NNPDF

- Partons: Issues
- NNPDF2.1
- Users Guide
- Reweighting

RDB, Valerio Bertone, Francesco Cerutti, Luigi Del Debbio,
Stefano Forte, Alberto Guffanti, Jose Latorre, Juan Rojo, Maria Ubiali
(Barcelona, Edinburgh, Copenhagen, Milan, Aachen)

Atlas/CERN May 2011

PDFs for LHC

To fully exploit LHC data, we need:

- Precise reliable faithful PDFs
  - Extract from DIS+hadronic data: “global fit”
- No theoretical bias beyond (N)NLO pQCD, etc.
  - No bias due to functional form
  - No bias due to improper statistical procedure
- Genuine statistical confidence level
  - Full inclusion of correlations in exp systematics
  - Uniform treatment of uncertainties

PDFs for LHC

To fully exploit LHC data, we need:

- Precise reliable faithful PDFs
  - Extract from DIS+hadronic data: “global fit”
- No theoretical bias beyond (N)NLO pQCD, etc.
  - No bias due to functional form
  - No bias due to improper statistical procedure
- Genuine statistical confidence level
  - Full inclusion of correlations in exp systematics
  - Uniform treatment of uncertainties

Is this actually possible?

Traditional PDF Fitting (eg MSTW,CTEQ,...)

Duke & Owens 1982

- Choose a functional form for each PDF: eg

$$f(x) = x^{\alpha}(1-x)^{\beta}(1+\gamma x^{1/2}+\delta x)$$

- Find the best fit to the data by minimising χ^2
  (using eg MINUIT: » 25 params)
- Estimate uncertainties using Hessian matrix
  (diagonalize: gives » 25 eigenvector sets)
- Find uncertainties too small, because parametrization inflexible: increase uncertainties by inflating exp errors (T » 50-100)

Traditional PDF Fitting (eg MSTW,CTEQ,...)

Duke & Owens 1982

- Choose a functional form for each PDF: eg

$$f(x) = x^{\alpha}(1-x)^{\beta}(1+\gamma x^{1/2}+\delta x)$$

- Find the best fit to the data by minimising χ^2
  (using eg MINUIT: » 25 params)
- Estimate uncertainties using Hessian matrix
  (diagonalize: gives » 25 eigenvector sets)
- Find uncertainties too small, because parametrization inflexible: increase uncertainties by inflating exp errors (T » 50-100)

Problems:

Inflexible parametrization) theoretical bias

Better data/theory requires more parameters) instabilities

Monte Carlo PDFs (eg NNPDF)

Giele & Kosower 1998

Forte & Latorre 2002

- Choose a very flexible functional form for each PDF:

  (eg a neural network: » 250 params)
- Generate data replicas (» 100-1000) using exp uncertainties
- Find a good fit to each data replica by optimising χ^2

  (best fit useless – fitting statistical noise:

  instead use genetic algorithm + cross-validation)
- Treat resulting PDF replicas as statistical ensemble:

  each equally probable (importance sampling)

So simple averages give central values, uncertainties etc.

Monte Carlo PDFs (eg NNPDF)

Giele & Kosower 1998

Forte & Latorre 2002

- Choose a very flexible functional form for each PDF:

  (eg a neural network: » 250 params)
- Generate data replicas (» 100-1000) using exp uncertainties
- Find a good fit to each data replica by optimising χ^2

  (best fit useless – fitting statistical noise:

  instead use genetic algorithm + cross-validation)
- Treat resulting PDF replicas as statistical ensemble:

  each equally probable (importance sampling).

  So simple averages give central values, uncertainties etc.

Advantages:

No theoretical bias due to parametrization

Statistically meaningful uncertainties: no need for tolerance

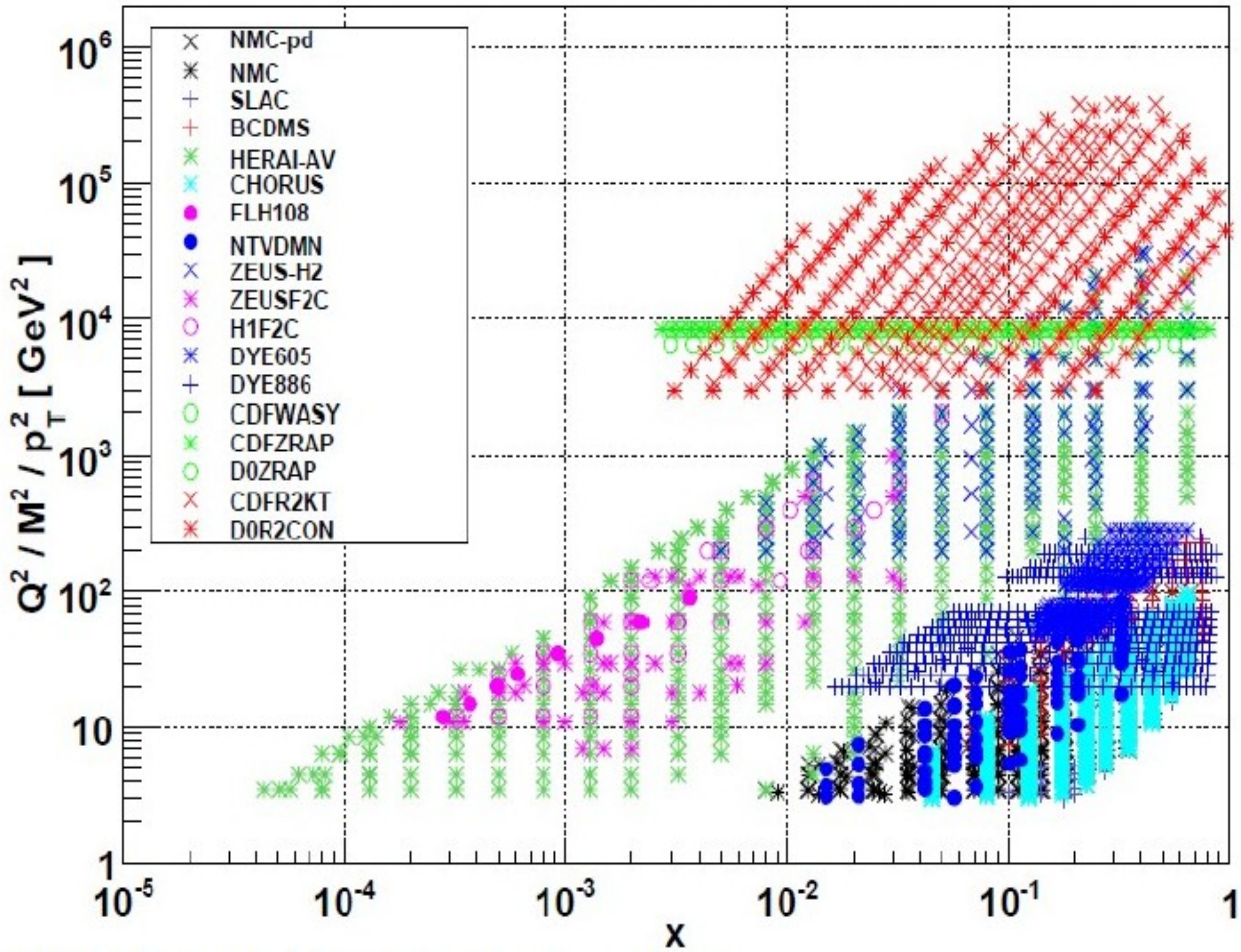
Technical stability: improved data/theory, same parametrization

NNPDF progress

- 2002: Structure Functions
- 2005: More Structure Functions
- 2007: Nonsinglet DIS partons
- 2008: First NLO DIS: NNPDF1.0
- 2009: Strange PDFs: NNPDF1.2
- 2010: First global NLO (DIS+DY+J): NNPDF2.0
- 2010: Reweighting (W-ev asymmetry)
- 2011: Global NLO + HQ: NNPDF2.1
- 2011: LO and NNLO (coming soon)

Major software development project

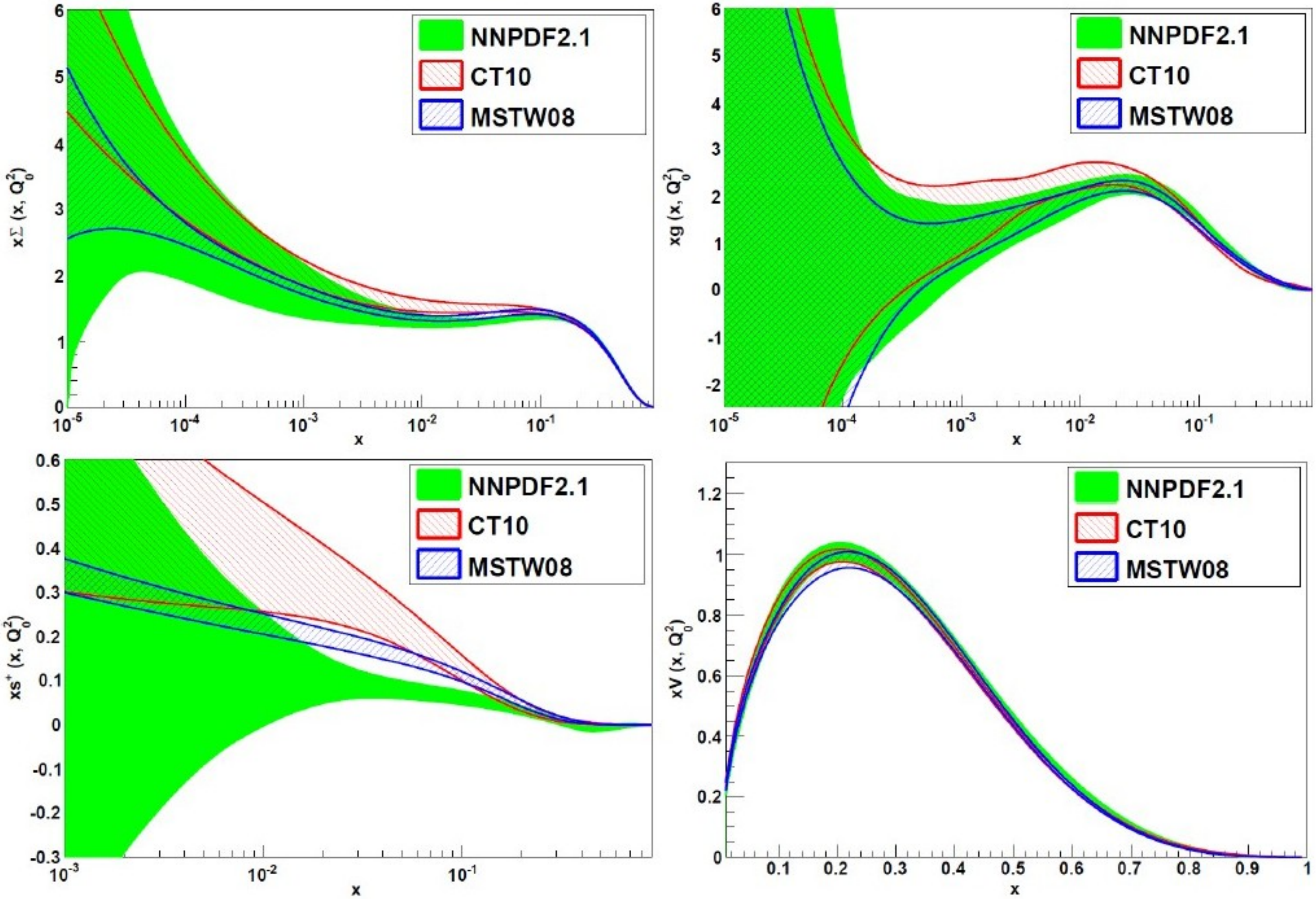
NNPDF2.1



- NLO pQCD
- no K-factors
- benchmarked
- 7 fitted PDFs (including $s, \bar{s}$)
- HQ FONLL-A
- No norm bias
- No param bias: 259 params
- 3477 data pts

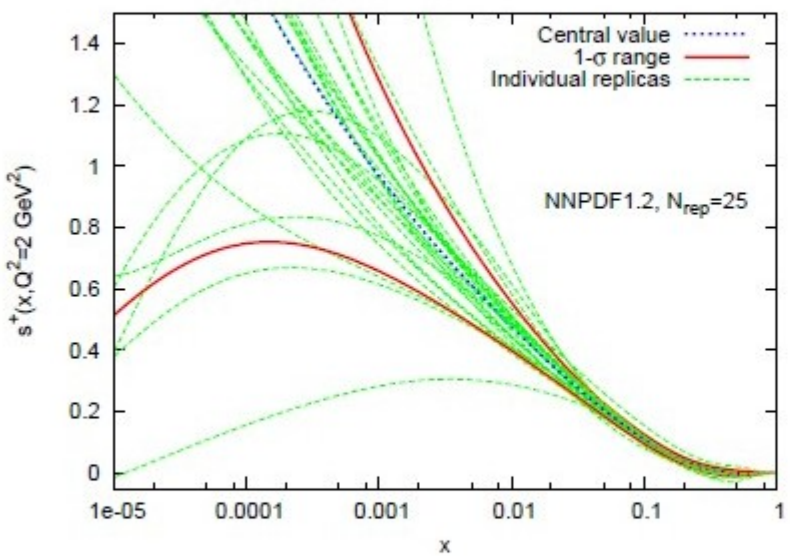
Sets with: 10 values of α_s, 3 values of m_c, 4 values of m_b, FFN scheme, etc,etc

NNPDF2.1 vs CT10 & MSTW08

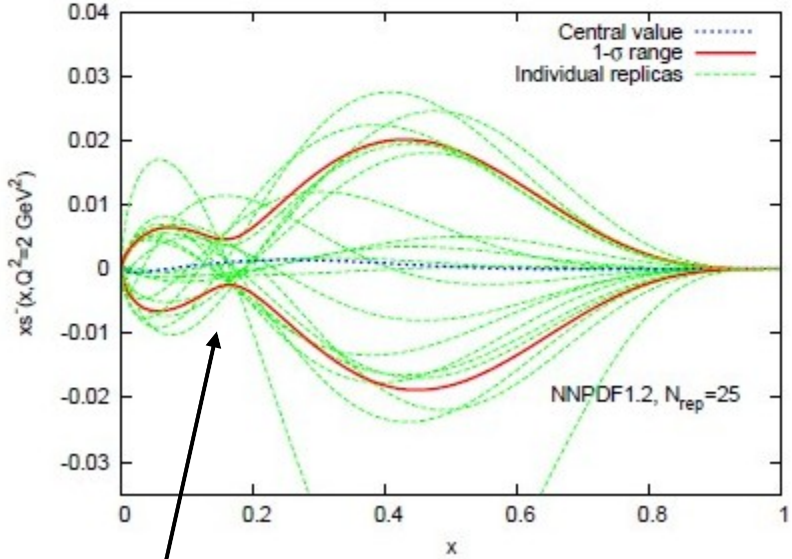


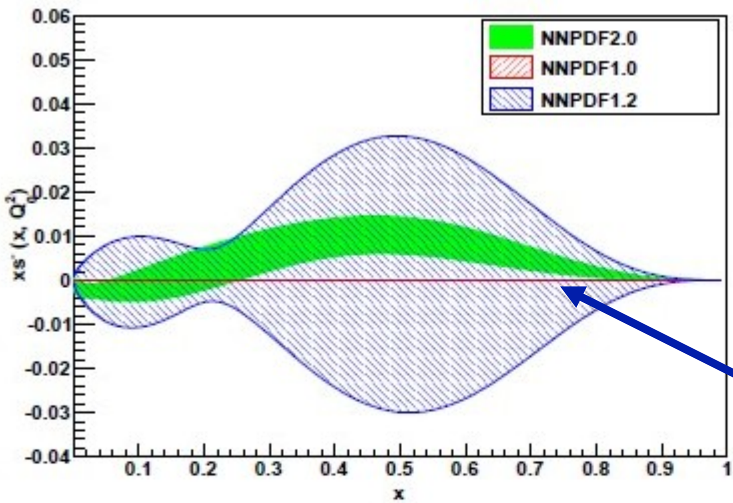
Strangeness

$s + \bar{s}$



$s - \bar{s}$





Crossings make s,sbar hard to fit:

CT10, NNPDF1.0: $s = \bar{s}$ 1 param

MSTW08: $s \neq \bar{s}$ 4 params

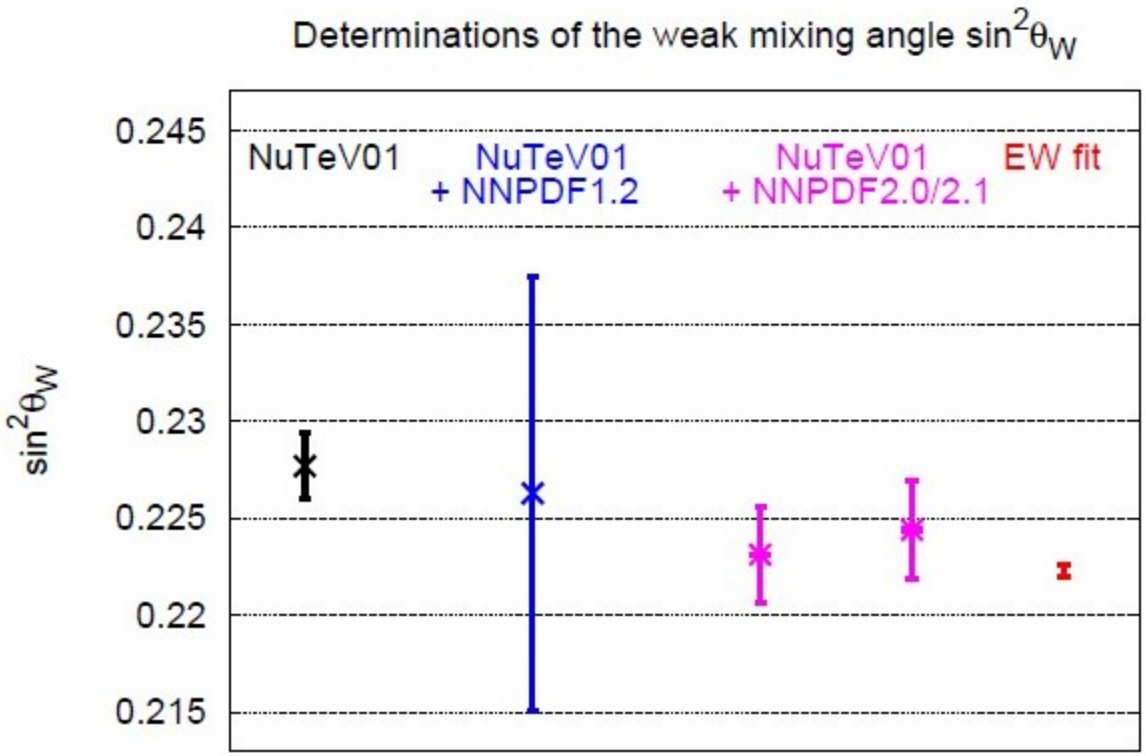
NNPDF1.2,2.0: $s \neq \bar{s}$ 74 params

Add DY and ν–DIS data (NNPDF2.0): constrains strangeness

The NuTeV Anomaly

Determination of $\sin^2\theta_W$ using neutrino DIS data: assumed $s = \bar{s}$:

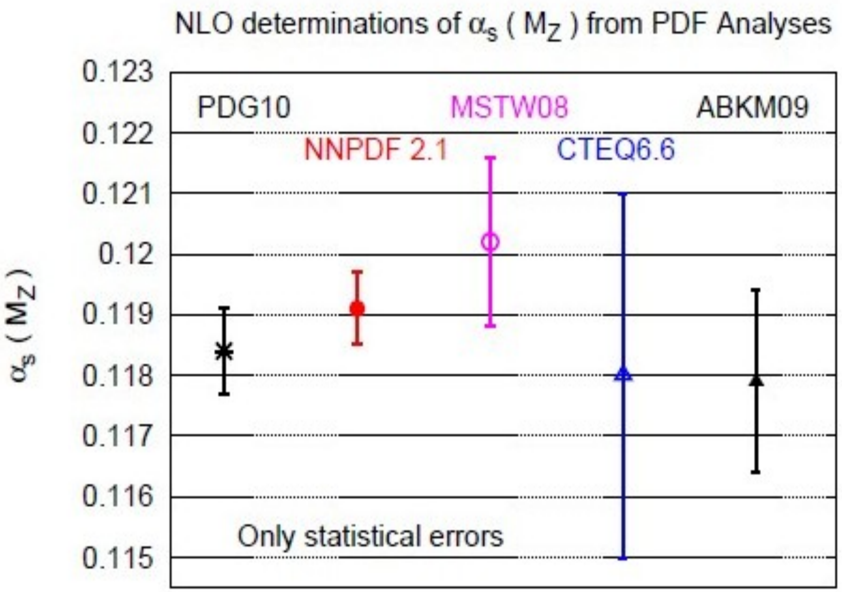
found 3-sigma discrepancy: new physics?

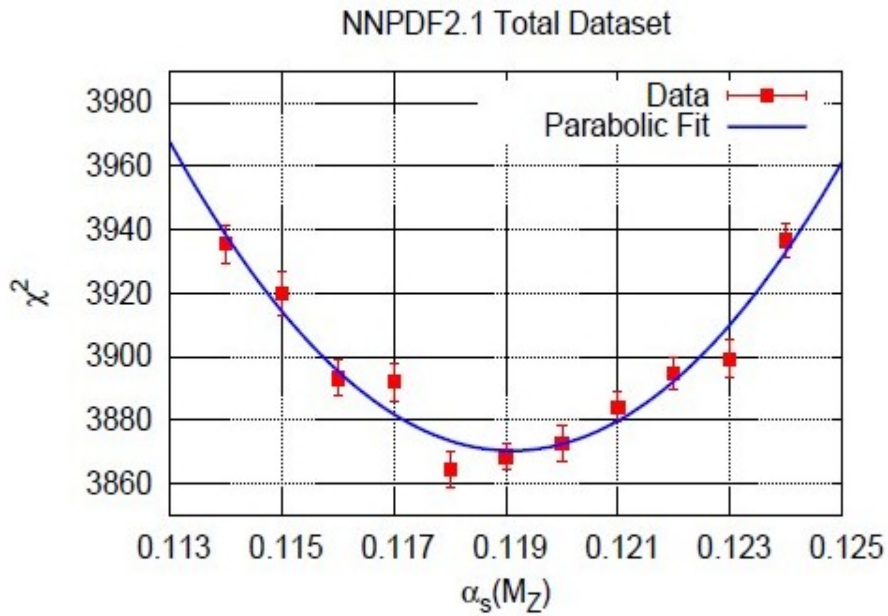


Include $s \neq \bar{s}$ using NNPDF1.2/2.0/2.1: discrepancy disappears!

Moral here for LHC....

Determining α_s





	$\alpha_s\,(M_Z)$
NNPDF2.1	$0.1191 \pm 0.0006^{\mathrm{stat}}$
NNPDF2.1 DIS–only	$0.1177 \pm 0.0009^{\mathrm{stat}}$

Uncertainty experimental only: theoretical uncertainties (NLO pQCD) rather larger

More flexible (NN)PDFs: more precise physics!

Summer 2010

PDF4LHC Recipe

The PDF4LHC group (CERN management mandate) coordinates studies and research in PDF determinations from different groups and is responsible for providing official recommendations for PDF use in LHC experiments

Current NLO recommendation for LHC analysis:

NLO Summary:

For the calculation of uncertainties at the LHC, use the envelope provided by the central values and PDF+α_s errors from the MSTW08, CTEQ6.6 and NNPDF2.0 PDFs, using each group's prescriptions for combining the two types of errors. We propose this definition of an envelope because the deviations between the predictions are currently greater than their uncertainties would strictly suggest. As a central value, use the midpoint of this

LHC experiments need to use NNPDFs

Users Guide

NNPDFs

- Download a set of NNPDFs (eg NNPDF2.1) from LHAPDF
- Each set of contains an ensemble of N 'replicas' (N=100,1000)
- Each replica f_k, k=1 … N is a set of PDFs: $\{g, u, \bar{u}, d, \bar{d}, \cdots\}$
  on a grid in x and Q^2 – just as usual
- Each replica f_k is equally probable as a candidate PDF.

For any observable O[f] depending on PDFs f:

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

"Master formula": all results are obtained using this

There are no "eigenvector sets" in NNPDF

Example 1: the PDFs

- Central values:

$$f_0 = \langle f \rangle = \frac{1}{N}\sum_{k=1}^{N} f_k$$

Note: f_0 is also given on LHAPDF as “set zero”

- Variances:

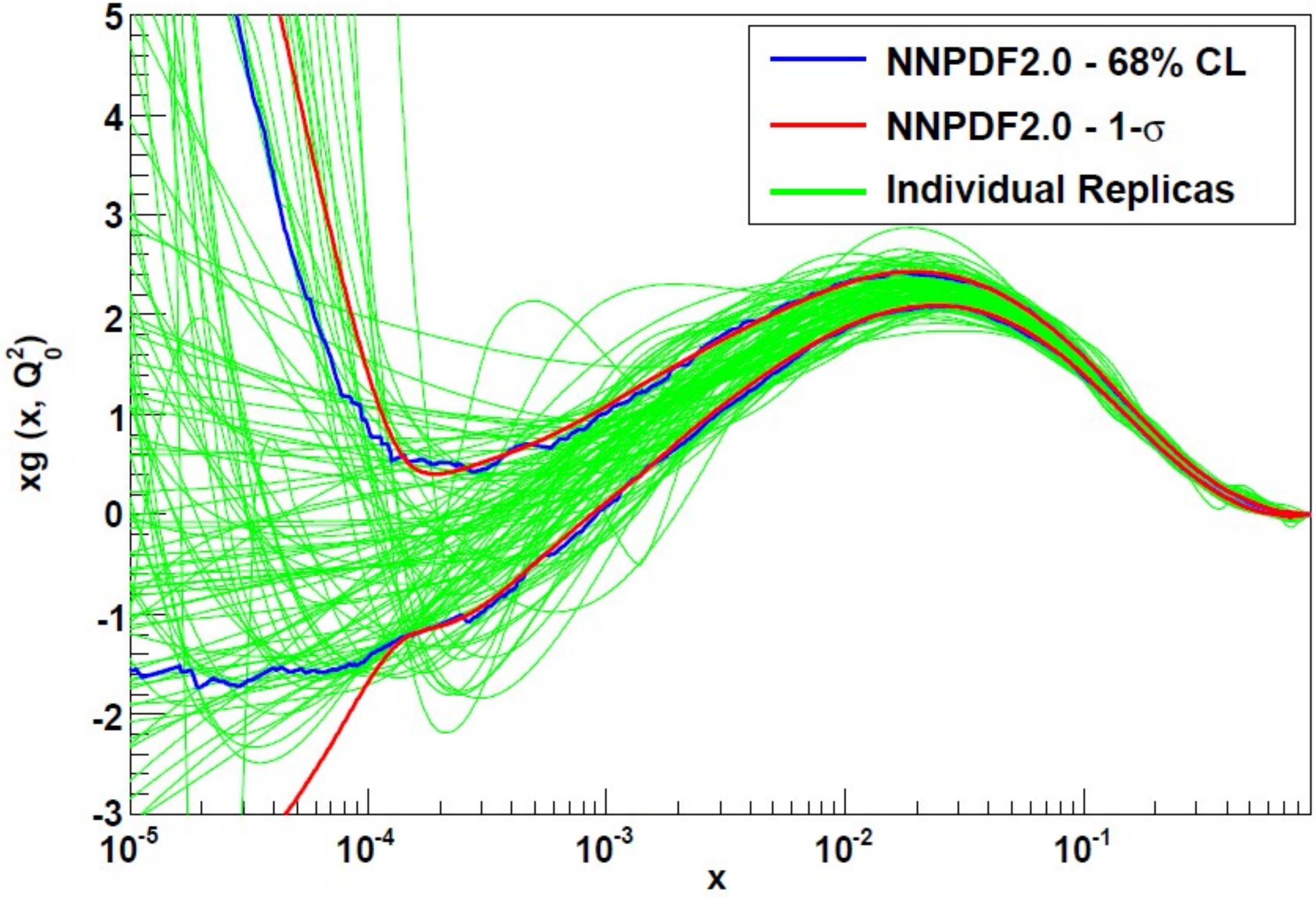
$$\mathrm{Var}[f] = \left\langle \left(f - \langle f \rangle\right)^2 \right\rangle = \frac{1}{N}\sum_{k=1}^{N} \left(f_k - f_0\right)^2$$

- Correlations: e.g.

$$\mathrm{Corr}[f, f'] = \langle \left(f - \langle f \rangle\right)\left(f' - \langle f' \rangle\right) \rangle = \frac{1}{N}\sum_{k=1}^{N} \left(f_k - f_0\right)\left(f'_k - f'_0\right)$$

- Confidence levels (e.g. interval with 68% replicas inside)
- etc, etc

1-sigma vs 68% CL



Non-Gaussianity at small-x (positivity constraints)

Example 2: DIS xsecs

DIS xsecs σ[f] depend linearly on the PDFs

- Central values:

$$E[\sigma] = \langle \sigma[f] \rangle = \frac{1}{N} \sum_{k=1}^{N} \sigma[f_k] = \sigma[f_0]$$

- Variances:

$$\begin{aligned} \mathrm{Var}[\sigma] &= \left\langle \left(\sigma[f] - \langle \sigma[f] \rangle \right)^2 \right\rangle \\ &= \frac{1}{N} \sum_{k=1}^{N} \left(\sigma[f_k] - \sigma[f_0] \right)^2 \end{aligned}$$

- etc, etc

σ[f] can be anything you like: str fn, red xsec, jet xsec, ...

Example 3: Hadronic xsecs

Hadronic xsecs σ[f,f] depend quadratically on the PDFs

- Central values:

$$E[\sigma_h] = \langle \sigma_h[f,f] \rangle = \frac{1}{N}\sum_{k=1}^{N} \sigma_h[f_k, f_k] \approx \sigma_h[f_0, f_0]$$

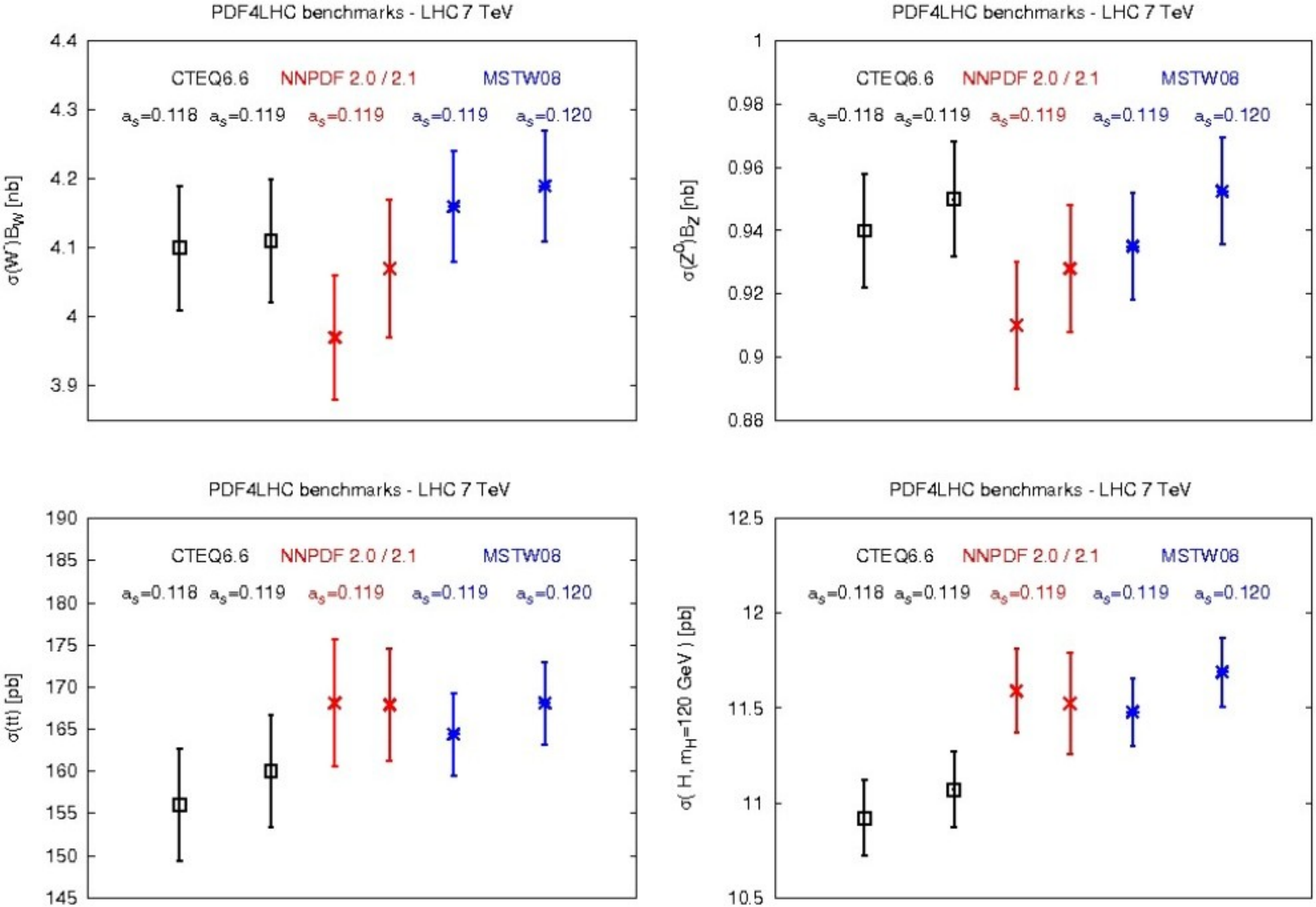
- Variances:

$$\begin{aligned}
\mathrm{Var}[\sigma_h] &= \left\langle \left(\sigma_h[f,f] - \langle \sigma_h[f,f] \rangle\right)^2 \right\rangle \\
&= \frac{1}{N}\sum_{k=1}^{N} \left(\sigma_h[f_k, f_k] - \frac{1}{N}\sum_{k=1}^{N} \sigma_h[f_k, f_k]\right)^2 \\
&\approx \frac{1}{N}\sum_{k=1}^{N} \left(\sigma_h[f_k, f_0] + \sigma_h[f_0, f_k] - 2\sigma_h[f_0, f_0]\right)^2 \qquad \text{if Gaussian}
\end{aligned}$$

The approximate expressions can be evaluated more quickly (smaller N)

LHC Standard Candles at 7TeV

MCFM



Three FAQs

Q: how many replicas N do I need?

A: depends on required accuracy: fluctuations fall as $1/\sqrt{N}$

typically use f_0 for central values, » 25-100 f_k for variances etc

Q: which replicas should I use?

A: any random selection! – all replicas are equally probable

Q: for hadronic xsecs, should I use the exact or approx formulae?

A: error from using approx is O(Var/E^2)

(so for typical 10% uncertainty, error is O(1%))

and variance formula neglects non-Gaussian errors

But: when you use MSTW or CTEQ, you do this all the time!

Reweighting

All replicas are equally probable (importance sampling):

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

Now add a new dataset $\{y_i, i = 1, \ldots n\}$

Q. What effect does this have on the PDFs?

A. The replicas are no longer equally probable: instead

$$\boxed{\langle \mathcal{O}[f] \rangle_{\mathrm{new}} = \frac{1}{N} \sum_{k=1}^{N} w_k \mathcal{O}[f_k]}$$

w_k are the 'weights': probability of replica f_k given new data:

No need to refit!

Giele & Kosower 1998

Calculating the weights

w_k are the probabilities of replica k given new data:

$$w_k \propto (\chi^2_k)^{\frac{1}{2}(n-1)} e^{-\frac{1}{2}\chi^2_k}$$

$$\frac{1}{N}\sum_{k=1}^{N} w_k = 1$$

$$\chi^2_k = \sum_{i,j=1}^{n} (y_i - y_i[f_k])\sigma^{-1}_{ij}(y_j - y_j[f_k]).$$

So... if you can plot the new data y_i and compare with the prediction $y_i[f_k]$, then you can compute w_k

Calculating the weights

w_k are the probabilities of replica k given new data:

$$w_k \propto (\chi^2_k)^{\frac{1}{2}(n-1)} e^{-\frac{1}{2}\chi^2_k} \qquad \frac{1}{N}\sum_{k=1}^{N} w_k = 1$$

$$\chi^2_k = \sum_{i,j=1}^{n} (y_i - y_i[f_k])\sigma^{-1}_{ij}(y_j - y_j[f_k]).$$

So... if you can plot the new data y_i and compare with the prediction $y_i[f_k]$, then you can compute w_k

You can do this at home!

Loss of efficiency: replicas no longer have equal probability

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

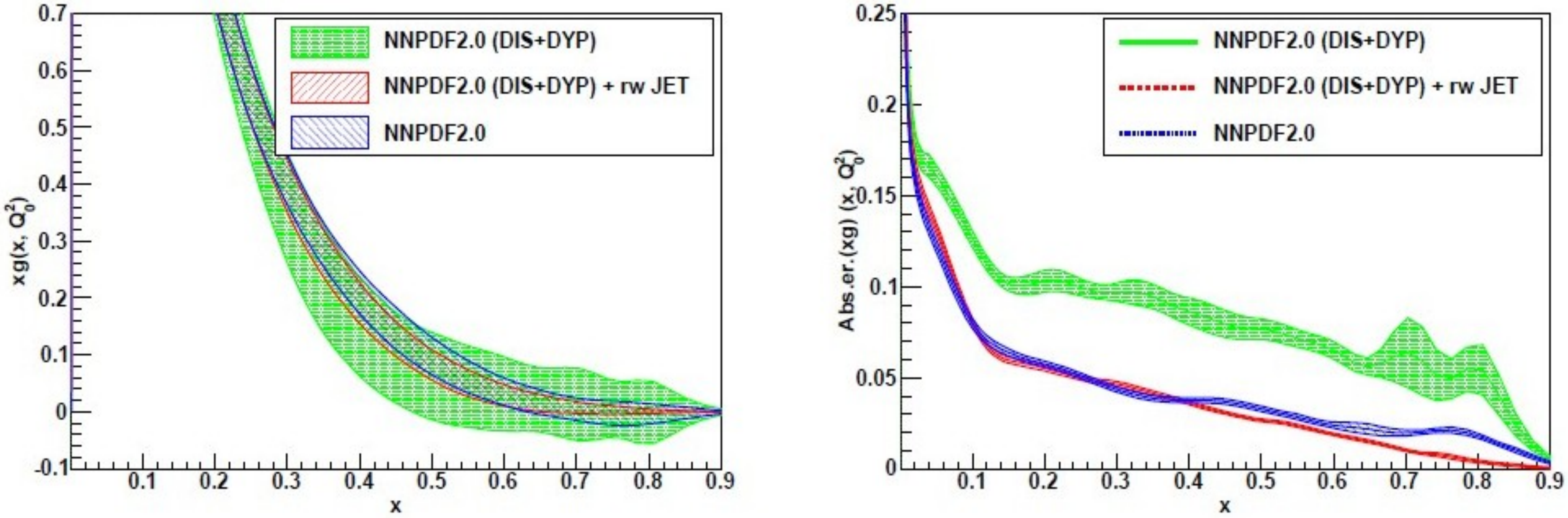
Shannon entropy

N_{eff}/N: gives measure of impact of new data

Does it work?

Example:

1) take fit of DIS+DY data only
2) add (CDF+D0) inclusive jet data by reweighting
3) compare to result of fit using all the data DIS+DY+jet



Impact of jet data: with N=1000, have N_{eff}=332 left: substantial

Note that if N_{eff} had been too small (say below 100), would need to refit (or start with more replicas)

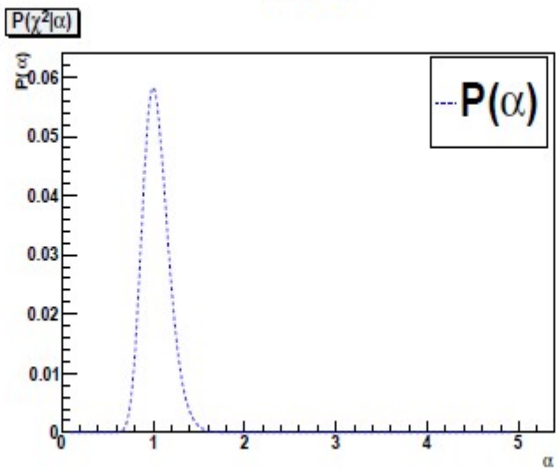
Are the new data consistent?

Rescale errors in new data: $\chi^2 \to \chi^2/\alpha$

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

CDF and D0

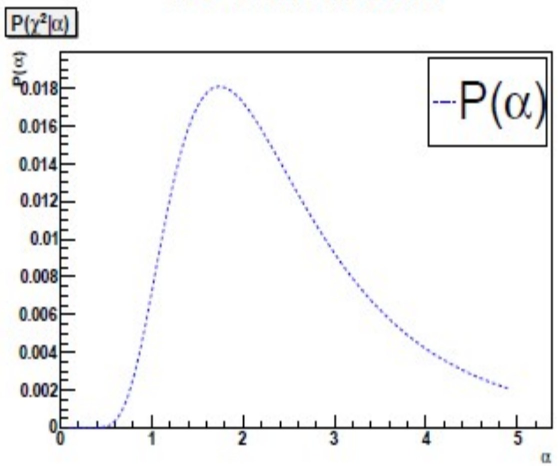
JETS



Yes

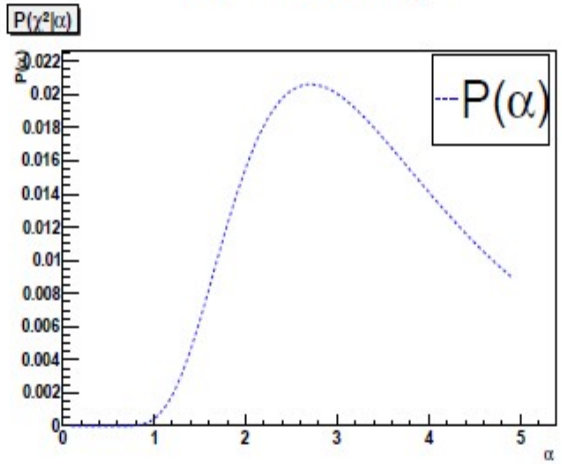
D0 W lepton asymmetry

D0 e INCLUSIVE



OK...

D0 e HIGH E_T

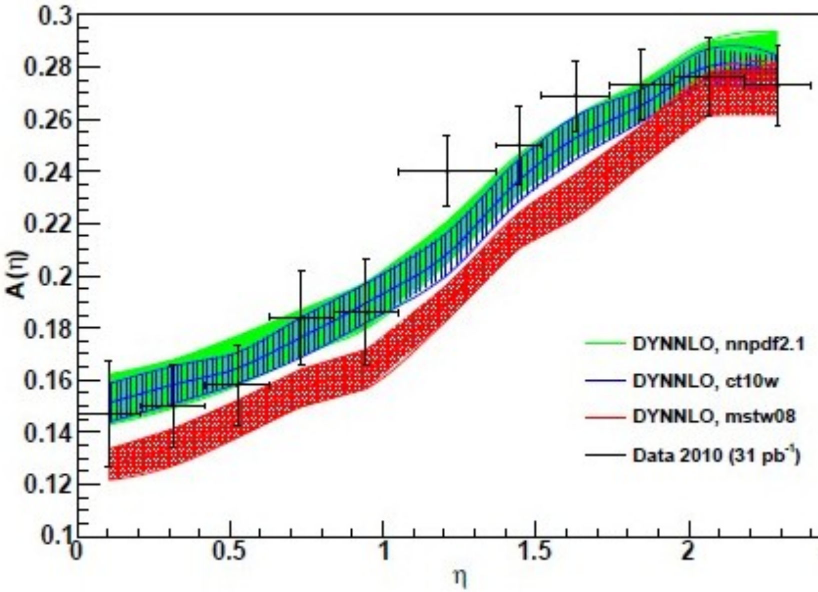


No!

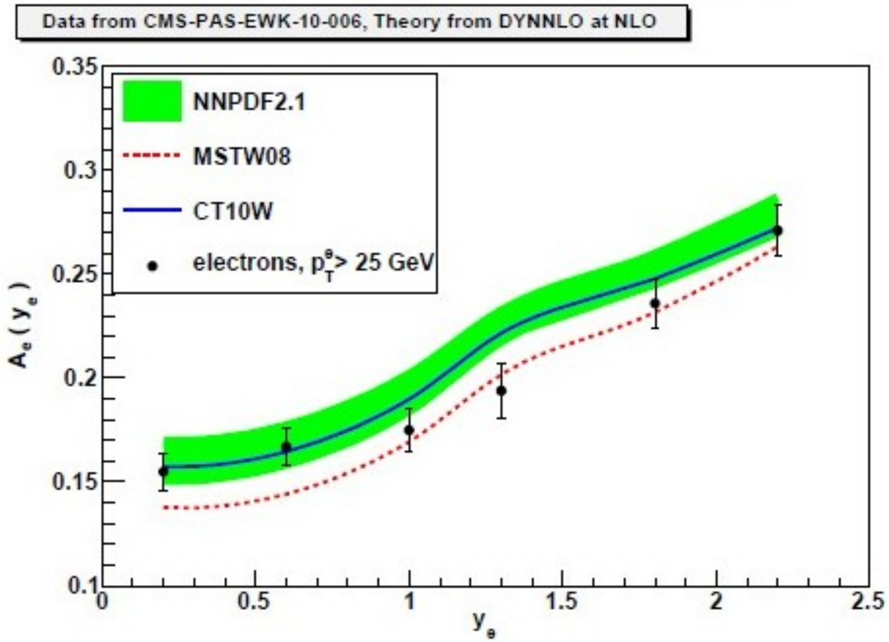
Mar 2011

W-lepton asymmetry

ATLAS



CMS

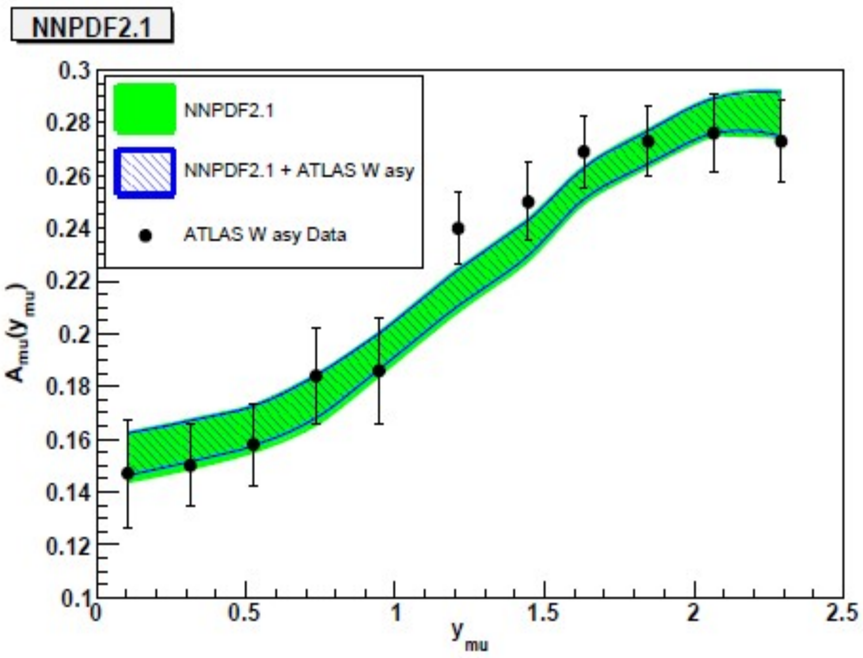


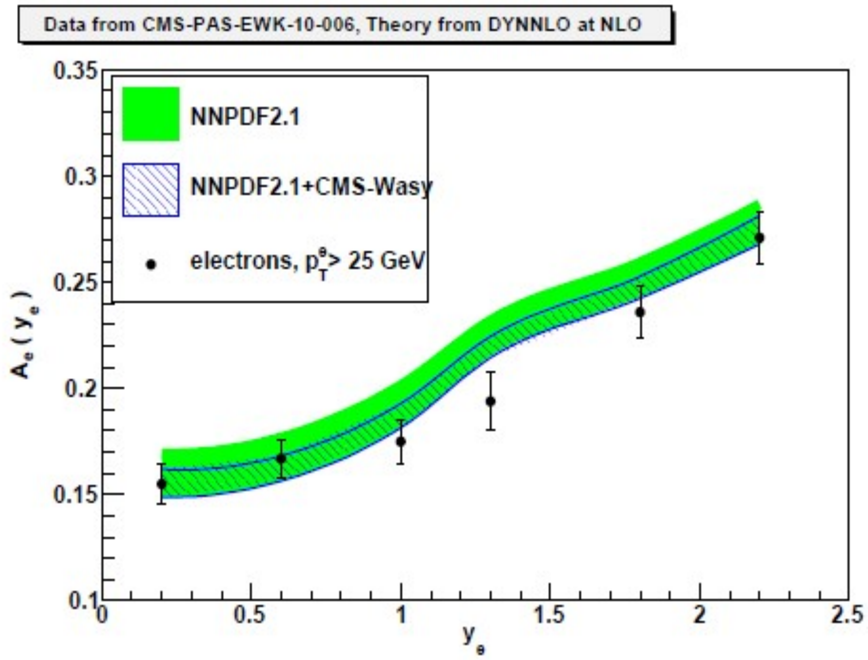
$\chi^2/N_{\rm dat}$ (EL)	ATLAS	CMS $p_T^l \geq 25$ GeV
NNPDF2.1	0.68	1.8
MSTW08	3.24	1.8
CT10W	0.81	1.2

Mar 2011

W-lepton asymmetry

Reweighting





Already good!

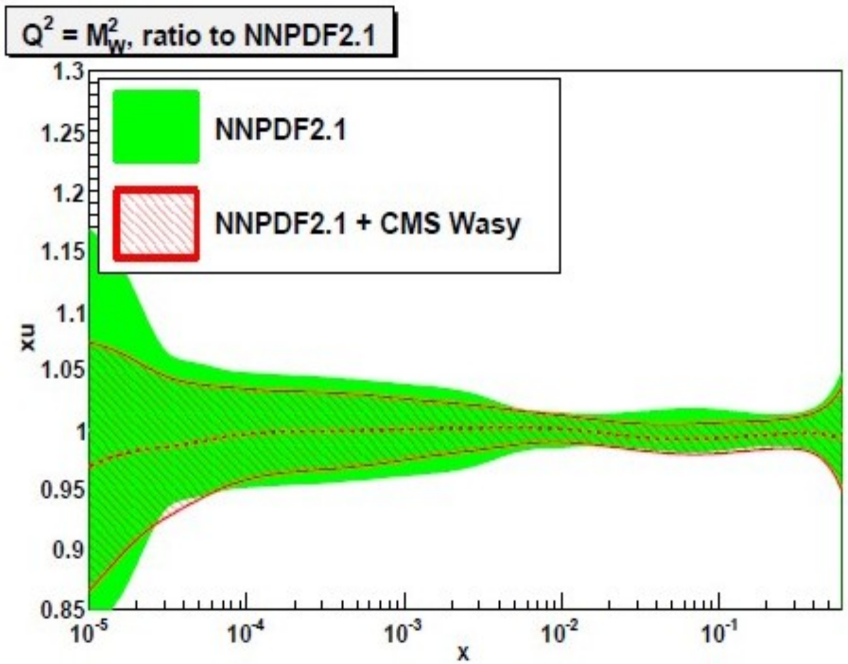
Now much better!

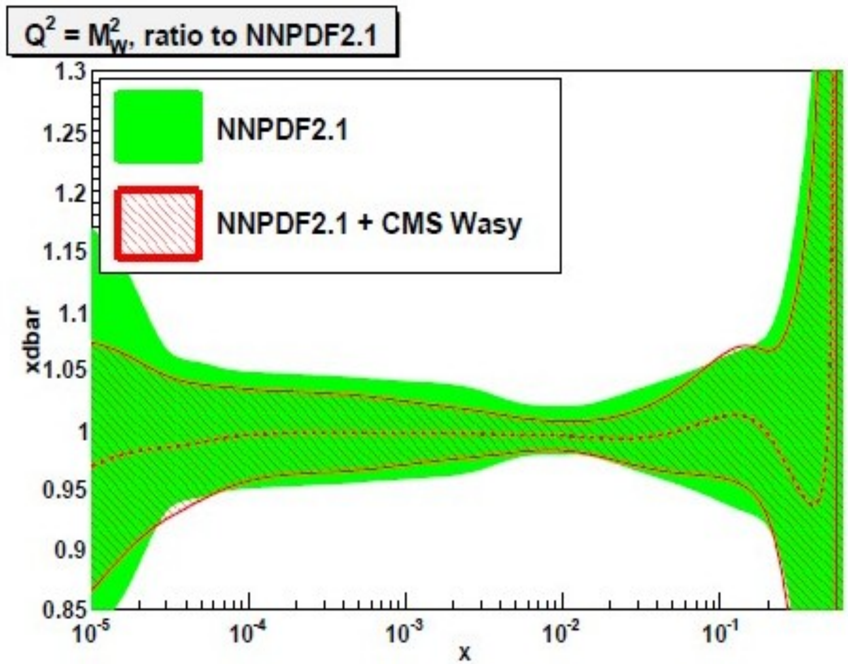
χ^2/N_{dat} (EL)	ATLAS	CMS $p_T^l \geq 25$ GeV
NNPDF2.1	0.68	1.8
NNPDF2.1+EACH DATASET	0.63	1.03

Mar 2011

W-lepton asymmetry

$$xu(x, M_W^2) \qquad\qquad x\bar{d}(x, M_W^2)$$





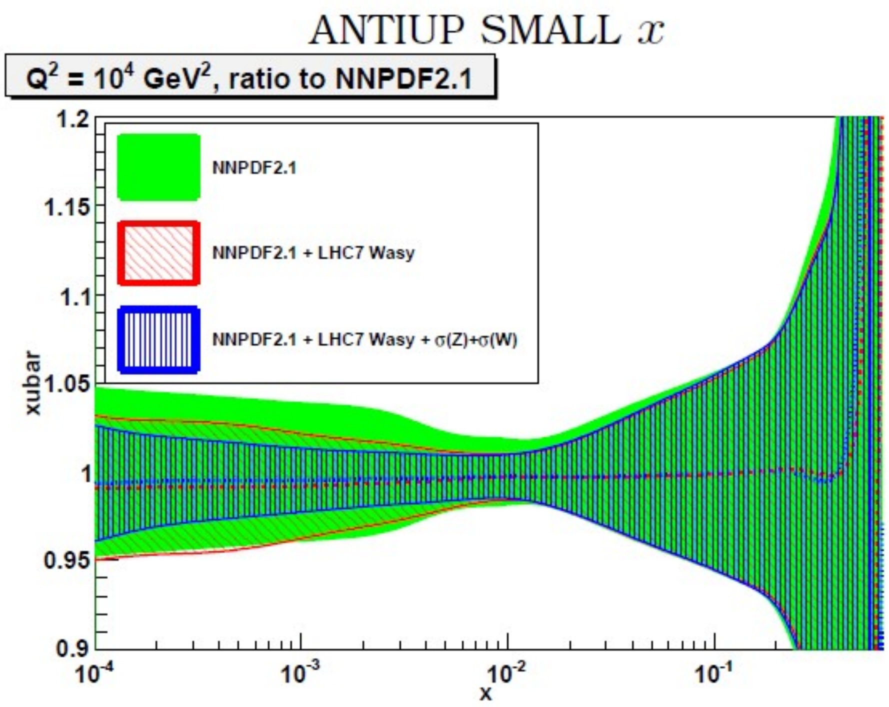
First ever use of LHC data to constrain PDFs

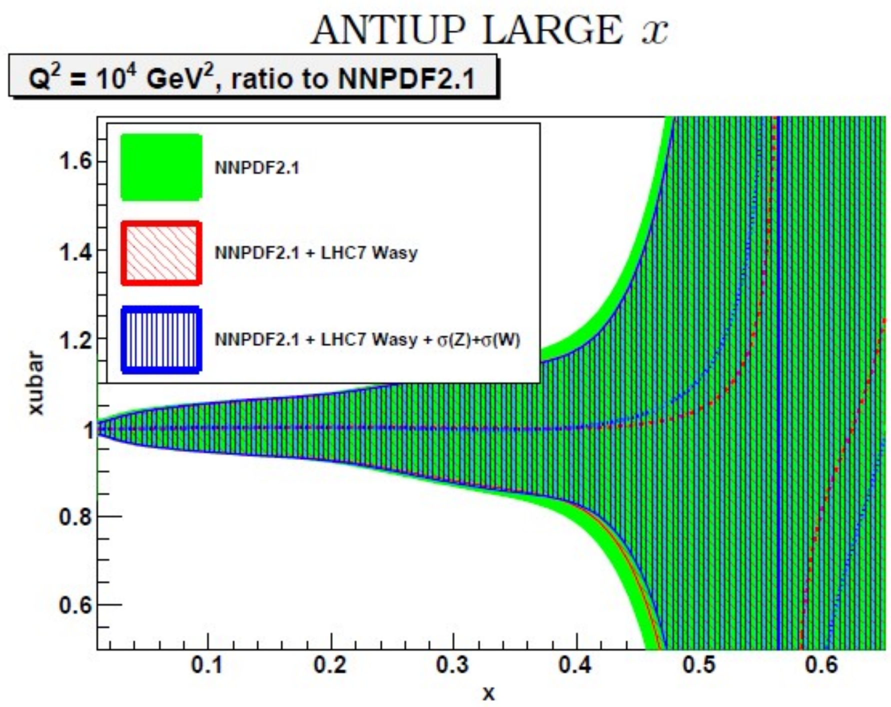
W-lepton asymmetry

ATLAS LHC7: future prospects

W lepton asymmetry measured to » 5% (kinematics courtesy A. Glazov)

W & Z total xsecs measured to 2%





Improvement seen for all flavours and antiflavours

WISHLIST ⇒ ROADMAP

- LHC CAN PROVIDE US PRECISION INFORMATION ON PDFS
- TOWARDS A "COLLIDER ONLY" HERA+LHC PDF FIT
  (TEVATRON DATA MIGHT BE SUPERFLUOUS)
  - MEDIUM & LARGE x GLUON
    - PROMPT PHOTONS AVAILABLE
    - (PRECISION) JETS IN PROGRESS
  - LIGHT FLAVORS AT MEDIUM @ SMALL x, FLAVOR SEPARATION @ SMALL x
    - LOW–MASS DRELL-YAN PRELIM.
    - Z RAPIDITY DISTRIBUTIONS PRELIM.
    - W ASYMMETRIES AVAILABLE
  - STRANGENES & HEAVY FLAVORS
    - STRANGENESS ⇒ $W + c$ FUTURE?
    - CHARM ⇒ $Z + c$, $\gamma + c$ FUTURE?
    - BOTTOM $Z + b$ IN PROGRESS
- PRECISION "LEP" PHYSICS POSSIBLE @ LHC!
  - NEW PHYSICS FROM EW SECTOR
  - NEW QCD EFFECTS

Summary & Outlook

- NNPDF works :

  1.0 (DIS), 1.2 (strange), 2.0 (global), 2.1 (HQ)……

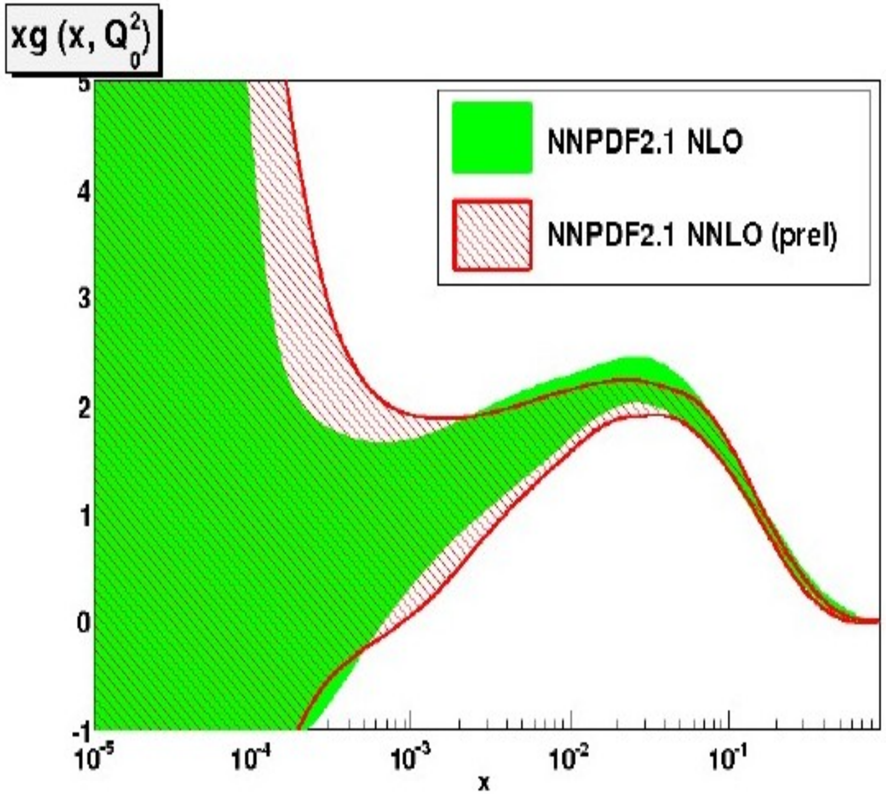
  See for yourself: http://projects.hepforge.org/lhapdf
- Reweighting:

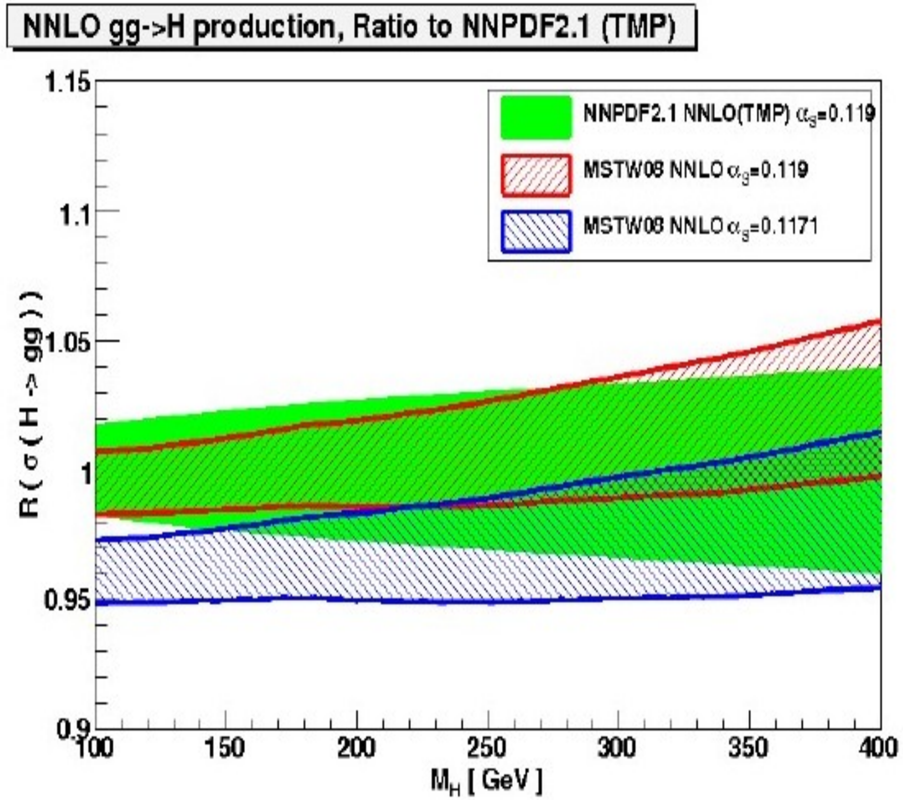
  You can update NNPDFs yourself: new tool
- For the future:

  LO, NNLO (soon), resummation, etc, etc,

  Lots and lots of new LHC data!

Preliminary NNPDF2.1 NNLO
(with FONLL-C heavy quarks)





Broad agreement with MSTW for Higgs...