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

3rd June 2015

University of Birmingham

Motivation

PDFs: why bother?

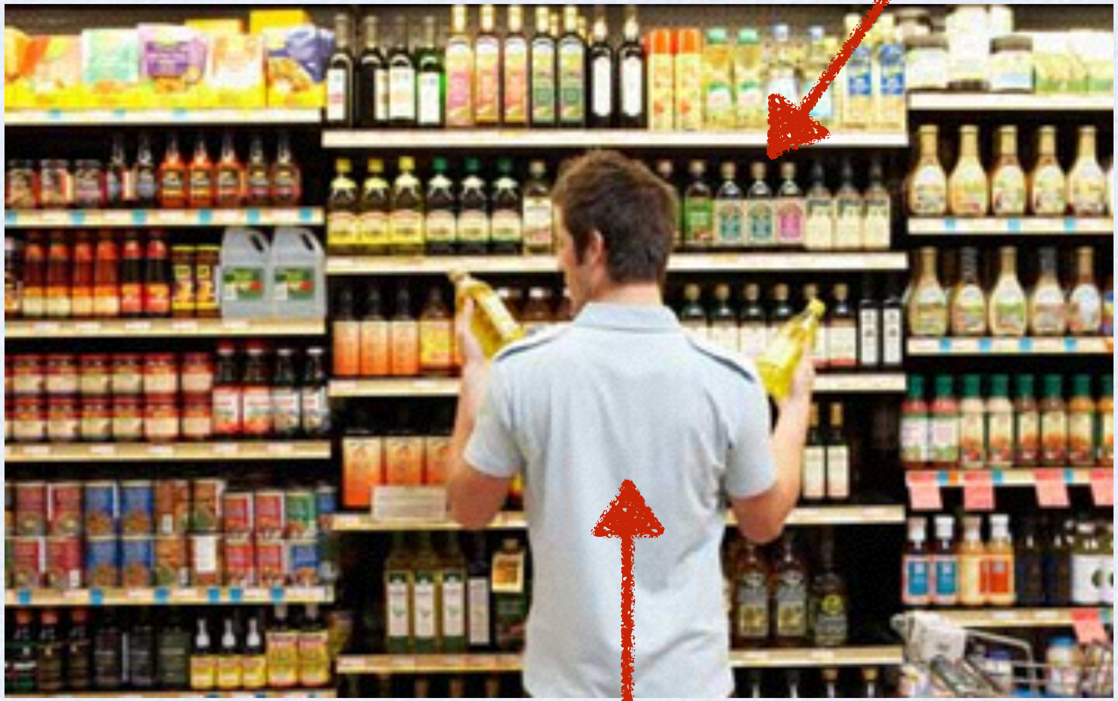
✦ A reliable understanding of PDF uncertainties plays a crucial role at hadron colliders

✦ Can we trust PDF uncertainties?

✦ How do we interpret the differences between PDF sets?

✦ Shall we just pick a set out of the PDFs "supermarket" shelf or take the envelope of ALL predictions?

LHAPDF

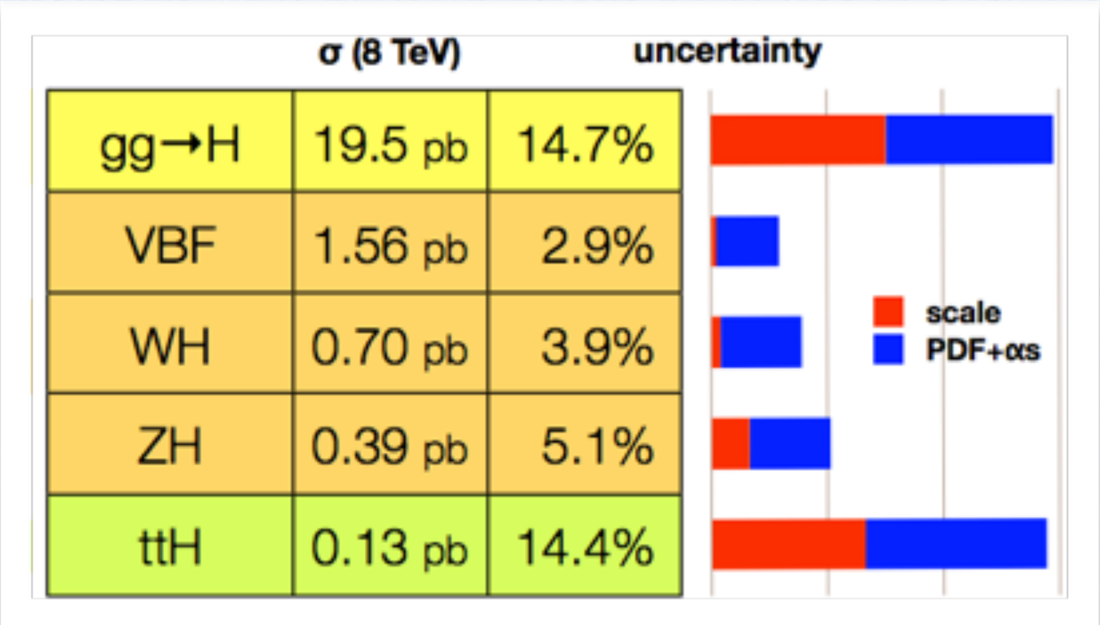


<physicist>

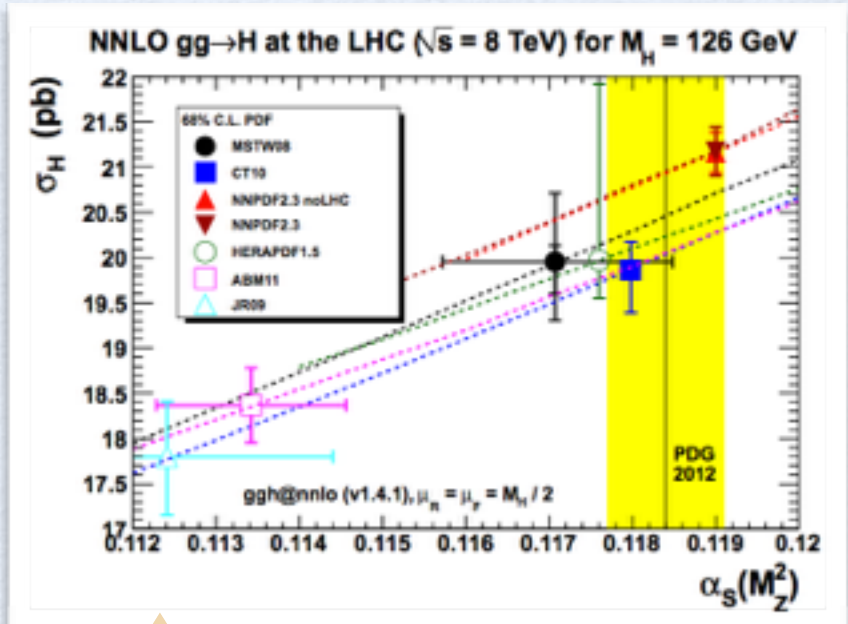
✦ PDF4LHC: huge effort in understanding differences & improving theoretical and statistical treatments in PDF analyses

Motivation

PDFs and LHC interplay



J. Campbell, ICHEP 2012



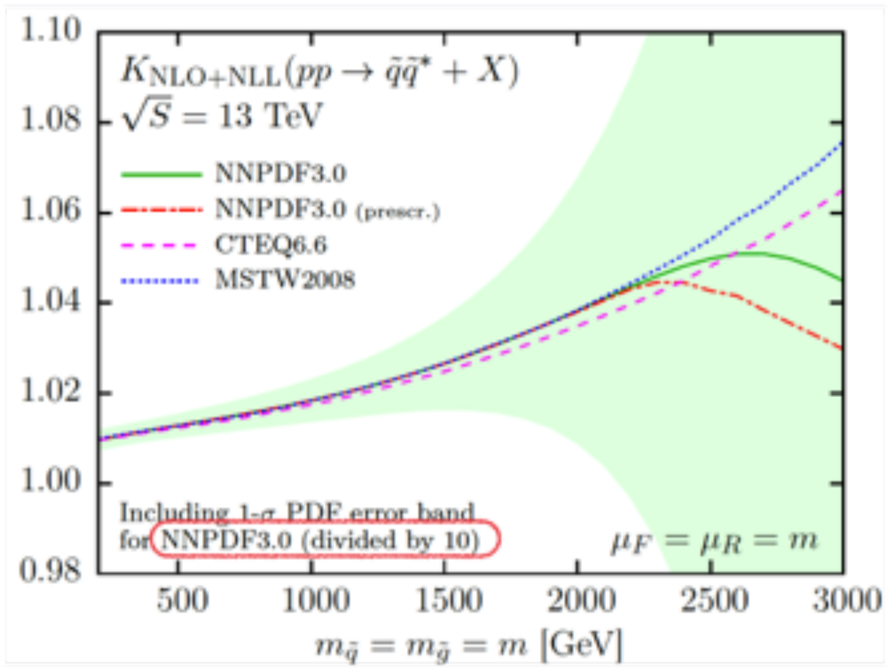
G. Watt, 2012

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

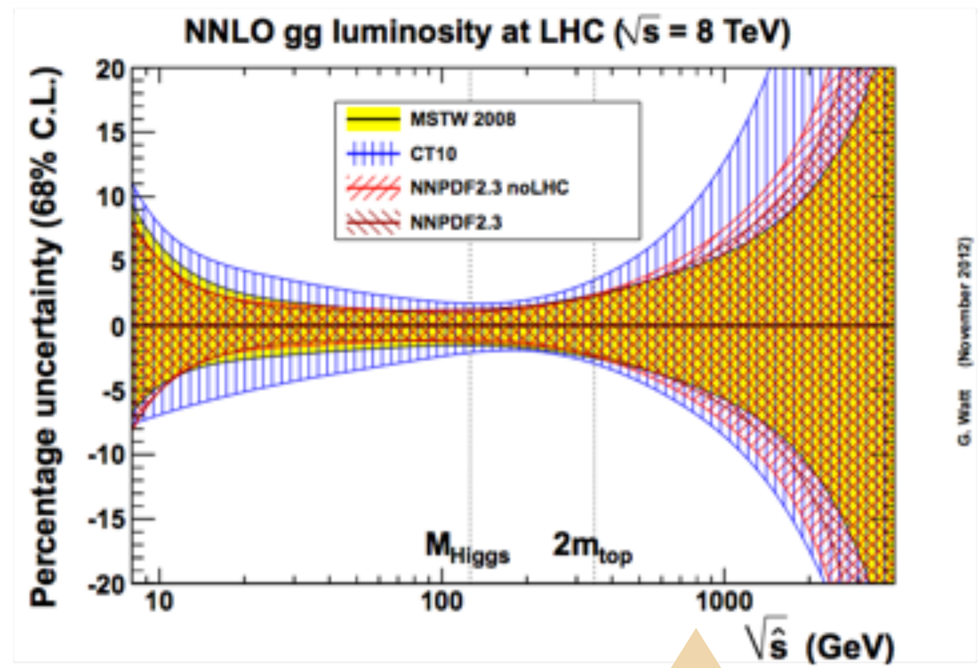
- ☑ PDF uncertainty of each PDF set
- ☑ Value of $\alpha_S(M_Z)$
- ☑ Combination of different PDF sets

Motivation

PDFs and LHC interplay



Courtesy of A. Kulesza

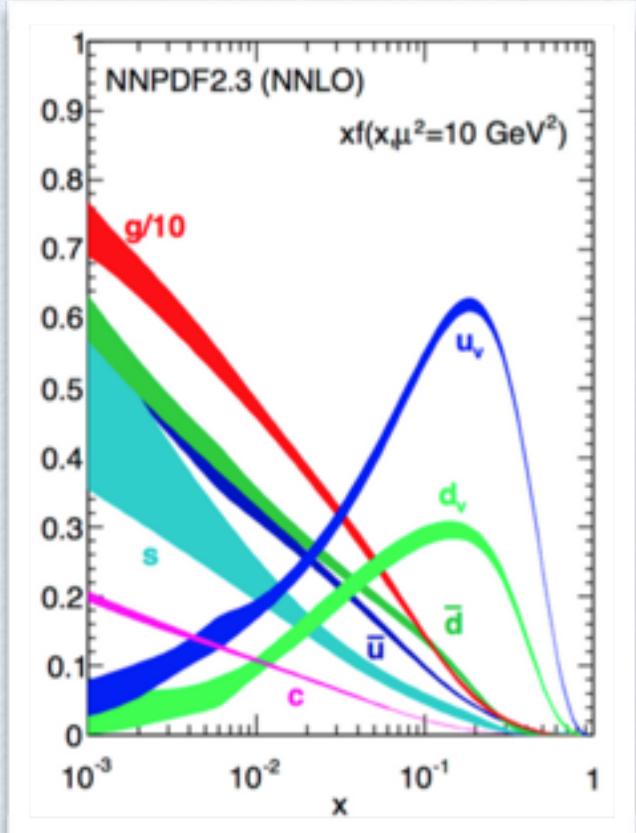


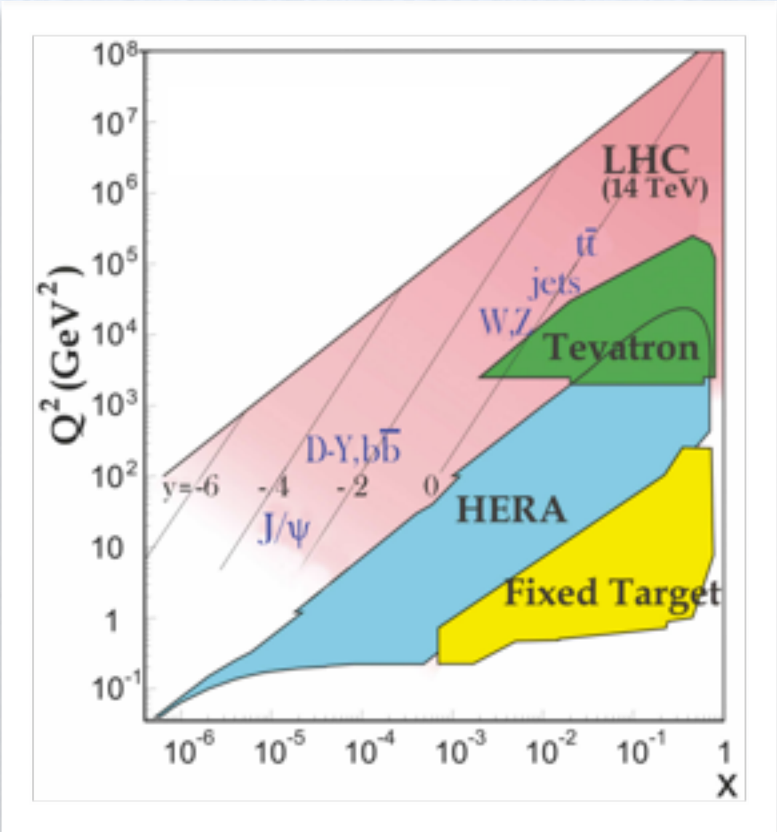
PDF uncertainties are a crucial input at the LHC, often being the limiting factor in the accuracy of theoretical predictions, both SM and BSM

- ☑ PDF uncertainty
- ☑ Lack of data
- ☑ Is theory precise enough?

Motivation

PDFs and LHC interplay





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

- PDFs for LHC run II
  - Collinear factorisation
  - Key ingredients in PDF determination
  - State of the art
- NNPDFs for LHC run II
  - The NNPDF approach
  - News: methodology, theory and data
  - PDF combinations
- Conclusions

PDFs for LHC run II

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

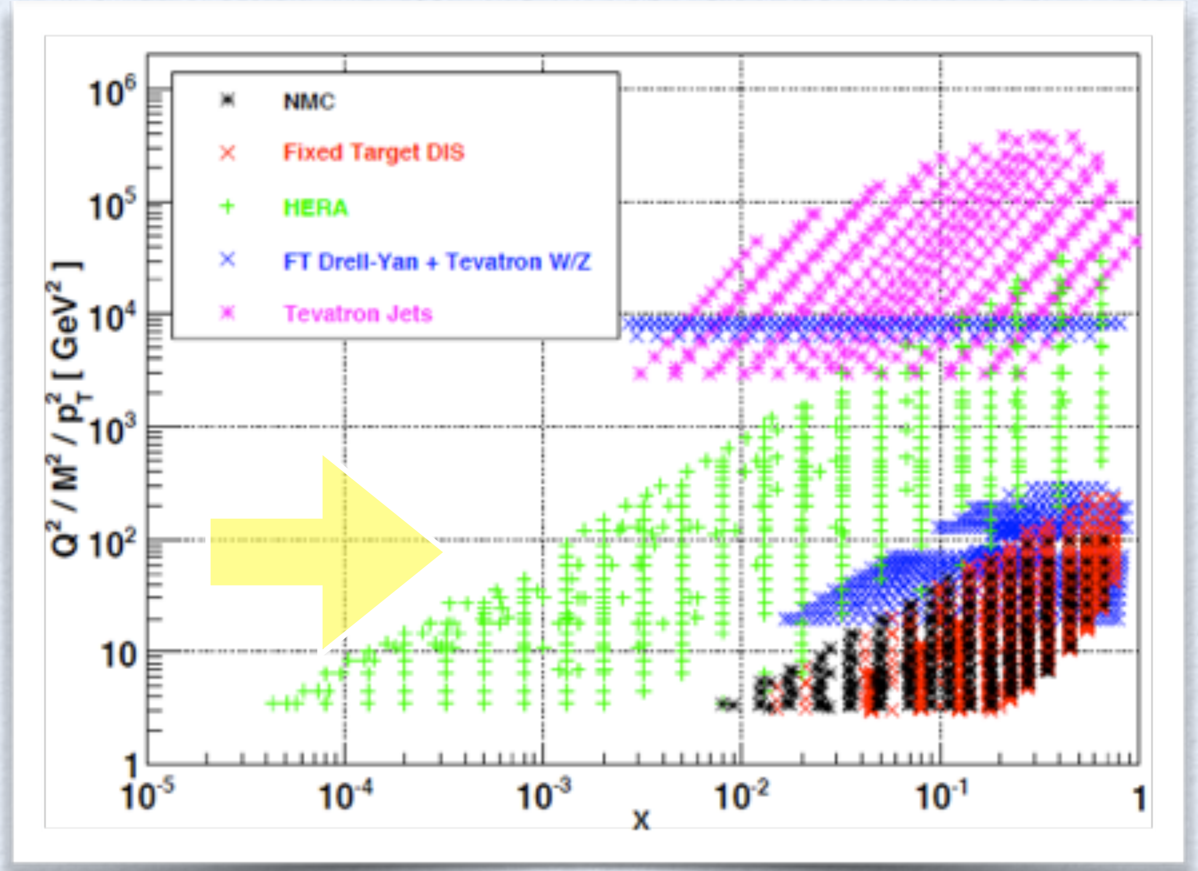
NNLO - Moch, Vermaseren, Vogt, 2004

PDFs

and collinear factorisation

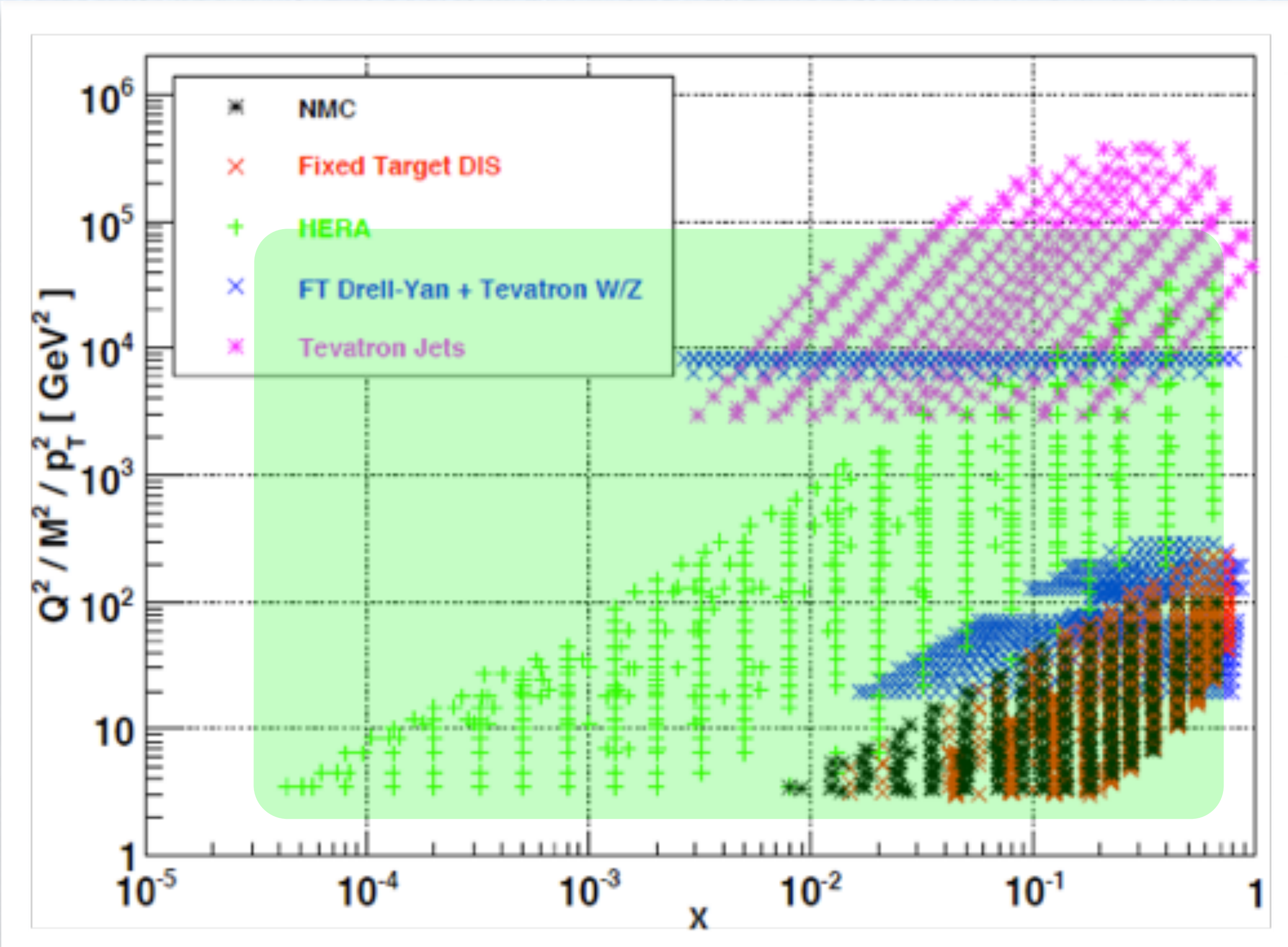
$$\frac{d\sigma_H^{pp\to ab}}{dX} = \sum_{i,j=1}^{N_f} f_i(x_1,\mu_F)\, f_j(x_2,\mu_F)\, \frac{d\sigma_H^{ij\to ab}}{dX}(x_1 x_2 S_{\rm had}, \alpha_s(\mu_R), \mu_F) + \mathcal{O}\left(\frac{\Lambda_{\rm QCD}^{2n}}{S_{\rm had}^n}\right)$$

- PDFs 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



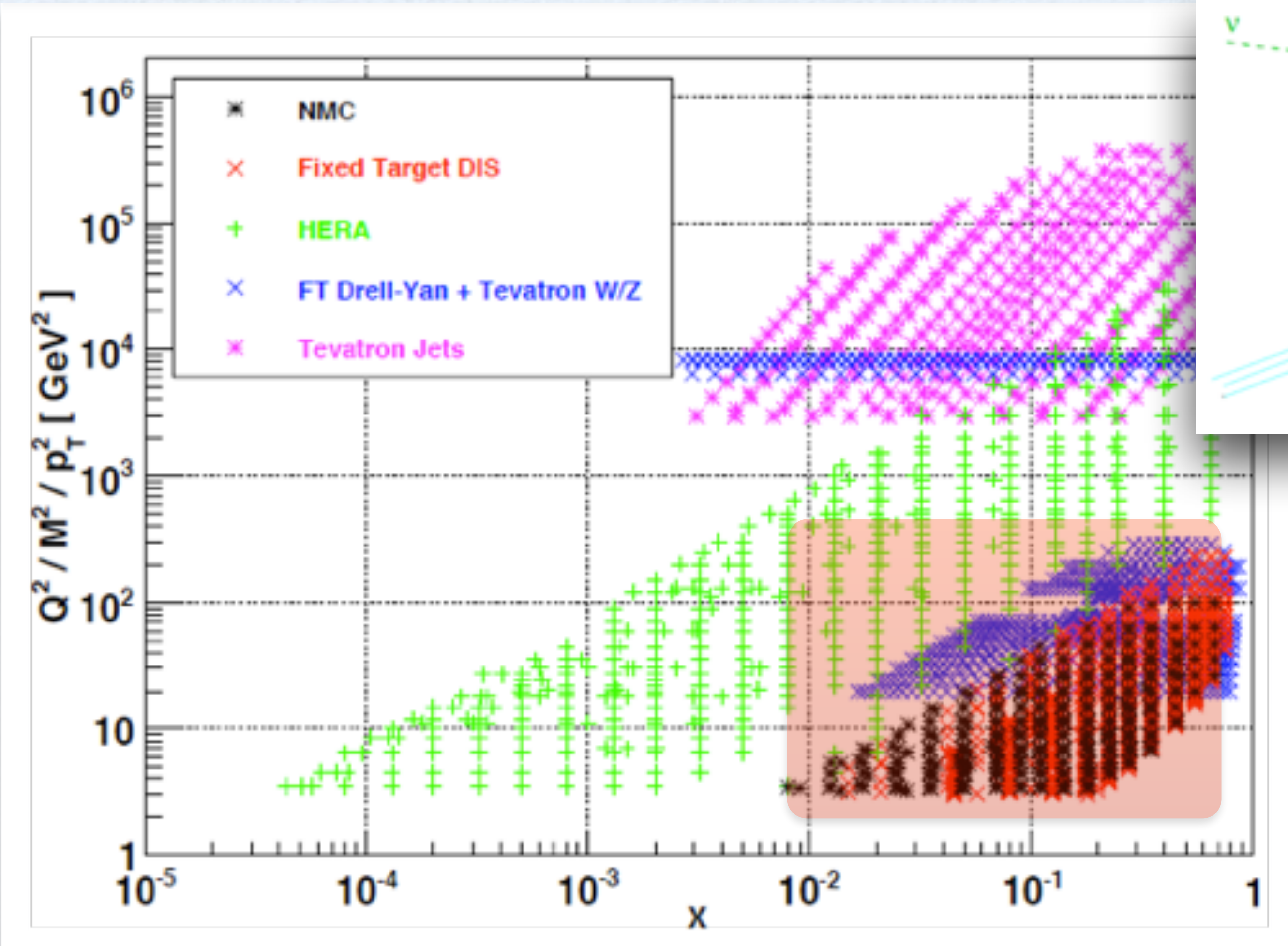
NC	$F_1^{\gamma,Z} = \sum_i e_i^2 (q_i + \bar{q}_i)$
CC	$F_1^{W^+} = \bar{u} + d + s + \bar{c}$
CC	$-F_3^{W^+}/2 = \bar{u} - d - s + \bar{c}$
	$F_2 = 2xF_1$

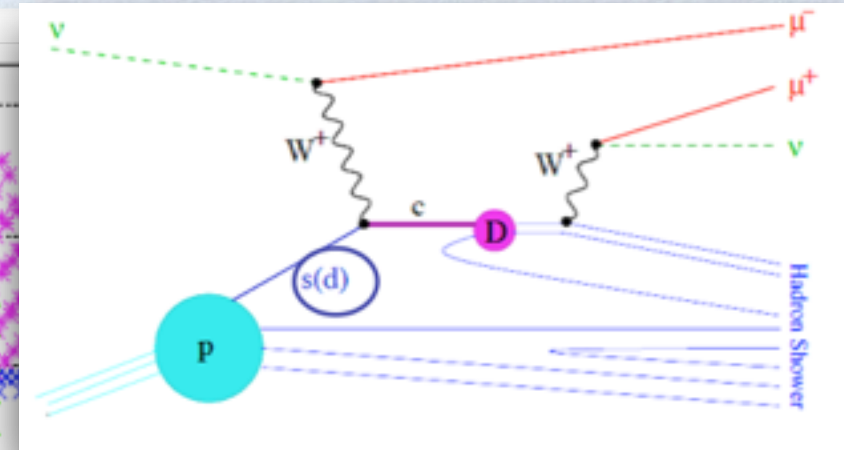
HERA DIS data

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

Constraints from data

before the LHC



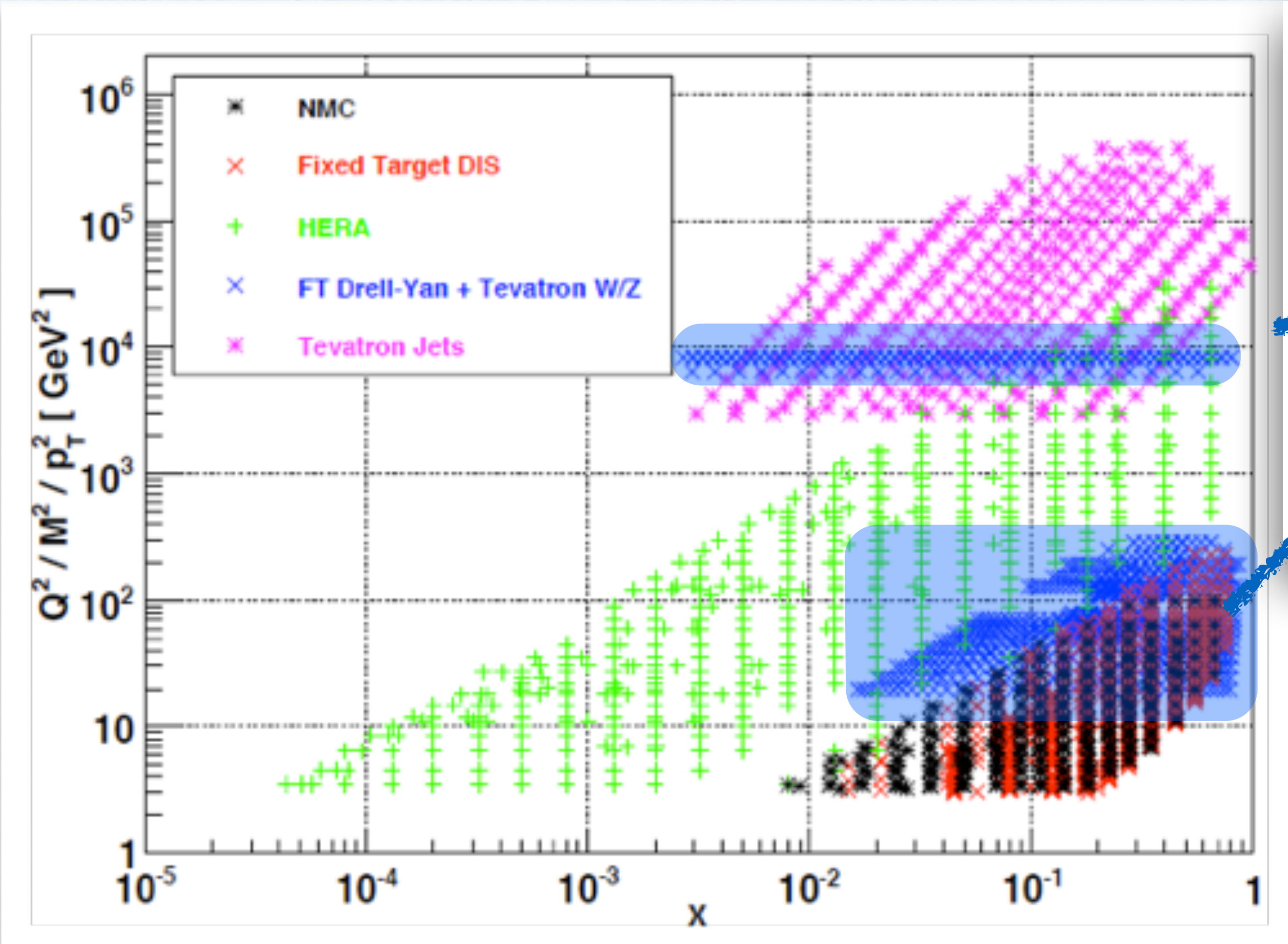


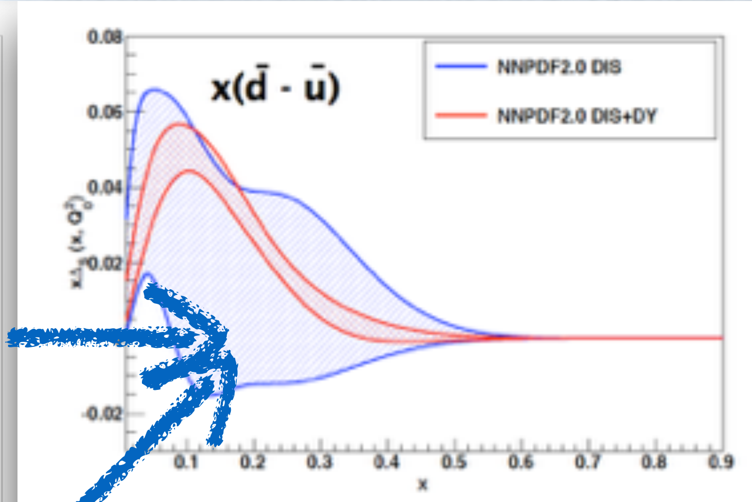
Fixed Target DIS data

- 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

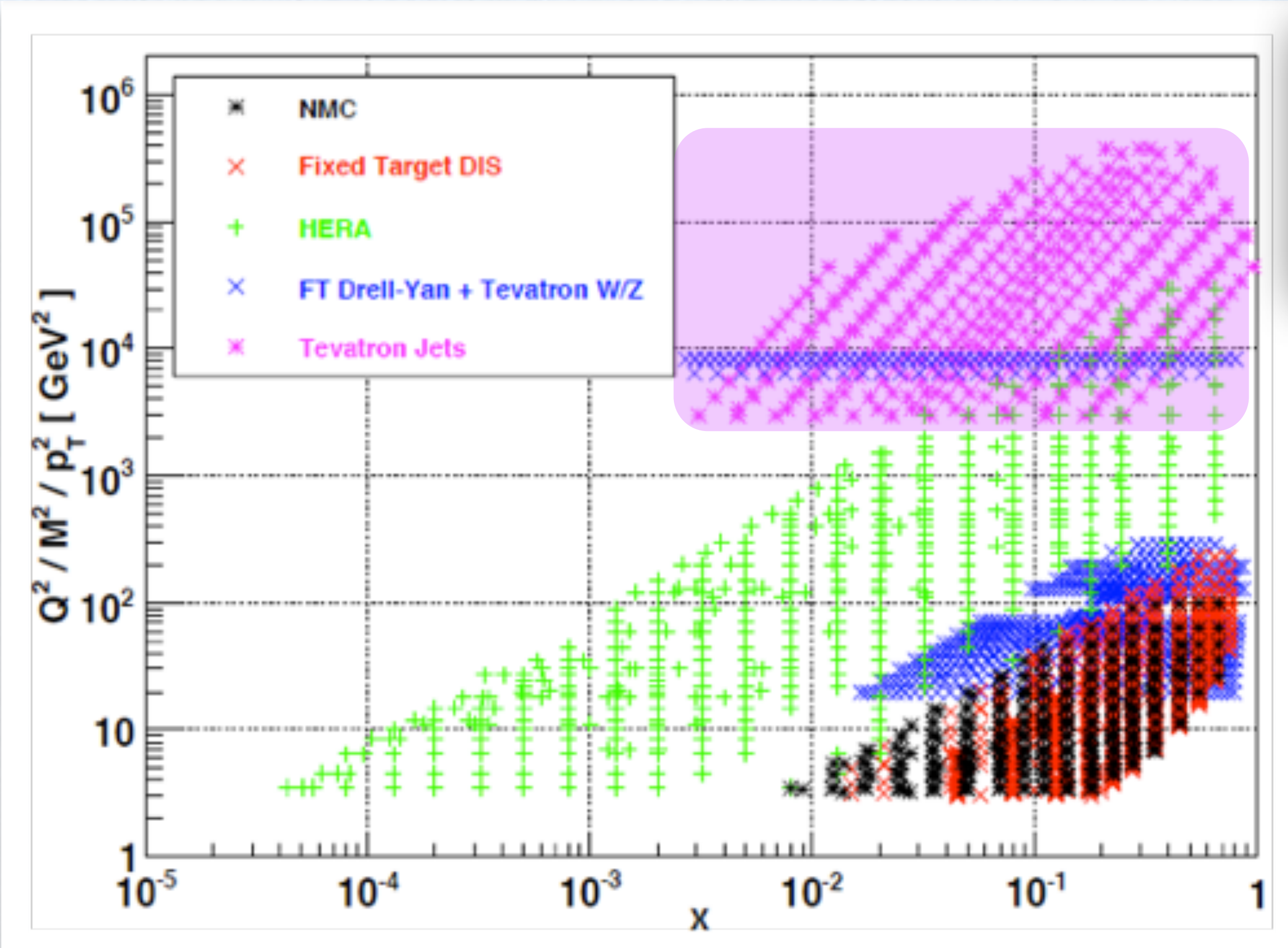
- light quark and antiquark separation

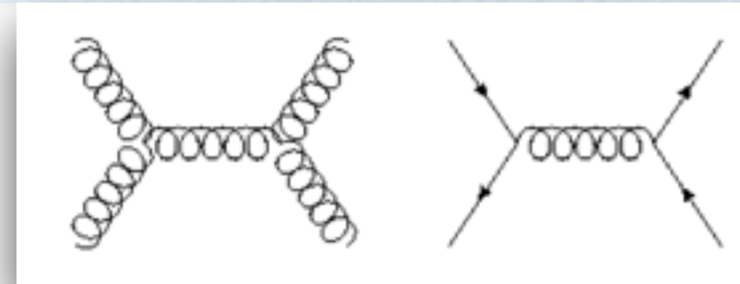
$$\sigma^{\mathrm{DY,p}} \propto u(x_1)\bar{u}(x_2) + d(x_1)\bar{d}(x_2)$$

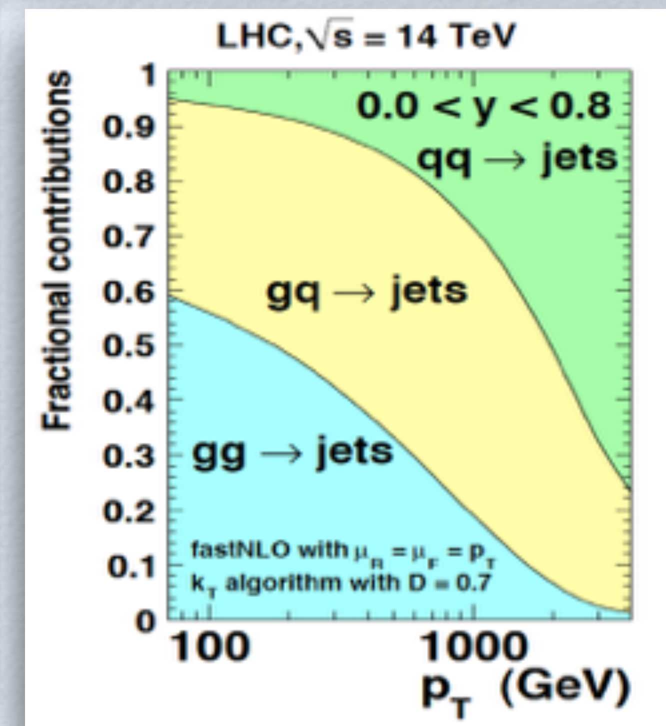
$$\sigma^{\mathrm{DY,d}} \propto u(x_1)(\bar{u} + \bar{d})(x_2) + d(x_1)(\bar{u} + \bar{d})(x_2)$$

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

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

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

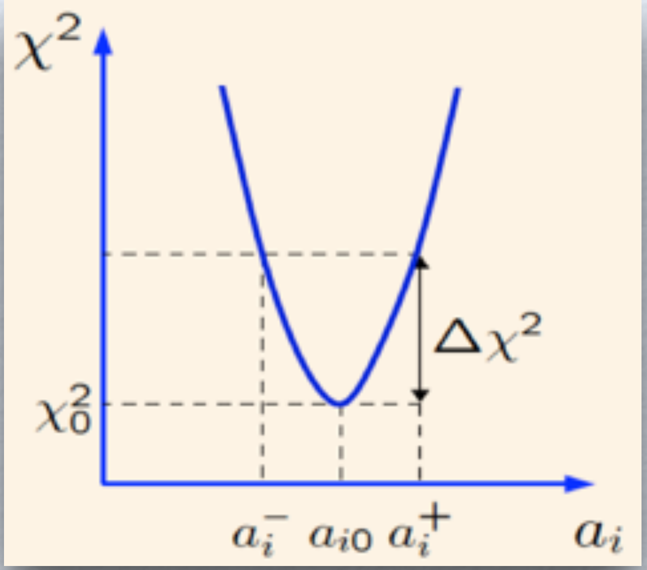
The name of the game

Not as simple as it may look - I: error propagation

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

- 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)

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



- 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

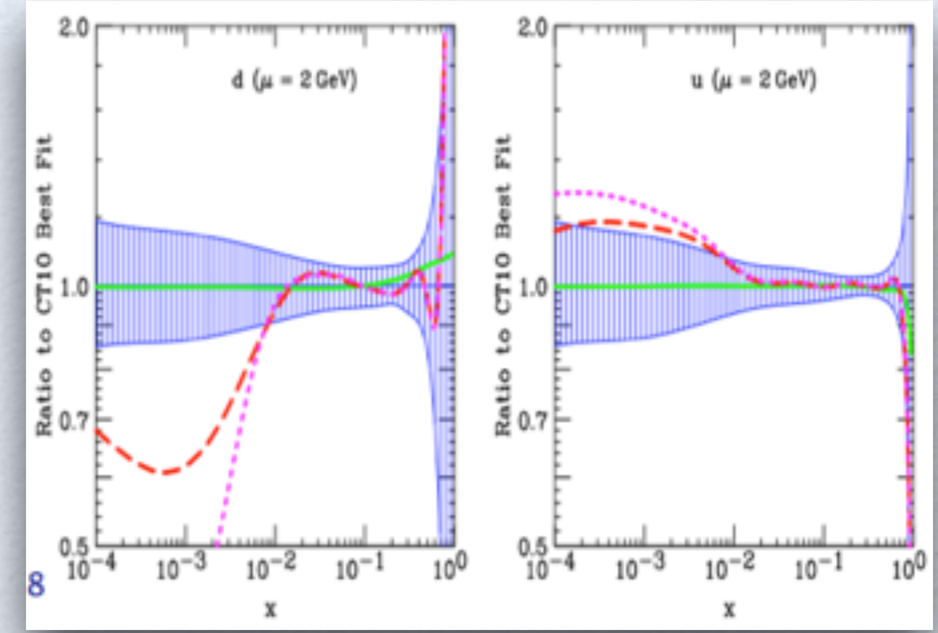
Not as simple as it may look - II: parametrisation bias

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

- 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

Progress in PDF determination

Looking back...

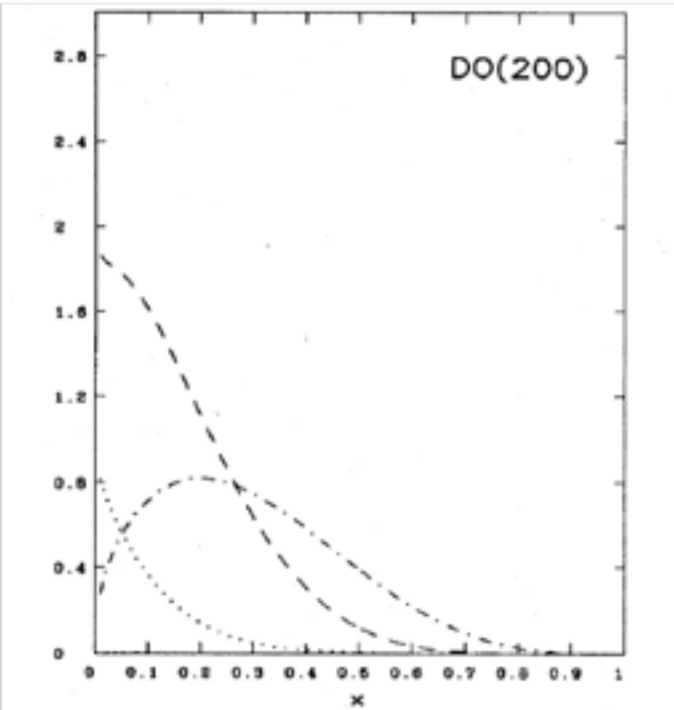
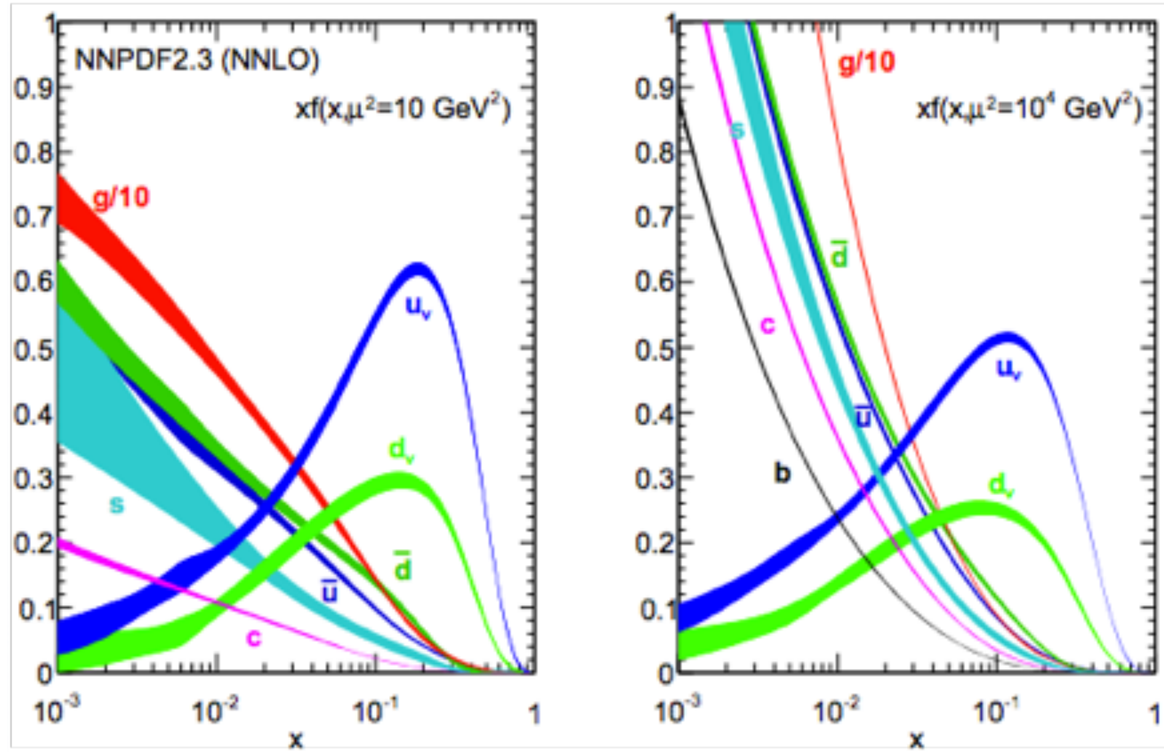


FIG. 27. "Soft-gluon" (Λ=200 MeV) parton distributions of Duke and Owens (1984) at $Q^2=5$ GeV^2: valence quark distribution $x[u_v(x)+d_v(x)]$ (dotted-dashed line), $xG(x)$ (dashed line), and $q_v(x)$ (dotted line).



PDG "Structure Functions" 2013

- **< 2002:** sets without uncertainty
- **2003-2004:** first MRST, CTEQ, Alekhin sets with uncertainties
- **2004-now:** huge progress made in statistical and theoretical understand, new players

Progress in PDF determination

A personal overview: victories of the past

PAST

THEORY

* Heavy quark scheme
* Parameters: α_S & m_Q
* (N)NLO corrections

DATA

* PDF uncertainty
* Treatment of correlated systematics

METH.

* Parametrisation bias
* Treatment of inconsistent data

Progress in PDF determination

A personal overview: challenges ahead

	PAST	PRESENT
THEORY	* Heavy quark scheme * Parameters: α_S & m_Q * (N)NLO corrections (!)	* NNLO corrections * QED/EW corrections * Resummations
DATA	* PDF uncertainty * Treatment of correlated systematics	* LHC data, combinations from HERA, Tevatron legacy, ...
METH.	* Parametrisation bias (!) * Treatment of inconsistent data	* Closure Tests * Combination of different PDF sets

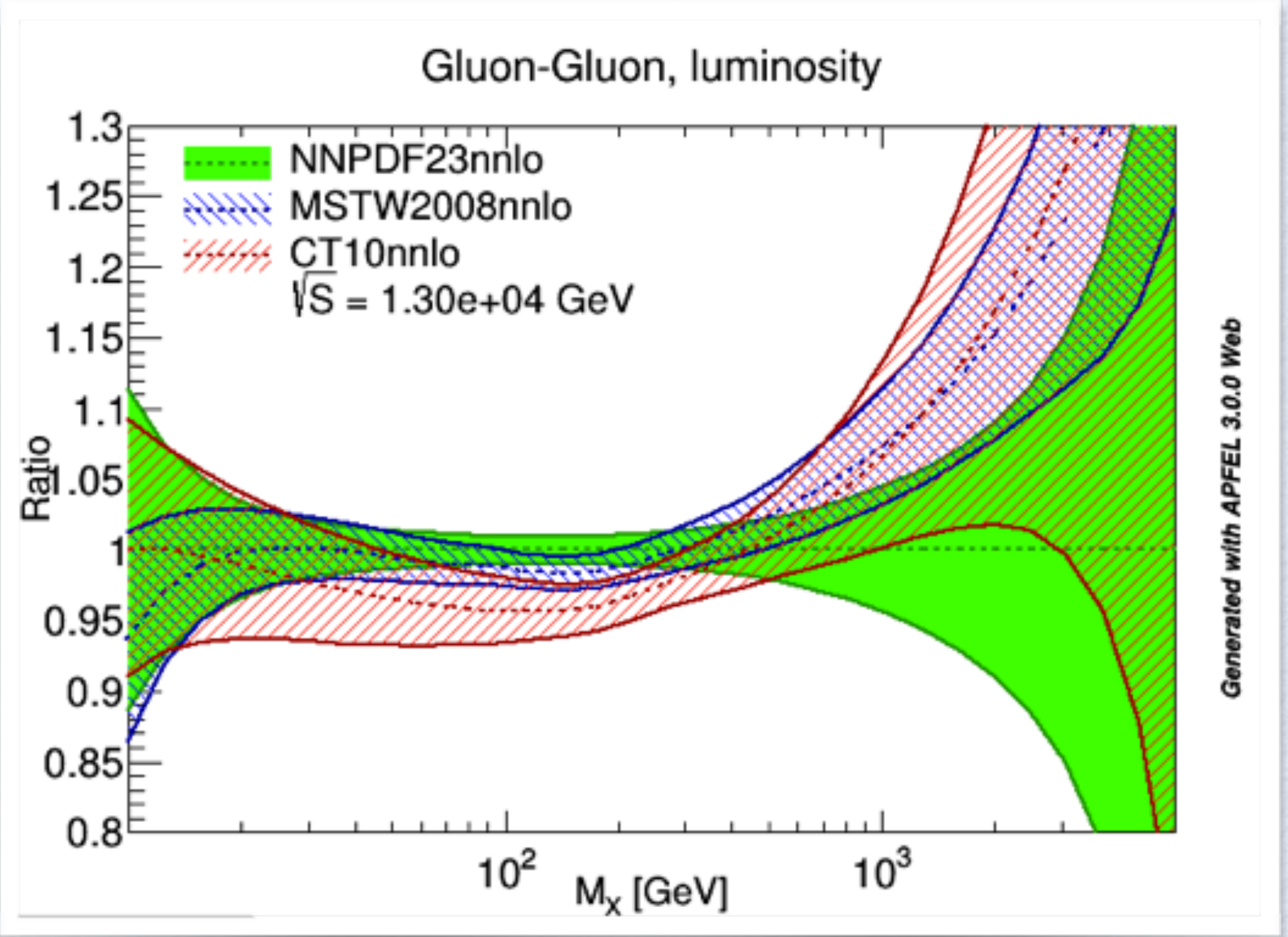
Progress in PDF determination

The state of the art

April 2015	Theory	Data	Methodology
CT14 preliminary	ACOT for HQ No LHC jets at NNLO APPLgrid/FastNLO Scale variation estimate	DIS Fixed Target DY Jets Top Quark LHC DY	Polynomial param (27 par.) Hessian eigenvectors Fixed Tolerance MC and Hessian Reweig.
MMHT arXiv:1410.3989	TR' for HQ No LHC jets at NNLO APPLgrid/FastNLO EW corrections Deuteron corrections	DIS Fixed Target DY Jets Top Quark LHC DY	Chebyshev pol. (37 par.) Hessian eigenvectors Dynamic Tolerance MC and Hessian Reweig.
NNPDF3.0 arXiv:1410.8849	FONLL for HQ NNLO approx for jets APPLgrid/FastNLO EW corrections	DIS Fixed Target DY Jets Top Quark LHC DY	Neural Network param MC replicas Bayesian Reweighting Closure tests
ABM12 arXiv:1310.3059	FFN for DIS VFN for LHC DY Fitted α_s	DIS Fixed Target DY LHC DY	Polynomial param (14 par.) Hessian eigenvectors No Tolerance
HERAPDF2.0 preliminary	TR' for HQ plus other schemes implemented	HERA-I HERA-II	Hessian eigenvectors MC representation Model & param uncertainty

News for LHC run II

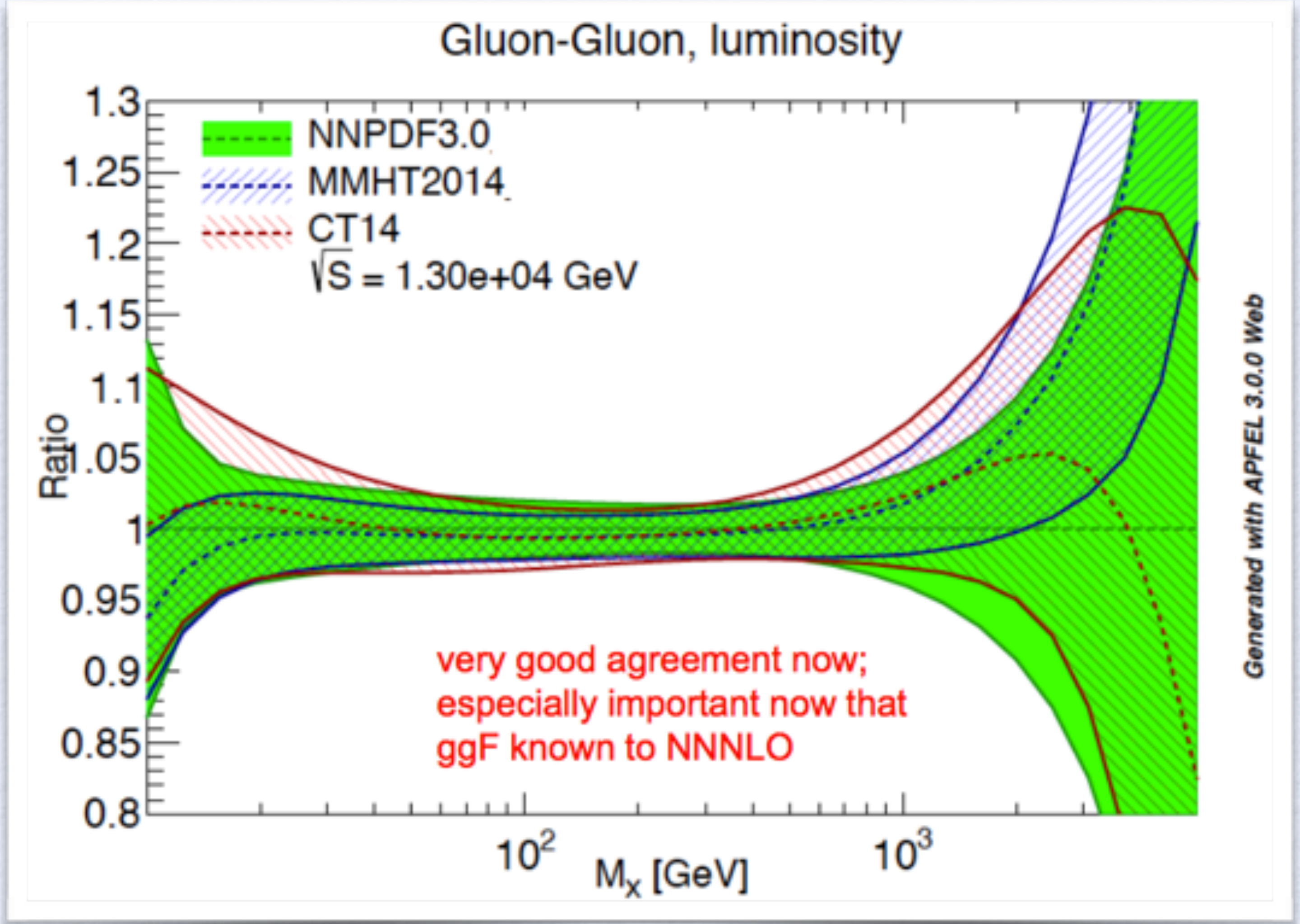
Gluon luminosity



2014

News for LHC run II

Gluon luminosity

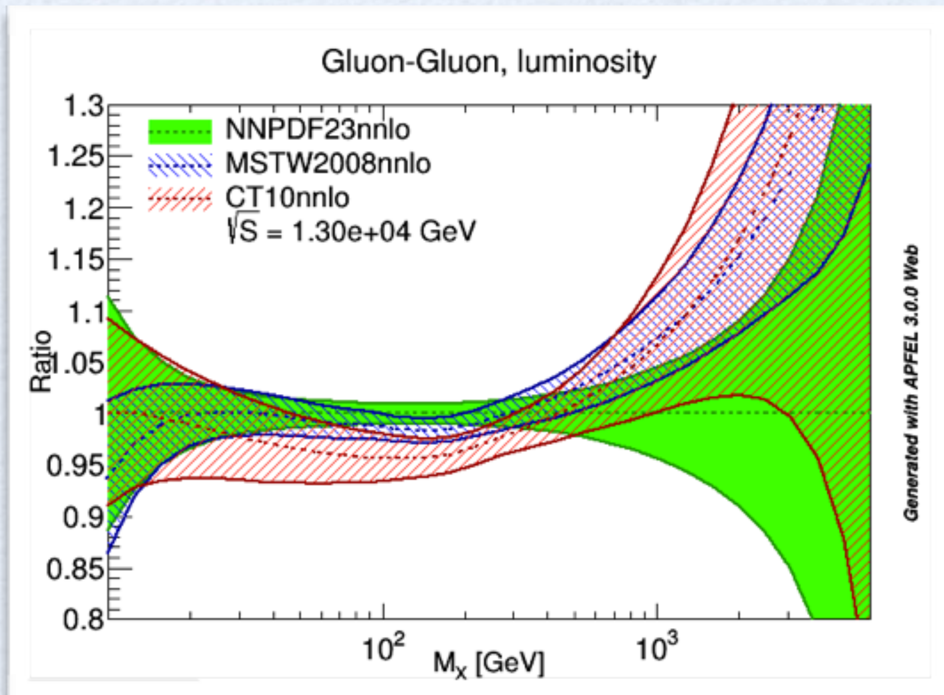


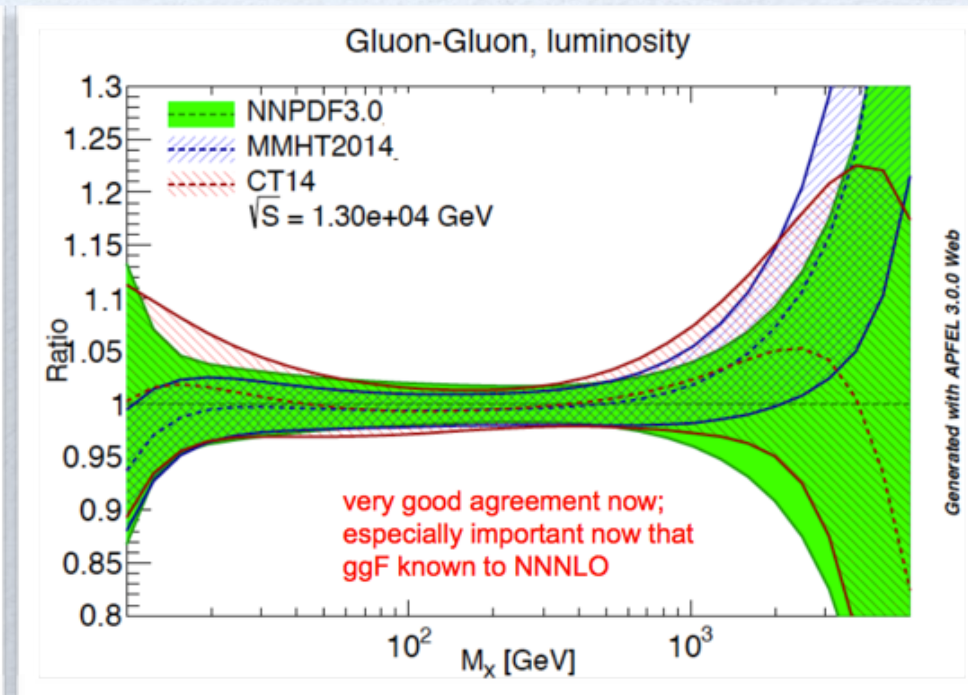
2015

J. Houston, PDF4LHC April 2015

News for LHC run II

Gluon luminosity and implications for Higgs cross section





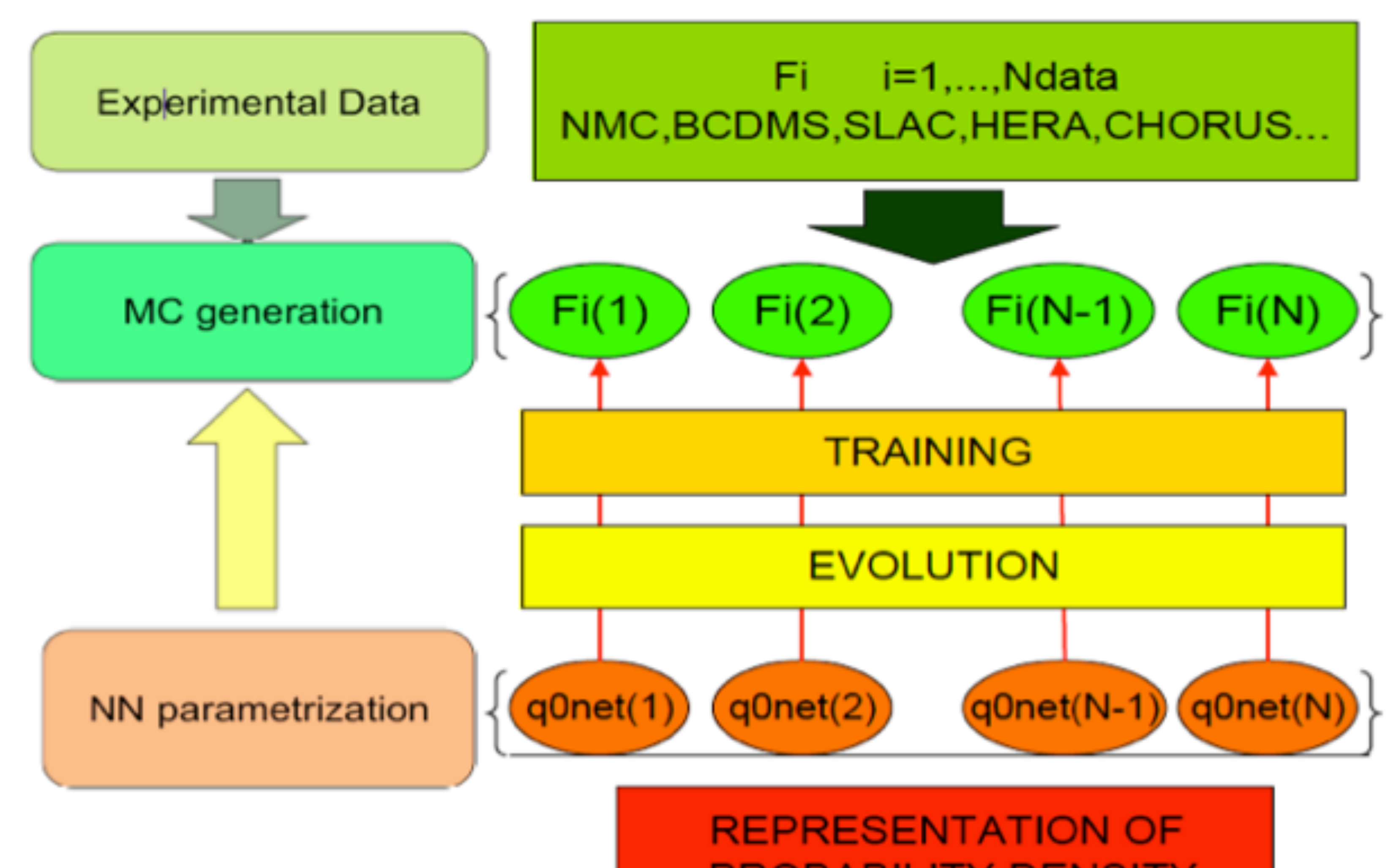
J. Houston, PDF4LHC April 2015

ggF @ NNLO (pb)	CT14	NNPDF3.0	MMHT2014
8 TeV	18.66	18.77	18.65
13 TeV	42.68	42.97	42.70

NNPDFs for run II

The NNPDF approach

Monte Carlo and Neural Network

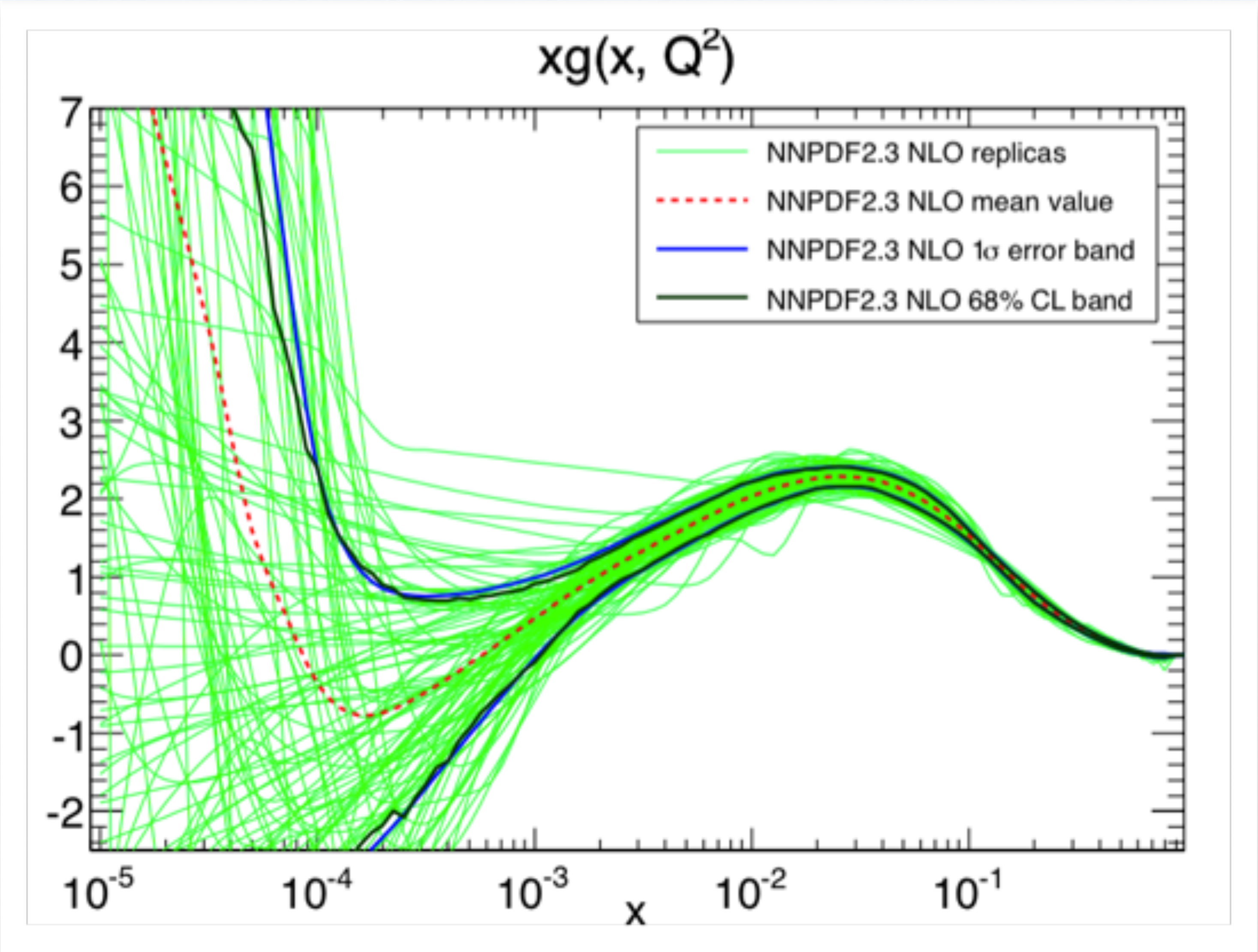


$$\left\langle F_i^{(\mathrm{net})} \right\rangle_{\mathrm{rep}} = \frac{1}{N_{\mathrm{rep}}} \sum_{k=1}^{N_{\mathrm{rep}}} F_i^{(\mathrm{net})(k)}$$

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

The NNPDF approach

Monte Carlo and Neural Network



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

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

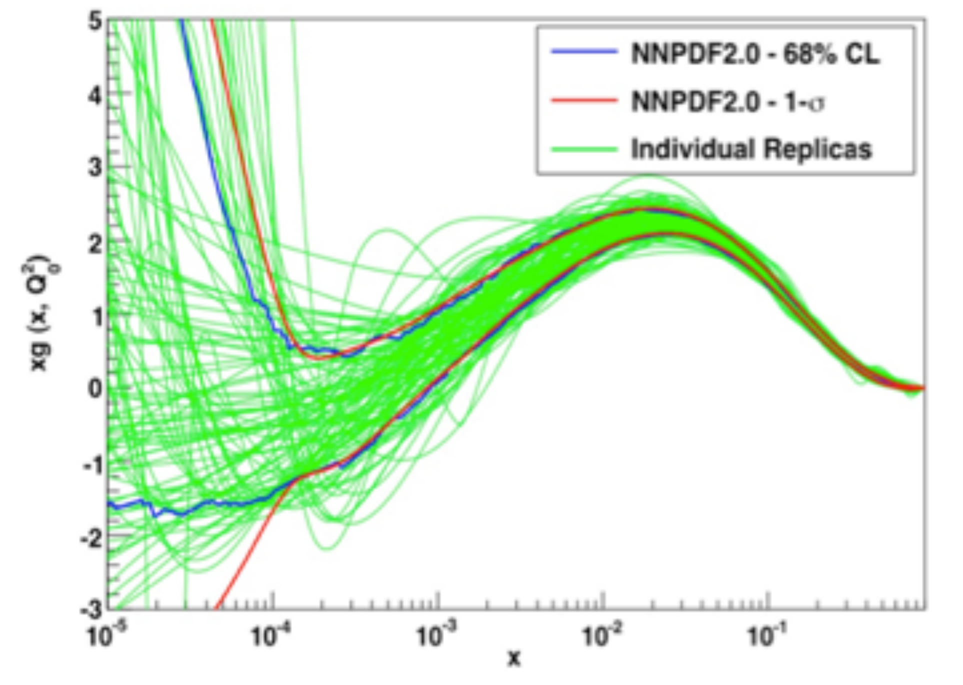
Generate a MC sampling in the parameter space? NOT SO PRACTICAL...

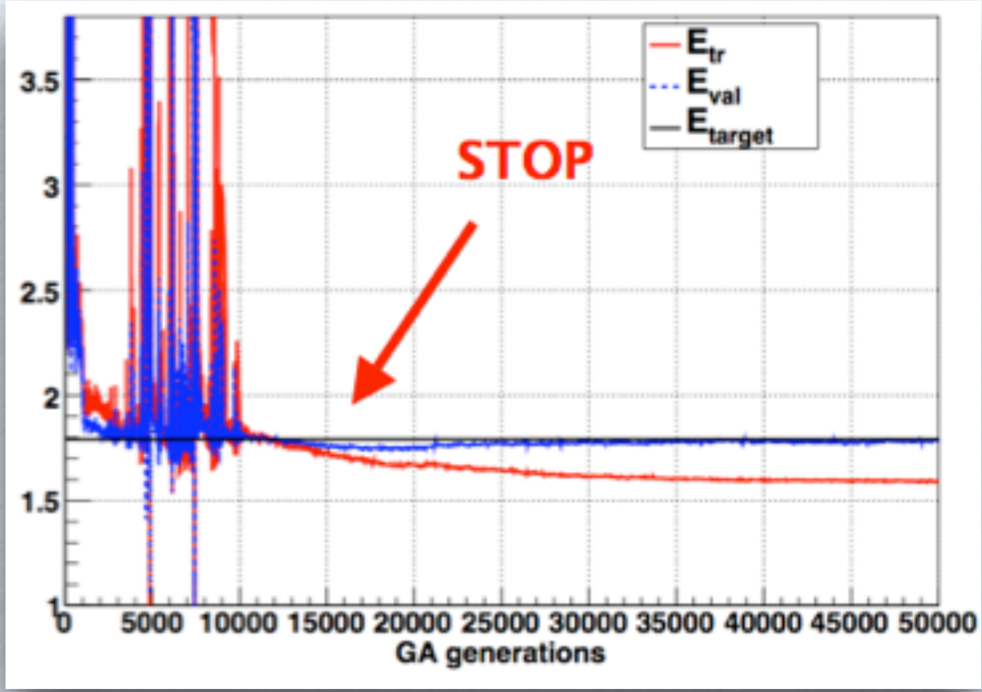
INSTEAD:
Choose replicas of the data, i.e. work in the space of data and project back into PDF space

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

The NNPDF approach

Monte Carlo and Neural Network



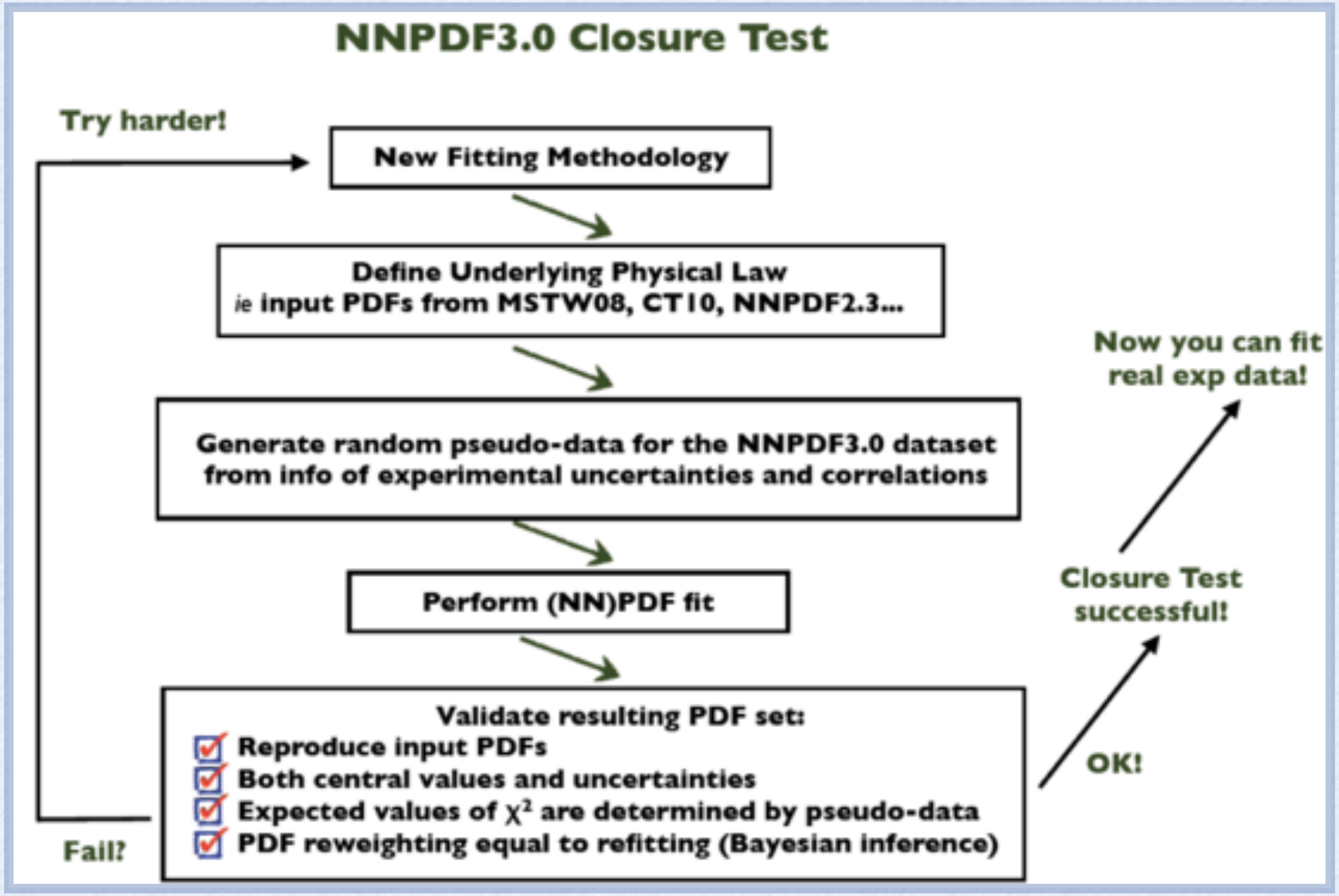


- Neural networks provide flexible and redundant parametrisation
- O(250) parameters versus O(25) parameters of fixed parametrisation
- Same parametrisation for all fits
- Can verify independence of parametrisation
- Cross-Validation method avoids over-learning of statistical fluctuations

Methodology

The closure test

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



Similar tests carried out by Thorne and Watt

Theory

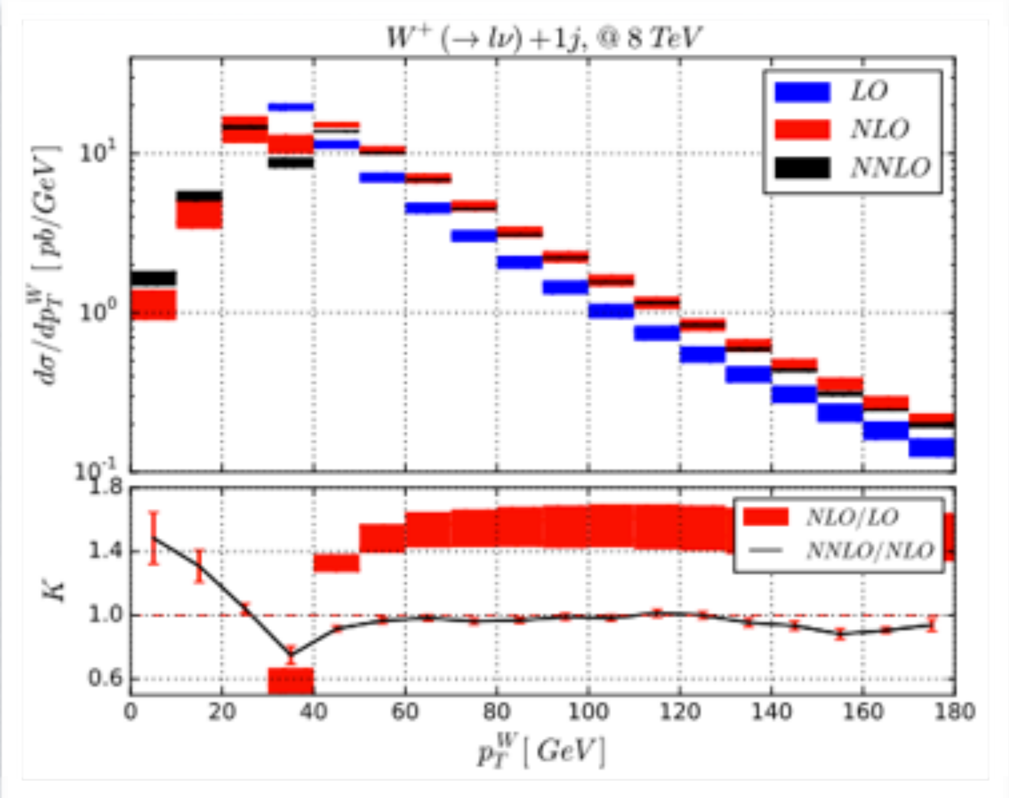
The NNLO revolution

- 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

✓ W+1j also available now at NNLO, soon Z+1j → gluon & quark separation

✓ NNLO inclusive jet production in the gluon gluon channel has been completed → gluon and quarks at large x

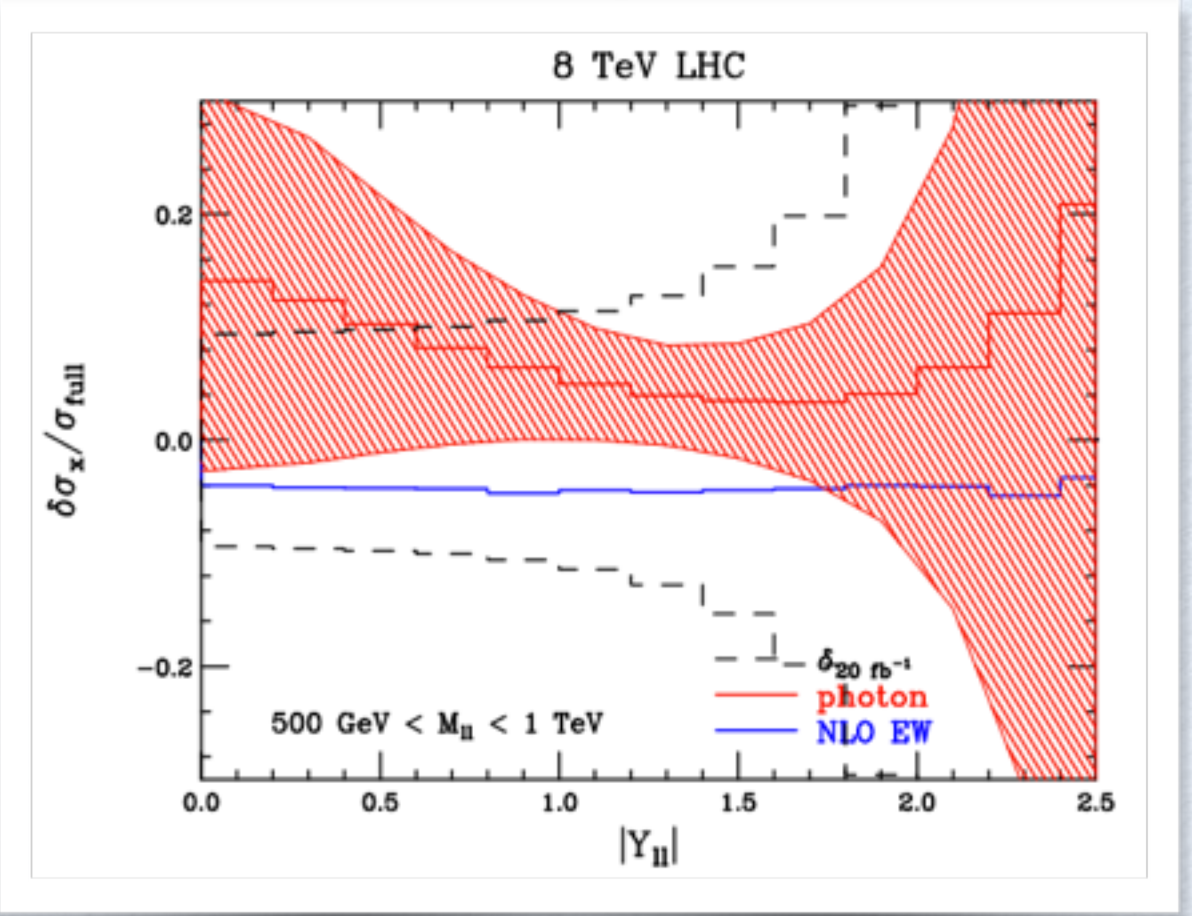


Czakon, Fiedler, Mitov PRL 110 (2013) 25
Boughezal et al, 1504.02131 (2015)
Gehrmann-De Ridder et al, Phys.Rev.Lett. 110 (2013) 16

Theory

QED and EW corrections

- EW corrections become relevant at the current precision level
- Several tools to compute them along with QCD correction [FEWZ3.1, Phys.Rev. D86 (2012) 094034]
- EW corrections can be sizeable especially at large invariant mass
- QED corrections affected by large uncertainty induced from uncertainty on photon PDF

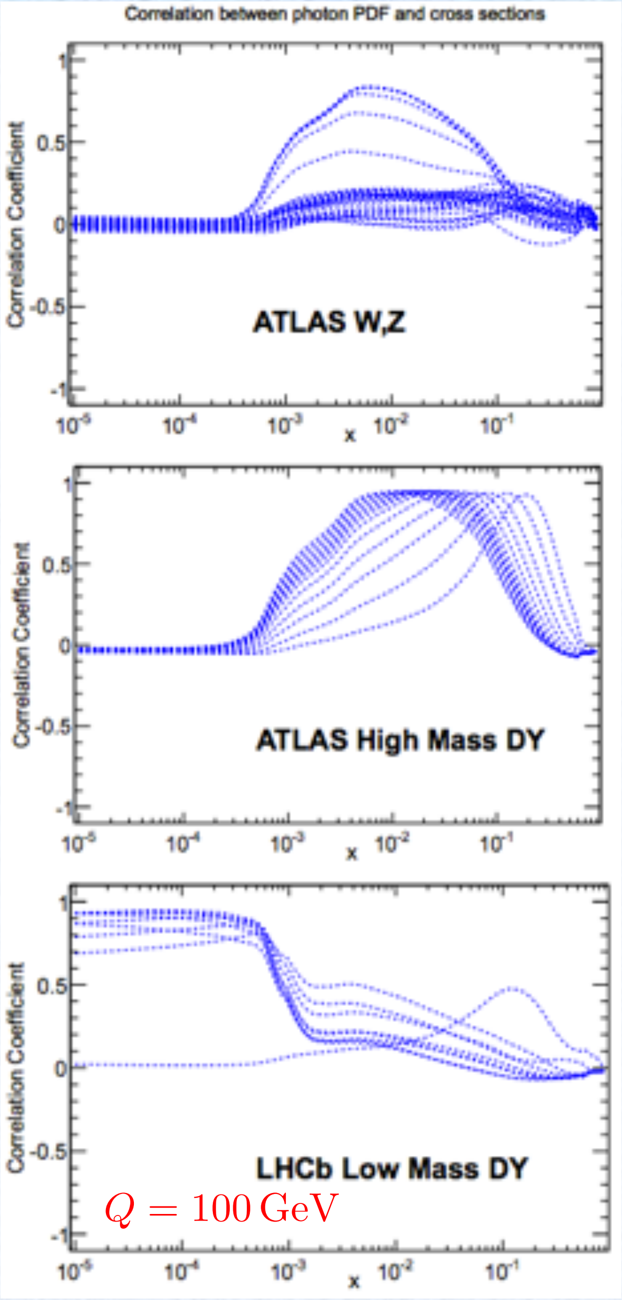


Boughezal, Li, Petriello, Phys.Rev. D89 (2014) 3, 034030

Theory

The photon PDF

- The inclusion of EW corrections requires PDF with QED effects
- **NNPDF23QED** is a recent PDF set with uncertainties which incorporates (N)NLO QCD + LO QED effects. MMHT QED set and CT14 sets expected soon
- Photon PDF fitted from DIS and DY data (on-shell W,Z production and low/high mass DY)
- Photon PDF is poorly determined from DIS data. Need hadron collider processes where photon contributes at LO!

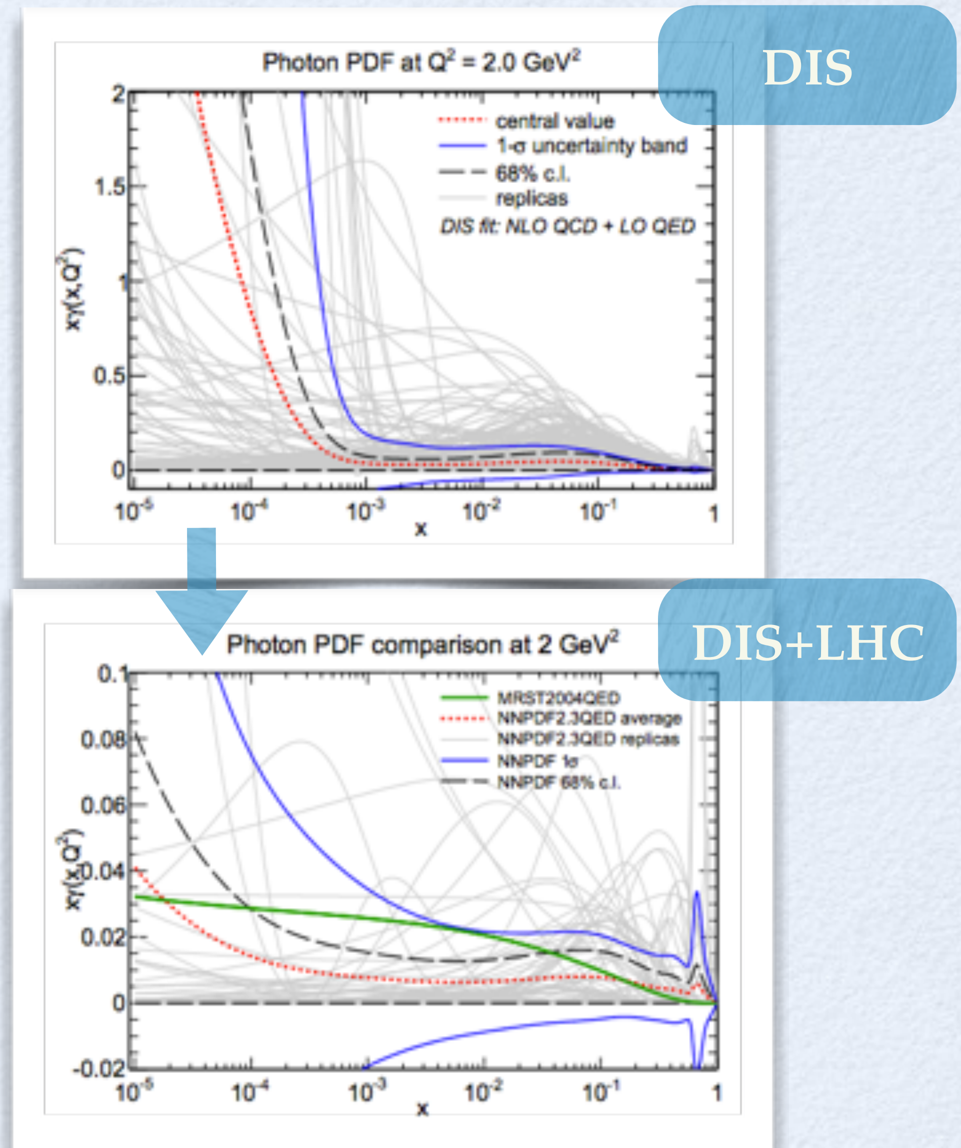


Ball et al, Nucl.Phys. B877 (2013) 290-320

Theory

The photon PDF

- The inclusion of EW corrections requires PDF with QED effects
- **NNPDF23QED** is a recent PDF set with uncertainties which incorporates (N)NLO QCD + LO QED effects. MMHT QED set and CT14 sets expected soon
- Photon PDF fitted from DIS and DY data (on-shell W,Z production and low/high mass DY)
- Photon PDF is poorly determined from DIS data. Need hadron collider processes where photon contributes at LO!

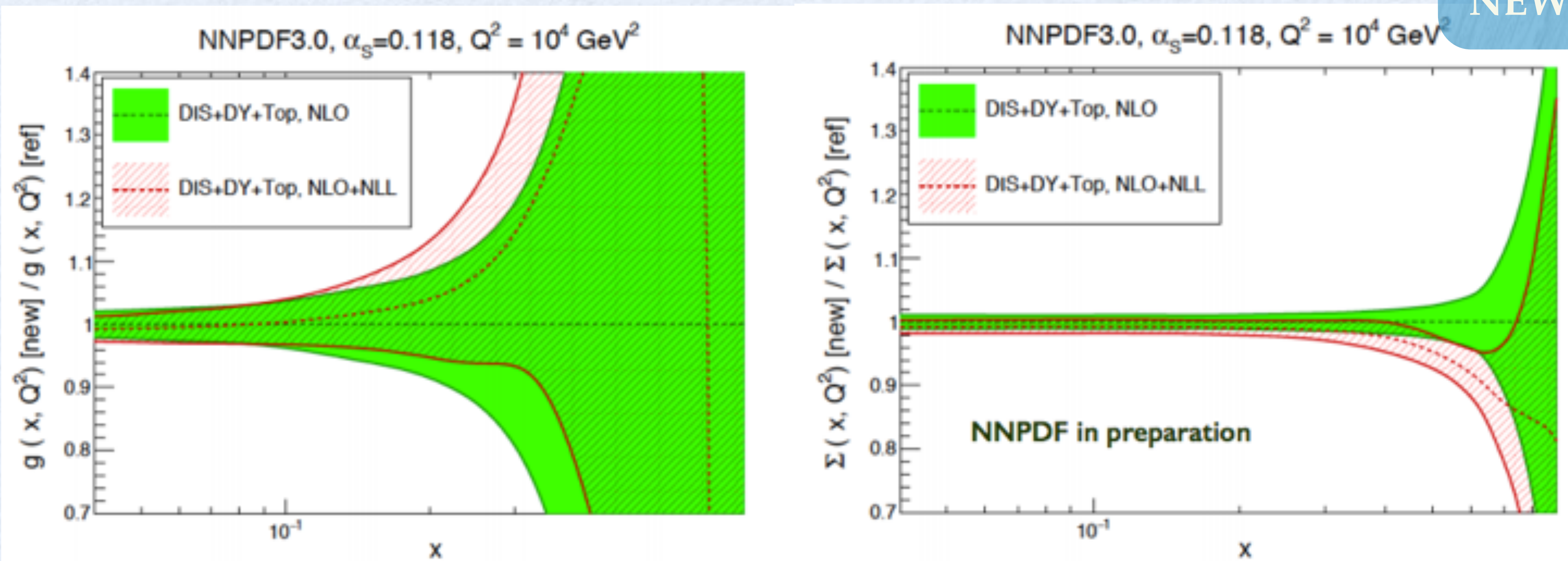


Ball et al, Nucl.Phys. B877 (2013) 290-320

Theory

Threshold resummation

NEW!



NNPDF coll., in progress

- Resummation included for the first time in PDF fit using public codes **ReDY** (Bonvini et al.), **TROLL** (successor of ResHiggs) and **TOP++** (Czakon et al.),
- In a NLO+NLL fit, effects can be large. Up to -20% for quark and +40% for gluons.
- Particularly crucial for high mass Drell-Yan [large-x] and predictions for heavy new physics particle

Data

Inclusion of LHC data

GLUON

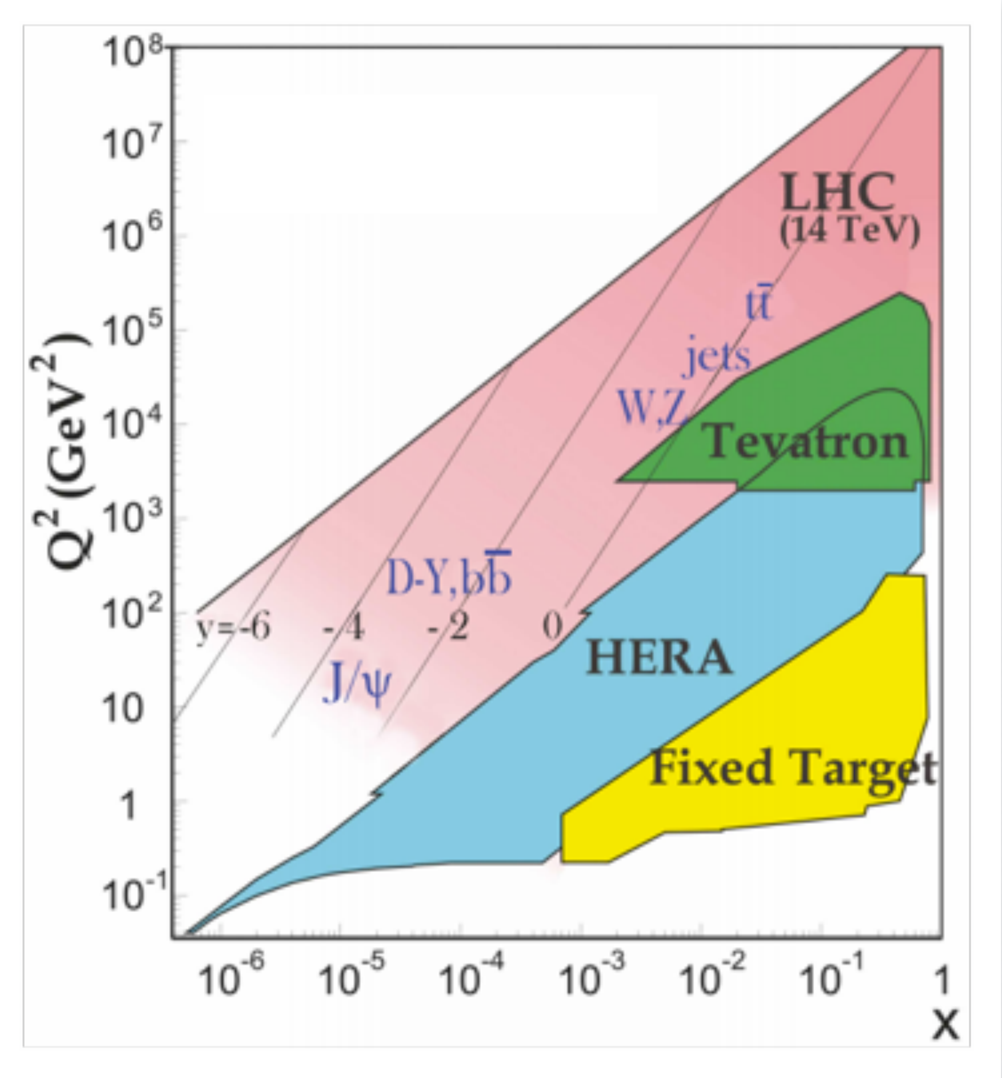
- Inclusive jets and dijets **(medium/large x)**
- Isolated photon and γ+jets **(medium/large x)**
- Top pair production **(large x)**
- High p_T Z(+jets) distribution **(small/medium x)**

QUARKS

- High p_T W(+jets) ratios **(medium/large x)**
- W and Z production **(medium x)**
- Low and high mass Drell-Yan **(small and large x)**
- Wc **(strangeness at medium x)**

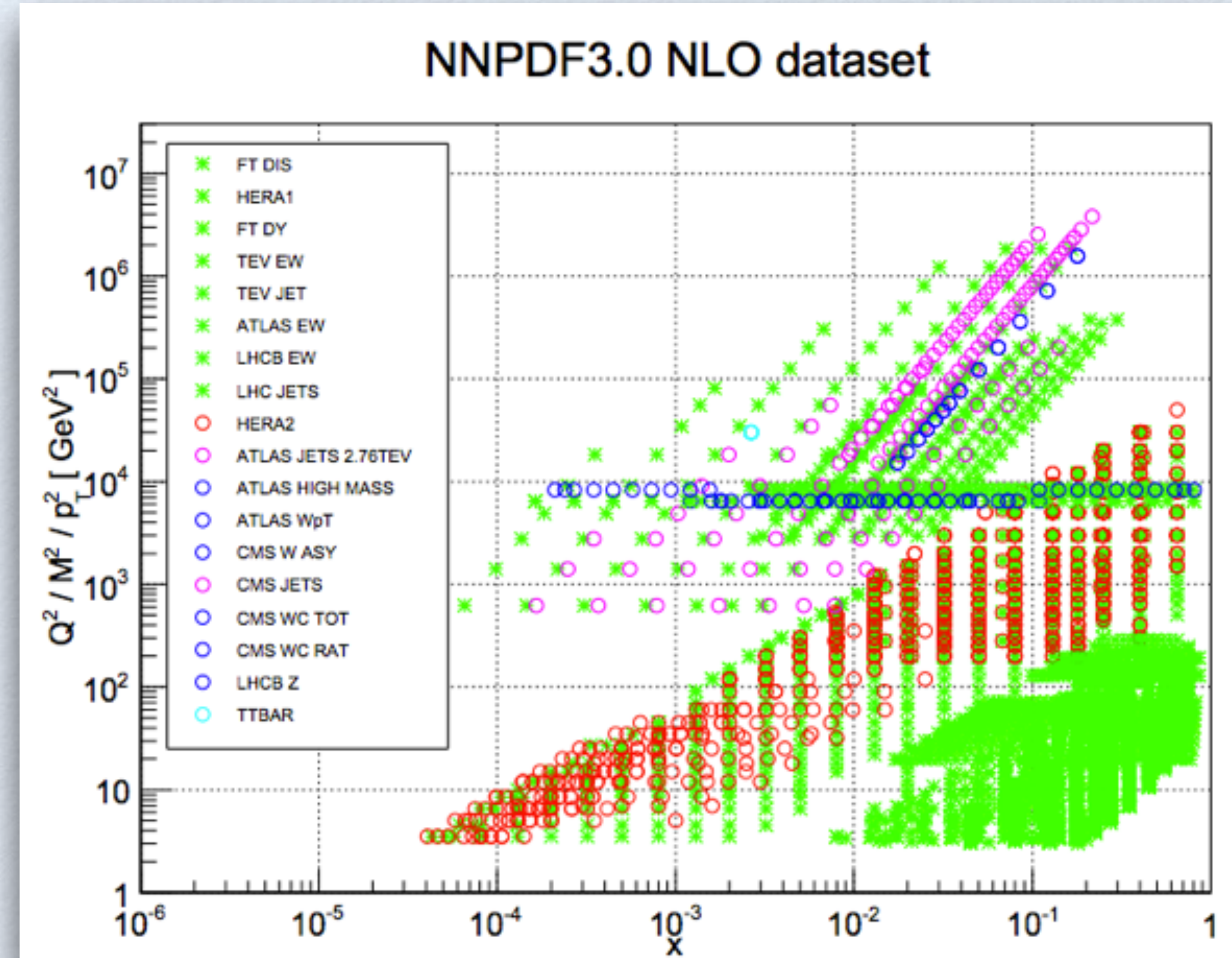
PHOTON

- Low and high mass Drell-Yan
- WW production



Data

The NNPDF3.0 set



NNPDF23

HERAII

new LHC EW

new LHC jets

LHC tt

Data

The NNPDF3.0 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)

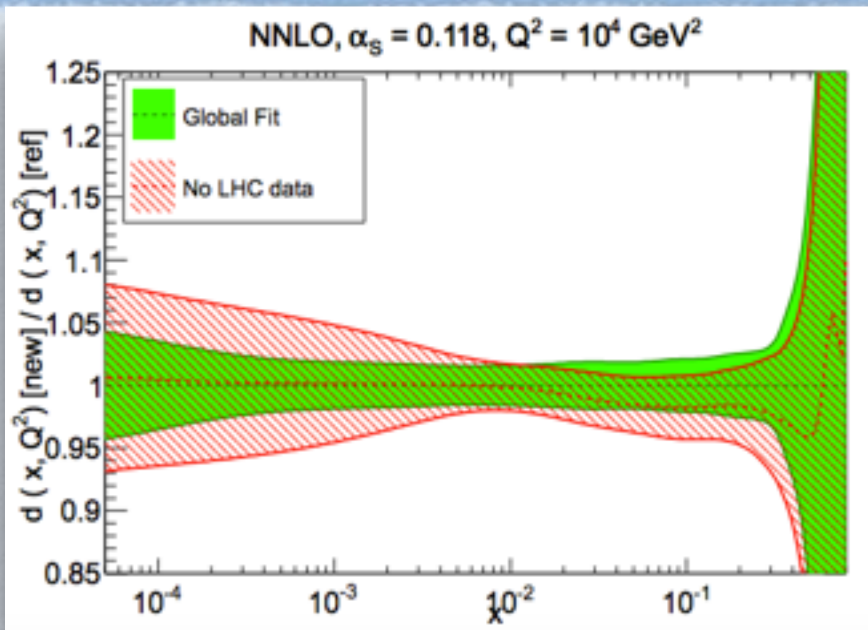
O(1000) NEW data points!

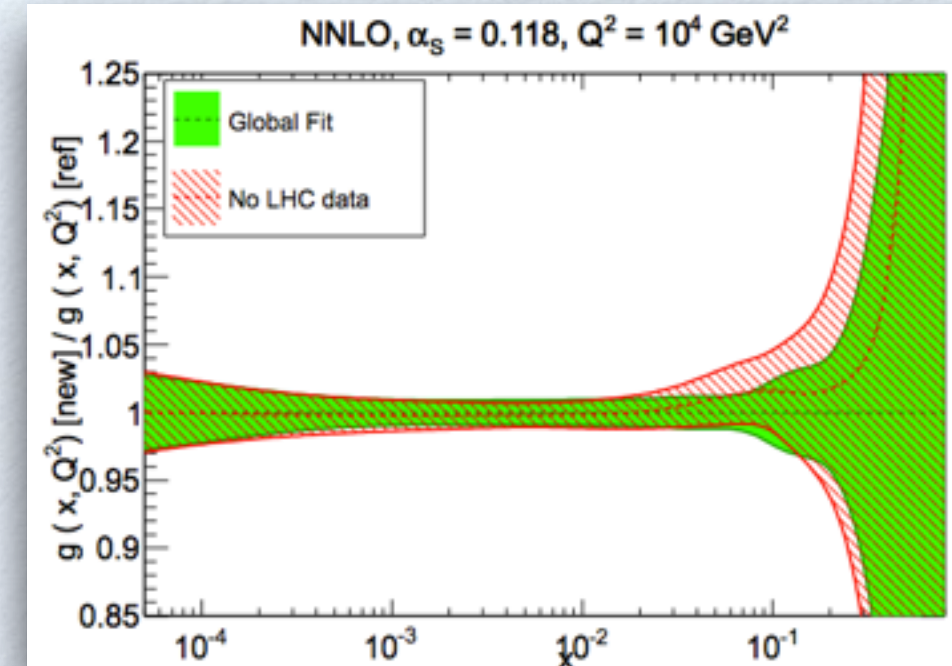
Over 4000 data points:

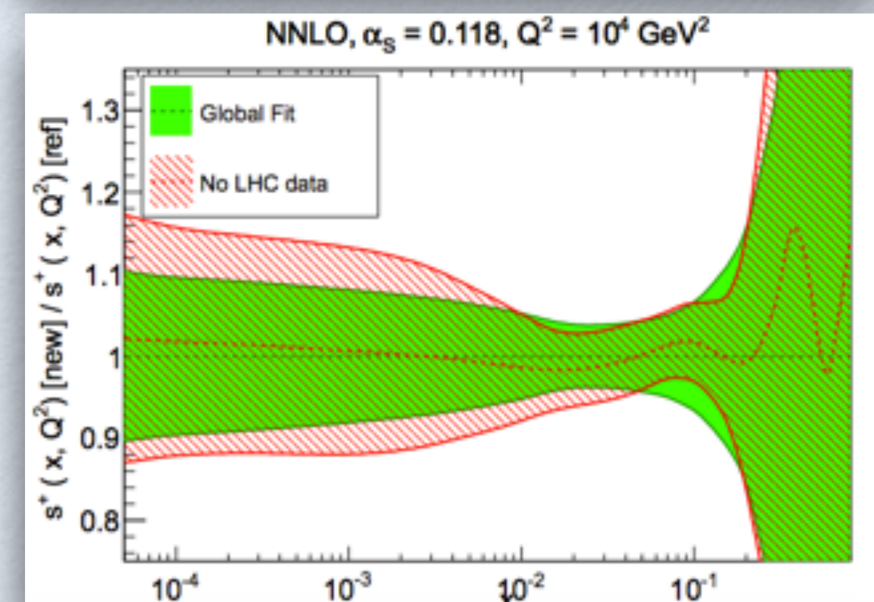
FastKernel + FASTNLO/APPLgrid systematically employed!

Data

The effect of LHC data on the NNPDF3.0 set



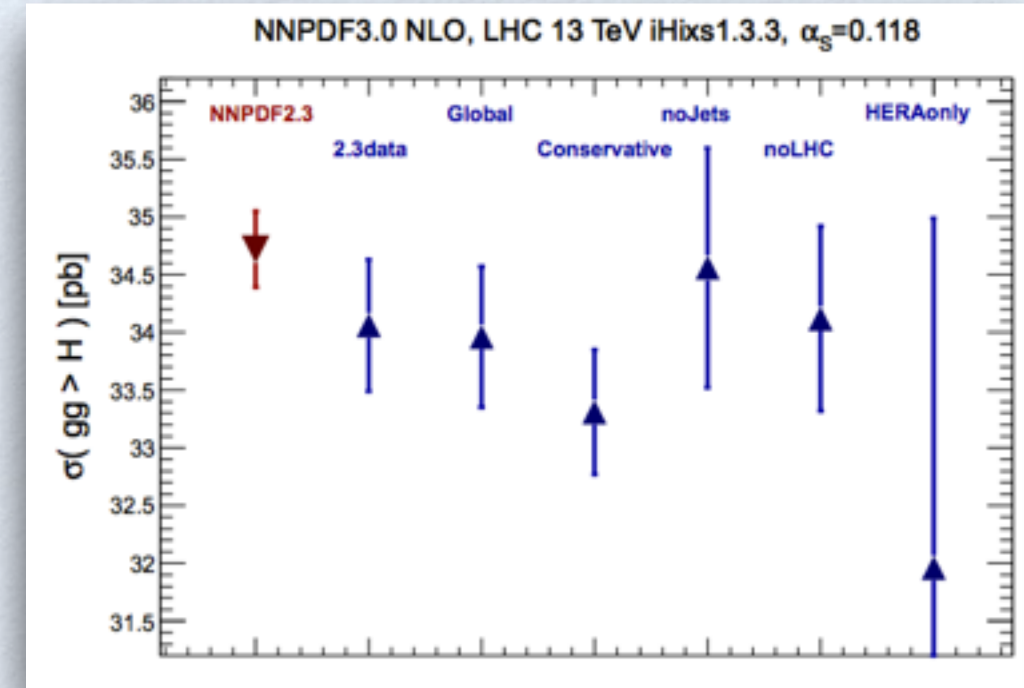


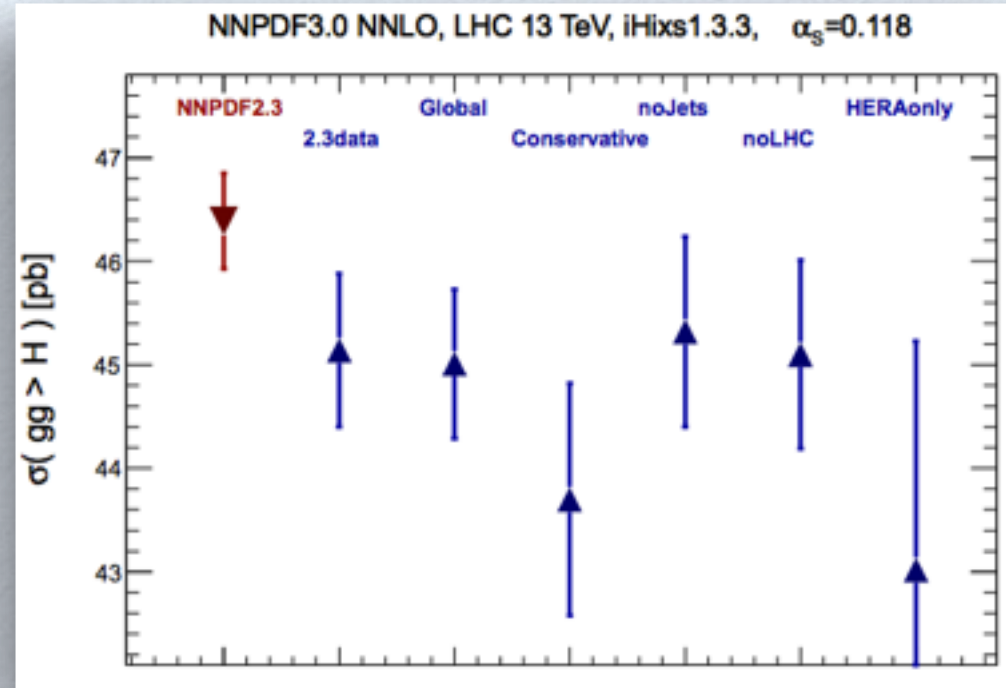


- PDF uncertainty of large-x gluon reduced by inclusion of jet and top quark data
- Uncertainty of light quarks at small x reduced by DY data and W+c
- Description of LHC data, already good with NNPDF2.3 improves in NNPDF3.0

Data

The effect of LHC data Higgs production via ggF

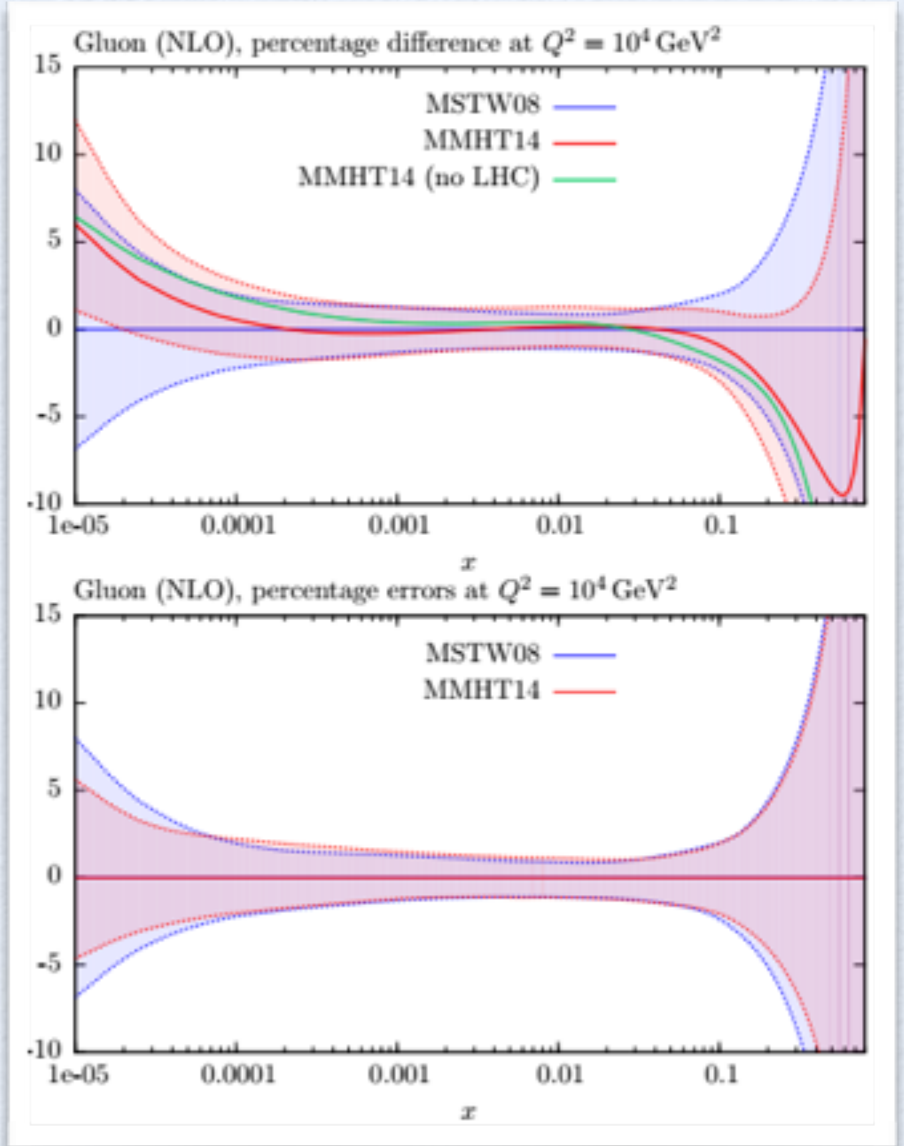


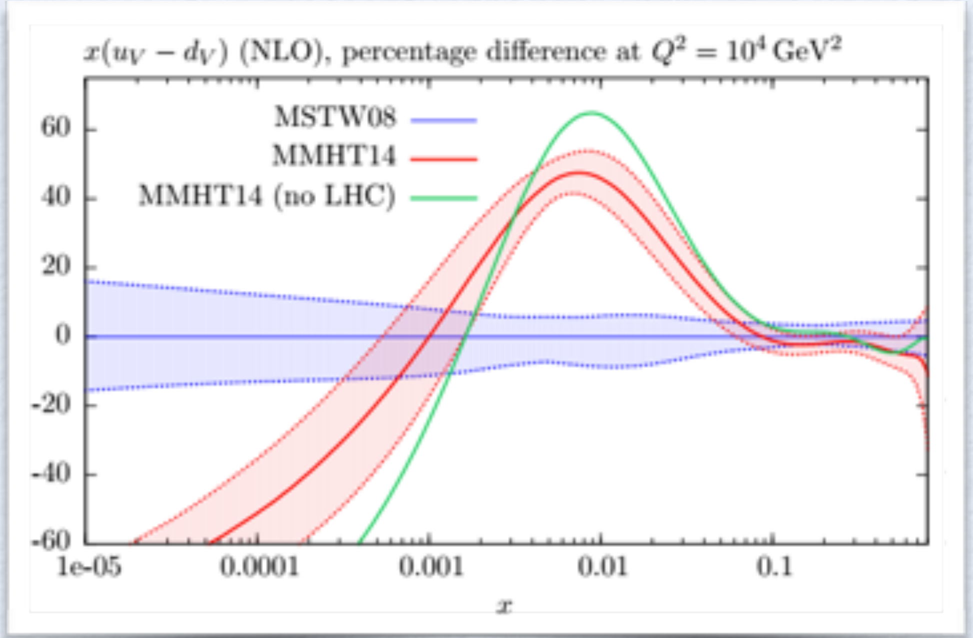


- Softer gluon-gluon luminosity leads to a decrease in the the ggH cross section at LHC 13 TeV
- The effect is most marked at NNLO rather than at NLO, with pull of ~ 1.5
- The ggH process is different from many other processes at LHC since there are no direct experimental constraints on the gluon at 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

Data

The effect of LHC data on the MMHT set



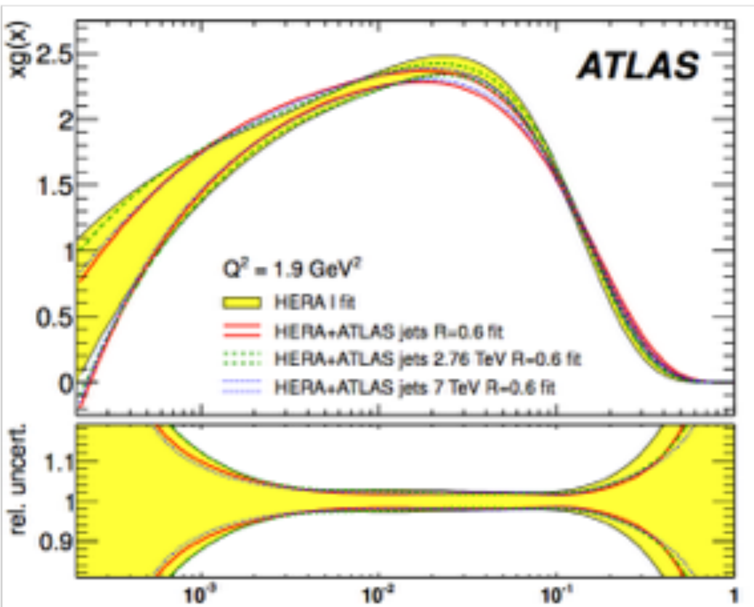


MMHT, arXiv:1412.3989

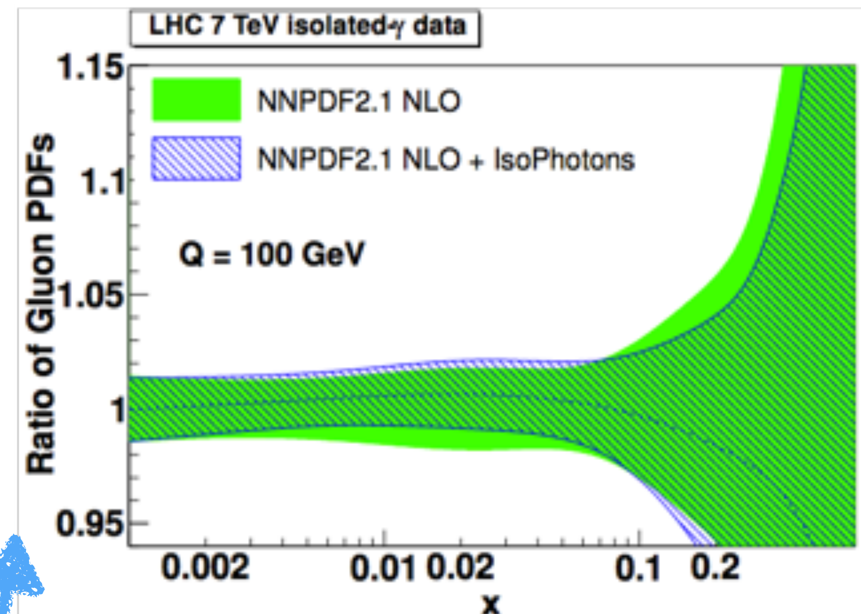
- Large effect on quark flavour decomposition
- Important to disentangle changes due to methodology (parametrisation and fitting) from effect of LHC data
- More studies from ex.collaborations

Data

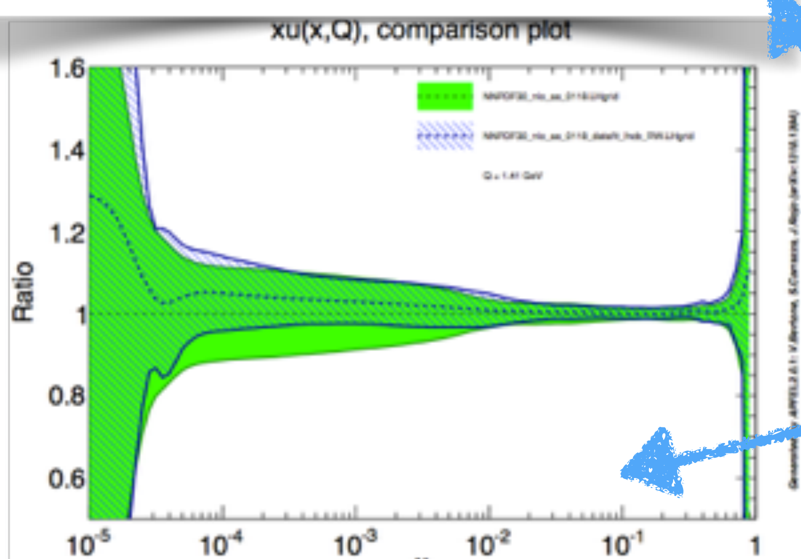
Inclusion of new LHC data in future sets



ATLAS, EPJC (2013) 73 2509



D'Enterria, Rojo, Nucl. Phys. B860 (2012), 311



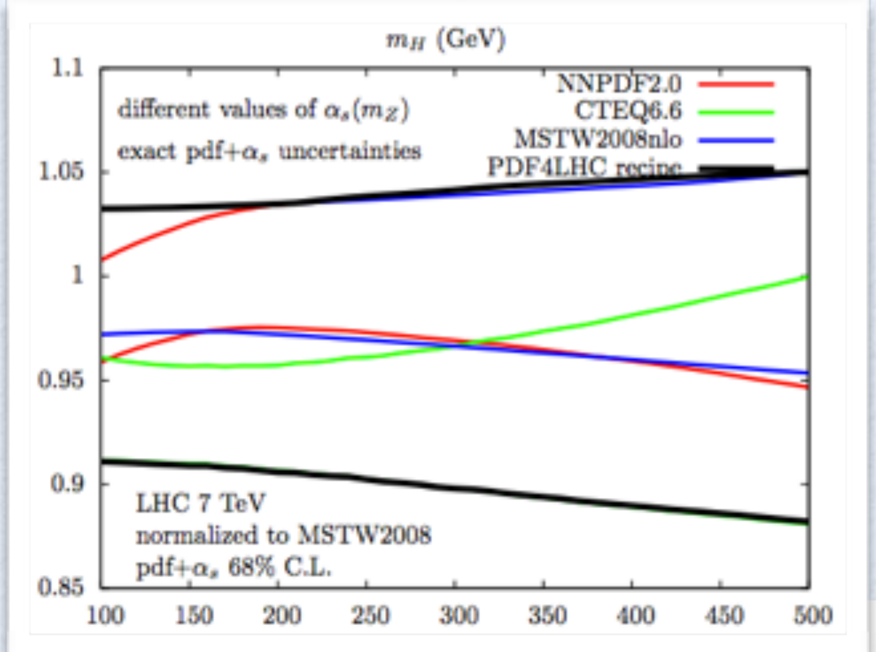
MU, LHCb workshop

- Isolated photon and γ+jets (medium/large x)
- Z/W pT distributions and ratios
- Top pair differential distributions
- Jet data and ratio
- New precise LHCb forward rapidity data
- High/Low-mass Drell-Yan pair production

+ combined HERA-I and HERA-II data, Tevatron legacy data...

Hot topic in the PDF4LHC WG

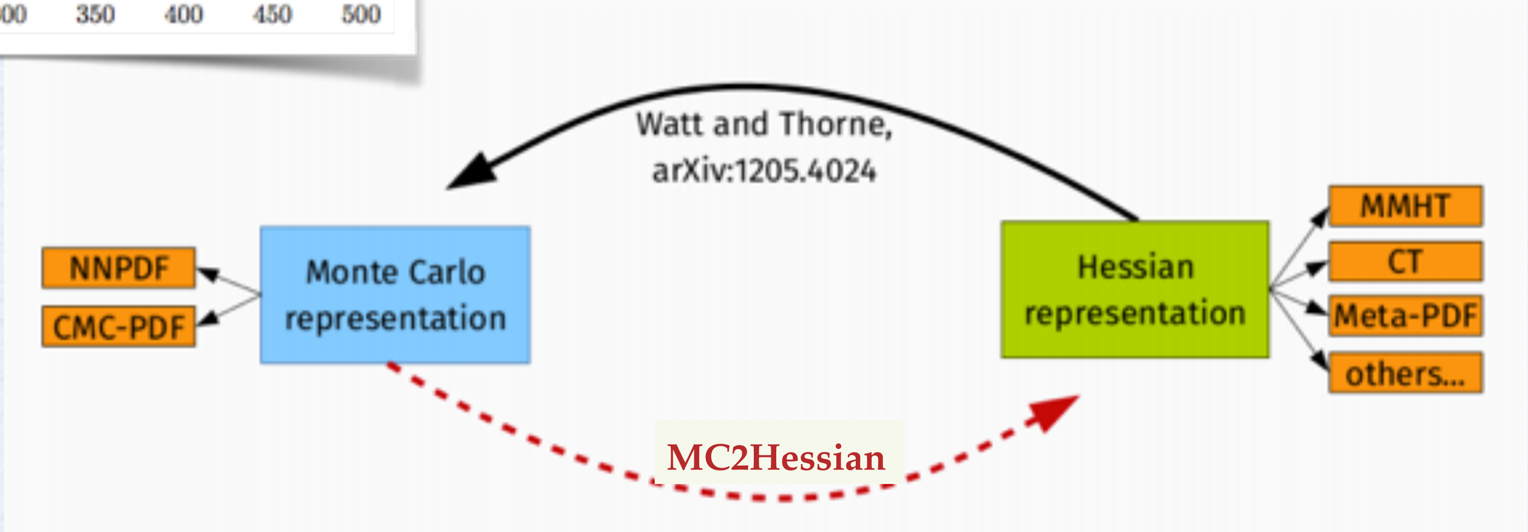
How to combine sets?



Moving forward from PDF4LHC envelope (2010):

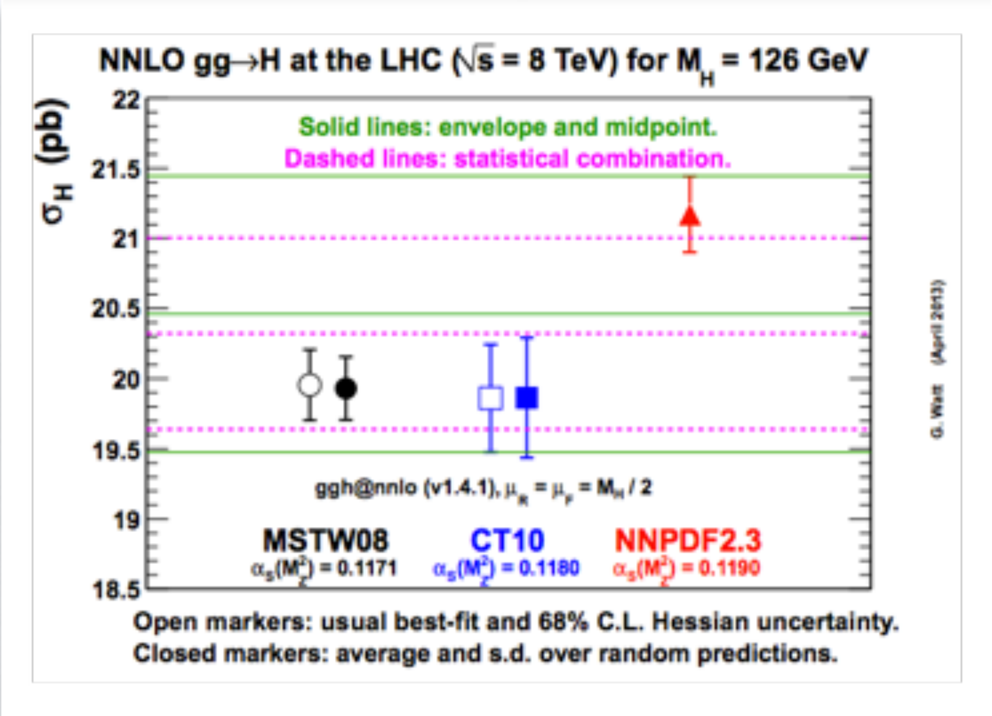
➡ Statistical combination from different PDF groups generating MC sets

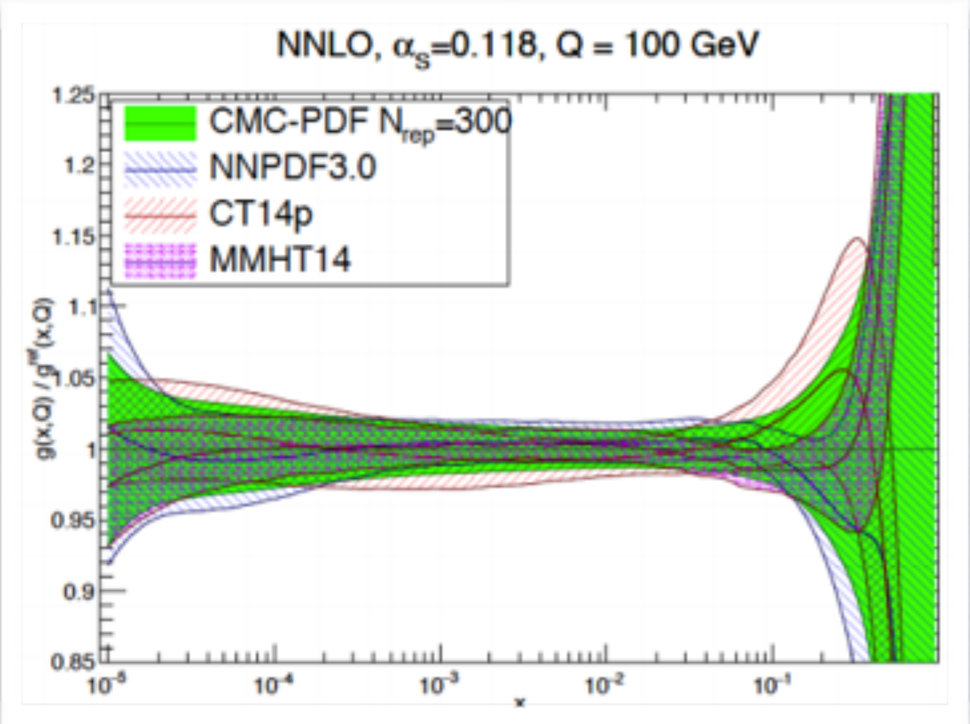
➡ Meta-PDFs: fit with input functional form the CT, MMHT and NNPDF shapes and combine in a unique consistent set



Hot topic in the PDF4LHC WG

How to combine sets?





- Monte Carlo combination of most recent global PDF sets [Forte, Watt]
- Each replica receives the same weight: uncertainty smaller than in the envelope, as in the latter outliers are given a larger weight
- New compression studies: N=40 replicas are virtually identical to the original 300 replicas from the point of view of correlation, standard deviation, observables [Carrazza et al.]
- Ongoing benchmark between compressed set of Monte Carlo replicas and meta-parametrisation

Conclusions and Outlook

- PDF uncertainties are still limiting factor in achieving precise predictions
- Fast progress in recent months, new PDF sets, inclusion of new data, more solid theory and methodology
- NNPDF3.0 is the first closure-test validated set available
- Still a lot of work ahead

* Fast interface to NNLO observables
* N(N)LO+NLO EW fits with initial photon
* Effect of parton shower resummation in PDF fits
* Small-x resummation
* Definition of theoretical uncertainties in PDF fits

* HERA I+II combination
* Loads of new data from LHC and new observables to be investigated

* Statistically-sound PDF combination
* Closure tests and measure of data consistency