

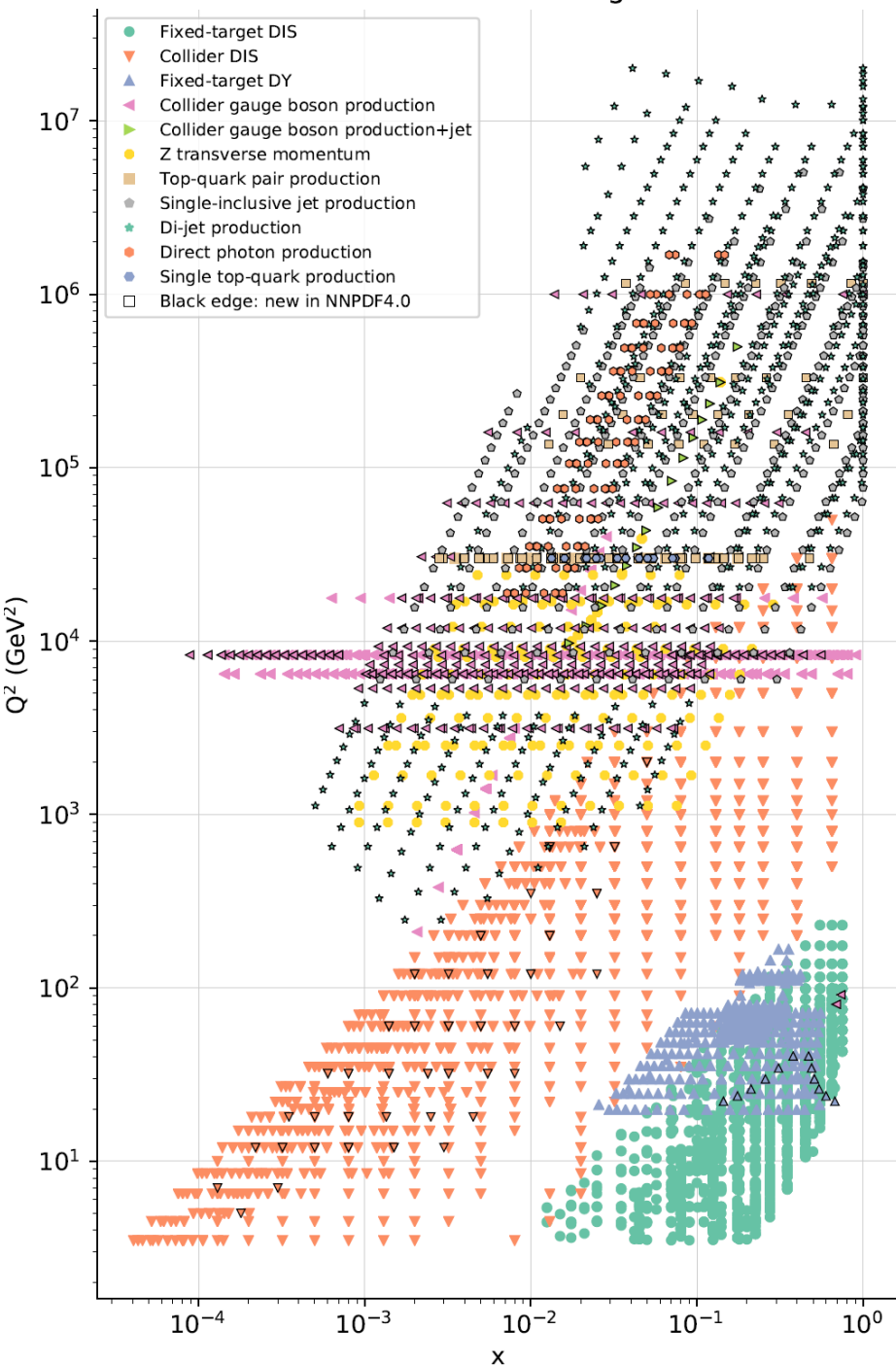
QCD@LHC2026
PDFs an Experimental Perspective
Amanda Cooper-Sarkar
Sep28th 2026

We=experimentalists, You= theoreticians

What can we do for you?

What can you do for us?

Kinematic coverage



1. Well we provide you with data

The HERA data are the ‘backbone of all PDF fits’

What other data have we used?

- Older fixed target DIS (E906 not old)
- Drell-Yan data from fixed target 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

2. **We make our own fits** What is the use of that?

i) **We produced the HERAPDF** based on a single ‘as consistent as possible’ data set across a broad range of the x , Q^2 plane.
And free from new physics effects.

But it is not enough, what could HERA not do?

High- x gluon and sea flavour detail s, c

And I don't think that adding just a single LHC data set to the HERA data is enough either, it doesn't tie down the PDF sufficiently

ii) **We** our interest in doing fits has led to our working very hard on our uncertainties- and you do need to know about that..

Particularly on correlated experimental uncertainties

Both to what extent are they actually correlated—suggested decorrelation scenarios within some of our data sets—jets, t -bar --- particularly where estimates of the systematic uncertainties are essentially a 2-point systematic

AND

To what extent are they correlated between data sets.

And this does matter?

ATLASpdf21 (2112.11266) is a fit to many different sorts of ATLAS data:

- W,Z boson production
- W,Z boson +jets
- T-tbar data
- Direct photon production
- Inclusive jet production

These are the processes which are known to have PDF sensitivity

It is a fit to NNLO in QCD and NLO in EW achieved either by direct NNLO grids or by NLO grids + k-factor corrections

The HERA data provided the backbone of the fit (as with ALL modern PDF fits) and provide good constraints from $\sim 10^{-4} < x < 0.1-0.4$ (depending on which PDF) and the ATLAS data provide

- Flavour separation in the sea W,Z, V+jets
- High-x gluon (Jets, t-tbar, direct- γ)
- in addition to further constraints on all PDFs $x > 0.001$

Data set	$\sqrt{s}$ [TeV]	Luminosity [fb^{-1}]	Decay channel	Observables entering the fit
Inclusive W, Z/ γ^* [9]	7	4.6	e, μ combined	η_l (W), y_Z (Z)
Inclusive Z/ γ^* [13]	8	20.2	e, μ combined	$\cos \theta$ in bins of $y_{\ell\ell}, M_{\ell\ell}$
Inclusive W [12]	8	20.2	μ	η_μ
$W^\pm$ + jets [23]	8	20.2	e	p_T^W
Z + jets [24]	8	20.2	e	p_T^{jets} in bins of $ y_{\text{jets}} $
$t\bar{t}$ [25, 26]	8	20.2	lepton + jets, dilepton	$m_{t\bar{t}}, p_T^t, y_{t\bar{t}}$
$t\bar{t}$ [15]	13	36	lepton + jets	$m_{t\bar{t}}, p_T^t, y_t, y_{t\bar{t}}$
Inclusive isolated γ [14]	8, 13	20.2, 3.2	-	E_T^γ in bins of η^γ
Inclusive jets [16–18]	7, 8, 13	4.5, 20.2, 3.2	-	p_T in bins of $ y_{\text{jets}} $

ATLASPDF USP:

- All data sets included have full information on correlated systematic uncertainties
- This allows us to consider **experimental correlations between data sets** as well as within data sets. This cannot be done by the global PDF fits unless we provide the information– but it matters if we are seeking an ultimate O(1%) accuracy on PDFs. ATLAS now makes this information public

Correlation of systematic sources between data sets

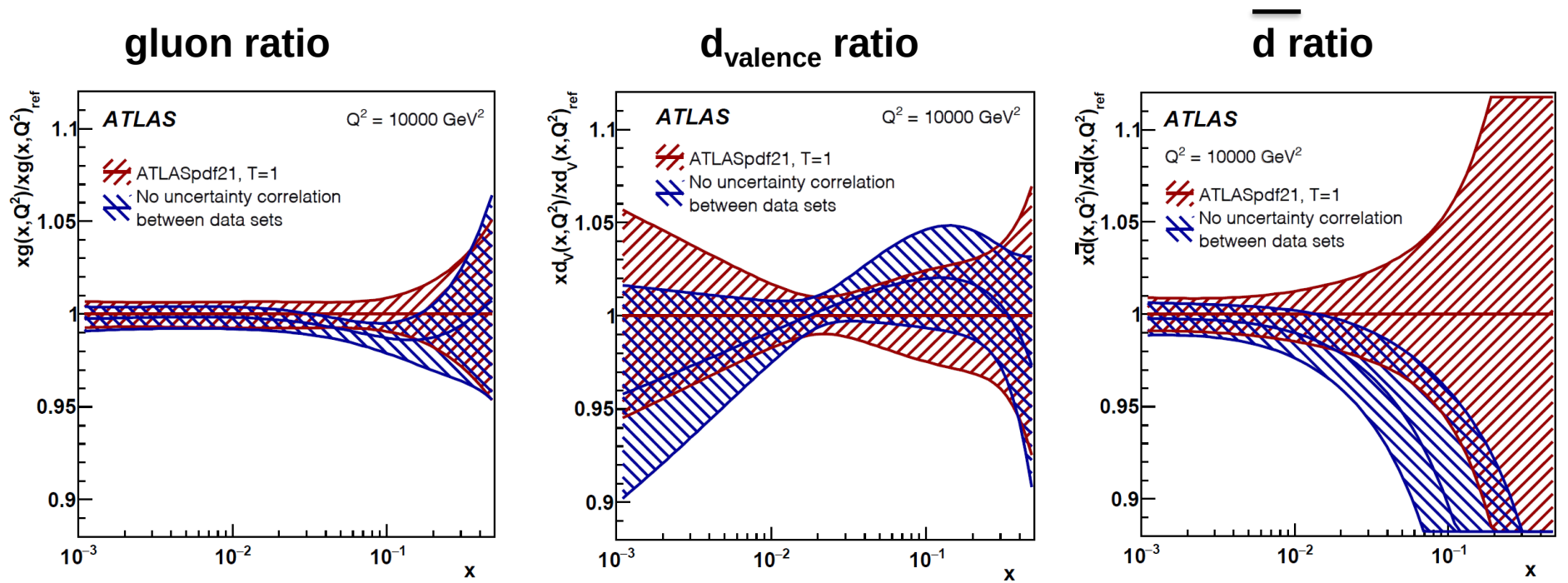
Note these are mostly the JES systematics, which are the largest -- lepton systematics are much smaller

Systematics	8 TeV W + jets	8 TeV Z + jets	8 TeV $t\bar{t}$ lepton + jets	13 TeV $t\bar{t}$ lepton + jets	8 TeV inclusive jets
Jet flavour response	JetScaleFlav2	Flavor Response	flavres-jes	JET29NP JET Flavour Response	syst JES Flavour Response*
Jet flavour composition	JetScaleFlav1Known	Flavor Comp	flavcomp-jes	JET29NP JET Flavour Composition	syst JES Flavour Comp
Jet punchthrough	JetScalepunchT	Punch Through	punch-jes	-	syst JES PunchThrough MC15
Jet scale	JetScalePileup2	PU OffsetMu	pileoffmu-jes	-	sys JES Pileup MuOffset
	-	PU Rho	pileoffrho-jes	JET29NP JET Pileup RhoTopology	sys JES Pileup Rho topology*
	JetScalePileup1	PU OffsetNPV	pileoffnpv-jes	JET29NP JET Pileup OffsetNPV	syst JES Pileup NPVOffset
	-	PU PtTerm	pileoffpt-jes	JET29NP JET Pileup PtTerm	syst JES Pileup Pt term
Jet JVF selection	JetJVFCut	JVF	jetvxfrac	-	syst JES Zjets JVF
B-tagged jet scale	-	btag-jes	JET29NP JET BJES Response	-	-
Jet resolution	-	jeten-res	JET JER SINGLE NP	-	-
Muon scale	-	-	mup-scale	MUON SCALE	-
Muon resolution	-	-	muonms-res	MUON MS	-
Muon identification	-	-	muid-res	MUON ID	-
Diboson cross-section	-	-	dibos-xsec	Diboson xsec	-
Z + jets cross section	-	-	zjet-xsec	Zjets xsec	-
Single- t cross section	-	-	singletop-xsec	st xsec	-

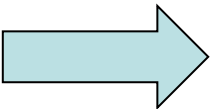
BUT NOTE that for some data sets decorrelations within the data set are needed
ATLAS 8 TeV inclusive jets, arXIV:1706.03912
ATLAS 8 TeV t - t bar spectra, arXIV: 2203.13923, ATL-PHYS-PUB-2018-027

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



The χ^2 of the fit is 30 units better when correlations are included
The difference in PDFs is small for the gluon
But can be large in the d-quark sector
Remember the goal for PDF precision is $\sim 1\%$ 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

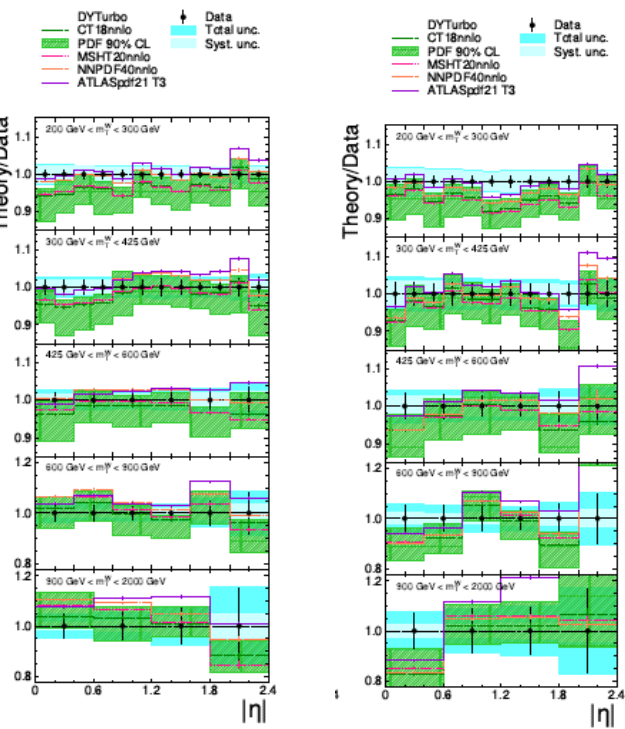


Correlations can be important

NEW ATLAS data is coming your way—some of it is already here...

W off shell –
arxiv:2502.21088

Electron and muon data
combined
eta spectra in m_t^W bins.

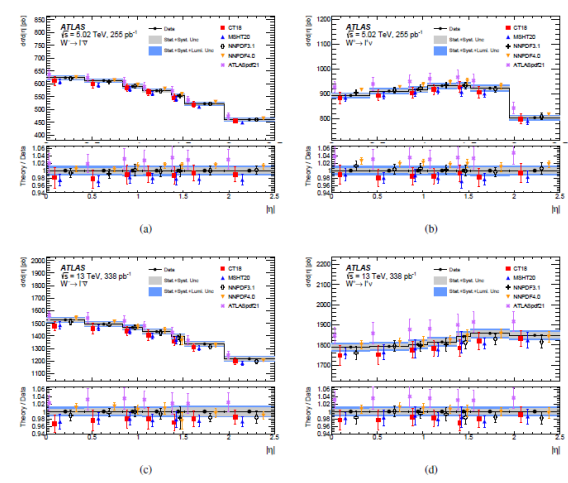


Lepton+
These data affect the d-type
sector.

Lepton –
These data affect the d-type
sector.

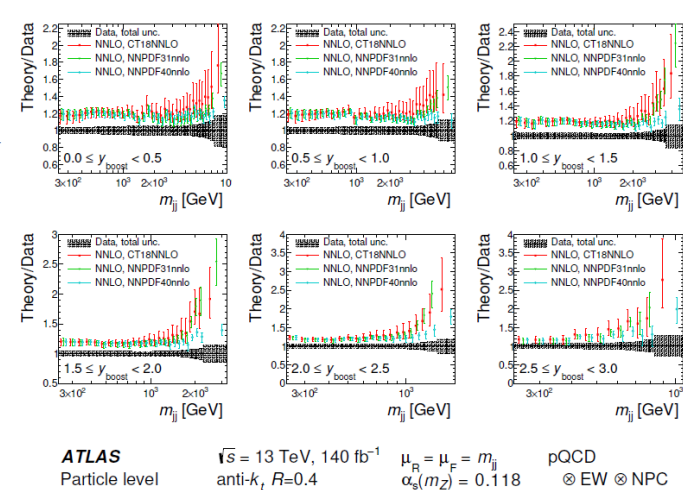
Low pile-up data
At 5 and 13 TeV --
arxiv:2502.09403

Eta distributions in p_T bins for
W+ and W- for both energies



These data affect the d-type
sector.

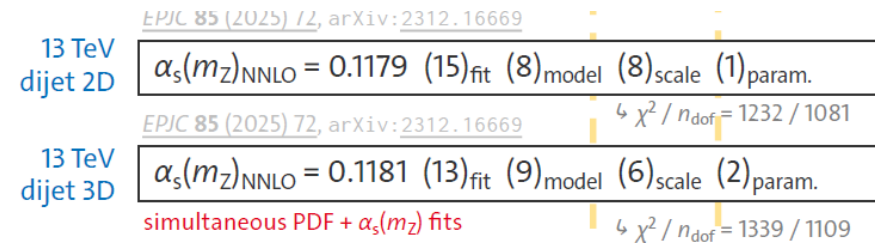
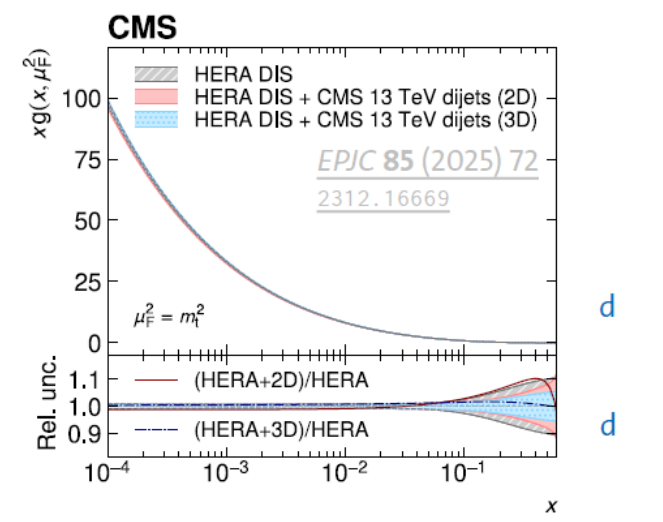
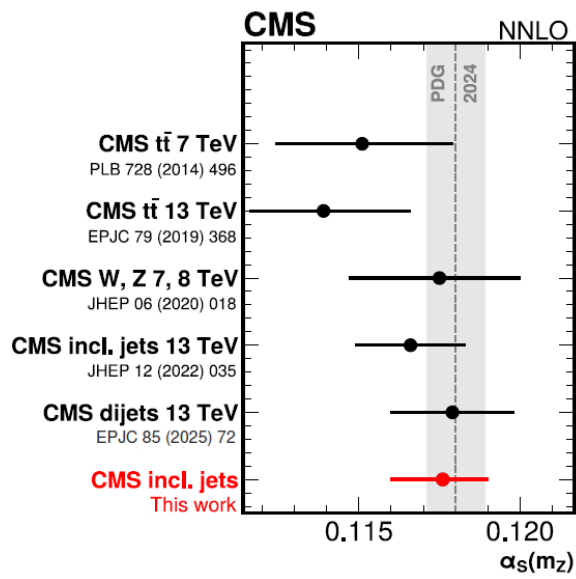
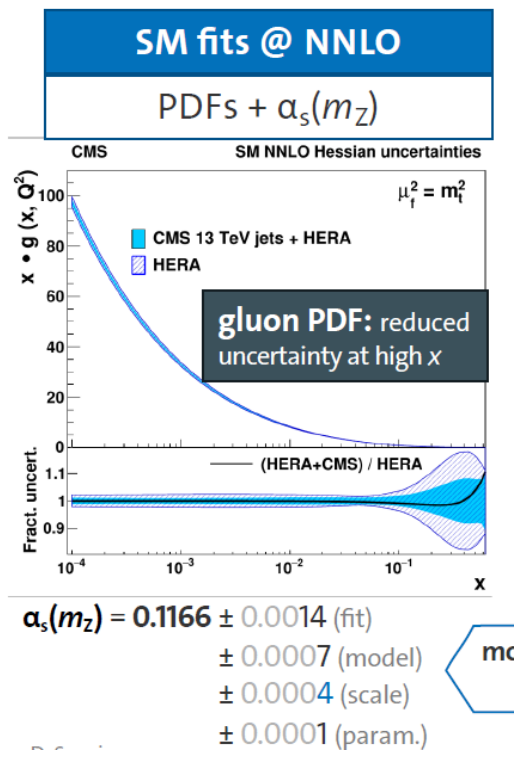
The new 13 TeV
dijet data– all Run2
140 pb-1
Arxiv:2512.19073 --
 m_{jj} distributions in
 y_{boost} bins



ATLAS
Particle level
 $\sqrt{s} = 13 \text{ TeV}, 140 \text{ fb}^{-1}$
anti- $k_t, R=0.4$
 $\mu_R = \mu_F = m_{jj}$
 $\alpha_s(m_Z) = 0.118$
pQCD
EW @ NPC

Obvious PDF sensitivity
These data affect the gluon
most strongly.

CMS usually present analyses with HERA+ one data set, (or class of data sets)
 Inclusive jets and t-tbar have been considered together
 But jets and dijets have been treated separately.
 Correlations between either statistical or systematic uncertainties are not provided



From 13 TeV 33.5 fb⁻¹
 Inclusive jets

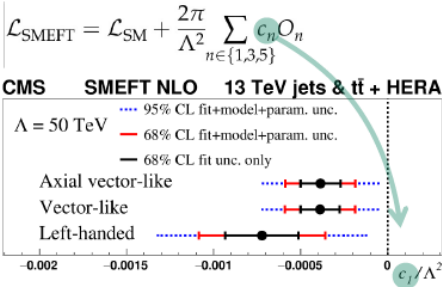
From 13 TeV all CMS
 Inclusive jets and t-tbar

From 13 TeV CMS 29.6 fb⁻¹
 Dijets

Also CMS Jets, t-tbar, W,Z have been used for joint $\alpha_s(M_W)$, m_t^{pole} , $\sin^2\Theta_W$ and even SMEFT coefficients (arxiv:2607.22390 **note not a CMS paper as such**)
 most ambitious study of all, but no correlations between experimental data sets.
Proof of principle of simultaneous PDF+SM parameter fits, NOTE T=1

SMEFT constraints

constrain *Wilson coefficients* for SM extensions involving four-quark contact interactions



top quark mass in good agreement with previous CMS results using only tt data

$m_t^{\text{(pole)}} [\text{GeV}] = 170.4 \pm 0.6 \text{ (fit)}$

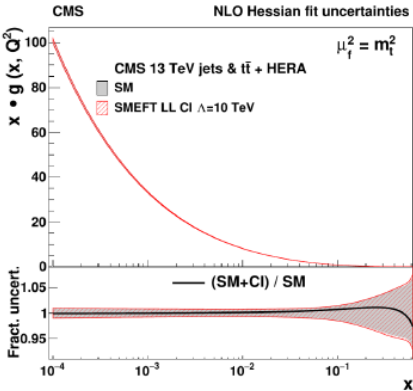
$\pm 0.1 \text{ (model)}$

$\pm 0.1 \text{ (scale)}$

precise NNLO result on jet cross sections $\pm 0.1 \text{ (param.)}$

combined fits with other data sets

PDFs + $\alpha_s(m_Z) + m_t^{\text{(pole)}}$



jet data “pins down” PDFs + $\alpha_s(m_Z)$

increased sensitivity to other parameters

$\alpha_s(m_Z) = 0.1188 \pm 0.0017 \text{ (fit)}$

$\pm 0.0004 \text{ (model)}$

$\pm 0.0025 \text{ (scale)}$

$\pm 0.0001 \text{ (param.)}$

Data set	Process / Observable	$\sqrt{s}$	Reference
<i>ep</i> DIS	$ep \rightarrow eX$: NC+CC inclusive cross sections (HERA I+II)	0.318 TeV	arXiv:1506.06042 [15]
<i>pp</i> jets	$pp \rightarrow \text{jet} + X$ $d\sigma/dp_T^{\text{jet}} d Y^{\text{jet}} $ ($R=0.7$)	2.76 TeV 7 TeV 8 TeV 13 TeV	arXiv:1512.06212 [16] arXiv:1410.6765 [17] arXiv:1609.05331 [18] arXiv:2111.10431 [13]
<i>pp</i> DY A_4	$pp \rightarrow Z/\gamma^* \rightarrow \ell^+ \ell^-$: A_4	13 TeV	arXiv:2408.07622 [2]
<i>pp</i> $W^\pm$ asymmetry	$pp \rightarrow W^\pm \rightarrow \ell^\pm \nu$: $d\sigma/d \eta_\ell $	13 TeV	arXiv:2008.04174 [19]
<i>pp</i> $t\bar{t}$ (ℓ +jet)	$pp \rightarrow t\bar{t}$ (ℓ +jet) $d\sigma/dm_{t\bar{t}}$	13 TeV	arXiv:2108.02803 [20]
<i>pp</i> $t\bar{t}$ (2ℓ)	$pp \rightarrow t\bar{t}$ (2ℓ) $d\sigma/dm_{t\bar{t}}$	13 TeV	arXiv:2402.08486 [21]

Table 1: Summary of the experimental measurements and their associated theoretical predictions.

The values of the SM parameters, obtained in our fit with the breakdown of uncertainties are summarized in Table 3.

Parameter	Central	Exp.	Model	QCD scale	NRQCD scale	Total
m_t^{pole} (GeV)	172.59	+0.35 −0.34	+0.04 −0.04	+0.35 −0.34	+0.12 −0.08	$172.59^{+0.51}_{-0.49}$
$\alpha_s(m_Z)$	0.1179	+0.0007 −0.0007	+0.00025 −0.00085	+0.0029 −0.0013	+0.0001 −0.0001	$0.1179^{+0.0030}_{-0.0016}$
$\sin^2 \theta_{\text{eff}}^\ell$	0.23142	+0.00023 −0.00023	+0.00022 −0.00004	+0.00008 −0.00018	> 0.00000	$0.23142^{+0.00032}_{-0.00029}$

	m_t^{pole} [GeV]	$\alpha_s(m_Z)$	$\sin^2 \theta_{\text{eff}}^\ell$
FO NNLO	172.03	0.1178	0.23145
FO NNLO + NLO NRQCD	172.48	0.1179	0.23143
FO NNLO + NLO NRQCD + NLL + EFT	172.80	0.1175	0.23153
FO NNLO + NLO NRQCD + NLL	$172.59^{+0.51}_{-0.49}$	$0.1179^{+0.0030}_{-0.0016}$	$0.23142^{+0.00032}_{-0.00029}$

Some consideration of how we judge goodness of fit.

PDF fits are not good by any strict statistical criterion..

We need certainty about correlations/decorrelations.

We often use nuisance parameters to formulate the χ^2 for systematic uncertainties rather than putting them into the covariance matrix.

$$\chi^2 = \sum_{i,k} \left(D_i - T_i \left(1 - \sum_j \gamma_{ij} b_j \right) \right) C_{\text{stat,uncor},ik}^{-1}(D_i, D_k) \left(D_k - T_k \left(1 - \sum_j \gamma_{kj} b_j \right) \right) + \sum_j b_j^2$$

BUT if the number of sources of systematic is larger than the number of data points, this becomes very inefficient

$$+ \sum_j b_j^2$$

This is the xFitter form of the χ^2 .

- The first term is the partial chisq of each data set taking into account correlations between systematic uncertainties using nuisance parameters b_j .
- The second term is the penalty term for nuisance parameters. This sum over the squares of the nuisance parameters, called the correlated term, tells us how far each systematic source has shifted from its notional value— this should not be more than $\sim 1\sigma$

The quality of a fit to individual data sets can only be judged by its contribution to both these terms

If you put the systematic uncertainties into the covariance matrix then you lose the information on the shift of each source. If you then regularise this matrix then you are messing about with systematics without knowing what you are actually doing and whether it is reasonable physically.

What can you do for us.. I will expand on this in the next few slides

Give us predictions

- hopefully with technology to make grids— a plea for two sets of predictions: $W+c$ or $W+D/D^*$ and direct photon
- NOT just NNLO at full colour—one day also at N3LO?
- Serious consideration of $\ln(pt)$ resummation for Z_{pt} spectra
- *EW corrections*
- *Non-perturbative corrections*

Depending on the data set to be fitted

Please think about grid accuracy, or give a way to estimate true uncertainty on the grids

Give us global PDFs

But also **Help to RESOLVE differences in PDFs,**

Questions of need for tolerance

Difference in FFPDF and NNPDF procedure

Closure tests, inconsistencies,

$T=1$ won't do either for profiling or for fitting,

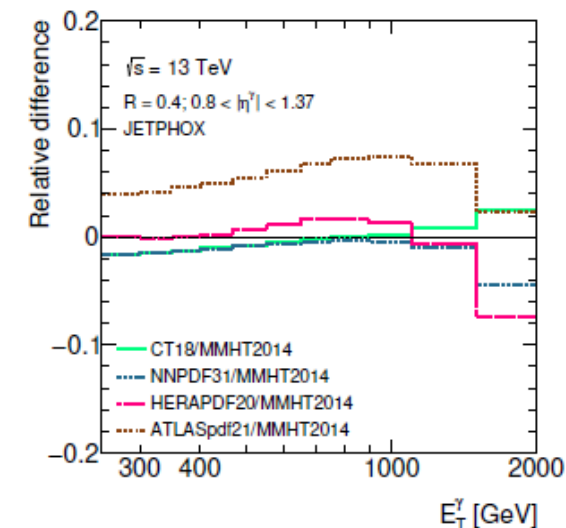
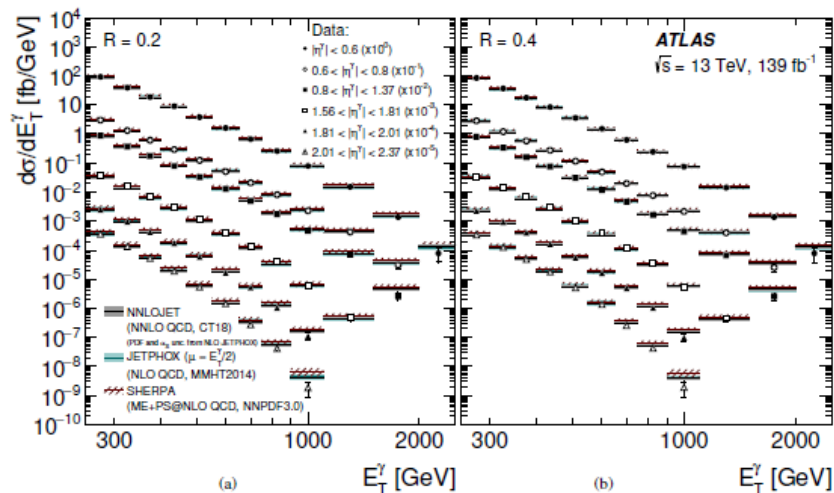
'errors on errors' approaches are interesting -- also not enough?

The 13 TeV direct photon data

/ arXIV:2302.00510

E_T^γ in 6 bins of η^γ

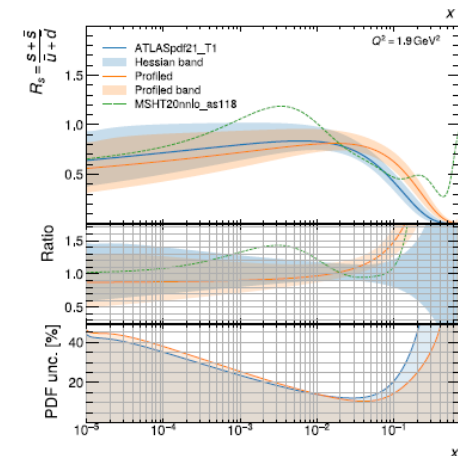
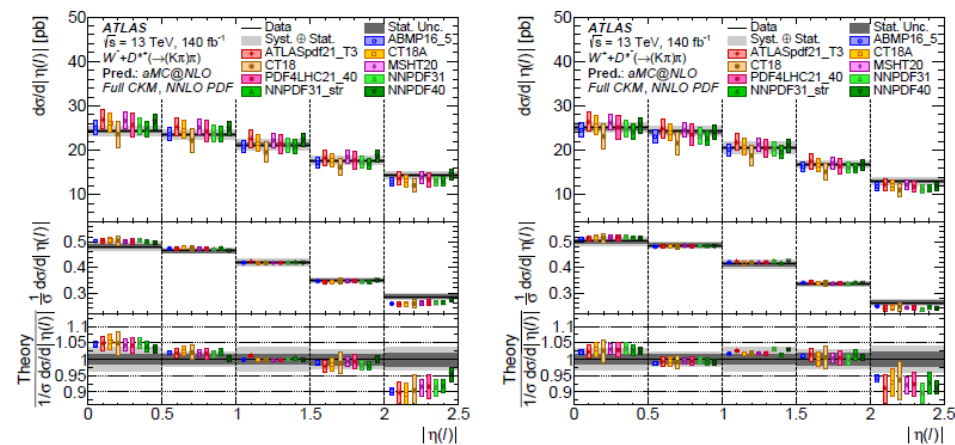
We need new grid predictions- since the bins have changed since the earlier data



Differences in the PDFs at high-x explain this

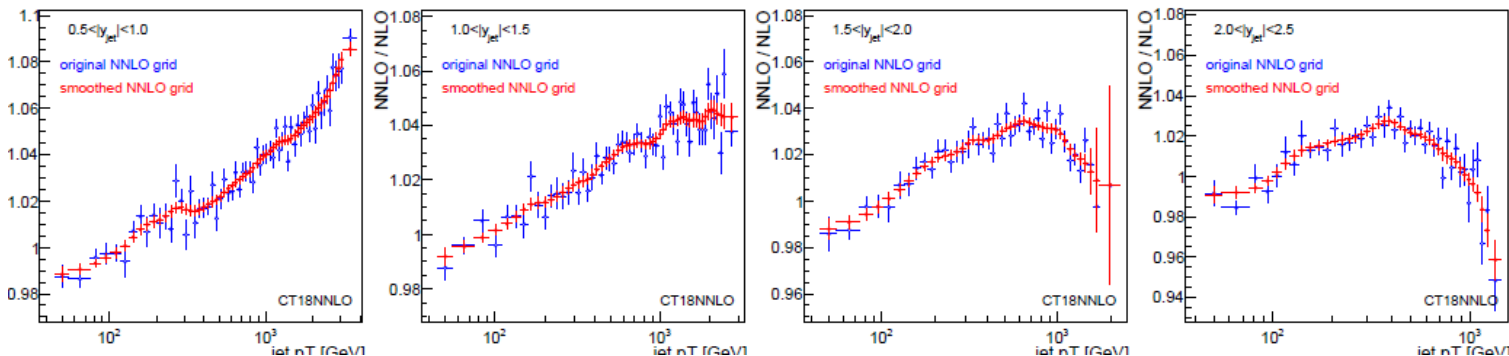
$W^\pm + D/D^*$ arXiv:2302.00336

p_T^D and $\eta(\text{lepton})$ spectra



Profiling ATLASpdf21 using NNLO predictions from Generet et al (2510.24525) suggest some impact at high-x **BUT we have no grid**

Blue is the
grid prediction
Red is a local
smoothing

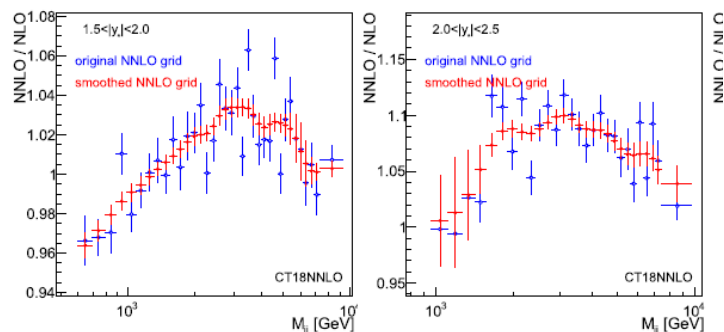
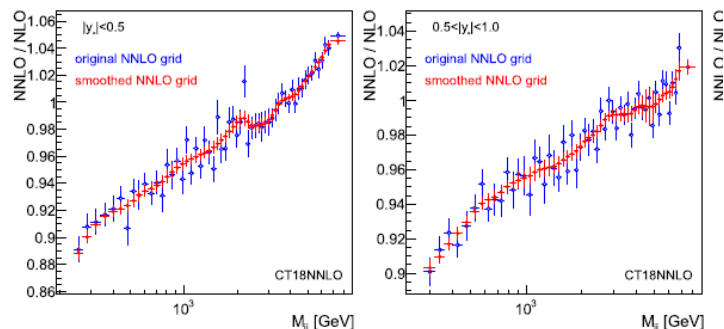


Inclusive jets

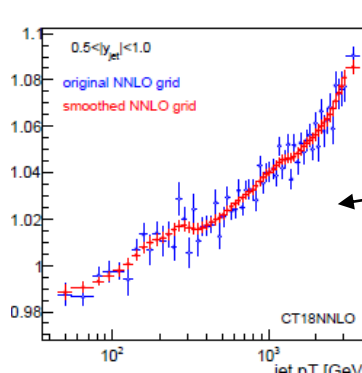
Predictions from the full colour grids are not very accurate---

~ 0.4% stat error for inclusive and
>~ 1% stat error for dijets

BUT are these meaningful statistical uncertainties?

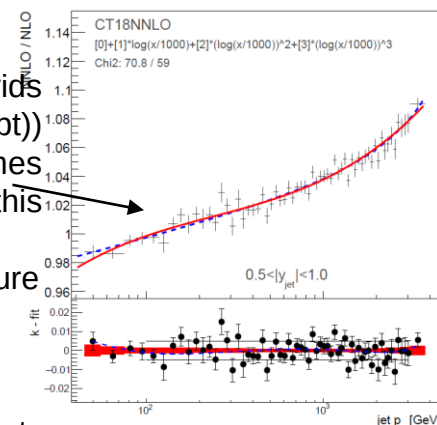


dijets



We fit the grids
(polynomial in $\ln(pt)$)
so that this becomes
this

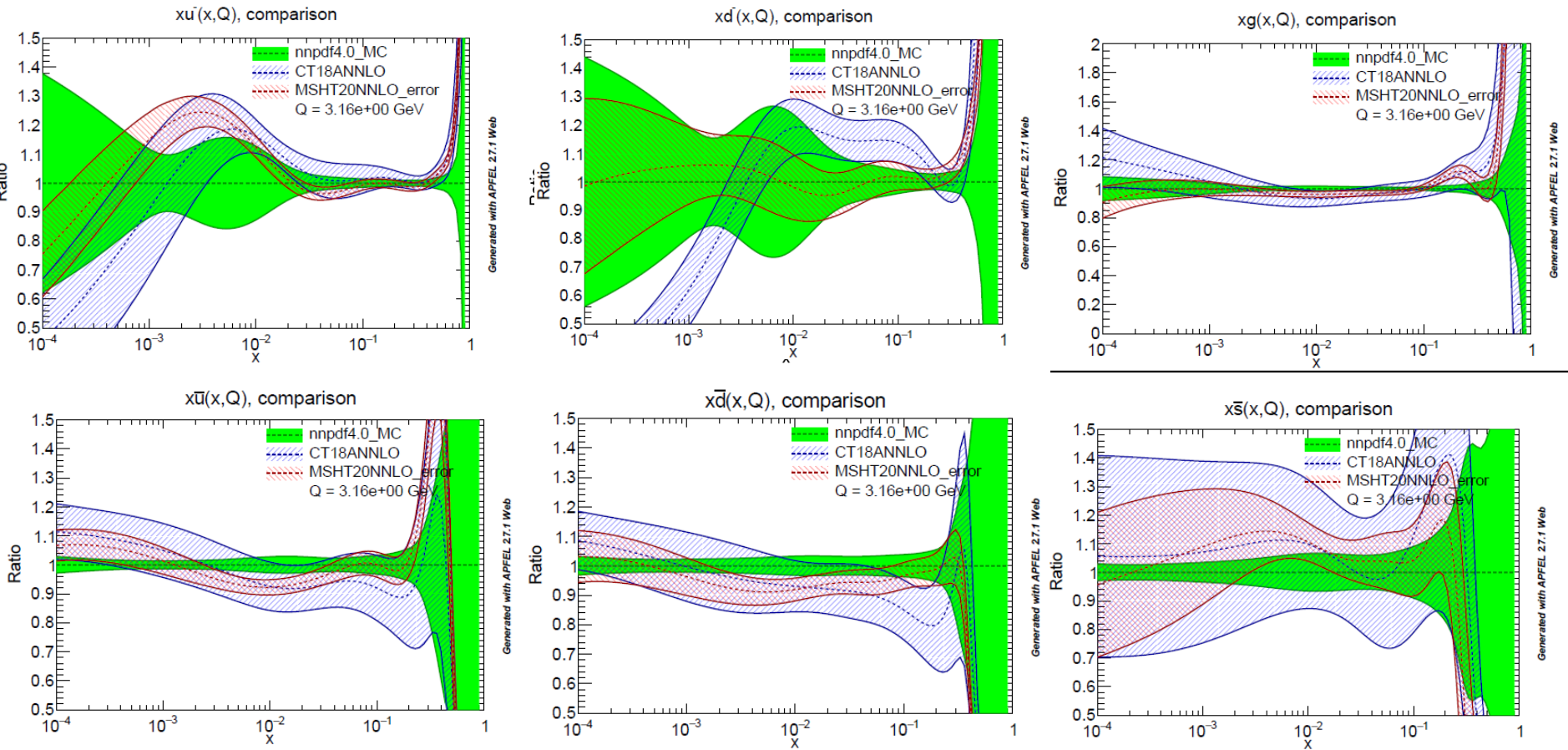
But what is a
reasonable measure
of uncertainty?



CT spend a lot of
time worrying about
this

[arxiv:2412.00350](https://arxiv.org/abs/2412.00350)

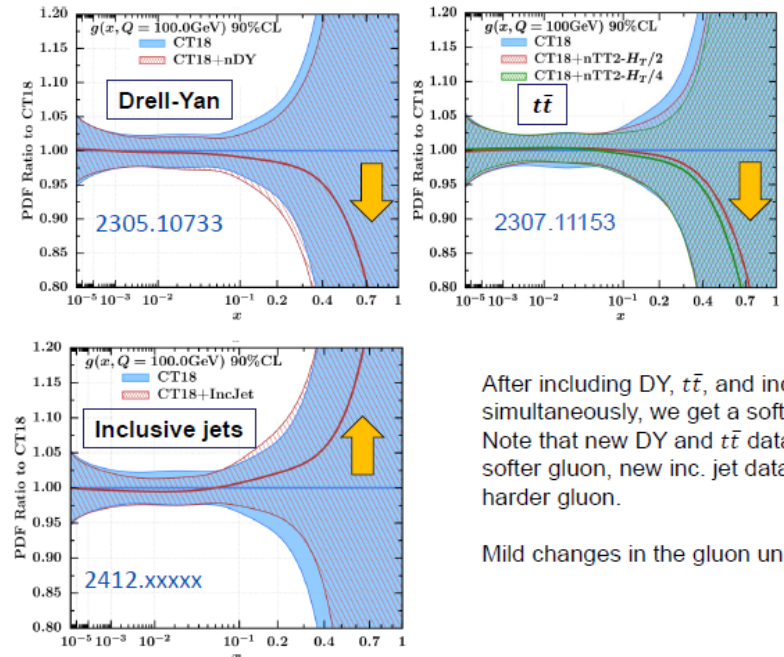
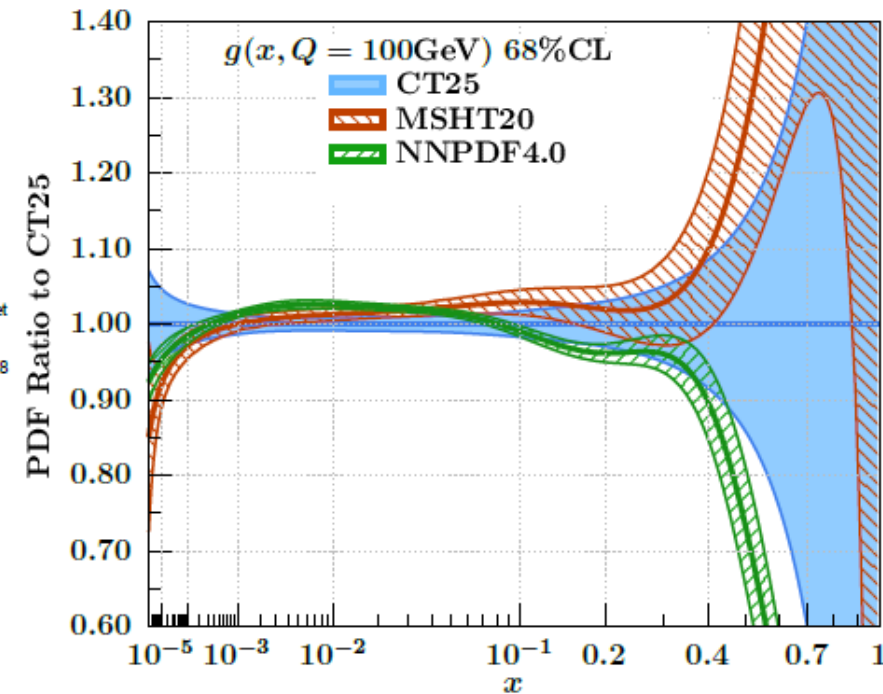
Differences in modern PDF sets –CT25 on next page---it is not way different from CT18A



Differences are more obvious in ratio
They are large at small- x and at high- x – as might be expected

Experimentalists are always asking 'who is right' or 'what shall I use'
Well the PDF4LHC21 is still a good compromise– but it is getting old
YES there will be a new one but these things take time

CT25 is nearly with us, you can get from the CT website, see a preview
[arxiv:2512.19779](https://arxiv.org/abs/2512.19779)



After including DY, $t\bar{t}$, and inc. jet data simultaneously, we get a softer gluon. Note that new DY and $t\bar{t}$ data favor a softer gluon, new inc. jet data prefer a harder gluon.

Mild changes in the gluon uncertainty

There is newer data from Drell-Yan, t - $\bar{t}$, jets (and fixed target SeaQuest on $d\bar{u}$)
 This results in a slightly softer gluon and mild decreases in gluon uncertainty

To note:

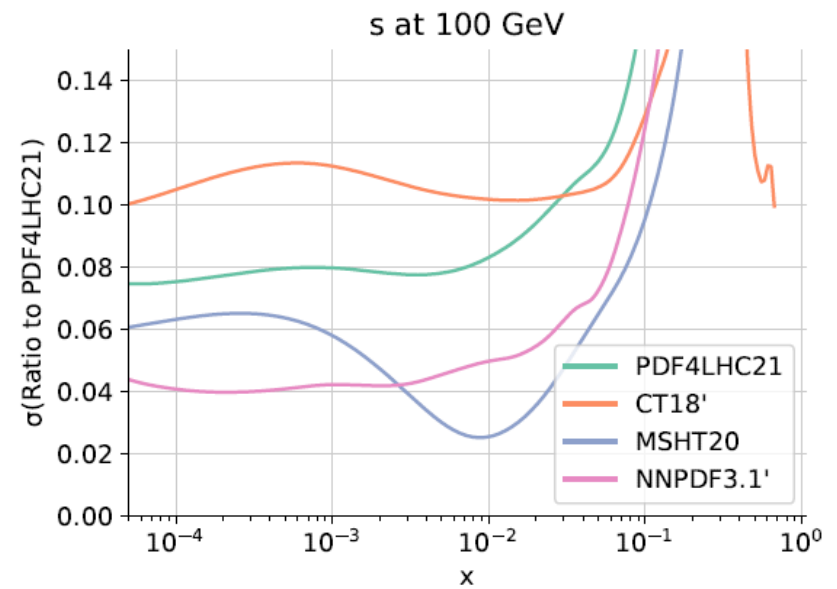
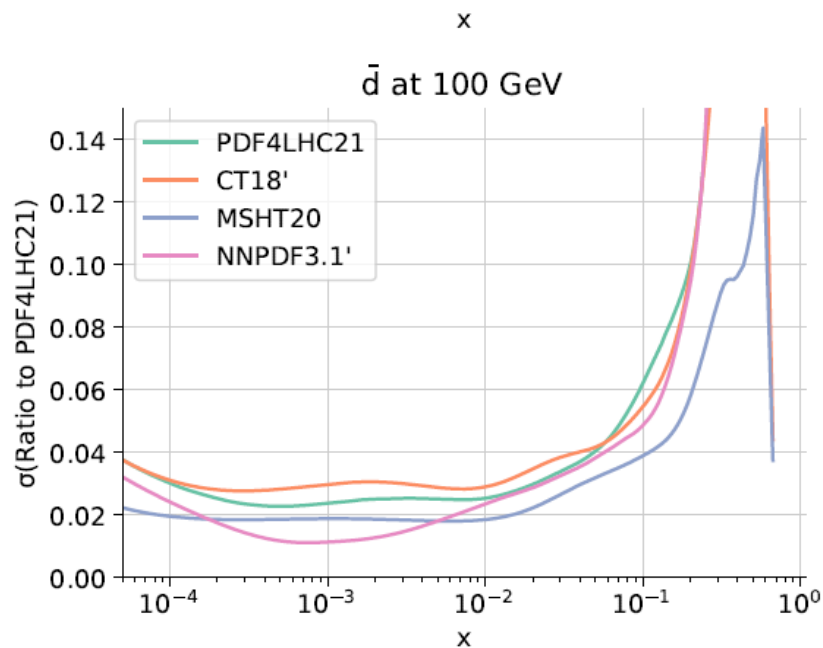
- For jets they use inclusive jets because these are more stable under scale change than dijets. Both changes of choice, like HT or PT, and changes of factors of 2
- With new DY data: [ATLAS Z3D](#) [ATLAS W8](#), [CMS13Z](#), [LHCb13W,Z](#), [ATLAS5W,Z](#) the enhancement of strangeness at low- x first observed by ATLAS becomes stronger

CT25 will be based on CT18A

PDF4LHC21

PDF4LHC21 actually combines variants of CT18 and NNPDF3.1 with MSHT20. Sets heavy quark masses to a common value and has a slight update of input data sets for NNPDF3.1.

The combination is a statistical combination without correlations between the three input PDFs. Where the three input PDFs 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



These plots represent ratios of uncertainties

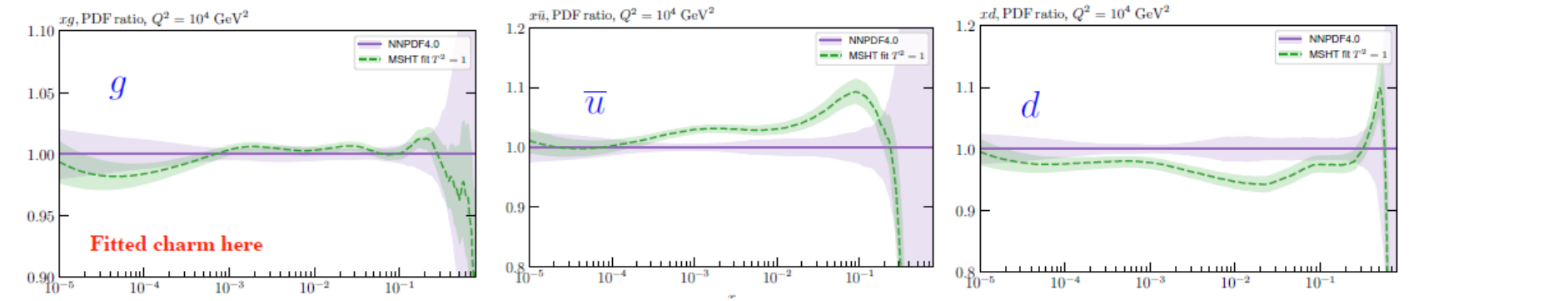
Efforts to understand differences in PDFs [arXiv:2407.07944](#)

Fit real NNPDF4.0 data set using NNPDF4.0 theory predictions but with MSHT20 parametrisation, the only difference is thus the MSHT methodology which is fixed parametrisation not neural net. MSHT also apply a tolerance

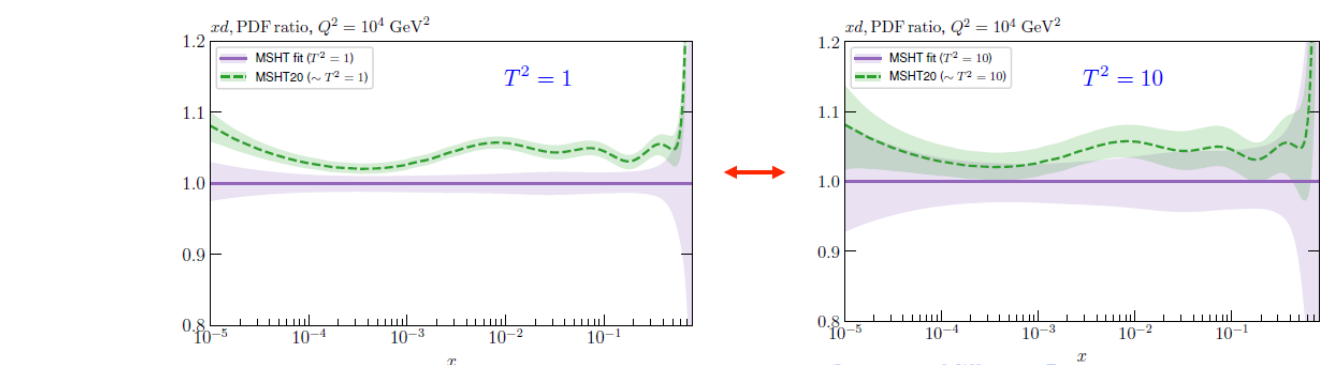
A very good fit is obtained, as good as NNPDF, for various choices;

	NNPDF4.0 pch	MSHT fit	MSHT fit (w positivity)
χ^2_{t0}	5928.3 (1.282)	5736.7 (1.240)	5837.8 (1.262)
$\Delta\chi^2_{t0}$		<u>-191.6 (-0.04)</u>	<u>-90.5 (-0.02)</u>

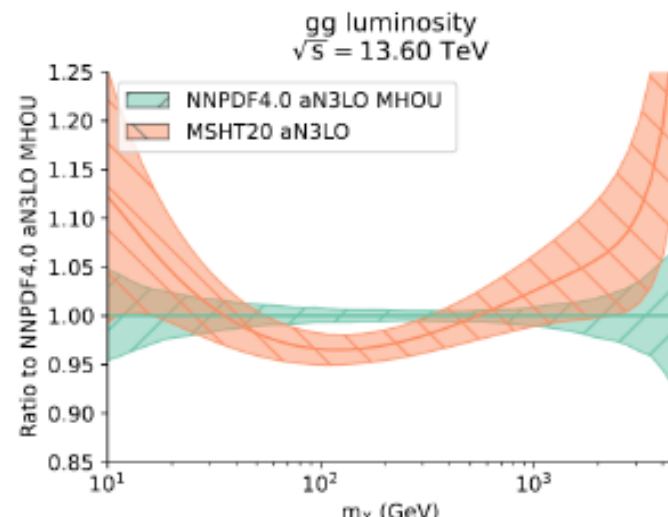
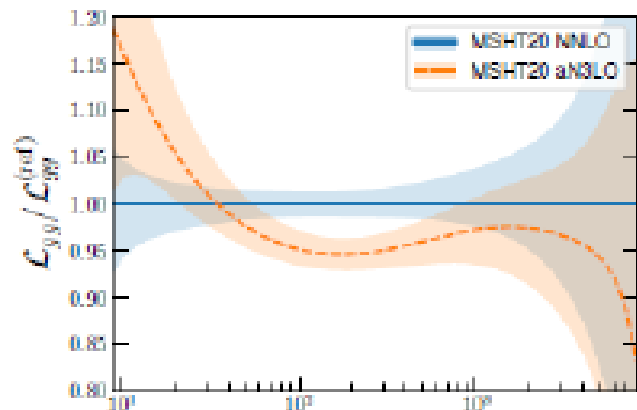
But if we compare these new MSHT PDFs to NNPDF, they are not consistent at T=1



Nor are they consistent with MSHT20 at T^2=1, but they are consistent within T^2=10



The difference between PDFs seemed to get worse at N3LO

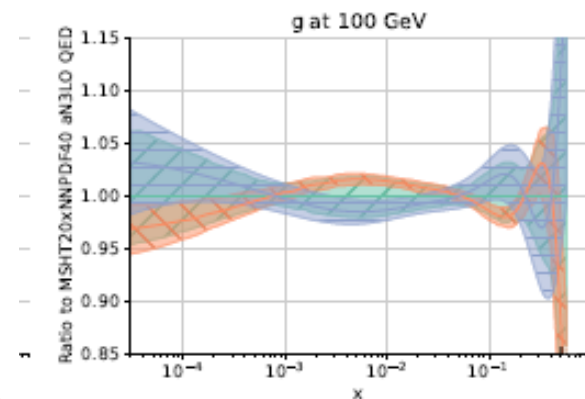


For published MSHT there is a 5% decrease in luminosity at the Higgs mass, for NNPDF this is more like 2%... BUT either way there is a significant difference.

BUT these come together a bit more with recent updates to N3LO and with QED corrections, which are larger for NNPDF than for MSHT.

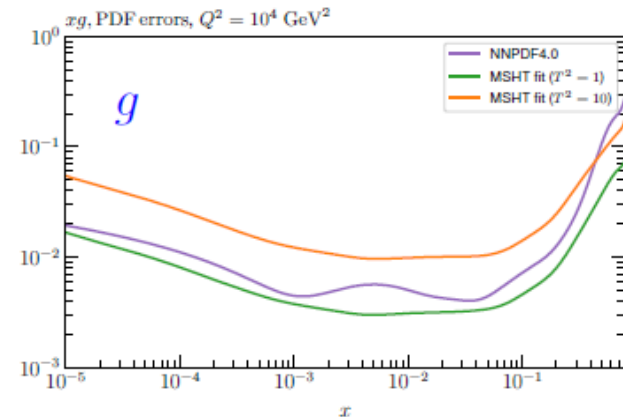
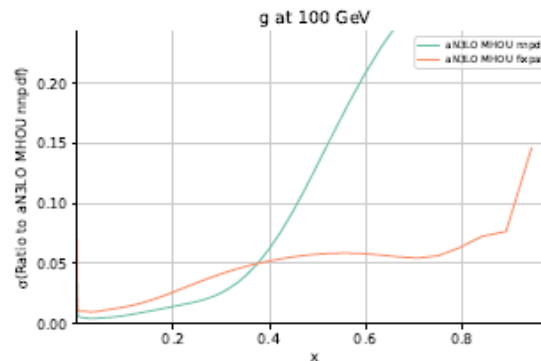
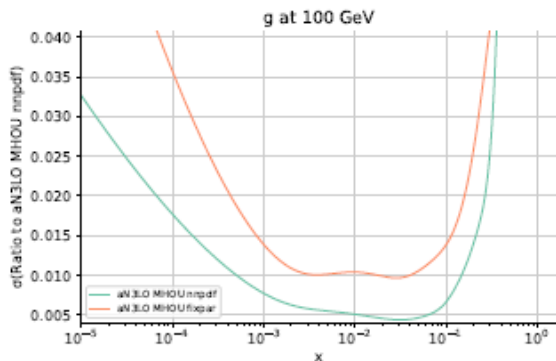
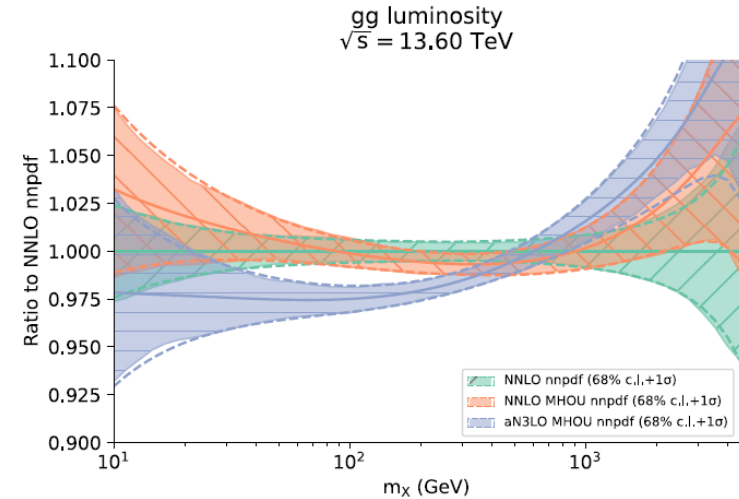
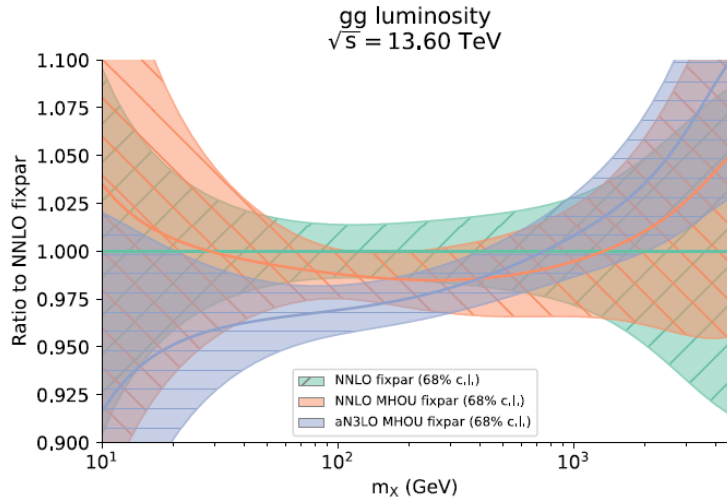
There is also a combination of NNPDF and MSHT aN3LO which covers their differences 2411.05373

Preliminary work of CT now confirm a shift of $\sim 2\%$ at the Higgs mass
(as of PDF4LHC N3LO benchmarking meeting June 2026)



We may worry that N3LO differences arise from methodology. Arxiv 2602.07118

An interesting exercise is to use the EXACTLY same data sets and the same theory settings with either the NNPDF neural net method or the fixed parametrisation method of MSHT. The effects of adding MHOU and moving to a N3LO are actually very similar in both cases – but the uncertainties of the fixed param method are larger ---due to enhanced tolerance

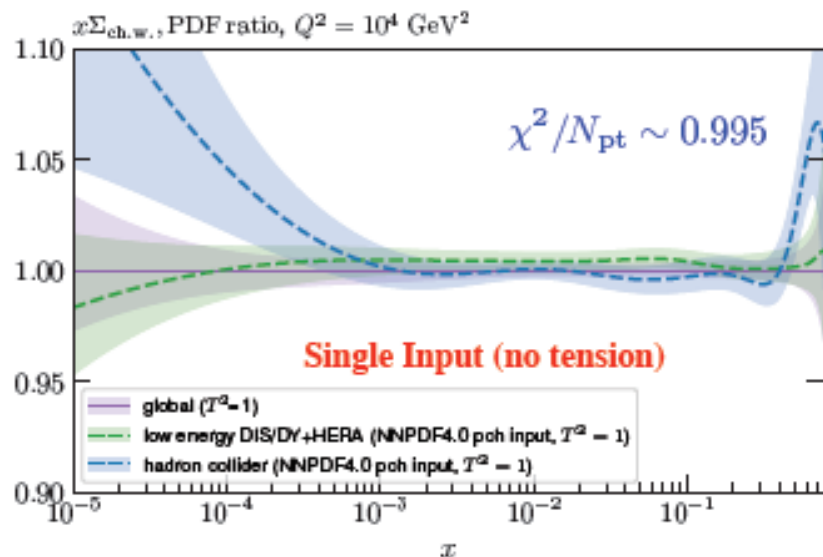


NNPDF is close to tolerance $T=1$, but not everywhere— high- x gluon!

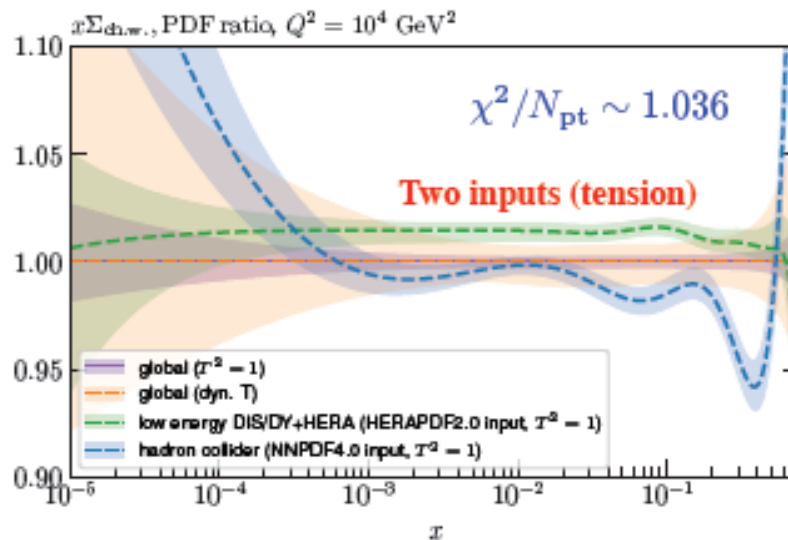
But is tolerance necessary?

CT [arxiv:2512.19779](https://arxiv.org/abs/2512.19779) and MSHT [Arxiv:2407.07944](https://arxiv.org/abs/2407.07944), [2408.11423](https://arxiv.org/abs/2408.11423)

certainly think so — one example that I find interesting injecting inconsistency into pseudodata to see what happens



Both low energy and high energy data generated from NNPDF, fit $\chi^2/\text{ndf} \sim 0.995$

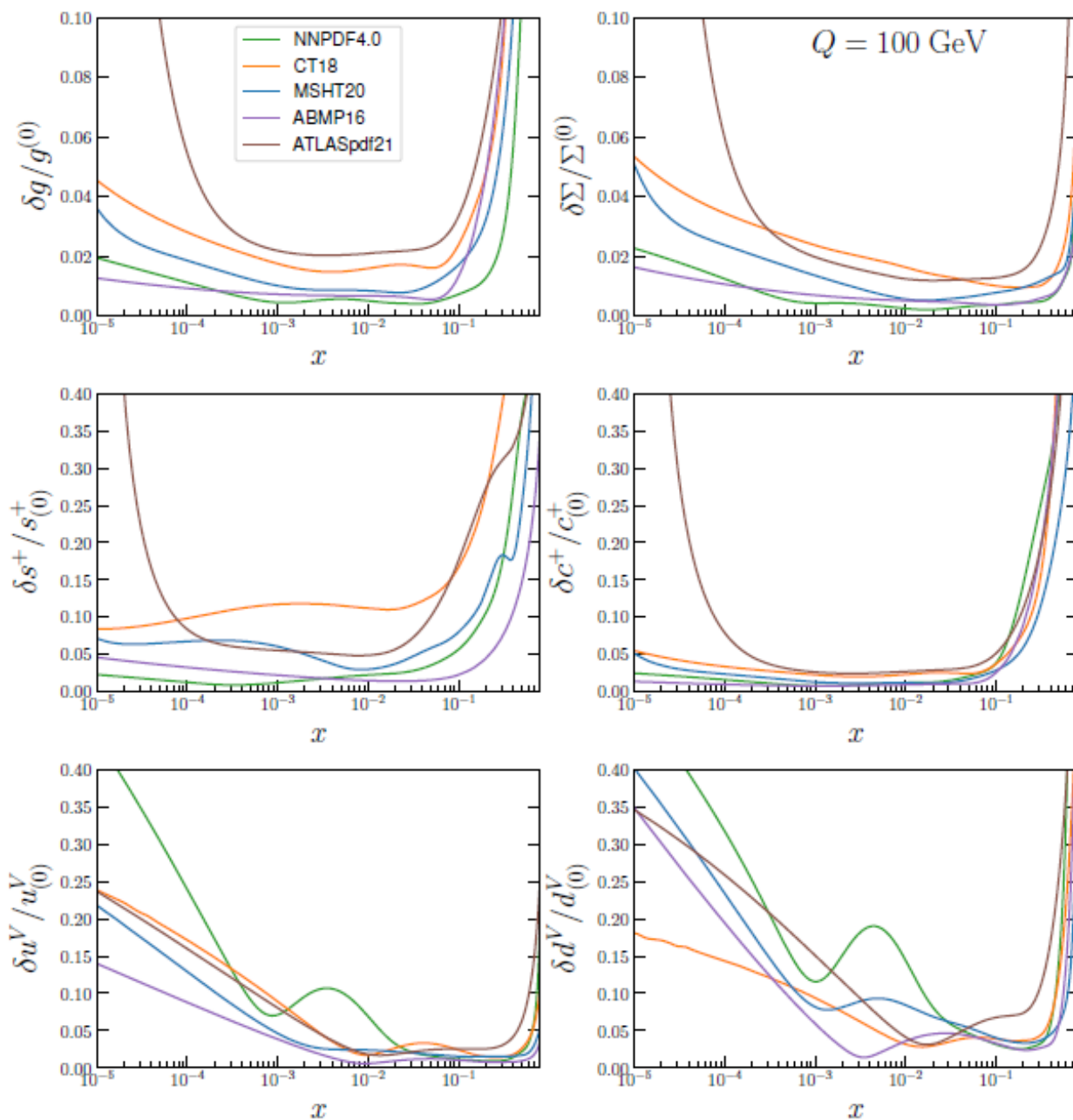


Low energy generated from HERAPDF, high energy data generated from NNPDF, fit $\chi^2/\text{ndf} \sim 1.035$

Interestingly the joint fit χ^2 does not reveal the inconsistency; the $\chi^2/\text{ndf} \sim 1.035$, is $< 2\sigma$ from 1. We often tolerate this (or worse) in PDF fits.

Only if we use a dynamic tolerance $T^2 \sim 10$, in the fit to both high and low energy data do we achieve consistency between fits to the two separate inputs and the 'global' result of fitting to both.

Now let's look at uncertainties for many PDFs



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
CT: $T \sim 6$, MSHT: $T \sim 3$ (still true for CT25)

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

So much work is going on on treatment of uncertainties

- CT25 alphas paper [arXiv:2512.23792](#) is a good read, summarising many methods
- MSHT have looked at 'errors on errors' – our uncertainties are rarely Gaussian. This gives a similar effect as using a tolerance $T=1.5$ – similar to the PDG procedure of reweighting uncertainties of each data set– but you still need an additional $T=2$ that may be attributed to inconsistencies between data sets
- NNPDF have worked on statistical estimators of when data sets are inconsistent and criteria for their acceptance/rejection. They also point out the importance of theory uncertainties-- once MHOU are considered most PDFs give similar quality of fits

Generally new data should make things better

- DIS data from the EIC should be the best- maybe we can replace old fixed target data
- Run 3 could tilt the balance between the our currently most accurate peak DY data and off-peak DY data at low and high mass. For jets /t-tbar it is less clear how much better we can do

But can we get better even before/without new data?

- There are obviously going to be correlations between the PDFs given that they use overlapping data sets—but by how much?.. Effort to assess this has stalled
- Can one just use DIS+DY for greater accuracy PDFs?

-----I doubt it ----

but if the definition of DY starts to include V+jets or Zpt then maybe..

Moving forward

We are all aware that PDFs are the limiting factor on precision SM measurements in the electroweak sector M_W , $\sin^2\Theta_W$ and the strong sector $\alpha_s(M_W)$

I don't think it is very easy to improve PDFs vastly right now -one always has to consider the difference between PDF sets

-- better data is always our best hope and better data is coming..slowly..

But we are only just finishing the precision SM Run2 analyses. Run3 is waiting.

Precision takes time

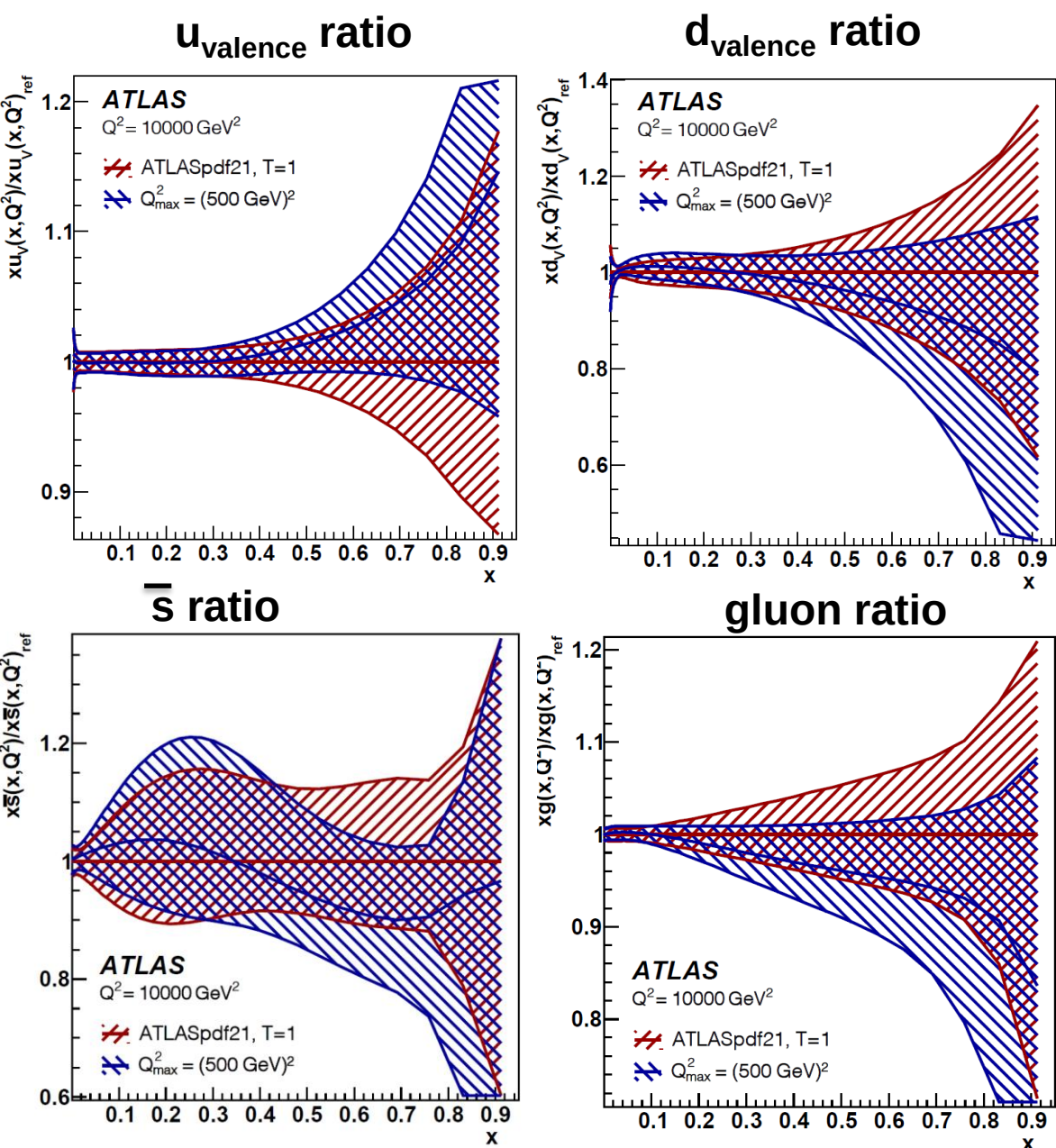
We need data with well understood systematic uncertainties that are also well specified
But the need to use data from different experiments, different runs, different teams remains and this always seems to lead to some level of inconsistency

Precision DY is our best data

- N3LO will become standard (slowly) and is already available for DIS and for DY
- LMDY brings low-x challenges– is full $\ln(1/x)$ resummation needed?
- HMDY brings the challenges of photon induced contributions, we will need QED evolution as well as QCD evolution for standard
- Low-pt needs $\ln(pt)$ resummation– and it seems this is needed until $pt \sim M_V$ (2607.19183)

Back-up

A fit was made 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



This cut mostly removes inclusive jet production data.

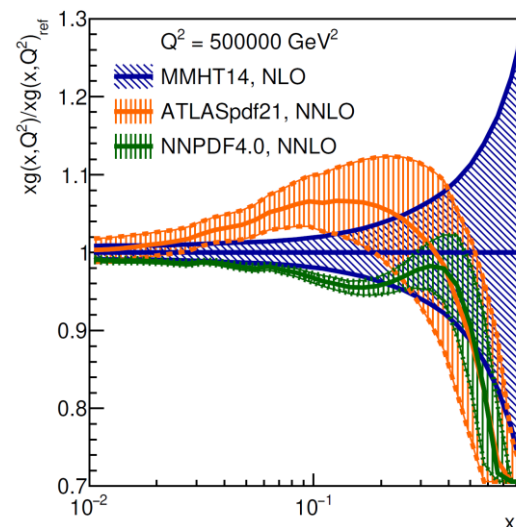
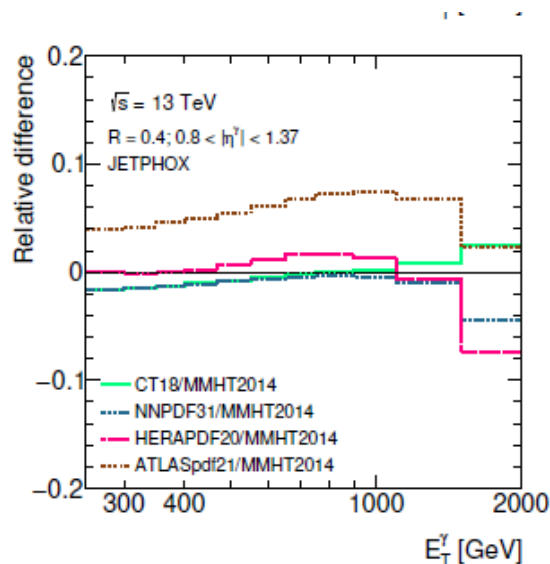
The effect is only seen at high x –note linear x scale- PDFs are not significantly changed

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

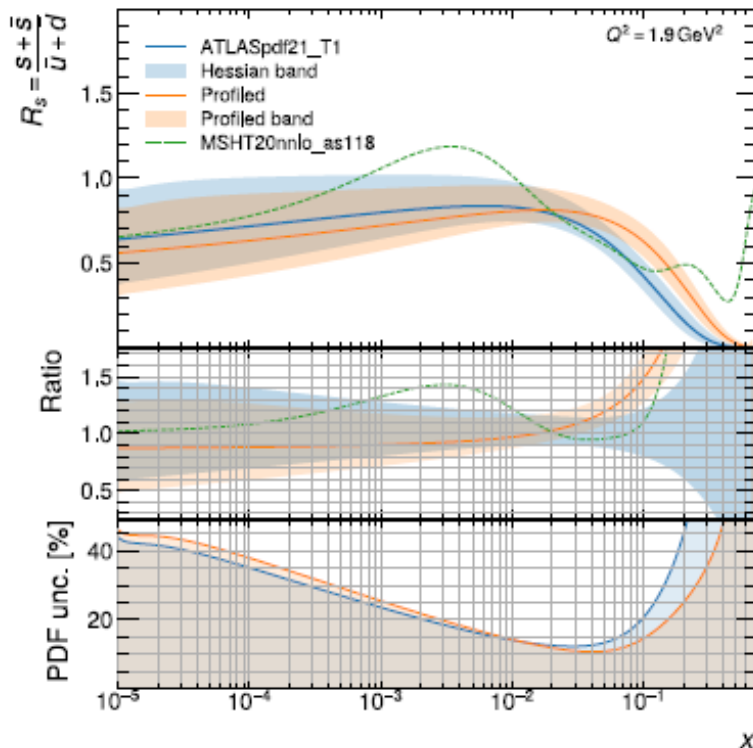
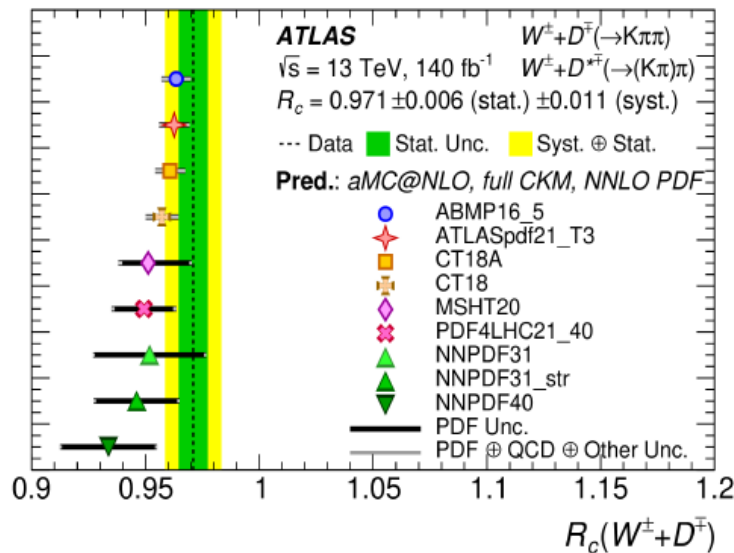
The 13 TeV direct photon data / [arXIV:2302.00510](https://arxiv.org/abs/2302.00510)

It is clear that these data are $\sim 5\%$ different from ATLASpdf21 predictions at $x \sim 0.1$ —this is a region where we have little other data to influence our fit.

We need new predictions- since the bins have changed since the earlier data These could come from either NLO*k-factors as last time or as an NNLO grid— we have already contacted the NNLOjet group—but this is not at the top of their ‘to-do’ list



Explanation in terms of our high- x gluon PDF at high scale— we always did admit to 2σ differences in this region



Ratio of W+/W- production suggests
 That these data prefer $s = \bar{s}$

Profiling ATLASpdf21 with all the W+D data
 with both signs and both D and D*

using NNLO predictions from Generet et al
 (2510.24525) BUT no grid

Suggests some impact at high x

But we lack a prediction grid to input these
 data to the fit

(Recent paper 2608.14268 is only NLO but with
 non zero charm mas)

Just a word on how correlated systematic uncertainties on data are treated in PDF fits. Our data are dominated by systematic uncertainties and these are often correlated point to point. In this case they cannot simply be added to the statistical and any uncorrelated systematics. An informative treatment is in terms of nuisance parameters

Where m^i is the theory prediction, depending on PDF parameters p , and μ^i is the measurement, each for data point i , and the measurement has statistical and uncorrelated systematic uncertainties $\delta\mu$, and correlated systematic uncertainties $\gamma\mu$ which are multiplied by nuisance parameters b for each source of uncertainty α . The uncertainties δ and γ are fractional.

$$\chi_{\text{exp}}^2(m, b) = \sum_i \frac{\left[m^i - \sum_{\alpha} \gamma_{\alpha}^i \mu^i b_{\alpha} - \mu^i\right]^2}{\left(\delta_{i,\text{stat}} \mu^i\right)^2 + \left(\delta_{i,\text{uncor}} \mu^i\right)^2} + \sum_{\alpha} b_{\alpha}^2.$$

This amounts to shifting the central values of the measurements by the estimates given by the systematic uncertainties, where $b=0$ is no shift and $b=\pm 1$ represent 1- σ shifts. This is appropriate for additive systematics, for multiplicative one can equivalently shift the predictions.

The fit determines both the PDF parameters p and the nuisance parameters b , but the calculation with respect to these is done analytically.

This is a more informative way to formulate the Chisq than using an overall covariance matrix because the effect of individual sources can be studied.

The xFitter manual is good source for study of alternative forms of the Chisq

Also beware—some of the PDF fitting groups chose to decorrelate some of our systematic uncertainties, when there are large Chisq issues... some of them consult us about it.. and some of them don't

You might want to assess the impact on the PDFs of some new data

Profiling varies nuisance parameters for the new data **and** for the eigenvectors of the PDF fit prediction
This assigns a different weight to each eigenvector variation such that a new PDF set can be constructed

$$f'_0 = f_0 + \sum_{\beta} b_{\beta,th}^{min} \left(\frac{f_{\beta}^+ - f_{\beta}^-}{2} - b_{\beta,th}^{min} \frac{f_{\beta}^+ + f_{\beta}^- - 2f_0}{2} \right),$$

where f_0 is the original central set and f_{β}^+ and f_{β}^- are the up and down eigenvectors

$$\chi^2(b_{exp}, b_{th}) = \sum_{i=1}^{N_{data}} \frac{(\sigma_i^{exp} + \sum_{\alpha} \Gamma_{i\alpha}^{exp} b_{\alpha,exp} - \sigma_i^{th} - \sum_{\beta} \Gamma_{i\beta}^{th} b_{\beta,th})^2}{\Delta_i^2} + \sum_{\alpha} b_{\alpha,exp}^2 + \sum_{\beta} b_{\beta,th}^2.$$

Profiling like this reweights the different eigenvectors to better fit your data. It cannot be trusted if changes are large, then you need a refit including the new data.

Also this form of profiling assumes $T=1$ for the new data set and thus assigns it a very large weight with respect to data in the fit, since the CT tolerance is usually $T^2 \sim 37$ and MSHT is $T^2 \sim 10$ for 68% CK. Also T should really be a function of β for dynamic tolerance.

Can use the **utility ePump** for profiling with Tolerance.

For NNPDF replicas **reweighting** is more appropriate than profiling (although you can use the equivalent Hessian set)

$$\langle O(\text{PDF}) \rangle = \frac{1}{N_{rep}} \sum_{k=1}^{N_{rep}} O(\text{PDF}_k) \longrightarrow \langle O^{new}(\text{PDF}) \rangle = \frac{1}{N_{rep}} \sum_{k=1}^{N_{rep}} w_k O(\text{PDF}_k)$$

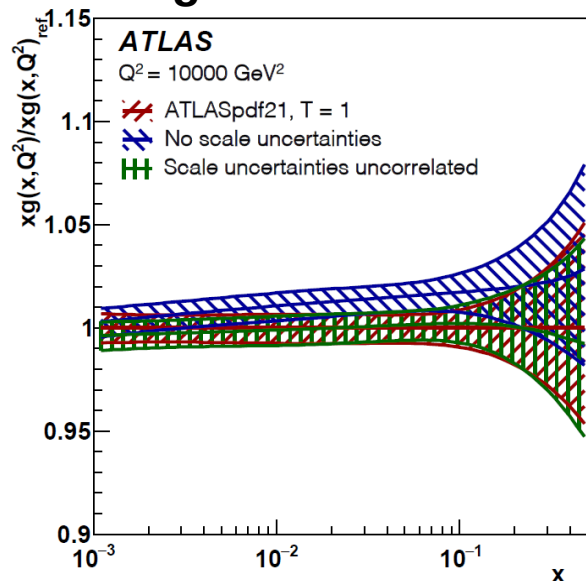
Where the weights are given by reweighting the best Chisq with Larger weights

$$w_k = \frac{(\chi_k^2)^{\frac{1}{2}(N_{data}-1)} \exp^{-\frac{1}{2}\chi_k^2}}{\frac{1}{N_{rep}} \sum_{k=1}^{N_{rep}} (\chi_k^2)^{\frac{1}{2}(N_{data}-1)} \exp^{-\frac{1}{2}\chi_k^2}},$$

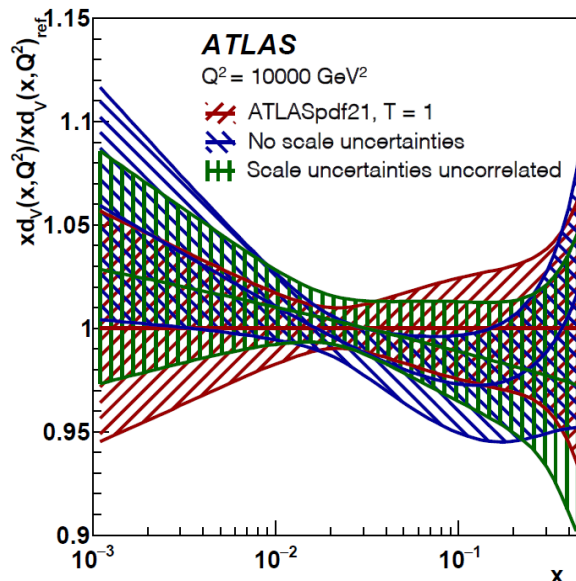
Impact of scale uncertainties

For the inclusive W,Z production at 7 and 8 TeV the experimental uncertainties are comparable to the scale uncertainties and thus the scale uncertainties are included as theoretical uncertainties in the fit. By default they are correlated between the W and Z data and between the 7 and 8 TeV data in the ATLASpdf21 fit.

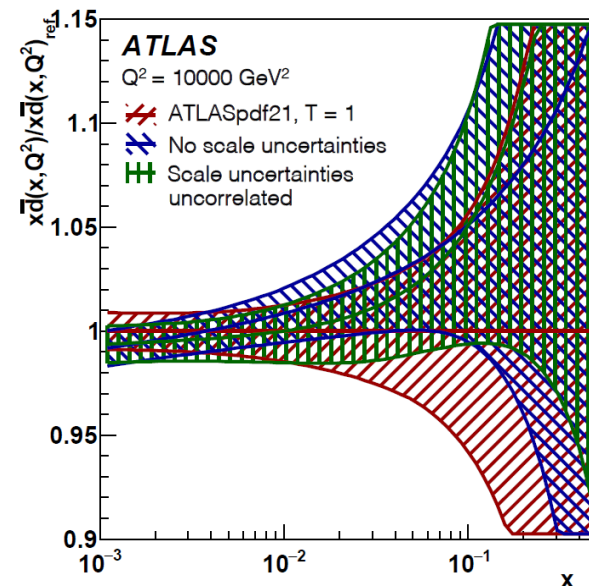
gluon ratio



d_{valence} ratio



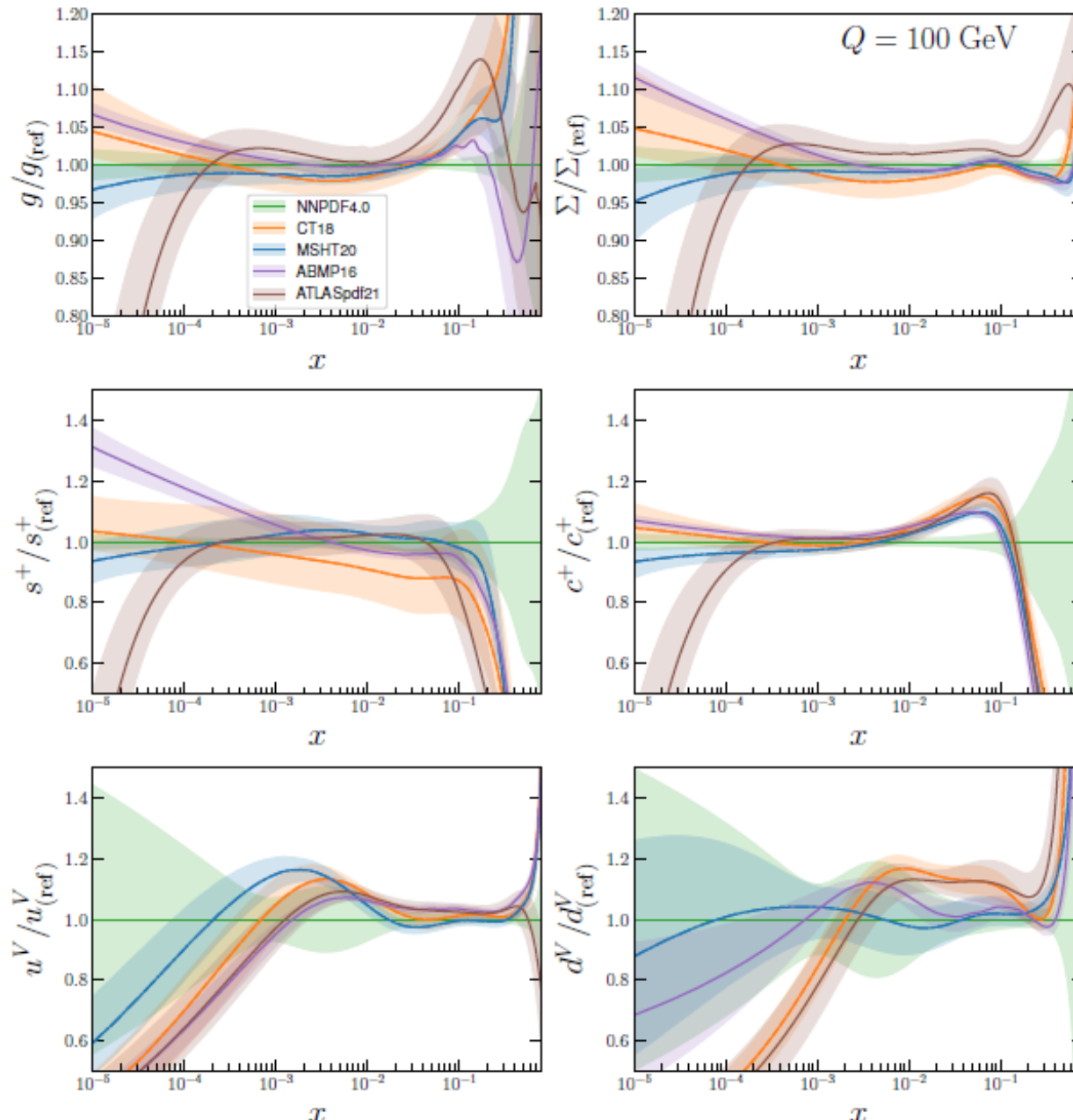
$\bar{d}$ ratio



Here we show the ratios of the gluon, d-valence and dbar PDFs with (red) and without (blue) these scale uncertainties included
In green we show the effect of including scale uncertainties but not correlating them between 7 and 8 TeV data

Clearly scale uncertainties can be important if 1% precision is sought

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



NOTE PDFs with less data like ABMP or ATLASpdf tend to be outliers at low and high-x

We are not so surprised by differences at high-x, though they can be outside uncertainties
(e.g. NNPDF has intrinsic charm. But it also less strange suppression at high-x)

Differences in low-x valence are also unsurprising, little is known on valence at very low-x and it is not very important

BUT beware of the low-x gluon differences