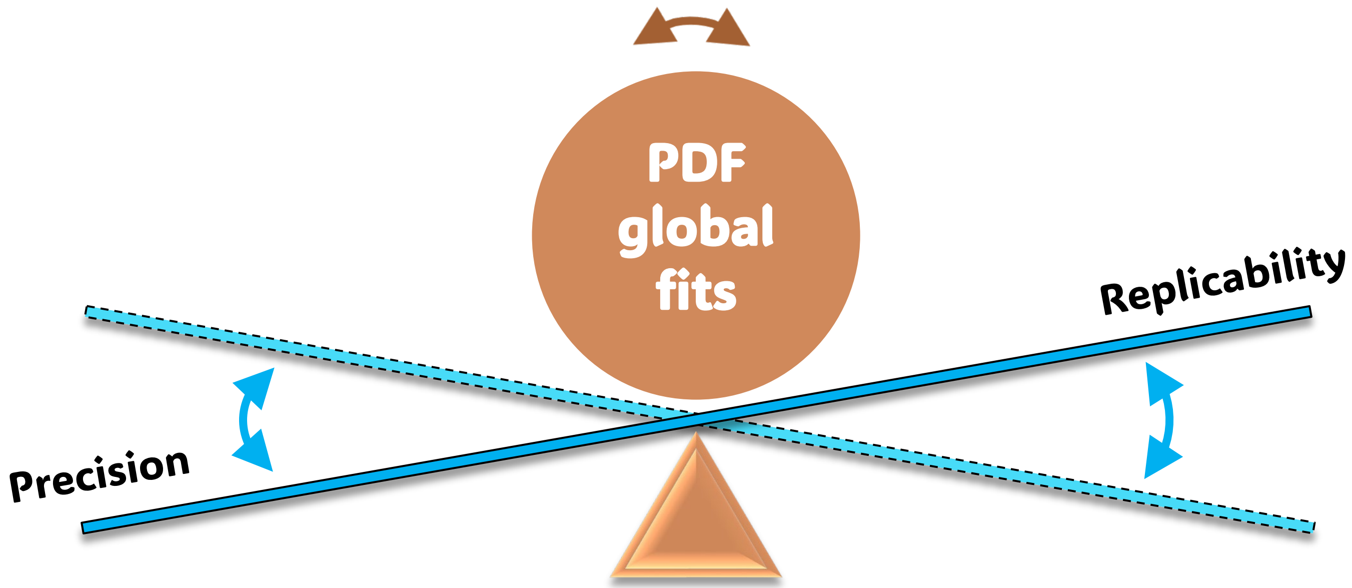
Unpolarized nucleon parton distributions: balancing precision and replicability

Pavel Nadolsky

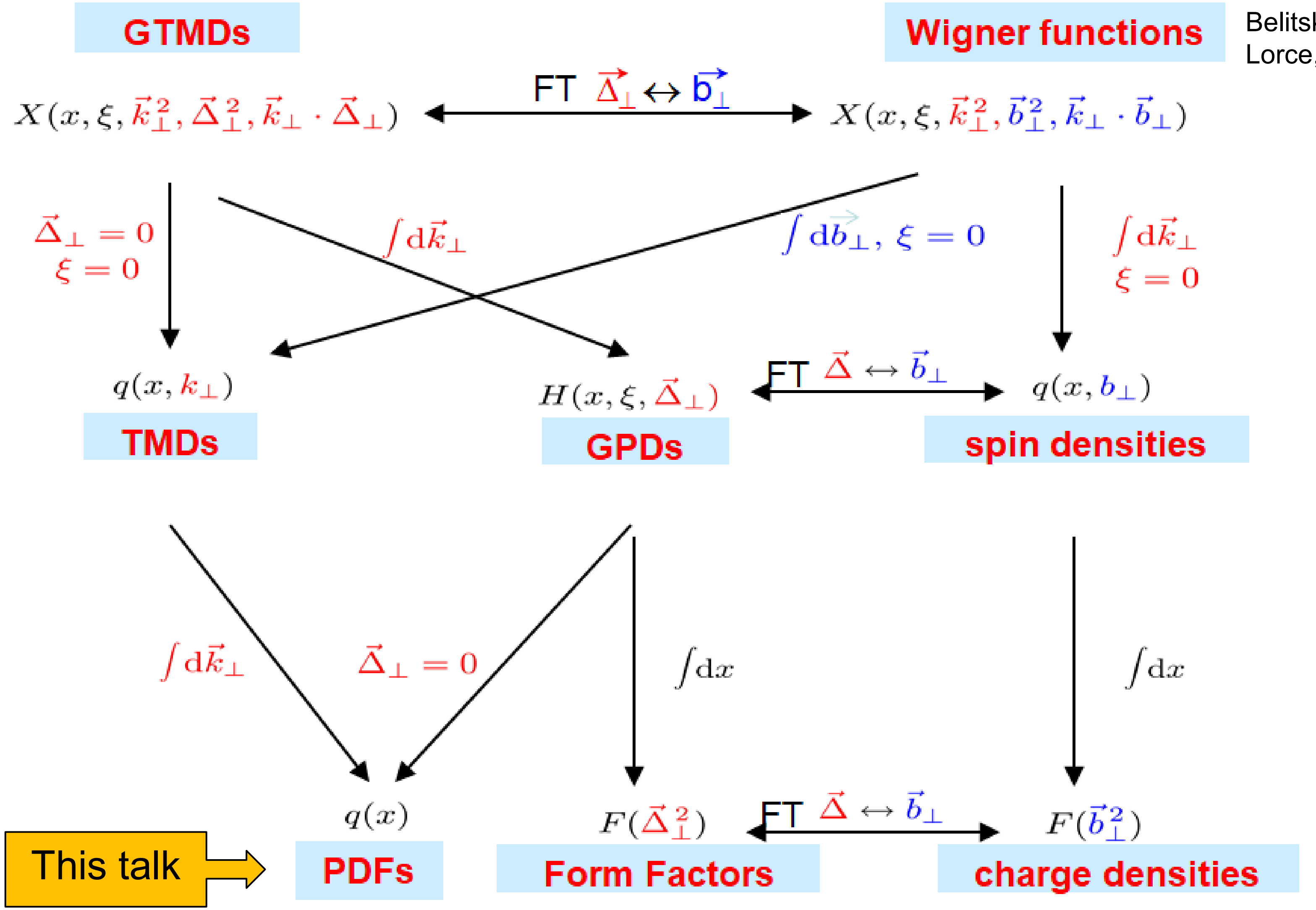
Southern Methodist University

With A. Courtoy, T. J. Hobbs, M. Guzzi, J. Huston, L. Kotz, F. Olness, M. Ponce Chavez, K. Xie, M. Yan, C.-P. Yuan,

and CTEQ-TEA (Tung Et Al.) Global QCD analysis group



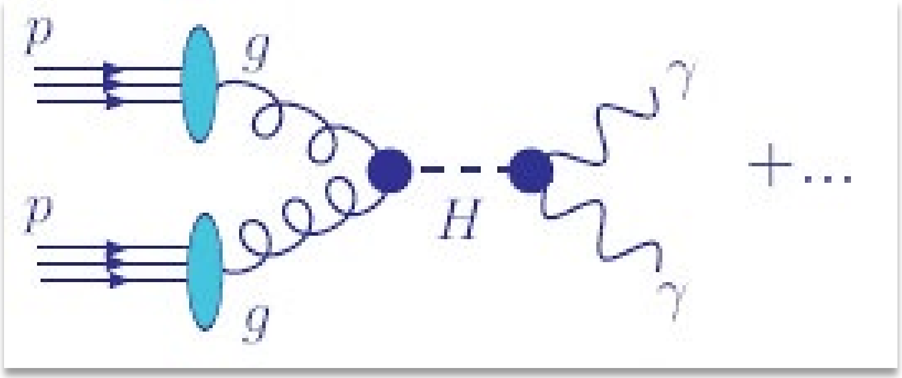
PDFs are the simplest nonperturbative functions for hadron structure



Belitsky, Ji, Yuan (2003)
Lorce, Pasquini (2011)

Figure: B. Pasquini, C. Lorcé

Parton distributions describe long-distance dynamics in high-energy collisions



$$\sigma_{pp\to H\to\gamma\gamma X}(Q) = \sum_{a,b=g,q,\bar{q}} \int_0^1 d\xi_a \int_0^1 d\xi_b \hat{\sigma}_{ab\to H\to\gamma\gamma}\left(\frac{x_a}{\xi_a}, \frac{x_b}{\xi_b}, \frac{Q}{\mu_R}, \frac{Q}{\mu_F}; \alpha_s(\mu_R)\right)$$
$$\times f_a(\xi_a,\mu_F) f_b(\xi_b,\mu_F) + O\left(\frac{\Lambda_{QCD}^2}{Q^2}\right)$$

$\hat{\sigma}$ is the hard cross section; computed order-by-order in $\alpha_s(\mu_R)$

$f_a(x,\mu_F)$ is the distribution for parton a with momentum fraction x, at scale μ_F

New insights about unpolarized parton distribution functions

Nonperturbative models and lattice QCD

NNLO global analyses of QCD data

CONNECTION?

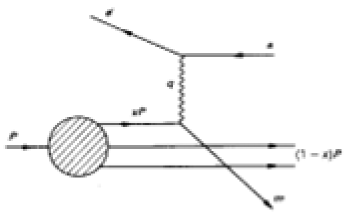
Global QCD analysis

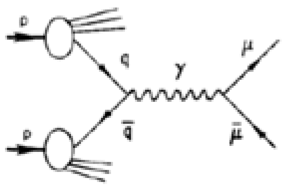
Precision tests at LHC, Jlab, EIC, AMBER, CERN FPF, …

PDFs in nonperturbative QCD

Relevant for processes at $Q^2 \approx 1\ GeV^2$?

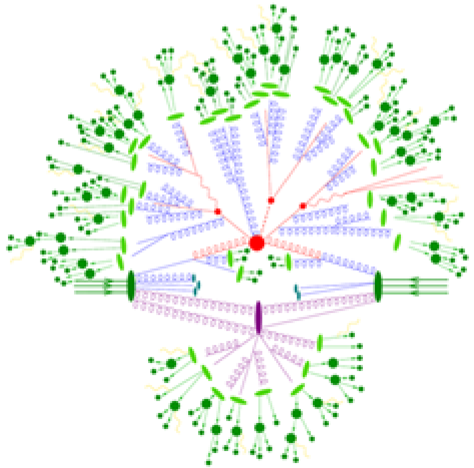
⇨ we can learn about nonperturbative dynamics by comparing predictions to data for the simplest scattering processes (DIS and DY)





Phenomenological PDFs

Determined from processes at $Q^2 \gg 1\ GeV^2$



⇨ pheno PDFs are determined from analyzing many processes with complex scattering dynamics

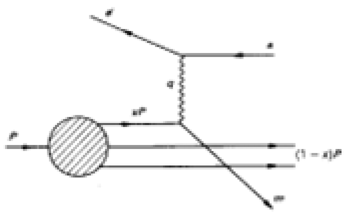
How to relate the x dependence of the perturbative and nonperturbative pictures?

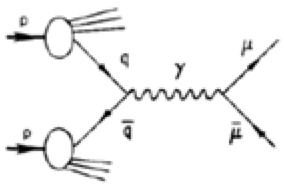
Does the evidence from primordial dynamics survive PQCD radiation?

PDFs in nonperturbative QCD

Relevant for processes
at $Q^2 \approx 1\ GeV^2$?

⇨ we can learn about nonperturbative dynamics by comparing predictions to data for the simplest scattering processes (DIS and DY)



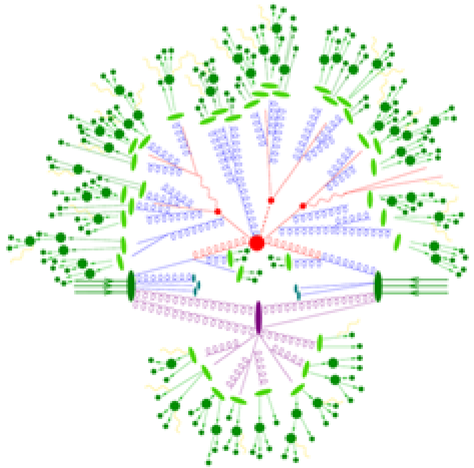


Success requires…

…all four!

Phenomenological PDFs

Determined from processes
at $Q^2 \gg 1\ GeV^2$



⇨ pheno PDFs are determined from analyzing many processes with complex scattering dynamics

How to relate the x dependence of the perturbative and nonperturbative pictures?

Does the evidence from primordial dynamics survive PQCD radiation?

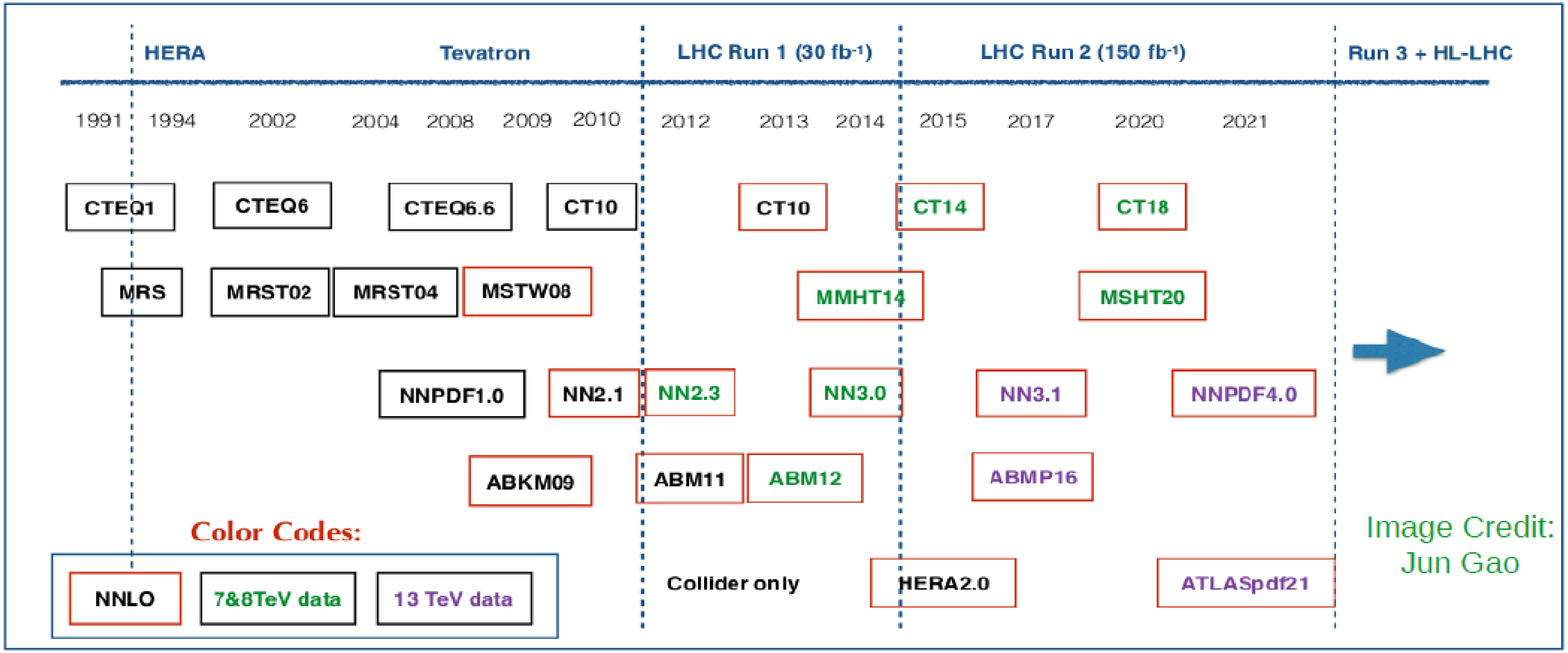
Snowmass'21 whitepaper:
Proton structrure at the precision frontier

S. Amoroso et al., Acta Physica Polonica B 53 (2022) 12, A1

A summary of recent trends in the global analysis of proton PDFs

1. Status of modern NNLO PDFs and their applications
2. Future experiments to constrain PDFs
3. Theory of PDF analysis at N2LO and N3LO
4. New methodological advancements
   - Experimental systematic uncertainties in PDF fits
   - Theoretical uncertainties in PDF fits
   - Machine learning/AI connections
5. Delivery of PDFs; PDF ensemble correlations in critical applications
6. PDFs and QCD coupling strength on the lattice
7. Nuclear, meson, transverse-momentum dependent PDFs
8. Public PDF fitting codes
9. Fast (N)NLO interfaces
10. PDF4LHC21 recommendation and PDF4LHC21 PDFs for the LHC analyses

Phenomenological PDF analyses for a nucleon

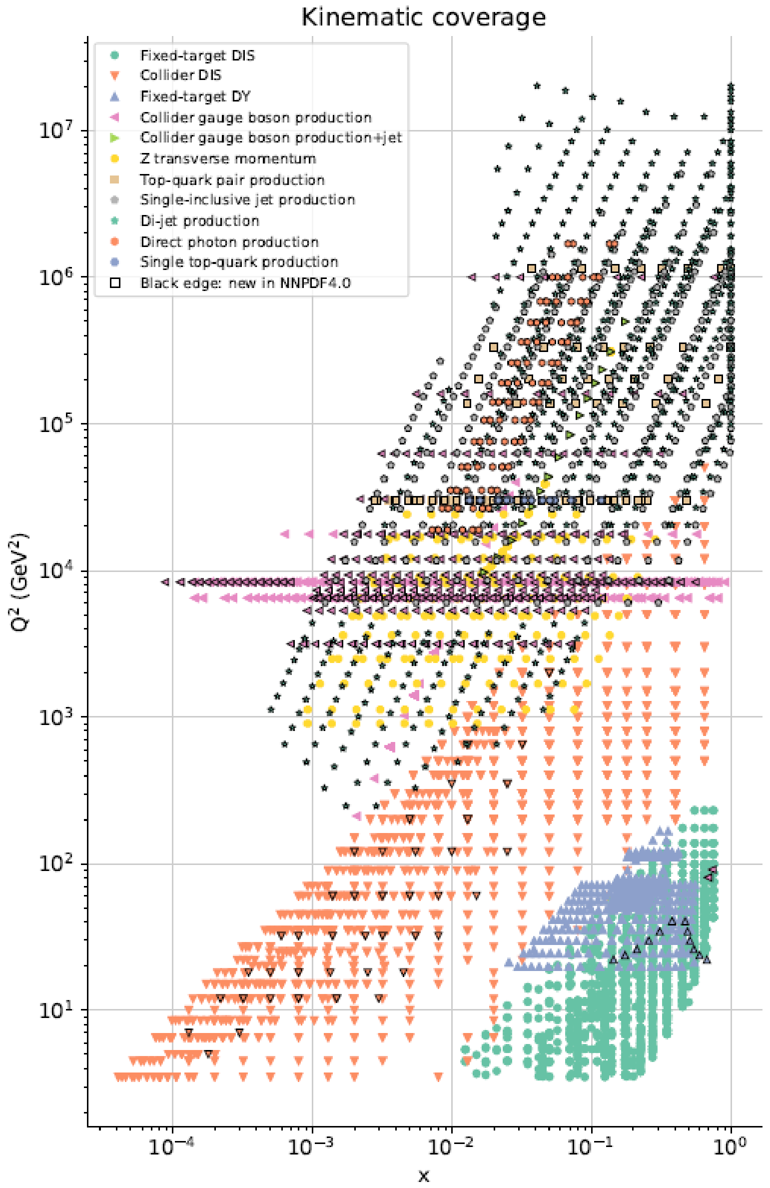


Pursued by several groups – ABM, ATLAS, **CTEQ-TEA (CT)**, CTEQ-Jlab, MSHT, NNPDF, JAM, …

Precision state-of-the art: NNLO QCD + NLO EW; partial N3LO results (NNPDF and MSHT groups)

Data from fixed-target experiments and colliders (HERA, Tevatron, LHC, …)

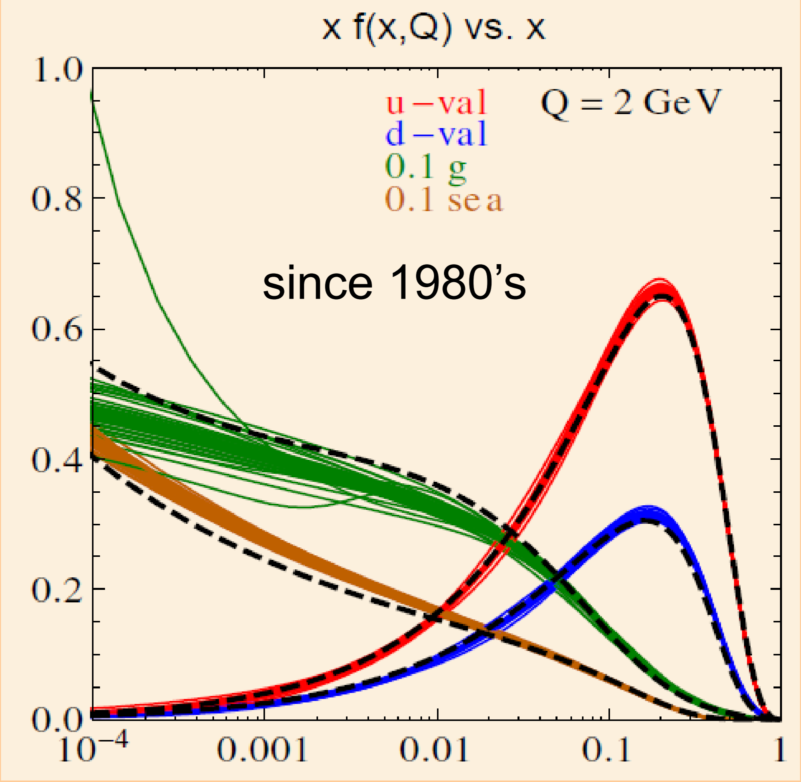
Data in the NNPDF4.0 fit



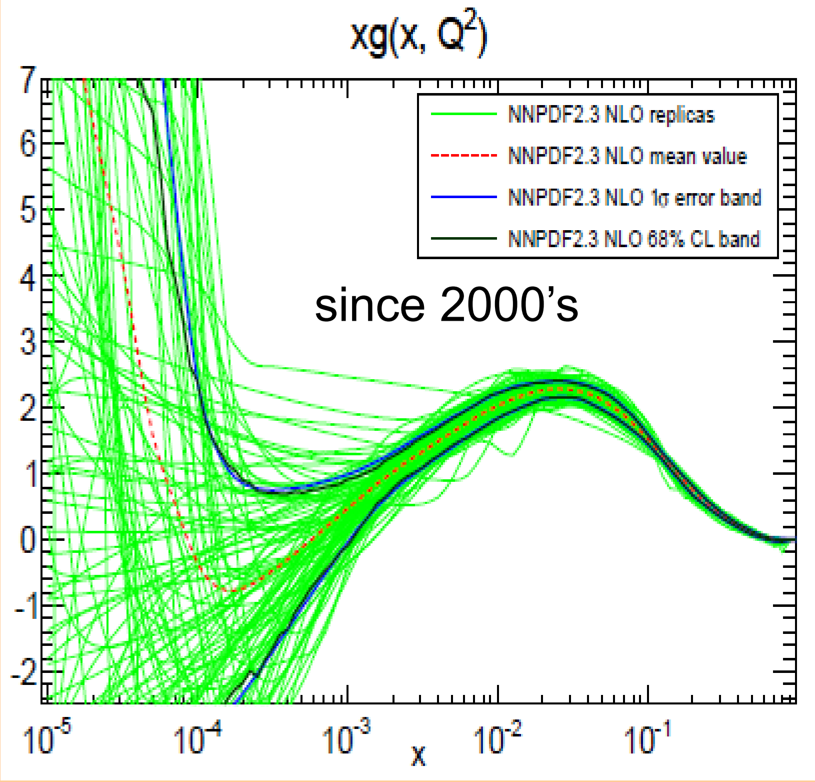
Similarly constraining data sets in the CT18 NNLO and MSHT'20 NNLO fits

Two types of modern error PDFs

Analytic parametrizations +
Hessian PDF eigenvector sets
(ABM, CTEQ, HERA, MSHT,…)



Neural network parameterizations
+ Monte Carlo PDF replicas
(NNPDF)



a textbook application of ML in particle physics

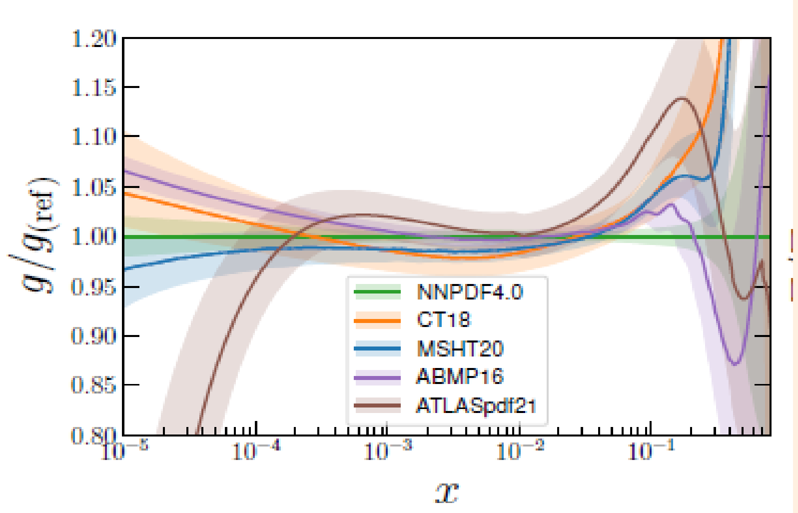
Two powerful, complementary representations for estimating PDF uncertainties
Hessian PDFs can be converted into MC ones, and vice versa.

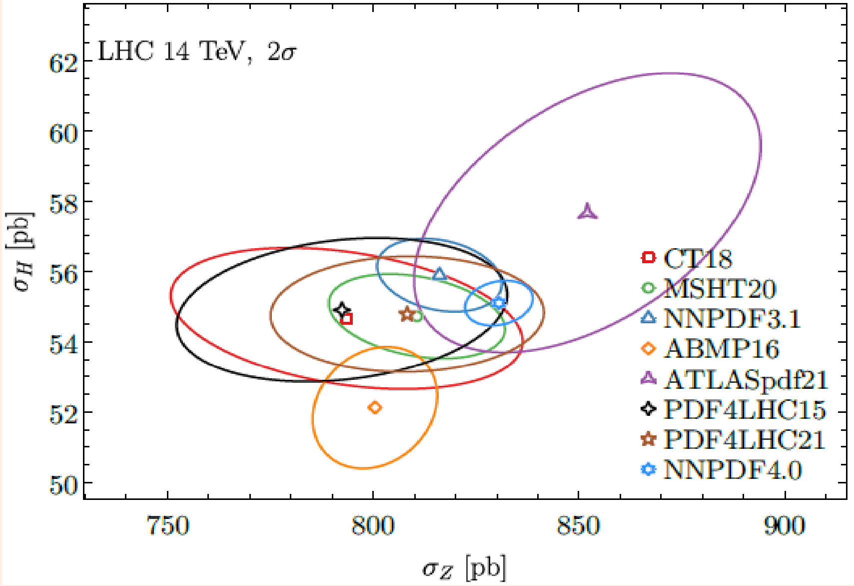
Progress in PDF analysis

Snowmass 2021 whitepaper: Proton structure at the precision frontier

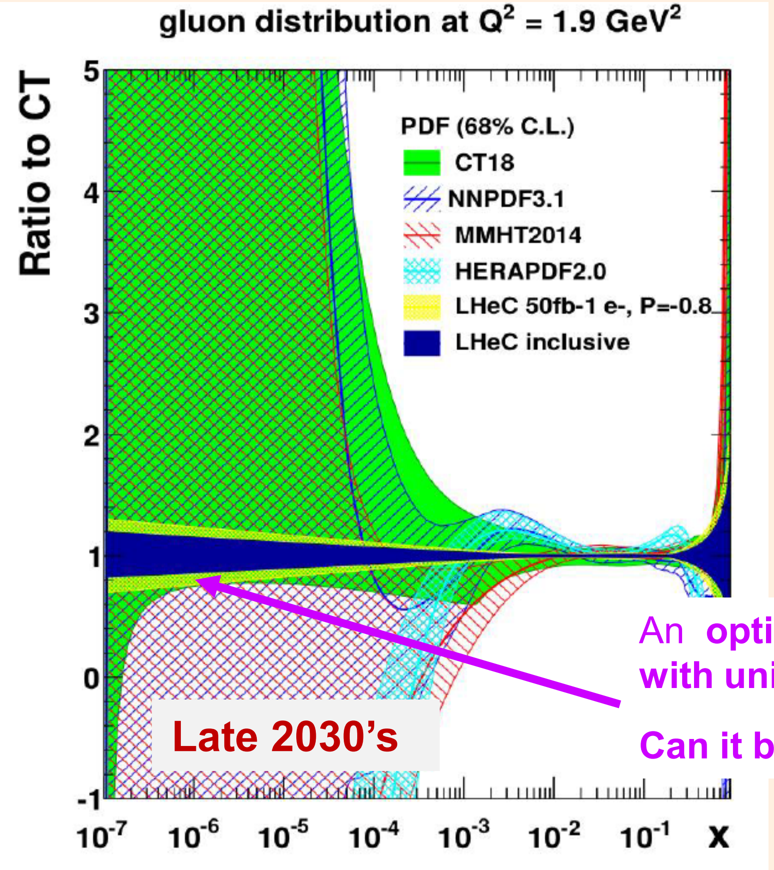
S. Amoroso et al., Acta Physica Polonica B 53 (2022) 12, A1

The current status **2022**





...and future prospects



An **optimistic** uncertainty **with unit tolerance**

Can it be reached?

Some ongoing developments

my biased selection

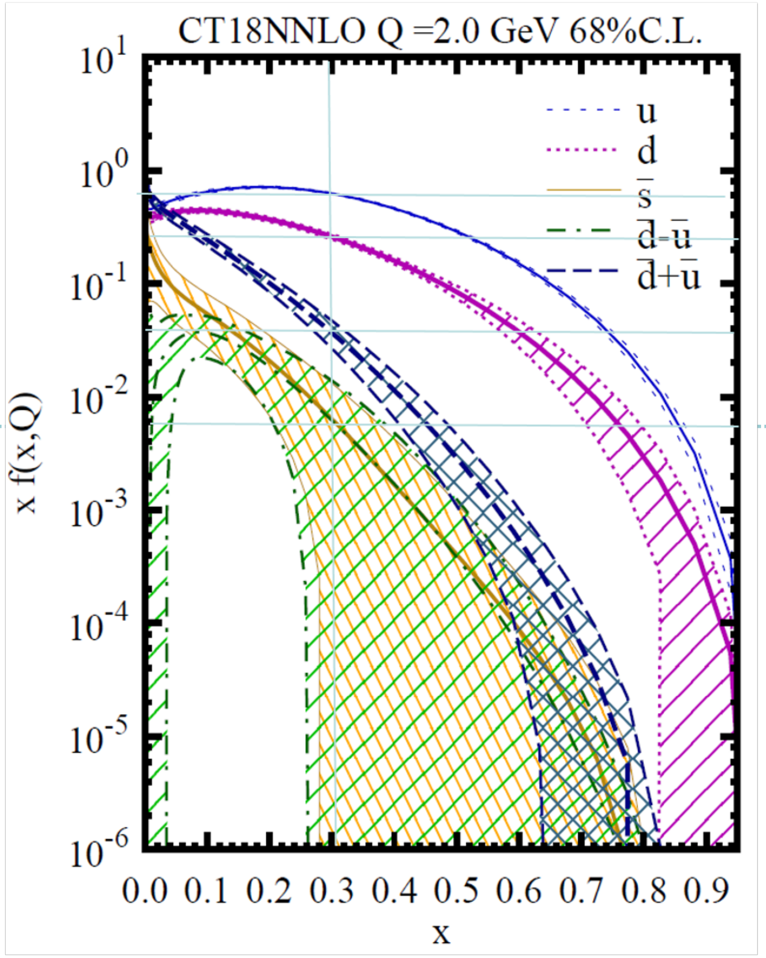
1. How do $\bar{u}$ and $\bar{d}$ PDFs behave at $x > 0.3$?

2. Are PDFs for the 2nd and 3rd quark generation $(s, c, \ldots)$ charge-symmetric?

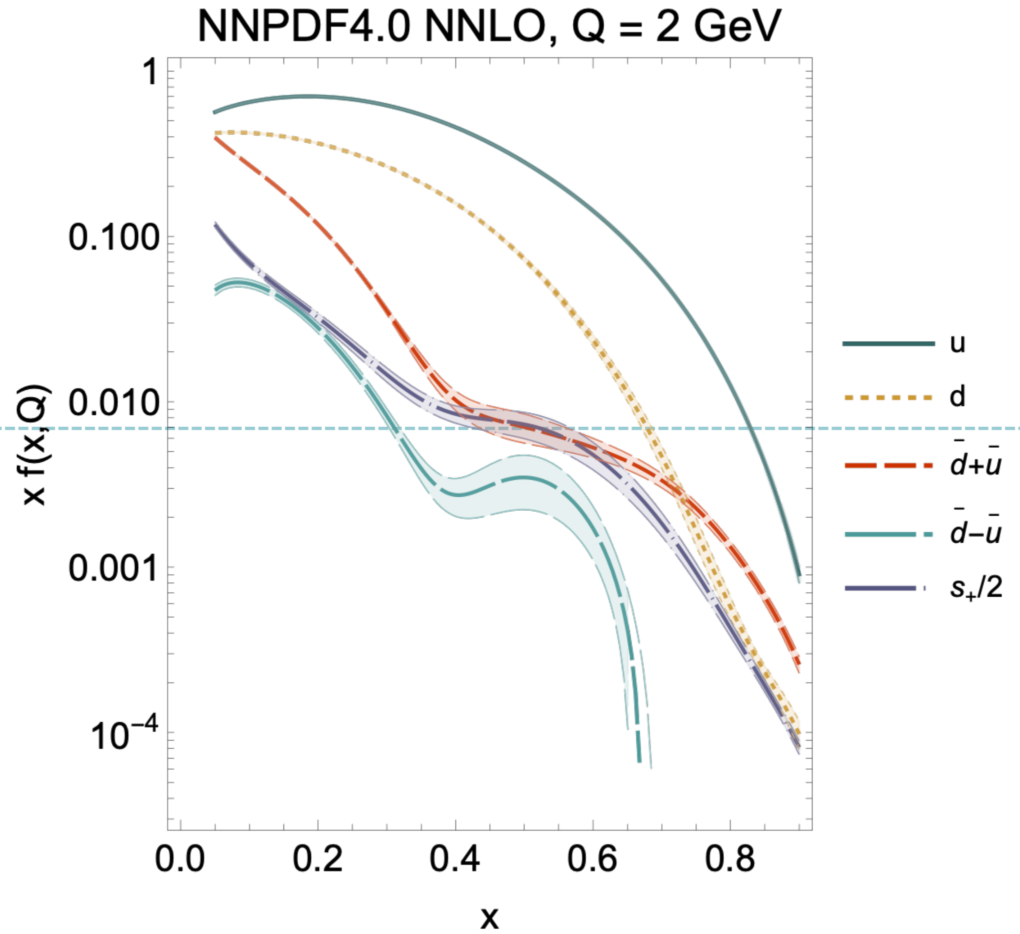
3. How large is fitted charm in the proton?

See also talks by Z. Meziani, M. Diehl

$\bar{u}, \bar{d}, \bar{s}$ PDFs are still poorly known at $x > 0.3$



Low-sea scenario with smooth light sea quarks.



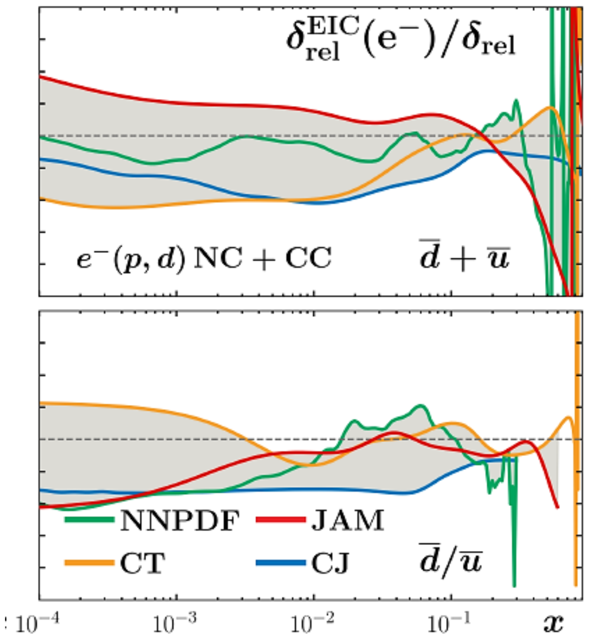
High-sea scenario with non-smooth light-sea quarks, with sea PDFs that can be larger than valence PDFs at large x.

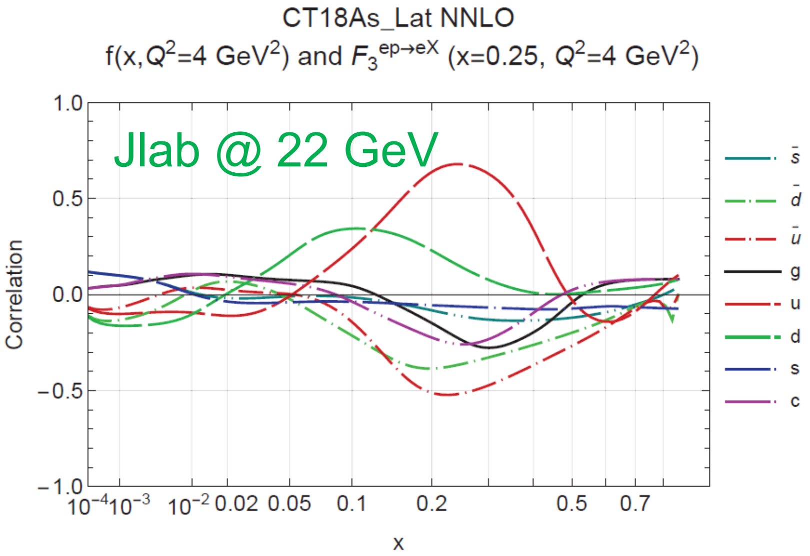
Both scenarios are compatible with current data. Uncertainties and even signs of PDFs vary among the groups (see, e.g., arXiv:2203.05506, 2205.10444)

Sensitivity of experiments to $\bar{u}$ and $\bar{d}$ at $x > 0.1$

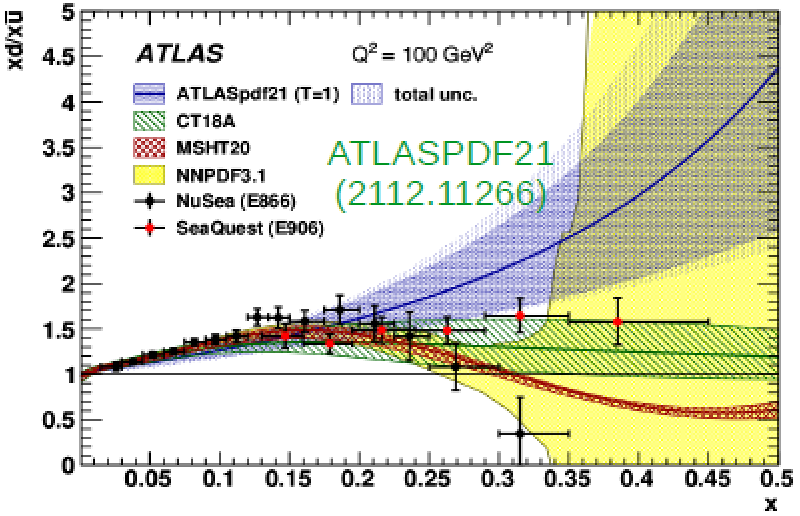
Inclusive DIS: sensitivity is limited as a result of dominance of u and d at $x \to 1$ (including in PVDIS at the EIC and SOLID). Sensitivity can be augmented by SIDIS data.

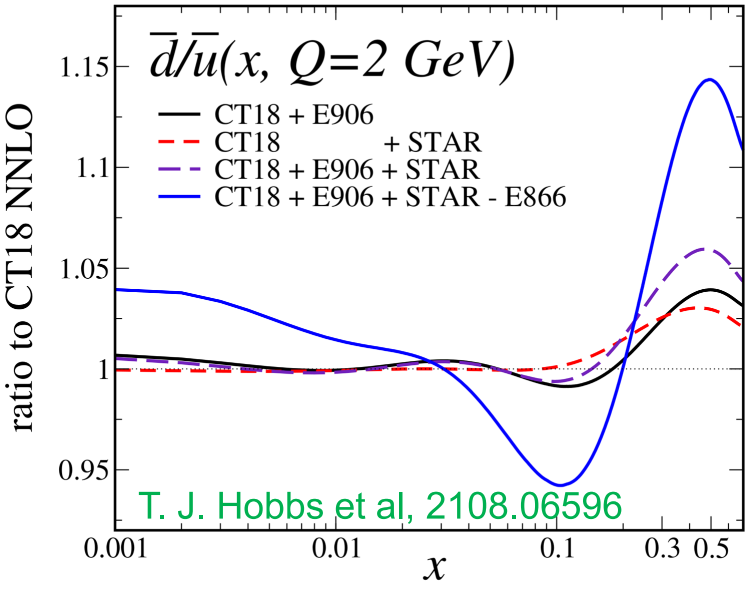
Projections for the EIC (2103.05419) and Jlab @ 22 GeV (2306.09360)



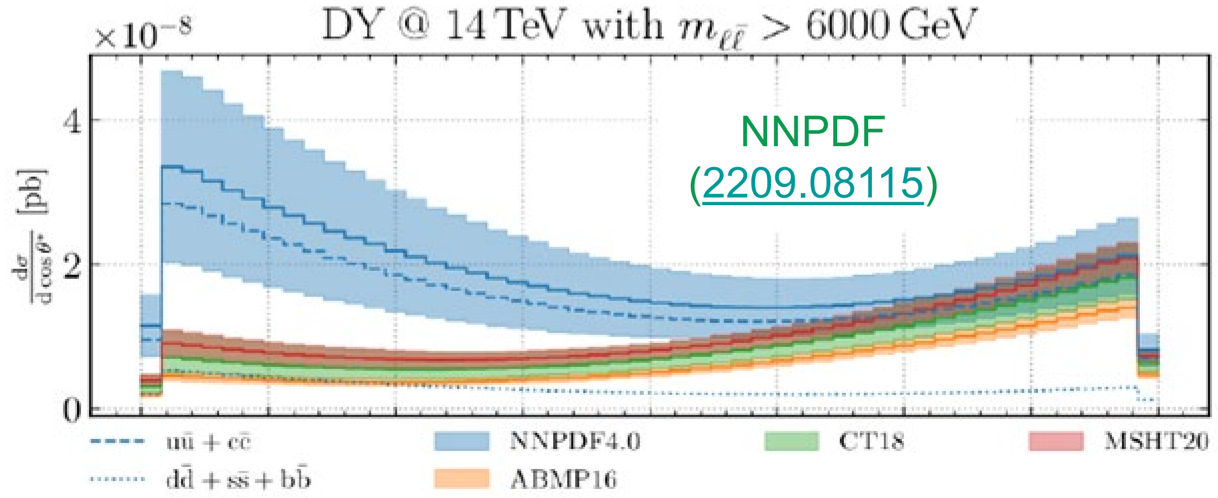


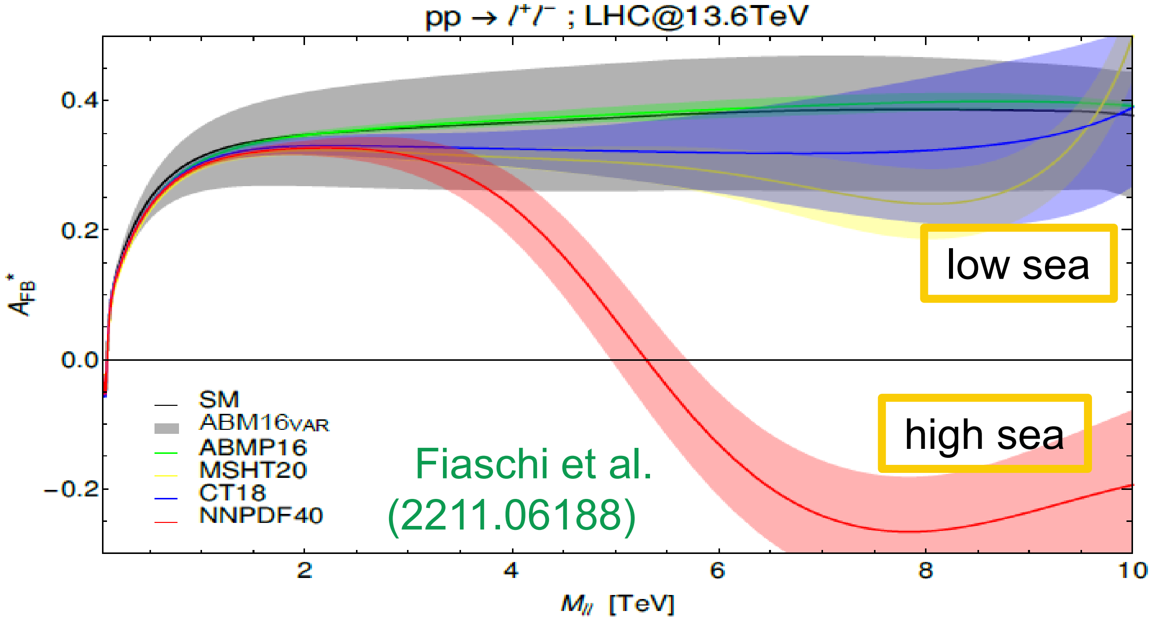
Drell-Yan process: good sensitivity. The σ_{pd}/σ_{pp} ratio by SeaQuest (E906) prefers $\bar{d} > \bar{u}$, is above the E866 ratio at $x > 0.25$. The net impact on PDFs is weak as a result of discrepancy. Uncertainties are still large



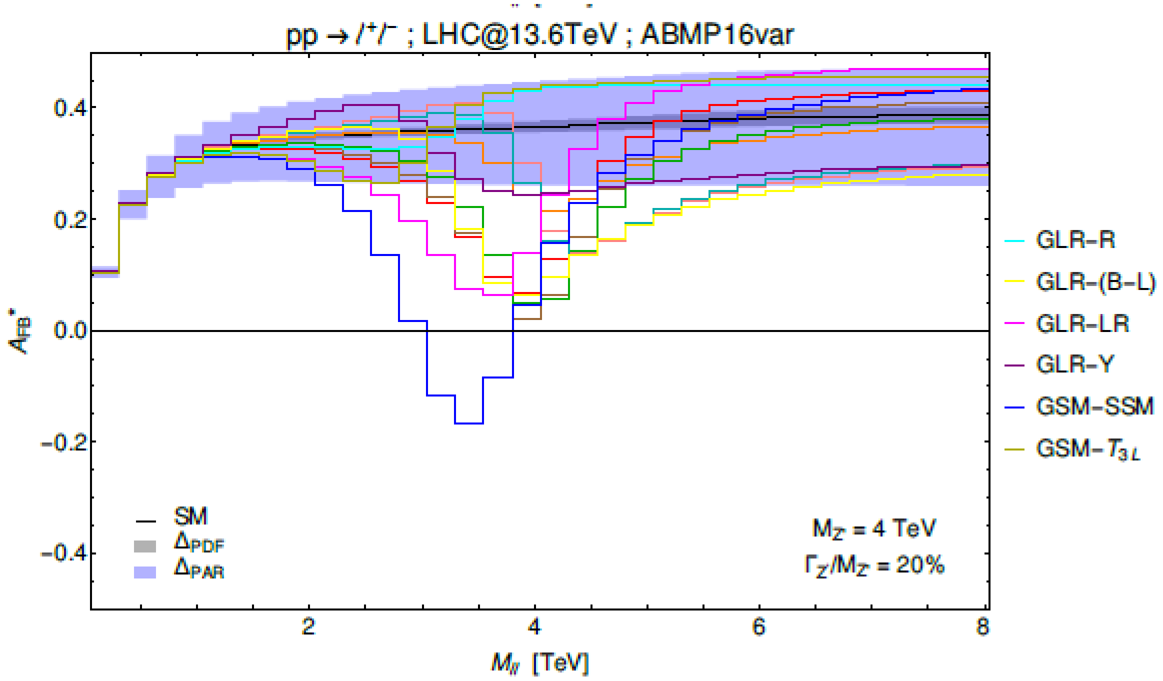


LHC high-mass Drell-Yan process probes $\bar{u}$ and $\bar{d}$





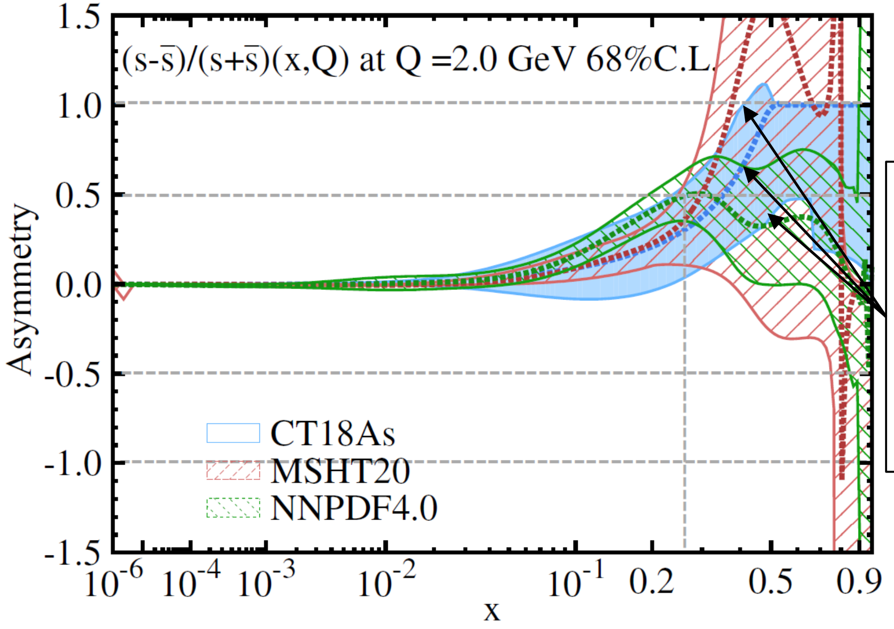
The forward-backward asymmetry for $M_{\ell\bar{\ell}} > 3$ TeV probes the $\bar{u}/u$ and $\bar{d}/d$ combinations at $x > 0.1$



Relevant in searches for BSM resonances

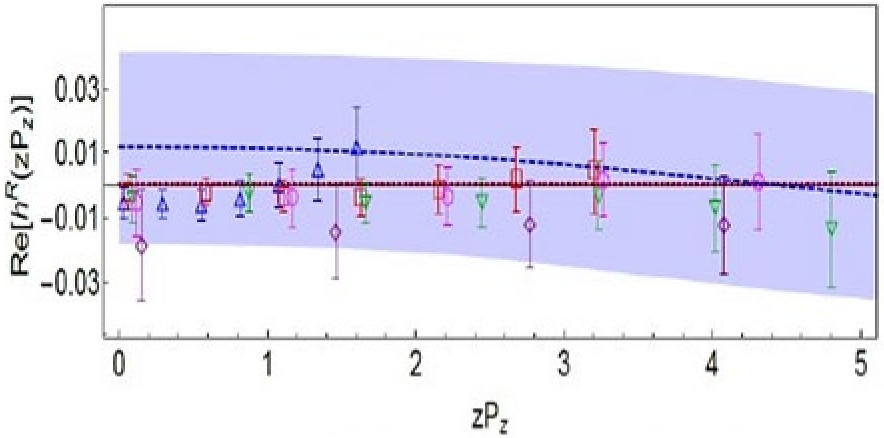
CT18As NNLO: Strangeness asymmetry with a lattice QCD constraint

T.-J. Hou et al., arXiv: 2204.07944



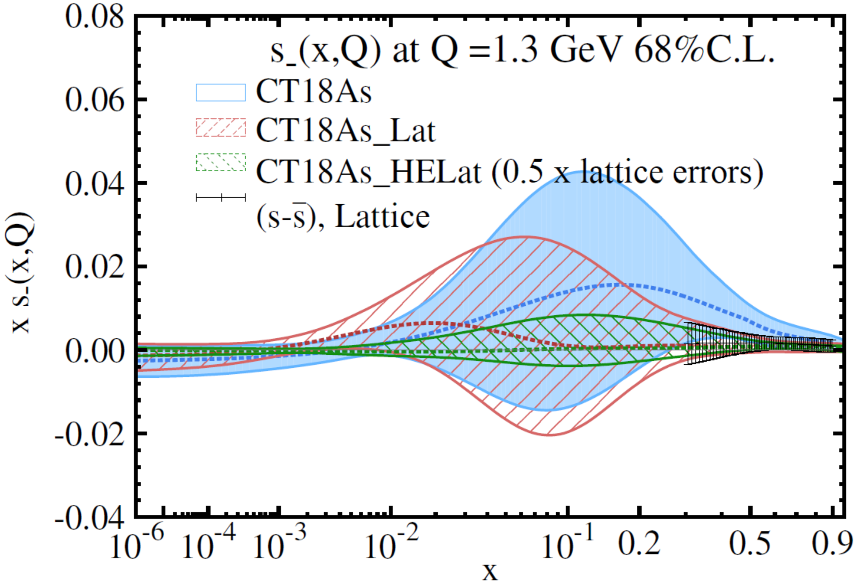
differences reflect the pulls of LHC and other experiments

Asymmetry nominally reaches $\approx 50\%$ at $x \approx 0.25$ in three global fits. **Is there a dynamical mechanism to produce it at such x?**

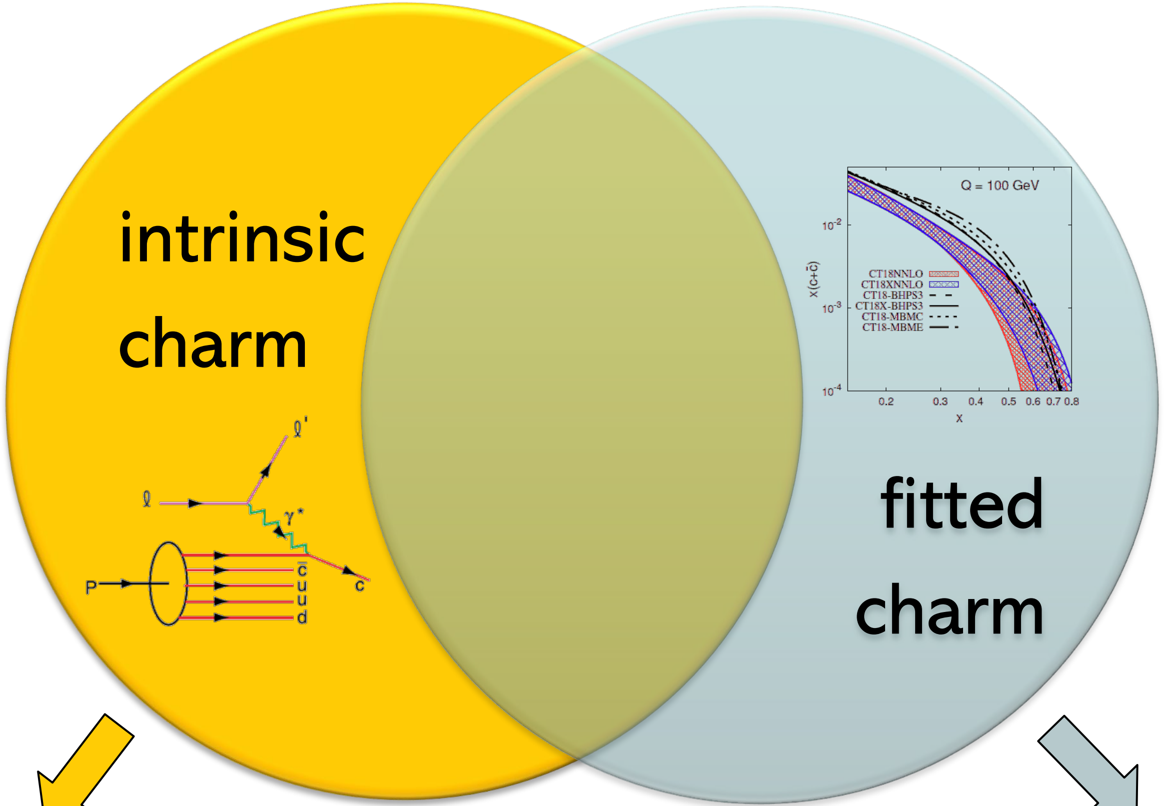


Include lattice data on s_- obtained by the MSULat/quasi-PDF method (2005.12015, Zhang, Lin, Yoon)

The lattice QCD prediction disfavors a large $s_-(x,Q)$ at $x > 0.3$ $\Rightarrow$ reduction in $s_-(x,Q)/s_+(x,Q)$ in CT18As_Lat fit



See also G. Magni



- The concept of nonperturbative methods
- Can refer to a component of the hadronic Fock state or the type of the hard process
- Predicts a typical enhancement of the charm PDF at $x \gtrsim 0.2$

Connection?

- A charm PDF parametrization at scale $Q_0 \approx 1$ GeV found by global fits [CT, NNPDF, …]
- Arises in perturbative QCD expansions over α_s and operator products
- May absorb process-dependent or unrelated radiative contributions

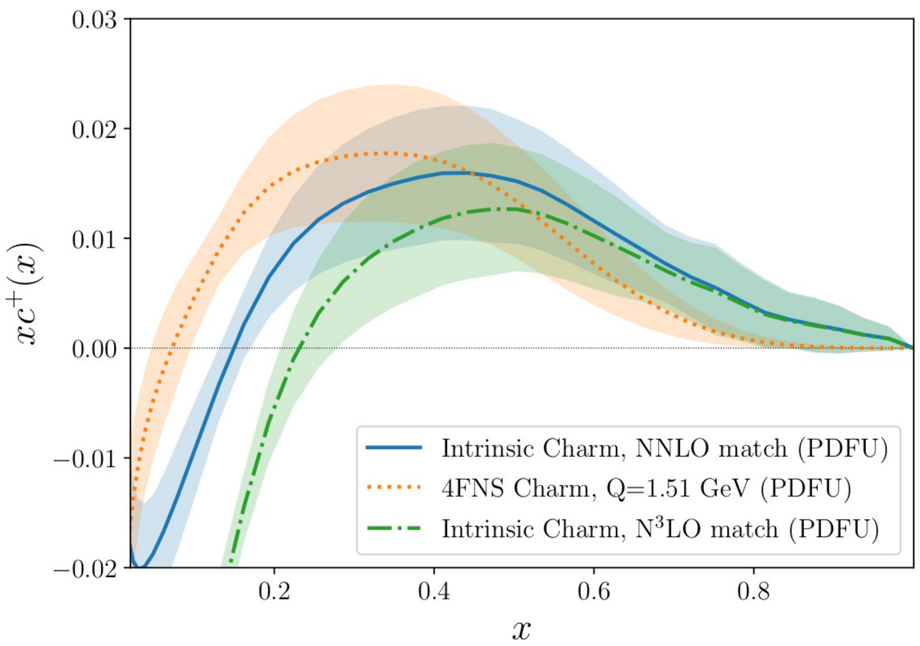
NNPDF IC PDFs and moments

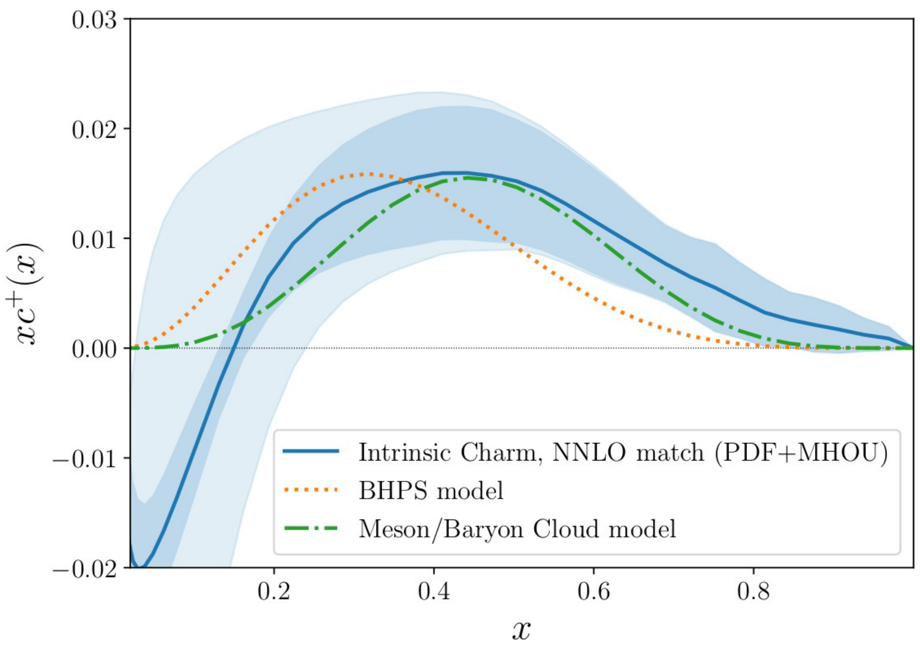
NNPDF, Nature 608 (2022) 7923, 483.

Preference for non-zero NNPDF IC PDF at $0.2 \lesssim x \lesssim 0.7$

Talk by G. Magni

Large perturbative instability from MHOU in DGLAP affects $x < 0.2$





→ MHOU excluded to obtain a nominal charm fraction, $\langle x \rangle_{\mathrm{FC}} = 0.62 \pm 0.28\%$

→ if MHOU is included, consistency with zero: $\langle x \rangle_{\mathrm{FC}} = 0.62 \pm 0.61\%$

CT18 FC charm PDFs *M. Guzzi, T. J. Hobbs, K. Xie, et al., PLB 843 (2023) 137975; update on T.-J. Hou et al. JHEP* 02 (2018) 059

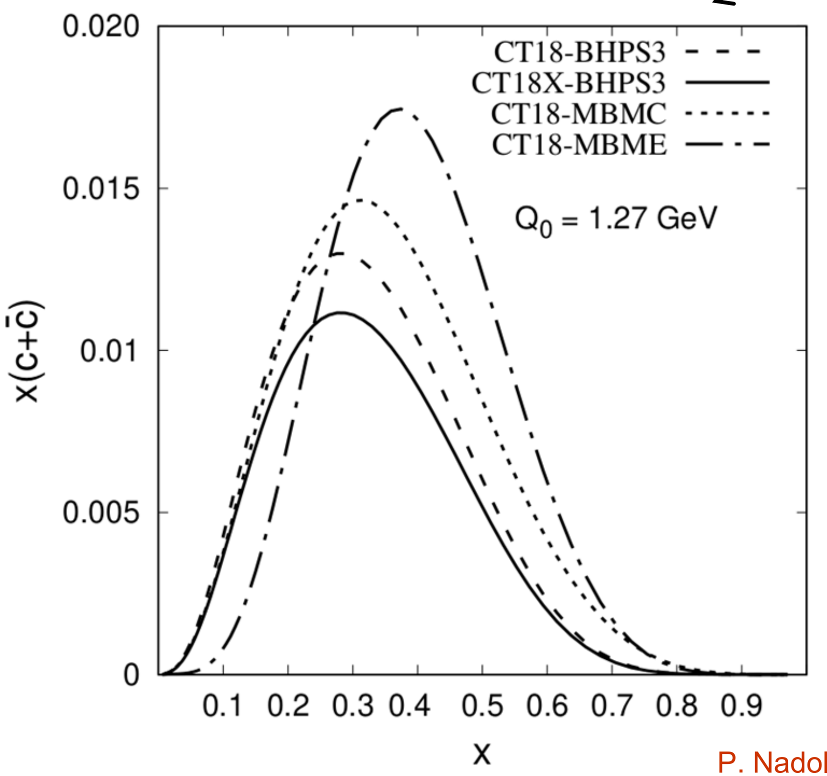
FC scenarios traverse range of high-*x* behaviors from IC models

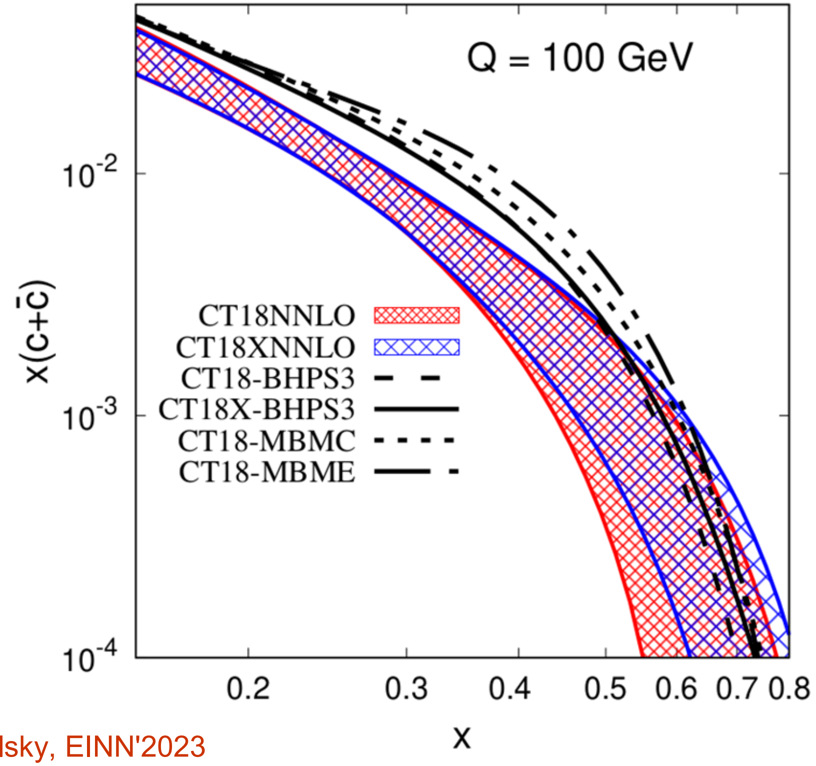
→ fit implementation of BHPS from CT14IC (BHPS3) on CT18 or CT18X (NNLO)

→ fit two MBMs: MBMC (confining), MBME (effective mass) on CT18

investigate constraints from newer LHC data in CT18

central fits

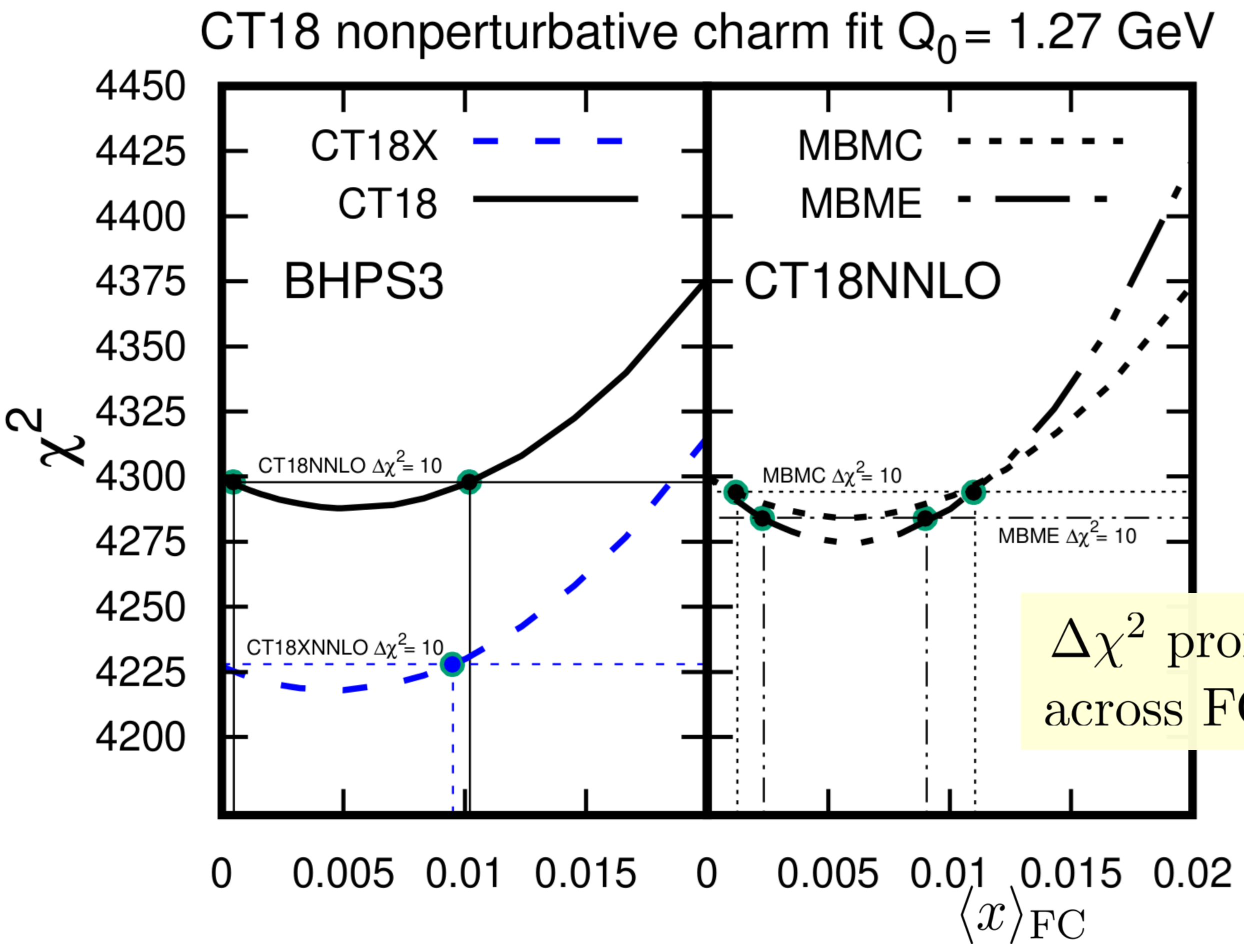




signal for FC in CT18 study, but with shallower $\Delta\chi^2$ than CT14 IC

CT14 IC

CT18 FC NNLO is compatible with zero



$\Delta\chi^2$ profiles similar across FC scenarios

$\langle x \rangle_{\rm FC}$

FC uncertainty quantified by normalization via $\langle x_{FC} \rangle$ for each input IC model

- $\langle x \rangle_{FC} \approx 0.5\%$ $(\Delta\chi^2 \gtrsim -25)$ in CT18 FC

  vs.

  $\langle x \rangle_{FC} \approx 0.8 - 1\%$ $(\Delta\chi^2 \gtrsim -40)$ in CT14 IC

FC PDF moments

even restrictive uncertainties give moments consistent with zero

→ broaden further for default CT tol.

→ lattice may give $\langle x \rangle_{c^+}$, $\langle x^2 \rangle_{c^-}$

$$\langle x \rangle_{\mathrm{FC}} \equiv \langle x \rangle_{\mathrm{c+}}[Q_0 = 1.27\,\mathrm{GeV}]$$

$$= 0.0048\,^{+0.0063}_{-0.0043}\,\left(^{+0.0090}_{-0.0048}\right),\ \mathrm{CT18\ (BHPS3)}$$

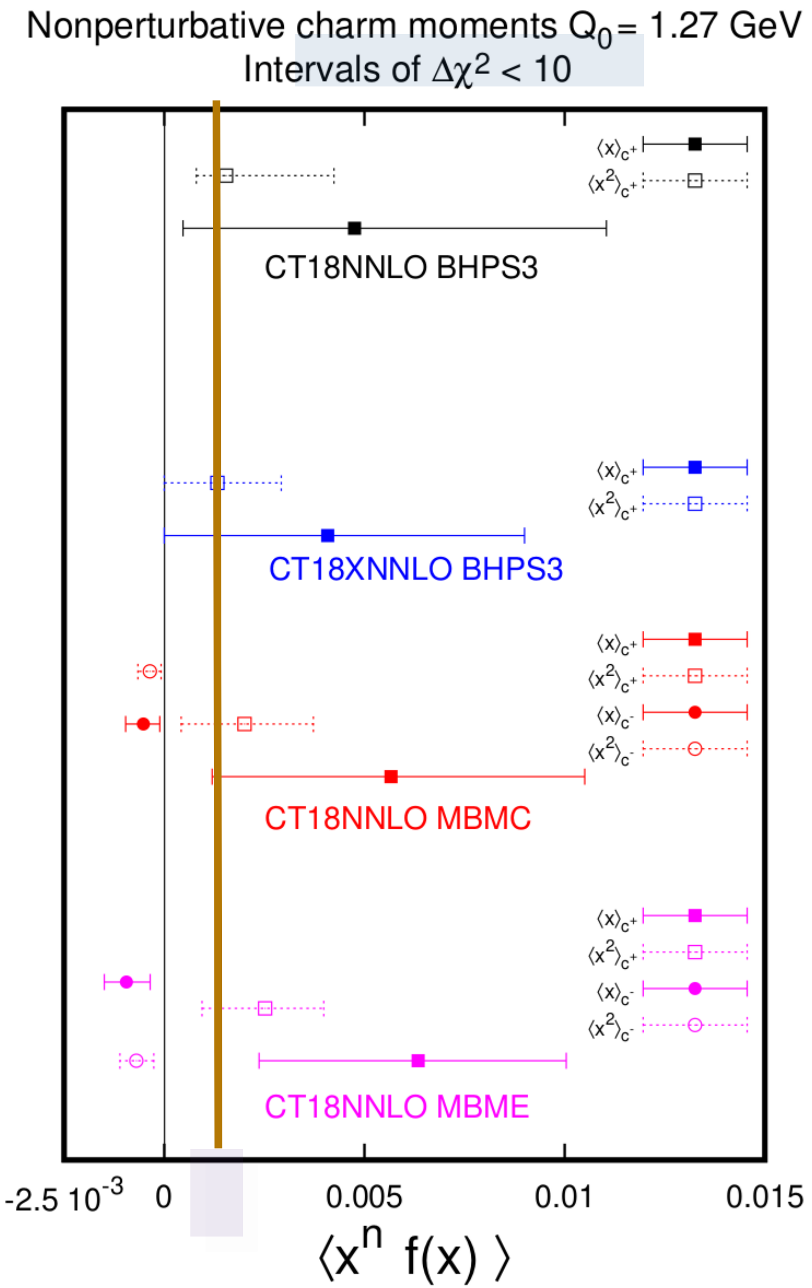
$$= 0.0041\,^{+0.0049}_{-0.0041}\,\left(^{+0.0091}_{-0.0041}\right),\ \mathrm{CT18X\ (BHPS3)}$$

$$= 0.0057\,^{+0.0048}_{-0.0045}\,\left(^{+0.0084}_{-0.0057}\right),\ \mathrm{CT18\ (MBMC)}$$

$$= 0.0061\,^{+0.0030}_{-0.0038}\,\left(^{+0.0064}_{-0.0061}\right),\ \mathrm{CT18\ (MBME)}$$

$\Delta\chi^2 \leq 10$ (restrictive tolerance)

$\Delta\chi^2 \leq 30$ (~CT standard tolerance)

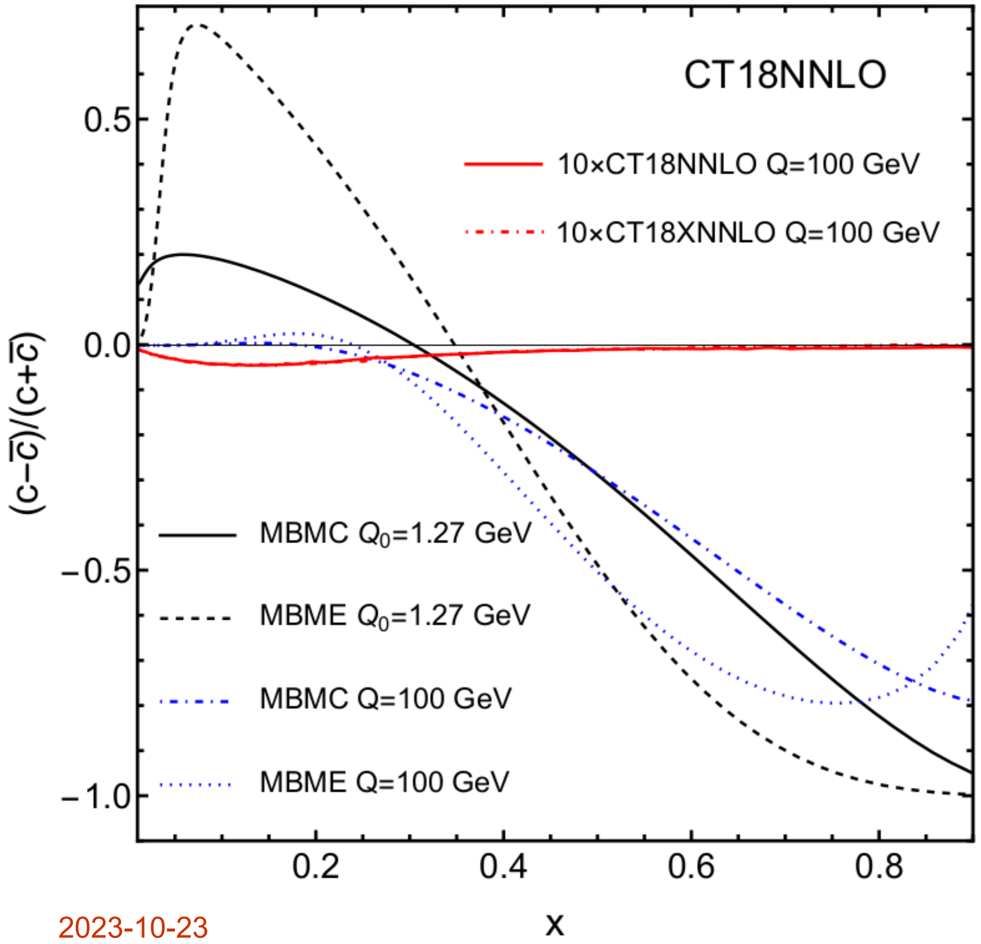


Charm-anticharm asymmetries can be a robust signal of nonperturbative charm

pQCD only very weakly breaks $c = \bar{c}$ through HO corrections

→ large(r) charm asymmetry would signal nonpert dynamics, IC

→ MBM breaks $c = \bar{c}$ through hadronic interactions

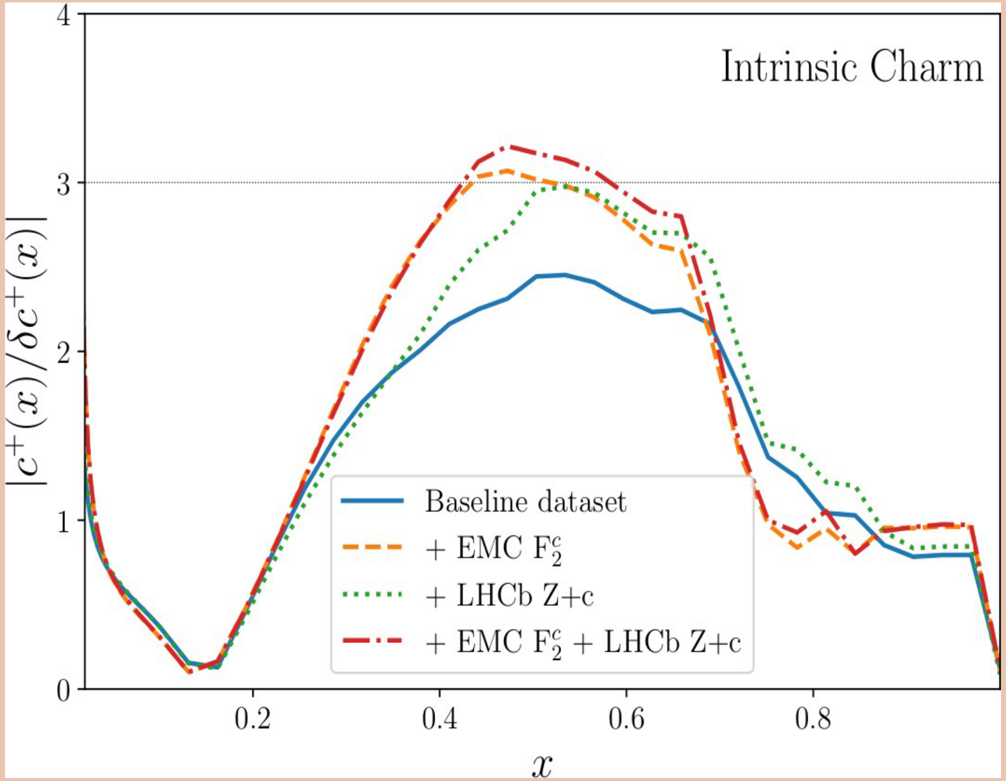


consider two MBM models as ***examples*** (<u>not predictions</u>)

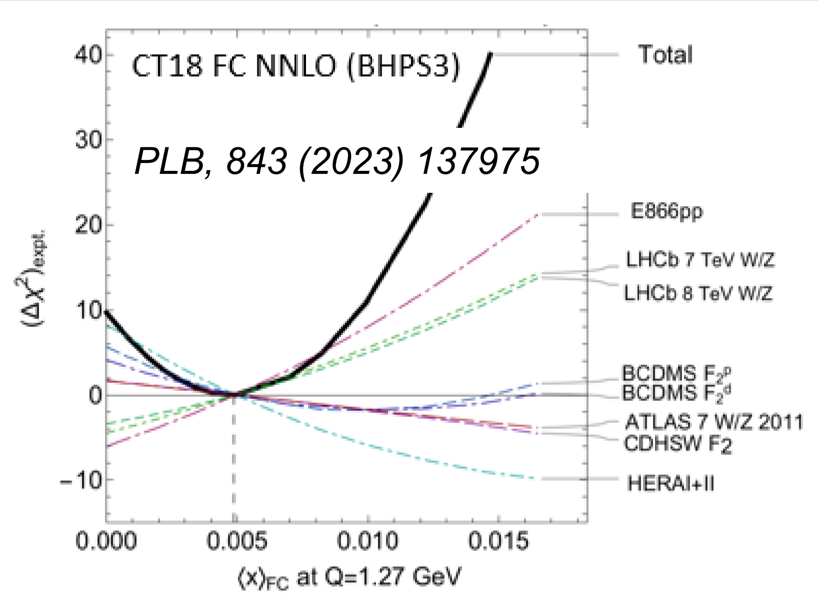
→ asymptotically small, but ratio can be bigger; will be hard to extract from data

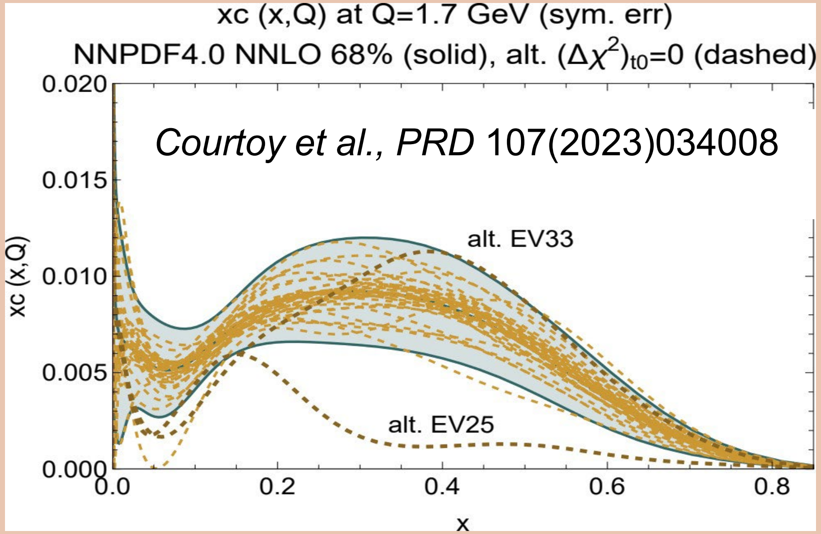
T. J. Hobbs, J. T. Londergan, W. Melnitchouk, *Phys. Rev. D* 89 (2014) 074008

How significant is the non-zero FC?



NNPDF states a 3σ evidence for $f_{IC}(x, Q_0) \neq 0$ based on the combined constraints from the baseline fit, LHCb $Z + c$ analysis, and EMC F_2^c data





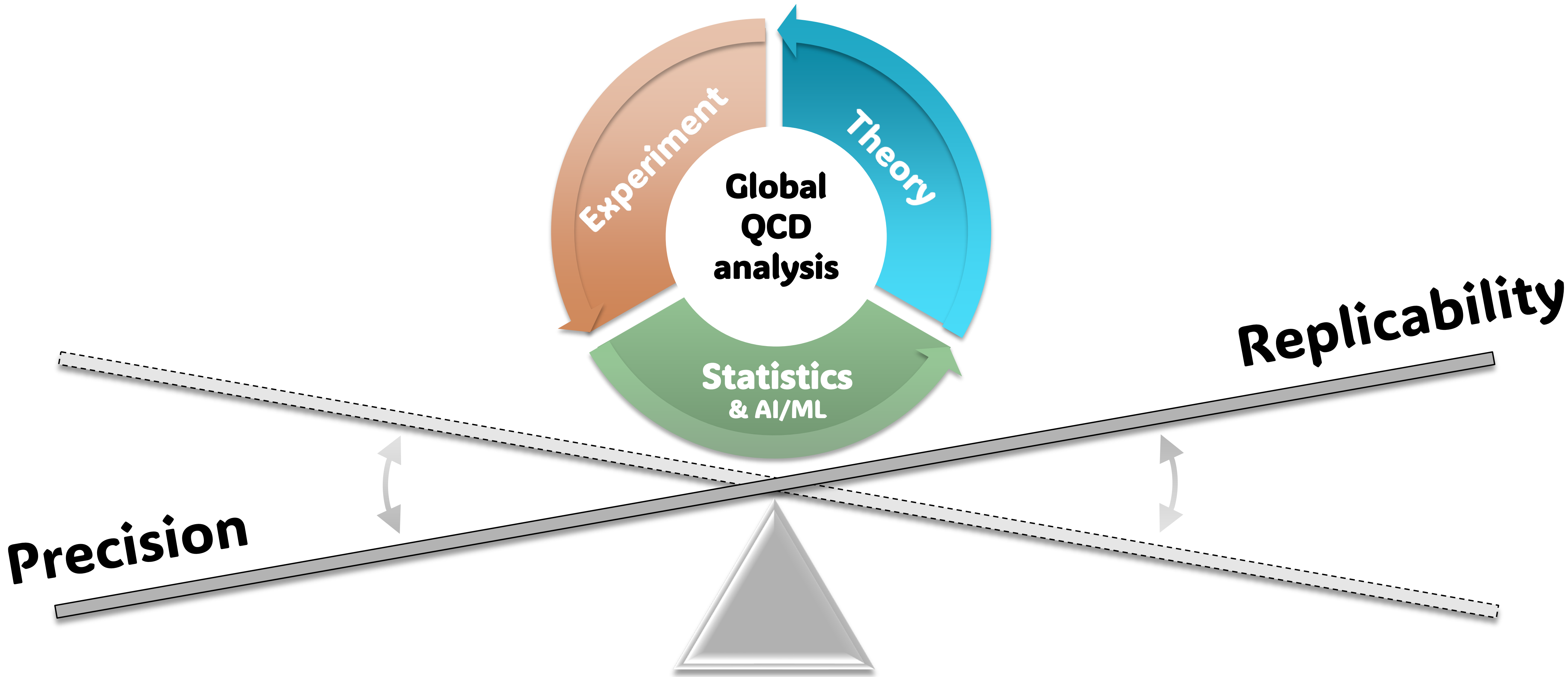
CTEQ-TEA authors find larger uncertainties in each of these sources.

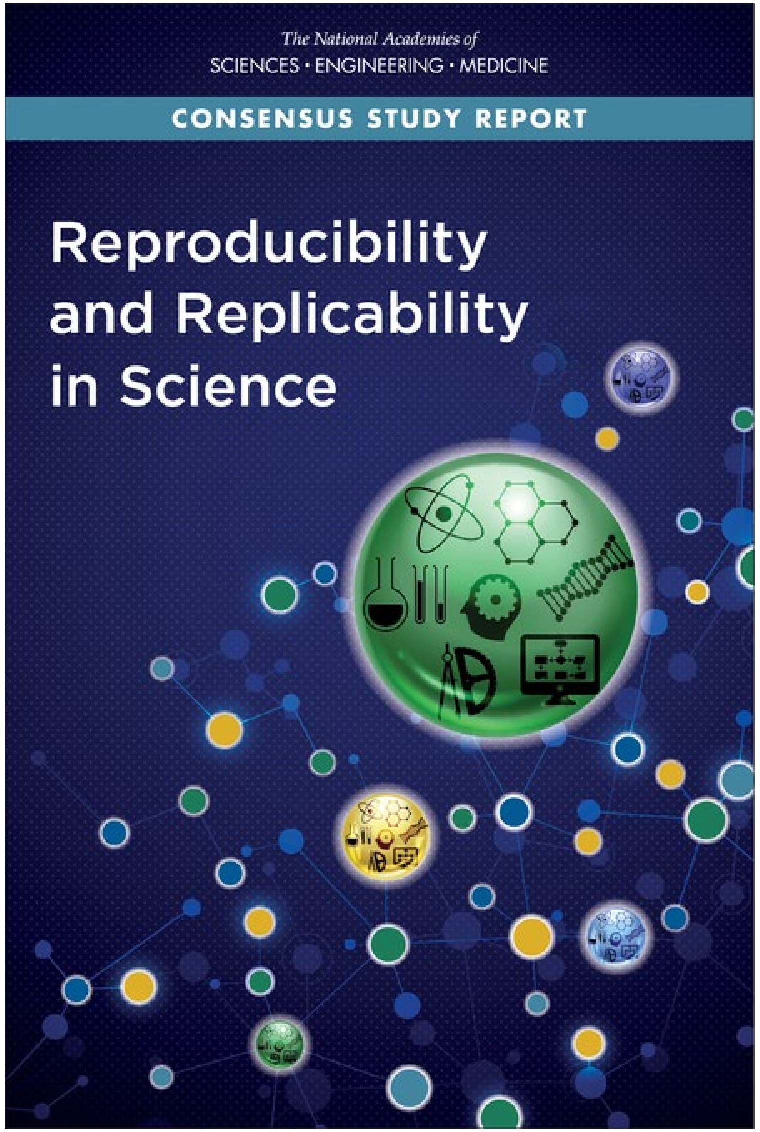
This conclusion is also supported by

- **Lagrange multiplier scans** in the CT18 FC fit [upper figure]
- **hopscotch sampling** of MC replicas in the NNPDF4.0 fitting code [lower figure].

Consequently, $f_{FC}(x, Q_0) \approx 0$ is allowed with high confidence.

Toward replicable PDF uncertainties





US National Academy of Sciences, Engineering, and Medicine, 2019, https://doi.org/10.17226/25303

Replicability risks for precision QCD

Replicability is a requirement of obtaining consistent results across studies aimed at answering the same scientific question, each of which has its own analysis strategy or data.

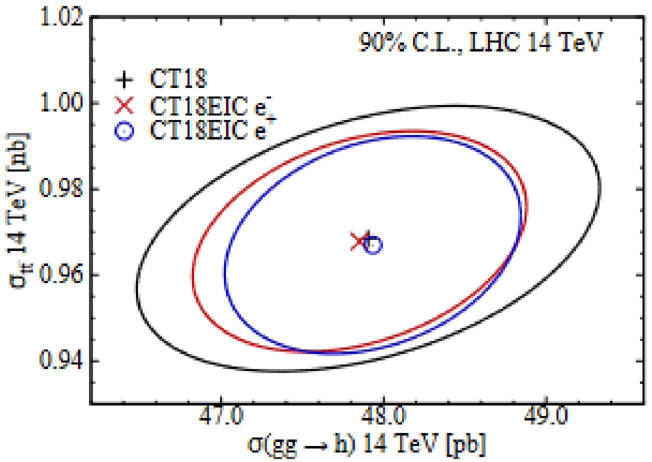
Nearly all complex STEM fields encounter replicability challenges.

Modern particle physics is not an exception.

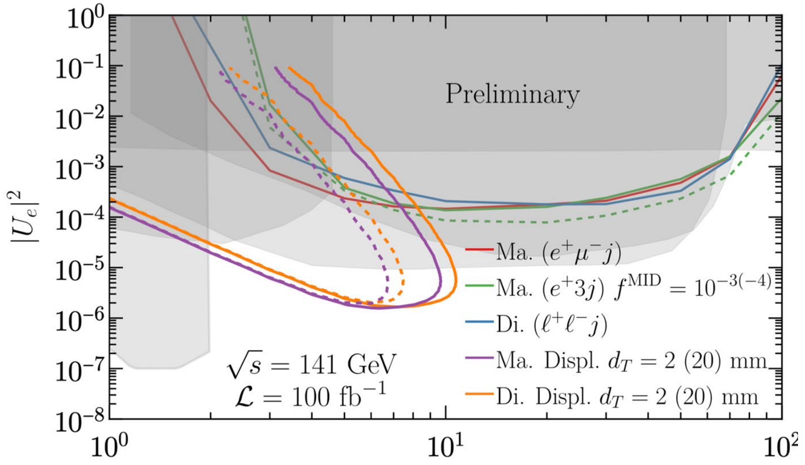
1. It is complex! Is it rigorous enough?
   - Many approaches, especially AI-based ones, increase complexity and are not rigorously understood
2. It often uses wrong prescriptions for estimating epistemic uncertainties
   - Tens to hundreds of systematic uncertainties affect measurements, phenomenology, and lattice QCD

Electron-Ion Collider: potentially a wealth of complex studies

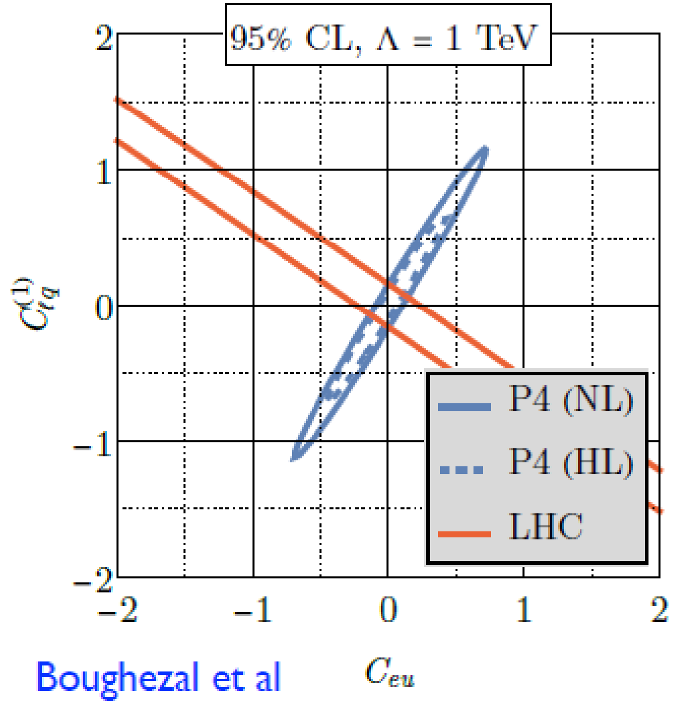
PDFs: arXiv:2103.05419



heavy neutral lepton searches arXiv: 2203.06705

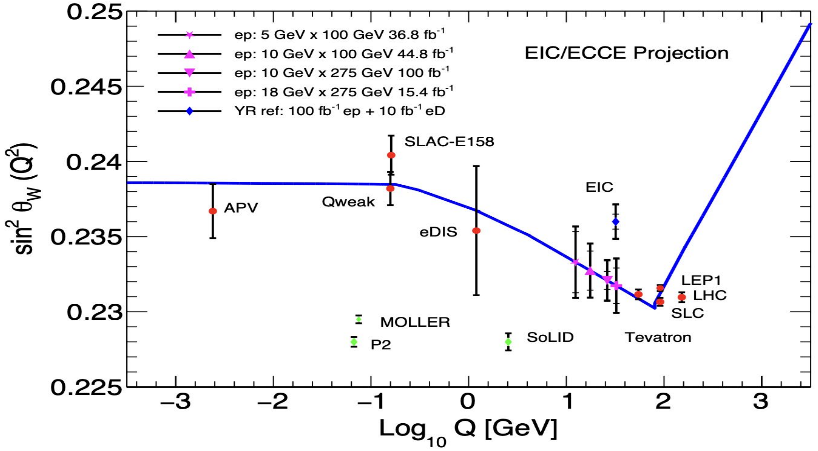


SMEFT Wilson coefficients



Boughezal et al
2004.00748, .2204.07557

weak mixing angle arXiv: 2203.13199



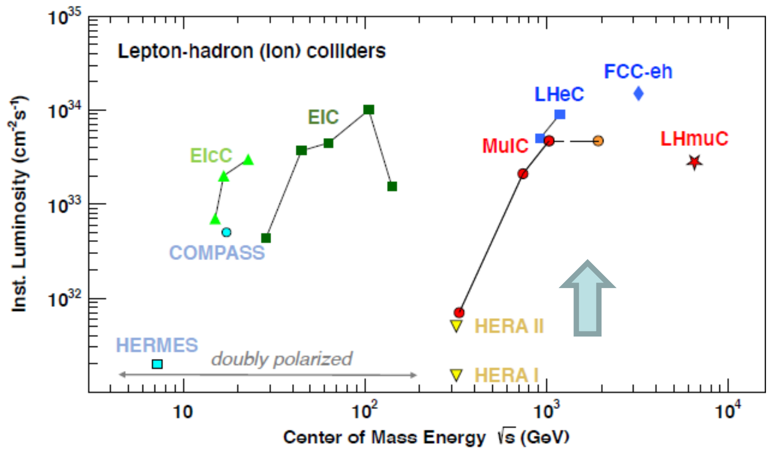
	EIC	LHC
$\lvert c_u^{XX} - c_u^{YY}\rvert$	0.37	15
$\lvert c_u^{XY}\rvert$	0.13	2.7
$\lvert c_u^{XZ}\rvert$	0.11	7.3
$\lvert c_u^{YZ}\rvert$	0.12	7.1
$\lvert a_{\mathrm{S}u}^{(5)TXX} - a_{\mathrm{S}u}^{(5)TYY}\rvert$	2.3	0.015
$\lvert a_{\mathrm{S}u}^{(5)TXY}\rvert$	0.34	0.0027
$\lvert a_{\mathrm{S}u}^{(5)TXZ}\rvert$	0.13	0.0072
$\lvert a_{\mathrm{S}u}^{(5)TYZ}\rvert$	0.12	0.0070

Lorentz/CPT violations
A. R. Vieira et al., 1911.04002

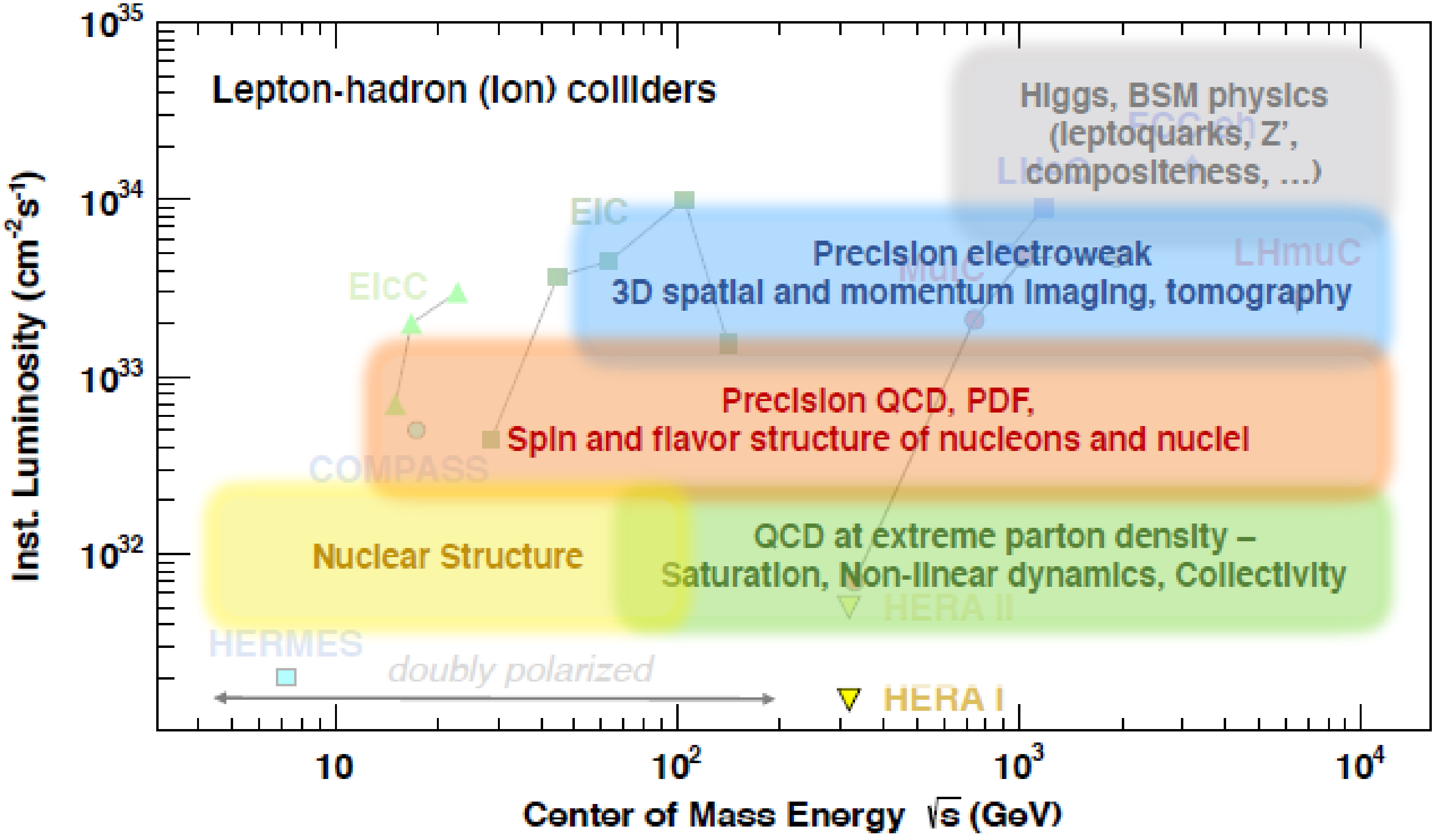
Abdul-Khalek et al., Snowmass 2021 whitepaper "EIC for HEP", 2203.13199

The Muon-Ion Collider, Large Hadron Electron Collider, FCC-eh

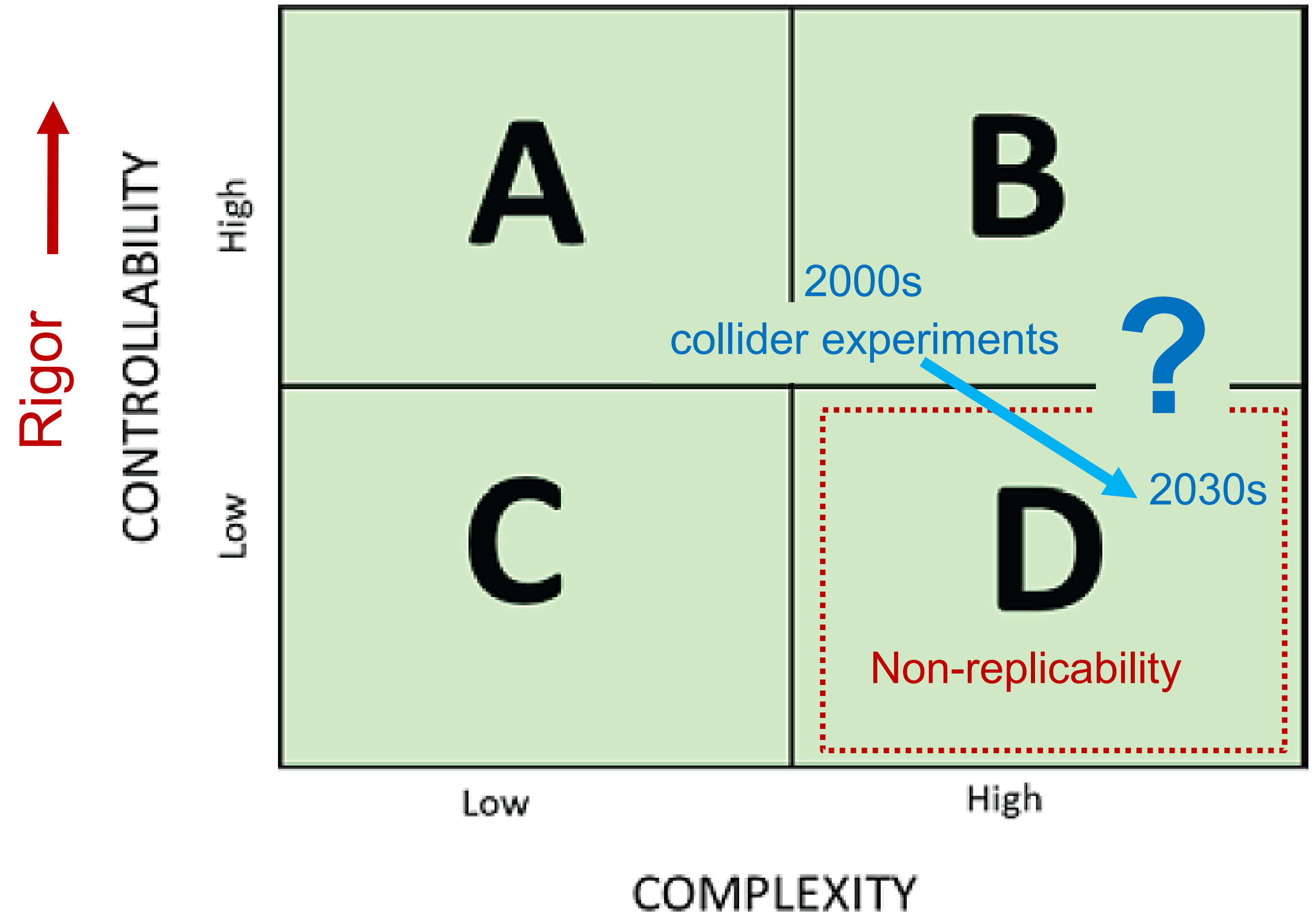
D. Acosta et al., "The Potential of a TeV-Scale Muon-Ion Collider," arXiv:2203.06258 [hep-ph]
LHeC, FCC-he Study Group, arXiv:1206.2913, 2007.14491



Exceptional machines for BSM discoveries, Higgs physics such as measurement of $\kappa_{H\to c\bar{c}}$, and SM tests at (sub)percent precision



Future scenarios for QCD precision analysis

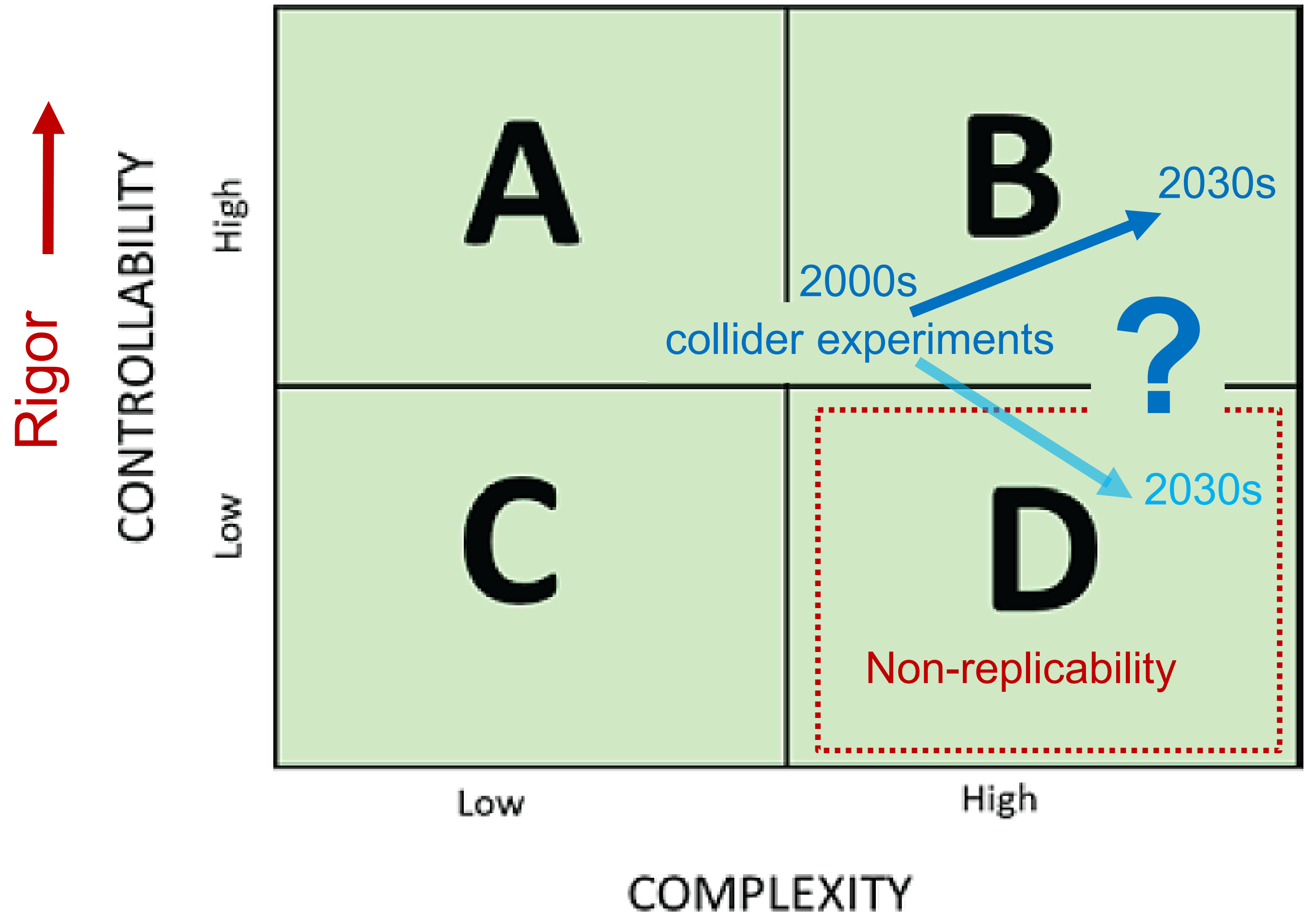


significant risk with current practices

similar to non-replicability issues in other STEM fields

Based on Fig. 5.2 in "REPRODUCIBILITY AND REPLICABILITY IN SCIENCE"

Future scenarios for QCD precision analysis



preferred scenario; requires a coordinated community strategy to adopt the **replicability mindset**

Based on Fig. 5.2 in "REPRODUCIBILITY AND REPLICABILITY IN SCIENCE"

Epistemic PDF uncertainty…

…reflects **methodological choices** such as PDF functional forms, NN architecture and hyperparameters, or model for systematic uncertainties

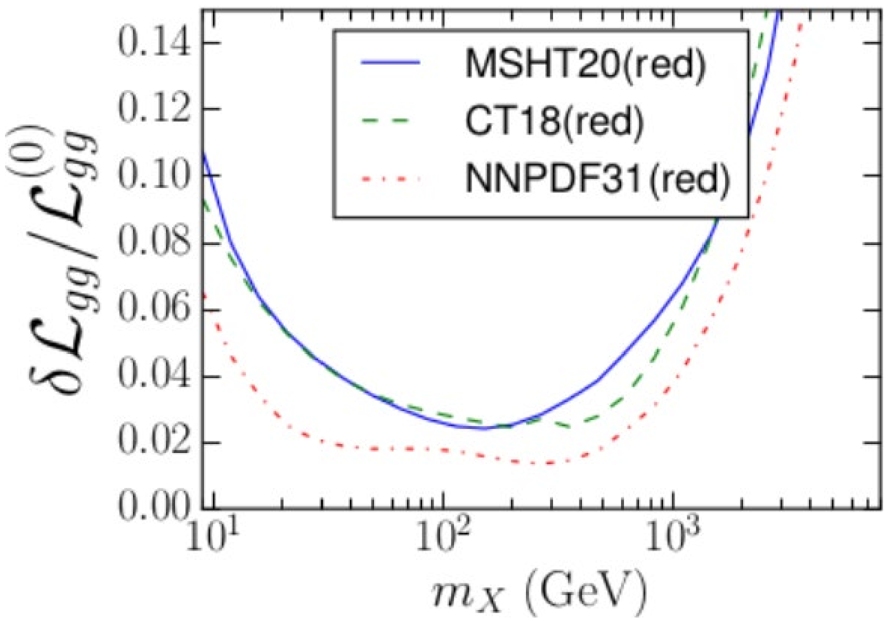
… can dominate the full uncertainty when experimental and theoretical uncertainties are small.

…is associated with the **prior probability**.

… can be estimated by **representative sampling** of the PDF solutions obtained with acceptable methodologies.

⇒ sampling over choices of experiments, PDF/NN functional space, models of correlated uncertainties…

⇒ in addition to sampling over data fluctuations



Epistemic uncertainties explain many of the differences among the sizes of PDF uncertainties by CT, MSHT, and NNPDF global fits to the same or similar data

Details in **arXiv:2203.05506**, **arXiv:2205.10444**

Two tips for improving replicability

1. With $O(10-1000)$ free parameters, including nuisance parameters, the $\Delta\chi^2 = 1$ criterion for 1σ PDF uncertainties is almost certainly incomplete. Stop using it “as is”. There are strong mathematical reasons.

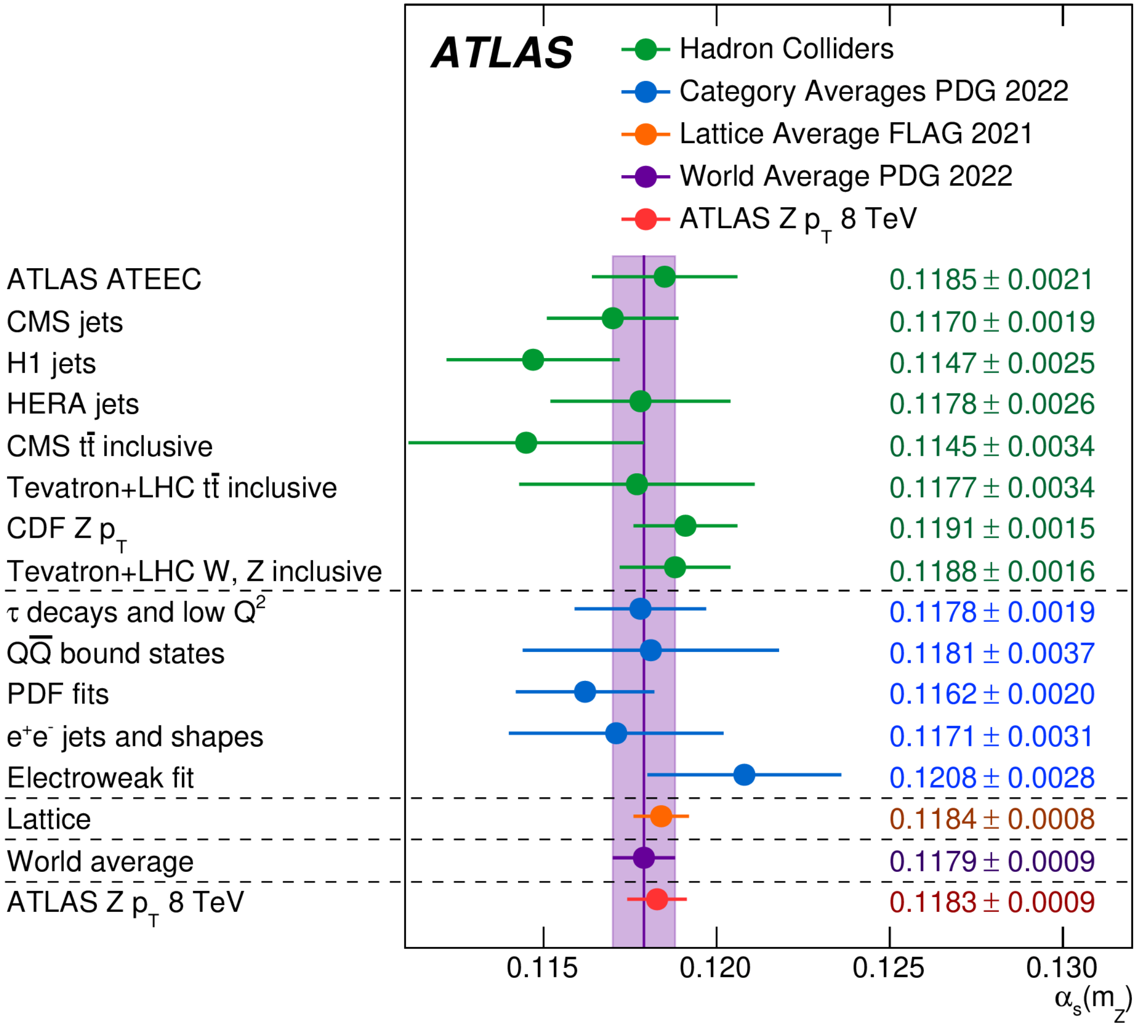
2. Thoroughly estimate the dependence on PDF parametrization forms, NN hyperparameters, and analysis settings when other uncertainties are small.

   - Public tools for this are increasingly available: xFitter, NNPDF code, ePump, Fantômas, MP4LHC,…

ATLAS measures strength of the strong force with record precision

The result showcases the power of the LHC to push the precision frontier and improve our understanding of nature

25 SEPTEMBER, 2023 arXiv:2309.12986



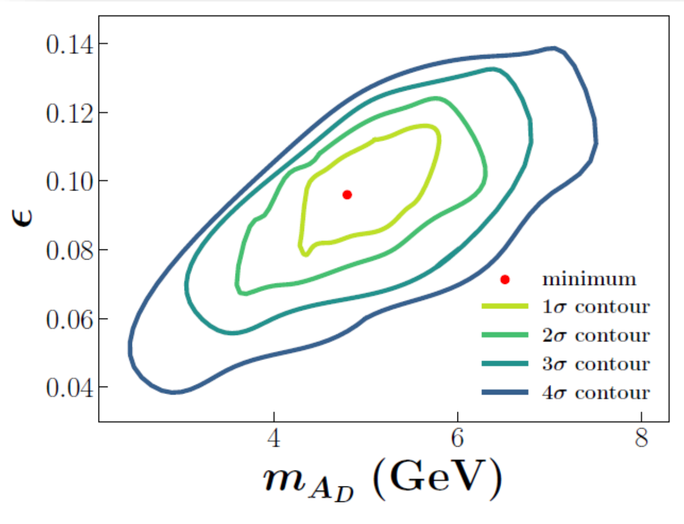
A novel determination of $\alpha_Z(M_Z)$ from $Z\ q_T$ data. However, the PDF uncertainties were not estimated properly

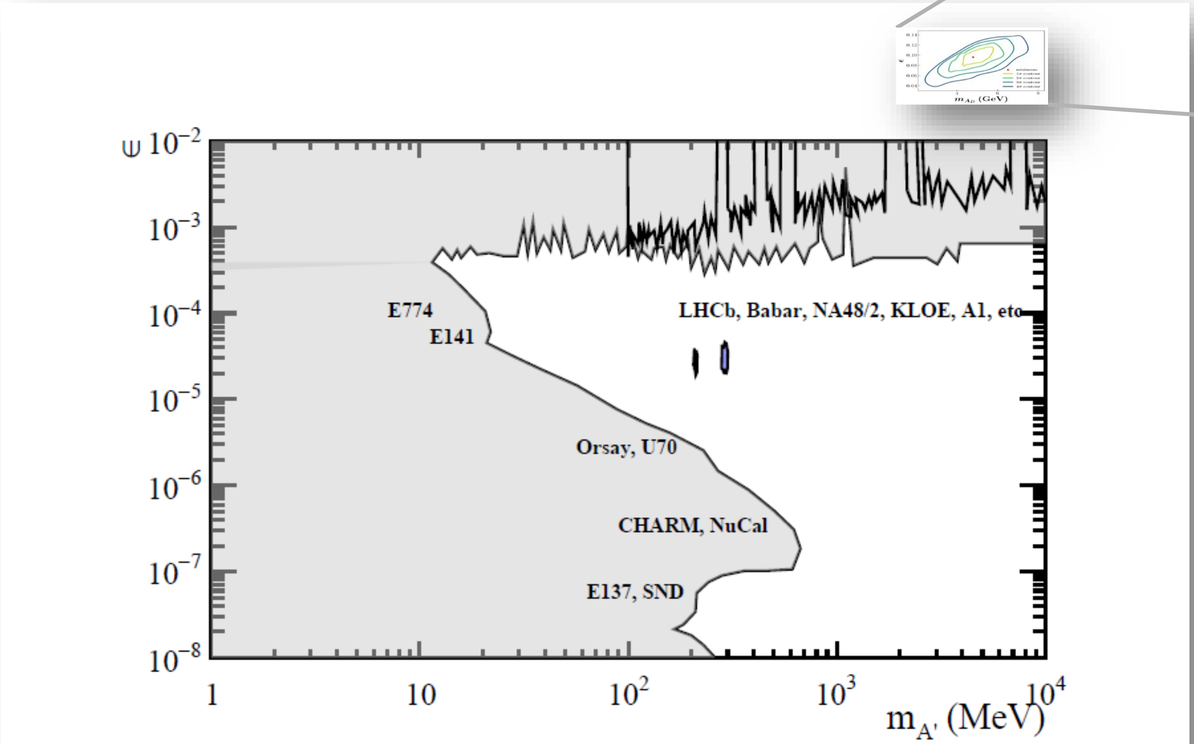
Profiling of global PDFs using $\Delta\chi^2 = 1$ ⇒ Underestimated uncertainties ⇒ Non-replicable result

[Details in T.J. Hou et al., 1912.10053, Appendix F]

A dark photon inside the proton?

N. T. Hunt-Smith, W. Melnitchouk, N. Sato, A. W. Thomas, X. G. Wang, M. J. White, *JHEP 09 (2023) 096*





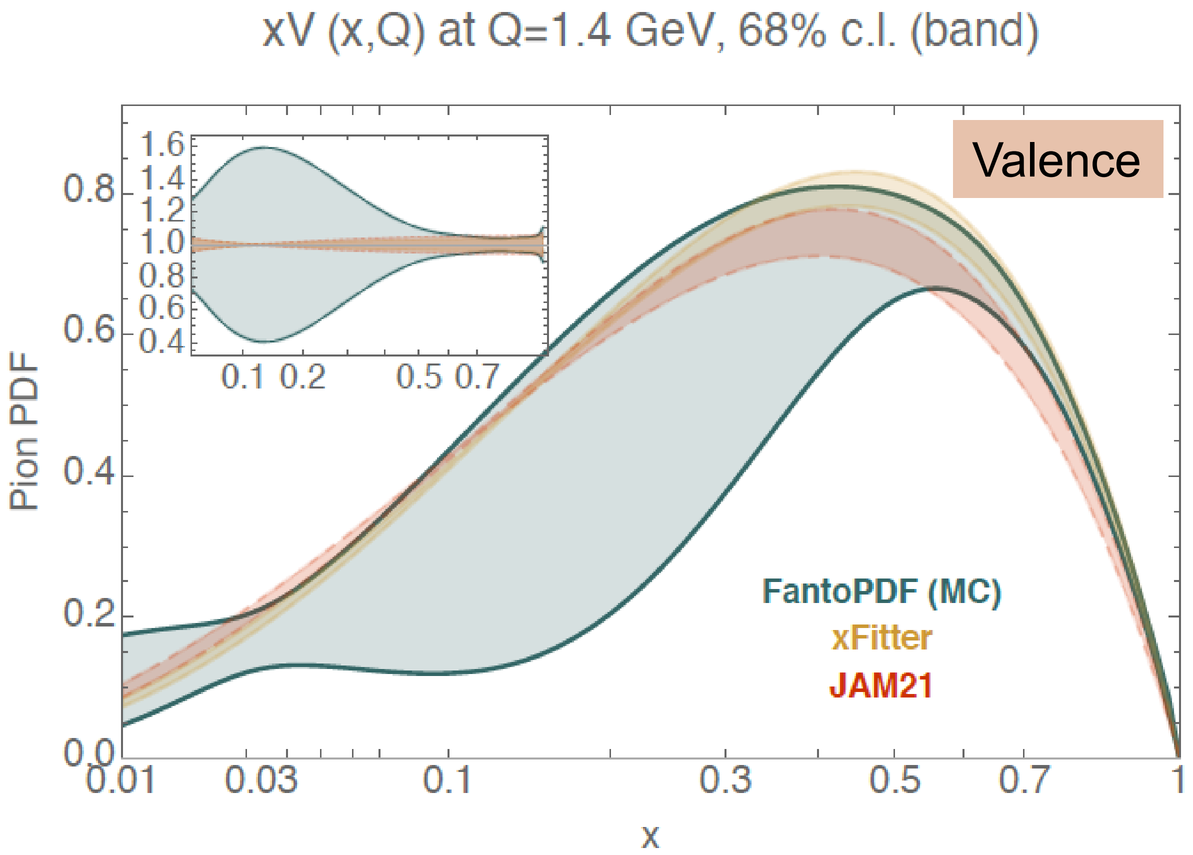
Current limits for Dark Photon in visible decays in the plane mixing strength ϵ versus mass of the Dark Photon $m_{A'}$. *arXiv:1901.09966.*

- An NLO global fit of proton data + muon $g-2$
- A fixed PDF parametrization,
$$f(x,Q_0^2) = Nx^{\alpha}(1-x)^{\beta} \times (1+\gamma\sqrt{x}+\eta x)$$
- States a 6.5σ preference for a dark photon model in the parameter region excluded by other experiments
- Used the $\Delta\chi^2 = 1$ criterion. The "dark photon" results in $\Delta\chi^2 \approx -66$ for $N_{dof} = 3283$. Alternatively, such $\Delta\chi^2$ is straightforwardly explained by the epistemic uncertainty that leads to $\Delta\chi^2 > 1$.

Fantômas: the parametrization uncertainty on the valence pion PDF

$\pi^{\pm}$ NEW

L. Kotz, A. Courtoy, M. Chavez, P. Nadolsky, F. Olness, and others, arXiv:2309.00152, arXiv:2311.XXXXX



We obtained an NLO PDF error ensemble for charged pions from experimental data in **xFitter** using a C++ module **Fantômas** to parameterize PDFs using **Bézier curves**

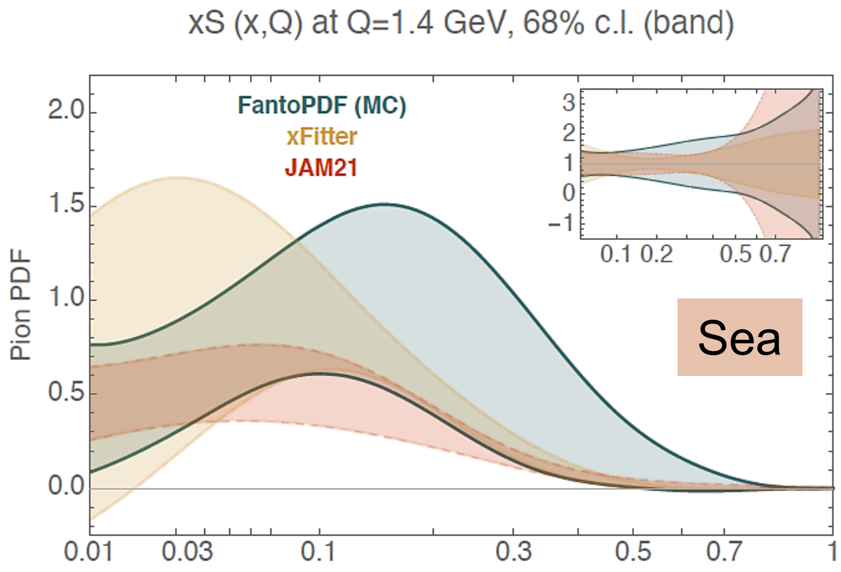
The Fantomas PDF error band is based on ~ 100 alternative parametrization forms with the same or better χ^2 as in the 2021 xFitter study *[Novikov et al., arXiv:2002.02902]*

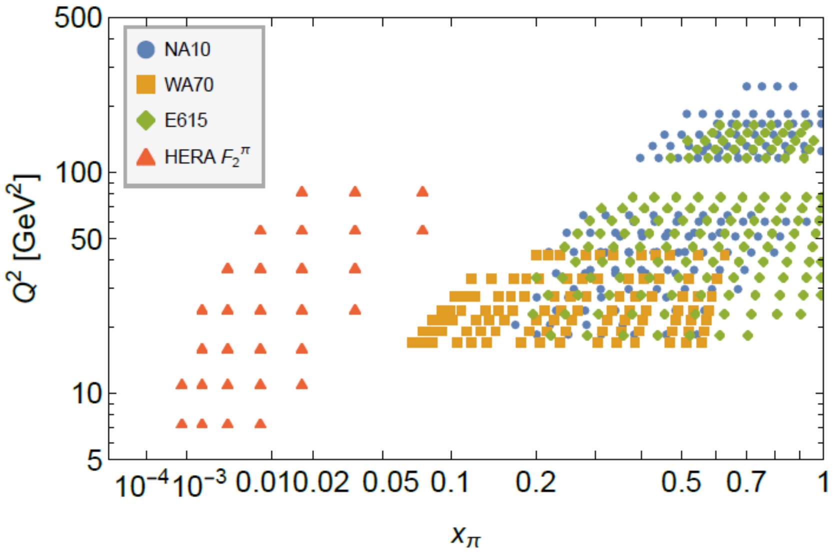
The PDF error bands are enlarged compared to xFitter'20 and JAM'21 due to estimating the parametrization uncertainty using the Fantômas & METAPDF *[arXiv:1401.00013]* techniques

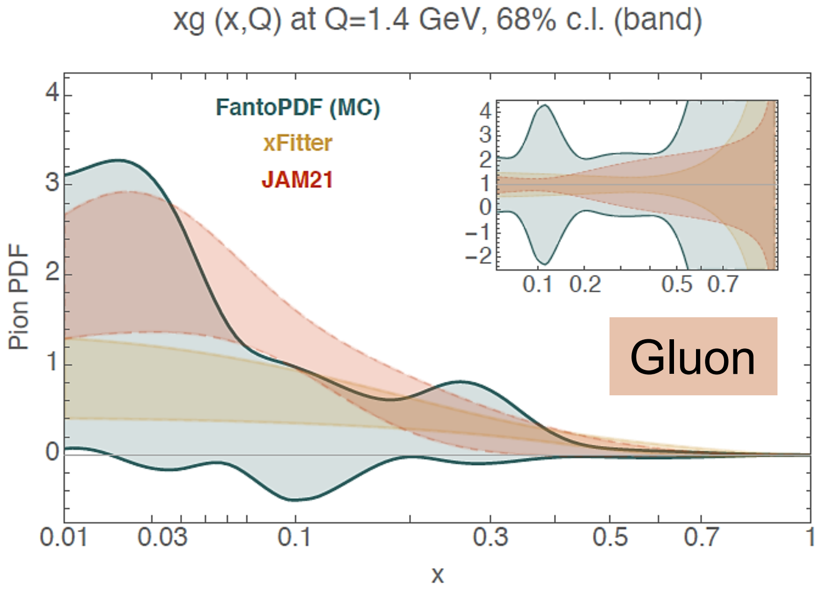
Fantômas pion PDFs: other results

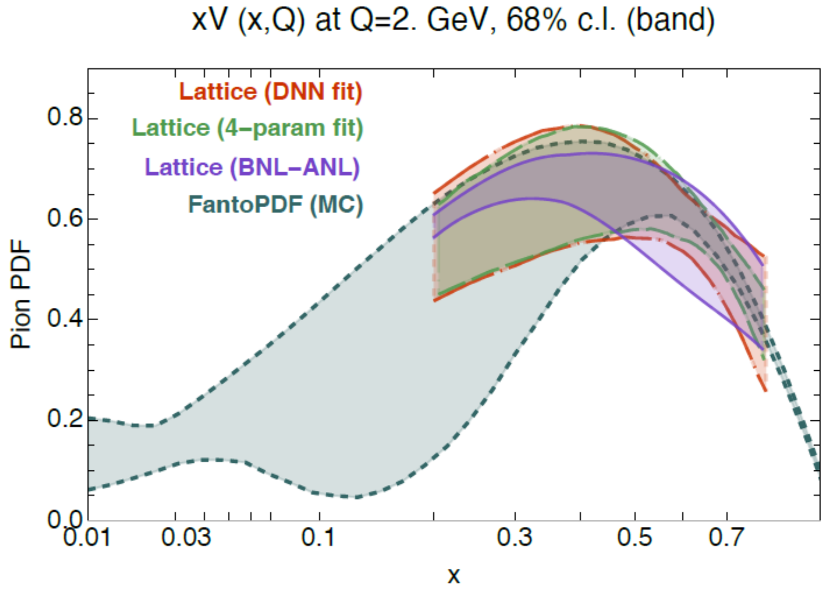
L. Kotz, A. Courtoy, M. Chavez, P. Nadolsky, F. Olness, and others, arXiv:2309.00152, arXiv:2311.XXXXX

$\pi^{\pm}$ NEW



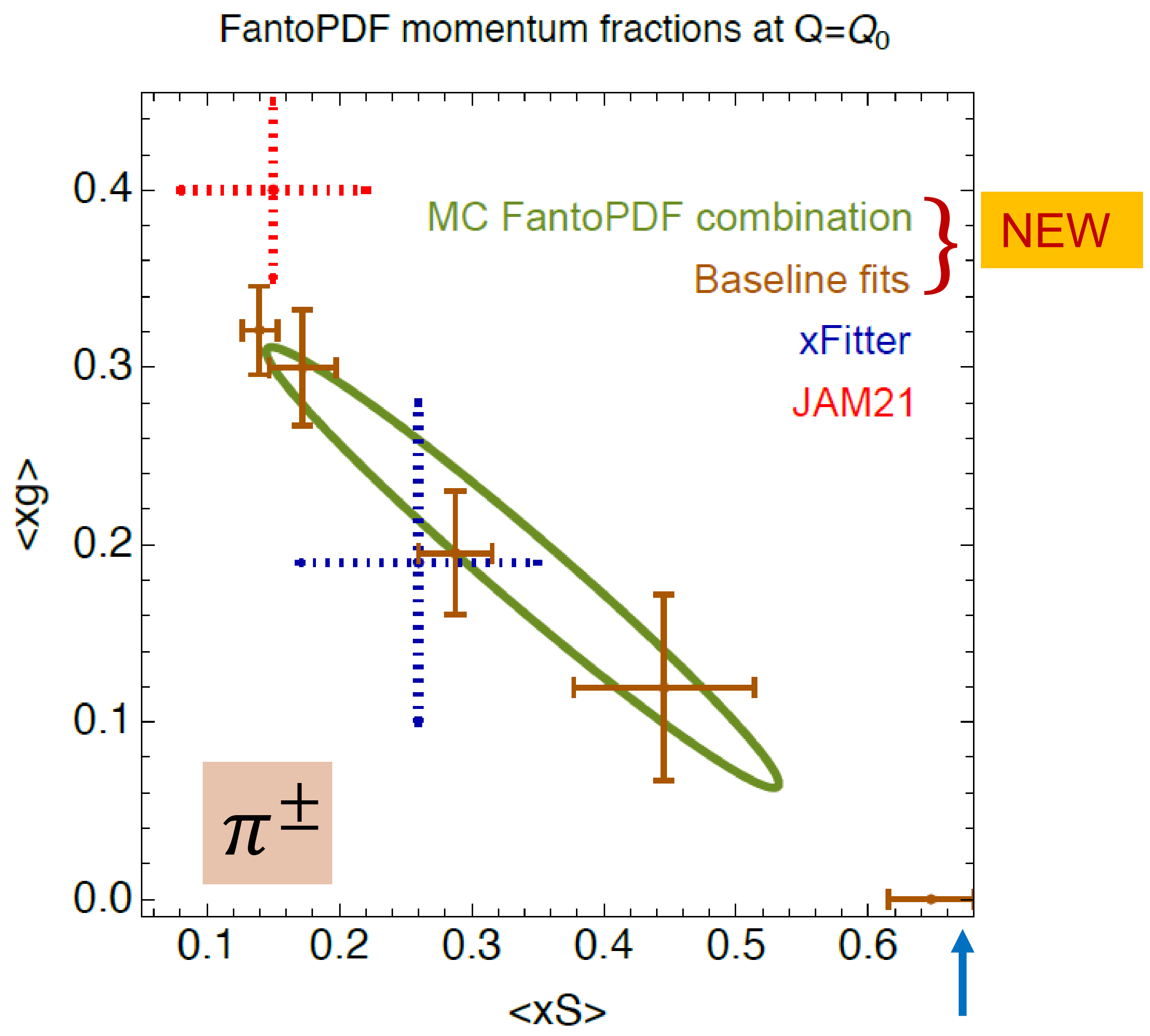






Fantômas pion PDFs: sea and gluon momentum fractions

L. Kotz, A. Courtoy, M. Chavez, P. Nadolsky, F. Olness, and others, arXiv:2309.00152, arXiv:2311.XXXXX



a zero-momentum gluon is experimentally allowed

What does lattice QCD say about the uncertainties on $\langle xf \rangle_{\pi^{\pm}}$?

Key points

1. High-luminosity Drell-Yan pair production at the LHC will distinguish between the low-sea and high-sea scenarios for light antiquarks at $x \to 1$

- Consequences for QCD models, BSM searches

2. Strangeness and charm charge asymmetries are valuable pheno probes of nonpertubative dynamics

- First lattice QCD calculations predict vanishing $s(x) - \bar{s}(x)$ at $x \to 1$
- Meson-baryon models predict large $c(x) - \bar{c}(x)$ at $x \to 1$

} More studies are needed

3. Size, shape of nonperturbatve charm remain uncertain

- unresolved theoretical ambiguities in connecting intrinsic charm (IC) and fitted charm (FC)
- need more NNLO and better showering calculations (e.g., for $Z + c$)
- promising experiments at the LHC; EIC; CERN FPF

4. Replicable PDF uncertainty quantification is central for the control over complexity

- Requires collaborations between theorists and experimentalists to establish the replicability mindset in the pheno and lattice QCD communities

Strategies for improving replicability and reproducibility

Preselection of planned studies based on their likely replicability

Detailed documentation of methods and uncertainty quantification in the publications

Training of researchers in relevant statistical methods

Journal policies that encourage replicability

Support from the funding agencies for the research infrastructure and collaborations focusing on replicability

Support for open publication of the analysis codes and key data, using agreed-upon formats

"Skin-in-the-game" incentives for researchers to produce replicable results

Based on "REPRODUCIBILITY AND REPLICABILITY IN SCIENCE"

Backup

QCD at 1% accuracy

N2LO and N3LO calculations	QCD infrastructure for these calculations	representative uncertainty estimates	systemwide processes and standards for accuracy control
Lots of promise in this area	Parton showers, fast NxLO interfaces, PDFs, ... must be comparably accurate	or The Importance of Being Earnest with Systematic Errors (experiment+theory; traditional or AI/ML)	This must be a part of the precision-focused community culture

Publishing statistical models: Getting the most out of particle physics experiments

Kyle Cranmer (New York U.), Sabine Kraml (LPSC, Grenoble), Harrison B. Prosper (Florida State U.), Philip Bechtle (Bonn U.), Florian U. Bernlochner (Bonn U.) Show All(33)

Sep 10, 2021

60 pages

Published in: *SciPost Phys.* 12 (2022) 1, 037, *SciPost Phys.* 12 (2022) 037

Published: Jan 25, 2022

e-Print: 2109.04981 [hep-ph]

2023 US DOE Funding Opportunity Announcement DE-FOA-0000315

*Advancing **Uncertainty Quantification** in Modeling, Simulation, and Analysis of **Complex Systems***

AI/ML techniques are superb for finding an excellent fit to data.

Are these techniques adequate for uncertainty estimation [exploring all good fits]?

A common resampling procedure used by experimentalists and theorists:

1. Train a neural network model T_i with $\mathrm{N_{par}}$ (hyper)parameters on a randomly fluctuated replica of discrete data D_i. Repeat $\mathrm{N_{rep}}$ times. In a typical application: $\mathrm{N_{par}} > 10^2$, $\mathrm{N_{rep}} < 10^4$.
2. Out of $\mathrm{N_{rep}}$ replicas T_i with "good" description of data [i.e., with a high likelihood $P(D_i|T_i) \propto e^{-\chi^2(D_i,T_i)/2}$], discard "badly behaving" (overfitted, not smooth, …) replicas
3. Estimate the uncertainties of T_i using the remaining "well-behaved" replicas

Is this procedure rigorous? How many N_{rep} replicas does one need?

A likelihood-ratio test of NN models T_1 and T_2

From Bayes theorem, it follows that

$$\frac{P(T_2|D)}{P(T_1|D)} = \frac{P(D|T_2)}{P(D|T_1)} \times \frac{P(T_2)}{P(T_1)}$$

$\equiv r_{\text{posterior}}$ $\equiv r_{\text{likelihood}}$ $\equiv r_{\text{prior}}$

aleatory epistemic + aleatory

Suppose replicas T_1 and T_2 have the same χ^2 $\left[r_{\text{likelihood}} = \exp\left(\frac{\chi_1^2 - \chi_2^2}{2}\right) = 1\right]$, but T_2 is disfavored compared to T_1 $[r_{\text{posterior}} \ll 1]$.

This only happens if $r_{\text{prior}} \ll 1$: T_2 is discarded based on its **prior** probability.

Statistics with many parameters is different!

In many applications, especially AI/ML ones:

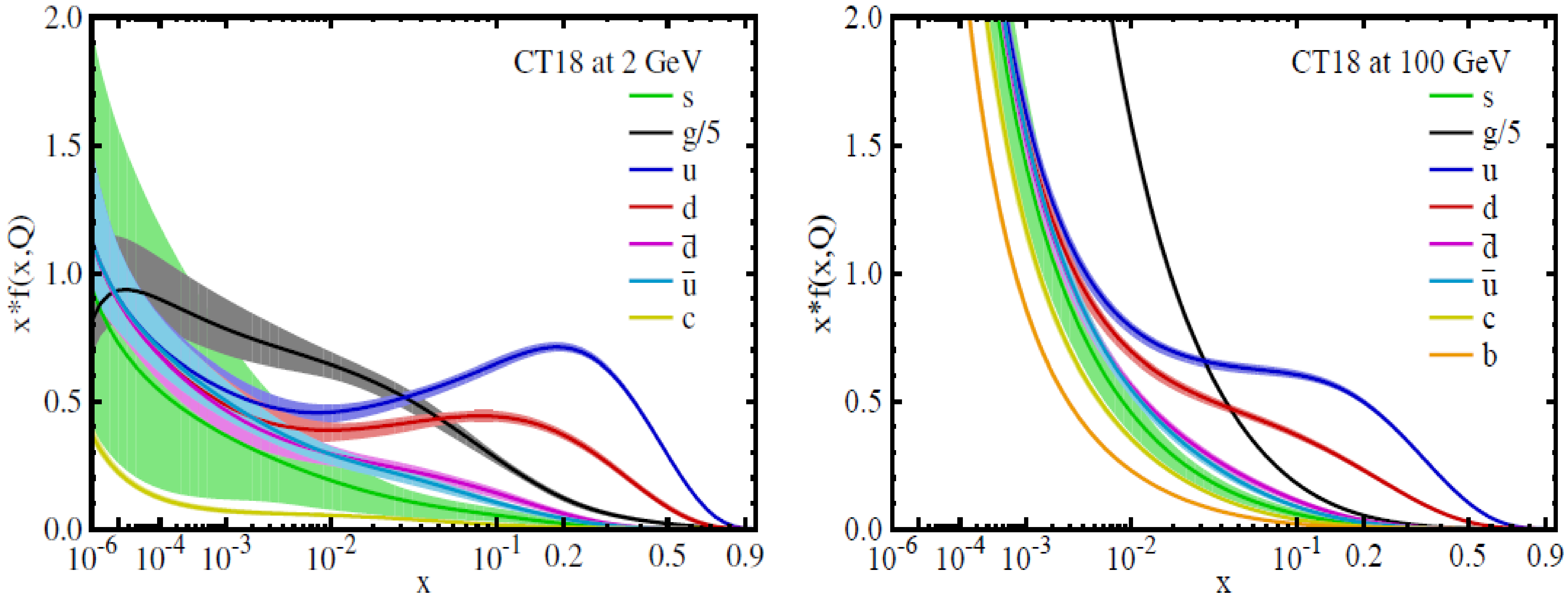
1. **There is no single global minimum of** χ^2 (or another cost function)
2. **The law of large numbers may not work**
   – uncertainty may not decrease as $1/\sqrt{N_{rep}}$, leading to the **big-data paradox** [Xiao-Li Meng, 2018]:

The bigger the data, the surer we fool ourselves.

3. **Replication of complex measurements is daunting**

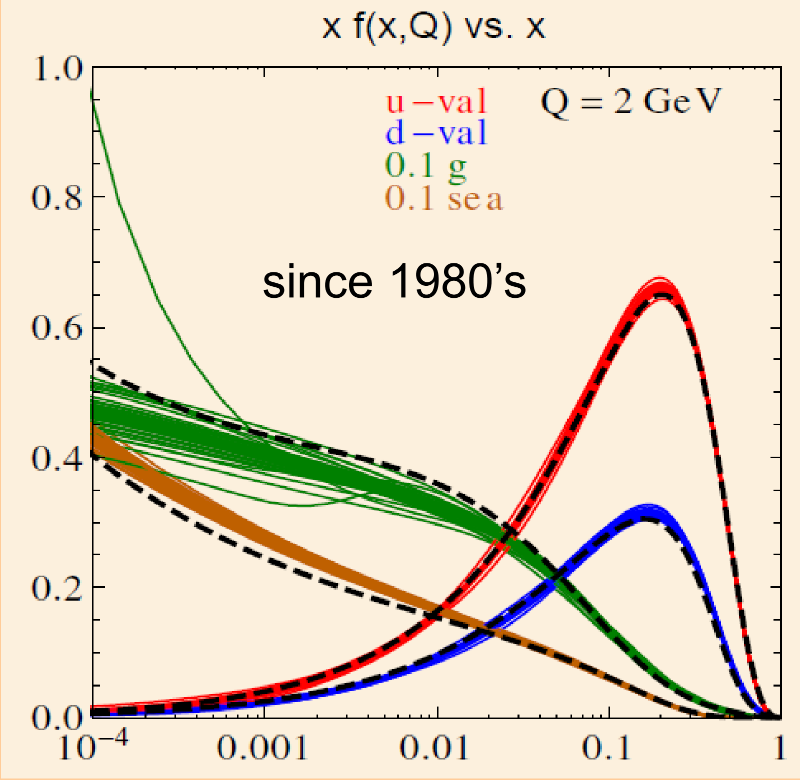
CT18 parton distributions

Recent PDFs from the CTEQ-TEA group arXiv:1912.10053 [hep-ph]

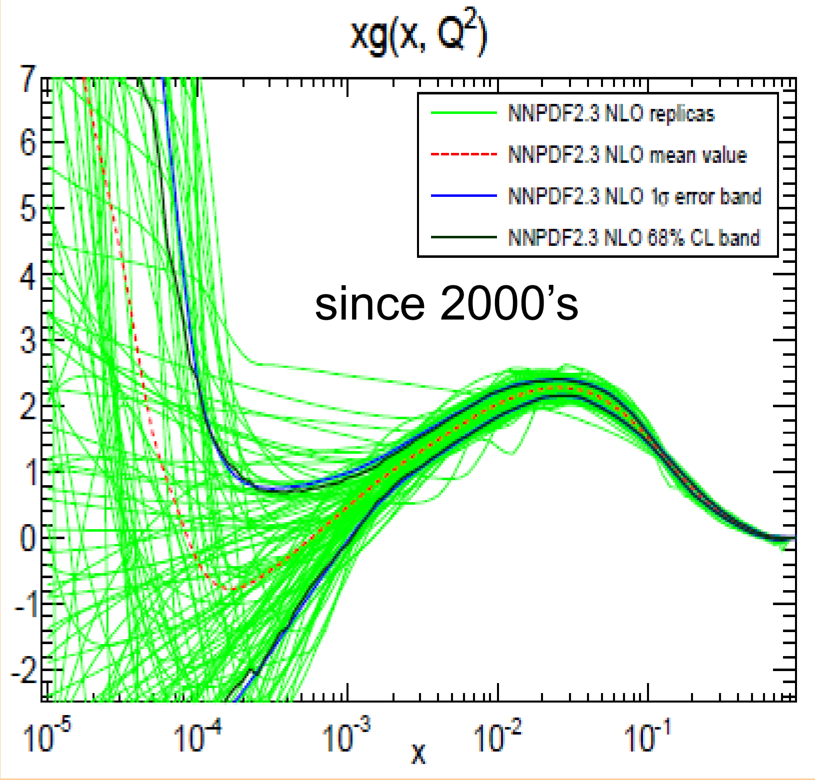


Two types of modern error PDFs

Analytic parametrizations +
Hessian PDF eigenvector sets
(ABM, CTEQ, HERA, MMHT,…)



Neural network parameterizations
+ Monte Carlo PDF replicas
(NNPDF)

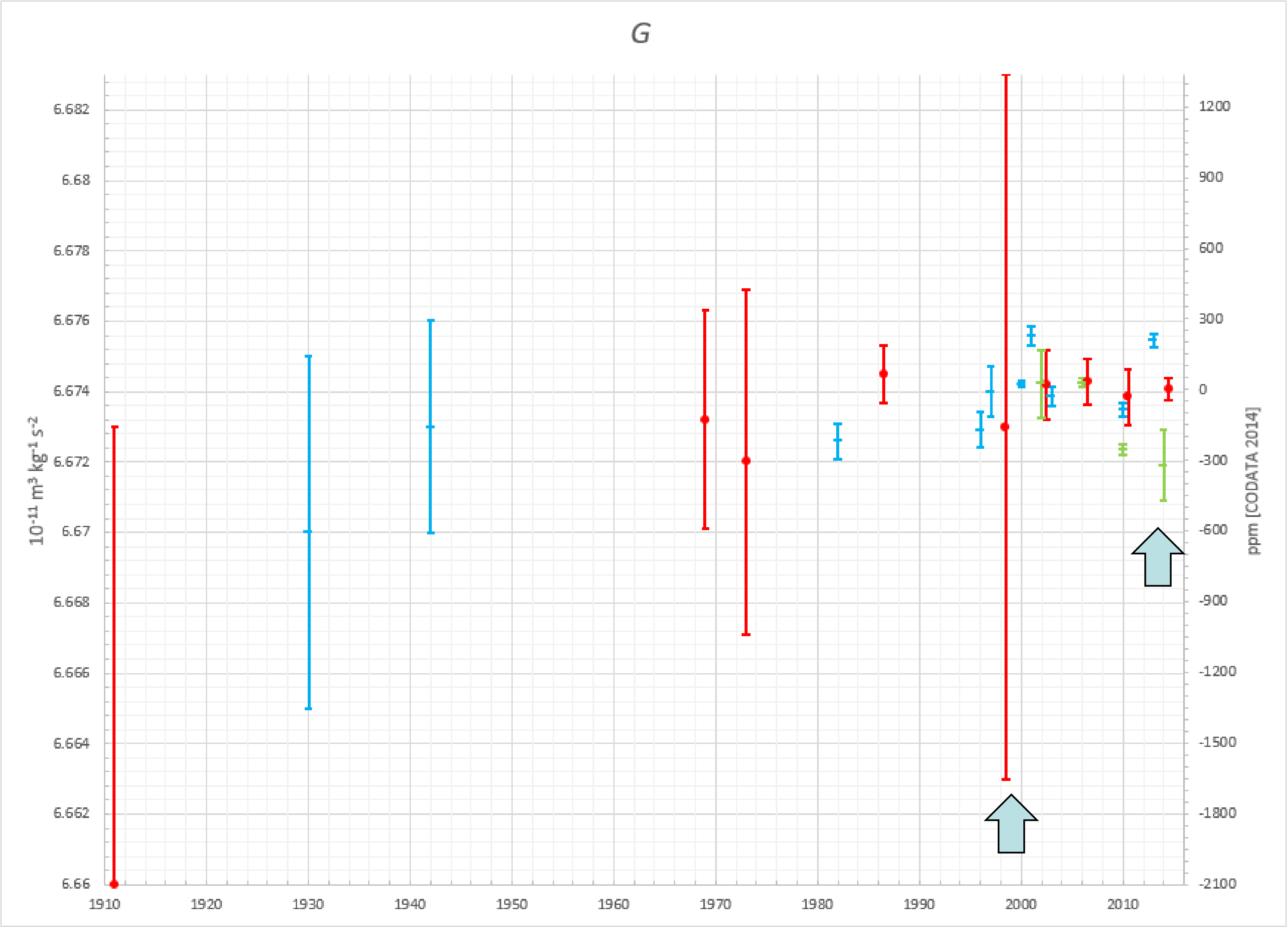


a textbook application of ML in particle physics

Two powerful, complementary representations.
Hessian PDFs can be converted into MC ones, and vice versa.

Example: determinations of QCD coupling α_s

World average for the gravitational constant



Timeline of measurements and recommended values for *G* since 1900: values recommended based on the NIST combination (red), individual torsion balance experiments (blue), other types of experiments (green).

The combination error bars are unstable after 1995

Some precise individual measurements are in a conflict among themselves and post-2014 combination

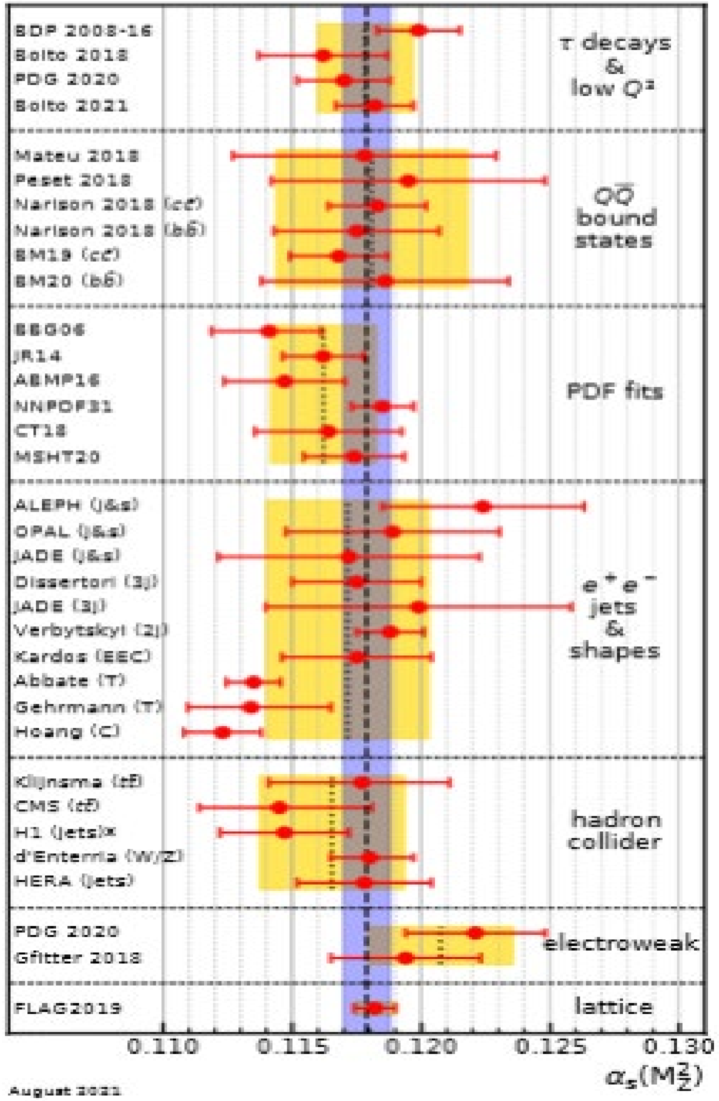
https://en.wikipedia.org/wiki/Gravitational_constant#Modern_value, retrieved on Oct. 22, 2023

Future measurements of the QCD coupling

individual α_s measurements can reach precision of $\sim 0.1\%$

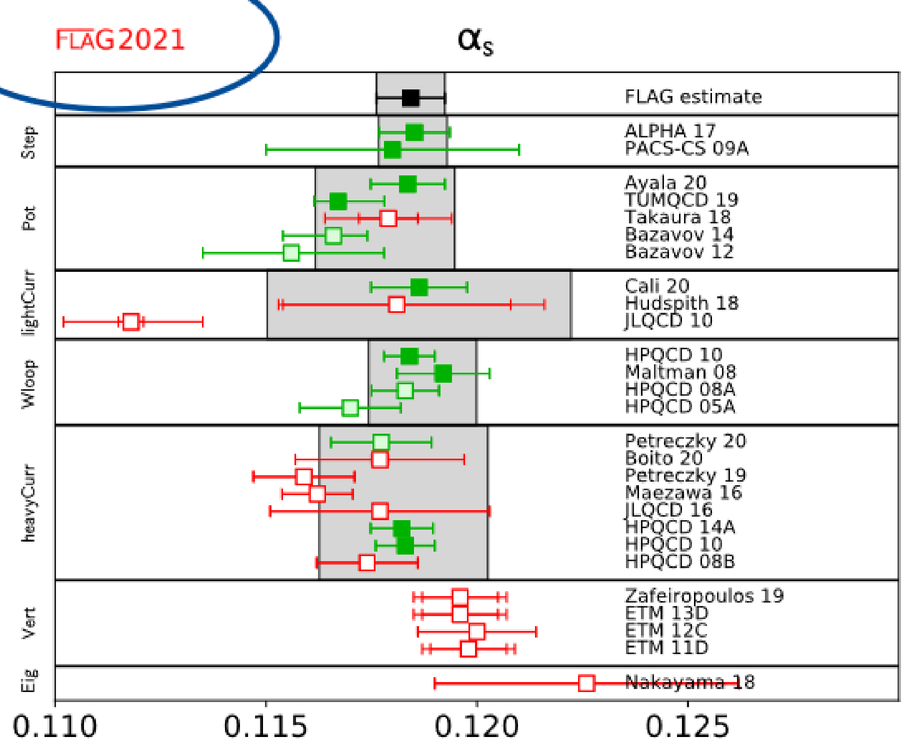
and symbols: CIPT='contour-improved perturbation theory', FOPT='fixed-order perturbation theory', NP='nonperturbative QCD', SF='structure functions', PS='Monte Carlo parton shower'.

Method	Relative $\alpha_s mZ$ uncertainty: Current theory & exp. uncertainties sources	Relative $\alpha_s mZ$ uncertainty: Near (long-term) future theory & experimental progress
(1) Lattice	0.7% Finite lattice spacing & stats. $\mathrm{N}^{2,3}$LO pQCD truncation	$\approx$ 0.3% (0.1%) Reduced latt. spacing. Add more observables Add $\mathrm{N}^{3,4}$LO, active charm (QED effects) Higher renorm. scale via step-scaling to more observ.
(2) τ decays	1.6% N^3LO CIPT vs. FOPT diffs. Limited τ spectral data	< 1.% Add N^4LO terms. Solve CIPT–FOPT diffs. Improved τ spectral functions at Belle II
(3) $Q\bar{Q}$ bound states	3.3% $\mathrm{N}^{2,3}$LO pQCD truncation $m_{c,b}$ uncertainties	$\approx$ 1.5% Add $\mathrm{N}^{3,4}$LO & more ($c\bar{c}$), ($b\bar{b}$) bound states Combined $m_{c,b} + \alpha_s$ fits
(4) DIS & PDF fits	1.7% $\mathrm{N}^{2,(3)}$LO PDF (SF) fits Span of PDF-based results	$\approx$ 1% (0.2%) N^3LO fits. Add new SF fits: $F_2^{p,d}$, g_i (EIC) Better corr. matrices, sampling of PDF solutions. More PDF data (EIC/LHeC/FCC-eh)
(5) e^+e^- jets & evt shapes	2.6% NNLO+$\mathrm{N}^{(1,2,3)}$LL truncation Different NP analytical & PS corrs. Limited datasets w/ old detectors	$\approx$ 1.5% (< 1%) Add $\mathrm{N}^{2,3}$LO+N^3LL, power corrections Improved NP corrs. via: NNLL PS, grooming New improved data at B factories (FCC-ee)
(6) Electroweak fits	2.3% N^3LO truncation Small LEP+SLD datasets	($\approx$ 0.1%) N^4LO, reduced param. uncerts. ($m_{W,Z}$, α, CKM) Add W boson. Tera-Z, Oku-W datasets (FCC-ee)
(7) Hadron colliders	2.4% NNLO(+NNLL) truncation, PDF uncerts. Limited data sets ($t\bar{t}$, W, Z, e-p jets)	$\approx$ 1.5% N^3LO+NNLL (for color-singlets), improved PDFs Add more datasets: Z p_T, p-p jets, σ_i/σ_j ratios,...
World average	0.8%	$\approx$ 0.4% (0.1%)

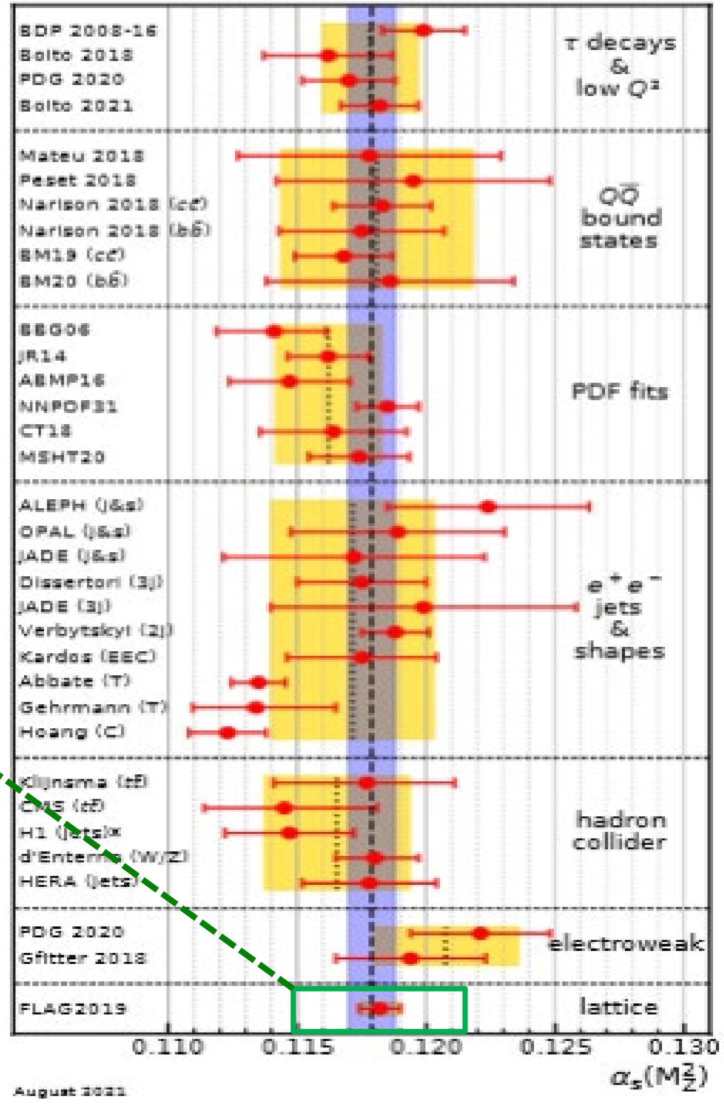


D. d'Enterria et al., EF QCD, arXiv:2203.08271

Lattice QCD & world-average α_s combination



A. Kronfeld



Lattice determinations of α_s in multiple channels are projected to be [far] more precise than many experiments. Several challenges with combining the eclectic α_s inputs with the current procedure.

Time to rethink how the world-average α_s combination is performed?

aleatory vs. **epistemic** uncertainties

Statistical uncertainty propagated from experiments
— reduced by increasing data size

Uncertainty due to lack of knowledge
or incomplete models
—bias (may be reduced by analysis improvements)

Two tips for improving replicability

1. With many parameters, the $\Delta\chi^2 = 1$ criterion for 1σ PDF uncertainties is almost certainly wrong. Stop using it "as is".

- Two common forms of χ^2 minimized w.r.t to the experimental or theoretical nuisance parameters λ_α,

$$\chi^2 = \sum_{i=1}^{N_{pt}} \frac{\left(D_i + \sum_\alpha \beta_{i,\alpha}\lambda_\alpha - T_i\right)^2}{s_i^2} + T^2 \sum_\alpha \lambda_\alpha^2 \quad \text{and} \quad \chi^2 = \sum_{i,j}^{N_{pt}} (D_i - T_i)(\mathrm{cov}^{-1})_{ij}\left(D_j - T_j\right),$$

rarely have $T^2 = 1$ and Gaussian form for $N_{par} \gg 1$ because of fundamental mathematical reasons

2. Thoroughly estimate the dependence on PDF parametrization forms or NN architecture when other uncertainties are small.

- Public tools for this are increasingly available: xFitter, NNPDF code, ePump, Fantômas, MP4LHC,…

Epistemic PDF uncertainty is important in W boson mass and α_s measurements

ATLAS-CONF-2023-004

PDF-Set	p_T^ℓ [MeV]	m_T [MeV]	combined [MeV]
CT10	$80355.6^{+15.8}_{-15.7}$	$80378.1^{+24.4}_{-24.8}$	$80355.8^{+15.7}_{-15.7}$
CT14	$80358.0^{+16.3}_{-16.3}$	$80388.8^{+25.2}_{-25.5}$	$80358.4^{+16.3}_{-16.3}$
CT18	$80360.1^{+16.3}_{-16.3}$	$80382.2^{+25.3}_{-25.3}$	$80360.4^{+16.3}_{-16.3}$
MMHT2014	$80360.3^{+15.9}_{-15.9}$	$80386.2^{+23.9}_{-24.4}$	$80361.0^{+15.9}_{-15.9}$
MSHT20	$80358.9^{+13.0}_{-16.3}$	$80379.4^{+24.6}_{-25.1}$	$80356.3^{+14.6}_{-14.6}$
NNPDF3.1	$80344.7^{+15.6}_{-15.5}$	$80354.3^{+23.6}_{-23.7}$	$80345.0^{+15.5}_{-15.5}$
NNPDF4.0	$80342.2^{+15.3}_{-15.3}$	$80354.3^{+22.3}_{-22.4}$	$80342.9^{+15.3}_{-15.3}$

Table 2: Overview of fitted values of the W boson mass for different PDF sets. The reported uncertainties are the total uncertainties.

ATLAS-CONF-2023-015

The statistical analysis for the determination of $\alpha_s(m_Z)$ is performed with the xFitter framework [60]. The value of $\alpha_s(m_Z)$ is determined by minimising a χ^2 function which includes both the experimental uncertainties and the theoretical uncertainties arising from PDF variations:

$$\chi^2(\beta_{\rm exp}, \beta_{\rm th}) = \sum_{i=1}^{N_{\rm data}} \frac{\left(\sigma_i^{\rm exp} + \sum_j \Gamma_{ij}^{\rm exp}\beta_{j,\rm exp} - \sigma_i^{\rm th} - \sum_k \Gamma_{ik}^{\rm th}\beta_{k,\rm th}\right)^2}{\Delta_i^2} + \sum_j \beta_{j,\rm exp}^2 + \sum_k \beta_{k,\rm th}^2 . \quad (1)$$

profiling of CT and MSHT PDFs requires to include a tolerance factor $T^2 > 10$ as in the ePump code

[T.J. Hou et al., 1912.10053, Appendix F]

Also the next slide.

Augmented likelihood for PDFs with global tolerance

1. Start by defining the correspondence between $\Delta\chi^2$ and cumulative probability level: 68% c.l. $\Leftrightarrow$ $\Delta\chi^2 = T^2$.
2. Write the **augmented** likelihood density for this definition:

$$P(D_i|T_i) \propto e^{-\chi^2/(2T^2)}$$

3. When profiling 1 new experiment with the prior imposed on PDF nuisance parameters $\lambda_{\alpha,th}$:

$$\chi^2(\vec{\lambda}_{\rm exp}, \vec{\lambda}_{\rm th}) = \sum_{i=1}^{N_{pt}} \frac{\left[D_i + \sum_\alpha \beta_{i,\alpha}^{\rm exp}\lambda_{\alpha,\rm exp} - T_i - \sum_\alpha \beta_{i,\alpha}^{\rm th}\lambda_{\alpha,\rm th}\right]^2}{s_i^2} + \sum_\alpha \lambda_{\alpha,\rm exp}^2 + \sum_\alpha T^2 \lambda_{\alpha,\rm th}^2. \qquad \beta_{i,\alpha}^{\rm th} = \frac{T_i(f_\alpha^+) - T_i(f_\alpha^-)}{2},$$

new experiment

priors on expt. systematics and PDF params

4. Alternatively, we can reparametrize $\chi^{2'} \equiv \chi^2/T^2$, so that 68% c.l. $\Leftrightarrow$ $\Delta\chi^{2'} = 1$. We have

$$P(D_i|T_i) \propto e^{-\chi^{2'}/2}$$

$$\chi^{2'}(\vec{\lambda}_{\rm exp}, \vec{\lambda}_{\rm th}) = \sum_{i=1}^{N_{pt}} \frac{\left[D_i + \sum_\alpha \beta_{i,\alpha}^{\rm exp}\lambda_{\alpha,\rm exp} - T_i - \sum_\alpha \beta_{i,\alpha}^{\rm th}\lambda_{\alpha,\rm th}\right]^2}{s_i^2\ T^2} + \sum_\alpha \frac{\lambda_{\alpha,\rm exp}^2}{T^2} + \sum_\alpha \lambda_{\alpha,\rm th}^2.$$

consistent redefinition

5. Inconsistent redefinitions:

$$\chi^{2'}(\vec{\lambda}_{\rm exp}, \vec{\lambda}_{\rm th}) = \sum_{i=1}^{N_{pt}} \frac{\left[D_i + \sum_\alpha \beta_{i,\alpha}^{\rm exp}\lambda_{\alpha,\rm exp} - T_i - \sum_\alpha \beta_{i,\alpha}^{\rm th}\lambda_{\alpha,\rm th}\right]^2}{s_i^2} + \sum_\alpha \lambda_{\alpha,\rm exp}^2 + \sum_\alpha \lambda_{\alpha,\rm th}^2.$$

and $P(D_i|T_i) \propto e^{-\chi^{2'}/2}$

or $P(D_i|T_i) \propto e^{-\chi^{2'}/(2T^2)}$

[equivalent to $s_i \to s_i/T$ or $\lambda_{\alpha,th} \to \lambda_{\alpha,th}T$ without $\beta_{i,\alpha,th} \to \beta_{i,\alpha,th}/T$]

Why **augmented** likelihood?

The term is accepted in lattice QCD to indicate that the log-likelihood contains **prior terms**

$$\chi^2(\vec{\lambda}_{\rm exp}, \vec{\lambda}_{\rm th}) = \sum_{i=1}^{N_{pt}} \frac{\left[D_i + \sum_\alpha \beta^{\rm exp}_{i,\alpha} \lambda_{\alpha,\rm exp} - T_i - \sum_\alpha \beta^{\rm th}_{i,\alpha} \lambda_{\alpha,\rm th}\right]^2}{s_i^2} + \sum_\alpha \lambda^2_{\alpha,\rm exp} + \sum_\alpha T^2 \lambda^2_{\alpha,\rm th}.$$

new experiment

priors on expt. systematics and PDF params

After minimization w.r.t. to $\lambda_{\alpha,exp}$, $\lambda_{\alpha,th}$, the prior terms are **hidden** inside the covariance matrix:

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

The usual χ^2 definition therefore contains a **prior** component, which may be handled differently by the various groups

Ongoing studies of systematic uncertainties are essential and still insufficient

- from the experiment side

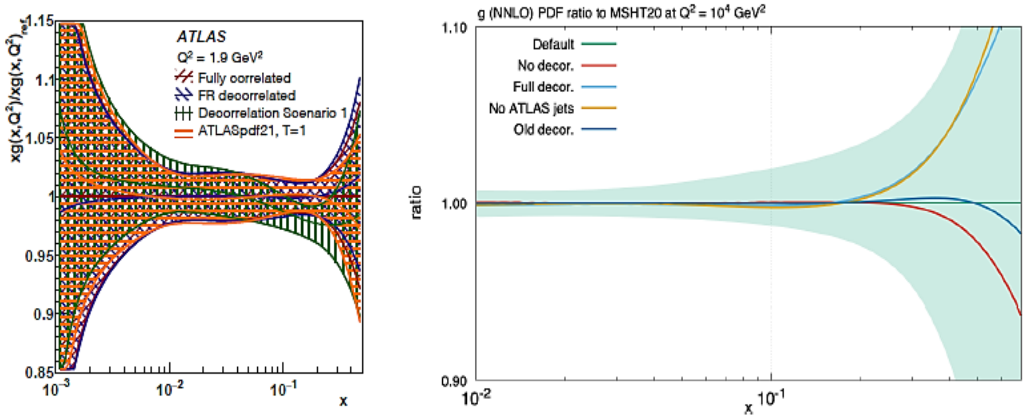


FIG. 9. Difference in the gluon PDF shown in ratio to the ATLASpdf21 (default) gluon(left). This default uses Decorrelation Scenario 2 and this is compared to the use of Full Correlation, Full decorrelation of the flavour response systematic and Decorrelation Scenario 1. The effect of no decorrelation, the default correlation of [9], the decorrelation in [362], and full decorrelation for the MSHT20 gluon (right).

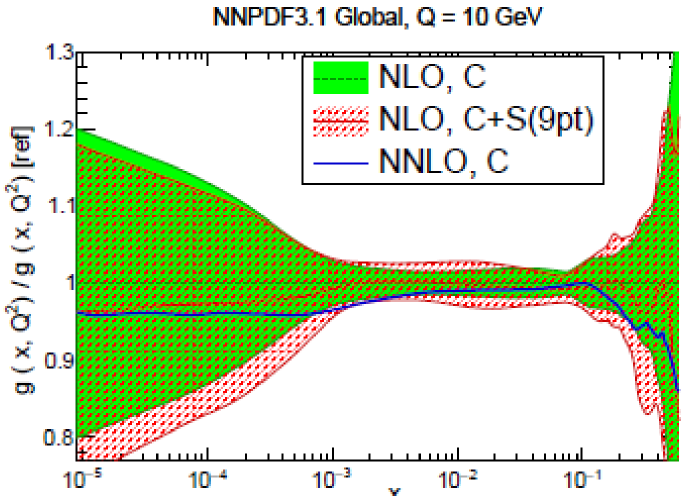
S. Amoroso et al., 2203.13923, Sec. 5.A

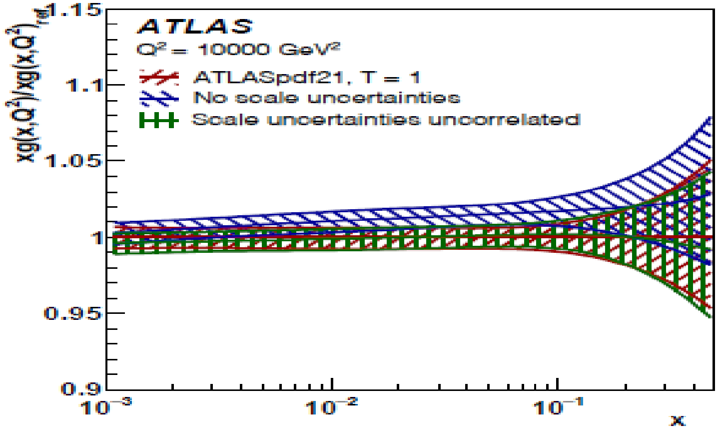
Strong dependence on the definition of corr. syst. errors raises a general concern:

Overreliance on Gaussian distributions and covariance matrices for poorly understood effects may produce very wrong uncertainty estimates

[N. Taleb, Black Swan & Antifragile]

- from the theory side





Examples: studies of theory uncertainties in the PDFs by NNPDF3.1 and ATLAS21

Two common forms of χ^2 in PDF fits

1. In terms of nuisance parameters $\lambda_{\alpha,exp}$

$$\chi^2 = \sum_{i=1}^{N_{pt}} \frac{\left[D_i + \sum_\alpha \beta_{i,\alpha}^{\exp} \lambda_{\alpha,\exp} - T_i\right]^2}{s_i^2} + \sum_\alpha \lambda_{\alpha,\exp}^2$$

2. In terms of the covariance matrix

$\Downarrow$ algebraic minimization of χ^2 with respect to $\lambda_{\alpha,exp}$

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

$$(\mathrm{cov})_{ij} \equiv s_i^2 \delta_{ij} + \sum_{\alpha=1}^{N_\lambda} \beta_{i,\alpha}\beta_{j,\alpha}, \qquad \beta_{i,\alpha} = \sigma_{i,\alpha} X_i,$$

D_i, T_i, s_i are the central data, theory, uncorrelated error
$\beta_{i,\alpha}$ is the correlation matrix for N_λ nuisance parameters.

Experiments publish $\sigma_{i,\alpha}$ (up to hundreds per data set). To reconstruct $\beta_{i,\alpha}$, we need to decide on the normalizations X_i. Possible choices:

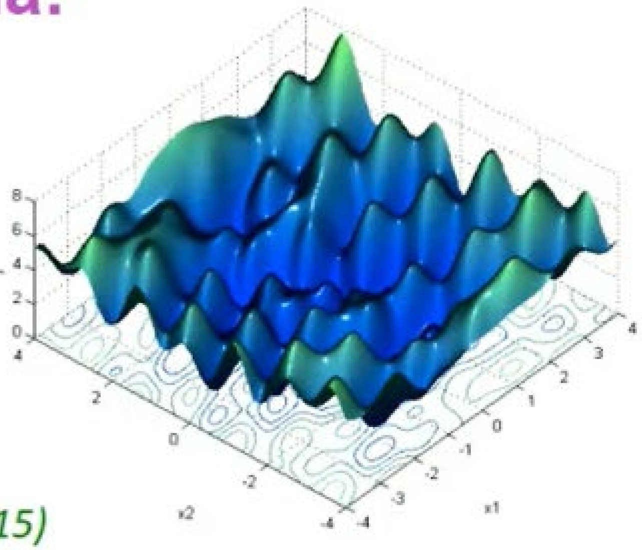
a. $X_i = D_i$: "**exp**erimental scheme"; can result in a bias
b. $X_i =$ fixed or varied T_i : "$\boldsymbol{t_0}$, T, extended T schemes"; can result in (different) biases

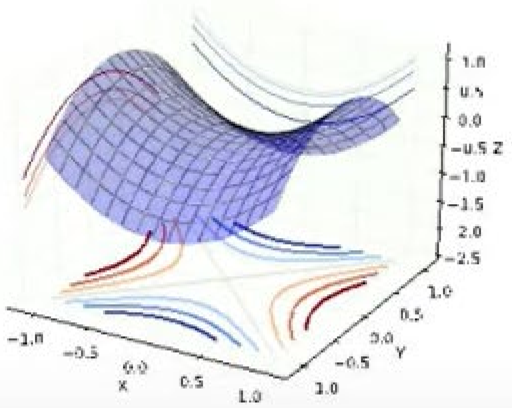
Not so terrible local minima: convexity is not needed

Myth busted:

- Local minima dominate in low-D, but saddle points dominate in high-D
- Most local minima are relatively close to the bottom (global minimum error)

(Dauphin et al NIPS'2014, Choromanska et al AISTATS'2015)





Global minimum: all $\frac{\partial^2\chi^2}{\partial a_i \partial a_j} > 0$ (improbable)

Saddle point: some $\frac{\partial^2\chi^2}{\partial a_i \partial a_j} > 0$ (probable)

An average global minimum: in properly chosen coordinates, $\frac{\partial^2\chi^2}{\partial z_i \partial z_j} > 0$ for dominant coordinate components

Y. Bengio, 2019 Turing lecture (YouTube)

Many dimensions introduce major difficulties with finding a global minimum…

The Loss Surfaces of Multilayer Networks
A. Choromanska, M. Henaff, M. Mathieu, G. Ben Arous, Y. LeCun PMLR 38:192-204, 2015

An important question concerns the distribution of critical points (maxima, minima, and saddle points) of such functions. Results from random matrix theory applied to spherical spin glasses have shown that these functions have a combinatorially large number of saddle points. Loss surfaces for large neural nets have many local minima that are essentially equivalent from the point of view of the test error, and these minima tend to be highly degenerate, with many eigenvalues of the Hessian near zero.

We empirically verify several hypotheses regarding learning with large-size networks:

- For large-size networks, most local minima are equivalent and yield similar performance on a test set.
- The probability of finding a "bad" (high value) local minimum is non-zero for small-size networks and decreases quickly with network size.
- Struggling to find the global minimum on the training set (as opposed to one of the many good local ones) is not useful in practice and may lead to overfitting.

Many dimensions introduce major difficulties with finding a global minimum…

…as well as with representative exploration of uncertainties

The Big Data Paradox in vaccine uptake

Article

Unrepresentative big surveys significantly overestimated US vaccine uptake

https://doi.org/10.1038/s41586-021-04198-4

Received: 18 June 2021

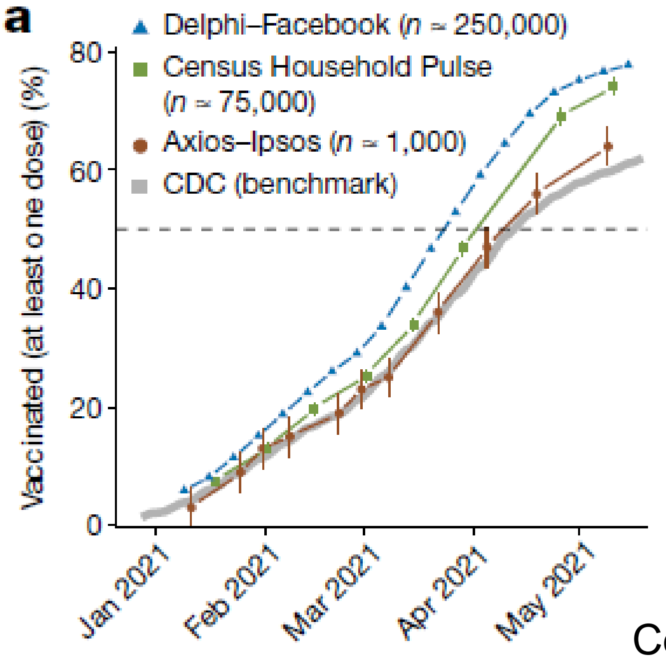
Accepted: 29 October 2021

Published online: 8 December 2021

Check for updates

Valerie C. Bradley[1,6], Shiro Kuriwaki[2,6], Michael Isakov[3], Dino Sejdinovic[1], Xiao-Li Meng[4] & Seth Flaxman[5]

Surveys are a crucial tool for understanding public opinion and behaviour, and their accuracy depends on maintaining statistical representativeness of their target populations by minimizing biases from all sources. Increasing data size shrinks confidence intervals but magnifies the effect of survey bias: an instance of the Big Data Paradox[1]. Here we demonstrate this paradox in estimates of first-dose COVID-19 vaccine uptake in US adults from 9 January to 19 May 2021 from two large surveys: Delphi–Facebook[2,3] (about 250,000 responses per week) and Census Household Pulse[4] (about 75,000 every two weeks). In May 2021, Delphi–Facebook overestimated uptake by 17 percentage points (14–20 percentage points with 5% benchmark imprecision) and Census Household Pulse by 14 (11–17 percentage points with 5% benchmark imprecision), compared to a retroactively updated benchmark the Centers for Disease Control and Prevention published on 26 May 2021. Moreover, their large sample sizes led to miniscule margins of error on the incorrect estimates. By contrast, an Axios–Ipsos online panel[5] with about 1,000 responses per week following survey research best practices[6] provided reliable estimates and uncertainty quantification. We decompose observed error using a recent analytic framework[1] to explain the inaccuracy in the three surveys. We then analyse the implications for vaccine hesitancy and willingness. We show how a survey of 250,000 respondents can produce an estimate of the population mean that is no more accurate than an estimate from a simple random sample of size 10. Our central message is that data quality matters more than data quantity, and that compensating the former with the latter is a mathematically provable losing proposition.



Nature v. 600 (2021) 695

Courtoy et al., PRD 107 (2023) 034008

Complexity and PDF tolerance

- **Bad news:** The tolerance puzzle is *intractable* in very complex fits
  - In a fit with N_{par} free parameters, the minimal number of PDF replicas to estimate the expectation values for ∀ χ^2 function grows as $N_{min} \geq 2^{N_{par}}$
  - Example: $N_{min} > 10^{30}$ for $N_{par} = 100$

[Sloan, Wo´zniakowski, 1997]

[Hickernell, MCQMC 2016, 1702.01487]

Good news: expectation values **for typical QCD observables** can be estimated with fewer replicas by reducing dimensionality of the problem or a targeted sampling technique.

Example: a "**hopscotch scan**", see 2205.10444