Reweighting NNPDFs

Nathan Hartland
University of Edinburgh

The NNPDF Collaboration:
Richard D. Ball, Valerio Bertone, Francesco Cerutti,
Luigi Del Debbio, Stefano Forte, Alberto Guffanti, N.H,
José I. Latorre, Juan Rojo and Maria Ubiali.

Young Theorists' Forum 2011
Durham University
14th December 2011

Outline

1. Parton distribution fitting
2. The NNPDF methodology
3. The reweighting method

Parton distributions for the LHC

$$\sigma_X = \sum_{a,b} \int_0^1 dx_1 dx_2 f_a(x_1, Q^2) f_b(x_2, Q^2) \sigma_{q_a q_b \to X} \left(x_1, x_2, Q^2\right)$$

- Need to have a reliable determination of PDFs for LHC physics!
- For many EW processes PDF uncertainty dominates - good uncertainty analysis is crucial.

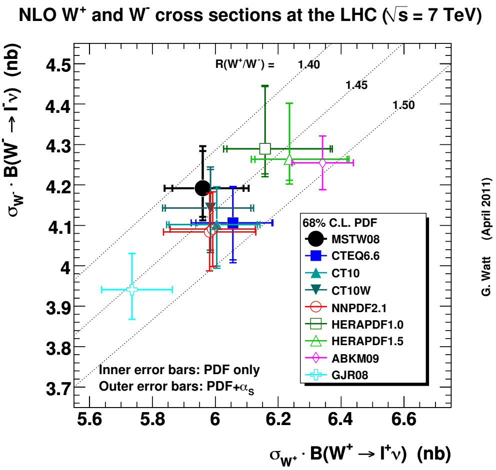
PDF x-dependence determined by a fit to experimental data.

PDF behaviour in Q^2: DGLAP Evolution.

Fitting groups

- MSTW
- CTEQ
- NNPDF
- HERAPDF
- ABKM
- GJR

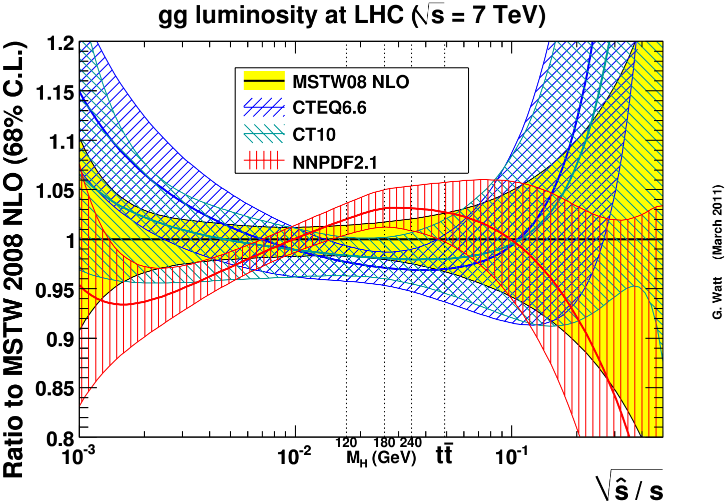
Parton distributions for the LHC



G. Watt [arXiv:1106.5788 [hep-ph]]

Standard Candles: Generally good agreement between global PDF fits.

Parton distributions for the LHC



G. Watt [arXiv:1106.5788 [hep-ph]]

Discrepancies: Use envelope of PDF uncertainties.

Parton distribution fitting - the standard approach

Choose a functional form for your PDFs.

Typical Parametrisations

- MSTW08

$$f(x, Q_0^2) \sim ax^b(1-x)^c(1+d\sqrt{x}+ex),$$

- CT10

$$f(x, Q_0^2) \sim ax^b(1-x)^c \exp\left(dx + ex^2 + f\sqrt{x}\right).$$

MSTW08, CT10 fits have in total 20 - 26 free parameters.

Evolve to the required scale, and compute physical observables.
Minimise some measure of fit quality w.r.t. the data.

$$\chi^2(a) = \frac{1}{N_{dat}} \sum_{i,j=1}^{N_{dat}} (D_i - T_i(a))(\sigma_{ij})^{-1}(D_j - T_j(a)).$$

Standard approach to parton fitting - uncertainties

How to propagate uncertainties from the experimental data to the PDFs?
Standard Approach: Linear propagation of uncertainties by Hessian Method.

- For a set of fit parameters $\{a\}$ define a tolerance in χ^2:

$$\Delta\chi^2(a) \equiv \chi^2(a) - \chi^2(a^{\min}) = \sum_{i,j=1}^{n} H_{ij}(a_i - a_i^{\min})(a_j - a_j^{\min}).$$

- Determine PDFs on surface of constant $\Delta\chi^2 = T$ in parameter space.
  - Numerical difficulties: Need to use rescaled eigenvectors of H.
- Obtain $2n$ PDF sets $S_i^{\pm}$, where n is the number of free parameters in the fit.

- Uncertainty in an observable $\mathcal{O}$ given by:

$$\mathrm{Var}[\mathcal{O}] = \frac{1}{2}\sum_{i=0}^{n}(\mathcal{O}[S_i^+] - \mathcal{O}[S_i^-])^2.$$

Difficulties with the standard approach

- Determining correct tolerance T.
  - Ideal value $T = 1$ leads to unrealistically small errors.
  - CTEQ choice $T \sim 100$
  - MSTW - dynamical procedure: $T \sim 5 - 20$.

- Parametrisation Bias
  - Choice of functional form may lead to biased fit.
  - Inflexible parameterisation may artificially constrain PDF errors.
  - PDF errors may *increase* when data is added due to the need to add more parameters.

Can we do things differently?

Outline

1 Parton distribution fitting

2 The NNPDF methodology

3 The reweighting method

NNPDF parameterisation

Idea: Use artificial neural networks to parameterise the PDFs.

- Neural nets: just another functional form.
- Each neural net in the NNPDF fit - 37 free parameters.

Robust, massively redundant parameterization:

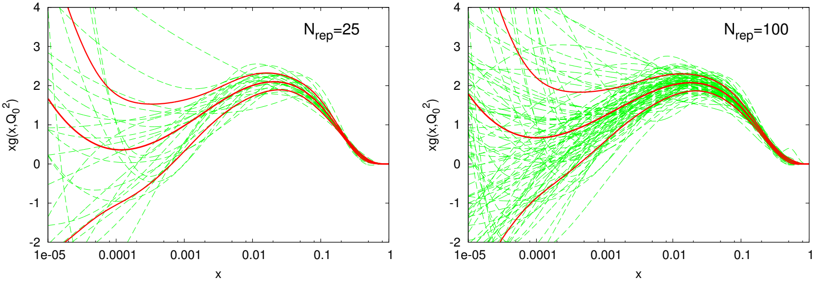
- An NNPDF parton fit has in total 259 fit parameters.
- Neural nets are extremely flexible, unbiased function approximators.

NNPDF functional form

$$f(x) = (1-x)^a x^{-b} \mathrm{NN}(x).$$

Monte Carlo uncertainty determination

- Form an ensemble of N artificial data 'replicas' by importance sampling the original data set.
- The ensemble of artificial data replicas forms a representation of the probability distribution in data.
- Perform a separate fit to each data replica, obtain an ensemble of PDF replicas.



NNPDF User Guide

Central value predictions

$$\langle \mathcal{O} \rangle = \frac{1}{N} \sum_{k=1}^{N} \mathcal{O}[f_k]\,.$$

Uncertainties

$$\mathrm{Var}[\mathcal{O}] = \frac{1}{N} \sum_{k=1}^{N} (\mathcal{O}[f_k] - \langle \mathcal{O} \rangle)^2.$$

Uncertainties in the PDFs faithfully represent the experimental data.
No tolerance criterion is required.

Outline

1. Parton distribution fitting
2. The NNPDF methodology
3. The reweighting method

Including new experimental data

How can we add new data to an existing parton set?

- Full Refit
  Time consuming, can only be done by the fitting collaboration.
- Reweight existing Monte Carlo parton set. Giele, Keller [hep-ph/9803393]

If the new data is statistically independent of the data in the prior set:

$$\mathcal{P}_{\rm new}(f) = \mathcal{N}_\chi \mathcal{P}(\chi^2|f)\, \mathcal{P}_{\rm old}(f),$$

$$\langle \mathcal{O} \rangle_{\rm new} = \int \mathcal{O}[f]\, \mathcal{P}_{\rm new}(f)\, Df = \frac{1}{N} \sum_{k=1}^{N} w_k \mathcal{O}[f_k].$$

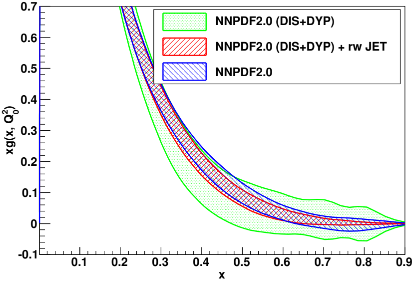
Weights determined by statistical inference

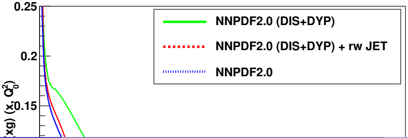
$$w_k = \mathcal{N}_\chi \mathcal{P}(\chi^2|f_k) = \frac{(\chi^2_k)^{(n-1)/2} e^{-\frac{1}{2}\chi^2_k}}{\frac{1}{N}\sum_{k=1}^{N} (\chi^2_k)^{(n-1)/2} e^{-\frac{1}{2}\chi^2_k}}\,.$$

R. D. Ball *et al.* Nucl. Phys. B **849** 112 [arXiv:1012.0836].

Verification of the reweighting method

- Produce a 1000 replica NNPDF fit to DIS and DY data only.
  - NNPDF2.0(DIS+DY)
- Reweight this set with Tevatron inclusive jet data.
- Compare reweighted set to the full fit.
  - NNPDF2.0





Error rescaling parameter

Useful tool for analysing experimental uncertainties.

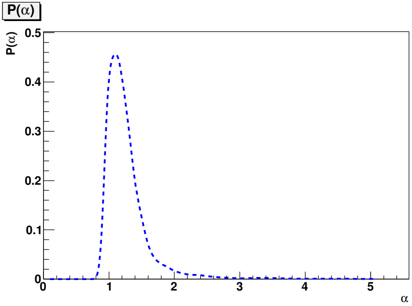
- Rescale uncertainties by a factor α.
- Compute a new weight $w_k(\alpha)$ with these uncertainties.
- Average over all replicas $\rightarrow$ probability of rescaling uncertainties by α.

$$\chi^2_{k,\alpha} = \chi^2_k/\alpha^2,$$

$$w_k(\alpha) = (\chi^2_{k,\alpha})^{(n-1)/2}\mathrm{e}^{-\chi^2_{k,\alpha}/2},$$

$$\mathcal{P}(\alpha) \propto \frac{1}{\alpha}\sum_{k=1}^{N} w_k(\alpha).$$

Tevatron Jets $\mathcal{P}(\alpha)$

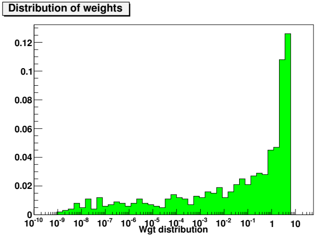


Ensemble Efficiency

- Prior set: maximally efficient representation of probability distribution.
- Reweighted set: loss of efficiency due to very small weights.
- Many replicas no longer contribute to the ensemble.

Loss of information quantified by the Shannon entropy:

$$N_{\text{eff}} \equiv \exp\left(\frac{1}{N_{\text{rep}}}\sum_{k=1}^{N_{\text{rep}}} w_k \ln(N_{\text{rep}}/w_k)\right)$$



A very low N_{eff} means data is either:

- Very constraining
- Inconsistent with prior

Tevatron Jet reweighting: $N_{\text{eff}} = 334.5$.

Producing unweighted PDFs

Aim: to produce a set of PDFs without weights, but equivalent to a reweighted set.

Method: Deterministically sample with replacement weighted replica distribution.

- Form new set with N'_{rep} replicas.
- Replicas with small weights not included in new set.
- Replicas with very high weights counted repeatedly.

Weights in reweighted set represented as multiplicities in unweighted set.

Allows us to perform further checks on the consistency of the procedure.

Consistency check - Multiple Reweighting

Reweighting with two data sets can be performed by:

- Reweighting with the combined dataset.
- Reweighting with one data set, and then the other.

To check consistency of reweighting, need to demonstrate

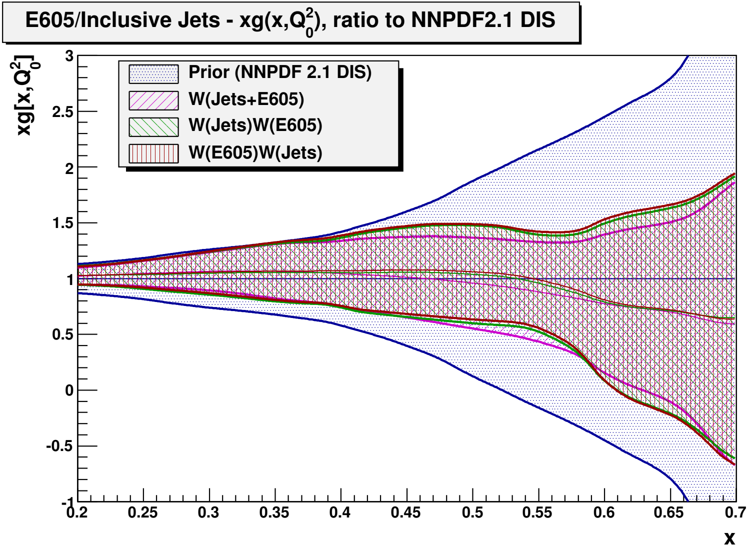
$$\hat{U}\hat{R}_{12} = \hat{U}\hat{R}_2\hat{U}\hat{R}_1 = \hat{U}\hat{R}_1\hat{U}\hat{R}_2.$$

Test Case: We reweight a DIS only fit (NNPDF2.1 DIS) with

- E605 Fixed target Drell-Yan data.
- Tevatron Run II inclusive jet data.

	Jets	E605	Jets +E605
Data points	186	119	305
$N_{\rm eff}$	627.1	59.5	63.7

Consistency check - Multiple Reweighting



Application: NNPDF2.2 Parton Set .

First global PDF set including data from the LHC

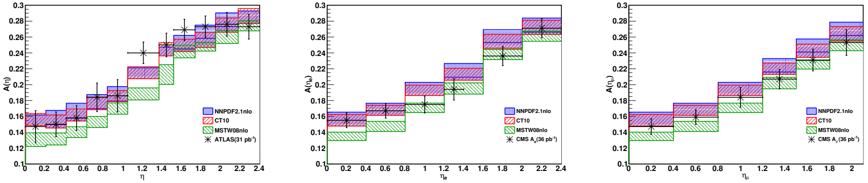
New data added by reweighting NNPDF2.1 Fit: W leptonic charge asymmetry.

R. D. Ball *et al*, Nucl. Phys. B **855** 608 [arXiv:1108.1758] .

Defined in terms of $W^{\pm} \to l^{\pm}\nu_l$ differential cross-sections $d\sigma_{l^\pm}/d\eta_l$

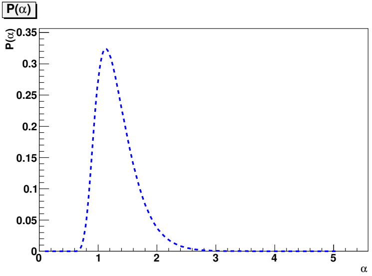
$$A_W^l = \frac{d\sigma_{l^+}/d\eta_l - d\sigma_{l^-}/d\eta_l}{d\sigma_{l^+}/d\eta_l + d\sigma_{l^-}/d\eta_l},$$

- ATLAS μ charge asymmetry. [arXiv:1103.2929]
- CMS $e + \mu$ charge asymmetry. [arXiv:1103.3470]
- D0 $e + \mu$ charge asymmetry. [arXiv:0709.4254]

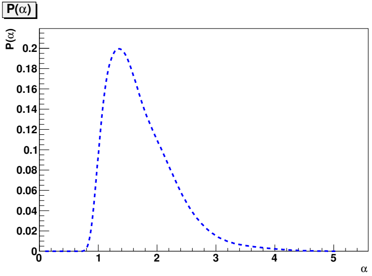


Fit quality

NNPDF2.1 LHC

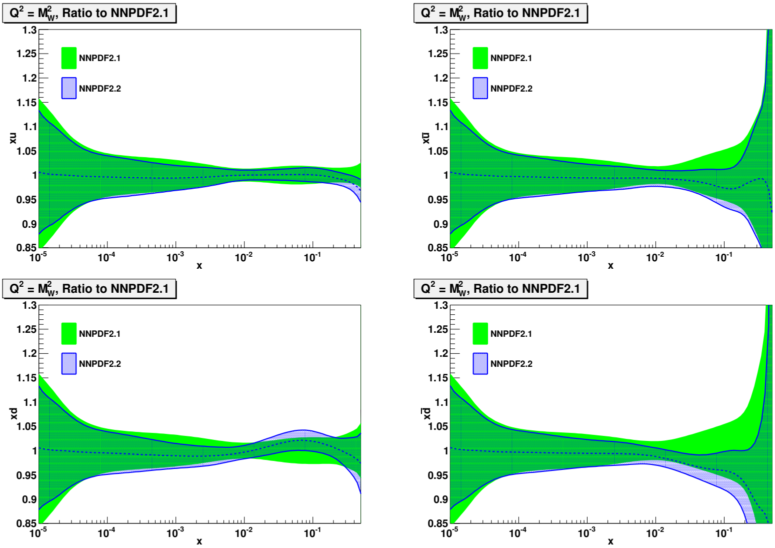


NNPDF2.2



Experiment	N_{dat}	NNPDF2.1	NNPDF2.1 LHC	NNPDF2.2
ATLASmuASY	11	[0.77]	0.97	1.07
CMSeASY	6	[1.83]	1.23	1.08
CMSmuASY	6	[1.24]	0.63	0.56
D0eASY	12	[4.39]	[3.46]	1.38
D0muASY	10	[1.48]	[1.17]	0.35
Full Dataset		1.165	1.158	1.157

First LHC constraints on parton distributions



Summary

- NNPDF Parton Sets
  - Neural Network parametrisation of PDFs.
    Redundant parametrisation for an unbiased fit.
  - Monte Carlo uncertainty determination.
    Faithful representation of the experimental uncertainties.

- Bayesian Reweighting
  - Powerful technique for including new data into existing parton fits.
    Fast assessment of data impact.

- NNPDF2.2
  - First PDF set including LHC data.
    W lepton asymmetry data provides small constraint on light quark PDFs.

NNPDF2.1 and NNPDF2.2 PDF Sets available through LHAPDF