

Precision and replicability of parton distributions

A proton at a collider

Pavel Nadolsky

Southern Methodist University

CTEQ-TEA global analysis group



CONAHCYT
CONSEJO NACIONAL DE HUMANIDADES
CIENCIAS Y TECNOLOGÍAS

I·AN Network of Networks
Inter-American QCD

Contents

Week 1

1. Dynamic images of hadrons in the AI era
2. Collinear PDFs, their applications and determinations
3. **RRR**: rigor, reproducibility, replicability in the PDF analysis

Week 4

4. Uncertainty quantification on parton distributions
 - Tolerance puzzle
 - Less known aspects of multivariate statistics

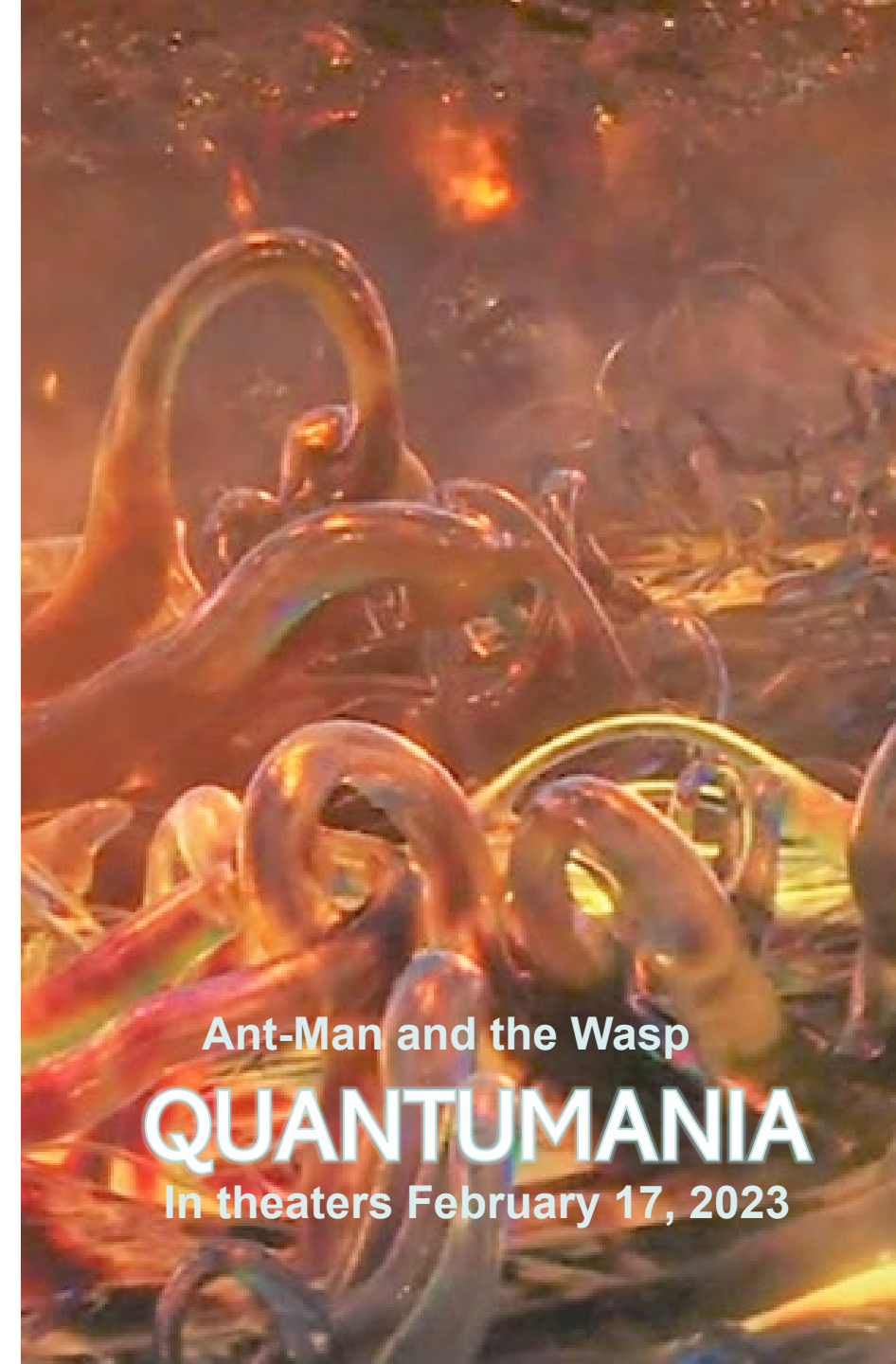
Visualizations of subatomic dynamics

1. Rigorous, mathematical

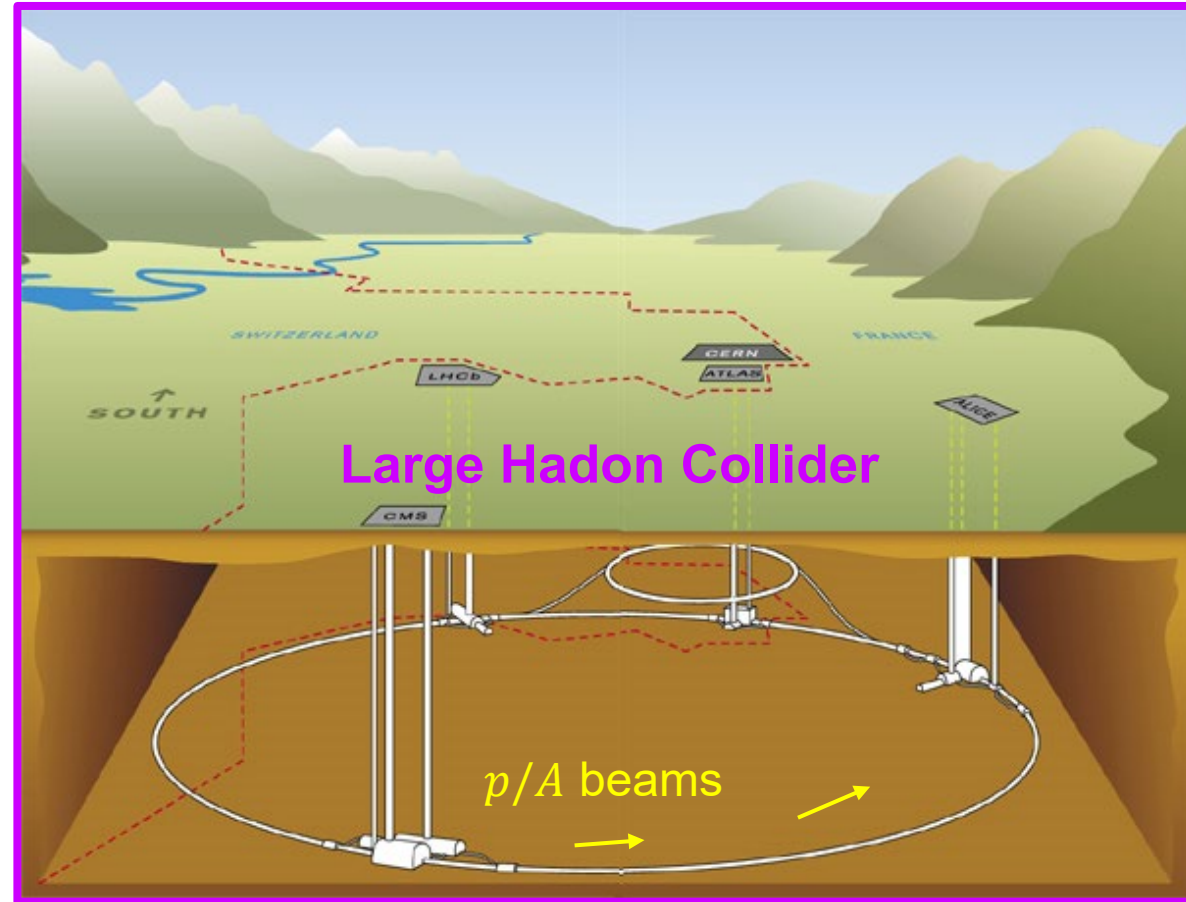
- QCD, perturbative and nonperturbative methods, factorization theorems, Monte-Carlo simulations,...
- Quantitative control of accuracy
- Mostly in energy-momentum (E, \vec{p}) coordinates

2. Simplified, conceptual

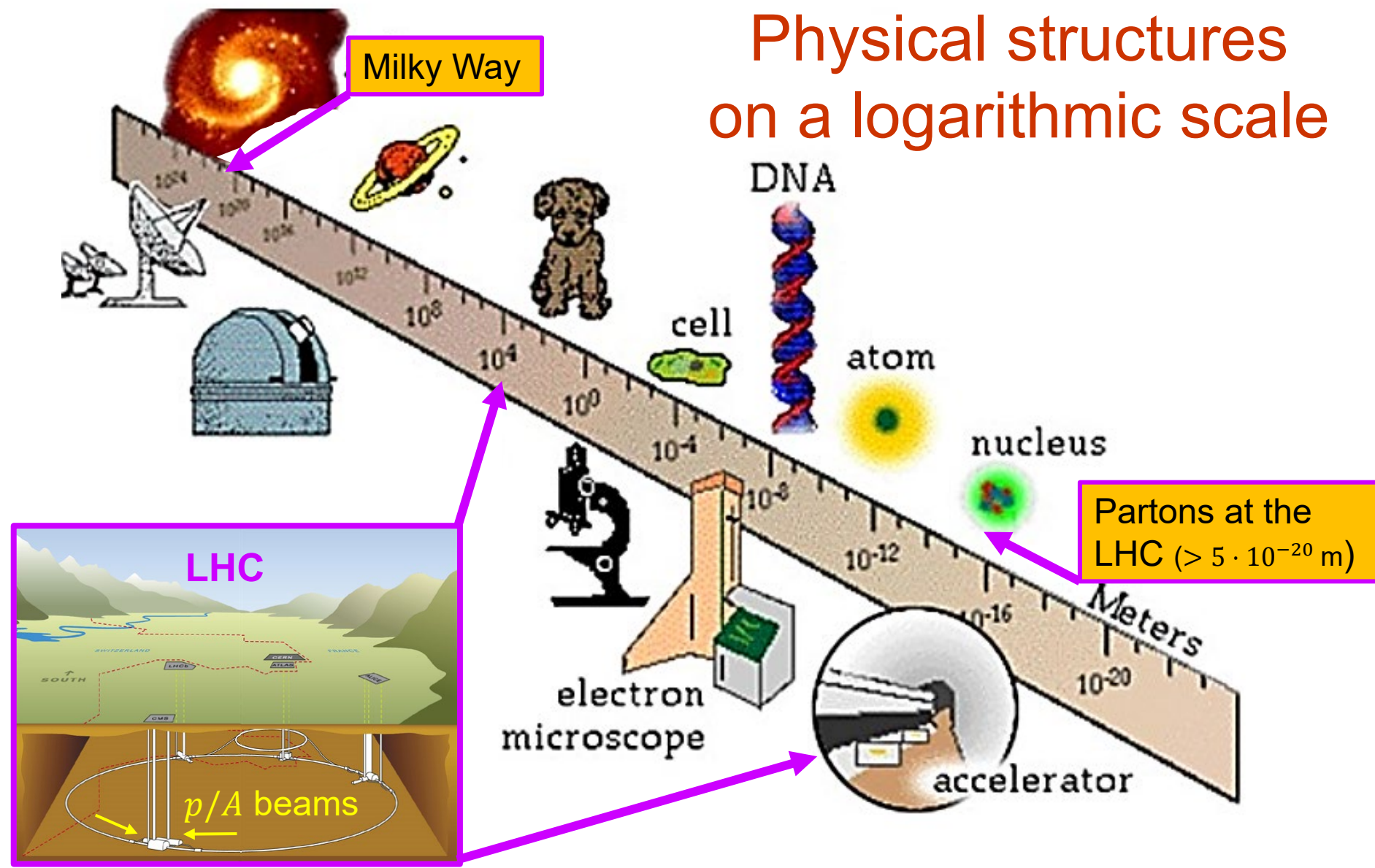
- for general education, cross-domain collaborations, outreach...
- Parton model, quantum Minecraft, movies,...
- inherently approximate
- Mostly in time-space (t, \vec{r}) coordinates



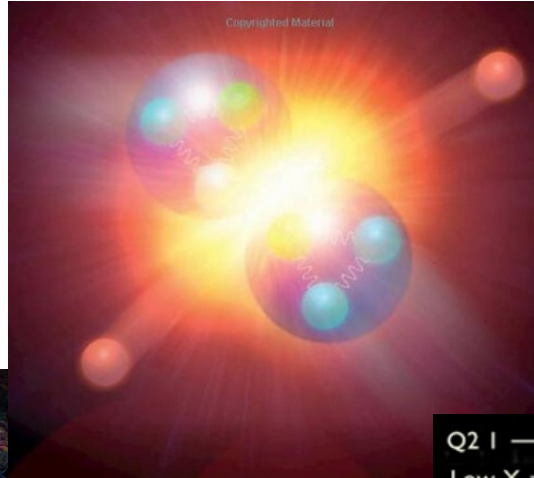
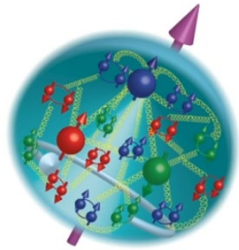
Particle colliders are the most powerful microscopes



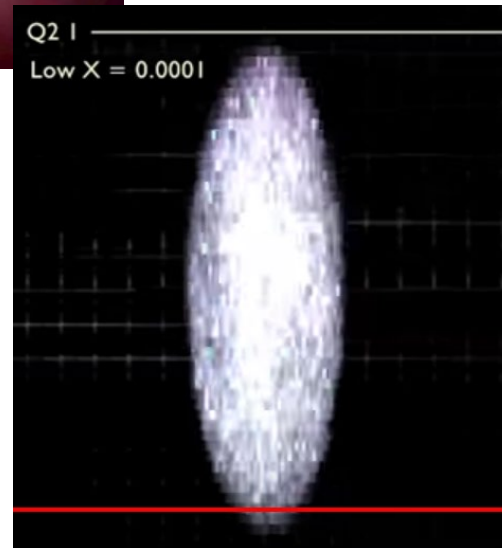
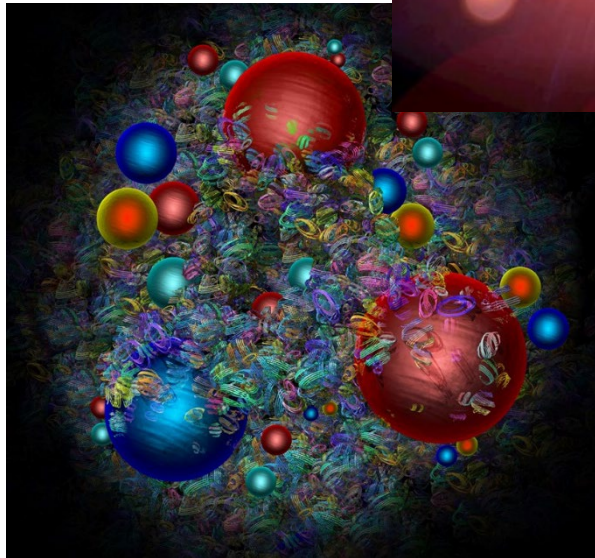
Physical structures on a logarithmic scale



Picturing a fast-moving proton



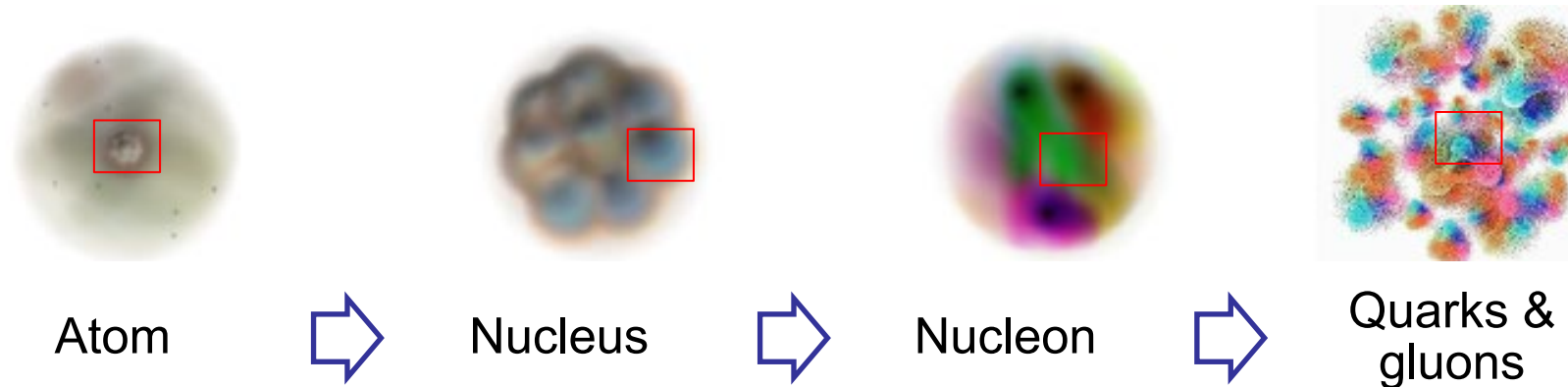
How does its instant photo in our TeV-band microscope look like?



???

Simplified

Zooming in on the atom and atomic nucleus



A short-distance probe (virtual photon, heavy boson, gluon) resolves increasingly small structures inside the nucleon.

Simplified

Quantum Minecraft



Nature is simulated using a discrete cycle time Δt and elastic 3-dimensional elementary blocks of size $\Delta \vec{r}$

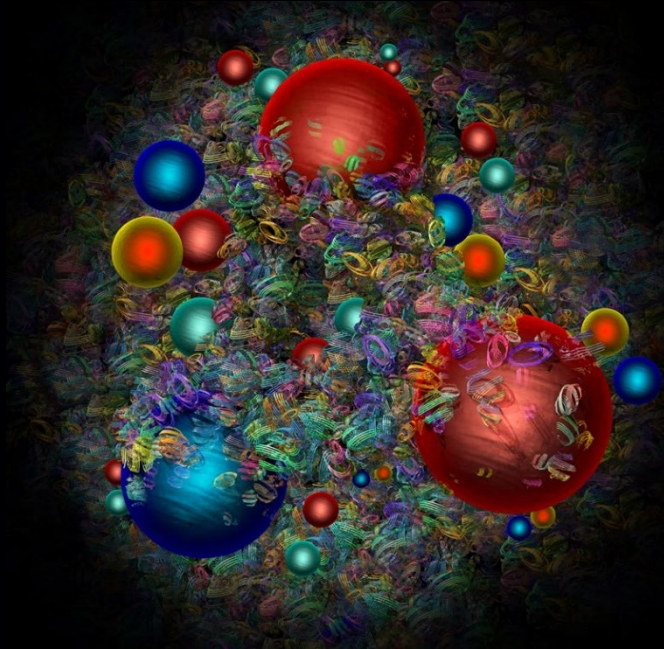
The record of each block is given by its quantum numbers (position \vec{r} , size $\Delta \vec{r}$, energy E , charges, etc.)

Important quantum numbers (charges) are conserved (=their global or local sums do not change)

Image dynamics is limited by two fundamental constants, \hbar (“precision of the record”) and c (“connectivity speed between the blocks”)

A proton at rest

$$V \approx 0$$



At the LHC, the proton appearance is very different!

- Proton structure is described by nonperturbative and lattice quantum chromodynamics

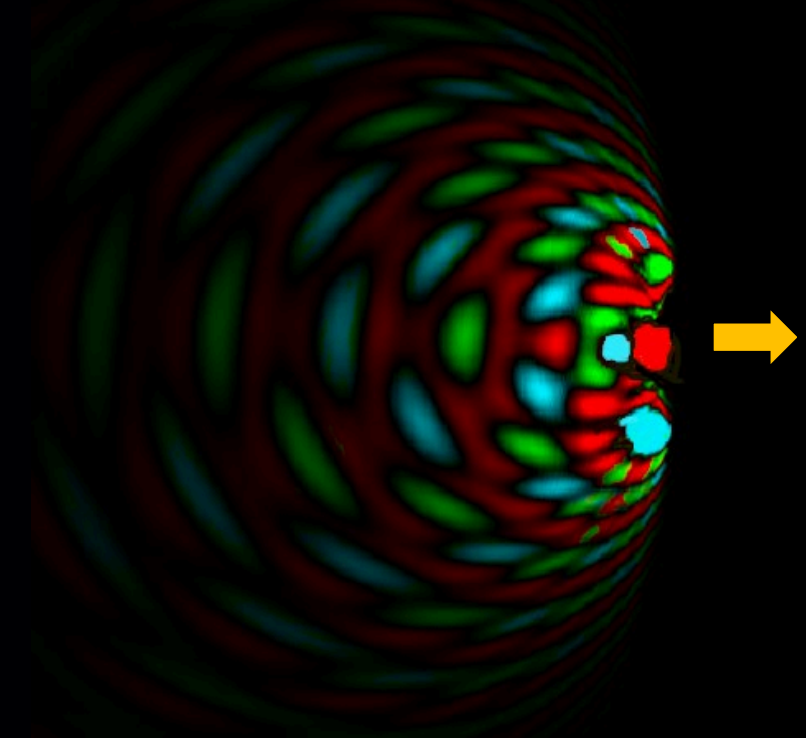
Simplified

A proton at a collider

moving with speed $V \approx c$ to the right

In the collider detector reference frame:

- Proton structure is practically frozen in time
- A proton appears as a narrow high-density shock wave followed by an extended low-density afterglow [relativistic Doppler effect]



- **Parton distribution functions (PDFs) describe densities of quarks and gluons in the shock-wave (light-cone) region**

Simplified

A hydrogen atom flying toward the viewer



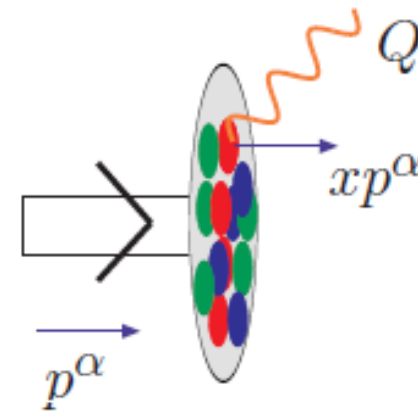
Reality: valence quarks are like flickering, rotating quantum vortices with color predominantly at their centers

The proton spins. Quark's internal angular momenta contribute $< 30\%$ of the net longitudinal proton spin, the rest is due the gluons and orbital angular momentum

Simplified

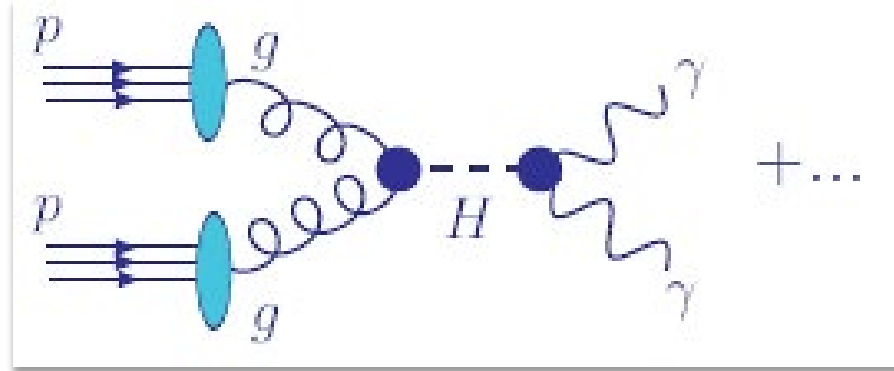
$$f_{a/h}(x, Q)$$

Unpolarized collinear parton distributions $f_{a/h}(x, Q)$ are associated with probabilities for finding a parton a with the “+” momentum xp^+ in a hadron h with the “+” momentum p^+ for $p^+ \rightarrow \infty$, at a resolution scale $Q > 1 \text{ GeV}$



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

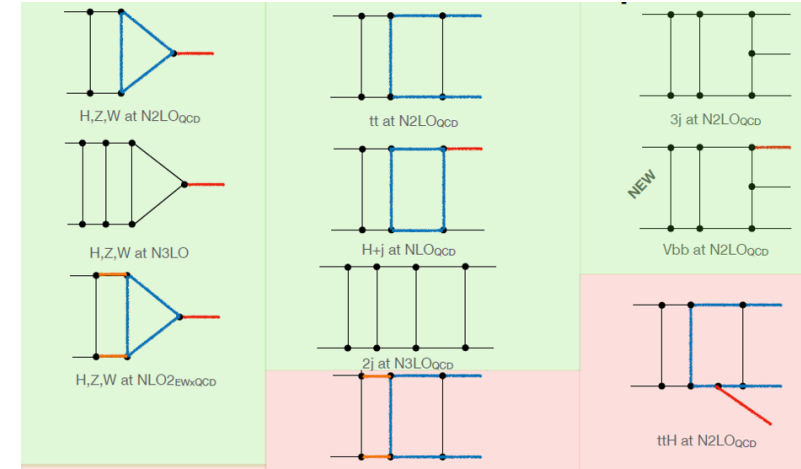
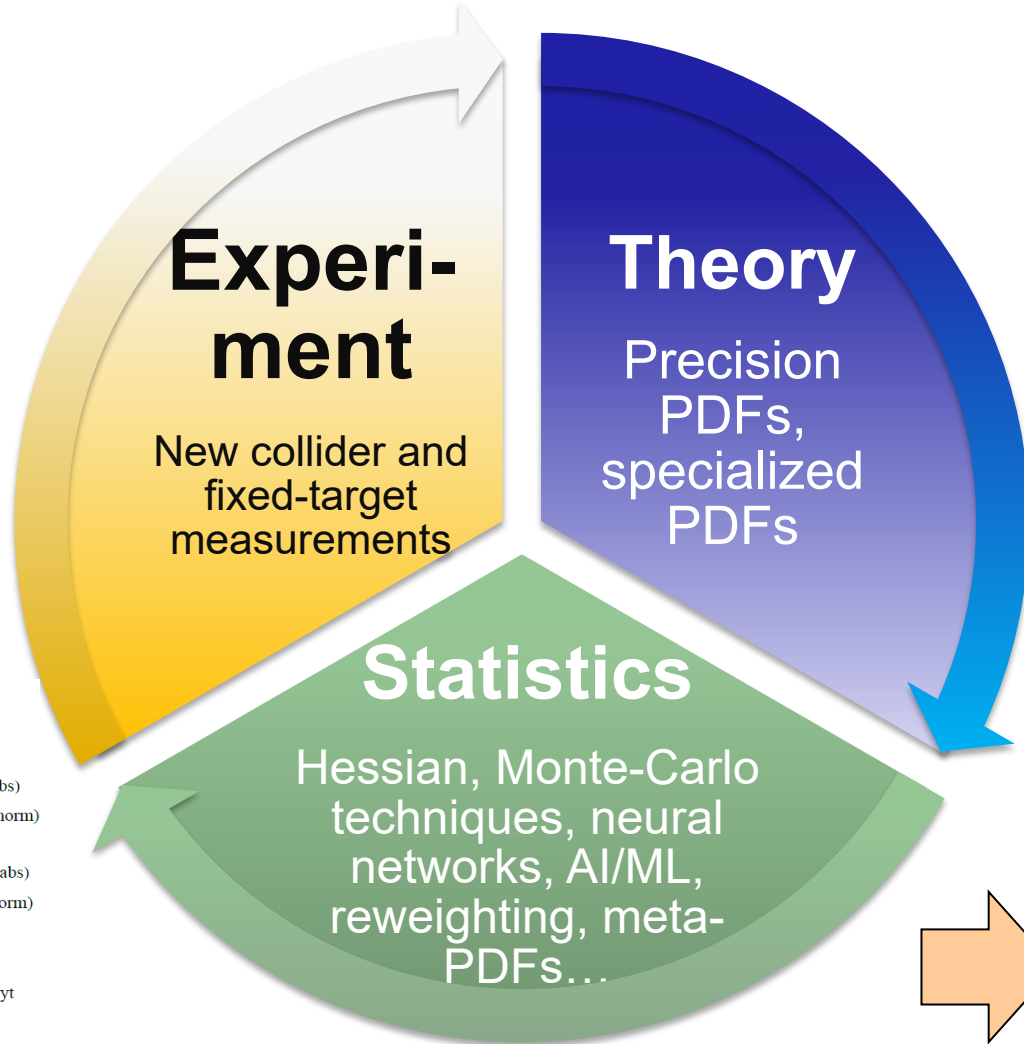
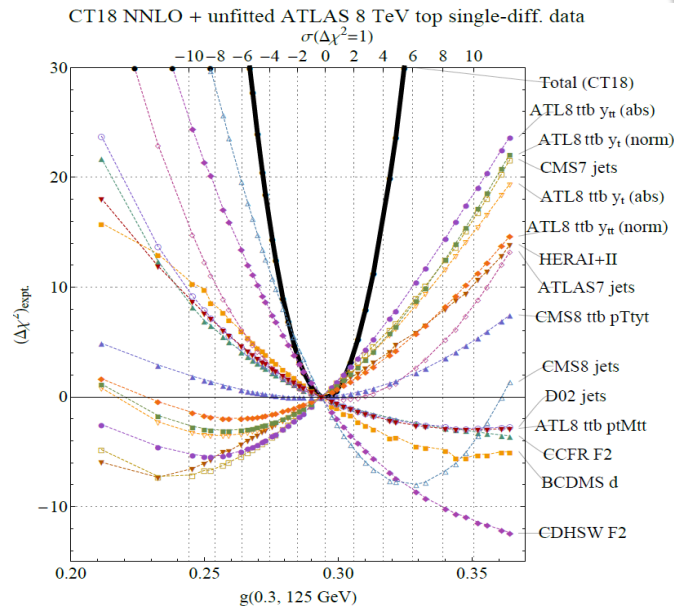
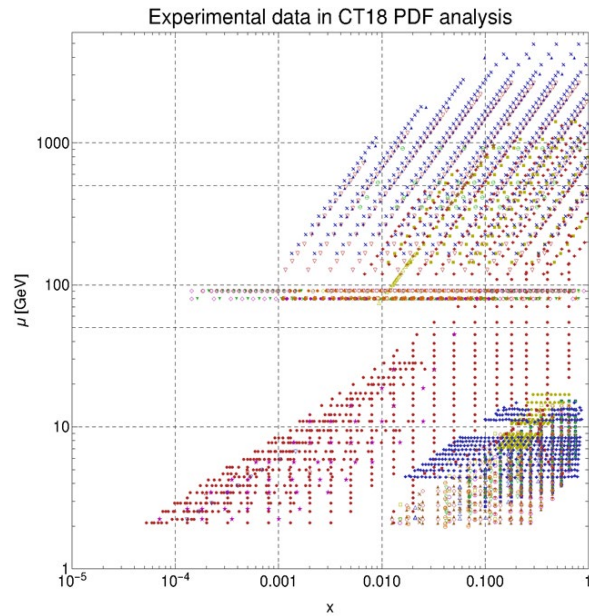
Rigorous



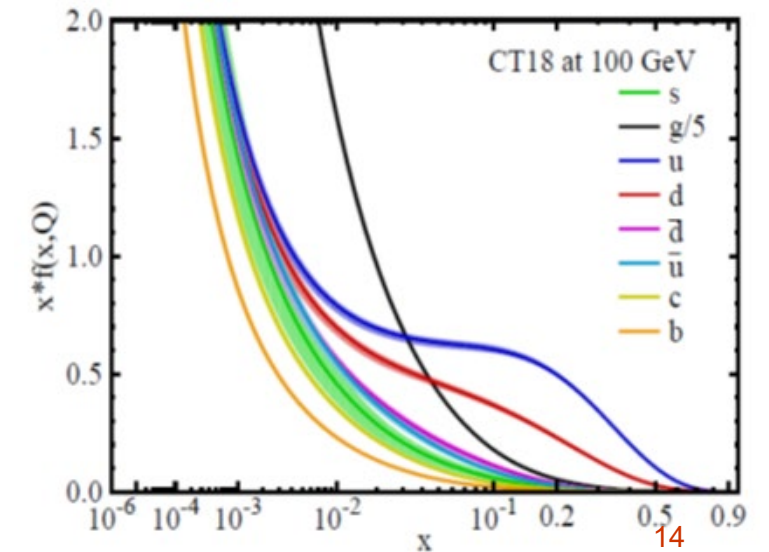
$$\sigma_{pp \rightarrow H \rightarrow \gamma\gamma X}(Q) = \sum_{a,b=g,q,\bar{q}} \int_0^1 d\xi_a \int_0^1 d\xi_b \hat{\sigma}_{ab \rightarrow H \rightarrow \gamma\gamma} \left(\frac{x_a}{\xi_a}, \frac{x_b}{\xi_b}, \frac{Q}{\mu_R}, \frac{Q}{\mu_F}; \alpha_s(\mu_R) \right) \\ \times f_a(\xi_a, \mu_F)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

$\hat{\sigma}$ 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

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

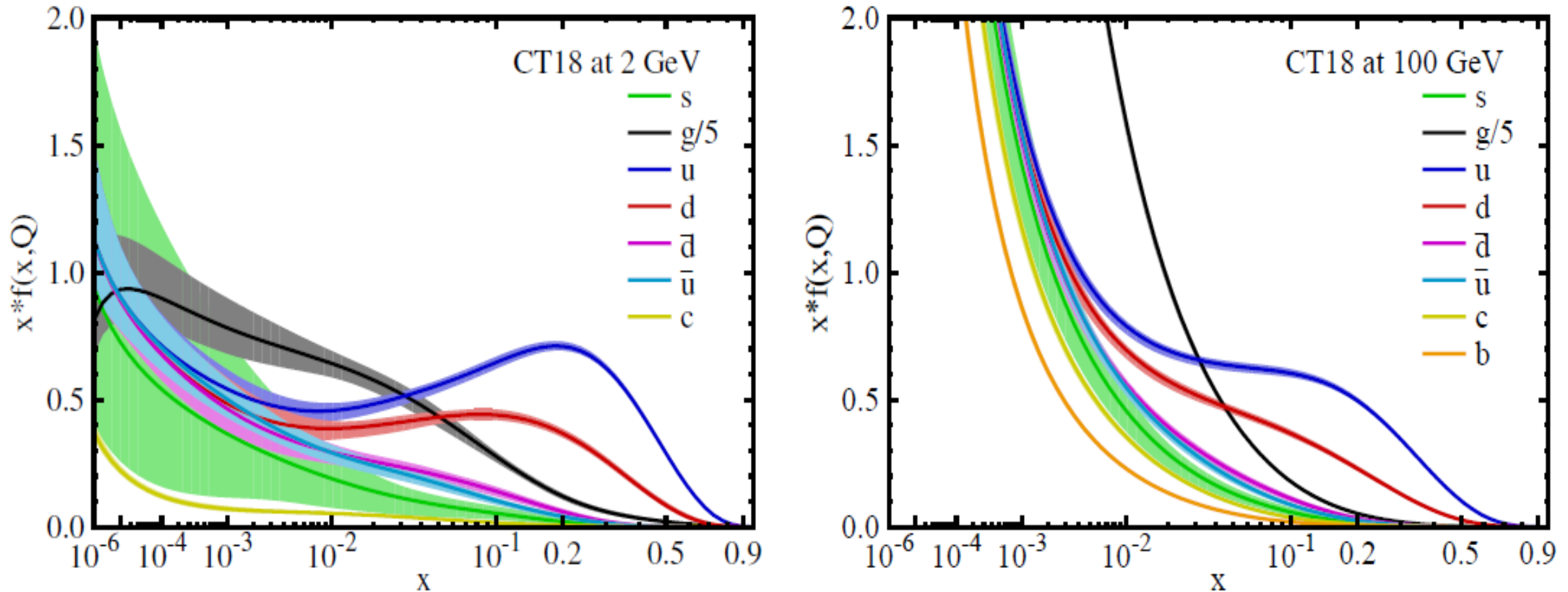


Parton distribution functions with uncertainties



CT18 NNLO parton distributions

Recent PDFs from the CTEQ-TEA group arXiv:[1912.10053](https://arxiv.org/abs/1912.10053) [hep-ph]



PDF parametrizations provide central (best-fit) PDFs and uncertainty estimates for parton flavors (quarks, antiquarks, gluons). PDF dependence on the energy scale (Q) is computed using perturbative evolution equations.

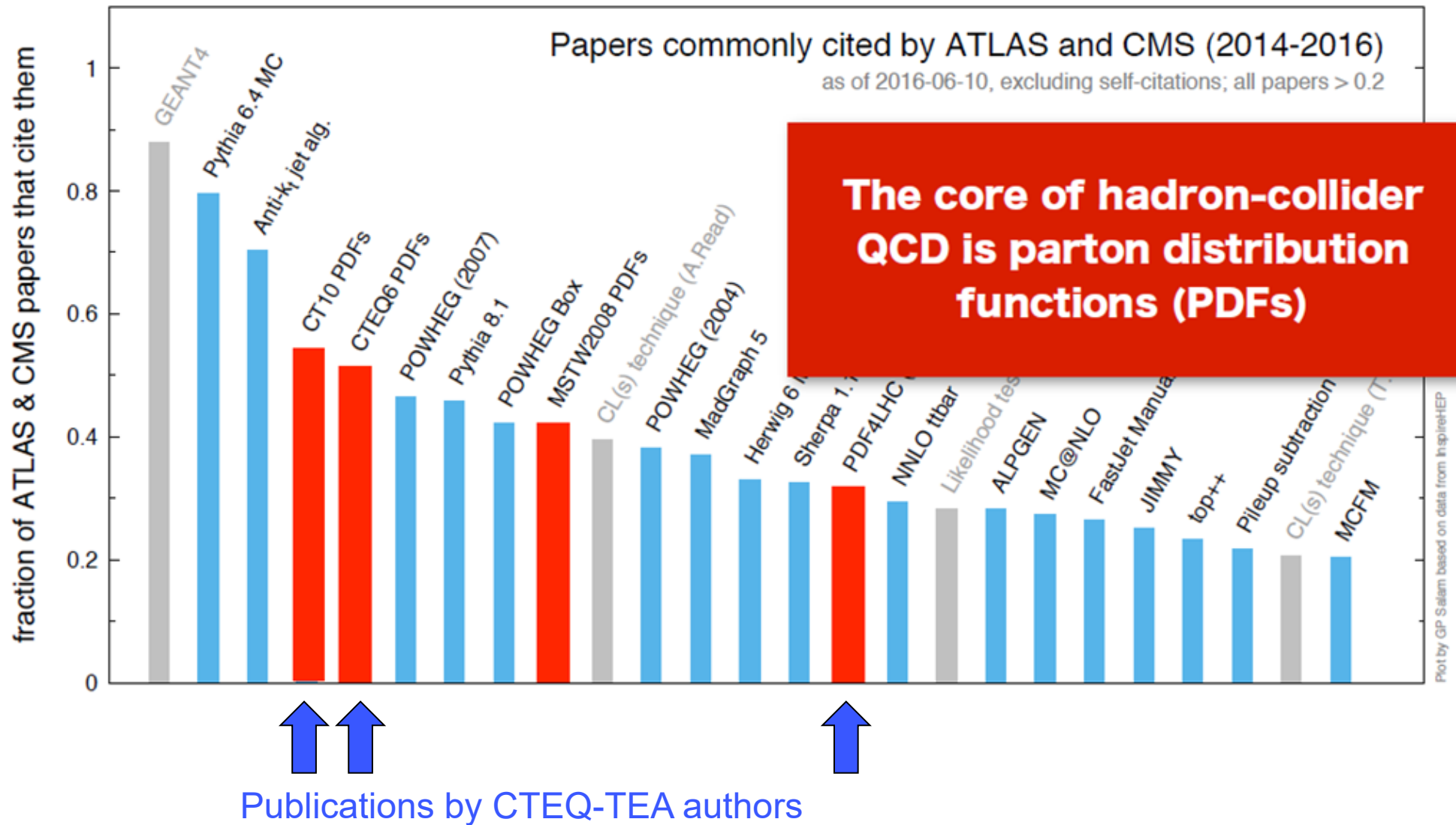
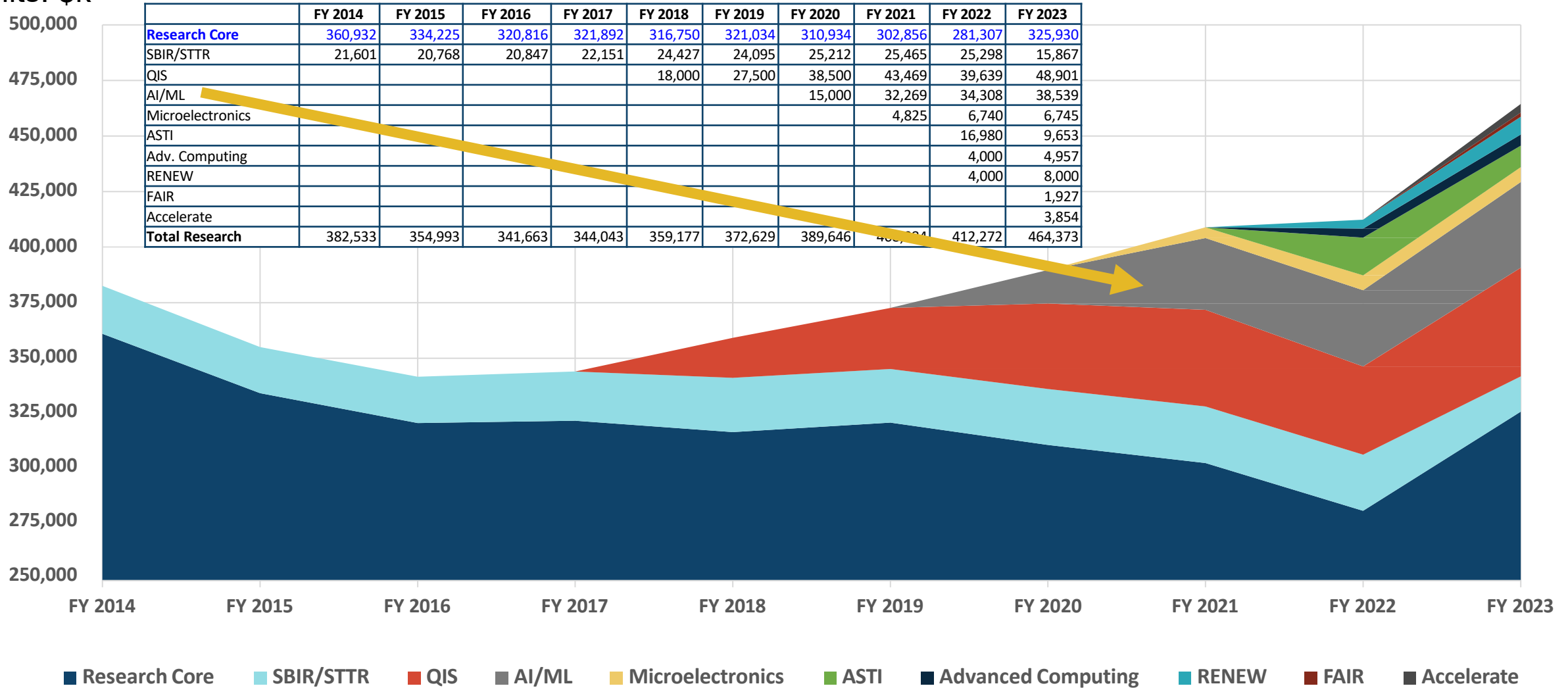


Figure credit: G. Salam

HEP Research increasingly invests in AI/ML

Units: \$k



CTEQ-TEA group has 20+ years of experience in **uncertainty quantification** for multivariate/AI-based inference

Particle physics & probabilistically rigorous AI

“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. 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

PCAST report “Supercharging Research: Harnessing AI to Meet Global Challenges”

https://www.whitehouse.gov/wp-content/uploads/2024/04/AI-Report_Upload_29APRIL2024_SEND-2.pdf

Executive Office of the President
President’s Council of Advisors on
Science and Technology

APRIL 2024



Will AI/AGI hurt or help replicability?

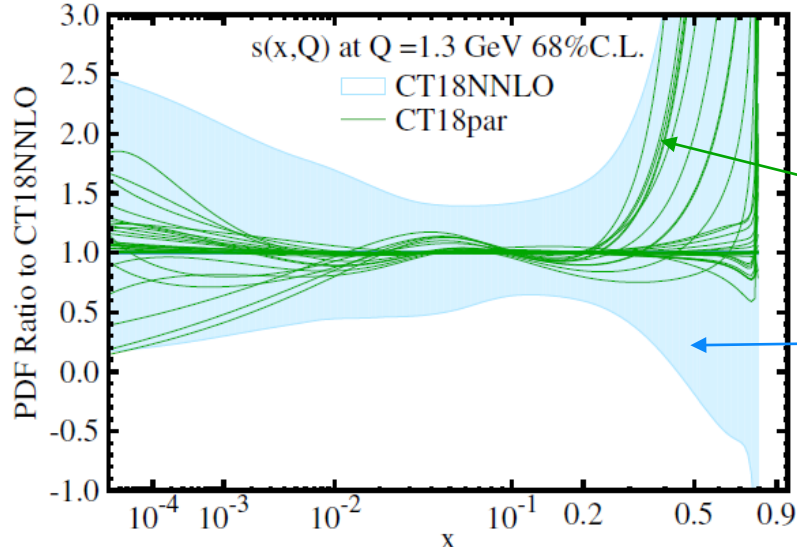
Replicability across STEM fields

“... 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](#)

Hessian PDFs, uncertainties (CT18 PDFs)

“Bayesian exploration with Gaussian emulation”



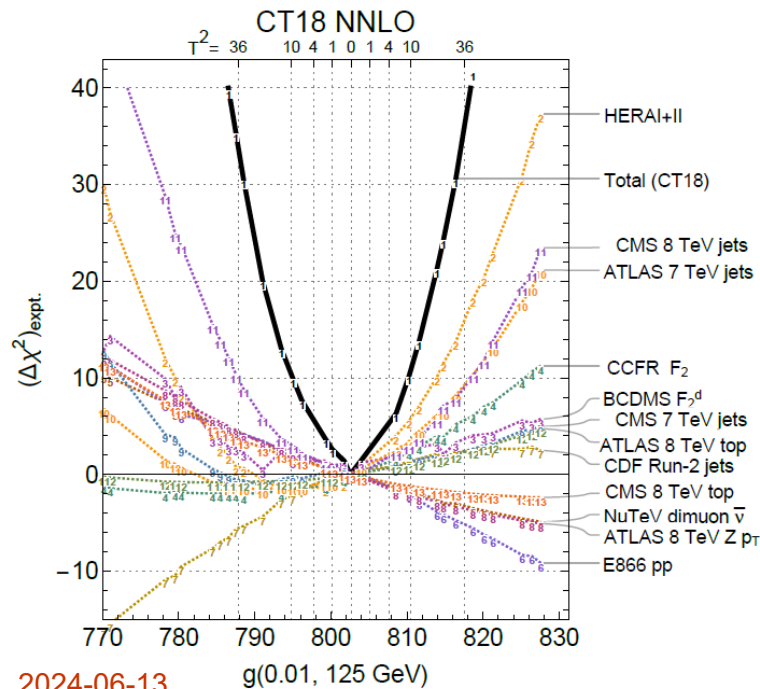
preliminary PDFs for alternative parametrizations

final uncertainty with one parametrization

Preliminary fits explore experimental, theoretical, parametrization, methodological uncertainties

The final PDFs are released as one quasi-Gaussian (**Hessian**) error set (50-60) that approximates the total uncertainty due to the above factors.

These error sets are constructed with a fixed choice of polynomial parametrization forms. The totality of error sources (not only experimental) is emulated by introducing **tolerance** T : the final 1σ uncertainty corresponds to $\Delta\chi^2 = T^2 \sim 10 - 30$, rather than $T^2 = 1$.



An alternative: Neural-network PDFs

Use **bootstrap** to estimate **aleatory** data fluctuations for a fixed training methodology (called “importance sampling” by NNPDF)

Parametrize PDFs using CNNs with optimized hyperparameters and restricted by prior conditions (positivity of cross sections, etc.)

[The whole fit is based on one “tuned” NN architecture]

All fitting codes, especially the NNPDF one, employ grid techniques for fast integration

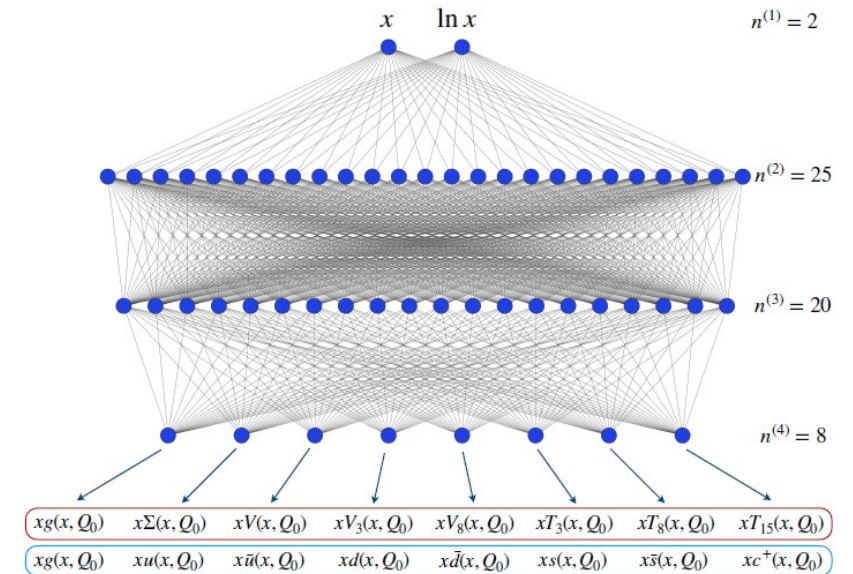
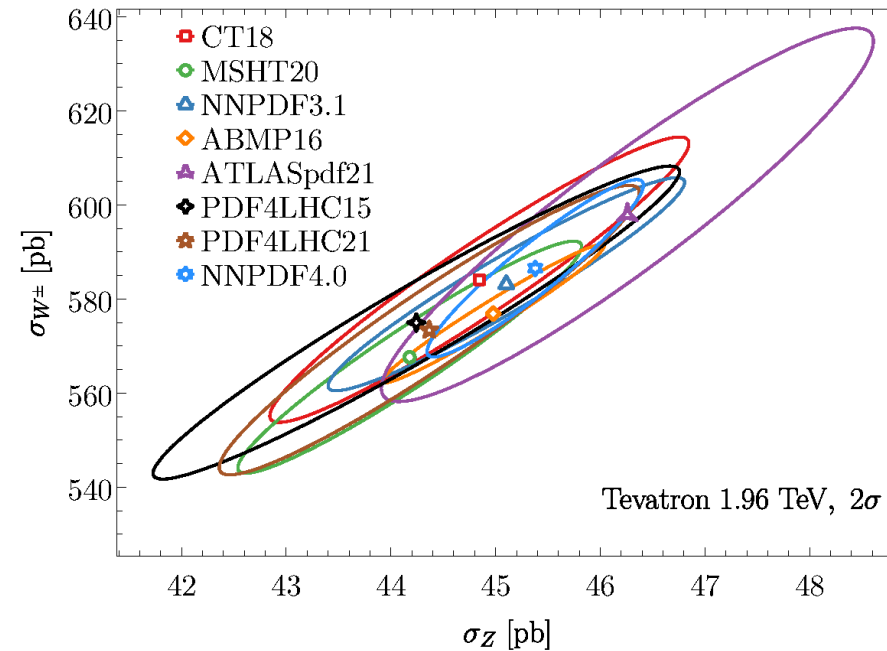
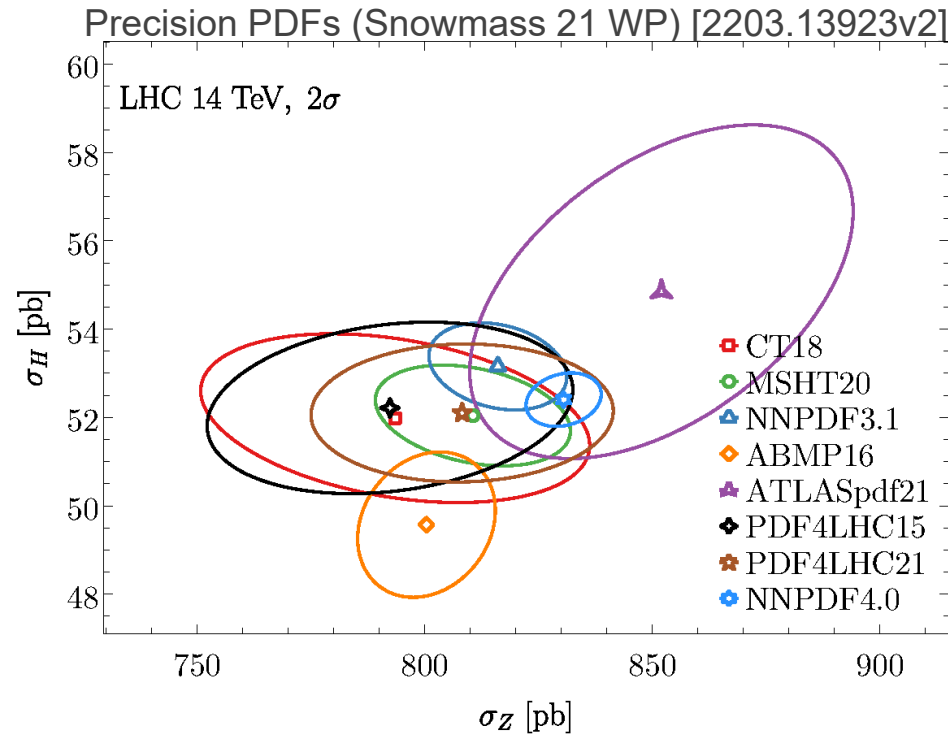


Figure 3.9. The neural network architecture adopted for NNPDF4.0. A single network is used, whose eight output values are the PDFs in the evolution (red) or the flavor basis (blue box). The architecture displayed corresponds to the optimal choice in the evolution basis; the optimal architecture in the flavor basis is different as indicated by Table 3.3).

NNPDF4.0 PDF ensemble,
R. Ball et al., arXiv:2109.02653

The tolerance puzzle

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



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

Comparisons of the latest PDF sets

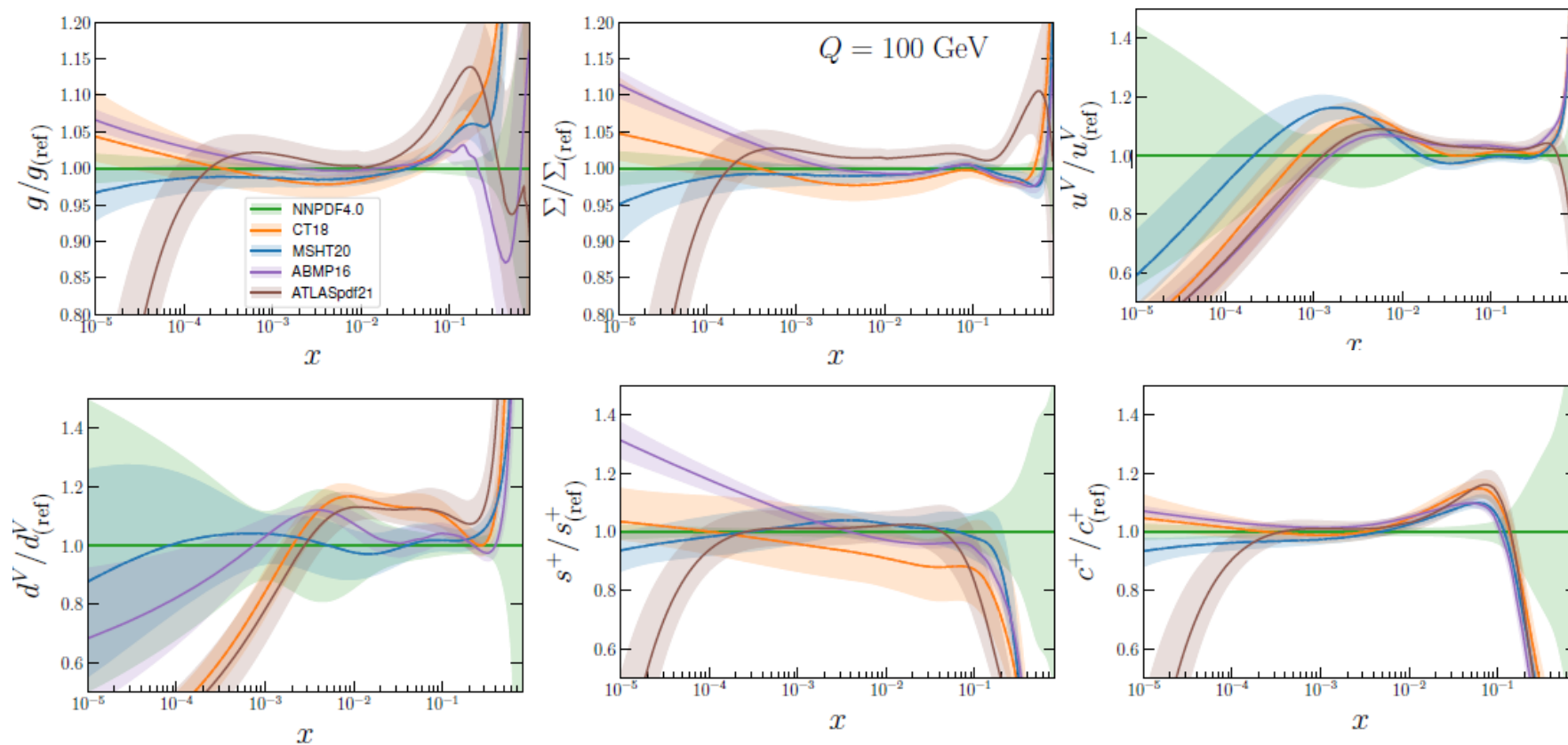
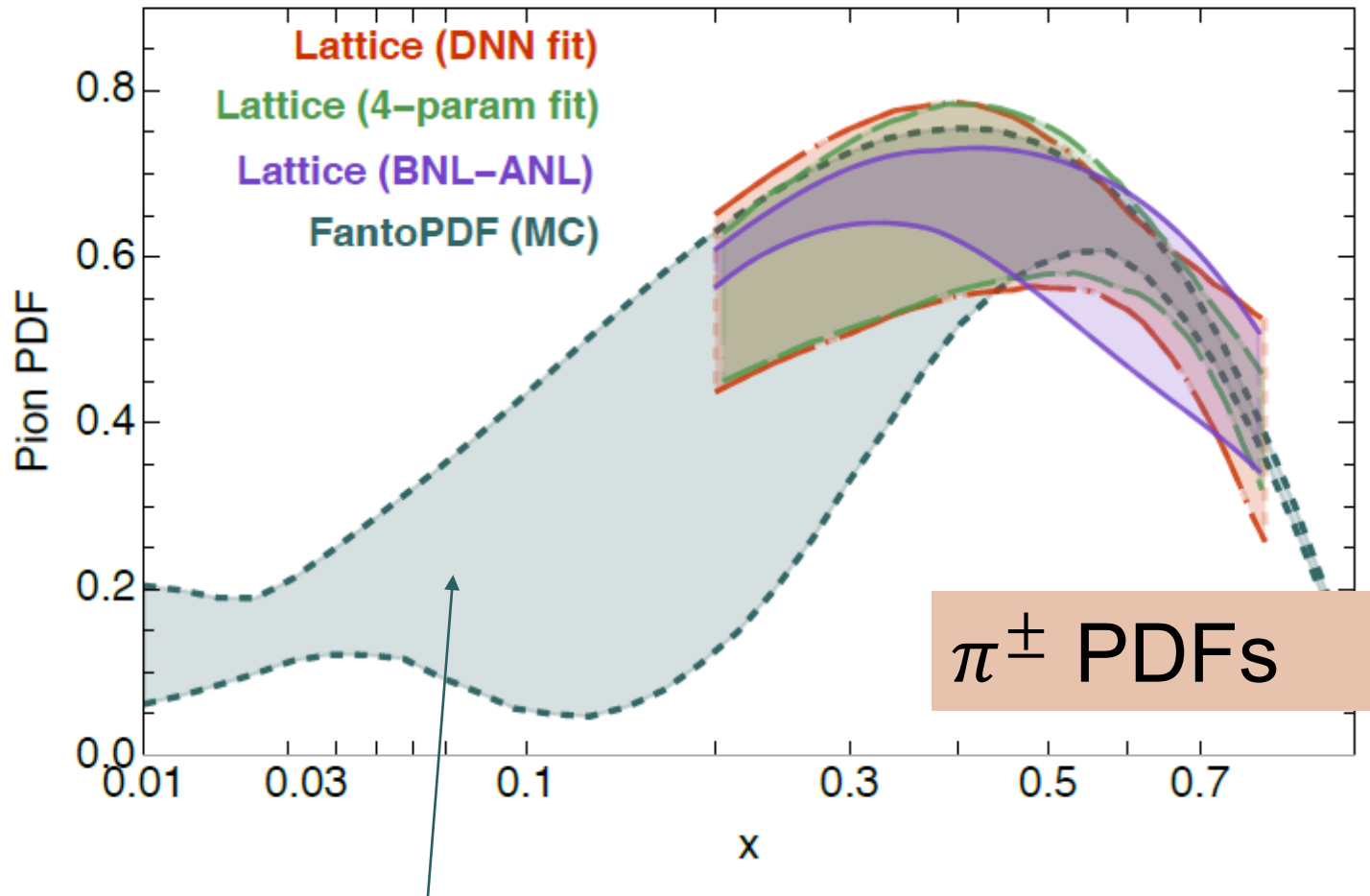


FIG. 2. Comparison of the PDFs at $Q = 100$ GeV. The PDFs shown are the N2LO sets of NNPDF4.0, CT18, MSHT20, ABMP16 with $\alpha_s(M_Z) = 0.118$, and ATLASpdf21. The ratio to the NNPDF4.0 central value and the relative 1σ uncertainty are shown for the gluon g , singlet Σ , total strangeness $s^+ = s + \bar{s}$, total charm $c^+ = c + \bar{c}$, up valence u^V and down valence d^V PDFs.

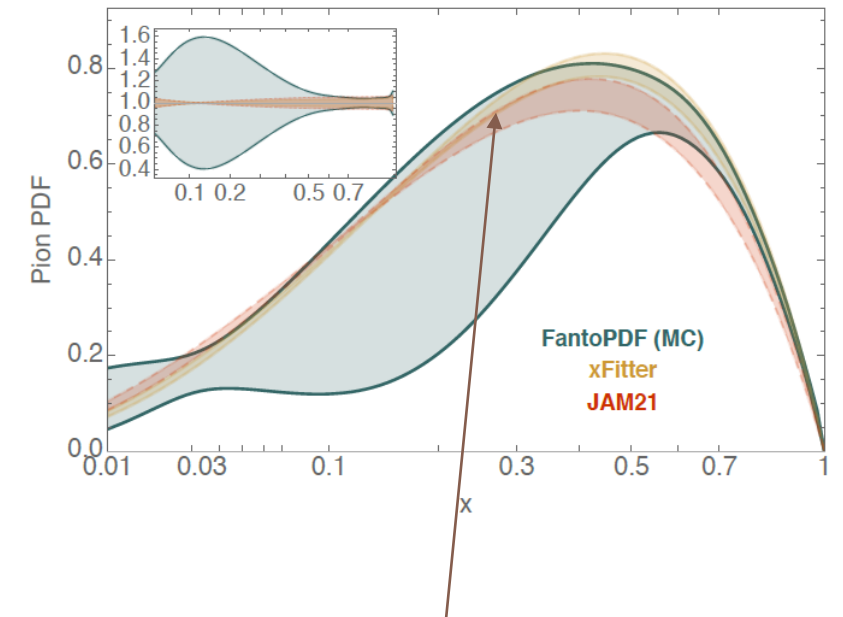
Lattice QCD already predicts some features of PDFs from first principles

$xV(x, Q)$ at $Q=2. \text{ GeV}$, 68% c.l. (band)



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

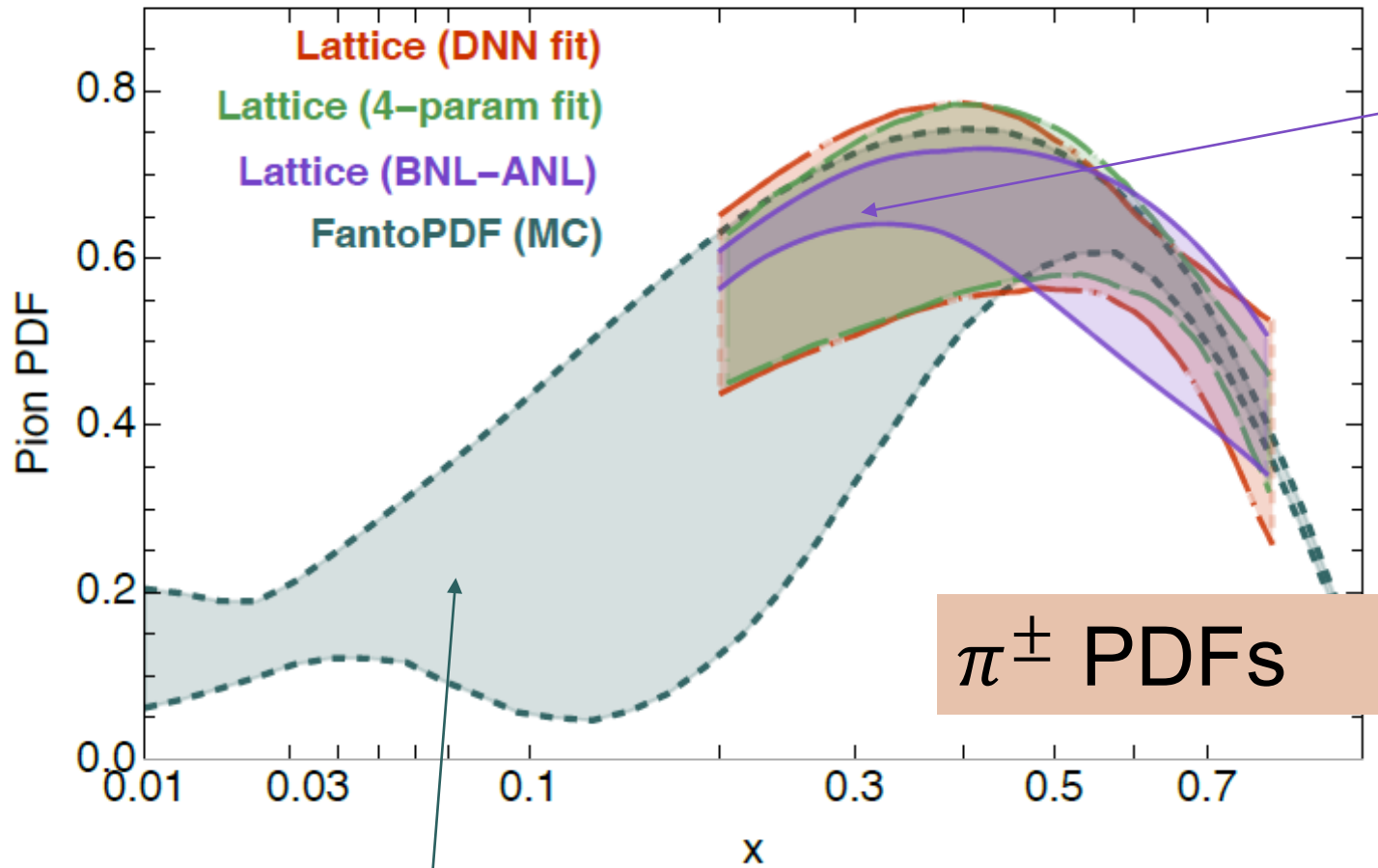
$xV(x, Q)$ at $Q=1.4 \text{ GeV}$, 68% c.l. (band)



without parametrization
dependence

Lattice QCD already predicts some features of PDFs from first principles

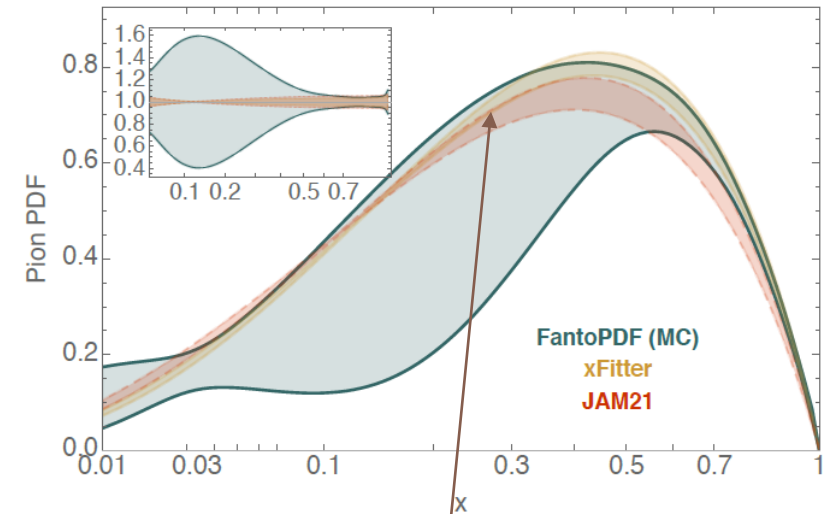
$xV(x, Q)$ at $Q=2. \text{ GeV}$, 68% c.l. (band)



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

are the lattice uncertainties fully estimated?

$xV(x, Q)$ at $Q=1.4 \text{ GeV}$, 68% c.l. (band)



without parametrization dependence

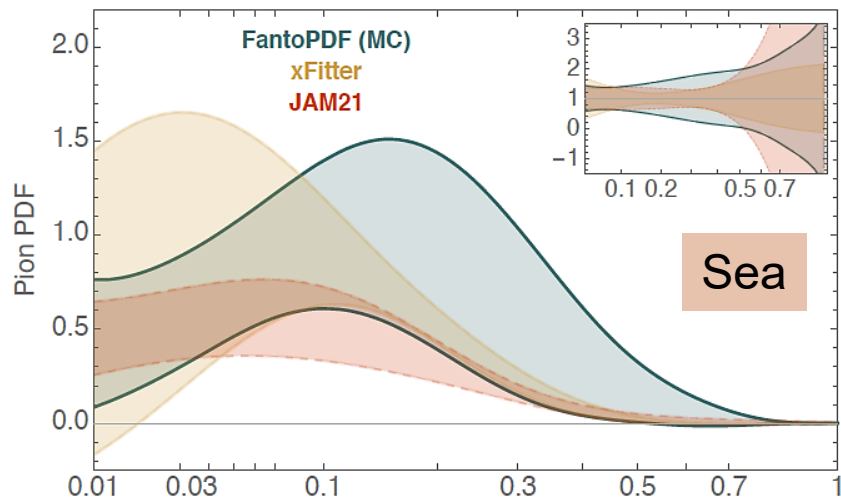
Fantômas pion PDFs: other results

π^\pm

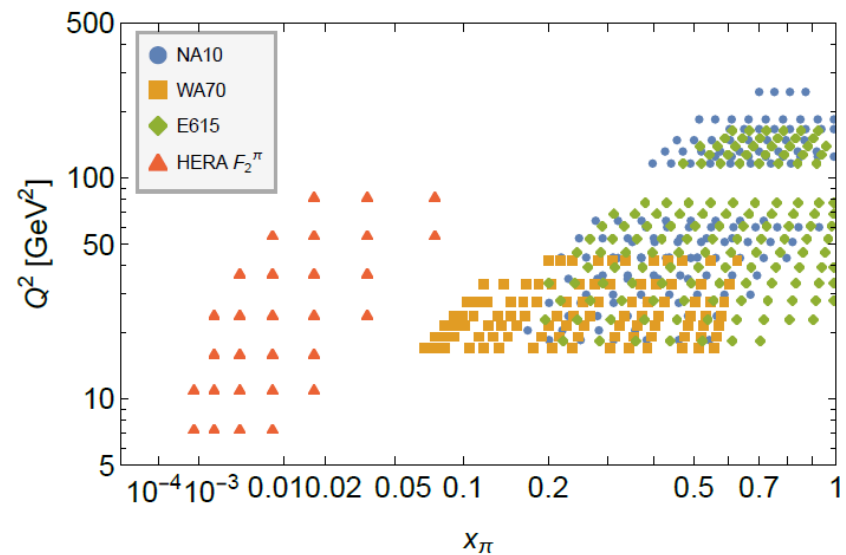
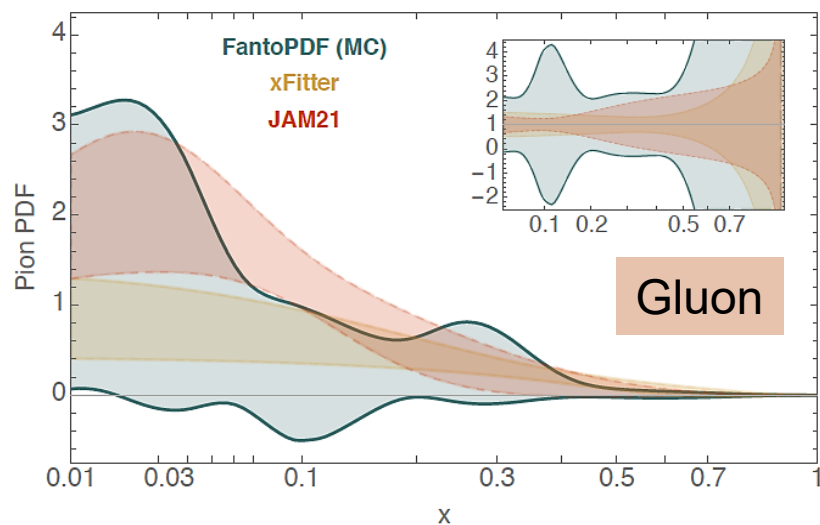
NEW

L. Kotz, A. Courtoy, M. Chavez, P. Nadolsky, F. Olness, arXiv:2311.08447

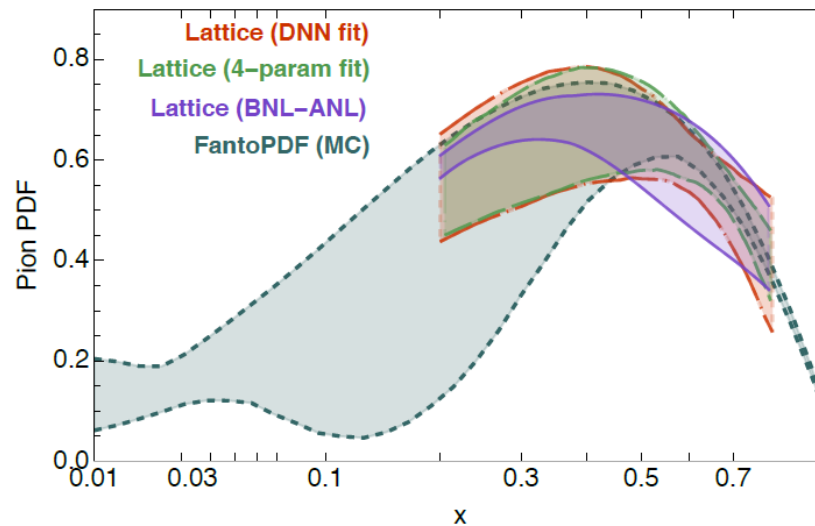
$xS(x,Q)$ at $Q=1.4$ GeV, 68% c.l. (band)



$xg(x,Q)$ at $Q=1.4$ GeV, 68% c.l. (band)

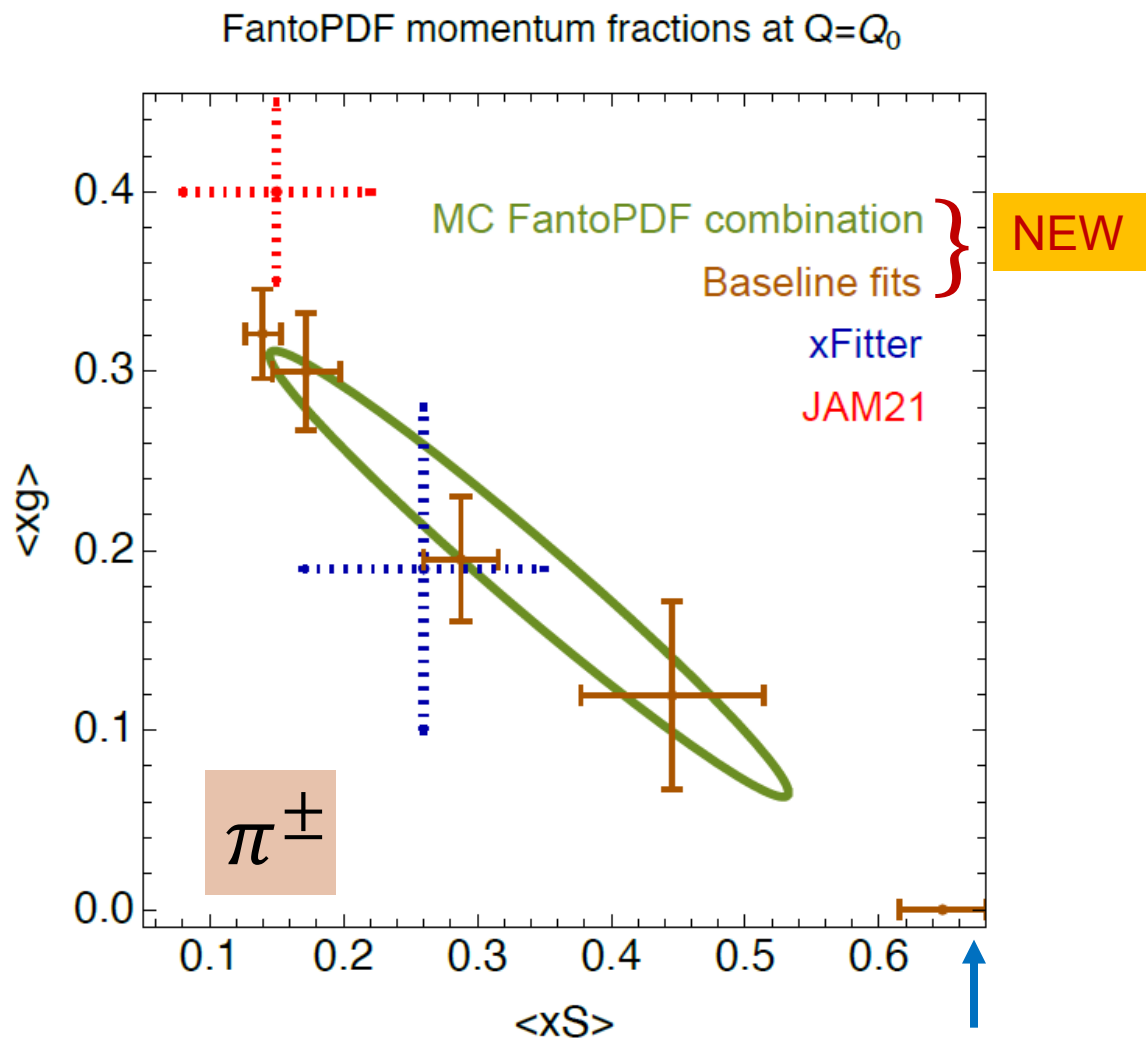


$xV(x,Q)$ at $Q=2.$ GeV, 68% c.l. (band)

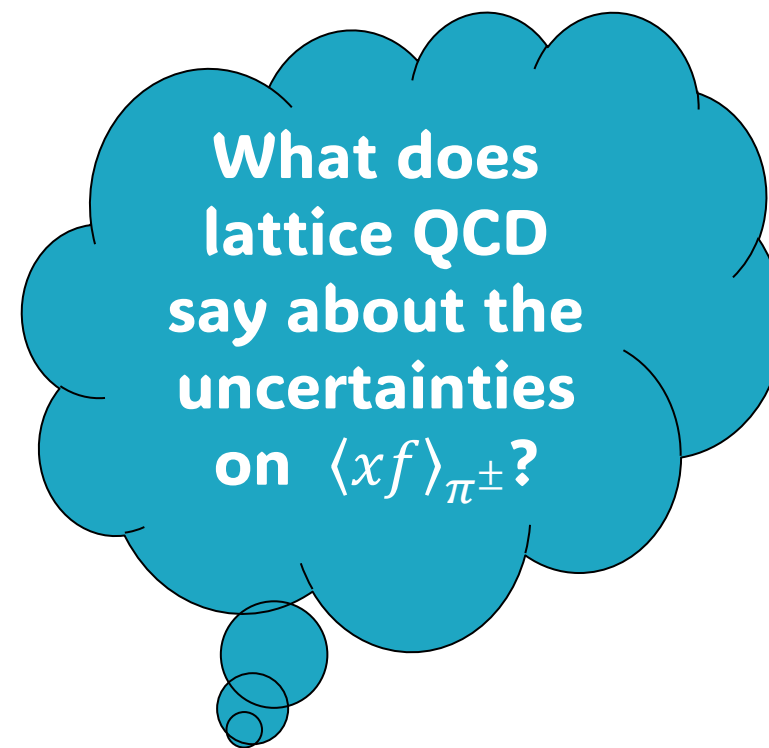


Fantômas pion PDFs: sea and gluon momentum fractions

L. Kotz, A. Courtoy, M. Chavez, P. Nadolsky, F. Olness, arXiv:2311.08447

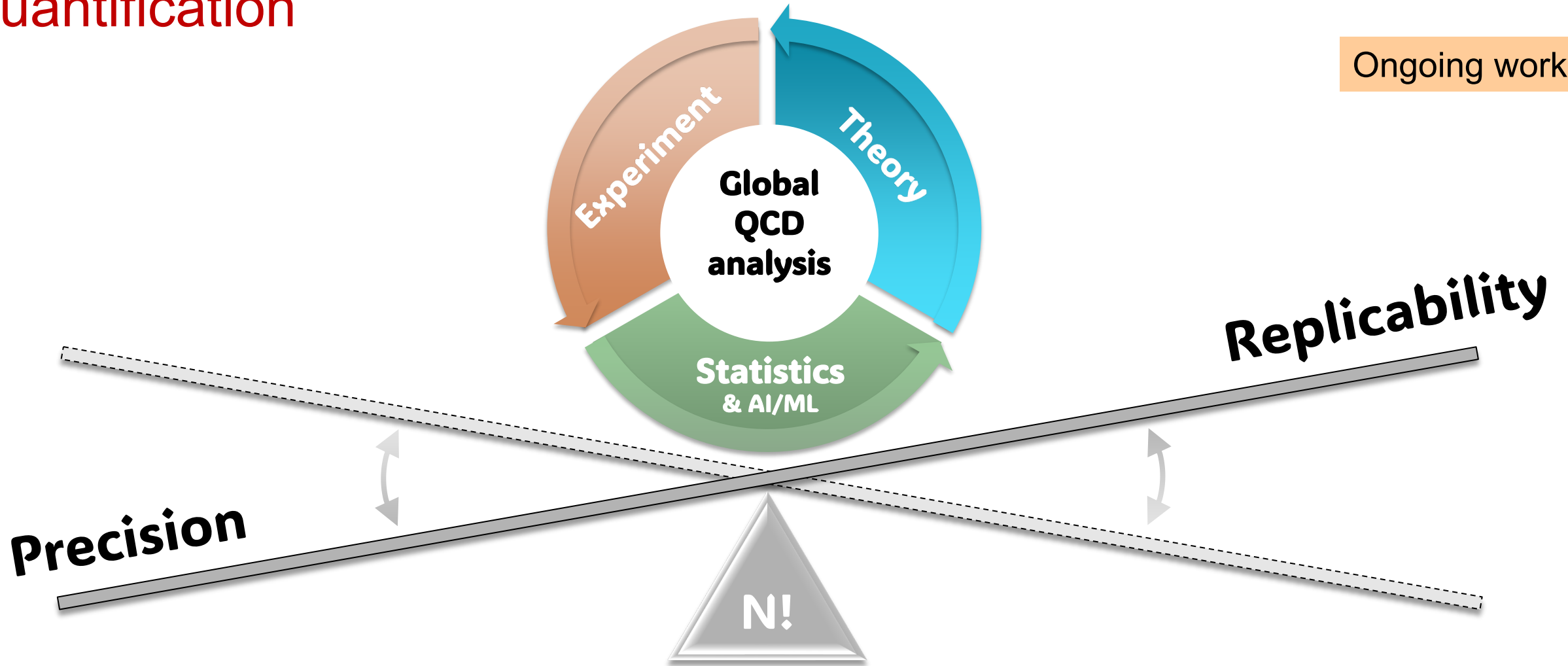


a zero-momentum gluon is experimentally allowed

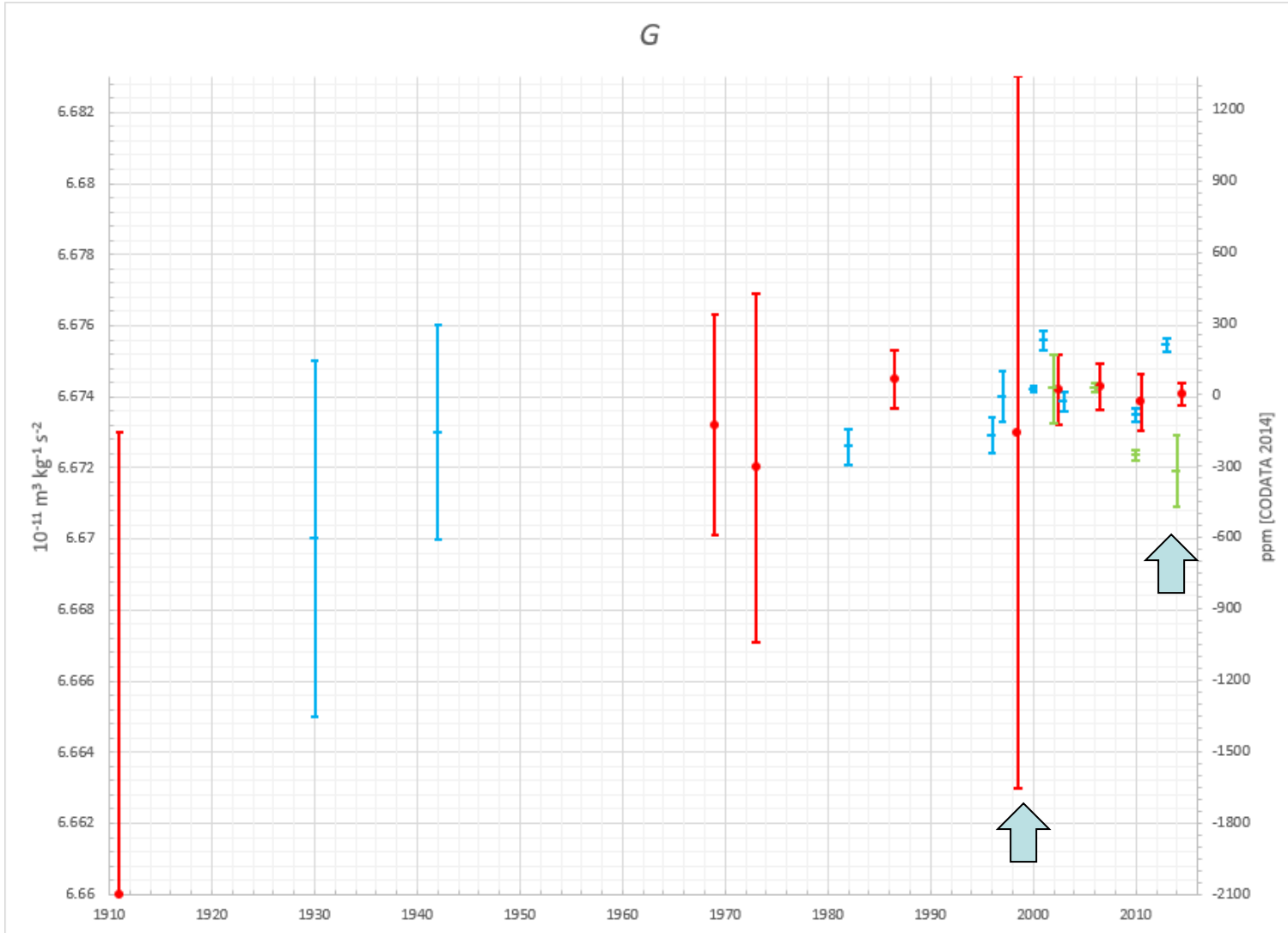


Balancing precision and replicability in PDF uncertainty quantification

Ongoing work



World average for the gravitational constant

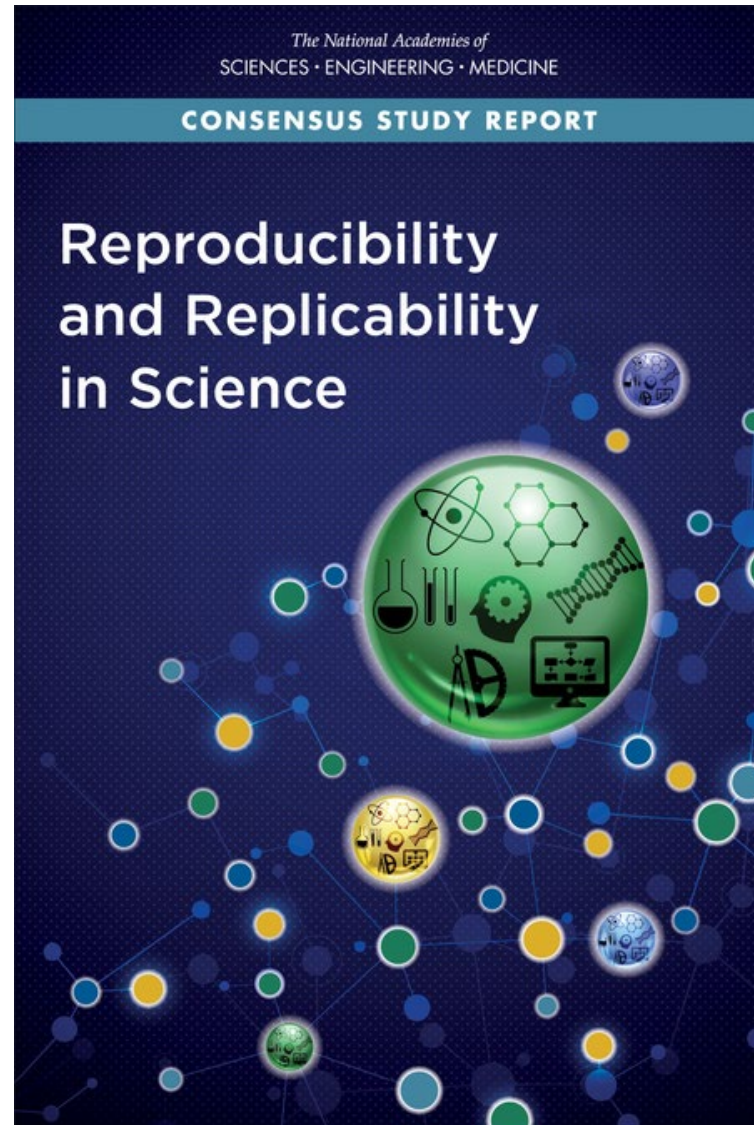


Timeline of measurements and recommended values for G since 1900: values recommended based on the NIST combination (red), individual torsion balance experiments (blue), other types of experiments (green).

The combination error bars are unstable after 1995

Some 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



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

Reproducibility, Replicability, Rigor: definitions

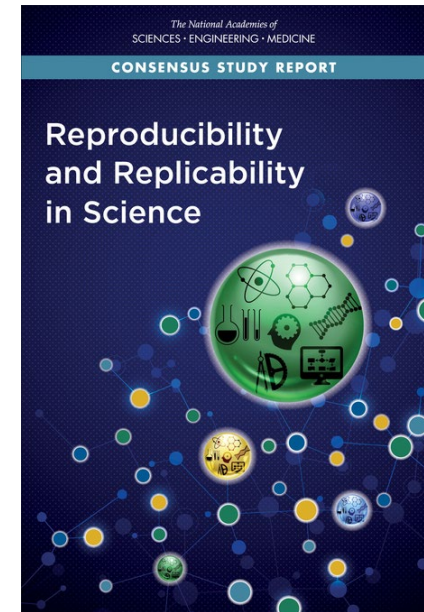
Reproducibility is obtaining consistent results using the same input data; computational steps, methods, and code; and conditions of analysis.

Replicability is obtaining consistent results across studies aimed at answering the same scientific question, each of which has obtained its own data.

Rigor -- the strict application of the scientific method to ensure robust and unbiased experimental design -- makes replication of a study more likely

RRR

Definitions adopted from “*REPRODUCIBILITY AND REPLICABILITY IN SCIENCE*”, Conclusion 3.1
National Academy of Sciences, 2019, <https://doi.org/10.17226/25303>



Replicability risks for precision QCD

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

⇒ **Part 2**

Are W boson mass measurements replicable?

For instance, W boson mass measurements at the Tevatron and LHC

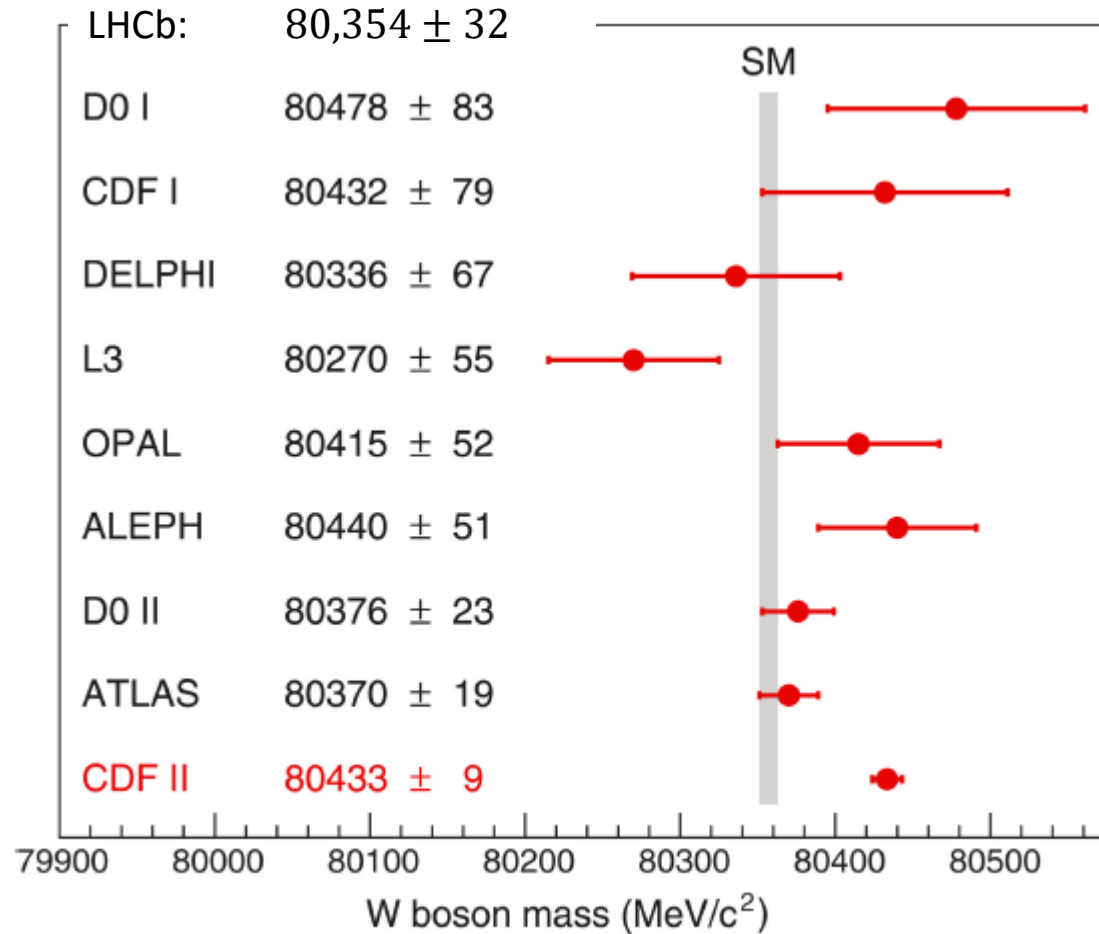
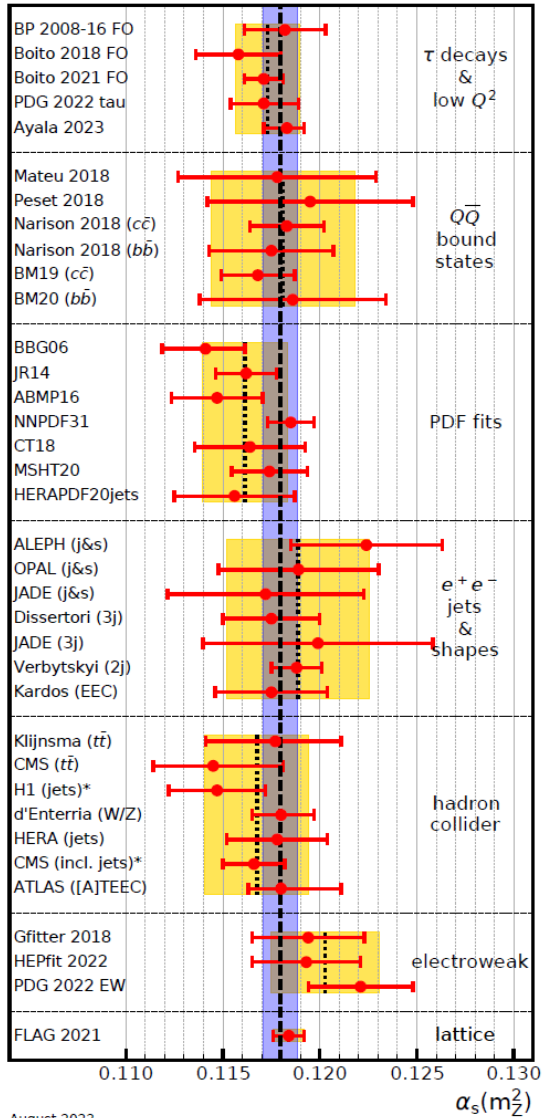
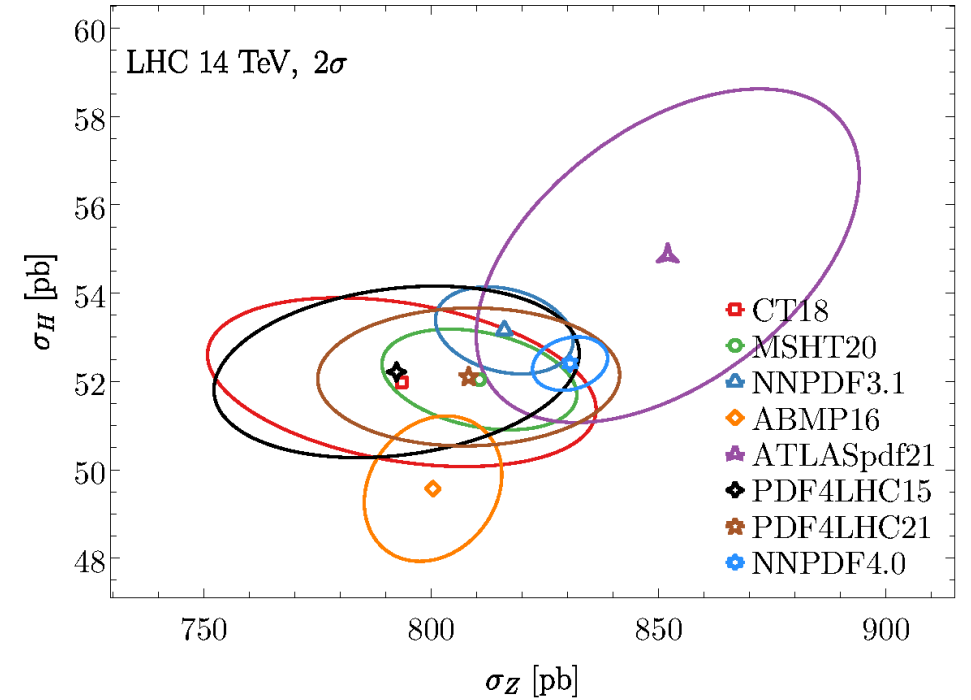


Figure reproduced from CDF-II measurement (Science 376, 170).

Replicability and PDF uncertainties



August 2023

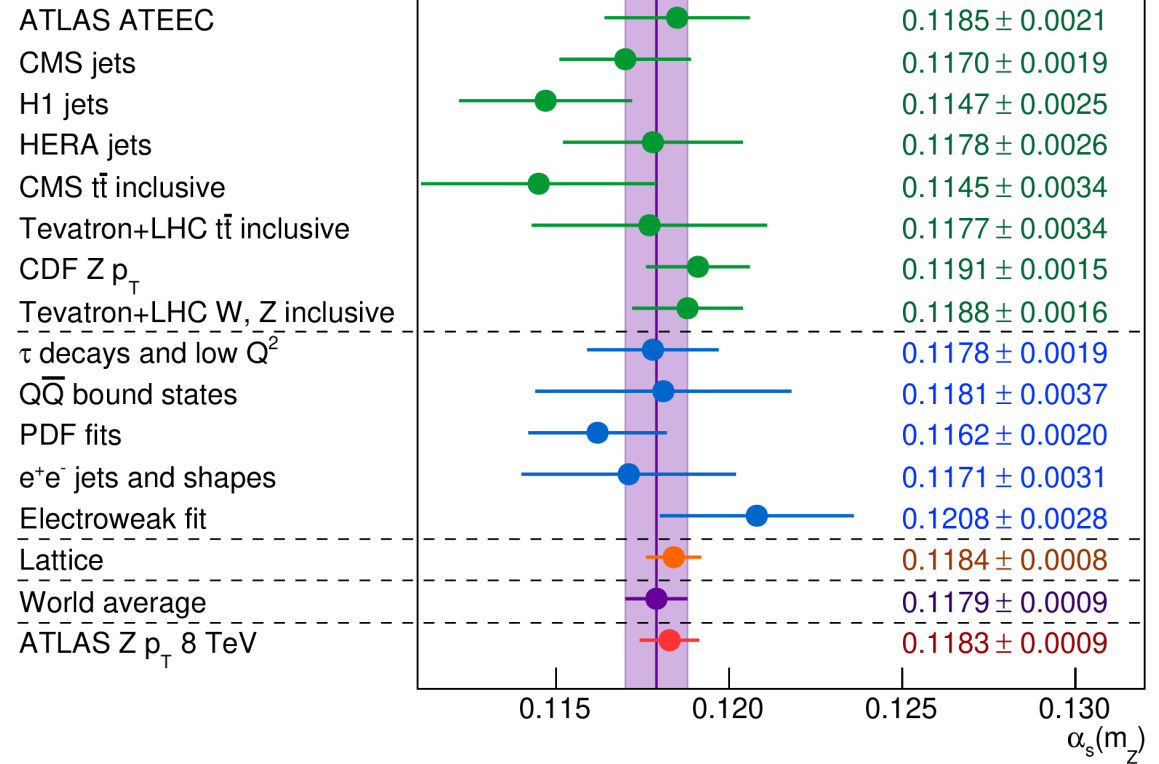


Quantification of **epistemic** PDF uncertainties is a central factor affecting **replicability** of upcoming determinations of the QCD coupling constant α_s , Higgs couplings, mass of weak bosons.

ATLAS measures strength of the strong force with record precision

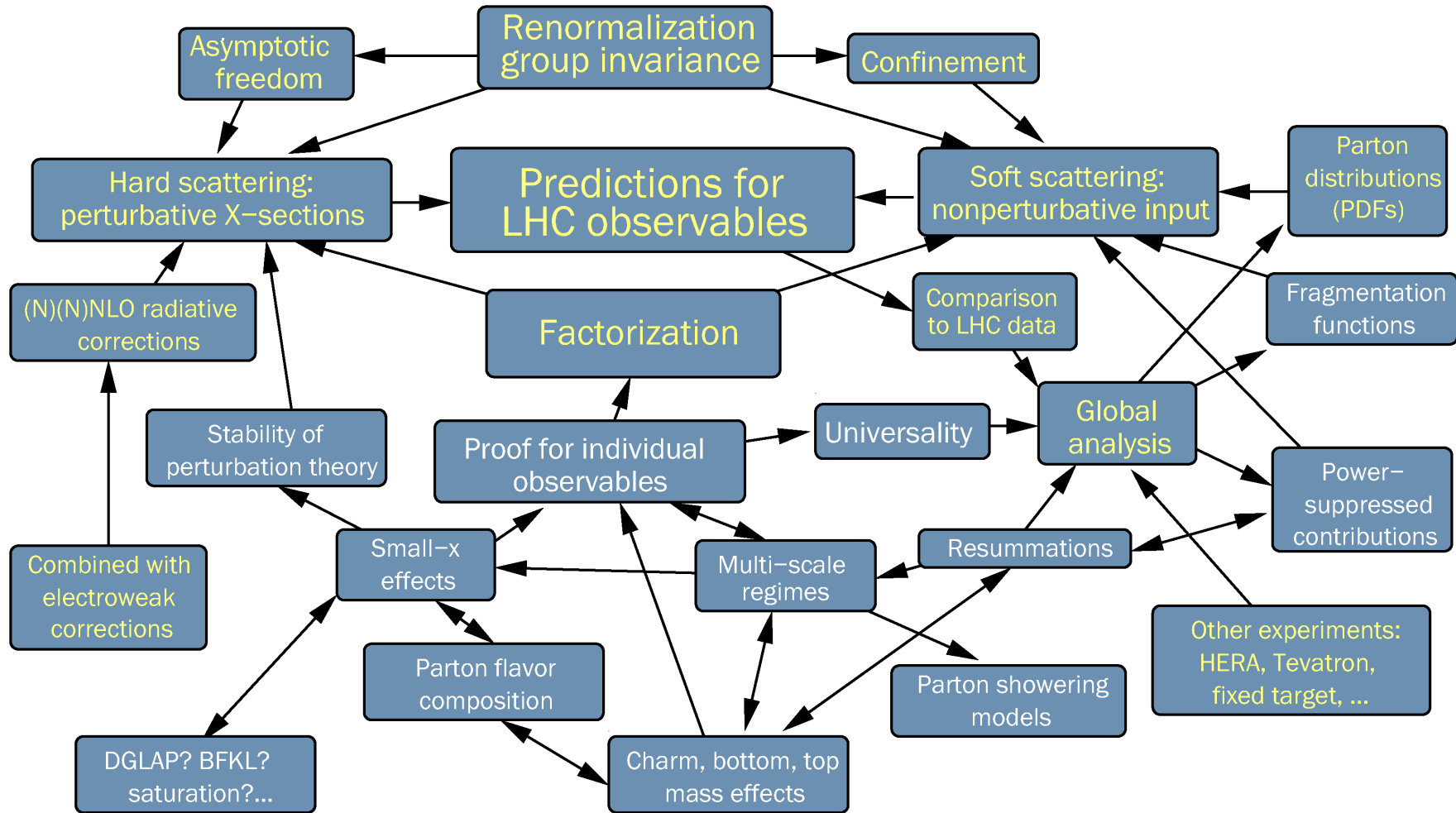
The result showcases the power of the LHC to push the precision frontier and improve our understanding of nature

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 $\Delta\chi^2 = 1 \Rightarrow$ Underestimated uncertainties \Rightarrow Non-replicable result



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

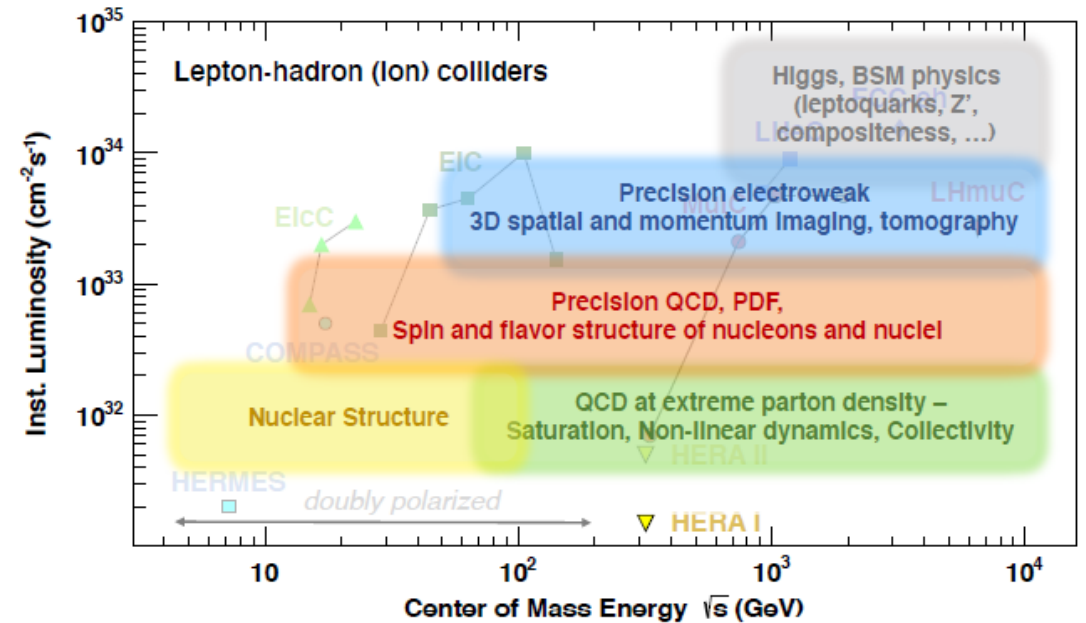
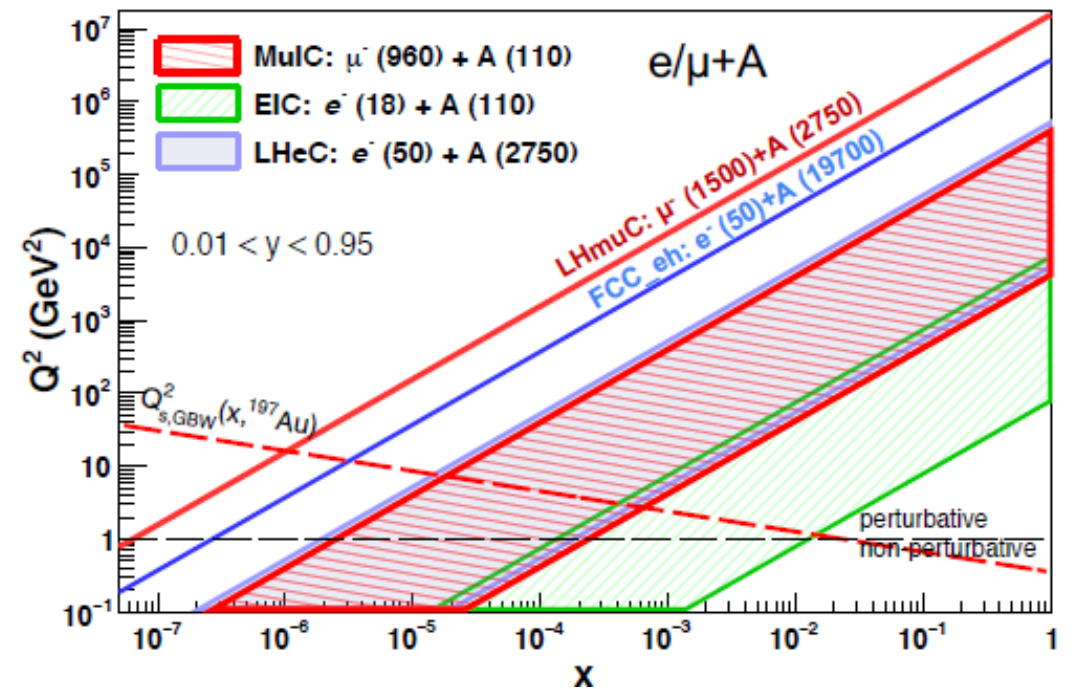
Parton distribution functions in future DIS experiments

Prospects

- Expansive reach of EIC/MuIC precision experiments in x and Q for μp and μA DIS
- In-depth SM studies and BSM searches
- 3-dimensional tomography of multiple hadron species

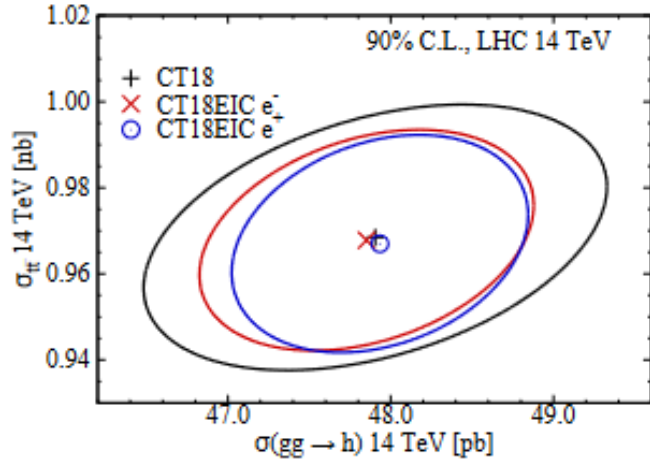
Challenges

- Cross-disciplinary nature of QCD phenomena at EIC/MuIC
- Reproducibility and replicability of future results

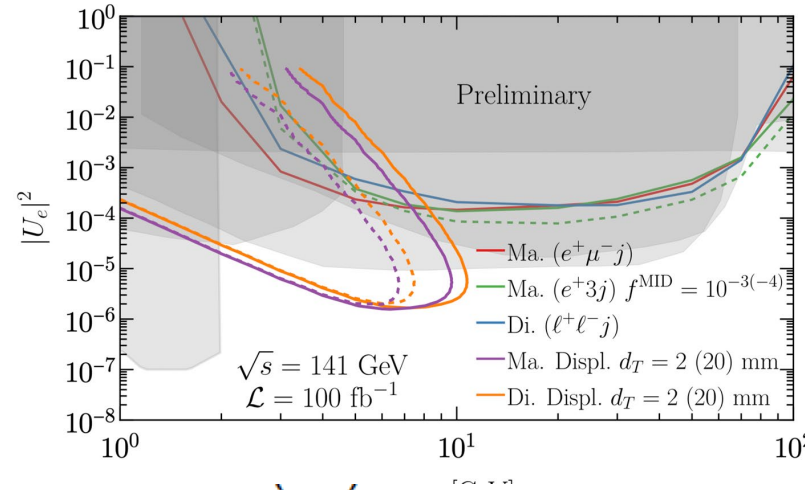


Electron-Ion Collider: potentially a wealth of complex studies

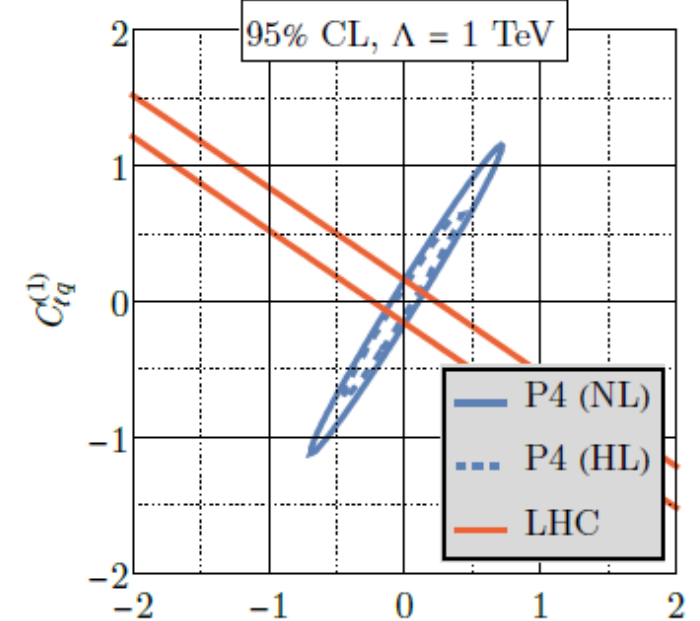
PDFs: arXiv:2103.05419



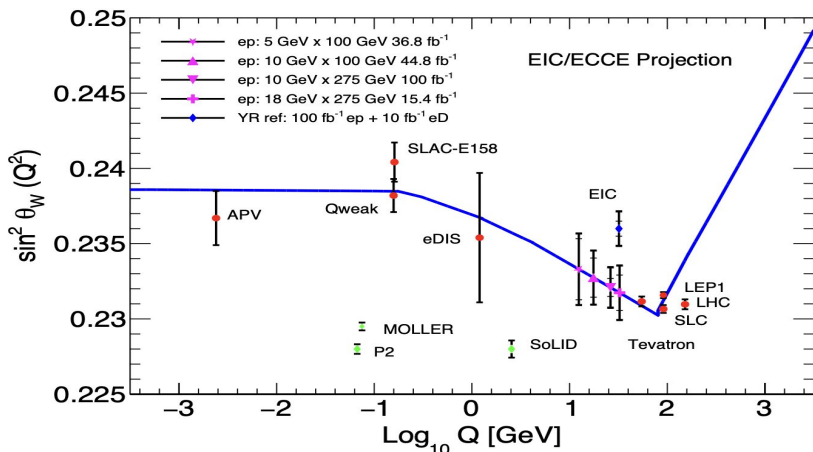
heavy neutral lepton searches arXiv: 2203.06705



SMEFT Wilson coefficients



weak mixing angle arXiv: 2203.13199



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

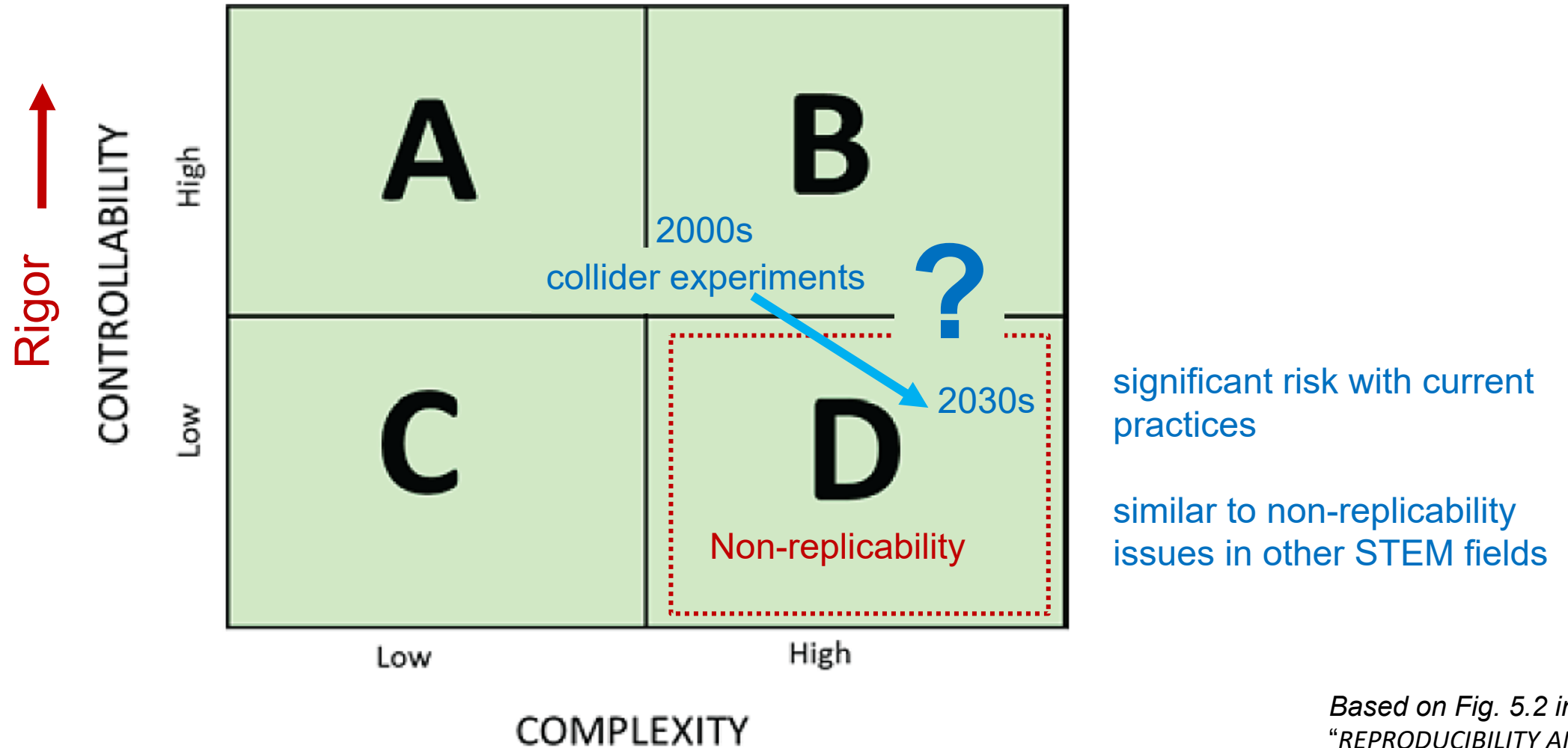
Boughezal et al
2004.00748, .2204.07557

Lorentz/CPT violations

A. R. Vieira et al., 1911.04002

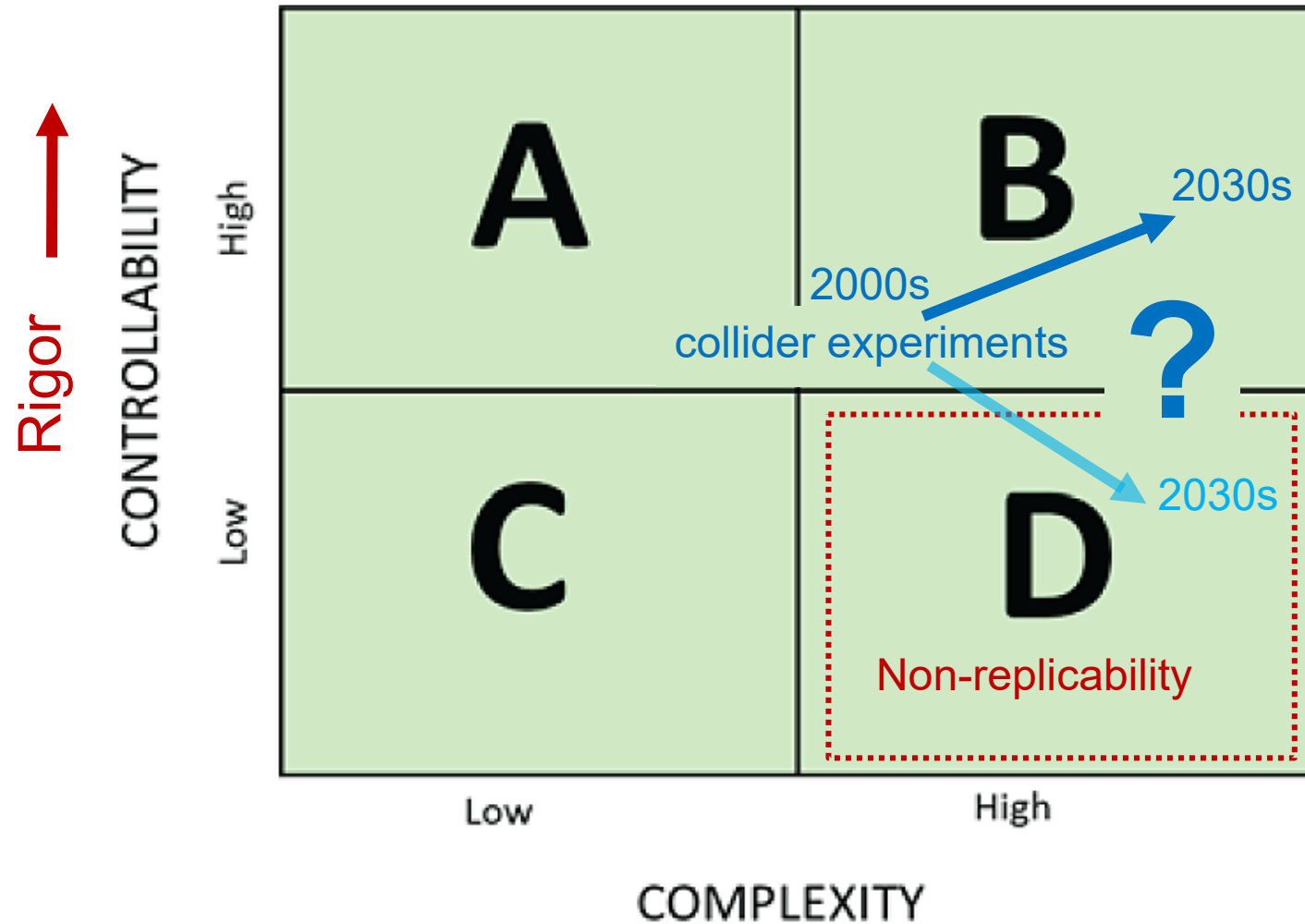
Abdul-Khalek et al., Snowmass 2021 whitepaper
"EIC for HEP", [2203.13199](https://arxiv.org/abs/2203.13199)

Future scenarios for QCD precision analysis



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

Future scenarios for QCD precision analysis



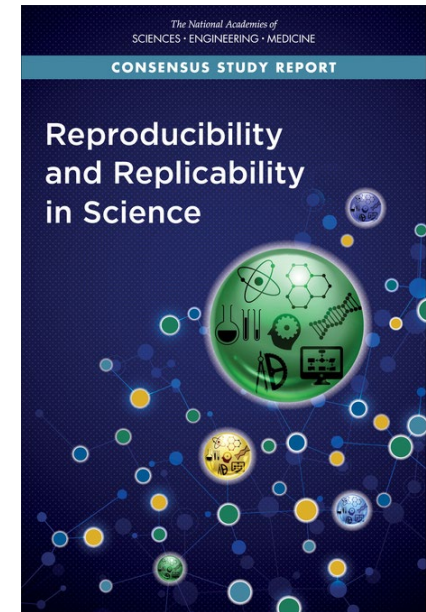
preferred scenario; requires a coordinated community strategy to adopt the **replicability mindset**

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

Universal factors affecting replicability

- complexity of the system under study;
- understanding of the number and relations among variables within the system under study;
- ability to control the variables;
- levels of noise within the system (or signal to noise ratios);
- mismatch of scale of the phenomena and the scale at which it can be measured;
- stability across time and space of the underlying principles;
- fidelity of the available measures to the underlying system under study (e.g., direct or indirect measurements);
- prior probability (pre-experimental plausibility) of the scientific hypothesis.

From “*REPRODUCIBILITY AND REPLICABILITY IN SCIENCE*”
National Academy of Sciences, 2019, <https://doi.org/10.17226/25303>



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

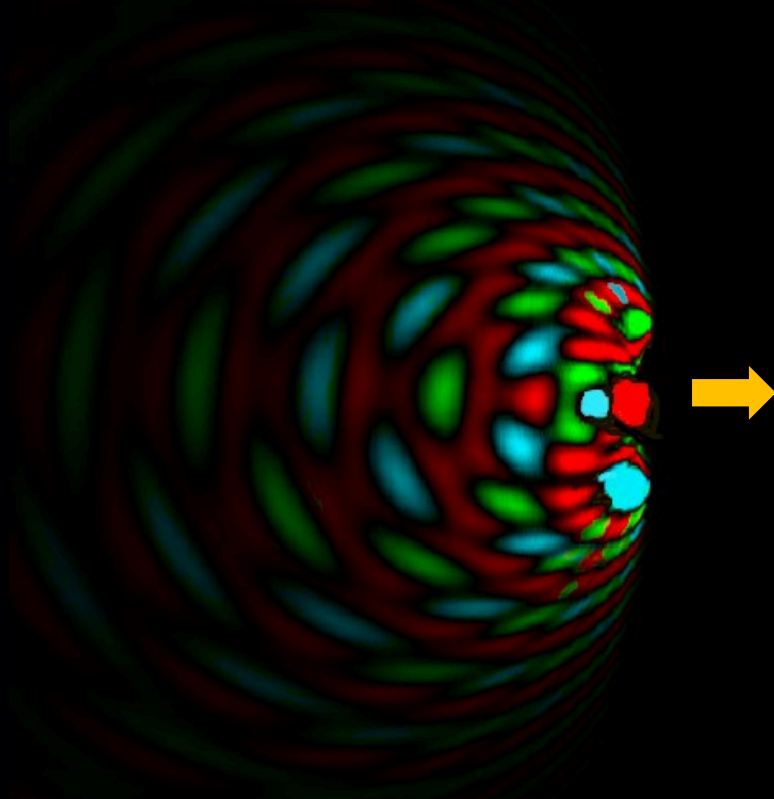
Support from the funding agencies for the research infrastructure and collaborations focusing on replicability

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

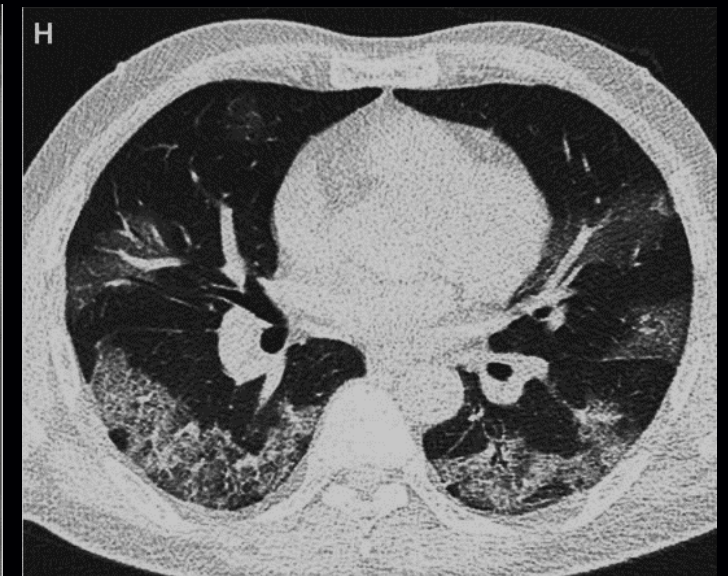
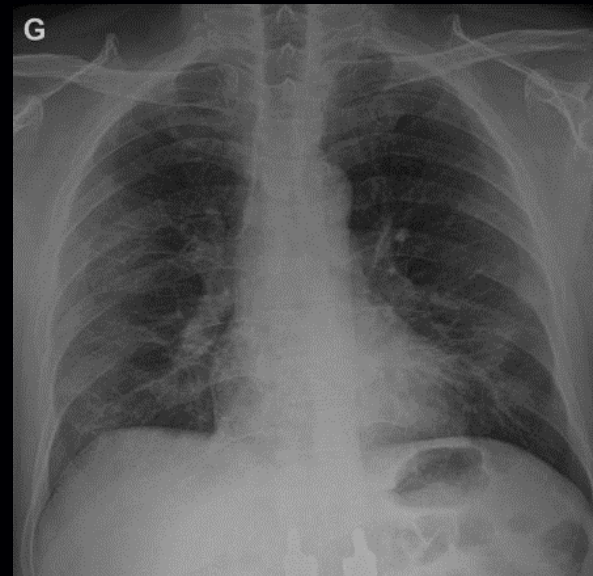
A community model of the replicability mindset

Particle physicists and radiology doctors infer from complex images



A proton at an *ep* collider moving with speed $V \approx c$ to the right

3-dim hadron femtography at the EIC



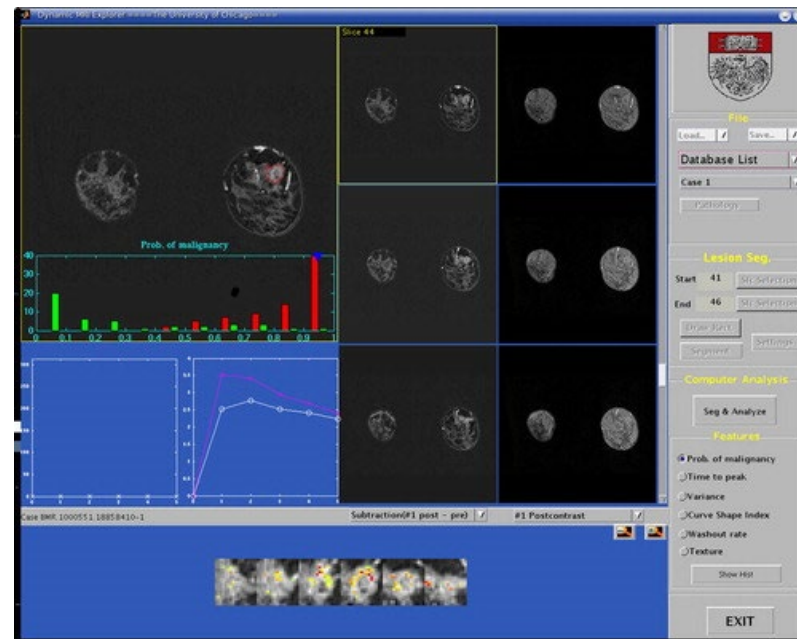
A 3-dim tomographic image of a COVID-19 patient

Particle physicists and radiology doctors address analogous questions in statistics

> Radiology. 2011 Mar;258(3):696-704. doi: 10.1148/radiol.10100409. Epub 2011 Jan 6.

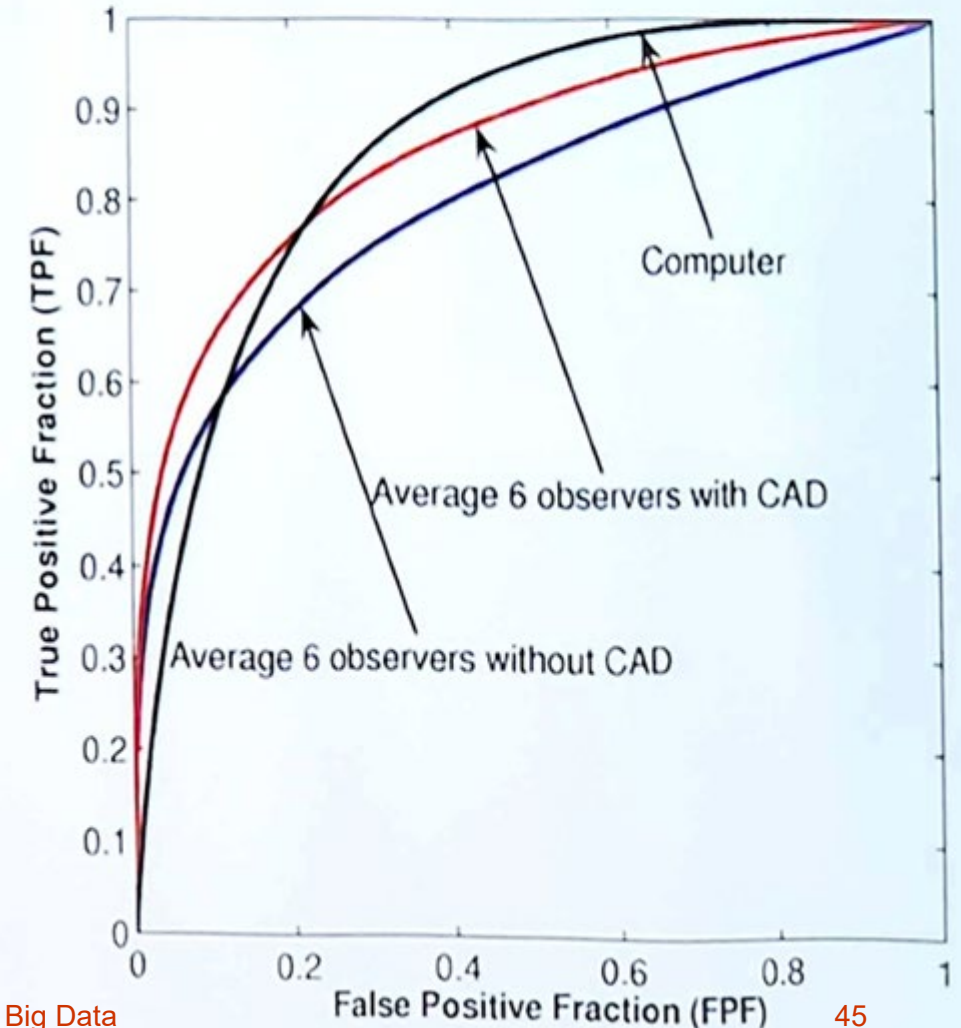
Evaluation of clinical breast MR imaging performed with prototype computer-aided diagnosis breast MR imaging workstation: reader study

Akiko Shimauchi¹, Maryellen L Giger, Neha Bhooshan, Li Lan, Lorenzo L Pesce, John K Lee, Hiroyuki Abe, Gillian M Newstead



CADx: Task of Distinguishing between Malignant & Benign Lesions on **Breast MRI**

Performance of the End User



M. Giger (U Chicago), FNAL,
July 20, 2023

2024-06-13

P. Nadolsky, QCD at the Femtoscale in the Era of Big Data

AI and replicability in radiology

“Within health care, the US Food and Drug Administration has already cleared 523 devices that use AI—75% of them for use in radiology.”

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

Statistical inference from hadron scattering data and medical images bear many similarities. The medical community working on AI is very large and well-funded.

During the COVID-19 pandemic, thousands of medical AI applications were developed to diagnose and cure the disease. Most have failed.

In response to this replicability crisis, the US medical community took numerous actions to implement systemwide infrastructure, standards, and procedures for organizing the data and quantifying uncertainties in AI-assisted analyses.

What went wrong with AI/ML methods for COVID imaging?

MIT
Technology
Review

Featured Topics Newsletters Events Podcasts

Artificial intelligence / Machine learning

400+

Hundreds of AI tools have been built to catch covid. None of them helped.

Some have been used in hospitals, despite not being properly tested. But the pandemic could help make medical AI better.

nature machine intelligence

Explore content ▾ About the journal ▾ Publish with us ▾

nature > nature machine intelligence > analyses > article

Analysis | Open Access | Published: 15 March 2021

Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans

Michael Roberts , Derek Driggs, Matthew Thorpe, Julian Gibbey, Michael Yeung, Stephan Ursprung, Angelica I. Aviles-Rivero, Christian Etmann, Cathal McCague, Lucian Beer, Jonathan R. Weir-McCall, Zhongzhao Teng, Effrossyni Gkrania-Klotsas, AIX-COVNET, James H. F. Rudd, Evis Sala & Carola-Bibiane Schönlieb

Nature Machine Intelligence 3, 199–217 (2021) | Cite this article

1) Poor quality of COVID imaging data

- Mislabeled data
- Multiple unknown sources
- Duplicate data (training and testing)
- No traceability, limited quality control
- Lack of external validation

2) Lack of collaboration/communication between AI/ML experts and biomedical experts

- Need for valid ground truth
- Need for independent test set

3) Bias and diversity

- Data collected for a specific clinical task
- Specific populations, lack of diversity
- Single expert score, data sources correlated with 'truth', ...



MIDRC

MEDICAL IMAGING AND DATA RESOURCE CENTER.

<https://www.midrc.org/>

Medical Imaging Community in response to the COVID-19 pandemic

M. Giger (U Chicago), FNAL, July 20, 2023



Medical Imaging and Data Resource Center

Established August 2020



NIBIB COVID-19 Contract 75N92020D00021

- An open, curated, diverse image data commons
- A partnership between the AAPM, ACR and RSNA, supported by NIBIB, hosted at University of Chicago, and on the Gen3 data platform

Two scientific components of MIDRC:

1. Open Discovery Data Commons
2. Machine Intelligence Computational Capabilities

3. The center uses a private subset of data to validate statistical rigor and replicability of the proposed (AI-assisted or not) algorithms

MIDRC by the Numbers

309,270

Imaging Studies
Ingested

152,772

Imaging Studies
released to the
Public

156,498

Imaging Studies
undergoing quality &
harmonization

377

Total Data
Downloads
this month

60,801

Cases

13.27 TB

Total size
Published

47

Publications

115+

Presentations

29

Algorithms

586

Registered
Users

100+

Investigators

416

Collaborating
Institutions

To date, MIDRC has focused on medical imaging and data of COVID-19 patients, and the imaging studies made available to the public have mainly been chest imaging. Currently, however, in order to keep up with developments in the pandemic, imaging studies associated with post acute sequelae of COVID-19 (PASC, 'long COVID') are actively being collected as well as different imaging modalities and various additional organ systems (such as the heart or brain).

Future plans include expansion to become a wider comprehensive resource for all Institutes at NIH, with focused medical imaging data commons of chronic disease (e.g., diabetes, chronic liver disease, coronary artery disease, COPD and emphysema), and other infectious pandemics. Additionally, the sequestering of some of the MIDRC data in a separate data commons not accessible to the public for validation and testing will provide a valuable resource for data science challenges and a path to long-term sustainability through industry support for translation to - and approval of - clinical use which will impact public health worldwide. [Learn more...](#)

<https://www.midrc.org/>, accessed on 2023-09-17

New HL-LHC and DIS experiments open unique opportunities to understand nonperturbative QCD. They also create a strong synergistic effect in both SM and BSM studies

Progress on this program, especially in precision measurements, increasingly depends on cross-cutting research and replicability of complex measurements

Precision QCD results must be replicable. The experience from radiology and other fields suggests community-wide strategies for achieving it.



Quantification of PDF uncertainties

Contents

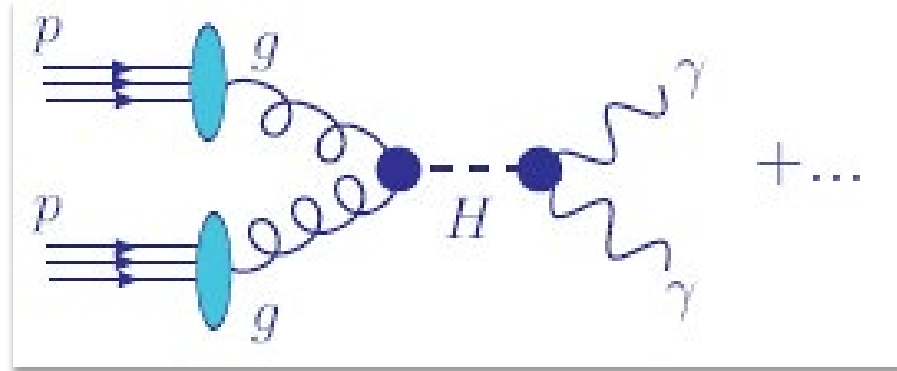
Week 1

1. Dynamic images of hadrons in the AI era
2. Collinear PDFs, their applications and determinations
3. **RRR**: rigor, reproducibility, replicability in the PDF analysis

Week 4

4. Uncertainty quantification on parton distributions
 - Tolerance puzzle
 - Less known aspects of multivariate statistics

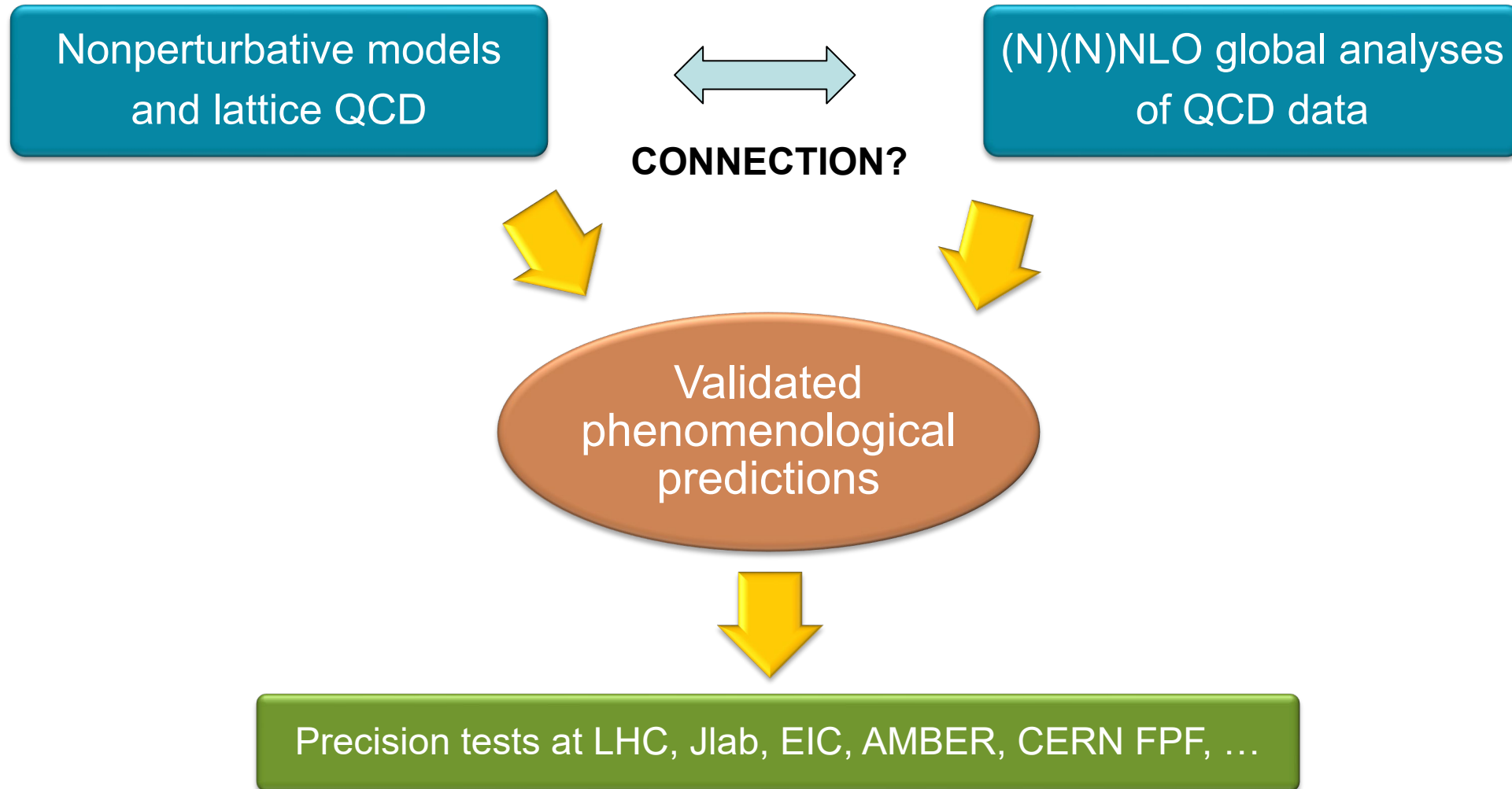
Parton distributions describe long-distance dynamics in high-energy collisions



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

$\hat{\sigma}$ 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

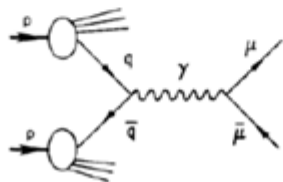
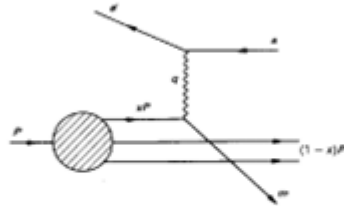
New insights about 3-dimensional structure of hadrons



PDFs in nonperturbative QCD

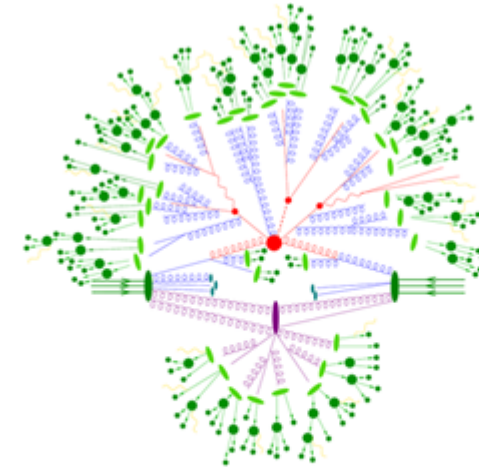
Relevant for processes
at $Q^2 \approx 1 \text{ GeV}^2$?

⇒ 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 $Q^2 \gg 1 \text{ GeV}^2$



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

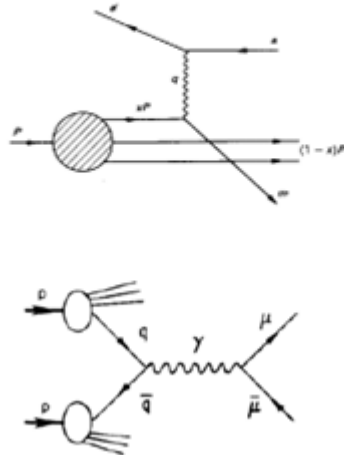
How to relate the x dependence of the perturbative and nonperturbative pictures?

Does the evidence from primordial dynamics survive PQCD radiation?

PDFs in nonperturbative QCD

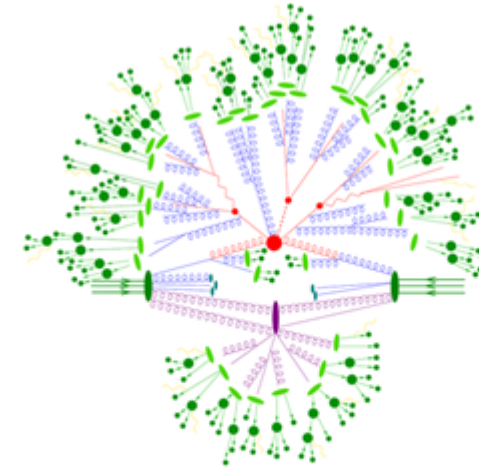
Relevant for processes
at $Q^2 \approx 1 \text{ GeV}^2$

⇒ 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 $Q^2 \gg 1 \text{ GeV}^2$



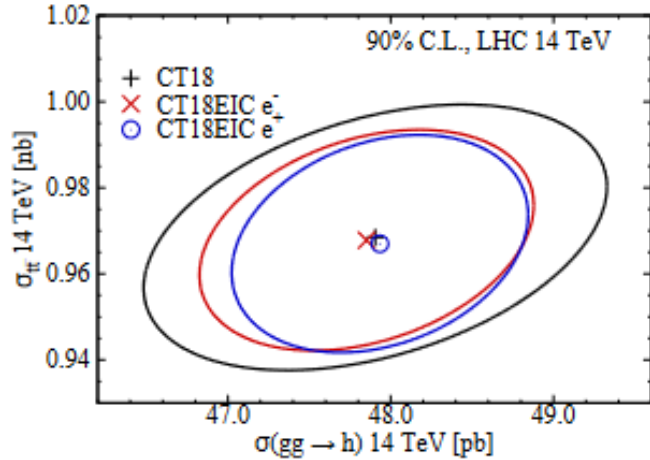
⇒ pheno PDFs are determined from analyzing many processes with complex scattering dynamics

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

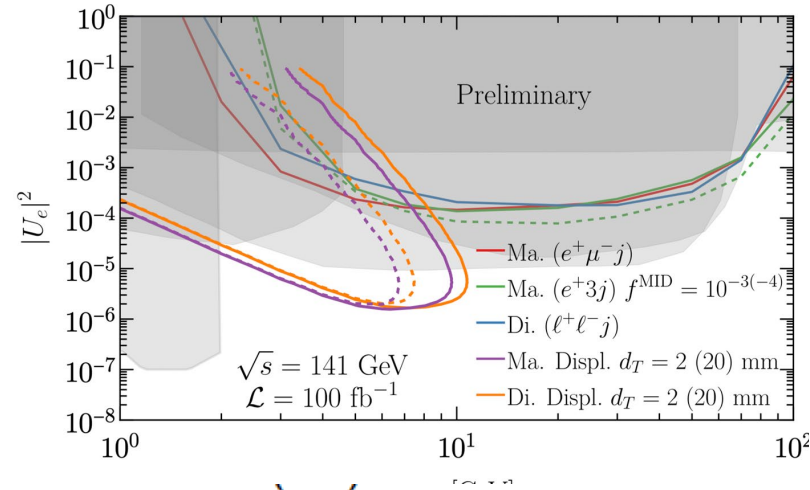
Does the evidence from primordial dynamics survive PQCD radiation?

Electron-Ion Collider: potentially a wealth of complex studies

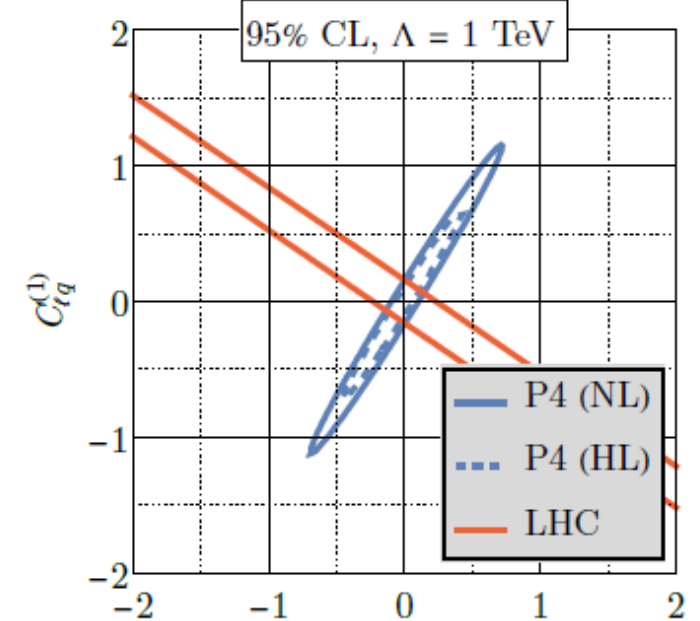
PDFs: arXiv:2103.05419



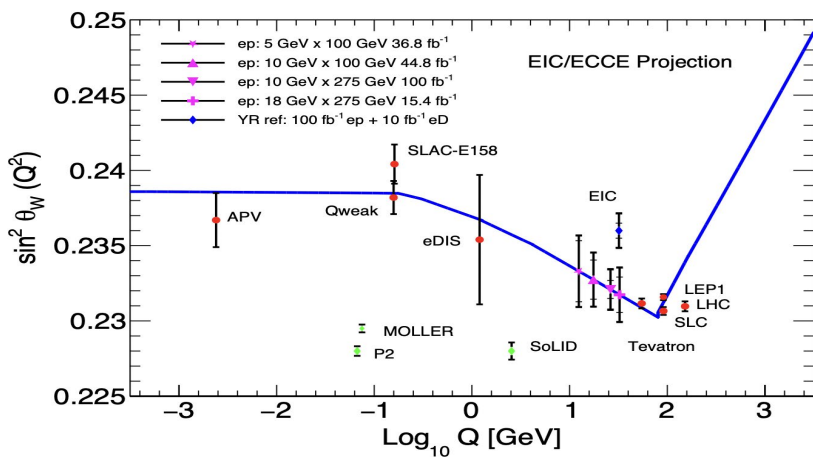
heavy neutral lepton searches arXiv: 2203.06705



SMEFT Wilson coefficients



weak mixing angle arXiv: 2203.13199

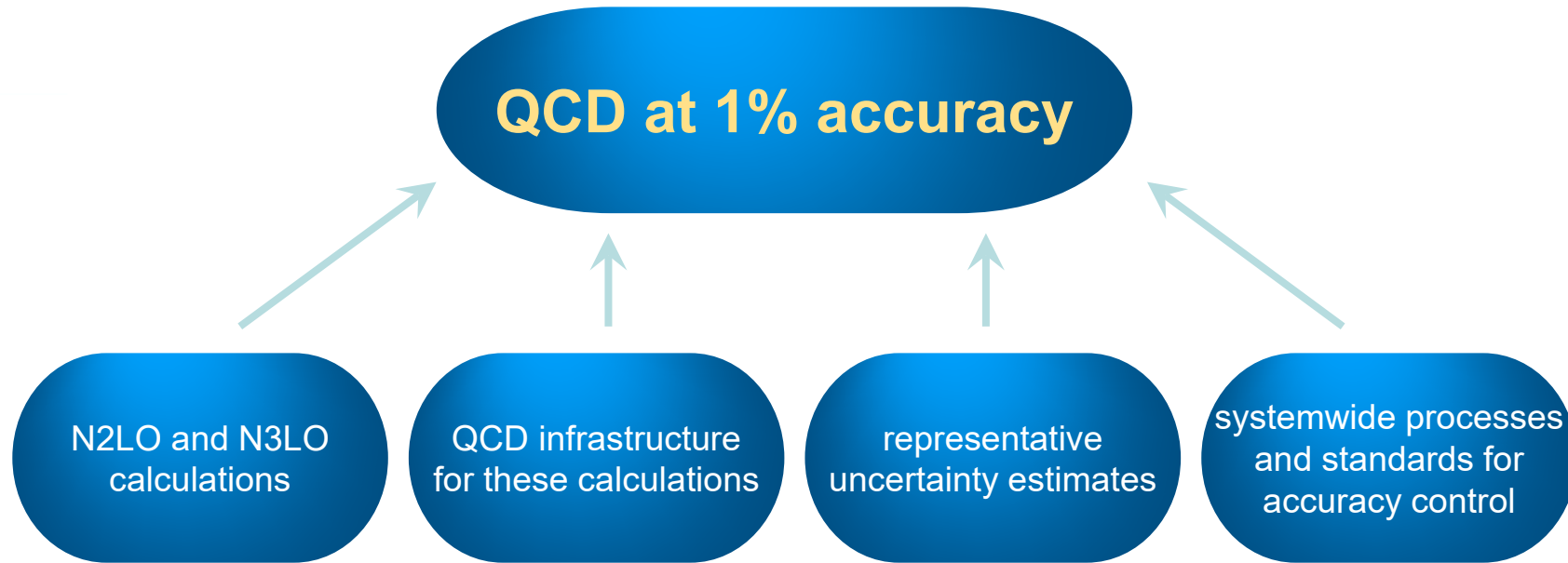


	EIC	LHC
$ c_u^{XX} - c_u^{YY} $	0.37	15
$ c_u^{XY} $	0.13	2.7
$ c_u^{XZ} $	0.11	7.3
$ c_u^{YZ} $	0.12	7.1
$ a_{Su}^{(5)TXX} - a_{Su}^{(5)TYY} $	2.3	0.015
$ a_{Su}^{(5)TXY} $	0.34	0.0027
$ a_{Su}^{(5)TXZ} $	0.13	0.0072
$ a_{Su}^{(5)TYZ} $	0.12	0.0070

Boughezal et al
2004.00748, .2204.07557

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

Abdul-Khalek et al., Snowmass 2021 whitepaper
"EIC for HEP", [2203.13199](https://arxiv.org/abs/2203.13199)



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)

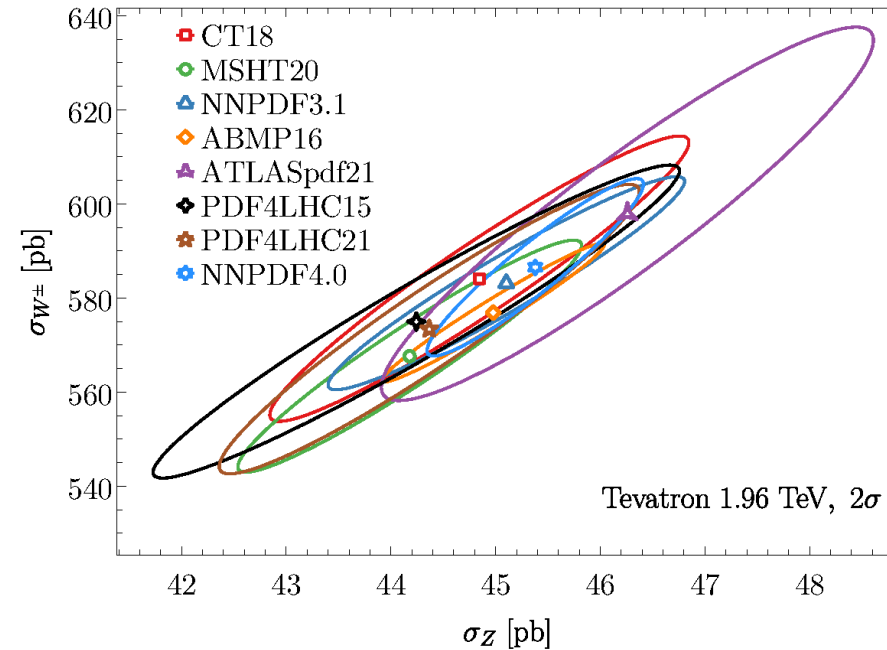
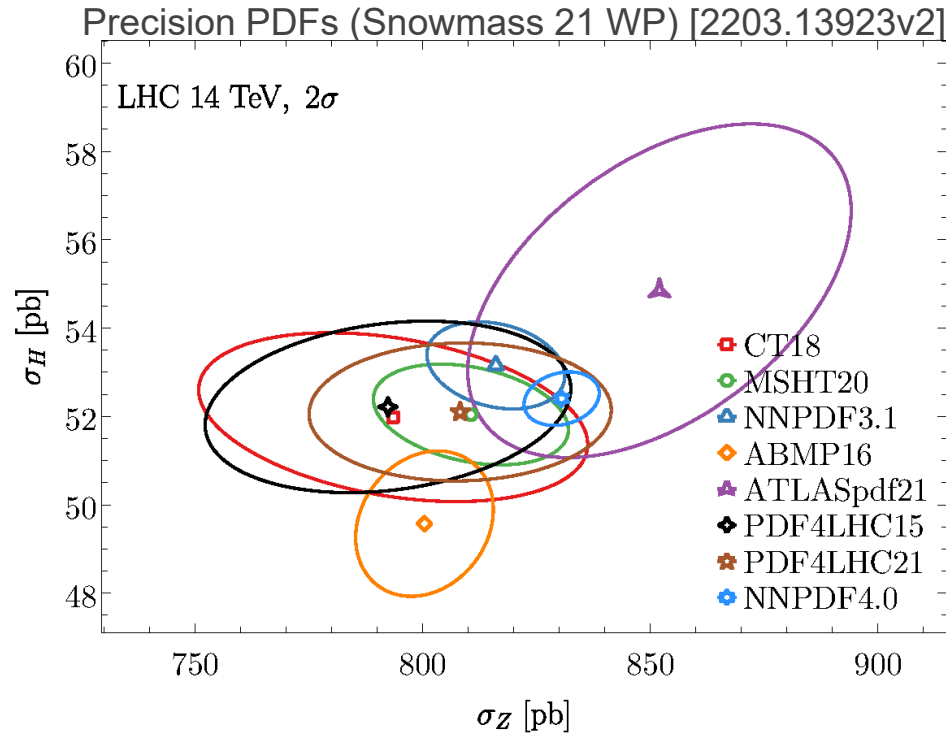
This must be a part of the precision-focused community culture

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

The tolerance puzzle

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



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

Comparisons of the latest PDF sets

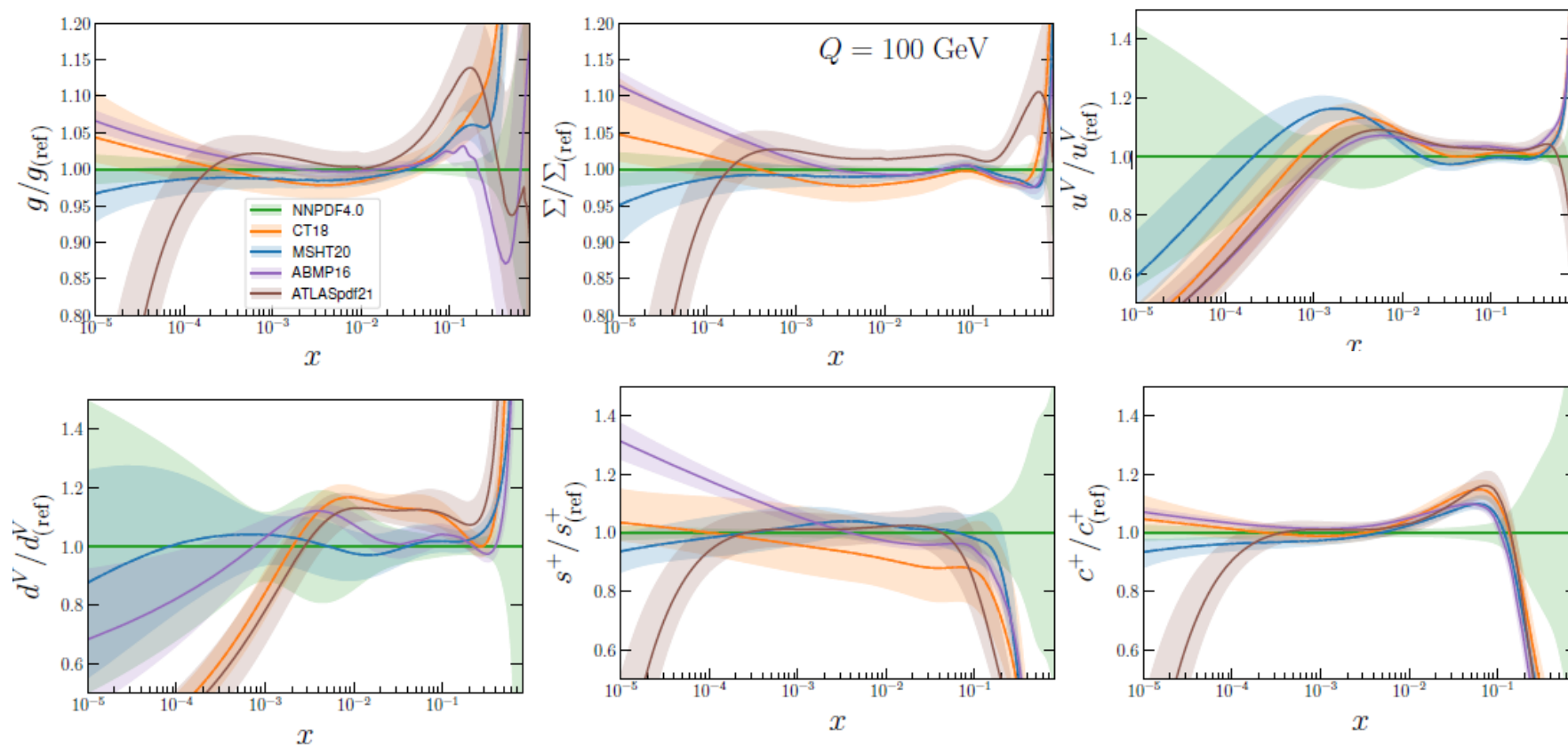
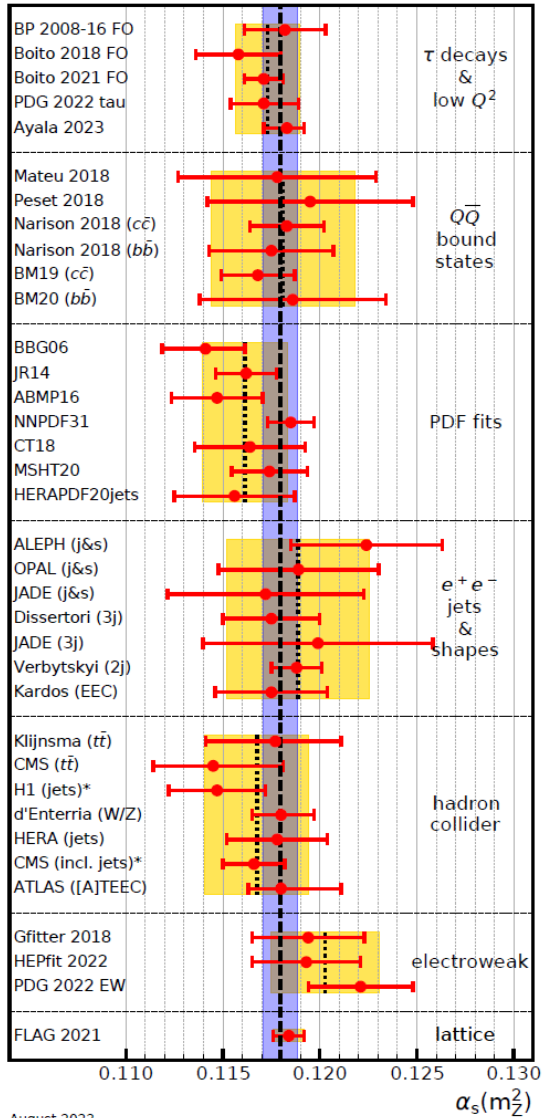
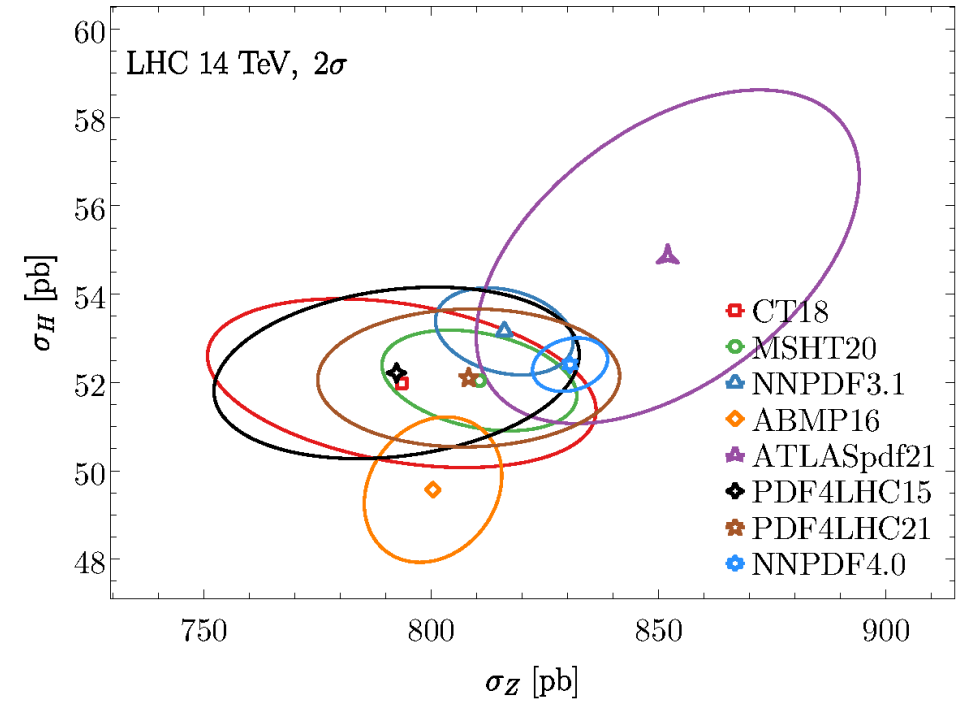


FIG. 2. Comparison of the PDFs at $Q = 100$ GeV. The PDFs shown are the N2LO sets of NNPDF4.0, CT18, MSHT20, ABMP16 with $\alpha_s(M_Z) = 0.118$, and ATLASpdf21. The ratio to the NNPDF4.0 central value and the relative 1σ uncertainty are shown for the gluon g , singlet Σ , total strangeness $s^+ = s + \bar{s}$, total charm $c^+ = c + \bar{c}$, up valence u^V and down valence d^V PDFs.

Replicability and PDF uncertainties

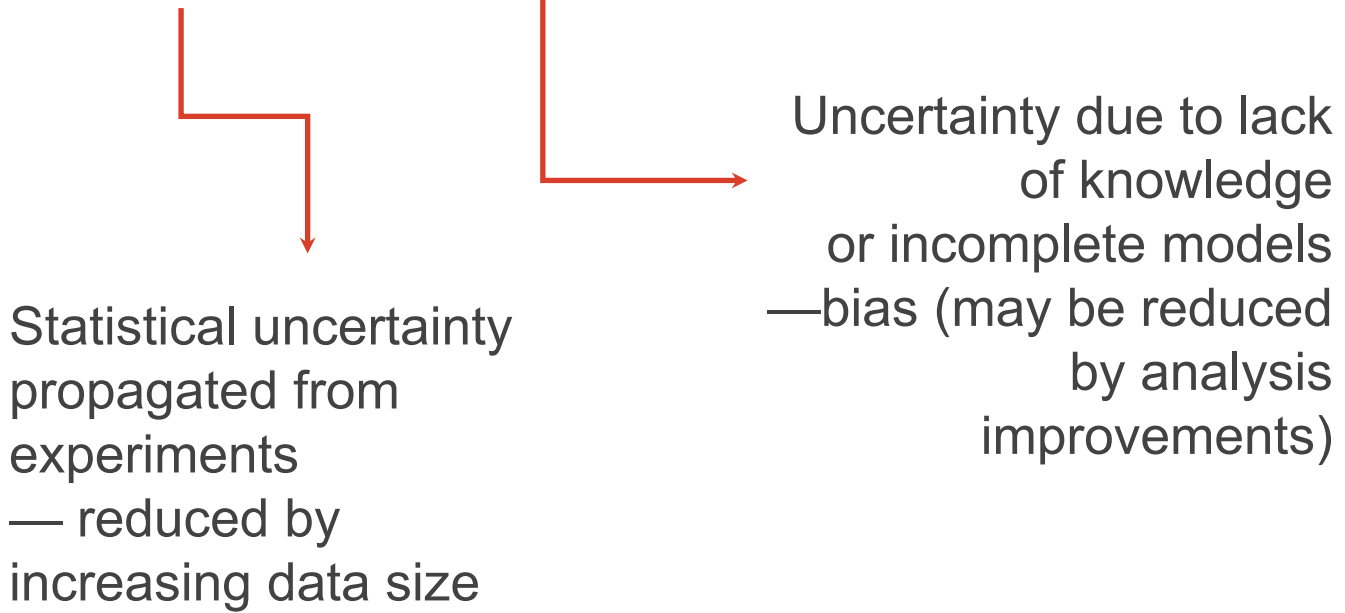


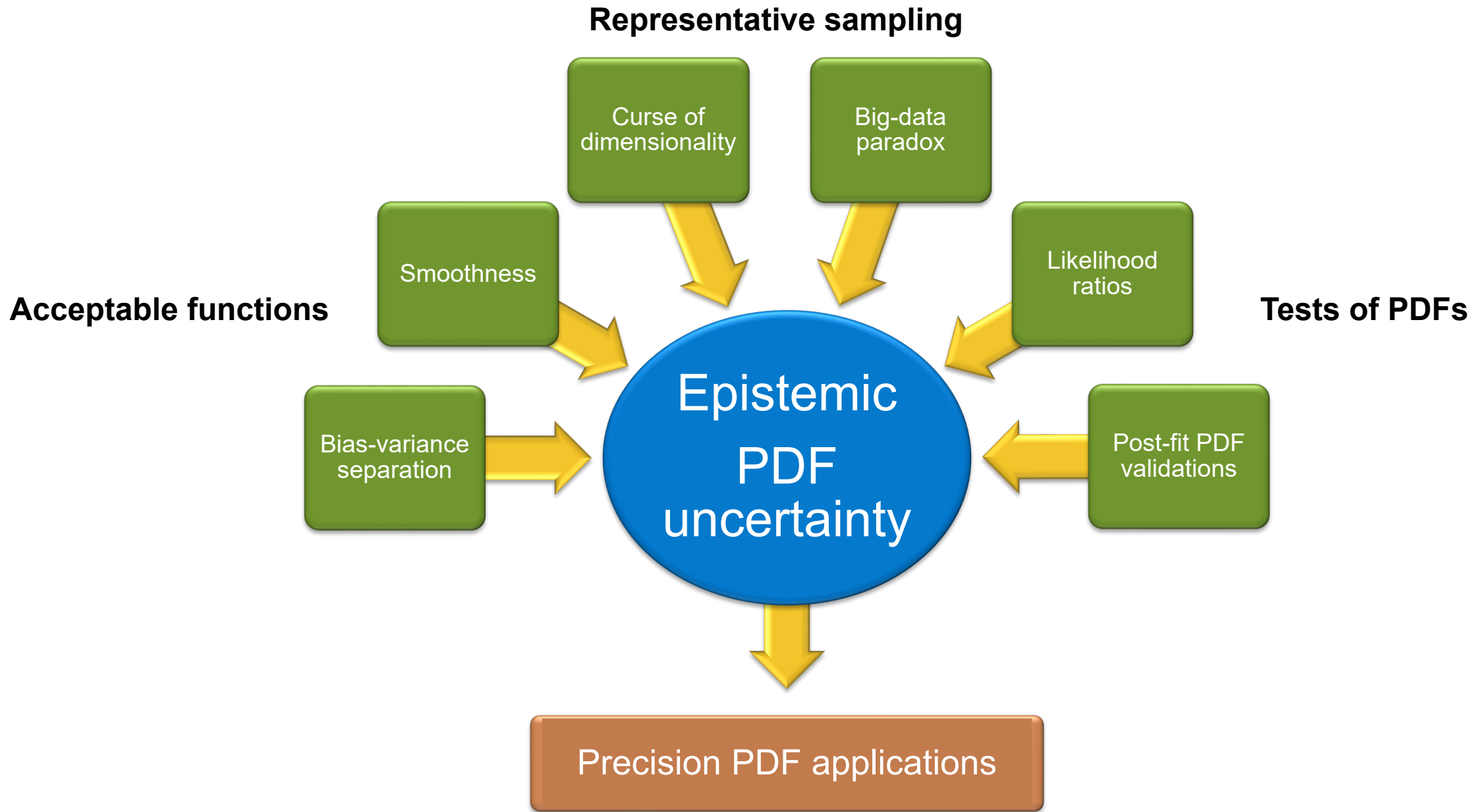
August 2023



Quantification of **epistemic** PDF uncertainties is a central factor affecting **replicability** of upcoming determinations of the QCD coupling constant α_s , Higgs couplings, mass of weak bosons.

aleatory vs. epistemic uncertainties





Epistemic PDF uncertainty...

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

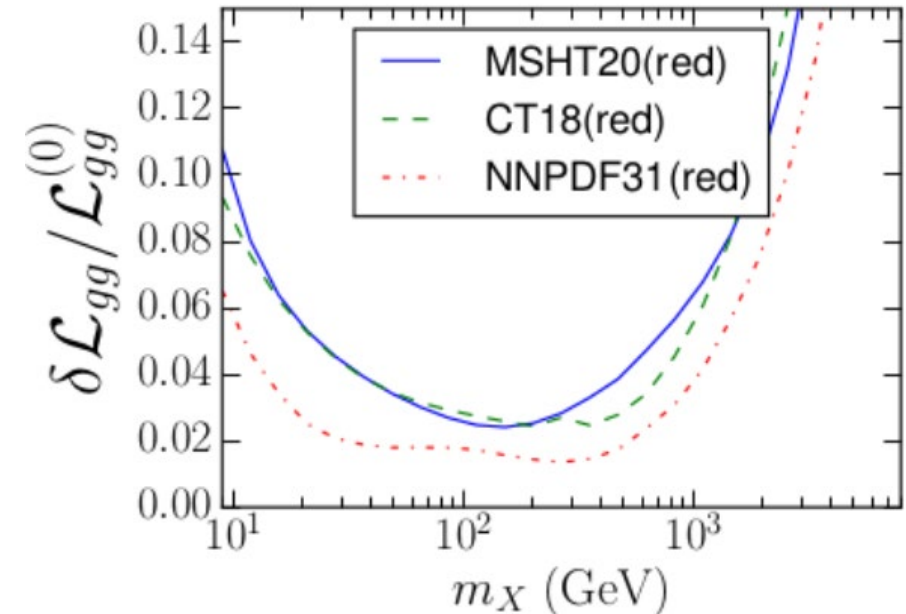
... can dominate the full uncertainty when experimental and theoretical uncertainties are small.

...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...

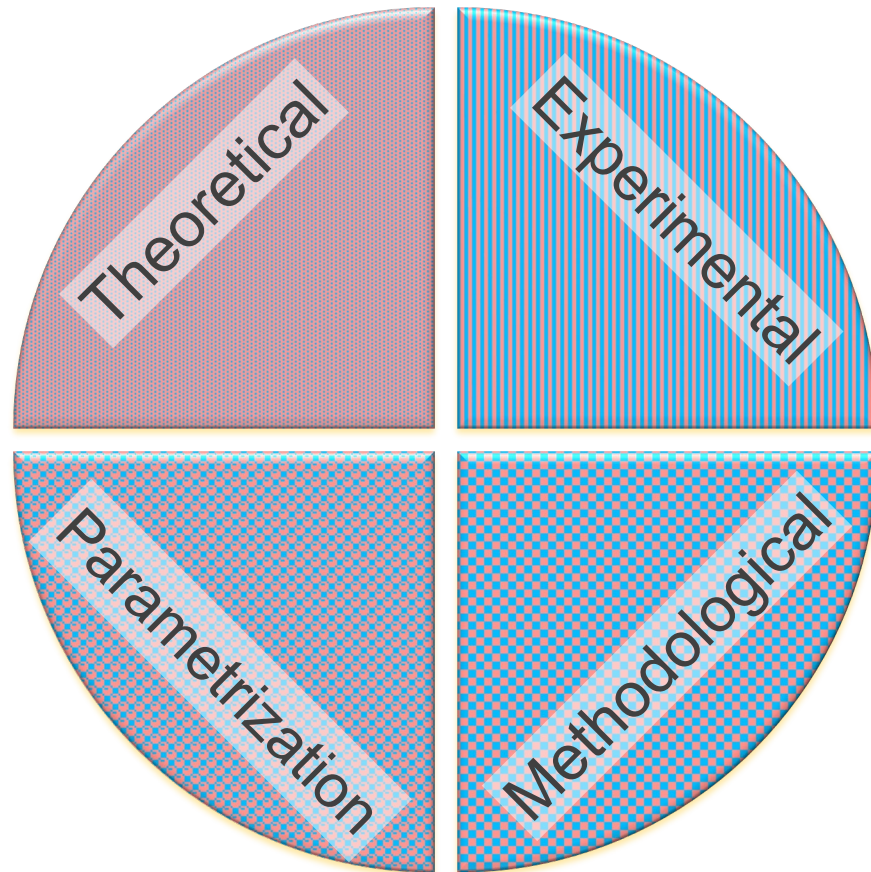
⇒ in addition to sampling over data fluctuations




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](https://arxiv.org/abs/2203.05506), [arXiv:2205.10444](https://arxiv.org/abs/2205.10444)

Components of PDF uncertainty



In each category, one must maximize

 **PDF fitting accuracy**
(accuracy of experimental, theoretical and other inputs)

 **PDF sampling accuracy**
(adequacy of sampling in space of possible solutions)

Fitting/sampling classification is borrowed from the statistics of large-scale surveys [Xiao-Li Meng, *The Annals of Applied Statistics*, Vol. 12 (2018), p. 685]

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,...

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 $1/\sqrt{N_{\text{rep}}}$, 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**

Ongoing studies of systematic uncertainties are essential and still insufficient

- from the experiment side

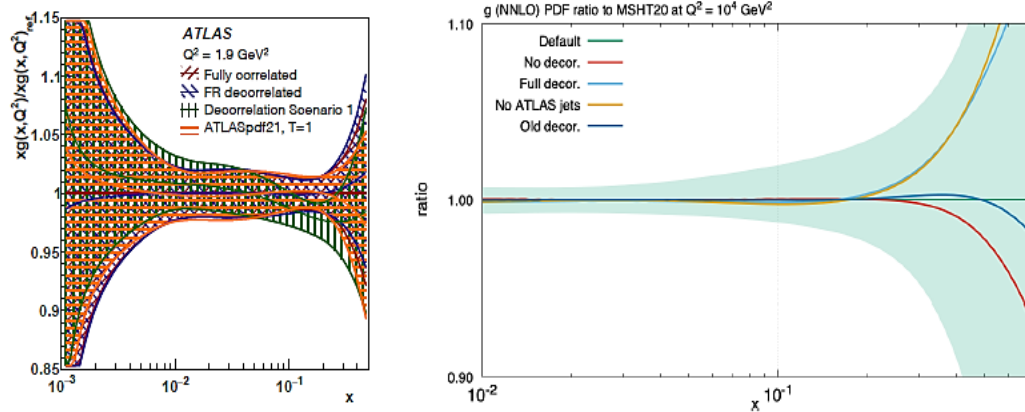


FIG. 9. Difference in the gluon PDF shown in ratio to the ATLASpdf21 (default) gluon (left). This default uses Decorrelation Scenario 2 and this is compared to the use of Full Correlation, Full decorrelation of the flavour response systematic and Decorrelation Scenario 1. The effect of no decorrelation,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

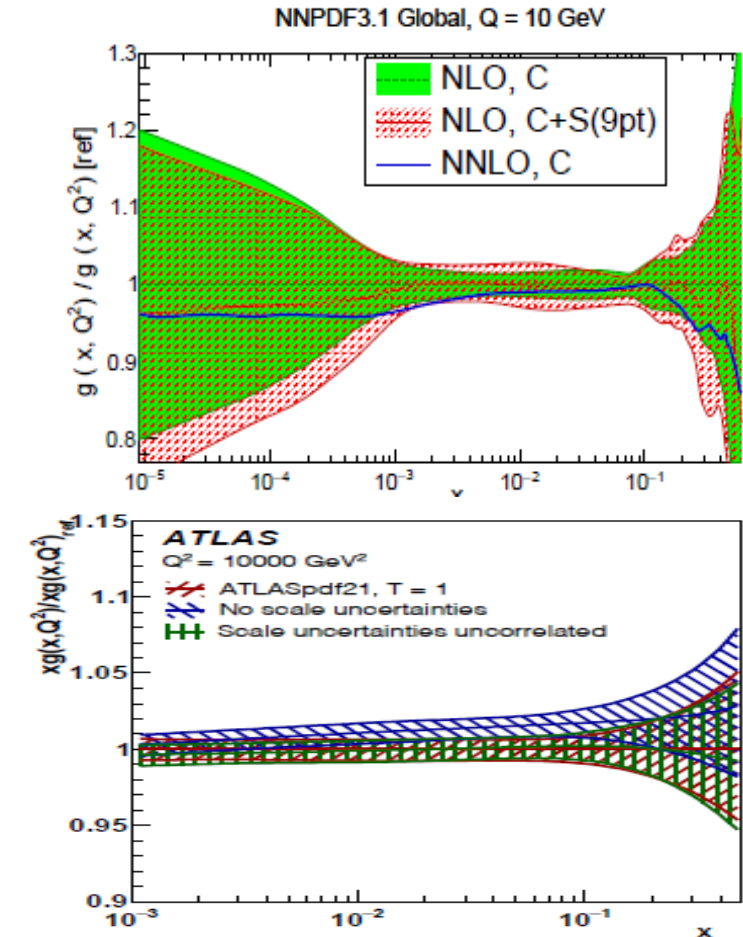
S. Amoroso et al., 2203.13923, Sec. 5.A

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 $\lambda_{\alpha,exp}$

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

2. In terms of the covariance matrix

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

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

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

$$\beta_{i,\alpha} = \sigma_{i,\alpha} X_i$$

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:

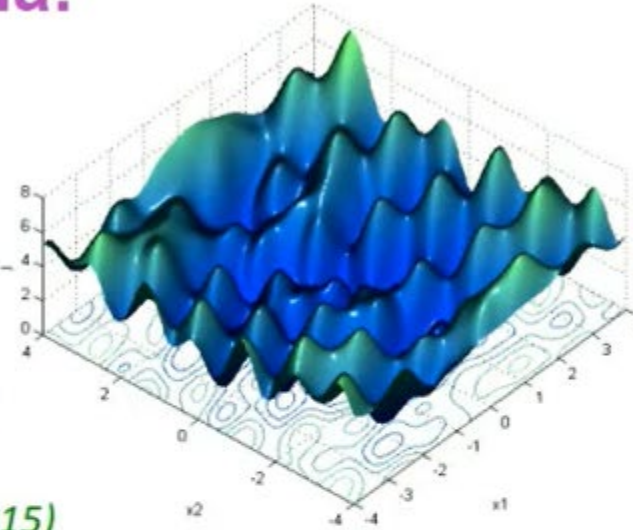
- $X_i = D_i$: “**experimental scheme**”; can result in a bias
- $X_i = \text{fixed or varied } T_i$: “ **t_0, T , extended T schemes**”; can result in (different) biases

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)

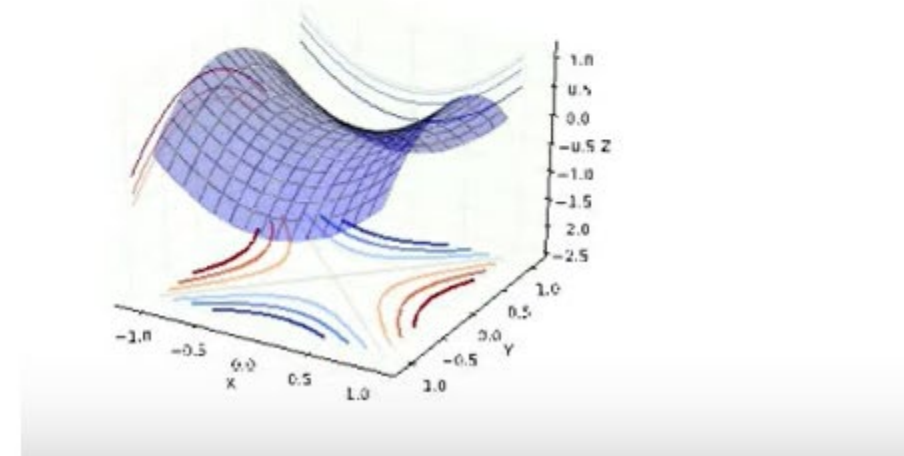
(Dauphin et al NIPS'2014, Choromanska et al AISTATS'2015)



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...

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.

The Loss Surfaces of Multilayer Networks

A. Choromanska, M. Henaff, M. Mathieu, G. Ben Arous, Y. LeCun PMLR 38:192-204, 2015

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

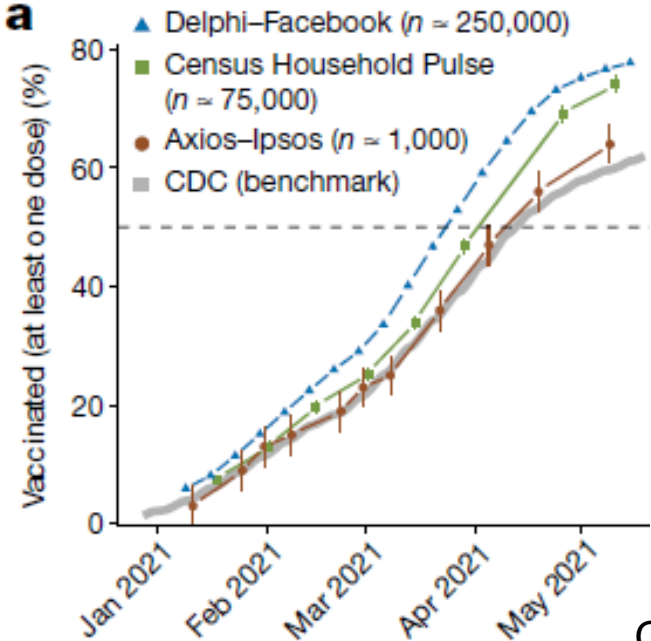
Check for updates

Valerie C. Bradley^{1,2}, Shiro Kuriwaki^{2,3}, Michael Isakov³, Dino Sejdinovic³, Xiao-Li Meng⁴ & Seth Flaxman⁵✉

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 Paradox¹. 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 Pulse⁴ (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 minuscule margins of error on the incorrect estimates. By contrast, an Axios–Ipsos online panel⁵ with about 1,000 responses per week following survey research best practices⁶ provided reliable estimates and uncertainty quantification. We decompose observed error using a recent analytic framework⁷ 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.

Many dimensions introduce major difficulties with finding a global minimum...

...as well as with representative exploration of uncertainties



Nature v. 600 (2021) 695

Courtoy et al., PRD 107 (2023) 034008

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 N_{par} (hyper)parameters on a randomly fluctuated replica of discrete data D_i . Repeat N_{rep} times. In a typical application: $N_{\text{par}} > 10^2$, $N_{\text{rep}} < 10^4$.
2. Out of 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 [$r_{\text{likelihood}} = \exp\left(\frac{\chi_1^2 - \chi_2^2}{2}\right) = 1$], 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.

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 $\forall \chi^2$ function grows as $N_{min} \geq 2^{N_{par}}$
 - Example: $N_{min} > 10^{30}$ for $N_{par} = 100$

[Sloan, Woźniakowski, 1997]

[Hickernell, MCQMC 2016, 1702.01487]

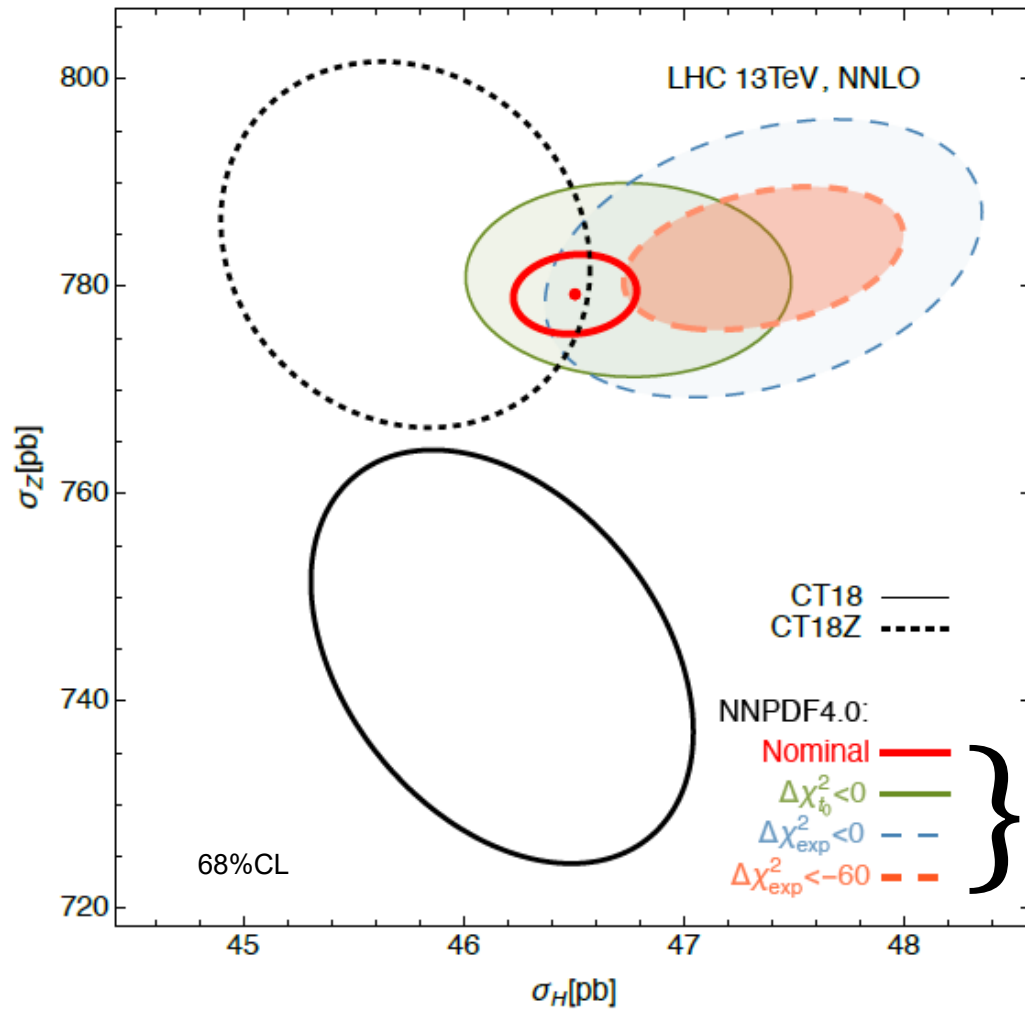
Good news: expectation values for **typical QCD observables** can be estimated with fewer replicas by reducing dimensionality of the problem or a targeted sampling technique.

Example: a “**hopscotch scan**”, see 2205.10444



Example: the impact of epistemic uncertainty on NNLO Higgs and Z cross sections

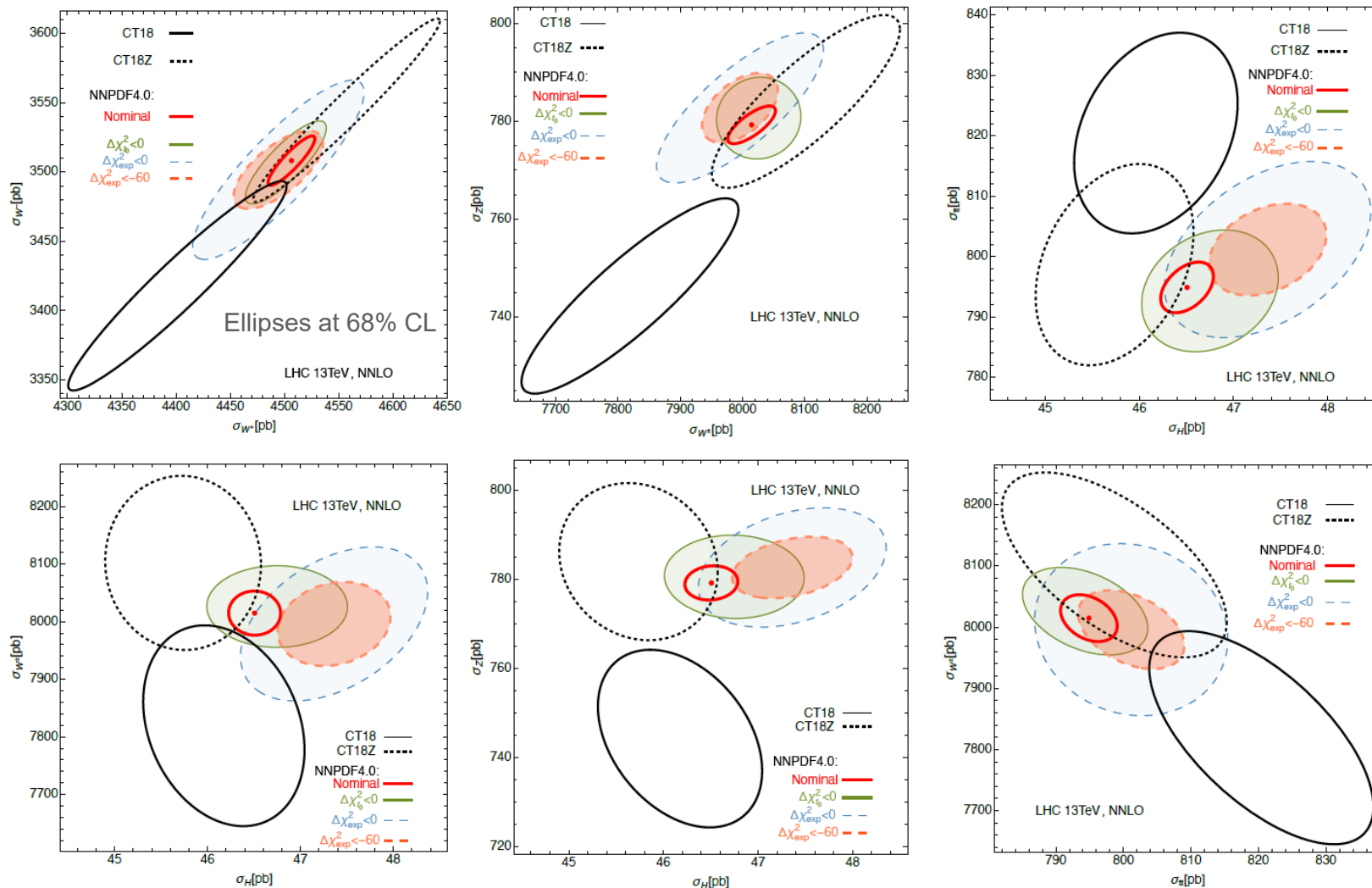
Details in
A. Courtoy et al.,
[arXiv:2205.10444](https://arxiv.org/abs/2205.10444)



obtained with the same NNPDF4.0 fitting code
using a “**hopscotch scan**” of the PDF param. space

all ellipses contain acceptable predictions
according to the likelihood-ratio test
Nominal NN4.0 uncertainty is too small!

Impact of epistemic uncertainties on other cross sections



The ellipses are projections of 68% c.l. ellipsoids in N_{par} -dim. spaces

$N_{par} = 28$ and 50 for CT18 and NNPDF4.0 Hessian PDFs

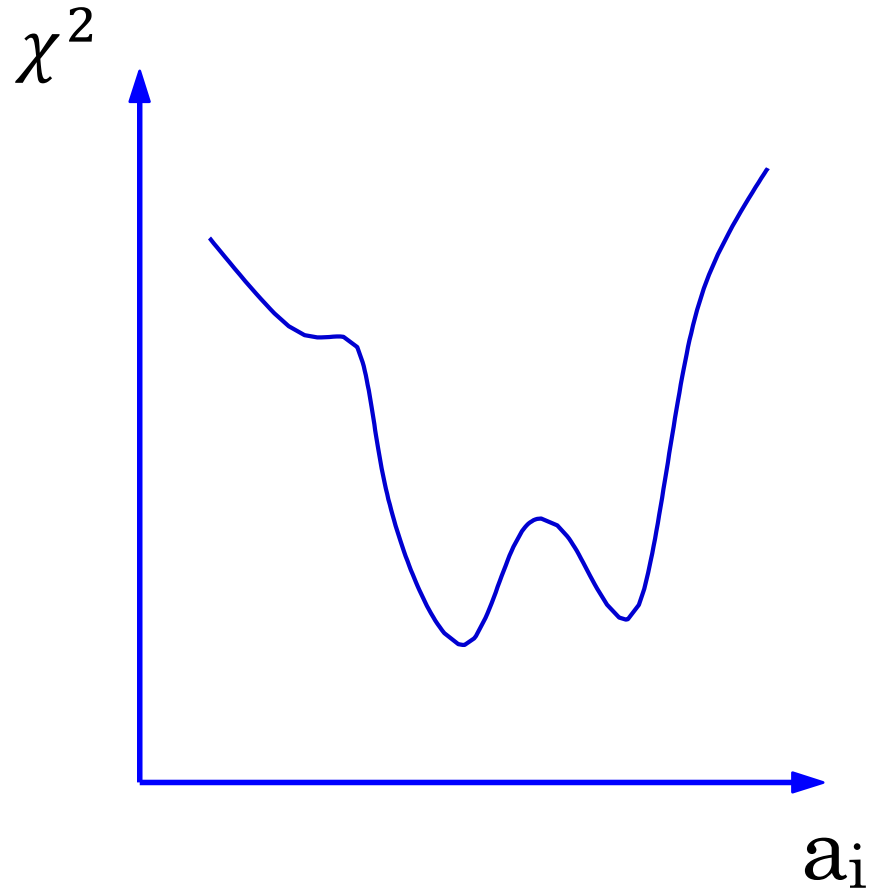
We knew about the PDF mini-landscape (now viewed as a saddle-point manifold) for 20+ years!

Profound implications for uncertainty quantification

Justification of the PDF tolerance due to

- incomplete agreement of experiments
- epistemic uncertainty

Multi-dimensional PDF error analysis

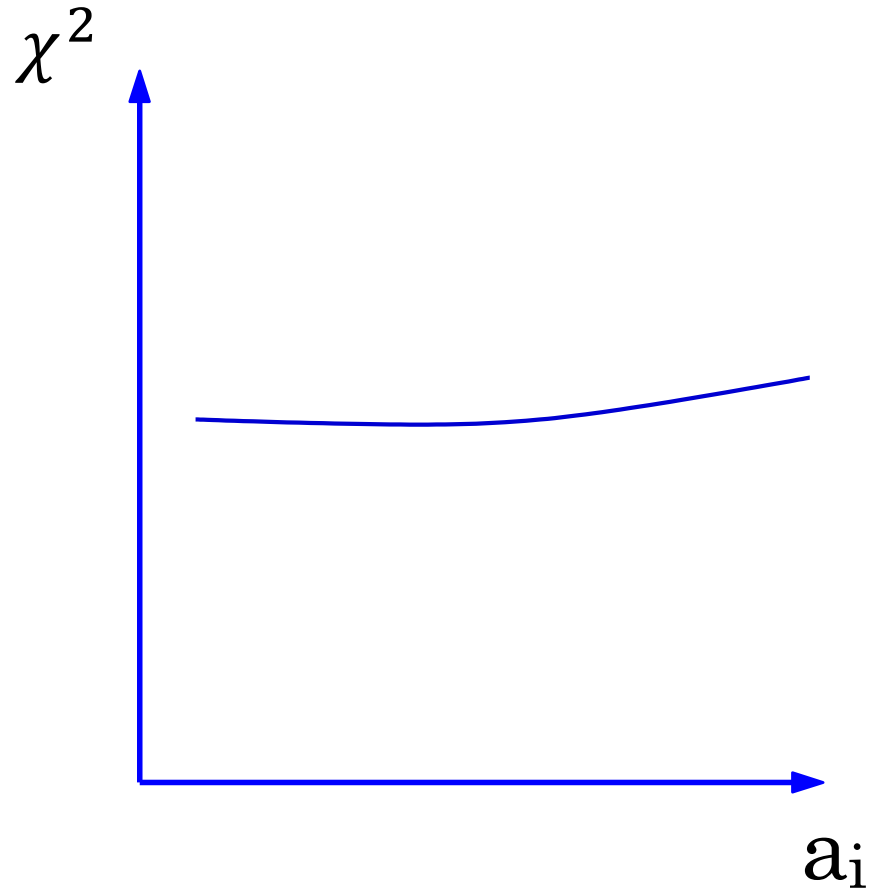


Pitfalls to avoid

- “Landscape”
 - disagreements between the experiments

<https://online.kitp.ucsb.edu/online/lhc08/nadolsky/>

Multi-dimensional PDF error analysis

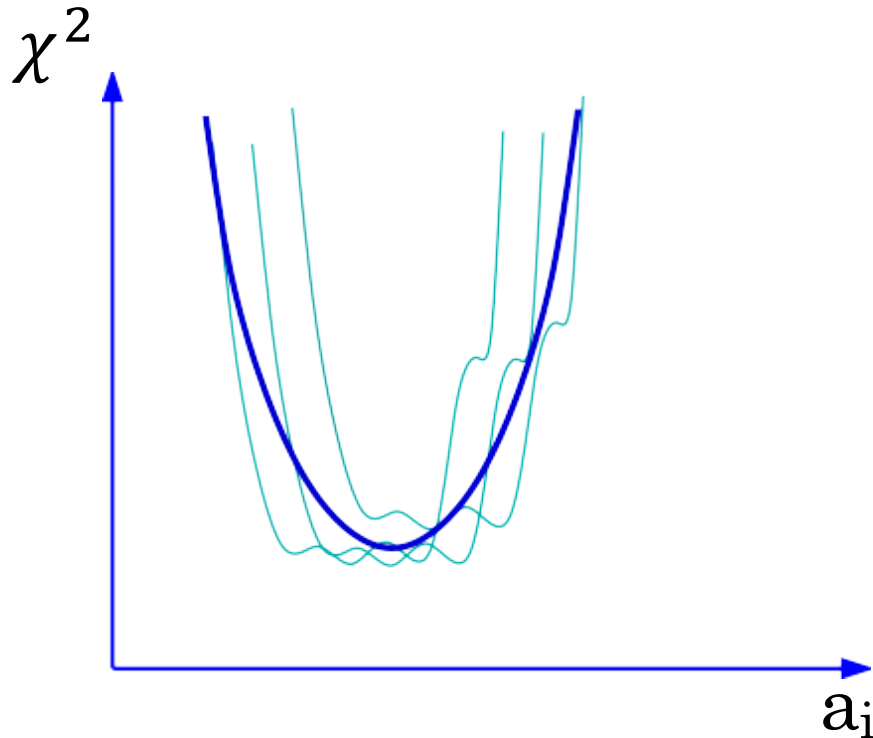


Pitfalls to avoid

- Flat directions
 - unconstrained combinations of PDF parameters

<https://online.kitp.ucsb.edu/online/lhc08/nadolsky/>

Multi-dimensional PDF error analysis

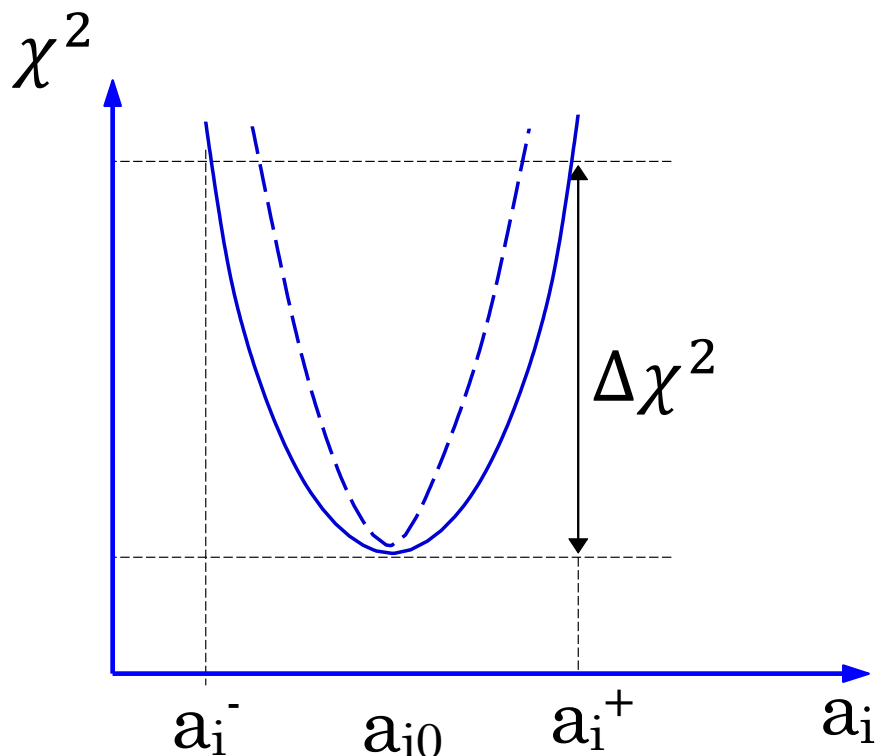


The actual χ^2 function shows

- a well pronounced global minimum near χ_0^2
- weak tensions between data sets in the vicinity of χ_0^2 (mini-landscape)
- some dependence on assumptions about flat directions

The likelihood is approximately described by a quadratic χ^2 with a revised tolerance condition $\Delta\chi^2 \leq T^2$

Multi-dimensional PDF error analysis



The actual χ^2 function shows

- a well pronounced global minimum near χ_0^2
- weak tensions between data sets in the vicinity of χ_0^2 (mini-landscape)
- some dependence on assumptions about flat directions

The likelihood is approximately described by a quadratic χ^2 with a revised tolerance condition $\Delta\chi^2 \leq T^2$

Weak and strong goodness-of-fit criteria

Kovarik, P. N., Soper, **arXiv:1905.06957**

Weak (common) goodness-of-fit (GOF) criterion

Based on the global χ^2

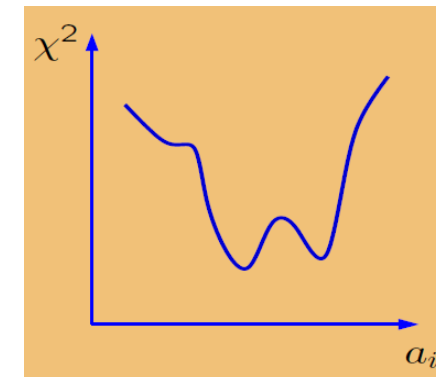
A fit of a PDF model to N_{exp} experiments with N_{pt} points ($N_{pt} \gg 1$) is good at the probability level p if $\chi_{global}^2 \equiv \sum_{n=1}^{N_{exp}} \chi_n^2$ satisfies

$$P(\chi^2 \geq \chi_{global}^2, N_{pt}) \geq p; \quad e.g.$$

$$|\chi_{global}^2 - N_{pt}| \lesssim \sqrt{2N_{pt}} \quad \text{for } p = 0.68$$

Even when the weak GOF criterion is satisfied, parts of data can be poorly fitted

Then, **tensions between experiments** may lead to **multiple solutions** or **local χ^2 minima** for some PDF combinations



An excellent fit requires more than a good global χ^2

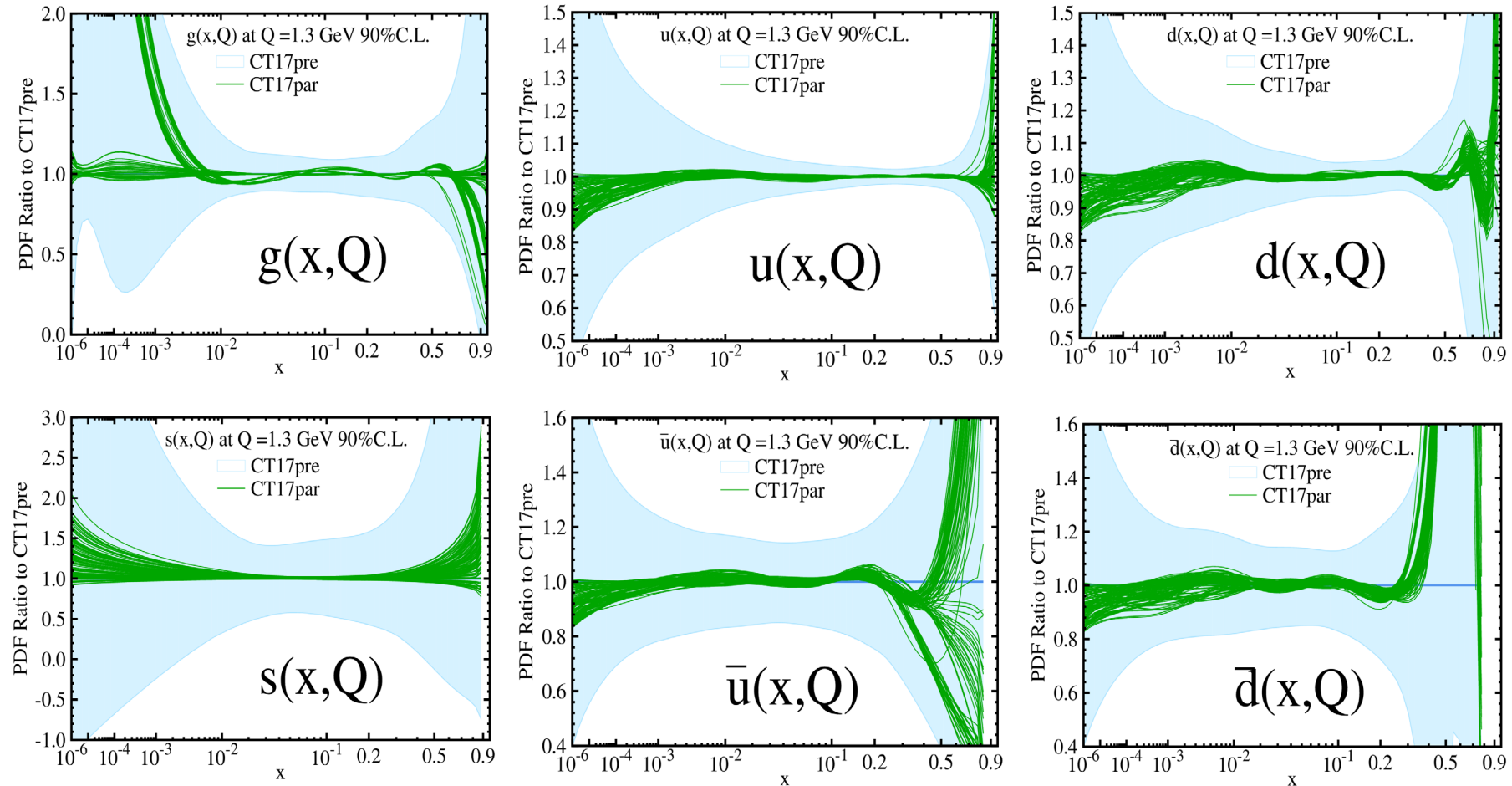
It passes a number of quality tests, called together the **strong set of goodness-of-fit criteria**

1. Each possible partition n of the global data set has a good χ^2
 - differences between theory and data for this partition are indistinguishable from random fluctuations
 - $P(\{\chi_n^2\}) \geq 0.68$ for the distribution of χ_n^2 over N_{part} partitions
2. Best-fit nuisance parameters obey the expected probability distribution
3. **Resampling test:** the data are neither underfitted nor overfitted
4. A closure test is passed, such as the one used in NNPDF 3.x
5. ...

Functional forms of PDFs and resampling test

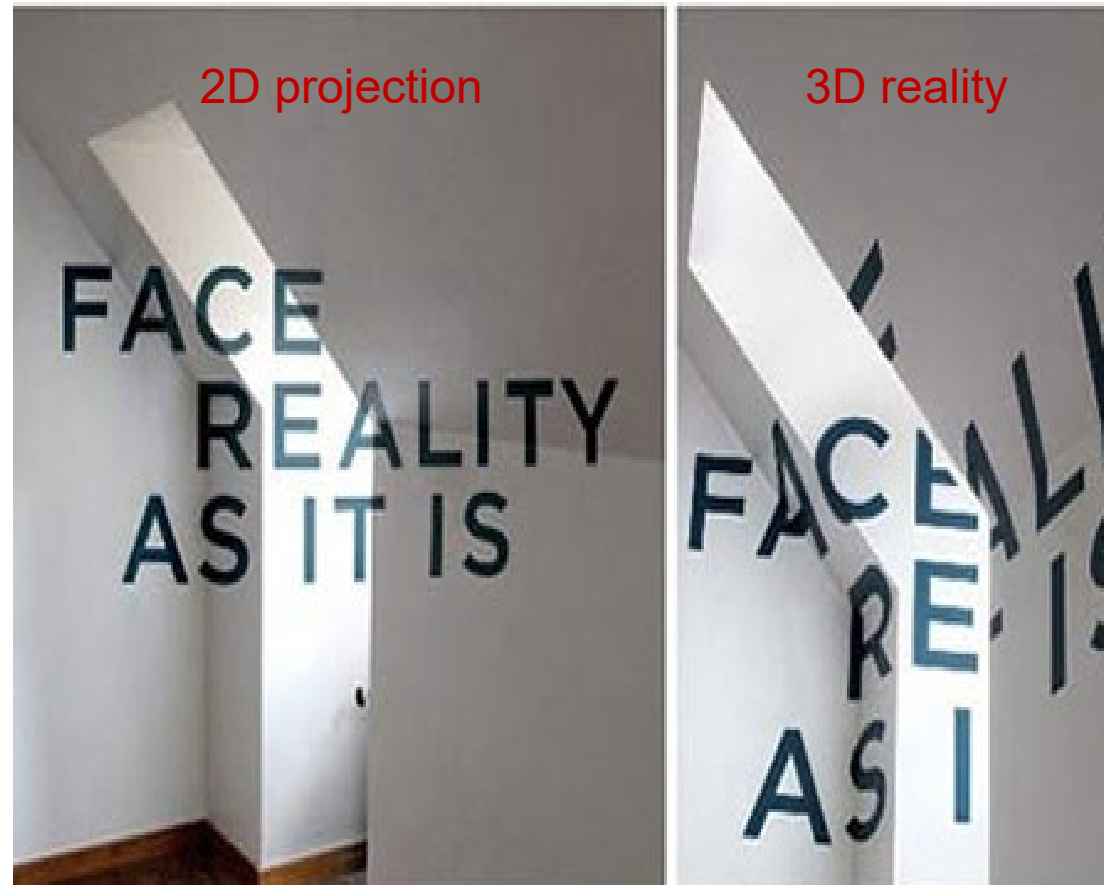
The uncertainty due to the PDF functional form contributes as much as 50% of the total PDF uncertainty in CT fits. The CT18 analysis estimates this uncertainty using 100 trial functional forms.

Explore various non-perturbative parametrization forms of PDFs



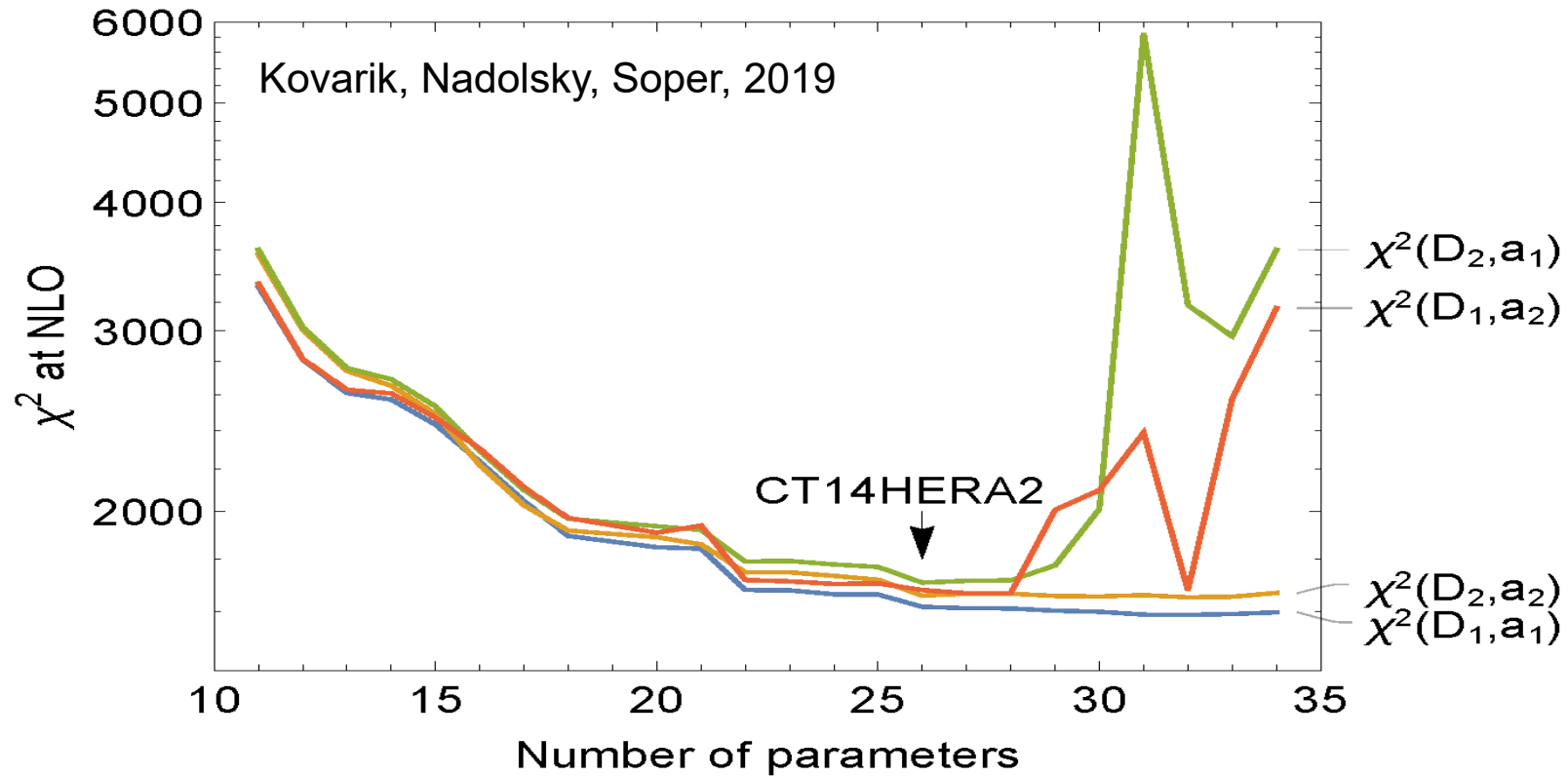
- CT17par – sample result of using various non-perturbative parametrization forms.
- No data constrain very large x or very small x regions.

If too few parameters



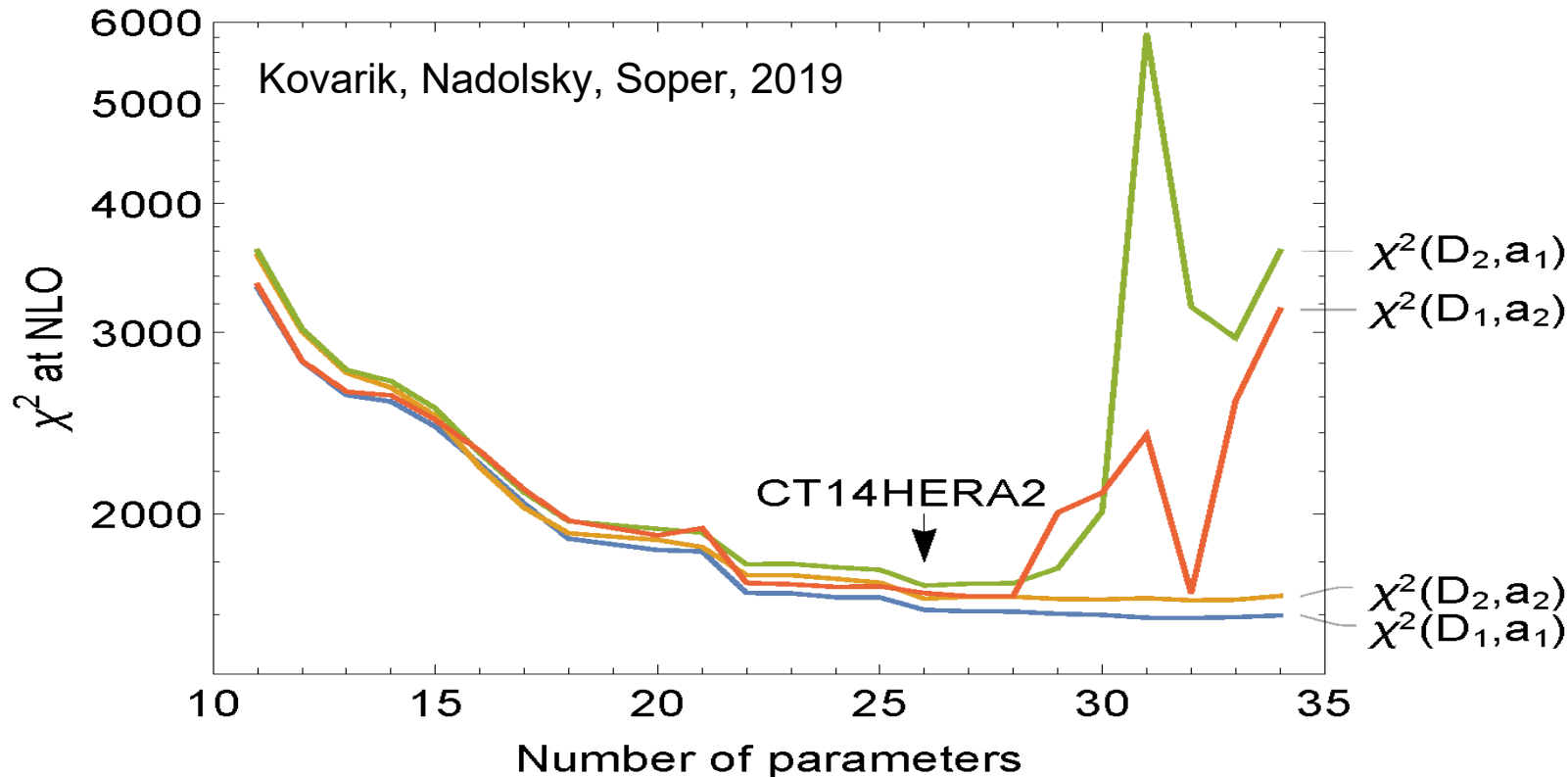
The solution can be consistent and false

If too many parameters



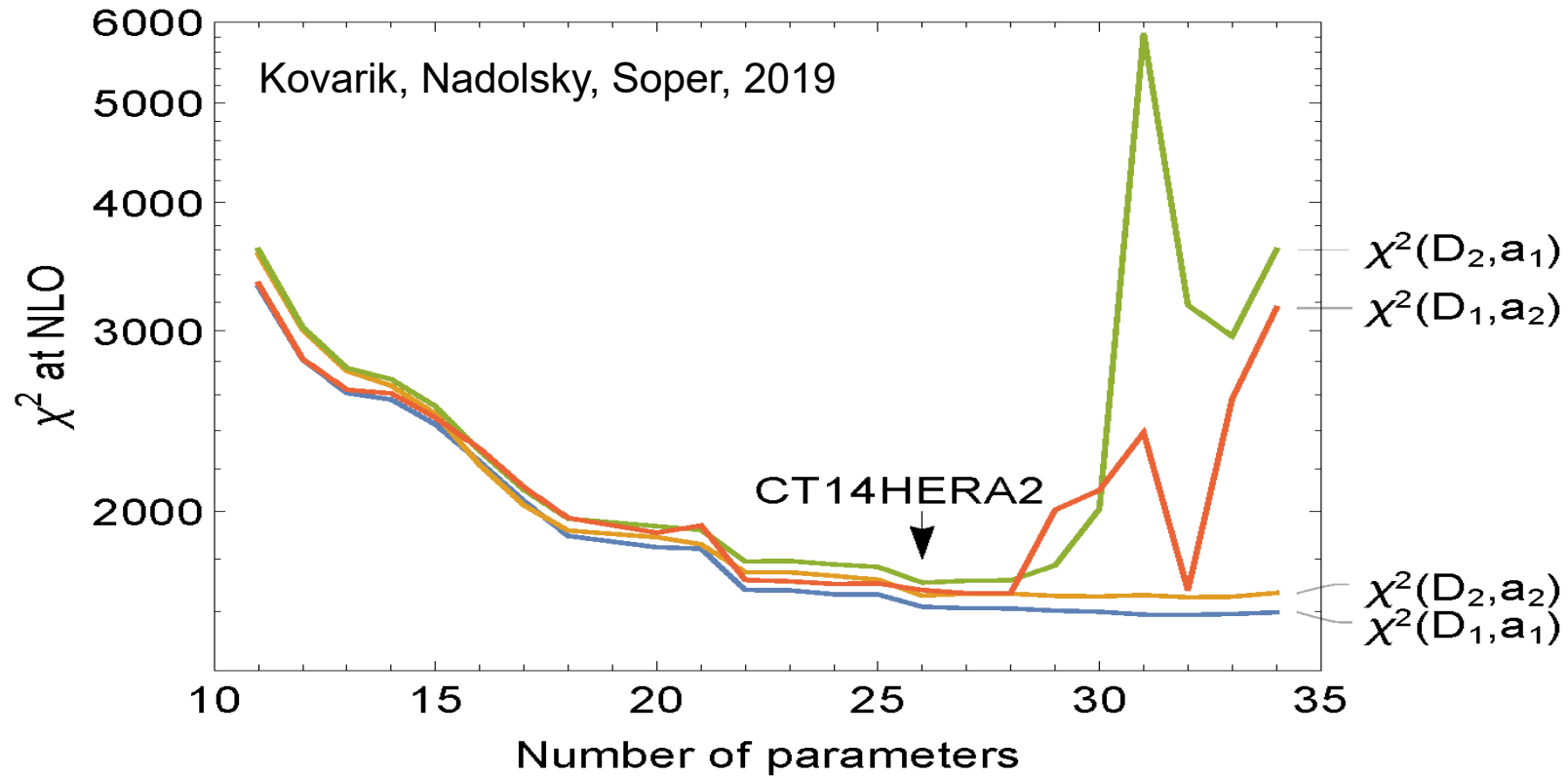
- Randomly split the CT14HERA data set into two halves, D_1 and D_2
- Find parameter vectors a_1 and a_2 from the best fits for D_1 and D_2 , respectively

If too many parameters



- **Fitted samples:** $\chi^2(D_1, a_1)$ and $\chi^2(D_2, a_2)$ uniformly decrease with the number of parameters; eventually the fits become unstable (“fitting noise”)
- **Control samples:** $\chi^2(D_2, a_1)$ and $\chi^2(D_1, a_2)$ fluctuate when the number of parameters is larger than about 30

If too many parameters

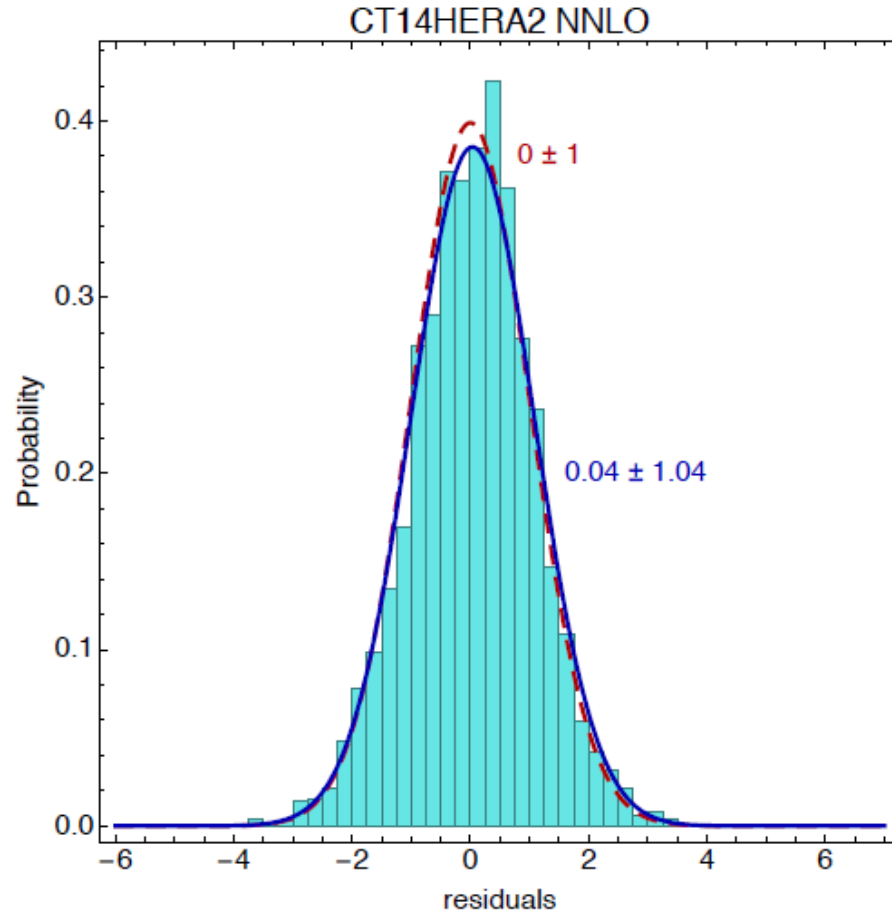


≈ 30 parameters (26 in CT14HERA2) is optimal for describing the CT14HERA2 data set. 15 parameters or less is optimal for nuclear PDFs

How well are the data described?

Note: It is convenient to define $S_n(\chi^2, N_{pt})$ that approximately obeys the standard normal distribution (mean=0, width=1) independently of N_{pt}

Example: data residuals r_n

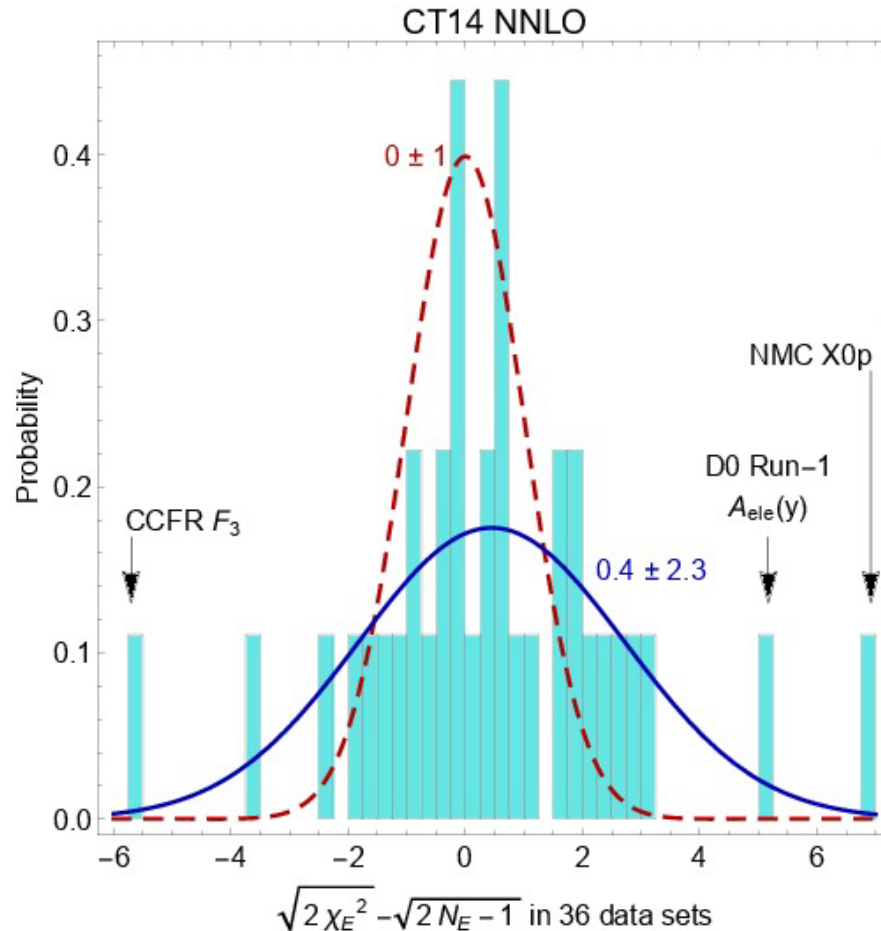


$$r_n \equiv \frac{T_n(\{a\}) - D_n^{shifted}(\{a\})}{\sigma_n^{uncorrelated}}$$

The distribution of residuals is consistent with the standard normal distribution

Full definition of r_n in the backup slides

Example: individual experiments



Define

$$S_n(\chi^2, N_{pt}) \equiv \sqrt{2\chi^2} - \sqrt{2N_{pt} - 1}$$

$S_n(\chi_n^2, N_{pt,n})$ are Gaussian distributed with mean 0 and variance 1 for $N_{pt,n} \geq 10$

[R.A.Fisher, 1925]

Even more accurate (χ^2, N_{pt}) :

T.Lewis, 1988

An empirical S_n distribution can be compared to $N(0,1)$ visually or using a statistical (KS or related) test

Effective Gaussian variables

