Parton Distributions with LHC Data

Nathan Hartland
University of Edinburgh

The NNPDF Collaboration:
R. D. Ball, V. Bertone, S. Carrazza,
F. Cerutti, C. Deans, L. Del Debbio, S. Forte, N.H
A. Guffanti, J.I. Latorre, J. Rojo and M. Ubiali.

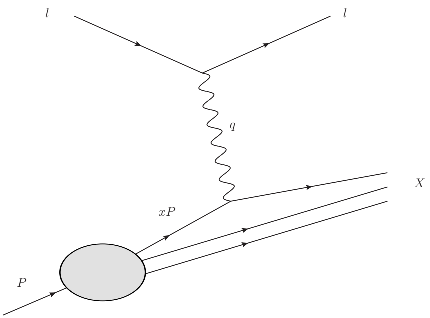
HEP phenomenology joint Cavendish-DAMTP seminar University of Cambridge
Thursday 29th November 2012

What is a parton distribution function?

QCD Factorisation:

When considering a scattering process with a single hadron in the initial state, the calculation may be factorized into a soft part and a perturbatively calculable hard part.

$$\sigma_X(Q^2) = \sum_a \int_0^1 dx\ f_a(x, \mu^2)\sigma_{q_a \to X}\left(x, \frac{Q^2}{\mu^2}\right)$$



$\sigma_{q_a \to X}$ - perturbative

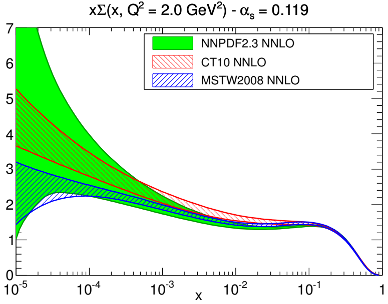
Hard cross section for lepton scattering off a parton of flavour a, carrying a fraction x of the parent hadron's momentum.

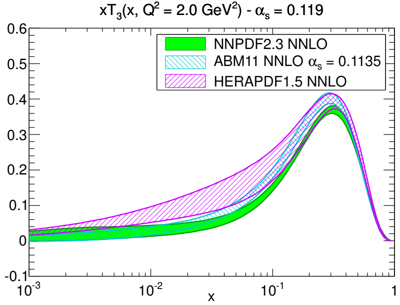
$f_a(x, \mu^2)$ - non-perturbative

Parton distribution function describing nonperturbative dynamics of target hadron. At LO can be interpreted as the probability of finding a parton of flavour a with momentum fraction x inside the target hadron.

Parton distributions - what's on the market?

- Many different determinations of parton distributions are available.





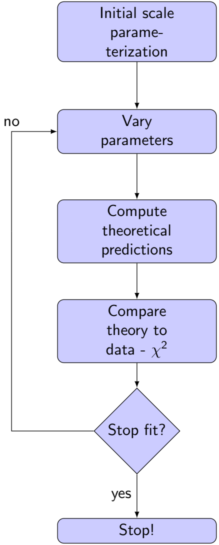
$$\Sigma(x) = \sum_q q(x) + \bar{q}(x)$$

$$\begin{aligned} T_3(x) &= u_V(x) - d_V(x). \\ &= (u(x) - \bar{u}(x)) - (d(x) - \bar{d}(x)) \end{aligned}$$

Collaborations: MSTW, CTEQ, NNPDF, HERAPDF, ABM, GJR

How can we determine proton PDFs?

1. Theoretical input
   - (N)NLO QCD, α_S, HQ Treatment
2. PDF Parameterization
   - What is a suitable choice of functional form?
3. Theoretical predictions
   - How can we make fast pQCD predictions for experimental data while including higher order corrections?
4. Comparison to data
   - What does the (LHC) data tell us about proton structure?



In this talk $\rightarrow \chi^2[f] = \frac{1}{N_{\rm dat}} \sum_{i,j}^{N_{\rm dat}} (D_i - T_i[f]) \, \sigma_{ij}^{-1} \, (D_j - T_j[f])$.

What do we know from theory?

Factorisation scale dependance:

PDFs evolve with scale according to the DGLAP equations.

$$\mu^2 \frac{\partial f(x,\mu^2)}{\partial \mu^2} = \int_x^1 \frac{dy}{y} P\left(\frac{x}{y}\right) f(y,\mu^2)$$

Where P are the perturbatively calculable splitting functions.

Theoretical constraints:

- PDF Sum Rules

$$\int_0^1 dx\, x(\Sigma(x) + g(x)) = 1, \qquad \sum_q \int_0^1 dx\, (q(x) - \bar{q}(x)) = 3$$

- Approx asymptotic behaviour

$$x \to 0: \quad f_V(x) \sim x^{-\alpha}$$
$$x \to 1: \quad f_V(x) \sim (1-x)^{\beta}$$

- Positivity of *physical observables* (F, σ)
  - Beyond LO pdfs are not restricted to be positive.

Aside from these constraints,
x dependence must be determined by fitting to experimental data!

Parton distribution fitting - initial scale parameterization

- PDFs at the initial scale are parametrized by some functional form

Typical Parameterizations

- MSTW08 $\sim$ 28 total PDF parameters

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

- CT10 $\sim$ 26 total PDF parameters

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

- HERAPDF $\sim$ 10 total PDF parameters

$$f_v(x) \sim ax^b(1-x)^c \exp\left(1+dx+ex^2\right).$$

NNPDF functional form

- NNPDF $\sim$ 259 total PDF parameters

$$f(x) \sim ax^b(1-x)^c \mathrm{NN}(x).$$

- Attempt to minimise figure of merit by varying $(a..f)$.
- Choice of functional form: Parameterization *bias*

NNPDF Strategy

- Minimise bias by choosing extremely flexible functional form
- Each PDF parametrized by a 2-5-3-1 Neural Network
- 259 Free parameters $\rightarrow$ **massively redundant** parameterization
- $b, c \rightarrow$ randomised preprocessing.

Theoretical Predictions

Calculate theoretical predictions for comparison with experimental data.
Evolve to required scale and perform convolution with hard coefficients.

DIS data: cross sections parametrized in terms of **structure functions**:

$$F_i(x, Q^2) = \int \frac{dy}{y} C_i^j(y, \alpha_s(Q^2)) f_j\left(\frac{x}{y}, Q^2\right)$$

Hadron Collider data: perform double convolution over PDFs

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

Hadronic data dependant upon PDFs through parton-parton luminosities:

$$\Phi_{ij}(\tau, M_X^2) = \frac{1}{s} \int_\tau^1 \frac{dx_1}{x_1} f_i(x_1, M_X^2) f_j(\tau/x_1, M_X^2)$$

Minimisation and Stopping in NNPDF

Minimisation by genetic algorithms

Problem: Very large parameter space, $\chi 2$ highly nonlocal.

- Minimisation is challenging.

Solution: Genetic Algorithms (GA)

- Generate mutations of fit parameters.
- Select those mutations that minimise figure of merit.

Dynamical fit stopping by cross-validation

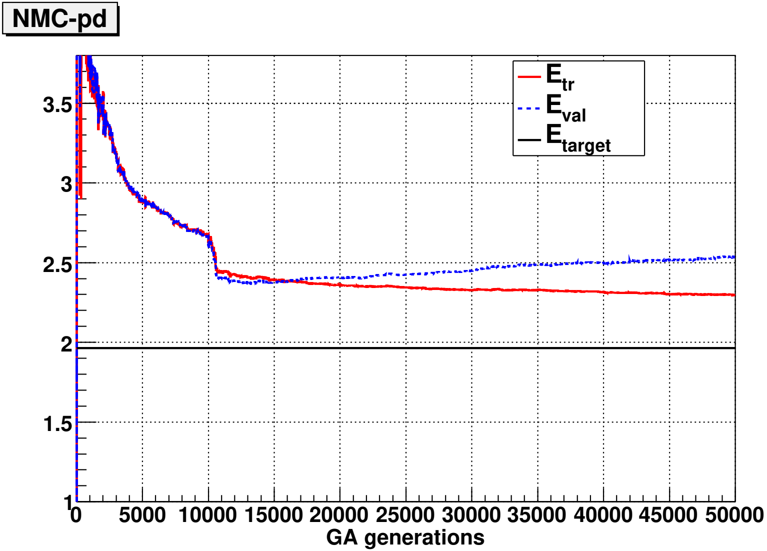
Problem: extremely flexible parameterisations are prone to *overfitting*.

- Fit has so many parameters, the minimum χ^2 corresponds to a fit not only to the data, but also statistical noise.

Solution: dynamical stopping by *Cross Validation*.

- Split the dataset into a training set and a validation set.
- Use the training set for minimisation, monitor the χ^2 to the validation set.
- Stop the fit when the χ^2 to the validation set starts to increase while the χ^2 to training set is still decreasing.

Cross Validation



Parton distribution fitting - Datasets

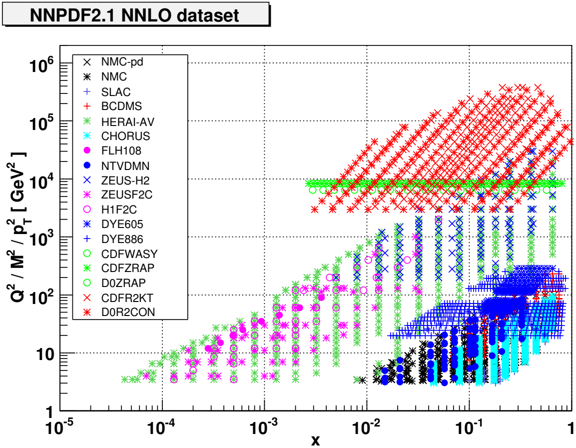
- PDF determination datasets divided into global and restricted determinations.

PDF Fit Datasets

Fit	Dataset
MSTW08	**global**
CT10	**global**
NNPDF2.1	**global**
HERAPDF1.5	HERA DIS
ABM11	DIS+DY

NNPDF2.1 NNLO Dataset

Experiment	Datapoints
DIS (Fixed Target)	1952
DIS (HERA)	834
Fixed Target DY	318
Tevatron WZ	70
Tevatron Jets	186
Total	3360



Reliable propagation of experimental error is crucial

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:

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

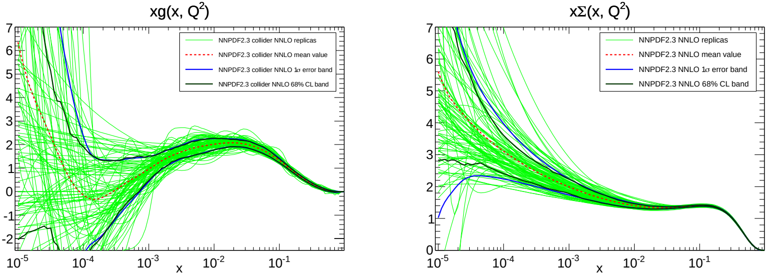
- Determine PDFs on surface of constant $\Delta\chi^2 = T$ in parameter space.
  - Numerical difficulties: Need to use rescaled eigenvectors of H.
- Obtain $2n$ PDF sets $S_i^{\pm}$, where n is the number of free parameters in the fit.

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

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

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.



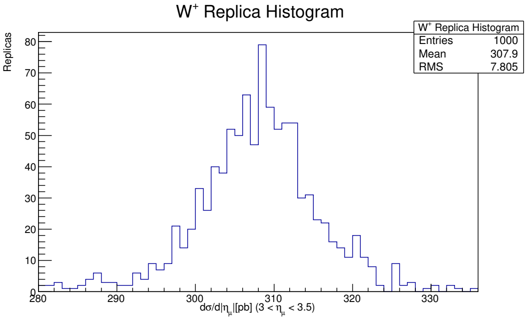
NNPDF User Guide

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.

Parton distributions for the LHC

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

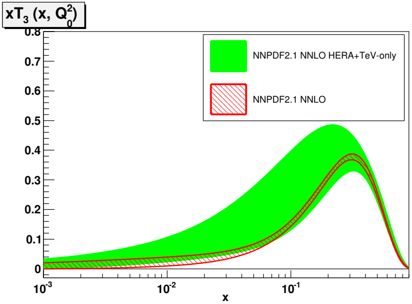
The LHC for parton distributions

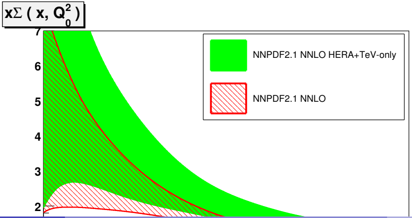
- LHC can access PDFs in new kinematic regions $\rightarrow$ high Q^2 and small-x.
- LHC measurements starting to be able to discriminate between PDF sets.

NNPDF collider only fits

Target: An NNPDF Fit based only upon collider data

- Free of contamination from higher twist corrections.
- Includes only proton data - no nuclear corrections required

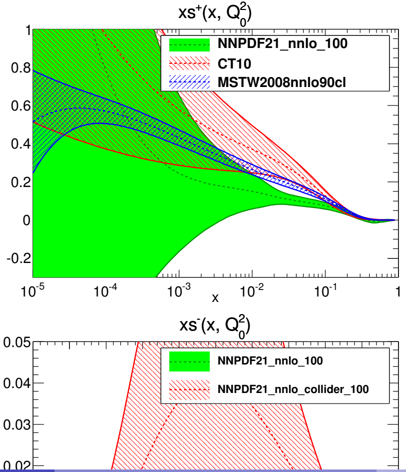




Strange content of the proton

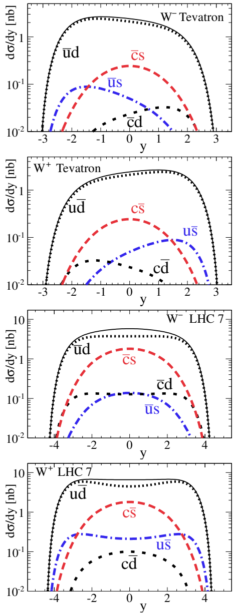
- Strange PDFs are relatively poorly constrained by existing data.
  - Primary constraints from dimuon production in Neutrino DIS

$$s^{+}(x) = s(x) + \bar{s}(x), \qquad s^{-}(x) = s(x) - \bar{s}(x).$$



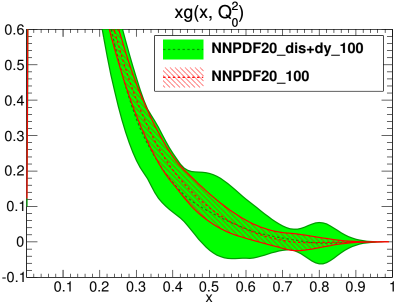
LHC Data for PDFs - EW vector boson production

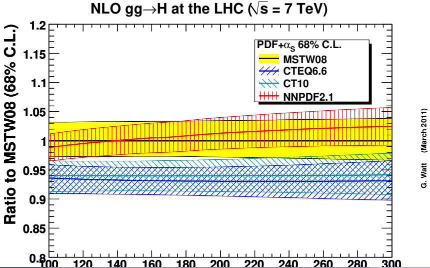
- $pp \to W + X \to l\nu + X$
  $pp \to Z + X \to l^+l^- + X$
  - Potentially a vital constraint upon light flavour separation.
  - Strange contribution **much** more important that at Tevatron.
- $pp \to W + c + X$
  - Direct probe of strange distribution ($|V_{cs}| > |V_{cd}|$)



LHC Data for PDFs - Inclusive Jets

- Precision inclusive jet data important for constraining the gluon distribution.





Including new experimental data

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

- Perform a new fit.
  (Difficult if a fast implementation of the theoretical prediction is not available).
- 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

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

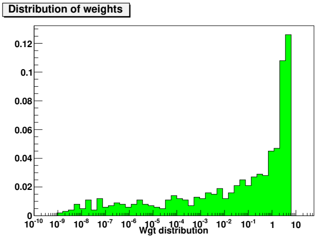
R. D. Ball *et al.* Nucl. Phys. B **849** 112 [arXiv:1012.0836].

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:

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



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

- Very constraining
- Inconsistent with prior

Error rescaling parameter

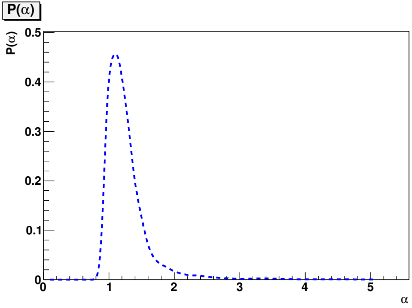
Useful tool for analysing experimental uncertainties. Differentiates between inconsistent and constraining data when N_{eff} is small.

- 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.

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

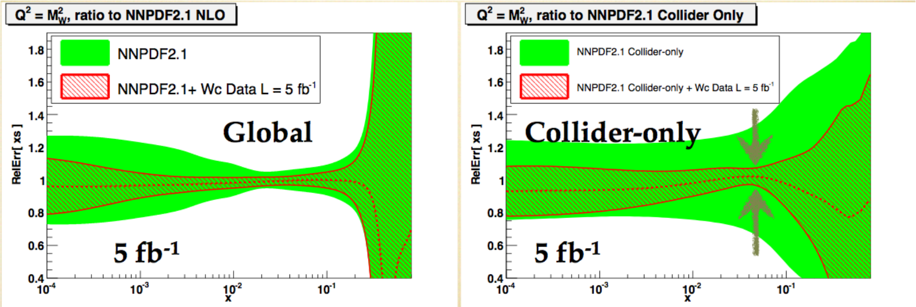
$$w_k(\alpha) = (\chi^2_{k,\alpha})^{(n-1)/2} \mathrm{e}^{-\chi^2_{k,\alpha}/2},$$

$$\mathcal{P}(\alpha|\chi^2) \propto \frac{1}{\alpha} \sum_{k=1}^{N} w_k(\alpha).$$



Reweighting Application: W+c Pseudodata

- Generate MC pseudo data for W+c events with CMS kinematics
- Reweight NNPDF2.1 fit with artificial data to investigate possible impact of high statistics measurements.



- Suggests substantial constrain possible with 5fb^{-1}
- Brings collider only strangeness to $\sim$ same accuracy as global fit.
- After 5fb^{-1} uncertainties are systematics dominated $\rightarrow$ W+jets.

(J.Rojo)

Reweighting Application: NNPDF2.2 Parton Set .

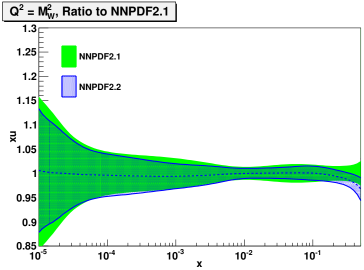
New data added by reweighting NNPDF2.1 Fit: W leptonic charge asymmetry.

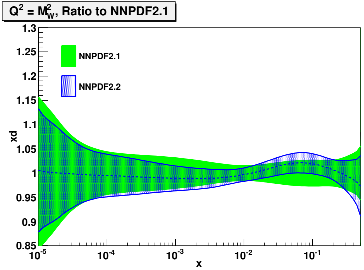
R. D. Ball *et al*, Nucl. Phys. B **855** 608 [arXiv:1108.1758] .

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]





Including new experimental data - fitting

How can we efficiently include LHC data into a full fit?
Tools: APPLgrid/FastNLO projects

- Precompute and store MC weights on an interpolation grid in x and Q^2:

$$\sigma = \sum_p \sum_l^{N_{\rm sub}} \int_0^1 dx_1 dx_2 \; \hat{\sigma}^{(p)(l)}\left(x_1, x_2, Q^2\right) F^{(l)}\left(x_1, x_2, Q^2\right) \rightarrow$$

$$\sigma = \sum_p \sum_l^{N_{\rm sub}} \sum_{\alpha,\beta}^{N_x} \sum_\tau^{N_Q} W^{(p)(l)}_{\alpha\beta\tau} F^{(l)}\left(x_\alpha, x_\beta, Q^2_\tau\right) \qquad (1)$$

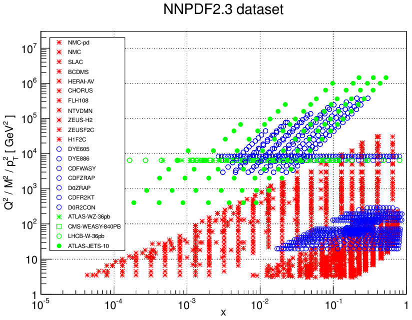
PDF Evolution in the FastKernel method is a similar procedure, Idea: Combine weight grids with evolution grids

$$f_i(x_\alpha, Q^2_\tau) = \sum_\beta^{N_x} \sum_j^{N_{\rm pdf}} A^\tau_{\alpha\beta ij} N^0_j(x_\beta) \quad \rightarrow \quad \sigma = \sum_{\alpha,\beta}^{N_x} \sum_{i,j}^{N_{\rm pdf}} \sigma_{\alpha\beta ij} N^0_i(x_\alpha) N^0_j(x_\beta)$$

- Precomputing all Q^2 dependence leads to extremely efficient calculations.

NNPDF2.3 - LHC Data

- The NNPDF2.3 dataset contains all relevant LHC data with covariance matrices
  - 36 pb^{-1} ATLAS Inclusive jet measurements [arxiv:1112.5141]
  - 35 pb^{-1} ATLAS W lepton and Z differential distributions [arxiv:1109.5141]
  - 37 pb^{-1} LHCb W lepton differential distribution [arxiv:1204.1620]
  - 840 pb^{-1} CMS W electron asymmetry [arxiv:1206.2598]



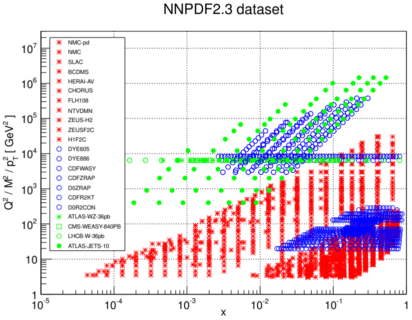
Methodological Improvements

Improved dynamical stopping.
Expanded genetic algorithm minimisation.
Improved training/validation partitioning.
Correction to NuTeV dimuon cross section.

	NNPDF2.1		**NNPDF2.3**	
	NLO	NNLO	NLO	NNLO
Fit Quality	1.15	1.16	1.10	1.14

NNPDF2.3 - LHC Data

- The NNPDF2.3 dataset contains all relevant LHC data with covariance matrices
  - 36 pb^{-1} ATLAS Inclusive jet measurements [arxiv:1112.5141]
  - 35 pb^{-1} ATLAS W lepton and Z differential distributions [arxiv:1109.5141]
  - 37 pb^{-1} LHCb W lepton differential distribution [arxiv:1204.1620]
  - 840 pb^{-1} CMS W electron asymmetry [arxiv:1206.2598]



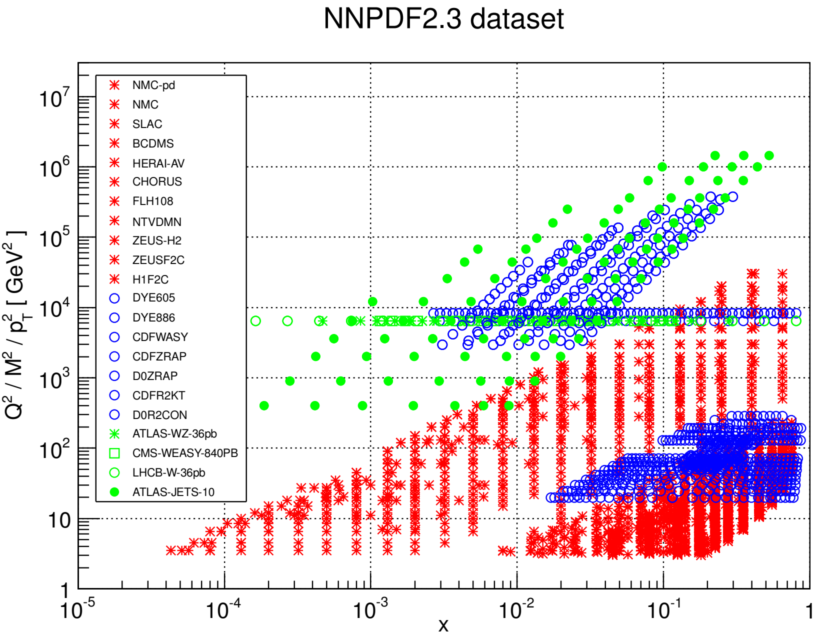
Methodological Improvements

Improved dynamical stopping.
Expanded genetic algorithm minimisation.
Improved training/validation partitioning.
Correction to NuTeV dimuon cross section.

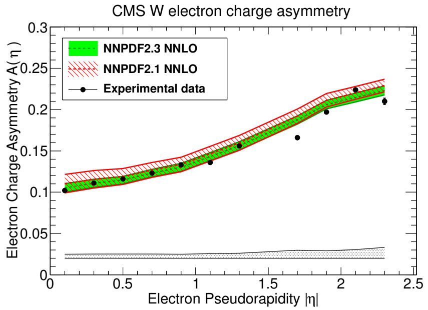
	NNPDF2.1		NNPDF2.3	
	NLO	NNLO	NLO	NNLO
Fit Quality	1.15	1.16	1.10	1.14

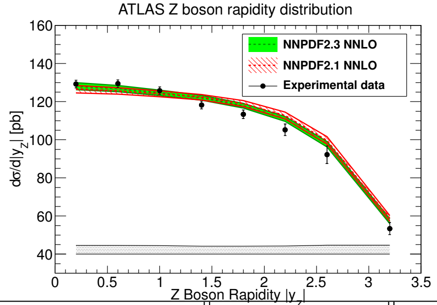
- Also in the NNPDF2.3 family
  - NNPDF2.3 noLHC: same dataset as NNPDF2.1, with improved methodology.
  - NNPDF2.3 Collider only: dataset restricted to HERA, Tevatron and LHC data.

NNPDF2.3 Dataset



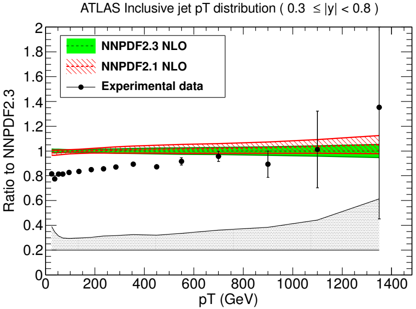
Impact of LHC EW vector boson data

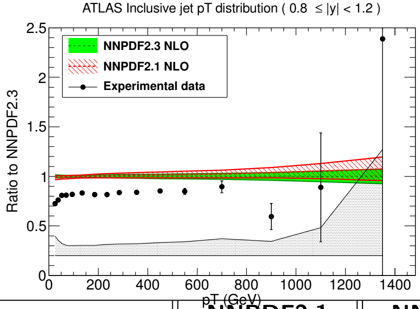




	NNPDF2.1	NNPDF2.3

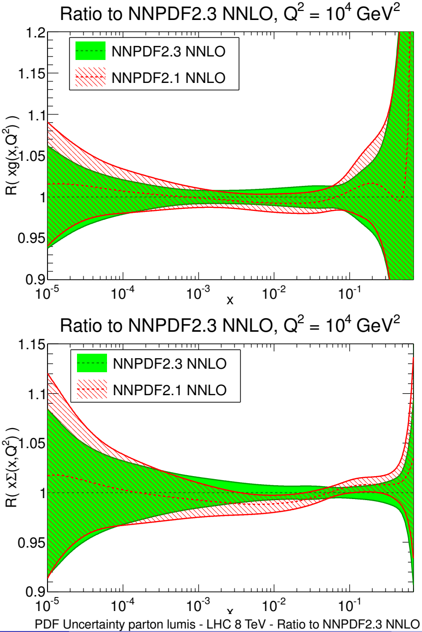
Impact of ATLAS inclusive jet data





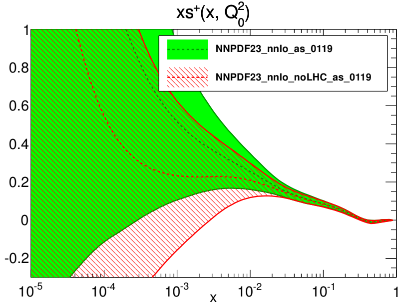
NNPDF2.1	NNPDF2.3

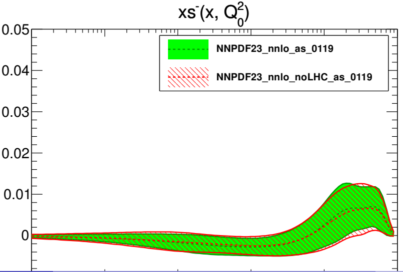
NNPDF2.3 vs NNPDF2.1



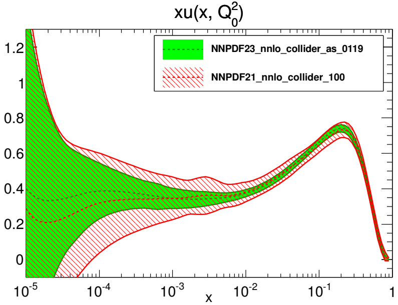
NNPDF2.3 vs NNPDF2.1 - strangeness

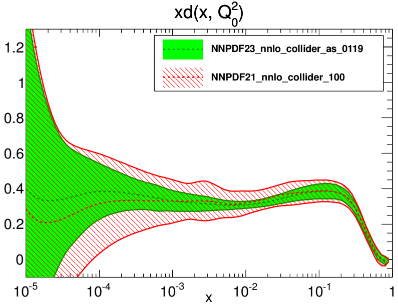
- Isolate impact of LHC data upon strangeness - noLHC vs global





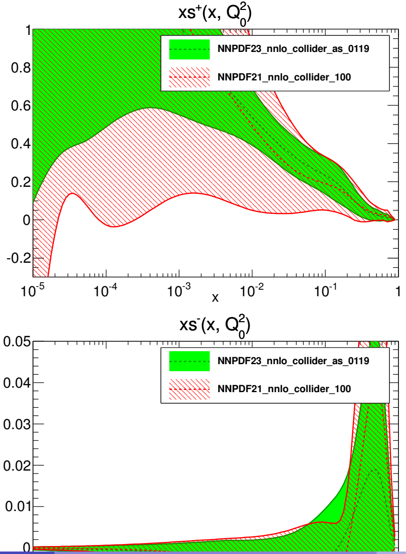
Collider only PDFs with LHC data





Collider only PDFs with LHC data - strangeness

- Substantial constraint upon strange sea distribution for the collider only fit.
- Data once again insensitive to $s - \bar{s}$.



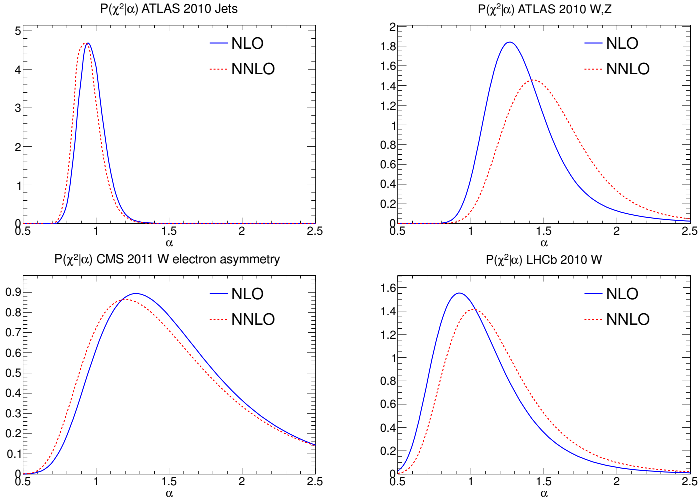
Consistency Checks - χ^2

- Consider χ^2 to each dataset in global fit, before and after inclusion of LHC data.
- Significant improvement in description of LHC data, without deterioration in fit quality to existing datasets.
- Demonstrates consistency of the global QCD analysis.

	NNPDF2.1		**NNPDF2.3**	
Experiment	NLO	NNLO	NLO	NNLO
Total	1.145	1.162	1.101	1.139
NMC-pd	0.97	0.93	0.95	0.95
NMC	1.68	1.58	1.61	1.59
SLAC	1.34	1.04	1.24	1.00
BCDMS	1.21	1.29	1.20	1.28
CHORUS	1.10	1.08	1.10	1.07
NTVDMN	0.70	0.50	0.43	0.56
HERAI-AV	1.04	1.04	1.00	1.01
FLH108	1.34	1.23	1.29	1.20
ZEUS-H2	1.21	1.21	1.20	1.22
ZEUS F_2^c	0.75	0.81	0.82	0.90
H1 F_2^c	1.50	1.44	1.59	1.53
DYE605	0.94	1.08	0.86	1.04
DYE886	1.42	1.69	1.27	1.58
CDF W asy	1.88	1.63	1.57	1.64
CDF Z rap	1.77	2.38	1.80	2.03
D0 Z rap	0.57	0.67	0.56	0.61
ATLAS W,Z	[1.57]	[2.21]	1.26	1.43
CMS W el asy	[2.02]	[1.27]	0.82	0.81
LHCb W	[0.89]	[1.13]	0.67	0.83
CDF RII k_T	0.68	0.65	0.60	0.68
D0 RII cone	0.90	0.98	0.84	0.94
ATLAS jets	[1.06]	[0.95]	1.00	0.94

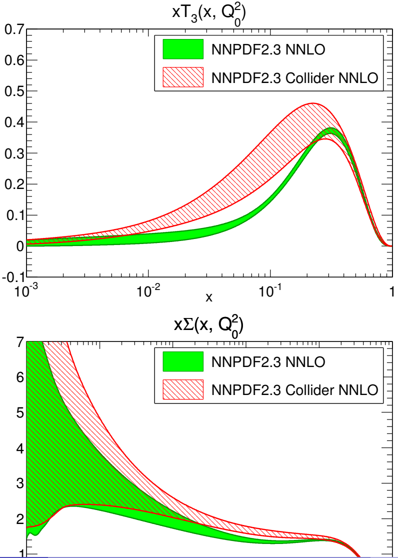
Consistency Checks - $P(\alpha)$

- (In)consistency of LHC data with existing dataset can be studied with $P(\alpha)$.



Should we be using collider only PDFs yet?

- LHC data providing excellent constraints upon previous collider only determination.



Summary

NNPDF Methodology and Reweighting

- NNPDF methodology provides a PDF determination free of parameterization bias, and with a robust propagation of experimental uncertainty.
- Reweighting MC parton distributions can provide valuable insight into constraints offered by LHC data, regardless of the availability of a fast observable code.

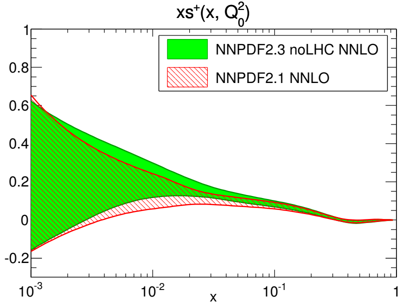
NNPDF2.3

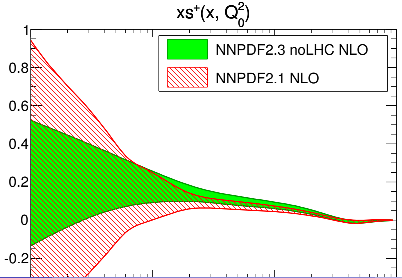
- New NNPDF parton set including constraints from LHC measurements of W/Z production and inclusive jet data.
- Addition of LHC data to the fit enabled by the FK method for collider observables $\rightarrow$ extremely fast observable computation.
- Strong constraints upon NNPDF collider only determination, however uncertainties remain large.

BACKUPS

NNPDF2.1 vs NNPDF2.3noLHC - strangeness

- Corrected NuTeV dimuon cross section





Unweighting procedure