Representative Monte Carlo sampling for Parton Distributions

Aurore Courtoy

Instituto de Física
Universidad Nacional Autónoma de México (UNAM)

Mini-workshop on MC methods

ANL (HEP)

05/18/23

Consejo Nacional de Ciencia y Tecnología

Dirección General de Asuntos del Personal Académico

Instituto de Física UNAM

Uncertainty quantification for parton distributions

Parton Distribution Functions (PDFs) encapsulate probabilities for finding quarks, antiquarks, and gluons in hadrons participating in high-energy collisions. A PDF is a function of a variable $x \in [0,1]$, and the truth expression of that function is unknown from theory. The shape of PDFs is hence extracted from data.

I will discuss uncertainty quantification for PDFs mainly based on

"Parton distributions need representative sampling"
[Phys.Rev.D 107] ***arXiv version more complete***

CTEQ-TEA collaboration
China: S. Dulat, J. Gao, T.-J. Hou, I. Sitiwaldi, M. Yan, and collaborators
Mexico: A. Courtoy
USA: T.J. Hobbs, M. Guzzi, J. Huston, P. Nadolsky, C. Schmidt, D. Stump, K. Xie, C.-P. Yuan

Application of concept of epistemic PDF uncertainties —Fantômas team
Mexico: A. Courtoy, D.M. Ponce-Chávez
USA: L. Kotz, P. Nadolsky, F. Olness

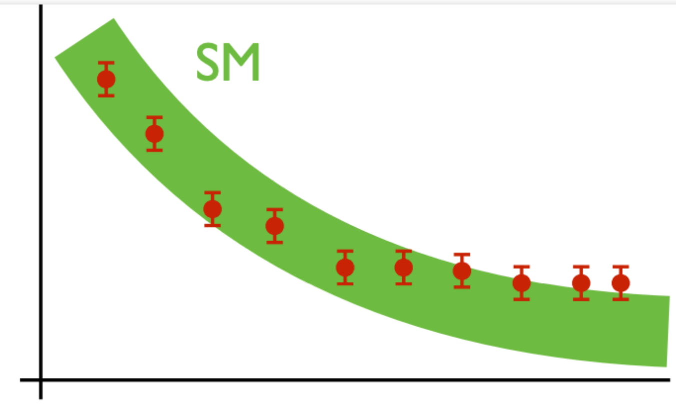
based on [AC & Nadolsky, Phys.Rev.D 103].

Challenges in global analyses of PDFs

Keynote talks at DIS'23 (April 23)

Daniel de Florian:
need for precision

- Most likely look for "new interactions"
- Small deviations from SM : PRECISION
- EFT description / BSM model



- **Precision** is the name of the game for the next decades (Higgs sector)

Marteen Boonekamp:
need for accuracy

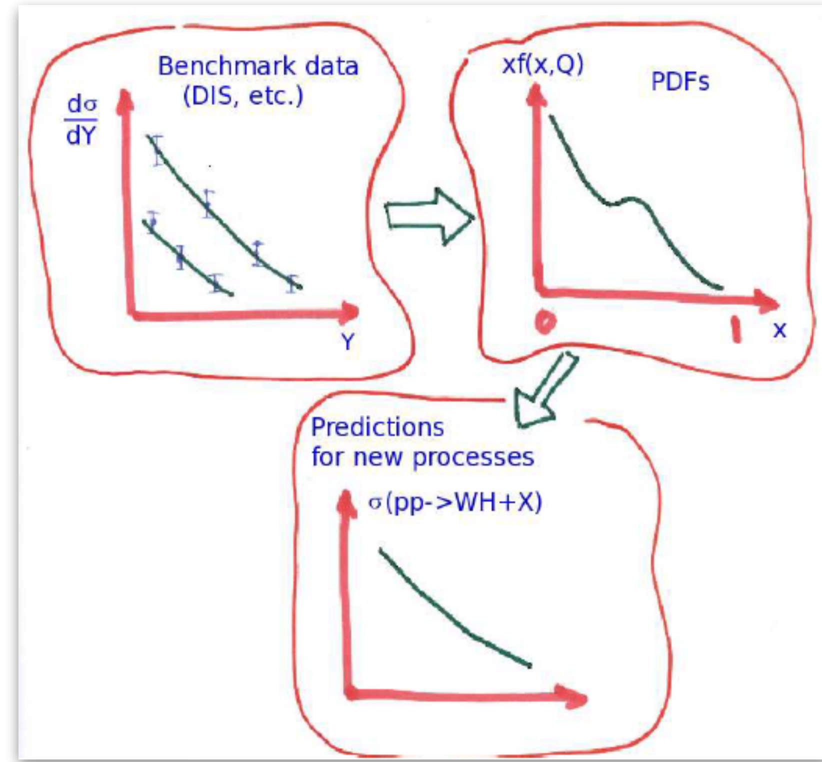
- Experiments WELCOME the ongoing inclusion of theoretical uncertainties in PDF fits.
- Still, very difficult to understand the significance of differences between results obtained using different PDF sets
  - Very interesting discussion in WG1
  - better uncertainty decomposition required

- M_W is such an active field, all of a sudden!
- Uncertainty propagation for this measurement currently *almost* broken by the PDFs – we should improve, and the discussions this week were extremely helpful

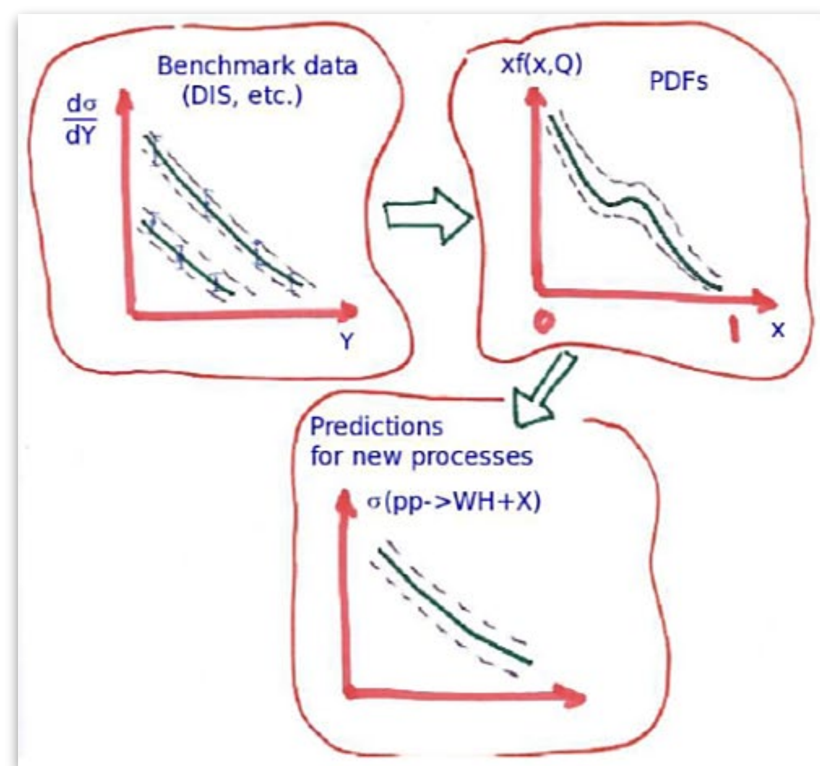
The shape of parton distributions

Parton distributions are functions of the momentum fraction x, they are extracted from data that are sensitive to specific PDF flavors, etc.

⇨ finding the shape in x is the goal of PDF analyses



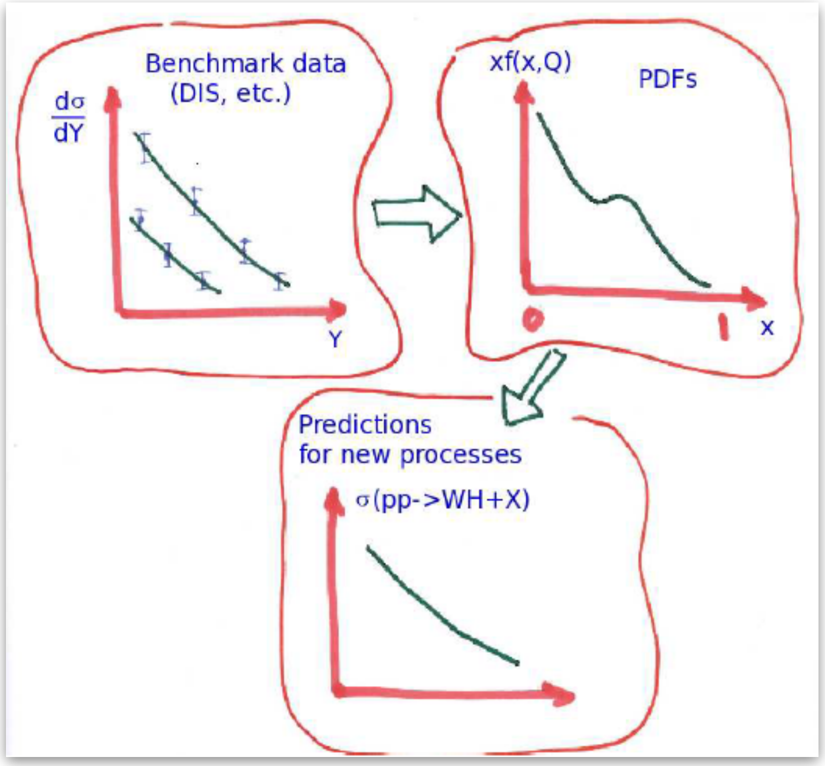
Uncertainty propagates from data and methodology to the PDF determination.
There are two classes of them,



The shape of parton distributions

Parton distributions are functions of the momentum fraction x, they are extracted from data that are sensitive to specific PDF flavors, etc.

⇨ finding the shape in x is the goal of PDF analyses

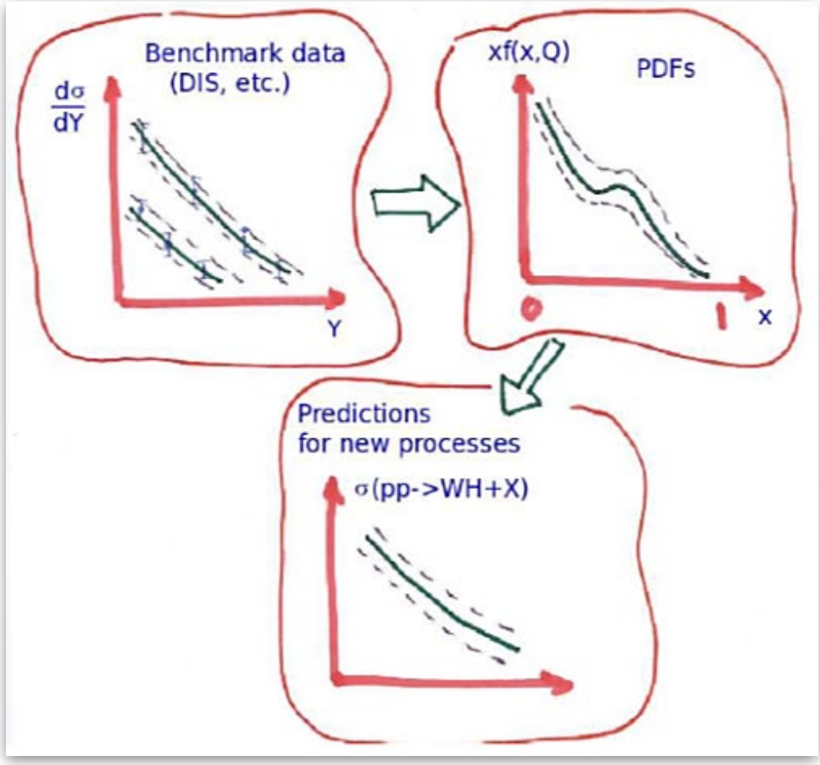


Uncertainty propagates from data and methodology to the PDF determination.
There are two classes of them,

epistemic vs. aleatory uncertainties

Statistical uncertainty propagated from experiments — irreducible

Uncertainty due to lack of knowledge —bias (may be reduced)



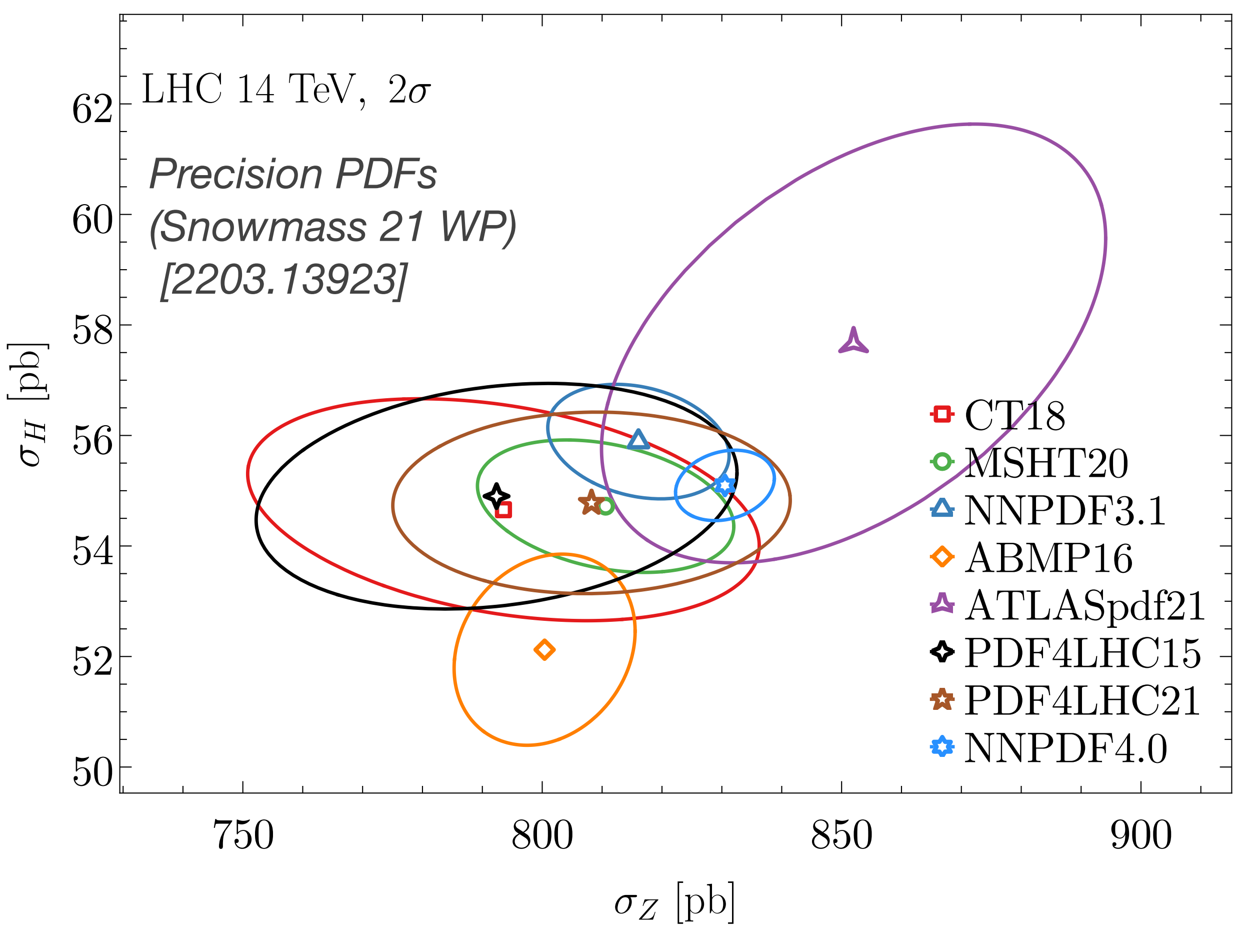
Criteria for PDF uncertainties

Recent advancements in the determination of unpolarized PDFs:
CT18, MSHT20, NNPDF4.0, ATLASpdf21 as well as PDF4LHC21.

PDF4LHC21:

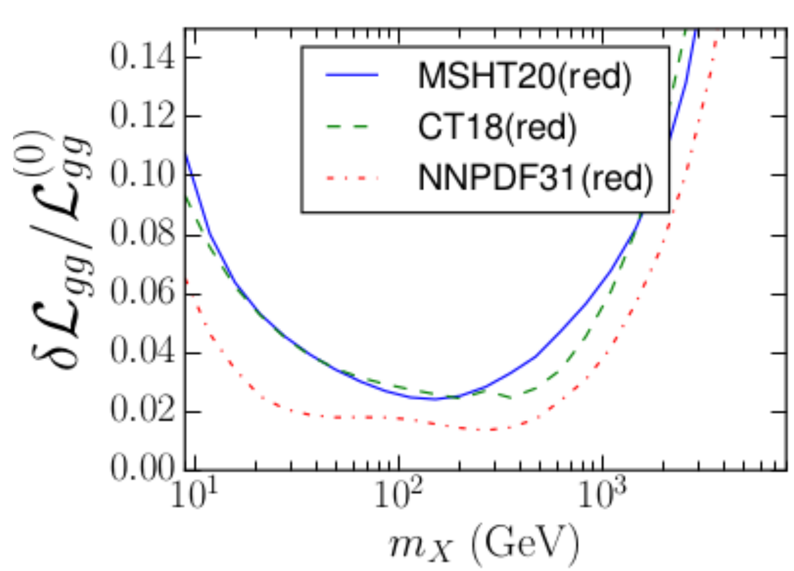
benchmarking and combination of the leader PDF sets, CT, MSHT & NNPDF, for the run III of the LHC.

[Ball, et al, J.Phys.G 49 (2022)]

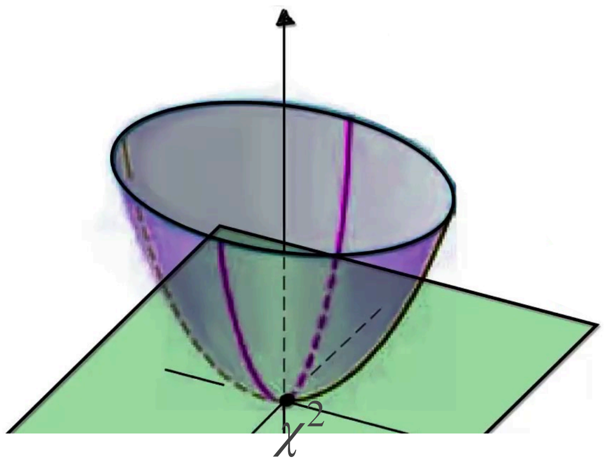


What is the origin of the differences in size of correlation ellipses among various fits?

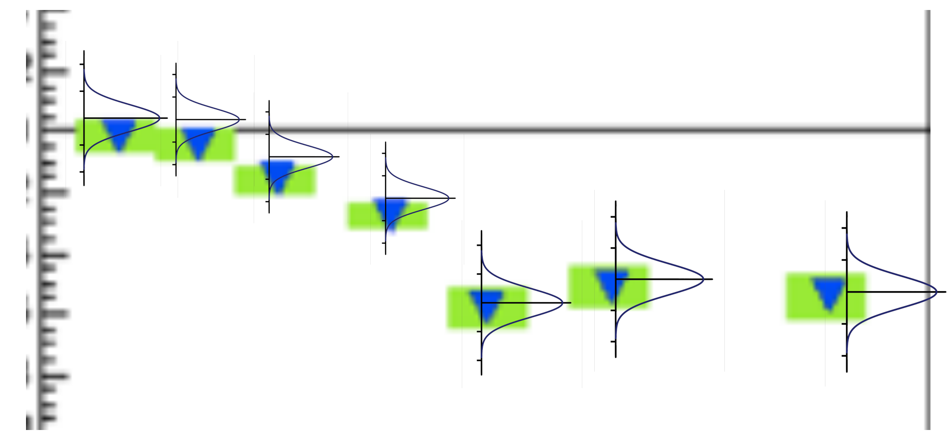
PDF4LHC21 excercise highlights the role of methodology. Monte Carlo-based analysis (NNPDF) gives smaller uncertainties.



Hessian and MC frameworks

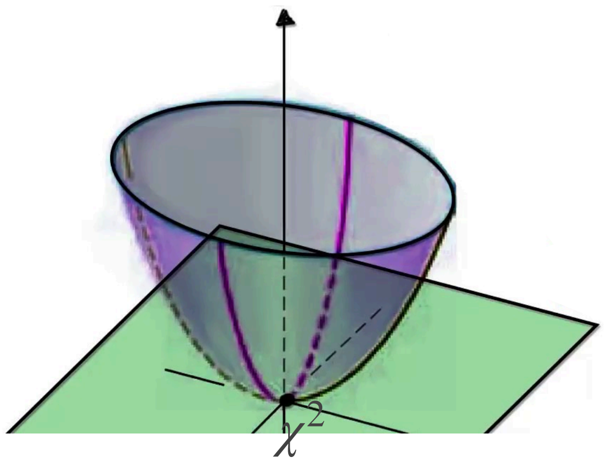


Hessian methodology finds the global minimum and explores the parameter space.
PDFs are represented by functional forms ; a set of Hessian PDFs depend on ~30 params.

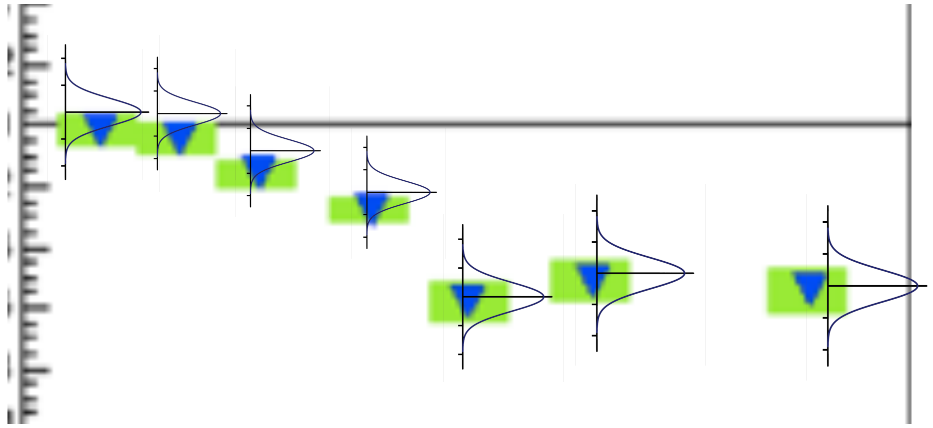


Monte-Carlo methodology (neural network, AI/ML) replicates fluctuated data, then optimizes each replica (up to training).
A set of MC PDFs depends on up to several hundreds of latent parameters, as well as hyperparameters

Hessian and MC frameworks



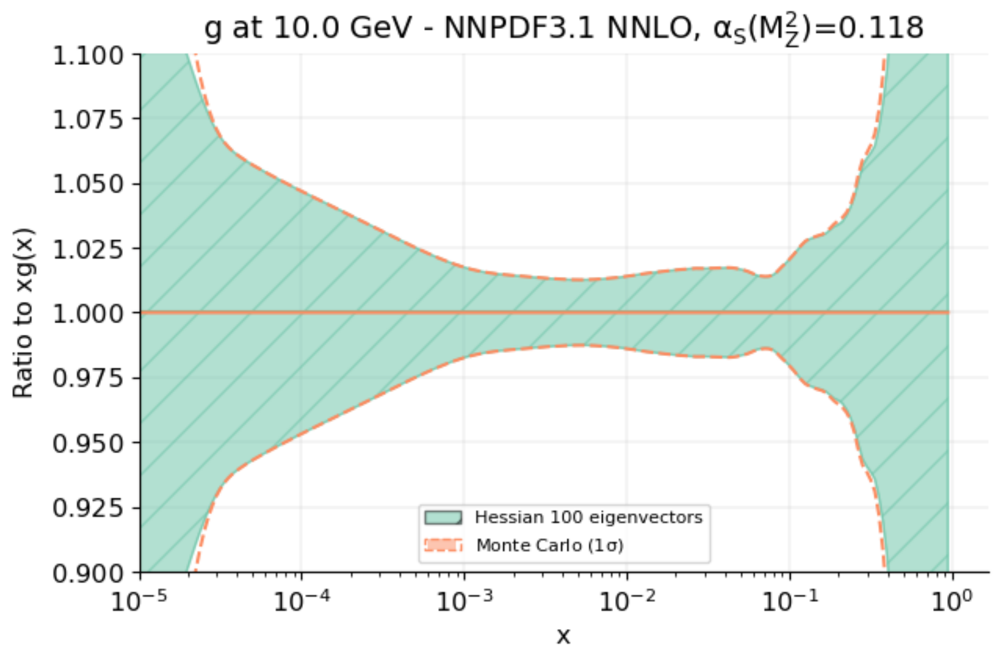
Hessian methodology finds the global minimum and explores the parameter space.
PDFs are represented by functional forms ; a set of Hessian PDFs depend on ~30 params.



Monte-Carlo methodology (neural network, AI/ML) replicates fluctuated data, then optimizes each replica (up to training).
A set of MC PDFs depends on up to several hundreds of latent parameters, as well as hyperparameters

Hessian and Monte Carlo representations of given PDF sets are shown to be compatible — conversions exist in both ways

[Gao & Nadolsky, JHEP 07 (2014) 035]
[Carrazza et al.,Eur.Phys.J.C 75 (2015) 8, 369]



[Lambri thesis, U. Milano]

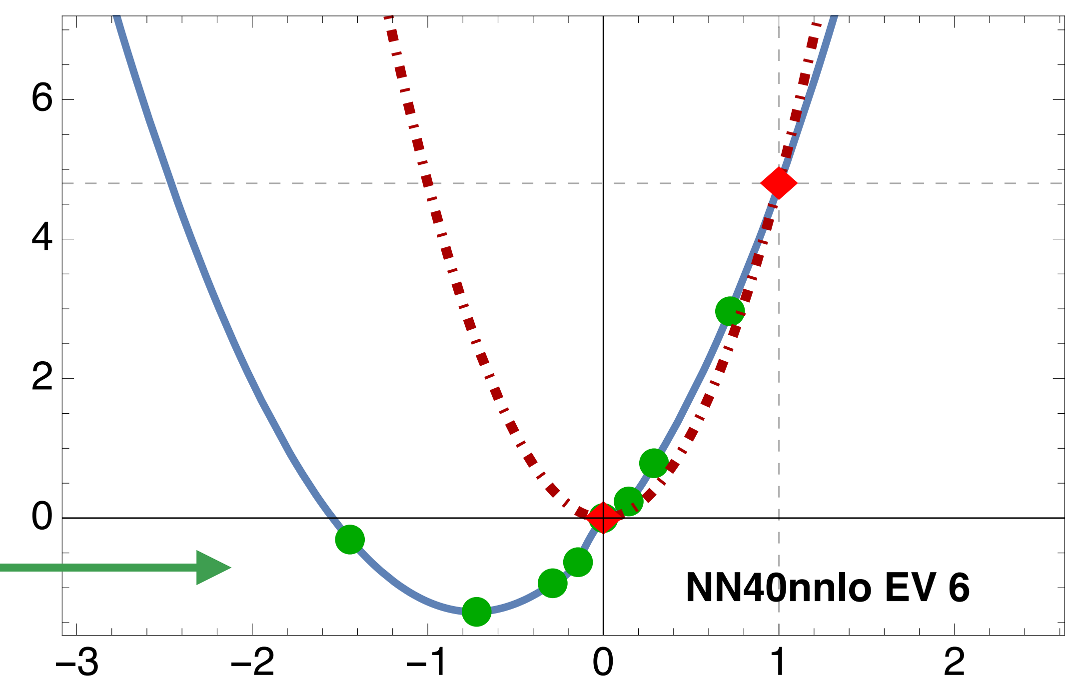
Hence, a chi-square paraboloid can also be defined for Monte Carlo-based analyses.

Equivalence of Hessian and MC frameworks

In multivariate analyses, sampling occurs at various levels — parameter space, bootstrap but also priors, … In large-dimensional problems, sampling is complex.

The conversion to Hessian is convenient to explore the efficiency of the Monte Carlo method in terms of χ^2 behavior in parameter space.

The space of N_{rep} is reduced to a 50-dimensional space, referred to by 50 eigenvector (EV) directions. The shape of each projected χ^2 is a parabola.

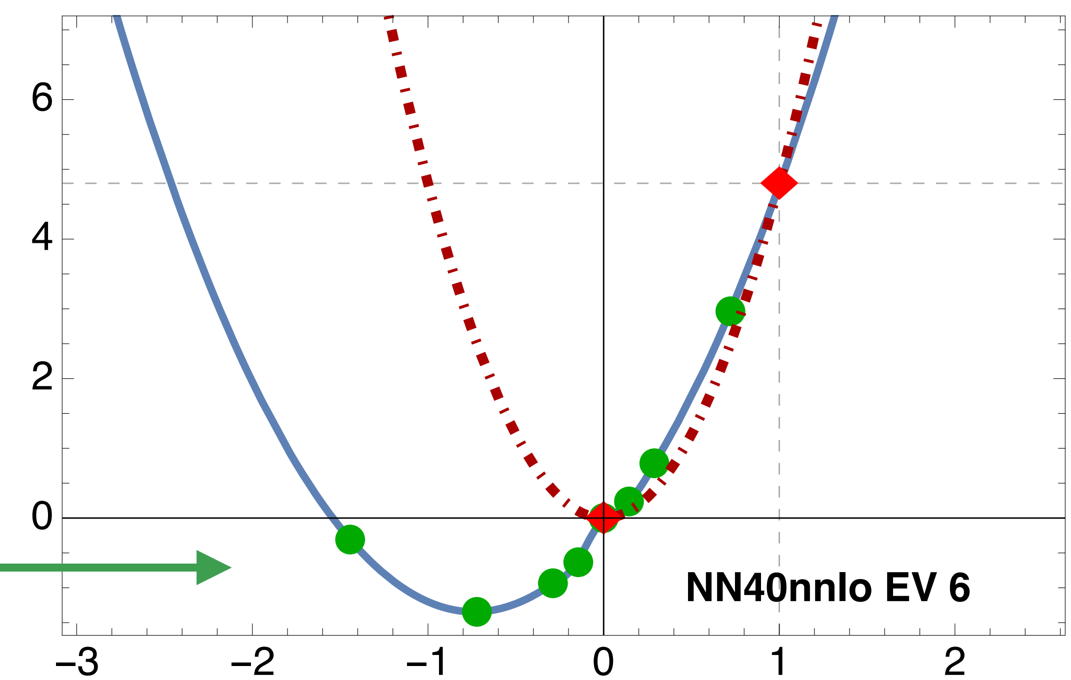


Equivalence of Hessian and MC frameworks

In multivariate analyses, sampling occurs at various levels — parameter space, bootstrap but also priors, … In large-dimensional problems, sampling is complex.

The conversion to Hessian is convenient to explore the efficiency of the Monte Carlo method in terms of χ^2 behavior in parameter space.

The space of N_{rep} is reduced to a 50-dimensional space, referred to by 50 eigenvector (EV) directions. The shape of each projected χ^2 is a parabola.



We can reconstruct EV directions for the neural network PDF collaboration (NNPDF4.0) that provide a Hessian set, in blue. There are second crossing points with$\Delta\chi^2 = 0$, for all 50 EV directions.
This indicates a larger paraboloid than the red curve provided by NNPDF4.0.

Key role played by methodology

The depth and position of the paraboloid obtained by the MC-to-Hessian converted PDFs suggest an incomplete span of the space of solutions.

Outside of HEP/NP, there is significant interest in statistical problems that are similar to the assessment of uncertainties for PDF.
These studies introduce a fundamental distinction between the fitting uncertainty and sampling uncertainty, often overlooked in the PDF fits.

Article

Unrepresentative big surveys significantly overestimated US vaccine uptake

Nature **v. 600** (2021) 695

https://doi.org/10.1038/s41586-021-04198-4
Received: 18 June 2021

Valerie C. Bradley, Shiro Kuriwaki, Michael Isakov, Dino Sejdinovic, Xiao-Li Meng & Seth Flaxman

SCIENCE ADVANCES | RESEARCH ARTICLE

MATHEMATICS

Models with higher effective dimensions tend to produce more uncertain estimates

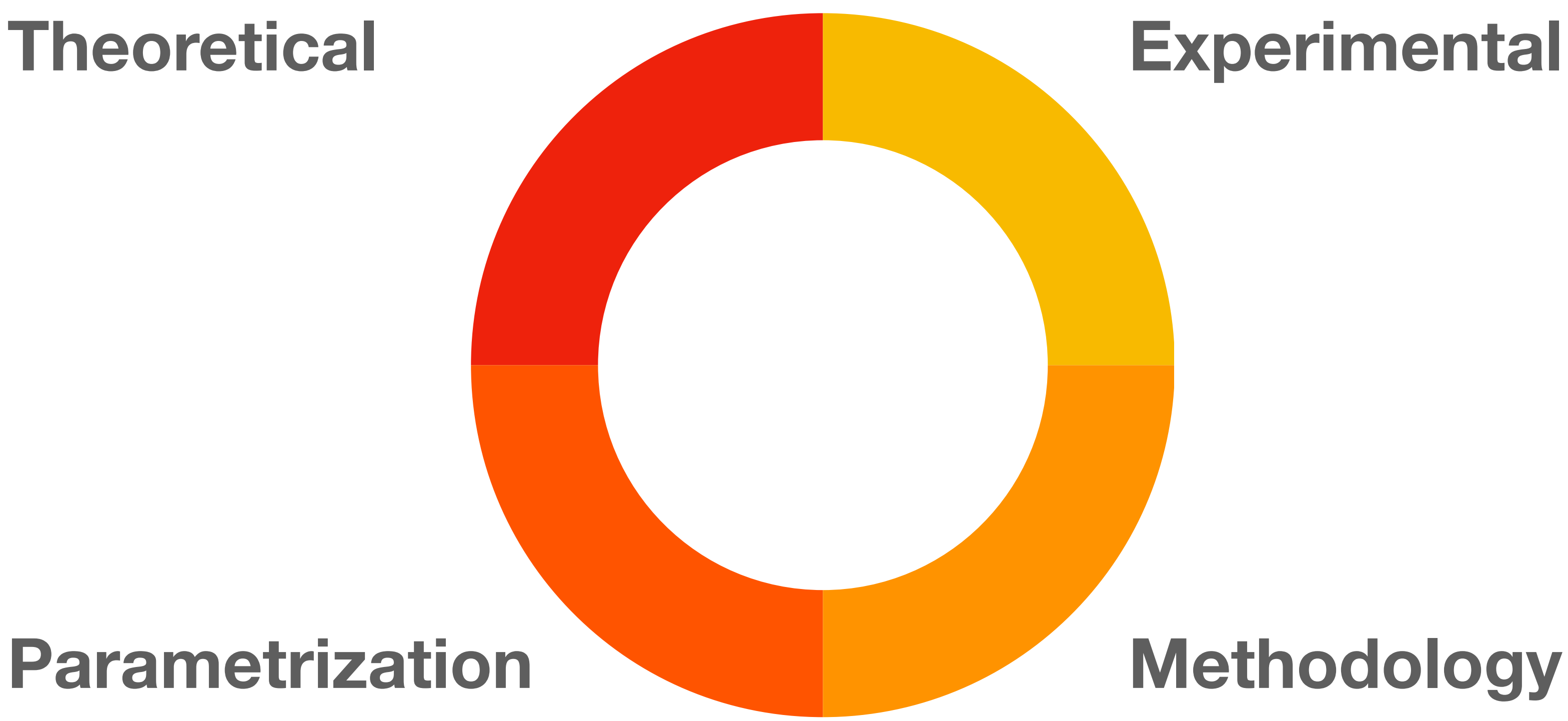
Arnald Puy[1,2,3]*, Pierfrancesco Beneventano[4], Simon A. Levin[2], Samuele Lo Piano[5], Tommaso Portaluri[6], Andrea Saltelli[3,7]

The Big Data Paradox in Clinical Practice

Pavlos Msaouel

To cite this article: Pavlos Msaouel (2022) The Big Data Paradox in Clinical Practice, Cancer Investigation, 40:7, 567-576, DOI: 10.1080/07357907.2022.2084621

On uncertainty quantification



In all four categories of uncertainties, we can further distinguish *PDF fitting accuracy* from *PDF sampling accuracy*.

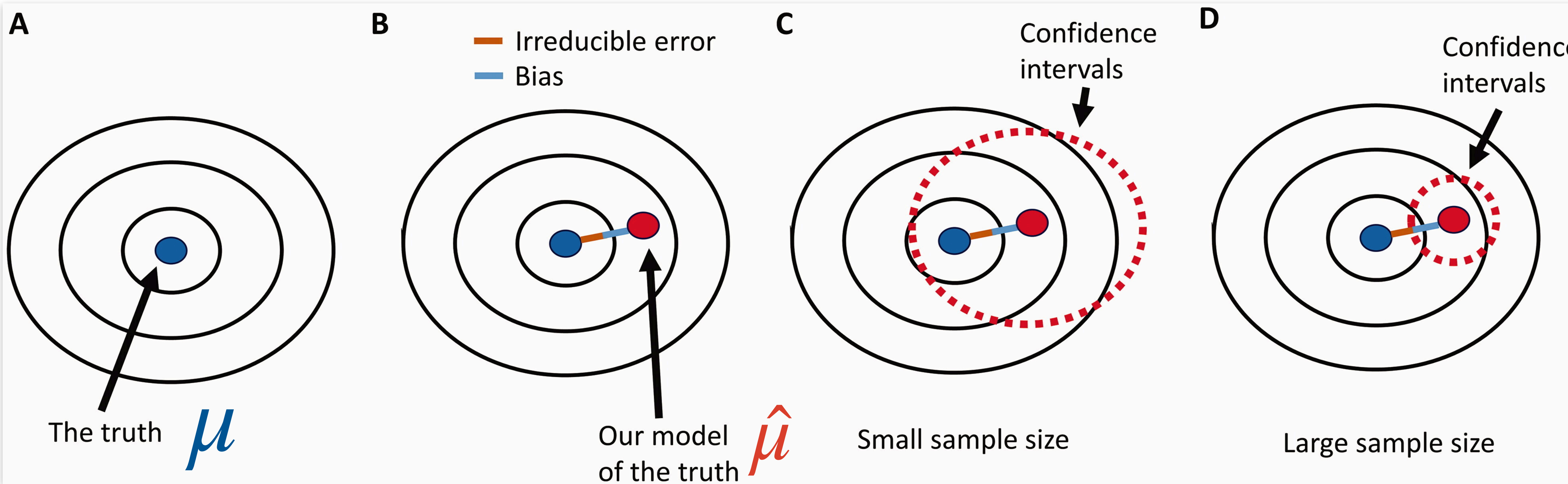
Goodness-of-fit applies to an individual best fit.

[Kovarik et al, Rev.Mod.Phys. 92 (2020)]

Sampling accuracy applies either to the tolerance or the number of error sets in a PDF ensemble.

This talk.

Sampling bias and big-data paradox

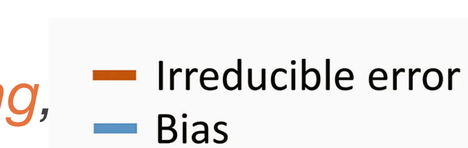


Pavlos Msaouel (2022)
Cancer Investigation, 40:7, 567-576

With an increasing size of sample $n \to \infty$, under a set of hypotheses, it is usually expected that the *deviation* on an observable decreases like $\left(\sqrt{n}\right)^{-1}$.

That's the law of large numbers.

What uncertainties keep us from including *the truth*, μ?

The law of large numbers disregards the *quality of the sampling*, 

.

Xiao-Li Meng
The Annals of Applied Statistics
Vol. 12 (2018), p. 685

Trio identity

Xiao-Li Meng
The Annals of Applied Statistics
Vol. 12 (2018), p. 685

$$\boxed{\mu - \hat{\mu} = \text{(data+sampling defect)} \times \text{(measure discrepancy)} \times \text{(inherent problem difficulty)}}$$

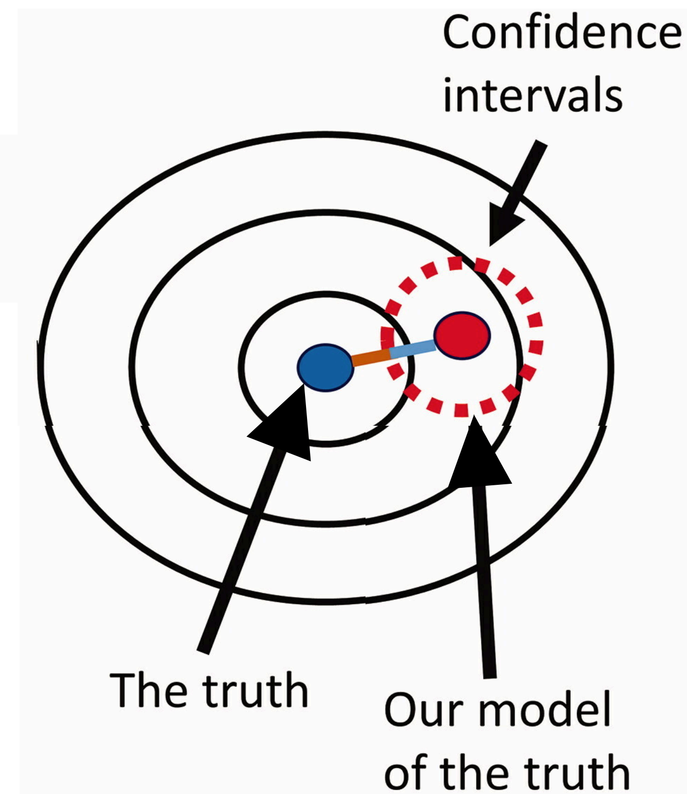
depends on the sampling algorithm

can tend to $\left(\sqrt{n}\right)^{-1}$ for random sampling

Irreducible error
Bias

$\equiv$ statistical model, quality of data,…

Large sample size



For a sample of n items from the population of size N, we can consider an array built by the random spanning of the binary responses of the $N - n$ (0) and n (1) items, so that

$$\boxed{\mu - \hat{\mu} = \mathrm{Corr}[\text{observable, sampling quality}] \times \sqrt{\frac{N}{n} - 1} \times \sigma(\text{observable})}$$

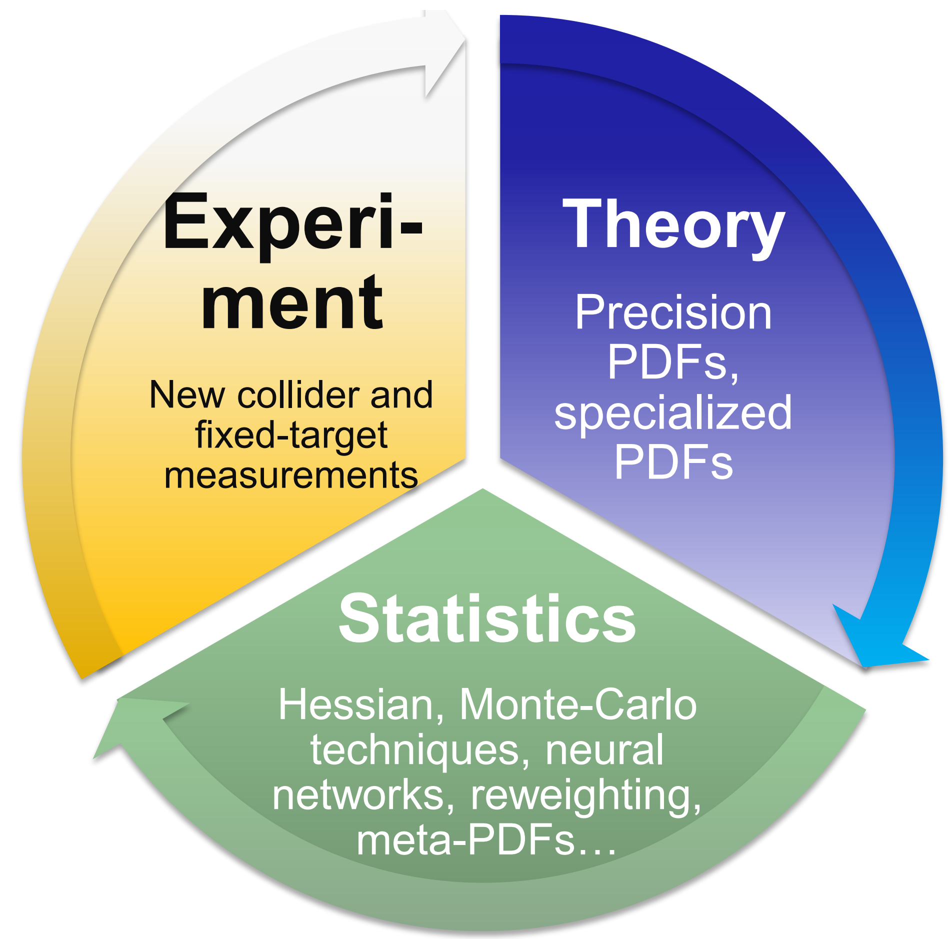
Sampling bias in PDF global analyses—I

How do we know the "data+sampling defect=confounding correlation" of our analysis?

Methodological choices are reflected in the epistemic uncertainty, including biases from sampling.

Priors, including choice of functional form or Bayesian *priors*, influence the sampling algorithm.

Representative sampling accounts for the confounding correlation, and can ultimately be used to optimize its contribution, e.g. through the study of largest effective dimensions.



⇨ dimensionality reduction (effective dimensions) vs. phase space reduction (priors)

Sampling bias in PDF global analyses — I

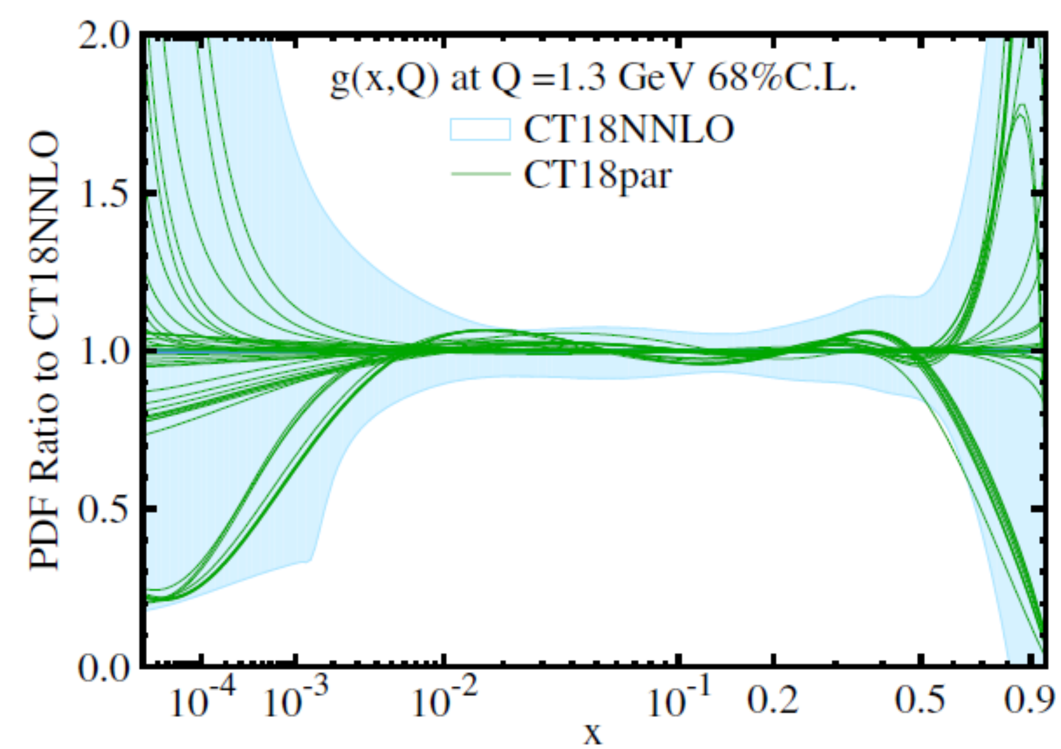
How do we know the "data+sampling defect=confounding correlation" of our analysis?

Hessian-based analysis:
objective function includes penalties, establishing the **tolerance criteria.**

Size of uncertainties reflect a series of confounding sources —selection of fitted experiments, treatment of correlated systematic errors, functional forms of PDFs, …

Verification that proper spanning of parameter space is compatible with total uncertainties (*a posteriori*). >300 functional forms are tested in CT18.

Dimensions of the problem given by the number of parameters=eigenvector (EV) directions.



Hou et al, Phys.Rev.D 103 (2021)

Sampling bias in PDF global analyses—II

How do we know the “data+sampling defect=confounding correlation” of our analysis?

Monte Carlo-based analysis:
optimization implies selection of hyperparameters (see NNPDF)

There are still many choices associated with the optimization:

- Number and width of the layers
- Activation functions and initialization
- Optimization algorithm (and associated parameters
- Training length, stopping condition
- Physics consstraints (positivity, integrability)

Do we understand sampling for QCD global analyses?

Sampling of multidimensional spaces ($d \gg 20$) is exponentially inefficient and may require $n > 2^d$ replicas to obtain a convergent expectation value.

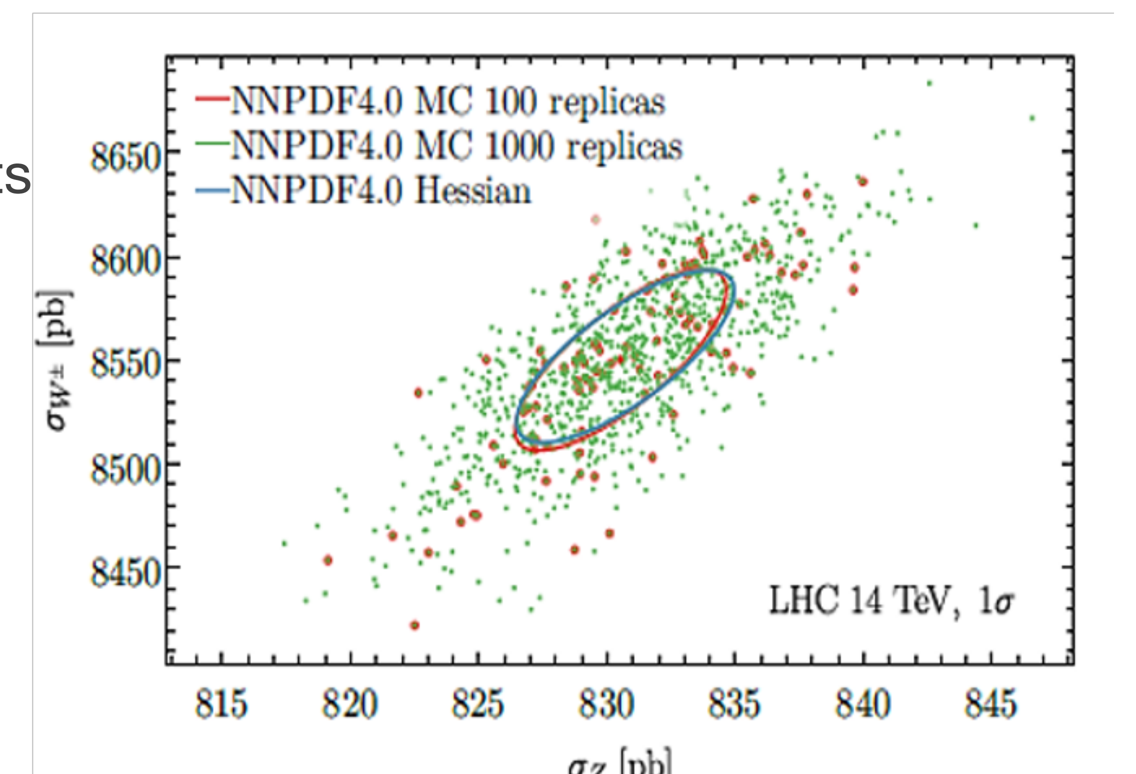
In general, an intractable problem.

[Hickernell, MCQMC 2016, 1702.01487]
[Sloan, Wo´zniakowski, 1997]

1. Justification for tolerance criteria for Hessian-based PDF fits

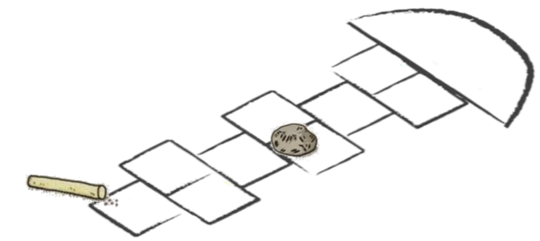
2. How is sampling achieved in Monte Carlo-based PDF fits?

Importance sampling, as defined by NNPDF
- =bootstrap/resampling of random fluctuations in data
- expectations are then unweighted averages over replica fits



Such sampling does not include sampling over hyperparameters and priors.

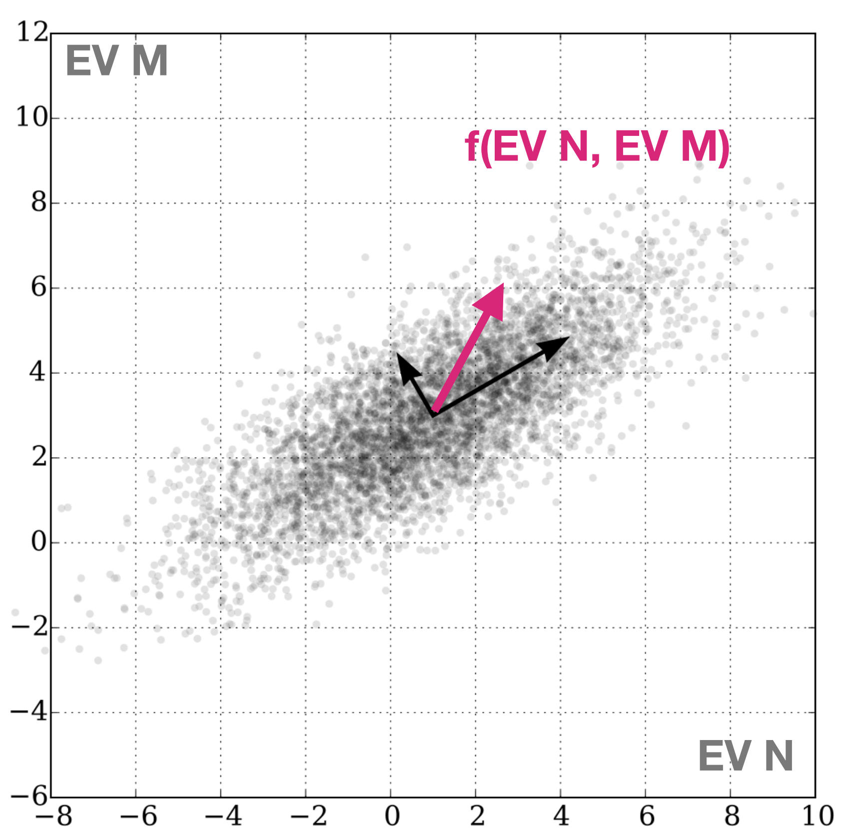
Observable-oriented effective dimensions



To efficiently span the 50-dimensional Hessian space, we identify the largest dimensions affecting chosen pairs of cross-sections.

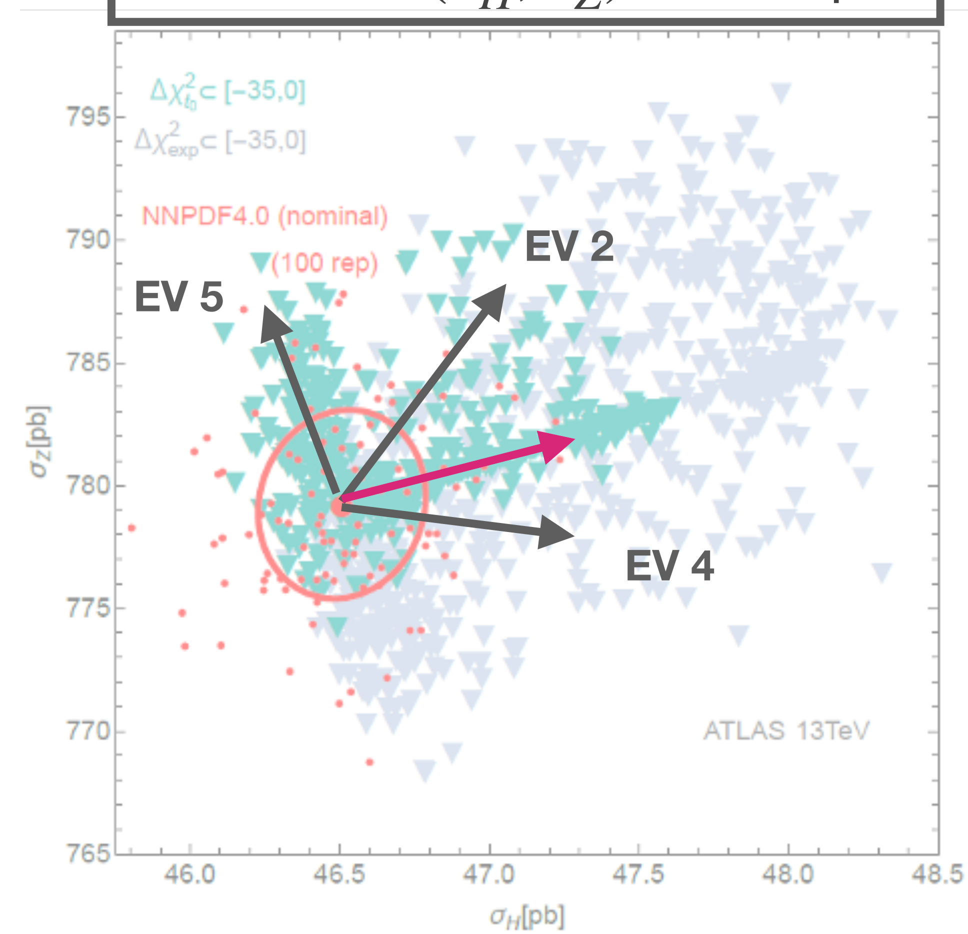
Based on the most relevant EV directions for given observables, we explore the directions in χ^2 space with $\chi^2 < \chi_0^2$ and, in turn, they sample the space of observables with more plausible solutions. That's the the hopscotch algorithm.

χ^2 space



° represent χ^2 for bootstrapped distribution

Observable (σ_H, σ_Z) scatter plot

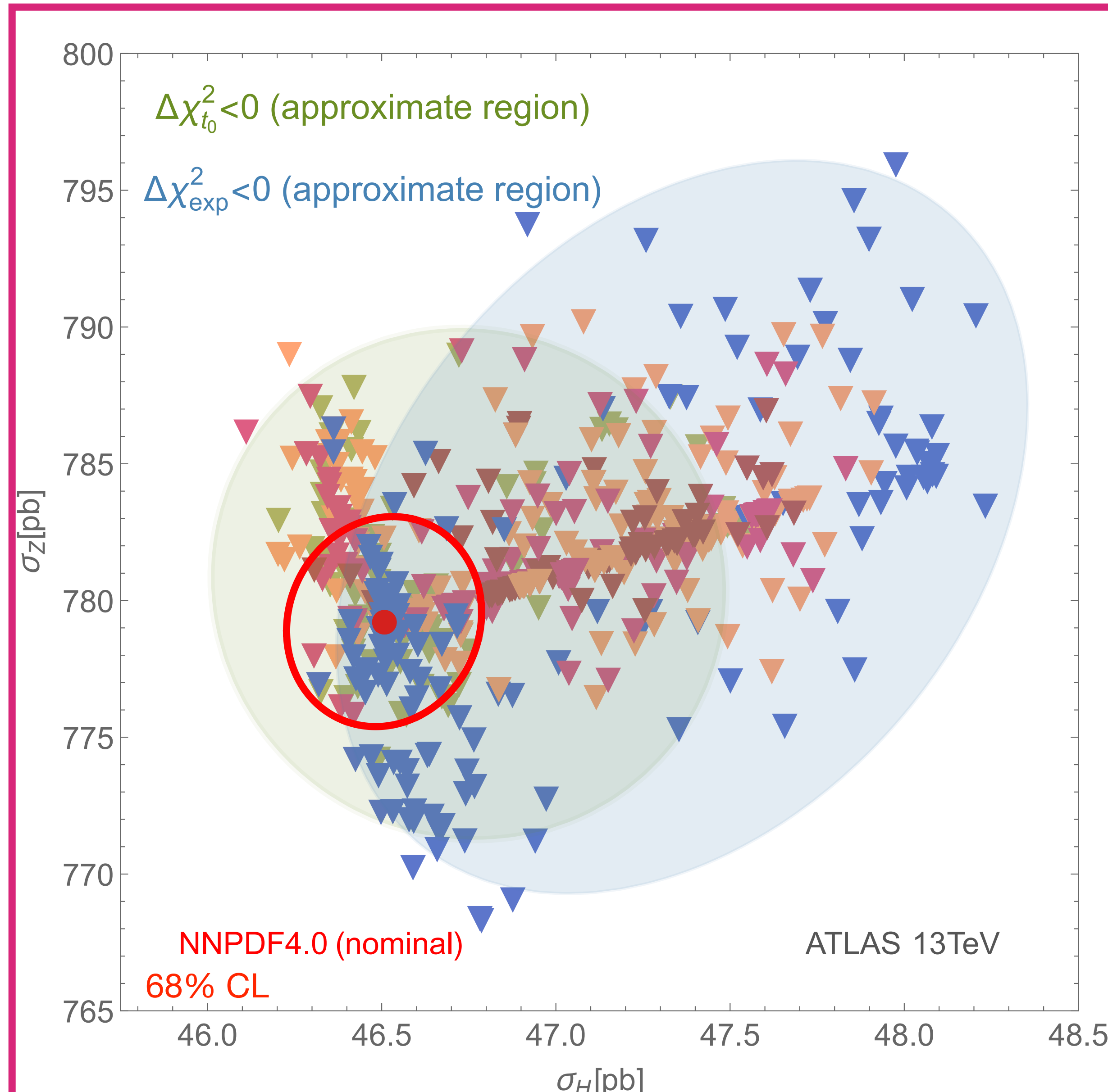


A hopscotch scan of LHC cross sections for NNPDF4.0 PDFs

The green and blue ellipses (constructed using a convex hull method) are approximate region containing all found replicas with $\Delta\chi^2 < 0$.

They have the statistical meaning related to a likelihood-ratio test.

The green and blue areas are larger than the nominal NNPDF4.0 uncertainty (red ellipse).

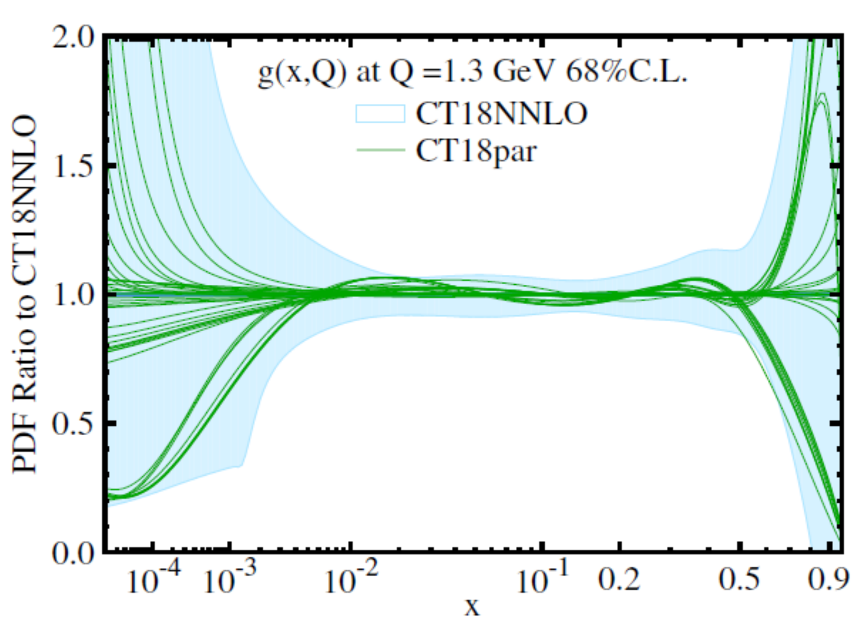


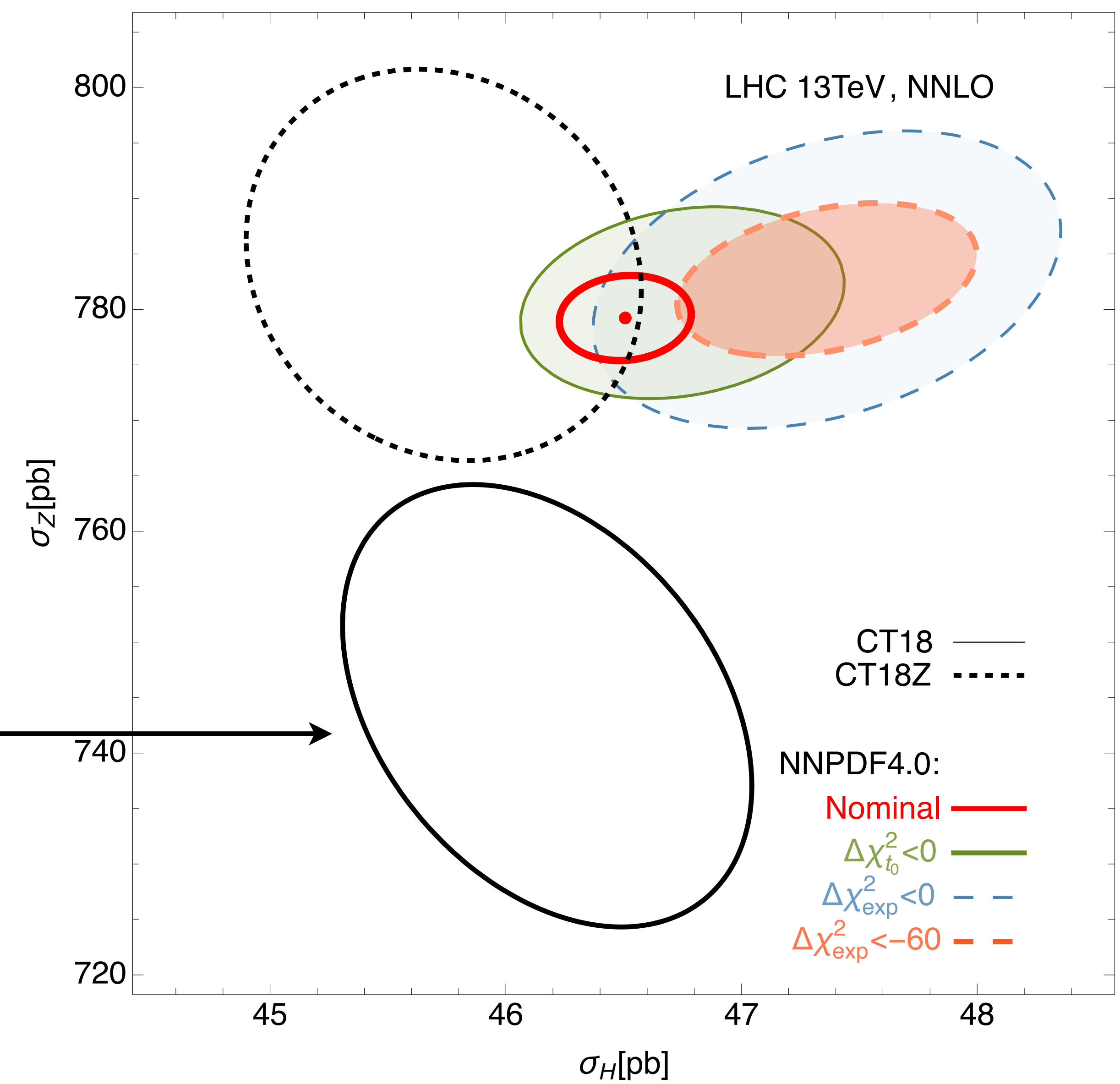
Representative sampling — the hopscotch algorithm

[AC, Huston, Nadolsky, Xie, Yan & Yuan, 2205.10444, PRD107]

The hopscotch algorithm more densely samples the effective dimensions relevant for the chosen observables. The resulting uncertainty is larger than the nominal one, shown here for (σ_H, σ_Z).

Monte Carlo uncertainties from sampling bias found through our dimensional reduction method play a similar role as sampling of parameter space in Hessian uncertainties.

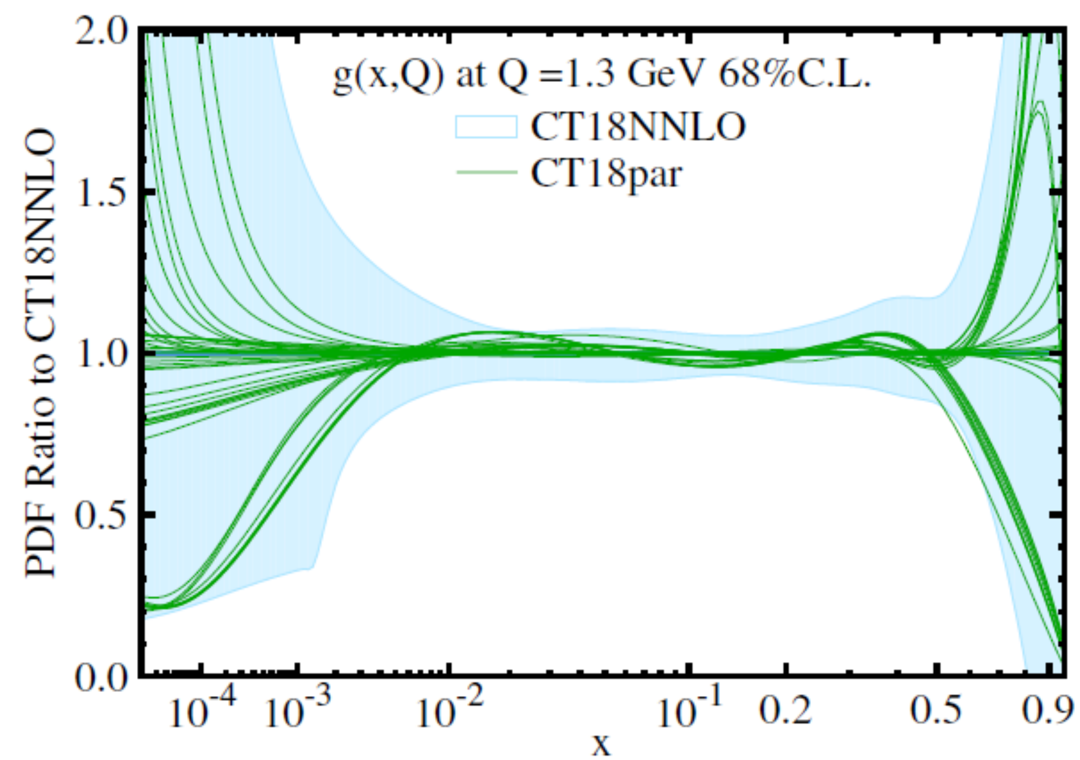




Epistemic uncertainties in global analyses

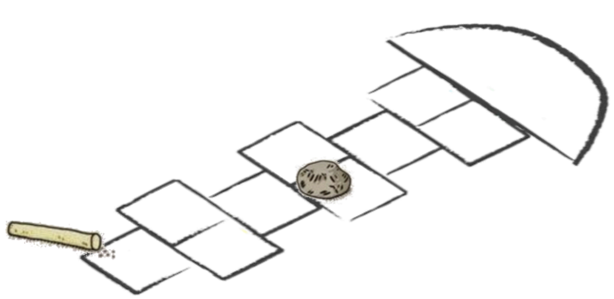
CT18 PDF uncertainty:

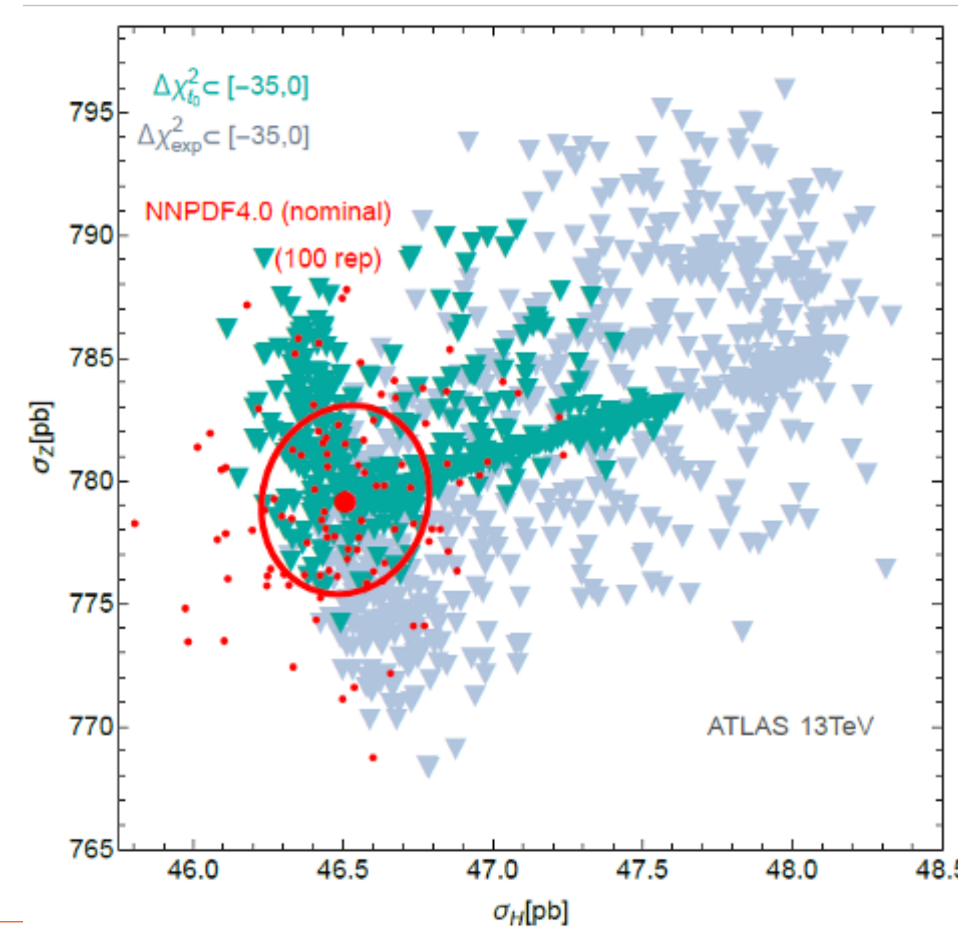
Hessian-based methodology.
Inclusive of sampling bias/lack of knowledge.



Monte Carlo-based PDF uncertainty:

Higher-dimensional space.
The "hopscotch" algorithm gives a lower bound
on the bias due to lack of knowledge.

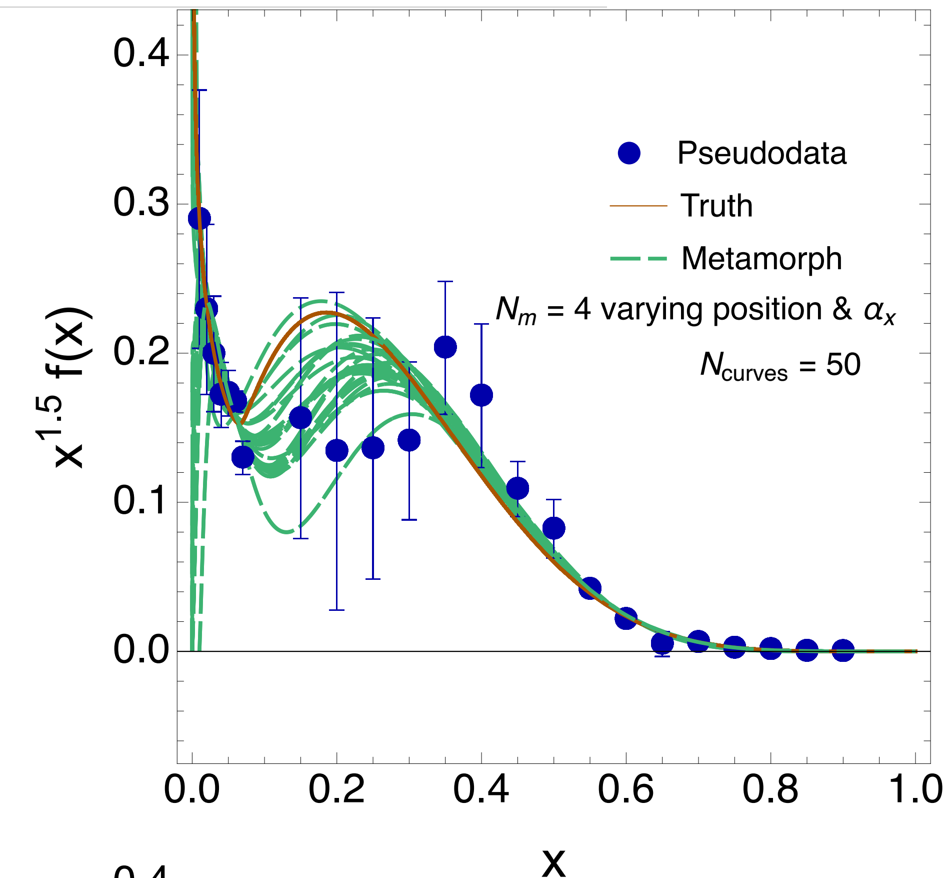




Epistemic uncertainties: Fantômas

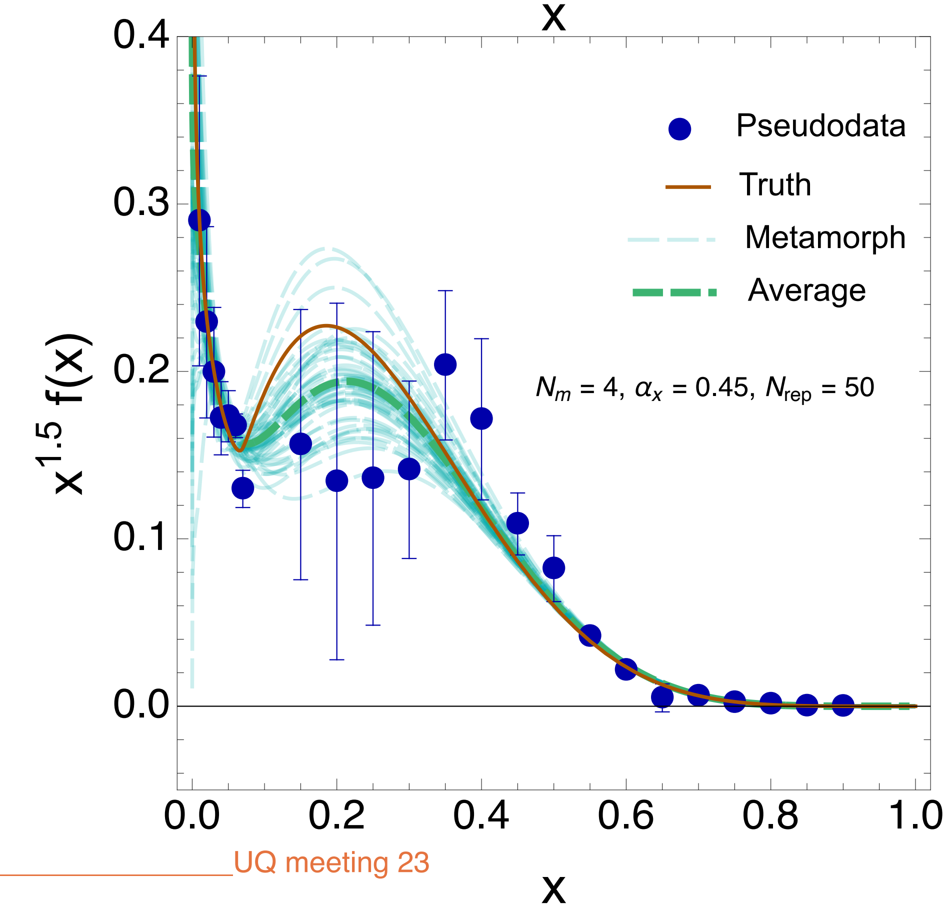
Epistemic uncertainty:

On-the-fly generation of functional forms for which the parameters of interest reflect the pulls on few specific points.



Aleatory uncertainty:

Here illustrated as a bootstrapped generation of replicas for a given functional form (probability distribution on the bundle of replicas).



[Kotz, Ponce-Chávez, AC, Nadolsky, Olness, very soon]

Hypothesis testing and parton distributions

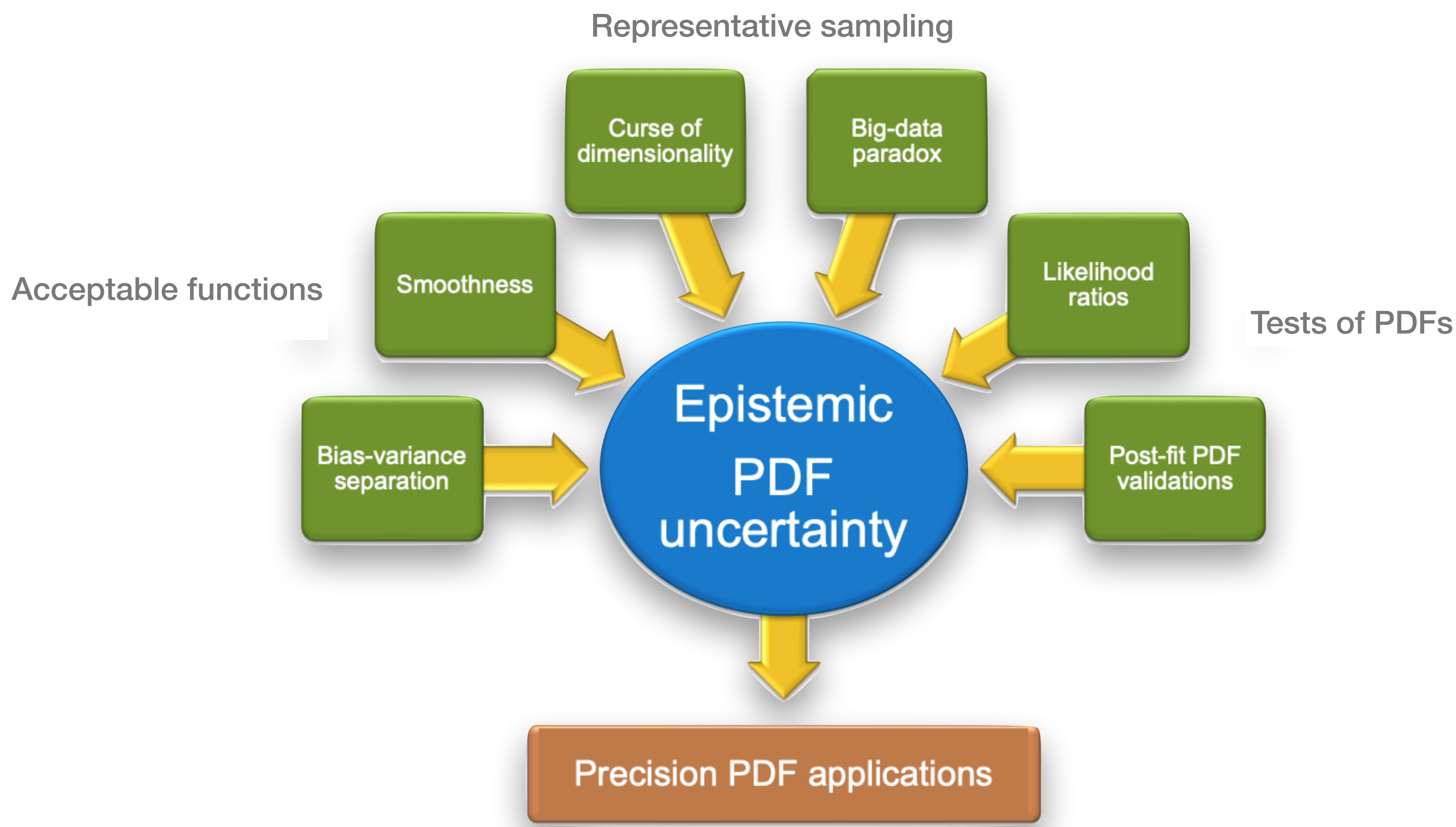


diagram by P. Nadolsky [DIS2023]

Quantification of the epistemic uncertainty

The hopscotch algorithm does not necessarily find the global minimum.

We currently cannot quantify the epistemic uncertainty, especially in the study case of the hopscotch.

Quantification of the epistemic uncertainty

The hopscotch algorithm does not necessarily find the global minimum.

We currently cannot quantify the epistemic uncertainty, especially in the study case of the hopscotch.

⇨ Scientific case for providing full statistical models for experimental data ; would affect correlation matrix for nuisance parameters.

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

[Cranmer et al, SciPost Phys.12, 037]

⇨ Case for clearly defining statistical models of global analyses: properly identify contributions to counfounding correlation for both MC and Hessian frameworks, role of priors.

Efforts that starts with our hopscotch study, will be expanded for CT2X release.

Can be generalized to broader PDF applications.

⇨ Epistemic uncertainty can only be optimized if it is understood —though irreducible in certain cases.

Likelihood and sampling — I

What is the adequate objective function for PDF analyses?

$$P(a|D) \propto P(D|a)\, P(a)$$

$$\Leftrightarrow \exp(-\chi^2_{\rm aug}/2) \propto \exp(-\chi^2/2)\, \exp(-\chi^2_{\rm prior}/2)$$

$$\Rightarrow \chi^2_{\rm aug} = \chi^2 + \chi^2_{\rm prior}$$

[Lepage et al., NPB Proc.Suppl.106(2002) 12-20]

Parameters: parameters of interest a and nuisance parameters.

Likelihood: “augmented” likelihood contains constraints/priors/penalties as well as the minimal likelihood. Identify priors on a.

The Bayesian statistics expression may require integration over a large space: needs for improved MC integrations — see F. Hickernell’s talk.

The possibility of using MC integration for expectation values was pointed out long ago, but the approach was deemed computationally inefficient.

Likelihood and sampling — II

On which basis are PDFs accepted or rejected?

Likelihood ratios:

two replicas can be ordered according to their relative likelihood or relative prior.

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

aleatory   epistemic + aleatory   probabilities

Prior: replica can be discarded based on $P(T_2) < P(T_1)$ even for $r_{likelihood} \sim 1$

Likelihood: replica can be accepted based on $r_{likelihood} = \dfrac{P(D\,|\,T_2)}{P(D\,|\,T_1)} \sim 1$ when $P(T_2) \sim P(T_1)$

Key role played by priors

Priors have been identified to reduce the phase space

Constraints fit exploits the benefits of well-controlled priors

"Constrained curve fitting," [Lepage et al., Nucl.Phys.B Proc.Suppl.106(2002) 12-20] — lattice oriented

⇨ similarly for polarized PDF analysis [Benel et al, EPJC]

Solutions may be prejudiced by strong priors

Some publications show how strong priors have affected results (that has led to important claims)

⇨ Proton structure: "Parton distributions need representative sampling"+ communication with NNPDF

[us]

⇨ Neutrino physics: "Neutrino mass and mass ordering: no conclusive evidence for normal ordering"

[Stefano Gariazzo et al JCAP10(2022)010]

Conclusions

A new avenue to the explain the differences among UQ from various PDF fitting groups is proposed through the study of the sampling uncertainties — a complementing source to the fitting uncertainty.

Highlights on the sampling uncertainties:

1. Epistemic uncertainty plays an important role in precise PDF determinations.
2. Concept of effective large dimensions. Difficult to sample the full parameter space with many parameters without biases. A **hopscotch scan** intelligently reduces dimensionality of the relevant PDF parameter space for an observable under consideration.
3. Define sources of counfounding correlations from the likelihood/statistical model. We learn and then aim to optimize it (where irreducible).
4. MC sampling for marginalization, more during this workshop

Moving toward epistemic PDF uncertainty!

Back up

A hopscotch scan of LHC cross sections for NNPDF4.0 PDFs

Step 1

The NNPDF4.0 Hessian set ($n = 50$) defines a coordinate system on a manifold corresponding to the largest variations of the PDF uncertainty —red dots and curve.

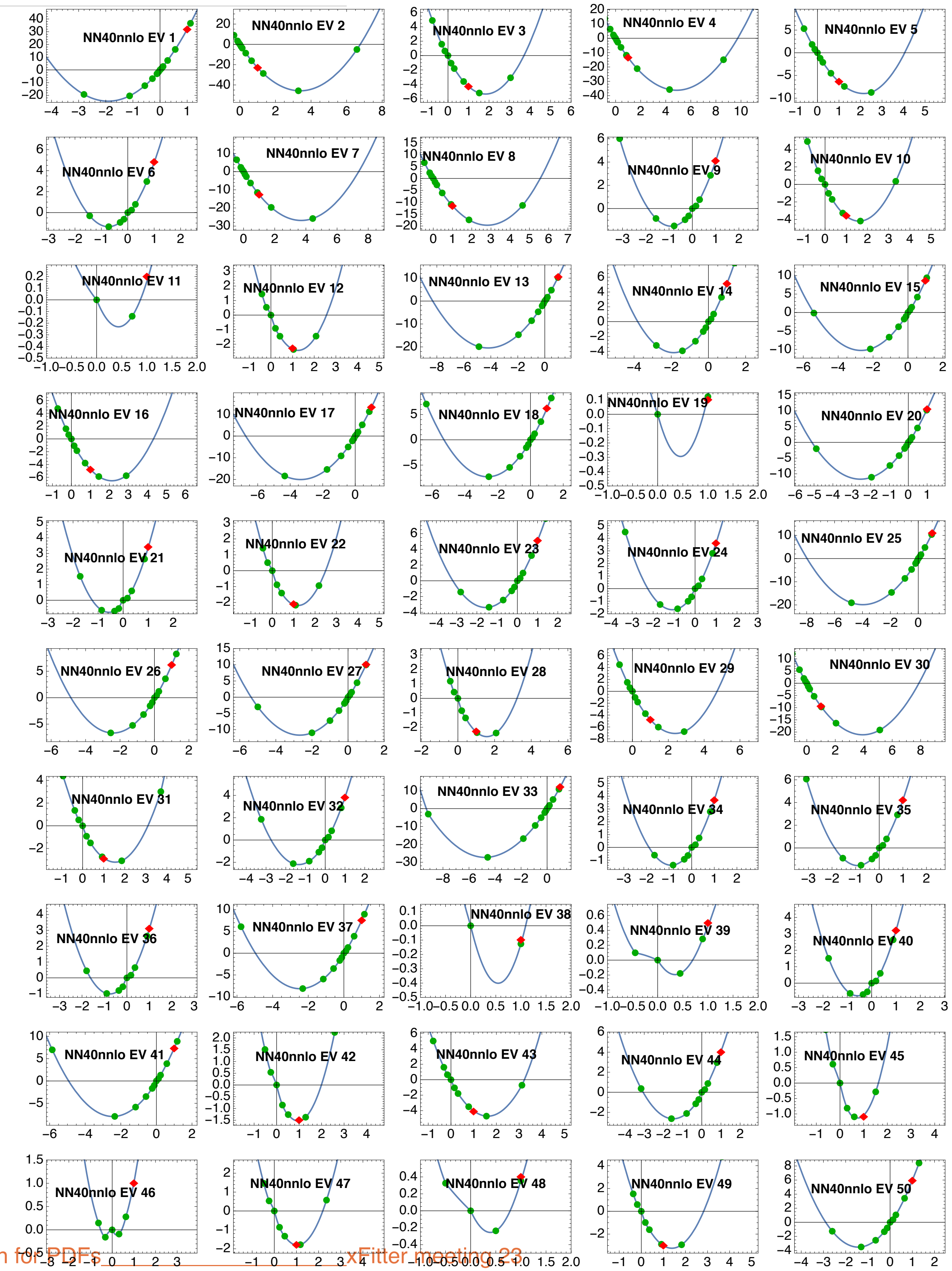
[NNPDF, 2109.02653]

Step 2

Using the public NNPDF code, scan χ^2_{tot} along the 50 EV directions to identify a hypercube corresponding to $\Delta\chi^2 \leq T^2$ (where $T^2 > 0$ is a user-selected value).

Lagrange multiplier scan confirms the approximate Gaussian profiles, but suggest that there exist solutions with lower χ^2 — green dots and blue curve.

No fitting involved.



A hopscotch scan of LHC cross sections for NNPDF4.0 PDFs

Step 1

The NNPDF4.0 Hessian set ($n = 50$) defines a coordinate system on a manifold corresponding to the largest variations of the PDF uncertainty —red dots and curve.

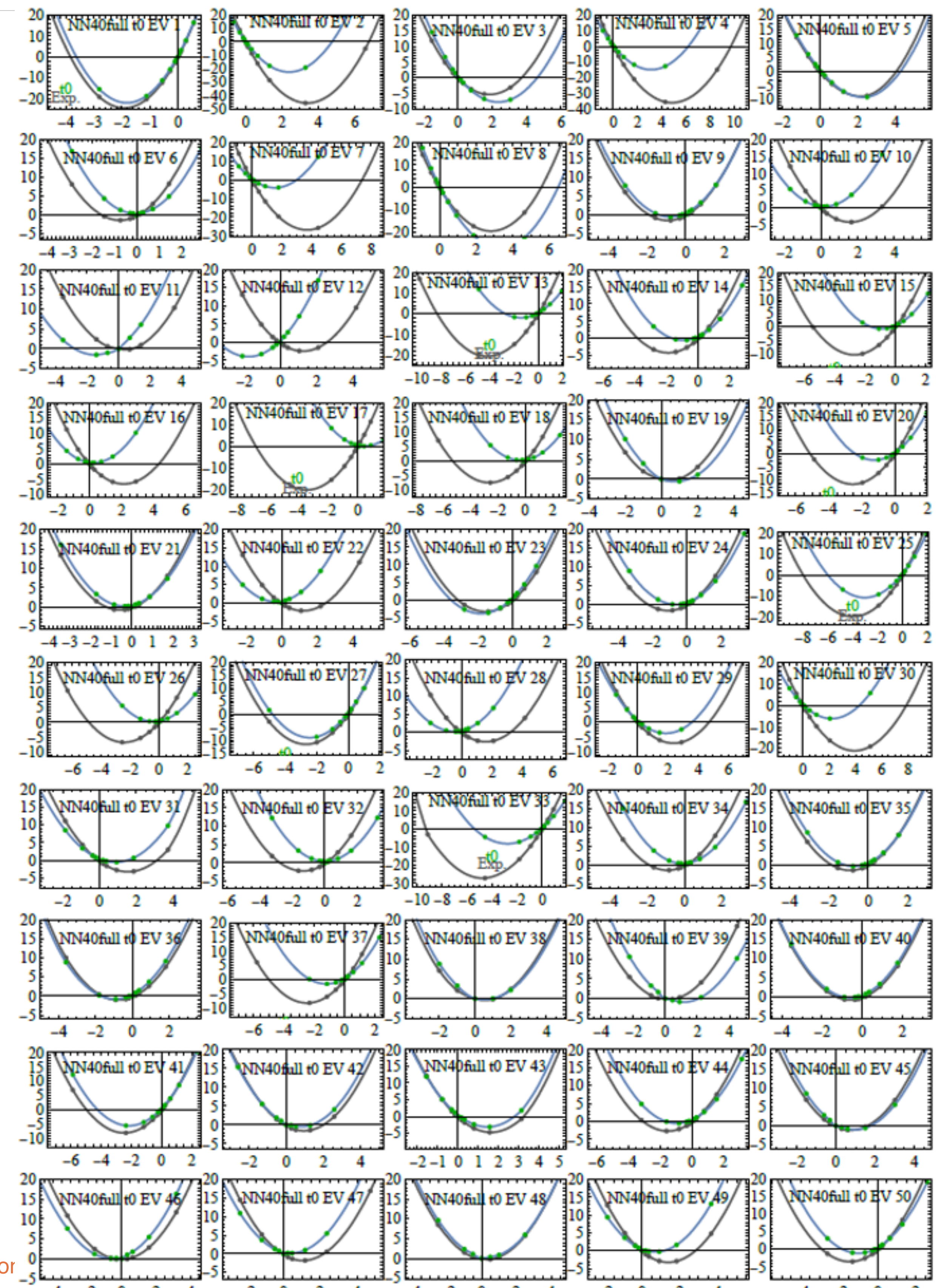
[NNPDF, 2109.02653]

Step 2

Using the public NNPDF code, scan χ^2_{tot} along the 50 EV directions to identify a hypercube corresponding to $\Delta\chi^2 \leq T^2$ (where $T^2 > 0$ is a user-selected value).

Lagrange multiplier scan confirms the approximate Gaussian profiles, but suggest that there exist solutions with lower χ^2 — green dots and blue curve.

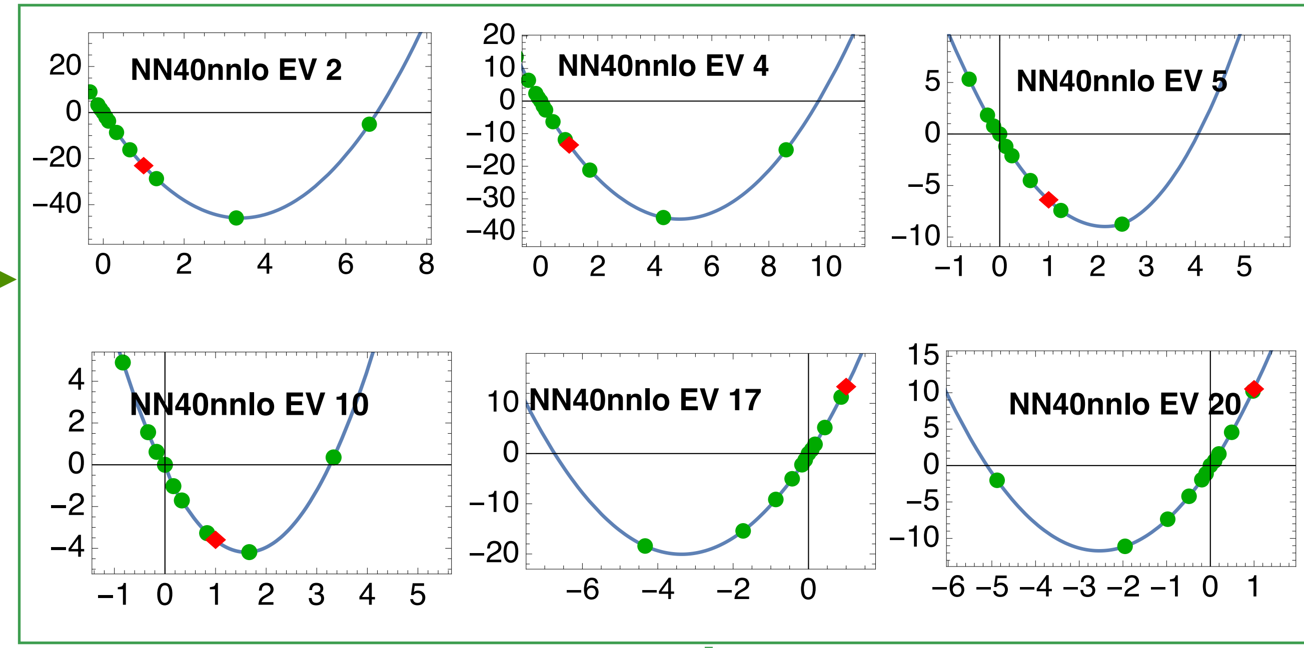
No fitting involved.



A hopscotch scan of LHC cross sections for NNPDF4.0 PDFs

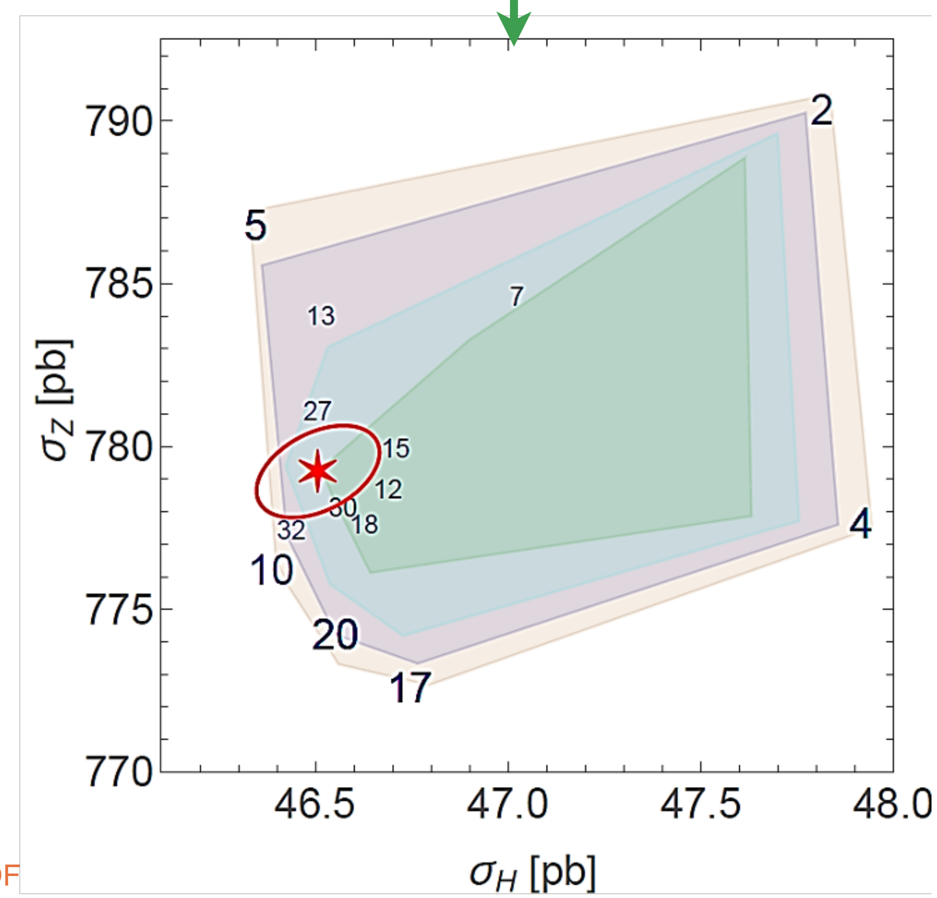
Step 3

Guidance from specific cross sections:
we identify 4-7 EV directions that give the largest displacements for a given $\Delta\chi^2$ per pair.



Large EV directions are shared among various pairs of cross sections.

Construct the convex hulls for $\Delta\chi^2 = +10, 0, -10, -20$ *w.r.t.* NNPDF4.0 replica 0 (red).



A hopscotch scan of LHC cross sections for NNPDF4.0 PDFs

<u>Step 4</u>

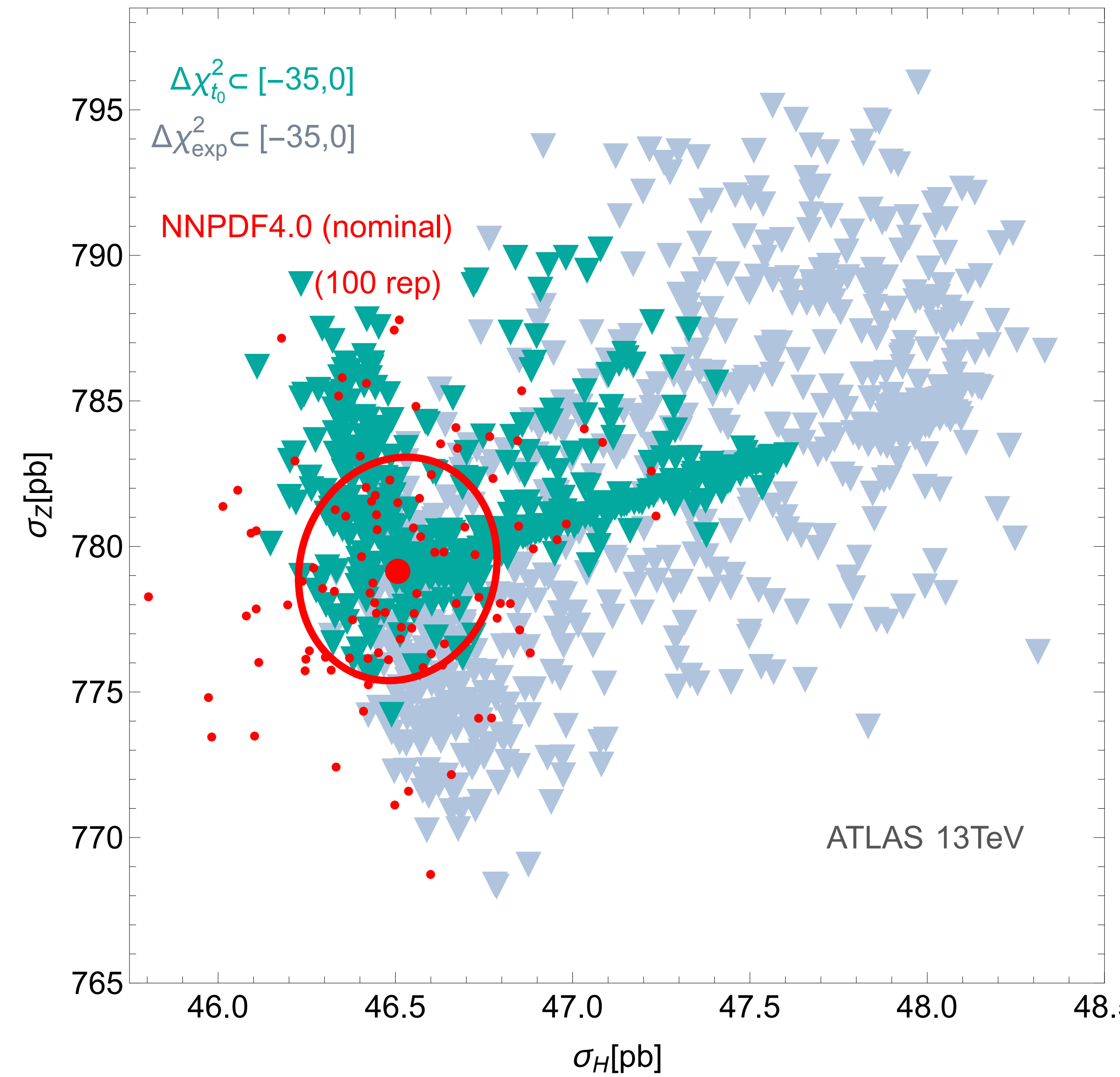
For each pair of cross sections, we generate 300 replicas by sampling uniformly along the "large" EV directions.

Sort the $n_{pairs} \times 300$ resulting replicas according to their $\Delta\chi^2$ *w.r.t.* to NN40 replica 0.

Hopscotch replicas are linear combinations of NNPDF4.0 Hessian EV.

Each of the solutions is an acceptable PDF set from the NNPDF4.0 fit.

High-density MC sampling of a span of a few EV directions that drive the specific PDF uncertainty.



Likelihoods in PDF analyses

$$\chi^2(a,\lambda^{exp}) = \sum_{i=1}^{N_{pts}} \left(\frac{T_i(a) - D_i + \sum_{\alpha=1}^{N_\lambda} \beta_{i,\alpha}\lambda_\alpha^{exp}}{\sigma_i} \right)^2 + \sum_{\alpha=1}^{N_\lambda} \lambda_\alpha^{exp,2}$$

Simple algebraic eq. $\frac{d}{d\lambda^{exp}}\chi^2(a,\lambda^{exp}) = 0 \Rightarrow \bar{\lambda}^{exp}$

$$\chi^2(a) = \sum_{i=1}^{N_{pts}} \left(\frac{T_i(a) - D_i + \sum_{\alpha=1}^{N_\lambda} \beta_{i,\alpha}\bar{\lambda}_\alpha^{exp}}{\sigma_i} \right)^2 + \sum_{\alpha=1}^{N_\lambda} \bar{\lambda}_\alpha^{exp,2} + \chi^2_{prior}(a)$$

a is the vector of parameters of interest

β is the correlation matrix for nuisance parameters

Figures of merit in the NNPDF4.0 analysis I

1. χ^2 with respect to the central experimental values

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

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

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

Experiments publish $\sigma_{i,\alpha}$. To reconstruct $\beta_{i,\alpha}$, we need to decide on the normalizations X_i.

NNPDF4.0 use:

a. $X_i = D_i$: "**exp**erimental scheme"; can result in a bias
b. $X_i = \text{fixed } T_i$: "$\boldsymbol{t_0}$ scheme"; can result in a (different) bias

Figures of merit in the NNPDF4.0 analysis II

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

NNPDF4.0 use:

a. $X_i = D_i$: **exp**erimental scheme; can result in a bias

b. $X_i = \text{fixed } T_i$: $\boldsymbol{t_0}$ scheme; can result in a (different) bias

The conventions are neither complete nor unique. Ambiguity affects all groups. See Appendix in 1211.5142.

2. **NNPDF4.0 trains MC replicas with χ^2 for fluctuated D_i, t_0 scheme, and replica selection (prior) conditions:**

$$\text{Cost} = \chi^2_{t_0}(T_i, D_i^{fluctuated}) + \chi^2_{prior}$$

3. NNPDF4.0 quotes the final unfluctuated χ^2 in the "exp" scheme.

Experimental scheme:
$\chi^2_{tot}/N_{pt} = 1.160\,.$

t_0 scheme:
$\chi^2_{tot}/N_{pt} = 1.233\,.$

$\chi^2(\text{exp}) - \chi^2(t_0) = -340$ for 4618 data points

The hopscotch scan counterbalances the bias of the nominal replica ensemble

6.2 Creating a less biased sub-sample

The basic idea is to use such partial information about the selection bias to design a *biased* sub-sampling scheme to *counterbalance* the bias in the original sample, such that the resulting sub-samples have a *high likelihood* to be less biased than the original sample from our target population. That is, we create a sub-sampling indicator S_I, such that with high likelihood, the correlation between $S_I R_I$ and G_I is reduced, compared to the original $\rho_{R,G}$, to such a degree that it will compensate for the loss of sample size and hence reduce the MSE of our estimator (e.g., the sample average). We say with *high likelihood*, in its non-technical meaning, because without full information on the response/recording mechanism, we can never guarantee such a counterbalance sub-sampling (CBS) would always do better. However, with judicious execution, we can reduce the likelihood of making serious mistakes.

X.-L. Meng, Survey Methodology, Catalogue 12-001-X, vol. 48 (2022), #2

Priors and optimal sampling parameters

statistical estimate of an arbitrary function of the parameters using

$$\langle f(\rho)\rangle = B^{-1}\int \mathrm{e}^{-\chi^2_{\mathrm{aug}}(\rho)/2} f(\rho)\, d^n\rho \quad (15)$$

where

$$B \equiv \int \mathrm{e}^{-\chi^2_{\mathrm{aug}}(\rho)/2}\, d^n\rho, \quad (16)$$

and the variance is $\sigma_f^2 \equiv \langle f^2\rangle - \langle f\rangle^2$, as usual. In practice these integrals are quite difficult to evaluate for all but the simplest of fits. This is because $P(\rho|\overline{G})$ is typically very sharply peaked about its maximum. For smaller problems, adaptive Monte Carlo integrators, such as `vegas`, are effective. For larger problems Monte Carlo simulation techniques, such as the Metropolis or hybrid Monte Carlo methods, can be effective. Still the cost of evaluating the integrals is often prohibitive, particularly when there are lots of poorly constrained parameters (which lead to long, narrow, high ridges in the probability distribution). Consequently efficient approximations are useful.

The distribution obtained from this modified bootstrap algorithm is not precisely the Bayes distribution $P(\rho|\overline{G})$. It has additional factors such as $\sqrt{\det g_{ij}}$ where

$$g_{ij} \equiv \sum_{t,t'} \sigma^{-2}_{t,t'} \frac{\partial G(t;\rho)}{\partial \rho_i}\frac{\partial G(t';\rho)}{\partial \rho_j} \quad (19)$$

is a metric induced on ρ space [5]. These factors become constants for sufficiently high statistics and so make no difference in that limit. This particular factor is interesting, however, because it makes the measure in ρ space invariant under reparameterizations. This suggests that

$$P'(\rho|\overline{G}) \propto \sqrt{\det g_{ij}}\, \mathrm{e}^{-\chi^2_{\mathrm{aug}}/2} \quad (20)$$

might be a better choice for our Bayesian probability.

"Constrained curve fitting," [Lepage et al., Nucl.Phys.B Proc.Suppl.106(2002) 12]

The possibility of using MC integration for expectation values was pointed out long ago, but the approach was deemed computationally inefficient.

Quasi-MC integration and dimensionality reduction may help, as well as parameter transformations to sample using a non-informative (e.g., Jeffrey's) prior