First global NNPDF analysis

Maria Ubiali

University of Edinburgh
& Université Catholique de Louvain

Rencontres de Moriond, QCD session
19$^{\text{th}}$ 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.[1,5]

[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
[5] CP3, Université Catholique de Louvain, Chemin du Cyclotron, 1348 Louvain-la-Neuve, Belgium

Outline

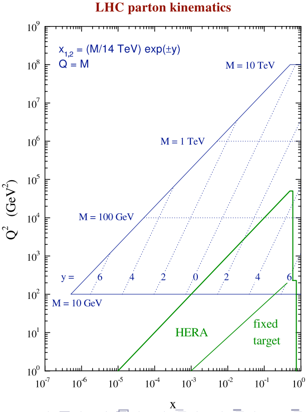
- Factorization Theorem ($Q^2 \gg \Lambda^2_{\mathrm{QCD}}$):

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

- 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.



- 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
$$q_i(x, Q_0^2) = A_i x^{b_i} (1-x)^{c_i} (1 + \ldots)$$
2. Determine best-fit values of parameters which define the functions
3. Errors determined via gaussian linear error propagation and large tolerance
$$\Delta\chi^2 \gg 1$$

Issues

1. Parametrization: how can we know that it is flexible enough and does not introduce a theoretical bias?
2. Large tolerance $T = \sqrt{\Delta\chi^2}$ means that errors are blown up by
$$S = \sqrt{\Delta\chi^2/2.7}$$
3. Benchmark partons do not agree with global partons within errors

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

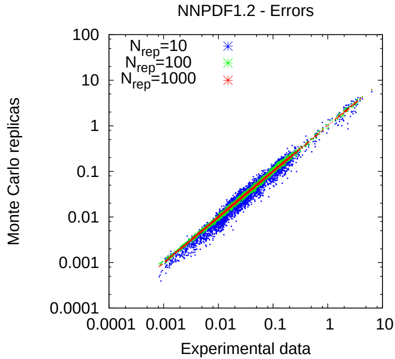
Use of neural network as redundant and unbiased parametrization

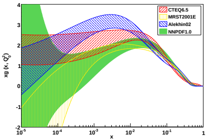
- Structure functions [hep-ph/0501067]
- Non-singlet PDF $q^- = u + d - (\bar{u} + \bar{d})$ [hep-ph/0701127]
- DIS global analysis: NNPDF1.0 [arXiv:0808.1231]
- Determination of the strange content: NNPDF1.2 [arXiv:0906.1958]
- Global (DIS+DY+JET) analysis: NNPDF2.0 [arXiv:1002.4407]

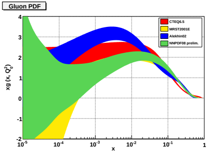
All sets are available in the LHAPDF interface

1) Monte Carlo behaves in a statistically consistent way

- Generate a Monte Carlo ensemble in the space of data
- *N* pseudo–data reproduce the probability distribution of the original experimental data.
- Precision of estimators scales as the size of the sample.







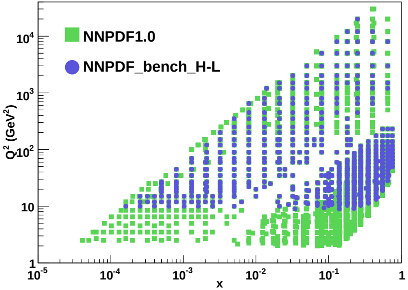
2) Results are shown to be independent of the parametrization and even stable upon the addition of independent PDFs parametrizations

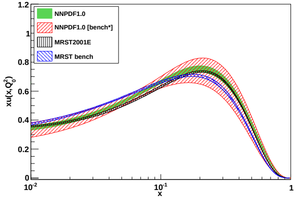
7x37 pars → **7x31 pars**

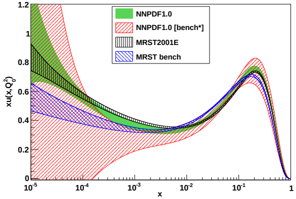
Motivation

What's nice about NNPDF #2

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

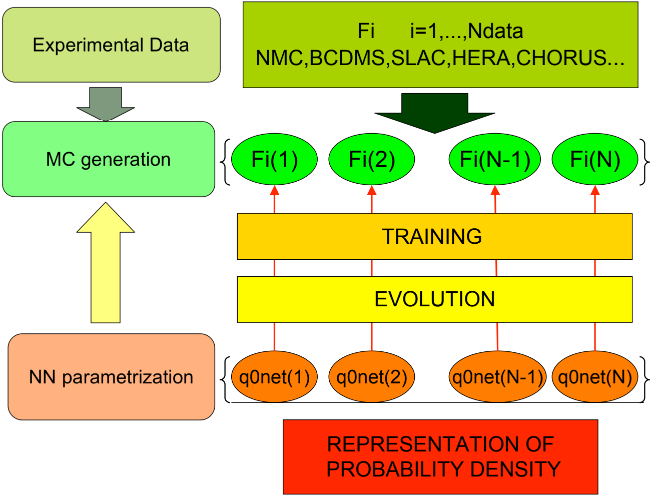






4) Control on PDFs uncertainties: NuteV anomaly solved AND V_{cs} precise determination at the same time [NNPDF1.2 [arXiv:0906.1958]]

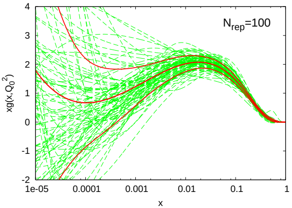
$$F_i^{\rm rep}(k) \;=\; S_{i,N}^{(k)}\, F_i^{\rm exp}\left(1 + r_i^{(k)}\sigma_i^{\rm stat} + \sum_{j=1}^{N_{\rm sys}} r_{i,j}^{(k)}\sigma_{i,j}^{\rm sys}\right) \qquad r_{i,j}^{(k)} = r_{i',j}^{(k)} \text{ if i and i' correlated}$$



1. Monte Carlo determination of errors:

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

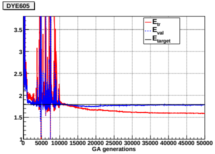
$$\sigma_{\mathcal{F}[f(x)]} = \sqrt{\langle \mathcal{F}[f(x)]^2 \rangle - \langle \mathcal{F}[f(x)] \rangle^2}$$



2. Neural Networks as redundant and unbiased parametrisation of PDFs:

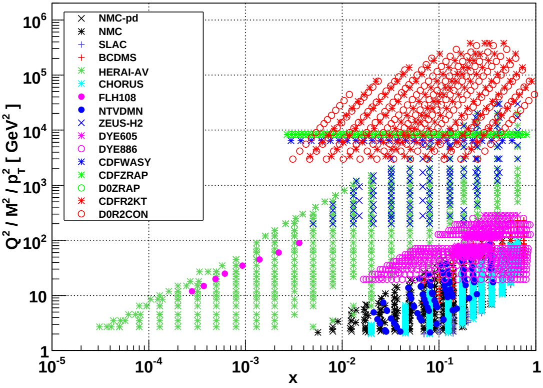
$\Sigma(x)$	$\longmapsto \mathrm{NN}_{\Sigma}(x)$	2-5-3-1	37 pars
$g(x)$	$\longmapsto \mathrm{NN}_{g}(x)$	2-5-3-1	37 pars
$V(x)$	$\longmapsto \mathrm{NN}_{V}(x)$	2-5-3-1	37 pars
$T_3(x)$	$\longmapsto \mathrm{NN}_{T3}(x)$	2-5-3-1	37 pars
$\Delta_S(x) \equiv \bar{d}(x) - \bar{u}(x)$	$\longmapsto \mathrm{NN}_{\Delta}(x)$	2-5-3-1	37 pars
$s^+(x) \equiv (s(x) + \bar{s}(x))/2$	$\longmapsto \mathrm{NN}_{(s+)}(x)$	2-5-3-1	37 pars
$s^-(x) \equiv (s(x) - \bar{s}(x))/2$	$\longmapsto \mathrm{NN}_{(s-)}(x)$	2-5-3-1	37 pars

3. Dynamical stopping criterion in order to fit data and not statistical noise (259 pars).



* Divide data in two sets: training and validation.
* 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 $\rightarrow$ STOP.

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
$d\sigma^{\mathrm{DY}}/dM^2dy$	E605
$d\sigma^{\mathrm{DY}}/dM^2dx_F$	E886
W asymmetry	CDF
Z rap. distr.	CDF,D0
incl. $\sigma^{(\mathrm{jet})}$	D0(cone) Run II
incl. $\sigma^{(\mathrm{jet})}$	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

NNPDF2.0

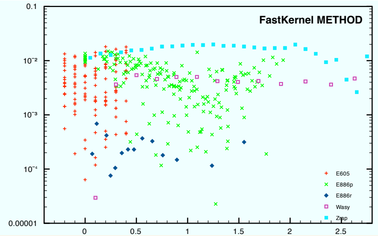
FastKernel

- NLO computation of hadronic observables too slow for parton global fits.
- MSTW08 and CTEQ include Drell-Yan NLO 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.

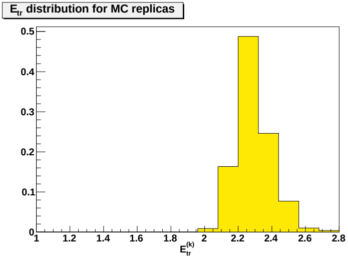


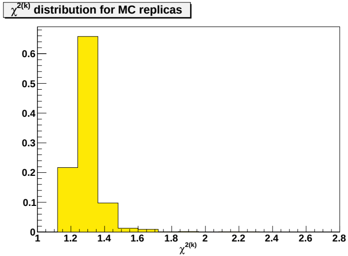
- Both PDFs evolution and double convolution sped up
- Use high-orders polynomial interpolation
- Precompute all Green Functions

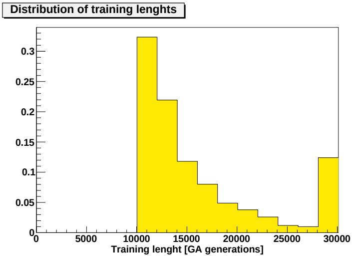
$$\int_{x_{0,1}}^{1} dx_1 \int_{x_{0,2}}^{1} dx_2\, f_a(x_1) f_b(x_2) C^{ab}(x_1, x_2) \to \sum_{\alpha,\beta=1}^{N_x} f_a(x_{1,\alpha}) f_b(x_{2,\beta}) \int_{x_{0,1}}^{1} dx_1 \int_{x_{0,2}}^{1} dx_2\, \mathcal{I}^{(\alpha,\beta)}(x_1, x_2) C^{ab}(x_1, x_2)$$

Results

Statistical features: global χ^2



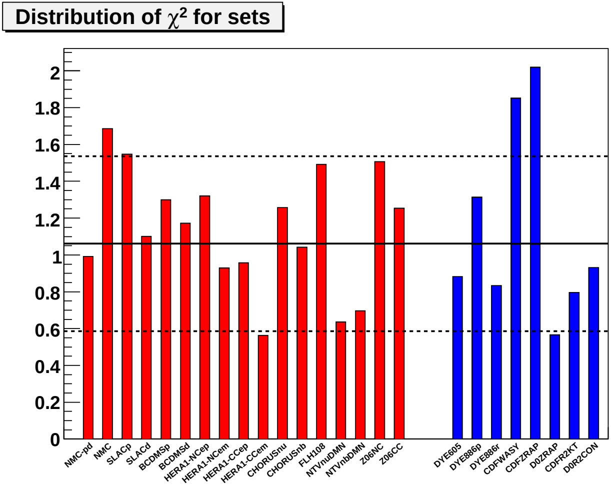




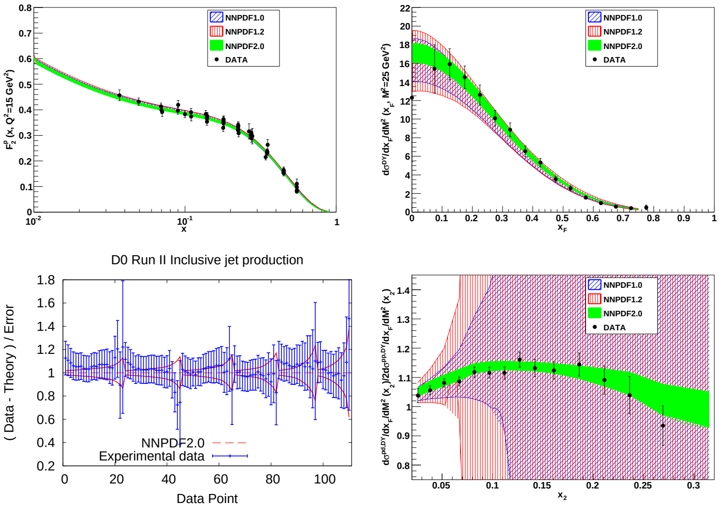
χ^2_{tot}	1.21
$\langle E \rangle \pm \sigma_E$	2.32 ± 0.10
$\langle E_{\mathrm{val}} \rangle \pm \sigma_{E_{\mathrm{val}}}$	2.29 ± 0.11
$\langle E_{\mathrm{val}} \rangle \pm \sigma_{E_{\mathrm{val}}}$	2.35 ± 0.12
$\langle TL \rangle \pm \sigma_{TL}$	16175 ± 6275
$\langle \chi^{2(k)} \rangle \pm \sigma_{\chi^2}$	1.29 ± 0.09

Results

Statistical features: individual experiments



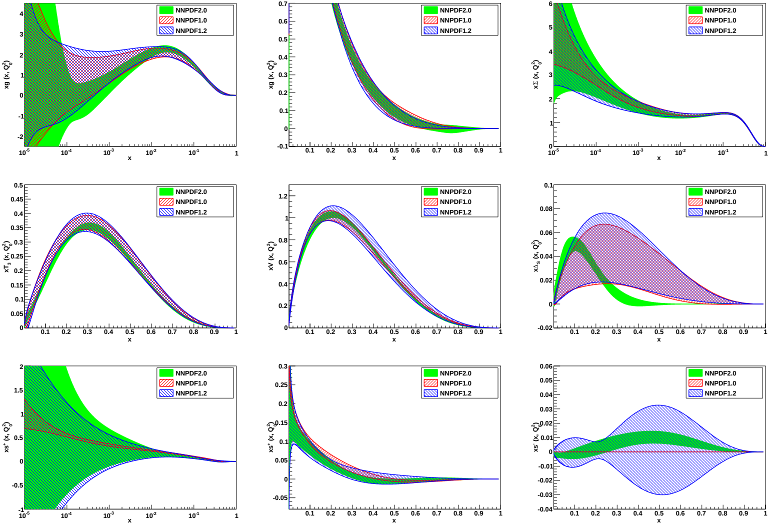
No obvious data incompatibility



NNPDF1.0
NNPDF1.2
NNPDF2.0
DATA
F_2^p (x, Q^2=15 GeV^2)
x
$d\sigma^{DY}/dx_F/dM^2$ (x_F, M^2=25 GeV^2)
x_F
D0 Run II Inclusive jet production
(Data - Theory) / Error
Data Point
NNPDF2.0
Experimental data
$d\sigma^{pd,DY}/dx_F/dM^2$ (x_2)/$2d\sigma^{pp,DY}/dx_F/dM^2$ (x_2)
x_2

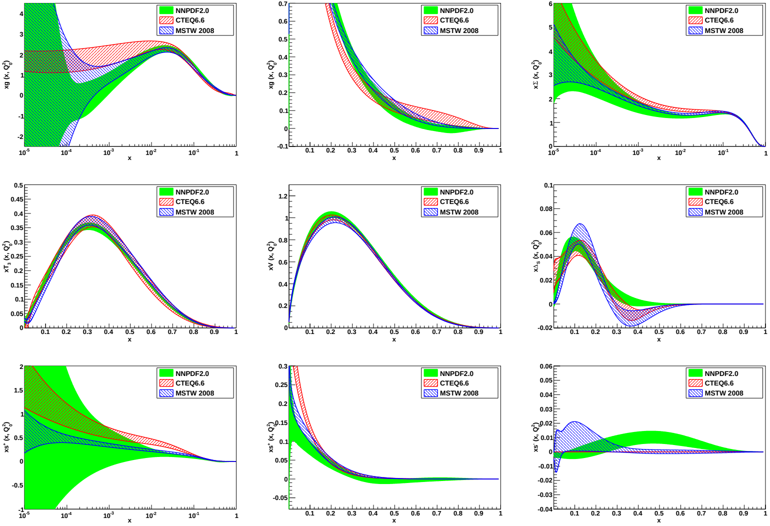
Results

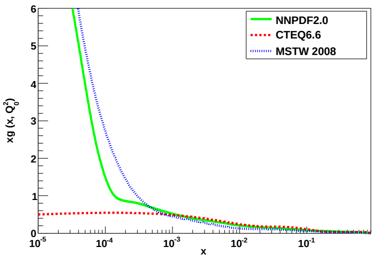
Partons

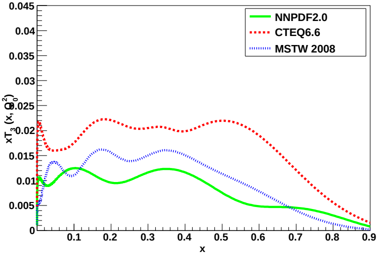


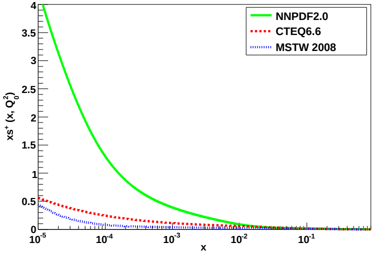
Results

Partons









- Reduction of uncertainties with respect to older NNPDF sets.
- Small uncertainties due to excellent data compatibility.
- Uncertainty larger than other groups when MSTW/CTEQ parametrizations are too restrictive.

- A quantitative assessment is possible

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

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

- Comparisons performed in NNPDF2.0 analysis

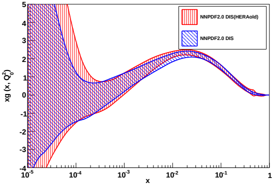
  1. Start from NNPDF1.2
  2. NNPDF1.2 vs. NNPDF1.2 + minimization/training improvements
  3. Improved NNPDF1.2 vs. Improved NNPDF1.2 + t_0-method
  4. Fit to DIS dataset with H1/ZEUS data vs. Fit with HERA-I combined
  5. Fit to DIS dataset vs. Fit to DIS+JET
  6. Fit to DIS+JET vs. NNPDF2.0 final

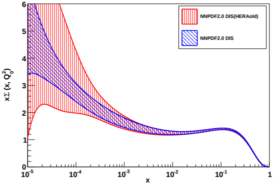
Impact of modifications

HERA-I combined dataset

Fit	NNPDF1.2	NNPDF1.2+IGA	NNPDF1.2+IGA+t_0	2.0 DIS
χ^2_{tot}	1.32	1.16	1.12	1.20
$\langle E \rangle$	2.79	2.41	2.24	2.31
$\langle \chi^{2(k)} \rangle$	1.60	1.28	1.21	1.29
HERAI	1.05	0.98	0.96	1.13

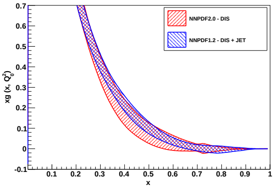
- HERA-I combined more precise.
- Quality of other data unchanged.
- Overall fit quality to the whole dataset is good
  - σ^{+}_{NC} dataset has relatively high $\chi^2 \sim 1.3$
  - σ^{-}_{CC} dataset has very low $\chi^2 \sim 0.55$
- Same χ^2-pattern observed in the HERAPDF1.0 analysis
- Impact on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

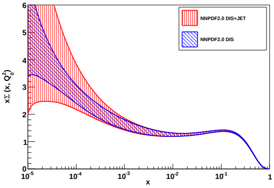




Fit	NNPDF1.2	NNPDF1.2+IGA	NNPDF1.2+IGA+t_0	2.0 DIS	2.0 DIS+JET
$\chi^2_{\rm tot}$	1.32	1.16	1.12	1.20	1.18
$\langle E \rangle$	2.79	2.41	2.24	2.31	2.28
$\langle \chi^{2(k)} \rangle$	1.60	1.28	1.21	1.29	1.27
CDFR2KT	*1.10*	*0.95*	*0.78*	*0.91*	0.79
D0R2CON	*1.18*	*1.07*	*0.94*	*1.00*	0.93

- Tevatron Run-II inclusive jet data provide a valuable constrain on large-x gluon.
- No incompatibility.
- Run-I data not included but compatibility with the outcome of the fit has been checked.



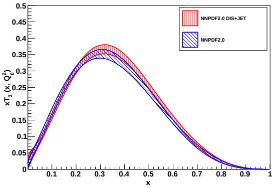


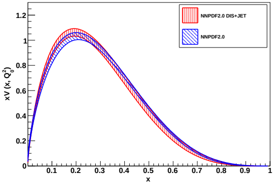
Impact of modifications

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
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
$\langle \chi^{2(k)} \rangle$	1.60	1.28	1.21	1.29	1.27	1.29
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

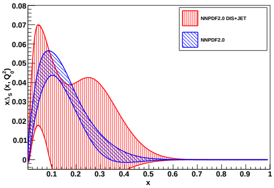
- Good description of fixed target Drell-Yan data (E605 proton and E886 proton and p/d ratio)
- Vector boson production at colliders (CDF W-asymmetry and Z rapidity distribution) harder to fit
- All valence-type PDF combinations are affected by these data
- Sizable reduction in the uncertainty of the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

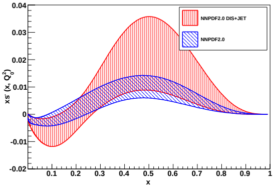




Fit	NNPDF1.2	NNPDF1.2+IGA	NNPDF1.2+IGA+t_0	2.0 DIS	2.0 DIS+JET	NNPDF2.0
$\chi^2_{\rm tot}$	1.32	1.16	1.12	1.20	1.18	1.21
$\langle E \rangle$	2.79	2.41	2.24	2.31	2.28	2.32
$\langle \chi^{2(k)} \rangle$	1.60	1.28	1.21	1.29	1.27	1.29
DYE605	*11.19*	*22.89*	*8.21*	*7.32*	*10.35*	0.88
DYE866	*53.20*	*4.81*	*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

- 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)



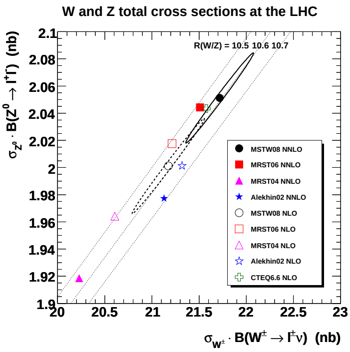


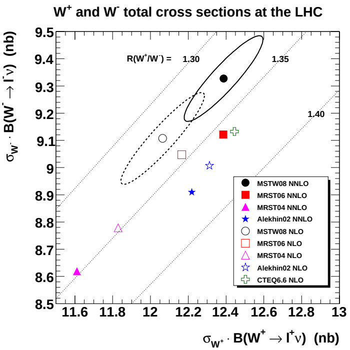
Conclusion and outlook

- 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 the data, not from bias due to functional form.
  - 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 signs of incompatibility between data!!!!
- Available on the common LHAPDF interface (http://projects.hepforge.org/lhapdf)
- The NNPDF2.X with FONLL (see Ref.ArXiv:1001.2312) with better treatment of heavy quarks mass will be soon available.
- The NNPDF2.Y NNLO fit is a work in progress.

THANK YOU!!

BACK-UP





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%

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

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

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

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

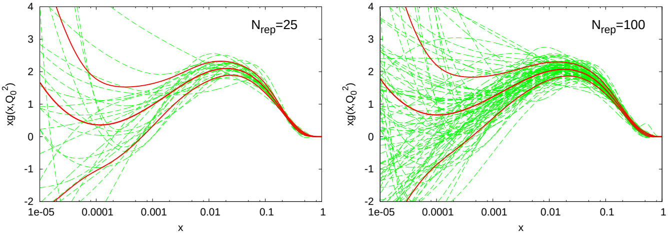
Correlations are properly taken into account.

NNPDF approach

Ingredient ♯1: Monte Carlo Errors

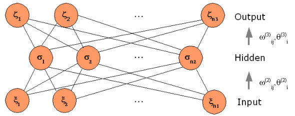
$$\langle \mathcal{F}[f(x)] \rangle = \frac{1}{N_{\rm rep}} \sum_{k=1}^{N_{\rm rep}} \mathcal{F}[f^{(k)({\rm net})}(x)]$$

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



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: $Q_0^2 = 2$ 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.

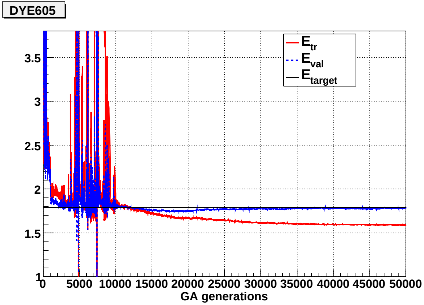
Ingredients of fitting procedure

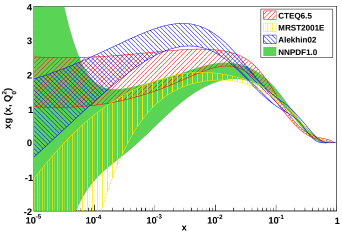
1. Flexible and redundant parametrization
2. Genetic Algorithm minimization
3. Dynamical stopping criterion: cross-validation technique

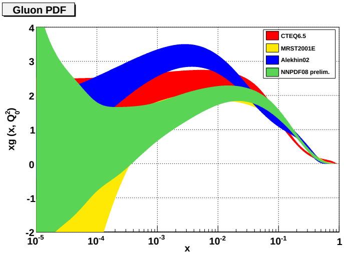
* GA is monotonically decreasing by construction.
* The best fit is not given by the absolute minimum.
* The best fit is given by an optimal training beyond which the figure of merit improves only because we are fitting statistical noise of the data.

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.



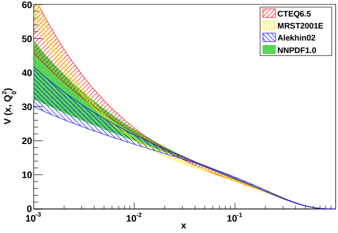


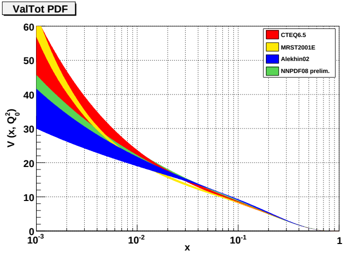


Results are statistically independent of the choice of the parametrization

	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
$\Sigma(x, Q_0^2)$	$5\ 10^{-4} \le x \le 0.1$	$10^{-5} \le x \le 10^{-4}$
$\langle d[f] \rangle$	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

PDFs are stable upon the addition of new 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

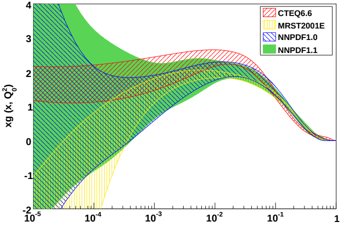
Strangeness valence : $s^-(x) \equiv (s(x) - \bar{s}(x))/2 \longmapsto \mathrm{NN}_{(s-)}(x)$ 2-5-3-1 37 pars

- Added two unconstrained PDFs.

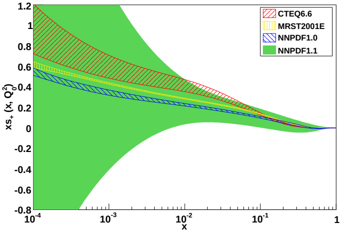
185 $\rightarrow$ 259 parameters

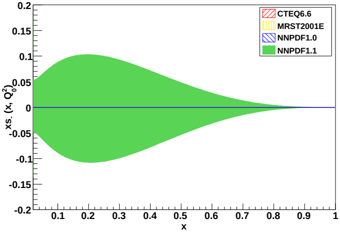
- Only strange (and Σ) affected. Gluon and statistical features remain unchanged.

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.





- 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

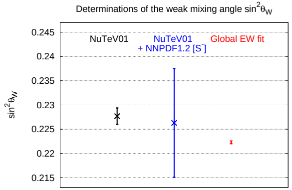
$\sin^2\theta_W = 0.2223 \pm 0.0002$

NuTeV

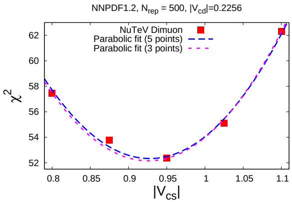
$\sin^2\theta_W = 0.2276 \pm 0.0014$

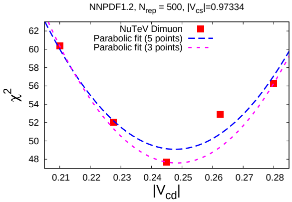
$$\delta_s \sin^2\theta_W \sim -0.240 \frac{[S^-]}{[Q^-]}$$

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



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

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

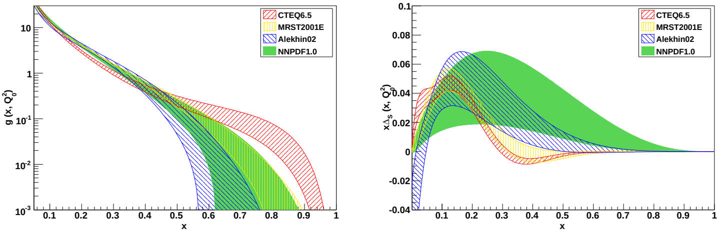
NNPDF1.2 analysis

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

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

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

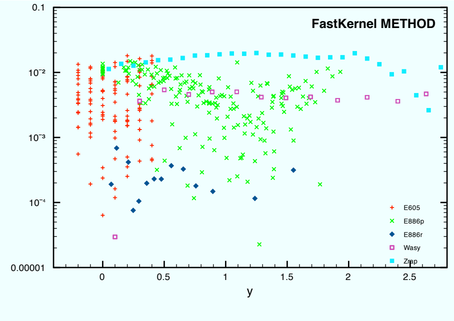
$$A^W = \frac{d\sigma^+_W/dy - d\sigma^-_W/dy}{d\sigma^+_W/dy - d\sigma^-_W/dy} \propto \frac{u\bar{d} - d\bar{u}}{u\bar{d} + d\bar{u}}$$



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

1. GLOBAL: Includes 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. NLO: Fast computation of Drell-Yan observables based on higher-order interpolating polynomials.
6. Enforced positivity constraints

- 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)	$e_{\rm rel}(u_V)$	$e_{\rm rel}(\Sigma)$	$e_{\rm rel}(g)$
$1 \cdot 10^{-7}$	$2.1 \cdot 10^{-4}$	$2.7 \cdot 10^{-5}$	$4.7 \cdot 10^{-6}$
$1 \cdot 10^{-6}$	$8.9 \cdot 10^{-5}$	$3.0 \cdot 10^{-5}$	$2.1 \cdot 10^{-5}$
$1 \cdot 10^{-5}$	$9.3 \cdot 10^{-5}$	$2.3 \cdot 10^{-5}$	$2.0 \cdot 10^{-5}$
$1 \cdot 10^{-4}$	$4.5 \cdot 10^{-5}$	$4.4 \cdot 10^{-5}$	$4.2 \cdot 10^{-5}$
$1 \cdot 10^{-3}$	$3.0 \cdot 10^{-5}$	$4.0 \cdot 10^{-5}$	$3.5 \cdot 10^{-5}$
$1 \cdot 10^{-2}$	$7.9 \cdot 10^{-5}$	$4.5 \cdot 10^{-5}$	$5.8 \cdot 10^{-5}$
$1 \cdot 10^{-1}$	$1.7 \cdot 10^{-4}$	$1.6 \cdot 10^{-5}$	$3.9 \cdot 10^{-5}$
$3 \cdot 10^{-1}$	$9.1 \cdot 10^{-6}$	$1.1 \cdot 10^{-5}$	$1.9 \cdot 10^{-7}$
$5 \cdot 10^{-1}$	$2.4 \cdot 10^{-5}$	$2.2 \cdot 10^{-5}$	$2.2 \cdot 10^{-5}$
$7 \cdot 10^{-1}$	$9.1 \cdot 10^{-5}$	$7.8 \cdot 10^{-5}$	$1.2 \cdot 10^{-4}$
$9 \cdot 10^{-1}$	$1.0 \cdot 10^{-3}$	$8.0 \cdot 10^{-4}$	$2.8 \cdot 10^{-3}$

- 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;

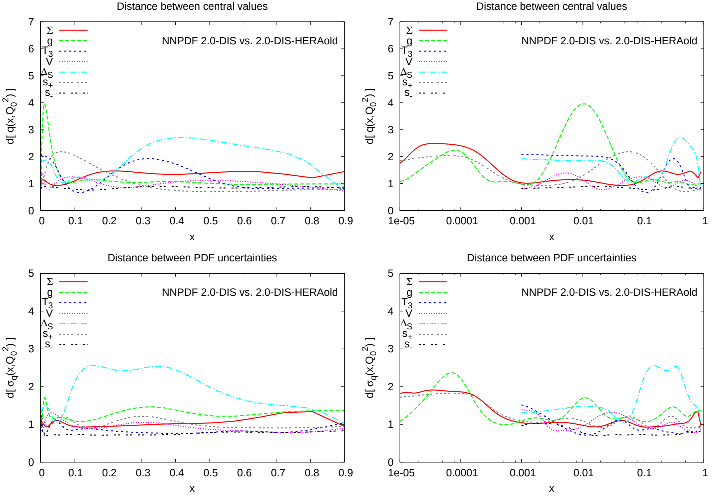
A truly NLO analysis

Impact of modifications

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

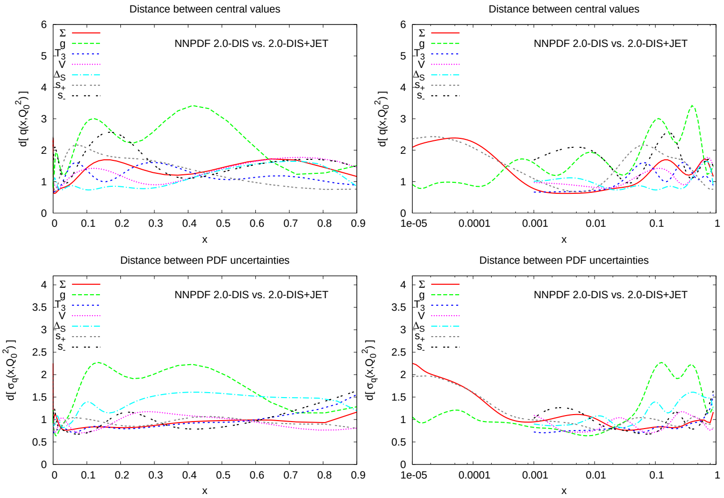
Impact of modifications

HERA-I combined dataset



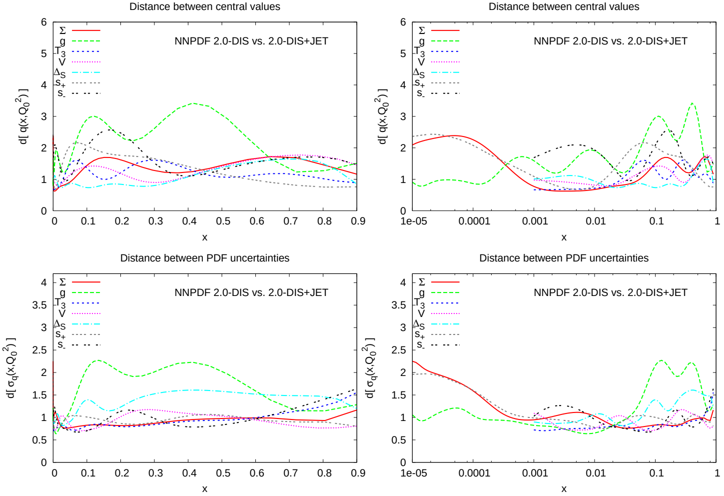
Impact of modifications

Drell-Yan and Vector Boson production data



Impact of modifications

Tevatron inclusive Jet data



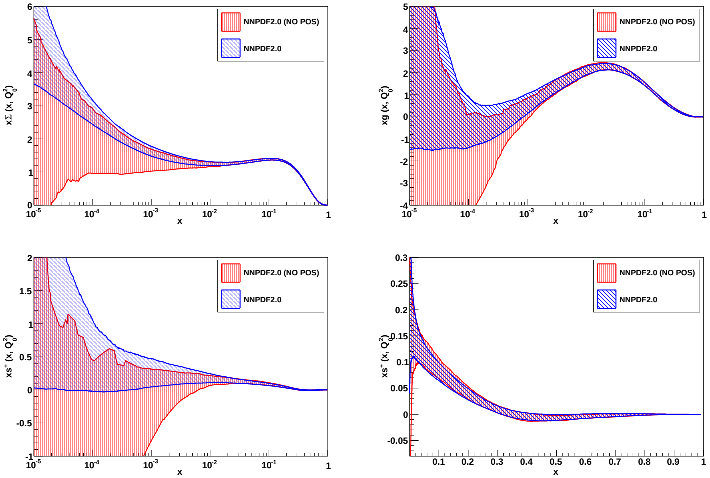
- We want the fitting procedure to explore only the subspace of acceptable physical solutions.
- We want
  - F_L positive.
  - Dimuon cross–section positive.
  - Momentum and valence sum rules.
- Modify the training function witha ddition of a Lagrangian multiplier:

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

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

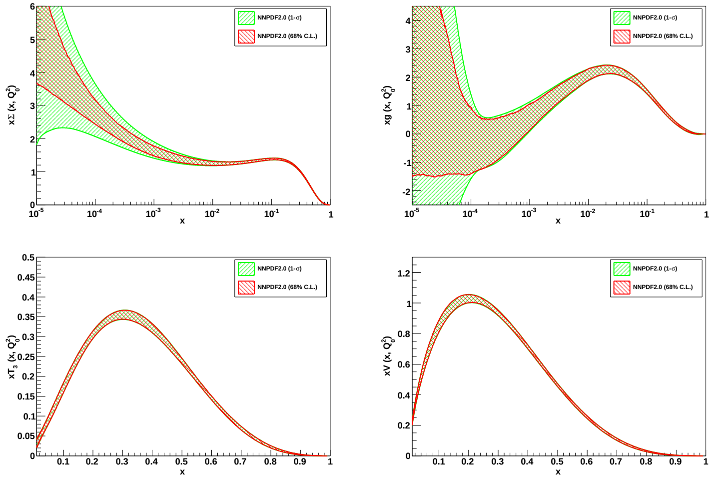
Results

Impact of positivity



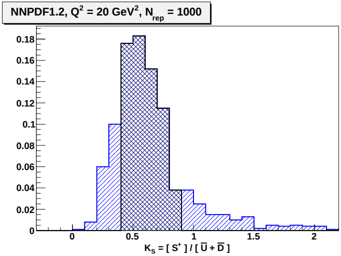
Results

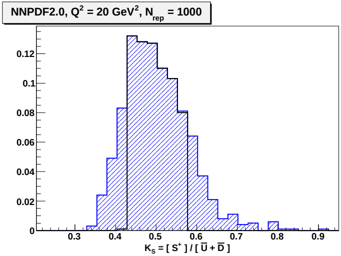
Gaussian behaviour

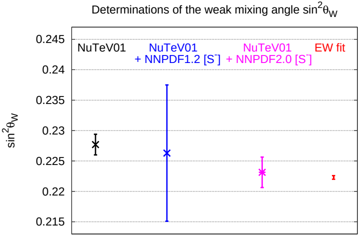


Some phenomenology

The proton strangeness revisited







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

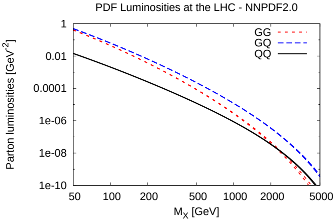
$$R_S = \begin{cases} 0.006 \pm 0.045^{\mathrm{stat}} \pm 0.010^{\mathrm{syst}} & \text{(NNPDF1.2)} \\ 0.019 \pm 0.008^{\mathrm{stat}} & \text{(NNPDF2.0)} \end{cases}$$

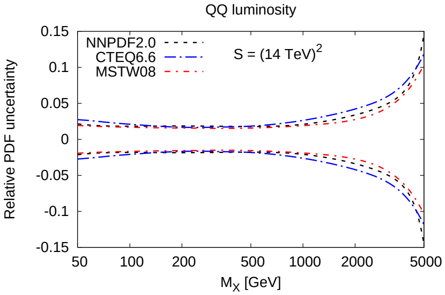
- Uncertainty reduced by addiction of DY data
- Striking agreement with EW fits

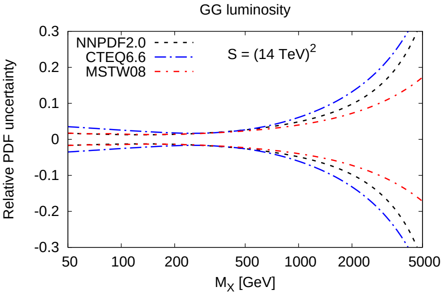
$$\Phi_{gg}\left(M_X^2\right) = \frac{1}{s}\int_\tau^1 \frac{dx_1}{x_1} g\left(x_1, M_X^2\right) g\left(\tau/x_1, M_X^2\right)$$

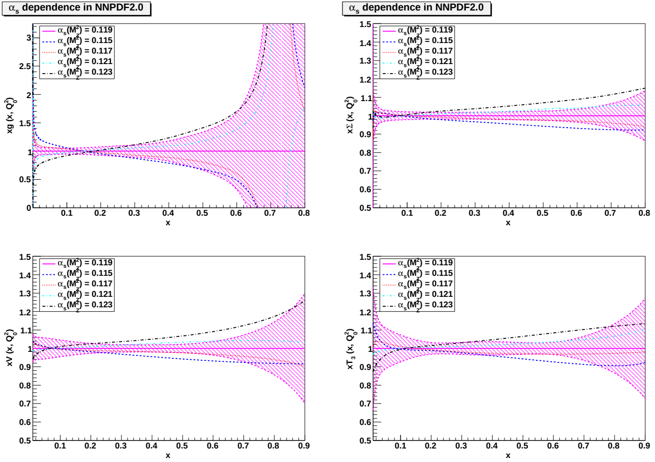
$$\Phi_{gq}\left(M_X^2\right) = \frac{1}{s}\int_\tau^1 \frac{dx_1}{x_1} \left[g\left(x_1, M_X^2\right) \Sigma\left(\tau/x_1, M_X^2\right) + (1 \to 2)\right]$$

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

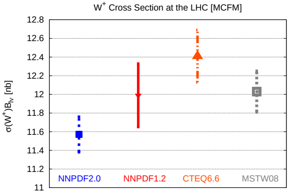


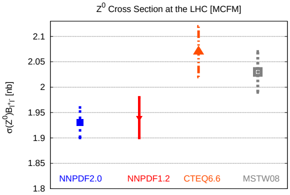


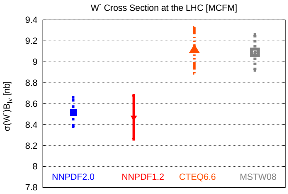


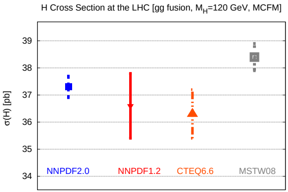


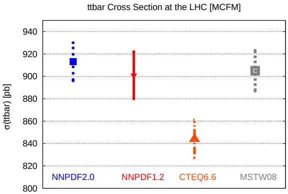
- Greater sensitivity to α_s than NNPDF1.2
- Greater NLO corrections for Drell-Yan observables.





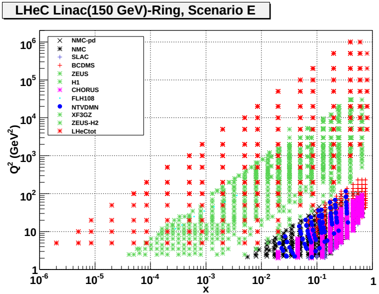


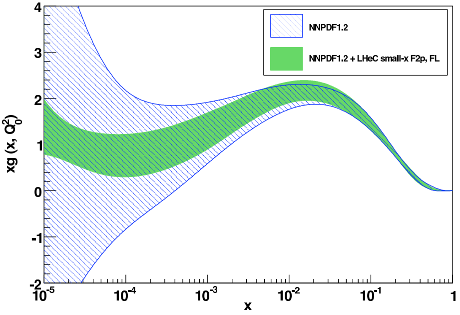




- HQ treatment?
- Impact of K factor approximation?
- Impact of rigid parametrization?

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





- Same settings: no tuning
- HIgh predictivity.

Impact of modifications

To conclude...

- 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!!!