The Neural Network approach to PDF fitting

The NNPDF Collaboration:
Luigi Del Debbio[1], Stefano Forte[2],
José I. Latorre[3], Andrea Piccione[4, 5] and Joan Rojo[3],

[1] Theory Division, CERN,
[2] Dipartimento di Fisica, Università di Milano and INFN, Sezione di Milano
[3] Departament d'Estructura i Constituents de la Matèria, Universitat de Barcelona
[4] INFN Sezione di Torino
[5] Dipartimento di Fisica Teorica, Università di Torino

Deep Inelastic Scattering 2005,
Structure Functions Working Group ,
Madison, April 2005

Motivation

- The problem: parametrization of parton distributions from data.
- Shortcomings of standard approaches to pdf global fits:
  1. A priori bias introduced by the choice of a fixed functional form.
  2. Problems with estimation of uncertainties:
     - The pdf parametrization affects (in an unknown way) the representation of uncertainties.
     - Incompatible data, non gaussian errors: (arbitrary) Tolerance criteria.
  3. Correct implementation of error propagation: Some methods use (not trustable) linear approximations.
- Very relevant problem: In parallel with the determination of best-fit PDFs an equally important front in global analysis has been opened ... the development of quantifiable uncertainties on the PDFs ... Much progress has been made, many useful results have been obtained, but there are no unambiguous conclusions, W.K. Tung, hep-ph/0410139

What is the problem?

- For a single quantity $\rightarrow$ 1 sigma errors
- For a pair of numbers $\rightarrow$ 1 sigma ellipse
- For a function $\rightarrow$ We need the probability measure $\mathcal{P}[f]$ in the space of functions $f(x)$

Expectation values $\rightarrow$ Functional integrals

$$\langle \mathcal{F}[f(x)] \rangle = \int \underbrace{\mathcal{D}f \; \mathcal{P}[f(x)]}_{\text{Integ. meas.}} \underbrace{\mathcal{F}[f(x)]}_{\text{Observable}}$$

The problem: Determine an infinite-dimensional object (a function) from finite set of data points.

$\rightarrow$ **Mathematically ill-posed** problem.

What is the problem?

Parton distributions $\rightarrow$ DIS structure functions

$$F(x, Q^2) = \sum_f C_f \left(\alpha_s \left(Q^2\right)\right) \otimes q_f(Q^2) + C_g \left(\alpha_s \left(Q^2\right)\right) \otimes g_f(Q^2)$$

- Trivial complication: disentangle quark flavors and gluon, evolution, deconvolution.
- Serious complication: determine error on pdfs $f(x)$, $f = q_i, g$

A (marginally) simpler problem: Determine the structure function $F(x, Q^2)$ with associated errors.
The NNPDF program: proceed in two steps:

1. Determination of structure functions: Completed
2. Determination of parton distributions: Preliminary results

Step 1:
Determination of Structure Functions

(Completed)

The NNPDF Collaboration approach:

Use neural networks as unbiased universal interpolants to construct a probability measure in the space of structure functions $\mathcal{P}\left[F(x,Q^2)\right]$ from experimental data.

General strategy:

1. Monte Carlo sampling of data (Generation of replicas of experimental data):
   Faithful representation of uncertainties

2. Neural network training over Monte Carlo replicas:
   Unbiased parametrization.

The probability measure $\mathcal{P}\left[F\right]$ contains all information from experimental data (central values, errors, correlations) with the only assumption of smoothness.
Expectation values $\rightarrow$ Functional integrals over probability measure

$$\left\langle \mathcal{F}\left[F(x,Q^2)\right]\right\rangle = \int \underbrace{\mathcal{D}F\ \mathcal{P}\left[F(x)\right]}_{\text{Int. meas}}\ \underbrace{\mathcal{F}\left[F(x)\right]}_{\text{Observable}} = \frac{1}{N_{rep}}\sum_{k=1}^{N_{rep}} \mathcal{F}\left(F^{(net)(k)}(x,Q^2)\right)$$

1.- Monte Carlo sampling of experimental data

Generate N_{rep} Monte Carlo sets of 'pseudo-data', replicas of the original N_{dat} data points $F_i^{(exp)}$

$$F_i^{(art)(k)} \quad k = 1, \ldots, N_{rep}, \quad i = 1, \ldots, N_{dat}$$

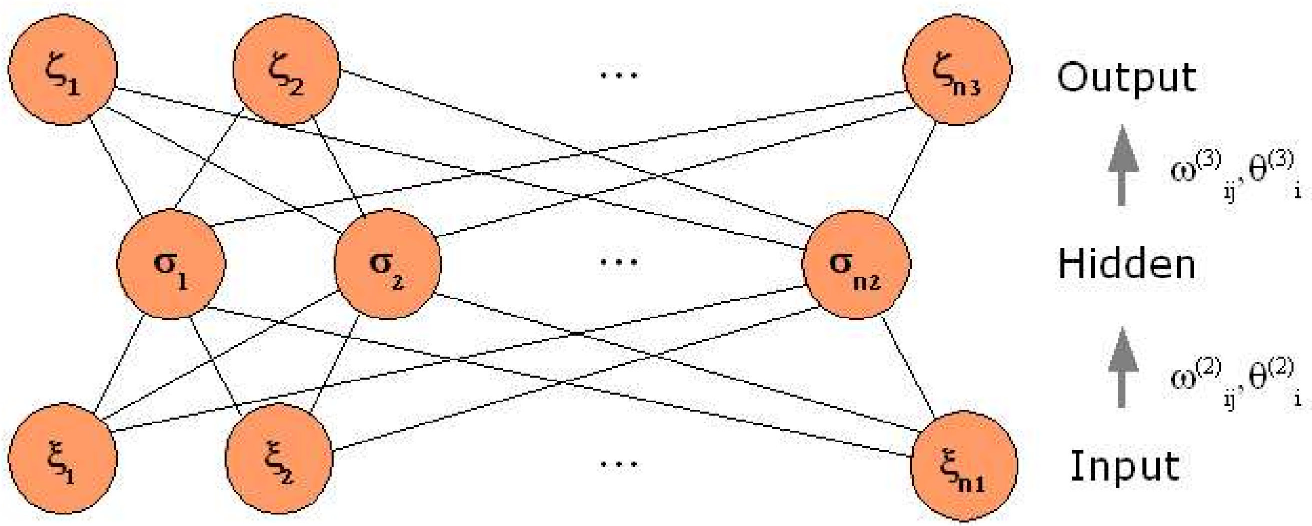
using full information on experimental errors and correlations:

$$F_i^{(art)(k)} = \left(1 + r_N^{(k)} \sigma_N\right) \left[F_i^{(exp)} + r_i^s \sigma_i^{stat} + \sum_{l=1}^{N_{sys}} r^{l,(k)} \sigma_i^{sys,l}\right]$$

$r^{(k)}$: Gaussian random numbers.
Size of set of replicas $\left\{F^{(art)(k)}\right\}$ large enough to reproduce central values, errors and correlations of exp. data.
Similar to the Bayesian Monte Carlo approach (Giele ,Kosower, Keller 01).

2.- Neural network replica training

Neural network: highly nonlinear mapping between input and output patterns, defined by its parameters (weights $\omega_{ij}^{(l)}$ and thresholds $\theta_i^{(l)}$)



Neural networks are suitable to parametrize PDFs as

- Are the most unbiased prior.
- Robust, unbiased universal approximants
- Interpolate between data points with only assumption → smoothness.

2.- Neural network replica training

Perceptrons: feed-forward multilayer neural networks

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

Choose redundant architecture $\longrightarrow$ No smoothing bias

Neural network training (PDF fitting):

Minimization of χ^2 with experimental covariance matrix.

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

Training method $\longrightarrow$ Genetic Algorithms

GA $\longrightarrow$ effective to find the global minimum, but slow convergence rate

Set of trained nets $\left\{F^{(net)(k)}(x,Q^2)\right\} \equiv$ **Probability measure** $\mathcal{P}\left[F(x,Q^2)\right]$

$\rightarrow$ Compute observables with errors and correlations from weighted averages.

Ex.1 : Average and error of structure function for arbitrary (x,Q^2):

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

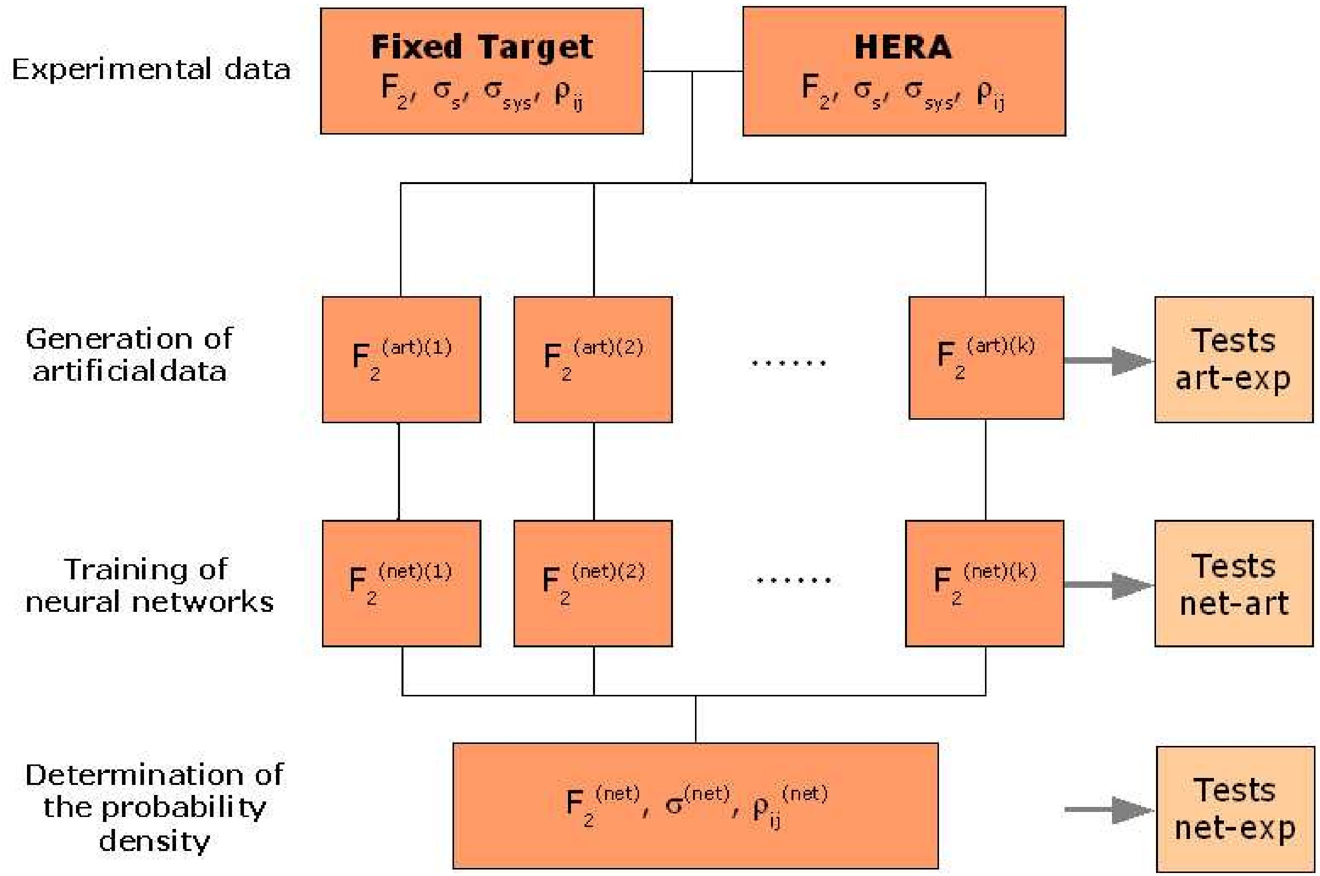
$$\sigma(x,Q^2) = \sqrt{\langle F(x,Q^2)^2\rangle - \langle F(x,Q^2)\rangle^2}$$

No need of linear approximations in error propagation.

Ex.2 : Correlations between (arbitrary) pairs of points:

$$\left\langle F(x_1,Q_1^2)F(x_2,Q_2^2)\right\rangle = \frac{1}{N_{rep}}\sum_{k=1}^{N_{rep}} F^{(net)(k)}(x_1,Q_1^2)F^{(net)(k)}(x_2,Q_2^2)$$

Summary of the NNPDF strategy



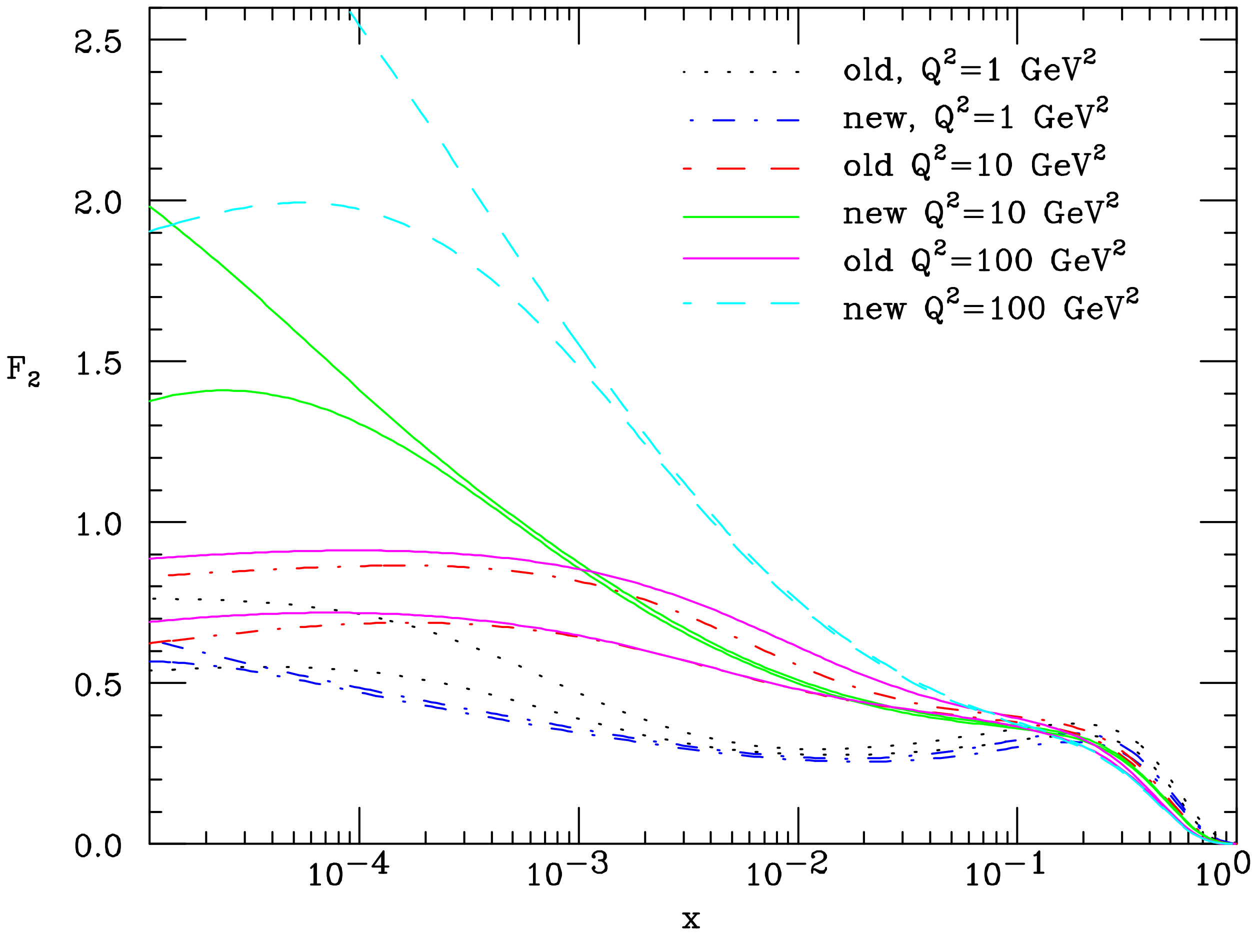
Previous work

- S. Forte, L. Garrido, J. I. Latorre and A. Piccione, "Neural network parametrization of deep-inelastic structure functions," JHEP **0205** (2002) 062 [arXiv:hep-ph/0204232].
  $\longrightarrow$ Determination of F_2^p, F_2^d, F_2^{NS} from NMC and BCDMS data
- L. Del Debbio, S. Forte, J. I. Latorre, A. Piccione and J. Rojo [NNPDF Collaboration], "Unbiased determination of the proton structure function F_2^p with faithful uncertainty estimation" , arXiv:hep-ph/0501067.
  $\longrightarrow$ Determination of F_2^p from all available data (including HERA)
  $\longrightarrow$ Incorporates data from 13 experiments in very different kinematical regions.

Source code, driver program and graphical web interface for F_2 plots and numerical computations available

http://sophia.ecm.ub.es/f2neural

Comparing old and new fits of $F_2^p(x, Q^2)$



Features:

- Compatibility old & new
- Extrapolation
- Faithful uncertainty est.

Step 2
Determination of Parton Distributions

(preliminary results)

The neural network approach to pdf fitting

Same strategy as with structure functions + Altarelli-Parisi evolution

1. Monte Carlo sampling of structure functions data $\rightarrow$ Faithful estimation of uncertainties
2. Parametrize parton distributions with neural networks $\rightarrow$ Unbiased parametrization.
3. Evolution of parton distributions to experimental data scale and training over Monte Carlo replica sample.

The probability measure $\mathcal{P}[q]$ contains all information from experimental data (central values, errors, correlations) with the only assumption of smoothness.

The neural network approach to pdf fitting

Expectation values → Functional integrals over probability measure

$$\langle \mathcal{F}\left[q(x)\right]\rangle = \int \mathcal{D}q \mathcal{F}\left[q(x)\right]\mathcal{P}\left[q(x)\right] = \frac{1}{N_{rep}} \sum_{k=1}^{N_{rep}} \mathcal{F}\left(q^{(net)(k)}(x)\right)$$

Monte Carlo sampling → Compute correlations between pairs of different parton distributions at different points:

Ex. : Correlation of quark pdf $q_f(x)$ and gluon pdf $g(x)$

$$\langle q_f(x_1) g(x_2)\rangle = \frac{1}{N_{\rm rep}} \sum_{k=1}^{N_{\rm rep}} q_f^{(net)(k)}(x_1, Q_0^2) g^{(net)(k)}(x_2, Q_0^2)$$

→ Extremely important for computation of physical processes

Example → Correlation between u and d quark pdfs.

Strategies in PDF global fits

The standard approach:

1.- PDFs parametrized by functional forms $q(x, Q_0^2) = x^{\alpha}(1-x)^{\beta}P(x)$

2.- Uncertainties: representation as ranges in parameters, estimation with different methods (offset, Hessian, Lagrange multiplier,..)

3.- Error propagation (sometimes) in linearized approximation, depends on parametrization.

The NNPDF approach:

1.- PDFs $q(x, Q_0^2)$ parametrized by neural networks. $\longrightarrow$ no bias due to functional form.

2.- Monte Carlo sampling of experimental data $\longrightarrow$ Faithful representation of errors and correlations.

3.- Monte Carlo sampling of experimental data $\longrightarrow$ Exact error propagation.

Parton distribution evolution

PDFs parametrized by a neural network → Mellin inversion of N-space evolution kernel (no complex neural networks):

$$q(N, Q^2) = q(N, Q_0^2)\Gamma\left(N, \alpha_s\left(Q^2\right), \alpha_s\left(Q_0^2\right)\right)$$

$$\Gamma\left(x, \alpha_s\left(Q^2\right), \alpha_s\left(Q_0^2\right)\right) \equiv \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} x^{-N}\Gamma\left(N, \alpha_s\left(Q^2\right), \alpha_s\left(Q_0^2\right)\right)$$

$\Gamma(x)$ is a distribution, diverges at $x = 1$.

Regulating the $\Gamma(x)$ distribution → PDF evolution equation:

$$q(x, Q^2) = q(x, Q_0^2)\int_x^1 \Gamma(y)dy + \int_x^1 \frac{dy}{y}\Gamma(y)\left(q\left(\frac{x}{y}, Q_0^2\right) - yq(x, Q_0^2)\right)$$

Note: the neural network $q(x, Q_0^2)$ must learn a convolution.

Details of PDF evolution (I)

- At higher orders → Wilson coefficients $C(N, \alpha_s\left(Q^2\right))$ through a modified evolution factor

$$\tilde{\Gamma}\left(x, \alpha_s\left(Q^2\right), \alpha_s\left(Q_0^2\right)\right) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} x^{-N} C(N, \alpha_s\left(Q^2\right)) \Gamma\left(N\right)$$

- Mellin transform inversion of evolution factor $\Gamma(N)$ with Fixed Talbot algorithm (very efficient).
- Evolution formalism benchmarked against the Les Houches parton distribution evolution benchmark results, (G. Salam, A. Vogt, hep-ph/0204316)

Details of PDF evolution (II)

- During pdf fitting (neural network training) only $\Gamma(x)$ is required
  Compute $\Gamma(x)$ and its integral

  $$\gamma(x) = \int_0^x dy\, \Gamma(y)$$

  before the fit (hard numerical task) and interpolate them.

- Interpolation of $\Gamma(x)$ non trivial ($\Gamma(x)$ is a distribution, diverges at $x = 1$)
  $\rightarrow$ Benchmark evolution with interpolated $\Gamma(x)$.

- Much faster evolution with interpolated $\Gamma(x)$.

- Resulting formalism: Fast and efficient parton evolution.

The nonsinglet parton distribution

First application of the method:

Determination of the nonsinglet parton distribution $q_{NS}(x,Q_0^2)$ from the NS structure function $F_2^{NS}(x,Q^2)$.

At Leading Order:

$$F_2^{NS}(x,Q^2) \equiv 2\left(F_2^p - F_2^d\right)(x,Q^2) = \frac{x}{6}\left(u+\bar{u}-d-\bar{d}\right)(x,Q^2) \equiv x q_{NS}(x,Q^2)$$

Evolution equations:

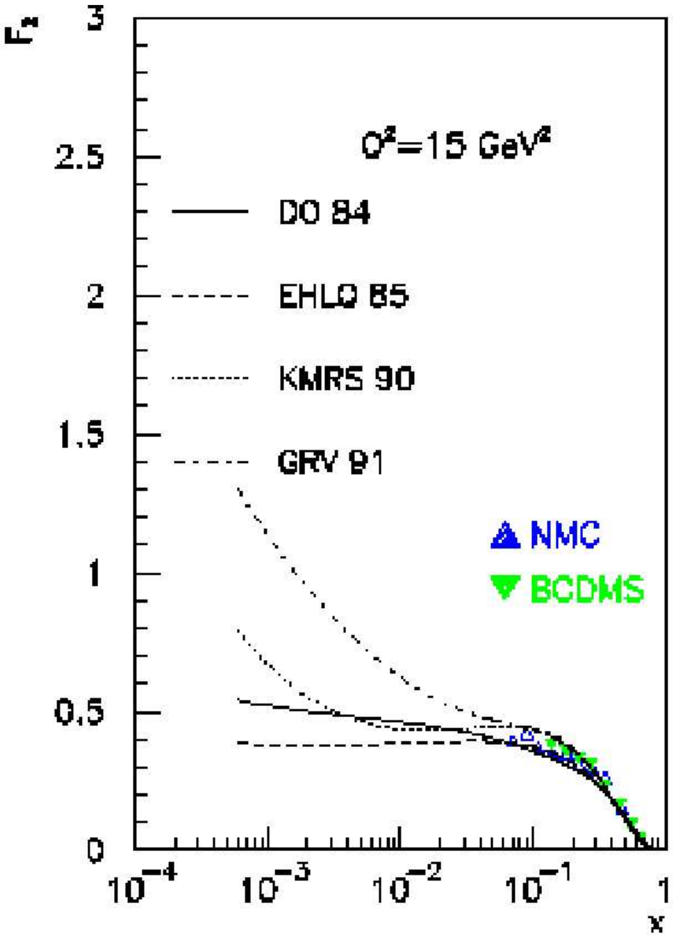
$$q_{NS}(x,Q^2) = q_{NS}(x,Q_0^2) + \int_x^1 \frac{dy}{y}\Gamma(y)\left(q_{NS}\left(\frac{x}{y},Q_0^2\right) - y q_{NS}(x,Q_0^2)\right)$$

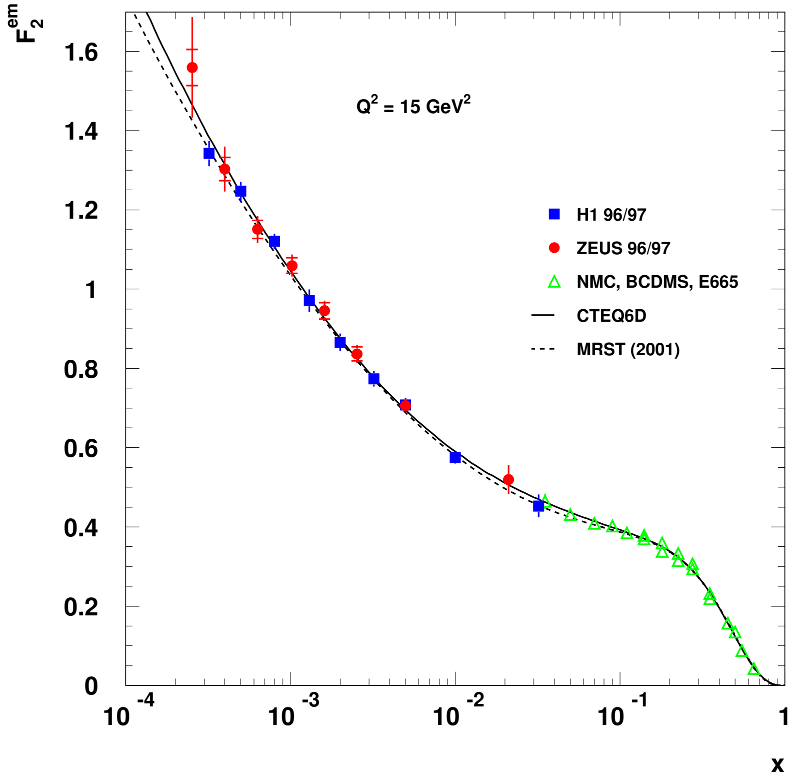
$$-q(x,Q_0^2)\int_0^x q_{NS}(x,Q_0^2)$$

where $q_{NS}(x,Q_0^2)$ is parametrized by a neural network.

The nonsinglet pdf at small x (I)

Historical reminder: $F_2(x, Q^2)$ (singlet) before and after HERA data.[a]





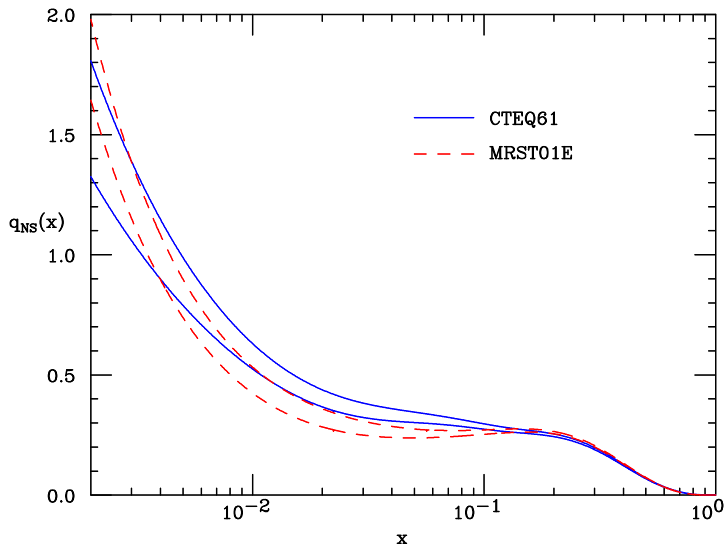
→ Only experimental data could discriminate the correct low x behavior.

[a] Thanks to C. Gwenlan for the plots.

The nonsinglet pdf at small x (II)

Does the nonsinglet parton distribution $q_{NS}(x)$ grow at low x?

- Global fits of parton distributions show a rising $q_{NS}(x)$ at small x.



- Theoretical arguments appear to point in this direction:
  1. Regge theory: $q_{NS}(x) \sim x^{-0.5}$ (A_2 Reggeon)
  2. Low$-x$ resummations: Ex. $q_{NS}(x) \sim x^{-\omega^+/2}$, $\omega^+ = 0.38$, B. Ermolaev et al. ph/0503019

However, experimental data for F_2^{NS} stops at $x \sim 10^{-2}$ and has large errors

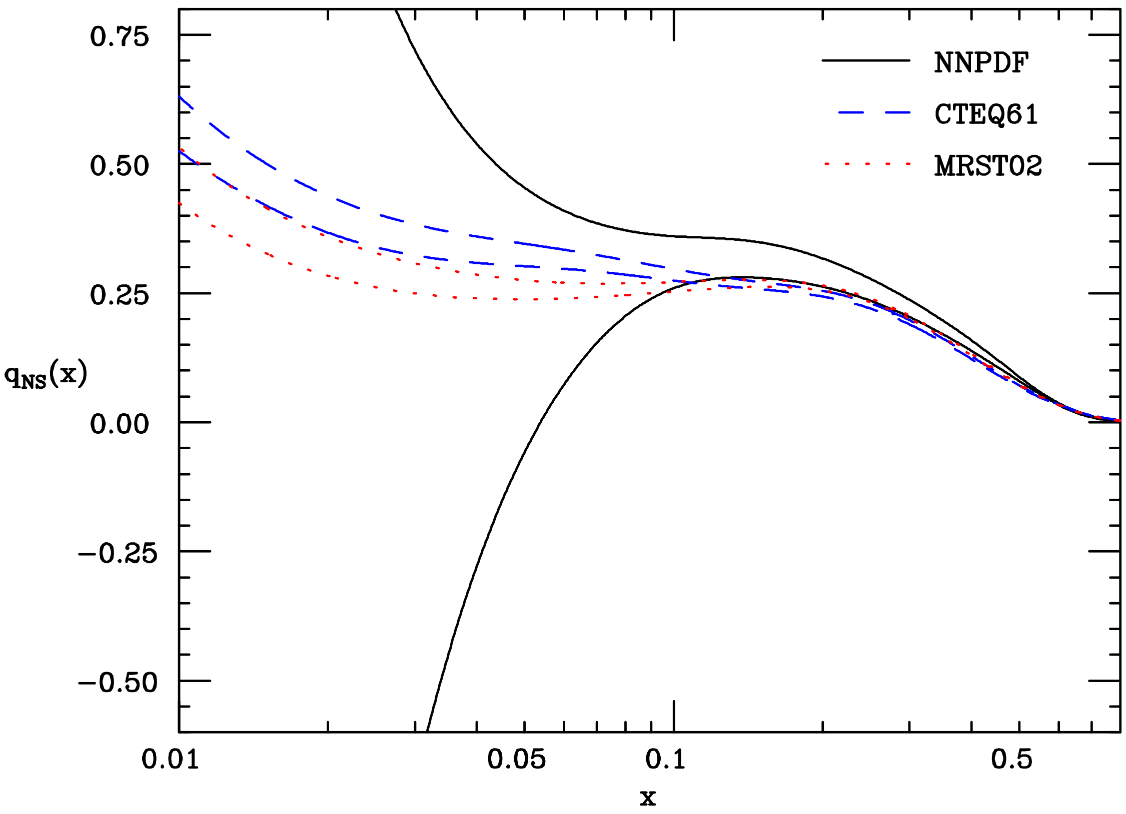
Is the small x growth implied by current data?

Details of the fit

- Experimental data: $F_2^{NS}(x, Q^2)$ from the NMC and BCDMS Collaborations: 347 points.
- Kinematical cuts: $Q^2 \geq 9$ GeV^2, $W^2 \geq 6.25$ GeV^2
- Starting evolution scale: $Q_0^2 = 2$ GeV^2
- Only assumption $\rightarrow$ $q_{NS}(x = 1, Q_0^2) = 0$.
- Perturbative order: NLO
- Neural network architecture: 2-2-2-1.
- Strong coupling $\alpha_s\left(Q^2\right)$ determined from world average value:

$$\alpha_s\left(M_Z^2\right) = 0.118$$

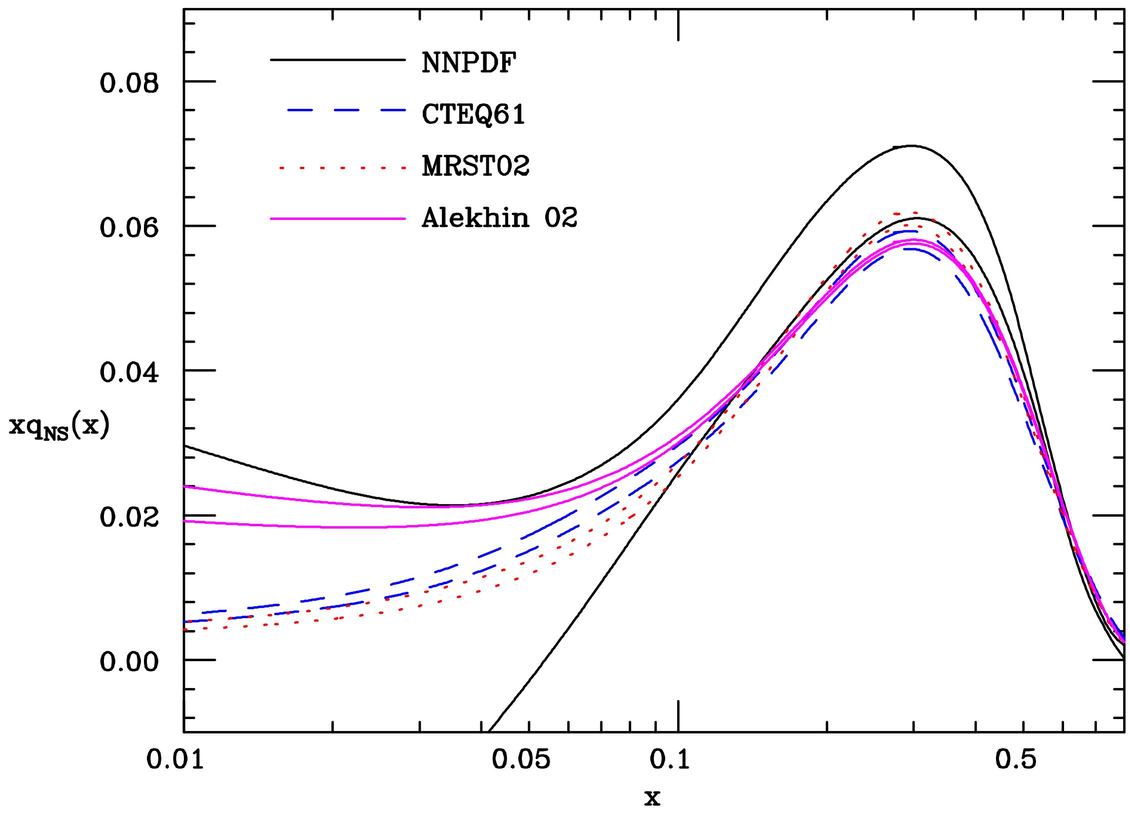
Results: $q_{NS}(x, Q_0^2)$ NNPDF at NLO



Very large uncertainties at small x:

The growth of $q_{NS}(x)$ is **allowed** but **not implied** by (available) experimental data.

Results: $xq_{NS}(x, Q_0^2)$ NNPDF at NLO



Very large uncertainties at small x:

Other $xq_{NS}(x)$ within the NNPDF fit error band.

Summary

- Unbiased determination of structure functions with faithful estimation of uncertainties.
- Successful implementation of neural parton fitting: Determination of nonsinglet parton distribution at NLO with fully correlated uncertainties from $F_2^{NS}(x, Q^2)$.
- The uncertainties at small x are very large: to settle the issue of the low x behavior of the nonsinglet we need additional experimental data.
- Ideal scenario: A run with deuterons at HERA II.
  See discussions of the HERA-LHC workshop,
  http://www.desy.de/~ heralhc

Outlook

- Construct full set of NNPDF parton distributions from all available data.
- Estimate impact of theoretical uncertainties.
- Assess impact of uncertainties of PDFs for relevant observables at LHC.
- Make formalism compatible with standard interface: LHAIPDFv4: NNPDF partons available for use in Monte Carlo generators.