NNPDF partons for LHC analyses

Strangeness in the nucleon: solving the NuTeV anomaly.

Maria Ubiali

School of Physics, University of Edinburgh

London Workshop on SM discoveries with early LHC data
University College London
London, 30 March 2009

Work in collaboration

NNPDF collaboration

R.D.Ball[1], L.Del Debbio[1], S.Forte[2], A.Guffanti[3], J.I.Latorre[4], A. Piccione[2], J. Rojo[2], M.U.[1]

[1] PPT Group, School of Physics, University of Edinburgh
[2] Dipartimento di Fisica, Università di Milano
[3] Physikalisches Institut, Albert-Ludwigs-Universität Freiburg
[4] Departament d'Estructura i Constituents de la Matèria, Universitat de Barcelona

NNPDF collaboration, Nucl. Phys. B 809, 1 (2009) [arXiv:0808.1231] NNPDF1.0
NNPDF collaboration, [arXiv:0811.2288] NNPDF1.1
NNPDF collaboration, in preparation NNPDF1.2
NNPDF collaboration, in preparation NNPDF2.0

Outline

1. Introduction
   - Parton fit
   - NNPDF approach: the main ingredients
2. Results
   - NNPDF1.0
   - NNPDF1.1
   - NNPDF1.2
   - NNPDF2.0
3. Conclusions

Outline

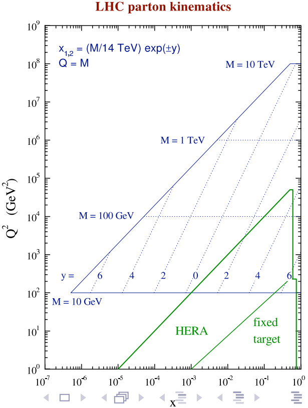
Parton Distribution Functions

- Factorization Theorem ($Q^2 \gg \Lambda^2_{\rm QCD}$):

$$\frac{d\sigma_H}{dX} = \sum_{a,b} \int dx_1 dx_2 f_a(x_1, \mu_f) f_b(x_2, \mu_f) \otimes \frac{d\hat{\sigma}}{dX}(\alpha_s(\mu_r), \mu_r, \mu_f, x_1, x_2, Q^2)$$

- DGLAP equations:

$$\frac{d}{dt}\begin{pmatrix} q \\ g \end{pmatrix} = \frac{\alpha_s}{2\pi}\begin{pmatrix} P_{qq} & P_{qg} \\ P_{gq} & P_{gg} \end{pmatrix} \otimes \begin{pmatrix} q \\ g \end{pmatrix} + O(\alpha_s^2)$$



PDFs and their associated uncertainties will play a crucial role in the full exploitation of the LHC physics potential.
For some processes PDFs errors will provide dominant contribution to systematic uncertainties.

Parton fits

* Need robust input for analyses at LHC.
* Need statistically reliable interpretation for PDFs error bars.

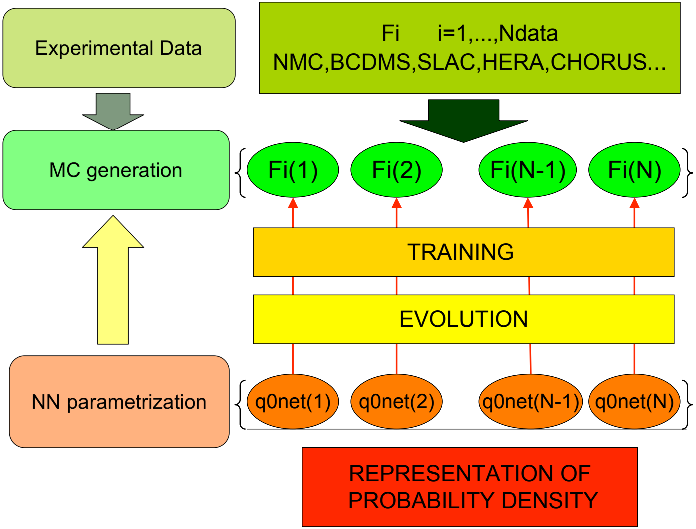
NNPDF approach

Determination of unbiased PDFs with faithful estimation of their uncertainties.

$$\langle \mathcal{F}[f_i(x)] \rangle = \int [\mathcal{D}f_i]\, \mathcal{F}[f_i(x)] \mathcal{P}[f_i(x)] \to \frac{1}{N_{\rm rep}} \sum_{k=1}^{N_{\rm rep}} \mathcal{F}[f_i^{(k)(\rm net)}(x)]$$

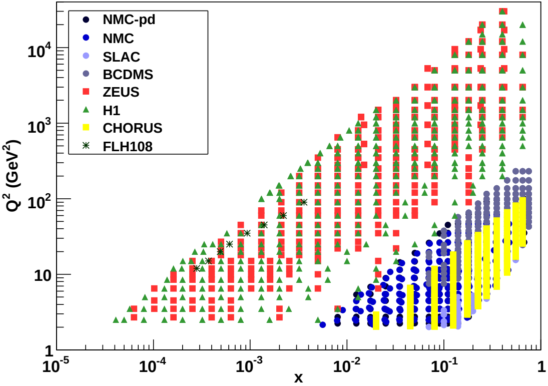
* The measure $\mathcal{P}[f_i(x)]$ in space of PDFs is determined with a MC method.
* Use all information contained in experiments.
* Redundant parametrization of PDFs: reduce bias.
* Statistic estimators to assess errors, correlations, stability and size of systematics.
* Results show to behave as expected when comparing full and benchmark analyses [HERA-LHC and PDF4LHC workshops]

NNPDF approach



Outline

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.

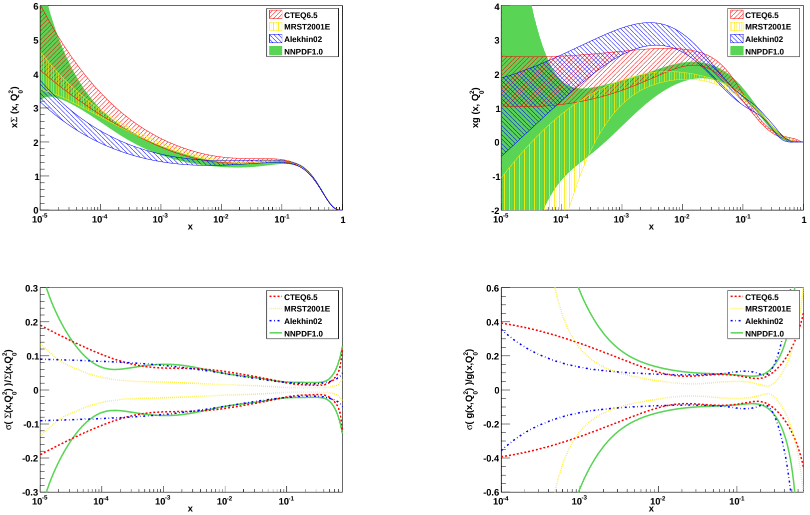
NNPDF1.0: Parametrization

Parametrization of 5 combinations of PDFs at $Q_0^2 = 2\ \mathrm{GeV}^2$

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: Partons



NNPDF1.1: A consistency check

- NNPDF1.0: flavor assumptions, symmetric strange sea proportional to non strange sea according to $C_s \sim 0.5$ suggested by neutrino DIS data.

$$s(x) = \bar{s}(x) \qquad \bar{s}(x) = \frac{C_s}{2}(\bar{u}(x) + \bar{d}(x))$$

- NNPDF1.1: independent parametrization of the strange content of the nucleon.

Total strangeness : $s^+(x) \equiv (s(x) + \bar{s}(x))/2 \longmapsto \mathrm{NN}_{(s+)}(x)$ 2-5-3-1 37 pars

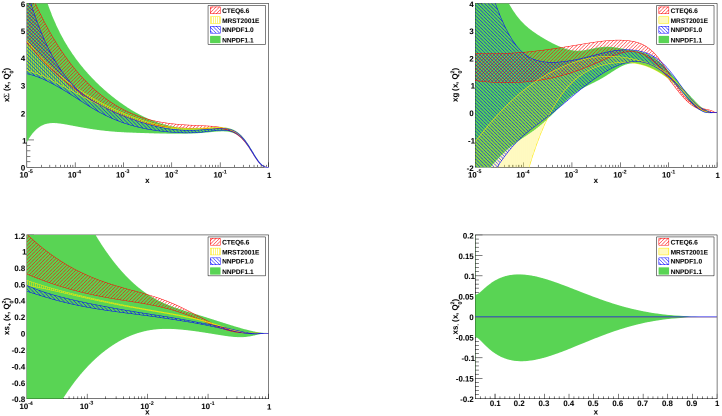
Strangeness valence : $s^-(x) \equiv (s(x) - \bar{s}(x))/2 \longmapsto \mathrm{NN}_{(s-)}(x)$ 2-5-3-1 37 pars

- Added two unconstrained PDFs.

185 $\rightarrow$ **259** parameters

- Randomized preprocessing.

NNPDF1.1: A consistency check

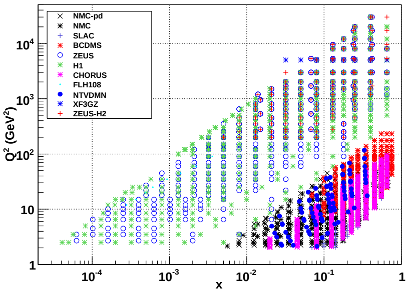


- Large uncertainty for strange PDFs. Bigger uncertainties for singlet PDFs.
- Same χ^2 and statistical features of the fit. Same gluon shape and error band.
- Check of stability and consistency of our statistically-sound approach.

NNPDF1.2: Constrain the strange distribution

- Direct determination of both s and $\bar{s}$ allowed by recent NuTeV data, via

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



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

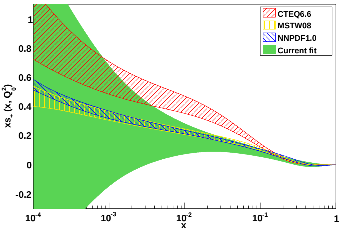
* Neutrino and anti-neutrino dimuon production from NuTeV.
* HERA-II ZEUS data on NC and CC reduced xsec at large-Q^2.
* HERA-II ZEUS data on $xF_3^{\gamma Z}$.

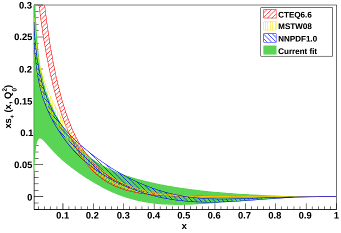
NNPDF1.2: Theoretical issues

- A theoretical constraint on strange PDFs comes from valence sum rule, enforced to a 10^{-7} accuracy without introducing bias on strange shape.
- Mass effects: many data have $Q^2 \gtrapprox m_c^2$, charm mass effects are important for NuTeV dimuon data.
- We implemented the I-ZM-VFN scheme [Thorne, Tung, ArXiv:0809.0714].
- Massless coefficients with correct kinematics of heavy quark production which account for dominant mass effects of the full GM-VFN treatment [Nadolsky, Tung, ArXiv:0903.2667]
- NuTeV dimuon data (and CHORUS data) is taken on a nuclear target: nuclear corrections applied according to various models and study of their impact.[Hirai, Kumano, Nagai - de Florain, Sassot]

NNPDF1.2: Strangeness determination (preliminary results)

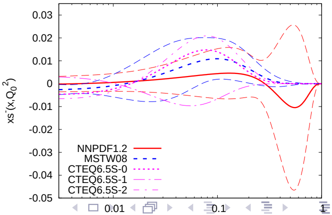
Total strangeness (log scale) ↓ (lin scale)→





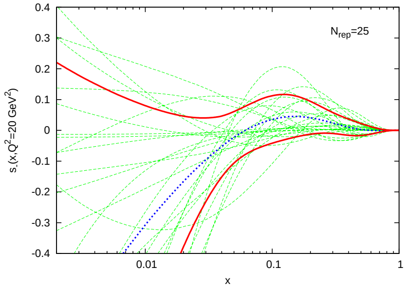
Strange valence →

	N_{dat}	χ^2
Global	3382	1.29
NuTeV $\nu + \bar{\nu}$	84	0.60
NuTeV ν	43	0.45
NuTeV $\bar{\nu}$	41	0.71

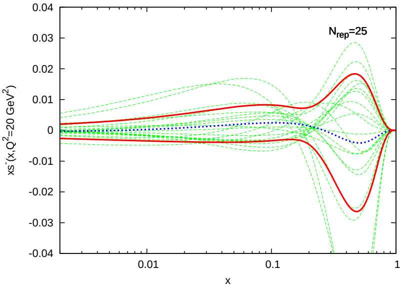


NNPDF1.2: Strangeness determination (preliminary results)

Total strangeness ↓



Strange valence ↓



- No bias on the shape or normalization of strange valence and total strange.
- The only constraint comes from strange valence sum rule.
- There must be at least one crossing but neither the crossing point or the sign are enforced by fixed parametrization.
- Faithful determination of uncertainties.

NNPDF1.2: Impact on NuTeV anomaly (preliminary results)

- Define second momentum of PDFs f: $[F] = \int_0^1 dx\, x\, f(x, Q^2)$.
- Discrepancy $\geq 3\sigma$ between indirect and direct determination from NuTeV measurement assuming $[S^-] = 0$ and isospin symmetry.

EW fit

$\sin^2\theta_W = 0.2223 \pm 0.0002$

NuTeV

$\sin^2\theta_W = 0.2276 \pm 0.0014$

$$\sin^2\theta_W\Big|_{\mathrm{NuTeV}} - \sin^2\theta_W\Big|_{\mathrm{EW}} = 0.0053$$

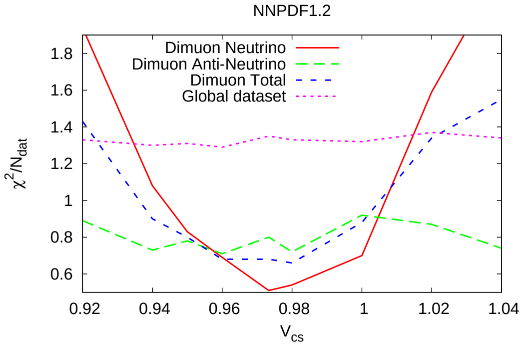
- If we consider $[S^-] \neq 0$:

$$\delta_s \sin^2\theta_W \sim -0.240\frac{[S^-]}{[Q^-]}$$

$$\delta_s \sin^2\theta_W = -0.0005 \pm 0.0096^{\mathrm{PDFs}} \pm sys$$

- Central value compatible with zero, but uncertainty large enough to remove NuTeV anomaly!!!

NNPDF1.2: Direct V_{cs} determination (preliminary results)



CKM global fit

$$V_{cs} = 0.97334 \pm 0.00023, \ \Delta V_{cs} \sim 0.02\%$$

Direct determination-D and B decays

$$V_{cs} = 1.04 \pm 0.06, \qquad \Delta V_{cs} \sim 6\%$$

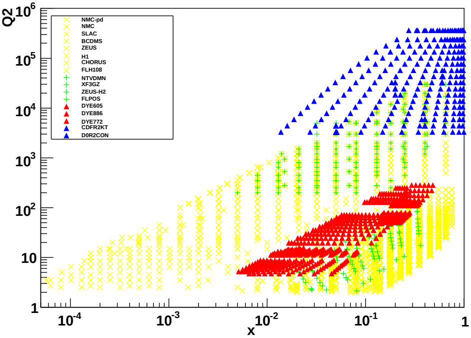
[Amsler et al, Phys. Lett. B67(2008) 1.]

- Commonly assumed that no info on V_{cs} comes from DIS fits due to s uncertainty.
- Best determination from DIS fits $V_{cs} > 0.59$ at 90% confidence level.
- Fit quality for dimuon neutrino data degenerates dramatically when moving from V_{cs}^{CKM}: direct determination from DIS analysis with an uncertainty better than few percents!!!

$$\Delta V_{cs}\Big|_{\rm CKM} \ll \Delta V_{cs}\Big|_{\rm nnpdf1.2} \ll \Delta V_{cs}\Big|_{\rm direct}$$

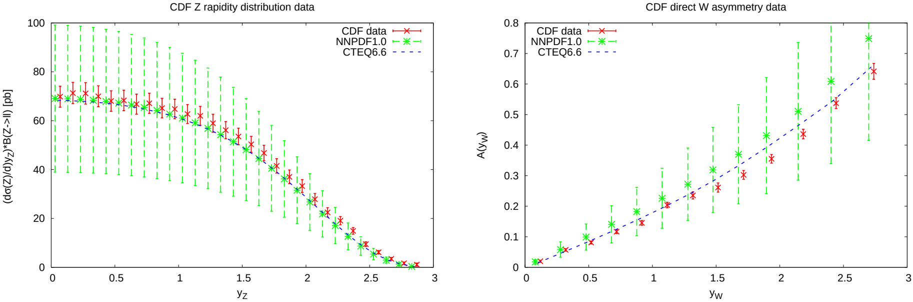
NNPDF2.0: Experimental data

- The inclusion of hadronic data is necessary to constrain large-x gluon behavior, sea quarks, u/d ratio at large x.
- Upcoming NNPDF2.0 is the first NNPDF global fit: inclusion of fixed target Drell-Yan data, Tevatron electroweak gauge boson production, Run II inclusive jet data from Tevatron, 1000 new data.



OBS	Data set
$d\sigma^{\rm DY}/dM^2dy$	E605
$d\sigma^{\rm DY}/dM^2dx_F$	E772
$d\sigma^{\rm DY}/dM^2dx_F$	E886
W asym.	D0/CDF
Z rap. distr.	D0/CDF
incl. $\sigma^{\rm jet}$	CDF(k_T)
incl. $\sigma^{\rm jet}$	D0(cone)

NNPDF2.0: Predictions from previous fits



- Predictions evaluated with NNPDF1.0 error sets.
- Large error bands on predictions, compatible with data.

NNPDF2.0: FastNLO-like evolution

- The NLO computation of hadronic observables might be too slow for parton global fits.
- Many parton fits rely on K-factor approximation, relatively fast.
- K-factor depends on PDFs and it is not always a good approximation.

* NNPDF2.0 includes full NLO calculation of hadronic observables.
* Use available fastNLO interface for jet inclusive cross-sections.[hep-ph/0609285]
* Built up our own **fastNLO-like evolution for Drell-Yan** observables, not available in literature.
* Fast code easy to benchmark versus other slow codes.

Outline

Conclusions

- The first NNPDF1.0 parton set [arXiv:0808.1231] from a comprehensive DIS anaysis is available on the common LHAPDF interface (http://projects.hepforge.org/lhapdf), the NNPDF1.1 is available on the NNPDF website (http://sophia.ecm.ub.es/nnpdf/)
- Inclusion of NuTev data constrains the strange distribution in the upcoming NNPDF1.2 fit.
- Faithful estimation of strange content of the nucleon solves the NuTeV anomaly.
- First direct determination of V_{cs} CKM matrix element from DIS analysis.
- Inclusion of hadronic data (DY, jets, W asymmetry): first global NNPDF2.0 fit.
- Implementation of a full fastNLO-like evolution strategy for hadronic observable, including Drell-Yan.

Outlook

For first global fit results stay tuned to DIS2009 in Madrid.

EXTRA MATERIAL

NNPDF1.2: Normalization and Sum Rules

$$
\begin{aligned}
\Sigma(x,Q_0^2) &= (1-x)^{m_\Sigma} x^{-n_\Sigma} \mathrm{NN}_\Sigma(x) \; , \\
V(x,Q_0^2) &= A_V (1-x)^{m_V} x^{-n_V} \mathrm{NN}_V(x) \; , \\
T_3(x,Q_0^2) &= (1-x)^{m_{T_3}} x^{-n_{T_3}} \mathrm{NN}_{T_3}(x) \; , \\
\Delta_S(x,Q_0^2) &= A_{\Delta_S} (1-x)^{m_{\Delta_S}} x^{-n_{\Delta_S}} \mathrm{NN}_{\Delta_S}(x) \; , \\
g(x,Q_0^2) &= A_g (1-x)^{m_g} x^{-n_g} \mathrm{NN}_g(x) \\
s^+(x,Q_0^2) &= (1-x)^{m_s^+} x^{-n_s^+} NN_{s^+}(x) \\
s^-(x,Q_0^2) &= (1-x)^{m_s^-} x^{-n_s^-} NN_{s^-}(x) - A_{s^-} [x^{r_{s^-}} (1-x)^{m_{t^-}}]
\end{aligned}
$$

Normalization $\longrightarrow$ Fixed by valence and momentum sum rules

$$
\begin{aligned}
\int_0^1 dx \; x \left(\Sigma(x) + g(x)\right) &= 1 \\
\int_0^1 dx \; \left(u(x) - \bar{u}(x)\right) &= 2 \\
\int_0^1 dx \; \left(d(x) - \bar{d}(x)\right) &= 1 \\
\int_0^1 dx \; \left(s(x) - \bar{s}(x)\right) &= 0
\end{aligned}
$$

NNPDF1.2: Sum Rules

- For instance

$$A_V = \frac{3}{\int_0^1 dx\,\left((1-x)^{m_V}\, x^{-n_V}\mathrm{NN}_V(x)\right)}$$

- For the strange sum rule it is slightly different:

$$A_{s^-} = \frac{\Gamma\left(r_{s^-}+t_{s^-}+2\right)}{\Gamma\left(r_{s^-}+1\right)\Gamma\left(t_{s^-}+1\right)}\int_0^1 dx\,\left((1-x)^{m_{s^-}}\, x^{-n_{s^-}}\mathrm{NN}_{s^-}(x)\right)$$

- When $A_{s^-}=0$ the valence sum rule constraint is removed.

Preprocessing exponents

- Polynomial preprocessing functions are introduced in order to speed up the training but should not affect final results.
- Default values for the preprocessing exponents, $\chi^2 = 1.34$.

	m	n
Σ	3	1.2
g	4	1.2
T_3	3	0.3
V	3	0.3
Δ_S	3	0.

- Stability checks under variation of exponents:

Valence sector		Singlet sector	
	χ^2		χ^2
$n_{T_3} = n_V = 0.1$	1.38	$n_\Sigma = n_g = 0.8$	1.39
$n_{T_3} = n_V = 0.5$	1.34	$n_\Sigma = n_g = 1.6$	1.52
$m_{T_3} = m_V = 2$	1.55	$m_\Sigma = m_g - 1 = 2$	1.37
$m_{T_3} = m_V = 4$	1.28	$m_\Sigma = m_g - 1 = 4$	1.41

Stability estimator: distance between MC ensembles.

* All features of the NNPDF parton set can be assessed by using standard statistical tools.
* Distances between two probability distributions:
  Quark $\left\{ f_{ik}^{(1)} = f_k^{(1)}(x_i, Q_0^2) \right\}$

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

* With:

$$\langle f_i \rangle_{(1)} \equiv \frac{1}{N_{\rm rep}^{(1)}} \sum_{k=1}^{N_{\rm rep}^{(1)}} f_{ik}^{(1)} \ ,$$

$$\sigma^2[f_i^{(1)}] \equiv \frac{1}{N_{\rm rep}^{(1)}(N_{\rm rep}^{(1)}-1)} \sum_{k=1}^{N_{\rm rep}^{(1)}} \left(f_{ik}^{(1)} - \langle f_i \rangle_{(1)} \right)^2$$

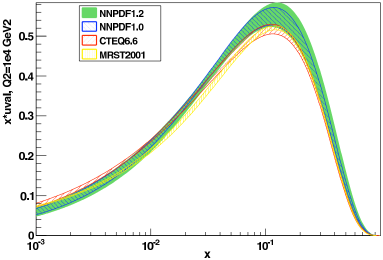
* For statistically equivalent PDF sets: $\langle d[f] \rangle \sim \langle d[\sigma_f] \rangle \sim 1$

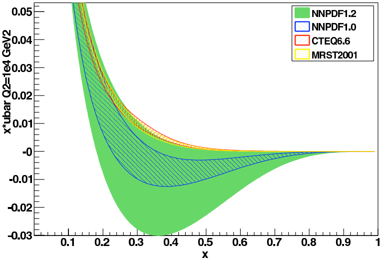
Stability versus preprocessing exponents

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
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

NNPDF1.2: Randomized preprocessing

- Remarkable stability: in most cases variations are within 90% C.L.
- Exception given by valence and triplet: deviation $\sim 1.4\sigma$ from central value when varying exponents.
- Uncertainty on V and T_3 underestimated by factor between 1 and 2.
- Note that we have full control on that!
- NNPDF1.2: Randomized preprocessing!

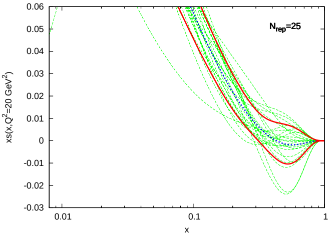


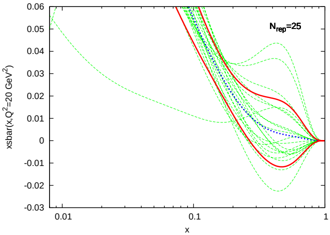


- Bigger uncertainty on $\bar{u}$ and u_v! Will be reduced by DY data.

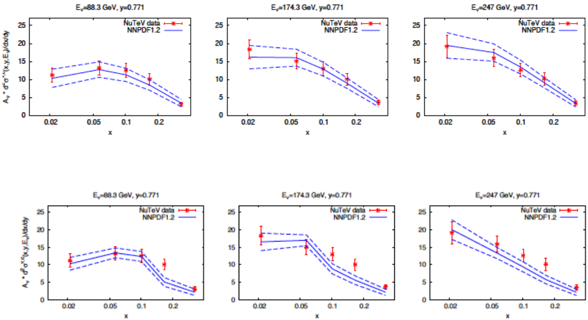
NNPDF1.2: Strangeness determination

- Individual replicas for strange an anti-strange.
- Bigger uncertainty for $\bar{s}$ due to larger uncertainties of anti-neutrino data.





Vcs scan



- Comparison data vs predictions for neutrino dimuon data.
- Top: $V_{cs} = 1.04$, Bottom: $V_{cs} = 0.97$.
- V_{cs} cannot be reabsorbed in normalization, deterioration of χ^2 depends on shape.

NNPDF APPROACH

Monte Carlo sample

Generate a $N_{\rm rep}$ Monte Carlo sets of artificial data, or "pseudo-data" of the original N_{data} data points

$$F_i^{(exp)}(x_p, Q_p^2) \equiv F_{i,p}^{(exp)} \to F_i^{(art)(k)}(x_p, Q_p^2) \equiv F_{i,p}^{(art)(k)} \qquad \begin{array}{rcl} i &=& 1,\ldots,N_{\rm data} \\ k &=& 1,\ldots,N_{\rm rep} \end{array}$$

Multi-gaussian distribution centered on each data point:

$$F_{i,p}^{(art)(k)} = S_{p,N}^{(k)} F_{i,p}^{(\rm exp)} \left(1 + r_p^{(K)} \sigma_p^{\rm stat} + \sum_{j=1}^{N_{\rm sys}} r_{p,j}^{(k)} \sigma_{p,j}^{\rm sys} \right)$$

If two points have correlated systematic uncertainties

$$r_{p,j}^{(k)} = r_{p',j}^{(k)}$$

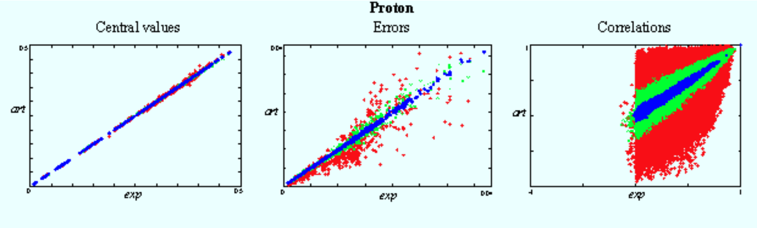
Correlations are properly taken into account.

Validation of the MC sample

Experiment	ZEUS	CHORUS	Total
$\langle PE\left[\langle F^{(\mathrm{art})}\rangle_{\mathrm{rep}}\right]\rangle_{\mathrm{dat}}$	$8.5\cdot 10^{-4}$	$1.8\cdot 10^{-3}$	$7.1\cdot 10^{-5}$
$r\left[F^{(\mathrm{art})}\right]$	1.000	1.000	0.980
$\langle PE\left[\langle \sigma^{(\mathrm{art})}\rangle_{\mathrm{rep}}\right]\rangle_{\mathrm{dat}}$	$9.6\cdot 10^{-3}$	$1.8\cdot 10^{-2}$	$3.0\cdot 10^{-3}$
$\langle \sigma^{(\mathrm{exp})}\rangle_{\mathrm{dat}}$	0.0607	0.1088	0.0556
$\langle \sigma^{(\mathrm{art})}\rangle_{\mathrm{dat}}$	0.0603	0.1109	0.0562
$r\left[\sigma^{(\mathrm{art})}\right]$	1.000	0.998	0.980
$\langle \rho^{(\mathrm{exp})}\rangle_{\mathrm{dat}}$	0.079	0.650	0.145
$\langle \rho^{(\mathrm{art})}\rangle_{\mathrm{dat}}$	0.082	0.657	0.146
$r\left[\rho^{(\mathrm{art})}\right]$	0.982	0.996	0.996
$\langle \mathrm{cov}^{(\mathrm{exp})}\rangle_{\mathrm{dat}}$	$1.53\cdot 10^{-4}$	$2.03\cdot 10^{-2}$	$1.07\cdot 10^{-3}$
$\langle \mathrm{cov}^{(\mathrm{art})}\rangle_{\mathrm{dat}}$	$1.57\cdot 10^{-4}$	$2.11\cdot 10^{-2}$	$1.01\cdot 10^{-3}$
$r\left[\mathrm{cov}^{(\mathrm{art})}\right]$	0.996	0.998	0.997

A MC sample with $\mathcal{O}(1000)$ replicas reproduces mean values, variances, correlations of experimental data within 1% accuracy.

Convergence rate increases with N_{rep}.



From sample to Monte Carlo errors

For each replica $^{(k)}$ of the experimental data we fit a set of independent PDFs

Ensemble of fitted replicas of PDFs: representation of the probability distribution in the space of PDFs

Uncertainties, central values and any other statistical property (e. g. correlations) of the PDFs (or any function of them) can be evaluated using standard statistical methods.

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

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

$$\rho[f_a(x_1,Q_1^2), f_b(x_2,Q_2^2)] = \frac{\langle f_a(x_1,Q_1^2) f_b(x_2,Q_2^2) \rangle - \langle f_a(x_1,Q_1^2) \rangle \langle f_b(x_2,Q_2^2) \rangle}{\sigma_a(x_1,Q_1^2)\sigma_b(x_2,Q_2^2)}$$

How PDFs uncertainties must be evaluated

- Monte Carlo prescription (NNPDF)

$$\sigma_{\mathcal{F}} = \left(\frac{N_{\rm set}}{N_{\rm set}-1} \left(\langle \mathcal{F}[\{f\}]^2 \rangle - \langle \mathcal{F}[\{f\}] \rangle^2 \right) \right)^{1/2}$$

- HEPDATA prescription (CTEQ and MRST/MSTW)

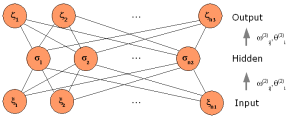
$$\sigma_{\mathcal{F}} = \frac{1}{2C_{90}} \left(\sum_{k=1}^{N_{\rm set}/2} \left(\mathcal{F}[\{f^{(2k-1)}\}] - \mathcal{F}[\{f^{(2k)}\}] \right)^2 \right)^{1/2} , \quad C_{90} = 1.64485$$

C_{90} accounts for the fact that the upper and lower parton sets correspond to 90% confidence levels rather than to one-σ uncertainties.

- HEPDATA* prescription (Alekhin)

$$\sigma_{\mathcal{F}} = \left(\sum_{k=1}^{N_{\rm set}} \left(\mathcal{F}[\{f^{(k)}\}] - \mathcal{F}[\{f^{(0)}\}] \right)^2 \right)^{1/2} .$$

What are neural networks?



* Each neuron receives input from neurons in preceding layer.
* Activation determined by weights and thresholds according to a non linear function:

$$\xi_i = g\Big(\sum_j \omega_{ij}\xi_j - \theta_i\Big), \qquad g(x) = \frac{1}{1+e^{-x}}$$

In a simple case (1-2-1) we have,

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

7 parameters

...Just a convenient functional form which provides a redundant and flexible parametrization.

We want the best fit to be independent of any assumption made on the parametrization.

Fitting Strategy

Our fitting strategy is very different from that of normally used: instead of a set of basis functions with a small number of pars, we have an unbiased basis of functions parameterized by a very large and redundant set of pars.

CTEQ,MSTW,AKP

$\mathcal{O}(20)$ parm

NNPDF

$\mathcal{O}(200)$ parm

Not trivial because ...

1. A redundant parametrization might accommodate also random fluctuations of statistical data.
2. Very large space of parameters

Ingredients of fitting procedure

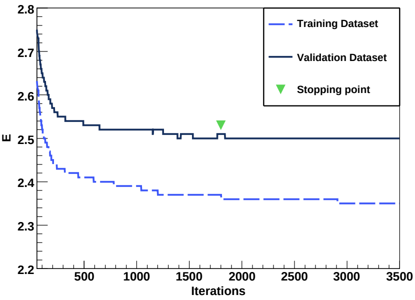
1. Flexible and redundant parametrization
2. Genetic Algortihm minimization
3. Dynamical stopping criterion

Dynamical Stopping Criterion

* GA is monotonically decreasing by construction.
* The best fit is given by an optimal training beyond which the figure of merit improves only because we are fitting statistical noise of the data.

Cross-validation method

* Divide data in two sets: training and validation.
* Random division for each replica ($f_t = f_v = 0.5$).
* Minimisation is performed only on the training set. The validation χ^2 for the set is computed.
* When the training χ^2 still decreases while the validation χ^2 stops decreasing → STOP.



Definition of χ^2

- Fully correlated χ^2:

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

- The covariance matrix $\overline{\rm cov}^{(k)}$ is defined from the experimental covariance matrix which does not include normalization errors.

$$\left(\overline{\rm cov}^{(k)}\right)_{ij} = \left(\overline{\rm cov}^{(exp)}\right)^{-1}_{ij} S_{iN}^{(k)} S_{jN}^{(k)}$$

$$S_{pN}^{(k)} = \prod_{n=1}^{N_a} \left(1 + r_{p,n}^{(k)} \sigma_{p,n}\right) \prod_{i=1}^{N_r} \sqrt{1 + r_{p,i}^{(k)} \sigma_{p,i}}$$

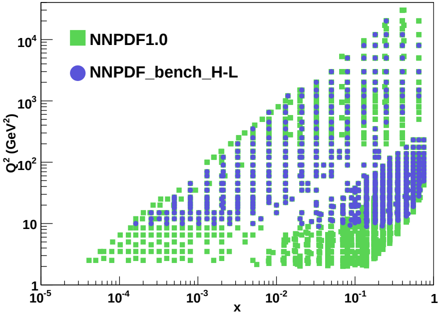
- $F_i^{(net)}$ is computed from PDFs using NLO, ZM-VFN scheme.
- α_s kept fixed.
- N_{rep}= 100-1000 to obtain accurate description of data.

BENCHMARK PARTONS

Dependence on data sets

HERA-LHC benchmark

Benchmark PDF fit to a reduced consistent set of DIS data.(hep-ph/0511119)



Set	$N_{\rm dat}$
BCDMSp	322
NMC	95
NMC-pd	73
Z97NC	206
H197lowQ^2	77

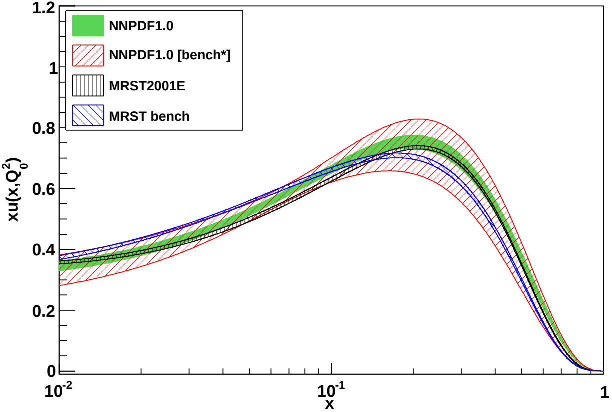
3163 data $\longrightarrow$ **773** data

$$Q^2 > 9\,\mathrm{GeV}^2$$
$$W^2 > 15\,\mathrm{GeV}^2$$

Dependence on data sets

HERA-LHC benchmark

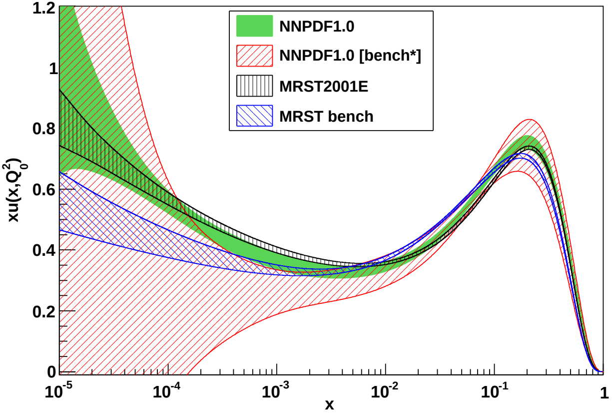
Comparison between collaborations and between benchmark/global partons.
$u(x, Q^2 = 2\mathrm{GeV}^2)$: Data region



Dependence on data sets

HERA-LHC benchmark

Comparison between collaborations and between benchmark/global partons.
$u(x, Q^2 = 2\mathrm{GeV}^2)$: Extrapolation Region



Dependence on data sets

HERA-LHC benchmark

- MRST01: benchmark partons and global partons do not agree within error!
- Input parametrization, flavor assumptions and statistical treatment ($\Delta\chi^2_{\mathrm{global}} = 50$, $\Delta\chi^2_{\mathrm{bench}} = 1$) are tuned to data.
- This is not satisfactory especially to predict the behaviour of PDFs in the extrapolation region (LHC)

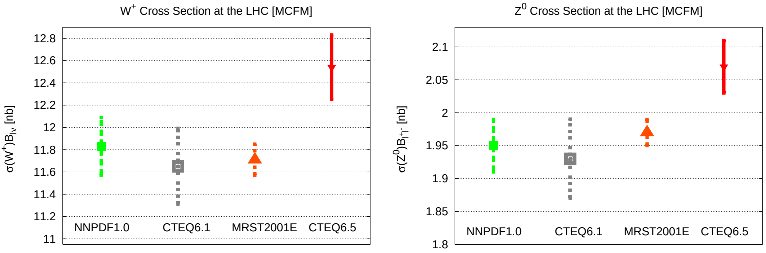
- NNPDF1.0 is consistent with MRST01 global fit.
- NNPDFbench is consistent with NNPDF1.0 and MRST01.
- Same parametrization and flavour assumption.
- Same statistical treatment.
- Underestimation of the error in the standard approach.

NNPDF1.0

Prediction on LHC standard candle processes

- Gauge boson production at the LHC.
- All quantities have been computed at NLO with MCFM (http://mcfm.fnal.gov)
- Quoted uncertainties are the one-σ bands due to the PDF uncertainty only.

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



Statistical estimator: distance between MC ensembles.

* All features of the NNPDF parton set can be assessed by using standard statistical tools.
* Distances between two probability distributions:
  Quark $\left\{ f_{ik}^{(1)} = f_k^{(1)}(x_i, Q_0^2) \right\}$

$$\langle d[f] \rangle = \sqrt{\left\langle \frac{\left(\langle f_i \rangle_{(1)} - \langle f_i \rangle_{(2)} \right)^2}{\sigma^2[f_i^{(1)}] + \sigma^2[f_i^{(2)}]} \right\rangle_{\mathrm{pts}}}$$

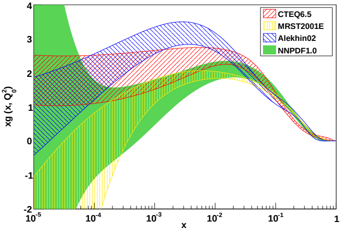
* With:

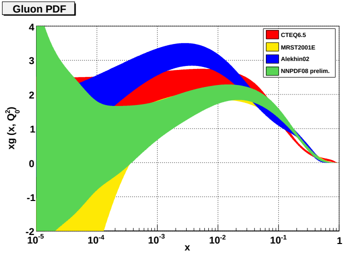
$$\langle f_i \rangle_{(1)} \equiv \frac{1}{N_{\mathrm{rep}}^{(1)}} \sum_{k=1}^{N_{\mathrm{rep}}^{(1)}} f_{ik}^{(1)} \, ,$$

$$\sigma^2[f_i^{(1)}] \equiv \frac{1}{N_{\mathrm{rep}}^{(1)}(N_{\mathrm{rep}}^{(1)} - 1)} \sum_{k=1}^{N_{\mathrm{rep}}^{(1)}} \left(f_{ik}^{(1)} - \langle f_i \rangle_{(1)} \right)^2$$

* For statistically equivalent PDF sets: $\langle d[f] \rangle \sim \langle d[\sigma_f] \rangle \sim 1$

Stability under variation of the parametrization

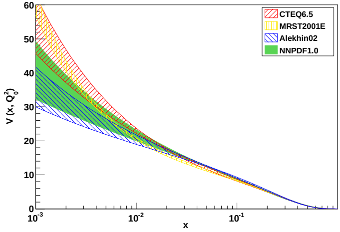


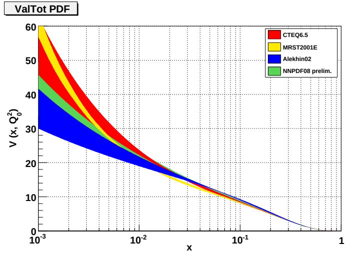


	Data	Extrapolation
$\Sigma(x, Q_0^2)$	$5\ 10^{-4} \leq x \leq 0.1$	$10^{-5} \leq x \leq 10^{-4}$
$\langle d[f] \rangle$	0.98	1.25
$\langle d[\sigma] \rangle$	1.14	1.34
$g(x, Q_0^2)$	$5\ 10^{-4} \leq x \leq 0.1$	$10^{-5} \leq x \leq 10^{-4}$
$\langle d[f] \rangle$	1.52	1.15
$\langle d[\sigma] \rangle$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

* Stability under change of architecture of the nets: **37 pars** $\rightarrow$ **31 pars**
* Independence on the parametrization!

Stability under variation of the parametrization





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

* Stability under change of architecture of the nets: **37 pars** $\longrightarrow$ **31 pars**
* Independence on the parametrization!