European Research Council
Established by the European Commission

MARIA UBIALI
UNIVERSITY OF CAMBRIDGE

PARTON DISTRIBUTION FUNCTIONS

European Research Council
Established by the European Commission

MARIA UBIALI
UNIVERSITY OF CAMBRIDGE

PARTON DISTRIBUTION FUNCTIONS

– A PERSONAL OVERVIEW

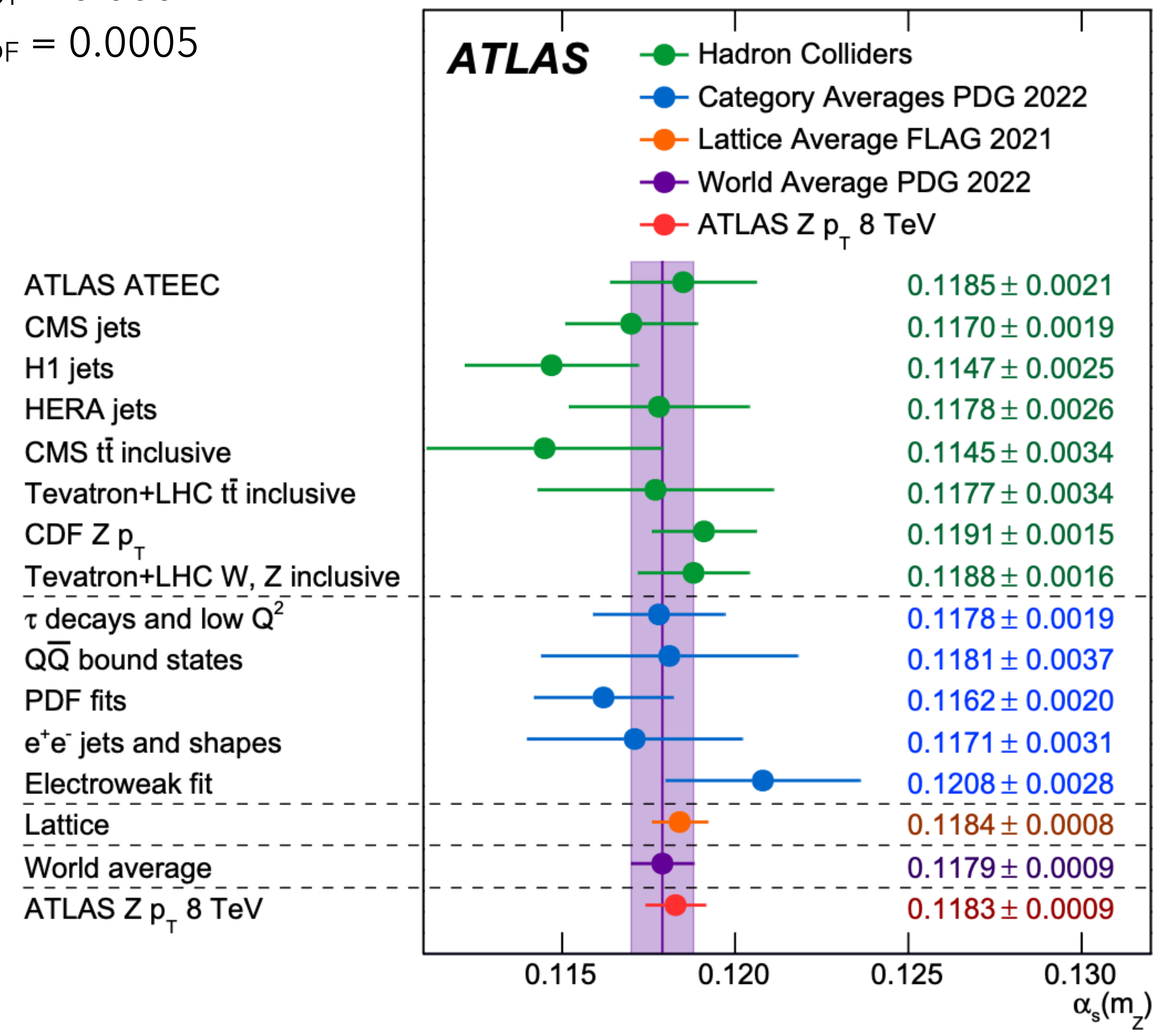
PDFS CRUCIAL INPUT FOR PRECISION PHYSICS

$$d\sigma^{pp\to ab} = \sum_{i,j} \boxed{f_i \otimes f_j} \otimes \boxed{d\hat{\sigma}^{ij\to ab}} + \ldots$$

ATLAS α_S determination

Δ_{TOT} = 0.0009

Δ_{PDF} = 0.0005

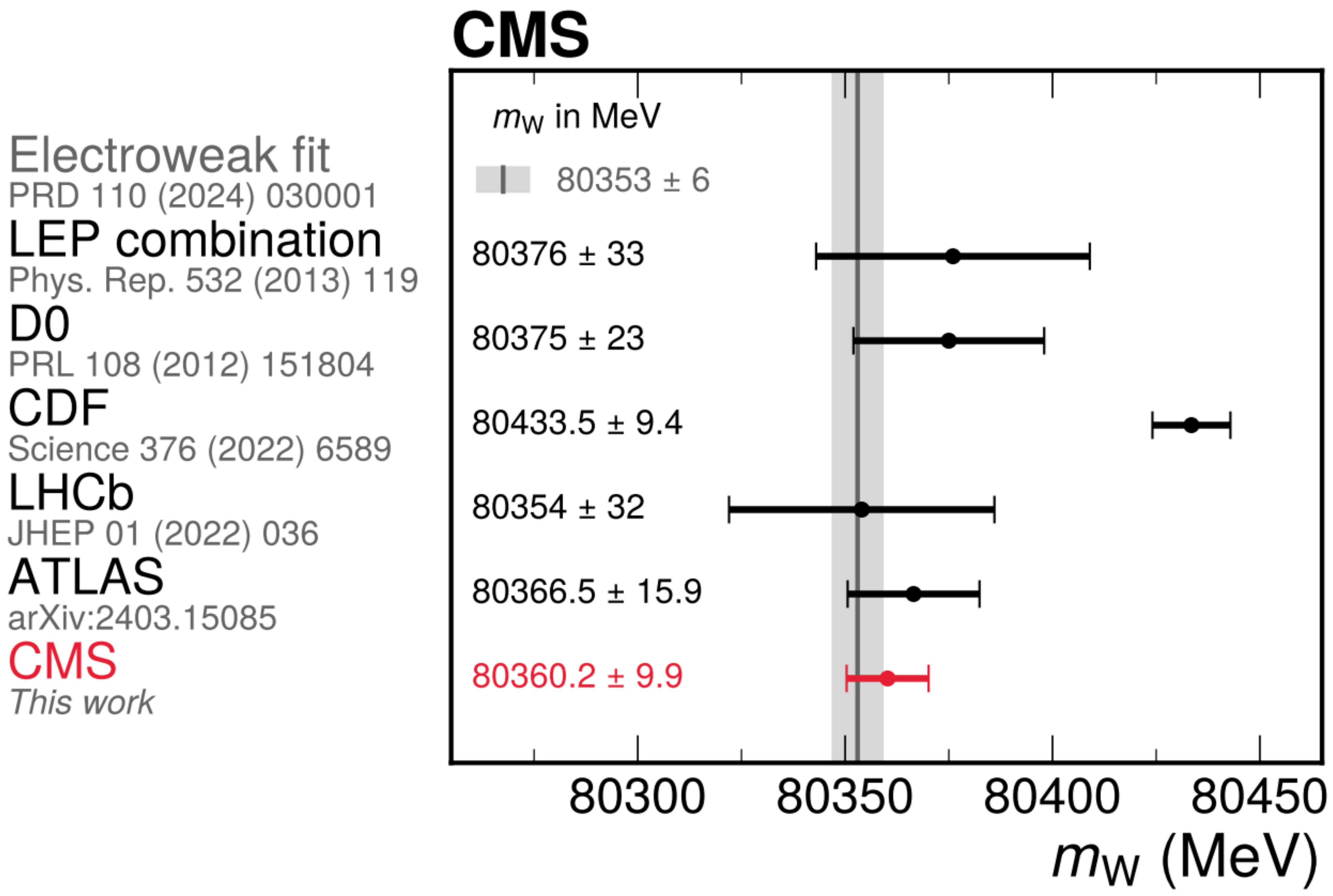


ATLAS collaboration, arXiv: 2309.12986

CMS M_W determination

Δ_{TOT} = 9.9 MeV

Δ_{PDF} = 4.4 MeV



CMS collaboration, arXiv: 2412.13872

OUTLINE

➡ Part I: progress and highlights on recent developments in PDF determination

➡ Part II: precision vs accuracy: new challenges and frontiers

PDF DETERMINATION

- Huge progress from first PDF determinations to where we stand now
  - ➡ Experimental data: many more precise measurements constraining PDFs
  - ➡ Theory: heavy quark schemes, aN3LO PDFs, inclusion of MHOUs in PDF error bands, EW and QED corrections, initial photon…
  - ➡ Methodology: uncertainty propagation, parametrisation bias, closure tests,…

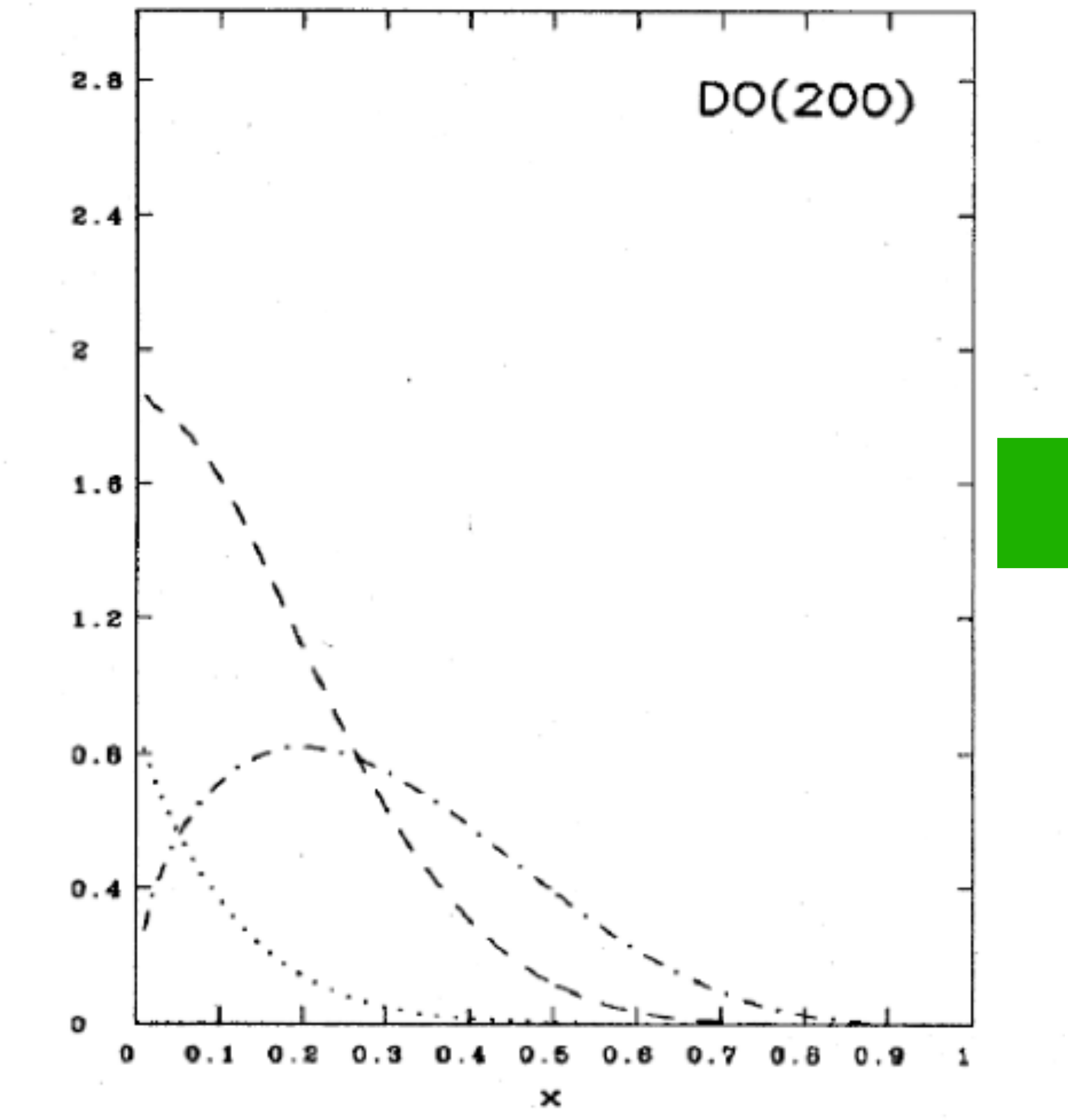
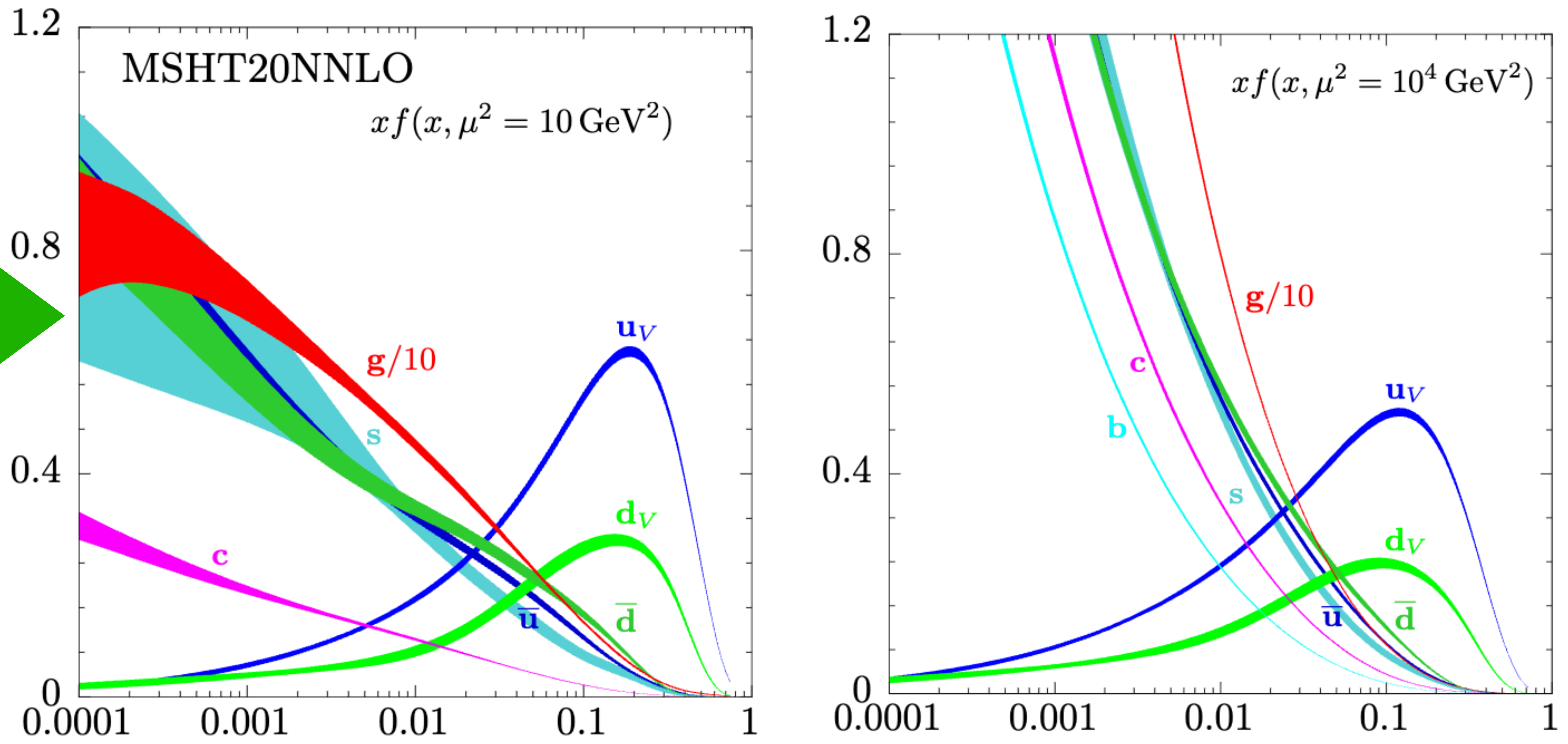


FIG. 27. "Soft-gluon" (Λ=200 MeV) parton distributions of Duke and Owens (1984) at Q^2=5 GeV^2: valence quark distribution $x[u_v(x)+d_v(x)]$ (dotted-dashed line), $xG(x)$ (dashed line), and $q_v(x)$ (dotted line).

40 years

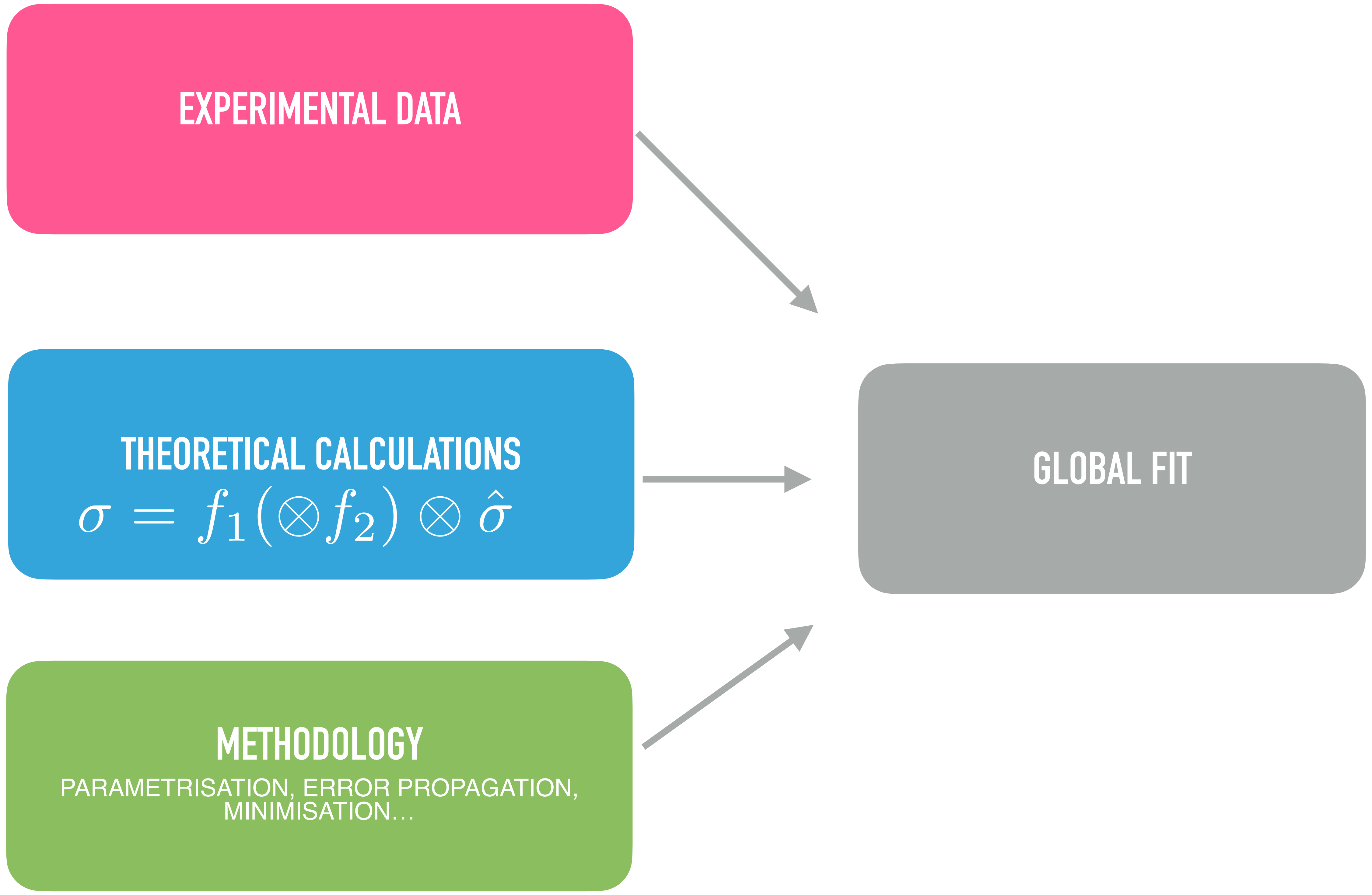


Rev. Mod. Phys 1984

PDG 2024

PDF DETERMINATION

- Huge progress from first PDF determinations to where we stand now
  - ➡ Experimental data: many more precise measurements constraining PDFs
  - ➡ Theory: heavy quark schemes, aN3LO PDFs, inclusion of MHOUs in PDF error bands, EW and QED corrections, initial photon…
  - ➡ Methodology: uncertainty propagation, parametrisation bias, closure tests,…



Accuracy of PDF determinations has a limit that is determined by the accuracy of all ingredients entering a global fit

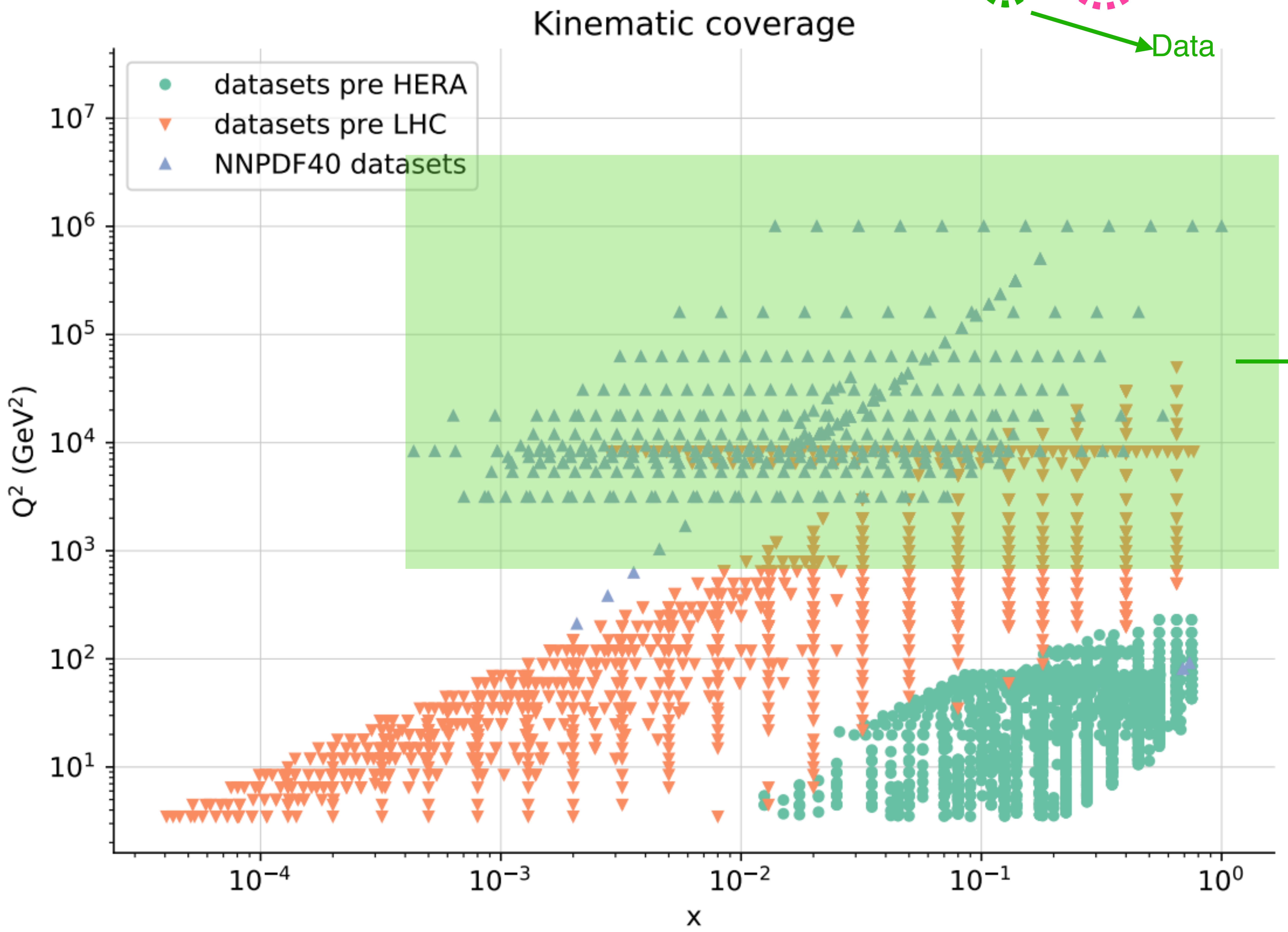
PROGRESS: DATA

$$f_i(x,\mu)$$

Perturbative QCD

Data

Splitting functions P_{ab} known up to approximate N3LO
[Blumlein, Moch, Gehrman, von Manteufel, Sotnikov, Yang, Davies, Vogt, Bonvini, Marzani,…]



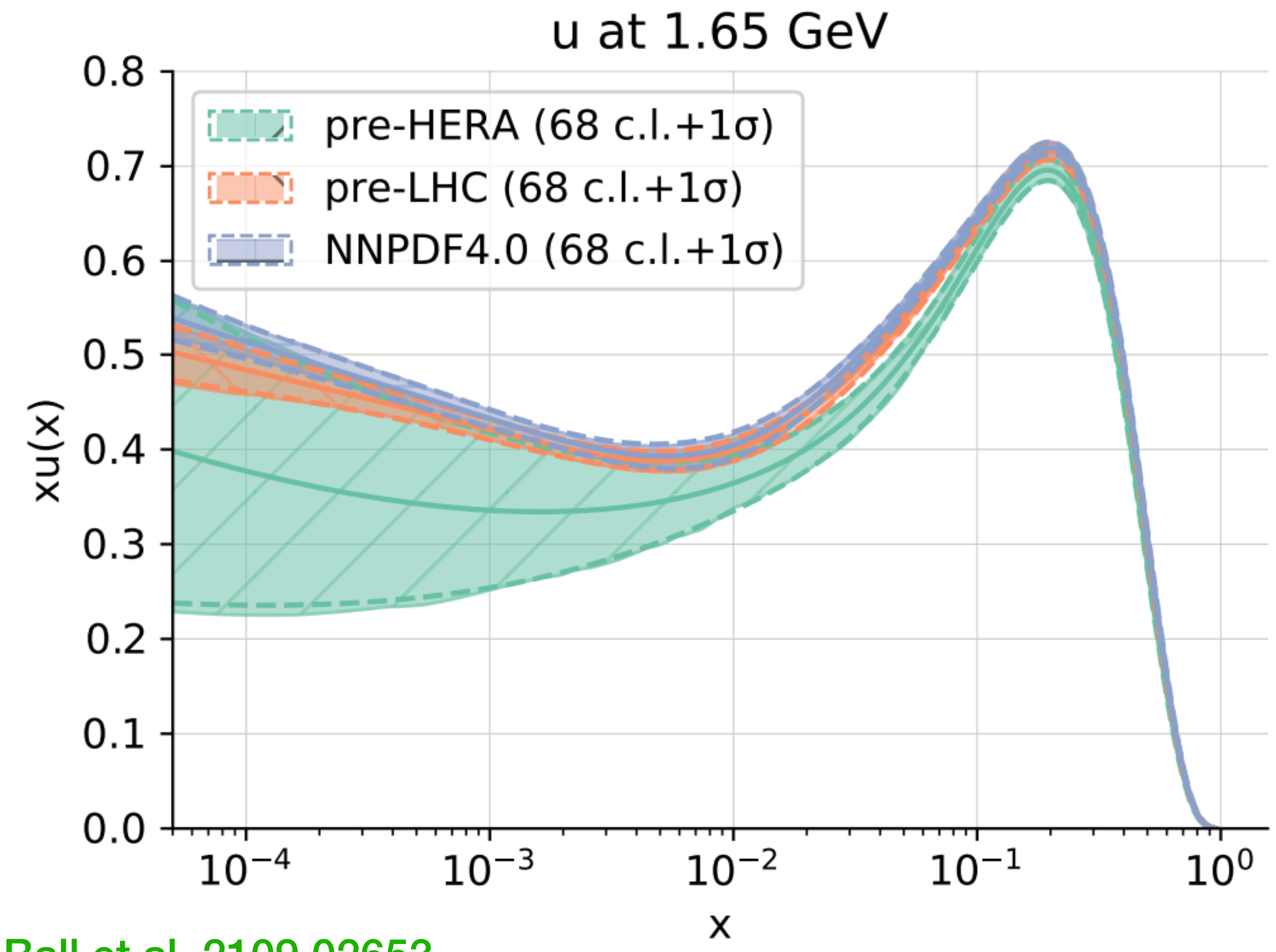
NNPDF4.0:
O(5000) datapoints,
about 30% from LHC

- Modern data show increasingly small statistical uncertainty and dominating correlated systematic uncertainties make global PDF fits challenging

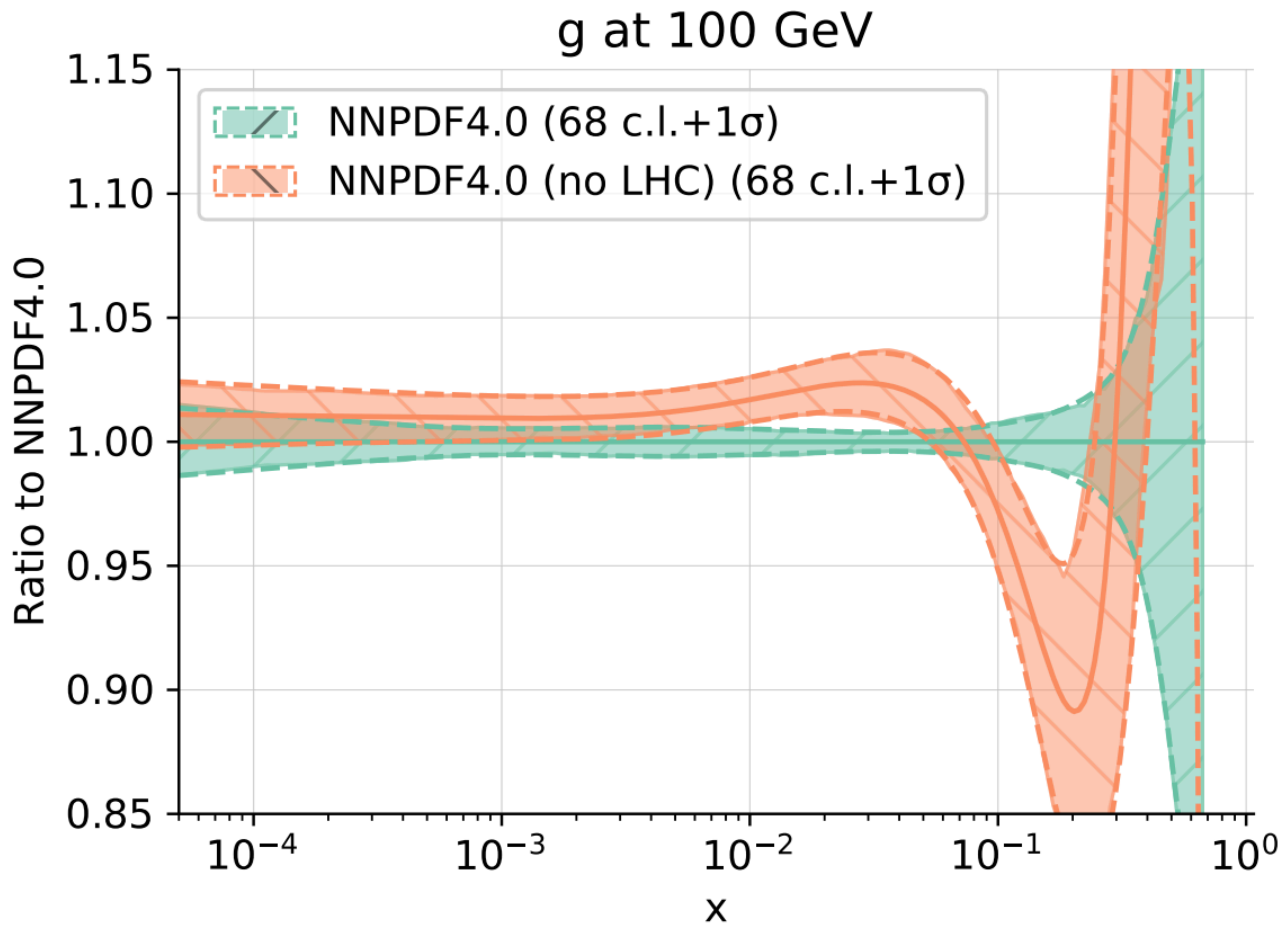
Kassabov, Nocera, Wilson 2207.00690
Bailey, Harland-Lang 1909.10541

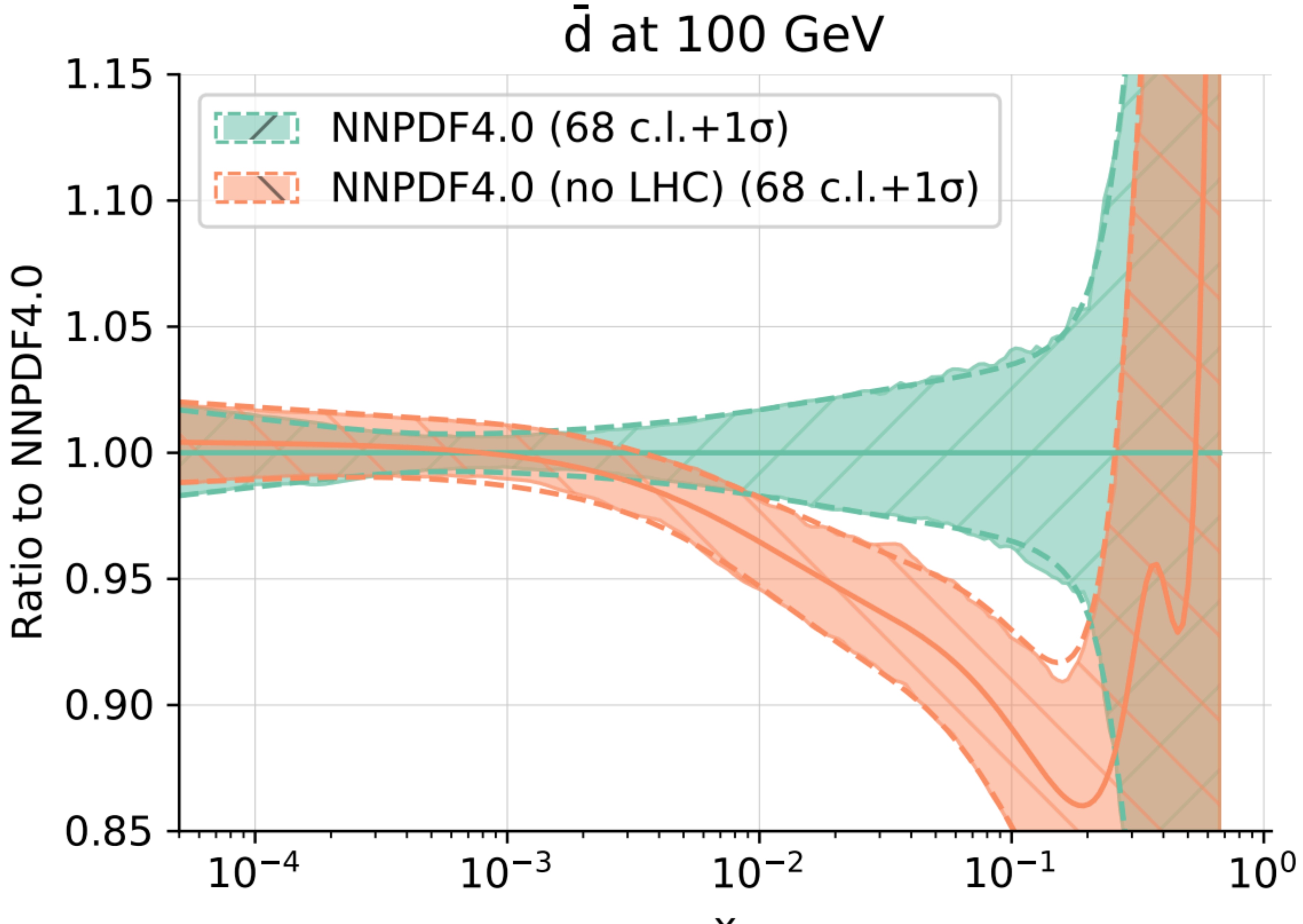
Ball et al, 2109.02653

PROGRESS: DATA



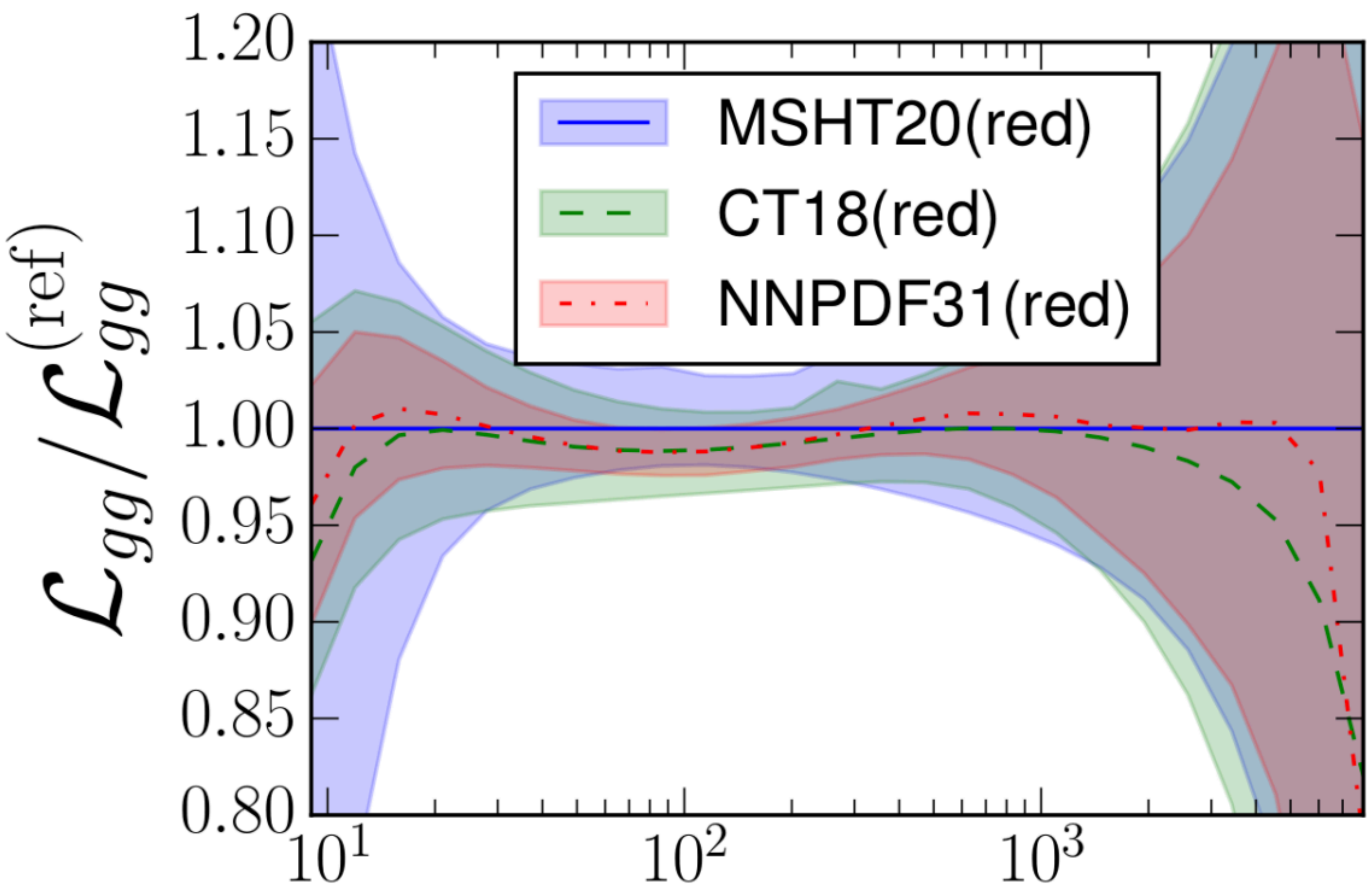
Ball et al, 2109.02653

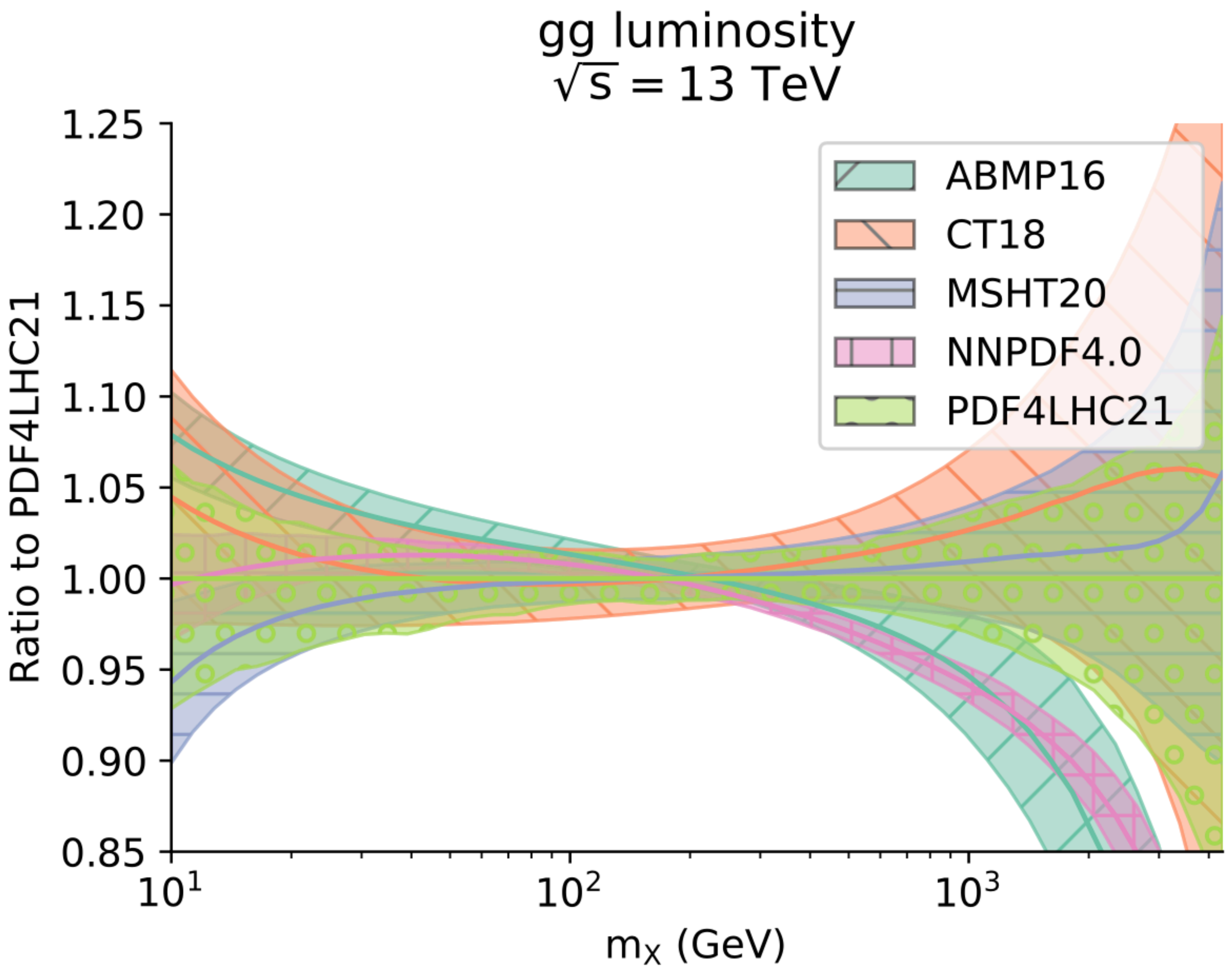




- **HERA** data crucial to constrain quark valence (up and to less extent down valence) across intermediate to small x and gluon at small x
- **LHC** high energy data crucial to provide additional constraints to PDFs, in particular in medium- to large-x gluon and quarks.
- Some **tension** with older fixed-target Drell-Yan and DIS data visible in the large-x region (especially gluon and anti-quarks)

DATA-DRIVEN PROGRESS





- In recent updates from global PDF fitting collaborations **effect of LHC data driving PDF uncertainties down**
  ➜ PDF4LHC21 combination of NNPDF31′, CT18′ and MSHT20′ smaller uncertainties than PDF4LHC15
- Benchmark among NNPDF3.1, MSHT20 and CT18: overall agreement, which improves once common dataset is used, **differences in uncertainties** associated with methodologies with ΔCT18 ≳ ΔMSHT20 ≳ ΔNNPDF31
- Increased precision of the data and their strong correlation demands methodological improvements (e. g. **more flexible parametrisation** that shifts PDFs outside nominal error band) and **>1 PDF fits** to accomodate some tension (CT18 vs CT18Z)

PROGRESS: TESTING THE METHODOLOGY

- Closure tests: key study to **statistically validate** various PDF fitting methodologies and **assess faithfulness of PDF uncertainties**
- Imagine we knew the law of Nature **G(w⁰)** and the true PDFs **w⁰** is fitting methodology able to reproduce it? Is the uncertainty faithful?
- Can also be used to check the effect of inconsistencies in the data

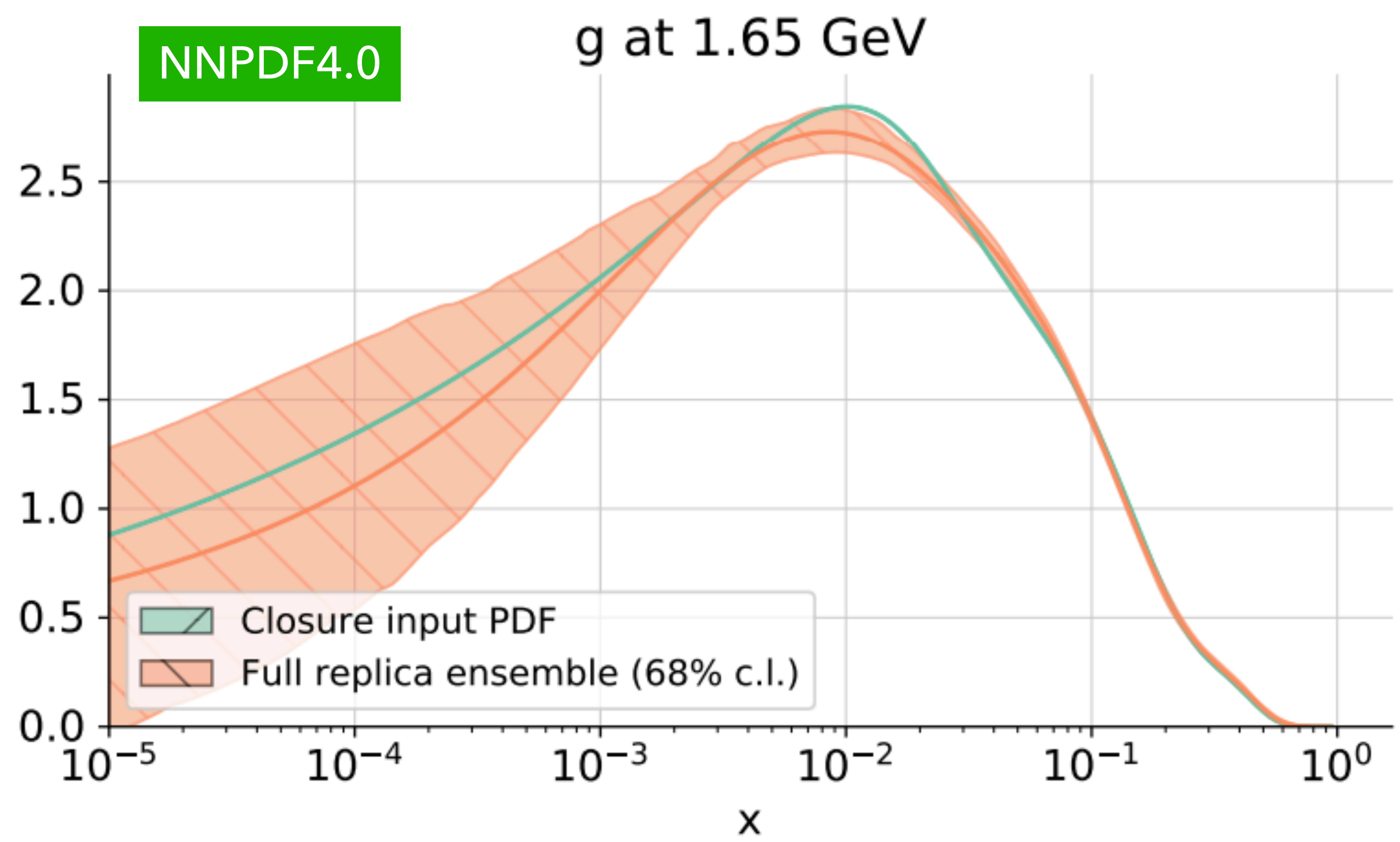
[Harland-Lang et al, arXiv:2407.07944, Costantini at al: arXiv:2503.17447]

$$y_0 = \mathcal{G}(w^0) + \eta$$

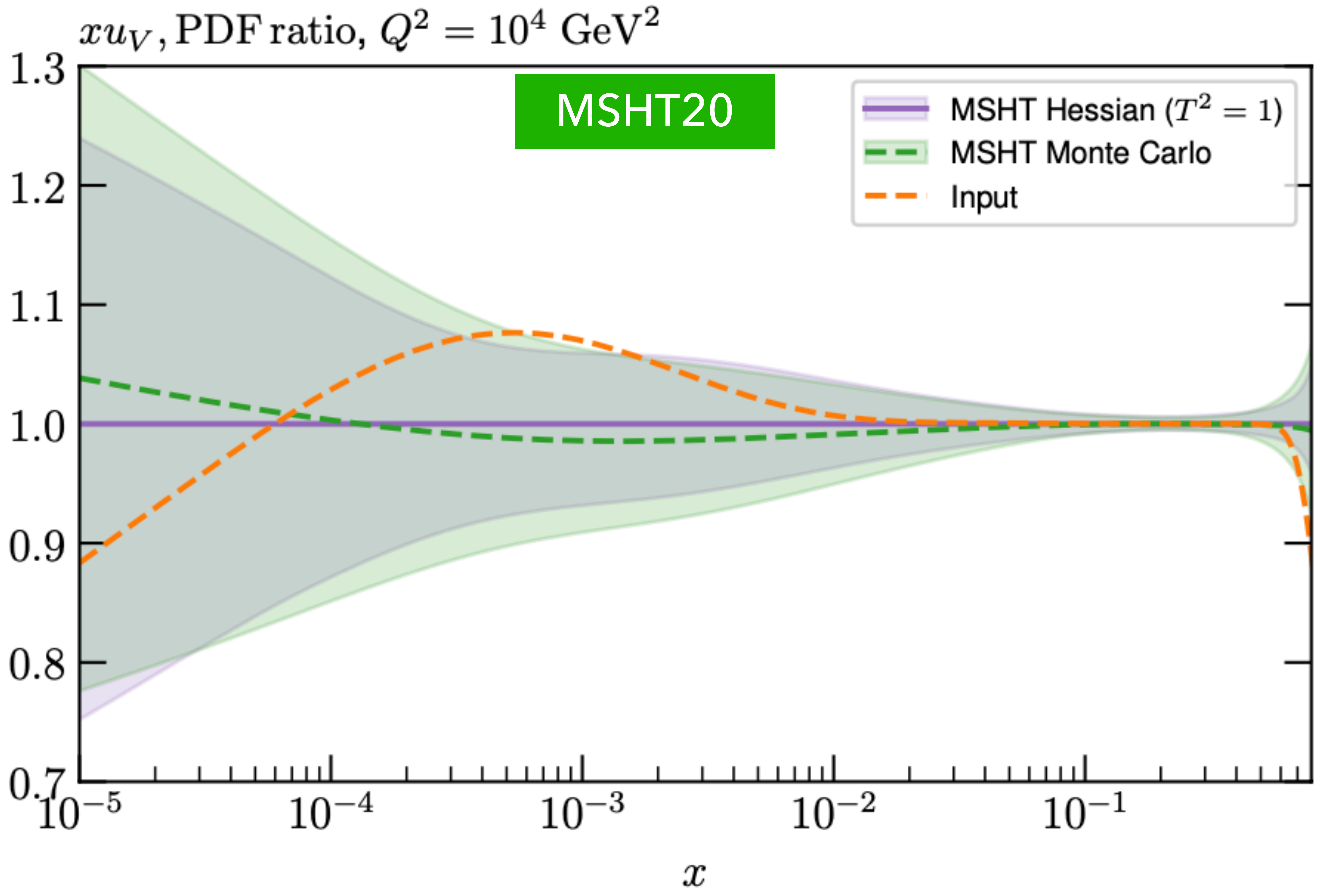
Experimental noise

Law of Nature

$\hat{\sigma}_{\rm NNLO} \otimes (f \otimes f)_{\text{“true” PDF}}$



Del Debbio et al, [arXiv: 2111.05787]



Harland-Lang et al, arXiv:2407.07944

PROGRESS: TESTING PDF PERFORMANCE ON NEW DATA

- To discriminate among PDF sets, test all global PDF sets agains **new precise** data from LHC data & DIS HERA jets data. How well do various PDF sets describe data that are **not yet** included in the fit?

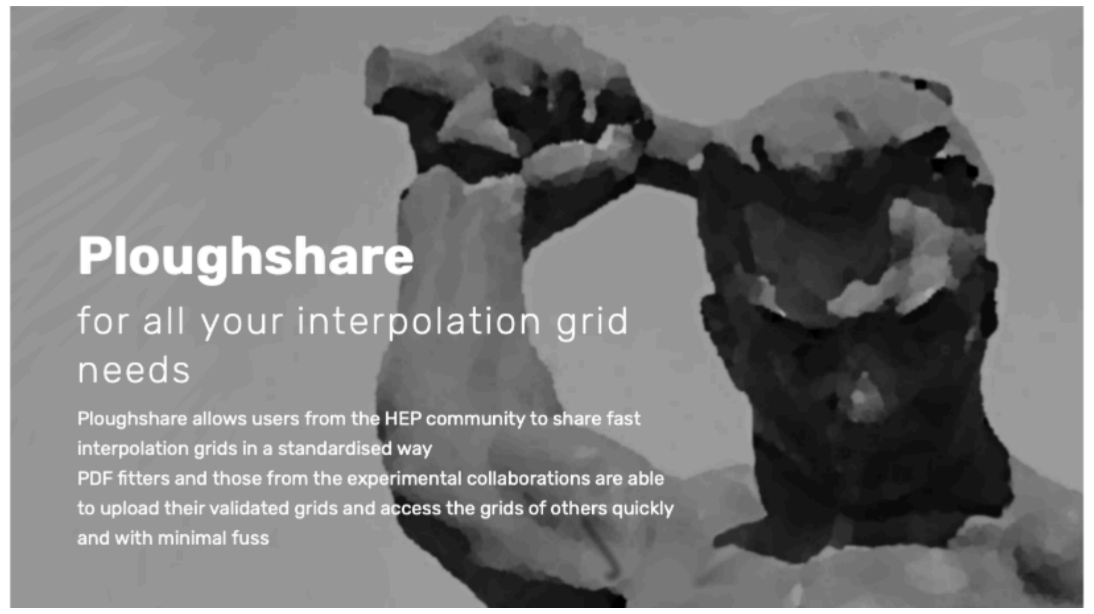
- **Generalisation test** is crucial and possible using exact NNLO predictions (no k-factor approximation) thanks to PineAPPL, NNLOJET, MATRIX and Ploughshare

PineAPPL

Carrazza et al: 2008.12789
Barontini et al: 2302.12124

NNLOJET

Gehrmann-De Ridder et al: 1507.02850, 1605.04295,2301.11827, Britzger et al: 2207.13735,Cruz-Martinez et al: 2501.13167



Ploughshare project, APPLgrid, FastNLO
Kluge et al: hep-ph/0609285
Carli et al: 0911.2985
Czakon et al: 1704.08551

MATRIX

Munich -- the MUlti-chaNnel Integrator at swiss (CH) precision --
Automates qT-subtraction and Resummation to Integrate X-sections

Grazzini et al: 1711.06631

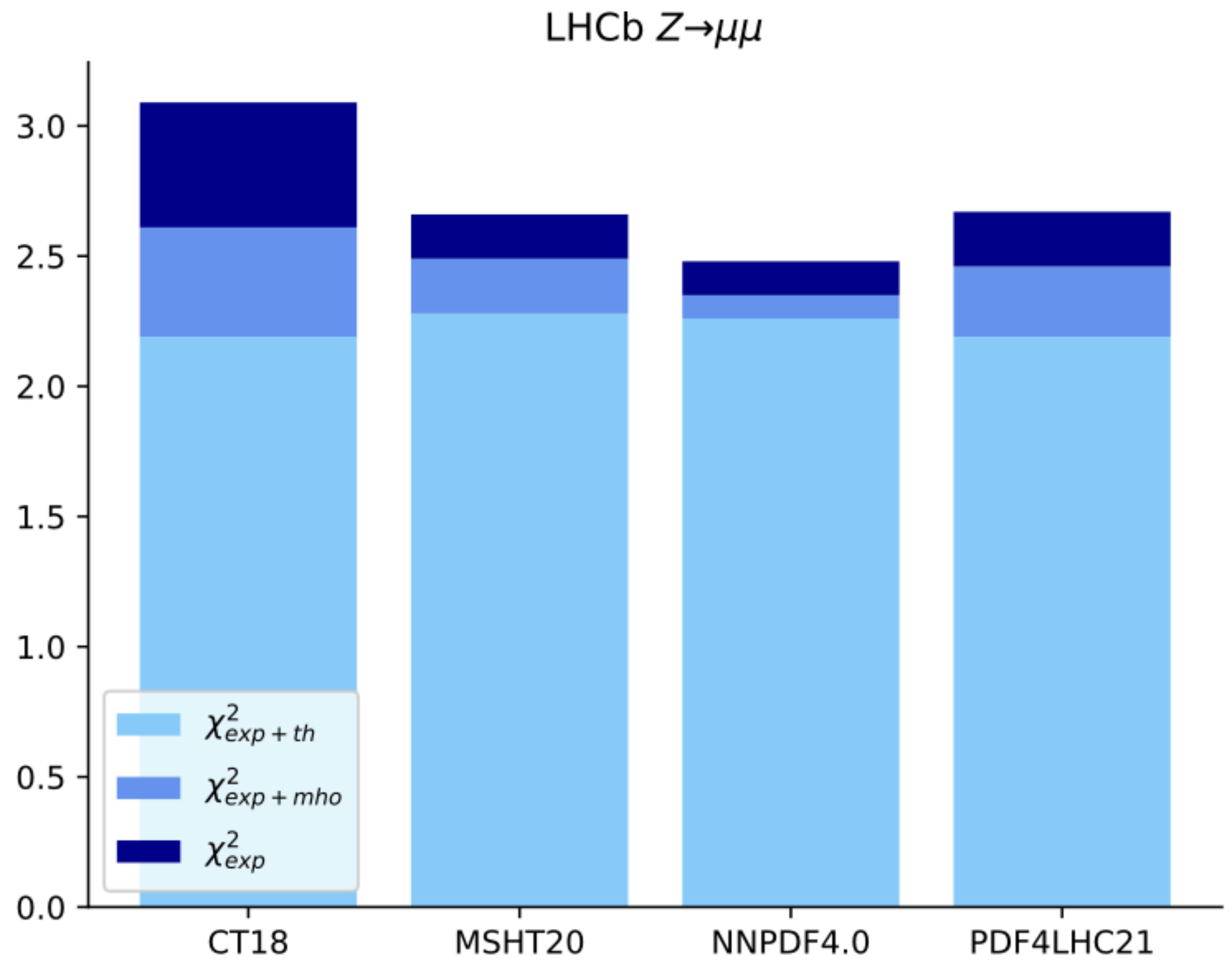
- Data-theory agreement with unseen data must be measured by considering **all sources of uncertainties**: experimental, missing higher order (MHOU), PDF uncertainties, α_s uncertainties by including all bin-by-bin correlations

$$\chi^2 = \sum_{i,j=1}^{N_{\rm dat}} \left(T_i^{(0)} - D_i\right) \left(\mathrm{cov}^{-1}\right)_{ij} \left(T_j^{(0)} - D_j\right)$$

$$\boxed{\chi^2_{\rm exp+mho+pdf+as} \leftrightarrow \mathrm{cov} = \mathrm{cov}_{\rm exp} + \mathrm{cov}_{\rm mho} + \mathrm{cov}_{\rm pdf} + \mathrm{cov}_{\rm as}}$$

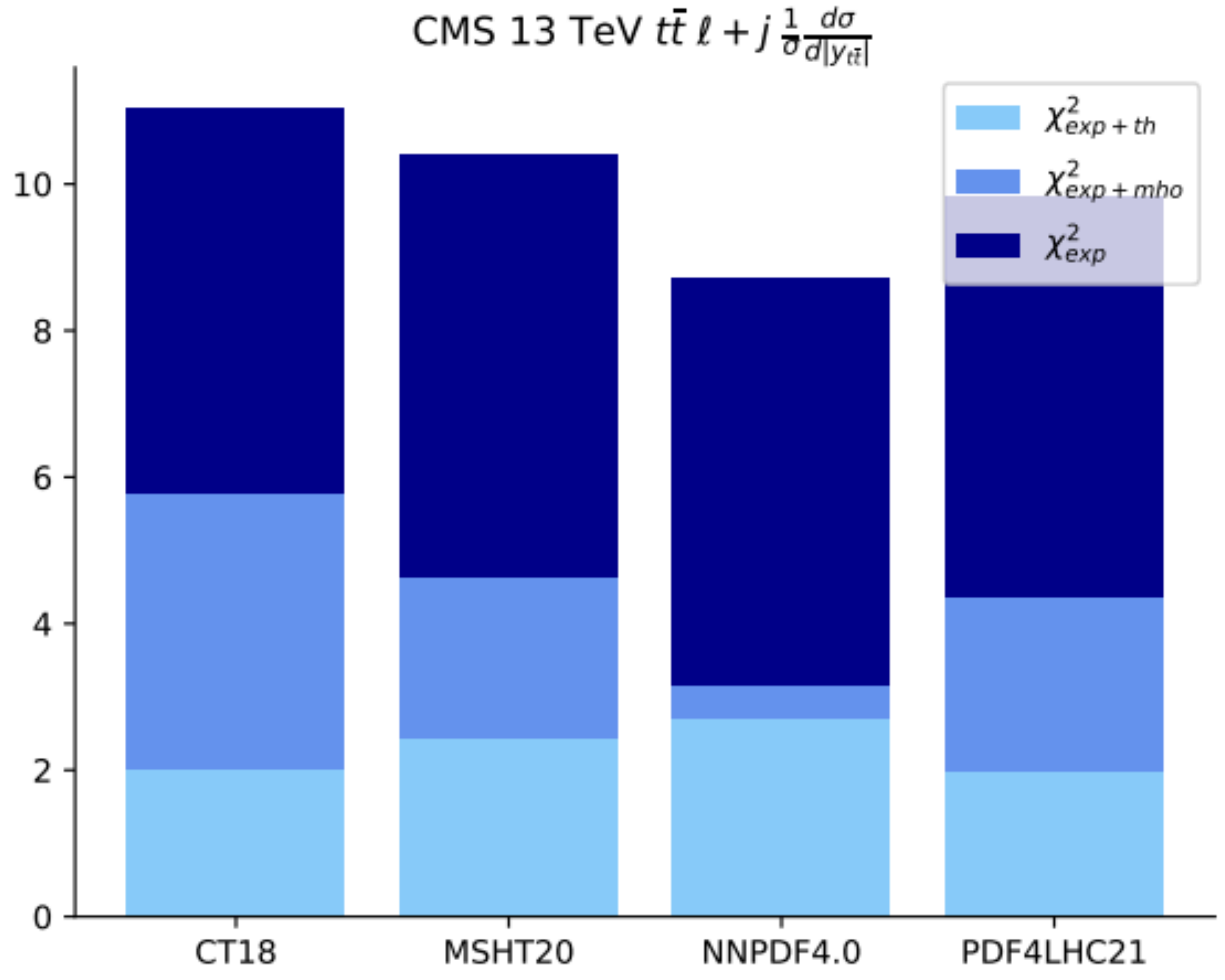
HOW WELL DO MODERN PDF SETS PERFORM?

Chiefa et al: 2501.10359



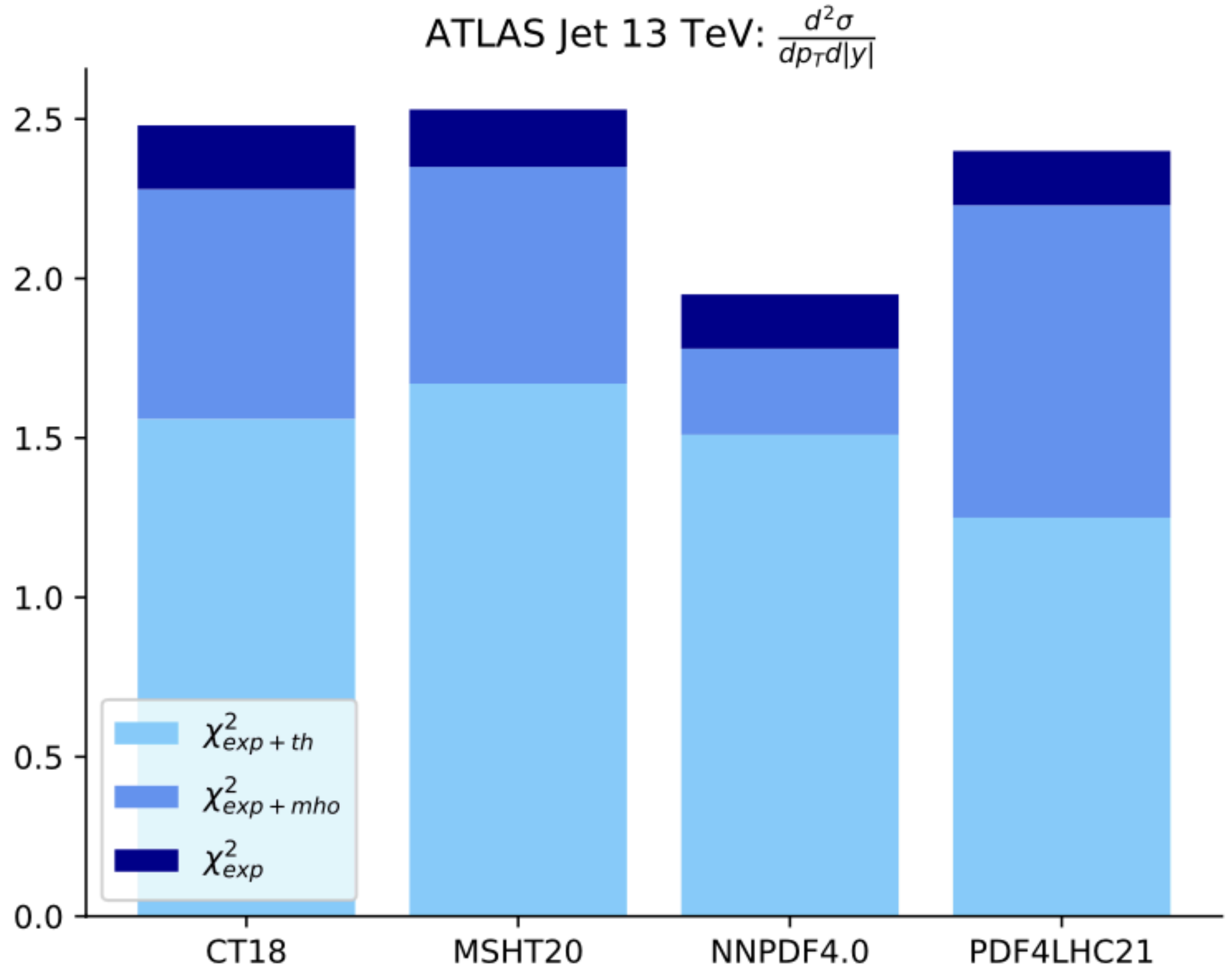
LHCb Z forward production at 13 TeV

LHCb collaboration [arXiv:2112.07458]



CMS tt~ production at 13 TeV, l+jet

CMS collaboration [arXiv:2108.02803]



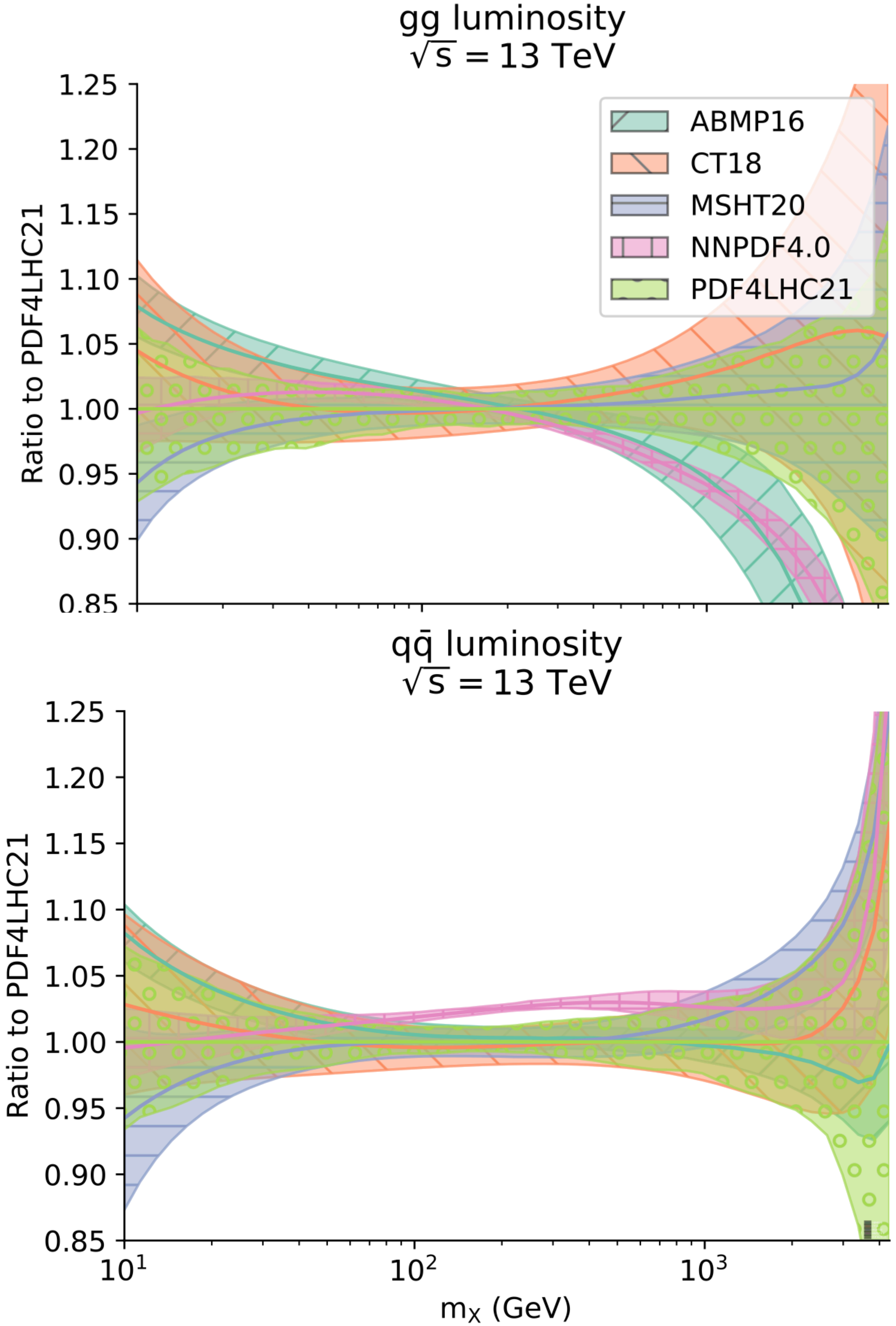
ATLAS incl. jet sec at 13 TeV, R=0.6, anti-kT

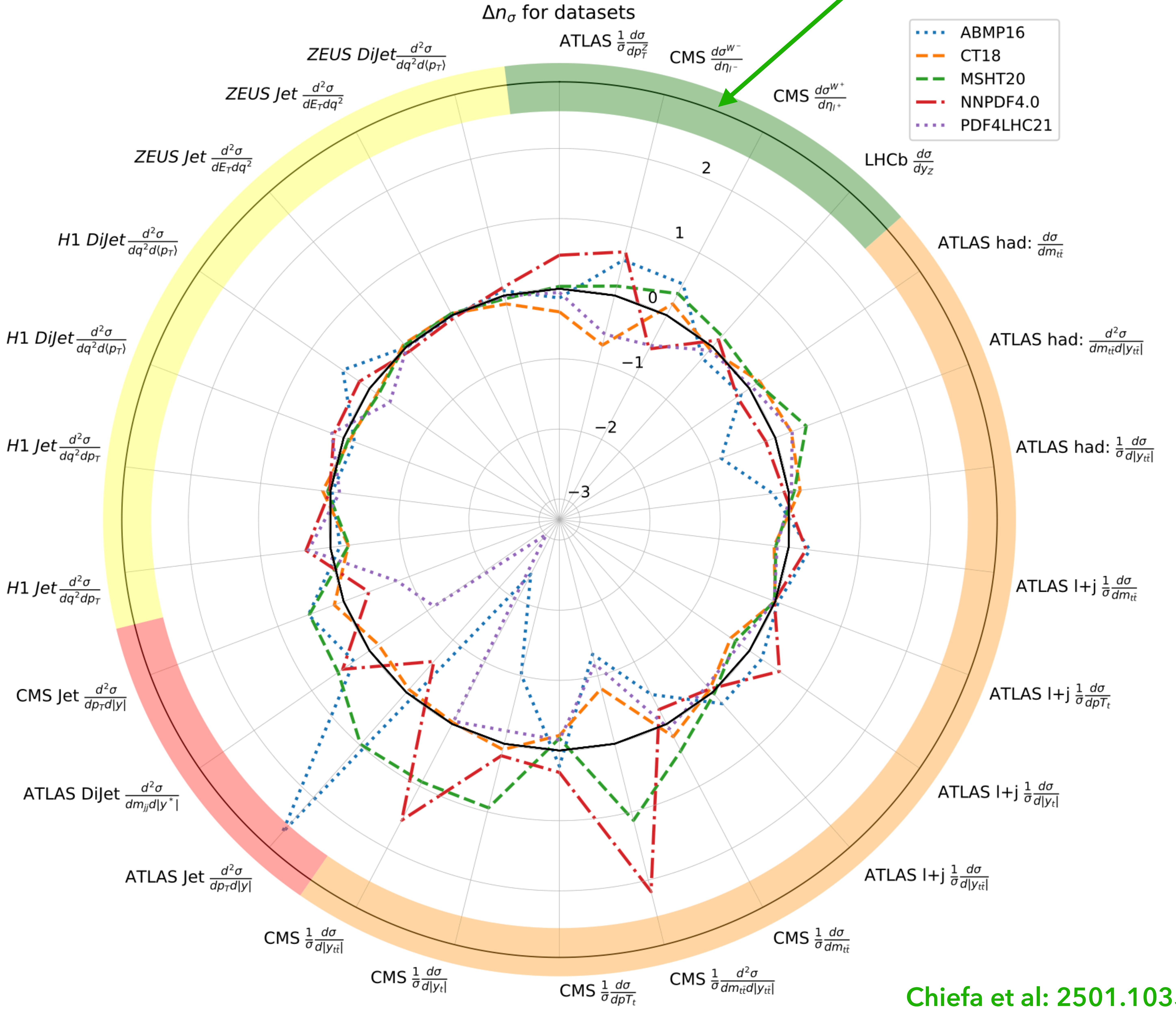
ATLAS collaboration [arXiv:1711.02692]

- Recently analysed 25 experimental distributions LHC@13 TeV and HERA jets measurements not yet included in any PDF analysis: inclusive jets, dijets, top differential distributions, W and Z production, Z pT, DIS jets [Chiefa et al, arXiv:2501.10359]
- Crucial to keep into account all sources of uncertainties and all bin-by-bin corrections.
- For some observables the inclusion of MHO uncertainties or PDF uncertainties changes the picture as compared to pure experimental chi2
- LHC@13 TeV measurements do not strongly discriminate among PDF sets so far.
- NNPDF4.0, despite being the most precise (smallest PDF uncertainties in data region) yields good agreement with data!

HOW WELL DO MODERN PDF SETS PERFORM?

$$\Delta n_\sigma^{(i)} = \frac{\chi^{2(i)}_{\mathrm{exp+th}} - \left\langle \chi^2_{\mathrm{exp+th}} \right\rangle_{\mathrm{pdfs}}}{\sqrt{2/n_{\mathrm{data}}}}$$



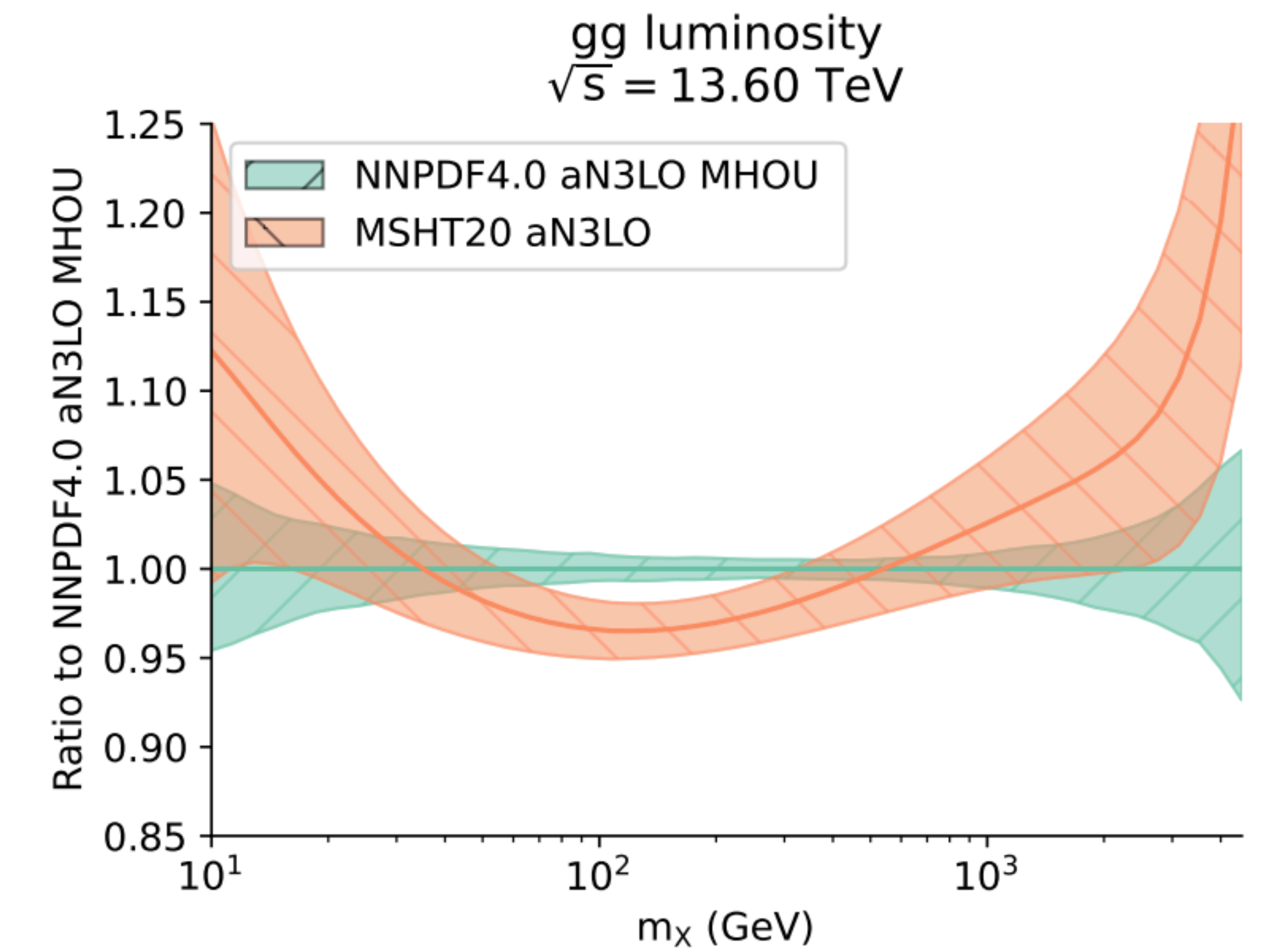


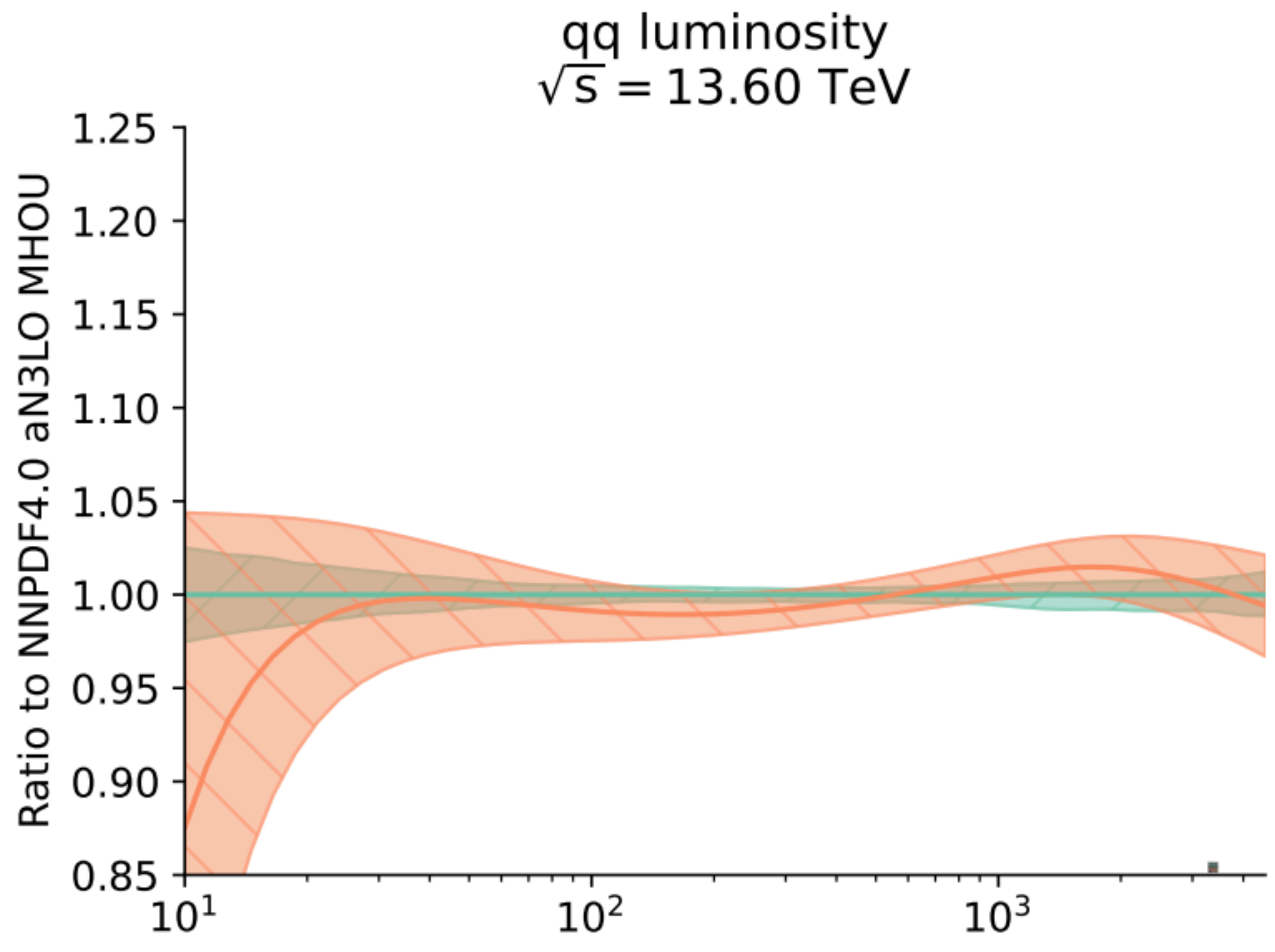
Chiefa et al: 2501.10359

TOWARDS N3LO PDFS

First N3LO PDF sets by MSHT [McGowan et al 2207.04739] and NNPDF [Ball et al 2402.18635] collaborations

- Order-by-order convergence improved in both PDF sets and in phenomenological predictions
- Rather good agreement between collaborations' results considering different input included and strategies for the approximation
- aN3LO necessary to match the order of computation of partonic cross sections



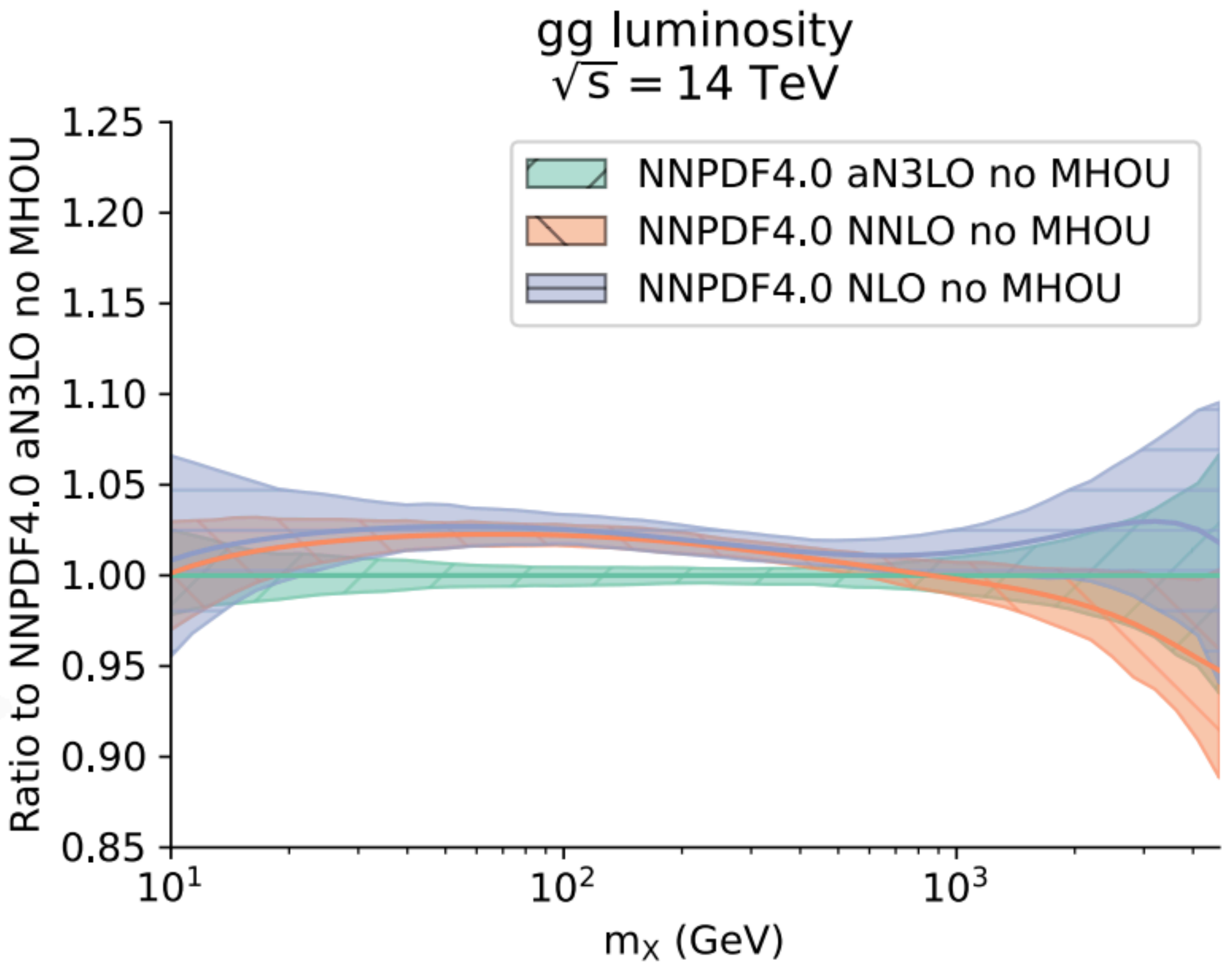


Cridge et al, arXiv:2411.05373

TOWARDS N3LO PDFS

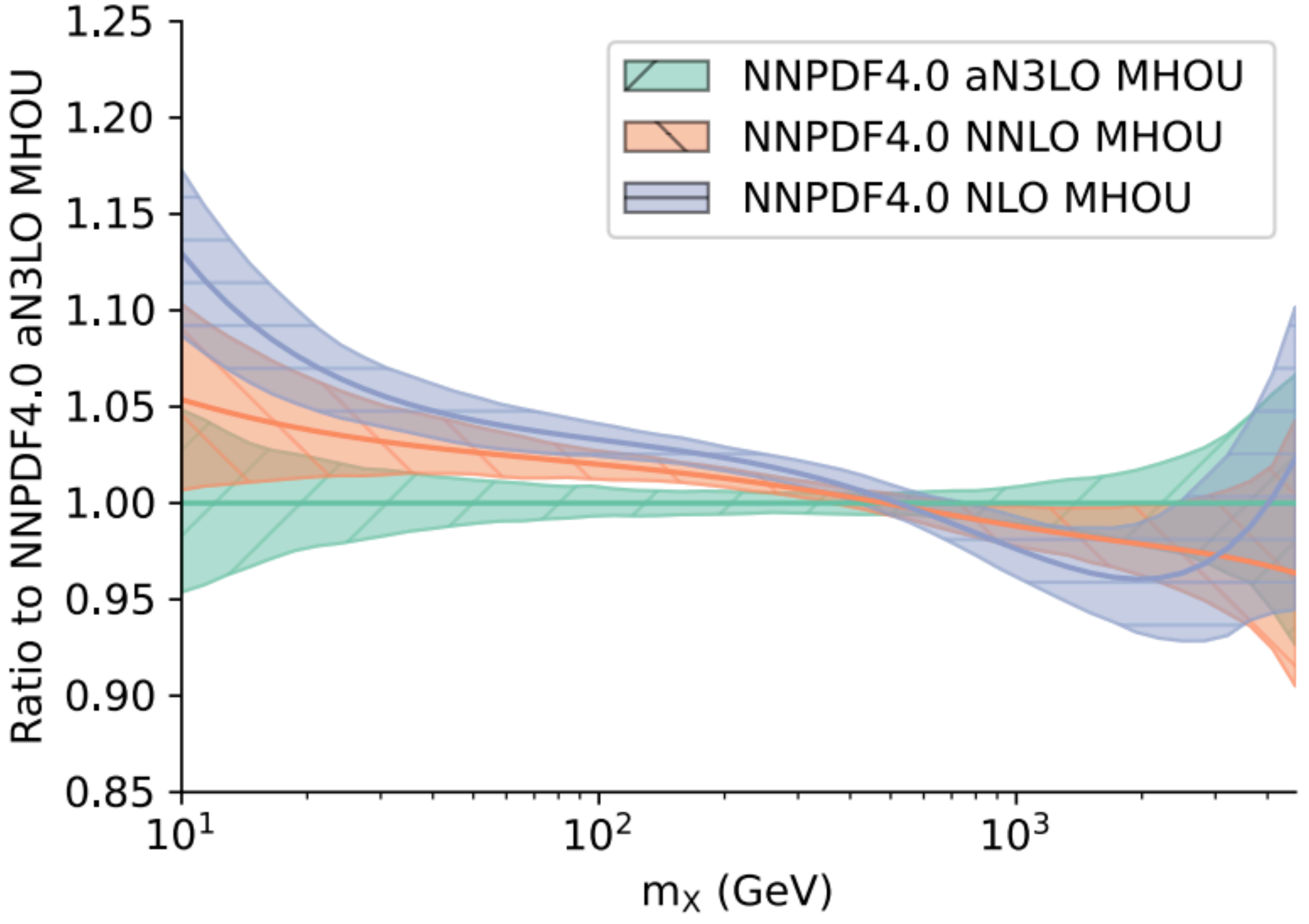
First N3LO PDF sets by MSHT [McGowan et al 2207.04739] and NNPDF [Ball et al 2402.18635] collaborations

- Order-by-order convergence improved in both PDF sets and in phenomenological predictions
- Rather good agreement between collaborations' results considering different input included and strategies for the approximation
- aN3LO necessary to match the order of computation of partonic cross sections



Results are not exact N3LO but rather aN3LO (approximated)

- Splitting functions not fully known, but known in enough limits and enough momenta to build a reasonable approximation [benchmark and references in arXiv:2406.16188]
- Fiducial pp predictions only accessible at NNLO (some only through K-factors)
- Effect of MHOUs estimated though scale variation greatly improves convergence



TOWARDS N3LO PDFS

Error estimate before aN3LO PDFs

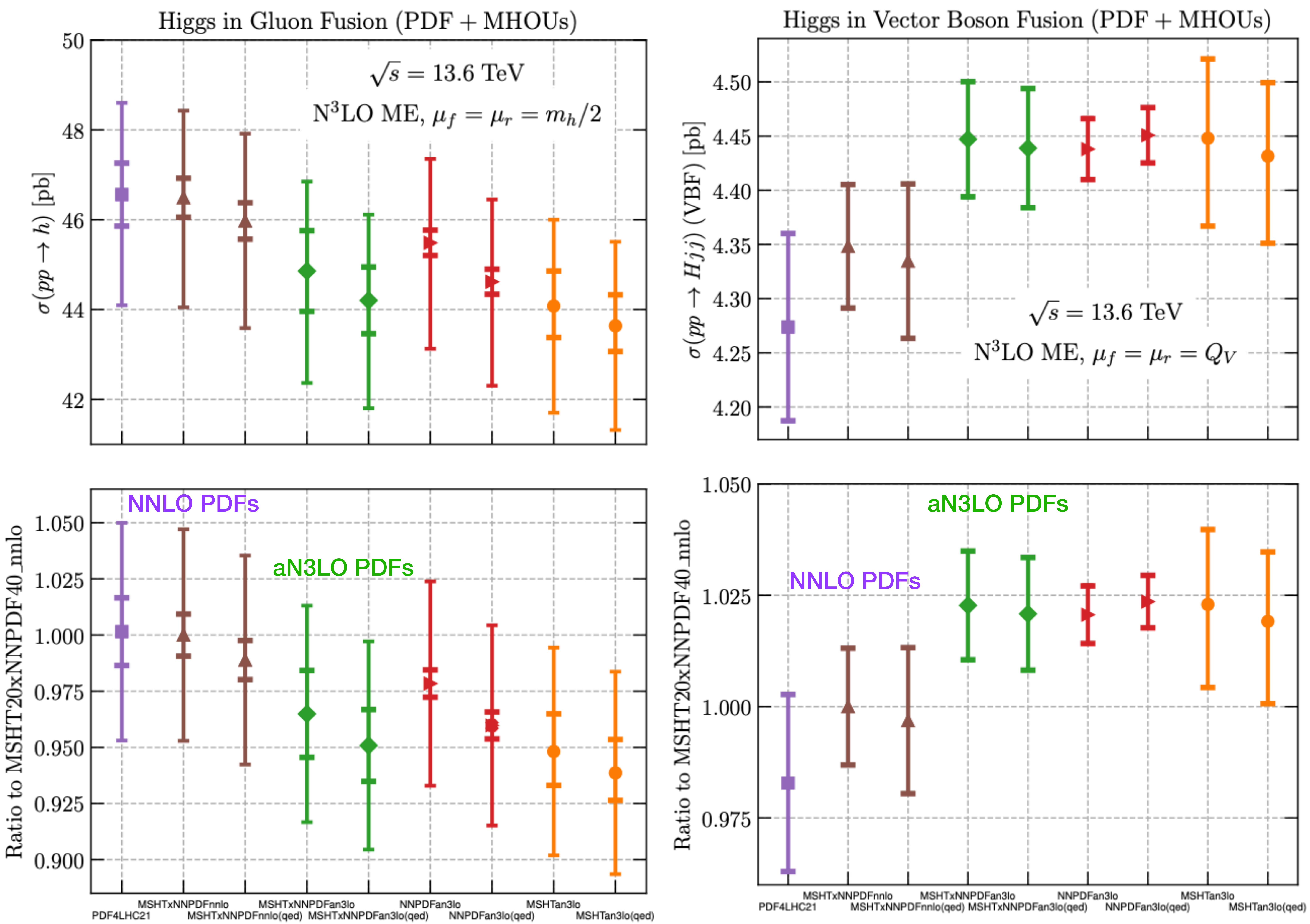
$$\Delta^{\rm app}_{\rm NNLO} \equiv \frac{1}{2}\left|\frac{\sigma^{\rm NNLO}_{\rm NNLO-PDF} - \sigma^{\rm NNLO}_{\rm NLO-PDF}}{\sigma^{\rm NNLO}_{\rm NNLO-PDF}}\right|$$

Actual error using aN3LO PDFs

$$\Delta^{\rm exact}_{\rm NNLO} \equiv \left|\frac{\sigma^{\rm N^3LO}_{\rm N^3LO-PDF} - \sigma^{\rm N^3LO}_{\rm NNLO-PDF}}{\sigma^{\rm N^3LO}_{\rm N^3LO-PDF}}\right|.$$

	ggF	VBF
$\Delta_{\rm NNLO}$ (app)	1.6%	0.5%
$\Delta_{\rm NNLO}$ (actual)	3.3%	2.3%

Previous estimates of the effect of NNLO/N3LO mismatch were **optimistic**



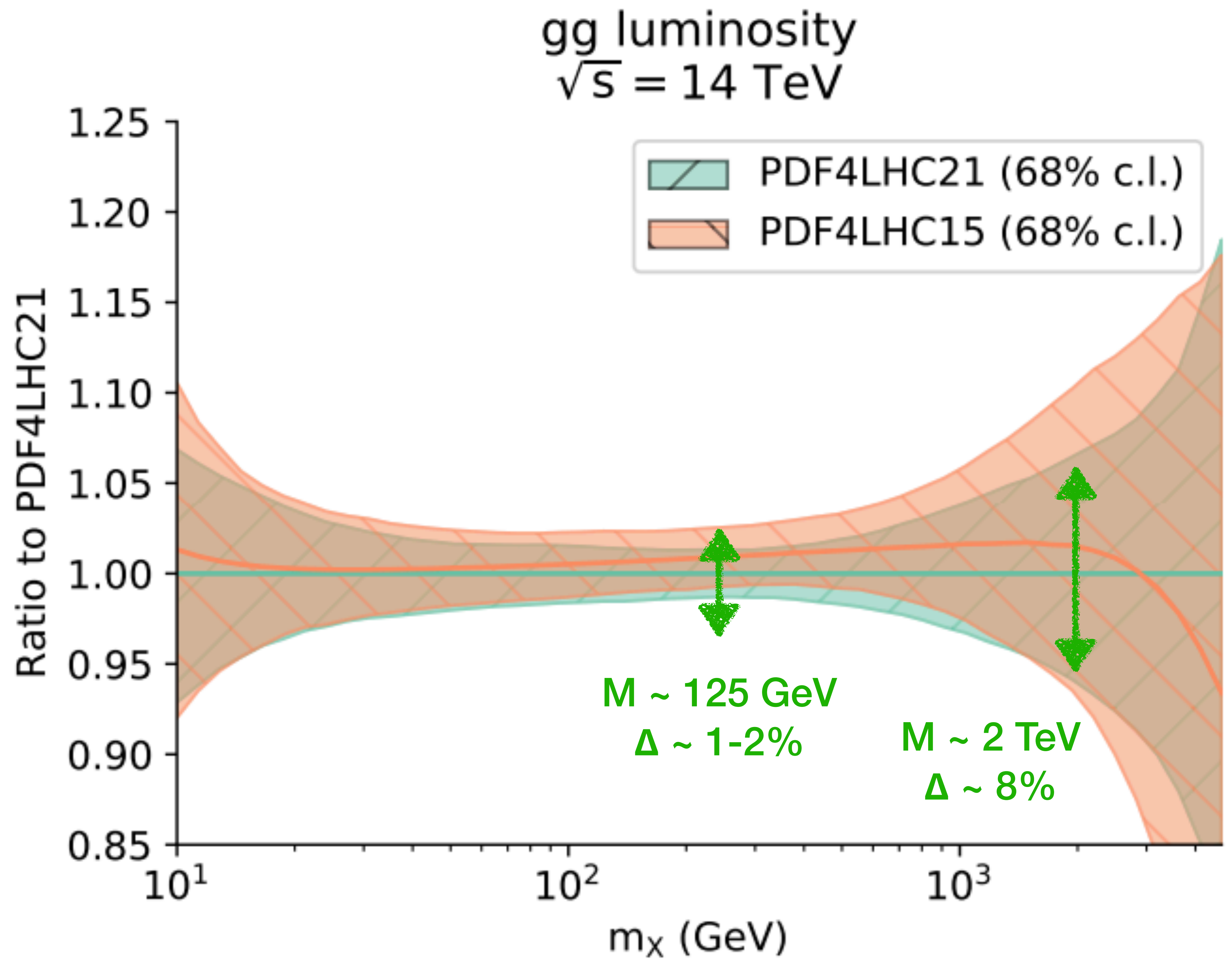
Cridge et al, arXiv:2411.05373

PART II: PRECISION VS ACCURACY – NEW CHALLENGES, NEW FRONTIERS

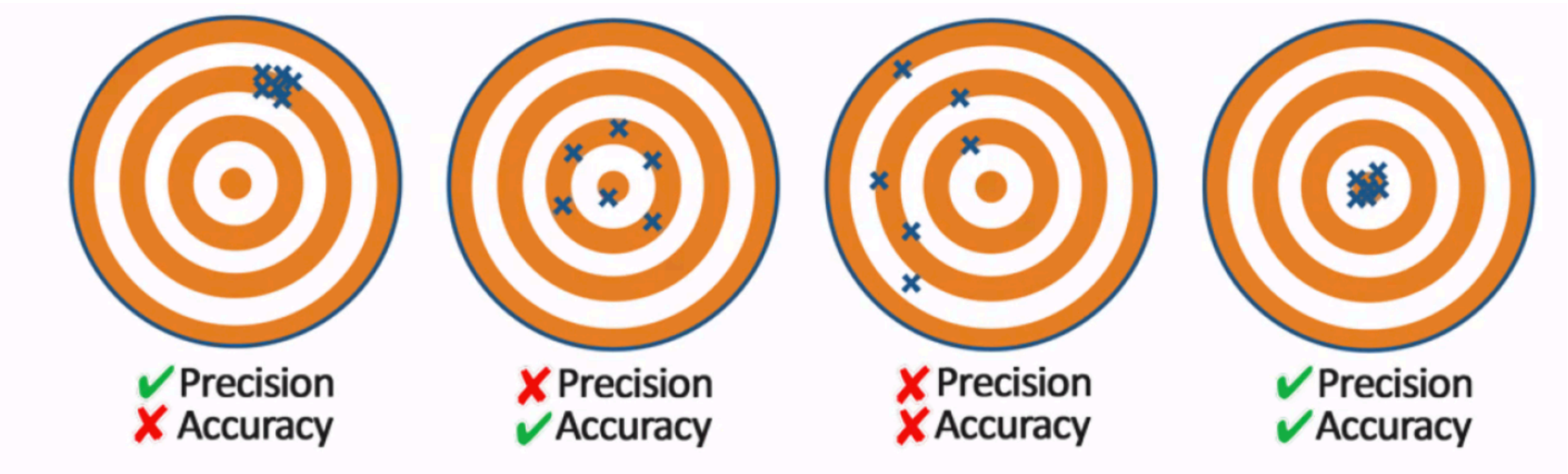
THE PRECISION VERSUS ACCURACY CHALLENGE

At a time of precision physics, it is time for precision PDF determination, does precision match accuracy?
Highly non trivial challenges

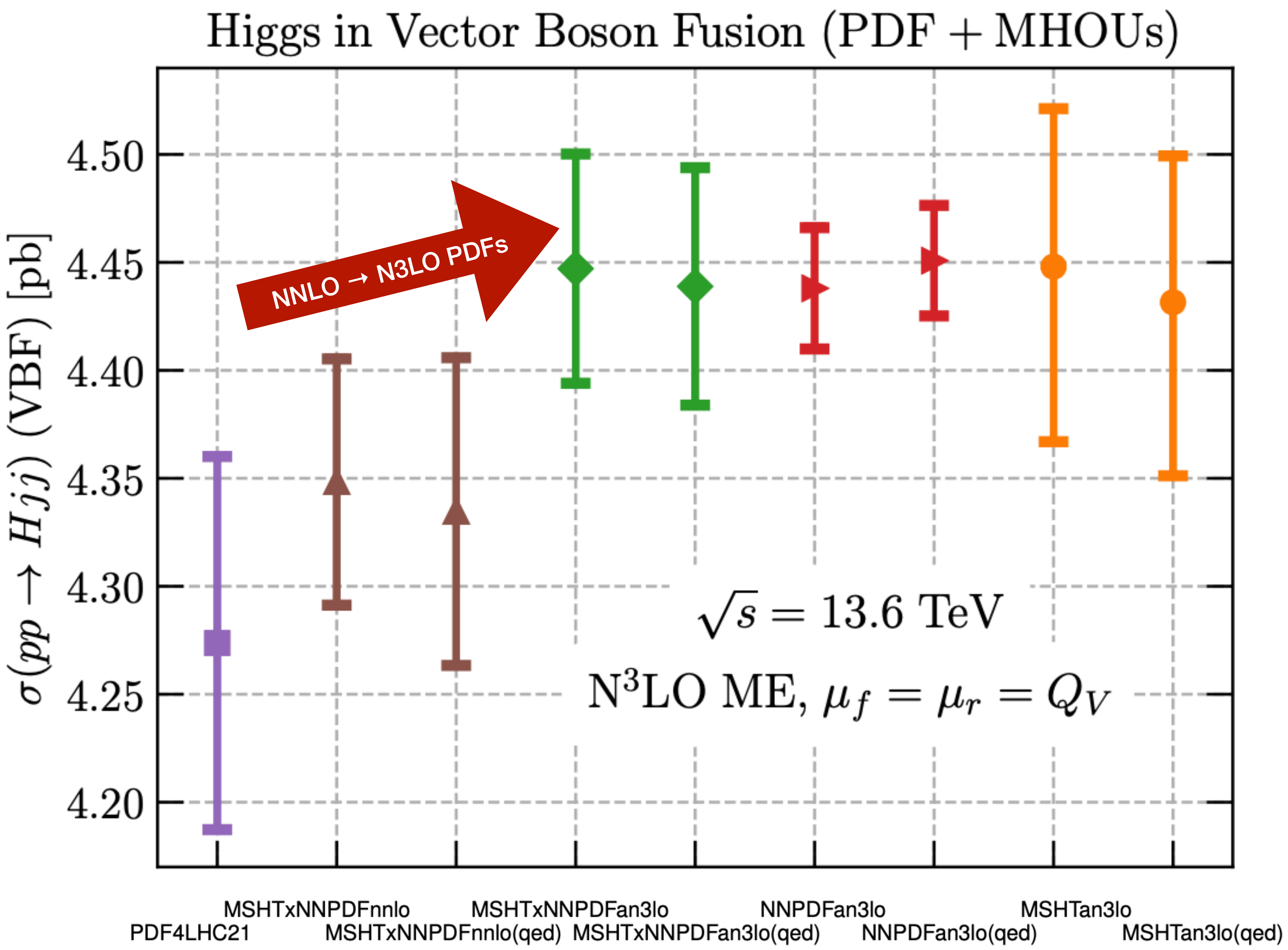
- Genuine inconsistencies in experimental inputs (possible underestimated, over-correlated systematics…)
- Imperfect fitting methodology (parametrisation, error propagation…)
- Inaccurate theoretical framework
  - ➡ Missing higher orders (exact N3LO, resummation…)
  - ➡ Other corrections (electroweak, nuclear, higher-twists, other non-perturbative effects…)
  - ➡ BSM effects in the high energy tails of distributions



PDF4LHC21 study 2203.05506



CHALLENGE #1: N3LO PDFS AND MISSING HIGHER ORDER UNCERTAINTIES



e.g. Higgs in VBF @ N3LO

MSHT+NNPDF arXiv:2411.05373

- Shift between results obtained with NNLO PDFs and with aN3LO PDFs more significant than expected
- Next steps:
    - As more N3LO theory ingredients become available, **aN3LO PDFs updated** by improving convergence of results obtained with different methodologies
    - **Exact N3LO evolution**, exact NNLO grids for hadronic coefficients and N3LO k-factors
- Every step of the way becomes computationally more costly and complex for a smaller and smaller improvements. But ***extremely needed!***
- What about other effects? Resummation, higher twists, power correction…
- Estimate of incomplete N3LO terms and MHO uncertainties still computed via scale variations: is there a better way?

CHALLENGE #2: DETERMINATION OF SM PARAMETERS

PDF set	$\alpha_s(m_Z)$	PDF uncertainty	g [GeV^2]	q [GeV^4]
MSHT20 [37]	0.11839	0.00040	0.44	−0.07
NNPDF4.0 [84]	0.11779	0.00024	0.50	−0.08
CT18A [29]	0.11982	0.00050	0.36	−0.03
HERAPDF2.0 [65]	0.11890	0.00027	0.40	−0.04

ATLAS collaboration, arXiv:2309.12986

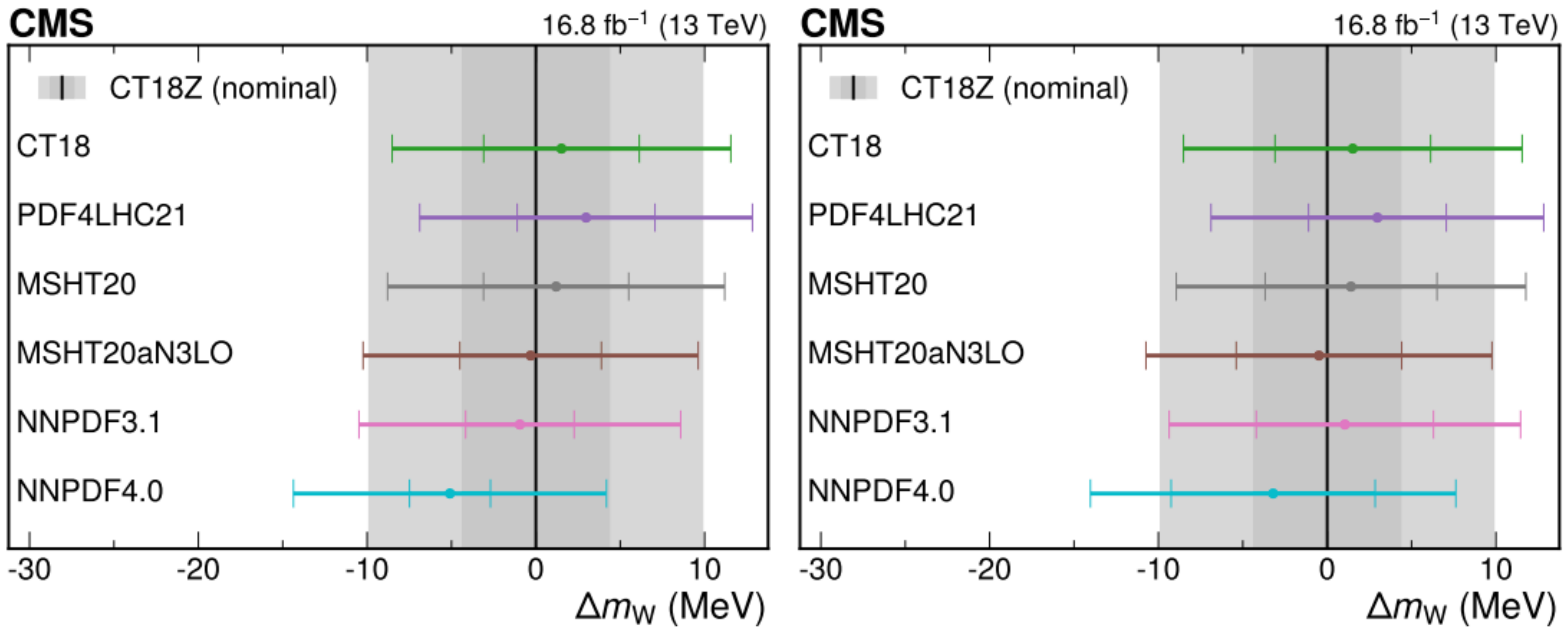
Significantly different results obtained when using different input PDFs.

- ATLAS determination of α_S from neutral-current Drell-Yan Z pT measurement most precise α_S determination ever performed from a single experiment
  $\Delta_{\alpha,\mathrm{TOT}} = 9 \times 10^{-4}$ with $\Delta_{\alpha,\mathrm{PDF}} = 5 \times 10^{-4}$ (MSHT20aN3LO)
  But
  $\delta_{\alpha,\mathrm{PDF}} = 20 \times 10^{-4}$ (shift CT18A versus NNPDF4.0)

CHALLENGE #2: DETERMINATION OF SM PARAMETERS

PDF set	$\alpha_s(m_Z)$	PDF uncertainty	g [GeV^2]	q [GeV^4]
MSHT20 [37]	0.11839	0.00040	0.44	−0.07
NNPDF4.0 [84]	0.11779	0.00024	0.50	−0.08
CT18A [29]	0.11982	0.00050	0.36	−0.03
HERAPDF2.0 [65]	0.11890	0.00027	0.40	−0.04

ATLAS collaboration, arXiv:2309.12986



CMS collaboration, arXiv: 2412.13872

Significantly different results obtained when using different input PDFs.

- ATLAS determination of α_S from neutral-current Drell-Yan Z pT measurement most precise α_S determination ever performed from a single experiment
  $\Delta_{\alpha,TOT} = 9 \times 10^{-4}$ with $\Delta_{\alpha,PDF} = 5 \times 10^{-4}$ **(MSHT20aN3LO)**
  But
  $\delta_{\alpha,PDF} = 20 \times 10^{-4}$ **(shift CT18A versus NNPDF4.0)**

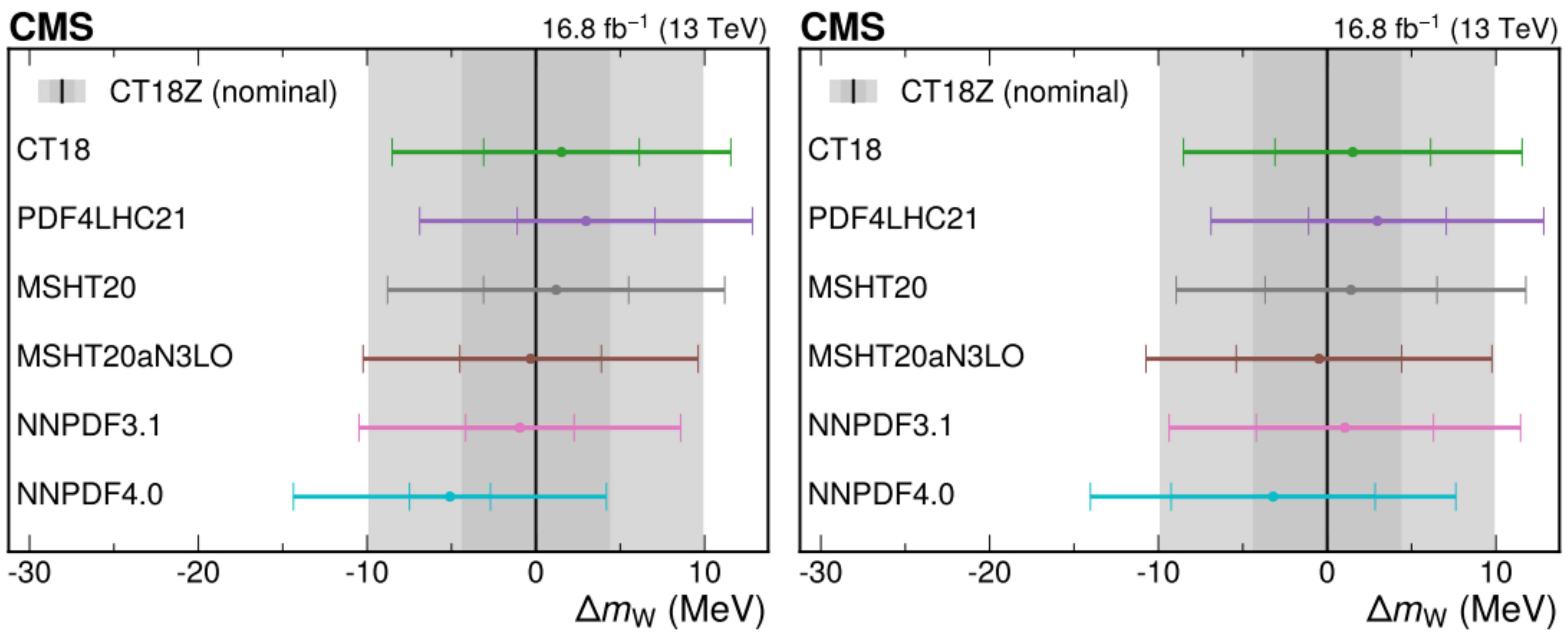
- Similar situation in Mw determination **[ATLAS-CONF-2023-004]**
  $\Delta_{MW,PDF} = 7.7$ (14.6) MeV in $p_{T,l}$ (m_T) channel **(CT18)**
  But
  $\delta_{MW,PDFS} = 17$ (21) MeV **(shift MSHT20 versus NNPDF4.0)**

- CMS Mw: inflate PDF uncertainties by a factor and profile with CMS data decreases shift between PDF set (but effectively giving much larger weight to a given measurement as compared to all experiments that determine PDFs in global fit)

CHALLENGE #2: DETERMINATION OF SM PARAMETERS

PDF set	$\alpha_s(m_Z)$	PDF uncertainty	g [GeV^2]	q [GeV^4]
MSHT20 [37]	0.11839	0.00040	0.44	−0.07
NNPDF4.0 [84]	0.11779	0.00024	0.50	−0.08
CT18A [29]	0.11982	0.00050	0.36	−0.03
HERAPDF2.0 [65]	0.11890	0.00027	0.40	−0.04

ATLAS collaboration, arXiv:2309.12986



CMS collaboration, arXiv: 2412.13872

Can a better solution be achieved?
Simultaneous PDF-SM fits? Profiling over all data?

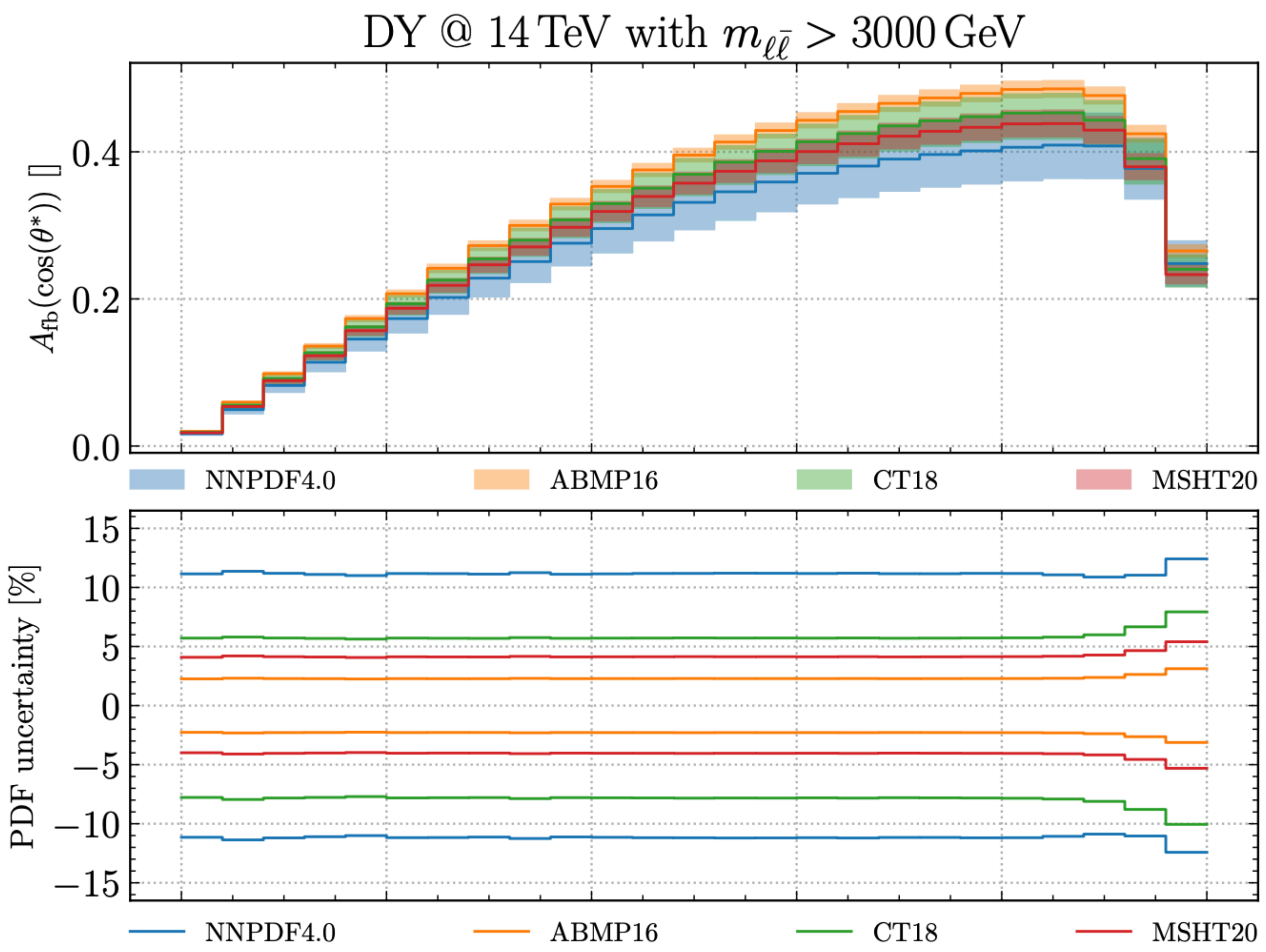
Significantly different results obtained when using different input PDFs.

- ATLAS determination of α_s from neutral-current Drell-Yan Z pT measurement most precise α_s determination ever performed from a single experiment
  $\Delta_{\alpha,\mathrm{TOT}} = 9 \times 10^{-4}$ with $\Delta_{\alpha,\mathrm{PDF}} = 5 \times 10^{-4}$ **(MSHT20aN3LO)**
  But
  $\delta_{\alpha,\mathrm{PDF}} = 20 \times 10^{-4}$ **(shift CT18A versus NNPDF4.0)**

- Similar situation in Mw determination **[ATLAS-CONF-2023-004]**
  $\Delta_{\mathrm{MW,PDF}} = 7.7\ (14.6)$ MeV in $p_{T,l}\,(m_T)$ channel **(CT18)**
  But
  $\delta_{\mathrm{MW,PDFS}} = 17\ (21)$ MeV **(shift MSHT20 versus NNPDF4.0)**

- CMS Mw: inflate PDF uncertainties by a factor and profile with CMS data decreases shift between PDF set (but effectively giving much larger weight to a given measurement as compared to all experiments that determine PDFs in global fit)

CHALLENGE #3: PRECISE PDFS FOR NEW PHYSICS SEARCHES

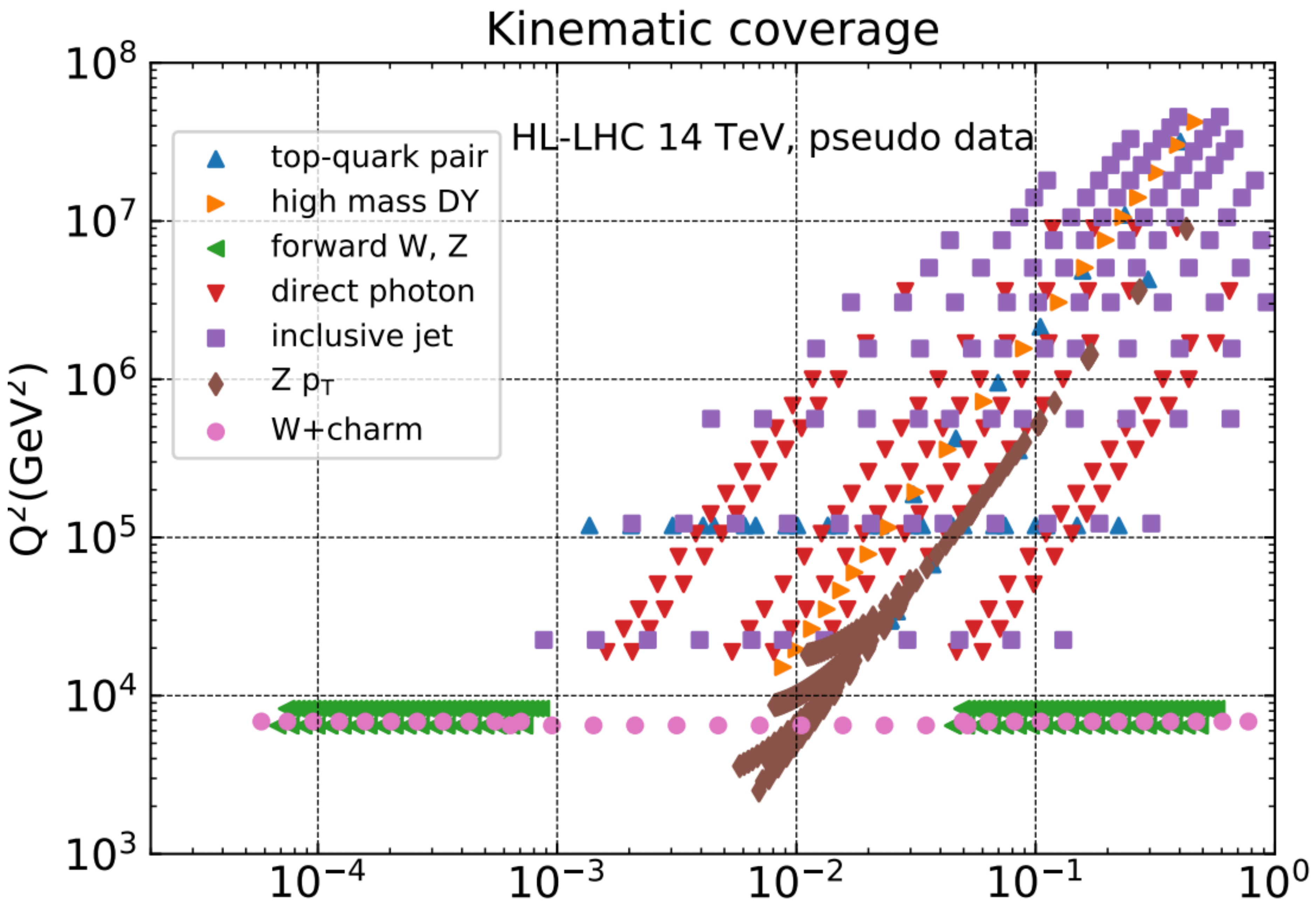


e. g. Heavy Z′ search in DY Forward-Backward asymmetry

Ball et al arXiv: 2209.08115

- Large-x PDF affected by large uncertainties
- Reducing PDF uncertainties in large-x region is essential to characterise any possible new physics and to interpret deviations in the tails of experimental distributions
- Looking forward: **HL-LHC (and FCC-hh)**

Large x ⟷ Large E and/or Large Y



Khalek et al, 1810.03639

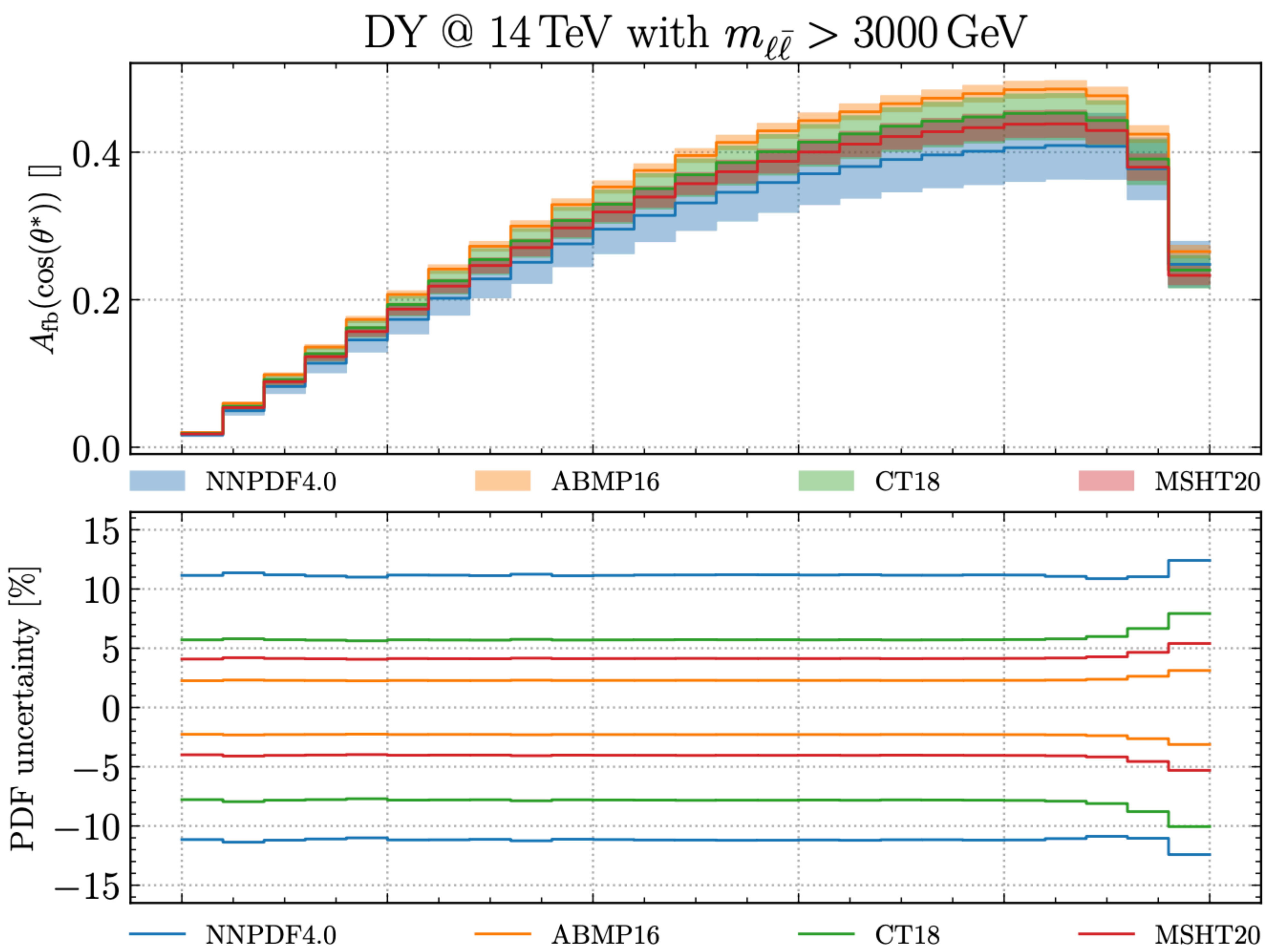
$$x \approx \frac{M}{\sqrt{S}}$$

High-mass final states

⇔

Large-x PDFs

CHALLENGE #3: PRECISE PDFS FOR NEW PHYSICS SEARCHES



e. g. Heavy Z′ search in DY Forward-Backward asymmetry

Ball et al arXiv: 2209.08115

$$\boxed{x \approx \frac{M}{\sqrt{S}}}$$

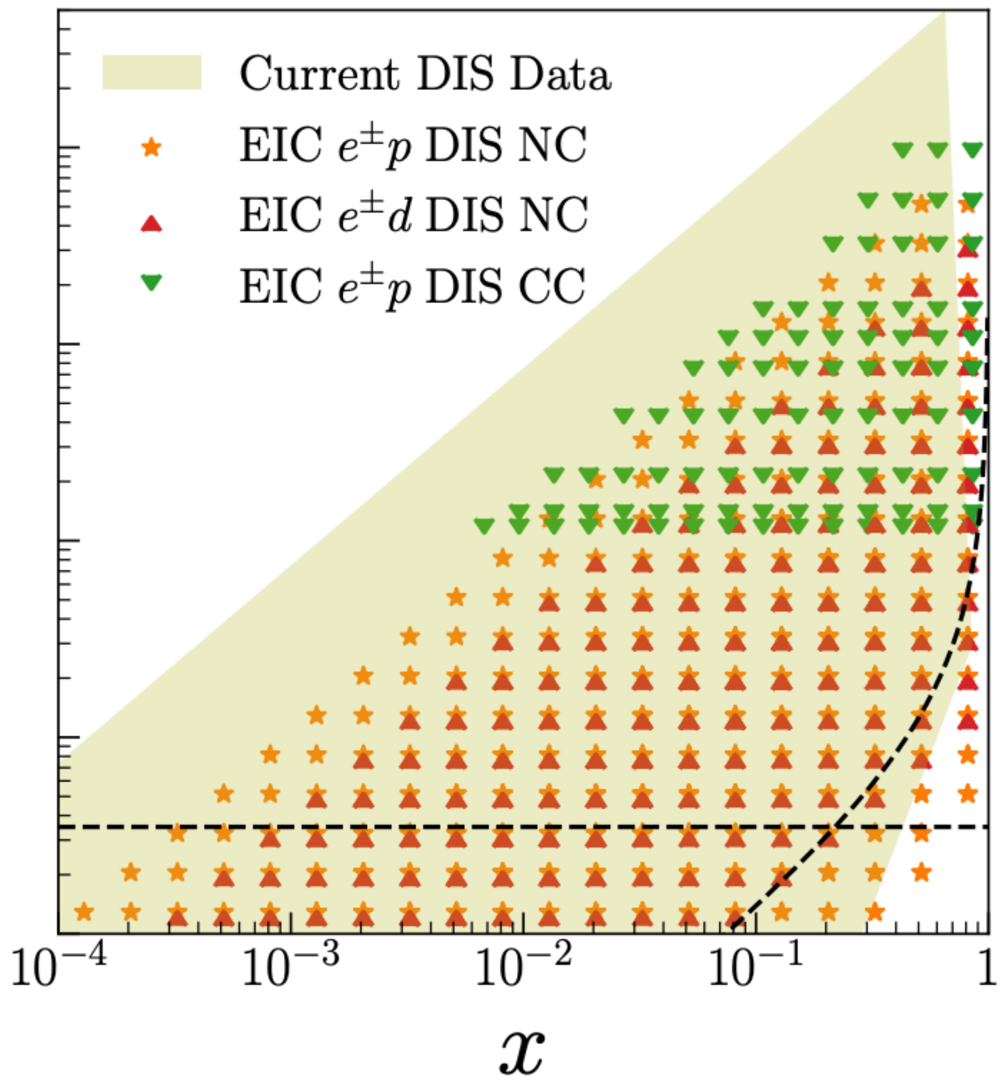
High-mass final states

⇔

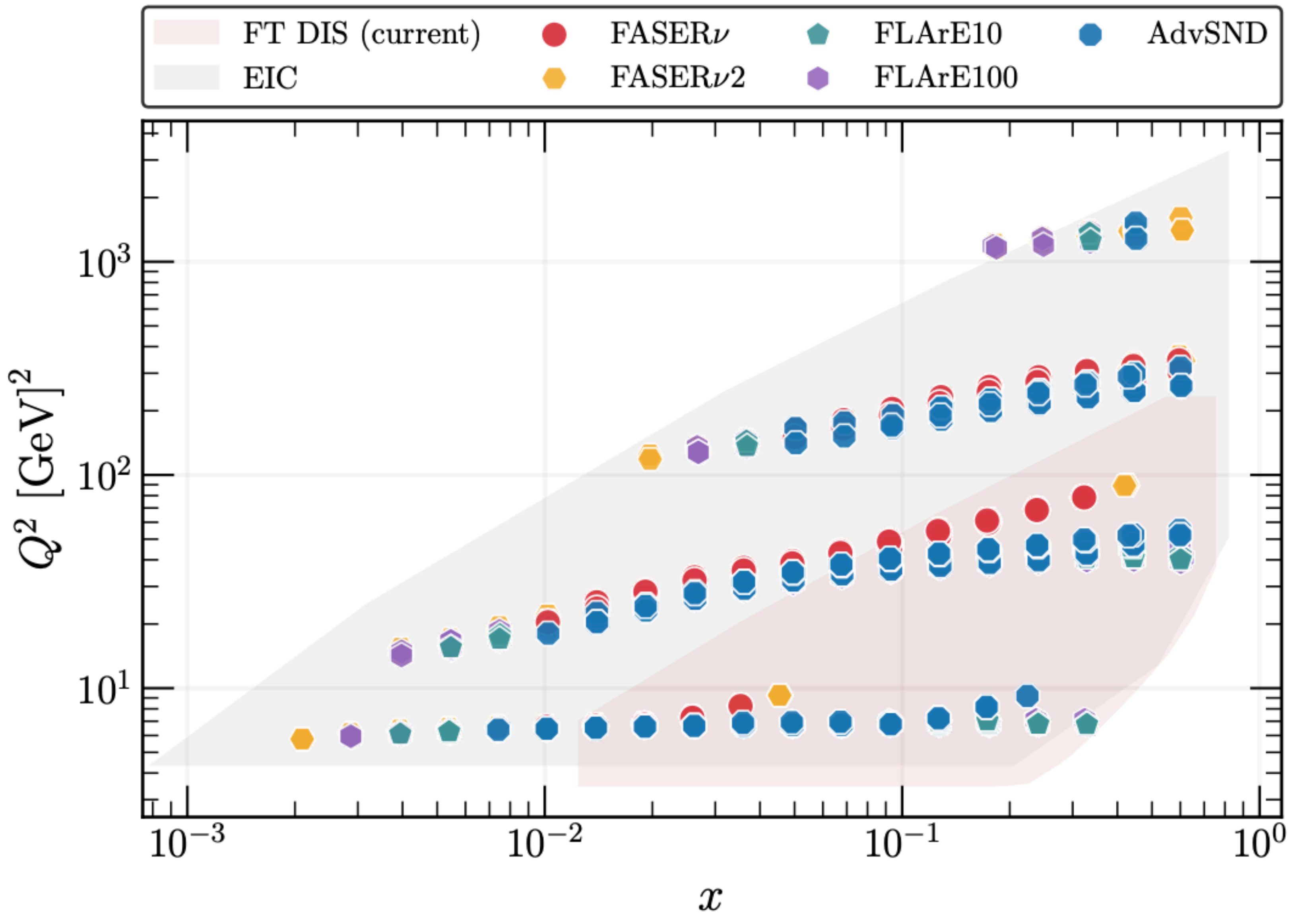
Large-x PDFs

- Large-x PDF affected by large uncertainties
- Reducing PDF uncertainties in large-x region is essential to characterise any possible new physics and to interpret deviations in the tails of experimental distributions
- Looking forward: **JLab, EIC, Forward Physics Facilities…**

Large x ⟷ Low/Medium E

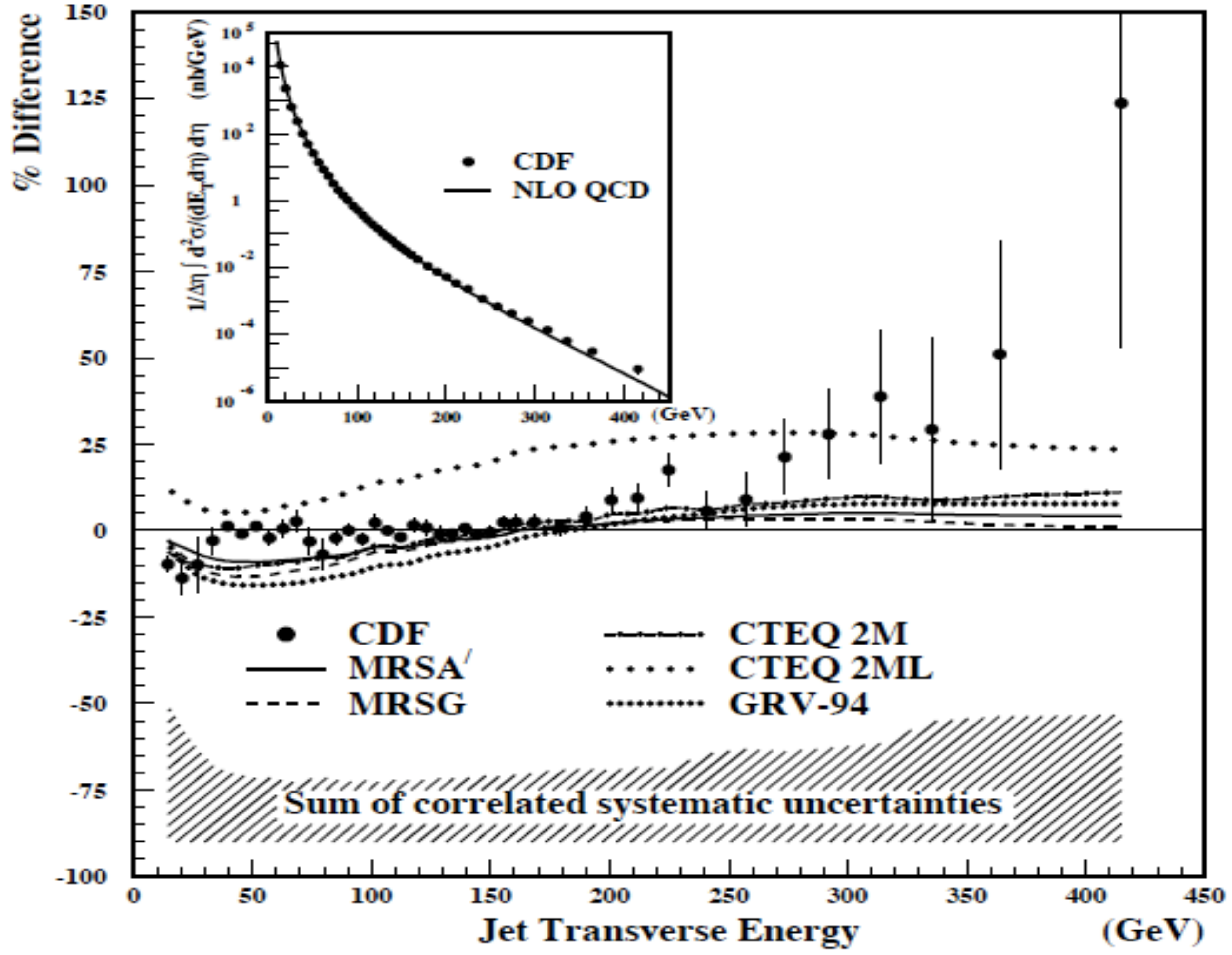


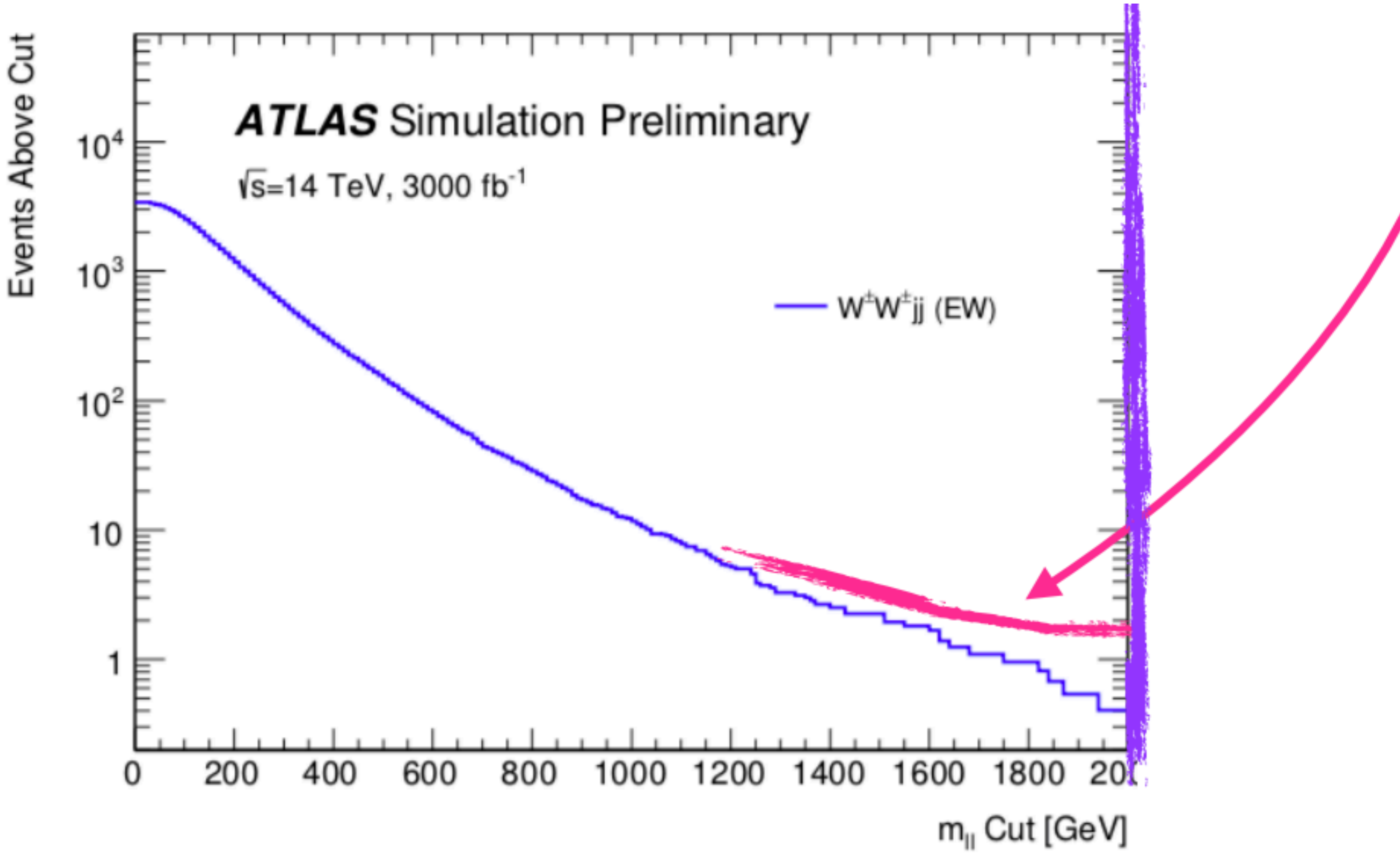
Khalek et al, 2102.00018



J Cruz-Martinez et al 2309.09581

CHALLENGE #3: PRECISE PDFS FOR NEW PHYSICS SEARCHES





CA Lee, HL/HE-LHC Jamboree, 1 March 2019

➡ Discrepancy between QCD calculation and CDF jets data (1995)

At that time no information on PDF uncertainties and theory predictions strongly depended on gluon shape at x>0.1. Once data included in the CTEQ fit, discrepancy disappeared.

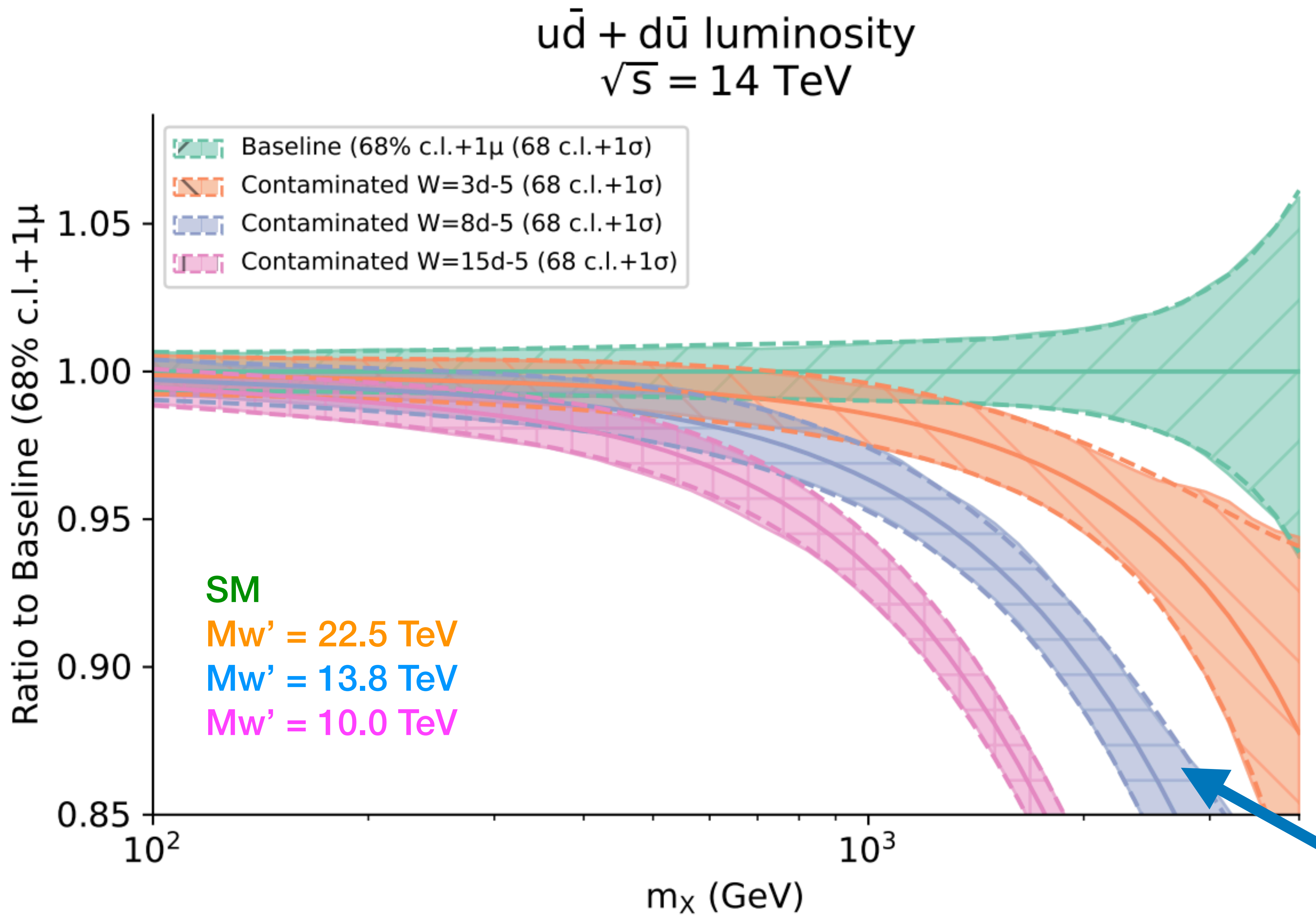
➡ Deviations from SM predictions in high energy tails (>2025)

New physics or limited understanding of proton structure?

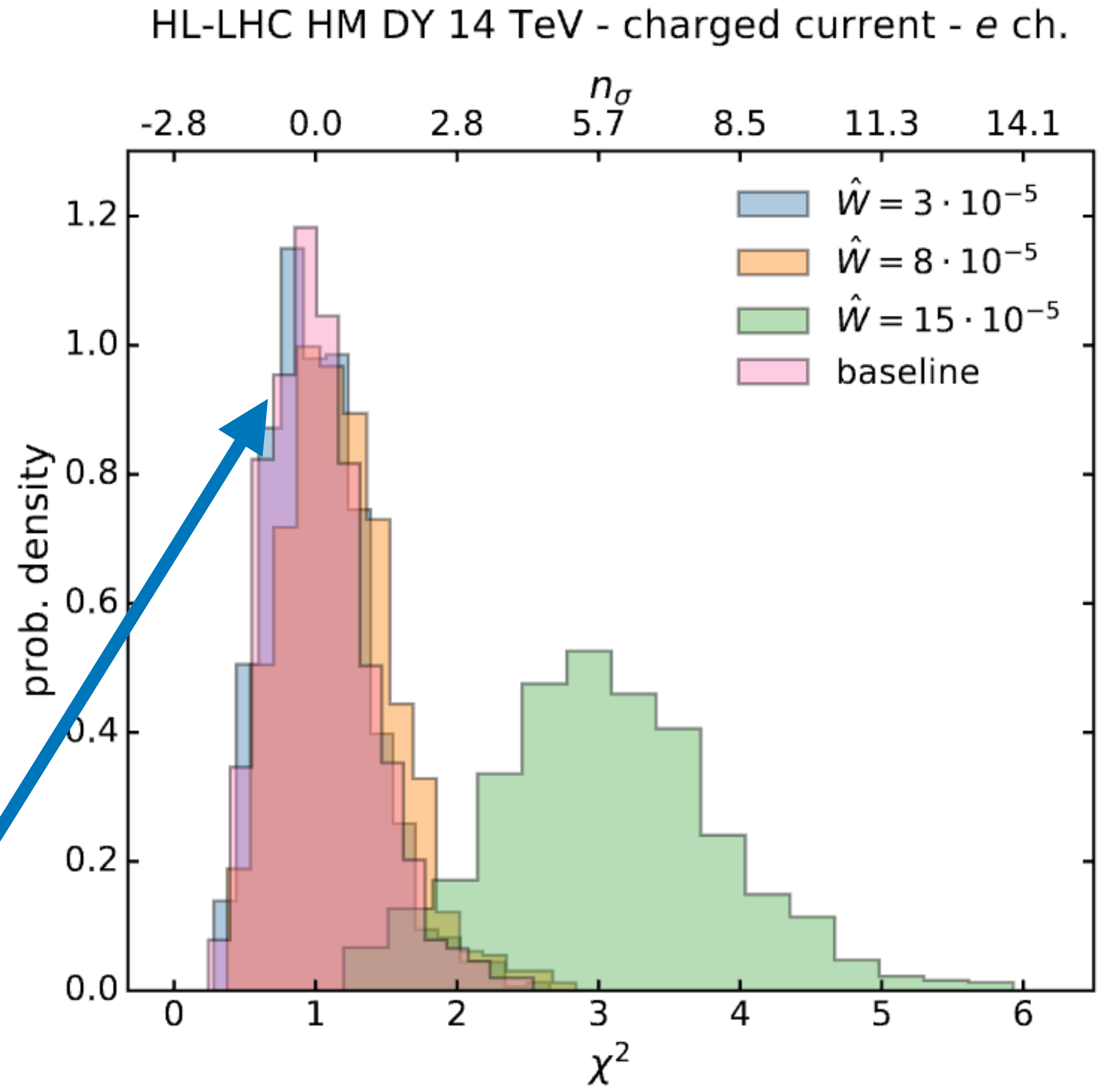
CHALLENGE #3: PRECISE PDFS FOR NEW PHYSICS SEARCHES

- Inject the "true" NP model and "true" PDFs in artificial data included in global fit
- Generate **HL-LHC** pseudo-data assuming
  "true" law of nature = "true" PDFs + "true" UV model
- Fit PDFs assuming SM
- **Can PDFs absorb signs of new physics? Yes**
- Important to tackle the issue (simultaneous fits, complementary constraints, test observables) and broaden understanding (gluon, strange)

Z'	W'
Selection test: ✗	Selection test: ✓
➡ Excluded from PDF fit	➡ Included in PDF fit
No impact on PDFs	**PDFs contaminated**



Max BSM bias allowed by global fit Without spoiling χ^2



Hammou, Madigan, Mangano, Mantani, Moore, Morales, MU 2307.10370

CONCLUSIONS AND OUTLOOK

- In an era of precision at LHC, need precise and accurate PDFs
- Impressive advances in theory inputs and experimental inputs, pressure to validate and understand methodological differences.
- Tools like closure tests and generalisation tests are crucial to test methodology robustness, effects of possible experimental inconsistencies and even possible effects of new physics in the high energy tails
- Crucial reproducibility & open-source tools (and fitting codes)

- Much to look forward to at DIS2025 as far as PDFs are concerned

Updates from collaborations - R. Thorne, O Zenaiev, T Hobbs, T R Rabemananjara WG1
HQ schemes - V. Bertone WG1
Determination of αS and PDFs - R Stegeman, T Cridge WG1
aN3LO combination - T. Cridge WG1
HL-LHC and other experiments synergies - K Wichmann, MU WG1/6
Higher twists - A Accardi WG1
Parametrisation studies - A Courtoy WG1
Closure tests with inconsistent data - MU WG1
Bayesian fits of PDFs - L Mantani WG1

… and much more!!