Using Monte Carlo methods and Neural Networks for the determination of Parton Distribution Functions

Alberto Guffanti

Albert-Ludwigs-Universität Freiburg

On behalf of the NNPDF Collaboration:

R. D. Ball, L. Del Debbio, M. Ubiali (Edinburgh), S. Forte, A. Piccione, J. Rojo (Milano), J. I. Latorre (Barcelona)

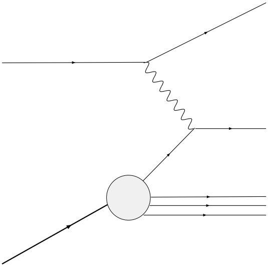
CAUTION

This talk contains (mostly) unpolarized physics

INTRODUCTION

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?

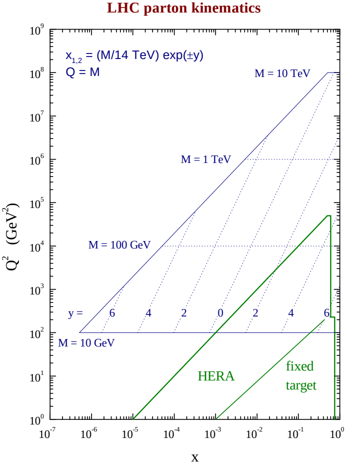
- The initial condition cannot be computed in Perturbation Theory (Lattice? In principle yes, but ...)
- The evolution with the energy scale is given by Altarelli-Parisi evolution equations (or DGLAP or renormalization group equations for mass factorization)

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

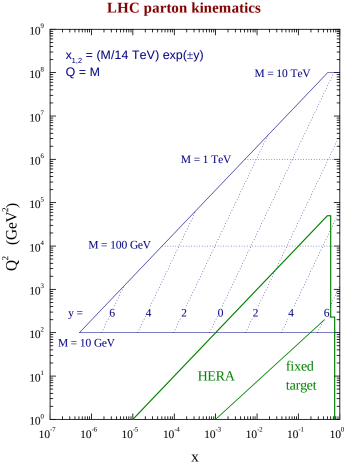
where $\otimes$ denotes the Mellin convolution

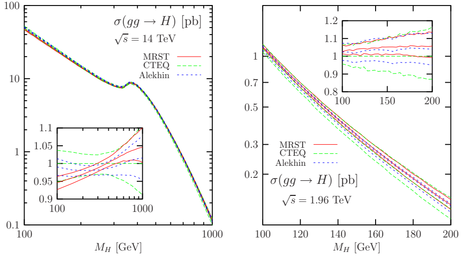
- The splitting functions P can be computed in PT and are known up to 3-loop (NNLO)

Should we care about PDFs (and their uncertainties)?



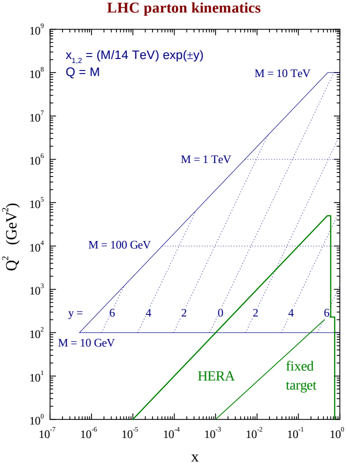
Should we care about PDFs (and their uncertainties)?

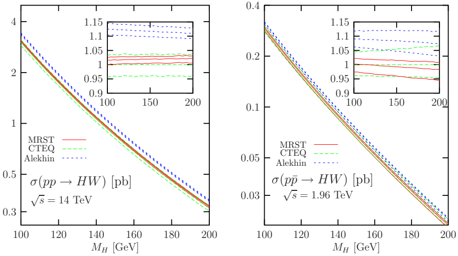




[A. Djouadi and S. Ferrag, hep-ph/0310209]

Should we care about PDFs (and their uncertainties)?





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

Should we care about PDFs (and their uncertainties)?

- Errors on PDFs are in some cases the dominating theoretical error on precision observables

Ex. $\sigma(Z^0)$ at the LHC: $\delta_{PDF} \sim 3\%$, $\delta_{NNLO} \sim 2\%$

[J. Campbell, J. Huston and J. Stirling, (2007)]

Should we care about PDFs (and their uncertainties)?

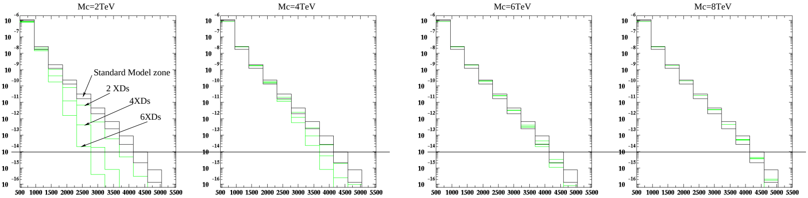
- Errors on PDFs are in some cases the dominating theoretical error on precision observables

Ex. $\sigma(Z^0)$ at the LHC: $\delta_{PDF} \sim 3\%$, $\delta_{NNLO} \sim 2\%$

[J. Campbell, J. Huston and J. Stirling, (2007)]

- Errors on PDFs might reduce sensitivity to New Physics

Ex. Extra Dimensions discovery in dijet cross section at the LHC:



[S. Ferrag (ATLAS), hep-ph/0407303]

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 an infinite-dimensional object (a function) from a finite set of data points ... mathematically ill-defined problem.

Solution

Standard Approach

- Introduce a simple functional form with enough free parameters

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

- Fit parameters minimizing χ^2.

Solution

Standard Approach

- Introduce a simple functional form with enough free parameters

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

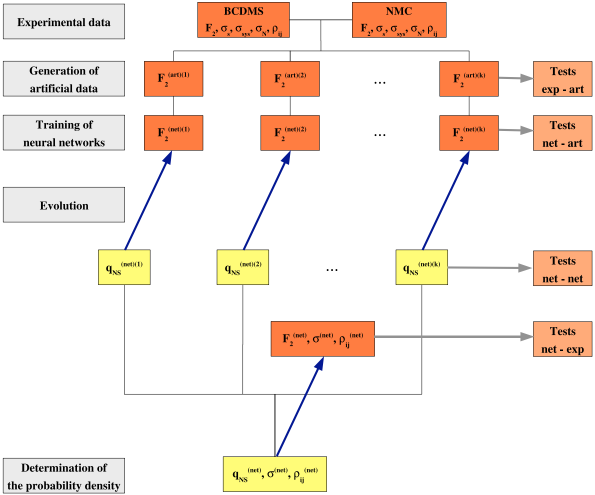
- Fit parameters minimizing χ^2.

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.

THE NNPDF METHODOLOGY

The NNPDF methodology



The Neural Network Approach in a Nutshell

- Generate N_{rep} Monte-Carlo replicas of the experimental data.
- Fit a set of Parton distribution functions on each replica, thus defining a sampling of probability density on the space of the 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.

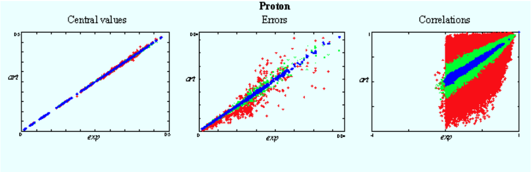
Monte Carlo replicas generation

- 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

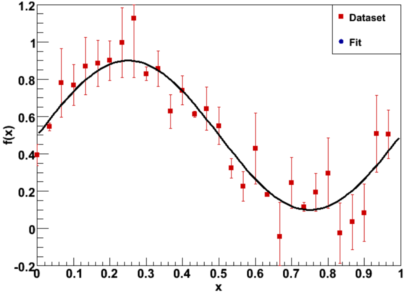
- Validate Monte Carlo replicas against experimental data
(statistical estimators, faithful representation of errors, convergence rate increasing N_{rep})



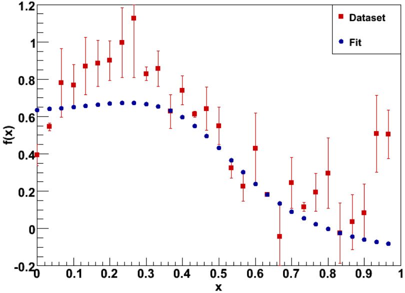
- $\mathcal{O}(1000)$ replicas needed to reproduce correlations to percent accuracy

Proper Fitting avoiding Overlearning

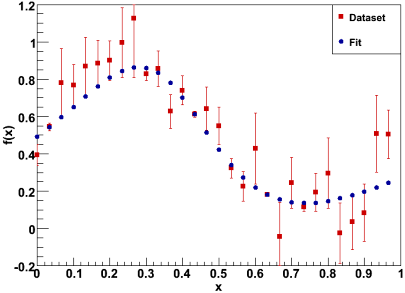
- Let's see how proper fitting works in a toy model



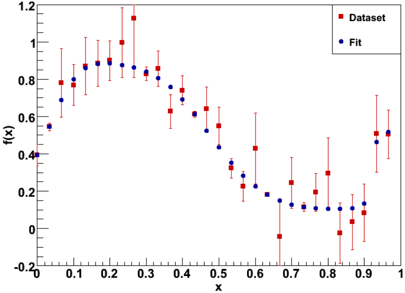
Proper Fitting avoiding Overlearning



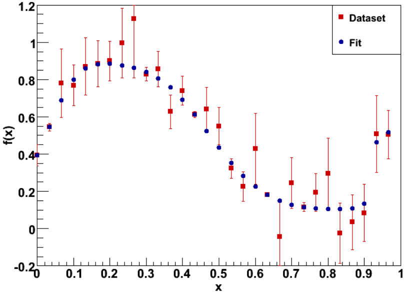
Proper Fitting avoiding Overlearning



Proper Fitting avoiding Overlearning



Proper Fitting avoiding Overlearning



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

How to avoid Overlearning

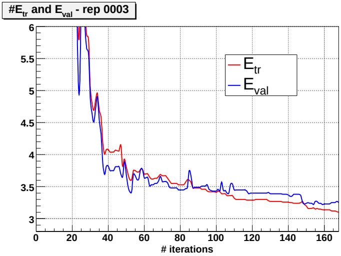
Stopping criterion based on Training-Validation separation

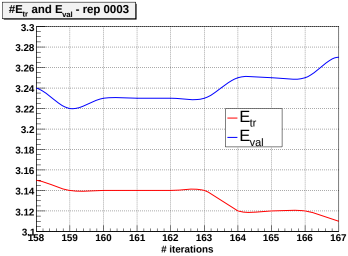
- Divide the data in two sets: Training and Validation
- Minimize the χ^2 of the data in the Training set
- Compute the χ^2 for the data in the Validation set
- When validation χ^2 stops decreasing, **STOP** the fit

How to avoid Overlearning

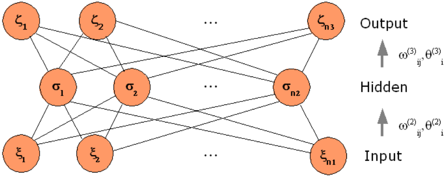
Stopping criterion based on Training-Validation separation

- 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





Why use Neural Networks?



- Neural Networks are non-linear statistical tools.
- Any continuous function can be approximated with neural network with one internal layer and non-linear neuron activation function.
- Efficient minimization algorithms for complex parameter spaces.
- They provide a parametrization which is redundant and robust against variations.

Neural Networks

... just another basis of functions

Multilayer feed-forward networks

- Each neuron receives input from neurons in preceding layer and feeds output to neurons in subsequent layer
- Activation determined by weights and thresholds

$$\xi_i = g\left(\sum_j \omega_{ij}\xi_j - \theta_i\right)$$

- Sigmoid activation function

$$g(x) = \frac{1}{1 + e^{-\beta x}}$$

Neural Networks

... just another basis of functions

Multilayer feed-forward networks

- Each neuron receives input from neurons in preceding layer and feeds output to neurons in subsequent layer

- Activation determined by weights and thresholds

$$\xi_i = g\left(\sum_j \omega_{ij}\xi_j - \theta_i\right)$$

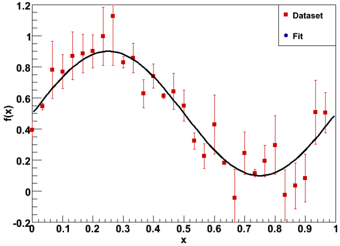
- Sigmoid activation function

$$g(x) = \frac{1}{1 + e^{-\beta x}}$$

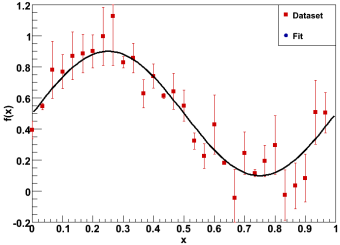
A 1-2-1 NN:

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

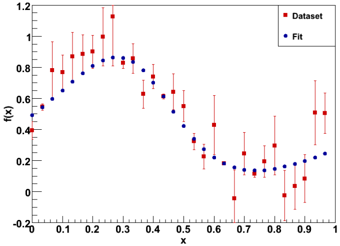
Neural Network vs. Polynomial form



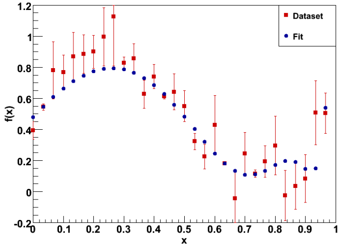
Neural Network vs. Polynomial form



Neural Net Fit $(\chi^2 = 1)$

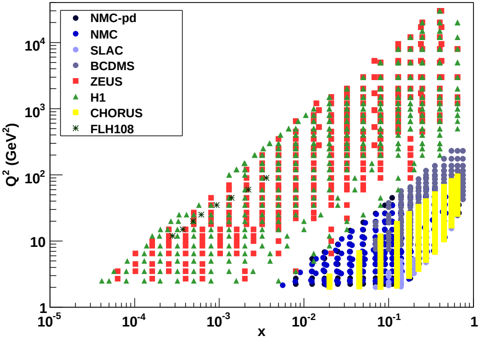


Polynomial form Fit $(\chi^2 = 1)$



RESULTS

NNPDF1.0: Experimental 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	σ_ν , $\sigma_{\bar{\nu}}$	CHORUS
F_2^d / F_2^p	NMC-pd	F_L	H1

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

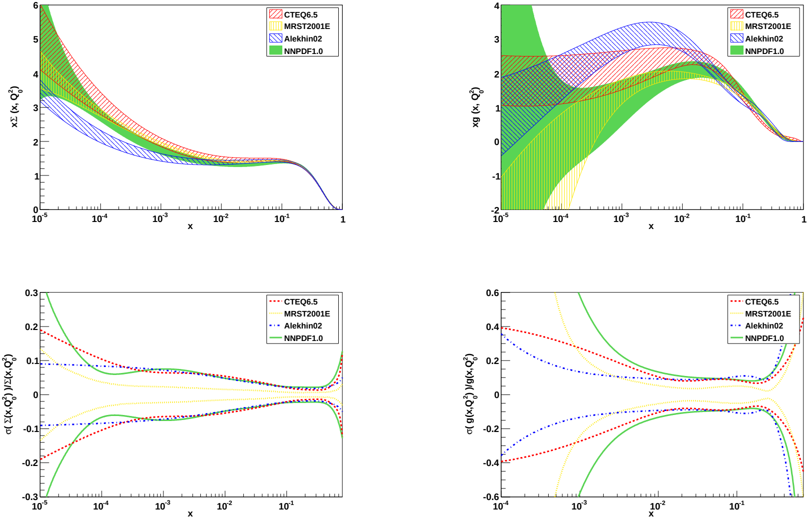
NNPDF1.0: Parametrization

Parametrization of 5 combinations of PDFs at $Q_0^2 = 2$ 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

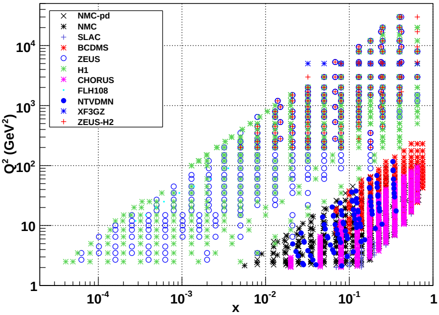
NNPDF1.0 Results



NNPDF1.2: Constraining the strange distribution

- Determination of both s and $\bar{s}$ allowed by inclusion of NuTeV dimuon data

$$\frac{1}{E_\nu}\frac{d^2\sigma^{\nu(\bar{\nu}),2\mu}}{dx\,dy}(x,y,Q^2) \equiv \frac{1}{E_\nu}\frac{d^2\sigma^{\nu(\bar{\nu}),c}}{dx\,dy}(x,y,Q^2)\cdot\langle \mathrm{Br}\,(D\to\mu)\rangle\cdot\mathcal{A}\,(x,y,E_\nu)$$

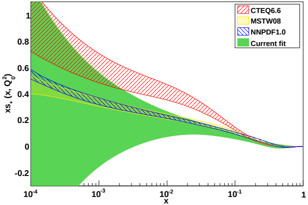


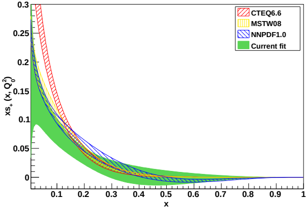
$$\begin{aligned}
\tilde{\sigma}^{\nu(\bar{\nu}),c} &\propto \left(F_2^{\nu(\bar{\nu}),c},\,F_3^{\nu(\bar{\nu}),c},\,F_L^{\nu(\bar{\nu}),c}\right)\\
F_2^{\nu,c} &= x\left[C_{2,q}\otimes 2|V_{cs}|^2 s+\frac{1}{n_f}C_{2,g}\otimes g\right]\\
F_2^{\bar{\nu},c} &= x\left[C_{2,q}\otimes 2|V_{cs}|^2 \bar{s}+\frac{1}{n_f}C_{2,g}\otimes g\right]
\end{aligned}$$

- Neutrino and anti-neutrino dimuon production from NuTeV.
- HERA-II ZEUS data on NC and CC reduced xsec at large-Q^2.

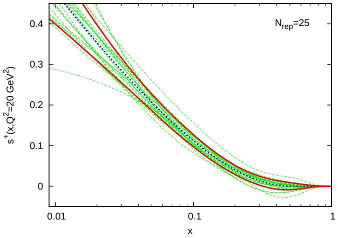
NNPDF 1.2 Results

Total strangeness determination



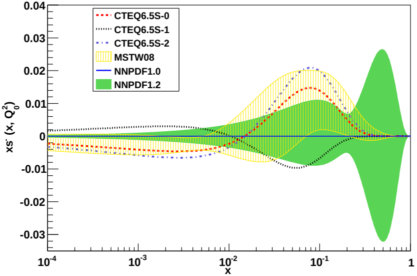


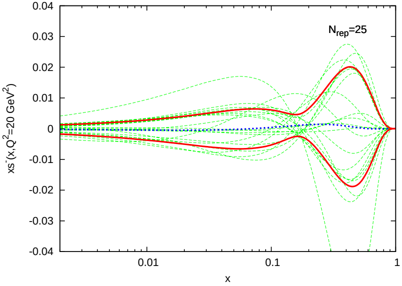
- Data region
  $\rightarrow$ Moderate uncertainties, larger than CTEQ6.6/MSTW08
- Extrapolation region
  $\rightarrow$ Blow-up of uncertainties due to lack of experimental constraints
- Difference with NNPDF1.0 is a signal of parametrization bias



NNPDF 1.2 Results

Strange Asymmetry determination: $s^-(x, Q^2)$





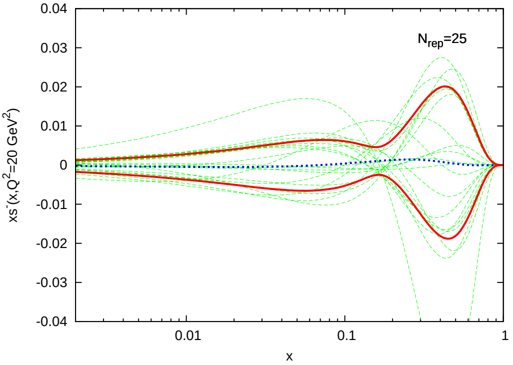
Analysis	$[S^-](Q^2 = 20\,\mathrm{GeV}^2)\cdot 10^3$
NNPDF1.2	0 ± 10
MSTW08	1.4 ± 1.2
CTEQ6.5s	1.2 ± 1.1
AKP08	1.0 ± 1.3
NuTeV07	1.3 ± 0.8

NNPDF uncertainty on $[S^-]$ is large enough to explain the NuTeV (non-)Anomaly.

NNPDF 1.2 Results

Strange Asymmetry determination: $s^-(x, Q^2)$

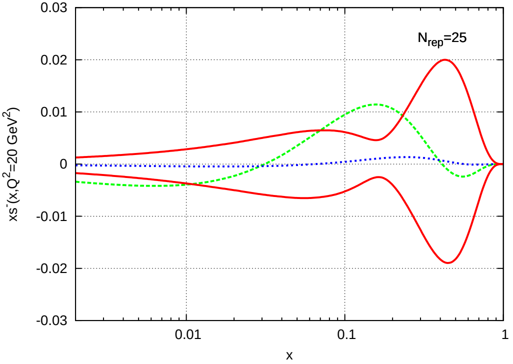
- Only theoretical constraints on $s^-(x, Q_0^2)$ is the valence sum rule.
- At least one crossing required by sum rule, but some replicas have two crossings.



NNPDF 1.2 Results

Strange Asymmetry determination: $s^-(x, Q^2)$

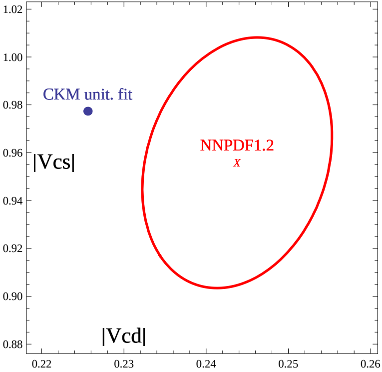
- Only theoretical constraints on $s^-(x, Q_0^2)$ is the valence sum rule.
- At least one crossing required by sum rule, but some replicas have two crossings.



NNPDF 1.2 Results

CKM matrix elements determination

- Joint determination of $|V_{cd}|$ and $|V_{cs}|$ CKM matrix elements



- Result for the combined fit

$$
\begin{aligned}
|V_{cs}| &= 0.96 \pm 0.07 \\
|V_{cs}| &= 0.244 \pm 0.019 \\
\rho[V_{cs}, V_{cd}] &= 0.21
\end{aligned}
$$

- $|V_{cs}|$ most accurate direct determination
- $|V_{cd}|$ accuracy comparable to other determinations from dimuons

- Ability to disentangle (large) uncertainties on PDFs from (small) uncertainties on physical parameters

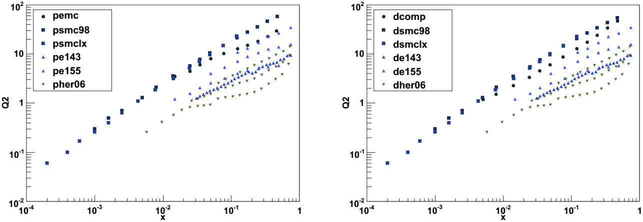
Towards NNPDF 2.0

The first NNPDF Global Fit

- Inclusion of hadronic data (Drell-Yan EW gauge boson production and Jets) crucial to constrain certain PDFs.
- NLO computation for hadronic observables too slow to be used in a parton fit $\Rightarrow$ many parton fits use $K-$factors (impact on the accuracy difficult to assess).
- We use *fastNLO* to include jet observables and develop our own *fastDY* for DY-like observables.
- First preliminary fits look promising expect the set to be public in Fall '09

[L. Del Debbio, A. Piccione and AG, in preparation]

- The NNPDF methodology is used to obtain a bias-free parametrization of the polarized DIS structure functions g_1^p and g_1^d from asymmetry data and a revalutation of the Bjorken sum rule.
- We use the available data for the virtual photon asymmetry A_1

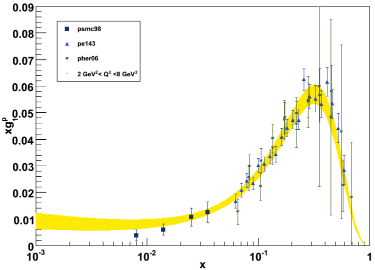


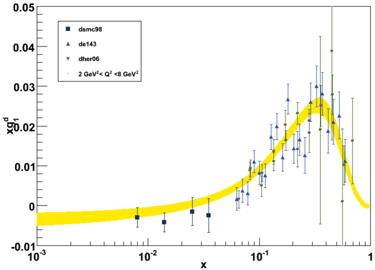
- We assess the impact of different assumptions on the extraction of g_1 from A_1 data.

g_1 and Bjorken sum rule from Neural Networks

A first step in the polarized territory

- We obtained a bias-free parametrization of g_1^p and g_1^d.





- We are now working on the extraction of the couplings α_S, g_A and the higher-twist term from the computation of the Bjorken sum rule.

Conclusions

- An accurate determination of Parton Densities and a faithful determination of their errors will play a crucial role in the success of the LHC program.

- The NNPDF Approach, based on combining Monte Carlo techinques and Neural Networks is effective in tackling problems affecting traditional PDF fits (parametrization bias, data incompatibility).

- The methodology is especially effective estimating the uncertainies in situations where experimental information are scarce.

Project status

- Global DIS fit (NNPDF1.0) completed and available in LHAPDF.
- Dedicated strangeness analysis (NNPDF1.2) yields interesting results, to be released soon.
- Work in progress on the first global fit including hadronic data (NNPDF2.0).