First global NNPDF analysis

Maria Ubiali

University of Edinburgh
& Université Catholique de Louvain

University of Manchester
10^{th} March 2010

"A first unbiased global NLO determination of parton distribution functions"
arXiv:1002.4407

The NNPDF Collaboration
R.D.Ball[1], L.Del Debbio[1], S.Forte[2], A.Guffanti[3], J.I.Latorre[4], 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

Outline

1. Introduction
   - Standard Approach
   - NNPDF
2. NNPDF method
   - Monte Carlo Determination of Errors
   - Neural Network as unbiased and redundant parametrization
   - Dynamical Stopping Criterion
3. NNPDF2.0: a global fit
   - New Features
   - Results
   - Impact of modifications
4. Conclusions and outlook

Outline

1. Introduction
   - Standard Approach
   - NNPDF
2. NNPDF method
   - Monte Carlo Determination of Errors
   - Neural Network as unbiased and redundant parametrization
   - Dynamical Stopping Criterion
3. NNPDF2.0: a global fit
   - New Features
   - Results
   - Impact of modifications
4. Conclusions and outlook

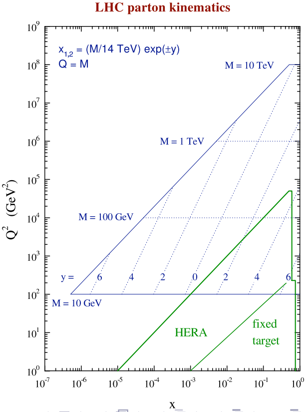
- Factorization Theorem ($Q^2 \gg \Lambda^2_{\mathrm{QCD}}$):

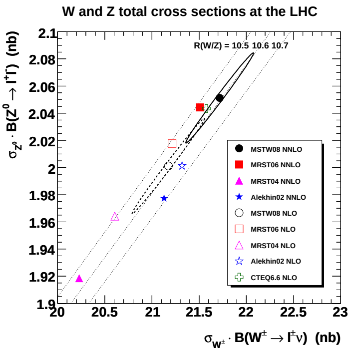
$$\frac{d\sigma_H}{dX} = \sum_{a,b} \int dx_1 dx_2 f_a(x_1,\mu_f) f_b(x_2,\mu_f) \otimes \frac{d\hat{\sigma}_{ab}}{dX}(\alpha_s(\mu_r),\mu_r,\mu_f,x_1,x_2,Q^2)$$

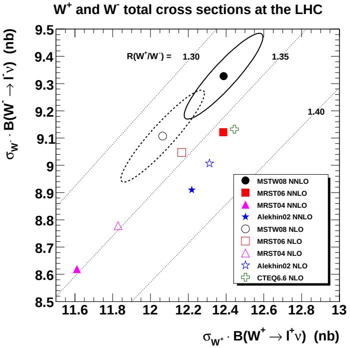
- DGLAP equations:

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

PDFs and their associated uncertainties will play a crucial role in the full exploitation of the LHC physics potential.







James Stirling ArXiv:0812.2341

- For some standard candle processes at LHC the uncertainty on PDFs is the dominant one.
- $\sigma(Z^0)$ at LHC: $\delta\sigma_{\rm PDG} \sim$ 2-3%, $\delta\sigma_{\rm pert} \sim$ 1%
- $\sigma(W^{+,-})$ at LHC: $\delta\sigma_{\rm PDG} \sim$ 3-4%

- Given a set of data points we must determine a set of functions with errors.
- We need an error band in the space of functions, i.e. a probability density $\mathcal{P}[f(x)]$ in the space of PDFs, f(x). For an observable $\mathcal{F}$ depending on PDFs :

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

Standard approach

1. Choose a specific functional form: the ∞-dimensional space of function reduces to a finite-dimensional space of parameters.

$$q_i(x, Q_0^2) = A_i x^{b_i} (1-x)^{c_i} (1 + ...)$$

2. Determine best-fit values of parameters which define the functions.
3. Errors determined via gaussian linear error propagation and large tolerance

$$\Delta\chi^2 \gg 1$$

PDF4LHC meetings

1. The error associated to the choice of parametrization is difficult to assess. How can we know that it is flexible enough and does not introduce a theoretical bias?
2. Non trivial propagation of errors when in presence of non-gaussian errors and incompatible data.
3. Large tolerance $T = \sqrt{\Delta\chi^2}$ means that error on experimental measurements and on PDFs themselves is blown up by a factor $S = \sqrt{\Delta\chi^2/2.7}$.
   $S_{\rm CTEQ} \sim 6$; $S_{\rm MSTW} \sim 4.5$: is that factor mandatory?
4. Both parametrization and statistical treatment $(\Delta\chi^2)$ need to be tuned to exp. data $\rightarrow$ No predictivity in extrapolation region and for future experiments.
5. Benchmark partons do not agree with global partons within errors: incompatible experiments, not enough flexibility in the parametrization or both?

Monte Carlo representation of the probability measure in the space of functions

Use of neural network as redundant and unbiased parametrization

- Structure functions [hep-ph/0501067]
- Non-singlet PDF $q^- = u + d - (\bar{u} + \bar{d})$ [hep-ph/0701127]
- DIS global analysis: NNPDF1.0 [arXiv:0808.1231]
- Determination of the strange content: NNPDF1.2 [arXiv:0906.1958]
- Global (DIS+DY+JET) analysis: NNPDF2.0 [arXiv:1002.4407]

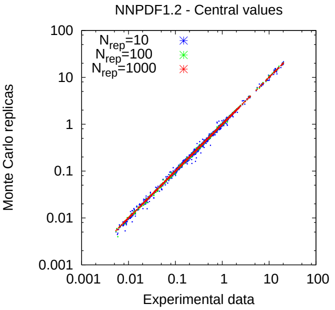
All sets are available in the LHAPDF interface

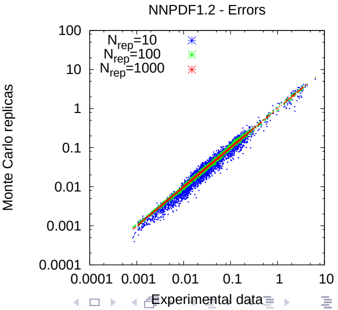
- **Importance sampling:** find PDF ensemble $\{f_k\}$ such that

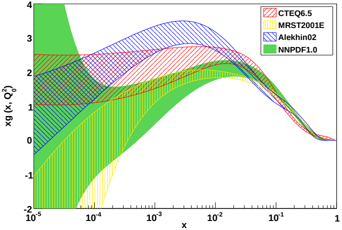
$$\langle \mathcal{O}[f] \rangle \sim \frac{1}{N} \sum_{n=1}^{N} \mathcal{O}[f_n]$$

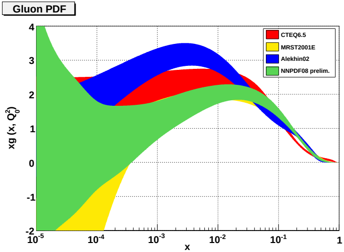
- $\mathcal{O}[f]$ can be xsec, variance, correlation...
- Generate ensemble in the space of data: N pseudo–data reproduce the probability distribution of the original experimental data; uncertainty scales as the size of the sample.

Monte Carlo behaves in a statistically consistent way





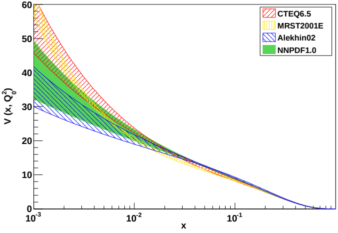


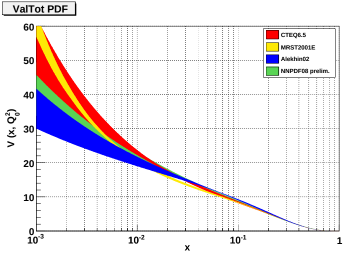


Results are statistically independent of the choice of the parametrization

	Data	Extrapolation
$\Sigma(x, Q_0^2)$	$5\ 10^{-4} \le x \le 0.1$	$10^{-5} \le x \le 10^{-4}$
$\langle d[f] \rangle$	0.62	0.88
$\langle d[\sigma] \rangle$	0.87	0.95
$g(x, Q_0^2)$	$5\ 10^{-4} \le x \le 0.1$	$10^{-5} \le x \le 10^{-4}$
$\langle d[f] \rangle$	1.07	0.87
$\langle d[\sigma] \rangle$	0.86	0.78
$T_3(x, Q_0^2)$	$0.05 \le x \le 0.75$	$10^{-3} \le x \le 10^{-2}$
$\langle d[f] \rangle$	1.00	1.11
$\langle d[\sigma] \rangle$	1.24	1.61
$V(x, Q_0^2)$	$0.1 \le x \le 0.6$	$3\ 10^{-3} \le x \le 3\ 10^{-2}$
$\langle d[f] \rangle$	1.30	0.90
$\langle d[\sigma] \rangle$	0.90	0.78
$\Delta_S(x, Q_0^2)$	$0.1 \le x \le 0.6$	$3\ 10^{-3} \le x \le 3\ 10^{-2}$
$\langle d[f] \rangle$	0.84	1.02
$\langle d[\sigma] \rangle$	1.02	1.12

37 pars $\rightarrow$ 31 pars





Results are statistically independent of the choice of the parametrization

	Data	Extrapolation
$\Sigma(x, Q_0^2)$	$5\ 10^{-4} \leq x \leq 0.1$	$10^{-5} \leq x \leq 10^{-4}$
$\langle d[f] \rangle$	0.62	0.88
$\langle d[\sigma] \rangle$	0.87	0.95
$g(x, Q_0^2)$	$5\ 10^{-4} \leq x \leq 0.1$	$10^{-5} \leq x \leq 10^{-4}$
$\langle d[f] \rangle$	1.07	0.87
$\langle d[\sigma] \rangle$	0.86	0.78
$T_3(x, Q_0^2)$	$0.05 \leq x \leq 0.75$	$10^{-3} \leq x \leq 10^{-2}$
$\langle d[f] \rangle$	1.00	1.11
$\langle d[\sigma] \rangle$	1.24	1.61
$V(x, Q_0^2)$	$0.1 \leq x \leq 0.6$	$3\ 10^{-3} \leq x \leq 3\ 10^{-2}$
$\langle d[f] \rangle$	1.30	0.90
$\langle d[\sigma] \rangle$	0.90	0.78
$\Delta_S(x, Q_0^2)$	$0.1 \leq x \leq 0.6$	$3\ 10^{-3} \leq x \leq 3\ 10^{-2}$
$\langle d[f] \rangle$	0.84	1.02
$\langle d[\sigma] \rangle$	1.02	1.12

37 pars $\rightarrow$ 31 pars

PDFs are even stable upon the addition of independent PDFs parametrizations

- 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

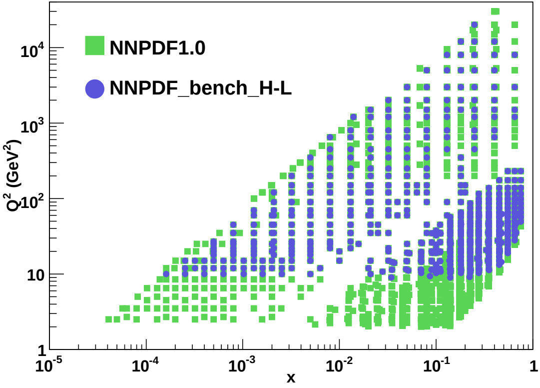
Strangeness valence : $s^{-}(x) \equiv (s(x) - \bar{s}(x))/2 \longmapsto \mathrm{NN}_{(s-)}(x)$ 2-5-3-1 37 pars

- Added two unconstrained PDFs.

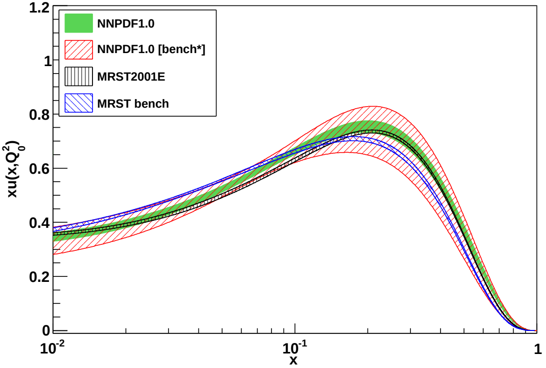
185 → 259 parameters

- Only strange (and Σ) affected. Gluon and statistical features remain unchanged.

PDFs behave as expected upon the addition of new data
HERA-LHC benchmark: DIS data

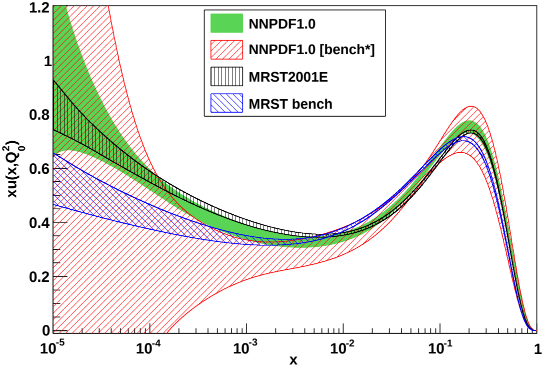


PDFs behave as expected upon the addition of new data
HERA-LHC benchmark: DIS data



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

PDFs behave as expected upon the addition of new data
HERA-LHC benchmark: DIS data



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

Control on PDFs uncertainties: NuteV anomaly solved AND precision studies at the same time

- 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

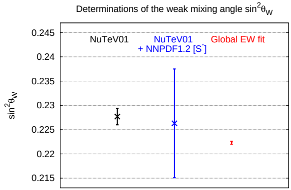
$\sin^2\theta_W = 0.2223 \pm 0.0002$

NuTeV

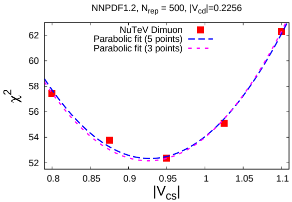
$\sin^2\theta_W = 0.2276 \pm 0.0014$

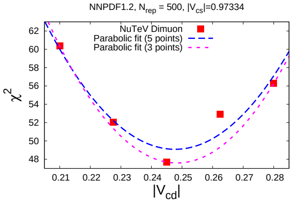
$$\delta_s \sin^2\theta_W \sim -0.240 \frac{[S^-]}{[Q^-]}$$

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



Control on PDFs uncertainties: NuteV anomaly solved AND precision studies at the same time





CKM fit

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

Direct Determination

$$\begin{array}{lcll} V_{cs} & = & 1.04 \pm 0.06, \ \Delta V_{cs} \sim 6\% & \text{D/B decays} \\ V_{cs} & > & 0.59 & \text{DIS fit} \end{array}$$

NNPDF1.2 analysis

$$V_{cs} = 0.97 \pm 0.07, \ \Delta V_{cs} \sim 6\%$$

Outline

1 Introduction
- Standard Approach
- NNPDF

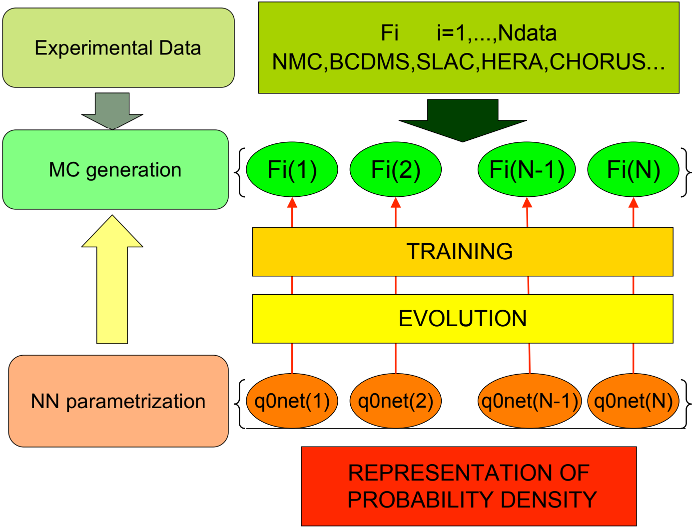
2 NNPDF method
- Monte Carlo Determination of Errors
- Neural Network as unbiased and redundant parametrization
- Dynamical Stopping Criterion

3 NNPDF2.0: a global fit
- New Features
- Results
- Impact of modifications

4 Conclusions and outlook

NNPDF approach

General scheme



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

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

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

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

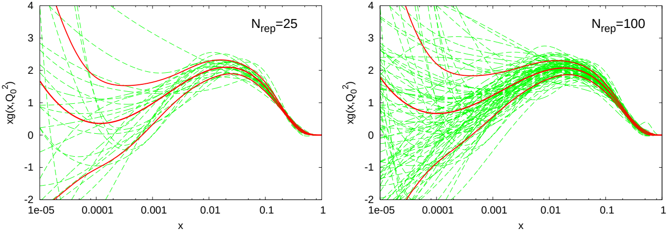
Correlations are properly taken into account.

NNPDF approach

Ingredient ♯1: Monte Carlo Errors

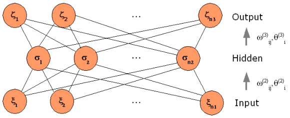
$$\langle \mathcal{F}[f(x)] \rangle = \frac{1}{N_{\rm rep}} \sum_{k=1}^{N_{\rm rep}} \mathcal{F}[f^{(k)(\rm net)}(x)]$$

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



Even though individual replicas may fluctuate significantly, average quantities such as central values and error bands are smooth inasmuch as stability is reached due to the dimension of the ensemble increasing.

Each independent PDF at the initial scale $Q_0^2 = 2\mathrm{GeV}^2$ is parameterized by an individual NN.



* 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.

Basis set: $Q_0^2 = 2\ \mathrm{GeV}^2$

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
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

259 parameters

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

CTEQ,MSTW,AL

$\mathcal{O}(20)$ parm

NNPDF

$\mathcal{O}(200)$ parm

Not trivial because ...

A redundant parametrization might adapt not only to physical behavior but also to random statistical fluctuations of data.

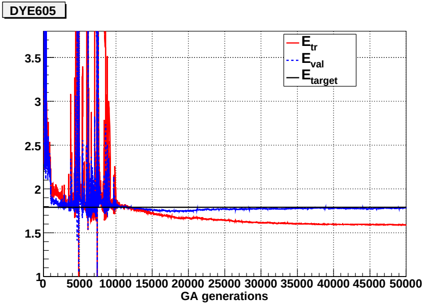
Ingredients of fitting procedure

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

* GA is monotonically decreasing by construction.
* The best fit is not given by the absolute minimum.
* 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.



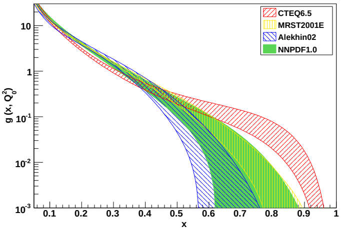
Outline

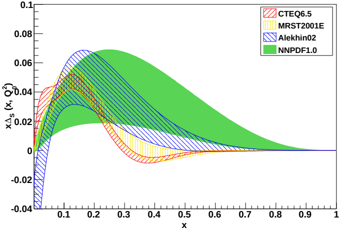
1. Introduction
   - Standard Approach
   - NNPDF
2. NNPDF method
   - Monte Carlo Determination of Errors
   - Neural Network as unbiased and redundant parametrization
   - Dynamical Stopping Criterion
3. NNPDF2.0: a global fit
   - New Features
   - Results
   - Impact of modifications
4. Conclusions and outlook

- DIS data are insufficient to determine accurately PDFs.
- Flavor decomposition of quark–antiquark sea and large–x gluon distribution.

$$R^{pd} = \frac{d\sigma^d/dM^2 dx_F}{d\sigma^p/dM^2 dx_F} \quad \propto \quad (1 + \bar{d}/\bar{u})$$

$$A^W = \frac{d\sigma_W^+/dy - d\sigma_W^-/dy}{d\sigma_W^+/dy - d\sigma_W^-/dy} \quad \propto \quad \frac{u\bar{d} - d\bar{u}}{u\bar{d} + d\bar{u}}$$

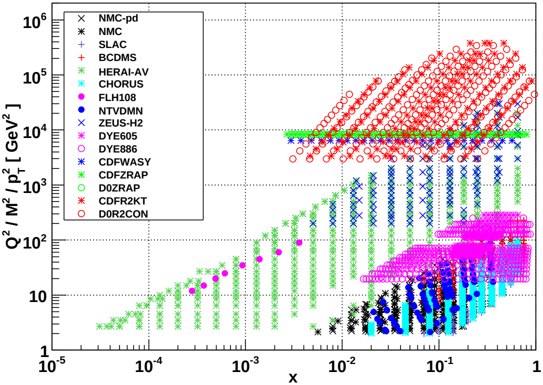




- NNPDF2.0 includes most of the available hadronic data.

1. Included fixed target Drell-Yan, Tevatron jets and weak bosons data.
2. Improvements in training/stopping
   - Target Weighted Training
   - Improved stopping
3. Improved treatment of normalization errors (t_0 method)
   - For details see [arXiv:0912.2276]
4. Fast DGLAP evolution based on higher-order interpolating polynomials
5. Fast computation of Drell-Yan observables based on higher-order interpolating polynomials. Full NLO computation.
6. Enforced positivity constraints

NNPDF2.0 dataset



3477 data points

For comparison MSTW08 includes 2699 data points

OBS	Data sets
F_2^p	NMC,SLAC,BDCMS
F_2^d	SLAC,BCDMS
F_2^d / F_2^p	NMC-pd
σ_{NC}	HERA-I AV, ZEUS-H2
σ_{CC}	HERA-I AV, ZEUS-H2
F_L	H1
σ_ν , $\sigma_{\bar{\nu}}$	CHORUS
dimuon prod.	NuTeV
$d\sigma^{\rm DY}/dM^2dy$	E605
$d\sigma^{\rm DY}/dM^2dx_F$	E886
W asymmetry	CDF
Z rap. distr.	CDF,D0
incl. $\sigma^{(\rm jet)}$	D0(cone) Run II
incl. $\sigma^{(\rm jet)}$	CDF(k_T) Run II

- Kinematical cuts on DIS data
  Q^2 >2 GeV^2
  $W^2 = Q^2(1-x)/x$ >12.5 GeV^2
- No cuts on hadronic data

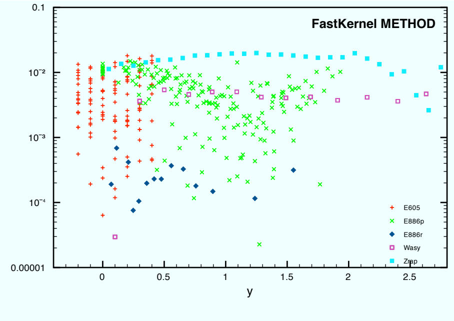
- The NLO computation of hadronic observables too slow for parton global fits.
- Often higher order corrections are included as (local) K factors rescaling the LO cross section
- K-factor depends on PDFs and it is not always a good approximation.

* 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 **FastKernel** computation of DY observables.

- Both PDFs evolution and double convolution are sped up by:
  - Use of high-orders polynomial interpolation
  - Pre–computing all Green Functions

$$\int_{x_{0,1}}^{1} dx_1 \int_{x_{0,2}}^{1} dx_2\, f_a(x_1) f_b(x_2) C^{ab}(x_1, x_2) \rightarrow \sum_{\alpha,\beta=1}^{N_x} f_a(x_{1,\alpha}) f_b(x_{2,\beta}) \int_{x_{0,1}}^{1} dx_1 \int_{x_{0,2}}^{1} dx_2\, \mathcal{I}^{(\alpha,\beta)}(x_1, x_2) C^{ab}(x_1, x_2)$$

- New strategy to solve DGLAP evolution equation
- Implementation benchmarked against the Les Houches tables
- Gain in speed by a factor 30 (for a fit to 3000 datapoints)



x (50 pts)	$e_{\rm rel}(u_V)$	$e_{\rm rel}(\Sigma)$	$e_{\rm rel}(g)$
$1\cdot10^{-7}$	$2.1\cdot10^{-4}$	$2.7\cdot10^{-5}$	$4.7\cdot10^{-6}$
$1\cdot10^{-6}$	$8.9\cdot10^{-5}$	$3.0\cdot10^{-5}$	$2.1\cdot10^{-5}$
$1\cdot10^{-5}$	$9.3\cdot10^{-5}$	$2.3\cdot10^{-5}$	$2.0\cdot10^{-5}$
$1\cdot10^{-4}$	$4.5\cdot10^{-5}$	$4.4\cdot10^{-5}$	$4.2\cdot10^{-5}$
$1\cdot10^{-3}$	$3.0\cdot10^{-5}$	$4.0\cdot10^{-5}$	$3.5\cdot10^{-5}$
$1\cdot10^{-2}$	$7.9\cdot10^{-5}$	$4.5\cdot10^{-5}$	$5.8\cdot10^{-5}$
$1\cdot10^{-1}$	$1.7\cdot10^{-4}$	$1.6\cdot10^{-5}$	$3.9\cdot10^{-5}$
$3\cdot10^{-1}$	$9.1\cdot10^{-6}$	$1.1\cdot10^{-5}$	$1.9\cdot10^{-7}$
$5\cdot10^{-1}$	$2.4\cdot10^{-5}$	$2.2\cdot10^{-5}$	$2.2\cdot10^{-5}$
$7\cdot10^{-1}$	$9.1\cdot10^{-5}$	$7.8\cdot10^{-5}$	$1.2\cdot10^{-4}$
$9\cdot10^{-1}$	$1.0\cdot10^{-3}$	$8.0\cdot10^{-4}$	$2.8\cdot10^{-3}$

- Drell–Yan fast computation exploits linear interpolation
- Accuracy below 1% for all points included in the fit
- Increasing number of points in the grid one can improve accuracy;

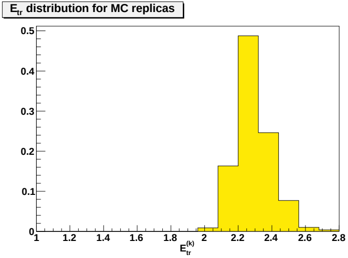
A truly NLO analysis

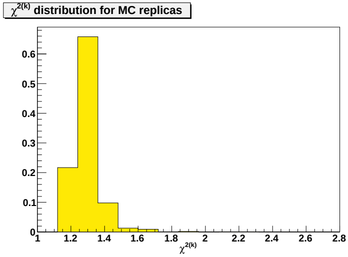
- We want the fitting procedure to explore only the subspace of acceptable physical solutions.
- We want
  - F_L positive.
  - Dimuon cross–section positive.
  - Momentum and valence sum rules.
- Modify the training function witha ddition of a Lagrangian multiplier:

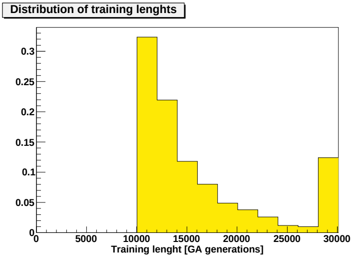
$$E^{(k)} \longrightarrow E^{(k)} - \lambda_{\rm pos} \sum_{I=1}^{N_{\rm dat,pos}} \Theta\left(F_I^{(net)(k)}\right) F_I^{(net)(k)}$$

- $N_{\rm dat,pos}$: number of pseudodata points used to implement positivity constraints.
- $\lambda_{\rm pos}$: associated Lagrangian multiplier (10^{10})

Statistical features: global χ^2



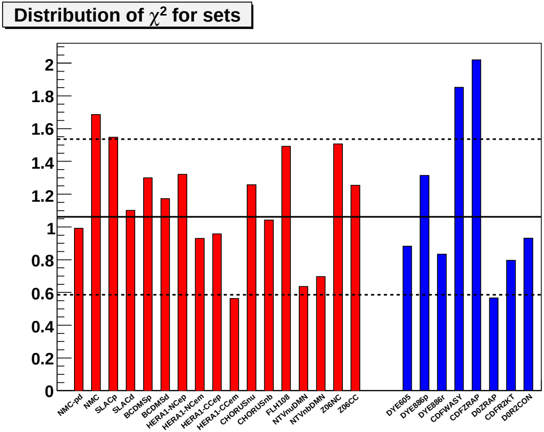




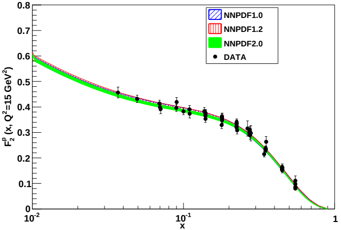
$\chi^2_{\rm tot}$	1.21
$\langle E \rangle \pm \sigma_E$	2.32 $\pm$ 0.10
$\langle E_{\rm val} \rangle \pm \sigma_{E_{\rm val}}$	2.29$\pm$ 0.11
$\langle E_{\rm val} \rangle \pm \sigma_{E_{\rm val}}$	2.35$\pm$ 0.12
$\langle TL \rangle \pm \sigma_{TL}$	16175 $\pm$ 6275
$\langle \chi^{2(k)} \rangle \pm \sigma_{\chi^2}$	1.29$\pm$ 0.09
$\langle \sigma^{(\rm exp)} \rangle_{\rm dat} (\%)$	11.4
$\langle \sigma^{(\rm net)} \rangle_{\rm dat} (\%)$	6.0

Results

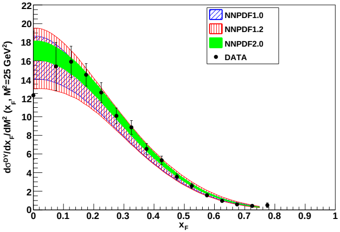
Statistical features: individual experiments



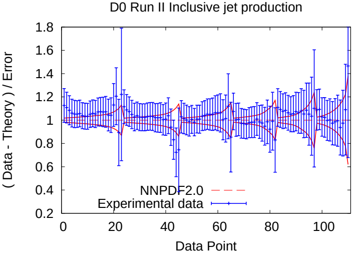
- NuTeV does not provide covariance matrix
- HERAI: larger χ^2 for more precise NC data. Same pattern as HERA analysis.
- Poor compatibility of W-asymmetry data.
- Z rapidity distribution: same χ^2 if fitted off–line. No much info added.



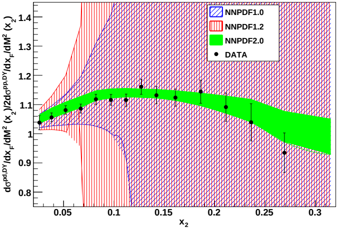
NNPDF1.0
NNPDF1.2
NNPDF2.0
DATA
$F_2^p (x, Q^2=15\ GeV^2)$
x
10^{-2}
10^{-1}
1



NNPDF1.0
NNPDF1.2
NNPDF2.0
DATA
$d\sigma^{DY}/dx_F/dM^2\ (x_F,\ M^2=25\ GeV^2)$
x_F



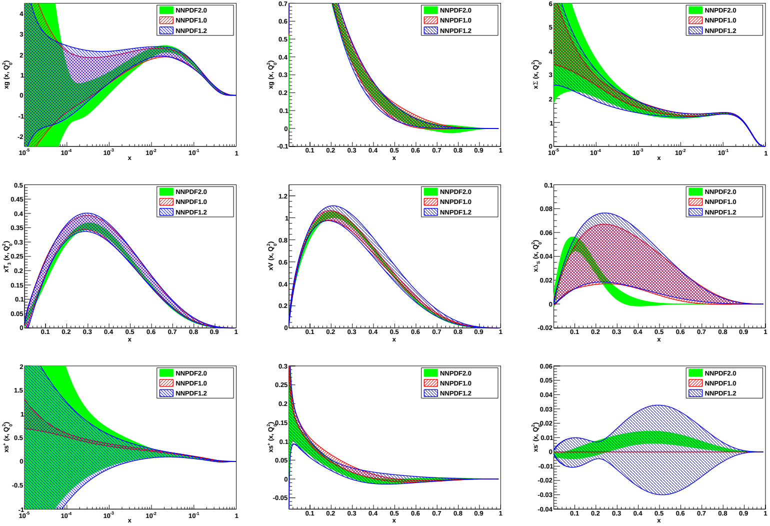
D0 Run II Inclusive jet production
(Data - Theory) / Error
Data Point
NNPDF2.0
Experimental data



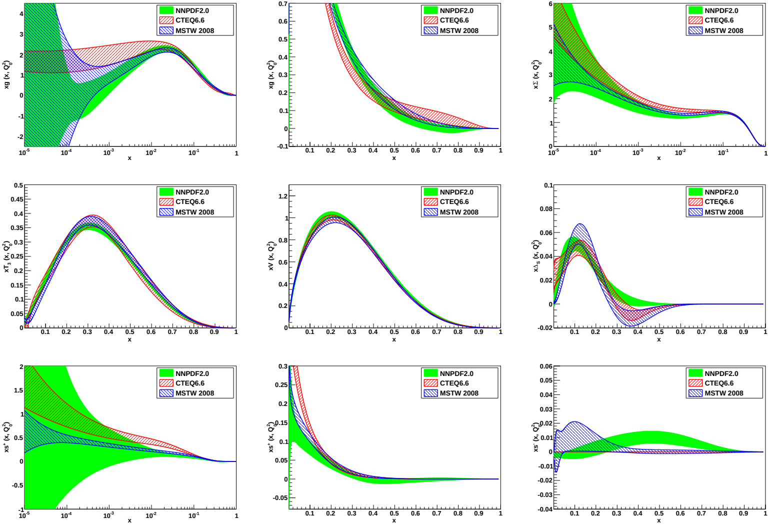
NNPDF1.0
NNPDF1.2
NNPDF2.0
DATA
$d\sigma^{pd,DY}/dx_F/dM^2\ (x_2)/2d\sigma^{pp,DY}/dx_F/dM^2\ (x_2)$
x_2

Results

Partons

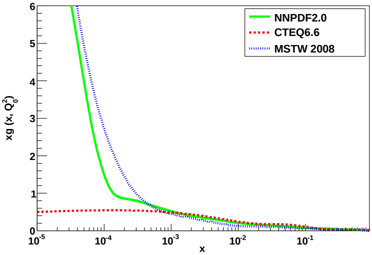


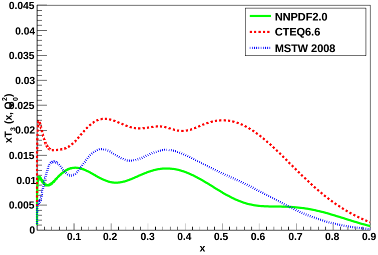
Results
Partons

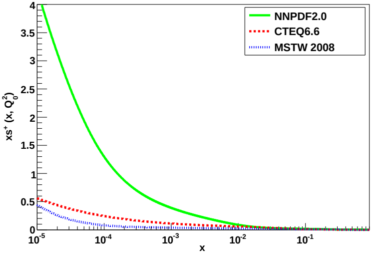


Results

Partons





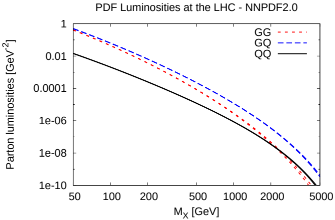


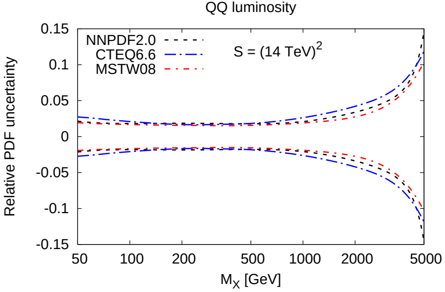
- Reduction of uncertainties with respect to older NNPDF sets (consistency).
- Uncertainties competitive with results from other groups.
- Smaller for some PDFs, larger when parametrization is too rigid.

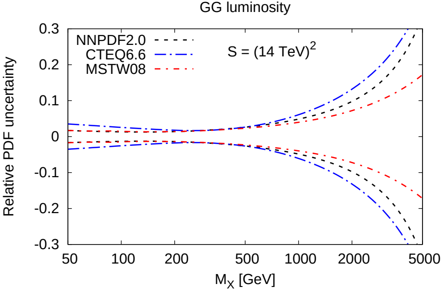
$$\Phi_{gg}\left(M_X^2\right) = \frac{1}{s}\int_\tau^1 \frac{dx_1}{x_1} g\left(x_1, M_X^2\right) g\left(\tau/x_1, M_X^2\right)$$

$$\Phi_{gq}\left(M_X^2\right) = \frac{1}{s}\int_\tau^1 \frac{dx_1}{x_1} \left[g\left(x_1, M_X^2\right) \Sigma\left(\tau/x_1, M_X^2\right) + (1 \to 2)\right]$$

$$\Phi_{qq}\left(M_X^2\right) = \frac{1}{s}\int_\tau^1 \frac{dx_1}{x_1} \sum_{i=1}^{N_f} \left[q_i\left(x_1, M_X^2\right) \bar{q}_i\left(\tau/x_1, M_X^2\right) + (1 \to 2)\right]$$







- A quantitative assessment is possible

$$d(q_j) = \sqrt{\left\langle \frac{\left(\langle q_j\rangle_{(1)} - \langle q_j\rangle_{(2)}\right)^2}{\sigma_1^2[q_j] + \sigma_2^2[q_j]} \right\rangle_{N_{\rm part}}}$$

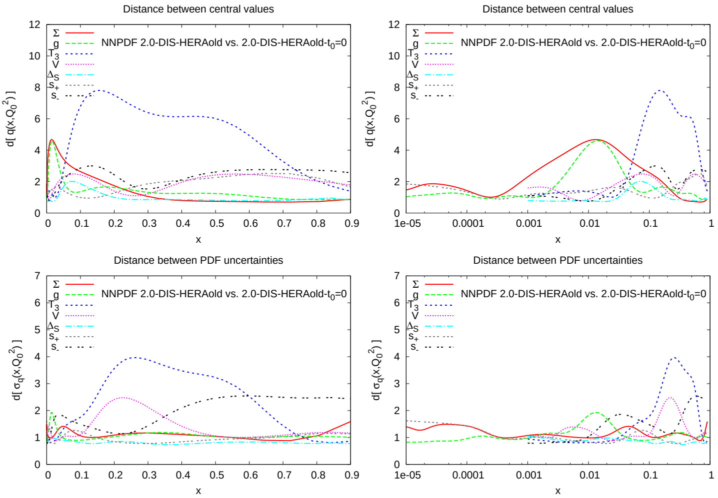
$$d(\sigma_j) = \sqrt{\left\langle \frac{\left(\langle \sigma_j\rangle_{(1)} - \langle \sigma_j\rangle_{(2)}\right)^2}{\sigma_1^2[\sigma_j] + \sigma_2^2[\sigma_j]} \right\rangle_{N_{\rm part}}}$$

- Comparisons performed in NNPDF2.0 analysis

  1. Start from NNPDF1.2
  2. NNPDF1.2 vs. NNPDF1.2 + minimization/training improvements
  3. Improved NNPDF1.2 vs. Improved NNPDF1.2 + t_0-method
  4. Fit to DIS dataset with H1/ZEUS data vs. Fit with HERA-I combined
  5. Fit to DIS dataset vs. Fit to DIS+JET
  6. Fit to DIS+JET vs. NNPDF2.0 final

Impact of modifications

HERA-I combined dataset

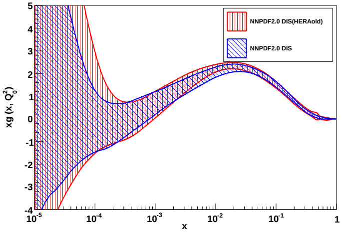


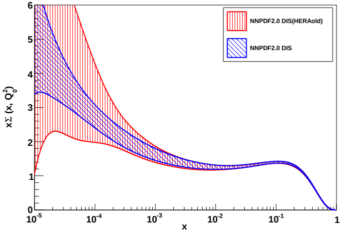
Fit	NNPDF1.2	NNPDF1.2+IGA	NNPDF1.2+IGA+t_0	2.0 DIS	2.0 DIS+JET	NNPDF2.0
$\chi^2_{\rm tot}$	1.32	1.16	1.12	1.20	1.18	1.21
$\langle E\rangle$	2.79	2.41	2.24	2.31	2.28	2.32
$\langle E_{\rm tr}\rangle$	2.75	2.39	2.20	2.28	2.24	2.29
$\langle E_{\rm val}\rangle$	2.80	2.46	2.27	2.34	2.32	2.35
$\langle\chi^{2(k)}\rangle$	1.60	1.28	1.21	1.29	1.27	1.29
NMC-pd	1.48	0.97	0.87	0.85	0.86	0.99
NMC	1.68	1.72	1.65	1.69	1.66	1.69
SLAC	1.20	1.42	1.33	1.37	1.31	1.34
BCDMS	1.59	1.33	1.25	1.26	1.27	1.27
HERAI	1.05	0.98	0.96	1.13	1.13	1.14
CHORUS	1.39	1.13	1.12	1.13	1.11	1.18
FLH108	1.70	1.53	1.53	1.51	1.49	1.49
NTVDMN	0.64	0.81	0.71	0.71	0.75	0.67
ZEUS-H2	1.52	1.51	1.49	1.50	1.49	1.51
DYE605	*11.19*	*22.89*	*8.21*	*7.32*	*10.35*	0.88
DYE866	*53.20*	*4.81*	*2.46*	*2.24*	*2.59*	1.28
CDFWASY	*26.76*	*28.22*	*20.32*	*13.06*	*14.13*	1.85
CDFZRAP	*1.65*	*4.61*	*3.13*	*3.12*	*3.31*	2.02
D0ZRAP	*0.56*	*0.80*	*0.65*	*0.65*	*0.68*	0.47
CDFR2KT	*1.10*	*0.95*	*0.78*	*0.91*	0.79	0.80
D0R2CON	*1.18*	*1.07*	*0.94*	*1.00*	0.93	0.93

Impact of modifications

HERA-I combined dataset

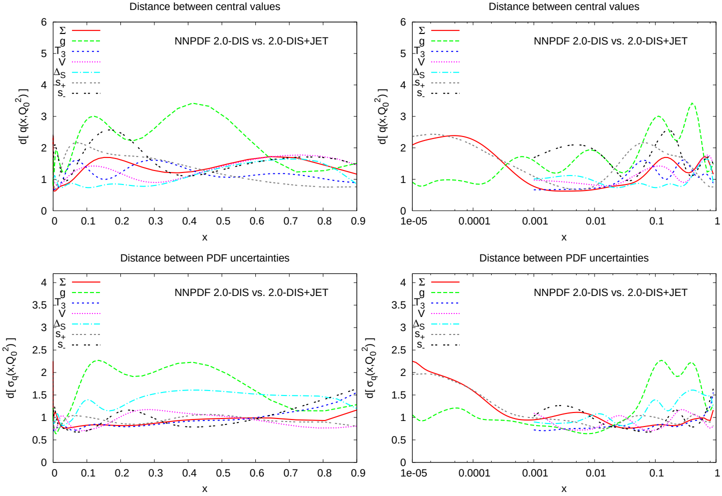
- HERA-I combined more precise:cross-calibration $\chi^2 = 1.12$ vs $\chi^2 = 1.21$
- Quality of other data unchanged
- Overall fit quality to the whole dataset is good ($\chi^2 = 1.13$)
  - $\sigma^+_{\rm NC}$ dataset has relatively high $\chi^2 \sim 1.3$
  - $\sigma^-_{\rm CC}$ dataset has very low $\chi^2 \sim 0.55$
- Same χ^2-pattern observed in the HERAPDF1.0 analysis
- Impact on PDFs, mainly Singlet and Gluon at small-x





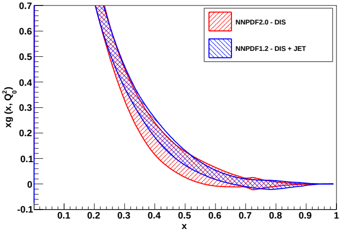
Impact of modifications

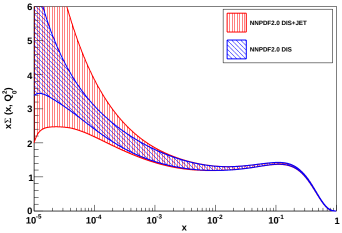
Tevatron inclusive Jet data



Fit	NNPDF1.2	NNPDF1.2+IGA	NNPDF1.2+IGA+t_0	2.0 DIS	2.0 DIS+JET	NNPDF2.0
$\chi^2_{\rm tot}$	1.32	1.16	1.12	1.20	1.18	1.21
$\langle E \rangle$	2.79	2.41	2.24	2.31	2.28	2.32
$\langle E_{\rm tr} \rangle$	2.75	2.39	2.20	2.28	2.24	2.29
$\langle E_{\rm val} \rangle$	2.80	2.46	2.27	2.34	2.32	2.35
$\langle \chi^{2(k)} \rangle$	1.60	1.28	1.21	1.29	1.27	1.29
NMC-pd	1.48	0.97	0.87	0.85	0.86	0.99
NMC	1.68	1.72	1.65	1.69	1.66	1.69
SLAC	1.20	1.42	1.33	1.37	1.31	1.34
BCDMS	1.59	1.33	1.25	1.26	1.27	1.27
HERAI	1.05	0.98	0.96	1.13	1.13	1.14
CHORUS	1.39	1.13	1.12	1.13	1.11	1.18
FLH108	1.70	1.53	1.53	1.51	1.49	1.49
NTVDMN	0.64	0.81	0.71	0.71	0.75	0.67
ZEUS-H2	1.52	1.51	1.49	1.50	1.49	1.51
DYE605	*11.19*	*22.89*	*8.21*	*7.32*	*10.35*	0.88
DYE866	*53.20*	*4.81*	*2.46*	*2.24*	*2.59*	1.28
CDFWASY	*26.76*	*28.22*	*20.32*	*13.06*	*14.13*	1.85
CDFZRAP	*1.65*	*4.61*	*3.13*	*3.12*	*3.31*	2.02
D0ZRAP	*0.56*	*0.80*	*0.65*	*0.65*	*0.68*	0.47
CDFR2KT	*1.10*	*0.95*	*0.78*	*0.91*	0.79	0.80
D0R2CON	*1.18*	*1.07*	*0.94*	*1.00*	0.93	0.93

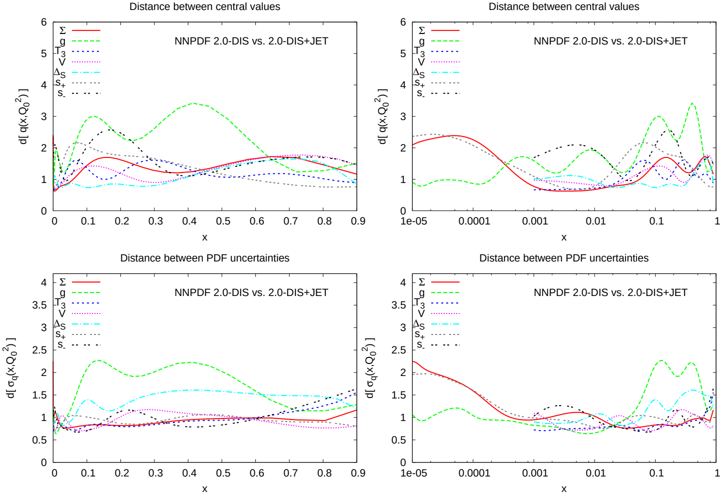
- We include Tevatron Run-II inclusive jet data
- They provide a valuable constrain on large-x gluon
- Run-I data not included but compatibility with the outcome of the fit has been checked. χ^2 close to CTEQ66.





Impact of modifications

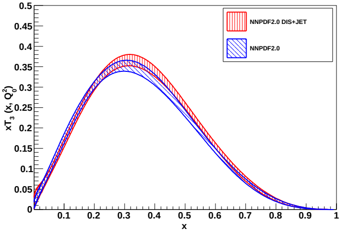
Drell-Yan and Vector Boson production data

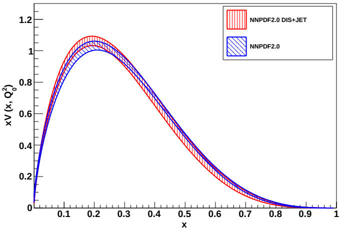


Drell-Yan and Vector Boson production data

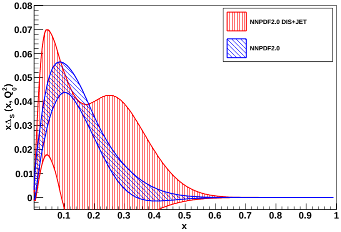
Fit	NNPDF1.2	NNPDF1.2+IGA	NNPDF1.2+IGA+t_0	2.0 DIS	2.0 DIS+JET	NNPDF2.0
$\chi^2_{\rm tot}$	1.32	1.16	1.12	1.20	1.18	1.21
$\langle E \rangle$	2.79	2.41	2.24	2.31	2.28	2.32
$\langle E_{\rm tr} \rangle$	2.75	2.39	2.20	2.28	2.24	2.29
$\langle E_{\rm val} \rangle$	2.80	2.46	2.27	2.34	2.32	2.35
$\langle \chi^{2(k)} \rangle$	1.60	1.28	1.21	1.29	1.27	1.29
NMC-pd	1.48	0.97	0.87	0.85	0.86	0.99
NMC	1.68	1.72	1.65	1.69	1.66	1.69
SLAC	1.20	1.42	1.33	1.37	1.31	1.34
BCDMS	1.59	1.33	1.25	1.26	1.27	1.27
HERAI	1.05	0.98	0.96	1.13	1.13	1.14
CHORUS	1.39	1.13	1.12	1.13	1.11	1.18
FLH108	1.70	1.53	1.53	1.51	1.49	1.49
NTVDMN	0.64	0.81	0.71	0.71	0.75	0.67
ZEUS-H2	1.52	1.51	1.49	1.50	1.49	1.51
DYE605	*11.19*	*22.89*	*8.21*	*7.32*	*10.35*	0.88
DYE866	*53.20*	*4.81*	*2.46*	*2.24*	*2.59*	1.28
CDFWASY	*26.76*	*28.22*	*20.32*	*13.06*	*14.13*	1.85
CDFZRAP	*1.65*	*4.61*	*3.13*	*3.12*	*3.31*	2.02
D0ZRAP	*0.56*	*0.80*	*0.65*	*0.65*	*0.68*	0.47
CDFR2KT	*1.10*	*0.95*	*0.78*	*0.91*	0.79	0.80
D0R2CON	*1.18*	*1.07*	*0.94*	*1.00*	0.93	0.93

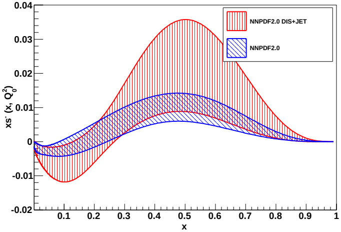
- Good description of fixed target Drell-Yan data (E605 proton and E886 proton and p/d ratio)
- Vector boson production at colliders (CDF W-asymmetry and Z rapidity distribution) harder to fit
- All valence-type PDF combinations are affected by these data
- Sizable reduction in the uncertainty of the strange valence (possible impact on NuTeV anomaly)





- Good description of fixed target Drell-Yan data (E605 proton and E886 proton and p/d ratio)
- Vector boson production at colliders (CDF W-asymmetry and Z rapidity distribution) harder to fit
- All valence-type PDF combinations are affected by these data
- Sizable reduction in the uncertainty of the strange valence (possible impact on NuTeV anomaly)

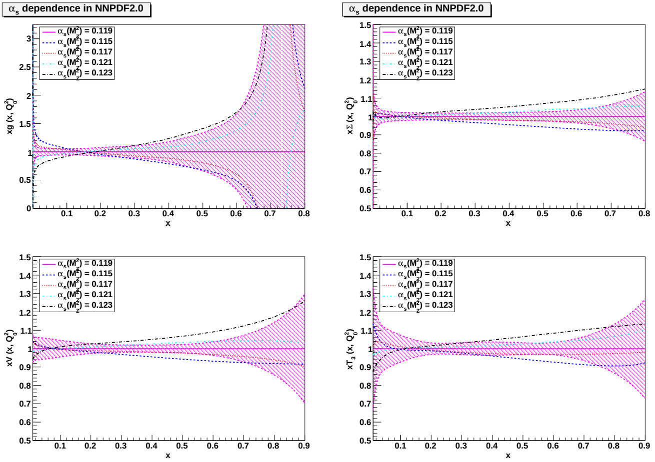




* Precise HERA data
  $\rightarrow$ Small x gluon & singlet
* Tevatron W asymmetry data
  $\rightarrow$ Small x flavor separation
* Fixed Target DIS data, Drell-Yan, neutrino inclusive
  $\rightarrow$ Small x flavor separation
* Neutrino dimuon
  $\rightarrow$ Strangeness
* Tevatron jets
  $\rightarrow$ Large x gluon

No signs of tension between datasets included in the analysis!!!

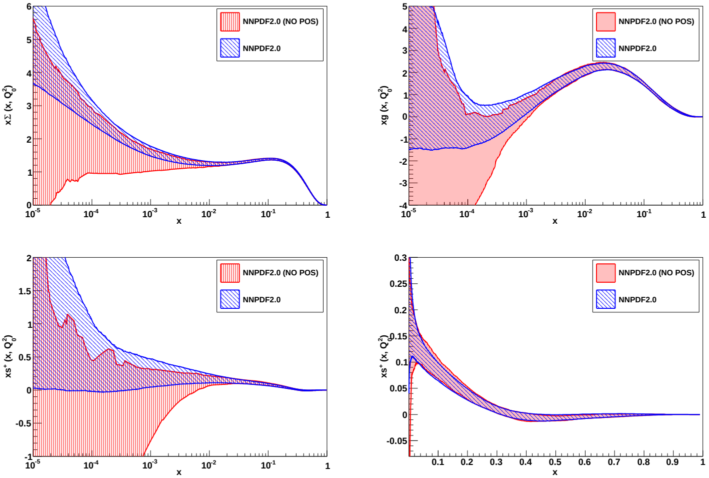
Results
Impact of α_s



- Greater sensitivity to α_s than NNPDF1.2
- Greater NLO corrections for Drell-Yan observables.

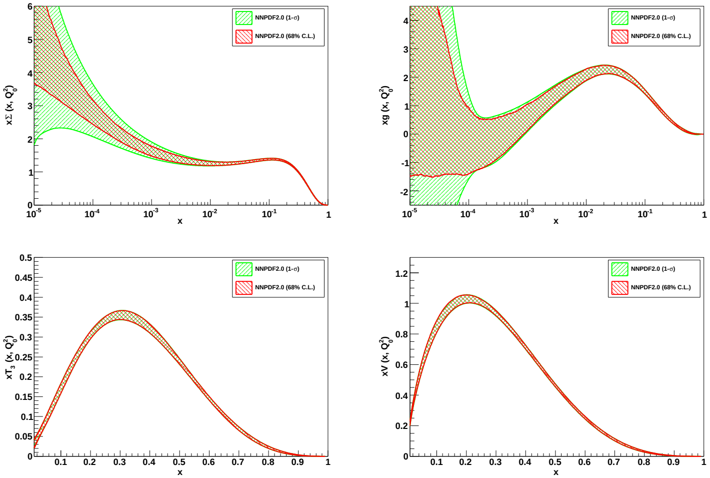
Results

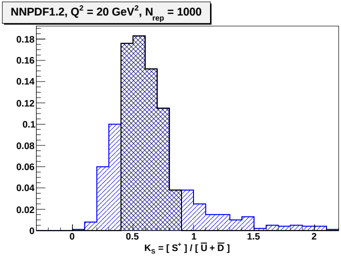
Impact of positivity

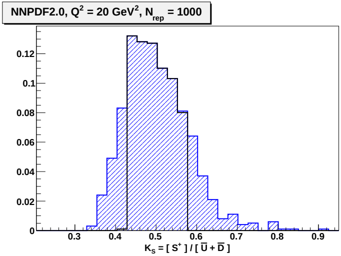


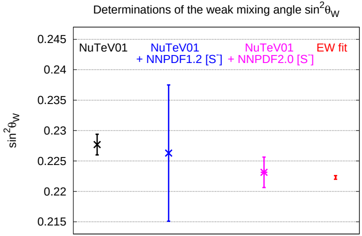
Results

Gaussian behaviour





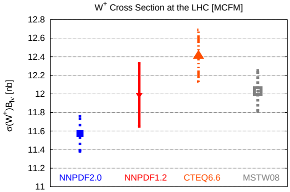


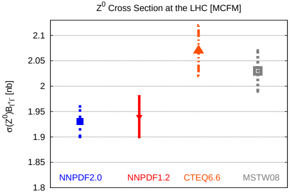


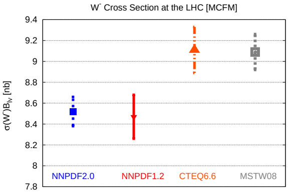
$$K_S = \begin{cases} 0.71^{+0.19}_{-0.31}\,{}^{\rm stat} \pm 0.26^{\rm syst} & \text{(NNPDF1.2)} \\ 0.503 \pm 0.075^{\rm stat}; & \text{(NNPDF2.0)} \end{cases}$$

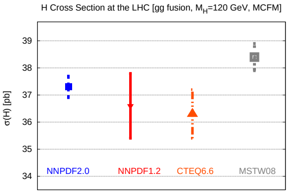
$$R_S = \begin{cases} 0.006 \pm 0.045^{\rm stat} \pm 0.010^{\rm syst} & \text{(NNPDF1.2)} \\ 0.019 \pm 0.008^{\rm stat} & \text{(NNPDF2.0)} \end{cases}$$

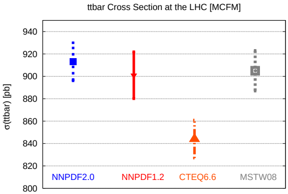
- Uncertainty reduced by addiction of DY data
- Striking agreement with EW fits





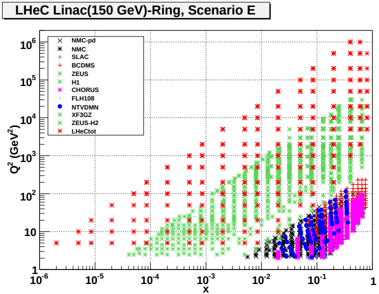


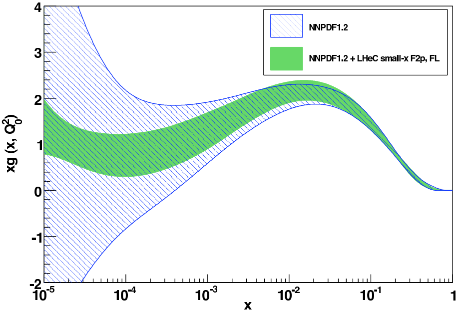




- HQ treatment?
- Impact of K factor approximation?
- Impact of rigid parametrization?

- Generate LHeC pseudo-data.
- Add them to the data set.
- Fit them (or reweight)





- Same settings: no tuning
- HIgh predictivity.

Outline

1. Introduction
   - Standard Approach
   - NNPDF
2. NNPDF method
   - Monte Carlo Determination of Errors
   - Neural Network as unbiased and redundant parametrization
   - Dynamical Stopping Criterion
3. NNPDF2.0: a global fit
   - New Features
   - Results
   - Impact of modifications
4. Conclusions and outlook

Conclusion and outlook

- Monte Carlo ensemble
  - Any statistical property of PDFs can be calculated using standard statistical methods.
  - No need of any tolerance criterion.
- The Neural Network parametrization
  - Small uncertainties come from an underlying physical law, not from par bias.
  - Inconsistent data or underestimated uncertainties do not require a separate treatment and are automatically signalled by a larger value of the χ^2.
- The NNPDF2.0 is the fist unbiased global NLO fit [FastKernel]
- Same consistent statistical behaviour under addition of hadronic data. No incompatibilities.
- Available on the common LHAPDF interface (http://projects.hepforge.org/lhapdf)
- The NNPDF2.X with FONLL (see Ref.ArXiv:1001.2312) inclusion of HQ mass effects soon available.
- The NNPDF2.Y NNLO fit is a work in progress.

BACKUP SLIDES

Pseudodata

$r\,[F]$	1.00000
$\langle\sigma^{(\mathrm{exp})}\rangle_{\mathrm{dat}}(\%)$	11.3
$\langle\sigma^{(\mathrm{gen})}\rangle_{\mathrm{dat}}(\%)$	11.4
$r\left[\sigma^{(\mathrm{gen})}\right]$	0.99996
$\langle\rho^{(\mathrm{exp})}\rangle_{\mathrm{dat}}$	0.176
$\langle\rho^{(\mathrm{gen})}\rangle_{\mathrm{dat}}$	0.179
$r\left[\rho^{(\mathrm{gen})}\right]$	0.99676

Experiment	$r[F]$	$\langle\sigma^{(\mathrm{exp})}\rangle_{\mathrm{dat}}$ (%)	$\langle\sigma^{(\mathrm{gen})}\rangle_{\mathrm{dat}}$ (%)	$r[\sigma]$
NMC-pd	1.000	1.78	1.72	0.999
NMC	1.000	4.91	4.89	0.998
SLAC	1.000	4.20	4.16	0.999
BCDMS	1.000	5.73	5.70	0.999
HERAI-AV	1.000	7.52	7.53	1.000
CHORUS	1.000	14.83	14.92	0.999
FLH108	1.000	71.90	70.78	1.000
NTVDMN	1.000	21.22	21.10	0.998
ZEUS-H2	1.000	13.79	13.56	1.000
DYE605	1.000	22.60	23.11	1.000
DYE866	1.000	20.76	20.73	1.000
CDFWASY	1.000	5.99	6.06	0.999
CDFZRAP	1.000	11.51	11.52	1.000
D0ZRAP	1.000	10.23	10.50	0.999
CDFR2KT	1.000	22.97	22.92	1.000
D0R2CON	1.000	16.82	17.18	1.000

Ingredient number 3: Evolution code

* To train NN we need to evolve from Q_0^2 to the experimental scales.

$$f_i(x, Q^2) = \sum_j \Gamma_{ij}(x, \alpha_s, \alpha_s^0) \otimes f_j(x, Q_0^2)$$

* Observables are a convolution over x of PDFs and Coefficient Functions.

$$F_I(x, Q^2) = \sum_j C_{Ij}(x, \alpha_s) \otimes f_j(x, Q^2) = \sum_{j,k} C_{Ij}(x, \alpha_s) \otimes \Gamma_{jk}(x, \alpha_s, \alpha_s^0) \otimes f_k(x, Q_0^2)$$

We want: Mellin space evolution

$$K_{Ik}(N, \alpha_s, \alpha_s^0) = \sum_j C_{Ij}(N, \alpha_s) \Gamma_{jk}(N, \alpha_s, \alpha_s^0)$$

We do not want: Complex Neural Networks

$$\begin{aligned} K_{Ik}(y, \alpha_s, \alpha_s^0) &= \frac{1}{2\pi i} \int_{\mathcal{C}} dN\, y^{-N} K_{Ik}(N, \alpha_s, \alpha_s^0) \\ F_I(x, Q^2) &= \sum_k \int_x^1 \frac{dy}{y} K_{Ik}(y, \alpha_s, \alpha_s^0)\, f_k(\frac{x}{y}, Q_0^2) \end{aligned}$$

F_2 proton structure function

$$F_2^p = x\{\frac{5}{18}C^s_{2,q}\otimes\Sigma + \frac{1}{6}C_{2,q}\otimes(T_3+\frac{1}{3}(T_8-T_{15})+\frac{1}{5}(T_{24}-T_{35})) + \langle e_q^2\rangle C_{2,g}\otimes g\}$$

$$F_2^p = x\{K_{\mathrm{F2},\Sigma}\otimes\Sigma_0 + K_{\mathrm{F2},g}\otimes g_0 + K_{\mathrm{F2},+}\otimes\left(T_{3,0}+\frac{1}{3}(T_{8,0}-T_{15,0})\right)\}$$

In Mellin space

$$K_{\mathrm{F2},\Sigma} = \frac{5}{18}C^s_{2,q}\Gamma^{qq}_{\mathrm{S}} + \frac{1}{30}C_{2,q}(\Gamma^{24,q}_{\mathrm{S}} - \Gamma^{35,q}_{\mathrm{S}}) + \langle e_q^2\rangle C_{2,g}\Gamma^{gq}_{\mathrm{S}}$$

$$K_{\mathrm{F2},g} = \frac{5}{18}C^s_{2,q}\Gamma^{qg}_{\mathrm{S}} + \frac{1}{30}C_{2,q}(\Gamma^{24,g}_{\mathrm{S}} - \Gamma^{35,g}_{\mathrm{S}}) + \langle e_q^2\rangle C_{2,g}\Gamma^{gg}_{\mathrm{S}}$$

$$K_{\mathrm{F2},+} = \frac{1}{6}C_{2,q}\Gamma^{+}_{\mathrm{NS}}$$

FastKernel

Schematically, take

$$\sum_{i,j=1}^{N_q} a_{ij} \int_{x_1}^{1} \frac{dy_1}{y_1} \int_{x_2}^{1} \frac{dy_2}{y_2} f_i(y_1) f_j(y_2)\, C^{\rm ij}(y_1, y_2)$$

Define an interpolation grid:

$$x_{\rm min} \equiv x_1 < x_2 < \ldots < x_{N_x - 1} < x_{N_x} \equiv 1$$

Around each of these grid points let us define a set of interpolating functions $\mathcal{I}^{(\alpha)}$:

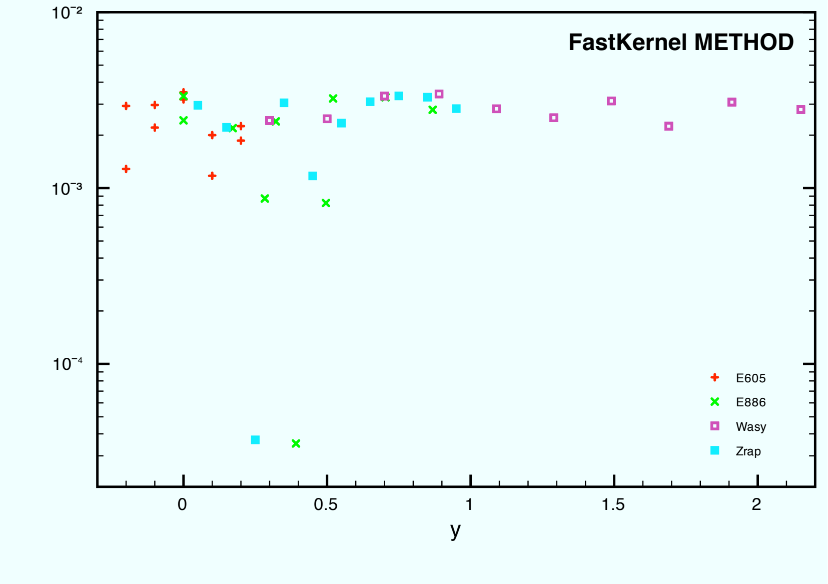
$$\mathcal{I}^{(\alpha)}(x_\alpha) = 1 \qquad \mathcal{I}^{(\alpha)}(x_\beta) = 0 \qquad \sum_{\alpha=1}^{N_x} \mathcal{I}^{(\alpha)}(y) = 1\,,\, \forall y$$

$$f_j(y) = \sum_{\alpha=1}^{N_x} f_j(x_\alpha)\, \mathcal{I}^{(\alpha)}(y)$$

Therefore, like in FastNLO we get:

$$\sum_{i,j=1}^{N_q} a_{ij} \sum_{\alpha,\beta=1}^{N_x} f_i(x_1^\alpha) f_j(x_2^\beta) \int_{x_1}^{1} \frac{dy_1}{y_1} \int_{x_2}^{1} \frac{dy_2}{y_2} \mathcal{I}^\alpha(y_1) \mathcal{I}^\beta(y_2)\, C^{\rm QQ}(y_1, y_2) = \sum_{\alpha,\beta=1}^{N_x} f_i(x_1^\alpha) f_j(x_2^\beta) c_{ij\alpha\beta}(x_1, x_2)$$

with $c_{ij\alpha\beta}$ and evolution precomputed (analogously).


FastKernel METHOD
10^{-2}
10^{-3}
10^{-4}
0
0.5
1
1.5
2
y
E605
E886
Wasy
Zrap

Distances

Thanks to the MC method, 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:

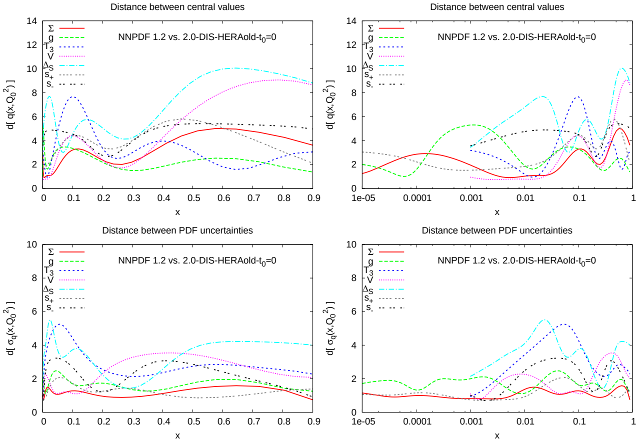
$$\langle f_i \rangle_{(1)} \equiv \frac{1}{N_{\mathrm{rep}}^{(1)}} \sum_{k=1}^{N_{\mathrm{rep}}^{(1)}} f_{ik}^{(1)} \, ,$$

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

* For statistically equivalent PDF sets: $\langle d[f] \rangle \sim \langle d[\sigma_f] \rangle \sim 1$
* 1σ distance when: $\langle d[f] \rangle \sim \langle d[\sigma_f] \rangle \sim 10$

Impact of modifications

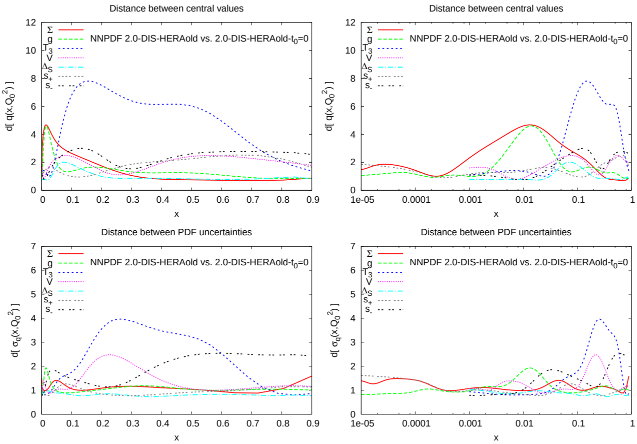
Improved minimization



- Improved quality of fixed–target DIS experiments (NMC, BCDMS, CHORUS)
- Impact on valence PDFs especially.

Fit	NNPDF1.2	NNPDF1.2+IGA	NNPDF1.2+IGA+t_0	2.0 DIS	2.0 DIS+JET	NNPDF2.0
χ^2_{tot}	1.32	1.16	1.12	1.20	1.18	1.21
$\langle E \rangle$	2.79	2.41	2.24	2.31	2.28	2.32
$\langle E_{\mathrm{tr}} \rangle$	2.75	2.39	2.20	2.28	2.24	2.29
$\langle E_{\mathrm{val}} \rangle$	2.80	2.46	2.27	2.34	2.32	2.35
$\langle \chi^{2(k)} \rangle$	1.60	1.28	1.21	1.29	1.27	1.29
NMC-pd	1.48	0.97	0.87	0.85	0.86	0.99
NMC	1.68	1.72	1.65	1.69	1.66	1.69
SLAC	1.20	1.42	1.33	1.37	1.31	1.34
BCDMS	1.59	1.33	1.25	1.26	1.27	1.27
HERAI	1.05	0.98	0.96	1.13	1.13	1.14
CHORUS	1.39	1.13	1.12	1.13	1.11	1.18
FLH108	1.70	1.53	1.53	1.51	1.49	1.49
NTVDMN	0.64	0.81	0.71	0.71	0.75	0.67
ZEUS-H2	1.52	1.51	1.49	1.50	1.49	1.51
DYE605	*11.19*	*22.89*	*8.21*	*7.32*	*10.35*	0.88
DYE866	*53.20*	*4.81*	*2.46*	*2.24*	*2.59*	1.28
CDFWASY	*26.76*	*28.22*	*20.32*	*13.06*	*14.13*	1.85
CDFZRAP	*1.65*	*4.61*	*3.13*	*3.12*	*3.31*	2.02
D0ZRAP	*0.56*	*0.80*	*0.65*	*0.65*	*0.68*	0.47
CDFR2KT	*1.10*	*0.95*	*0.78*	*0.91*	0.79	0.80
D0R2CON	*1.18*	*1.07*	*0.94*	*1.00*	0.93	0.93

Experiment	χ^2	$\langle E \rangle$	$\langle \sigma^{(\mathrm{exp})} \rangle_{\mathrm{dat}}(\%)$	$\langle \sigma^{(\mathrm{net})} \rangle_{\mathrm{dat}}(\%)$	$\langle \rho^{(\mathrm{exp})} \rangle_{\mathrm{dat}}$	$\langle \rho^{(\mathrm{net})} \rangle_{\mathrm{dat}}$
NMC-pd	0.99	2.05	1.8	0.5	0.03	0.36
NMC	1.69	2.79	4.9	1.7	0.16	0.77
SLAC	1.34	2.42	4.2	1.9	0.31	0.84
BCDMS	1.27	2.40	5.7	2.6	0.47	0.55
HERAI-AV	1.14	2.25	7.5	1.3	0.06	0.44
CHORUS	1.18	2.32	14.8	12.8	0.09	0.38
FLH108	1.49	2.51	71.9	3.3	0.65	0.68
NTVDMN	0.67	1.90	21.1	14.6	0.03	0.63
ZEUS-H2	1.51	2.66	13.6	1.2	0.29	0.58
DYE605	0.88	1.85	22.6	8.3	0.47	0.75
DYE866	1.28	2.35	20.8	9.1	0.20	0.45
CDFWASY	1.85	3.09	6.0	4.3	0.52	0.72
CDFZRAP	2.02	2.96	11.5	3.5	0.83	0.65
D0ZRAP	0.57	1.65	10.2	3.0	0.53	0.69
CDFR2KT	0.80	2.22	23.0	5.2	0.78	0.67
D0R2CON	0.93	1.92	16.2	6.0	0.78	0.64



- DIS fixed–target experiment have larger normalization uncertainties.
- Impact on valence PDFs especially.

- Many independent data sets with a different overall normalization uncertainty.

$$\sigma_{\rm norm}^{\rm HERA} \sim 0.5\% \qquad \sigma_{\rm norm}^{\rm E605} \sim 15\%$$

- Methods in common use (penalty trick) lead to systematic biases.
- Introduce the $t0$ prescription [ArXiv:0912.2276]

$$(\mathrm{cov}_{t0})_{IJ} = \left(\sum_{l=1}^{N_c} \sigma_{I,l}\sigma_{J,l} + \delta_{IJ}\sigma_{I,s}^2 \right) F_I F_J + \left(\sum_{n=1}^{N_a} \sigma_{I,n}\sigma_{J,n} + \sum_{n=1}^{N_r} \sigma_{I,n}\sigma_{J,n} \right) F_I^{(0)} F_J^{(0)}$$

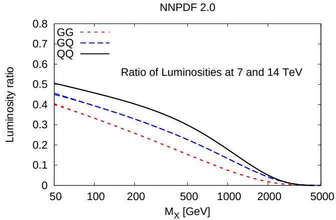
- F_I : measured central values for a give observable.
- $F_I^{(0)}$: corresponding observables computed recursively from previous fit.

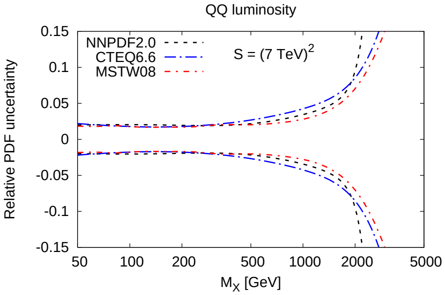
Fit	NNPDF1.2	NNPDF1.2+IGA	NNPDF1.2+IGA+t_0	2.0 DIS	2.0 DIS+JET	NNPDF2.0
$\chi^2_{\rm tot}$	1.32	1.16	1.12	1.20	1.18	1.21
$\langle E \rangle$	2.79	2.41	2.24	2.31	2.28	2.32
$\langle E_{\rm tr} \rangle$	2.75	2.39	2.20	2.28	2.24	2.29
$\langle E_{\rm val} \rangle$	2.80	2.46	2.27	2.34	2.32	2.35
$\langle \chi^{2(k)} \rangle$	1.60	1.28	1.21	1.29	1.27	1.29
NMC-pd	1.48	0.97	0.87	0.85	0.86	0.99
NMC	1.68	1.72	1.65	1.69	1.66	1.69
SLAC	1.20	1.42	1.33	1.37	1.31	1.34
BCDMS	1.59	1.33	1.25	1.26	1.27	1.27
HERAI	1.05	0.98	0.96	1.13	1.13	1.14
CHORUS	1.39	1.13	1.12	1.13	1.11	1.18
FLH108	1.70	1.53	1.53	1.51	1.49	1.49
NTVDMN	0.64	0.81	0.71	0.71	0.75	0.67
ZEUS-H2	1.52	1.51	1.49	1.50	1.49	1.51
DYE605	*11.19*	*22.89*	*8.21*	*7.32*	*10.35*	0.88
DYE866	*53.20*	*4.81*	*2.46*	*2.24*	*2.59*	1.28
CDFWASY	*26.76*	*28.22*	*20.32*	*13.06*	*14.13*	1.85
CDFZRAP	*1.65*	*4.61*	*3.13*	*3.12*	*3.31*	2.02
D0ZRAP	*0.56*	*0.80*	*0.65*	*0.65*	*0.68*	0.47
CDFR2KT	*1.10*	*0.95*	*0.78*	*0.91*	0.79	0.80
D0R2CON	*1.18*	*1.07*	*0.94*	*1.00*	0.93	0.93

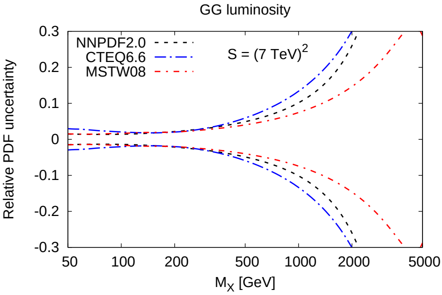
$$\Phi_{gg}\left(M_X^2\right) = \frac{1}{s}\int_\tau^1 \frac{dx_1}{x_1} g\left(x_1, M_X^2\right) g\left(\tau/x_1, M_X^2\right)$$

$$\Phi_{gq}\left(M_X^2\right) = \frac{1}{s}\int_\tau^1 \frac{dx_1}{x_1} \left[g\left(x_1, M_X^2\right) \Sigma\left(\tau/x_1, M_X^2\right) + (1 \to 2)\right]$$

$$\Phi_{qq}\left(M_X^2\right) = \frac{1}{s}\int_\tau^1 \frac{dx_1}{x_1} \sum_{i=1}^{N_f}\left[q_i\left(x_1, M_X^2\right) \bar{q}_i\left(\tau/x_1, M_X^2\right) + (1 \to 2)\right]$$

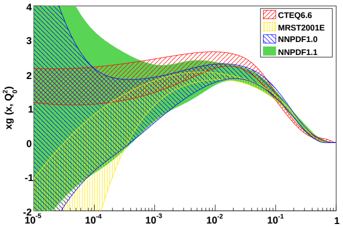




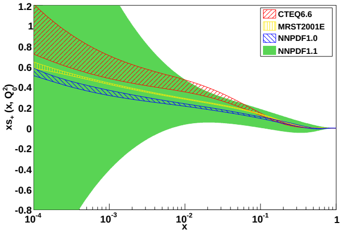


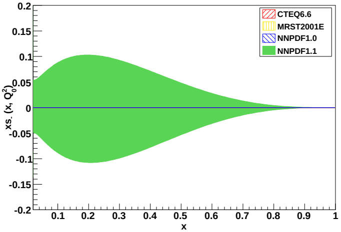
NNPDF1.1

PDFs are stable upon the addition of new independent PDFs parametrizations



- Larger in extrapolation region due to more flexibility (+ 60 pars).
- Same χ^2 and statistical features of the fit. Same gluon shape and error band.





	$\sigma(W^+)\mathrm{Br}\left(W^+ \to l^+\nu_l\right)$	$\sigma(W^-)\mathrm{Br}\left(W^- \to l^+\nu_l\right)$	$\sigma(Z^0)\mathrm{Br}\left(Z^0 \to l^+l^-\right)$
NNPDF1.2	11.99 ± 0.34 nb	8.47 ± 0.21 nb	1.94 ± 0.04 nb
NNPDF2.0	11.57 ± 0.19 nb	8.52 ± 0.14 nb	1.93 ± 0.03 nb
CTEQ6.6	12.41 ± 0.28 nb	9.11 ± 0.22 nb	2.07 ± 0.05 nb
MSTW08	12.03 ± 0.22 nb	9.09 ± 0.17 nb	2.03 ± 0.04 nb

	$\sigma(t\bar{t})$	$\sigma(H, m_H = 120\,\mathrm{GeV})$
NNPDF1.2	901 ± 21 pb	36.6 ± 1.2 pb
NNPDF2.0	913 ± 17 pb	37.3 ± 0.4 pb
CTEQ6.6	844 ± 17 pb	36.3 ± 0.9 pb
MSTW08	905 ± 18 pb	38.4 ± 0.5 pb