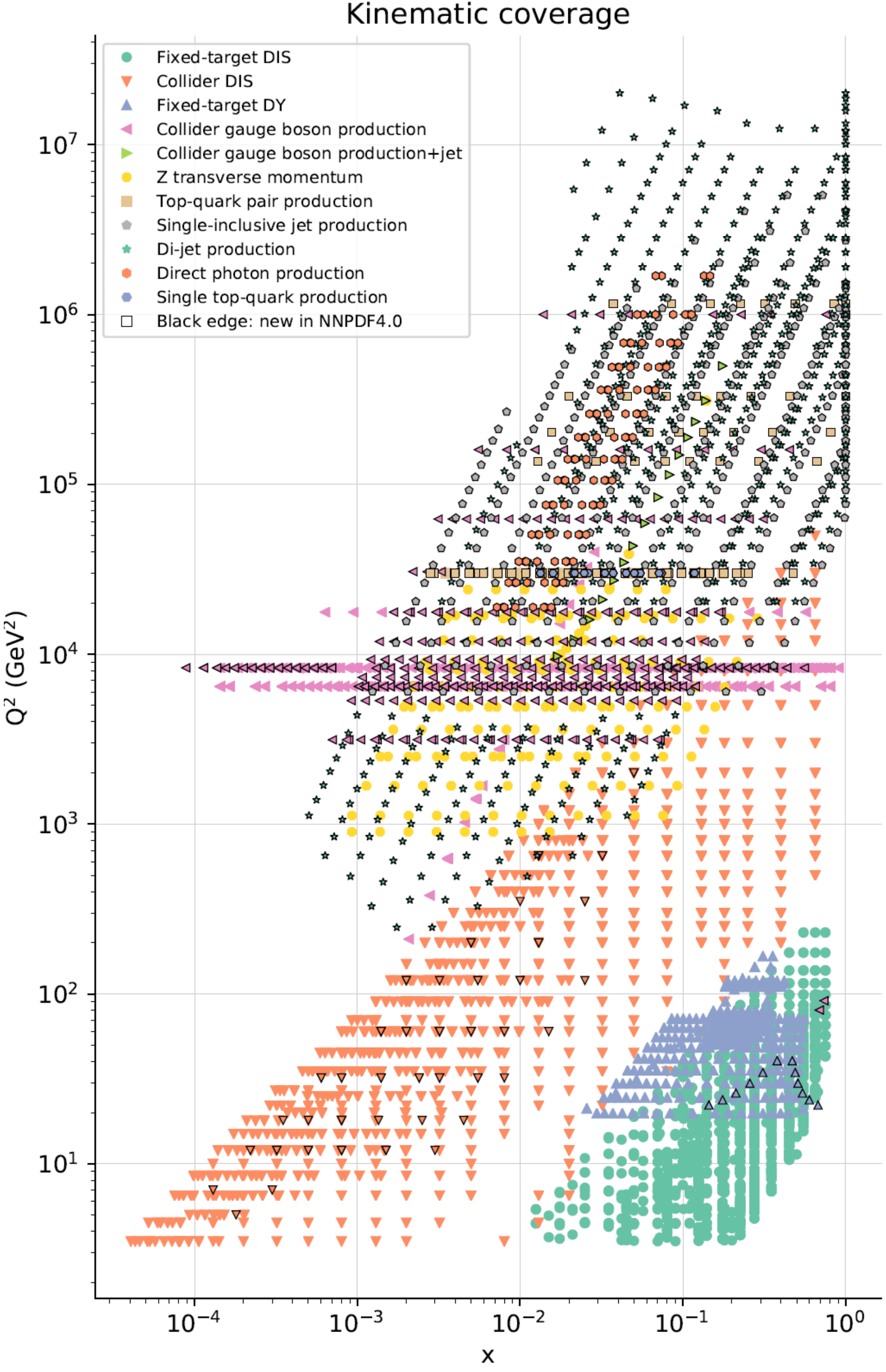
PDF comparisons

Low-x, September2023

A M Cooper-Sarkar

A brief tour of modern PDF sets

and some questions that arise…



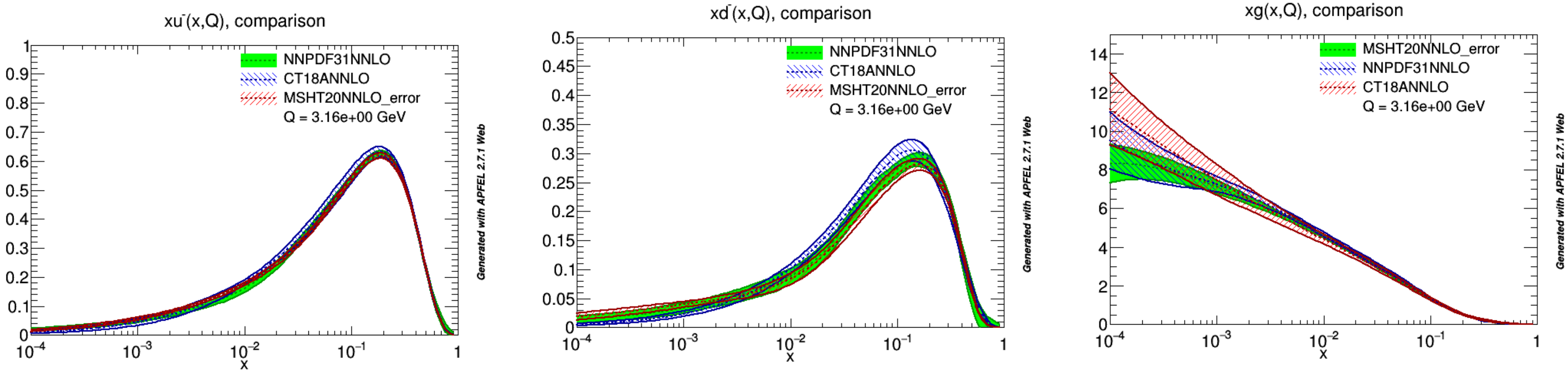
The HERA data are the 'backbone of all PDF fits

BUT what could HERA not do?
High-x gluon and sea flavour detail s,c
What other data can we use?

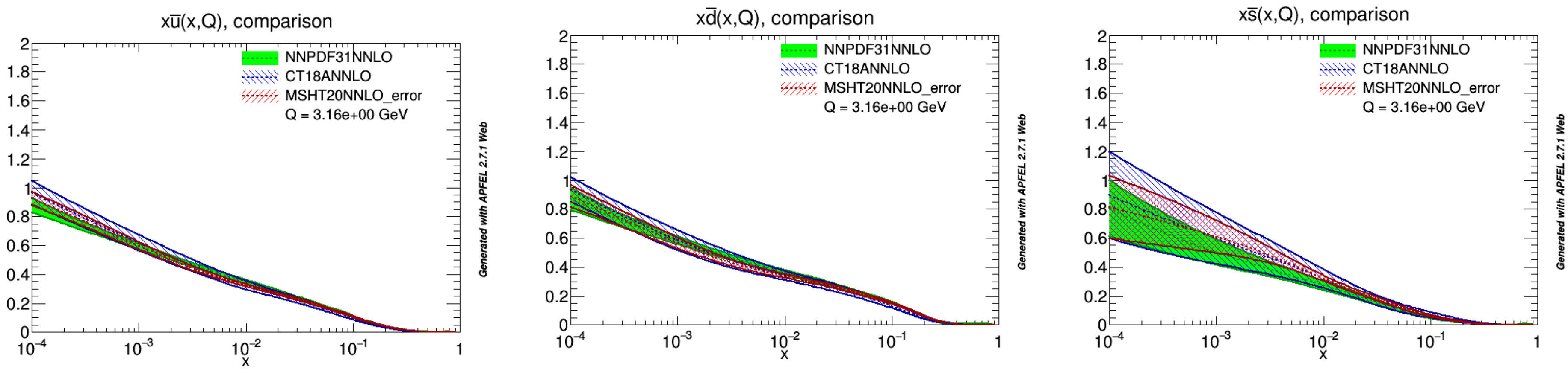
- Drell-Yan data from fixed target DIS and the Tevatron and LHC
- W,Z rapidity spectra from Tevatron and LHC
- Jet pT spectra from Tevatron and LHC
- Top-anti-top differential cross-sections from LHC
- W and Z +jet spectra, or Z pt spectra from LHC
- W and Z +heavy flavours from LHC

- Beware: 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

Let’s see how we are doing--- PDF comparisons at NNLO in pQCD

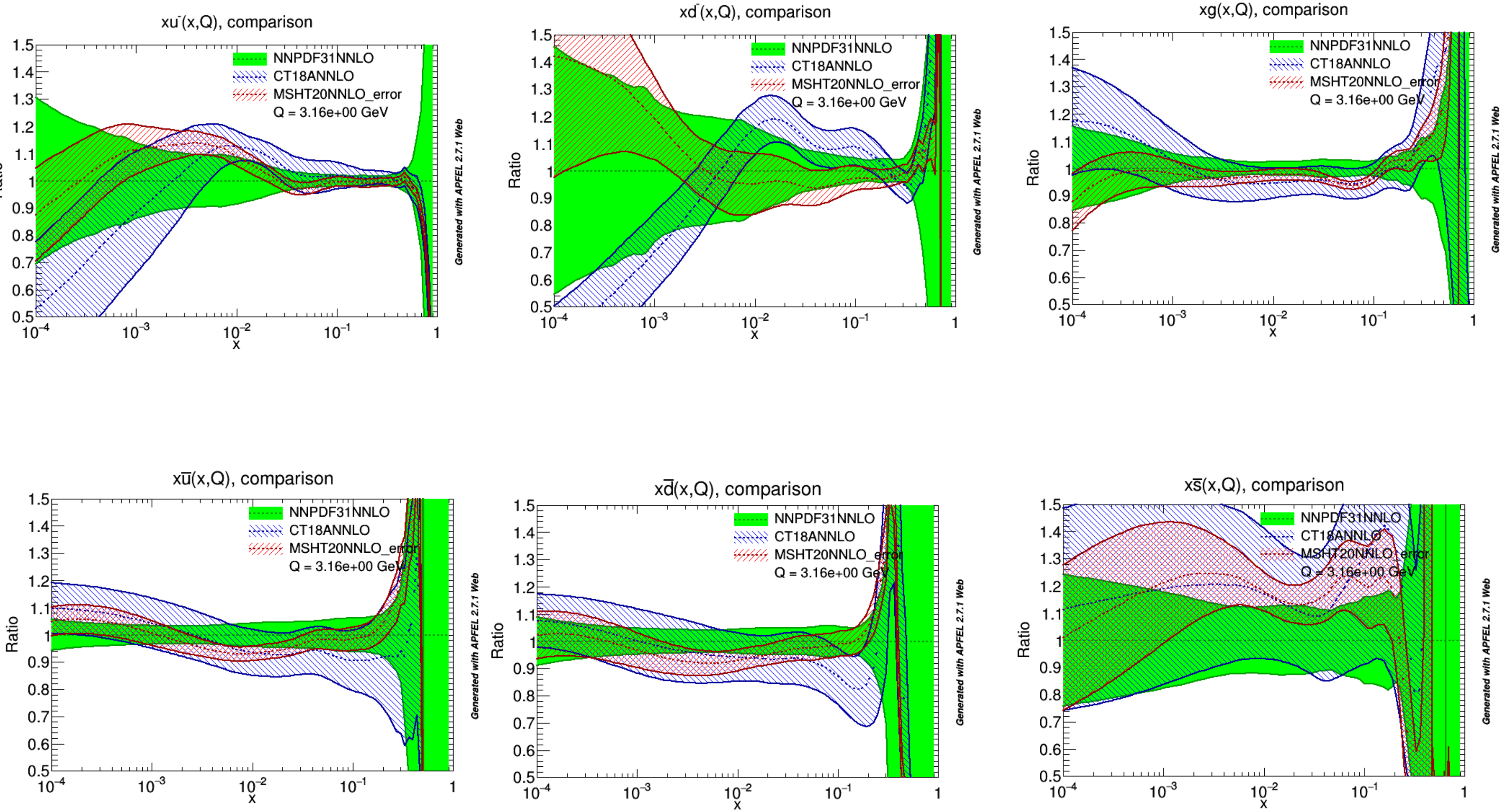


I will come to NNPDF4.0 later



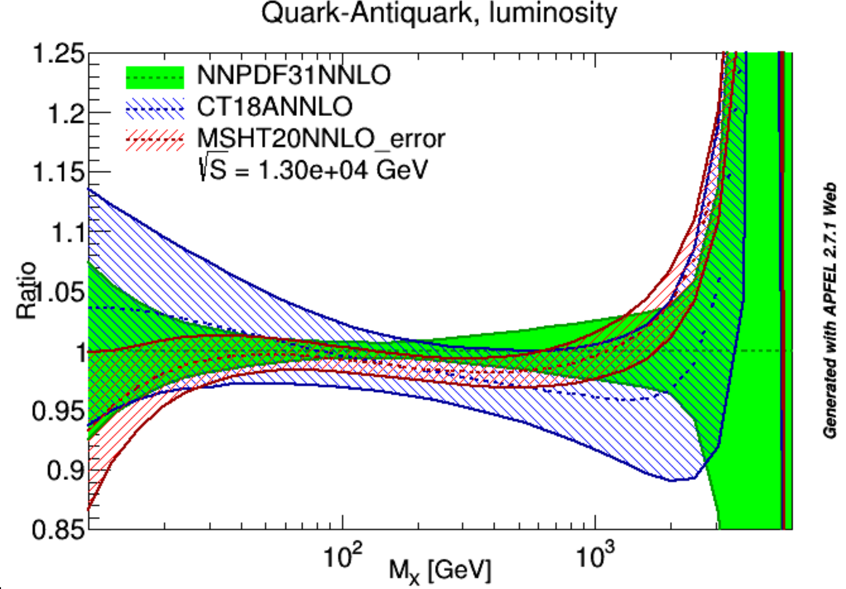
Several groups extract PDFs and there are significant differences because of slightly different choices:

- Exact choice of data entering fit
- Choice of heavy quark masses, heavy quark schemes
- Choice of starting scale for QCD evolution, choice of parametrisation…etc, etc



Differences are more obvious in ratio
They are large at small-x and at high-x

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.

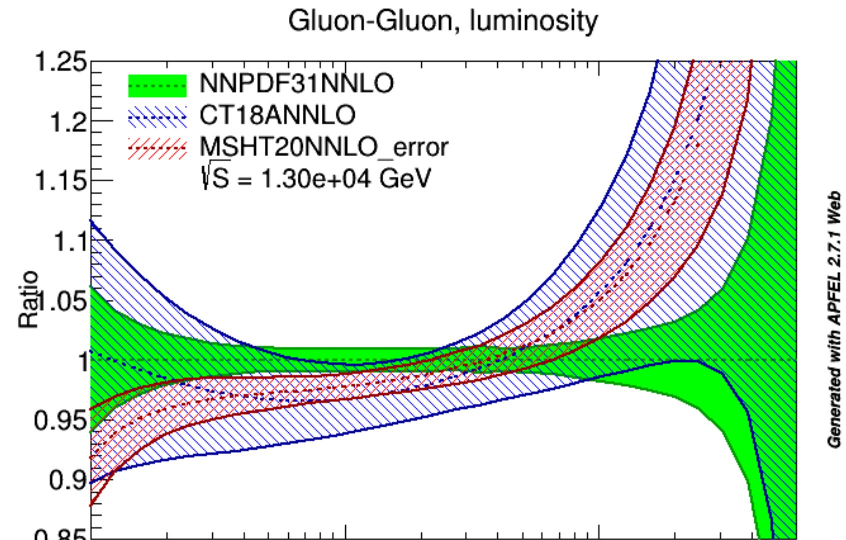


The quark-antiquark and gluon-gluon luminosities for various PDFs are compared here for 13 TeV LHC running in terms of the centre of mass energy of the parton sub- process M_X

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

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

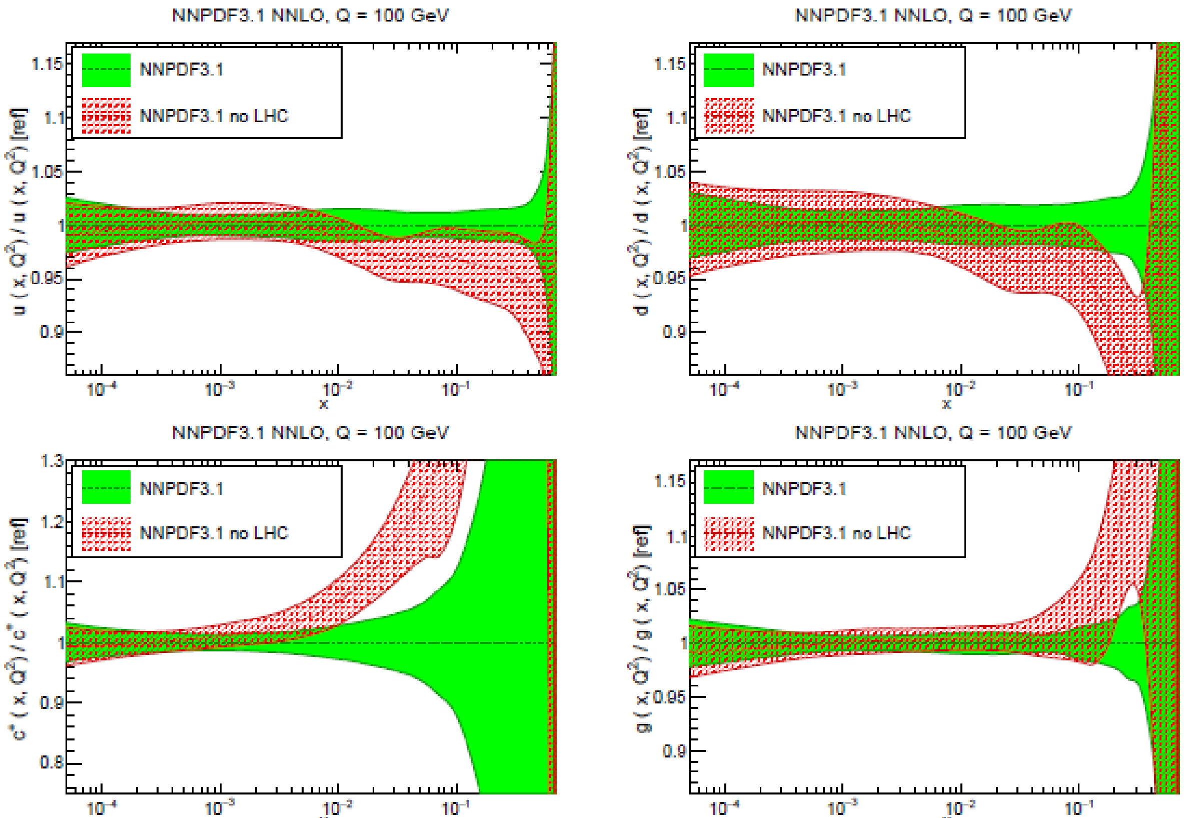
So for quark-antiquark production of W or Z bosons ----at Mx ~80,90 GeV
Or for gluon-gluon production of Higgs at ---Mx~125 GeV
the parton-parton luminosities are fairly well known….but not as well known as we'd like
This is much worse for higher mass particles that could be produced by 'Beyond' Standard Model (BSM) physics



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

Let's see how much LHC data has improved PDFs

NNPDF3.1 includes modern LHC data on W,Z + jets + top + Zpt from 7 and 8 TeV running. Compare PDFs with and without LHC



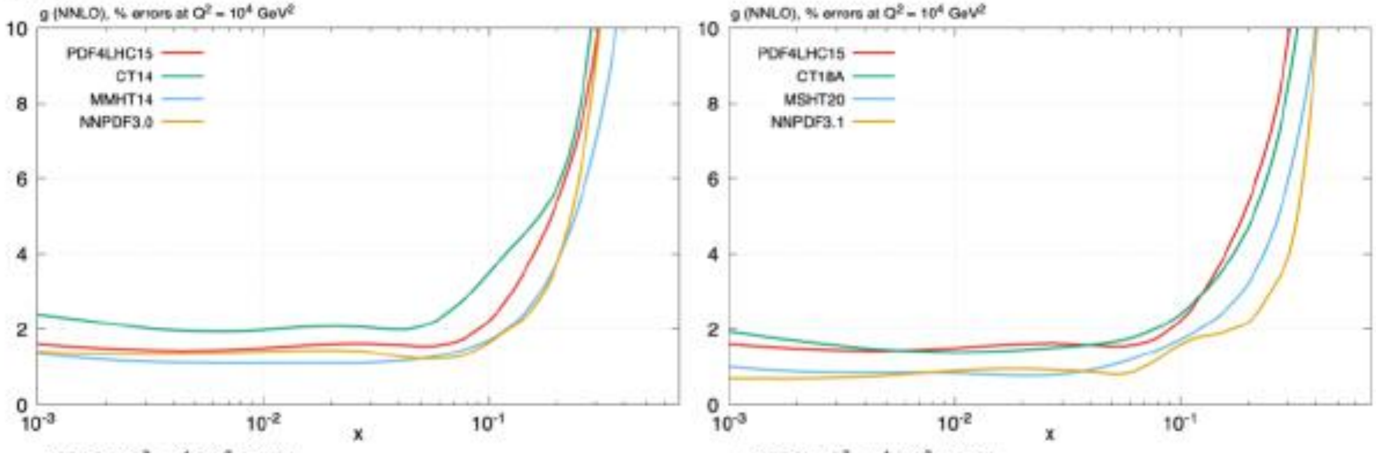
Some of the data input to NNPDF3.1 –like the ATLAS W,Z data have already reached a limit of how accurate they could be. The experimental uncertainties of O(1%) are limited by experimental systematics not by statistics. This will not get better in the foreseeable future e.g. with the High Luminosity LHC

FURTHERMORE, 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—such differences could even increase the total uncertainty of the PDF4LHC PDF combinations---due to differences in central values between PDF sets

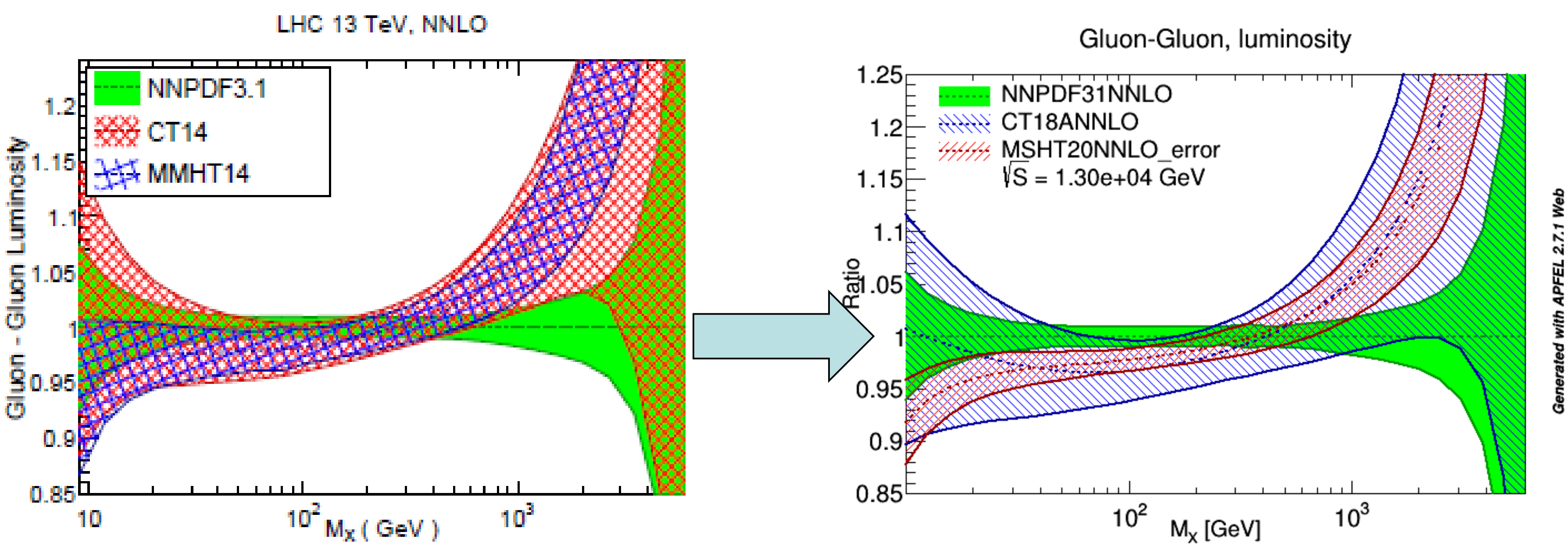
(I will come back to NNPDF4.0)

Gluon



IS THERE PROGRESS?-look at the uncertainties in the present and previous generations of PDFsets

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

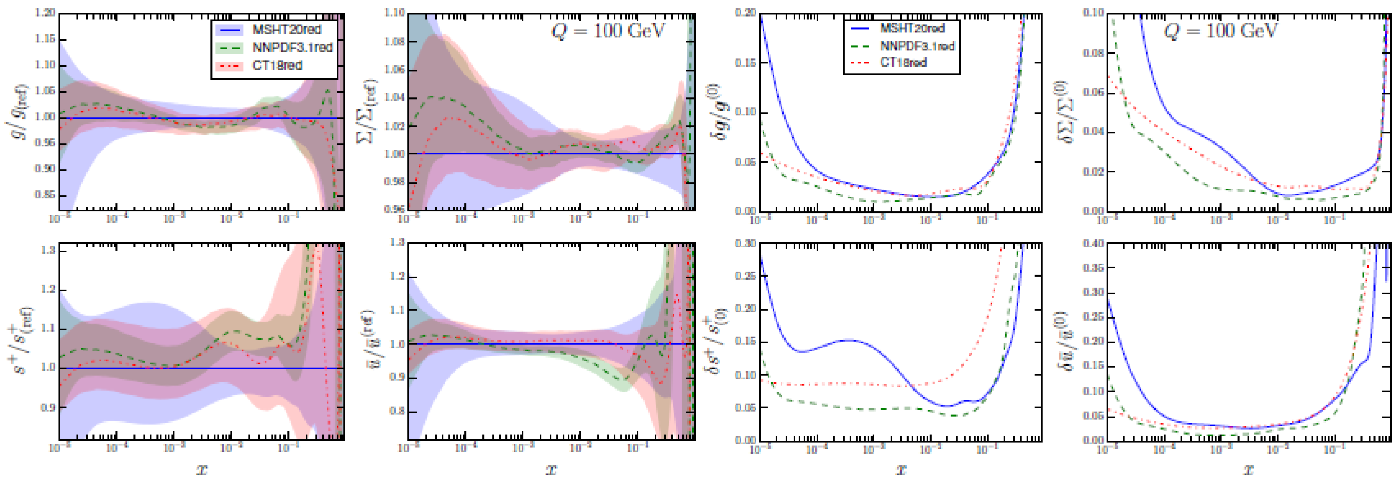


The PDF4LHC group makes combinations of the PDFs from the three main fitting groups NNPDF, CT and MSHT

First 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.

BUT 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 a reason!

New tools to asses data sensitivity/inconsistency arXIV:2306.03918

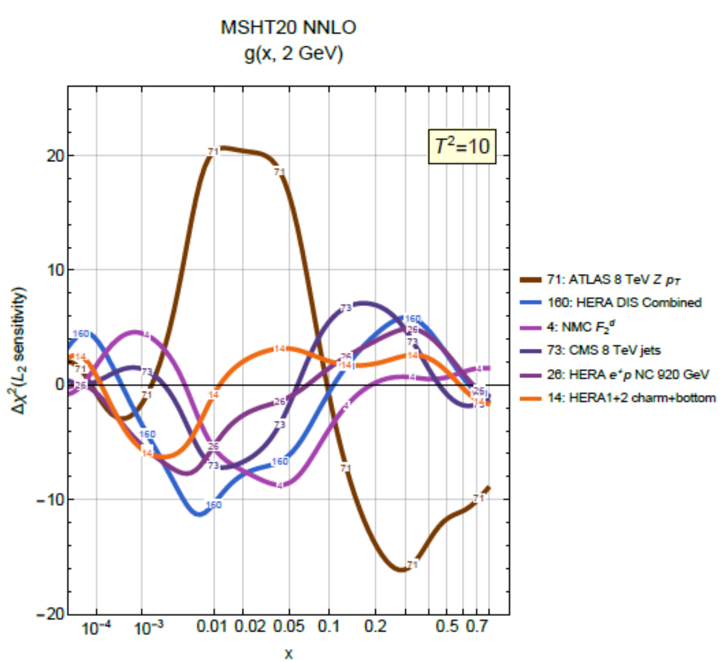
L2 sensitivity combines information on the sensitivity of a measurement to a PDF in principle, and the precision of a particular measurement

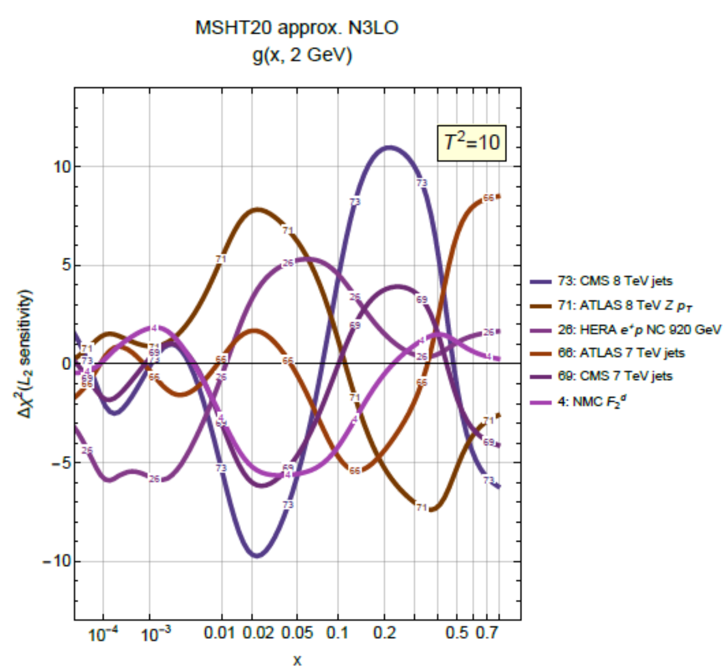
$$S^{\mathrm{H}}_{f,L2}(E) \equiv \frac{\vec{\nabla}\chi^2_E \cdot \vec{\nabla} f}{\delta_{\mathrm{H}} f} = \left(\delta_{\mathrm{H}}\chi^2_E\right)\ C_{\mathrm{H}}(f,\chi^2_E)\ , \tag{5}$$

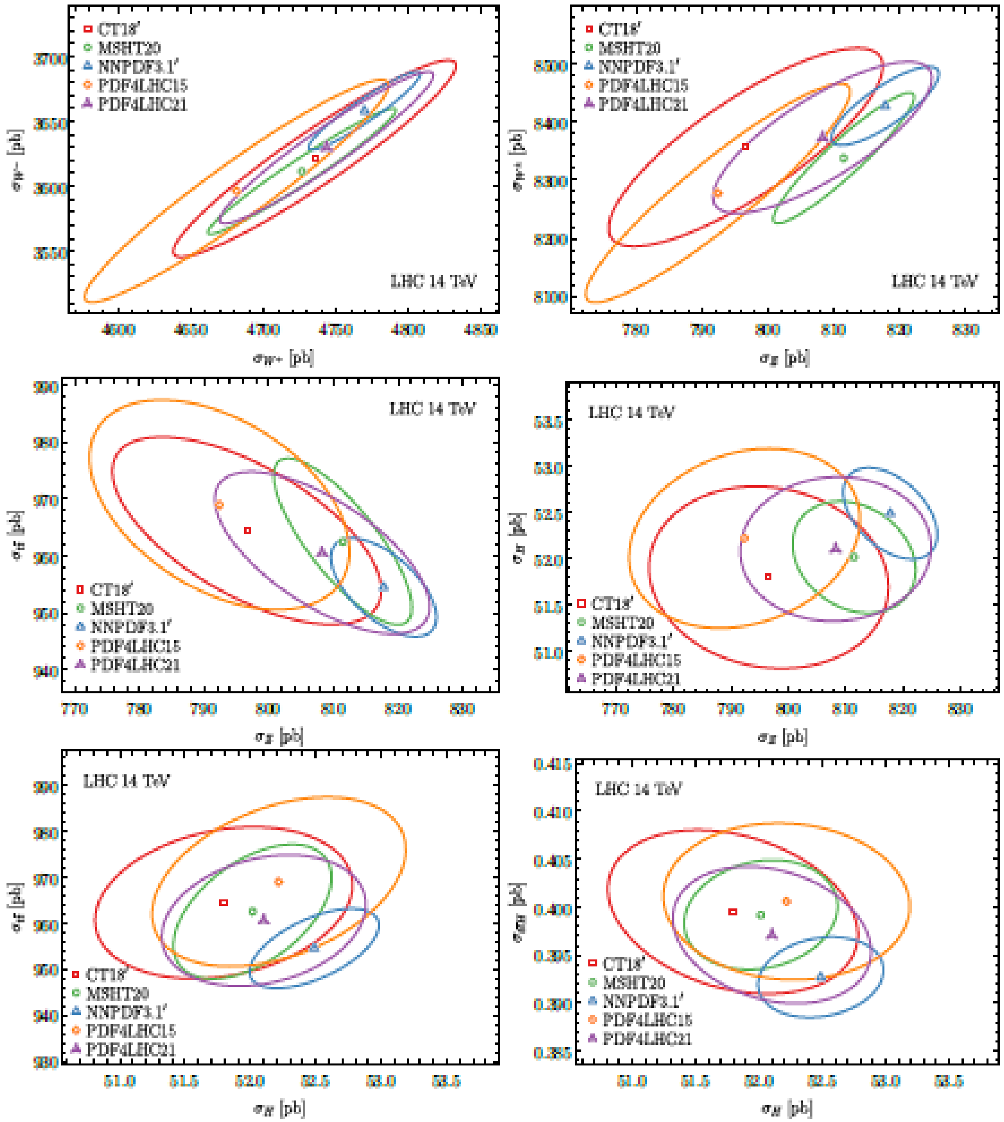
where $C_{\mathrm{H}}(f,\chi^2_E)$ represents the cosine of the correlation angle between f and the χ^2 for experiment E, evaluated over the $2D$ Hessian eigenvector sets.

Sensitivity of the gluon to different data sets in the MSHT fit

Zpt VERY sensitive at NNLO………………less so at N3LO—we will come back to N3LO



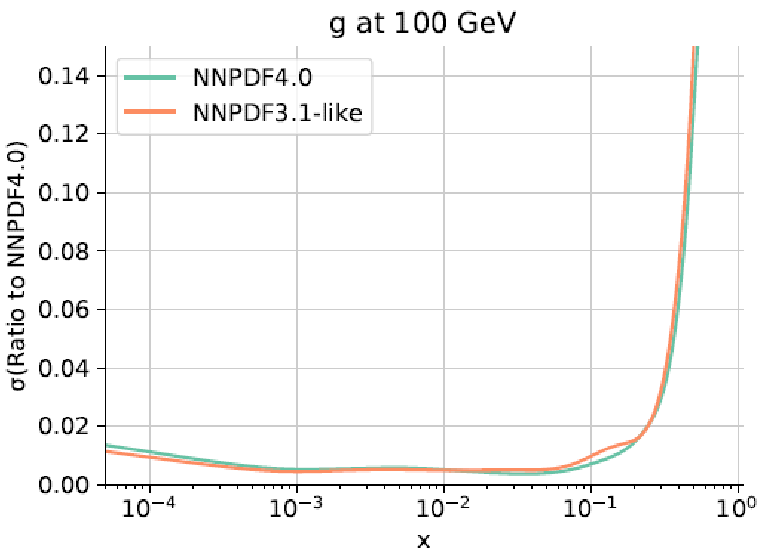




The PDF4LHC group makes combinations of the PDFs from the three main fitting groups NNPDF, CT and MSHT

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

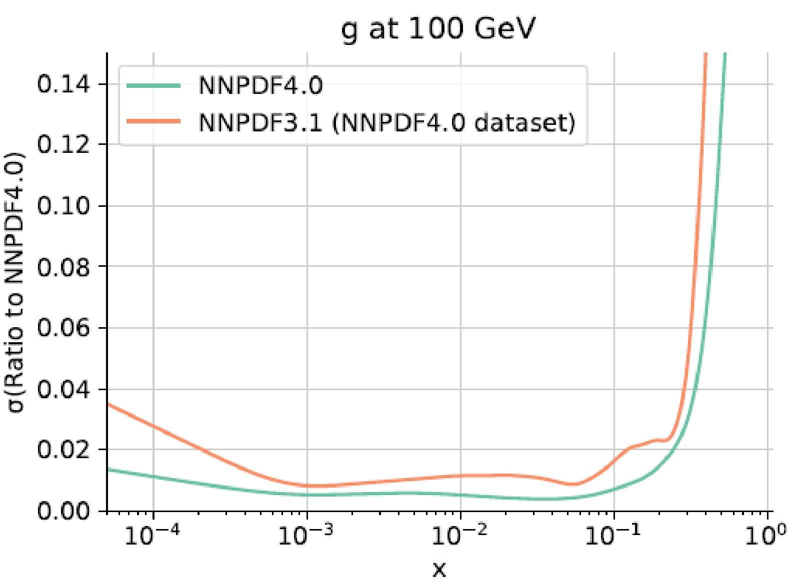
There IS an improvement in uncertainty BUT this is not enough to reduce the PDF uncertainty on on LHC measurement of say m_W sufficiently to compete with the CDF uncertainty- we need more than this…



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

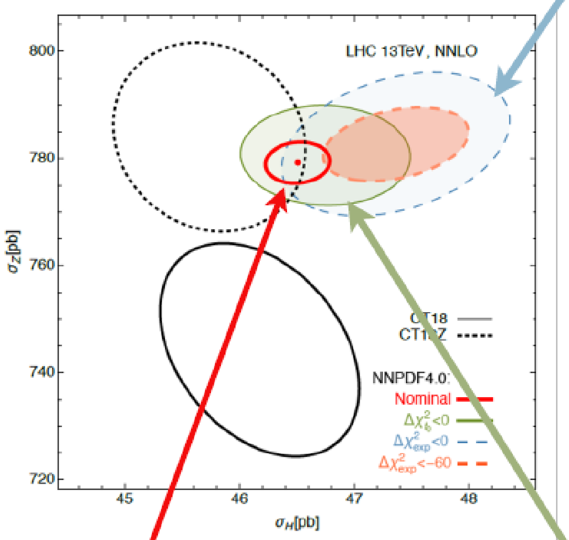
There is currently a lot of debate in the PDF community over the new methodology. But even if it is accepted this does not help much if one is trying to combine with other PDFs MSHT20 and CT18 with different central values and larger uncertainties.

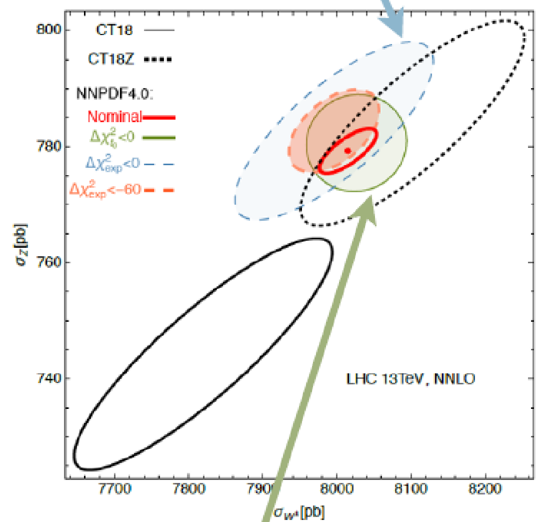
Monte-Carlo sampling sensitivity for PDFs

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

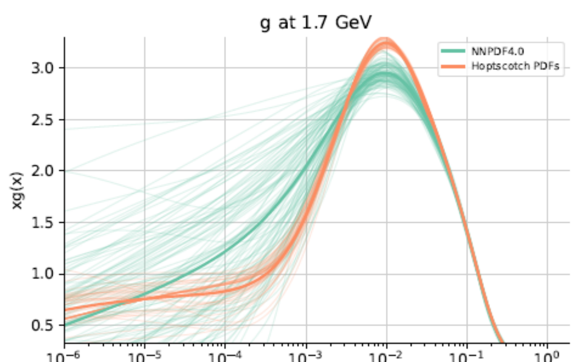
Region containing good solutions according to the most recent t_0 form of χ^2 (used to train NN4.0 replicas)

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.

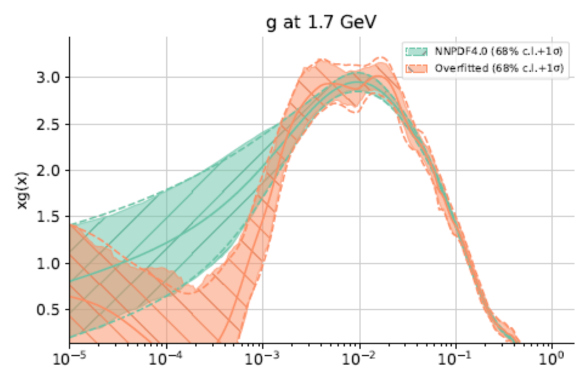
Observation 1

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



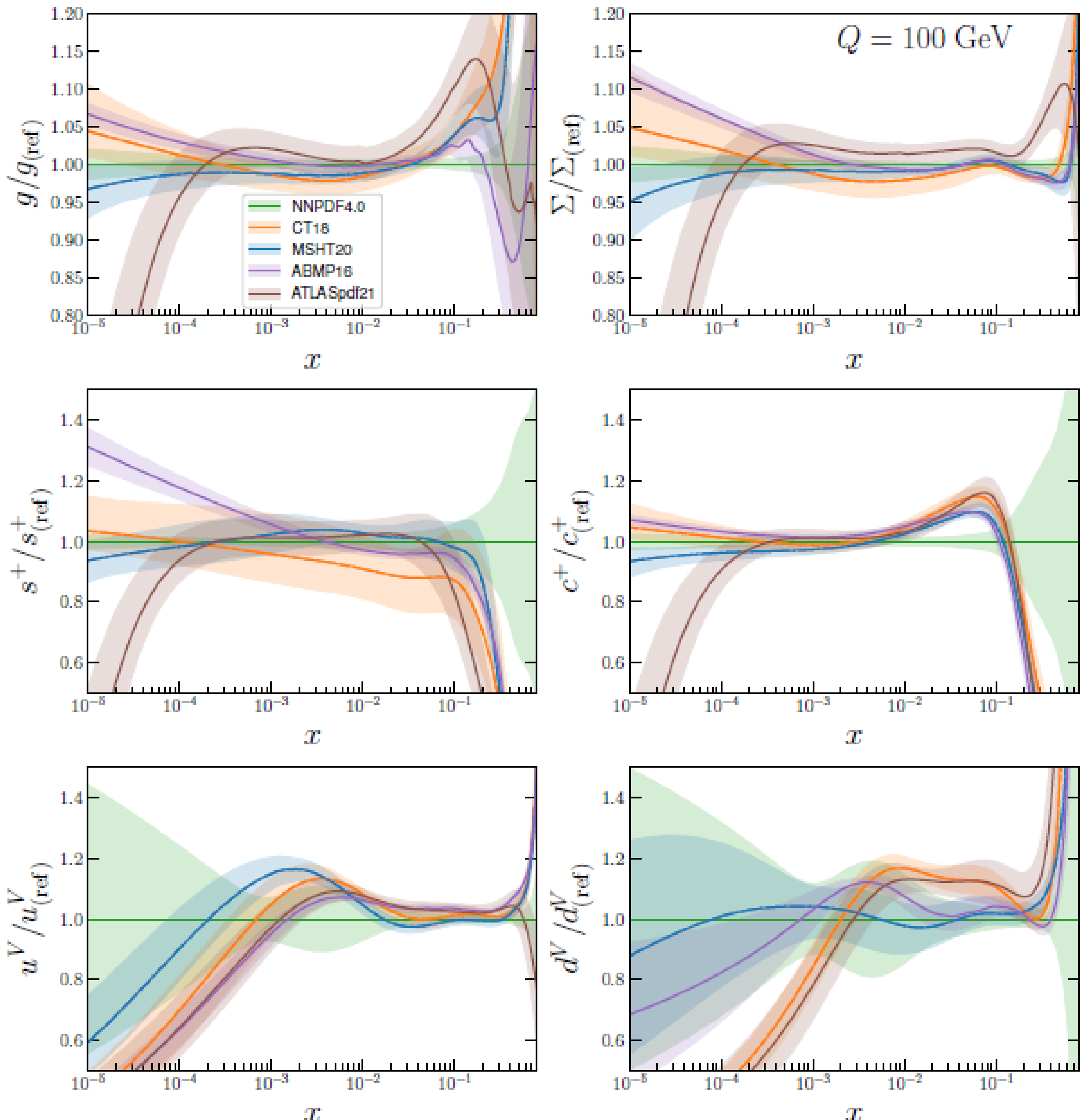
Observation 2

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



NNPDF reply at PDF4LHC2022… but it goes on

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



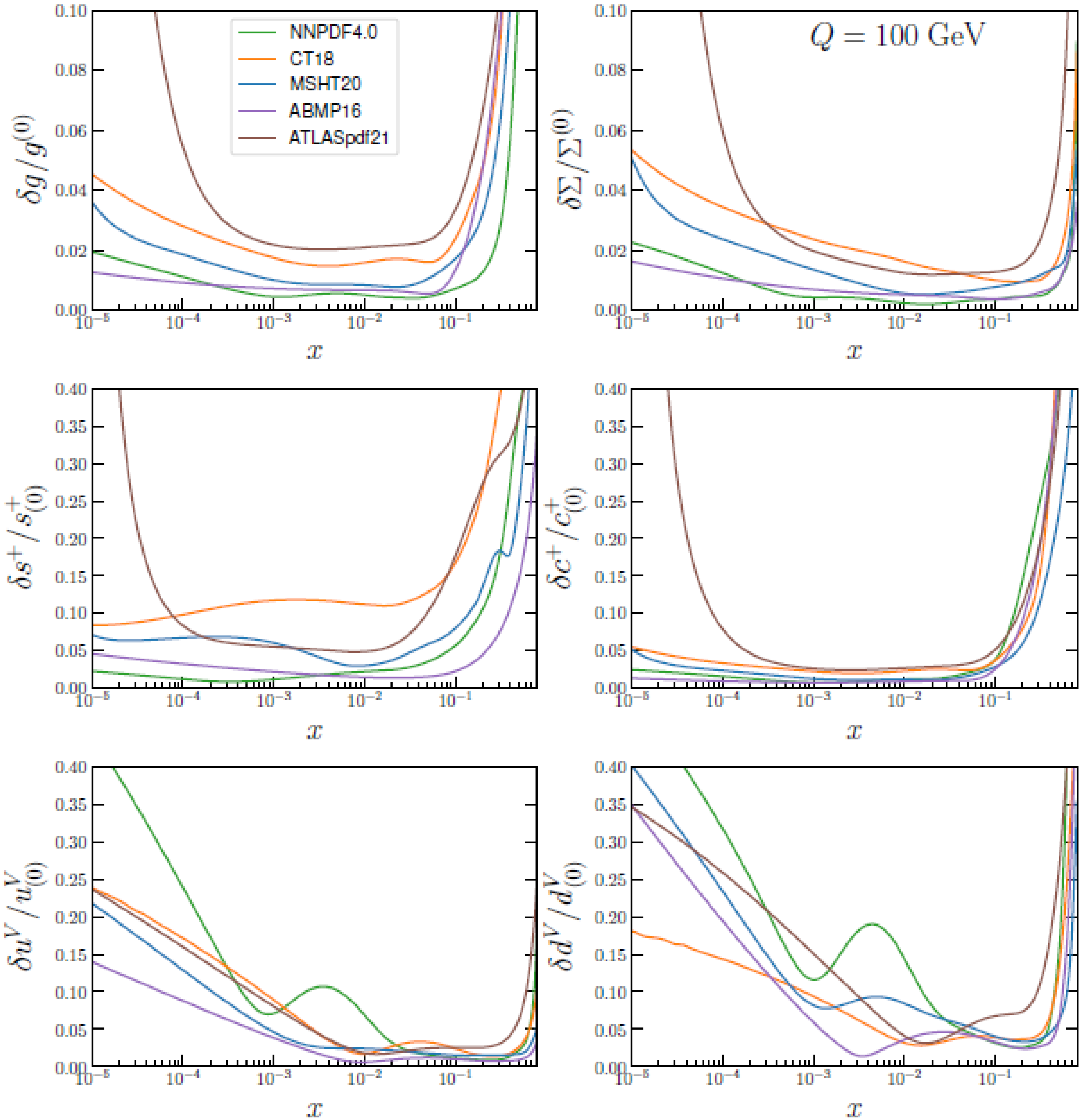
In ratio to NNPDF4.0

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

BUT I mostly want to talk about the low-x gluon



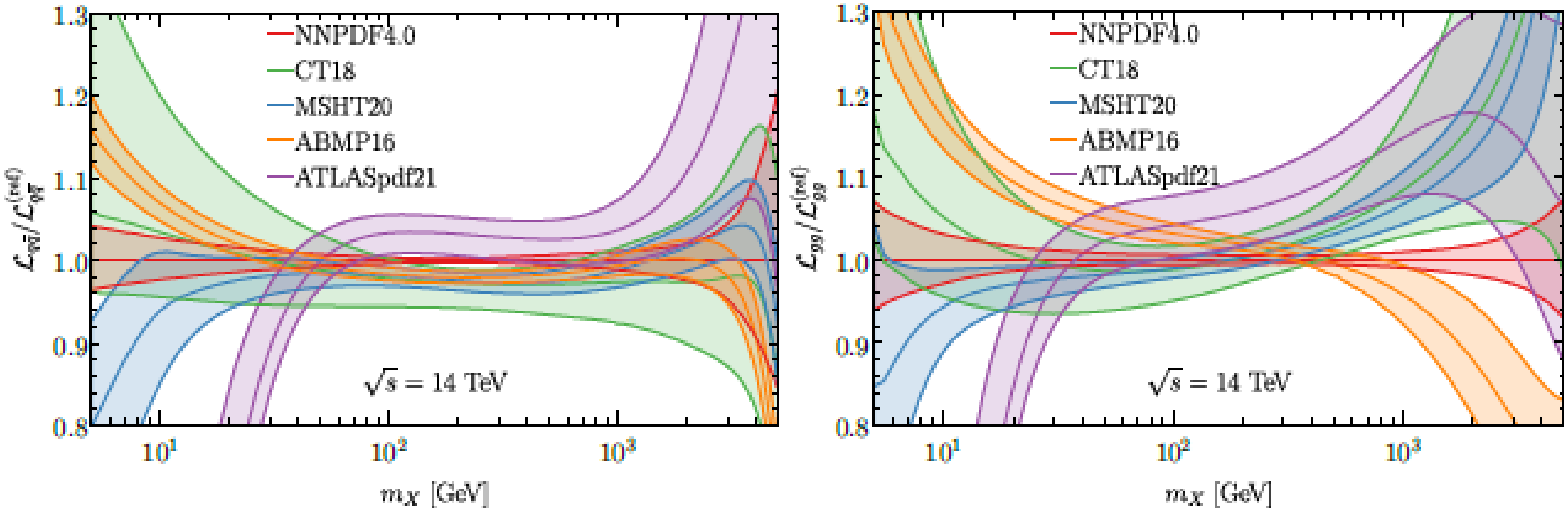
Before that let us look at uncertainties

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..



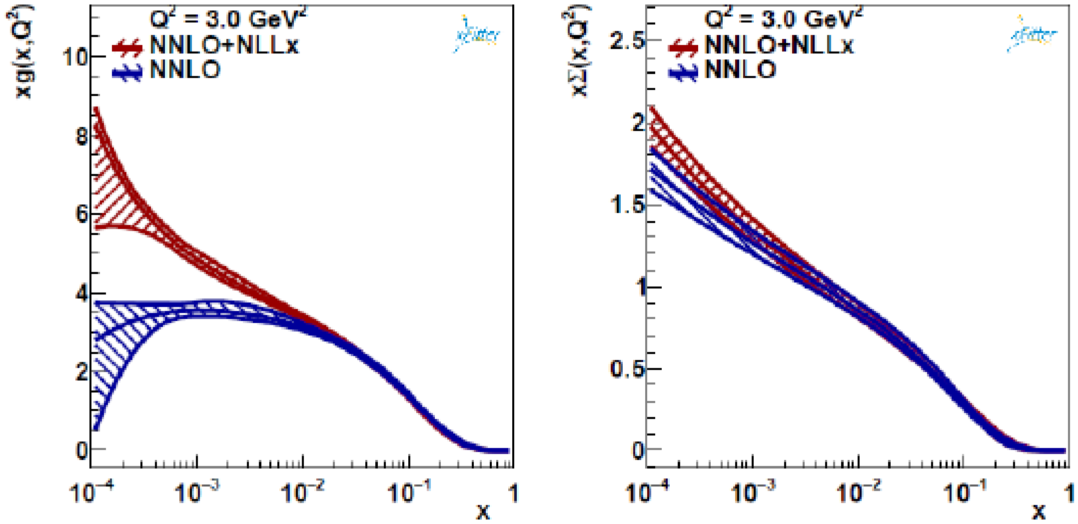
Now let us see the consequences of differences at low-x for LHC luminosities:
q-qbar left, g-g right in ratio to NNPDF
We don’t often worry about mX < 30 GeV at the LHC

But if we did.. We need to worry about the low-x theory

i) In the ATLASpdf21 fit the HERA data at Q^2 < 10 GeV^2 (x < ~10^{-4}) were cut precisely to avoid this problematic region– (the HERA data are still the main data which probe this region) but it turns out that this is almost exactly the wrong thing to do at NNLO– a better approximation to ‘the truth’ would be got by fitting down to lower Q^2 and putting up with the larger χ2– as is done by MSHT (who have a similar gluon parametrisation), CT, NNPDF

What do I mean by ‘the truth’– well I mean what one might get at higher order, or with BFKL ln(1/x) resummation

There has long been an issue that at low-x one should probably be resuming ln(1/x) terms as well as ln(Q^2) terms –this NLLx is BFKL resummation and is beyond DGLAP
This has been done by NNPDF- in NNPDF3.1sx arXIV:1710.05935
And by the HERAPDF using xFitter arXIV:1802.00064 (using HELL from Bonvini



What does NLLx do?
It turns blue into red– dramatic change on the low-x gluon

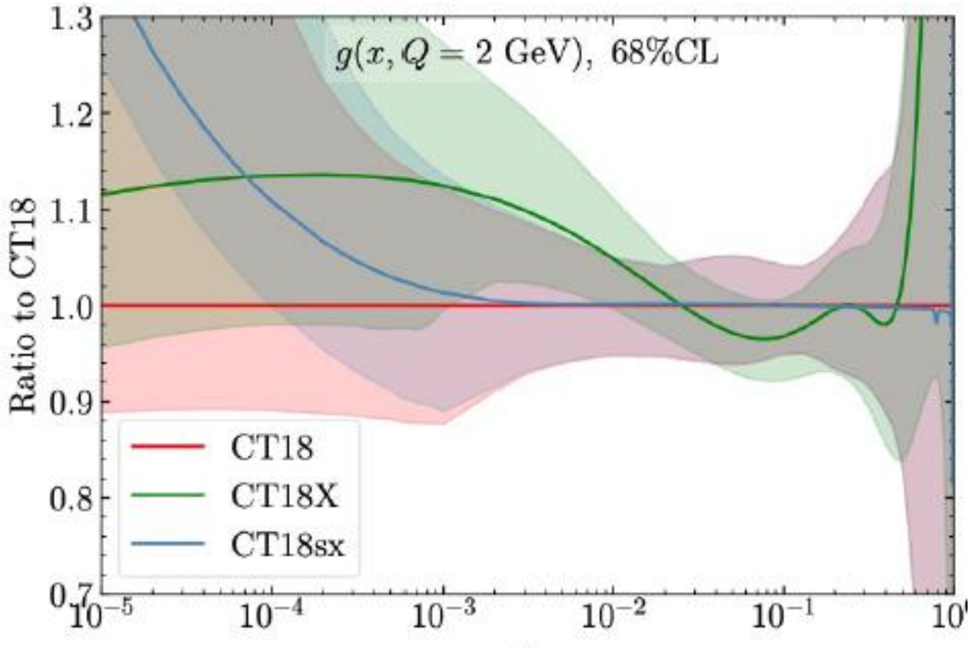
Gets a better χ2 for the low-x,Q2 HERA data by ~70 units

CT have now also done this in CT18sx

But there is another thing one needs to consider– **high density effects** when the gluon gets large such that gluons may recombine, as well as split, and this may lead to gluon saturation. CT have modelled this with an x dependent scale for DIS in CT18X
Not Q^2 BUT

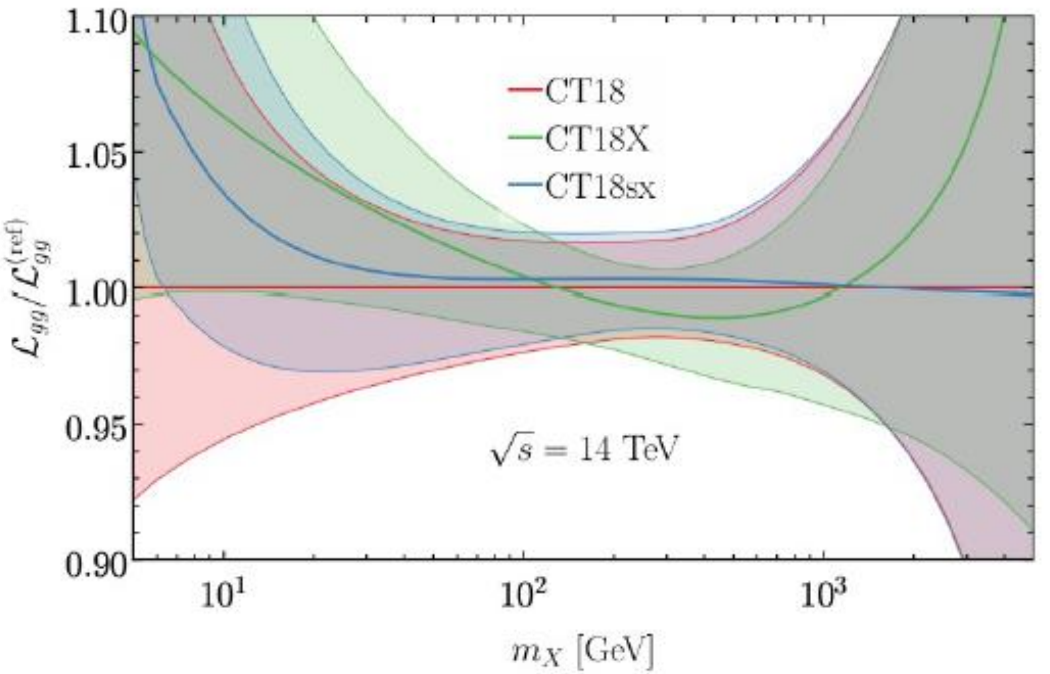
$$\mu_{DIS,X}^2 = 0.8^2 \left(Q^2 + \frac{0.3\ GeV^2}{x^{0.3}} \right)$$

This also enhances the low-x gluon—--
And it gives a similar decrease in χ2 for the low-x,Q2 HERA data by ~70 units



Compare gluon shapes at low scale

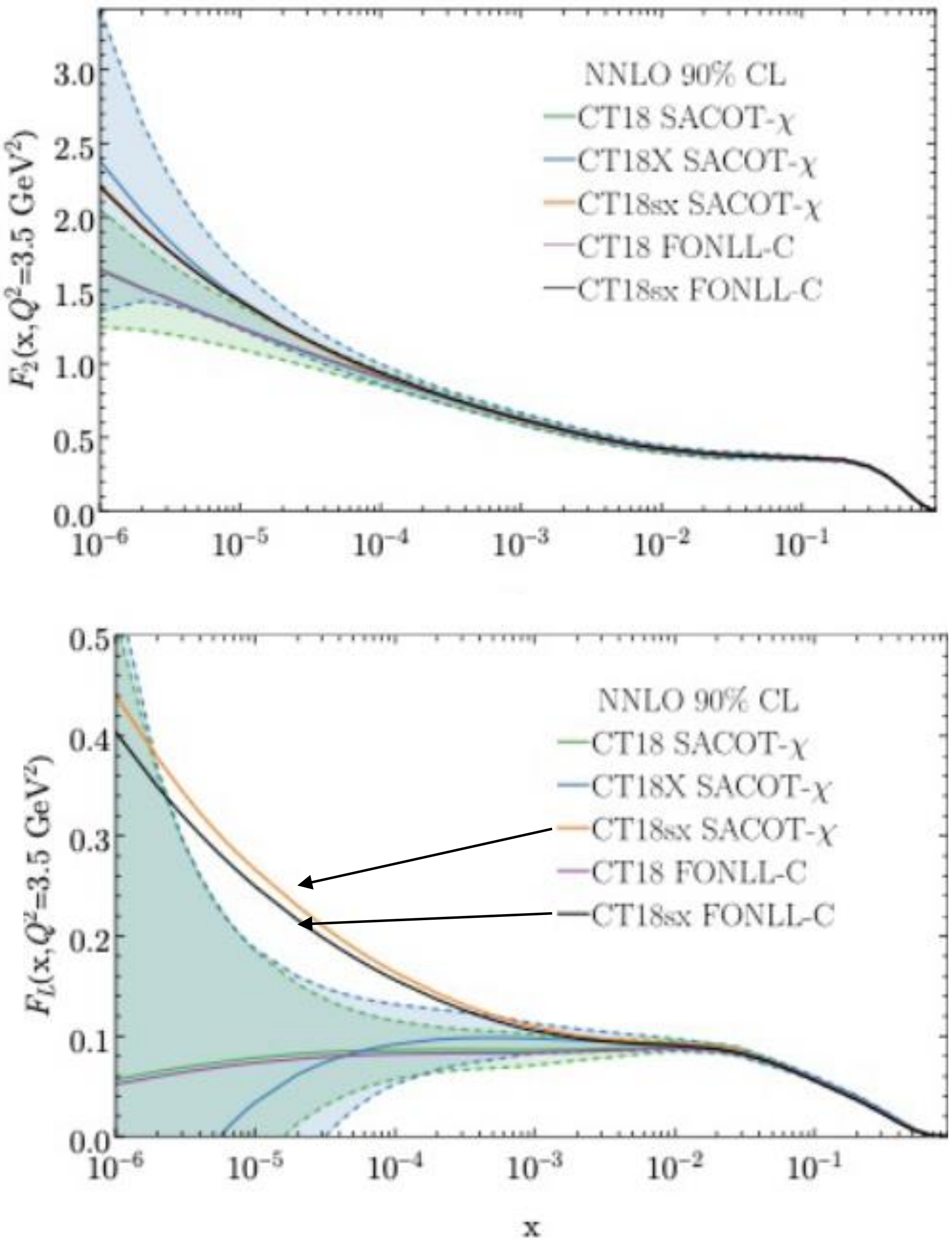
CT18X is a variant which uses a scale intended to mimic saturation
CT18sx is a variant with ln(1/x) resummation



And then compare gluon-gluon luminosities for the LHC

Hard to distinguish within uncertainties BUT..

arXIV: 2108.06596

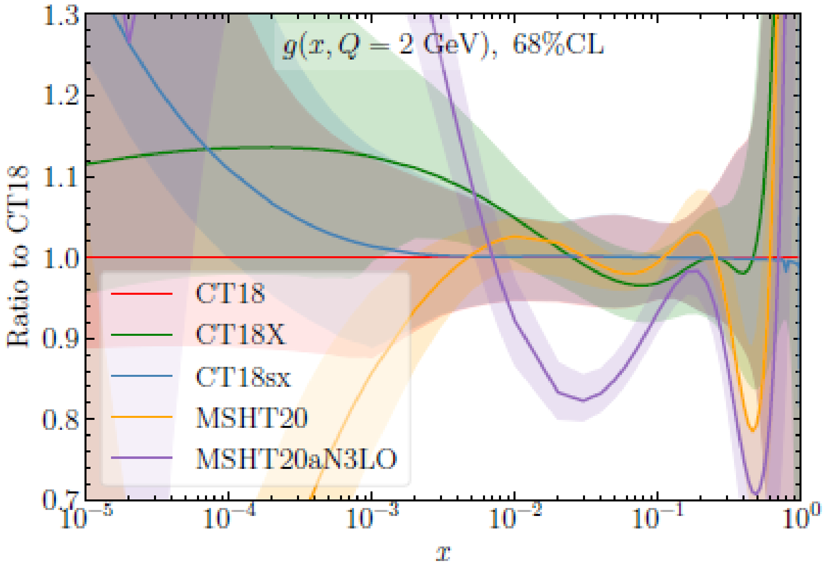


DIS structure function FL at low Q
Can maybe tell difference between CT18/CT18X and CT18sx

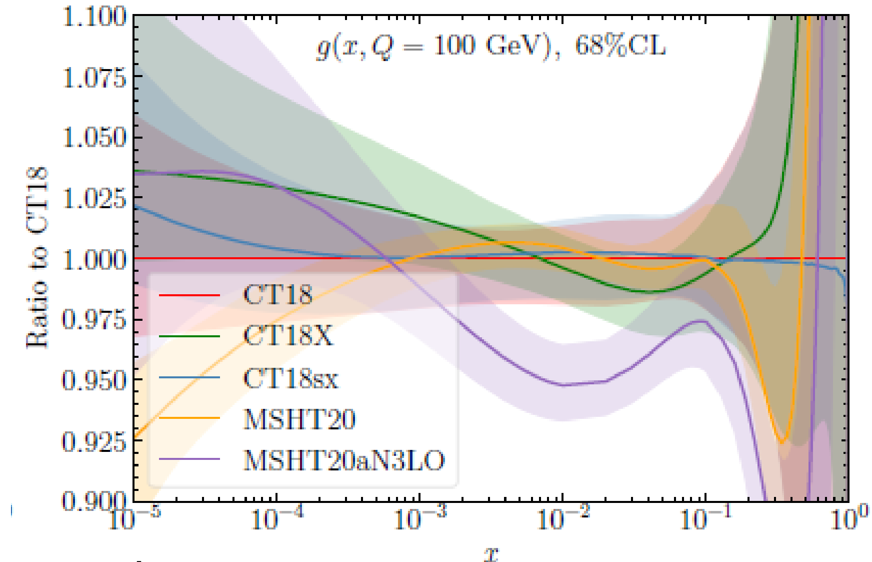
Returning to the LHC there has been a parallel development --- N3LO

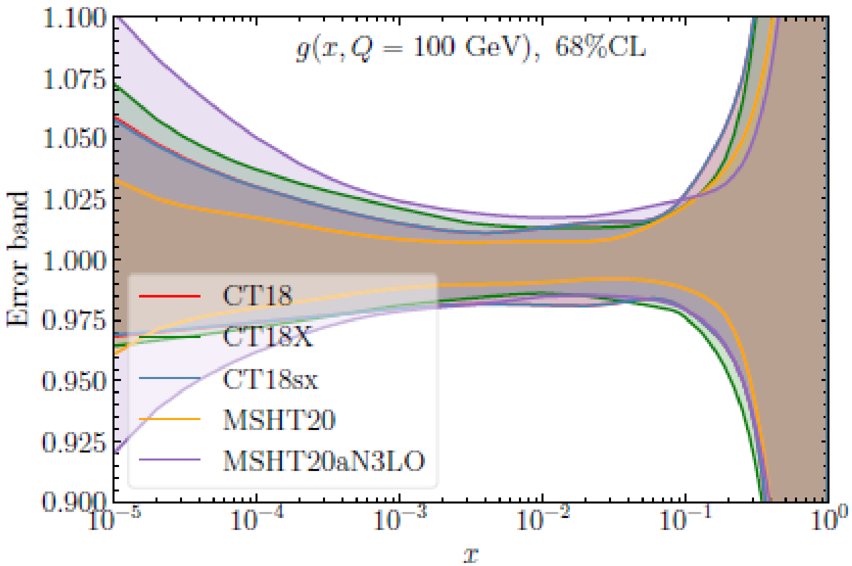
Well at least approximately

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



Which persists to LHC scales

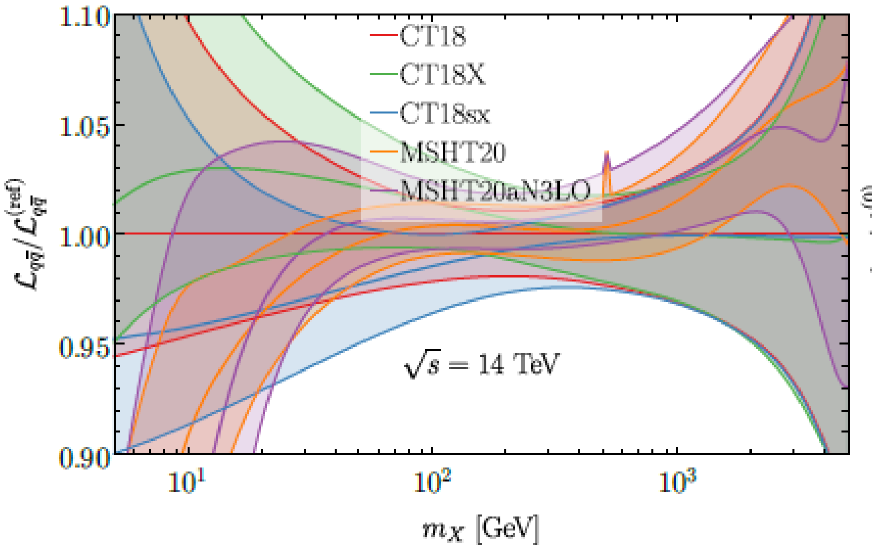


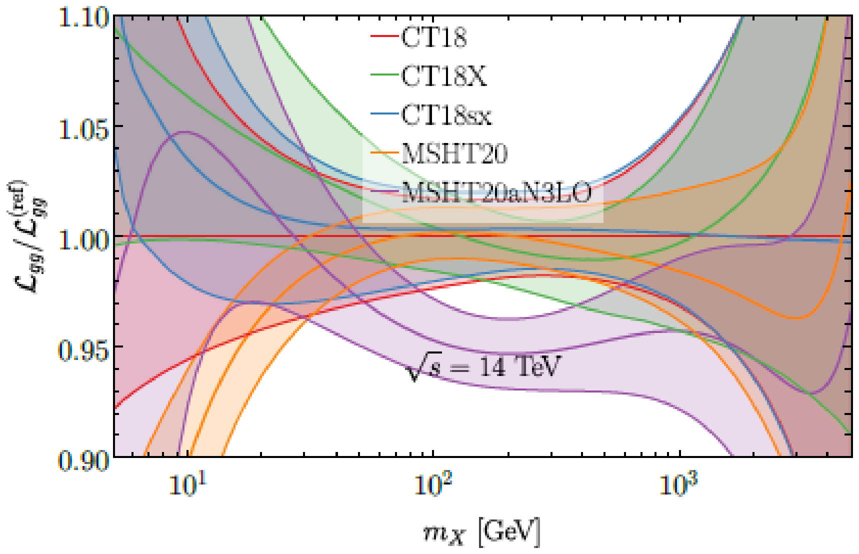


Contrast the MSHT20 NNLO
With the MSHT20aN3LO
But also note it is much stronger than the changes of CT18 to either CT18sx or CT18X– although there are similarities in a rise of the low-x gluon

And note in passing that the uncertainties are larger because there is an attempt to include uncertainty from yet higher orders

How about LHC parton-parton luminosities in these newer variants?





q-qbar left and g-g right, in ratio to CT18

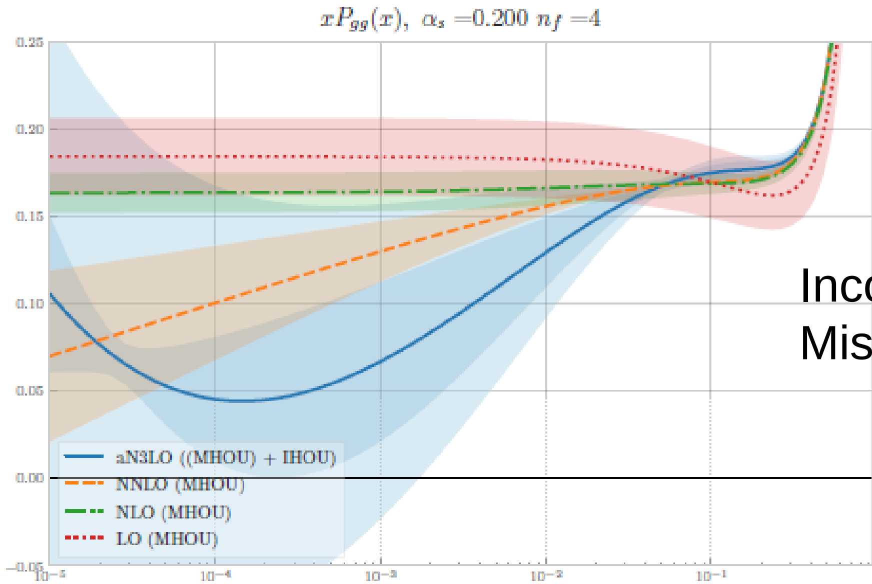
Note especially the ‘knock-on’ effect of the rise in gluon luminosity at low scale for N3LO to a decrease of ~5% at the Higgs scale…

Clearly we are going to have to consider low-x both

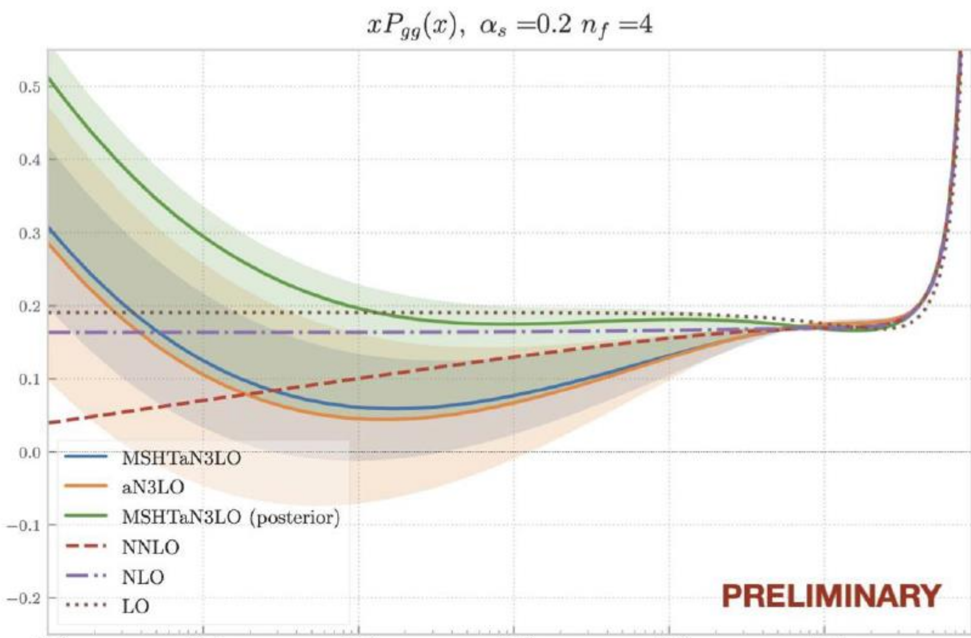
- If we go forward to FPF
- And if we go to higher energy—when the kinematic region moves to lower-x

But ALSO – we need to worry about the consequences at the Higgs scale

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)



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

Summary/ Things to worry about

- PDF improvement is not just a matter of more data
- Consistency of data matters
- Knowledge of common systematic uncertainties matters,if we wish to reach O(1%) precision
- Real data are always more problematic than pseudo-data projections

- Differences in the PDFs are not just about choice of data set—PDF4LHC reduced data sets still have some difference
- There are irreducible methodological differences between the PDFs
- Sometime this is just a matter of model choices that can be made consistent e.g. heavy quark masses, $\alpha_s(M_Z)$.
- But sometimes the choices are made for 'ideological reasons'—parameterisations, NNs, heavy quark treatment/intrinsic charm
- Greatest differences in definitions of how to set uncertainties – choice of $\chi 2$ tolerance /NN method

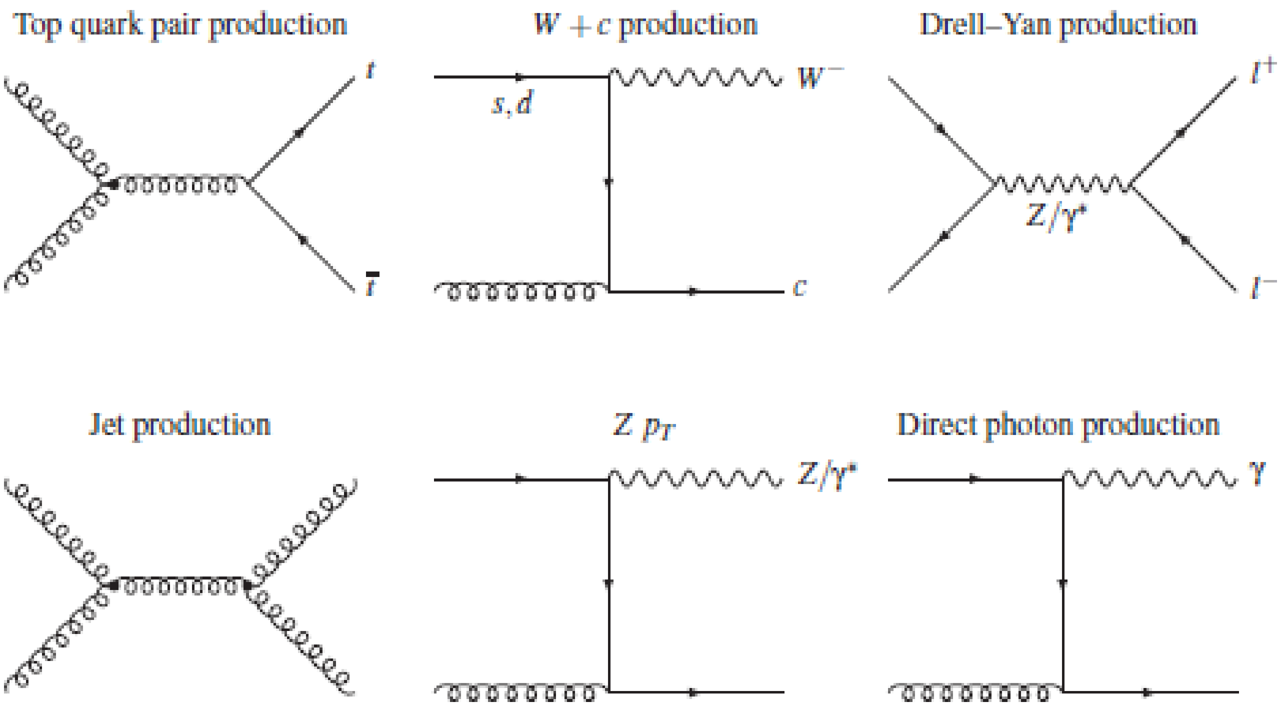
- N3LO /N3LO benchmarking
- Ln(1/x) resummation

Clearly we are going to have to consider these both

- If we go forward to FPF
- And if we go to higher energy—when the kinematic region moves to lower-x
- But ALSO – we need to worry about the consequences at the Higgs scale

Further comments on future projections

A study of potential improvements has been made using processes for which are now statistics limited, where the High-Luminosity LHC (HL-LHC) should help



Pseudo-data is generated for these processes assuming luminosity of 3 ab^{-1} for CMS and ATLAS and 0.3 ab^{-1} for LHCb

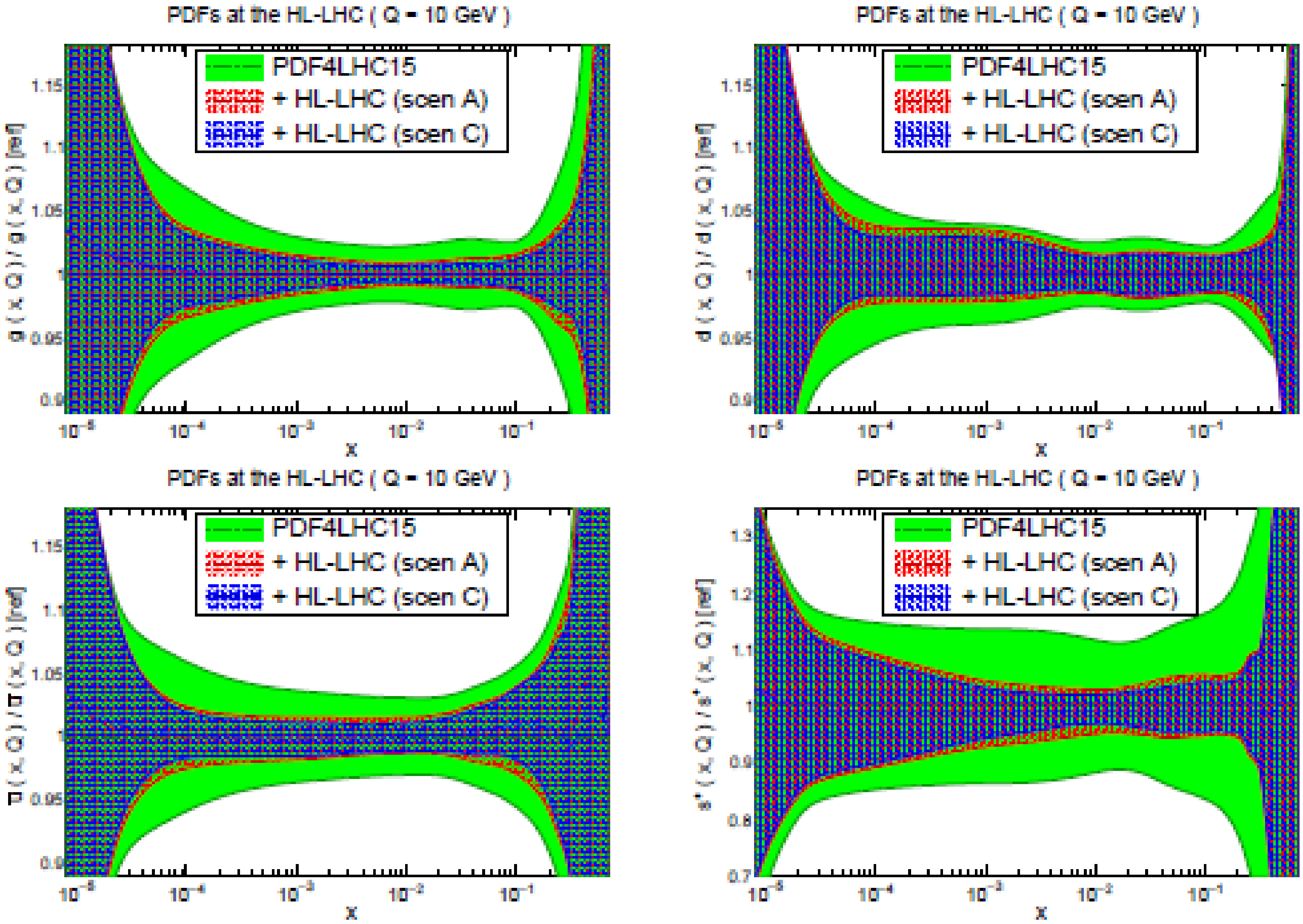
Pessimistic and Optimistic assumptions are made about systematic uncertainties based on experience with real data

Both about the effect of correlations-- typically, f_{corr} = 1, 0.25

And about possible reduction in uncertainty typically, f_{red} = 1, 0.4

This is about as good as you can do with pseudo-data but let's not forget that this is a somewhat ideal situation

So we see potential improvements in the PDFs



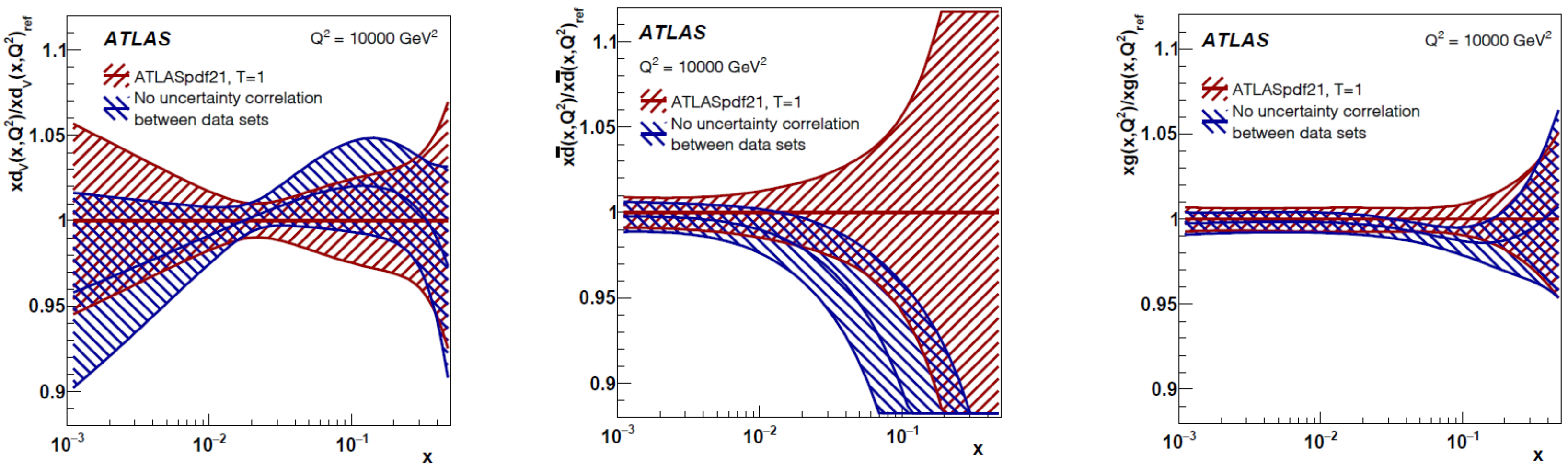
Where scenario A is pessimistic and scenario C is optimistic

--Such improvements could give up to a factor 2 improvement in the PDF uncertainty on something like m_W ....but such estimates are unlikely to be fully realistic…

Are we being a little too optimistic with these pseudo-data projections?

One of the issues with LHC data is that realistically **it involves the combination of many data sets analysed by different groups and with differing procedures for the evaluation of systematic uncertainties,** which makes cross-correlating them difficult. Such correlations are not usually known/applied

But recent work by **ATLAS** uses many different types of LHC data, (W,Z, jets, direct photon, top-antitop, V+jets), evaluating the largest correlations (arXiv:2112.11266)



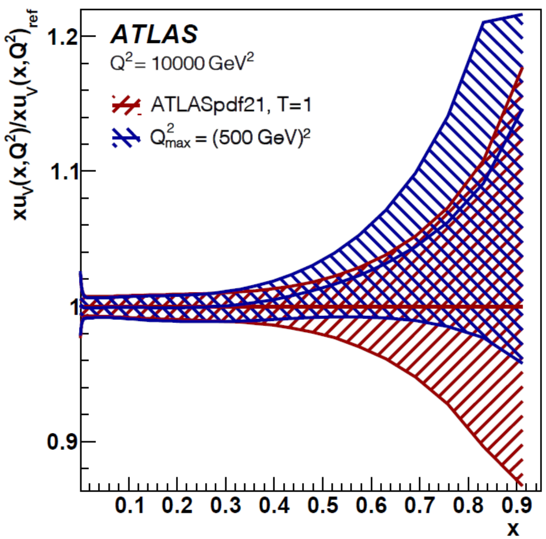
The larger systematics come from jet energy scales and are correlated between data sets such as: inclusive jets, W or Z boson +jets, t-tbar in lepton+jets mode

The difference between accounting for the correlations or not doing so is the shift from red to blue—shown in ratio here

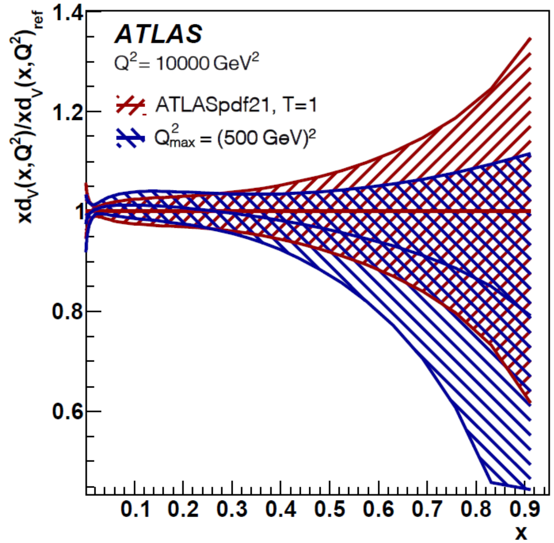
It is not a big effect, but if you want ~1% accuracy on PDFs then it matters

What if we are fitting away new physics effects ?
ATLAS made a fit cutting data for which the scale > 500 GeV, to check if the PDFs differ if we cut out possible hidden new physics in the high scale data

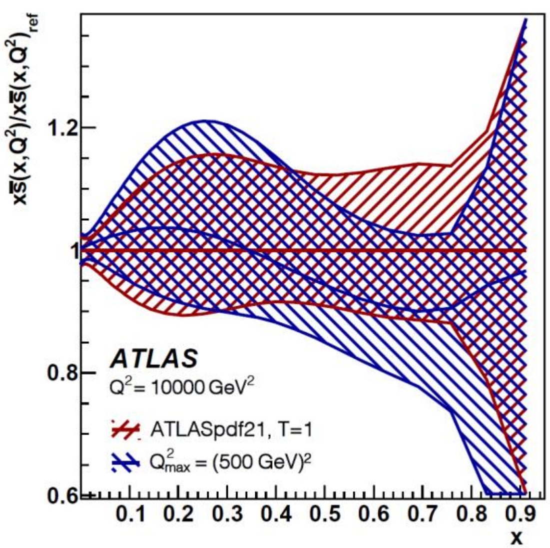
$u_{valence}$ ratio



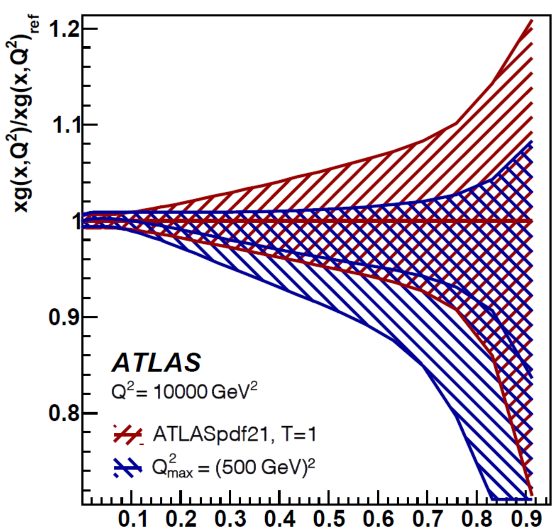
$d_{valence}$ ratio



$\bar{s}$ ratio



gluon ratio

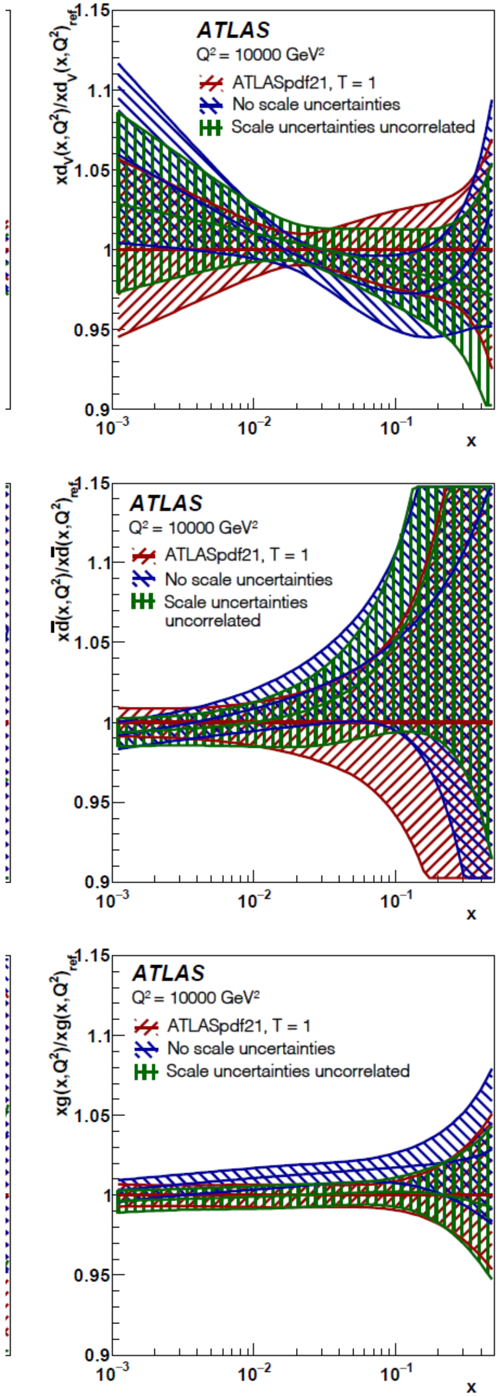


This cut mostly removes inclusive jet production data.

The effect is only seen at high x –note linear x scale- PDFs are not significantly changed

These changes would barely show up on our usual log scale in x

See arXiv:2307.10370
For strategies



There is a further issue: Impact of scale uncertainties

PDFs are extracted at finite order, the current state of the art is NNLO

How much difference does this make?

We use the variation of uncertainties on the choice of scale for the process as a measure of the missing higher order corrections (e.g. N3LO and higher)

The natural scale for W,Z boson production is the mass of the boson. This is varied by a factor of two to evaluate the scale uncertainty.

The plots show the change in the PDFs when including or not including scale uncertainty for W, Z boson production under two assumptions:

- Scale uncertainties correlated between W and Z and between data taken at 7 and 8 TeV
- Scale uncertainties correlated between W and Z but not between data taken at 7 and 8 TeV

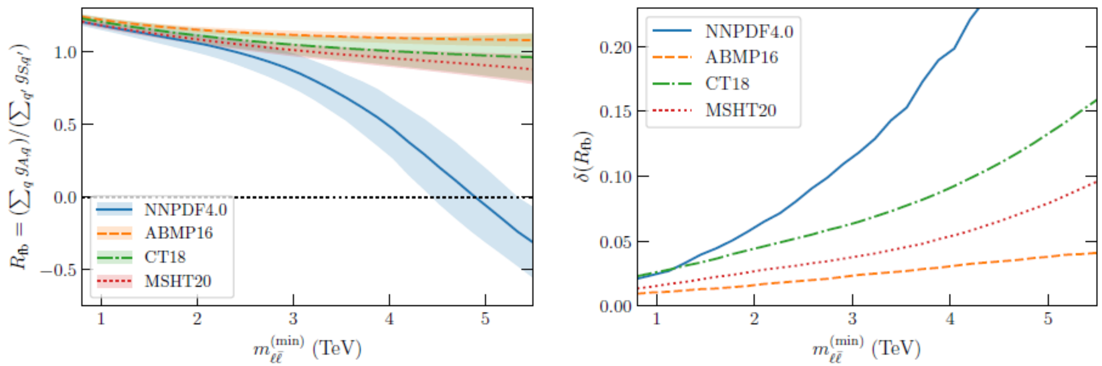
Again this is not a very big effect but it matters if we are striving for ultimate accuracy

Now being taken into account together with the N3LO studies

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?

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



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_{fb} vanishes at large $m_{\ell\bar{\ell}}$