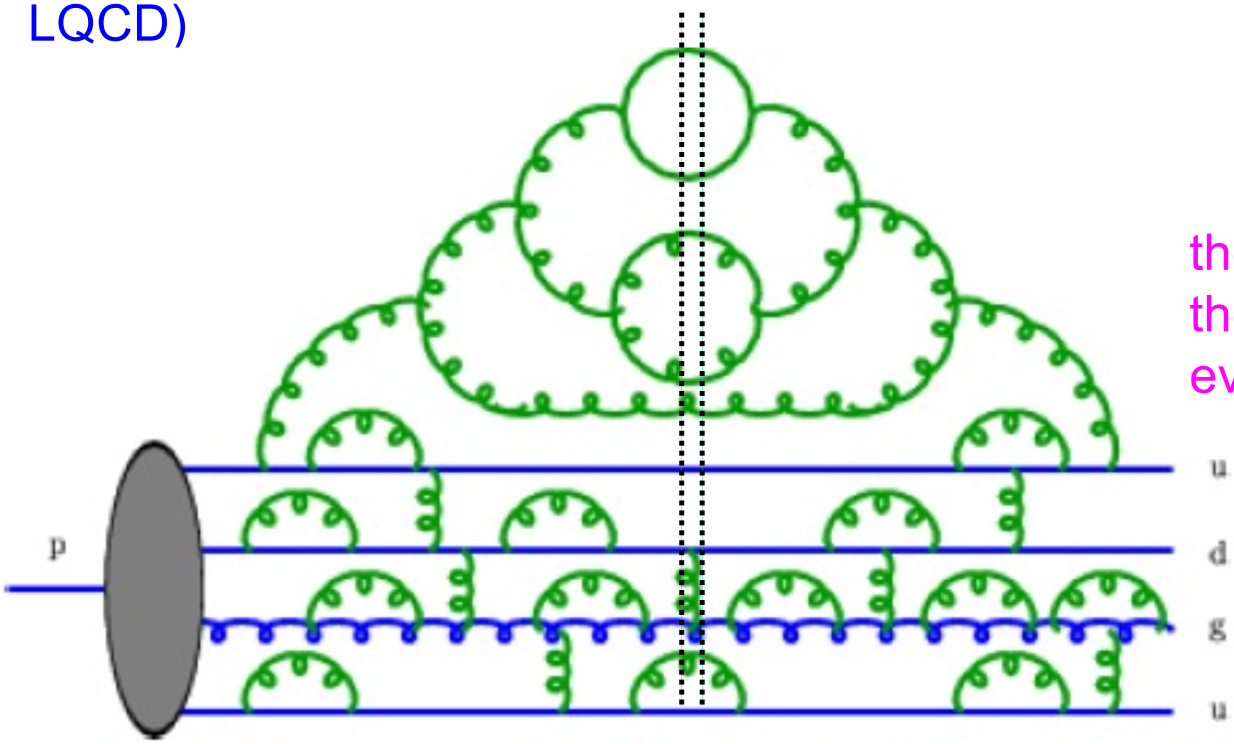
Status Overview of Global PDF Fits (with a concentration on tools for understanding PDF uncertainty)

J. Huston

Michigan State University

PDFs

- The proton is a dynamical object; the structure observed depends on the time-scale (Q^2) of the observation
- But we know how to calculate this variation (DGLAP) at LO, NLO, NNLO and now at aN3LO
- We *just* have to determine the starting points from fits to data (or from LQCD)



$f_i(x, Q^2)$ = number density of partons i
at momentum fraction x and probing scale Q^2

the higher the value of Q^2, the more detail of the evolution we sample

PDFs

- Determined from global fits to data from a wide variety of processes, both from fixed target and collider experiments, with an increasing contribution from the LHC itself
- The 3 main PDF groups are CTEQ-TEA (CT), MSHT (new acronym) and NNPDF; other fits by ATLAS, CMS
- Each global fit uses on order of 4000 data points to determine the best fit PDFs and their uncertainties
  - with CT and MSHT using a Hessian formalism and NNPDF using a neural net formalism
- Each group provides regularly updated sets of PDFs

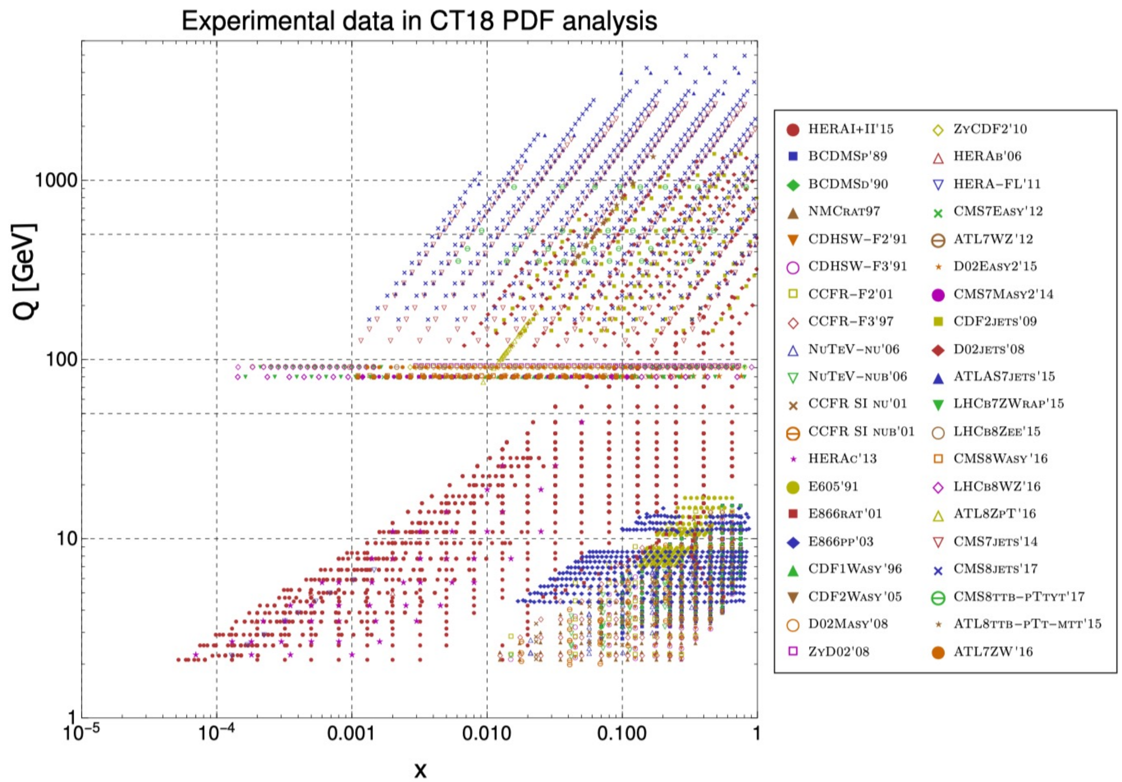


FIG. 1: The CT18 data set, represented in a space of partonic (x, Q), based on Born-level kinematical matchings, $(x, Q) = (x_B, Q)$, in DIS, *etc.*. The matching conventions used here are described in Ref. [20]. Also shown are the ATLAS 7 TeV W/Z production data (ID=248), labeled ATL7WZ'12, fitted in CT18Z.

because of the difference, there can sometimes be difficulties in comparing/understanding results

to better understand similarities and differences, it is useful to periodically perform benchmarking exercises, and to construct analytical tools

PDFs

- Determined from global fits to data from a wide variety of processes, both from fixed target and collider experiments, with an increasing contribution from the LHC itself
- The 3 main PDF groups are CTEQ-TEA (CT), MSHT (new acronym) and NNPDF; other fits by ATLAS, CMS
- Each global fit uses on order of 4000 data points to determine the best fit PDFs and their uncertainties
  - with CT and MSHT using a Hessian formalism and NNPDF using a neural net formalism
- Each group provides regularly updated sets of PDFs

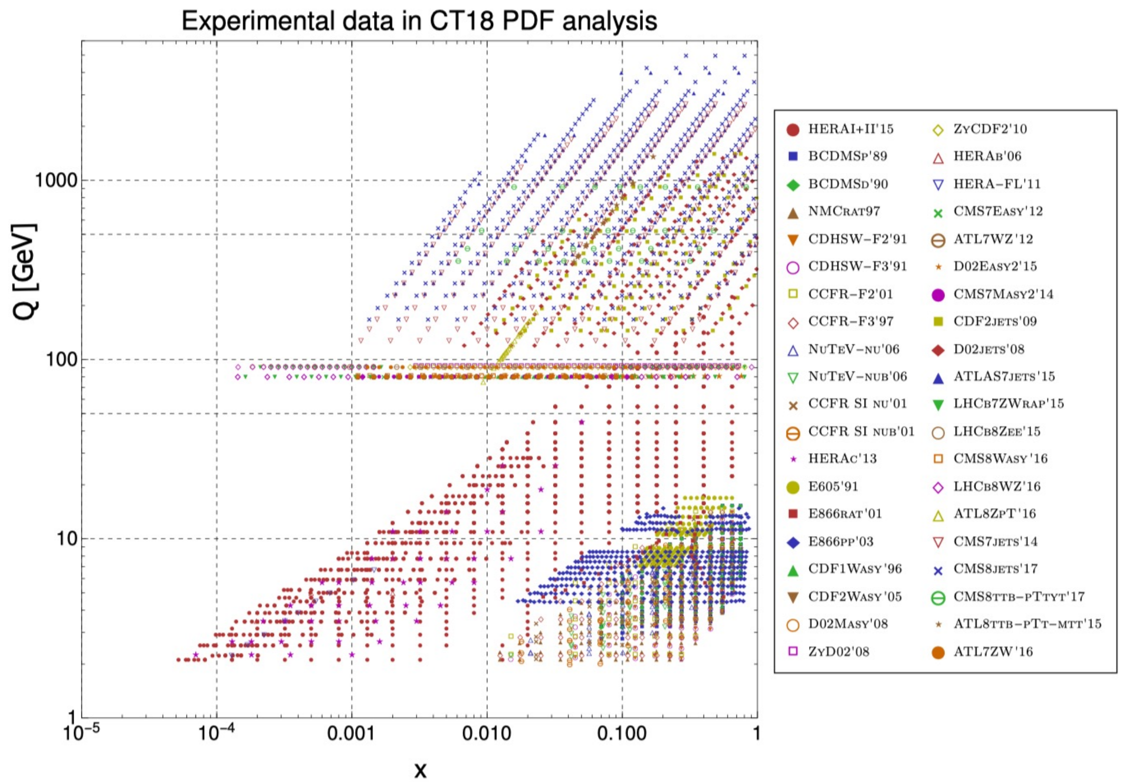
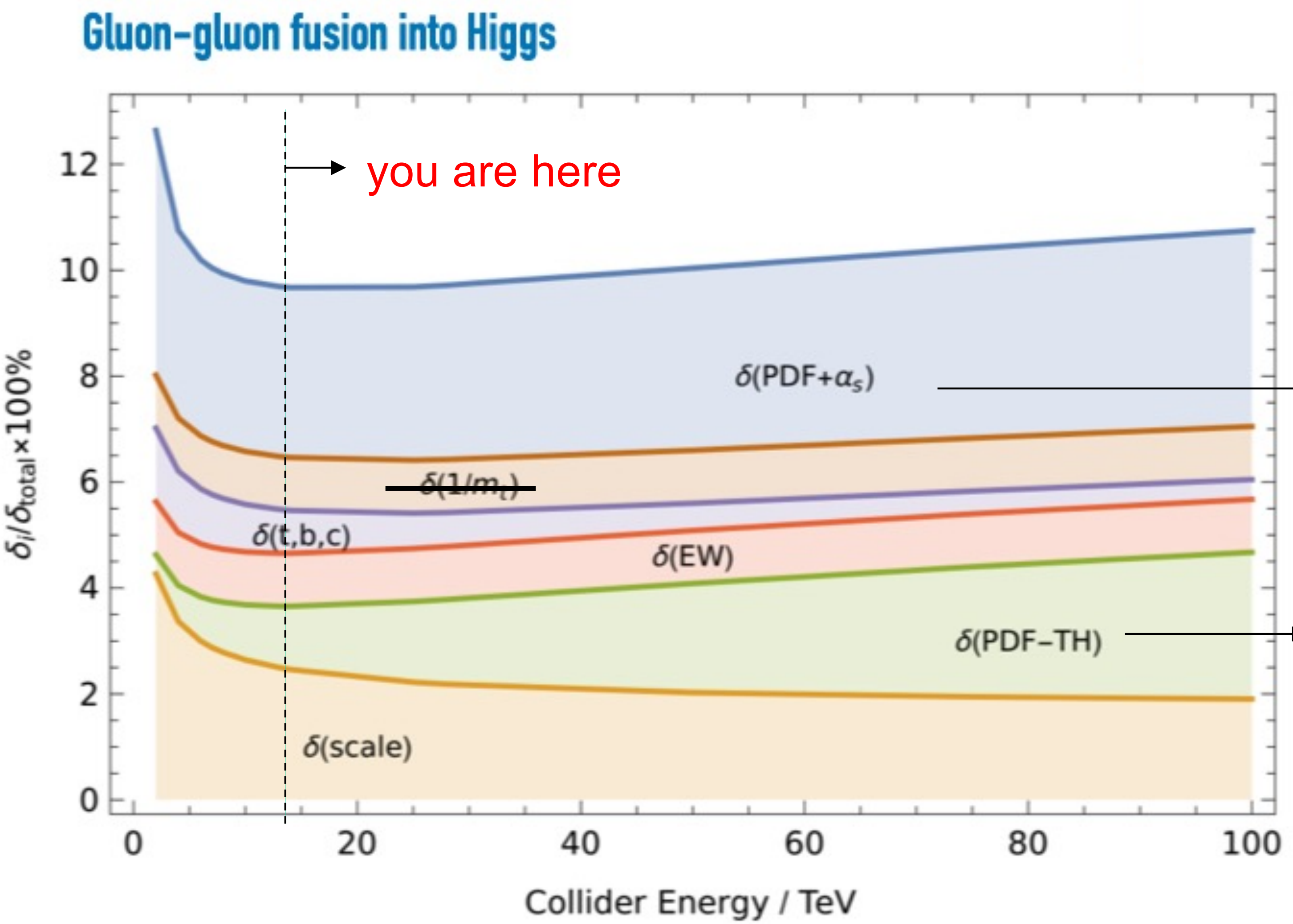


FIG. 1: The CT18 data set, represented in a space of partonic (x, Q), based on Born-level kinematical matchings, $(x, Q) = (x_B, Q)$, in DIS, *etc.*. The matching conventions used here are described in Ref. [20]. Also shown are the ATLAS 7 TeV W/Z production data (ID=248), labeled ATL7WZ'12, fitted in CT18Z.

The comparison of these two very different techniques allow us to obtain a better understanding of the PDFs and their uncertainties.

Better precision for PDFs easy to motivate



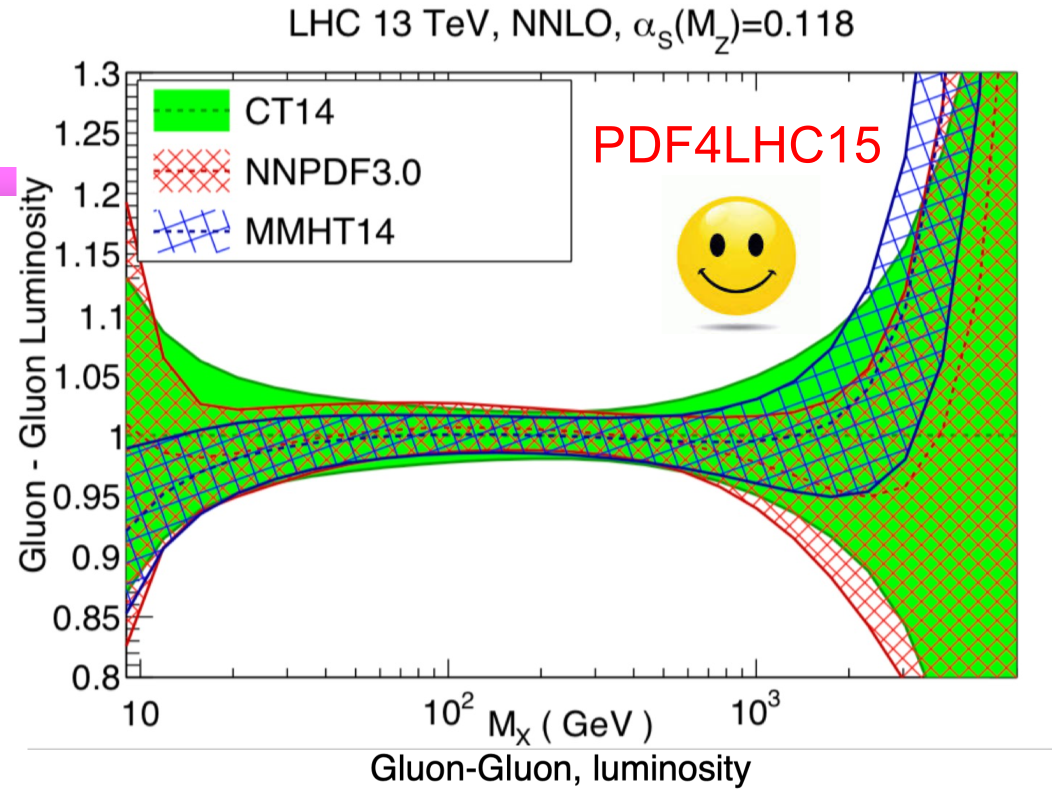
α_s uncertainties can be as important as PDF uncertainties, for an α_s-rich process such as ggF

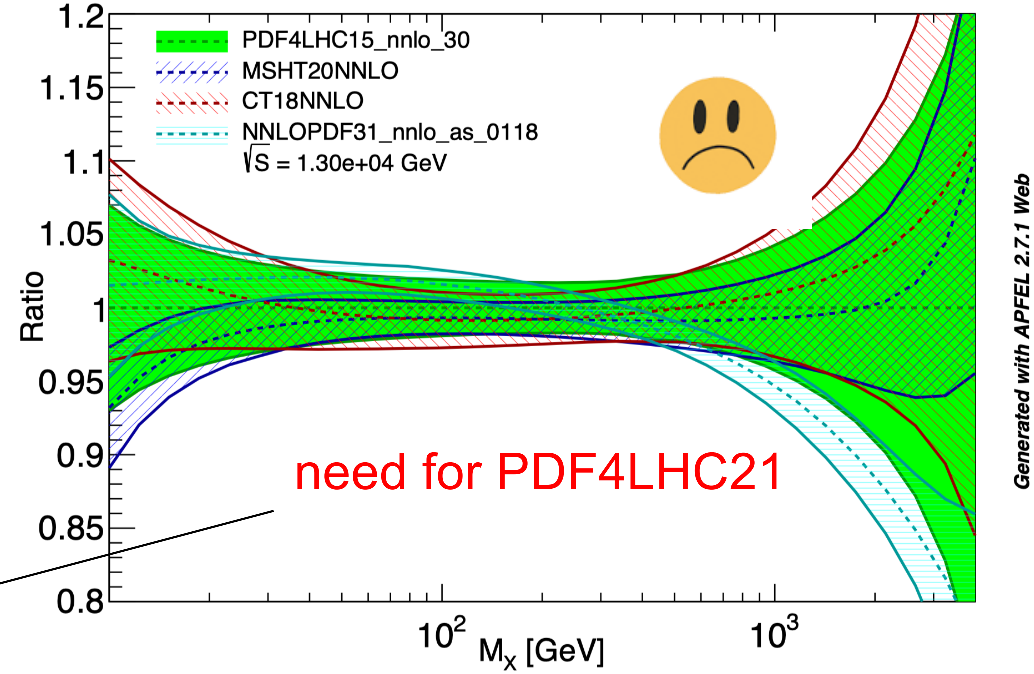
PDFs are only known to NNLO
So far; a few cross sections, such as ggF and DY, are known to N3LO

When can we produce N3LO PDFs?

Global PDF fits

- CT and MSHT both use a Hessian-based approach (for the determination of the central PDF and the uncertainties), while NNPDF uses a Monte Carlo replica approach (although the Monte Carlo replica basis can be converted into a Hessian basis, and indeed this is often the format that allows the easiest understanding of the uncertainties, *IMHO*)
- One of the crucial services that can be provided by the global PDF fitters is to try to provide a better understanding of the central values and uncertainties obtained by each of the fits





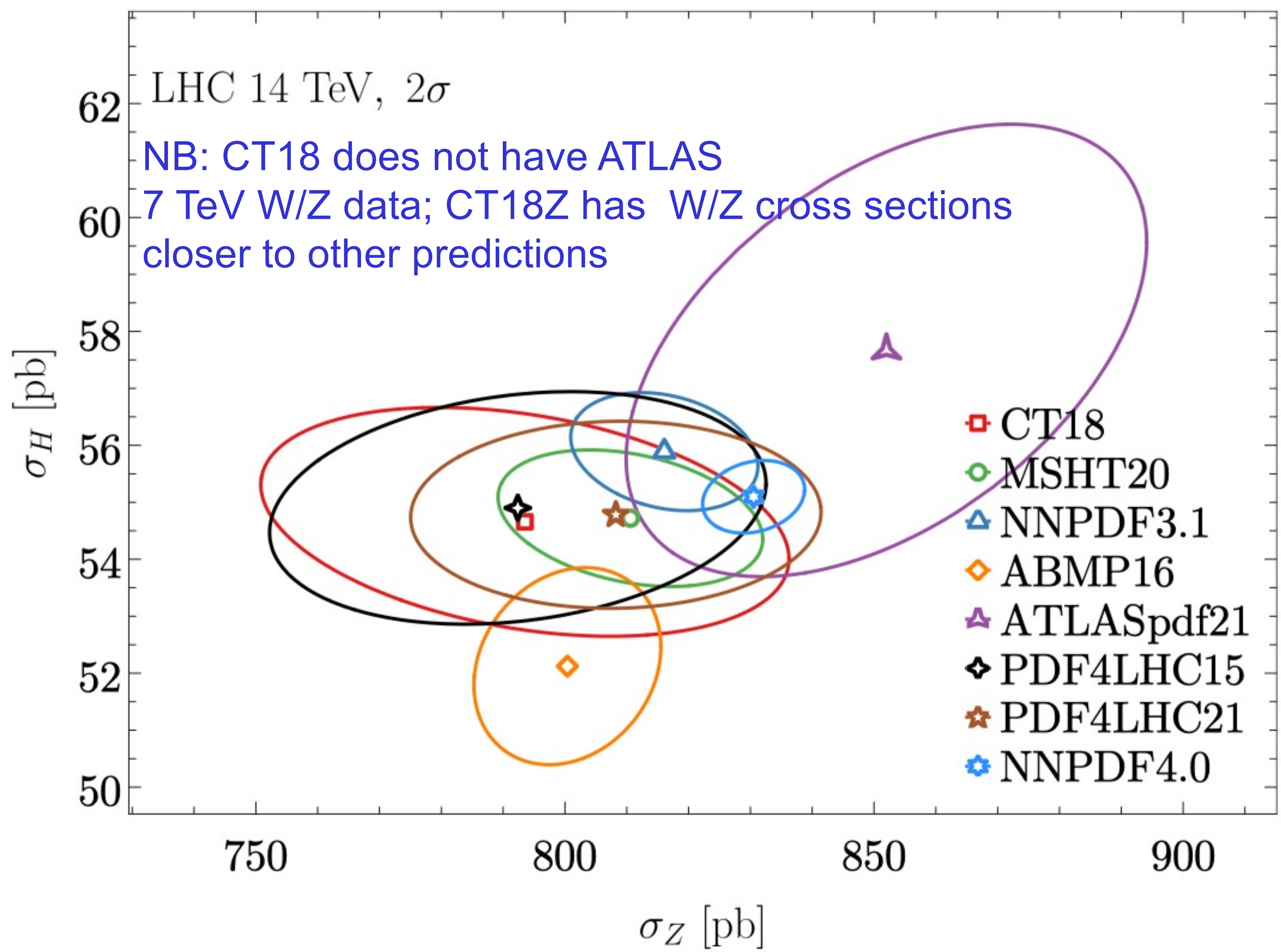
J.Phys.G 49 (2022) 8, 080501
e-Print: 2203.05506

Aside: uncertainties

- PDF uncertainties depend first of all on the experimental uncertainties of the data
- Data from two measurements, or even from within the same measurement, can both be very precise, but the result of adding both to the PDF fit can be an increase in the PDF uncertainty (or more likely) a smaller decrease in uncertainty than expected) if the data are in tension with each other
- The resultant PDF uncertainty relies on the definition of a tolerance, i.e. what is a significant increase from the global minimum χ^2, i.e. PDF uncertainty can be adjusted by changing the tolerance
- $\Delta\chi^2=1$ is not applicable for ~4000 data points from different experiments
- NB: all groups see tensions; the relevant χ^2 values show that the fits do not correspond to zero tension (see tables in PDF4LHC21 doc)
- NB: CT (Tier 2) and MSHT (dynamic tolerance) have introduced criteria to restrict the pull of data sets that disagree with global fit
  - MSHT criterion is sometimes stricter
- All groups sometimes throw away data sets that produce very large χ^2; some data sets need extra de-correlations provided by the exptl collaborations to produce reasonable χ^2 values

Useful to look at pairs of cross sections

Precision PDFs (Snowmass 21 WP) [2203.13923]



NNPDF3.1 uncertainty is smaller than either CT18 or MSHT20

NNPDF4.0 uncertainty is smaller still

NNPDF4.0 Higgs σ is closer to CT18, MSHT20

What is the best estimate of the uncertainty? and the central value?

PDF4LHC21 exercise: reduced data set

- Diverse enough to provide information for all PDFs
- Sparse enough that uncertainties should be very similar for all 3 PDFs
- Origins of differences of PDFs
  - due to variations of experimental input, treatment of systematic errors, different theory settings, fitting methodologies?
  - so for benchmarking, use common theory settings (i.e. perturbative charm, m_{charm}=1.4 GeV, s=sbar at input scale, $\alpha_s(m_Z)$=0.118, positive-definite PDFs, no deuteron or nuclear corrections…)
  - add several data sets to NNPDF3.1->3.1' (closer to 4.0)

Dataset	Reference	Dataset	Reference
BCDMS proton, deuteron DIS	[155, 156]	LHCb 8 TeV $Z \to ee$	[62]
NMC deuteron to proton ratio DIS	[157]	ATLAS 7 TeV high precision W, Z (2016)	[63]
NuTeV νN dimuon	[158]	D0 Z rapidity	[159]
HERA I+II inclusive DIS	[60]	CMS 7 TeV electron asymmetry	[160]
E866 Drell-Yan ratio pd/pp DIS	[161]	ATLAS 7 TeV W, Z rapidity (2011)	[149]
LHCb 7, 8 TeV W, Z rapidity	[61, 65]	CMS 8 TeV inclusive jet	[69]

Table 3.1. The measurements included in the initial round of PDF fits to a reduced dataset, together with the corresponding publication reference. This dataset is chosen as the largest subset of data fit by CT18, MHST20, and NNPDF3.1 in an (almost) identical manner.

Reduced fits

- Central values agree reasonably well
- …as do uncertainties at higher x
- There are some differences, for example at low x for the gluon distribution; this is a region nominally not well constrained by data

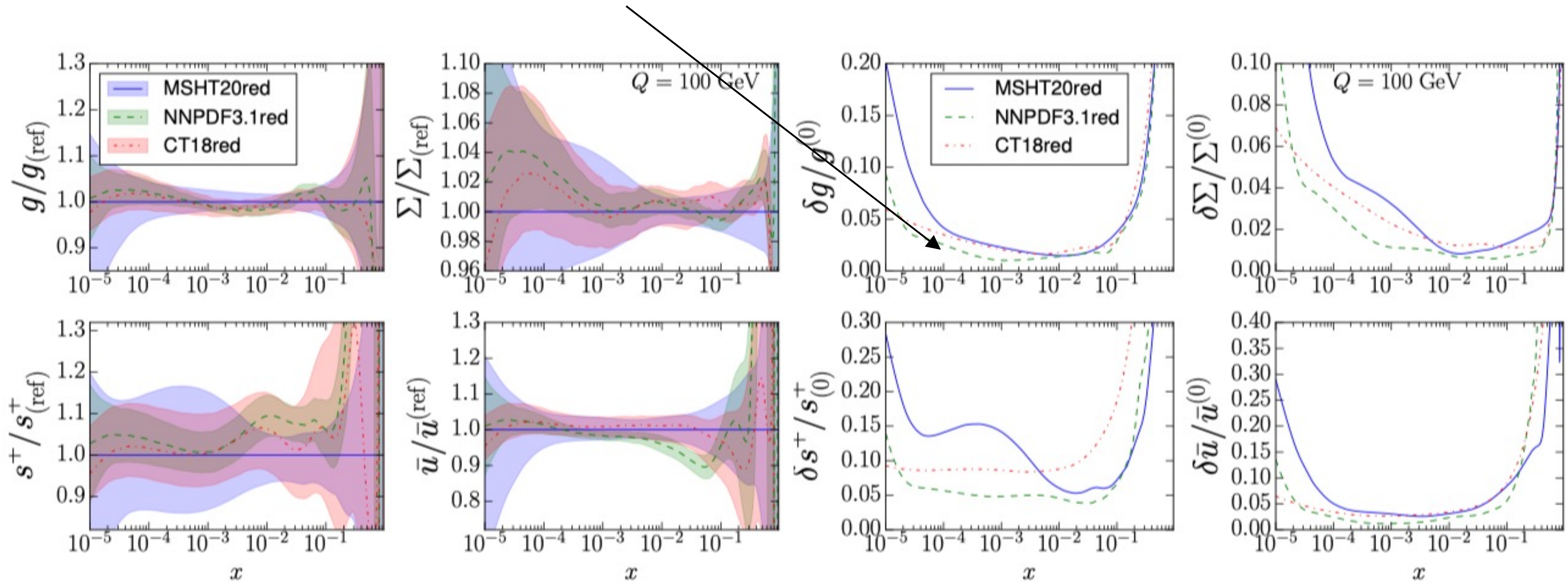
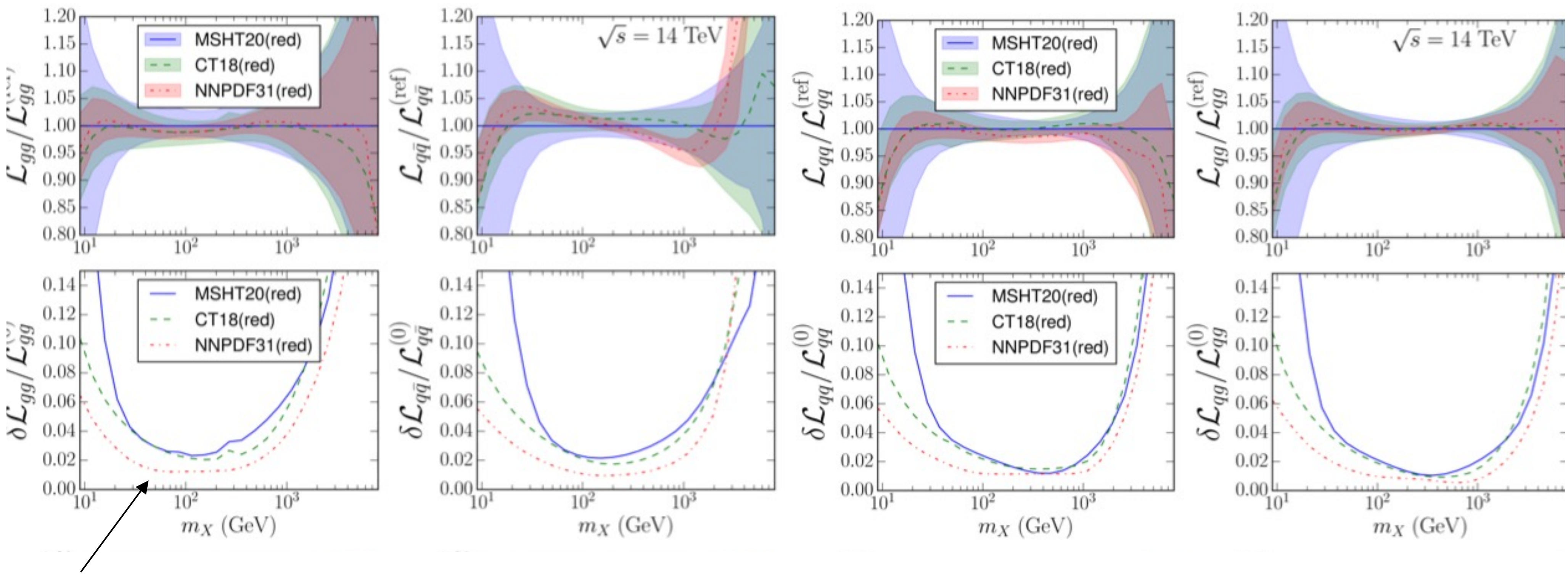


Figure 3.4. Comparison between the reduced PDF fits from the three groups, in the same format as in Fig. 3.1. For the three groups, PDF errors correspond to 1σ intervals. In the left panels, PDFs are displayed normalised to the central value of the MSHT20 reduced PDF set.

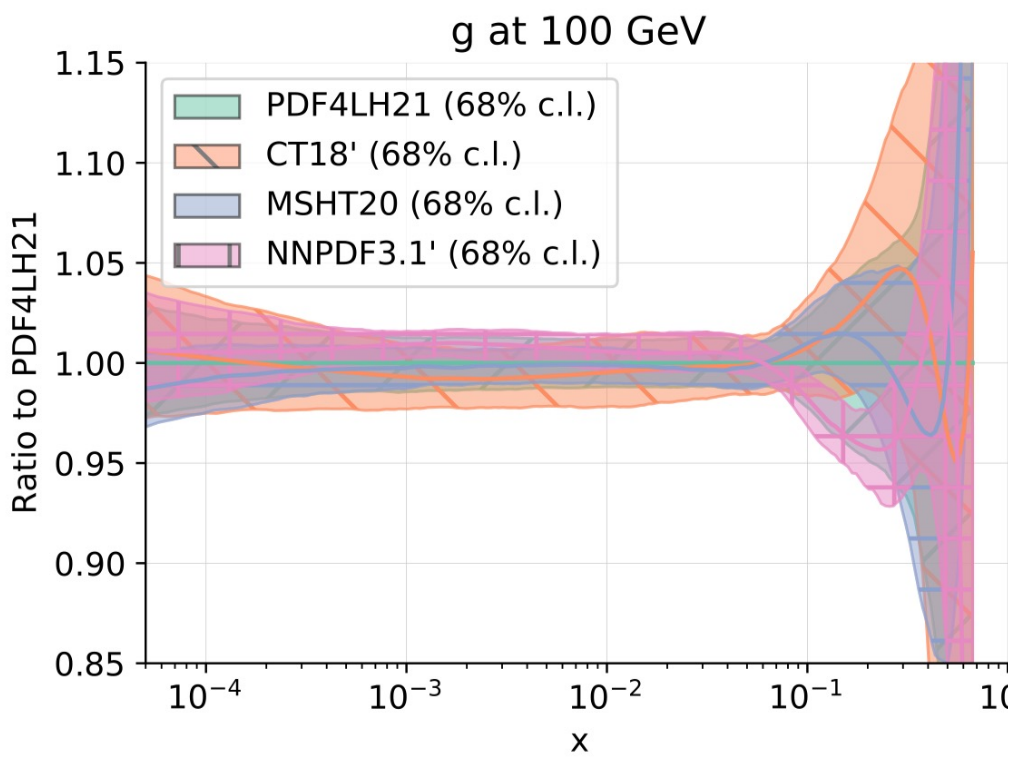
PDF luminosities for reduced fits

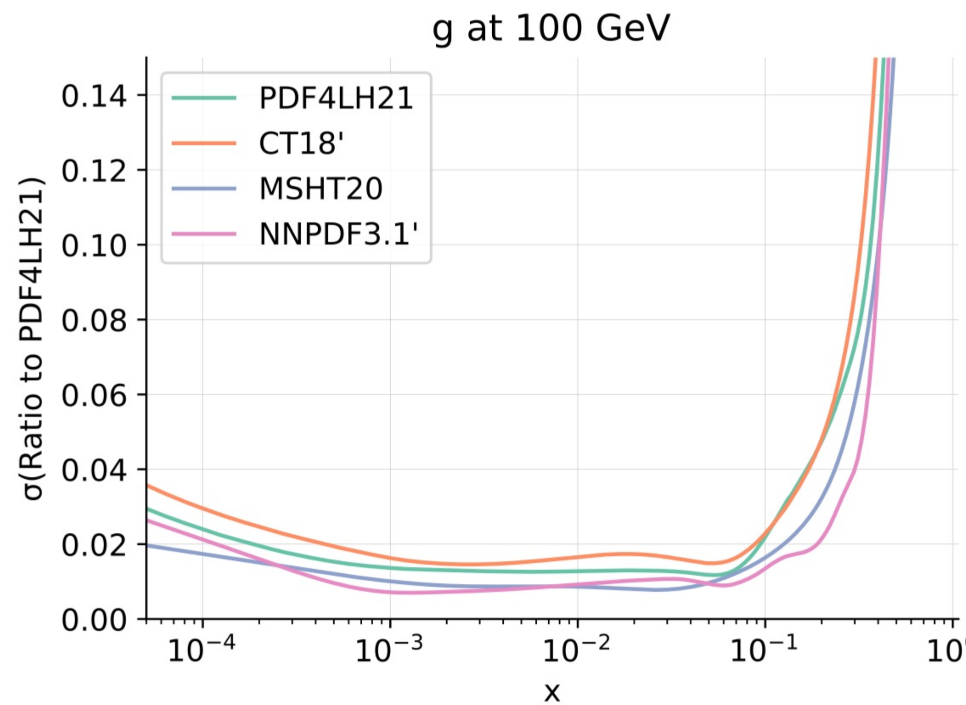


NNPDF3.1 has significantly reduced gg uncertainty using the same set of data; this implies their effective tolerance (for the same data information) is smaller than for CT or MSHT; the effect is even larger with NNPDF4.0. Due just to smaller gluon uncertainty? Maybe correlations are also different?

Figure 3.5. Comparison of the partonic luminosities between the CT18, MSHT20, and NNPDF3.1 reduced fits at $\sqrt{s} = 14$ TeV as a function of the invariant mass of the produced final state m_X. From left to right we show the gluon-gluon, quark-antiquark, quark-quark and quark-gluon luminosities, normalised to the central value of the MSHT20 prediction, together with the associated 1σ relative PDF uncertainties. The upper panels display the luminosities evaluated without any restriction on the final-state rapidity y_X, while the bottom panels instead account for a rapidity cut of $|y_X| < 2.5$ which restricts the produced final state to lie within the ATLAS/CMS central acceptance region.

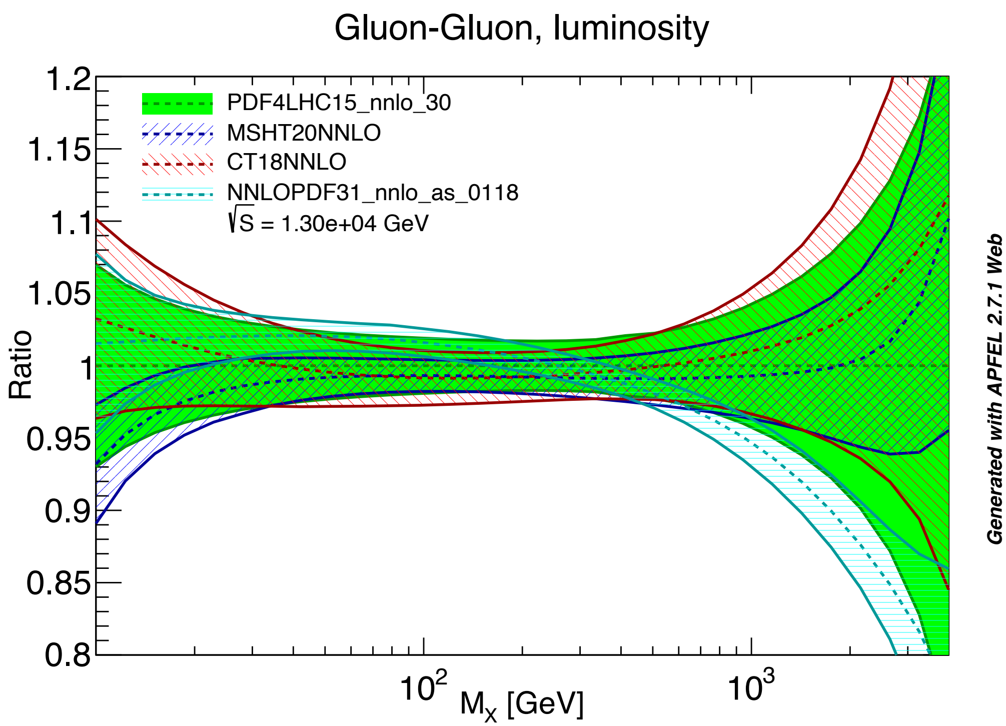
Gluon for PDF4LHC21

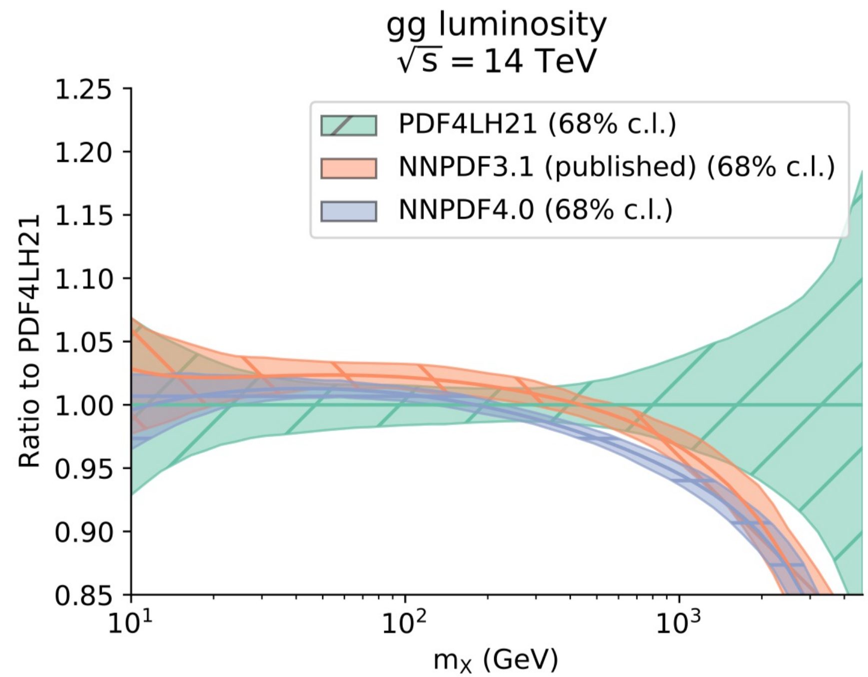




The prime signifies modifications from the original PDF needed for combination; in the CT18’ case, use mc=1.4 GeV instead of 1.3 GeV): in the NNPDF3.1’ case, several major new datasets added (which came after the publication of NNPDF3.1) -> “halfway to NNPDF4.0”

PDF luminosities for full fits





NNPDF4.0 has a larger data set than 3.1, but the crucial data sets are already in 3.1' used for the PDF4LHC21 combination (which are common with CT18 and MSHT20). Note that small datasets may create a problem with the use of a sampling technique. Because of the small number of data points, all data points are typically used for both sampling and testing.

Additional tools for understanding uncertainty

- For data to influence the PDF fit in a particular region of x and Q^2, two conditions must be met
  - the parton-level dynamics must depend on a particular PDF (say that of the gluon), as manifested in a statistical correlation
  - the data must have sufficient resolving power to contribute to the PDF likelihood analysis
- The L_2 sensitivity incorporates both of these features
- The L_2 sensitivity is a way of viewing the pulls of all of the experiments used in a global PDF fit, for a particular parton flavor, as a function of a kinematic variable, such as parton x
  - or, when plotted for a PDF luminosity, as a function of the mass
- The fit value for a particular PDF(x,Q) is determined by the sum of these pulls

What is the L_2 sensitivity?

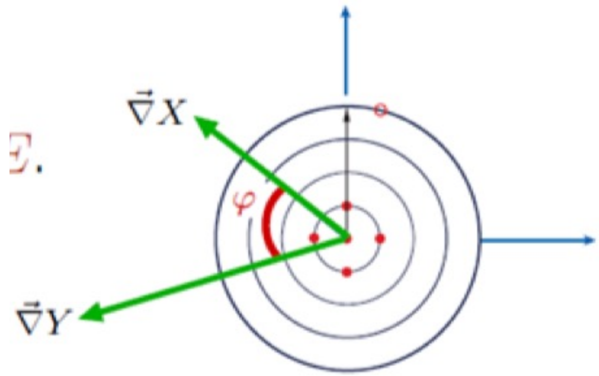
- The L_2 sensitivity provides a visualization of what is happening inside the PDF fit
- It can be considered as a faster version of Lagrange Multiplier scans (but dependent on the Gaussian approximation)
- The L_2 sensitivity streamlines comparisons among independent analyses, using the log-likelihood (χ^2) values for the fitted experiments and the error PDFs
- Both the L_2 and LM methods explore the parametric dependence of the χ^2 function in the vicinity of the global minimum
- The L_2 sensitivity has been used internally by CT (in CT18), by the PDF4LHC21 benchmarking group (to determine which data sets should be in the reduced PDF fit used for benchmarking), and now by CT, MSHT and ATLASpdf in a common paper arXiv:2306.03918

L_2 sensitivity

$$S^{\mathrm{H}}_{f,L2}(E) \equiv \frac{\vec{\nabla}\chi^2_E \cdot \vec{\nabla} f}{\Delta^{\mathrm{H}} f} = \left(\Delta^{\mathrm{H}}\chi^2_E\right)\; C^{\mathrm{H}}(f, \chi^2_E)$$

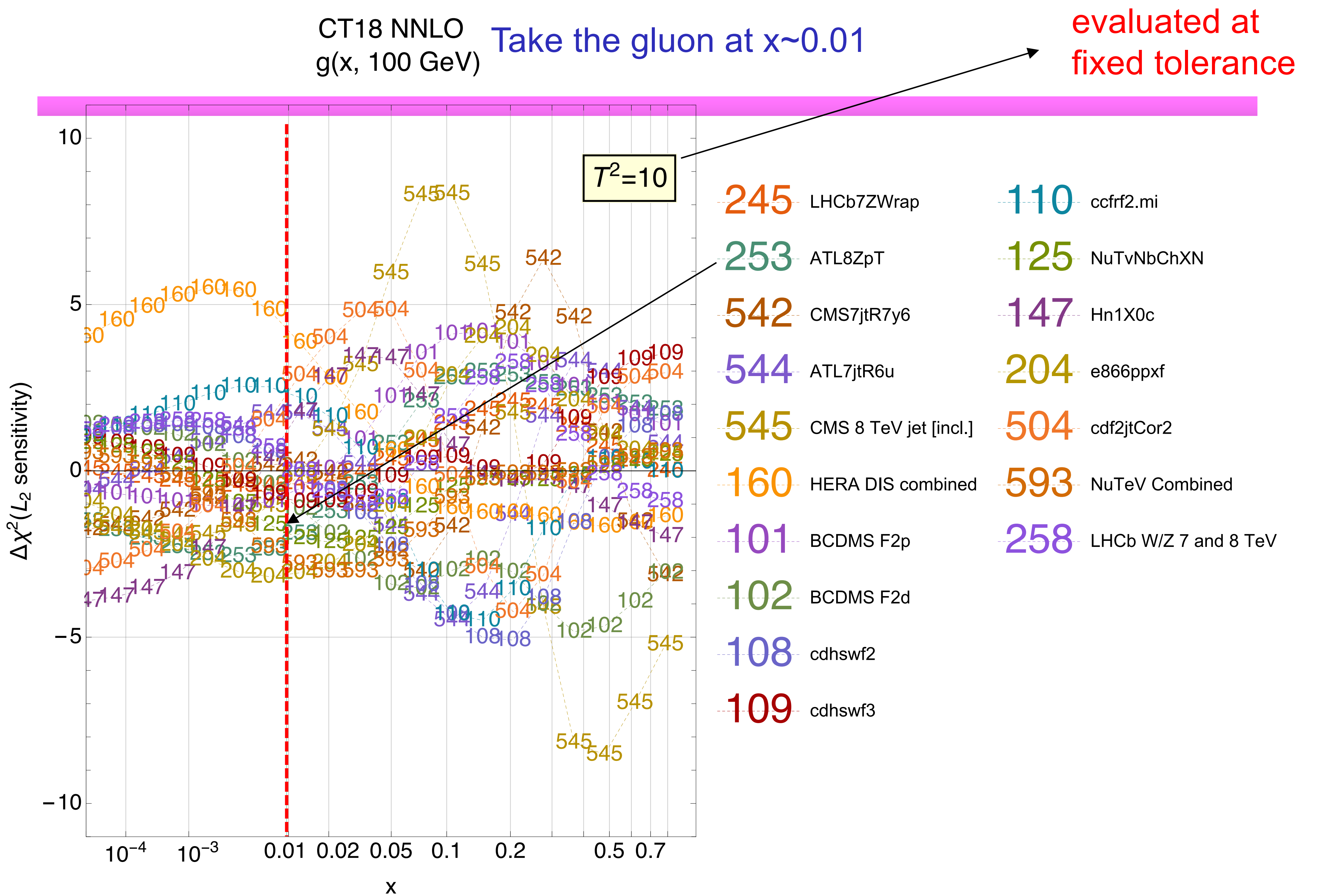
2^{nd} Lagrangian technique

- C^H represents the cosine of the correlation angle between PDF flavor f (or any defined quantity) and experimental χ^2

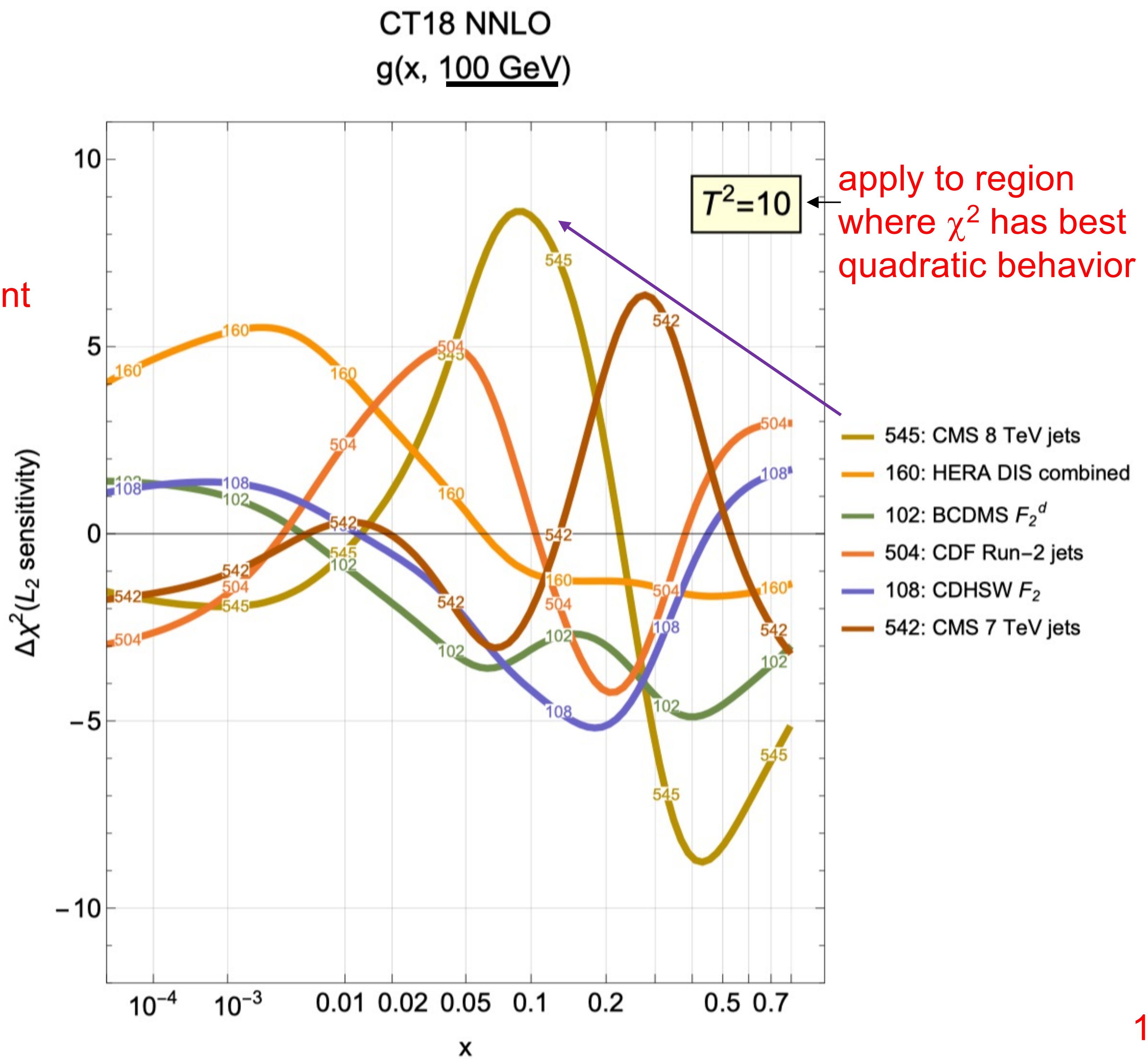


The importance of an experiment for a particular PDF depends not only on the correlation of the cross section with that PDF, but the degree to which the cross section can determine that PDF.

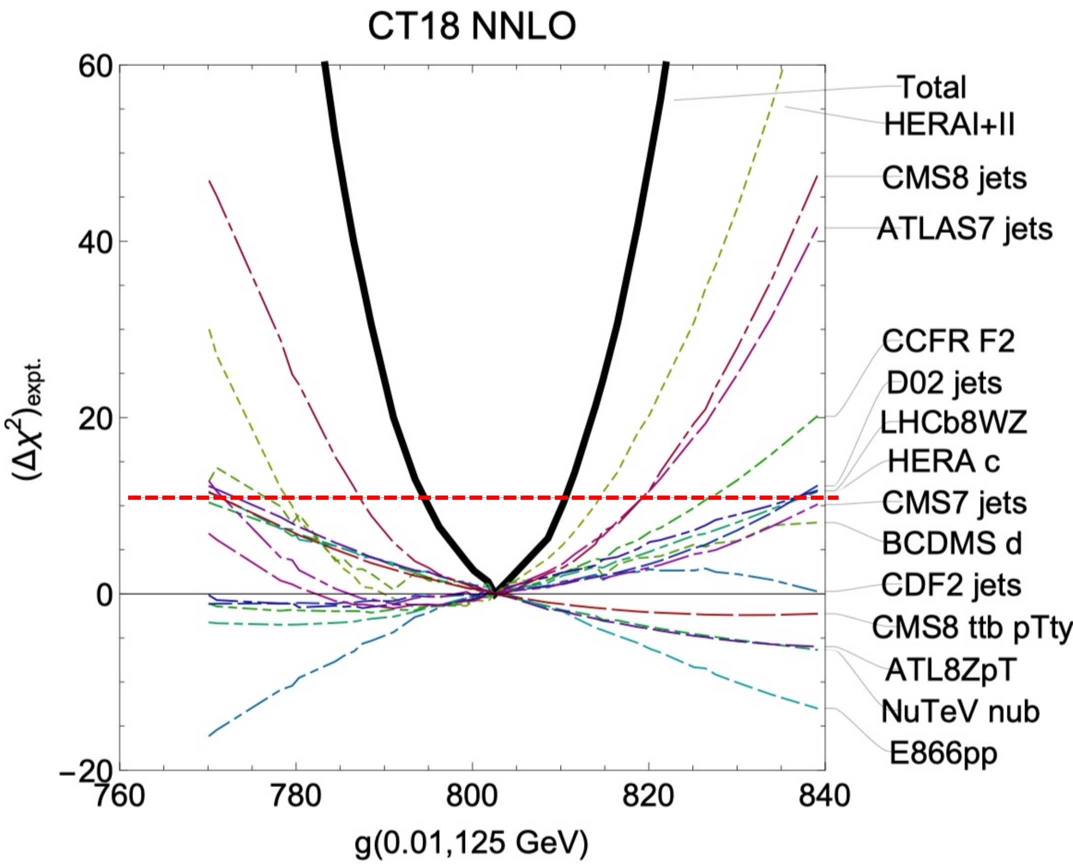
- Can also be defined for the MC PDF approach



HERA DIS wants to pull the gluon down, a number of other experiments want to pull it up show only 6 most important experiments



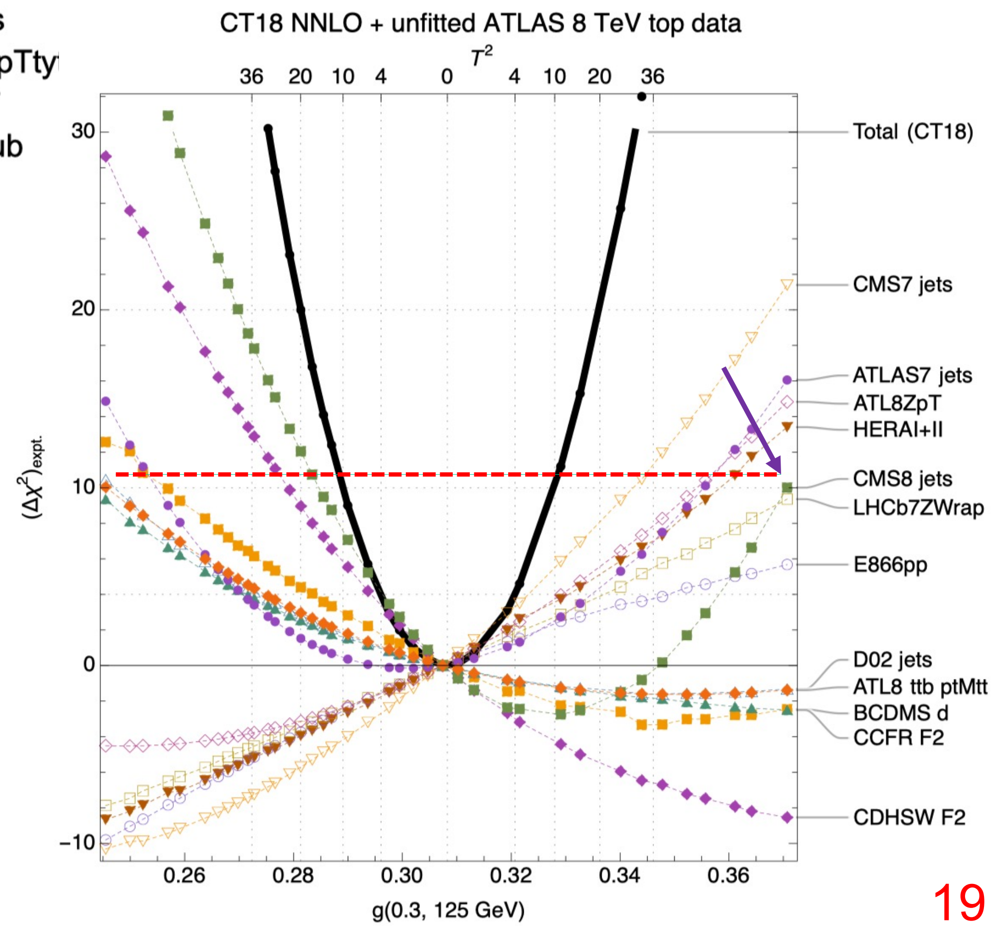
apply to region where χ^2 has best quadratic behavior

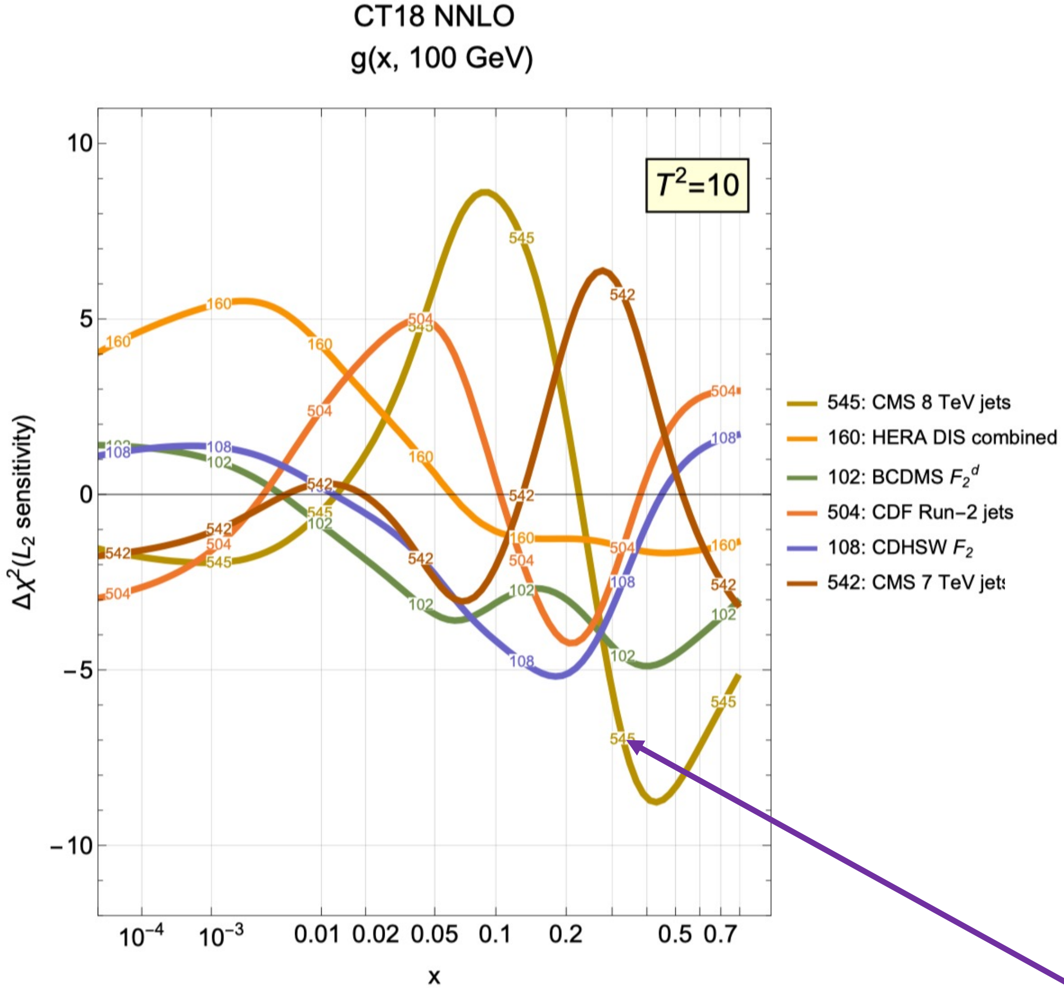


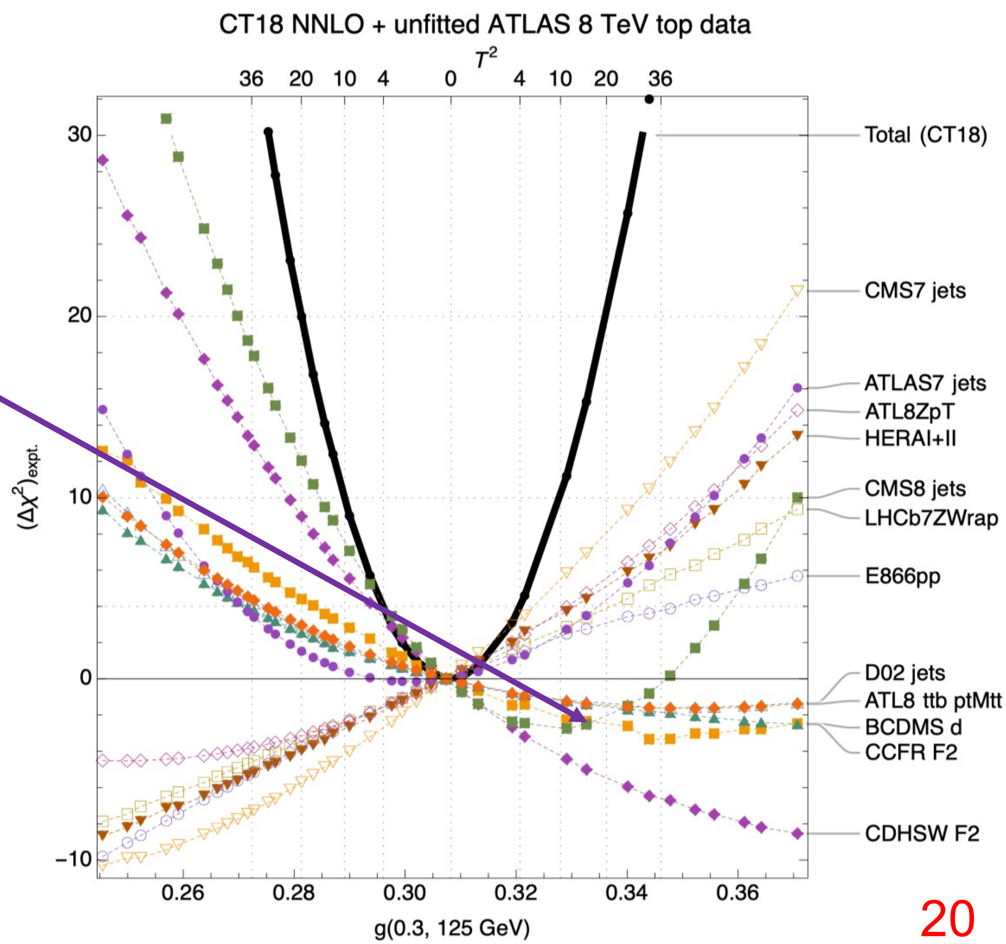
Lagrange Multplier Scans

compare to LM scans of the gluon at x=0.01 and x=0.3

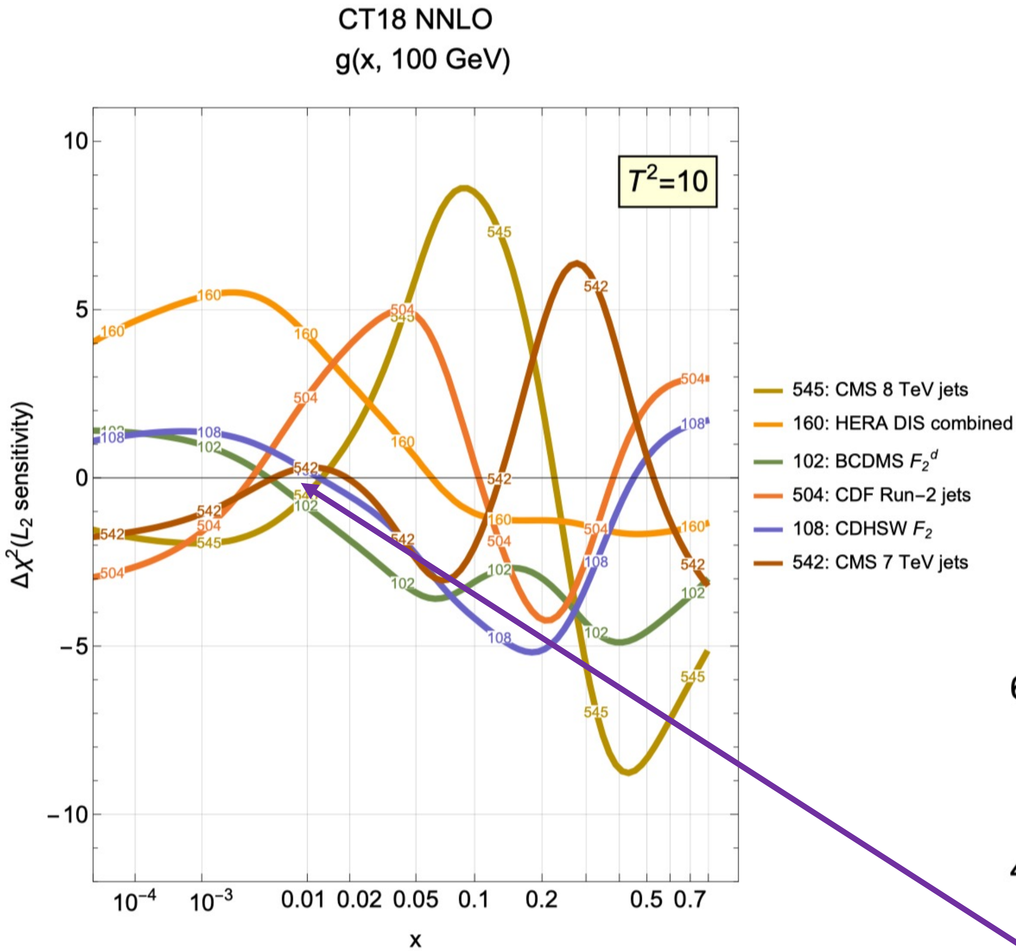
varied preferences for the gluon distribution from the different data sets, but the net results are reasonably parabolic

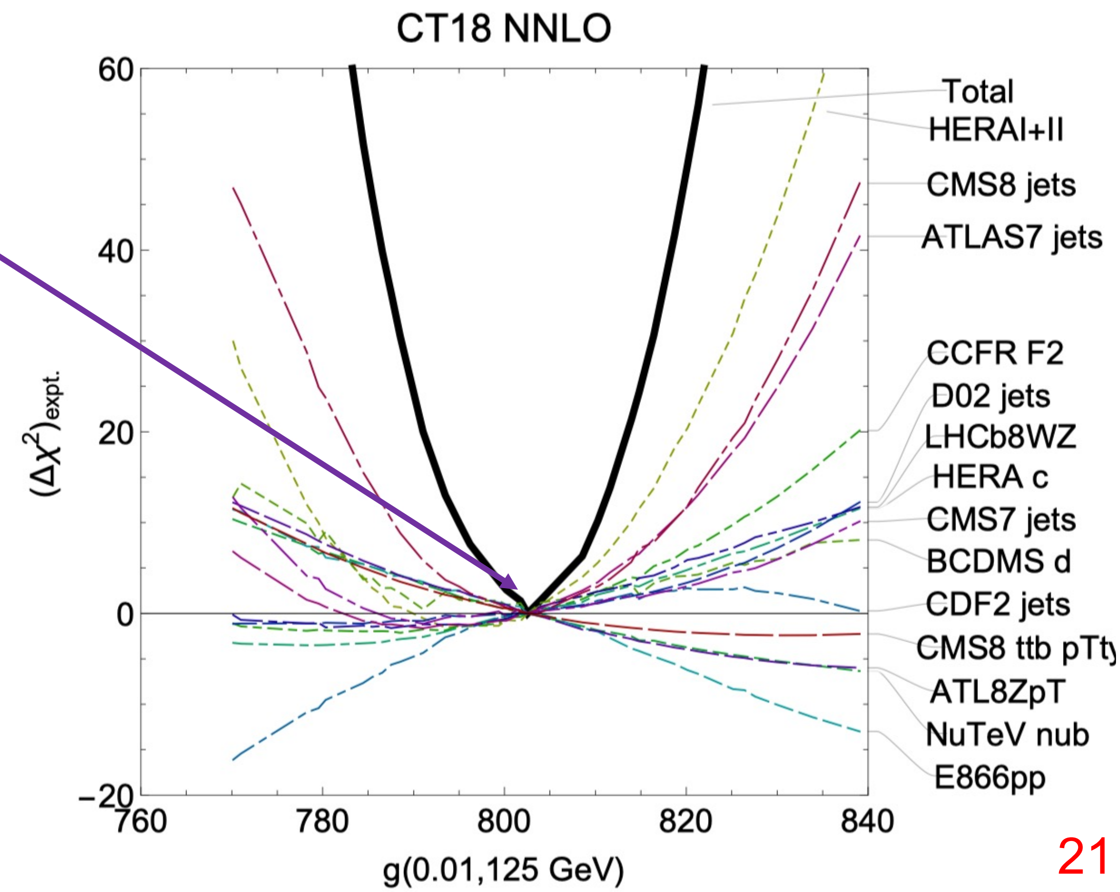






The CMS 8 TeV jet data prefer a harder gluon at x=0.3, but are outvoted by the combination of the other experiments

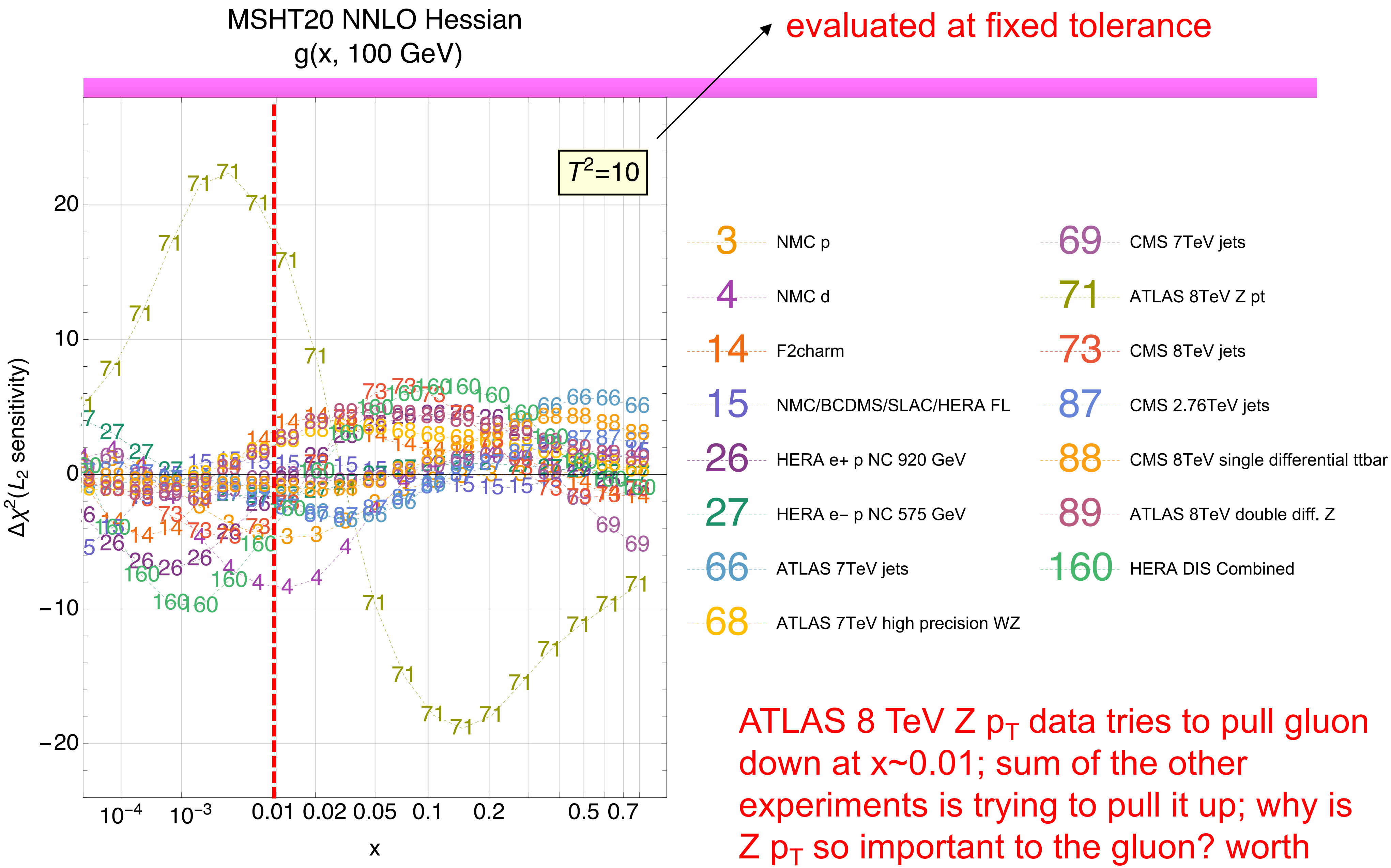




At x=0.01, the CMS 8 TeV jet data have an L2 sensitivity near zero. Why, given the precision of that data?

Because they've already won.

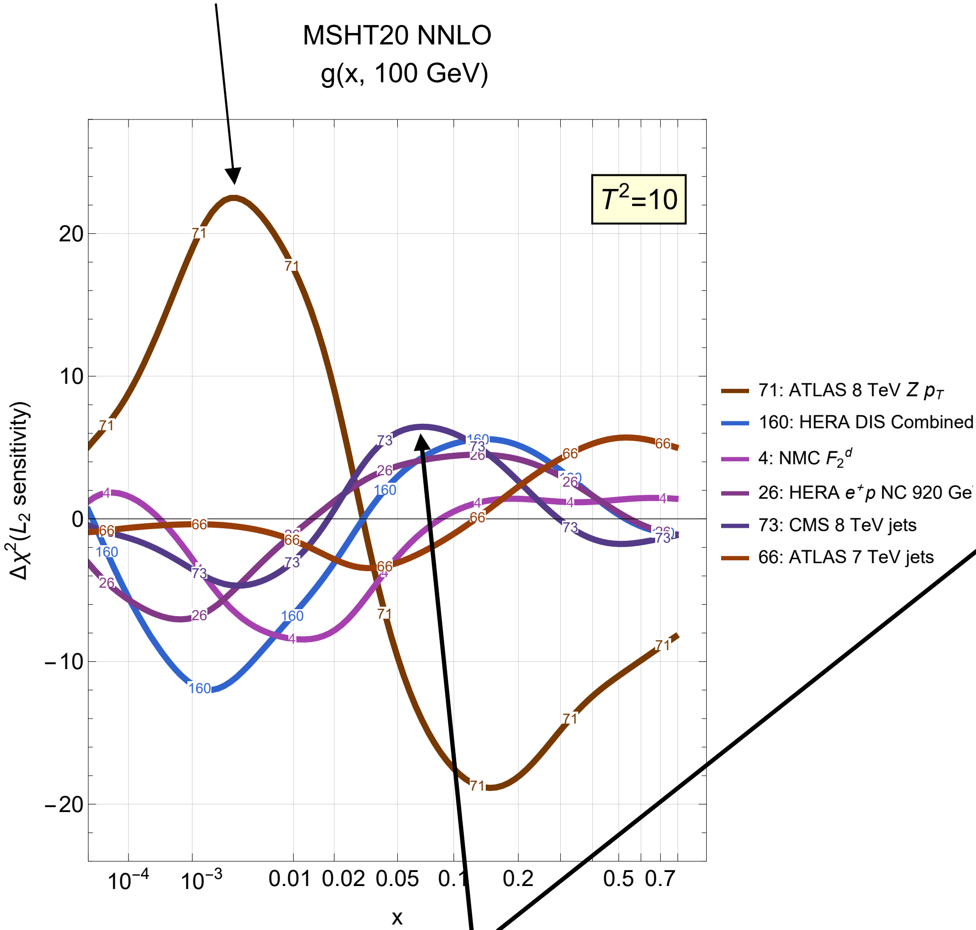
evaluated at fixed tolerance

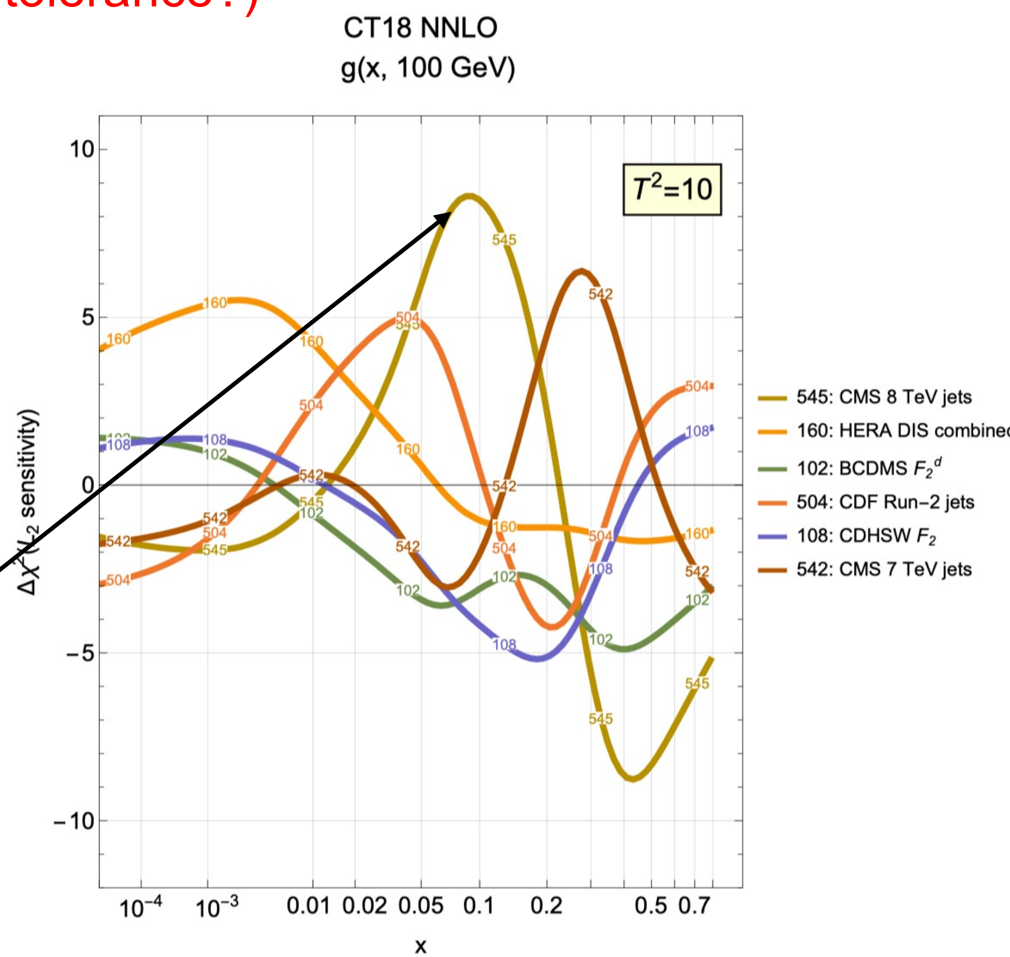


ATLAS 8 TeV Z p_T data tries to pull gluon down at x~0.01; sum of the other experiments is trying to pull it up; why is Z p_T so important to the gluon? worth investigating further

MSHT20 and CT18

Note importance of ATLAS Z p_T data
(also, Z p_T data poorly fit at NNLO; dynamic tolerance?)





CMS 8 TeV jet data play a similar role as in CT18

ATLAS Z p_T not one of 6 most important experiments (more restrictive kinematic region)

aN3LO PDFs: the prequel

- Previous steps in this direction by looking at theory uncertainties in NNLO (by definition probing N3LO effects) cross sections in global PDF fits
- NNPDF->tie factorization scales together and renormalization scales of similar processes; vary scales over a reasonable range and examine impact on PDF fit

  Eur.Phys.J.C 79 (2019) 11, 931

  e-Print: 1906.10698

- MSHT->tie to physical cross sections

  Eur.Phys.J.C 79 (2019) 3, 225

  e-Print: 1811.08434

"To do this, we use the fact that a PDF fit may be recast in a physical basis, where the PDFs themselves are bypassed entirely, and one instead relates measured observables to predicted ones."

What do we need to know for N3LO PDFs?

- Need 4 ingredients. Current Knowledge (schematic summary):

Theory	Utility	Order required	What's known?
1. Splitting functions $P_{ab}^{(3)}(x)$	PDF evolution	4-loop	Mellin moments[3–5], leading small-x behaviour[3,6–11], plus some leading large-x in places[3]
2. Transition matrix elements $A_{ab,H}^{(3)}(x)$	Transitions between number of flavours in PDFs at mass thresholds	3-loop	Mellin moments[12], leading small-x behaviour[13–14], plus some leading large-x in places[14,15].
3. Coefficient functions (NC DIS) $C_{H,a}^{VF,(3)}$	Combine with PDFs and Transition Matrix Elements to form Structure Functions (NC DIS)	N3LO	Some approximations to FFNS (low Q^2) coefficient functions at α_S^3 (with exact LL pieces at low x, NLL unknown)[13,16–17], ZM-VFNS (high Q^2) N3LO coefficient functions known exactly[18]. Therefore GM-VFNS interpolation not completely known.
4. Hadronic Cross-sections (K-factors)	Determine cross-sections at N3LO	N3LO	Very little (none in usable form for PDFs)

- None of these are completely known, but a lot of information already.
- How to construct approximate N3LO PDFs given theory info. not fully known? Include known info. + theory nuisance parameters.

How can we incorporate N3LO knowledge into PDFs?

- Consider usual PDF fit probability:

Theory, Data, Hessian matrix - contains uncorrelated (s_k) and correlated uncertainties (β_k)

$$P(T|D) \propto \exp(-\chi^2) \propto \exp(-\frac{1}{2}(T-D)^T H_0 (T-D))$$

Experimental Nuisance parameters

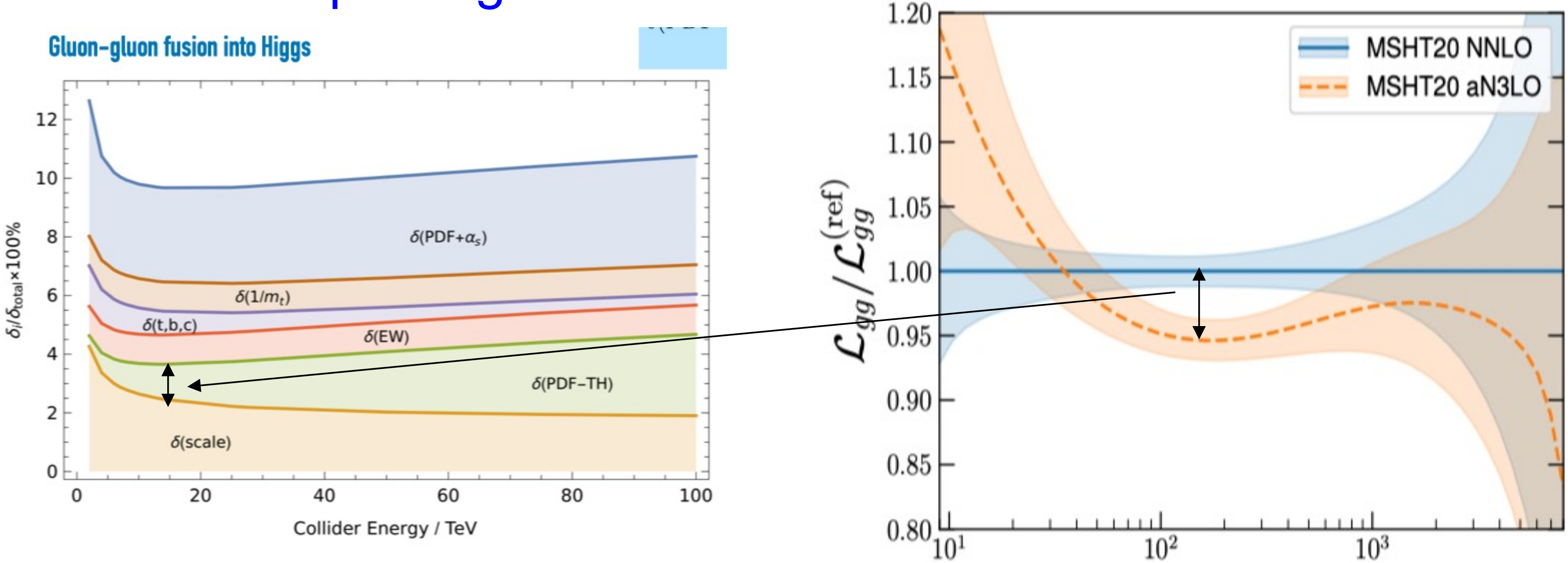
$$\propto \exp(-\frac{1}{2}\sum_{k=1}^{N_{pt}} \frac{1}{s_k^2}(D_k - T_k - \sum_{\alpha=1}^{N_{corr}} \beta_{k,\alpha}\lambda_\alpha)^2 + \sum_{\alpha=1}^{N_{corr}} {\lambda_\alpha}^2)$$

- Include known N3LO pieces (tu) + parameterise remaining unknown pieces ⇒ theory nuisance parameters (θ').
- Now theory $T' = T + tu + (\theta - t)u = T_0' + \theta' u$, i.e. use known info. to shift theory to N3LO central value then allow to vary by θ'.

- Why this approach and theory nuisance parameters (TNPs):
  1. TNPs probe precisely the missing pieces, not lower orders.
  2. Allow inclusion of known N3LO pieces without risk of MHOU probing known info.
  3. Can be updated as and when new N3LO info is available.
  4. Scale variations in PDF fit and predictions need to be correlated.

Impact of aN3LO

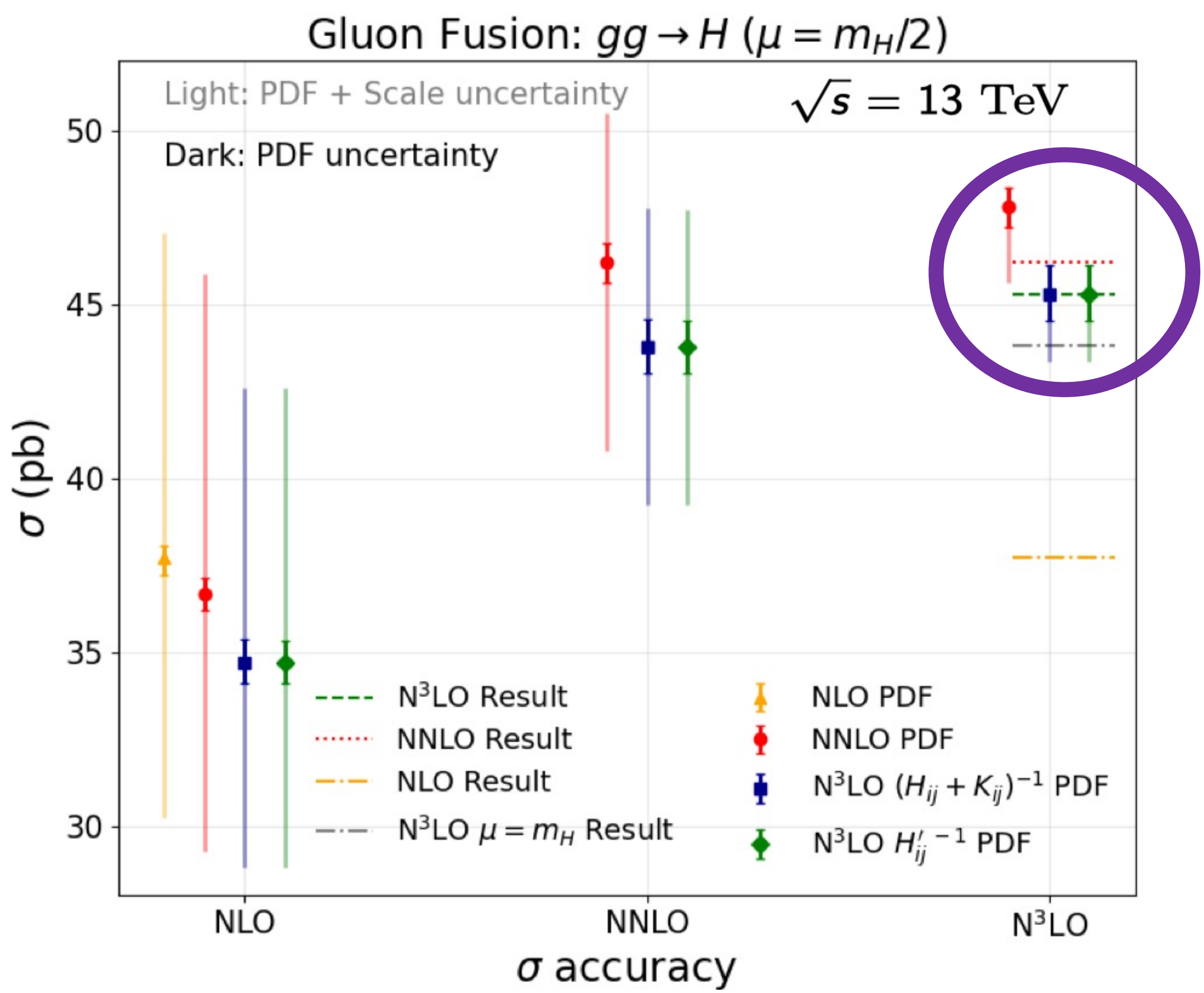
- gg PDF luminosity at aN3LO at Higgs mass ~5% lower than nominal NNLO MSHT20 (large impact from P_{gg})
- If correct, then our benchmark cross sections for ggF would need updating



- How robust are the aN3LO PDFs, and in particular the splitting functions?

…but on the bright side

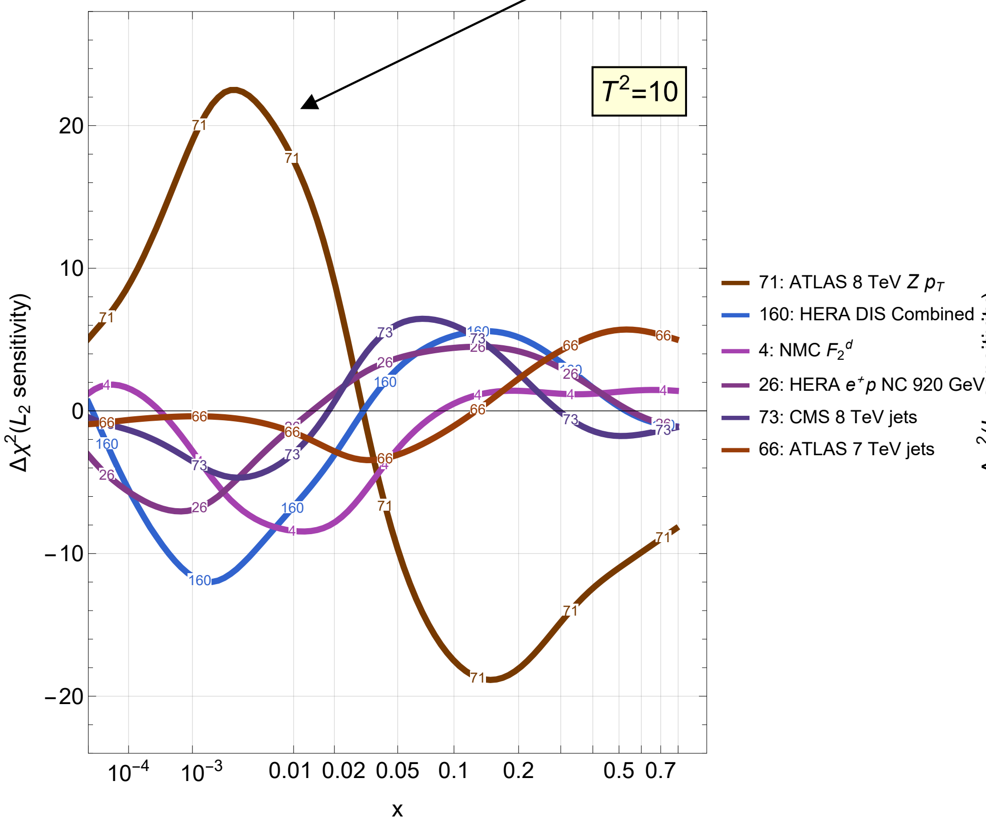
- This would mean that the ggF cross section is more convergent



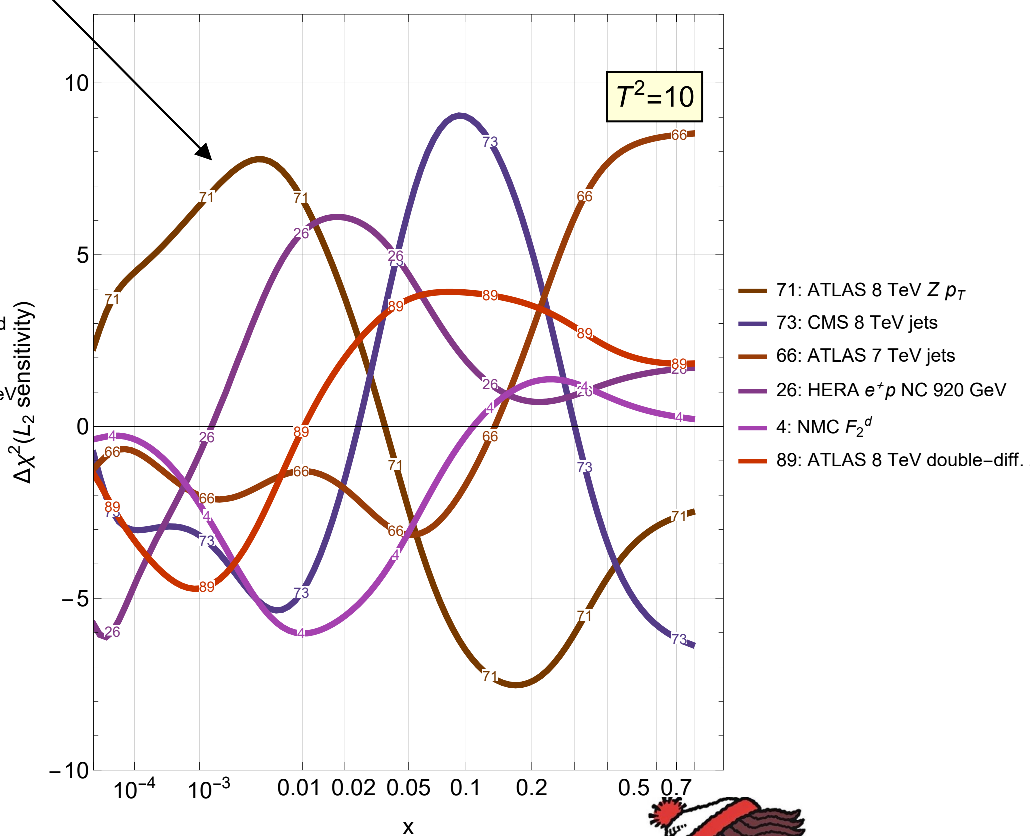
MSHT20 NNLO and aN3LO

shape of L_2 sensitivity similar for two PDFs, but absolute value of Z p_T decreased by almost a factor of 3; significant change in low x gluon

MSHT20 NNLO
g(x, 100 GeV)



MSHT20 approx. N3LO
g(x, 100 GeV)

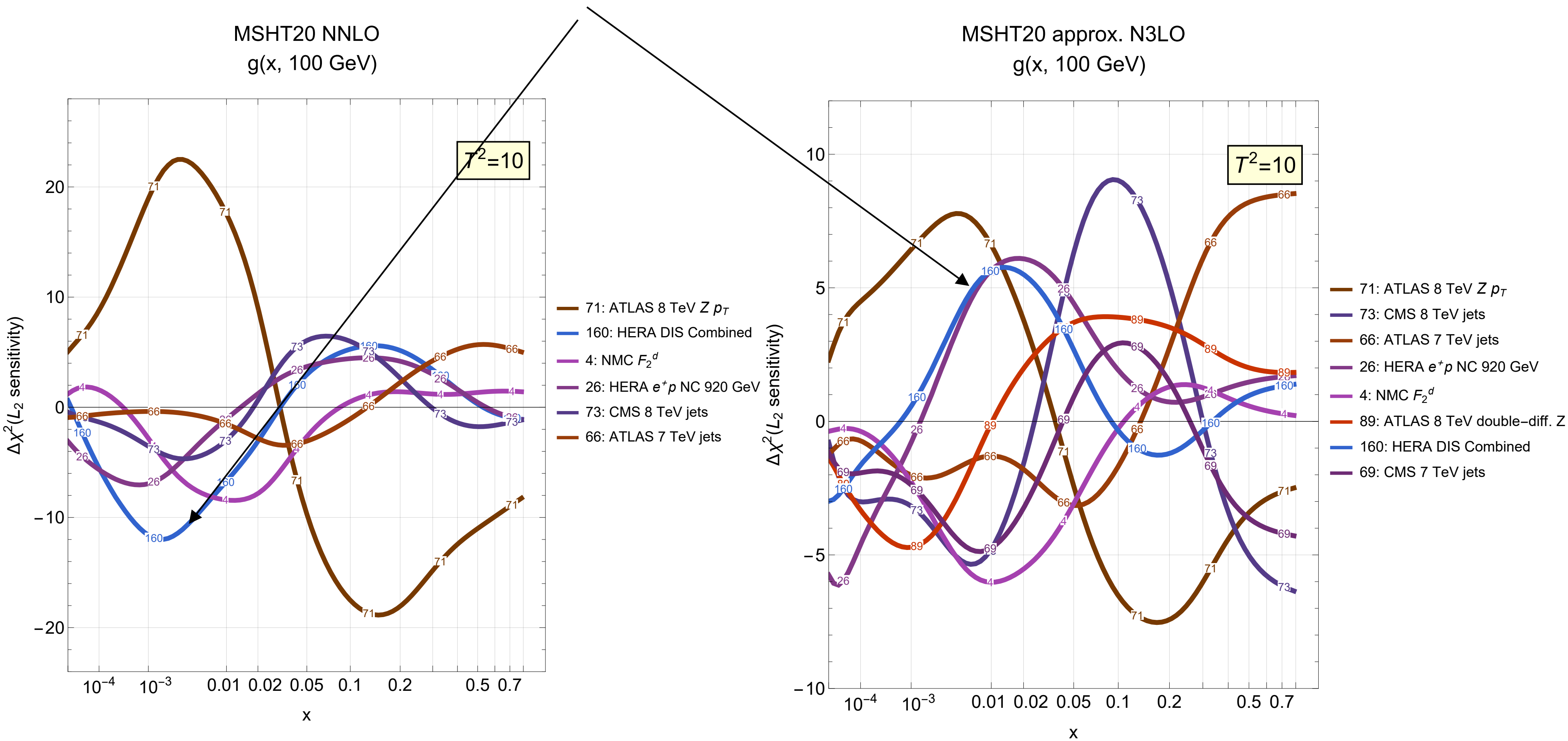


Perhaps not so surprising that a process with large NNLO corrections would have a sensitivity to N3LO corrections

where's expt 160 HERA-DIS?

MSHT20 NNLO and aN3LO

at aN3LO, the two experiments now on same side; aN3LO needed for HERA



160 fell out of the top 6; seemingly the aN3LO gluon released some tensions

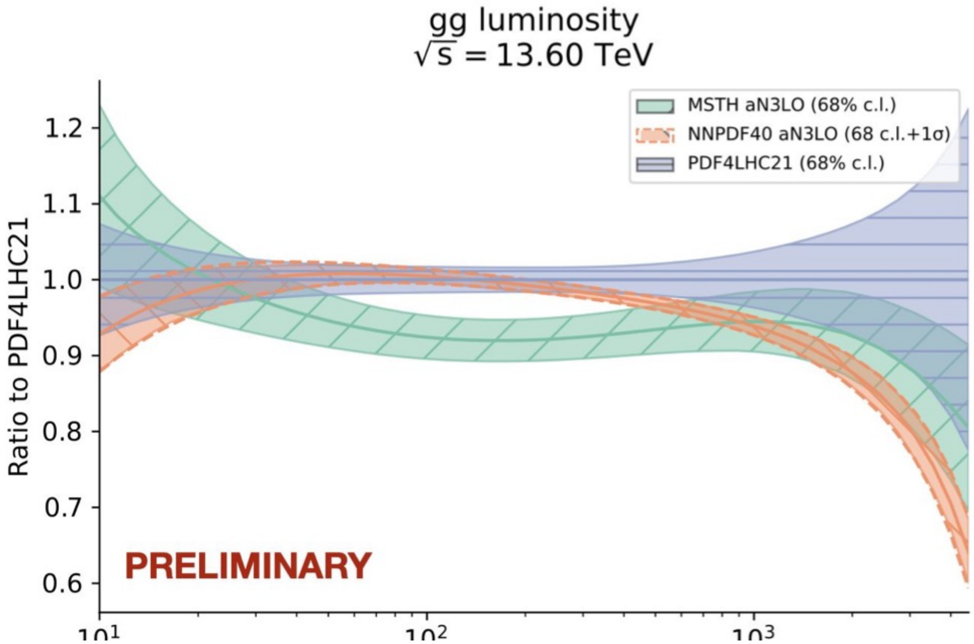
Les Houches 2023

Preliminary results from an aN3LO analysis from NNPDF presented by Stefano Forte

#a3nlo-pdf-ggh

aN3LO PDFs & gluon-fusion Higgs production (VI)

Some differences between aN3LO sets by **MSHT** & **NNPDF** in gg luminosity



Followup studies:

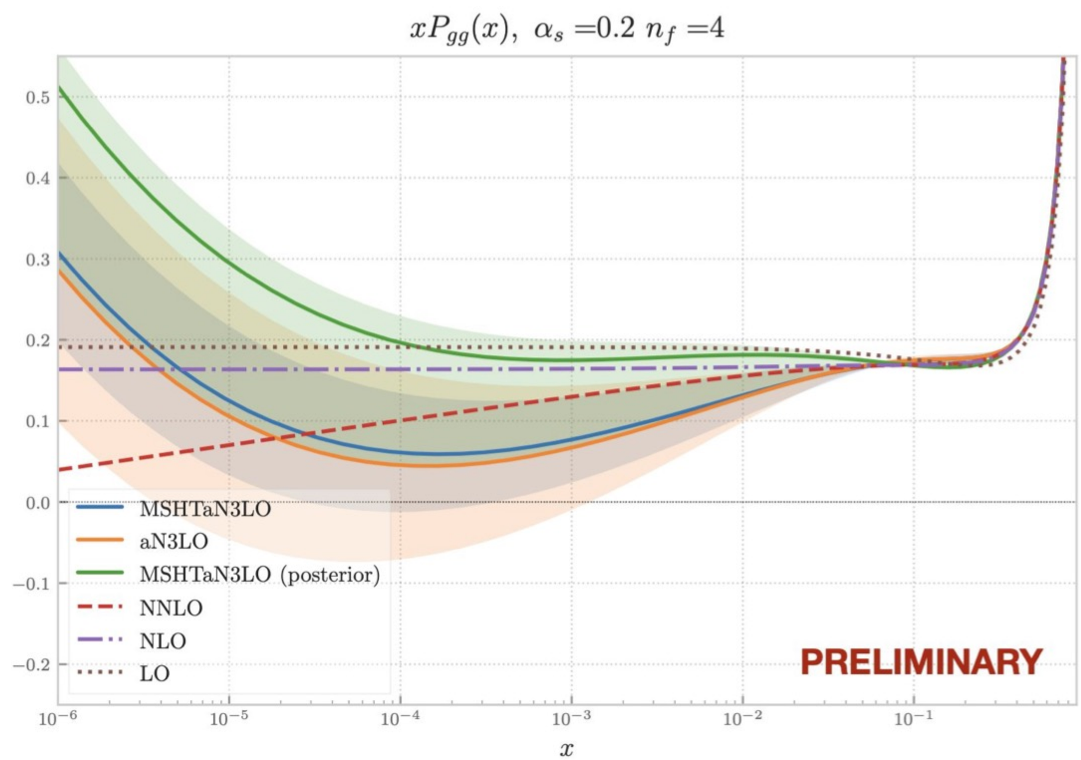
- Understand origin of differences (impact from prior → posterior? treatment of MHO uncertainties & other N3LO inputs? difference in methodology? …)
- Compare evolution of toy PDFs
- Cross-section level comparisons
- …

It’s mostly about the splitting functions

#a3nlo-pdf-ggh

aN3LO PDFs & gluon-fusion Higgs production (IV)

With two independent aN3LO sets, a more detailed look into approximated splitting functions



MSHT (prior) ≃ **NNPDF**

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

…it is closer to low-x resummed gluon distributions

…but wait, there's more!

Four-loop results

10 moments now calculated for singlet
Sven Moch QCD seminar CERN June 23

- Moments $N = 2, \ldots 20$ for pure-singlet anomalous dimension $\gamma_{\rm ps}^{(3)}(N)$

$$\begin{aligned}
\gamma_{\rm ps}^{(3)}(N=2) &= -691.5937093\, n_f + 84.77398149\, n_f^2 + 4.466956849\, n_f^3\,, \\
\gamma_{\rm ps}^{(3)}(N=4) &= -109.3302335\, n_f + 8.776885259\, n_f^2 + 0.306077137\, n_f^3\,, \\
\gamma_{\rm ps}^{(3)}(N=6) &= -46.03061374\, n_f + 4.744075766\, n_f^2 + 0.042548957\, n_f^3\,, \\
\gamma_{\rm ps}^{(3)}(N=8) &= -24.01455020\, n_f + 3.235193483\, n_f^2 - 0.007889256\, n_f^3\,, \\
&\ldots \\
\gamma_{\rm ps}^{(3)}(N=20) &= -0.442681568\, n_f + 0.805745333\, n_f^2 - 0.020918264\, n_f^3\,.
\end{aligned}$$

- Results $N \leq 8$ agree with inclusive DIS S.M., Ruijl, Ueda, Vermaseren, Vogt '21 (also for $N = 10$ and $N = 12$)
- Quartic color terms $d_R^{abcd} d_R^{abcd}$ agree with S.M., Ruijl, Ueda, Vermaseren, Vogt '18
- Large-n_f parts agree with all-N results Davies, Vogt, Ruijl, Ueda, Vermaseren '17;
- ζ_4 terms in $\gamma_{\rm ps}^{(3)}(N)$ agree with Davies, Vogt '17 based on no-π^2 theorem Jamin, Miravitllas '18; Baikov, Chetyrkin '18
- Renormalization constants involving alien operators (required to three loops) agree with Gehrmann, von Manteuffel, Yang '23

.

Approximations in x-space

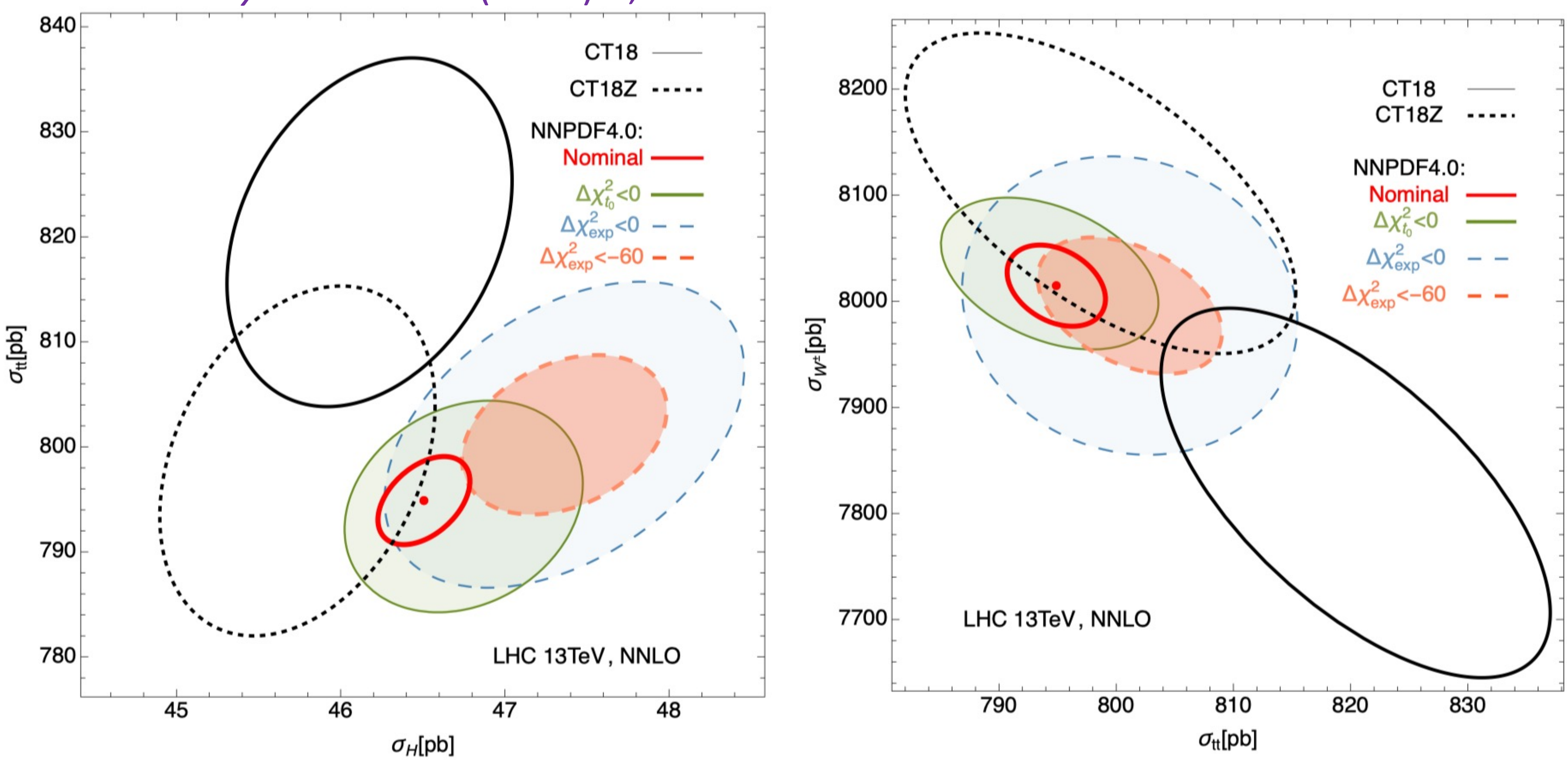
- Large- and small-x information about $P^{(3)}_{\rm gq}(x)$ and $P^{(3)}_{\rm gg}(x)$
  - leading logarithm $(\ln^3 x)/x$ Fadin, Kuraev, Lipatov '75; Balitsky, Lipatov '78
    next-to-leading logarithm $(\ln^2 x)/x$ for $P^{(3)}_{\rm gg}(x)$ Fadin, Lipatov '98
  - sub-dominant logarithms $\ln^k x$ with $k = 6, 5, 4$ Davies, Kom, S.M., Vogt '22
  - leading large-x terms for $P^{(3)}_{\rm gg}(x)$
    $$P^{(3)}_{gg}(x) \;=\; \frac{A_{4,\rm g}}{(1-x)_+} + B_{4,\rm g}\,\delta(1-x) + C_{4,\rm g}\ln(1-x) + D_{4,\rm g}$$
  - sub-leading large-x terms $(1-x)^j \ln^k(1-x)$ with $j \geq 1$ and $k \leq 4$
    with $k = 5, 4$ known Soar, S.M., Vermaseren, Vogt '09
- Approximations of four-loop splitting function with suitable ansatz
  - main uncertainity unknown leading small-x terms: $(\ln x)/x$, $1/x$

Now

- Approximations for $P^{(3)}_{\rm gq}(x)$ and $P^{(3)}_{\rm gg}(x)$ based on moments $N = 2, \ldots 10$
  - higher moments $N = 12, \ldots 20$ with improved accuracy to come

Hopscotch scans: approximate error ellipses

Phys.Rev.D 107 (2023) 3, 034008 arXiv:2205.10444



There are PDF solutions with equal or better χ^2 than present in the nominal NNPDF4.0 uncertainty ellipses. Do these mean anything? Some would be rejected because they are too wiggly, or are too close to going negative.

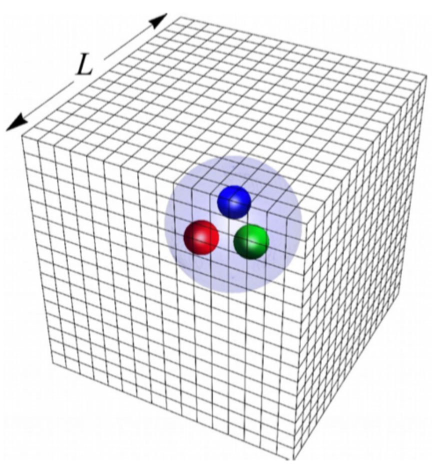
This is a selection that can affect the allowed uncertainty for the PDFs and may need to be further studied in order to better understand Hessian vs Monte Carlo.

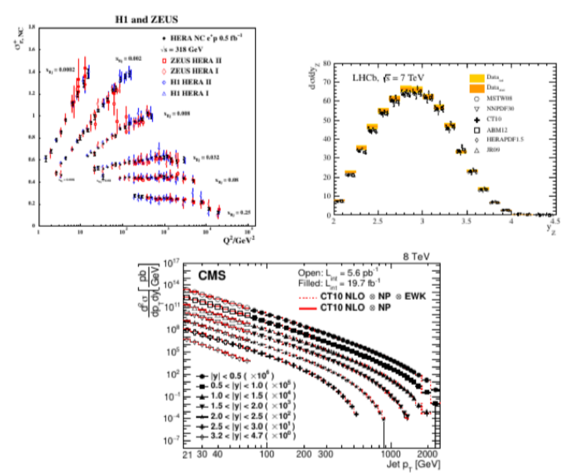
Can the lattice save us?

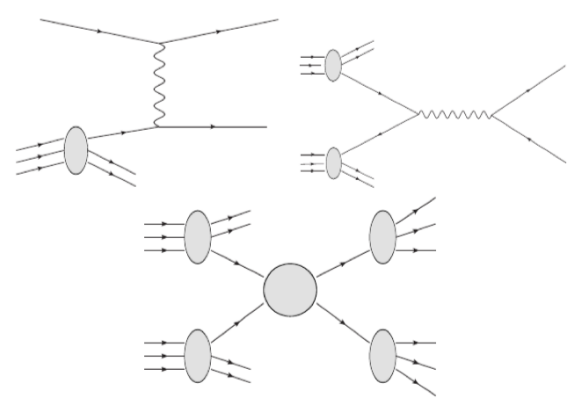
…as it does for the determination of α_s

What about incorporating LQCD into global analysis?

$$\sigma = f_a(x_1, Q^2) \otimes f_b(x_2, Q^2) \otimes \hat{\sigma}$$

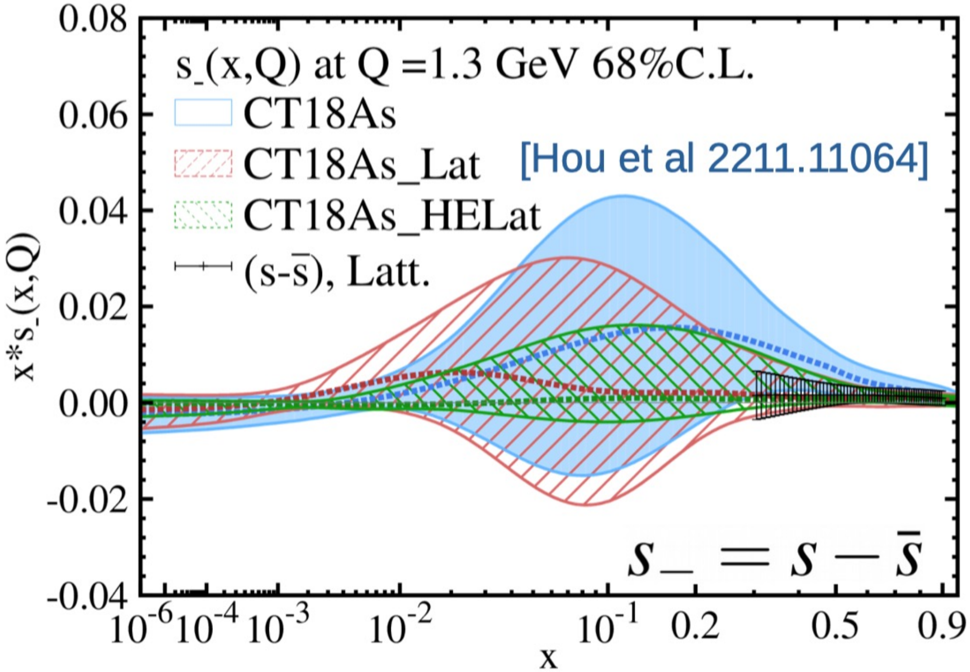






T.J. Hou DIS2023

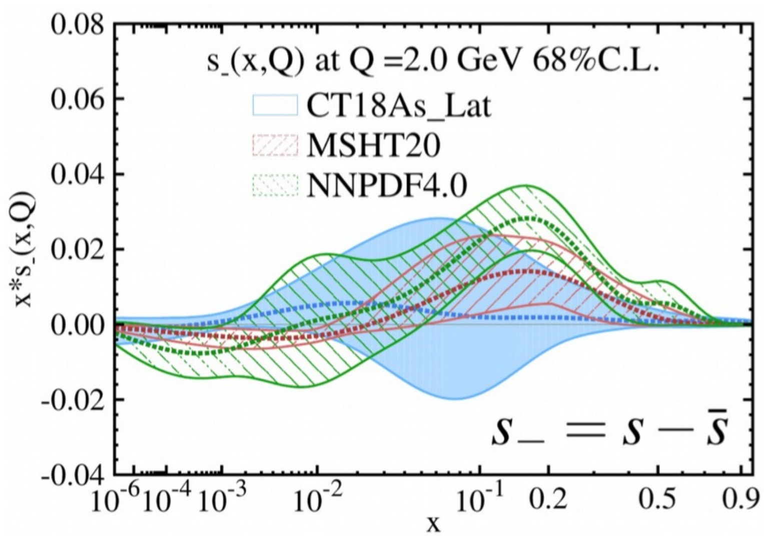
…in general, not yet



CT18As: CT18A with strangeness asymmetry at Q_0 = 1.3 GeV.

CT18As_Lat: PDFs with lattice input.

CT18As_HELat: PDFs with the lattice errors reduced by half.



- Lattice QCD calculation provide prediction at 0.3 < x < 0.8, while the di-muon data constraint strangeness at 0.015 < x < 0.336.
- Lattice input improves the determination of strangeness asymmetry.
- LQCD can improve heavy flavor decomposition.

For other heavy flavor topics, i.e. charm, see talk by Pavel Nadolsky

Summary

- Determination of central PDF values and of uncertainties has come a long way
- A great deal of LHC data has made it into the global PDF fits, with much more to be added at 13 TeV, and now, 13.6 TeV
- There is still work to do to provide a more rigorous understanding, especially of the different techniques used to determine the central values and the errors on PDFs
- Paradoxically, increasing the data sample and the parametric space **may** increase the sample expectation deviation
- New PDF tools can help us to better understand exactly how the PDF fits are formed, from both the experimental side and the theory side
- N3LO is the frontier, and will remain the frontier for some time; so far the only matrix elements known, used in PDF fits, are for Drell-Yan
- However, for the gluon, the dominant impact should be from the P_{gg} splitting functions and those are closer to a better understanding with the new moments calculated

https://metapdf.hepforge.org/L2

- **2023: L2 sensitivities** for ATLAS21, CT18, MSHT20 PDFs
- **mcgen**
- **META PDFs**
  - **PDF4LHC15 gallery**
- PDFSense tool and results:
  - **Maps of sensitivities**
  - **Quasi-PDFs and Mellin moments**

L2 sensitivities for global QCD analyses

Constraints on parton distributions and their combinations

This website collects figures of L2 sensitivities for experimental data sets obtained for ATLAS21 NNLO, CT18/CT18As/CT18As_Lat NNLO, and MSHT20 NNLO and approximate N3LO global QCD analyses.

Citation policy

If you use the programs or results from this website, please cite

Quantifying the interplay of experimental constraints in analyses of parton distributions

X. Jing, A. Cooper-Sarkar, A. Courtoy, T. Cridge, F. Giuli, L. Harland-Lang, T. J. Hobbs, J. Huston, P. Nadolsky, K. Xie, R. S. Thorne, and C.-P. Yuan

arXiv:2306.03918

L2 sensitivities can be computed using a C++ program **L2LHAexplorer** and Hessian PDFs from the LHAPDF library. See 00README included in the .zip file. Alternatively, a Mathematica notebook to plot the L2 sensitivities collected on this website can be requested.

The L2 sensitivities can be plotted in two styles:

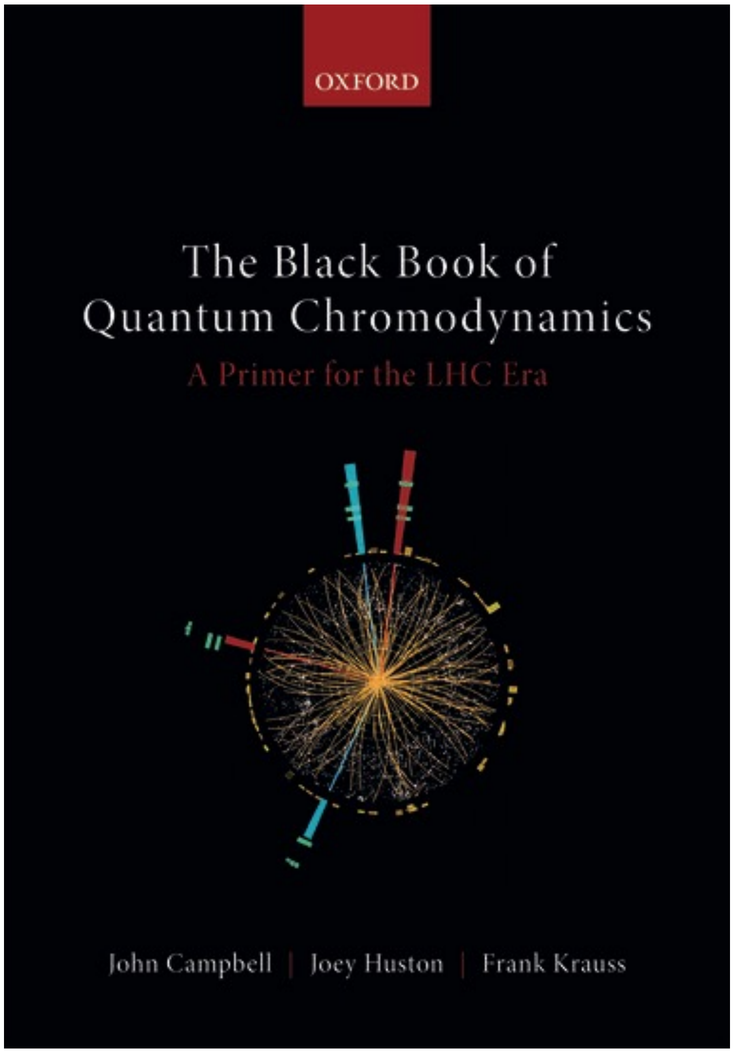
- **L2 sensitivities for a chosen PDF flavor or PDF combination and the most sensitive fitted experiments**
- **L2 sensitivities for a chosen fitted experiment and all PDF flavors or several PDF combinations**

Support: **Pavel Nadolsky**

…now available as free download

OAPEN

https://library.oapen.org › 9780199652747_Print



Extras

TABLE I. Datasets included in the CT18(Z) NNLO global analyses. Here we directly compare the quality of fit found for CT18 NNLO vs CT18Z NNLO on the basis of χ^2_E, $\chi^2_E/N_{pt,E}$, and S_E, in which $N_{pt,E}$, χ^2_E are the number of points and value of χ^2 for experiment E at the global minimum. S_E is the effective Gaussian parameter [38,42,56] quantifying agreement with each experiment. The ATLAS 7 TeV 35 pb^{-1} W/Z dataset, marked by ‡‡, is replaced by the updated one (4.6 fb^{-1}) in the CT18A and CT18Z fits. The CDHSW data, labeled by †, are not included in the CT18Z fit. The numbers in parentheses are for the CT18Z NNLO fit.

Exp. ID#	Experimental dataset		$N_{pt,E}$	χ^2_E	$\chi^2_E/N_{pt,E}$	S_E
160	HERAI + II 1 fb^{-1}, H1 and ZEUS NC and CC $e^{\pm}p$ reduced cross sec. comb.	[30]	1120	1408 (1378)	1.3 (1.2)	5.7 (5.1)
101	BCDMS F_2^p	[57]	337	374 (384)	1.1 (1.1)	1.4 (1.8)
102	BCDMS F_2^d	[58]	250	280 (287)	1.1 (1.1)	1.3 (1.6)
104	NMC F_2^d/F_2^p	[59]	123	126 (116)	1.0 (0.9)	0.2 (−0.4)
108†	CDHSW F_2^p	[60]	85	85.6 (86.8)	1.0 (1.0)	0.1 (0.2)
109†	CDHSW $x_B F_3^p$	[60]	96	86.5 (85.6)	0.9 (0.9)	−0.7 (−0.7)
110	CCFR F_2^p	[61]	69	78.8 (76.0)	1.1 (1.1)	0.9 (0.6)
111	CCFR $x_B F_3^p$	[62]	86	33.8 (31.4)	0.4 (0.4)	−5.2 (−5.6)
124	NuTeV $\nu\mu\mu$ SIDIS	[63]	38	18.5 (30.3)	0.5 (0.8)	−2.7 (−0.9)
125	NuTeV $\bar{\nu}\mu\mu$ SIDIS	[63]	33	38.5 (56.7)	1.2 (1.7)	0.7 (2.5)
126	CCFR $\nu\mu\mu$ SIDIS	[64]	40	29.9 (35.0)	0.7 (0.9)	−1.1 (−0.5)
127	CCFR $\bar{\nu}\mu\mu$ SIDIS	[64]	38	19.8 (18.7)	0.5 (0.5)	−2.5 (−2.7)
145	H1 σ_r^b	[65]	10	6.8 (7.0)	0.7 (0.7)	−0.6 (−0.6)
147	Combined HERA charm production	[66]	47	58.3 (56.4)	1.2 (1.2)	1.1 (1.0)
169	H1 F_L	[33]	9	17.0 (15.4)	1.9 (1.7)	1.7 (1.4)
201	E605 Drell-Yan process	[67]	119	103.4 (102.4)	0.9 (0.9)	−1.0 (−1.1)
203	E866 Drell-Yan process $\sigma_{pd}/(2\sigma_{pp})$	[68]	15	16.1 (17.9)	1.1 (1.2)	0.3 (0.6)
204	E866 Drell-Yan process $Q^3 d^2\sigma_{pp}/(dQ dx_F)$	[69]	184	244 (240)	1.3 (1.3)	2.9 (2.7)
225	CDF run-1 lepton A_{ch}, $p_{T\ell} > 25$ GeV	[70]	11	9.0 (9.3)	0.8 (0.8)	−0.3 (−0.2)
227	CDF run-2 electron A_{ch}, $p_{T\ell} > 25$ GeV	[71]	11	13.5 (13.4)	1.2 (1.2)	0.6 (0.6)
234	DØ run-2 muon A_{ch}, $p_{T\ell} > 20$ GeV	[72]	9	9.1 (9.0)	1.0 (1.0)	0.2 (0.1)
260	DØ run-2 Z rapidity	[73]	28	16.9 (18.7)	0.6 (0.7)	−1.7 (−1.3)
261	CDF run-2 Z rapidity	[74]	29	48.7 (61.1)	1.7 (2.1)	2.2 (3.3)
266	CMS 7 TeV 4.7 fb^{-1}, muon A_{ch}, $p_{T\ell} > 35$ GeV	[75]	11	7.9 (12.2)	0.7 (1.1)	−0.6 (0.4)
267	CMS 7 TeV 840 pb^{-1}, electron A_{ch}, $p_{T\ell} > 35$ GeV	[76]	11	4.6 (5.5)	0.4 (0.5)	−1.6 (−1.3)
268‡‡	ATLAS 7 TeV 35 pb^{-1} W/Z cross sec., A_{ch}	[77]	41	44.4 (50.6)	1.1 (1.2)	0.4 (1.1)
281	DØ run-2 9.7 fb^{-1} electron A_{ch}, $p_{T\ell} > 25$ GeV	[78]	13	22.8 (20.5)	1.8 (1.6)	1.7 (1.4)
504	CDF run-2 inclusive jet production	[79]	72	122 (117)	1.7 (1.6)	3.5 (3.2)
514	DØ run-2 inclusive jet production	[80]	110	113.8 (115.2)	1.0 (1.0)	0.3 (0.4)

TABLE II. Like Table I, for newly included LHC measurements. The ATLAS 7 TeV W/Z data (4.6 fb^{-1}), labeled by ‡, are included in the CT18A and CT18Z global fits, but not in CT18 and CT18X.

Exp. ID#	Experimental dataset		$N_{pt,E}$	χ^2_E	$\chi^2_E/N_{pt,E}$	S_E
245	LHCb 7 TeV 1.0 fb^{-1} W/Z forward rapidity cross sec.	[81]	33	53.8 (39.9)	1.6 (1.2)	2.2 (0.9)
246	LHCb 8 TeV 2.0 fb^{-1} $Z \to e^-e^+$ forward rapidity cross sec.	[82]	17	17.7 (18.0)	1.0 (1.1)	0.2 (0.3)
248‡	ATLAS 7 TeV 4.6 fb^{-1}, W/Z combined cross sec.	[39]	34	287.3 (88.7)	8.4 (2.6)	13.7 (4.8)
249	CMS 8 TeV 18.8 fb^{-1} muon charge asymmetry A_{ch}	[83]	11	11.4 (12.1)	1.0 (1.1)	0.2 (0.4)
250	LHCb 8 TeV 2.0 fb^{-1} W/Z cross sec.	[84]	34	73.7 (59.4)	2.1 (1.7)	3.7 (2.6)
253	ATLAS 8 TeV 20.3 fb^{-1}, Z p_T cross sec.	[85]	27	30.2 (28.3)	1.1 (1.0)	0.5 (0.3)
542	CMS 7 TeV 5 fb^{-1}, single incl. jet cross sec., $R = 0.7$ (extended in y)	[86]	158	194.7 (188.6)	1.2 (1.2)	2.0 (1.7)
544	ATLAS 7 TeV 4.5 fb^{-1}, single incl. jet cross sec., $R = 0.6$	[9]	140	202.7 (203.0)	1.4 (1.5)	3.3 (3.4)
545	CMS 8 TeV 19.7 fb^{-1}, single incl. jet cross sec., $R = 0.7$, (extended in y)	[87]	185	210.3 (207.6)	1.1 (1.1)	1.3 (1.2)
573	CMS 8 TeV 19.7 fb^{-1}, $t\bar{t}$ norm. double-diff. top p_T and y cross sec.	[88]	16	18.9 (19.1)	1.2 (1.2)	0.6 (0.6)
580	ATLAS 8 TeV 20.3 fb^{-1}, $t\bar{t}$ p_T^t and $m_{t\bar{t}}$ abs. spectrum	[89]	15	9.4 (10.7)	0.6 (0.7)	−1.1 (−0.8)

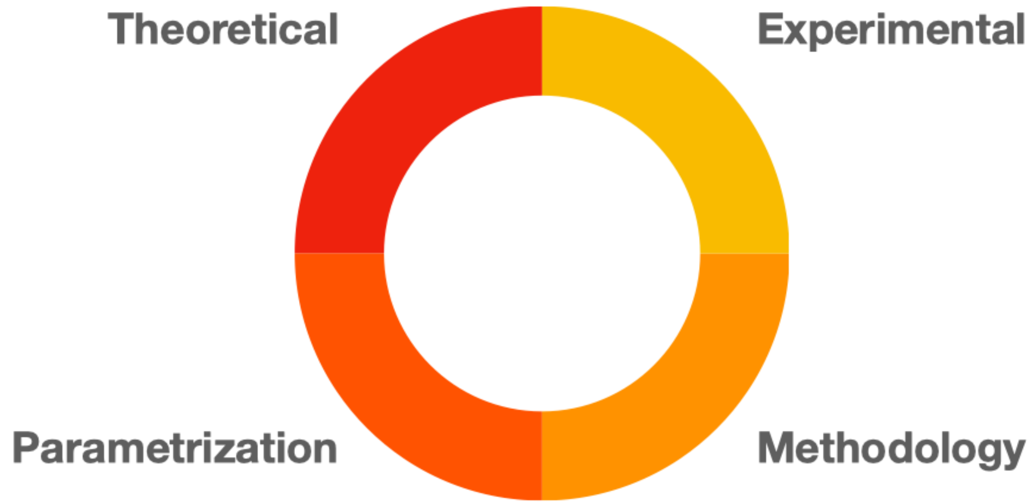
since first derivative of χ^2 vanishes at the global minimum, the sum of the L_2 sensitivities must be zero within uncertainties

$$0 < \sum_E S_{f,L_2} \ll T^2 < \sum_E |S_{f,L_2}|$$

Another too for studying uncertainty: Hopscotch

- Contributions to PDF uncertainties include
  - experimental errors of the data
  - parametrization uncertainties (CT18 uncertainty incorporates effect of trying out hundreds of parametrization forms)
  - theoretical uncertainties/limitations
  - methodology, including <u>sampling accuracy for Monte Carlo fitting</u>
    - the sampling accuracy has typically been ignored
    - ->hopscotch scans

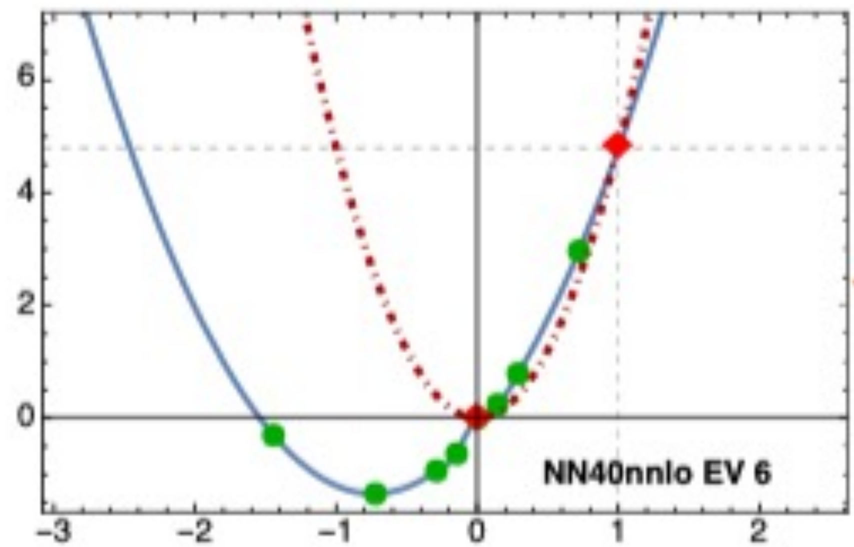
arXiV:2205.10444



Control of sampling biases in the determination of PDFs can play a critical role

Eigenvectors

- Sampling of multi-dimensional spaces (d>>20) can be exponentially inefficient and require $n > 2^d$ replicas for reasonable convergence
- A study of this multi-dimensional space for NNPDF is possible due to the public release of the NNPDF4.0 code
- Use published NNPDF4.0 Hessian basis (n=50), converted from MC replicas
  - total χ^2 of each eigenvector set varies, as large as +35 and as low as -25 (wrt replica 0); the majority no larger than 5-10 units in magnitude; only 1 error set per EV
- Can determine χ^2 at green points, where, for some eigenvectors, lower χ^2 solutions evident (displaced from 0)

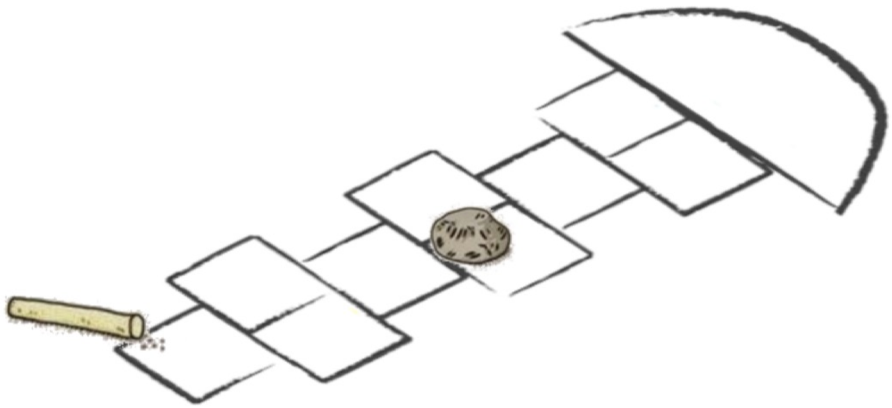


red points correspond to replica 0 and EV6

evaluate χ^2 at 16 points per eigenvector; quadratic behavior observed, i.e. Gaussian uncertainties

Hopscotch scans

- Scan along 50 EV directions to identify a hypercube corresponding to $\Delta\chi^2<T^2$ (T is the tolerance, user-chosen)
- Confirm Gaussian profiles in each eigenvector direction with LM scans



- Concentrate on 4-8 large dimensions in the PDF eigenvector space controlling the large variations of the cross sections under investigation
- Generate replicas varying primarily in these directions; this is not a search for the true global minimum

finding the displaced global minimum in the whole 50-dimensional space is computationally expensive; replica generation is a stochastic exploration; the minimum lies within error ellipses

The sausage-making of α_s

- We (PDG) divide the determinations into 7 categories and take an unweighted fit for each category.
- The 6 non-lattice measurements are then averaged with the lattice average provided by the FLAG group

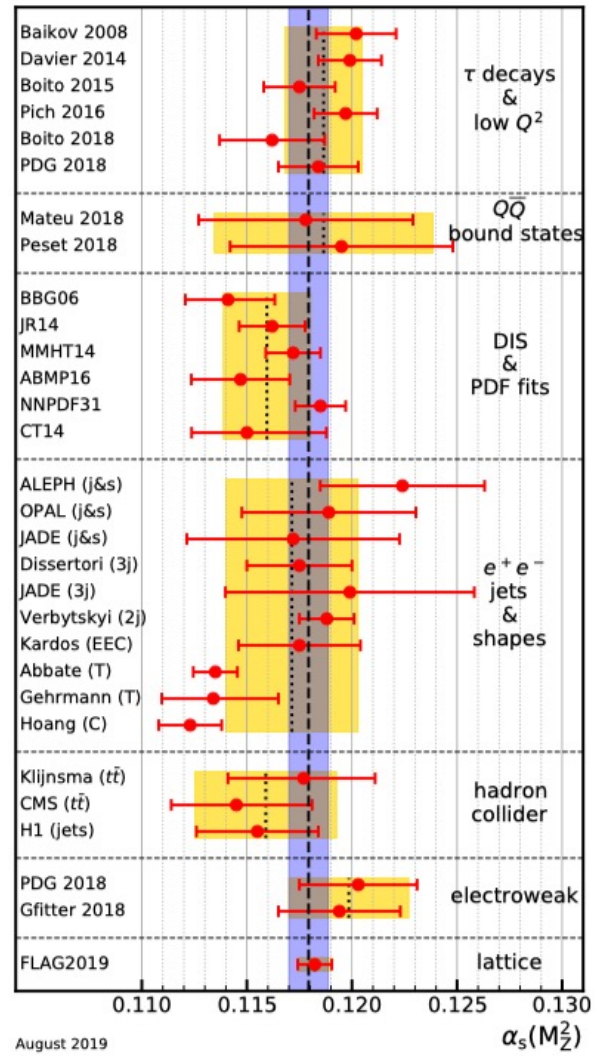
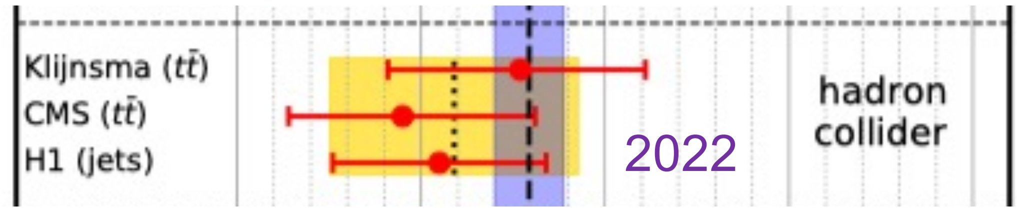


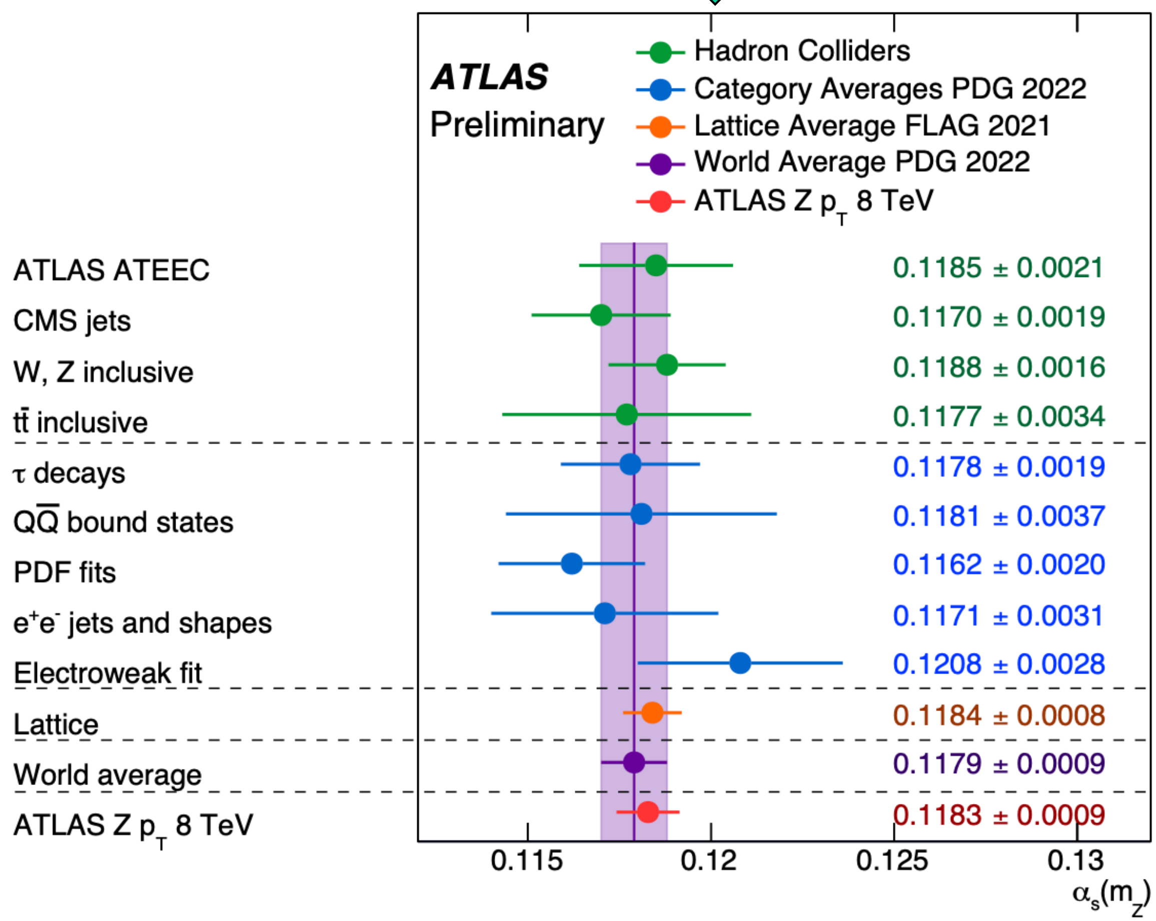
Figure 9.2: Summary of determinations of $\alpha_s(M_Z^2)$ from the seven sub-fields discussed in the text. The yellow (light shaded) bands and dotted lines indicate the pre-average values of each sub-field. The dashed line and blue (dark shaded) band represent the final world average value of $\alpha_s(M_Z^2)$.

Collider measurements of α_s

- As the number of NNLO calculations has increased, there have been a growing number of determinations of $\alpha_s(m_Z)$ at that order (or higher) from the LHC experiments that have nominal uncertainties that rival the full world average uncertainty
  - Z p_T
  - event shapes
- It would be nice to understand those uncertainties better, especially if PDF uncertainties are taken into account



N^3LL+N^3LO →



Compare relative luminosity uncertainties

good agreement in size of uncertainties between the 3 global PDFs

larger uncertainties of HERAPDF1.5 apparent

ABM11 uncertainties smaller at high mass

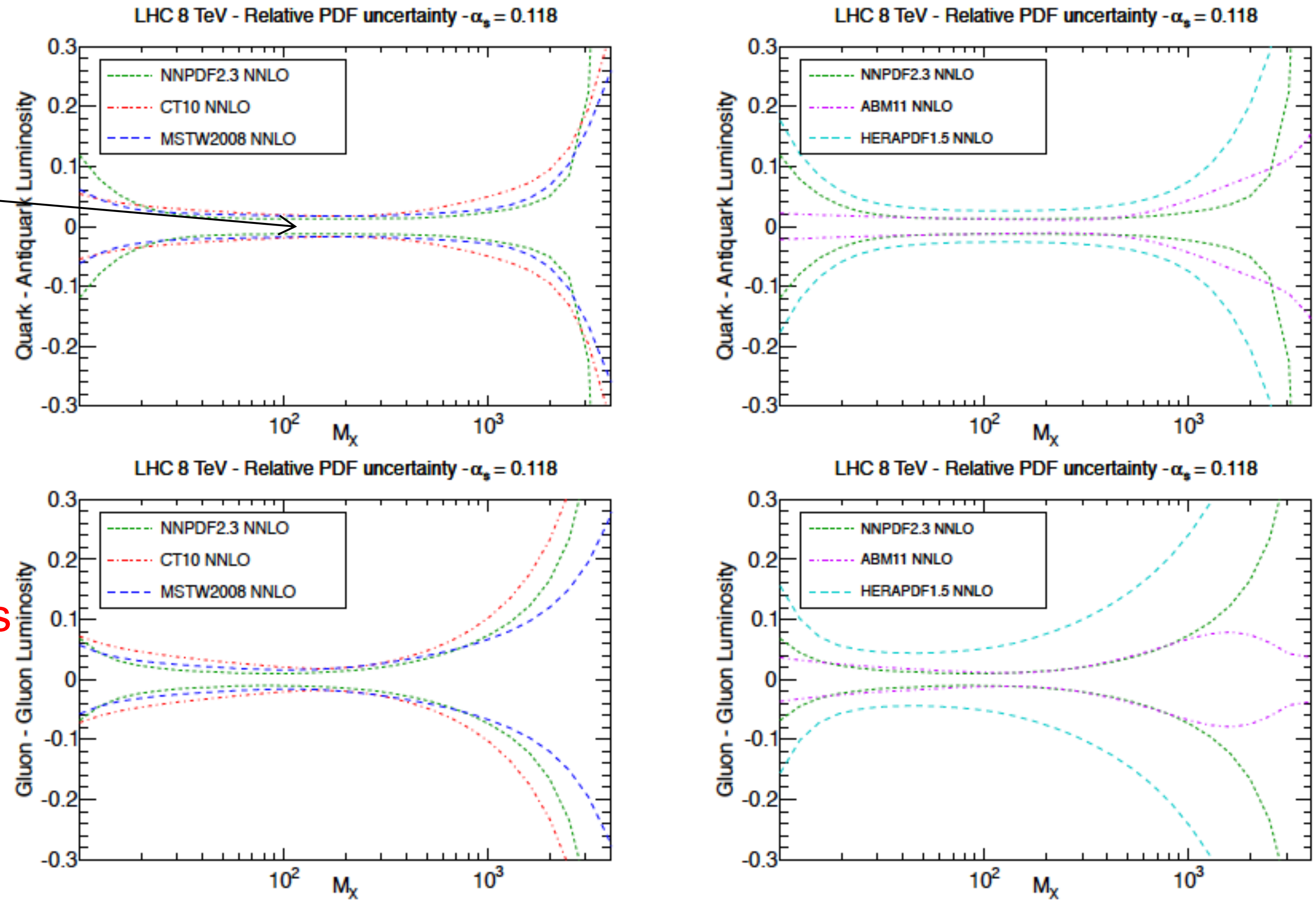


Figure 8: The relative PDF uncertainties in the quark-antiquark luminosity (upper plots) and in the gluon-gluon luminosity (lower plots), for the production of a final state of invariant mass M_X (in GeV) at the LHC 8 TeV. All luminosities are computed at a common value of $\alpha_s = 0.118$.

CT18

Experimental data set E		N_{pt}	χ^2/N_{pt}	S
LHCb 7 TeV 1.0 fb^{-1} W/Z forward rapidity	[61]	33	1.63 (1.21)	2.3 (0.9)
LHCb 8 TeV 2.0 fb^{-1} $Z \to e^- e^+$ forward rapidity	[62]	17	1.04 (1.06)	0.2 (0.3)
ATLAS 7 TeV 4.6 fb^{-1}, W/Z combined‡	[63]	34	8.45 (2.61)	16 (5.1)
CMS 8 TeV 18.8 fb^{-1} muon charge asymmetry A_{ch}	[64]	11	1.04 (1.10)	0.2 (0.3)
LHCb 8 TeV 2.0 fb^{-1} W/Z cross sec.	[65]	34	2.17 (1.75)	4.0 (2.7)
ATLAS 8 TeV 20.3 fb^{-1}, Z p_T cross sec.	[66]	27	1.12 (1.05)	0.5 (0.2)
CMS 7 TeV 5 fb^{-1}, single incl. jets, $R = 0.7$	[67]	158	1.23 (1.19)	2.0 (1.7)
ATLAS 7 TeV 4.5 fb^{-1}, single incl. jets, $R = 0.6$	[68]	140	1.45 (1.45)	3.4 (3.4)
CMS 8 TeV 19.7 fb^{-1}, single incl. jets, $R = 0.7$, (extended)	[69]	185	1.14 (1.12)	1.3 (1.2)
CMS 8 TeV 19.7 fb^{-1}, $t\bar{t}$ norm. double-diff. top p_T and y	[70]	16	1.18 (1.19)	0.6 (0.6)
ATLAS 8 TeV 20.3 fb^{-1}, $t\bar{t}$ p_T^t and $m_{t\bar{t}}$ abs. spectrum	[71]	15	0.63 (0.71)	-1.1 (-0.8)

Table 2.1. Numbers of points, χ^2/N_{pt}, and the effective Gaussian variables for the newly added LHC measurements in the CT18 and CT18Z NNLO fits. The ATLAS 7 TeV W/Z data (4.6 fb^{-1}), labelled by ‡, are included in the CT18A and CT18Z global fits, but not in CT18 and CT18X.

Spartyness, a variable that describes the goodness of fit, taking into account the number data points; expect S to be in the range of -1 to 1.

If S>>1, that means the data is poorly fit; if S<<1, that means the fit is too good, and possibly the errors are overestimated

$$S_E = \sqrt{2\chi_E^2} - \sqrt{2N_{pt,E} - 1}$$

MSHT20

Experimental data set	N_{pt}	χ^2/N_{pt}	S
D0 W asymmetry [106]	14	0.86	-0.3
$\sigma_{t\bar{t}}$ Tevatron +CMS+ATLAS 7,8 TeV [107]- [108]	17	0.85	-0.4
LHCb 7+8 TeV $W + Z$ [61,62]	67	1.48	2.6
LHCb 8 TeV e [65]	17	1.54	1.5
CMS 8 TeV W [64]	22	0.58	-1.5
ATLAS 7 TeV jets $R = 0.6$ [68]	140	1.59	4.4
CMS 7 TeV $W + c$ [102]	10	0.86	-0.2
ATLAS 7 TeV W, Z [63]	61	1.91	4.3
CMS 7 TeV jets $R = 0.7$ [67]	158	1.11	1.0
ATLAS 8 TeV Zp_T [66]	104	1.81	5.0
CMS 8 TeV jets [69]	174	1.50	4.2
ATLAS 8 TeV $t\bar{t} \to l + j$ single-diff [71]	25	1.02	0.1
ATLAS 8 TeV $t\bar{t} \to l^+ l^-$ single-diff [109]	5	0.68	-0.4
ATLAS 8 TeV high-mass Drell-Yan [110]	48	1.18	0.9
ATLAS 8 TeV $W^{+,-}$ + jet [111]	32	0.60	-1.7
CMS 8 TeV $(d\sigma_{t\bar{t}}/dp_{T,t}dy_t)/\sigma_{t\bar{t}}$ [70]	15	1.50	1.3
ATLAS 8 TeV W^+, W^- [100]	22	2.61	4.2
CMS 2.76 TeV jets [112]	81	1.27	1.7
CMS 8 TeV $t\bar{t}$ y_t distribution [113]	9	1.47	1.0
ATLAS 8 TeV double differential Z [99]	59	1.45	2.3

Table 2.2. Numbers of points, fit qualities χ^2/N_{pt} and S values for new collider data added to the NNLO MSHT20 fit.

Note the trouble fitting the ATLAS W/Z data

Definitions for CT18'/NNPDF3.1'

- CT18->CT18': m_c=1.4 GeV,m_b=4.75 GeV
- NNPDF3.1->NNPDF3.1': same as above plus some additions to the data set (in some ways NNPDF3.1' is a transition from 3.1-> 4.0)
- No MSHT20' since the above are the heavy quark mass values they normally use

	NNPDF3.1 [15]			NNPDF3.1′		
Experimental data set	$N_{\rm pt}$	$\chi^2/N_{\rm pt}$	S	$N_{\rm pt}$	$\chi^2/N_{\rm pt}$	S
D0 W electron asymmetry [121]	8	2.70	+2.70	11	3.07	+3.64
D0 W muon asymmetry [122]	9	1.56	+1.18	9	1.58	+1.21
ATLAS low-mass DY 7 TeV [123]	6	0.90	−0.03	6	0.89	−0.05
ATLAS W,Z 7 TeV [63]	34	2.14	+3.88	61	1.99	+4.58
ATLAS Z p_T 8 TeV (p_T, $m_{\ell\ell}$) [66]	44	0.93	−0.28	44	0.94	−0.23
ATLAS Z p_T 8 TeV (p_T, y_Z) [66]	48	0.94	−0.25	48	0.95	−0.20
ATLAS single-inclusive jets 7 TeV ($R = 0.6$) [68]	31	1.07	+0.33	140	1.25	+2.00
ATLAS $\sigma_{t\bar{t}}^{\rm tot}$ 7, 8, 13 TeV [124,125]	3	0.86	+0.04	3	0.95	+0.15
ATLAS $t\bar{t}$ ℓ+jets 8 TeV ($1/\sigma\ d\sigma/dy_t$) [71]	9	1.45	+0.99	4	3.56	+2.69
CMS W rapidity 8 TeV [64]	22	1.01	+0.11	22	1.03	+0.17
CMS Z p_T 8 TeV [126]	28	1.32	+1.18	28	1.34	+1.25
CMS single-inclusive jets 2.76 TeV [112]	81	1.03	+0.23	—	—	—
CMS single-inclusive jets 8 TeV [69]	—	—	—	185	1.30	+2.72
CMS $\sigma_{t\bar{t}}^{\rm tot}$ 7, 8, 13 TeV [127,128]	3	0.20	−1.14	3	0.18	−1.20
CMS $t\bar{t}$ ℓ+jets 8 TeV ($1/\sigma\ d\sigma/dy_{t\bar{t}}$) [113]	9	0.94	−0.01	9	1.67	+1.36
CMS $t\bar{t}$ 2D 2ℓ 8 TeV ($1/\sigma\ d\sigma/dy_t dm_{t\bar{t}}$) [70]	—	—	—	16	0.81	−0.48
LHCb $W,Z \to \mu$ 7 TeV [61]	29	1.76	+1.55	29	1.96	+3.11
LHCb $W,Z \to \mu$ 8 TeV [65]	30	1.37	+1.39	30	1.36	+1.35

Note the trouble fitting the ATLAS W/Z data

important addition

Table 2.3. The numbers of points, $\chi^2/N_{\rm pt}$ and S values for new collider data in the NNPDF3.1 fit [15] and in the NNPDF3.1′ fit variant adopted in the present combination. The Tevatron and LHC data sets already included in NNPDF3.0 are kept in NNPDF3.1, but not necessarily in NNPDF3.1′. These are not indicated in the table. Note that, despite the number of LHC data points is larger in NNPDF3.1′ than that in NNPDF3.1, the total number of data points in the two analyses is similar, mainly because the Tevatron single-inclusive jet measurements (not indicated in the table) are no longer included in NNPDF3.1′. See text for details.

Combination

- Generate 300 MC replicas of each of the 3 PDFs and combine

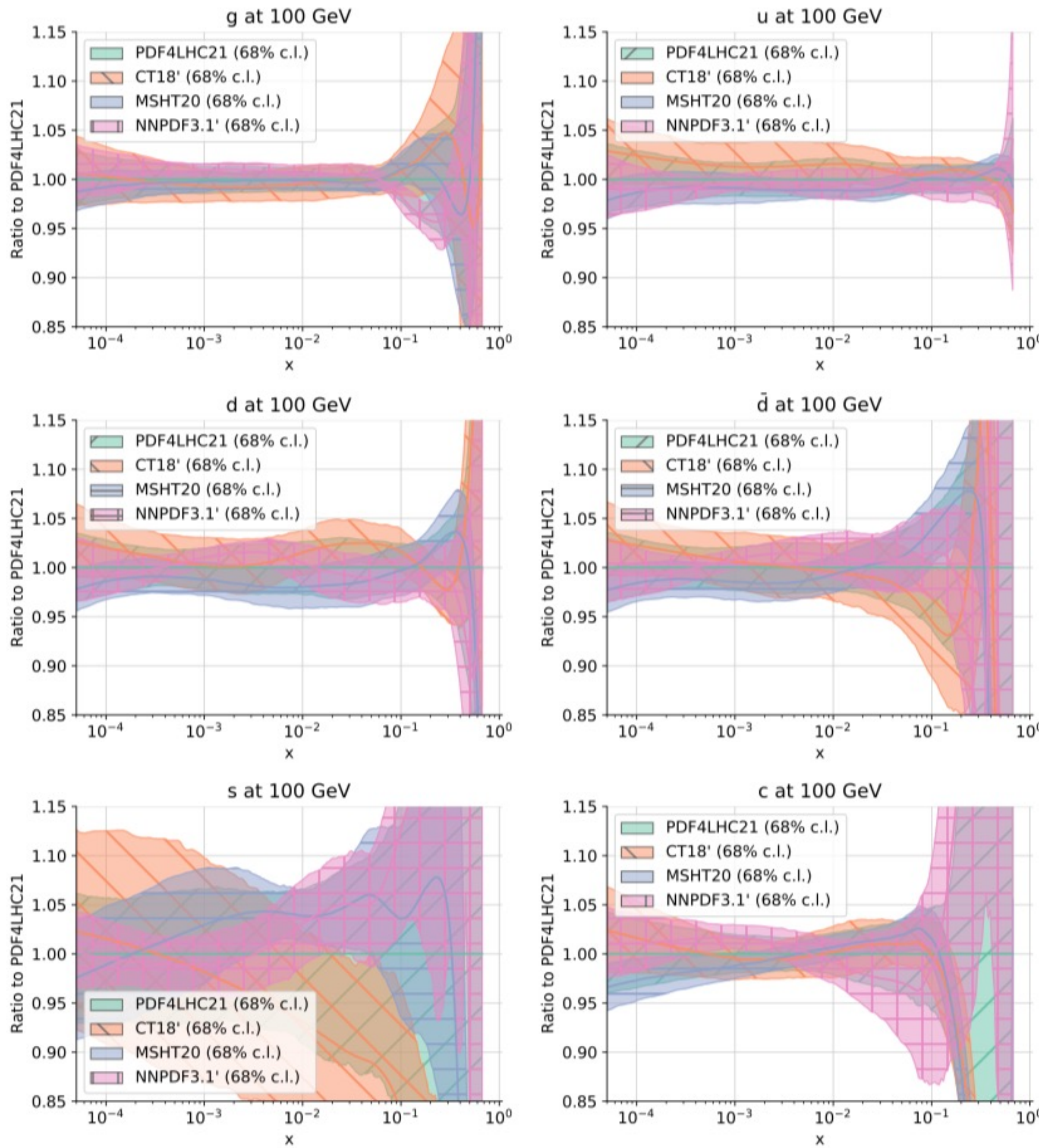


Figure 4.1. Comparison of the PDF4LHC21 combination (composed by $N_{\rm rep} = 900$ replicas) with the three constituent sets at $Q = 100$ GeV, normalised to the central value of the former and with their respective 68%CL uncertainty bands. In the case of the Hessian sets (CT18′ and MSHT20) we display their Monte Carlo representation composed by $N_{\rm rep} = 300$ replicas generated according to Eq. (4.3). The NNPDF3.1′ band is also constituted by $N_{\rm rep} = 300$ (native) replicas.

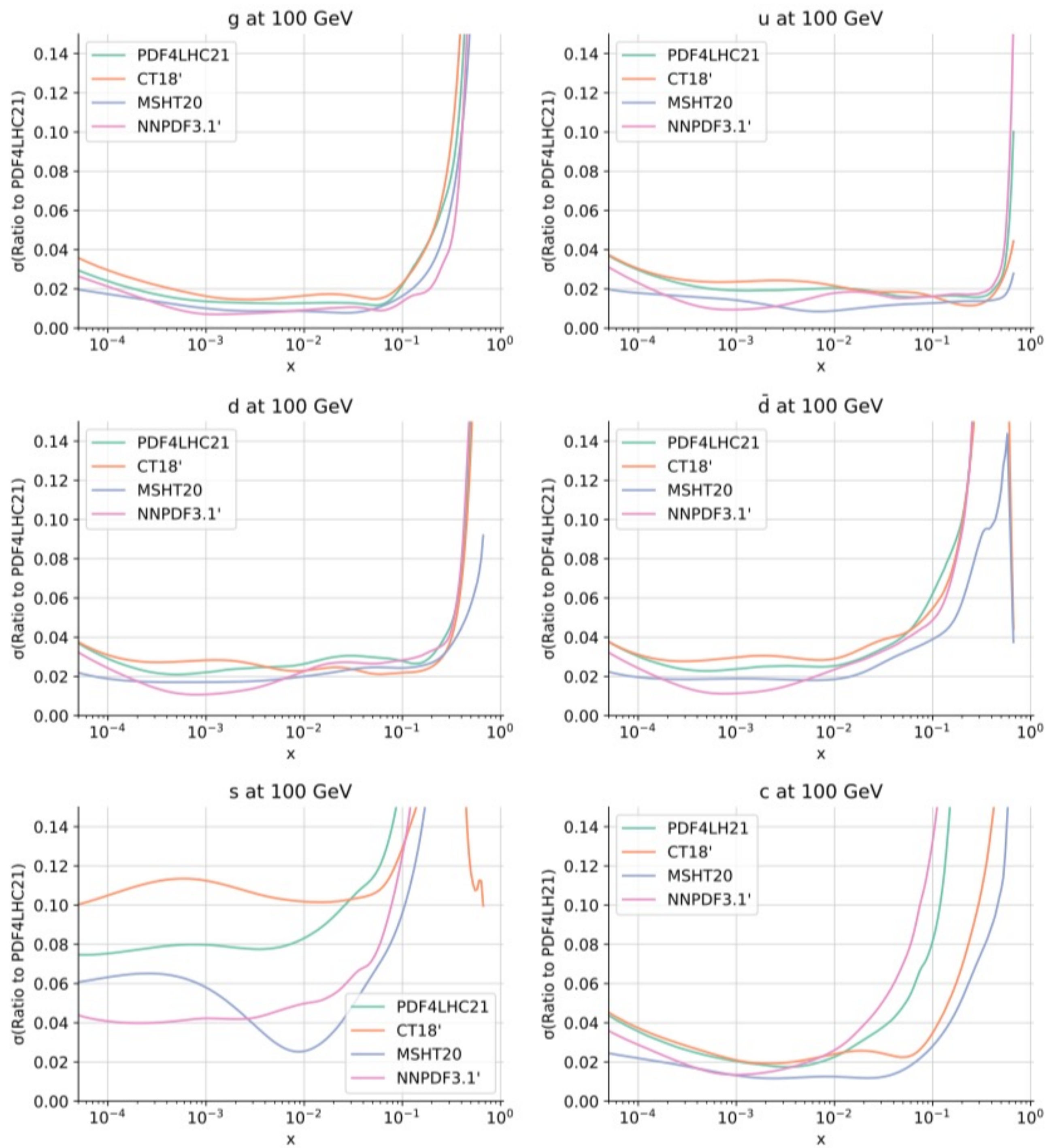


Figure 4.2. Same as Fig. 4.1 now showing the relative PDF 68% CL uncertainties (normalised to the PDF4LHC21 central value) of the four PDF sets.

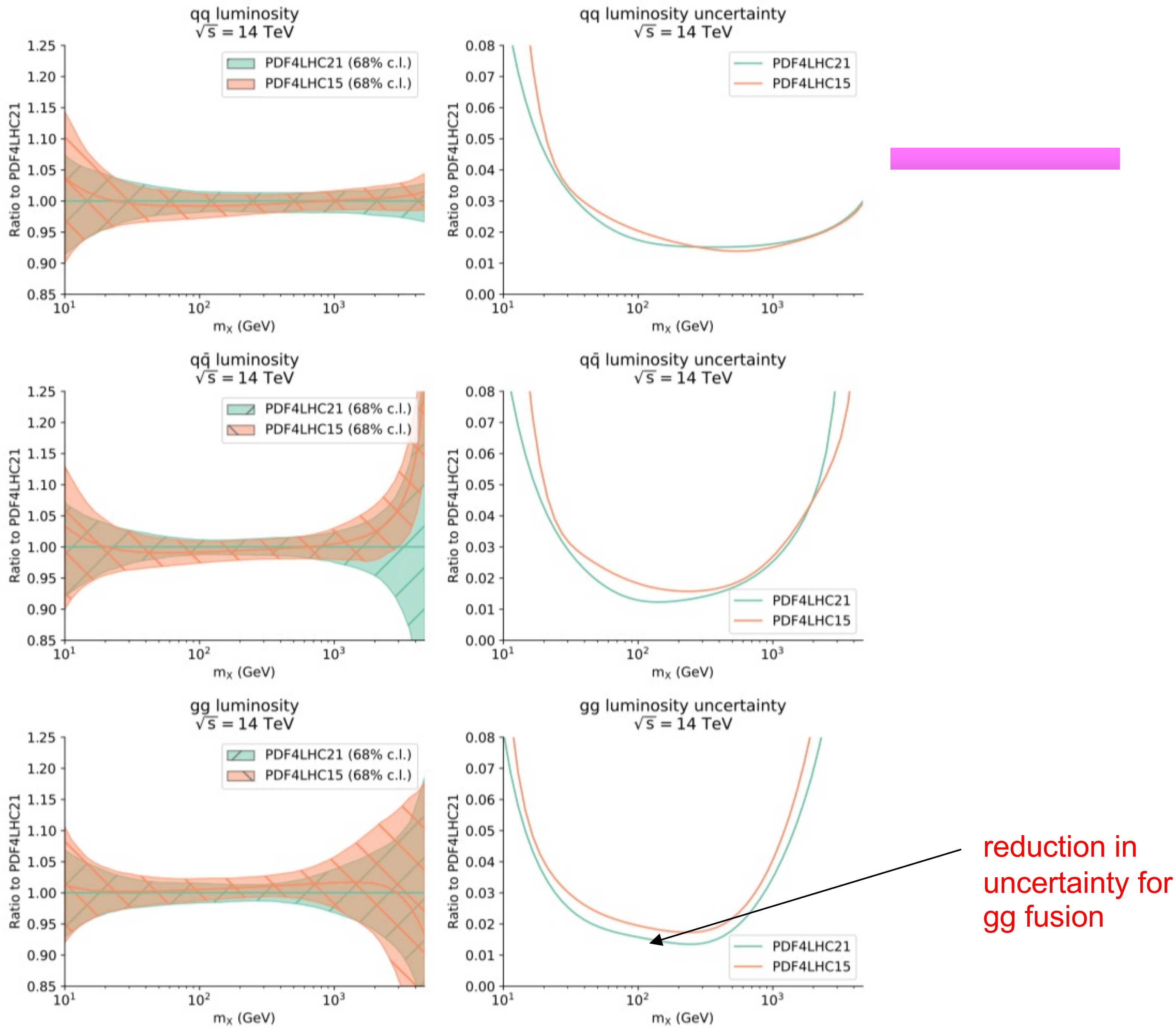


Figure 4.11. Comparison of the partonic luminosities at $\sqrt{s}$ = 14 TeV between PDF4LHC15 and PDF4LHC21. In both cases, the original sets with $N_{\rm rep}$ = 900 have been used. Results are shown for the quark-quark, quark-antiquark, and gluon-gluon luminosities as a function of the final state invariant mass m_X, and normalised to the central value of the PDF4LHC21 prediction. The right panels display the corresponding 68% CL relative uncertainties.

- It can be useful to look at 2-D ellipses comparing cross sections

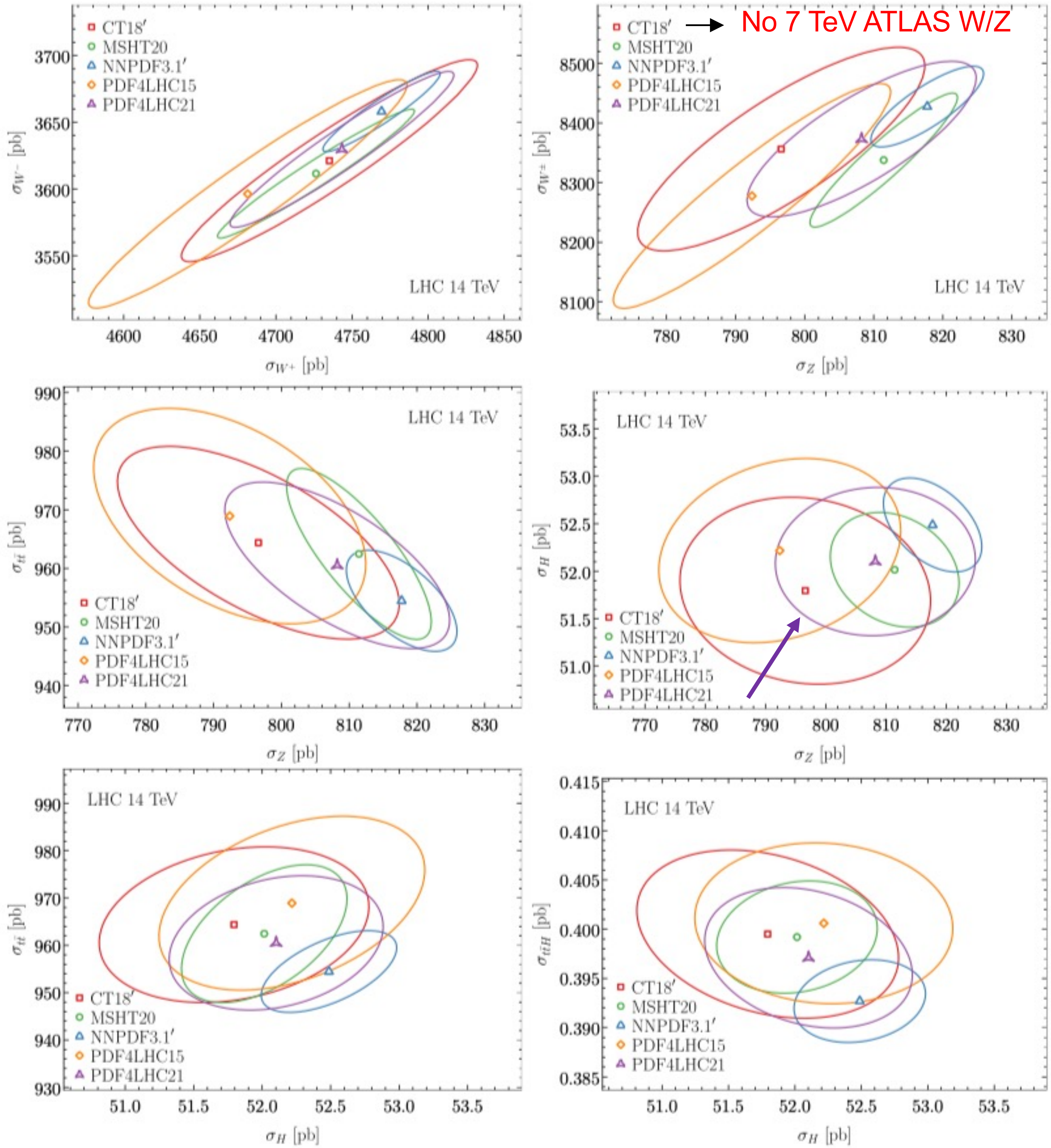


Figure 5.2. The 1σ ellipses for pairs of inclusive cross sections among $W^\pm, Z, t\bar{t}, H, t\bar{t}H$ production at the LHC 14 TeV. The $W^\pm/Z$ cross sections are defined in the ATLAS 13 TeV fiducial volume [170], while others correspond to the full phase space. See text for details of the theory calculations.

Error PDFs

- ATLAS, CT and MSHT groups adopt the Hessian format for their PDF error sets
- D error PDFs are used to determine the PDF uncertainty (assuming the probability distribution is approximately Gaussian)
- Consider an expansion of a function X of the parameters R in the vicinity of the global χ^2 minimum X_o

$$X(\vec{R}) = X_0 + \sum_{i=1}^{D} \left.\frac{\partial X}{\partial R_i}\right|_{\vec{R}=\vec{0}} R_i + \frac{1}{2}\sum_{i,j=1}^{D} \left.\frac{\partial^2 X}{\partial R_i \partial R_j}\right|_{\vec{R}=\vec{0}} R_i R_j + \ldots .$$

$$\left.\frac{\partial X}{\partial R_i}\right|_{\vec{R}=\vec{0}} \approx \frac{X_{+i} - X_{-i}}{2}.$$

use symmetrized form for first order derivative

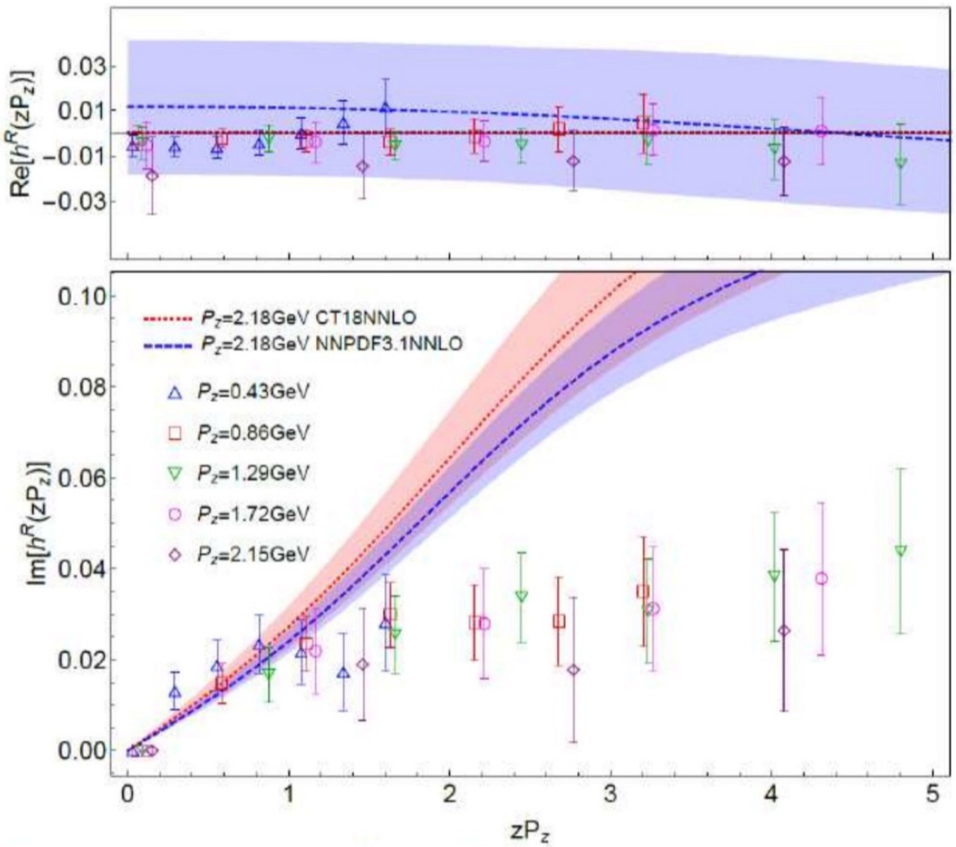
$$\Delta^{\mathrm{H}} X = \left|\vec{\nabla} X\right| = \frac{1}{2}\sqrt{\sum_{i=1}^{D} [X_{+i} - X_{-i}]^2}.$$

define 68% CL hypersphere

$$C^{\mathrm{H}}(X,Y) = \frac{1}{4\Delta^{\mathrm{H}} X \Delta^{\mathrm{H}} Y} \sum_{i=1}^{D} (X_{+i} - X_{-i})(Y_{+i} - Y_{-i})$$

define correlation between 2 variables X and Y

From quasi-PDF to PDF



- Due to the large uncertainty in strangeness PDF from global analysis, lattice QCD calculation is able to provide more information.

$$\mathrm{Re}[h(z)] \propto \int dx\,(s(x)-\bar{s}(x))\cos(xzP_z)$$

$$\mathrm{Im}[h(z)] \propto \int dx\,(s(x)+\bar{s}(x))\sin(xzP_z)$$

- MSULat/quasi-PDF method
- Clover on 2+1+1 HISQ 0.12-fm 310-MeV QCD vacuum
- RI/MOM renormalization
- Extropolartion to M_pi = 140 MeV

[Zhang et al, 2005.12015]