The neural network approach to parton fitting

NNPDF Collaboration

Luigi Del Debbio[1], Stefano Forte[2],
José I. Latorre[3], A. P.[4] and Joan Rojo[3]

[1] Particle Physics Theory Group, School of Physics, University of Edinburgh
[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] Dipartimento di Fisica Teorica, Università di Torino and INFN, Sezione di Torino

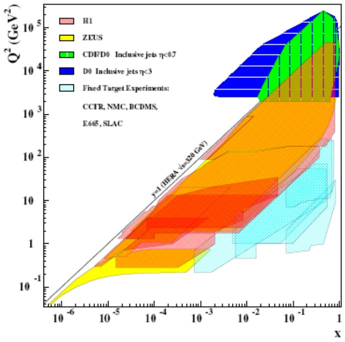
How do we describe hadrons?

- QCD describes interactions between quarks and gluons. Experimentally we observe only hadrons $\rightarrow$ Confinement
- Perturbative QCD is not trustable at low energies ($\sim$ GeV). We can not solve QCD in the non-perturbative region, but on a lattice ...
- We can extract information on the proton structure from a process with only one initial proton (DIS at HERA). Then we can use these as an input for a process where two initial protons are involved (DY at LHC) $\rightarrow$ Factorization

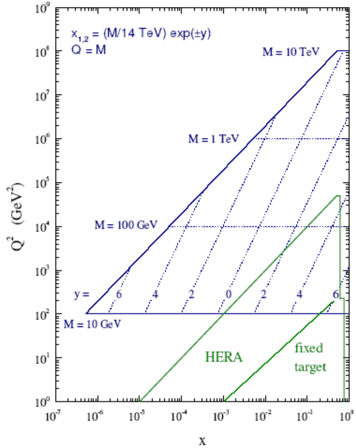
How do we describe hadrons?

- QCD describes interactions between quarks and gluons.
  Experimentally we observe only hadrons → Confinement
- Perturbative QCD is not trustable at low energies ($\sim$ GeV). We can not solve QCD in the non-perturbative region, but on a lattice . . .
- We can extract information on the proton structure from a process with only one initial proton (DIS at HERA).
  Then we can use these as an input for a process where two initial protons are involved (DY at LHC) → Factorization

Kinematics



H1
ZEUS
CDF/D0 Inclusive jets η<0.7
D0 Inclusive jets η<3
Fixed Target Experiments:
CCFR, NMC, BCDMS,
E665, SLAC
y=1 (HERA √s=320 GeV)
Q² (GeV²)
x



x1,2 = (M/14 TeV) exp(±y)
Q = M
M = 10 TeV
M = 1 TeV
M = 100 GeV
M = 10 GeV
y = 6 4 2 0 2 4 6
HERA
fixed
target
Q² (GeV²)
x

Deep Inelastic Scattering

- The cross section

$$\frac{d^2\sigma}{dxdQ^2} = \frac{4\pi\alpha^2}{Q^4}\left[[1+(1-y)^2]F_1 + \frac{1-y}{x}(F_2 - 2xF_1)\right]$$

- The structure function

$$F_2(x,\ Q^2) = x\left[\sum_{q=1}^{n_f} e_q^2\, C^q \otimes q_q(x,\ Q^2) + 2n_f\, C^g \otimes g(x,\ Q^2)\right]$$

- Parton distribution evlution is described by DGLAP equations

$$Q^2\frac{d}{dQ^2}q(x,Q^2) = \frac{\alpha_s(Q^2)}{2\pi}(P \otimes q)(x,Q^2)$$

Deep Inelastic Scattering and QCD

- The cross section

$$\frac{d^2\sigma}{dxdQ^2} = \frac{4\pi\alpha^2}{Q^4}\left[[1+(1-y)^2]F_1 + \frac{1-y}{x}(F_2 - 2xF_1)\right]$$

- The structure function

$$F_2(x,\ Q^2) = x\left[\sum_{q=1}^{n_f} e_q^2\,\mathcal{C}^q \otimes q_q(x,Q^2) + 2n_f\,\mathcal{C}^g \otimes g(x,Q^2)\right]$$

- Parton distribution evlution is described by DGLAP equations

$$Q^2\frac{d}{dQ^2}q(x,Q^2) = \frac{\alpha_s(Q^2)}{2\pi}(P\otimes q)(x,Q^2)$$

Deep Inelastic Scattering and QCD

- The cross section

$$\frac{d^2\sigma}{dxdQ^2} = \frac{4\pi\alpha^2}{Q^4}\left[[1+(1-y)^2]F_1 + \frac{1-y}{x}(F_2 - 2xF_1)\right]$$

- The structure function

$$F_2(x,\ Q^2) = x\left[\sum_{q=1}^{n_f} e_q^2\, \mathcal{C}^q \otimes q_q(x,\ Q^2) + 2n_f\, \mathcal{C}^g \otimes g(x,\ Q^2)\right]$$

- Parton distribution evlution is described by DGLAP equations

$$Q^2 \frac{d}{dQ^2} q(x,\ Q^2) = \frac{\alpha_s(Q^2)}{2\pi}(P \otimes q)(x,\ Q^2)$$

The problem

- For a single quantity → 1 sigma error
- For a pair of numbers → 1 sigma ellipse
- For a function → We 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 → Functional integrals

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

Determine an infinite-dimensional object (a function) from finite set of data points → Mathematically ill-posed problem

The standard approach

1. Choose a simple functional form with enough free parameters

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

2. Fit parameters by minimizing χ^2

Open problems:

- Errors combination and propagation from data to parameters and from parameters to observables is not trivial
- Theoretical bias due to the choice of a parametrization is difficult to assess (effects can be large if data are not precise or hardly compatible)
- NNLO *vs.* NLO+Resummations

The standard approach

1. Choose a simple functional form with enough free parameters

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

2. Fit parameters by minimizing χ^2

Open problems:

- Errors combination and propagation from data to parameters and from parameters to observables is not trivial
- Theoretical bias due to the choice of a parametrization is difficult to assess (effects can be large if data are not precise or hardly compatible)
- NNLO *vs.* NLO+Resummations

The standard approach

1. Choose a simple functional form with enough free parameters

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

2. Fit parameters by minimizing χ^2

Open problems:

- Errors combination and propagation from data to parameters and from parameters to observables is not trivial
- Theoretical bias due to the choice of a parametrization is difficult to assess (effects can be large if data are not precise or hardly compatible)
- NNLO *vs.* NLO+Resummations

The standard approach

1. Choose a simple functional form with enough free parameters

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

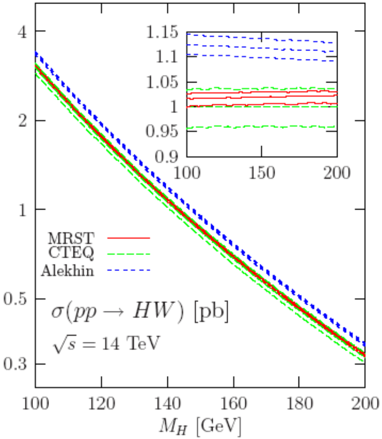
2. Fit parameters by minimizing χ^2

Open problems:

- Errors combination and propagation from data to parameters and from parameters to observables is not trivial
- Theoretical bias due to the choice of a parametrization is difficult to assess (effects can be large if data are not precise or hardly compatible)
- NNLO *vs.* NLO+Resummations

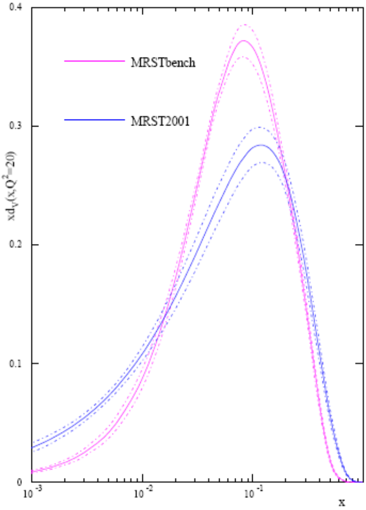
The standard approach - Limitations

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

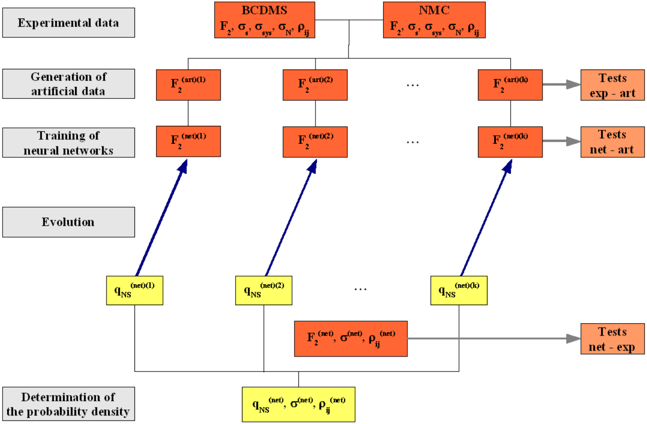


The standard approach - Limitations

[R. S. Thorne, hep-ph/0511119]

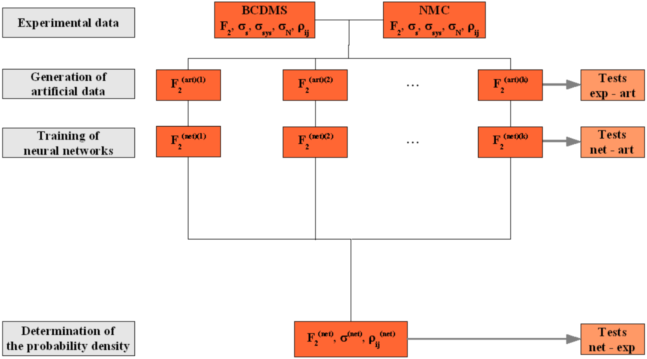


The NNPDF approach



The NNPDF approach

[S. Forte et al., hep-ph/0204232 - A. P., hep-ph/0207204 - L. Del Debbio et al., hep-ph/0501067]



Faithful error propagation: Data → Parametrization

- Monte Carlo sampling of data (generation of replicas of experimental data)

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

where σ_i are the experimantal errors, and r_i are random numbers choosen accordingly to the experimental correlation matrix.

Faithful error propagation: Parametrization → Observables

- Expectation values:

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

- Errors:

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

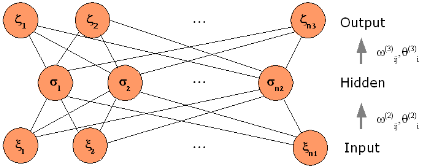
- Correlations between pairs of different parton distributions at different points:

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

Unbiased parametrization

- A neural network is trained over each MC replica
- Neural networks are a class of algorithms very suitable to fit incomplete or noisy data [for HEP applications see ACAT 2005]
- Any continuous function can be uniformly approximated by a continuous neural network having only one internal layer, and with an arbitrary continuous sigmoid non-linearity [G. Cybenko (1989)]

Unbiased parametrization



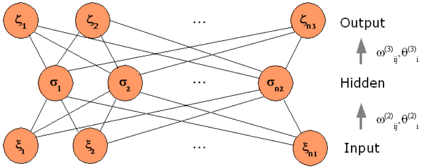
- Activation function:

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

- As an example, in a very simple case (1-2-1) we have

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

Unbiased parametrization



- Activation function:

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

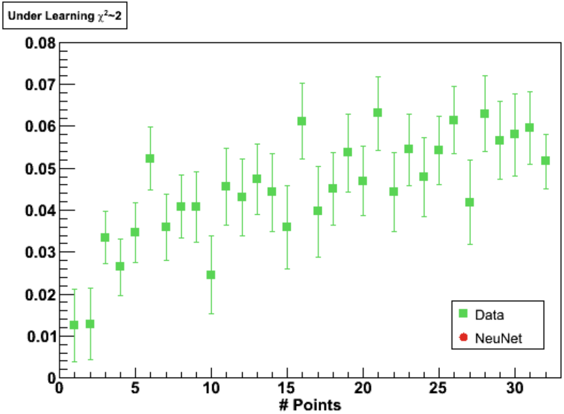
- As an example, in a very simple case (1-2-1) we have

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

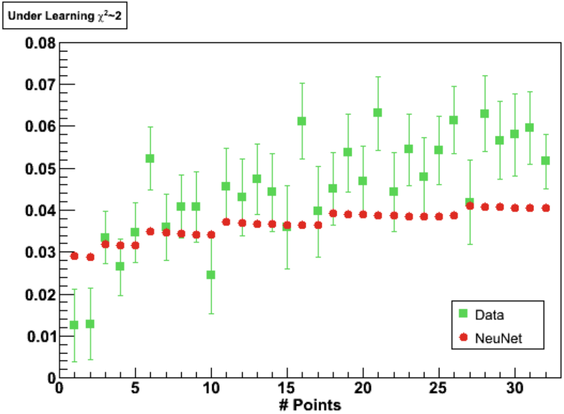
Minimization with a Genetic Algorithm

1. Set the parameters randomly.
2. Make clones of the set of parameters.
3. Mutate randomly each clone.
4. Evaluate χ^2 for all the clones.
5. Select clones with the lowest χ^2.
6. Back to 2, till $\chi^2 \sim \bar{\chi}^2$.

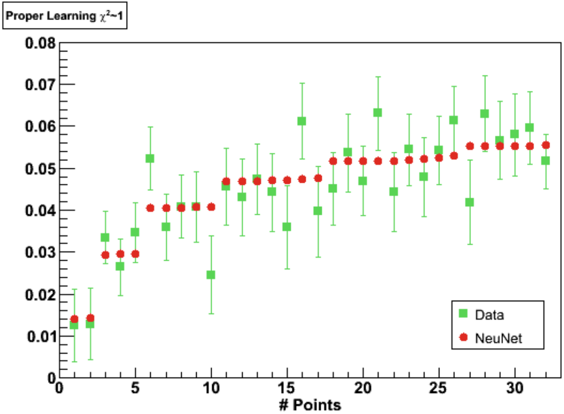
Learning of data



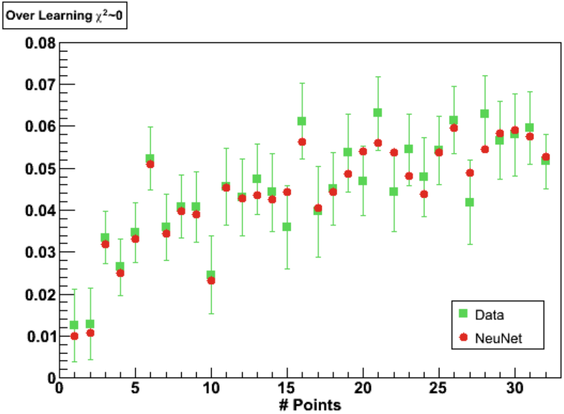
Learning of data



Learning of data

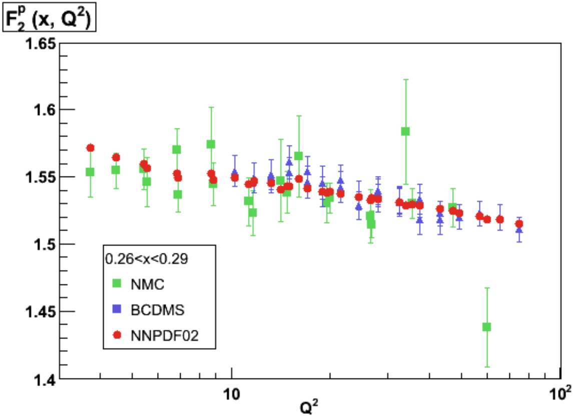


Learning of data



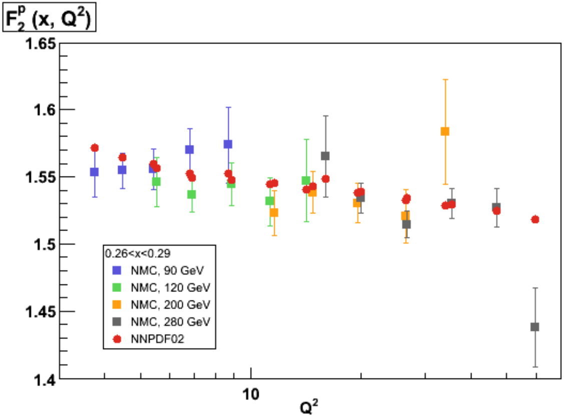
Incompatible data

[S. Forte et al., hep-ph/0204232 - A. P., hep-ph/0207204]



Incompatible data

[S. Forte et al., hep-ph/0204232 - A. P., hep-ph/0207204]



A new framework

- We want Mellin space evolution:

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

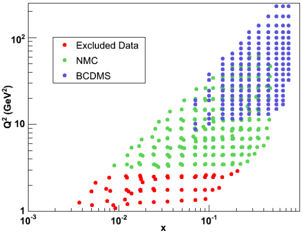
- We do not want complex neural networks:

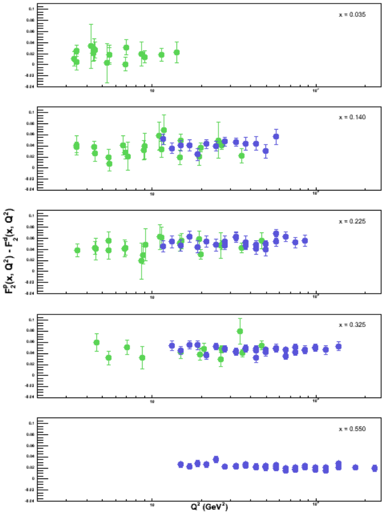
$$\Gamma\left(x, \alpha_s\left(Q^2\right), \alpha_s\left(Q_0^2\right)\right) \equiv \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} dN\, x^{-N}\Gamma\left(N, \alpha_s\left(Q^2\right), \alpha_s\left(Q_0^2\right)\right)$$

- The evolved PDF is given by

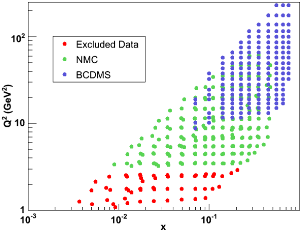
$$q(x, Q^2) = \int_x^1 \frac{dy}{y}\Gamma\left(y, \alpha_s\left(Q^2\right), \alpha_s\left(Q_0^2\right)\right) q\left(\frac{x}{y}, Q_0^2\right)$$

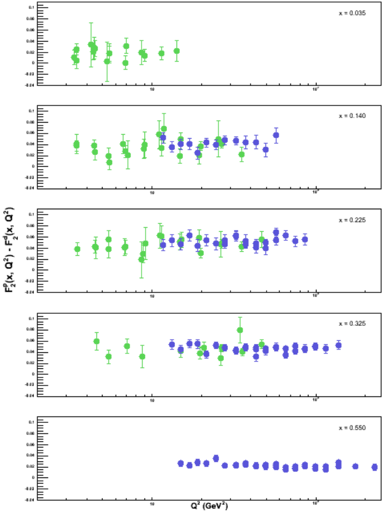
Some details





Some details





- Neural network architecture: 2-5-3-1 (37 parameters).
- $x\, q(x, Q_0^2) = NN(x, \log x)(1-x)^a$
  $q(x, Q_0^2) = (u + \bar{u} - d - \bar{d})(x, Q_0^2)$
  $Q_0^2 = 2\ GeV^2$
- TMC: Georgi-Politzer,
  F_2 integral evaluated with NN F_2
- ZM-VFN:
  $m_c = 1.4 GeV$
  $m_b = 4.5 GeV$
  $m_t = 175 GeV$

Delivery

- # MC reps: 1000
- Strong coupling: $\alpha_s\left(M_Z^2\right) = 0.118 \pm 0.002$
- Perturbative order: LO, NLO, NNLO
- LHAPDF interface
- With $\alpha_s = 0.118$ @ NLO we have:

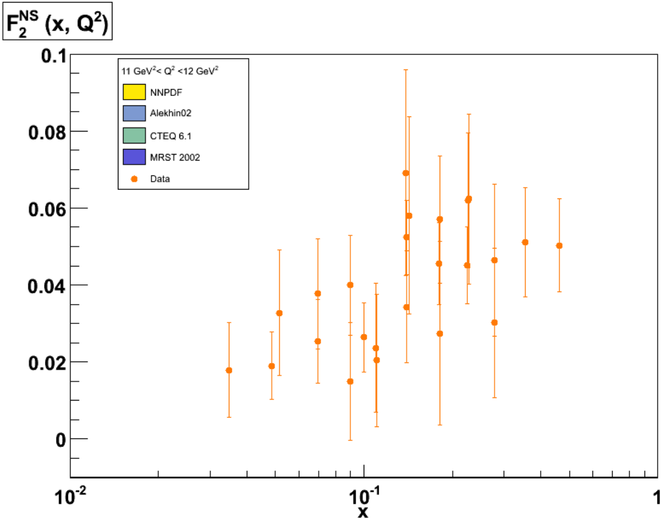
	Total	NMC	BCDMS
χ^2/d.o.f.	0.95	0.92	0.97

Delivery

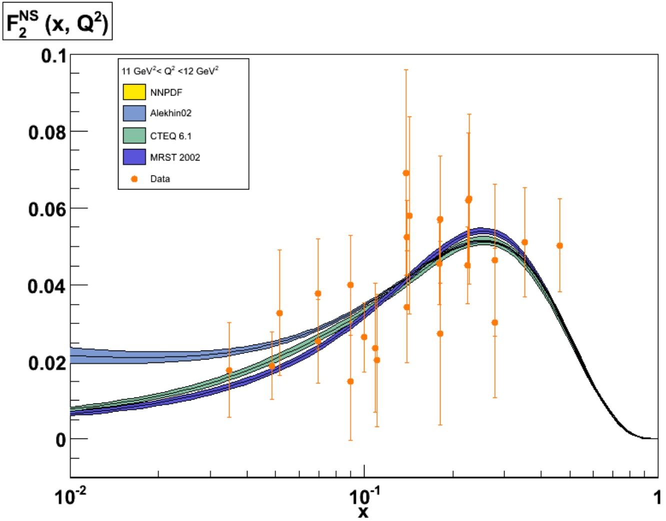
- # MC reps: 1000
- Strong coupling: $\alpha_s\left(M_Z^2\right) = 0.118 \pm 0.002$
- Perturbative order: LO, NLO, NNLO
- LHAPDF interface
- With $\alpha_s = 0.118$ @ NLO we have:

	Total	NMC	BCDMS
$\chi^2/\mathrm{d.o.f.}$	0.95	0.92	0.97

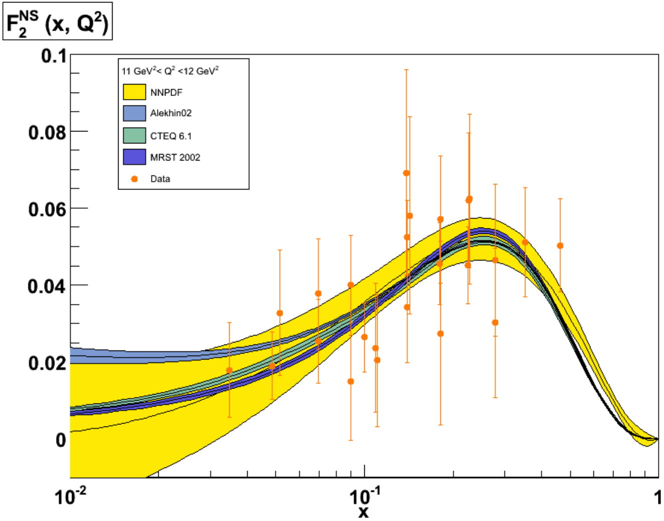
Reconstructing F_2^{NS} @ NLO with errors



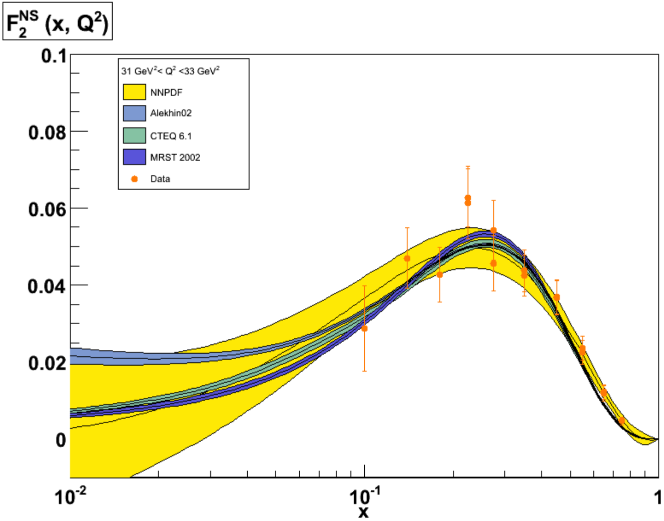
Reconstructing F_2^{NS} @ NLO with errors



Reconstructing F_2^{NS} @ NLO with errors

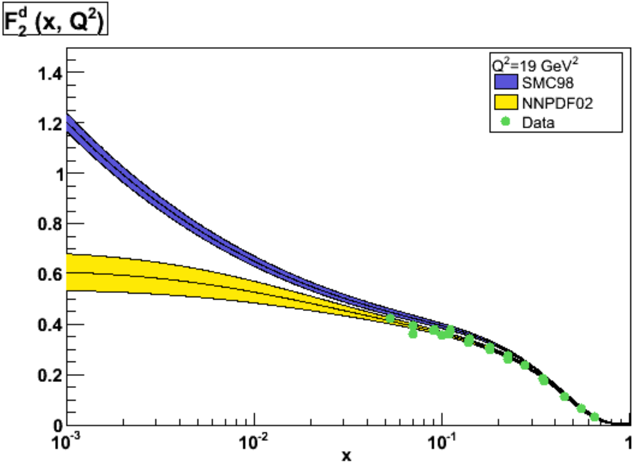


Reconstructing F_2^{NS} @ NLO with errors



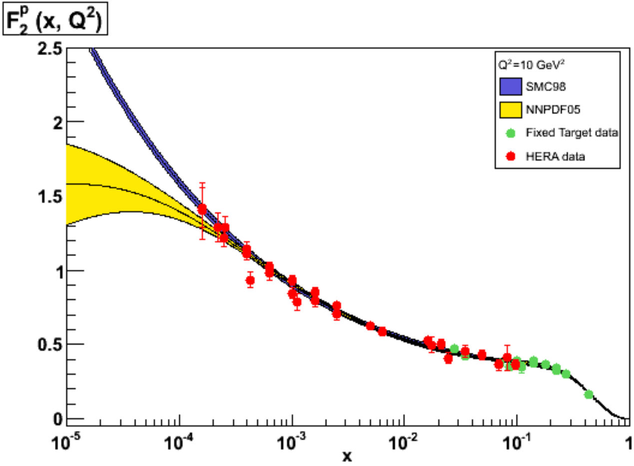
Fit of $F_2^d(x, Q^2)$

[S. Forte et al., hep-ph/0204232 - A. P., hep-ph/0207204]



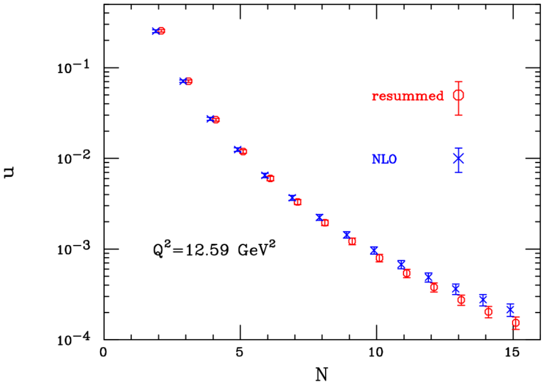
Fit of $F_2^p(x, Q^2)$

[L. Del Debbio et al., hep-ph/0501067]



Resummations

[G. Corcella and L. Magnea, hep-ph/0506278]



$$q^{NLO}(N, Q^2) = \frac{F(N-1, Q^2)}{C^{NLO}(N, \alpha_s(Q^2))} \qquad q^{res}(N, Q^2) = \frac{F(N-1, Q^2)}{C^{res}(N, \alpha_s(Q^2))}$$

Re-evaluation of the Gottfried sum rule

[R. Abbate and S. Forte, hep-ph/0511231]

- NMC:

$$S_G(0.004 < x < 0.8,\ 4\ \mathrm{GeV}^2) = 0.2281 \pm 0.0201$$

- NNPDF:

$$S_G(0.004 < x < 0.8,\ 4\ \mathrm{GeV}^2) = 0.2281 \pm 0.0437$$

- The two estimations perfectly agree for all $x_{min} < x < 0.8$ ranges, but the for the smallest $x_{min} = 0.004$.
- NMC uncertainty at the boundary of the measured region is evaluated assuming that the error is linear across the bins, and this results in an underestimation of the error on the last bin.
- The inclusion of the (assumed/unknown) small-x contribution yields

$$S_G(1.5\ \mathrm{GeV}^2 < Q^2 < 4.5\ \mathrm{GeV}^2) = 0.244 \pm 0.045$$

Re-evaluation of the Gottfried sum rule

[R. Abbate and S. Forte, hep-ph/0511231]

- NMC:
$$S_G(0.004 < x < 0.8,\ 4\ \mathrm{GeV}^2) = 0.2281 \pm 0.0201$$
- NNPDF:
$$S_G(0.004 < x < 0.8,\ 4\ \mathrm{GeV}^2) = 0.2281 \pm 0.0437$$
- The two estimations perfectly agree for all $x_{min} < x < 0.8$ ranges, but the for the smallest $x_{min} = 0.004$.
- NMC uncertainty at the boundary of the measured region is evaluated assuming that the error is linear across the bins, and this results in an underestimation of the error on the last bin.
- The inclusion of the (assumed/unknown) small-x contribution yields

$$S_G(1.5\ \mathrm{GeV}^2 < \mathrm{Q}^2 < 4.5\ \mathrm{GeV}^2) = 0.244 \pm 0.045$$

Results

We have developed a tool to fit data that

- provides a faithful combination of experimental errors;
- allows a faithful propagation of errors on computed observables;
- handles incompatibilities among experiments without assumptions;
- avoids theoretical biases on the used parametrization.

This approach is general and can be applied to different problems:

- Parton Distribution Functions;
- nucleon Structure Functions;
- b-meson Shape Function (?);
- any other idea? Let's try ...

Results

We have developed a tool to fit data that

- provides a faithful combination of experimental errors;
- allows a faithful propagation of errors on computed observables;
- handles incompatibilities among experiments without assumptions;
- avoids theoretical biases on the used parametrization.

This approach is general and can be applied to different problems:

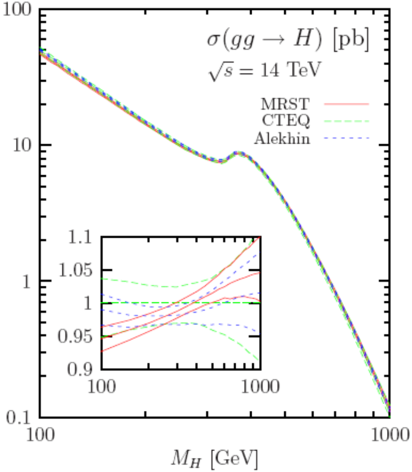
- Parton Distribution Functions;
- nucleon Structure Functions;
- b-meson Shape Function (?);
- any other idea? Let's try ...

Perspectives

- Go ahead ...
- ... a singlet set from DIS data (December 2006?)
- ... a singlet set from DIS+DY data (April 2007?)

The standard approach - Limitations

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



The standard approach

- MRST: 15 parms. - $\Delta\chi^2 = 50$ - NC and CC DIS, DY, W-asym, jets

$$xq(x, Q_0^2) = A(1-x)^{\eta}(1+\epsilon x^{0.5}+\gamma x)x^{\delta}, \quad x[\bar{u}-\bar{d}](x, Q_0^2) = A(1-x)^{\eta}(1+\gamma x+\delta x^2)x^{\delta}.$$

$$xg(x, Q_0^2) = A_g(1-x)^{\eta_g}(1+\epsilon_g x^{0.5}+\gamma_g x)x^{\delta_g} - A_-(1-x)^{\eta_-}x^{-\delta_-},$$

- CTEQ: 20 parms. - $\Delta\chi^2 = 100$ - NC and CC DIS, DY, W-asym, jets

$$x\,f(x, Q_0) = A_0\,x^{A_1}\,(1-x)^{A_2}\,e^{A_3 x}\,(1+e^{A_4}\,x)^{A_5}$$

  with independent params for combinations $u_v \equiv u-\bar{u}$, $d_v \equiv d-\bar{d}$, g, and $\bar{u}+\bar{d}$, $s=\bar{s}=0.2\,(\bar{u}+\bar{d})$ at Q_0; norm. fixed by sum rules

- Alekhin: 17 parms. - $\Delta\chi^2 = 1$ - NC DIS (+ DY)

$$xu_V(x, Q_0) = \frac{2}{N_u^V}x^{a_u}(1-x)^{b_u}(1+\gamma_2^u x); \qquad xu_S(x, Q_0) = \frac{A_S}{N_S}\eta_u x^{a_s}(1-x)^{b_{su}}$$

$$xd_V(x, Q_0) = \frac{1}{N_d^V}x^{a_d}(1-x)^{b_d}; \qquad xd_S(x, Q_0) = \frac{A_S}{N^S}x^{a_s}(1-x)^{b_{sd}},$$

$$xs_S(x, Q_0) = \frac{A_S}{N^S}\eta_s x^{a_s}(1-x)^{(b_{su}+b_{sd})/2}; \qquad xG(x, Q_0) = A_G x^{a_G}(1-x)^{b_G}(1+\gamma_1^G\sqrt{x}+\gamma_2^G x),$$

SF Details

Architecture: 4-5-3-1

- Inputs: x, $\log x$, Q^2, $\log Q^2$
- Output: $F_2(x, Q^2)$

Minimization strategy:

- Back Propagation ($\sim 10^8$ training cycles):

$$\chi^{2\,(k)}_{\mathrm{diag}} = \frac{1}{N_{\mathrm{dat}}} \sum_{i=1}^{N_{\mathrm{dat}}} \frac{\left(F_i^{(\mathrm{art})(k)} - F_i^{(\mathrm{net})(k)}\right)^2}{\sigma_{i,t}^{(\mathrm{exp})2}}$$

- Genetic Algorithm ($\sim 10^4$ generations):

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

Mellin Inversion with the Fixed Talbot algorithm

$$f(t) = \frac{1}{2\pi i}\int_C ds\ e^{ts}\tilde{f}(s), \quad t = -\ln x$$

$$s(\theta) = r\theta\left(\cot\theta + i\right), \quad -\pi \leq \theta \leq \pi$$

$$f(t) = \frac{r}{\pi}\int_0^{\pi} d\theta\ Re\left[\exp(ts(\theta))\tilde{f}(s(\theta))(1 + i\sigma(\theta))\right]$$

$$\sigma(\theta) = \theta + (\theta\cot\theta - 1)cot\theta$$

$$f(t, M) = \frac{r}{M}\left[\frac{1}{2}\tilde{f}(r)e^{rt} + \sum_{k=1}^{M-1} Re\left[\exp(ts(\theta_k))\tilde{f}(s(\theta_k))(1 + i\sigma(\theta_k))\right]\right]$$

$$r = \frac{2M}{5t}, \qquad \theta_k = \frac{k\pi}{M}$$