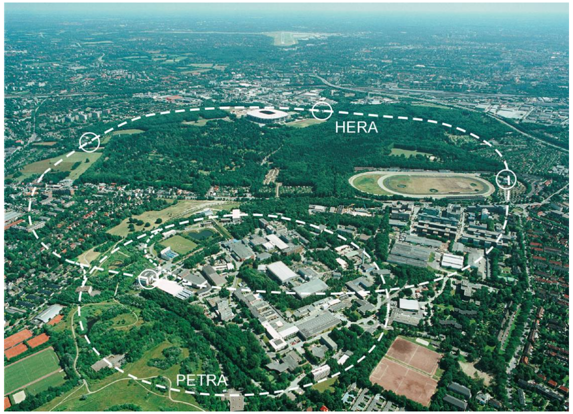
Precision PDFs

- Background, Status, State of the Art



Terascale alliance Monte Carlo school
DESY, Hamburg, Germany.

Thomas Cridge, DESY, 20th February 2024

Outline

1) **Motivation** – Need for Precision PDFs.
2) **Introduction** – PDFs, evolution and Factorisation.
3) **PDF Fitting** and Parameterisation
4) **PDF constraints – Datasets** → Fixed Target, DIS Structure Functions, neutrino scattering, DY, Jets, Top, ZpT.
5) **PDF methodology and Uncertainties**
6) **PDF comparison and PDF4LHC**
7) **PDF state of the art** - QED improved PDFs, theory uncertainties, aN3LO
8) **Strong coupling**

Many details are schematic given time!

Much you may already know, particularly in introduction.

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!

1. Motivation

Motivation

1

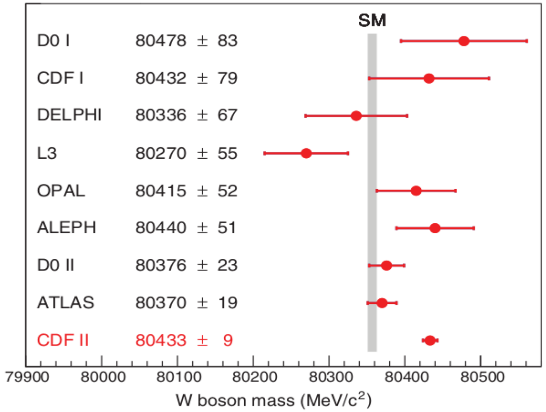
- Key input to very many calculations/measurements at colliders → Need both **accuracy and precision.** Moreover, often a dominant contribution to *uncertainty*.

1) Precision Standard Model (SM) Measurements –

(a) Electroweak Precision:

- W boson mass (M_W):

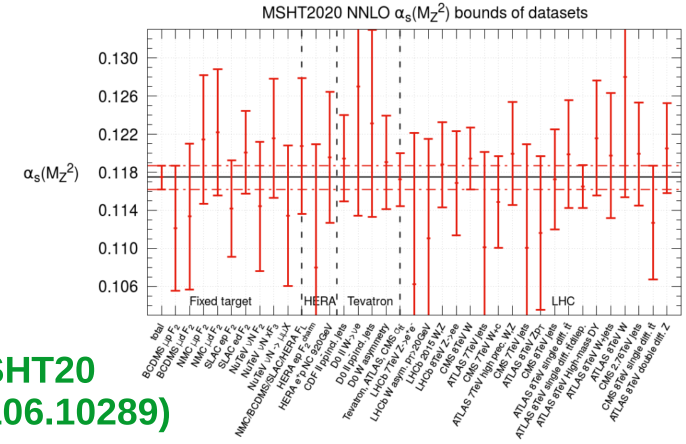
CDF (2022)



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

ATLAS (CONF-2023-004)

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



MSHT20 (2106.10289)

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

See also Snowmass review (2203.08271)

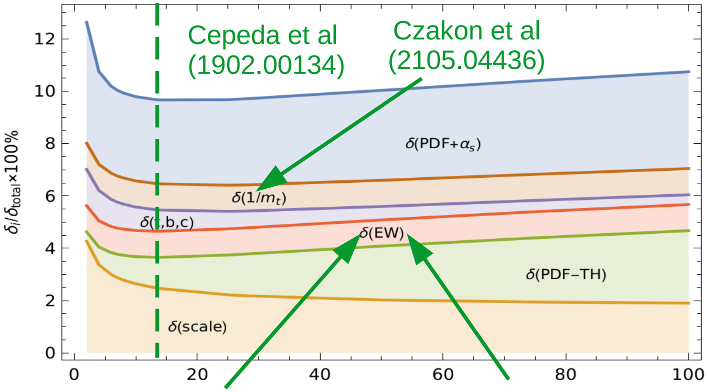
Motivation

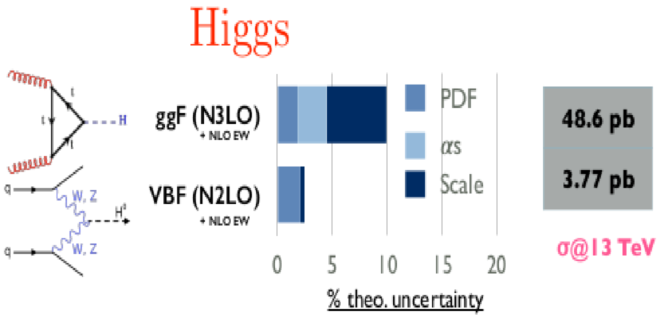
- Key input to very many calculations/measurements at colliders → Need both **accuracy and precision.** Moreover, often a dominant contribution to *uncertainty.*

2) Higgs Measurements:

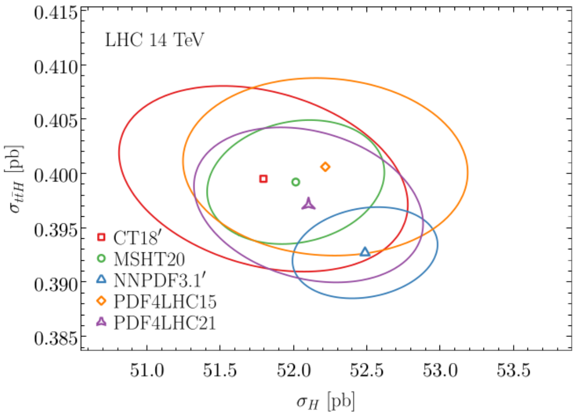
- PDF and related uncertainties (α_s, PDF-TH from NNLO – N3LO mismatch) dominant in ggF Higgs production. Also large in other production mechanisms.

LHC Higgs XSWG 2019





PDF4LHC21 (2203.05506)



Motivation

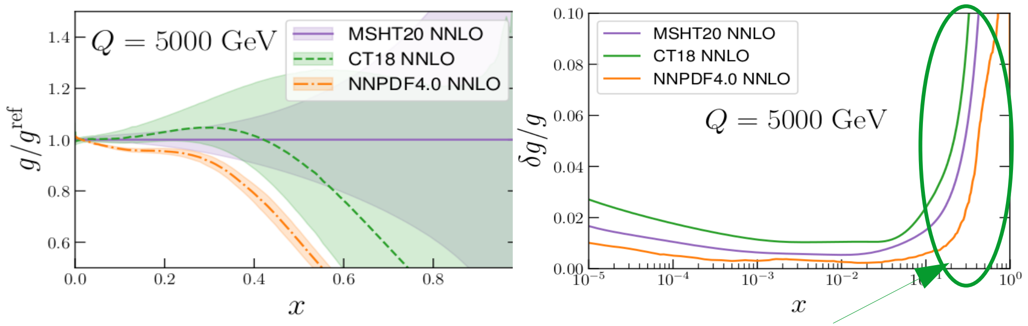
1

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

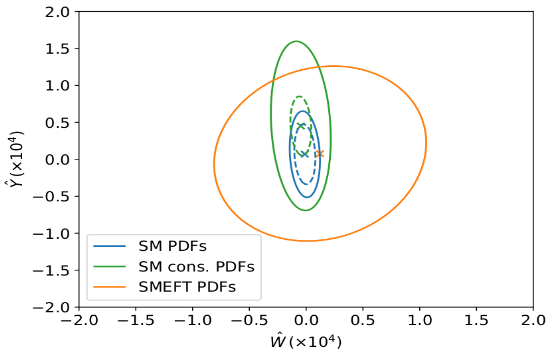
3) Beyond Standard Model (BSM) Searches:

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

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



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



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

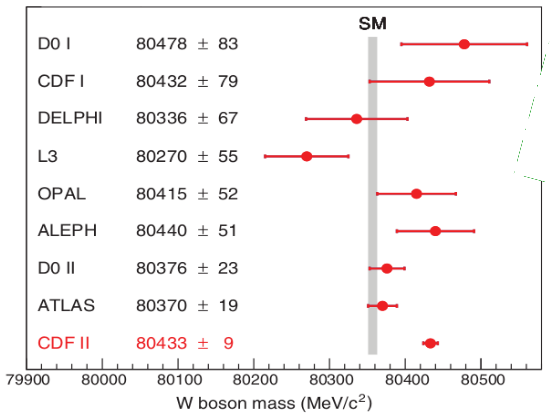
Motivation

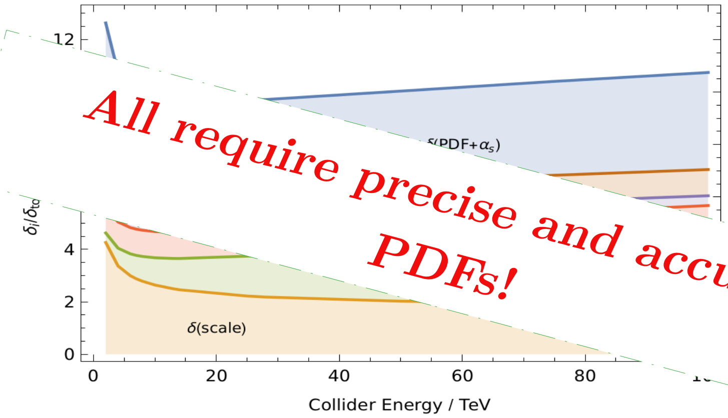
- Key input to very many calculations/measurements at colliders → Need both **accuracy and precision.** Moreover, often a dominant contribution to *uncertainty.*

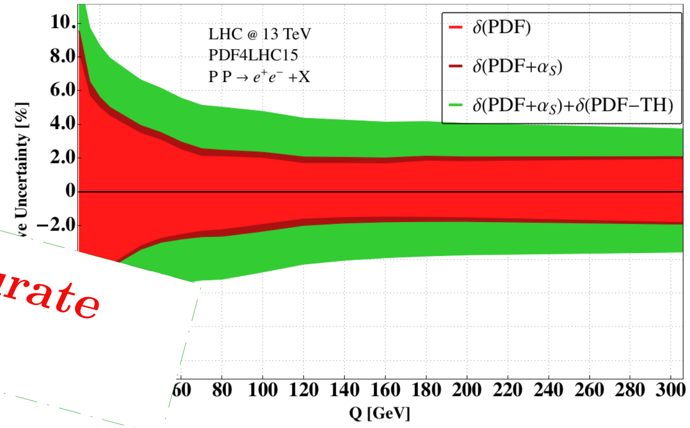
1) Precision Standard Model (SM) Measurements: M_W, $\sin^2\Theta_W$, $\alpha_s(M_Z^2)$, etc.

2) Higgs Measurements

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







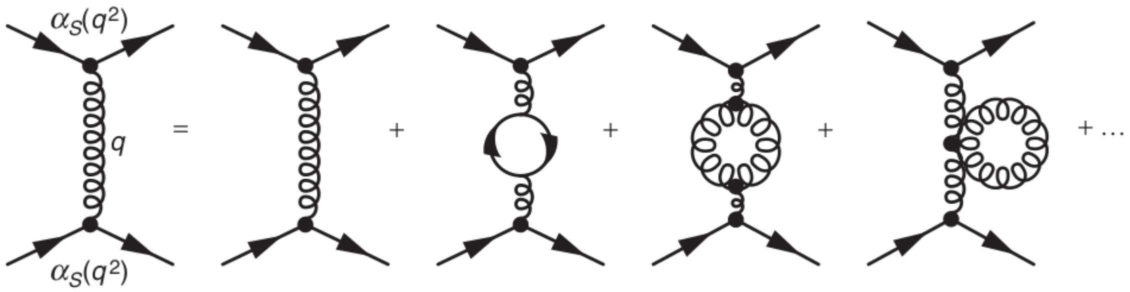
All require precise and accurate PDFs!

2. Introduction

Introduction – QCD Coupling

- *Parton Distribution Functions (PDFs)* are a crucial input and key output of collider physics. Encode non-perturbative content of the proton.
- Strong coupling runs with energy scale:

Outline – seen already in T. Sjöstrand's Lectures



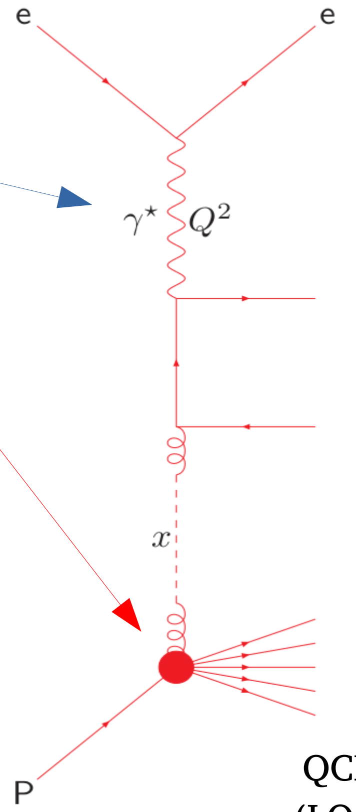
$$\alpha_S(Q^2) = \frac{\alpha_S(\mu^2)}{1 + \frac{(11N_c - 2N_f)}{12\pi}\alpha_S(\mu^2)\log(Q^2/\mu^2)}$$

+ve

- Strong coupling grows at low energies – non-perturbative physics, but reduces at high energies – asymptotic freedom.
- Makes first principles analytic calculations of hadronic physics difficult – instead rely on separation of short (collider energy) and long distance (proton content) physics – Factorisation.

Introduction - Factorisation

- Consider electron scattering off a proton:
- Can separate short distance perturbative physics in coefficient functions and hard cross-sections from non-perturbative long distance PDFs.
- Based on **Quark Parton Model** -
  - In "infinite momentum frame", electron scatters off independent partons inside proton.
  - No transverse motion, difference in energy scales means no quark interactions over this time.
- Separate short distance perturbative physics in scattering from non-perturbative physics determining parton distributions in the proton.
- Parton Distribution Functions (PDFs) are **universal**.



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

Schematic! See references for more detail!

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

QCD splittings alter this (LO) picture somewhat – *QCD improved parton model*

Introduction - Factorisation

- Therefore write DIS cross-section in factorised form:

$$\sigma(ep \to eX) = \sum_i C_i^P(x, \alpha_s(Q^2)) \otimes \tilde{f}_i(x, Q^2, \alpha_s(Q^2)) + \mathcal{O}\left(\frac{\Lambda^2_{QCD}}{Q^2}\right)$$

$$\text{where} \qquad (a(x) \otimes b(x) = \int_x^1 a(z) b(x/z)\, dz)$$

Corrections to this separation.

- "PDFs", $f_i(x, Q^2)$ represent probability of finding a parton of type i carrying momentum fraction x of the proton, independent of process!

- Coefficient functions, $C_i^P(x, Q^2)$ are process (P) dependent and perturbative – expand as power series in strong coupling:

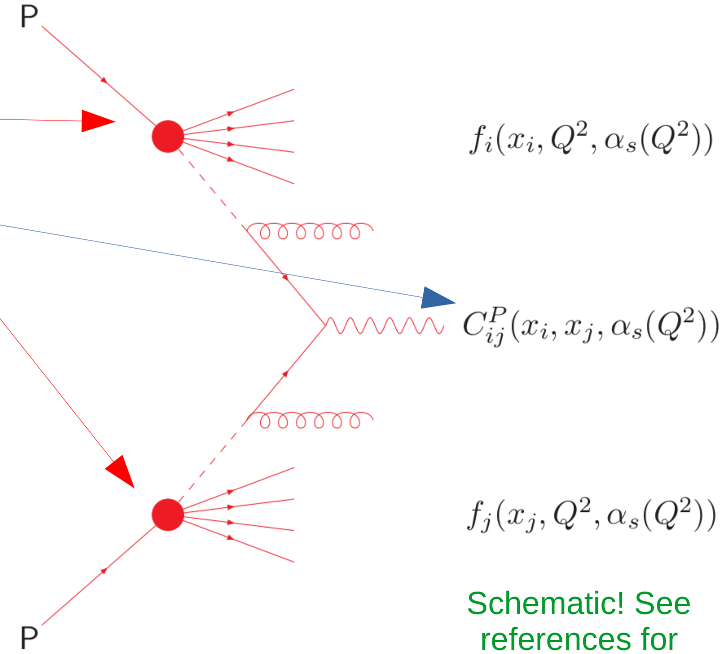
Schematic! See references for more detail!

$$C_i^P(x, \alpha_s(Q^2)) = \sum_k C_i^{P,k}(x) \alpha_s^k(Q^2).$$

Introduction - Factorisation

- Same applies for pp collisions (LHC):

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



Schematic! See references for more detail!

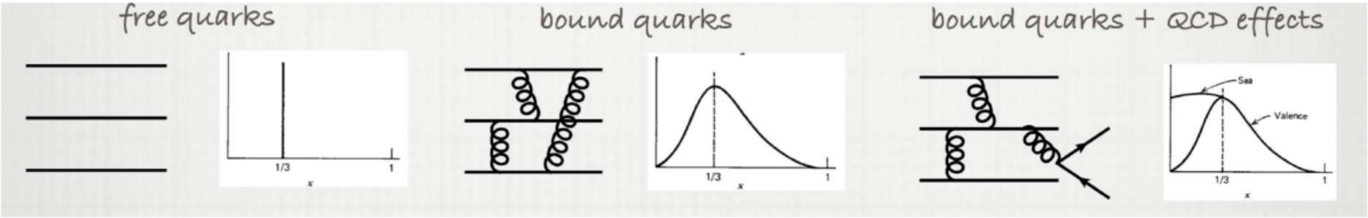
- Formally what we're doing is doing is absorbing collinear emissions from coefficient functions into redefinitions of the parton distributions.

- We factor any emission with transverse momentum <some scale (factorisation scale, μ_F) into PDFs.

- In this process we absorb collinear divergences from initial state radiation.

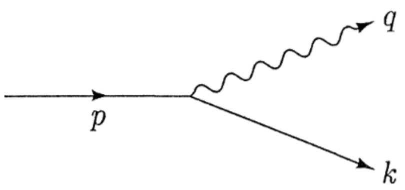
Introduction – PDF evolution

- What do the PDFs look like?



So QCD splittings affect PDFs → transfer momentum between partons.

- Consider e→γ splitting (proxy for q→g), following Peskin and Schroder:



$$q = (zp, p_\perp, 0, zp - \frac{p_\perp^2}{2zp}),$$

$$k = ((1-z)p, -p_\perp, 0, (1-z)p + \frac{p_\perp^2}{2zp}).$$

$$i\mathcal{M} = \bar{u}_L(k)(-ie\gamma_\mu)u_L(p)\epsilon_T^{*\mu}(q),$$

$$i\mathcal{M}(e_L^- \to e_L^-\gamma_R) = ie\frac{\sqrt{2(1-z)}}{z}p_\perp.$$

$$i\mathcal{M}(e_L^- \to e_L^-\gamma_L) = ie\frac{\sqrt{2(1-z)}}{z(1-z)}p_\perp.$$

Parity violation → unchanged if all helicities flipped

$$\frac{1}{2}\sum_{\text{pols.}} |\mathcal{M}|^2 = \frac{2e^2p_\perp^2}{z(1-z)}\left[\frac{1+(1-z)^2}{z}\right].$$

Contains $P_{e\gamma}$ splitting function

$$P_{\gamma\leftarrow e}(z) = \frac{1+(1-z)^2}{z}$$

Introduction – PDF evolution

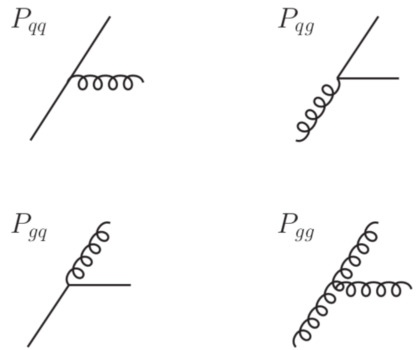
- Similar for other QED splitting functions, can then convert to QCD via correct colour factors (except new g→gg splitting), obtain LO Splitting functions:

Interpreted as emission “probabilities” in parton showers – seen T. Sjöstrand’s Lectures

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

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

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



- These connect partons at different x and Q^2, via DGLAP equations (2Nf+1 coupled integro-differential equations):

Saw already in T. Sjöstrand’s Lectures

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

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!

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

Singlet

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

$$q_V(x,Q^2) = \sum_{i=1}^{n_f} [q_i(x,Q^2) - \bar{q}(x,Q^2)]$$
$$q_{ij}^{\pm}(x,Q^2) = (q_i \pm \bar{q}_i) - (q_j \pm \bar{q}_j)$$

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).

- Gets much more complicated at higher orders… :

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

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

$$\begin{aligned}
P^{(1)}_{\mathrm{ps}}(x) &= 4\, C_F n_f \left(\frac{20}{9}\frac{1}{x} - 2 + 6x - 4\mathrm{H}_0 + x^2\left[\frac{8}{3}\mathrm{H}_0 - \frac{56}{9}\right] + (1+x)\left[5\mathrm{H}_0 - 2\mathrm{H}_{0,0}\right]\right)
\end{aligned}$$

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

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

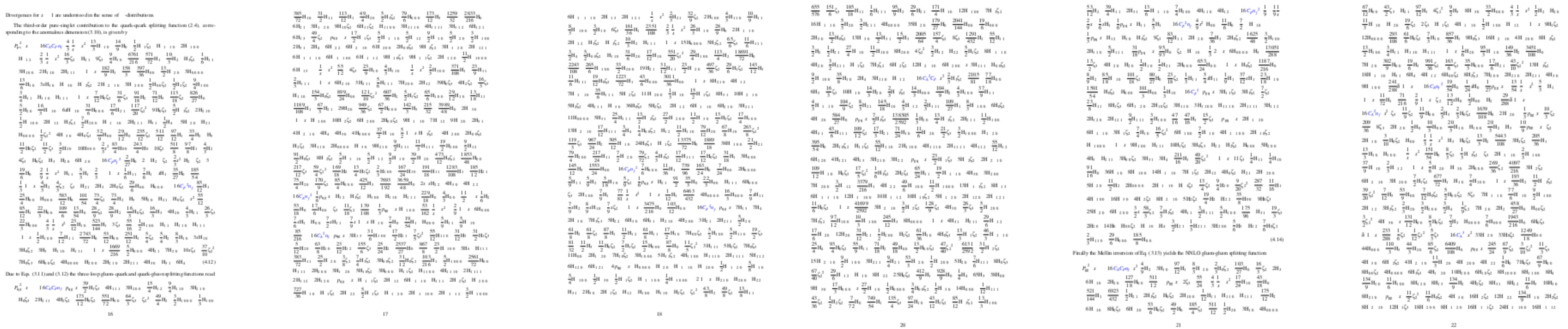
$$\begin{aligned}
P^{(1)}_{\mathrm{gg}}(x) &= 4\, C_A n_f \left(1 - x - \frac{10}{9}p_{\mathrm{gg}}(x) - \frac{13}{9}\left(\frac{1}{x} - x^2\right) - \frac{2}{3}(1+x)\mathrm{H}_0 - \frac{2}{3}\delta(1-x)\right) + 4\, {C_A}^2 \left(27\right. \\
&+(1+x)\left[\frac{11}{3}\mathrm{H}_0 + 8\mathrm{H}_{0,0}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

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

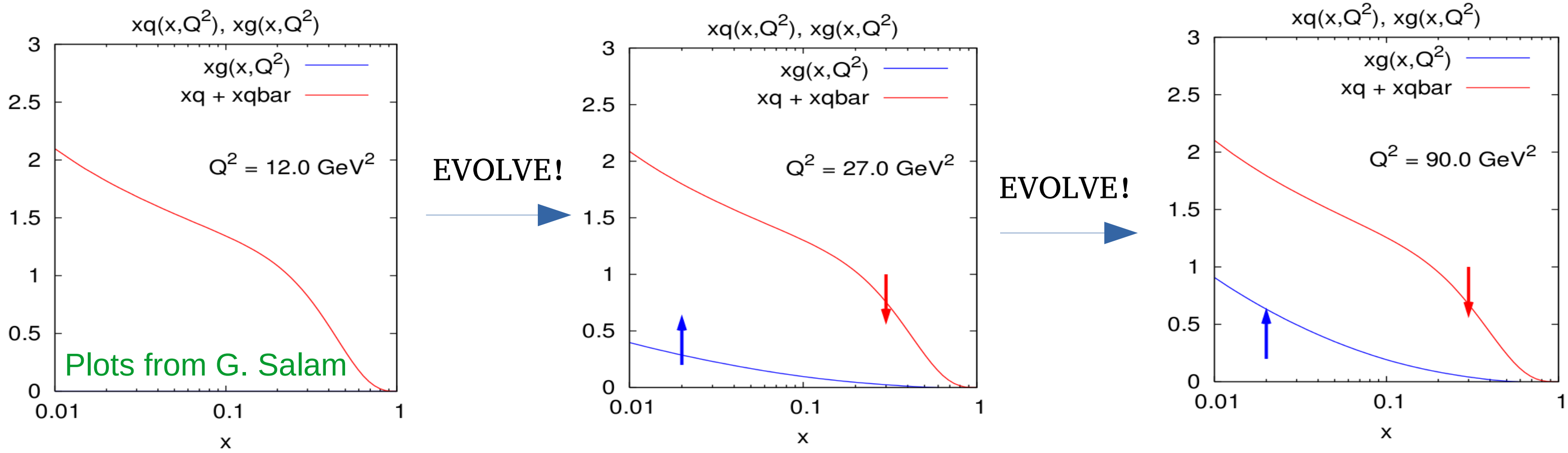
2

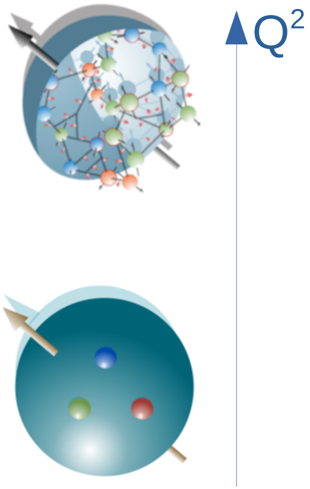
- How does DGLAP affect PDFs?
- P_{gg}, P_{gq} diverge at small x, means PDFs evolve to lower x with increasing Q^2.
- Can visualise this with LO example, take just quarks:

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

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

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





Introduction – PDF Fitting

2

- This means, once we know the PDFs at one scale (Q^2) for all x, we know them for all (perturbative) (x,Q^2)!
- PLUS – we saw from factorisation that PDFs are universal → fit to one/one set of process(es) and use for predictions for others → PDF (global) fitting.
- How does this work?

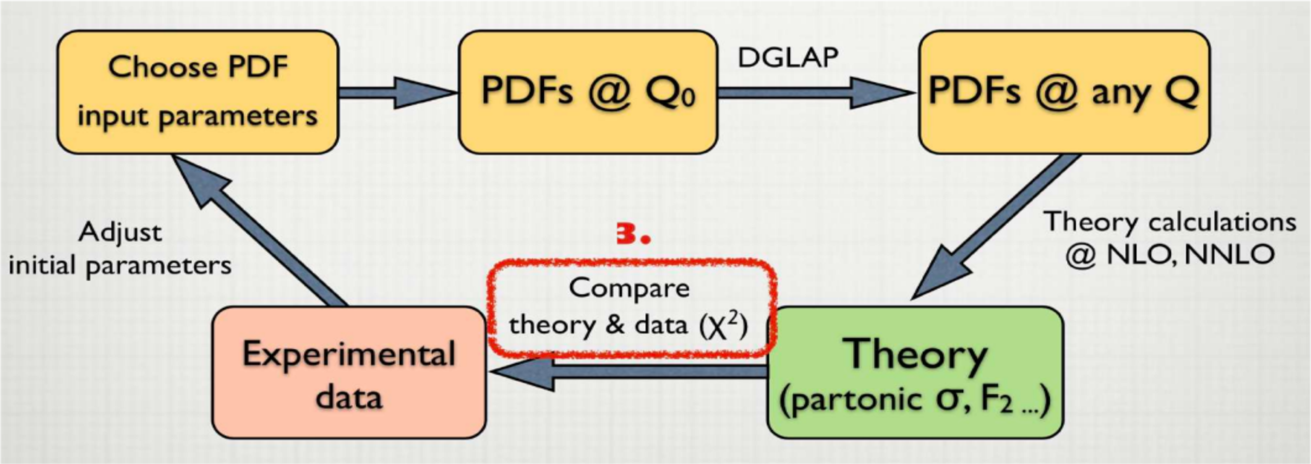


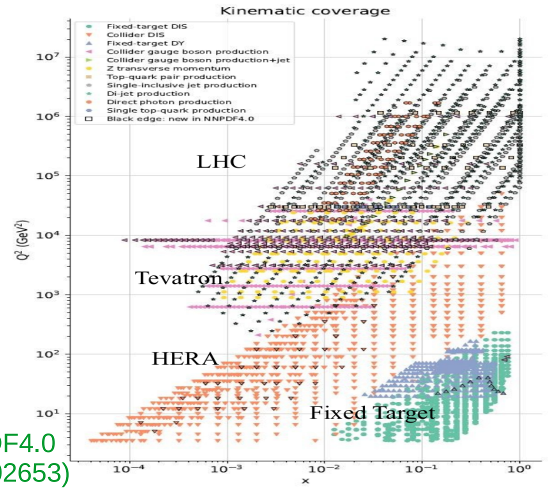
Figure from M. Guzzi

3. Global PDF Fitting

Global PDF Fitting

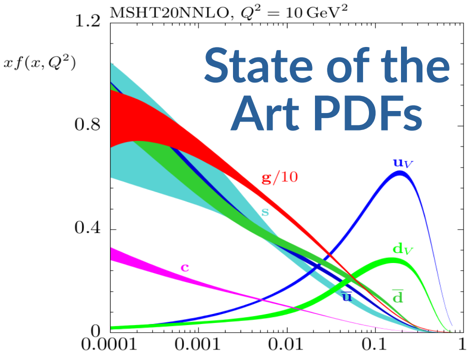
1) Experiment

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



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

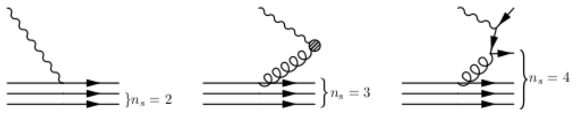
PDF Parameterisation

3

- 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



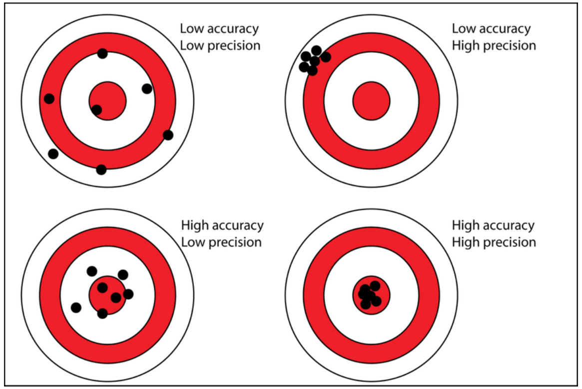
$$f(x) \xrightarrow[x_B \to 1]{} (1 - x_B)^3 \qquad (1 - x_B)^5 \qquad (1 - x_B)^7$$

- 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:

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

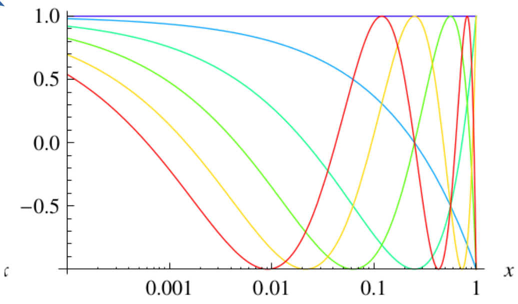
Momentum sum rule:

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

- Cannot constrain all partons independently (see later).
- 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": $\boxed{T_i\left(y(x) = 1 - 2\sqrt{x}\right)}$
- 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...

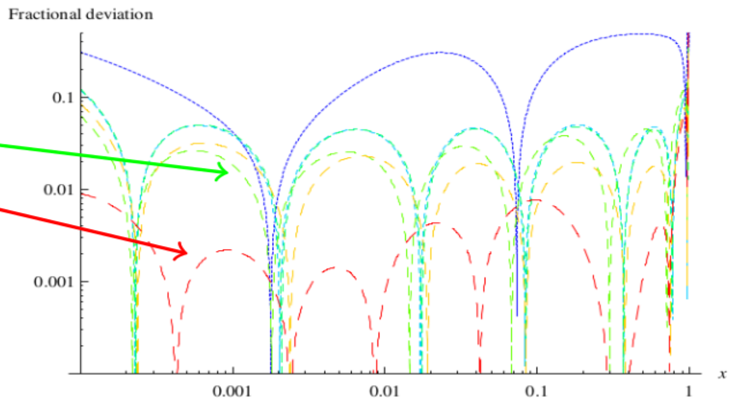
More info in MMSTWW (1211.1215)



PDF Parameterisation (MSHT)

3

- MSHT use Chebyshev polynomials $T_i(1-2x^{0.5})$ to parameterise PDFs.
- MMHT used 4 Chebyshevs, MSHT now uses 6 Chebyshevs $\Rightarrow$ enables fitting to $< 1\%$ if data allows.
- Parameterise $\bar{d}/\bar{u}$ instead of $\bar{d}-\bar{u}$, with $\bar{d}/\bar{u} \rightarrow$ constant as $x \rightarrow 0$.



MMHT: 1211.1215.

More info in MSHT (2012.04684)

Study using pseudo-data of deviation of fit from truth → (much) less than 1%.

New 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 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 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 fixed by $\int_0^1 s^-\,dx = 0$, δ_{s-} 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

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

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

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

4. PDF Constraints

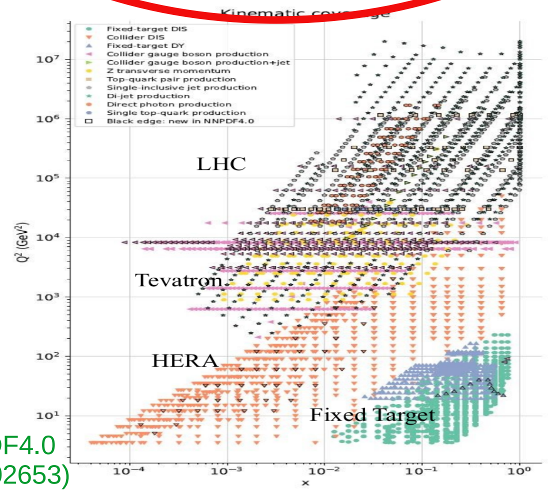
Global PDF Fitting

3

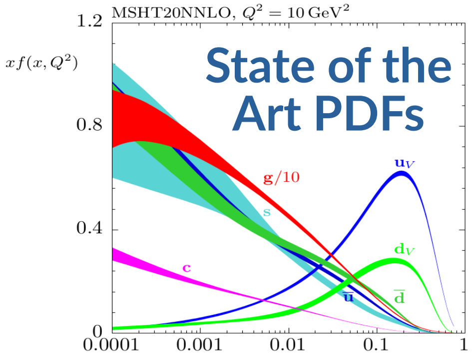
1) Experiment
- Latest experimental data
- Fixed target, collider DIS, Tevatron, LHC, etc
- EW boson, jets, top, ...
- Large range in x, Q^2

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

Global PDF fit



NNPDF4.0 (2109.02653)



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

PDF Constraints - What data?

- 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 and more!.
- Constrains central values and uncertainites.

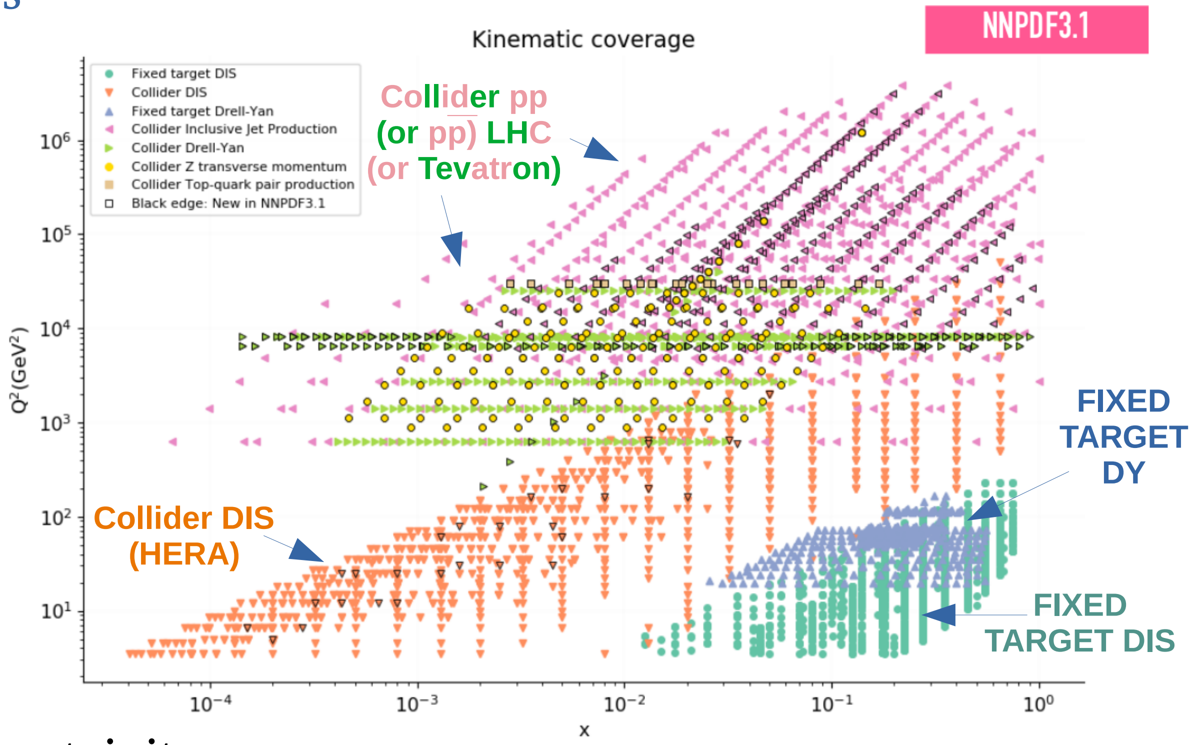


Figure adapted from NNPDF3.1 (1706.00428)

PDF Constraints (DIS reminder)

4

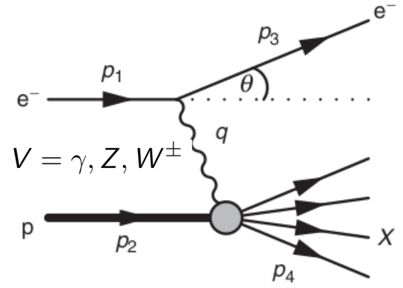
- How do we constrain all these PDFs and parameters? (>50 in MSHT!)
- Backbone of PDFs is (still) Deep Inelastic Scattering (DIS), i.e. ep → eX:
- Lepton scattering off a proton via vector boson exchange:
- Can write differential xsec as:

LeptonicTensor (EW physics)

Hadronic Tensor (Non-perturbative)

$$\frac{d^2\sigma}{dxdy} = x(s-m^2)\frac{d^2\sigma}{dxdQ^2} = \frac{2\pi y\alpha^2}{Q^4}\sum_j \eta_j L_j^{\mu\nu} W^j_{\mu\nu}$$

Propagators and couplings, $\eta_\gamma = 1$



- Most general form of hadronic tensor:

$$W_{\mu\nu} = (-g_{\mu\nu} + \frac{q_\mu q_\nu}{q^2})F_1(x,Q^2) + \frac{\hat{p}_\mu \hat{p}_\nu}{p.q}F_2(x,Q^2) - i\epsilon_{\mu\nu\alpha\beta}\frac{q^\alpha p^\beta}{2p.q}F_3(x.Q^2)$$

Structure Functions

Parity Violating (0 for QED)

- Obtain overall:

$$\frac{d^2\sigma}{dxdy} = \frac{4\pi\alpha^2}{xyQ^2}\eta^i\left((1-y-\frac{x^2y^2M^2}{Q^2})F_2^i + y^2xF_1^i \mp (y-y^2/2)xF_3^i\right)$$

PDF Constraints (DIS reminder)

4

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

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

- Compare with last page (simplified – forget last term as QED):

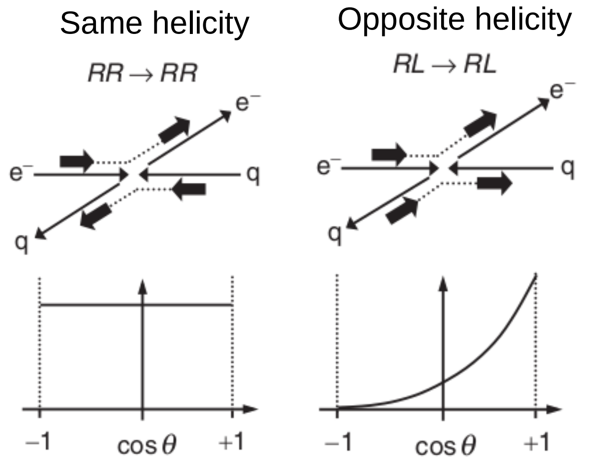
$$\frac{d^2\sigma}{dxdQ^2} = \frac{4\pi\alpha^2}{Q^4}\left((1-y)F_2(x,Q^2)/x + y^2 F_1(x,Q^2)\right)$$

- Callan Gross relation and Bjorken scaling:

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

Aside:

Can understand $1+(1-y^2)$ factor by helicity:



All angles equally possible

Angular distribution goes like 0.5(1+cosθ) = 0.5(1- y)

Sum all 4 possibilities = $1 + (1-y)^2$

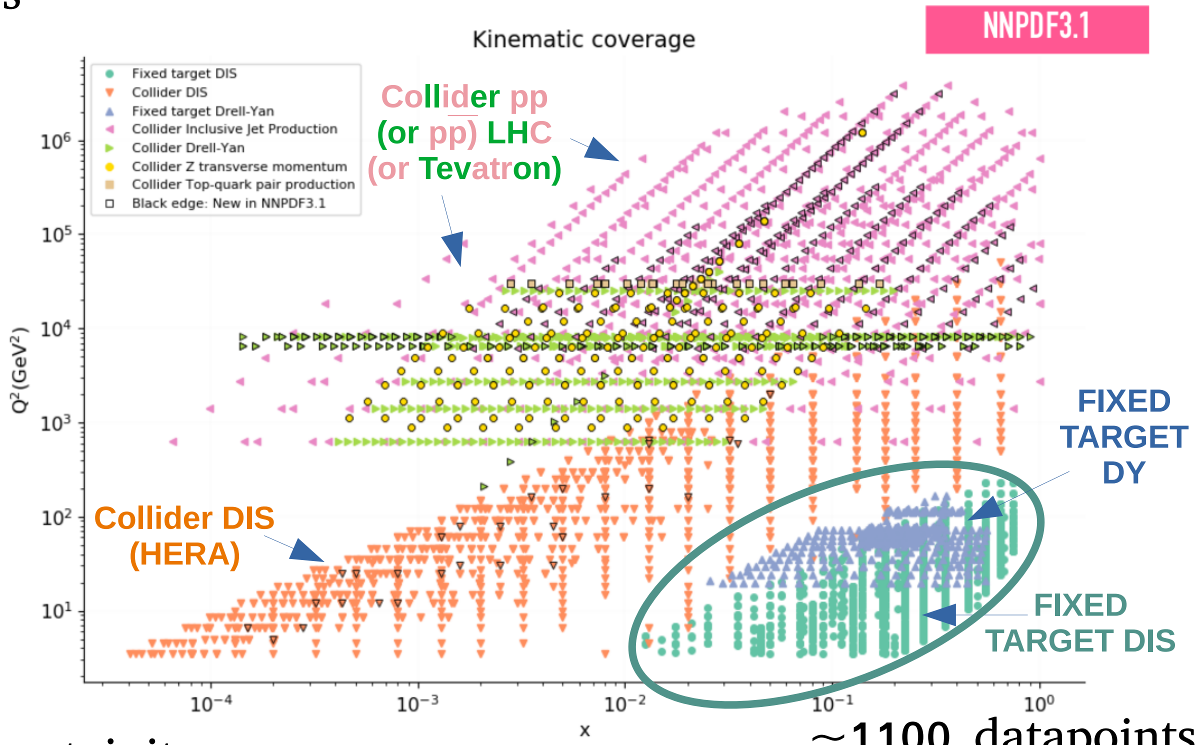
PDF Constraints - Fixed Target DIS ④

- 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.



~1100 datapoints

Figure adapted from NNPDF3.1 (1706.00428)

- Constrains central values and uncertainites.

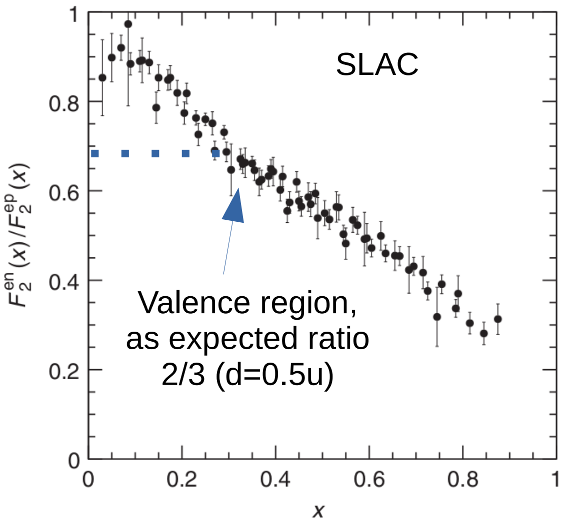
PDF Constraints (Fixed Target DIS) ④

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

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

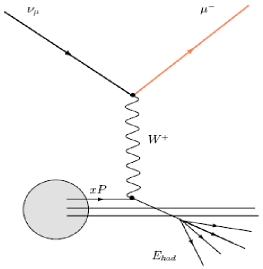
- Structure function sensitive to charge-weighted sum of quarks.
- Measure at high x in fixed target experiments (BCDMS, SLAC, NMC, etc) → high x quarks!
- If also measure on neutrons then you get different charge-weighted sum → break some degeneracy in flavour decomposition (u vs d).
- Ratio tells us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

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



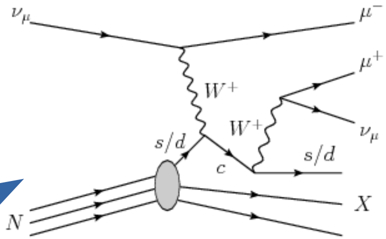
PDF Constraints (Fixed Target DIS) 4

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



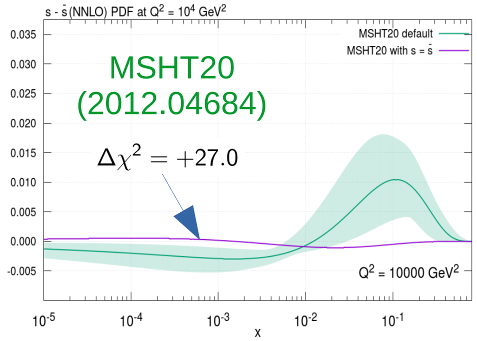
$$
\begin{aligned}
F_2^{\nu} &= 2x[d + s + \bar{u} + \bar{c}] \\
F_2^{\bar{\nu}} &= 2x[u + c + \bar{d} + \bar{s}] \\
xF_3^{\nu} &= 2x[d + s - \bar{u} - \bar{c}] \\
xF_3^{\bar{\nu}} &= 2x[u + c - \bar{d} - \bar{s}].
\end{aligned}
$$

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



- 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.

NuTeV experiment



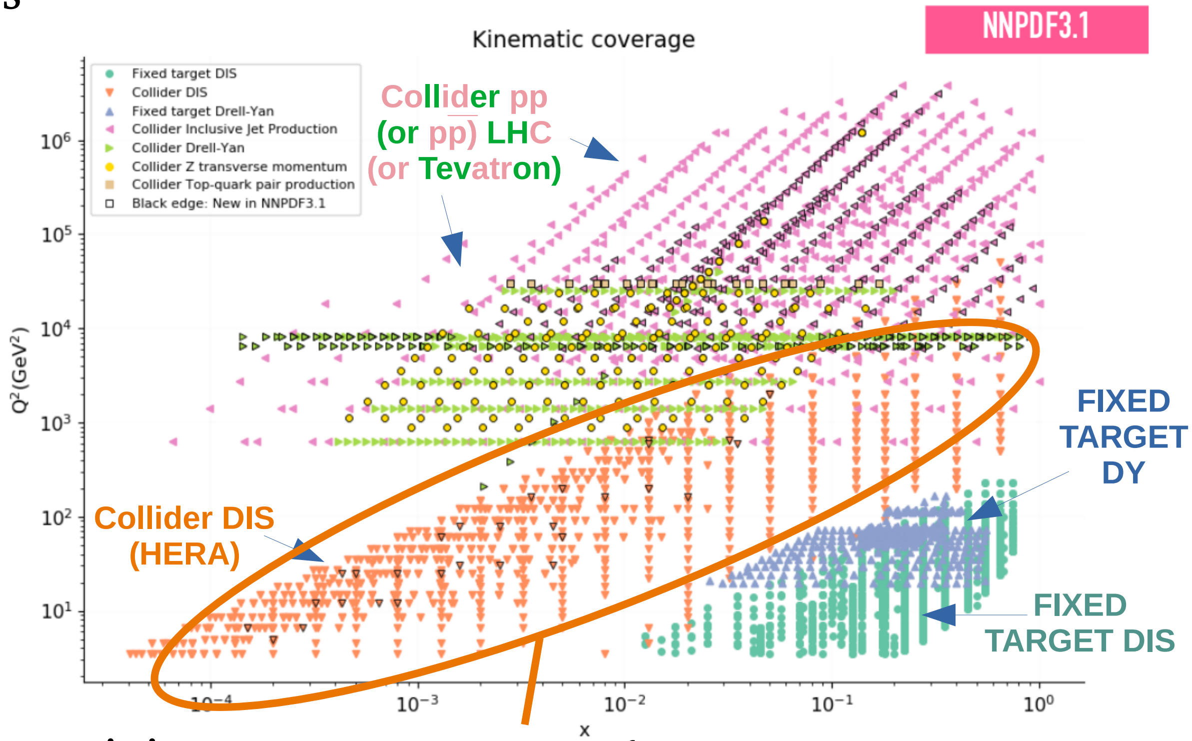
- One of main constraints on strangeness asymmetry s-s̄ in PDFs.

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)

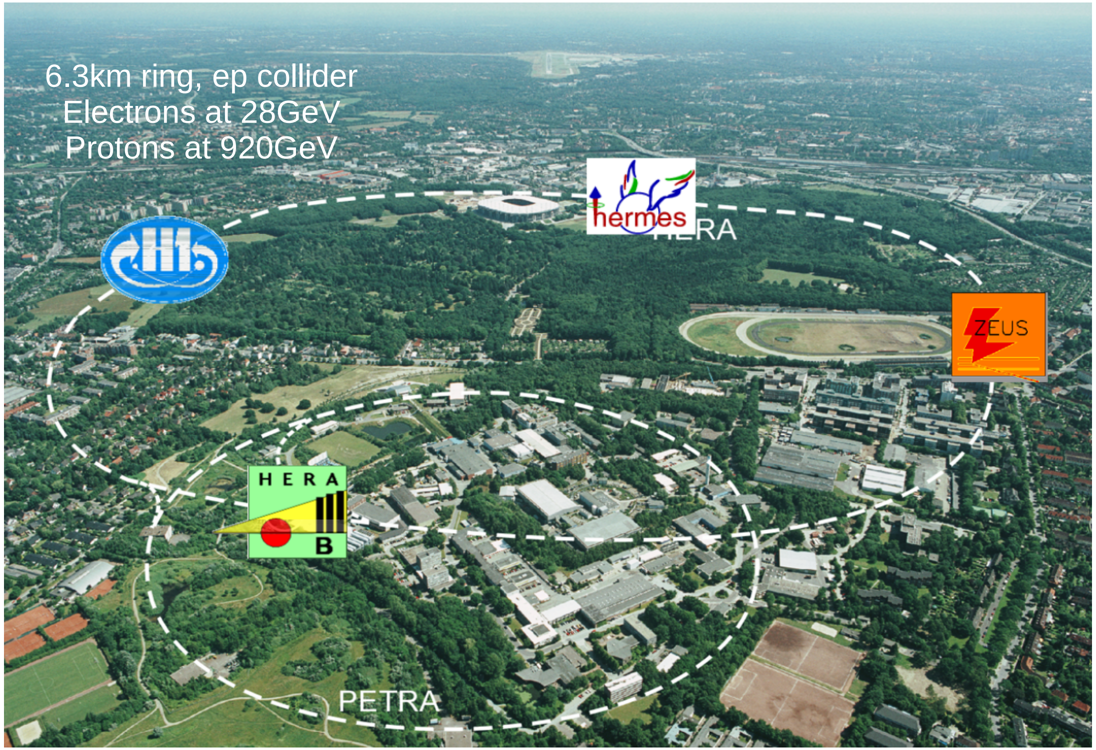


Huge amount of data! Very wide range of x, Q^2

~**1500** datapoints

Collider DIS - HERA

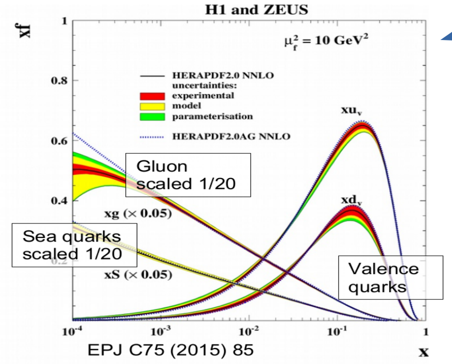
- HERA at DESY (1992-2007).
- One of aims was precise measurement of structure functions at lower x and higher Q^2.
- Range of Q^2 gives access partons over wide range.
- Also observe scaling violations – access indirectly gluon PDF.

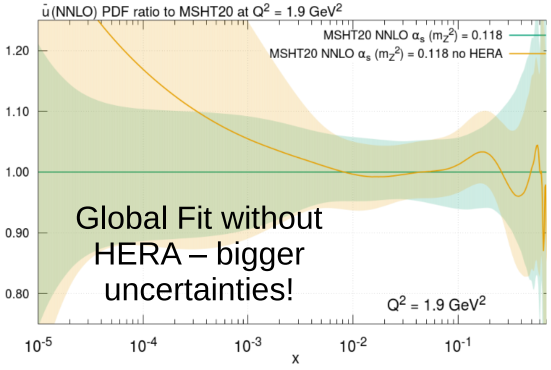


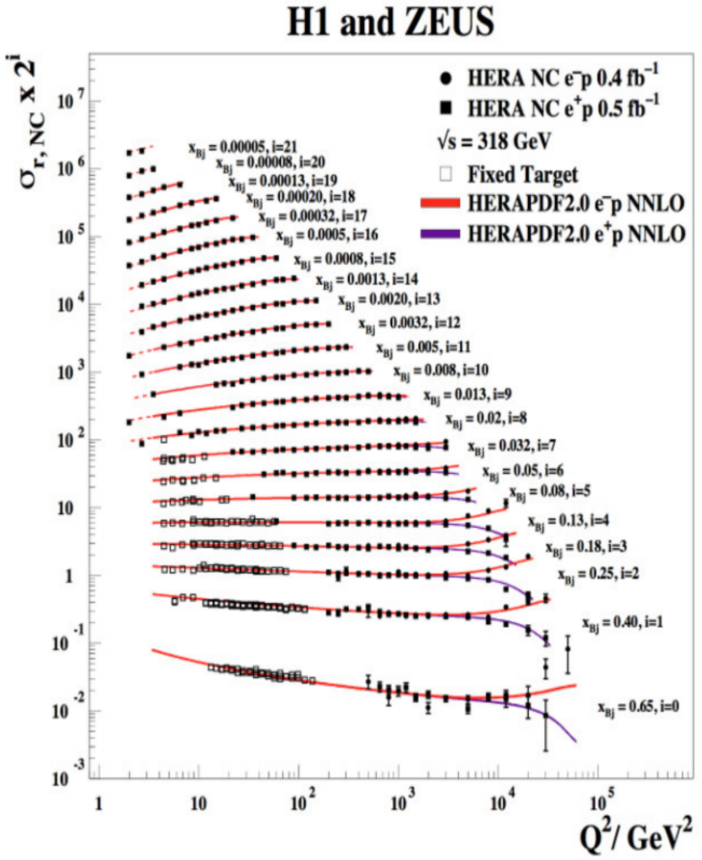
- Backbone of PDF fits! ~1500 datapoints.

PDF Constraints - HERA

- 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.
- F_3 (parity violating component) comes in at high Q^2.
- Can do PDF fits using only HERA – HERAPDF sets.





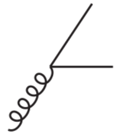


- Also measure heavy quark components – F_2^c, F_2^b, constrain heavy quarks.

Heavy quarks in PDFs

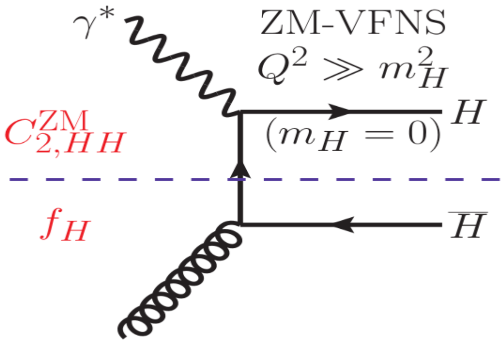
4

- 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:

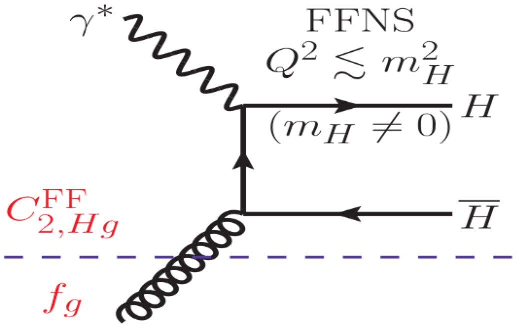


High $Q^2 \gg m_b^2, m_c^2$

Zero-Mass Variable Flavour Number Scheme (ZMVFNS)



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



Fixed Flavour Number Scheme (FFNS)

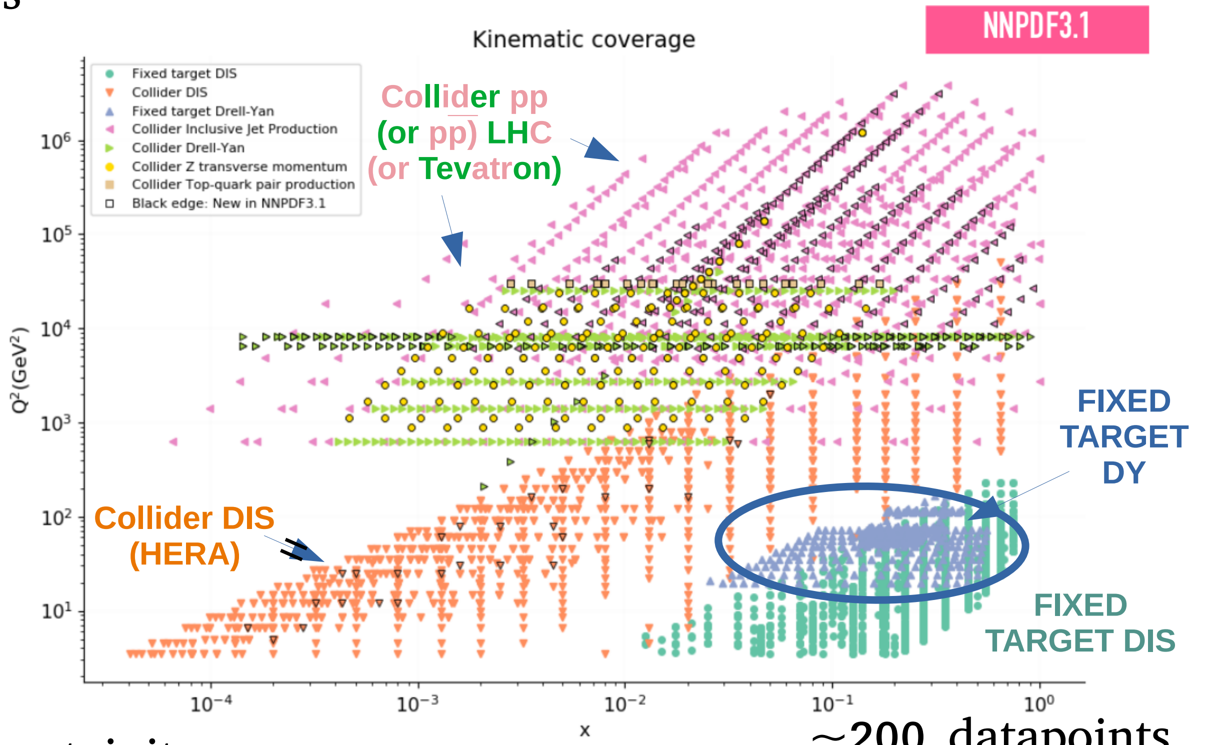
- Connect both regimes to get description over whole Q^2 range – General Mass Variable Flavour Number Scheme (GMVFNS). Measurement of F_2^c, F_2^b, constrains heavy quarks.

PDF Constraints - Fixed Target DY

4

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.

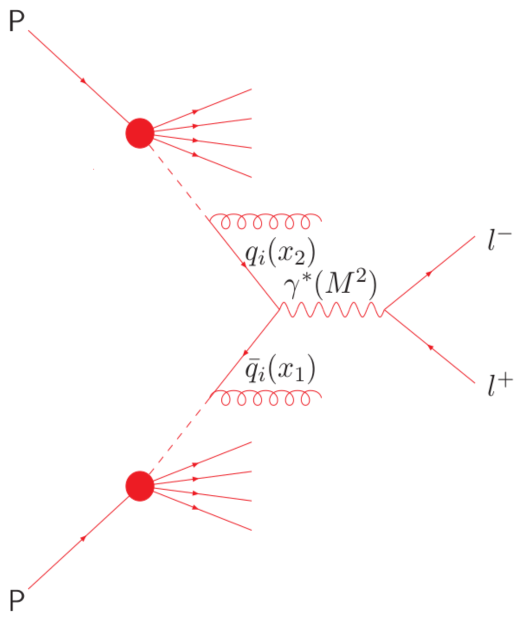


~200 datapoints

- Constrains central values and uncertainites.

PDF Constraints – Fixed Target DY ④

- Only sensitive to certain combinations of quark PDFs in DIS.
- Need other type of data to break degeneracy and constrain "flavour decomposition" – i.e. how many of each type of quark and antiquark rather than sums.
- Fixed Target Drell-Yan (e.g.) at Fermilab E605, 772, 866 (NuSea), 905(Seaquest) experiments.



$$\frac{d\sigma}{dM^2 dx_F} \propto \sum e_q^2 (q(x_1)\bar{q}(x_2) + q(x_2)\bar{q}(x_1)).$$

Photon mediated

- Q² relatively low (γ* exchange)– probe u_V, d_V, $\bar{u}$, $\bar{d}$ at high x.
- Higher Q² would be sensitive to other combinations via Z, W:

$$L_{ij}(x_1, x_2) = q_i(x_1)\bar{q}_j(x_2)$$

$$\gamma^* : \quad \frac{d\sigma}{dy dM^2} = \frac{4\pi\alpha^2}{9M^2 S} \sum_i e_i^2 L_{ij}(x_1, x_2)$$

$$Z : \quad \frac{d\sigma}{dy} = \frac{\pi G_F M_V^2 \sqrt{2}}{3S} \sum_i (v_{iZ}^2 + a_{iZ}^2) L_{ij}(x_1, x_2)$$

$$W : \quad \frac{d\sigma}{dy dM^2} = \frac{\pi G_F M_V^2 \sqrt{2}}{3S} \sum_{ij} |V_{ij}^{\mathrm{CKM}}|^2 L_{ij}(x_1, x_2)$$

PDF Constraints – Fixed Target DY

4

- Difference between $\bar{u}$, $\bar{d}$ at high x still difficult to disentangle.
- Therefore again can consider both proton and deuteron targets!

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

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

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

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

$$\boxed{\begin{aligned}\frac{pd}{2pp} &= (pp+pn)/2pp \\ &\approx 0.5\left(1+\frac{\bar{d}(x2)}{\bar{u}(x2)}\right)\end{aligned}} \left(1+\cancel{0.25\frac{d(x_1)}{u(x_1)}}]\right) / \left(1+\cancel{0.25\frac{d(x_1)\bar{d}(x_2)}{u(x_1)\bar{u}(x_2)}}\right)$$

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

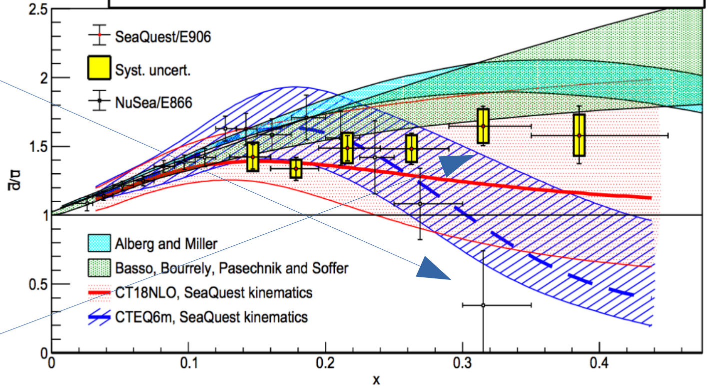
- What does Fixed Target DY at E866(NuSea) tell us then?
- Remember

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

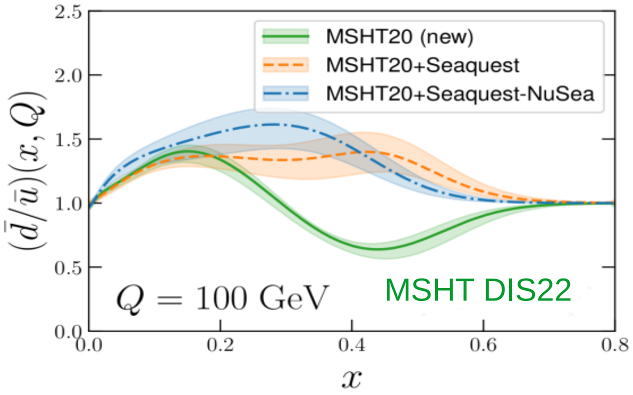
→ data implies $\bar{d}<\bar{u}$ at high x

- 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$\bar{\text{p}}$

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$\bar{\text{p}}$) colliders.
- Constrains central values and uncertainites.

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

$\sim$**1800** datapoints.

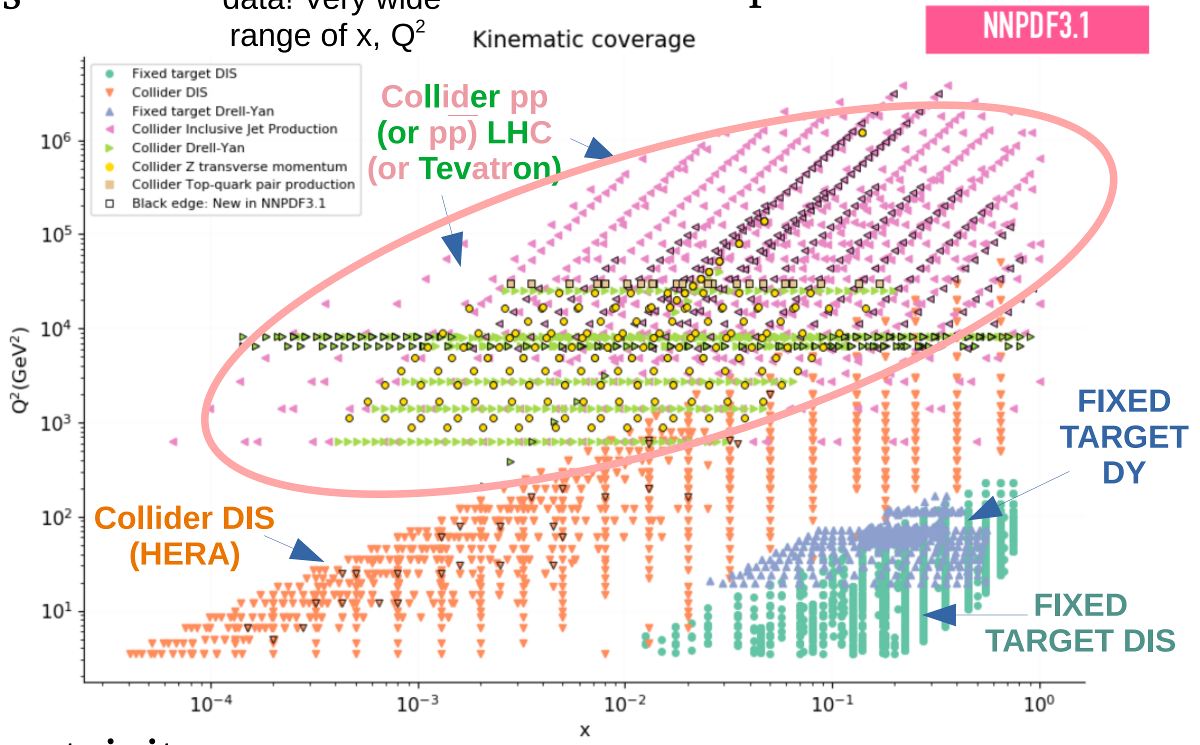


Figure adapted from NNPDF3.1 (1706.00428)

PDF Constraints – Tevatron DY

- Proton colliders also provide a lot of information for modern global fit PDFs.
- Tevatron at Fermilab – $p\bar{p}$ collider at $\sqrt{s}$=1.96TeV from 1983-2011.
- Higher $Q^2 \rightarrow Z, W$ mediated (rather than γ) and lower x probed.
- W^+ W^- Asymmetry gives further info:

Also Tevatron jets data – discussed later for LHC jets!

$$Z: \frac{d\sigma}{dy} = \frac{\pi G_F M_V^2 \sqrt{2}}{3S} \sum_i (v_{iZ}^2 + a_{iZ}^2) L_{ij}(x_1, x_2)$$

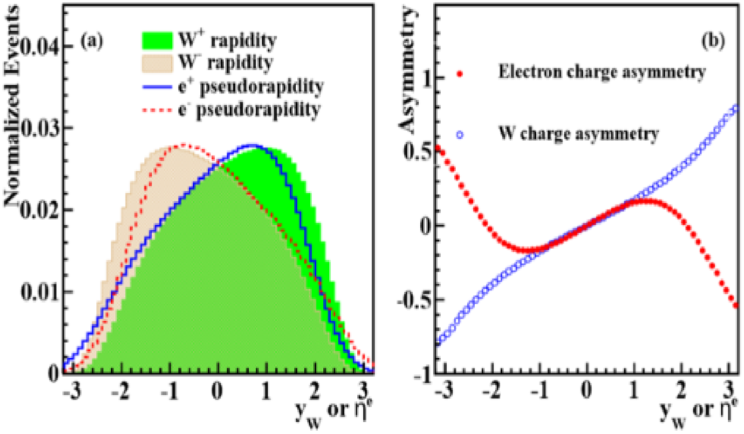
$$W: \frac{d\sigma}{dy dM^2} = \frac{\pi G_F M_V^2 \sqrt{2}}{3S} \sum_{ij} |V_{ij}^{\mathrm{CKM}}|^2 L_{ij}(x_1, x_2)$$

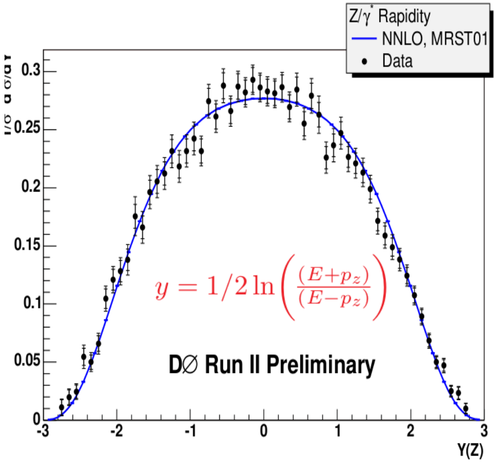
V-A structure of lepton decay means e+/- emitted opposite to W+/-.

Assume u from p and $\bar{d}$ from $\bar{p}$ to simplify

W+/- produced dominantly in p/$\bar{p}$ direction

$$A_W(y) = \frac{d\sigma(W^+)/dy - d\sigma(W^-)/dy}{d\sigma(W^+)/dy + d\sigma(W^-)/dy} \approx \frac{u(x_1)d(x_2) - d(x_1)u(x_2)}{u(x_1)d(x_2) + d(x_1)u(x_2)},$$





• PDF Constraints – ATLAS/CMS DY ④

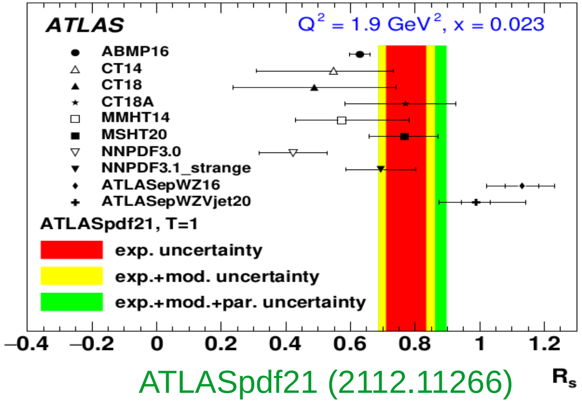
- Since 2010 lots of LHC data has been produced, adds further constraints on PDFs!
- Includes Drell-Yan of course, higher Q^2 still so lower x and now pp so complementary PDF sensitivity.

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

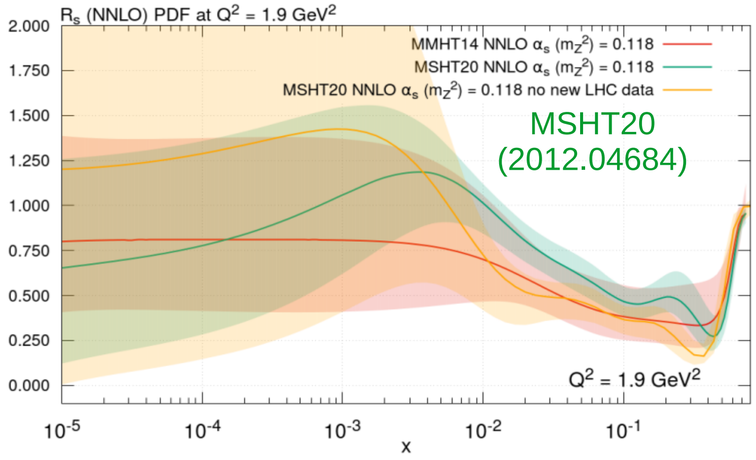
- ATLAS 7TeV W,Z is high precision. (CMS results as well)
- Up and down quarks/antiquarks already quite constrained from non-LHC data, therefore precision means **strangeness is most constrained.**

Strangeness enhancement!

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



ATLASpdf21 (2112.11266)



MSHT20 (2012.04684)

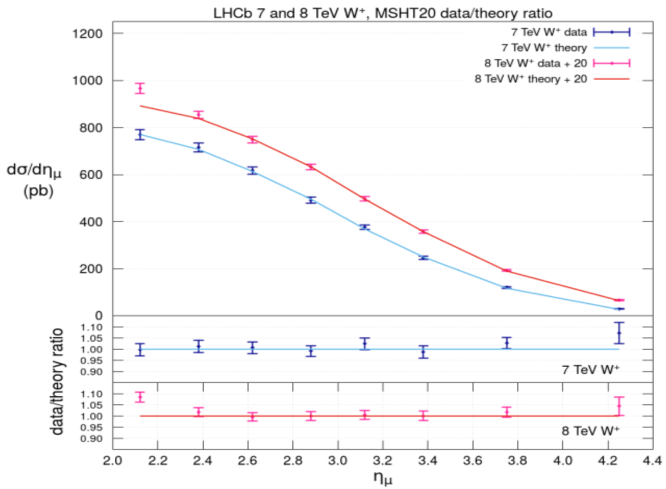
• PDF Constraints – LHCb DY

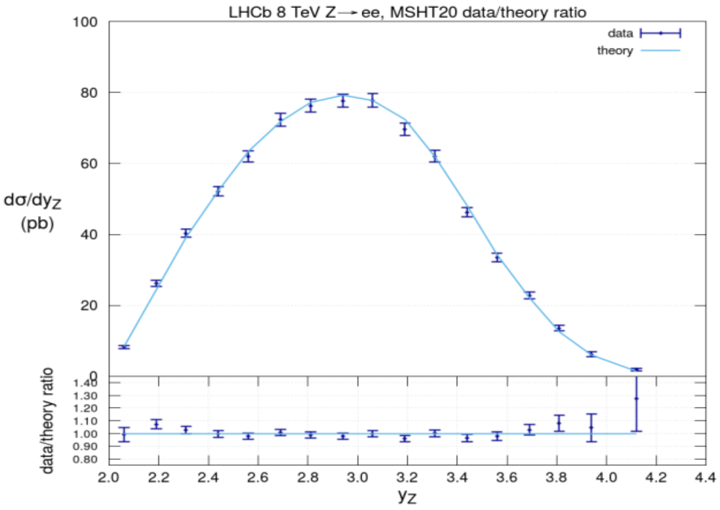
4

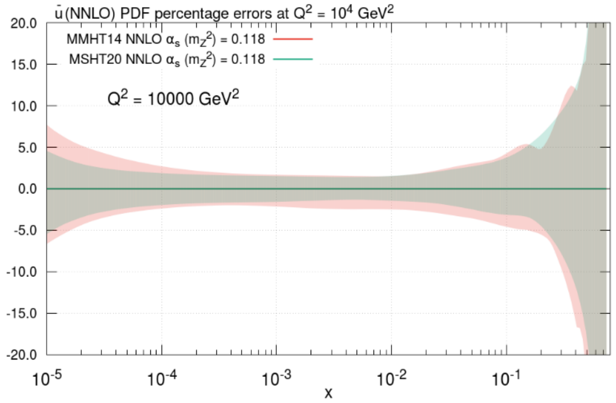
- LHCb is a forward detector experiment, rapidities in range **2-4.5** → help to constrain valence quarks at higher x and in particular sea quarks at lower x than possible ATLAS/CMS.
  → good fits obtained in global PDF fits:

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

Figs from MSHT20 (2012.04684)



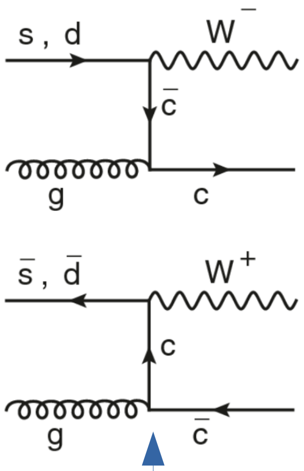




• PDF Constraints – W/Z + charm

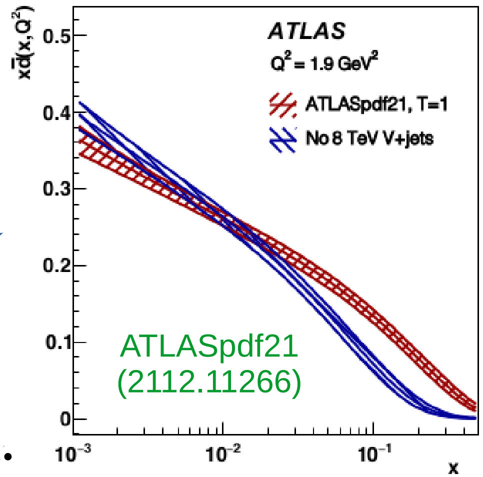
④

- 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.

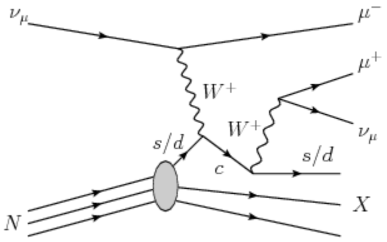
Need IRC safe observable for theory calculation.

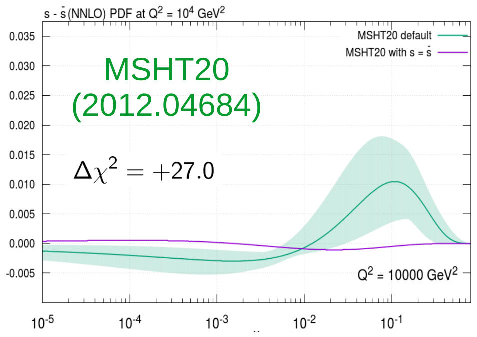


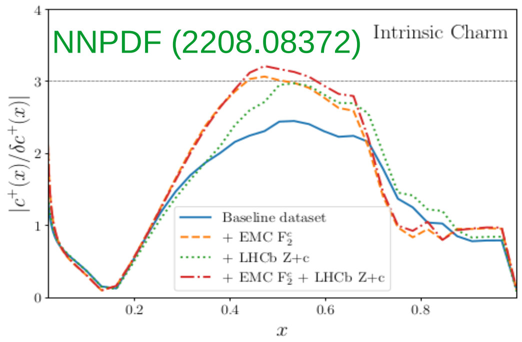
• PDF Constraints – Intrinsic charm? 4

- 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 – no evidence of asymmetry yet.
- 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)
- 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 ④

- 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).

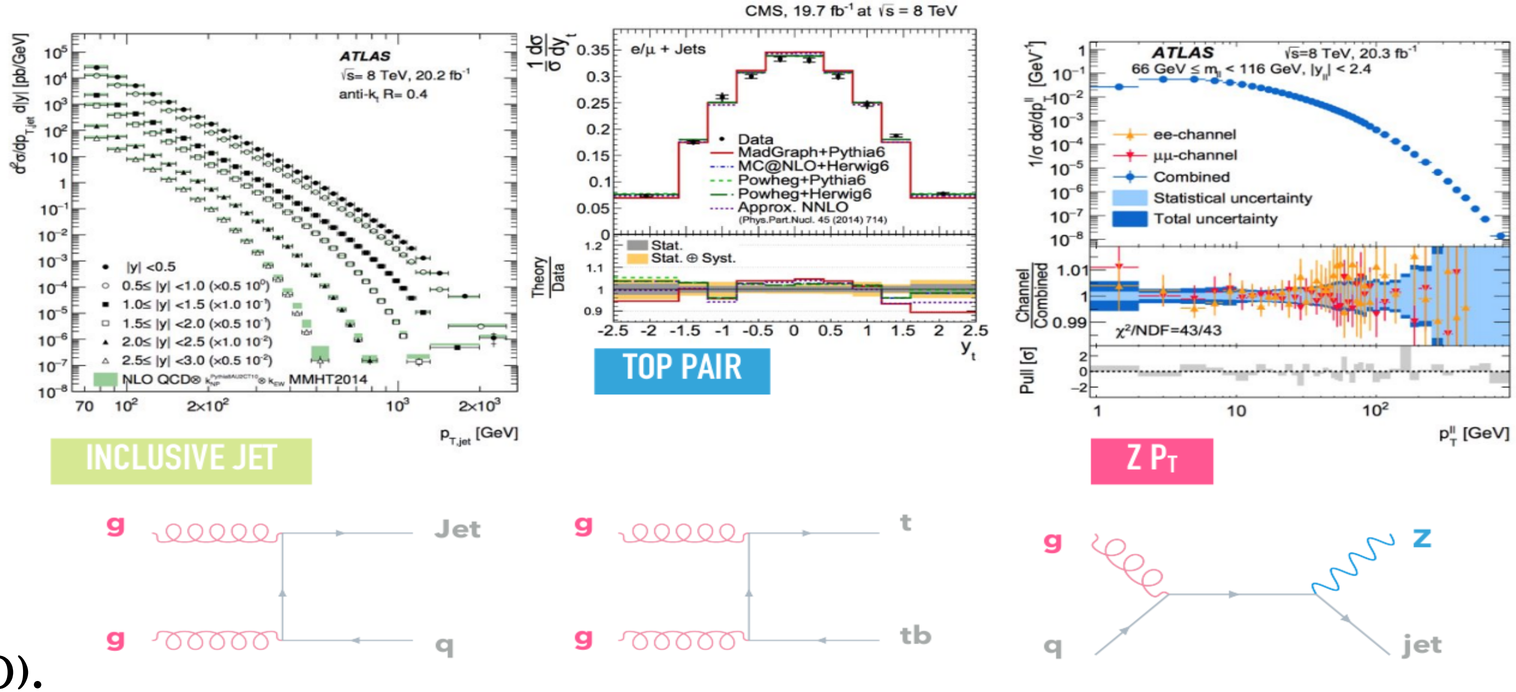
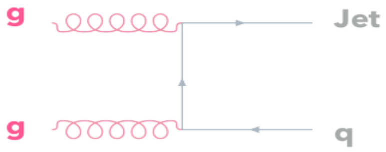


Fig from M. Ubiali

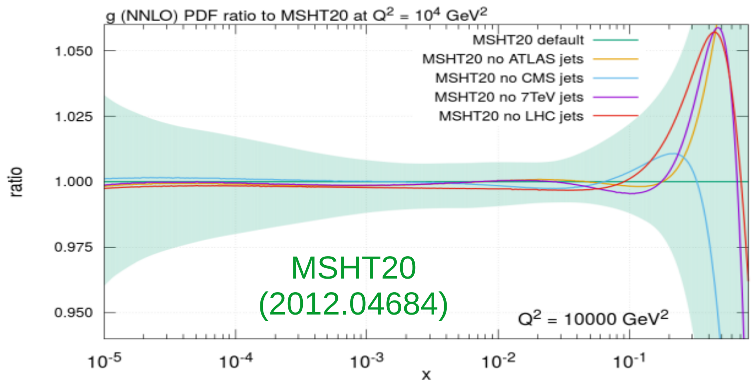
• PDF Constraints – LHC jets

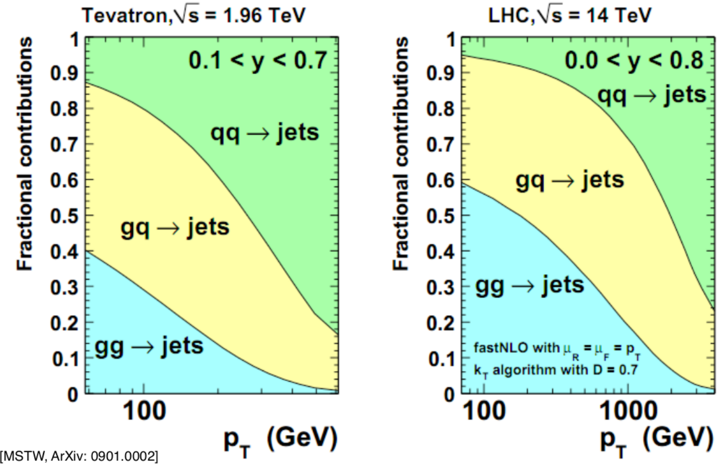
4

- Inclusive jet data sensitive to high x gluon, more so at LHC than Tevatron.
- ATLAS and CMS 7 and 8TeV inclusive jet data impact gluon.
- Some tensions observed between datasets.
- Also some issues in fitting the data: "2-point systematics" where two MCs used and difference taken as fully correlated uncertainty
- This causes issues in PDF fit → need to decorrelate.



More info in ATLAS: 1706.03192

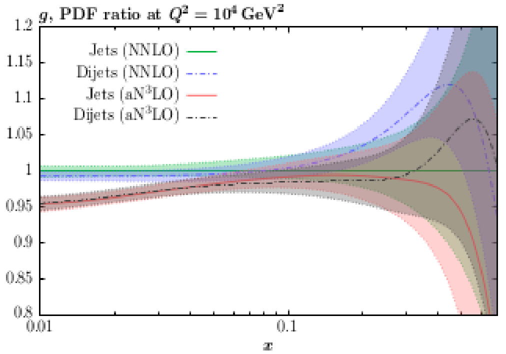




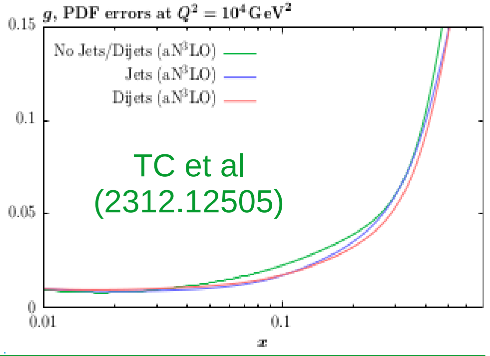
[MSTW, ArXiv: 0901.0002]

• PDF Constraints – Jets/Dijets?

- Dijets may have some advantages here - 3D measurement now possible, non-unitary nature of inclusive jets, etc
- We have also investigated dijets instead:
  - Obtain better fit quality at NNLO and aN3LO than inclusive jets.
  - Moreover, dijet fit quality improves further slightly at aN3LO.



Inclusive Jets	N_{pts}	χ^2/N_{pts} NNLO	χ^2/N_{pts} aN3LO	Dijets	N_{pts}	χ^2/N_{pts} NNLO	χ^2/N_{pts} aN3LO
Total	472	1.39	1.43	Total	266	1.12	1.04
Total (+ATLAS 8 TeV jets)	643	1.67	1.61	Total	266	1.12	1.04

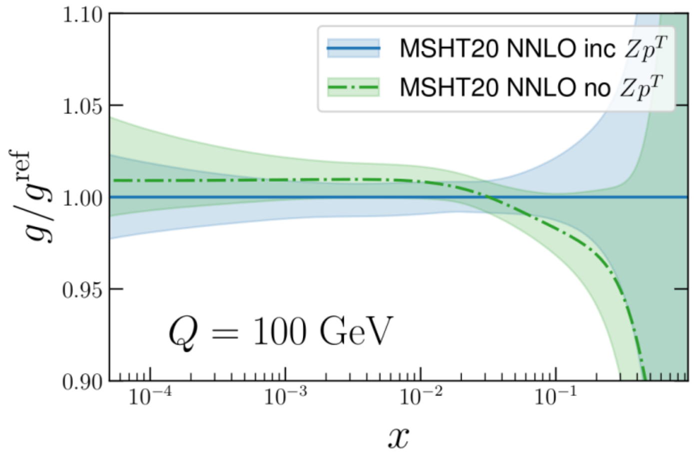


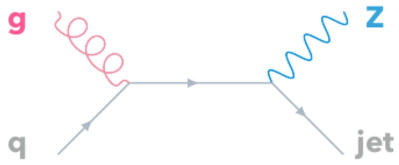
- Limited effect on PDFs at aN3LO - gluon consistent between dijets/inclusive jets. Dijets slightly more constraining on gluon.
- Results here leading colour, full colour effects limited on PDFs.

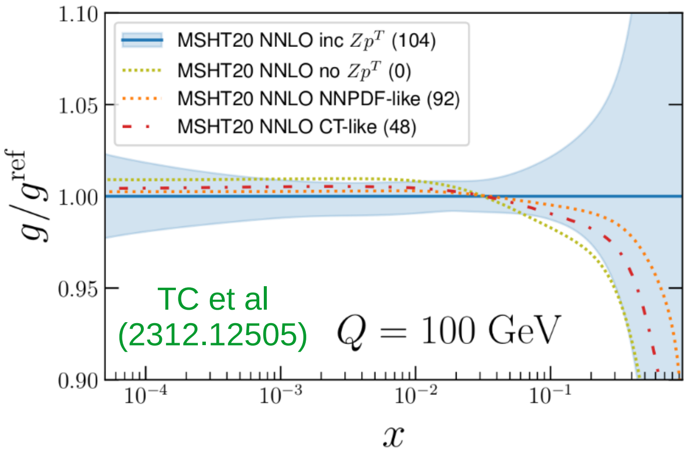
• PDF Constraints – ZpT data

4

- ATLAS 8TeV Z pT data also sensitive to gluon.
- Very precise data ~1% uncertainties out to large pT.
- Has large NNLO corrections.
- Different global fit PDF groups fit different amounts of data and with different assumptions and uncertainties applied → see slightly different impacts.
- MSHT see largest impact – upwards pull on high x gluon – and fit most data (104 datapoints). Also recent evidence of improved fit at aN3LO (see later!)



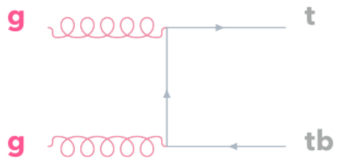




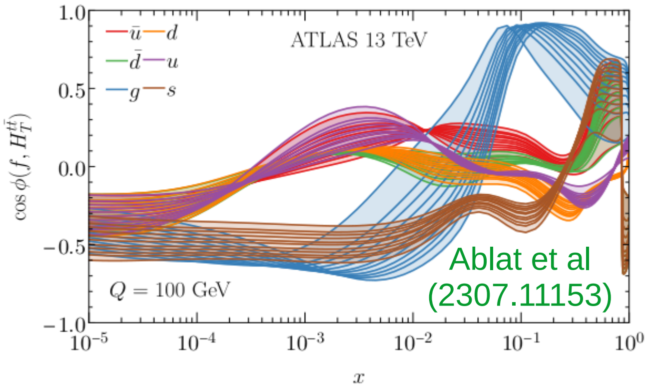
• PDF Constraints – Top data

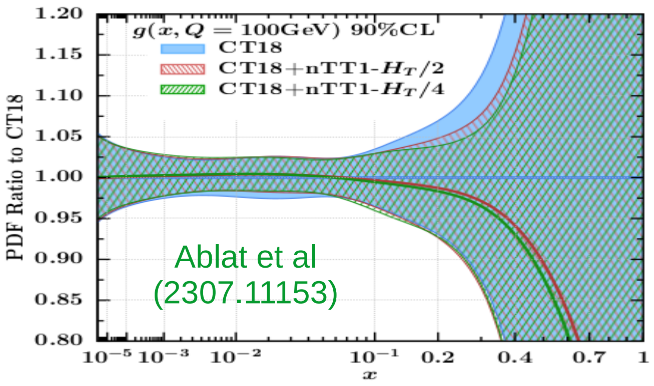
4

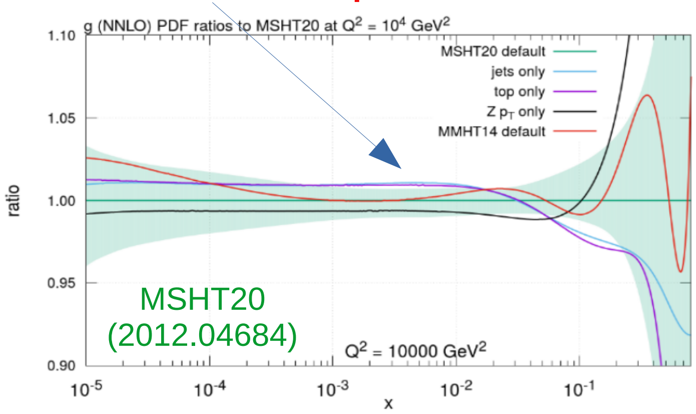
- Top total/differential cross-sections also sensitive to gluon at high x.
- CMS single/double differential top quark data and ATLAS multi-differential data provided with correlations.
- Also some issues with systematics observed in latter.
- Generally reduce uncertainty on high x gluon.
- Some tensions between Top, ZpT, Jets data – global fit is balance of pulls.



Harland-Lang et al 1909.10541



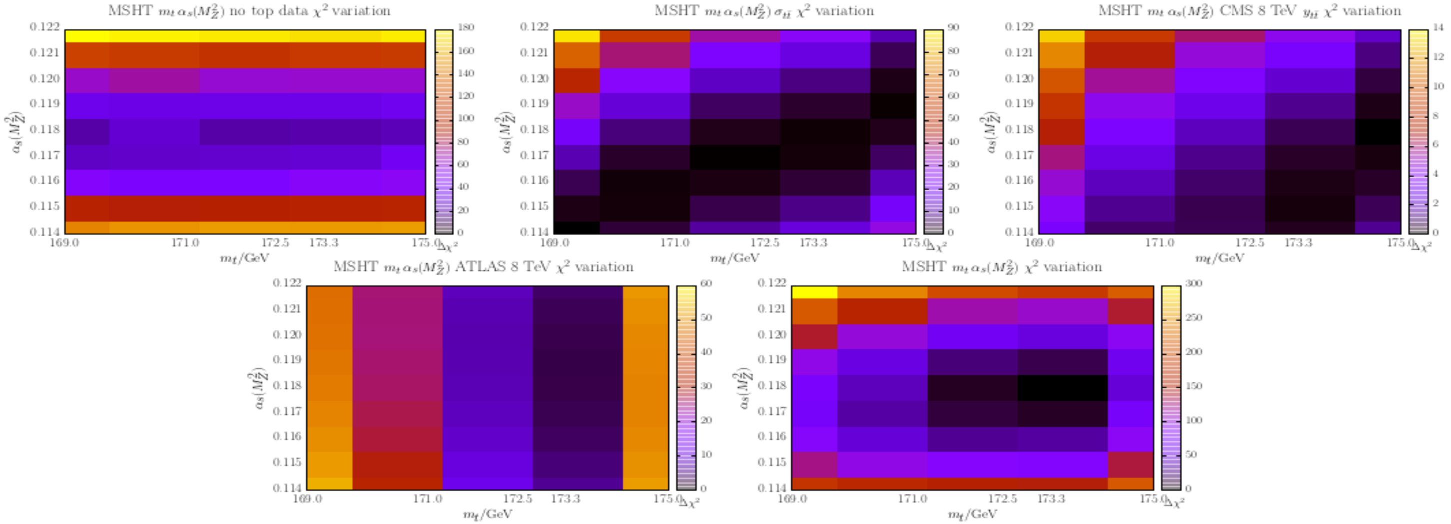




• PDF Constraints – Top data - mt

4

- 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 α_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:

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

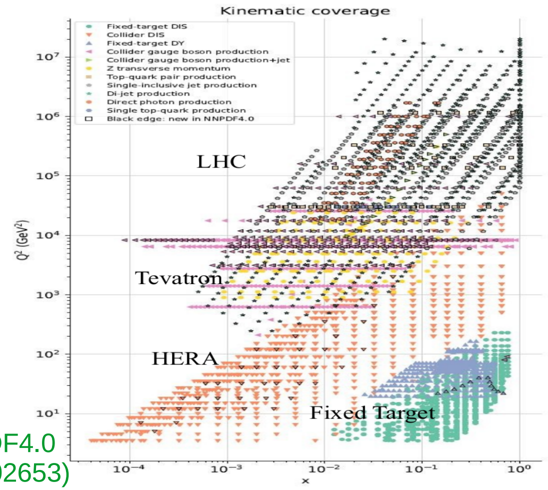
5. PDF Methodology and Uncertainties

Global PDF Fitting

3

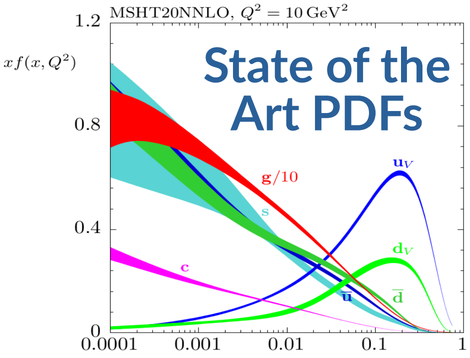
1) Experiment

- Latest experimental data
- Fixed [...] collider DIS, T[...] HC, etc
- EW [...] 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 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)

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

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

Experimental Nuisance parameters

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

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

- Must call theory predictions at each step of minimisation – use theory “grids” → Applgrid, FastNLO, etc, available at NLO + NNLO K-factors, or increasingly NNLO.
- 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

- 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 to set 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

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

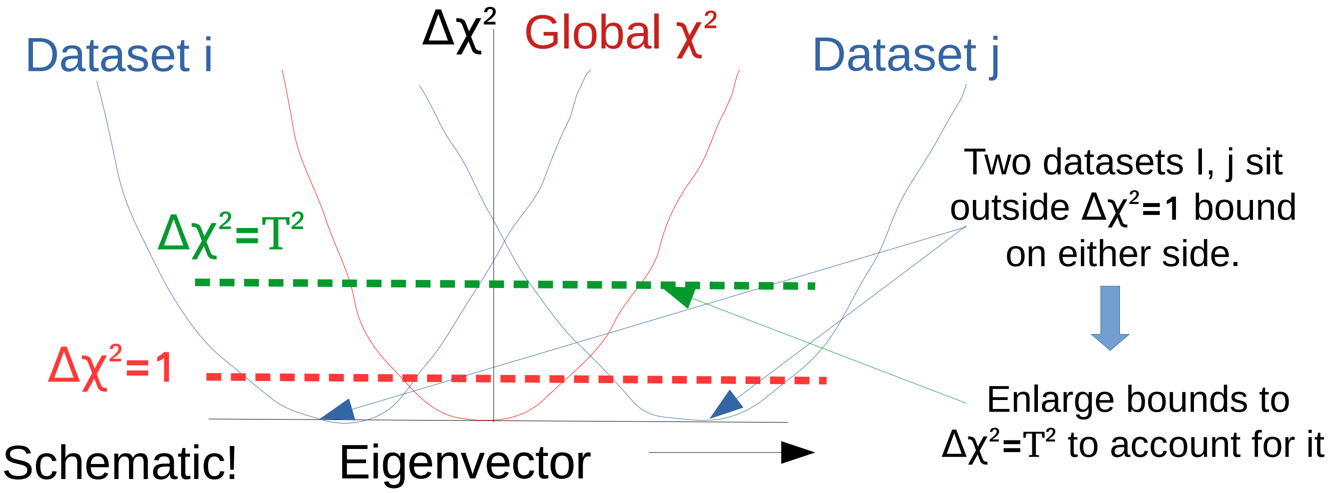
Can convert between the two forms (1205.4024,1401.0013, 1505.06736)

PDF Methodology - Tolerance

5

- How are dataset tensions accounted for in the uncertainty?
- In replica approach these mean you have a non-Gaussian distribution with outliers, *may* enlarge uncertainty or *may not*.
- In Hessian approach you can enlarge $\Delta\chi^2$ criterion used to reflect dataset tensions. $\Delta\chi^2=1$ may not be appropriate due to dataset tensions, issues of systematics, missing theory contributions etc → CT and MSHT use a "tolerance", $\Delta\chi^2=T^2$.

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



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

Dataset Tension - L2 Sensitivity

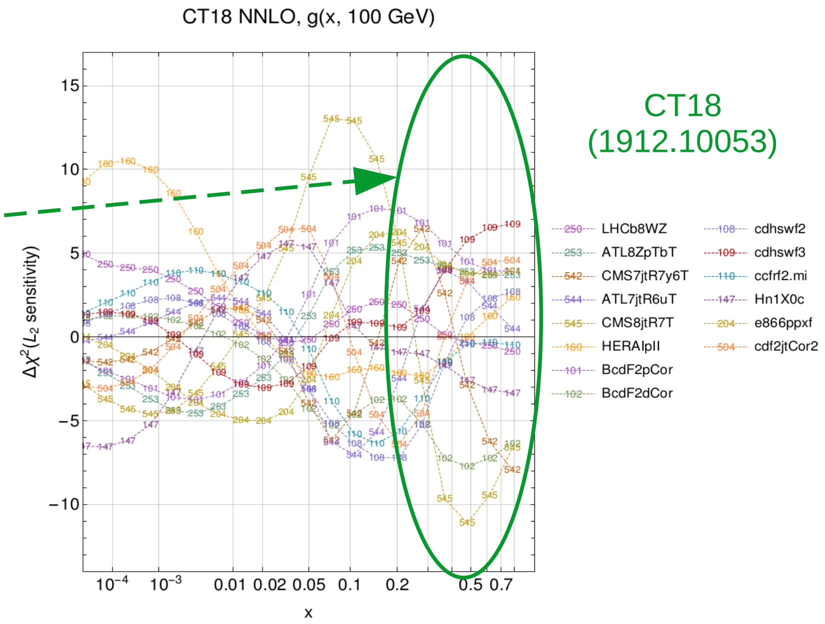
4

- Seen a few examples now of "dataset tensions".
- May be due to several effects; including fluctuations, data/systematic issues, missing higher order theory, etc.
- Limits reduction of PDF uncertainty.
- Can visualise via L2 sensitivity – think of as $\Delta\chi^2$ of dataset upon moving PDF by "1σ".
- Illustrates dataset tensions, e.g. high x gluon

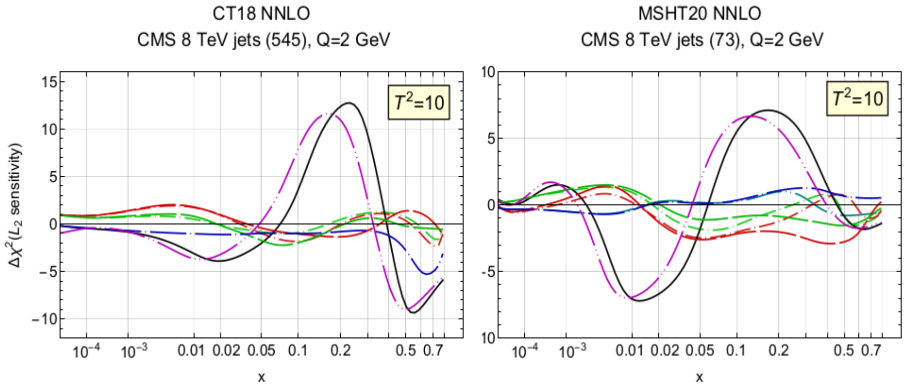
Publicly available program to calculate it.

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

X Jing et al (2306.03918)



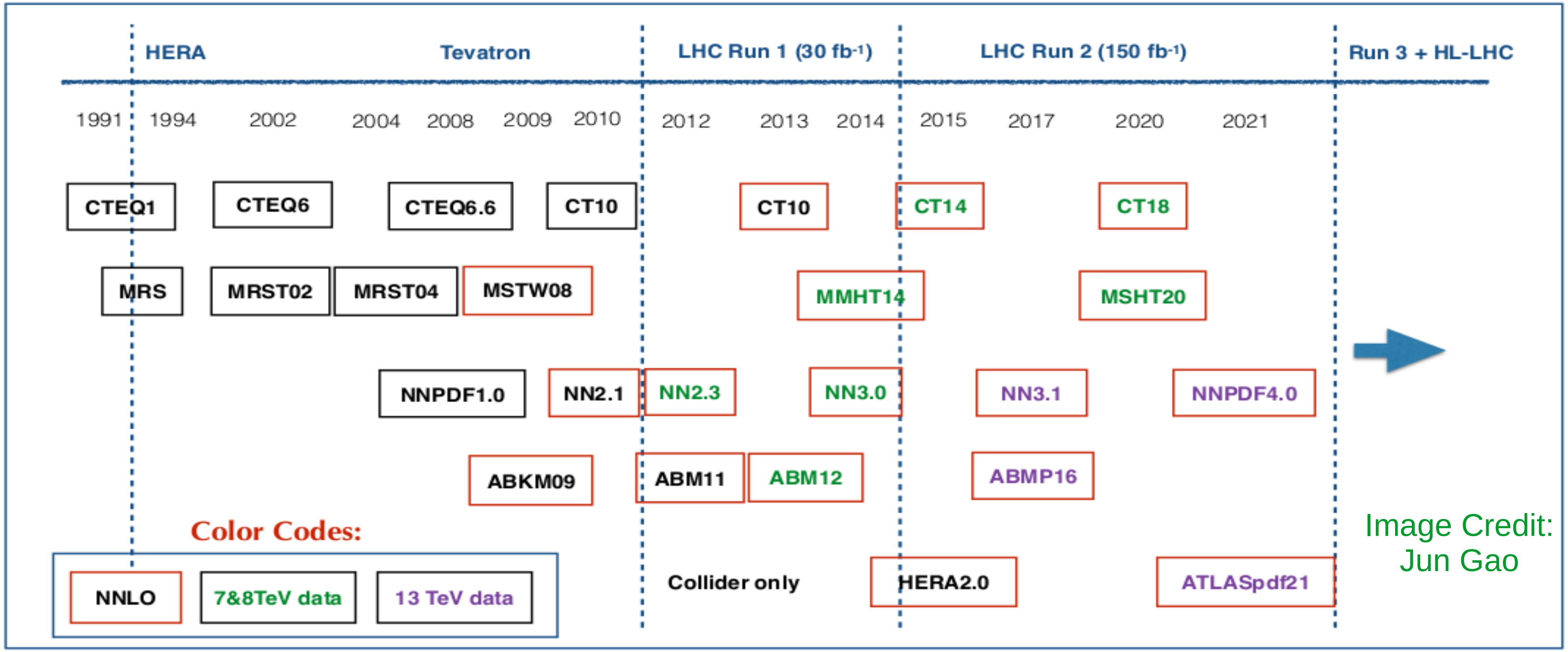
CT18 (1912.10053)



6. Current PDF Landscape

Several 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

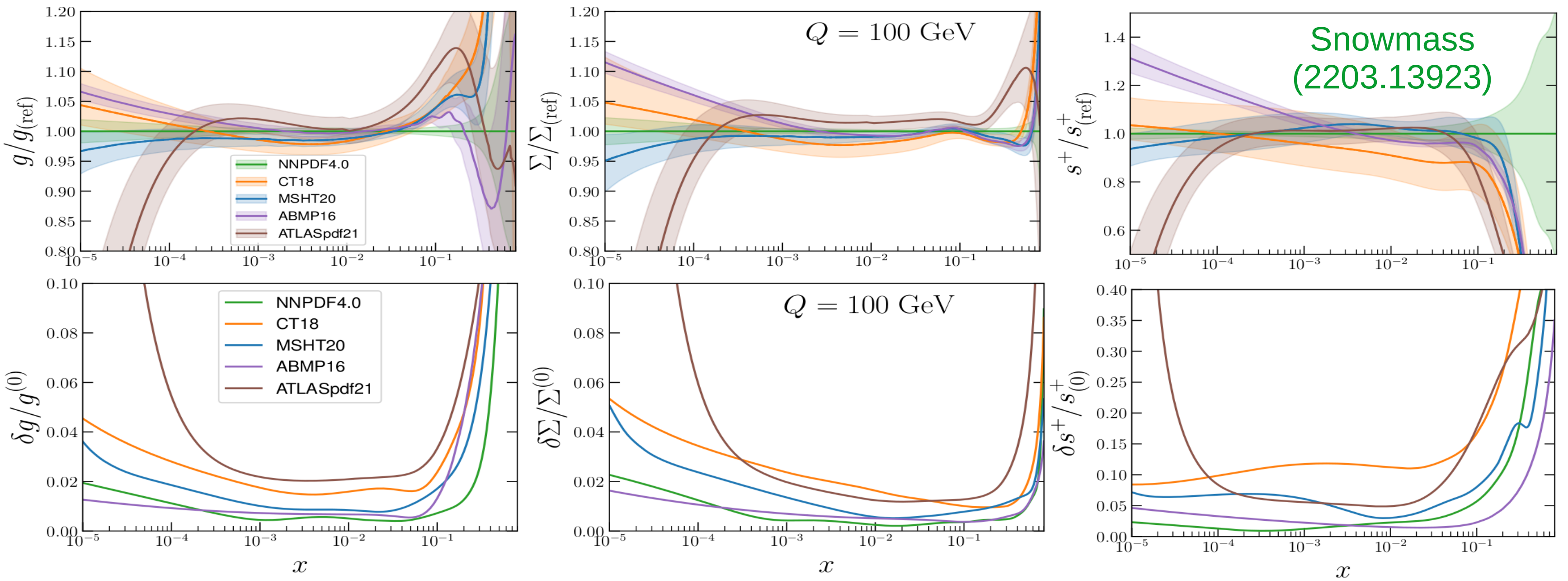
NNLO QCD + NLO EW and latest LHC and other data

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

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

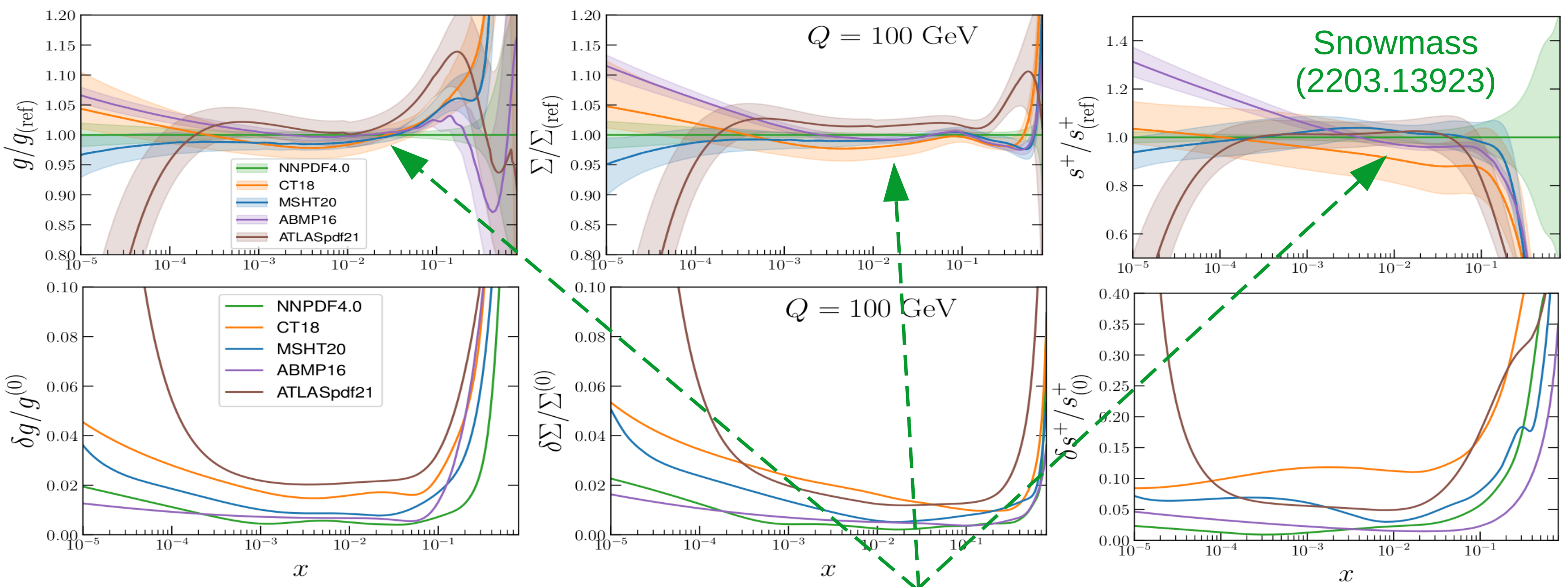
PDF Comparison

- 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:

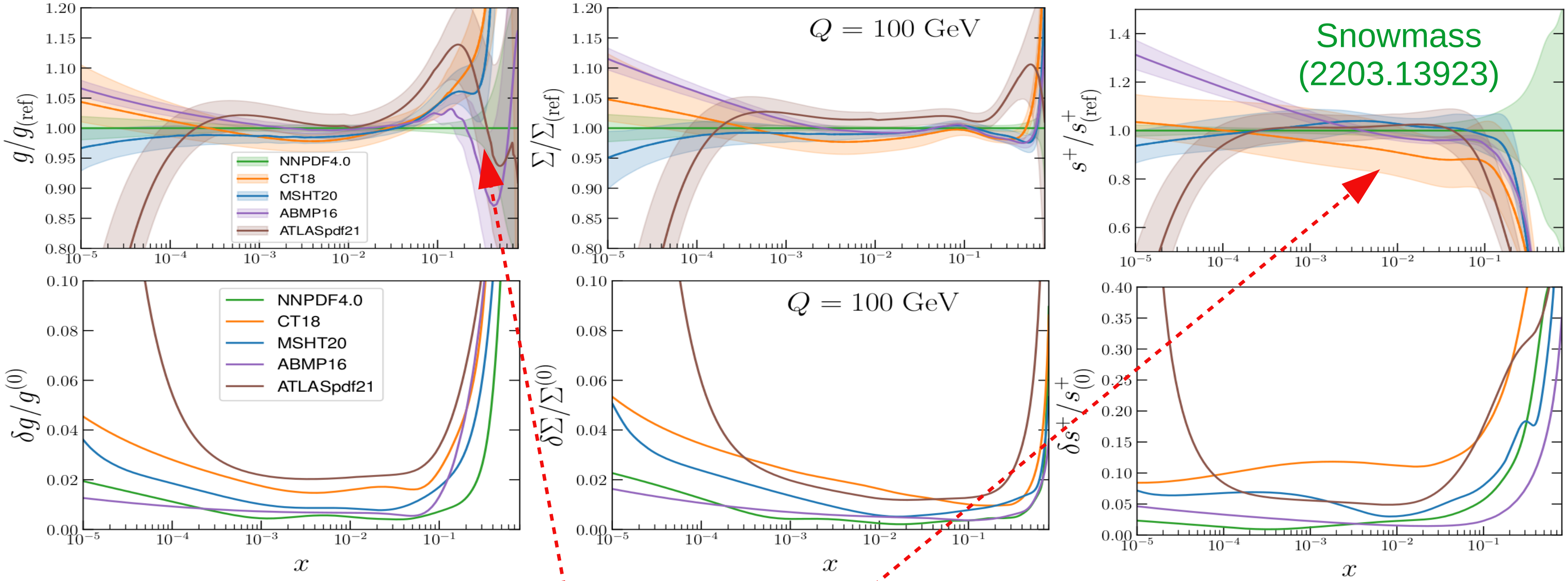


- General agreement over data range ($10^{-4} < x < 10^{-1}$) within uncertainties.

PDF Comparison

6

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

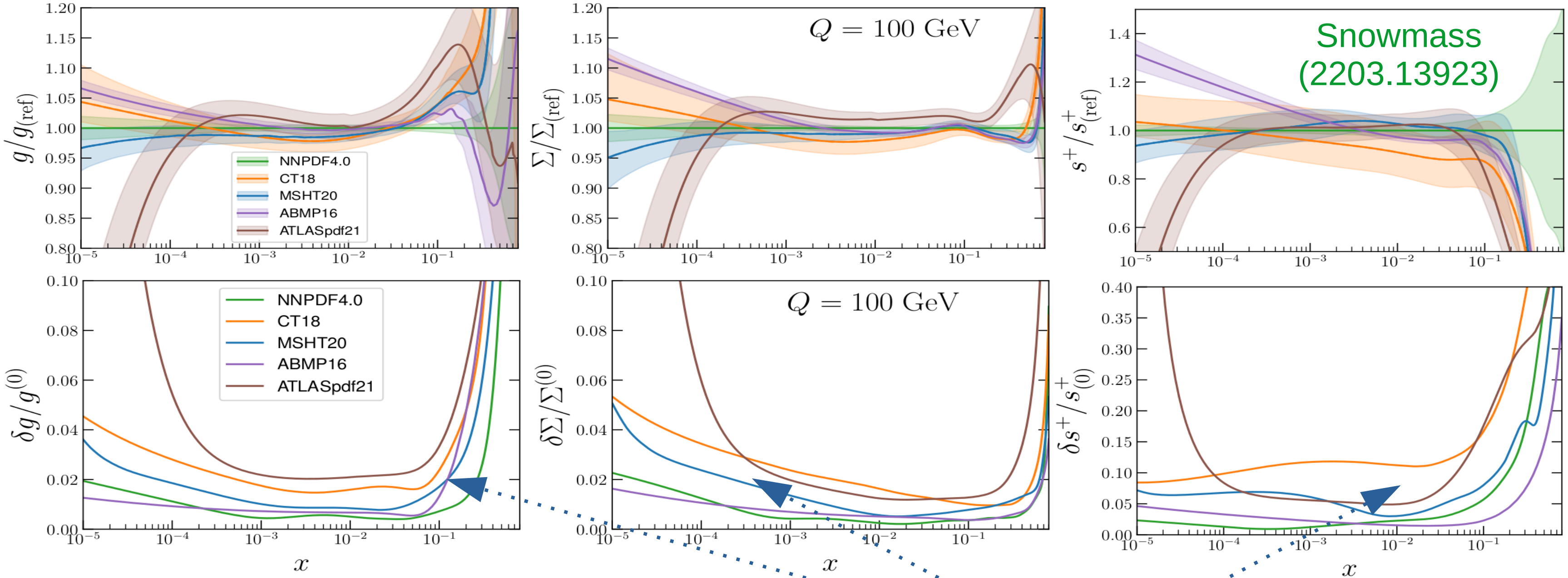


- General agreement over data range ($10^{-4} < x < 10^{-1}$) within uncertainties.
- Differences exist (high x gluon, strangeness, …).

PDF Comparison

6

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

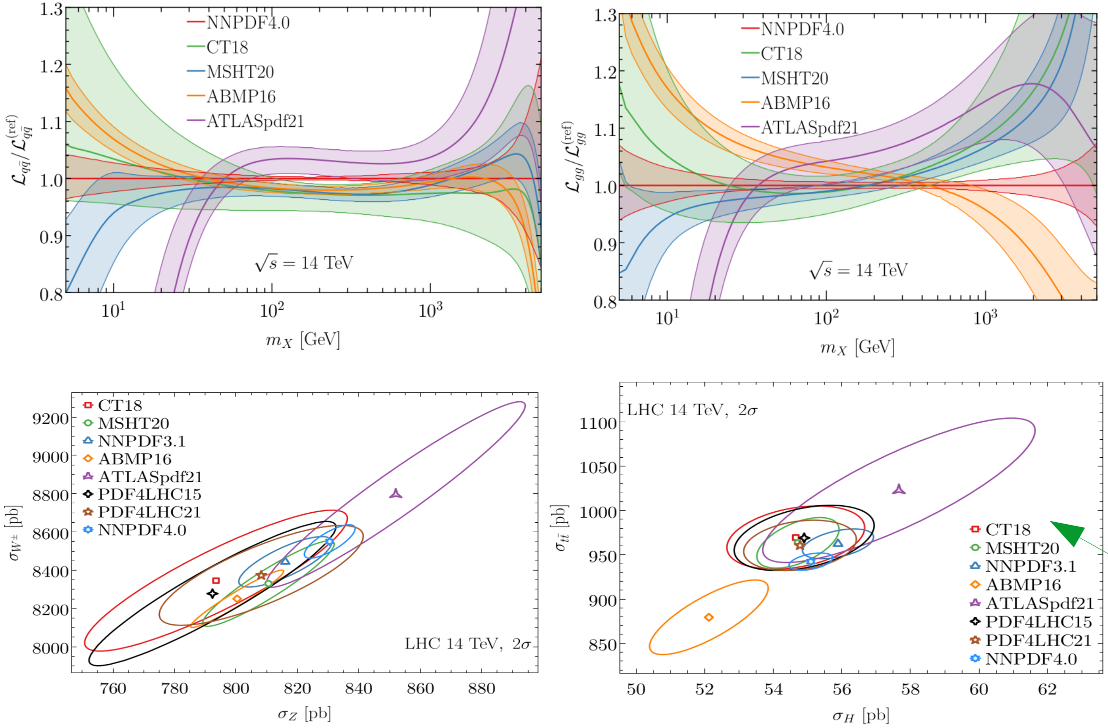


- General agreement over data range ($10^{-4} < x < 10^{-1}$) within uncertainties.
- Differences exist (high x gluon, strangeness, uncertainty sizes).

Parton luminosities and Xsecs

6

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



Snowmass (2203.13923)

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

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

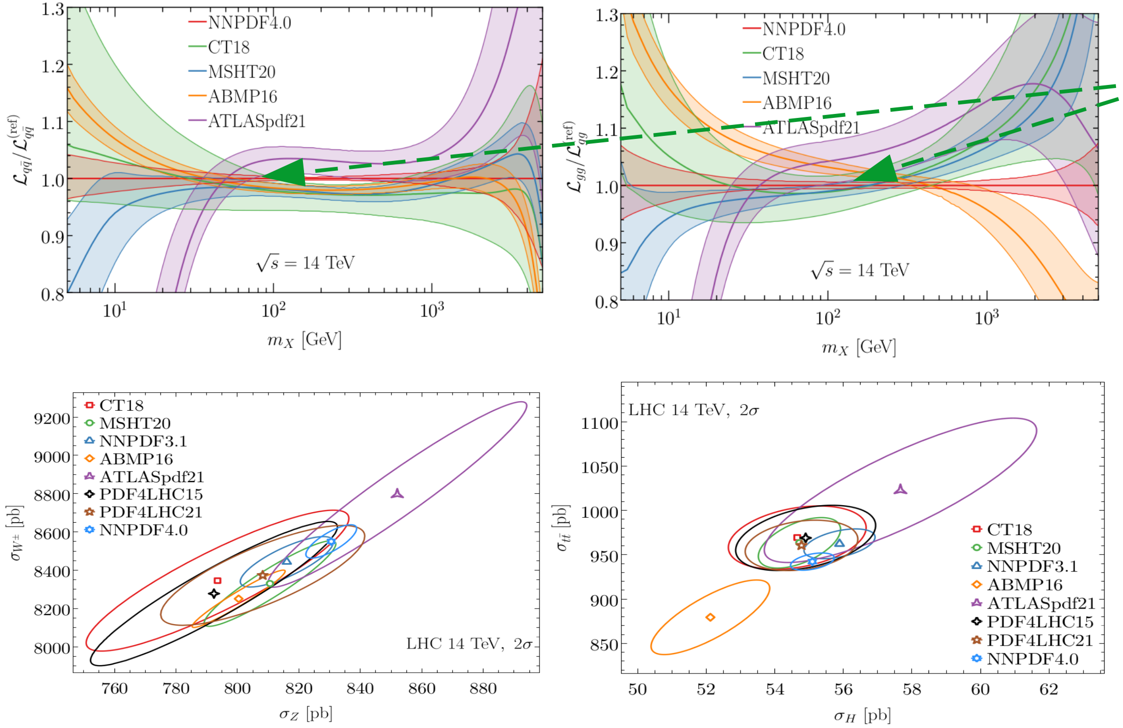
$$\sigma = \sum_{a,b=q,\bar{q},g} \int_{M^2}^{s} \frac{d\hat{s}}{\hat{s}} \mathcal{L}_{ab}(\hat{s}, \mu_F^2)\, \hat{s}\hat{\sigma}_{ab}(\hat{s}, M^2, \mu_R^2, \mu_F^2)$$

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

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

Parton luminosities and Xsecs

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

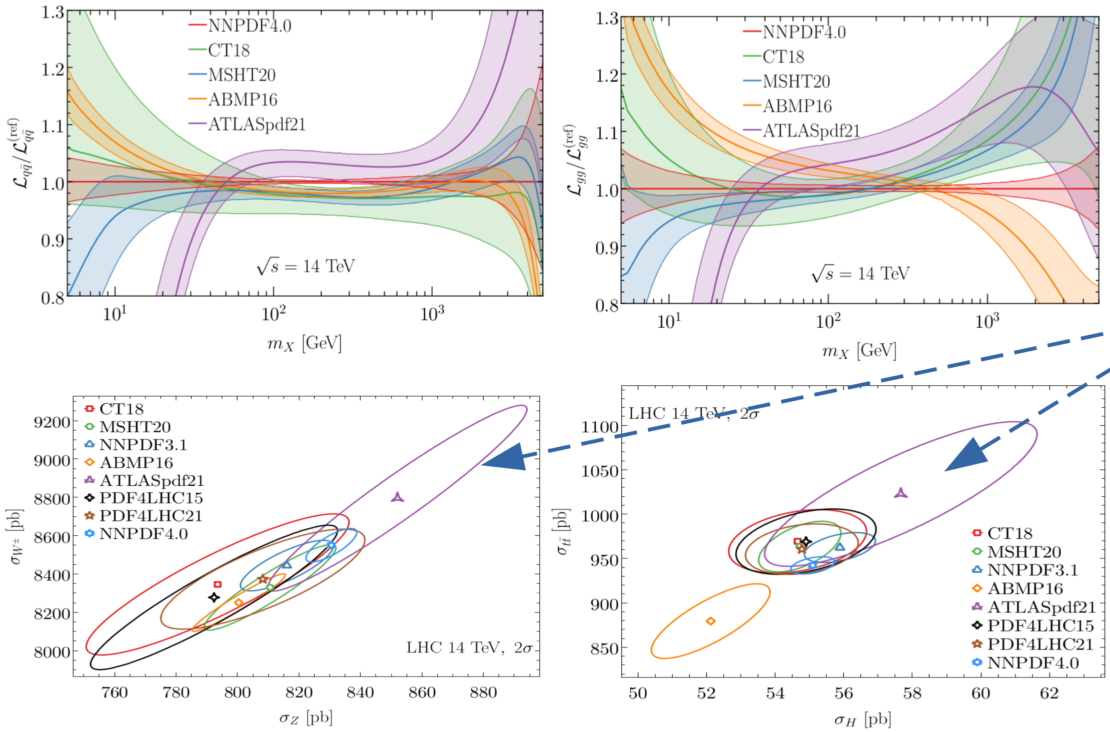


Snowmass (2203.13923)

- 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:



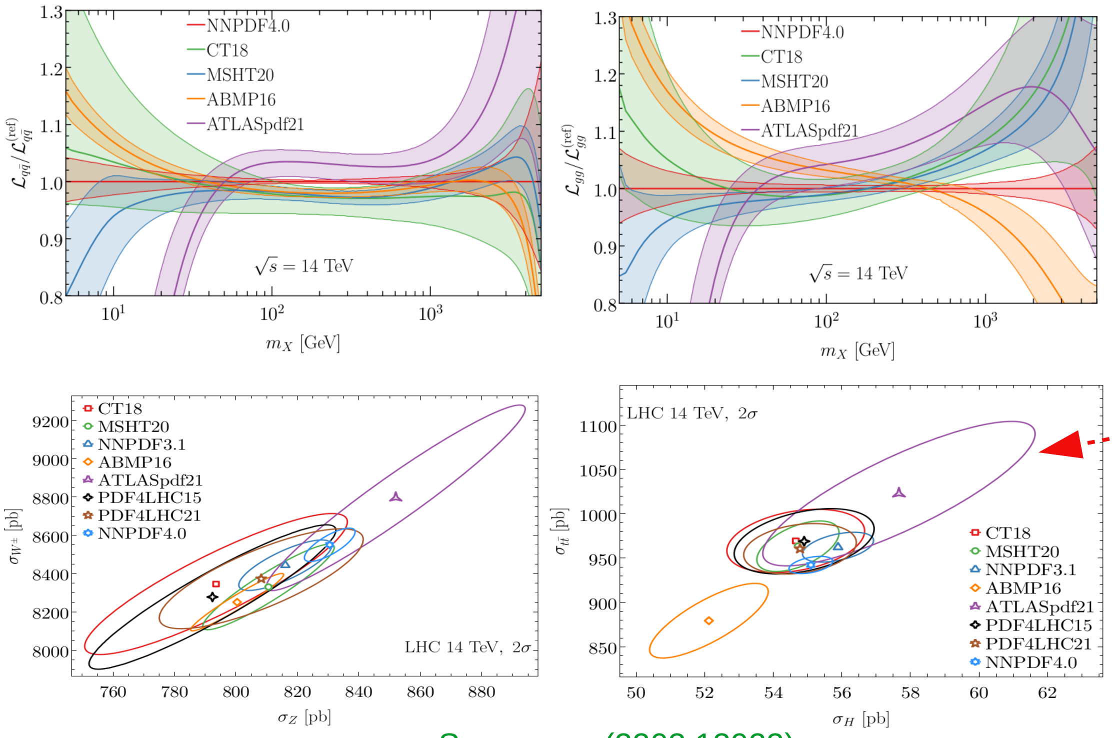
Snowmass (2203.13923)

- 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

6

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



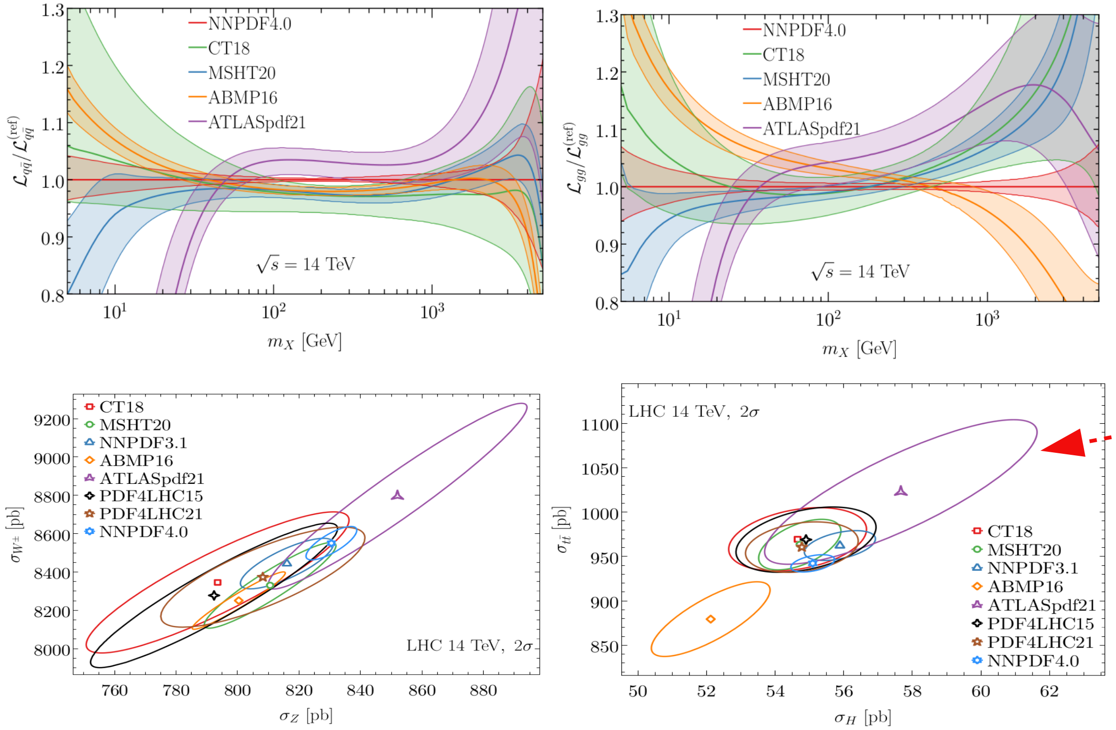
Snowmass (2203.13923)

- 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

6

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

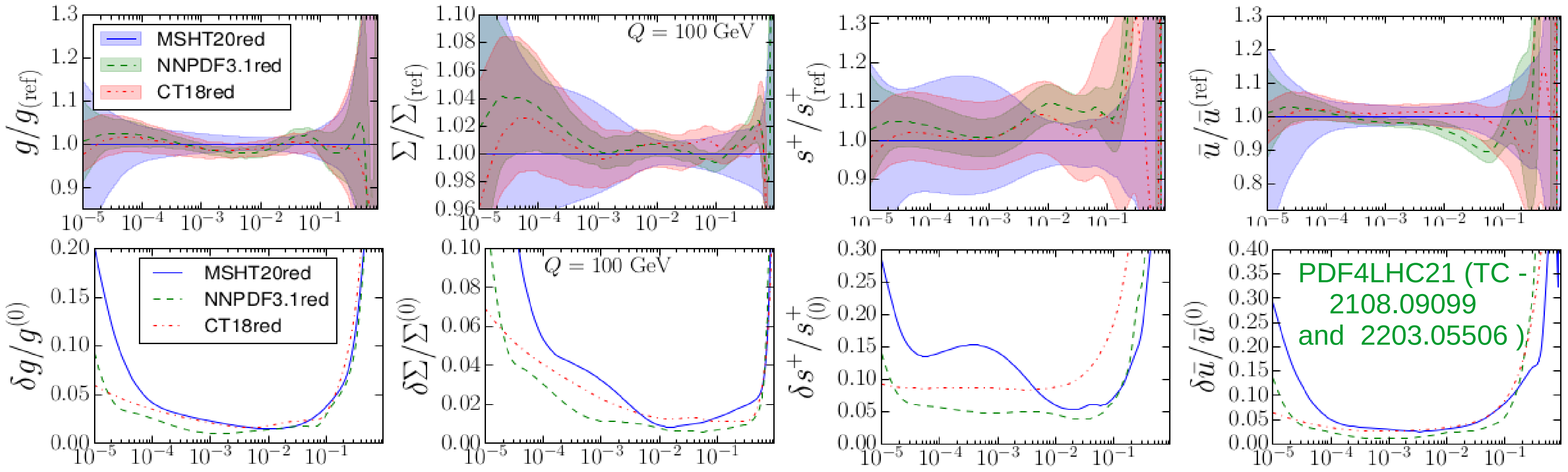


Snowmass (2203.13923)

- 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

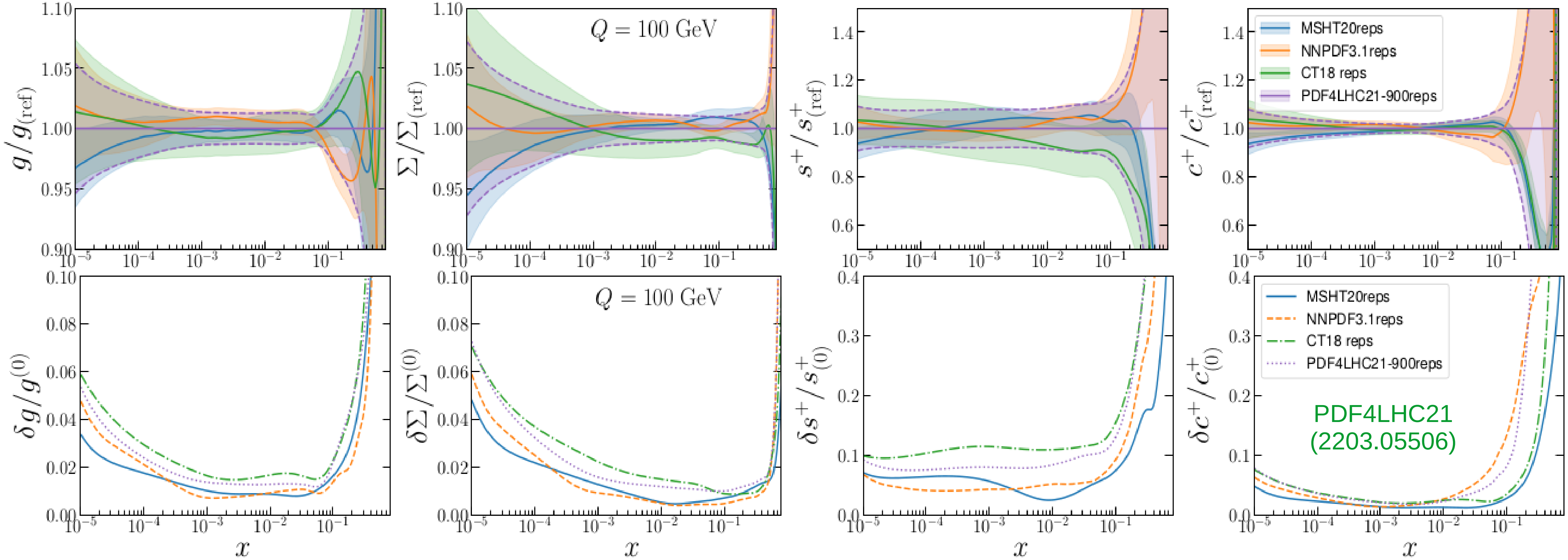
- Observe differences in central values and uncertainties between groups → is it down to data included, methodology applied, theory settings or other differences?
- Take "Reduced Fit" PDFs, using common data and "same" theory (where possible):



- 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

- 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.

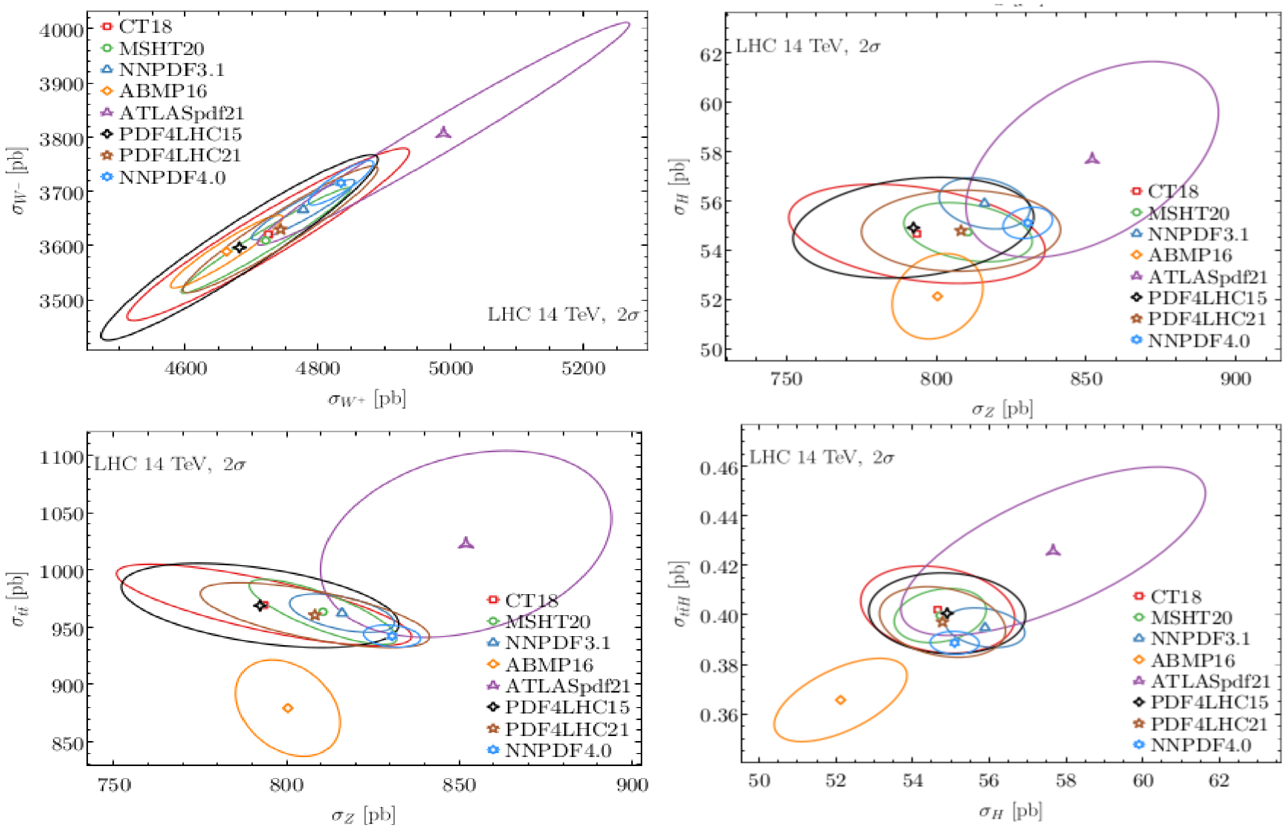
See PDF4LHC21 (2203.05506) for more info!

- 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:



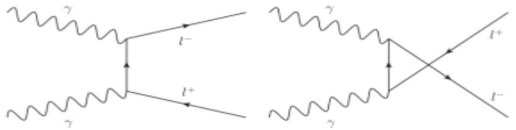
- PDF4LHC21 uncertainty encompasses central values of CT18, MSHT20, NNPDF3.1 here.
- Some PDFs (ABMP, ATLASpdf21) can differ notably, different settings and input data.

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

7. PDF State of the Art

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.

- Idea is to use the NC expression for DIS at low Q^2, then rewrite to obtain the photon PDF from experimentally measured structure functions:

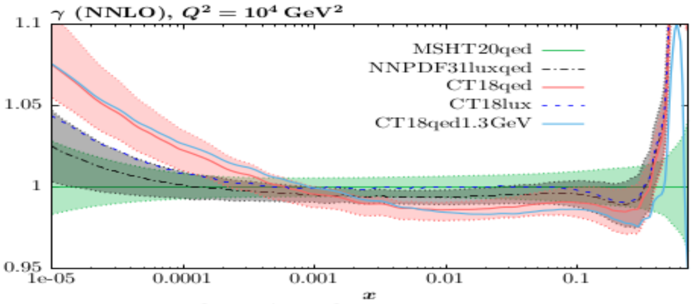
$$x\gamma(x, Q_0^2) = \frac{1}{2\pi\alpha(Q_0^2)} \int_x^1 \frac{dz}{z} \Big\{ \int_{\frac{x^2 m_p^2}{1-z}}^{\frac{Q_0^2}{1-z}} \frac{dQ^2}{Q^2} \alpha^2(Q^2) \Big[\Big(z P_{\gamma,q}(z) + \frac{2x^2 m_p^2}{Q^2} \Big) F_2(x/z, Q^2) - z^2 F_L(x/z, Q^2) \Big] - \alpha^2(Q_0^2) z^2 F_2(x/z, Q_0^2) \Big\},$$

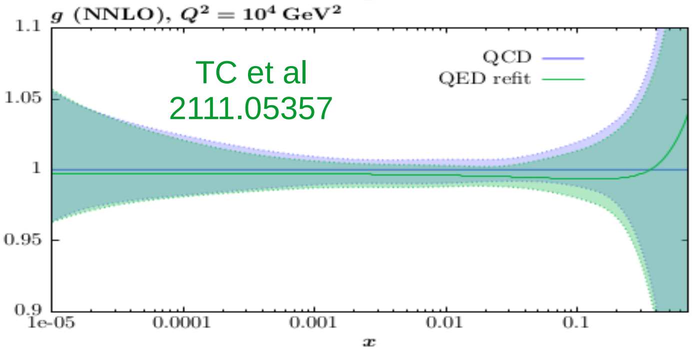
Manohar et al, 1708.01256

Harland-Lang et al 1907.02750

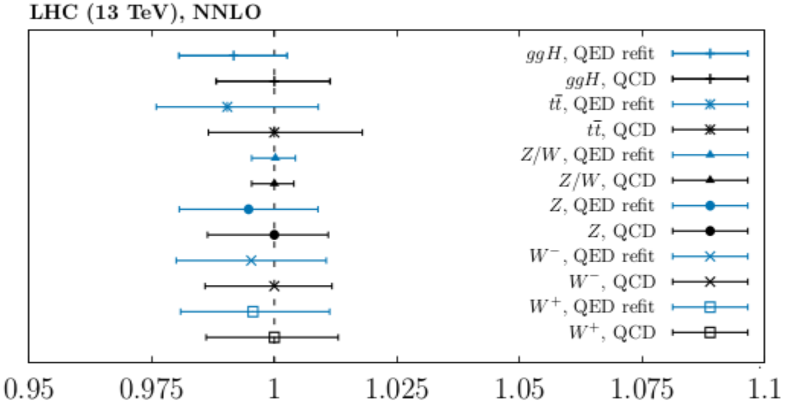
QED PDFs

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





- 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.



aN3LO PDFs - Motivation

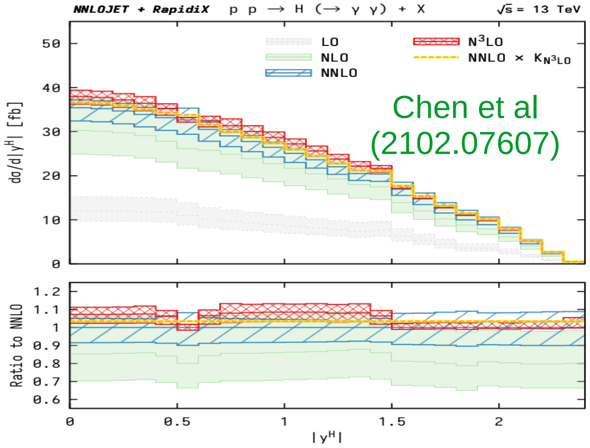
Confronting Precise Data

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

Need for Higher Orders (N3LO):

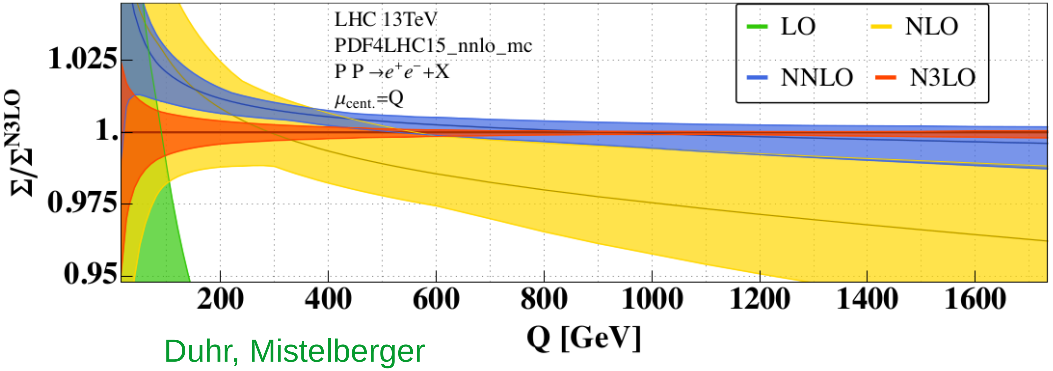
- Progress in recent years on N3LO cross-sections for key processes, e.g. Higgs, DY:

ggF



Chen et al (2102.07607)

NC DY Z/photon



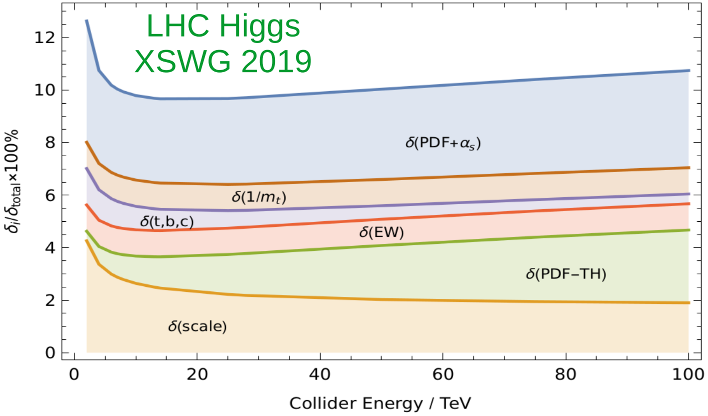
Duhr, Mistelberger (2111.10379)

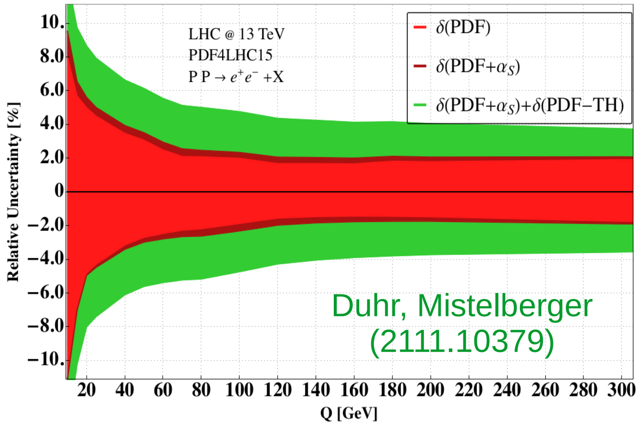
aN3LO PDFs - Motivation

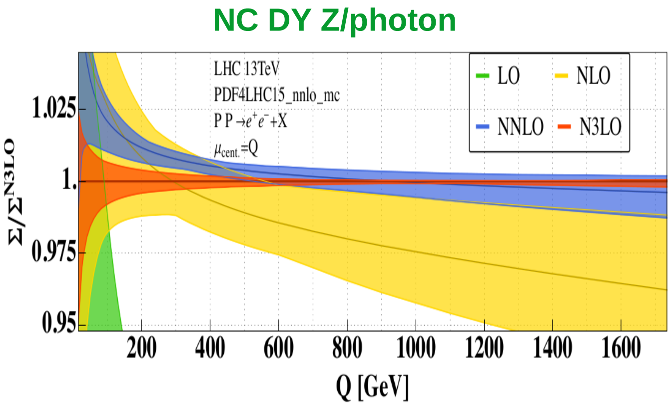
7

Need for Higher Orders

- Only NNLO PDFs have been available – often PDF errors significant + there's mismatch between cross-section and PDF order.





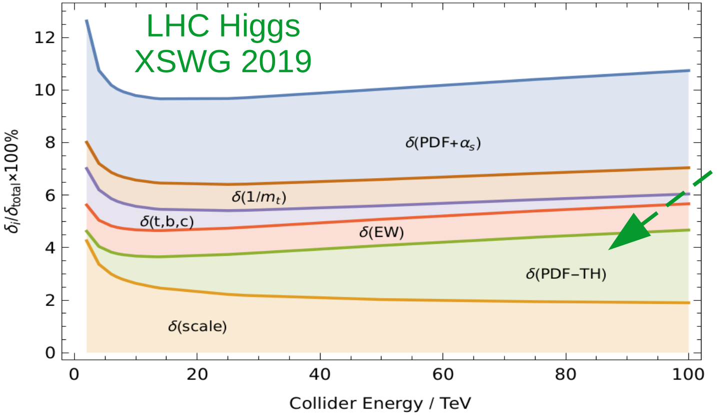


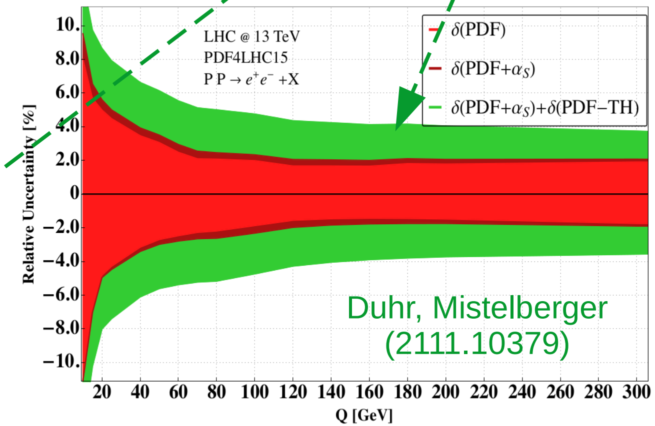
aN3LO PDFs - Motivation

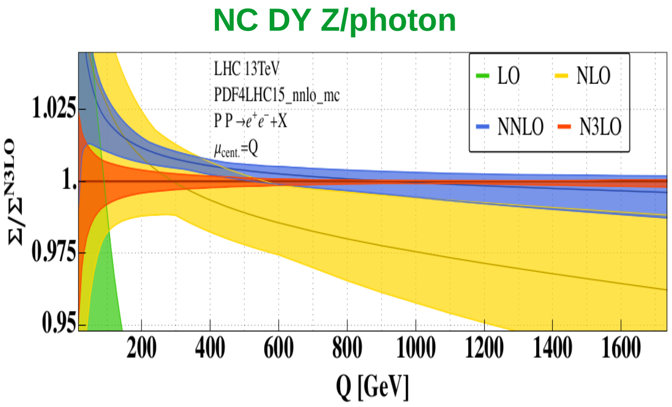
7

<u>Need for Higher Orders</u>

- Only NNLO PDFs have been available – often PDF errors significant + there's mismatch between cross-section and PDF order. Reflected in a **"PDF-th"** uncertainty.



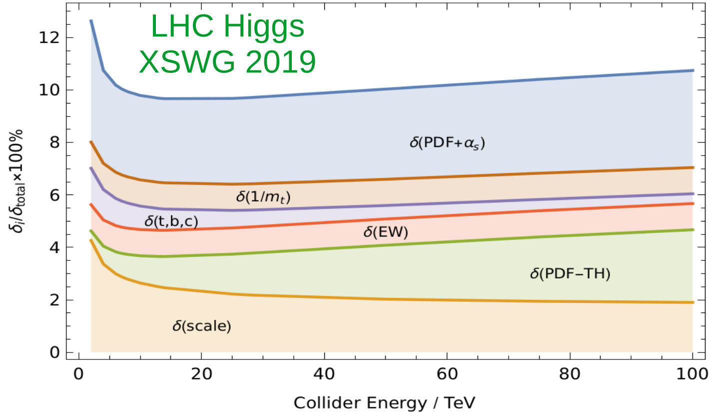


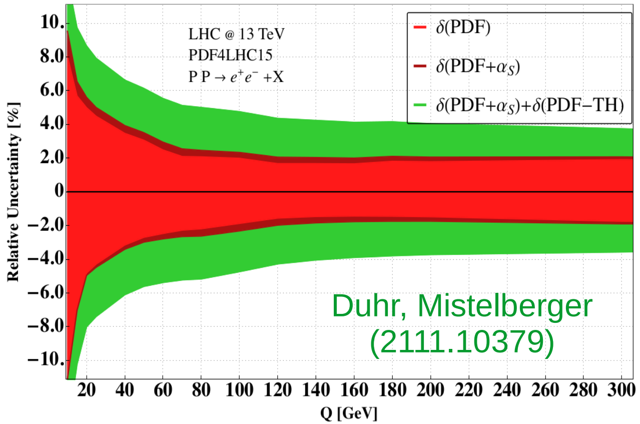


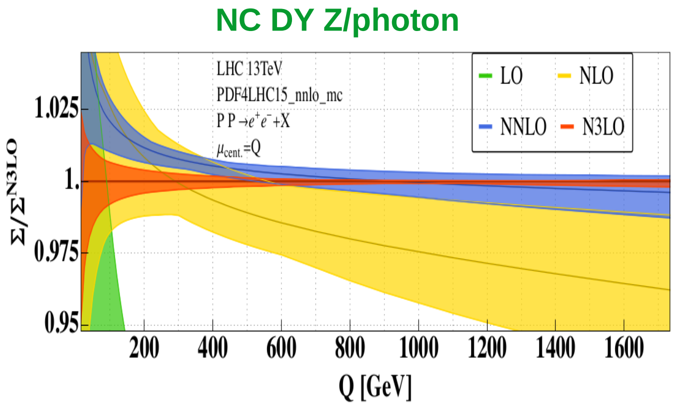
aN3LO PDFs - Motivation

Need for Higher Orders

- Only NNLO PDFs have been available – – often PDF errors significant + there's mismatch between cross-section and PDF order. Reflected in a **"PDF-th"** uncertainty.
- Without consideration of this you cannot estimate the full theoretical uncertainty.
- Only way to remove these bands and properly understand associated uncertainties is determining **N3LO PDFs**, plus inclusion of PDF theoretical uncertainties.





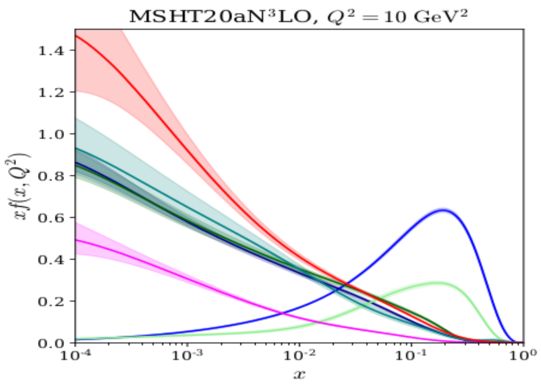


aN3LO PDFs - Idea

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

  ⇒ we can address both in one go!

  ⇒ **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).

aN3LO PDFs – What's required?

Lots of effort by theory community! References in backup

- 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):

Theory	Utility	Order required	What's known?
1. **Splitting functions** $P_{ab}^{(3)}(x)$	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** $A_{ab,H}^{(3)}(x)$	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) $C_{H,a}^{VF,(3)}$	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

- 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

- Include known N3LO pieces + parameterise remaining unknown pieces $\Rightarrow$ theory nuisance parameters (θ') and allow to vary $\rightarrow$ uncertainty.
  - Probes precisely the missing higher order terms. ✓
  - Allows inclusion of known N3LO information (a lot) without needing to wait for remaining few pieces. ✓
  - Avoids scale variations - can underestimate MHOU, issue of correlation between PDF fit and use, no need to raise Q^2 cut on data to enable downwards scale variations. ✓

L.A. Harland-Lang, R.S. Thorne 1811.08434

NNPDF 2207.07616

NNPDF theory uncertainties on NNLO (not aN3LO) via scale variations - 2401.10319

(Theoretical Nuisance Parameters more generally $\rightarrow$ F. Tackmann SCET Workshop 2019)

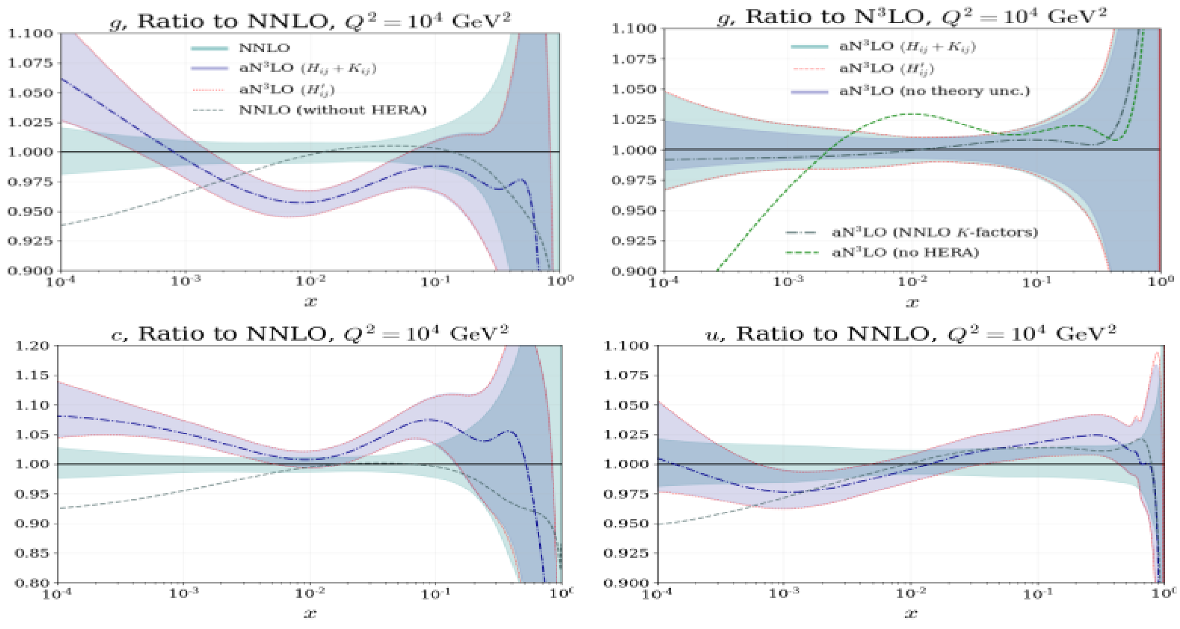
aN3LO PDFs – PDF Impacts

7

- 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.

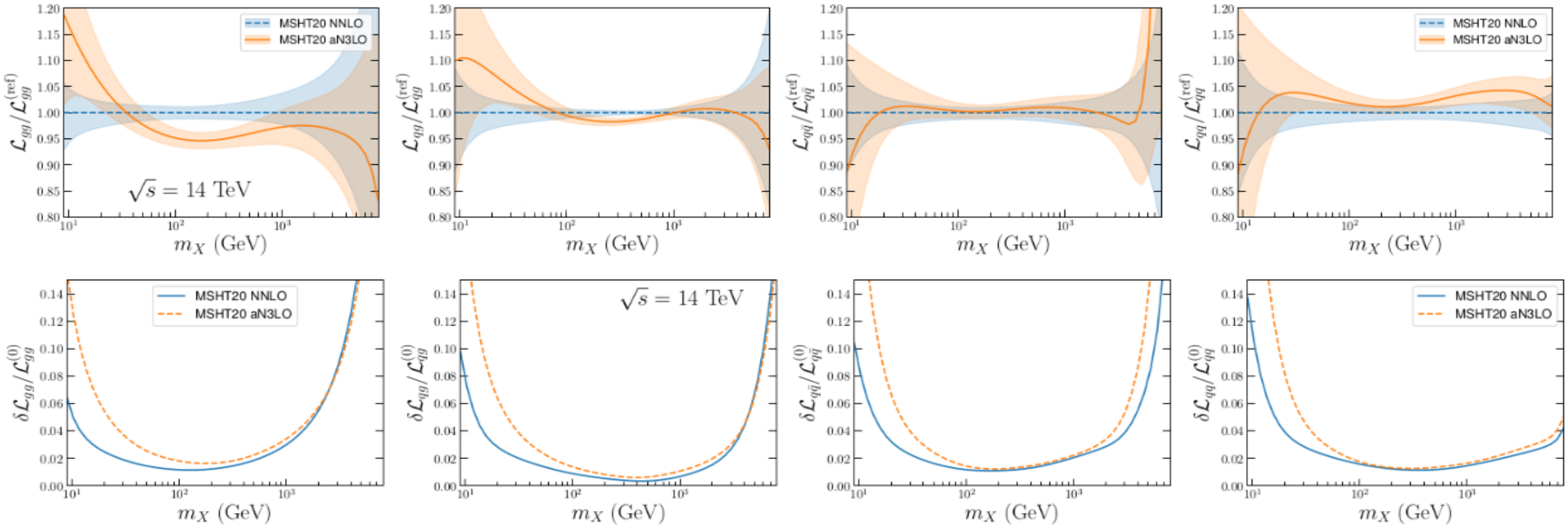


- Gluon PDF raises and its uncertainty increases at low x.
- Heavy quarks also increase across x, purely perturbative.
- Other PDFs much more mildy affected.

J. McGowan et al (2207.04739)

aN3LO PDF Luminosities

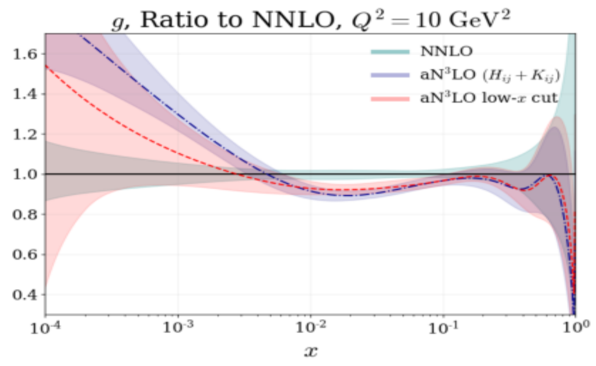
1



- PDF changes have implications for PDF luminosities for phenomenology.
- gg luminosity reduced around 100GeV and increased at 10GeV.
- Luminosity uncertainties enlarged (and more so at lower invariant masses) due to inclusion of aN3LO and PDF theory uncertainties.

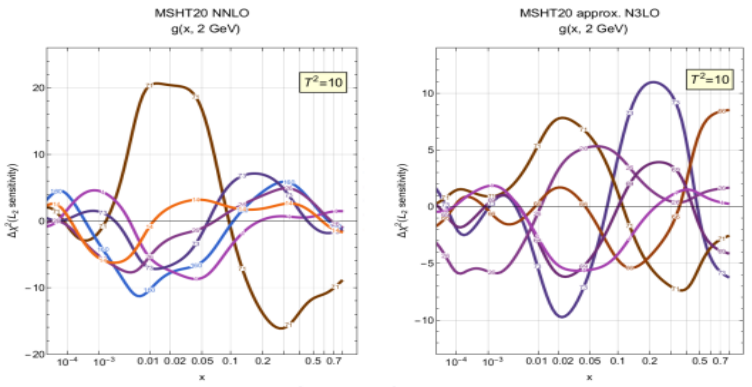
aN3LO PDFs – PDF Impacts

- Reduced tensions between some datasets seen at aN3LO.
- Small x - high x data tension reduced.
- Precise ATLAS 8 TeV Zp_T data fit quality at NNLO is poor, but at aN3LO is good:



Order	NNLO	aN3LO
ATLAS 8 TeV Zp_T	1.87	1.04
Total	1.22	1.17

Table: Fit qualities χ^2/N_{pts}.

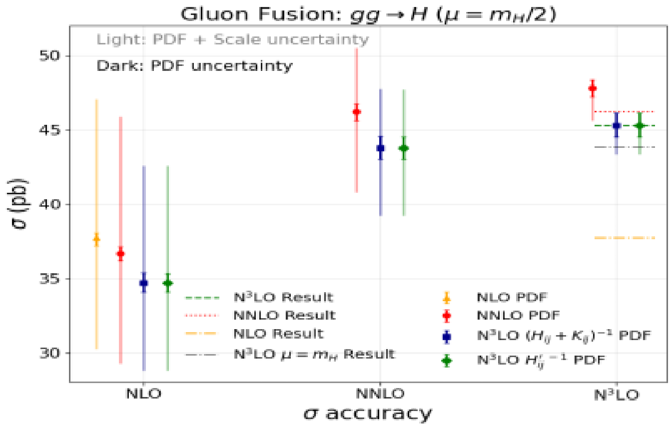


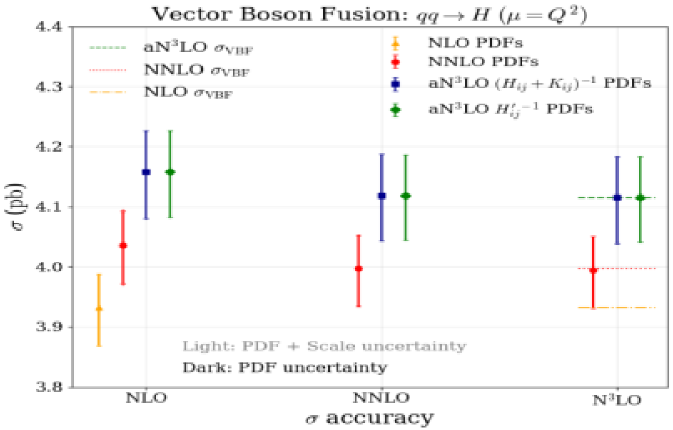
L2 study 2306.03918

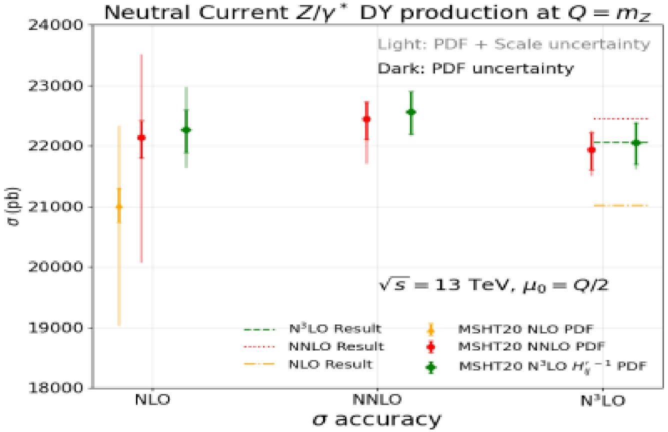
- Tensions between ATLAS 8TeV Zp_T and other data reduced at aN3LO.
- High precision data requires high precision theory.

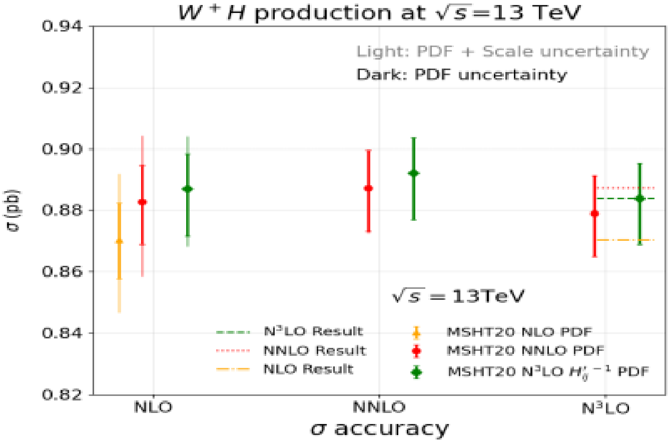
aN3LO PDFs – Xsec Impacts

7









- How does aN3LO affect xsecs?
- ggF Higgs – increase from N3LO xsec balanced by reduction from g PDF at aN3LO.
- Increase in VBF xsec due to increase in heavy quarks.
- Only small change in DY or VH as more quark dominated.

J. McGowan et al (2207.04739)

aN3LO QCD + QED PDFs (bonus!)

7

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

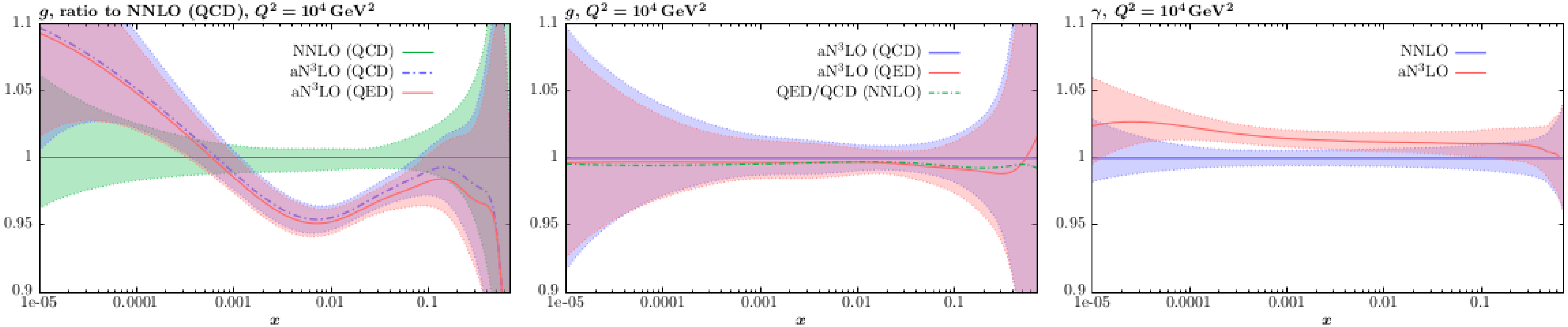
QED $$P_{ij} = \frac{\alpha}{2\pi} P_{ij}^{(0,1)} + \frac{\alpha\alpha_S}{(2\pi)^2} P_{ij}^{(1,1)} + \left(\frac{\alpha}{2\pi}\right)^2 P_{ij}^{(0,2)}$$

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

TC et al (2312.07665)

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

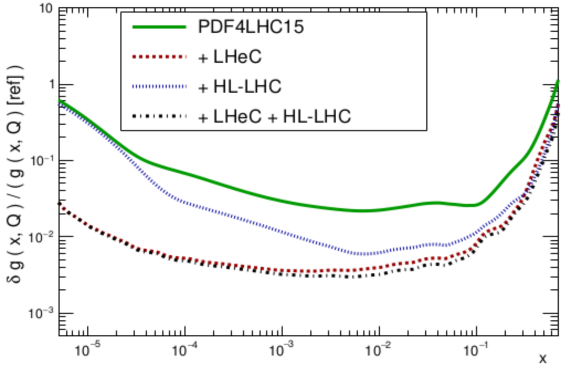


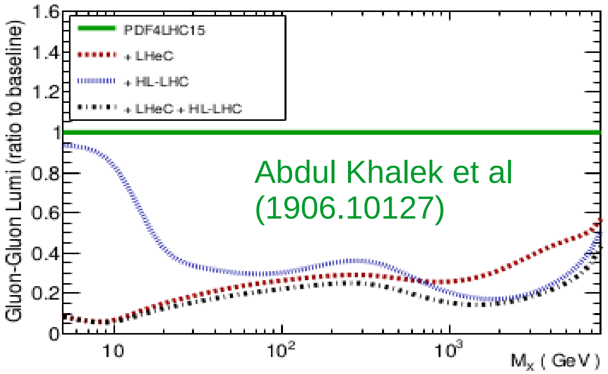
PDF Constraints – Future LHC

4

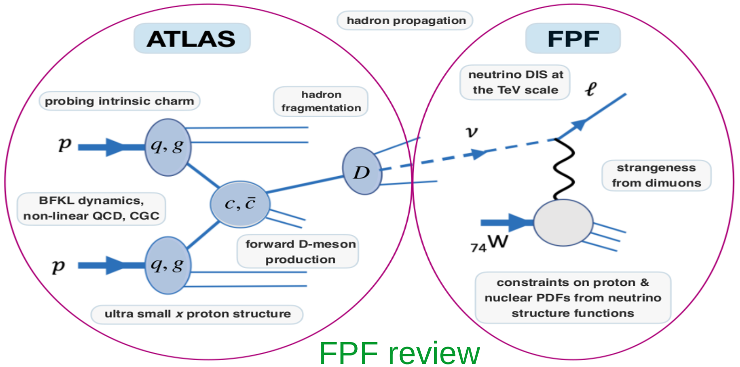
- **HL-LHC** – Reduce PDF uncertainties where processes currently statistically limited/coverage can be extended, e.g. high x gluon and gg luminosity.
- **LheC** – Inclusive/Semi-inclusive DIS data constrain intermediate/small x.

LHeC review (2007.14491)





Abdul Khalek et al (1906.10127)



FPF review (2203.05090)

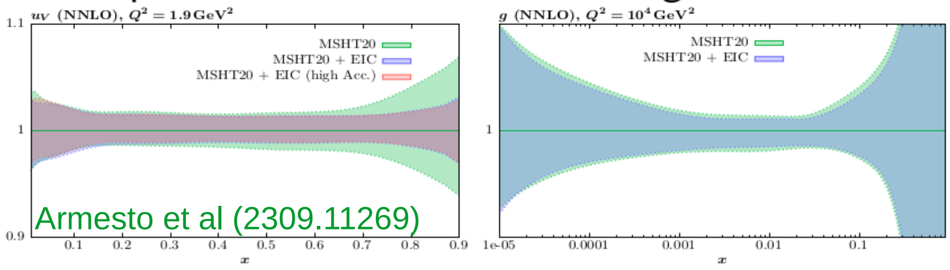
- **FPF** – Very forward neutrino production → intrinsic charm at high x, very low x gluon dynamics. Then Neutrino CC DIS → flavour separation.

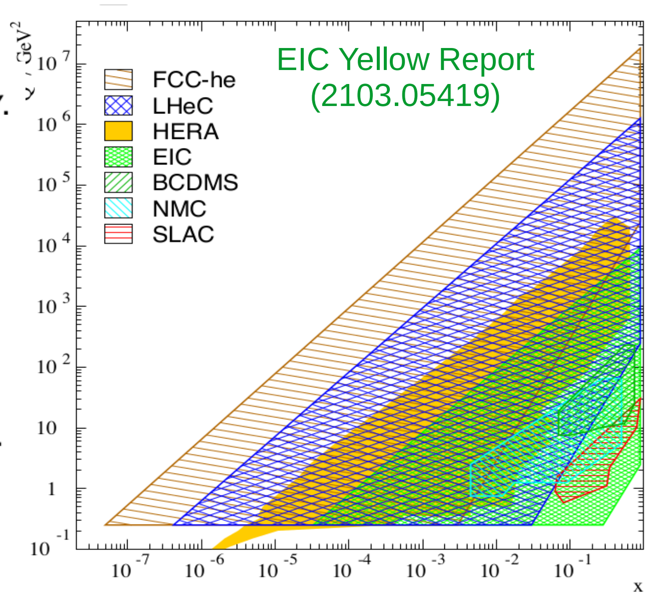
Complementarity between all future experiments and knock on effects for physics goals!

PDF Constraints – Future EIC

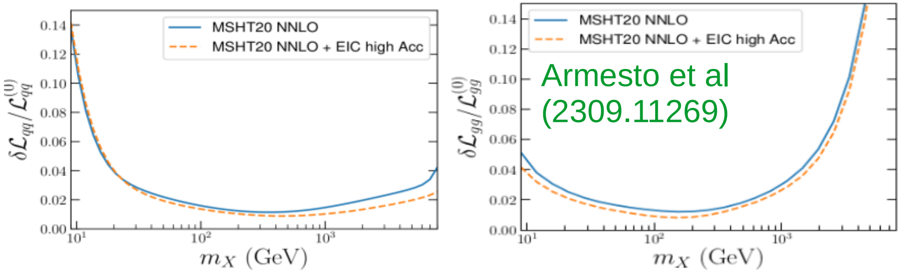
4

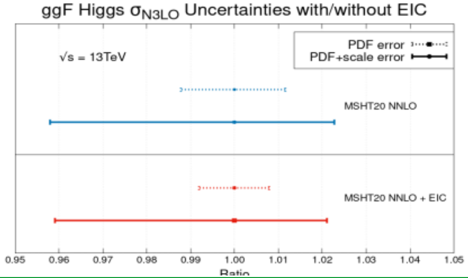
- Electron Ion Collider to run at Brookhaven in 2030s.
- Impact of EIC pseudodata on MSHT20 - *high x lower Q^2 sensitivity.*

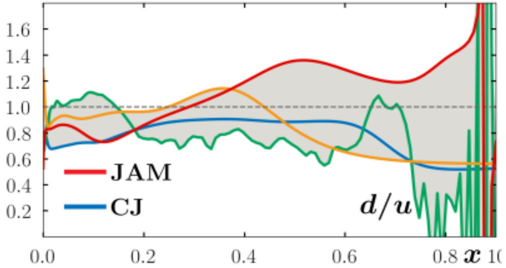




- Effect on up valence larger due to charge-squared γ coupling in DIS.
- Gluon uncertainty nonetheless reduced across range of x.
- Impact on luminosity uncertainties - $\delta\mathcal{L}_{qq}$ reduced at high x.
- $\delta\mathcal{L}_{gg}$ reduced across x causes smaller PDF uncertainty for $gg \to H$.







EIC Yellow Report (2103.05419)

8. Strong Coupling Determination (if time!)

Strong Coupling sensitivity

- 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.

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

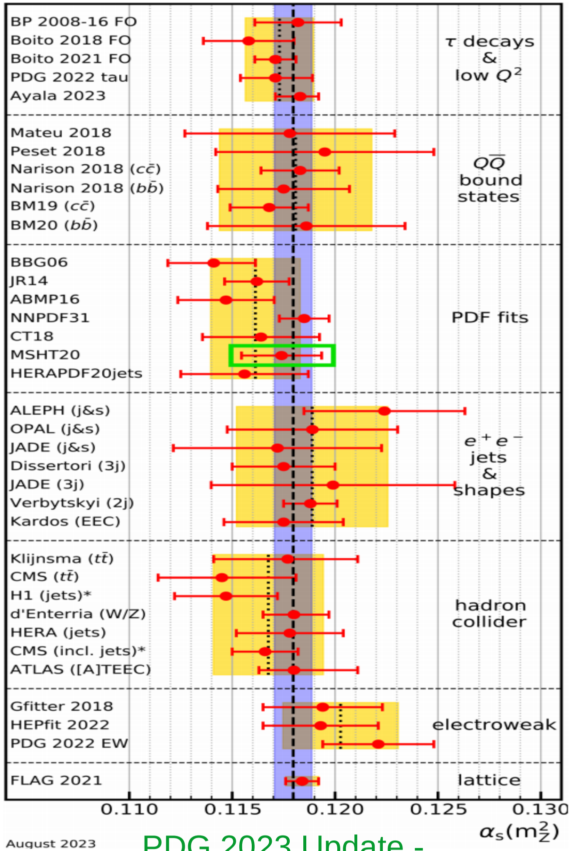
$$\mathbf{C}(\alpha_S) = \alpha_S^i[\mathbf{C}_0 + \alpha_S \mathbf{C}_1 + \alpha_S^2 \mathbf{C}_2 + \alpha_S^3 \mathbf{C}_3 + \ldots] \quad \chi^2$$

  - 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 χ^2 → best fit $\alpha_S(M_Z^2)$



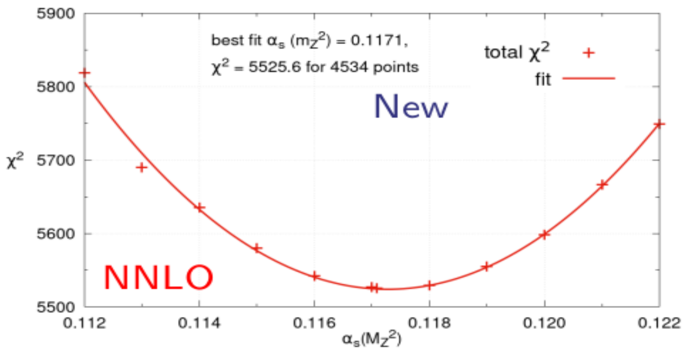
PDG 2023 Update - 2312.14015.

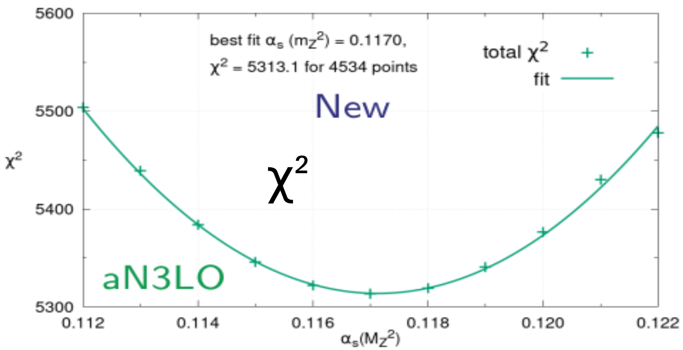
Strong Coupling sensitivity

- Good perturbative convergence of α_S determination.

More in backup!

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





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 – upcoming!

9. Conclusions

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

Summary

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

Summary

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

Thankyou! Any Questions?

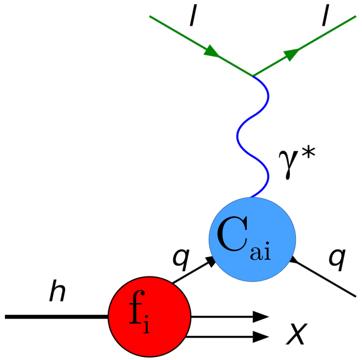
Backup

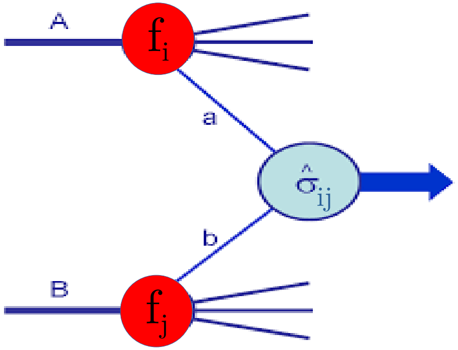
TC is funded from a project of the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No. 101002090 COLORFREE).

Introduction

- Consider electron scattering off a hadron (proton):
- Collider physics relies on QCD Collinear factorisation:

DIS: Hadron-Hadron (pp):





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

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

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

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

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

PDF Constraints (Fixed Target DIS) 4

- Integral over x gives total u, d momentum (f_u, f_d) in proton: $F_2^{en} = F_2^{ep}(u \leftrightarrow d)$

$$\int_0^1 F_2^{ep}(x)dx = \int_0^1 x\left(\frac{4}{9}(u(x)+\bar{u}(x)) + \frac{1}{9}(d(x)+\bar{d}(x))\right) = \frac{4}{9}f_u + \frac{1}{9}f_d$$

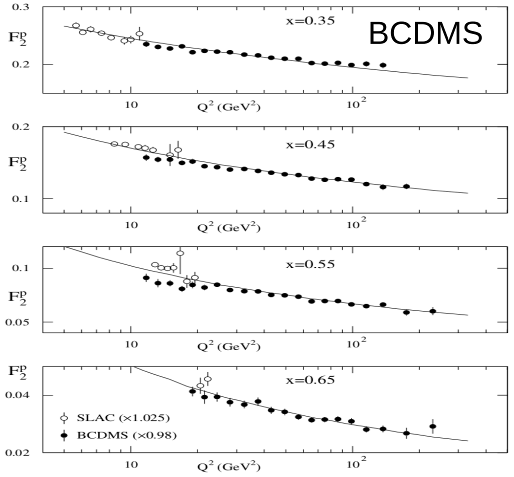
Momentum contribution of other light quarks/ antiquarks is very small

- Measure and solve → $f_u \approx$ **0.36**, $f_d \approx$ **0.36**, rest in gluon!
- Can also measure as function of Q^2 - Scaling violations?!
- In QCD improved parton model we have also quark splittings (reduces high x partons with Q^2):

$$\frac{d\tilde{f}^{NS}(x,Q^2)}{d\ln Q^2} = P^{NS}(x,\alpha_s(Q^2)) \otimes \tilde{f}^{NS}(x,Q^2)$$

- Therefore expected and also seen at HERA (later).
- Also means evolution of high x structure functions is sensitive to strong coupling – see later!

At higher Q^2, sensitivity instead to Z couplings – different linear combination of PDFs!



MRST (hep-ph/9803445)

Heavy quarks in PDFs (HERA)

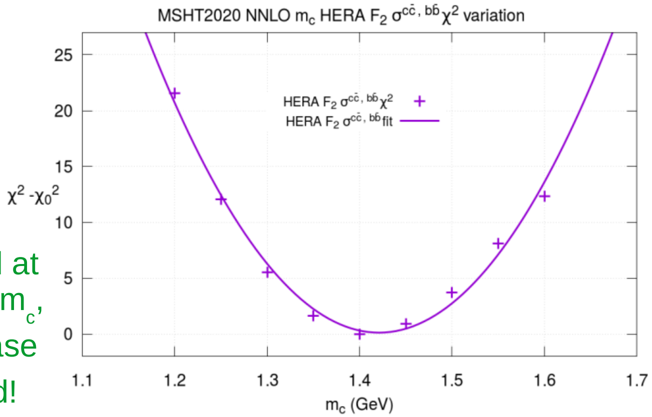
4

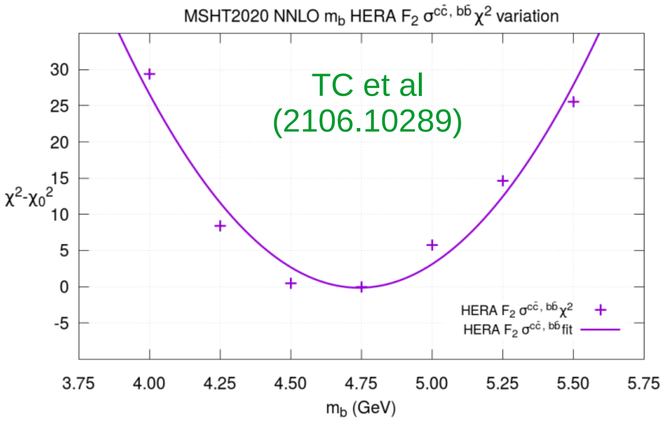
- PDFs then have **3** flavours up to m_c, **4** between there and m_b, and **5** flavours above that – as expected.
- PDFs "switched on" at m_c, m_b by tranisition matrix elements.

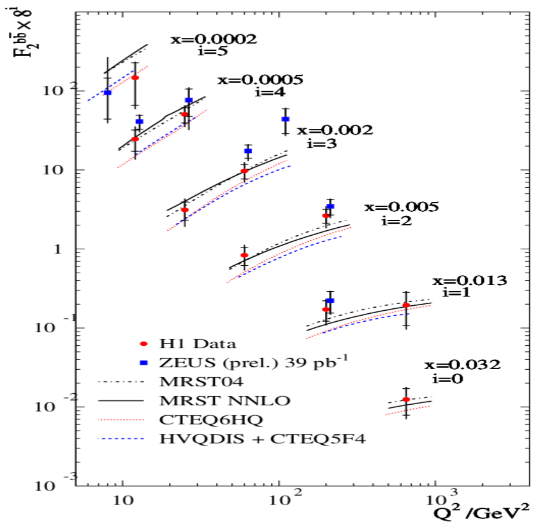
$$f_\alpha^{n_f+1}(x,Q^2) = \left[A_{\alpha i}(Q^2/m_h^2) \otimes f_i^{n_f}(Q^2)\right](x)$$

- Good fit to HERA heavy flavour structure function data with GMVFNS (right).
- Also induces sensitivity to m_c, m_b into PDF fit.

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):

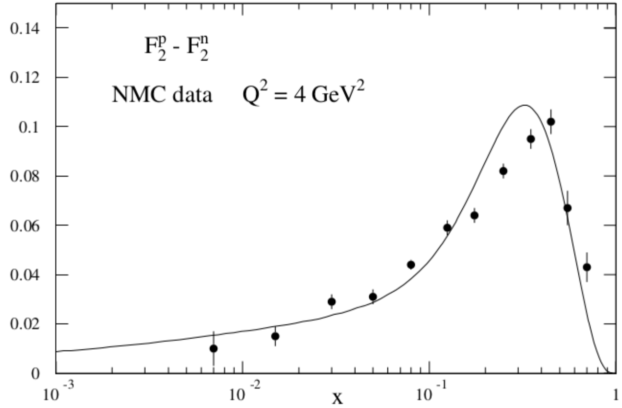
$$S_G = \int_0^1 [F_2^p(x) - F_2^n(x)]dx/x = \frac{1}{3}\int_0^1 [u(x) + \bar{u}(x) - d(x) - \bar{d}(x)]$$
$$= \frac{1}{3}\int_0^1 [u_V(x) - d_V(x) + 2(\bar{u}(x) - \bar{d}(x)] = \frac{1}{3} + \frac{2}{3}\int_0^1 [\bar{u}(x) - \bar{d}(x)]$$

- 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!

Assumes measure whole x range → not true!

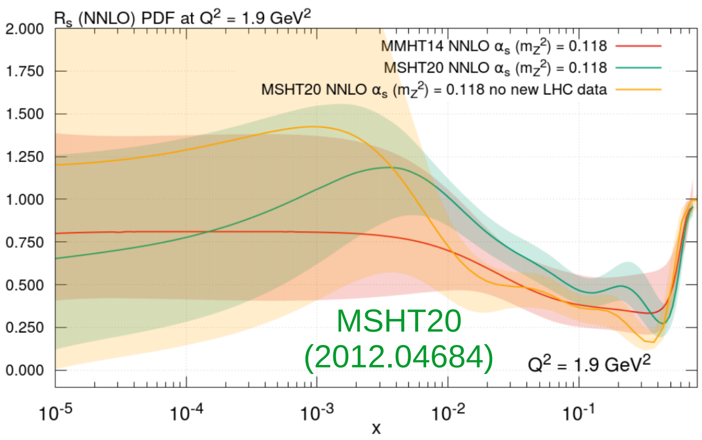


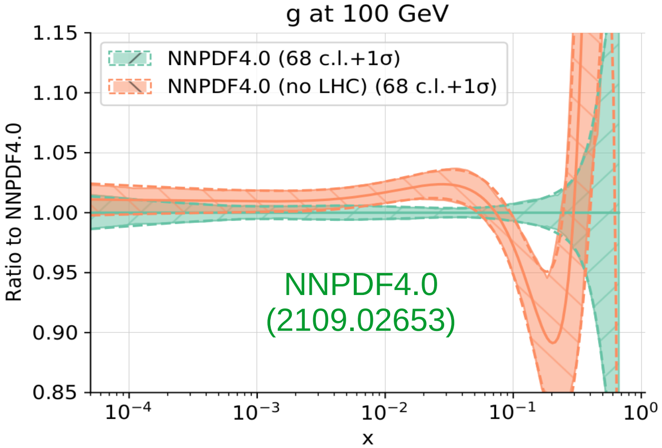
- What does Fixed Target DY at E866 tell us then?

Experimental Advances

Confronting Precise Data

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

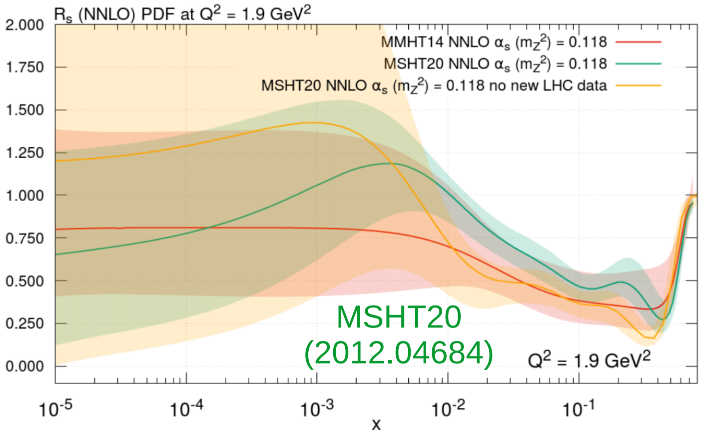


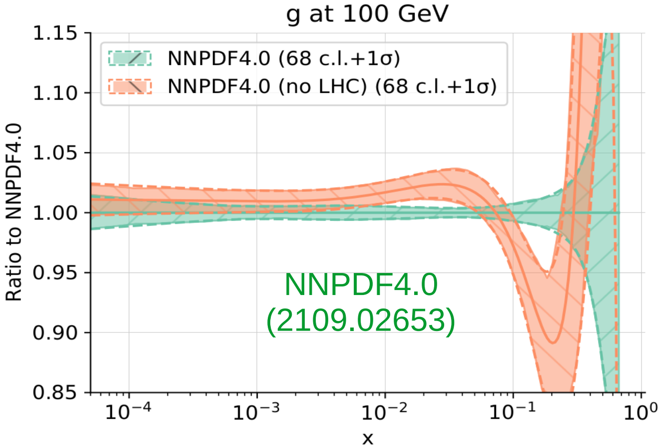


Experimental Advances

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

MSHT20 (2012.04684)

3. Challenges and Developments

3

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

Experimental Challenges

4

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

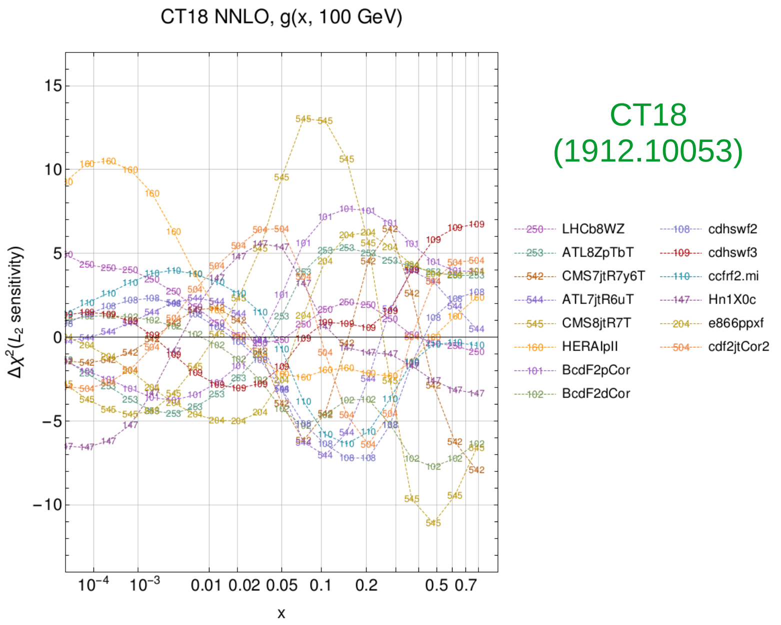
- Issues can arise in fitting some datasets - **poor fit qualities** χ^2/N_{pts}.
- Two frequent (experimental/methodological) causes:

1) Dataset tensions

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

→ Antiquark isospin asymmetry

→ High x gluon (jets, Zp_T, top)



CT18 (1912.10053)

Experimental Challenges

4

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

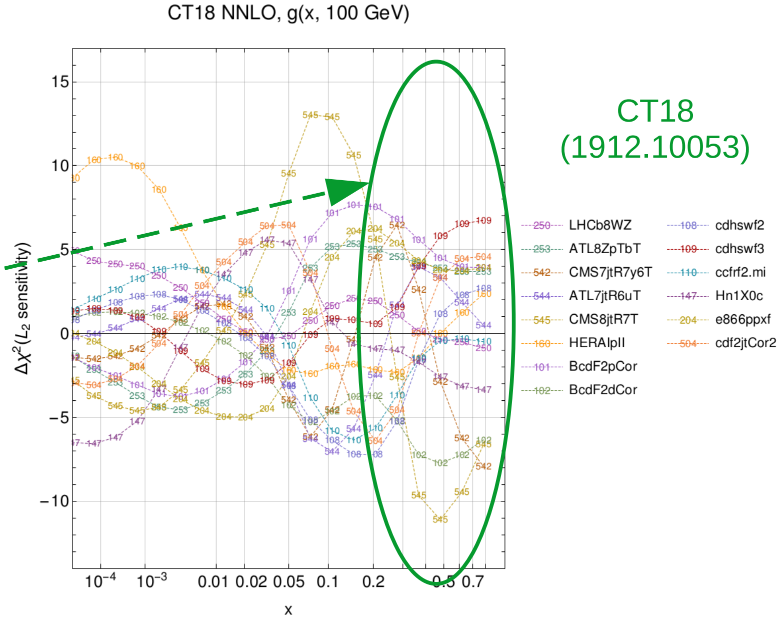
- Issues can arise in fitting some datasets - **poor fit qualities** χ^2/N_{pts}.

- 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)



CT18 (1912.10053)

Experimental Challenges

4

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

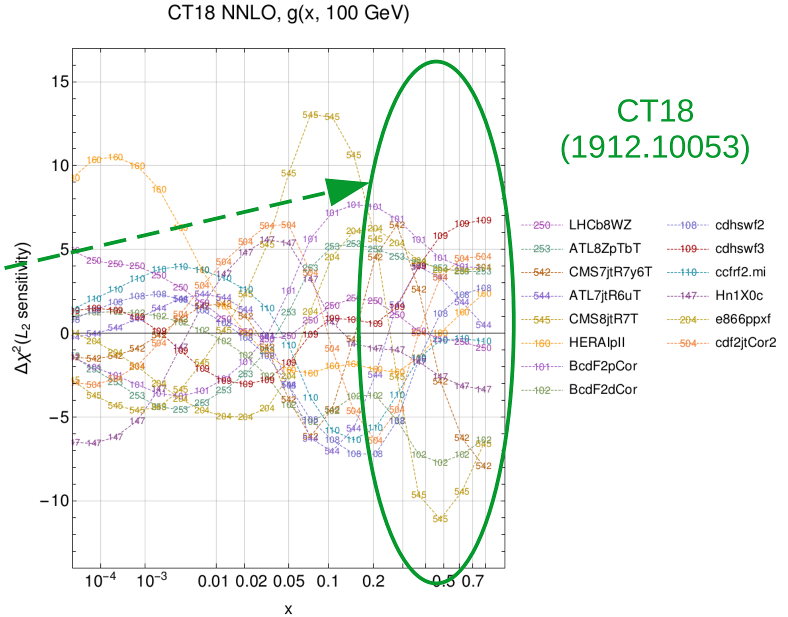
- Issues can arise in fitting some datasets - **poor fit qualities** χ^2/N_{pts}.
- Two frequent (experimental/methodological) causes:

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

→ Antiquark isospin asymmetry

→ High x gluon (jets, Zp_T, top)

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



CT18 (1912.10053)

Experimental Challenges

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

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

- Two frequent (experimental/methodological) causes:

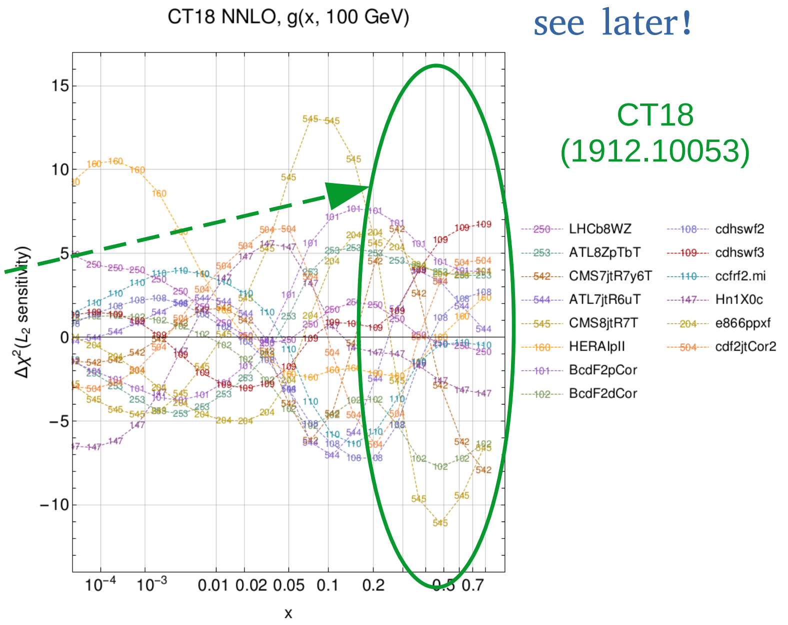
1) Dataset tensions
– Different datasets have conflicting pulls on the PDFs. Examples include

→ Antiquark isospin asymmetry

→ High x gluon (jets, Zp_T, top)

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

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



CT18 (1912.10053)

Experimental Challenges

4

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

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

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

Theoretical Review in Peng et al (1402.1236)

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

Experimental Challenges

4

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

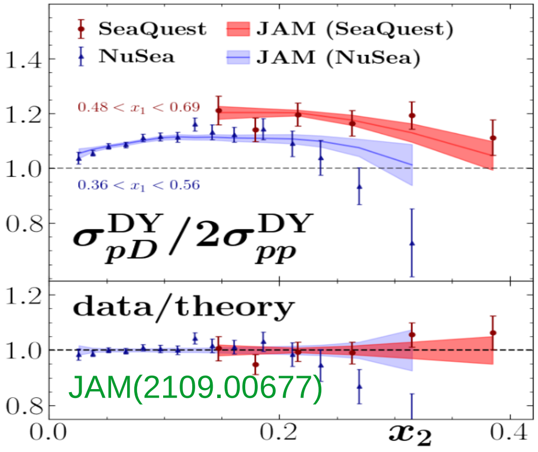
- High x $\bar{d}/\bar{u}$: - Theoretical models (e.g. pion cloud) generally favour $\bar{d}>\bar{u}$ at high x.

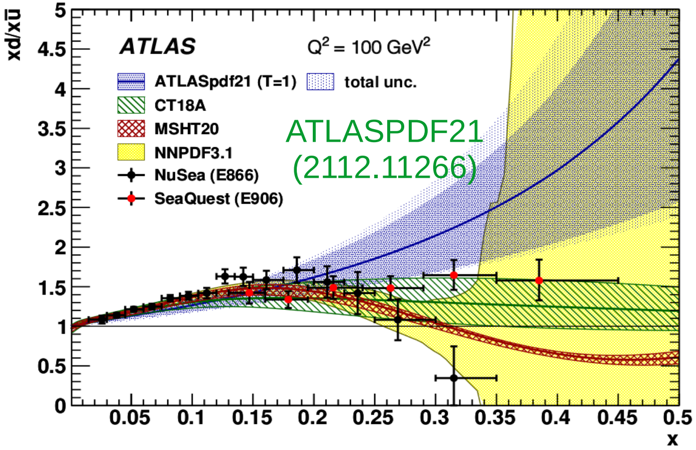
NuSea (hep-ex/0103030)
Seaquest (2103.04024)

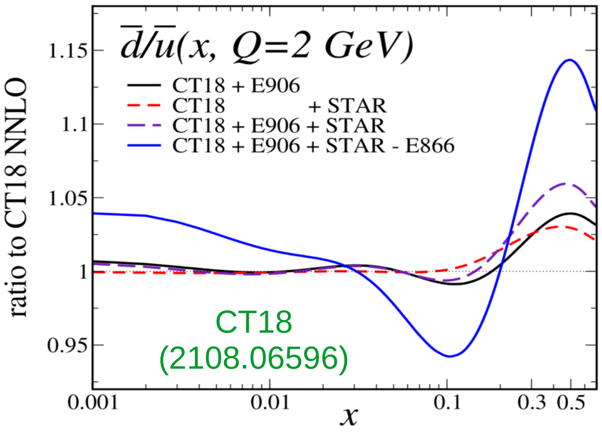
- Gottfried sum rule – NMC found $\int_0^1 [F_2^p(x) - F_2^n(x)]dx/x < \frac{1}{3}$

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

- New Seaquest/E906 data instead favour $\bar{d}>\bar{u}$.







Experimental Challenges

4

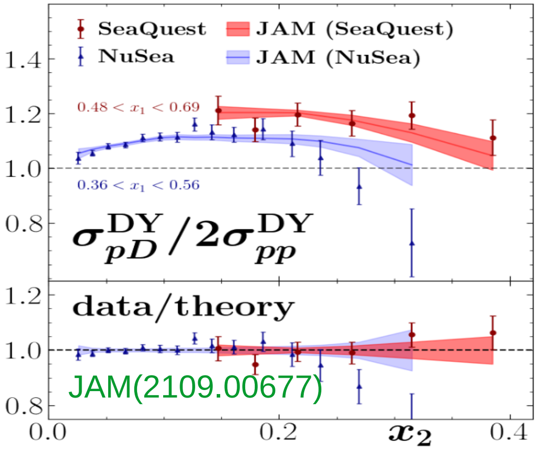
Dataset Tensions – *Can reflect experimental, methodological or theoretical issues!*

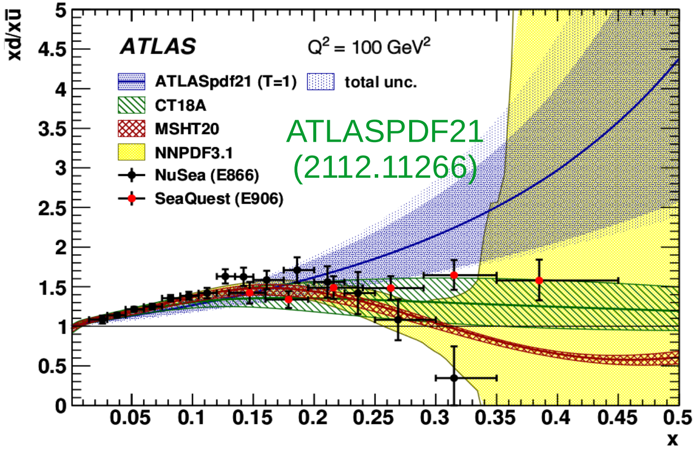
- High x $\bar{d}/\bar{u}$: - Theoretical models (e.g. pion cloud) generally favour $\bar{d}>\bar{u}$ at high x.

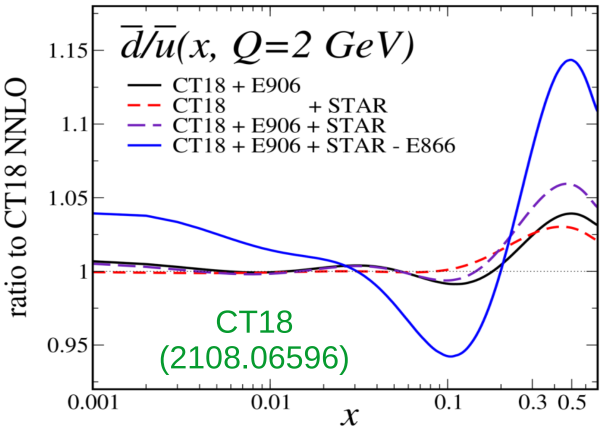
NuSea (hep-ex/0103030)
Seaquest (2103.04024)

- Gottfried sum rule – NMC found $\int_0^1 [F_2^p(x) - F_2^n(x)]dx/x < \frac{1}{3}$
- E866/NuSea data favoured $\bar{d}<\bar{u}$ at high x.
- New Seaquest/E905 data instead favour $\bar{d}>\bar{u}$.

How should this be interpreted in a global PDF fit?







Experimental Challenges

Confronting Precise Data

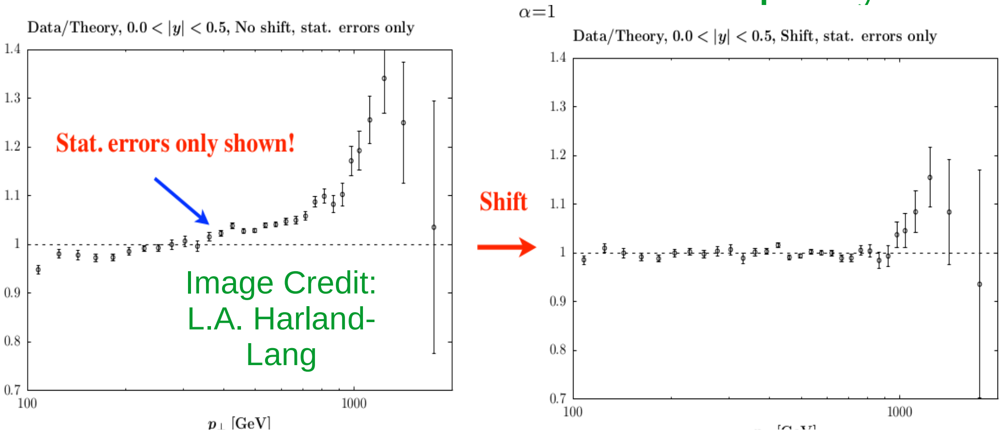
2) **Correlated Systematics** – Issues occur for 2 of the 3 dataset high x gluon data types – top and jets. Consider ATLAS 7 TeV jets:

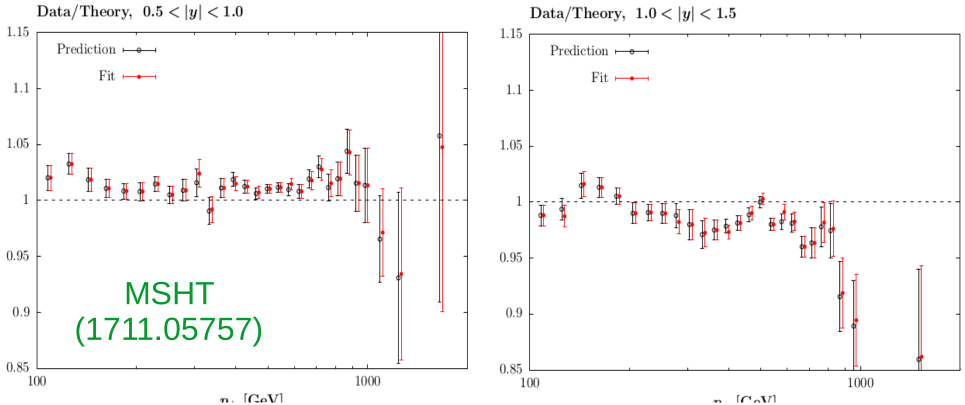
ATLAS 7TeV jets (1410.8857)

ATLAS 8TeV jets (1706.03192)

Can fit individual rapidity bins

But not multiple rapidity bins!





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

Experimental Challenges

4

Confronting Precise Data

[ATLAS study - 1706.03192]

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

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

Experimental Challenges

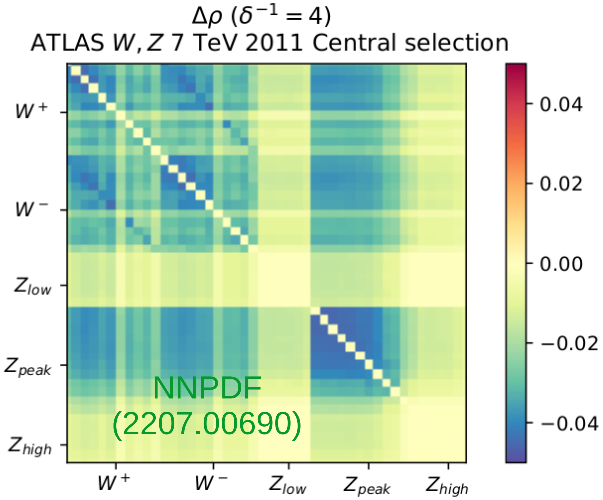
4

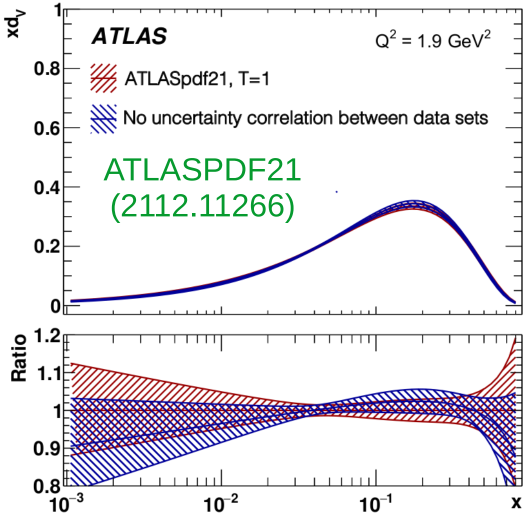
Confronting Precise Data

[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.



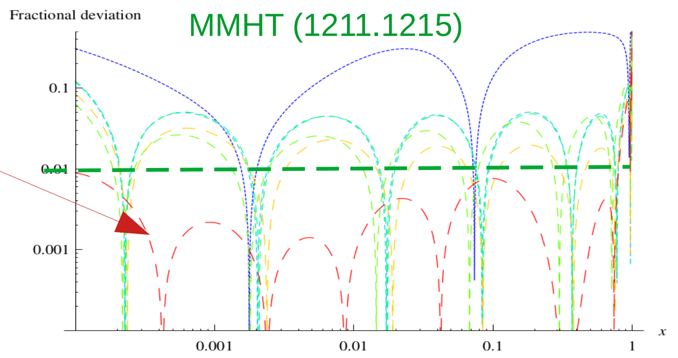


Methodological Challenges

5

Confronting Precise Data

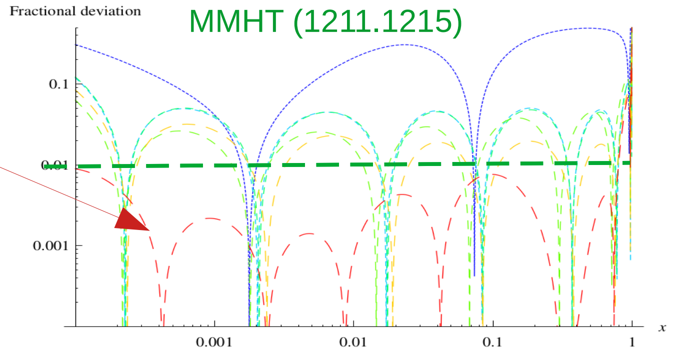
- PDF fitting groups must continually evolve fitting methodology.
- Extended parameterisations, investigate different forms:
  - MSHT20 → 51 parton parameters to fit to $< 1\%$ if data allows. Gives Net $\Delta\chi^2_{\text{global}} = -73$.

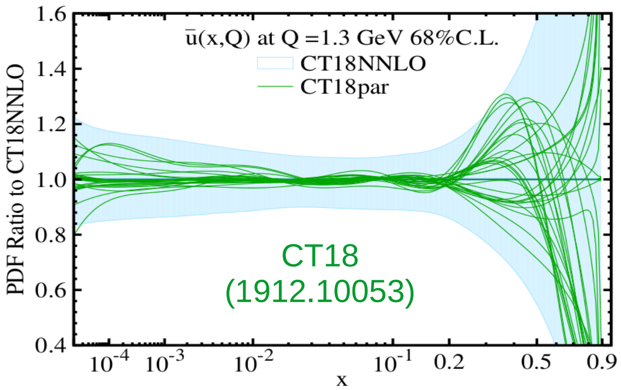


Methodological Challenges

Confronting Precise Data

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



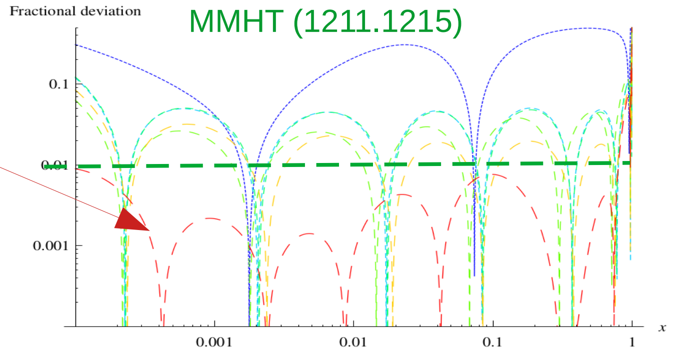


Methodological Challenges

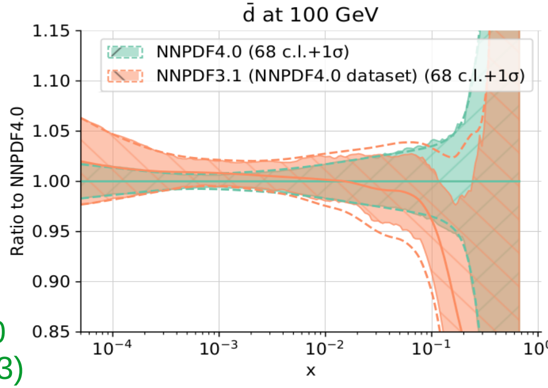
⑤

Confronting Precise Data

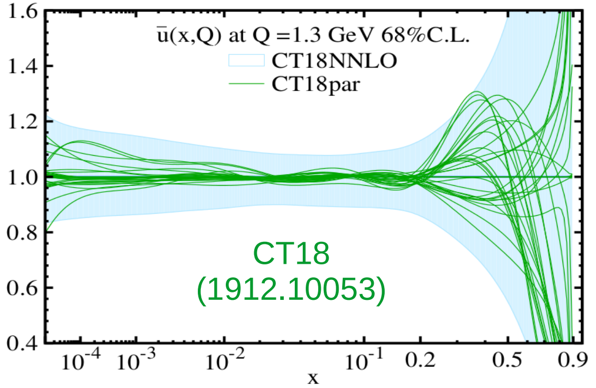
- PDF fitting groups must continually evolve fitting methodology.
- Extended parameterisations, investigate different forms:
  - MSHT20 → **51** parton parameters to fit to $<$ **1%** if data allows. Gives Net $\Delta\chi^2_{\rm global} = -73$.
  - CT18 → Investigation of different functional forms.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



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



NNPDF4.0 (2109.02653)



Methodological Challenges

New Codes and Tools

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

1) Lagrange Multiplier (LM) scans

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

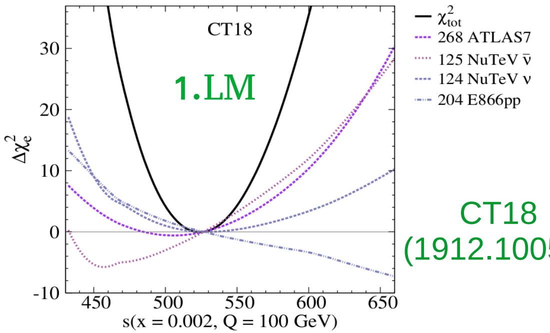
3) Weighted fits

4) Effective Gaussian Variables (S_E)

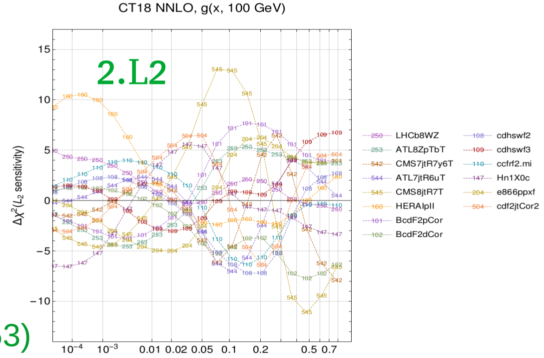
$$S_E = \sqrt{2\chi^2_E} - \sqrt{2N_E - 1}$$

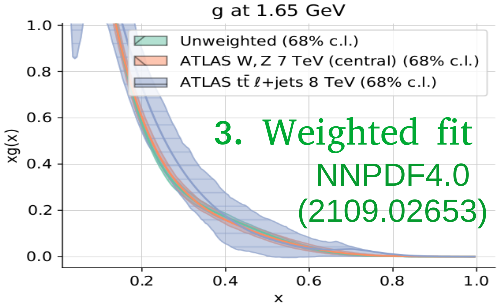
"Spartyness"

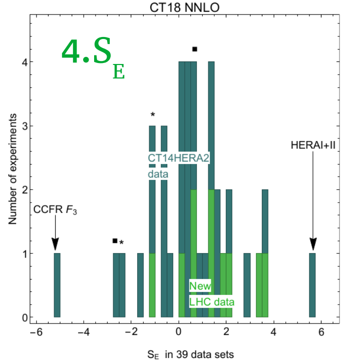
5) Many more….



CT18 (1912.10053)





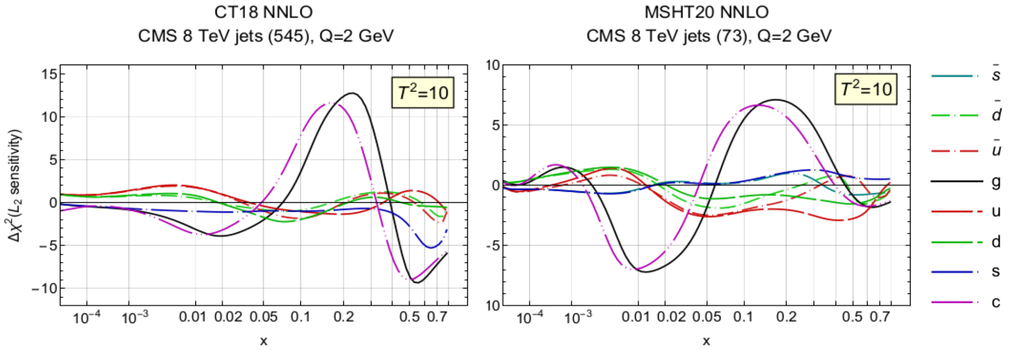


Methodological Challenges

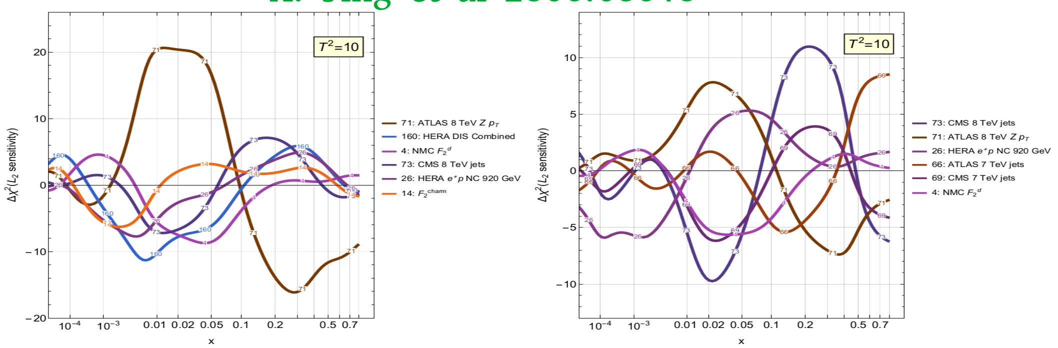
<u>Understanding Data Pulls in Different PDF sets/group:</u>

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

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



X. Jing et al 2306.03918



Methodological Challenges

5

<u>Uncertainties</u>

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

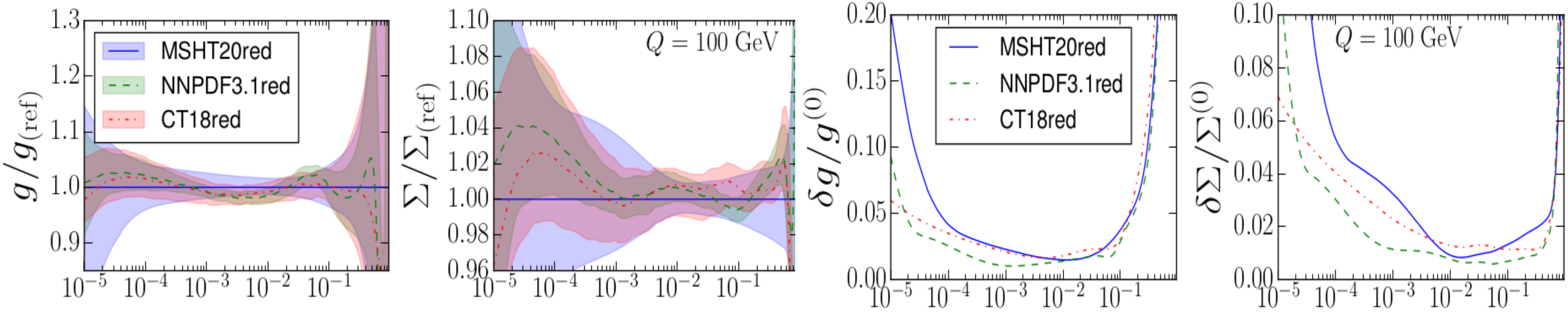
Methodological Challenges

5

Uncertainties

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

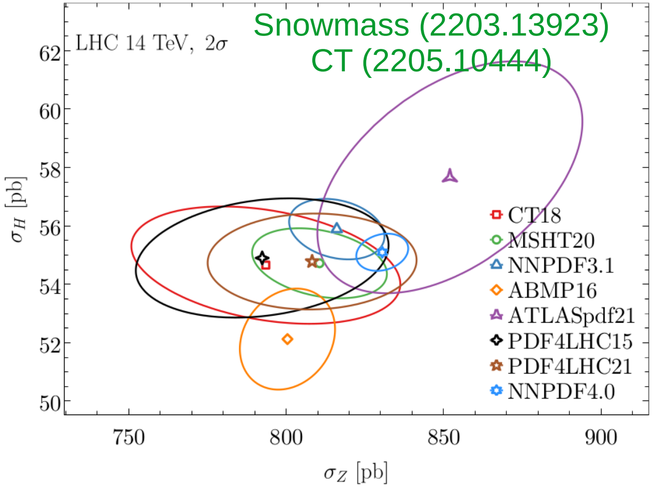
PDF4LHC21 (2203.05506)



Methodological Challenges

Uncertainties

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



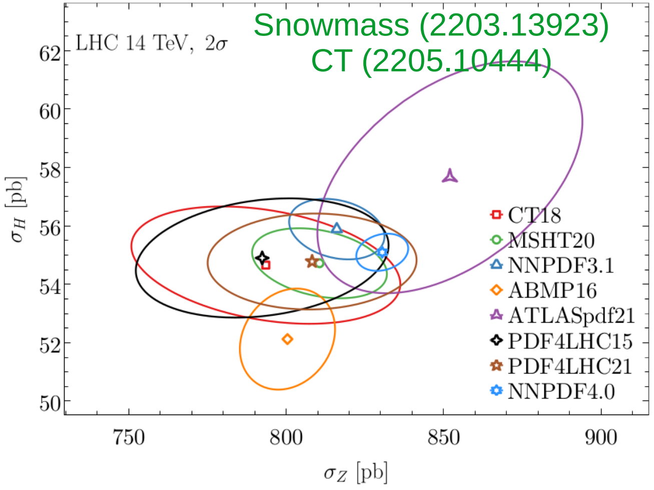
Methodological Challenges

Uncertainties

- Different groups see varying sizes of PDF uncertainties. Other explanations?
- Data sampling – Sampling large multidimensional parameter spaces is difficult.
- Various ways to test this
  - Closure/future testing – use artificial/restricted data to test for bias and uncertainty sizes.
  - Parameter space scan, look for additional solutions and compare with uncertainties (e.g. "hopscotch").

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

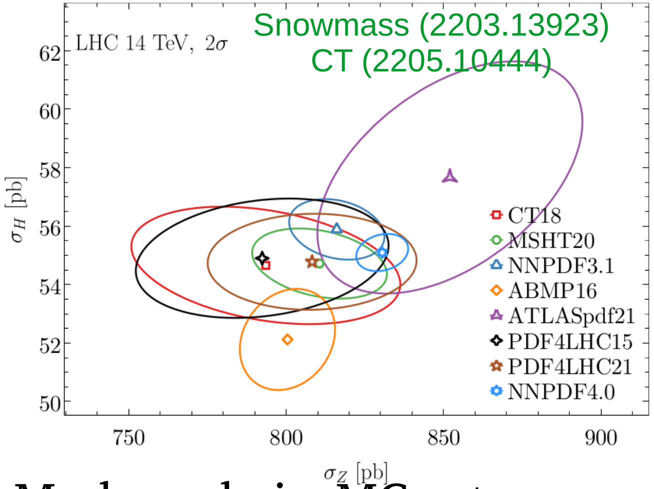
Snowmass (2203.13923)
CT (2205.10444)



Methodological Challenges

<u>Uncertainties</u>

- Different groups see varying sizes of PDF uncertainties. Other explanations?
- Data sampling – Sampling large multidimensional parameter spaces is difficult.
- Various ways to test this
  - Closure/future testing – use artificial/restricted data to test for bias and uncertainty sizes.
    MSTW(1205.4024)
    NNPDF (2103.08606)
    NNPDF4.0 (2109.02653)
  - Parameter space scan, look for additional solutions and compare with uncertainties (e.g. "hopscotch").



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

Hunt-Smith et al (2206.10782)

Methodological Challenges

5

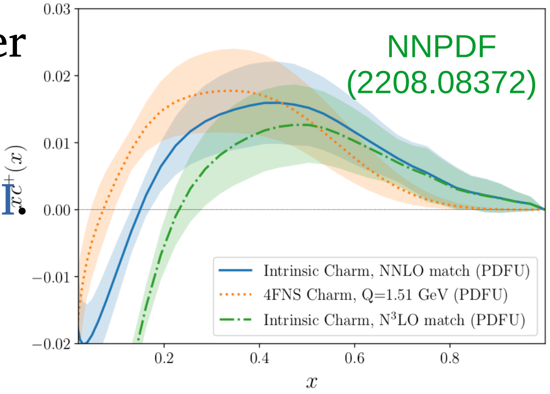
Is it there and can we see it?

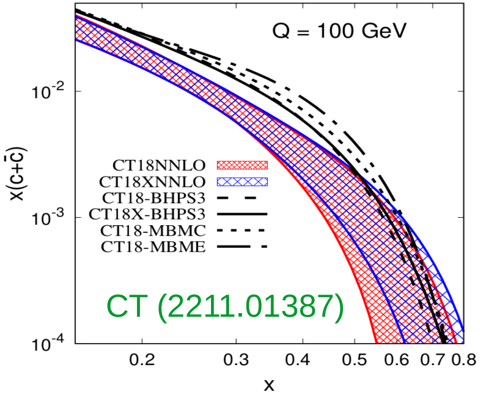
Intrinsic Charm (IC)

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



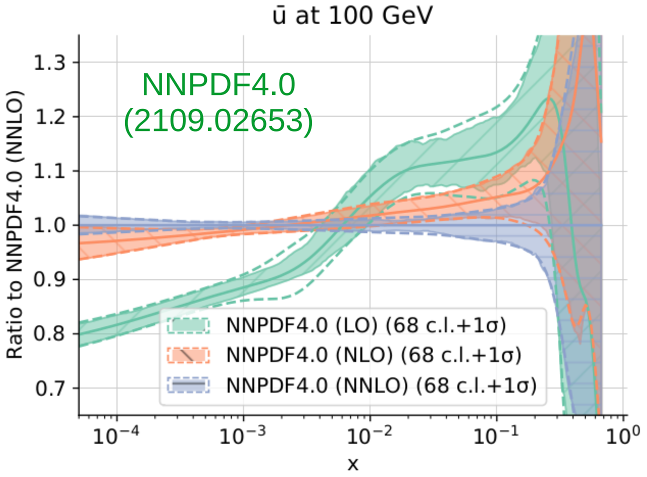


Theoretical Challenges

Theoretical Uncertainties - MHOUs:

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

Missing Higher Order Uncertainties (MHOUs).



Theoretical Challenges

Theoretical Uncertainties - MHOUs:

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

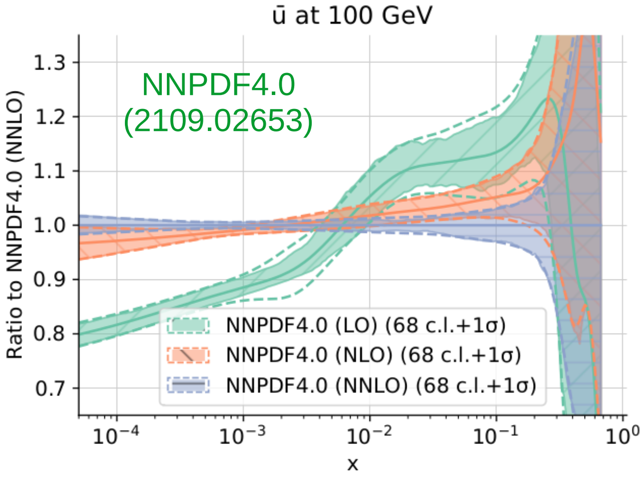
  Missing Higher Order Uncertainties (MHOUs).

- Three main approaches:

  1) Scale variation/joint fits

  2) Bayesian approaches

  3) Theoretical Nuisance Parameters



Theoretical Challenges

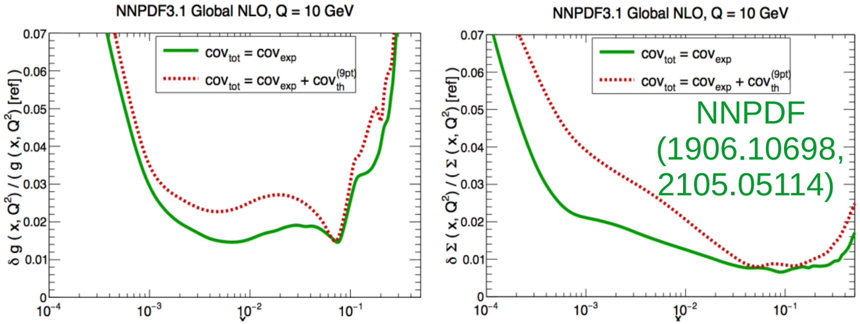
6

Theoretical Uncertainties - MHOUs:

1) Scale variation - Include scale variations as proxy.

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

MMHT (1811.08434)



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

Theoretical Challenges

6

Theoretical Uncertainties - MHOUs:

1) Scale variation - Include scale variations as proxy.

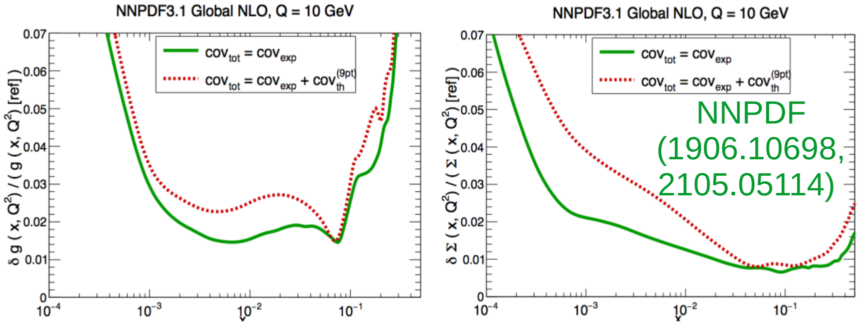
- NNPDF have done NLO, using "theory covariance matrix", S.
- Get small improvements in χ^2/N and larger uncertainties.
- Potential issue of double counting scale variations in PDFs and cross-sections.
- Degree of variation used is arbitrary, only probes (N)NLO terms.

MMHT (1811.08434)

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

- Bonvini and Cacciari Houdeau models.

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



NNPDF (1906.10698, 2105.05114)

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

Theoretical Challenges

Theoretical Uncertainties - MHOUs:

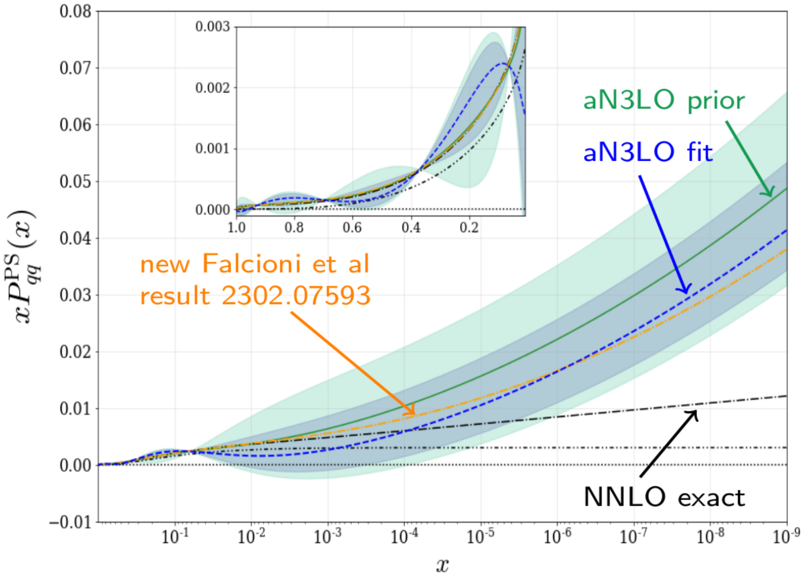
MSHT (2207.04739)

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.**

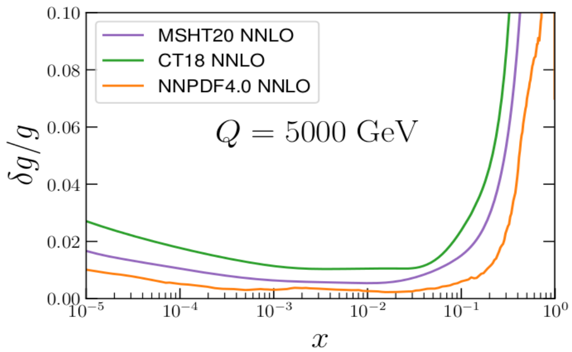


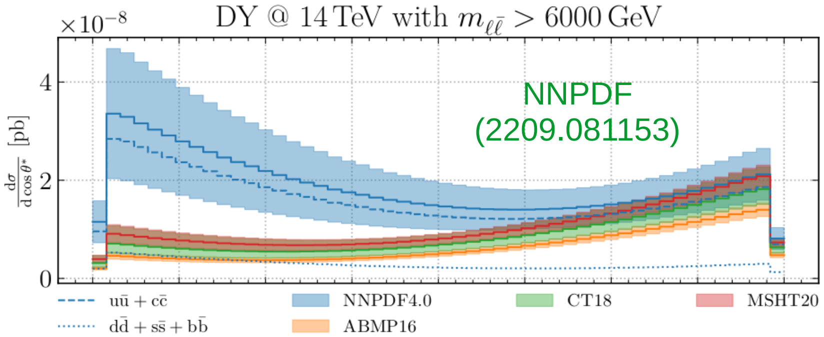
Theoretical Challenges

6

PDFs and Beyond Standard Model Physics

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

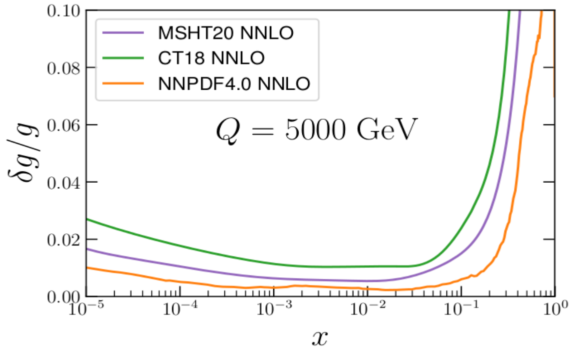


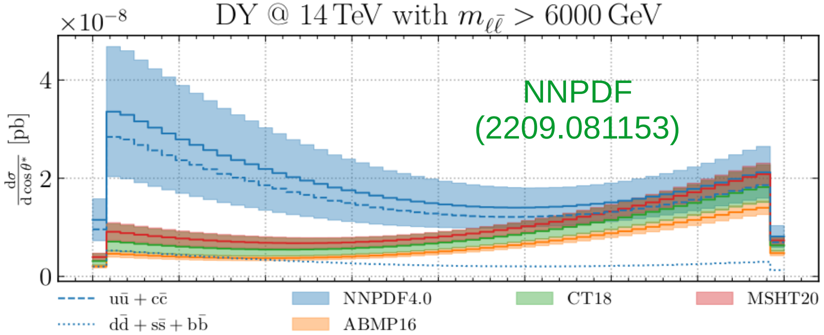


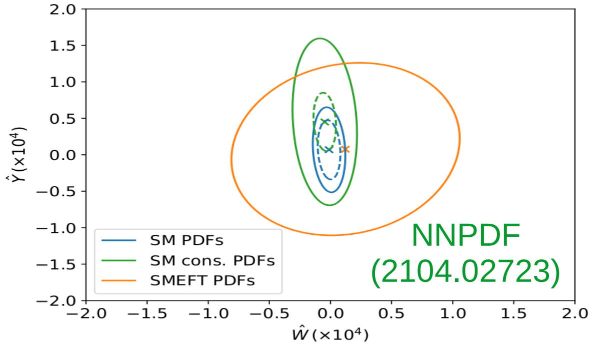
Theoretical Challenges

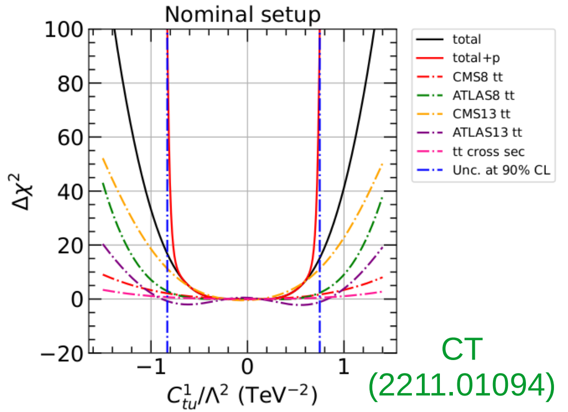
PDFs and Beyond Standard Model Physics

- 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.









Methodological Challenges

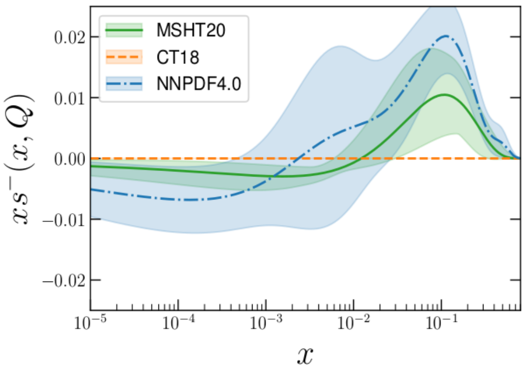
More info on positivity or otherwise of MS PDFs:
NNPDF (2006.07377)
Collins et al (2111.01170)

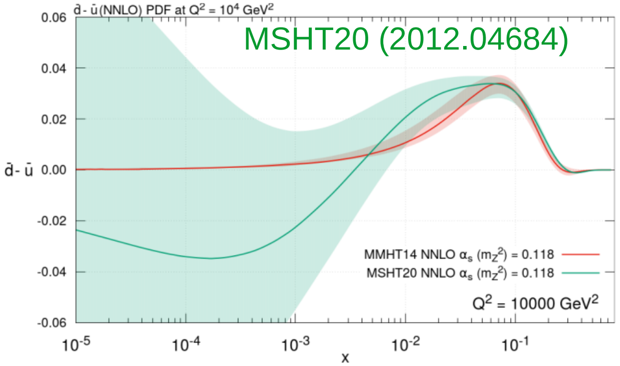
Additional Constraints

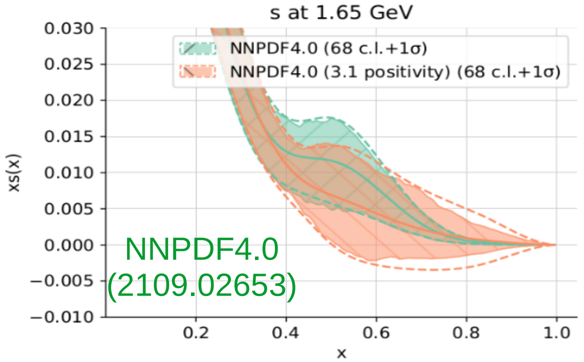
- In order to ensure physical PDFs, often additional constraints are added. Many different types and methods:

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

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





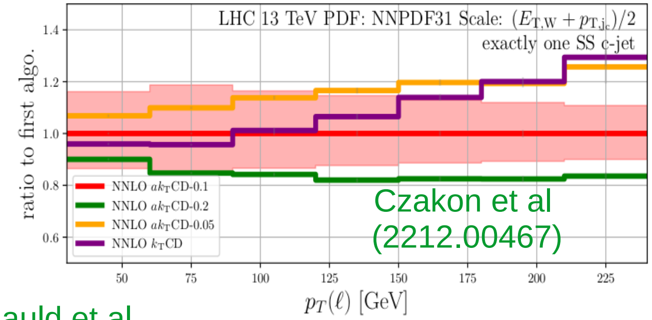


Methodological Challenges

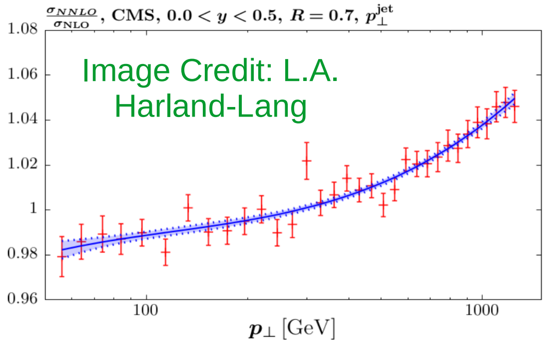
Theoretical Grids

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

Czakon et al (1912.08801)



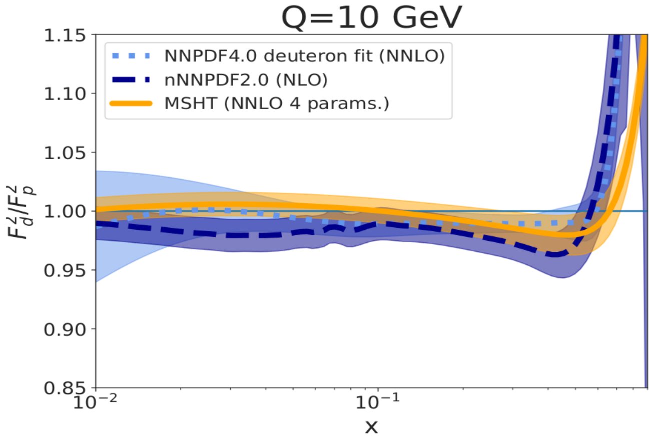
Gauld et al (2208.11138)

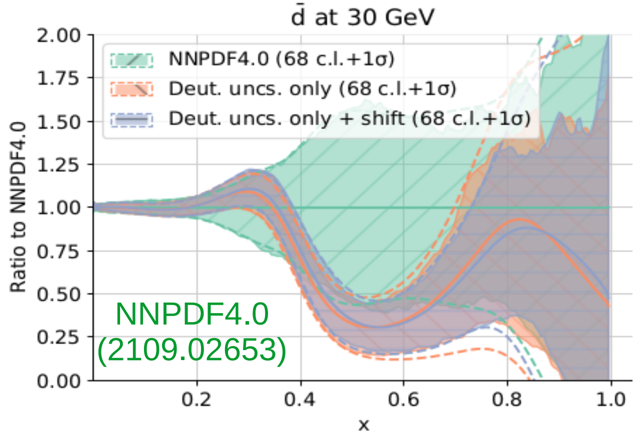


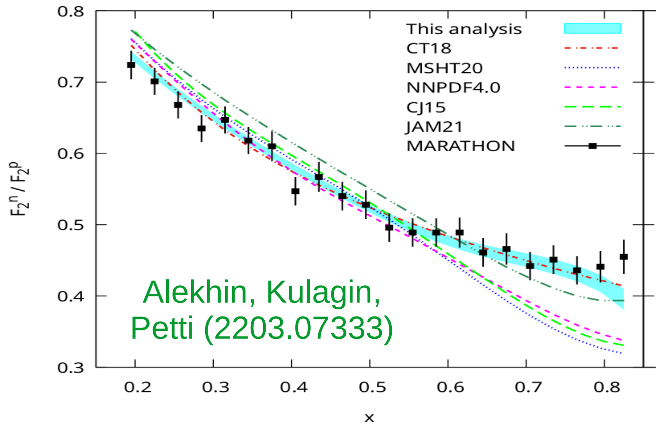
Theoretical Challenges

Deuteron and nuclear corrections

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







Some(!) aN3LO References...

①

[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.

aN3LO – How does it work?

1

- 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:

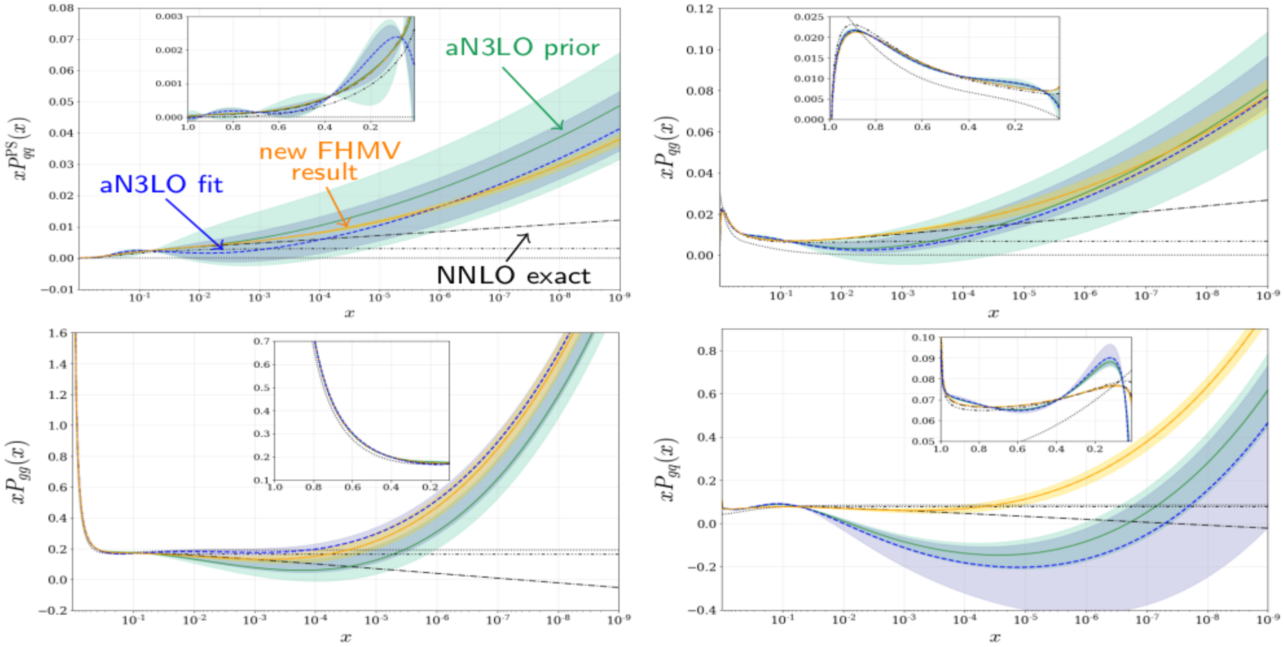
$$P_{ab}^{(3)}(x) = \sum_{i=1}^{k} A_i f_i(x) + f_e(x, \rho_{ab})$$

1 TNP per Splitting Function = 5 TNPs.

aN3LO – How accurate is it?

1

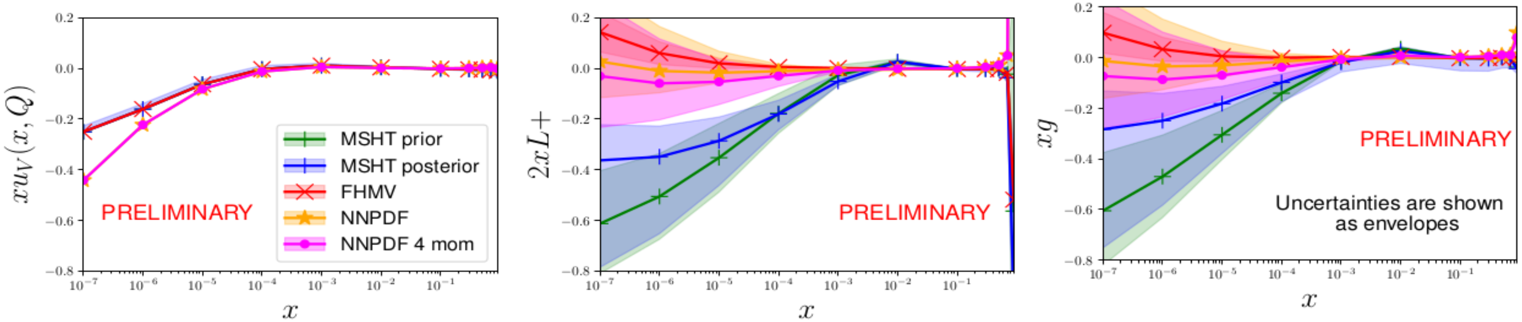
- Now 5 moments for P_{gg}, P_{gq} [12,13] and 10 for P_{qq}^{PS}, P_{qg} [14,15].
- Largely good agreement with MSHT determinations in central values.
- Exception is P_{gq}, least well determined (one extra low x log unknown).
- Reduction in P_{qq}^{PS}, P_{qg} uncertainties. Impacts reduced once in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



aN3LO evolution benchmarking

1

- aN3LO evolution benchmarking - use toy PDFs, no fit, no other complications and check impacts, as in hep-ph/0511119 (NNLO).
- Difference relative to NNLO evolution:



- Agreement between groups down to 10^{-3}, i.e. over data region.
- Up to few percent level effects on PDFs here due to N3LO evolution.
- Differences outside this with larger uncertainties at (very) low x.
- New information provides some additional constraints but still consistent with previous determinations.
- Different groups agree when using the same splitting functions.

Alphas bounds – NNLO

1

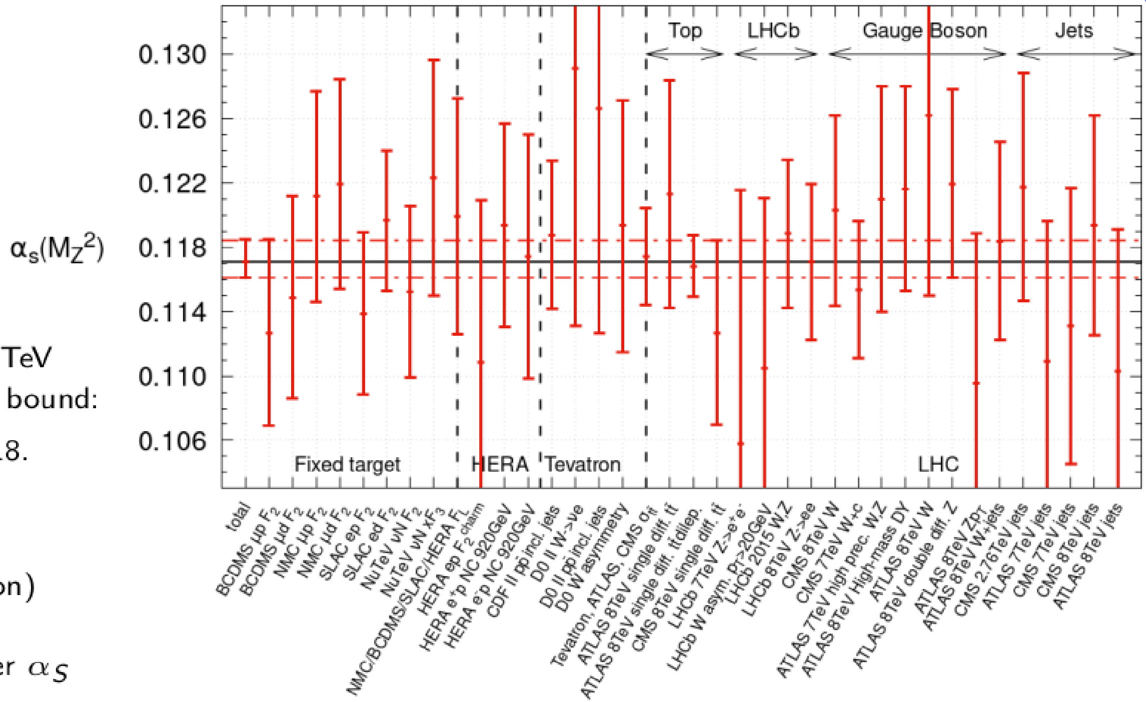
MSHT20 α_S bounds - NNLO

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.

Preliminary!

Alphas bounds – aNN3LO

MSHT20 α_S bounds - aN3LO

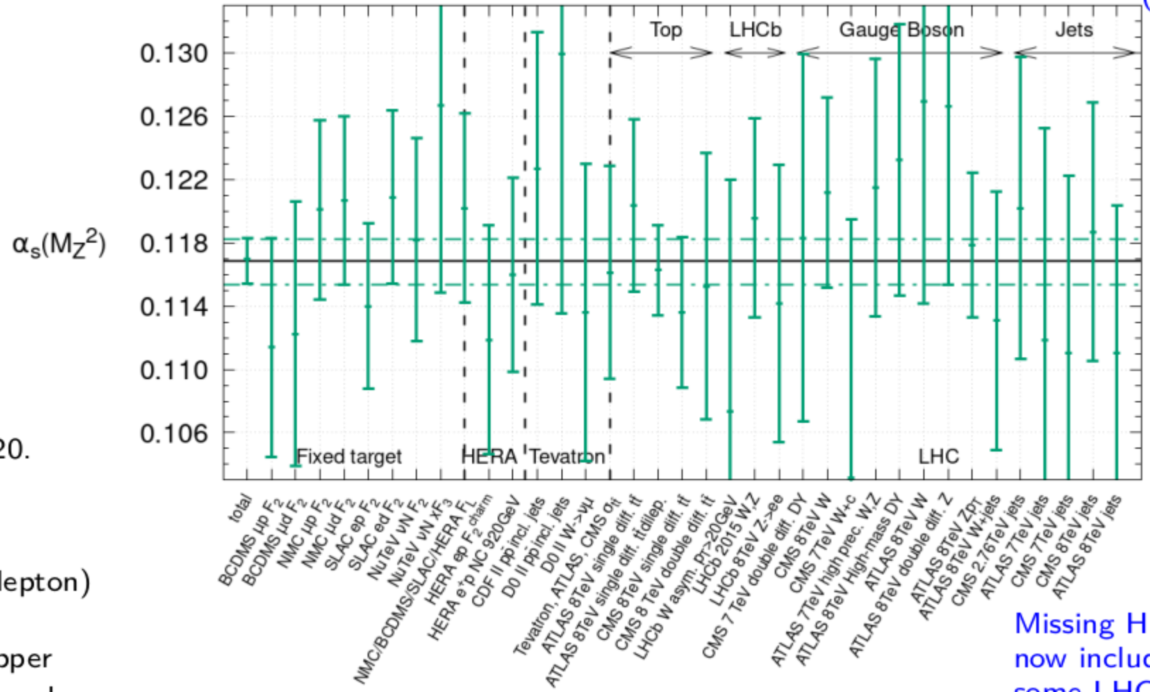
Preliminary!

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 $\Delta\alpha_S(M_Z^2) = -0.0017$.

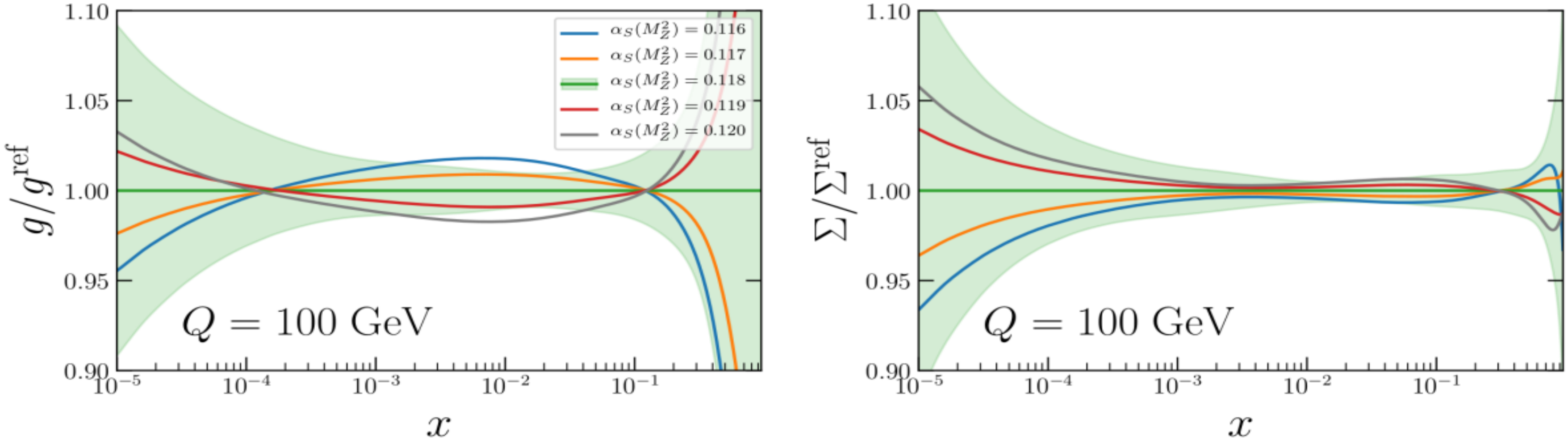
Missing Higher Order Uncertainties now included, in particular causes some LHC bounds to weaken as unknown N3LO K-factors.

Alphas correlation with PDFs... ①

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.

PDF + Strong Coupling for Xsecs

1

- 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.

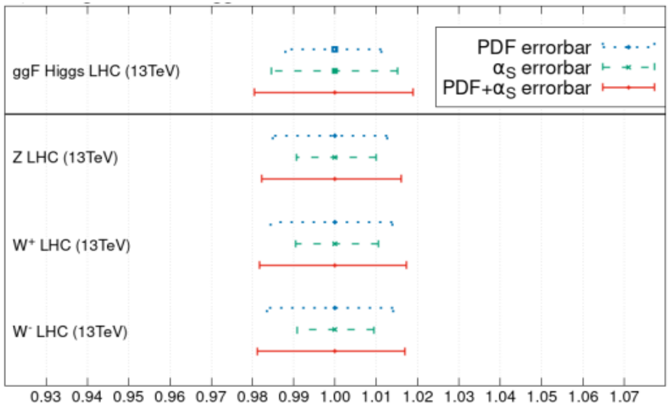
$$\Delta\sigma = \sqrt{(\Delta\sigma_{\mathrm{PDF}})^2 + (\Delta\sigma^2_{\alpha_S(M_Z^2)})}$$

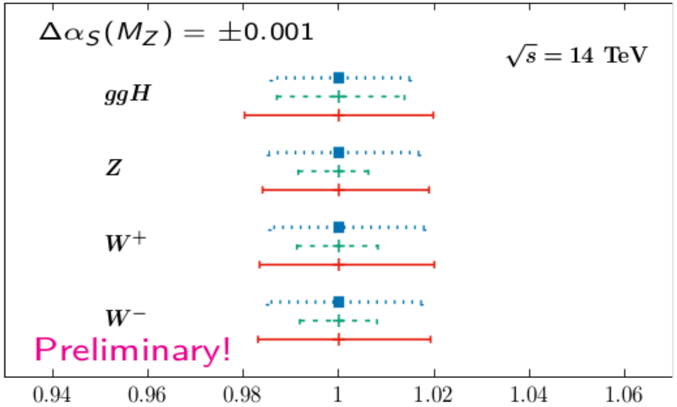
- Works provided central PDFs are best fit PDFs with $\alpha_S(M_Z^2)$ free.
- Choice of $\Delta\alpha_S(M_Z^2)$ up to user but recommended to be close to our 1σ bounds, e.g. ± 0.001 for simplicity and near that of world average.

PDF + Strong Coupling for Xsecs

1

- 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.

Dynamic Tolerance and PDFs

1

- 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 $\Delta\chi 2=T2$) 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. $\xi_{68} \gtrsim N_{pts}$.