The NNPDF approach to parton distributions: ideas, strangeness, and the JLAB12 program

Juan Rojo

Dipartimento di Fisica, Università degli Studi di Milano

Probing Strangeness in Hard Processes 2010
Laboratorio Nazionale di Frascati, INFN

Work in collaboration

NNPDF collaboration

R.D.Ball[1], V. Bertone[3], F.Cerutti[4], L.Del Debbio[1], S.Forte[2], A.Guffanti[3], J.I.Latorre[4], **J. R.**[2] and M. Ubiali[1]

[1] PPT Group, School of Physics, University of Edinburgh
[2] Dipartimento di Fisica, Università di Milano and INFN, Sezione Milano
[3] Physikalisches Institut, Albert-Ludwigs-Universität Freiburg

[4] Departament d'Estructura i Constituents de la Matèria, Universitat de Barcelona

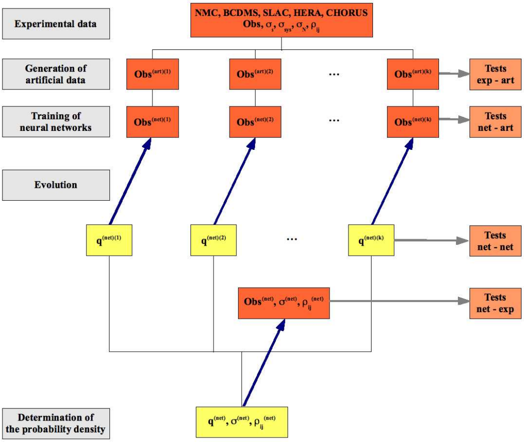
Outline

Outline

Motivation

- The accurate determination of unpolarized parton distributions (PDFs) is a crucial ingredient of the LHC program
- In the recent years a new approach, the NNPDF approach, has been developed that bypasses all the problems present in the standard approach to PDF determination: bias due to restrictive input functional forms, gaussian/linear approximations, lack of rigorous statistical interpretation of uncertainties ...
- While in this talk we will present results (mostly) about unpolarized NNPDFs, the methodology generalizes straightforwardly to other cases: fragmentation functions, TMDs, nuclear PDFs,...
- For example, see the NNPDF-inspired determination of Generalized Parton Distributions in K. Kumericki and D. Mueller, arXiv:1008.2762

The NNPDF approach to PDF determination



Monte Carlo errors

Non-gaussian errors and non trivial error propagation.

Neural Networks

Avoid bias from a restrictive fixed functional form.

Dynamical Stopping

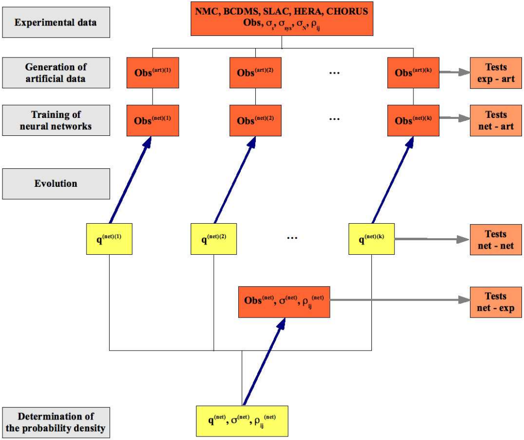
No looking for absolute minimum but learning from data.

Bayesian reweighting

Include new experimental information without refitting

N. B.: The general strategy holds also for polarized PDFs, fragmentation functions, nuclear PDFs, ...

Step 1: Monte Carlo Errors



Monte Carlo errors

Non-gaussian errors and non trivial error propagation.

Monte Carlo sample

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

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

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.
No need of linear, gaussian assumptions (unlike Hessian approach), exact error propagation

Monte Carlo Errors

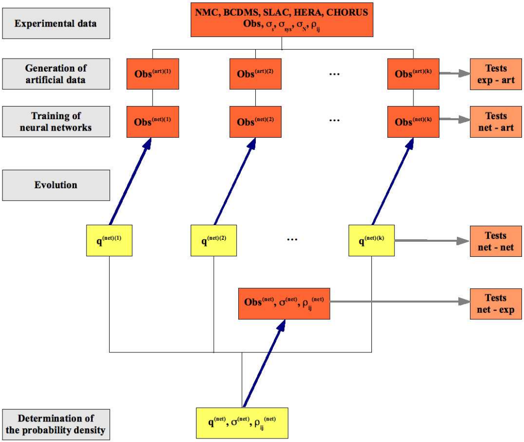
For each replica $^{(k)}$ of the experimental data we fit a set of independent PDFs

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

Uncertainties, central values, correlations ... of PDFs and functions of them evaluated using textbook statistical methods.
Rigorous statistics, no need of arbitrary *tolerances* $T \equiv \Delta\chi^2 \gg 1$

$$
\begin{aligned}
\langle \mathcal{F}[f(x)] \rangle &= \frac{1}{N_{\rm rep}} \sum_{k=1}^{N_{\rm rep}} \mathcal{F}[f^{(k)({\rm net})}(x)] \\
\sigma_{\mathcal{F}[f(x)]} &= \sqrt{\langle \mathcal{F}[f(x)]^2 \rangle - \langle \mathcal{F}[f(x)] \rangle^2} \\
\rho[f_a(x_1,Q_1^2), f_b(x_2,Q_2^2)] &= \frac{\langle f_a(x_1,Q_1^2) f_b(x_2,Q_2^2) \rangle - \langle f_a(x_1,Q_1^2) \rangle \langle f_b(x_2,Q_2^2) \rangle}{\sigma_a(x_1,Q_1^2)\sigma_b(x_2,Q_2^2)}
\end{aligned}
$$

Step 2: Neural Network as unbiased and redundant parametrization



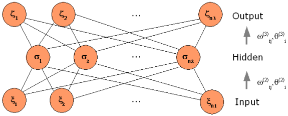
Neural Networks

Avoid bias from a restrictive fixed functional form.

Recall: in the standard approach, restrictive input shape for PDFs based on theoretical prejudices

What are neural networks?

Each independent PDF at the initial scale 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:

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

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

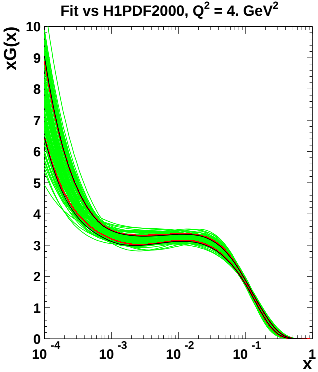
$$\xi_1^{(3)} = \frac{1}{1 + e^{\theta_1^{(3)} - \frac{\omega_{11}^{(2)}}{1+e^{\theta_1^{(2)} - \xi_1^{(1)}\omega_{11}^{(1)}}} - \frac{\omega_{12}^{(2)}}{1+e^{\theta_2^{(2)} - \xi_1^{(1)}\omega_{21}^{(1)}}}}}$$

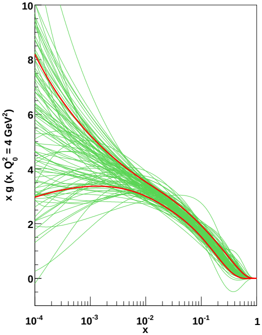
7 parameters

...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.

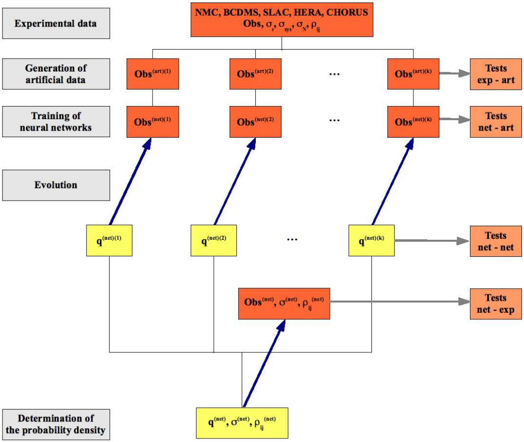
Simple functional froms vs. NeuralNets





- PDFs parametrized with simple functional forms ⟶ May result in systematic underestimation of PDF uncertainties
- The use of an universal interpolant like Artificial Neural Networks avoids any theoretical bias from choice of input PDF functional form
- Compare $\mathcal{O}$(300) parms in NNPDF with $\mathcal{O}$(10-25) parms in CTEQ,MSTW, DSSV, ...

Step 3: Training and dynamical stopping

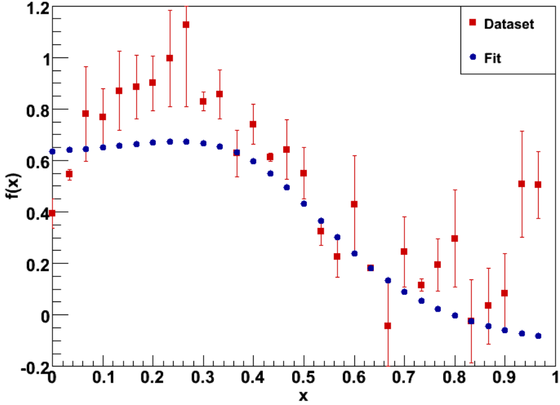


Dynamical Stopping

No looking for absolute minimum but learning from data.

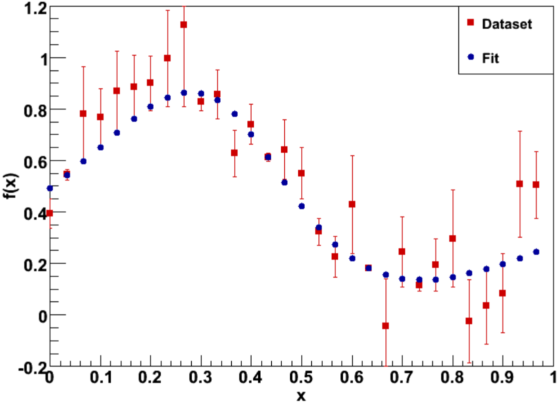
Neural network learning

We need to train to avoid under-learning ...



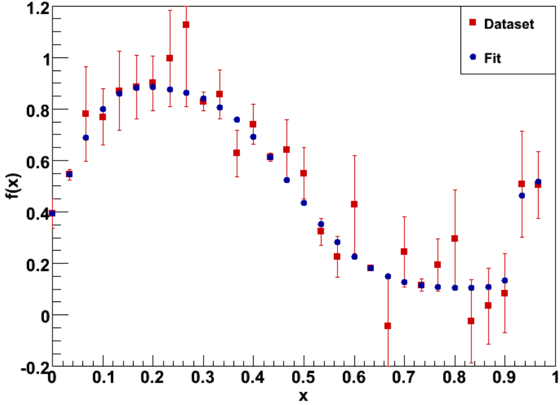
Neural network learning

... until we arrive to proper learning ....

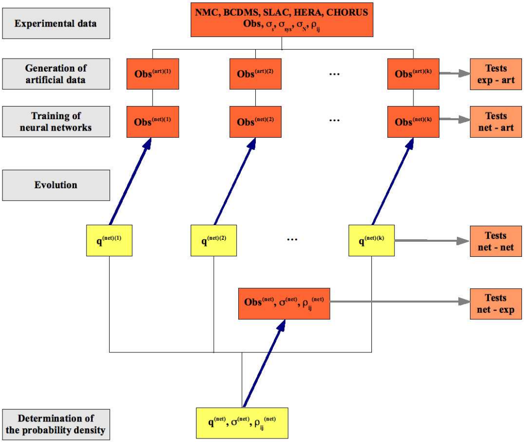


Neural network learning

... but be careful to avoid overlearning!



Step 4: PDF reweighting: Include new data without refitting



Bayesian reweighting

Include new experimental information without refitting

Bayesian reweighting

- Given the probability distribution of the PDFs represented by $N_{\rm rep}$ instances, $\{\rm PDF_i\}$, $\rm i = 1, N_{rep}$
- Given a set of new experimental data (or simulated pseudo-data) $y = \{y_1, y_2, \cdots, y_n\}$
- Given a general functional of the PDFs, $\mathcal{O}[\{\rm PDF\}]$

The impact of the new data y on the PDFs can be determined using Bayesian inference without refitting : PDF reweighting

$$\langle\mathcal{O}\rangle_{\rm old} = \sum_{k=1}^{N_{\rm rep}} \frac{1}{N_{\rm rep}} \mathcal{O}[f_k] \quad \rightarrow \quad \langle\mathcal{O}\rangle_{\rm new} = \sum_{k=1}^{N} w_k\, \mathcal{O}[f_k]$$

$$w_k \equiv \mathcal{N}_\chi(\chi^2(y, f_k))^{n/2-1} e^{-\frac{1}{2}\chi^2(y,f_k)},$$

$\chi^2(y, f_k) \rightarrow \chi^2$ of new data for the prediction of $\rm PDF_k$
No need of any refitting! Applications to the JLAB12 program later ...

Bayesian reweighting

- Given the probability distribution of the PDFs represented by $N_{\rm rep}$ instances, $\{\rm PDF_i\}$, $\rm i = 1, N_{rep}$
- Given a set of new experimental data (or simulated pseudo-data) $y = \{y_1, y_2, \cdots, y_n\}$
- Given a general functional of the PDFs, $\mathcal{O}\left[\{\rm PDF\}\right]$

The impact of the new data y on the PDFs can be determined using Bayesian inference without refitting : PDF reweighting

$$\langle \mathcal{O} \rangle_{\rm old} = \sum_{k=1}^{N_{\rm rep}} \frac{1}{N_{\rm rep}} \mathcal{O}[f_k] \quad \rightarrow \quad \langle \mathcal{O} \rangle_{\rm new} = \sum_{k=1}^{N} w_k \, \mathcal{O}[f_k]$$

$$w_k \equiv \mathcal{N}_\chi (\chi^2(y, f_k))^{n/2-1} e^{-\frac{1}{2}\chi^2(y,f_k)},$$

$\chi^2(y, f_k) \rightarrow \chi^2$ of new data for the prediction of $\rm PDF_k$

No need of any refitting! Applications to the JLAB12 program later ...

Bayesian reweighting

- Given the probability distribution of the PDFs represented by $N_{\rm rep}$ instances, $\{\rm PDF_i\}$, $\rm i = 1, N_{rep}$
- Given a set of new experimental data (or simulated pseudo-data) $y = \{y_1, y_2, \cdots, y_n\}$
- Given a general functional of the PDFs, $\mathcal{O}\left[\{\rm PDF\}\right]$

The impact of the new data y on the PDFs can be determined using Bayesian inference without refitting : PDF reweighting

$$\langle \mathcal{O} \rangle_{\rm old} = \sum_{k=1}^{N_{\rm rep}} \frac{1}{N_{\rm rep}} \mathcal{O}[f_k] \quad \rightarrow \quad \langle \mathcal{O} \rangle_{\rm new} = \sum_{k=1}^{N} w_k \, \mathcal{O}[f_k]$$

$$w_k \equiv \mathcal{N}_\chi (\chi^2(y, f_k))^{n/2-1} e^{-\frac{1}{2}\chi^2(y, f_k)},$$

$\chi^2(y, f_k) \rightarrow \chi^2$ of new data for the prediction of $\rm PDF_k$

No need of any refitting! Applications to the JLAB12 program later ...

Bayesian reweighting

- Given the probability distribution of the PDFs represented by $N_{\rm rep}$ instances, $\{\rm PDF_i\}$, $\rm i = 1, N_{rep}$
- Given a set of new experimental data (or simulated pseudo-data) $y = \{y_1, y_2, \cdots, y_n\}$
- Given a general functional of the PDFs, $\mathcal{O}\left[\{\rm PDF\}\right]$

The impact of the new data y on the PDFs can be determined using Bayesian inference without refitting : PDF reweighting

$$\langle \mathcal{O} \rangle_{\rm old} = \sum_{k=1}^{N_{\rm rep}} \frac{1}{N_{\rm rep}} \mathcal{O}[f_k] \quad \to \quad \langle \mathcal{O} \rangle_{\rm new} = \sum_{k=1}^{N} w_k \, \mathcal{O}[f_k]$$

$$w_k \equiv \mathcal{N}_\chi (\chi^2(y, f_k))^{n/2-1} e^{-\frac{1}{2}\chi^2(y,f_k)},$$

$\chi^2(y, f_k) \to \chi^2$ of new data for the prediction of $\rm PDF_k$

No need of any refitting! Applications to the JLAB12 program later ...

Bayesian reweighting

- Given the probability distribution of the PDFs represented by $N_{\rm rep}$ instances, $\{\rm PDF_i\}$, $\rm i=1,N_{rep}$
- Given a set of new experimental data (or simulated pseudo-data) $y=\{y_1,y_2,\cdots,y_n\}$
- Given a general functional of the PDFs, $\mathcal{O}\left[\{\rm PDF\}\right]$

The impact of the new data y on the PDFs can be determined using Bayesian inference without refitting : PDF reweighting

$$\langle\mathcal{O}\rangle_{\rm old}=\sum_{k=1}^{N_{\rm rep}}\frac{1}{N_{\rm rep}}\mathcal{O}[f_k] \quad\rightarrow\quad \langle\mathcal{O}\rangle_{\rm new}=\sum_{k=1}^{N}w_k\,\mathcal{O}[f_k]$$

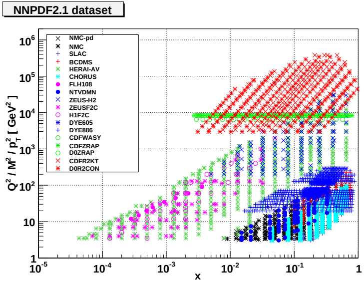
$$w_k\equiv\mathcal{N}_\chi(\chi^2(y,f_k))^{n/2-1}e^{-\frac{1}{2}\chi^2(y,f_k)},$$

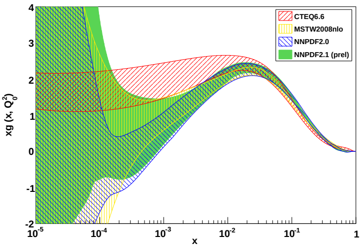
$\chi^2(y,f_k)\rightarrow\chi^2$ of new data for the prediction of $\rm PDF_k$
No need of any refitting! Applications to the JLAB12 program later ...

Outline

Unpolarized NNPDF: Status of the Art

- **NNPDF2.1** includes all relevant datasets from DIS, Drell-Yan, vector-boson production and inclusive jet production
- Only **red true NLO global fit**: consistent use of NLO pQCD (`FastKernel`, `FastNLO`), no K-factor approximations
- **NNPDF2.1** improves on any existing NLO PDF set

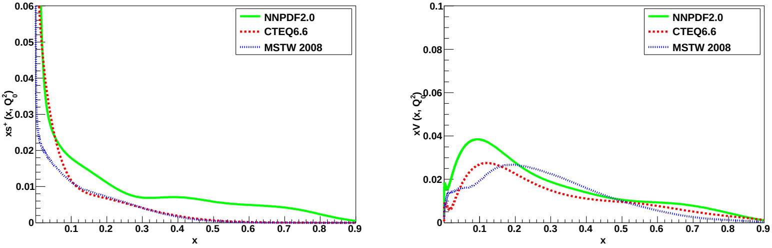




Strange PDF $s^+(x, Q^2)$

- The strange PDF is the less constrained of all light quark PDFs
- Experimental constraints: neutrino charm production and partly Drell-Yan
- Much larger uncertainties than u, d PDFs

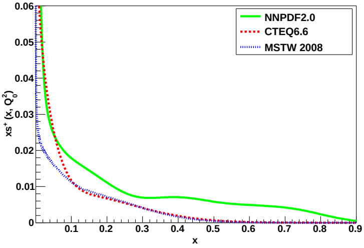
Absolute PDF uncertainties from modern PDF sets:

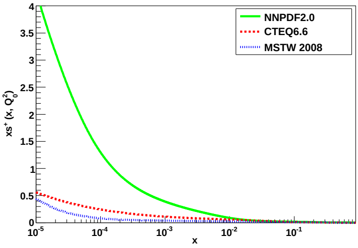


Strange PDF $s^+(x, Q^2)$

- The strange PDF is the less constrained of all light quark PDFs
- Experimental constraints: neutrino charm production and partly Drell-Yan
- Much larger uncertainties than u, d PDFs

Absolute PDF uncertainties from modern PDF sets:



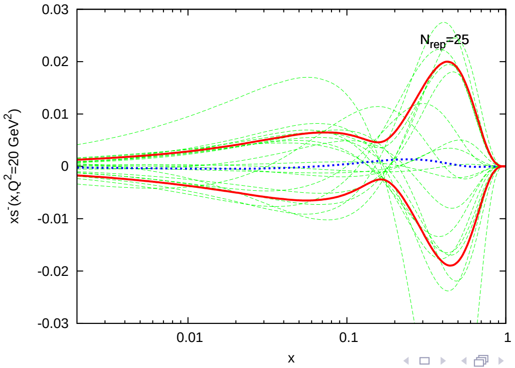


Note functional form biases in CTEQ6.6/MSTW08: no experimental constraints on xs^+ for $x \leq 10^{-2}$

Strange asymmetry PDF: $s^-(x, Q^2)$ in NNPDF1.2 (DIS-only fit)

- No theoretical constraints on $s^-(x, Q_0^2)$ apart from valence sum rule
- At least one crossing required by sum rule, but some replicas have two crossings
- Compare with more restrictive parametrizations

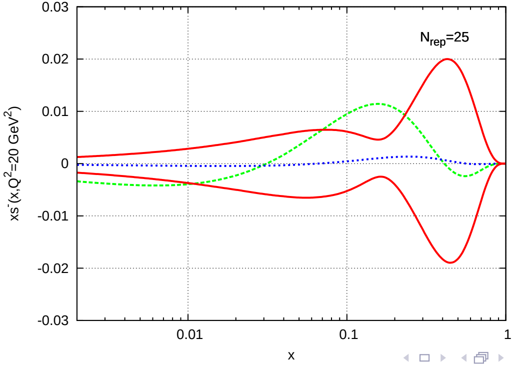
$$xs^-_{\rm mstw} = A_- x^{0.2} (1-x)^{\eta_-} (1 - x/x_0)$$



Strange asymmetry PDF: $s^-(x, Q^2)$ in NNPDF1.2 (DIS-only fit)

- No theoretical constraints on $s^-(x, Q_0^2)$ apart from valence sum rule
- At least one crossing required by sum rule, but some replicas have two crossings
- Compare with more restrictive parametrizations

$$xs^-_{\rm mstw} = A_- x^{0.2} (1-x)^{\eta_-} (1 - x/x_0)$$



Impact on NuTeV anomaly

Accurate determination of xs^- $\to$ Important phenomenological implications:

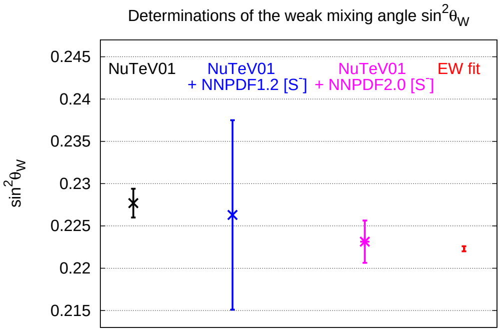
NuTeV anomaly: Discrepancy ($\geq 3\sigma$) between indirect (global fit)and direct (NuTeV neutrino scattering) determinations of $\sin^2\theta_W$

EW fit

$\sin^2\theta_W = 0.2223 \pm 0.0003$

NuTeV (assumes $[S^-]=0$)

$\sin^2\theta_W = 0.2277 \pm 0.0017$



NuTeV + NNPDF1.2 $[S^-]$

$\sin^2\theta_W = 0.2263 \pm 0.0017^{\rm exp} \pm 0.0107^{\rm PDFs}$

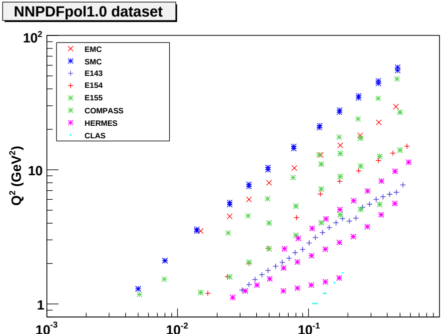
NuTeV + NNPDF2.0 $[S^-]$

$\sin^2\theta_W = 0.22314 \pm 0.00251$

Polarized NNPDF: Status of the Art

- Unbiased NNPDF analysis crucial for pPDFs, with much less experimental constraints than in unpolarized
- Inclusive polarized structure function data $g_1(x, Q^2)$ on proton, deuteron and neutron targets from spin asymmetries

$$g_1(x, Q^2) = A_1(x, Q^2)\frac{F_2(x, Q^2)}{2x(1+R(x, Q^2))}\left(1+\gamma^2\right) \ , \qquad \gamma^2 \equiv \frac{4M_N^2 x^2}{Q^2}$$



Theoretical constraints:

- * Sum rules

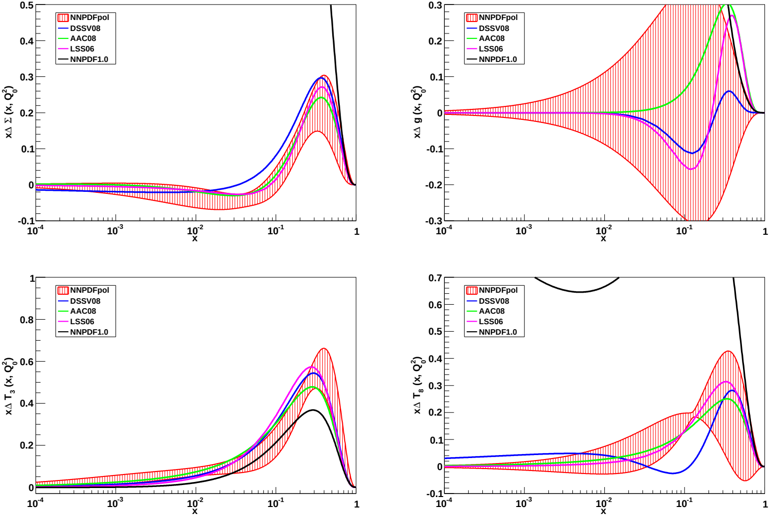
$$\left[\Delta T_3(Q_0^2)\right] \equiv \int_0^1 dx\ \Delta T_3(x, Q_0^2) = a_3 \ ,$$

$$\left[\Delta T_8(Q_0^2)\right] \equiv \int_0^1 dx\ \Delta T_8(x, Q_0^2) = a_8 \ ,$$

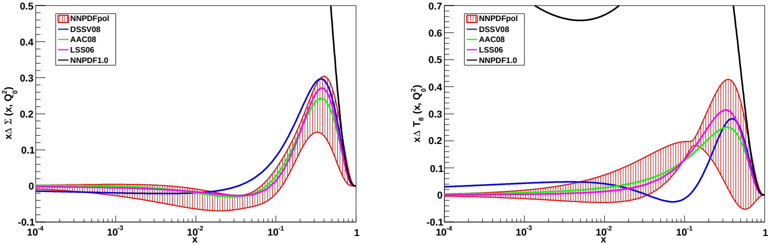
- * Positivity of polarized PDFs $\longrightarrow$ Constraints on polarized SFs: $|g_1(x, Q^2)| \leq F_1(x, Q^2)$ for all targets

- F_1^p, F_1^d, F_1^n computed consistently from NNPDF1.0

Polarized NNPDFs (Preliminary)



Polarized NNPDFs (Preliminary)



- The polarized gluon $\Delta g(x)$ is essentially unconstrained from inclusive polarized DIS only
- Polarized strangeness (in ΔT_8) poorly constrained, info from a_8 sum rule in inclusive DIS fit, large uncertainties
- More work required for quantitative phenomenology

Outline

NNPDF and the JLAB12 program

- The JLAB12 upgrade offers a rich program on the polarized and unpolarized proton structure at large-x
- Relevant both for our understanding of QCD but also for their implications to precision LHC physics
- Bayesian reweighting specially suitable to study impact of projected data onto PDFs without refitting
- Unbiased PDFs are crucial to faithfully assess impact of new data in extrapolation regions (non functional bias)
- Here we will discuss only two applications: constraining the d/u ratio from F_2^n/F_2^p data and constraining strangeness using kaon-tagged SIDIS

N. B.: Not an expert in JLAB physics – excuse inaccuracies ...

NNPDF and the JLAB12 program

- The JLAB12 upgrade offers a rich program on the polarized and unpolarized proton structure at large-x
- Relevant both for our understanding of QCD but also for their implications to precision LHC physics
- Bayesian reweighting specially suitable to study impact of projected data onto PDFs without refitting
- Unbiased PDFs are crucial to faithfully assess impact of new data in extrapolation regions (non functional bias)
- Here we will discuss only two applications: constraining the d/u ratio from F_2^n/F_2^p data and constraining strangeness using kaon-tagged SIDIS

N. B.: Not an expert in JLAB physics – excuse inaccuracies ...

NNPDF and the JLAB12 program

- The JLAB12 upgrade offers a rich program on the polarized and unpolarized proton structure at large-x
- Relevant both for our understanding of QCD but also for their implications to precision LHC physics
- Bayesian reweighting specially suitable to study impact of projected data onto PDFs without refitting
- Unbiased PDFs are crucial to faithfully assess impact of new data in extrapolation regions (non functional bias)
- Here we will discuss only two applications: constraining the d/u ratio from F_2^n/F_2^p data and constraining strangeness using kaon-tagged SIDIS

N. B.: Not an expert in JLAB physics – excuse inaccuracies ...

NNPDF and the JLAB12 program

- The JLAB12 upgrade offers a rich program on the polarized and unpolarized proton structure at large-x
- Relevant both for our understanding of QCD but also for their implications to precision LHC physics
- Bayesian reweighting specially suitable to study impact of projected data onto PDFs without refitting
- Unbiased PDFs are crucial to faithfully assess impact of new data in extrapolation regions (non functional bias)
- Here we will discuss only two applications: constraining the d/u ratio from F_2^n/F_2^p data and constraining strangeness using kaon-tagged SIDIS

N. B.: Not an expert in JLAB physics – excuse inaccuracies ...

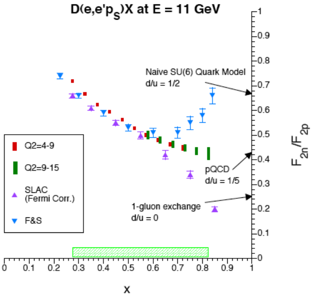
NNPDF and the JLAB12 program

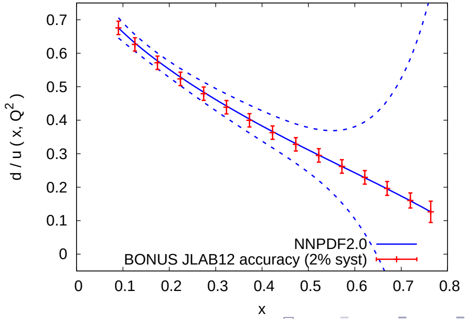
- The JLAB12 upgrade offers a rich program on the polarized and unpolarized proton structure at large-x
- Relevant both for our understanding of QCD but also for their implications to precision LHC physics
- Bayesian reweighting specially suitable to study impact of projected data onto PDFs without refitting
- Unbiased PDFs are crucial to faithfully assess impact of new data in extrapolation regions (non functional bias)
- Here we will discuss only two applications: constraining the d/u ratio from F_2^n/F_2^p data and constraining strangeness using kaon-tagged SIDIS

N. B.: Not an expert in JLAB physics – excuse inaccuracies ...

The d/u ratio at large-x

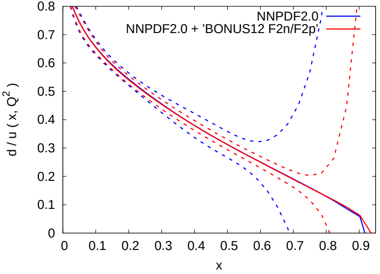
- Large-x PDFs affected by large uncertainties from lack of experimental constraints
- The d/u ratio on top might suffer from deuterium nuclear uncertainties
- Both problems solved by an accurate measurement of F_2^n at large-x
- Relevant for non-perturbative QCD, but also for LHC physics

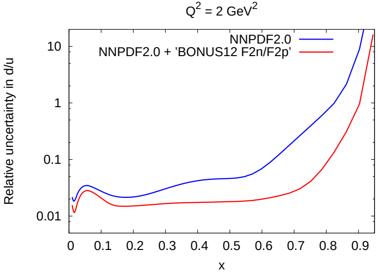




The d/u ratio at large-x

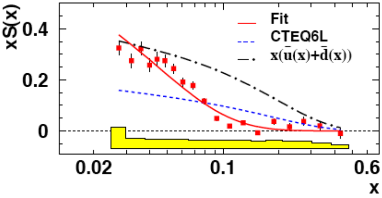
- Assess impact of pseudo-data from BONUS12 F_2^n/F_2^p measurements using PDF reweighting (data from E. Bueltmann)
  $\sigma_{\rm syst} = 2\%$ assumed, $W \geq 2$ GeV, $\mathcal{L} = 2 \cdot 10^{34} cm^{-2} s^{-1}$
- BONUS $F_2^n/F_2^p \rightarrow$ Huge reduction of PDF uncertainties in d/u ratio!
- Implications for LHC physics to be explored

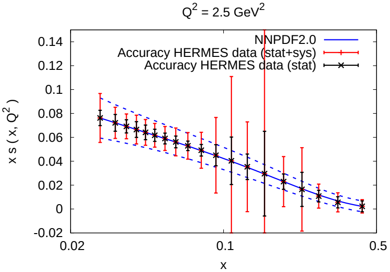




Unpolarized strangeness via SIDIS

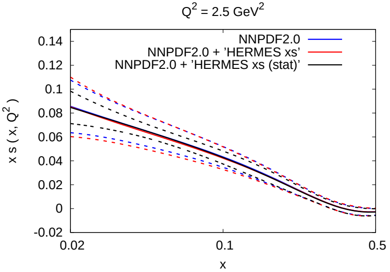
- HERMES has published measurements of kaon production in SIDIS (PLB 666(2008)446)
- At LO, knowledge of fragmentation functions leads to strangeness determination
- As a playground for JLAB12, assume HERMES kinematics and uncertainties for the SIDIS strangeness measurements

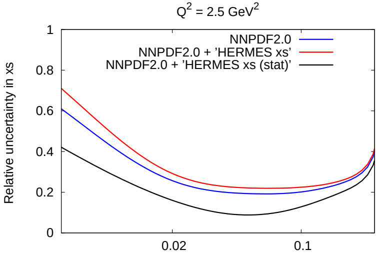




Unpolarized strangeness via SIDIS

- Include *"xs"* pseudo-data in NNPDF2.0 via reweighting
- Crude approximations (missing NLO corrections, assumes D_u^K, D_s^K), analysis should be improved for realistic estimates
- Potential for sizable reductions of unpolarized strangeness from fixed target SIDIS data. Impact on LHC physics? Non-perturbative nucleon models? Polarized strangeness?
- Redo with JLAB12 realistic pseudo-data





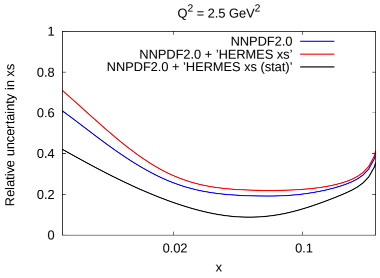
Outline

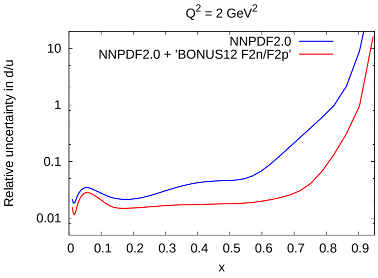
Conclusions

- The NNPDF approach provides an unbiased, statistically rigorous methodology to determine PDFs from global analysis
- Unpolarized NNPDFs fully competitive, essential ingredient of LHC physics (PDF4LHC recommendation)
- Ongoing work on polarized NNPDFs. NNPDF fragmentation functions, nuclear PDFs, Generealized PDFs, ... $\longrightarrow$ medium-long term projects
- Bayesian reweighting allows to determine impact of new data on PDFs without refitting
- JLAB12 program $\longrightarrow$ Potential to substantially improve our knowledge of unpolarized and polarized proton structure

Conclusions

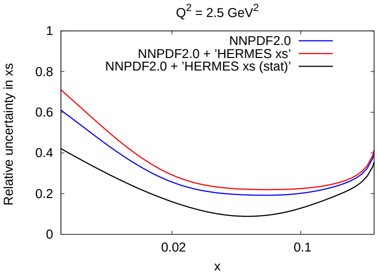
- JLAB12 program → Potential to improve substantially our knowledge of unpolarized and polarized proton structure
  This potential can be **quantified** using PDF reweighting
  (Both for unpolarized and polarized PDFs)

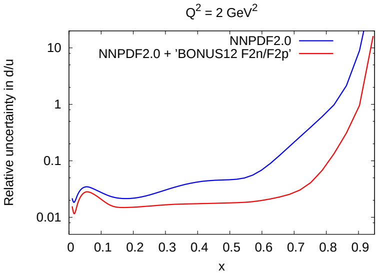




Conclusions

- JLAB12 program → Potential to improve substantially our knowledge of unpolarized and polarized proton structure
  This potential can be **quantified** using PDF reweighting



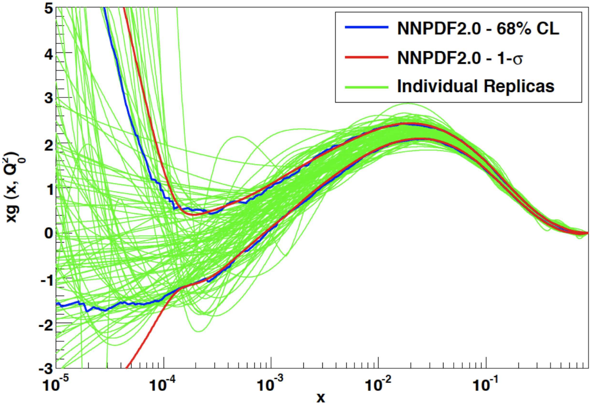


Thanks for your attention!

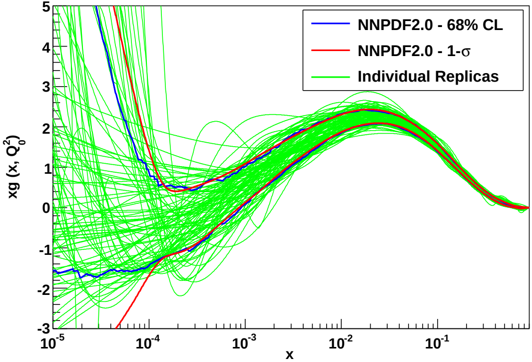
Outline

Confidence Levels

- Easy to determine arbitrary confidence levels on PDFs and physical observables
- CLs can be very different from gaussian 1–sigma errors

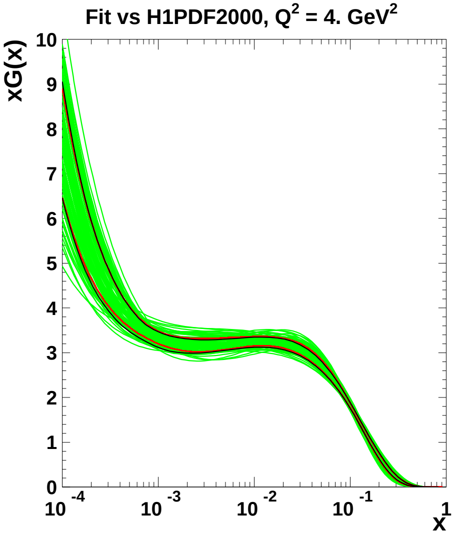


Individual replicas vs Average quantities



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 number of replicas increasing.

Monte Carlo vs. Hessian PDF uncertainties



HERA-LHC 2009 PDF benchmarks

- H1PDF2000 fit done with Hessian method and with Monte Carlo method
- The standard deviation of the 100 PDF replicas - MC method - in perfect agreement with **Hessian errors with** $\Delta\chi^2 = 1$
- The Monte Carlo method to estimate PDF uncertainties reproduces Hessian result when global χ^2 is quadratic

Genetic Algorithm

- Set neural network parameters randomly.
- Make clones of the parameter vector and mutate them.
- Evaluate the figure of merit for each clone:

Error function

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

${\rm cov}^{(t_0)}$ defined from an experimental covariance matrix which to include normalization errors with the t_0 method (arXiv:0912.2276)

$${\rm cov}_{ij}^{(t_0)} = \sigma_i^{\rm stat,2} F_i^{(\rm exp)2} + \sum_k^{N_{\rm sys}} \sigma_i^{\rm sys,k} \sigma_j^{\rm sys,k} F_i^{(\rm exp)} F_j^{(\rm exp)} + \sigma_i^{\rm N} \sigma_j^{\rm N} F_i^{(0)} F_j^{(0)},$$

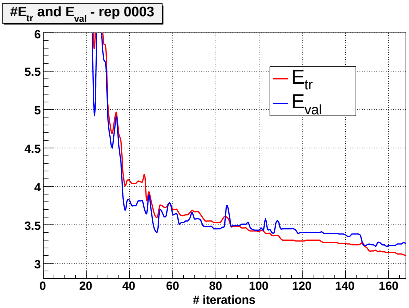
- Select the best ones and iterate the procedure until a stability is reached.

Dynamical Stopping Criterion

* Genetic Algorithms are 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 ($f_t = f_v = 0.5$).
* Minimization 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.

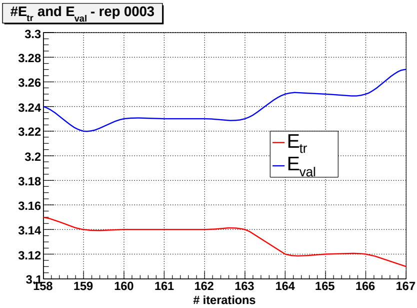


Dynamical Stopping Criterion

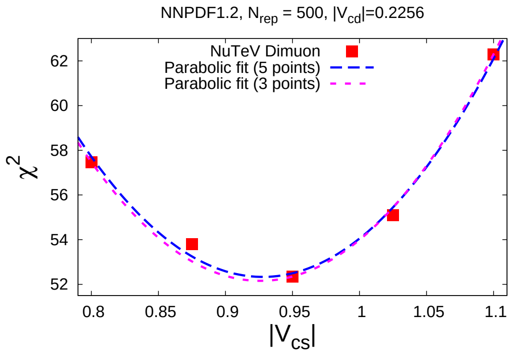
* Genetic Algorithms are monotonically decreasing by construction.
* The best fit is not given by the absolute minimum.
* The best fit is given by an optimal training beyond which the figure of merit improves only because we are fitting statistical noise of the data.

Cross-validation method

* Divide data in two sets: training and validation.
* Random division for each replica ($f_t = f_v = 0.5$).
* Minimization 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.



Direct $|V_{cs}|$ determination



CKM global fit

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

Direct determination-D and B decays

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

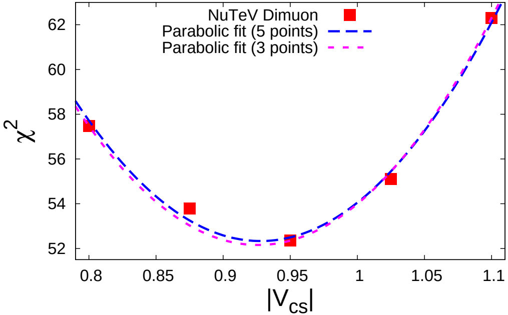
Direct det from ν−DIS (CCFR)

$$V_{cs} \geq 0.74 \quad (90\%\mathrm{CL})$$

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

Direct $|V_{cs}|$ determination

NNPDF1.2, N_{rep} = 500, $|V_{cd}|$=0.2256



Direct det from $\nu-$DIS (CCFR)

$$V_{cs} \geq 0.74 \quad (90\%\mathrm{CL})$$

Direct determination-D and B decays

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

Direct det NNPDF1.2

$$V_{cs} = 0.96 \pm 0.07, \; \Delta V_{cd} \sim 7\%$$

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

- $|V_{cs}|$ determination from neutrino DIS affected by $s^+(x)$ uncertainties
- Unbiased parametrizations for PDFs allow to discriminate variations in $s^+(x)$ from variations in CKM matrix elements

NNPDF: Input datasets

NNPDF releases available based on reduced datasets

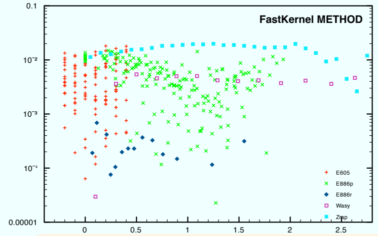
	Global	DIS	DIS+DY	DIS+JET
NNPDF	Y	Y	Y	Y
CT	Y	N	N	N
MSTW	Y	N	N	N
ABKM	N	N	Y	N
HERAPDF	N	Y	N	N

Only differ in data set, all other settings identical

Available in LHAPDF: `NNPDF20_dis_100.LHgrid`, `NNPDF20_dis+dy_100.LHgrid`, ...

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



* 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

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

Input PDF basis

Polarized PDFs are parametrized at $Q_0^2 = 1$ GeV^2 in the basis:

- Singlet $\Delta\Sigma(x) \equiv \sum_{i=1}^{n_f} (\Delta q_i(x) + \Delta\bar{q}_i(x))$,
- Triplet $\Delta T_3(x) \equiv (\Delta u(x) + \Delta\bar{u}(x)) - (\Delta d(x) + \Delta\bar{d}(x))$,
- Octet
  $\Delta T_8(x) \equiv (\Delta u(x) + \Delta\bar{u}(x)) + (\Delta d(x) + \Delta\bar{d}(x)) - 2(\Delta s(x) + \Delta\bar{s}(x))$,
- Gluon $\Delta g(x)$.

PDFs are parametrized with Artificial Neural Networks

$$
\begin{aligned}
\Delta\Sigma(x, Q_0^2) &= (1-x)^{m_{\Delta\Sigma}} x^{-n_{\Delta\Sigma}} \mathrm{NN}_{\Delta\Sigma}(x) \,, \\
\Delta T_3(x, Q_0^2) &= A_{\Delta T_3}(1-x)^{m_{\Delta T_3}} x^{-n_{\Delta T_3}} \mathrm{NN}_{\Delta T_3}(x) \,, \\
\Delta T_8(x, Q_0^2) &= A_{\Delta T_8}(1-x)^{m_{\Delta T_8}} x^{-n_{\Delta T_8}} \mathrm{NN}_{\Delta T_8}(x) \,, \\
\Delta g(x, Q_0^2) &= (1-x)^{m_{\Delta g}} x^{-n_{\Delta g}} \mathrm{NN}_{\Delta g}(x) \,.
\end{aligned}
$$

Preprocessing makes learning more efficient

$A_{\Delta T_3}, A_{\Delta T_8}$ determined from sum rules

Fitting Strategy

NNPDF strategy is very different from the standard approach (CTEQ/MSTW... polarized, DSSV/LSS/... polarized)
Instead of a set of basis functions with a small number of parameters → unbiased basis of functions parameterized by a very large and redundant set of parameters

Standard approach

$\mathcal{O}$(10-20) parm (pol)
$\mathcal{O}$(20-30) parm (unpol)

NNPDF

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