Parton Distribution Functions and their uncertainties: the polarized case

3rd Workshop on the QCD Structure of the Nucleon - Bilbao, Spain

Emanuele R. Nocera
in collaboration with R.D. Ball, S. Forte, A. Guffanti, G. Ridolfi and J. Rojo

Università degli studi di Milano & INFN, sezione di Milano

October, 24 2012

Outline

1. Motivation
   - Issues in Standard PDF determination
2. The NNDPF procedure
   - A general overview
   - Monte Carlo sampling, Neural Networks and minimization
3. Towards `NNPDFpol1.0`
   - Experimental dataset, PDF parametrization, theoretical constraints
   - Preliminary results: the `NNPDFpol1.0` parton set
4. Conclusions
   - Summary and outlook

1. Motivation

Issues in standard PDF determination

- Extraction of a set of functions with error bands from a set of data points.
- We need an error band, i.e. a probability density $\mathcal{P}[\Delta q(x)]$ in the space of PDFs:

$$\langle \mathcal{O} \rangle = \int \mathcal{D}\Delta q \, \mathcal{P}[\Delta q] \mathcal{O}[\Delta q]$$

$$\sigma^2_{\mathcal{O}} = \int \mathcal{D}\Delta q \, \mathcal{P}[\Delta q] (\mathcal{O}[\Delta q] - \langle \mathcal{O} \rangle)^2$$

Standard approach

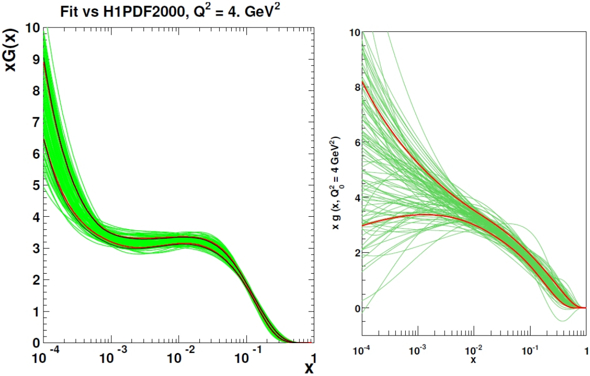
1. Choose a fixed functional form like
$$\Delta q_i(x, Q_0^2) = A_i x^{b_i} (1-x)^{c_i} (1+\ldots)$$
2. Determine best-fit parameters
3. Errors determined via Gaussian linear error propagation

But...

1. Is the parametrization flexible enough?
2. What is the error associated to any particular choice?
3. Need to rely on linear error propagation

Simple functional forms vs Neural Networks

HERA-LHC 2009 PDF benchmarks



simple functional forms Neural Networks

- Simple functional forms $\Delta q(x) = Ax^b(1-x)^c P(x)$
  $\longrightarrow$ systematic underestimation of uncertainties $\Rightarrow$ tolerance
- Artificial Neural Networks as universal interpolants
  $\longrightarrow$ reduce theoretical bias from choice of PDF functional form

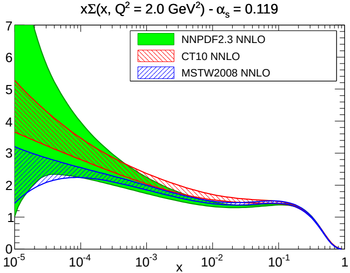
PDF fitting: a new approach

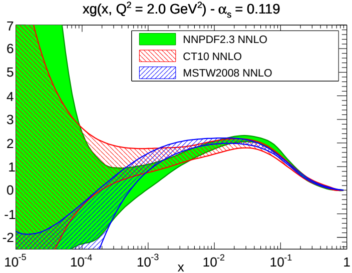
NNPDF: a new approach to PDF fitting based on
Monte Carlo sampling and Neural Networks

	NNPDF1.0	NNPDF1.2	NNPDF2.0	NNPDF2.1 NLO	NNPDF2.1 LO and NNLO	NNPDF2.2/2.3
DIS	✔	✔	✔	✔	✔	✔
Drell-Yan data	✘	✘	✔	✔	✔	✔
Jet data	✘	✘	✔	✔	✔	✔
LHC data	✘	✘	✘	✘	✘	✔
Independent param of the strange and anti-strange	✘	✔	✔	✔	✔	✔
Heavy Quark masses	✘	✘	✘	✔	✔	✔
NNLO	✘	✘	✘	✘	✔	✔

PDF fitting: a new approach

NNPDF: a new approach to PDF fitting based on
Monte Carlo sampling and Neural Networks





The NNPDF Collaboration, arXiv:1207.1303, accepted for publication in Nucl.Phys.B

Routinely used in LHC data analysis and theory prediction
Most recent global fit: NNPDF2.3 available at NLO and NNLO
The first fit including all available LHC data

NNPDF available products: technical details

1. NNPDF are available as .LHgrid files in which we provide the value of each PDF member at gridded values (x, Q^2) suitable for spline interpolation.
2. Different .LHgrid files are available for different values of $\alpha_s(M_Z^2)$ and datasets.
3. .LHgrid files can be downloaded from the LHAPDF web page

   http://lhapdf.hepforge.org/pdfsets

   and can be used through the LHAPDF interface
4. We also provide a Mathematica interface at the NNPDF web page

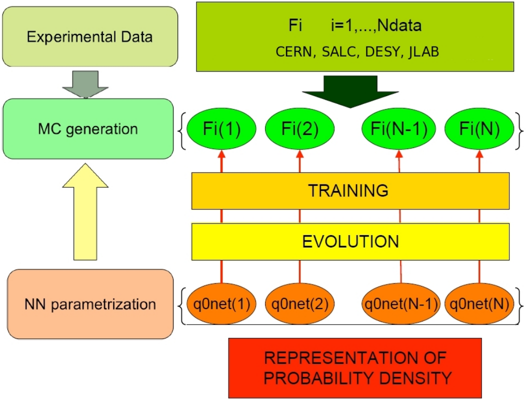
   http://nnpdf.hepforge.org/html/mathematica.html

Further NNPDF projects

1. Spin-dependent PDF fits: NNPDFpol1.0
2. Nuclear PDFs with NNPDF methodology: Nuclear NNPDF1.0

2. The NNPDF approach

A general overview on the methodology



Ingredients:
Monte Carlo sampling and Neural Networks

Ingredient 1: Monte Carlo sampling of experimental data

MONTE CARLO SAMPLING

- Sample the probability density $\mathcal{P}[\Delta q]$ in the space of functions assuming multi-Gaussian data probability distribution

$$g_{1,p}^{(\mathrm{art}),(k)}(x,Q^2) = \left[1 + r_{s,p}^{(k)}\sigma_{s,p}\right] g_{1,p}^{(\mathrm{exp})}(x,Q^2)$$

$\sigma_{s,p}$: statistical errors (also uncorrelated systematics) $r_{s,p}^{(k)}$: Gaussian random numbers

- Generate MC ensemble of N_{rep} replicas with the data probability distribution

Ingredient 1: Monte Carlo sampling of experimental data

MONTE CARLO SAMPLING

- Sample the probability density $\mathcal{P}[\Delta q]$ in the space of functions assuming multi-Gaussian data probability distribution

$$g_{1,p}^{(\mathrm{art}),(k)}(x, Q^2) = \left[1 + r_{s,p}^{(k)}\sigma_{s,p}\right] g_{1,p}^{(\mathrm{exp})}(x, Q^2)$$

$\sigma_{s,p}$: statistical errors (also uncorrelated systematics) $r_{s,p}^{(k)}$: Gaussian random numbers

- Generate MC ensemble of N_{rep} replicas with the data probability distribution

MAIN FEATURES

- Expectation values for observables are Monte Carlo integrals

$$\langle \mathcal{O}[\Delta q] \rangle = \frac{1}{N_{\mathrm{rep}}} \sum_{k=1}^{N_{\mathrm{rep}}} \mathcal{O}[\Delta q_k]$$

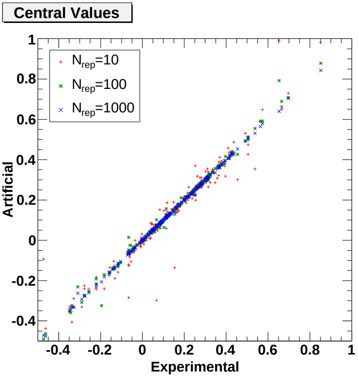
...and the same is true for errors, correlations etc.

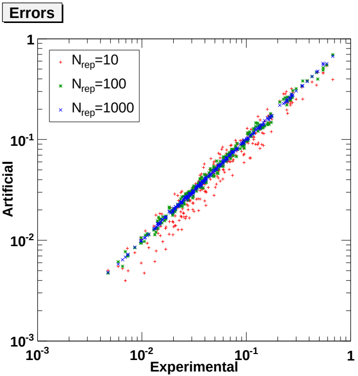
- No need to rely on linear propagation of errors
- Possibility to test for non-Gaussian behaviour in fitted PDFs

Ingredient 1: Monte Carlo sampling of experimental data

DETERMINING SAMPLE SIZE

- Require the average over the replicas reproduces central values and errors of the original experimental data to desired accuracy



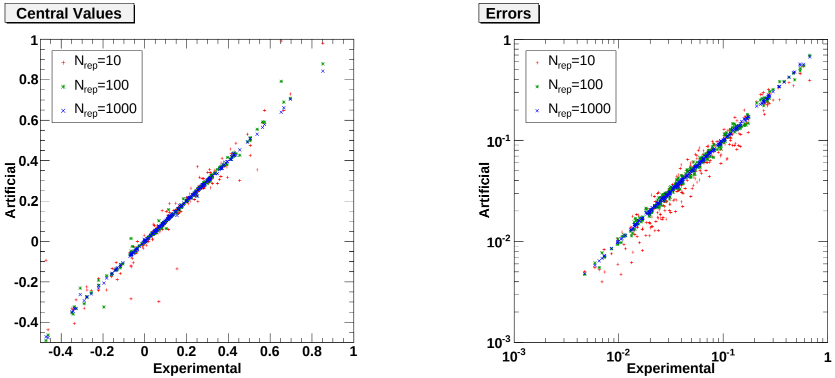


Accuracy of few % requires ∼ 100 replicas

Ingredient 1: Monte Carlo sampling of experimental data

DETERMINING SAMPLE SIZE

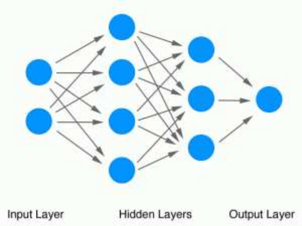
- Require the average over the replicas reproduces central values and errors of the original experimental data to desired accuracy



Accuracy of few % requires ∼ 100 replicas

Ingredient 2: Neural Networks

A convenient functional form
providing redundant and flexible parametrization
used as a generator of random functions in the PDF space

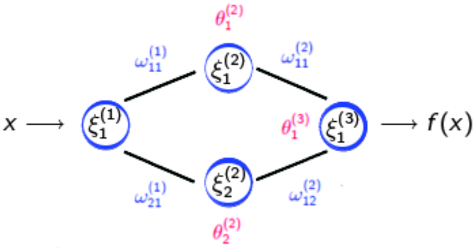


$$\xi_i^{(l)} = g\left(\sum_j^{n_l-1} \omega_{ij}^{(l-1)} \xi_j^{(l-1)} - \theta_i^{(l)}\right)$$

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

- made of neurons grouped into layers (define the architecture)
- each neuron receives input from neurons in preceding layer (feed-forward NN)
- activation determined by parameters (weights and thresholds)
- activation determined according to a non-linear function (except the last layer)

EXAMPLE: THE SIMPLEST 1-2-1 NN



$$f(x) \equiv \xi_1^{(3)} = \left\{ 1 + \exp\left[\theta_1^{(3)} - \frac{\omega_{11}^{(2)}}{1 + e^{\theta_1^{(2)} - x\omega_{11}^{(1)}}} - \frac{\omega_{12}^{(2)}}{1 + e^{\theta_2^{(2)} - x\omega_{21}^{(1)}}} \right] \right\}^{-1}$$

Recall: $$\xi_i^{(l)} = g\left(\sum_j^{n_l - 1} \omega_{ij}^{(l-1)} \xi_j^{(l-1)} - \theta_i^{(l)} \right) \quad ; \quad g(x) = \frac{1}{1 + e^{-x}}$$

Ingredient 2: Neural Networks

NEURAL NETWORKS

- Parametrize each polarized PDF replica with flexible Neural Network

DSSV, AAC, LSS, BB
$\mathcal{O}(10-20)$ parameters

NNPDFpol
$\mathcal{O}(200)$ parameters

- Train NN to determine the best fit for each replica
- Compute an ensemble of observables and compare to experimental data

Ingredient 2: Neural Networks

NEURAL NETWORKS

- Parametrize each polarized PDF replica with flexible Neural Network

DSSV, AAC, LSS, BB	NNPDFpol
$\mathcal{O}(10-20)$ parameters	$\mathcal{O}(200)$ parameters

- Train NN to determine the best fit for each replica
- Compute an ensemble of observables and compare to experimental data

MAIN FEATURES

- Only require smoothness of the fitted function
- Do not require any other prejudice on *a priori* functional form
- Reduce the bias associated to the choice of some functional form

One more ingredient: minimization and stopping

GENETIC ALGORITHM

Standard minimization unefficient owing to the large parameter space
and non-local x-dependence of the observables
Genetic algorithm provides better exploration of the whole parameter space

- Set Neural Network parameters randomly
- Make clones of the parameter vector and mutate them
- Define a figure of merit or error function for the k-th replica

$$E^{(k)} = \frac{1}{N_{\rm rep}} \sum_{i,j=1}^{N_{\rm rep}} \left(g_{1,i}^{({\rm art})(k)} - g_{1,i}^{({\rm net})(k)} \right) \left(({\rm cov})^{-1} \right)_{ij} \left(g_{1,j}^{({\rm art})(k)} - g_{1,j}^{({\rm net})(k)} \right)$$

$g_{1,i}^{({\rm art})(k)}$: generated from Monte Carlo sampling
$g_{1,i}^{({\rm net})(k)}$: computed from Neural Network PDFs

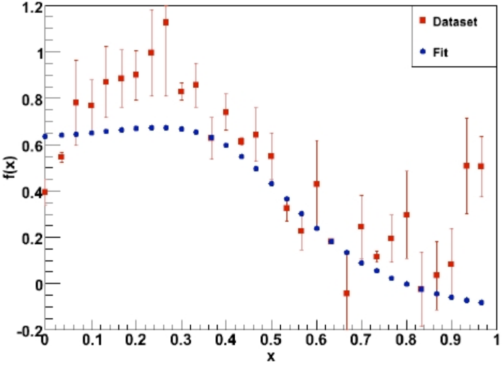
- Select the best set of parameters and perform other manipulations (crossing, mutating, ...) until stability is reached.

One more ingredient: minimization and stopping

DRAWBACK

- NN can learn fluctuations owing to their flexibility

UNDERLEARNING

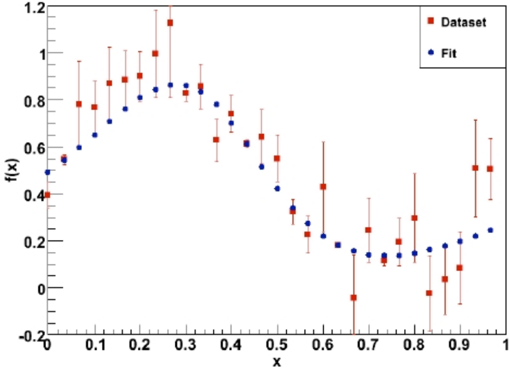


One more ingredient: minimization and stopping

DRAWBACK

- NN can learn fluctuations owing to their flexibility

PROPER LEARNING

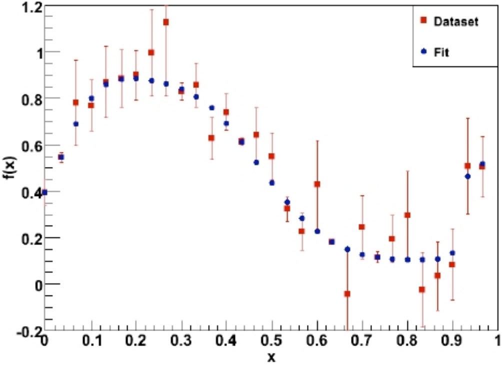


One more ingredient: minimization and stopping

DRAWBACK

- NN can learn fluctuations owing to their flexibility

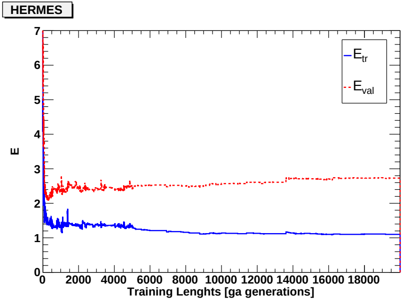
OVERLEARNING



One more ingredient: minimization and stopping

CROSS-VALIDATION METHOD

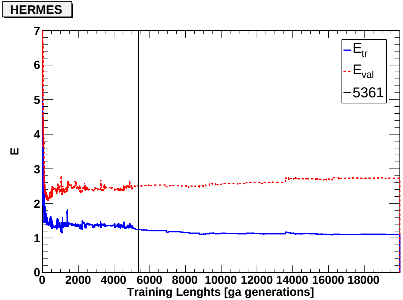
- divide data into two subsets (training & validation)
- train the NN on training subset and compute χ^2 for each subset
- stop when χ^2 of validation subset no longer decreases
  (NN are learning fluctuations!)



One more ingredient: minimization and stopping

CROSS-VALIDATION METHOD

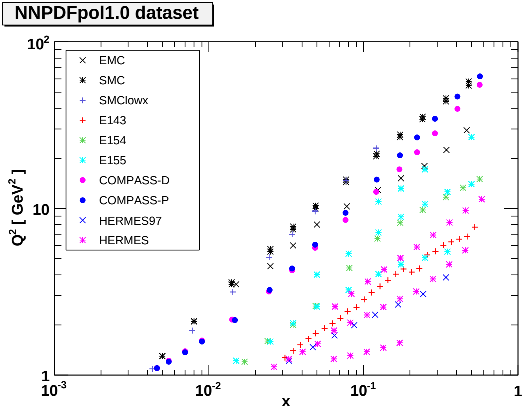
- divide data into two subsets (training & validation)
- train the NN on training subset and compute χ^2 for each subset
- stop when χ^2 of validation subset no longer decreases (NN are learning fluctuations!)



The best fit does not coincide with the χ^2 absolute minimum

3. Towards NNPDFpol1.0

Experimental dataset



1. All inclusive polarized DIS data on proton, neutron and deuteron targets
2. Kinematical cuts to remove the sensitivity to dynamical higher-twist
   - $Q^2 > 1Gev^2$
   - $W^2 = Q^2(1-x)/x \geq 6.25GeV^2$ (C. Simolo, Ph.D. Thesis. arXiv:0807.1501)

(higher twist terms added to observables and fitted to data become compatible with zero)

Input polarized PDF basis

Four polarized PDFs (gluon + linear combinations of light quarks)

1. singlet $\Delta\Sigma(x) \equiv \Sigma_{i=1}^{n_f} \Delta q_i(x)$
2. gluon $\Delta g(x)$
3. triplet $\Delta T_3(x) \equiv \Delta u(x) - \Delta d(x)$
4. octet $\Delta T_8(x) \equiv \Delta u(x) + \Delta d(x) - 2\Delta s(x)$

$$\Delta q_i(x,Q^2) = q_i^{\uparrow\uparrow}(x,Q^2) + \bar{q}_i^{\uparrow\uparrow}(x,Q^2) - q_i^{\uparrow\downarrow}(x,Q^2) + \bar{q}_i^{\uparrow\downarrow}(x,Q^2)$$

$$\Delta g(x,Q^2) \equiv g^{\uparrow\uparrow}(x,Q^2) - g^{\uparrow\downarrow}(x,Q^2)$$

Inclusive neutral-current DIS data do not allow disentangling the contributions from q and $\overline{q}$.
In our notation, Δq takes into account flavor plus anti-flavor contributions.

- At initial scale $Q_0^2 = 1 GeV^2$
- Assume all heavy quarks are generated radiatively
- Adopt $\alpha_s(M_Z^2) = 0.119$, $m_c = 1.4$ GeV, $m_b = 4.75$ GeV

PDF Parametrization

$$\Delta\Sigma(x, Q_0^2) = (1-x)^{m_{\Delta\Sigma}} x^{-n_{\Delta\Sigma}} NN_{\Delta\Sigma}(x)$$

$$\Delta g(x, Q_0^2) = (1-x)^{m_{\Delta g}} x^{-n_{\Delta g}} NN_{\Delta g}(x)$$

$$\Delta T_3(x, Q_0^2) = A_{\Delta T_3}(1-x)^{m_{\Delta T_3}} x^{-n_{\Delta T_3}} NN_{\Delta T_3}(x)$$

$$\Delta T_8(X, Q_0^2) = A_{\Delta T_8}(1-x)^{m_{\Delta T_8}} x^{-n_{\Delta T_8}} NN_{\Delta T_8}(x)$$

1. Each polarized PDF parametrized with a multi-layer feed-forward NN (2-5-3-1)
2. Parametrization supplemented with a preprocessing polynomial:
   $\longrightarrow$ exponents m and n randomly chosen in fixed intervals;
   $\longrightarrow$ intervals must be sufficient large not to introduce a bias on the fit
   $\longrightarrow$ check *a posteriori* by studying asymptotic exponents
3. Overall normalization constant factored out for triplet and octet.

$$A_{\Delta T_3} = \frac{a_3}{\int_0^1 dx[(1-x)^{m_{\Delta T_3}} x^{-n_{\Delta T_3}} NN_{\Delta T_3}(x)]}$$

$$A_{\Delta T_8} = \frac{a_8}{\int_0^1 dx[(1-x)^{m_{\Delta T_8}} x^{-n_{\Delta T_8}} NN_{\Delta T_8}(x)]}$$

PDF Parametrization

$$\Delta\Sigma(x, Q_0^2) = (1-x)^{m_{\Delta\Sigma}} x^{-n_{\Delta\Sigma}} NN_{\Delta\Sigma}(x)$$

$$\Delta g(x, Q_0^2) = (1-x)^{m_{\Delta g}} x^{-n_{\Delta g}} NN_{\Delta g}(x)$$

$$\Delta T_3(x, Q_0^2) = A_{\Delta T_3}(1-x)^{m_{\Delta T_3}} x^{-n_{\Delta T_3}} NN_{\Delta T_3}(x)$$

$$\Delta T_8(X, Q_0^2) = A_{\Delta T_8}(1-x)^{m_{\Delta T_8}} x^{-n_{\Delta T_8}} NN_{\Delta T_8}(x)$$

1. Each polarized PDF parametrized with a multi-layer feed-forward NN (2-5-3-1)
2. Parametrization supplemented with a preprocessing polynomial:
   → exponents m and n randomly chosen in fixed intervals;
   → intervals must be sufficient large not to introduce a bias on the fit
   → check *a posteriori* by studying asymptotic exponents
3. Overall normalization constant factored out for triplet and octet.

$$A_{\Delta T_3} = \frac{a_3}{\int_0^1 dx[(1-x)^{m_{\Delta T_3}} x^{-n_{\Delta T_3}} NN_{\Delta T_3}(x)]}$$

$$A_{\Delta T_8} = \frac{a_8}{\int_0^1 dx[(1-x)^{m_{\Delta T_8}} x^{-n_{\Delta T_8}} NN_{\Delta T_8}(x)]}$$

PDF Parametrization

$$\Delta\Sigma(x, Q_0^2) = (1-x)^{m_{\Delta\Sigma}} x^{-n_{\Delta\Sigma}} NN_{\Delta\Sigma}(x)$$

$$\Delta g(x, Q_0^2) = (1-x)^{m_{\Delta g}} x^{-n_{\Delta g}} NN_{\Delta g}(x)$$

$$\Delta T_3(x, Q_0^2) = A_{\Delta T_3}(1-x)^{m_{\Delta T_3}} x^{-n_{\Delta T_3}} NN_{\Delta T_3}(x)$$

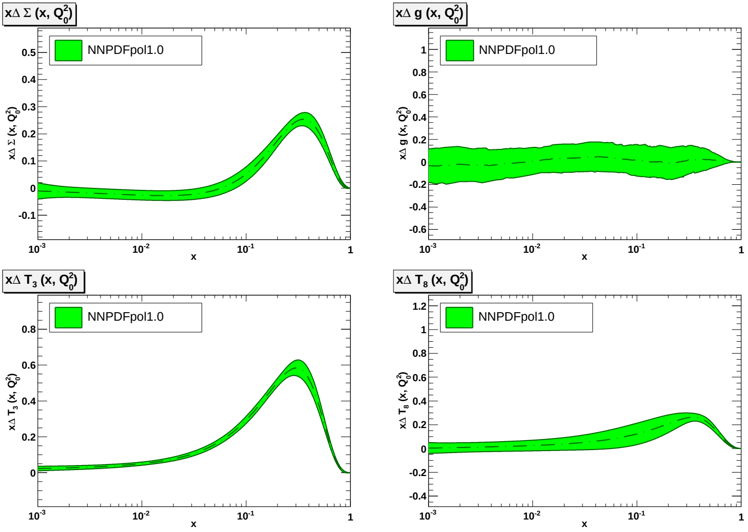
$$\Delta T_8(X, Q_0^2) = A_{\Delta T_8}(1-x)^{m_{\Delta T_8}} x^{-n_{\Delta T_8}} NN_{\Delta T_8}(x)$$

1. Each polarized PDF parametrized with a multi-layer feed-forward NN (2-5-3-1)
2. Parametrization supplemented with a preprocessing polynomial:
   → exponents m and n randomly chosen in fixed intervals;
   → intervals must be sufficient large not to introduce a bias on the fit
   → check *a posteriori* by studying asymptotic exponents
3. Overall normalization constant factored out for triplet and octet.

$$A_{\Delta T_3} = \frac{a_3}{\int_0^1 dx[(1-x)^{m_{\Delta T_3}} x^{-n_{\Delta T_3}} NN_{\Delta T_3}(x)]}$$

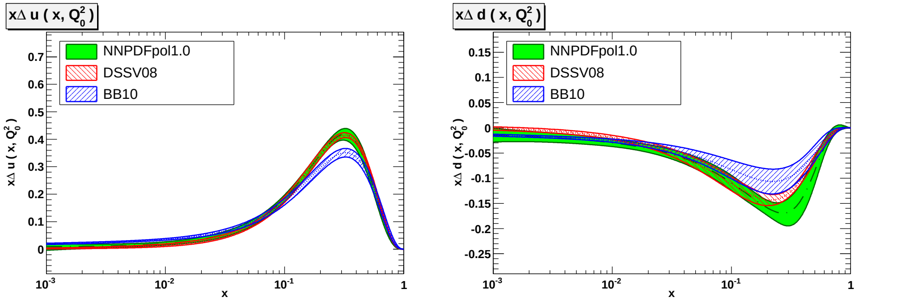
$$A_{\Delta T_8} = \frac{a_8}{\int_0^1 dx[(1-x)^{m_{\Delta T_8}} x^{-n_{\Delta T_8}} NN_{\Delta T_8}(x)]}$$

The NNPDFpol1.0 parton set: parametrization basis



Comparison with other parton sets: flavor basis

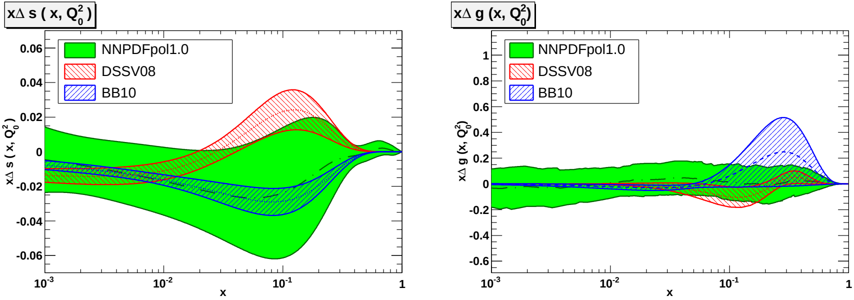
DSSV09: D. de Florian et al., Phys.Rev. D80 (2009) 034030
BB10: J. Blumlein and H. Bottcher, Nucl.Phys. B841 (2010) 205



- NNPDF Δu and Δd uncertainties comparable with other PDF sets
- Much larger error band for strangeness and gluon PDFs

Comparison with other parton sets: flavor basis

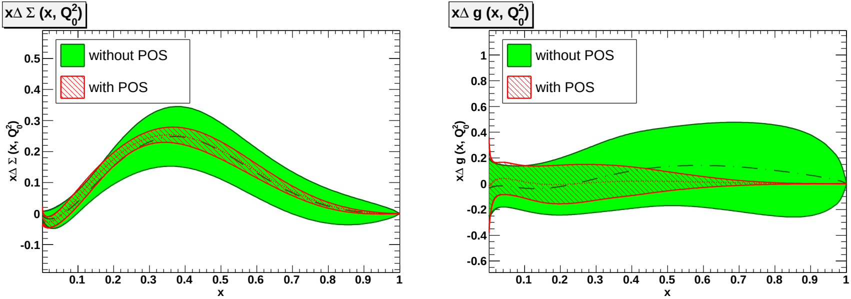
DSSV09: D. de Florian et al., Phys.Rev. D80 (2009) 034030
BB10: J. Blumlein and H. Bottcher, Nucl.Phys. B841 (2010) 205



- NNPDF Δu and Δd uncertainties comparable with other PDF sets
- Much larger error band for strangeness and gluon PDFs

Theoretical constraints: positivity

Polarized PDFs are only loosely constrained by the data.
At high-x positivity constraints are as important as data!



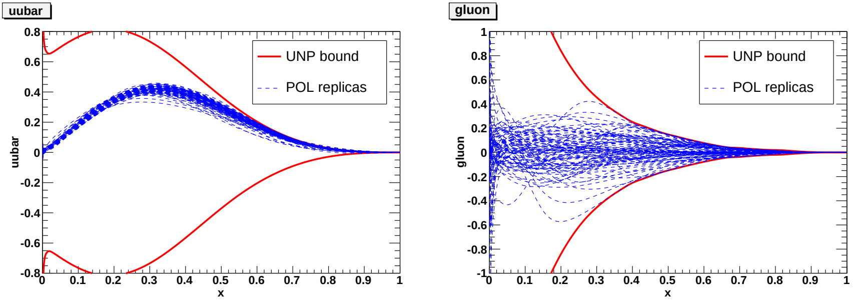
$$\chi^2_{\rm tot} = 0.72 \text{ (no positivity constraints)}$$
$$\chi^2_{\rm tot} = 0.75 \text{ (with positivity constraints)}$$

Theoretical constraints: positivity

We impose LO positivity, i.e.

$$\left|\delta f(x, Q_0^2)\right| \leq f(x, Q_0^2)$$

at the parametrization scale $Q_0 = 1GeV^2$ separately for each flavor (u,d,s) and gluon (g).



The spin content of the proton

Singlet and Gluon first moments in $\overline{\text{MS}}$ scheme at $Q_0^2 = 4$ GeV^2

	NNPDFpol1.0	DSSV08 (tr)	AAC08 (pos Δg)	BB10	LSS10 (pos Δg)
$[\Delta\Sigma]$	0.26 ± 0.16	0.37 ± 0.02	0.24 ± 0.07	0.19 ± 0.08	0.21 ± 0.03
$[\Delta g]$	0.1 ± 1.7	-0.03 ± 0.16	0.63 ± 0.81	0.46 ± 0.43	0.32 ± 0.19

Notice the large uncertainty on the first moments:
Singlet between two and four times
Gluon almost one order of magnitude

DSSV09: D. de Florian et al., Phys.Rev. D80 (2009) 034030
AAC08: M. Hirai and S. Kumano, Nucl.Phys. B813 (2009) 106
BB10: J. Blumlein and H. Bottcher, Nucl.Phys. B841 (2010) 205
LSS10: E. Leader, A.V. Sidorov and D.B. Stamenov, Phys.Rev. D82 (2010) 114018

4. Conclusions

Final remarks

Summary

1. The NNPDF methodology provides a statistically sound procedure for PDF fitting based on:
   - Monte Carlo sampling of experimental data;
   - Neural Networks used as unbiased interpolants.
2. `NNPDFpol1.0` is the first polarized parton determination using the NNPDF methodology:
   - Δu and Δd quark combinations have uncertainties similar to other parton sets;
   - Δs and Δg have larger uncertainties (both can be compatible with zero)

     $\Rightarrow$ what is their (true) contribution to the proton spin content?

Outlook

1. Include data sets from other processes (open charm and jet production with fixed target, inclusive jet production, W boson production at RHIC, ...) in order to test how uncertainty bands can be improved, in particular Δs and Δg.

Final remarks

Summary

1. The NNPDF methodology provides a statistically sound procedure for PDF fitting based on:
   - Monte Carlo sampling of experimental data;
   - Neural Networks used as unbiased interpolants.
2. `NNPDFpol1.0` is the first polarized parton determination using the NNPDF methodology:
   - Δu and Δd quark combinations have uncertainties similar to other parton sets;
   - Δs and Δg have larger uncertainties (both can be compatible with zero)

     $\Rightarrow$ what is their (true) contribution to the proton spin content?

Outlook

1. Include data sets from other processes (open charm and jet production with fixed target, inclusive jet production, W boson production at RHIC, ...) in order to test how uncertainty bands can be improved, in particular Δs and Δg.

Thank you for your attention!

5. Backup

PDF fitting: state of the art

1. First stage: first moments of polarized PDFs and polarized sum rules (last 25 years)
   $\longrightarrow$ "historical" experimental collaborations (at CERN, SLAC, DESY, JLAB):
   EMC, SMC, E142, E143, E154, E155, COMPASS, HERMES, CLAS, ...
2. Second stage: polarized PDF fits from global NLO QCD analysis (last $\sim$15 years)
   $\longrightarrow$ different choice of datasets, parton parametrization, treatement of higher twists, ...
   ABFR (arXiv:hep-ph/9803237, 1998), BB (arXiv:1005.3113, 2010) (DIS only);
   AAC (arXiv:0808.0413, 2008), LSS (arXiv:1010.0574, 2010) (DIS+SIDIS);
   DSSV (arXiv:0904.3821, 2009) (DIS+SIDIS+pp)
3. Third stage: provide uncertainties on polarized PDFs (last $\sim$10 years)
   $\longrightarrow$ Gaussian error propagation, Lagrange multiplier + Hessian method; fit with orthogonal polynomials (arXiv:1011.4873, 2010)

Polarized PDF evolution

In Mellin space the DGLAP equations

$$\begin{aligned} \mu^2 \frac{\partial}{\partial \mu^2} \Delta q_{NS}^{\pm,\nu}(N,\mu^2) &= \Delta\gamma_{NS}^{\pm,\nu} q_{NS}^{\pm,\nu}(N,\mu^2) \\ \mu^2 \frac{\partial}{\partial \mu^2} \begin{pmatrix} \Delta\Sigma \\ \Delta g \end{pmatrix} (N,\mu^2) &= \begin{pmatrix} \Delta\gamma_{qq}(N,\alpha_s(Q^2)) & \Delta\gamma_{qg}(N,\alpha_s(Q^2)) \\ \Delta\gamma_{gq}(N,\alpha_s(Q^2)) & \Delta\gamma_{gg}(N,\alpha_s(Q^2)) \end{pmatrix} \begin{pmatrix} \Delta\Sigma \\ \Delta g \end{pmatrix} \end{aligned}$$

can be solved analitically

$$\Delta q_{NS}^{\pm,\nu}(N,Q^2) = \Gamma_{NS}^{\pm,\nu}(N,a_s,a_0)\Delta q_{NS}^{\pm,\nu}(N,Q_0^2) \text{ ,} a_s \equiv \alpha_s/2\pi$$

where, at NLO,

$$\Gamma_{NS,NLO}^{\pm,\nu}(N,a_s,a_0) = \exp\left\{ \frac{U_1^{\pm,\nu}}{b_1} \ln\left(\frac{1+b_1 a_s}{1+b_1 a_0} \right) \right\} \left(\frac{a_s}{a_0} \right)^{-R_0^{NS}}$$

Polarized PDF evolution

NNPDF NLO polarized PDF evolution (**Fast Kernel method**) benchmarked with the Les Houches PDF benchmarks (G. Salam and a. Vogt, hep-ph/0511119)

x	$\epsilon_{\rm rel}\,(\Delta u_V)$	$\epsilon_{\rm rel}\,(\Delta d_V)$	$\epsilon_{\rm rel}\,(\Delta\Sigma)$	$\epsilon_{\rm rel}\,(\Delta g)$
10^{-3}	$1.1\,10^{-4}$	$9.2\,10^{-5}$	$9.9\,10^{-5}$	$1.1\,10^{-4}$
10^{-2}	$1.4\,10^{-4}$	$1.9\,10^{-4}$	$3.5\,10^{-4}$	$9.3\,10^{-5}$
0.1	$1.2\,10^{-4}$	$1.6\,10^{-4}$	$5.4\,10^{-6}$	$1.7\,10^{-4}$
0.3	$2.3\,10^{-6}$	$1.1\,10^{-5}$	$7.5\,10^{-6}$	$1.7\,10^{-5}$
0.5	$5.6\,10^{-6}$	$9.6\,10^{-6}$	$1.6\,10^{-5}$	$2.5\,10^{-5}$
0.7	$1.2\,10^{-4}$	$9.2\,10^{-7}$	$1.6\,10^{-4}$	$7.8\,10^{-5}$
0.9	$3.5\,10^{-3}$	$1.1\,10^{-2}$	$4.1\,10^{-3}$	$7.8\,10^{-3}$

Very accurate evolution!

Target mass corrections

- Extracting both structure functions $g_1(x, Q^2)$ and $g_2(x, Q^2)$ from data requires measuring longitudinal and transverse spin asymmetries $A_{\parallel}$ and $A_{\perp}$
- Experimental information on $A_{\perp}$ is rather poor (in most cases only $A_{\parallel}$ is measured), thus $g_1(x, Q^2)$ and $g_2(x, Q^2)$ are related

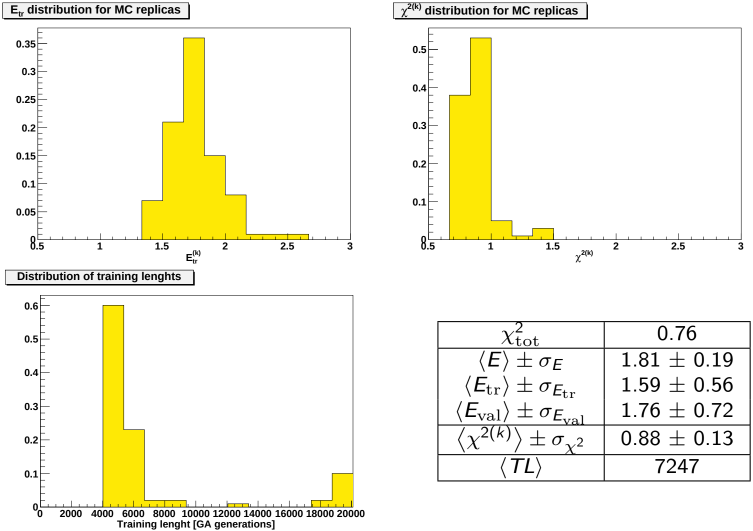
$$g_1(x, Q^2) = \frac{F_1(x, Q^2)}{1+\gamma\eta}\frac{A_{\parallel}}{D} + \frac{\gamma(\gamma-\eta)}{\gamma\eta+1} g_2(x, Q^2)$$

$$\gamma = \frac{2m_N x}{Q}\ ; \quad \eta = \frac{\epsilon\gamma y}{1-\epsilon(1-y)}\ ; \quad D = \frac{1-(1-y)\epsilon}{1+\epsilon R(x, Q^2)}\ ; \quad \epsilon = \frac{4(1-y)-\gamma^2 y^2}{2y^2+4(1-y)+\gamma^2 y^2}\ ; \quad y = 1-\frac{E'}{E}$$

MUST MAKE SOME ASSUMPTION ON $g_2(x, Q^2)$

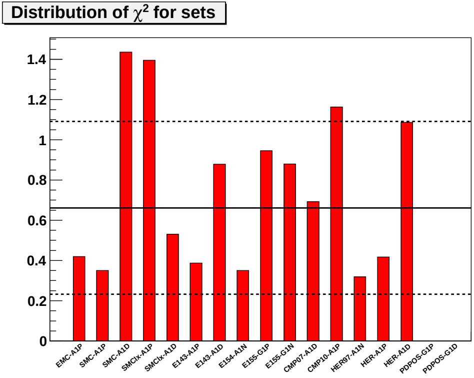
$g_2 = 0$ OR $g_2 = g_2^{WW}$ (relate g_2 to g_1 by means of the Wandzura-Wilczek relation)

- Target mass corrections implemented iteratively during the minimization procedure



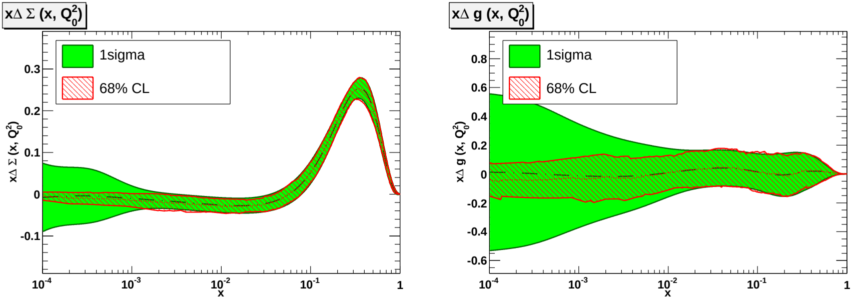
$\chi^2_{\rm tot}$	0.76
$\langle E \rangle \pm \sigma_E$	1.81 $\pm$ 0.19
$\langle E_{\rm tr} \rangle \pm \sigma_{E_{\rm tr}}$	1.59 $\pm$ 0.56
$\langle E_{\rm val} \rangle \pm \sigma_{E_{\rm val}}$	1.76 $\pm$ 0.72
$\langle \chi^{2(k)} \rangle \pm \sigma_{\chi^2}$	0.88 $\pm$ 0.13
$\langle TL \rangle$	7247

NNPDFpol1.0: individual experiments χ^2



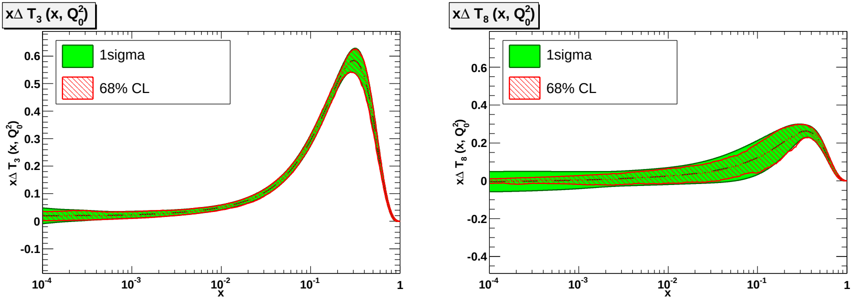
- No evidence of any specific dataset being inconsistet with each other
- Distribution of individual χ^2 values broadly consistent with statistical expectations

NNPDFpol1.0: 68% confidence levels



comparison between 1σ error bands and 68% confidence level
test for non-Gaussian behaviour
sizable deviations from Gaussian behaviour in the extrapolation region

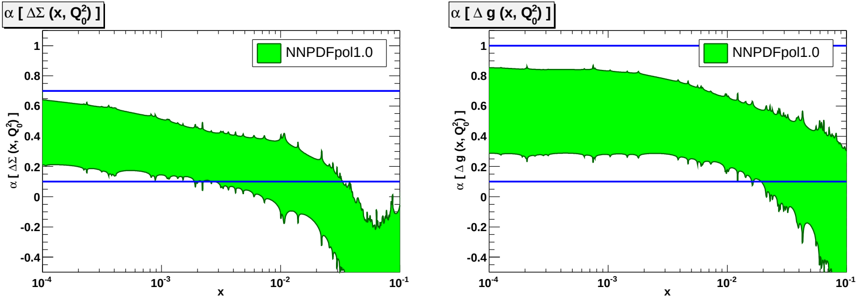
NNPDFpol1.0: 68% confidence levels



comparison between 1σ error bands and 68% confidence level
test for non-Gaussian behaviour
sizable deviations from Gaussian behaviour in the extrapolation region

Preprocessing: effective asymptotic exponents

$0.0 \leq n_{\Delta\Sigma} \leq 0.6$ $\qquad$ $0.0 \leq n_{\Delta g} \leq 0.6$

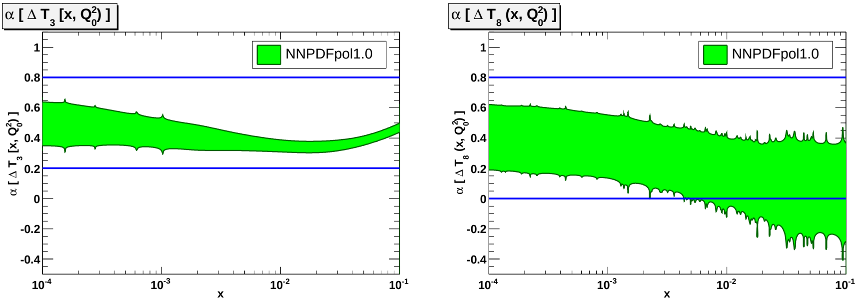


$$\alpha_{\rm eff}(x,Q^2) \equiv \ln(\Delta q(x,Q^2))/\ln(1/x) \quad \text{at } Q^2 = Q_0^2 = 1 GeV^2$$

Effective exponents always contained in the preprocessing exponents range
The polarized PDF is driven only by experimental data

Preprocessing: effective asymptotic exponents

$0.0 \leq n_{\Delta\Sigma} \leq 0.6$ $\qquad$ $0.0 \leq n_{\Delta g} \leq 0.6$

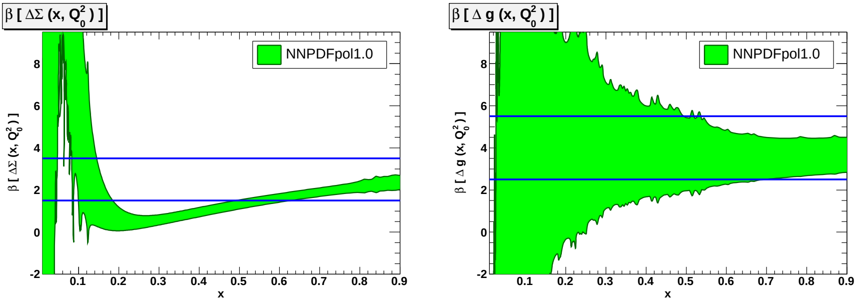


$$\alpha_{\rm eff}(x, Q^2) \equiv \ln(\Delta q(x, Q^2))/\ln(1/x) \quad \text{at } Q^2 = Q_0^2 = 1 GeV^2$$

Effective exponents always contained in the preprocessing exponents range
The polarized PDF is driven only by experimental data

Preprocessing: effective asymptotic exponents

$0.5 \leq m_{\Delta\Sigma} \leq 5.0$ $\quad$ $0.5 \leq m_{\Delta g} \leq 4.0$



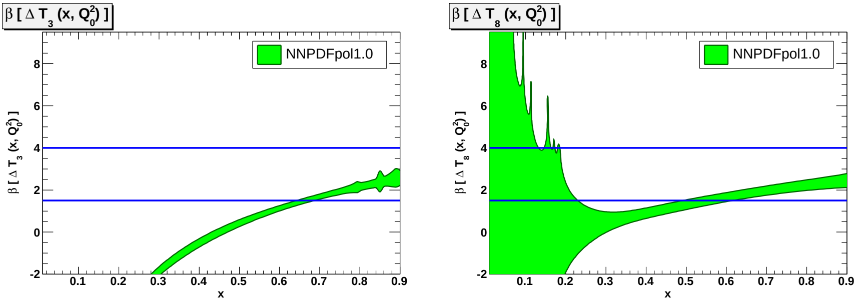
$$\beta_{\text{eff}}(x, Q^2) \equiv \ln(\Delta q(x, Q^2))/\ln(1-x) \quad \text{at } Q^2 = Q_0^2 = 1 GeV^2$$

Effective exponents always contained in the preprocessing exponents range

The polarized PDF is driven only by experimental data

Preprocessing: effective asymptotic exponents

$0.5 \leq m_{\Delta\Sigma} \leq 4.0$ $\quad$ $0.5 \leq m_{\Delta g} \leq 7.0$



$$\beta_{\text{eff}}(x, Q^2) \equiv \ln(\Delta q(x, Q^2))/\ln(1-x) \quad \text{at } Q^2 = Q_0^2 = 1 GeV^2$$

Effective exponents always contained in the preprocessing exponents range
The polarized PDF is driven only by experimental data

Distances

Compare two sets of $N_{\rm rep}^{(1)}$ and $N_{\rm rep}^{(1)}$ replicas coming from different fits
Do they have belong to the same underlying probability distribution?

MEAN VALUE

$$d^2\left(\langle q^{(k)}\rangle_{(1)}, \langle q^{(k)}\rangle_{(2)}\right) = \frac{\left(\langle q^{(k)}\rangle_{(1)} - \langle q^{(k)}\rangle_{(2)}\right)^2}{\sigma^2\left[\langle q^{(k)}\rangle_{(1)}\right] + \sigma^2\left[\langle q^{(k)}\rangle_{(2)}\right]}$$

$$\langle q^{(k)}\rangle_{(i)} = \frac{1}{N_{\mathrm{rep}(i)}}\sum_{l=1}^{N_{\mathrm{rep}(i)}} q_l^{(k)} \qquad \sigma^2\left[\langle q^{(k)}\rangle_{(i)}\right] = \frac{1}{N_{\mathrm{rep}(i)}}\sigma^2\left[q_{(i)}^{(k)}\right] = \frac{1}{N_{\mathrm{rep}(i)}-1}\sum_{l=1}^{N_{\mathrm{rep}(i)}}\left(q_{l,(i)} - \langle q\rangle_{(i)}\right)^2$$

UNCERTAINTY

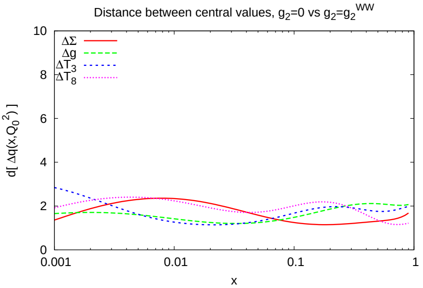
$$d^2\left(\sigma^2_{(1)}, \sigma^2_{(2)}\right) = \frac{\left(\bar\sigma^2_{(1)} - \bar\sigma^2_{(2)}\right)}{\sigma^2\left[\sigma^2_{(1)}\right] + \sigma^2\left[\sigma^2_{(2)}\right]}$$

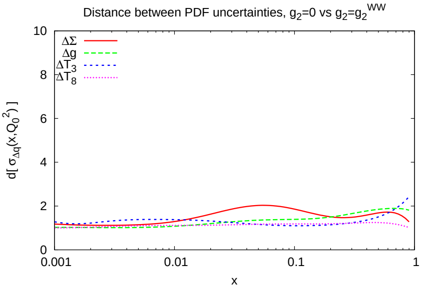
$$\bar\sigma^2_{(i)} \equiv \sigma^2\left[q_{(i)}^{(k)}\right] \qquad \sigma^2\left[\sigma^2_{(i)}\right] = \frac{1}{N_{\mathrm{rep}(i)}}\left[\frac{1}{N_{\mathrm{rep}(i)}}\sum_{l=1}^{N_{\mathrm{rep}(i)}}\left(q_{l,(i)} - \langle q\rangle_{(i)}\right)^4 - \frac{N_{\mathrm{rep}(i)}-3}{N_{\mathrm{rep}(i)}-1}\left(\bar\sigma^2_{(i)}\right)^2\right]$$

By definition, the distances have a χ^2 probability distribution with one degree of freedom
mean $\langle d\rangle = 1$ and $d \lesssim 2.3$ at 90% confidence level

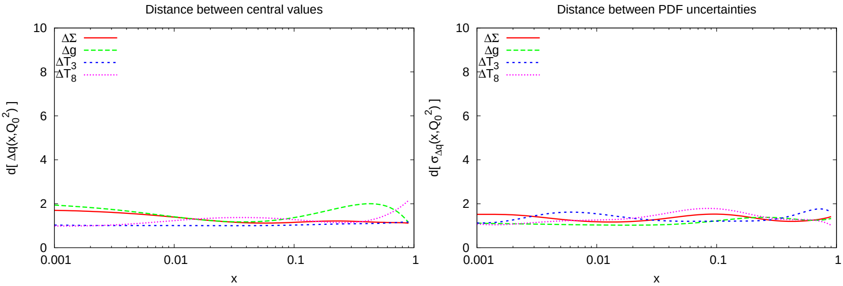
Distances

Comparison between different inclusion of Target Mass Corrections



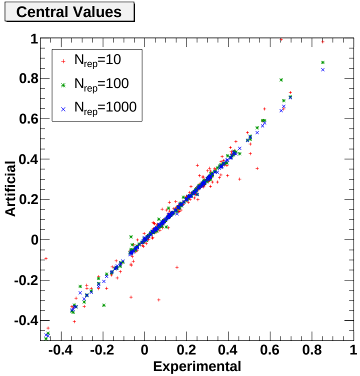


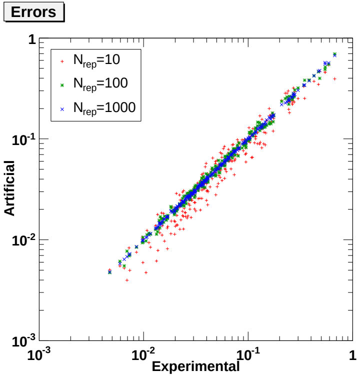
Impact of uncertainty on a8 coupling



DETERMINING SAMPLE SIZE

- Require the average over the replicas reproduces central values and errors of the original experimental data to desired accuracy





DETERMINING SAMPLE SIZE

- Require the average over the replicas reproduces central values and errors of the original experimental data to desired accuracy

Quantitative approach: devise proper statistical estimators

	$\left\langle \text{PE}\left[\langle g_1^{(\text{art})}\rangle\right]\right\rangle$ [%]			$r\left[g_1^{(\text{art})}\right]$		
N_{rep}	10	100	1000	10	100	1000
EMC	23.7	3.5	2.9	.76037	.99547	.99712
SMC	19.4	5.6	1.2	.94789	.99908	.99993
...	...	...	...	...	...	...

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

Percentage Error

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

Scatter Correlation

Ingredient 1: Monte Carlo sampling of experimental data

DETERMINING SAMPLE SIZE

- Require the average over the replicas reproduces central values and errors of the original experimental data to desired accuracy

Quantitative approach: devise proper statistical estimators

	$\left\langle \mathrm{PE}\left[\langle \delta g_1^{(\mathrm{art})}\rangle\right]\right\rangle$ [%]			$r\left[\delta g_1^{(\mathrm{art})}\right]$		
N_{rep}	10	100	1000	10	100	1000
EMC	12.8	4.9	2.0	.97397	.99521	.99876
SMC	22.4	5.4	1.7	.96585	.99489	.99980
...	...	...	...	...	...	...

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

Percentage Error

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

Scatter Correlation

Ingredient 1: Monte Carlo sampling of experimental data

DETERMINING SAMPLE SIZE

- Require the average over the replicas reproduces central values and errors of the original experimental data to desired accuracy

Accuracy of few % requires $\sim$ 100 replicas