Exploring the structure of the nucleon with Neural Networks

Andrea Piccione

Genova, 12 Luglio 2005

The NNPDF Collaboration

Luigi Del Debbio[1], Stefano Forte[2],
José I. Latorre[3], Andrea Piccione[4] and Joan Rojo[3]

[1] Theory Division, CERN
[2] Dipartimento di Fisica, Università di Milano and INFN, Sezione di Milano
[3] Departament d'Estructura i Constituents de la Matèria, Universitat de Barcelona
[4] Dipartimento di Fisica Teorica, Università di Torino and INFN, Sezione di Torino

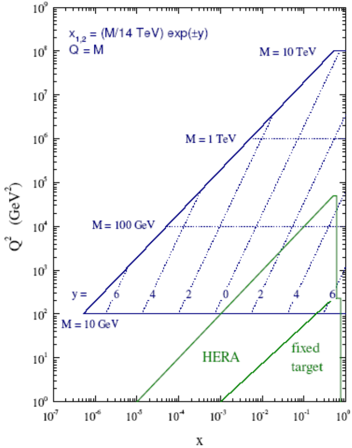
QCD and Hadrons

- QCD describes interactions between quarks and gluons. Experimentally we observe only hadrons → Confinement
- Perturbative QCD is not trustable at low energies (∼ GeV). We can not solve QCD in the non-perturbative region, but on a lattice . . .
- We can extract information on the proton structure from a process with only one initial proton (DIS at HERA). Then we can use these as an input for a process where two initial protons are involved (DY at LHC) → Factorization

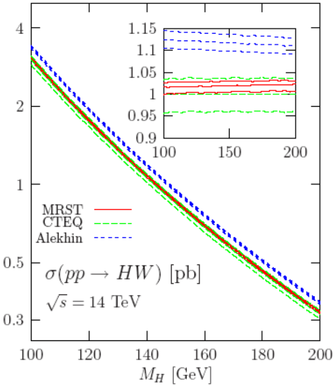
QCD and Hadrons

- QCD describes interactions between quarks and gluons.
  Experimentally we observe only hadrons $\longrightarrow$ Confinement
- Perturbative QCD is not trustable at low energies ($\sim$ GeV).
  We can not solve QCD in the non-perturbative region, but on a lattice ...
- We can extract information on the proton structure from a process with only one initial proton (DIS at HERA).
  Then we can use these as an input for a process where two initial protons are involved (DY at LHC) $\longrightarrow$ Factorization

HERA and the LHC

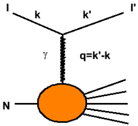


PDFs and Higgs production



[Djouadi and Ferrag 2003]

Deep Inelastic Scattering



The cross section

$$\frac{d^2\sigma}{dxdQ^2} = \frac{4\pi\alpha^2}{Q^4}\left[[1+(1-y)^2]F_1 + \frac{1-y}{x}\left(F_2 - 2xF_1\right)\right]$$

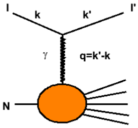
where

$$Q^2 = -(k-k')^2 \qquad \nu = p\cdot q \qquad x = \frac{Q^2}{2\nu} \qquad y = \frac{q\cdot p}{k\cdot p}$$

In the Bjorken limit ($Q^2, \nu \to \infty$ at fixed x):

$$F_i\left(x, Q^2\right) \to F_i(x)$$

Deep Inelastic Scattering



The cross section

$$\frac{d^2\sigma}{dxdQ^2} = \frac{4\pi\alpha^2}{Q^4}\left[[1+(1-y)^2]F_1 + \frac{1-y}{x}(F_2 - 2xF_1)\right]$$

where

$$Q^2 = -(k-k')^2 \qquad \nu = p\cdot q \qquad x = \frac{Q^2}{2\nu} \qquad y = \frac{q\cdot p}{k\cdot p}$$

In the Bjorken limit ($Q^2, \nu \to \infty$ at fixed x):

$$F_i\left(x, Q^2\right) \to F_i(x)$$

QCD and the parton model - I

- The structure function

$$F_2(x,\ Q^2) = x\left[\sum_{q=1}^{n_f} e_q^2\,\mathcal{C}^q \otimes q_q(x,Q^2) + 2n_f\,\mathcal{C}^g \otimes g(x,Q^2)\right]$$

where

$$(f \otimes g)(x) = \int_x^1 \frac{dy}{y} f(y) g\left(\frac{x}{y}\right)$$

- Parton distribution evlution is described by DGLAP equations

$$Q^2 \frac{d}{dQ^2} q(x,Q^2) = \frac{\alpha_s(Q^2)}{2\pi}(P \otimes q)(x,Q^2)$$

QCD and the parton model - I

- The structure function

$$F_2(x,\ Q^2) = x\left[\sum_{q=1}^{n_f} e_q^2\, \mathcal{C}^q \otimes q_q(x,Q^2) + 2n_f\, \mathcal{C}^g \otimes g(x,Q^2)\right]$$

where

$$(f \otimes g)(x) = \int_x^1 \frac{dy}{y} f(y) g\left(\frac{x}{y}\right)$$

- Parton distribution evlution is described by DGLAP equations

$$Q^2 \frac{d}{dQ^2} q(x,Q^2) = \frac{\alpha_s(Q^2)}{2\pi} (P \otimes q)(x,Q^2)$$

QCD and the parton model - II

- The singlet distribution evolves coupled with the gluon, while the non-singlet distribution evolves decoupled from the gluon:

$$\Sigma(x, Q^2) = \sum_{i=1}^{n_f} q_i(x, Q^2)$$

$$q_{NS}(x, Q^2) = \sum_{i \neq j = 1}^{n_f} (q_i(x, Q^2) - q_j(x, Q^2))$$

- For numerical implementations the evolution is solved in the Mellin space:

$$f_n = \int_0^1 dx\, x^{n-1} f(x) \qquad \rightarrow \qquad Q^2 \frac{d}{dQ^2} q_n(Q^2) = \frac{\alpha_s(Q^2)}{2\pi} P_n\, q_n(Q^2)$$

QCD and the parton model - II

- The singlet distribution evolves coupled with the gluon, while the non-singlet distribution evolves decoupled from the gluon:

$$\begin{aligned} \Sigma(x, Q^2) &= \sum_{i=1}^{n_f} q_i(x, Q^2) \\ q_{NS}(x, Q^2) &= \sum_{i \neq j=1}^{n_f} \left(q_i(x, Q^2) - q_j(x, Q^2)\right) \end{aligned}$$

- For numerical implementations the evolution is solved in the Mellin space:

$$f_n = \int_0^1 dx\, x^{n-1} f(x) \qquad \rightarrow \qquad Q^2 \frac{d}{dQ^2}\, q_n(Q^2) = \frac{\alpha_s(Q^2)}{2\pi} P_n\, q_n(Q^2)$$

The problem - I

- Structure function (or Xsec) is a convolution over x of PDFs and perturbative cross section $\rightarrow$ Deconvolution
- Each structure function (or Xsec) is a linear combination of many PDFs ($2n_f$ quarks + gluon) $\rightarrow$ Different processes
- Data are given at various scales, and we want PDFs as functions of x at a common scale Q^2 $\rightarrow$ Evolution
- TH uncertainties: resummation, nuclear corrections, higher twist, heavy quark thresholds, ...

Which is the uncertainty associated with a PDFs set?
[Frixione and Mangano 2004, Tung 2004, HERA and the LHC Workshop 2004-2005]

The problem - I

- Structure function (or Xsec) is a convolution over x of PDFs and perturbative cross section $\rightarrow$ Deconvolution
- Each structure function (or Xsec) is a linear combination of many PDFs ($2n_f$ quarks + gluon) $\rightarrow$ Different processes
- Data are given at various scales, and we want PDFs as functions of x at a common scale Q^2 $\rightarrow$ Evolution
- TH uncertainties: resummation, nuclear corrections, higher twist, heavy quark thresholds, ...

Which is the uncertainty associated with a PDFs set?
[Frixione and Mangano 2004, Tung 2004,
HERA and the LHC Workshop 2004-2005]

The problem - II

- For a single quantity → 1 sigma error
- For a pair of numbers → 1 sigma ellipse
- For a function → We need an "error band" in the space of functions (*i.e.* the probability density $\mathcal{P}[f]$ in the space of functions $f(x)$)

Expectation values → Functional integrals

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

Determine an infinite-dimensional object (a function) from finite set of data points → Mathematically ill-posed problem

The problem - II

- For a single quantity $\rightarrow$ 1 sigma error
- For a pair of numbers $\rightarrow$ 1 sigma ellipse
- For a function $\rightarrow$ We need an "error band" in the space of functions (*i.e.* the probability density $\mathcal{P}[f]$ in the space of functions $f(x)$)

Expectation values $\rightarrow$ Functional integrals

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

Determine an infinite-dimensional object (a function) from finite set of data points $\rightarrow$ Mathematically ill-posed problem

The standard approach

1. Choose a simple functional form with enough free parameters

$$q(x, Q_0^2) = x^{\alpha}(1-x)^{\beta} P(x; \lambda_1, \ldots, \lambda_n)$$

2. Fit parameters by minimizing χ^2

Some difficulties arise:

- Errors and correlations of parameters require at least fully correlated analysis of data errors
- Error propagation to observables is difficult: many observables are nonlinear/nonlocal functional of parameters
- Theoretical bias due to choice of parametrization is difficult to assess (effects can be large if data are not precise or hardly compatible)

The standard approach

1. Choose a simple functional form with enough free parameters

$$q(x, Q_0^2) = x^{\alpha}(1-x)^{\beta}P(x; \lambda_1, \ldots, \lambda_n)$$

2. Fit parameters by minimizing χ^2

Some difficulties arise:

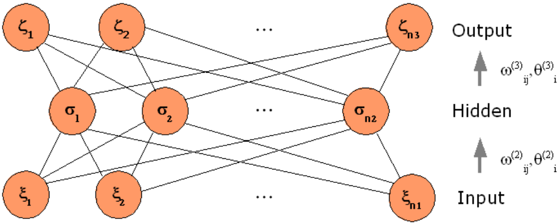
- Errors and correlations of parameters require at least fully correlated analysis of data errors
- Error propagation to observables is difficult: many observables are nonlinear/nonlocal functional of parameters
- Theoretical bias due to choice of parametrization is difficult to assess (effects can be large if data are not precise or hardly compatible)

The NNPDF approach

- Determination of the Structure Functions:
  this is the easiest case, since no evolution is required, but only data fitting. A good application to test the technique → Done
- Determination of the Parton Distributions:
  the real stuff → Working on it ...

Structure - I

Neural networks: a class of algorithms providing robust, universal, unbiased approximants to incomplete or noisy data



Structure - II

- Building blocks: neurons, *i. e.* input/output units characterized by sigmoid activation

$$\xi_i^{(l)} = g\left(\sum_{j=1}^{n_l-1} \omega_{ij}^{(l-1)} \xi_j^{(l-1)} - \theta_i^{(l)}\right) \qquad g(x) = \frac{1}{1+e^{-x}}$$

- Parameters: weights $\omega_{ij}^{(l)}$ and thresholds $\theta_i^{(l)}$.
- Architecture: multilayer feed-forward NN. Each neuron receives input from neurons in preceding layer and feeds output to neurons in successive layer
- Assumption: smooth function (step functions are not allowed)

Training - Back Propagation

1. Set the parameters randomly.
2. Present an input and calculate the output.
3. Evaluate χ^2.
4. Modify the weights to reinforce correct decisions and discourage incorrect ones:

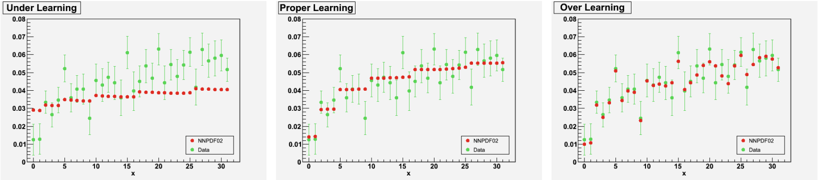
$$\omega_{ij} \rightarrow \omega_{ij} - \eta \frac{\partial \chi^2}{\partial \omega_{ij}}$$

   where η is the learning rate.
5. Back to 2, till the stability of χ^2 is reached

Training - Genetic Algorithm

1. Set the parameters randomly.
2. Make clones of the set of parameters.
3. Mutate each clone.
4. Evaluate χ^2 for all the clones.
5. Select the clone that has the lowest χ^2.
6. Back to 2, till the stability of χ^2 is reached.

Learning *vs.* overlearning



General strategy - I

- Monte Carlo sampling of data (generation of replicas of experimental data) $\rightarrow$ Faithful representation of uncertainties

$$F_i^{(art)(k)} = \left(1 + r_N^{(k)}\sigma_N\right)\left[F_i^{(exp)} + r_i^s \sigma_i^{stat} + \sum_{l=1}^{N_{sys}} r^{l,(k)} \sigma_i^{sys,l}\right]$$

- NN training over MC replicas $\rightarrow$ Unbiased parametrization

Expectation values $\rightarrow$ Sum over the Nets

$$\langle \mathcal{F}\left[F(x,Q^2)\right]\rangle = \frac{1}{N_{rep}} \sum_{k=1}^{N_{rep}} \mathcal{F}\left(F^{(net)(k)}(x,Q^2)\right)$$

General strategy - I

- Monte Carlo sampling of data (generation of replicas of experimental data) → Faithful representation of uncertainties

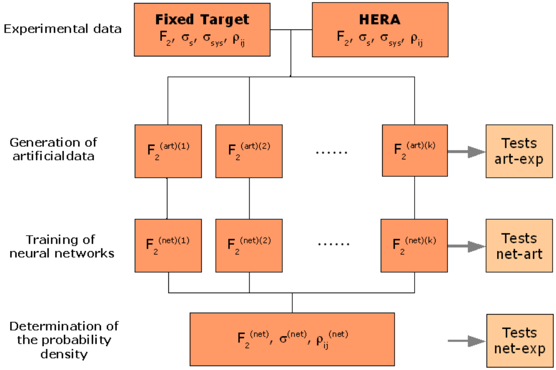
$$F_i^{(art)(k)} = \left(1 + r_N^{(k)}\sigma_N\right)\left[F_i^{(exp)} + r_i^s\sigma_i^{stat} + \sum_{l=1}^{N_{sys}} r^{l,(k)}\sigma_i^{sys,l}\right]$$

- NN training over MC replicas → Unbiased parametrization

Expectation values → Sum over the Nets

$$\left\langle \mathcal{F}\left[F(x,Q^2)\right]\right\rangle = \frac{1}{N_{rep}}\sum_{k=1}^{N_{rep}} \mathcal{F}\left(F^{(net)(k)}(x,Q^2)\right)$$

General strategy - II



Training - I

Architecture: 4-5-3-1

- Inputs: x, $\log x$, Q^2, $\log Q^2$
- Output: $F_2(x, Q^2)$

Minimization strategy:

- Back Propagation ($\sim 10^8$ training cycles):

$$\chi^{2\,(k)}_{\rm diag} = \frac{1}{N_{\rm dat}} \sum_{i=1}^{N_{\rm dat}} \frac{\left(F_i^{(\rm art)(k)} - F_i^{(\rm net)(k)}\right)^2}{\sigma_{i,t}^{(\rm exp)2}}$$

- Genetic Algorithm ($\sim 10^4$ generations):

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

Training - I

Architecture: 4-5-3-1

- Inputs: x, $\log x$, Q^2, $\log Q^2$
- Output: $F_2(x, Q^2)$

Minimization strategy:

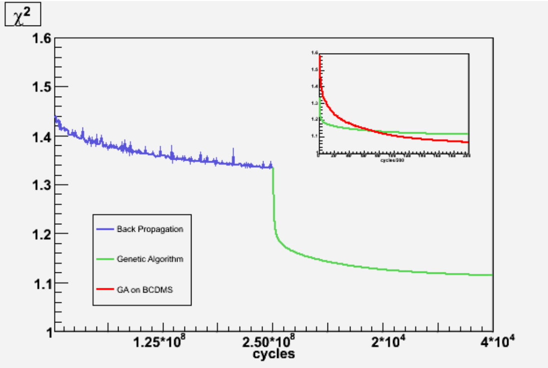
- Back Propagation ($\sim 10^8$ training cycles):

$$\chi^{2\,(k)}_{\mathrm{diag}} = \frac{1}{N_{\mathrm{dat}}} \sum_{i=1}^{N_{\mathrm{dat}}} \frac{\left(F_i^{(\mathrm{art})(k)} - F_i^{(\mathrm{net})(k)}\right)^2}{\sigma_{i,t}^{(\mathrm{exp})2}}$$

- Genetic Algorithm ($\sim 10^4$ generations):

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

Training - II


χ^2
1.6
1.5
1.4
1.3
1.2
1.1
1
1.25*10^8
2.50*10^8
2*10^4
4*10^4
cycles
Back Propagation
Genetic Algorithm
GA on BCDMS

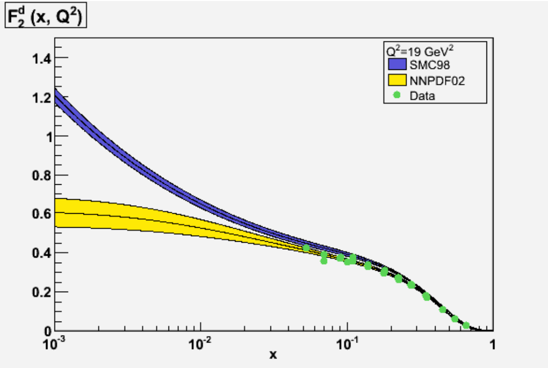
Credits

- S. Forte, L. Garrido, J. I. Latorre and A. P., *"Neural network parametrization of deep-inelastic structure functions,"* JHEP05 (2002) 062 [arXiv:hep-ph/0204232]
- L. Del Debbio, S. Forte, J. I. Latorre, A. P. and J. Rojo [NNPDF Collaboration], *"Unbiased determination of the proton structure function F_2^p with faithful uncertainty estimation"*, JHEP03 (2005) 080 [arXiv:hep-ph/0501067]

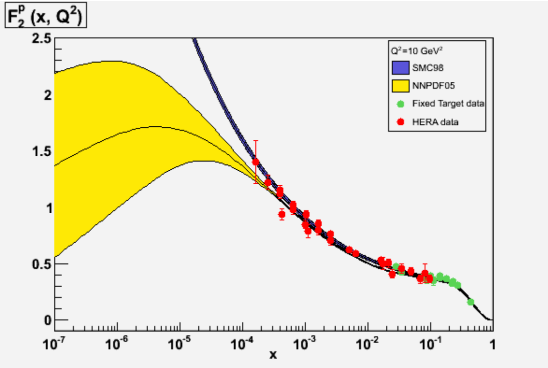
Source code, driver program and graphical web interface for F_2 plots and numerical computations available @

http://sophia.ecm.ub.es/f2neural

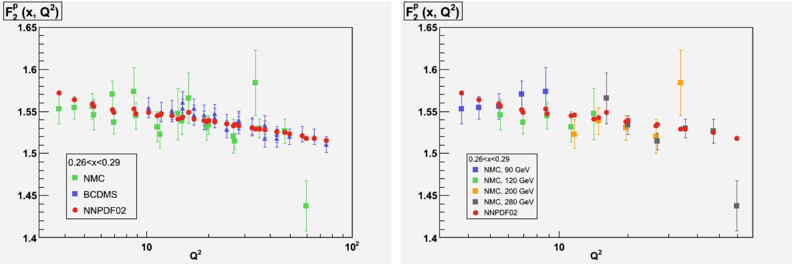
Fit of $F_2^d(x, Q^2)$ *[NNPDF 2002]*



Fit of $F_2^p(x, Q^2)$ *[NNPDF 2005]*



Incompatible data *[NNPDF 2002]*



Strategy

Same strategy as with SF + Altarelli-Parisi evolution

- Monte Carlo sampling of data
- Parametrize parton distributions with neural networks
- Evolution of parton distributions to experimental data scale and training over Monte Carlo replica sample

Examples

- Expectation values:

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

- Errors:

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

- Correlations between pairs of different parton distributions at different points:

$$\langle u(x_1)d(x_2)\rangle = \frac{1}{N_{rep}} \sum_{k=1}^{N_{rep}} u^{(net)(k)}(x_1, Q_0^2)\, d^{(net)(k)}(x_2, Q_0^2)$$

Evolution kernel

- We want Mellin space evolution:

$$q(N, Q^2) = q(N, Q_0^2)\Gamma\left(N, \alpha_s\left(Q^2\right), \alpha_s\left(Q_0^2\right)\right)$$

- We do not want complex neural networks:

$$\Gamma\left(x, \alpha_s\left(Q^2\right), \alpha_s\left(Q_0^2\right)\right) \equiv \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} dN\, x^{-N}\Gamma\left(N, \alpha_s\left(Q^2\right), \alpha_s\left(Q_0^2\right)\right)$$

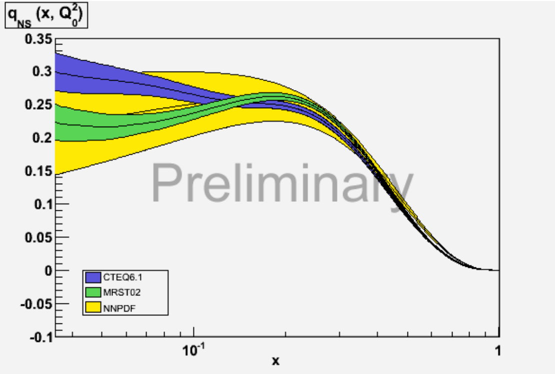
- $\Gamma(x)$ is a distribution $\rightarrow$ must be regularized at $x = 1$:

$$q(x, Q^2) = q(x, Q_0^2)\int_x^1 dy\, \Gamma(y) + \int_x^1 \frac{dy}{y}\Gamma(y)\left(q\left(\frac{x}{y}, Q_0^2\right) - yq(x, Q_0^2)\right)$$

Details

- $q_{NS}(x, Q^2) \equiv \frac{1}{6}\left(u + \bar{u} - d - \bar{d}\right)(x, Q^2)$
- Experimental data: NMC (94 pts) and BCDMS (253 pts)
- Kinematical cuts: $Q^2 \geq 9\ GeV^2$, $W^2 \geq 6.25\ GeV^2$
- Neural network architecture: 2-2-2-1 (15 params.)
- Strong coupling: $\alpha_s\left(M_Z^2\right) = 0.1182$
- Perturbative order: NNLO
- VFN: $m_c = 1.5 GeV$, $m_b = 4.5 GeV$, $m_t = 175 GeV$
- TMC: F_2 integral evaluated with NN F_2
- # replica: 25 (should be 1000)

Non-Singlet


$q_{NS}(x, Q_0^2)$
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
-0.05
-0.1
10^{-1}
1
x
Preliminary
CTEQ6.1
MRST02
NNPDF

Summary

- Unbiased determination of structure functions with faithful estimation of uncertainties
- Successful implementation of neural parton fitting at NNLO

Outlook

- Construct full set of NNPDF parton distributions from all available data
- Estimate impact of theoretical uncertainties
- Assess impact of uncertainties of PDFs for relevant observables at LHC
- Perform a benchmark set of pdfs, to compare the different fitting programs (CTEQ,MRST, Alekhin)
- Make formalism compatible with standard interfaces (LHAPDF, PDFLIB) $\rightarrow$ NNPDF partons available for use in Monte Carlo generators