PDFs from LHeC/FCCeh vs LHC

Amanda Cooper-Sarkar, Oxford

arXiv:1206.2913

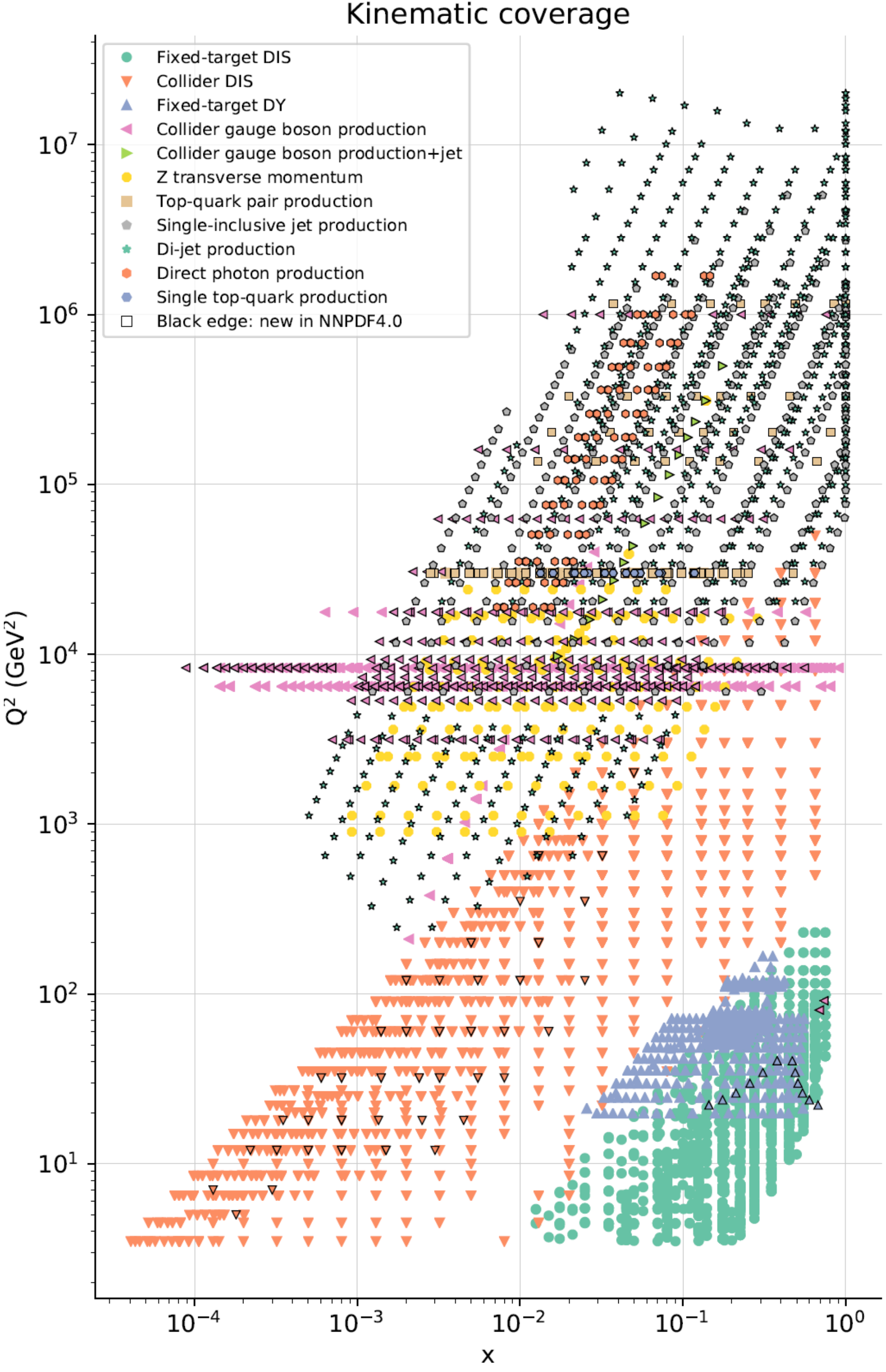
see also, FCC CDR, vols 1 and 3:

physics, EPJ C79 (2019), 6, 474

FCC with eh integrated, EPJ ST 228 (2019), 4, 755

J. Phys. G 48 (2021) 11, 110501

(arXiv:2007.14491)



The HERA data are the 'backbone of all PDF fits BUT what could HERA not do?

High-x gluon and sea flavour detail s,c

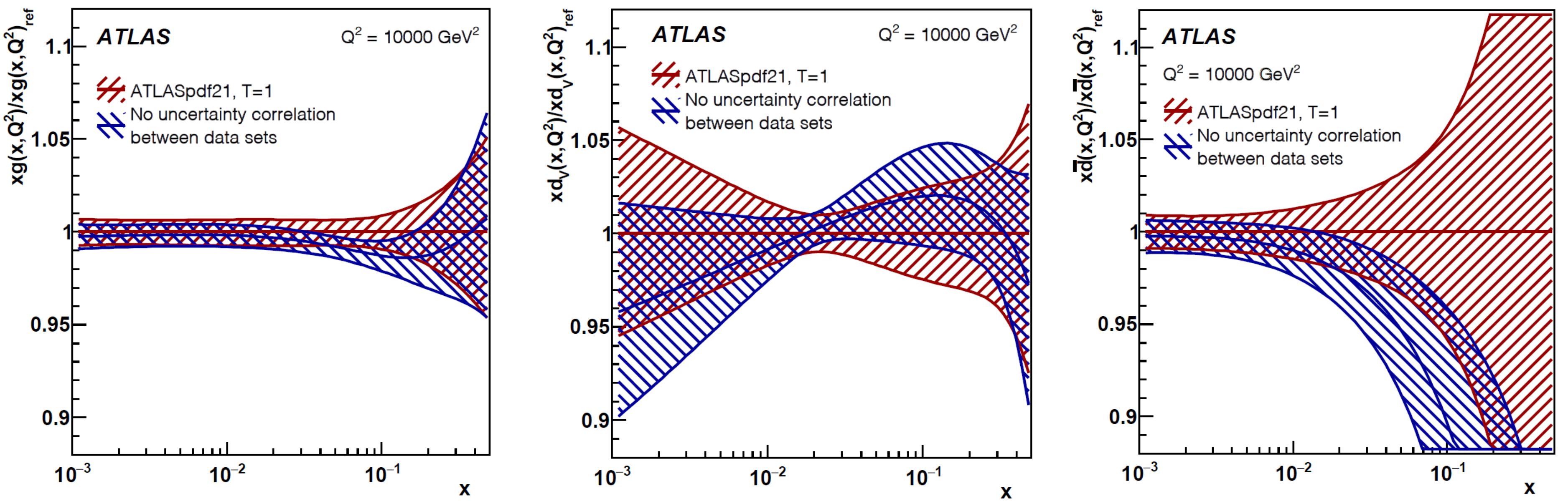
What other data can we use?

- Drell-Yan data from fixed target DIS and the Tevatron and LHC
- W,Z rapidity spectra from Tevatron and LHC
- Jet pT spectra from Tevatron and LHC
- Top-anti-top differential cross-sections from LHC
- W and Z +jet spectra, or Z pt spectra from LHC
- W and Z +heavy flavours from LHC
- **Beware**: IS the factorisation theorem proven?-only for DY!
- **Beware**: there may be new physics at high scale that we 'fit away'
- **Further warning**, this additional information comes from many different groups– often there is **no clarity on the correlations of experimental systematic uncertainties between differing LHC measurements-** and **there are tensions-- this can scupper the goal of 1% accuracy**

Effect of correlations between data sets

Lets look at a scale relevant for LHC physics and focus on the middling x range where W,Z and Higgs are produced

gluon ratio **$d_{valence}$ ratio** **$\bar{d}$ ratio**

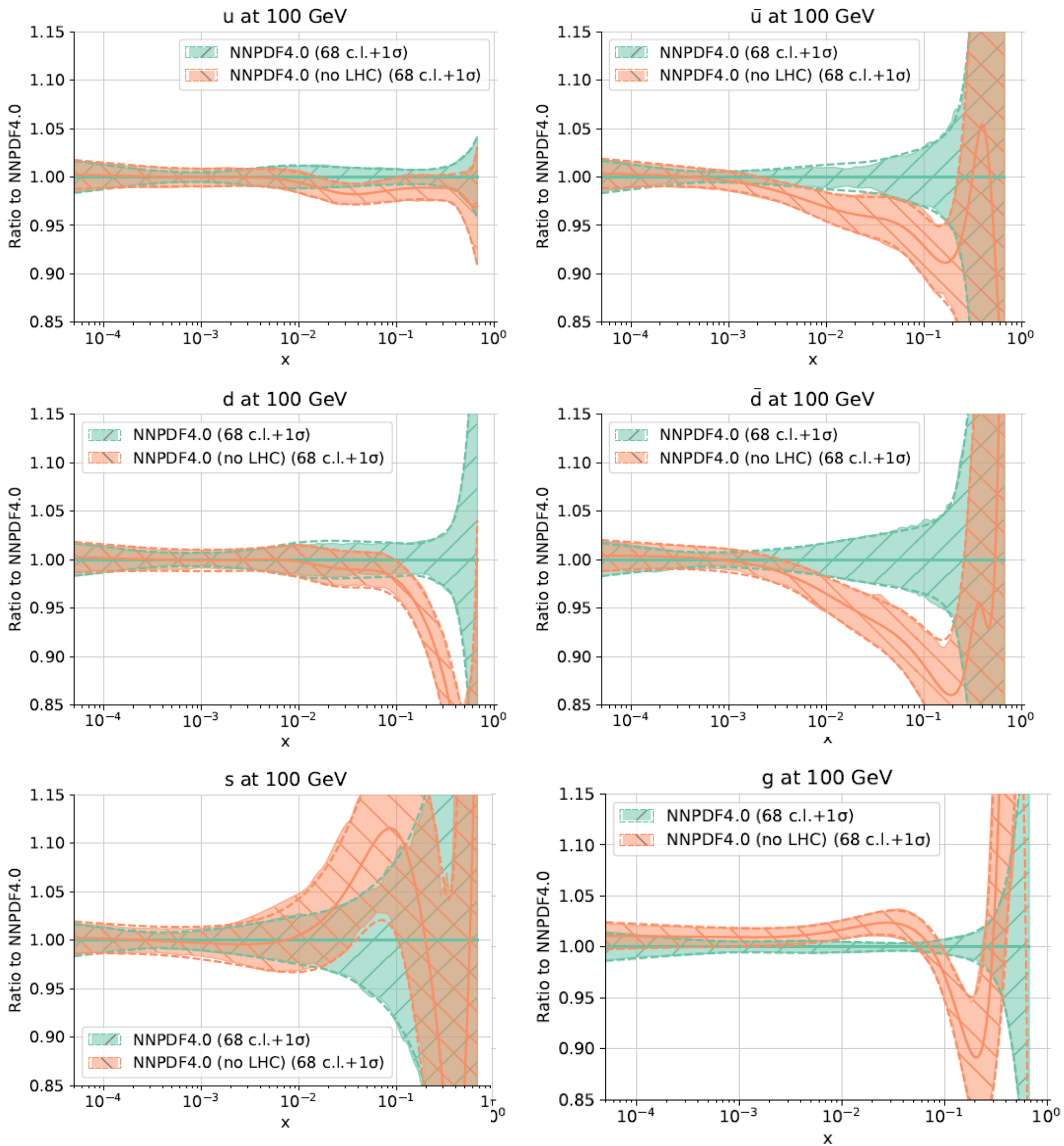


The χ2 of the fit is 30 units better when correlations are included
The difference in PDFs is small for the gluon
But can be larger in the d-quark sector
Remember the goal for PDF precision is ~1% for M_W and $\sin^2\theta_W$ measurements if BSM effects are to be seen by the deviations of these parameters from their SM values

Correlations can be important

Let's see how much LHC data has improved PDFs

NNPDF4.0 includes modern LHC data on W,Z + jets + top + Zpt from 7 and 8 TeV running. Compare PDFs with and without LHC



This looks good BUT specific choices were made by NNPDF e.g which top-quark differential distributions are used and of which jet data distributions are used etc., and what are the correlations between systematic uncertainties Other PDF groups are making slightly different choices—and such differences increase the total uncertainty due to differences between PDF sets

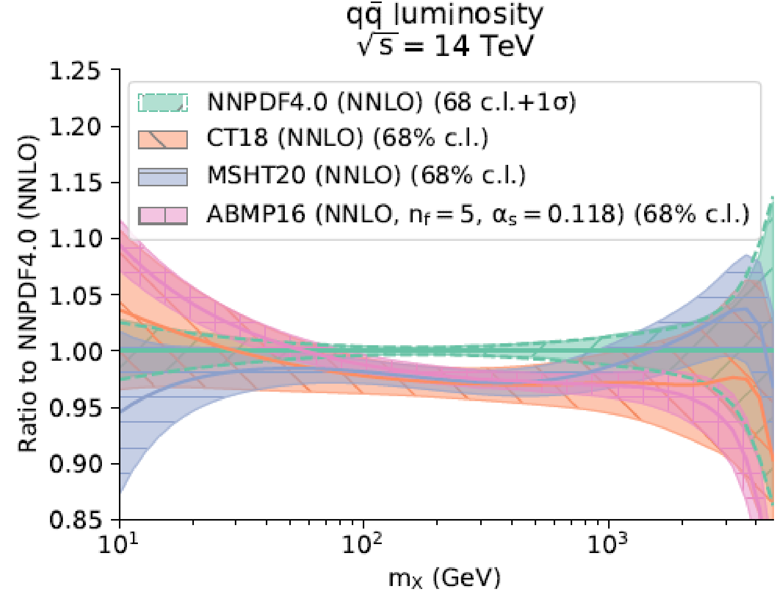
There are also differences in parametrisations, model choices (e.g.heavy quark treatment, start point of evolution..), treatment of tolerance etc….

How well do we know PDFs today ?

One way to see the impact of the uncertainties on the parton distribution functions at the LHC is in terms of parton-parton luminosities, which are the convolution of the purely partonic part of the sub-process cross-section.

The quark-antiquark and gluon-gluon luminosities for various PDFs are compared here for 14 TeV LHC running in terms of the centre of mass energy of the parton sub- process M_X

Small M_X corresponds to small x and Large M_X to large x



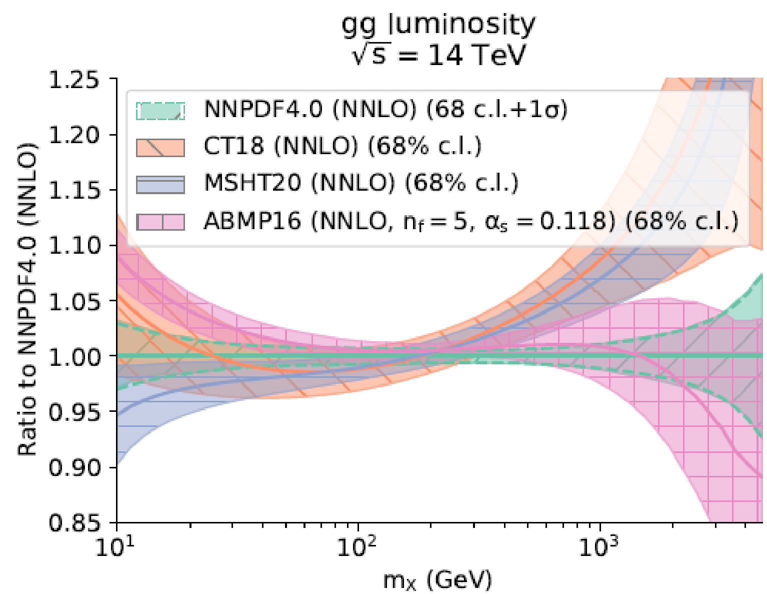
$$\frac{\partial \mathcal{L}_{\Sigma_q(q\bar{q})}}{\partial \hat{s}} = \frac{1}{s}\int_\tau^1 \frac{dx}{x} \sum_{q=d,u,s,c,b} \left[f_q(x,\hat{s}) f_{\bar{q}}(\tau/x,\hat{s}) + f_{\bar{q}}(x,\hat{s}) f_q(\tau/x,\hat{s})\right],$$

So for quark-antiquark production of W or Z bosons ----at Mx ~80,90 GeV

Or for gluon-gluon production of Higgs at ---Mx~125 GeV

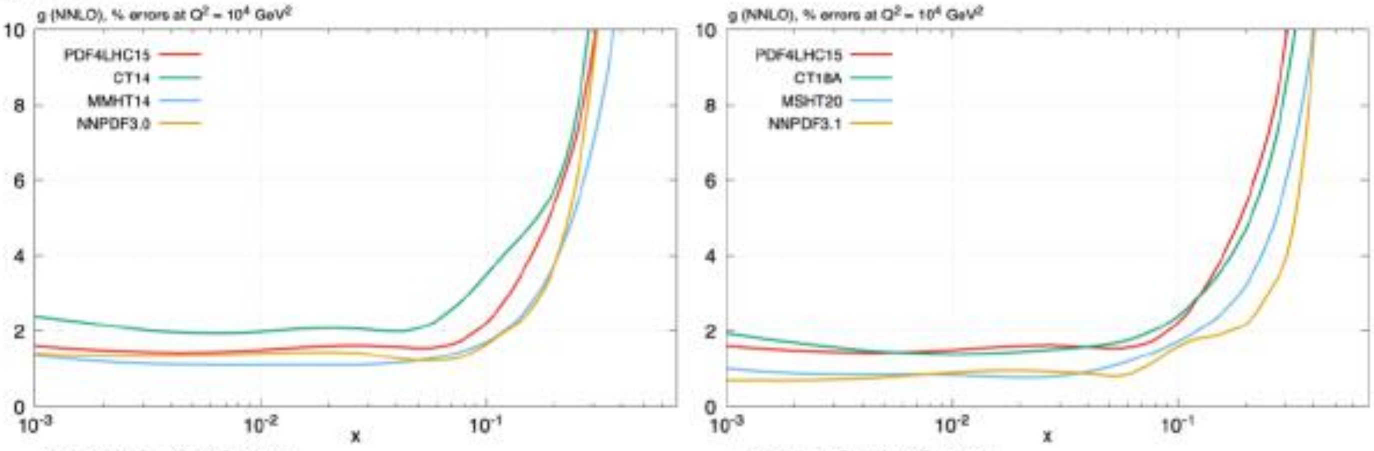
the parton-parton luminosities are fairly well known….but not as well known as we'd like for the backgrounds to SM measurements of $M_W/\sin^2\theta_W$

This is much worse for higher mass particles that could be produced by 'Beyond' Standard Model (BSM) physics



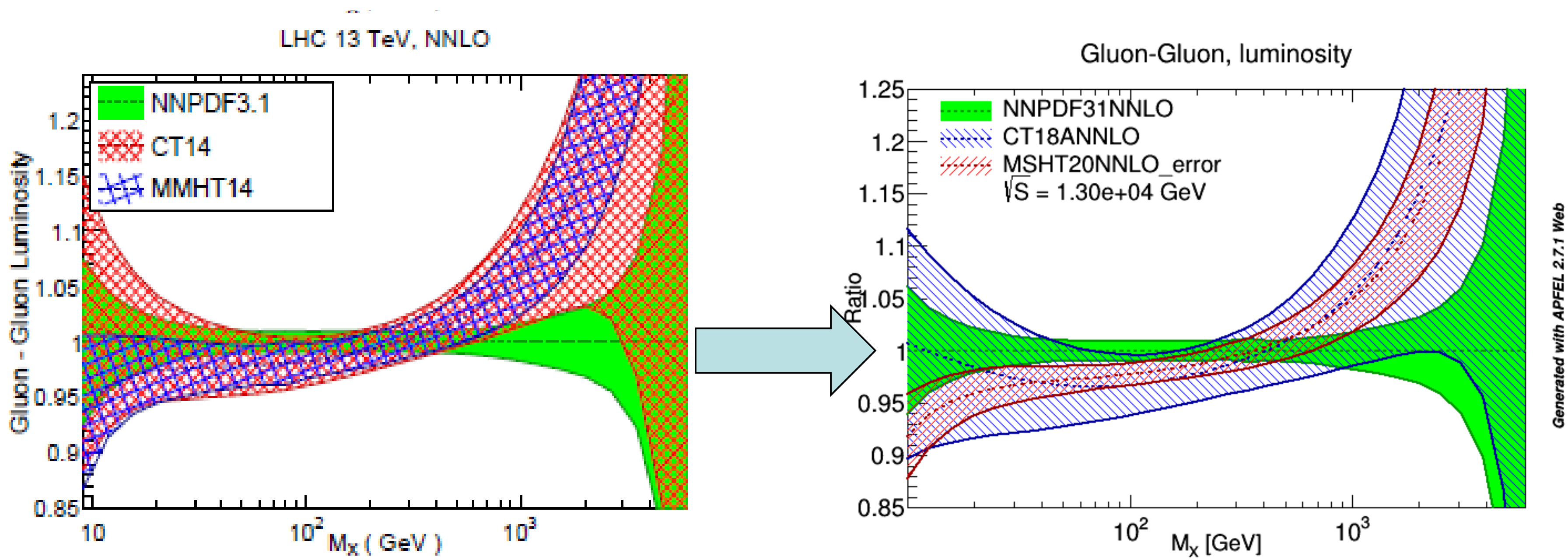
$$\frac{\partial \mathcal{L}_{gg}}{\partial \hat{s}} = \frac{1}{s}\int_\tau^1 \frac{dx}{x} f_g(x,\hat{s}) f_g(\tau/x,\hat{s}),$$

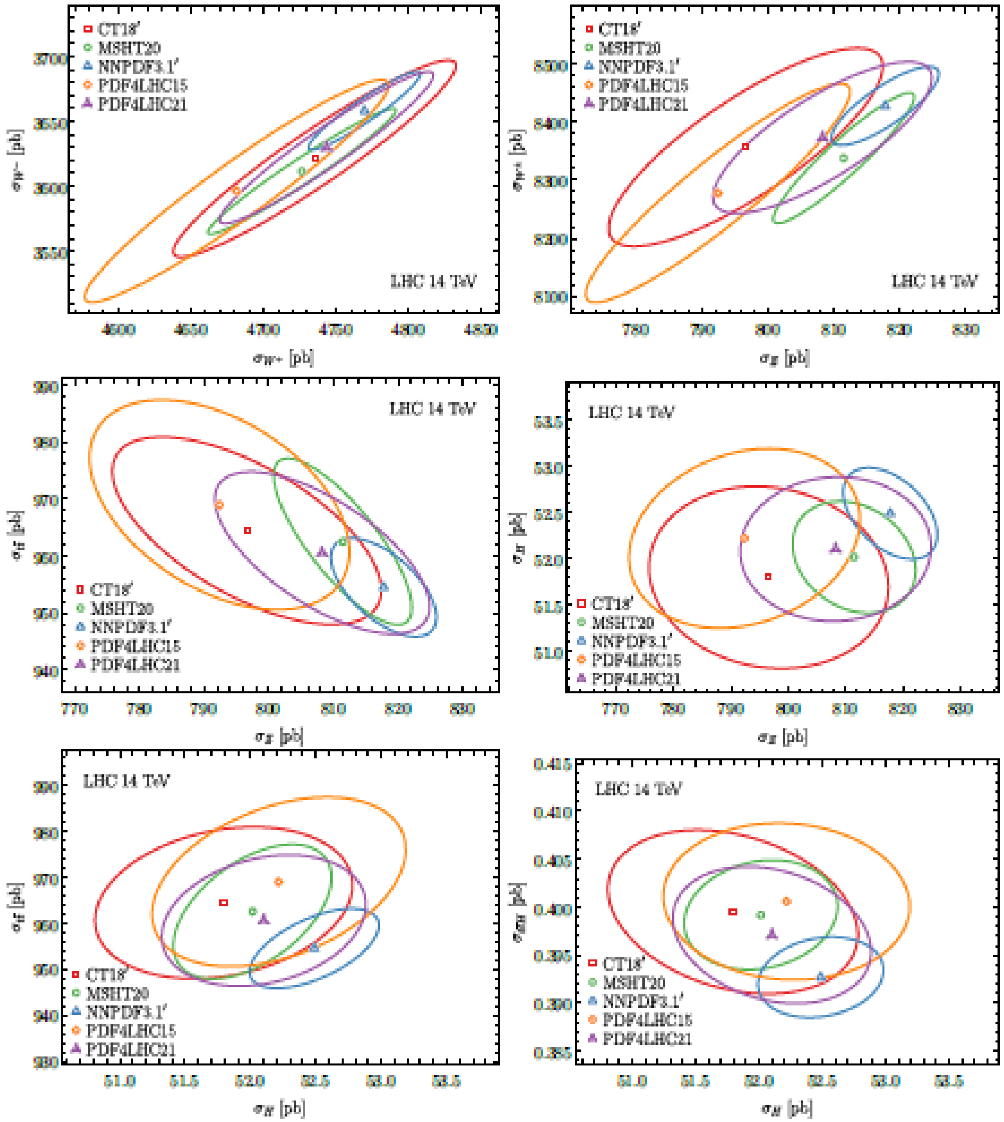
Gluon



IS THERE PROGRESS?

As the uncertainties of each individual PDF decrease with the input of more information, the divergence of the PDFs from each other has increased





The PDF4LHC group makes combinations of the PDFs from the three main fitting groups NNPDF, CT and MSHT

The PDF4LHC15 combination has now been superseded by the PDF4LHC21 combination (issued in 2022!) arxiv: 2203.05506

There IS an improvement in uncertainty BUT this is not enough e.g.

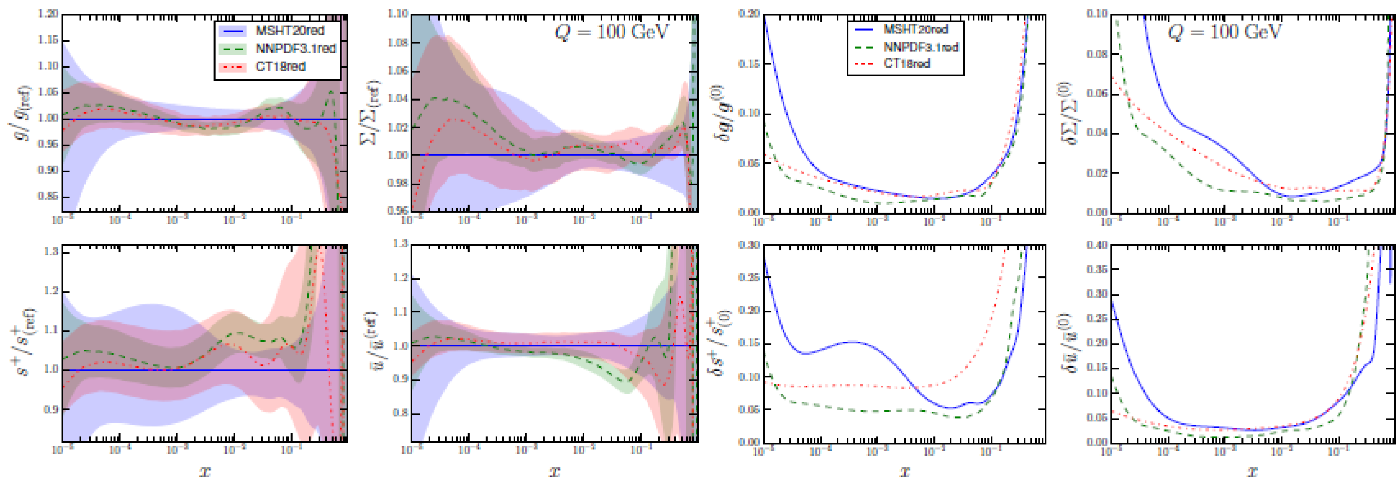
1. to reduce the PDF uncertainty on on LHC measurements of SM parameters such as M_W, sufficiently to compete with the CDF uncertainty- we need more than this…
2. PDF uncertainty at high x still limits our ability to see new physics

The PDF4LHC group makes combinations of the PDFs from the three main fitting groups NNPDF, CT and MSHT

But first we try to understand differences by using a common data set and common settings for heavy quark masses and alphas

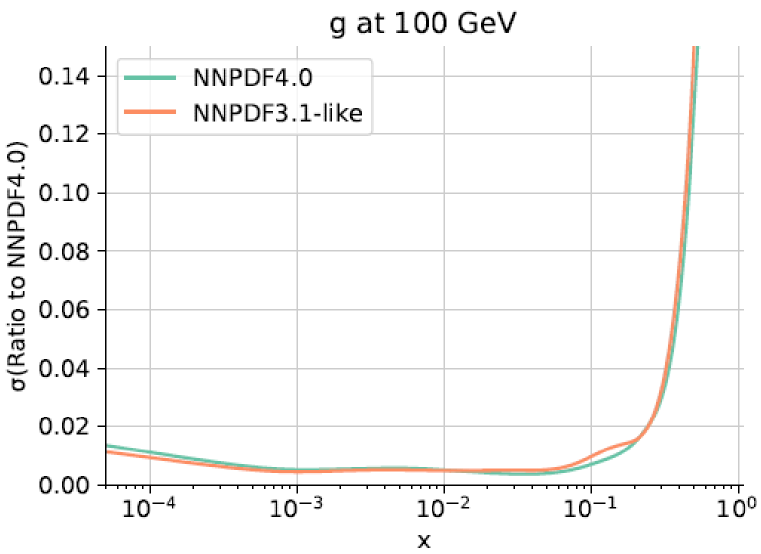
PDF Benchmarking: Reduced Fits

- Use fits to reduced common datasets and common theory settings.



- Very good agreement within uncertainties, including gluon.
- Similar size uncertainties in data regions, differences outside this, reflecting remaining methodological and other choices.
- Agreement much improved relative to global PDFs.
- Same data and theory settings → consistent PDFs. Smaller remaining differences, e.g. in errors, reflect methodological choices.

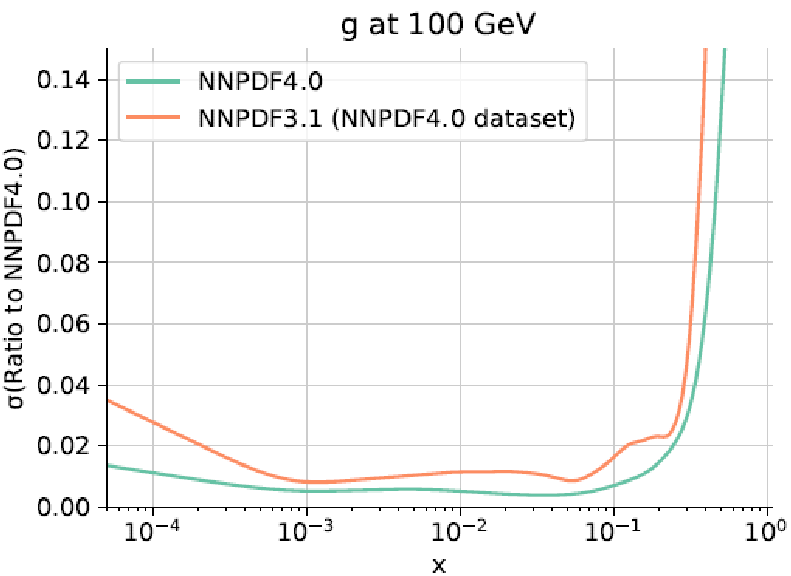
However, it is not recommended to use these reduced fits, greater consistency does not mean greater accuracy—the differences in the main fits are there for reasons that each group will support



Since the issue of PDF4LHC21 there has been a new PDF set from NNPDF4.0
This has a lot of new data from the LHC

Nevertheless the improvements in uncertainty are not much due to these data, they are more **due to improvements in their procedure**

The top plot compares the uncertainties of NNPDF4.0 and 3.1 data sets using the **SAME new methodology**



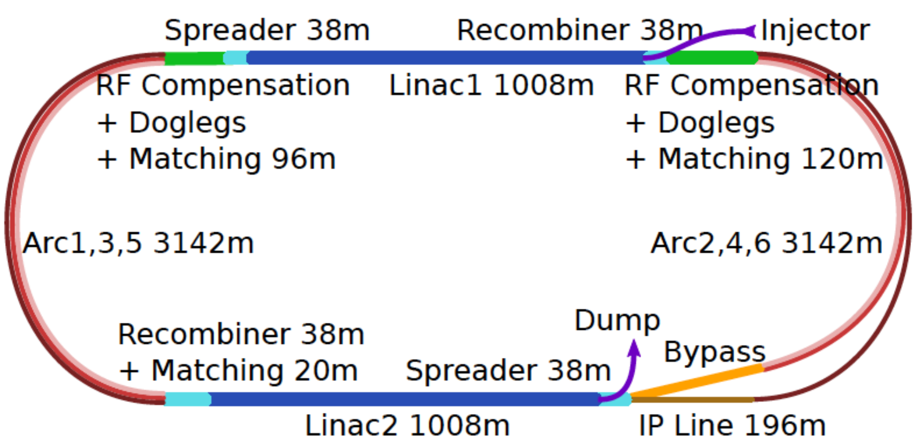
The bottom plot shows the impact of the methodology on the **SAME new data** set
4.0 shows new methodology and 3.1 here shows old methodology on new data-set

There is currently some debate in the PDF community over the new NNPDF uncertainties
But even if it is accepted this does not help much when combining with other PDFs such as MSHT20 and CT18 with different central values and larger uncertainties.

So how can we improve?

Better data, ie more consistent, more accurate data over large kinematic range, with sounder theoretical predictions

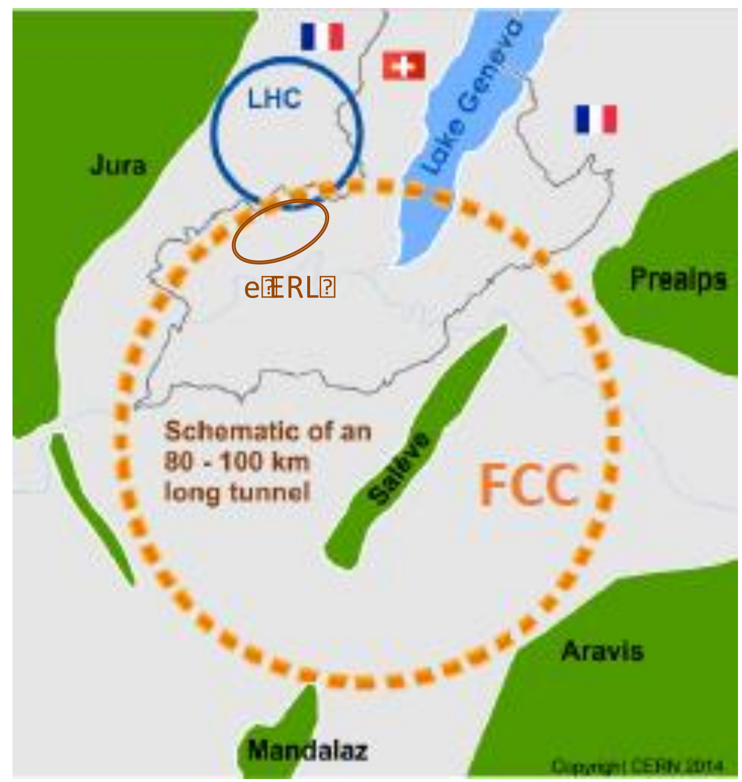
LHeC and FCCeh



energy recovery LINAC

e beam: up to 60 GeV

Lint ⟶ 1 ab^{-1} (1000× HERA ; per 10 yrs)

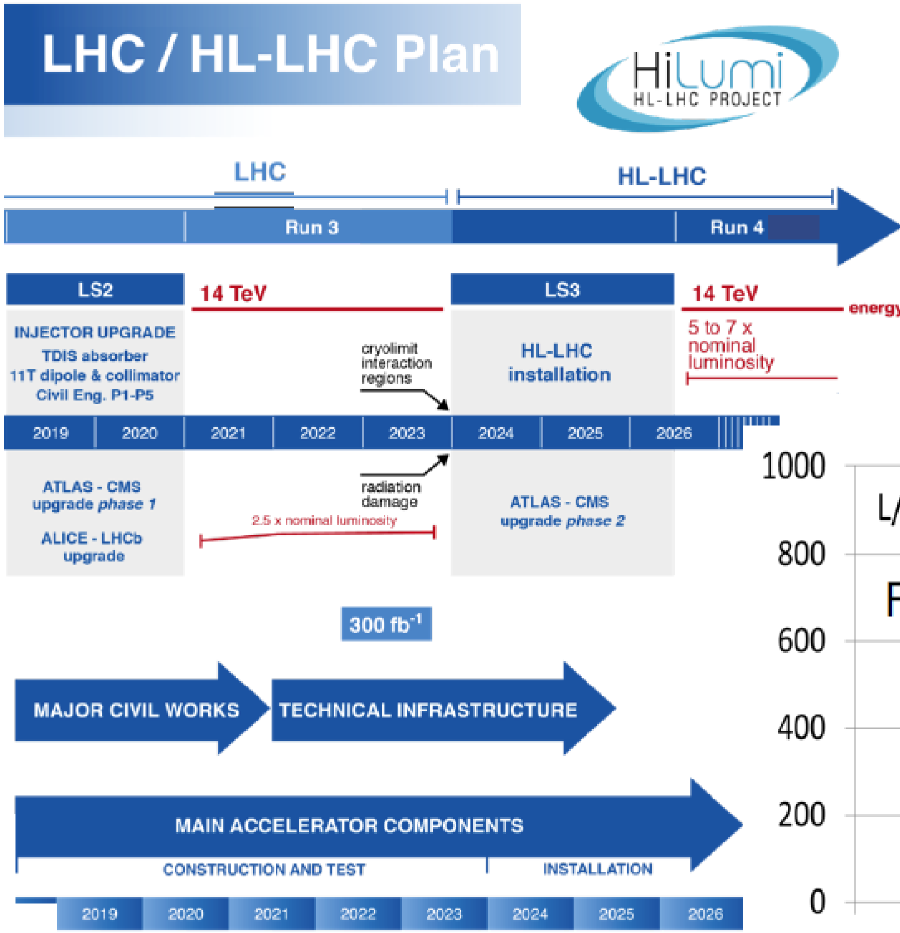


operating **synchronously** :

- with **HL-LHC** (or HE-LHC)
  p: 7 (14) **TeV, √s ≈ 1.3** (1.8) **TeV**
- and/or later with an **FCC** (A)
  p: 50 (20) **TeV, √s ≈ 3.5** (2.2) **TeV**

† **FCC (A):** a lower energy configuration that could operate earlier, in an FCC tunnel, using current magnet technology

Timelines



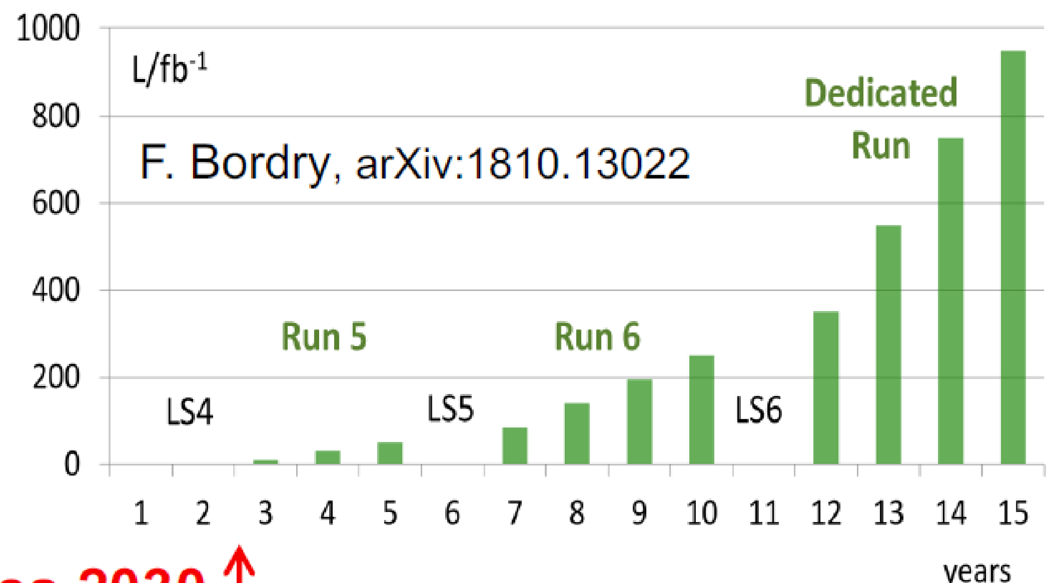
↑
today

LHeC 1st run, Lint approx. 50 fb^{-1}

total Lint ⟶ 1 ab^{-1}

(HERA total 1 fb $^{-1}$ in 15 years, 2 expts)

LHeC projected Integrated Luminosity:

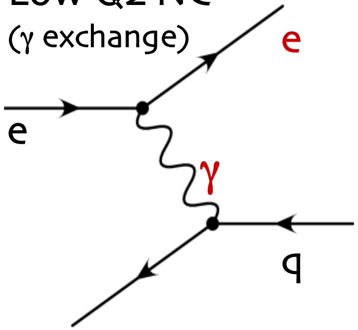


circa 2030 ↑

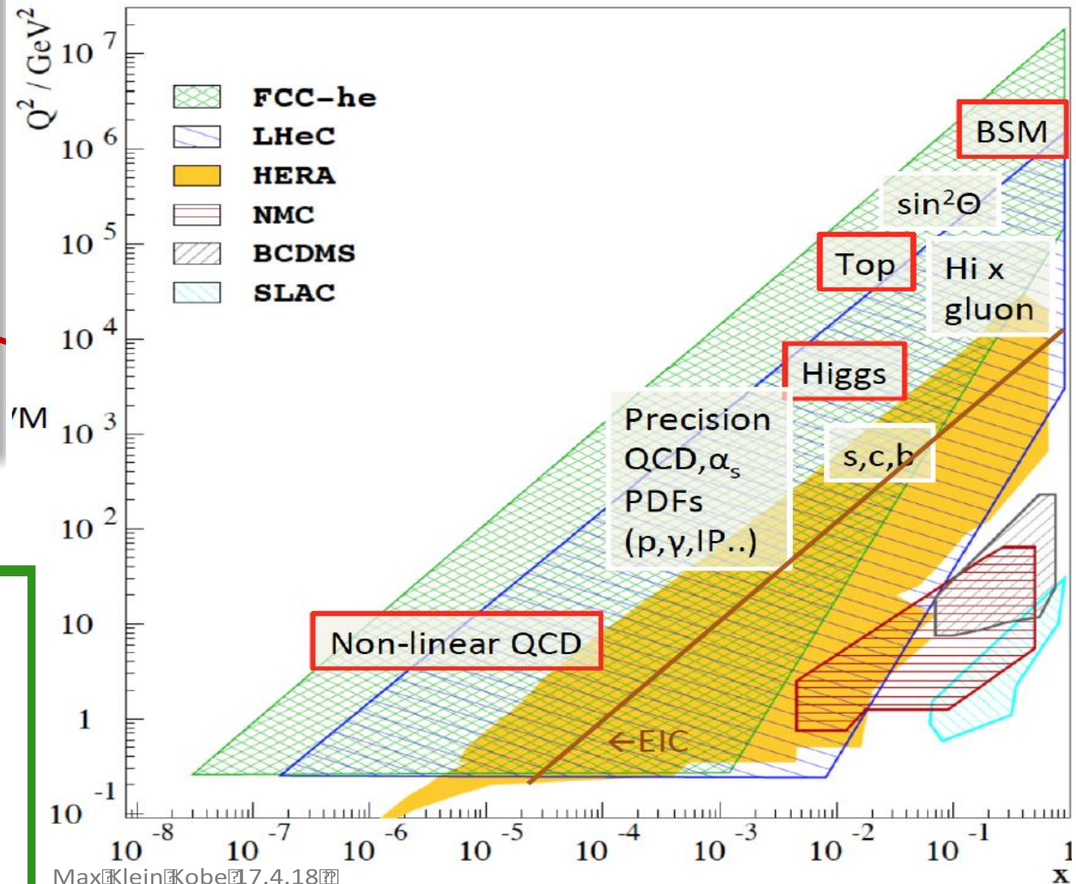
50 fb-1 can be achieved in 3years before LS5 and long before the end of HL-LHC running

Where does the information come from?

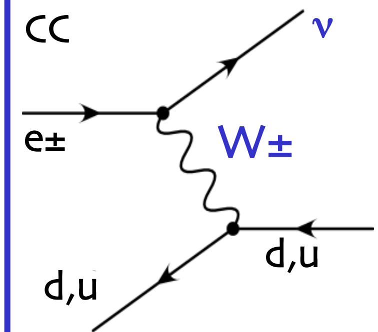
Low Q2 NC
(γ exchange)



F2 ~ Σx (q+qbar)
dF2/dlnQ2 ~ $\alpha_s \cdot g$

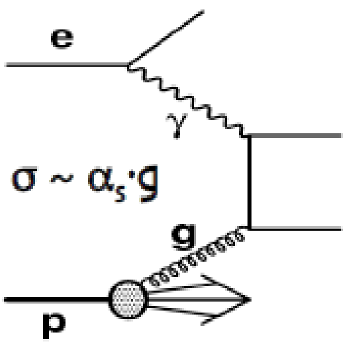


CC

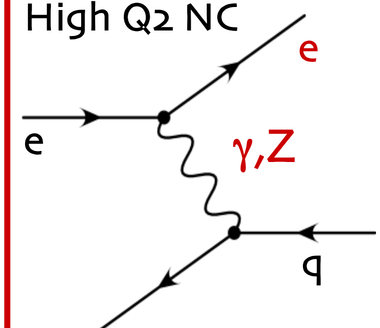


flavour composition
e+: d e-: u

Final States:
(Jets, Charm, ...)



High Q2 NC



Z → Parity Violation
xF3 ~ Σx (q-qbar)
valence

A future DIS machine would be a vast improvement on HERA in both luminosity and kinematic reach

×**15**/**120** extension in Q^2,1/x reach vs **HERA**

LHeC simulated data and QCD fits

NEW: **LHeC simulations** (e: **50 GeV***, p: 7 TeV†)

simulation: M. Klein

dataset	e charge	e pol.	lumi (fb-1)	
NC/CC	–	–0.8	5,50,1000	**luminosity**
NC/CC	+	0	1,10	**positron**
NC/CC	–	0	50	**polarisation** (important for EW)
NC/CC	–	+0.8	10,50	
NC/CC	–	0	1	**low-E** (p: 1 TeV)

uncert. assumptions:
elec. scale: 0.1%
hadr. scale 0.5%
radcor: 0.3%
γp at high y: 1%
uncorrelated uncert.: 0.5%
CC syst.: 1.5%
luminosity: 0.5%

*corresponds to possibility of smaller ERL cf. previous 60 GeV simulations †except for low-E

various combinations studied;
shown frequently in following slides:

LHeC 1st Run (50 fb^{-1} e– only; 3 yrs)

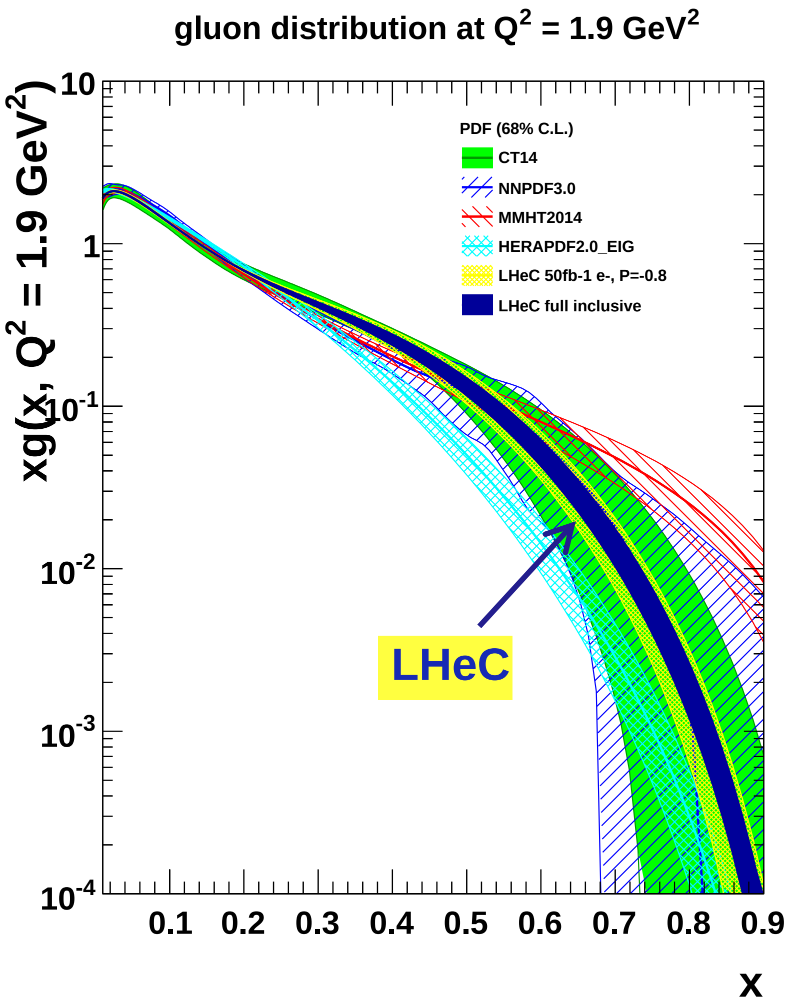
LHeC full inclusive

QCD analysis a la HERAPDF2.0, except **more flexible**, notably in **NO constraint** requiring dbar=ubar at small x;

4+1 xuv, xdv, xUbar, xDbar and **xg** (**14 free parameters**, cf. 10 by default in CDR)

5+1 xuv, xdv, xUbar, xdbar, xsbar and **xg** (if strange and HQ included; **17 free parameters**)

Gluon at large x



gluon at large x is small and currently very poorly known;

crucial for new physics searches

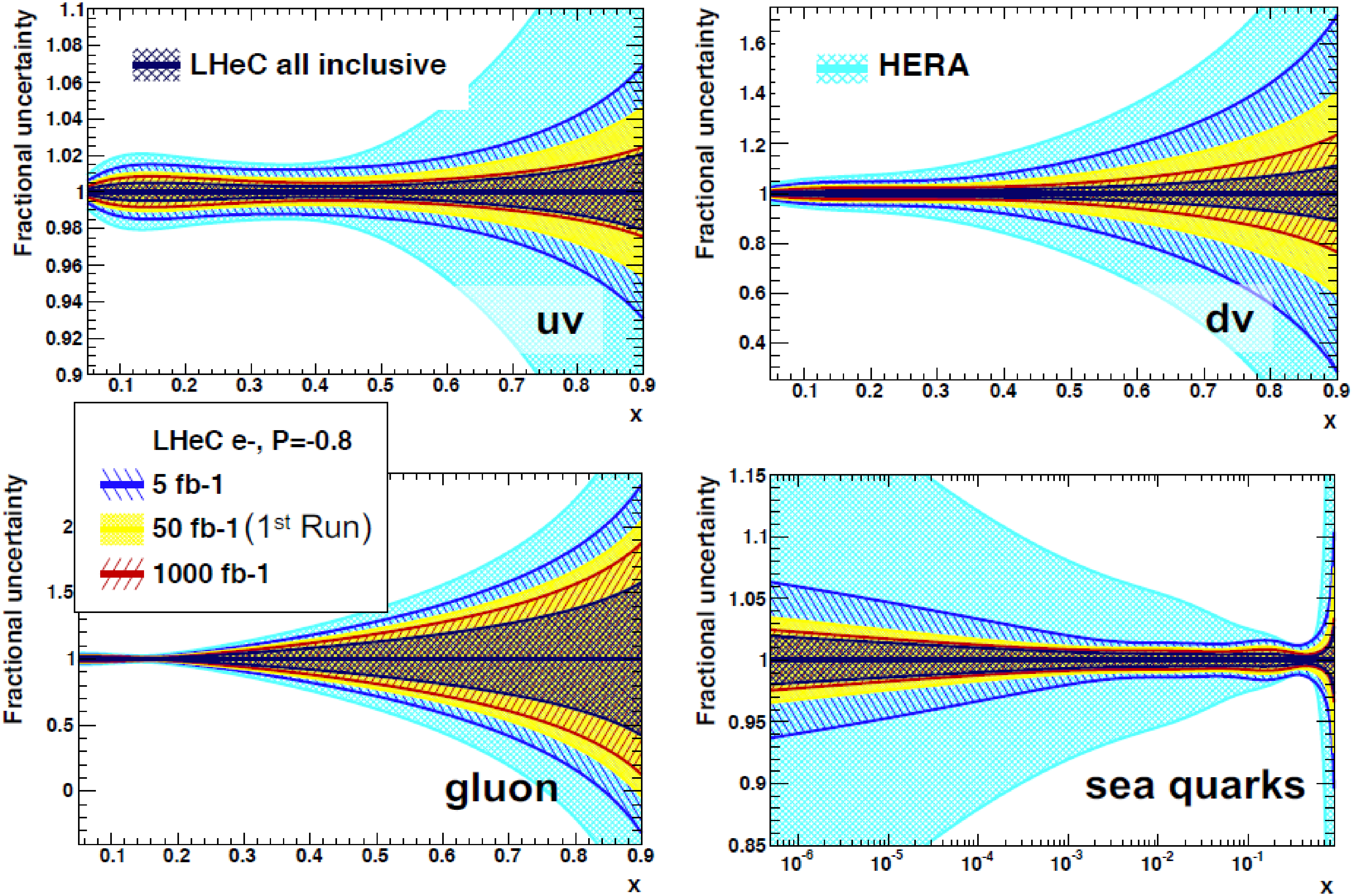
LHeC sensitivity at large x comes as part of overall package

high luminosity (×50–1000 HERA);
fully constrained quark pdfs; small x;
momentum sum rule

gluon and sea intimately related

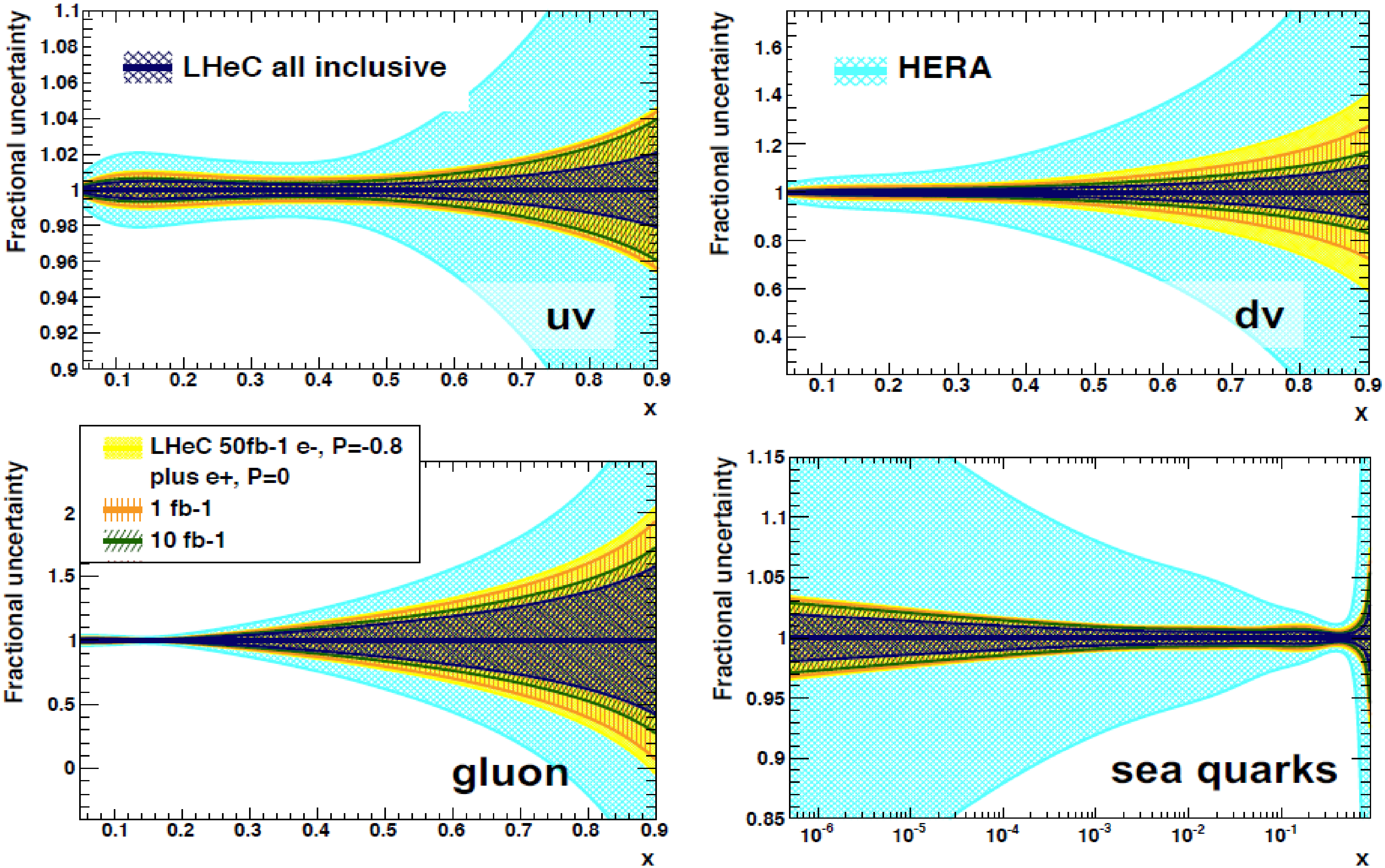
LHeC can disentangle sea from valence quarks at large x, with precision measurements of **CC** and **NC** $F2^{\gamma Z}$, $xF3^{\gamma Z}$

Impact of luminosity on PDFs



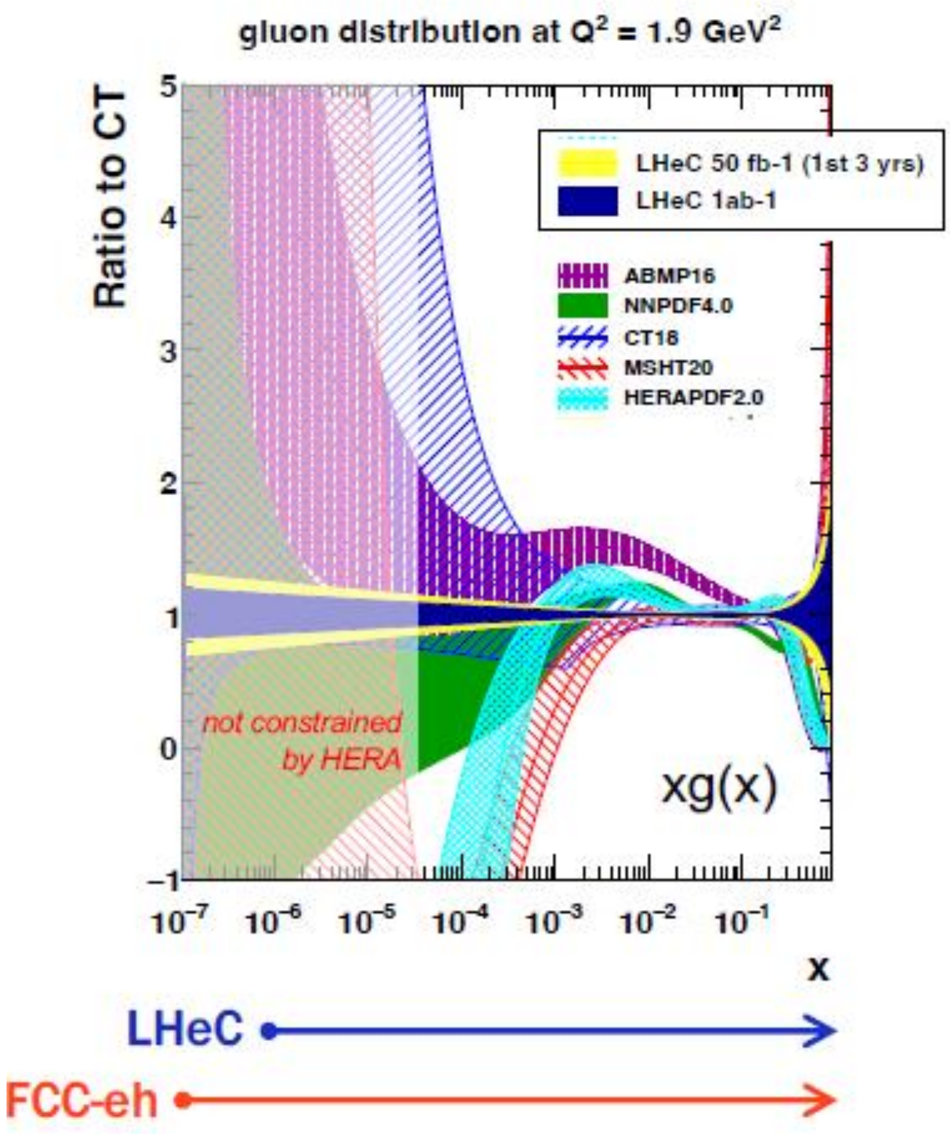
small and medium x quickly constrained (5 fb-1 ≡ ×5 HERA ≡ 1 year LHeC)

Impact of positrons on PDFs



CC: e+ sensitive to d; **NC**: e± asymmetry gives $xF3^{\gamma Z}$, sensitive to valence

Gluon at small x



no current data much below $x=5\times10^{-5}$

LHeC provides single, precise and unambiguous dataset down to $x=10^{-6}$

FCC-eh probes to even smaller $x=10^{-7}$

explore low x QCD:

DGLAP vs BFKL; non-linear evolution; gluon saturation; implications for ultra high energy neutrino cross sections

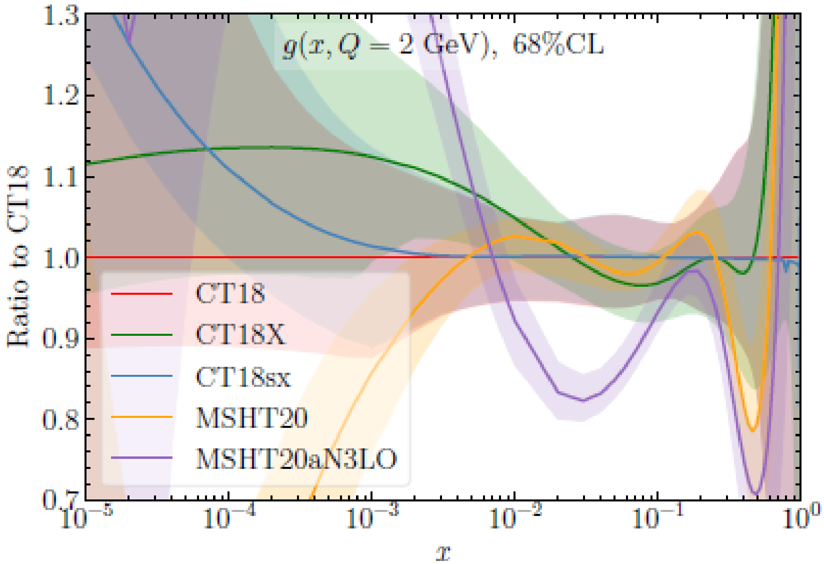
Do not be complacent in thinking that this region does not affect you…
PDFs are going to N3LO –
where the first of the BFKL (ln(1/x) resummation) terms matter..

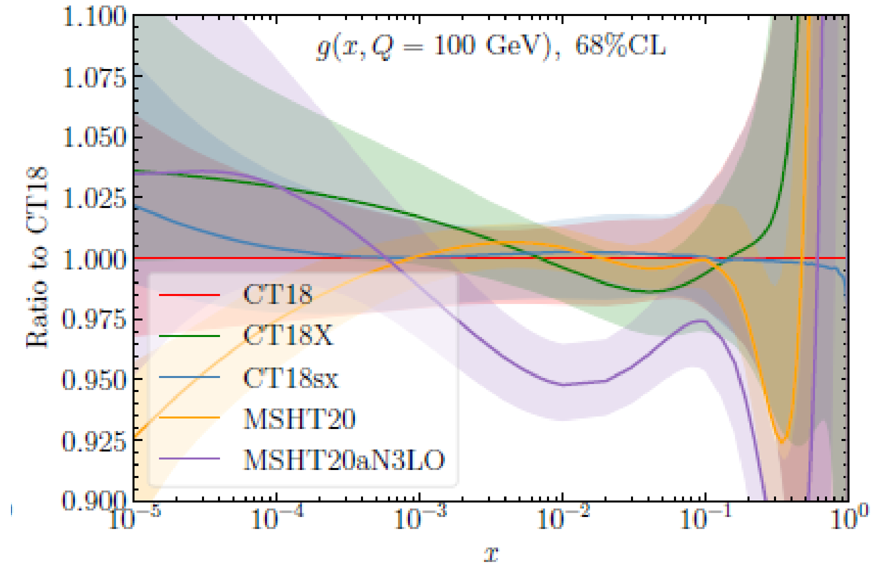
We now have N3LO predictions..

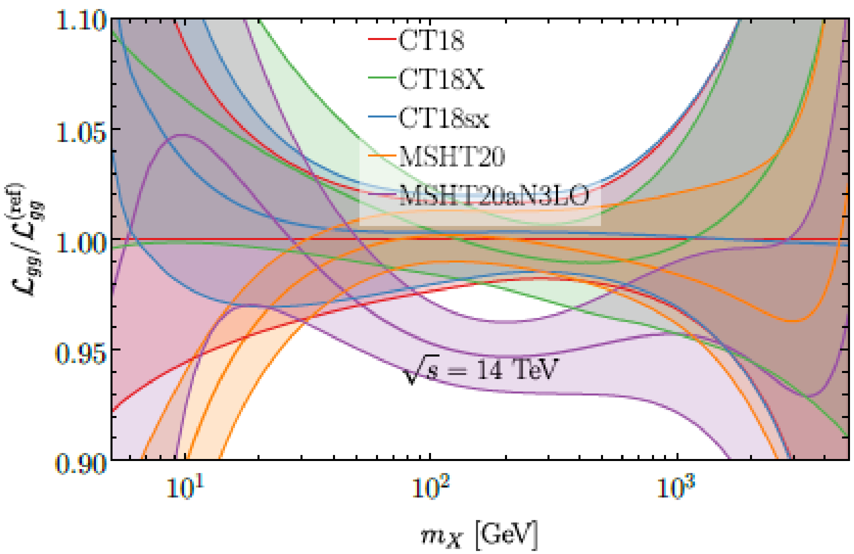
-----Well at least approximately

This has an astounding effect on the low-x gluon at low scales

Which persists to LHC scales







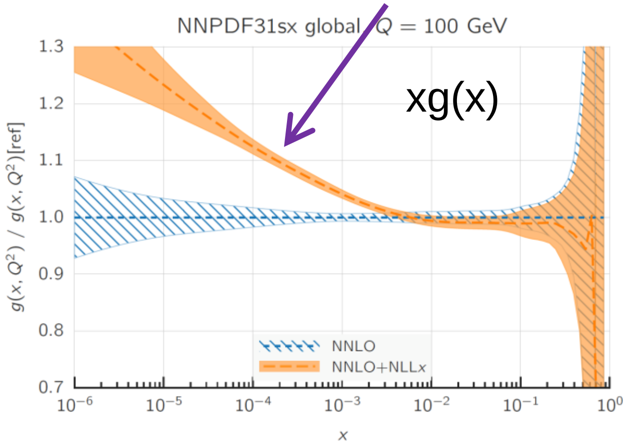
Contrast the MSHT20 NNLO
With the MSHT20aN3LO

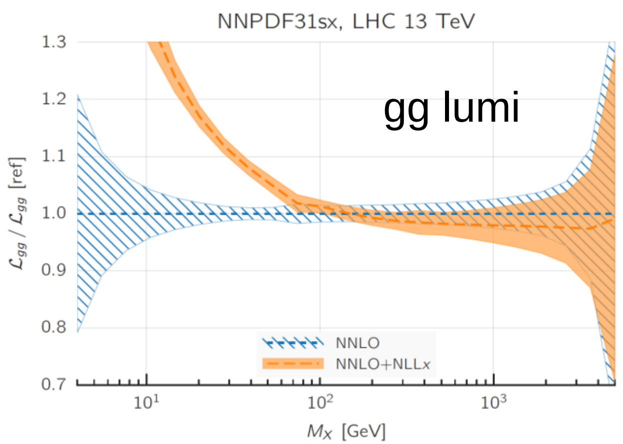
More alarming is the ‘knock-on’ effect on the gluon-gluon luminosity -- a decrease of ~5% at the Higgs scale

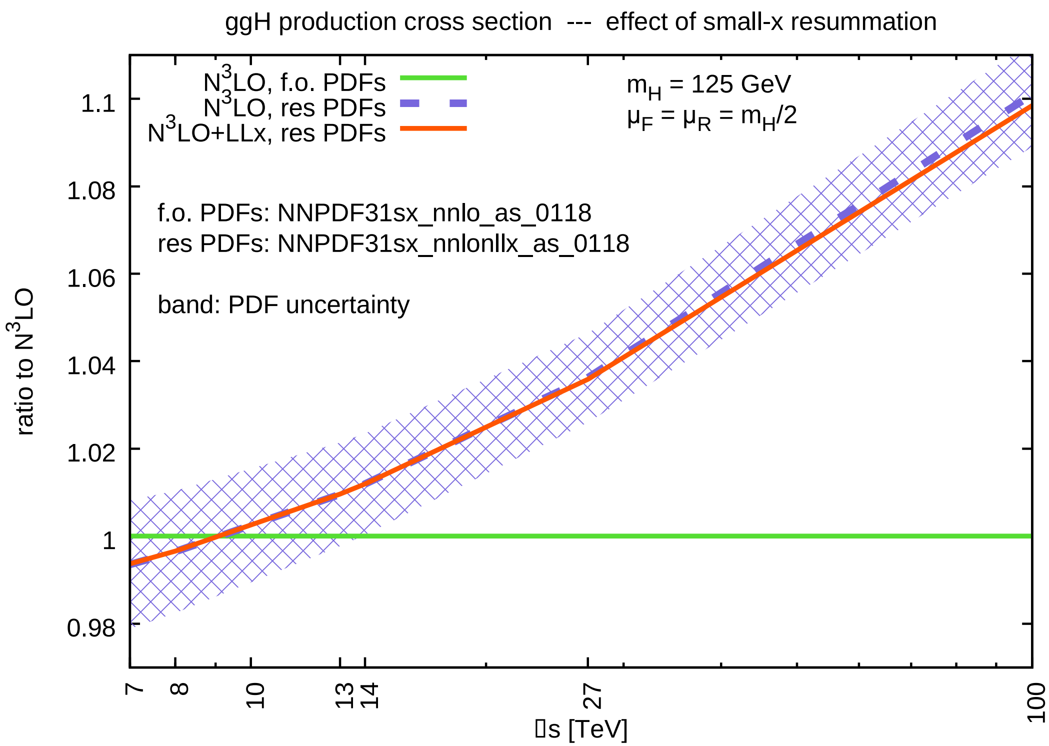
Full ln(1/x) resummation

R. Ball et al, arXiv:1710.05935

effect of small x resummation



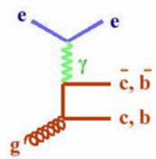




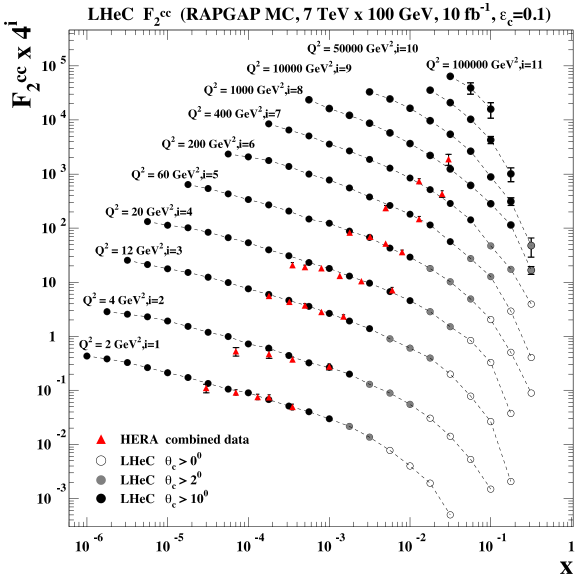
effect of small x resummation on ggH cross section for LHC, HE-LHC, FCC

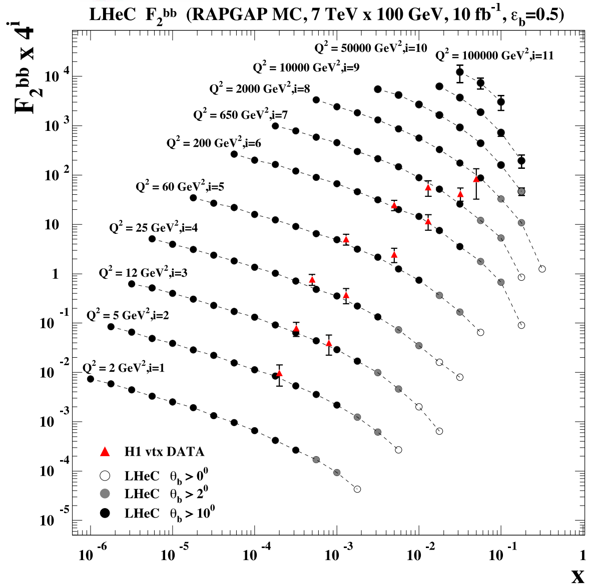
confirmed in xFitter study, arXiv:1802.00064

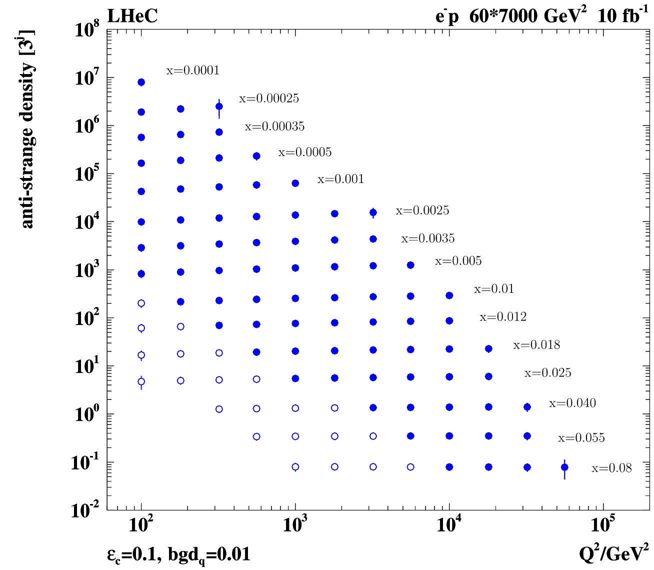
c, b quarks



strange







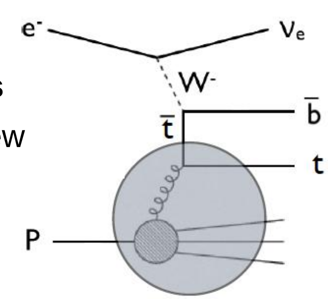
LHeC: enormously extended range and much improved precision c.f. HERA

- **δMc = 50 (HERA)** to **3 MeV**: impacts on αs, regulates ratio of charm to light, crucial for precision t, H
- **δMb** to **10 MeV**; MSSM: Higgs produced dominantly via bb → A

-

also top PDF

top quark becomes light at large Q^2: new field of research opens for top PDFs!

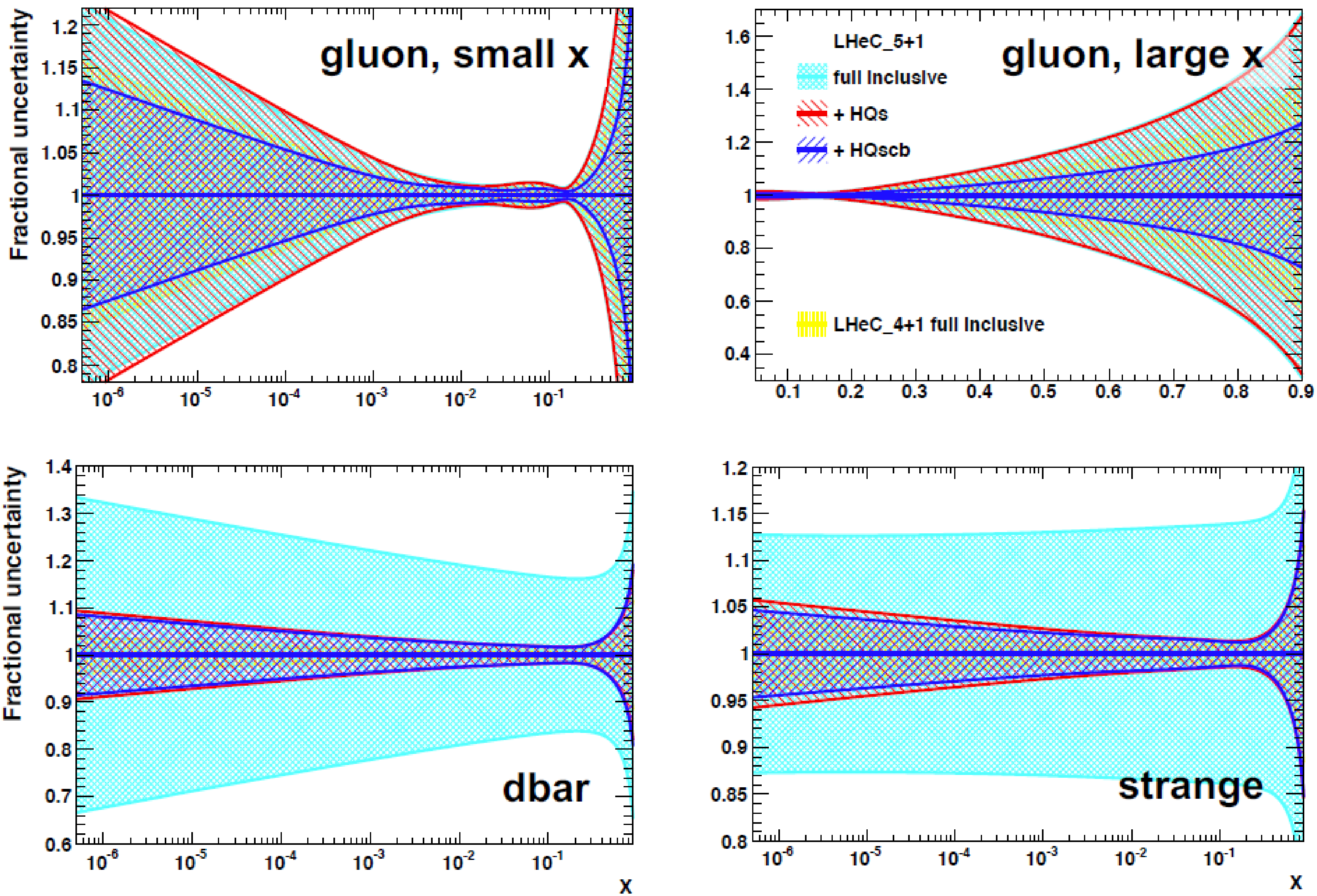


strange pdf poorly known; how suppressed cf. other light quarks? s ≠ sbar ?

LHeC: direct sensitivity to **strange** via W+s → c

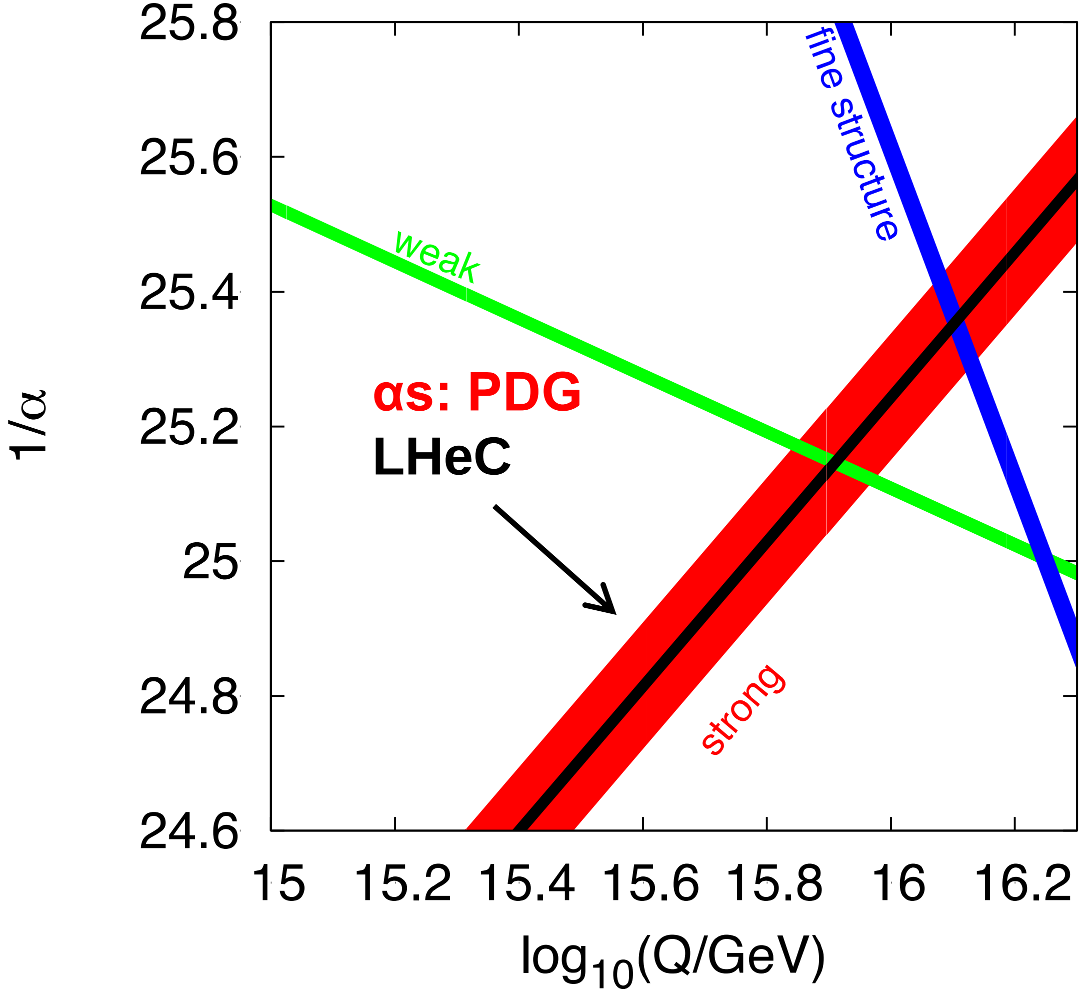
(x,Q^2) mapping of (anti) strange for first time

impact of HQ data on LHeC pdfs



more flexible parameterisation (5+1): **xuv, xdv, $x\bar{U}$, $x\bar{d}$, $x\bar{s}$** and **xg**

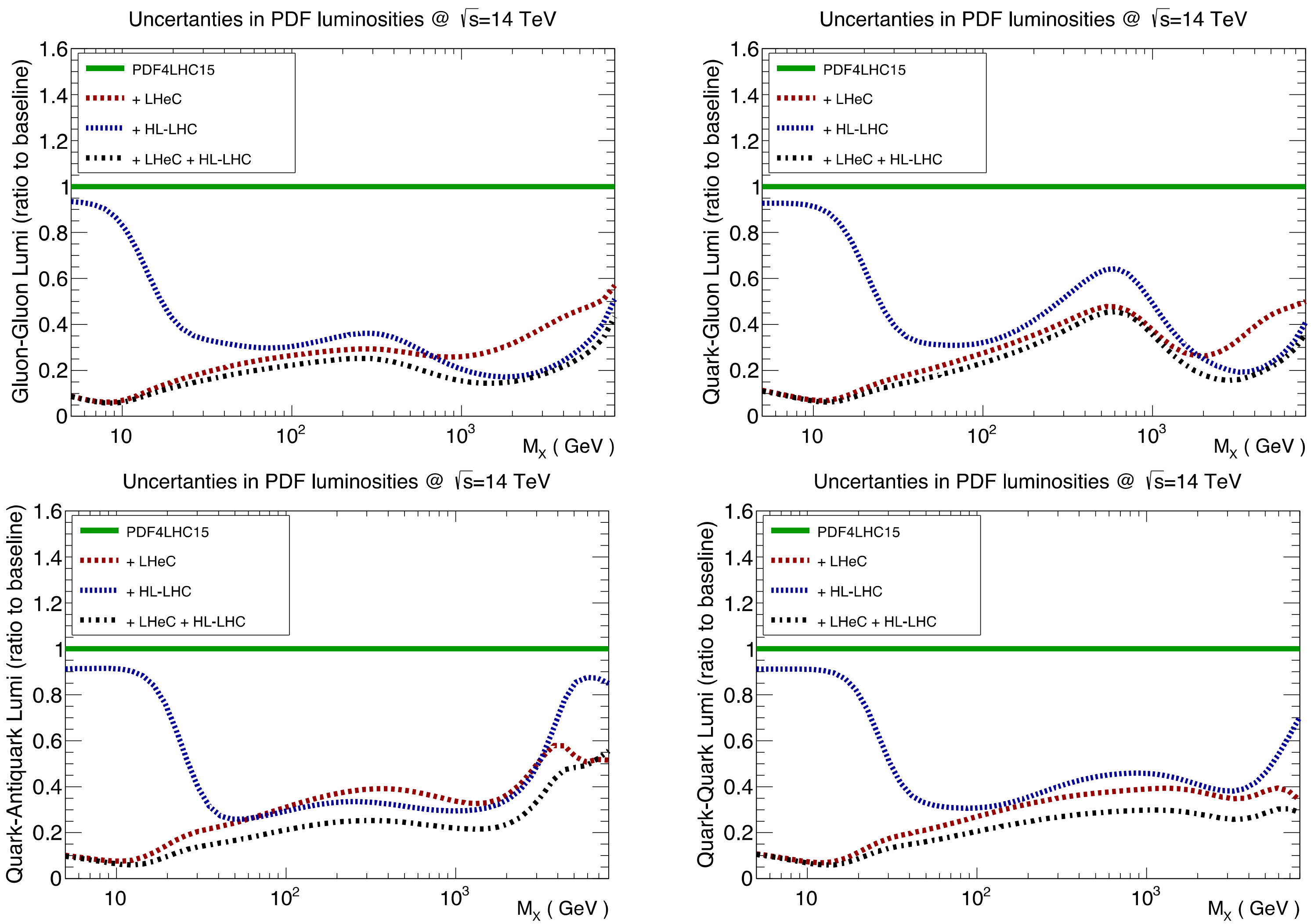
And there will be further information from jet production at the LHeC…… which will mostly contribute to the precision of the gluon PDF and thus to future determinations of strong coupling, α_S (MZ)



precise α_s needed: to constrain GUT scenarios; for cross section predictions, including Higgs; …

LHeC: permille precision possible in **combined QCD fit for pdfs+α_s**

Just in case you worry that a study of LHeC improvements based on a simple HERAPDF procedure may be optimistic. A study was done comparing future improvements from the HL-LHC to those from the LHeC in an 'apples to apples' manner. Profiling the PDF4LHC15 with HL-LHC pseudo-data or LHeC pseudo-data With consistent tolerance T=3



Abdul Khalek et al
arXiv:1810.03639
+ 1906.10127

Summary/ Things to think about

- PDF improvement is not just a matter of more data
- Consistency of data matters
- Knowledge of common systematic uncertainties matters
- Real data are always more problematic than pseudo-data projections
- Real data from a DIS machine would be more self consistent – a single team would analyse the whole kinematic region producing a consistent set of correlated systematic uncertainties----we have learnt our lessons at HERA
- and theoretically cleaner + less subject to new physics contamination at high scale

- Differences in the PDFs are not just about choice of data set—PDF4LHC comparisons of MSHT,CT,NN using the same reduced data sets still have some differences--
- There are some irreducible methodological differences between the PDFs
- But supplying new high precision DIS data set would certainly bring them closer

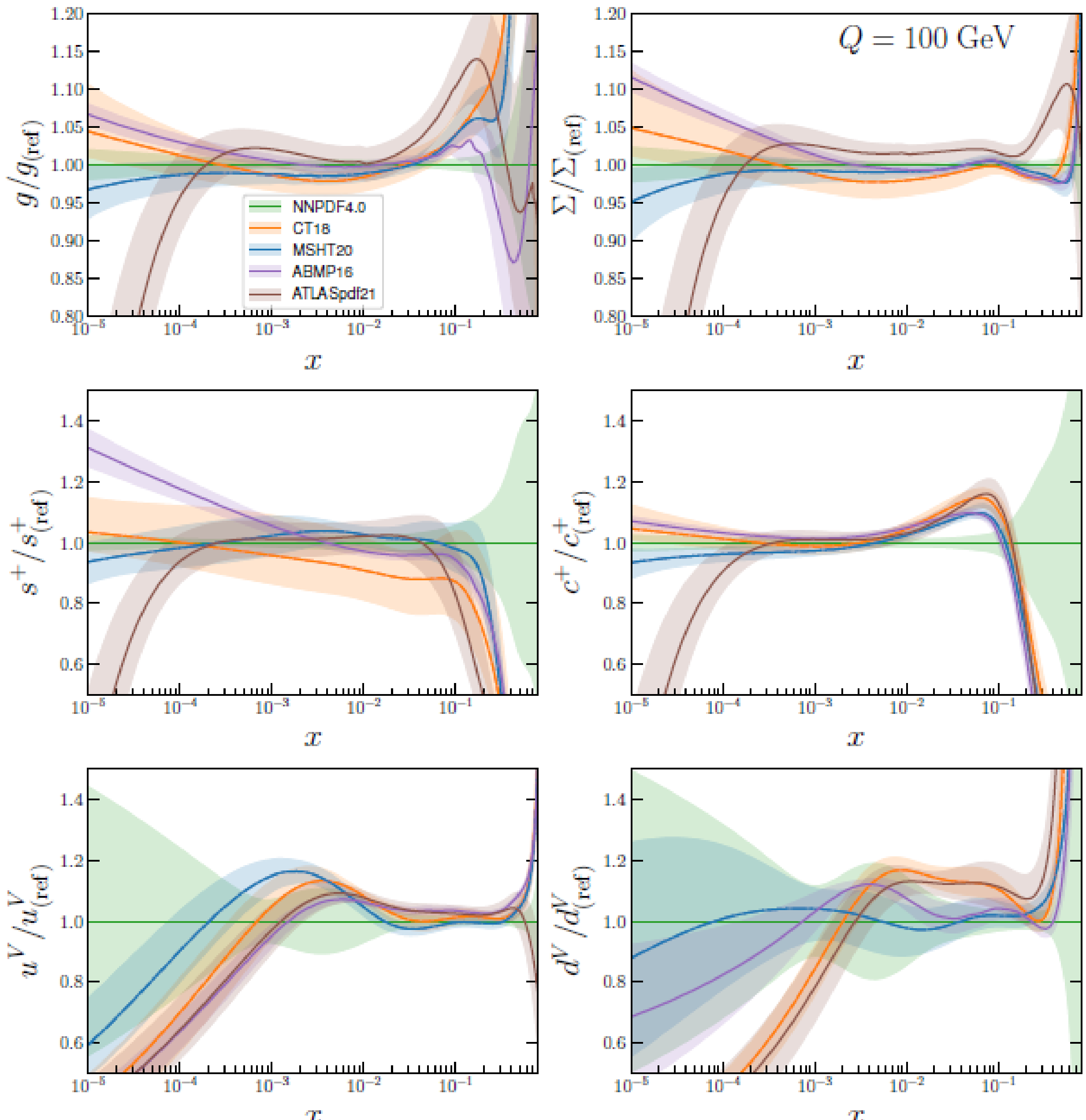
- N3LO, Ln(1/x) resummation, recombination/saturation

Clearly we are going to have to consider all these

- and it is NOT irrelevant at the Higgs scale

Backup

A closer look at modern PDFs going down to VERY low-x for Q=100, central LHC probes only down to $x \sim 10^{-3}$



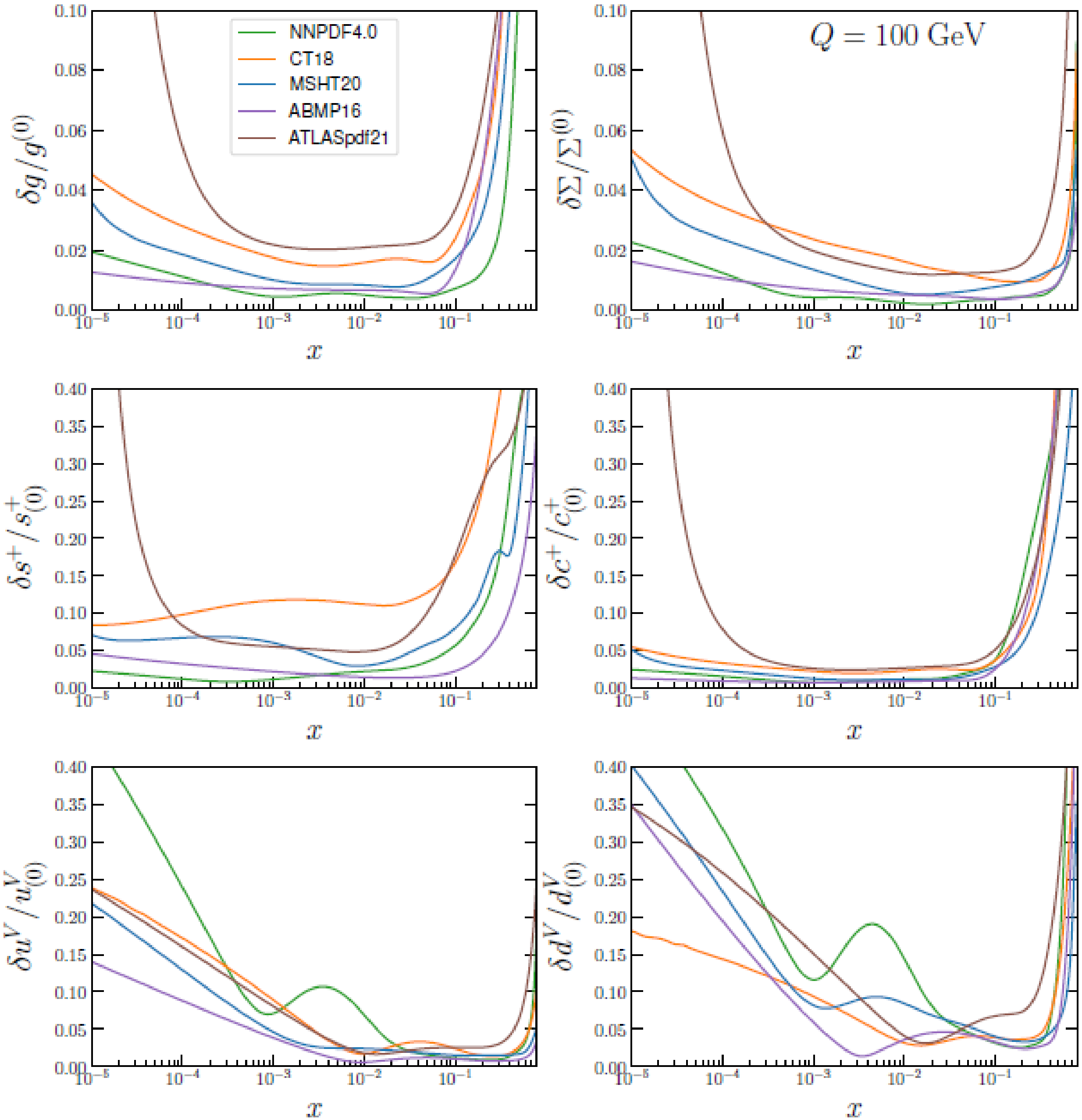
In ratio to NNPDF4.0

We are not so surprised by differences at high-x, though they can be outside uncertainties

e.g.NNPDF has intrinsic charm. But also less strange suppression

Differences in low-x valence are also unsurprising, when little is known on valence at very low-x

Let us look at low-x gluon



But first look at uncertainties

NOTE ABMP16 is relatively small in regions where similar amounts of data are used, because $\Delta\chi 2=1$ is used rather than a higher tolerance

ATLASpdf21 is larger at low and small x because less data are used

CT18 is often the larger of CT, MSHT because of a larger tolerance than MSHT

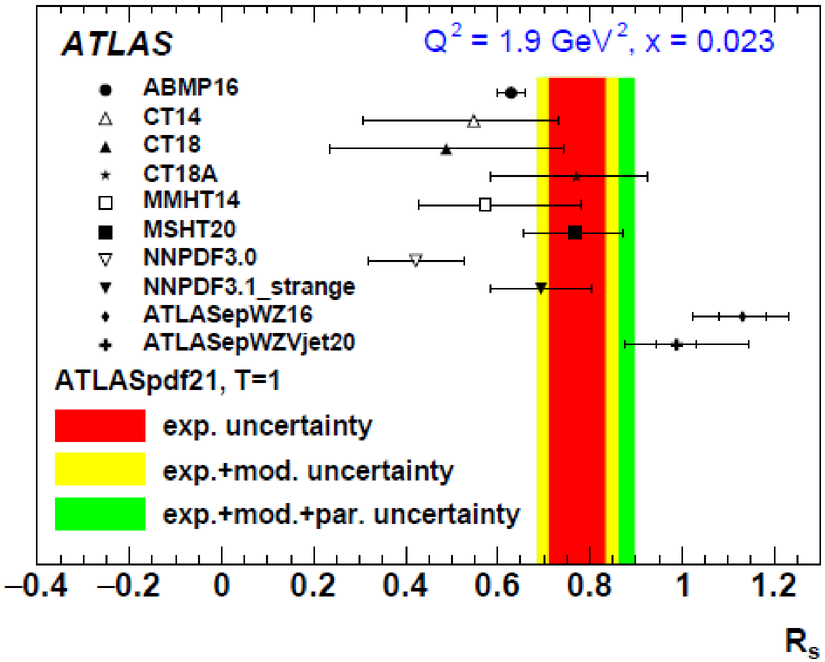
NNPDF4.0 has generally very small uncertainties in the data region--- new procedure, positivity, integrability etc..

Strangeness

The information on strangeness has often been presented at a single x,Q^2 point and compared to the result of global PDFs

Note that older PDFs CT14, MMHT14,NNPDF3.0 all had Rs~0.5 at low scale (Q^2=1.9GeV^2) **BUT** this has moved up to ~0.8 for CT18A, MSHT20, and NNPDF3.1_strange after ATLAS W,Z 7 TeV data was included (not for CT18 which does not include these data)

ATLAS older fits had Rs~1.0 and have moved down to Rs~0.8 due to input of new data, V+jets and W,Z 8 TeV and greater flexibility of low-x parametrisation



The history of these changes is:

ATLASepWZ16 to ATLASepWZVjets20
Input of V+jets data suppresses Rs at high-x and this has a knock-on effect at x=0.023 so that Rs~ 1.15 →1.0

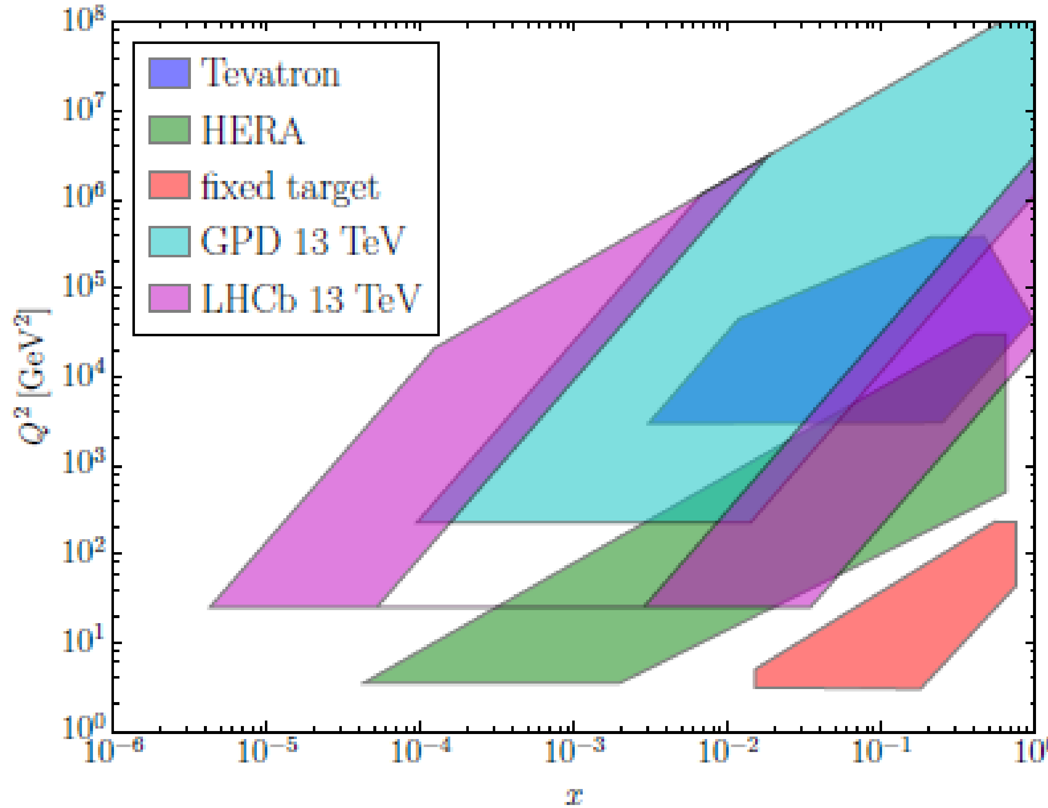
ATLASepWZVJets20 to ATLASpdf21

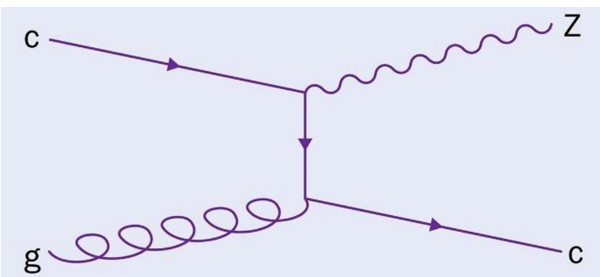
- More flexible low-x parametrisation corresponds to lower edge of Vjets20 error-band Rs~ 1.0 →0.85
- Addition of W,Z 8 TeV data Rs~ 0.85 →0.8

Flavour 2. Charm

The kinematic reach of LHCb goes to both higher and lower rapidity and hence to higher and lower x than CMS or ATLAS (labelled as GPD general purpose detectors)

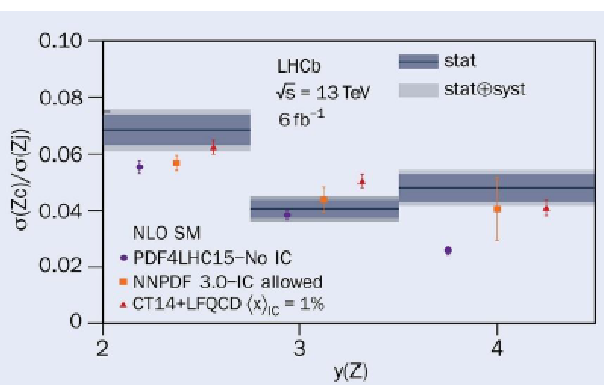
Hence they may be able to look into intrinsic charm in the nucleon

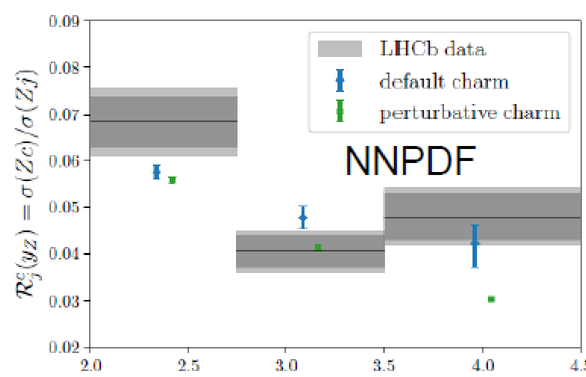




Clearly you need Z+c or γ+c

It's probably smart to take ratios like Z+c/Z+jets and do it at high rapidity to reach high x





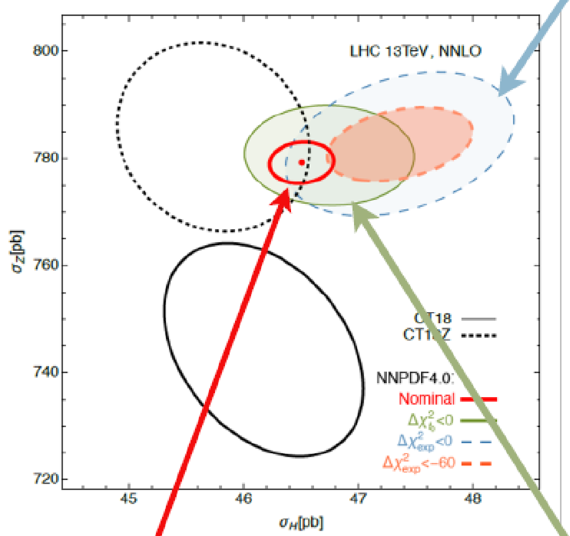
14

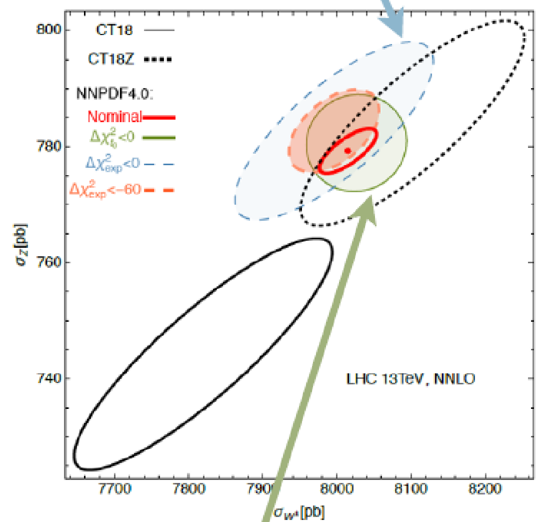
Monte-Carlo sampling sensitivity for PDFs

Debate on NNPDF4.0 uncertainties

Regions containing (very) good solutions according to the experimental form of χ^2

(is used in χ^2 summary tables of the NN4.0 article, was a default in the NN4.0 public code)





CT criticism
ArxiV:2205.10444

Nominal NN4.0 Hessian or MC 68%cl

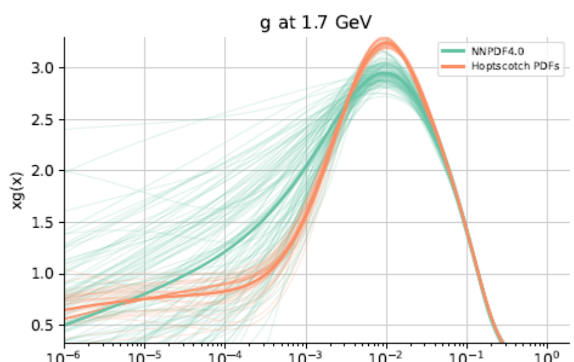
Region containing good solutions according to the most recent t_0 form of χ^2 (used to train NN4.0 replicas)

NNPDF prior disfavors some of the most displaced hopscotch replicas.
[M. Ubiali, HP2 2022 workshop, Durham, 2022/09]

The hopscotch replicas pass CT18 criteria for good individual solutions.

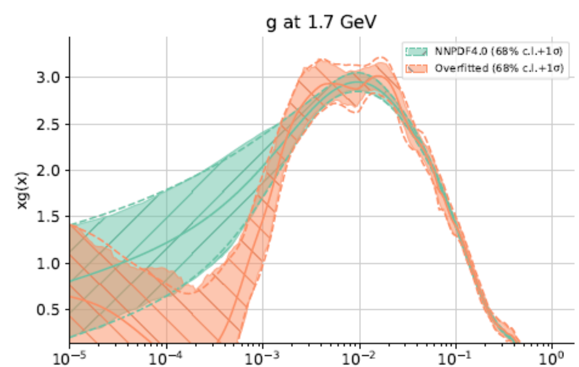
Observation 1

A comparison between NNPDF replicas and HS PDFs reveals a peculiar kink in HS PDFs



Observation 2

An overfitted gluon can be obtained with $\chi^{2(0,c)}$ smaller than NNPDF4.0 by $0.08 \times N_{\rm dat}$

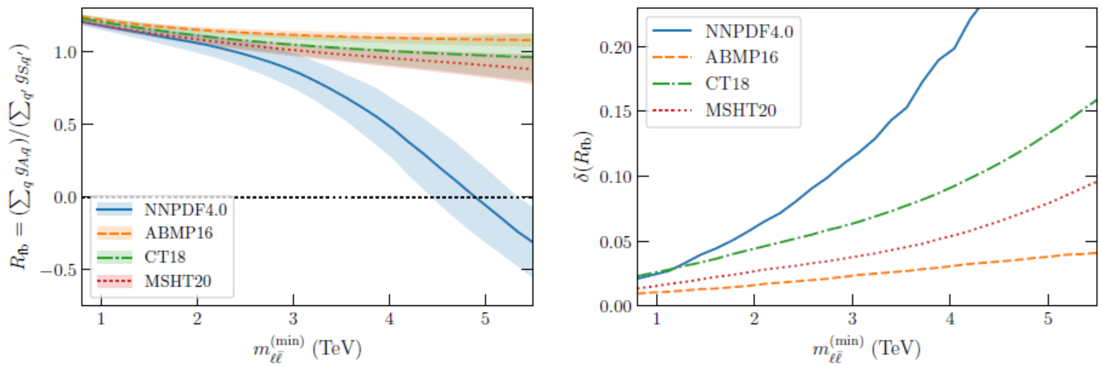


NNPDF reply at PDF4LHC2022… but it goes on

Comparisons at very high-x / High scale
AFB is very different for NNPDF4.0
NNPDF4.0 uncertainties remain large/largest beyond the current data region– but not large enough to cover this

Positive or negative asymmetry?

$$A_{\rm fb}(\cos\theta^*) \overset{\rm LO}{=} \frac{\cos\theta^*}{1+\cos^2(\theta^*)}\frac{g_A}{g_{S,q'}} \qquad g_{A,S} \propto \int \frac{dm_{\ell\bar{\ell}}}{m_{\ell\bar{\ell}}}\frac{dx_1}{x_1}\mathcal{L}_{A,S}(m_{\ell\bar{\ell}},x_1)$$



As the scale increases, larger x is probed

ABMP16, CT18 and MSHT20:
approximately scale independent

NNPDF: the coupling depends on the scale, larger uncertainty

Expectation: for NNPDF4.0, $A_{\rm fb}$ vanishes at large $m_{\ell\bar{\ell}}$

QCD fit parameterisation

QCD fit ansatz based on HERAPDF2.0, with following differences

much more relaxed sea ie. no requirement that ubar=dbar at small x

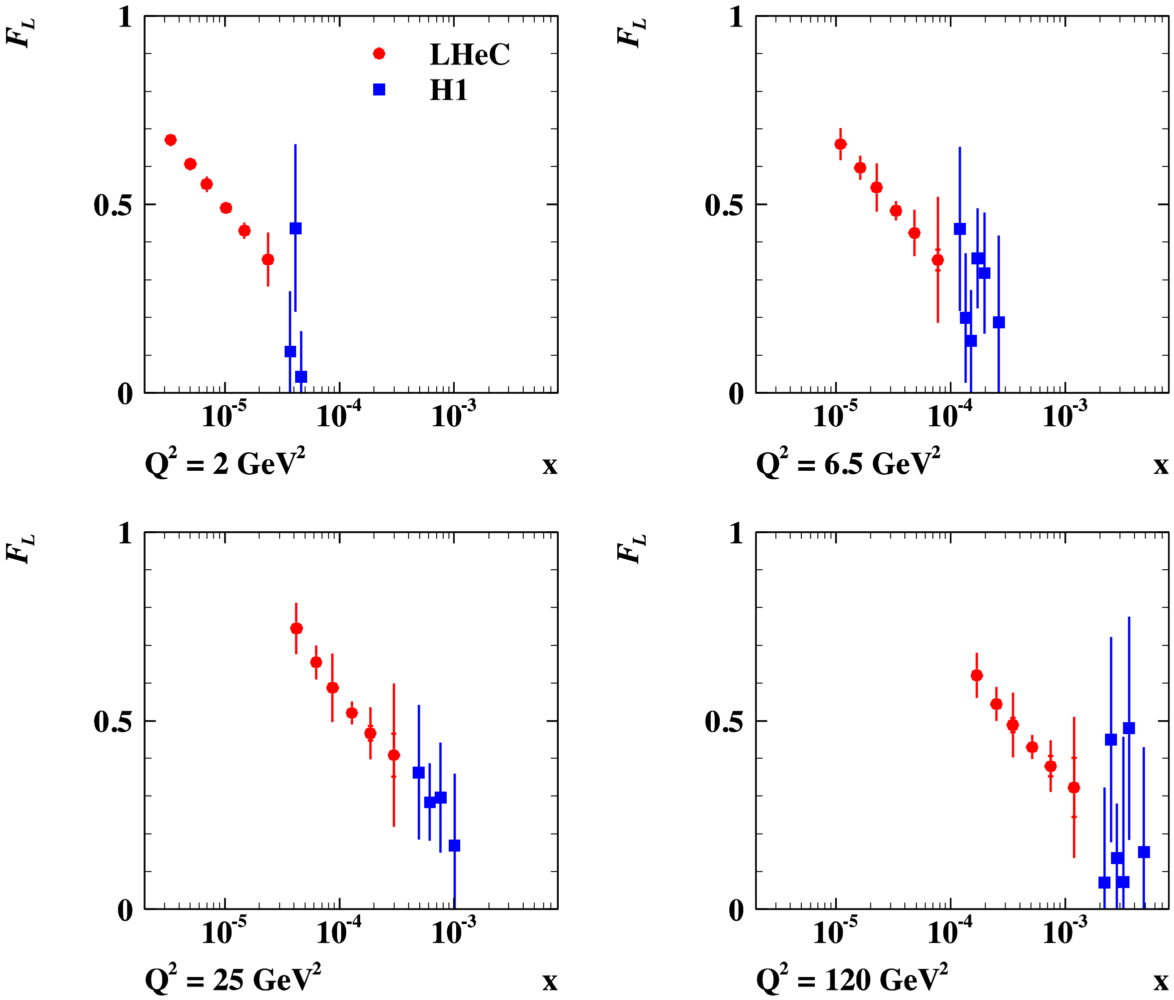
no negative gluon term (simply for the aesthetics of ratio plots – it has been checked that this does not impact size of projected uncertainties)

$$
\begin{aligned}
xg(x) &= A_g x^{B_g} (1-x)^{C_g} (1 + D_g x) \\
xu_v(x) &= A_{u_v} x^{B_{u_v}} (1-x)^{C_{u_v}} (1 + E_{u_v} x^2) \\
xd_v(x) &= A_{d_v} x^{B_{d_v}} (1-x)^{C_{d_v}} \\
x\bar{U}(x) &= A_{\bar{U}} x^{B_{\bar{U}}} (1-x)^{C_{\bar{U}}} \\
x\bar{D}(x) &= A_{\bar{D}} x^{B_{\bar{D}}} (1-x)^{C_{\bar{D}}}
\end{aligned}
$$

4+1 pdf fit (above) has 14 free parameters

5+1 pdf fit for HQ studies parameterises dbar and sbar separately, and has 17 free parameters

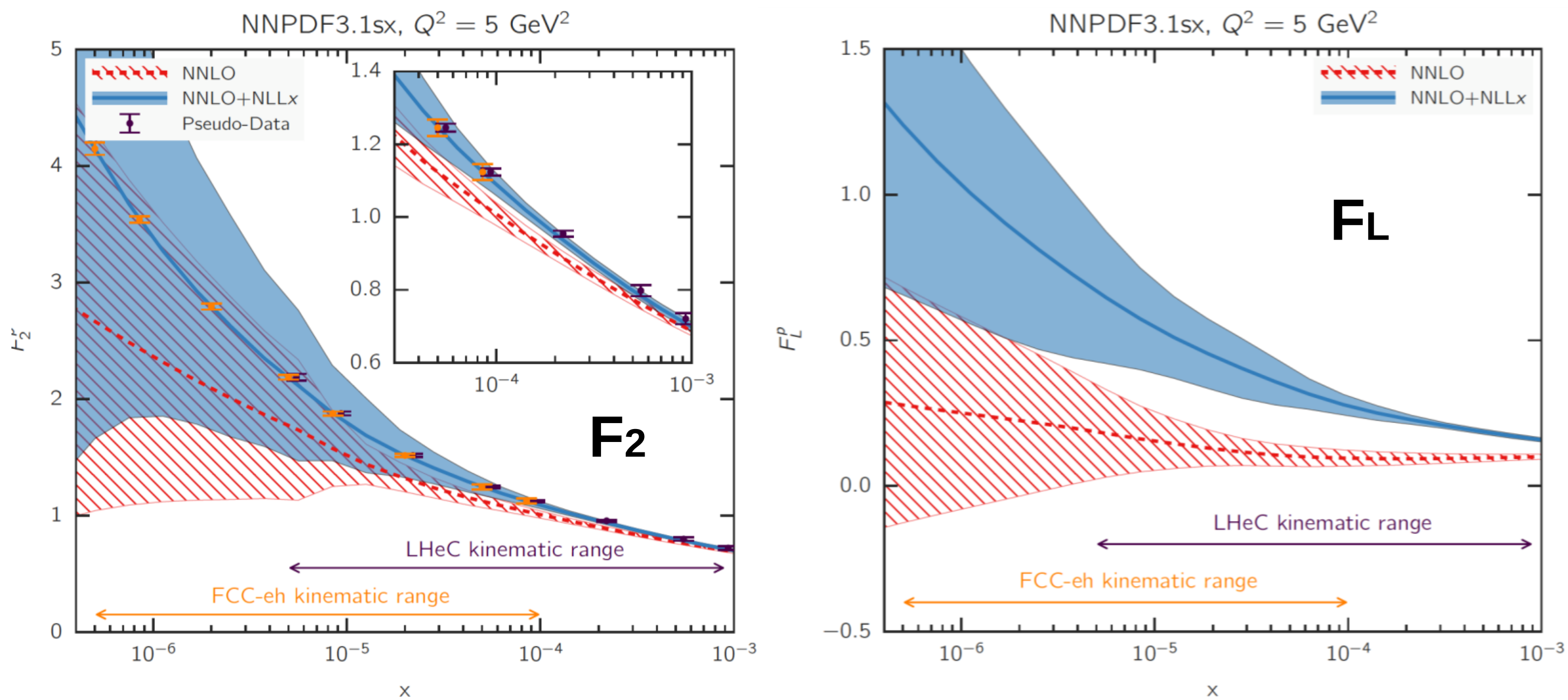
FL at LHeC



M. Klein, arXiv:1802.04317

gluon at small x

arXiv:1710.05935



F2 and FL predictions for simulated kinematics of **LHeC** and **FCC-eh**

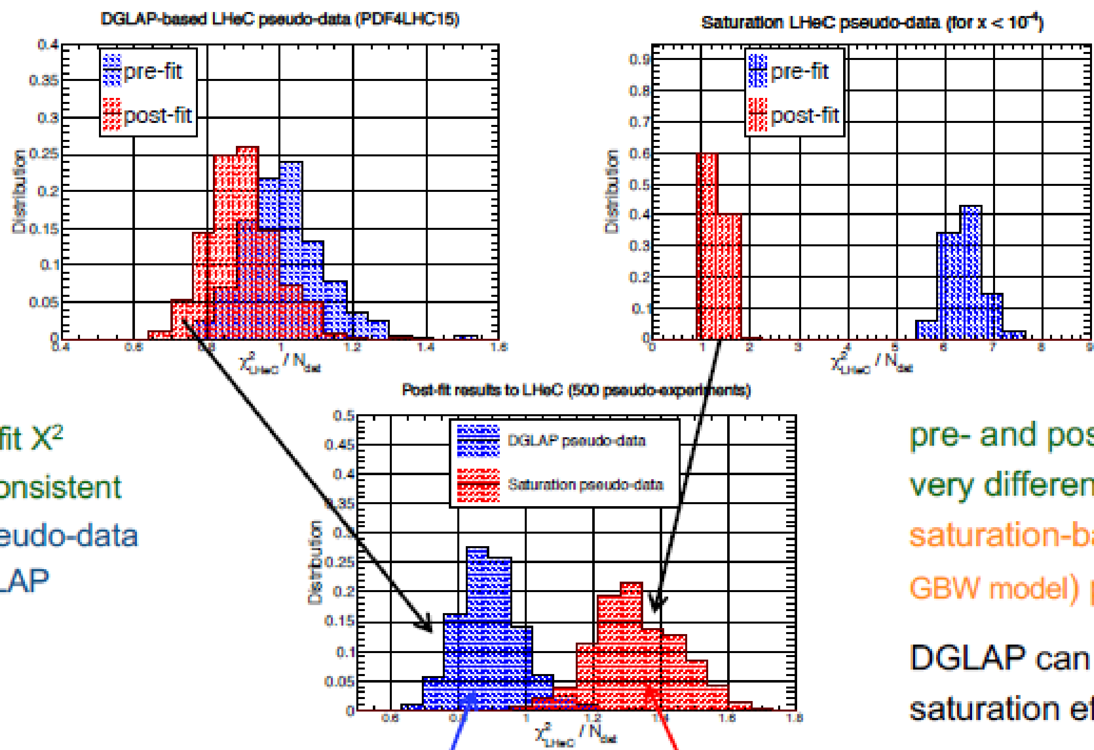
ep simulated data very precise – significant constraining power to discriminate between theoretical scenarios of small x dynamics

measurement of FL has a critical role to play

see also M. Klein, arXiv:1802.04317

Novel dynamics at small x: saturation

- studies show linear evolution **cannot accommodate saturation**, even at NNLO or NNLO+NLLx
- EG, **DGLAP-** vs **saturation-** based simulated data fitted with NNLO DGLAP



pre- and post-fit X^2 distributions consistent for DGLAP pseudo-data fitted with DGLAP

pre- and post-fit distributions very different for DGLAP fit to saturation-based ($x \leq 10^{-4}$, GBW model) pseudo-data

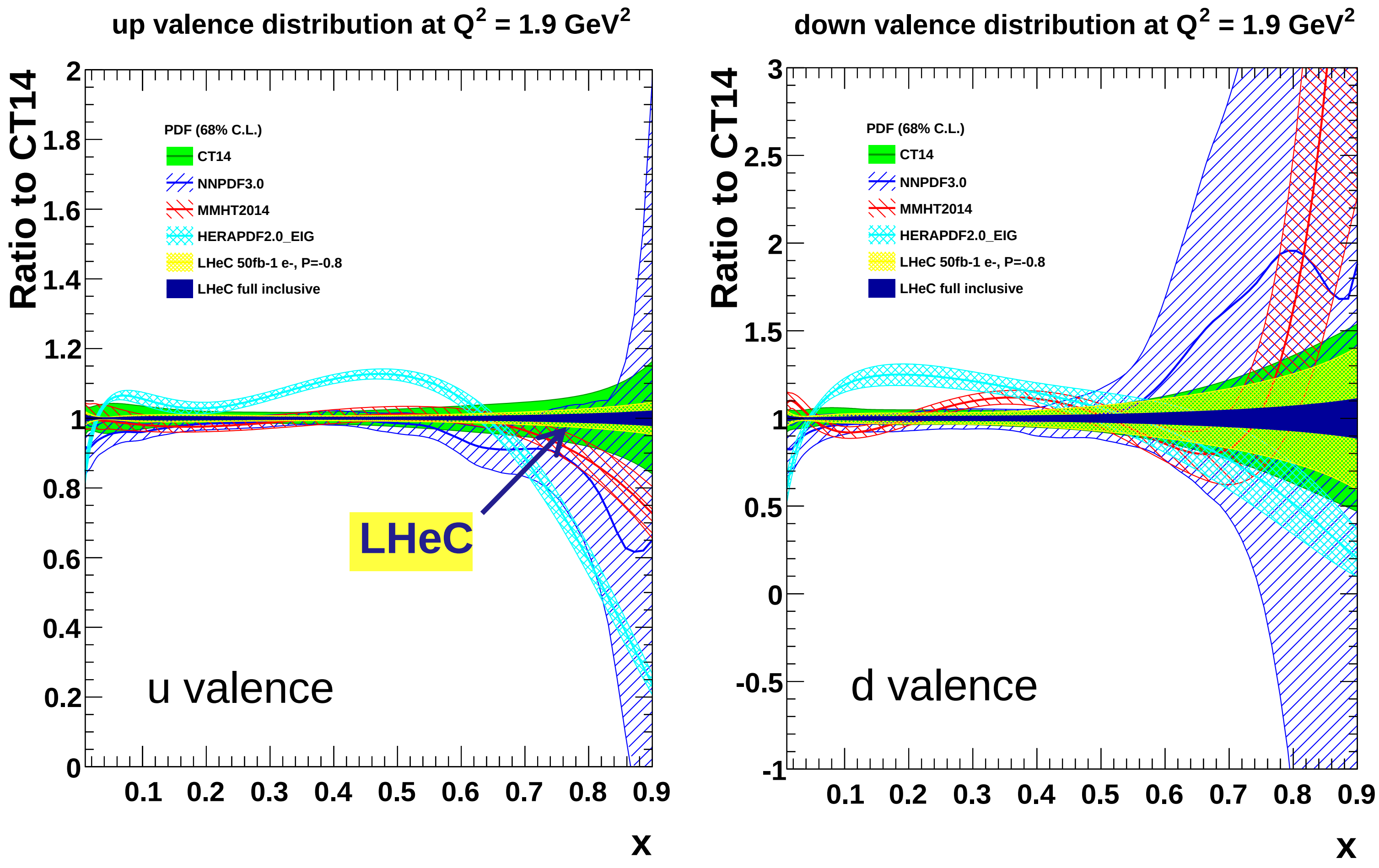
DGLAP can not absorb all saturation effects

LHeC can distinguish between **DGLAP** and **saturation**

(NB, large lever arm in Q^2 crucial, see also arXiv:1702.00839)

arXiv:2007.14491
(more detail in EXTRAS)

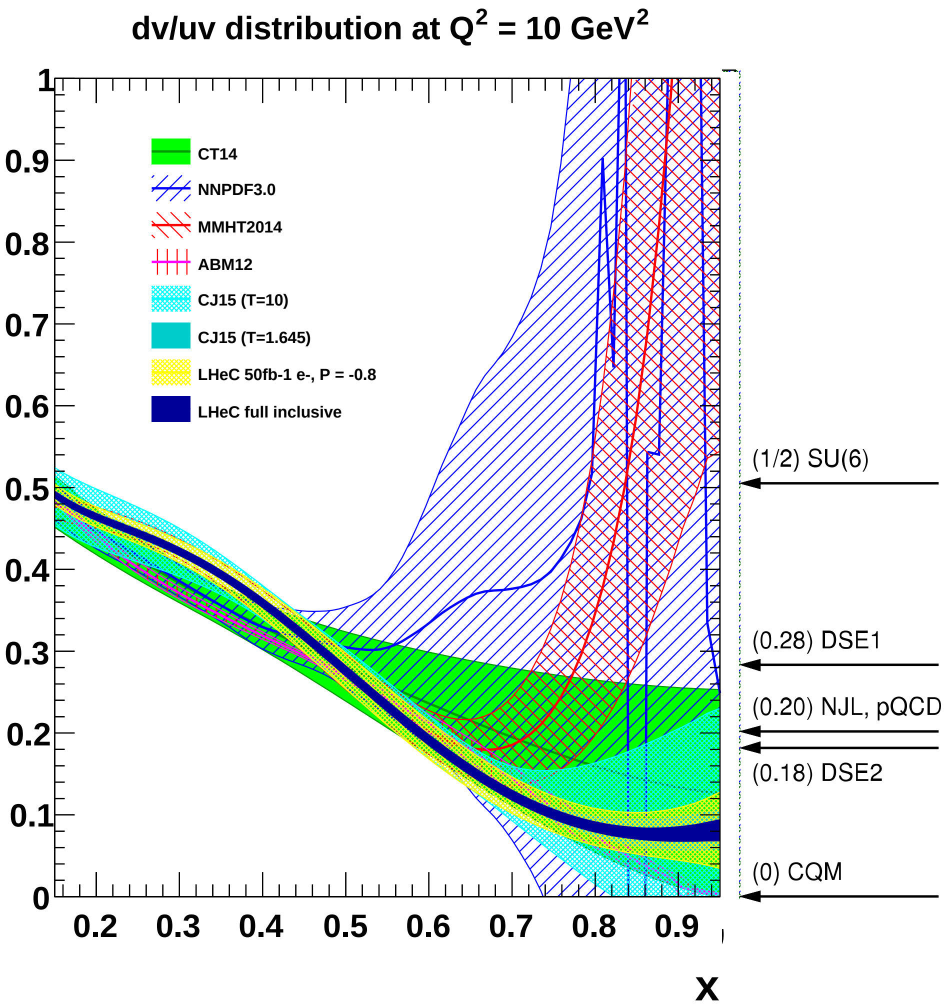
valence quarks from LHeC



precision determination, free from higher twist corrections and nuclear uncertainties

large x crucial for HL/HE–LHC and FCC searches; also relevant for DY, MW etc.

d/u at large x

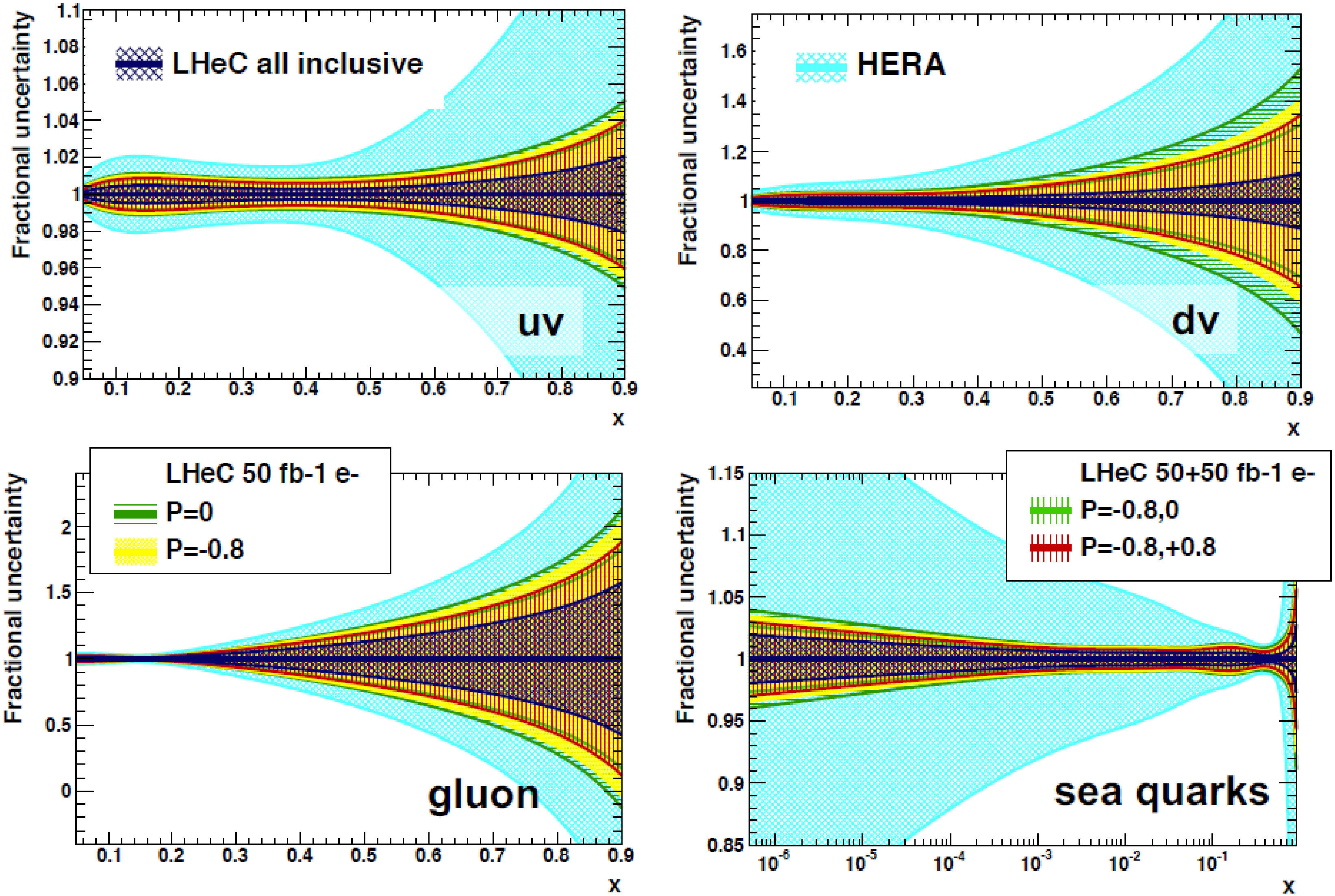


d/u essentially unknown at large x

no predictive power from current pdfs;
conflicting theory pictures;
data inconclusive, large nuclear uncerts.

resolve long-standing mystery of d/u ratio at large x

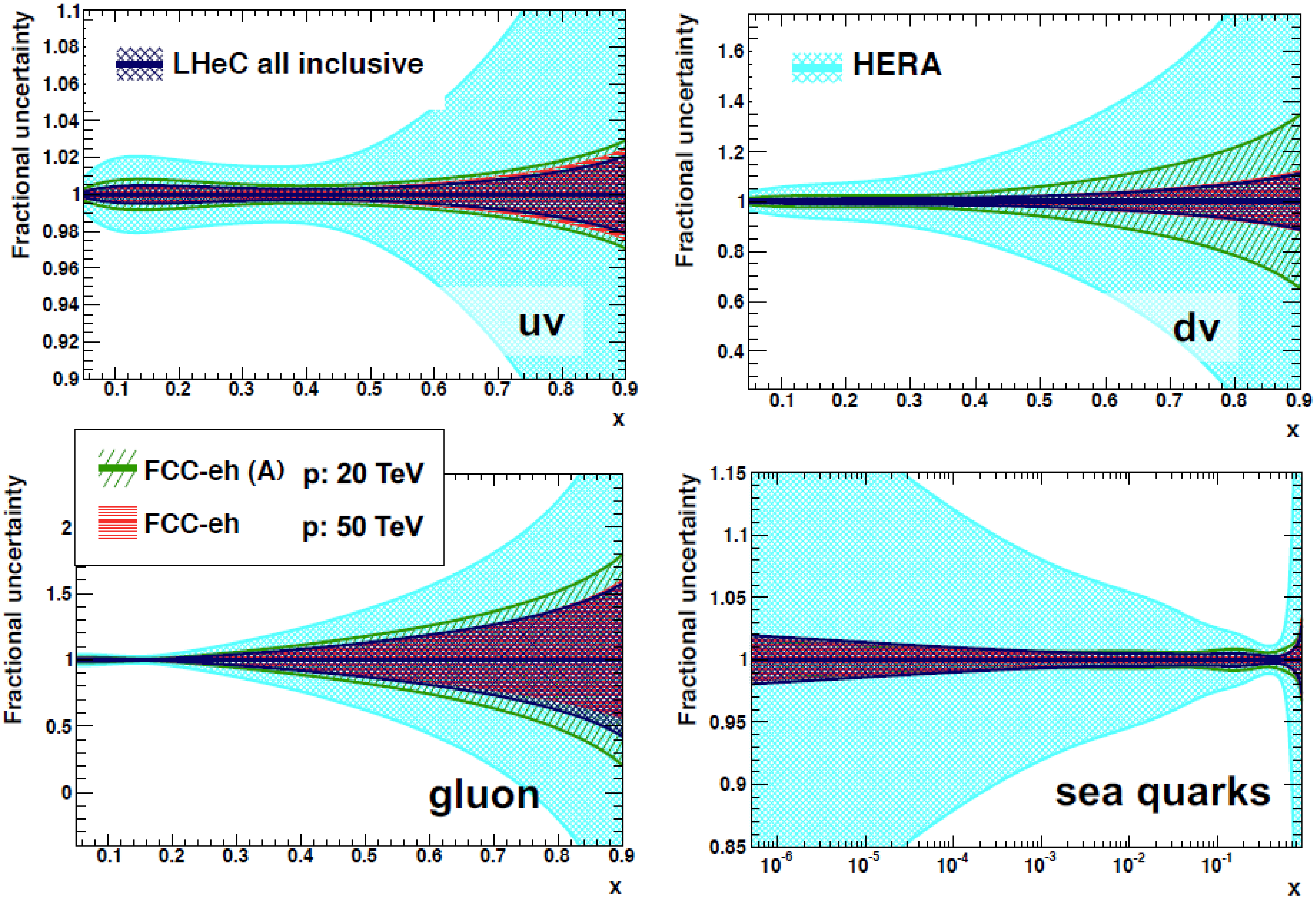
impact of polarisation on LHeC pdfs



impact of polarisation on pdfs generally small (but pol. important for ew)

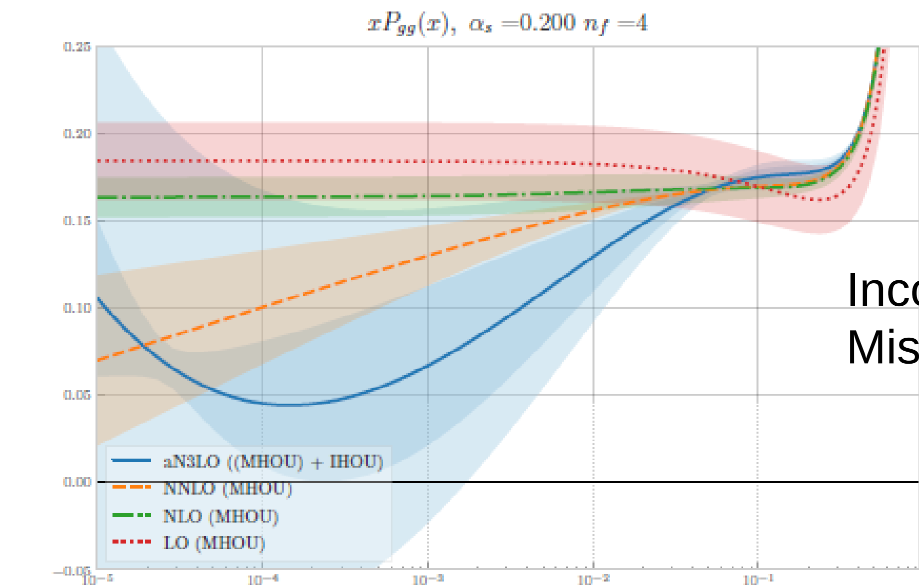
(**CC**: σ(e±) scales as (1±P) ; **NC**: effects subtle; pol. asym. gives access to $F2^{\gamma Z}$, new quark combinations)

Collider configurations

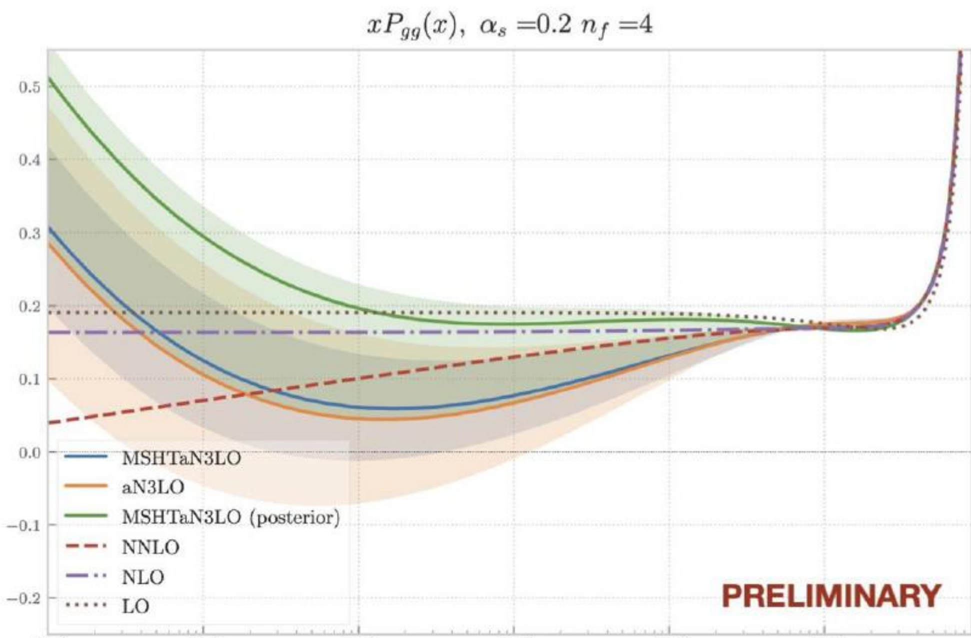


FCC-eh (A): new preliminary simulation with 2 ab^{-1} polarised e– (NO e+ yet; impact especially in d at large x)

Now NNPDF have issued some work at N3LO arXIV:2306.15294 and discussion at Les Houches 2023



Incomplete Higher Order Uncertainties (light)
Missing Higher Order Uncertainties (dark)



There is not complete agreement on N3LO between MSHT and NNPDF – may rescue the dramatic effect on gg luminosity somewhat

MSHT (prior) ≃ **NNPDF**

MSHT (posterior) shifts within uncertainty band
(absorbs some low-x logs?)

We need N3LO benchmarking