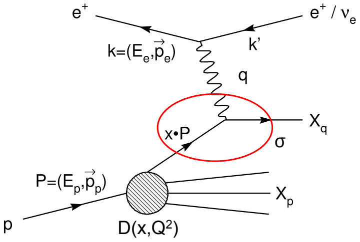
Parton Distribution Functions and Neural Networks

Alberto Guffanti
Albert-Ludwigs-Universität Freiburg

PSI Villigen,
March 11, 2010

What are Parton Distribution Functions?

- Consider a process with one hadron in the initial state



- According to the Factorization Theorem we can write the cross section as

$$d\sigma = \sum_a \int_0^1 \frac{d\xi}{\xi} D_a(\xi, \mu^2) d\hat{\sigma}_a \left(\frac{x}{\xi}, \frac{\hat{s}}{\mu^2}, \alpha_s(\mu^2) \right) + \mathcal{O}\left(\frac{1}{Q^p} \right)$$

What are Parton Distribution Functions?

- The initial condition cannot be computed in Perturbation Theory (Lattice? In principle yes, but ...)
- ... but the energy scale dependence is governed by DGLAP evolution equations

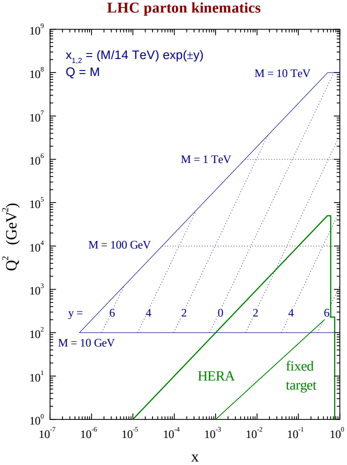
$$\frac{\partial}{\ln Q^2} q^{NS}(\xi, Q^2) = P^{NS}(\xi, \alpha_s) \otimes q^{NS}(\xi, Q^2)$$

$$\frac{\partial}{\ln Q^2} \begin{pmatrix} \Sigma \\ g \end{pmatrix} (\xi, Q^2) = \begin{pmatrix} P_{qq} & P_{qg} \\ P_{gq} & P_{gg} \end{pmatrix} (\xi, \alpha_s) \otimes \begin{pmatrix} \Sigma \\ g \end{pmatrix} (\xi, Q^2)$$

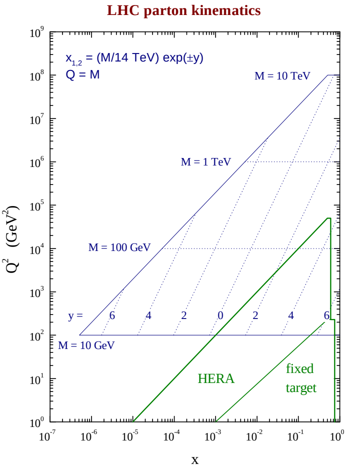
- ... and the **splitting functions** P can be computed in PT and are known up to **NNLO**

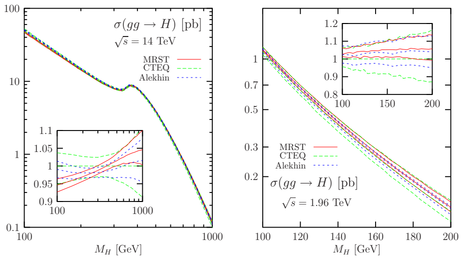
(**LO** - Dokshitzer; Gribov, Lipatov; Altarelli, Parisi; 1977)
(**NLO** - Floratos, Ross, Sachrajda; Gonzalez-Arroyo, Lopez, Yndurain; Curci, Furmanski, Petronzio, 1981)
(**NNLO** - Moch, Vermaseren, Vogt; 2004)

Why care about PDFs (and their uncertainties)?



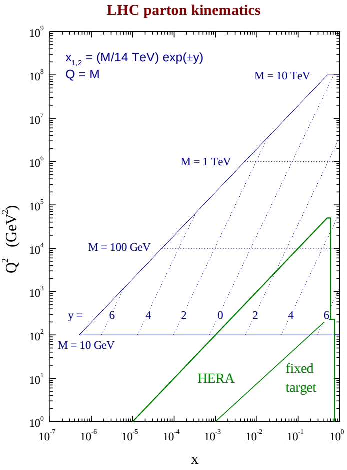
Why care about PDFs (and their uncertainties)?

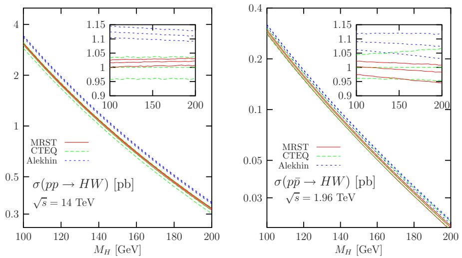




[A. Djouadi and S. Ferrag, hep-ph/0310209]

Why care about PDFs (and their uncertainties)?





[A. Djouadi and S. Ferrag, hep-ph/0310209]

Why care about PDFs (and their uncertainties)?

- Errors on PDFs are in some cases the dominating theoretical error on precision observables

Ex. $\sigma(Z^0)$ at the LHC: $\delta_{PDF} \sim 3\%$, $\delta_{NNLO} \sim 2\%$

[J. Campbell, J. Huston and J. Stirling, (2007)]

Why care about PDFs (and their uncertainties)?

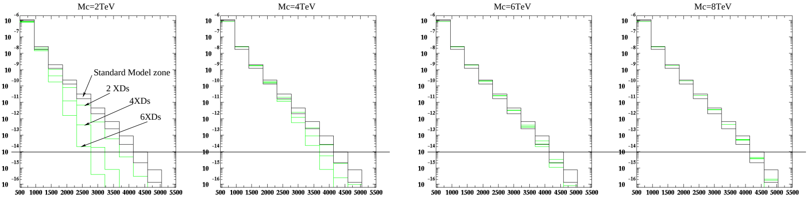
- Errors on PDFs are in some cases the dominating theoretical error on precision observables

Ex. $\sigma(Z^0)$ at the LHC: $\delta_{PDF} \sim 3\%$, $\delta_{NNLO} \sim 2\%$

[J. Campbell, J. Huston and J. Stirling, (2007)]

- Errors on PDFs might reduce sensitivity to New Physics

Ex. Extra Dimensions discovery in dijet cross section at the LHC:



[S. Ferrag (ATLAS), hep-ph/0407303]

Problem

Faithful estimation of errors on PDFs

- Single quantity: 1-σ error
- Multiple quantities: 1-σ contours
- Function: 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 are Functional integrals

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

Problem

Faithful estimation of errors on PDFs

- Single quantity: 1-σ error
- Multiple quantities: 1-σ contours
- Function: 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 are 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 a finite set of data points ... mathematically ill-defined problem.

Solution

Standard Approach

- Introduce a simple functional form with enough free parameters

$$q(x, Q^2) = x^\alpha (1 - x)^\beta P(x; \lambda_1, ..., \lambda_n).$$

- Fit parameters minimizing χ^2.

Solution

Standard Approach

- Introduce a simple functional form with enough free parameters

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

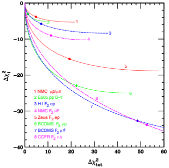
- Fit parameters minimizing χ^2.

Open problems:

- **Error propagation** from data to parameters and from parameters to observables is **not trivial**.
- **Theoretical bias** due to the chosen **parametrization** is difficult to assess.

- Standard $\Delta\chi^2 = 1$ criterion is **too restrictive** to account for large discrepancies among experiments.

[Collins & Pumplin, 2001]

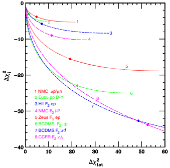


Shortcomings of the Standard approach

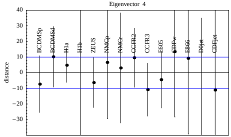
What is the meaning of a one-σ uncertainty?

- Standard $\Delta\chi^2 = 1$ criterion is **too restrictive** to account for large discrepancies among experiments.

[Collins & Pumplin, 2001]



- Introduce a **TOLERANCE** criterion, i.e. take the envelope of uncertainties of experiments to determine the $\Delta\chi^2$ to use for the global fit (CTEQ).

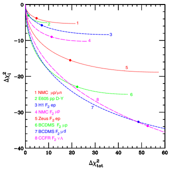


Shortcomings of the Standard approach

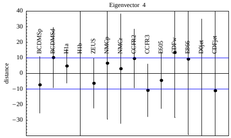
What is the meaning of a one-σ uncertainty?

- Standard $\Delta\chi^2 = 1$ criterion is **too restrictive** to account for large discrepancies among experiments.

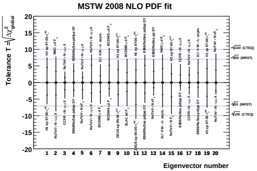
  [Collins & Pumplin, 2001]



- Introduce a **TOLERANCE** criterion, i.e. take the envelope of uncertainties of experiments to determine the $\Delta\chi^2$ to use for the global fit (CTEQ).

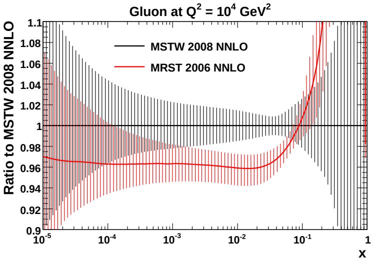


- Make it **DYNAMICAL**, i.e. determine $\Delta\chi^2$ separately for each hessian eigenvector (MSTW).



- Uncertainties in standard fits often increase when adding new data to the fit.
- Related to the need of extending the parametriztion in order to accomodate the new data

Smaller high-x gluon (and slightly smaller α_S) results in larger small-x gluon – now shown at NNLO.



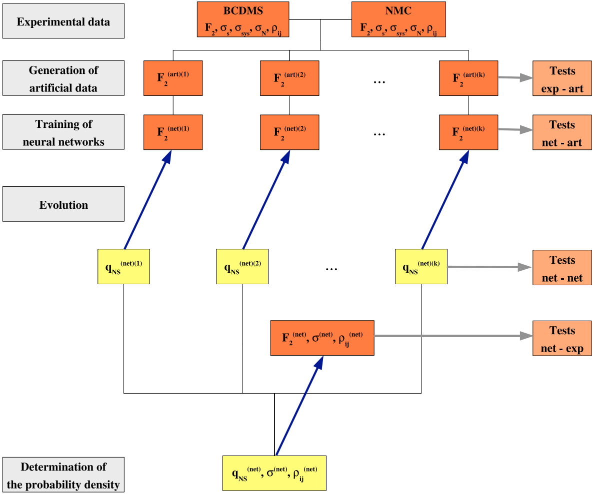
Larger small-x uncertainty due to extrat free parameter.

[R. Thorne, PDF4LHC]

THE NNPDF METHODOLOGY

[R. D. Ball, L. Del Debbio, S. Forte, J. I. Latorre, A. Piccione, J. Rojo, M. Ubiali and AG]

The NNPDF methodology



The Neural Network Approach in a Nutshell

- Generate N_{rep} **Monte-Carlo replicas** of the experimental data.
- Fit a set of Parton Distribution Functions on each replica, thus defining a sampling of probability density on the space of the PDFs.
- **Expectation values** for observables are **Monte Carlo integrals**

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

... the same is true for errors, correlations, etc.

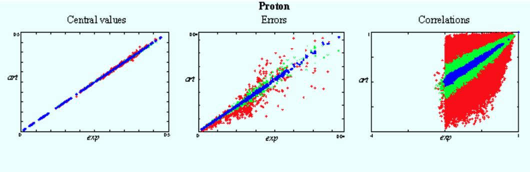
Monte Carlo replicas generation

- Generate artificial data according to distribution

$$O_i^{(art)\,(k)} = (1 + r_N^{(k)}\,\sigma_N)\left[O_i^{(exp)} + \sum_{p=1}^{N_{sys}} r_p^{(k)}\,\sigma_{i,p} + r_{i,s}^{(k)}\,\sigma_s^i\right]$$

where r_i are univariate gaussian random numbers

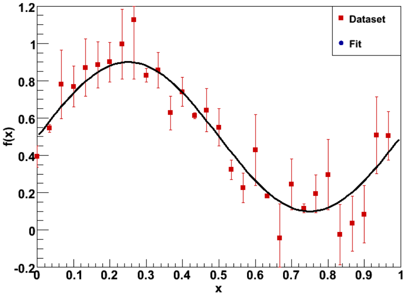
- Validate Monte Carlo replicas against experimental data (statistical estimators, faithful representation of errors, convergence rate increasing N_{rep})



- $\mathcal{O}(1000)$ replicas needed to reproduce correlations to percent accuracy

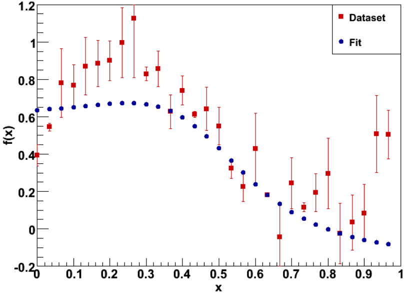
Proper Fitting avoiding Overlearning

- Let's see how proper fitting works in a toy model



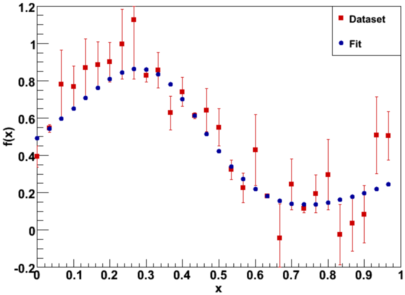
Proper Fitting avoiding Overlearning

- Let's see how proper fitting works in a toy model



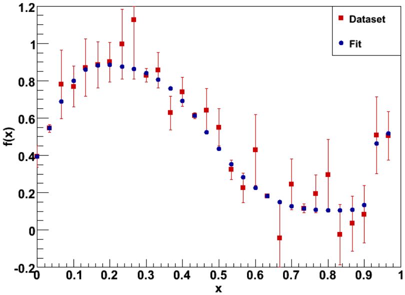
Proper Fitting avoiding Overlearning

- Let's see how proper fitting works in a toy model



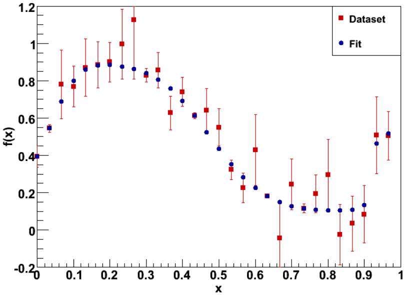
Proper Fitting avoiding Overlearning

- Let's see how proper fitting works in a toy model



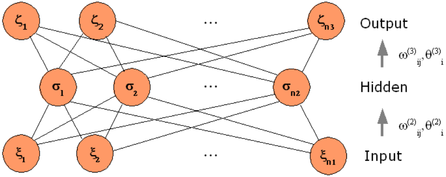
Proper Fitting avoiding Overlearning

- Let's see how proper fitting works in a toy model



- Need a **redundant parametrization** to avoid parametrization bias.
- Need a way of **stopping the fit before overlearning** sets in to avoid fitting statistical noise.

Why use Neural Networks?



- Neural Networks are non-linear statistical tools.
- Any continuous function can be approximated with neural network with one internal layer and non-linear neuron activation function.
- Efficient minimization algorithms for complex parameter spaces.
- They provide a parametrization which is redundant and robust against variations.

Neural Networks

... just another basis of functions

Multilayer feed-forward networks

- Each neuron receives input from neurons in preceding layer and feeds output to neurons in subsequent layer
- Activation determined by weights and thresholds

$$\xi_i = g\left(\sum_j \omega_{ij}\xi_j - \theta_i\right)$$

- Sigmoid activation function

$$g(x) = \frac{1}{1 + e^{-\beta x}}$$

Neural Networks

... just another basis of functions

Multilayer feed-forward networks

- Each neuron receives input from neurons in preceding layer and feeds output to neurons in subsequent layer
- Activation determined by weights and thresholds

$$\xi_i = g\left(\sum_j \omega_{ij}\xi_j - \theta_i\right)$$

- Sigmoid activation function

$$g(x) = \frac{1}{1+e^{-\beta x}}$$

A 1-2-1 NN:

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

Genetic Algorithm

1. Set network 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, until stability in χ^2 is reached.

Genetic Algorithm

1. Set network 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, until stability in χ^2 is reached.

Pros:

- Allows to minimize the fully correlated χ^2
- Explores the full parameter space, reducing the risk of being trapped in a local minimum

Cons:

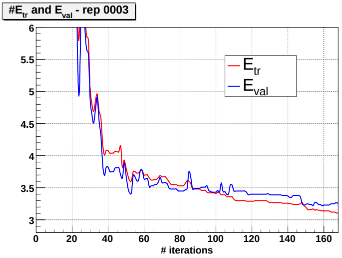
- Slow convergence
- χ^2 decreases monotonically - need to find a suitable stopping criterion

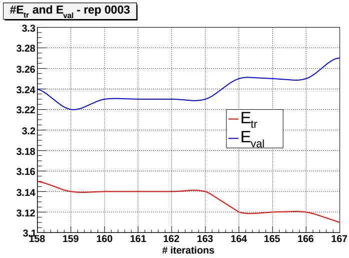
Stopping criterion based on Training-Validation separation

- Divide the data in two sets: Training and Validation
- Minimize the χ^2 of the data in the Training set
- Compute the χ^2 for the data in the Validation set
- When validation χ^2 stops decreasing, **STOP** the fit

Stopping criterion based on Training-Validation separation

- Divide the data in two sets: Training and Validation
- Minimize the χ^2 of the data in the Training set
- Compute the χ^2 for the data in the Validation set
- When validation χ^2 stops decreasing, **STOP** the fit





RESULTS

The Past

NNPDF1.0/1.2

- **NNPDF 1.0**

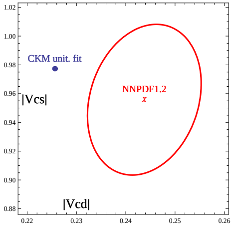
  [R. D. Ball et al., arXiv:0808.1231]

  - Global DIS fit
  - First application of the full NNPDF Methodology (multiple exps., multiple PDFs)

- **NNPDF 1.2**

  [R. D. Ball et al., arXiv:0906.1958]

  - Constraining strangeness (dimuon data)
  - Extraction of physical parameters (CKM matrix elements)



- Result for the combined fit

$$
\begin{aligned}
|V_{cs}| &= 0.96 \pm 0.07 \\
|V_{cd}| &= 0.244 \pm 0.019 \\
\rho[V_{cs}, V_{cd}] &= 0.21
\end{aligned}
$$

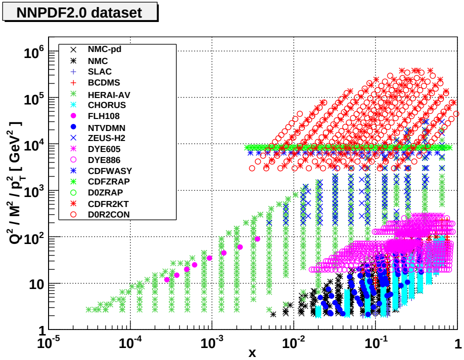
NNPDF 2.0

Technical improvements

- Fast DGLAP evolution based on higher-order interpolating polynomials
- Improved treatment of normalization errors (t_0 method)
  - For details see [R. D. Ball et al., arXiv:0912.2276]
- Improvements in training/stopping
  - Target Weighted Training
  - Improved stopping for avoiding under-/over-learning
- For all the deatils see: [R. D. Ball et al., arXiv:1002.4407]

NNPDF2.0

Dataset



- 3477 data points

  (for comparison MSTW08 includes 2699 data points)

OBS	*Data set*
Deep Inelastic Scattering	
F_2^d / F_2^p	NMC-pd
F_2^p	NMC SLAC BCDMS
F_2^d	SLAC BCDMS
σ_{NC}^{+}	ZEUS H1
σ_{NC}^{-}	ZEUS H1
F_L	H1
$\sigma_\nu , \sigma_{\bar\nu}$	CHORUS
dimuon prod.	NuTeV
Drell-Yan & Vector Boson prod.	
$d\sigma^{\rm DY}/dM^2dy$	E605
$d\sigma^{\rm DY}/dM^2dx_F$	E866
W asymm.	CDF
Z rap. distr.	D0/CDF
Inclusive jet prod.	
Incl. $\sigma^{\rm (jet)}$	CDF (k_T) - Run II
Incl. $\sigma^{\rm (jet)}$	D0 (cone) - Run II

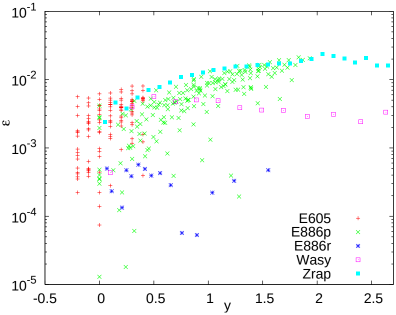
NNPDF2.0

Proper inclusion of NLO corrections

- Inclusion of higher order corrections to hadronic processes in parton fits is often too expensive
- Often higher order corrections are included as (local) K factors rescaling the LO cross section
- We use FastNLO for inclusive jet cross section

  [T. Kluge et al.,hep-ph/0609285]
- We developed our own FastDY for fixed target Drell-Yan and vector boson production at colliders

Relative Accuracy w.r.t to Exact calculation



NNPDF2.0

Parametrization

- We parametrize 7 PDF combinations at the initial scale with Neural Networks

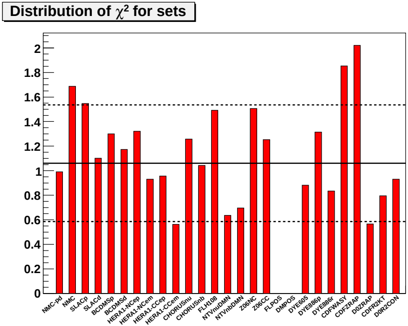
Parton Distributions Combination		**NN architechture**
Singlet ($\Sigma(x)$)	$\Longrightarrow$	2-5-3-1 (37 pars)
Gluon ($g(x)$)	$\Longrightarrow$	2-5-3-1 (37 pars)
Total valence ($V(x) \equiv u_V(x) + d_V(x)$)	$\Longrightarrow$	2-5-3-1 (37 pars)
Non-singlet triplet ($T_3(x)$)	$\Longrightarrow$	2-5-3-1 (37 pars)
Sea asymmetry ($\Delta_S(x) \equiv \bar{d}(x) - \bar{u}(x)$)	$\Longrightarrow$	2-5-3-1 (37 pars)
Total Strangeness ($s^+(x) \equiv (s(x) + \bar{s}(x))/2$)	$\Longrightarrow$	2-5-3-1 (37 pars)
Strange valence ($s^-(x) \equiv (s(x) - \bar{s}(x))/2$)	$\Longrightarrow$	2-5-3-1 (37 pars)

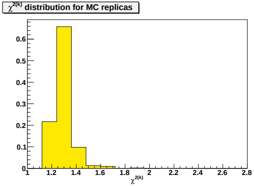
259 parameters

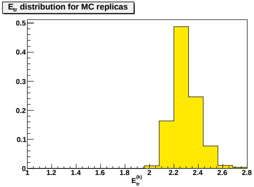
NNPDF2.0

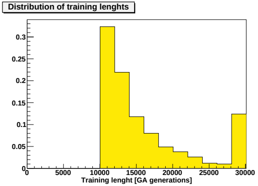
Results - General features of the fit

$\chi^2_{\rm tot}$	1.21
$\langle E \rangle \pm \sigma_E$	2.32 ± 0.10
$\langle E_{\rm tr} \rangle \pm \sigma_{E_{\rm tr}}$	2.29 ± 0.11
$\langle E_{\rm val} \rangle \pm \sigma_{E_{\rm val}}$	2.35 ± 0.12
$\langle {\rm TL} \rangle \pm \sigma_{\rm TL}$	16175 ± 6257
$\langle \chi^{2(k)} \rangle \pm \sigma_{\chi^2}$	1.29 ± 0.09



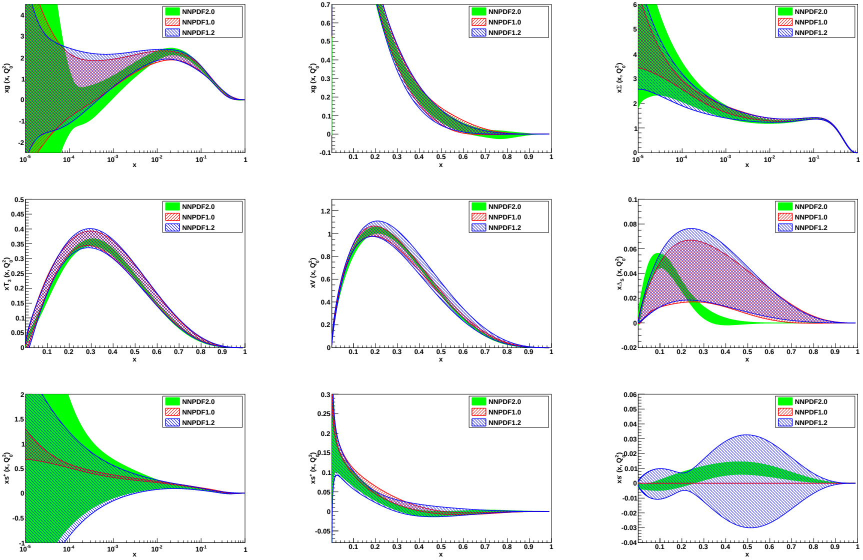






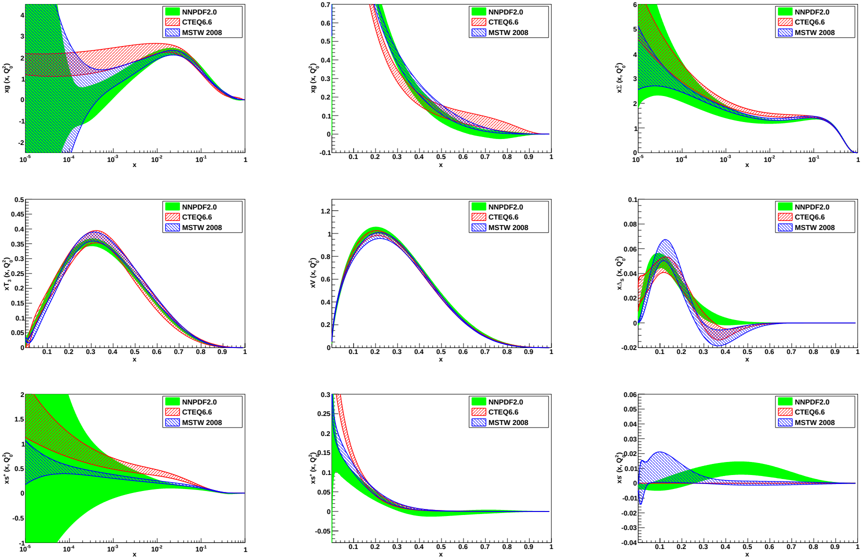
NNPDF2.0

Results - Partons - Comparison to older NNPDF set



NNPDF2.0

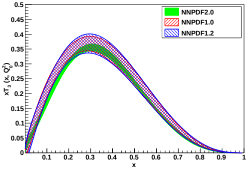
Results - Partons - Comparison to other global fits



NNPDF2.0

Results - Partons - A couple of upshots

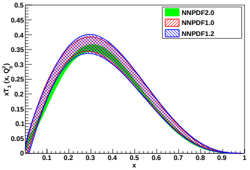
- **Reduction of uncertainties** with respect to older NNPDF sets due to **inclusion of new data**



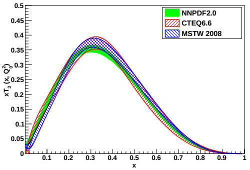
NNPDF2.0

Results - Partons - A couple of upshots

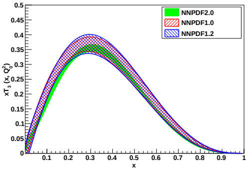
- **Reduction of uncertainties** with respect to older NNPDF sets due to **inclusion of new data**



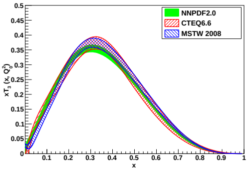
- **Uncertainties** on PDFs **competitive** with results from other groups ...



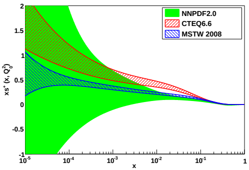
- **Reduction of uncertainties** with respect to older NNPDF sets due to **inclusion of new data**



- **Uncertainties** on PDFs **competitive** with results from other groups ...



- ... but still retain **unbiasedness** in regions where there are little or no experimental constraints



NNPDF2.0

Results - Quantitative assesment of impact of modifications

- We define the **distance** between central values of PDFs

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

and the similarly for Standard Deviations.

NNPDF2.0

Results - Quantitative assesment of impact of modifications

- We define the **distance** between central values of PDFs

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

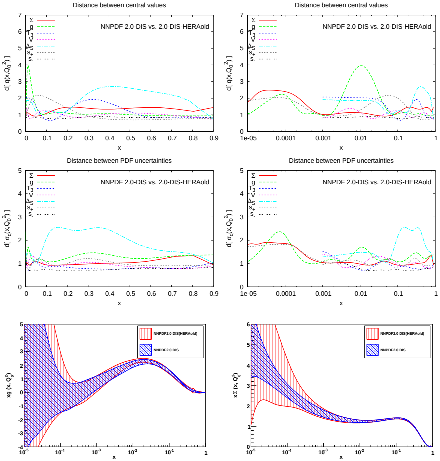
  and the similarly for Standard Deviations.

- Comparisons we have performed for NNPDF2.0
  - NNPDF1.2 vs. NNPDF1.2 + minimization/training improvements
  - Improved NNPDF1.2 vs. Improved NNPDF1.2 + t_0-method
  - Fit to DIS dataset with H1/ZEUS data vs. Fit with HERA-I combined
  - Fit to DIS dataset vs. Fit to DIS+JET
  - Fit to DIS+JET vs. NNPDF2.0 final

Results

Impact HERA-I combined dataset

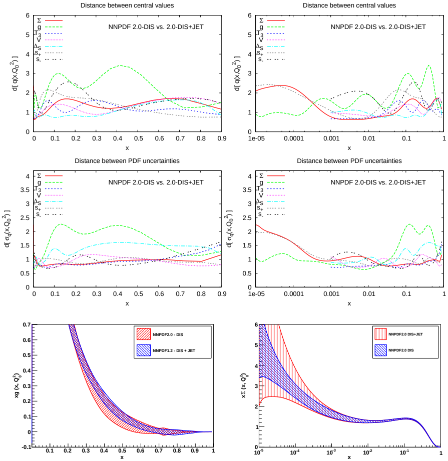
- Overall fit quality to the whole dataset is good ($\chi^2 = 1.14$)
  - $\sigma^{+}_{\rm NC}$ dataset has relatively high $\chi^2 \sim 1.3$
  - $\sigma^{-}_{\rm CC}$ dataset has very low $\chi^2 \sim 0.55$
- Same χ^2-pattern observed in the HERAPDF1.0 analysis
- Impact on PDFs is moderate, affecting mainly Singlet and Gluon at small-x



Results

Impact of Tevatron inclusive Jet data

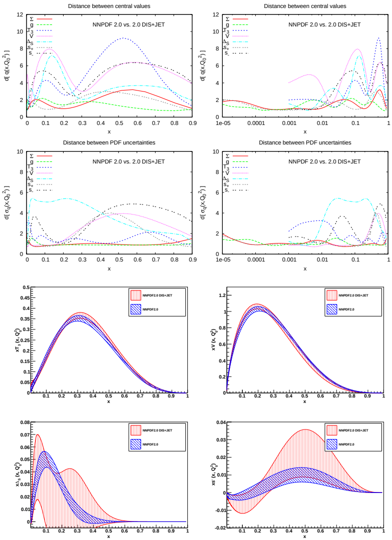
- We include Tevatron Run-II inclusive jet data
- They provide a valuable constrain on large-x gluon
- No signs of tension with other datasets included in the analysis
- Run-I data not included but compatibility with the outcome of the fit has been checked



Results

Impact of Drell-Yan and Vector Boson production data

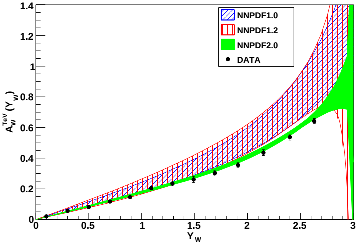
- Good description of fixed target Drell-Yan data (E605 proton and E886 proton and p/d ratio)
- Vector boson production at colliders (CDF W-asymmetry and Z rapidity distribution) harder to fit
- All valence-type PDF combinations are affected by these data
- Sizable reduction in the uncertainty of the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



Results

Vector Boson production at colliders

- Z rapidity distribution:
  - Very good description of D0 data ($\chi^2 = 0.57$)
  - Poor description of CDF data ($\chi^2 = 2.02$)
  - MSTW08 has the same pattern
  - Possible inconsistency of the two datasets?
- CDF W-asymmetry
  - We fit the direct W-asymmetry data, not the leptoinc asymmetry
  - Poor description of the data ($\chi^2 = 1.85$)



Results

Phenomenological implications

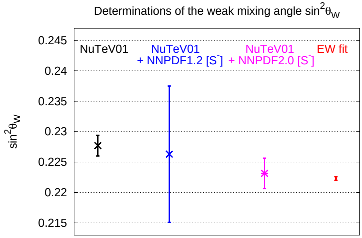
- LHC standard Candles

	$\sigma(W^+)\mathrm{Br}\left(W^+ \to l^+\nu_l\right)$	$\sigma(W^-)\mathrm{Br}\left(W^- \to l^+\nu_l\right)$	$\sigma(Z^0)\mathrm{Br}\left(Z^0 \to l^+l^-\right)$
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

- LHC standard Candles

	$\sigma(W^+)\mathrm{Br}\left(W^+ \to l^+\nu_l\right)$	$\sigma(W^-)\mathrm{Br}\left(W^- \to l^+\nu_l\right)$	$\sigma(Z^0)\mathrm{Br}\left(Z^0 \to l^+l^-\right)$
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

- Impact on NuTeV determination of $\sin^2\theta_W$



Conclusions

- A **reliable** estimation of **PDF uncertainties** is crucial in order to exploit the full physics potential of the LHC experiments.
- The **NNPDF methodology** based on using **Monte Carlo** techniques and **Neural Networks** is well suited to address problems of standard fits.
- **No** sign of **strong tension** among different datasets
- Officially **released** NNPDF sets (NNPDF 1.0/1.2/2.0) are available within the **LHAPDF** interface.
- Next steps:
  - Improved treatment of Heavy Flavour contributions, NNPDF 2.x
  - Inclusion of higher order contributions (QCD/EW effects), NNPDF x.x
  - ...

BACKUP SLIDES

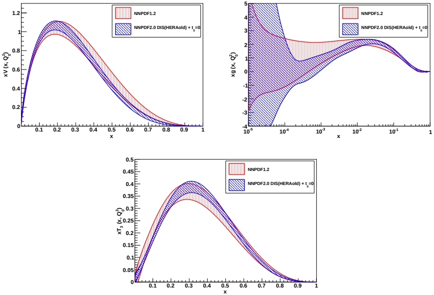
NLO QCD

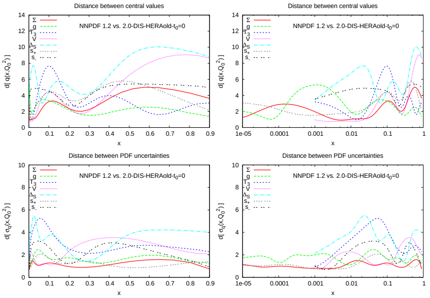
Fast Evolution

- Implementation of a new strategy to solve DGLAP evolution equation
- Evolution is performed as interpolation using higher-oder interpolating polynomials (Hermite polyonomials)
- Implementation benchmarked against the Les Houches tables
- Gain in speed by a factor 30 (for a fit to 3000 datapoints)
- Speed of the evolution scales with number of points in the interpolating grid (compare to older implementations which scaled with number of datapoints).

Methodology

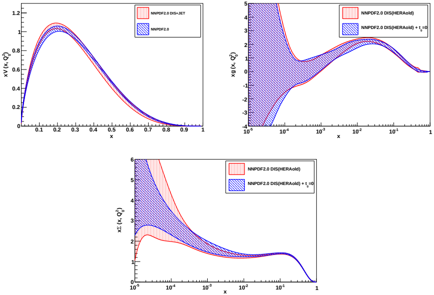
Impact of improved trainig/stopping

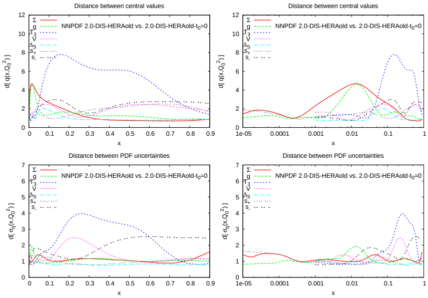




Methodology

Impact of t_0-method





Results

Some more phenomenological implications

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