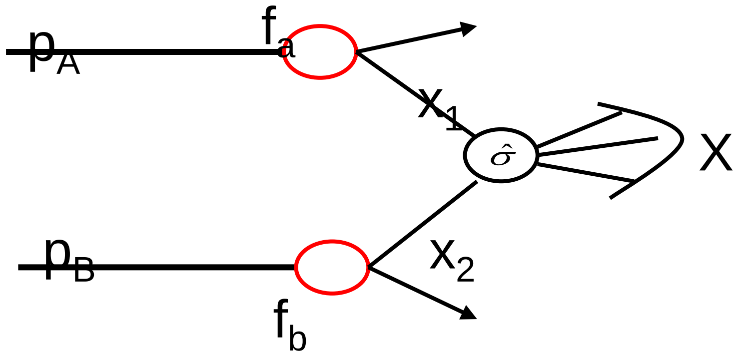
Lecture-3: Modern Parton Distribution Functions

And why they matter at the LHC

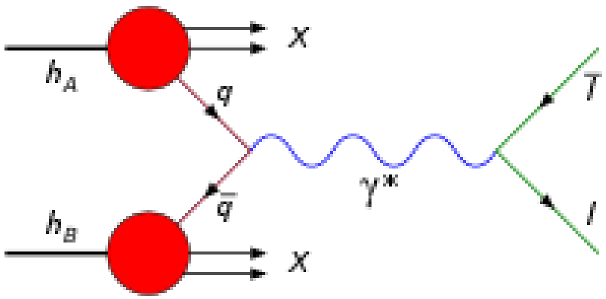
- Why do PDFs differ, what are the sources of uncertainty and what is the consequence for the LHC measurements?
- How well are we doing?
- Is there progress and how does the LHC contribute?
- How can we improve in future?



$$\sigma_X = \sum_{a,b} \int_0^1 dx_1 dx_2\, f_a(x_1, \mu_F^2)\, f_b(x_2, \mu_F^2) \times \hat{\sigma}_{ab \to X}\left(x_1, x_2, \{p_i^\mu\}; \alpha_S(\mu_R^2), \alpha(\mu_R^2), \frac{Q^2}{\mu_R^2}, \frac{Q^2}{\mu_F^2}\right)$$

When a collision happens at the LHC a parton from one of the protons (A) takes a fraction x_1 of the momentum of this proton and a parton from the other proton (B) takes a fraction x_2 of the momentum of this proton, such that **the centre-of-mass energy squared of this collision is not s =(13 TeV)^2 it is x_1x_2s**

Thus the energy involved in each collision– its scale- is different
AND the probability of each collision depends on the joint probability that proton A contained a parton of momentum fraction x_1 , $f_a(x_1)$, and proton B contained a parton of momentum fraction x_2 , $f_b(x_2)$, and that these two partons were of the right type, or flavour to interact to make final state X (as embodied in the cross section for interaction $\sigma_{ab\to X}$).
The probabilities * momentum fractions: $x_1f_a(x_1)$, $x_2f_b(x_2)$, are the parton momentum distributions or PDFs



Drell-Yan is one of the simplest processes

$$\sigma(pp \to cdX) = \sum \iint dx_a dx_b f_a(x_a, Q^2) f_b(x_b, Q^2) \hat{\sigma}(ab \to cd)$$

- **Parton kinematics**

$$\vec{p}_A = x_A\left(\sqrt{s}/2\right); \vec{p}_B = -x_B\left(\sqrt{s}/2\right)$$

$$\hat{s} = (E_A + E_B)^2 - (\vec{p}_A - \vec{p}_B)^2$$

$$\hat{s} = 4p_A p_B = x_A x_B s = M^2$$ Invariant mass squared of lepton pair

Q^2: scale of parton –parton reaction, e.g. M^2

83

- **$\mu^+\mu^-$ via virtual photon: parton-parton cross section.**

$$\hat{\sigma}(q_a \bar{q}_a \to \mu^+\mu^-) = \frac{e_a^2}{3}\frac{4\pi\alpha^2}{3\hat{s}}$$

- **Master formula:**

$$\frac{d^2\sigma(\mu^+\mu^-)}{dx_a dx_b} = [f_A(x_a) f_B(x_b) + \leftrightarrow]\frac{e_a^2}{3}\frac{4\pi\alpha^2}{3\hat{s}}$$

$$\tau = M^2/s = x_a x_b$$

$$2y = \ln\left(\frac{x_a}{x_b}\right)$$

- **Change variables**

$$y = \frac{1}{2}\ln\left(\frac{E + p_z}{E - p_z}\right); \tau = \hat{s}/s = M^2/s$$

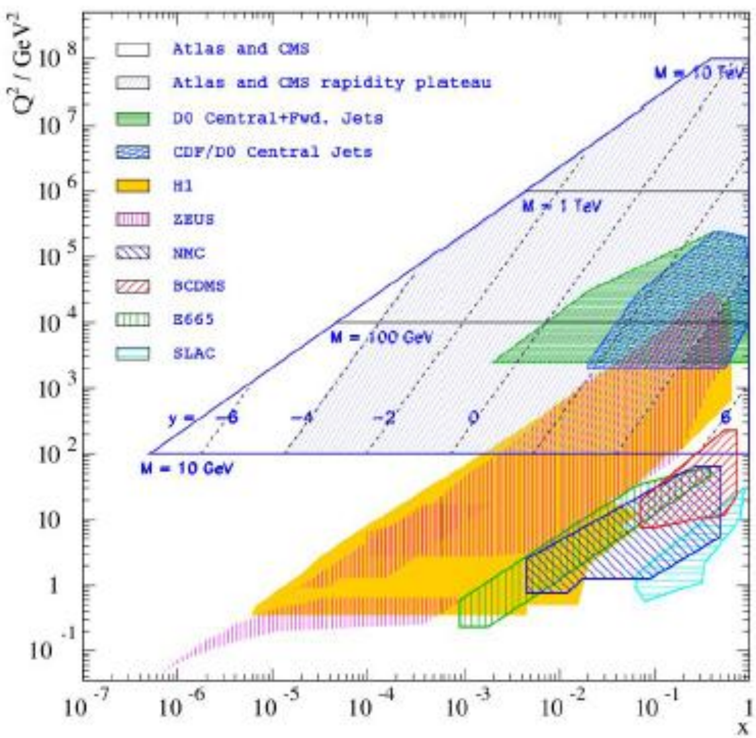
- **Scaling cross section:**

$$\frac{d^2\sigma(\mu^+\mu^-)}{d\tau dy} = [f_A(x_a) f_B(x_b) + \leftrightarrow]\frac{e_a^2}{3}\frac{4\pi\alpha^2}{3\hat{s}}$$

$$x_{a,b} = \frac{M}{\sqrt{S}}\exp(\pm y)$$

We evolve the PDFs from the HERA region up to the LHC region

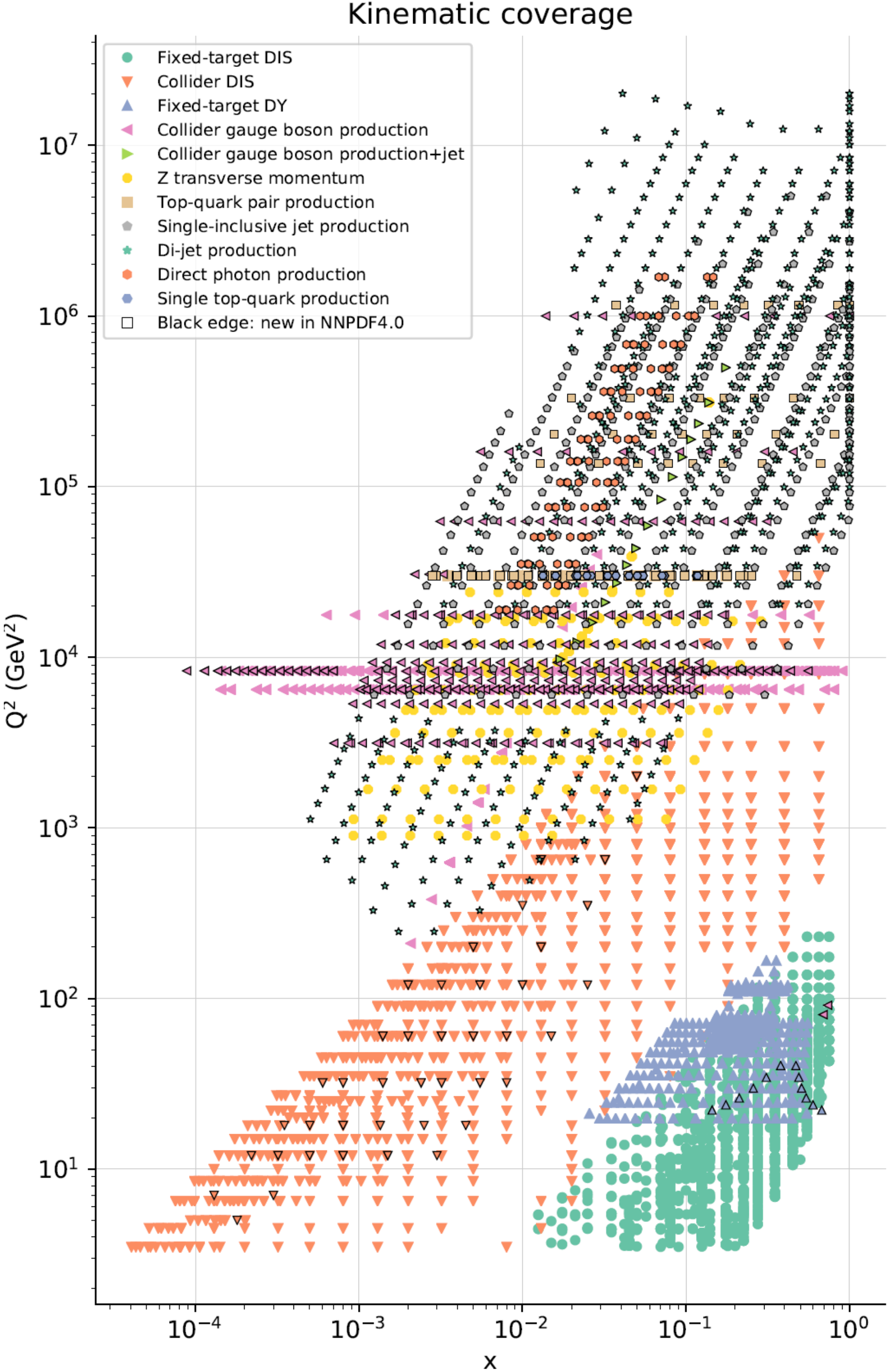
LHC kinematics



$$x_{1,2} = \frac{M}{\sqrt{S}} \exp(\pm y)$$

- To create an object with a mass *M* and rapidity *y* for the LHC center of mass energy *S* we need to have two partons with momentum fraction of corresponding proton of x_1 and x_2
- At electroweak scale, LHC needs PDFs with $x > 10^{-4}$

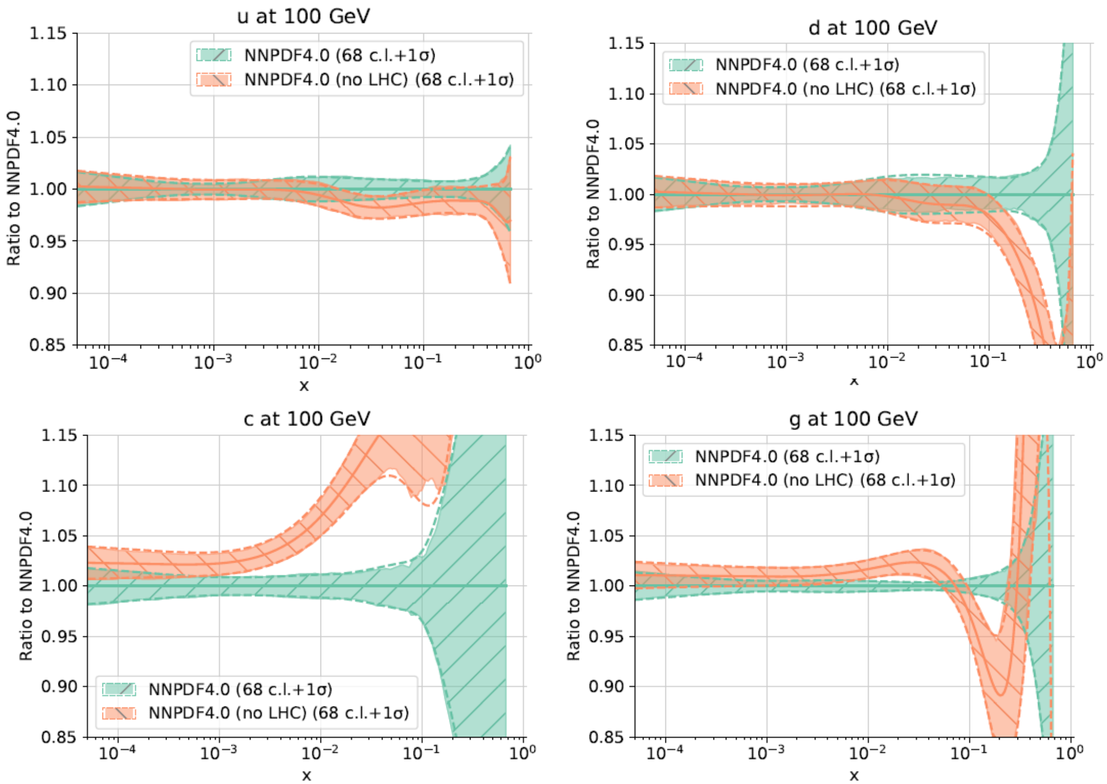
And assuming the PDFs are universal we can predict LHC cross sections.
We can also use SM LHC processes to improve the PDFs….



We now use many other processes than deep-inelastic scattering for the determination of PDFs
We use many SM processes that can be reliably calculated to NNLO in fixed order QCD

- Drell-Yan data from fixed targets 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
- W and Z +jet spectra, or W,Z pt spectra
- W and Z +heavy flavours
- Beware: there may be new physics at high scale that we 'fit away'

How do the PDFs look before/after LHC data?



Some change in shape
Some reduction in uncertainty

Illustration from NNPDF

But several groups extract PDFs and there are significant differences between them because of slightly different model choices:

- Exact choice of data entering fit and cuts impose on them
- Choice of heavy quark masses, heavy quark schemes
- Choice of starting scale for QCD evolution, Value of $\alpha_S(M_Z)$

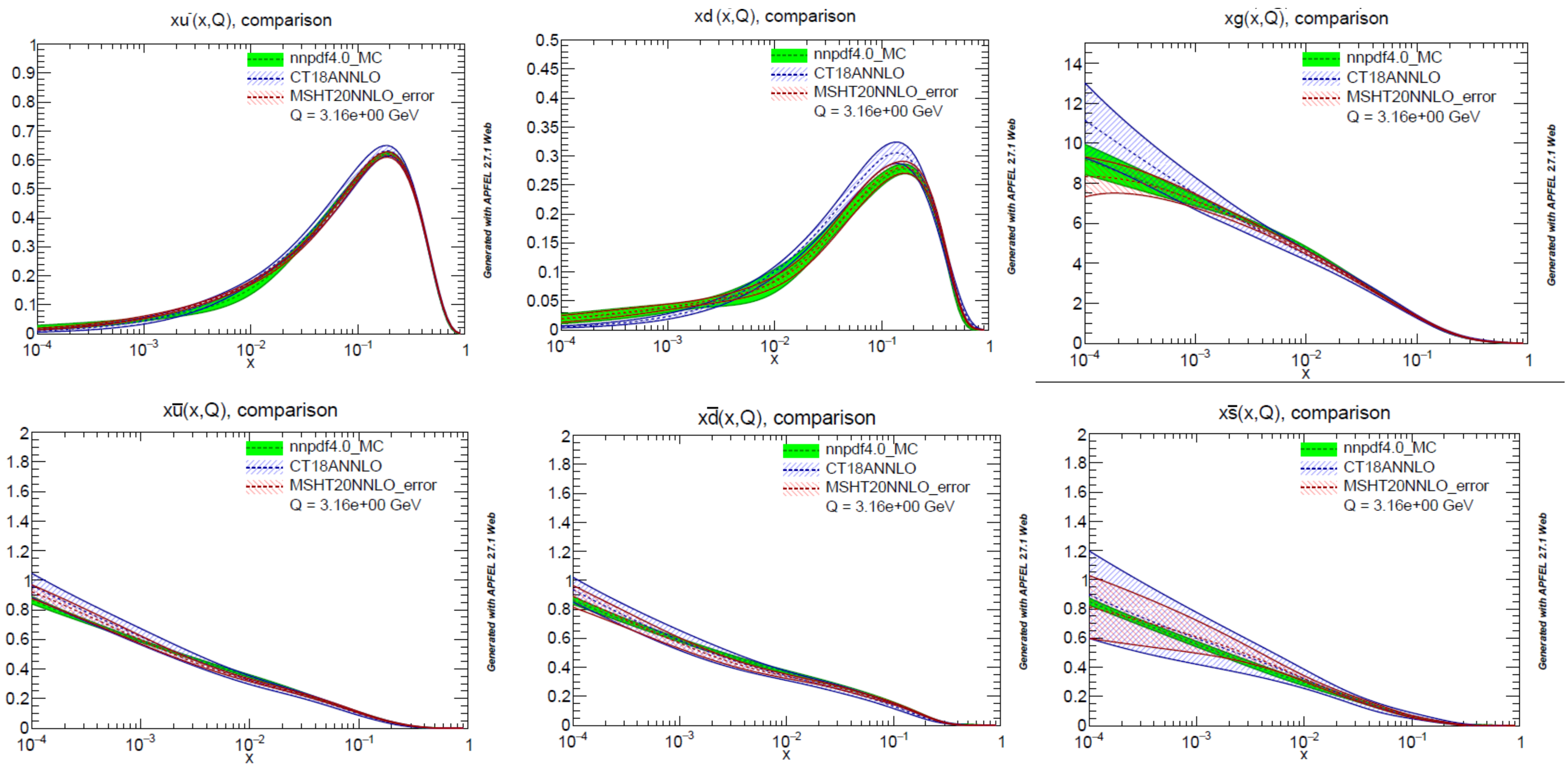
And differing methodology

- Use of parametrization, what parametrization
- Use of NN, what architecture, training, stopping etc
- How are uncertainties determined

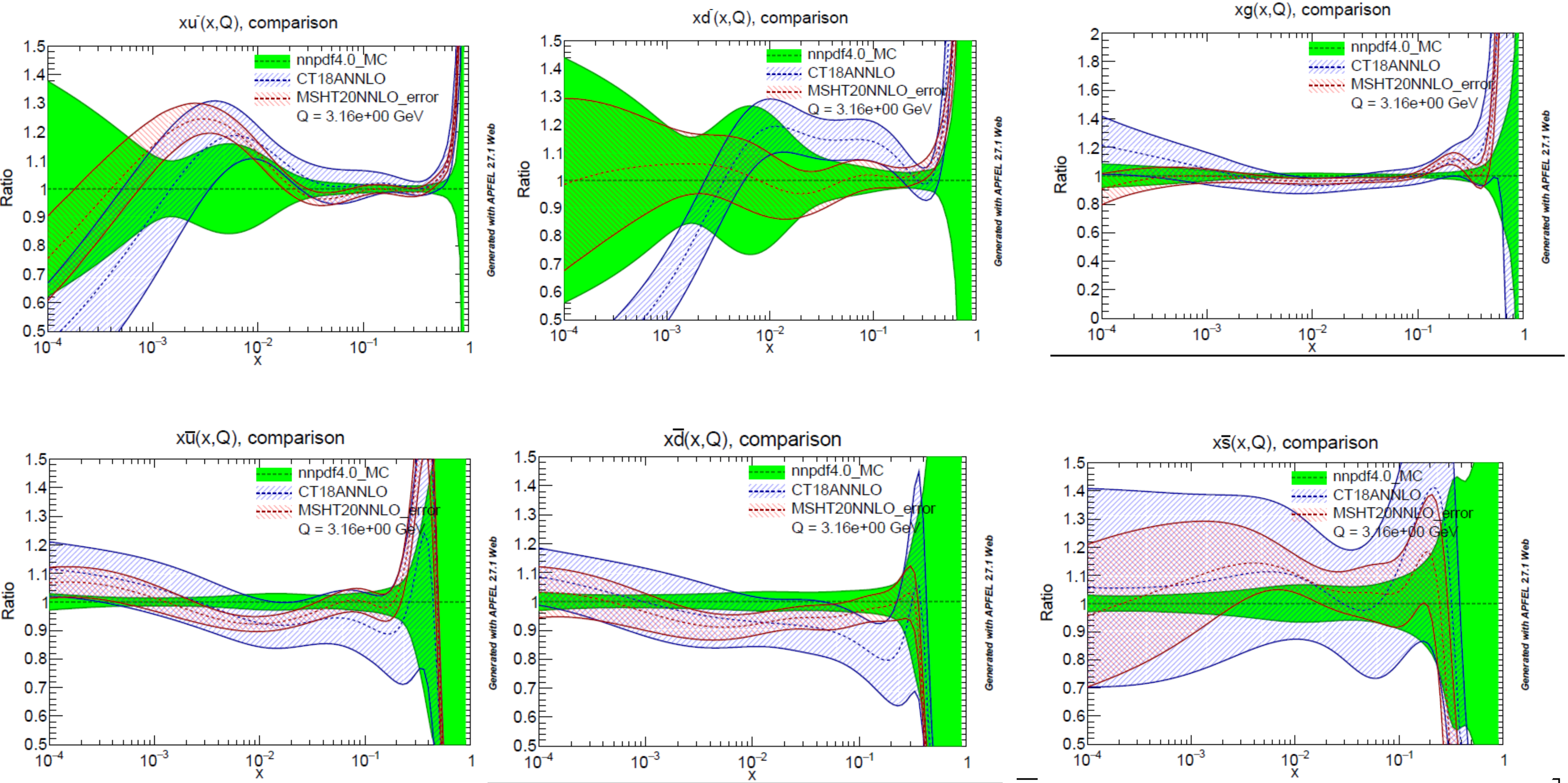
The ‘big three’ of PDF fitting groups are CT, NNPDF and MSHT

The latest CT18, NNPDF4.0, MSHT20. Be agnostic in the choice between these.

The one with the smallest uncertainties is not necessary the best. We want accuracy as much as we want precision



This looks as if we are doing reasonably well –but look at ratios….



Differences are more obvious in ratio. They are large at small-x and at high-x, where there is less data. The differing model choices matter. Different methodologies matter.

So also do theory choices, standard is NNLO, often using NNLO/NLO ‘k-factors’ but such calculations can differ. Even when using direct NNLO grids - treatment of grid uncertainties can differ. Now there is also progress on N3LO PDF analyses.

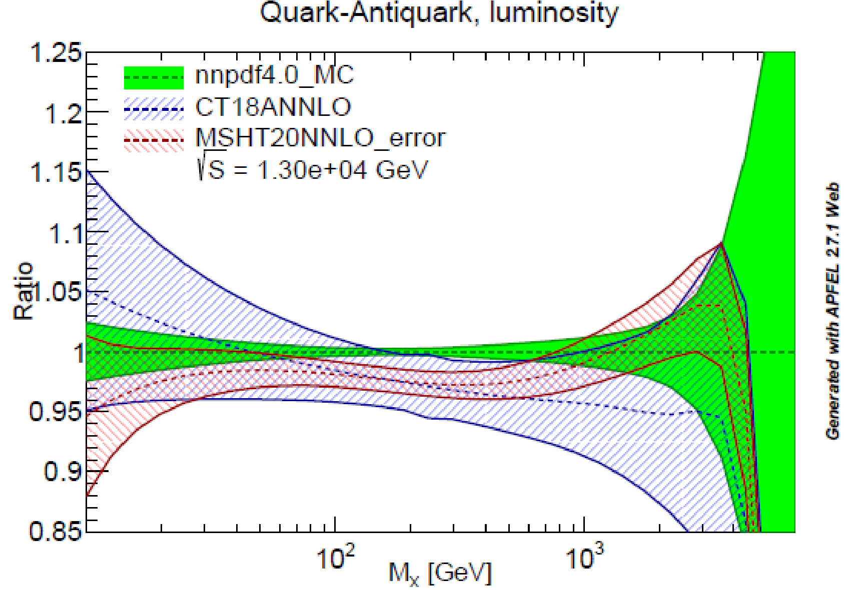
PDFs also differ in how they evaluate their uncertainties some use enhanced $\chi 2$ tolerances -- closer to the hypothesis testing criterion– but this is a whole lecture series in itself.

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

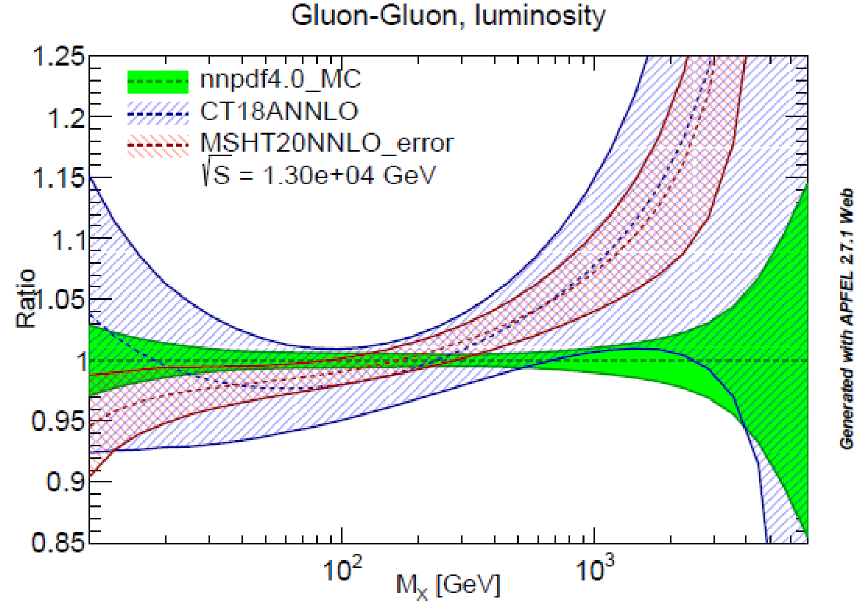
The quark-antiquark and gluon-gluon luminosities for various PDFs are compared here for 13 TeV LHC running in terms of the centre of mass energy of the parton sub- process M_X

Small M_X corresponds to small x and

Large M_X to large x



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



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

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

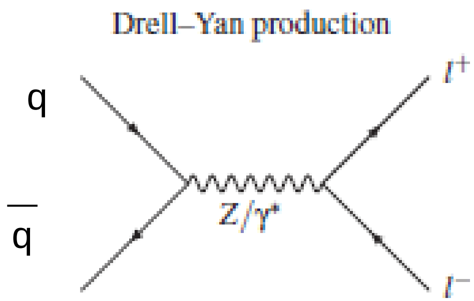
Or for gluon-gluon production of Higgs at ---Mx~125 GeV

the parton-parton luminosities are fairly well known- well this is as good as it gets--

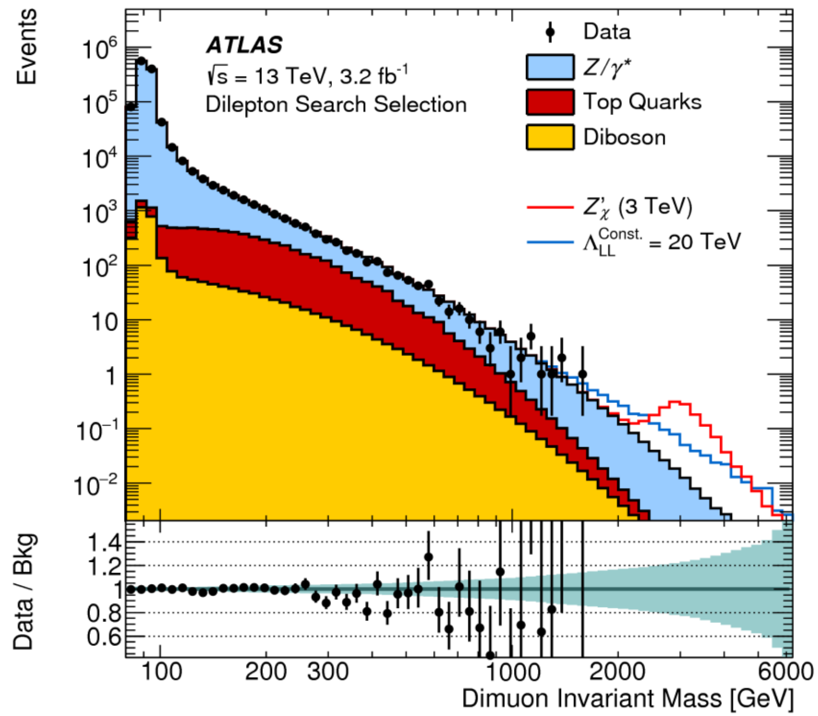
This is not so for higher mass particles that could be produced by ‘Beyond’ Standard Model (BSM) physics

Uncertainty in the high-x sea?-one example

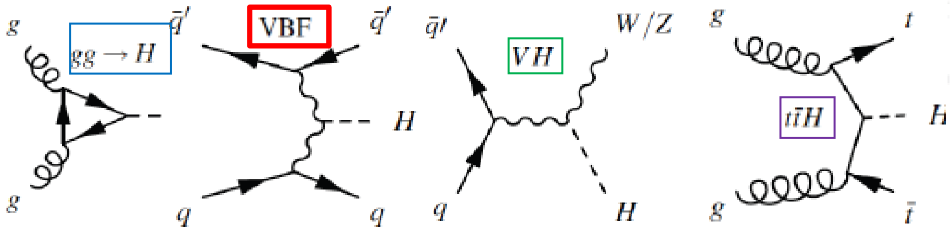
Current BSM searches in High Mass Drell-Yan are limited by high-x antiquark uncertainties as well as by high-x valence quark uncertainties



Drell-Yan is a term for q-qbar → $\mu^+ \mu^-$ collisions mediated by Z or virtual γ,Z bosons.
Some new theories predict higher mass Z' states, these have been excluded up to 2 TeV
The main reason we cannot do better is that the PDF uncertainty on the 'normal' Standard Model background is too big.

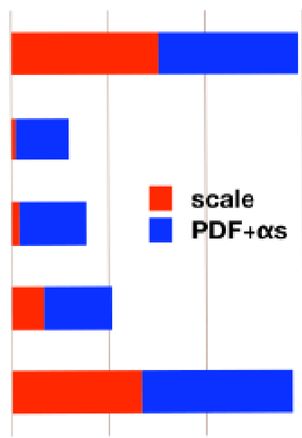


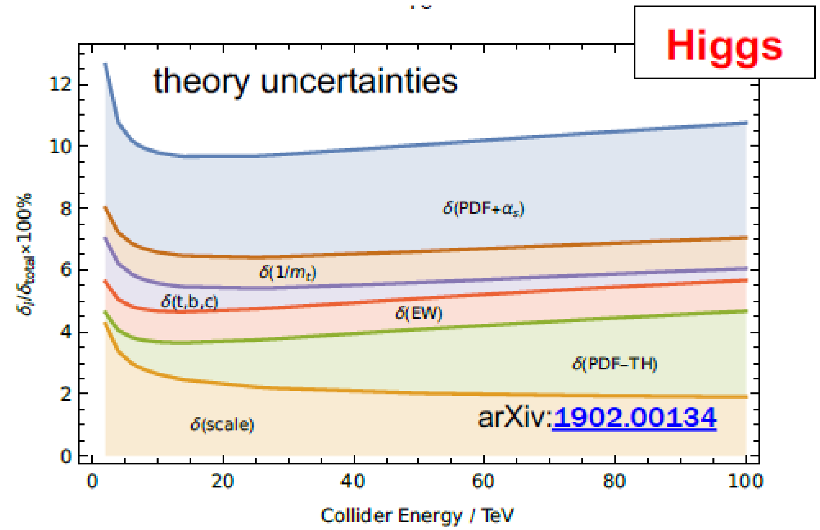
What about the Higgs? Gluon-gluon to Higgs is the dominant channel



PDFs and the Higgs

		σ (8 TeV)	uncertainty
NNLL QCD +NLO EW	gg→H	19.5 pb	14.7%
NNLO QCD +NLO EW	VBF	1.56 pb	2.9%
	WH	0.70 pb	3.9%
	ZH	0.39 pb	5.1%
NLO QCD	ttH	0.13 pb	14.4%





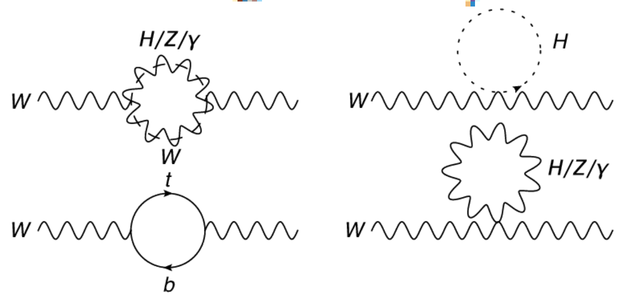
The uncertainty on the calculation for the SM Higgs cross section has substantial uncertainty from the PDFs– mainly the gluon PDF

All indications are that the Higgs boson that we have seen is the Higgs boson of the Standard Model (SM) , but measurements are not as yet very accurate.

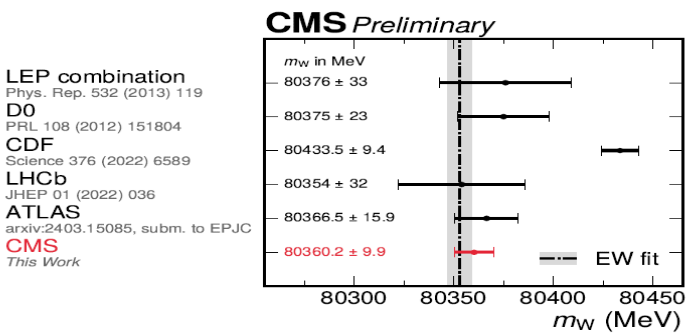
Beyond Standard Model (BSM) physics can manifest itself as small discrepancies from SM predictions. We need to reduce uncertainties on our predictions.

As well as limiting our ability to identify BSM effects at high M_X , uncertainties on PDFs also limit indirect observations of new physics which we may hope to make by measuring discrepancies from the Standard Model (SM) values for fundamental parameters such as m_W – the W mass

$$m_W^2\left(1-\frac{m_W^2}{m_Z^2}\right)=\frac{\pi\alpha}{\sqrt{2}G_\mu}(1+\Delta r),$$



The W mass is predicted in the SM in terms of other SM parameters like the fine structure constant and the weak coupling G, but Δr represents higher order loops in the diagrams which are presently calculated with known particles like the top quark or Higgs, but could also contain BSM particles.

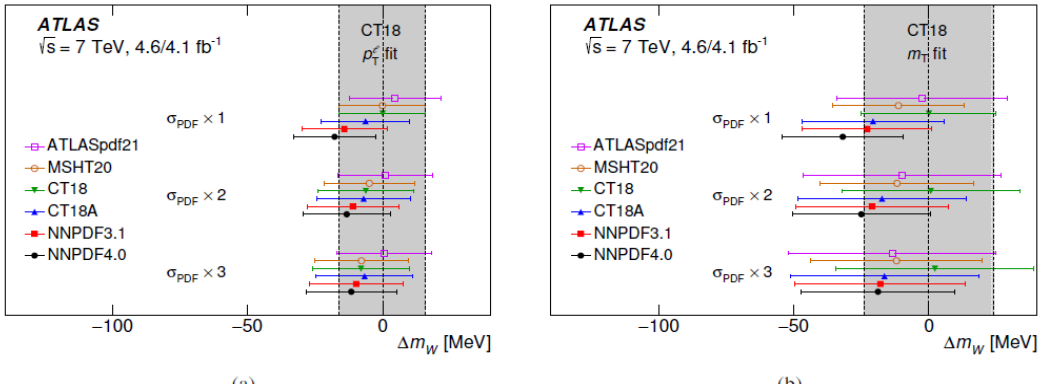


In that case the value of m_W would differ from its SM value
And indeed that is what we saw in the 2022 CDF measurement from FermiLab

However, this has been checked at the LHC- the new CMS measurement agrees with SM
The plot shows the latest measurements from ATLAS and CMS (2024).
A major contribution to the uncertainty of these measurements is the PDF uncertainty, which comprises half the uncertainty in both cases. LHC uses p-p not p-pbar and its kinematic reach is such that most collisions producing W are sea-quark collisions.

What is our problem?

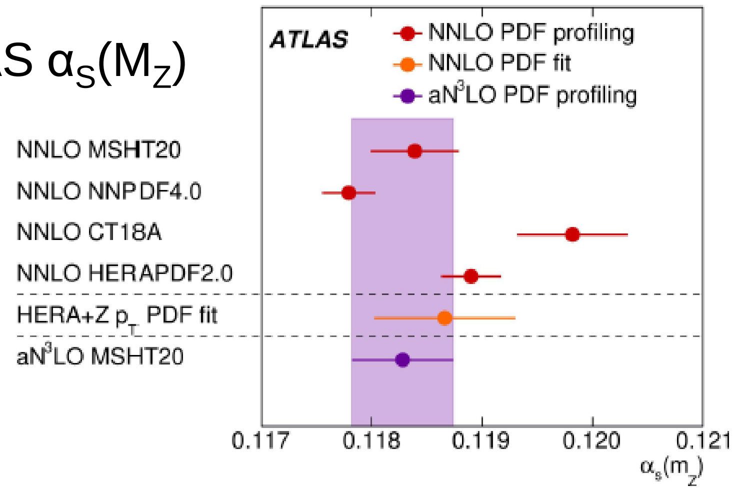
ATLAS Mw



CMS Mw

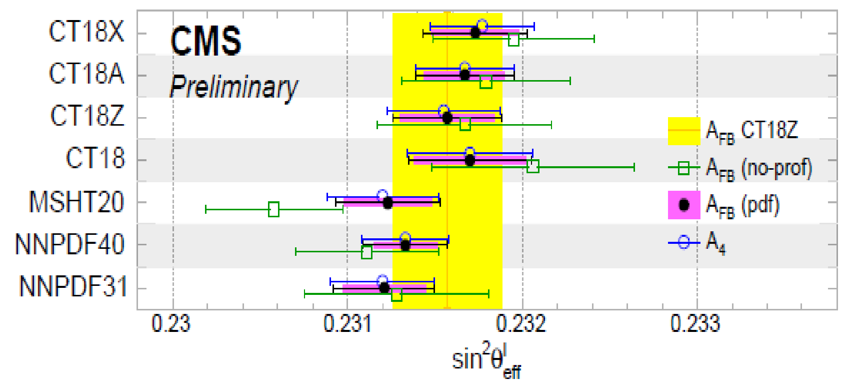
PDF set	Extracted m_W (MeV) Original σ_{PDF}	Scaled σ_{PDF}
CT18Z	$80\,360.2 \pm 9.9$	
CT18	$80\,361.8 \pm 10.0$	
PDF4LHC21	$80\,363.2 \pm 9.9$	
MSHT20	$80\,361.4 \pm 10.0$	$80\,361.7 \pm 10.4$
MSHT20aN3LO	$80\,359.9 \pm 9.9$	$80\,359.8 \pm 10.3$
NNPDF3.1	$80\,359.3 \pm 9.5$	$80\,361.3 \pm 10.4$
NNPDF4.0	$80\,355.1 \pm 9.3$	$80\,357.0 \pm 10.8$

ATLAS $\alpha_S(M_Z)$



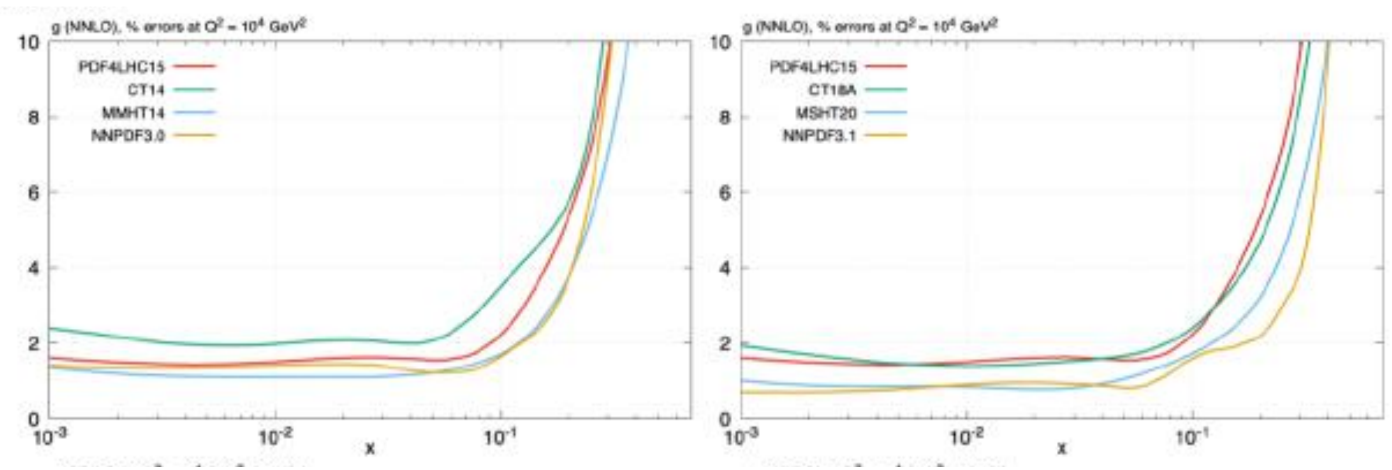
PDF set	$\alpha_s(m_Z)$	PDF uncertainty
MSHT20 [37]	0.11839	0.00040
NNPDF4.0 [84]	0.11779	0.00024
CT18A [29]	0.11982	0.00050
HERAPDF2.0 [65]	0.11890	0.00027

CMS $\sin^2\theta_W$

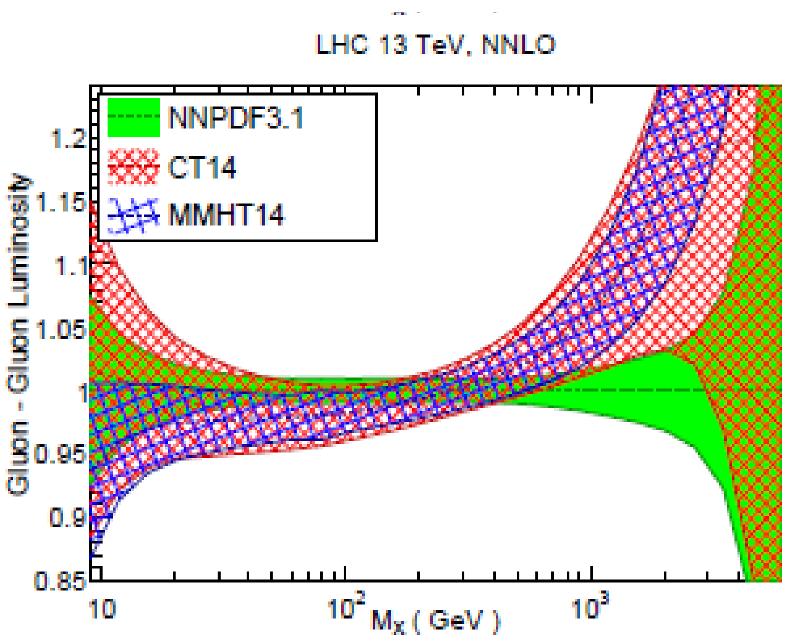


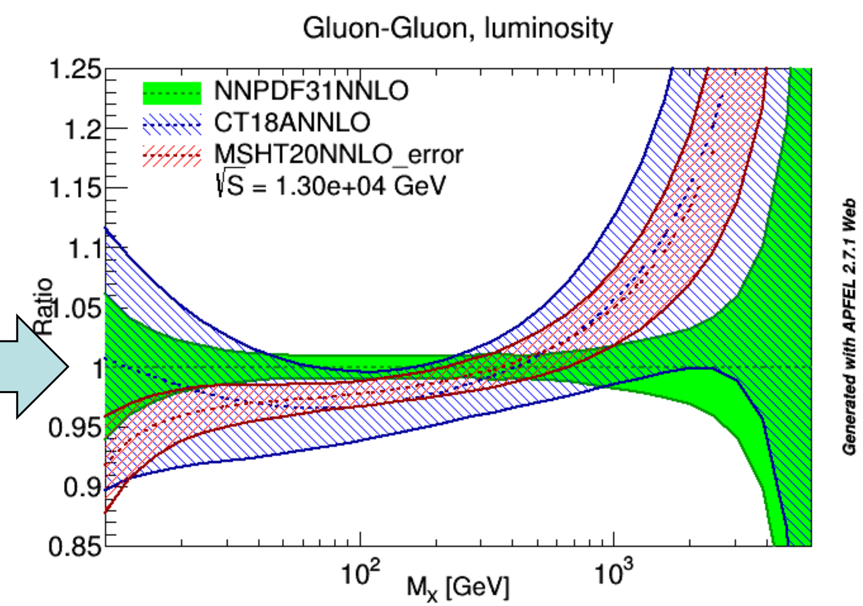
Has there been progress in recent years?

Gluon Uncertainties

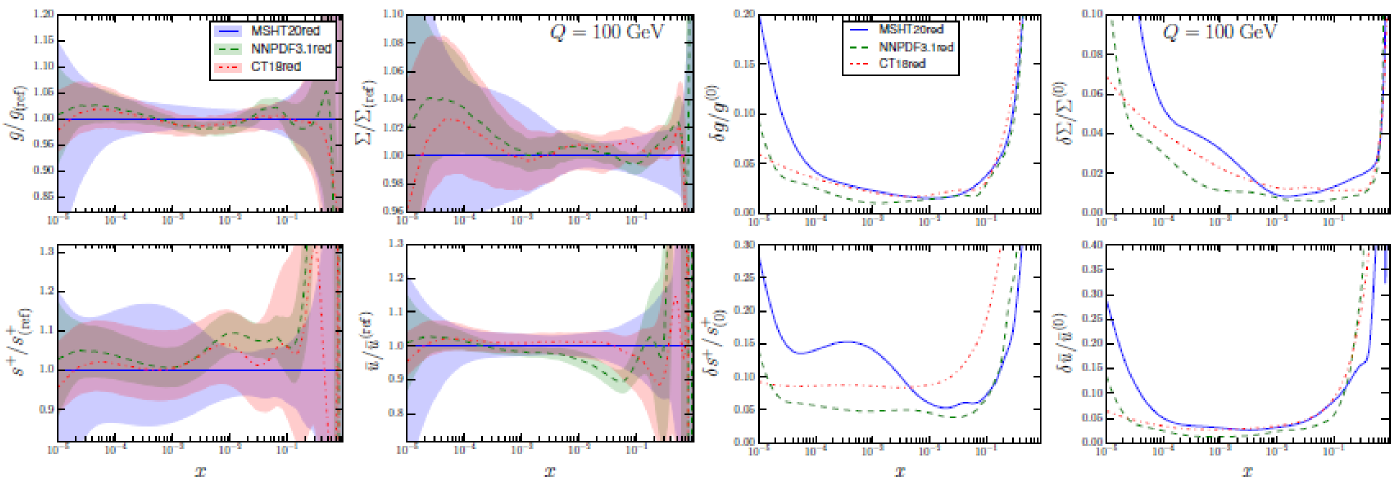


As the uncertainties of each individual PDF decrease with the input of more information, the divergence of the PDFs from each other has increased





The PDF4LHC group makes combinations of the PDFs from the three main fitting groups NNPDF, CT and MSHT (arxiv:2203.05506)
First try to understand differences by using a common data set and common settings for heavy quark masses and $\alpha_S(M_Z)$



- 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.

BUT It is not recommended to use these reduced fits,
greater consistency does not mean greater accuracy—
the methodological differences in the main fits are there for a reason!

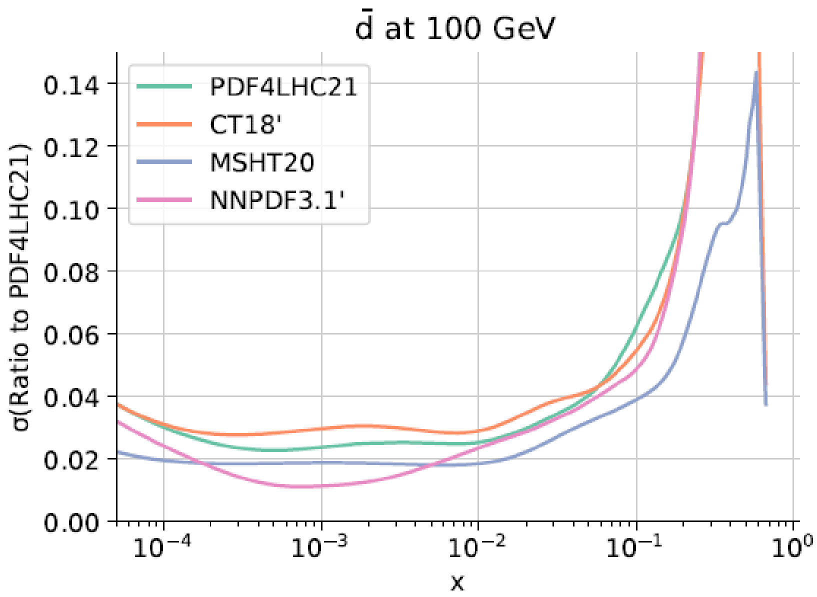
PDF4LHC21 combination uses the published PDFs with very little modification:
modified CT' (HQ masses), NNPDF3.1'(closer to common data set) and MSHT (as published)
but does not strive for common methodology

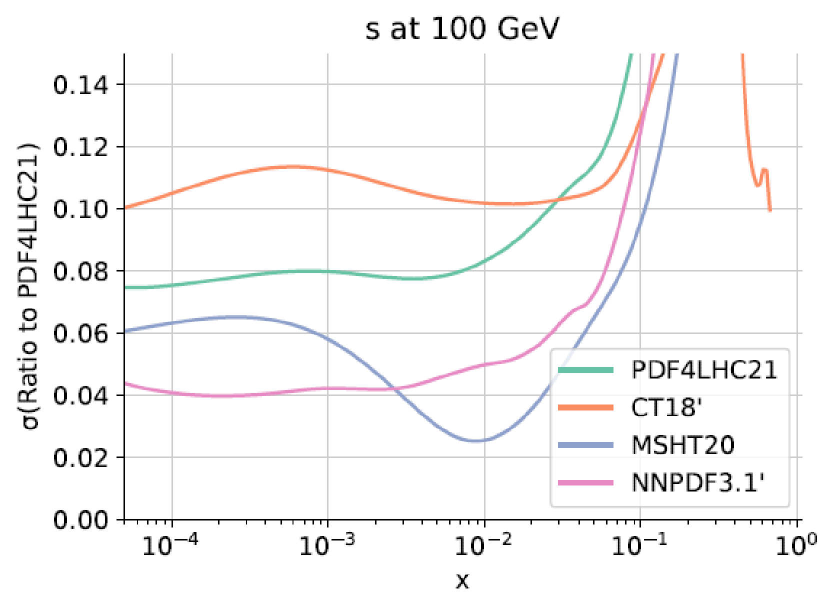
The combination is a statistical combination of the three input PDFs.

Where the they are consistent the resulting PDF4LHC uncertainty represents an average of the the PDF set uncertainties—generally closest to, though smaller than, the largest uncertainty of the three, namely CT18.

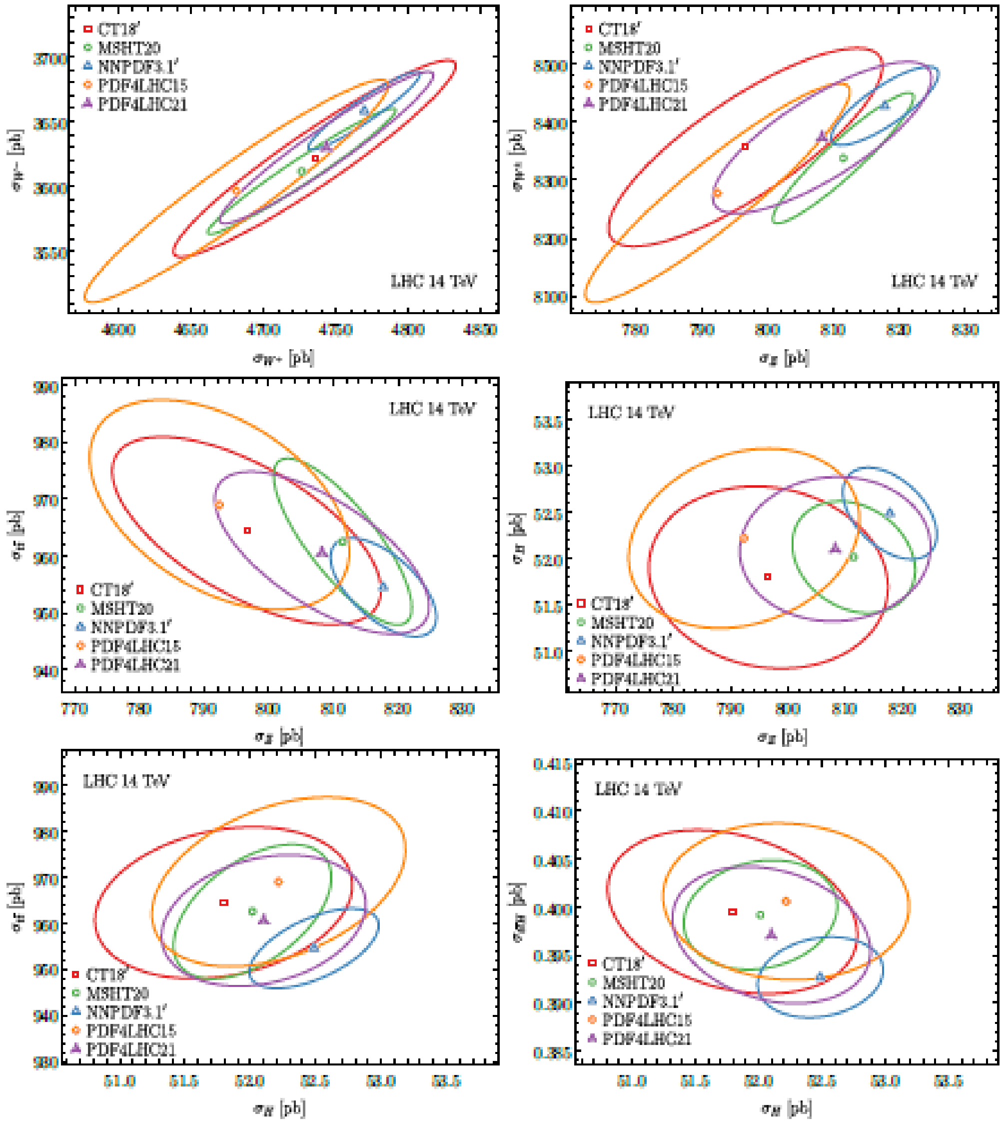
But where there are discrepancies the PDF4LHC uncertainty can be larger than those of any of the individual sets since it includes the spread in the central prediction

x





These plots represent ratios of uncertainties

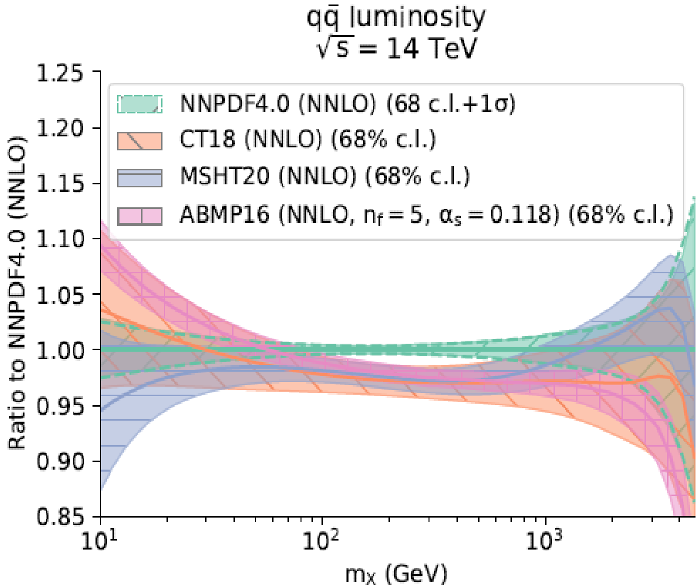


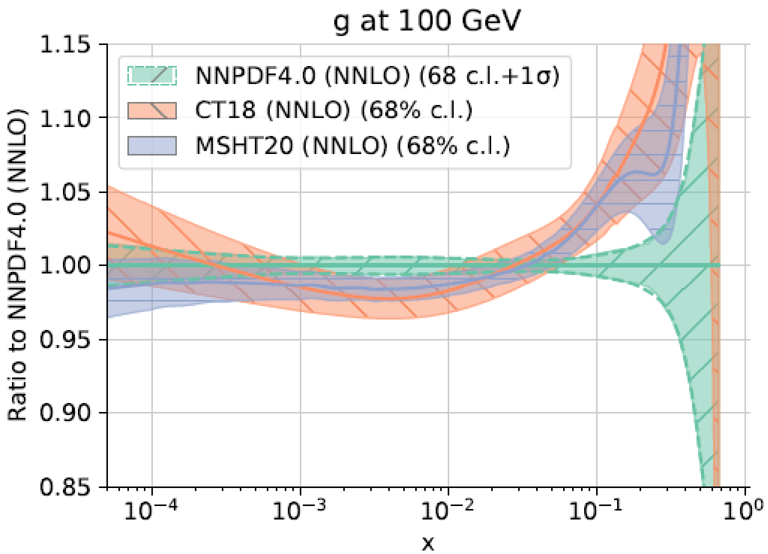
The PDF4LHC group makes combinations of the PDFs from the three main fitting groups NNPDF, CT and MSHT

The PDF4LHC15 combination has just been superseded by the PDF4LHC21 combination

There IS an improvement in uncertainty BUT this is not enough to reduce the PDF uncertainty on LHC measurement of m_W, $\sin^2\theta_W$, $\alpha_S(M_Z)$ dramatically

Since the issue of PDF4LHC21 there has been the new PDF set from NNPDF4.0
This has a lot of new data from the LHC and considerable decrease in uncertainty, with respect to NNPDF3.1.
BUT the improvements in uncertainty are **not so much due to the new data**, they are more **due to improvements in their procedure.**
Unfortunately decrease in uncertainty of a single PDF does not help much if there are discrepancies with other PDFs.
The uncertainty on combination of PDFs will remain higher than the uncertainty of any individual PDF set





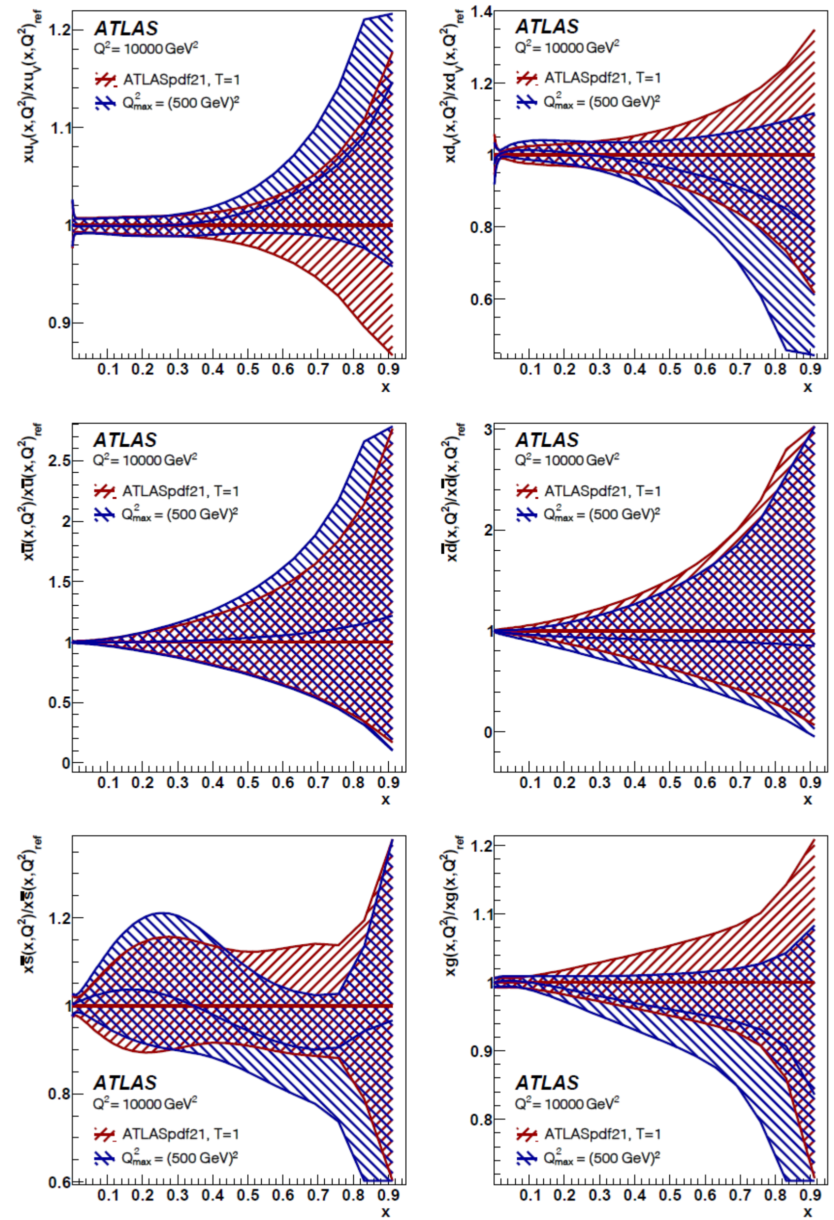
And there are other PDFs eg ABMP, which does not use jet data..
This could be important if we think that high energy jet data may actually contain hidden new physics.

There is a danger when fitting high scale data—such as high pT jet production– of ‘fitting away’ the very BSM effects you would like to look for, ie including the deviations BSM from SM in your PDF fit.

Thus ATLAS also cut all data at scale > 500 GeV, ie Q^2 > 250000 GeV^2
From the fit and re-evaluated the PDFs.

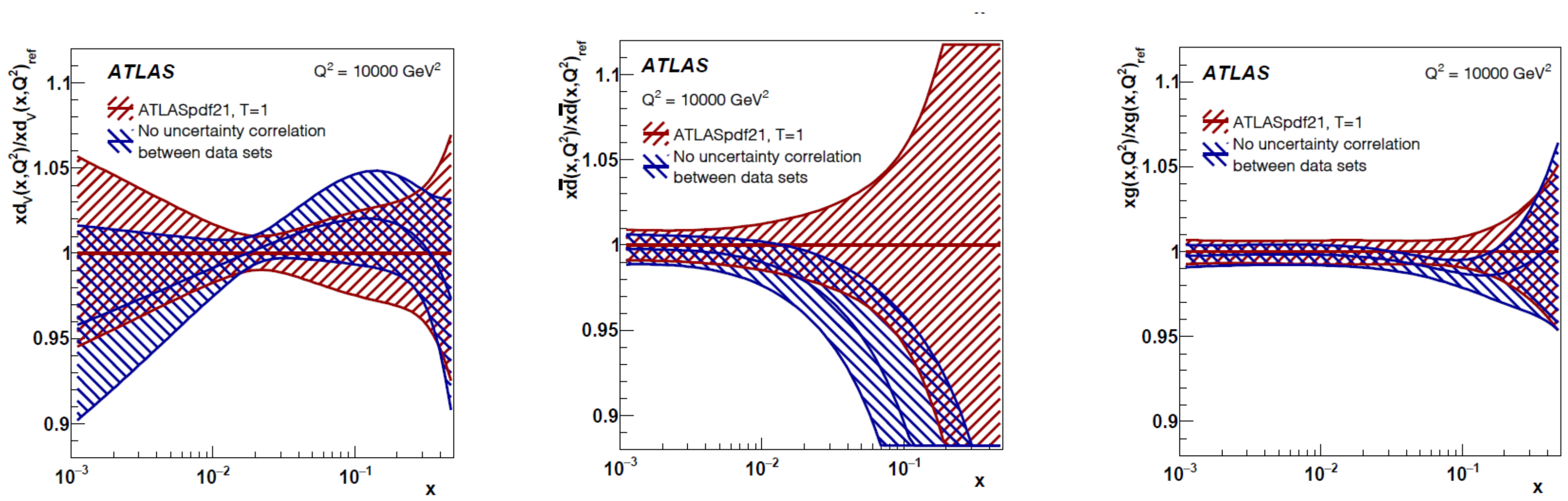
We can see the difference if we look at very high-x.
Differences only exceed 5% for x>0.5.
We cannot claim discovery of new physics on this basis, but we must be vigilant.

There are studies in which the new-physics effect is fitted at the same tie as the PDF--- so far no definitive evidence.

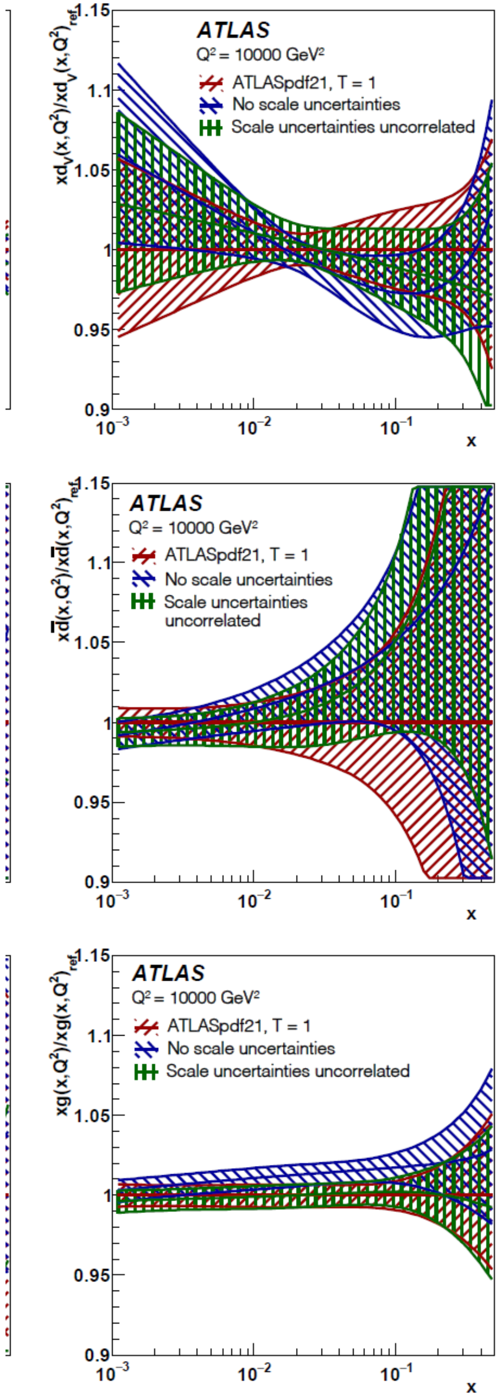


There are further issues in the use of LHC data to fit PDFs

LHC data today consistents of many data sets analysed by different groups, with differing procedures for the evaluation of systematic uncertainties, which makes correlating systematic uncertainties between different measurements difficult.
Such correlations are not usually known/applied
Recent work by ATLAS uses many different types of LHC data, evaluating the largest correlations



The larger correlations come 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



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

How much difference does this make?

We use scale uncertainty as a measure of the missing higher order corrections.

What does this mean?

The scale for DIS is Q, for other processes we make choices

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.

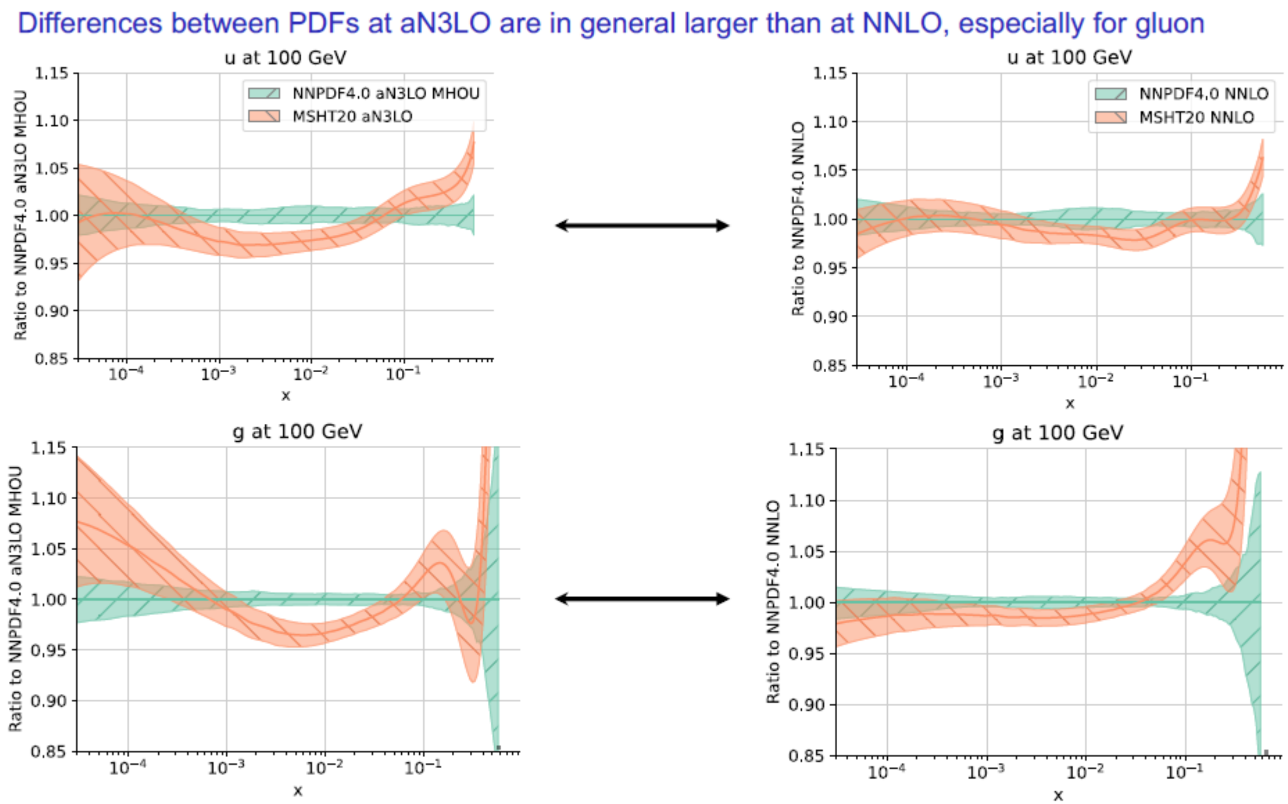
The plots show the change in the PDFs when including or not including scale uncertainty for W, Z boson production under two assumptions:

- 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

This is not a very big effect but it matters if we are striving for ultimate accuracy O(1%)

AND recently there has been a move to N3LO..**which is not fully computed yet** so MSHT and NNPDF have given **approximate** PDFs.

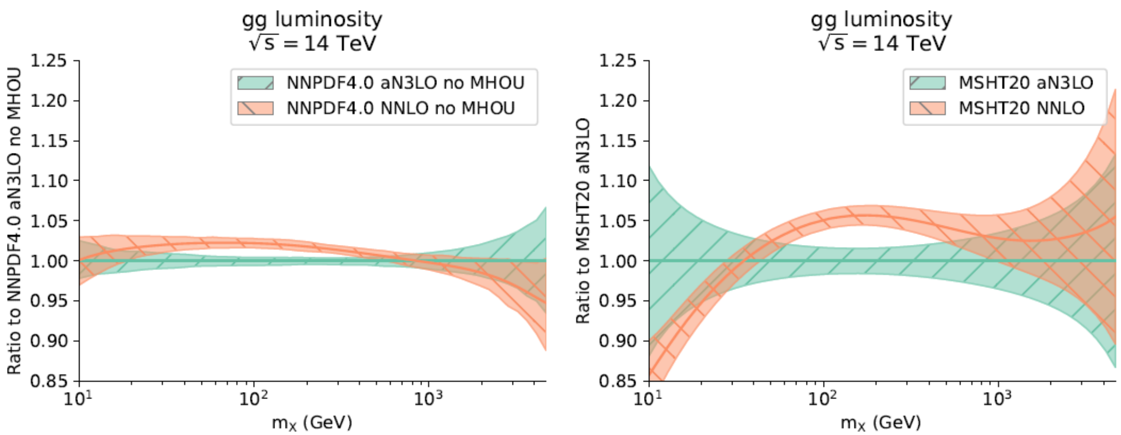
The differences between NNLO and aN3LO are larger than our estimates from MHOUs led us to believe, but also the two groups do not agree very well with each other, especially at low-x

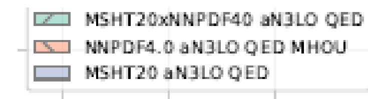


N3LO

NNLO

And that translates to an effect on the low Mx region for the gluon-gluon luminosity and has a ‘knock-on effect’ on the luminosity in the Higgs region Mx =125 GeV





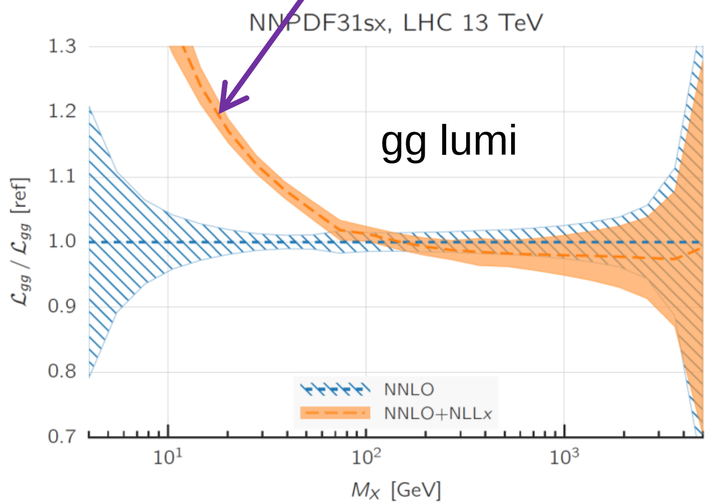
We now have two approximate N3LO PDFs from NNPDF and MSHT
Differing groups have different ways of implementing the aN3LO

For MSHT there is a 5% decrease in luminosity at the Higgs mass, for NNPDF this is more like 2%... BUT either way there is a significant difference.
These come together a bit more with recent updates on N3LO and with QED corrections, which are larger for NNPDF than for MSHT.

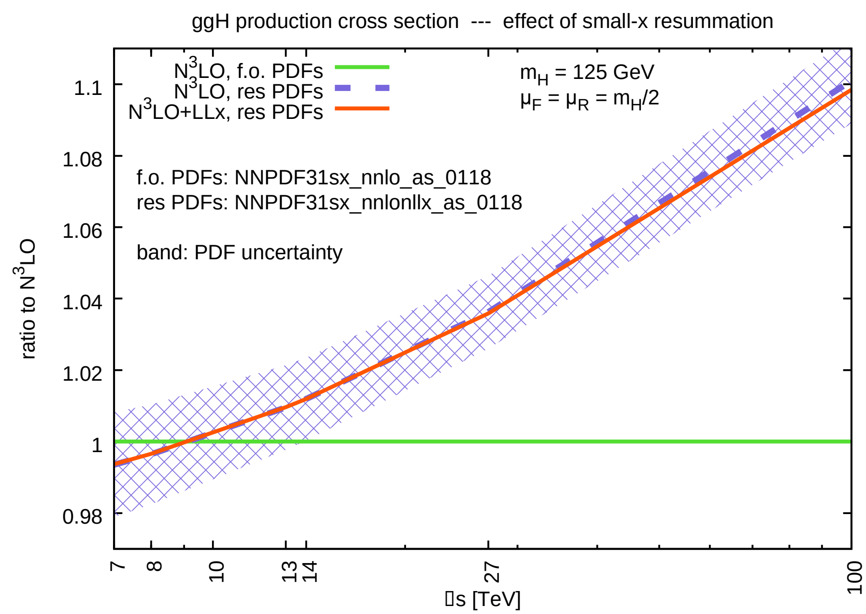
Furthemore aN3LO includes one more term of the ln(1/x) resummation, but what if we considered more terms? What are the consequences at the LHC?

Full ln(1/x) BFKL resummation

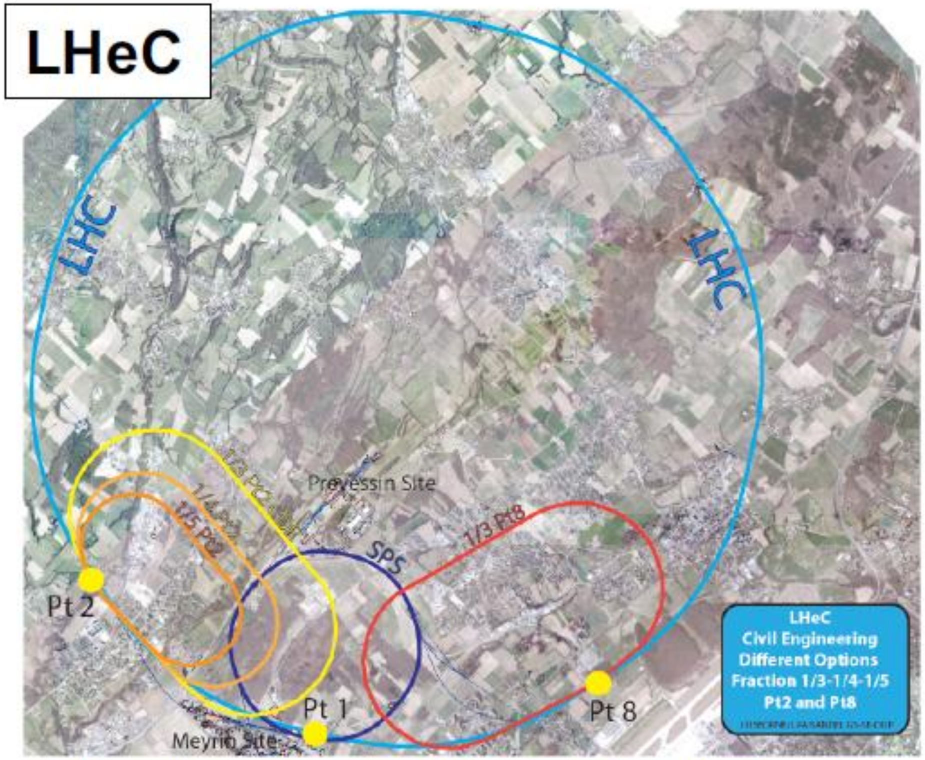
effect of small x resummation on LHC luminosity



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



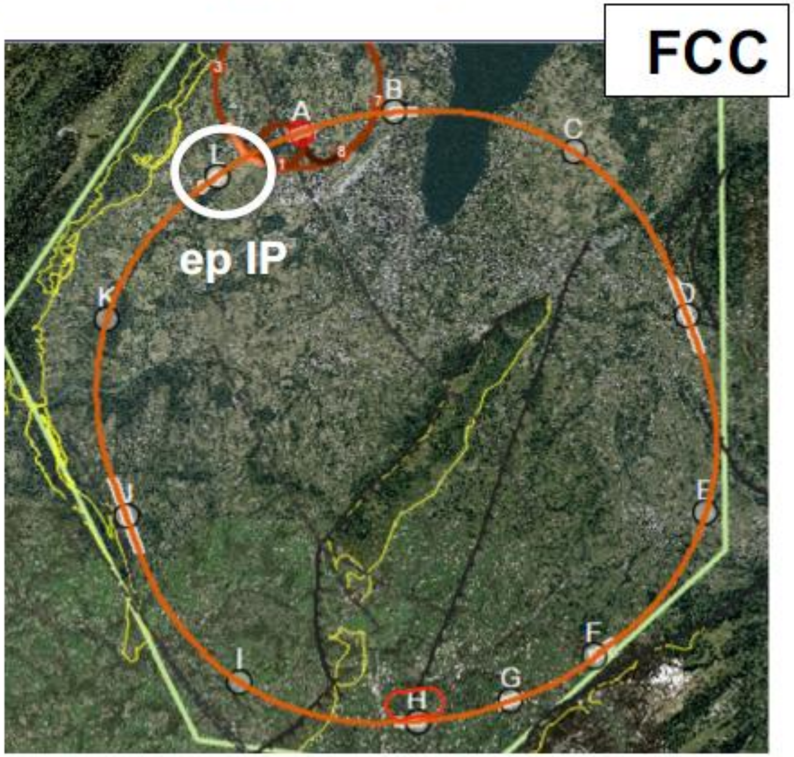
SO HOW MAY WE ACTUALLY DO MUCH BETTER?
One thing that has always made PDF fitters agree better with each other is better data. We need New DIS machines…



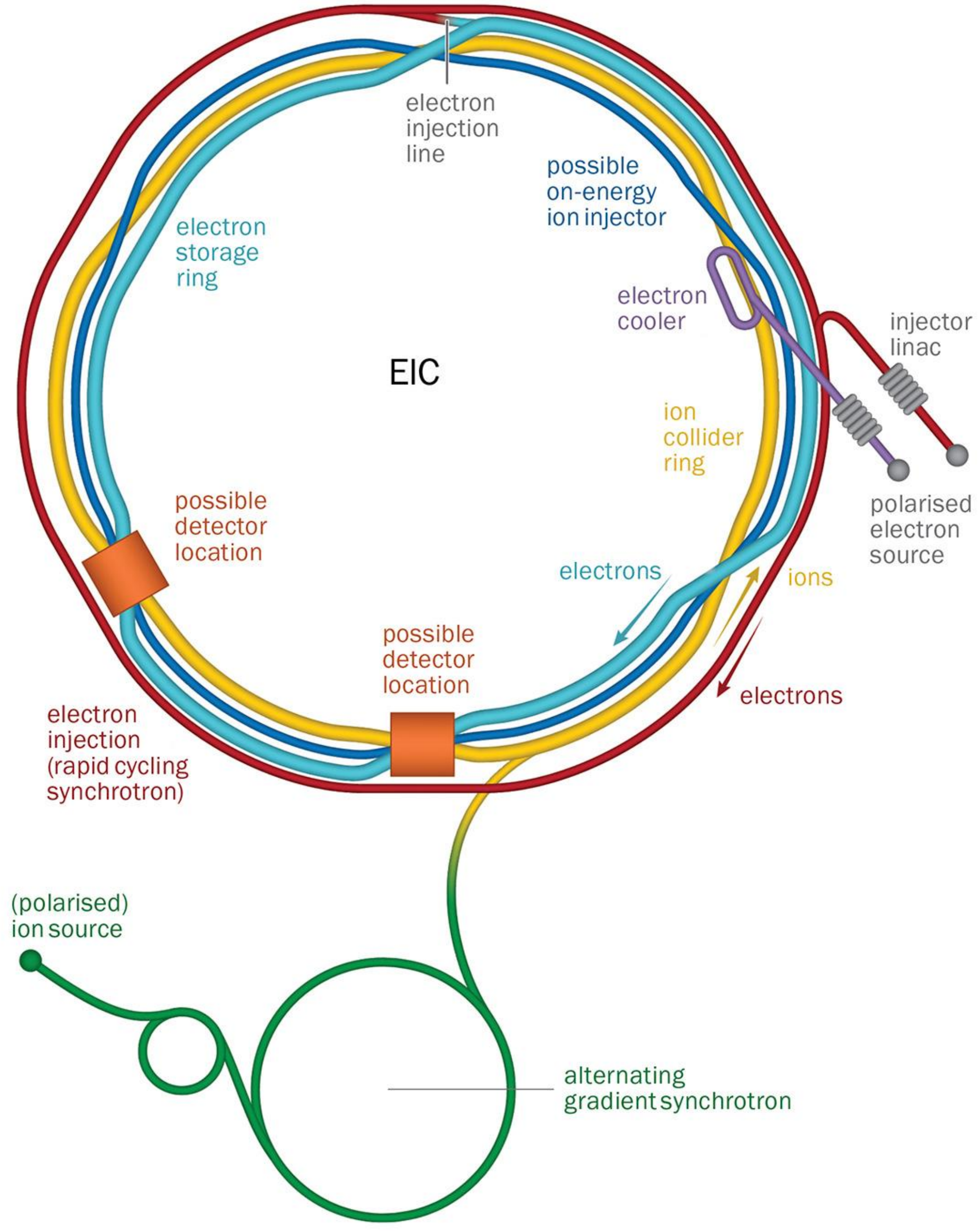
(M Klein, Rencontre du Vietnam, Sept 2017)

LHeC may be considered as a 'bridge' project to any possible future FCC arXiv:2503.17727

LHeC and **FCC-eh**
energy recovery LINAC
e-beam: 60 GeV
$L_{int} \rightarrow 1\ ab^{-1}$



LHeC (FCC-eh) complementary to, synchronous with, **HL-LHC (FCC)**



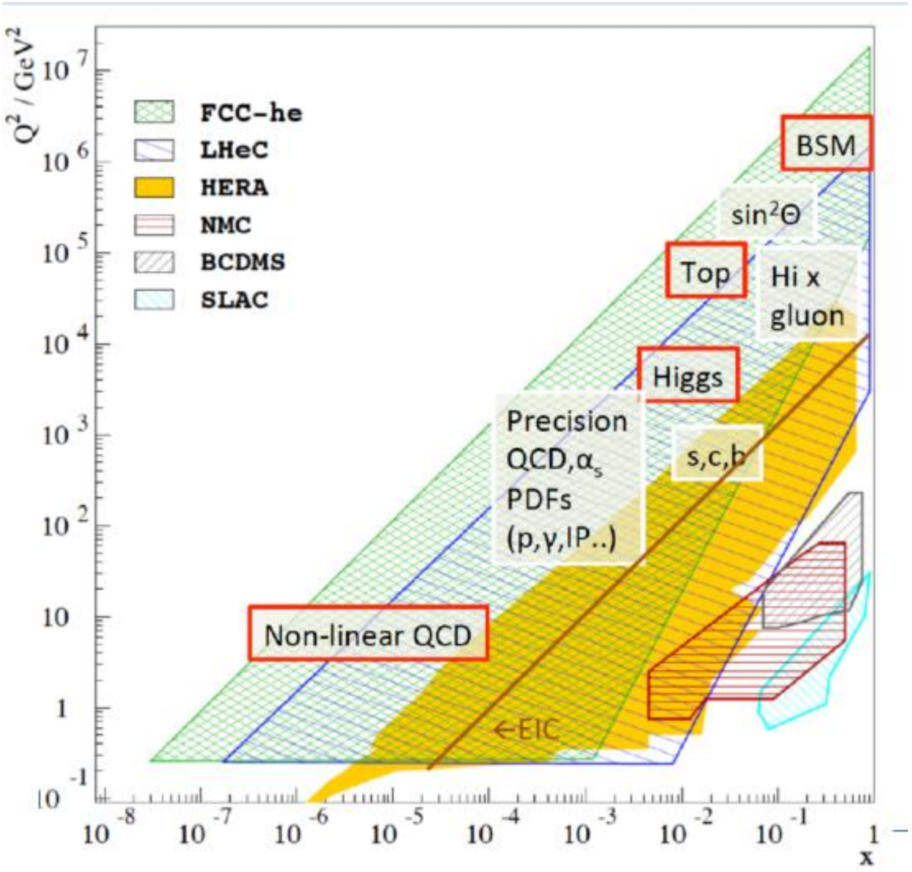
The Electron Ion Collider at Brookhaven

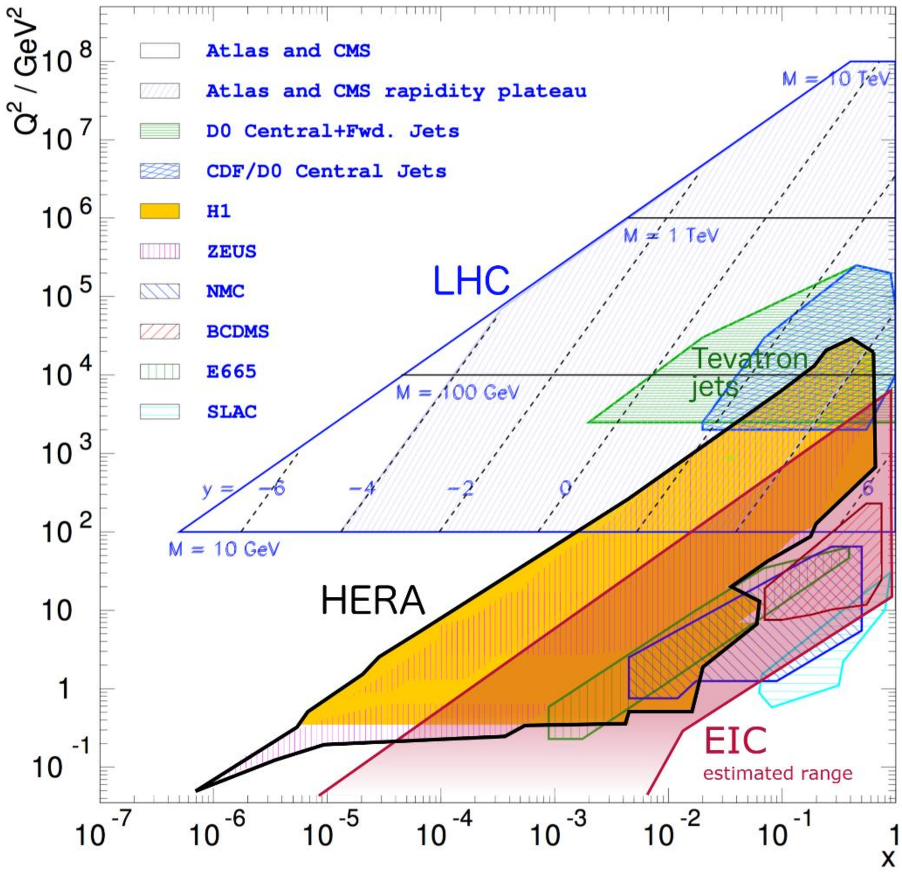
AND this one WILL definitely happen !

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.

A new DIS machine would instead give a consistent data set across a wide x, Q^2 range

Consider the kinematic reach of each of these

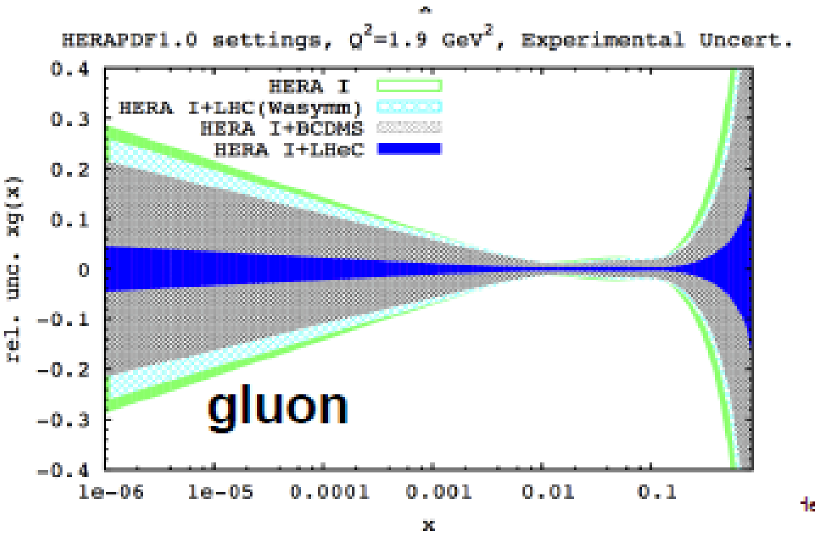




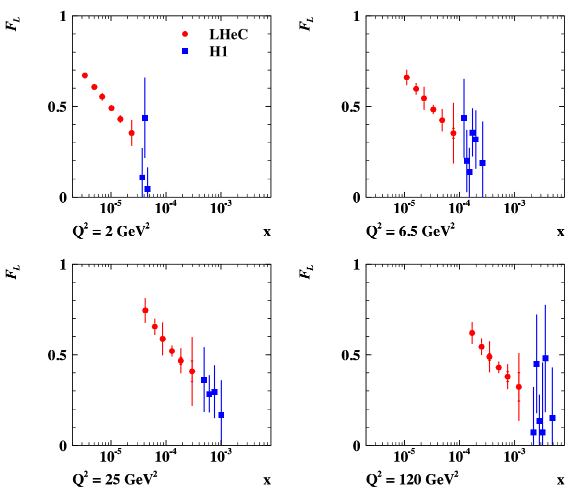
The proposed LHeC and FCC-eh machines reach lower x than HERA could reach

The EIC will reach higher x than HERA could reach

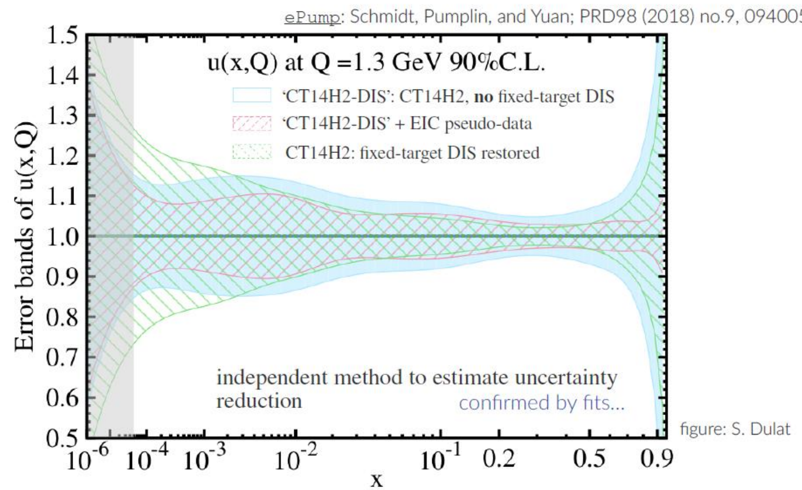
The LHeC would extend sensitivity to gluon and sea at low x



- FL measurement will improve
- Explore low-x QCD DGLAP vs BFKL or non-linear evolution



Consider the EIC This time the big impact is at high x



- **inclusive EIC may surpass total impact of fixed-target DIS in modern fits**

And this is only to mention the impact on Parton Distribution Functions which are afunction of the fractional longitudinal momentum of the parton.

The EIC is about much more, transverse momentum distributions, spin, generalized parton distribution functions etc

Summary

What have we learnt about/from the deep structure of the proton in the last 40 years?

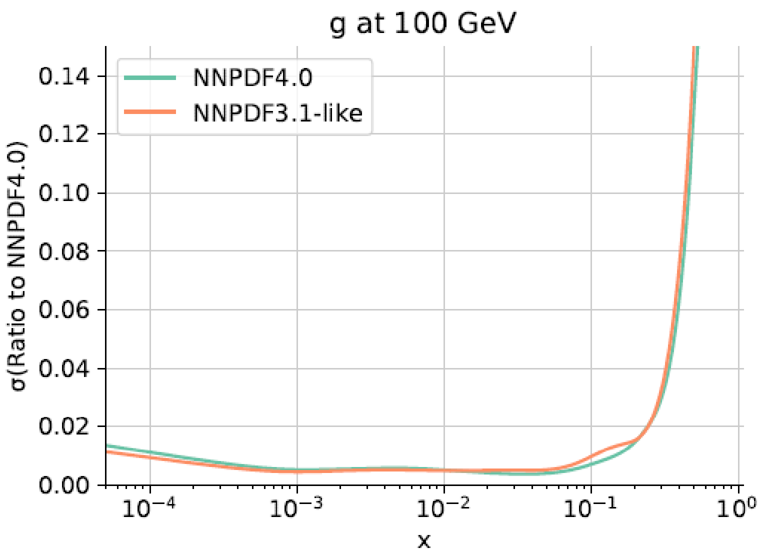
- It's full of partons- quarks, antiquarks and gluons
- The harder you look the more of them you see
- Established Quantum-Chromo-Dynamics as the theory of the strong interaction
- Measurement of its essential parameters: Parton Distribution Functions, the strong coupling parameter, $\alpha_s(M_Z)$ and the running of α_s with scale
- Sets the background for discovery physics at the LHC

But the PDF uncertainties need to be reduced

- Precision PDFs are needed reduce the background in searches for BSM physics- both at the LHC and any FCC-hh
- They are also needed for precision measurements of SM parameters, where small deviations from SM values may indicate BSM physics
- The measurements from the High Luminosity –LHC should improve on our current knowledge
- But a dedicated Deep Inelastic Scattering machine such as an LHeC/FCCeh or EIC could do better --- and EIC will definitely happen!

- And I have only spoken about the longitudinal momemtum distributions of the partons, the EIC will also probe transverse momentum and spin distributions
- And we should learn more about QCD beyond the DGLAP formalism– gluon saturation, BFKL resummation.

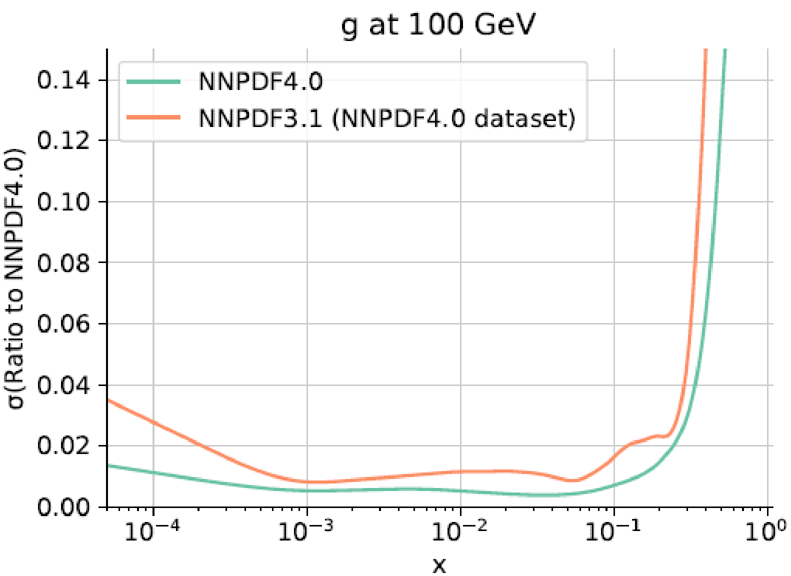
extras



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**

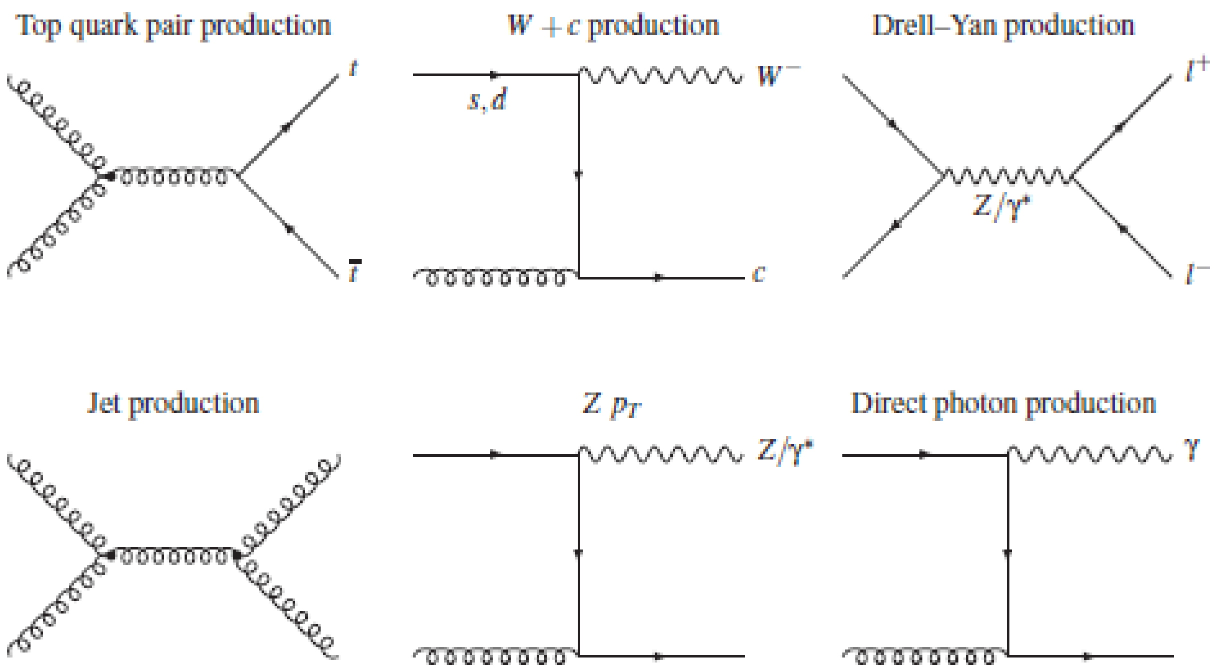
The top plot compares the uncertainties of NNPDF4.0 and 3.1 data sets using the **SAME new methodology**



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 has been a lot of debate in the PDF community over the new methodology.
But if we just accept it this still does not help much if one is trying to combine with other PDFs MSHT20 and CT18 with different central values

SO how can we improve PDFs?-A study of potential improvements has been made using processes 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

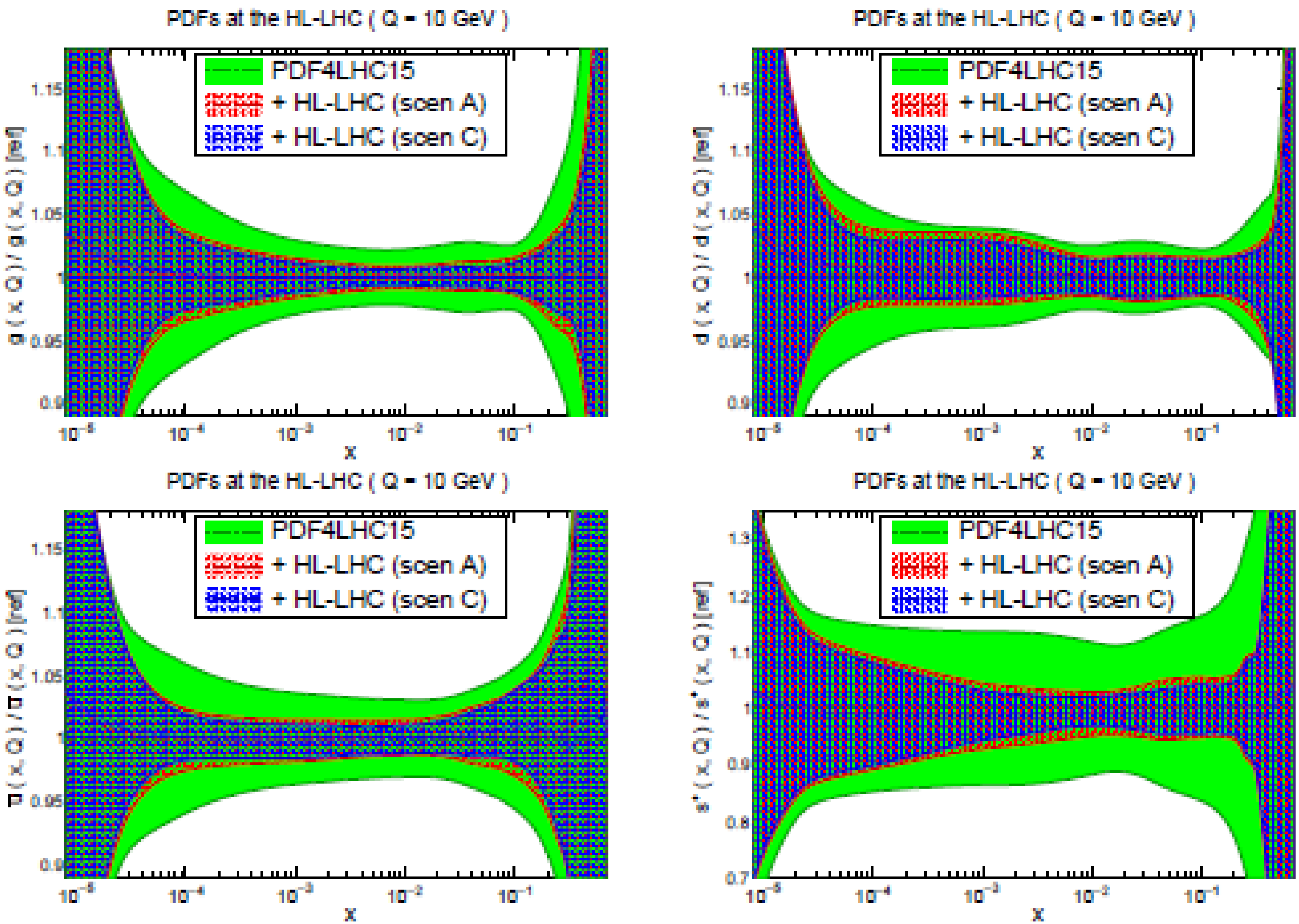
Pessimistic and Optimistic assumptions are made about systematic uncertainties based on experience with real data

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

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

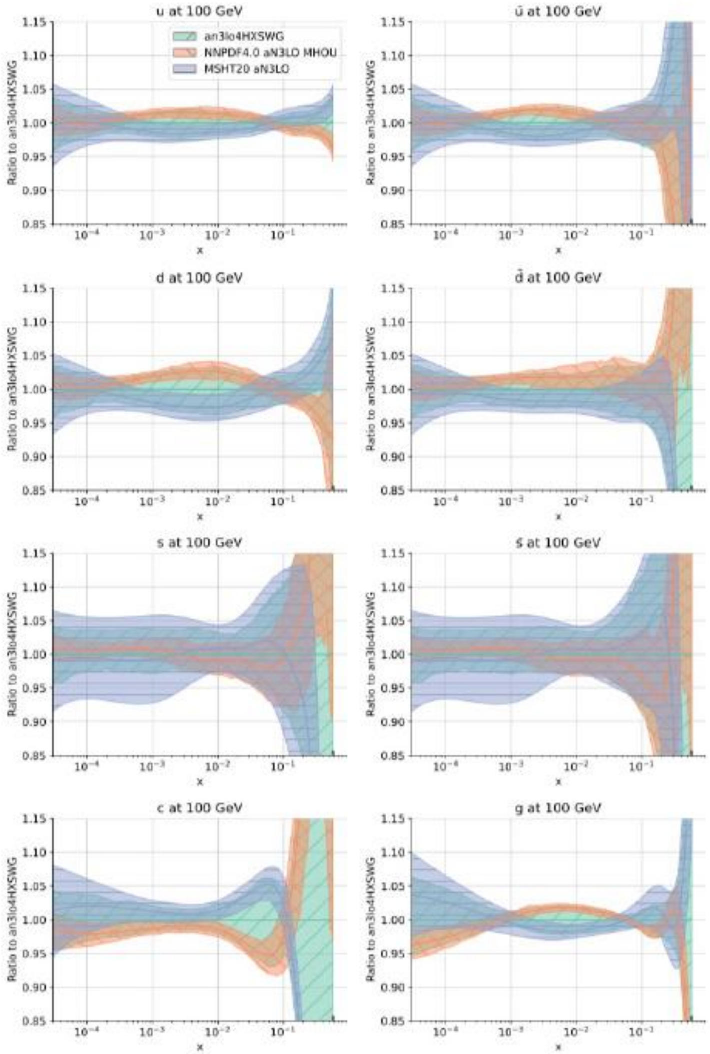
This is about as good as you can do with pseudo-data but let's not forget that this is a somewhat ideal situation

So we see potential improvements in the PDFs at the HL-LHC



Where scenario A is pessimistic and scenario C is optimistic

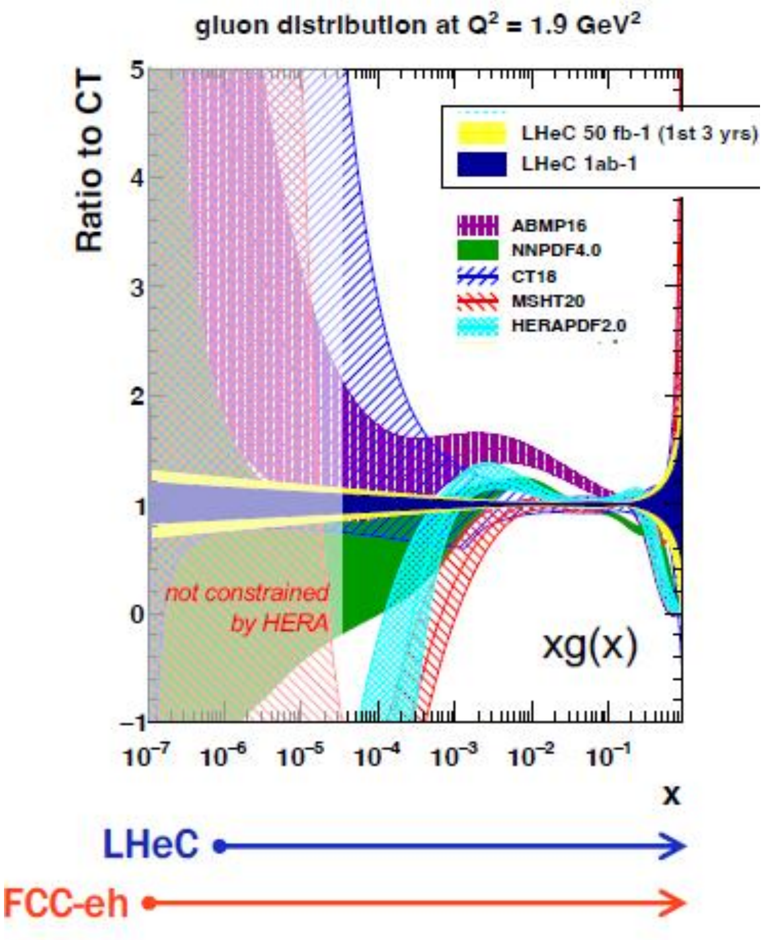
--Such improvements could give up to a factor 2 improvement in the PDF uncertainty on something like m_W



2411.05373

Paper provides a combination of the two aN3LO PDFs to produce aN3LOHXSWG; note this is not a benchmarked study, just a combination of the two aN3LO PDFs generating the same number of MC replicas for each.

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

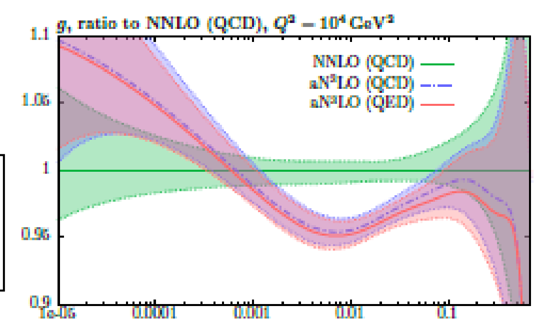
for ultra high energy neutrino cross sections

Even if your specific interest is not in low-x physics do not be complacent in thinking that this region does not affect you…

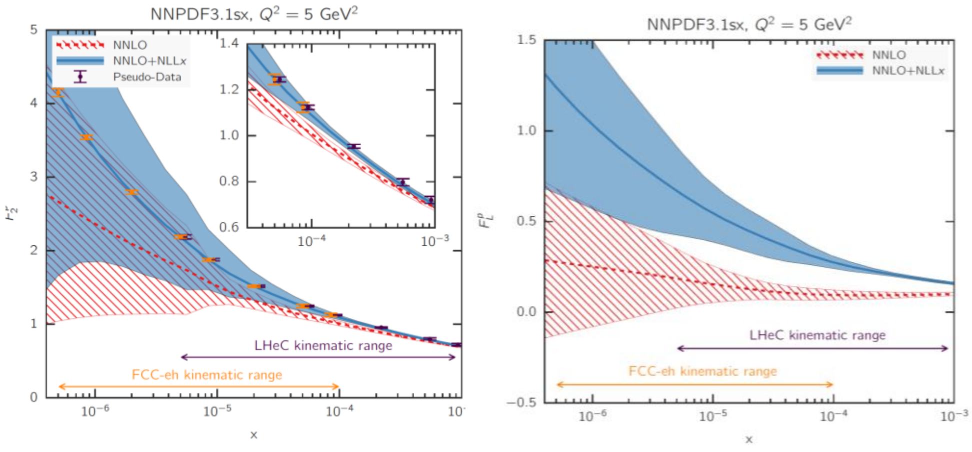
PDFs are going to N3LO –

where the first of the BFKL (ln(1/x) resummation) terms matter..

This has a significant effect on the low-x gluon



Effect of small x resummation on predictions for DIS F2 and FL



Prospects for FL measurement at LHeC