Credibility and uncertainty in PDF analyses

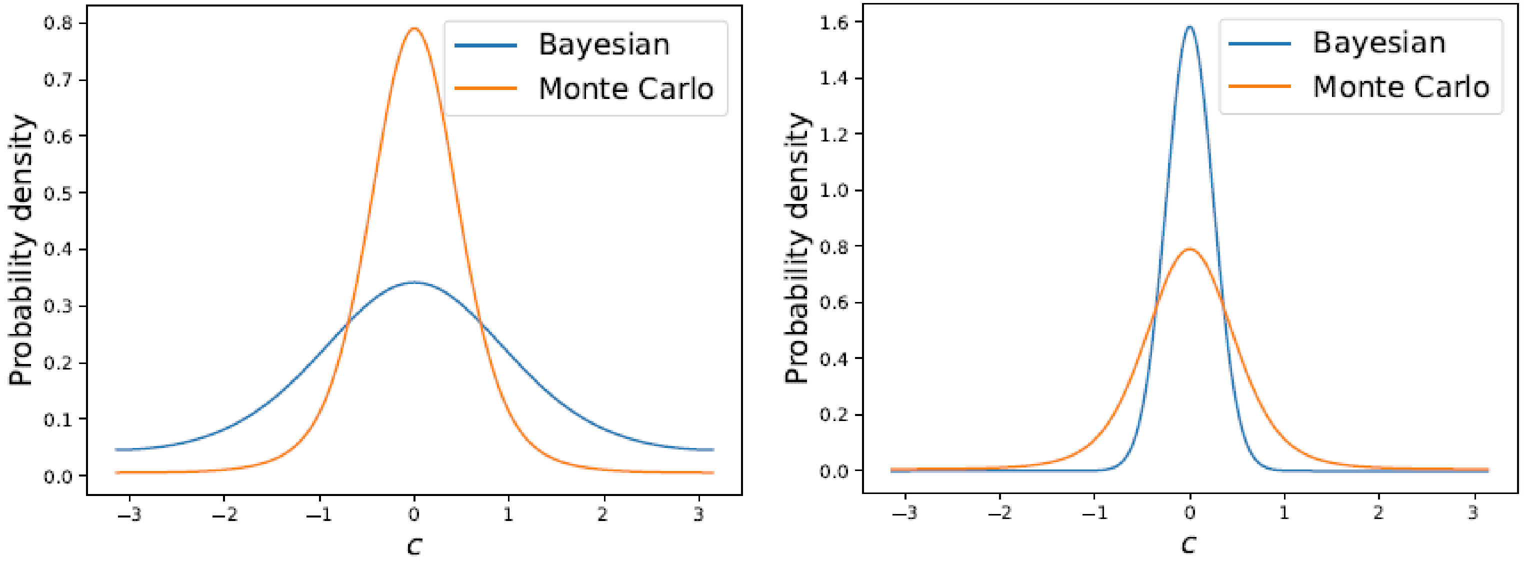
Pavel Nadolsky

Southern Methodist University

& Michigan State University

With inputs from A. Courtoy, T. Hobbs, A. Kronfeld, C.-P. Yuan, Y. Zhao, and

CTEQ-TEA global analysis group

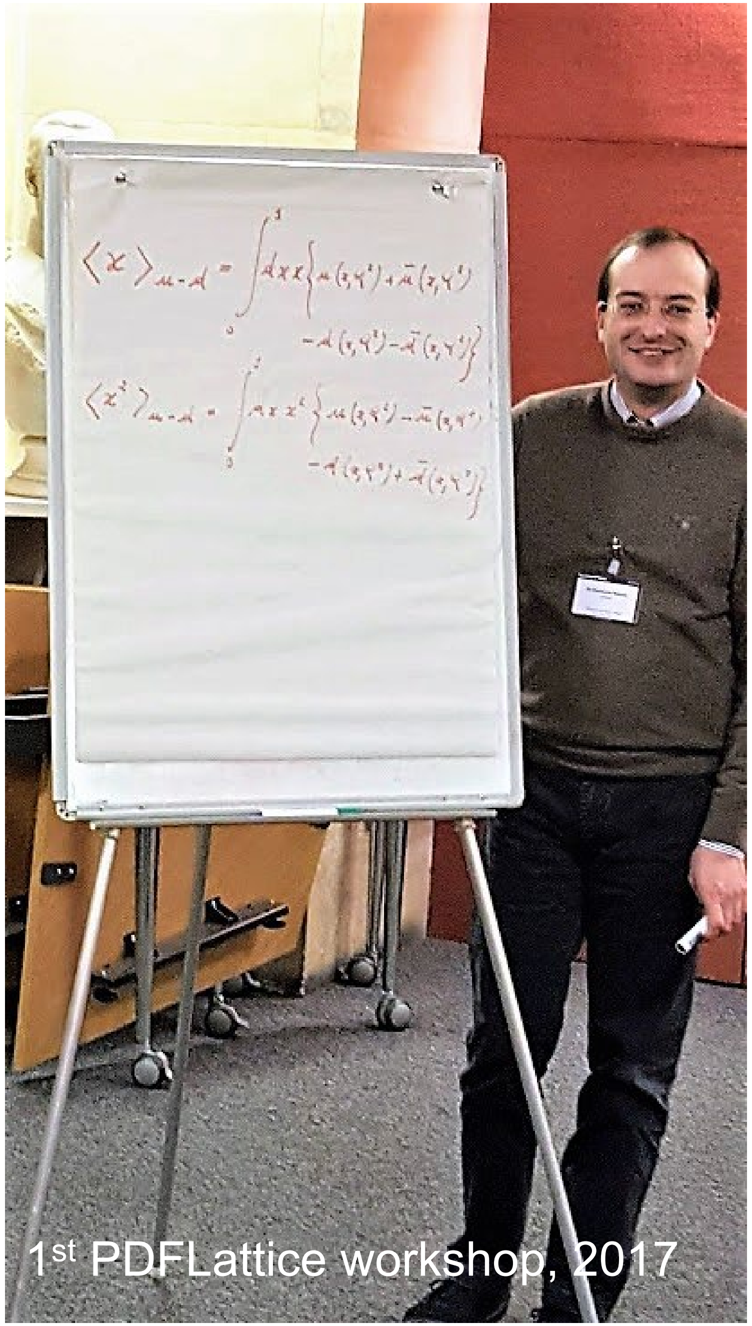


M. Costantini et al., 2404.10056

What is this workshop about?

The focus of this workshop is on Uncertainty Quantification
A **list of Key Questions** is attached to the Indico page of the workshop

1. Accessing PDFs: global analyses and lattice computations
   $\longrightarrow$ How does PDF determination work in global analyses and lattice QCD?
2. Global QCD analyses: inverse problem and objective functions
   $\longrightarrow$ How is the inverse problem entailed by PDF determination addresed?
3. Lattice QCD: considerations on the validity of the perturbative matching
   $\longrightarrow$ How is the equivalence between zP_z and $\xi^- P^+$ defined?
4. Setting up a common language: definitions and benchmarks
   $\longrightarrow$ How to benchmark lattice moments and quasi-/pseudo-PDFs with global analyses?
5. Combining lattice and experimental data to determine PDFs
   $\longrightarrow$ What are the efforts/limitations to incorporate lattice data in PDF determinations?
6. Uncertainty quantification and bias/variance trade-off
   $\longrightarrow$ How are aleatoric and epistemic uncertainties combined? How is a model chosen?


1st PDFLattice workshop, 2017

New insights about unpolarized parton distribution functions

Nonperturbative models and lattice QCD

CONNECTION?

(N)NNLO global analyses of QCD data

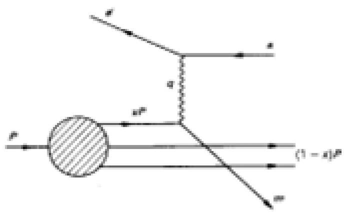
Global QCD analysis

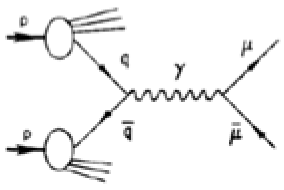
Precision tests at LHC, Jlab, EIC, AMBER, CERN FPF, …

PDFs in nonperturbative QCD

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

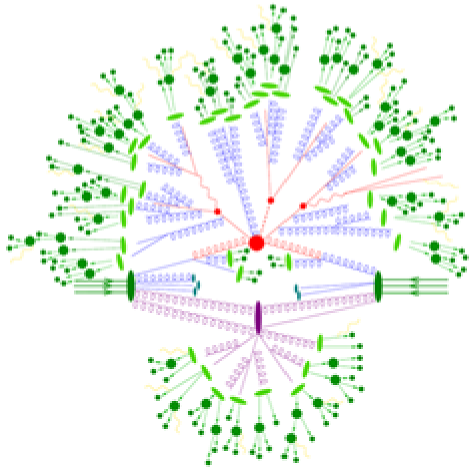
⇨ we can learn about nonperturbative dynamics by comparing predictions to data for the simplest scattering processes (DIS and DY)





Phenomenological PDFs

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



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

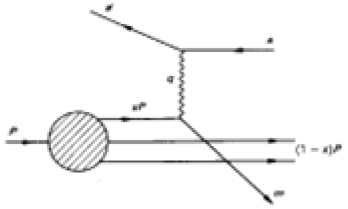
How to relate the x dependence of the perturbative and nonperturbative pictures?

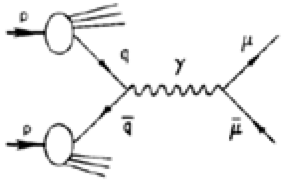
Does the evidence from primordial dynamics survive PQCD radiation?

PDFs in nonperturbative QCD

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

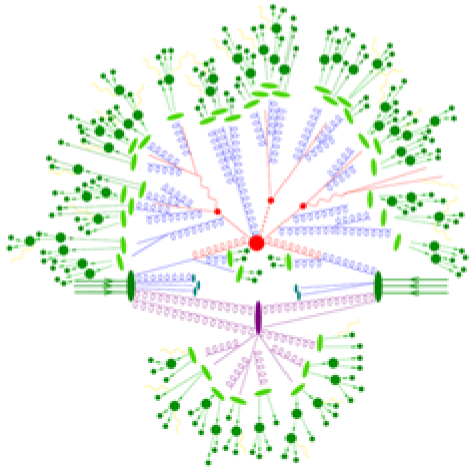
⇨ we can learn about nonperturbative dynamics by comparing predictions to data for the simplest scattering processes (DIS and DY)





Phenomenological PDFs

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



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

Robust uncertainty quantification is crucial for relating these pictures.
It involves foundational issues both in physics and information theory.

Global fits of proton scattering data at (N)NNLO accuracy

A profound **inverse problem** with many parameters and a wide range of implications

Multiloop QCD and EW computations

Exploration of most complex experimental data sets

Accurate and fast high-performance computing

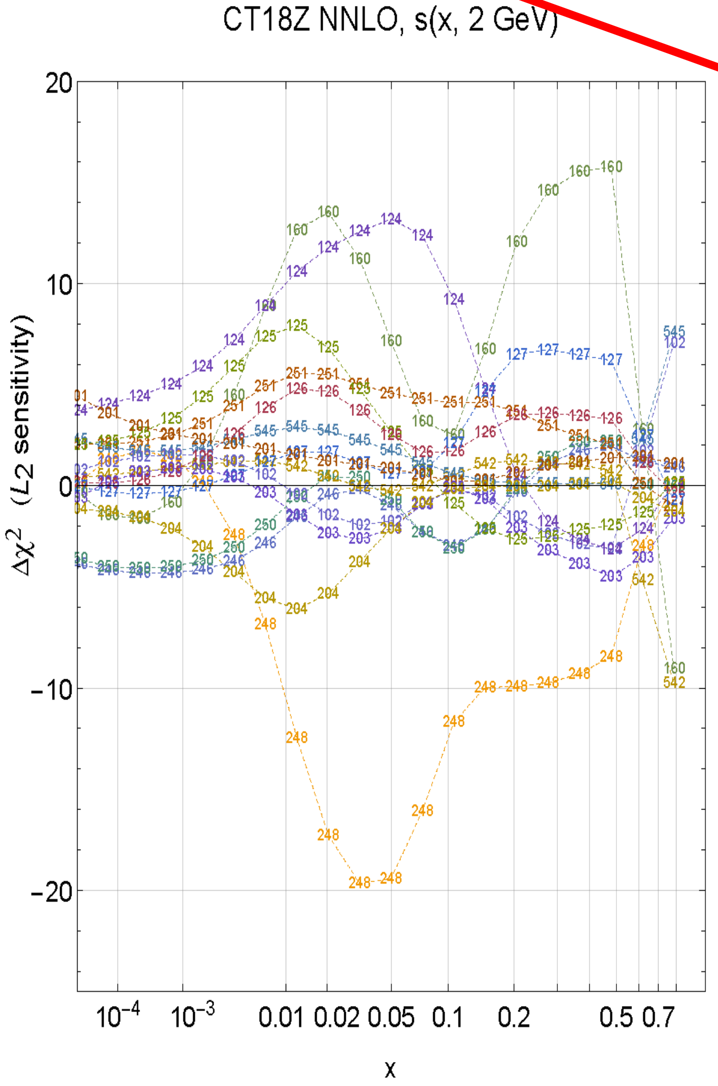
A testing bed for multidimensional uncertainty quantification, ML/AI, …

Which strangeness PDF is preferred by lattice QCD?

Updates?

Unpolarized strangeness $s(x,Q)$ is the least known in global fits; apparent contradictions in preferred $s(x,Q)$ values from various experiments and the lattice

Moment	Lattice QCD	Global Fit (PDF4LHC)	Global fit (uw avg)
$\langle x \rangle_{s^+}$	$0.092(41)^\dagger$	0.036(6)	0.037(3)

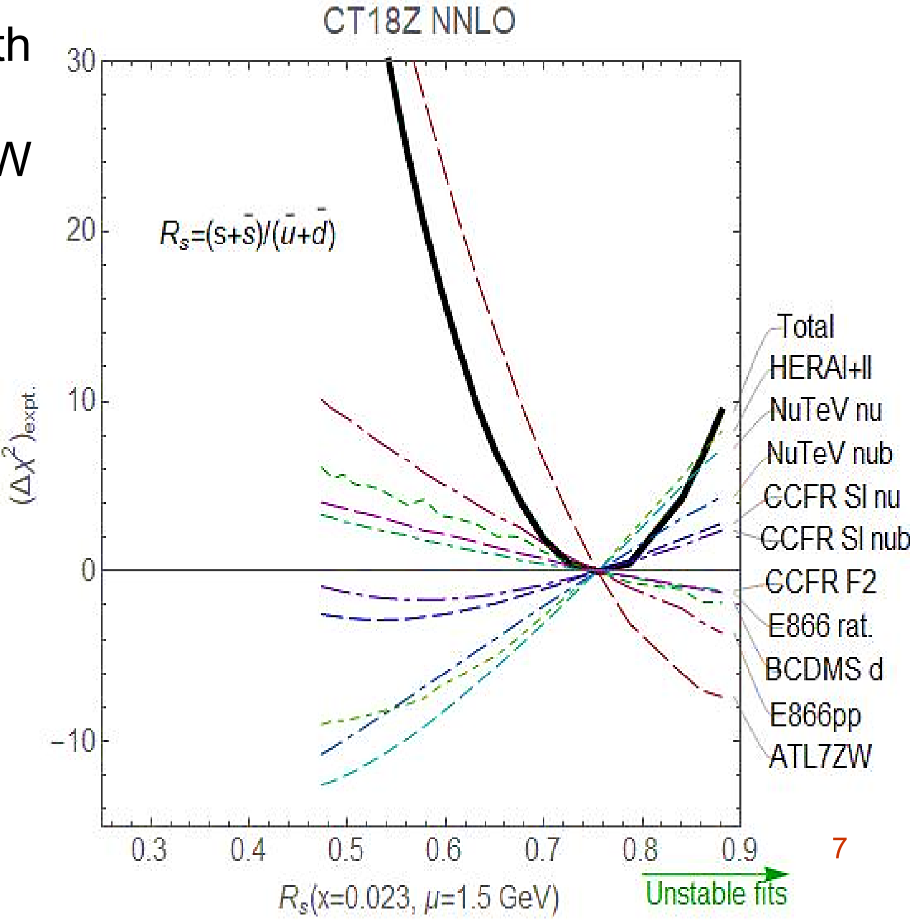


- $R_s \approx 0.5$ at $x \sim 0.02$ preferred both dimuon SIDIS and inclusive DIS
- $R_s > 0.7$ preferred by ATLAS 7 Z/W production and some LHC experiments

Unstable fits for $R_s > 0.9$

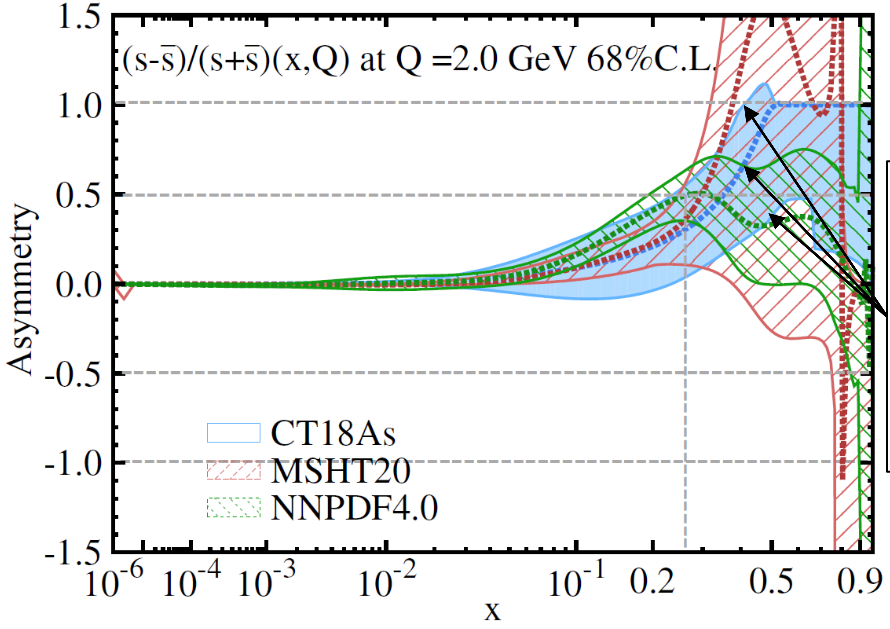
Most sensitive experiments

246 LHCb8Zeer	124 NuTvNuChXN
248 ATL7ZW.xF	125 NuTvNbChXN
250 LHCb8WZ	126 CcfrNuChXN
251 ATL8DY	127 CcfrNbChXN
542 CMS7jtR7y6T	201 e605
545 CMS8jtR7T	203 e866f
160 HERAIpII	204 e866ppxf
102 BcdF2dCor	



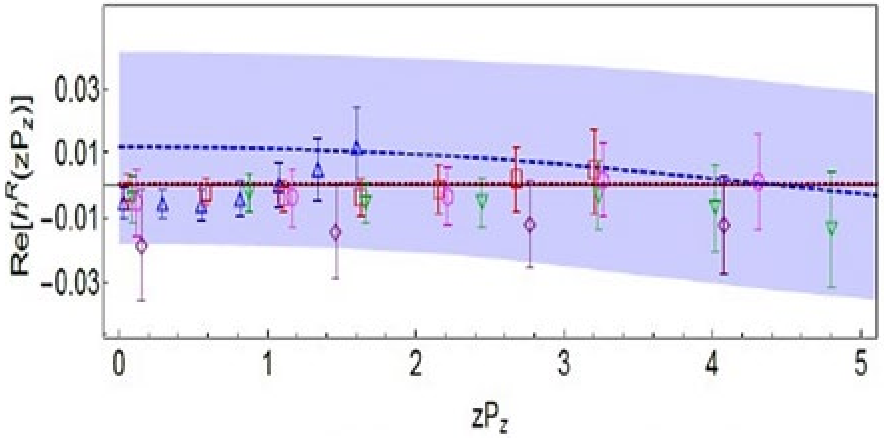
CT18As NNLO: Strangeness asymmetry with a lattice QCD constraint

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



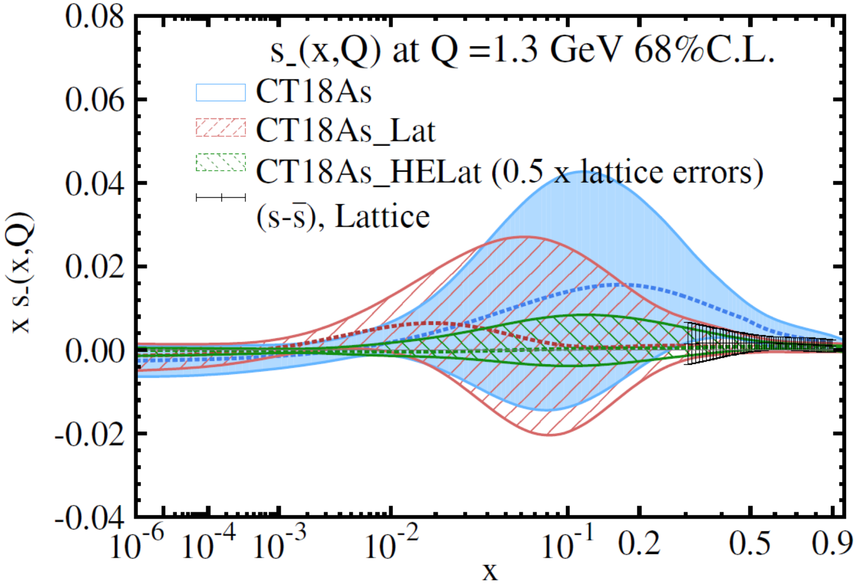
differences reflect the pulls of LHC and other experiments

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?**

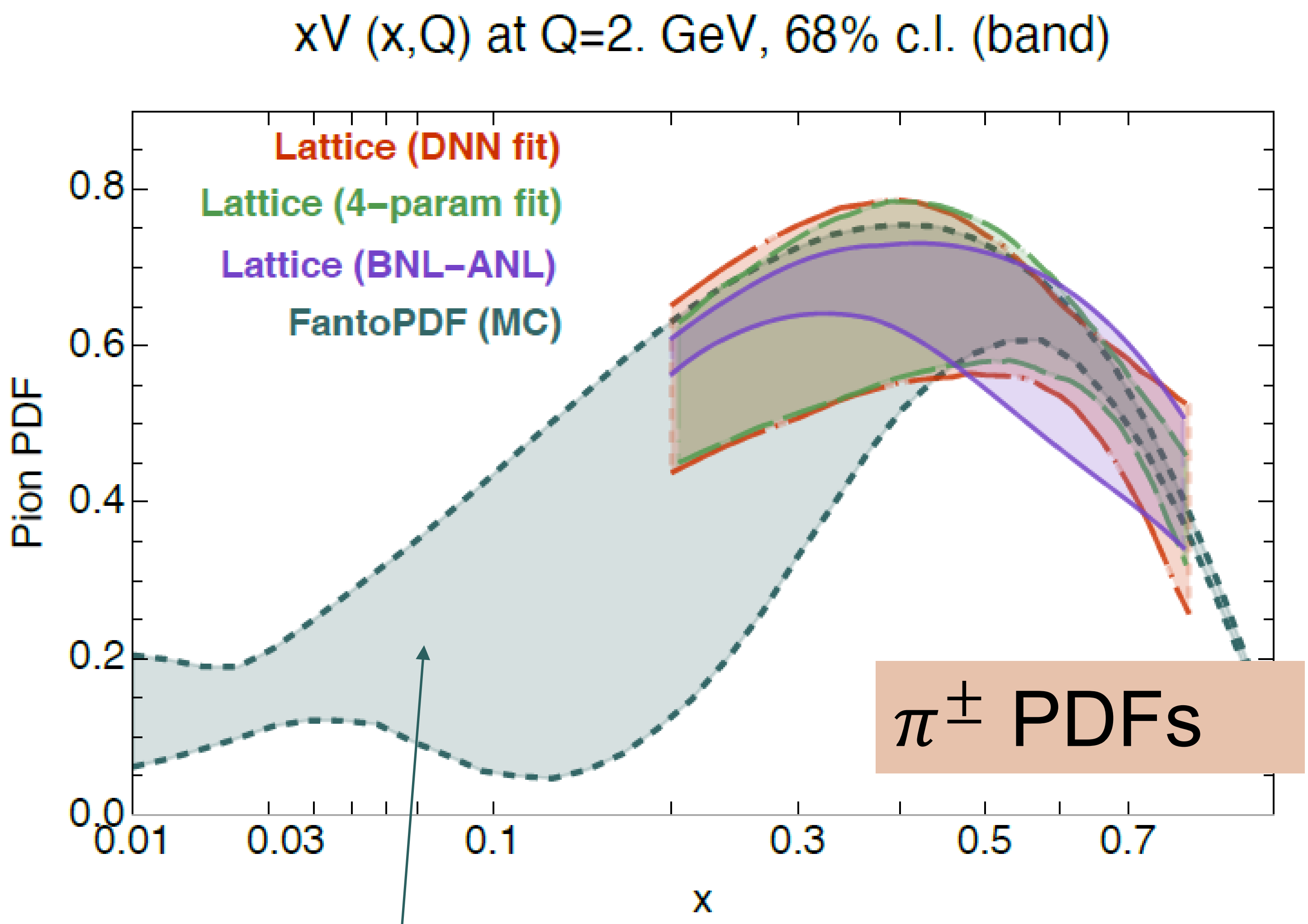


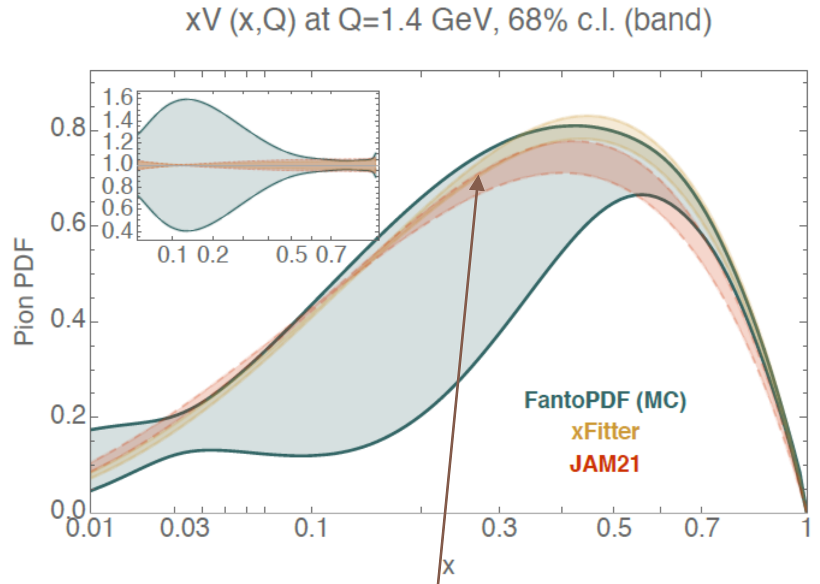
Include lattice data on s_- obtained by the MSULat/quasi-PDF method (2005.01124, Zhang, Lin, Yoon)

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



Lattice QCD already predicts some features of PDFs from first principles

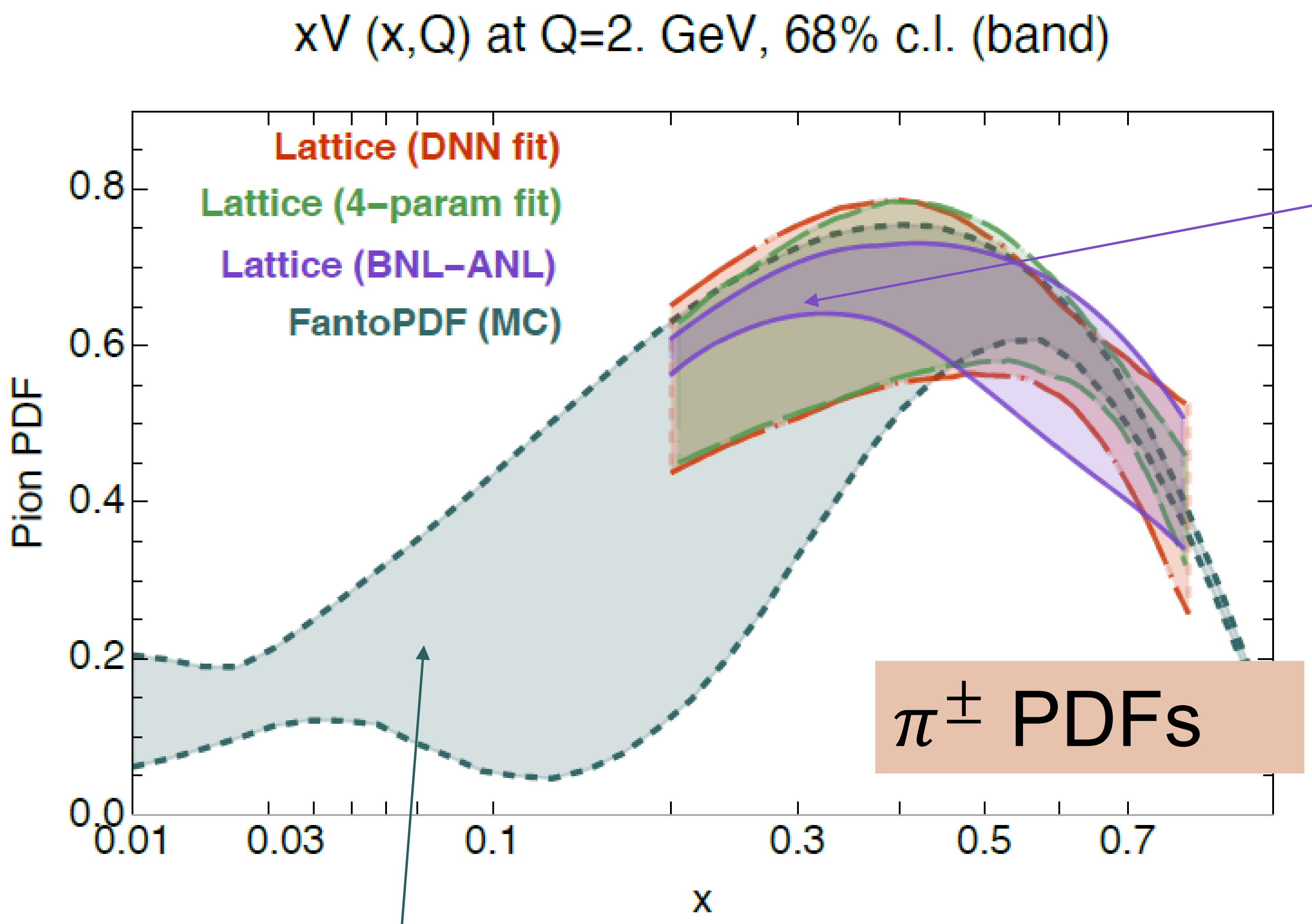




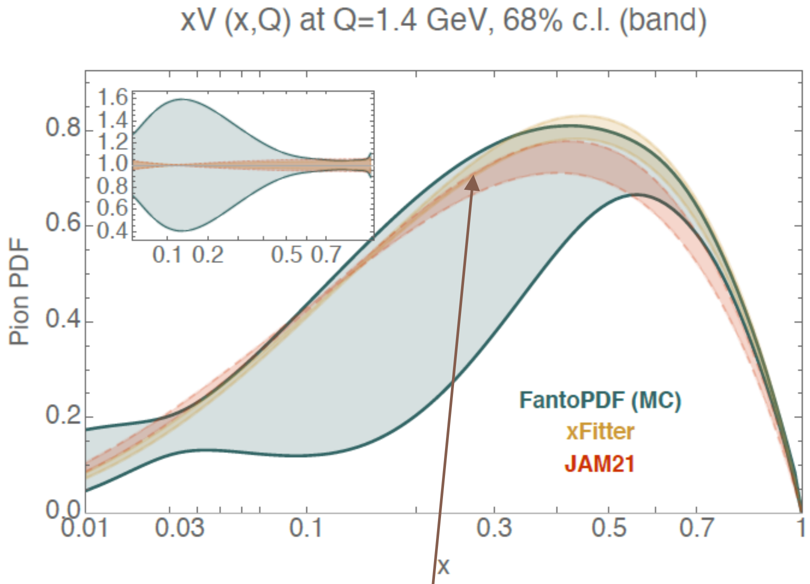
Phenomenological analysis, including the parametrization dependence
L. Kotz, A. Courtoy, M. Chavez, P. N., F. Olness, arXiv:2311.08447

without parametrization dependence

Lattice QCD already predicts some features of PDFs from first principles



are the lattice uncertainties fully estimated?



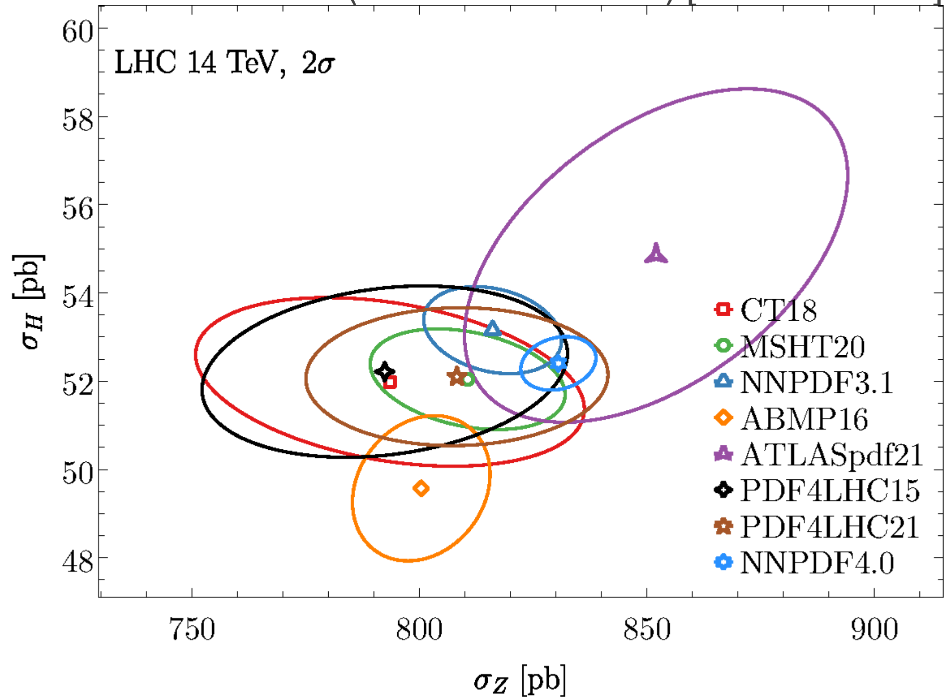
Phenomenological analysis, including the parametrization dependence
L. Kotz, A. Courtoy, M. Chavez, P. N., F. Olness, arXiv:2311.08447

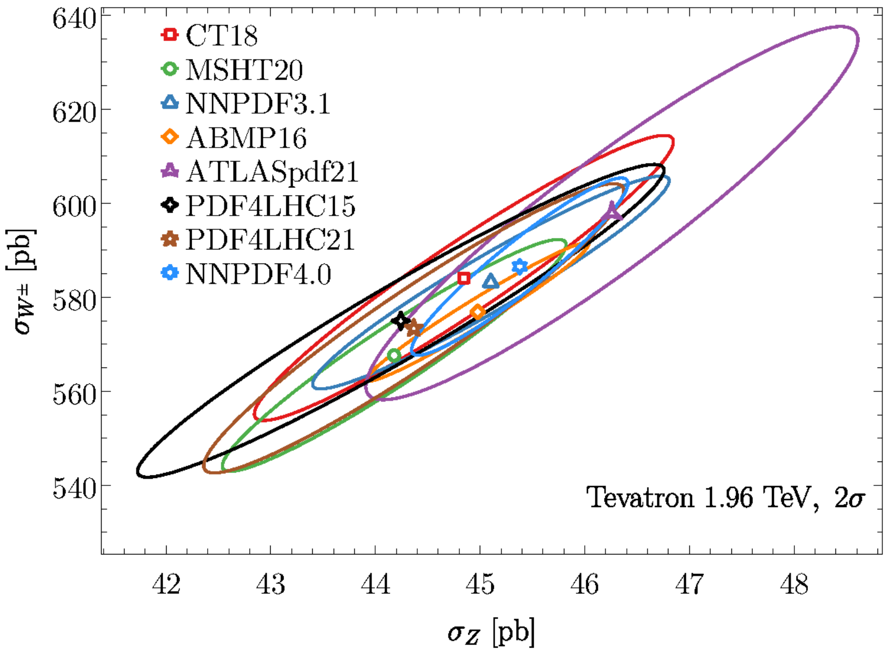
without parametrization dependence

The tolerance puzzle

Why do groups fitting similar data sets obtain different PDF uncertainties?

Precision PDFs (Snowmass 21 WP) [2203.13923v2]





The answer has direct implications for high-stake experiments such as 3D femtography, W boson mass measurement, tests of nonperturbative QCD models and lattice QCD, high-mass BSM searches, etc.

PDF uncertainty: pheno classification

1. Experimental uncertainties, e.g., statistical, correlated and uncorrelated systematic uncertainties of each experimental data set;

2. Theoretical uncertainties due to the absent radiative contributions, approximations in parton showering simulations

3. Parameterization uncertainties associated with the choice of the PDF functional form or AI/ML replica training algorithm
   – contribute at least a half of the CT18 total PDF uncertainty

4. Methodological uncertainties associated with the selection of experimental data sets, fitting procedures, and goodness-of-fit criteria.

associated with the **epistemic** uncertainty; explain several differences among the PDF fits

Kovarik et al., arXiv: 1905.06957

PDF uncertainty: lattice classification

1. LATTICE-specific uncertainties: … TO BE FILLED IN
2. + many PHENO uncertainties from the previous slides

To do:

1. Identify the full error budget for lattice PDF calculations
2. Designate a few calculations (1-2 Mellin moments? pion PDFs? …) as the first targets for FLAG-like validation
3. Do a UQ benchmarking study for these calculations ⇒ A. Courtoy
4. Understand model averaging for PDFs ⇒ E. Neil

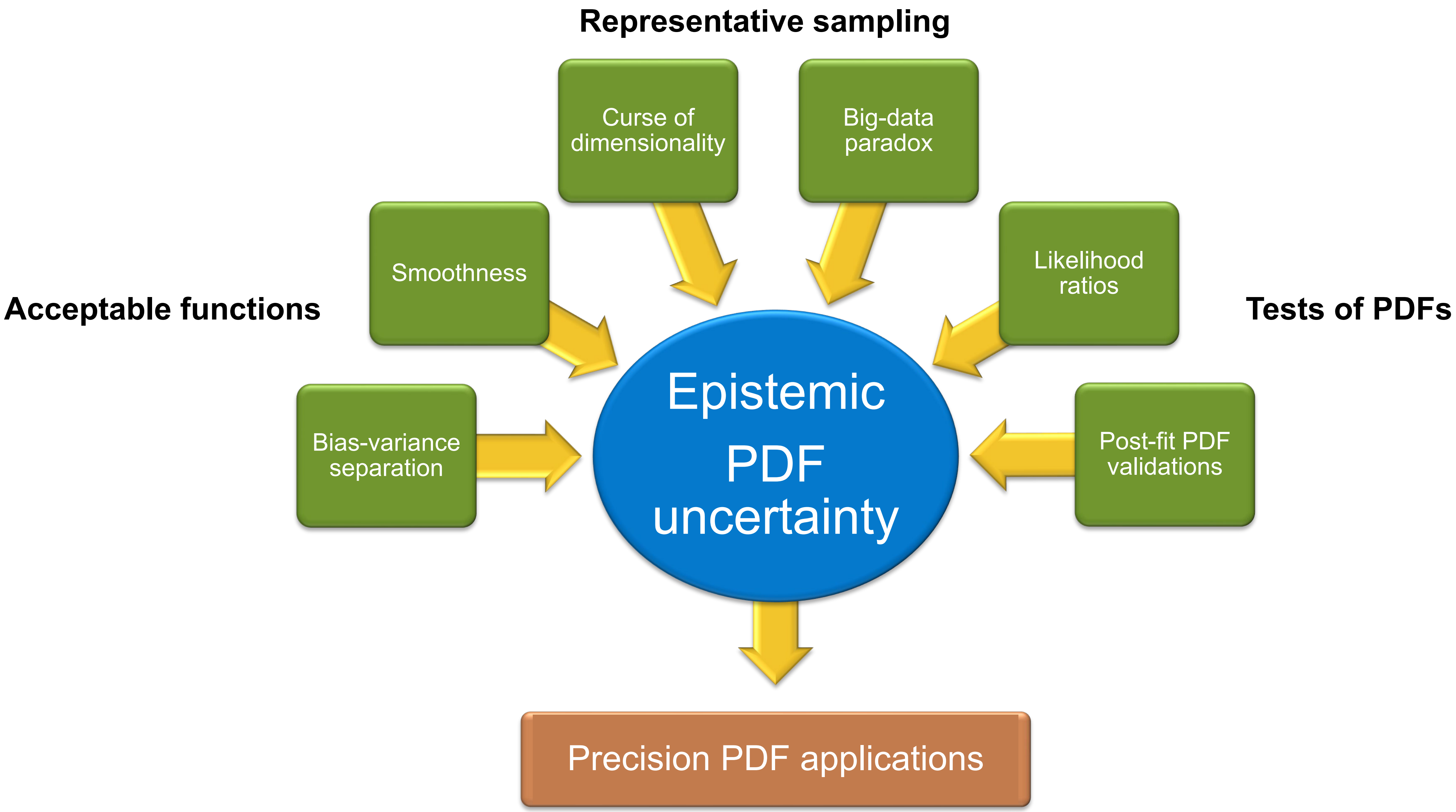
PDF uncertainty: information theory classification

Malinin, Gales, 2018
PN, Courtoy, et al, 2022-24
Hobbs, Kriesten, Gomprecht, 2023-24

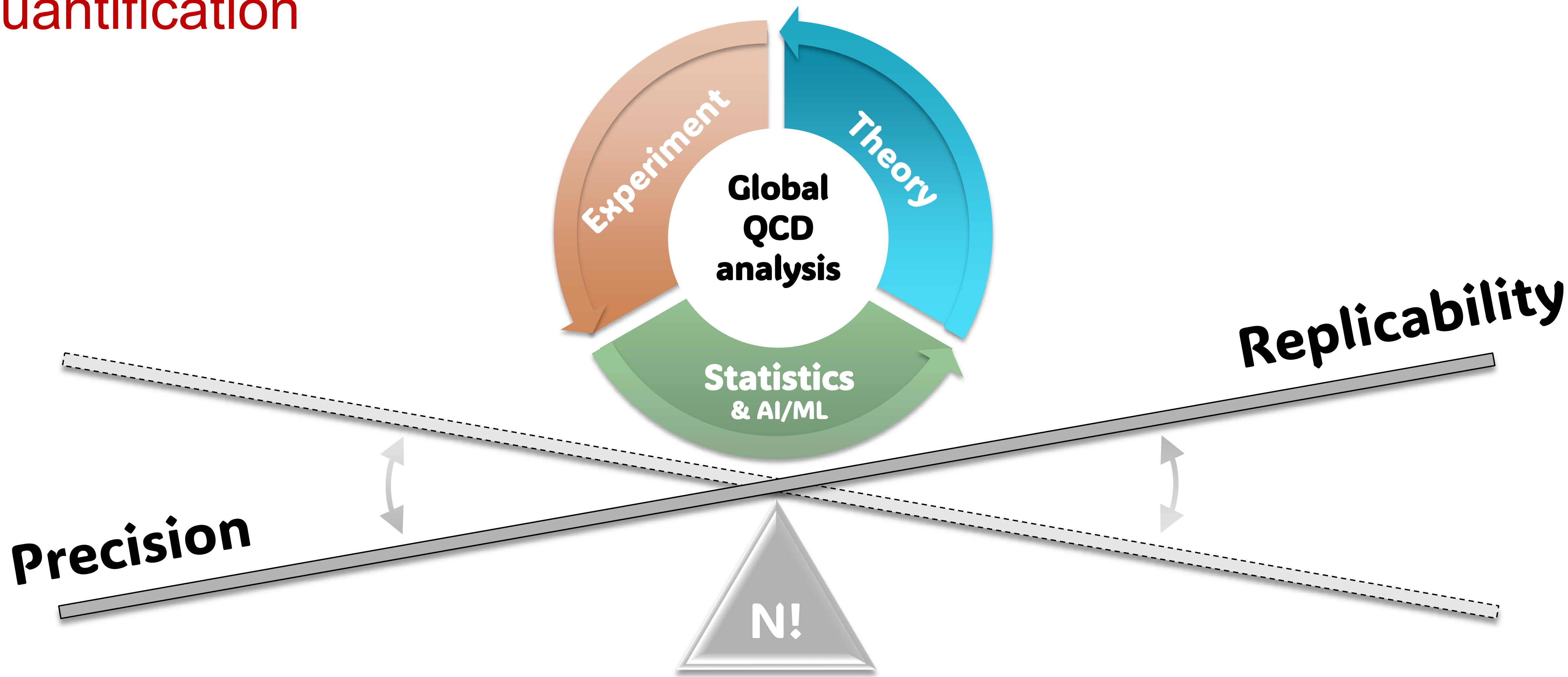
Aleatory (dicey) uncertainty: statistical, propagated from experiments, reduced by increasing data size

Epistemic uncertainty: due to lack of knowledge, bias

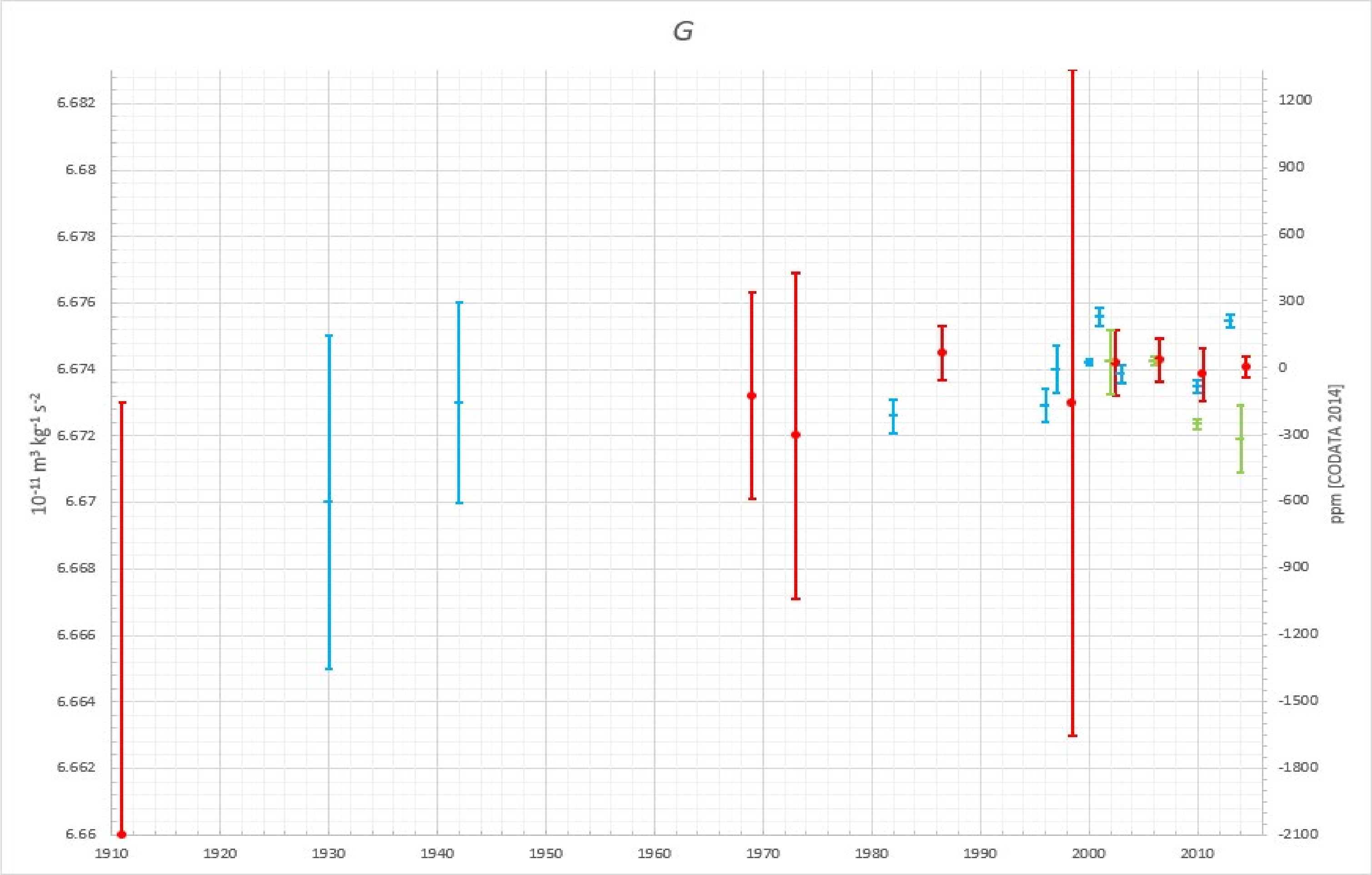
- model uncertainty: reduced by improving the model
- distributional uncertainty: reduced by representative sampling


Representative sampling
Curse of dimensionality
Big-data paradox
Acceptable functions
Smoothness
Likelihood ratios
Tests of PDFs
Bias-variance separation
Epistemic PDF uncertainty
Post-fit PDF validations
Precision PDF applications

Balancing precision and replicability in PDF uncertainty quantification



A life cycle of a precision measurement

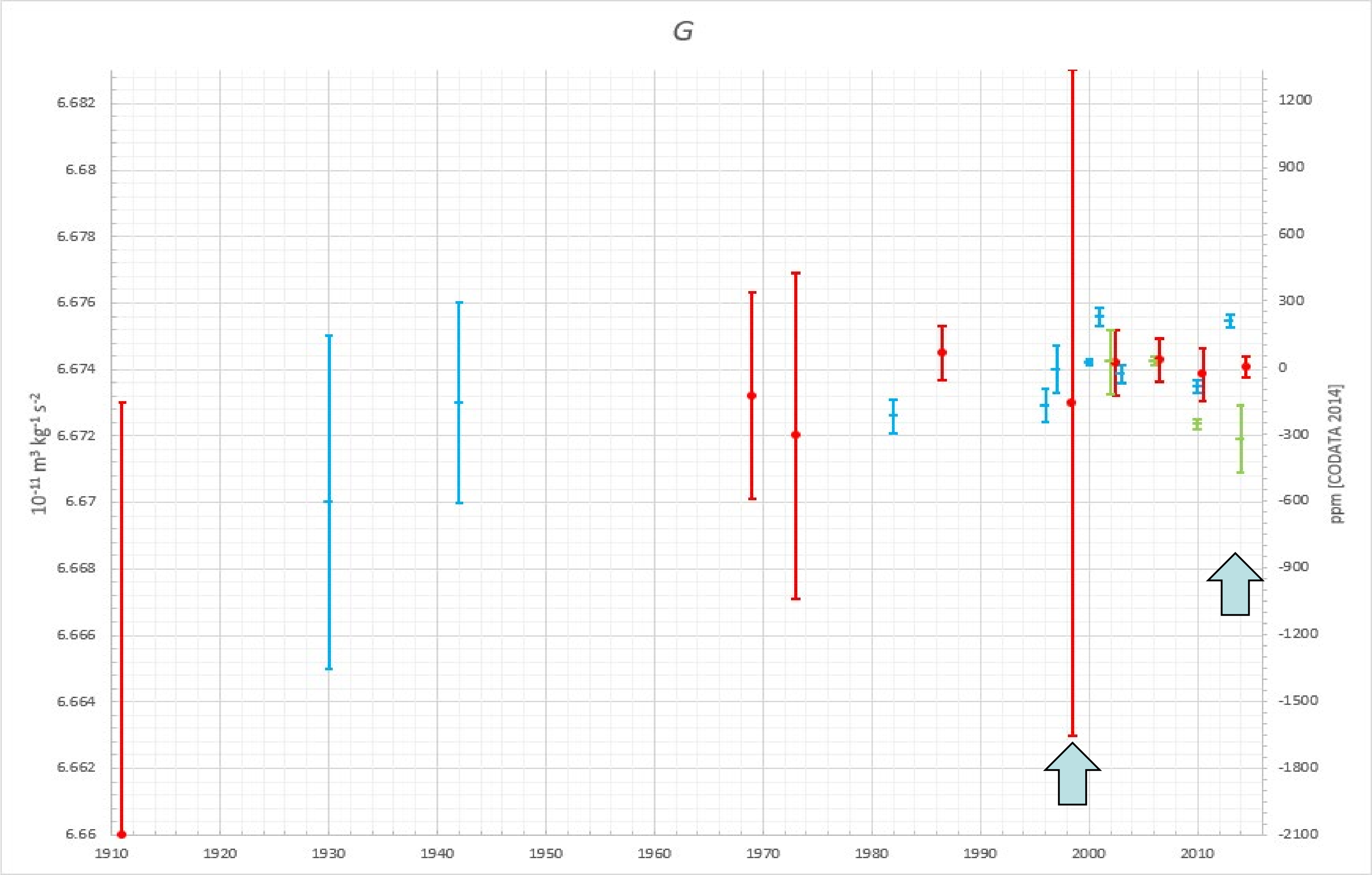


Beginning → Entropy

Example: measurements of the gravitational constant

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

A life cycle of a precision measurement



Entropy

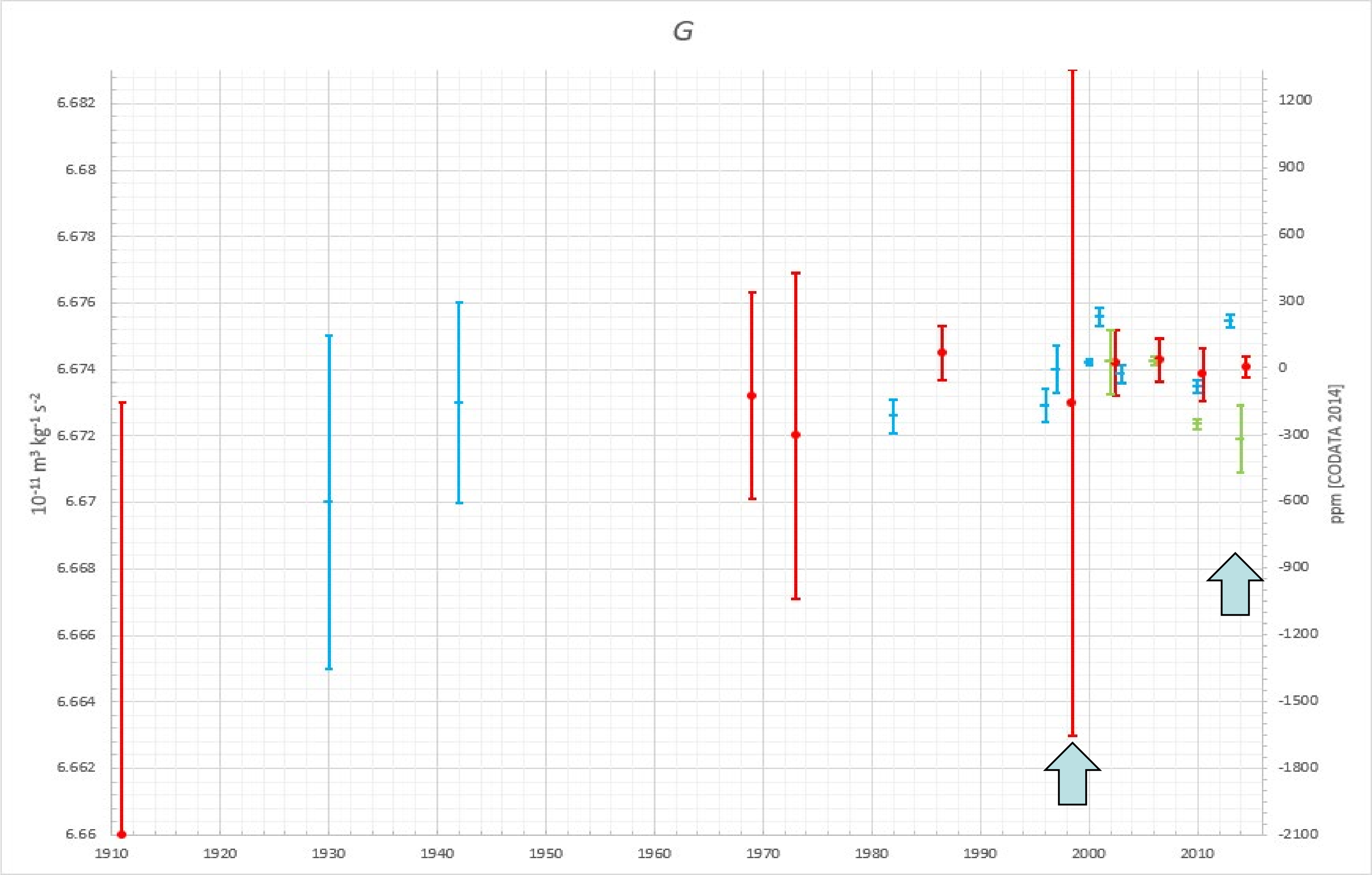
Timeline of measurements and recommended values for *G* since 1900: values recommended based on the NIST combination (red), individual torsion balance experiments (blue), other types of experiments (green).

The combination error bars are unstable after 1995

Some latest precise measurements are in a conflict among themselves and with the post-2014 combination

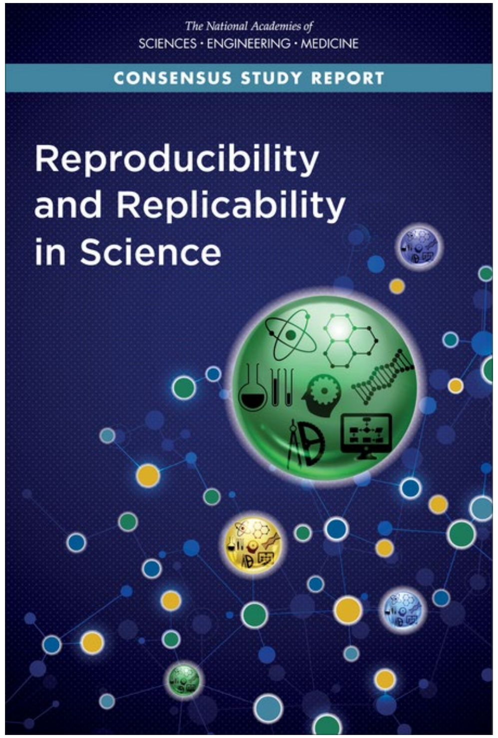
https://en.wikipedia.org/wiki/Gravitational_constant#Modern_value, retrieved on Oct. 22, 2023

A life cycle of a precision measurement



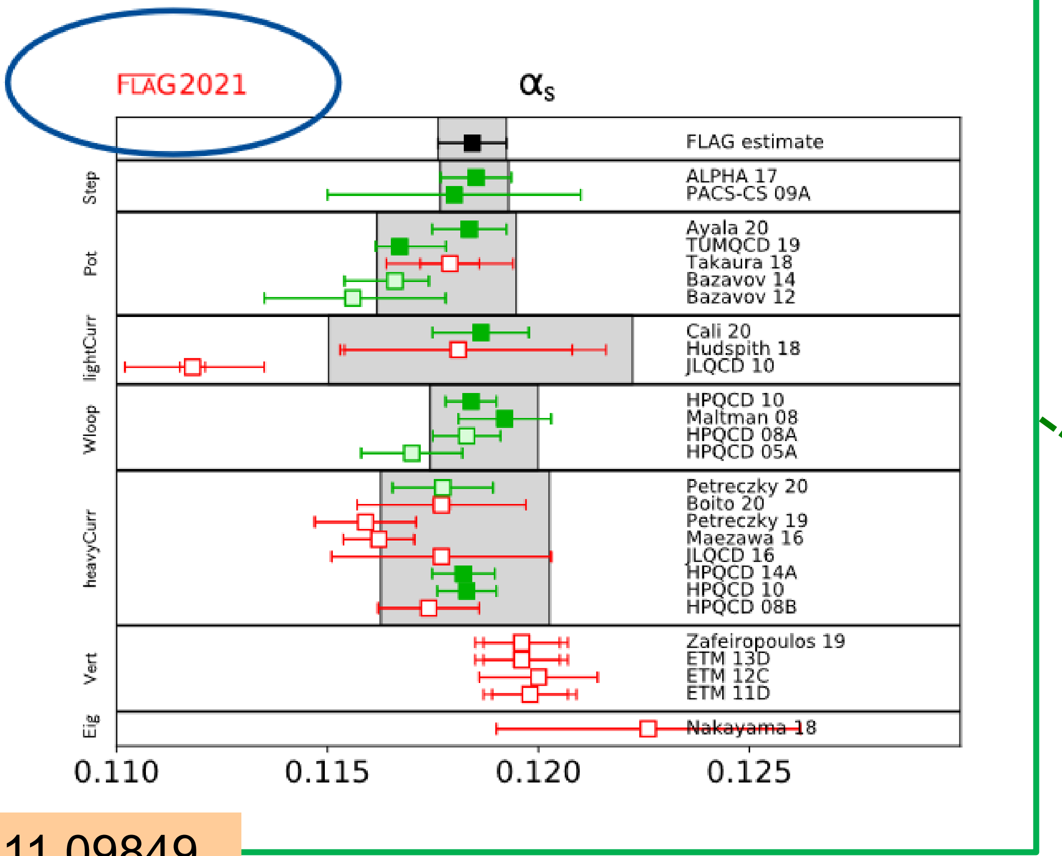
Entropy

The entropy stage can be delayed by adopting the **replicability mindset** for all components of the analysis

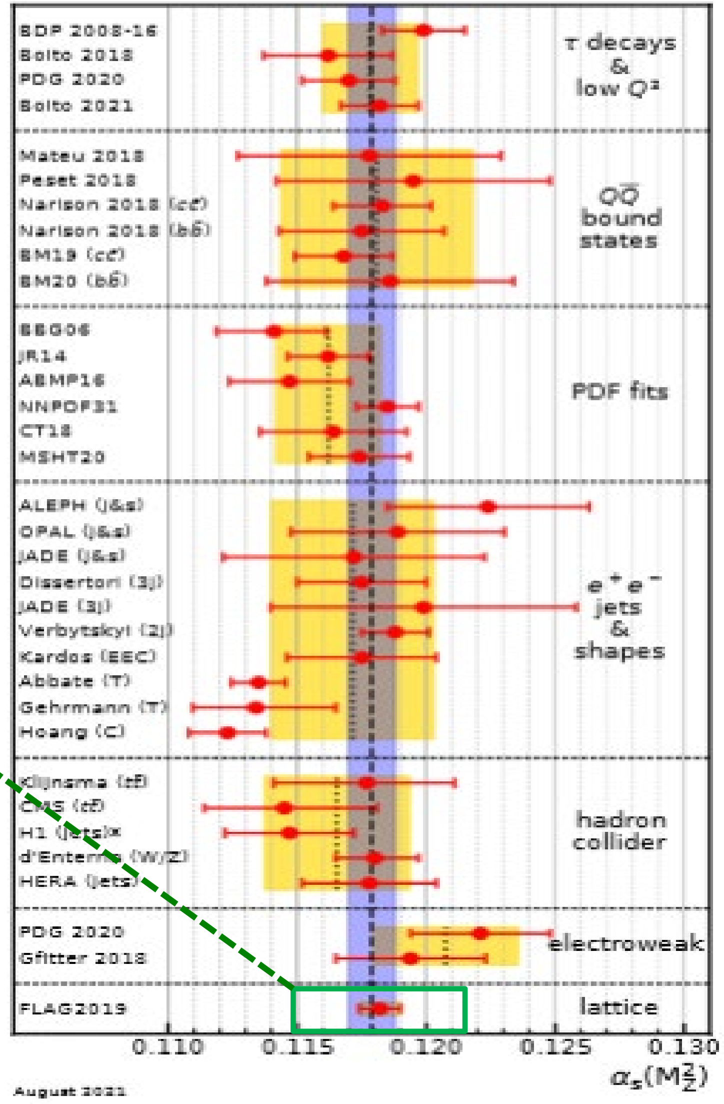


US National Academy of Sciences, Engineering, and Medicine, 2019, https://doi.org/10.17226/25303

Lattice QCD & world-average α_s combination



arXiv:2111.09849



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.

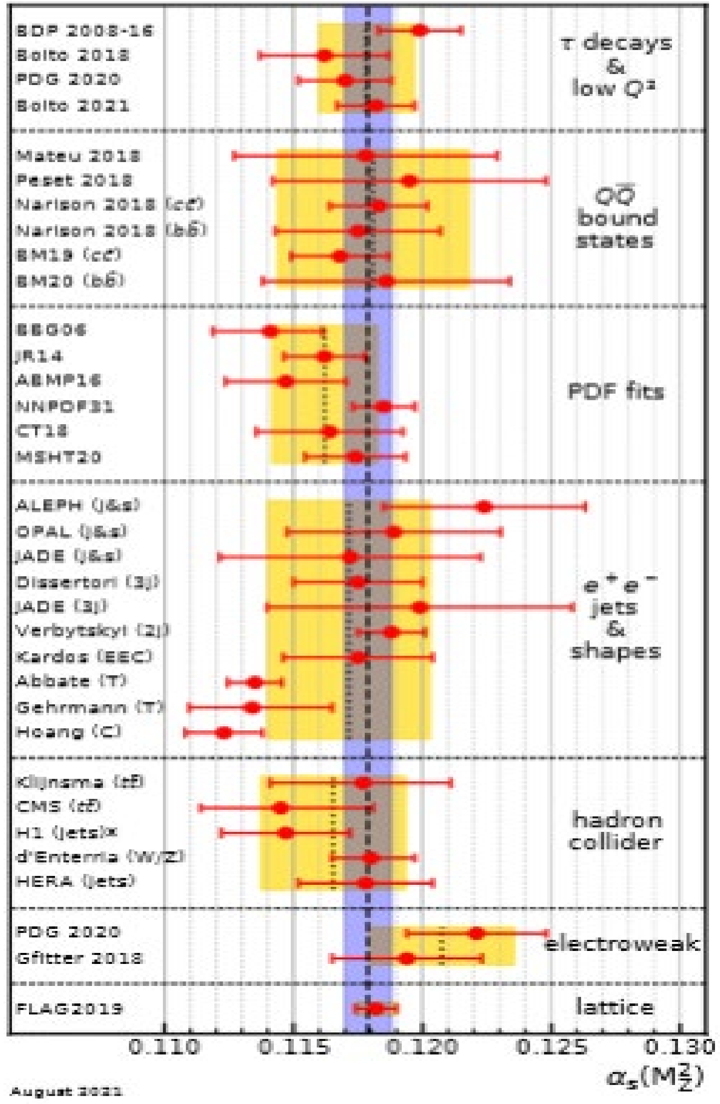
Time to rethink how the world-average α_s combination is performed?

Future measurements of the QCD coupling

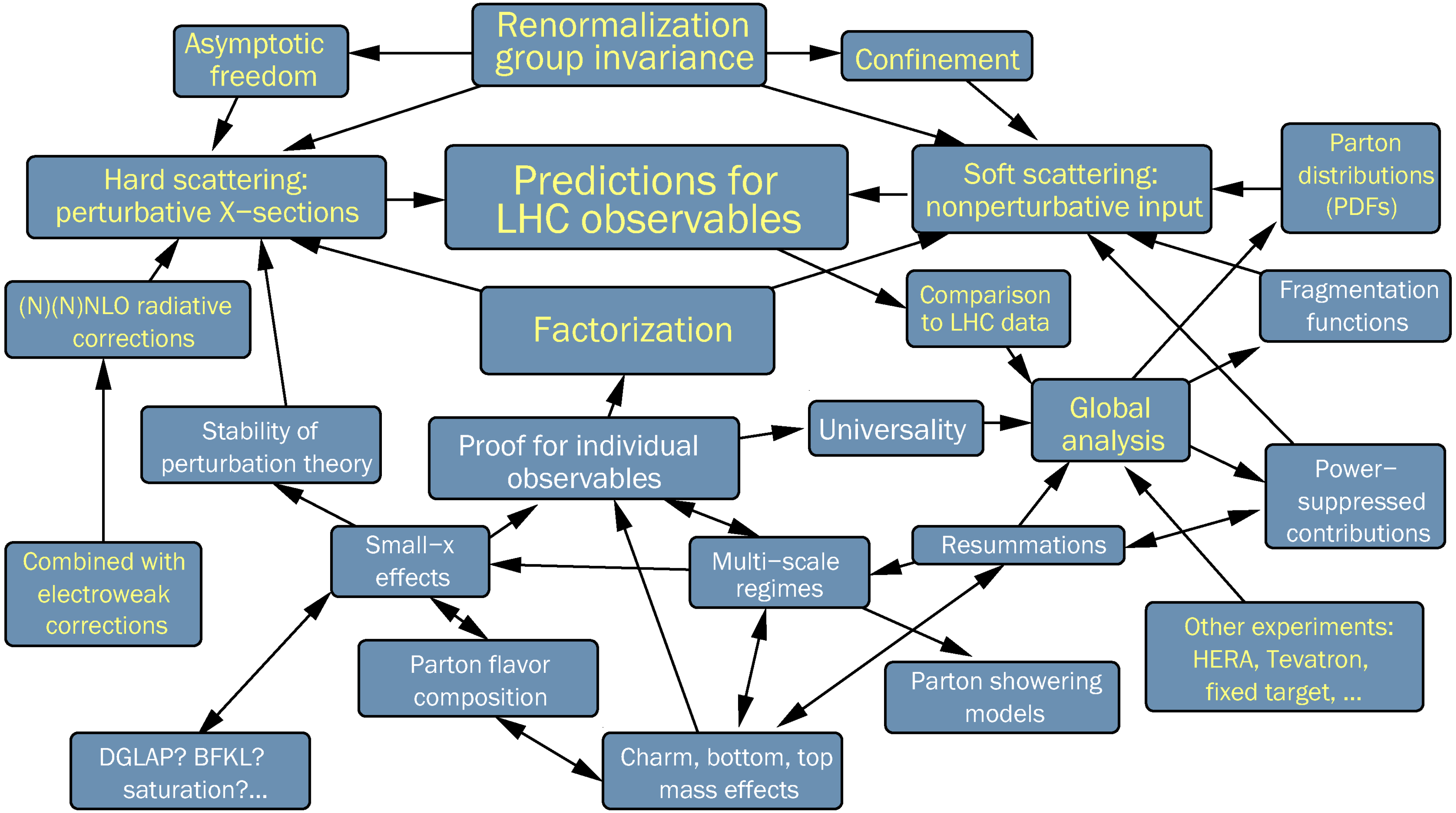
individual α_s measurements can reach precision of $\sim 0.1\%$

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. $N^{2,3}LO$ pQCD truncation	$\approx$ 0.3% (0.1%) Reduced latt. spacing. Add more observables Add $N^{3,4}LO$, 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% $N^{2,3}LO$ pQCD truncation $m_{c,b}$ uncertainties	$\approx$ 1.5% Add $N^{3,4}LO$ & more ($c\bar{c}$), ($b\bar{b}$) bound states Combined $m_{c,b} + \alpha_s$ fits
(4) DIS & PDF fits	1.7% $N^{2,(3)}LO$ 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+$N^{(1,2,3)}$LL truncation Different NP analytical & PS corrs. Limited datasets w/ old detectors	$\approx$ 1.5% ($<$ 1%) Add $N^{2,3}LO+N^3LL$, 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%)

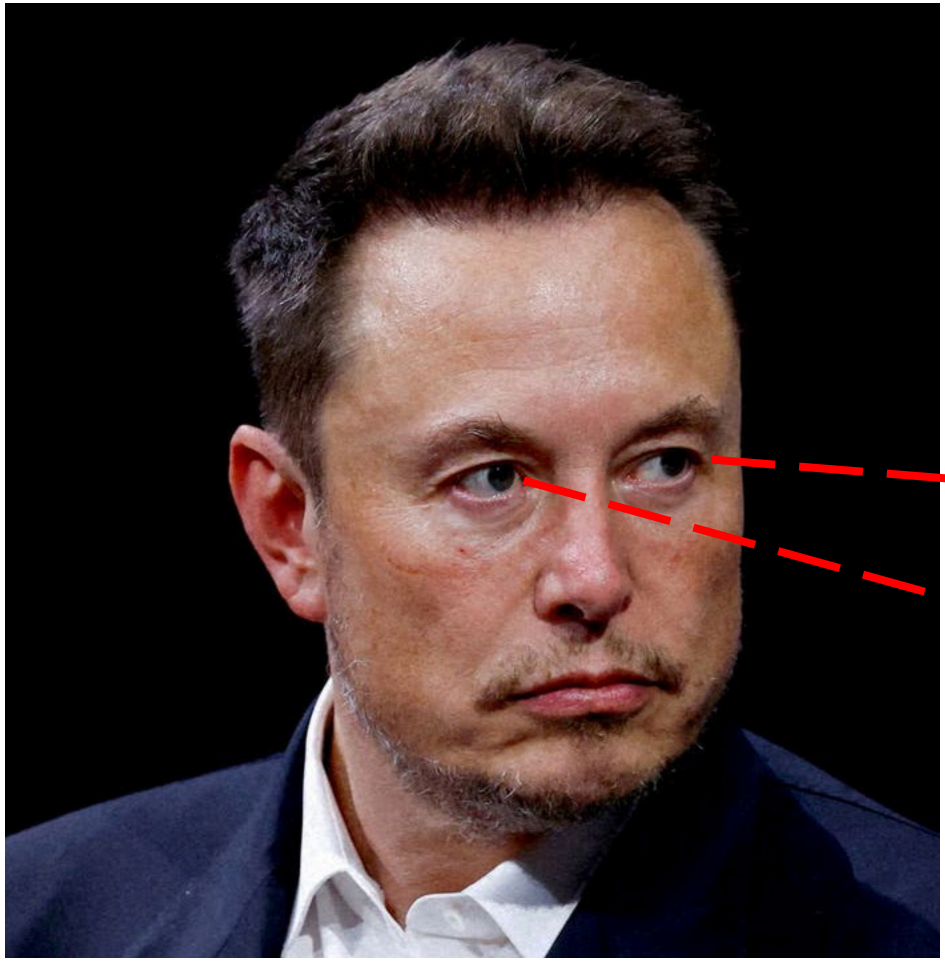


D. d'Enterria et al., EF QCD, arXiv:2203.08271



No analysis is an island *entire of itself*

- Accuracy is determined both by the individual calculation and its ambient connections
- **Aleatory** and **epistemic** uncertainties both play a role



Nonreplicability

is

CONSPICUOUSLY

inefficient

Fitting = learning

Fitting the data is equivalent to learning the probability distribution. In global fits, we also explore statistical foundations of AI/ML. This has an impact on UQ and replicability with AI-based techniques.

"… AI can help verify what we already know by addressing **science's replicability crisis. Around 70% of scientists report having been unable to reproduce another scientist's experiment—a disheartening figure.** As AI lowers the cost and effort of running experiments, it will in some cases be easier to replicate results or conclude that they can't be replicated, contributing to a greater trust in science."

Eric Schmidt, This is how AI will transform the way science gets done,
MIT Technology Review, 2023-07-05

REPORT TO THE PRESIDENT

Supercharging Research: Harnessing Artificial Intelligence to Meet Global Challenges

Executive Office of the President

President's Council of Advisors on Science and Technology

APRIL 2024

Fundamental physics and cosmology are built on statistical analyses of data to test theory, so they require a deep understanding of the probabilities in the interpretation of data. This requirement is driving the mathematical development of AI that can handle probabilistic rigor. … For a measurement of a key number, it would provide a range of possible values that are, say, 68% likely, 95% likely, or 99.9% likely. **Assessing uncertainties is crucial for fundamental physics, and probabilistically rigorous AI would be a game changer for many other fields of science as well, in addition to being invaluable for applications beyond science.**

Sec. 3.4. Revealing the Fundamental Physics of the Universe

Possible to-dos for this workshop

To agree: a common UQ glossary

To review: foundations of multivariate fits

1. Fitting as learning
   a. Meaning of uncertainties: Bayesian, frequentist, Hessian, Monte-Carlo,…
   b. Goodness-of-fit criteria: χ^2 is not the only measure!
   c. Wilks' theorem: the likelihood ratio as the fundamental quantity for hypothesis/parameter testing
   d. Aleatoric, model, distributional uncertainties in an ML-based approach — B. Kriesten
   e. Averaging over model uncertainty — E. Neil
   f. Fitting the likelihood and priors (a Gaussian model mixture) — K. Mohan
2. Dependence on the number of parameters N_{par}
   a. Parsimony: Occam's razor, information criteria, naturalness…
   b. Curse of dimensionality
   c. Big-data paradox in sampling with many N_{par}
3. Fundamental limitations
   a. Dominance of saddle points in non-convex optimization with many N_{par}
   b. Bias-variance ambiguity
   c. Impact on systematic uncertainties
   d. "No free lunch" theorems

…TO BE CONTINUED

Fitting = learning

Fitting the data implies learning the probability distribution $\mathrm{P}(\mathrm{a}\,|\mathrm{D}, \mathrm{T}(\mathrm{a}))$.

The key steps:

1. Assume a probability distribution $P(\{D_k\}|\{\langle D_k\rangle\})$ due to random fluctuations of D_k around $\langle D_k\rangle$. Construct the covariance matrix $\mathrm{cov}_{ij}^{-1} \equiv \langle (D_i - \langle D_i\rangle)\left(D_j - \langle D_j\rangle\right)\rangle$.

2. Minimize $|T_k(a) - \langle D_k\rangle|$. It can be done using several forms of χ^2
   - A closure test: check that the objective function does not bias the probability from step 1 [L. Harland-Lang]; such tests for PDFs are complex and still somewhat limited

3. A hypothesis test: are the deviations $T_k(a) - \langle D_k\rangle$ consistent with random data fluctuations estimated in step 1?
   - Here we work with the χ^2 distribution for $N_{pt} - N_{par}$ degrees of freedom
   - Weak and strong goodness-of-fit criteria

4. A parameter test: what variations of parameters a in $T_k(a)$ do not violate the acceptance of hypothesis in step 3?
   - Here we also work with χ^2. Tolerance $\Delta\chi^2 = T^2 > 1$ to account for hidden errors.

A likelihood-ratio test of models T_1 and T_2

From Bayes theorem, it follows that

Soper, Collins, hep-ph/9411214
Kovarik, Nadolsky, Soper, 1905.06957
Courtoy et al., 2205.10444

$$\underbrace{\frac{P(T_2|D)}{P(T_1|D)}}_{\equiv r_{\text{posterior}}} = \underbrace{\frac{P(D|T_2)}{P(D|T_1)}}_{\equiv r_{\text{likelihood}}} \times \underbrace{\frac{P(T_2)}{P(T_1)}}_{\equiv r_{\text{prior}}}$$

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.

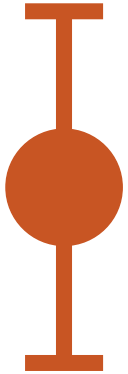
Estimating the epistemic uncertainty is hard because statistics with many parameters is different!

In typical applications, especially AI/ML ones:

1. **As a rule, there is no single global minimum of** χ^2 (or another cost function)
   – "Best fits" are dominated by saddle points with the same low χ^2
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; Courtoy et al.,]:

The bigger the data, the surer we fool ourselves.

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

To agree: the meaning of 

Is the PDF uncertainty…

1. Bayesian (a credibility interval)?

2. Frequentist (a confidence interval)?

3. Both?

4. None?