NNPDF3.0:

A next generation parton set for the LHC

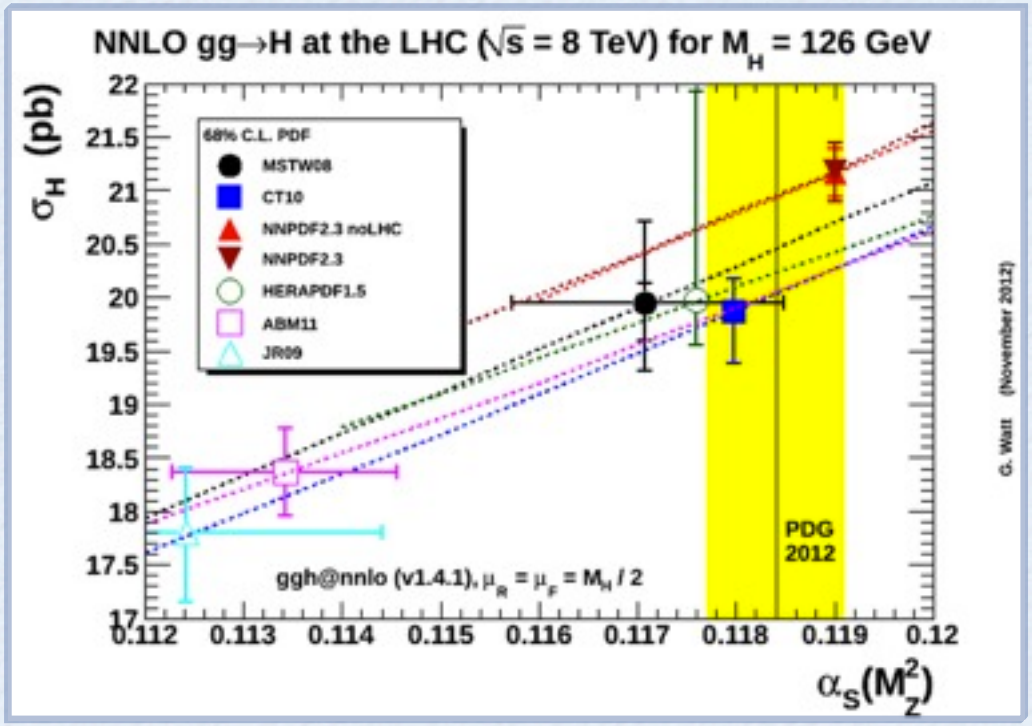
M. Ubiali (University of Cambridge)

for the NNPDF collaboration:
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

29th April 2014
DIS2014 workshop, Warsaw, Poland

Motivation

PDFs and LHC interplay

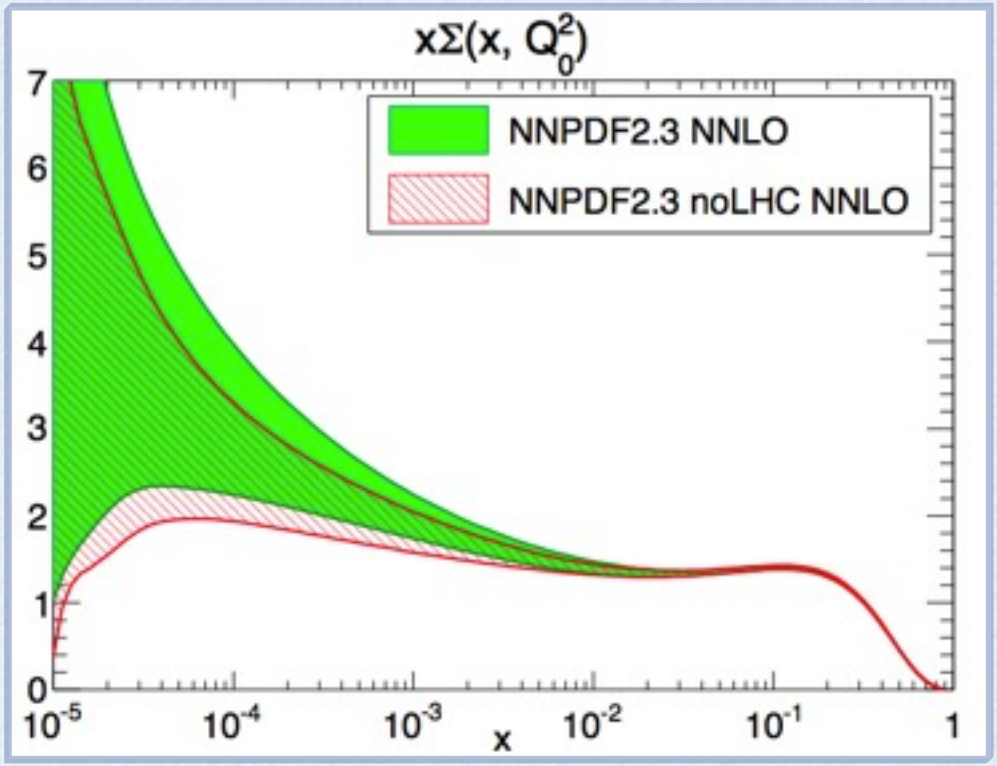


S. Forte, G. Watt, ArXiv:1301.6754

- LHC data provides huge amount of information in kinematical regions where PDFs are not well constrained or rely on fixed-target data: must fully exploit its physics potential

- PDFs are crucial input for the LHC: a faithful determination of PDF uncertainties (experimental, statistical and theoretical) is vital for precision physics and exclusion/discovery in BSM searches

NNPDF, Nucl.Phys. B867 (2013) 244-289



Outlook

- Introduction
- The NNPDF3.0 parton set
  - New methodology
  - Closure test
  - New data
- Results
  - Description of the LHC data
  - The NNPDF3.0 partons
- Conclusions

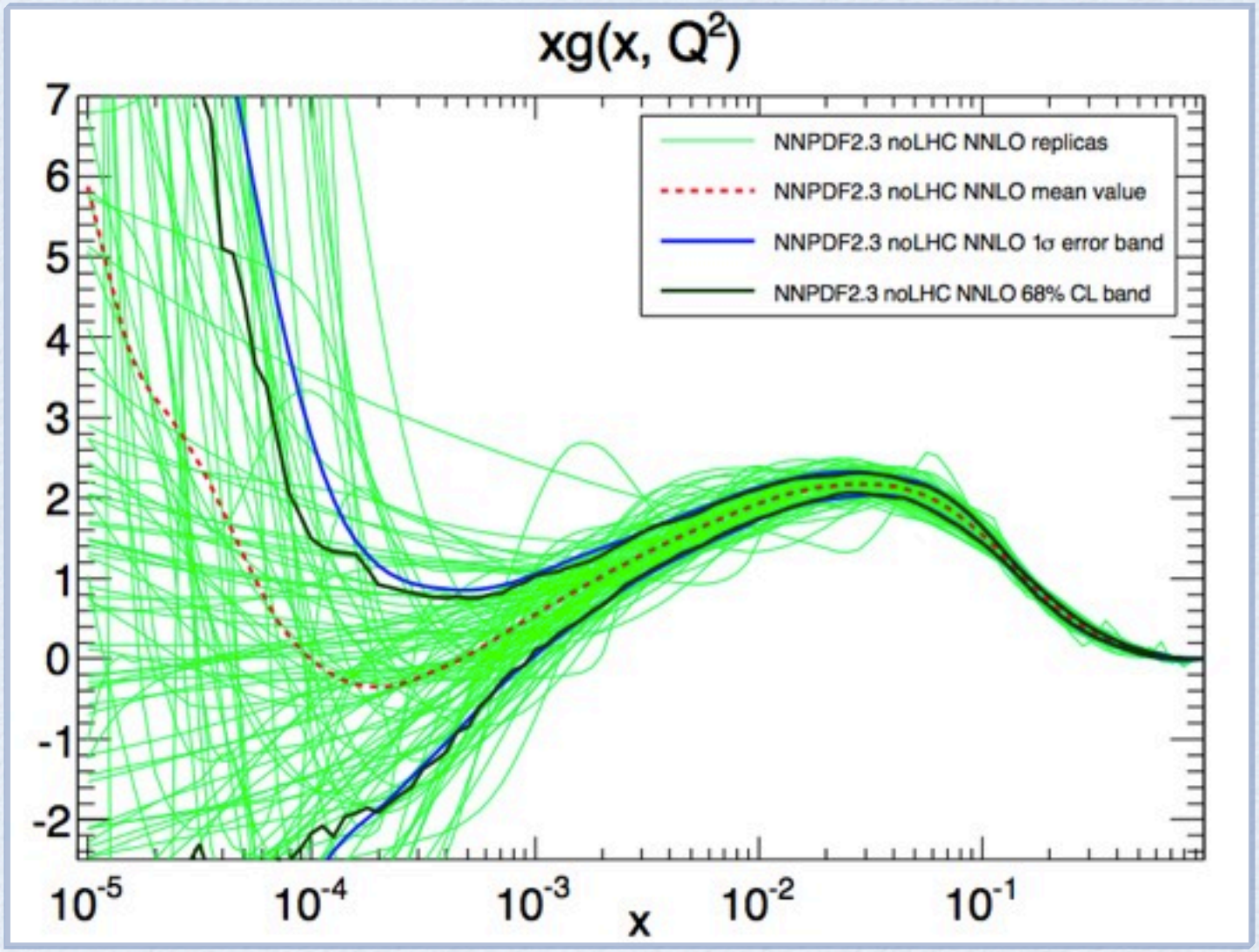
Introduction

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

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

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



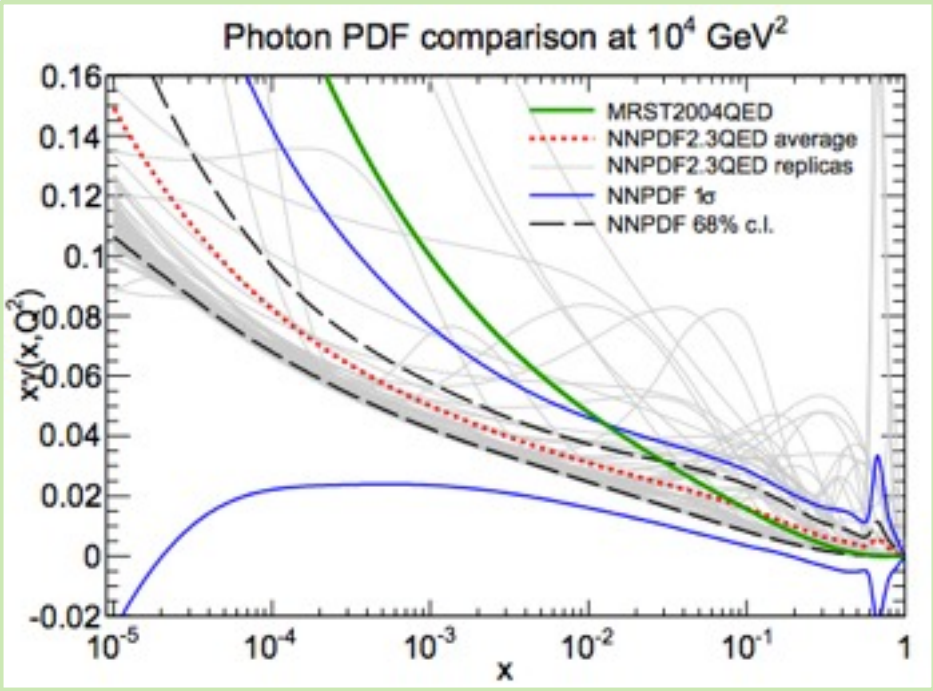
Introduction

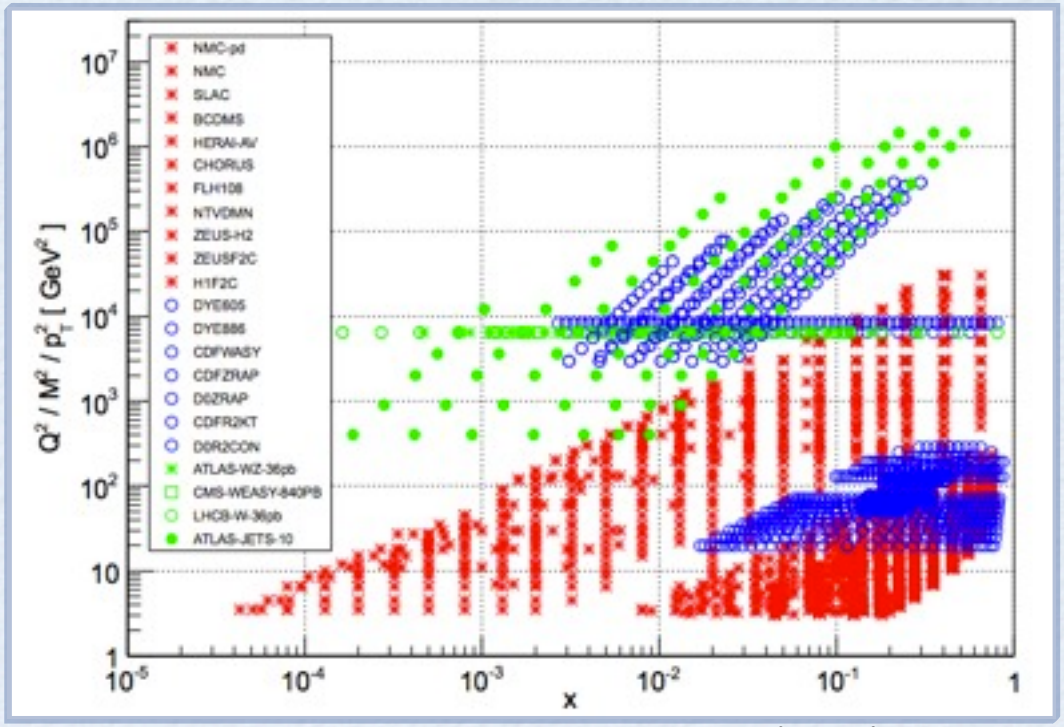
The NNPDF latest releases

NNPDF2.3: improved genetic algorithm and inclusion of early 2010 LHC data with all information on correlations

- ATLAS W and Z lepton rapidity 35 pb^{-1}
- ATLAS inclusive jets 36 pb^{-1}
- LHCb W rapidity at large Y 36 pb^{-1}
- CMS W electron asymmetry 36 pb^{-1}

NNPDF, Nucl.Phys. B877 (2013) 2, 290-320





NNPDF, Nucl.Phys. B867 (2013) 244-289

NNPDF2.3QED: PDF set with QED corrections, uncertainty on photon PDF and inclusion of photon-initiated processes at the LHC

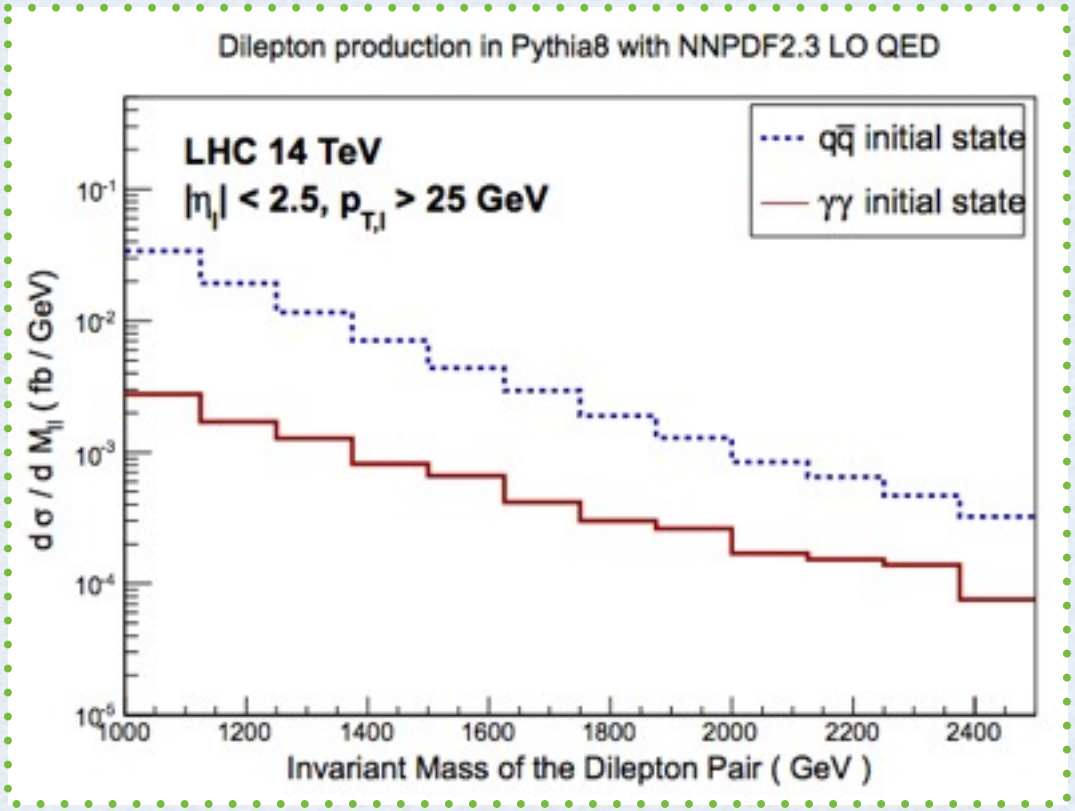
- ATLAS W and Z lepton rapidity 35 pb^{-1}
- ATLAS γ /Z high mass 4.9 fb^{-1}
- LHCb γ /Z low mass 37 pb^{-1}

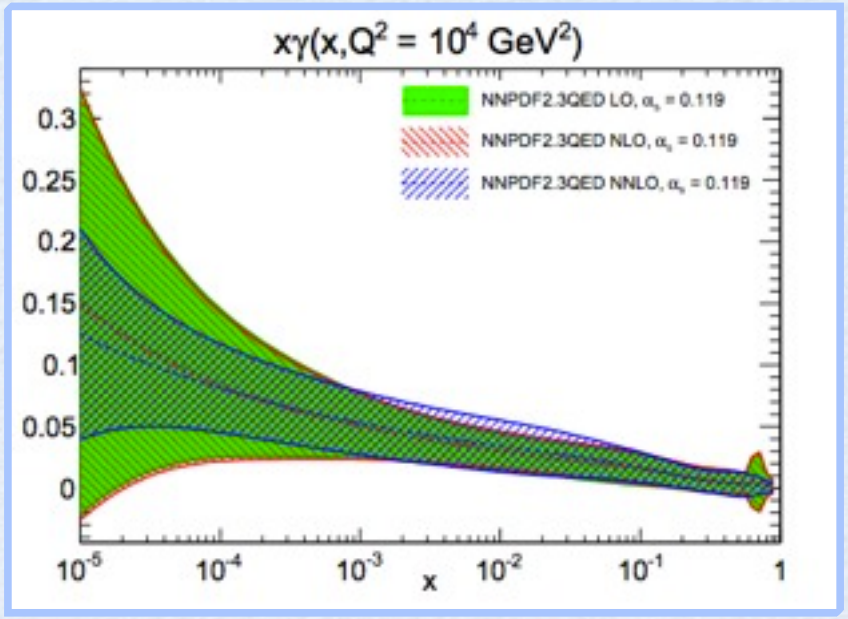
Introduction

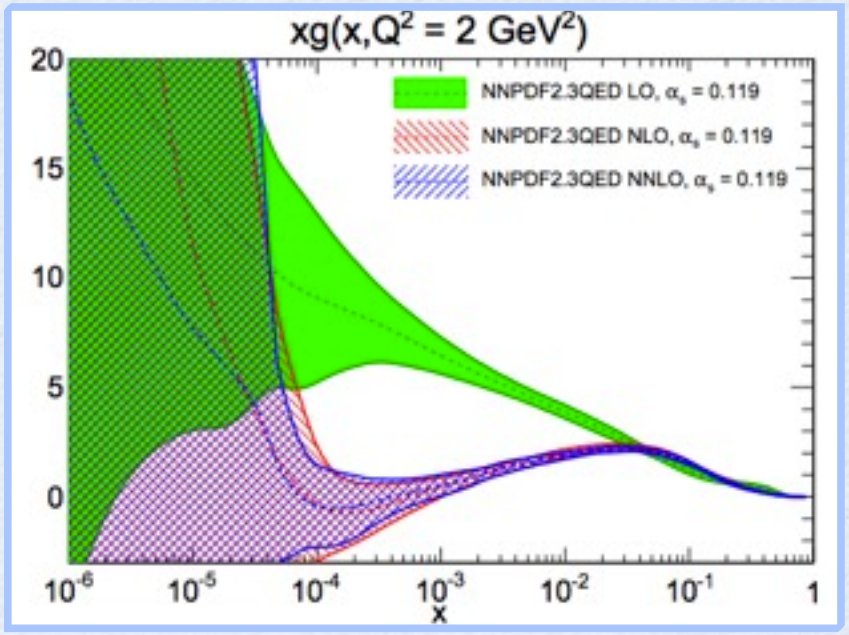
The NNPDF latest releases

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]



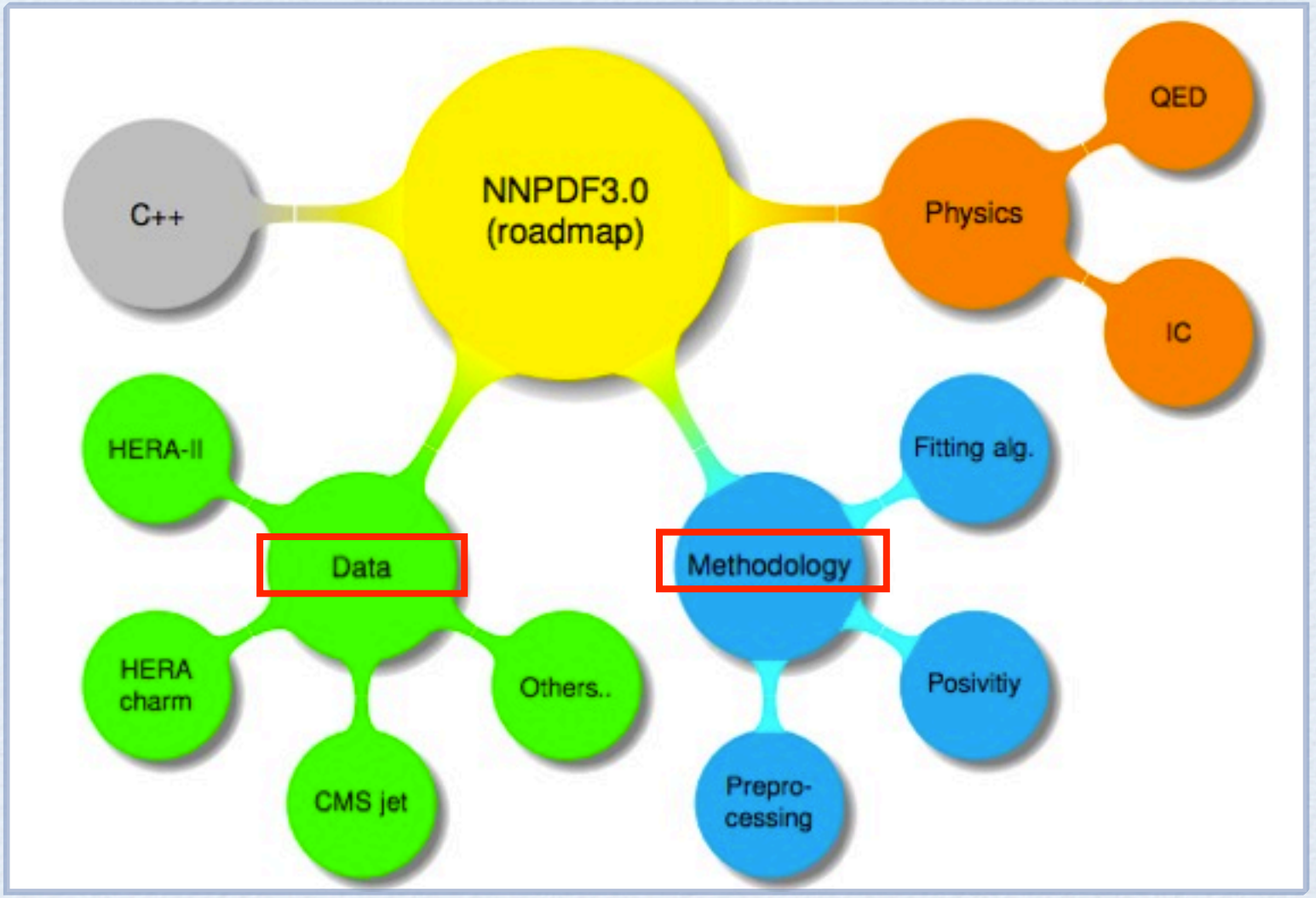




Introduction

Moving forward: the NNPDF3.0 set

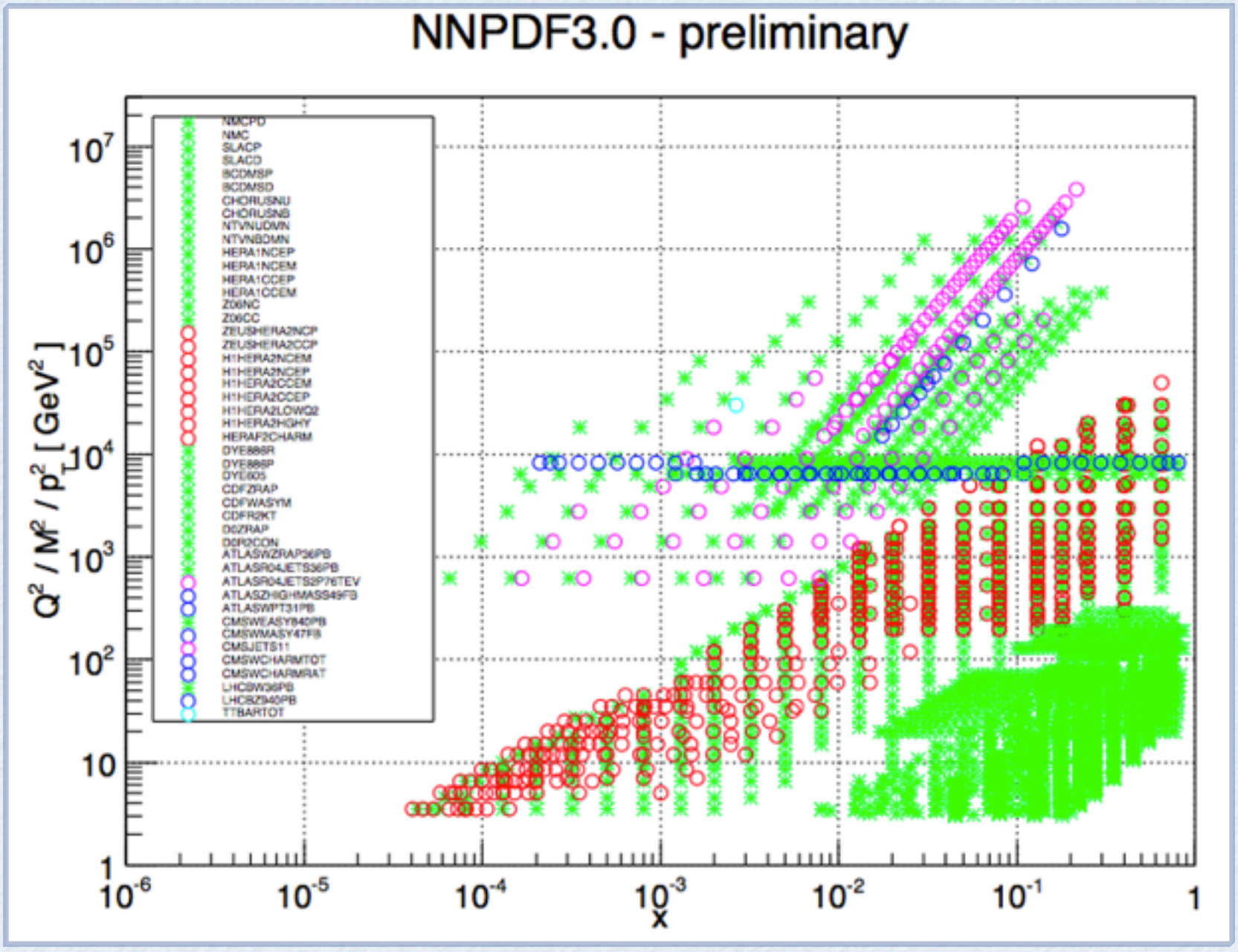
- Major update
- Code completely re-written in c++
- Completely re-designed fitting methodology based on closure test with known underlying physical law
- Tested Weight Penalty method based on iterative Bayesian regularization
- More than 1000 new data points from HERA II and LHC



S. Carrazza, PDF4LHC 2013

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!

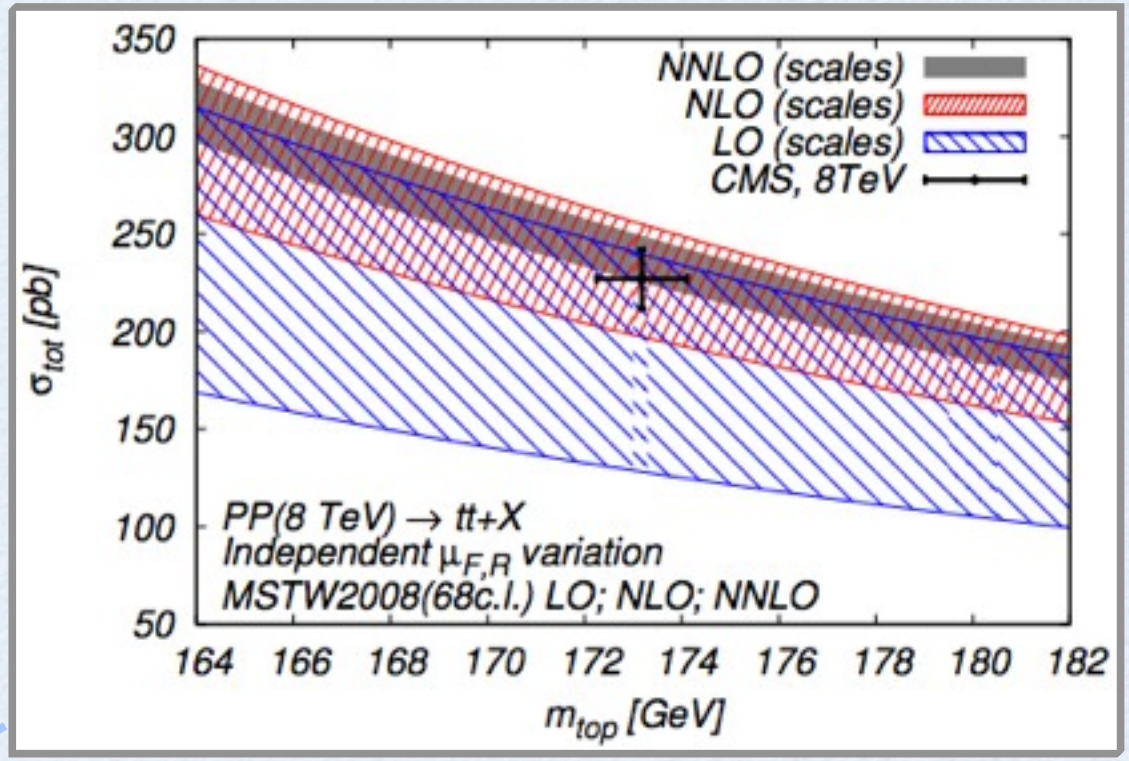
The NNPDF3.0 set

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

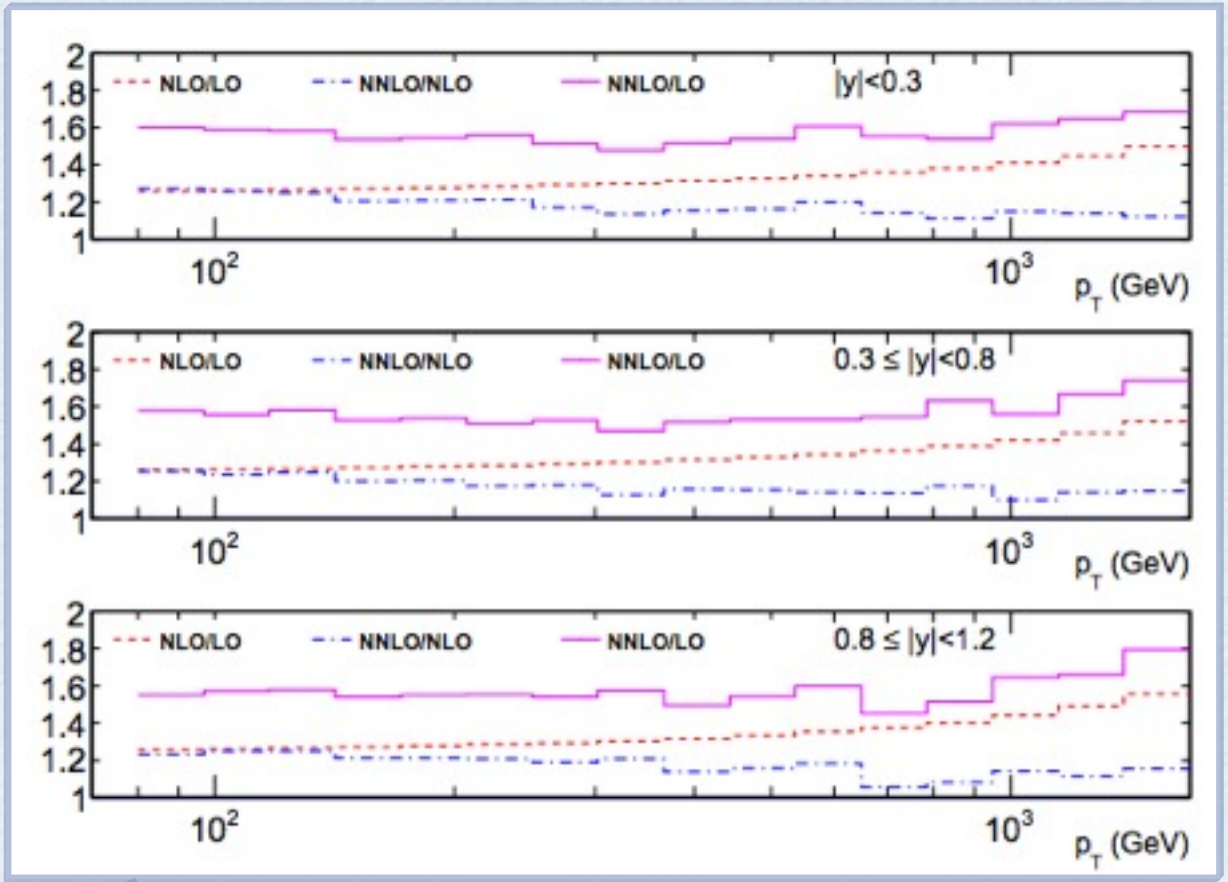
Theoretical aspects: higher order corrections

- 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

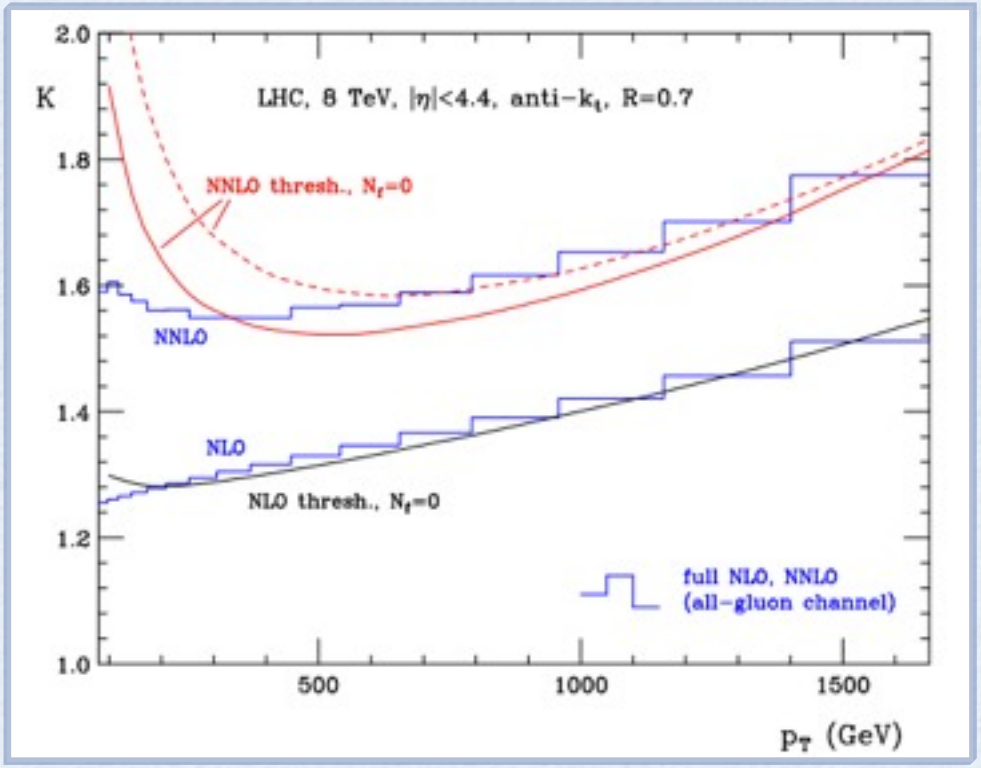
For jets full NNLO calculation is not yet available but...

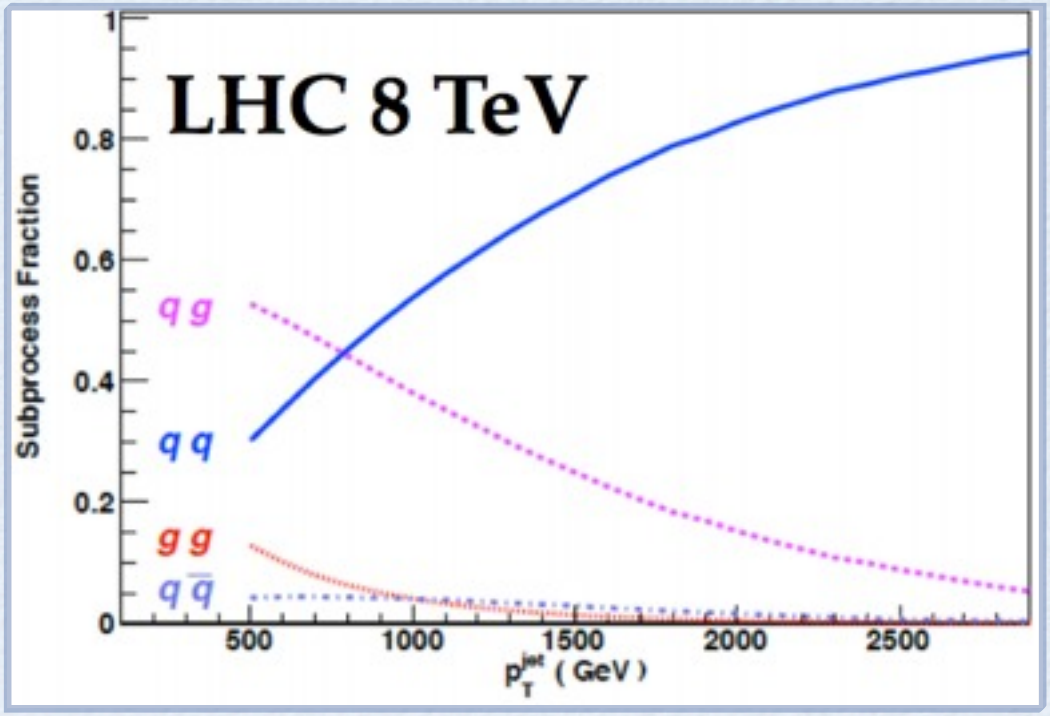
In gg channel up to 20-25% enhancement of NNLO wrt NLO result

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**]





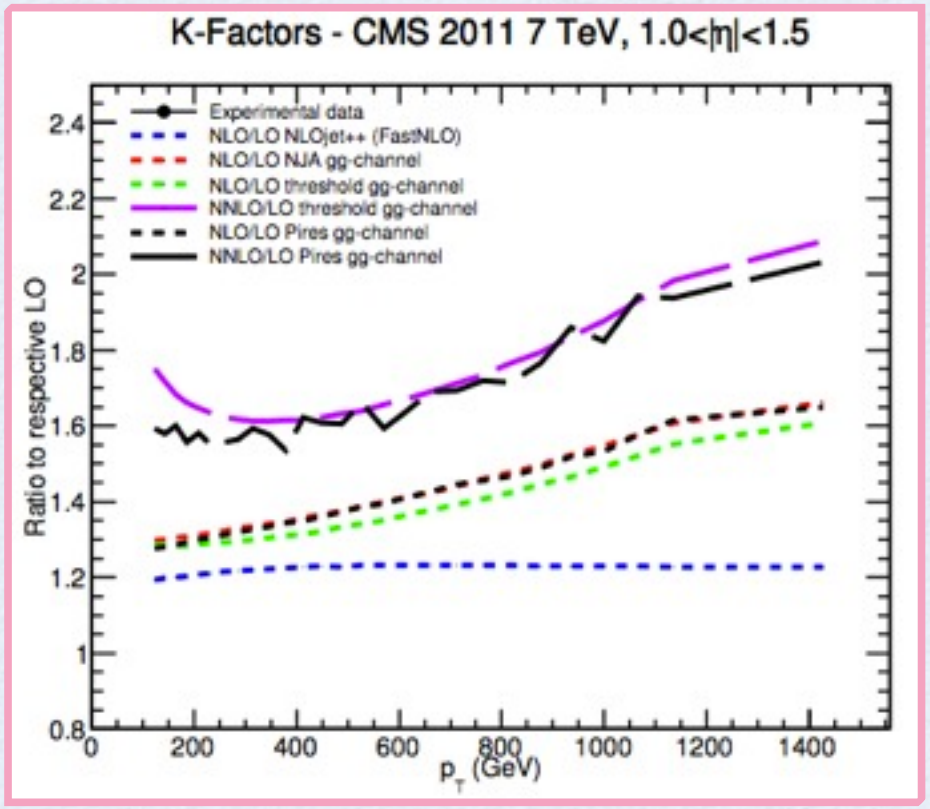
[De Florian et al, Phys.Rev.Lett. 112 (2014) 082001]

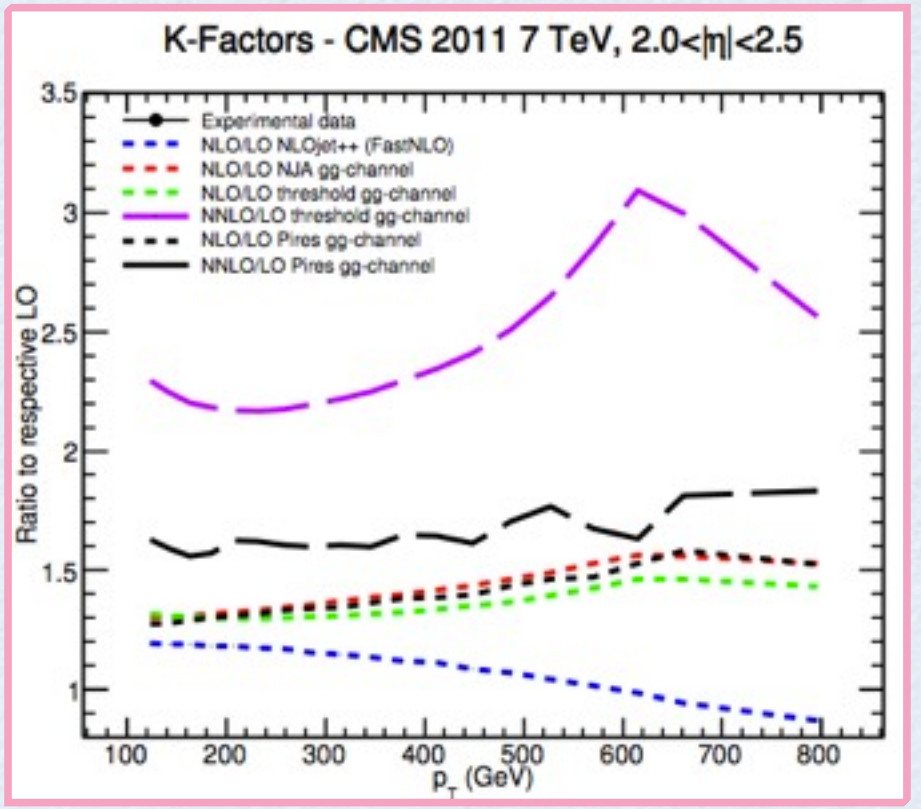
- Comparison between NNLO approx and full NNLO in the gg channel can determine for which value of pT and η NNLO approx can be trusted
- This assumes NNLO K-factors similar in all channels

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

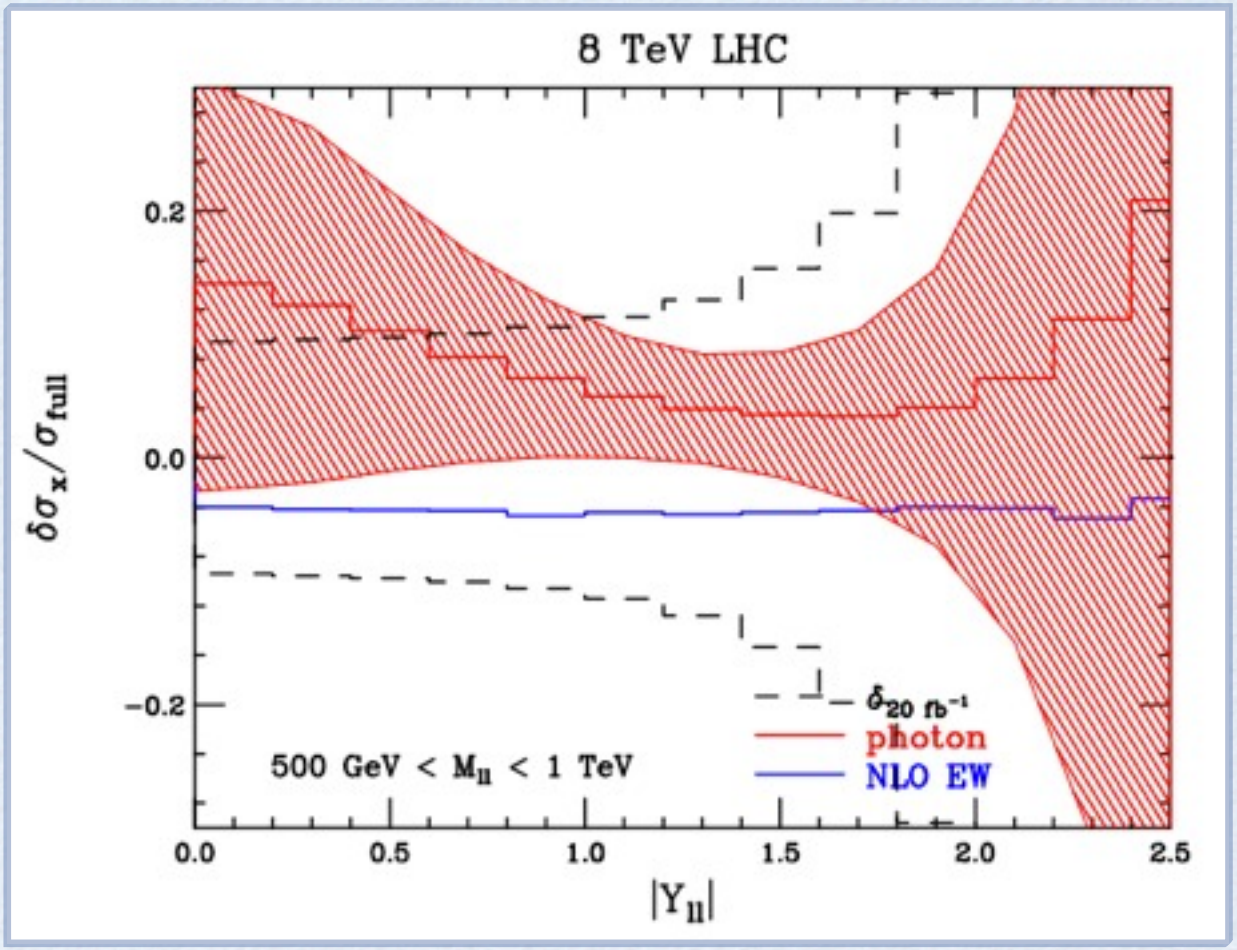
Theoretical aspects: higher order corrections

- 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

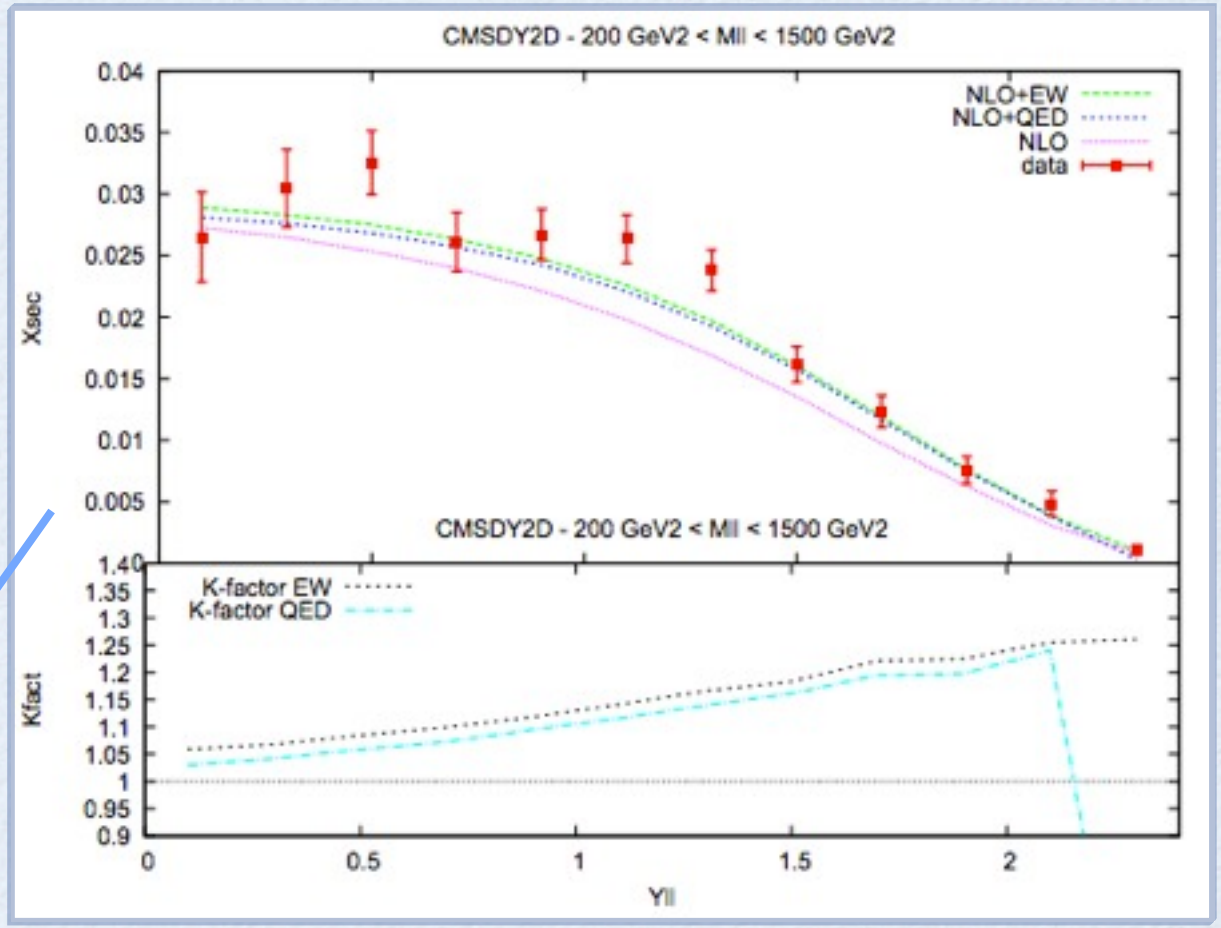
Theoretical aspects: higher order corrections

- 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



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

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

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**]

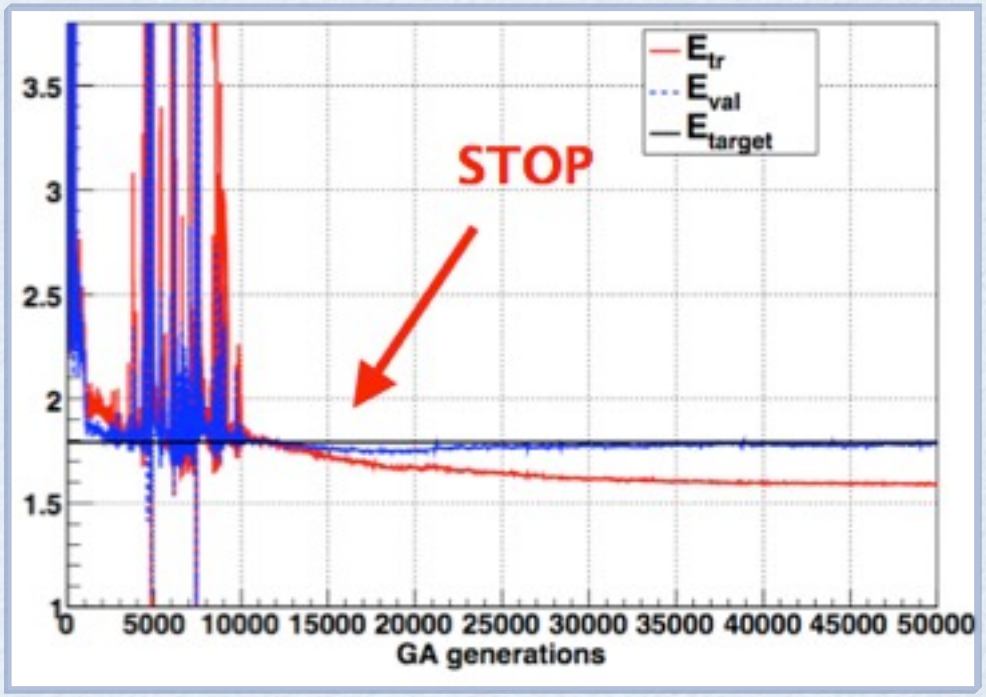
$$E[d,f] = \tfrac{1}{2}\chi^2[d,f] + \alpha\Delta[f]$$

penalty function related to the complexity of each NN

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

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

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



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

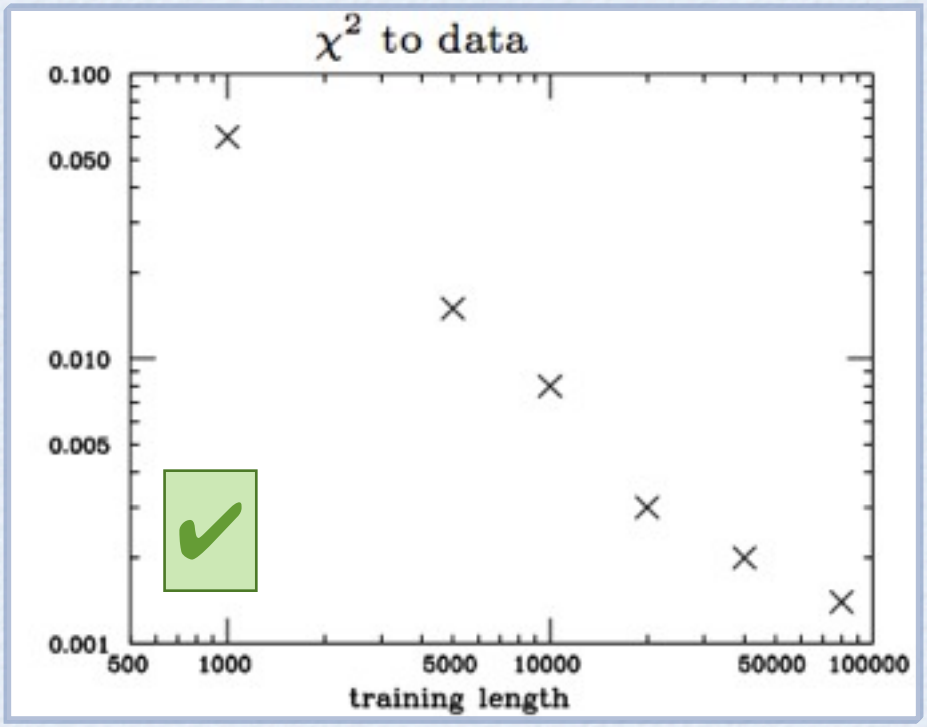
The NNPDF3.0 set

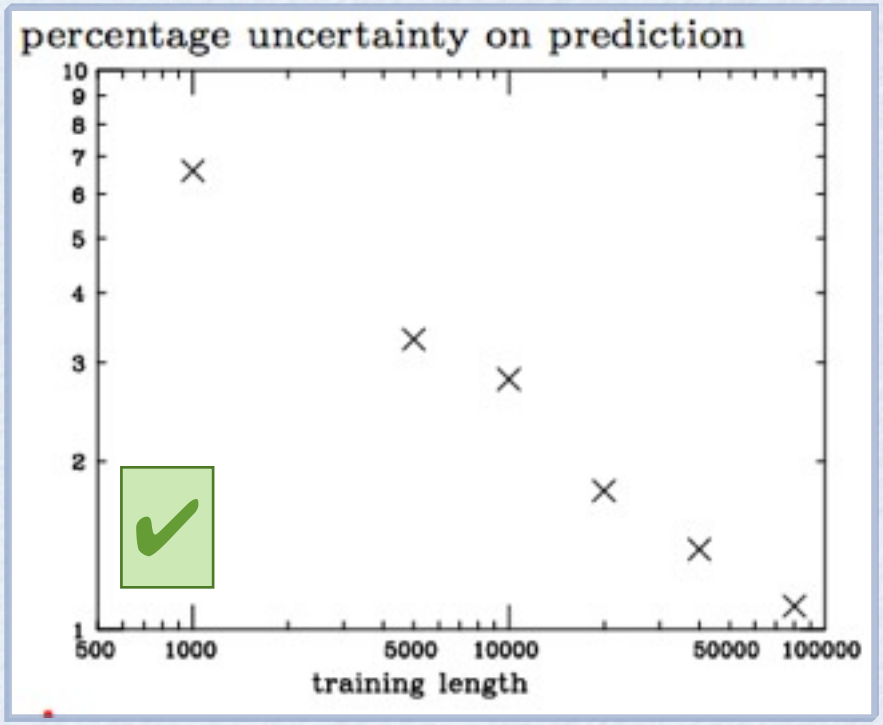
Improved methodology: closure test

How do we determine the best minimization strategy?

(i) Assume underlying PDFs known (say, MSTW2008)
(ii) Generate data with given statistical and correlated systematics
(iii) Perform a fit and compare to the "truth"

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



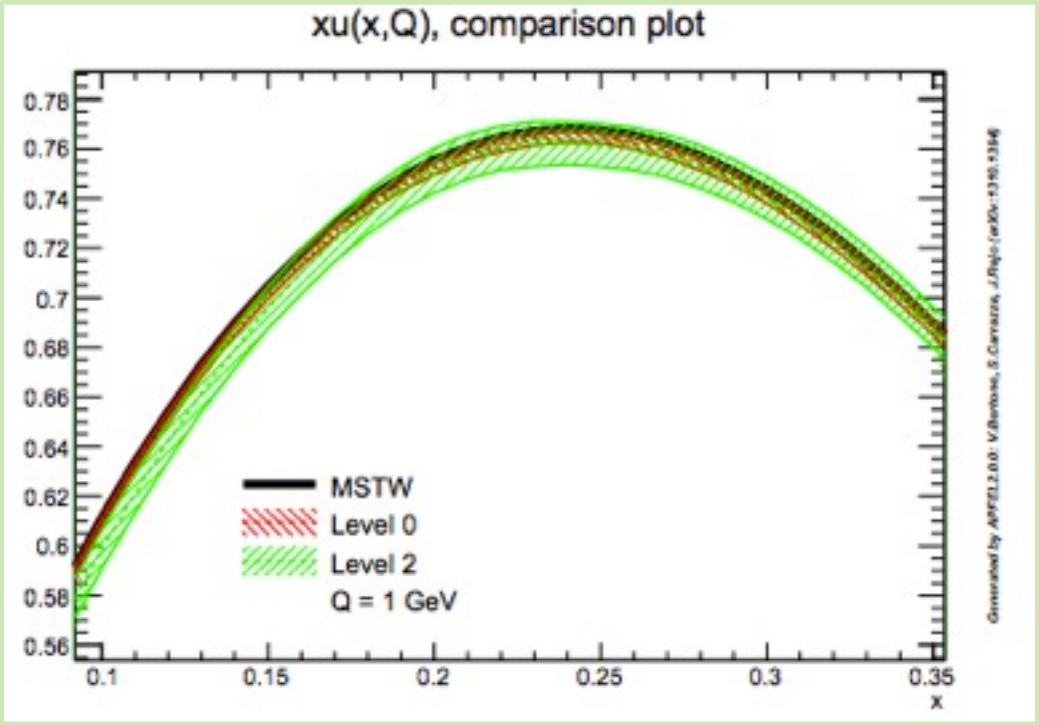


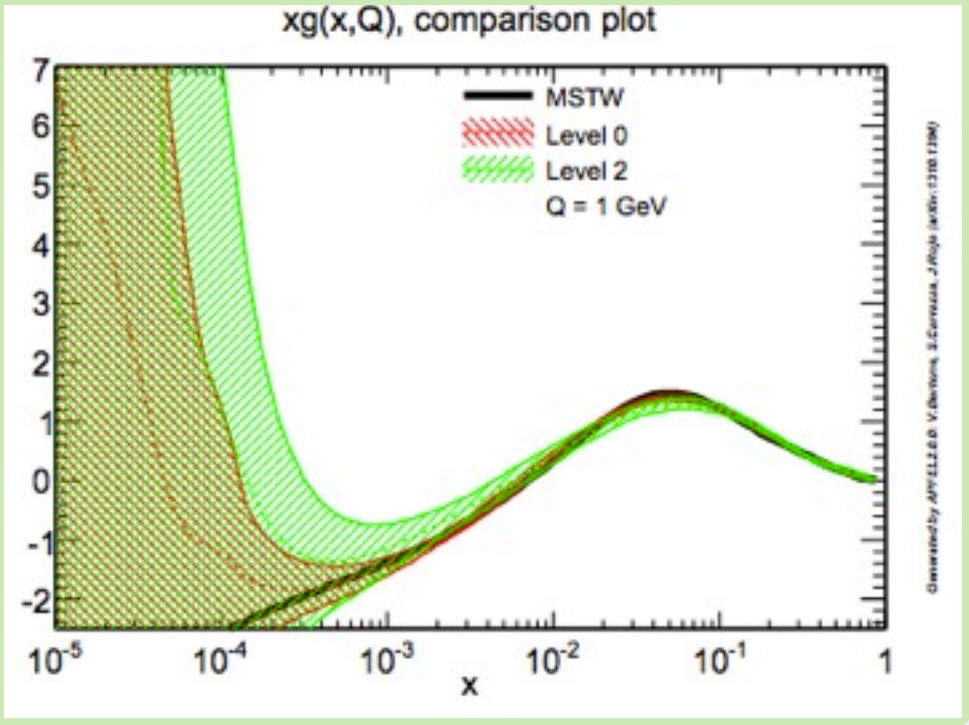
The NNPDF3.0 set

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





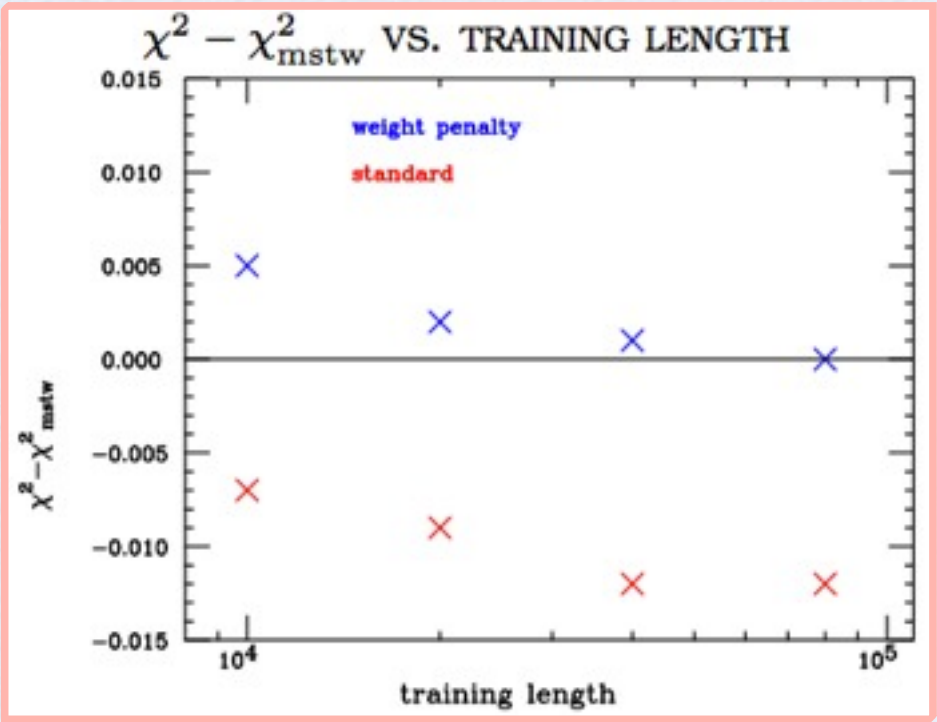
✔ Truth is within 1σ error band!

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 → $\chi^2 = 0.96, \langle E \rangle = 2.0$ (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

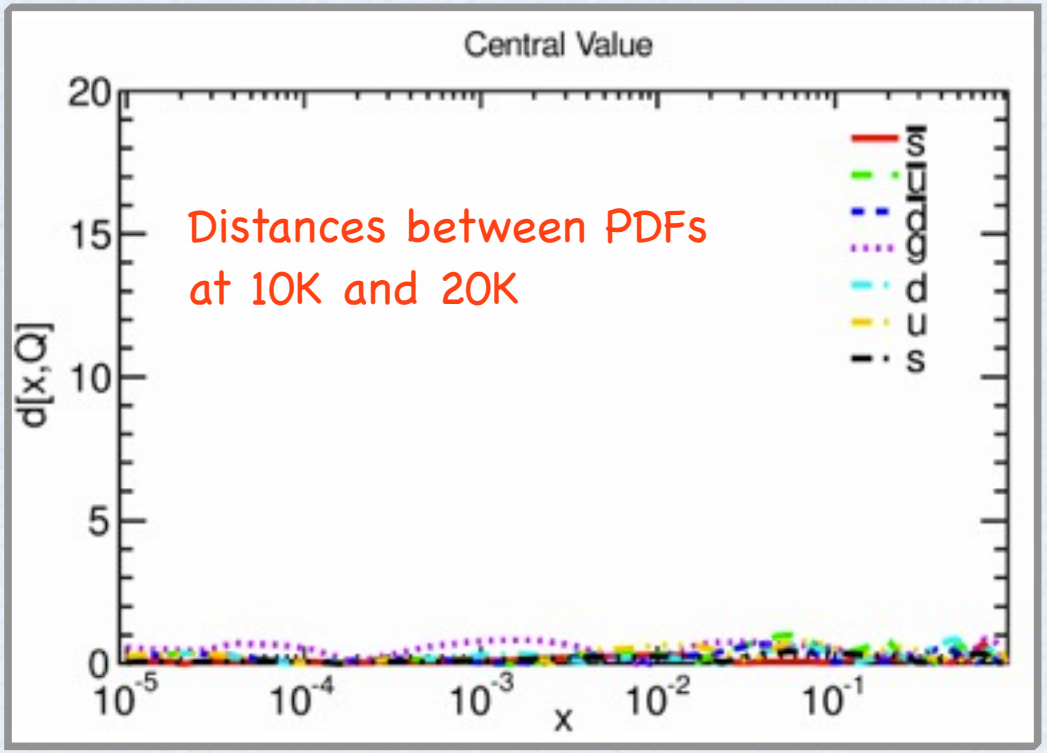
$$\Delta\chi^2 \ll \sigma_{\chi^2}$$

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 → $\chi^2 = 0.96, \langle E \rangle = 2.0$ (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

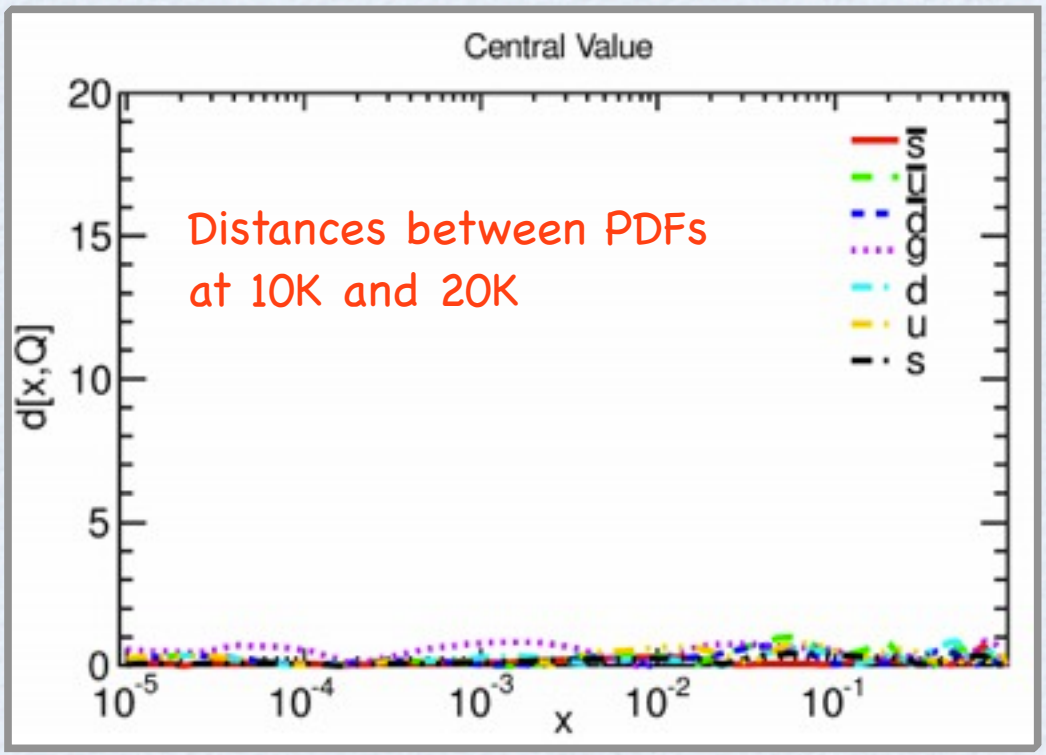
$$\Delta\chi^2 \ll \sigma_{\chi^2}$$

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

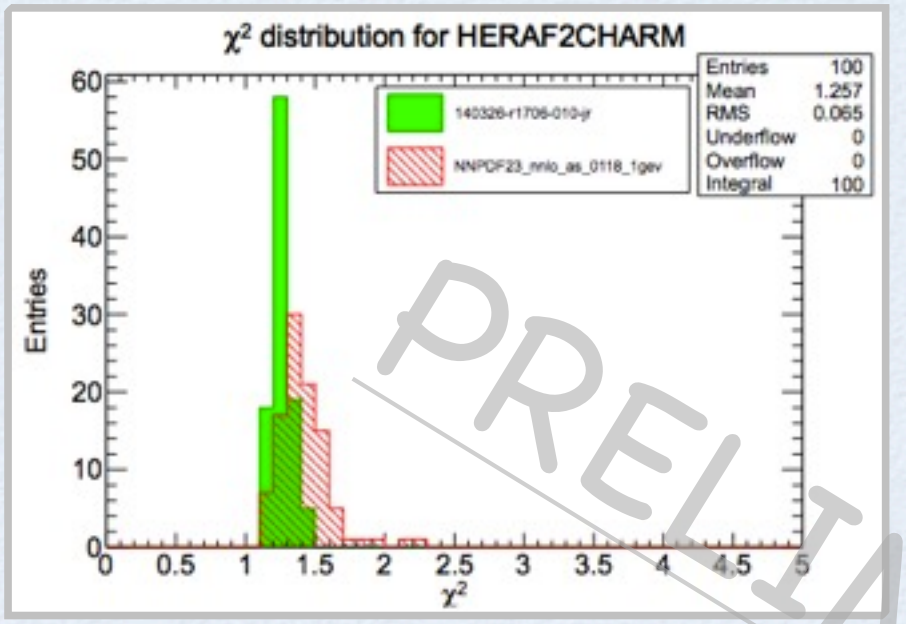


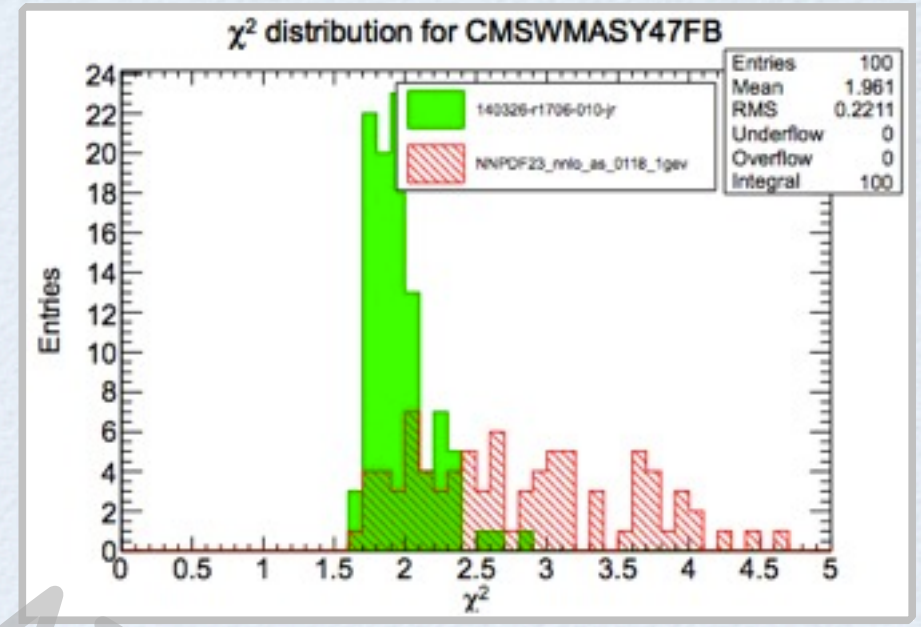
Preliminary conclusions

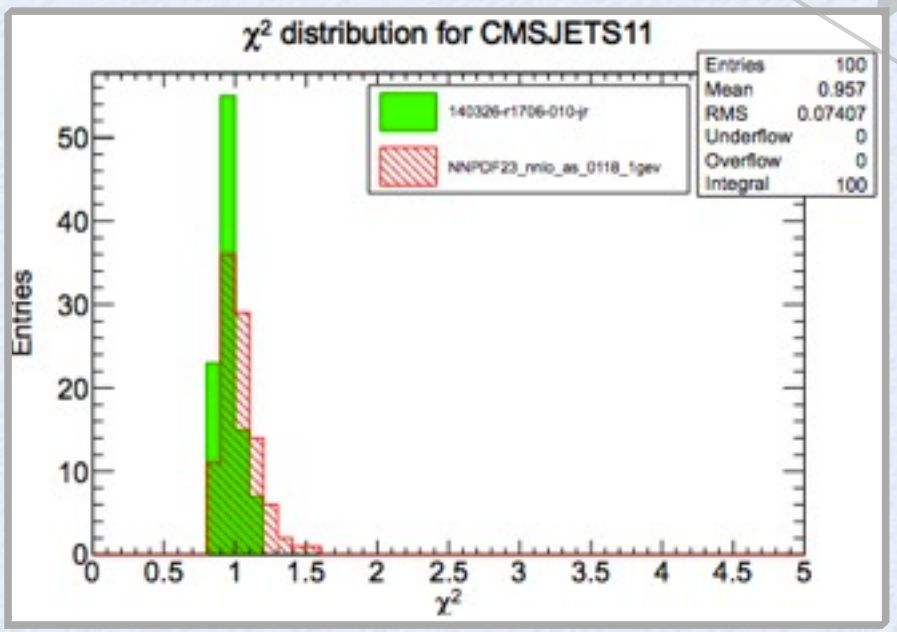
- Fixed-Length fit fully adequate
- No overlearning in global fit due to large number of data
- Over-learning observed in fits to reduced datasets
- Effect of Weigh-Penalty moderate

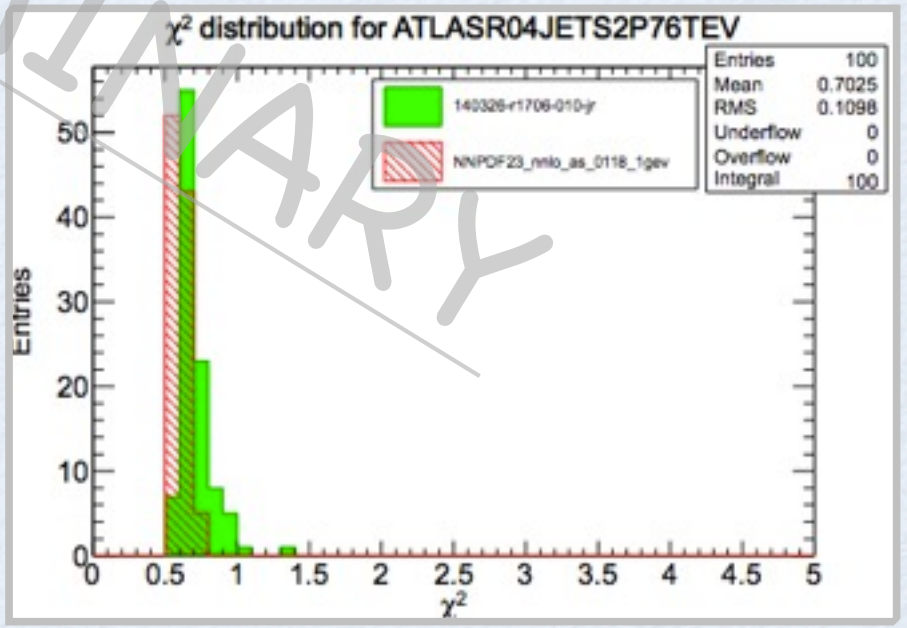
Results

Description of the new data



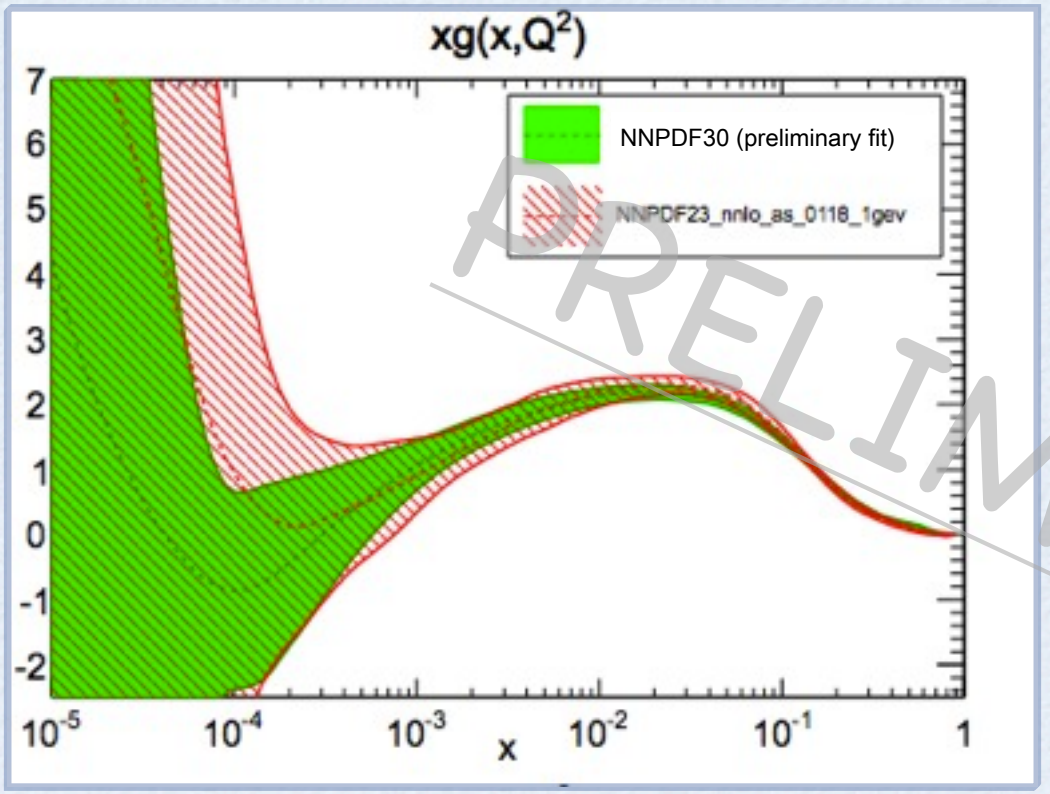


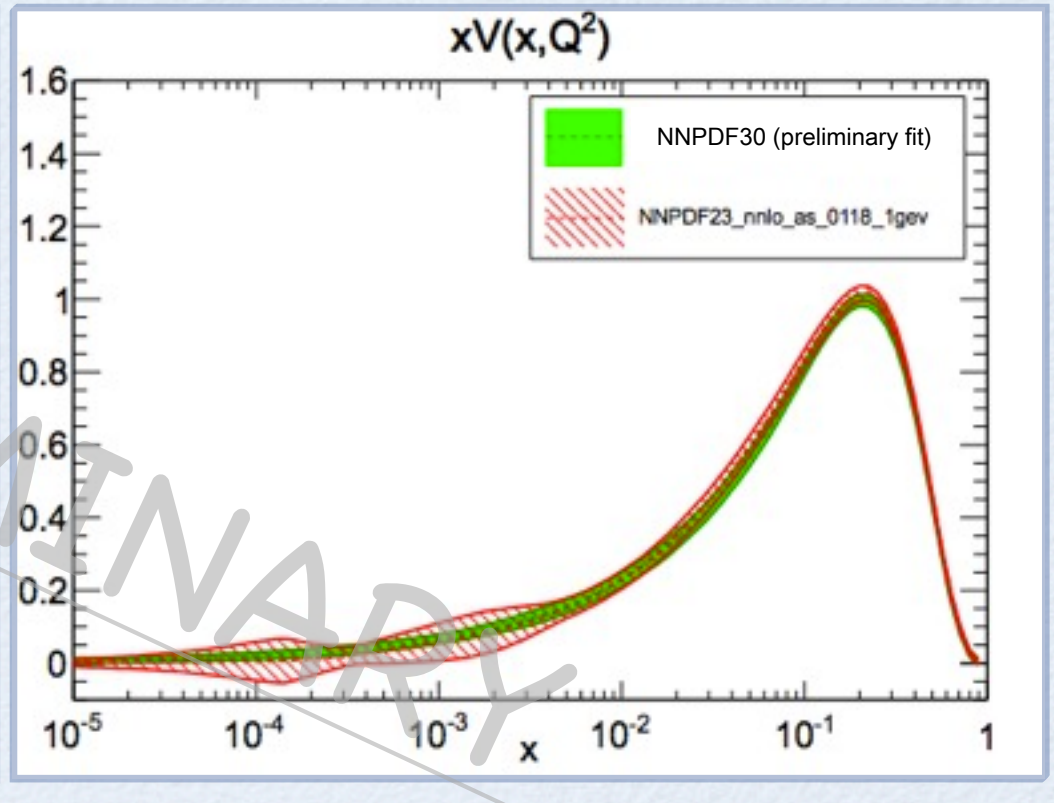




Results

The NNPDF3.0 partons





Preliminary fit:

- some data (ATLAS W pT) not yet included
- some theoretical refinement (C-factors for EW corrections and some NNLO C-factors)

Conclusions and Outlook

- NNPDF23 is the first public fit to include the effect of the LHC data.
  NNPDF23QED and NNPDF23 for MC widely used

- NNPDF is the only unpolarized and polarized set available in LHAPDF
  [see Emanuele Nocera's talk in WG6]
- Fit to fragmentation functions within similar framework soon available!

- Upcoming NNPDF30 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, larger impact expected than NNPDF23
  - Improved positivity (SUSY observables and large x gluons and quarks)

- Release in LHAPDF expected by summer 2014

  - NNPDF30 will be available at LO, NLO, NNLO
  - Soon after NNPDF30QED and NNPDF30IC with intrinsic charm
  - In the near future NNPDF30 including N3LO approx and resummations

  based on Ball, Bonvini, Forte, Marzani, Ridolfi et al, NP B874 (2013)

Conclusions and Outlook

- NNPDF23 is the first public fit to include the effect of the LHC data. NNPDF23QED and NNPDF23 for MC widely used

- NNPDF is the only unpolarized and polarized set available in LHAPDF
  [see Emanuele Nocera's talk in WG6]
- Fit to fragmentation functions within similar framework soon available!

- Upcoming NNPDF30 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, larger impact expected than NNPDF23
  - Improved positivity (SUSY observables and large x gluons and quarks)

THANKS for your attention and STAY TUNED!

- Release in LHAPDF expected by summer 2014

  - NNPDF30 will be available at LO, NLO, NNLO
  - Soon after NNPDF30QED and NNPDF30IC with intrinsic charm
  - In the near future NNPDF30 including N3LO approx and resummations

  based on Ball, Bonvini, Forte, Marzani, Ridolfi et al, NP B874 (2013)