# Parton Distribution Functions - Opportunities and Challenges

Robert Thorne

September 25th 2024

University College London

# Introduction to Parton Distribution Functions (PDFs)

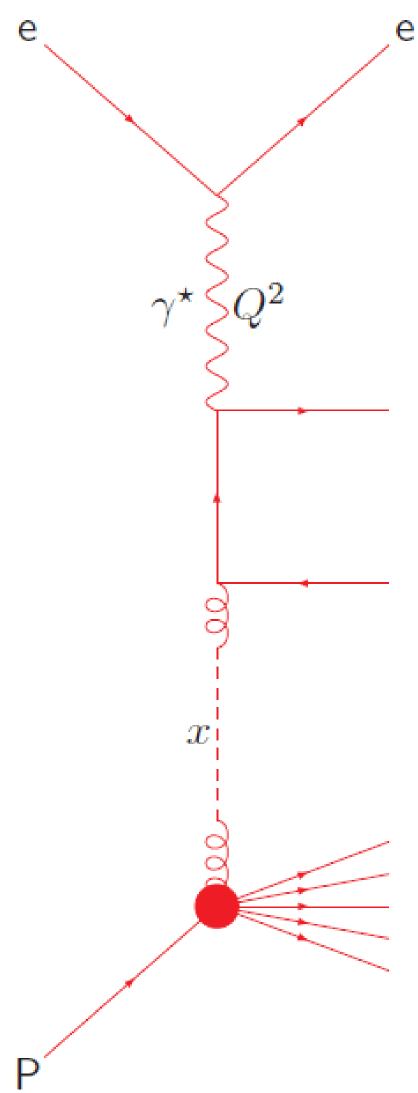
Strong force makes it difficult to perform analytic calculations of scattering processes involving hadronic particles.

The weakening of $\alpha_S(\mu^2)$ at higher scales $\longrightarrow$ the **Factorization Theorem**.

Hadron scattering with an electron factorizes.

$Q^2$ – Scale of scattering

$x = \frac{Q^2}{2m\nu}$ – Momentum fraction of Parton ($\nu$=energy transfer)



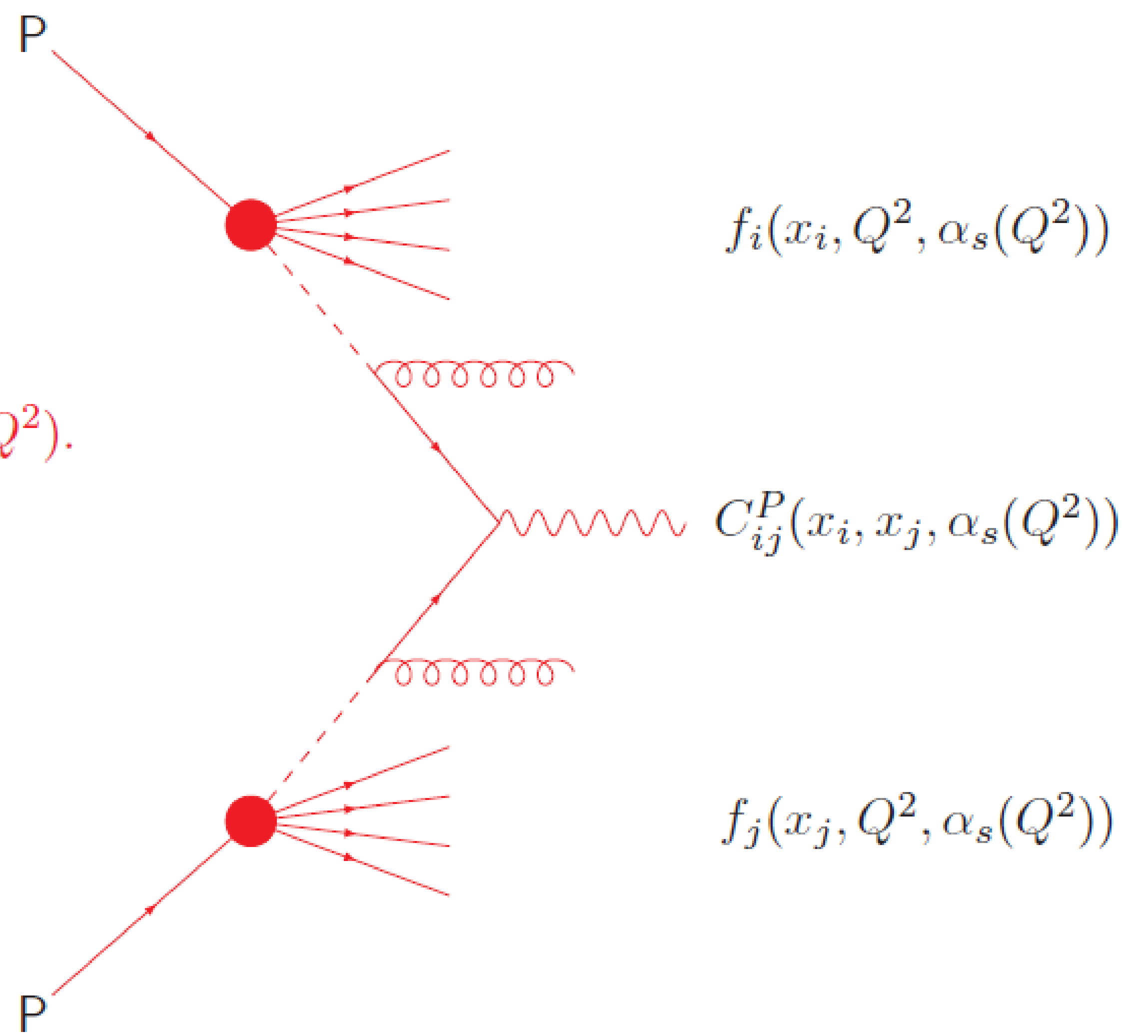

perturbative
calculable
coefficient function
$C_i^P(x, \alpha_s(Q^2))$

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

The coefficient functions $C_i^P(x, \alpha_s(Q^2))$ are process dependent (new physics) but are calculable as a power-series in $\alpha_s(Q^2)$.

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

Since the parton distributions $f_i(x, Q^2, \alpha_s(Q^2))$ are process-independent, i.e. **universal**, and evolution with scale is calculable, once they have been measured at one experiment, one can predict many other scattering processes.



## Obtaining PDF sets – General procedure.

Start parton evolution at low scale $Q_0^2 \sim 1\mathrm{GeV}^2$. 6 independent PDF combinations, or 7 if we assume $s = \bar{s}$.

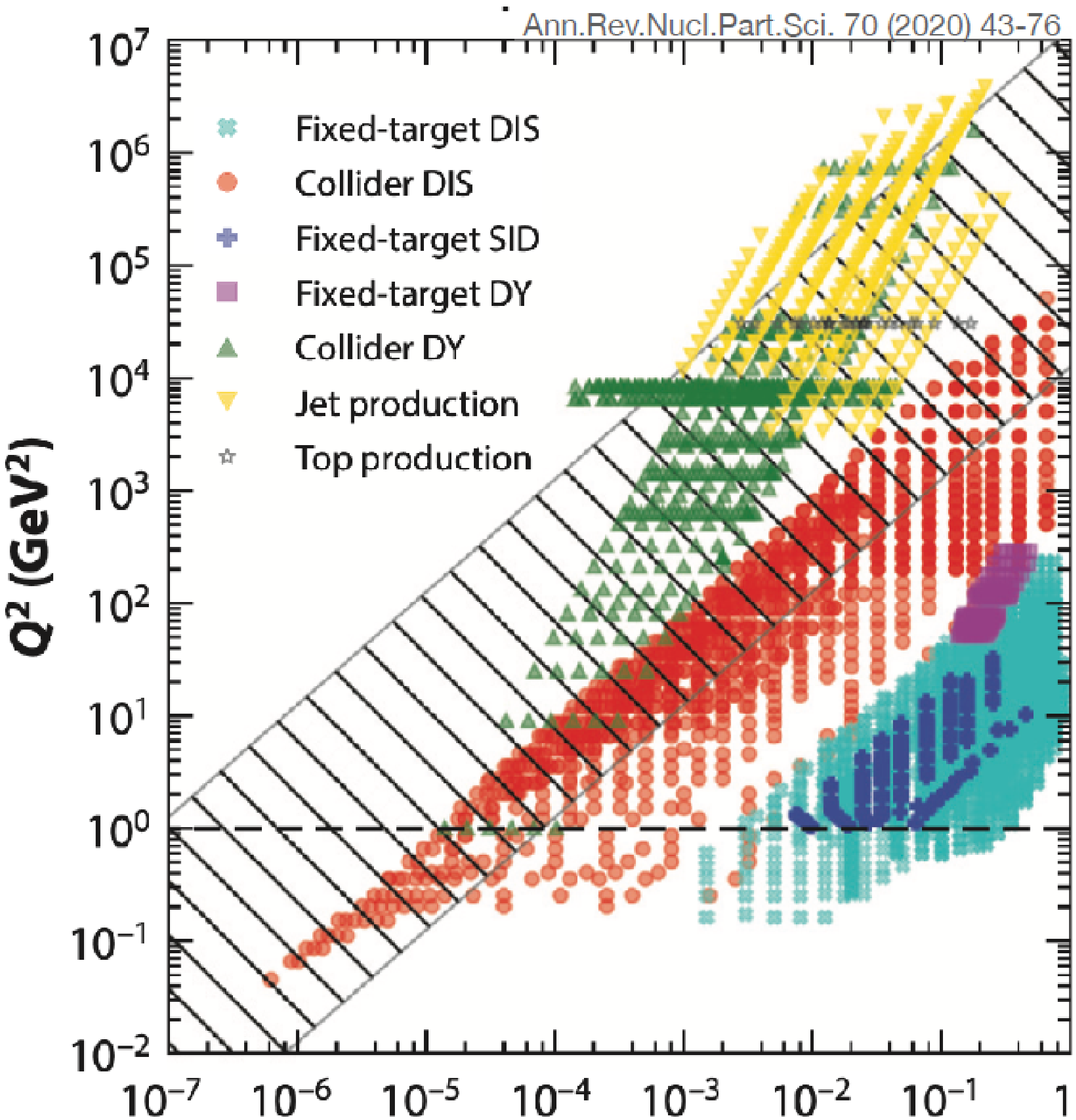
May also consider independent (nonperturbative) to contributions heavy quarks $c, b$.

Evolve partons upwards using LO, NLO or NNLO DGLAP equations.

$$\frac{df_i(x,Q^2,\alpha_s(Q^2))}{d\ln Q^2} = \sum_j P_{ij}(x,\alpha_s(Q^2)) \otimes f_j(x,Q^2,\alpha_s(Q^2))$$

Fit data above $\sim 2\mathrm{GeV}^2$. Need very many types for full determination.

# Range of Data Sets used



# Methodology

- Two distinct methodologies on the market to parameterising PDFs: **Neural Nets** (NNPDF) or **Explicit Parameterisation** (CT, MSHT).

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

$$\searrow\ \mathrm{NN}_i(x) \quad \textbf{NNPDF}$$

✦ MSHT: **52** free parameters in terms of Chebyshev polynomials.

$$u_V(x,Q_0^2) = A_u(1-x)^{\eta_u} x^{\delta_u}\left(1+\sum_{i=1}^{6} a_{u,i} T_i(y(x))\right)$$

$$d_V(x,Q_0^2) = A_d(1-x)^{\eta_d} x^{\delta_d}\left(1+\sum_{i=1}^{6} a_{d,i} T_i(y(x))\right)$$

$$S(x,Q_0^2) = A_S(1-x)^{\eta_S} x^{\delta_S}\left(1+\sum_{i=1}^{6} a_{S,i} T_i(y(x))\right)$$

$$s_+(x,Q_0^2) = A_{s_+}(1-x)^{\eta_{s+}} x^{\delta_S}\left(1+\sum_{i=1}^{6} a_{s_+,i} T_i(y(x))\right)$$

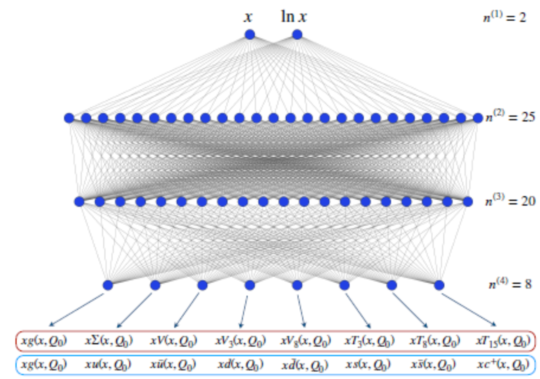
$$g(x,Q_0^2) = A_g(1-x)^{\eta_g} x^{\delta_g}\left(1+\sum_{i=1}^{4} a_{g,i} T_i(y(x))\right) + A_{g-}(1-x)^{\eta_{g-}} x^{\delta_{g-}}$$

$$s_-(x,Q_0^2) = A_{s\_}(1-x)^{\eta_{s-}}(1-x/x_0)x^{\delta_{s-}}$$

$$(\bar{d}/\bar{u})(x,Q_0^2) = A_\rho(1-x)^{\eta_\rho}\left(1+\sum_{i=1}^{6} a_{\rho,i} T_i(y(x))\right)$$

✦ Less flexible in general - need to be sure flexible enough! Allows direct handle on uncertainties in Hessian framework.

★ NNPDF: **763** free parameter Neural Net.




★ Increased flexibility, but needs robust optimisation + stopping (avoid over and under fitting).

Most groups use a parton parameterization and Hessian approach.

NNPDF fit to data replicas obtaining PDF replicas $q_i^{(net)(k)}$.

# Importance of perturbative precision.

NLO is very far from sufficient. Large change in some cross sections at NNLO.

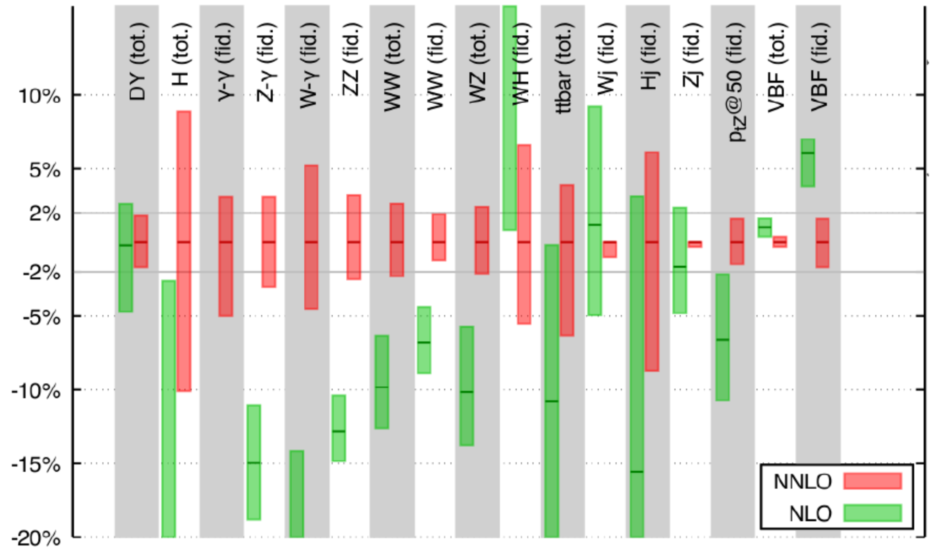



Image Credit: G. Salam

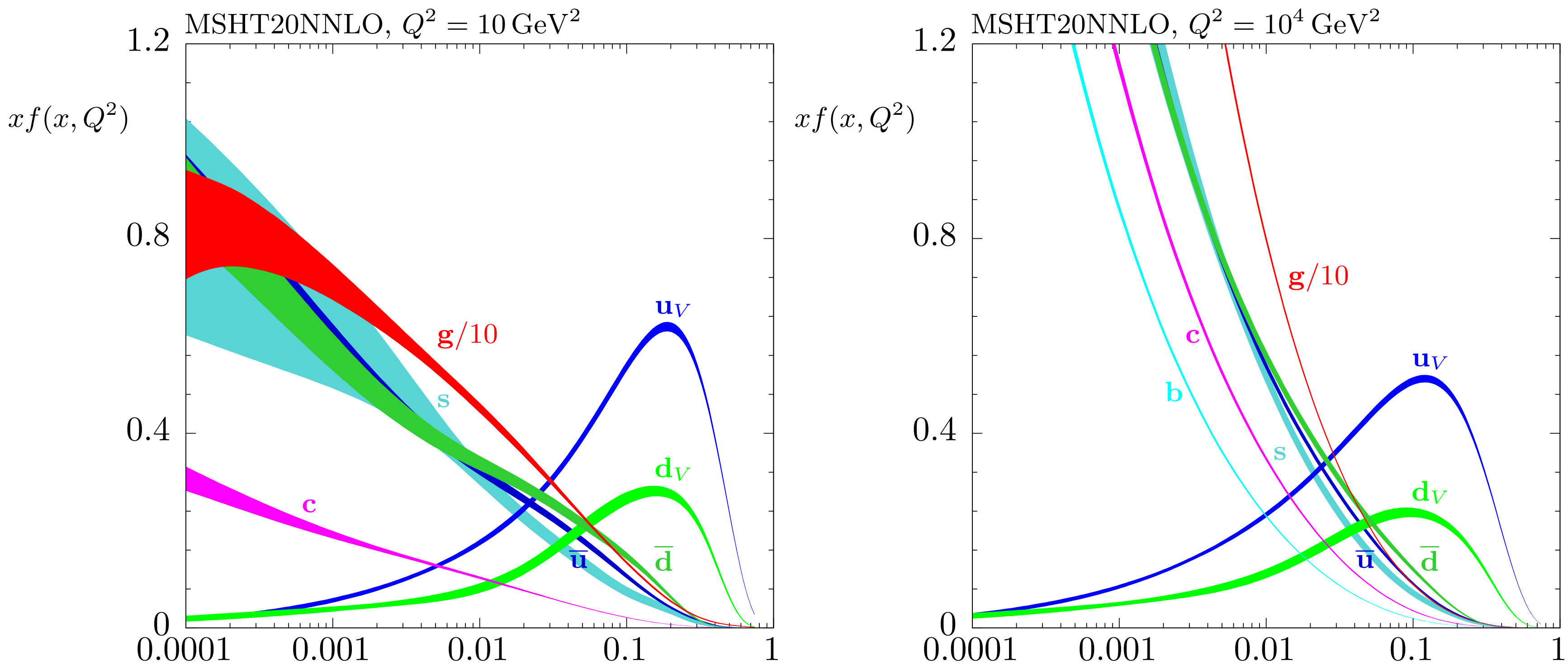
Also, a very poor fit at NLO to some data sets.

**MSHT20**

| Data set | $N_{\text{pts}}$ | NLO $\chi^2/N_{\text{pts}}$ | NNLO $\chi^2/N_{\text{pts}}$ |
|---|---|---|---|
| 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 |

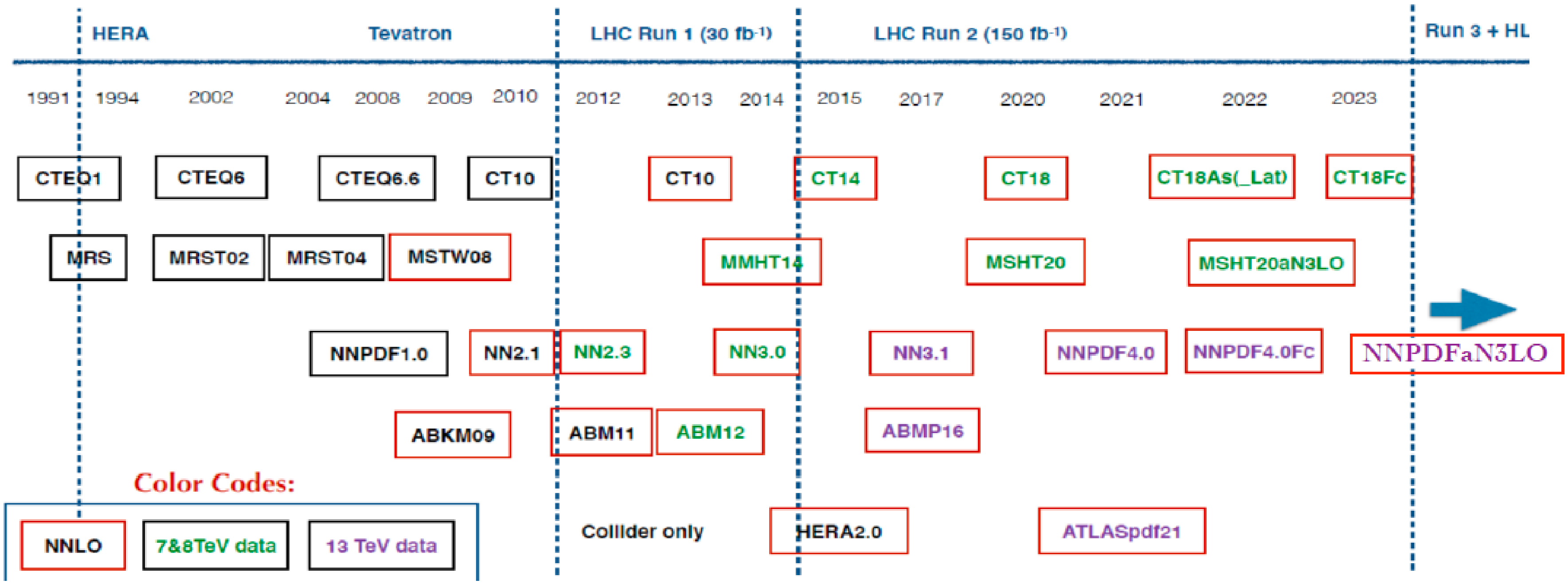
Image Credit: Tom Cridge

This procedure is generally successful and is part of a large-scale, ongoing project. Results in partons of the general form shown.




Various choices of PDF – MSHT, CT, NNPDF, ABM(P), HERA/ATLAS, CJ *et al* *etc.*. All cross-sections in hadron collisions rely on our understanding of these partons.

## **Major PDF Analyses** – slide from J. Gao



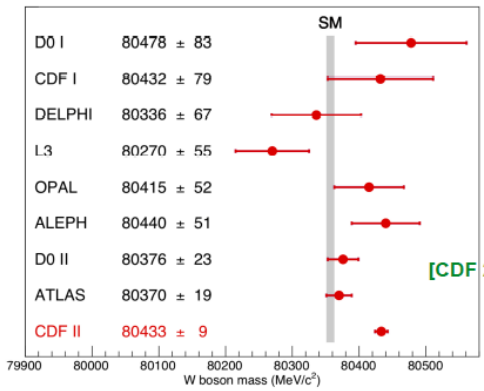

Many updates over the last few years.

Changes in methodology , not just more data included.

# Importance of PDFs.

**W mass**

W boson mass from different experiments




[CDF 2022]

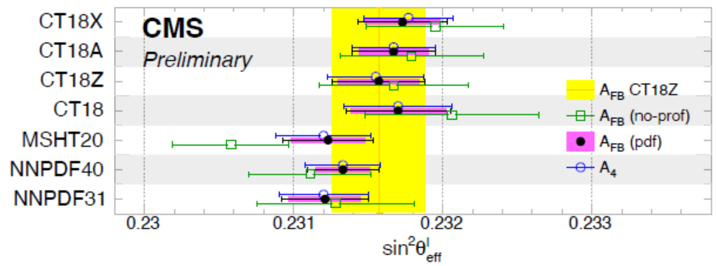
Fig. 5

SM expectation: $M_W = 80{,}357 \pm 4_{\text{inputs}} \pm 4_{\text{theory}}$ (PDG 2020)

LHCb measurement : $M_W = 80{,}354 \pm 23_{\text{stat}} \pm 10_{\text{exp}} \pm 17_{\text{theory}} \pm 9_{\text{PDF}}$ *[JHEP 2022, 36 (2022)]*

PDF unc. of CDF / ATLAS / LHCb: 3.9 / 8 / 9 MeV

$\sin^2\theta^l_{\text{eff}}$

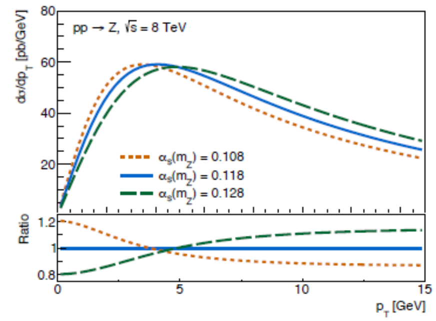



CMS PAS SMP-22-010

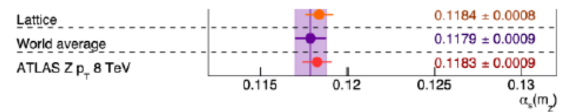
$\sin^2\theta^\ell_{\text{eff}} = 0.23157 \pm 0.00010(\text{stat}) \pm 0.00015(\text{syst}) \pm 0.00009(\text{theo}) \pm 0.00027(\text{PDF})$,

$\sigma_{\text{PDF}} \sim \sigma_{\text{tot}}$

$\alpha_S$




ATLAS, 2309.12986



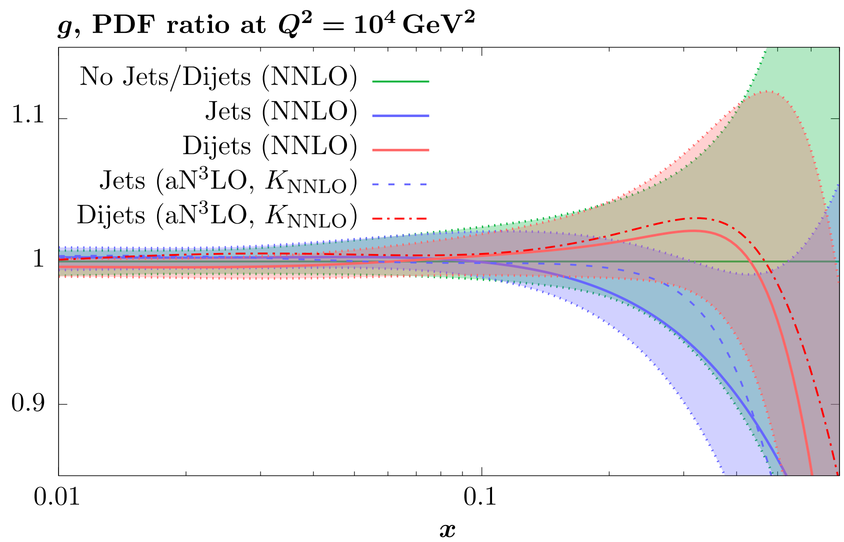

$\sigma_{\text{PDF}} \sim \sigma_{\text{tot}}/2$

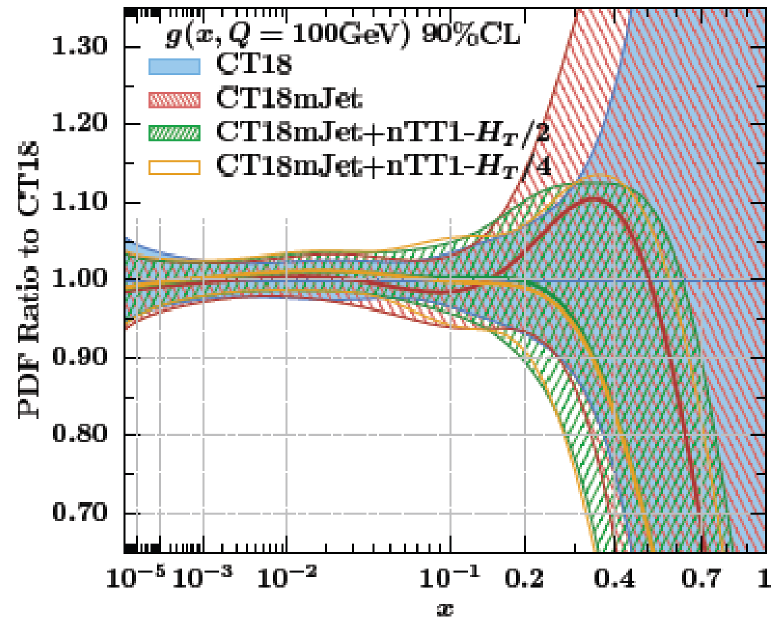
Beyond their fundamental interest in understanding QCD and the structure of the proton PDFs vitally important in particle physics.

Input to cross-sections.

Also major uncertainty in extracting seemingly unrelated fundamental parameters, e.g. $M_W, \sin^2\theta_W$. Fundamentally correlated with the strong coupling in many determinations.

## Impact of Recent Data.



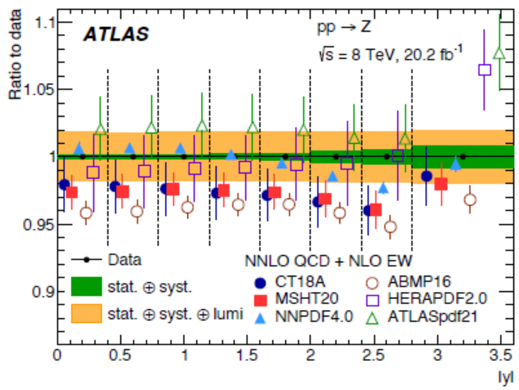




Recently many new types of data, but just a couple of examples.

Dijets and inclusive jets pull slightly differently. The latter are a better fit and are more stable under corrections.

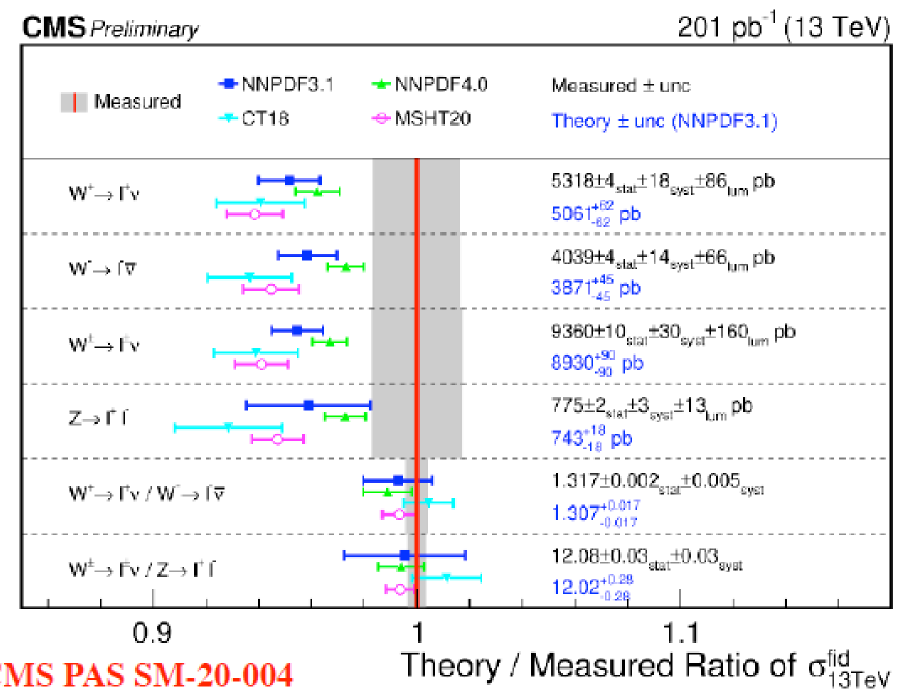
Top and jet data tend to pull the high-$x$ gluon differently.

# New Data Sets.



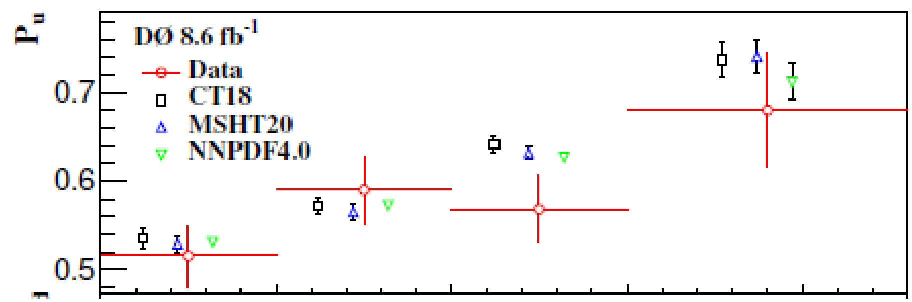

| PDF set | Total $\chi^2$ / d.o.f. | $\chi^2$ p-value | Pull on luminosity |
|---|---|---|---|
| MSHT20aN$^3$LO [58] | 13/8 | 0.11 | $1.2 \pm 0.6$ |
| CT18A [59] | 12/8 | 0.17 | $0.9 \pm 0.7$ |
| MSHT20 [60] | 10/8 | 0.26 | $0.9 \pm 0.6$ |
| NNPDF4.0 [61] | 30/8 | 0.0002 | $0.0 \pm 0.2$ |
| ABMP16 [62, 63] | 30/8 | 0.0002 | $1.8 \pm 0.4$ |
| HERAPDF2.0 [64] | 22/8 | 0.005 | $-1.3 \pm 0.8$ |
| ATLASpdf21 [65] | 20/8 | 0.01 | $-1.1 \pm 0.8$ |

ATLAS, arXiv:2309.09318




CMS PAS SM-20-004

Some high precision or new types of data recently presented (and soon to appear) will have an important impact.

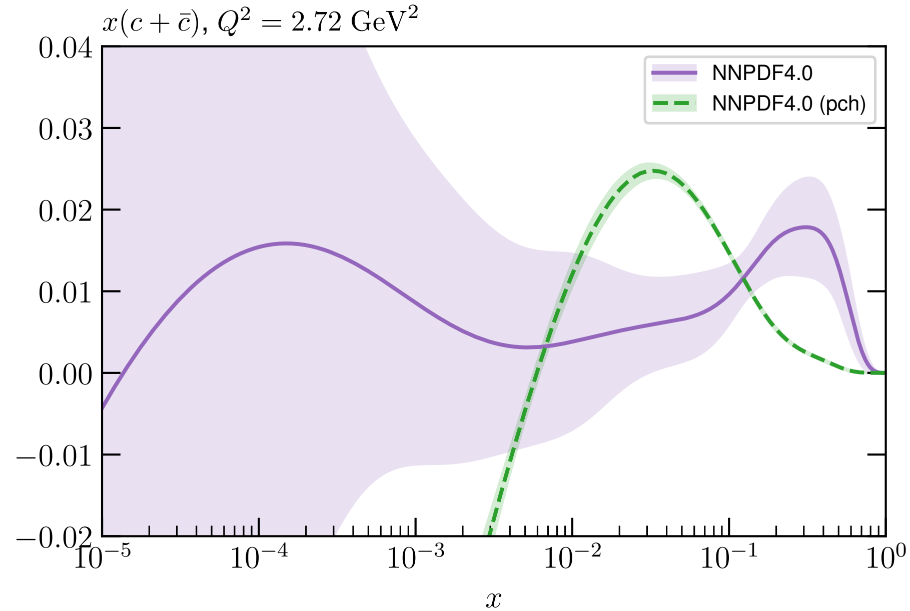


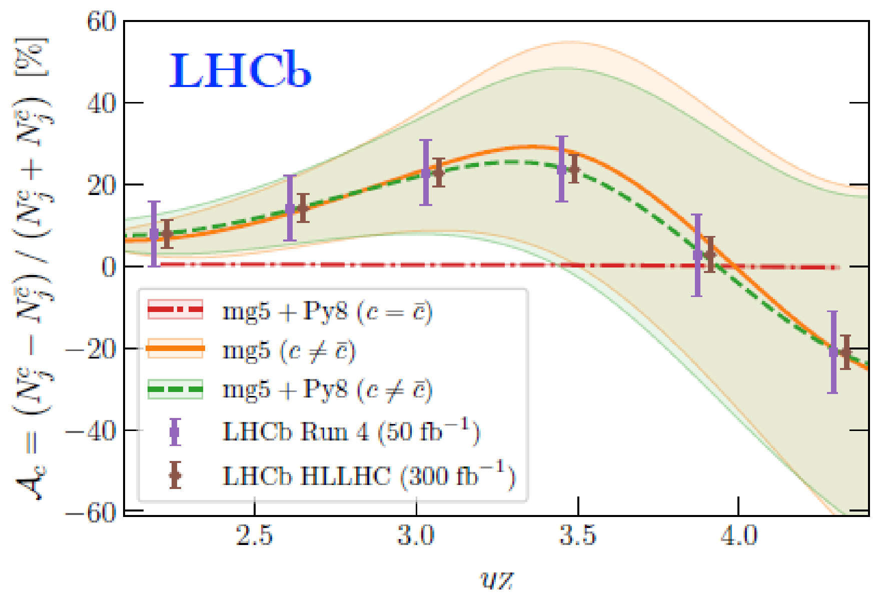
# Implications for Charm PDF.

Suggestions from structure function data and LHC collider data that fitted charm required.

Not overwhelming, and “fitted” and “intrinsic” difficult to disentangle.

Even some suggestion of intrinsic valence charm.




$$\mathcal{A}_c(y_Z) \equiv \frac{N_j^c(y_Z) - N_j^{\bar{c}}(y_Z)}{N_j^c(y_Z) + N_j^{\bar{c}}(y_Z)}$$

## aN$^3$LO PDFs and Theory Uncertainties

Leading source of uncertainties is from from Missing Higher Orders in perturbation theory. Numerous sources of this for e.g structure functions, i.e. splitting functions

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

but also heavy flavour transition matrix elements and cross-sections (coefficient functions)

$$F_2(x,Q^2) = \sum_{\alpha\in\{H,q,g\}} \Big( C_{q,\alpha}^{\mathrm{VF},\ n_f+1} \otimes A_{\alpha i}(Q^2/m_h^2) \otimes f_i^{n_f}(Q^2) + C_{H,\alpha}^{\mathrm{VF},\ n_f+1} \otimes A_{\alpha i}(Q^2/m_h^2) \otimes f_i^{n_f}(Q^2) \Big)\,,$$

Current knowledge is up to NNLO, with full higher orders unknown.

Already lots of progress in calculating features at N$^3$LO [2-13]. Since PDFs appeared also [14-18]

## N$^3$LO - What do we know?

Zero-mass structure function N$^3$LO coefficient functions are known [2], and flavour transition matrix elements for partons recently finalised [17,18].

For splitting functions some information from leading terms in the small $x$ and large $x$ regime [3-12], e.g.

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

Some numerical constraints (Low-integer Mellin moments) [3-12], and intuition from lower orders and expectations from perturbation theory.

Splitting Functions at aN$^3$LO – $N_m$ Mellin moments and small-$x$ constraints can be used to define

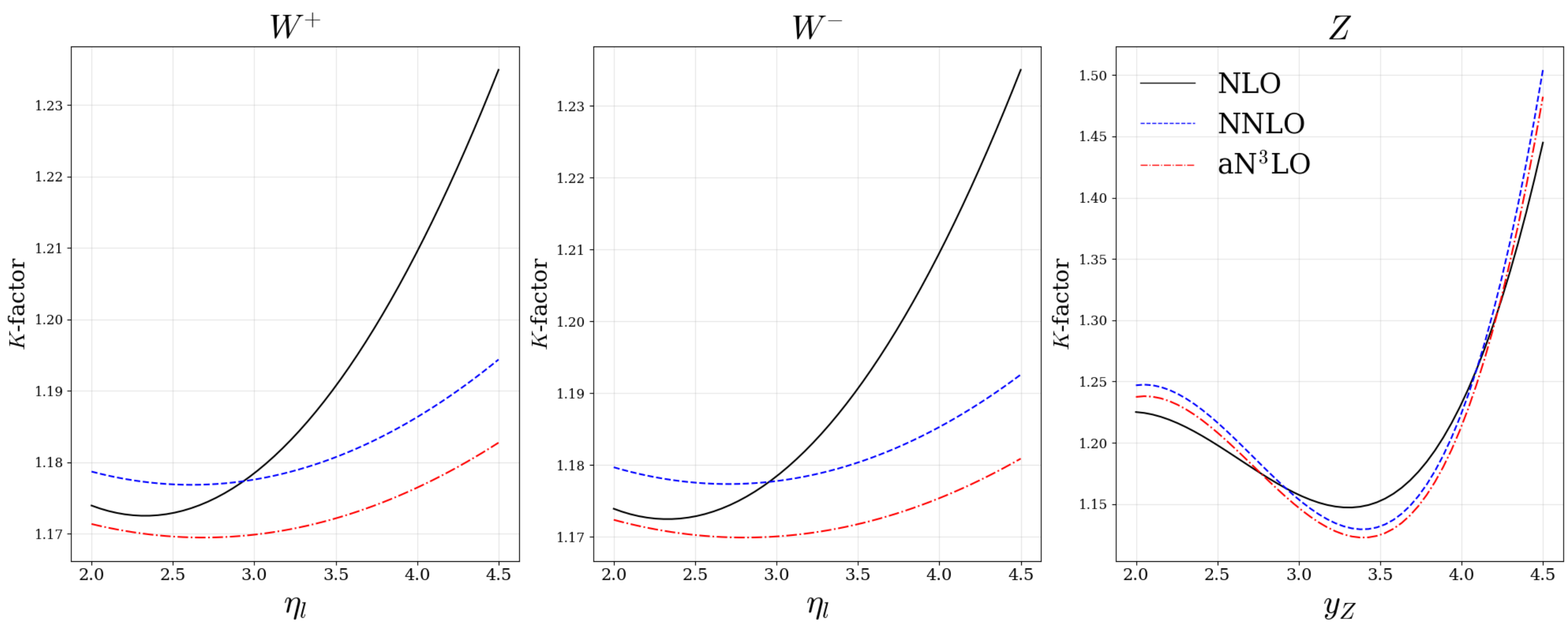
$$F(x) = \sum_{i=1}^{N_m} A_i f_i(x) + f_e(x).$$

Choose a set of relevant functions and solve for $A_i$.

Very little about many cross-sections (K-factors). e.g. parametrize the N$^3$LO K-factor as a superposition of both NNLO and NLO K-factors.

$$K(y) = 1 + \frac{\alpha_s}{\pi} D(y) + \left(\frac{\alpha_s}{\pi}\right)^2 E(y) + \left(\frac{\alpha_s}{\pi}\right)^3 F(y) + \mathcal{O}(\alpha_s^4).$$

$$K^{\mathrm{N^3LO/LO}} = K^{\mathrm{NNLO/LO}} \left(1 + \alpha_s^3 \hat{a}_1 \frac{\mathcal{N}^2}{\pi} D + \alpha_s^3 \hat{a}_2 \frac{\mathcal{N}}{\pi^2} E\right).$$



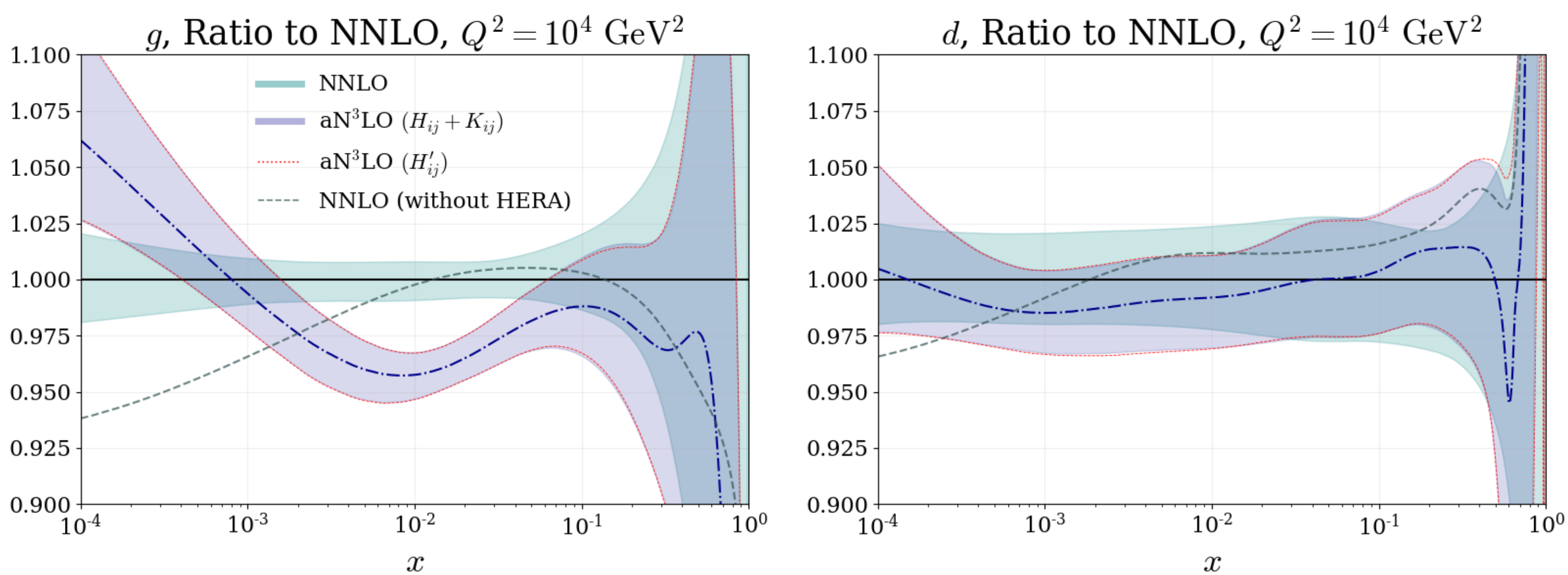

Calculations of N$^3$LO Drell Yan production now exist [19-21], but not in practical terms so far.

# Global Fit Quality at aN$^3$LO

The overall $\chi^2$ follows the general trend one may expect from perturbation theory.

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

Including aN$^3$LO has reduced tensions between small and large-$x$.




The gluon is enhanced at small-$x$ due to the large logarithms present at higher orders. Light quarks enhanced slightly at high $x$.

**NNPDF study recently completed [16].** Similar in numerous respects. More up-to-date inputs.

Approximate N$^3$LO splitting functions as

$$\gamma_{ij}^{(3)} = \gamma_{ij,n_f^3}^{(3)} + \gamma_{ij,N\to\infty}^{(3)} + \gamma_{ij,N\to 0}^{(3)} + \tilde{\gamma}_{ij}^{(3)}$$

Parametrise $\tilde{\gamma}_{ij}^{(3)} = \sum_l a_{ij}^{(l)} G_l(N)$

– $G_1$ for the leading unknown large-$N$ term

– $G_2$ for the leading unknown small-$N$ term

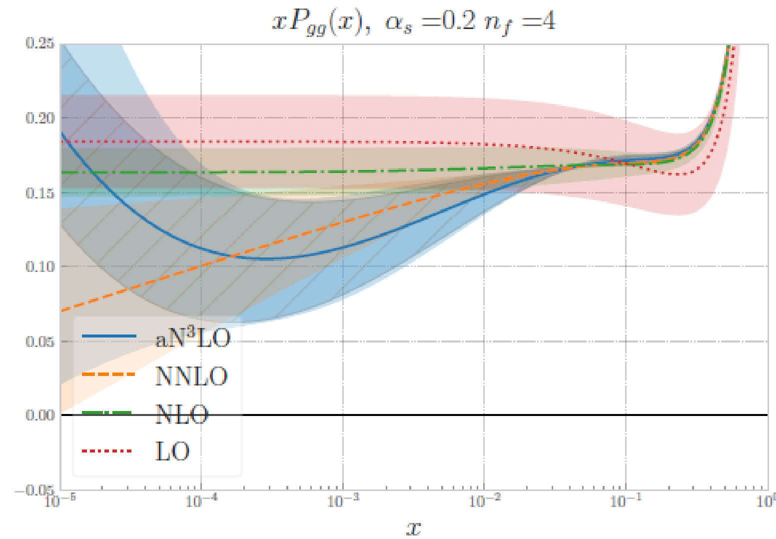
– 3 or 8 $G_l$ for the sub-leading unknown small- and large-$N$ contributions

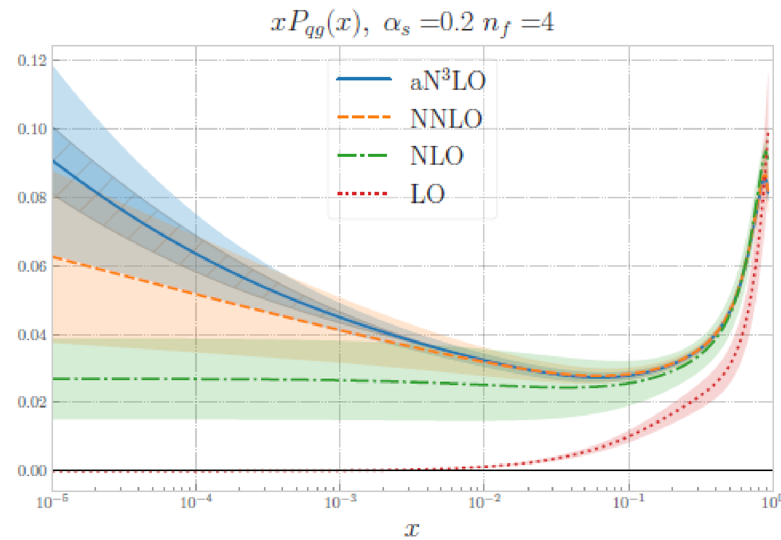
– vary the functions $G_l$ to generate a variety of approximations and estimate IHOU

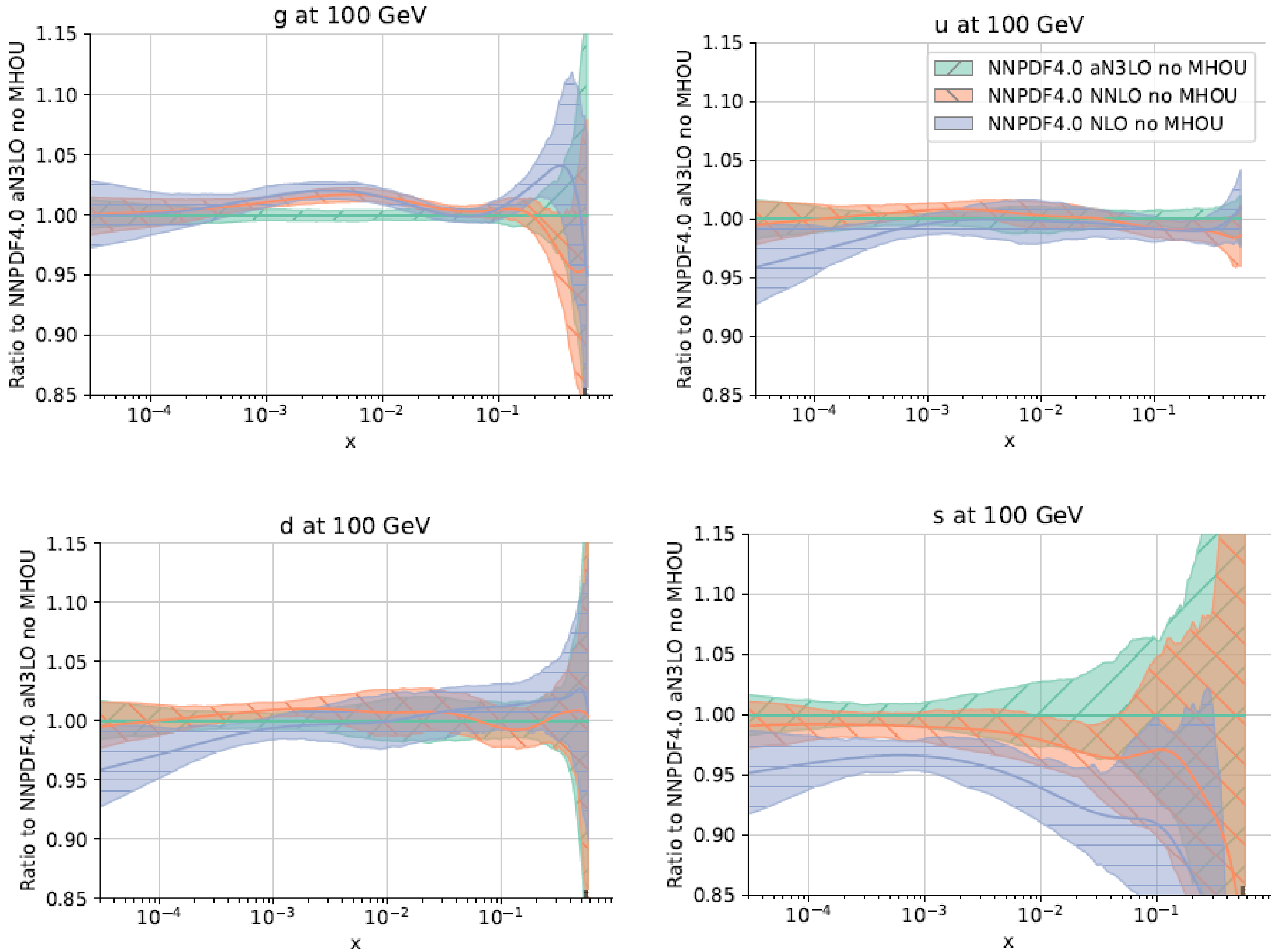
– determine the coefficients $a_{ij}^{(l)}$ with known moments and momentum conservation

*Adopted basis function for $\tilde{\gamma}_{qq}^{(3)}$*

| | |
|---|---|
| $G_1(N)$ | $\mathcal{M}[(1-x)\ln^2(1-x)]$ |
| $G_2(N)$ | $-\frac{1}{(N-1)^2} + \frac{1}{N^2}$ |
| $G_3(N)$ | $\frac{1}{N^4},\ \frac{1}{N^3},\ \mathcal{M}[(1-x)\ln(1-x)]$ <br> $\mathcal{M}[(1-x)^2\ln(1-x)^2],\ \frac{1}{N-1} - \frac{1}{N},\ \mathcal{M}[(1-x)\ln(x)]$ |
| $G_4(N)$ | $\mathcal{M}[(1-x)(1+2x)],\ \mathcal{M}[(1-x)x^2],$ <br> $\mathcal{M}[(1-x)x(1+x)],\ \mathcal{M}[(1-x)]$ |

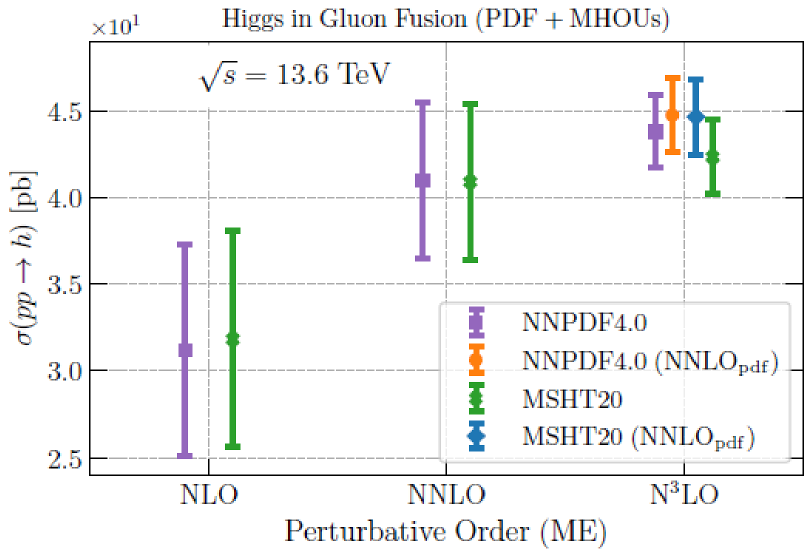


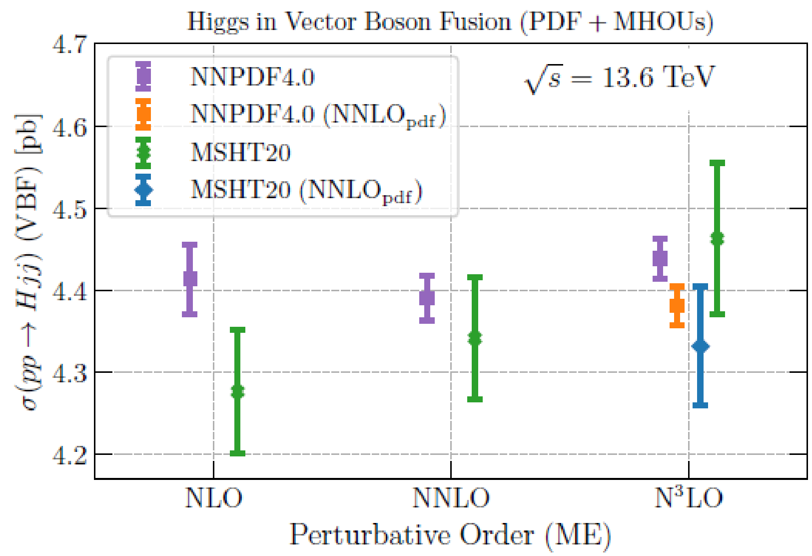






PDFs - main change in $g$

# Consequences for Higgs Cross Sections.



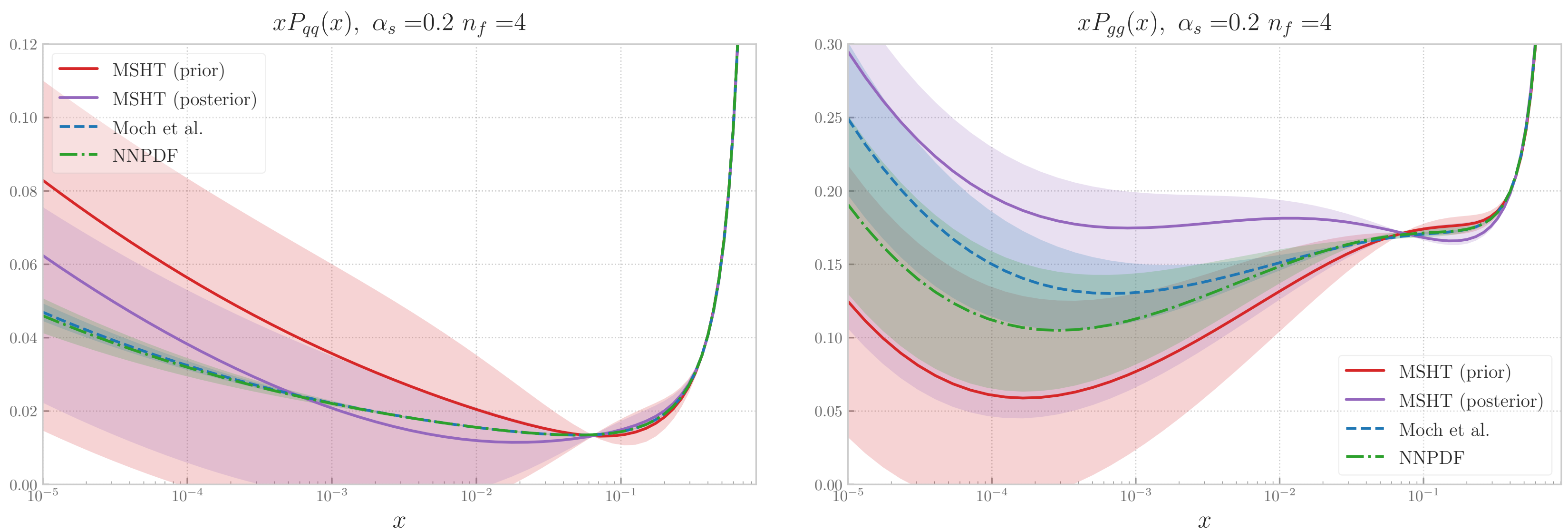




Changes in N$^3$LO cross section relative to use of NNLO PDFs obvious. Smaller for NNPDF than MSHT.

# Recent improvements in knowledge of splitting functions.

Now 5 moments available for $P_{gg}$. Allows improved constraint provided by [17] (Moch et al.)

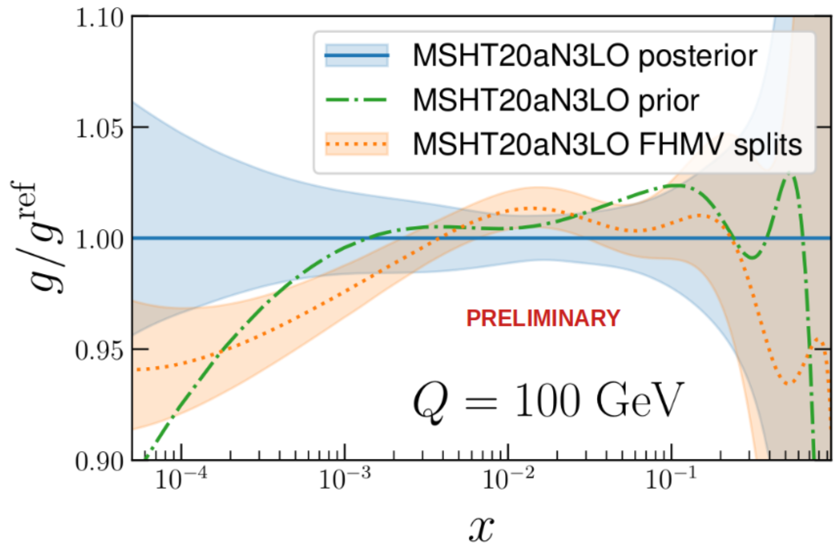
Now 10 moments for $P_{qq}^{PS}$ and $P_{qg}$, and very recently $P_{gq}$. Allows much improved constraint in [17,18] (Falconi, et al.).

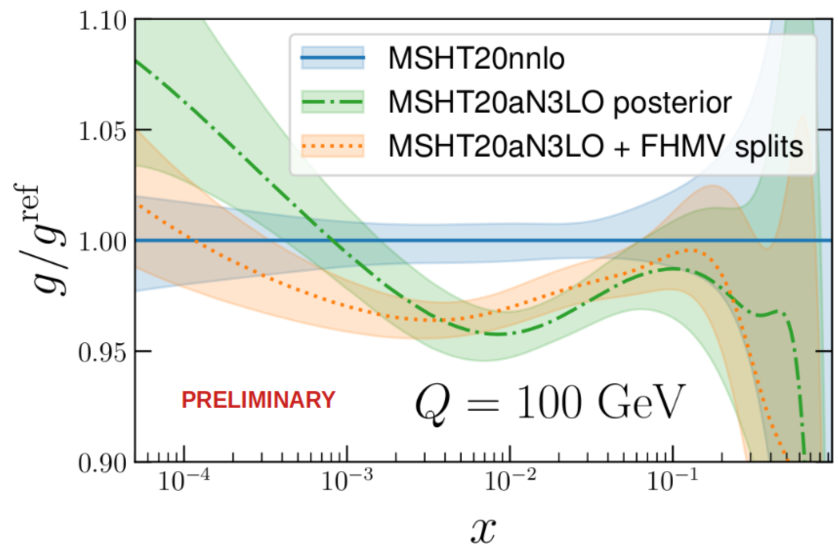



Range of allowed total splitting splitting functions.

We are close to the level at N$^3$LO that we were when NNLO PDFs were becoming standard $\sim 2008$.

## Effect of MSHT fits with improved [14-16] splitting functions.



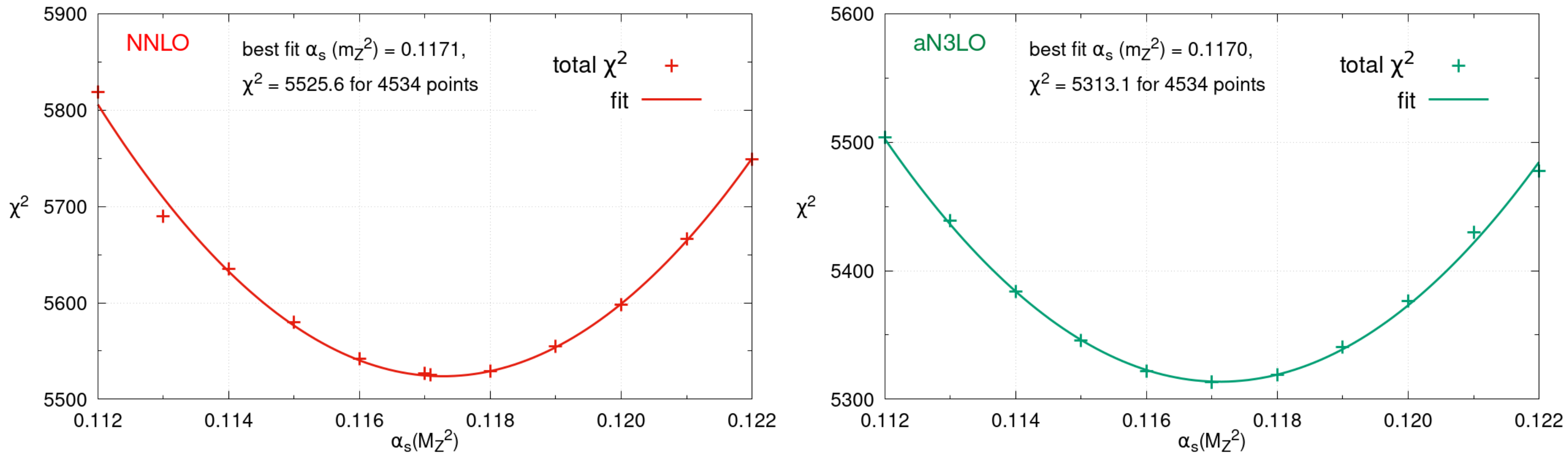

Note - only central value.

$\chi^2$ still over 100 lower than NNLO, very largely at small $x$ - would improve further once uncertainty accounted for.

aN$^3$LO gluon changes a little compared to published MSHT PDFs, raising $1.5\%$ near $x = 0.01$.

Main features of aN$^3$LO comparison to NNLO remain the extremely similar.

## Best fit value of $\alpha_S(M_Z^2)$ at aN$^3$LO possible 2404.02964




Previously [21] we found at NNLO that $\alpha_S(M_Z^2) = 0.1174 \pm 0.0013$.

Repeat analysis at NNLO with new baseline (ATLAS 8 TeV inclusive jet data) and also at aN$^3$LO.

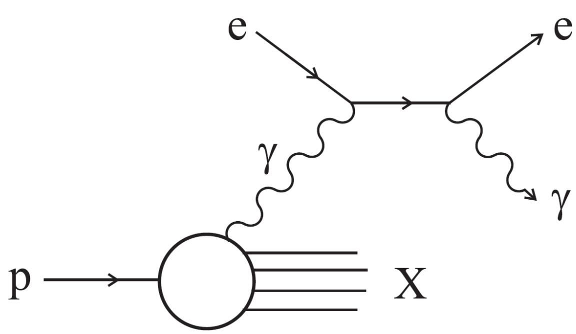
$$\alpha_S(M_Z^2) = 0.1171 \pm 0.0014 \text{ NNLO}$$

$$\alpha_S(M_Z^2) = 0.1170 \pm 0.0016 \text{ aN}^3\text{LO}$$

Determine uncertainty by dynamical tolerance procedure, same as for eigenvector uncertainties. Uncertainty corresponds to $\Delta\chi^2 = 13$ NNLO, $\Delta\chi^2 = 16$ N$^3$LO.

# PDFs up to aN$^3$LO with QED corrections 2312.07665

At the level of accuracy we are now approaching it is important to account for electroweak corrections. For a consistent treatment we need PDFs which incorporate QED into the evolution, i.e. the inclusion of the photon PDF $\gamma(x, Q^2)$ [22-24].

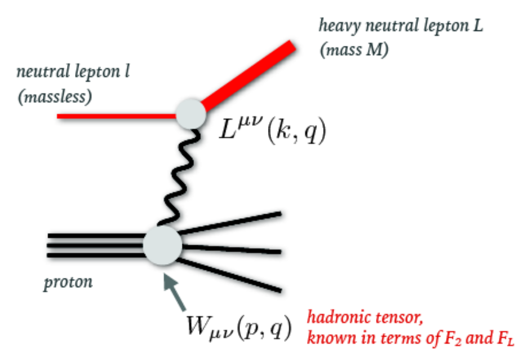



$$\frac{\partial \gamma(x,Q^2)}{\partial \log Q^2} = \frac{\alpha}{2\pi}\int_x^1 \frac{dy}{y}\left(P_{\gamma\gamma}\otimes\gamma + \sum_1 e_i^2 P_{\gamma q}\otimes q_i\right)$$

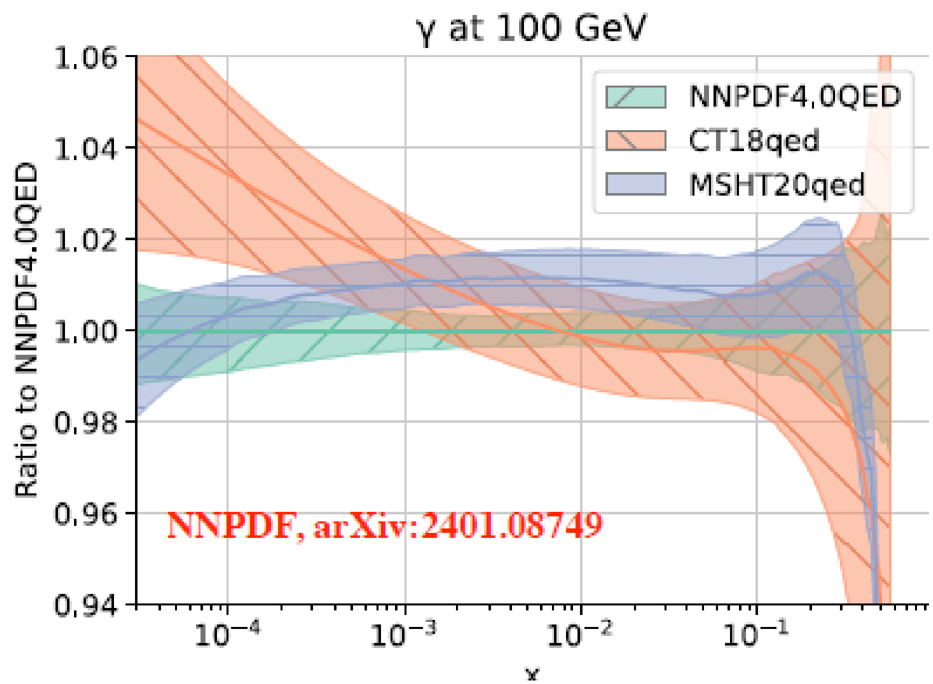
Put on truly quantitative footing in LUXqed photon PDF [25]. Relates photon to structure functions, and uncertainty of at most a few percent.

$$x f_{\gamma/p}(x,\mu^2) = \frac{1}{2\pi\alpha(\mu^2)}\int_x^1 \frac{dz}{z}\left\{ \int_{\frac{x^2 m_p^2}{1-z}}^{\frac{\mu^2}{1-z}} \frac{dQ^2}{Q^2}\alpha^2(Q^2) \left[\left(z p_{\gamma q}(z) + \frac{2x^2 m_p^2}{Q^2}\right) F_2(x/z, Q^2) - z^2 F_L\left(\frac{x}{z}, Q^2\right)\right] - \alpha^2(\mu^2) z^2 F_2\left(\frac{x}{z},\mu^2\right)\right\}, \quad (6)$$

22

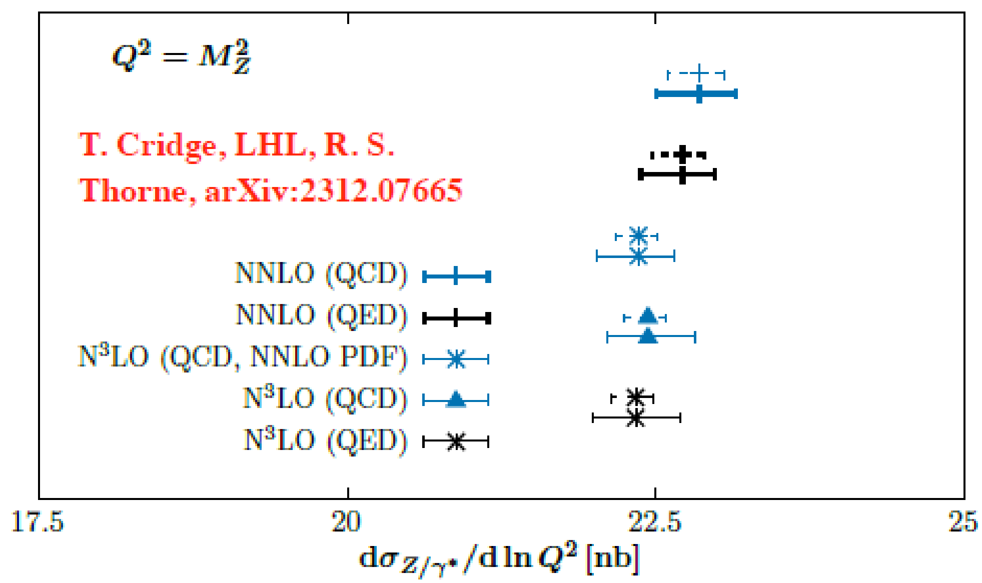


This results in a largely consistent set of photon PDFs between groups.




Differences similar to differences in quark PDFs, couple of percent uncertainties.

Direct impact due to input photon PDF in some cases, e.g. $Z + H$.



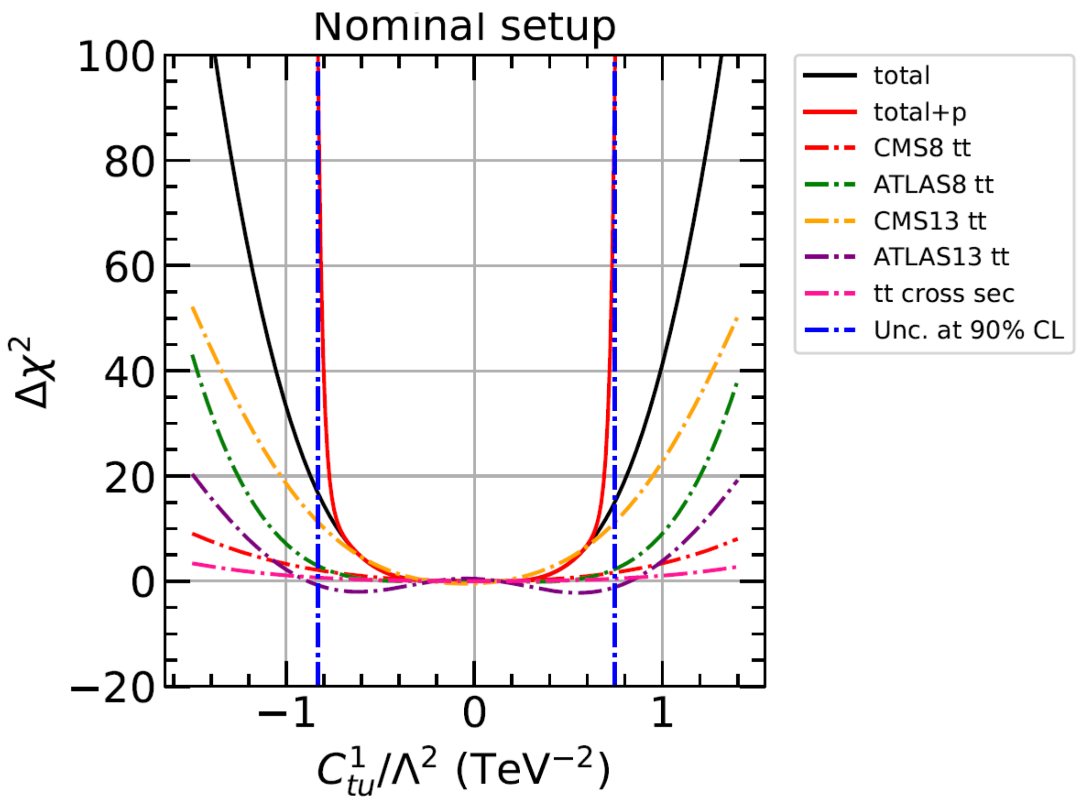

Generally small impact (reduction) due to other PDF modifications.

## New Physics and PDFs.

Look for indirect signs of new physics in standard high energy data.

Parametrize in SMEFT

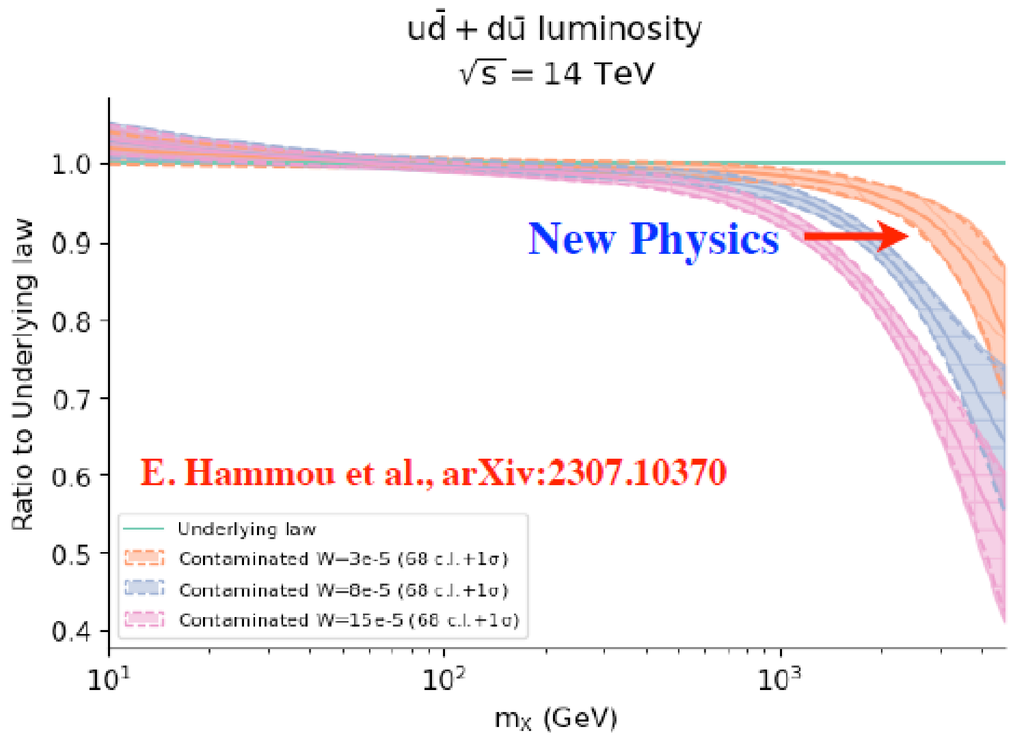
$$\mathcal{L}_{SMEFT} = \mathcal{L}_{SM} + \sum_i \frac{C_i O_i^{(6)}}{\Lambda^2} + \cdots .$$

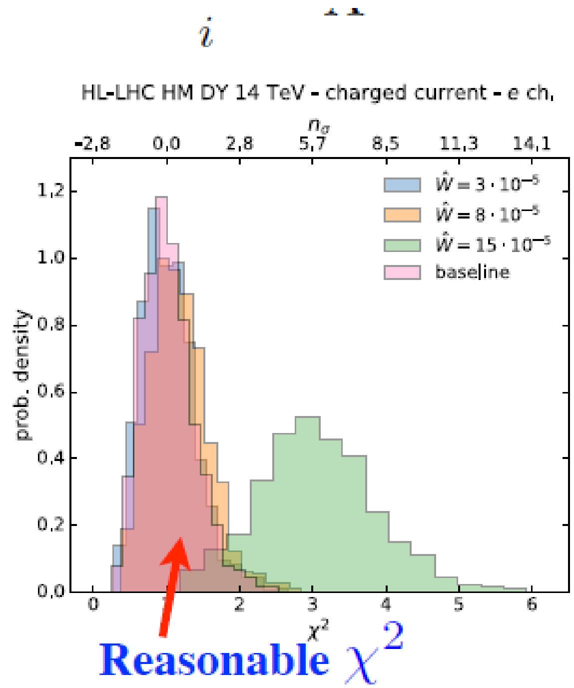



Need to account for interplay of BSM Physics with PDFs.

Example, CT study 2211.01094, joint fit to SMEFT + PDF parameters.

Similar study by NNPDF. Also consider HL-LHC pseudodata, and if it can be absorbed in PDF fit 2307.10370.



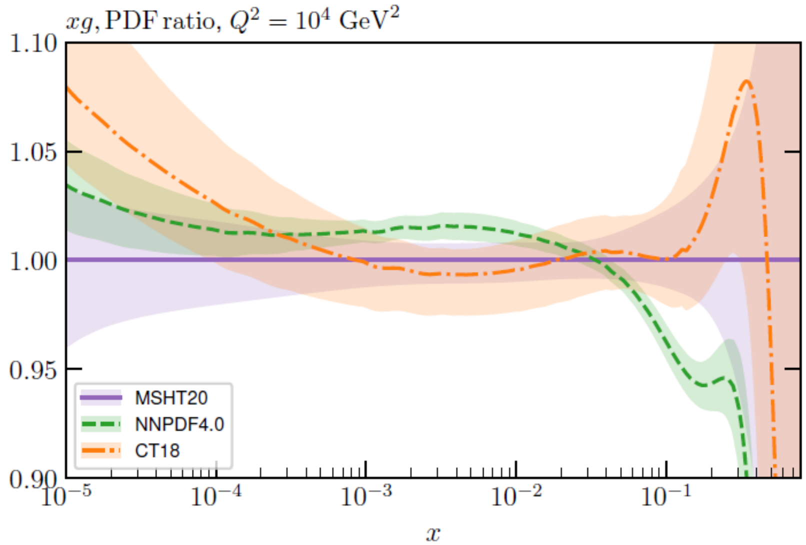



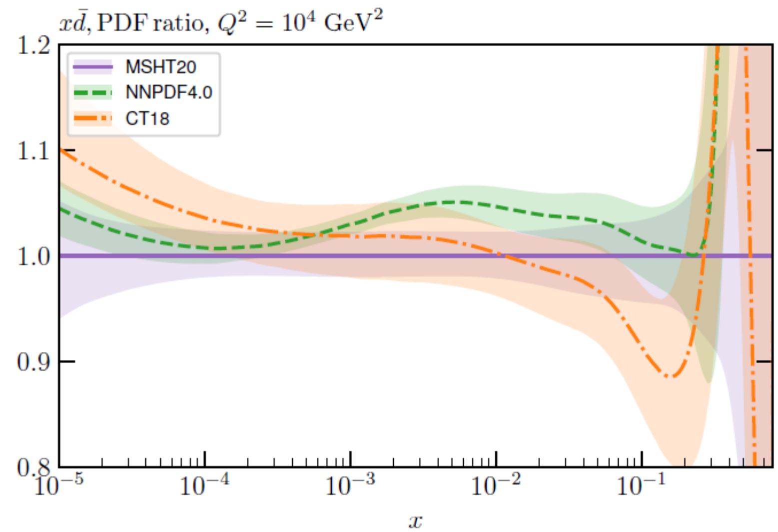

For some models, depending on new physics scale, it can.

Can reduce effects by looking a cross section ratios.

Much better to constrain PDFs from both high and lower energy constraints, e.g. HL-LHC and EIC. Highlights need for global fits.

# Current PDFs at NNLO



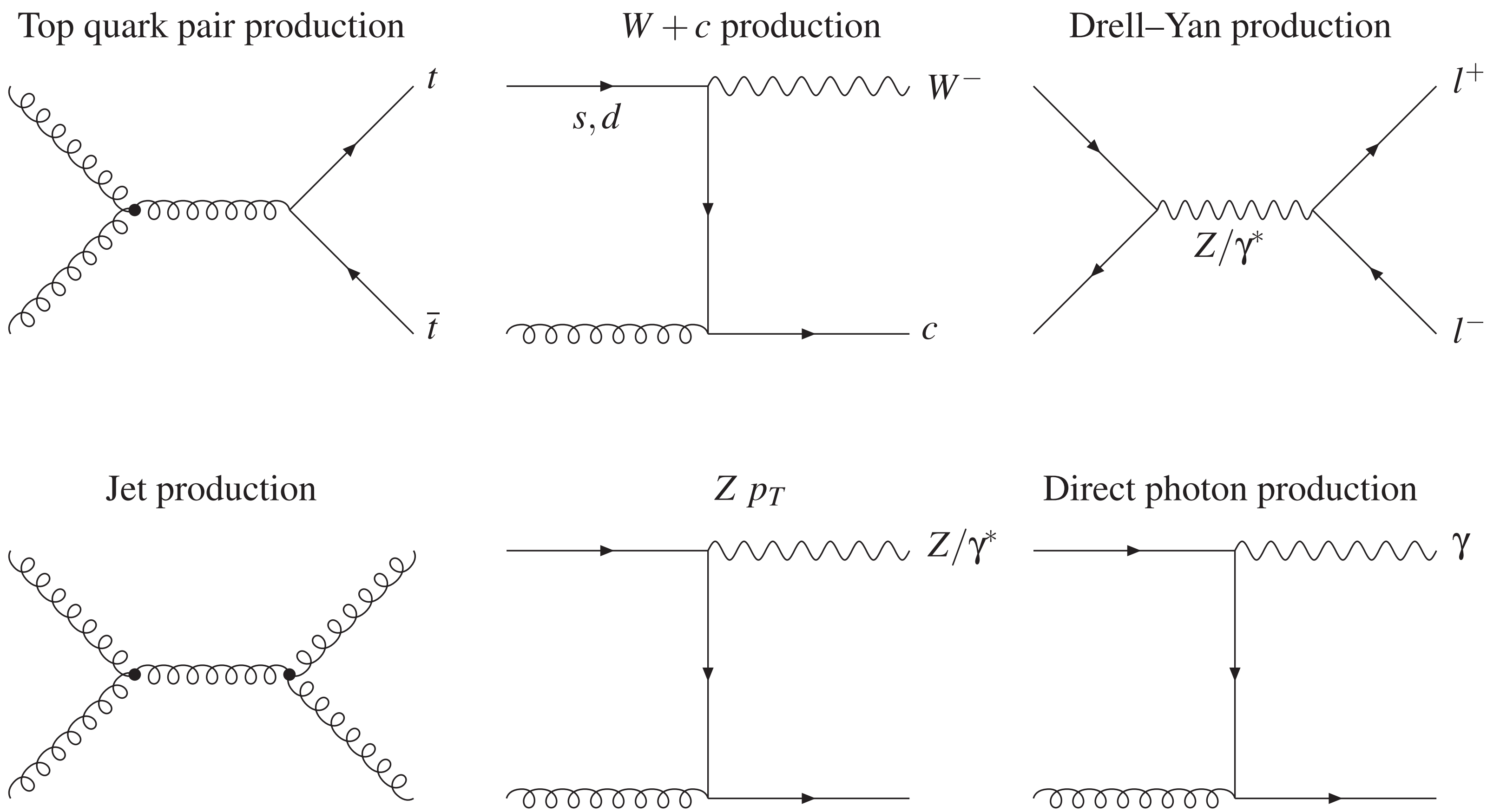




An illustration of a current comparison of PDFs.

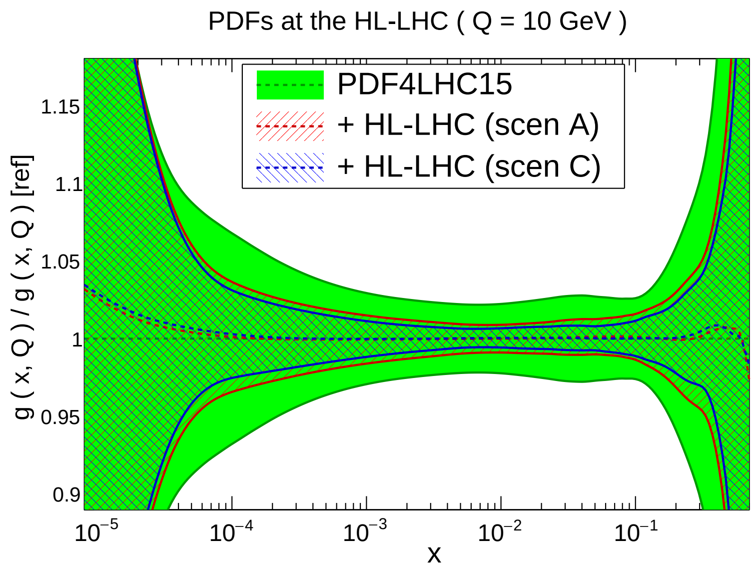
Uncertainties and agreement has improved with time.

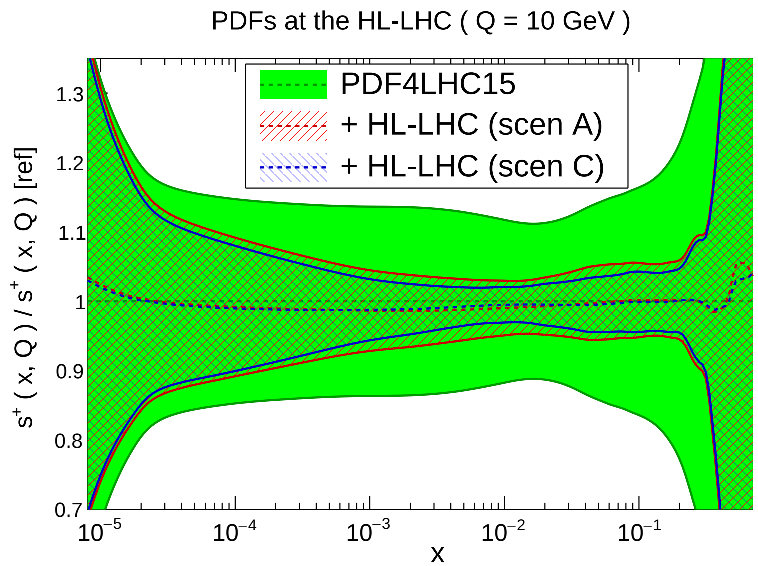
However, still disagreement in some central values, and quite markedly in some uncertainties.

# **Future Experiment Impacts – HL-LHC** - study in 1810.03639.




Types of data sets used in the study. Chosen to maximise impact on currently less well-known PDFs, regions.

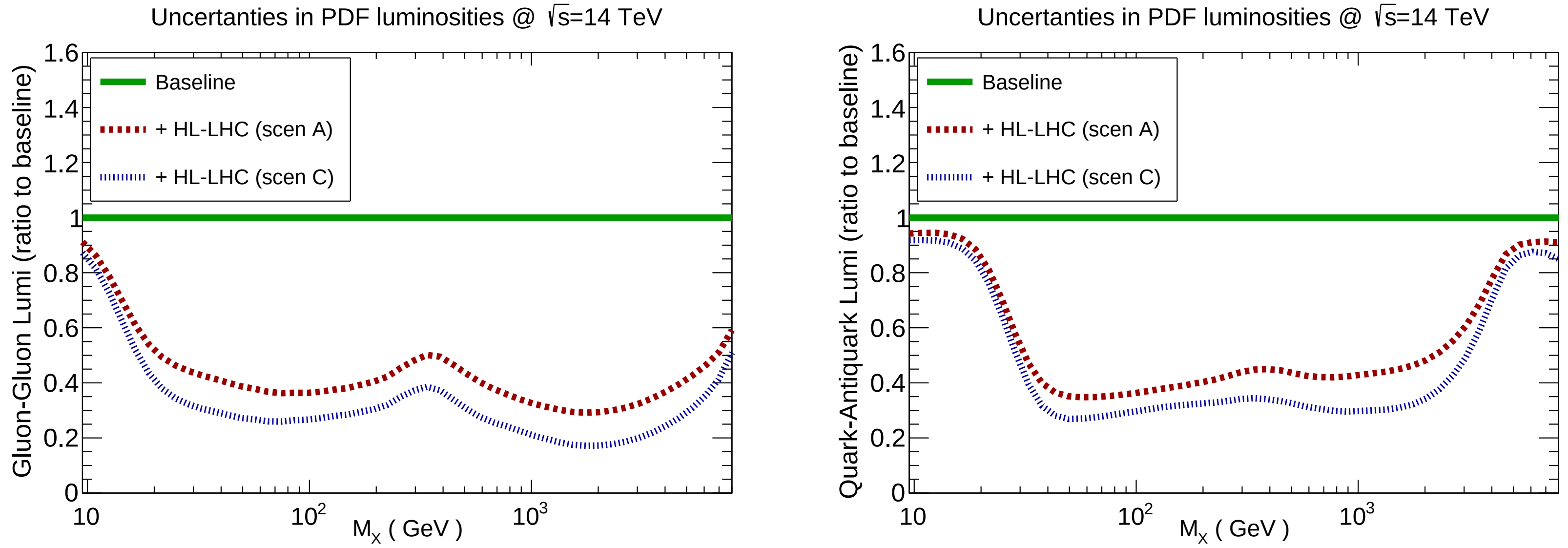



Impact in more or less optimistic (regarding systematic uncertainties) scenarios.

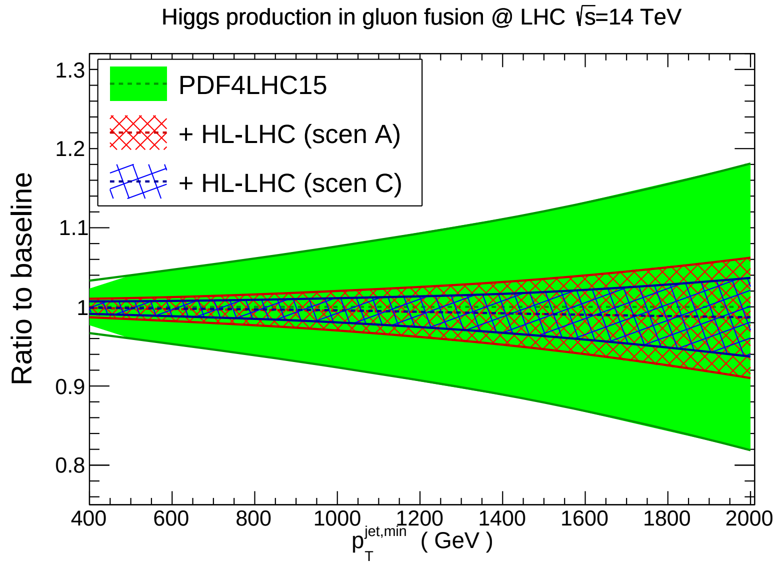
Note, “tolerance” in $\Delta\chi^2 = 9$ has been used to be roughly consistent with global fit procedure.

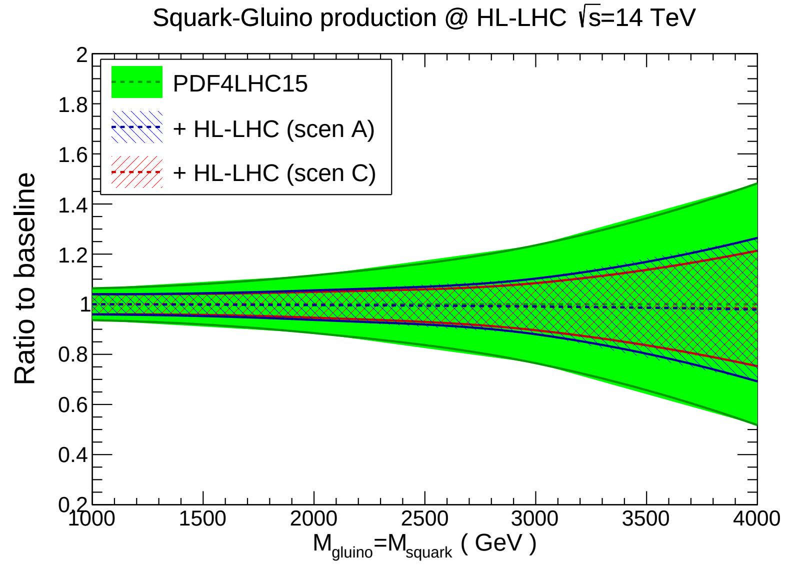
Smallest effect at low $x$ where HERA constraints remain dominant.




Impact on related quantities of parton luminosity functions.

Smallest effects at smaller $M_X$, i.e. small $x$ PDFs.



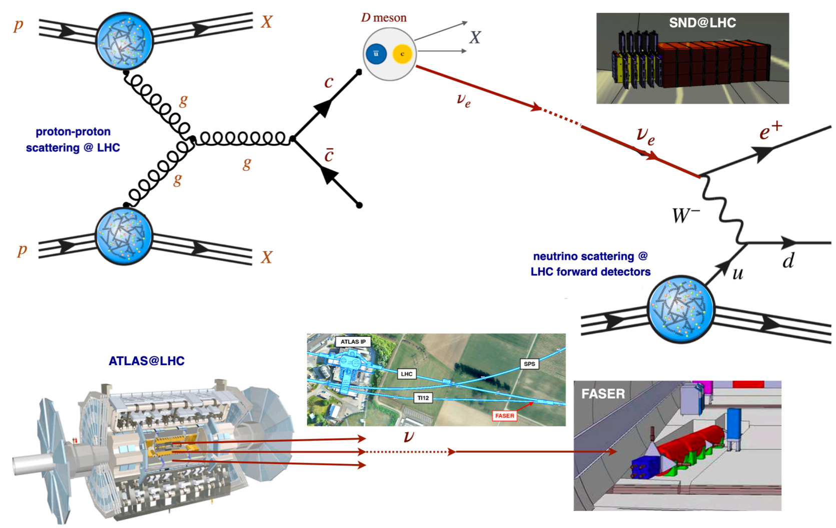




This translates into similar reductions on uncertainties for cross section predictions.

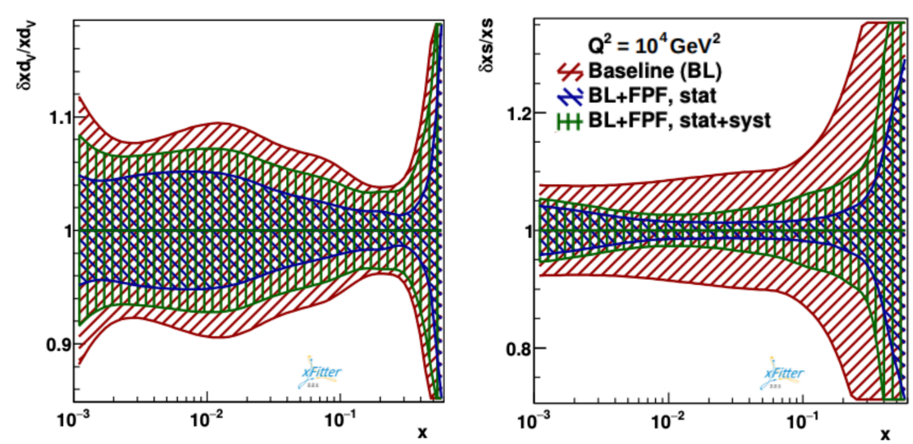
## Future Experiment Impacts – LHC -FPF - J. Rojo, DIS2024

The LHC also acts as a source of high energy neutrinos from decays of produced mesons.

Detected at e.g. FASER$\nu$, SND.




Mid-range energy ($10^2 - 10^3\mathrm{GeV}$) muon neutrinos from pion and kaon decay provide improved information of flavour separation in PDFs.



## Charm meson production probed by neutrino production.

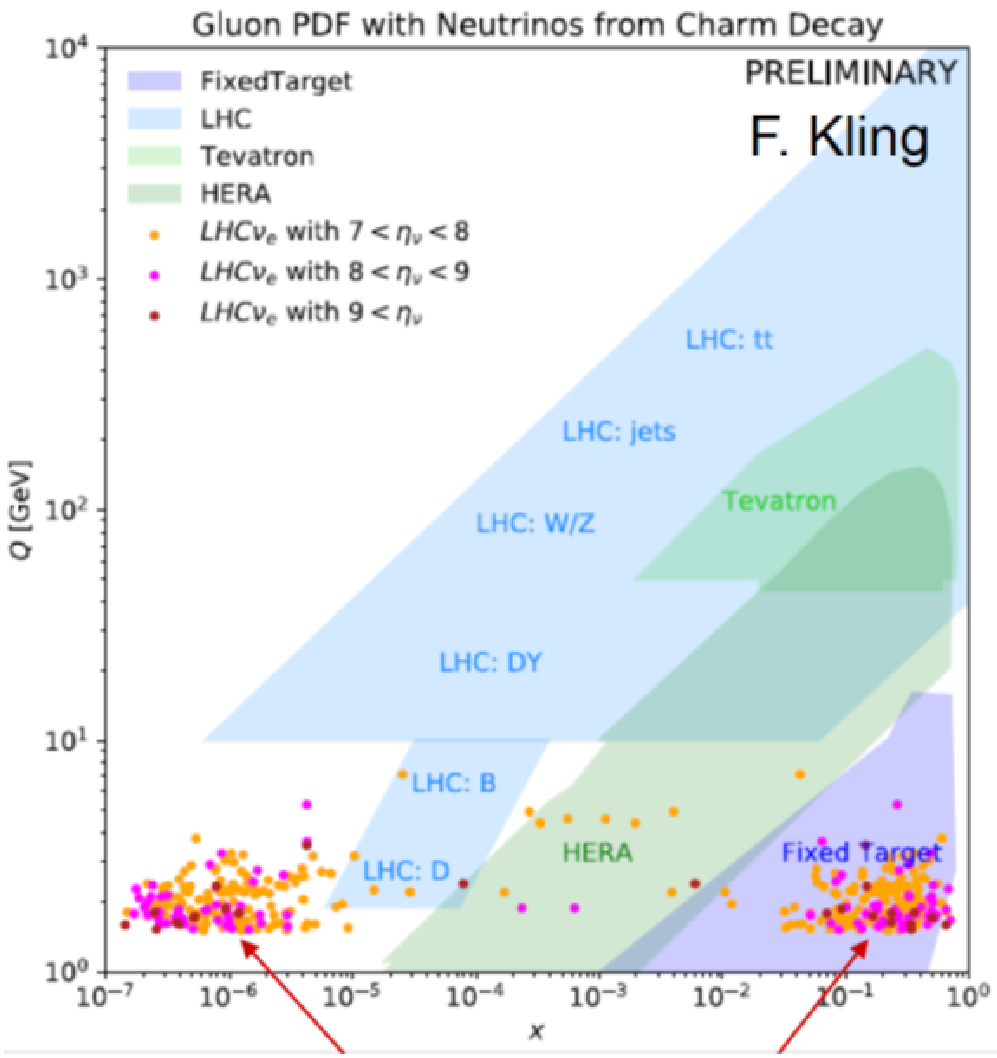
Charm mesons dominant source of neutrino flux, or more specifically interactions at SND.

Charm produced in LHC collisions at pseudo-rapidity $\sim 7-9$.

Corresponds to $x_1 \sim 0.04 - 0.8$ and $x_2 \sim 10^{-7} - 5 \times 10^{-5}$.

On the edge of LHCb lower limit.

Provides info. on small-$x$ physics and high-$x$ intrinsic charm.

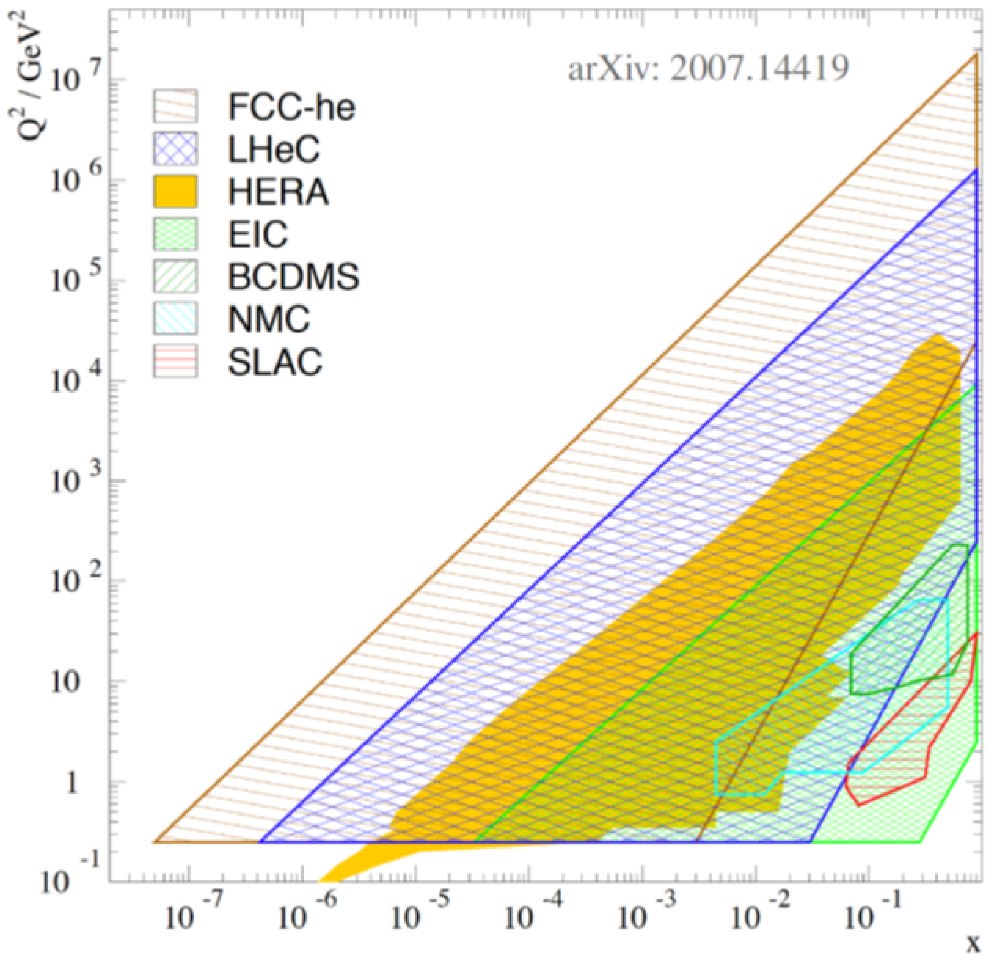


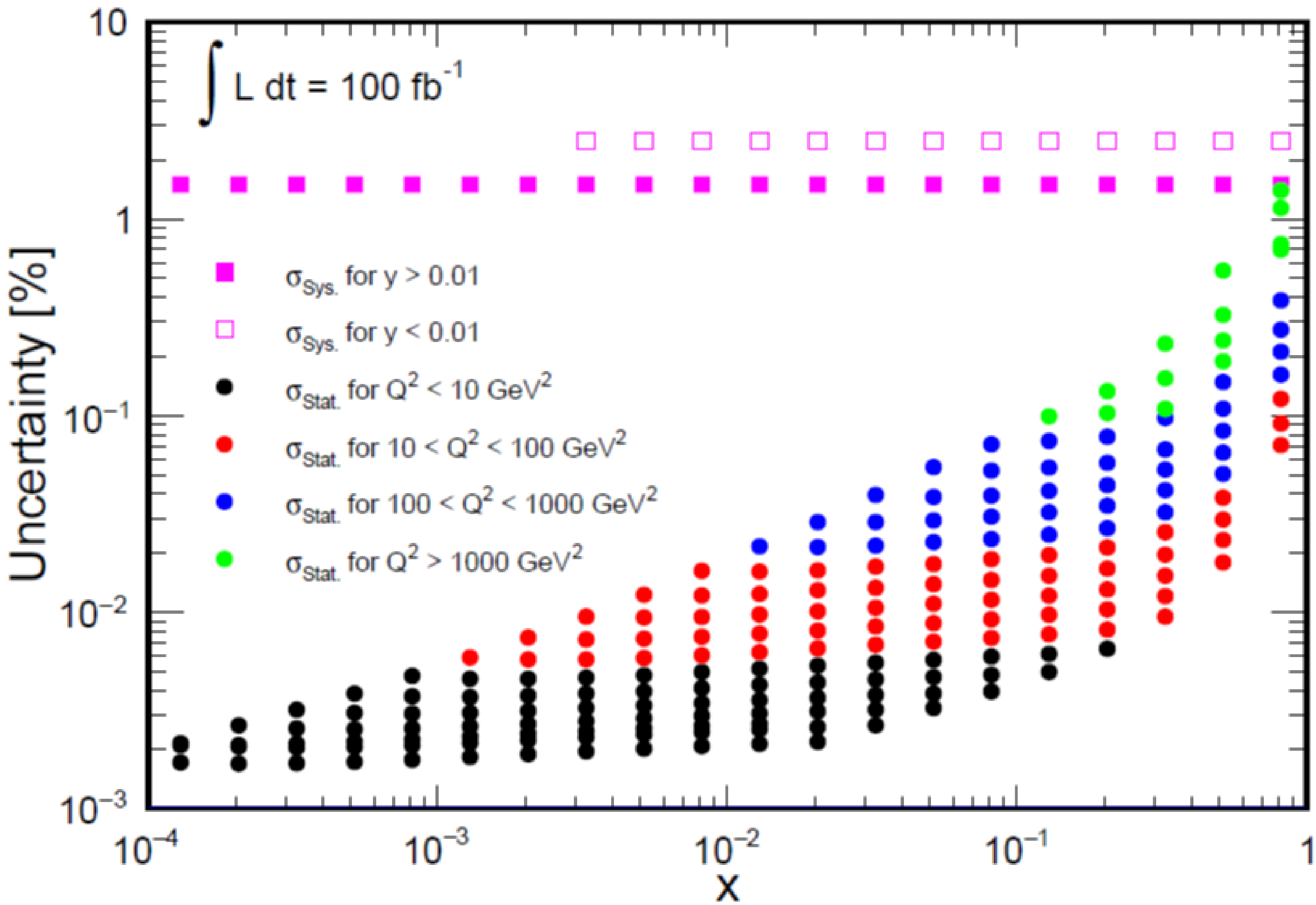
# EIC constraints on PDFs.

Kinematic range compared with other experiments providing constraints.

Much better precision than previous experiments at high $x$.

Not too low ($\leq 15\mathrm{GeV}^2$) $W^2$ – less higher twist contamination (compare with JLAB).



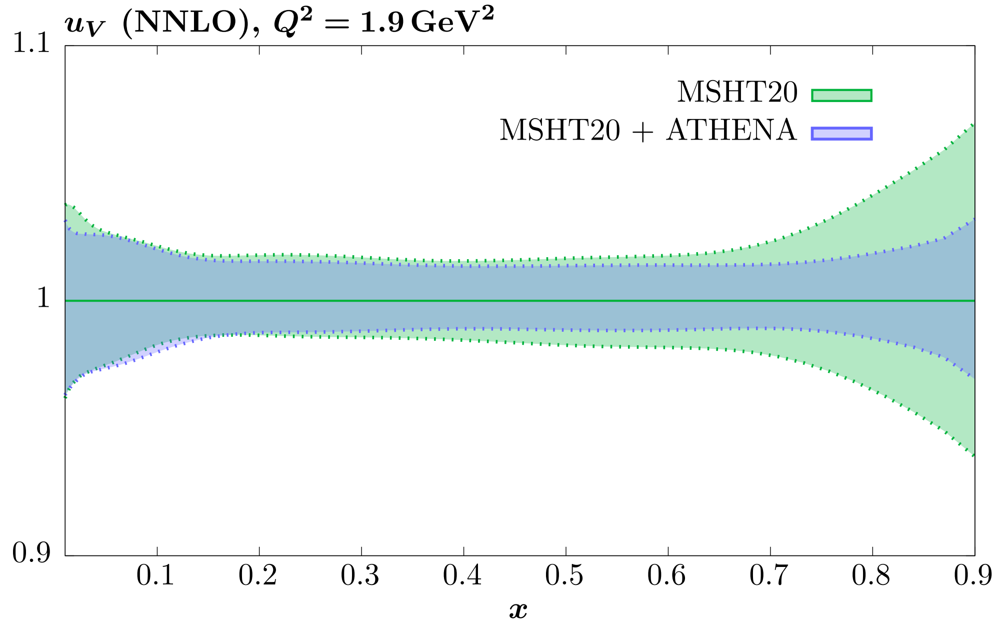



Dominated by systematic rather than statistical uncertainties. Latter projected to be much better than previous DIS experiments.

## Impact of EIC proton structure function data.

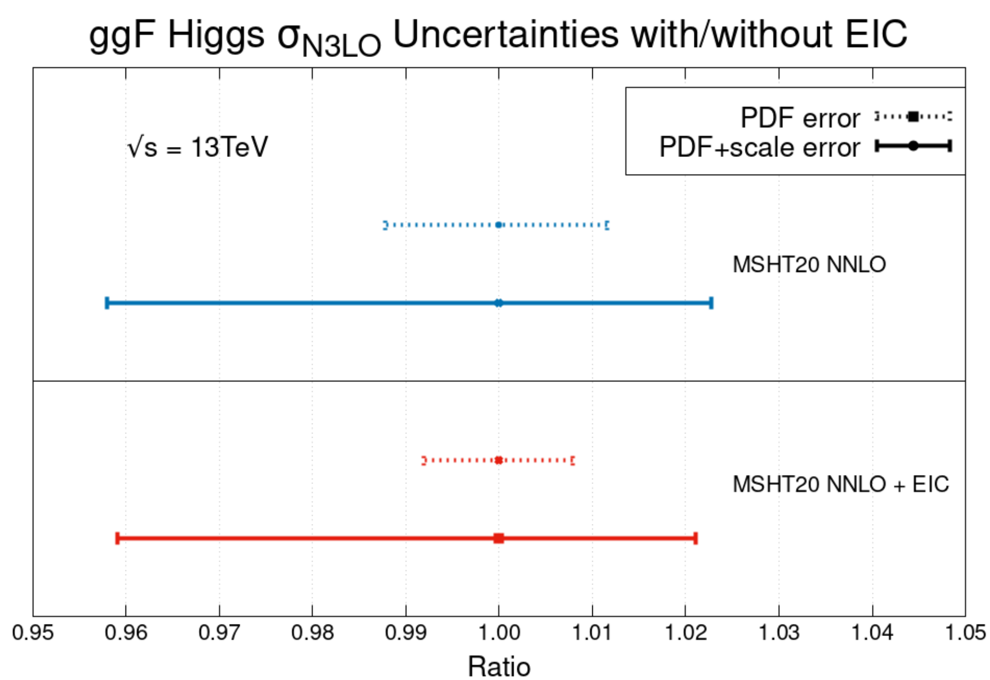
Consider this within a global fit (e.g. MSHT - impact clearly larger in DIS-only fit. 2309.11269.

Main impact on high-$x$ quarks, but some also on gluon distribution.

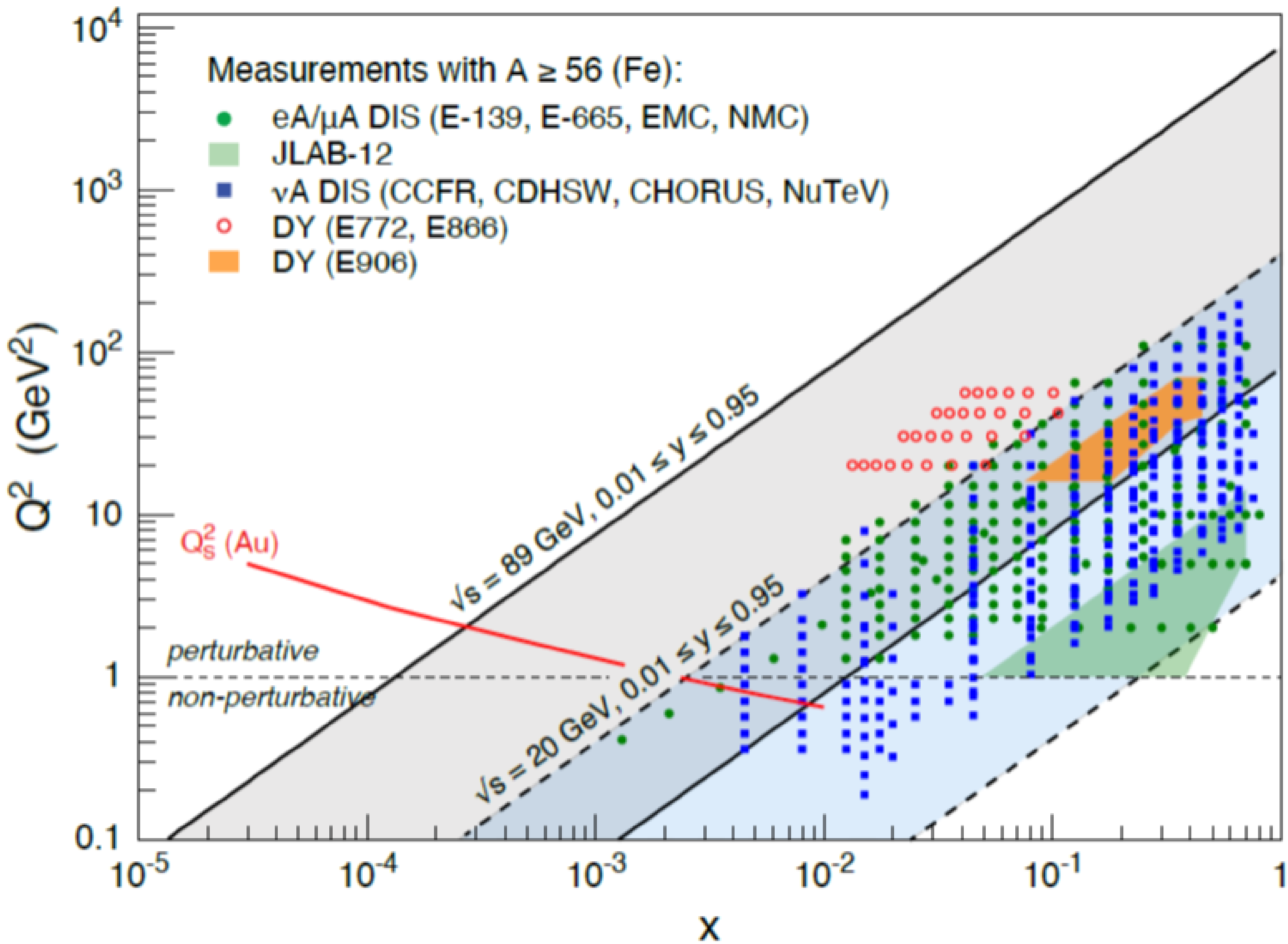



Has a noticeable effect on Higgs cross-section via gluon-gluon fusion at the LHC.

Extra improvements possible from deuteron data, especially with neutron tagging.

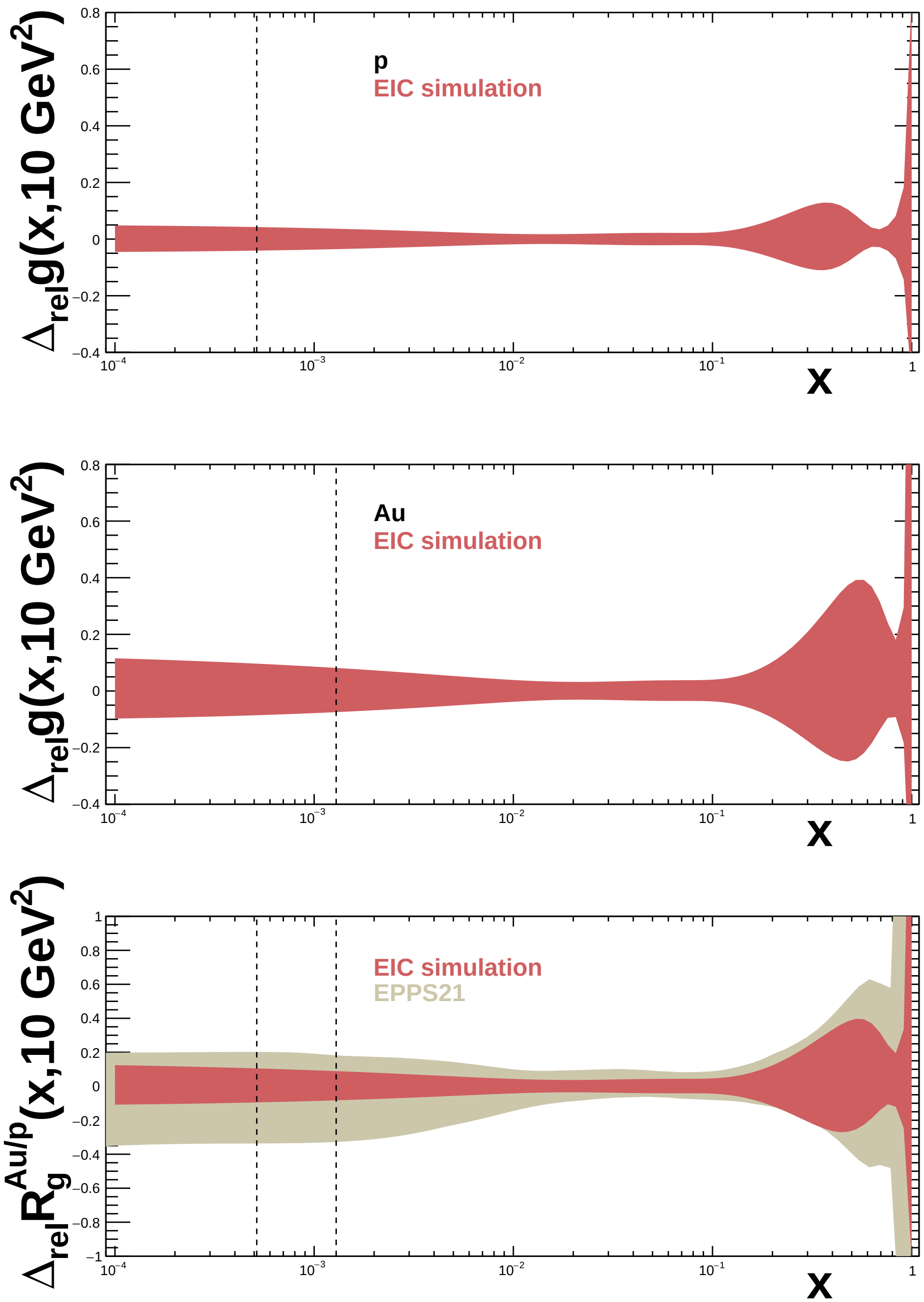


# Big impact on Nuclear PDFS



Test the nuclear corrections in much more detail, e.g. 2309.11269.

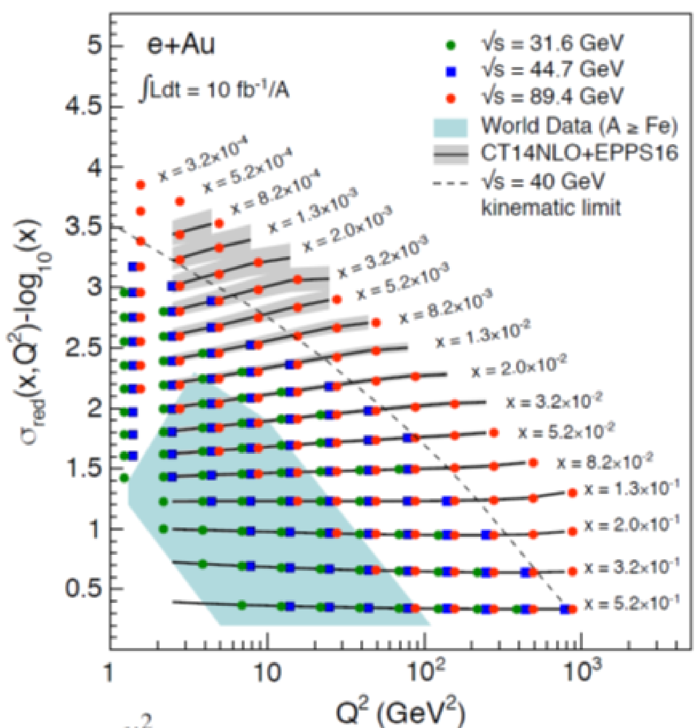
Significant reduction in gluon uncertainty at small $x$.

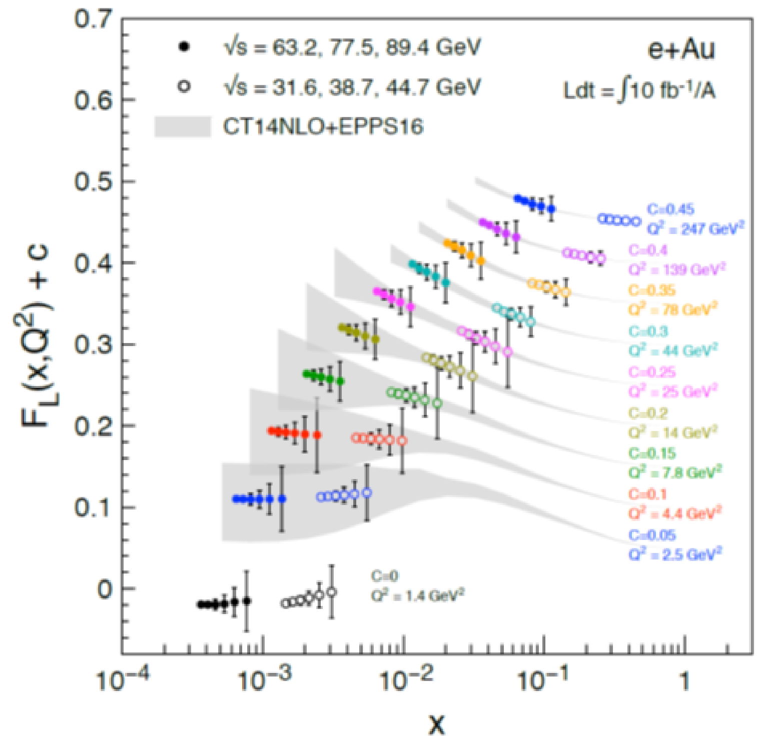


# Projection for possible EIC heavy flavour data.

- Inclusive eA scattering measurements

arXiv:1708.01527



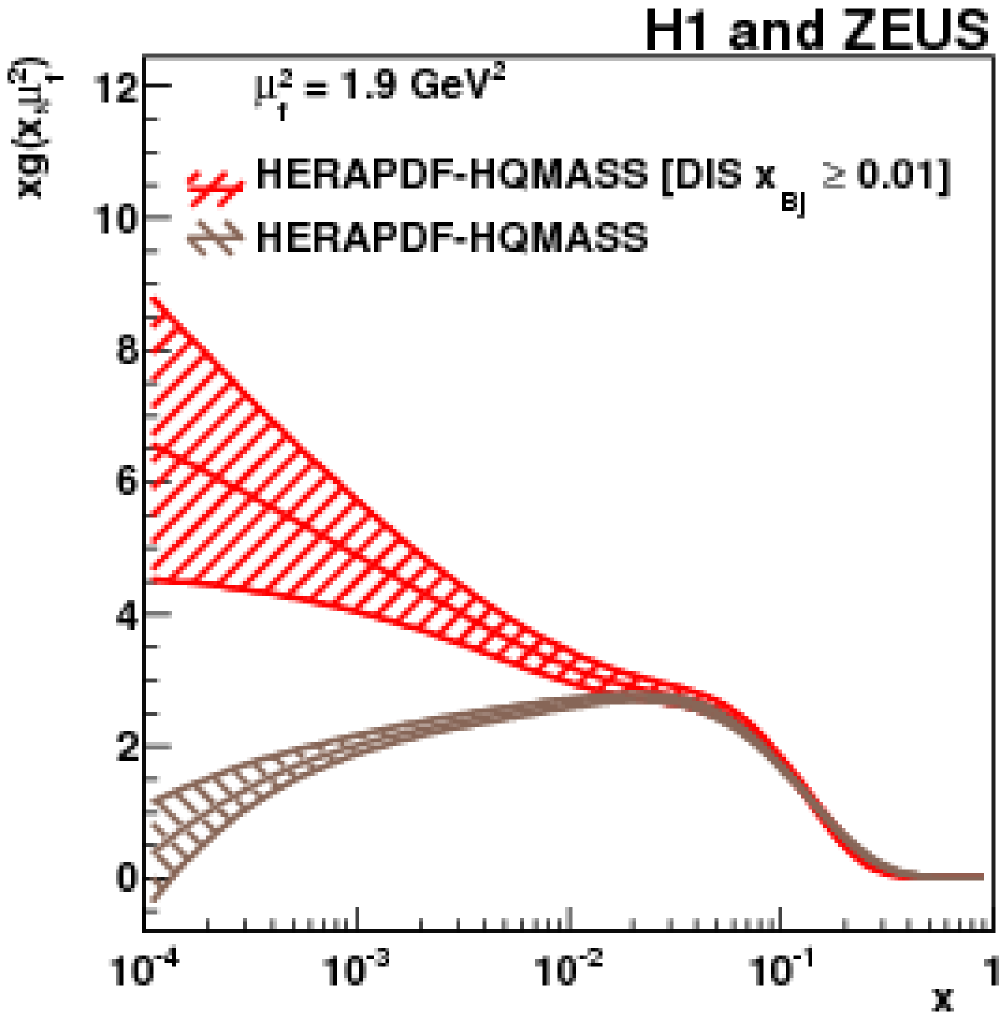




$$\sigma_{\rm red} = F_2 - \frac{y^2}{Y_+} F_L$$

$$\left(\frac{d^2\sigma}{dx\,dQ^2}\right) = \frac{2\pi\alpha^2 Y_+}{xQ^4}\left(F_2 - \frac{y^2}{Y_+}F_L\right)$$

Tensions between inclusive structure function data and heavy flavour.

Latter prefers steeper gluon – (EPJC 78 (2018) 6, 473).



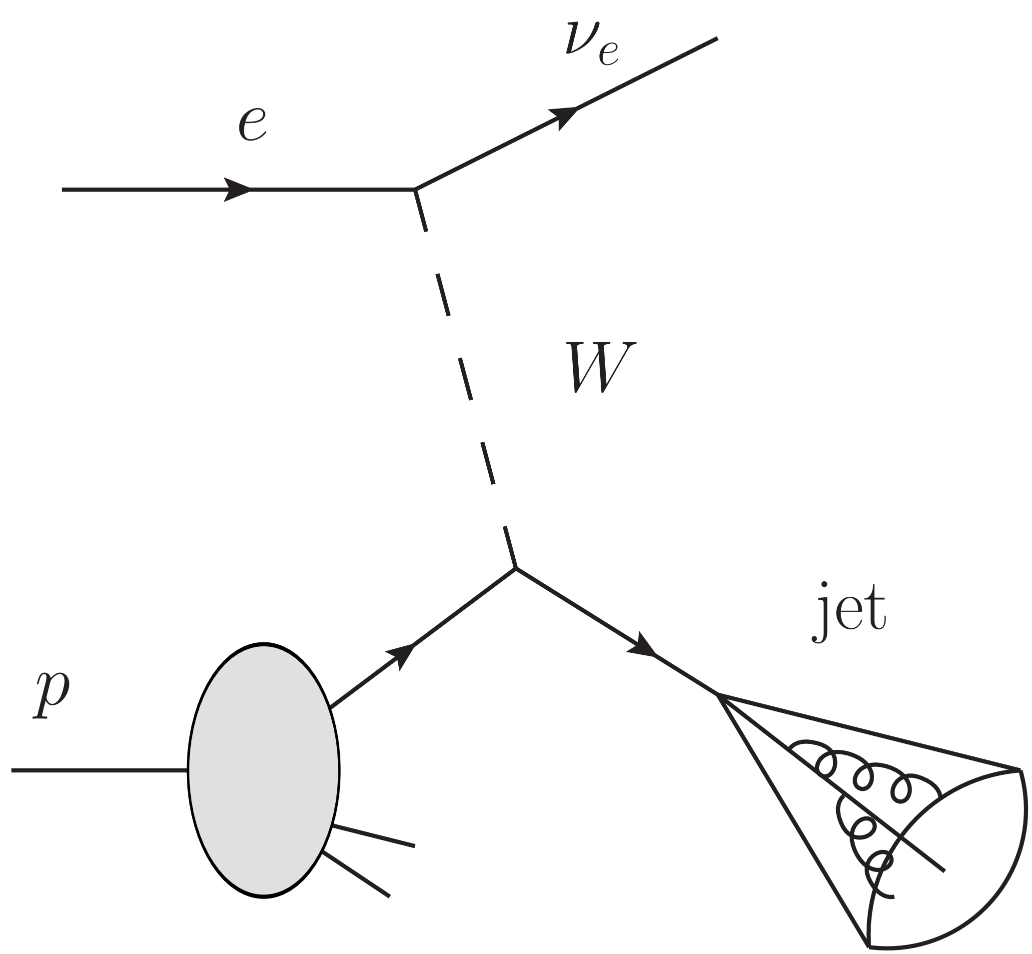

The EIC will get into this range where tension is seen.

## High-$x$ Strange Quark.

There is also the possibility of looking at the less well know strange quark via charm quark jets.

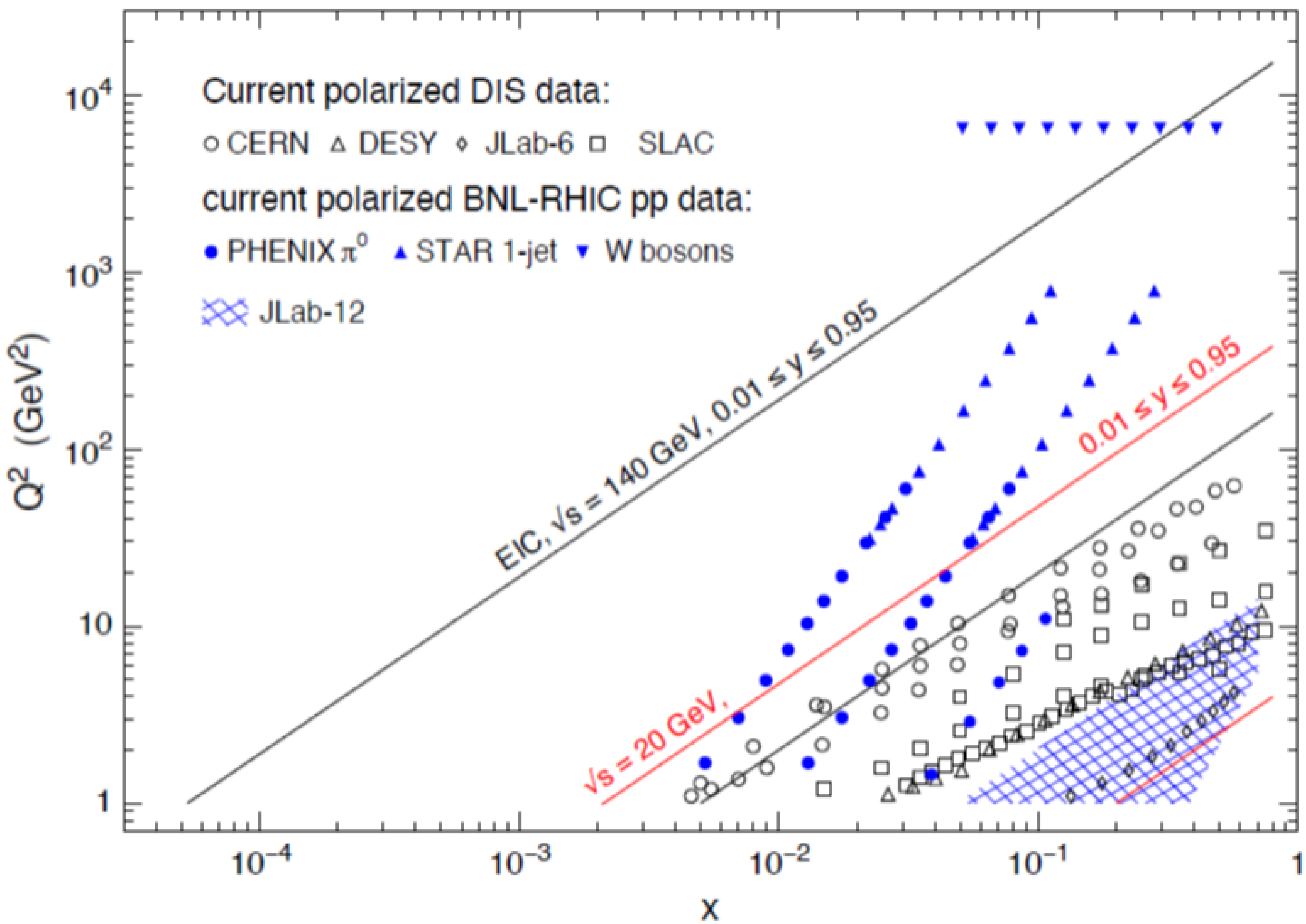
Requires dealing with fragmentation (but so do some current methods at some level).

Similar type of data from neutrino scattering on iron targets from CCFR/NuTeV already used.



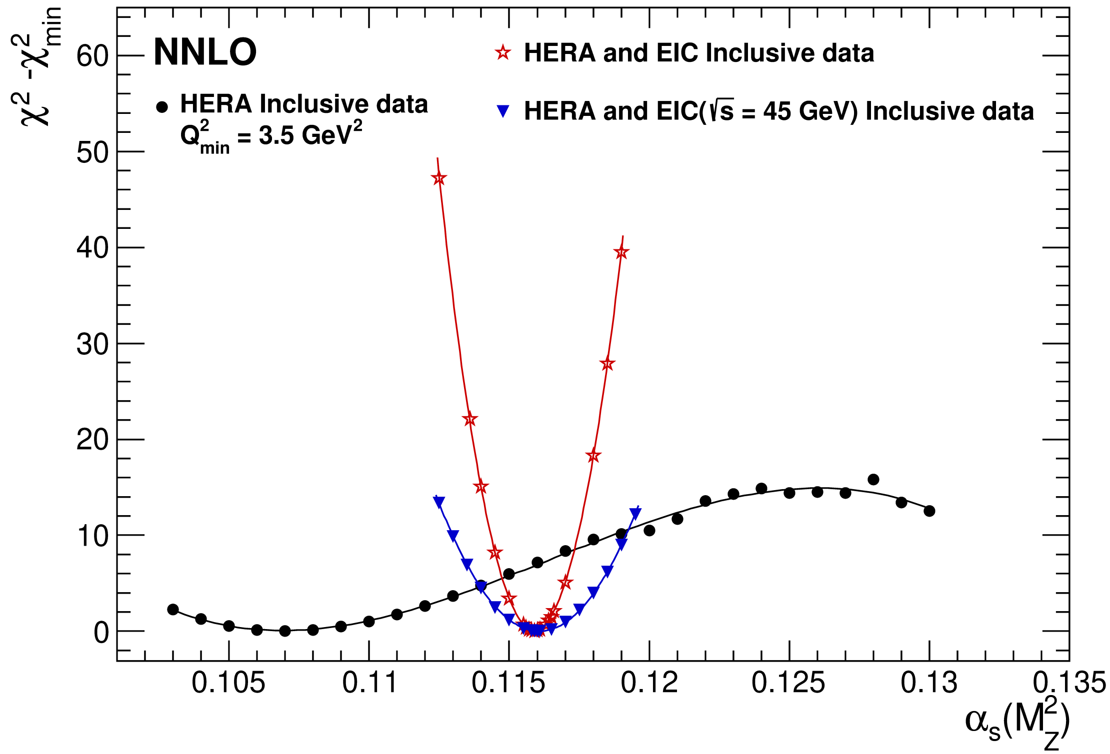

Plot from https://arxiv.org/abs/2006.12520.

## Polarized PDFs.




Very significant expansion of coverage to small $x$, as well as higher $Q^2$ at high $x$. $(\partial g_1(x,Q^2)/\partial \ln Q^2) \sim -\Delta g(x,Q^2)$. Significant impact on PDFs and spin sum rule.

# $\alpha_S(M_Z^2)$ determinations from structure functions.

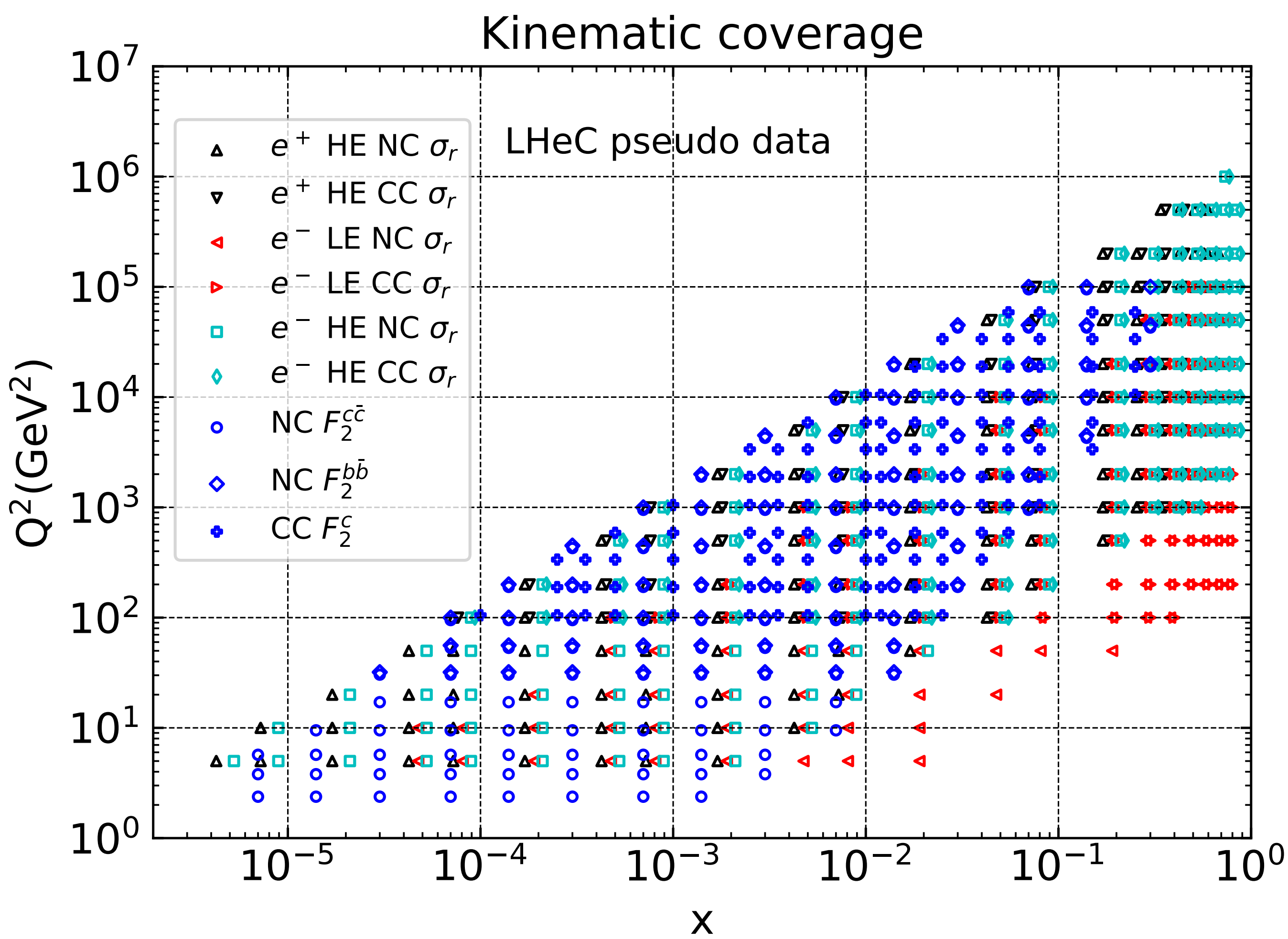



High-$x$, not too low $Q^2, W^2$ data very clean probe of $\alpha_S(M_Z^2)$ via non-singlet structure function evolution.

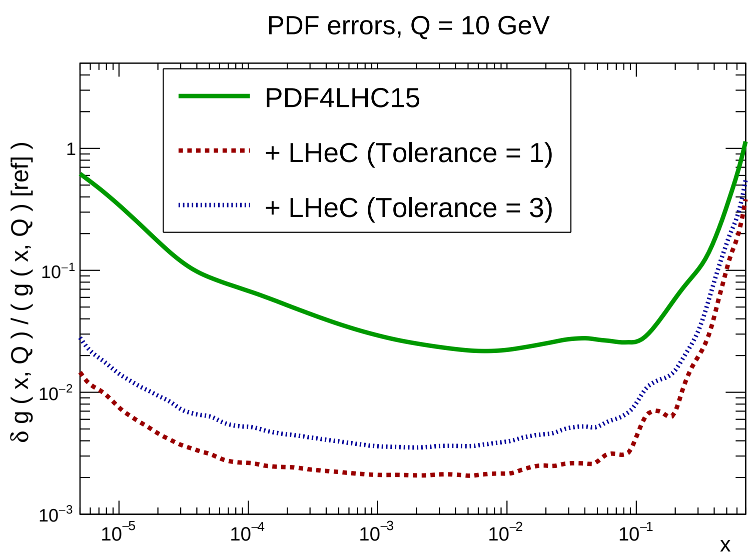
EIC data can improve constraints dramatically 2307.01183.

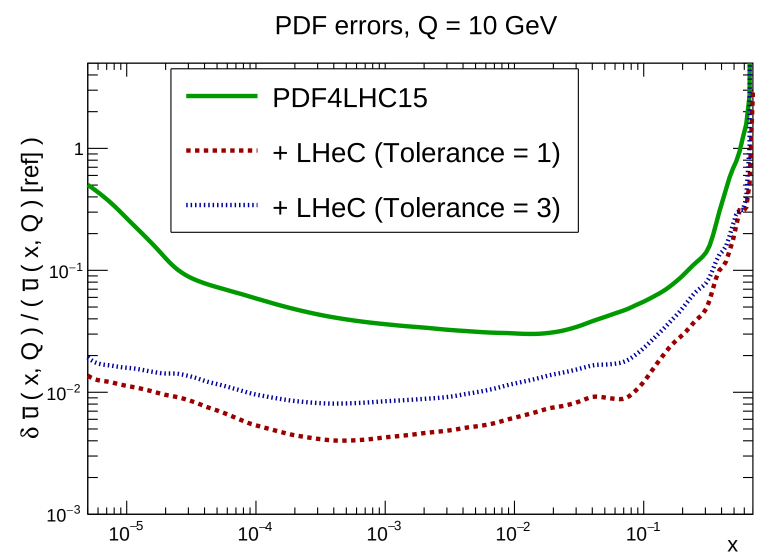
Need also to consider theoretical/methodological uncertainties.

# Future Experiment Impacts – LHeC.



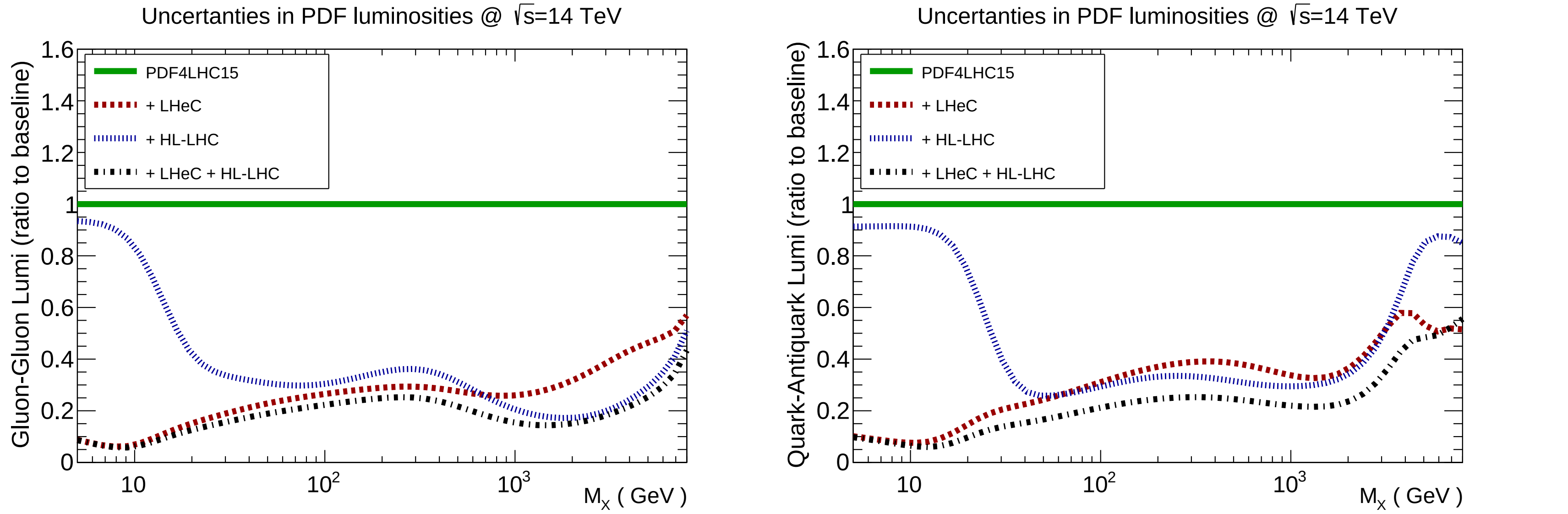

Example, impact of pseudodata shown above considered in 1906.10127.




Very considerable impact on PDFs seen, particularly at small $x$ (contrast with HL-LHC).

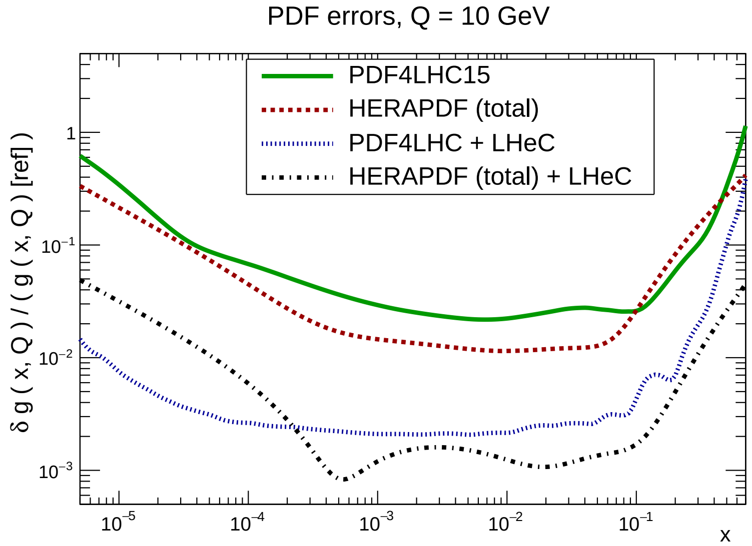
Impact of taking "tolerance" $\Delta\chi^2 = 9$ for consistency with global PDF fit compared to $\Delta\chi^2 = 1$ - not a naive factor of 3 increase in uncertainty.

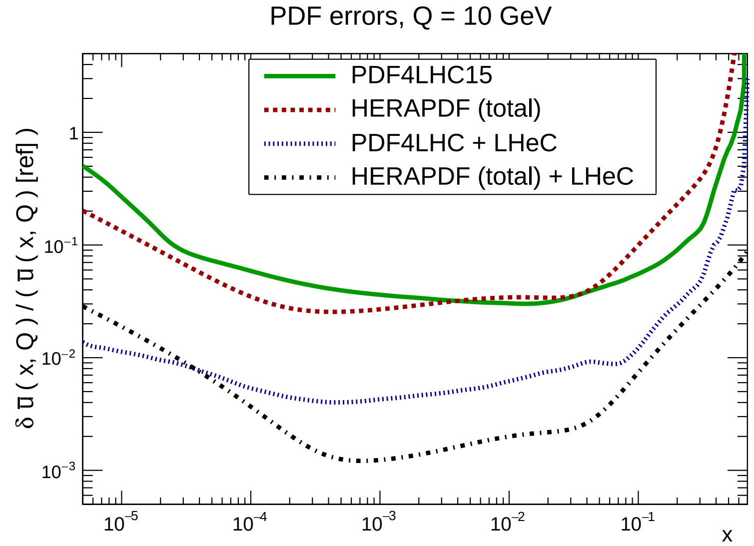



Impact of just LHeC pseudodata, just HL-LHC or both.

Clearly different regions where one or the other dominates.

Excellent improvement in a very wide range of places when both can be applied.




Note PDF parameterization must allow the same level of precision as that achieved by the constraint in order for results to be reliable.

Potentially the constraint can be limited by the parameterization not the data, more so as the data precision improves.

Using exactly same procedure the improvement at high $x$ for simple version of HERAPDF with 14 parameters is much greater than for 100 eigenvector PDF4LHC.

Much more similar at very small $x$.

## Non-collider experiments.

PDFs matter for neutrino experiments, not only at ultra high-energy.

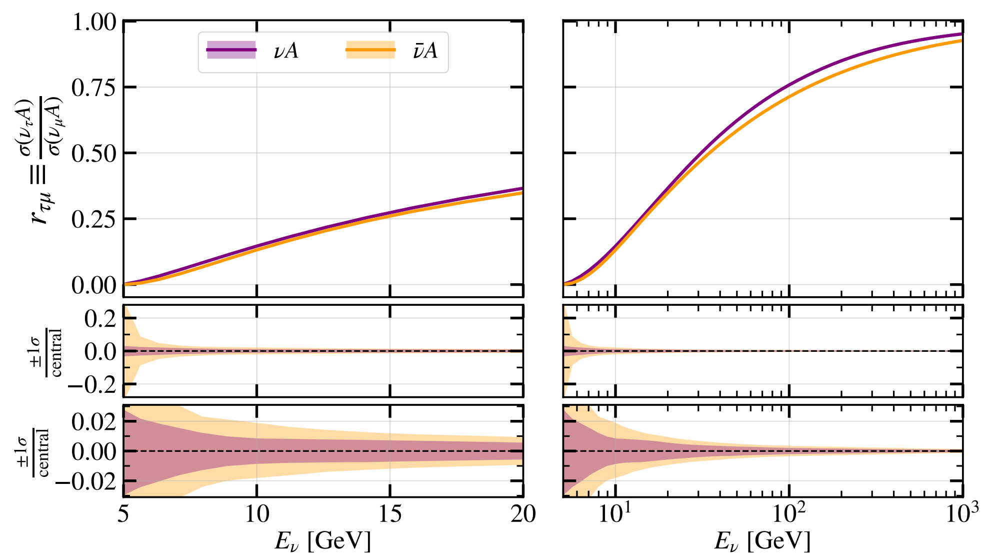
See e.g. 2409.01258. Cross sections depend on different combinations of structure functions due to different lepton masses.

Particularly for $\nu_\tau$.

$$\begin{aligned}\frac{\mathrm{d}^2\sigma^{\nu_\ell N(\bar\nu_\ell N)}}{\mathrm{d}x\mathrm{d}y} = \frac{G_F^2 M_N E_\nu}{\pi(1+Q^2/M_W^2)^2}\Bigg\{ &\left(y^2 x + \frac{m_\ell^2 y}{2E_\nu M_N}\right) F_1 \\ &+ \left[\left(1-\frac{m_\ell^2}{4E_\nu^2}\right) - \left(1+\frac{M_N x}{2E_\nu}\right) y\right] F_2 \\ &\pm \left[xy\left(1-\frac{y}{2}\right) - \frac{m_\ell^2 y}{4E_\nu M_N}\right] F_3 \qquad (1) \\ &+ \frac{m_\ell^2(m_\ell^2+Q^2)}{4E_\nu^2 M_N^2 x} F_4 - \frac{m_\ell^2}{E_\nu M_N} F_5 \Bigg\}.\end{aligned}$$

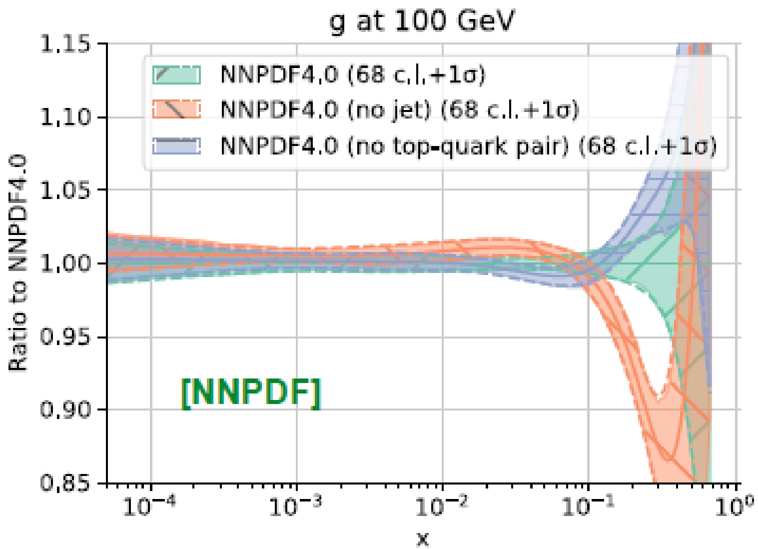
Correlations between cross sections then depends on PDFs.

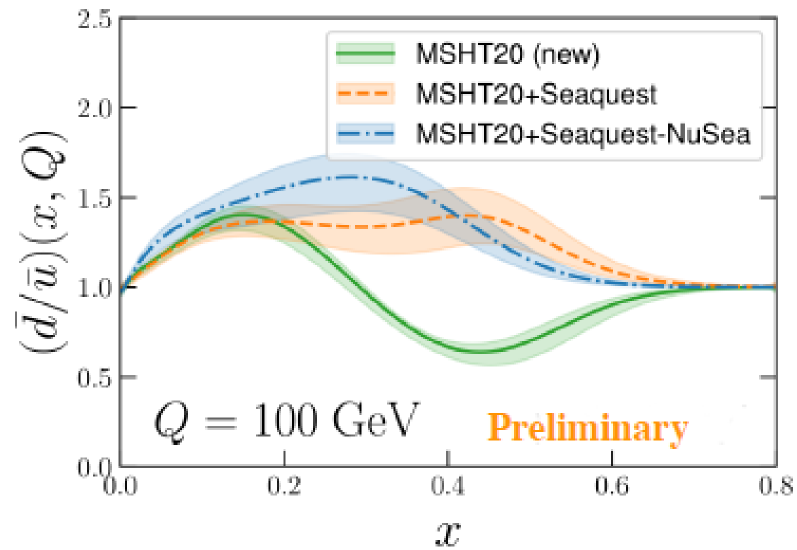
Uncertainty of $\sigma_{\nu_\tau}/\sigma_{\nu_\mu}$ at e.g. IceCube-DeepCore, KM3NET or DUNE, shown.



## Challenges - Improving/understanding Methodology.

### Tensions

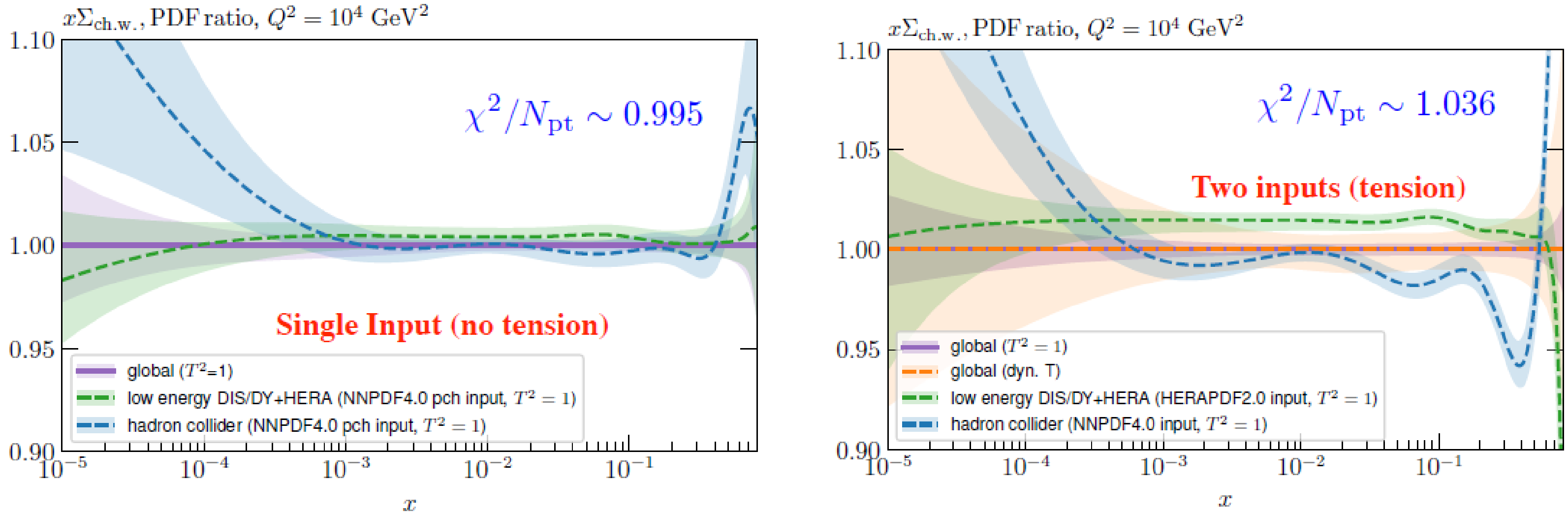



Clearly tensions on pulls of PDFs from different data sets.

Sometimes different types of data, e.g. LHC jets and top data seems to prefer different high-$x$ gluons - could be for variety of reasons.

Sometimes from identical types of data, e.g. Seaquest and Nusea Drell-Yan asymmetry - less scope for reasons.

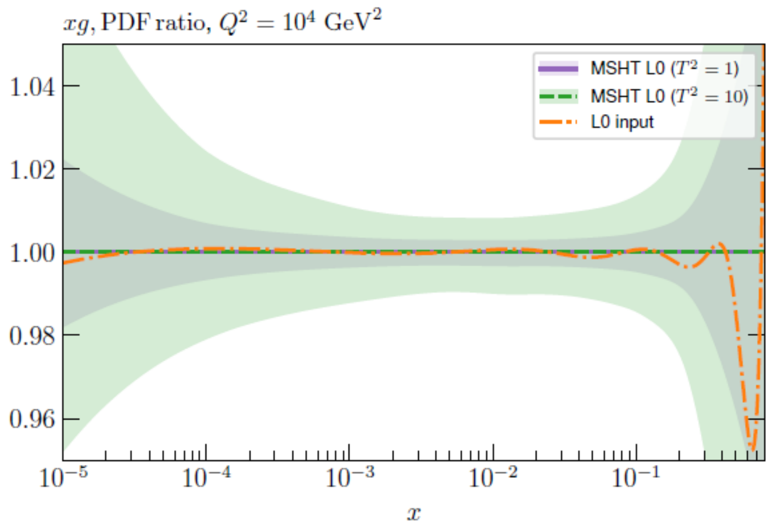


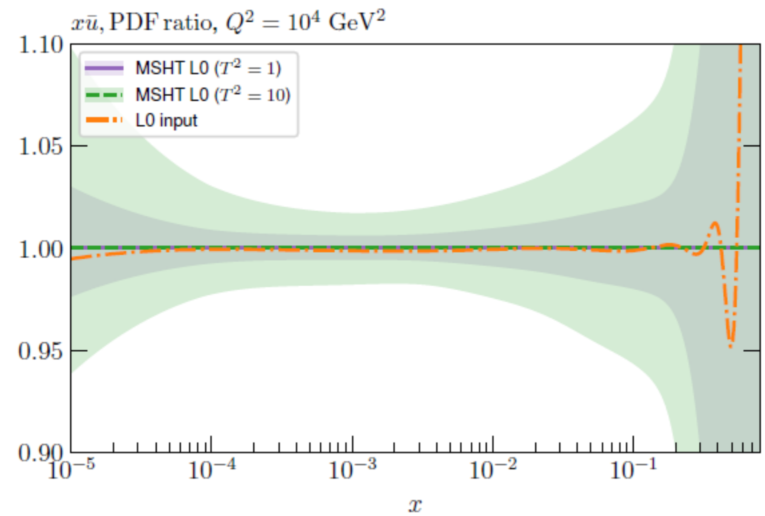

Look at effect when generating purposefully inconsistent sets of data 2407.07944.

Inconsistent data does not lead to an increase in PDF uncertainty using conventional treatments.

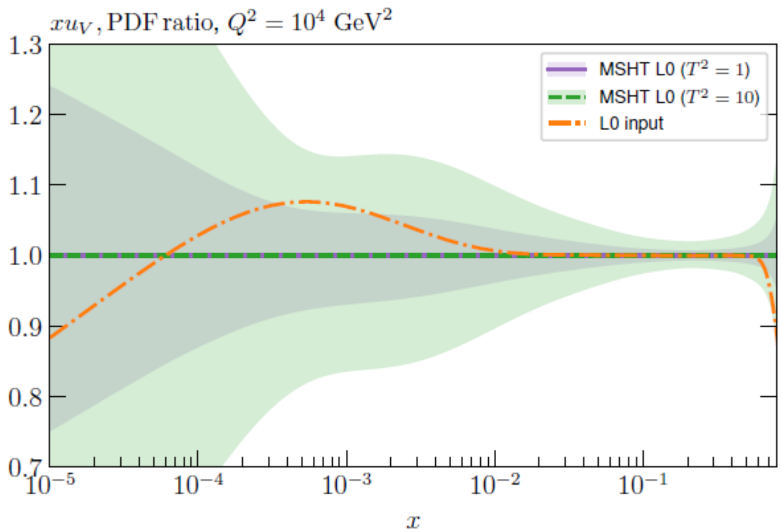
Clear inconsistencies in pulls from different data types, but global fit $\chi^2$ is not “that bad”.

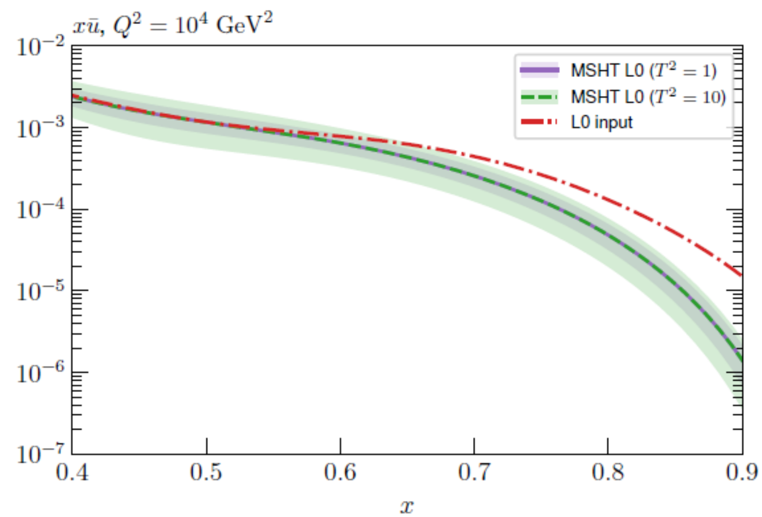
# Issues of Parameterization limitations – Closure Test.



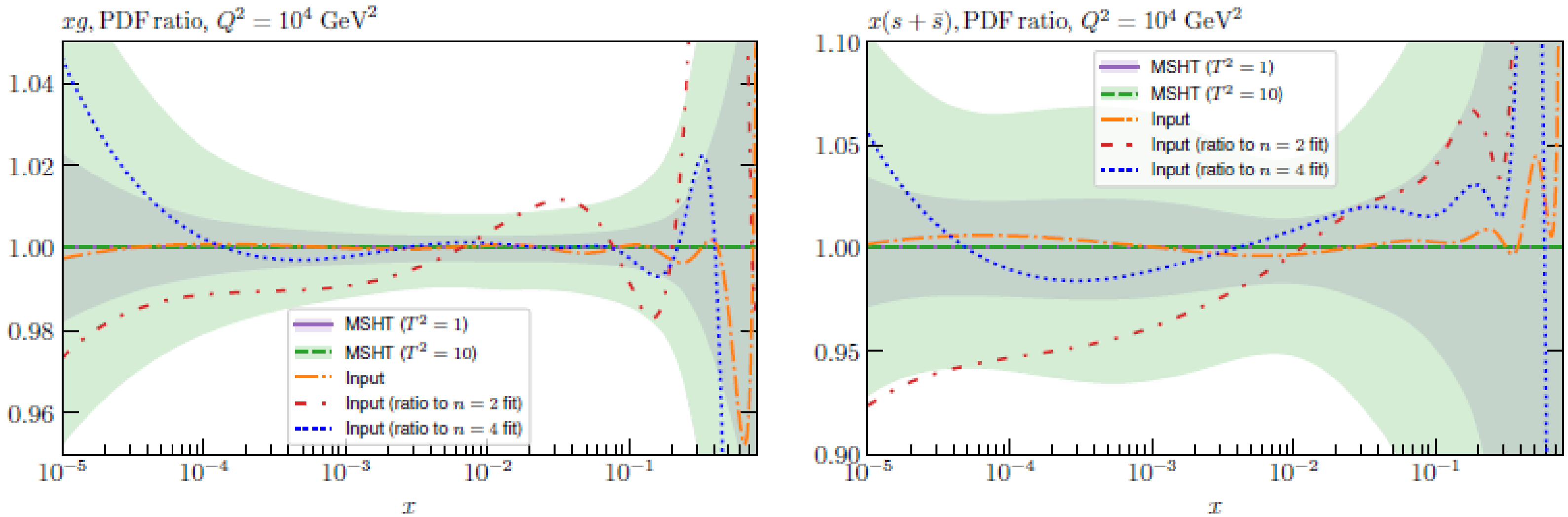

- Ratio of (NNPDF4.0pch) L0 input to fit result, including PDF uncertainties with $T^2 = 1$ and 10 that come from the closure test fit. Latter is ~ result of dynamic tolerance used in MSHT20 (checked here).




- In less well constrained regions deviation larger, e.g for $u_V, d_V$ at low and high $x$ and the $\bar{u}, \bar{d}$ at high $x$ .
- Hence in extrapolation region L0 input not always consistent within uncertainties

Relies of having enough parameters (and a suitably well-chosen form of parameterization) to be successful for a given number of PDFs over a given range in $x$ and at a certain level of precision.



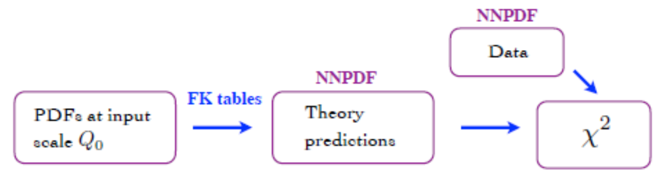

Limitations from less flexible parametrizations can lead to inaccuracies in PDFs consistent with $\geq \Delta\chi^2 = 10$ over a range of PDFs.

**Note** – CT justify part of their (larger "tolerance") by comparing variation between parameter choices in their PDF fits.

# Methodology uncertainties - compare extremely like-to-like fits.

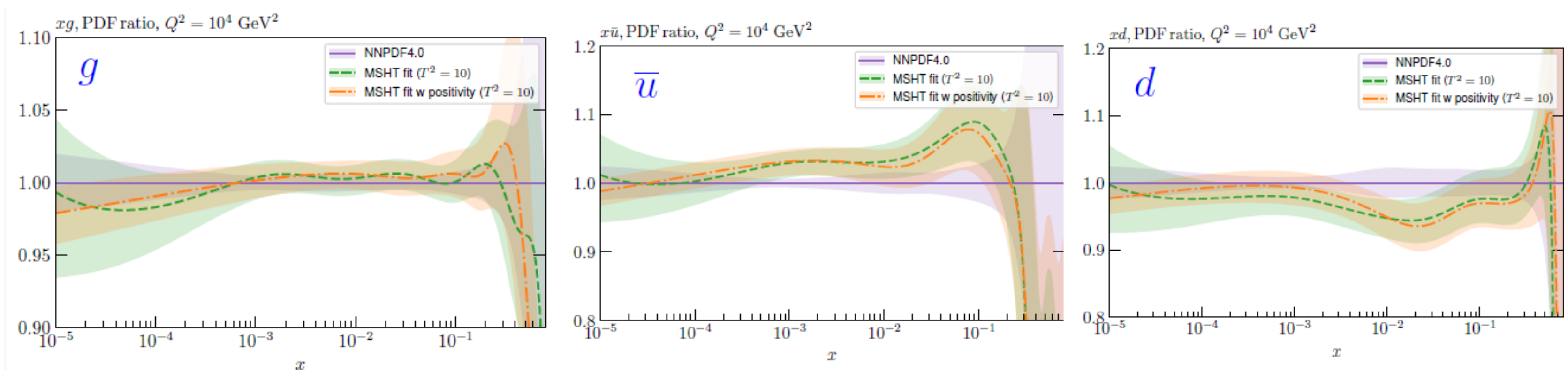
- Readily extend previous study to include **fitted charm**. NNPDF theory inputs change accordingly, while PDFs parametrised at $Q_0 = 1.65\,\mathrm{GeV}$ ($> m_c$) rather than 1 GeV, and parameterise charm:

$$xc_+(x,Q_0) = A_{c_+} x^{\delta_{c+}} (1-x)^{\eta_{c+}} \left(1 + \sum_{i=1}^{6} a_{c,i} T_i(y(x))\right)$$
$$xc_-(x,Q_0) = 0$$




- Find:

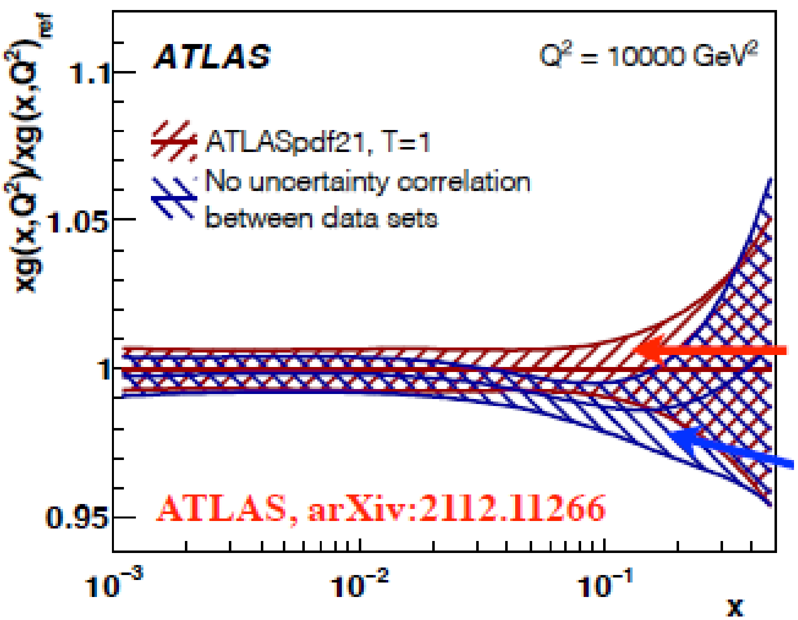
| | NNPDF4.0 | MSHT fit | MSHT fit (w positivity) |
|---|---|---|---|
| $\chi^2_{t0}$ | 5692.1 (1.233) | 5645.2 (1.222) | 5651.0 (1.224) |
| $\Delta\chi^2_{t0}$ : | | −46.9 (0.011) | −41.1 (0.009) |

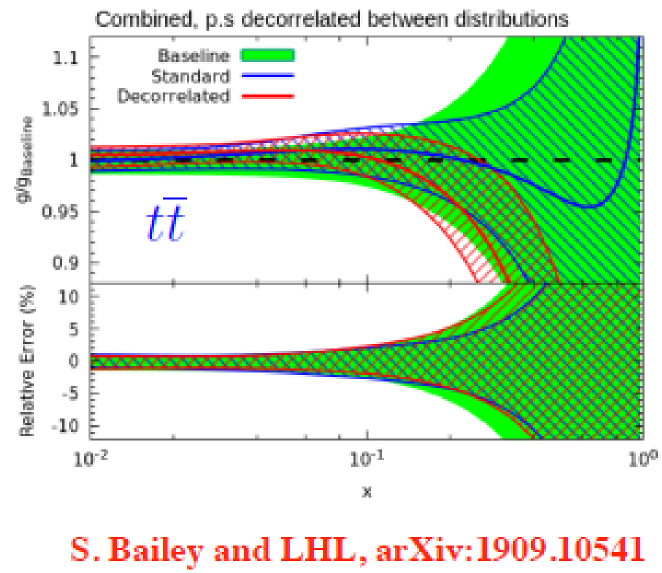



Clearly differences of order the uncertainties. Not always clear why.

## PDF changes due to treatment of (correlated) systematics

Issues related to correlations between data sets and due to correlations within data sets.




| $n_{dat}$ | default | stat. uncorr. | p.s. uncorr |
|---|---|---|---|
| 25 | 7.00 | 3.28 | 1.80 |

Systematics related to variations in Monte Carlo generators likely to not fully represent correlations or absolute size of the uncertainty.

Potential issues due to lack of complete knowledge of what the error is.

Need to consider the "error on the error".

- Model the estimated variances, $v_i$, of $\sigma^2_{u_i}$, as Gamma distributed gives:

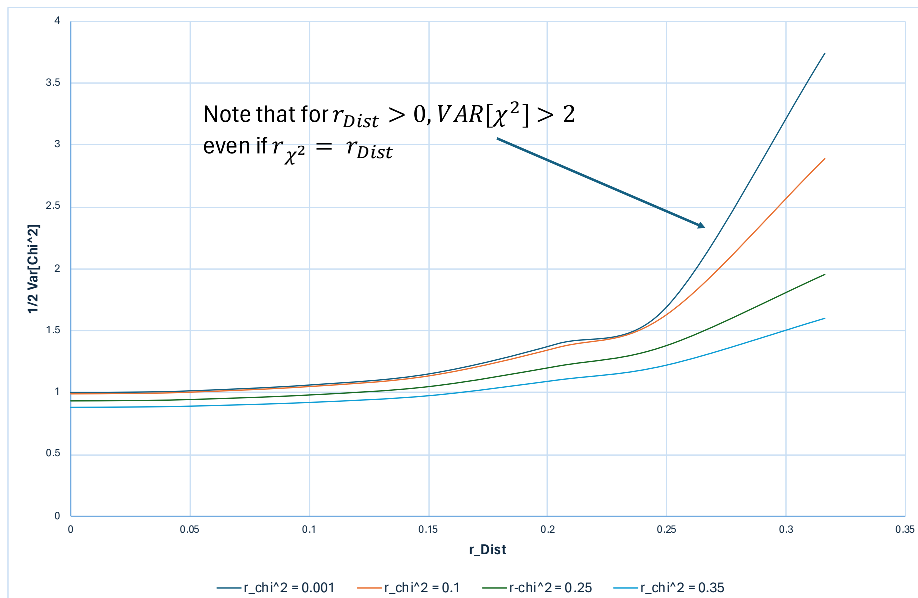
$$L(\mu, \theta, \sigma^2_{u_i}) = P(y|\mu, \theta) \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma_{u_i}} e^{-(u_i-\theta_i)^2/2\sigma^2_{u_i}} \frac{\beta_i^{\alpha_i}}{\Gamma(\alpha_i)} v_i^{\alpha_i - 1} e^{-\beta_i v_i}$$

$$\alpha_i = \frac{1}{4r_i^2} \qquad \beta_i = \frac{1}{4r_i^2 \sigma^2_{u_i}}$$

- $r_i$ is defined as the relative uncertainty in the estimate of the systematic error. The parameters $r_i$ can therefore be referred to as the "error on errors".

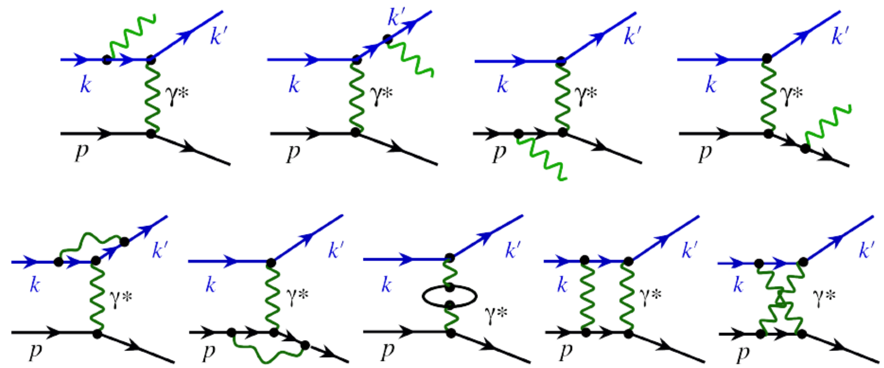
[2] Cowan arXiv:1809.05778v3

Variance increases if $r_i > 0$ especially (but not only) if $\chi^2$ calculation not modifies to acknowledge this, e.g. 2408.12922.



## Taking Account of Lepton Sector at EIC - example of complicated systematic.

Take properly into account photon radiation from both the lepton and the quarks 2408.08377.

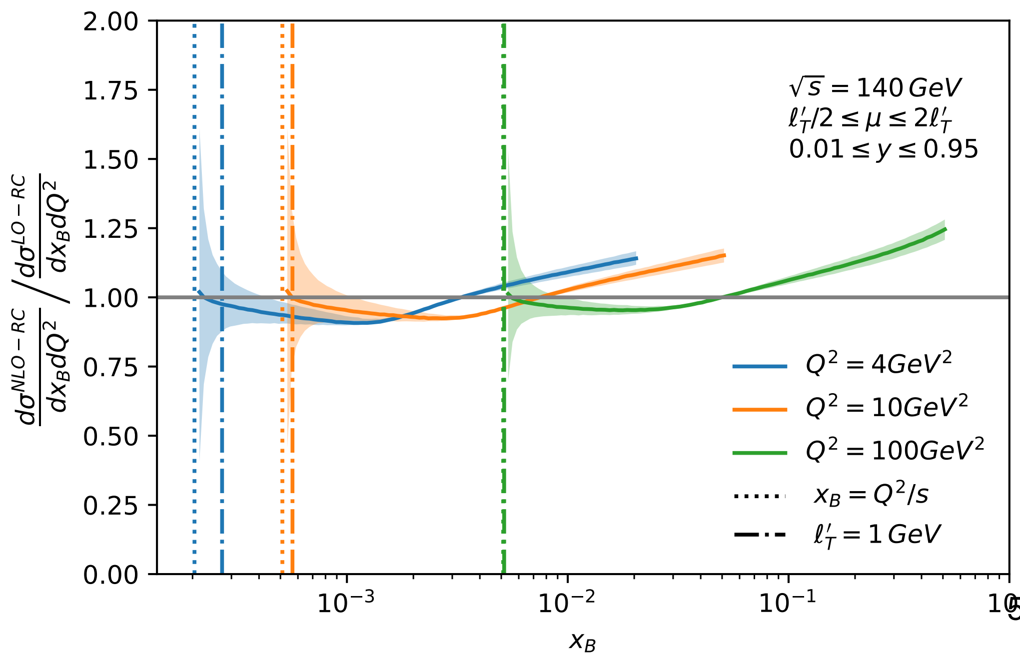



Strictly, need to account for lepton distributions functions and fragmentation.

$$E'\frac{d\sigma_{\ell P\to\ell' X}}{d^3\ell'} \approx \frac{1}{2s}\sum_{ija}\int_{\zeta_{\min}}^{1}\frac{d\zeta}{\zeta^2}D_{e/j}(\zeta,\mu^2)\int_{\xi_{\min}}^{1}\frac{d\xi}{\xi}f_{i/e}(\xi,\mu^2) \times \int_{x_{\min}}^{1}\frac{dx}{x}f_{a/h}(x,\mu^2)\,\hat{H}_{ia\to jX}(\xi\ell, xP, \ell'/\zeta,\mu^2) + O\left(1/\ell_T'\right)^{\alpha}$$

Simpler perturbative models currently used at e.g. HERA.

Modifies the kinematics and can affect the cross section at specified points.



## Summary.

PDFs fully (in practice) established at NNLO, and are approaching uncertainties of $1\%$ in many regions - flavour decomposition not at this level in general.

Both improving theory, e.g. (reasonably) complete N$^3$LO and future experiments can drive us to sub percent accuracy over most PDF flavours and kinematic regions. **But ....**

– Require very wide range of experiments to achieve this, particularly if some constraints on PDFs are in the very high energy limit.

– We are at the point where all methodological uncertainties in PDF determination need to be understood and quantified better.

– This interacts with experiment - uncertainties need to be presented in a comprehensive and genuinely realistic manner.

– Vital to understand variation in not only PDF central values but uncertainties in order to both understand QCD with precision and to utilize PDFs in other studies - e.g. CMS inflating PDF uncertainties by varying amounts for $M_W$ determination.

# References


[1] - S. Bailey et. al., Eur. Phys. J. C 81 341 (2021).

[2] - J. Vermaseren, A. Vogt, and S. Moch, Nuclear Physics B, 724, 3182 (2005)

[3] - S. Moch, B. Ruijl, T. Ueda, J. A. M. Vermaseren, and A. Vogt, Journal of High Energy, 1653, Physics, 2017, (2017)

[4] - A. Vogt et al., PoS LL2018, 050 (2018), 1808.08981

[5] - S. Moch, B. Ruijl, T. Ueda, J. A. M. Vermaseren, and A. Vogt, (2021), 2111.15561

[6] - S. Moch, B. Ruijl, T. Ueda, J. A. M. Vermaseren, and A. Vogt, Journal of High Energy, 1664, Physics, 2017, (2017)

[7] - I. Bierenbaum, J. Blumlein, and S. Klein, Nuclear Physics B, 820, 417 (2009)

[8] - M. Bonvini and S. Marzani, Journal of High Energy Physics, 2018, (2018)

[9] - J. Ablinger et al., Nucl. Phys. B, 886, 733 (2014), 1406.4654.

10] - J. Ablinger et al., Nuclear Physics B, 890, 48151 (2015)

[11] - J. Ablinger et al., Nuclear Physics B, 882, 263288 (2014)

[12] - H. Kawamura, N. A. Lo Presti, S. Moch, and A. Vogt, Nucl. Phys. B, 864, 399 (2012),1689

[13] - J. Blumlein et al., PoS, QCDEV2017, 031 (2017), 1711.07957

[14] - G. Falcioni, F. Herzog, S. Moch and A. Vogt, Phys. Lett. B 842 (2023).

[15] - G. Falcioni, F. Herzog, S. Moch and A. Vogt, Phys. Lett. B 846 (2023).

[16] - S. Moch, B. Ruijl, T. Ueda, J. Vermaseren and A. Vogt.

[17] - J. Ablinger et al., JHEP 12 (2022) 134, 2211.05462

[18] - J. Ablinger, et al., 2403.00513.

[19] - C. Duhr, F. Dulat and B. Mistlberger, JHEP 11 (2020), 143.

[20] - C. Duhr and B. Mistlberger, JHEP 03 (2022), 116.

[21] X. Chen, T. Gehrmann, N. Glover, A. Huss, T. Z. Yang and H. X. Zhu, Phys. Rev. Lett. 128 (2022) no.5, 052001.

[22] - A. De Rujula *et. al.* NPB154 (1979) 394,

[23] - J. Kripfganz and H. Perlt, ZPC41 (1988) 319,

[24] - J. Blümlein, ZPC47 (1990) 89.

[25] - A. Manohar et al., Physical Review Letters 117, 242002 (2016), JHEP 1712, 046 (2017)

[26] - T. Cridge, L.A. Harland-Lang, A.D. Martin and R.S. Thorne, Eur. Phys. J. C **81** (2021) no.8, 744.

# Back-Up Slides

Parts unknown at N$^3$LO estimated using existing covariance matrix/scale variation approach.

$$\Delta_m(ij,k) = T_m(ij,k) - \bar{T}_m,$$

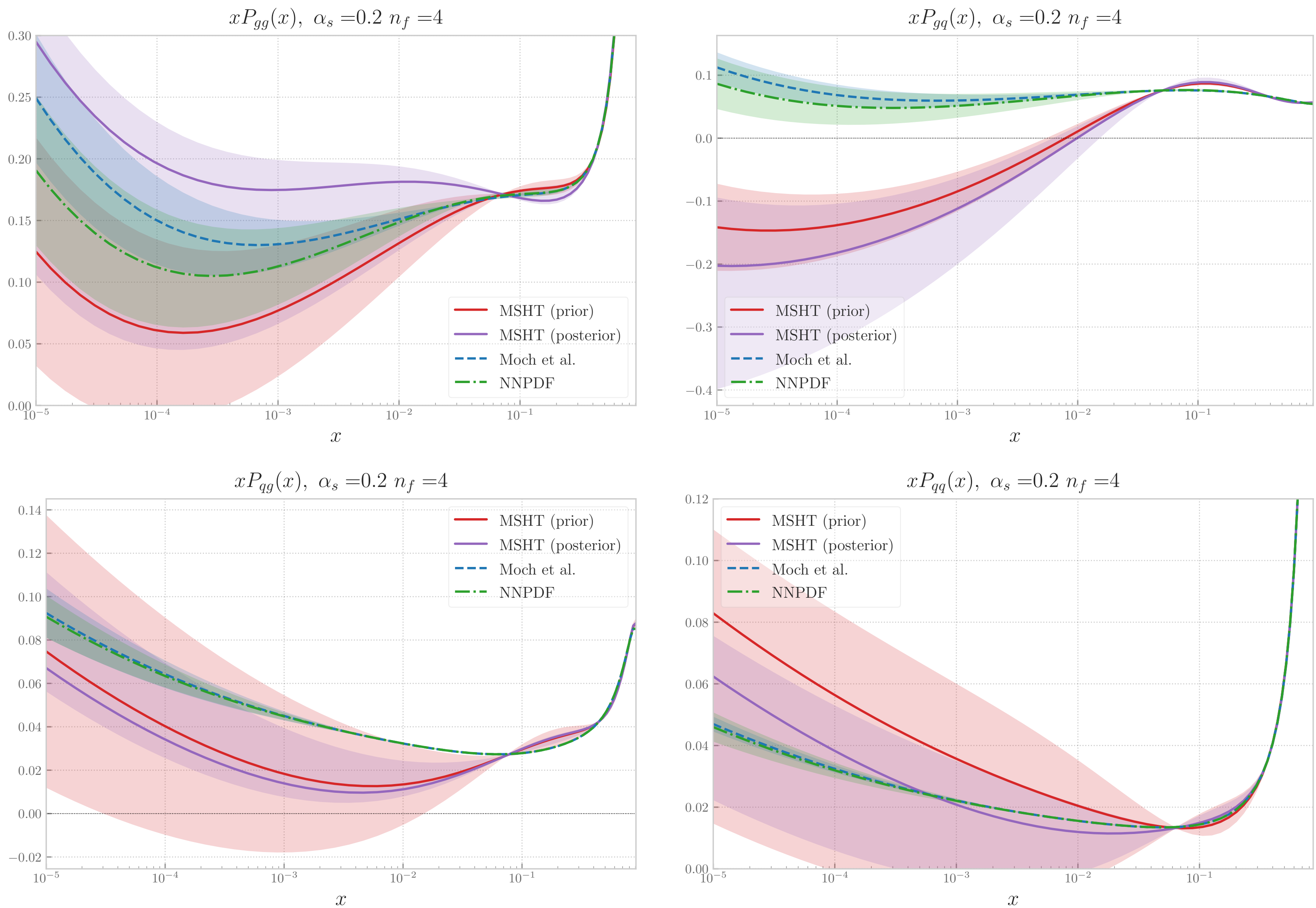
$$\mathrm{cov}_{mn}^{(ij)} = \frac{1}{\widetilde{N}_{ij}-1}\sum_{k=1}^{\widetilde{N}_{ij}} \Delta_m(ij,k)\Delta_n(ij,k).$$

$$\mathrm{cov}_{mn}^{\mathrm{IHOU}} = \mathrm{cov}_{mn}^{(gg)} + \mathrm{cov}_{mn}^{(gq)} + \mathrm{cov}_{mn}^{(qg)} + \mathrm{cov}_{mn}^{(qq)}.$$

$$(\sigma_{ij}(N))^2 = \frac{1}{\widetilde{N}_{ij}-1}\sum_{k=1}^{\widetilde{N}_{ij}} \left(\gamma_{ij}^{(3),\,(k)}(N) - \gamma_{ij}^{(3)}(N)\right)^2.$$

Gives uncertainty on splitting functions, similar approach for other quantities.

# Comparison with MSHT and NNPDF versions



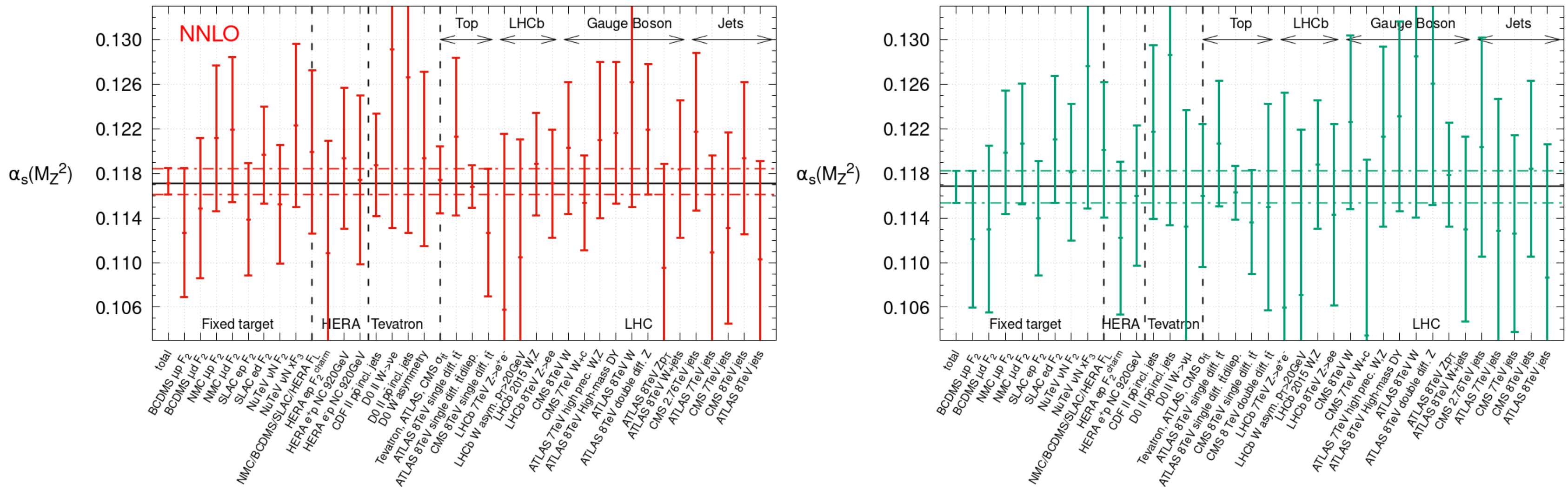
# Future Experiment Impacts – HL-LHC - study in 1810.03639

Details of data sets used in the study.

Chosen to maximise impact on currently less well-known PDFs, regions.

| Process | Kinematics | $N_{\rm dat}$ | $f_{\rm corr}$ | $f_{\rm red}$ | Baseline |
|---|---|---|---|---|---|
| $Z$ $p_T$ | $20\,{\rm GeV} \le p_T^{ll} \le 3.5$ TeV<br>$12\,{\rm GeV} \le m_{ll} \le 150$ GeV<br>$\|y_{ll}\| \le 2.4$ | 338 | 0.5 | $(0.4, 1)$ | [52] (8 TeV) |
| high-mass Drell-Yan | $p_T^{l1(2)} \ge 40(30)$ GeV<br>$\|\eta^l\| \le 2.5$, $m_{ll} \ge 116$ GeV | 32 | 0.5 | $(0.4, 1)$ | [47] (8 TeV) |
| top quark pair | $m_{t\bar{t}} \simeq 5$ TeV, $\|y_t\| \le 2.5$ | 110 | 0.5 | $(0.4, 1)$ | [50] (8 TeV) |
| $W$+charm (central) | $p_T^\mu \ge 26$ GeV, $p_T^c \ge 5$ GeV<br>$\|\eta^\mu\| \le 2.4$ | 12 | 0.5 | $(0.2, 0.5)$ | [24] (13 TeV) |
| $W$+charm (forward) | $p_T^\mu \ge 20$ GeV, $p_T^c \ge 20$ GeV<br>$p_T^{\mu+c} \ge 20$ GeV<br>$2 \le \eta^\mu \le 4.5$, $2.2 \le \eta^c \le 4.2$ | 10 | 0.5 | $(0.4, 1)$ | LHCb projection |
| Direct photon | $E_T^\gamma \lesssim 3$ TeV, $\|\eta_\gamma\| \le 2.5$ | 118 | 0.5 | $(0.2, 0.5)$ | [55] (13 TeV) |
| Forward $W, Z$ | $p_T^l \ge 20$ GeV, $2.0 \le \eta^l \le 4.5$<br>$60\,{\rm GeV} \le m_{ll} \le 120$ GeV | 90 | 0.5 | $(0.4, 1)$ | [49] (8 TeV) |
| Inclusive jets | $\|y\| \le 3$, $R = 0.4$ | 58 | 0.5 | $(0.2, 0.5)$ | [61] (13 TeV) |
| Total | | 768 | | | |

**Table 2.1.** Summary of the features of the HL–LHC pseudo–data generated for the present study. For each process we indicate the kinematic coverage, the number of pseudo–data points used across all detectors $N_{\rm dat}$, the values of the correction factors $f_{\rm corr}$ and $f_{\rm red}$; and finally the reference from the 8 TeV or 13 TeV measurement used as baseline to define the binning and the systematic uncertainties of the HL–LHC pseudo–data, as discussed in the text.




Determine uncertainty by dynamical tolerance procedure, same as for eigenvector uncertainties.

Examine fit quality with varying $\alpha_S(M_Z^2)$ for each data set, and find most limiting set in each direction.

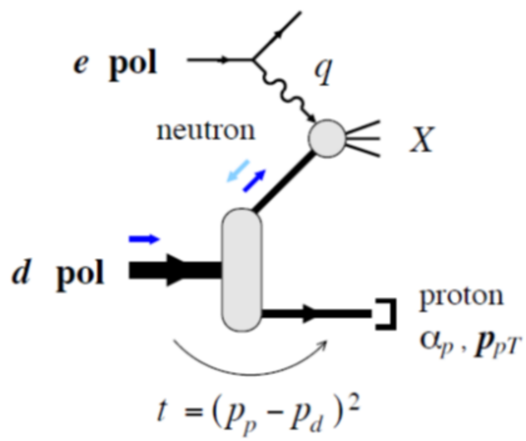
Find very similar constraints regarding datasets at each order, though slightly wider bounds at aN$^3$LO on data types with current N$^3$LO K-factors freedom. Better measure of true theoretical uncertainty.

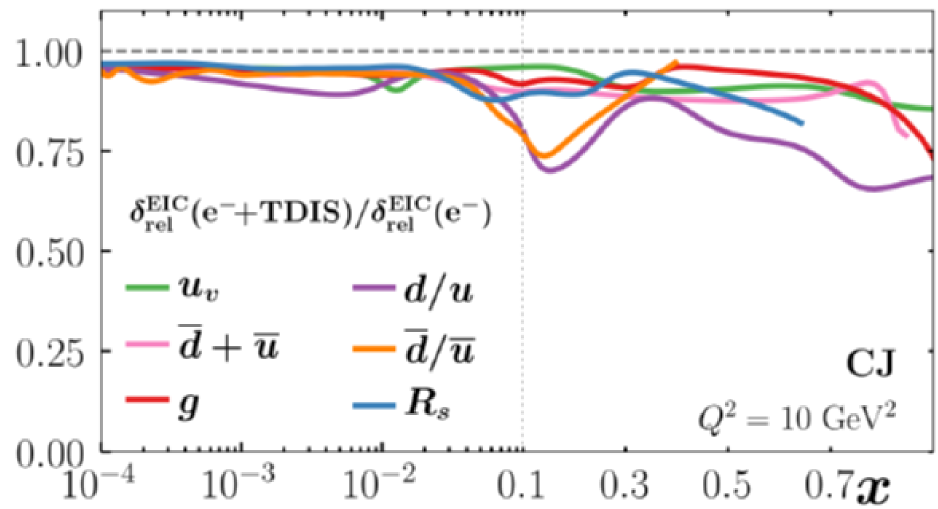
Uncertainty corresponds to $\Delta\chi^2 = 13$ NNLO, $\Delta\chi^2 = 16$ N$^3$LO.

## Deuteron structure Functions.

Possibility of tagging a final state neutron in deuteron scattering.

As $t - m_N^2 \rightarrow 0$ corresponds to scattering of neutron, i.e. "on-shell extrapolation".

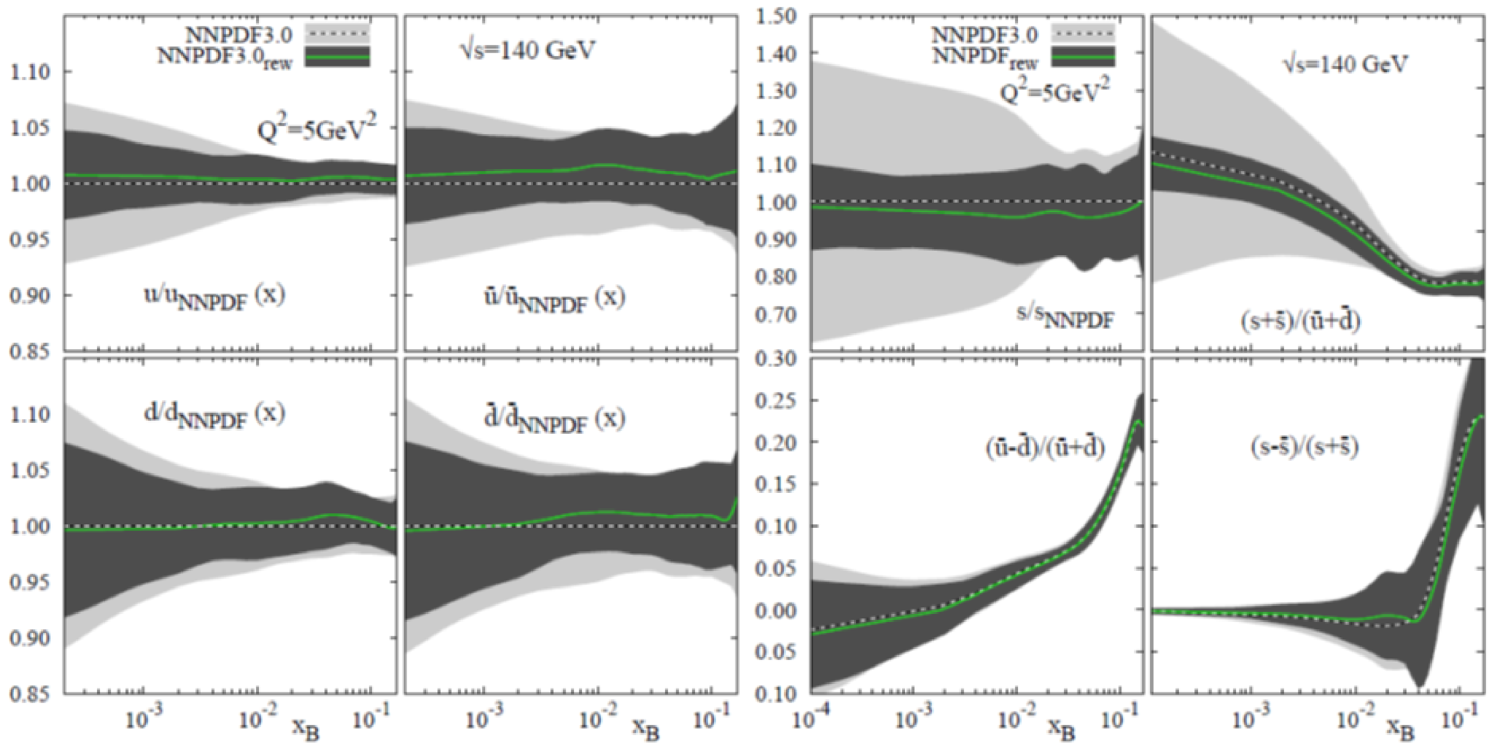



Eliminates uncertainty in deuteron correction to sum of free $p, n$ distributions.

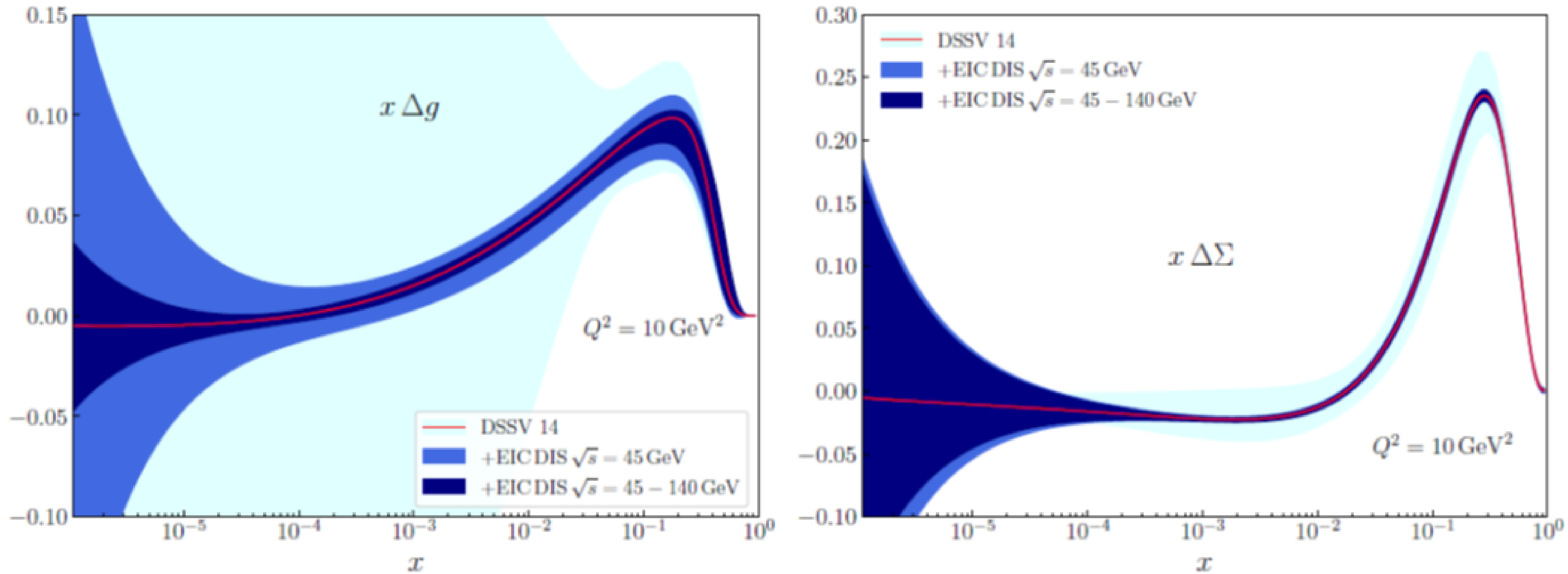
Can possibly start testing isospin symmetry.

# Semi-Inclusive DIS



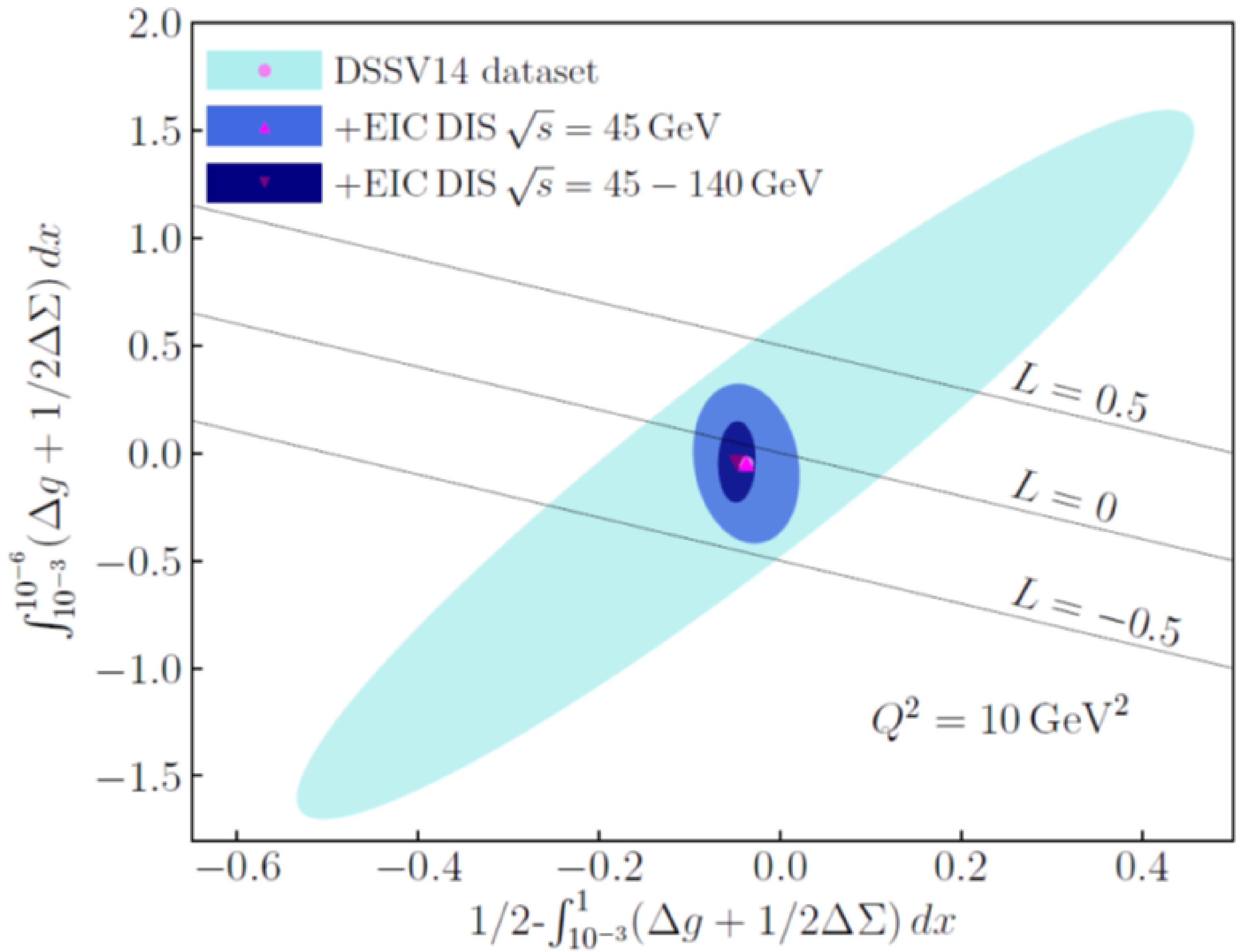

Also an impact, particularly on strange from kaon production, though tied to uncertainty on fragmentation functions Phys. Rev. D 990094004.

Leads directly to the below improvements in PDFs Borsa et al., Phys. Rev. D.102 (2020) 094018.

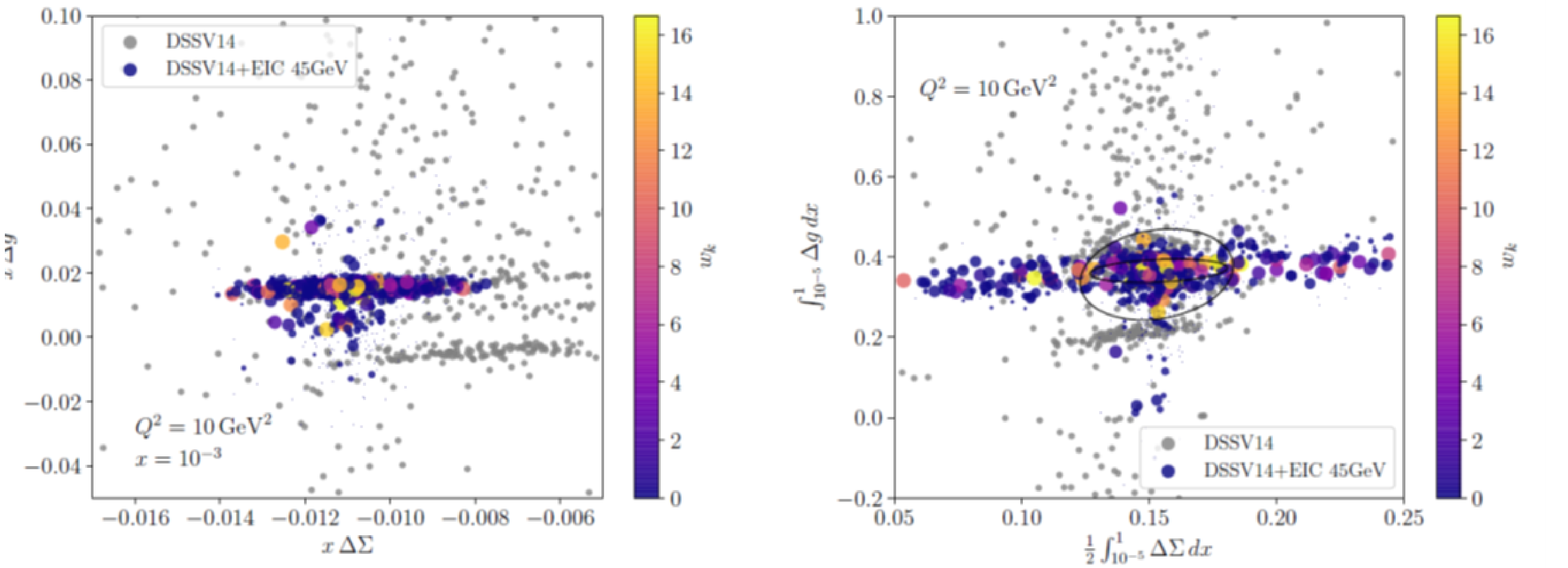



Slight caveat, need care that parameterization flexibility consistent with precision of data and pseudodata – PDFs have 5 parameters here, probably sufficient.

Related to this is the possible impact on the knowledge of the sum rule.



A direct example of how PDF replicas are selected and focussed by the potential new data.

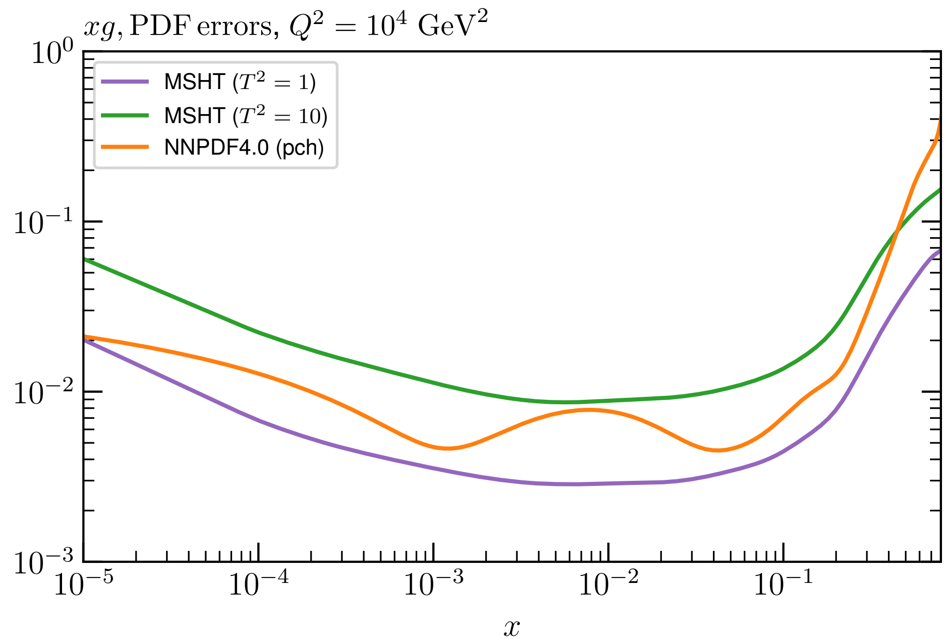



Distinct bunching, but also variation of weight within reduced set of PDFs.

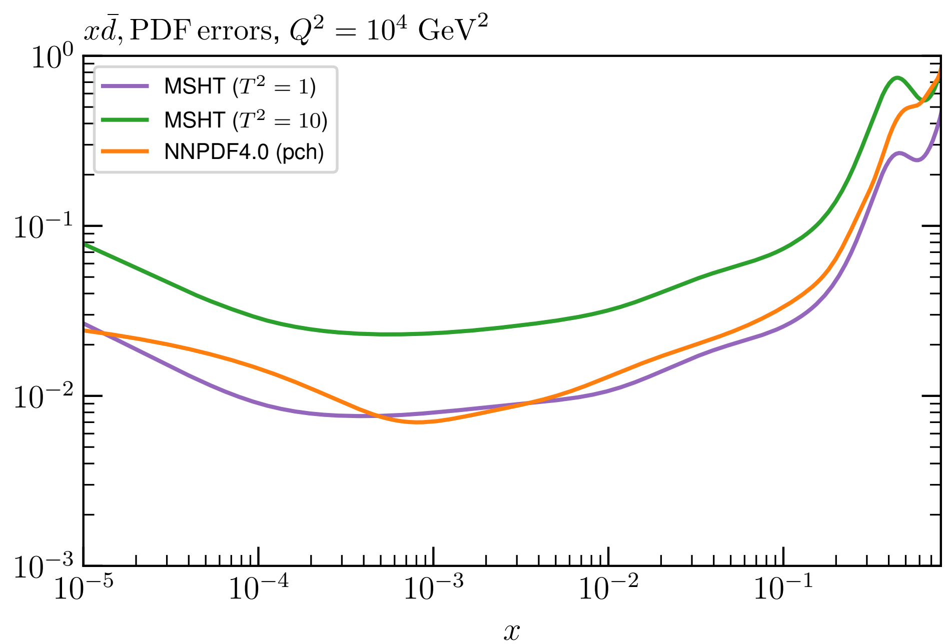
# Uncertainties on PDFs

Uncertainties from same data using various approaches.

Gluon distribution.

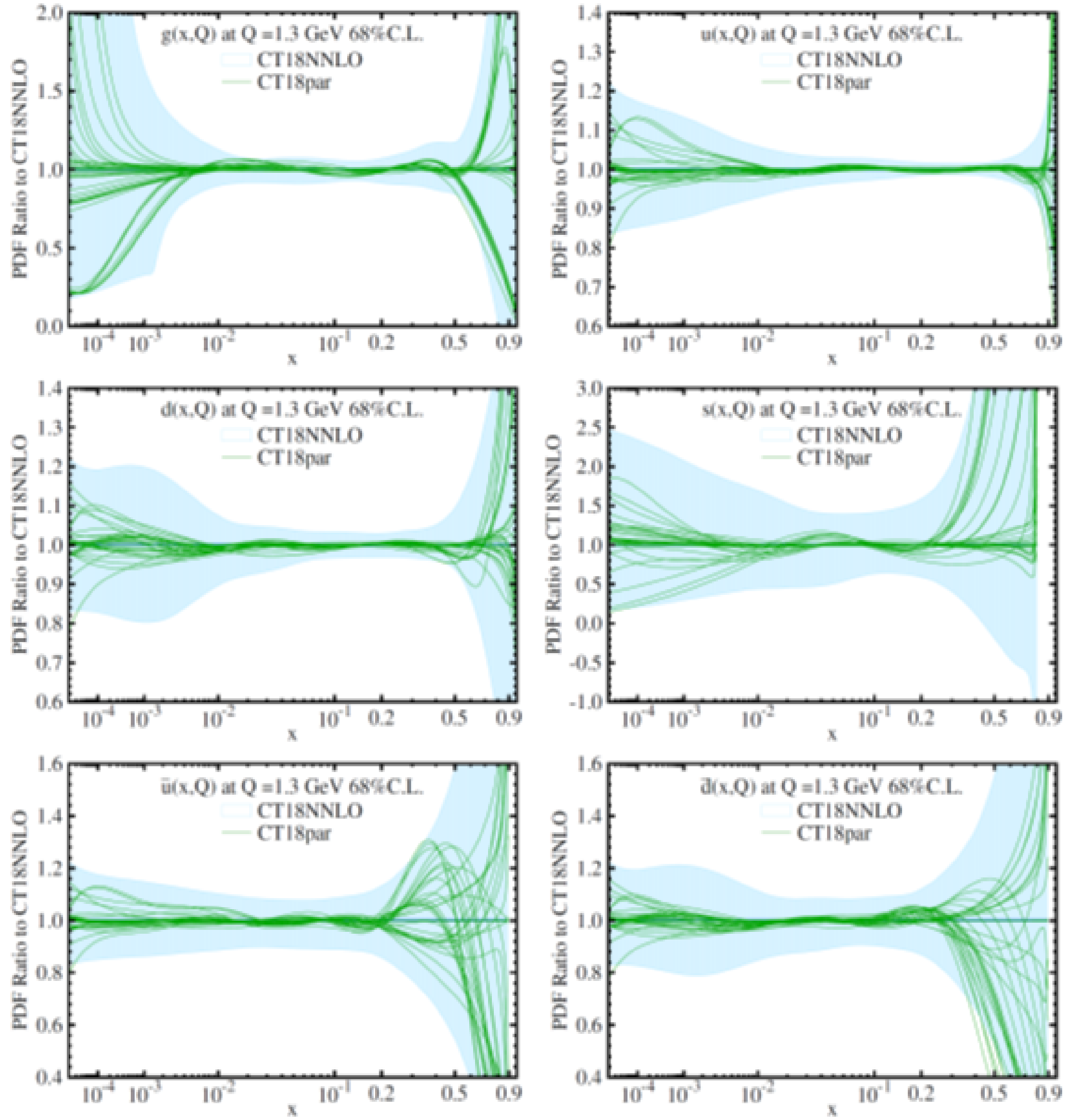



Anti-down distribution.



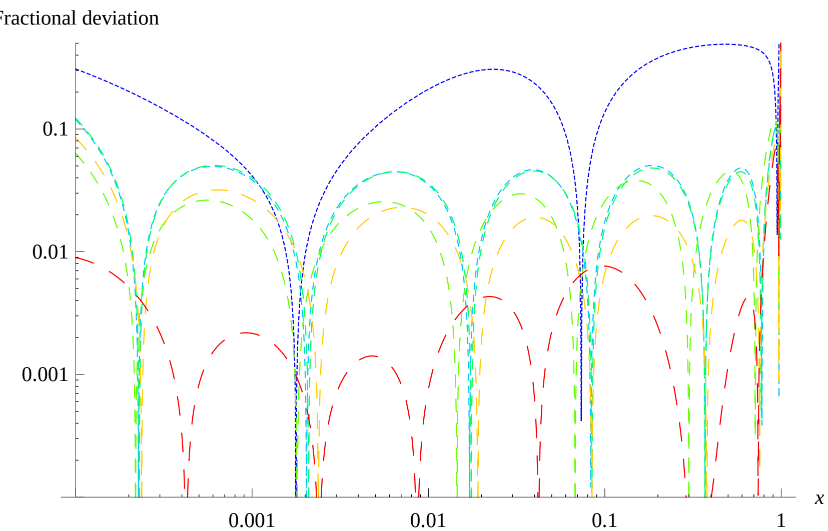
Use 26 parameters, but use various different choices and investigate PDF changes.

Require consistency with $\chi^2$ tolerance chosen.



## Extension of parameterisation.

General parameterisation used $A(1-x)^{\eta}x^{\delta}(1+\sum_{i=1}^{n}a_iT_i(1-2x^{\frac{1}{2}}))$, where $T_i(1-2x^{\frac{1}{2}}))$ are Chebyshev polynomials.



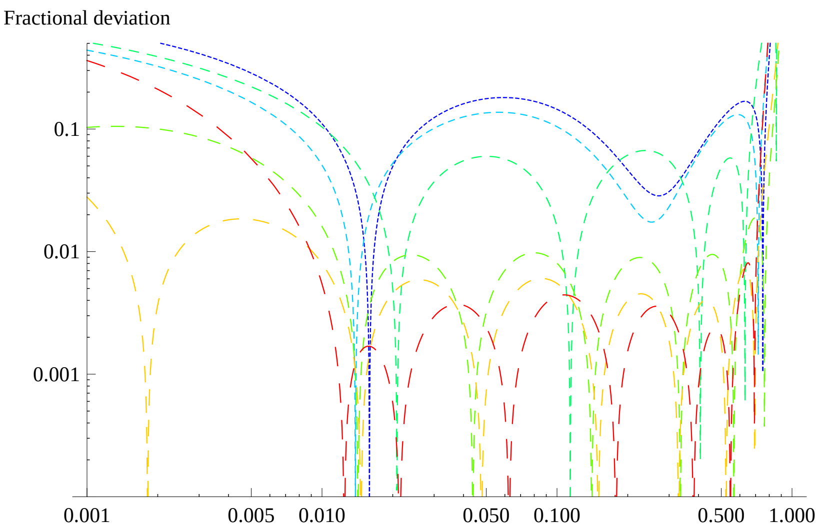



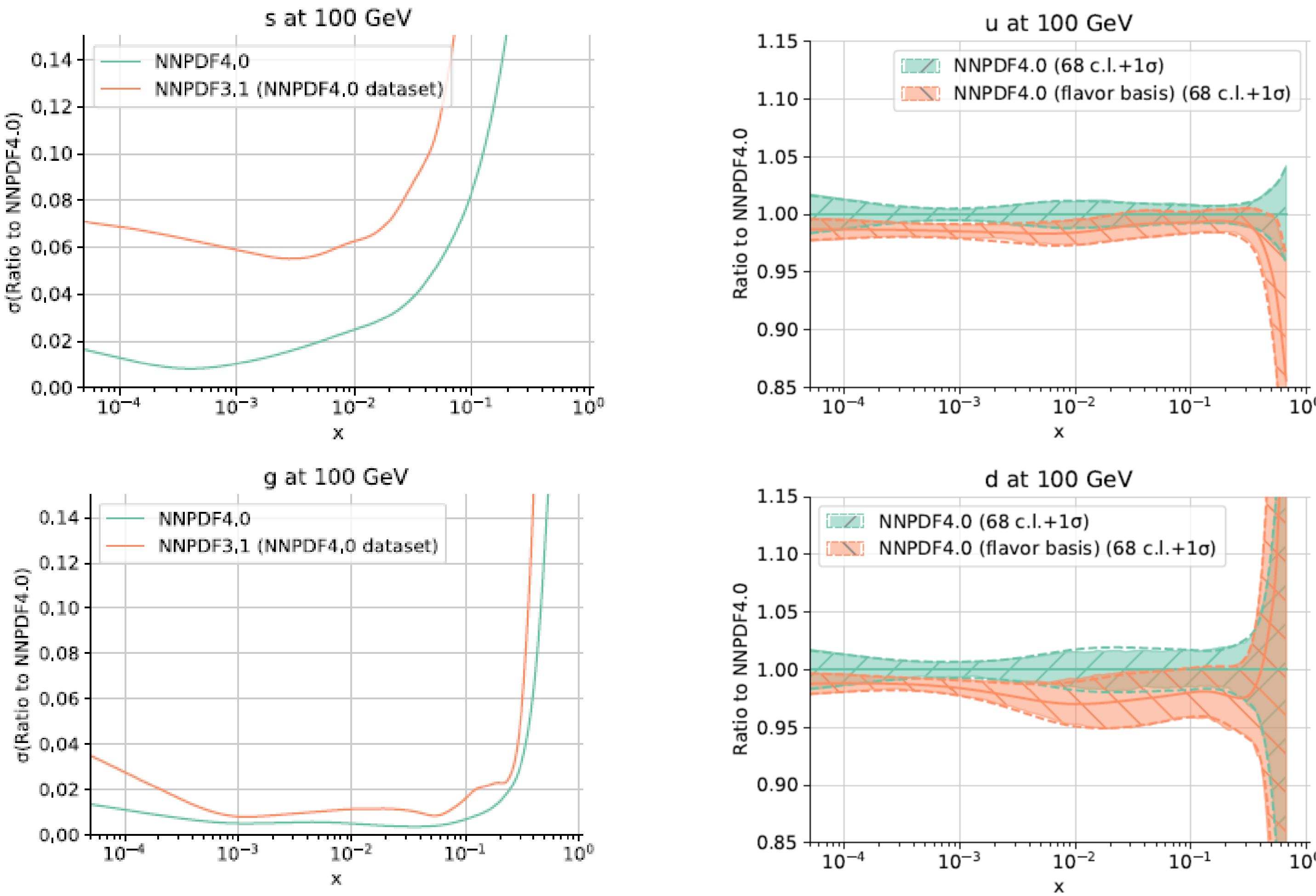

Illustration of precision possible with increasing $n$, sea-like (left) and valence-like (right) (where pseudo-data for $x > 0.01$). Using $n = 6$ would lead to much better than $1\%$ precision.

For most PDFs $n = 4$ default for MMHT2014 – 36 parameters.

Now extend to $n = 6$ – total of 51 parton parameters.

When determining uncertainties go from 25 eigenvector pairs to 32.

# Uncertainties on PDF - Methodology.




Comparison between uncertainty on NNPDF4.0 and NNPDF3.1 and comparison between bases for PDF input for NNPDF4.0 when using exactly the same data.