Neural Network Parton Distributions

Juan Rojo[5]

on behalf of 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. R.[5] and M. Ubiali[1]

[1]University of Edinburgh, [2] Università di Milano, [3]Albert-Ludwigs-Universität Freiburg, [4]Universitat de Barcelona, [5]LPTHE, UPMC Paris VI

International Symposium on Multiparticle Dynamics 08
18/09/2008, DESY Hamburg

Introduction

- After 40 years of QCD, still issues to be understood in the determination of parton distributions (G. Altarelli, LHeC workshop opening lecture)
- The standard approach to PDF determination (see J. Stirling's talk) has important drawbacks, summarized by the 2006 HERA-LHC PDF benchmark analysis
- The NNPDF Collaboration approach is a proposal to overcome various problems in PDF determination with statistically sound techniques
- A faithfully estimate of PDF uncertainties is of paramount importance for precision LHC studies, even for discovery! (see talks by M. Lancaster and T. Shears)
- NNPDF1.0 → First parton set from the NNPDF collaboration → *"A determination of parton distribution with faithful uncertainty estimation"*, arxiv:0808.1231

BENCHMARK PARTONS

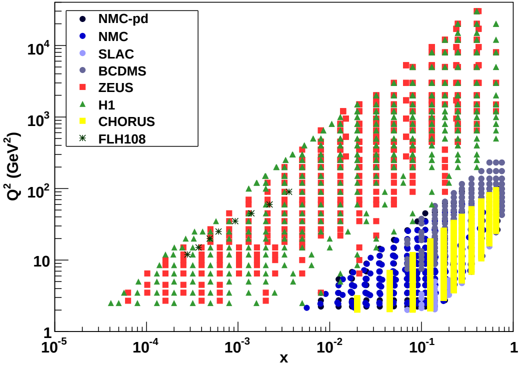
PDF benchmark analysis

- Proposed during the first HERA-LHC workshop → Benchmark PDF fit to a reduced, consistent DIS data set

Set	$N_{\rm dat}$	$x_{\rm min}$	$x_{\rm max}$	$Q^2_{\rm min}$	$Q^2_{\rm max}$
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

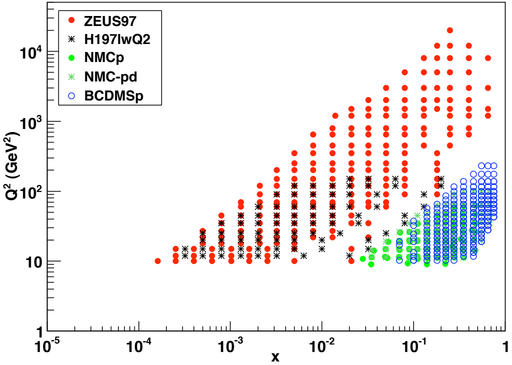
PDF benchmark analysis

- Proposed during the first HERA-LHC workshop → Benchmark PDF fit to a reduced, consistent DIS data set
- From a full DIS analysis data set ...



PDF benchmark analysis

- Proposed during the first HERA-LHC workshop → Benchmark PDF fit to a reduced, consistent DIS data set
- ... to the reduced PDF benchmark analysis data set

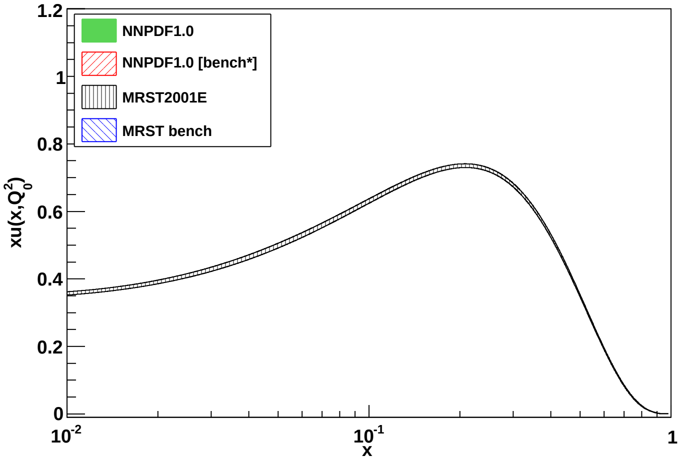


PDF benchmark analysis

- Proposed during the first HERA-LHC workshop → Benchmark PDF fit to a reduced, consistent DIS data set
- Compare results between PDF fitting collaborations and with global fits including more data
- Note for benchmark fit $\Delta\chi^2 = 1$, while for global fit $\Delta\chi^2_{\rm mrst} = 50, \Delta\chi^2_{\rm cteq} = 100$ → Statistical treatment is dataset dependent, also input parametrizations are different

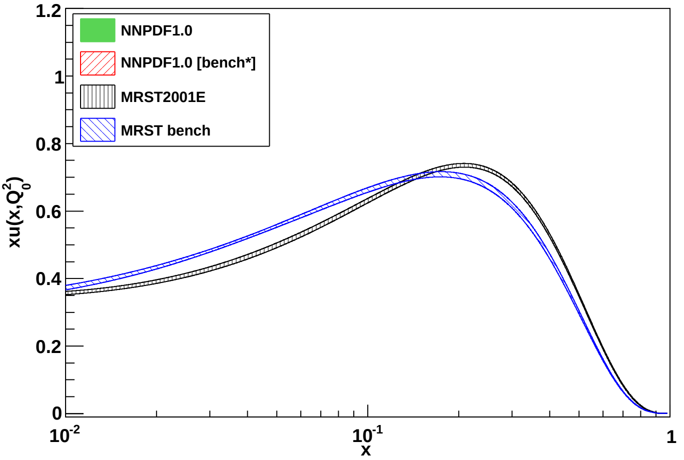
Benchmark partons

Compare $u(x, Q^2 = 2\ \text{GeV}^2)$ from MRST2001 global PDF determination ...



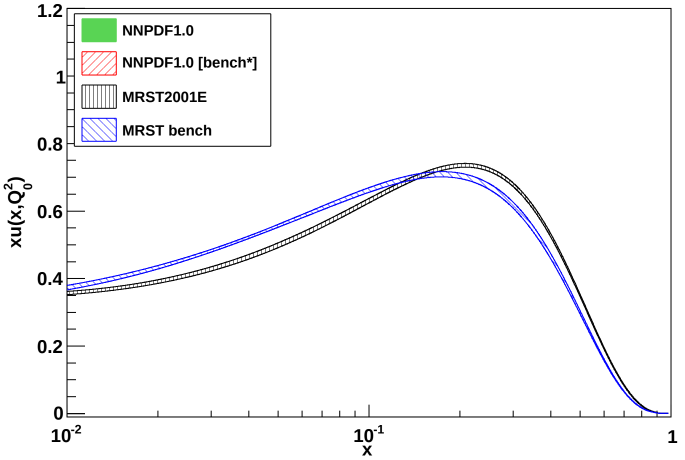
Benchmark partons

... with MRST HERA-LHC benchmark partons



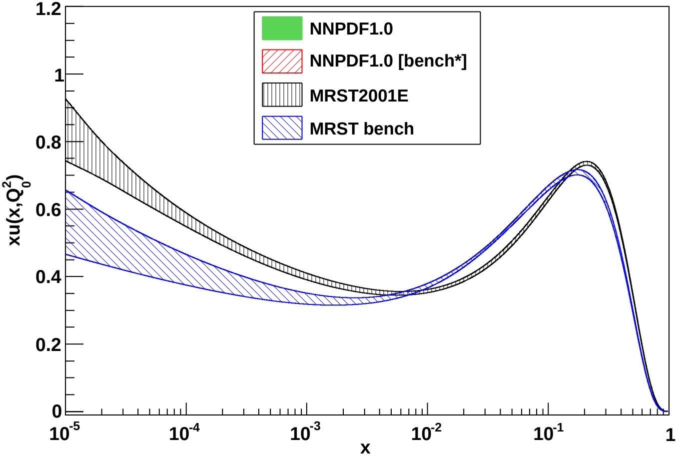
Benchmark partons

PDFs inconsistent by many σ! in data region



Benchmark partons

Similar inconsistencies in the extrapolation region



Problems in standard PDF determination approach

- Summary of HERA-LHC benchmark fit: Benchmark partons do not agree with global fit partons within uncertainties
- Implications → Both the PDF input parametrization (and flavour assumptions) and the statistical treatment (value of $\Delta\chi^2$) need to be tuned to experimental data set for standard approach
- Situation not satisfactory, specially problematic to predict behaviour of PDFs in extrapolation regions like for the LHC
- Global fits introduce large tolerances → Error blow-up by a factor $S = \sqrt{\Delta\chi^2/2.7}$ (B. Cousins, PDF4LHC) →$S_{\rm cteq} \sim 6$, $S_{\rm mstw} \sim 4.5$ both in input measurements and in output PDFs
- Need statistically reliable way to determine if such large values of S are indeed mandatory. Note $\Delta\chi^2 \sim 1$ in DIS+DY fits (Alekhin)

ISMD 2008 DESY

THE NNPDF APPROACH

The NNPDF approach

- Generate $N_{\rm rep}$ Monte Carlo replicas $F_i^{(\rm art)(k)}$ of the original data $F_i^{(\rm exp)}$

$$F_i^{(\rm art)(k)} = \left(1 + r_N^{(k)}\sigma_N\right)\left(F_i^{(\rm exp)} + \sum_{p=1}^{N_{\rm sys}} r_p^{(k)}\sigma_{i,p} + r_i^{(k)}\sigma_{i,s}\right)$$

- Evolve each PDF parametrized with Neural Nets $q_\alpha^{(\rm net)(k)}(x, Q_0^2)$

$$F_i^{(\rm net)(k)}(x, Q^2) = C_{i\alpha}(x, \alpha(Q^2)) \otimes q_\alpha^{(\rm net)(k)}\left(x, Q^2\right)$$

- Training: Minimize χ^2 using Genetic Algs. + Dynamical Stopping:

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

- Set of trained NNs $\rightarrow$ Representation of the PDFs probability density

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

The NNPDF approach

- Generate $N_{\rm rep}$ Monte Carlo replicas $F_i^{(\rm art)(k)}$ of the original data $F_i^{(\rm exp)}$

$$F_i^{(\rm art)(k)} = \left(1 + r_N^{(k)}\sigma_N\right)\left(F_i^{(\rm exp)} + \sum_{p=1}^{N_{\rm sys}} r_p^{(k)}\sigma_{i,p} + r_i^{(k)}\sigma_{i,s}\right)$$

- Evolve each PDF parametrized with Neural Nets $q_\alpha^{(\rm net)(k)}(x, Q_0^2)$

$$F_i^{(\rm net)(k)}(x, Q^2) = C_{i\alpha}(x, \alpha(Q^2)) \otimes q_\alpha^{(\rm net)(k)}\left(x, Q^2\right)$$

- Training: Minimize χ^2 using Genetic Algs. + Dynamical Stopping:

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

- Set of trained NNs $\to$ Representation of the PDFs probability density

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

The NNPDF approach

- Generate $N_{\rm rep}$ Monte Carlo replicas $F_i^{(\rm art)(k)}$ of the original data $F_i^{(\rm exp)}$

$$F_i^{(\rm art)(k)} = \left(1 + r_N^{(k)}\sigma_N\right)\left(F_i^{(\rm exp)} + \sum_{p=1}^{N_{\rm sys}} r_p^{(k)}\sigma_{i,p} + r_i^{(k)}\sigma_{i,s}\right)$$

- Evolve each PDF parametrized with Neural Nets $q_\alpha^{(\rm net)(k)}(x, Q_0^2)$

$$F_i^{(\rm net)(k)}(x, Q^2) = C_{i\alpha}(x, \alpha(Q^2)) \otimes q_\alpha^{(\rm net)(k)}\left(x, Q^2\right)$$

- Training: Minimize χ^2 using Genetic Algs. + Dynamical Stopping:

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

- Set of trained NNs $\rightarrow$ Representation of the PDFs probability density

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

The NNPDF approach

- Generate $N_{\rm rep}$ Monte Carlo replicas $F_i^{(\rm art)(k)}$ of the original data $F_i^{(\rm exp)}$

$$F_i^{(\rm art)(k)} = \left(1 + r_N^{(k)}\sigma_N\right)\left(F_i^{(\rm exp)} + \sum_{p=1}^{N_{\rm sys}} r_p^{(k)}\sigma_{i,p} + r_i^{(k)}\sigma_{i,s}\right)$$

- Evolve each PDF parametrized with Neural Nets $q_\alpha^{(\rm net)(k)}(x, Q_0^2)$

$$F_i^{(\rm net)(k)}(x, Q^2) = C_{i\alpha}(x, \alpha(Q^2)) \otimes q_\alpha^{(\rm net)(k)}\left(x, Q^2\right)$$

- Training: Minimize χ^2 using Genetic Algs. + Dynamical Stopping:

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

- Set of trained NNs → Representation of the PDFs probability density

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

The NNPDF approach

- Generate $N_{\rm rep}$ Monte Carlo replicas $F_i^{({\rm art})(k)}$ of the original data $F_i^{(\exp)}$

$$F_i^{({\rm art})(k)} = \left(1 + r_N^{(k)}\sigma_N\right)\left(F_i^{(\exp)} + \sum_{p=1}^{N_{\rm sys}} r_p^{(k)}\sigma_{i,p} + r_i^{(k)}\sigma_{i,s}\right)$$

- Evolve each PDF parametrized with Neural Nets $q_\alpha^{({\rm net})(k)}(x, Q_0^2)$

$$F_i^{({\rm net})({\rm k})}(x, Q^2) = C_{i\alpha}(x, \alpha(Q^2)) \otimes q_\alpha^{({\rm net})(k)}\left(x, Q^2\right)$$

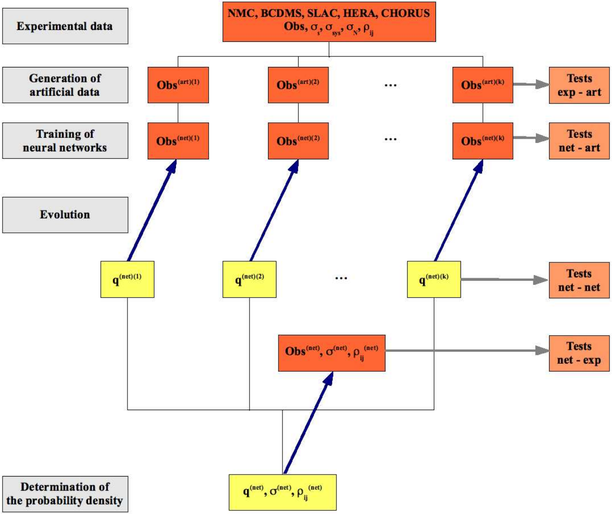
- Training: Minimize χ^2 using Genetic Algs. + Dynamical Stopping:

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

- Set of trained NNs → Representation of the PDFs probability density

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

The NNPDF approach



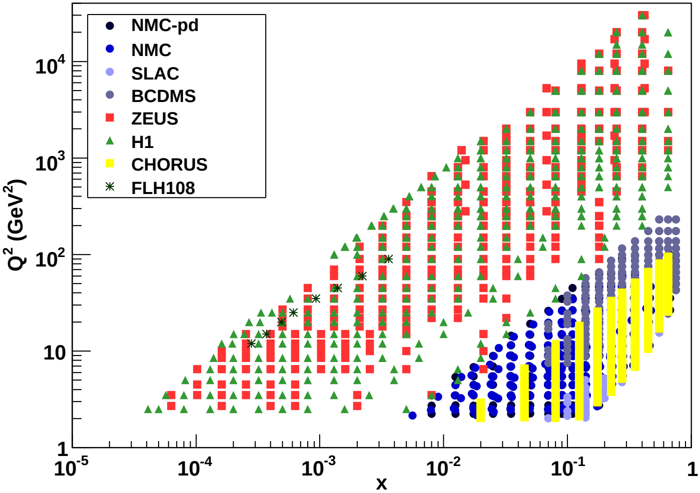
ISMD 2008

THE NNPDF DIS ANALYSIS: NNPDF1.0

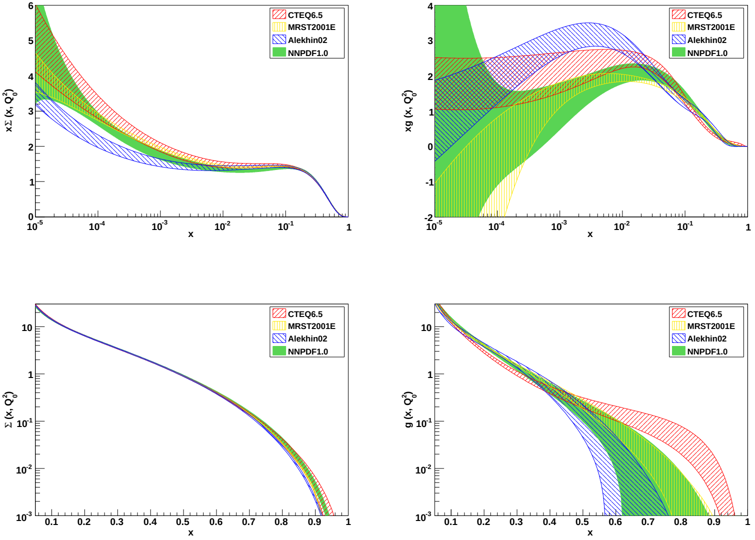
NNPDF1.0 - details

- NNPDF1.0 $\to$ PDF set determination from all relevant DIS experimental data (~ 3000 data points)
- 5 PDFs ($\Sigma(x), V(x), T_3(x), \Delta_S(x)$ and $g(x)$) parametrized with NNs at $Q_0^2 = 2$ GeV^2 (37 free params each)
- Valence and momentum sum rules incorporated
- Flavour assumptions $\to s(x) = \bar{s}(x) = C_s/2\left(\bar{u}(x) + \bar{d}(x)\right)$
- NLO evolution with ZM-VFN scheme for heavy quarks

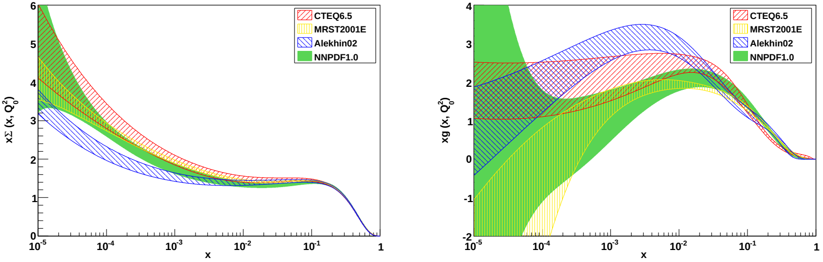
Data set


NMC-pd
NMC
SLAC
BCDMS
ZEUS
H1
CHORUS
FLH108
Q^2 (GeV^2)
10^4
10^3
10^2
10
1
10^{-5}
10^{-4}
10^{-3}
10^{-2}
10^{-1}
1
x

Results - Singlet PDFs

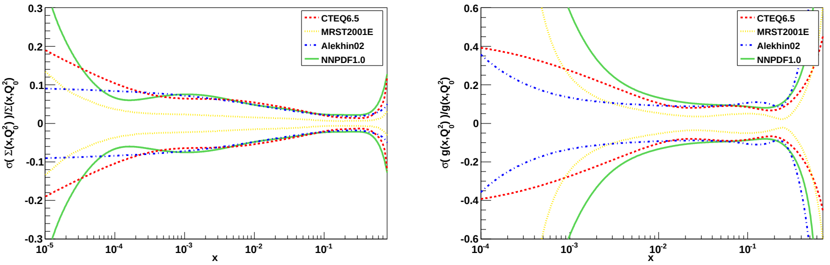


Results - Singlet PDFs



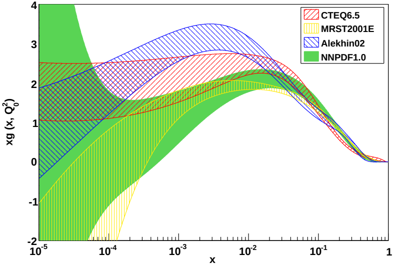
- NNPDF1.0 uncertainties faithfully determined
- PDF error larger than other PDF sets in some regions (extrapolation), smaller in others (not artificially inflated by large $\Delta\chi^2 \sim 50/100$)
- In general close to CTEQ6.5 in data region

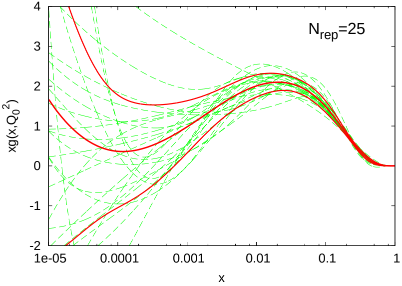
Results - Singlet PDFs



- NNPDF1.0 uncertainties faithfully determined
- PDF error larger than other PDF sets in some regions (extrapolation), smaller in others (not artificially inflated by large $\Delta\chi^2 \sim 50/100$)
- In general close to CTEQ6.5 in data region

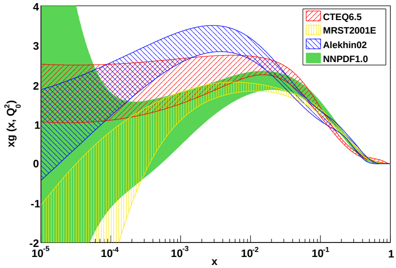
Results - Singlet PDFs

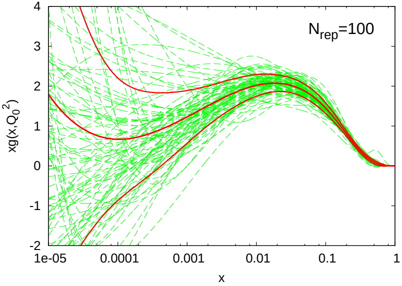




Individual PDF replicas (*i.e.* the gluon) span uncertainty range free from functional form biases ($N_{\rm rep} = 25$)

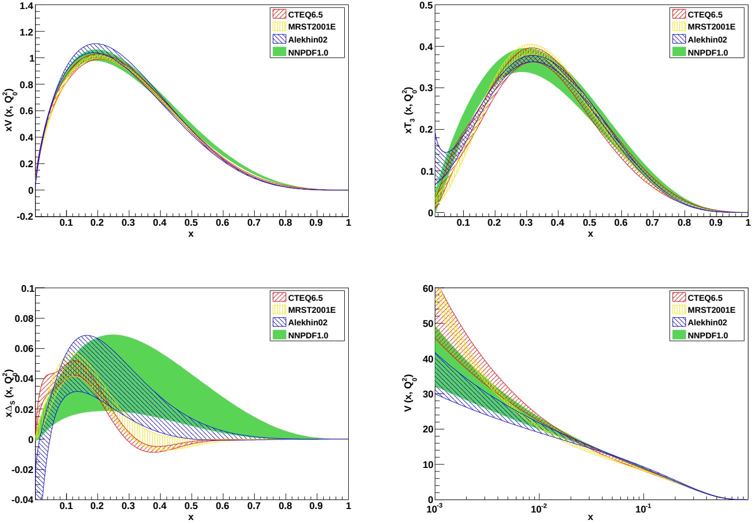
Results - Singlet PDFs



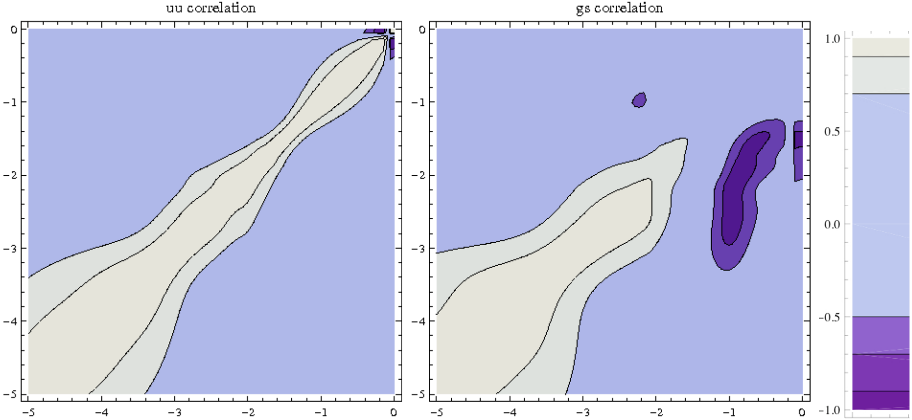


Individual PDF replicas (*i.e.* the gluon) span uncertainty range free from functional form biases ($N_{\rm rep} = 100$)

Results - Valence PDFs



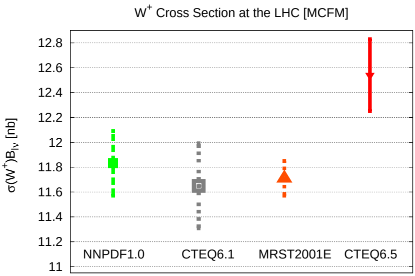
Parton correlations

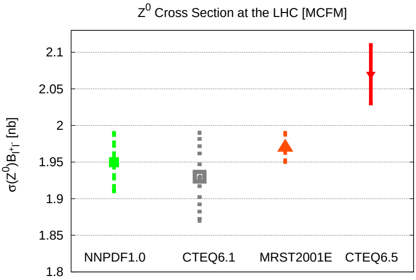


Compute parton-parton correlations using textbook statistics

$$\rho\left[q(x_1,Q_1^2)\widetilde{q}(x_2,Q_2^2)\right] = \frac{\left\langle q(x_1,Q_1^2)\widetilde{q}(x_2,Q_2^2)\right\rangle_{\rm rep} - \left\langle q(x_1,Q_1^2)\right\rangle_{\rm rep}\left\langle \widetilde{q}(x_2,Q_2^2)\right\rangle_{\rm rep}}{\sigma_q(x_1,Q_1^2)\sigma_{\widetilde{q}}(x_2,Q_2^2)}$$

Results - Predictions for LHC





	$\sigma_{W+}\mathcal{B}_{l+\nu_l}$	$\Delta\sigma_{W+}/\sigma_{W+}$	$\sigma_{W-}\mathcal{B}_{l-\nu_l}$	$\Delta\sigma_{W-}/\sigma_{W-}$	$\sigma_Z\mathcal{B}_{l+l-}$	$\Delta\sigma_Z/\sigma_Z$
NNPDF1.0	11.83 ± 0.26	2.2%	8.41 ± 0.20	2.4%	1.95 ± 0.04	2.1%
CTEQ6.1	11.65 ± 0.34	2.9%	8.56 ± 0.26	3.0%	1.93 ± 0.06	3.1%
MRST01	11.71 ± 0.14	1.2%	8.70 ± 0.10	1.1%	1.97 ± 0.02	1.0%
CTEQ6.5	12.54 ± 0.29	2.3%	9.19 ± 0.22	2.4%	2.07 ± 0.04	1.9%

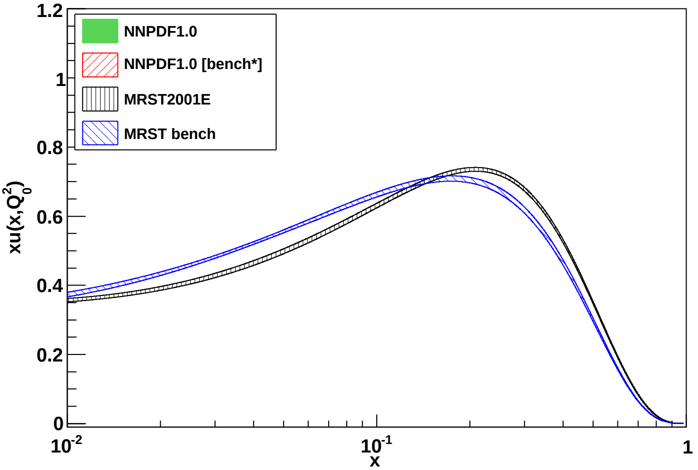
BENCHMARK PARTONS REVISITED

PDF benchmark analysis

- Does the NNPDF approach solve the problem with MRST benchmark partons?
- Compare NNPDF1.0 partons with a PDF set obtained from the reduced data set of the HERA-LHC workshop
- For a complete NNPDF version of the HERA-LHC PDF benchmark, see A. Piccione's talks at PDF4LHC meetings and HERA-LHC workshop proceedings

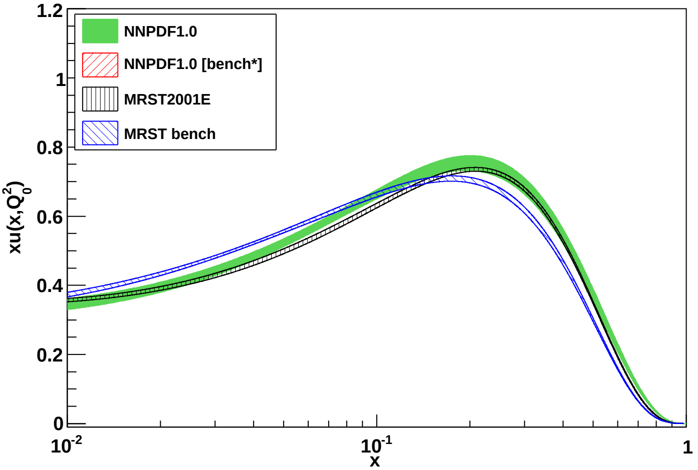
Benchmark partons revisited

PDFs inconsistent by many σ! in data region in standard approach ...



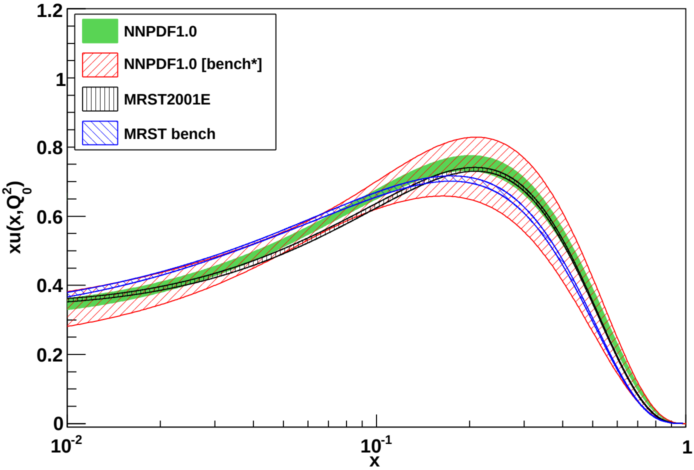
Benchmark partons revisited

... but not within the NNPDF approach: Full DIS fit

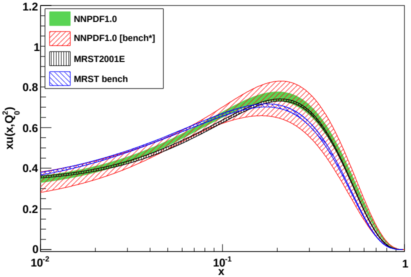


Benchmark partons revisited

... but not within the NNPDF approach: Benchlike fit



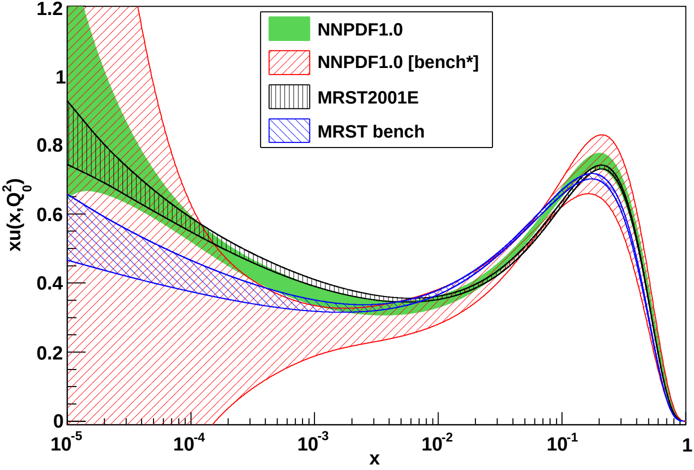
Benchmark partons revisited



- NNPDF1.0 consistent with MRST global fit
- NNPDF benchlike consistent with both NNPDF1.0 and MRST global and benchmark fits
- Error determination understimated in standard aproach to PDF determination (central values ok)

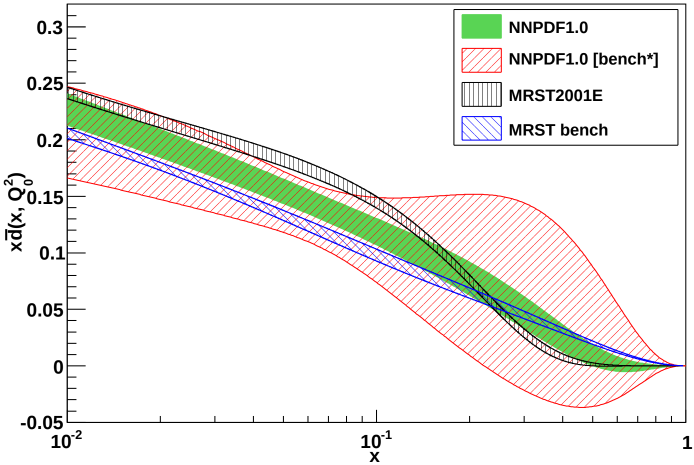
Benchmark partons revisited

Problems also cured in (low-x) extrapolation region



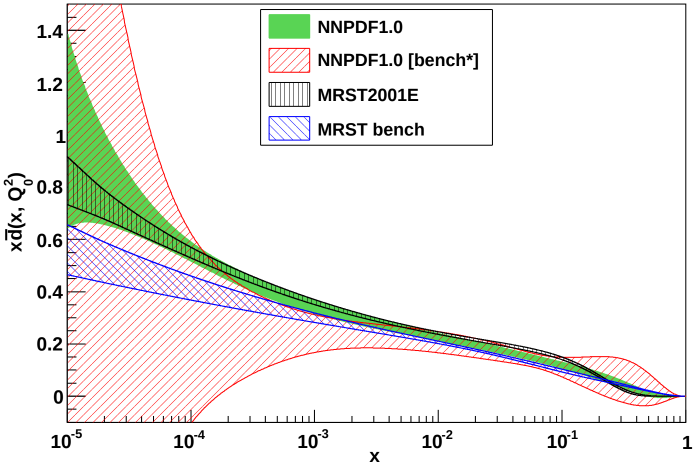
Benchmark partons revisited

Same for other PDFs - $\bar{d}(x, Q_0^2)$ in data region



Benchmark partons revisited

Same for other PDFs - $\bar{d}(x, Q_0^2)$ in extrapolation region



OUTLOOK

Outlook

- NNPDF1.0 $\longrightarrow$ DIS NNPDF set completed and available from the LHAPDF interface
- Faithful determination of uncertainties $\longrightarrow$ Suited to to precision LHC physics
- Work in progress $\longrightarrow$ More general flavour assumptions ($s(x)$ & $\bar{s}(x)$), addition of hadronic data and heavy quark effects, and detailed studies of PDF uncertainty impact on LHC physics

Thanks for your attention!

Outlook

- NNPDF1.0 → DIS NNPDF set completed and available from the LHAPDF interface
- Faithful determination of uncertainties → Suited to to precision LHC physics
- Work in progress → More general flavour assumptions ($s(x)$ & $\bar{s}(x)$), addition of hadronic data and heavy quark effects, and detailed studies of PDF uncertainty impact on LHC physics

Thanks for your attention!

EXTRA MATERIAL

Interpretation of benchmark PDFs

R. Thorne, HERA-LHC 2006 proceedings

errors, but these are relatively small. However, the partons extracted using a very limited data set are completely incompatible, even allowing for the uncertainties, with those obtained from a global fit with an identical treatment of errors and a minor difference in theoretical procedure. This implies that the inclusion of more data from a variety of different experiments moves the central values of the partons in a manner indicating either that the different experimental data are inconsistent with each other, or that the theoretical framework is inadequate for correctly describing the full range of data. To a certain extent both explanations are probably true. Some data sets are not entirely consistent with each other (even if they are seemingly equally reliable). Also, there are a wide variety of reasons why NLO perturbative QCD might require modification for some data sets, or in some kinematic regions [89]. Whatever the reason for the inconsistency between the MRST benchmark partons and the MRST01 partons, the comparison exhibits the dangers in extracting partons from a very limited set of data and taking them seriously. It also clearly illustrates the problems in determining the true uncertainty on parton distributions.

Parametrization independence

Quantify statistical differences between PDF sets $\longrightarrow$
Distances between two probability distributions which describe two sets of PDFs (*i.e.* the gluon $\{g^{(1)}_{ik} = g^{(1)}_k(x_i, Q_0^2)\}$):

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

$\langle d[g] \rangle \longrightarrow$ Distance between PDF in units of the variance of expectation value $\langle g \rangle$

For statistically equivalent PDF sets: $\langle d[g] \rangle \sim \langle d[\sigma_g] \rangle \sim 1$

Parametrization independence

Check stability for NNs arch. from 2-5-3-1 to 2-4-3-1 (6 params less per PDF)

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

Dynamical stopping

In a standard fit, look for minimum χ^2 for given parametrization.

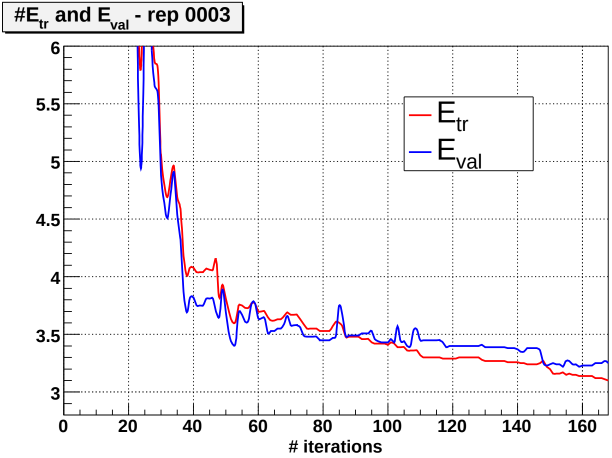
- If basis too large $\rightarrow$ convergence never reached
- If basis too small $\rightarrow$ parametrization bias

How can one obtain an unbiased compromise? For NNs, smoothness decreases as fit quality improves $\rightarrow$ Stop before fitting statistical noise (overlearning).

1. Divide the data set into training and validation sets
2. Minimize χ^2 of training set, monitor χ^2 of validation set
3. Stop minimization when validation χ^2 begins to rise (overlearning)

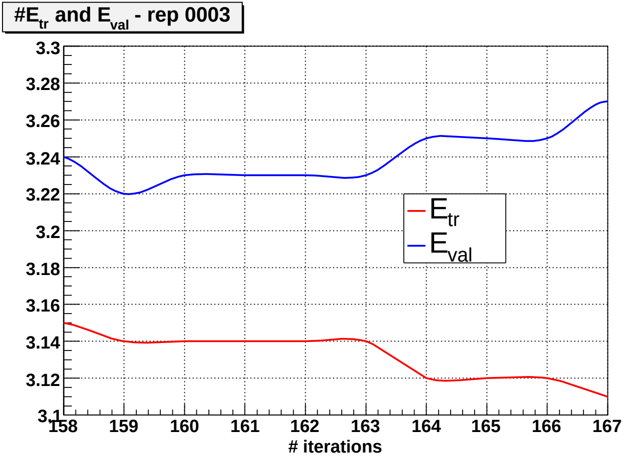
Dynamical stopping

Stop minimization when validation χ^2 begins to rise (overlearning)



Dynamical stopping

Stop minimization when validation χ^2 begins to rise (overlearning)



Problems in standard PDF determination approach

- Consensus (PDF4LHC workshop): serious problem in PDF fits
- Problem summarized by the HERA-LHC benchmark fit: Benchmark partons do not agree with global fit partons within errors
- Implications → either experiments are incompatible, or parametrizations not flexible enough, or both
- Global fit solution → Error blow-up by a factor $S = \sqrt{\Delta\chi^2/2.7}$ (B. Cousins, PDF4LHC) →$S_{\rm cteq} \sim 6$, $S_{\rm mstw} \sim 4.5$ both in input measurements and in output PDFs (very large!)
- Need statistically reliable way to determine if such large values of S are indeed mandatory. Note $\Delta\chi^2 \sim 1$ in DIS+DY fits (Alekhin)

Experimental data set

Experiment	Set	$N_{\rm dat}$	$x_{\rm min}$	$x_{\rm max}$	$Q^2_{\rm min}$	$Q^2_{\rm max}$	$\sigma_{\rm tot}$ (%)	F	Ref.
SLAC									
	SLACp	211 (47)	.07000	.85000	0.6	29.	3.6	F_2^p	[51]
	SLACd	211 (47)	.07000	.85000	0.6	29.	3.2	F_2^d	[51]
BCDMS									
	BCDMSp	351 (333)	.07000	.75000	7.5	230.	5.5	F_2^p	[47]
	BCDMSd	254 (248)	.07000	.75000	8.8	230.	6.6	F_2^d	[48]
NMC		288 (245)	.00350	.47450	0.8	61.	5.0	F_2^p	[50]
NMC-pd		260 (153)	.00150	.67500	0.2	99.	2.1	F_2^d/F_2^p	[49]
ZEUS									
	Z97lowQ2	80	.00006	.03200	2.7	27.	4.9	$\tilde{\sigma}^{NC,e^+}$	[56]
	Z97NC	160	.00080	.65000	35.0	20000.	7.7	$\tilde{\sigma}^{NC,e^+}$	[56]
	Z97CC	29	.01500	.42000	280.0	17000.	34.2	$\tilde{\sigma}^{CC,e^+}$	[57]
	Z02NC	92	.00500	.65000	200.0	30000.	13.2	$\tilde{\sigma}^{NC,e^-}$	[58]
	Z02CC	26	.01500	.42000	280.0	30000.	40.2	$\tilde{\sigma}^{CC,e^-}$	[59]
	Z03NC	90	.00500	.65000	200.0	30000.	9.1	$\tilde{\sigma}^{NC,e^+}$	[60]
	Z03CC	30	.00800	.42000	280.0	17000.	31.0	$\tilde{\sigma}^{CC,e^+}$	[61]
H1									
	H197mb	67 (55)	.00003	.02000	1.5	12.	4.9	$\tilde{\sigma}^{NC,e^+}$	[52]
	H197lowQ2	80	.00016	.20000	12.0	150.	4.2	$\tilde{\sigma}^{NC,e^+}$	[52]
	H197NC	130	.00320	.65000	150.0	30000.	13.3	$\tilde{\sigma}^{NC,e^+}$	[53]
	H197CC	25	.01300	.40000	300.0	15000.	29.8	$\tilde{\sigma}^{CC,e^+}$	[53]
	H199NC	126	.00320	.65000	150.0	30000.	15.5	$\tilde{\sigma}^{NC,e^-}$	[54]
	H199CC	28	.01300	.40000	300.0	15000.	27.6	$\tilde{\sigma}^{CC,e^-}$	[54]
	H199NChy	13	.00130	.01050	100.0	800.	9.2	$\tilde{\sigma}^{NC,e^-}$	[55]
	H100NC	147	.00131	.65000	100.0	30000.	10.4	$\tilde{\sigma}^{NC,e^+}$	[55]
	H100CC	28	.01300	.40000	300.0	15000.	21.8	$\tilde{\sigma}^{CC,e^+}$	[55]
CHORUS									
	CHORUSν	607 (471)	.02000	.65000	0.3	95.	11.2	$\tilde{\sigma}^{\nu}$	[63]
	CHORUS$\bar{\nu}$	607 (471)	.02000	.65000	0.3	95.	18.7	$\tilde{\sigma}^{\bar{\nu}}$	[63]
FLH108		8	.00028	.00360	12.0	90.	69.2	F_L	[62]
Total		3948 (3161)							

Statistical estimators

$\chi^2_{\rm tot}$	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
$\left\langle \sigma^{(\rm exp)} \right\rangle_{\rm dat}$	$5.6\ 10^{-2}$
$\left\langle \sigma^{(\rm net)} \right\rangle_{\rm dat}$	$1.4\ 10^{-2}$
$\left\langle \rho^{(\rm exp)} \right\rangle_{\rm dat}$	0.15
$\left\langle \rho^{(\rm net)} \right\rangle_{\rm dat}$	0.40
$\left\langle {\rm cov}^{(\rm exp)} \right\rangle_{\rm dat}$	$1.0\ 10^{-3}$
$\left\langle {\rm cov}^{(\rm net)} \right\rangle_{\rm dat}$	$1.6\ 10^{-4}$

Dependence with preprocessing

Data region								
	$n_v = 0.1$	$n_v = 0.5$	$m_v = 2$	$m_v = 4$	$n_s = 0.8$	$n_s = 1.6$	$m_s = 2$	$m_s = 4$
$\Sigma(x, Q_0^2)$								
$\langle d[q] \rangle$	1.34	1.25	1.37	2.14	1.72	1.38	1.45	1.64
$\langle d[\sigma] \rangle$	1.45	1.44	1.25	1.44	2.03	2.66	0.95	1.35
$g(x, Q_0^2)$								
$\langle d[q] \rangle$	1.31	1.30	2.69	1.15	3.06	2.08	1.20	1.74
$\langle d[\sigma] \rangle$	1.34	1.60	1.56	1.37	3.21	2.44	0.98	1.72
$T_3(x, Q_0^2)$								
$\langle d[q] \rangle$	1.97	2.48	8.35	9.74	1.31	3.23	1.03	1.41
$\langle d[\sigma] \rangle$	1.10	1.47	1.98	1.53	1.10	2.66	1.76	1.99
$V(x, Q_0^2)$								
$\langle d[q] \rangle$	11.03	1.55	3.61	5.60	0.94	2.12	1.25	3.54
$\langle d[\sigma] \rangle$	3.57	4.74	4.04	3.09	1.03	1.10	0.66	1.98
$\Delta_S(x, Q_0^2)$								
$\langle d[q] \rangle$	2.00	2.29	7.51	2.36	1.14	1.70	0.76	0.92
$\langle d[\sigma] \rangle$	1.25	5.20	1.17	3.50	1.00	1.98	0.97	2.05
Extrapolation								
	$n_v = 0.1$	$n_v = 0.5$	$m_v = 2$	$m_v = 4$	$n_s = 0.8$	$n_s = 1.6$	$m_s = 2$	$m_s = 4$
$\Sigma(x, Q_0^2)$								
$\langle d[q] \rangle$	1.06	1.69	1.49	1.84	7.72	4.67	0.87	3.15
$\langle d[\sigma] \rangle$	1.12	1.84	2.11	1.52	2.47	3.66	0.82	2.34
$g(x, Q_0^2)$								
$\langle d[q] \rangle$	1.41	2.32	2.33	1.34	1.62	4.73	1.04	3.49
$\langle d[\sigma] \rangle$	1.41	1.86	1.95	1.30	2.15	2.72	0.81	2.38
$T_3(x, Q_0^2)$								
$\langle d[q] \rangle$	1.71	2.70	7.40	1.60	1.36	2.37	0.78	0.91
$\langle d[\sigma] \rangle$	4.83	4.54	2.89	5.09	1.00	1.65	0.92	1.26
$V(x, Q_0^2)$								
$\langle d[q] \rangle$	14.85	3.23	3.75	2.55	0.86	2.52	1.26	1.34
$\langle d[\sigma] \rangle$	2.65	5.08	3.94	2.78	1.20	0.87	0.62	2.25
$\Delta_S(x, Q_0^2)$								
$\langle d[q] \rangle$	1.25	2.50	7.75	2.48	1.09	1.47	1.09	0.83
$\langle d[\sigma] \rangle$	1.80	2.85	1.50	2.28	0.90	2.01	0.90	1.64