^2\right\rangle_{\mathrm{dat}}-\left(\left\langle\left\langle F^{(\mathrm{art})}\right\rangle_{\mathrm{rep}}\right\rangle_{\mathrm{dat}}\right)^2}.$$

- $r\left[\sigma^{(\mathrm{art})}\right]$ $r\left[\rho^{(\mathrm{art})}\right]$ $r\left[\mathrm{cov}^{(\mathrm{art})}\right]$ are defined in the same way.
- The scatter correlation indicates the size of the spread of data around a straight line. Specifically $r\left[\sigma^{(\mathrm{art})}\right]=1$ implies that $\left\langle\sigma_i^{(\mathrm{art})}\right\rangle$ is proportional to $\sigma_i^{(\mathrm{exp})}$.

Covariance matrix

Once the systematics are known, the experimental covariance matrix for each experiment can be easily computed

$$\mathrm{cov}_{ij} = \left(\sum_{p=1}^{N_{\mathrm{sys}}} \sigma_{i,p}\sigma_{j,p} + F_i F_j \sigma_N^2 \right) + \delta_{ij}\sigma_{i,s}^2 \ ,$$

where F_i, F_j are central values, $\sigma_{i,p}$ are the N_{sys} correlated systematic errors, σ_N is the total normalization uncertainty, and $\sigma_{i,s}$ is the statistical uncertainty. The correlation matrix is given by

$$\rho_{ij} = \frac{\mathrm{cov}_{ij}}{\sigma_{i,\mathrm{tot}}\sigma_{j,\mathrm{tot}}} \ ,$$

where the total error $\sigma_{i,\mathrm{tot}}$ for the i-th point is given by

$$\sigma_{i,\mathrm{tot}} = \sqrt{\sigma_{i,s}^2 + \sigma_{i,c}^2 + F_i^2\sigma_N^2}$$

and the total correlated uncertainty $\sigma_{i,c}$ is the sum of all correlated systematics

$$\sigma_{i,c}^2 = \sum_{p=1}^{N_{\mathrm{sys}}} \sigma_{i,p}^2 \ .$$

Normalization bias

Normalization errors are not included in the covariance matrix on the same footing as other sources of systematics since this would bias the fit. Rather, normalization errors are included by rescaling all errors independently for each replica by a factor

$$\overline{\sigma}_{i,s}^{(k)} = (1 + r_N^{(k)}\sigma_N)\sigma_{i,s} \ ,$$
$$\overline{\sigma}_{i,p}^{(k)} = (1 + r_N^{(k)}\sigma_N)\sigma_{i,p} \qquad p = 1 \ , \ldots , N_{\rm sys} \ ,$$

where $r_N^{(k)}$ is the random variable associated to the normalization uncertainty. The covariance matrix is then given by

$$\overline{{\rm cov}^{(k)}}_{ij} = \left(\sum_{p=1}^{N_{\rm sys}} \overline{\sigma}^{(k)}{}_{i,p}\overline{\sigma}^{(k)}{}_{j,p} \right) + \delta_{ij}\overline{\sigma}^{(k)2}{}_{i,s} \ ,$$

in terms of the rescaled uncertainties.

Error function and χ^2

- Error function

$$E^{(k)}\left[\omega\right] = \frac{1}{N_{\mathrm{dat}}} \sum_{i,j=1}^{N_{\mathrm{dat}}} \left(F_i^{(\mathrm{art})(k)} - F_i^{(\mathrm{net})(k)}\right) \left(\left(\overline{\mathrm{cov}}\right)^{-1}\right)_{ij} \left(F_j^{(\mathrm{art})(k)} - F_j^{(\mathrm{net})(k)}\right),$$

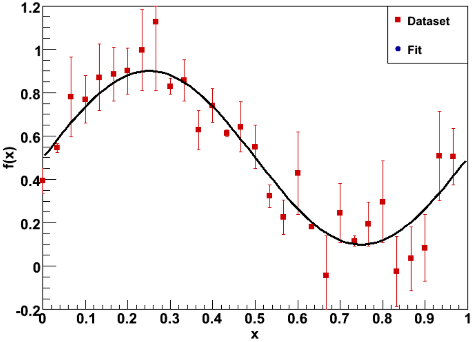
  The covariance matrix $\overline{\mathrm{cov}}^{(k)}$ does not include normalization errors.

- $E^{(k)}$ is a property of each individual replica, whereas the quality of the global fit is given by the χ^2 computed from the averages over the sample of trained neural networks.
- χ^2

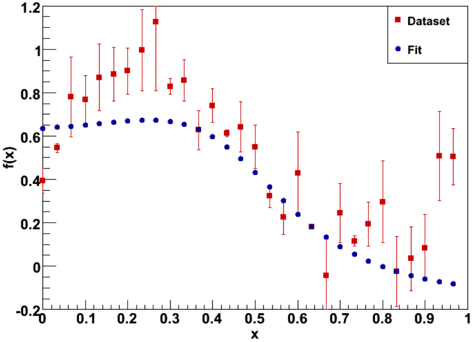
$$\chi^2 = \frac{1}{N_{\mathrm{dat}}} \sum_{i,j=1}^{N_{\mathrm{dat}}} \left(F_i^{(\mathrm{exp})} - \langle F_i^{(\mathrm{net})} \rangle_{\mathrm{rep}}\right) \left(\left(\mathrm{cov}\right)^{-1}\right)_{ij} \left(F_j^{(\mathrm{exp})} - \langle F_j^{(\mathrm{net})} \rangle_{\mathrm{rep}}\right),$$

  where now the covariance matrix includes normalization uncertainties.

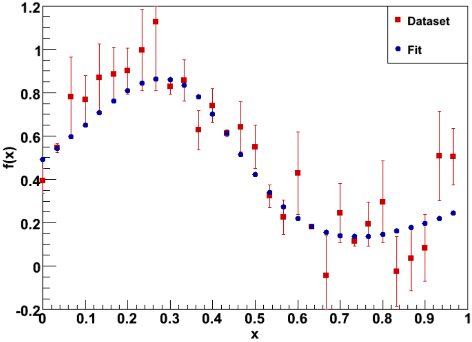
Proper Fitting avoiding Overlearning: an example



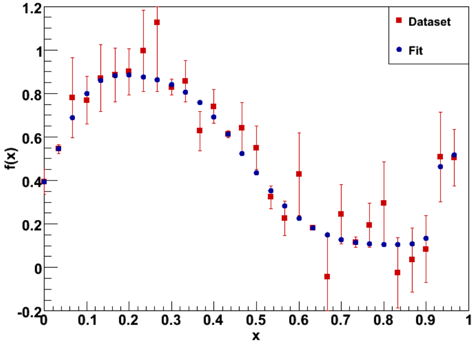
Proper Fitting avoiding Overlearning: an example



Proper Fitting avoiding Overlearning: an example



Proper Fitting avoiding Overlearning: an example

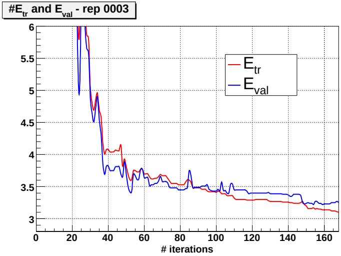


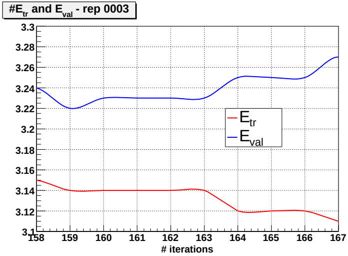
- Need a redundant parametrization to avoid excessive constraining
- Need a way of stopping the fit before overlearning sets in

How to avoid Overlearning?

Stopping criterion based on Training-Validation separation

* Divide data in two sets: training and validation.
* Minimisation is performed only on the training set. The validation χ^2 for the set is computed.
* When the training χ^2 still decreases while the validation χ^2 stops decreasing → STOP.



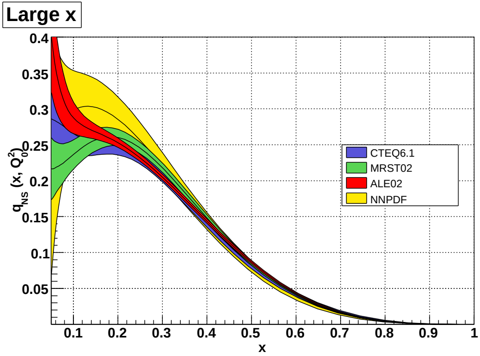


Non singlet fit

- Determination of

$$T_3(x, Q_0^2) \equiv (u + \bar{u} - d - \bar{d})(x, Q_0^2)$$

  at $Q_0^2 = 2\text{GeV}^2$ at LO, NLO, NNLO.
- DATA SETS: $F_2^p(x, Q^2) - F_2^d(x, Q^2)$ BCDMS and NMC



See hep-ph/0701127 for all technical details