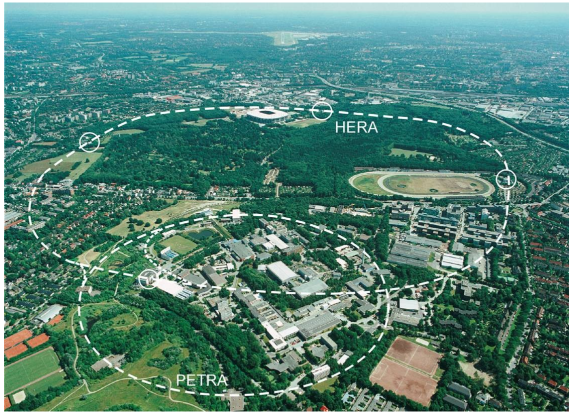
Extracting Precision PDFs

- Background, Status, State of the Art



YETI 2026 School: Precision x Collider
IPPP, Durham, U.K.

Thomas Cridge, 12th July 2026

THE ROYAL SOCIETY

MANCHESTER 1824
The University of Manchester

Outline

Many details are schematic given time!

Much you may already know, particularly in introduction.

1) **Motivation** – Need for Precision PDFs.
2) **Introduction** – PDFs, evolution and Factorisation.
3) **PDF Current Status and comparison**
4) **PDF constraints – Datasets** → Fixed Target, DIS Structure Functions, neutrino scattering, DY, Jets, Top, ZpT.
5) **PDF methodology and Uncertainties**
6) **PDF Theory state of the art** - QED improved PDFs, theory uncertainties, aN3LO
7) **Strong coupling, top mass**

Far from exhaustive, chosen my favourite examples, obviously often used MSHT

References

1

Many good references of varying levels of complication, several of which (and others) used in compiling this talk – *thanks to the authors*!

Far from exhaustive!

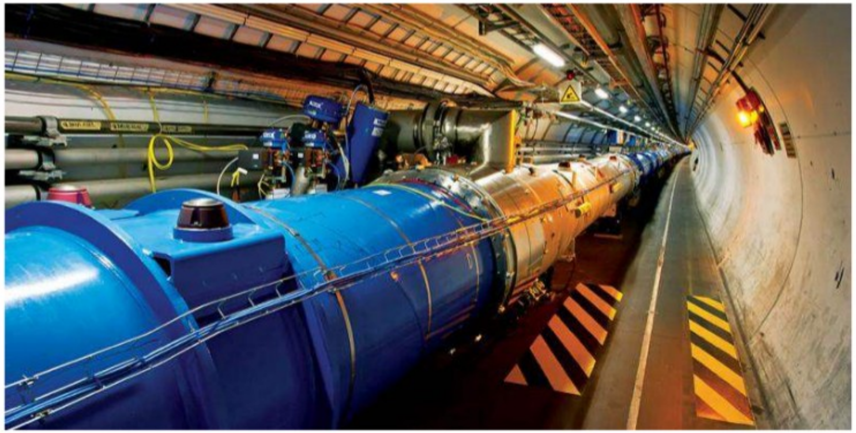
1) Black book of QCD – Campbell, Huston, Krauss
2) QCD and Collider Physics ("Pink book") – Ellis, Stirling, Webber
3) Modern Particle Physics - Thomson.
4) Review of Particle Physics (Sections 9, 15, 18 and others)
5) Handbook of Perturbative QCD – Sterman et al
6) Various review articles – e.g. Gao et al 1709.04922, Ridolfi, Dissertori et al, Kovarik et al 1905.06957, Forte and Watt (1301.6754), Accardi et al (1603.08906)
7) Many talks available online – CTEQ school, GGI lectures, etc (several used in compiling this talk!) and those by Thorne, Harland-Lang, Diehl, Guzzi, Salam, Martin, Nadolsky, Forte, Stump, Melnitchouk, Guffanti, Rojo, Ubiali and more!

Material, plots, slides taken from my own work + these refs + talk by Lucian Harland-Lang

1. Motivation

Motivation – PDFs

- Large Hadron Collider (LHC) is a proton-proton (pp) collider → Almost all results, analyses, data, theory predictions etc depend on proton content.



- *Parton Distribution Functions (PDFs)* are a crucial input and key output of collider physics. Encode non-perturbative content of the proton.
- "PDFs", represent probability of finding a parton of type i carrying momentum fraction x of the proton. Interpretation complicates at higher orders in QCD.

Motivation – PDFs

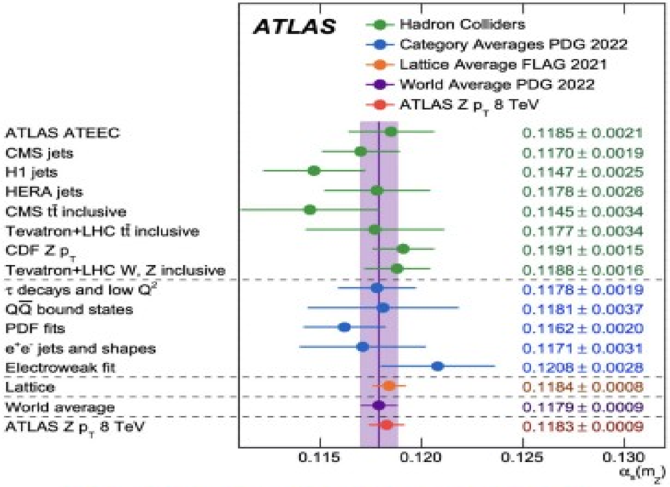
- Key input to very many calculations/measurements at colliders → Need both **accuracy and precision**. Moreover, often a dominant contribution to *uncertainty*.

1) Precision Standard Model (SM) Measurements –

Often $\delta_{\mathrm{PDF}} \sim \frac{1}{2}\delta_{\mathrm{tot}}$

ATLAS α_s determination

$\Delta_{TOT} = 0.0009$, $\Delta_{PDF} = 0.0005$

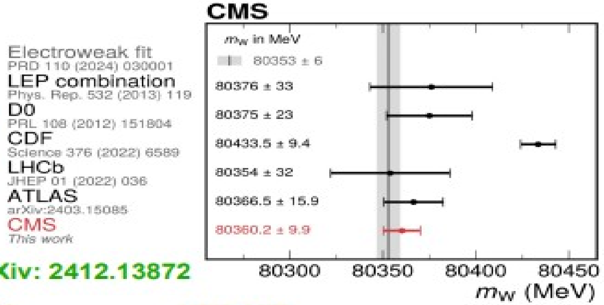


ATLAS collaboration, arXiv: 2309.12986

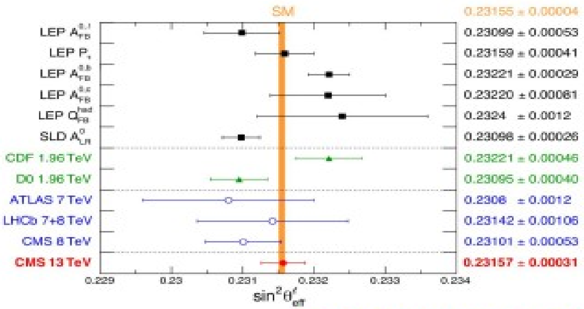
CMS M_W determination

Δ_{TOT} = 9.9 MeV

Δ_{PDF} = 4.4 MeV



CMS collaboration, arXiv: 2412.13872



CMS $\sin^2\theta_W$ determination

$\Delta_{TOT} = 31 \times 10^{-4}$

$\Delta_{PDF} = 27 \times 10^{-4}$

CMS collaboration, arXiv: 2408.07622

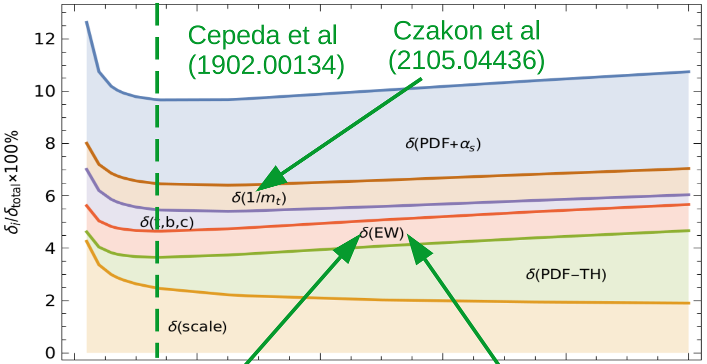
Motivation – PDFs

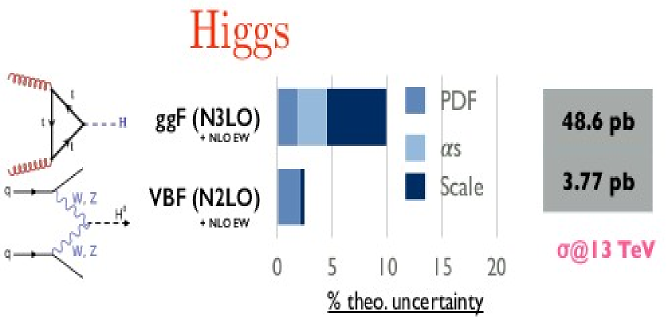
- Key input to very many calculations/measurements at colliders → Need both **accuracy and precision**. Moreover, often a dominant contribution to *uncertainty*.

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.

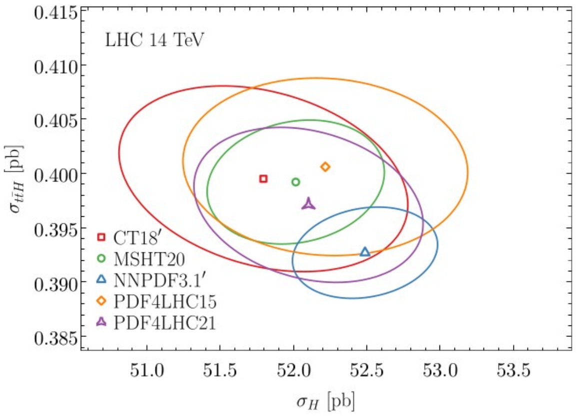
LHC Higgs XSWG 2019





Often $\delta_{\mathrm{PDF}} \sim \frac{1}{2}\delta_{\mathrm{tot}}$

PDF4LHC21 (2203.05506)



Motivation – PDFs

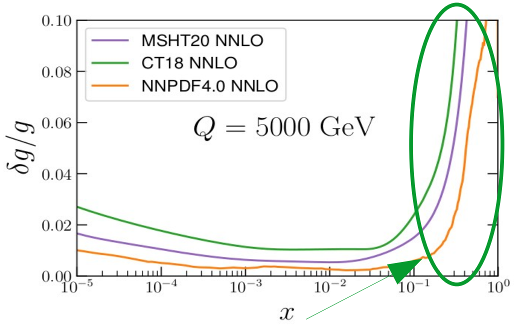
1

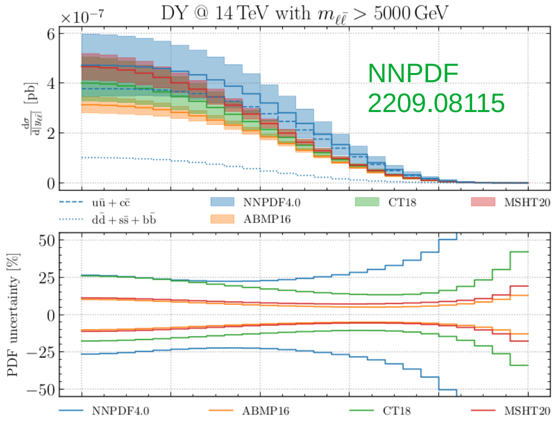
- Key input to very many calculations/measurements at colliders → Need both **accuracy and precision**. Moreover, often a dominant contribution to *uncertainty*.

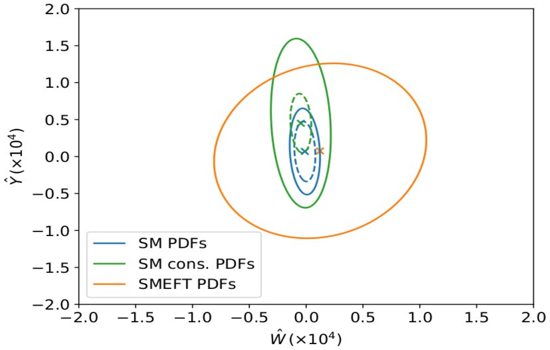
3) Beyond Standard Model (BSM) Searches:

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

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







If looking in high energy tails for new BSM effects, PDFs limit sensitivity
- (left) dijet searches gluon PDF uncertainty large
- (centre) DY AFB PDF uncertainty grows.

NNPDF 2209.08115

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

Motivation – PDFs

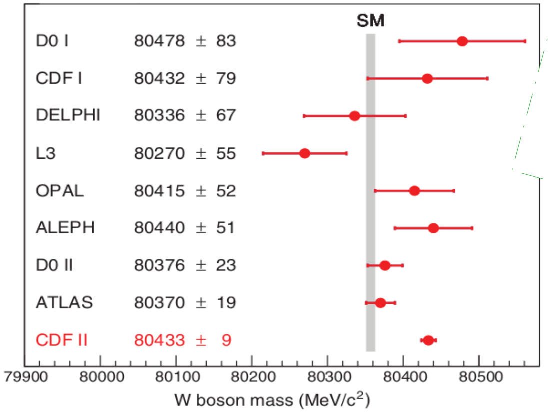
1

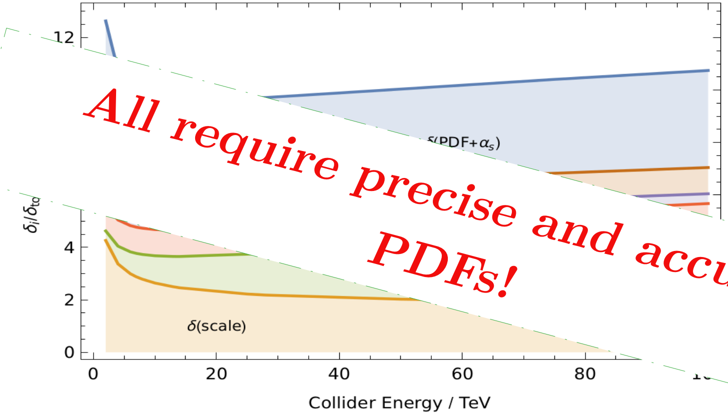
- Key input to very many calculations/measurements at colliders → Need both **accuracy and precision**. Moreover, often a dominant contribution to *uncertainty*.

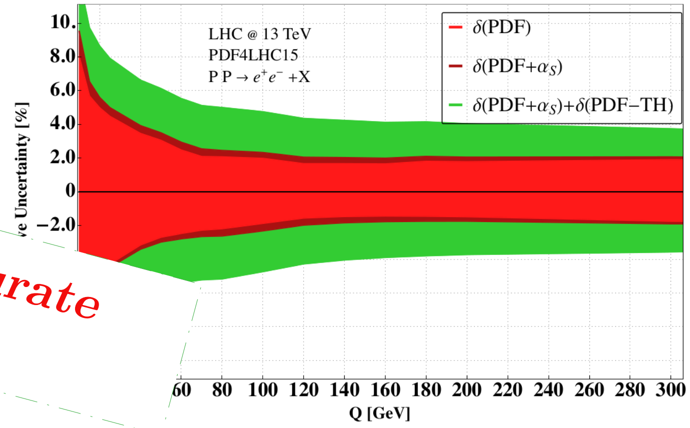
1) Precision Standard Model (SM) Measurements: M_W, $\sin^2\Theta_W$, $\alpha_S(M_Z^2)$, etc.

2) Higgs Measurements

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





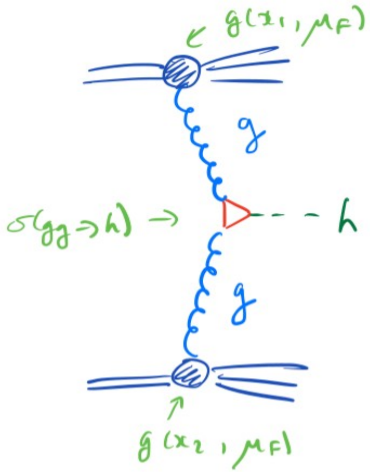


All require precise and accurate PDFs!

2. Introduction

Introduction – PDF Factorisation

2



- Consider dominant mode of Higgs production at LHC:

$$\sigma(pp \to h + X) \sim \sigma(gg \to h) \otimes g(x_1, Q^2) \otimes g(x_2, Q^2) \ ,$$

- Hadron-level cross section given in terms of:

$\sigma(gg \to h)$: parton-level cross section. $\alpha_S(m_h) \ll 1 \Rightarrow$ perturbative expansion in α_S :

$$\sigma(gg \to h) = \alpha_S(m_h)^2(\sigma_0 + \alpha_S(m_h)\sigma_1 + \cdots$$

$g(x, Q^2)$: PDF for gluon ~ probability for finding gluon with momentum fraction x at energy scale Q.

perturbative
calculable
coefficient function
$C_i^P(x, \alpha_s(Q^2))$

Schematic! See references for more detail!

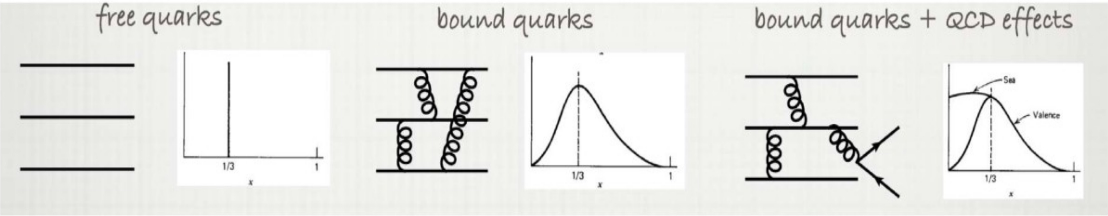
nonperturbative
incalculable
parton distribution
$f_i(x, Q^2, \alpha_s(Q^2))$

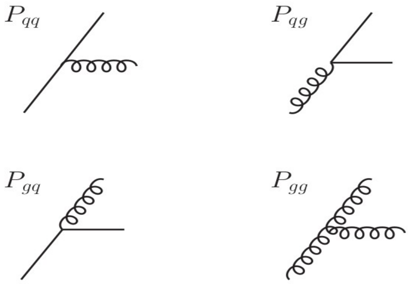
★ Parton-level cross section can be predicted with high precision, while PDFs (currently) cannot - need other way to determine!

- As strong coupling grows at low energies, PDFs are fundamentally non-perturbative, we cannot currently compute them (though progress in lattice QCD).

Introduction – PDF evolution

- What do the PDFs look like?





So QCD splittings affect PDFs → transfer momentum between partons.

- Described by Splitting functions:
- Connect partons at different x and Q^2, via DGLAP equations:

$$P_{qq} = \frac{4}{3}\left[\frac{1+x^2}{(1-x)}\right]_+ = \frac{4}{3}\left[\frac{1+x^2}{(1-x)_+}\right] + 2\delta(1-x)\ ,$$

$$P_{qg} = \frac{1}{2}\left[x^2 + (1-x)^2\right]\ ,\ P_{gq} = \frac{4}{3}\left[\frac{1+(1-x)^2}{x}\right]\ ,$$

$$P_{gg} = 6\left[\frac{1-x}{x} + x(1-x) + \frac{x}{(1-x)_+}\right] + \left[\frac{11}{2} - \frac{n_f}{3}\right]\delta(1-x),$$

DGLAP equations for PDF evolution:

$$\mu\frac{d}{d\mu}\begin{pmatrix} f_i(x,\mu) \\ f_g(x,\mu) \end{pmatrix} = \sum_j \frac{\alpha_s}{\pi}\int_x^1 \frac{d\xi}{\xi}\begin{pmatrix} P_{q_i q_j}(\frac{x}{\xi}) & P_{q_i g}(\frac{x}{\xi}) \\ P_{g q_j}(\frac{x}{\xi}) & P_{gg}(\frac{x}{\xi}) \end{pmatrix}\begin{pmatrix} f_j(\xi,\mu) \\ f_g(\xi,\mu) \end{pmatrix}$$

Can also obtain by requiring independence of structure functions of unphysical scale μ.

Introduction – PDF evolution

2

- Given gluon splitting generates quark-antiquark pair ("singlet"), only this couples to the gluon and other valence and "non-singlet" quantities drop out:

Schematic! See references for more detail!

$$g(x,Q^2)$$

$$q_S(x,Q^2) = \sum_{i=1}^{n_f}\left[q_i(x,Q^2) + \bar{q}(x,Q^2)\right]$$

Singlet

$$\frac{d}{d\log Q^2}\begin{pmatrix} q_S \\ g \end{pmatrix} = \begin{pmatrix} P_{qq} & P_{qg} \\ P_{gq} & P_{gg} \end{pmatrix} \otimes \begin{pmatrix} q_S \\ g \end{pmatrix}$$

$$q_V(x,Q^2) = \sum_{i=1}^{n_f}\left[q_i(x,Q^2) - \bar{q}(x,Q^2)\right]$$

$$q_{ij}^{\pm}(x,Q^2) = (q_i \pm \bar{q}_i) - (q_j \pm \bar{q}_j)$$

Non-Singlet

$$\frac{dq_{ij}^{\pm}}{d\log Q^2} = P_{\pm} \otimes q_{ij}^{\pm} \qquad \frac{dq_V}{d\log Q^2} = P_v \otimes q_V$$

- But, this is just the LO splittings, can expand as power series in strong coupling:

$$P_{ij}(x) = \frac{\alpha_S(\mu^2)}{4\pi} P_{ij}^{(0)}(x) + \frac{\alpha_S(\mu^2)}{4\pi}^2 P_{ij}^{(1)}(x) + \frac{\alpha_S(\mu^2)}{4\pi}^3 P_{ij}^{(2)}(x) + \ldots$$

- So whilst the form of the PDFs is non-perturbative their evolution between different (x,Q^2) is perturbative (for strong coupling in perturbative regime).

Introduction – PDF evolution

2

- Gets much more complicated at higher orders… :

  NLO (2-loop): (Curci, Furmanski, Petronzio '80):

$$P_{ij}(x) = \frac{\alpha_S(\mu^2)}{4\pi} P_{ij}^{(0)}(x) + \left(\frac{\alpha_S(\mu^2)}{4\pi}\right)^2 P_{ij}^{(1)}(x)$$

$$P^{(1)}_{\rm ps}(x) \;=\; 4\, C_F n_f \Big(\frac{20}{9}\frac{1}{x} - 2 + 6x - 4\mathrm{H}_0 + x^2\Big[\frac{8}{3}\mathrm{H}_0 - \frac{56}{9}\Big] + (1+x)\Big[5\mathrm{H}_0 - 2\mathrm{H}_{0,0}\Big]\Big)$$

$$\begin{aligned} P^{(1)}_{\rm qg}(x) \;=\;& 4\, C_A n_f \Big(\frac{20}{9}\frac{1}{x} - 2 + 25x - 2p_{\rm qg}(-x)\mathrm{H}_{-1,0} - 2p_{\rm qg}(x)\mathrm{H}_{1,1} + x^2\Big[\frac{44}{3}\mathrm{H}_0 - \frac{218}{9}\Big] \\ &+4(1-x)\Big[\mathrm{H}_{0,0} - 2\mathrm{H}_0 + x\mathrm{H}_1\Big] - 4\zeta_2 x - 6\mathrm{H}_{0,0} + 9\mathrm{H}_0\Big) + 4\, C_F n_f \Big(2p_{\rm qg}(x)\Big[\mathrm{H}_{1,0} + \mathrm{H}_{1,1} + \mathrm{H}_2 \\ &-\zeta_2\Big] + 4x^2\Big[\mathrm{H}_0 + \mathrm{H}_{0,0} + \frac{5}{2}\Big] + 2(1-x)\Big[\mathrm{H}_0 + \mathrm{H}_{0,0} - 2x\mathrm{H}_1 + \frac{29}{4}\Big] - \frac{15}{2} - \mathrm{H}_{0,0} - \frac{1}{2}\mathrm{H}_0\Big) \end{aligned}$$

$$\begin{aligned} P^{(1)}_{\rm gq}(x) \;=\;& 4\, C_A C_F \Big(\frac{1}{x} + 2p_{\rm gq}(x)\Big[\mathrm{H}_{1,0} + \mathrm{H}_{1,1} + \mathrm{H}_2 - \frac{11}{6}\mathrm{H}_1\Big] - x^2\Big[\frac{8}{3}\mathrm{H}_0 - \frac{44}{9}\Big] + 4\zeta_2 - 2 \\ &-7\mathrm{H}_0 + 2\mathrm{H}_{0,0} - 2\mathrm{H}_1 x + (1+x)\Big[2\mathrm{H}_{0,0} - 5\mathrm{H}_0 + \frac{37}{9}\Big] - 2p_{\rm gq}(-x)\mathrm{H}_{-1,0}\Big) - 4\, C_F n_f \Big(\frac{2}{3}x \\ &-p_{\rm gq}(x)\Big[\frac{2}{3}\mathrm{H}_1 - \frac{10}{9}\Big]\Big) + 4\, {C_F}^2 \Big(p_{\rm gq}(x)\Big[3\mathrm{H}_1 - 2\mathrm{H}_{1,1}\Big] + (1+x)\Big[\mathrm{H}_{0,0} - \frac{7}{2} + \frac{7}{2}\mathrm{H}_0\Big] - 3\mathrm{H}_{0,0} \\ &+1 - \frac{3}{2}\mathrm{H}_0 + 2\mathrm{H}_1 x\Big) \end{aligned}$$

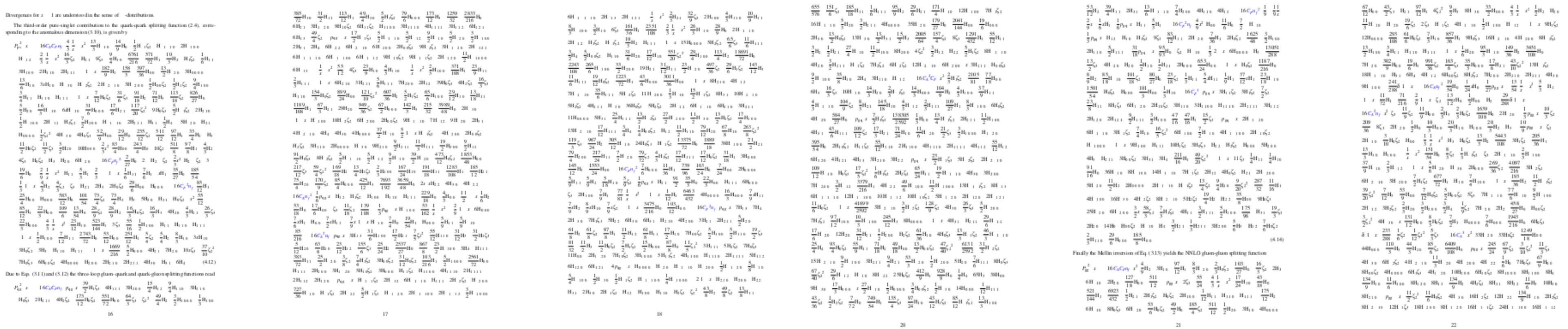
$$\begin{aligned} P^{(1)}_{\rm gg}(x) \;=\;& 4\, C_A n_f \Big(1 - x - \frac{10}{9}p_{\rm gg}(x) - \frac{13}{9}\Big(\frac{1}{x} - x^2\Big) - \frac{2}{3}(1+x)\mathrm{H}_0 - \frac{2}{3}\delta(1-x)\Big) + 4\, {C_A}^2 \Big(27 \\ &+(1+x)\Big[\frac{11}{3}\mathrm{H}_0 + 8\mathrm{H}_{0,0} - \frac{27}{2}\Big] + 2p_{\rm gg}(-x)\Big[\mathrm{H}_{0,0} - 2\mathrm{H}_{-1,0} - \zeta_2\Big] - \frac{67}{9}\Big(\frac{1}{x} - x^2\Big) - 12\mathrm{H}_0 \\ &-\frac{44}{3}x^2\mathrm{H}_0 + 2p_{\rm gg}(x)\Big[\frac{67}{18} - \zeta_2 + \mathrm{H}_{0,0} + 2\mathrm{H}_{1,0} + 2\mathrm{H}_2\Big] + \delta(1-x)\Big[\frac{8}{3} + 3\zeta_3\Big]\Big) + 4\, C_F n_f \Big(2\mathrm{H}_0 \\ &+\frac{2}{3}\frac{1}{x} + \frac{10}{3}x^2 - 12 + (1+x)\Big[4 - 5\mathrm{H}_0 - 2\mathrm{H}_{0,0}\Big] - \frac{1}{2}\delta(1-x)\Big)\ . \end{aligned}$$

Introduction – PDF evolution

- Gets much more complicated at higher orders… :

  NNLO (3-loop): (Moch, Vermaseren, Vogt ‘04):

Much recent progress on N3LO (4-loop) → see aN3LO PDFs later and references!



$$P_{ij}(x) = \frac{\alpha_S(\mu^2)}{4\pi} P_{ij}^{(0)}(x) + \frac{\alpha_S(\mu^2)}{4\pi}^2 P_{ij}^{(1)}(x) + \frac{\alpha_S(\mu^2)}{4\pi}^2 P_{ij}^{(2)}(x)$$

Introduction – PDF evolution

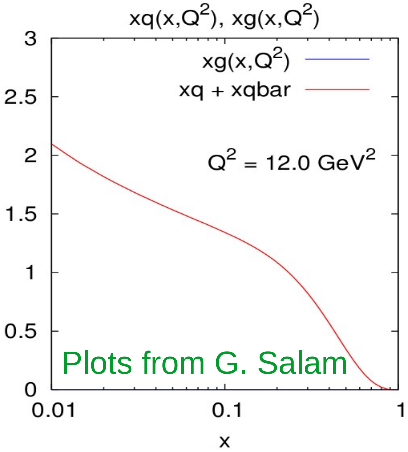
- How does DGLAP affect PDFs?
- P_{gg}, P_{gq} diverge at small x, means PDFs evolve to lower x with increasing Q^2.
- Can visualise this with LO example, take just quarks:

$$P_{qq} = \frac{4}{3}\left[\frac{1+x^2}{(1-x)}\right]_+ = \frac{4}{3}\left[\frac{1+x^2}{(1-x)_+}\right] + 2\delta(1-x)\,,$$

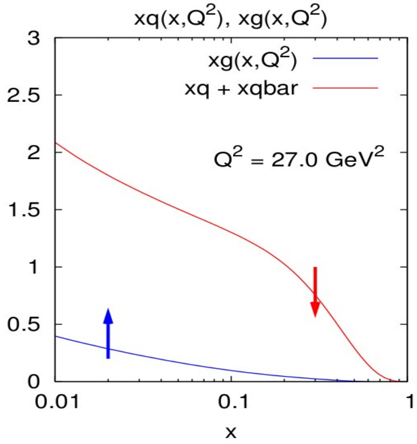
$$P_{qg} = \frac{1}{2}\left[x^2+(1-x)^2\right]\,,\; P_{gq} = \frac{4}{3}\left[\frac{1+(1-x)^2}{x}\right]$$

$$P_{gg} = 6\left[\frac{1-x}{x} + x(1-x) + \frac{x}{(1-x)_+}\right] + \left[\frac{11}{2} - \frac{n_f}{3}\right]\delta(1-x),$$

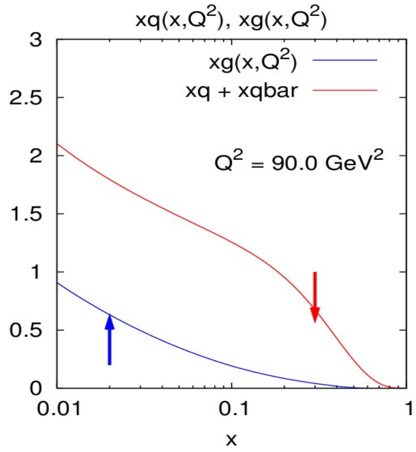
Proton structure depends on scale ("resolution") it's probed at!

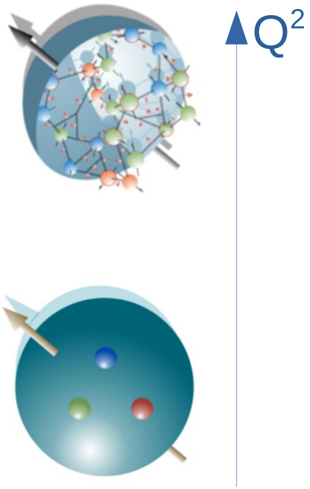


EVOLVE!



EVOLVE!





Introduction – PDF Key Properties ②

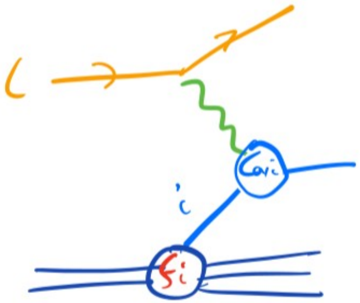
★ **Key facts:**

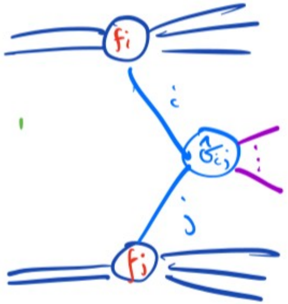
- PDFs satisfy (DGLAP) **evolution**. Predicts PDFs at higher scale from PDFs at lower scale.

$$\frac{\partial q(x,\mu)}{\partial \mu} = P_{qq} \otimes q(x,\mu) + P_{qg} \otimes g(x,\mu)$$
$$\frac{\partial g(x,\mu)}{\partial \mu} = P_{gq} \otimes q(x,\mu) + P_{gg} \otimes g(x,\mu)$$

$$f(x, Q_i^2) \rightarrow f(x, Q_f^2)$$

- QCD **factorization**: perturbative physics separated from **universal** non-perturbative PDFs





Schematic! See references for more detail!

$$\text{Factorization} \Rightarrow f_i^{\text{DIS}}(x, Q^2) \equiv \{f_i^{\text{Collider}}(x, Q^2)\} \leftarrow$$ **Drell Yan, Jets, Higgs...**

- Foundation of **global PDF fits**: use wide range of data at different scales and from different processes to extract PDFs.

PDFs are **universal!**

Introduction – PDF Fitting

- PDF Universality + known evolution → Determine PDFs by fitting to data.
- How does this work?

$$\text{Data} = \text{PDF} \otimes \sigma_H$$

measure ↗ **fit** ↑ **predict** ↖

but this is non-trivial **inverse problem**. No unique answer. Many ingredients enter!

Parameterise PDFs at low scale Q_0

↓

Select data & theory

↓

Perform fit

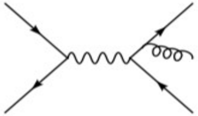
↓

Output PDFs

$f_i(x, Q_0)$: $A_f x^{a_f}(1-x)^{b_f} \times$ ⟶ $\sum_{i=1}^{n} \alpha_{f,i} P_i(y(x))$, **CT, MSHT…**

↘ $\mathrm{NN}_i(x)$ **NNPDF**

Perturbative QCD for parton-level theory



$+ \ldots$

Minimise $\chi^2(\{a\},\{\lambda\}) = \sum_{k=1}^{N_{pt}} \frac{1}{s_k^2}\left(D_k - T_k - \sum_{\alpha=1}^{N_\lambda} \beta_{k,\alpha}\lambda_\alpha\right)^2 + \sum_{\alpha=1}^{N_\lambda} \lambda_\alpha^2$,

DGLAP: $f(x, Q_0) \rightarrow f(x, \mu_{\text{data}})$

$$f(x, \mu) \pm \Delta(x, \mu)$$

3. Current PDF Landscape

Global PDF Fitting

3

2) Experiment

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

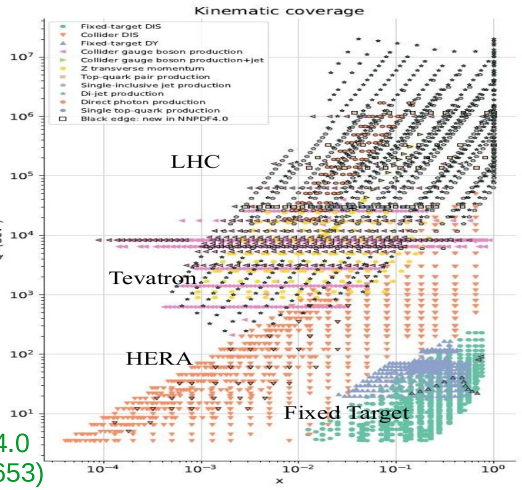
1) Methodology

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

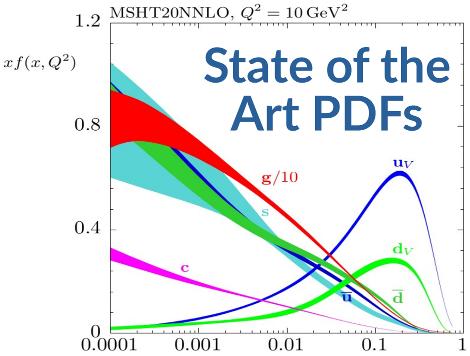
Global PDF fit

3) Theory

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



NNPDF4.0
(2109.02653)



Several Global PDF Fitting Groups

3

- Several different PDF analysis groups – ABM, ATLASPDF, CJ, CT, HERAPDF, JAM, MSHT, NNPDF and others. Not covered all here, naturally more MSHT examples.

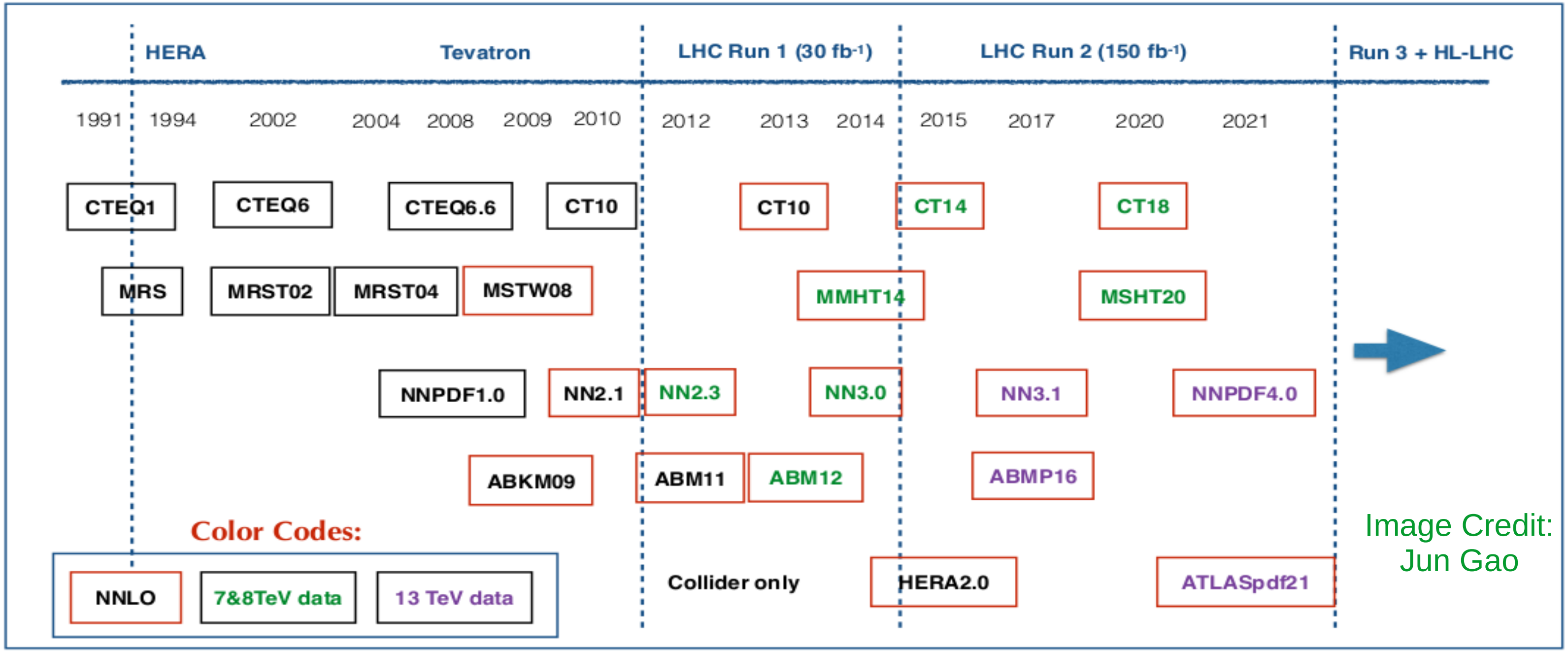


Image Credit: Jun Gao

Default now

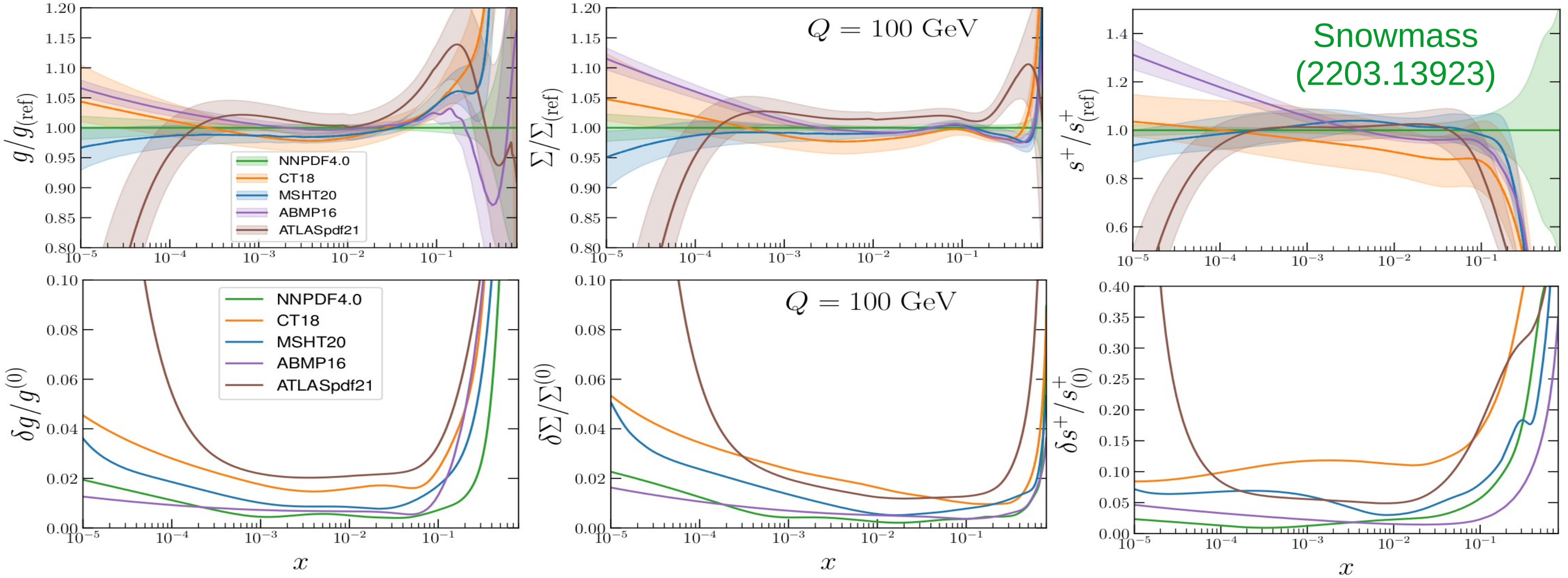
NNLO QCD + NLO EW and latest LHC and other data

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

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

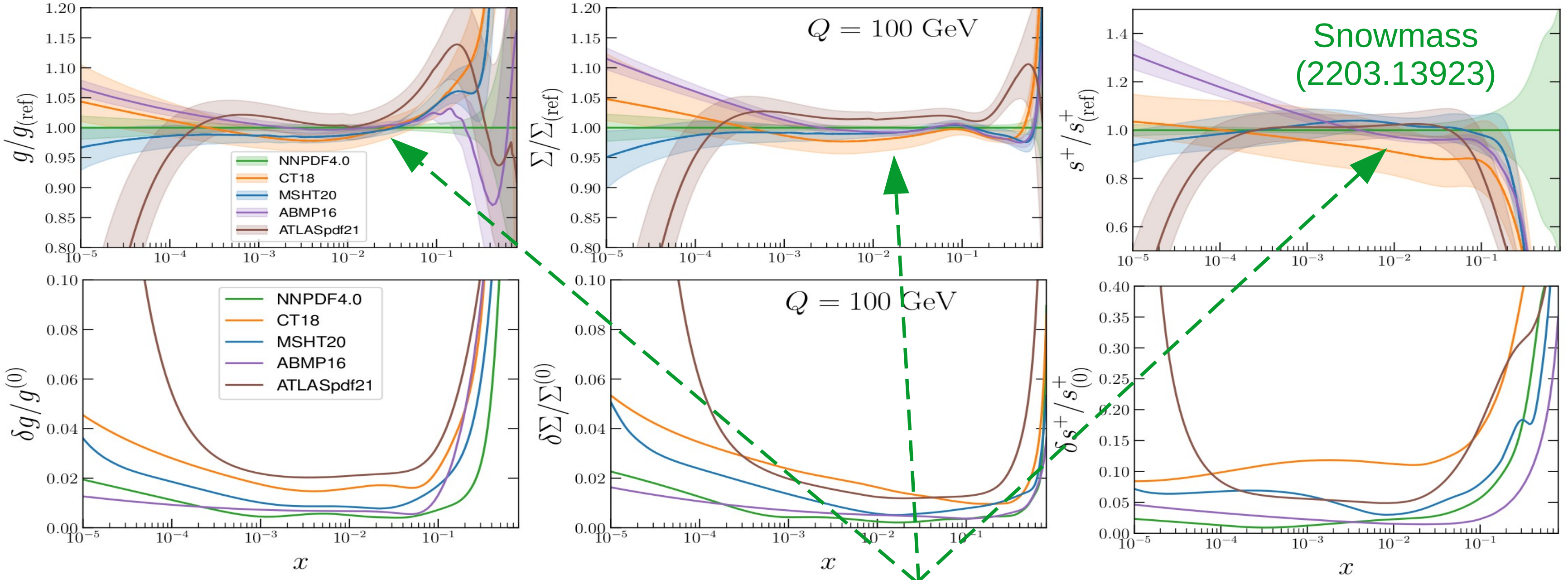
PDF Comparison

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



PDF Comparison

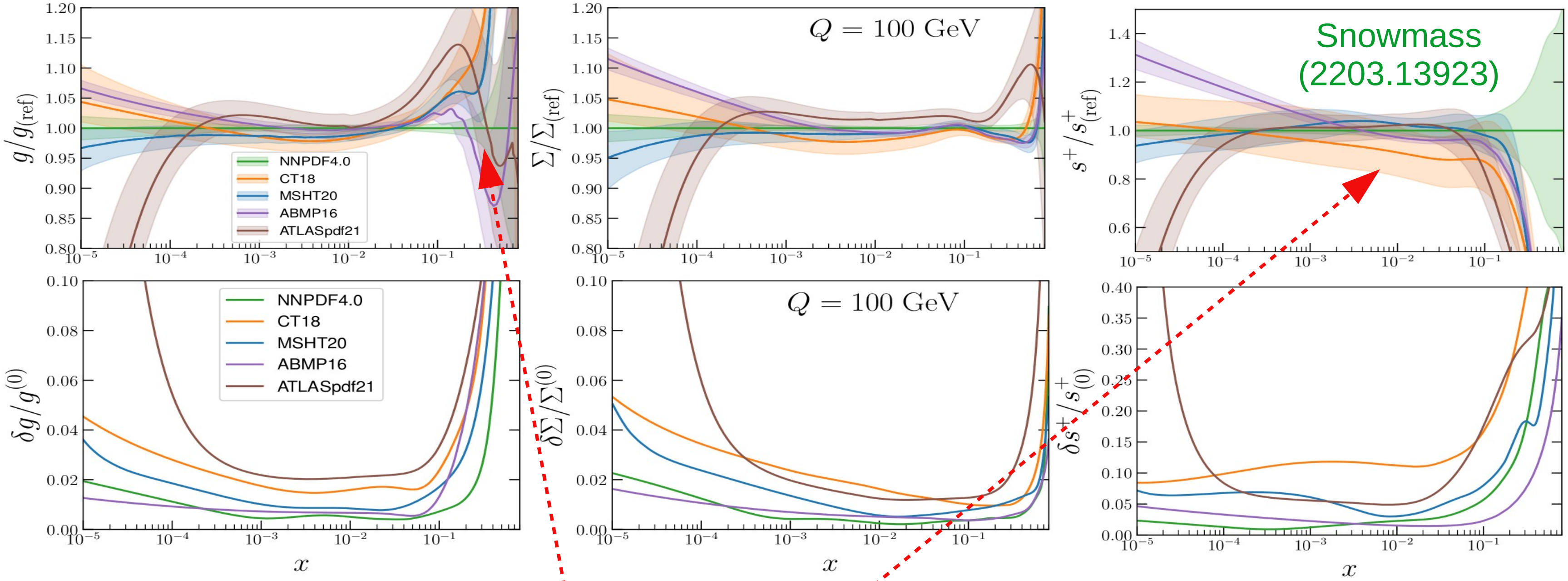
- Compare several of these at the level of the PDFs and uncertainties:



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

PDF Comparison

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

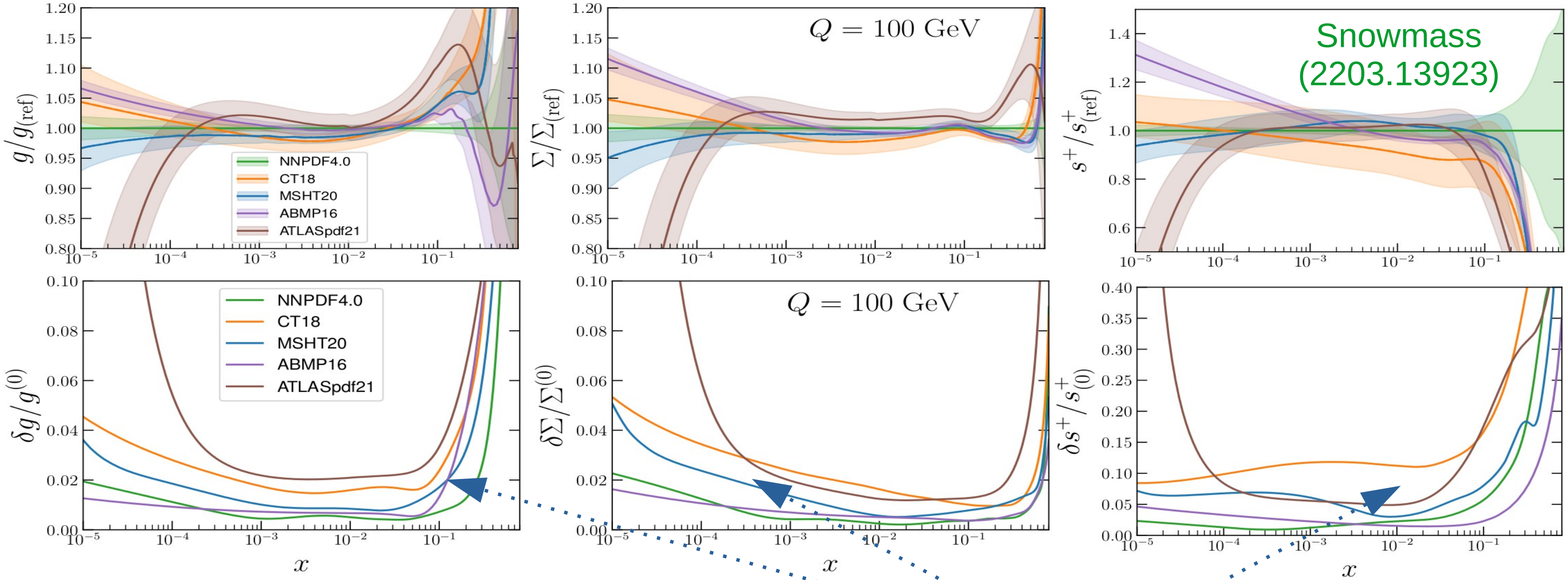


- General agreement over data range ($10^{-4} < x < 10^{-1}$) within uncertainties.
- Differences exist (high x gluon, strangeness, …).

PDF Comparison

3

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

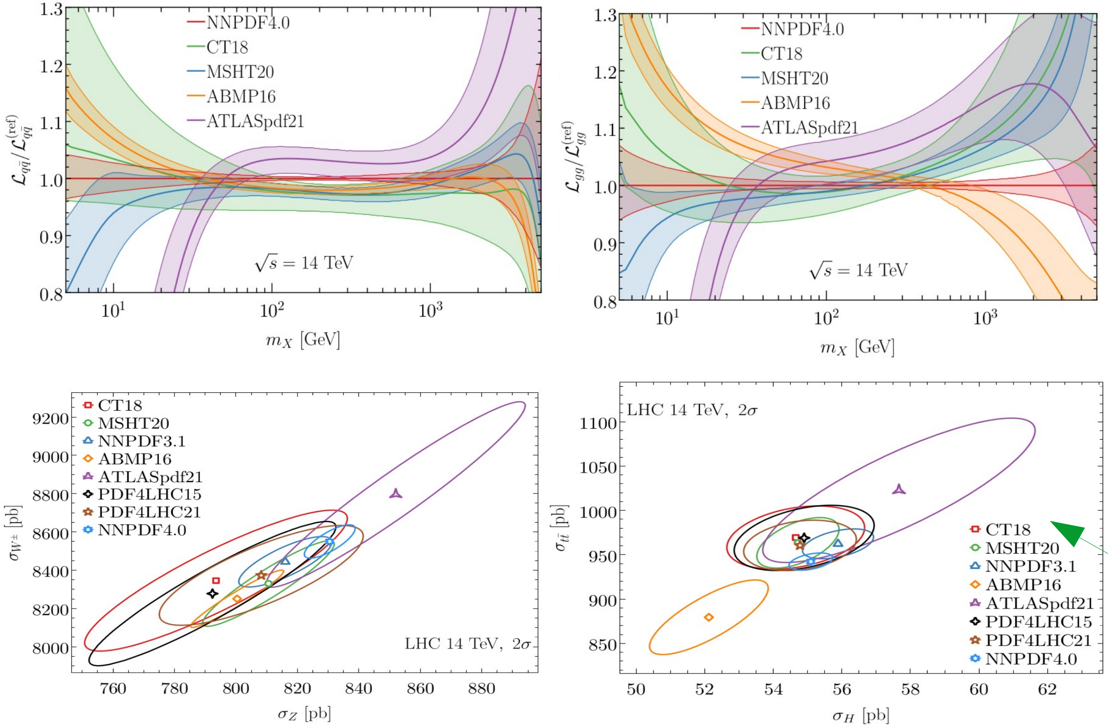


- General agreement over data range ($10^{-4} < x < 10^{-1}$) within uncertainties.
- Differences exist (high x gluon, strangeness, uncertainty sizes).

Parton luminosities and Xsecs

3

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



Snowmass (2203.13923)

Useful way to view PDFs is in terms of PDF luminosity:

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

$$\sigma = \sum_{a,b=q,\bar{q},g} \int_{M^2}^{s} \frac{d\hat{s}}{\hat{s}} \mathcal{L}_{ab}(\hat{s}, \mu_F^2)\, \hat{s}\hat{\sigma}_{ab}(\hat{s}, M^2, \mu_R^2, \mu_F^2)$$

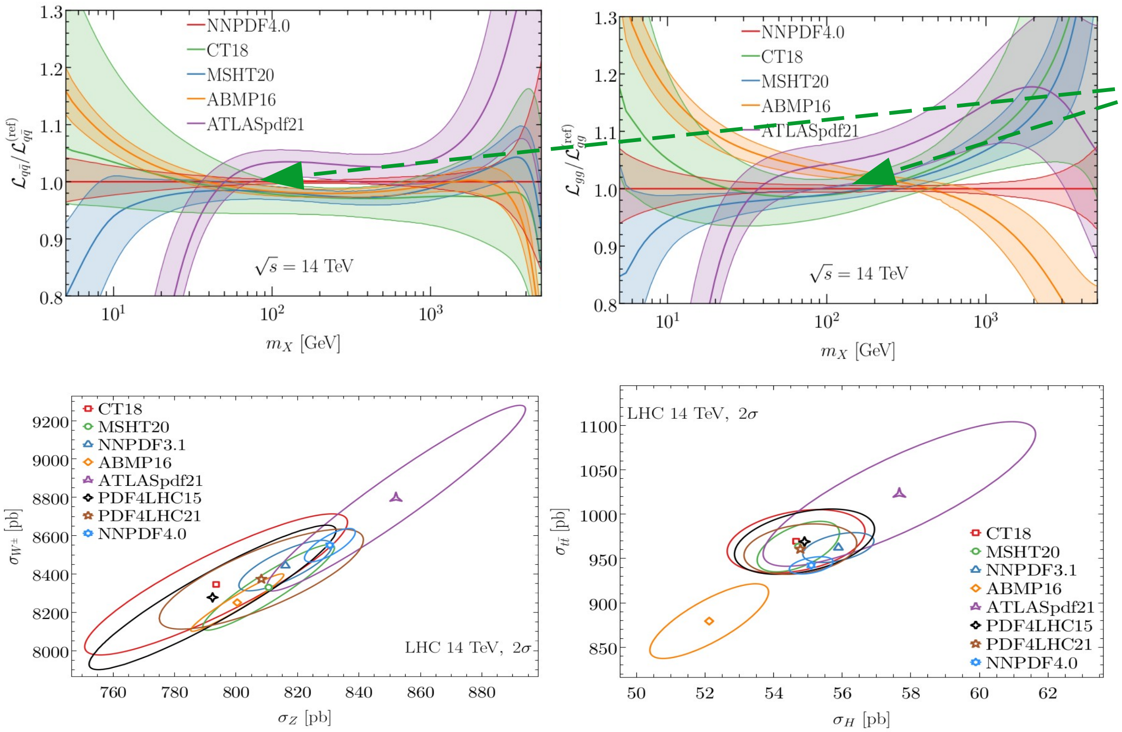
$$\mathcal{L}_{ab}(\tau, \mu_F^2) = \frac{1}{s} \int_{\tau/s}^{1} \frac{dx}{x} f_a(\tau/sx, \mu_F^2)\, f_b(x, \mu_F^2)$$

Bottom plots show cross-sections (central value and uncertainty for different PDF sets) and their correlations

Parton luminosities and Xsecs

3

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

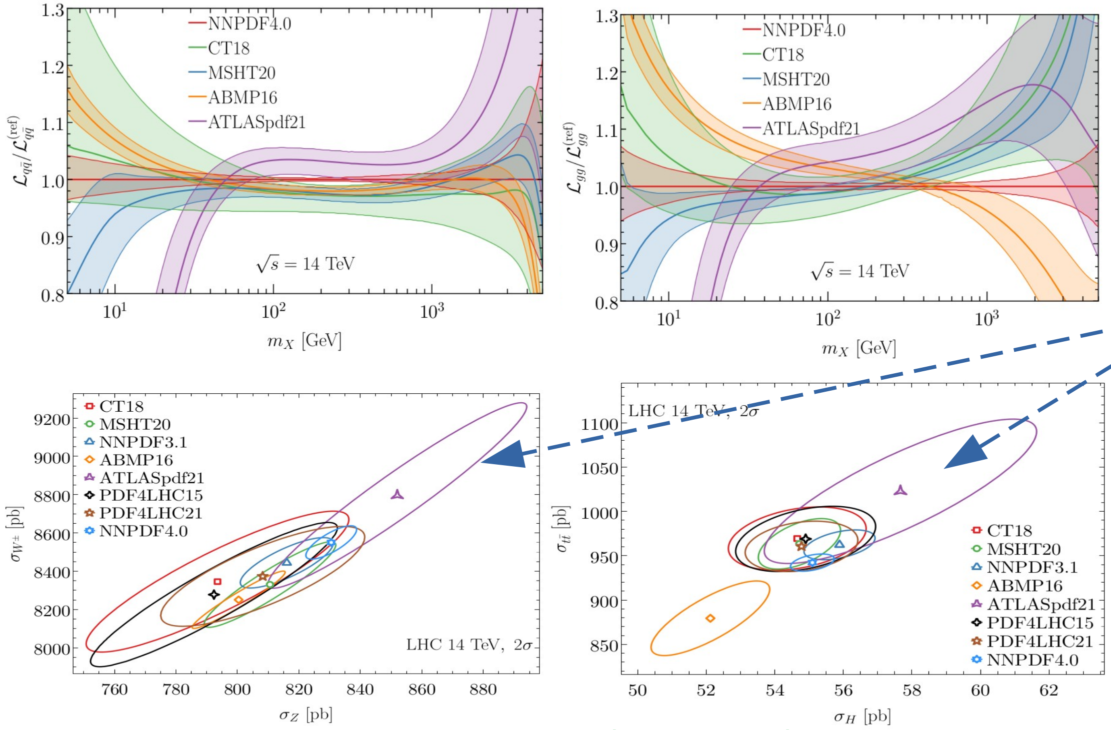


Snowmass (2203.13923)

- 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:



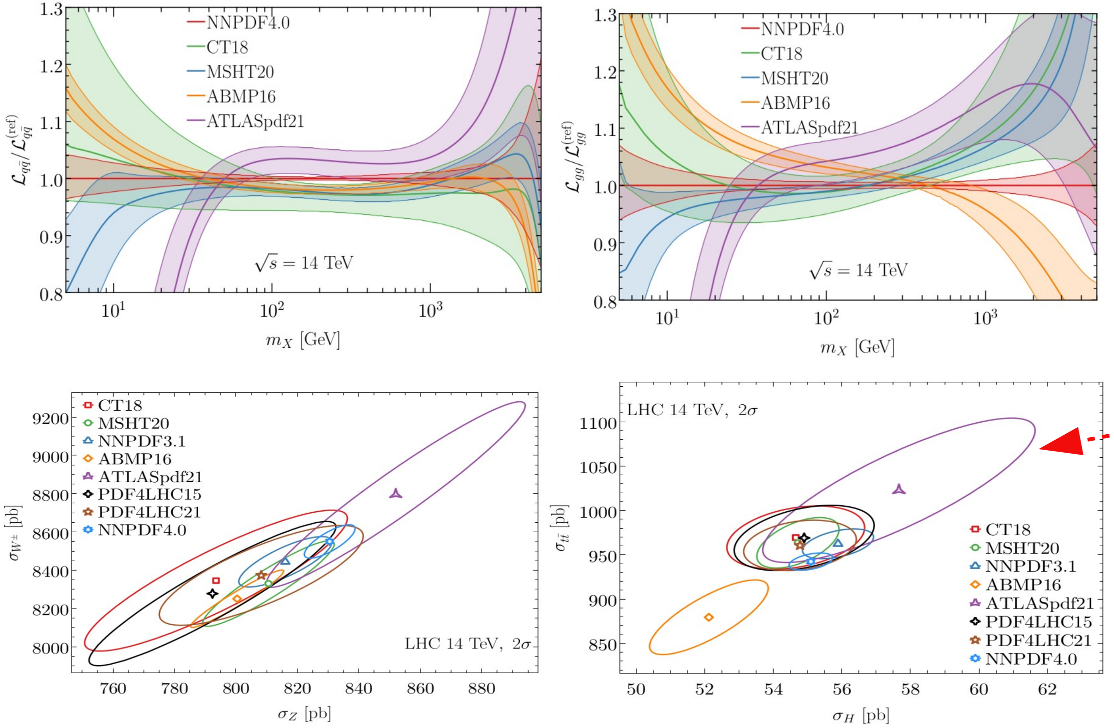
Snowmass (2203.13923)

- 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

3

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

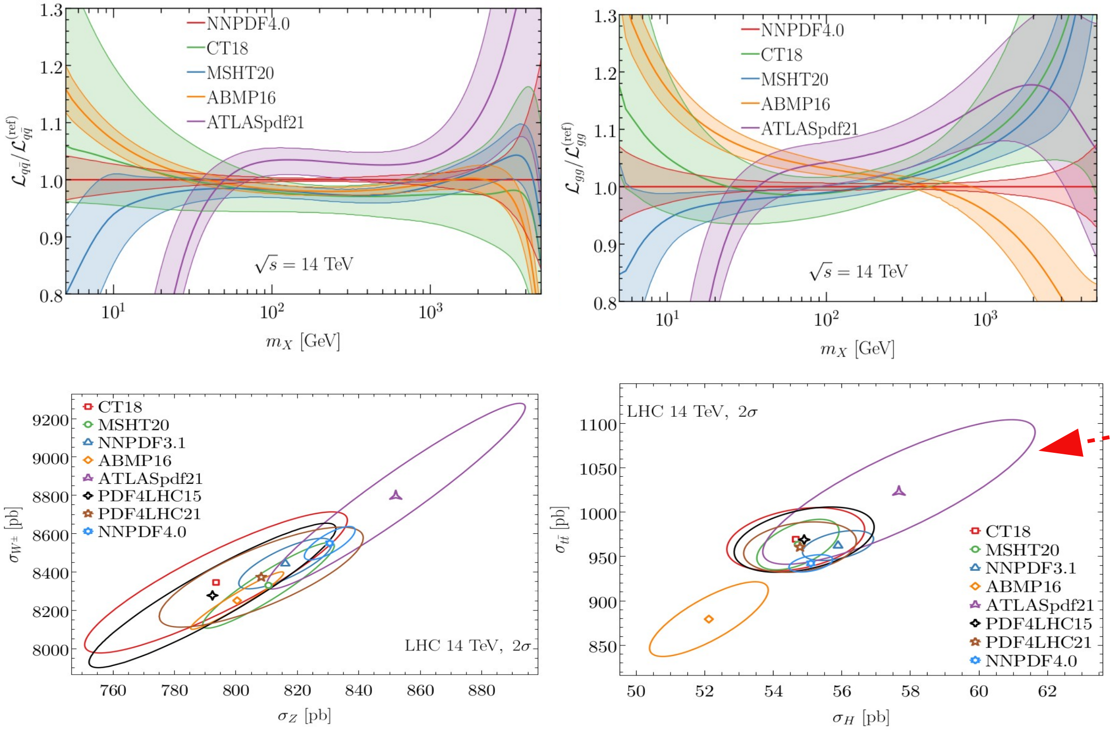


Snowmass (2203.13923)

- 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

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

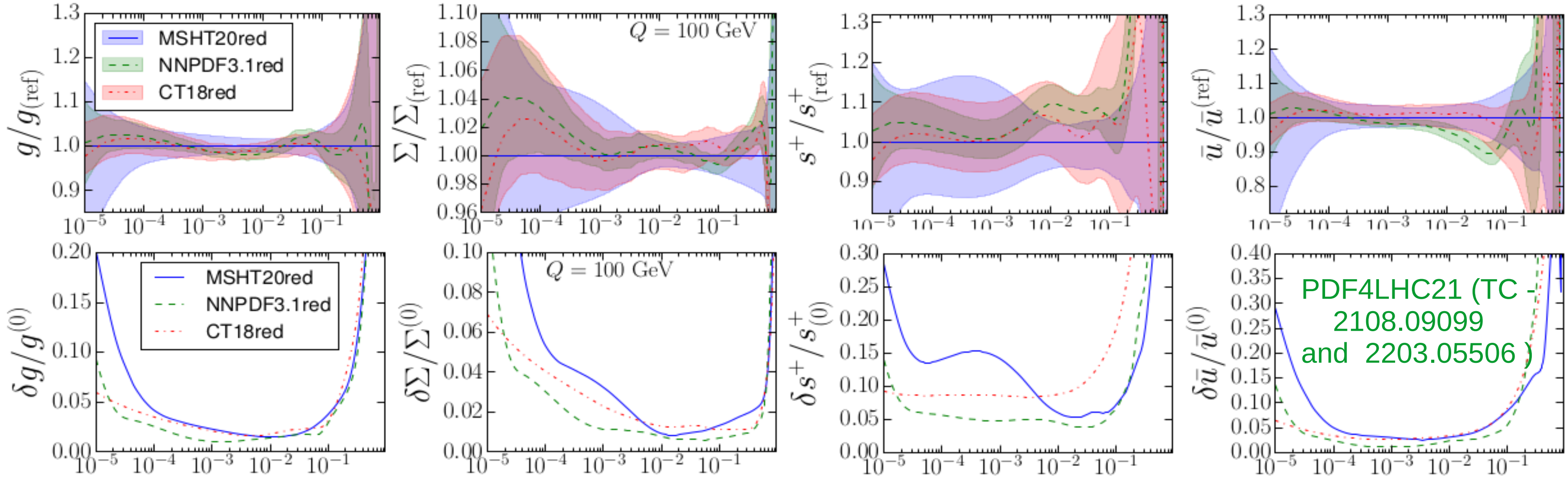


Snowmass (2203.13923)

- 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.
- Nonetheless we have the *most precise and accurate PDFs* yet.

PDF4LHC21 Benchmarking

- Observe differences in central values and uncertainties between groups → is it down to data included, methodology applied, theory settings or other differences?
- Take "Reduced Fit" PDFs, using common data and "same" theory (where possible):

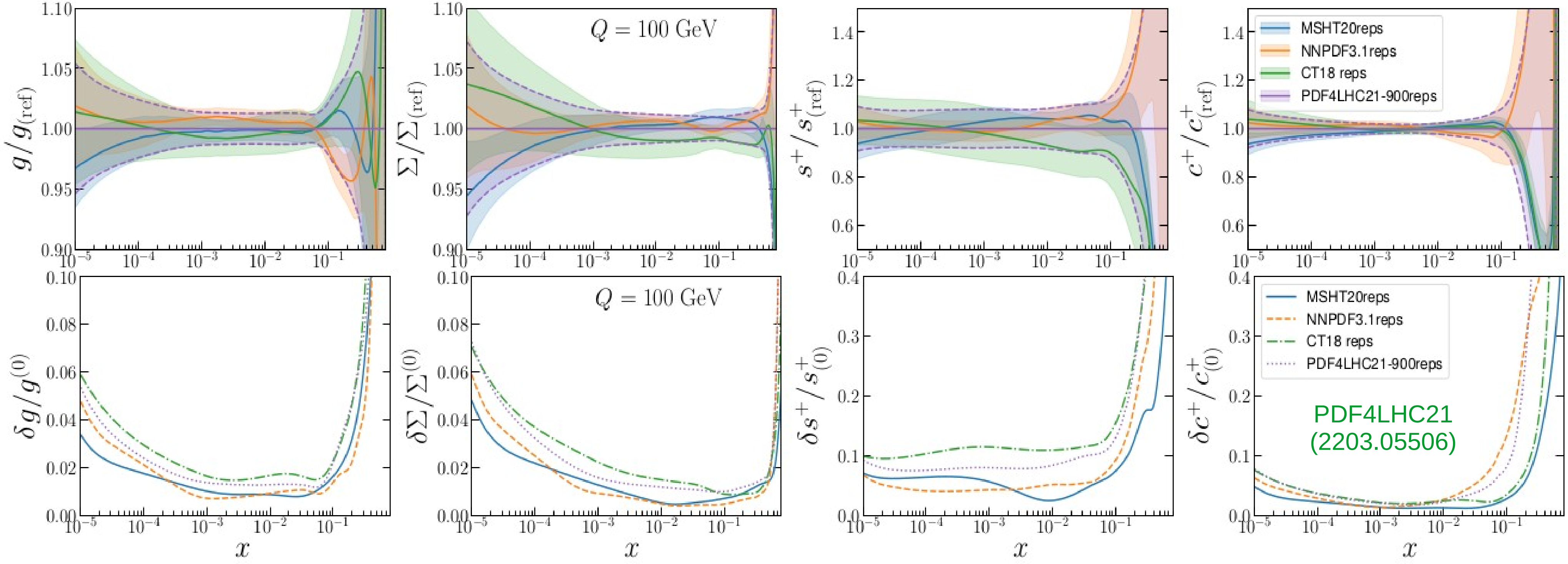


N.B. Smallest Uncertainty not necessarily most precise and accurate PDF.

- Common settings → consistent PDFs! Central values in agreement.
- Uncertainties still differ, reflecting underlying *methodological differences.*

PDF4LHC21

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



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

PDF4LHC21

3

- PDF4LHC21 combination of MSHT20, CT18', NNPDF3.1' global PDF sets.
- Provided in several forms – e.g. both Hessian and MC replica.
- Central values and uncertainties of all 3 PDFs reflected.

When to use? *Use your judgment*, but generic recommendations:

- Comparison between data and theory for SM measurements → recommend to use individual global fit group PDFs (and several of them).

- Search for BSM phenomena or measurements of lower precision SM observables → May use PDF4LHC21.

- Theoretical computations → May use PDF4LHC21.

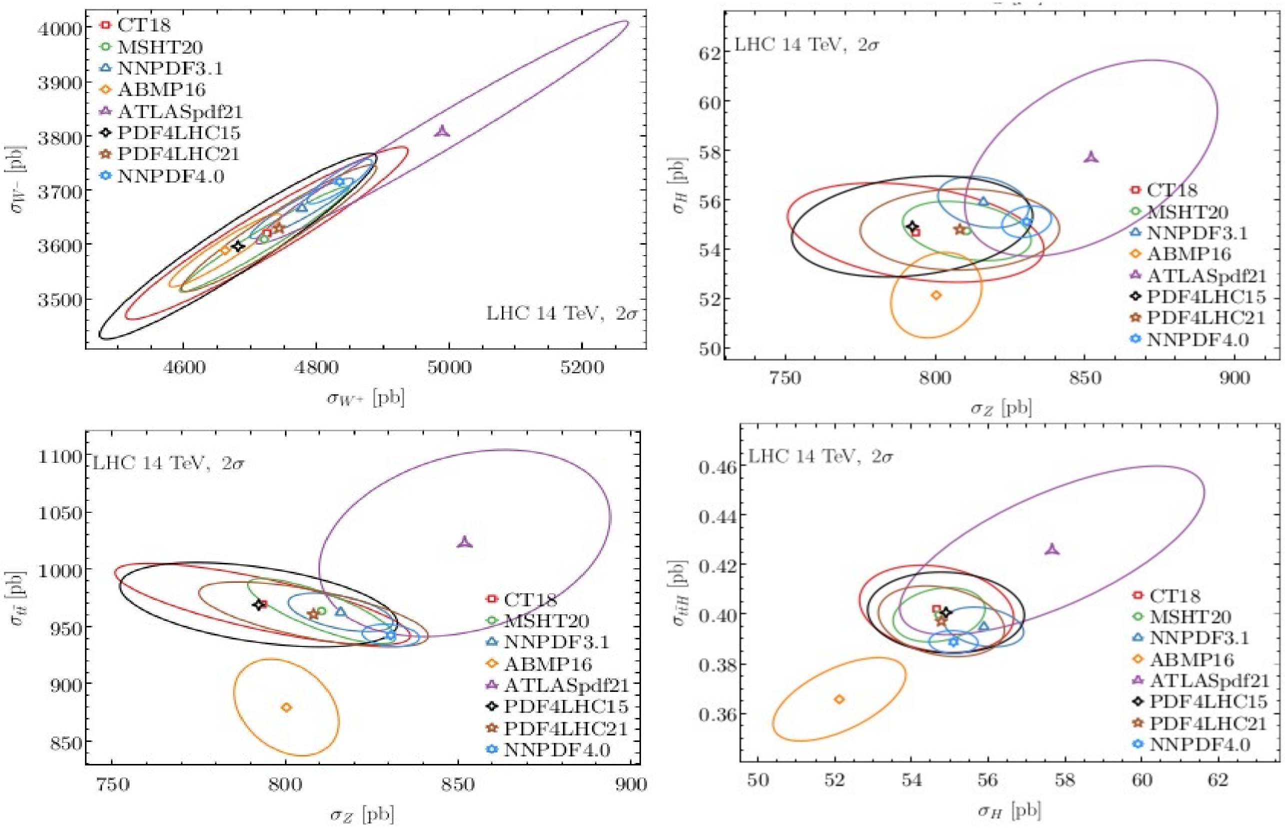
See PDF4LHC21 (2203.05506) for more info!

- Key point → PDF4LHC21 is a useful extra PDF set and doesn't preclude the use of individual global fit PDFs.

→ If large discrepancies observed, advise to use range of individual group PDF sets.

PDF Luminosity and Xsecs

- How do PDFs compare with each other and PDF4LHC21 for total cross-sections:



- PDF4LHC21 uncertainty encompasses central values of CT18, MSHT20, NNPDF3.1 here.
- Some PDFs (ABMP, ATLASpdf21) can differ notably, different settings and input data.

See PDF4LHC21 (2203.05506) for more info!
Plots from 2203.13923

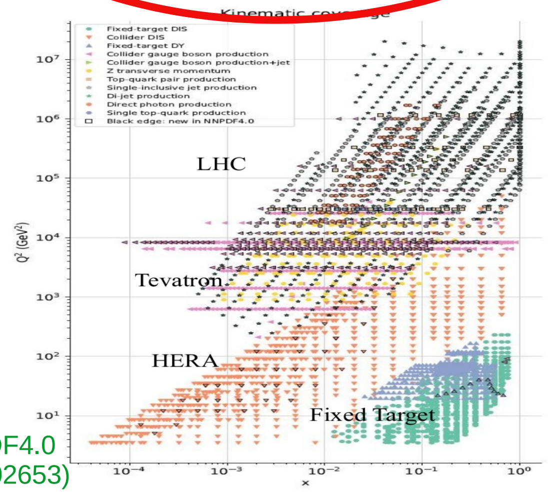
4. PDF Data + Constraints

Global PDF Fitting

4

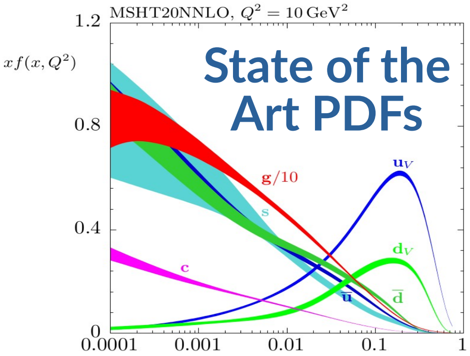
1) Experiment

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



NNPDF4.0 (2109.02653)

Global PDF fit



2) Methodology

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

3) Theory

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

PDF Constraints - What data?

4

- Must therefore constrain the PDFs with fits to data:
- Huge amount of data in global PDF fits – CT, MSHT, NNPDF.
- Typically $\geq$ 5000 datapoints.
- Fixed target DIS, collider DIS, Drell-Yan, jets, top from pp (or $p\bar{p}$) colliders and more!.

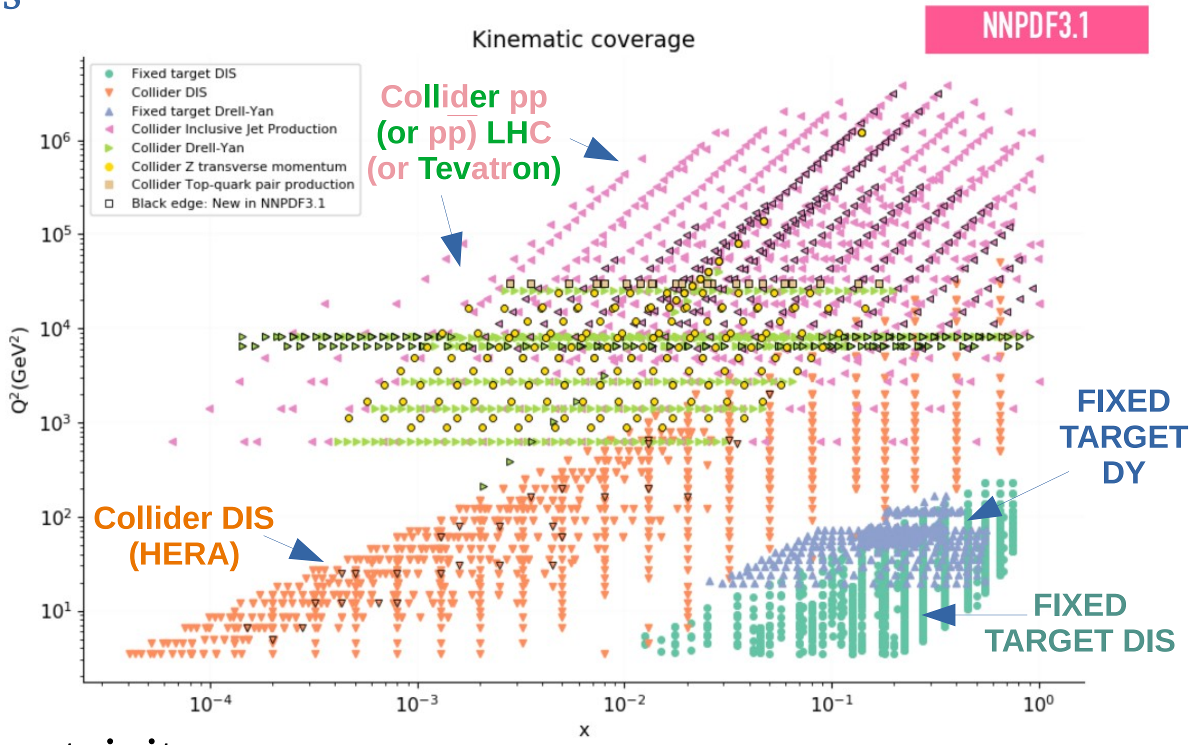
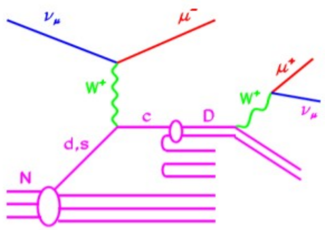


Figure adapted from NNPDF3.1 (1706.00428)

- Constrains central values and uncertainites.

PDF Constraints - What data?

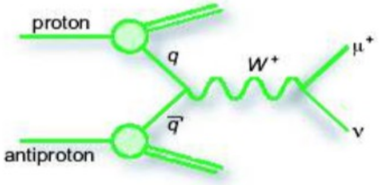
4

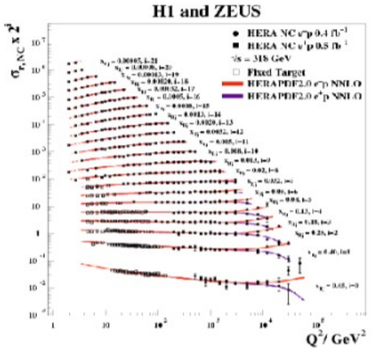


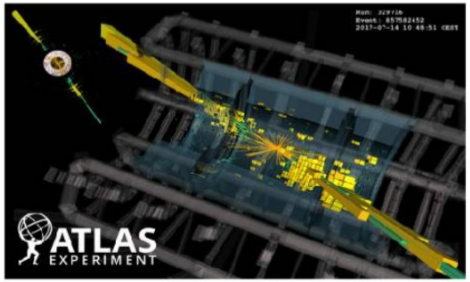
Global PDF fits: datasets

$$\text{Data} = \text{PDF} \otimes \sigma_H$$

	Process	Subprocess	Partons	x range
Fixed Target	$\ell^\pm\{p,n\} \to \ell^\pm + X$	$\gamma^* q \to q$	$q, \bar{q}, g$	$x \gtrsim 0.01$
	$\ell^\pm n/p \to \ell^\pm + X$	$\gamma^* d/u \to d/u$	d/u	$x \gtrsim 0.01$
	$pp \to \mu^+\mu^- + X$	$u\bar{u}, d\bar{d} \to \gamma^*$	$\bar{q}$	$0.015 \lesssim x \lesssim 0.35$
	$pn/pp \to \mu^+\mu^- + X$	$(u\bar{d})/(u\bar{u}) \to \gamma^*$	$\bar{d}/\bar{u}$	$0.015 \lesssim x \lesssim 0.35$
	$\nu(\bar{\nu})N \to \mu^-(\mu^+) + X$	$W^* q \to q'$	$q, \bar{q}$	$0.01 \lesssim x \lesssim 0.5$
	$\nu N \to \mu^-\mu^+ + X$	$W^* s \to c$	s	$0.01 \lesssim x \lesssim 0.2$
	$\bar{\nu} N \to \mu^+\mu^- + X$	$W^* \bar{s} \to \bar{c}$	$\bar{s}$	$0.01 \lesssim x \lesssim 0.2$
Collider DIS	$e^\pm p \to e^\pm + X$	$\gamma^* q \to q$	$g, q, \bar{q}$	$0.0001 \lesssim x \lesssim 0.1$
	$e^+ p \to \bar{\nu} + X$	$W^+\{d,s\} \to \{u,c\}$	d, s	$x \gtrsim 0.01$
	$e^\pm p \to e^\pm c\bar{c} + X$	$\gamma^* c \to c, \gamma^* g \to c\bar{c}$	c, g	$10^{-4} \lesssim x \lesssim 0.01$
	$e^\pm p \to e^\pm b\bar{b} + X$	$\gamma^* b \to b, \gamma^* g \to b\bar{b}$	b, g	$10^{-4} \lesssim x \lesssim 0.01$
	$e^\pm p \to \text{jet} + X$	$\gamma^* g \to q\bar{q}$	g	$0.01 \lesssim x \lesssim 0.1$
Tevatron	$p\bar{p} \to \text{jet} + X$	$gg, qg, qq \to 2j$	g, q	$0.01 \lesssim x \lesssim 0.5$
	$p\bar{p} \to (W^\pm \to \ell^\pm\nu) + X$	$ud \to W^+, \bar{u}\bar{d} \to W^-$	$u, d, \bar{u}, \bar{d}$	$x \gtrsim 0.05$
	$p\bar{p} \to (Z \to \ell^+\ell^-) + X$	$uu, dd \to Z$	u, d	$x \gtrsim 0.05$
	$p\bar{p} \to t\bar{t} + X$	$qq \to t\bar{t}$	q	$x \gtrsim 0.1$
LHC	$pp \to \text{jet} + X$	$gg, qg, q\bar{q} \to 2j$	g, q	$0.001 \lesssim x \lesssim 0.5$
	$pp \to (W^\pm \to \ell^\pm\nu) + X$	$u\bar{d} \to W^+, d\bar{u} \to W^-$	$u, d, \bar{u}, \bar{d}, g$	$x \gtrsim 10^{-3}$
	$pp \to (Z \to \ell^+\ell^-) + X$	$q\bar{q} \to Z$	$q, \bar{q}, g$	$x \gtrsim 10^{-3}$
	$pp \to (Z \to \ell^+\ell^-) + X, p_\perp$	$gq(\bar{q}) \to Zq(\bar{q})$	$g, q, \bar{q}$	$x \gtrsim 0.01$
	$pp \to (\gamma^* \to \ell^+\ell^-) + X$, Low mass	$q\bar{q} \to \gamma^*$	$q, \bar{q}, g$	$x \gtrsim 10^{-4}$
	$pp \to (\gamma^* \to \ell^+\ell^-) + X$, High mass	$q\bar{q} \to \gamma^*$	$\bar{q}$	$x \gtrsim 0.1$
	$pp \to W^+\bar{c}, W^-c$	$sg \to W^+c, \bar{s}g \to W^-\bar{c}$	$s, \bar{s}$	$x \sim 0.01$
	$pp \to t\bar{t} + X$	$gg \to t\bar{t}$	g	$x \gtrsim 0.01$
	$pp \to D, B + X$	$gg \to c\bar{c}, b\bar{b}$	g	$x \gtrsim 10^{-6}, 10^{-5}$
	$pp \to J/\psi, \Upsilon + pp$	$\gamma^*(gg) \to c\bar{c}, b\bar{b}$	g	$x \gtrsim 10^{-6}, 10^{-5}$
	$pp \to \gamma + X$	$gq(\bar{q}) \to \gamma q(\bar{q})$	g	$x \gtrsim 0.005$







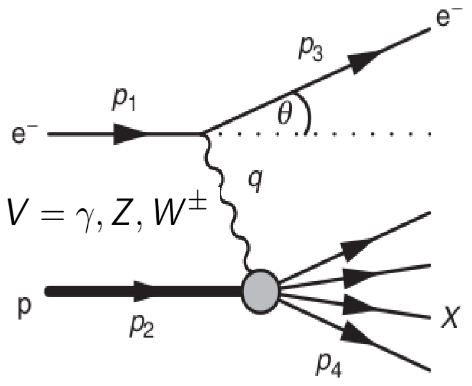
- From fixed target, to HERA DIS and collider LHC data.

$N_{\text{dataset}} \sim 50 - 60 \qquad N_{\text{pts}} \sim 4000 - 5000$

PDF Constraints (DIS reminder)

- Compute in Parton model – it's just eq scattering weighted by chance of interacting with each quark in proton (assume $Q^2 \ll M_Z^2$):

$$\frac{d^2\sigma}{dxdQ^2} = \frac{8\pi\alpha^2}{Q^4}[1+(1-y)^2]\sum_i e_q^2 q_i(x)$$



- Alternatively write in terms of leptonic (EW physics) and hadronic (non-perturbative) tensors and use symmetries to write in terms of "Structure Functions", F_i.

$$\frac{d^2\sigma}{dxdQ^2} = \frac{4\pi\alpha^2}{Q^4}\left((1-y)F_2(x,Q^2)/x + y^2F_1(x,Q^2)\right)$$

x = parton momentum fraction
Y = inelasticity
Q = momentum transfer

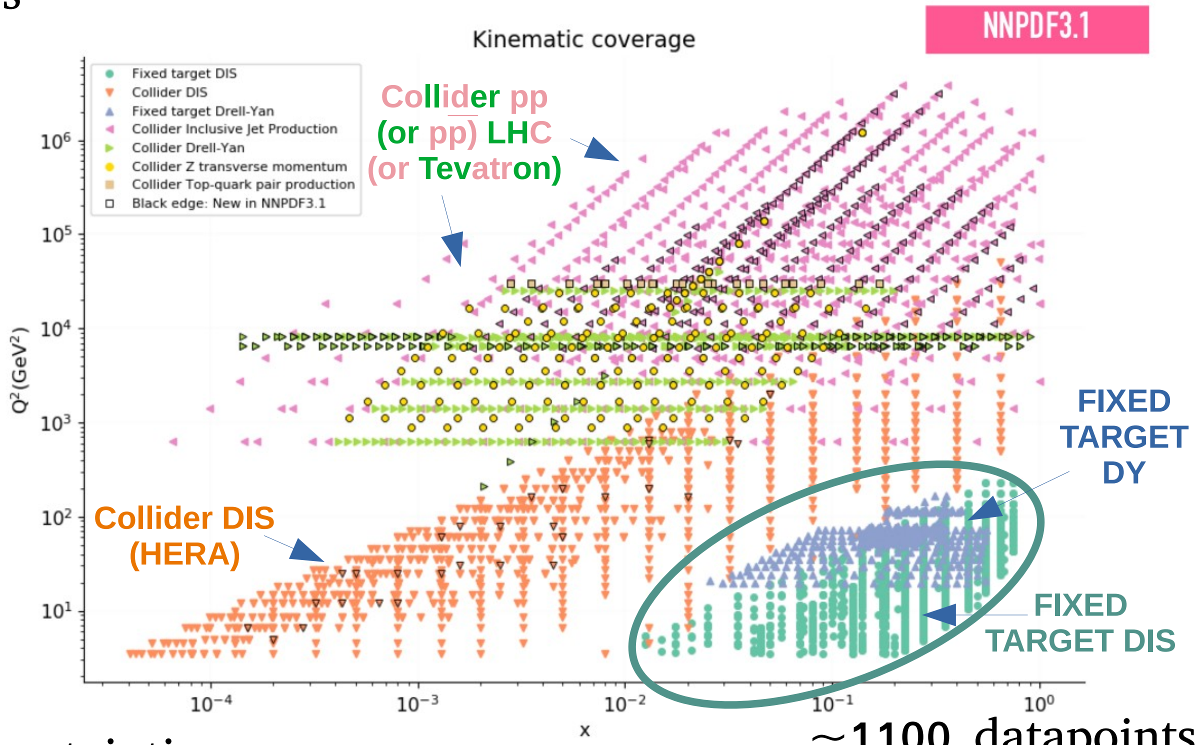
- Callan Gross relation and Bjorken scaling:

$$F_2(x,Q^2) = 2xF_1(x,Q^2) = x\sum_i e_q^2 q_i(x)$$

PDF Constraints - Fixed Target DIS

4

- Must therefore constrain the PDFs with fits to data:
- Huge amount of data in global PDF fits – CT, MSHT, NNPDF.
- More than **4000** datapoints.
- Fixed target DIS, collider DIS, Drell-Yan, jets, top from pp (or $p\bar{p}$) colliders.
- Constrains central values and uncertainties.



$\sim$**1100** datapoints

Figure adapted from NNPDF3.1 (1706.00428)

PDF Constraints - Fixed Target DIS

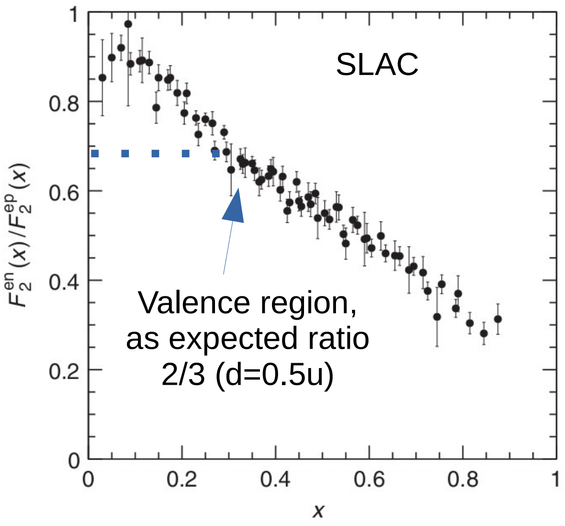
4

- So we have from parton model (LO) (and isospin symmetry: $F_2^{en} = F_2^{ep}(u \leftrightarrow d)$):

$$F_2^{ep} = x\left(\frac{4}{9}u(x) + \frac{1}{9}d(x) + \frac{1}{9}s(x) + ... + \frac{4}{9}\bar{u}(x) + \frac{1}{9}\bar{d}(x) + \frac{1}{9}\bar{s}(x) + ...\right)$$

- Structure function sensitive to charge-weighted sum of quarks.
- Measure at high x in fixed target experiments (BCDMS, SLAC, NMC, etc) → high x quarks!
- If also measure on neutrons then you get different charge-weighted sum → break some degeneracy in flavour decomposition (u vs d).
- Ratio tells us sea quarks (S(x)) dominate at low x, and dv/uv ratio → 0 (exclusion principle?) at high x as:

$$\frac{F_2^{en}(x)}{F_2^{ep}(x)} = \frac{4d_v(x) + u_v(x) + aS(x)}{4u_v(x) + d_v(x) + aS(x)} \to 1, 0.25 \text{ as } x \to 0, 1.$$



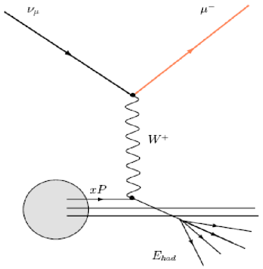
PDF Constraints - Fixed Target DIS

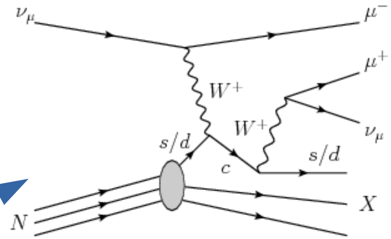
4

- This was Neutral Current DIS, also Charged Current – neutrino scattering
- Mediated by W bosons, therefore sensitive to differently-weighted combinations of quarks.

$$
\begin{aligned}
F_2^{\nu} &= 2x[d + s + \bar{u} + \bar{c}] \\
F_2^{\bar{\nu}} &= 2x[u + c + \bar{d} + \bar{s}] \\
xF_3^{\nu} &= 2x[d + s - \bar{u} - \bar{c}] \\
xF_3^{\bar{\nu}} &= 2x[u + c - \bar{d} - \bar{s}].
\end{aligned}
$$

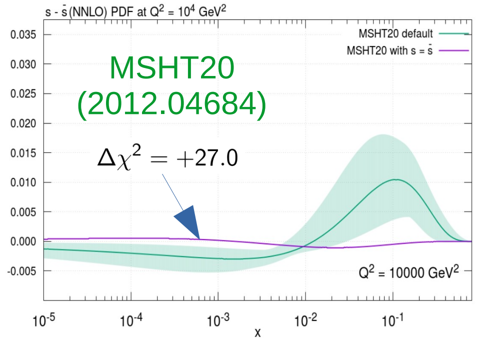
$$
\begin{aligned}
F_2^{\nu} + F_2^{\bar{\nu}} &= 2x\sum_i (q + \bar{q}) = \Sigma \\
F_3^{\nu} + F_3^{\bar{\nu}} &= u_V + d_V.
\end{aligned}
$$





- Determine total valence and total singlet (sea) quarks.
- How can we disentangle strange PDFs? CKM matrix |Vcs|~1, so if you produce charm you likely scattered off strange → Semi inclusive DIS Dimuon Processes.
- One of main constraints on strangeness asymmetry s-$\bar{s}$ in PDFs.

NuTeV experiment

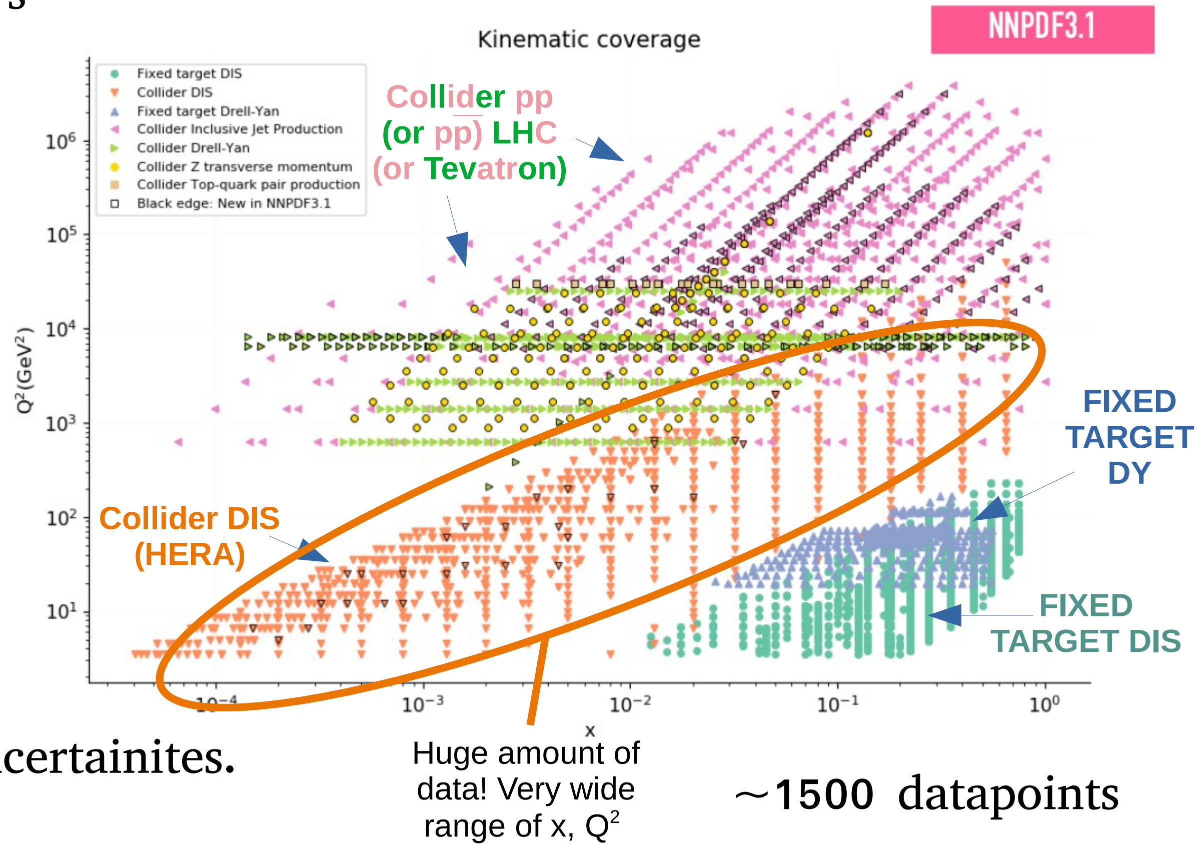


PDF Constraints - Collider DIS

4

- Must therefore constrain the PDFs with fits to data:
- Huge amount of data in global PDF fits – CT, MSHT, NNPDF.
- More than **4000** datapoints.
- Fixed target DIS, collider DIS, Drell-Yan, jets, top from pp (or $p\bar{p}$) colliders.
- Constrains central values and uncertainites.

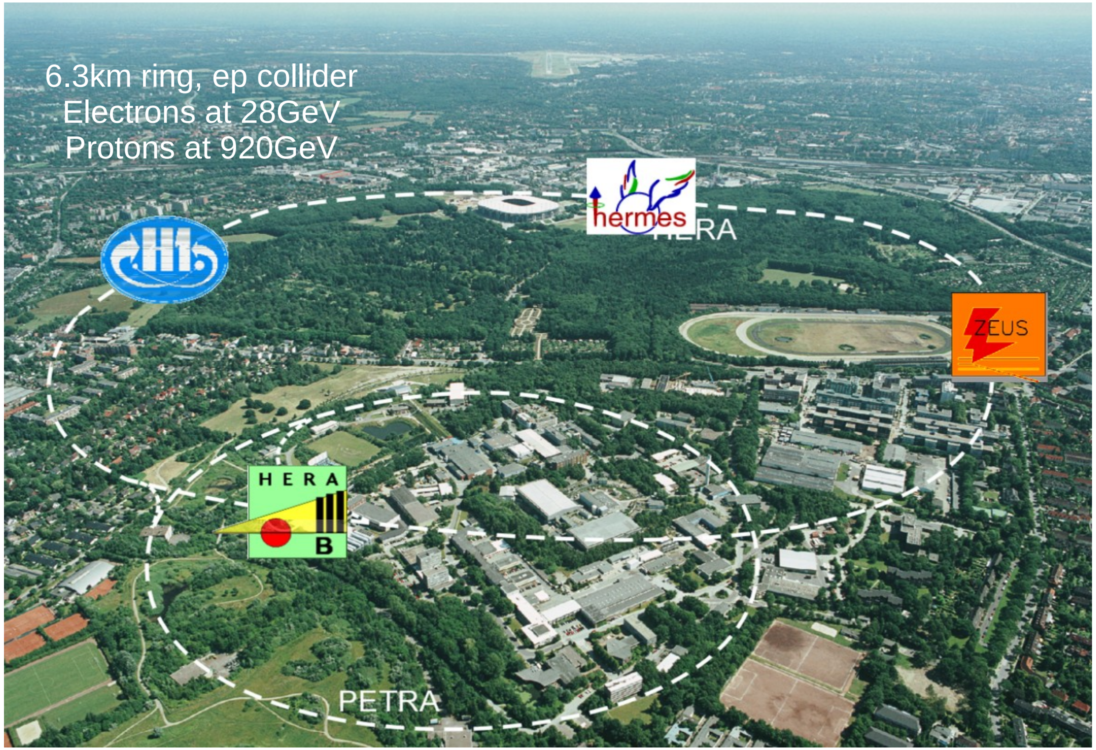
Figure adapted from NNPDF3.1 (1706.00428)



$\sim$**1500** datapoints

Collider DIS - HERA

- HERA at DESY (1992-2007).
- One of aims was precise measurement of structure functions at lower x and higher Q^2.
- Range of Q^2 gives access partons over wide range.
- Also observe scaling violations – access indirectly gluon PDF.

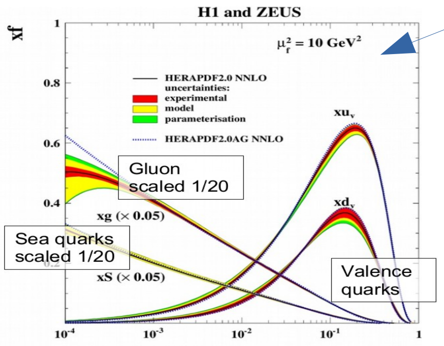


- Backbone of PDF fits! ~1500 datapoints.

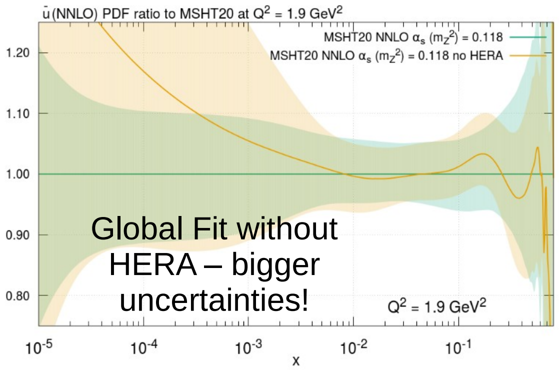
PDF Constraints - HERA

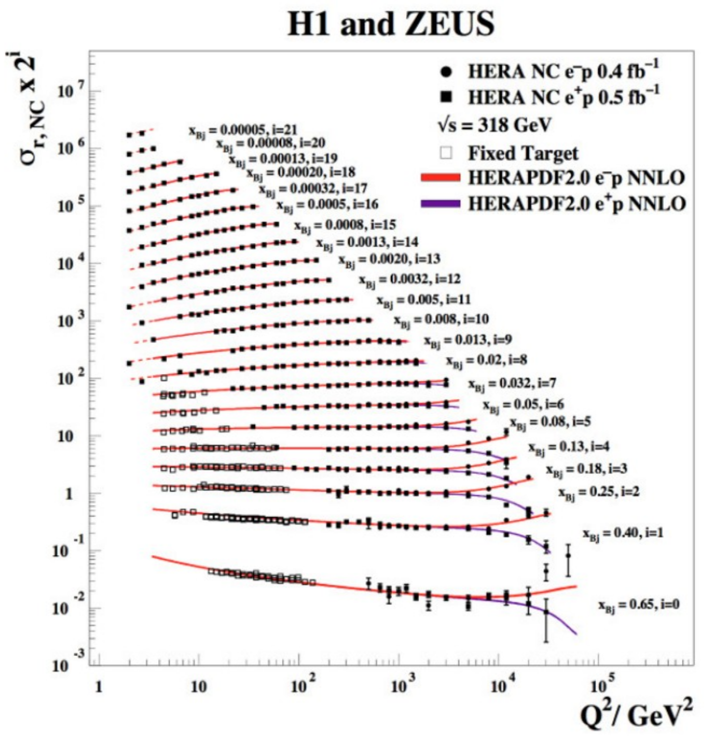
4

- Take $F_2(x,Q^2) = x\sum_q e_q^2\,[q(x,Q^2)+\bar{q}(x,Q^2)]$ → $\frac{\partial F_2}{\partial \ln Q^2} = x\sum_q e_q^2 \frac{\partial(q+\bar{q})}{\partial \ln Q^2} = \frac{\alpha_s}{2\pi}x\sum_q e_q^2\,[P_{qq}\otimes(q+\bar{q})+2\,P_{qg}\otimes g]\,.$
- Observe scaling violations – dF_2/dQ^2 depends on Q^2!
- QCD improved parton model means we have parton splittings → $dF_2/dQ^2 \sim \alpha_s g$. Indirect sensitivity to g PDF
- Can do PDF fits using only HERA – HERAPDF sets.



EPJ C75 (2015) 85



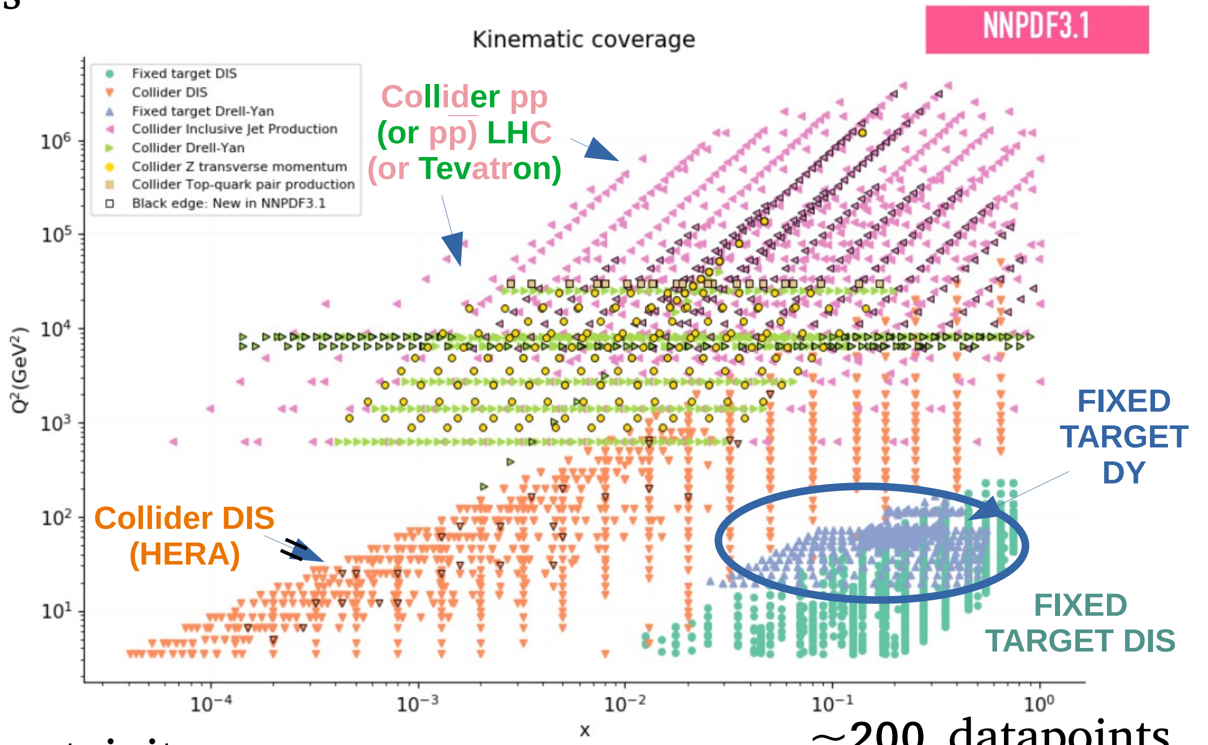


- Charm/bottom parts – F_2^c, F_2^b, constrain heavy quarks

PDF Constraints - Fixed Target DY

Figure adapted from NNPDF3.1 (1706.00428)

- Must therefore constrain the PDFs with fits to data:
- Huge amount of data in global PDF fits – CT, MSHT, NNPDF.
- More than 4000 datapoints.
- Fixed target DIS, collider DIS, Drell-Yan, jets, top from pp (or $p\bar{p}$) colliders.



$\sim$200 datapoints

- Constrains central values and uncertainites.

PDF Constraints – Fixed Target DY

4

- Difference of $\bar{u}$, $\bar{d}$ at high x difficult to disentangle in DIS:
- Easier in DY: Consider both proton and deuteron targets:

Can constrain via "Gottfried Sum Rule":

$$S_G(Q^2) = \int_0^1 \frac{dx}{x}\left[F_2^p(x,Q^2) - F_2^n(x,Q^2)\right] = \frac{1}{3} - \frac{2}{3}\int_0^1 (\bar{d} - \bar{u})\, dx.$$

Measured at NMC < 1/3 → $\bar{d} > \bar{u}$

$$pp \approx (4u(x_1)\bar{u}(x_2) + d(x_1)\bar{d}(x_2))$$

$$pn \approx (4u(x_1)\bar{d}(x_2) + d(x_1)\bar{u}(x_2))$$

Taking $u, d(x_1) \gg u, d(x_2)$

$d(x) \ll 4u(x)$

$$\frac{pd}{2pp} = (pp + pn)/2pp \approx 0.5\left(1 + \frac{\bar{d}(x2)}{\bar{u}(x2)}\right)$$

$$\left(1 + \cancel{0.25\frac{d(x_1)}{u(x_1)}}]\right) \Big/ \left(1 + 0.25\cancel{\frac{d(x_1)\bar{d}(x_2)}{u(x_1)\bar{u}(x_2)}}\right)$$

N.B. Here low Q^2 so γ → still charge-weighted but now products of PDFs.

Most complex models have $\bar{d} > \bar{u}$ at high x. Not clear though!

- Direct measure of $\bar{d}/\bar{u}$, at high x.
- What do we expect? Difficult Question – Pauli Exclusion → $\bar{d} > \bar{u}$?

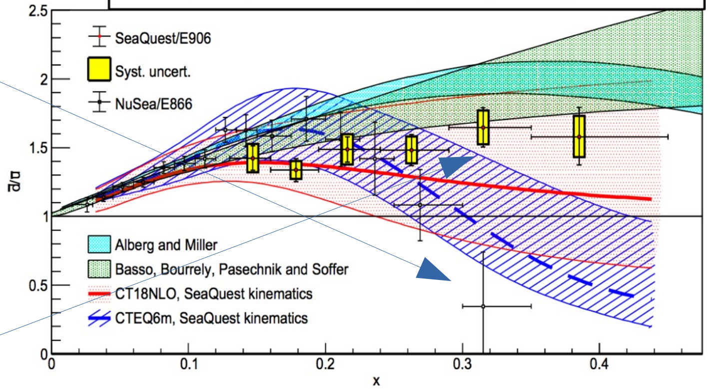
More u than d means more symmetrisation possibilities

- Pion cloud, p → π+ > p → π0 or p → π- → $\bar{d} > \bar{u}$? - Antisymmetrisation → $\bar{u} > \bar{d}$?

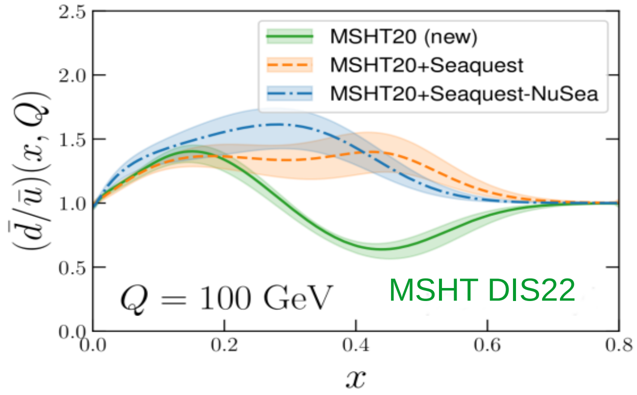
PDF Constraints – Fixed Target DY ④

- What does Fixed Target DY at E866(NuSea) tell us then?
- Remember $\frac{pd}{2pp} = (pp + pn)/2pp \approx 0.5\left(1 + \frac{\bar{d}(x2)}{\bar{u}(x2)}\right)$ → data implies $\bar{d} < \bar{u}$ at high x
- Puzzling – remeasured recently by E906 (Seaquest) Fermilab.
- Different results!
  → data implies $\bar{d} > \bar{u}$ at high x.
  → Not clear why?
- *How to deal with in Global PDF fit?*
  → example of **tensions between datasets** which is often an issue in PDF fits.

SeaQuest, Nature 590 (2021) 7847, 561-565



Unlike most theoretical models!



In better agreement with theory.

PDF Constraints - Collider pp/p$\bar{\text{p}}$

- Must therefore constrain the PDFs with fits to data:
- Huge amount of data in global PDF fits – CT, MSHT, NNPDF.
- More than **4000** datapoints.
- Fixed target DIS, collider DIS, Drell-Yan, jets, top from pp (or p$\bar{\text{p}}$) colliders.

Huge amount of data! Very wide range of x, Q^2

$\sim$**1800** datapoints.

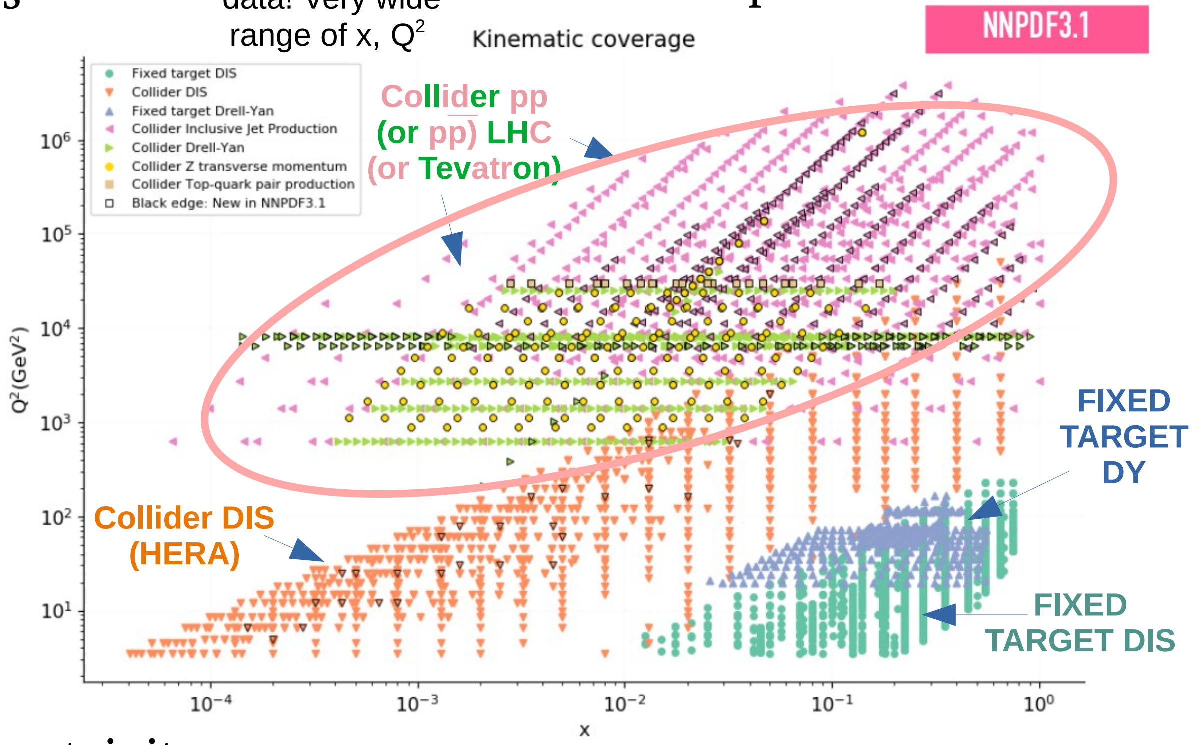


Figure adapted from NNPDF3.1 (1706.00428)

- Constrains central values and uncertainites.

PDF Constraints – ATLAS/CMS DY

4

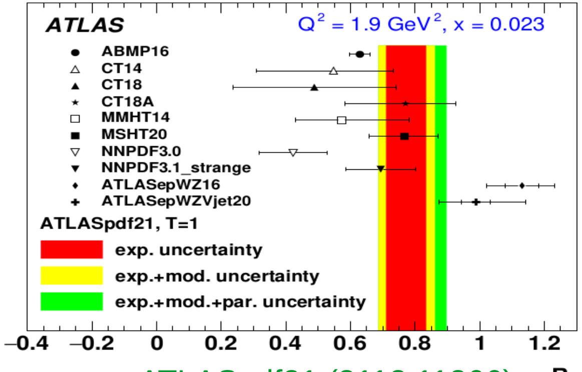
- Since 2010 lots of LHC data has been produced, adds further constraints on PDFs!
- Includes Drell-Yan of course, higher Q^2 still so lower x and now pp so complementary PDF sensitivity.

$$\sigma(pp \to Z) = u\bar{u} + d\bar{d} + s\bar{s}$$
$$\sigma(pp \to W^+) = u\bar{d} + c\bar{s}$$
$$\sigma(pp \to W^-) = d\bar{u} + s\bar{c}$$

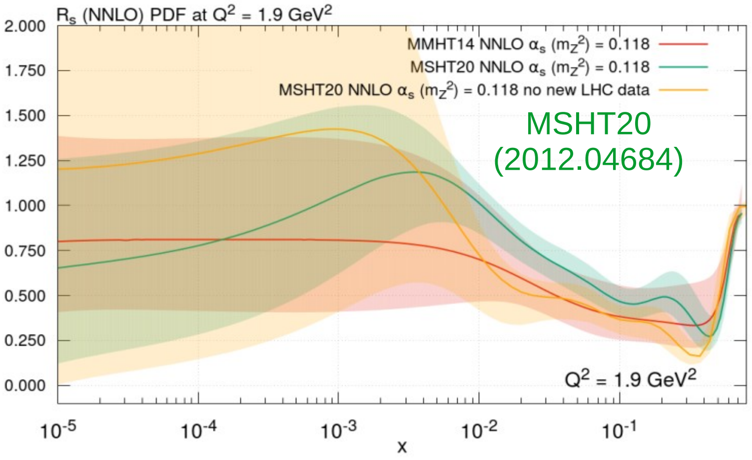
- ATLAS 7TeV W,Z is high precision. (CMS results as well)
- Up and down quarks/antiquarks already quite constrained from non-LHC data, therefore precision means **strangeness is most constrained.**

Strangeness enhancement!

$$R_s = \frac{s + \bar{s}}{\bar{u} + \bar{d}}$$



ATLASpdf21 (2112.11266)



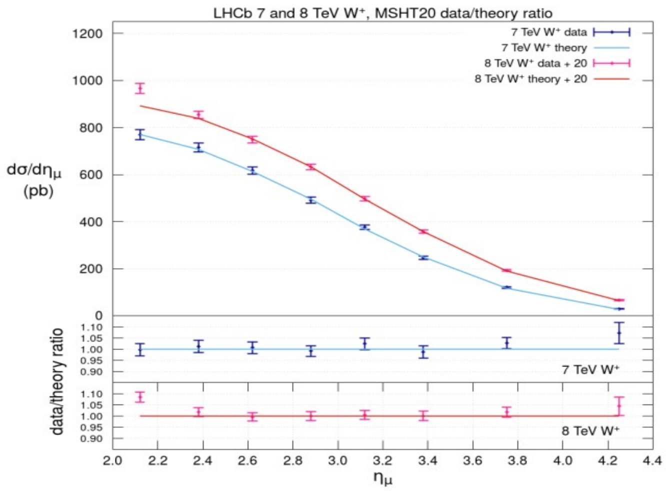
PDF Constraints – LHCb DY

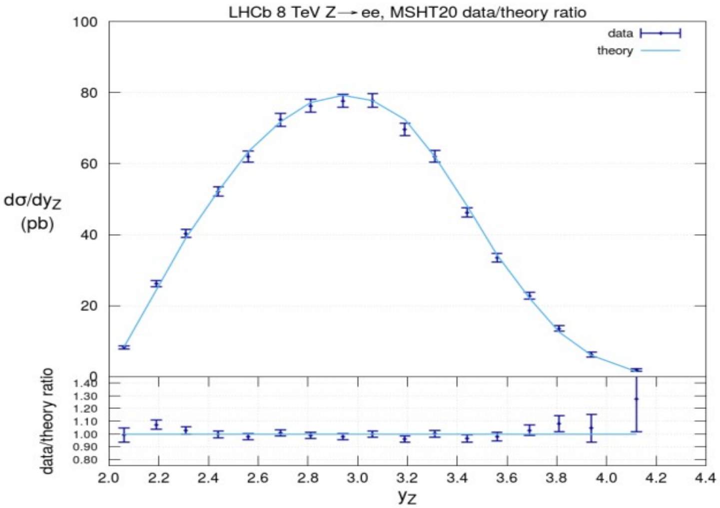
4

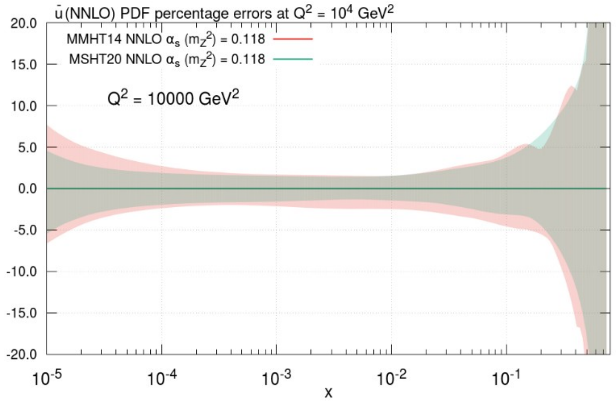
- LHCb is a forward detector experiment, rapidities in range **2-4.5** → help to constrain valence quarks at higher x and in particular sea quarks at lower x than possible ATLAS/CMS.
  → good fits obtained in global PDF fits:

$$x_{1,2} = \frac{Q}{\sqrt{s}} e^{\pm y}$$

Figs from MSHT20 (2012.04684)



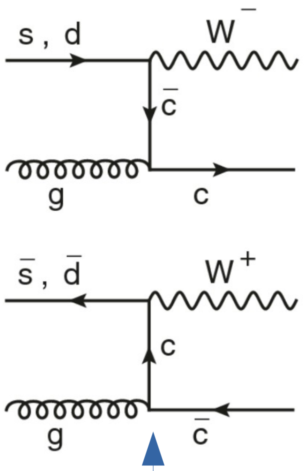




PDF Constraints – W/Z + charm

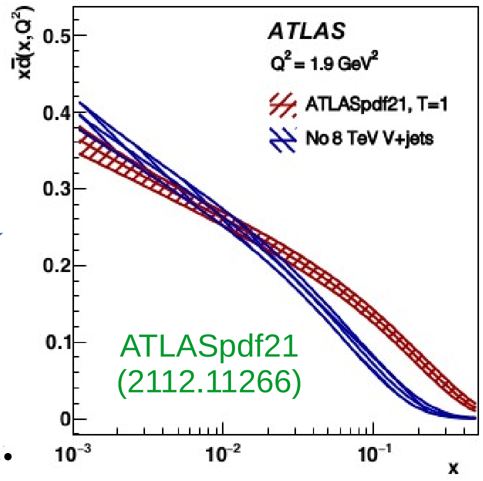
4

- Additionally more exclusive measurements can constrain strange/charm. E.g. CMS W+c data available at 7, 13TeV.
- Full NNLO calculation difficult due to mismatch of theory definition of final state charm jet and experimental measurement.
- Sensitive to strange content of proton.
- Prefers intermediate strangeness, not as enhanced as ATLAS 7TeV W,Z data.
- Also W/Z + jet data – additional probe of light quarks, higher Q^2.
- ATLASpdf21 see notable impact, more than in global PDF fits as less other data included.



Theory calculated using flavour kT algorithm vs experiment using anti-kT. More work recently on theory side with different jet algorithms.

Need IRC safe observable for theory calculation.

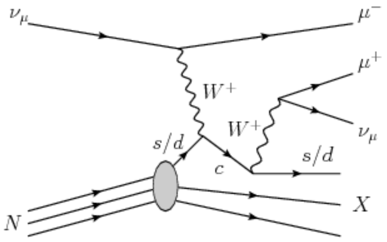


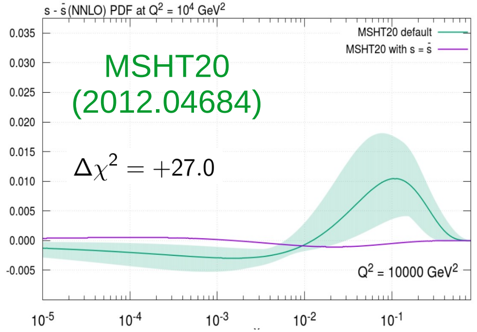
ATLASpdf21 (2112.11266)

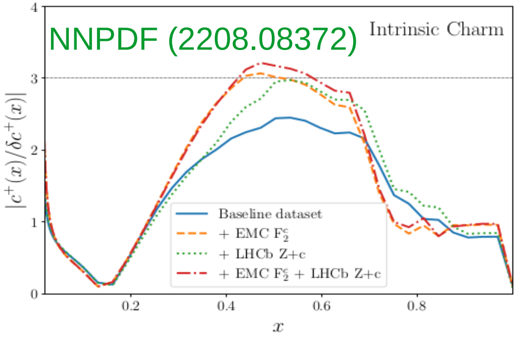
PDF Constraints – Intrinsic charm? ④

- Could there be a non-perturbative component of strange/charm in proton, like u, d?
- Clear sign would be quark-antiquark asymmetry, "cannot" be generated by $g \rightarrow q\bar{q}$.
- Weak evidence for strangeness asymmetry:
- Charm – little evidence of asymmetry so far.
- But could also be total charm > expected from perturbative splittings.
- NNPDF obtain "fitted charm" – parameterise charm like other PDFs + "subtract" off perturbative part.
- Difficult to separate from other effects (mc, gluon, higher orders, higher twist)
- Z + c jet – probes charm in proton.

NuTeV Dimuon SIDIS data







NNPDF claim evidence for "instrinsic" charm, CT same analysis don't find it.

PDF Constraints – High x gluon

- High x gluon still quite unconstrained, limited sensitivity from DIS data and DY only depends on gluon beyond LO – need further LHC data.
- How can we constrain it?
- Need processes which are gluon initiated.
  - Jets
  - Top
  - Z pT spectrum (latter beyond LO).

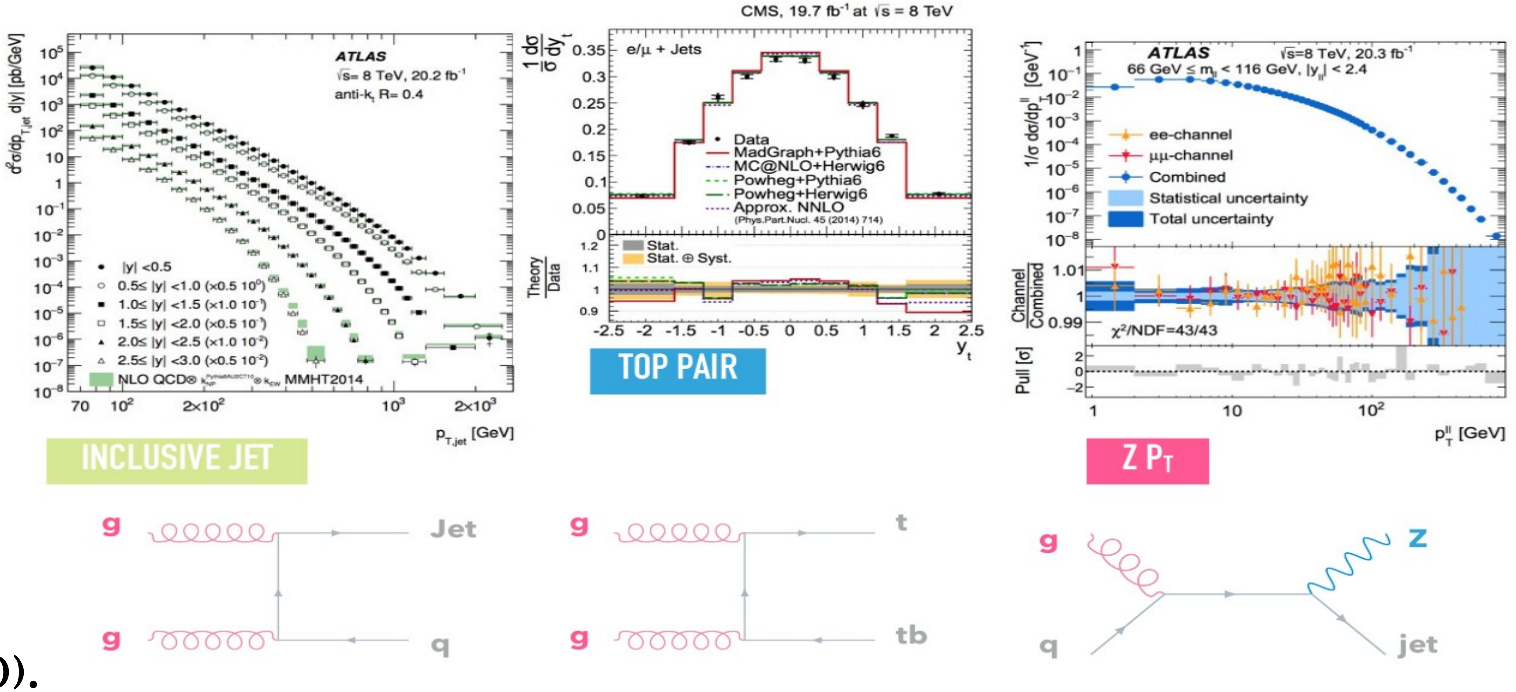
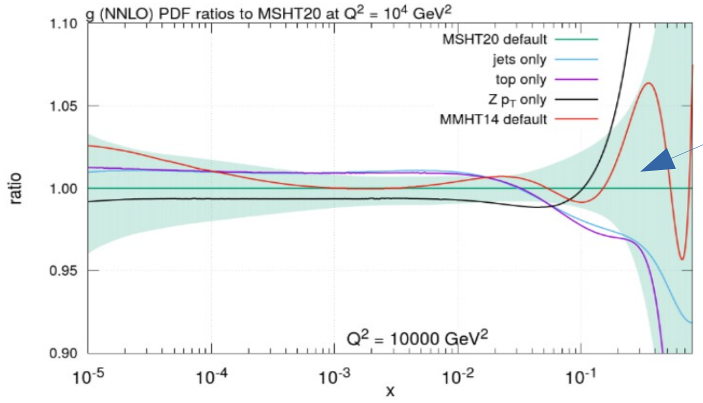


Fig from M. Ubiali

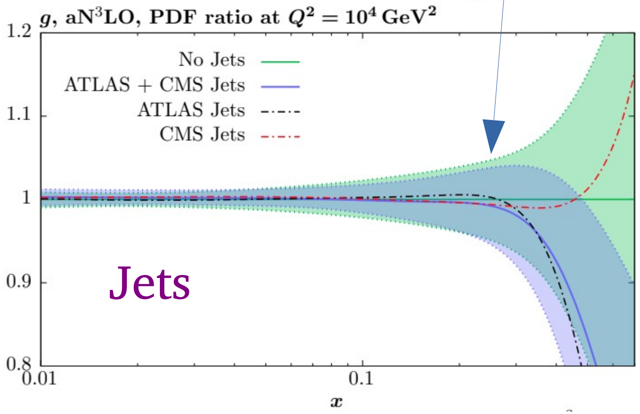
PDF Constraints – Jets, ZpT + Top

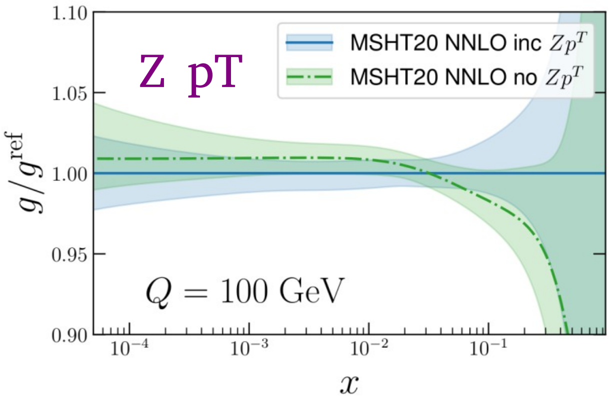
4

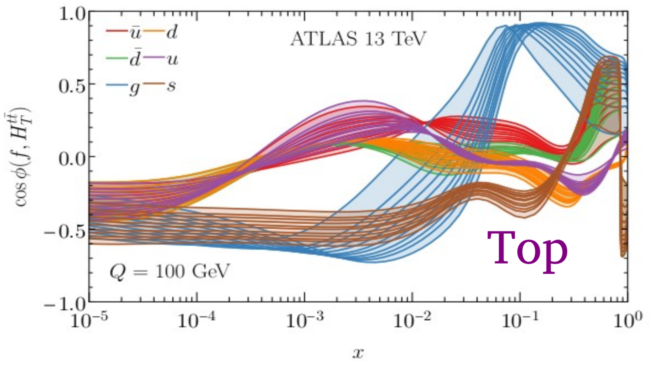
- Jet data clearly reduce uncertainty, but tension between ATLAS and CMS (7+8TeV)
- Further constraint from Z transverse momentum spectrum (pT > 30 GeV).
- Top data complementary – correlation with gluon PDF (bottom right).



Overall fit is balance of these three data types for high x gluon







5. PDF Methodology and Uncertainties

Global PDF Fitting

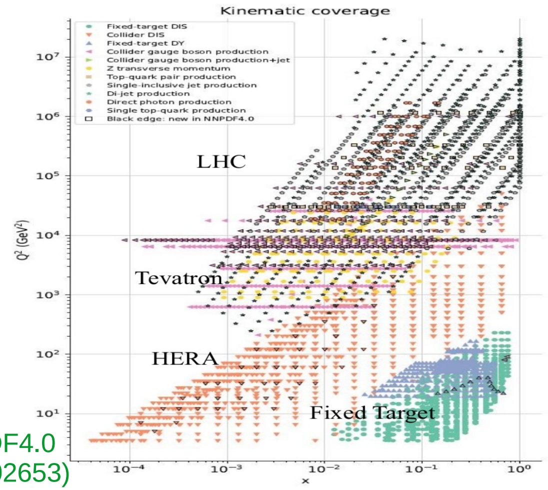
1) Experiment

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

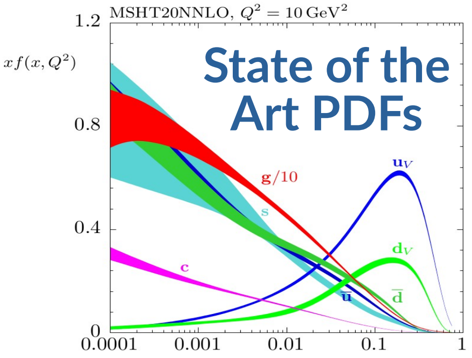
2) Methodology

- 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
- Approximate N3LO + theory uncertainties available

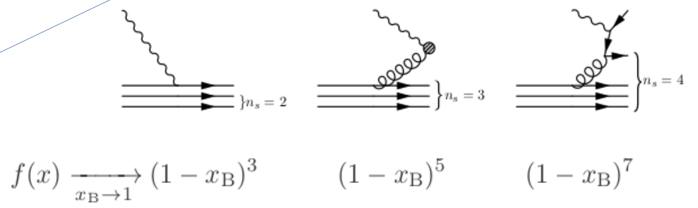
PDF Parameterisation

5

- First need to parameterise PDFs at input scale, $Q_0 \sim 1$GeV.
- 11 PDFs to consider:

$$u, \bar{u}, \quad d, \bar{d}, \quad s, \bar{s}, \quad c, \bar{c}, \quad b, \bar{b}, \quad g$$

η~3 for valence quarks, 5 for gluon, 7 for antiquarks. η~2*#"spectators"+1

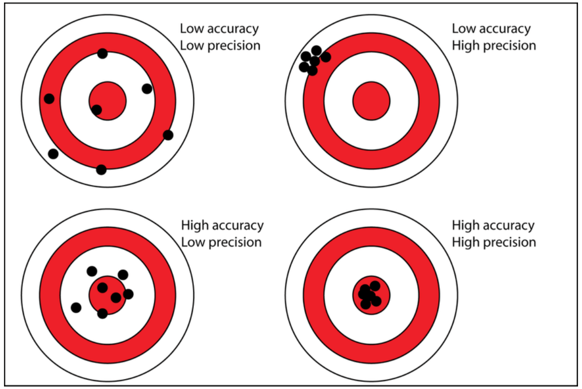


- Parameterise as:

For things that scale with gluon (singlet, gluon) δ ~ 0. Otherwise δ~0.5.

$$\boxed{xf_0(x, Q^2) = x^{\delta}(1 - x)^{\eta} F(x)}$$

- F(x) can then be fixed extendable parameterisation, neural network, etc.
- Must determine number of free parameters, neural network architecture, methodological settings to maximise accuracy and precision but avoid over-fitting/bias.



PDF Parameterisation

$$\Sigma = \sum_{i=1}^{5}(q_i + \bar{q}_i)$$

- Also have sum rules:

Valence sum rules:

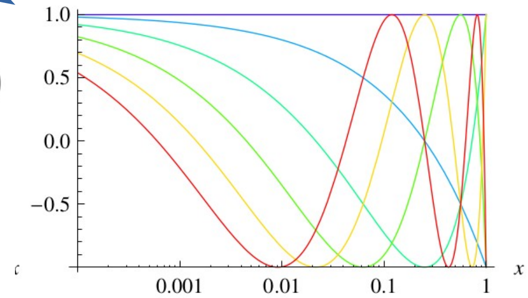
$$\int_0^1 u_V(x)\,dx = 2 \qquad \int_0^1 d_V(x)\,dx = 1$$

Momentum sum rule:

$$\int_0^1 x\Sigma(x) + xg(x)\,dx = 1.$$

- Cannot constrain all partons independently.
- What is F(x)? $xf_0(x, Q^2) = x^\delta(1-x)^\eta F(x)$
- In MSHT it's an orthogonal basis of functions – "Chebyshevs": $T_i\left(y(x) = 1 - 2\sqrt{x}\right)$
- In NNPDF it's a neural network (with constraints applied).
- In CT, "Bernstein polynomials": $P_{u_v} = d_0 p_0(y) + d_1 p_1(y) + d_2 p_2(y) + d_3 p_3(y) + d_4 p_4(y)$
- In HERAPDF it's standard polynomial: $1 + Dx + Ex^2$

→ **<u>All</u>** involve some choices, then investigate these in the fit…



PDF Parameterisation (MSHT)

5

After fit parameters are rearranged into orthogonal eigenvectors → PDF uncertainties and variations

More info in MSHT (2012.04684)

MSHT20 Parameterisation:

$$u_v(x,Q_0^2) = A_u(1-x)^{\eta_u}x^{\delta_u}(1+\textstyle\sum_{i=1}^{6} a_{i,u}T_i(1-2x^{\frac{1}{2}}));\ A_u \text{ fixed by } \int_0^1 u_v\,dx = 2$$

$$d_v(x,Q_0^2) = A_d(1-x)^{\eta_d}x^{\delta_d}(1+\textstyle\sum_{i=1}^{6} a_{i,d}T_i(1-2x^{\frac{1}{2}}));\ A_d \text{ fixed by } \int_0^1 d_v\,dx = 1$$

$$sea(x,Q_0^2) = A_S(1-x)^{\eta_S}x^{\delta_S}(1+\textstyle\sum_{i=1}^{6} a_{i,S}T_i(1-2x^{\frac{1}{2}}));$$

$$s^+(x,Q_0^2) = A_s(1-x)^{\eta_s}x^{\delta_S}(1+\textstyle\sum_{i=1}^{6} a_{i,s}T_i(1-2x^{\frac{1}{2}}));\ (a_{i,s} \neq a_{i,S}, i = 5,6)$$

$$(\bar{d}/\bar{u})(x,Q_0^2) = A_{\rm rat}(1-x)^{\eta_{\rm rat}}(1+\textstyle\sum_{i=1}^{6} a_{i,{\rm rat}}T_i(1-2x^{\frac{1}{2}}));$$

$$g(x,Q_0^2) = A_g(1-x)^{\eta_g}x^{\delta_g}(1+\textstyle\sum_{i=1}^{4} a_{i,g}T_i(1-2x^{\frac{1}{2}})) - A_{g-}(1-x)^{\eta_{g-}}x^{\delta_{g-}};$$

$$s^-(x,Q_0^2) = A_{s-}(1-x)^{\eta_{s-}}(1-x_0/x)x^{\delta_{s-}}.\ x_0 \text{ fixed by } \int_0^1 s^-\,dx = 0,\ \delta_{s-} \text{ fixed.}$$

51 parton parameters

(36 in MMHT14)

7 extra eigenvectors

- 1 extra in each of PDFs, except in s^-, 2 extra in s^+.

Net $\Delta\chi^2_{\rm global} = -73$.

MSHT20: 2012.04684

$$Sea,\ S(x) = 2(\bar{u}(x) + \bar{d}(x)) + s(x) + \bar{s}(x)$$

$$s^{\pm}(x) = s(x) \pm \bar{s}(x)$$

What about heavy quarks? c, b? (see later!)

Global PDF Fitting

5

1) Experiment

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

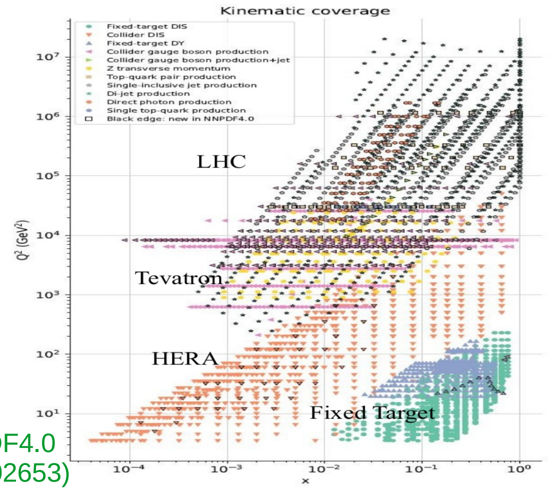
Global PDF fit

2) Methodology

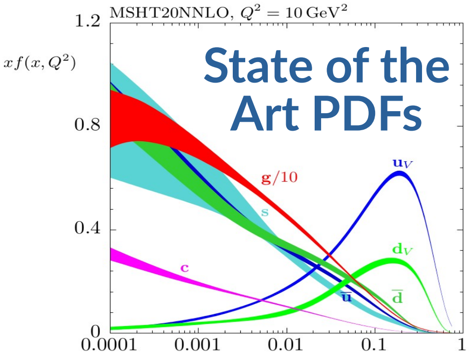
- Parameterise at low scale
- DGLAP, flavour schemes, …
- Minimisation of χ^2
- Uncertainty prescription

3) Theory

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



NNPDF4.0
(2109.02653)



PDF Methodology - χ^2

5

- How is this data used to give the global fit PDFs?
- Compare with theoretical predictions at NNLO in QCD (+NLO EW where relevant).
- Minimise χ^2, which measures difference of data and theory:

Consider usual PDF fit probability:

Theory Data Hessian matrix - contains uncorrelated (s_k) and correlated uncertainties (β_k)

$$P(T|D) \propto \exp(-\chi^2) \propto \exp\left(-\frac{1}{2}(T-D)^T H_0 (T-D)\right)$$

$$\propto \exp\left(-\frac{1}{2}\sum_{k=1}^{N_{pt}} \frac{1}{s_k^2}\left(D_k - T_k - \sum_{\alpha=1}^{N_{corr}} \beta_{k,\alpha}\lambda_\alpha\right)^2 + \sum_{\alpha=1}^{N_{corr}} {\lambda_\alpha}^2\right)$$

Experimental Nuisance parameters

N.B. Alternative is covariance matrix instead of nuisance parameters – equivalent!

$$\chi^2 = \frac{1}{2}(T-D)^T C^{-1}(T-D)$$

- Obtain "best fit" PDF with minimum χ^2.
- **PDFs then made available on LHAPDF** for use by community (and on group websites).

PDF Methodology - Uncertainty

5

Also Bayesian methods and MCMC explored:

$$p(\theta|D) = \frac{p(D|\theta)\,p(\theta)}{p(D)}$$

- Then two main ways used to obtain central PDF and uncertainty.

1) Hessian

- Minimise difference of real data and theory to obtain best fit PDF as central value.
- Obtain uncertainty by diagonalising at central value to obtain eigenvectors and using $\Delta\chi^2=1$ (or T) for PDF uncertainty.
- ~20-100 eigenvectors.
- CT, MSHT. (+HERAPDF, ATLASPDF21)

$$\Delta F = \frac{1}{2}\sqrt{\sum_{k=1}^{n}\left[F(S_k^+) - F(S_k^-)\right]^2},$$

2) Replica

- Fluctuate real data by its uncertainties → pseudo-data replicas.
- For each replica minimise difference of pseudo-data and theory to get PDF.
- Central value is average and uncertainty as 68% width of replica distribution.
- ~100-1000 replicas.
- NNPDF

$$\Delta F = 0.5(F(N_{rep,84\%}) - F(N_{rep,16\%}))$$

Can convert between the two forms (1205.4024,1401.0013, 1505.06736)

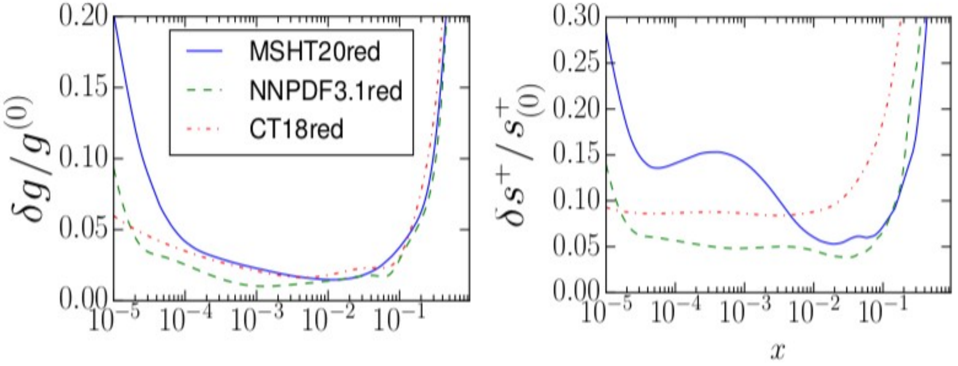
PDF Methodology - Uncertainty

5

- How do these methods compare? Perform fit to same (reduced) set of data:

Benchmark = similar data/settings

PDF4LHC21, arXiv:2203.05506



- With larger difference again expected in more recent NNPDF4.0 fits

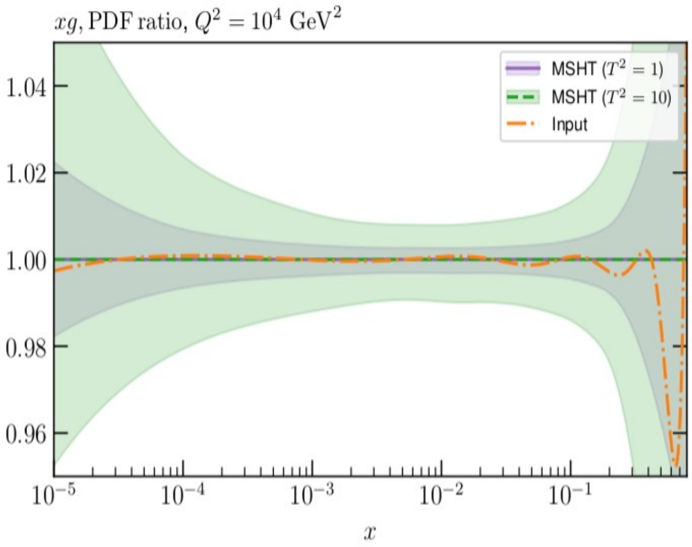
★ Suggests three possibilities:

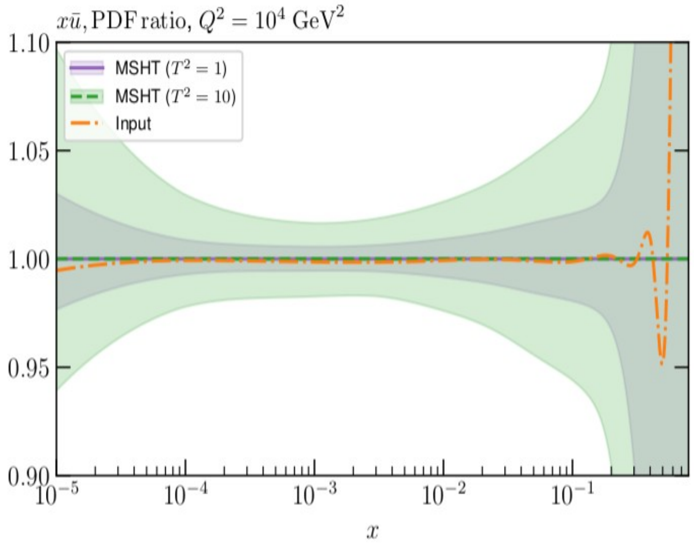
1. **NN** uncertainty not conservative enough (**too small**).
2. Fixed par. (**MSHT**) uncertainty too conservative (**too large**).
3. Fixed par. (**MSHT**) fit less accurate, due to parameterisation inflexibility, and hence enlarged errors needed (**less precise**).

PDF Methodology - Uncertainty

5

- To understand this we perform "closure test", generate pseudodata from NN fit (NNPDF) and fit with fixed parameterisation (MSHT).





Fixed parameterisation can fit just as well as NN (in fact slightly better!)

	χ^2/N_{pts}
MSHT (fixed param)	0.0005
Neural Net (NNPDF) **4.0 meth.**	0.002

L. Del Debbio, T. Giani an M. Wilson, arXiv:2111.057..

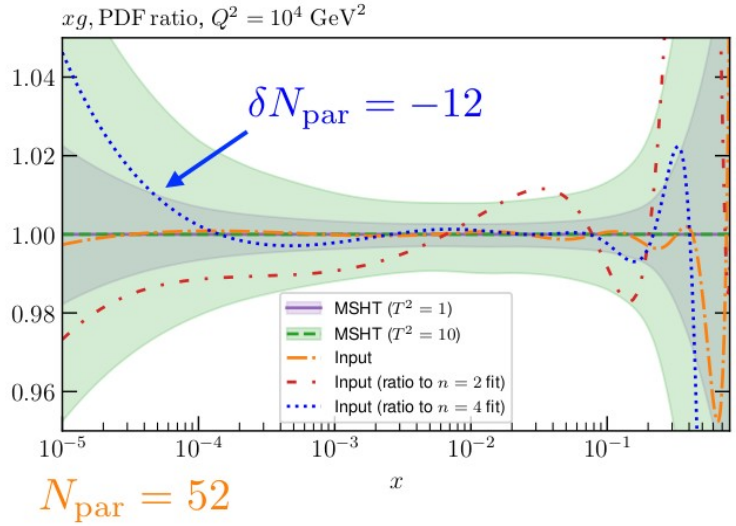
*Subtlety is NN stopping criterion and overfitting.

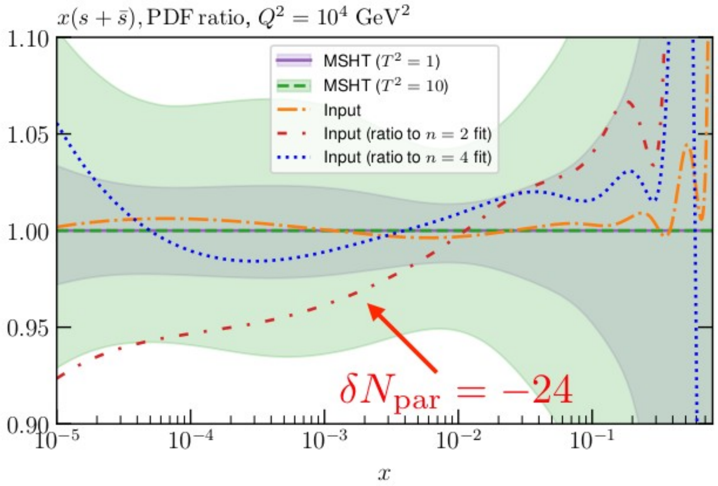
→ In data region input PDF matched very well, and much better than $T^2 = 1$ uncertainties. No evidence that the increased tolerance is driven by parameterisation inflexibility for MSHT.

PDF Methodology - Uncertainty

5

- Can also test with reduced parameterisations with fewer parameters – fail closure test! **Beware PDF fits with too few parameters** - sometimes done by experiments.





- Note this level of agreement is not automatic! Need flexible enough parameterisation: restricting number of free parameters gives much poorer agreement.
- Is also not coincidental: parameterisation chosen in order to provide 1% precision.

PDF Methodology - Uncertainty

- So why the difference in uncertainty size if not down to parameterisation?

- **Simple observations:** global fit quality bad strictly speaking, distribution amongst datasets not as expected.

$$\frac{\chi^2}{N_{\rm pts}} \gg 1 + \sigma(N_{\rm pts}) \sim 1.02$$

$$N_{\rm pts} \sim 4000 - 5000$$

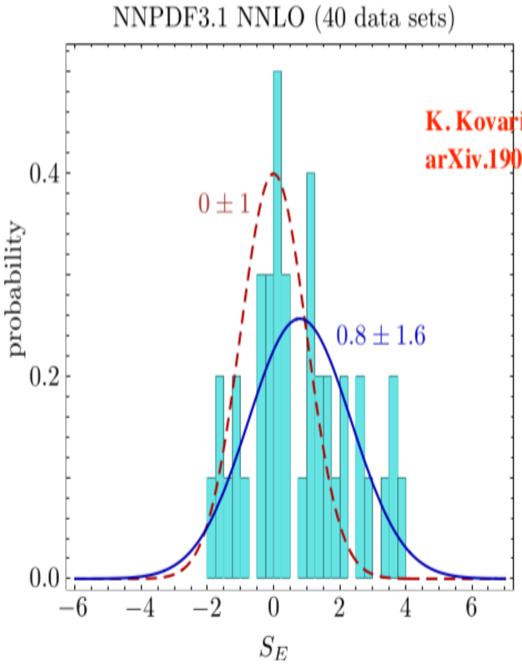
MSHTaN3LO

	LO	NLO	NNLO	N^{3}LO
$\chi^2_{N_{pts}}$	2.57	1.33	1.17	1.14

And similar for other fits!

- **Tensions** between different datasets often larger than naively expect.

G. Watt and R. Thorne, arXiv:1205.4024 M. Yan et al., arXiv.2406.01664
J. Pumplin, arXiv:0909.0268



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

K. Kovarich et al., arXiv.1905.06957

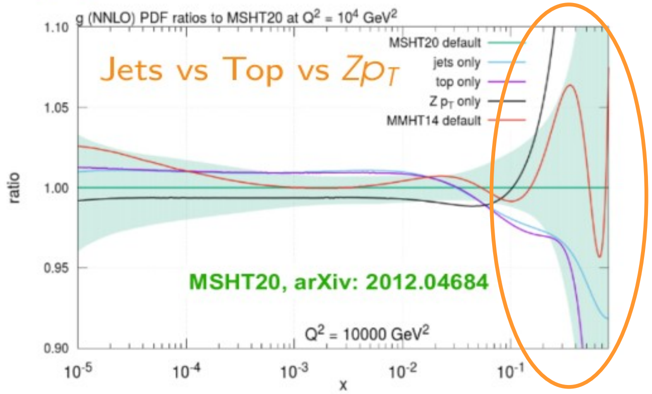
Fit is not a textbook case – fit quality > 1, one reason is dataset tensions….

- Given complete statistical compatibility, global PDF fit very constraining. Danger is claimed (high) precision will increasingly not match accuracy with $T = 1$. Motivates enlarged tolerance $T > 1$.

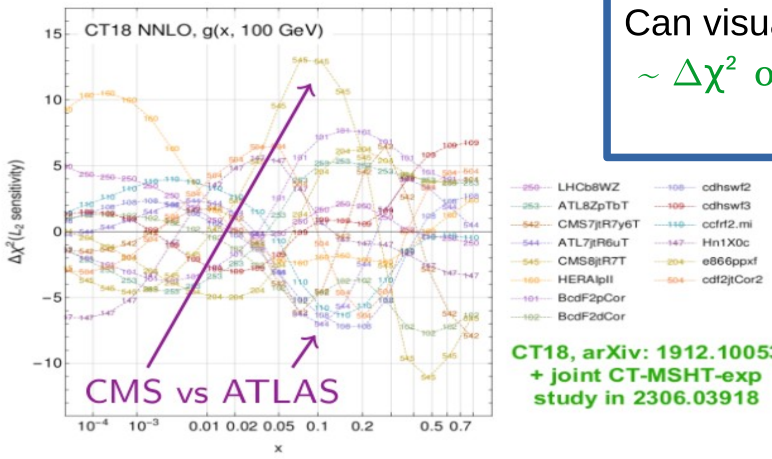
PDF Methodology – Data Tensions

5

- Wide range of data in global PDFs $\Rightarrow$ robust + reduces uncertainties.
- But global fits are adversial - datasets pull in different directions
$\Rightarrow$ data tensions both between different data types (e.g. jet vs top vs Zp_T) and different datasets of same type (e.g. ATLAS vs CMS).



MSHT global	LO	NLO	NNLO	aN3LO
χ^2/N_{pts} (4363)	2.57	1.33	1.17	1.14



CT18, arXiv: 1912.10053
+ joint CT-MSHT-exp
study in 2306.03918

Can visualise through "L2 Sensitivity":
$\sim \Delta\chi^2$ of dataset upon moving PDF by "1σ".

Publicly available program to calculate it.

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

X. Jing (inc TC) et al 2306.03918

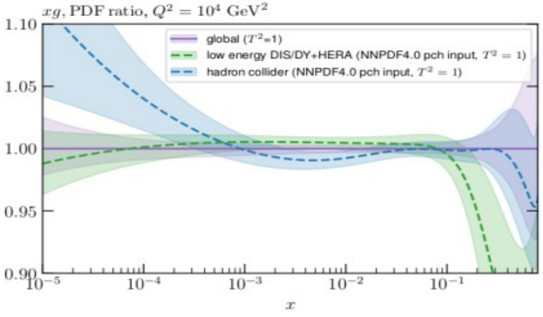
- Limits improvements in precision + causes $\chi^2/N_{\text{pts}} > 1$.
- Many reasons, sometimes connected with treatment of systematics.

PDF Methodology – Data Tensions

5

- How to quantify PDF uncertainties in presence of tensions?
- Size of uncertainties not affected at all by tensions using a "textbook" $\Delta\chi^2 = 1$ ("$T = 1$").

- $\Delta\chi^2 = 1$ would apply if:
  - Complete Statistical compatibility between all datasets.
  - Completely faithful evaluation of exp. uncertainties of each dataset.
  - Theoretical calculations to match these data exactly.



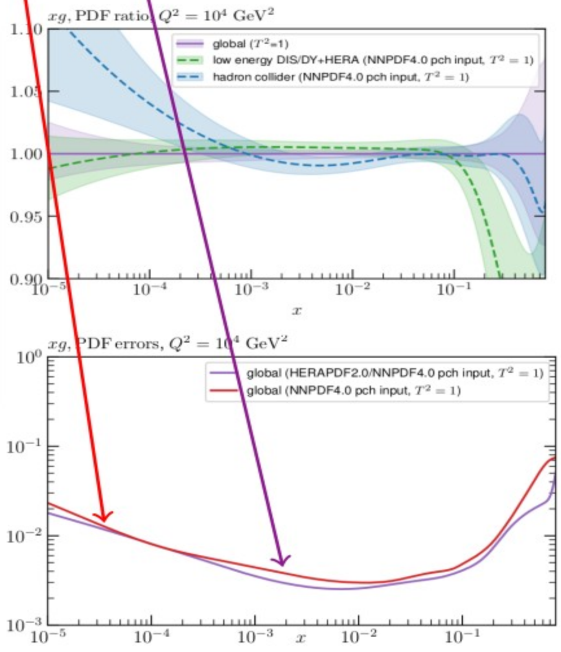
Harland-Lang, TC, R.S. Thorne 2407.07944

PDF Methodology – Data Tensions

5

- How to quantify PDF uncertainties in presence of tensions?
- Size of uncertainties not affected at all by tensions using a "textbook" $\Delta\chi^2 = 1$ ("$T = 1$"). E.g. generate data without/with tensions:

- $\Delta\chi^2 = 1$ would apply if:
  - Complete Statistical compatibility between all datasets.
  - Completely faithful evaluation of exp. uncertainties of each dataset.
  - Theoretical calculations to match these data exactly.



Harland-Lang, TC, R.S. Thorne 2407.07944

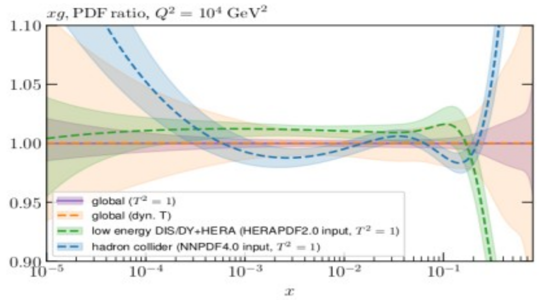
In textbook T=1 treatment, uncertainties unchanged by presence of tensions! I.e. tensions not reflected in uncertainty.

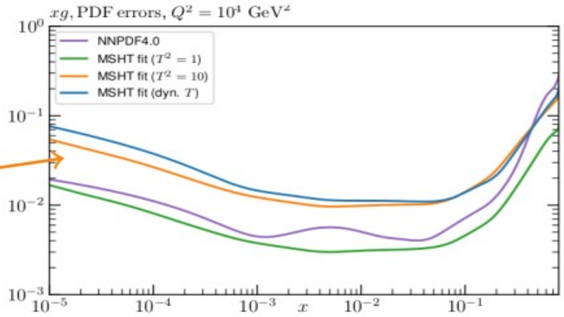
PDF Methodology – Data Tensions

5

- How to quantify PDF uncertainties in presence of tensions?
- Size of uncertainties not affected at all by tensions using a “textbook” $\Delta\chi^2 = 1$ (“$T = 1$”).

- $\Delta\chi^2 = 1$ would apply if:
  - Complete Statistical compatibility between all datasets.
  - Completely faithful evaluation of exp. uncertainties of each dataset.
  - Theoretical calculations to match these data exactly.





- MSHT (and CT) use a tolerance to account for dataset tensions + other effects (similar idea to PDG rescaling):

$$\Delta\chi^2 = T^2 \quad (T > 1).$$

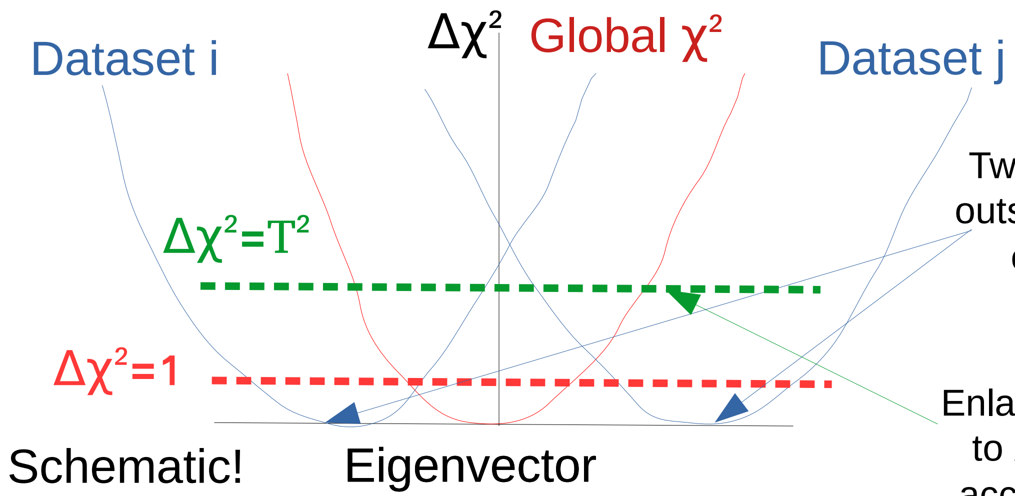
In textbook T=1 treatment, uncertainties unchanged by presence of tensions! I.e. tensions not reflected in uncertainty.

Harland-Lang, TC, R.S. Thorne 2407.07944

PDF Methodology – Data Tensions

5

- Tolerance therefore accounts for tensions between data, whether due to theory, parameterisation or experimental issues.



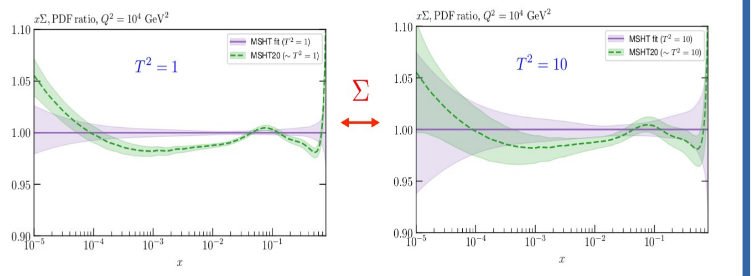
Idea essentially is to enlarge PDF uncertainty to account for dataset tensions

Two datasets I, j sit outside $\Delta\chi^2$=1 bound on either side.

Enlarge bounds to $\Delta\chi^2$=T^2 to account for it

- How much to enlarge?
- Different prescriptions, in MSHT enlarge so each dataset sits within 68% of expected χ^2 for Ndata.

Same parameterisation, different data/theory in MSHT → PDFs change notably



- Often find changing data in PDF fit moves results by $> \Delta\chi^2 = 1$ uncertainty. Covered by larger $\Delta\chi^2 = T^2$

6. PDF Theory + State of the Art

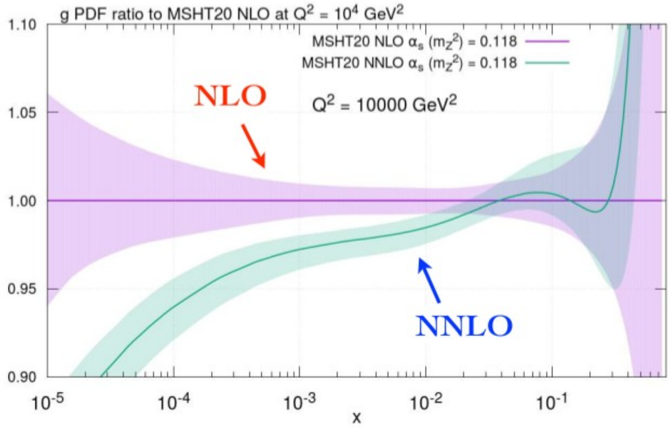
PDF Theory

- Standard is now **NNLO in QCD**, minimum required to describe LHC precision data:

$$O \sim \underbrace{a_0}_{LO} + \underbrace{a_1 \alpha_s}_{NLO} + \underbrace{a_2 \alpha_s^2}_{NNLO} + \dots$$

MSHT20

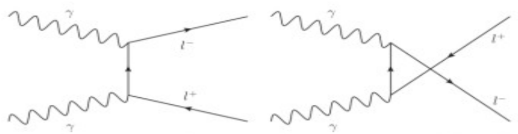
Data set	N_{pts}	NLO χ^2/N_{pts}	NNLO χ^2/N_{pts}
Total LHC data	1328	1.79	1.33
Total non-LHC data	3035	1.13	1.10
Total	4363	1.33	1.17



- PDF fits need to recompute theory predictions throughout fit – need precise predictions for range of processes and PDF evolution.
- PDF DGLAP evolution and DIS cross-sections typically done "in-house".
- LHC cross-sections are very computationally expensive to calculate, particularly if done in fiducial phase space (i.e. kinematic cuts) and multi-differential → use grids.

QED PDFs

- With NNLO QCD now standard, noting that $\alpha_{\mathrm{QED}}(M_Z) \sim \alpha_S^2(M_Z)$:
  $\Rightarrow$ important to consider EW effects, QED corrections are a key part.
- MSHT20 include EW corrections for:
  - Drell-Yan
  - inclusive jets
  - top
  - DIS.



Photon-Initiated contributions to Drell-Yan.

- QED corrections via QED modifications to DGLAP, via photon PDF and photon-initiated processes.
- Obtain $\gamma(x, Q^2)$ with $\mathcal{O}(\%)$ uncertainties via LUXQED-related method.

LUXQED idea: Use the NC expression for DIS at low Q^2, then rewrite to obtain the photon PDF from experimentally measured structure functions:

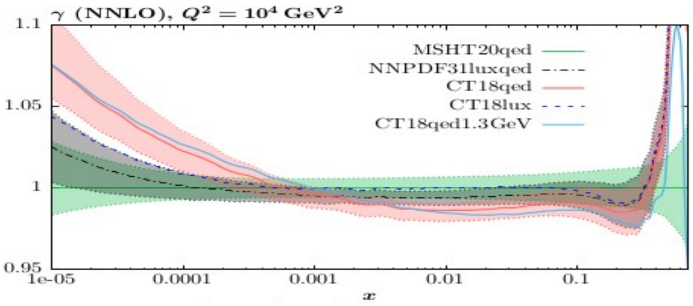
$$x\gamma(x,Q_0^2) = \frac{1}{2\pi\alpha(Q_0^2)} \int_x^1 \frac{dz}{z} \left\{ \int_{\frac{x^2 m_p^2}{1-z}}^{\frac{Q_0^2}{1-z}} \frac{dQ^2}{Q^2}\alpha^2(Q^2)\left[\left(zP_{\gamma,q}(z)+\frac{2x^2m_p^2}{Q^2}\right)F_2(x/z,Q^2) - z^2F_L(x/z,Q^2)\right] - \alpha^2(Q_0^2) z^2 F_2(x/z,Q_0^2)\right\},$$

Manohar et al, 1708.01256

Harland-Lang et al 1907.02750

QED PDFs

- General consistency with NNPDF, CT.
- Quarks reduced at high x by $q \to q\gamma$, gluon reduced by momentum sum rule.



PDF Set	QCD order	$\langle x\rangle_{\Sigma q\bar{q}}$	$\langle x\rangle_g$	$\langle x\rangle_\gamma$
CT18	NNLO	0.528	0.472	0
CT18$_{\text{QEDproton}}$	NNLO	0.526	0.471	0.00431
CT18$_{\text{QEDfit}}$	NNLO	0.526	0.469	0.00432
CT18$_{\text{LUX}}$	NNLO	0.527	0.470	0.00436
MSHT20	NNLO	0.530	0.470	0
MSHT20$_{\text{QED}}$	NNLO	0.528	0.468	0.00436
MSHT20$_{\text{QED_QCD}}$	aN3LO	0.538	0.460	0
MSHT20$_{\text{QED}}$	aN3LO	0.536	0.457	0.00441
NNPDF3.1	NNLO	0.529	0.470	0
NNPDF3.1$_{\text{QED}}$	NNLO	0.527	0.468	0.00435
NNPDF4.0	NNLO	0.534	0.466	0
NNPDF4.0$_{\text{QED}}$	NNLO	0.533	0.463	0.00433
NNPDF4.0	aN3LO	0.537	0.463	0
NNPDF4.0$_{\text{QED}}$	aN3LO	0.536	0.459	0.00433

- Adding a γ PDF requires momentum:

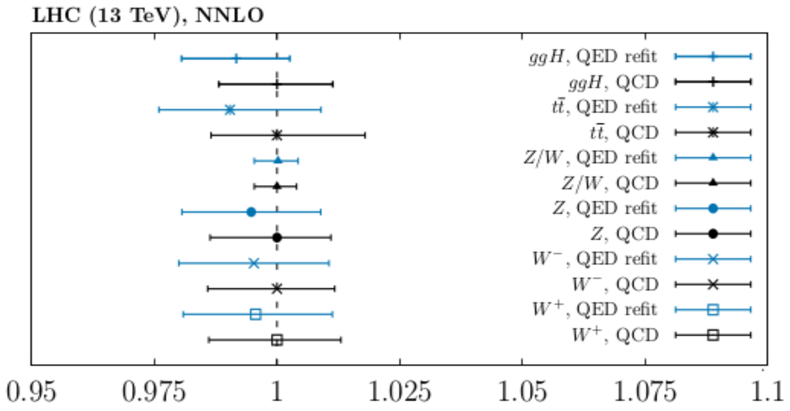
$$\int_0^1 dx(\Sigma(x,Q^2) + g(x,Q^2)) = 1 \Rightarrow \int_0^1 dx(\Sigma(x,Q^2) + g(x,Q^2) + \gamma(x,Q^2)) = 1$$

$\Rightarrow$ Total momentum of QCD partons reduces.

Total momentum by parton type in QCD only fits and once QED has been added!

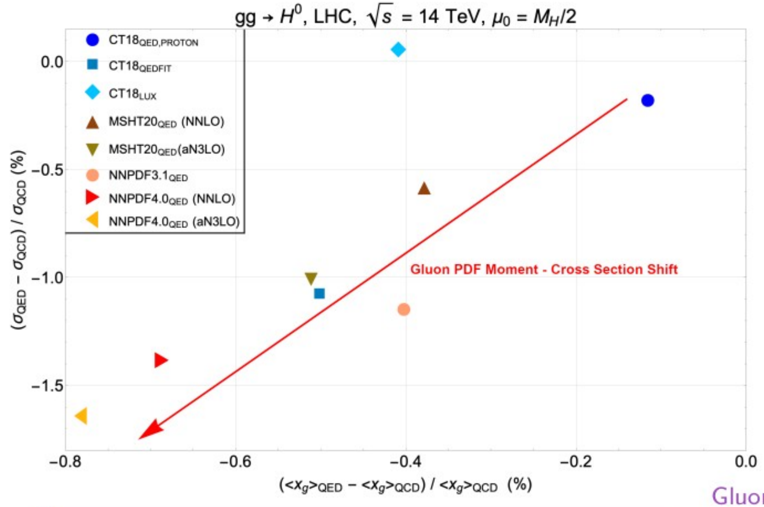
QED PDFs

- Gluon-initiated processes, lower by $\sim 1\%$ in QED case.
- W, Z production reduced by $q \to q\gamma$ splitting, W/Z ratio stable.
- Effect of QED $\lesssim$ PDF uncertainties.
- Uncertainties similar to QCD only.



Xsec % change

ggH

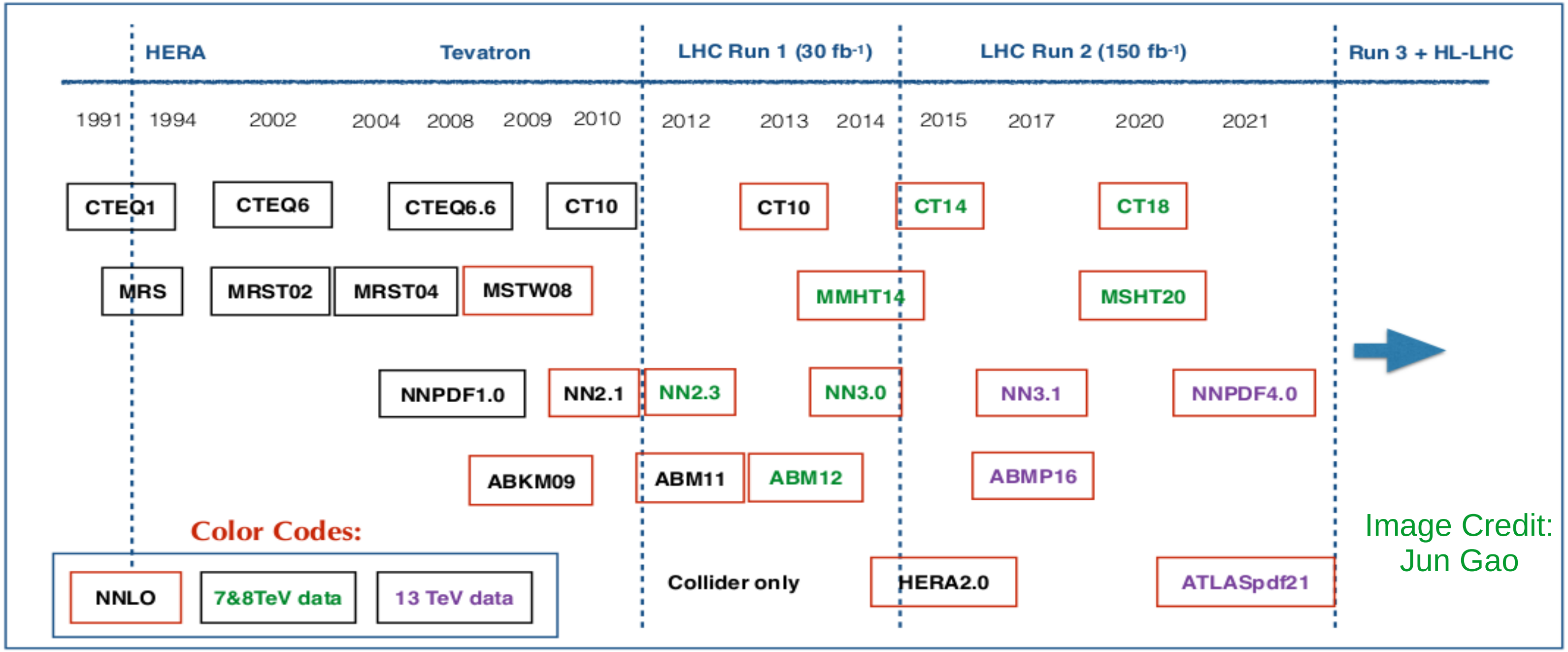


Gluon Momentum % change

QED Effects in PDFs important for precision studies! E.g. DY, Higgs, etc

Global PDF Fitting Groups

- Several different PDF analysis groups – ABM, ATLASPDF, CJ, CT, HERAPDF, JAM, MSHT, NNPDF and others. Not covered all here, naturally more MSHT examples.



Default now

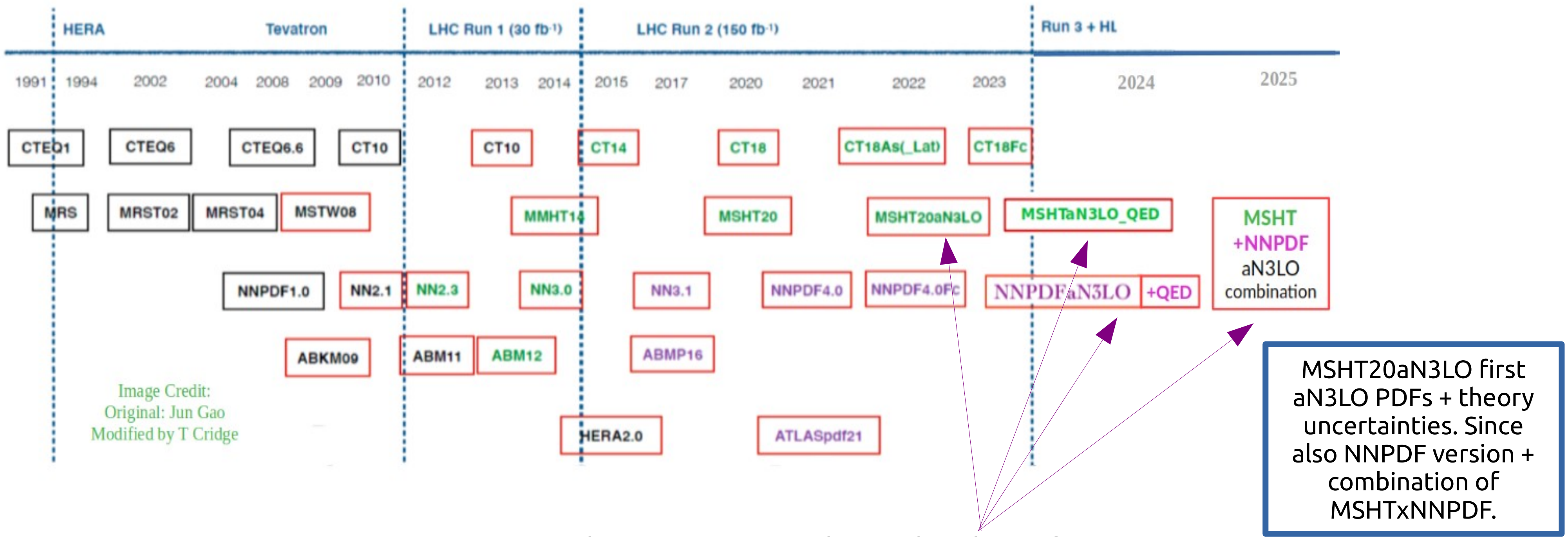
NNLO QCD + NLO EW and latest LHC and other data

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

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

Global PDF Fitting Groups – aN3LO! ③

- Several different PDF analysis groups – ABM, ATLASPDF, CJ, CT, HERAPDF, JAM, MSHT, NNPDF and others. Not covered all here, naturally more MSHT examples.

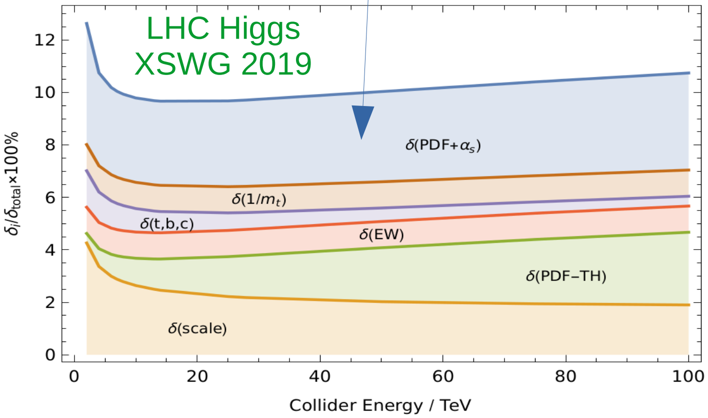


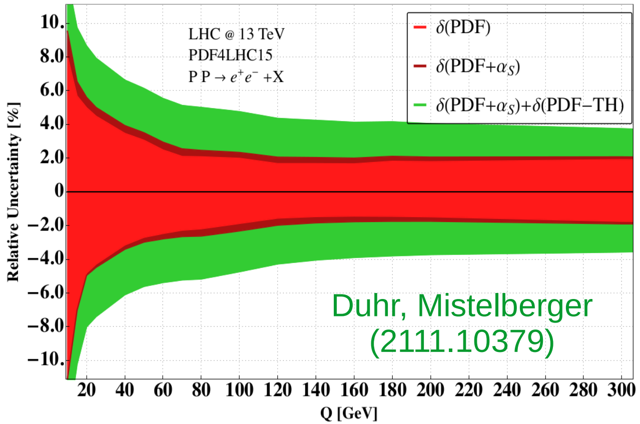
- **"Approximate N3LO" PDFs** have appeared in the last few years!

aN3LO PDFs - Motivation

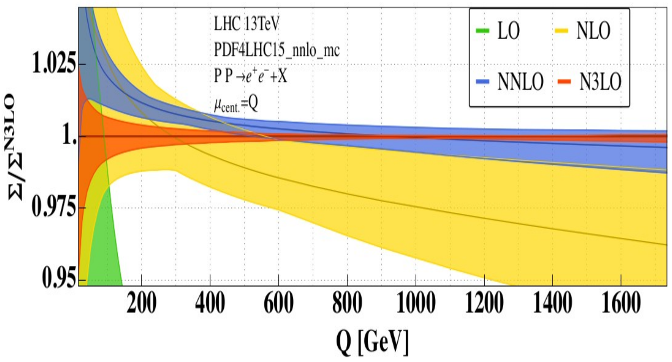
Need for Higher Orders

- N3LO becoming available for some xsecs – Higgs, DY. With only NNLO PDFs, PDF errors limiting + mismatch between xsec and PDF order.





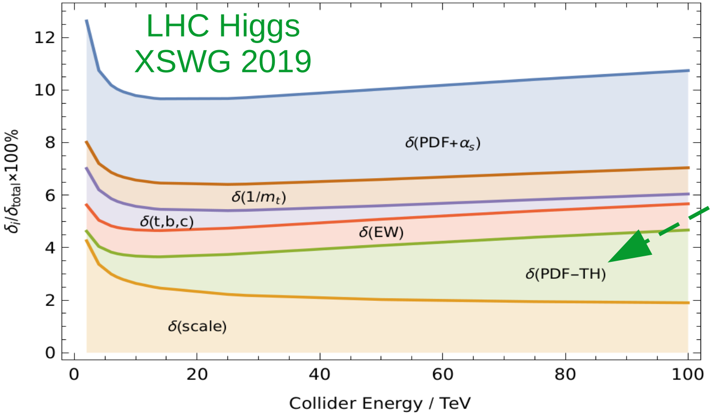
NC DY Z/photon

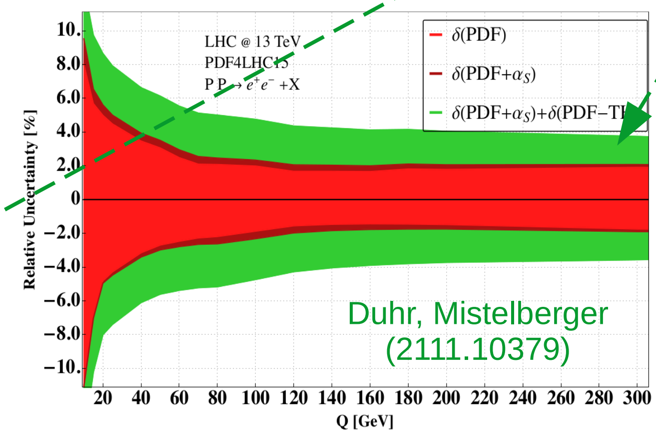


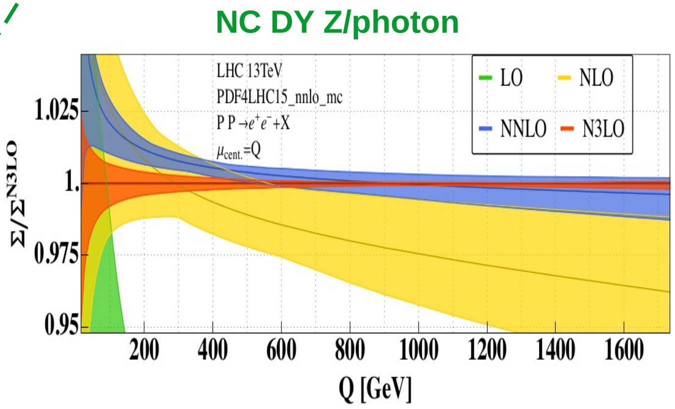
aN3LO PDFs - Motivation

<u>Need for Higher Orders</u>

- N3LO becoming available for some xsecs – Higgs, DY. With only NNLO PDFs, PDF errors limiting + mismatch between xsec and PDF order. → **"PDF-th"** uncertainty.





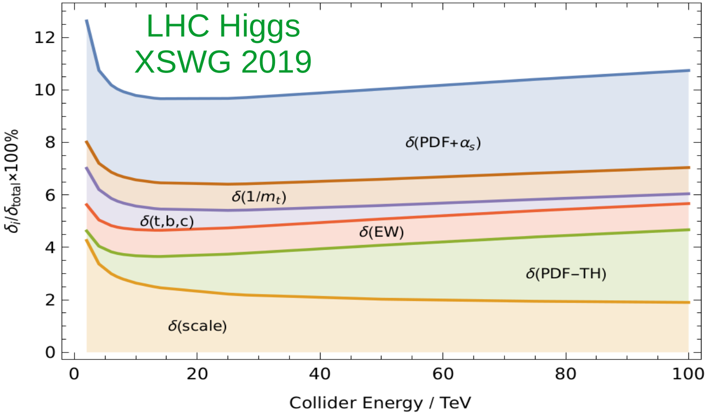


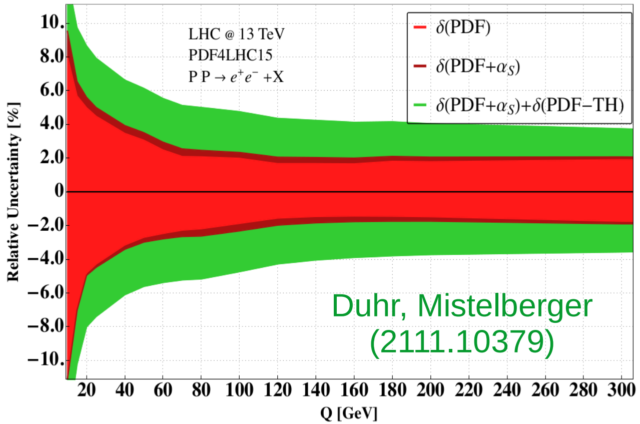
aNLO PDFs - Motivation

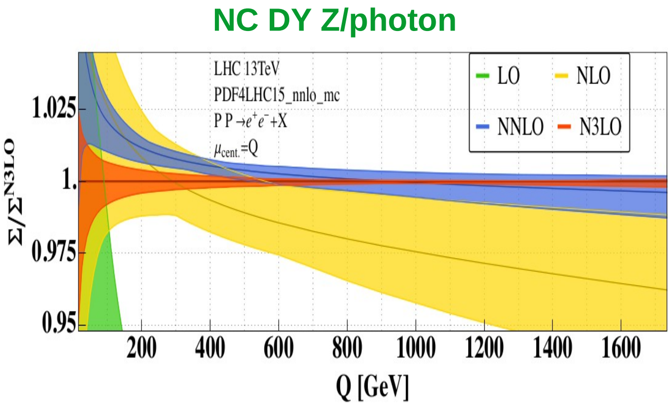
Need for Higher Orders

- N3LO becoming available for some xsecs – Higgs, DY. With only NNLO PDFs, PDF errors limiting + mismatch between xsec and PDF order. → **"PDF-th"** uncertainty.

- Only way to reduce/remove these uncertainties and properly estimate them is obtaining **N3LO PDFs** + including PDF theory uncertainties for missing higher orders.







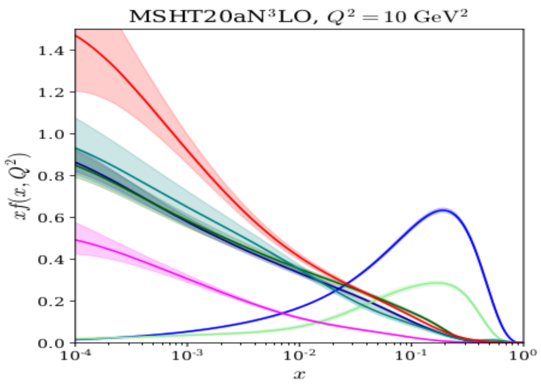
aNN3LO PDFs - Idea

- As PDFs become more precise two issues are more pressing:
  1. Moving to higher orders (N3LO).
  2. Inclusion of theoretical uncertainties.

  $\Rightarrow$ we can address both in one go!

  $\Rightarrow$ **MSHT20aN3LO PDFs**.

J. McGowan et al (2207.04739)



- Idea is to include known N3LO effects already into PDFs and to parameterise remaining unknown pieces via nuisance parameters.
- Variation of these remaining unknown N3LO pieces then provides a theoretical uncertainty within an approximate N3LO fit (aN3LO).

Approximate $\neq$ poorly known!

aN3LO PDFs – What's required?

6

- Full N3LO PDFs need all N3LO pieces for both PDFs and included cross-sections to be known, not yet possible as several pieces missing.
- Still, a lot of information is known already (schematic summary):

Lots of effort by theory community! References in papers

Theory	Utility	Order required	What's known?
1. **Splitting functions** $P^{(3)}_{ab}(x)$	PDF evolution	4-loop	Mellin moments[3–5], leading small-x behaviour[3,6–11], plus some leading large-x in places[3]. *Plus new* [12–15].
2. **Transition matrix elements** $A^{(3)}_{ab,H}(x)$	Transitions between number of flavours in PDFs at mass thresholds	3-loop	Mellin moments[16,17], leading small-x behaviour[18–19], plus some leading large-x in places[19,20]. *Plus new* [21–23].
3. **DIS Coefficient functions** (NC DIS) $C^{VF,(3)}_{H,a}$	Combine with PDFs and Transition Matrix Elements to form Structure Functions (NC DIS)	N3LO	Some approximations to FFNS (low Q^2) coefficient functions at α_S^3 (with exact LL pieces at low x, NLL unknown)[24–26], ZM-VFNS (high Q^2) N3LO coefficient functions known exactly[27]. Therefore GM-VFNS not completely known.
4. **Hadronic Coefficients (K-factors)**	Determine cross-sections at N3LO	N3LO	Very little (none in usable form for PDFs)

aN3LO PDFs - Methodology

6

- Consider PDF fit probability - add N3LO theory and theory uncertainty:

Experimental Nuisance parameters

$$P(T|D) \propto \exp\left(-\frac{1}{2}\sum_{k=1}^{N_{pt}}\frac{1}{s_k^2}\left(D_k - T_k - \sum_{\alpha=1}^{N_{corr}}\beta_{k,\alpha}\lambda_\alpha\right)^2 + \sum_{\alpha=1}^{N_{corr}}\lambda_\alpha^2\right)$$

$$\propto \exp\left(-\frac{1}{2}\sum_{k=1}^{N_{pt}}\frac{1}{s_k^2}\left(D'_k - \tilde{T}_k - \sum_{t=1}^{N_{TNPs}}u_{k,t}\theta'_t\right)^2 + \sum_{\alpha=1}^{N_{corr}}\lambda_\alpha{}^2 + \sum_{t=1}^{N_{TNPs}}\theta_t'^2\right)$$

Theory Nuisance Parameters

- Upgrade theory, $T \to \tilde{T}$ to now contain known N3LO info (aN3LO) and allow to vary by theory nuisance parameters, TNPs - θ'.
- Probes precisely the missing higher order terms.

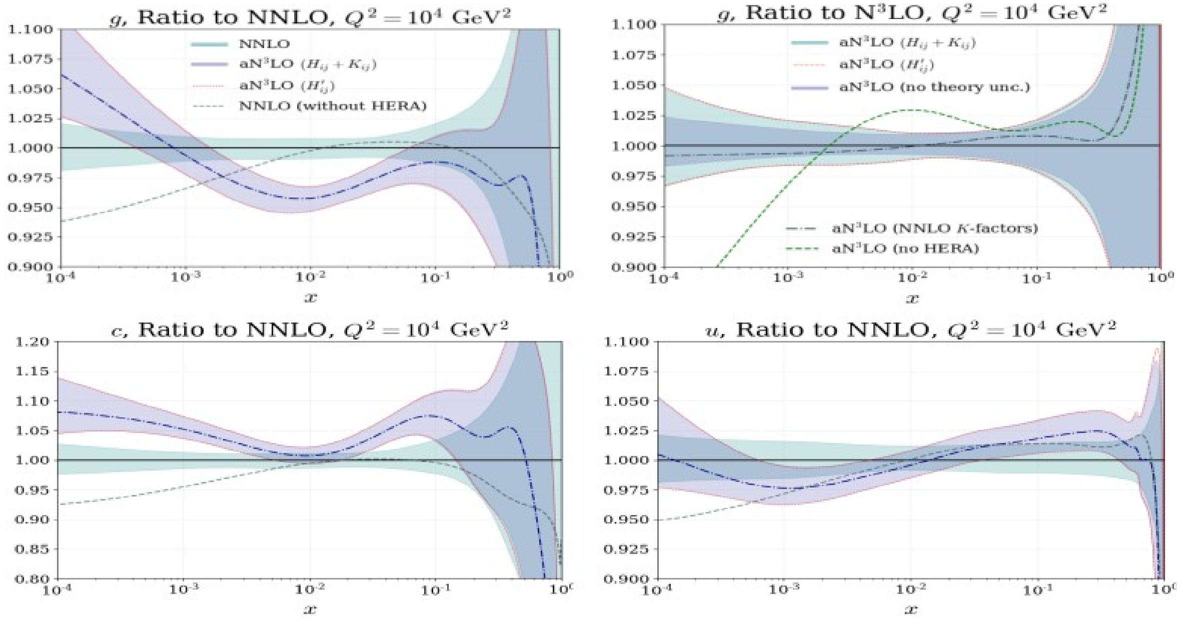
(Applications more widely - e.g. theory uncertainty for Z p_T spectrum and α_S - T.C., G. Marinelli, F. Tackmann 2506.13874.)

aN3LO PDFs – PDF Impacts

- Fit impact: perform aN3LO fit with identical dataset to NNLO:

Total Fit quality χ^2/N_{pts}	LO	NLO	NNLO	aN3LO
	2.57	1.33	1.17	1.14

Smooth fit improvement with order. $\Delta\chi^2 = -154.4$ from NNLO to aN3LO.



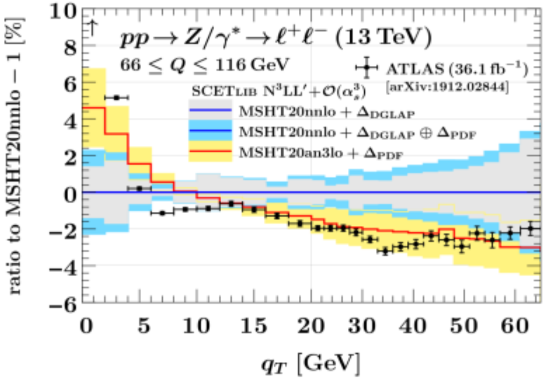
- Gluon PDF raises and its uncertainty increases at low x.
- Heavy quarks also increase across x, purely perturbative.
- Other PDFs much more mildy affected.

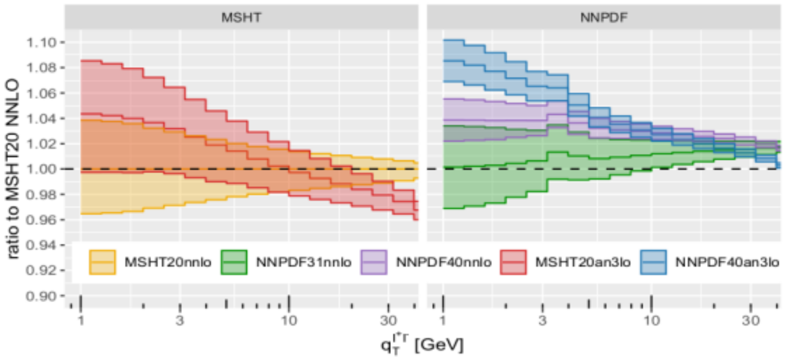
J. McGowan et al (2207.04739)

aN3LO PDFs – Pheno Impacts

6

- MSHT20aN3LO and NNPDFaN3LO PDFs have same impact on shape of q_T spectrum:





aN3LO PDFs improve agreement with data in precision Z at low qT

- Substantial aN3LO PDF effect on N3LL'/N4LL q_T spectrum.

Also (not shown) – better perturbative stability for DY and ggHiggs.

Left: `SCETlib` - Johannes Michel LHC EW WG meeting Sep 2022.

Centre: `CuTe-MCFM` - Tobias Neumann Loops and Legs March 2024

aN3LO QCD + QED PDFs

6

- Can also combine aN3LO QCD PDFs with QED sets for highest possible precision!
- Need to combine aN3LO QCD evolution and $\mathcal{O}(\alpha, \alpha\alpha_S, \alpha^2)$: (bonus!)

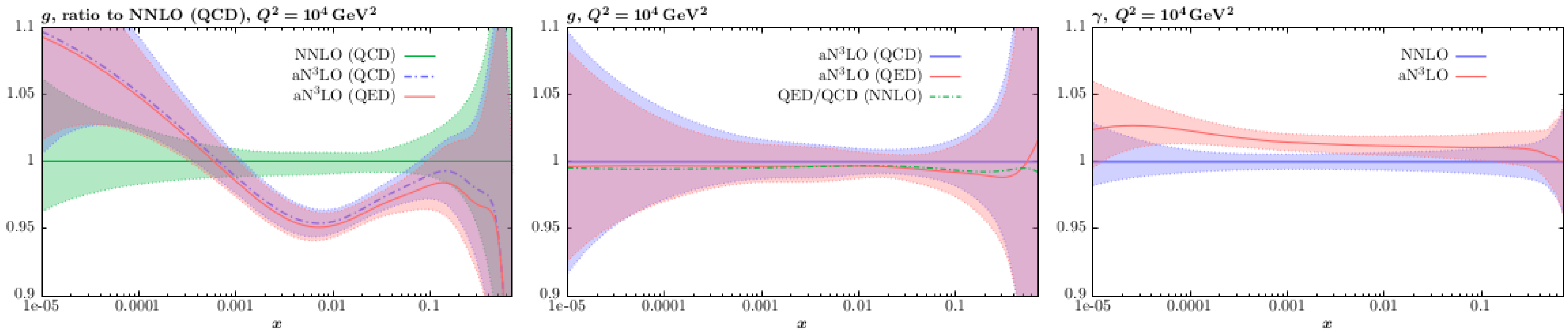
TC et al (2312.07665)

QED $$P_{ij} = \frac{\alpha}{2\pi} P_{ij}^{(0,1)} + \frac{\alpha\alpha_S}{(2\pi)^2} P_{ij}^{(1,1)} + \left(\frac{\alpha}{2\pi}\right)^2 P_{ij}^{(0,2)}$$

NNLO QCD $$+ \frac{\alpha_S}{2\pi} P_{ij}^{(1,0)} + \left(\frac{\alpha_S}{2\pi}\right)^2 P_{ij}^{(2,0)} + \left(\frac{\alpha_S}{2\pi}\right)^3 P_{ij}^{(3,0)}$$

aN3LO QCD $$+ \left(\frac{\alpha_S}{2\pi}\right)^4 P_{ij}^{(4,0)} \,.$$

- Effect of adding QED similar when applied to NNLO and aN3LO.



aN3LO PDFs Comparison to NNPDF ⑥

- Since original MSHT20aN3LO PDFs NNPDF have also produced aN3LO PDFs.
- Similar effects seen, can combine to "MSHT20xNNPDF4.0 aN3LO (+QED)".
- **aN3LO PDFs key for precision Higgs.**

Largest aN^{3}LO effect is on Higgs production in gluon-gluon fusion.

PDF		MSHT	NNPDF	MSHTxNNPDF
ggF xsec shift relative to NNLO	aN3LO	-5%	-2%	-4%
	aN3LO+QED	-6%	-4%	-5%

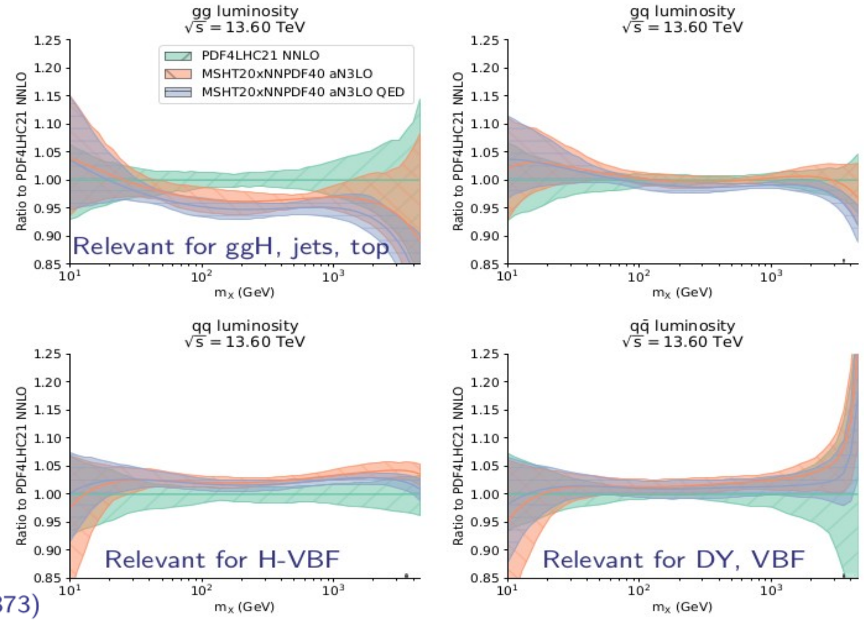
Smaller aN^{3}LO effects for quarks and hence for VBF, ZH, $W^{\pm}H$.

TC,et al

VBF: aN^{3}LO+QED result in +2.5%.

ZH and $W^{\pm}H$: all observe $\approx$ +0.5% from aN^{3}LO+QED.

Consistent trends in all cases between MSHT and NNPDF aN^{3}LO PDFs.



(See TC et al - MSHT and NNPDF - 2411.05373)

aNLO PDF updates

- New theory ingredients calculated for aN3LO PDFs.
- Can add and see impact.
- Find MSHT and NNPDF results become closer still.
- E.g. Gluon moves together, both see reduction of $\sim$3%

New in MSHT

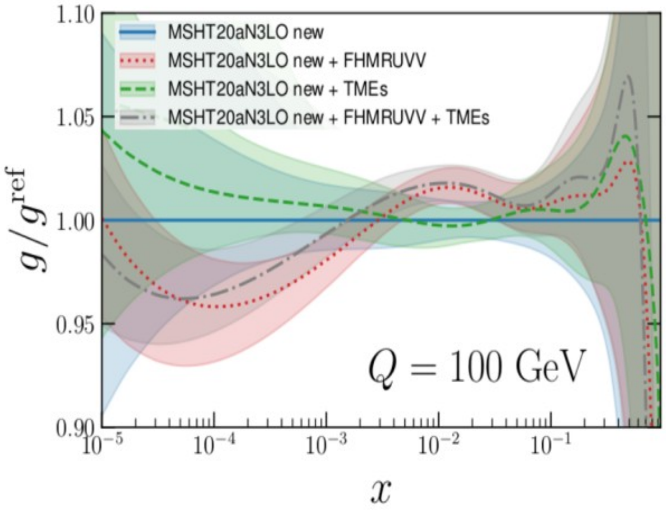
higher moments for singlet splitting funcs.

[PLB 842 (2023) 137944; 846 (2023) 138215; 848 (2024) 138351]

[PLB 856 (2024) 138906; 860 (2025) 139194]

computation of $A_{gg,H}$, $A_{H,g}$, $A_{qq,H}$, $A_{qg,H}$

[JHEP 12 (2022) 134; NPB 999 (2024) 116427]



[PoS DIS2025 (2025) 027, see also T. Cridge's talk DIS2026]

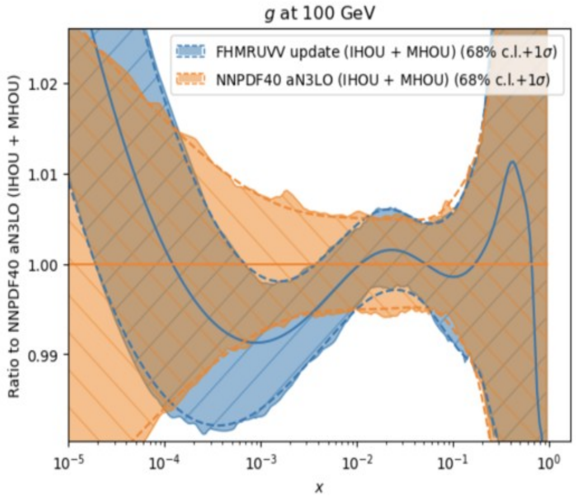
New in NNPDF

higher moments for singlet splitting funcs.

[PLB 856 (2024) 138906; 860 (2025) 139194]

improved parametrisation of $a_{H,g}^{(3)}$

[PLB 854 (2024) 138713]

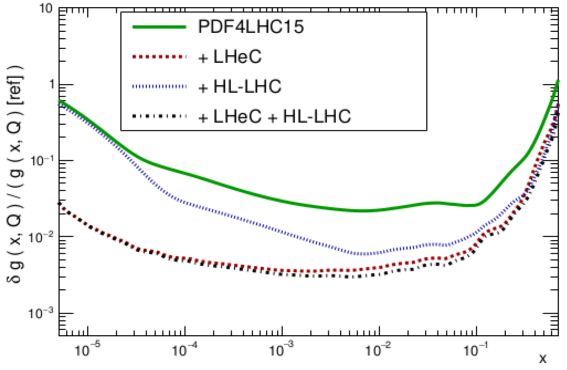


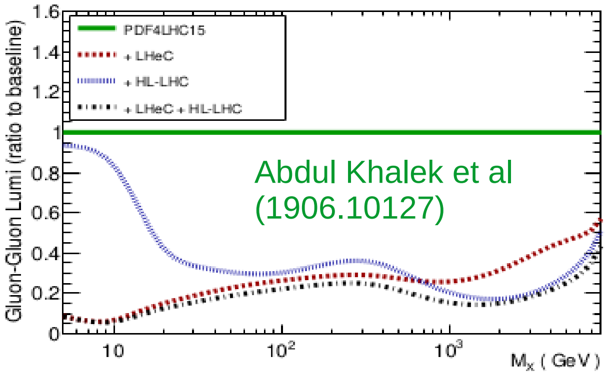
[See G. Magni's talk PDF4LHC 2024]

PDF Constraints – Future LHC

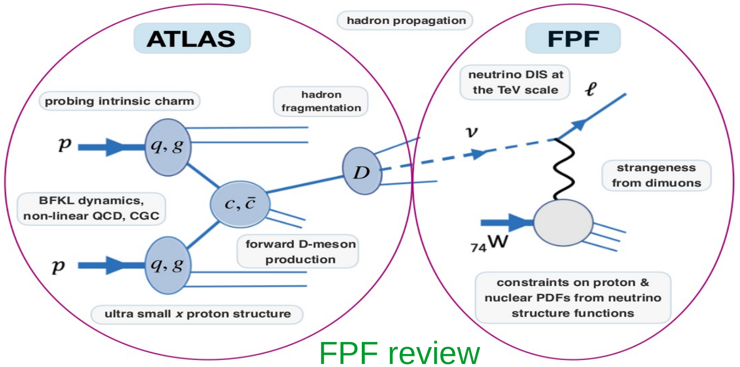
- **HL-LHC** – Reduce PDF uncertainties where processes currently statistically limited/coverage can be extended, e.g. high x gluon and gg luminosity.
- **LheC** – Inclusive/Semi-inclusive DIS data constrain intermediate/small x.

LHeC review (2007.14491)





Abdul Khalek et al (1906.10127)



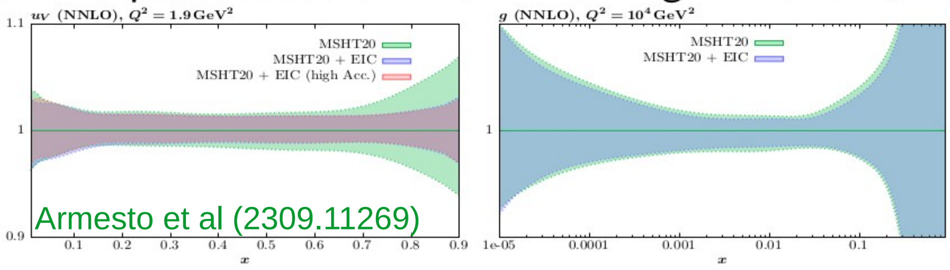
FPF review (2203.05090)

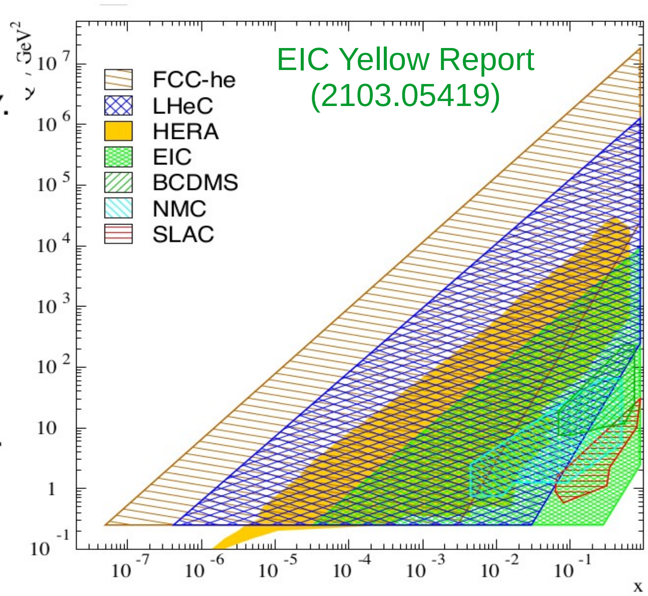
- **FPF** – Very forward neutrino production → intrinsic charm at high x, very low x gluon dynamics. Then Neutrino CC DIS → flavour separation.

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

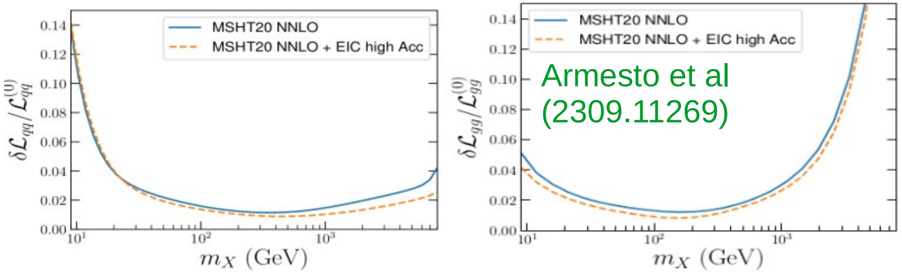
PDF Constraints – Future EIC

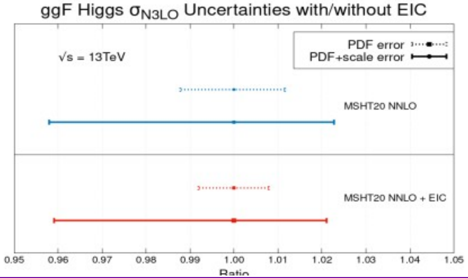
- Electron Ion Collider to run at Brookhaven in 2030s.
- Impact of EIC pseudodata on MSHT20 - *high x lower Q^2 sensitivity.*

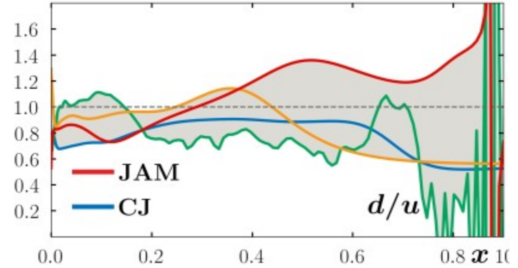




- Effect on up valence larger due to charge-squared γ coupling in DIS.
- Gluon uncertainty nonetheless reduced across range of x.
- Impact on luminosity uncertainties - $\delta\mathcal{L}_{qq}$ reduced at high x.
- $\delta\mathcal{L}_{gg}$ reduced across x causes smaller PDF uncertainty for $gg \to H$.







EIC Yellow Report (2103.05419)

7. Strong Coupling and mtop Determination (if time!)

Strong Coupling sensitivity

7

- PDFs sensitive to $\alpha_S(M_Z^2)$ → can determine it from PDF fit!
- $\alpha_S(M_Z^2)$ sensitivity in PDF fit comes from:
  - Direct $\alpha_S(M_Z^2)$ dependence in coefficient functions.

$$F(x,Q^2) = \sum_{i=q,\bar{q},g} \left[C_i(\alpha_s(Q^2)) \otimes f_i(Q^2) \right](x)$$

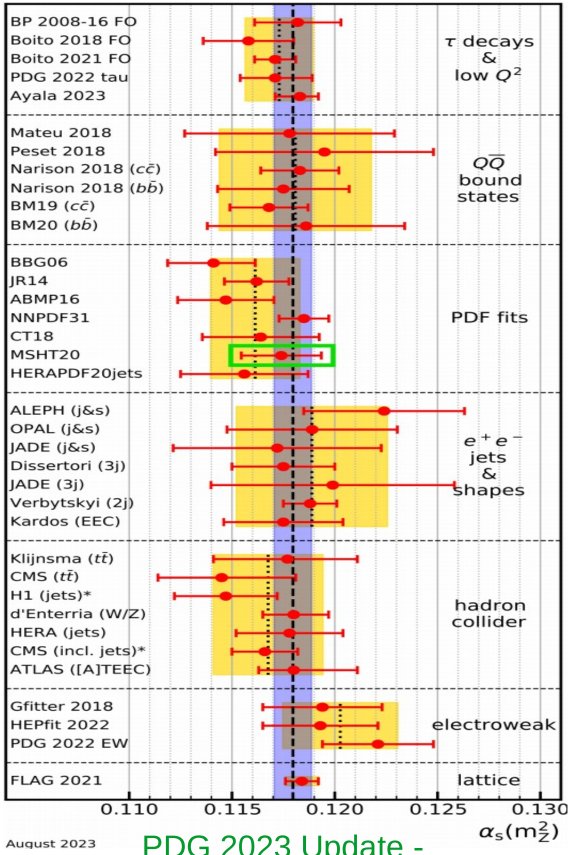
$$\mathbf{C}(\alpha_S) = \alpha_S^i[\mathbf{C}_0 + \alpha_S\mathbf{C}_1 + \alpha_S^2\mathbf{C}_2 + \alpha_S^3\mathbf{C}_3 + \ldots] \quad \chi^2$$

  - Indirect $\alpha_S(M_Z^2)$ dependence through PDF evolution.

$$\frac{d\boldsymbol{f}}{d\ln\mu_f^2} \equiv \frac{d}{d\ln\mu_f^2}\begin{pmatrix}\Sigma \\ g\end{pmatrix} = \begin{pmatrix}P_{qq} & n_f P_{qg} \\ P_{gq} & P_{gg}\end{pmatrix} \otimes \begin{pmatrix}\Sigma \\ g\end{pmatrix} \equiv \boldsymbol{P}\otimes\boldsymbol{f}$$

$$\boldsymbol{P}(x,\alpha_s) = \alpha_s\boldsymbol{P}^{(0)}(x) + \alpha_s^2\boldsymbol{P}^{(1)}(x) + \alpha_s^3\boldsymbol{P}^{(2)}(x) + \alpha_s^4\boldsymbol{P}^{(3)}(x) + \ldots$$

- Perform fit at different values of strong coupling and determine best fit χ² → best fit $\alpha_S(M_Z^2)$



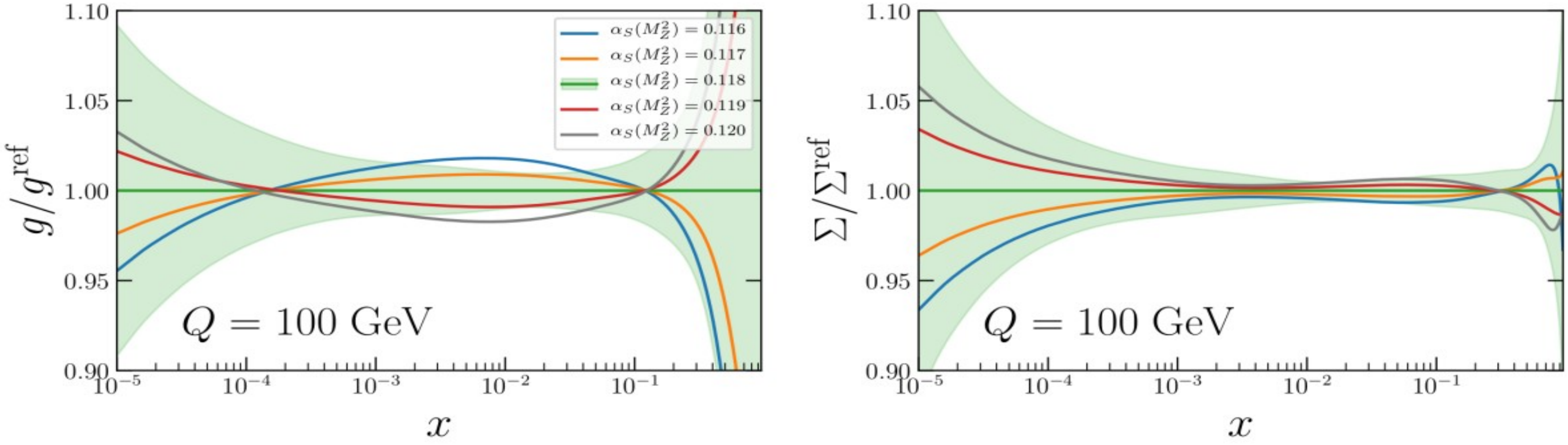
PDG 2023 Update - 2312.14015.

Alphas correlation with PDFs…

MSHT20 PDF α_S dependence

Forte, Kassabov: 2001.04986

- Correlations between PDFs and $\alpha_S \Rightarrow$ necessity of global fit.



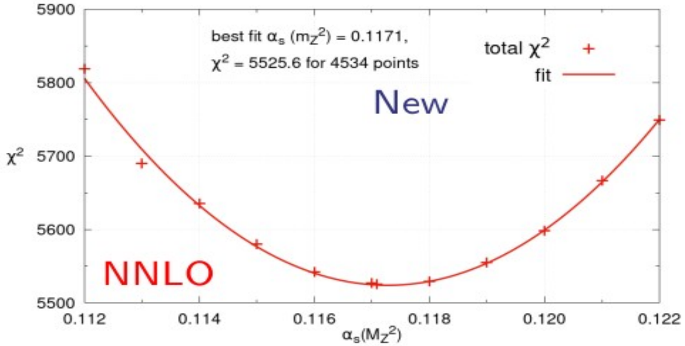
- Changes generally within PDF uncertainties for $\Delta\alpha_S(M_Z) \approx \pm 0.001$.
- Gluon anti-correlated with $\alpha_S(M_Z^2)$ for $x \lesssim 0.1$ as maintains $dF_2/dQ^2 \sim \alpha_S g$. Implies correlated at high $x \gtrsim 0.1$ by momentum sum rule.

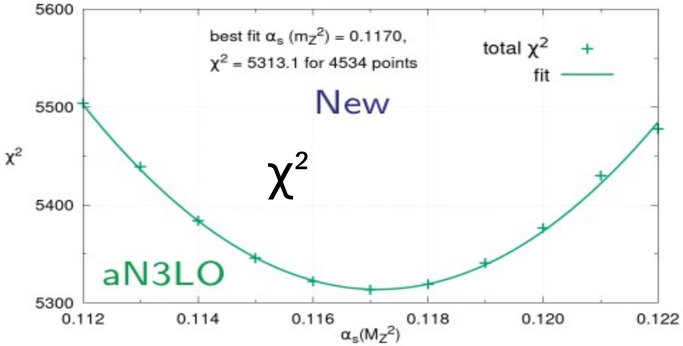
Strong Coupling sensitivity

- Good perturbative convergence of α_S determination.

$$\alpha^{new}_{S,\mathrm{NNLO}}(M_Z^2) = 0.1171 \qquad \alpha^{new}_{S,\mathrm{aN3LO}}(M_Z^2) = 0.1170$$

Previous NNLO result in PDG!





Nice Quadratic χ^2 profile ✓

Missing Higher Order Uncertainties now included, in particular causes some LHC bounds to weaken as unknown N3LO K-factors.

- Determine bounds on strong coupling exactly as those on PDFs (with tolerance):

$$\alpha_{S,\mathrm{NNLO}}(M_Z^2) = 0.1171 \pm 0.0014$$

Consistent with World Average of 0.1180 ± 0.0009.

$$\alpha_{S,\mathrm{aN3LO}}(M_Z^2) = 0.1170 \pm 0.0016$$

Consistent with (NNLO) World Average of 0.1180 ± 0.0009.

TC et al – 2404.02964

Alphas bounds – aN3LO

MSHT20 α_S bounds - aN3LO

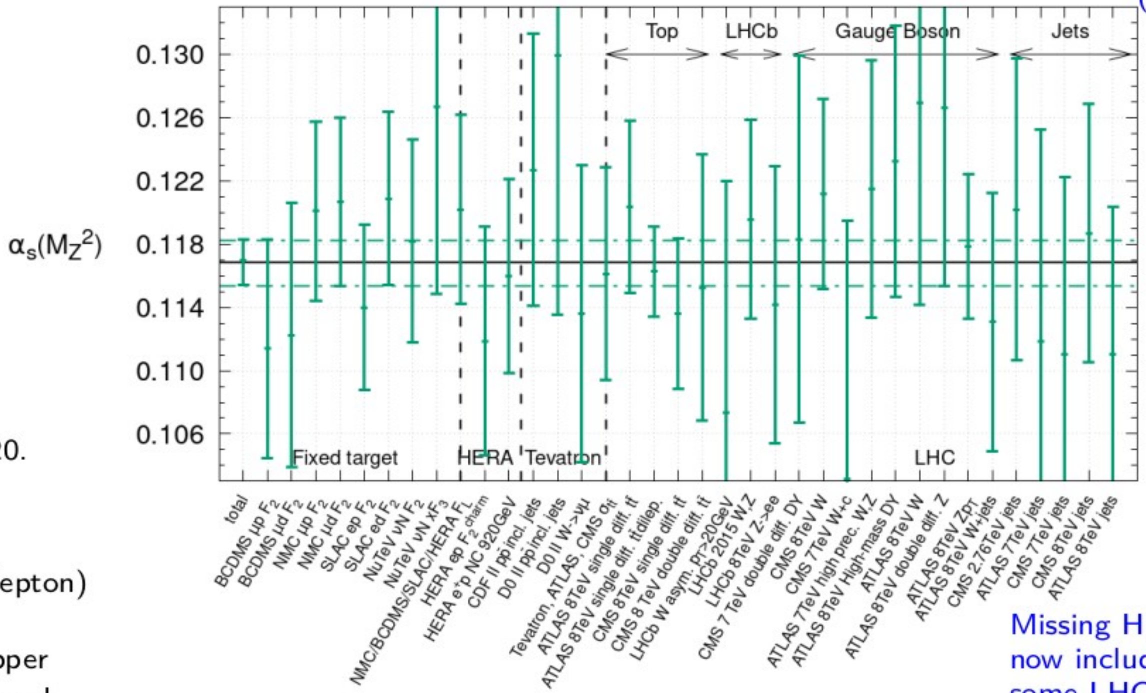
TC et al – 2404.02964

Consistent with α_S bounds seen in previous studies, and between orders (NNLO and aN3LO).

BCDMSp data strongest constraint upwards: $\Delta\alpha_S(M_Z^2) = +0.0013$.

F_2^c provides upwards bound of: $\Delta\alpha_S(M_Z^2) = +0.0020$.

CMS and ATLAS (dilepton) $t\bar{t}$ single diff. would give slightly higher upper α_S bounds, but not used.



SLAC deuteron data gives lower bound: $\Delta\alpha_S(M_Z^2) = -0.0016$.

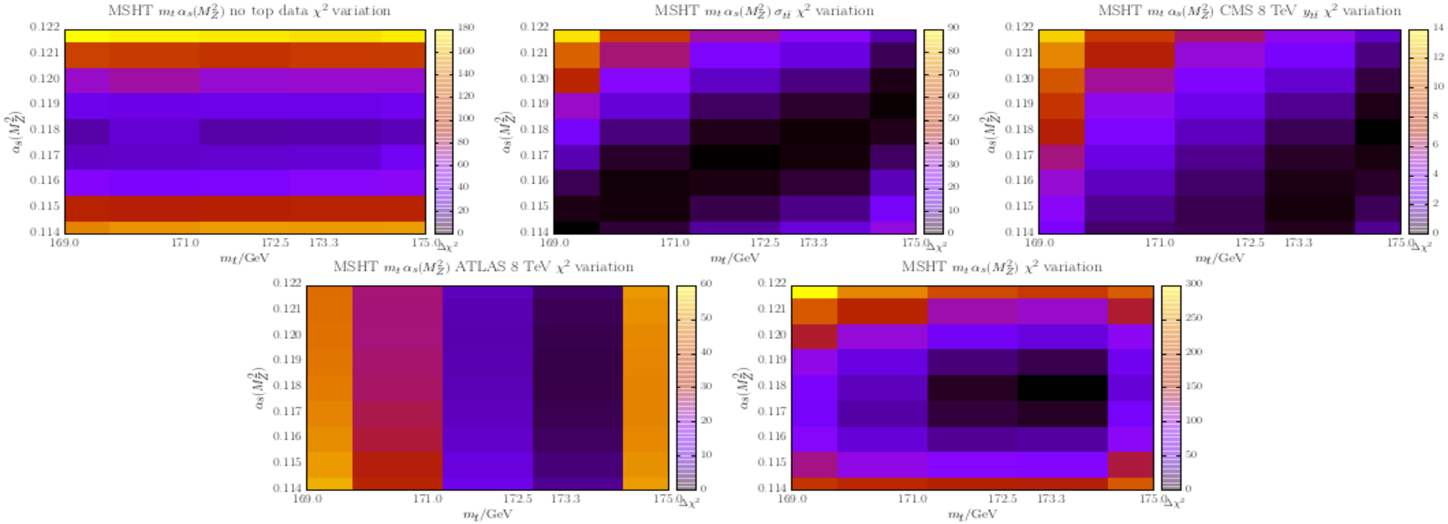
NMC deuteron, ATLAS 8 TeV Z both give lower bounds of $\Delta\alpha_S(M_Z^2) = -0.0017$.

Missing Higher Order Uncertainties now included, in particular causes some LHC bounds to weaken as unknown N3LO K-factors.

PDF Constraints – Top data - mtop

7

- How much correlation between m_t and α_S is there in a PDF fit?



Top data included also sensitive not only to gluon PDF but also strong coupling and top mass!

TC, M. Lim (2306.14885)

- Without top data, no m_t sensitivity and no m_t-α_S correlation.
- Total $\sigma_{t\bar{t}}$ data and rapidity differential data show significant m_t-α_S correlation. Reducing m_t and $\alpha_S \Rightarrow$ cross-sections $\approx$ unchanged.
- Data differential in p_T or m_{tt} much less correlated.
- Overall at level of total fit, limited correlation seen (only done at NNLO so far) $\Rightarrow$ can extract m_t at fixed α_S.

TC and M.A. Lim: arXiv:2306.14885.

Can therefore also fit top (pole) mass, MSHT obtained:

$$m_t^{\text{pole}} = 173.0 \pm 0.6 \text{ GeV}$$

8. Conclusions

T. Cridge wishes to thank the Royal Society for support
through his Royal Society University Research Fellowship: Grant URF\R1\251540..

Summary

- Precise and accurate PDFs are vital for physics at the LHC and beyond.
- Challenges faced on experimental, methodological and theoretical fronts.
- Experimentally - understanding of experimental uncertainties, tensions.
- Methodological - stress testing of methodology, parameterisation, uncertainties.
- Theoretically - higher orders in QCD, QED corrections, inclusion of theory uncertainties.
- Much work ongoing in PDF community, welcome collaboration with exp/theory colleagues.
- Complementarity of different groups is greatly beneficial for these aims.

Summary

- Precise and accurate PDFs are vital for physics at the LHC and beyond.
- Challenges faced on experimental, methodological and theoretical fronts.
- Experimentally - understanding of experimental uncertainties, tensions.
- Methodological - stress testing of methodology, parameterisation, uncertainties.
- Theoretically - higher orders in QCD, QED corrections, inclusion of theory uncertainties.
- Much work ongoing in PDF community, welcome collaboration with exp/theory colleagues.
- Complementarity of different groups is greatly beneficial for these aims.

Thankyou! Any Questions?

Backup

T. Cridge wishes to thank the Royal Society for support
through his Royal Society University Research Fellowship: Grant URF\R1\251540..

3. Challenges and Developments

4. Experimental Challenges

- Precision Data
- Data Tensions
- Correlated Systematics

Interconnected

Interconnected

Interconnected

5. Methodological Challenges

- New algorithms/methods
- PDF Tools
- Additional constraints
- Uncertainty Prescriptions
- Theory grids at NNLO and beyond

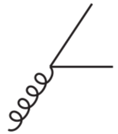
6. Theoretical Challenges

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

Heavy quarks in PDFs

4

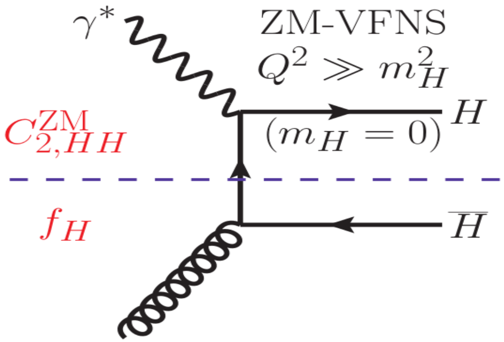
- What about heavy quarks? No parameterisation of heavy quarks?



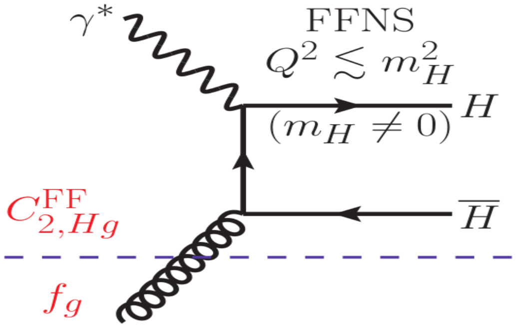
- No → at input we have $Q_0 \sim 1\text{GeV} < m_c, m_b$.
- Heavy quarks perturbatively generated by gluon splittings to quark-antiquark pairs.
- It's a bit more complicated in structure functions. Two regimes:

High $Q^2 \gg m_b^2, m_c^2$

Zero-Mass Variable Flavour Number Scheme (ZMVFNS)



Lower $Q^2 \sim m_b^2, m_c^2$



Fixed Flavour Number Scheme (FFNS)

- Connect both regimes to get description over whole Q^2 range – General Mass Variable Flavour Number Scheme (GMVFNS). Measurement of F_2^c, F_2^b, constrains heavy quarks.

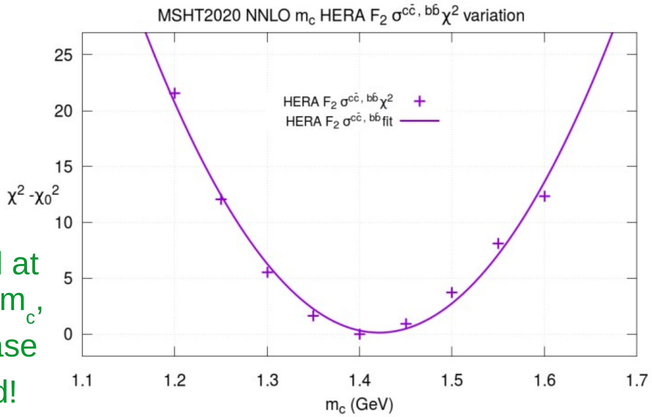
Heavy quarks in PDFs (HERA)

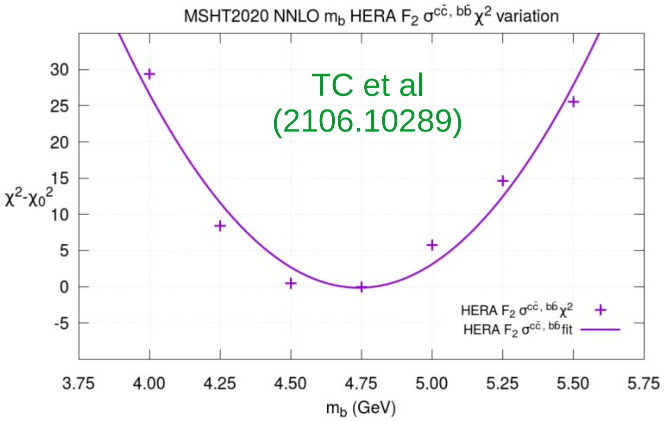
4

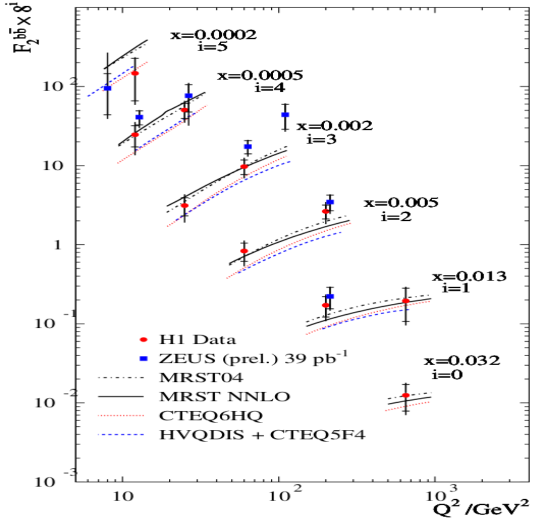
- PDFs then have **3** flavours up to m_c, **4** between there and m_b, and **5** flavours above that – as expected.
- PDFs "switched on" at m_c, m_b by tranisition matrix elements.
- Good fit to HERA heavy flavour structure function data with GMVFNS (right).
- Also induces sensitivity to m_c, m_b into PDF fit.

$$f_\alpha^{n_f+1}(x,Q^2) = \left[A_{\alpha i}(Q^2/m_h^2) \otimes f_i^{n_f}(Q^2)\right](x)$$

PDFs provided at different m_c, m_b in case needed!







PDF Constraints – $\bar{d}/\bar{u}$ at high x

4

- Also have some constraint from Fixed Target DIS earlier. Consider difference of proton and neutron structure functions (Gottfried Sum Rule):

$$S_G = \int_0^1 [F_2^p(x) - F_2^n(x)]dx/x = \frac{1}{3}\int_0^1 [u(x) + \bar{u}(x) - d(x) - \bar{d}(x)]$$

$$= \frac{1}{3}\int_0^1 [u_V(x) - d_V(x) + 2(\bar{u}(x) - \bar{d}(x)] = \frac{1}{3} + \frac{2}{3}\int_0^1 [\bar{u}(x) - \bar{d}(x)]$$

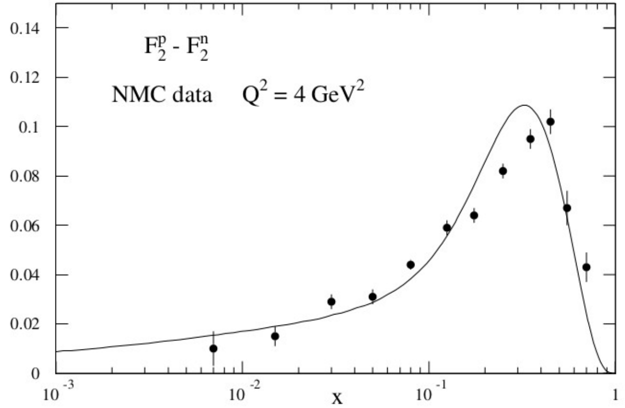
- S_G = 1/3 if SU(2) sea flavour symmetry.
- Measured by NMC: S_G = 0.235±0.026 <1/3

  → Integral of $\bar{u}$-$\bar{d}$ -ve, i.e. $\bar{d}>\bar{u}$.

  → Like most theoretical models!

- What does Fixed Target DY at E866 tell us then?

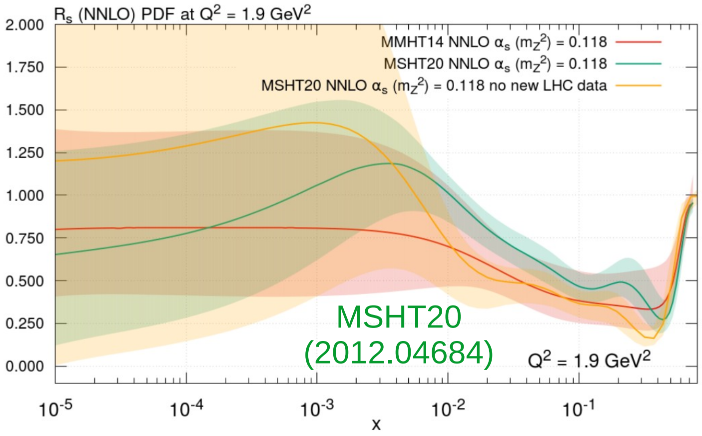
Assumes measure whole x range → not true!

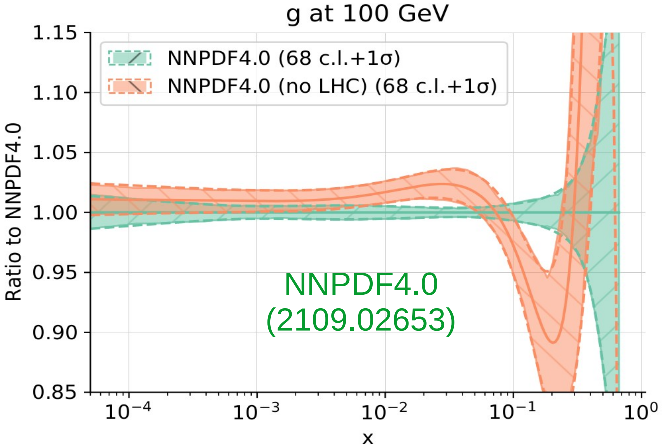


Experimental Advances

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.





Data set	$N_{\rm pts}$	NLO $\chi^2/N_{\rm pts}$	NNLO $\chi^2/N_{\rm pts}$
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)

Experimental Challenges

4

<u>Confronting Precise Data</u> *Can reflect experimental, methodological or theoretical issues!*

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

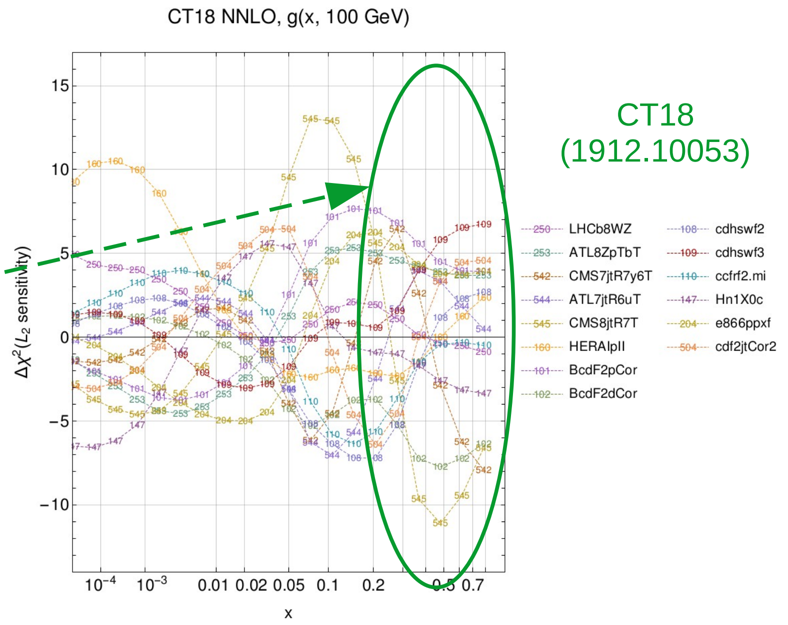
- Two frequent (experimental/methodological) causes:

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.



CT18 (1912.10053)

Experimental Challenges

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.

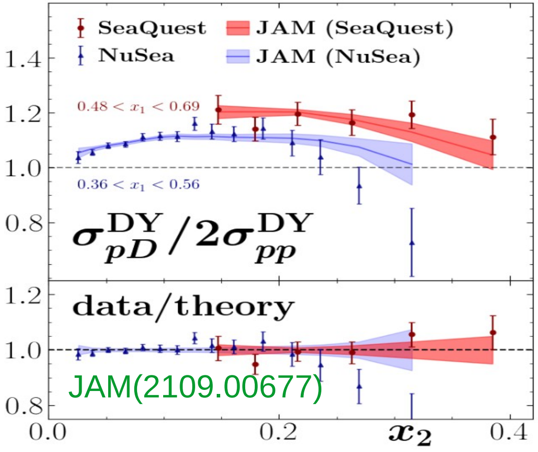
  - Gottfried sum rule – NMC found $\int_0^1 [F_2^p(x) - F_2^n(x)]dx/x < \frac{1}{3}$

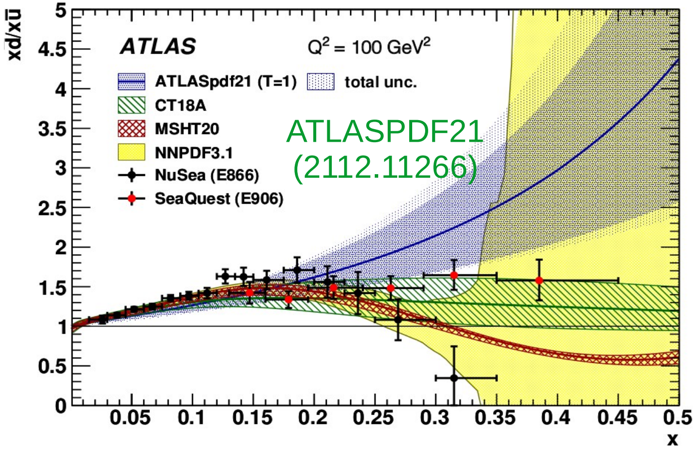
  - E866/NuSea data favoured $\bar{d}<\bar{u}$ at high x.

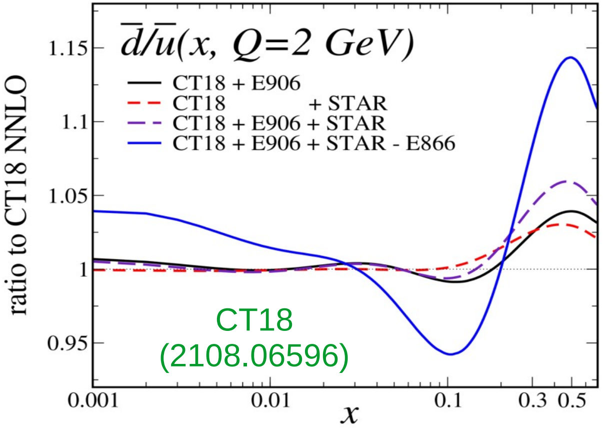
  - New Seaquest/E905 data instead favour $\bar{d}>\bar{u}$.

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

How should this be interpreted in a global PDF fit?







Experimental Challenges

4

Confronting Precise Data

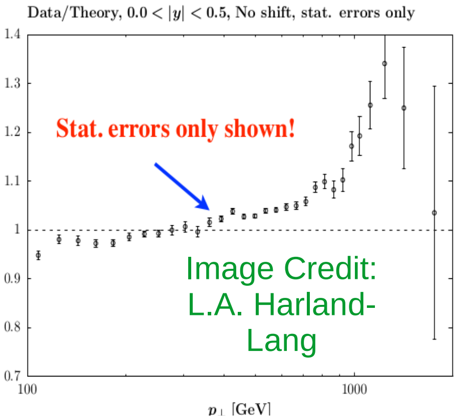
2) **Correlated Systematics** – Issues occur for 2 of the 3 dataset high x gluon data types – top and jets. Consider ATLAS 7 TeV jets:

ATLAS 7TeV jets (1410.8857)

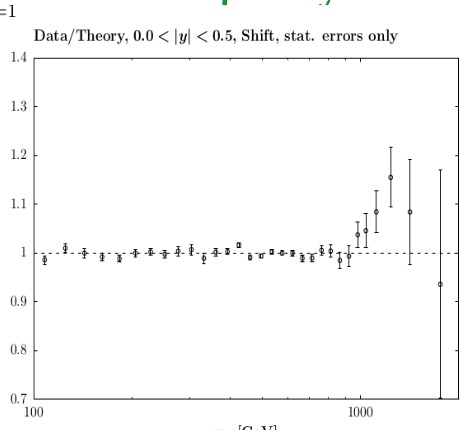
ATLAS 8TeV jets (1706.03192)

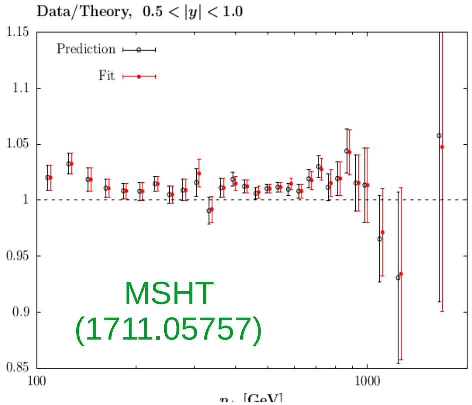
Can fit individual rapidity bins

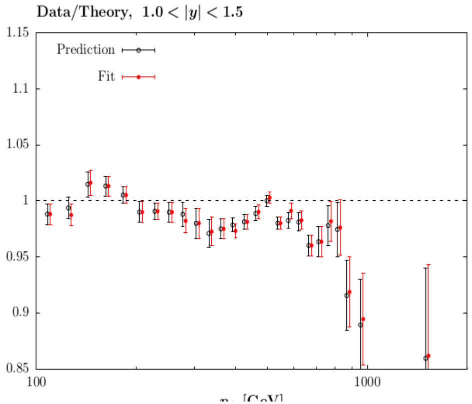
But not multiple rapidity bins!



Shift







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

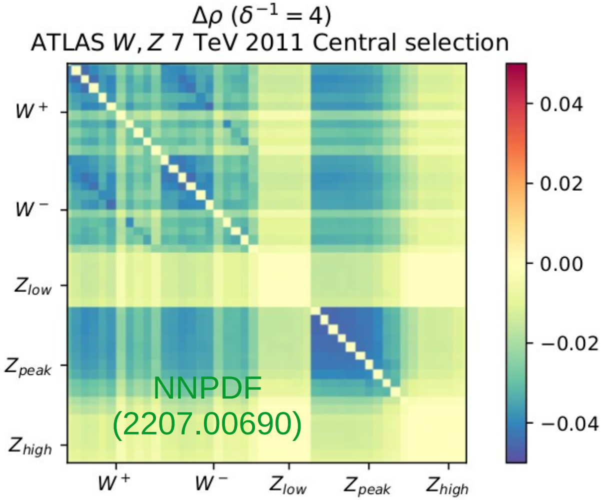
Experimental Challenges

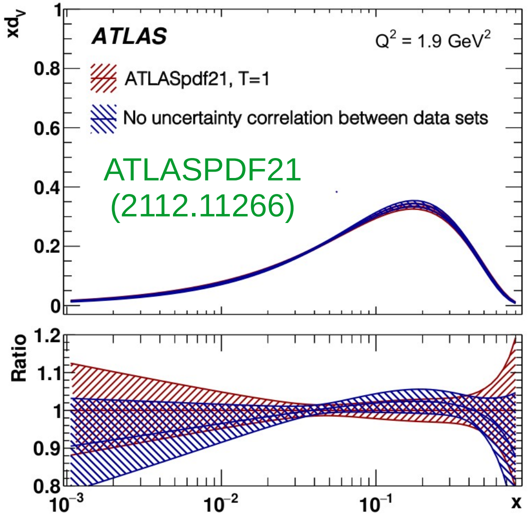
Confronting Precise Data

[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)
- Proposal to mitigate these systematic correlation issues by regularisation of the covariance matrix – NNPDF.
- Recent efforts to consider correlations between experiments – ATLASPDF21.

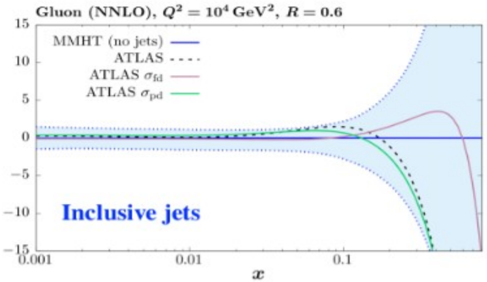




Exp Challenges - systematics

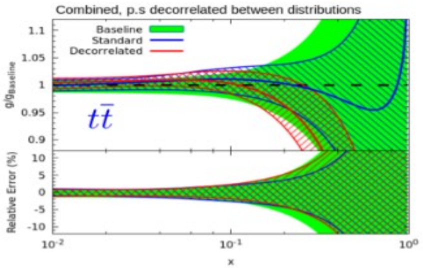
4

- As LHC datasets become increasingly systematics dominated, we become sensitive to precise treatment of systematics + their correlations.
- "Two point systematics" where two MCs/models used and difference taken as fully correlated are particularly problematic.



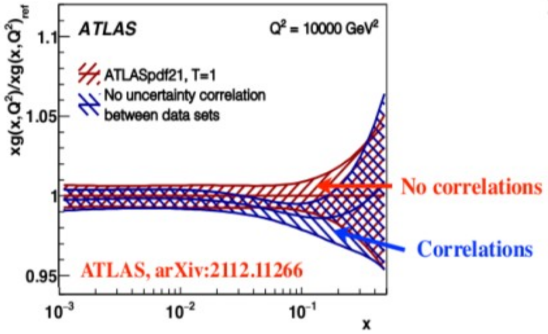
S. Bailey et al., arXiv:2012.04684

n_{dat}		default	part. decorr.	full decorr.
140	χ^2/N_{pts}	1.89	1.28	0.83



S. Bailey and LHL, arXiv:1909.10541

n_{dat}		default	stat. uncorr.	p.s. uncorr
25	χ^2/N_{pts}	7.00	3.28	1.80



- Currently investigating better correlation models...
- But how precisely do we know these systematic errors anyway?

$\Rightarrow$ "Errors on errors".

Exp Challenges - systematics

4

- "Error on error" framework - how uncertain are estimates of error.
- Model variances of uncertainty as Gamma-distributed:

$$L(\mu,\theta) = P(y|\mu,\theta) \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma_{u,i}} \exp(-(u_i-\theta_i)^2/2\sigma_{u,i}^2) \frac{\beta_i^{\alpha_i}}{\Gamma(\alpha_i)} v_i^{\alpha_i-1} \exp(-\beta_i v_i)$$

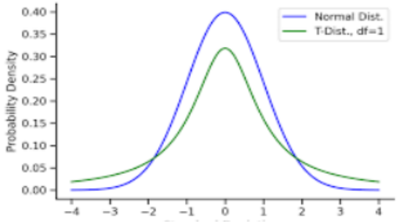
where $\alpha_i = 1/(4r_i^2)$, $\beta_i = 1/(4r_i^2\sigma_{u,i}^2)$,
r_i is relative uncertainty on systematic error $\Rightarrow$ "error on error".

- Model equivalent to Student's t-distribution:

$$L(\mu,\theta) = P(y|\mu,\theta) \prod_{i=1}^{N} \frac{\Gamma((\nu_i+1)/2}{\sqrt{\nu_i\pi}\Gamma(\nu_i/2)} (1 + \frac{t_i^2}{\nu_i})^{-(\nu_i+1)/2)}$$

where $t_i = (u_i - \theta_i)/\sqrt{\nu_i}$, $\nu_i = 1/(2r_i^2)$,

- So we can treat the nuisance parameters for correlated systematics as t-distributed!



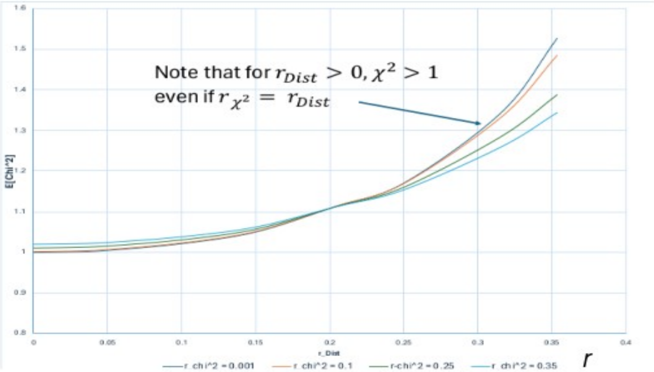
"Errors on errors" - G. Cowan 1809.05778.

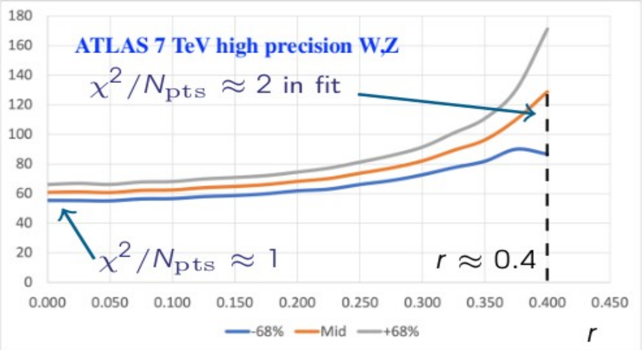
See M. Reader DIS24 for more details.

Exp Challenges - systematics

Preliminary!

- Taking Student's t-distribution compute expectation value of $\chi^2/N_{\rm pts}$ as r ("error on error") increases.
- $E[\chi^2/N_{\rm pts}]$ increases to > 1 as r increases, particularly if assuming $r = 0$ in fit, but even if taking "correct" $r \neq 0$.
- Would explain why (at least partly) PDF fits observe $\chi^2/N_{\rm pts} > 1$.
- Perhaps $T^2 \sim 2-5$ could result from such uncertainties?
- How does this impact χ^2/N? - Try on precision ATLAS 7 TeV W, Z data:
- Observed $\chi^2/N_{\rm pts} \approx 2$, requires $r \approx 0.4$.





M. Reader 2408.12922 and MSHT, Reader et al, in preparation.

Methodological Challenges

5

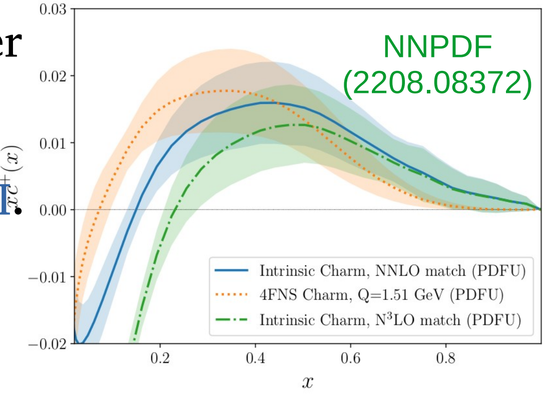
Is it there and can we see it?

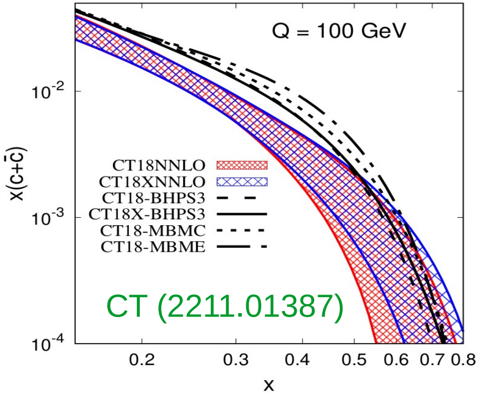
Intrinsic Charm (IC)

- 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.
- 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.





Methodological Challenges

5

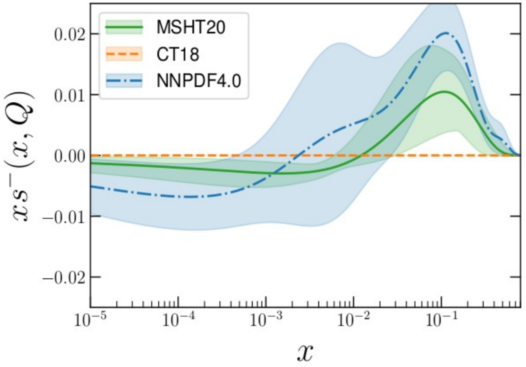
More info on positivity or otherwise of MS PDFs:
NNPDF (2006.07377)
Collins et al (2111.01170)

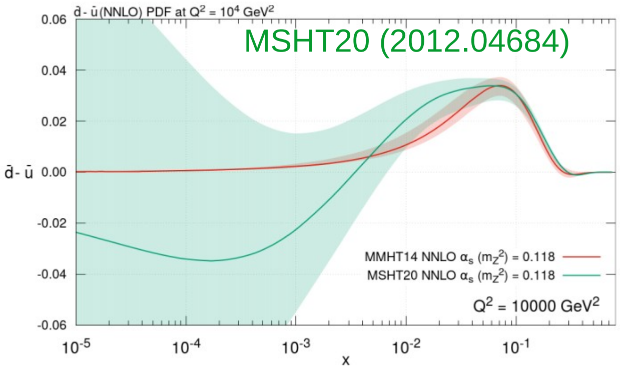
Additional Constraints

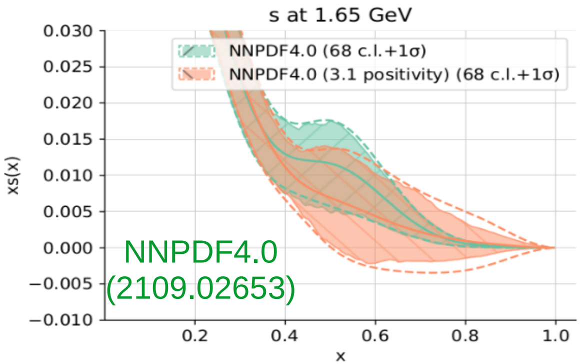
- In order to ensure physical PDFs, often additional constraints are added. Many different types and methods:

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

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.







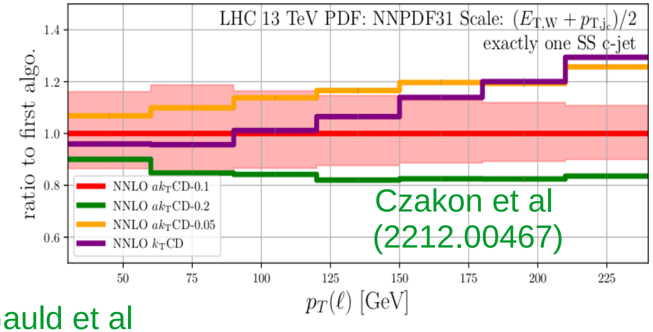
Methodological Challenges

5

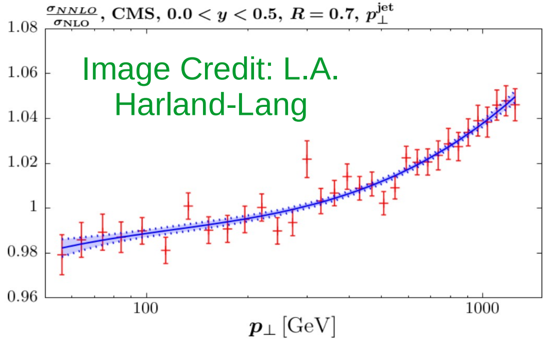
Theoretical Grids

- 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)



Dynamic Tolerance and PDFs

5

- How exactly is the size of the tolerance determined?
- Different prescriptions – could just expand $\Delta\chi^2$ for every eigenvector the same, e.g. CT in past have used $\Delta\chi^2=100$ for 90% CL (now do something more complex).
- MSHT look at each eigenvector and the dataset tensions it sees and set different tolerance → "dynamic tolerance".
- Consider χ^2(Ndatapoints) and rescale its $\Delta\chi^2$ by its 68% CL width. Then tolerance (I.e $\Delta\chi2=T2$) set for each eigenvector direction once one dataset exceeds this.
- Essentially apply weaker "Hypothesis Testing criteria" → rescale such that each dataset lies within its 68% CL for χ^2(Ndatapoints).

$$P_N(\chi^2) = \frac{(\chi^2)^{(N/2-1)}\exp(-\chi^2/2)}{2^{N/2}\Gamma(N/2)}$$

$$\chi_i^2 < \frac{\chi_{i,0}^2}{\xi_{50}}\,\xi_{68}$$

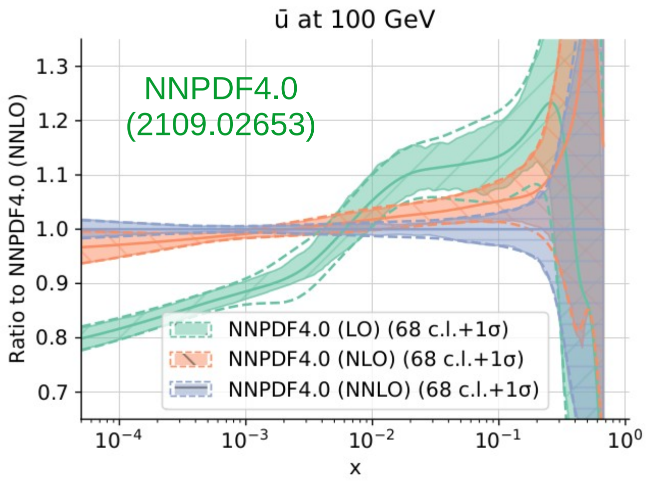
ξ_x gives N corresponding to fractional $(x/100)$ cumulant of distribution, e.g. $\xi_{68} \gtrsim N_{pts}$.

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

<u>Theoretical Uncertainties - MHOUs:</u>

1) Scale variation - Include scale variations as proxy.

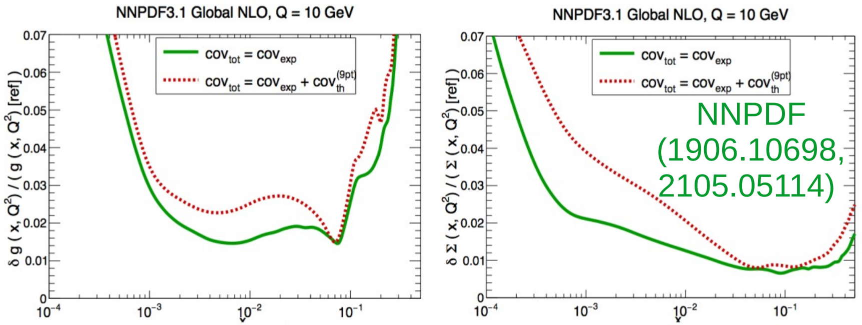
- NNPDF have done NLO, using "theory covariance matrix", S.
- Get small improvements in χ^2/N and larger uncertainties.
- Potential issue of double counting scale variations in PDFs and cross-sections.
- Degree of variation used is arbitrary, only probes (N)NLO terms.

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)

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

Theoretical Challenges

Theoretical Uncertainties - MHOUs:

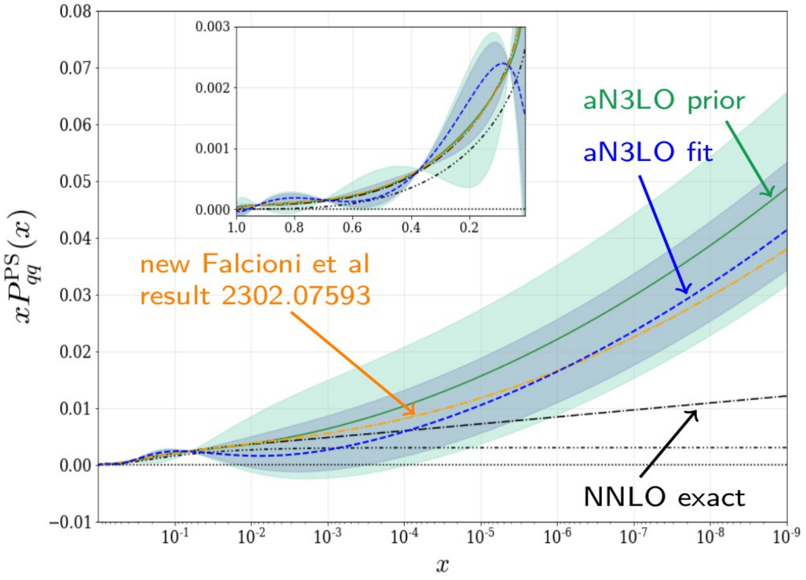
MSHT (2207.04739)

3) Theory Nuisance Parameters and known N3LO

- 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.**



Theory Challenges - Uncertainties

6

- For many processes NLO scale variations were not sufficient to incorporate NNLO result.

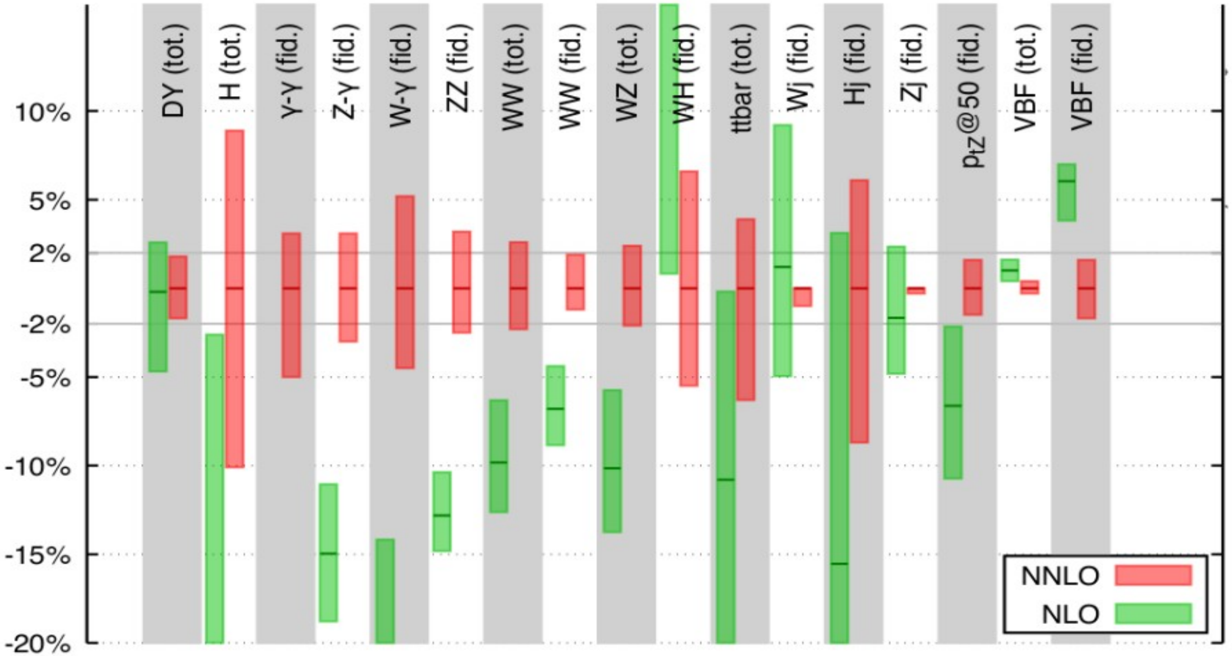


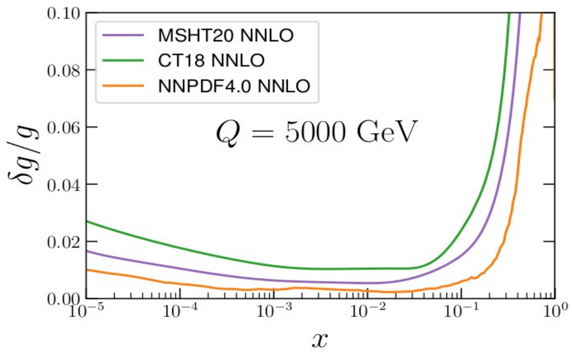
Image Credit: G. Salam

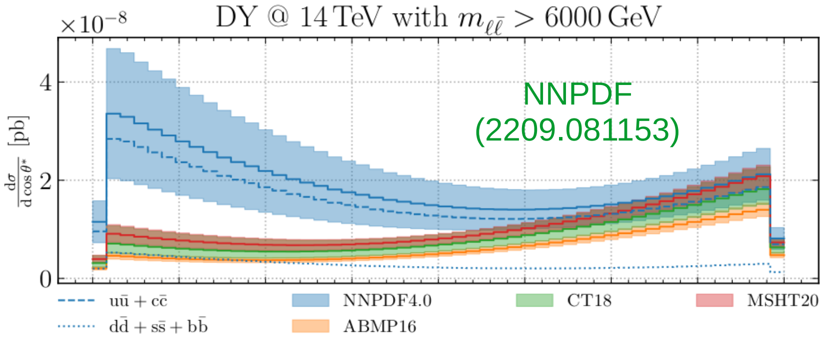
- Is there a better way to do this? → Theory Nuisance Parameters (TNPs)!

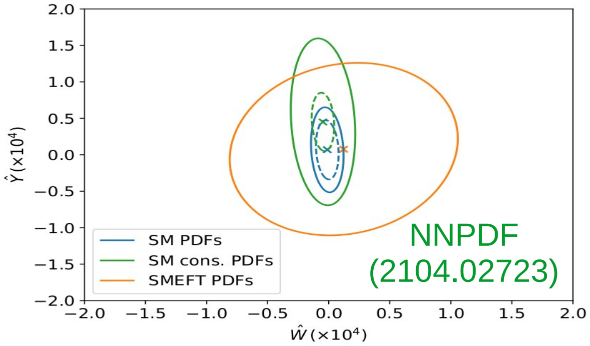
Theoretical Challenges

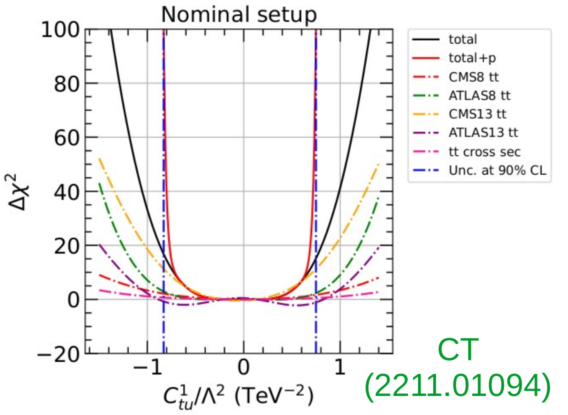
PDFs and Beyond Standard Model Physics

- 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.





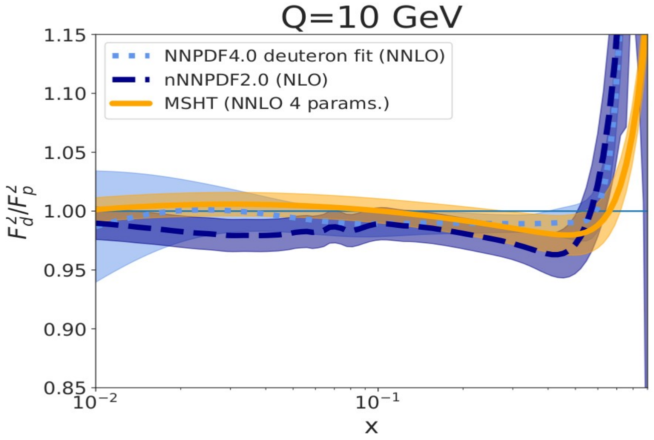


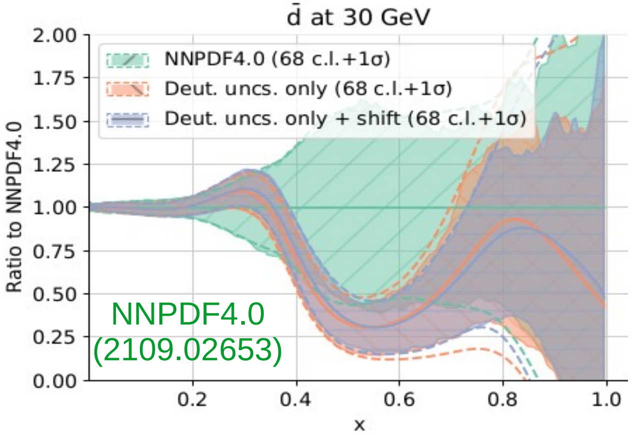


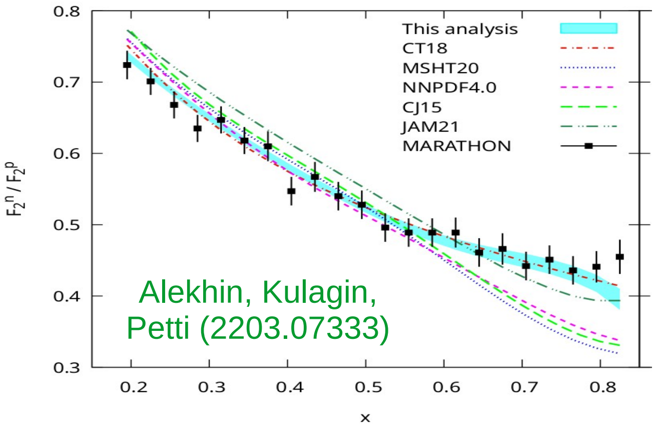
Theoretical Challenges

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.







aN3LO – How does it work?

6

- How do you include the aN3LO information and a theory uncertainty? Consider Pab:
- What do we know and how do we incorporate this information?:
  - Even low-integer N Mellin Moments (4-8) (now 5-10 known[12–15]) - constrain intermediate and high x via $\int_0^1 dx\, x^{N-1} P(x)$.
  - Parameterise $P_{ab}^{(3)}(x)$ with functions $f_{1,\ldots,k}$ where $k =$ No. of known moments and vary basis for uncertainty.
  - Exact LL form at low x from resummation - included in $f_e(x, \rho_{ab})$. E.g. for P_{qg}^3:

$$f_e(x, \rho_{qg}) = \frac{C_A^3}{3\pi^4}\left(\frac{82}{81} + 2\zeta_3\right)\frac{1}{2}\frac{\ln^2(1/x)}{x} + \rho_{qg}\frac{\ln 1/x}{x}$$

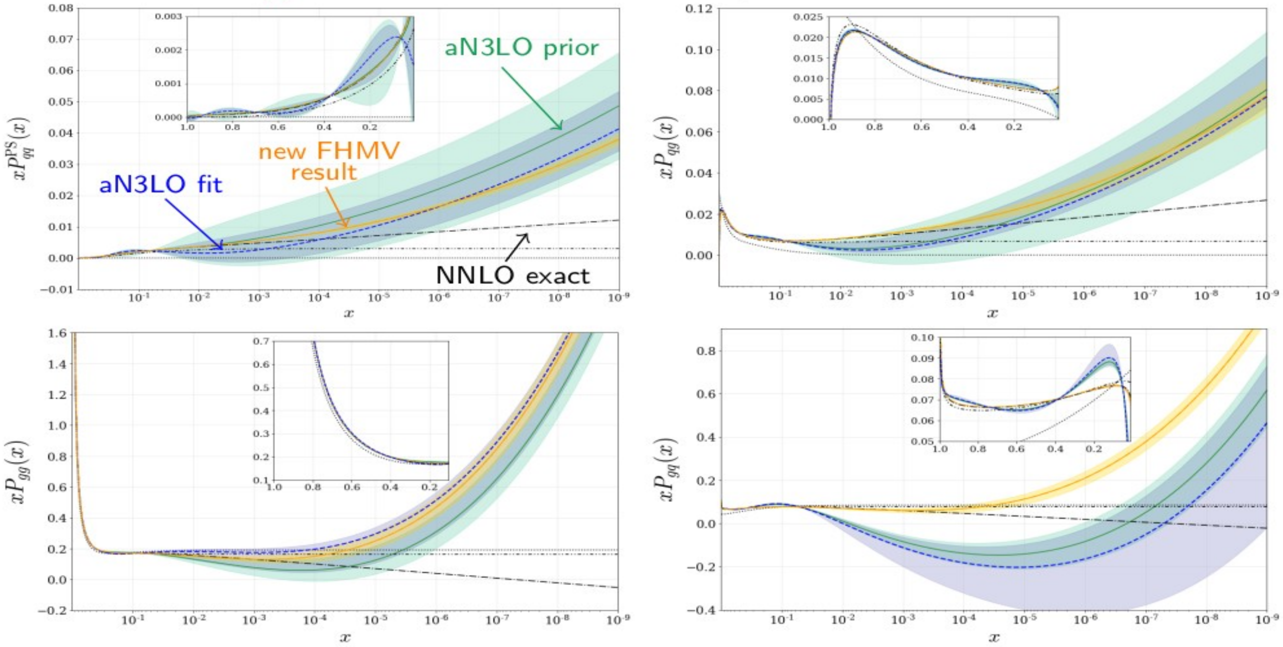
  - Uncertainty on this through coefficient of leading missing low x log as theory nuisance parameter (TNP) ρ_{ab}.
  - Include relevant high x known pieces also in $f_e(x)$.
- So overall:

$$P_{ab}^{(3)}(x) = \sum_{i=1}^{k} A_i f_i(x) + f_e(x, \rho_{ab})$$

1 TNP per Splitting Function = 5 TNPs.

aNLO – How accurate is it?

- Now 5 moments for P_{gg}, P_{gq} [12,13] and 10 for P^{PS}_{qq}, P_{qg} [14,15].
- Largely good agreement with MSHT determinations in central values.
- Exception is P_{gq}, least well determined (one extra low x log unknown).
- Reduction in P^{PS}_{qq}, P_{qg} uncertainties. Impacts reduced once in PDF fit.



Alphas bounds – NNLO

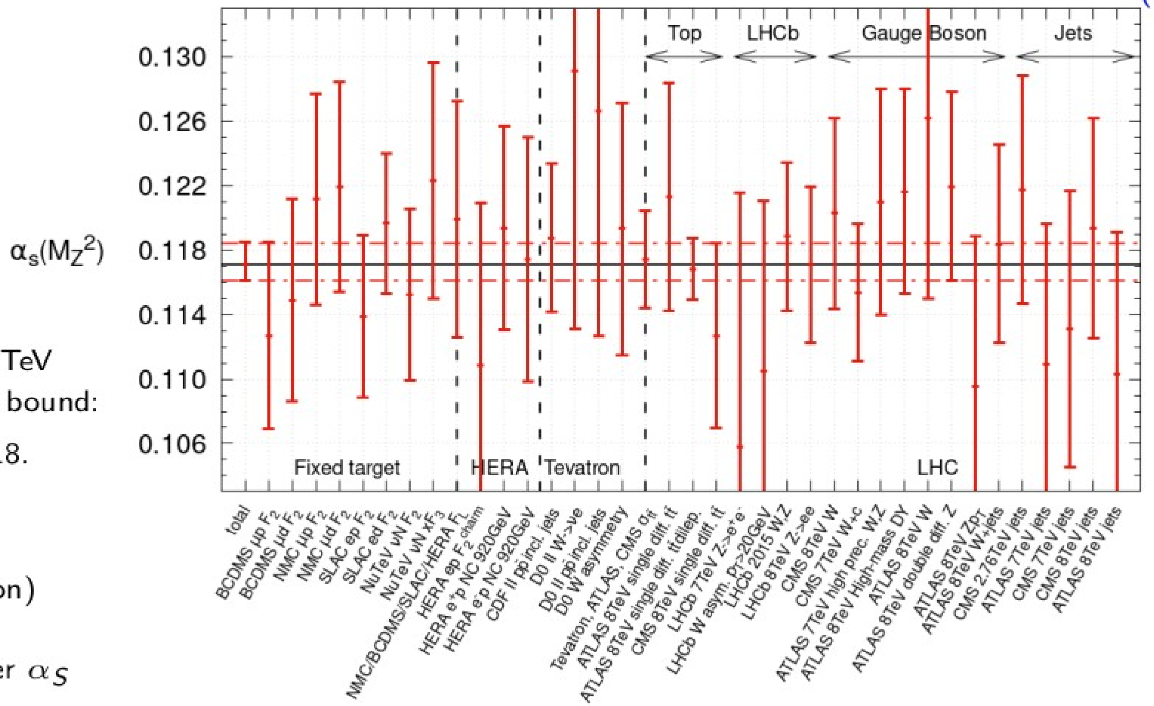
MSHT20 α_S bounds - NNLO

Consistent with α_S bounds seen in previous studies, and between orders (NNLO and aN3LO).

BCDMSp data strongest constraint upwards: $\Delta\alpha_S(M_Z^2) = +0.0014$.

SLACp and ATLAS 8TeV Zp_T both give upper bound: $\Delta\alpha_S(M_Z^2) = +0.0018$.

CMS/ATLAS (dilepton) $t\bar{t}$ single diff. would give lower/same upper α_S bound, but not used.



ATLAS 8 TeV Z data gives lower bound: $\Delta\alpha_S(M_Z^2) = -0.0010$.

NMC deuteron, ATLAS 8 TeV High Mass DY give lower bounds of $\Delta\alpha_S(M_Z^2)$ -0.0017, -0.0018.

Alphas bounds – aN3LO

8

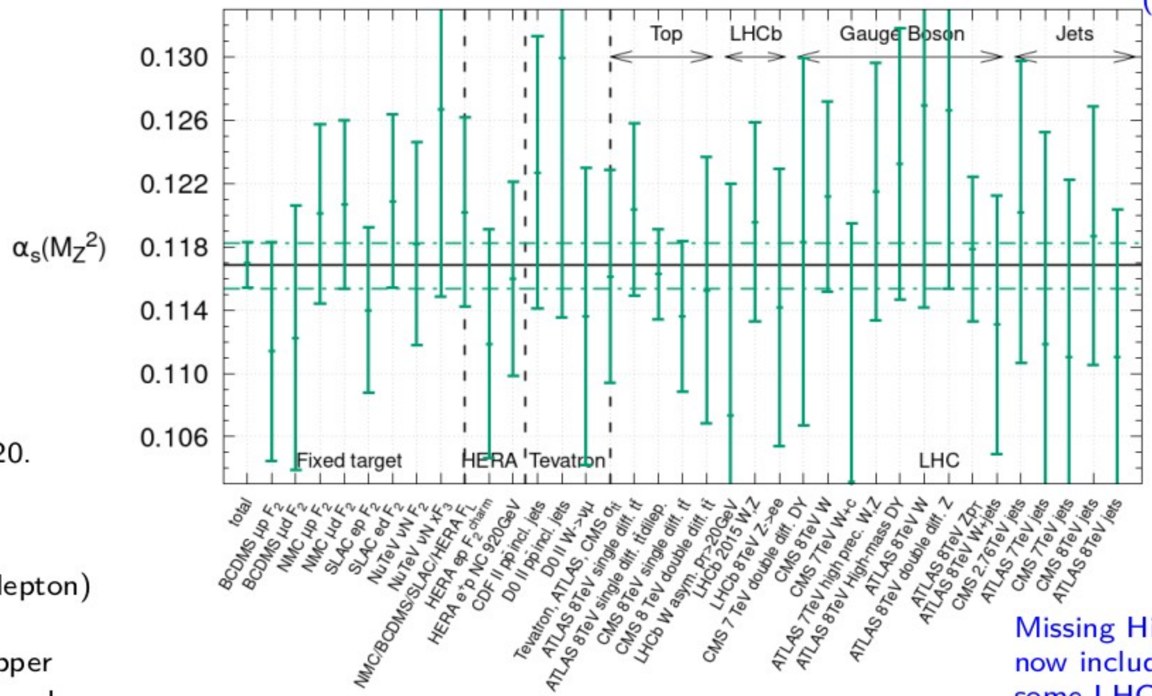
MSHT20 α_S bounds - aN3LO

Consistent with α_S bounds seen in previous studies, and between orders (NNLO and aN3LO).

BCDMSp data strongest constraint upwards: $\Delta\alpha_S(M_Z^2) = +0.0013$.

F_2^c provides upwards bound of: $\Delta\alpha_S(M_Z^2) = +0.0020$.

CMS and ATLAS (dilepton) $t\bar{t}$ single diff. would give slightly higher upper α_S bounds, but not used.



SLAC deuteron data gives lower bound: $\Delta\alpha_S(M_Z^2) = -0.0016$.

NMC deuteron, ATLAS 8 TeV Z both give lower bounds of $\Delta\alpha_S(M_Z^2) = -0.0017$.

Missing Higher Order Uncertainties now included, in particular causes some LHC bounds to weaken as unknown N3LO K-factors.

PDF + Strong Coupling for Xsecs

- Within Hessian approach to PDF uncertainties, correct manner to determine combined PDF+$\alpha_S(M_Z^2)$ uncertainty for any quantity, including correlations between PDFs and α_S is:

1. Take PDFs determined at $\alpha_S(M_Z^2) \pm \Delta\alpha_S(M_Z^2)$ and treat as additional pair of eigenvectors.
2. Determine quantity to obtain $\Delta\sigma_{\alpha_S}$.
3. Combine uncertainties in quadrature:

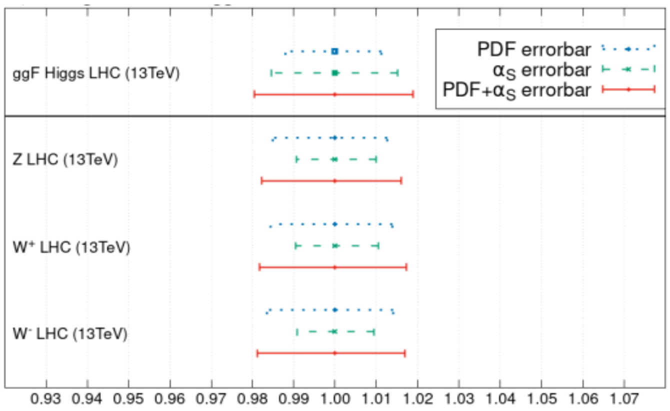
Quadrature as whilst central values correlated errors uncorrelated.
CT: 1004.4624.

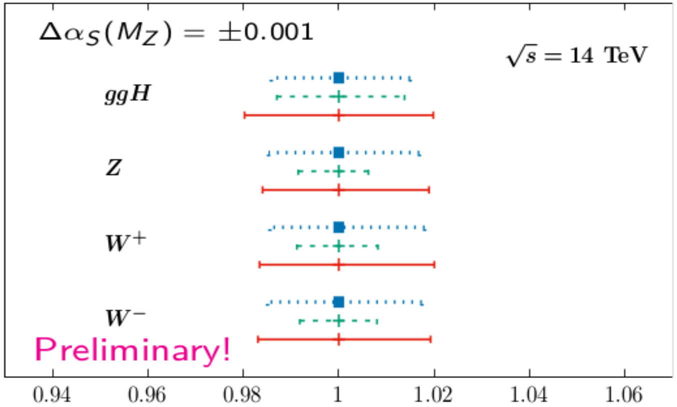
$$\Delta\sigma = \sqrt{(\Delta\sigma_{\mathrm{PDF}})^2 + (\Delta\sigma^2_{\alpha_S(M_Z^2)})}$$

- Works provided central PDFs are best fit PDFs with $\alpha_S(M_Z^2)$ free.
- Choice of $\Delta\alpha_S(M_Z^2)$ up to user but recommended to be close to our 1σ bounds, e.g. ± 0.001 for simplicity and near that of world average.

PDF + Strong Coupling for Xsecs

- Cross-section uncertainties at NNLO/aN3LO (left/right) at the LHC.





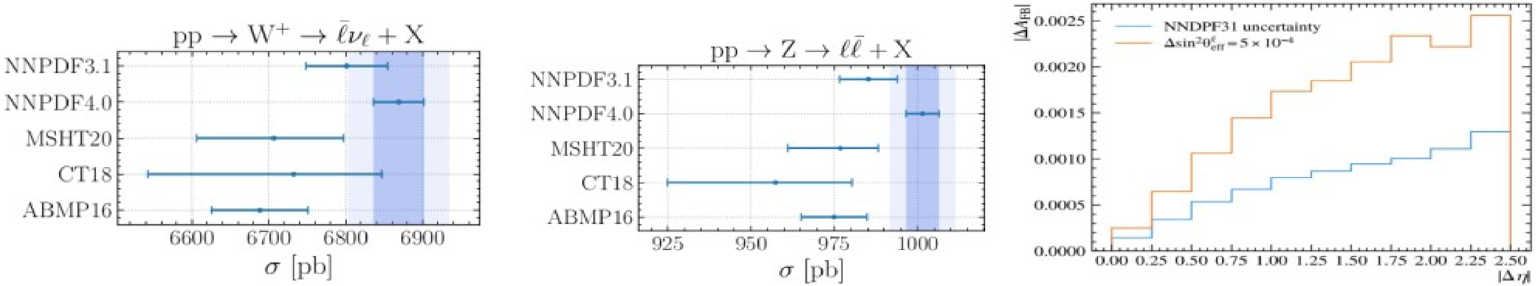
- Direct α_S uncertainty through xsec is small for DY. Total α_S sensitivity larger due to change of PDFs with α_S.
- Direct α_S uncertainty of ggF Higgs is larger ($\sim 2-3\%$), reduced by anti-correlation of gluon with α_S.
- Higher energies sample lower x quarks $\Rightarrow$ larger α_S uncertainties.
- Interplay of direct and indirect (through PDFs) effects $\Rightarrow$ importance of treating PDFs+α_S together.

PDF Uncertainties + PDF Profiling

- Step 1 - Evaluate PDF uncertainty for given PDF.
  Option 1:
- Scan/“offset” PDF eigenvectors: (LHCb m_W and $\sin^2\theta_W$, first ATLAS 7TeV m_W).

$$\delta\mathcal{O} = \sqrt{\sum_{i=1}^{N_{\rm ev}} [(\mathcal{O}_i^+ - \mathcal{O}_i^-)/2]^2}$$

- Typically used for PDF uncertainty on cross-sections:



- Can be used for parameter determination, e.g. LHCb $\sin^2\theta_W$.
- Now experiments often want to additionally constrain PDFs in situ
  $\Rightarrow$ PDF profiling to reduce PDF uncertainties a posteriori.

PDF Uncertainties + PDF Profiling

8

- Step 1 - Evaluate PDF uncertainty for given PDF.

Option 2:

- Profile PDFs - construct a joint χ^2 of new data + PDF:

$$\chi^2(\{\theta_k^{\rm exp}\}, \{\theta_m^{\rm PDF}\}, p) = \sum_i \frac{\left[(D_i + \sum_k \Gamma_{ik}^{\rm exp}\, \theta_k^{\rm exp}) - (T_i + \sum_m \Gamma_{im}^{\rm PDF}\, \theta_m^{\rm PDF})\right]^2}{\sigma_{i,\rm stat}^2} + \sum_k \left(\theta_k^{\rm exp}\right)^2 + \sum_m \left(\theta_m^{\rm PDF}\right)^2$$

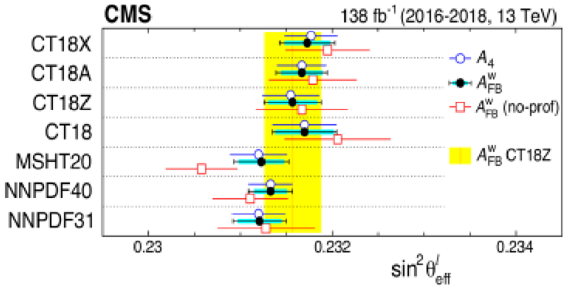
$$= \sum_{i,j} \left(D_i - T_i\right) \left(C_{\rm stat} + C_{\rm exp} + C_{\rm PDF}\right)^{-1}_{ij} \left(D_j - T_j\right)$$

Nuisance Parameter form

Covariance Matrix form

- Increasingly experimental collaborations profile PDFs:
  - See impact of new data on PDFs.
  - Extract SM parameters, e.g. $\alpha_S(M_Z)$.
- Allow fit to move PDF within uncertainty + to constrain PDFs simultaneously with POI (parameter of interest) $p \Rightarrow$ reduced $\Delta p^{\rm PDF}$.

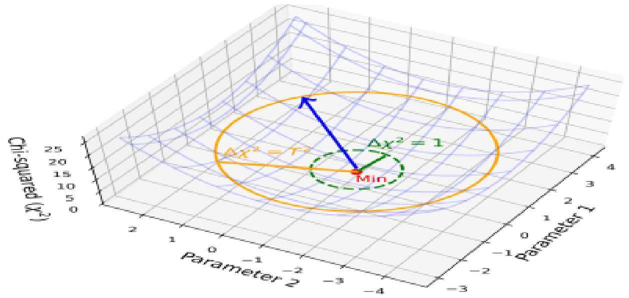
(CMS: 13TeV m_W, $\sin^2\theta_W$, ATLAS: updated 7TeV m_W, 8TeV $\sin^2\theta_W$).



PDF Uncertainties + PDF Profiling

8

- In profiling PDFs fit moves along eigenvectors in PDF parameter space to minimise joint χ^2 of new data + PDF.



Relies on 2 assumptions:

1. The PDFs are not altered significantly (i.e. central values moved $\lesssim 1\sigma$ and uncertainties not substantially changed). $\Rightarrow$ check PDFs pulls
2. PDF uncertainties, representing constraints of previous data in PDFs, are consistently included alongside the new data.

- Consider again joint χ^2 of new data + PDF (i.e. existing data):

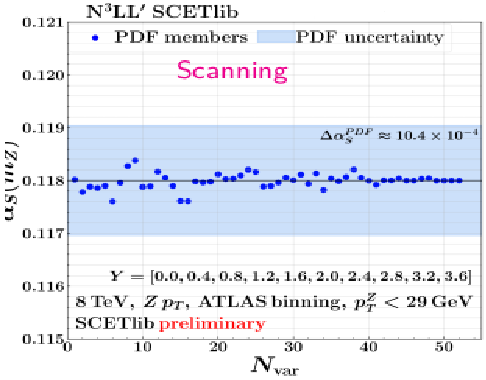
$$\chi^2_{\mathrm{total}} = \sum_{i,j=1}^{N_{\mathrm{new\ data}}} (D_i - E_i + \sum_k \Gamma^{PDF}_{ik}\theta^{PDF}_k) C^{-1}_{ij} (D_j - E_j + \sum_l \Gamma^{PDF}_{jl}\theta^{PDF}_l) + \sum_k T^2_k (\theta^{PDF}_k)^2.$$

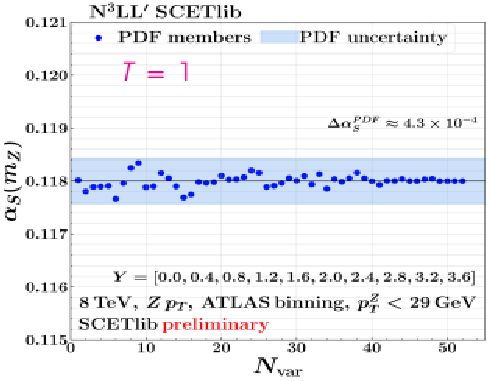
- Last term represents χ^2 penalty to move PDF from its best fit.
- But MSHT, CT PDFs use a "tolerance" so $\theta_k^{\mathrm{PDF}} = \pm 1$ is $\Delta\chi^2 = T_k^2 \neq 1$.

PDF Uncertainties + PDF Profiling

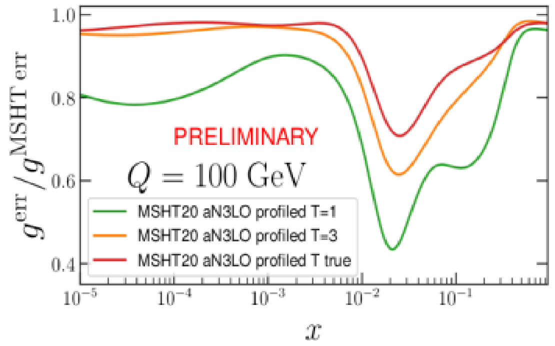
8

- Tolerance must be included in profiling PDFs, impact translates to POIs.
- PDF Uncertainty also significant on α_S from Z p_T, e.g. "scanning": Preliminary!





See Giulia's talk for setup and further info.



- But, new precise Z p_T data can also constrain PDFs $\Rightarrow$ profile to constrain PDFs simultaneously with α_S.
- If you profile PDFs *inconsistently* without tolerance, "T=1", you get a substantial uncertainty reduction.
- But you overestimate the reduction in the uncertainty on the PDFs which results in an incorrect PDF uncertainty on your POI, here α_S.

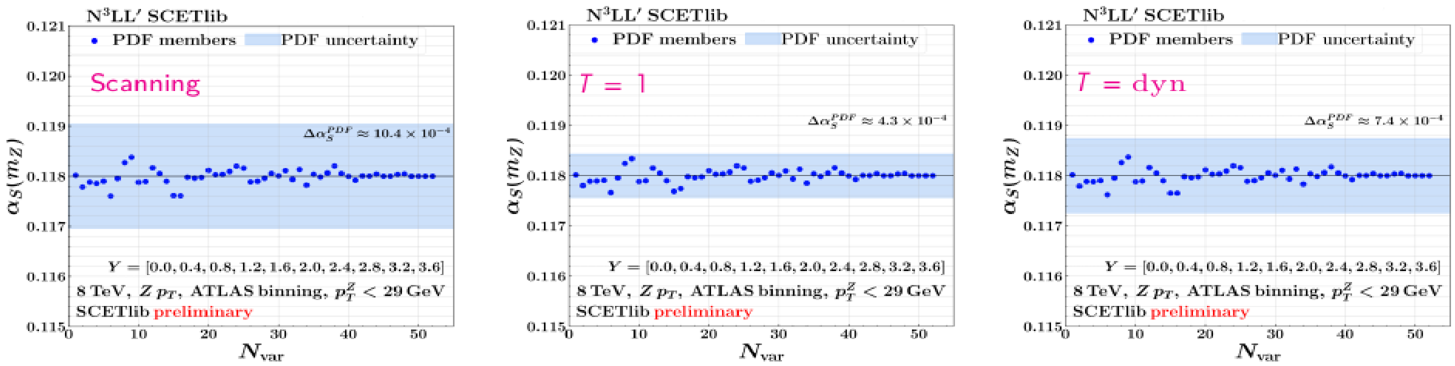
PDF Uncertainties + PDF Profiling

8

Preliminary!

- How can I propagate this to α_S uncertainty? Suggestion (WIP):

- Two-step profiling:
  1. First profile PDFs with factor of T^2 and $\Delta\chi^2 = T^2$ to correctly account for existing PDF uncertainties $\Rightarrow$ obtain consistently profiled PDFs.
  2. Then perform fit with $\Delta\chi^2 = 1$ for other uncertainties and obtain PDF uncertainty by scanning profiled PDFs as new input set.



- In terms of PDF uncertainty:
  $T = 1 \leq$ Two-step profiling $T = \mathrm{dyn} \leq$ Offset/scanned input PDFs.

Some(!) aN3LO References...

[1] M. Cepeda et al., 1902.00134.
[2] Duhr, Mistelberger, 2111.10379.
[3] S. Moch, B. Ruijl, T. Ueda, J. A. M. Vermaseren, and A. Vogt, 1707.08315.
[4] A. Vogt et al., 1808.08981.
[5] S. Moch et al., 2111.15561.
[6] S. Catani and F. Hautmann, Nucl. Phys. B 427, 475 (1994), hep-ph/9405388.
[7] L. N. Lipatov, Sov. J. Nucl. Phys. 23, 338 (1976).
[8] E. A. Kuraev, L. N. Lipatov, and V. S. Fadin, Sov. Phys. JETP 45, 199 (1977).
[9] I. I. Balitsky and L. N. Lipatov, Sov. J. Nucl. Phys. 28, 822 (1978).
[10] V. S. Fadin and L. N. Lipatov, hep-ph/9802290.
[11] M. Ciafaloni and G. Camici, hep-ph/9803389.
[12] G. Falcioni et al., 2310.01245.
[13] S. O. Moch et al., 2310.05744.
[14] G. Falcioni, A. Herzog, S. O. Moch, A. Vogt, 2302.07593.
[15] G. Falcioni, A. Herzog, S. O. Moch, A. Vogt, 2307.04158.
[16] I. Bierenbaum, J. Blümlein, and S. Klein, 0904.3563.
[17] J. Ablinger et al., 1406.4654.
[18] H. Kawamura, N. A. Lo Presti, S. Moch, and A. Vogt, 1205.5727.
[19] J. Ablinger et al., 1409.1135.
[20] J. Ablinger et al., 1402.0359.
[21] J. Ablinger et al., 2211.05462.
[22] T. Gehrmann et al., 2308.07958.
[23] J. Ablinger et al, 2311.00644.
[24] S. Catani, M. Ciafaloni, and F. Hautmann, Nucl. Phys. B 366, 135 (1991).
[25] E. Laenen and S.-O. Moch, hep-ph/9809550..
[26] H. Kawamura, N. A. Lo Presti, S. Moch, and A. Vogt, 1205.5727.
[27] J. Vermaseren, A. Vogt, and S. Moch, hep-ph/0504242.
[28] W. Van Neervan, A. Vogt, hep-ph/9907472.
[29] W. Van Neervan, A. Vogt, hep-ph/0006154.
[30] A. Martin, R.G. Roberts, W.J. Stirling, R.S. Thorne, hep-ph/0006154.
[31] A. Martin, R.G. Roberts, W.J. Stirling, R.S. Thorne, hep-ph/0201127.
[32] C. Anastasiou et al., 1602.00695.
[33] B. Mistlberger, 1802.00833.
[34] F.A. Dreyer and A. Karlberg, 1606.00840.
[35] J. Baglio, C. Duhr, B. Mistlberger, R. Szafron, 2209.06138.
[36] C. Duhr, F. Dulat and B. Mistlberger, 2001.07717.
[37] C. Duhr, F. Dulat and B. Mistlberger 2007.13313.
[38] X. Chen et al., 2107.09085.
[39] C. Duhr and B. Mistlberger, 2111.10379.
[40] N. Kidonakis, 2203.03698.
[41] M. Bonvini, 1812.01958.
[42] R.D. Ball et al, 1710.05935.
[43] H. Abdolmaleki et al, xFitter, 1802.00064.
[44] M. Bonvini, arXiv:1805.08785.
[45] M. Cacciari et al, 1506.02660.