NNPDF3.0:

parton distributions for the LHC run II

Maria Ubiali

University of Cambridge

in collaboration with:

R. D. Ball, V. Bertone, S. Carrazza, C. S. Deans, L. Del Debbio, S. Forte, A. Guffanti, N. P. Hartland, J. I. Latorre, J. Rojo, MU

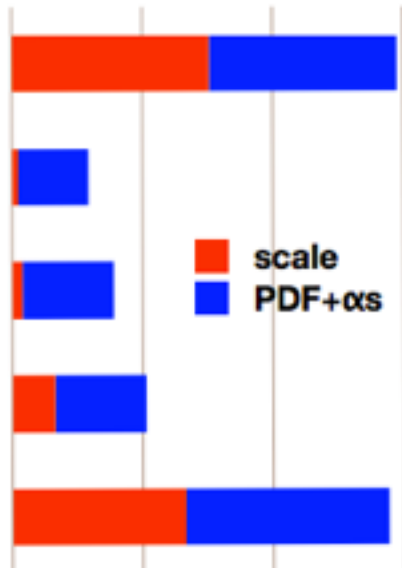
12th December 2014

University of Southampton

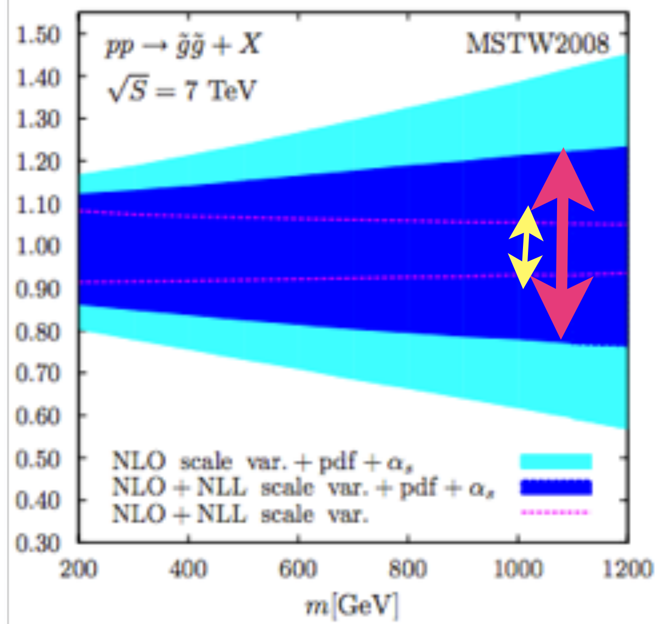
Motivation

PDFs and LHC interplay

	σ (8 TeV)	uncertainty
gg→H	19.5 pb	14.7%
VBF	1.56 pb	2.9%
WH	0.70 pb	3.9%
ZH	0.39 pb	5.1%
ttH	0.13 pb	14.4%



J. Campbell, ICHEP 2012



Beenakker et al (2011)

PDFs

PDF uncertainties are a crucial input at the LHC, often being the limiting factor in the accuracy of theoretical predictions, both SM and BSM

LHC

Exploit the power of precise LHC data to reduce PDF uncertainties and discriminate among PDF sets

Outline

- Introduction
  - Collinear factorisation
  - Key ingredients in PDF determination
- The NNPDF approach
  - Monte Carlo & NN
  - The closure test
- Results
  - The NNPDF2.3 QED partons set
  - The NNPDF3.0 set and implications for LHC phenomenology
- Conclusions

PDFs

and collinear factorisation

$$\frac{d\sigma_H^{pp\to ab}}{dX} = \sum_{i,j=1}^{N_f} f_i(x_1,\mu_F)\, f_j(x_2,\mu_F)\, \frac{d\sigma_H^{ij\to ab}}{dX}(x_1 x_2 S_{\rm had}, \alpha_s(\mu_R), \mu_F) + \mathcal{O}\left(\frac{\Lambda_{\rm QCD}^{2n}}{S_{\rm had}^n}\right)$$

PDFs cannot be computed in perturbative QCD but they are universal and their evolution with the scale is predicted by pQCD

$$\mu^2 \frac{\partial\, f(x,\mu^2)}{\partial\, \mu^2} = \int_z^1 \frac{dz}{z} \frac{\alpha_s}{2\pi} P(z)\, f\left(\frac{x}{z},\mu^2\right)$$

Dokshitzer, Gribov, Lipatov, Altarelli, Parisi renormalization group equations

LO - Dokshitzer; Gribov, Lipatov; Altarelli, Parisi, 1977

NLO - Floratos,Ross,Sachrajda; Floratos,Lacaze,Kounnas, Gonzalez-Arroyo,Lopez,Yndurain; Curci,Furmanski Petronzio, 1981

NNLO - Moch, Vermaseren, Vogt, 2004

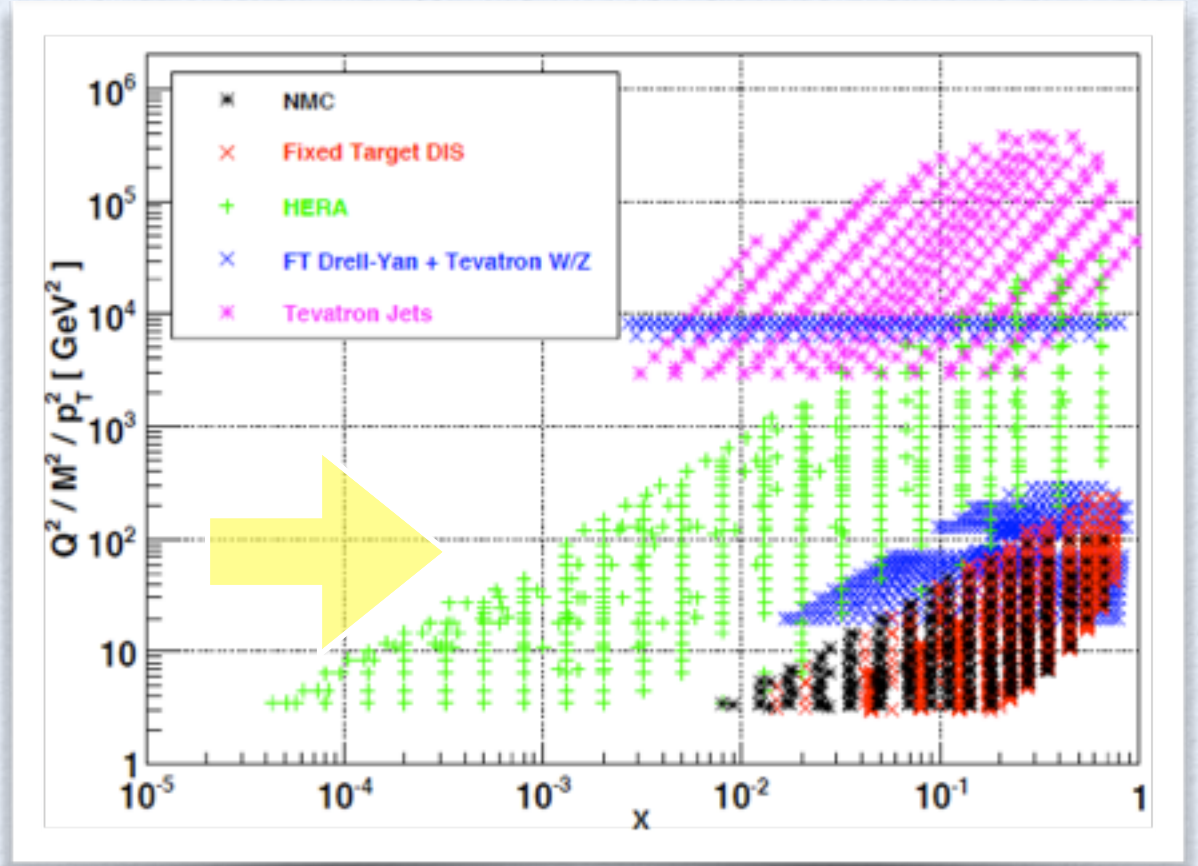
PDFs

and collinear factorisation

$$\frac{d\sigma_H^{pp\to ab}}{dX} = \sum_{i,j=1}^{N_f} f_i(x_1,\mu_F)\, f_j(x_2,\mu_F)\, \frac{d\sigma_H^{ij\to ab}}{dX}(x_1 x_2 S_{\rm had}, \alpha_s(\mu_R), \mu_F) + \mathcal{O}\left(\frac{\Lambda_{\rm QCD}^{2n}}{S_{\rm had}^n}\right)$$

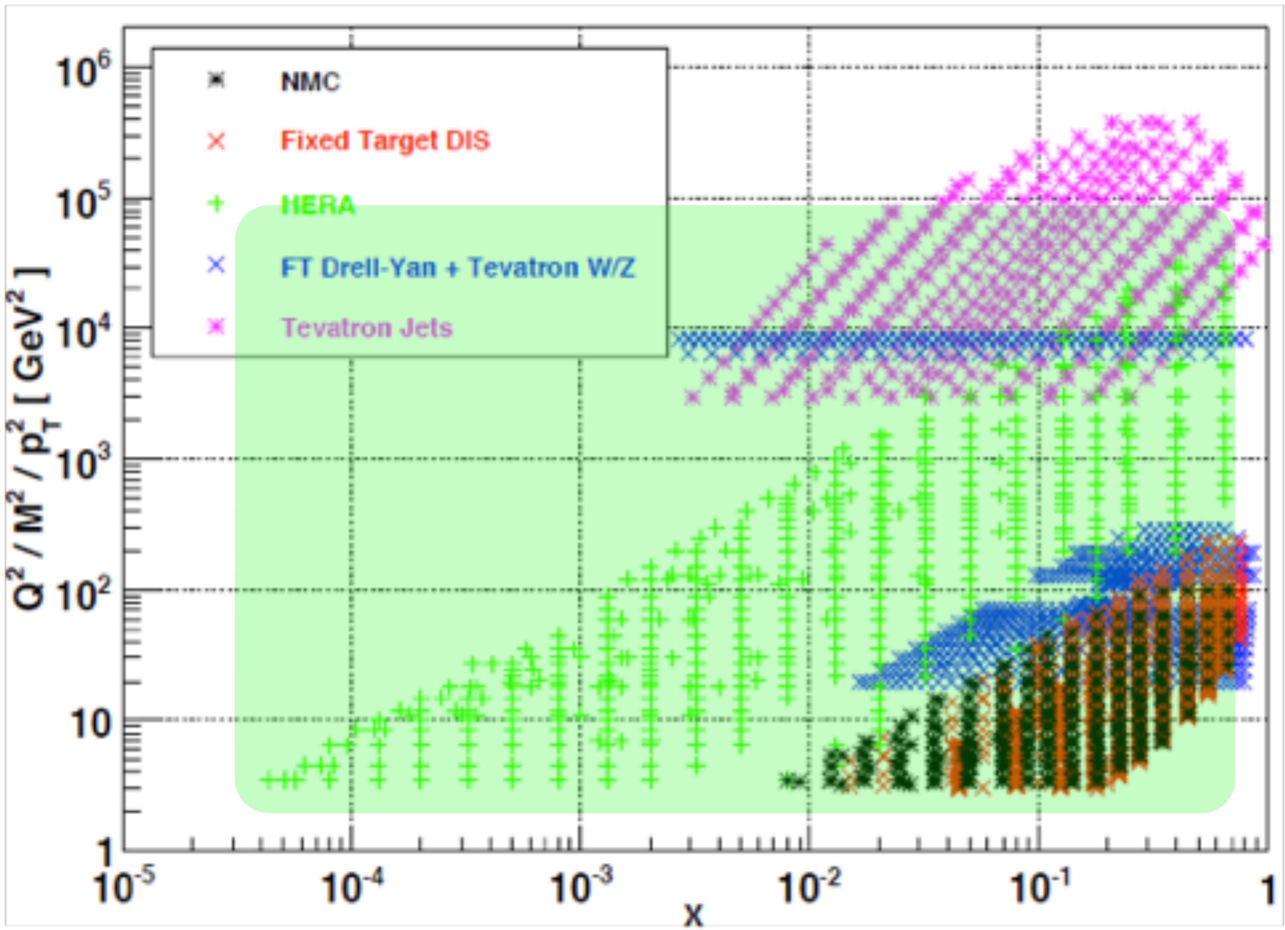
- They can be extracted from available experimental data and used as phenomenological input for theory predictions

- Different data constrain different parton combinations at different x



Constraints from data

before the LHC



NC $\quad F_1^{\gamma, Z} = \sum_i e_i^2 (q_i + \bar{q}_i)$

CC $\quad F_1^{W^+} = \bar{u} + d + s + \bar{c}$

CC $\quad -F_3^{W^+}/2 = \bar{u} - d - s + \bar{c}$

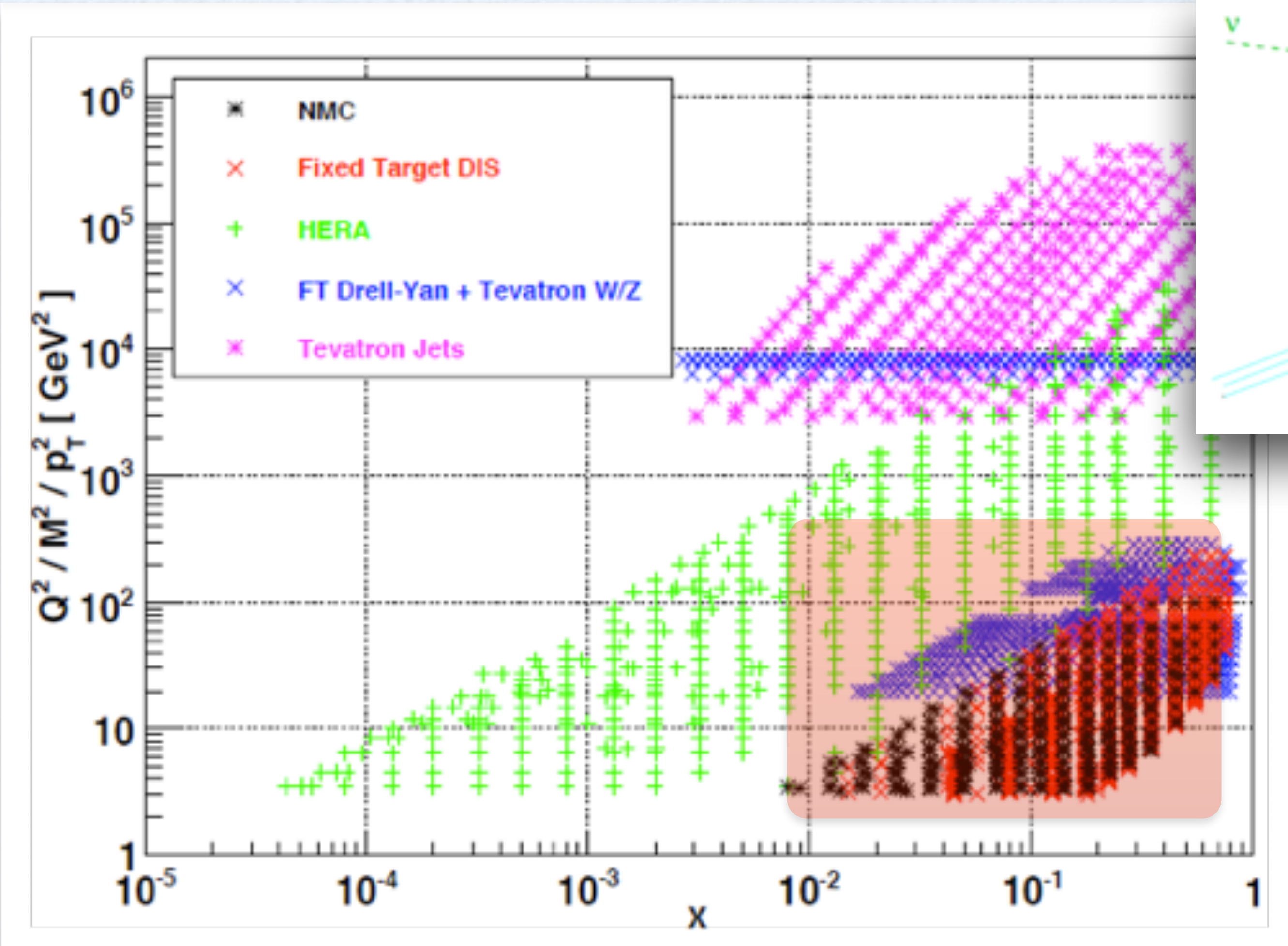
$$F_2 = 2x F_1$$

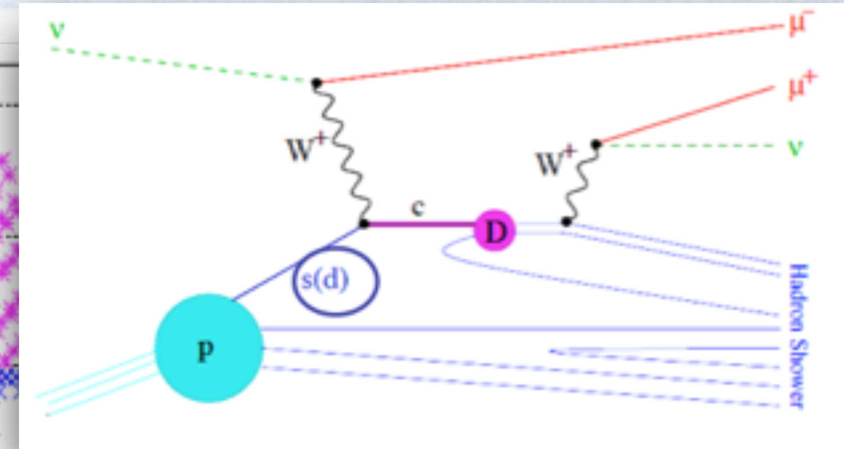
HERA DIS data

- backbone of any PDF fit
- q, qbar at 10^{-4}
- g at small and moderate x

Constraints from data

before the LHC



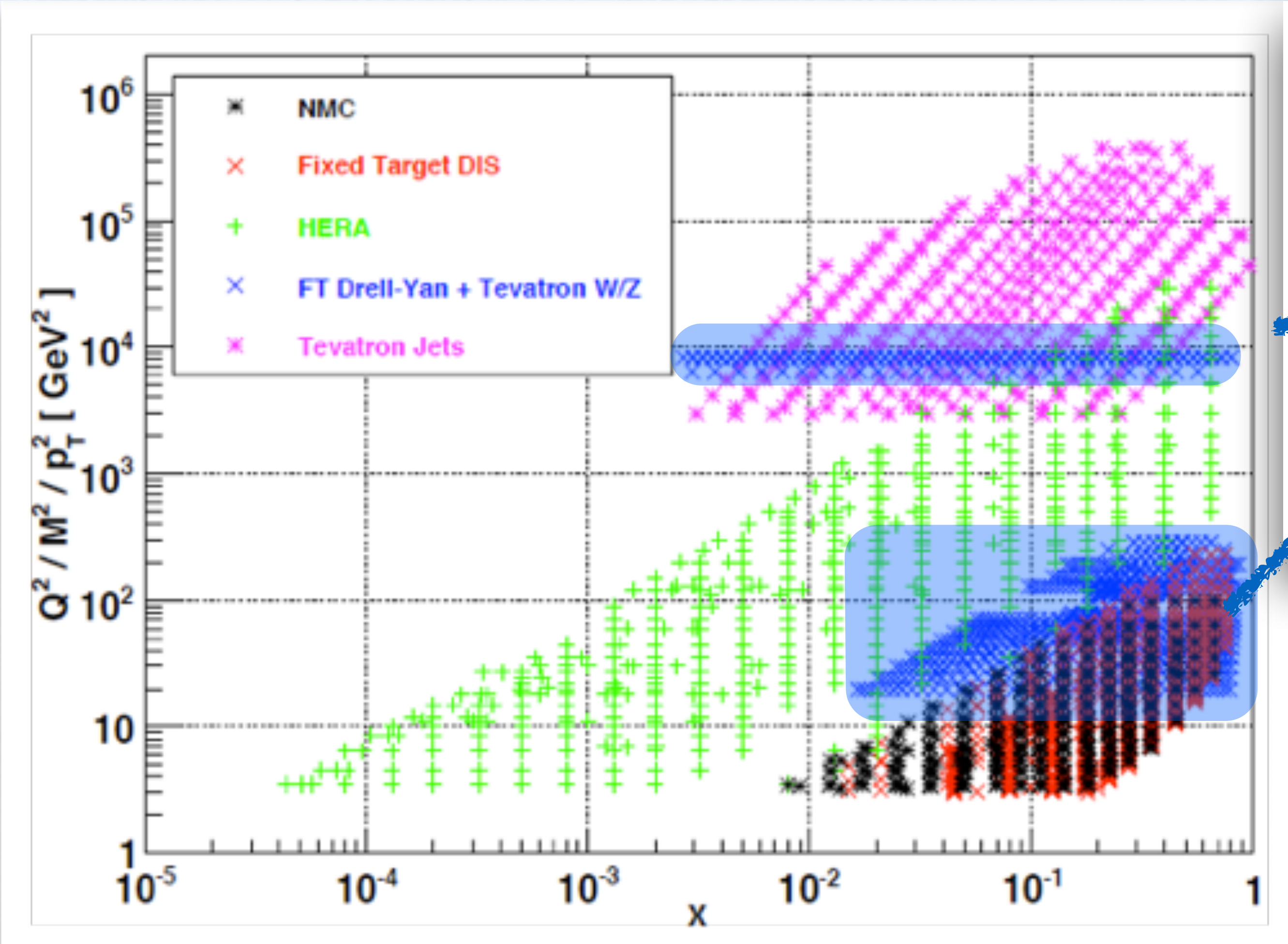


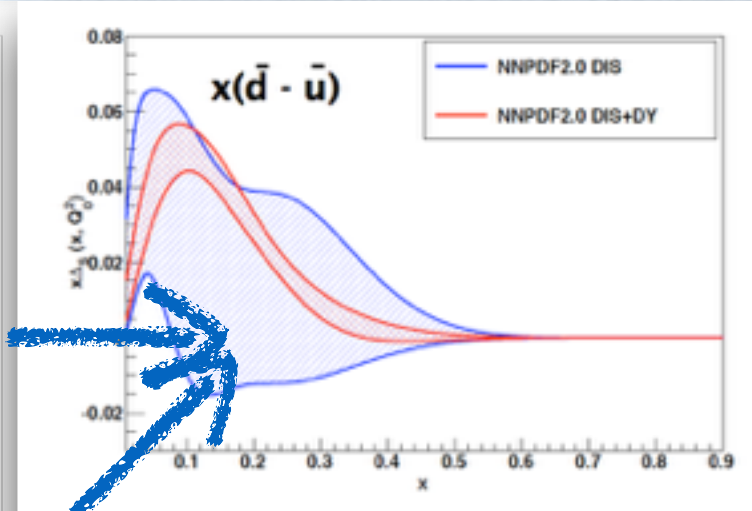
Fixed Target DIS data

- deuteron data: disentangle isospin triplet and singlet contributions
- strange and anti-strange at moderate $x > 10^{-2}$

Constraints from data

before the LHC





[NNPDF2.0, Ball et al, ArXiv: 1002.4407]

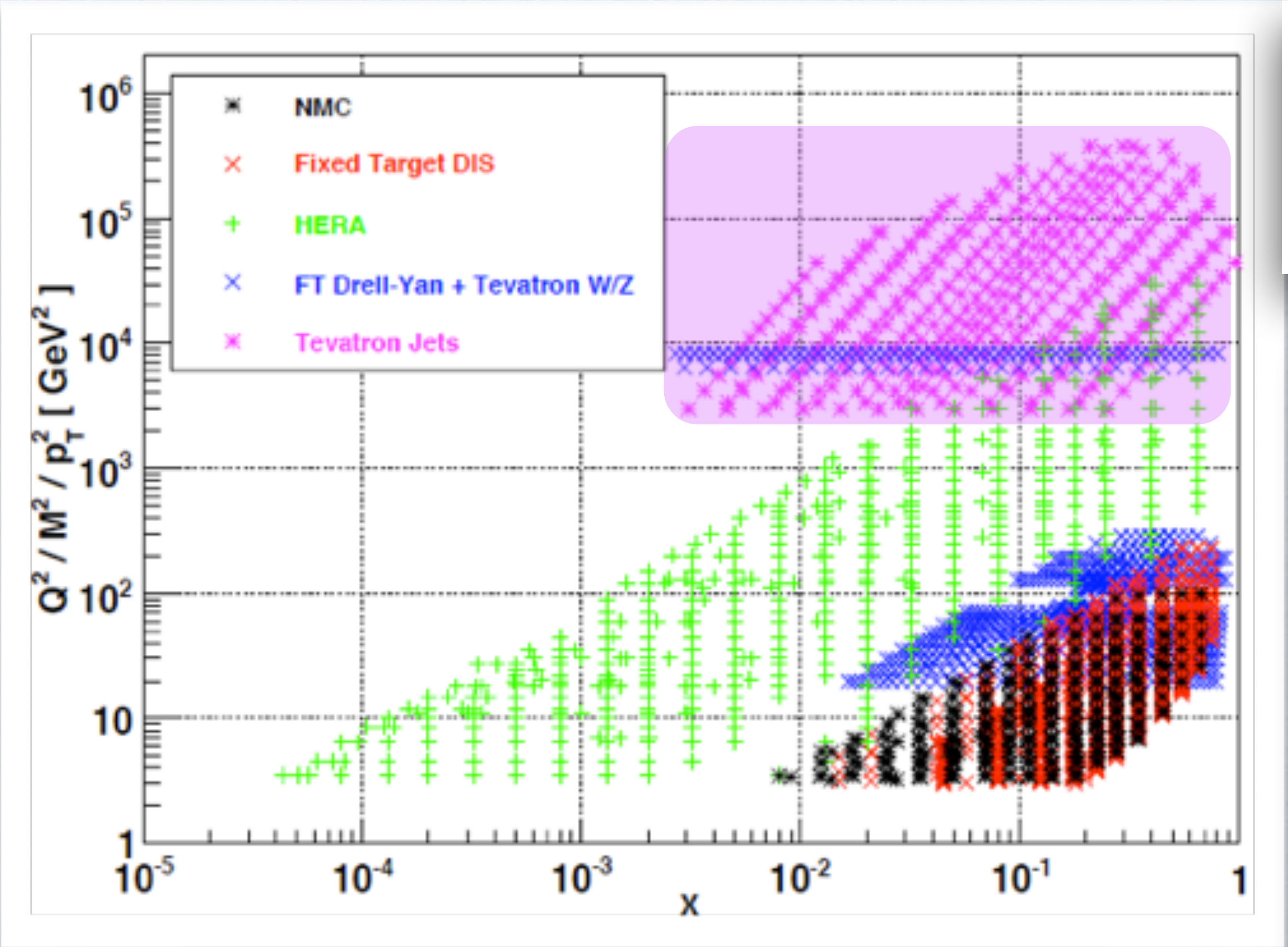
DY and EW vector boson data

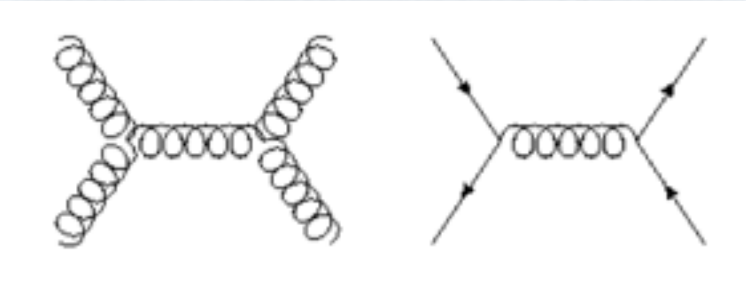
- light quark and antiquark separation

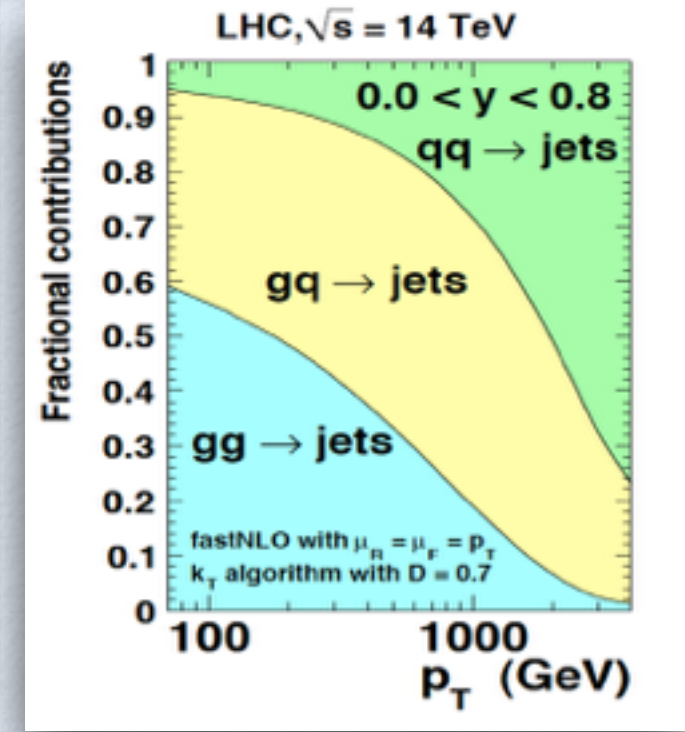
$$\sigma^{\mathrm{DY,p}} \propto u(x_1)\bar{u}(x_2) + d(x_1)\bar{d}(x_2)$$
$$\sigma^{\mathrm{DY,d}} \propto u(x_1)(\bar{u}+\bar{d})(x_2) + d(x_1)(\bar{u}+\bar{d})(x_2)$$

Constraints from data

before the LHC







- quarks and gluons at large x

The name of the game

How does it work?

- Choose **experimental data** to fit
- **Theory settings**: factorization scheme, perturbative order, heavy quark mass scheme, EW corrections
- Choose a starting scale where pQCD applies Q_0
- **Parametrise** quarks and gluon distributions at the starting scale
- Solve DGLAP equations from initial scale to scales of experimental data

  and build up **observables**
- Fit PDFs to data
- Provide **error sets** to compute PDF uncertainties

Hessian prescription

$$\sigma_{\mathcal{F}} = \left(\sum_{k=1}^{N_{\text{set}}} \left(\mathcal{F}[\{f^{(k)}\}] - \mathcal{F}[\{f^{(0)}\}] \right)^2 \right)^{1/2}$$

error sets

mem > 1

central set

mem = 0

LHAPDF interface

http://lhapdf.hepforge.org

```
call InitPDF(mem)

call evolvePDF(x,Q,f)
```

	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6
Parton	tbar	bbar	cbar	sbar	ubar	dbar	g	d	u	s	c	b	t

The name of the game

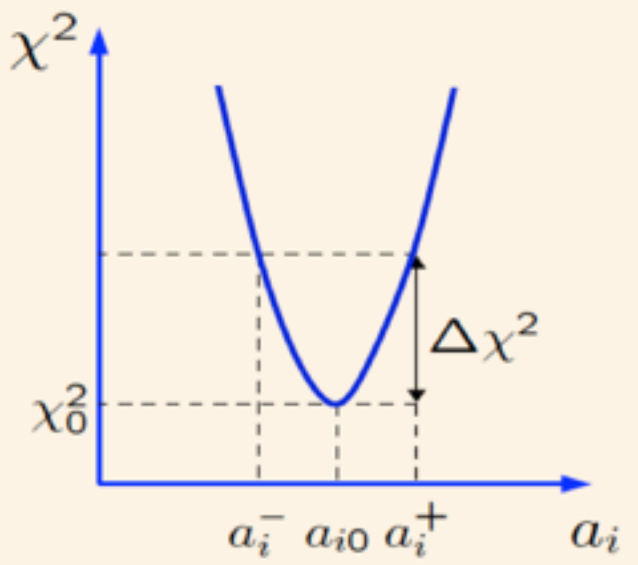
Not as simple as it may look

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

- Given a finite number of experimental point want a set of functions with error
- Standard approach: project into a n-dimensional space of parameters and use linear approximation around the minimum of the χ^2 (Hessian method)

$$f_i(x, Q_0^2) = a_0\, x^{a_1}\, (1-x)^{a_2}\, P(x, a_3, a_4, \dots)$$

- Possible issues:
  - (I) Linear approximation and Gaussian assumption
  - (II) Tolerance > 1 equivalent to blow up uncertainties



- $\Delta\chi^2 = 1$, ABKM fits and HERA (non global)
- $\Delta\chi^2 = 10$ [CT10], $\Delta\chi^2 \sim 7.5$ [MRST2001], dynamical tolerance [MSTW08], $3 < \Delta\chi^2 < 5$
- Uncertainty inflated by a factor 2/5?

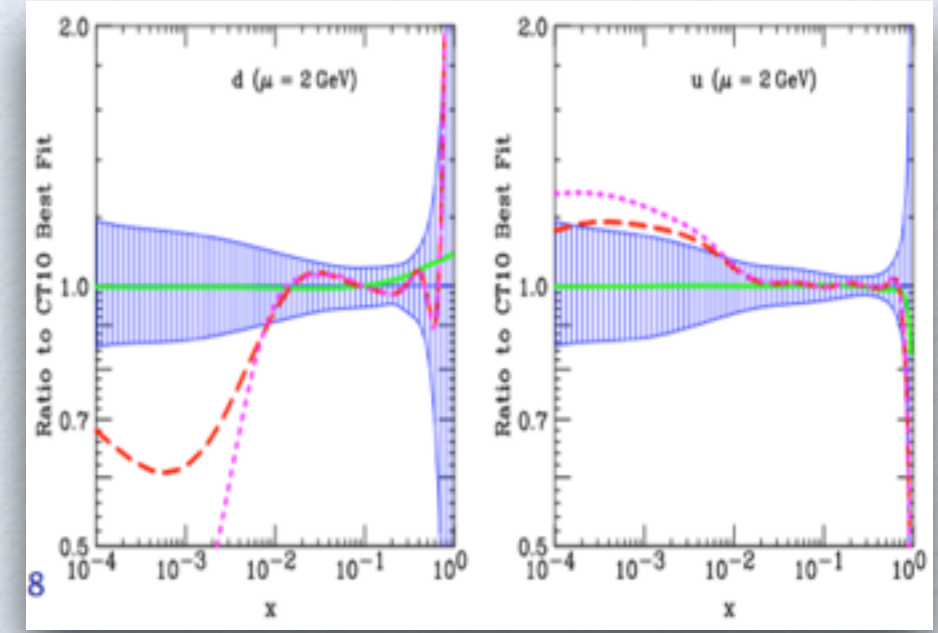
The name of the game

Not as simple as it may look

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

- Given a finite number of experimental point want a set of functions with error
- Standard approach: project into a n-dimensional space of parameters and use linear approximation around the minimum of the χ^2 (Hessian method)

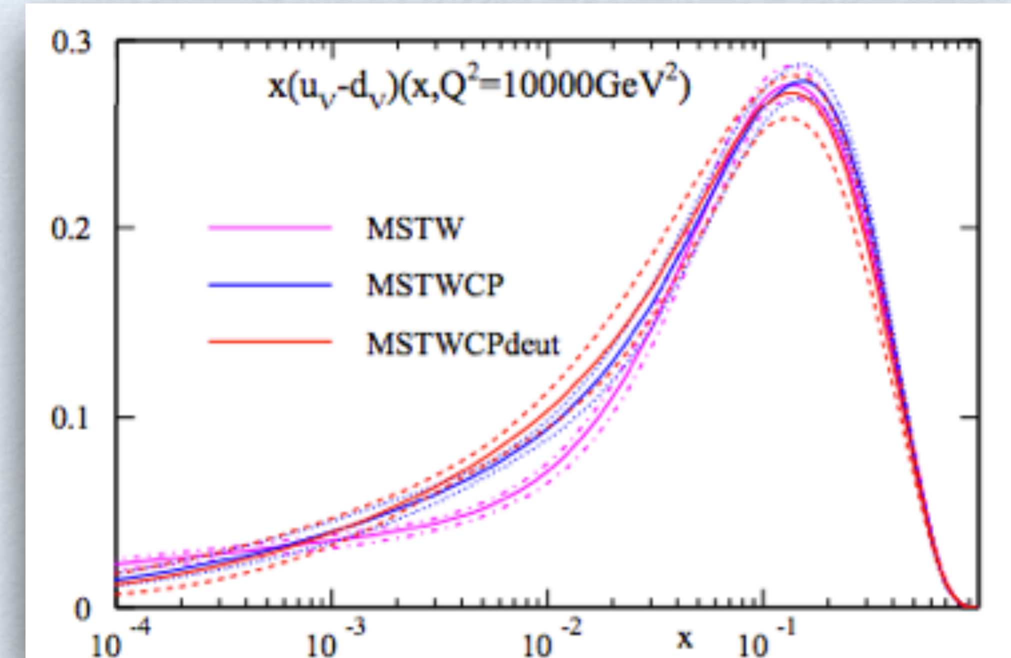
- Possible issues:
  - (III) Parametrisation: what is the error associated to a given functional form? If it is not flexible enough PDFs may be not able to adapt to new data or present unrealistically small errors where data do not constrain PDF uncertainties

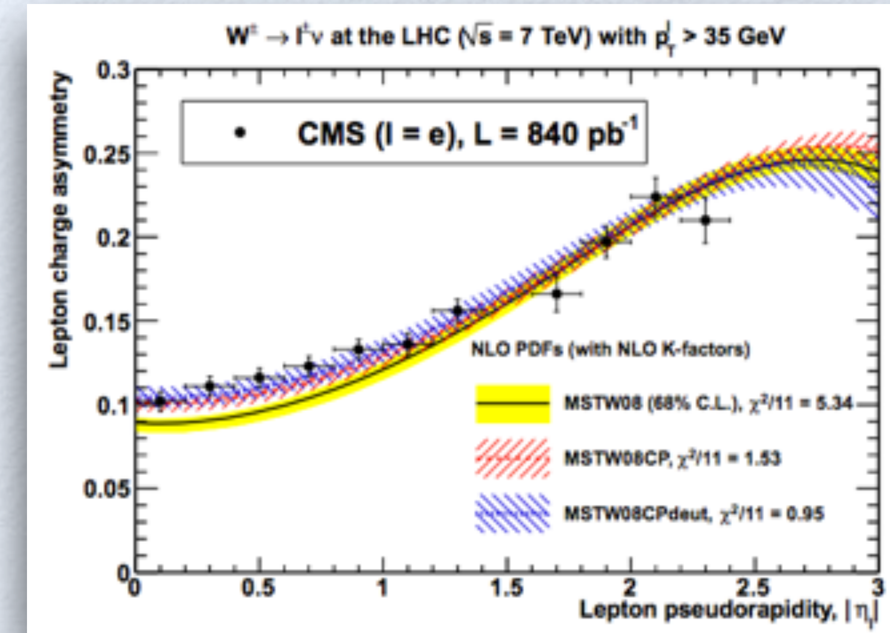


J. Pumplin ArXiv:0909.0268

The name of the game

Not as simple as it may look





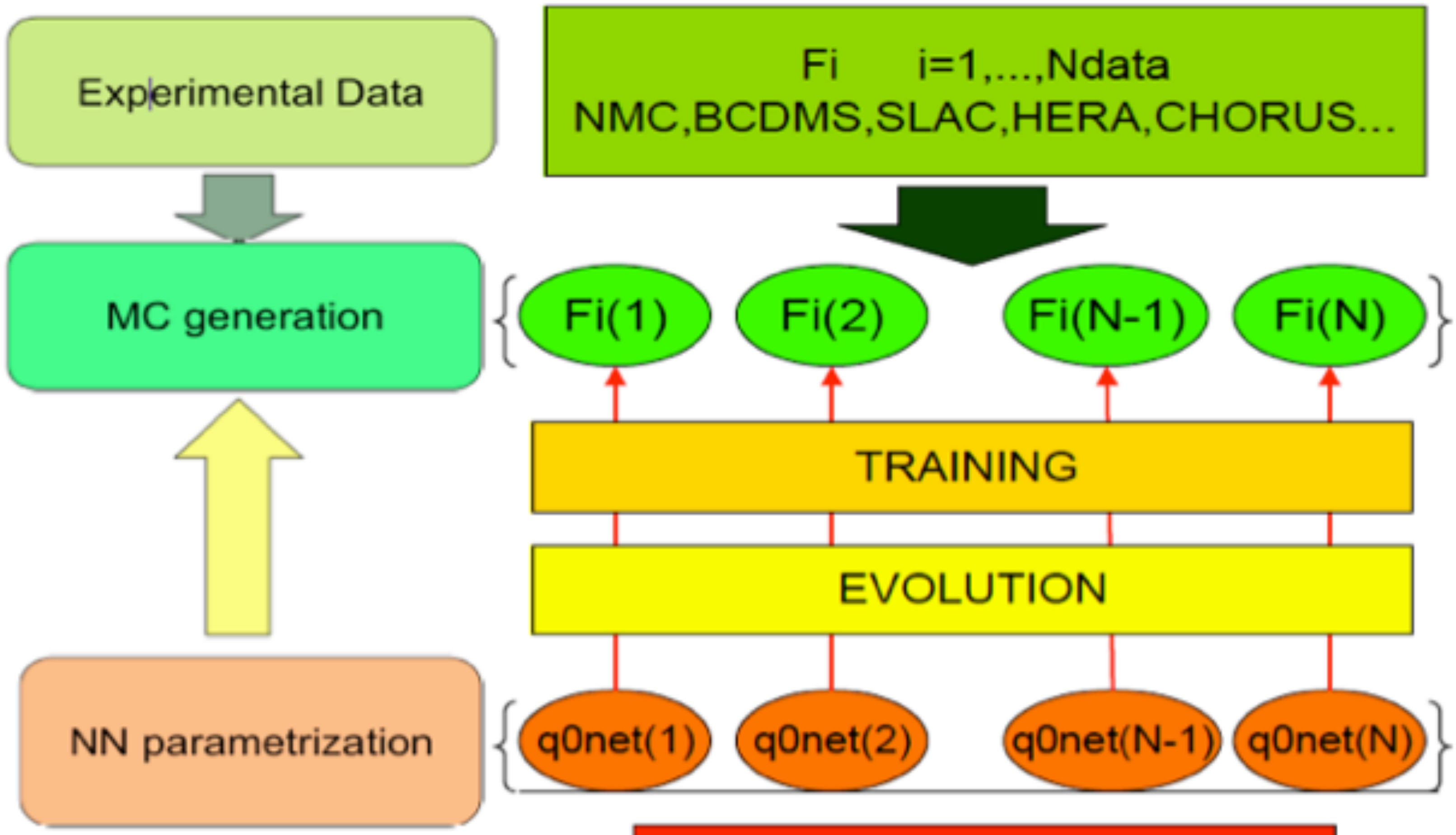
Martin, Mathijssen, Stirling, Thorne, Watt, Watt
ArXiv: 1211.1215 [Eur.Phys.J. C73 (2013) 2318]

- Recent study by MSTW collaboration by using large and flexible Chebyshev polynomials parametrization
- Spotted a restrictive u_v and d_v parametrization in MSTW2008 fit
- Larger parametrization needed to have an adequate description of W asymmetry data.

The NNPDF approach

The NNPDF solution

Monte Carlo and Neural Network

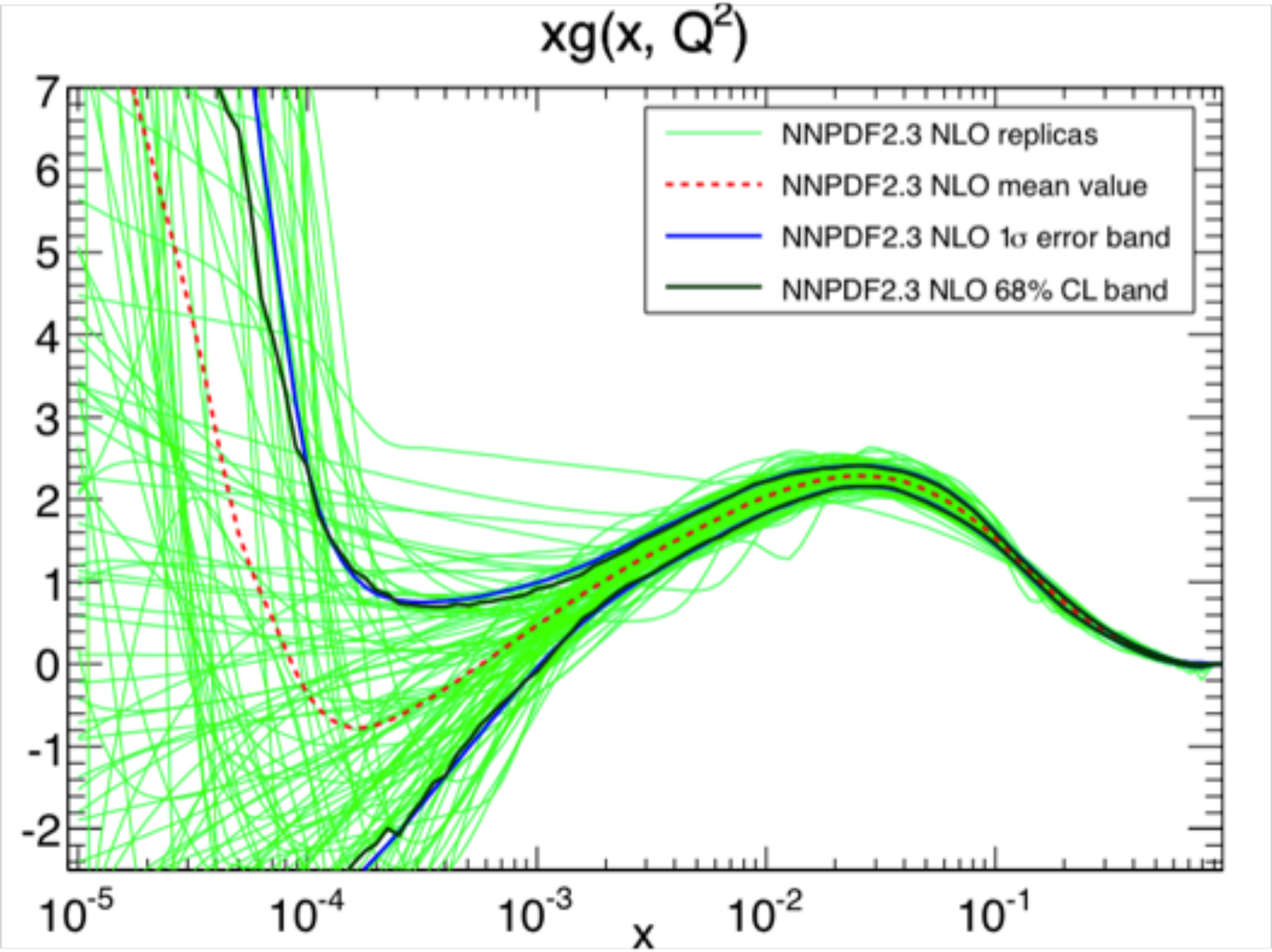


$$\left\langle F_i^{(\mathrm{net})} \right\rangle_{\mathrm{rep}} = \frac{1}{N_{\mathrm{rep}}} \sum_{k=1}^{N_{\mathrm{rep}}} F_i^{(\mathrm{net})(k)}$$

Ball, Del Debbio, Forte, Guffanti, Latorre, Rojo, MU, ArXiv:0808.1231

The NNPDF solution

Monte Carlo and Neural Network



$$\langle X \rangle = \int d\vec{a}\, X[\vec{a}]\, \mathcal{P}[\vec{a}]$$

$$\langle X \rangle \simeq \frac{1}{N_{\mathrm{rep}}} \sum_{i=1}^{N_{rep}} X(\vec{a}_i)$$

Generate a MC sampling in the parameter space? NO

INSTEAD:
Choose replicas of the data, i.e. work in the space of data and project back into PDF space

Ball, Del Debbio, Forte, Guffanti, Latorre, Rojo, MU, ArXiv:0808.1231

The NNPDF solution

Monte Carlo and Neural Network

- Generate N_{rep} sets of "pseudo-data" of the original N_{data} data points

$$F_i^{(art)(k)}(x_p, Q_p^2) \equiv F_{i,p}^{(art)(k)} \qquad \begin{aligned} i &= 1, \ldots, N_{\mathrm{data}} \\ k &= 1, \ldots, N_{\mathrm{rep}} \end{aligned}$$

- Multi-Gaussian distribution centered on each data point

$$F_{i,p}^{(art)(k)} = S_{p,N}^{(k)} F_{i,p}^{\mathrm{exp}} \left(1 + r_p^{(k)} \sigma_p^{\mathrm{stat}} + \sum_{j=1}^{N_{\mathrm{sys}}} r_{p,j}^{(k)} \sigma_{p,j}^{\mathrm{sys}} \right)$$

- If two points have correlated systematic uncertainties

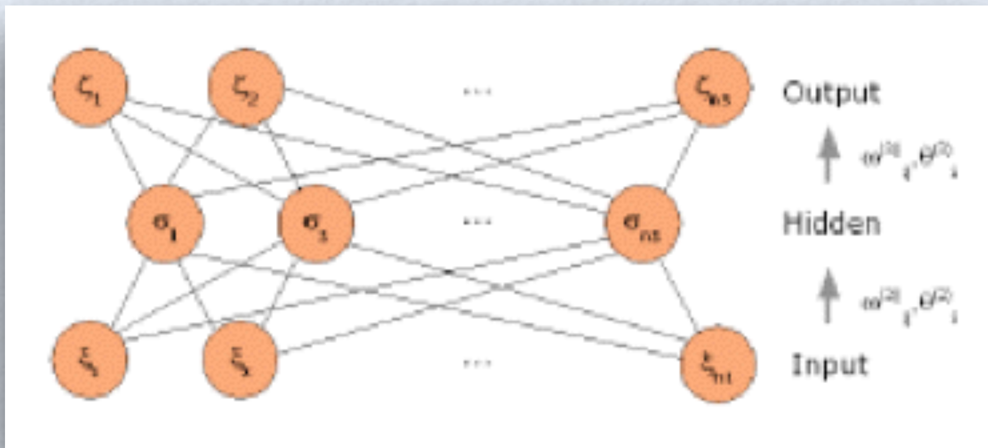
$$r_{p,j}^{(k)} = r_{p',j}^{(k)}$$

- Correlations are properly taken into account

The NNPDF solution

Monte Carlo and Neural Network

Each independent PDF at initial scale is parametrized by an individual NN



- Each neuron receives input from neurons in the preceding layer
- Activation determined by weights and thresholds according to non linear functions

$$\xi_i = g\Big(\sum_j \omega_{ij}\xi_j - \theta_i\Big), \qquad g(x) = \frac{1}{1+e^{-x}}$$

In a simple case (1-2-1) we have,

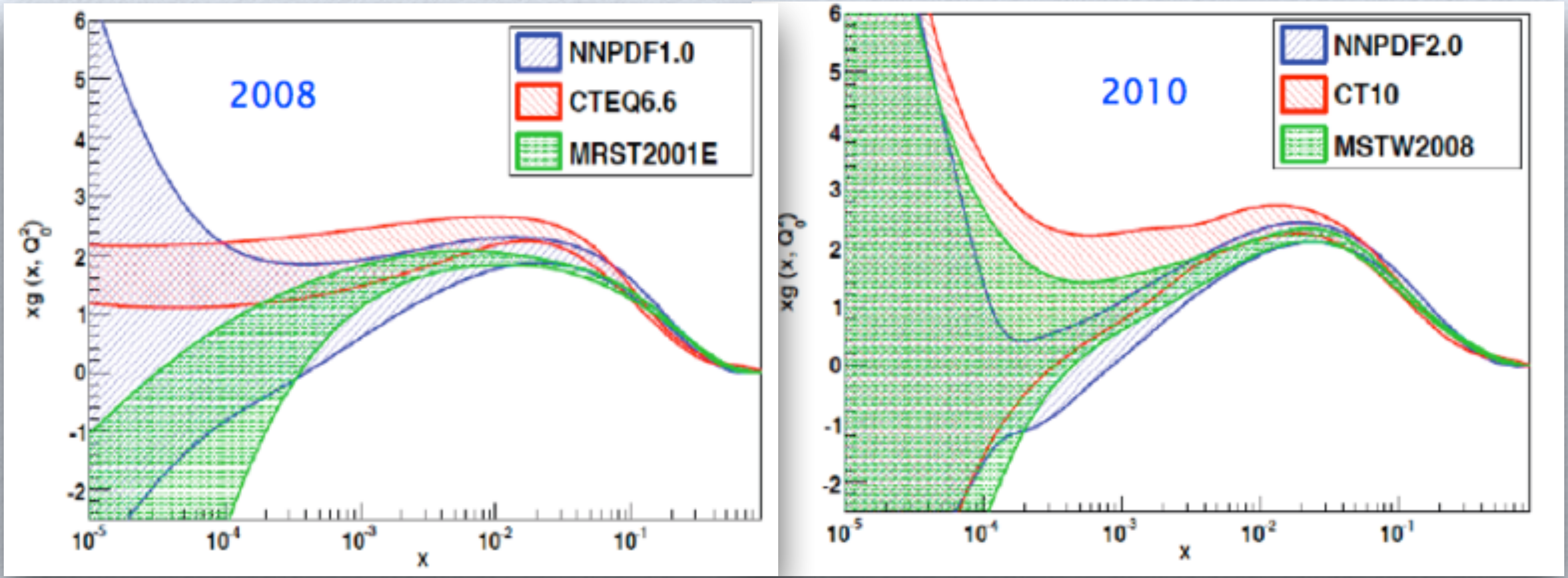
$$\xi_1^{(3)} = \frac{1}{1+e^{\theta_1^{(3)} - \frac{\omega_{11}^{(2)}}{1+e^{\theta_1^{(2)}-\xi_1^{(1)}\omega_{11}^{(1)}}} - \frac{\omega_{12}^{(2)}}{1+e^{\theta_2^{(2)}-\xi_1^{(1)}\omega_{21}^{(1)}}}}}$$

7 parameters

- Just a convenient functional form which provides a redundant and flexible parametrization

The NNPDF solution

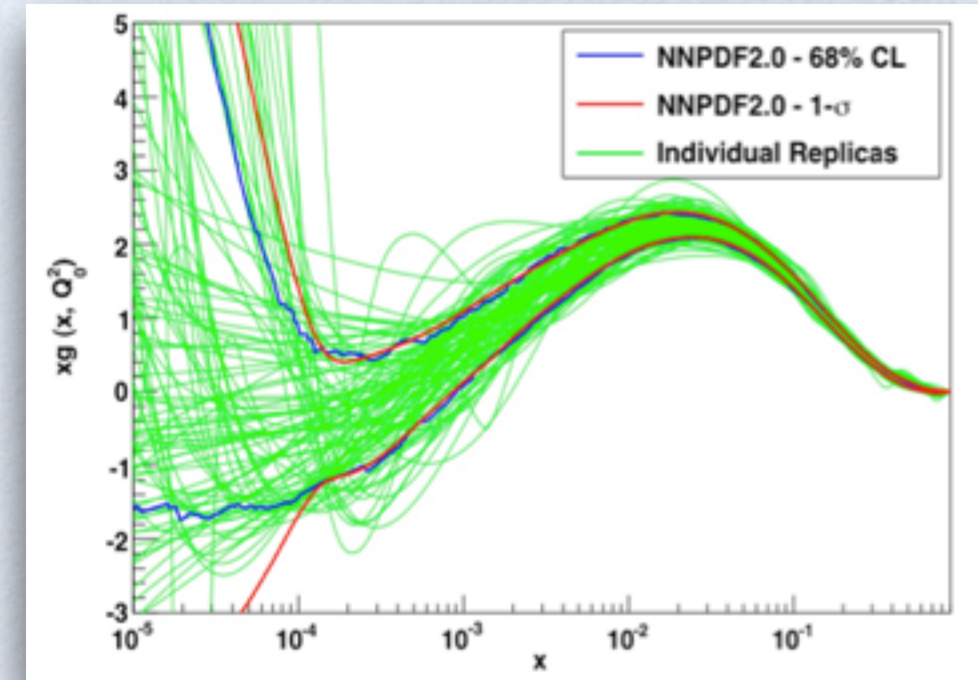
Monte Carlo and Neural Network

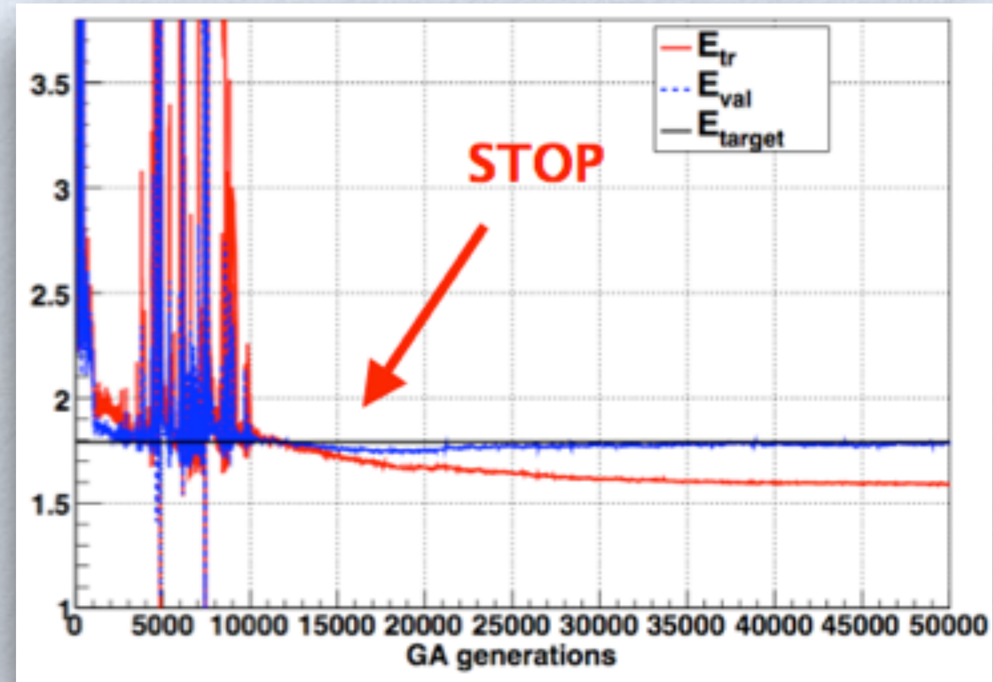


- MSTW and CT added parameters (e.g. an exponent in small-x region)
- NNPDF always uses same redundant parametrization

The NNPDF solution

Monte Carlo and Neural Network





- Neural networks provide flexible and redundant parametrization
- O(250) parameters versus O(25) parameters of fixed parametrization
- Same parametrization for all fits
- Can verify independence of parametrization
- Cross-Validation method avoids over-learning of statistical fluctuations

The NNPDF solution

Monte Carlo and Neural Network

	NNPDF1.0	NNPDF1.2	NNPDF2.0	NNPDF2.1	NNPDF2.3	MSTW08	CT10
DIS	✔	✔	✔	✔	✔	✔	✔
Drell-Yan	✘	✘	✔	✔	✔	✔	✔
Jet	✘	✘	✔	✔	✔	✔	✔
LHC	✘	✘	✘	✘	✔	✘	✘
strange	✘	✔	✔	✔	✔	✔	✔
Heavy Quark	✘	✘	✘	✔	✔	✔	✔
NNLO	✘	✘	✘	✔	✔	✔	✔

NNPDF2.3QED

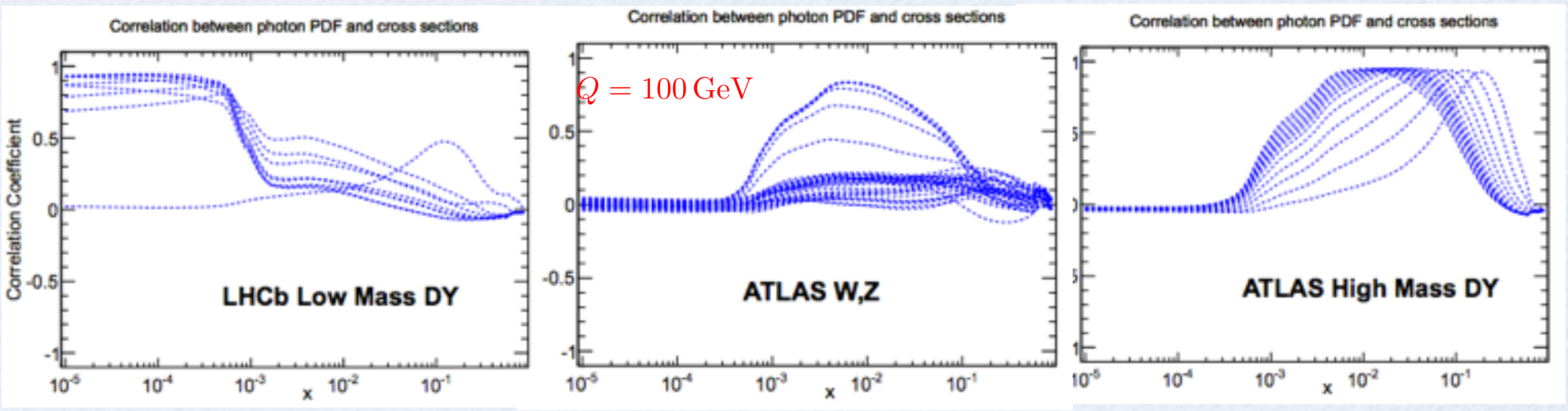
All details in arXiv:1308.0598

The NNPDF2.3 set

Features and data

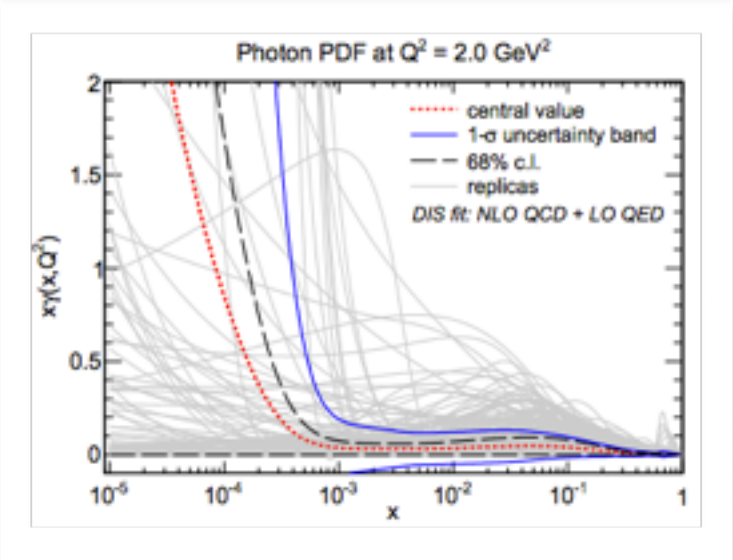
- EW corrections have become relevant at the current phenomenological precision level
- A consistent inclusion of EW corrections requires PDF with QED effects
- NNPDF23QED is new PDF set with uncertainties which incorporates (N)NLO QCD + LO QED effects
- Photon PDF fitted from DIS and DY data (on-shell W,Z production and low/high mass DY)
- DIS data fitted and DY data included via Bayesian re-weighting [Ball et al., Nucl.Phys. B855 (2012) 608-638]
- Photon PDF is poorly determined from DIS data. Need hadron collider processes where photon contributes at LO!

Dataset	Observable	$N_{\rm dat}$	$[\eta_{\rm min}, \eta_{\rm max}]$	$[M_{ll}^{\rm min}, M_{ll}^{\rm max}]$
LHCb γ^*/Z Low Mass	$d\sigma(Z)/dM_{ll}$	9	[2,4.5]	[5,120] GeV
ATLAS W, Z	$d\sigma(W^\pm, Z)/d\eta$	30	[-2.5,2.5]	[60,120] GeV
ATLAS γ^*/Z High Mass	$d\sigma(Z)/dM_{ll}$	13	[-2.5,2.5]	[116,1500] GeV

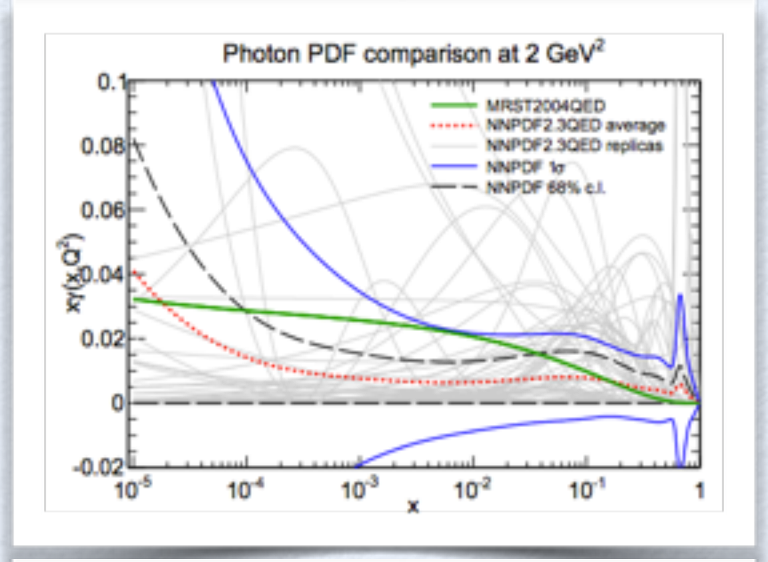


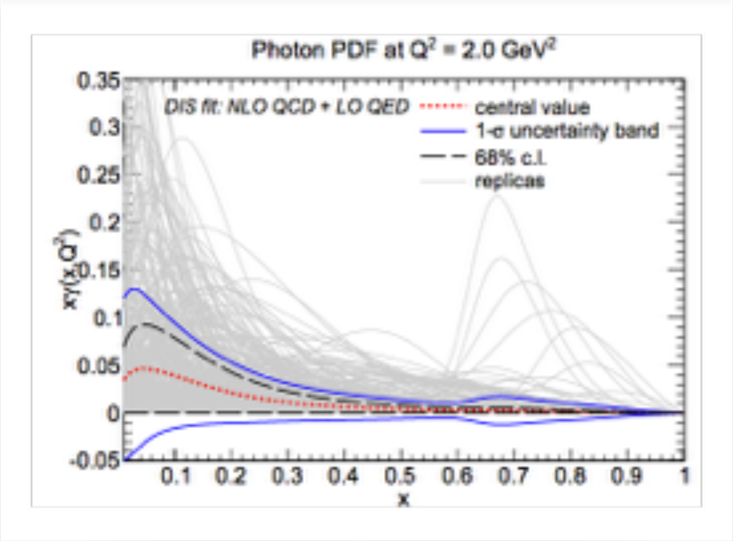
The NNPDF2.3 set

Constraints from the LHC

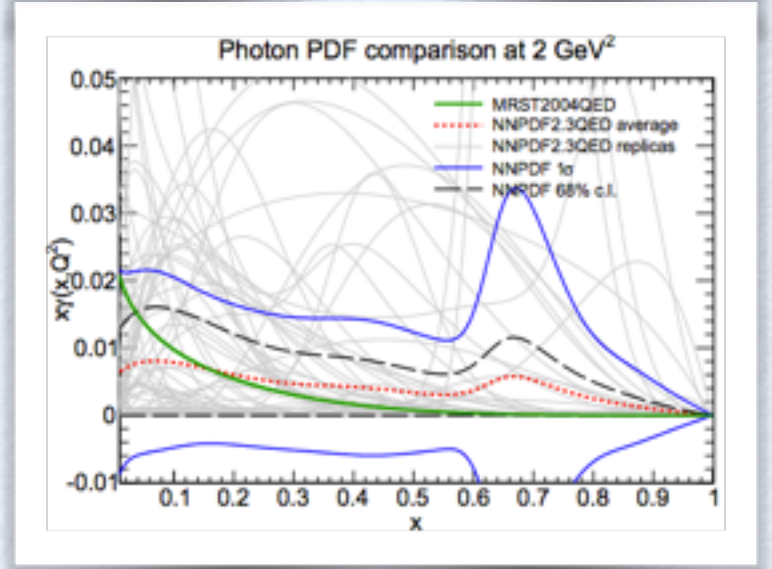


small x





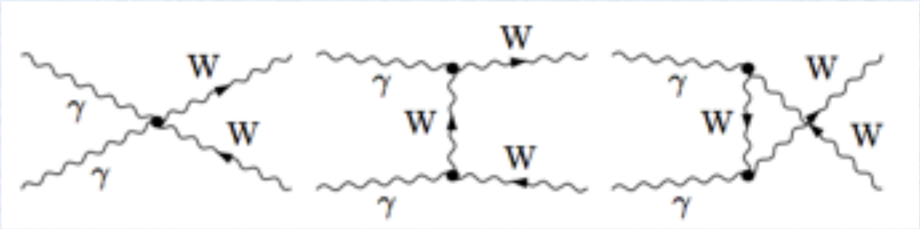
large x

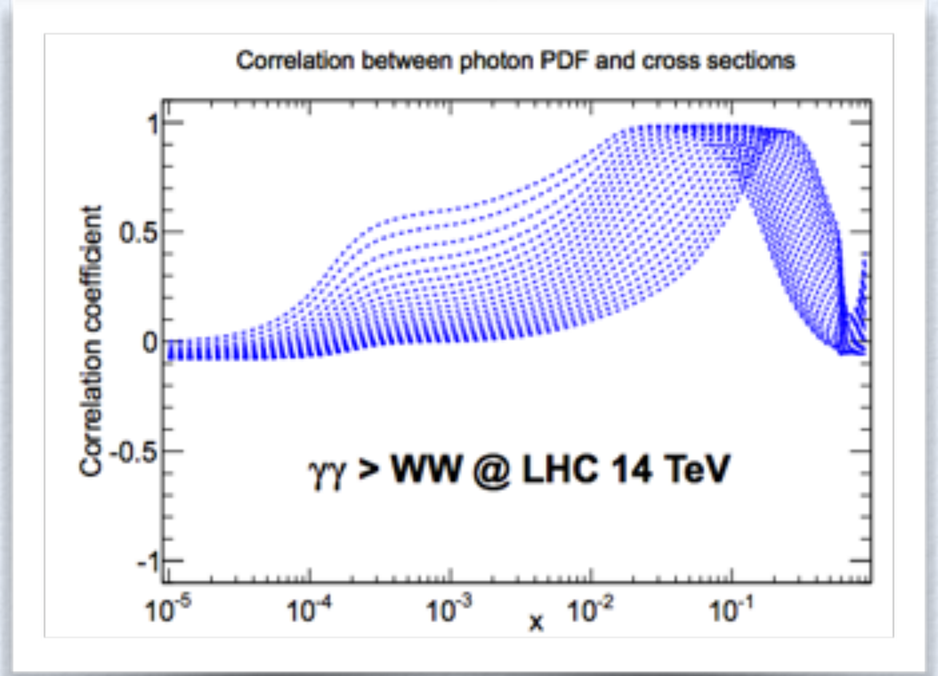


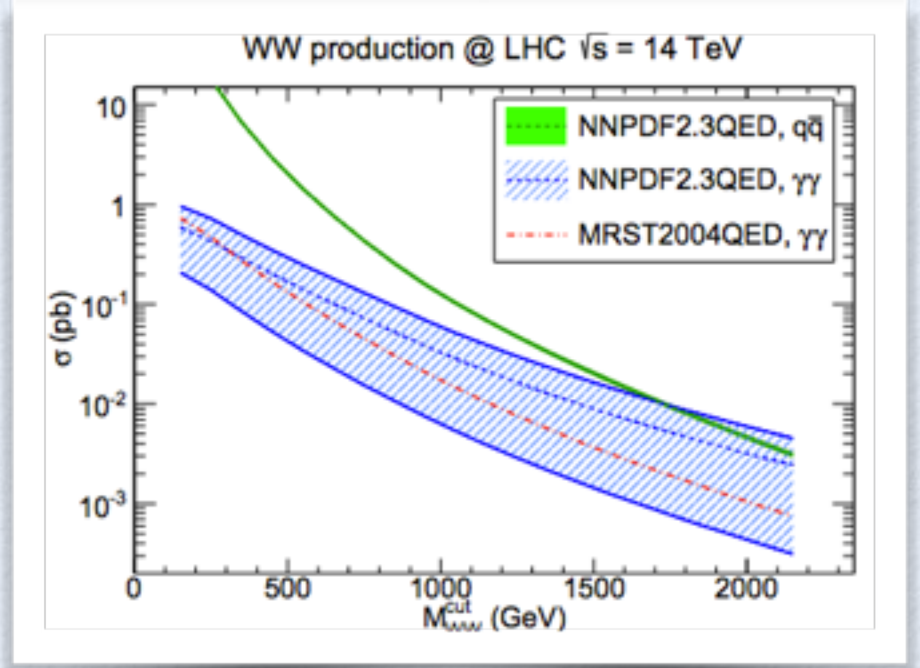
The NNPDF2.3 set

More constraints from the LHC

- WW production is phenomenologically relevant as a background for BSM searches
- At high M_{WW}, photon-induced contribution become relevant
- The large uncertainty at large M_{WW} comes from the large uncertainty of photon PDF for x > 0.1
- New LHC data give unique opportunity of constraining the photon in that region







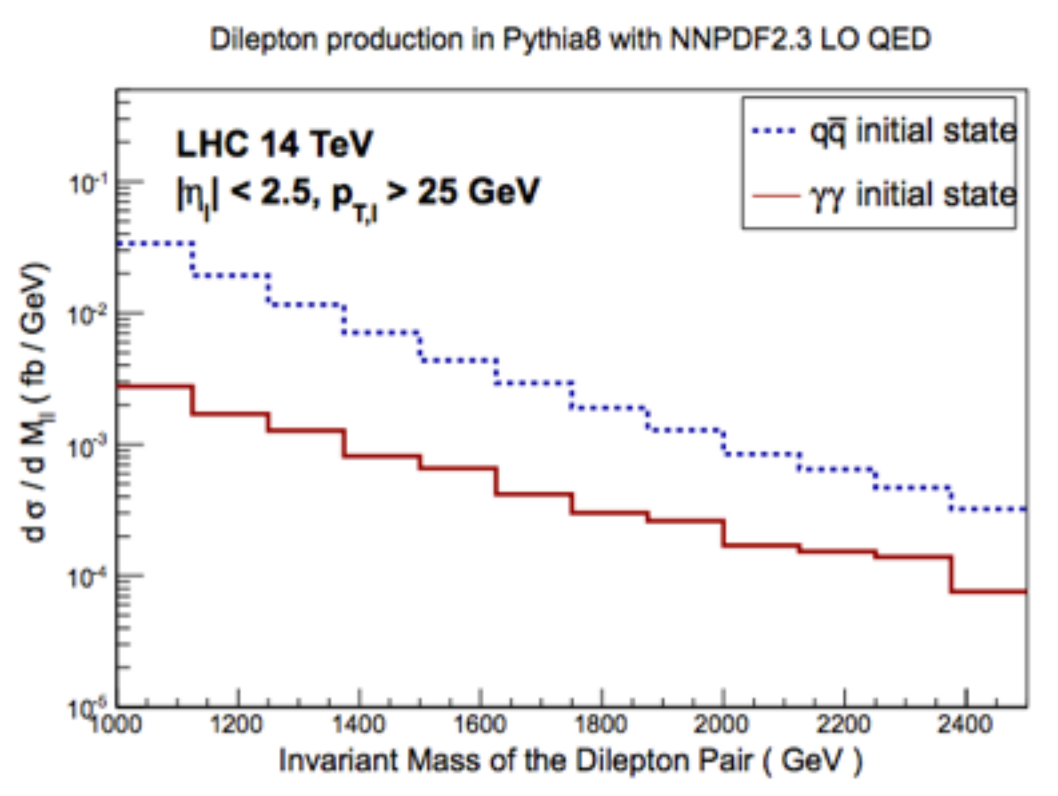
The NNPDF2.3 LO set

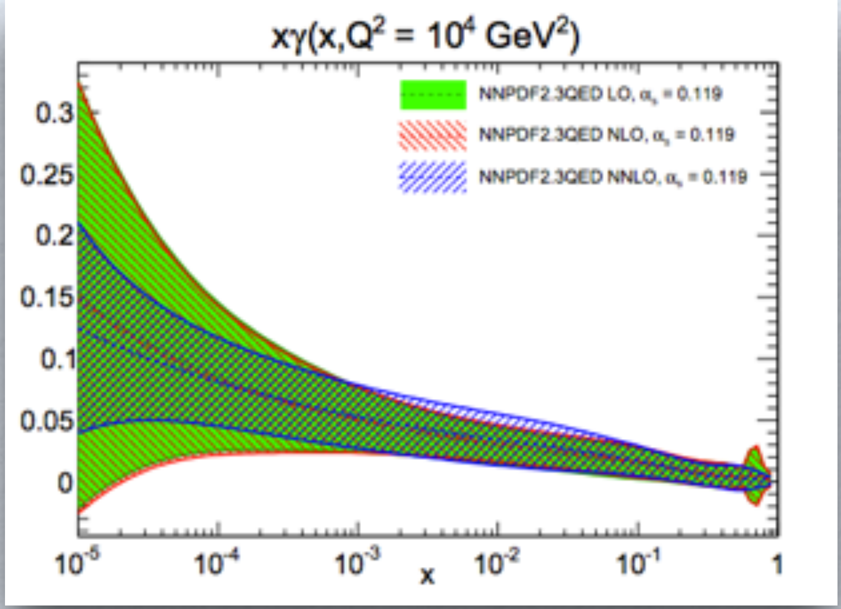
Monte Carlo event generators

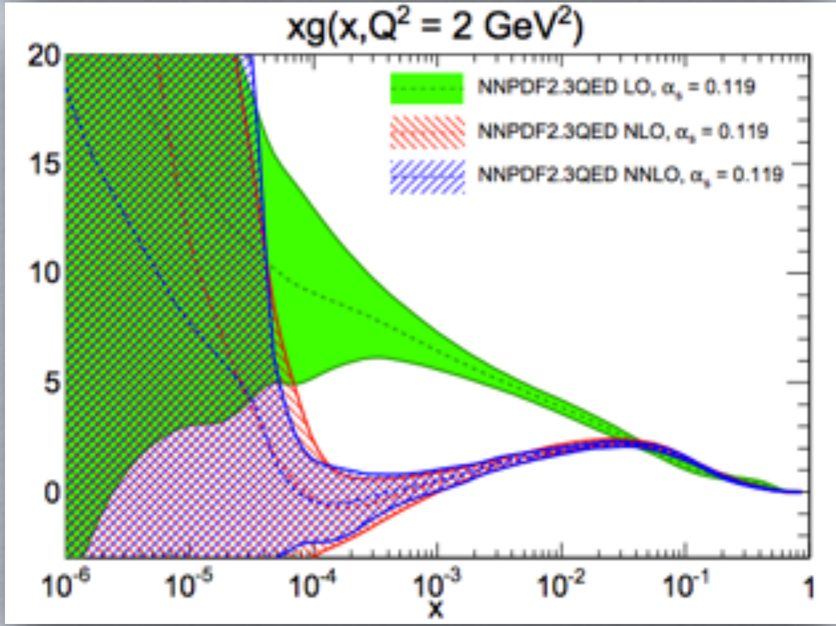
NNPDF2.3QED@LO : LO PDFs with QED corrections photon extracted from same data as NNPDF23 N(N)LO sets, internal set in Pythia8. [S. Carrazza et al, ArXiv: 1311.5887] Photon-initiated contribution relevant at large invariant mass. Important for new physics searches.

Employed in the **Monash 2013** tune of Pythia8

[P. Skands et al, ArXiv: 1404.5630]







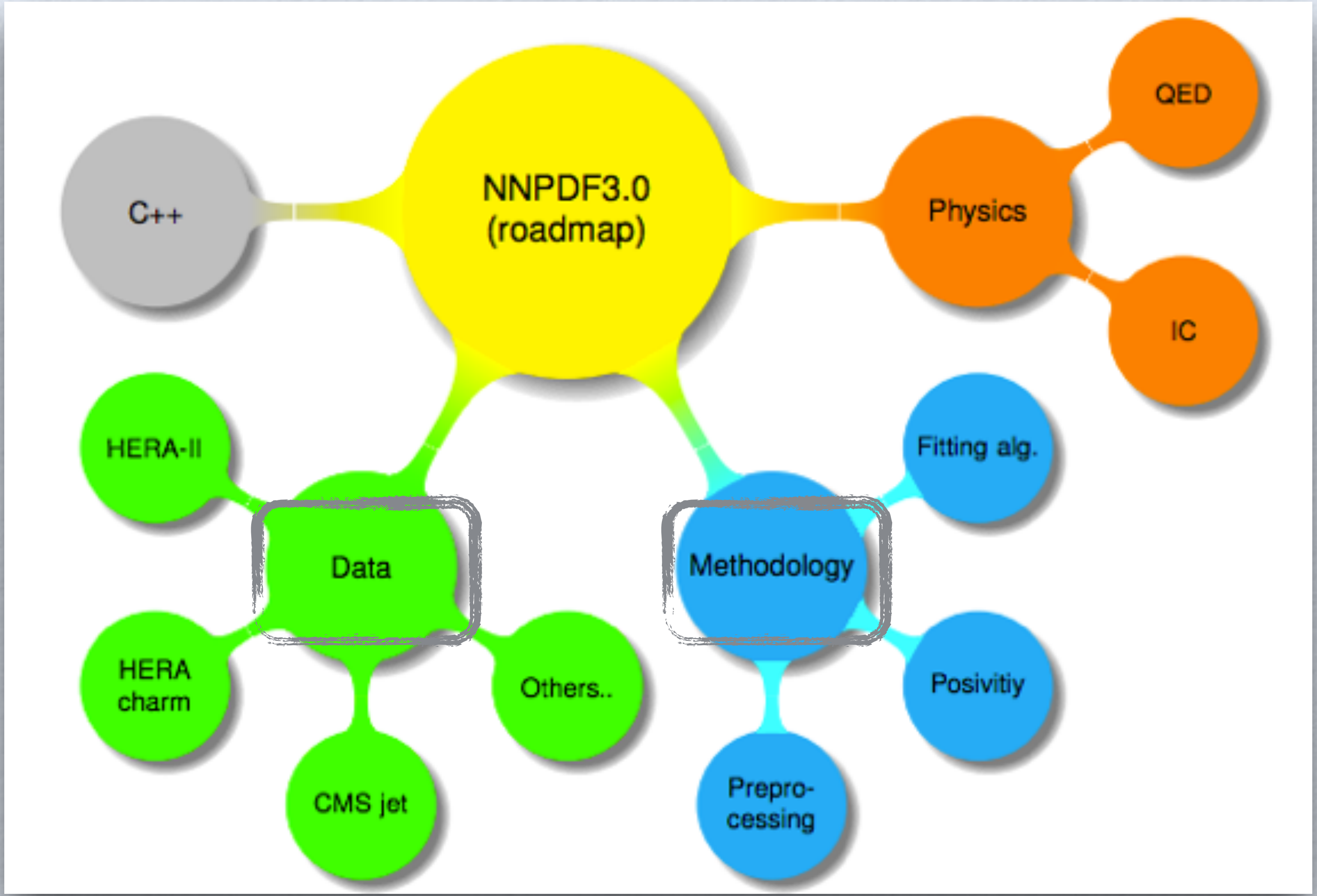
NNPDF3.0

All details in arXiv:1410.8849

The NNPDF3.0 set

Two years of hard work

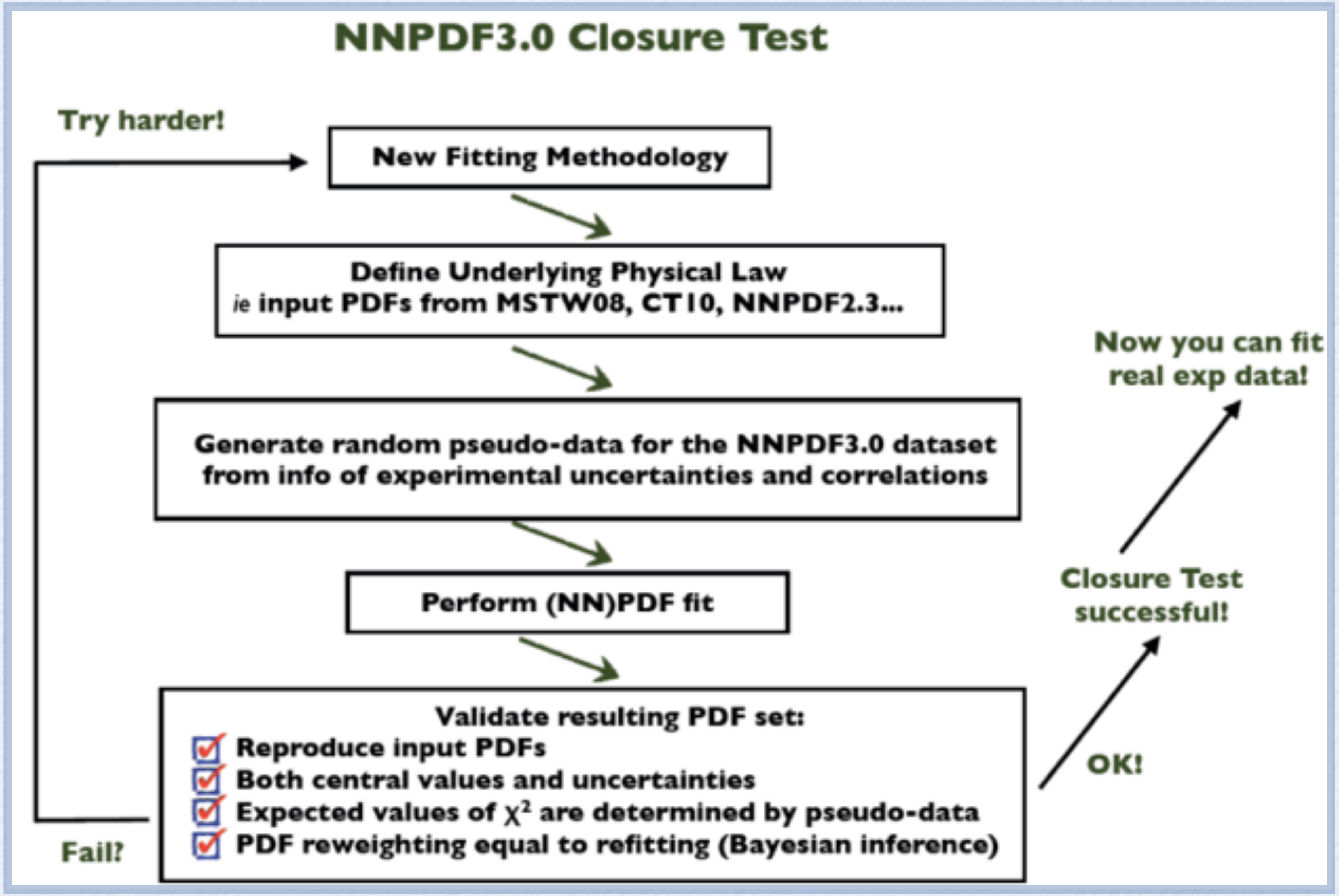
- Major update
- Code completely re-written in c++
- Completely re-designed fitting methodology based on closure test with known underlying physical law
- Tested Weight Penalty method based on iterative Bayesian regularization
- More than 1000 new data points from HERA II and LHC



Methodological uncertainty

The closure test

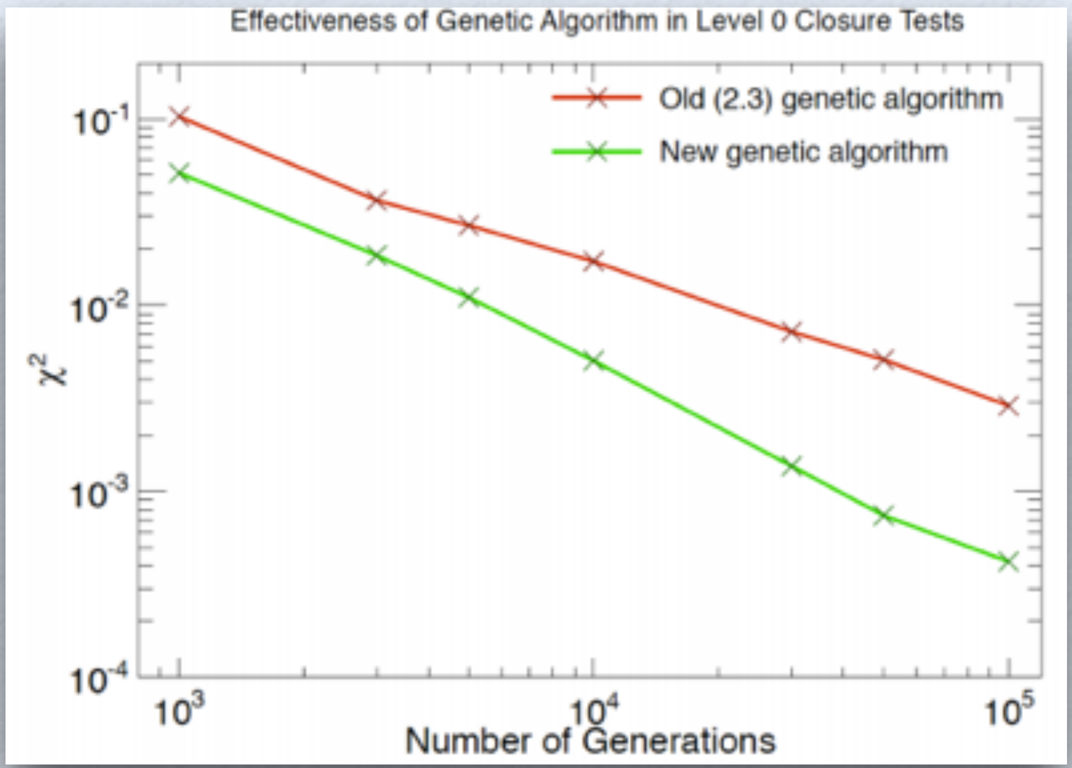
At current level of experimental precision, it is important to minimise and possibly kill methodological uncertainty. How?



Methodological uncertainty

The closure test

LEVEL 0: no fluctuation on pseudo-data, no Monte Carlo replica generation. Each datapoint equal to the MSTW true value and uncertainties assumed equal to experimental ones. Fit: must find $\chi^2 = 0$



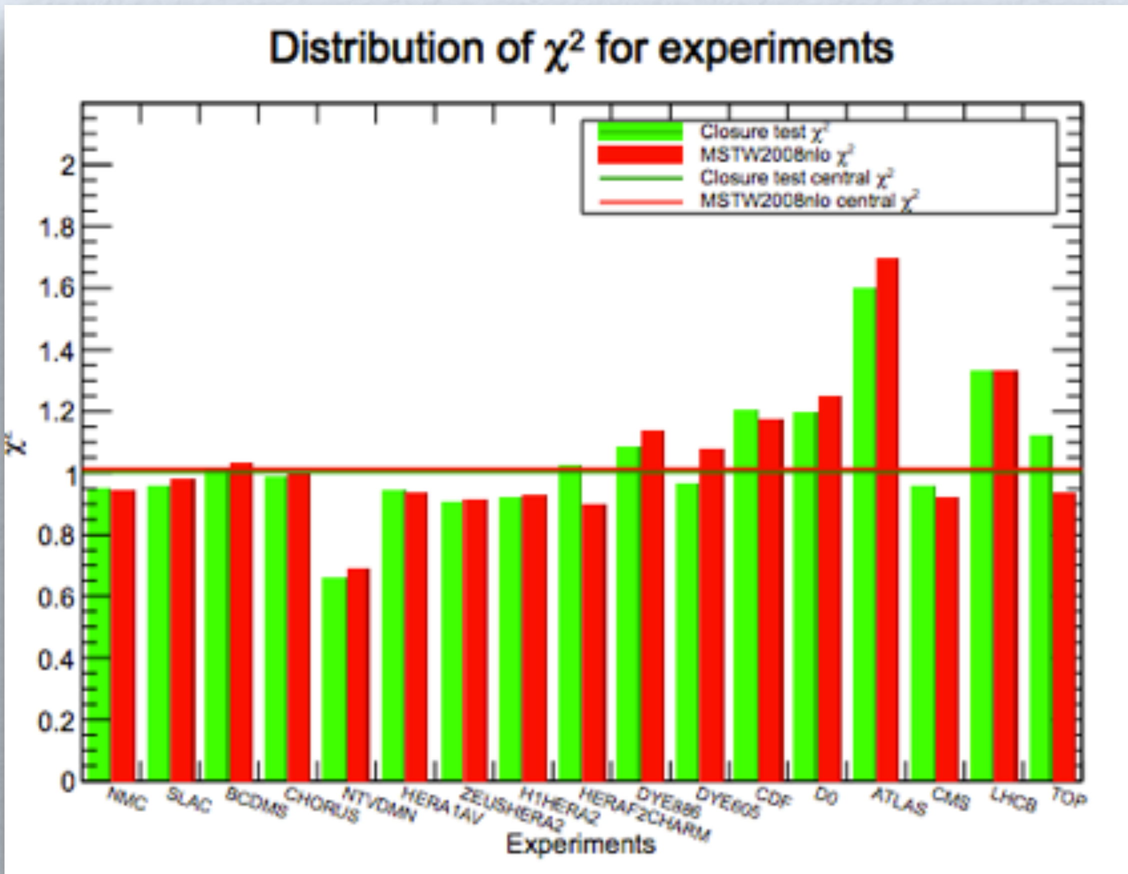
- ☑ Central values of input PDFs reproduced with arbitrary accuracy
- ☑ PDF uncertainties of fitted data points can become arbitrarily small
- ☑ Minimization in 3.0 more efficient that in 2.3

Methodological uncertainty

The closure test

LEVEL 1: fluctuation on pseudo-data, but no Monte Carlo replica generation.

LEVEL 2: fluctuation on pseudo-data and Monte Carlo replica generation.



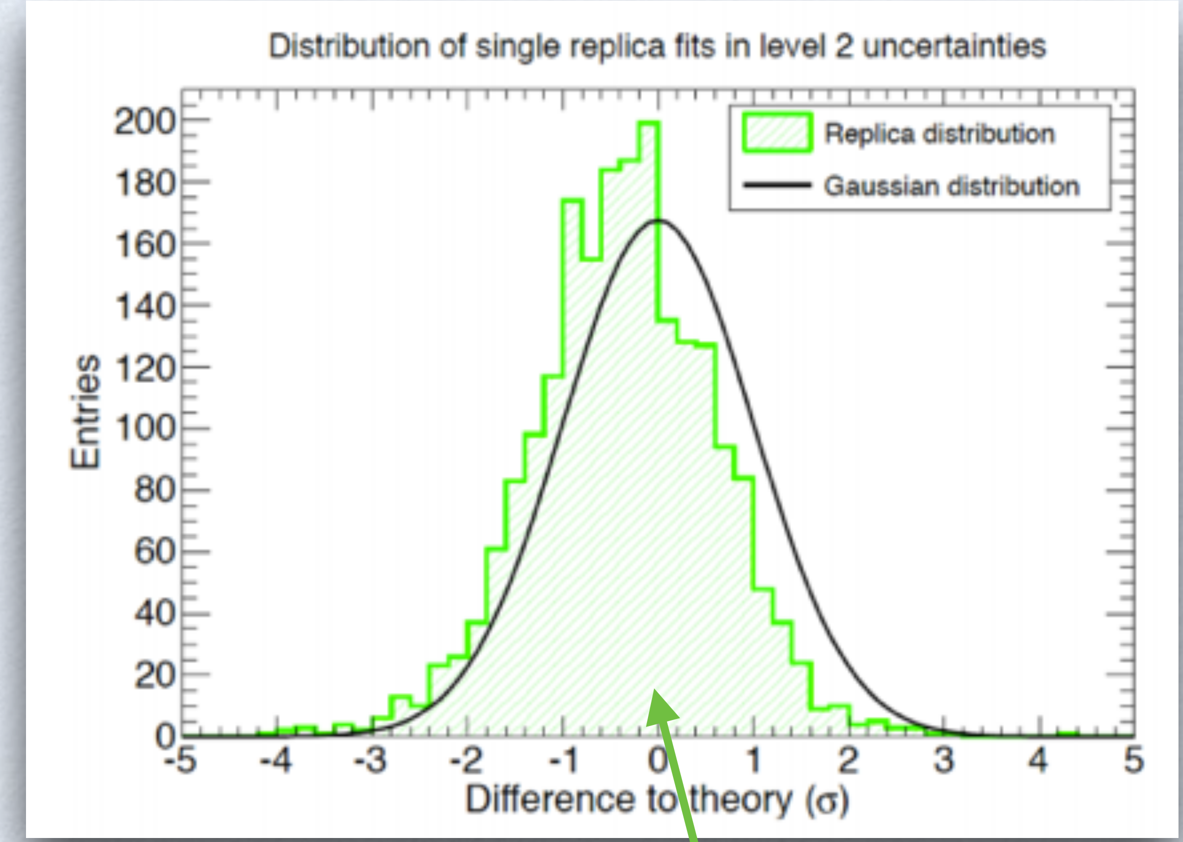
- Reproduce chi2 of input PDFs, both total and individual experiments
- Fitted PDFs central values fluctuate about input values by the same amount as expected from the size of the PDF uncertainties
- The central value of the fitted PDFs all in the one(two)sigma interval around 68%(95%) of the times (averaging over x and flavors)

Methodological uncertainty

The closure test

LEVEL 1: fluctuation on pseudo-data, but no Monte Carlo replica generation.

LEVEL 2: fluctuation on pseudo-data and Monte Carlo replica generation.

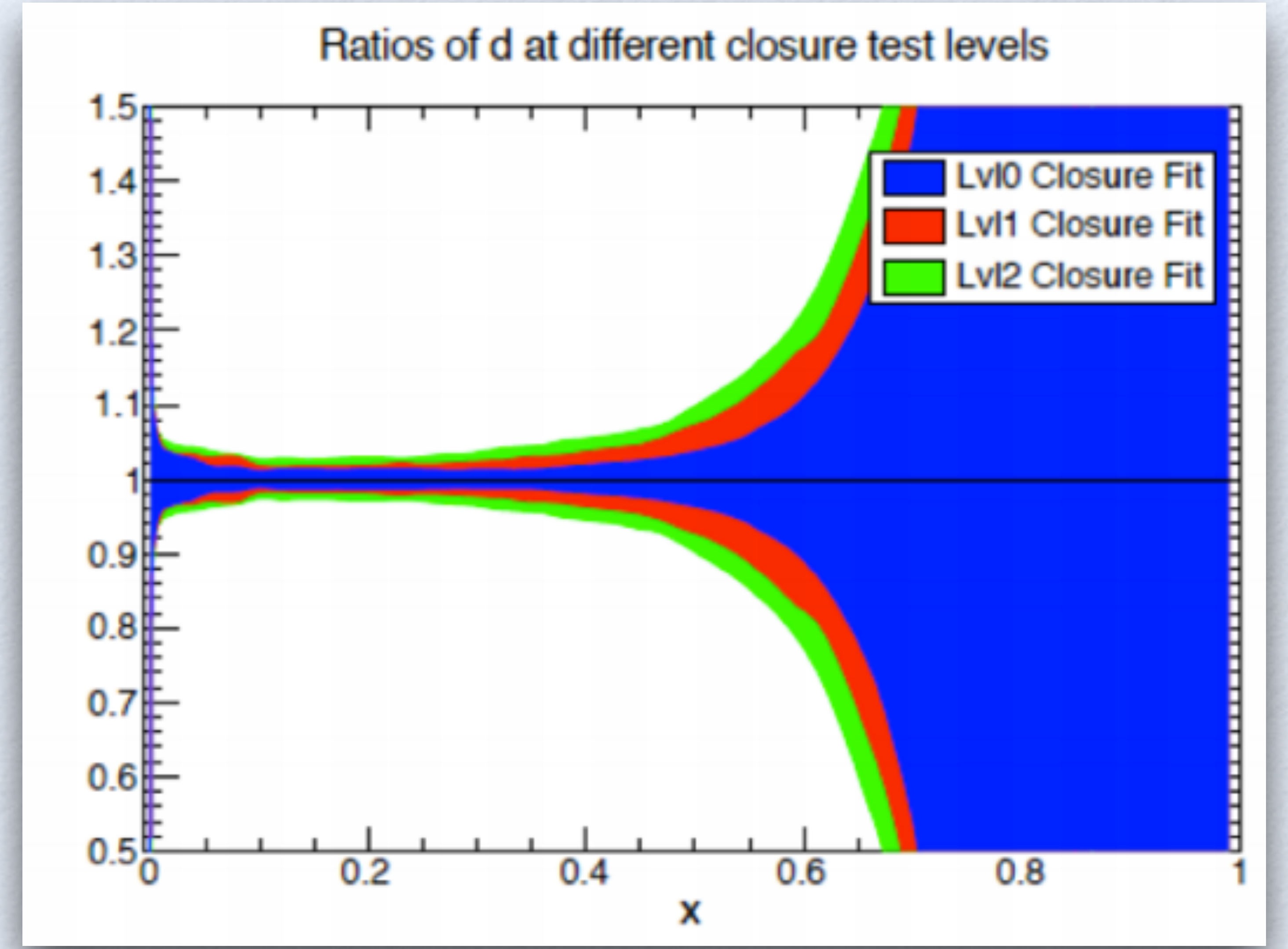


- Reproduce chi2 of input PDFs, both total and individual experiments
- Fitted PDFs central values fluctuate about input values by the same amount as expected from the size of the PDF uncertainties
- The central value of the fitted PDFs all in the one(two)sigma interval around 68%(95%) of the times (averaging over x and flavors)

Difference between fit and input PDF central values in unit of PDF uncertainties

Methodological uncertainty

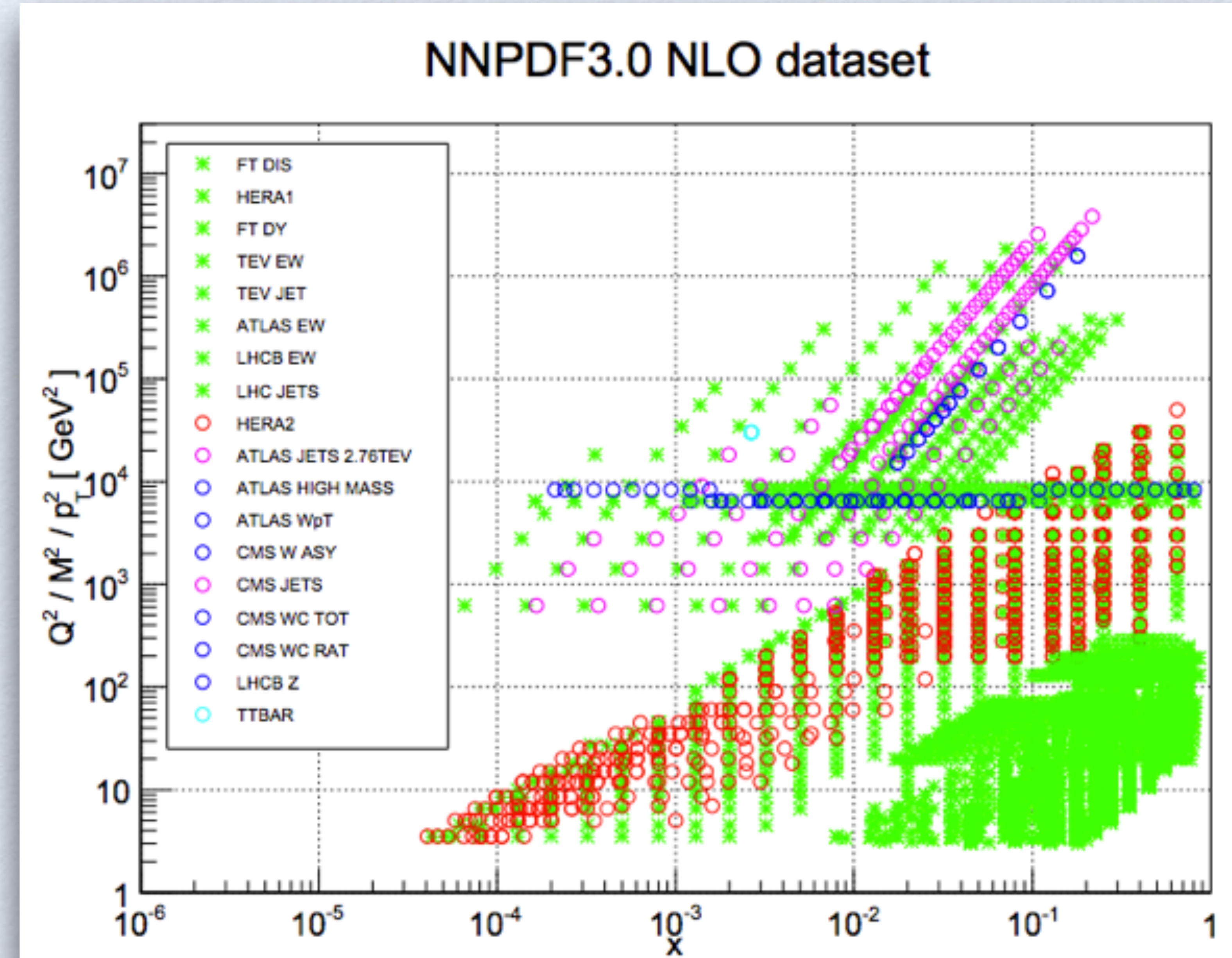
The closure test



- Comparing level 0, 1 and 2 closure tests provide a quantitative determination of components of the total PDF uncertainty
- L0: extrapolation uncertainty
- L1: functional uncertainty
- L2: experimental uncertainty

The NNPDF3.0 set

Data set



NNPDF23

HERAII

new LHC EW

new LHC jets

LHC tt

The NNPDF3.0 set

Data set

HERAII

- H1 high Q2 data [JHEP 1209 (2012) 061] -> quark at medium and large x
- H1 data at lower CoM energy (Ep = 460,575 460 GeV) [Eur.Phys.J. C71 (2011) 1579]
- H1 high inelasticity data [Eur.Phys.J. C71 (2011) 1579]
- Combined HERA charm production [Eur.Phys.J. C73 (2013) 2311] -> gluon at small/medium x
- ZEUS NC and CC with positron beams [Eur.Phys.J. C70 (2010) 945]

ATLAS

- Jets 2.76 TeV and 7 TeV [Eur.Phys.J. C73 (2013) 2509] -> stronger constraint
- High mass Drell-Yan [Phys.Lett. B725 (2013) 223] -> quark-antiquark separation at large x
- W pT distributions

CMS

- Jets 7 TeV 5fb^{-1} [Phys.Rev. D87 (2013) 112002] -> gluon at large x
- DY double differential distributions [JHEP 12 (2013) 30] -> flav. separation
- Muon charge asymmetry 4.7fb^{-1} [ArXiv:1312.6283]
- W + charm [JHEP 02 (2014) 013] -> strangeness

LHCb

- Large rapidity Z distributions [JHEP 1302 (2013) 106]

+ Total ttbar cross section from ATLAS and CMS (7 and 8 TeV)

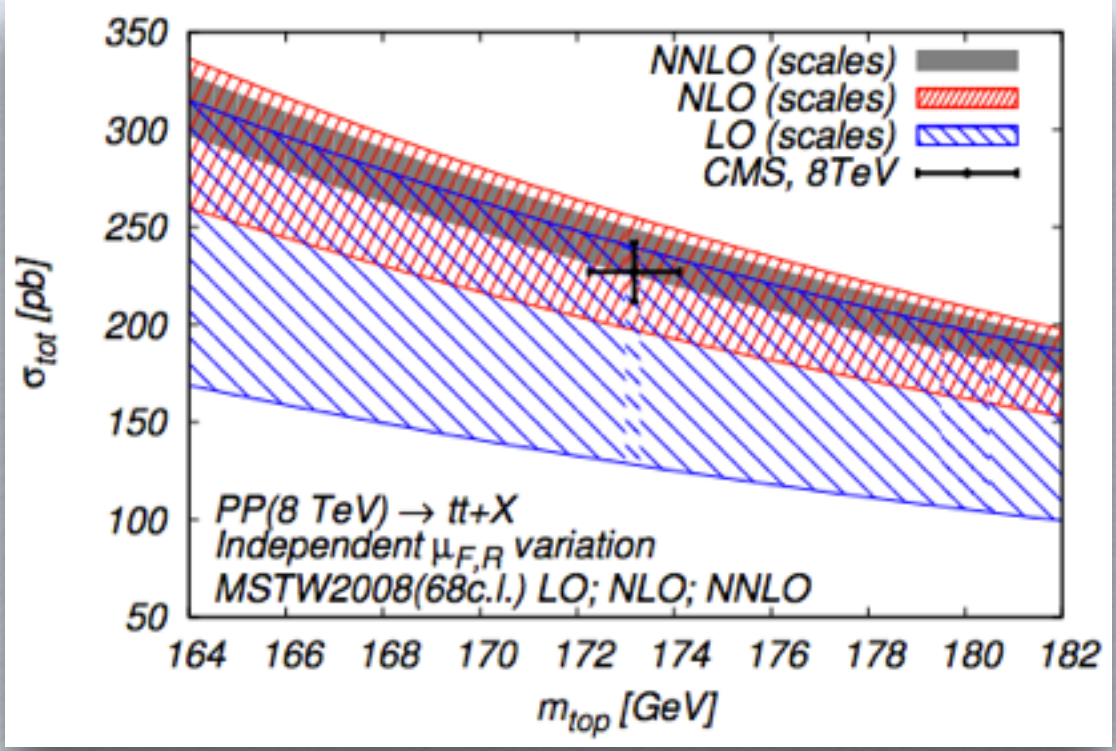
O(1000) NEW data points!
Over 4000 data points:
FastKernel + FASTNLO/APPLgrid systematically employed!

The NNPDF3.0 set

Theoretical aspects: higher order corrections

- NNLO calculations are essential to reduce theoretical uncertainties in PDF analyses
- Recently important progress has been made on some key processes

  - Full NNLO top quark production cross section is available (TOP++2.0) and differential distributions are expected soon -> gluon at large x
  - H+1j also available now at NNLO, important milestone towards Z,W+1j -> gluon & quark separation



Czakon et al., ArXiv:1305.3892

Czakon, Fiedler, Mitov PRL 110 (2013) 25

Boughezal et al, JHEP1306 (2013) 072

If NNLO calculations available, include NNLO corrections via C-factors

- Top quark very promising observable to provide constraint on the gluon

  Czakon et al JHEP 1307 (2013) 167

  Beneke et al JHEP 1207 (2012) 194

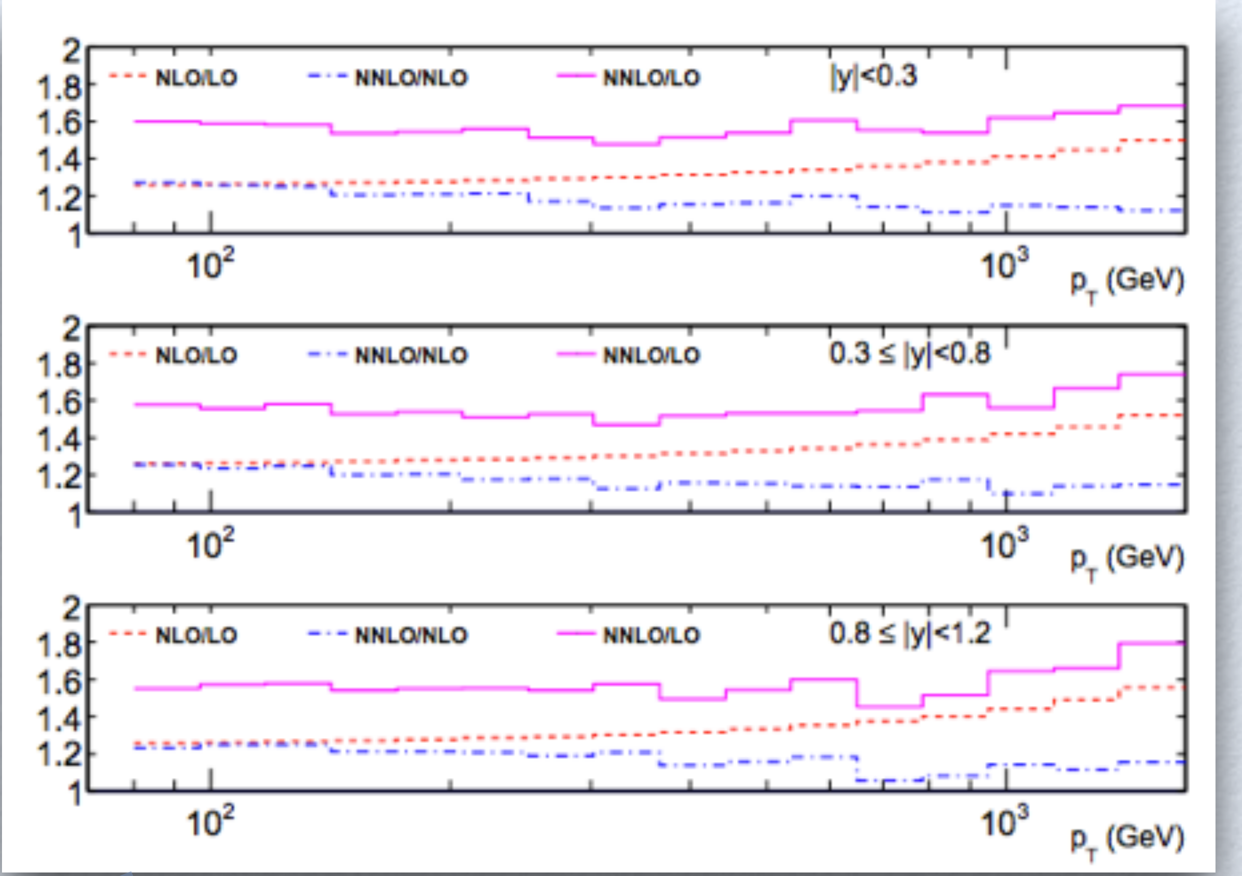
  Alekhin et al Phys.Rev. D89 (2014) 054028]

The NNPDF3.0 set

Theoretical aspects: higher order corrections

- NNLO calculations are essential to reduce theoretical uncertainties in PDF analyses
- Recently important progress has been made on some key processes

  - Full NNLO top quark production cross section is available (TOP++2.0) and differential distributions are expected soon -> gluon at large x
  - H+1j also available now at NNLO, important milestone towards Z,W+1j -> gluon & quark separation
  - NNLO inclusive jet production in the gg channel has been completed -> gluon & quark at large x



Gehrmann-De Ridder et al, Phys.Rev.Lett. 110 (2013) 16

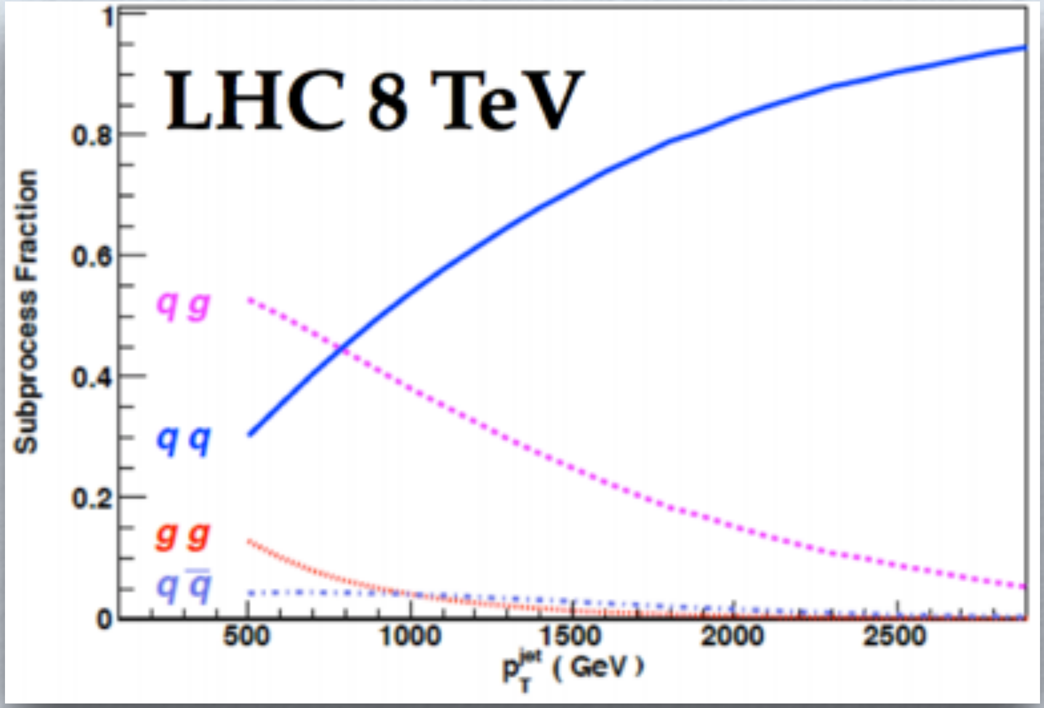
For jets full NNLO calculation is not yet available but...

In gg channel up to 20-25% enhancement of NNLO wrt NLO result

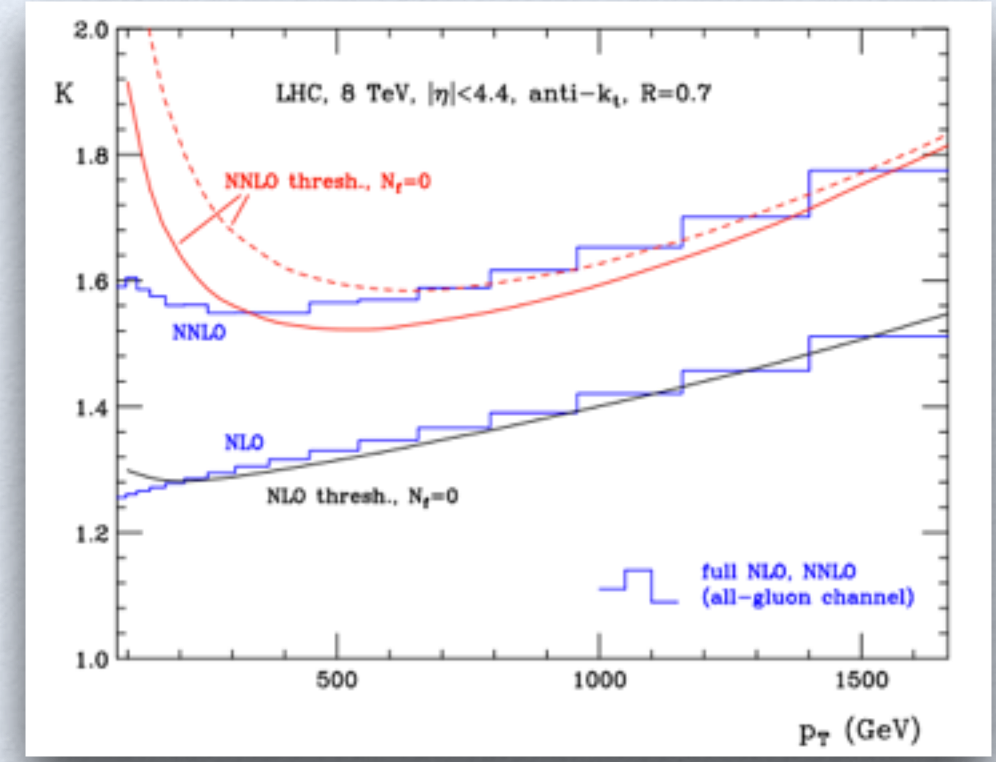
The NNPDF3.0 set

Theoretical aspects: jet cross section

- At the LHC gluon-gluon channel is small at medium-large pT
- Approximate NNLO results can be derived from the improved threshold calculation, reasonable at large pT and expected to break down at small pT
- Approx NNLO is an improved version of Kidonakis et al. [Phys.Rev. D63 (2001) 054019]



[De Florian et al, Phys.Rev.Lett. 112 (2014) 082001]

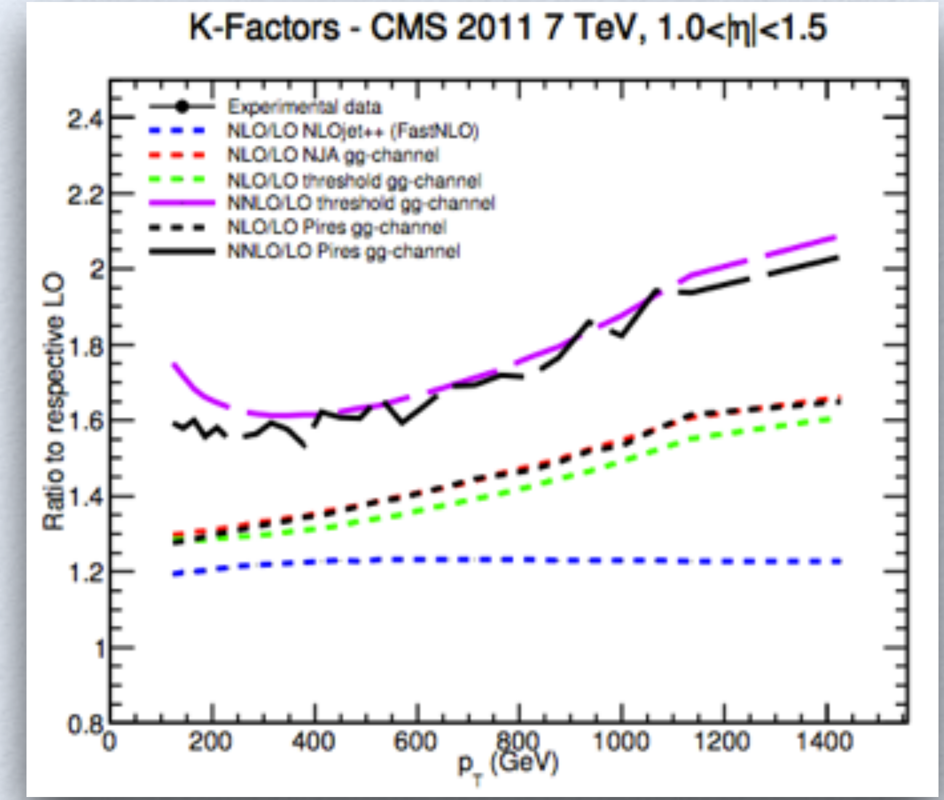


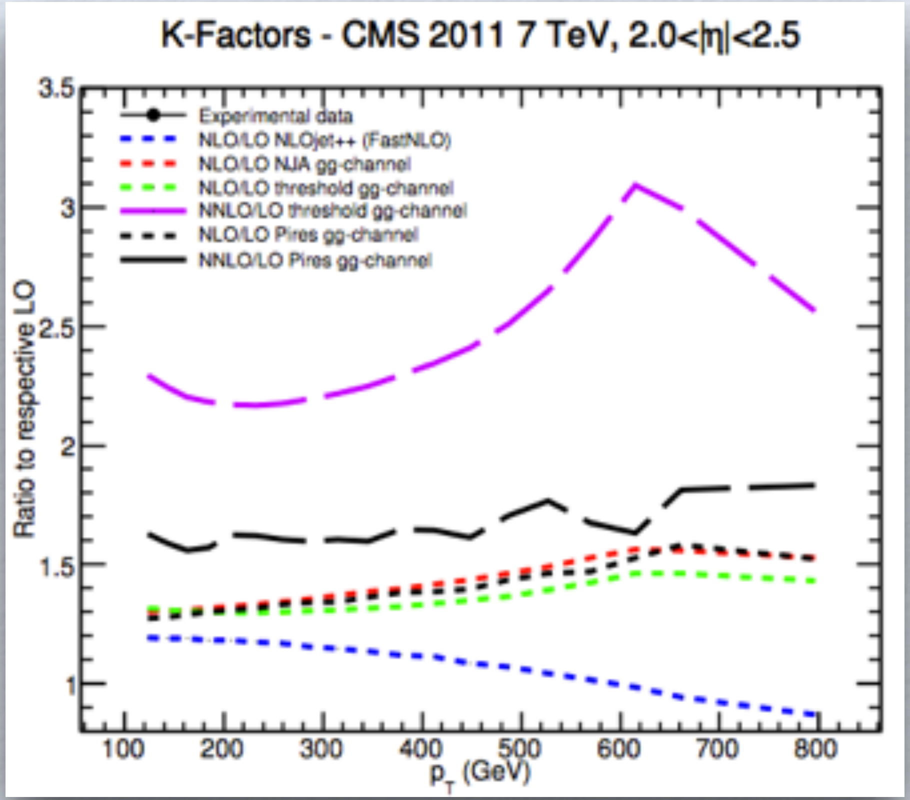
- Comparison between NNLO approx and full NNLO in the gg channel can determine for which value of pT and ηNNLO approx can be trusted
- This assumes NNLO K-factors similar in all channels

The NNPDF3.0 set

Theoretical aspects: jet cross section

Plots courtesy of J. Pires and S. Carrazza



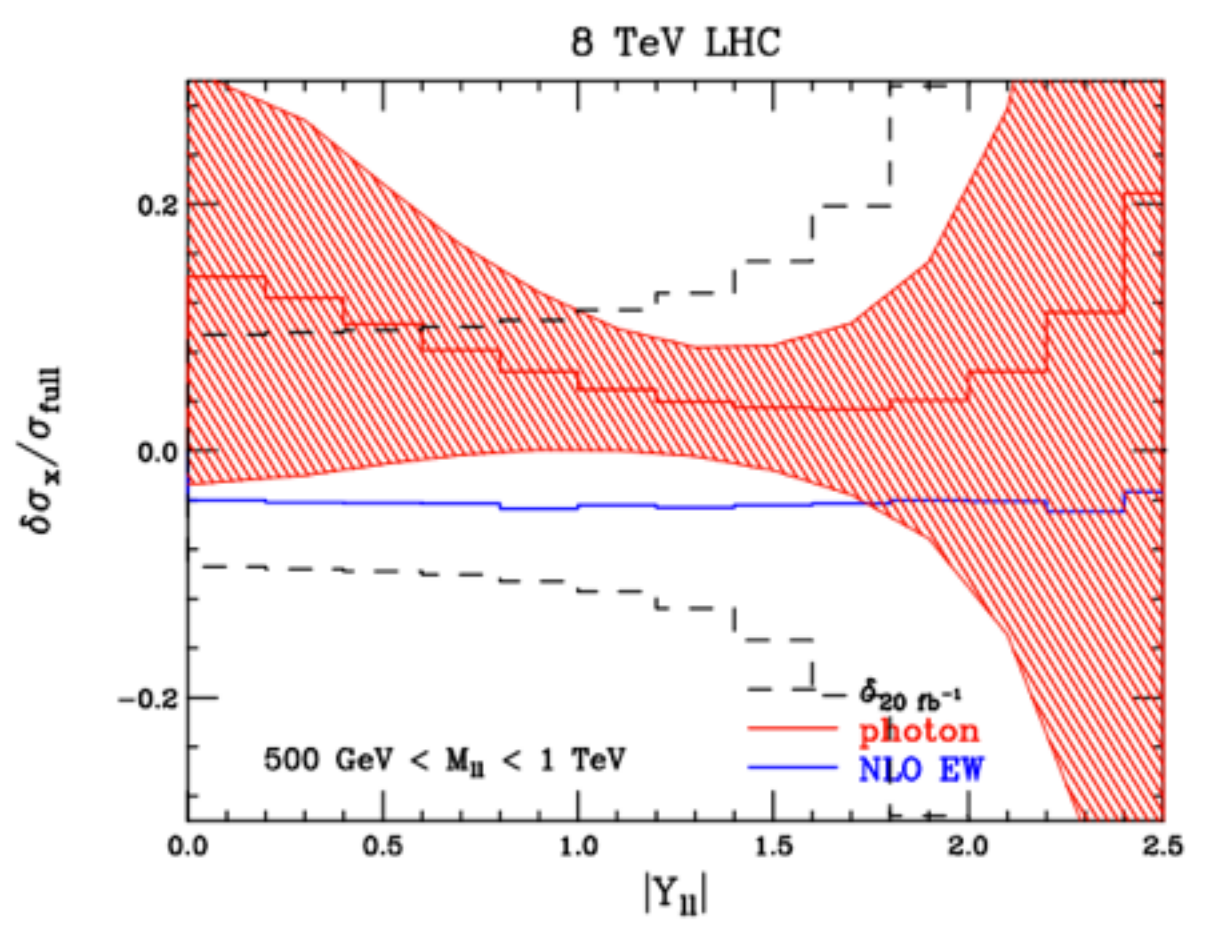


- Until exact NNLO result available, jet dat at small jet transverse momentum and large pseudo-rapidity have better been cut out from NNPDF30 NNLO fits as NNLO_threshold is not suitable in that region.
- Tevatron data and ATLAS 2010 data less affected due to different validity range and larger uncertainties
- Otherwise we include them by computing the NNLO_threshold/NLO C factors

The NNPDF3.0 set

Theoretical aspects: higher order corrections

- QED and EW corrections can also be easily computed with FEWZ3.1 **[Li, Petriello, Phys.Rev. D86 (2012) 094034]**
- They can be sizable especially at large invariant mass
- QED corrections affected by large uncertainty induced from uncertainty on photon PDF



Boughezal, Liu, Petriello, ArXiv:1312.4535

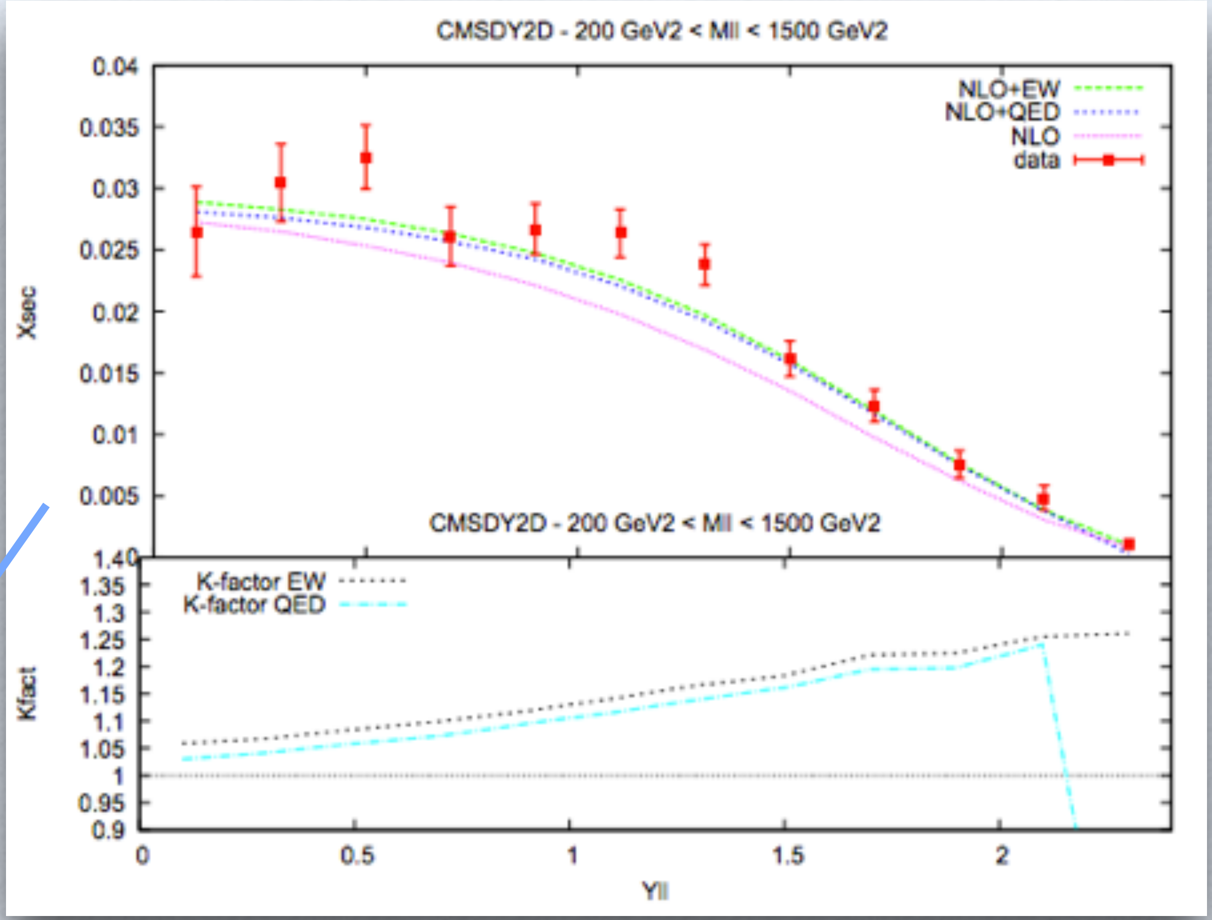
The NNPDF3.0 set

Theoretical aspects: higher order corrections

- QED and EW corrections can also be easily computed with FEWZ3.1
[Li, Petriello, Phys.Rev. D86 (2012) 094034]

- They can be sizable especially at large invariant mass

- QED corrections affected by large uncertainty induced from uncertainty on photon PDF



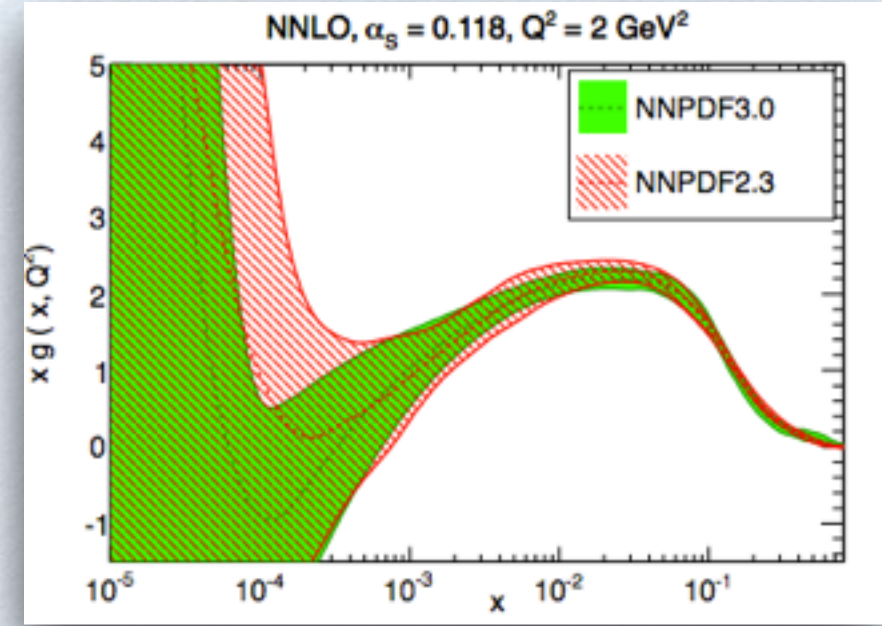
Pure EW C-factors included in theoretical predictions at NLO and NNLO in NNPDF30 fit

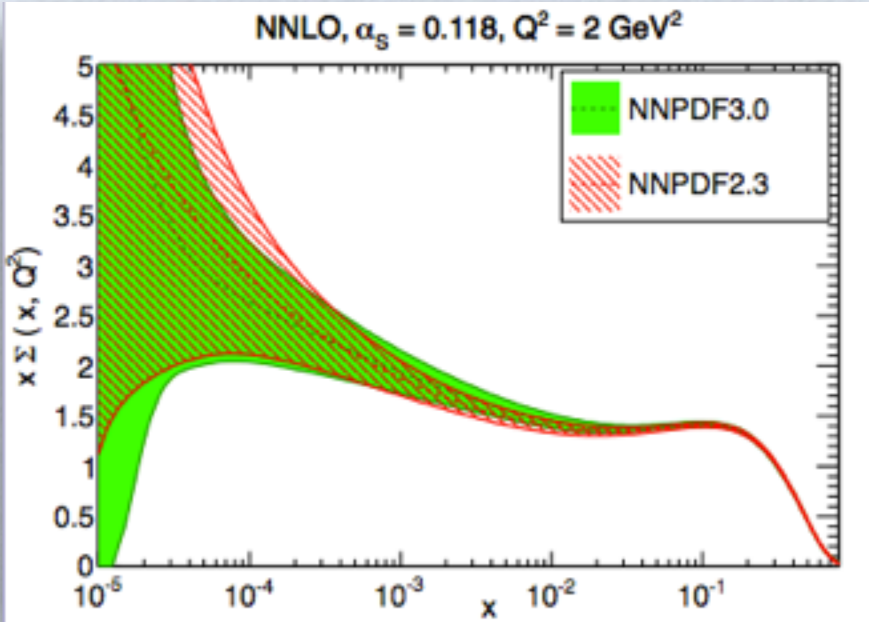
$$C^{\mathrm{NNLO}}_{\mathrm{fact}} = \frac{\hat{\sigma}_{\mathrm{NNLO}} \otimes f^i_{\mathrm{NNLO}}}{\hat{\sigma}_{\mathrm{NLO}} \otimes f^i_{\mathrm{NNLO}}}$$

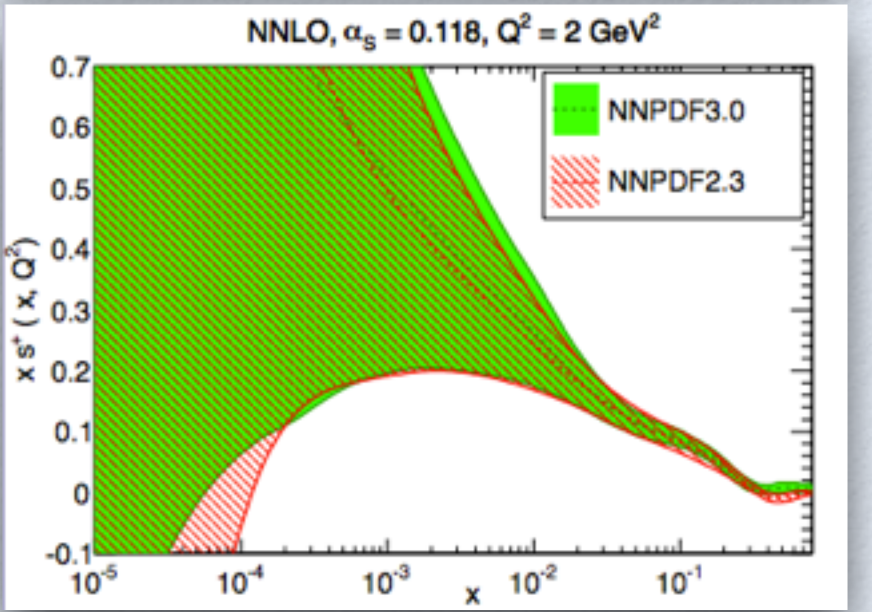
$$C^{\mathrm{EW}}_{\mathrm{fact}} = \frac{\hat{\sigma}_{\mathrm{NLO+EW}} \otimes f^i_{\mathrm{NLO}}}{\hat{\sigma}_{\mathrm{NLO}} \otimes f^i_{\mathrm{NLO}}}$$

The NNPDF3.0 set

Comparison with NNPDF2.3



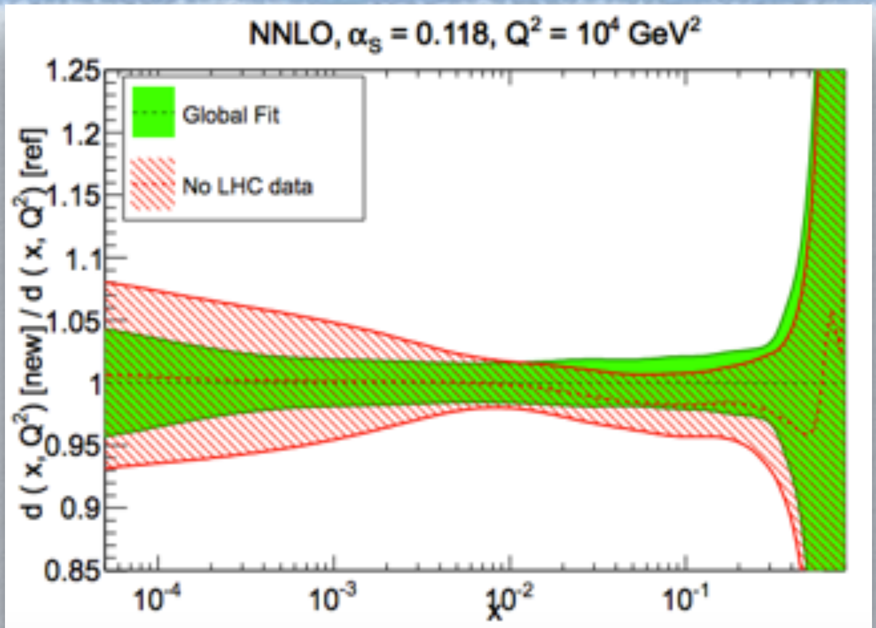


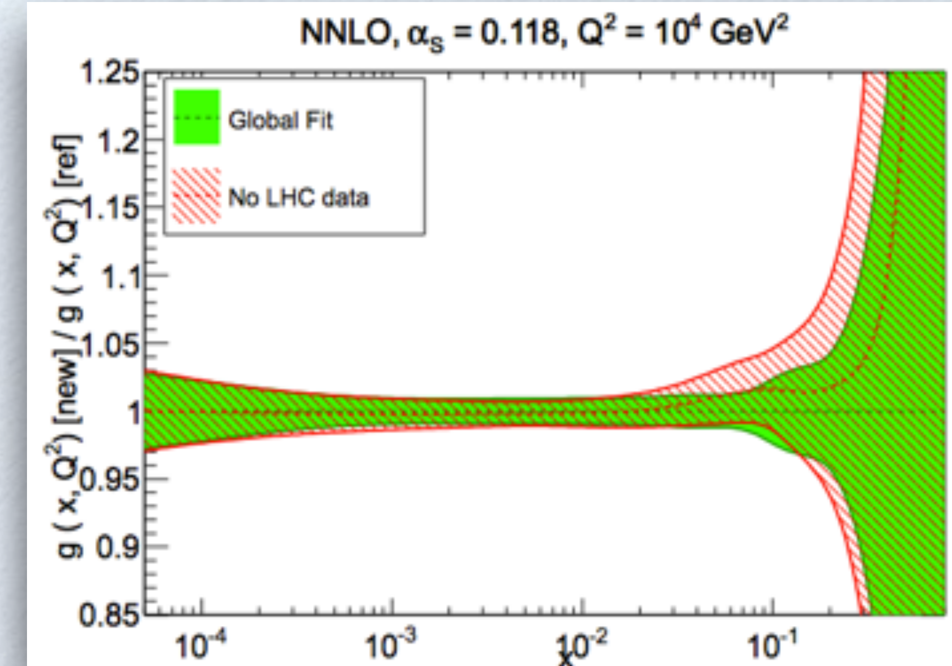


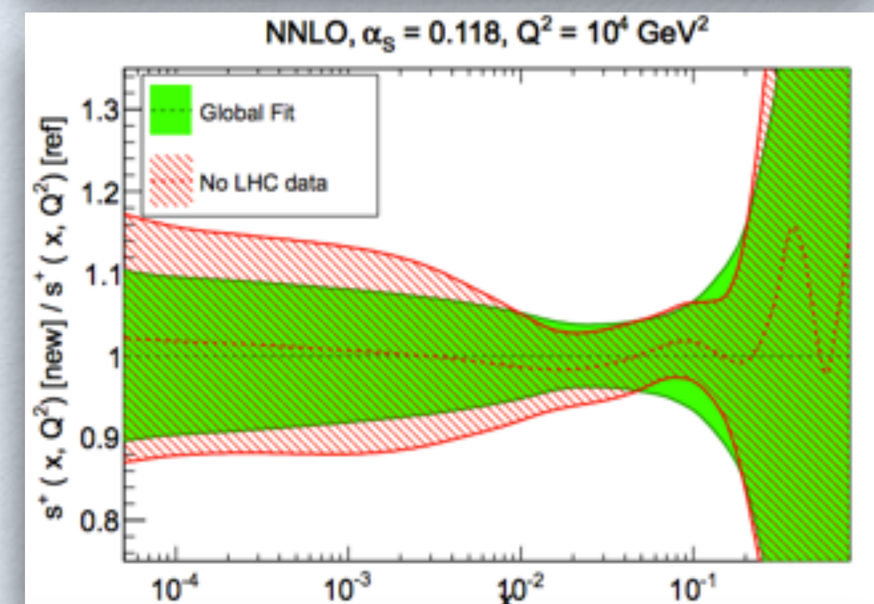
- Reasonable agreement with NNPDF2.3 and NNPDF3.0: expected given that all new HERA and LHC data are already well described by NNPDF2.3
- Differences between central values at 1σ level at most
- PDF uncertainties are reduced, effect most visible in gluon, down quark and strangeness

The NNPDF3.0 set

Effect of the LHC data



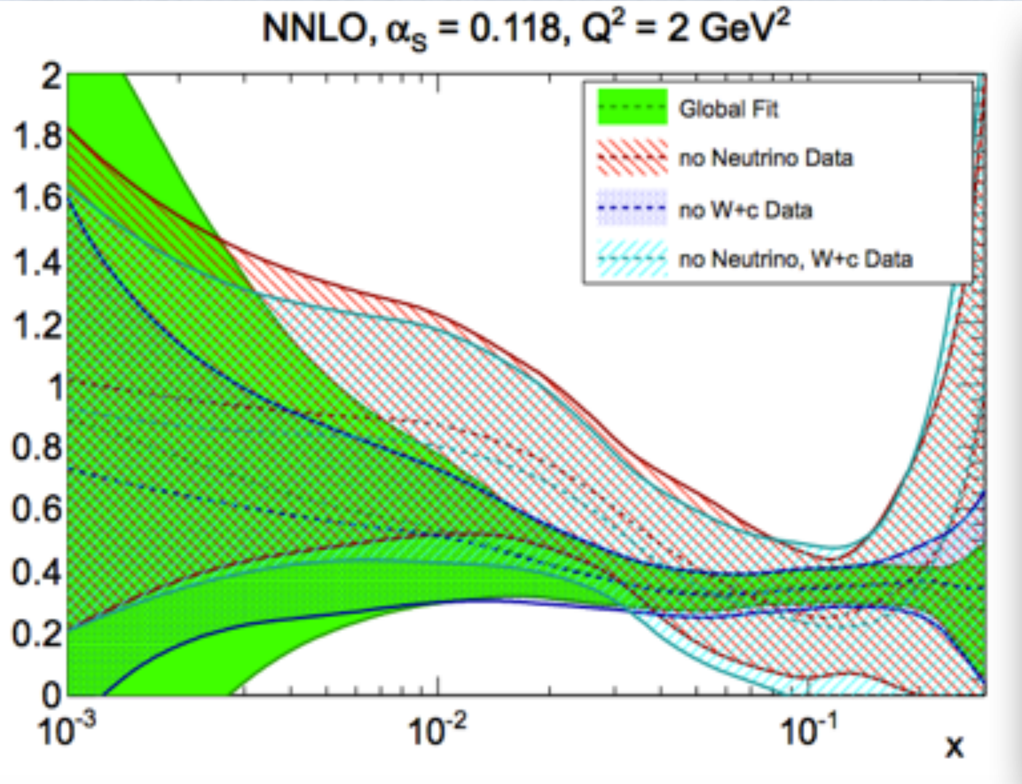


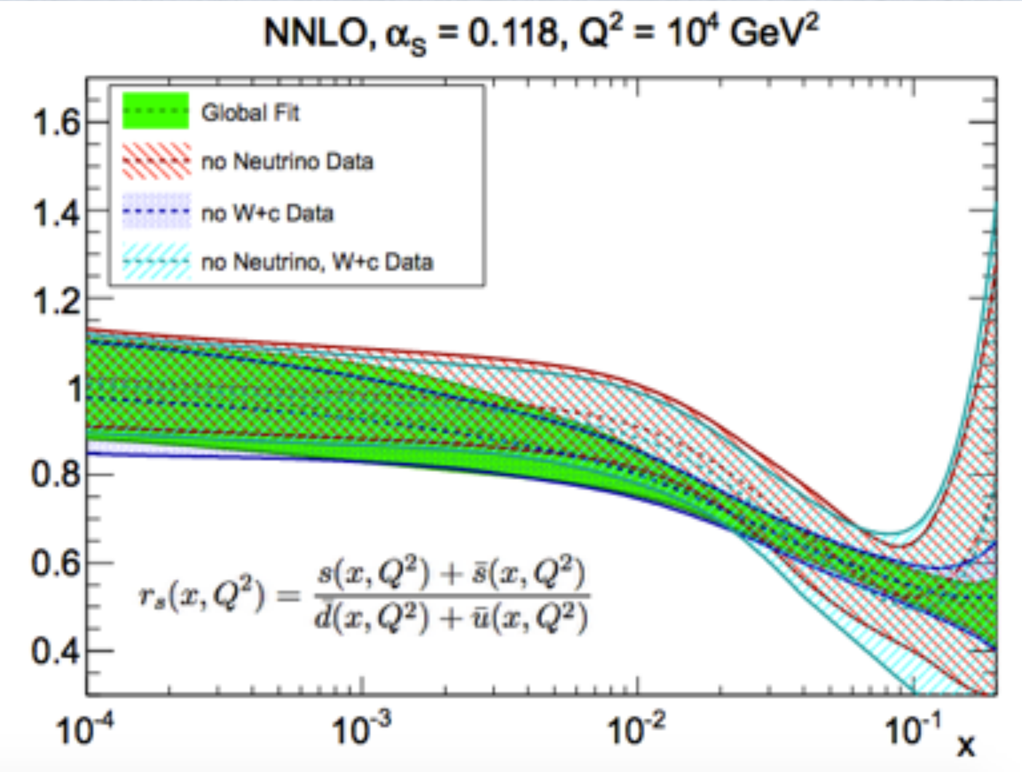


- PDF uncertainty of large-x gluon reduced by inclusion of jet and top quark data
- Uncertainty of light quarks at small x reduced by DY data and W+c
- Description of LHC data, already good with NNPDF2.3 improves in NNPDF3.0

The NNPDF3.0 set

Strangeness



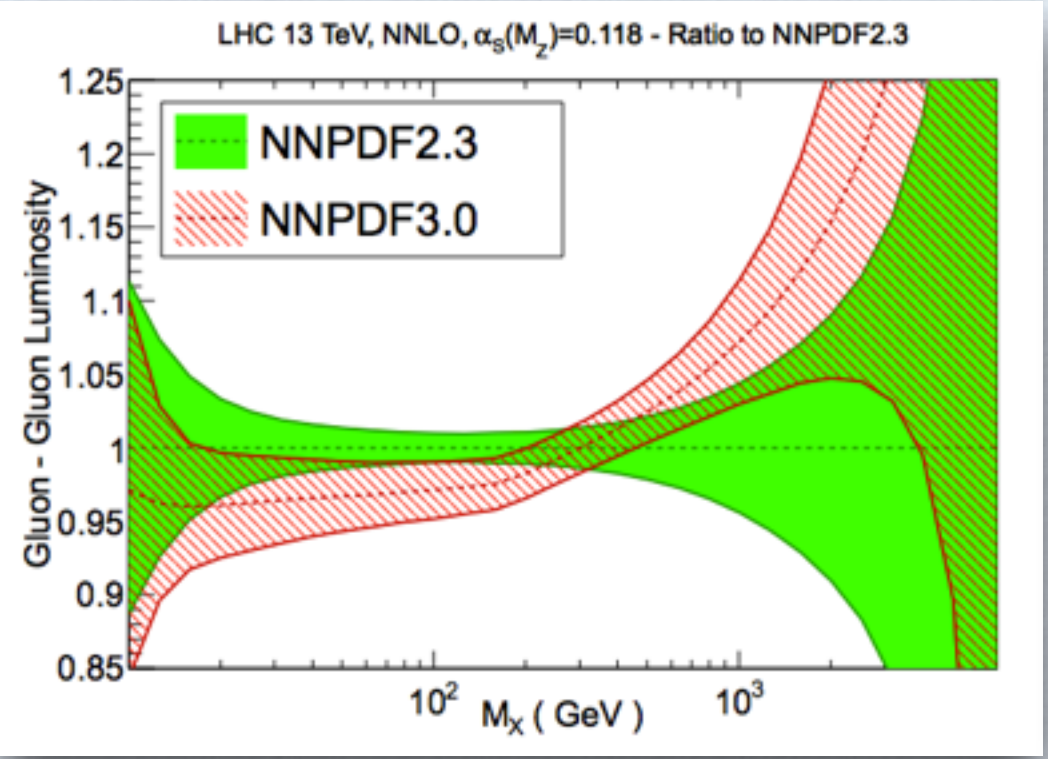


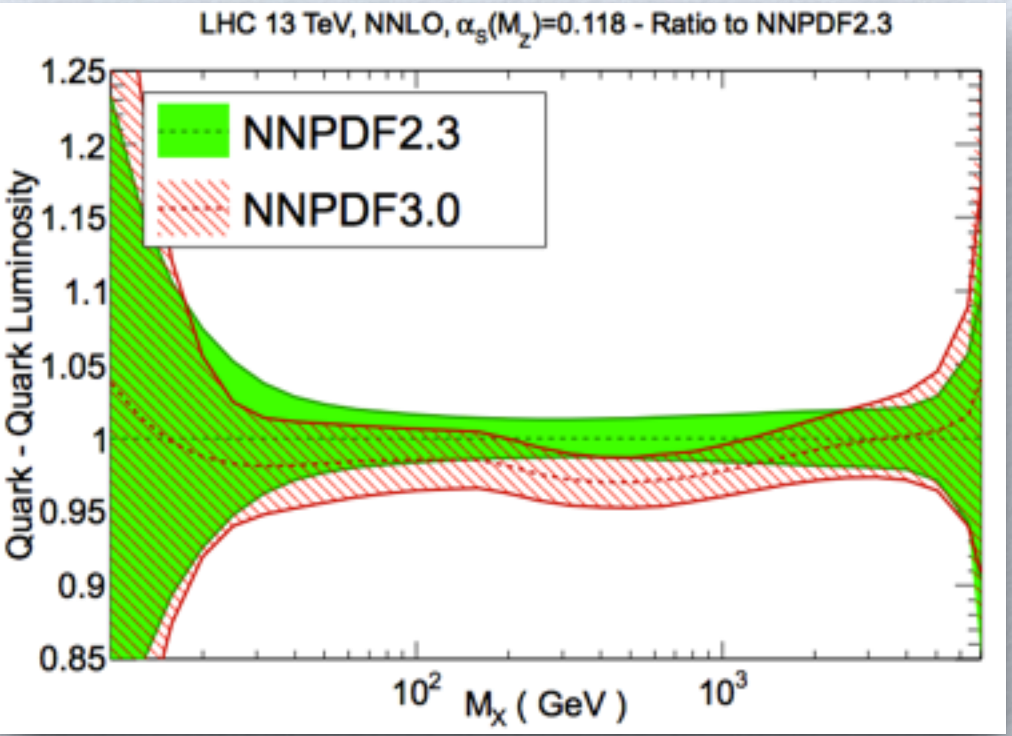
	$\chi^2_{\rm exp}$			
	Global	No neutrino	No $W+c$	No neutrino/$W+c$
CHORUS	1.13	*3.87*	1.09	*3.45*
NuTeV	0.62	*4.31*	0.66	*6.45*
ATLAS W, Z 2010	1.21	1.05	1.24	1.08
CMS $W+c$ 2011	0.86	0.50	*0.90*	*0.61*

No signs of tension between neutrino data and collider W+c data. Everything reconciled within large uncertainties

Phenomenology

Luminosities

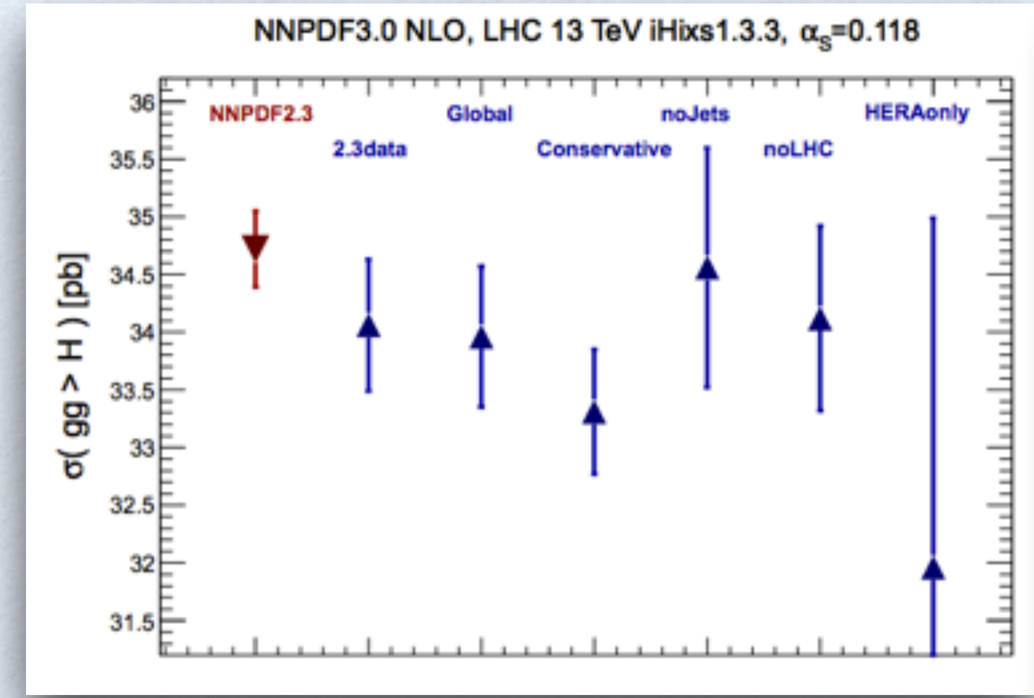


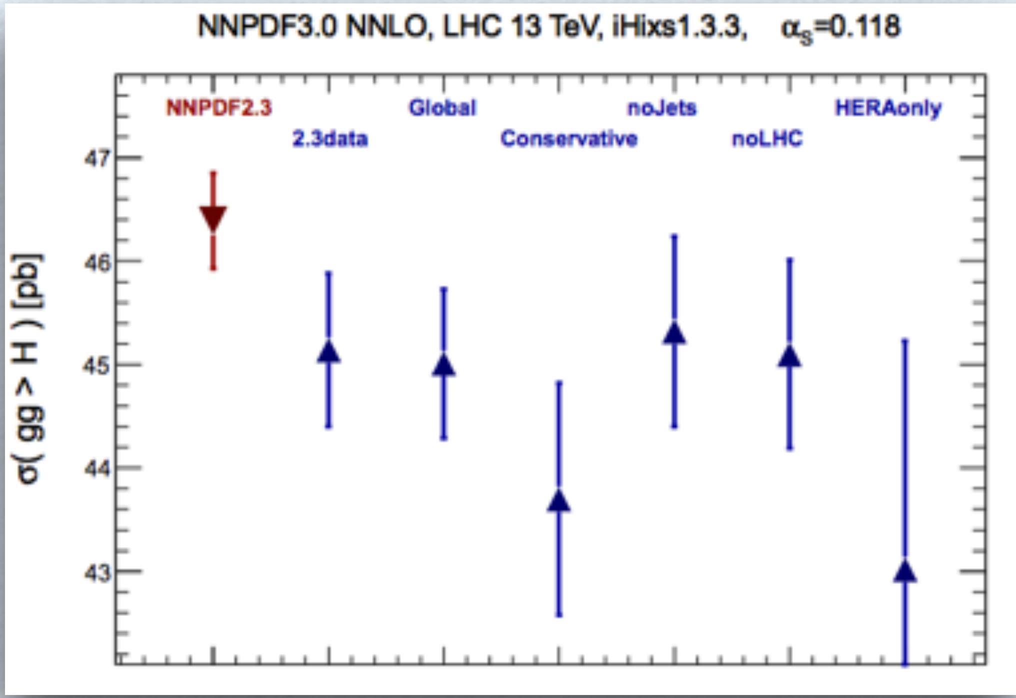


- PDF luminosities useful to translate differences in PDFs into differences in LHC cross sections
- QQ 3.0 luminosity softer for 300 GeV < M < 1 TeV → implication for heavy particle production
- GG 3.0 shifter down by 1σ for M < 200 → implication for gg > H

Phenomenology

Higgs production in gluon fusion

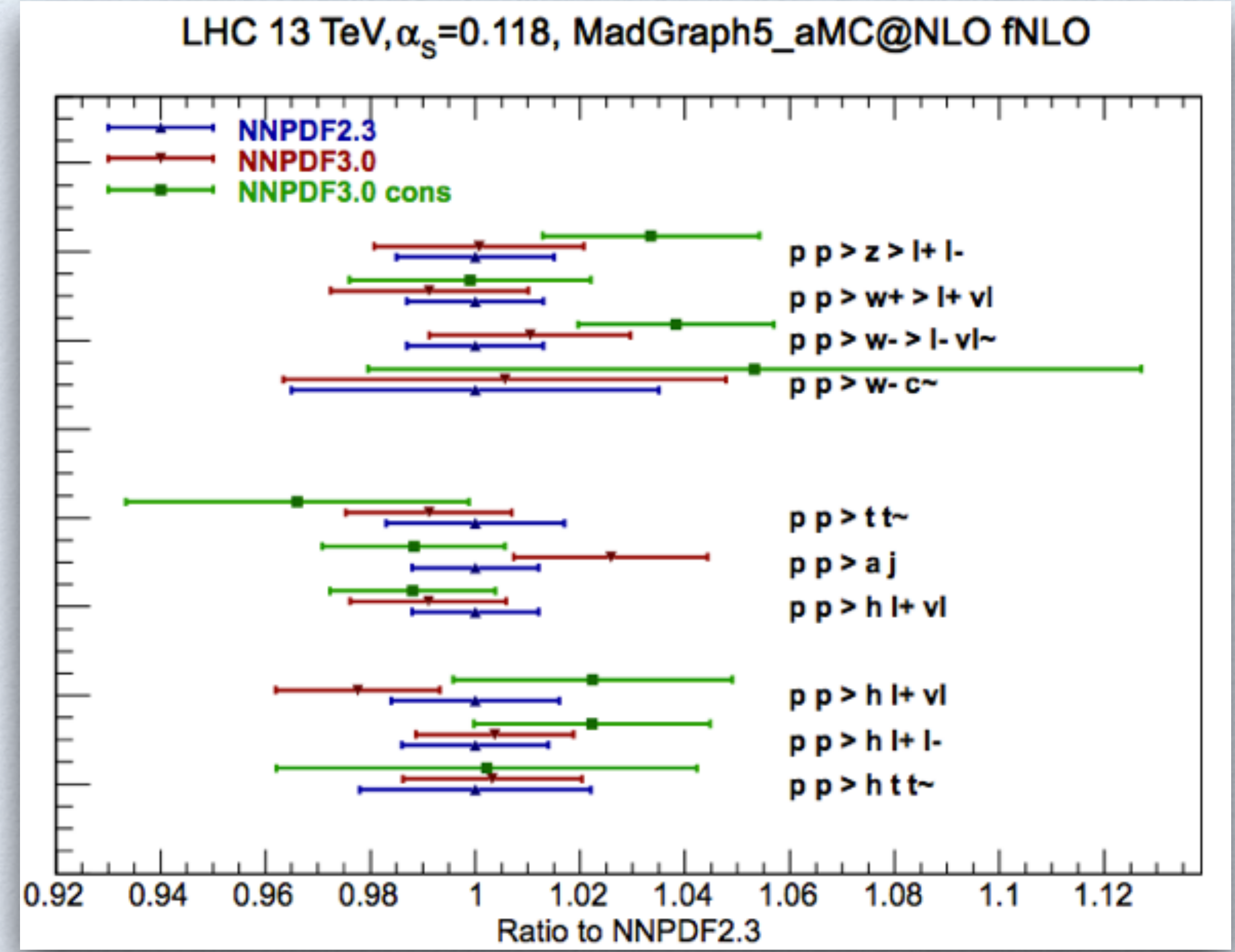




- Softer gluon-gluon luminosity leads to a decrease in the the ggH cross section at LHC 13 TeV
- The effect is most marked at NNLO rather than at NLO, with pull of ~ 1.5
- The ggH process is different from many other processes at LHC since there are no direct experimental constraints on the gluon at x ~ 0.01, thus predictions are very sensitive to methodology and choice of dataset
- In this case changes are most due to the change in methodology, now validated by closure tests

Phenomenology

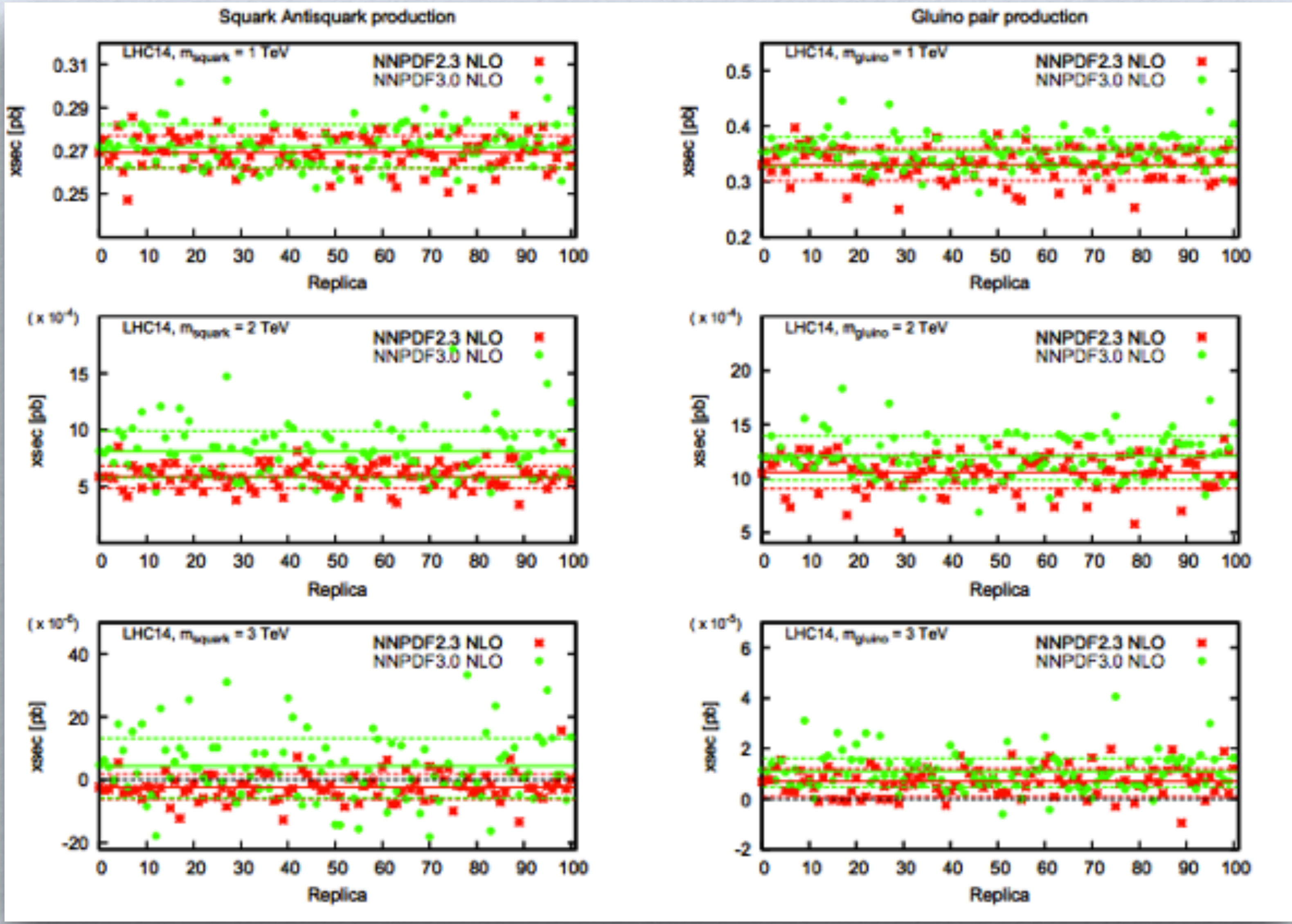
Other key processes



Phenomenology

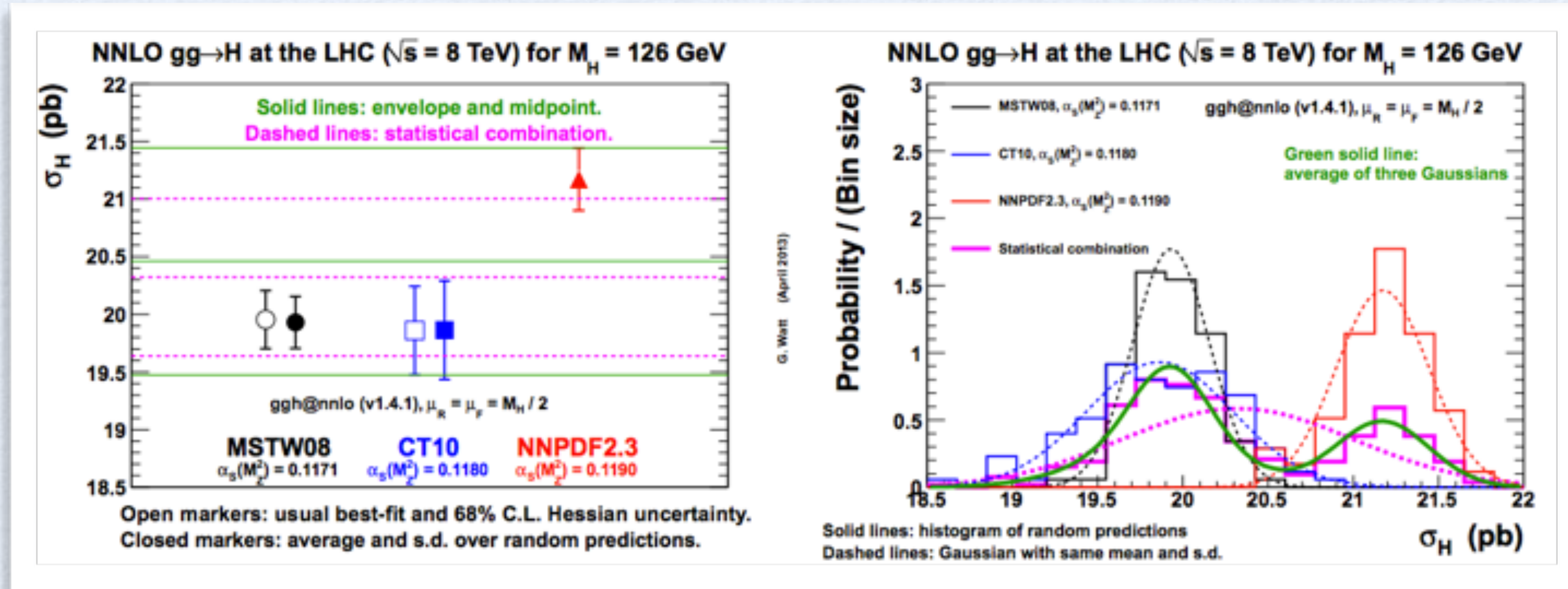
Positivity of BSM cross sections

Effect of extended positivity range in the fit via Lagrange multiplier: no more negative cross sections for heavy new particle production



Phenomenology

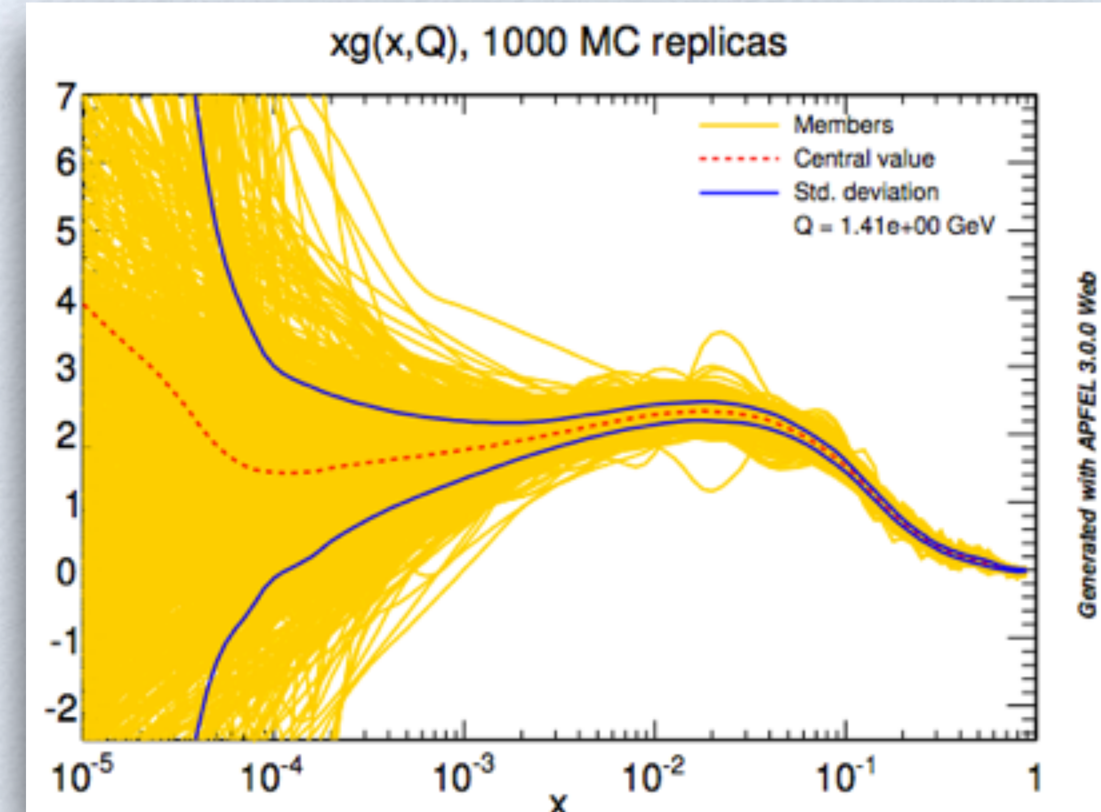
How to combine sets?



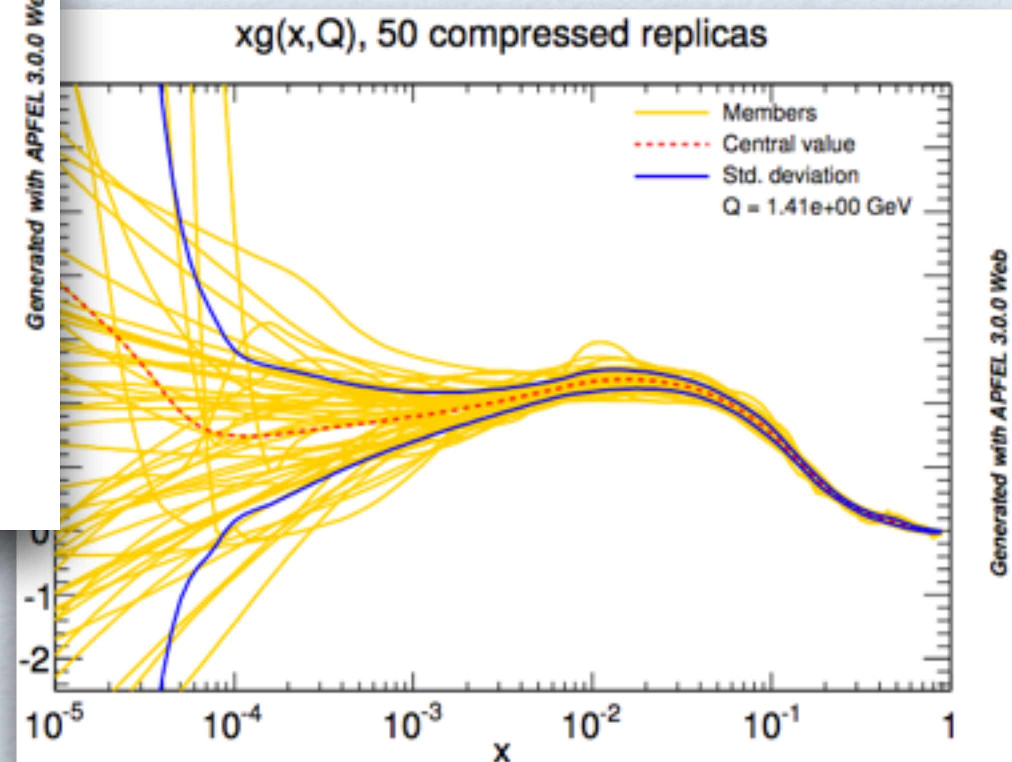
- Envelopes [PDF4LHC prescription **arXiv 1101.0538**]
- Statistical combination from different PDF groups generating MC sets. **[Forte, Watt, 2013]** Smaller uncertainty than envelope: 4.8% vs 3.4% for gg>H
- Meta-PDFs: fit with input functional form the CT, MSTW and NNPDF shapes and combine in a unique consistent set **[Gao, Nadolsky, 2014]**
- Crucial to decide optimal value of α_S and its uncertainty in combination

Phenomenology

How to combine sets?



- Is it possible to reduce the size of a PDF set of Monte Carlo replicas with no loss of information?



- Avoid bias in the extrapolation region
- Preserve physical requirements: positivity of xsec, sum rules and PDF correlations
- Complex procedure: work in progress by S. Carrazza and J.I. Latorre

Conclusions and Outlook

- NNPDF23QED and NNPDF23QED_LO for MC widely used
- The NNPDF3.0 release is a major upgrade
  - Totally rewritten code NNPDF++
  - Improved methodology and closure test validation
  - Proven independence of basis
  - More accurate theory settings: jets, EW corrections
  - Many more LHC data included, significant impact
  - Improved positivity (SUSY observables and large x gluons and quarks)
  - NNPDF30 is available at LO, NLO, NNLO, for several nf and as
- NNPDF is the only unpolarized and polarized set available in LHAPDF
- What's next?
  - Working on NNPDF30QED and NNPDF30IC with intrinsic charm
  - Fit to fragmentation functions within similar framework soon available!
  - In the near future NNPDF30 including N3LO approximation and resummations
    based on Ball, Bonvini, Forte, Marzani, Ridolfi et al, NP B874 (2013)

BACKUP

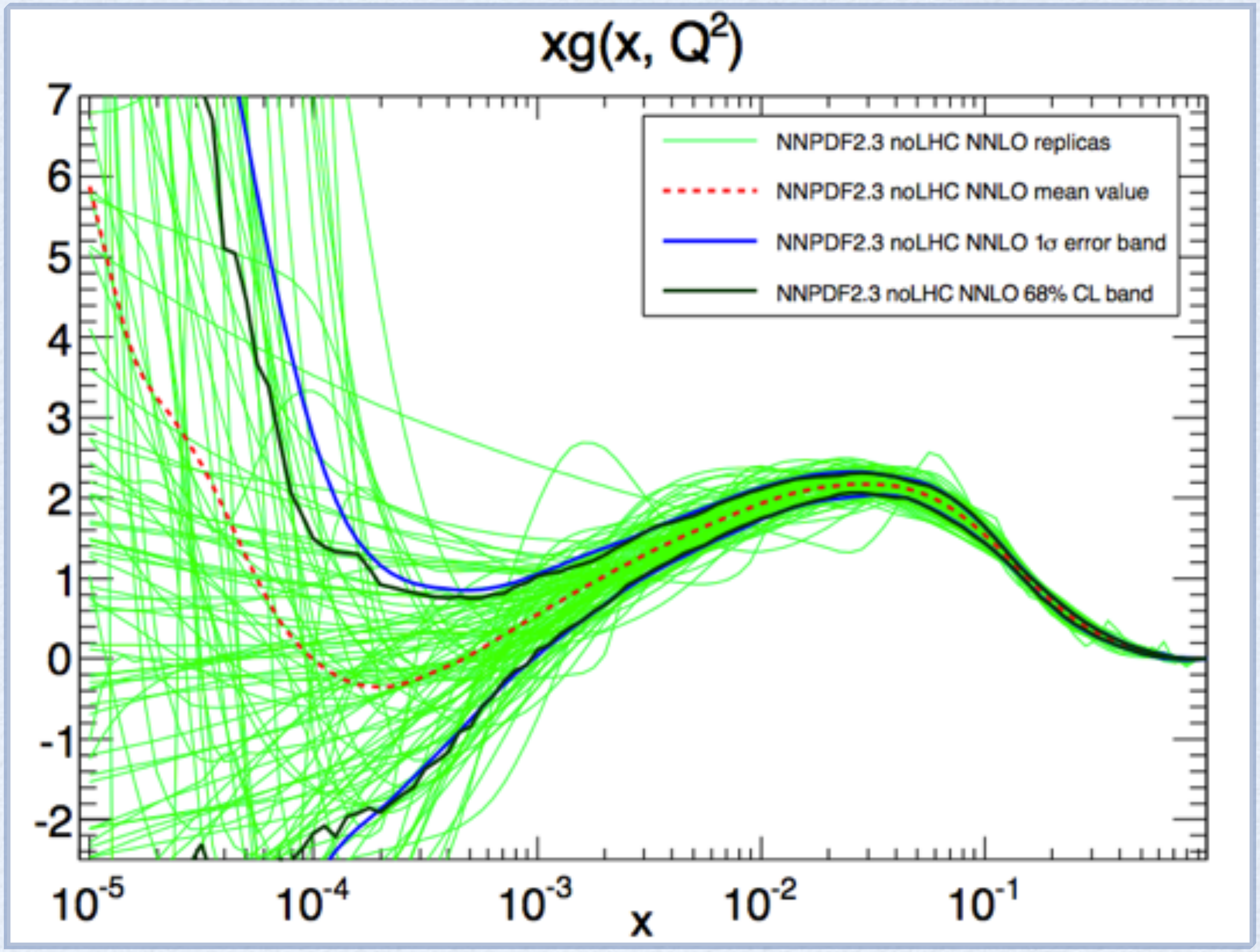
Introduction

The NNPDF approach

- Monte Carlo by importance sampling
- Neural Networks as interpolants
- Genetic algorithm for neural network training
- Cross-validation to stop of the minimization

$$\langle \mathcal{O} \rangle = \int \mathcal{O}[f]\,\mathcal{P}(f)\,Df$$

$$\frac{1}{N}\sum_{k=1}^{N} \mathcal{O}[f^{(k)}]$$



The NNPDF3.0 set

Improved methodology: Weight Penalty

- NNPDF optimal fitting has been determined so far by using CROSS-VALIDATION: data randomly divided in two sets: training (fitted) and validation (non-fitted).

- Alternatively one can introduce a penalty factor in the measure of goodness, designed to discriminate against functions that vary too fast [**Graczyk, Plonski, Sulej JHEP1009 (2010) 053**]

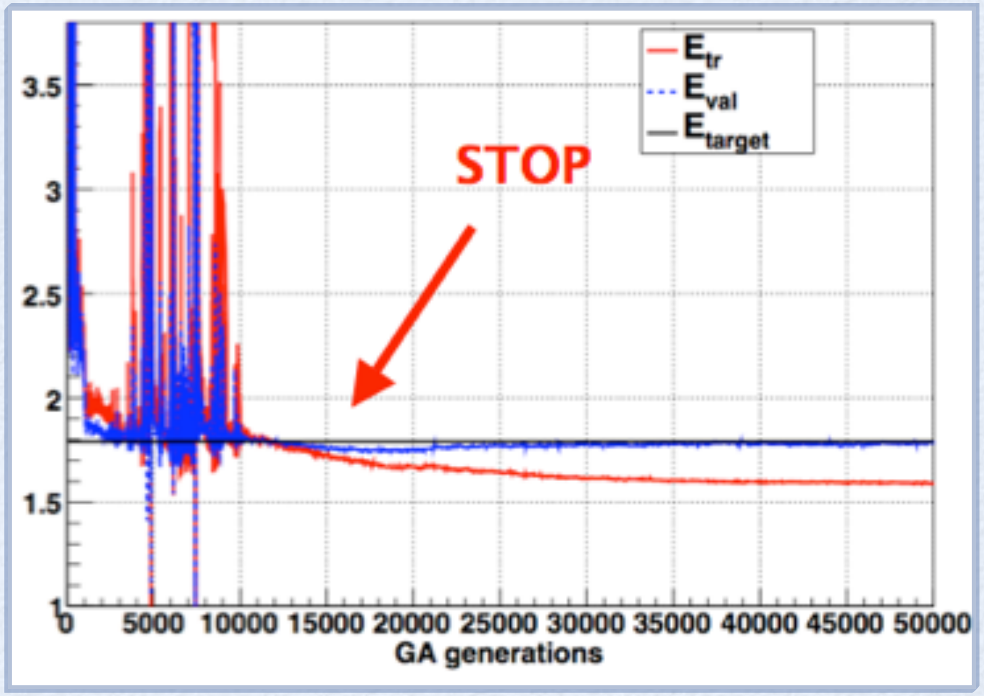
$$E[d,f] = \frac{1}{2}\chi^2[d,f] + \alpha\Delta[f]$$

constant determined by the expected complexity of each NN based on previous fits

penalty function related to the complexity of each NN

$$\Delta[f] = \sum_i w_i^2$$

$$\alpha_i = \left[\frac{\langle\Delta_i\rangle}{N_w}\right]^{-1}$$



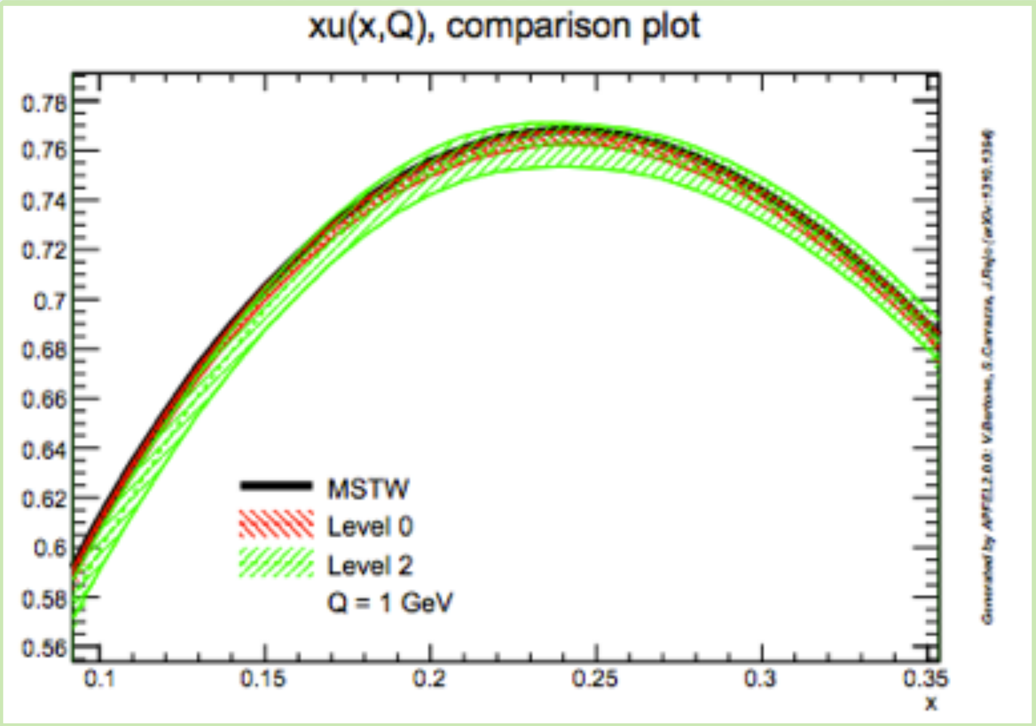
- Iterate till convergence
- Convergence is reached when network fit the data but are not too complex

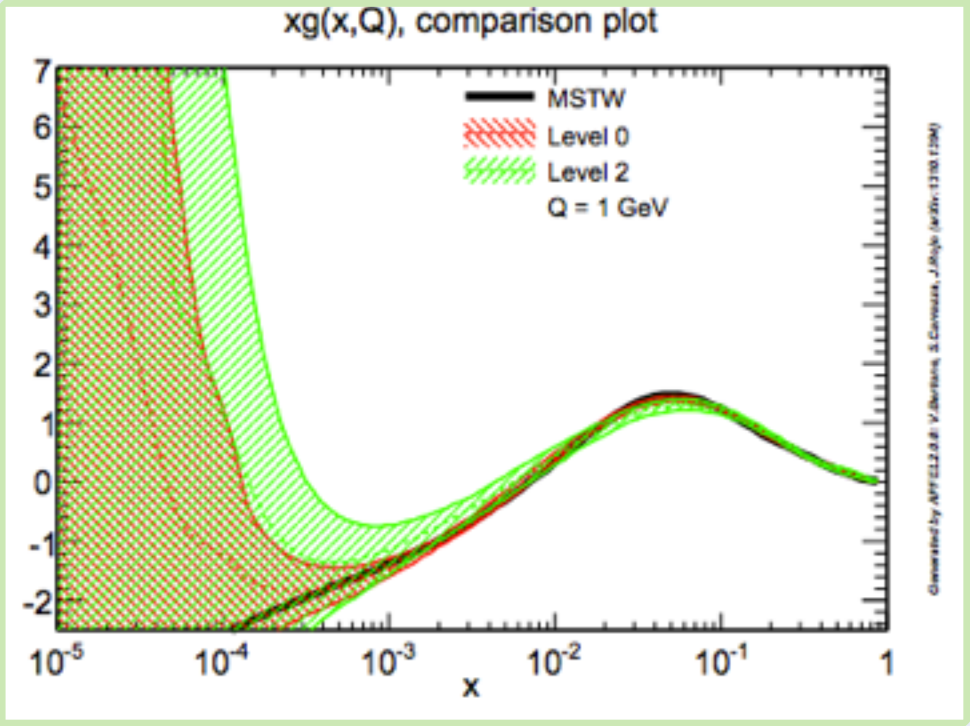
The NNPDF3.0 set

Improved methodology: closure test

LEVEL 2: each datapoint is obtained as a random fluctuation with given covariance matrix about the "truth". Generate pseudo-data replicas of these "data", then fit PDF replicas to pseudo-data replicas. Fit, must find $\chi^2 = 1$, (predictions-theory) compatible with 0 and within 1σ of MSTW "true" PDFs

Perform Fixed-Length fit to 100% data





✔ Truth is within 1σ error band!

The NNPDF3.0 set

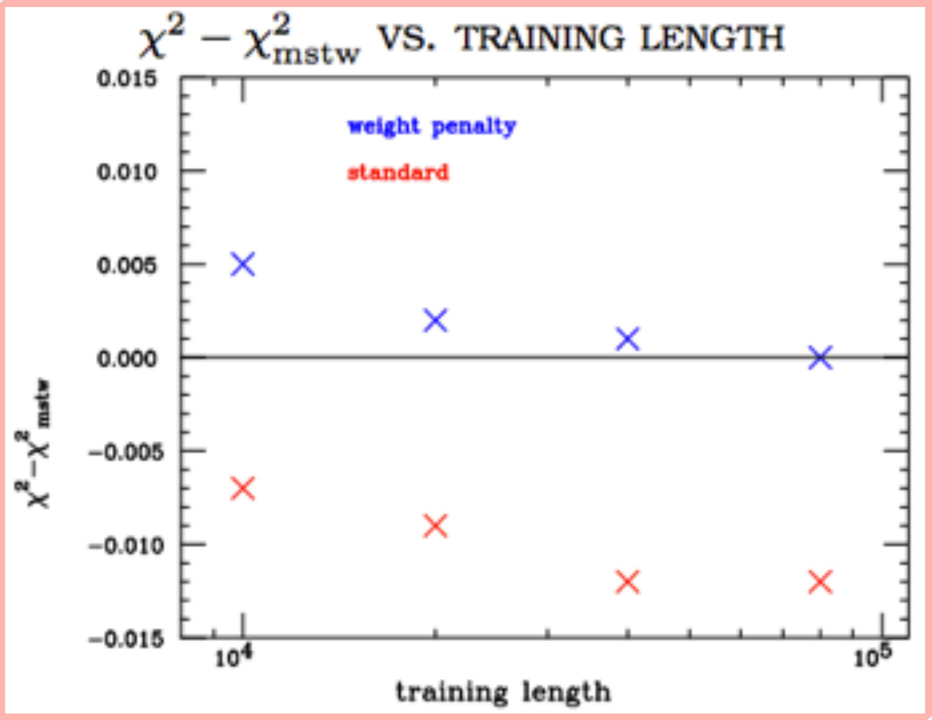
Improved methodology: closure test

LEVEL 2: each datapoint is obtained as a random fluctuation with given covariance matrix about the "truth". Generate pseudo-data replicas of these "data", then fit PDF replicas to pseudo-data replicas. Fit, must find $\chi^2 = 1$, (predictions-theory) compatible with 0 and within 1σ of MSTW "true" PDFs

- At 10K iterations

$$\chi^2 = 0.96,\ \langle E \rangle = 2.0\ (\text{NOTE}\ \chi^2_{mstw} = 0.96)$$

✔



- Chi2 within 0.1% accuracy!
- Same at 20K, 30K and 40K iterations.
- Non WP show signs of micro-overlearning around 10K iterations of GA
- WP does not overlearn up to 80K iterations
- However micro-overlearning is much smaller than statistical fluctuations

$$\Delta\chi^2 \ll \sigma_{\chi^2}$$

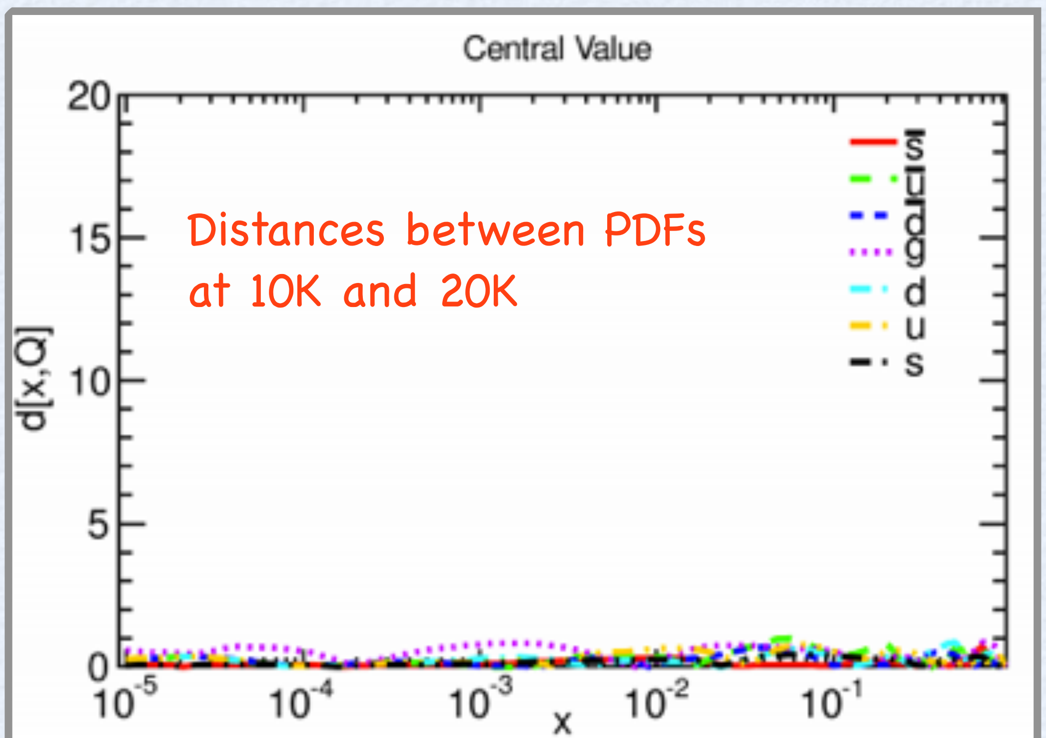
The NNPDF3.0 set

Improved methodology: closure test

LEVEL 2: each datapoint is obtained as a random fluctuation with given covariance matrix about the "truth". Generate pseudo-data replicas of these "data", then fit PDF replicas to pseudo-data replicas. Fit, must find $\chi^2 = 1$, (predictions-theory) compatible with 0 and within 1σ of MSTW "true" PDFs

- At 10K iterations

$$\chi^2 = 0.96,\ \langle E \rangle = 2.0\ (\text{NOTE } \chi^2_{mstw} = 0.96)$$



- Chi2 within 0.1% accuracy!
- Same at 20K, 30K and 40K iterations.
- Non WP show signs of micro-overlearning around 10K iterations of GA
- WP does not overlearn up to 80K iterations
- However micro-overlearning is much smaller than statistical fluctuations

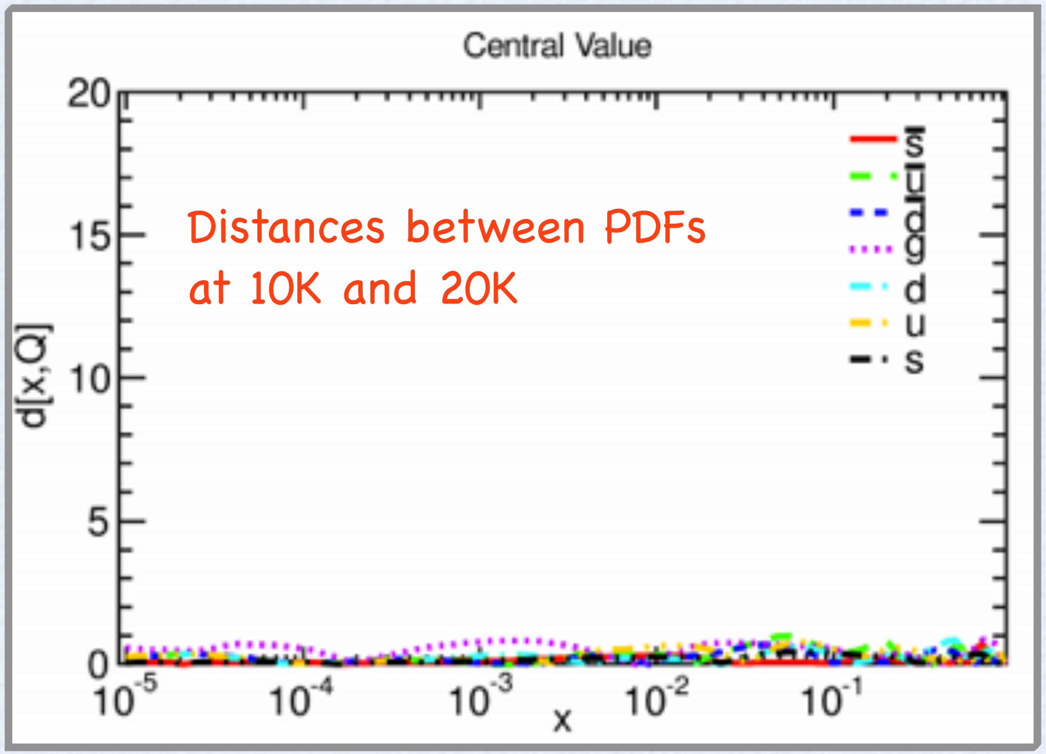
$$\Delta\chi^2 \ll \sigma_{\chi^2}$$

The NNPDF3.0 set

Improved methodology: closure test

LEVEL 0: each datapoint equal to the MSTW true value and uncertainties assumed equal to experimental ones. Fit: must find $\chi^2 = 0$ ✔

LEVEL 2: each datapoint is obtained as a random fluctuation with given covariance matrix about the "truth". Generate pseudo-data replicas of these "data", then fit PDF replicas to pseudo-data replicas. Fit, must find $\chi^2 = 1$, (predictions-theory) compatible with 0 and within 1σ of MSTW "true" PDFs ✔



Preliminary conclusions

- Fixed-Length fit fully adequate
- No overlearning in global fit due to large number of data
- Over-learning observed in fits to reduced datasets
- Effect of Weigh-Penalty moderate

NNPDF3.0 basis

$$
\begin{aligned}
\Sigma(x,Q_0^2) &= \left(u+\bar{u}+d+\bar{d}+s+\bar{s}\right)(x,Q_0^2)\\
T_3(x,Q_0^2) &= \left(u+\bar{u}-d-\bar{d}\right)(x,Q_0^2)\\
T_8(x,Q_0^2) &= \left(u+\bar{u}+d+\bar{d}-2s-2\bar{s}\right)(x,Q_0^2)\\
V(x,Q_0^2) &= \left(u-\bar{u}+d-\bar{d}+s-\bar{s}\right)(x,Q_0^2)\\
V_3(x,Q_0^2) &= \left(u-\bar{u}-d+\bar{d}\right)(x,Q_0^2)\\
V_8(x,Q_0^2) &= \left(u-\bar{u}+d-\bar{d}-2s+2\bar{s}\right)(x,Q_0^2),
\end{aligned}
$$

NNPDF2.3 basis

$$
\begin{aligned}
\Sigma(x,Q_0^2) &= \left(u+\bar{u}+d+\bar{d}+s+\bar{s}\right)(x,Q_0^2)\\
T_3(x,Q_0^2) &= \left(u+\bar{u}-d-\bar{d}\right)(x,Q_0^2)\\
V(x,Q_0^2) &= \left(u-\bar{u}+d-\bar{d}+s-\bar{s}\right)(x,Q_0^2)\\
\Delta_S(x,Q_0^2) &= \left(\bar{d}-\bar{u}\right)(x,Q_0^2)\\
s^+(x,Q_0^2) &= \left(s+\bar{s}\right)(x,Q_0^2)\\
s^-(x,Q_0^2) &= \left(s-\bar{s}\right)(x,Q_0^2),
\end{aligned}
$$

In NNPDF3.0 a **new input PDF parametrization basis** is used, directly related to the eigenvectors of DGLAP evolution

- **Checked robustness of the results** comparing fits in the ``NNPDF2.3" and ``NNPDF3.0" basis
- Verified that shapes of poorly known PDFs (like dbar-ubar or strange asymmetry) are **genuine results of the fit** and not artificial byproducts of the choice of basis