Reweighting

- NNPDF Users Guide
- Reweighting
- Unweighting
- NNPDF2.2

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

St Andrews Aug 2011

Monte Carlo PDFs (eg NNPDF)

Giele & Kosower 1998

Forte & Latorre 2002

- Choose a very flexible functional form for each PDF:
  (eg a neural network: $\sim$ 250 params)
- Generate data replicas ($\sim$ 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: $\sim$ 250 params)
- Generate data replicas ($\sim$ 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 NLO DIS+DY+J: NNPDF2.0
- 2010: Reweighting (W-ev asymmetry) arXiv:1012.0836
- 2011: Global NLO + HQ: NNPDF2.1 NLO arXiv:1101.1300
- 2011: NNPDF2.1 LO and NNLO arXiv:1107.2652
- 2011: First fit with LHC data: NNPDF2.2 arXiv:1108.1758

Major software development project

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$\cdots$ 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:

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

“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:

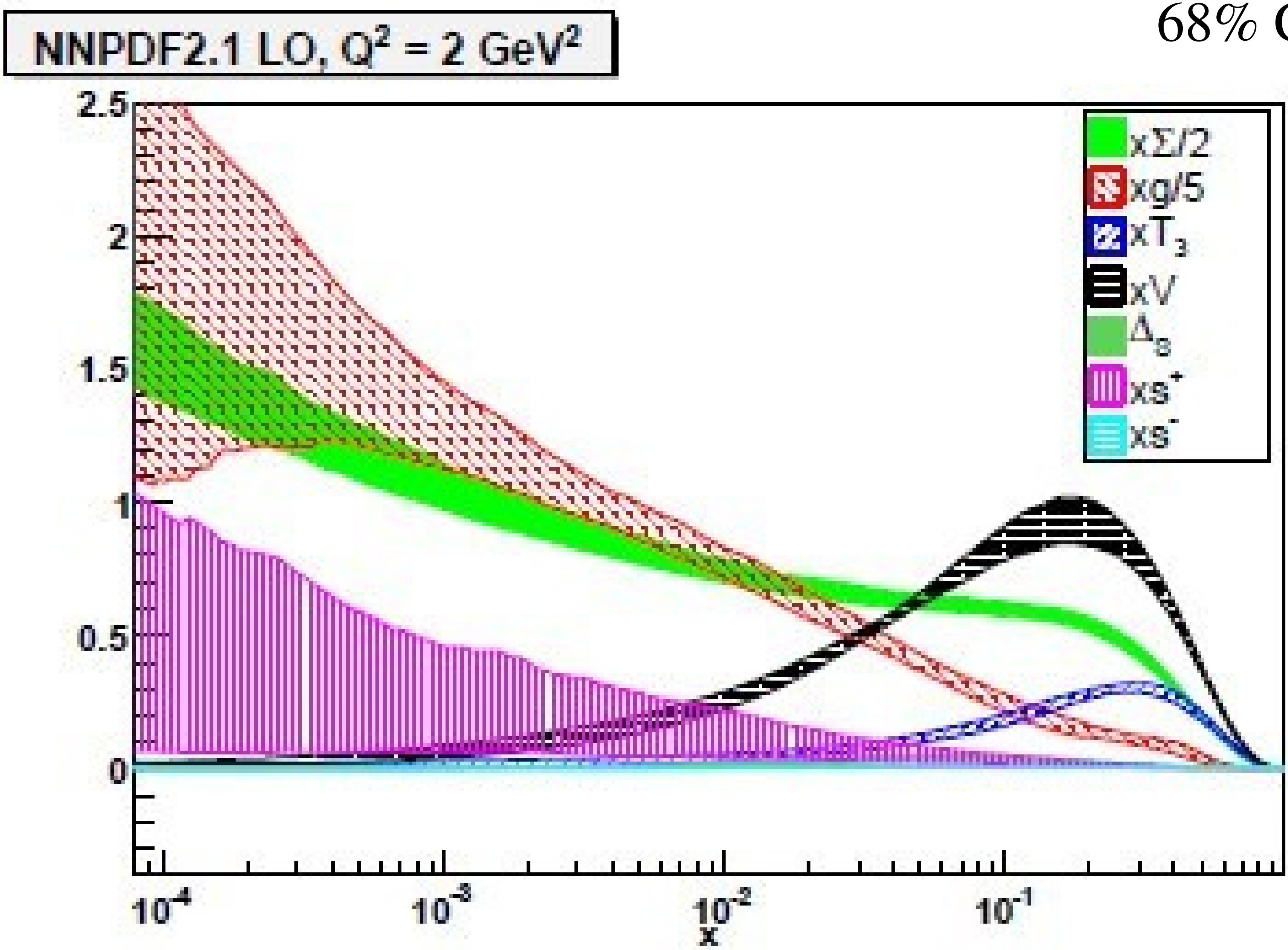
$$\mathrm{Var}[f] = \left\langle \left(f - \langle f \rangle \right)^2 \right\rangle = \frac{1}{N} \sum_{k=1}^{N} \left(f_k - f_0 \right)^2$$

- Correlations: e.g.

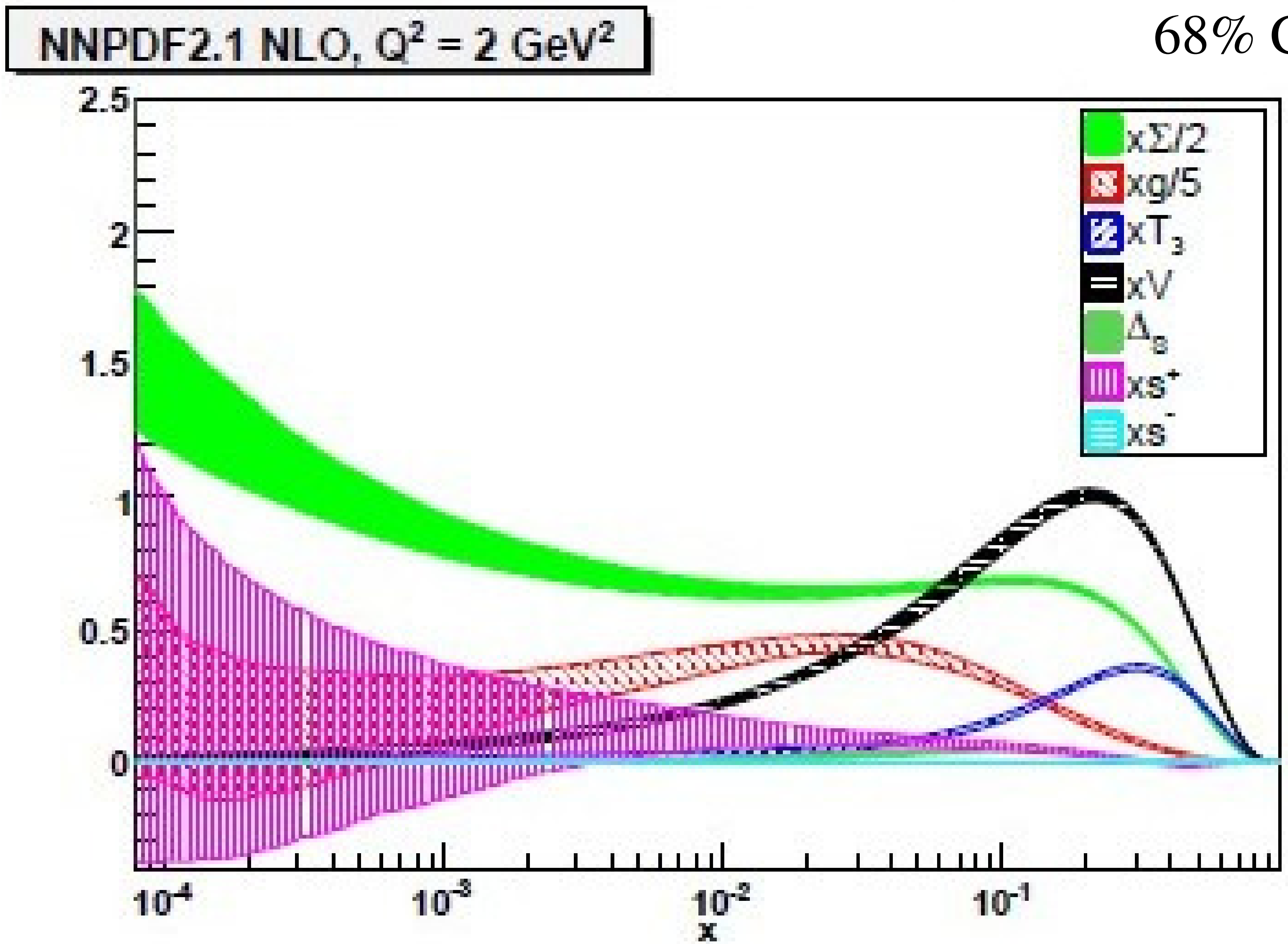
$$\mathrm{Corr}[f, f'] = \langle (f - \langle f \rangle) (f' - \langle f' \rangle) \rangle = \frac{1}{N} \sum_{k=1}^{N} (f_k - f_0) (f'_k - f'_0)$$

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

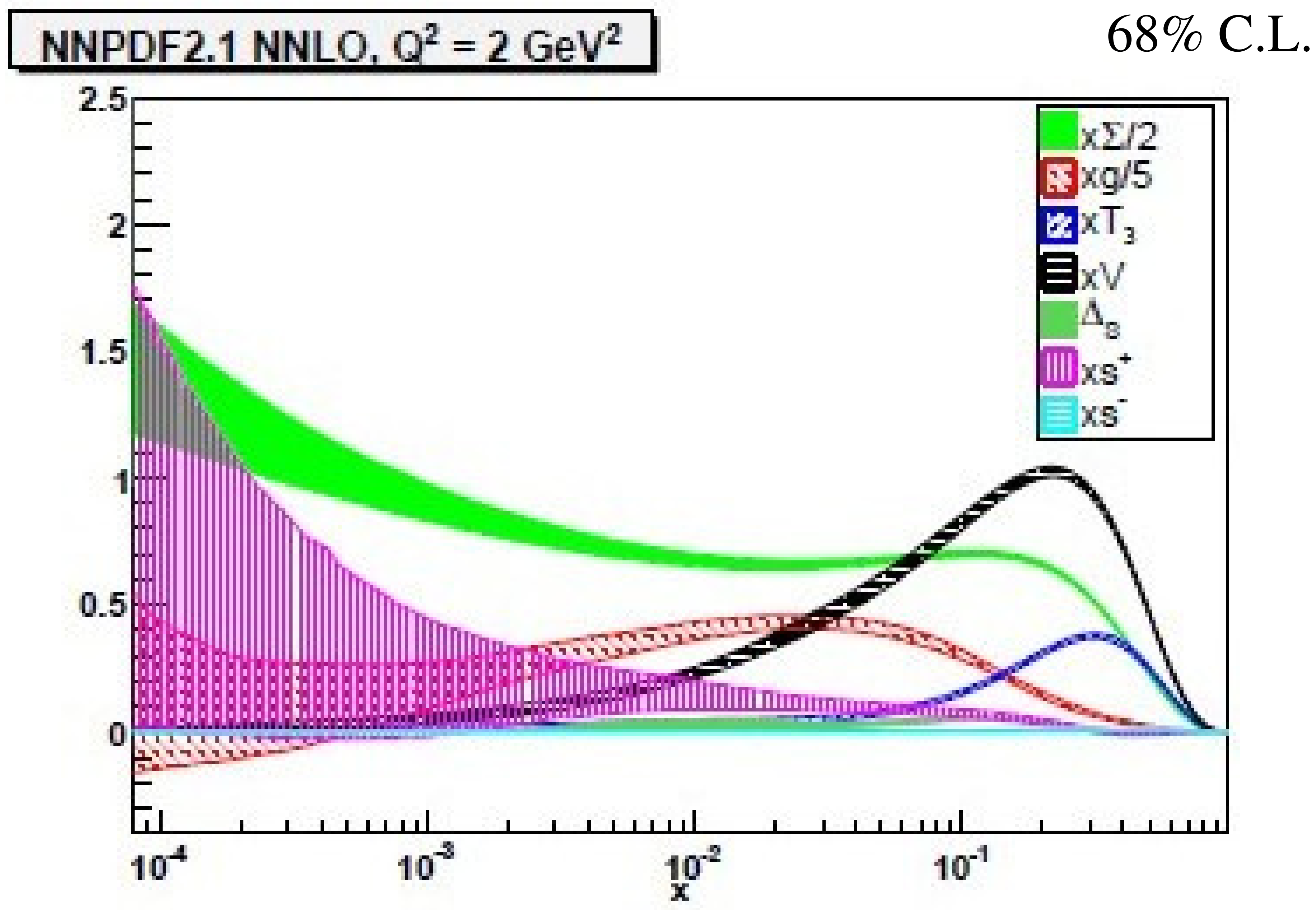
68% C.L.


NNPDF2.1 LO, Q^2 = 2 GeV^2
$x\Sigma/2$
$xg/5$
xT_3
xV
Δ_S
xs^+
xs^-
2.5
2
1.5
1
0.5
0
10^{-4}
10^{-3}
10^{-2}
10^{-1}
x

68% C.L.



NNPDF2.1 NLO, Q^2 = 2 GeV^2
2.5
2
1.5
1
0.5
0
10^{-4}
10^{-3}
10^{-2}
10^{-1}
x
$x\Sigma/2$
$xg/5$
xT_3
xV
Δ_S
xs^+
xs^-


NNPDF2.1 NNLO, Q^2 = 2 GeV^2
68% C.L.
$x\Sigma/2$
$xg/5$
xT_3
xV
Δ_S
xs^+
xs^-
2.5
2
1.5
1
0.5
0
10^{-4}
10^{-3}
10^{-2}
10^{-1}
x

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

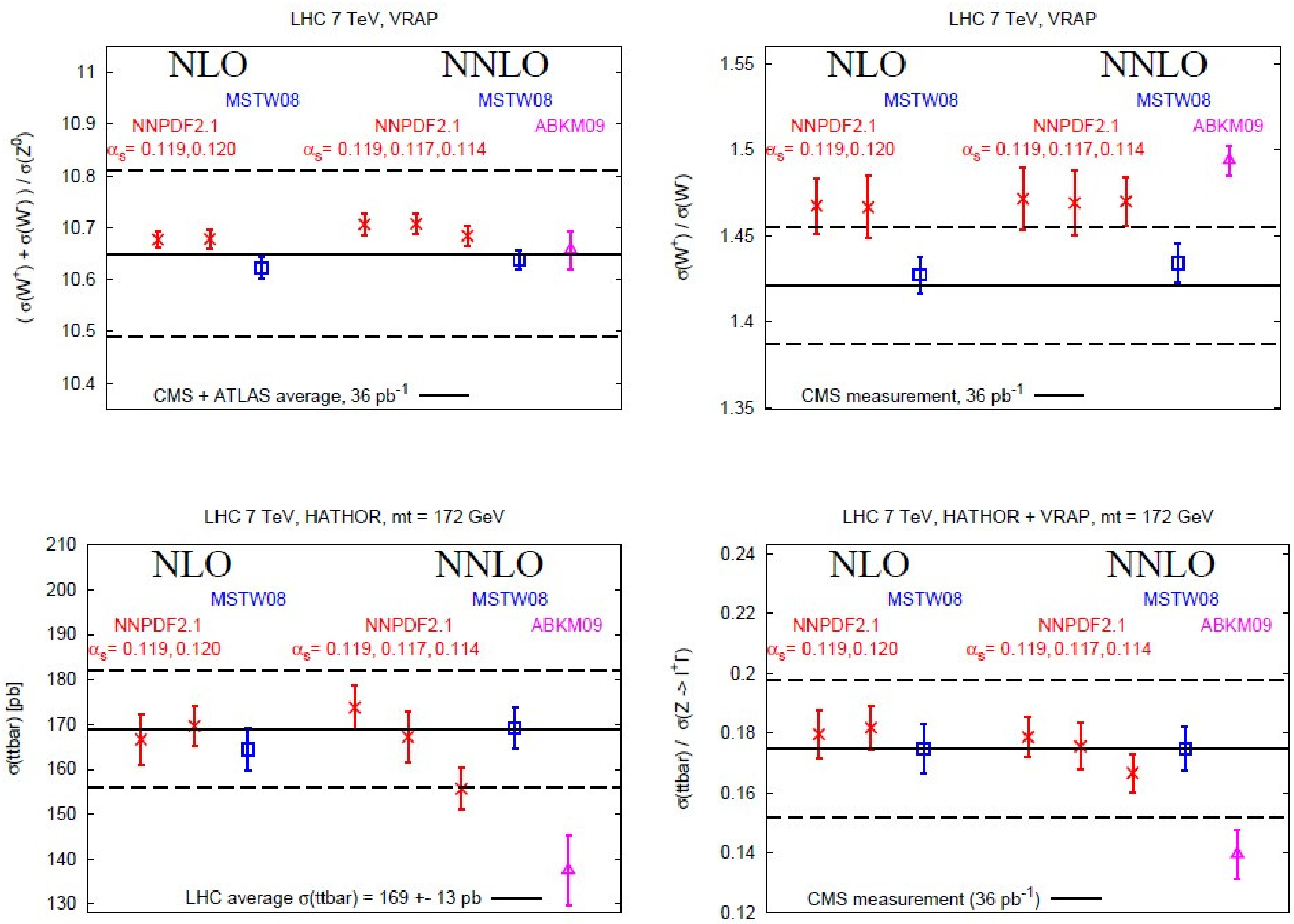
- Central values:

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

- Variances:

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

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



LHC 7 TeV, VRAP
NLO
NNLO
MSTW08
NNPDF2.1
ABKM09
α_s= 0.119,0.120
α_s= 0.119,0.117,0.114
(σ(W+) + σ(W-)) / σ(Z0)
CMS + ATLAS average, 36 pb-1
σ(W+) / σ(W-)
CMS measurement, 36 pb-1
LHC 7 TeV, HATHOR, mt = 172 GeV
σ(ttbar) [pb]
LHC average σ(ttbar) = 169 +- 13 pb
LHC 7 TeV, HATHOR + VRAP, mt = 172 GeV
σ(ttbar) / σ(Z -> l+l-)
CMS measurement (36 pb-1)

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, $\sim$ 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

Reweighting

All replicas are equally probable (importance sampling):

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

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

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

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 n 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 new data y_i and compare with theory $y_i[f_k]$, then you can compute w_k and thus effect of data on PDFs

DIY PDFs

Loss of efficiency: replicas no longer have equal probability

$$N_{\text{eff}} \equiv \exp\left[-\sum_{k=1}^{N} p_k \ln p_k\right] \qquad p_k \equiv w_k/N$$

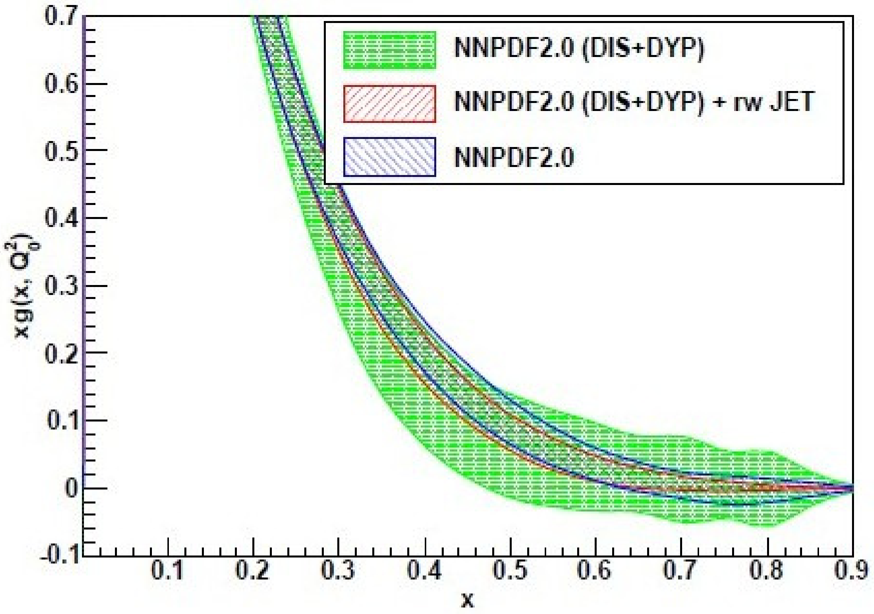
cf Shannon entropy

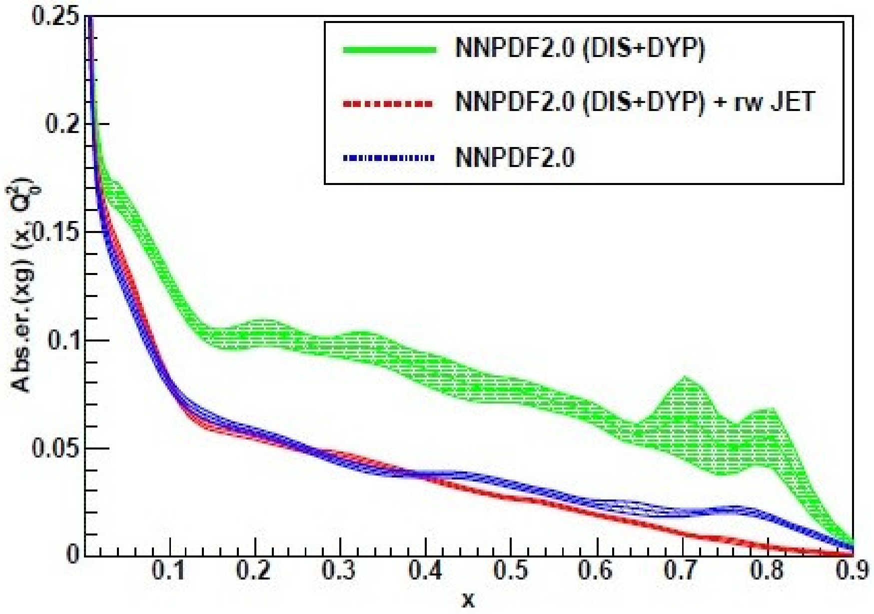
N_{eff}/N: gives measure of impact of new data

Does reweighting 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

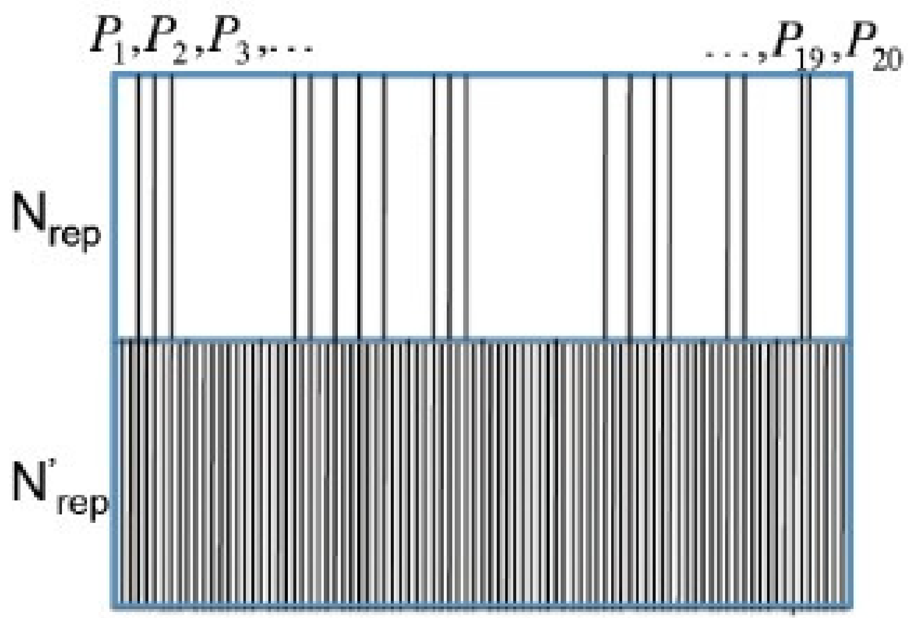
Note that if N_{eff} had been too small, would need to refit (or start with more replicas)

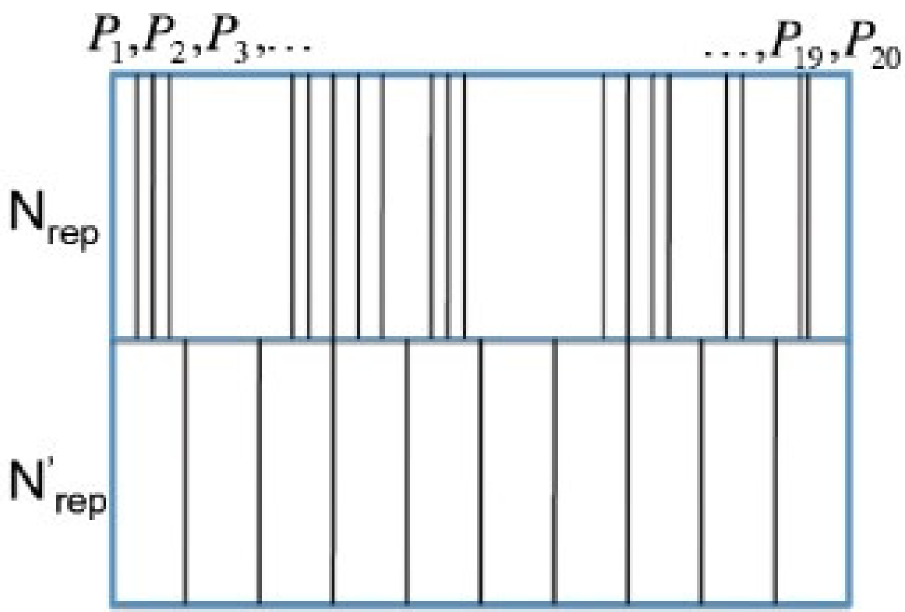
Unweighting

- Choose replicas according to weights, treated as probabilities
- Get subset of equally probable replicas equivalent to weighted set
- Treat just the same as a fitted set!
- In practice:

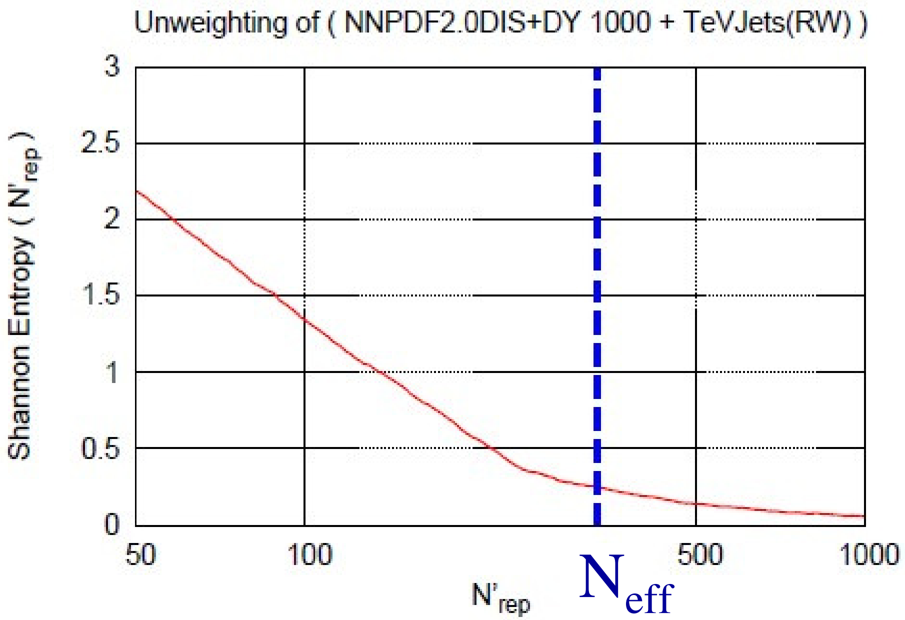
$$N'_{\rm rep} \lesssim N_{\rm eff} < N_{\rm rep}$$

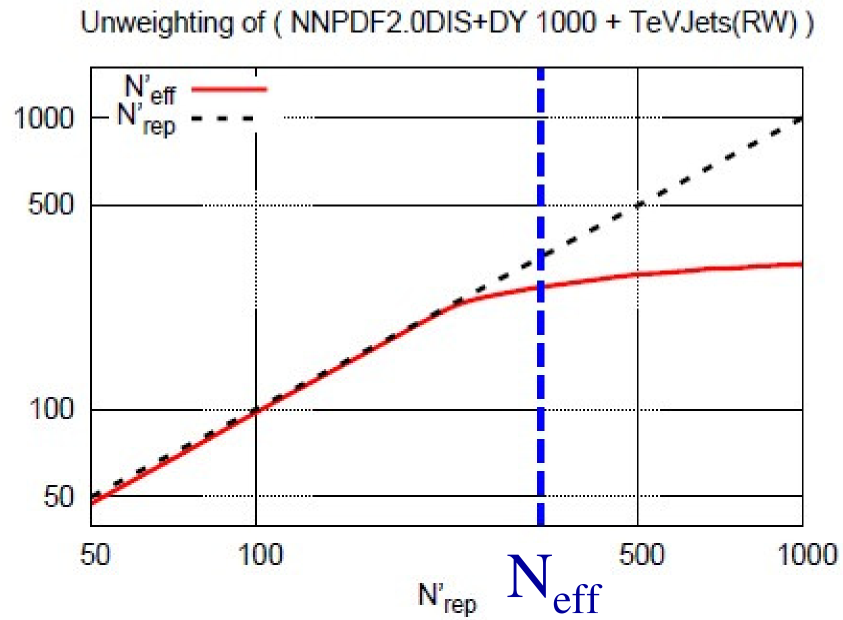
(so need N_{rep} large: typically $N_{rep} \sim 1000$, $N'_{rep} \sim 100$)





Does unweighting work?





$$H_R(N'_{\rm rep}) = \sum_{k=1}^{N_{\rm rep}} p'_k \ln \frac{p'_k}{p_k},$$

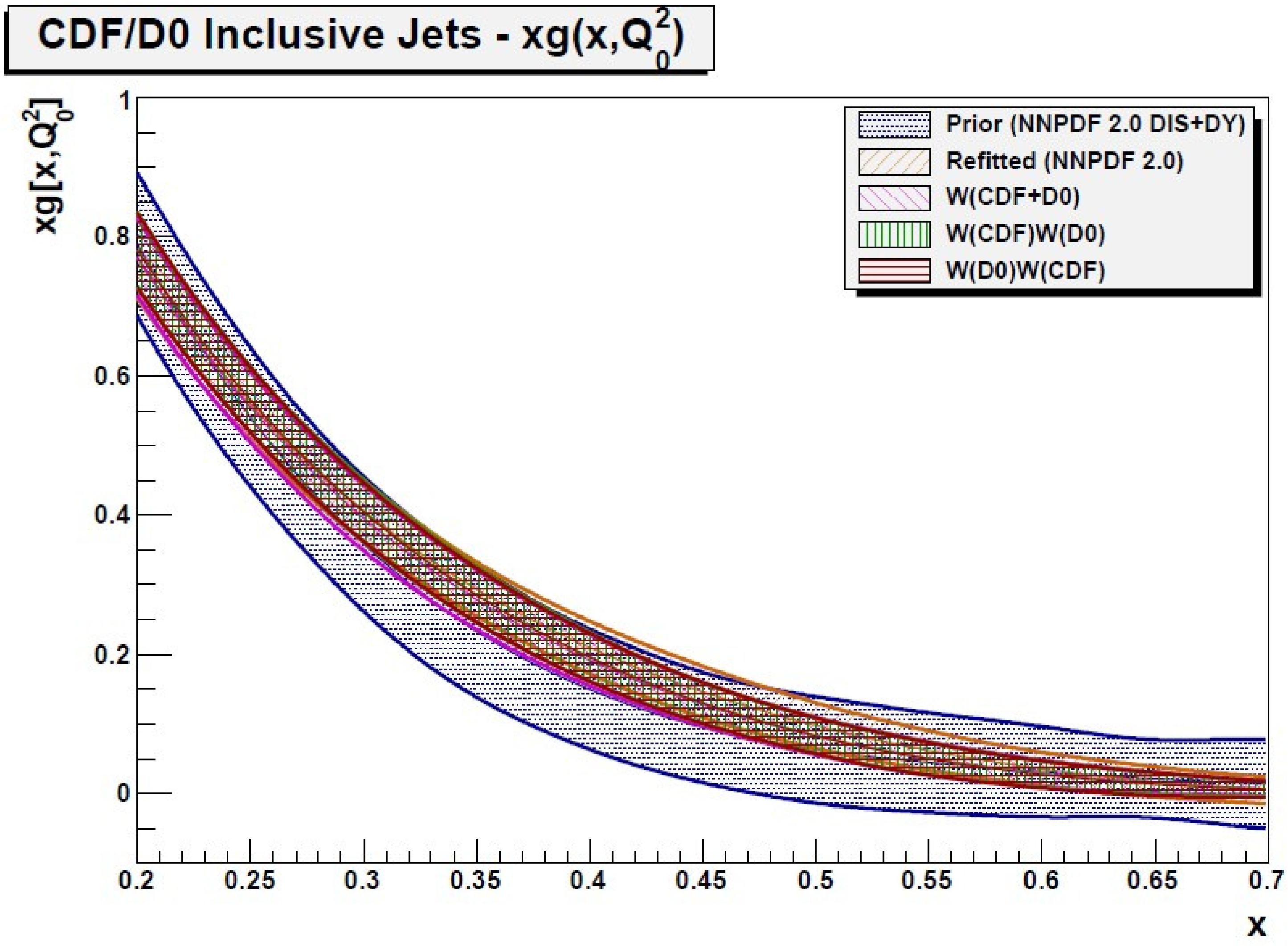
Relative entropy of unweighted and weighted

$$N'_{\rm eff} \equiv \exp\left[- \sum_{k=1}^{N_{\rm rep}} p'_k \ln p'_k \right]$$

Number of effective replicas in unweighted set

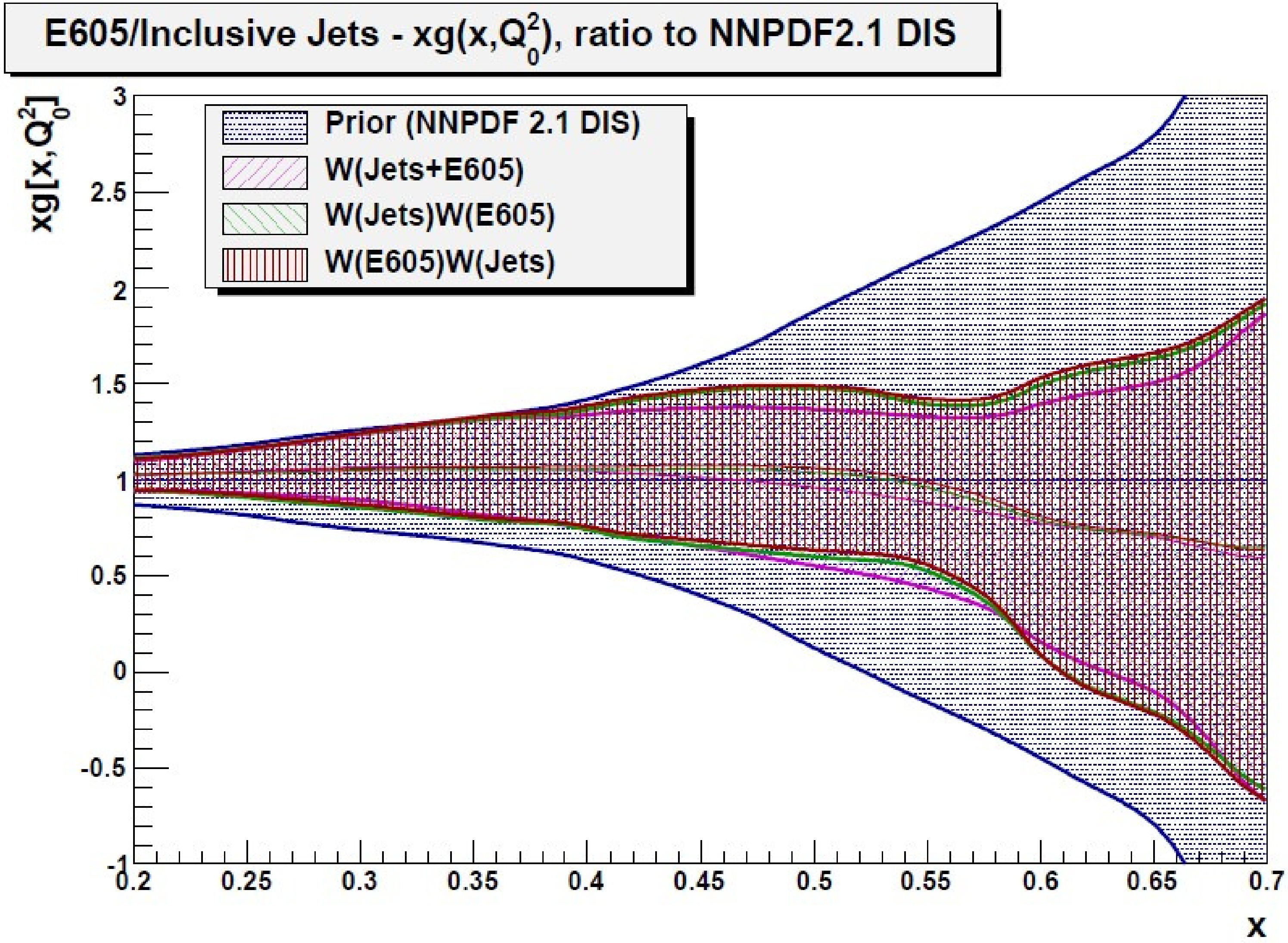
Consistency: two new sets, or one?

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



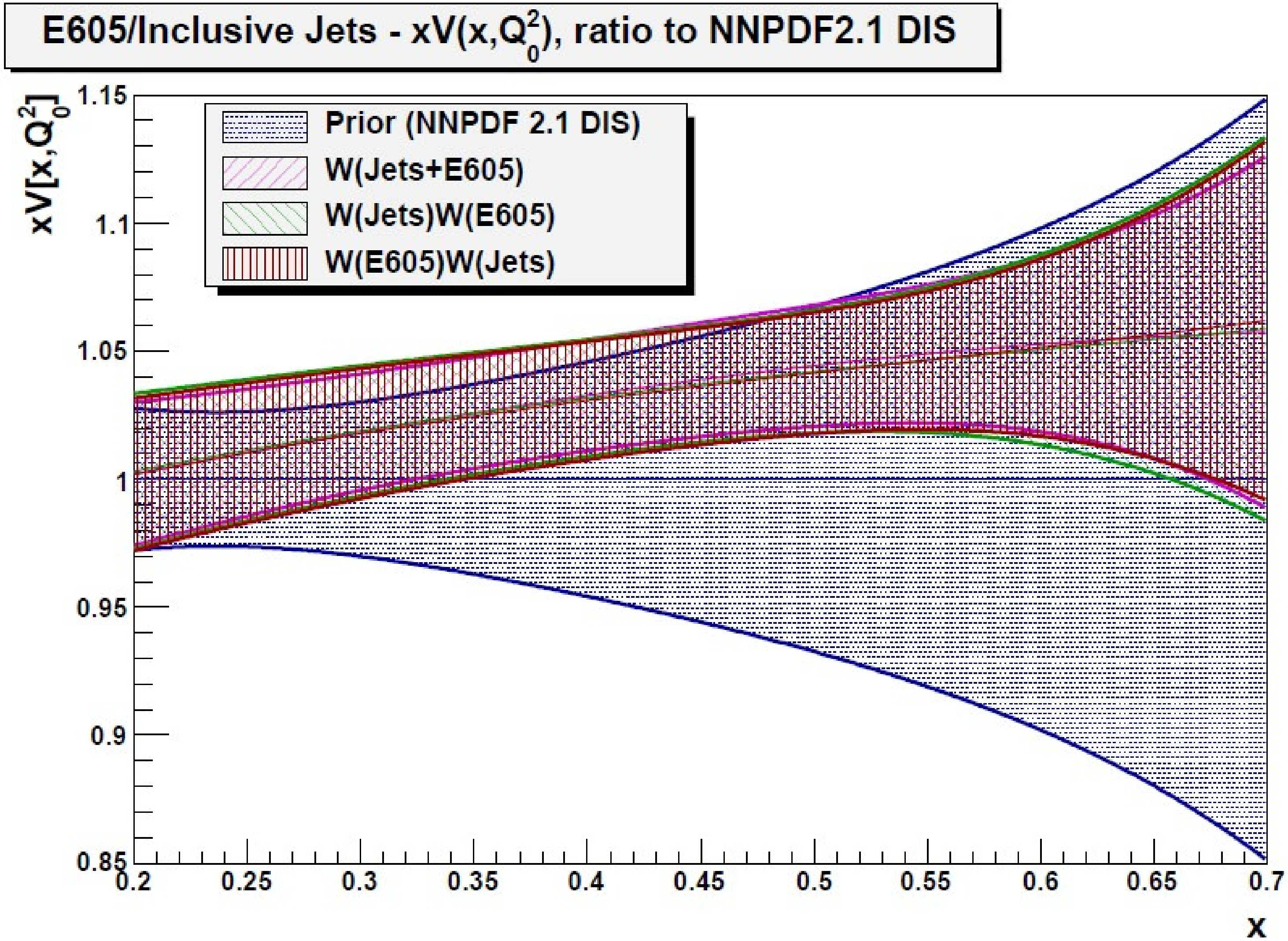
Consistency: two new sets, or one?

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



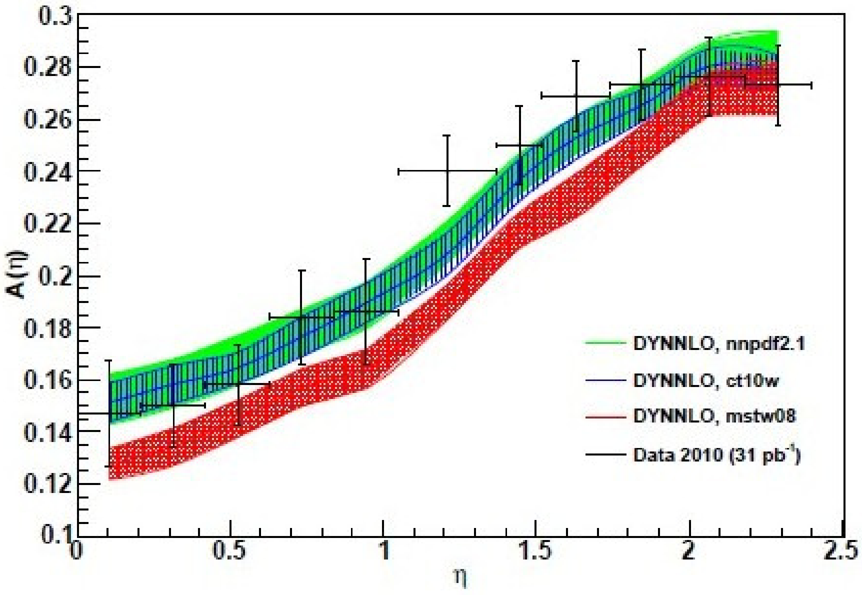
Consistency: two new sets, or one?

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

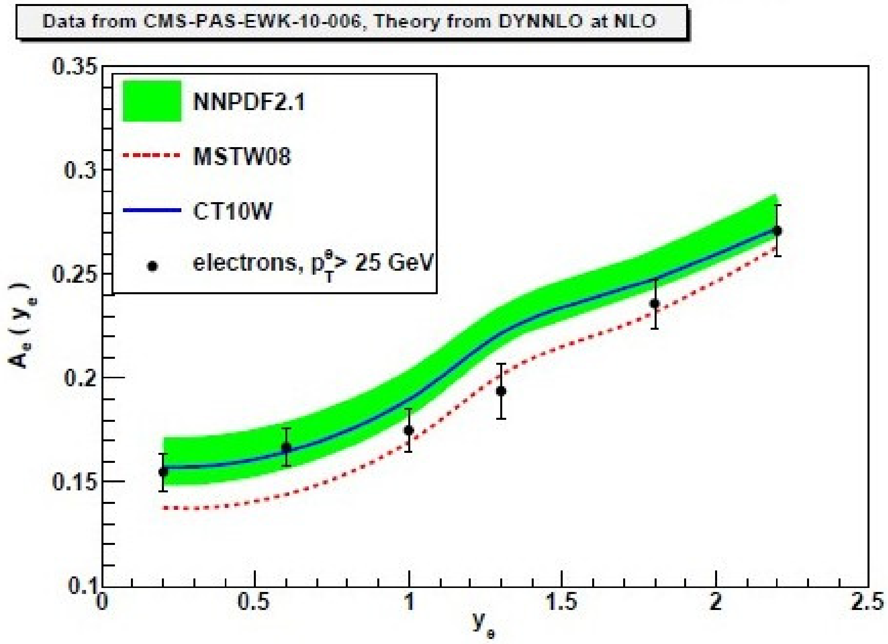


W-lepton asymmetry

ATLAS

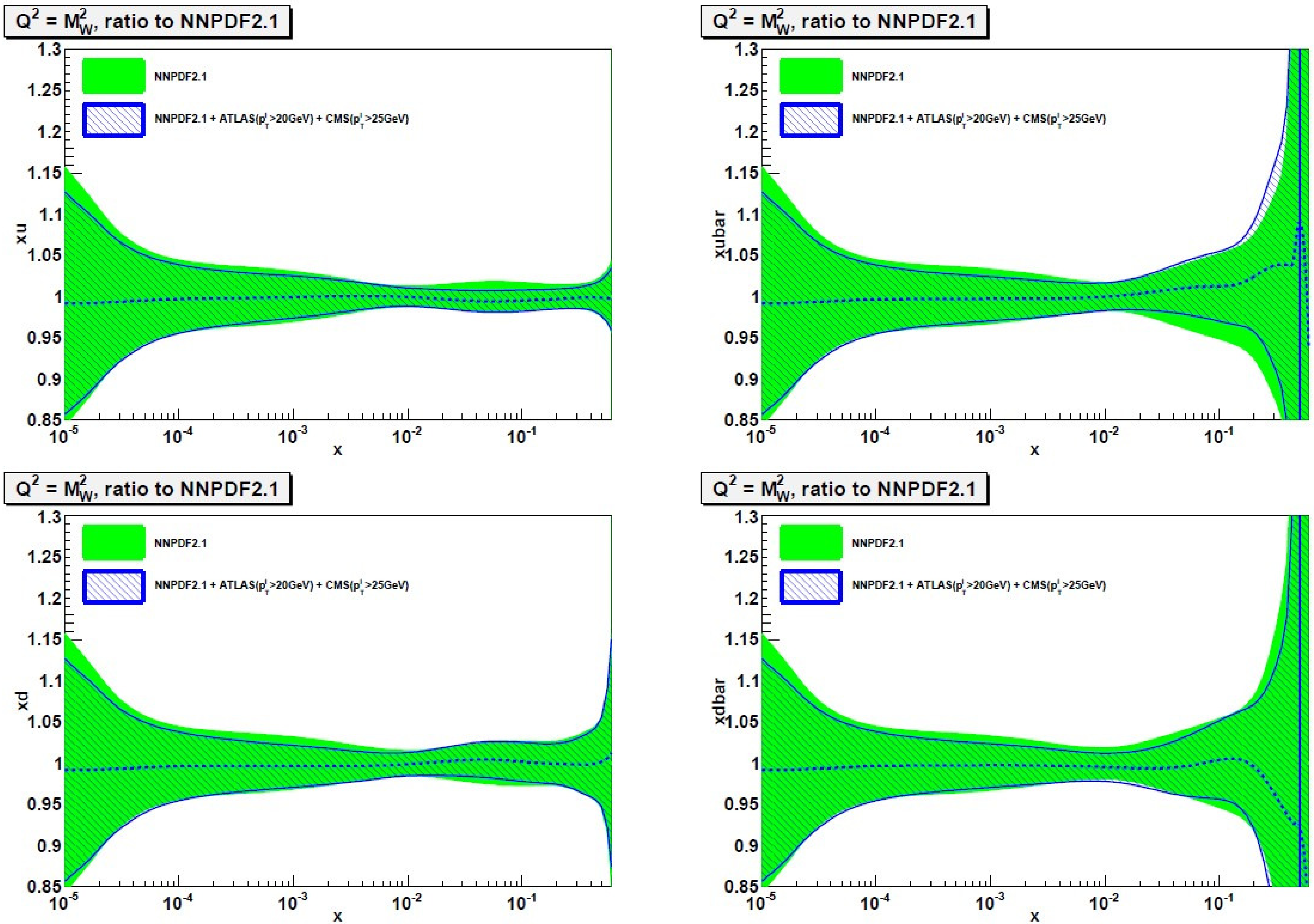


CMS



	$N_{\rm dat}$	NNPDF2.1	CT10	MSTW08
ATLAS($31 {\rm pb}^{-1}$)	11	0.76	0.77	3.32
CMS($36 {\rm pb}^{-1}$) electron $p_T > 25$ GeV	6	1.83	1.19	1.70
CMS($36 {\rm pb}^{-1}$) muon $p_T > 25$ GeV	6	1.24	0.73	0.77

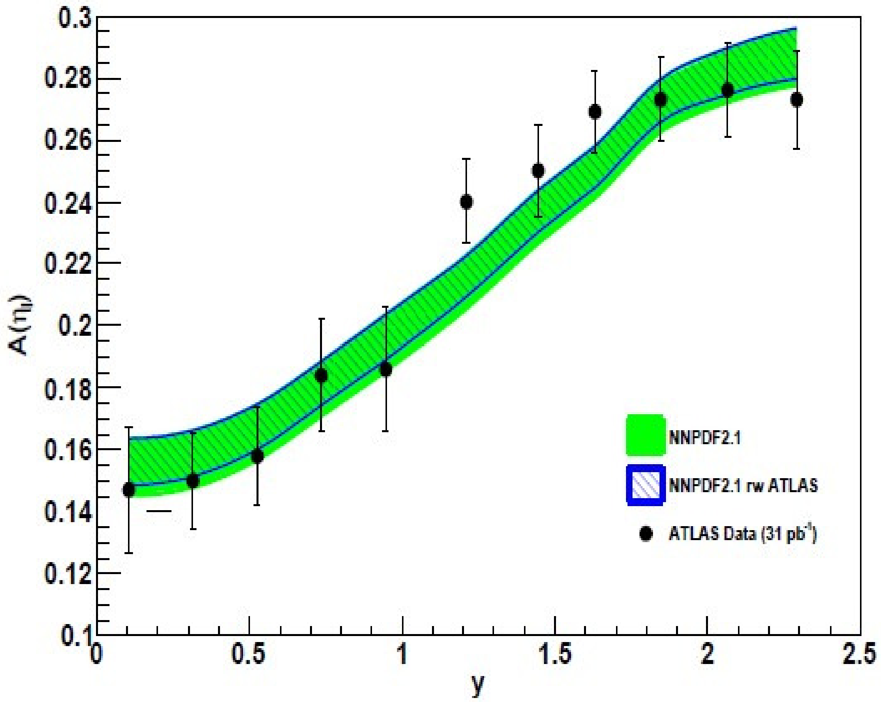
W-lepton asymmetry

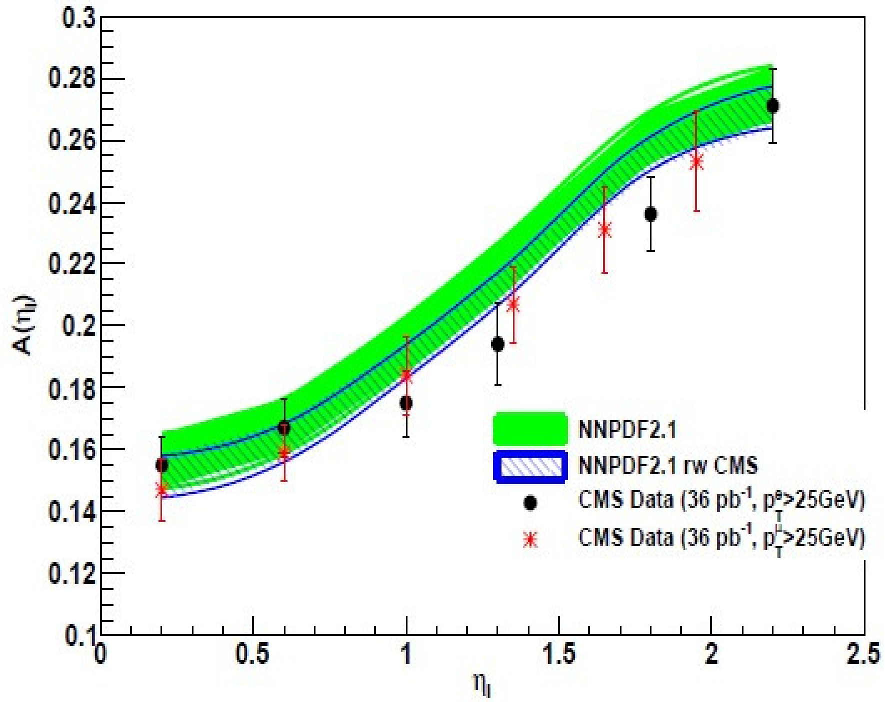


Mar 2011

W-lepton asymmetry

After Reweighting





	$N_{\rm dat}$	NNPDF2.1	+ LHC Wasy
ATLAS($31{\rm pb}^{-1}$)	11	0.76	0.97
CMS($36{\rm pb}^{-1}$) electron $p_T > 25$ GeV	6	1.83	1.23
CMS($36{\rm pb}^{-1}$) muon $p_T > 25$ GeV	6	1.24	0.63

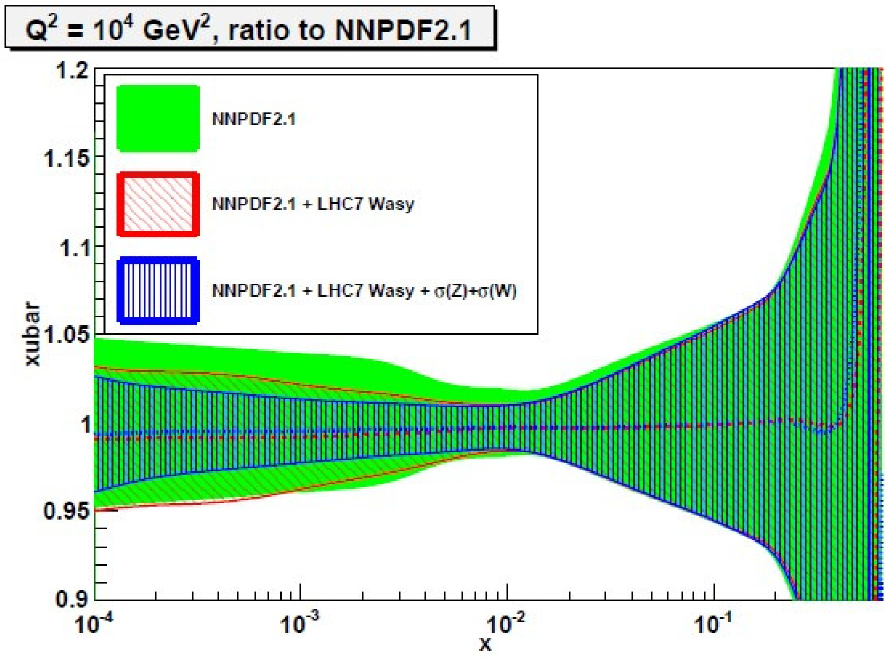
W-lepton asymmetry: future prospects

ATLAS LHC7

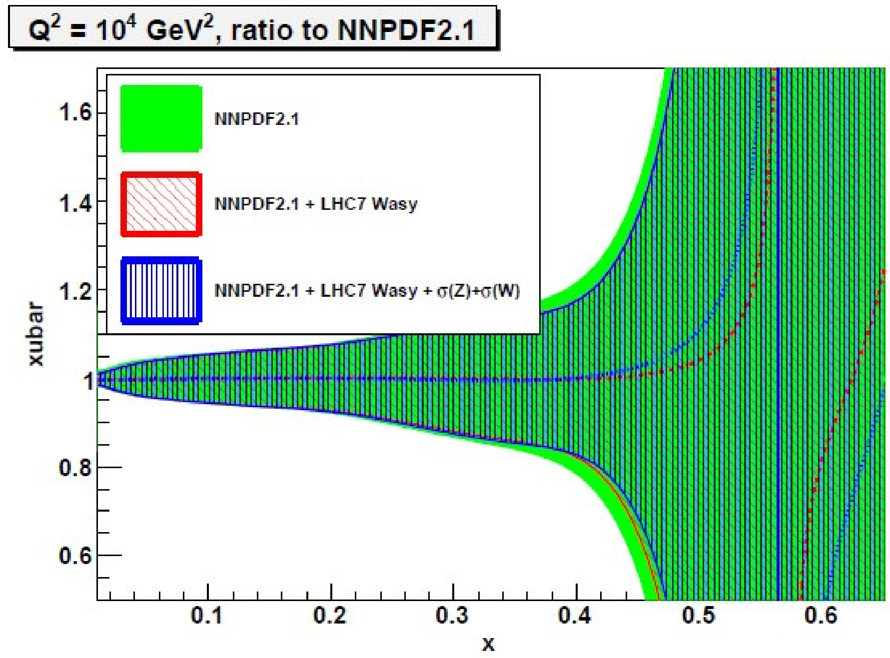
W lepton asymmetry measured to $\sim 5\%$ (kinematics courtesy A. Glazov)

W & Z total xsecs measured to 2%

ANTIUP SMALL x



ANTIUP LARGE x



Improvement seen for all flavours and antiflavours

NNPDF2.2

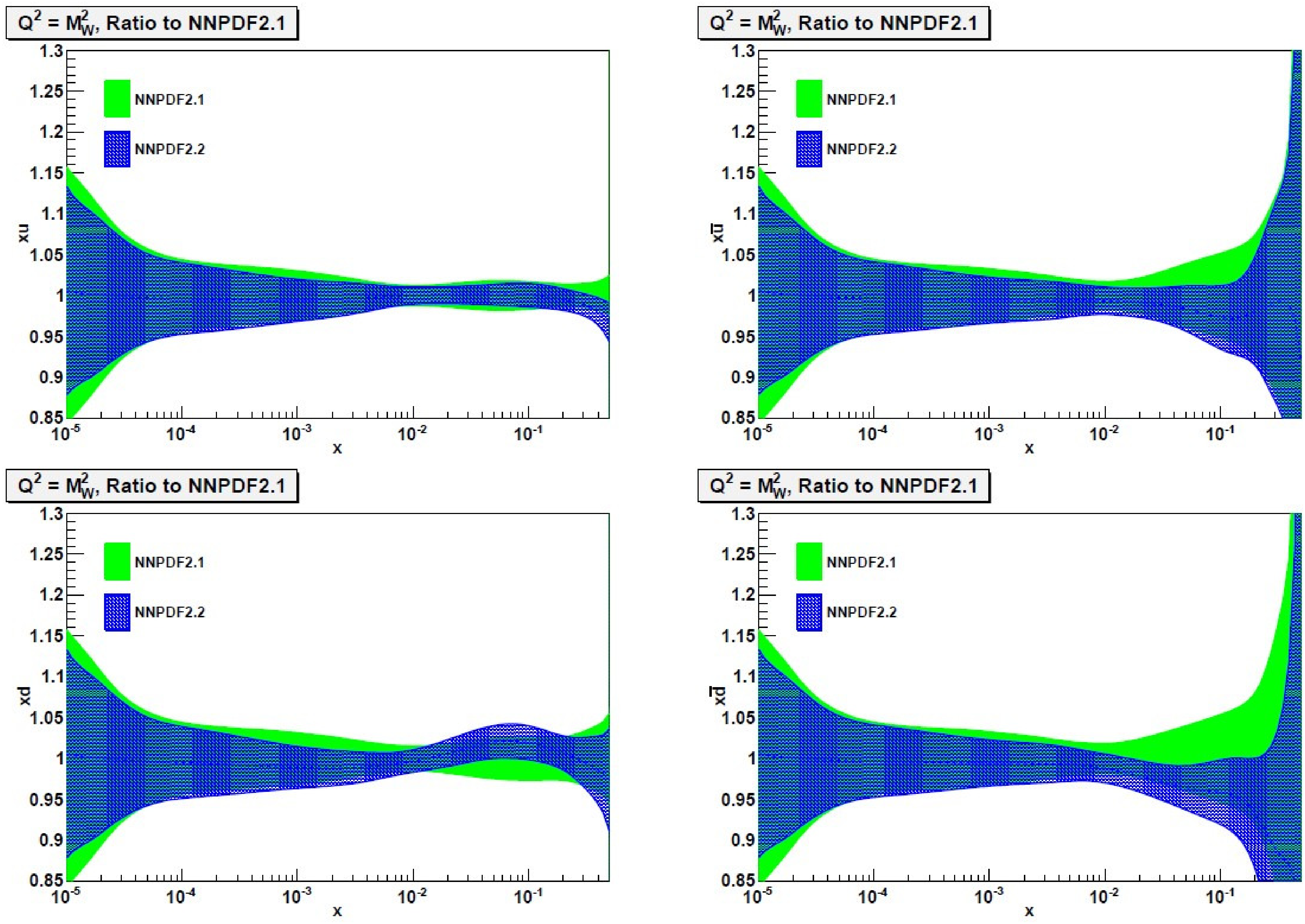
NNPDF2.2:

arXiv:1108.1758

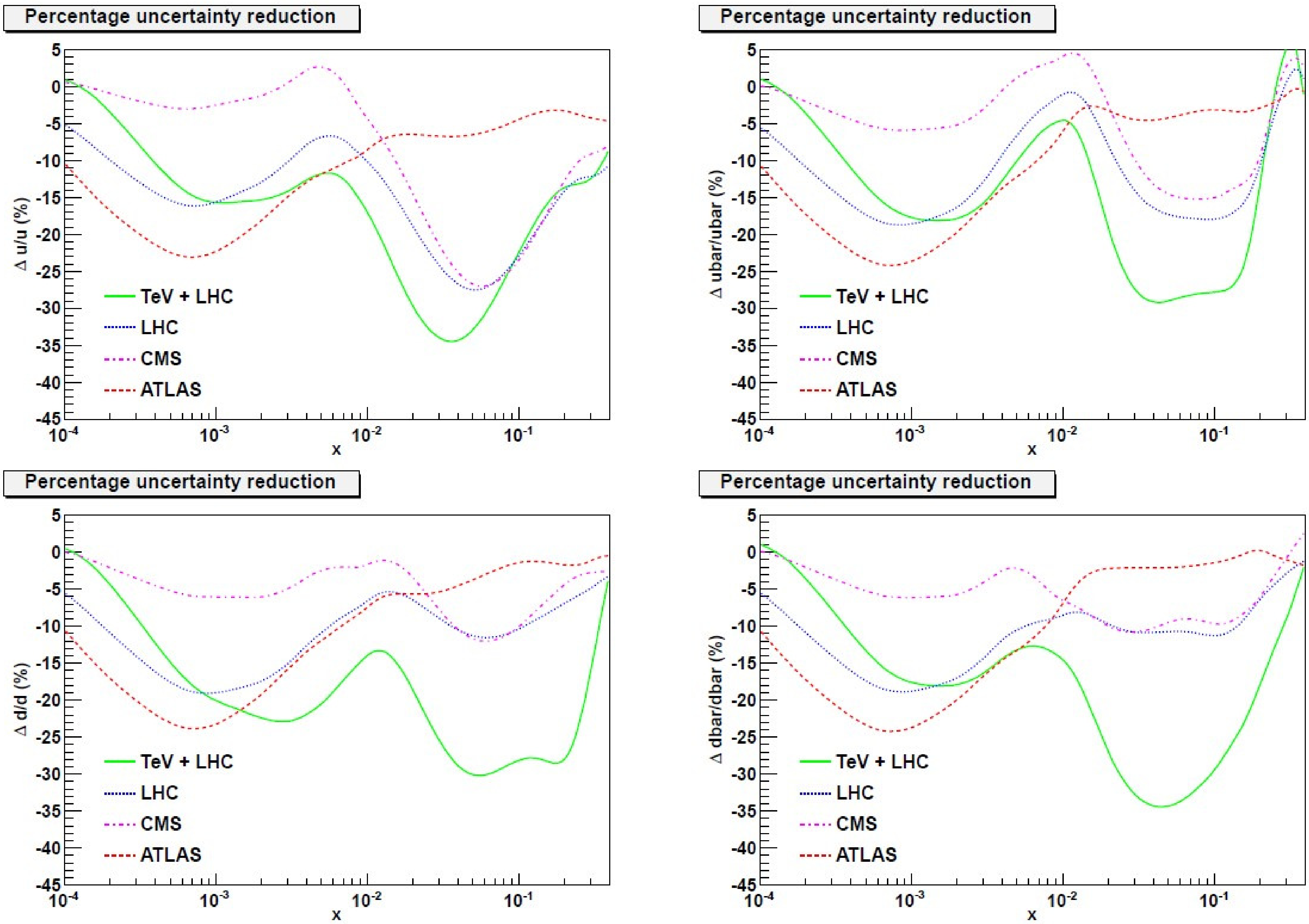
NNPDF2.1 + (Atlas+CMS+D0 W lepton asy)

Experiment	N_{dat}	NNPDF2.1	NNPDF2.1 LHC	NNPDF2.2
NMC-pd	132	0.97	0.95	0.97
NMC	221	1.73	1.72	1.72
SLAC	74	1.33	1.26	1.28
BCDMS	581	1.24	1.23	1.23
HERAI-AV	592	1.07	1.07	1.07
CHORUS	862	1.15	1.15	1.15
FLH108	8	1.37	1.37	1.37
NTVDMN	79	0.79	0.74	0.70
ZEUS-H2	127	1.29	1.28	1.28
ZEUSF2C	50	0.78	0.79	0.78
H1F2C	38	1.51	1.52	1.51
DYE605	119	0.84	0.84	0.86
DYE886	199	1.25	1.23	1.27
CDFWASY	13	1.85	1.81	1.81
CDFZRAP	29	1.66	1.61	1.70
D0ZRAP	28	0.60	0.60	0.58
CDFR2KT	76	0.98	0.98	0.96
D0R2CON	110	0.84	0.84	0.83
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
Total		1.165	1.158	1.157

NNPDF2.2: ups and downs



Percentage uncertainty reduction



Summary & Outlook

- NNPDF 2.1 LO, NLO, NNLO
  - See for yourself: http://projects.hepforge.org/lhapdf
- Reweighting:
  - You can update NNPDFs yourself: new tool
- NNPDF2.2:
  - First Global PDF fit with LHC data
- For the future:
  - Lots and lots of new LHC data!