Precision PDFs

- Developments and Challenges

XXX International Workshop on Deep-Inelastic Scattering and Related Subjects, Michigan State University, U.S.A.

Thomas Cridge, DESY, 27th March 2023

DESY.

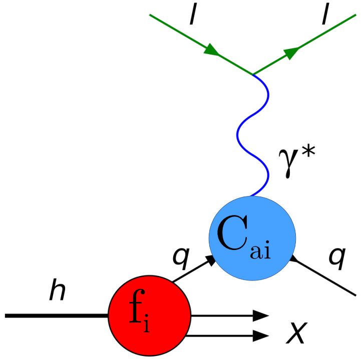
European Research Council

1. Introduction

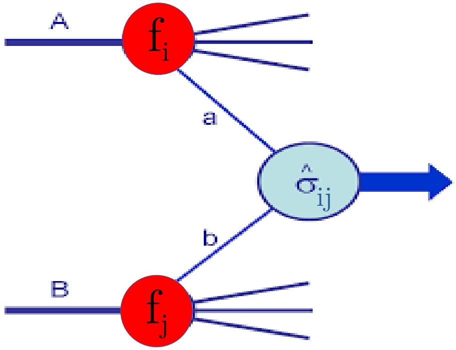
Introduction

- *Parton Distribution Functions (PDFs)* are a crucial input and key output of collider physics.
- Collider physics relies on QCD Collinear factorisation:

DIS:



Hadron-Hadron (pp):



$$F_a(x, Q^2) = \sum_{i=q,\bar{q},g} \int_0^1 \frac{dz}{z} f_i(z, Q^2) C_{a,i}\left(\frac{x}{z}, \alpha_S(Q^2)\right) + \mathcal{O}\left(\frac{\Lambda^2_{QCD}}{Q^2}\right)$$

$$\sigma = \sum_{ij} \int_{x_{min}}^1 dx_1 dx_2\, f_i(x_1, \mu_f^2) f_j(x_2, \mu_f^2) \hat{\sigma}_{ij}(x_1 p_1, x_2 p_2, Q, \mu_F^2)$$

- Separate short distance perturbative physics in coefficient functions and hard cross-sections from non-perturbative long distance PDFs.

$$\frac{\partial f_q^{NS}(x,\mu^2)}{\partial \log \mu^2} = \frac{\alpha_S}{2\pi} \int_x^1 \frac{dz}{z} f_q^{NS}(z,\mu^2) P_{qq}^{NS}(x/z)$$

- PDFs are **universal** and evolve between scales by DGLAP equations.

Global PDF fitting...

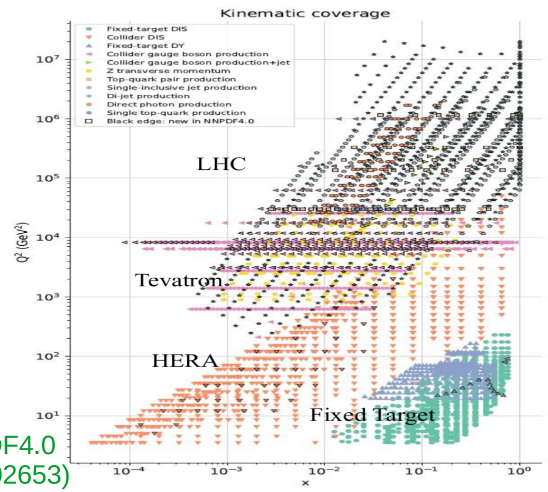
1) Experiment

- Latest experimental data
- Fixed target, collider
DIS, Tevatron, LHC, etc
- EW boson, jets, top, ...
- Large range in x, Q^2

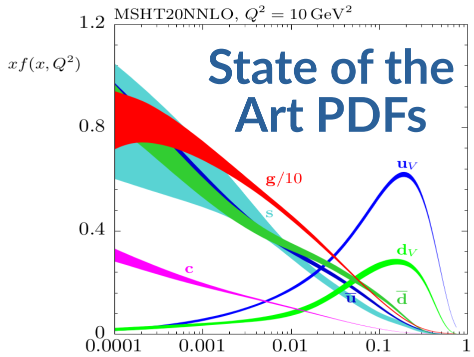
2) Methodology

- Parameterise at low scale
- DGLAP, flavour schemes, ...
- Minimisation of χ^2
- Uncertainty prescription

Global PDF fit



NNPDF4.0
(2109.02653)



3) Theory

- Most precise theoretical calculations available – usually grids + k-factors
- NNLO QCD + NLO EW standard
- Efforts to extend to approximate N3LO + theory uncertainties

Motivation

①

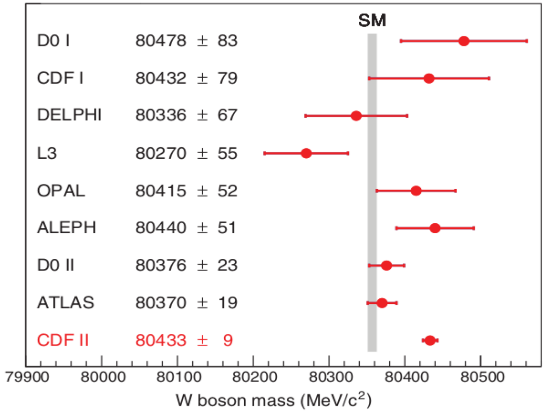
- Key input to almost all calculations/measurements at colliders → Need both **accuracy and precision.** Moreover, often a dominant contribution to *uncertainty*.

1) Precision Standard Model (SM) Measurements –

(a) Electroweak Precision:

- W boson mass (M_W):

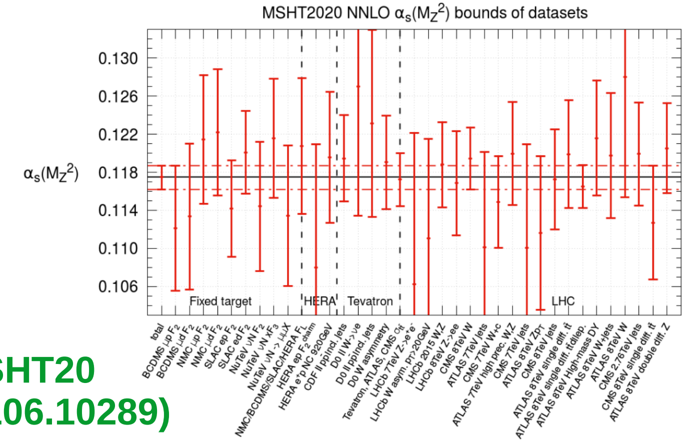
CDF (2022)



PDF-Set	p_T^ℓ [MeV]	m_T [MeV]	combined [MeV]
CT10	$80355.6^{+15.8}_{-15.7}$	$80378.1^{+24.4}_{-24.8}$	$80355.8^{+15.7}_{-15.7}$
CT14	$80358.0^{+16.3}_{-16.3}$	$80388.8^{+25.2}_{-25.5}$	$80358.4^{+16.3}_{-16.3}$
CT18	$80360.1^{+16.3}_{-16.3}$	$80382.2^{+25.3}_{-25.3}$	$80360.4^{+16.3}_{-16.3}$
MMHT2014	$80360.3^{+15.9}_{-15.9}$	$80386.2^{+23.9}_{-24.4}$	$80361.0^{+15.9}_{-15.9}$
MSHT20	$80358.9^{+13.0}_{-16.3}$	$80379.4^{+24.6}_{-25.1}$	$80356.3^{+14.6}_{-14.6}$
NNPDF3.1	$80344.7^{+15.6}_{-15.5}$	$80354.3^{+23.6}_{-23.7}$	$80345.0^{+15.5}_{-15.5}$
NNPDF4.0	$80342.2^{+15.3}_{-15.3}$	$80354.3^{+22.3}_{-22.4}$	$80342.9^{+15.3}_{-15.3}$

ATLAS (CONF-2023-004)

(b) Strong coupling ($\alpha_S(M_Z^2)$):



MSHT20 (2106.10289)

→ **Input to PDG determination (2021)**

See also Snowmass review (2203.08271)

Motivation

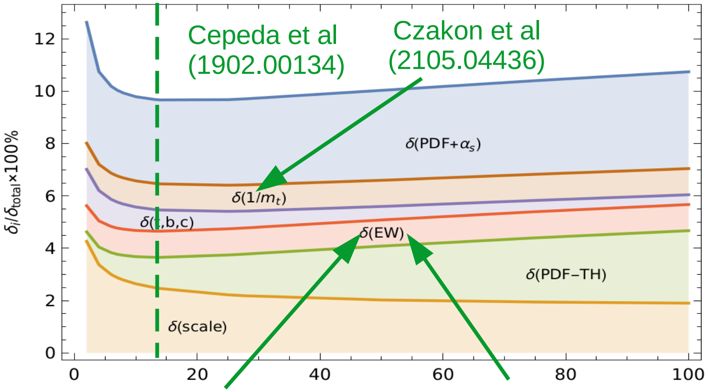
1

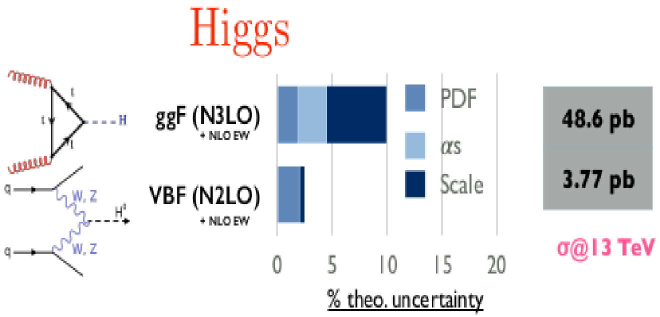
- Key input to almost all calculations/measurements at colliders → Need both **accuracy and precision.** Moreover, often a dominant contribution to *uncertainty.*

2) Higgs Measurements:

- PDF and related uncertainties (α_s, PDF-TH from NNLO – N3LO mismatch) dominant in ggF Higgs production. Also large in other production mechanisms.

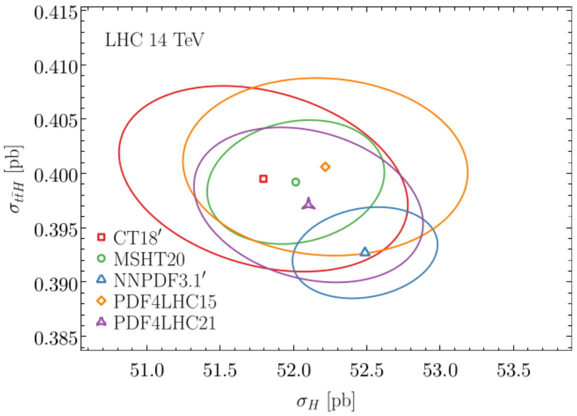
LHC Higgs XSWG 2019





See talk by F. Zhang this morning!

PDF4LHC21 (2203.05506)



Motivation

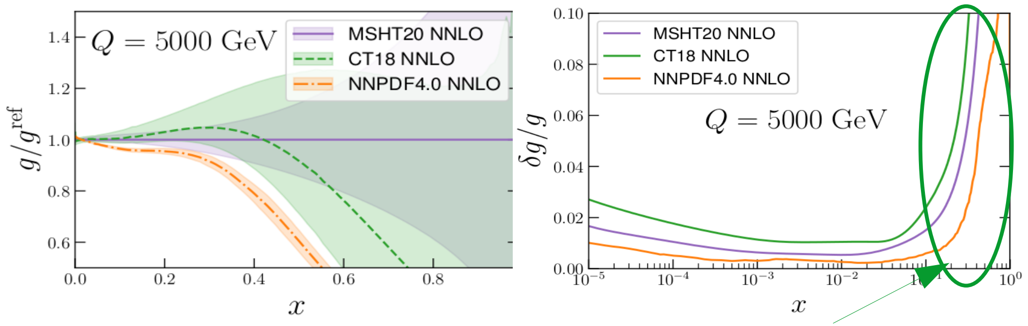
- Key input to almost all calculations/measurements at colliders → Need both **accuracy and precision.** Moreover, often a dominant contribution to *uncertainty.*

3) Beyond Standard Model (BSM) Searches:

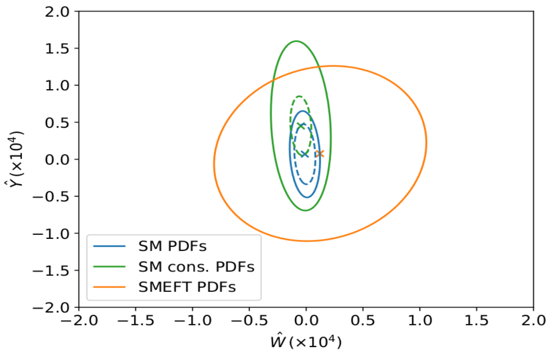
See talk by T.M.P. Tait this morning!

- Either look in high-energy tails of distributions → requires large x PDFs.

- Or look for small deviations from SM → requires precision PDFs.



Gluon, e.g. for dijet searches, at high x central values differ and uncertainty blows up → Lack of data constraint



PDF + SMEFT combined fit – Ubiali et al (2104.02723)

Motivation

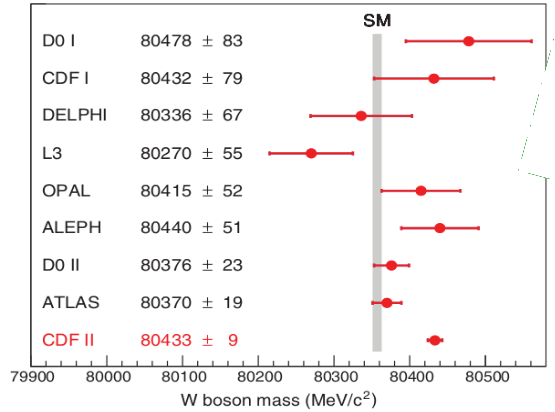
1

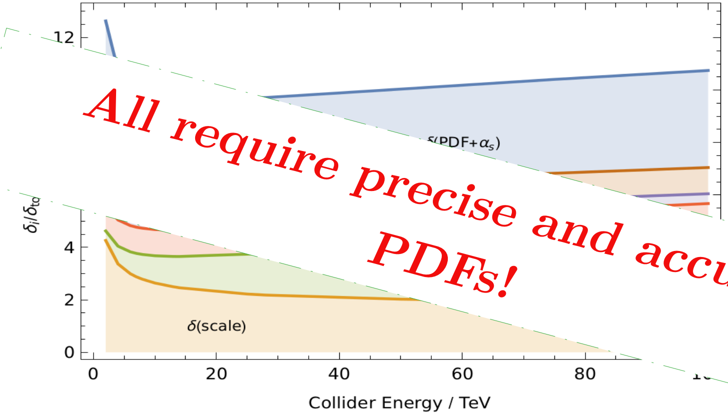
- Key input to almost all calculations/measurements at colliders → Need both **accuracy and precision.** Moreover, often a dominant contribution to *uncertainty.*

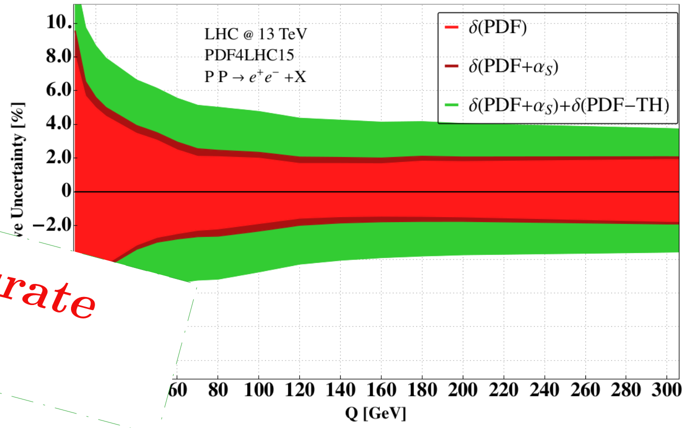
  1) Precision Standard Model (SM) Measurements: M_W, $\sin^2\Theta_W$, $\alpha_S(M_Z^2)$, etc.

  2) Higgs Measurements

  3) Beyond Standard Model (BSM) Searches: High energy, SMEFT, etc







All require precise and accurate PDFs!

2. Current PDF Landscape

Several Global PDF Fitting Groups ②

- Several different PDF analysis groups – ABM, ATLASPDF, CJ, CT, HERAPDF, JAM, MSHT, NNPDF and others. Will not be able to cover all here!

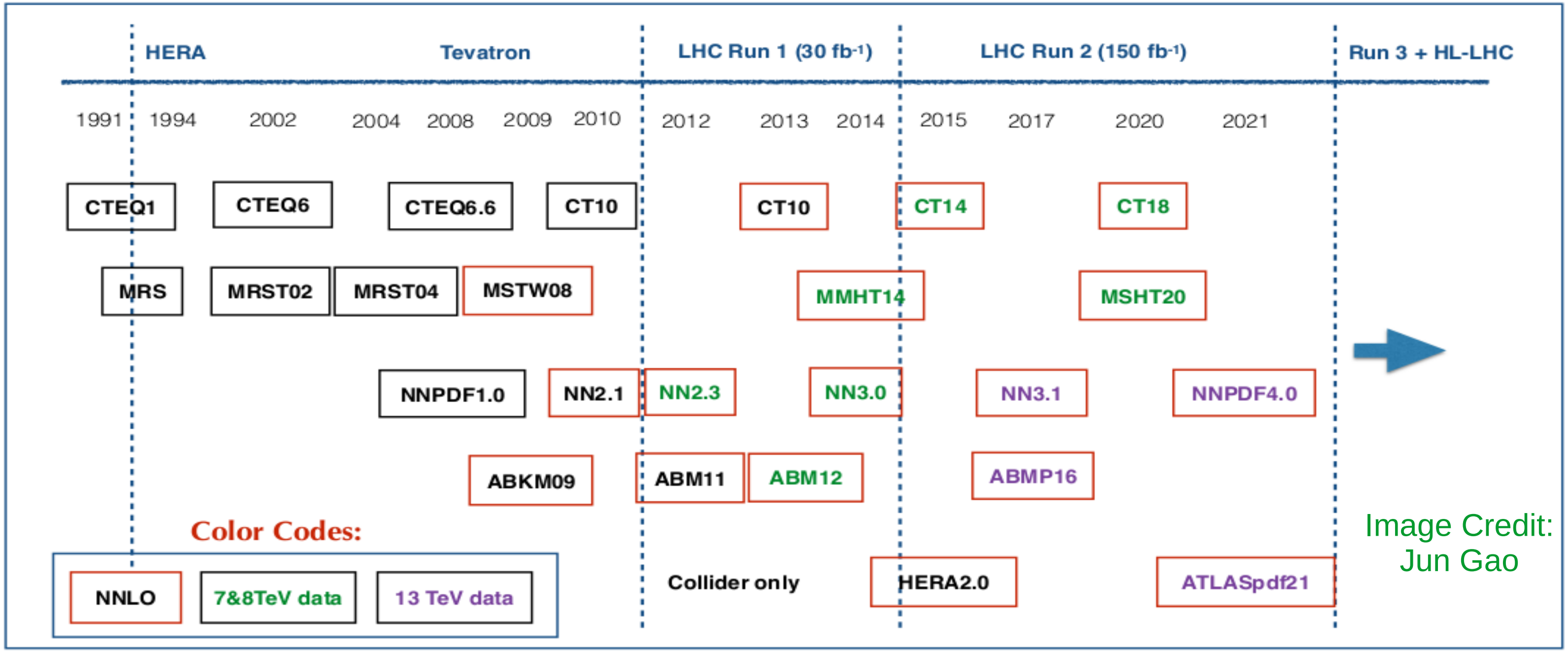


Image Credit: Jun Gao

Default now

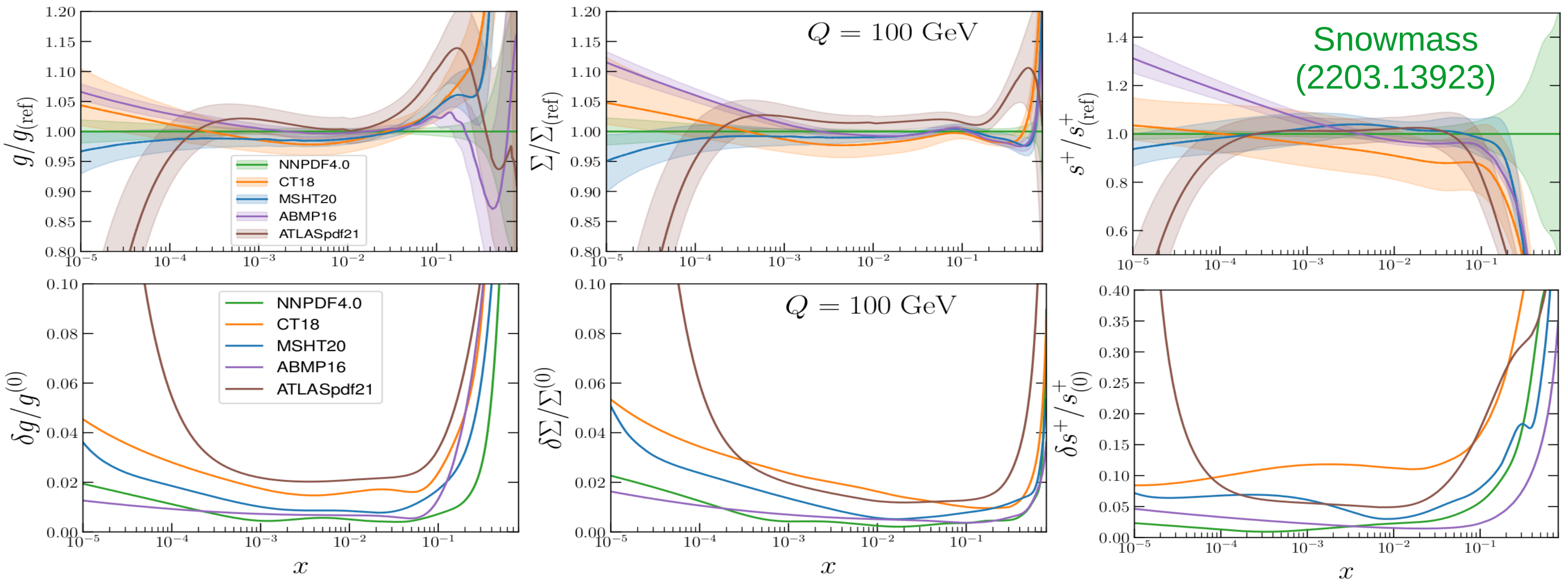
NNLO QCD + NLO EW and latest LHC and other data

+ PDF4LHC21 combination of MSHT20, CT18, NNPDF3.1 (2203.05506)

- Different focuses, methodologies, uncertainty prescriptions → **beneficial!**

PDF Comparison

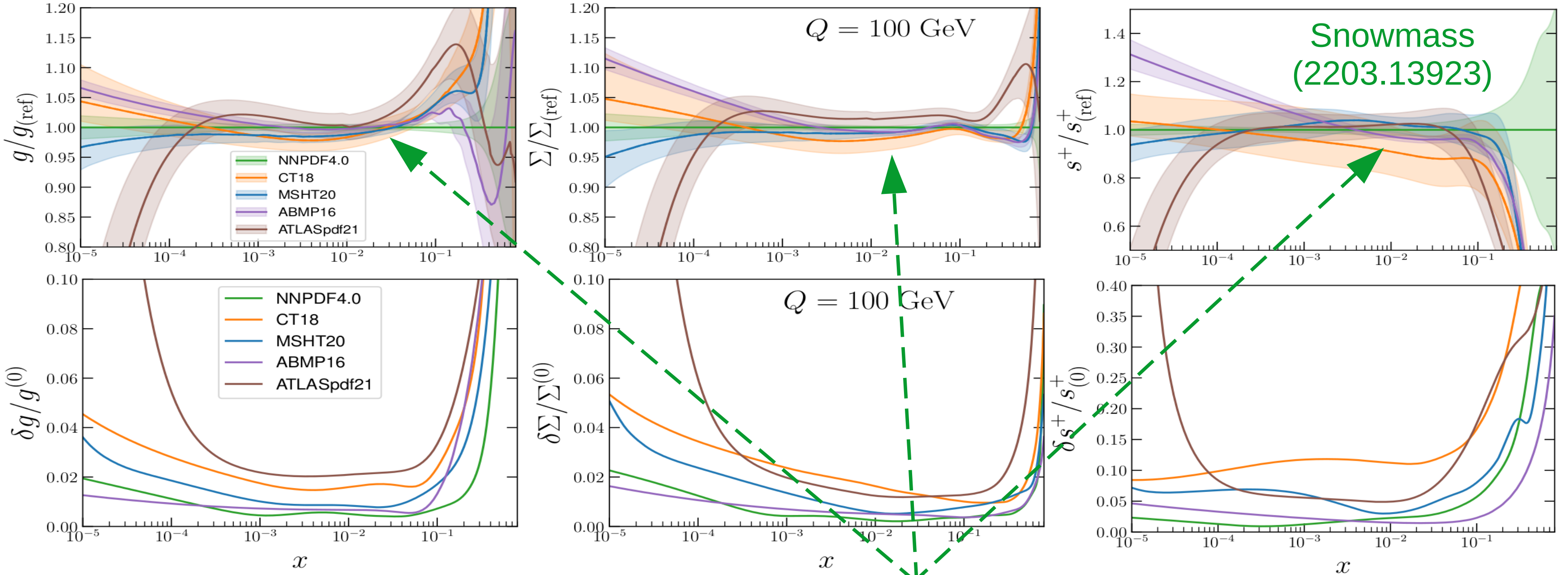
- Compare several of these at the level of the PDFs and uncertainties:



PDF Comparison

2

- Compare several of these at the level of the PDFs and uncertainties:

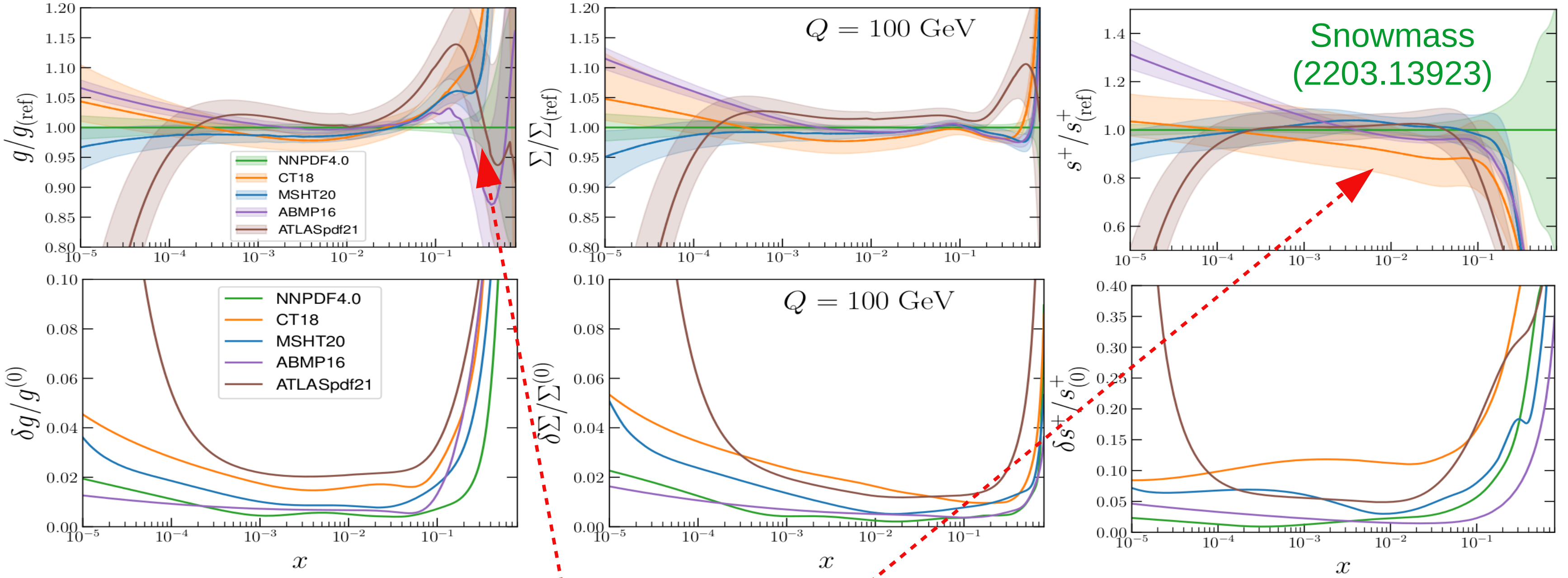


- General agreement over data range ($10^{-4} < x < 10^{-1}$) within uncertainties.

PDF Comparison

2

- Compare several of these at the level of the PDFs and uncertainties:

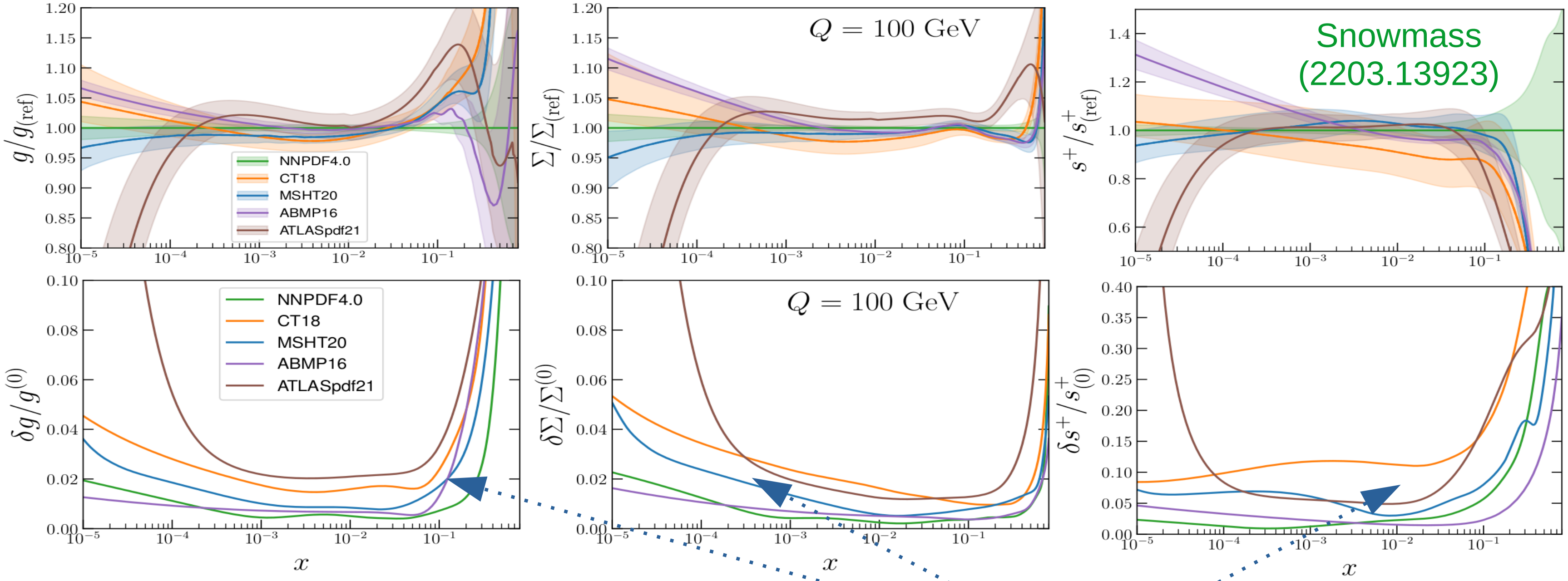


- General agreement over data range ($10^{-4} < x < 10^{-1}$) within uncertainties.
- Differences exist (high x gluon, strangeness, ...).

PDF Comparison

2

- Compare several of these at the level of the PDFs and uncertainties:

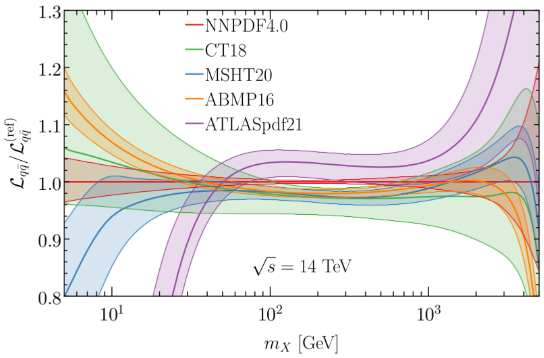


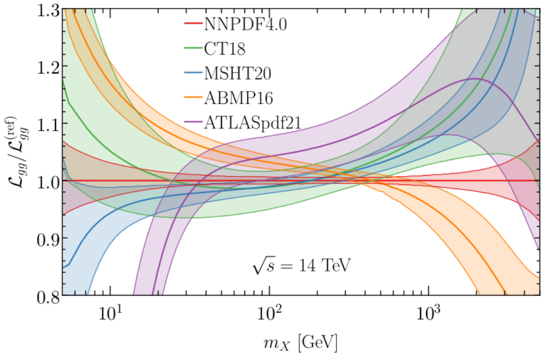
- General agreement over data range ($10^{-4} < x < 10^{-1}$) within uncertainties.
- Differences exist (high x gluon, strangeness, uncertainty sizes).

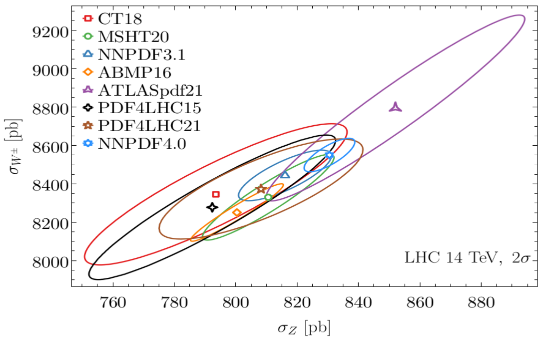
Parton luminosities and Xsecs

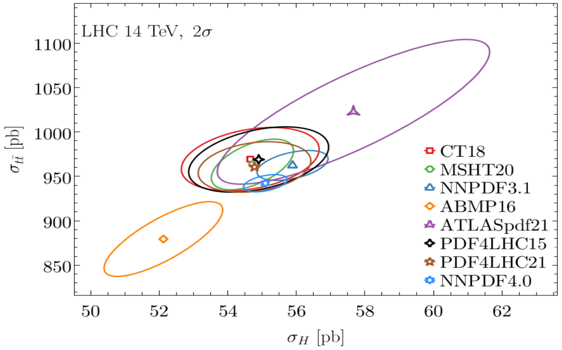
2

- Compare several of these at luminosity and cross-section level:







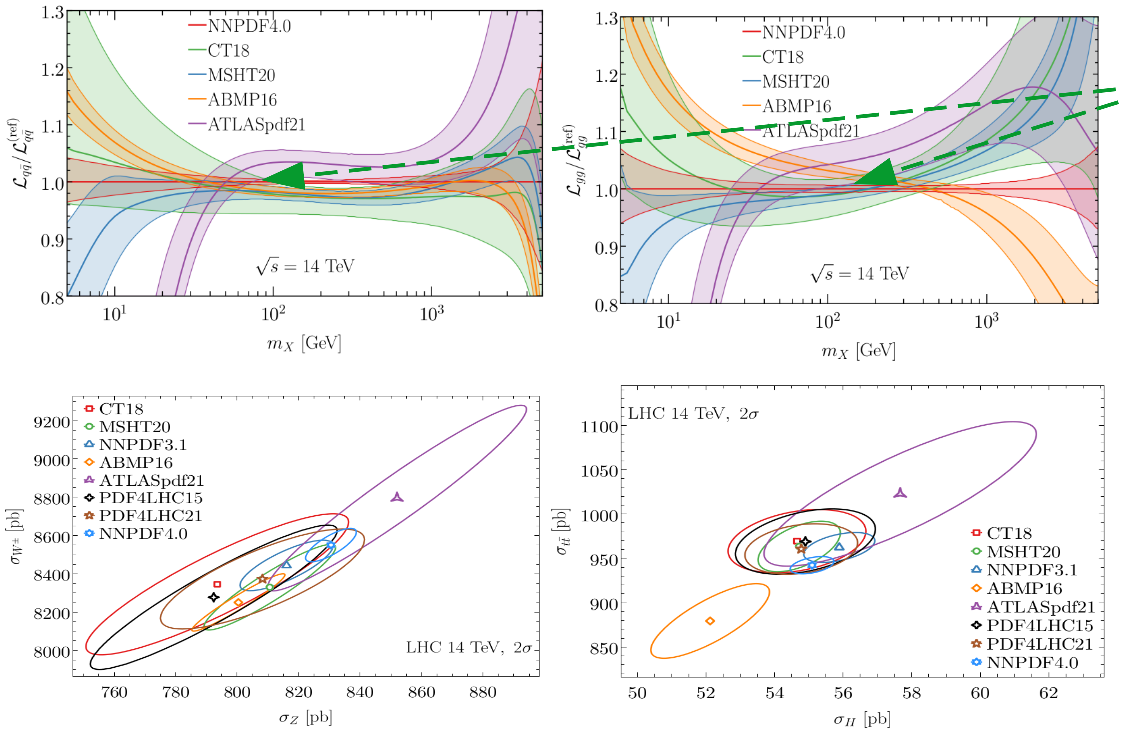


Snowmass (2203.13923)

Parton luminosities and Xsecs

2

- Compare several of these at luminosity and cross-section level:

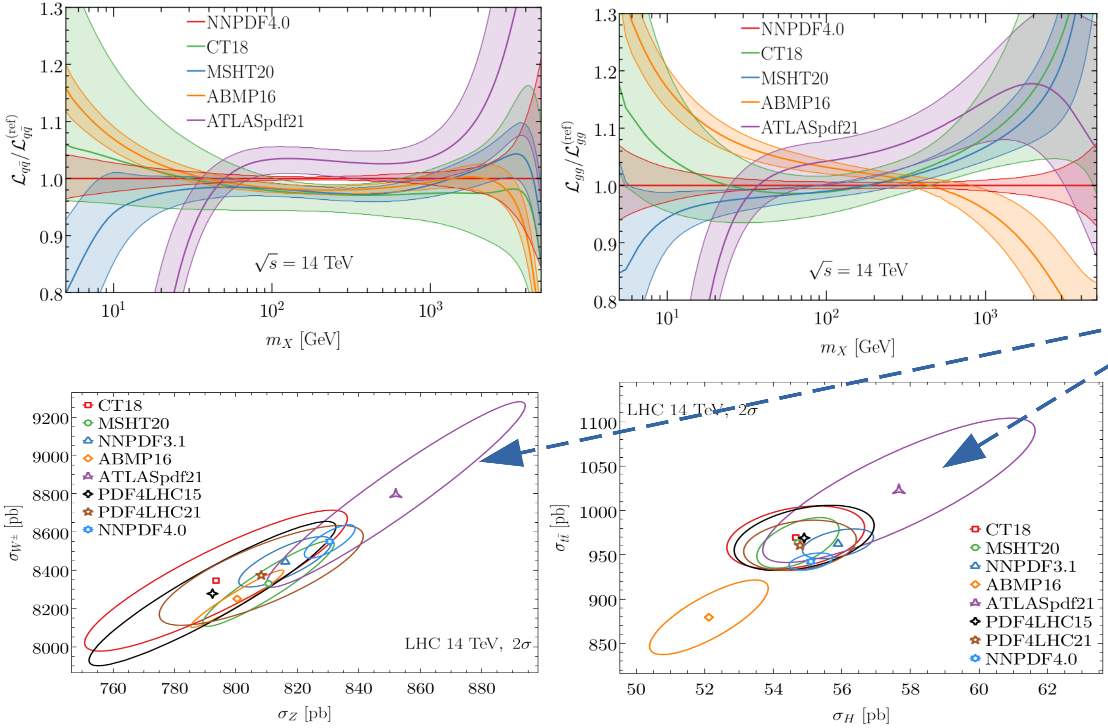


Snowmass (2203.13923)

- General agreement over intermediate invariant masses (10 GeV < M_x < 10^3 GeV).
- Xsecs show 2σ error ellipses, correlations in cross-sections visible.

Parton luminosities and Xsecs

- Compare several of these at luminosity and cross-section level:



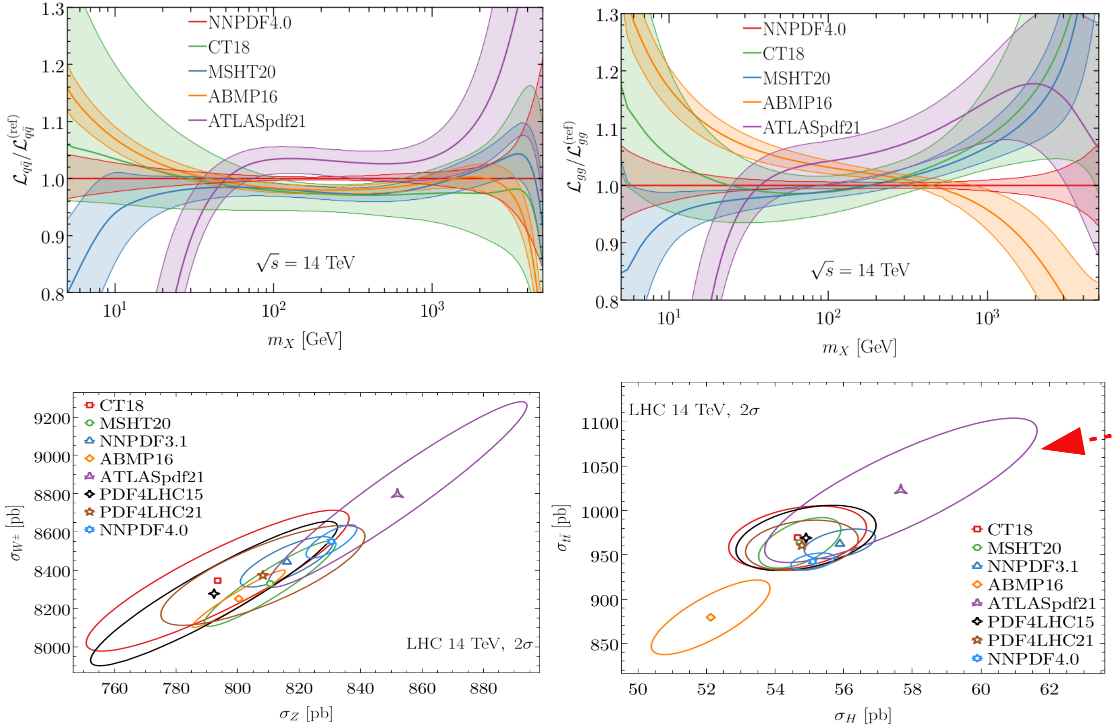
Snowmass (2203.13923)

- General agreement over intermediate invariant masses (10 GeV < M_x < 10^3 GeV).
- Xsecs show 2σ error ellipses, correlations in cross-sections visible.

Parton luminosities and Xsecs

②

- Compare several of these at luminosity and cross-section level:



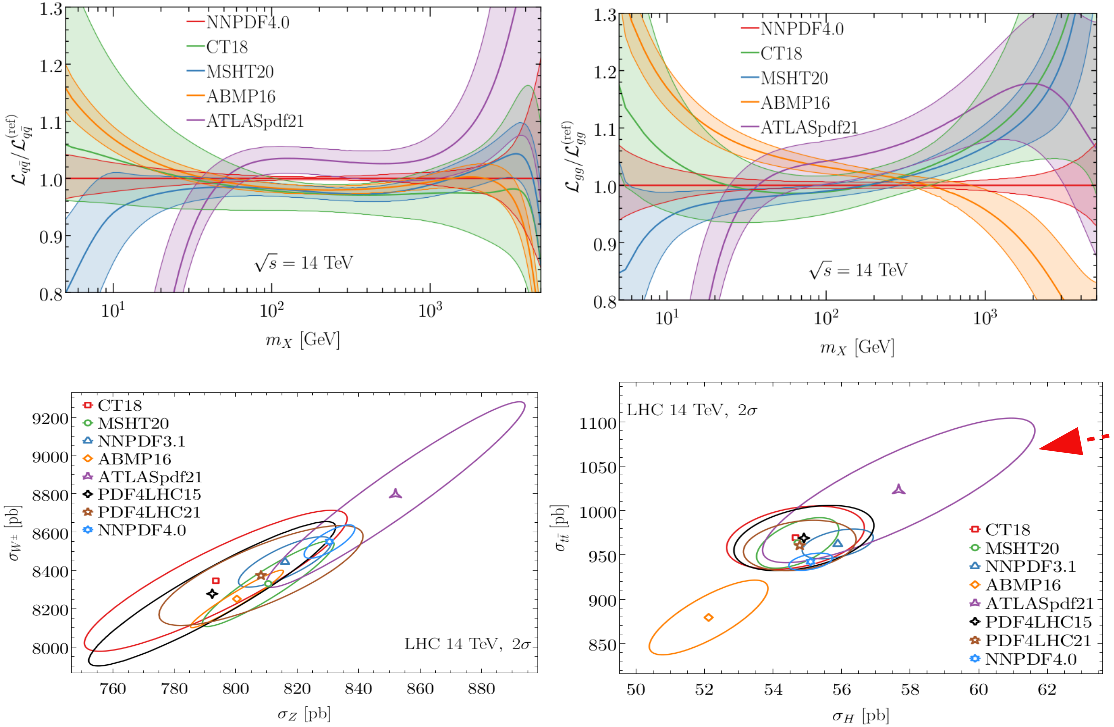
Snowmass (2203.13923)

- General agreement over intermediate invariant masses (10 GeV < M_x < 10^3 GeV).
- Xsecs show 2σ error ellipses, correlations in cross-sections visible.
- Differences exist in size of uncertainties, largely reflect experimental and methodological differences.

Parton luminosities and Xsecs

2

- Compare several of these at luminosity and cross-section level:



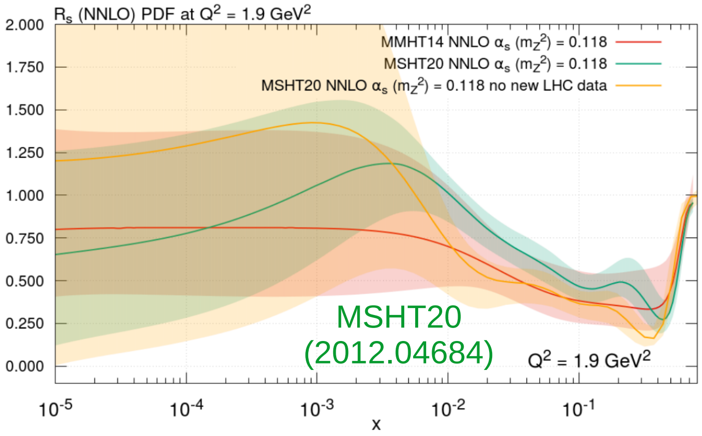
Snowmass (2203.13923)

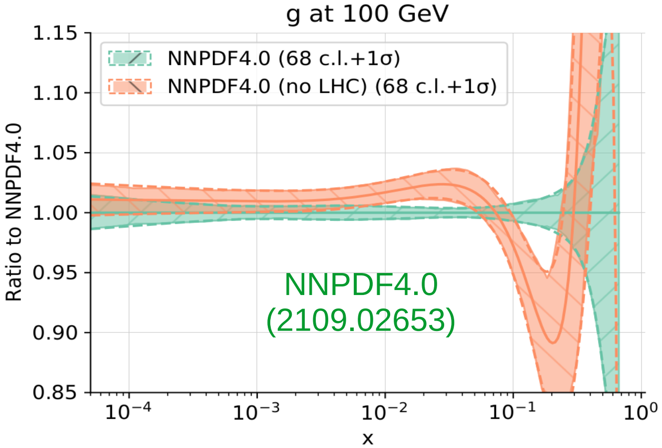
- General agreement over intermediate invariant masses (10 GeV < M_x < 10^3 GeV).
- Xsecs show **2**σ error ellipses, correlations in cross-sections visible.
- Differences exist in size of uncertainties, largely reflect experimental and methodological differences.
- Nonetheless we have the *most precise and accurate PDFs* yet.

Experimental Advances

Confronting Precise Data

- High precision, multi-differential data in more channels from LHC and elsewhere.
- Has improved our knowledge of PDFs in both **accuracy and precision**.
- Clear preference now for NNLO theory from precise LHC data.
- In order to exploit this data, more detailed analysis of experimental, methodological, and theoretical issues is required.



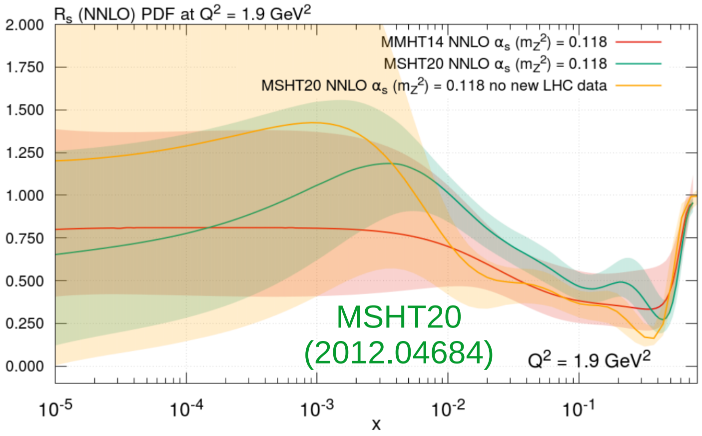


Experimental Advances

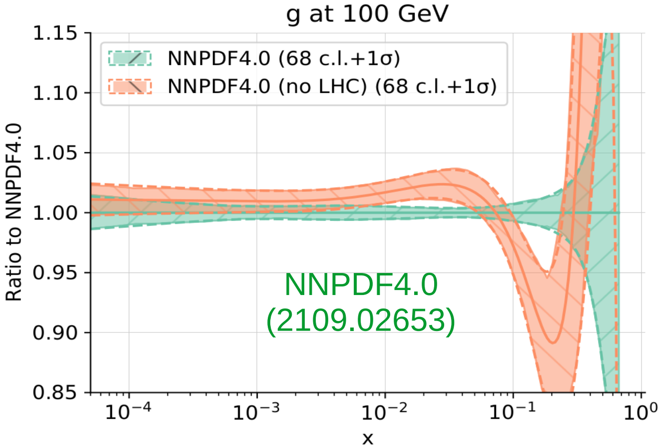
4

Confronting Precise Data

- High precision, multi-differential data in more channels from LHC and elsewhere.
- Has improved our knowledge of PDFs in both **accuracy and precision**.
- Clear preference now for NNLO theory from precise LHC data.
- In order to exploit this data, more detailed analysis of experimental, methodological, and theoretical issues is required.



MSHT20 (2012.04684)



NNPDF4.0 (2109.02653)

Data set	N_{pts}	NLO χ^2/N_{pts}	NNLO χ^2/N_{pts}
ATLAS 8 TeV s. diff $t\bar{t}$	25	1.56	0.98
CMS 8 TeV d. diff $t\bar{t}$	15	2.19	1.50
ATLAS 7 TeV W, Z	61	5.00	1.91
ATLAS 8 TeV W	22	3.85	2.61
ATLAS 8 TeV d. diff Z	59	2.67	1.45
ATLAS 8 TeV Z p_T	104	2.26	1.81
ATLAS 8 TeV W + jets	39	1.13	0.60
Total LHC data	1328	1.79	1.33
Total non-LHC data	3035	1.13	1.10
Total	4363	1.33	1.17

MSHT20 (2012.04684)

3-6. Challenges and Developments

3. Challenges and Developments

4. Experimental Challenges

- Precision Data
- Data Tensions
- Correlated Systematics

Interconnected

Interconnected

5. Methodological Challenges

- New algorithms/methods
- PDF Tools
- Additional constraints (positivity?)
- Theory grids at NNLO and beyond

Interconnected

6. Theoretical Challenges

- Inclusion of N3LO information
- Theoretical Uncertainties
- Deuteron/Nuclear corrections
- Small x resummed PDFs
- Lattice constraints

Will only cover a selection here!

4. Experimental Challenges

Experimental Challenges

4

Confronting Precise Data *Can reflect experimental, methodological or theoretical issues!*

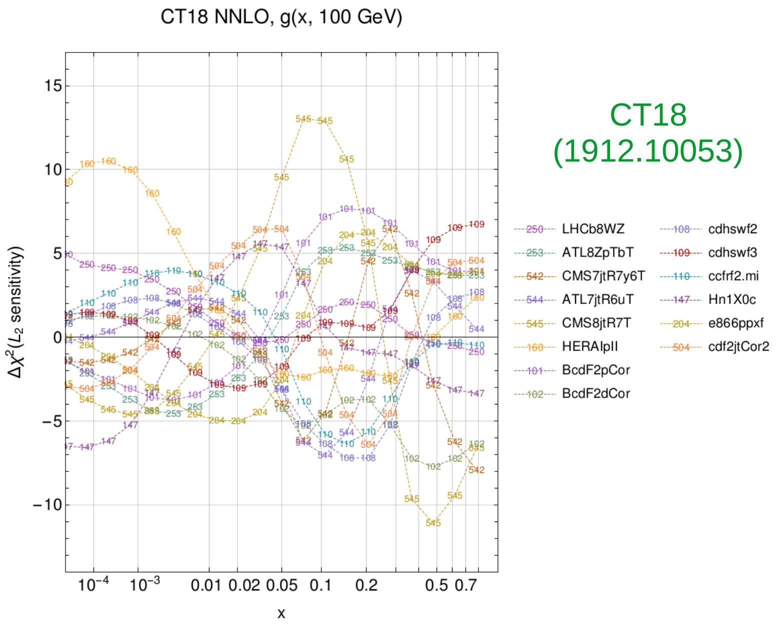
- Issues can arise in fitting some datasets - **poor fit qualities** χ^2/N_{pts}.
- Two frequent (experimental/methodological) causes:

1) Dataset tensions

– Different datasets have conflicting pulls on the PDFs. Examples include

→ Antiquark isospin asymmetry

→ High x gluon (jets, Zp_T, top)



CT18 (1912.10053)

Experimental Challenges

4

Confronting Precise Data *Can reflect experimental, methodological or theoretical issues!*

- Issues can arise in fitting some datasets - **poor fit qualities** χ^2/N_{pts}.

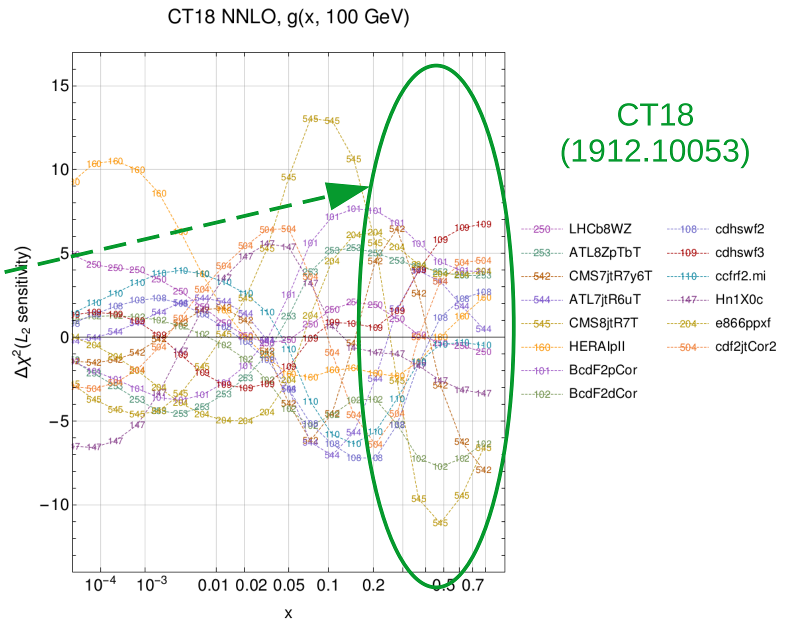
- Two frequent (experimental/methodological) causes:

1) Dataset tensions

– Different datasets have conflicting pulls on the PDFs. Examples include

→ Antiquark isospin asymmetry

→ High x gluon (jets, Zp_T, top)



CT18 (1912.10053)

Experimental Challenges

4

Confronting Precise Data Can reflect experimental, methodological or theoretical issues!

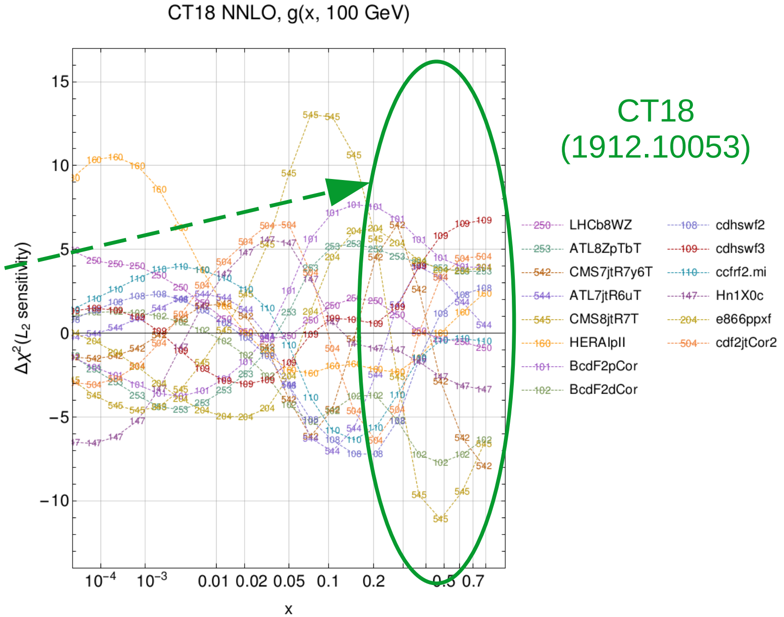
- Issues can arise in fitting some datasets - **poor fit qualities** χ^2/N_{pts}.
- Two frequent (experimental/methodological) causes:

1) Dataset tensions
– Different datasets have conflicting pulls on the PDFs. Examples include

→ Antiquark isospin asymmetry

→ High x gluon (jets, Zp_T, top)

2) Issues with systematic correlations – Often systematic errors now dominate, their less well-known correlations can notably affect fit quality.



Experimental Challenges

4

Confronting Precise Data

Can reflect experimental, methodological or theoretical issues!

- Issues can arise in fitting some datasets - **poor fit qualities** χ^2/N_{pts}.
- Two frequent (experimental/methodological) causes:

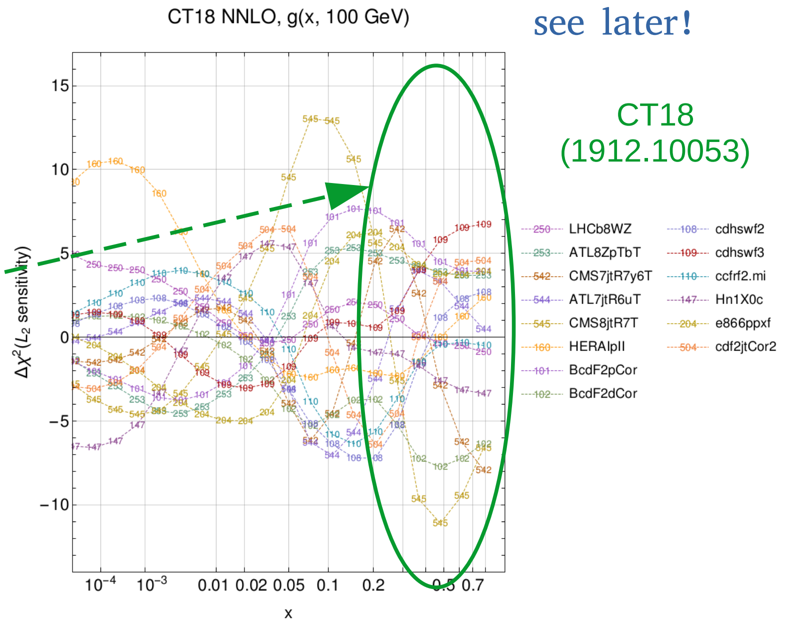
N.B. Also some evidence of effects from missing higher orders (MHOs) – see later!

1) Dataset tensions

– Different datasets have conflicting pulls on the PDFs. Examples include

→ Antiquark isospin asymmetry

→ High x gluon (jets, Zp_T, top)



CT18 (1912.10053)

2) Issues with systematic correlations – Often systematic errors now dominate, their less well-known correlations can notably affect fit quality.

Experimental Challenges

4

Dataset Tensions – *Can reflect experimental, methodological or theoretical issues!*

- High x $\bar{d}/\bar{u}$: - Theoretical models (e.g. pion cloud) generally favour $\bar{d}>\bar{u}$ at high x.

  - Gottfried sum rule – NMC found $\int_0^1 [F_2^p(x) - F_2^n(x)]dx/x < \frac{1}{3}$

Theoretical Review in Peng et al (1402.1236)

NMC (Phys. Rev. Lett. 66, 2712 (1991))

Experimental Challenges

4

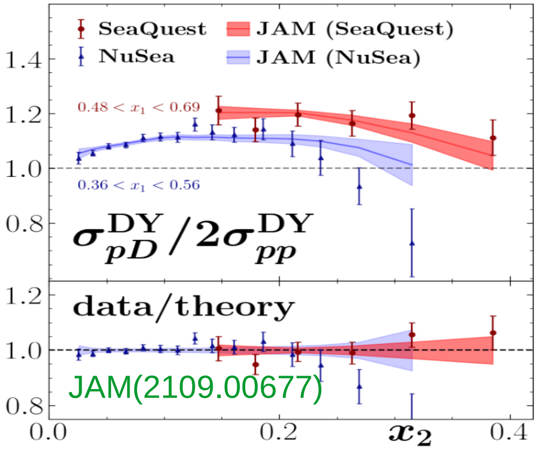
Dataset Tensions – *Can reflect experimental, methodological or theoretical issues!*

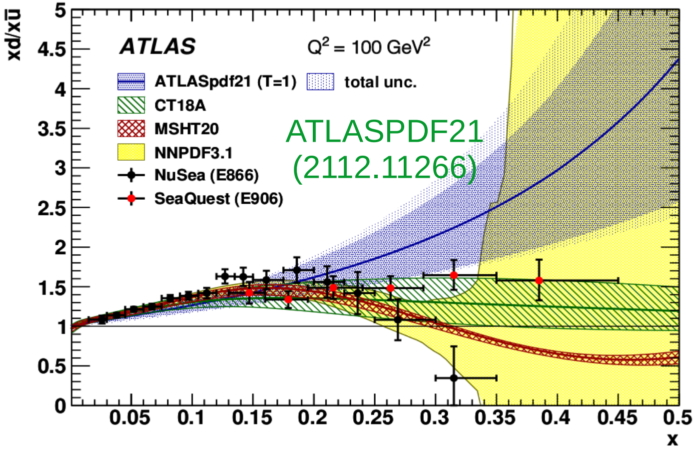
- High x $\bar{d}/\bar{u}$: - Theoretical models (e.g. pion cloud) generally favour $\bar{d}>\bar{u}$ at high x.

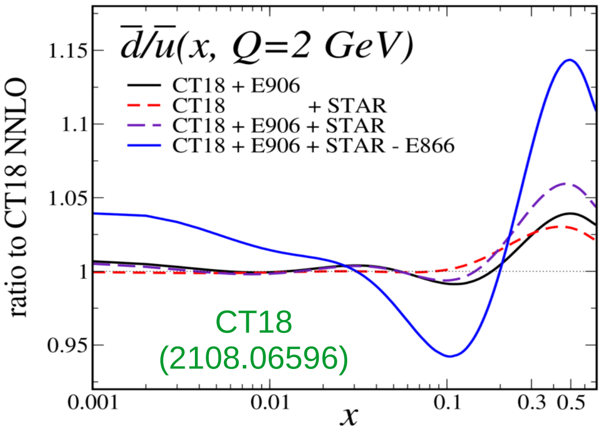
NuSea (hep-ex/0103030)
Seaquest (2103.04024)

- Gottfried sum rule – NMC found $\int_0^1 [F_2^p(x) - F_2^n(x)]dx/x < \frac{1}{3}$
- E866/NuSea data favoured $\bar{d}<\bar{u}$ at high x.
- New Seaquest/E906 data instead favour $\bar{d}>\bar{u}$.

See talk by Ching Him Leung







CJ22 - See talk by A. Accardi

Experimental Challenges

④

Dataset Tensions – *Can reflect experimental, methodological or theoretical issues!*

- High x $\bar{d}/\bar{u}$: - Theoretical models (e.g. pion cloud) generally favour $\bar{d}>\bar{u}$ at high x.

NuSea (hep-ex/0103030)
Seaquest (2103.04024)

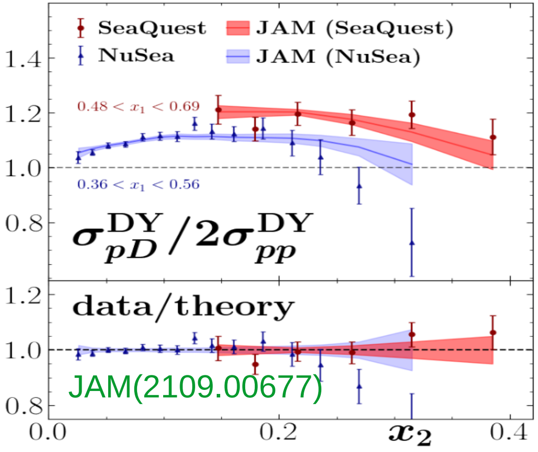
- Gottfried sum rule – NMC found $\int_0^1 [F_2^p(x) - F_2^n(x)]dx/x < \frac{1}{3}$

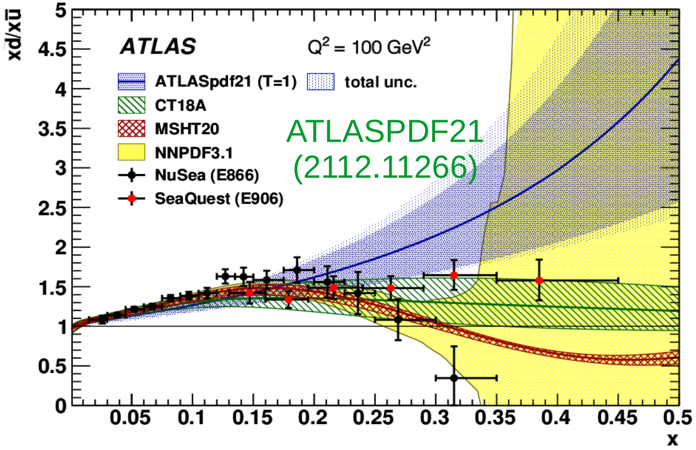
- E866/NuSea data favoured $\bar{d}<\bar{u}$ at high x.

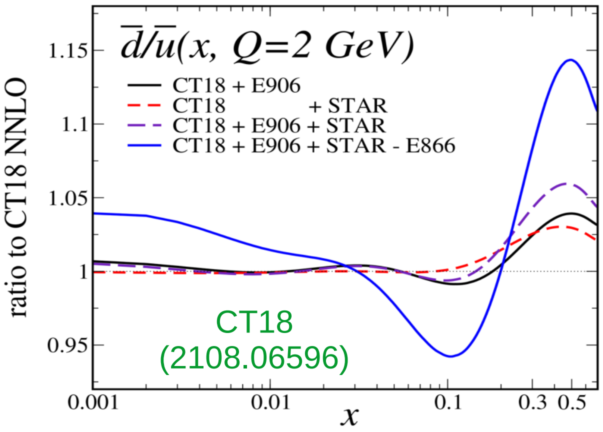
- New Seaquest/E905 data instead favour $\bar{d}>\bar{u}$.

See talk by Ching Him Leung

How should this be interpreted in a global PDF fit?







CJ22 - See talk by A. Accardi

Experimental Challenges

4

Confronting Precise Data

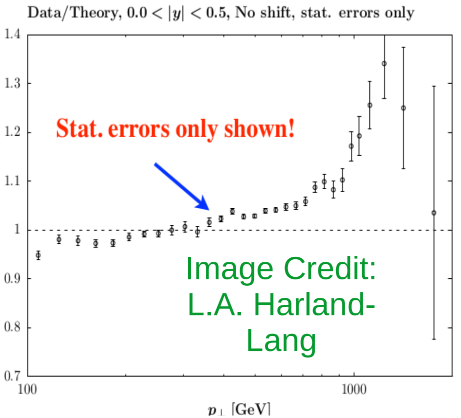
2) **Correlated Systematics –** Issues occur for 2 of the 3 dataset high x gluon data types – top and jets. Consider ATLAS 7 TeV jets:

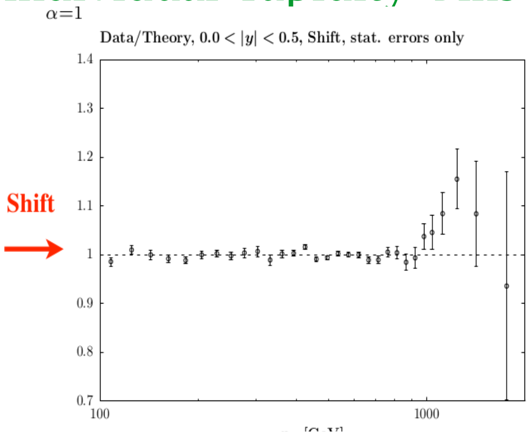
ATLAS 7TeV jets (1410.8857)

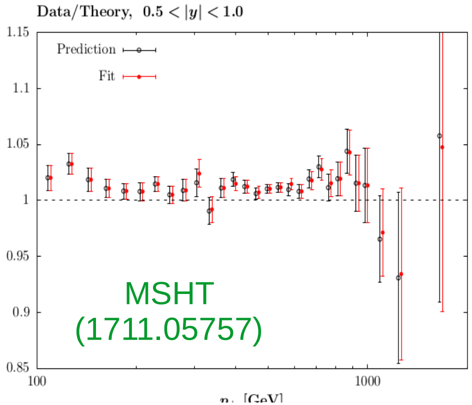
ATLAS 8TeV jets (1706.03192)

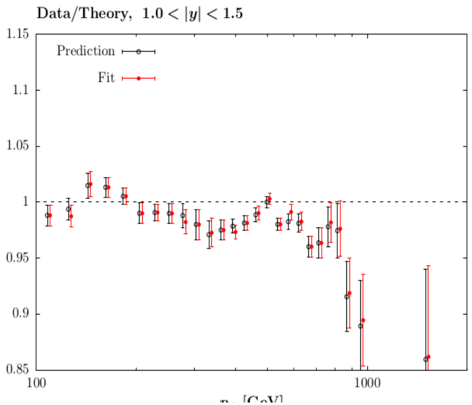
Can fit individual rapidity bins

But not multiple rapidity bins!









Systematic correlations between bins prevent a good fit being obtained, even for neighbouring bins sampling very similar x, Q2. Overly constraining? Decorrelate...

Experimental Challenges

Confronting Precise Data

[ATLAS study - 1706.03192]

2) **Correlated Systematics –** How to deal with issues?

- Experiments examine correlations more closely → guidance for ATLAS 7TeV jets. Useful to provide breakdown of systematics beyond covariance matrix or even full info on models used, broad community support for this. → Cranmer et al (2109.04981)

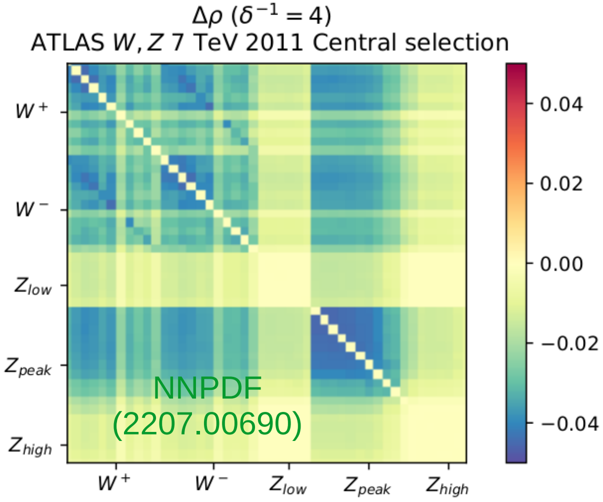
Experimental Challenges

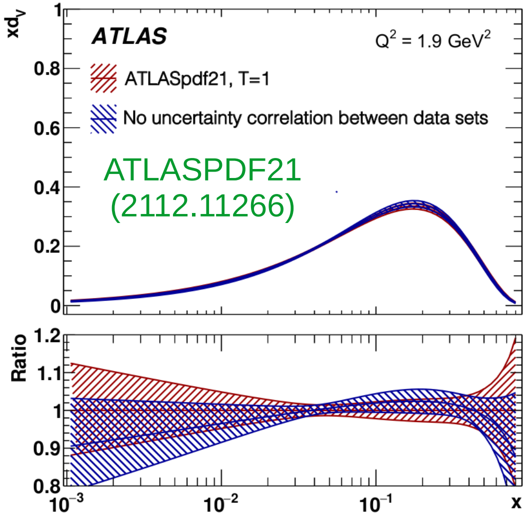
Confronting Precise Data

[ATLAS study - 1706.03192]

2) **Correlated Systematics** – How to deal with issues?

- Experiments examine correlations more closely → guidance for ATLAS 7TeV jets. Useful to provide breakdown of systematics beyond covariance matrix or even full info on models used, broad community support for this. → Cranmer et al (2109.04981)
- Proposal to mitigate these systematic correlation issues by regularisation of the covariance matrix – NNPDF.
- Recent efforts to consider correlations between experiments – ATLASPDF21.





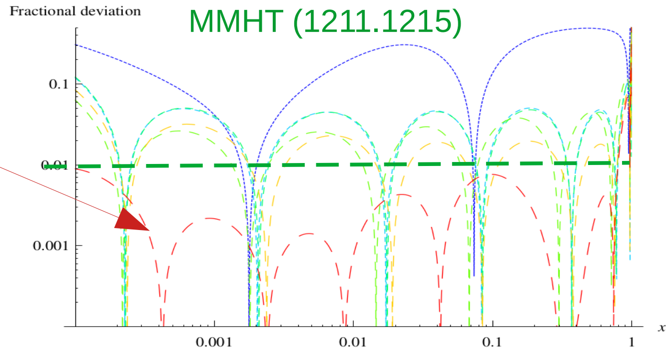
5. Methodological Challenges

Methodological Challenges

Confronting Precise Data

- PDF fitting groups must continually evolve fitting methodology.
- Extended parameterisations, investigate different forms:

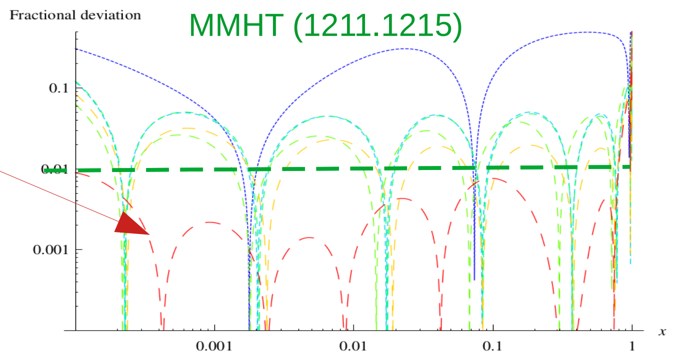
  - MSHT20 → 51 parton parameters to fit to < 1% if data allows. Gives Net $\Delta\chi^2_{\mathrm{global}} = -73$.

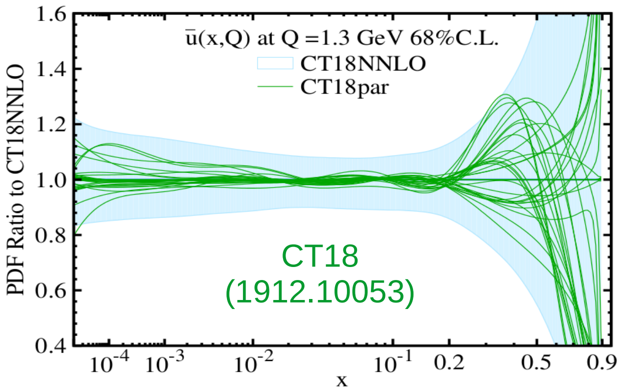


Methodological Challenges

Confronting Precise Data

- PDF fitting groups must continually evolve fitting methodology.
- Extended parameterisations, investigate different forms:
  - MSHT20 → 51 parton parameters to fit to < 1% if data allows. Gives Net $\Delta\chi^2_{\mathrm{global}} = -73$.
  - CT18 → Investigation of different functional forms.



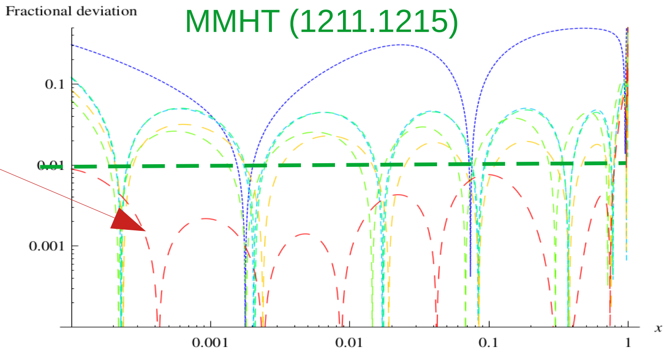


Methodological Challenges

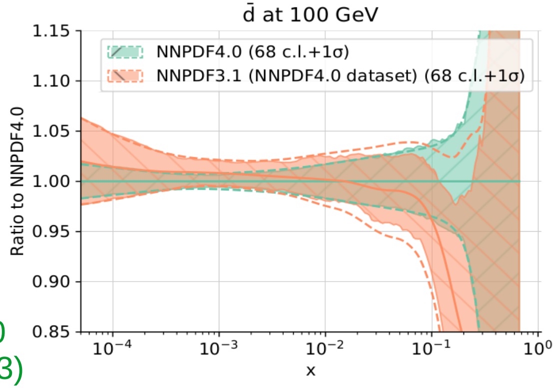
5

Confronting Precise Data

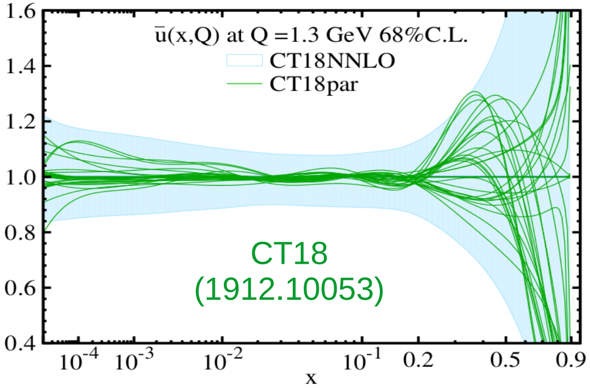
- PDF fitting groups must continually evolve fitting methodology.
- Extended parameterisations, investigate different forms:
  - MSHT20 → **51** parton parameters to fit to < **1%** if data allows. Gives Net $\Delta\chi^2_{\text{global}} = -73$.
  - CT18 → Investigation of different functional forms.
  - NNPDF4 → New algorithm



data set (N_{dat}) \ methodology	NNPDF3.1	NNPDF4.0
NNPDF3.1 (4093)	1.19	1.12
NNPDF4.0 (4491)	1.25	1.17



NNPDF4.0
(2109.02653)



Methodological Challenges

New Codes and Tools

- New tools, approaches can enhance our understanding of data pulls, tensions, etc.
- Tools for PDF studies (small selection given here):

1) Lagrange Multiplier (LM) scans

2) L2 Sensitivity $S_{f,L2}(E) = \vec{\nabla}\chi^2_E \cdot \frac{\vec{\nabla} f}{|\vec{\nabla} f|}$

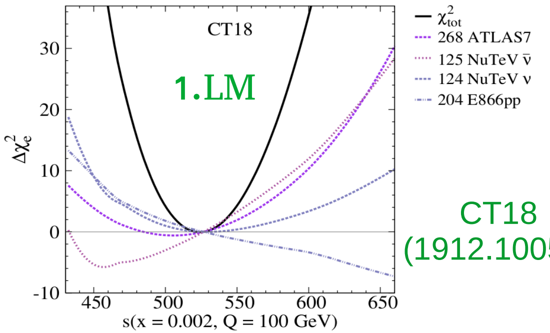
3) Weighted fits

4) Effective Gaussian Variables (S_E)

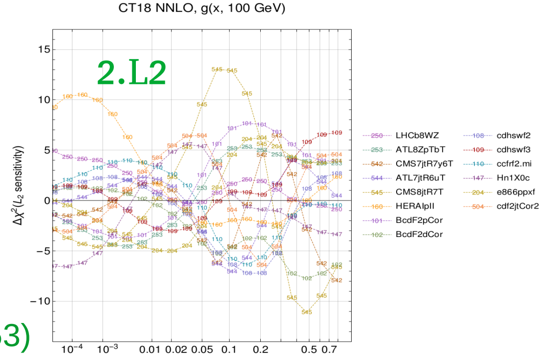
$$S_E = \sqrt{2\chi^2_E} - \sqrt{2N_E - 1}$$

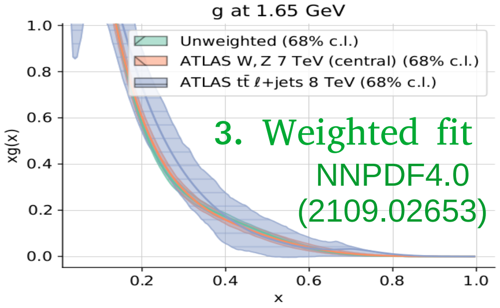
"Spartyness"

5) Many more….

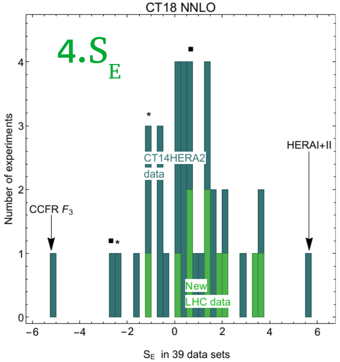


CT18 (1912.10053)





NNPDF4.0 (2109.02653)

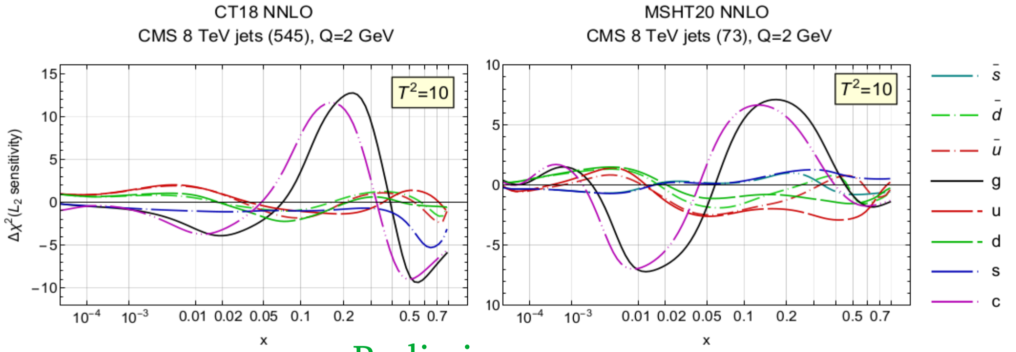


Methodological Challenges

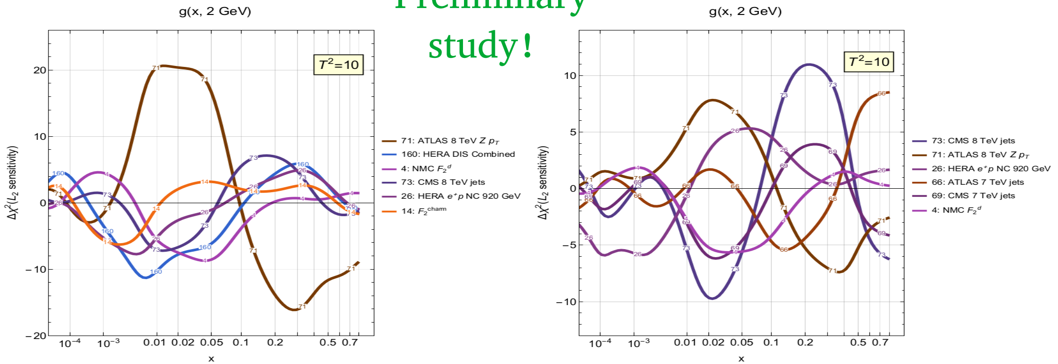
<u>Understanding Data Pulls in Different PDF sets/group:</u>

- Ongoing efforts to understand the effects of datasets in different PDF setups:
- Here using L2 measure:
- - CMS 8TeV jets pull PDFs similarly in CT18 and MSHT20 (top) at NNLO.

  - Pulls on gluon PDF in MSHT20 at NNLO and aN3LO (bottom).
- Useful for understanding effects of different data treatments and methodologies on output PDFs.



Preliminary study!



Methodological Challenges

Uncertainties

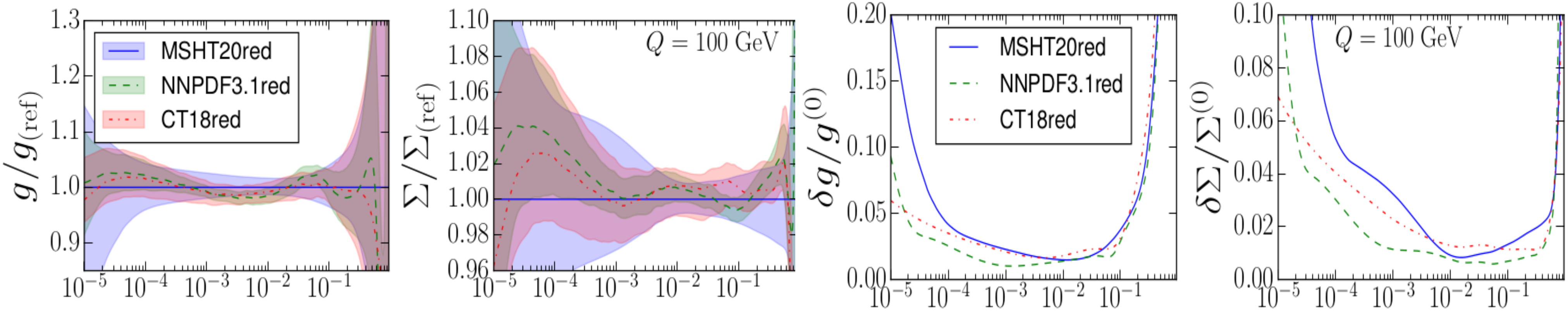
- Different groups see varying sizes of PDF uncertainties. *See talk by P. Nadolsky*
- Tolerance prescriptions of CT, MSHT, ATLASPDF21 account for data tensions, incomplete theory, other issues. ABMP and HERAPDF apply $\Delta\chi^2=1$ → smaller uncertainties.

Methodological Challenges

Uncertainties

- Different groups see varying sizes of PDF uncertainties. *See talk by P. Nadolsky*
- Tolerance prescriptions of CT, MSHT, ATLASPDF21 account for data tensions, incomplete theory, other issues. ABMP and HERAPDF apply $\Delta\chi^2=1$ → smaller uncertainties.
- Investigated further using reduced fits in PDF4LHC21
  - ***fit same data*** → consistent PDFs but differing uncertainties
  - further work ongoing to understand this by several groups...

PDF4LHC21 (2203.05506)

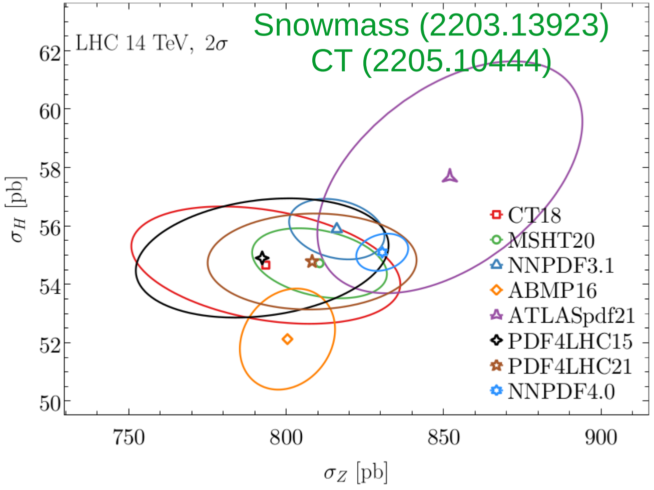


Methodological Challenges

Uncertainties

- Different groups see varying sizes of PDF uncertainties. Other explanations?
- Data sampling – Sampling large multidimensional parameter spaces is difficult.

See talk by P. Nadolsky



Methodological Challenges

Uncertainties

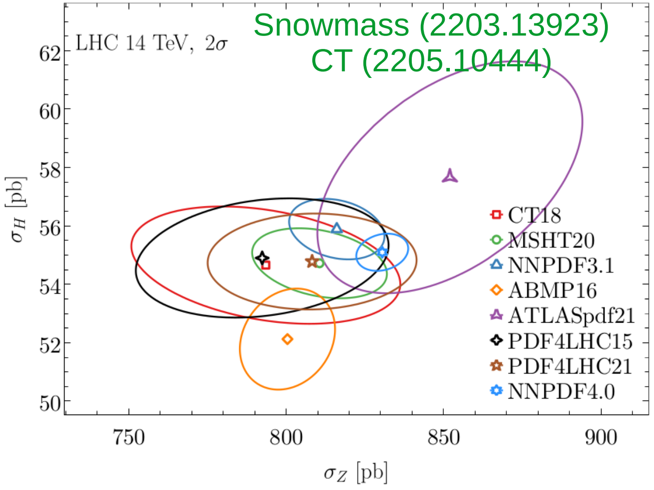
- Different groups see varying sizes of PDF uncertainties. Other explanations?
- Data sampling – Sampling large multidimensional parameter spaces is difficult.
- Various ways to test this

See talk by P. Nadolsky

- Closure/future testing – use artificial/restricted data to test for bias and uncertainty sizes.

MSTW(1205.4024)
NNPDF (2103.08606)
NNPDF4.0 (2109.02653)

- Parameter space scan, look for additional solutions and compare with uncertainties (e.g. "hopscotch").



Methodological Challenges

Uncertainties

- Different groups see varying sizes of PDF uncertainties. Other explanations?
- Data sampling – Sampling large multidimensional parameter spaces is difficult.
- Various ways to test this

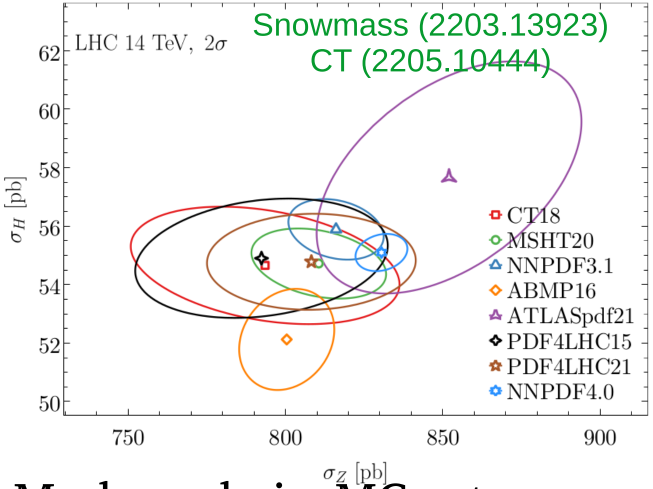
See talk by P. Nadolsky

- Closure/future testing – use artificial/restricted data to test for bias and uncertainty sizes.

MSTW(1205.4024)
NNPDF (2103.08606)
NNPDF4.0 (2109.02653)

- Parameter space scan, look for additional solutions and compare with uncertainties (e.g. "hopscotch").

Snowmass (2203.13923)
CT (2205.10444)



- CJ/JAM study – compared uncertainty estimates in toy model from Hessian, data resampling, nested sampling, Markov chain MC, etc

See talk by A. Accardi Hunt-Smith et al (2206.10782)

Methodological Challenges

5

Is it there and can we see it?

Intrinsic Charm (IC)

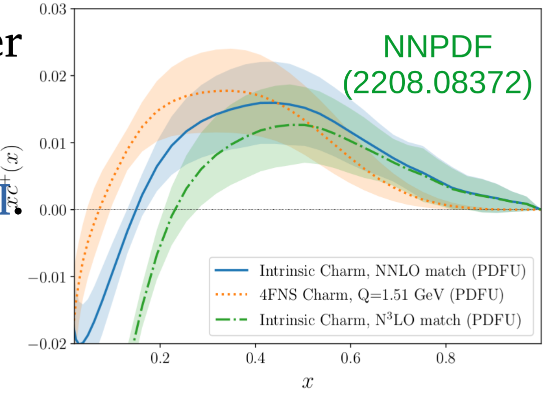
- Usual perturbative charm PDF generated from DGLAP splittings above m_c.
- Various theoretical models for IC, BHPS ("valence-like"), "sea-like", meson-baryon.
- NNPDF obtains "fitted charm" by fitting 4FNS c PDF and inverting at matching scale.
- Difficulty is separating IC from higher twist, process dep, higher order and other effects.

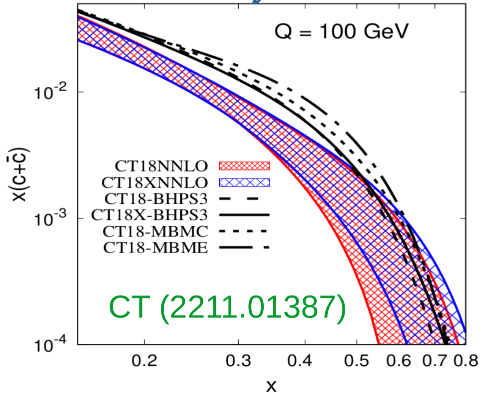
See talk by T.J. Hobbs

- Data on High x DIS, LHCb, etc may offer sensitivity.
- Issues of flavoured jets, NNLO QCD, MPI.

Gauld et al (2302.12844)

- Future measurements at EIC, FPF.





6. Theoretical Challenges

Theoretical Challenges

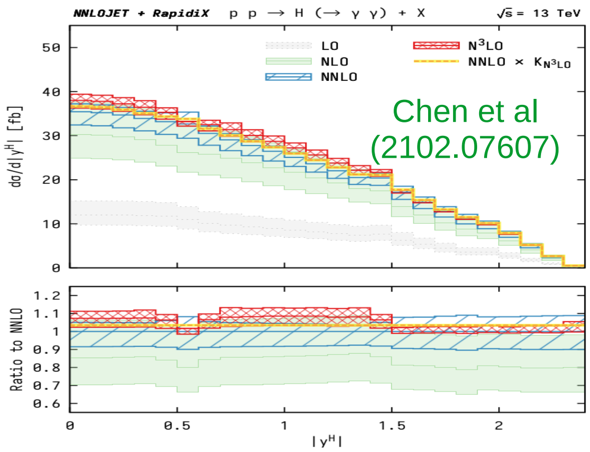
Confronting Precise Data

- To exploit precision data we need precision theory predictions. Must now consider higher orders and associated theoretical uncertainties, and other effects.

Need for Higher Orders (N3LO):

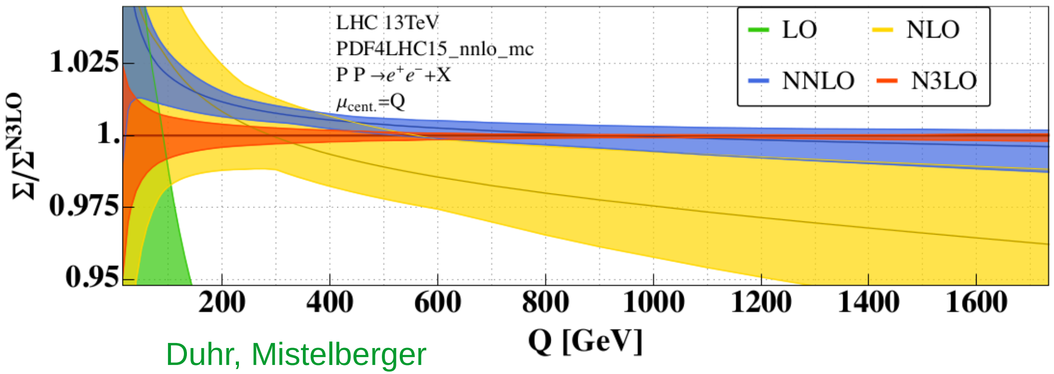
- Progress in recent years on N3LO cross-sections for key processes, e.g. Higgs, DY:

ggF



Chen et al (2102.07607)

NC DY Z/photon

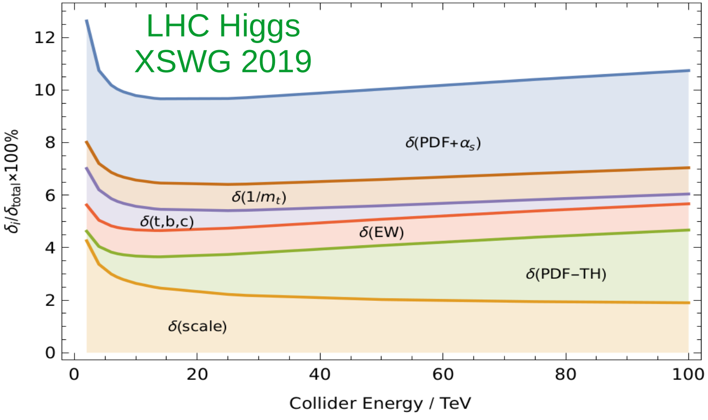


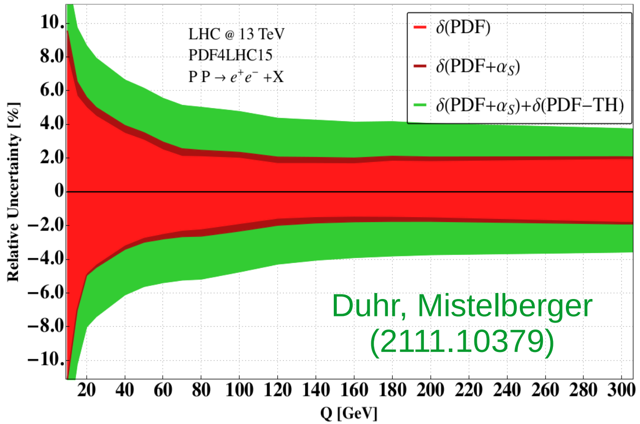
Duhr, Mistelberger (2111.10379)

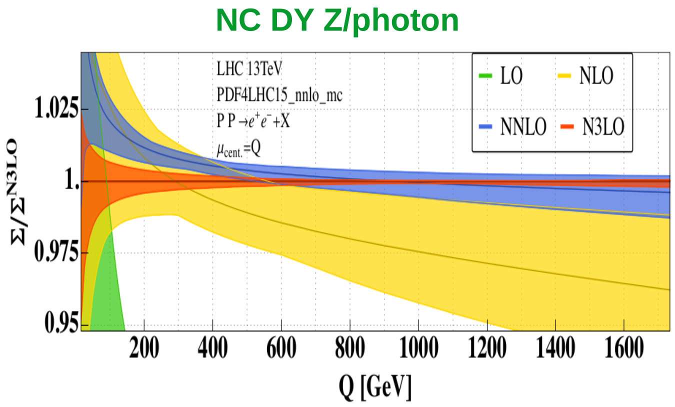
Theoretical Challenges

Need for Higher Orders

- Only NNLO PDFs have been available – mismatch between cross-section and PDF order.





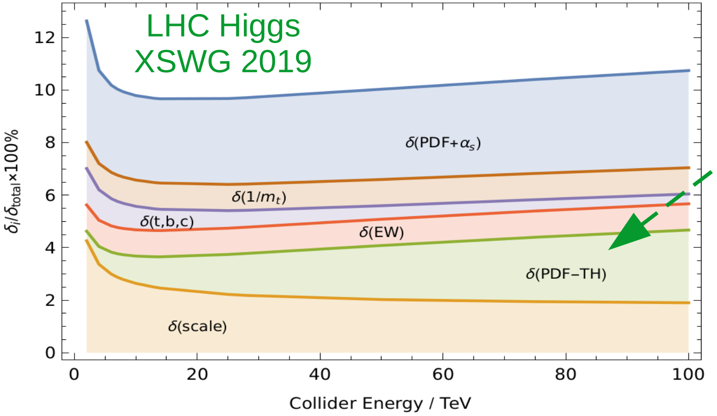


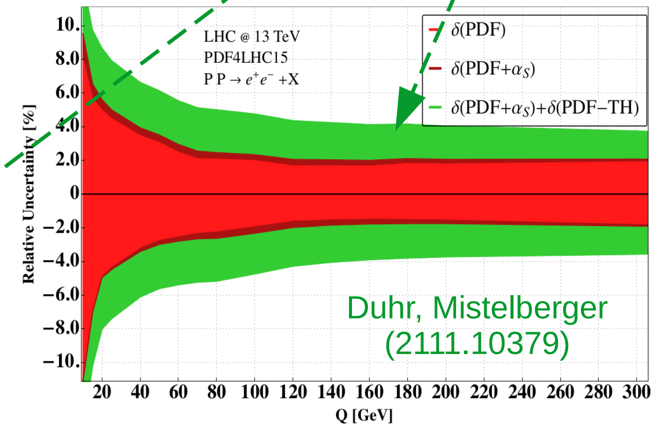
Theoretical Challenges

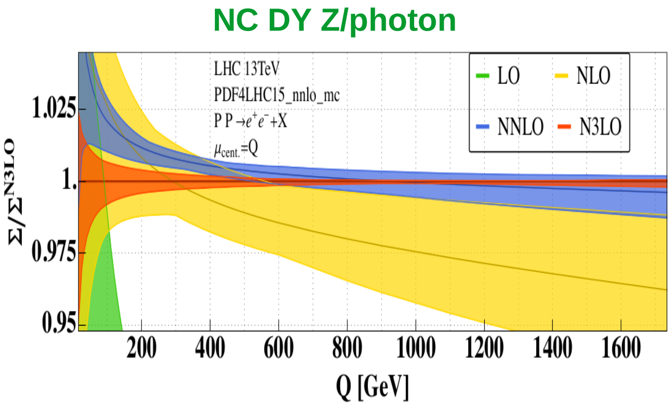
6

Need for Higher Orders

- Only NNLO PDFs have been available – mismatch between cross-section and PDF order. This adds a source of uncertainty – **“PDF-th”**.





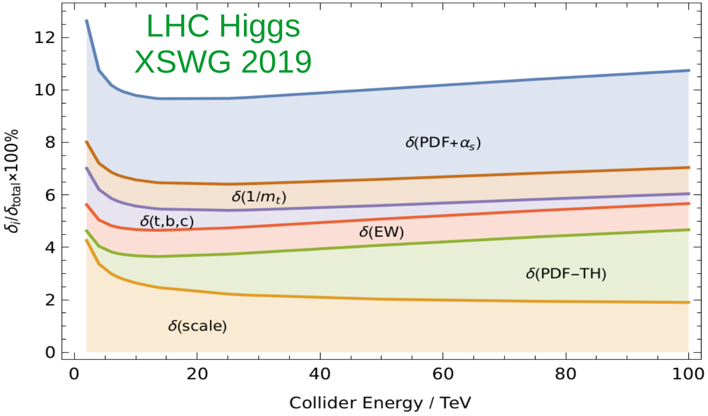


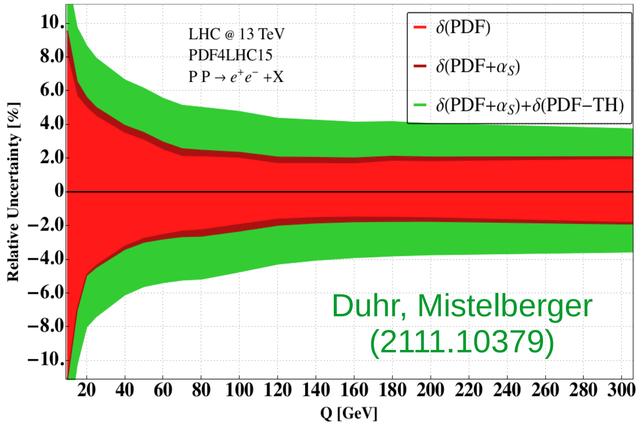
Theoretical Challenges

6

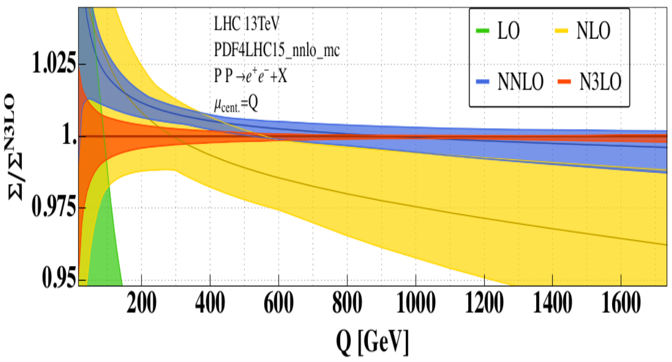
Need for Higher Orders

- Only NNLO PDFs have been available – mismatch between cross-section and PDF order. This adds a source of uncertainty – **“PDF-th”**.
- Without consideration of this you cannot estimate the full theoretical uncertainty.
- Only way to remove these bands and properly understand associated uncertainties is determining **N3LO PDFs**, plus inclusion of PDF theoretical uncertainties.





NC DY Z/photon



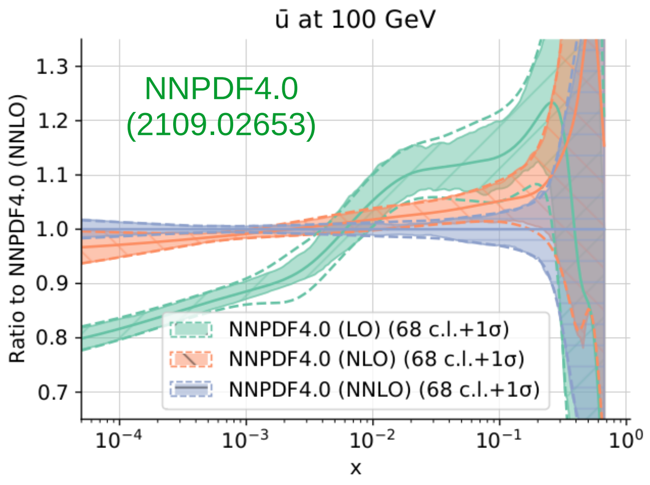
Theoretical Challenges

6

Theoretical Uncertainties - MHOUs:

- PDF uncertainties have up to now typically neglected theoretical uncertainties.
- In limit experimental systematics are perfectly known and statistical uncertainties reduce to 0 then $\chi^2 \to \infty$, as theory at fixed order will not match data.
- Need to add theoretical uncertainties into PDFs due to

Missing Higher Order Uncertainties (MHOUs).

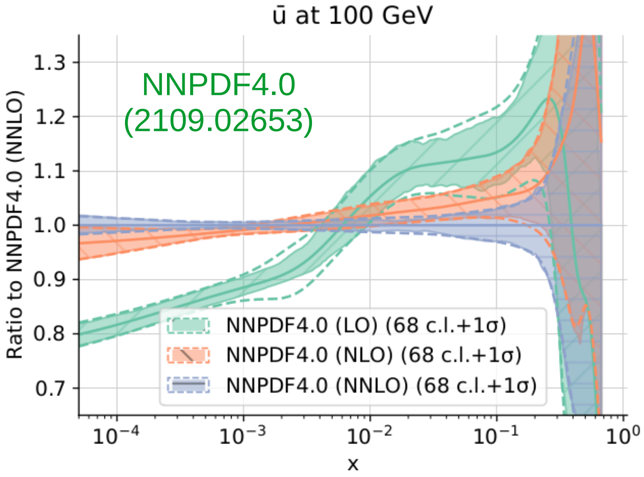


Theoretical Challenges

Theoretical Uncertainties - MHOUs:

- PDF uncertainties have up to now typically neglected theoretical uncertainties.
- In limit experimental systematics are perfectly known and statistical uncertainties reduce to 0 then $\chi^2 \to \infty$, as theory at fixed order will not match data.
- Need to add theoretical uncertainties into PDFs due to Missing Higher Order Uncertainties (MHOUs).
- Three main approaches:

1) Scale variation/joint fits
2) Bayesian approaches
3) Theoretical Nuisance Parameters



Theoretical Challenges

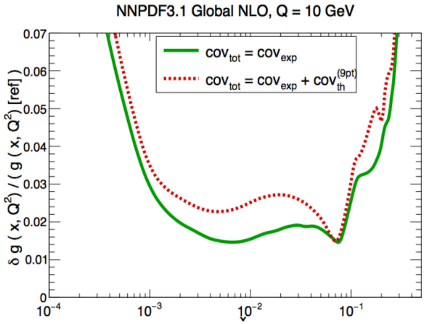
6

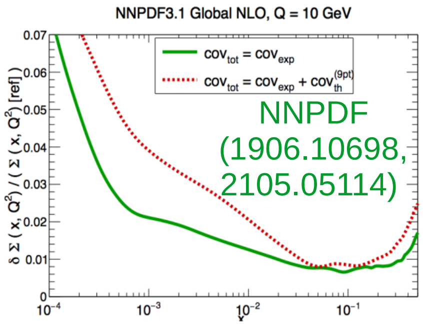
Theoretical Uncertainties - MHOUs:

1) Scale variation - Include scale variations as proxy.

- NNPDF have done NLO, using "theory covariance matrix", S.
- Get small improvements in χ^2/N and larger uncertainties.
- Potential issue of double counting scale variations in PDFs and cross-sections.
- Degree of variation used is arbitrary, only probes (N)NLO terms.

MMHT (1811.08434)





NNPDF (1906.10698, 2105.05114)

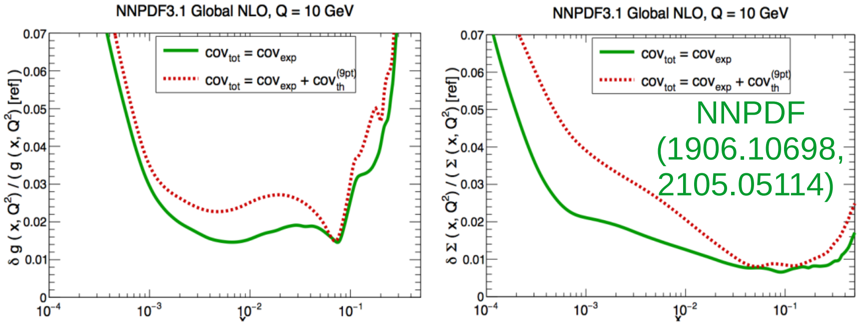
$\chi^2/N_{\rm dat}$ in the NNPDF3.1 global fits	C	$C+S^{(9pt)}$	$C+S^{(7pt)}$	$C+S^{(3pt)}$	$C+S^{(9pt)}_{\rm fit}$	$C+S^{(9pt)}_{\rm samp}$
	1.139	1.109	1.129	1.139	1.113	1.220

Theoretical Challenges

Theoretical Uncertainties - MHOUs:

1) Scale variation - Include scale variations as proxy.
   - NNPDF have done NLO, using "theory covariance matrix", S.
   - Get small improvements in χ^2/N and larger uncertainties.
   - Potential issue of double counting scale variations in PDFs and cross-sections.
   - Degree of variation used is arbitrary, only probes (N)NLO terms.

MMHT
(1811.08434)



$\chi^2/N_{\rm dat}$ in the NNPDF3.1 global fits	C	$C+S^{(9\rm pt)}$	$C+S^{(7\rm pt)}$	$C+S^{(3\rm pt)}$	$C+S^{(9\rm pt)}_{\rm fit}$	$C+S^{(9\rm pt)}_{\rm samp}$
	1.139	1.109	1.129	1.139	1.113	1.220

Theoretical Challenges

6

Theoretical Uncertainties - MHOUs:

1) Scale variation - Include scale variations as proxy.

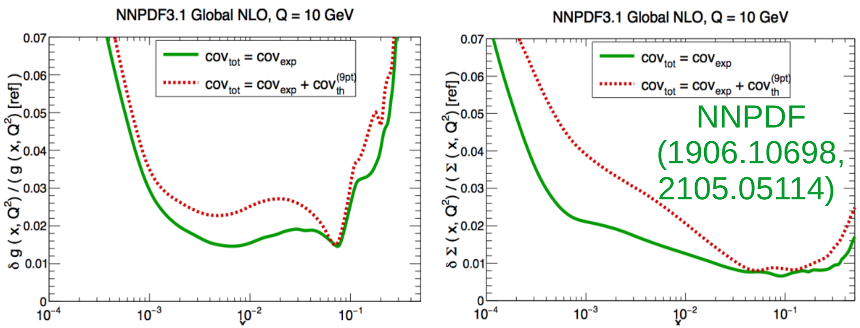
- NNPDF have done NLO, using "theory covariance matrix", S.
- Get small improvements in χ^2/N and larger uncertainties.
- Potential issue of double counting scale variations in PDFs and cross-sections.
- Degree of variation used is arbitrary, only probes (N)NLO terms.

MMHT (1811.08434)

2) Bayesian approach – Determine model dependence on order in statistically defined way. Not used in PDFs yet.

- Bonvini and Cacciari Houdeau models.

Bonvini (2006.16293)
Cacciari et al (1105.5152, 1409.5036)



NNPDF (1906.10698, 2105.05114)

$\chi^2/N_{\rm dat}$ in the NNPDF3.1 global fits	C	$C+S^{(9pt)}$	$C+S^{(7pt)}$	$C+S^{(3pt)}$	$C+S^{(9pt)}_{\rm fit}$	$C+S^{(9pt)}_{\rm samp}$
	1.139	**1.109**	**1.129**	**1.139**	**1.113**	**1.220**

Theoretical Challenges

6

Theoretical Uncertainties - MHOUs:

MSHT (2207.04739)

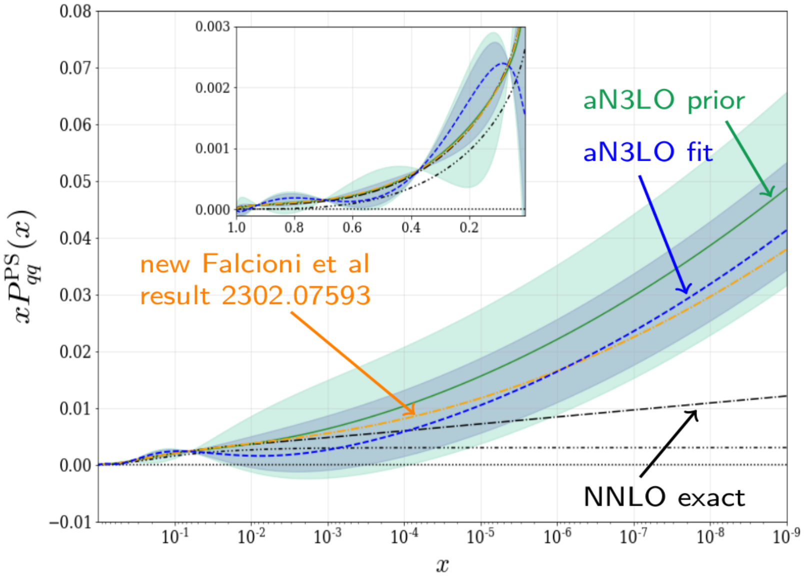
3) Theory Nuisance Parameters and known N3LO

See talk Tuesday morning

- Idea is to include known N3LO effects already into PDFs and to parameterise remaining unknown pieces via theoretical nuisance parameters.

- Variation of theoretical nuisance parameters then probes exactly the N3LO MHO terms + gives theoretical uncertainty on aN3LO PDF fit

→ **MSHT20aN3LO PDFs.**



See talks Thurs afternoon e.g. A. Pelloni, K. Schönwald

Theoretical Challenges

6

Approximate N3LO (aN3LO) PDFs + Theoretical Uncertainties

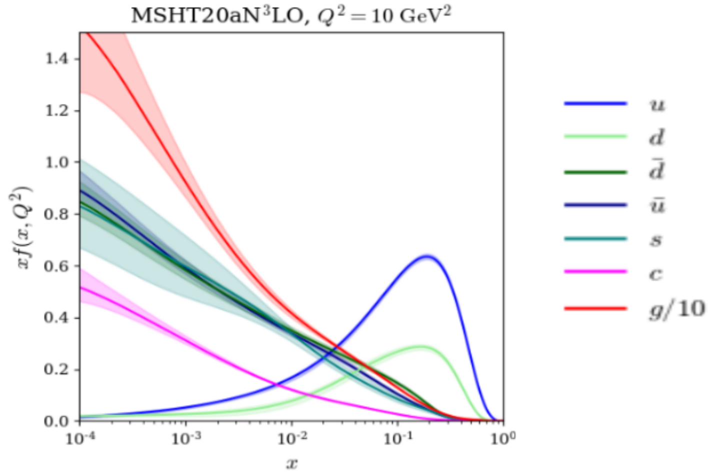
MSHT (2207.04739)

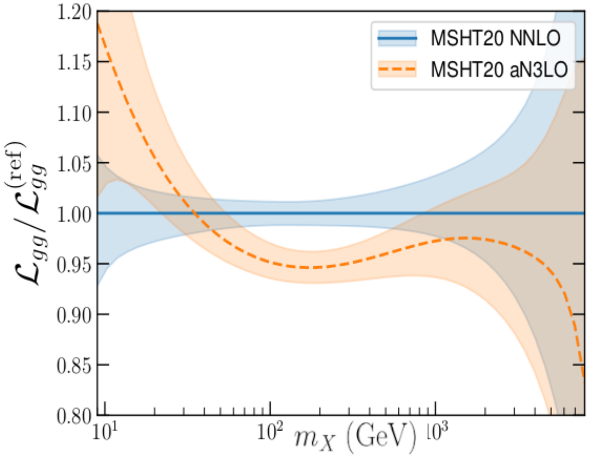
- Known N3LO info improves the fit qualities and alters the PDFs + uncertainties.
- Consequences for phenomenology:

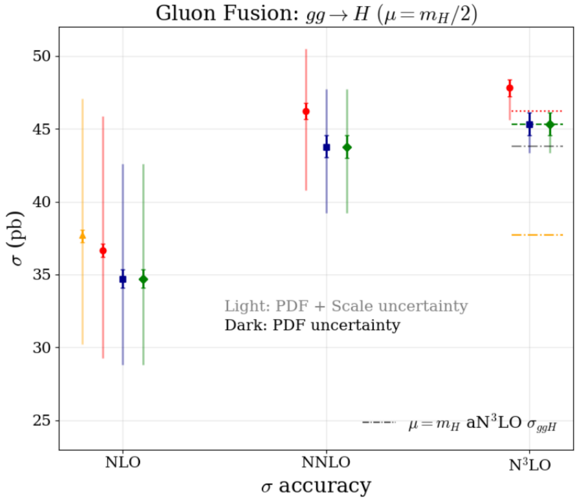
Also reduces dataset tensions

See talk Tuesday morning

- gg luminosity changes, reduction around 100GeV → affects Higgs ggF production.
- (slightly) increased uncertainties due to addition of theoretical MHOU component.





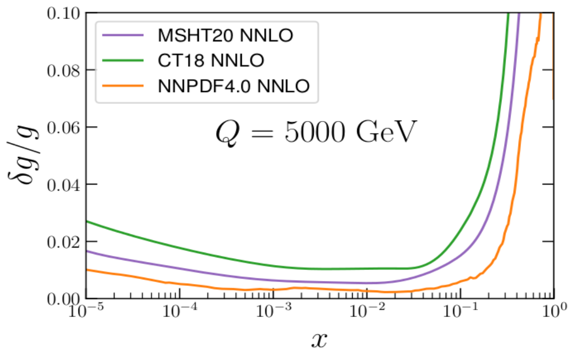


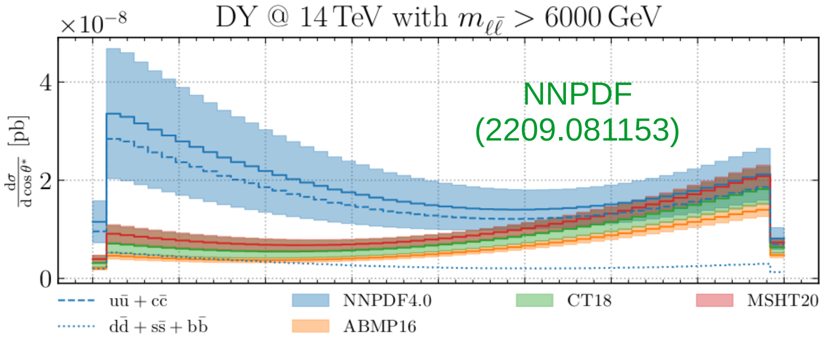
Theoretical Challenges

6

PDFs and Beyond Standard Model Physics

- PDF uncertainties grow rapidly at large x → limit searches for BSM at high mass.
- Parameterisation or other assumptions here also can have an affect e.g. in DY AFB.



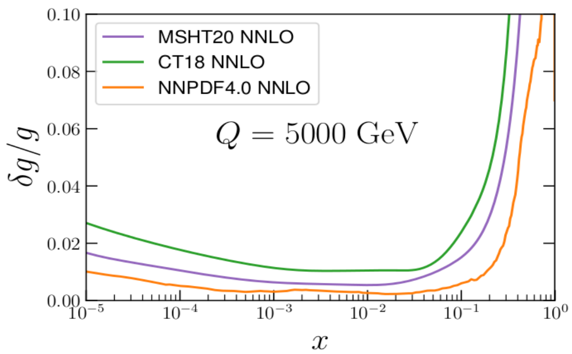


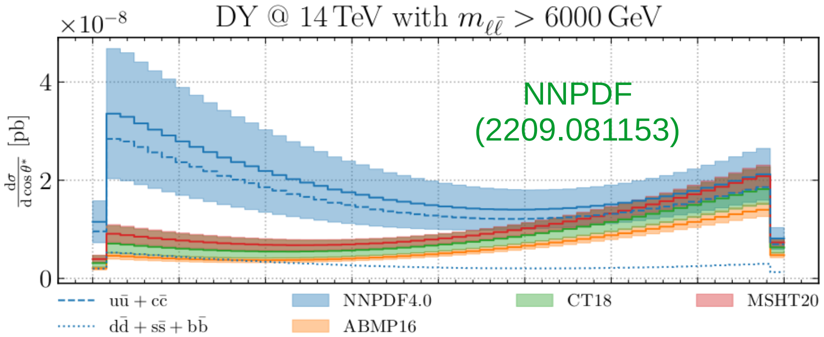
Theoretical Challenges

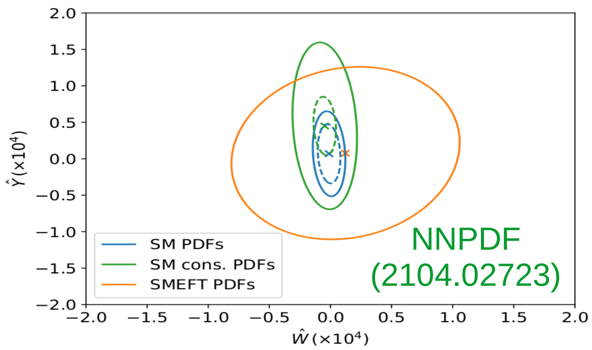
6

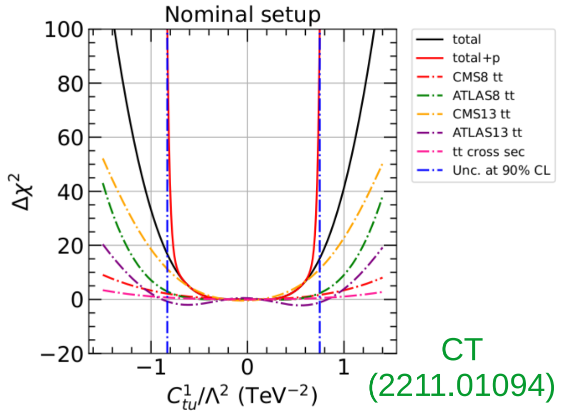
PDFs and Beyond Standard Model Physics

- PDF uncertainties grow rapidly at large x → limit searches for BSM at high mass.
- Parameterisation or other assumptions here also can have an affect e.g. in DY AFB.
- Meanwhile, for fitting of SMEFT parameters, there might be notable correlations between PDFs and the SMEFT → suggests doing a joint fit.









7. Future Progress

Future Progress

- **Short term** - progress on experimental, methodological, theoretical fronts – more precise data, methodological advances, N3LO + MHOUs, etc
- **Medium/long term** - several further projected improvements in PDF accuracy and precision from confirmed/proposed future experiments.

See WG6 talks

Complementarity between all future experiments and knock on effects for physics goals!

Future Progress

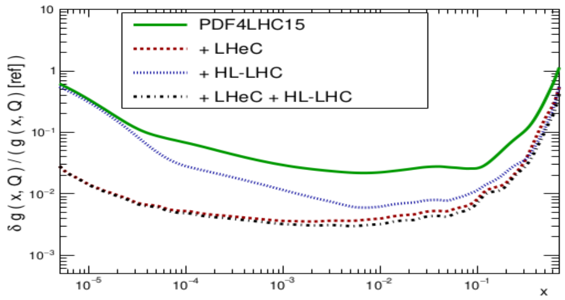
7

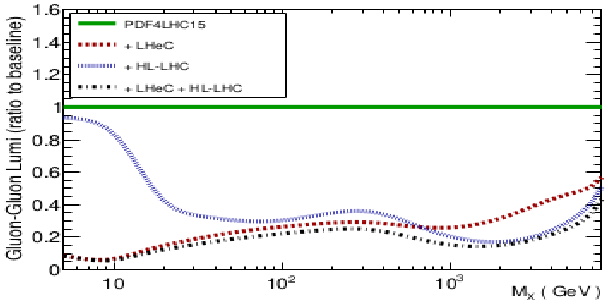
- **Short term** - progress on experimental, methodological, theoretical fronts – more precise data, methodological advances, N3LO + MHOUs, etc
- **Medium/long term** - several further projected improvements in PDF accuracy and precision from confirmed/proposed future experiments.

See WG6 talks

- **HL-LHC** – Reduce PDF uncertainties where processes currently statistically limited/coverage can be extended, e.g. high x gluon and gg luminosity.
- **LheC** – Inclusive/Semi-inclusive DIS data constrain intermediate/small x.

LHeC review (2007.14491)





Abdul Khalek et al (1906.10127)

Complementarity between all future experiments and knock on effects for physics goals!

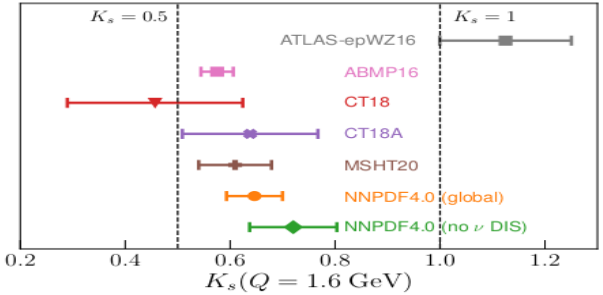
Future Progress

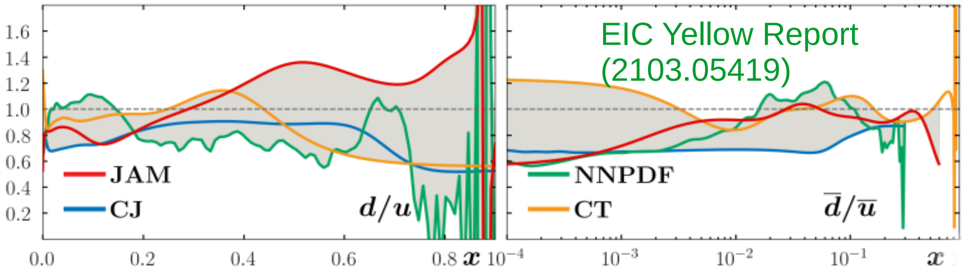
- **<u>Short term</u>** - progress on experimental, methodological, theoretical fronts – more precise data, methodological advances, N3LO + MHOUs, etc
- **<u>Medium/long term</u>** - several further projected improvements in PDF accuracy and precision from confirmed/proposed future experiments.

See WG6 talks

- **FPF** – Very forward neutrino production → intrinsic charm at high x, very low x gluon dynamics. Then Neutrino CC DIS → flavour separation.
- **EIC** – Constrain high x quarks via inclusive NC/CC DIS.

FPF review (2203.05090)





<u>Complementarity between all future experiments</u> and knock on effects for physics goals!

8. Conclusions

TC is funded from a project of the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No. 101002090 COLORFREE).

Summary

- Parton Distribution Functions remain a <u>crucial input for our goals at colliders</u>.
- At the same time they are <u>also a key output</u> of ongoing/future experiments.
- Thanks to global efforts from the experimental and theoretical communities **PDFs are currently more accurate and precise than ever before**.
- However they face challenges on experimental, methodological and theoretical fronts to keep pace with the demands. And we must be careful to ensure *accuracy and precision*.
- Recent significant progress on many issues from understanding dataset tensions, to examining uncertainties and including higher orders (approximate N3LO) and theoretical uncertainties.
- <u>Complementarity between different groups is greatly beneficial</u> for these aims.
- Several exciting talks across DIS23 and different WGs on a variety of these topics.

Summary

- Parton Distribution Functions remain a crucial input for our goals at colliders.
- At the same time they are also a key output of ongoing/future experiments.
- Thanks to global efforts from the experimental and theoretical communities **PDFs are currently more accurate and precise than ever before**.
- However they face challenges on experimental, methodological and theoretical fronts to keep pace with the demands. And we must be careful to ensure *accuracy and precision*.
- Recent significant progress on many issues from understanding dataset tensions, to examining uncertainties and including higher orders (approximate N3LO) and theoretical uncertainties.
- Complementarity between different groups is greatly beneficial for these aims.
- Several exciting talks across DIS23 and different WGs on a variety of these topics.

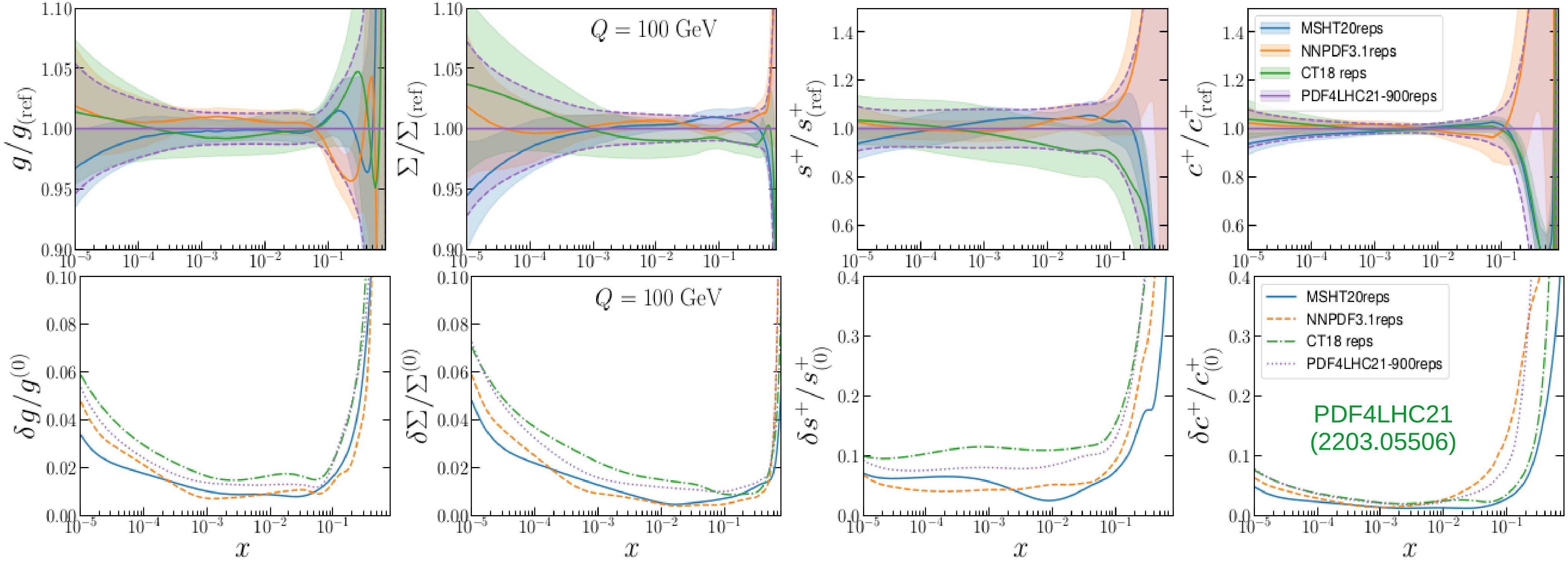
Thankyou! Any Questions?

Backup

TC is funded from a project of the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No. 101002090 COLORFREE).

PDF4LHC21

- PDF4LHC21 combination of MSHT20, CT18', NNPDF3.1' global PDF sets.



- Uncertainties reflect differences in central values as well as individual uncertainties.

Methodological Challenges

More info on positivity or otherwise of MS PDFs:
NNPDF (2006.07377)
Collins et al (2111.01170)

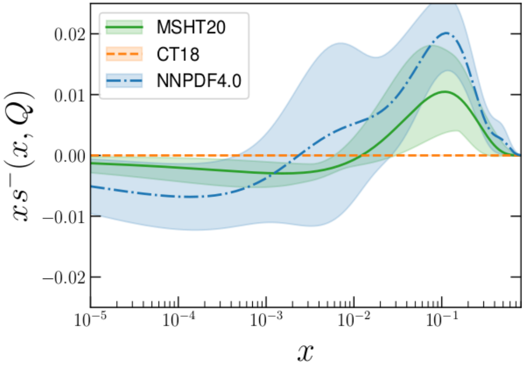
Additional Constraints

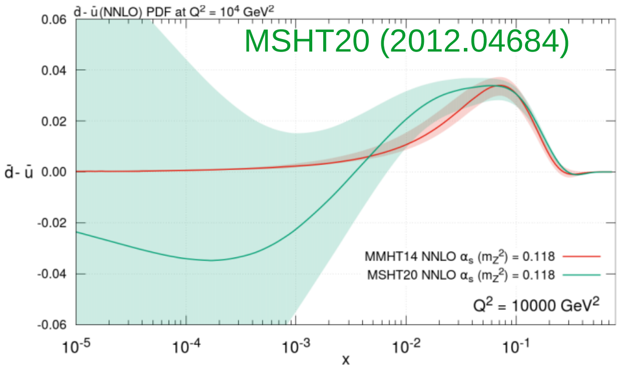
- In order to ensure physical PDFs, often additional constraints are added. Many different types and methods:

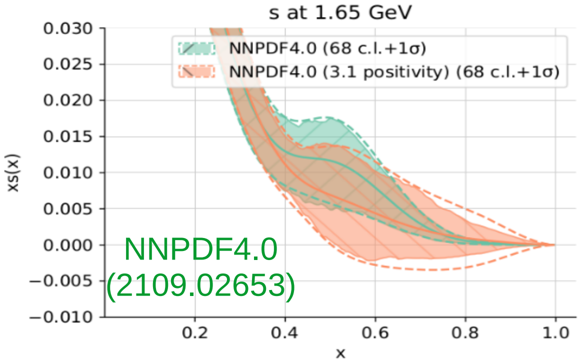
1) Parameterisation – Behaviour at low/high x where data limited. Can also be applied through pre-processing, or priors on parameters.

Also 3. Lattice constraints – see T. J. Hou Talk…

2) Positivity and integrability – Can require positivity of observables (DIS S.F.s), NNPDF4.0 also enforces positivity of g, light q PDFs via hard-wall χ^2 penalties.





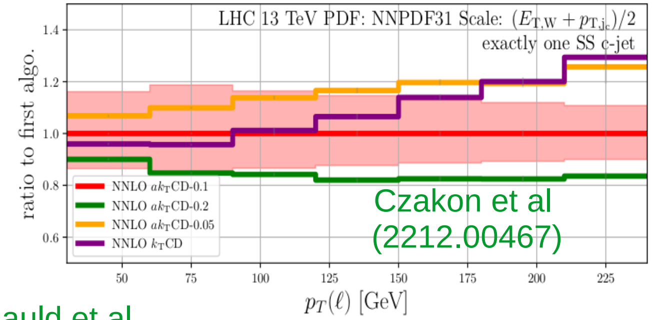


Methodological Challenges

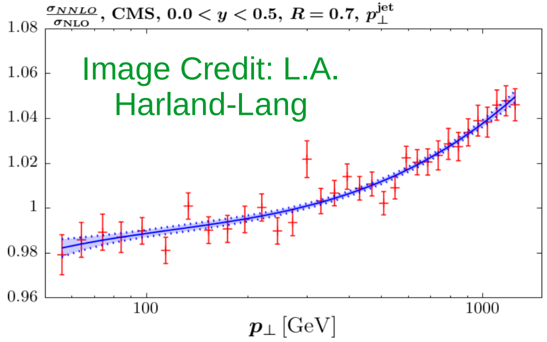
Theoretical Grids

- PDF fitting needs theoretical predictions – encoded by theory grids, produced once.
- Share grids via online repositories (applgrid, fastnlo, ploughshare).
- For most datasets, only NLO QCD grids + NNLO k-factors available.
- Differences also exist in treatment of Monte-Carlo errors in k-factors (right).
- However, full NNLO QCD grids becoming available for several processes – e.g. top
- Important if we wish to consider higher orders (see later!).
- Challenges to compute NNLO in when flavoured jet – W+c, Z+c data.

Czakon et al (1912.08801)



Gauld et al (2208.11138)



Theoretical Challenges

Deuteron and nuclear corrections

- Data of DIS scattering off deuteron/nuclear targets allows separation of u/d at high x and to examination of flavour decomposition via CC → used in PDF fits.
- Complications of dealing with corrections from deuteron/nuclear environment.
- Different groups use different treatments, generally % effects but more at high x.
- Connected issues of higher twists, target mass corrections, e.g. MARATHON.

See R. Petti talk