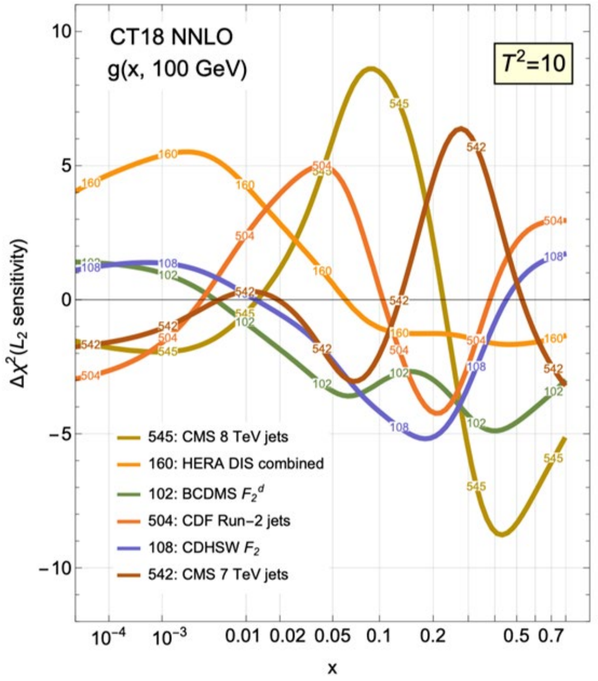
Uncertainty quantification for PDFs and machine learning

Pavel Nadolsky
Southern Methodist University



Based on

- **arXiv:2306.03918, arXiv:2205.10444 [PRD 107 (2023) 3, 034008]**
  with A. Cooper-Sarkar, A. Courtoy, T. Cridge, F. Giuli, L. Harland-Lang, T.J. Hobbs, J. Huston, X. Jing, R. S. Thorne, K. Xie, M. Yan, C.-P. Yuan

- studies with CTEQ-TEA, PDF4LHC working groups and Snowmass'21 EF06 Topical group

CTEQ

2023-06-15

What determines the size of PDF uncertainties?

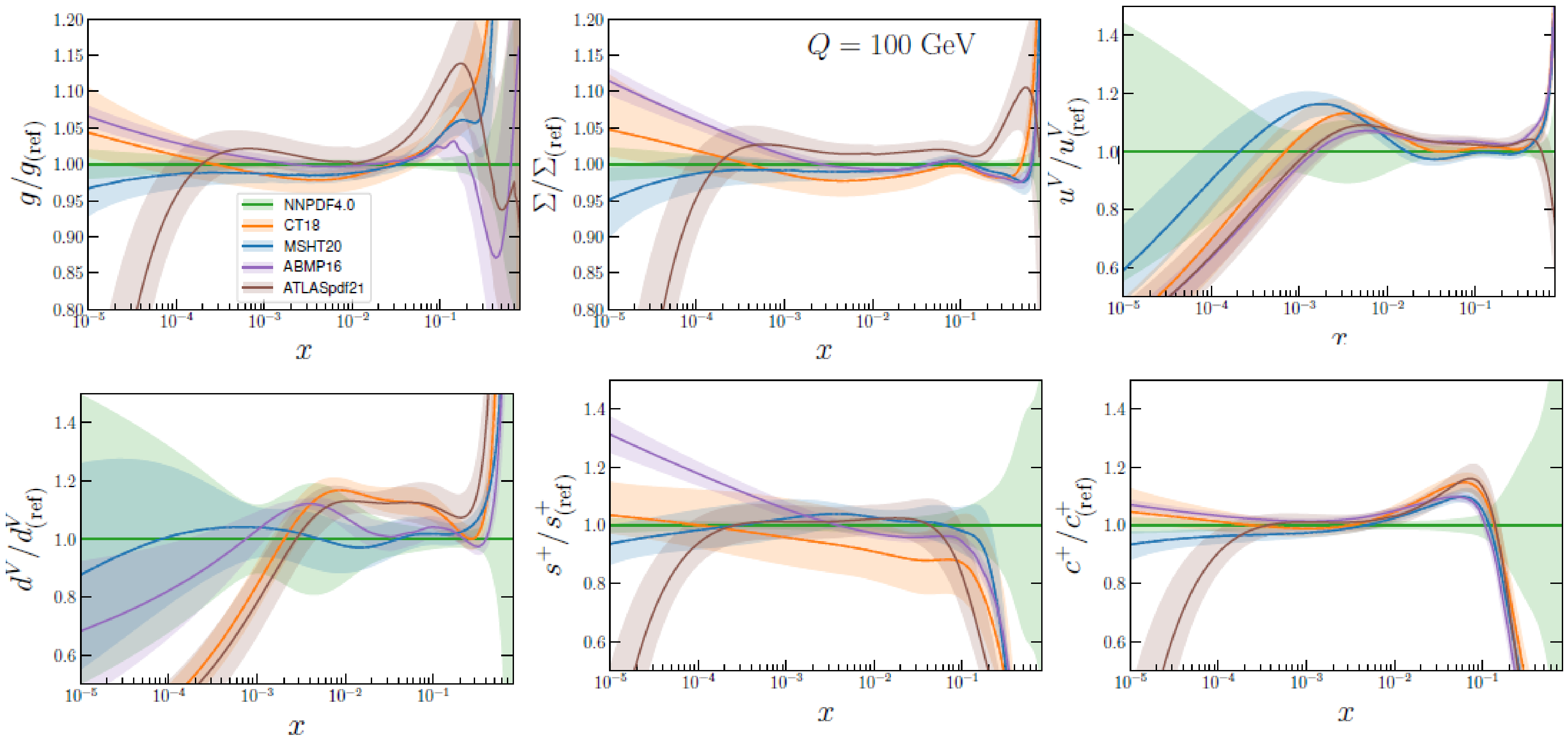
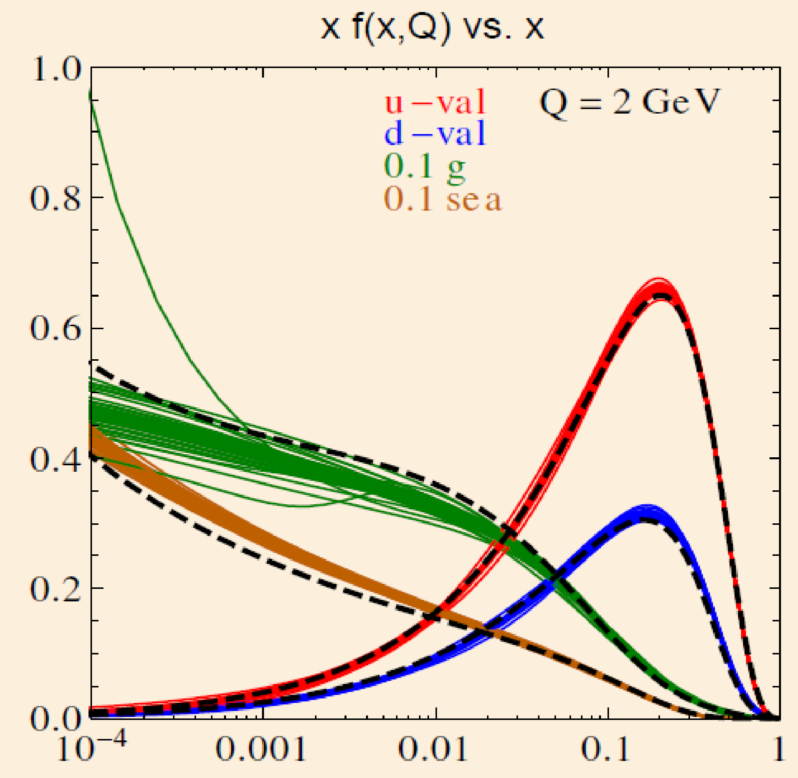


FIG. 2. Comparison of the PDFs at $Q = 100$ GeV. The PDFs shown are the N2LO sets of NNPDF4.0, CT18, MSHT20, ABMP16 with $\alpha_s(M_Z) = 0.118$, and ATLASpdf21. The ratio to the NNPDF4.0 central value and the relative 1σ uncertainty are shown for the gluon g, singlet Σ, total strangeness $s^+ = s +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

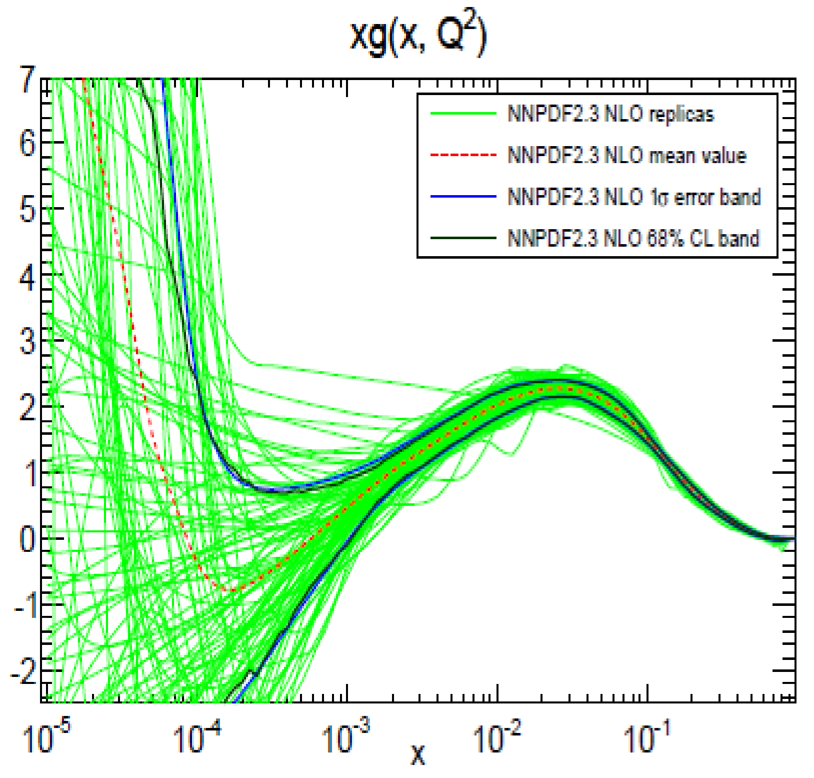
Figures from Snowmass'2021 whitepaper
"Proton structure at the precision frontier", arXiv:2203.13923

Two types of modern error PDFs

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



Neural network parameterizations
+ Monte Carlo PDF replicas
(NNPDF)

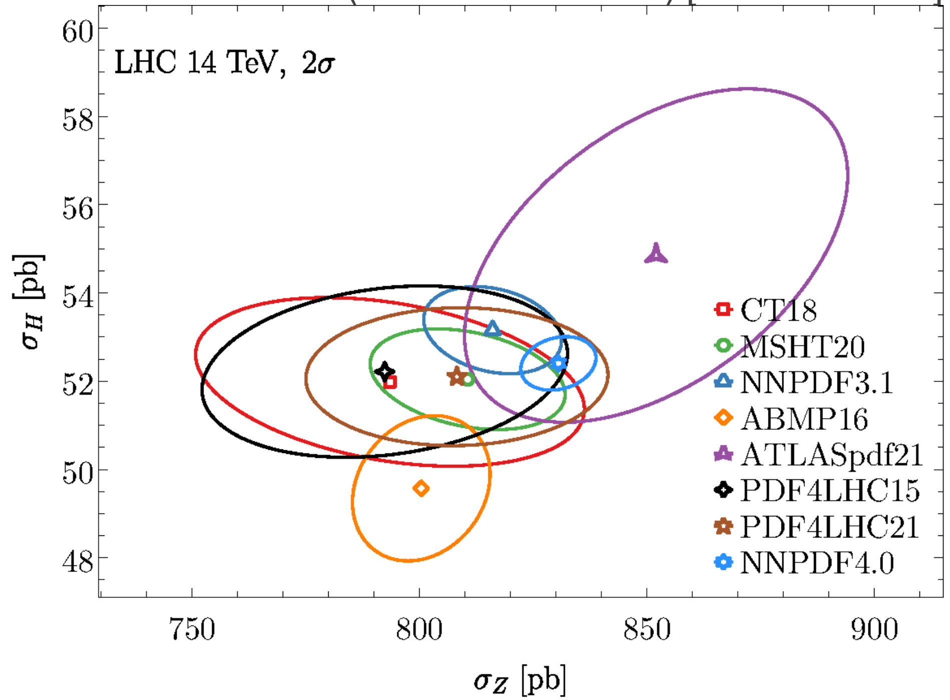


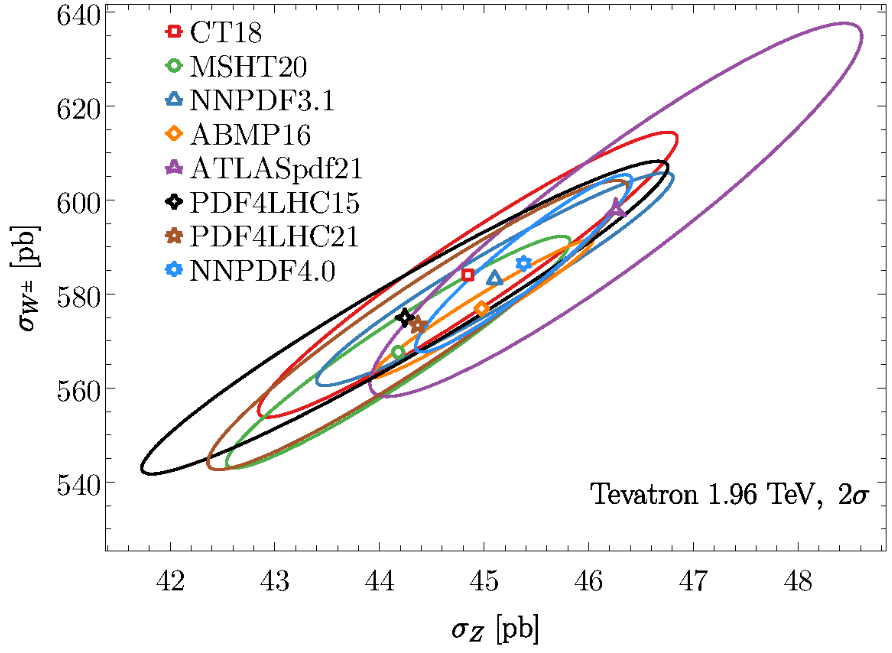
Two powerful, complementary representations.
Hessian PDFs can be converted into MC ones, and vice versa.

The tolerance puzzle

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

Precision PDFs (Snowmass 21 WP) [2203.13923v2]

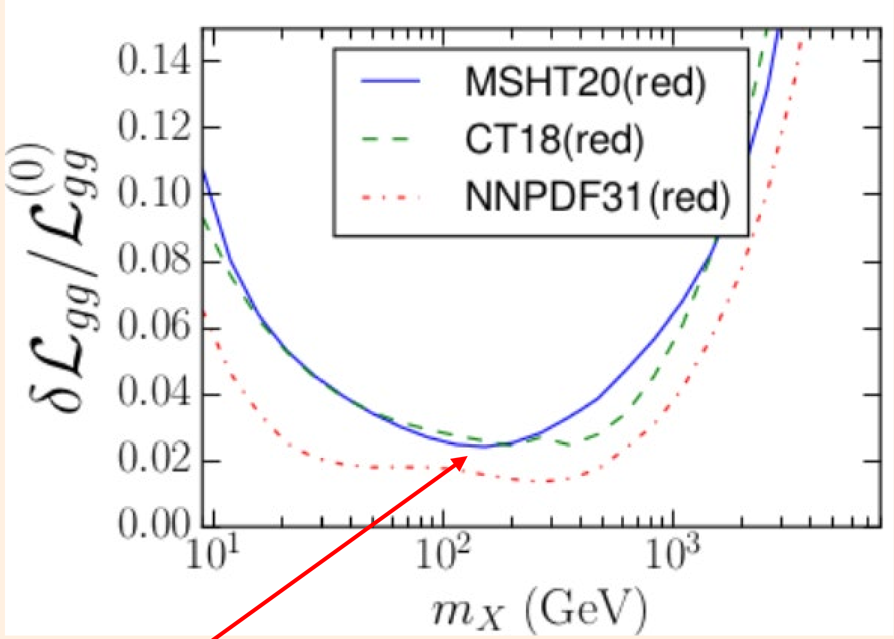




The answer has direct implications for high-stake experiments such as *W* boson mass measurement, tests of nonperturbative QCD models and lattice QCD, high-mass BSM searches, etc.

Tolerances explained by epistemic uncertainties

Relative PDF uncertainties on the gg luminosity at 14 TeV in three PDF4LHC21 fits to the **identical** reduced global data set



$\times$ 1.5 – 2 difference

While the fitted data sets are identical or similar in several such analyses, the differences in uncertainties can be explained by methodological choices adopted by the PDF fitting groups.

NNPDF3.1' and especially 4.0 (based on the NN's+ MC technique) tend to give smaller nominal uncertainties in data-constrained regions than CT18 or MSHT20

Epistemic uncertainties explain many such differences

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

Aspects of PDF uncertainties

1. **(Dis)agreement among fitted experimental data sets**
    - χ^2 **tensions,** L_2 **sensitivity**
2. **Modeling of systematic uncertainties**
3. Explicit and implicit priors in Hessian and NN/ML fits

associated with the **epistemic** uncertainty

explore using publicly available error PDFs and codes

Aspects of PDF uncertainties

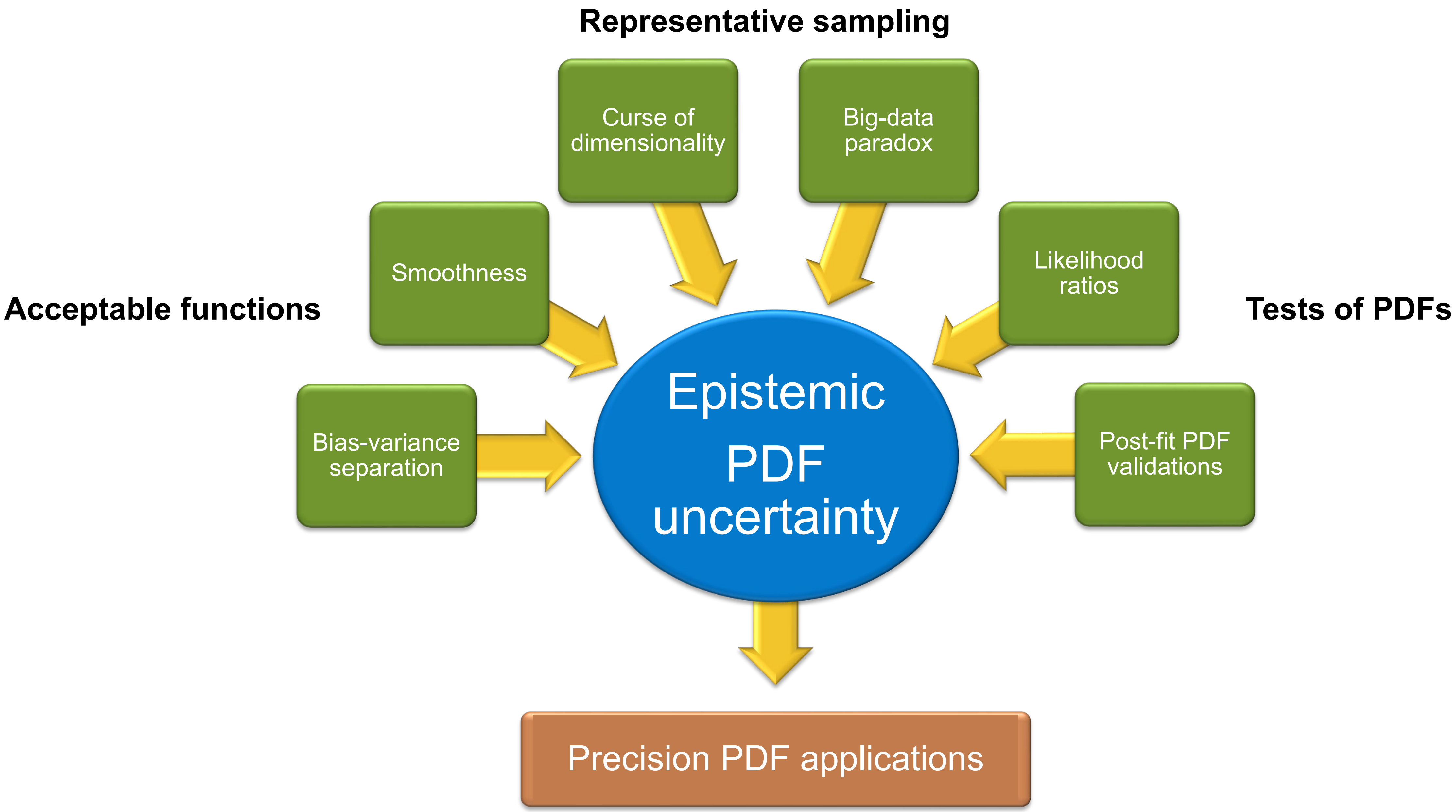
1. **(Dis)agreement among fitted experimental data sets**
   - χ^2 tensions, L_2 sensitivity
2. **Modeling of systematic uncertainties**
3. Explicit and implicit priors in Hessian and NN/ML fits

associated with the **epistemic** uncertainty

epistemic vs. **aleatory** uncertainties

Statistical uncertainty propagated from experiments — reduced by increasing data size

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


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

Epistemic PDF uncertainty…

…reflects **methodological choices** such as PDF functional forms or NN architecture and hyperparameters.

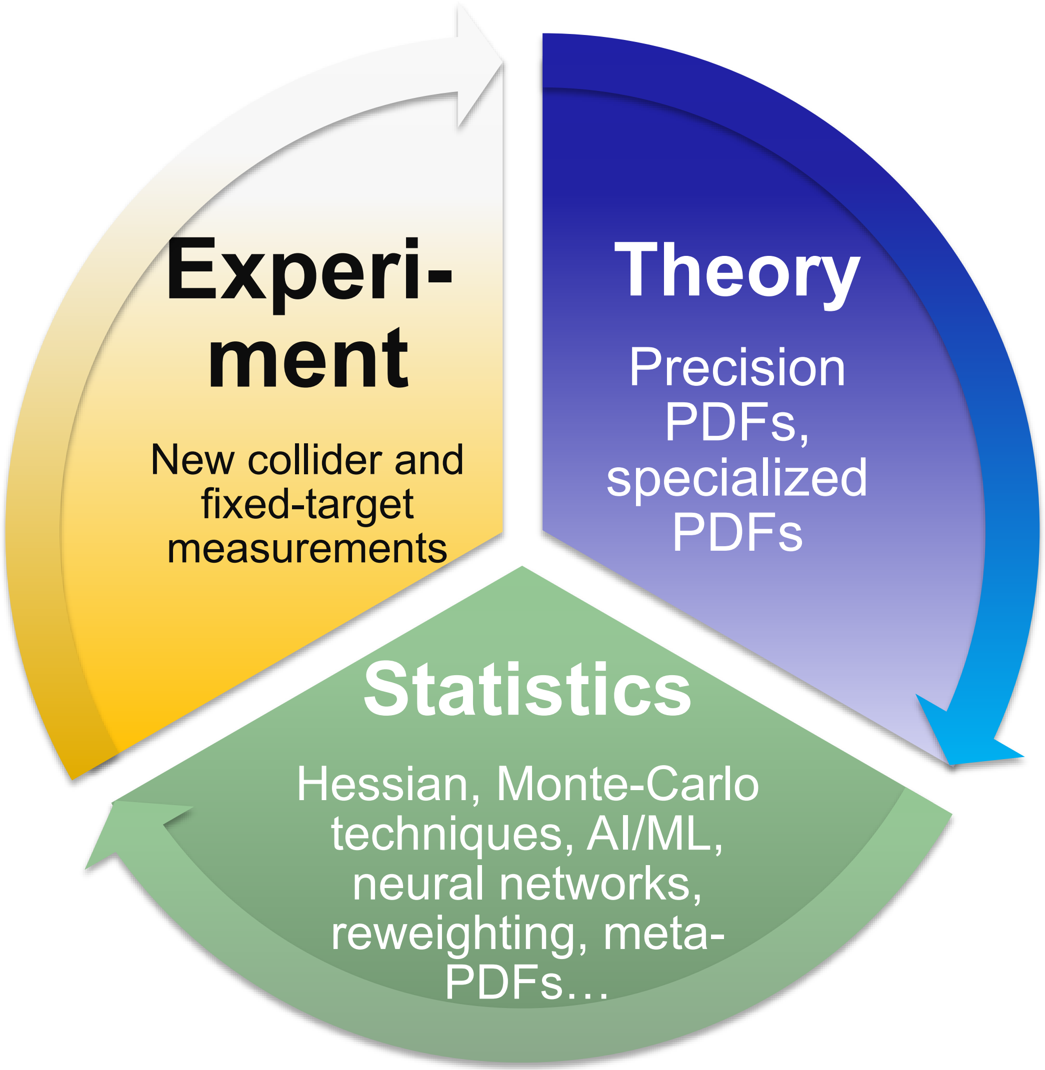
… can dominate the full uncertainty when experimental and theoretical uncertainties are small.

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

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

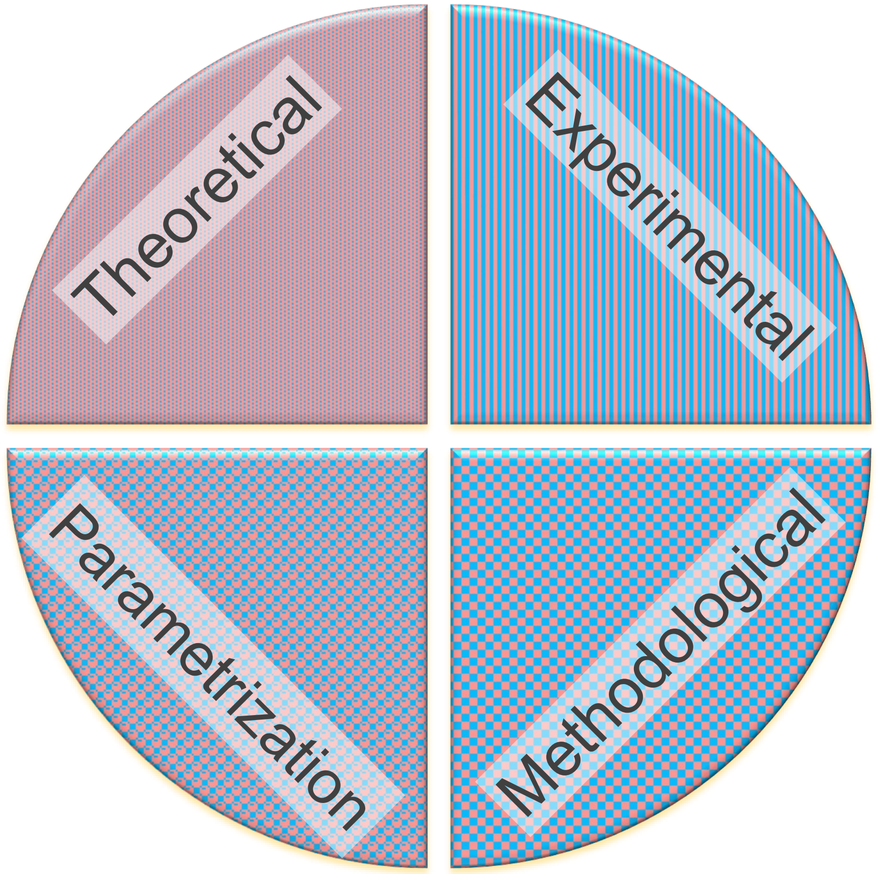
⇒ sampling over choices of experiments, PDF/NN functional space, models of correlated uncertainties…

⇒ in addition to sampling over data fluctuations



Components of a global QCD fit

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]

HEP is not alone

Various domains contend with **multi-dimensional non-probability samples**

Forecasting: presidential elections, financial markets, weather and climate, …
Meng, The Annals of Applied Statistics, 12(2), 685; Isakov and Kuriwaki, Harvard Data Science Review, 2(4), 2020

Political polling
M. R. Elliott, R. Valliant, Statistical Science, 32(2), 249 (2017)
M. A. Bailey, Polling at a Crossroads – Rethinking Modern Survey Research. Cambridge University Press, 2023

COVID-19 vaccination assessments and epidemiological studies
Bradley et al., https://doi.org/10.1038/s41586-021-04198-4
W. Dempsey, arXiv:2005.10425

Clinical trials of medical treatments
P. Msaouel, https://doi.org/10.1080/07357907.2022.2084621

Studies of biodiversity
R. Boyd et al., https://doi.org/10.1016/j.tree.2023.01.001

…

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

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

A common resampling procedure used by experimentalists and theorists:

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

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

A likelihood-ratio test of NN models T_1 and T_2

From Bayes theorem, it follows that

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

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

aleatory epistemic + aleatory probabilities

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.

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

ATLAS-CONF-2023-004

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

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

ATLAS-CONF-2023-015

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

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

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

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

Also the next slide.

1. Tensions among experiments

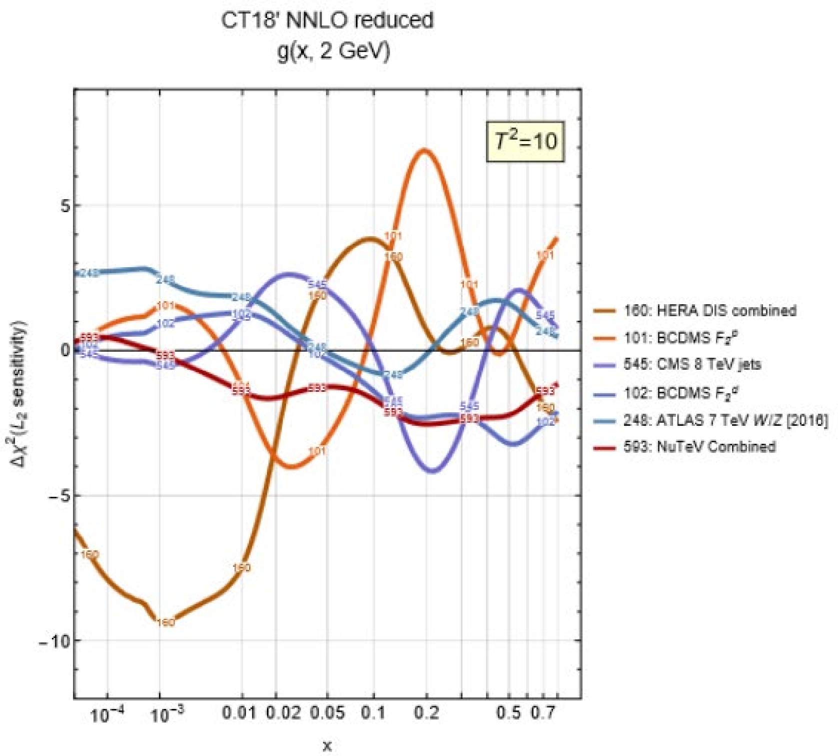
Explore using the L_2 sensitivity

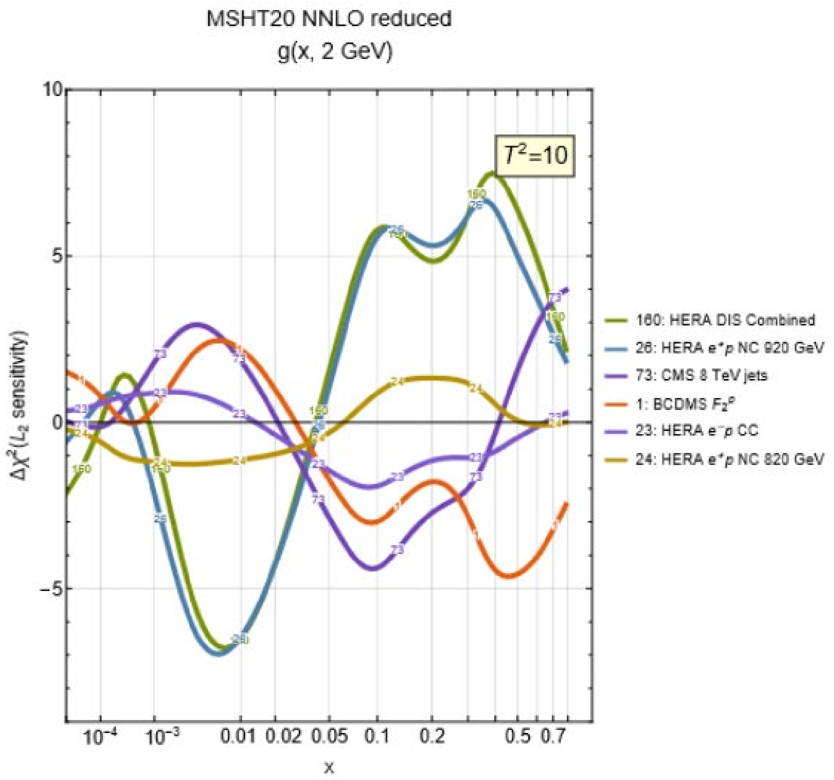
for Hessian PDFs

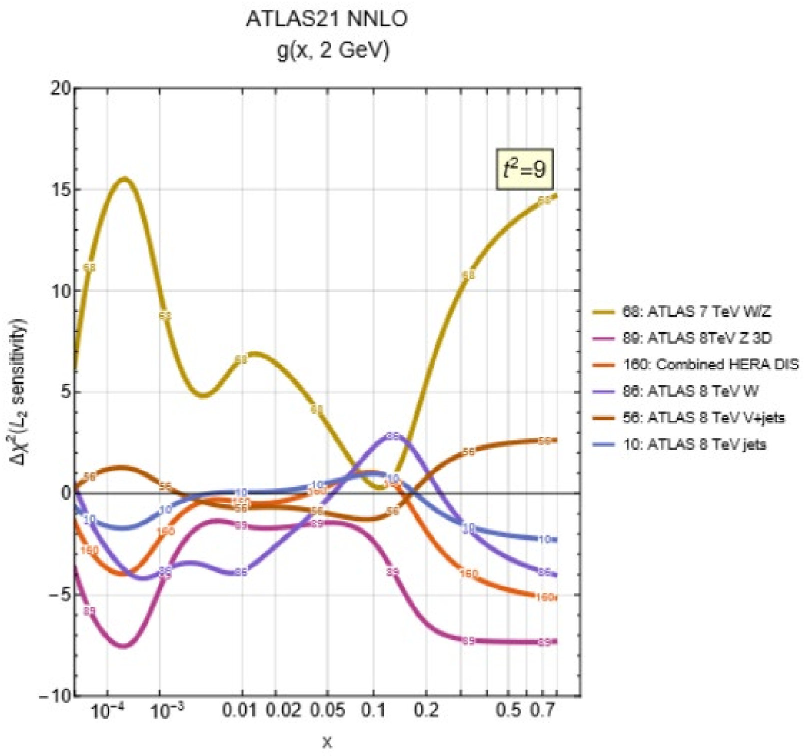
arXiv:2306.03918
by X. Jing, A. Cooper-Sarkar, A. Courtoy, T. Cridge, F. Giuli, L. Harland-Lang, T.J. Hobbs, J. Huston, P. N., R. S. Thorne, K. Xie, C.-P. Yuan

An ATLAS, CTEQ-TEA, and MSHT comparative study of NNLO and aN3LO PDF sensitivities

X.Jing et al.,arXiv:2306.03918







- Comparisons of strengths of constraints from individual data sets in 8 PDF analyses using the common L_2 sensitivity metric
- An interactive website (https://metapdf.hepforge.org/L2/) to plot such comparisons

[2070 figures in total]

Tolerance hypersphere in the PDF space

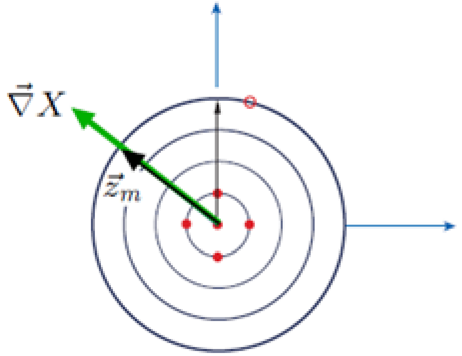
2-dim (i,j) rendition of N-dim (26) PDF parameter space

Hessian method: Pumplin et al., 2001

A symmetric PDF error for a physical observable X is given by

$$\Delta X = \vec{\nabla} X \cdot \vec{z}_m = \left|\vec{\nabla} X\right|$$

$$= \frac{1}{2}\sqrt{\sum_{i=1}^{N}\left(X_i^{(+)} - X_i^{(-)}\right)^2}$$



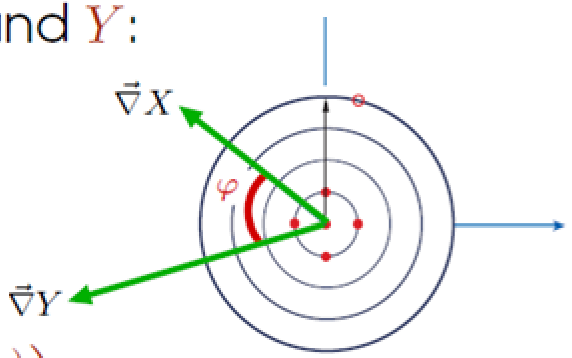
(b)
Orthonormal eigenvector basis

Correlation cosine for observables X and Y:

hep-ph/0110378

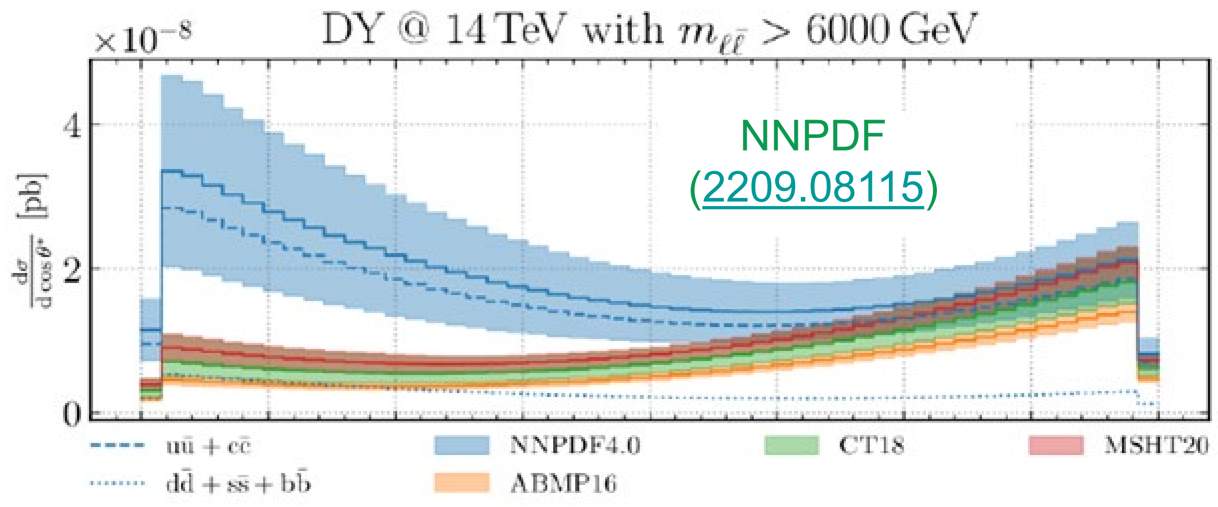
$$\cos\varphi = \frac{\vec{\nabla} X \cdot \vec{\nabla} Y}{\Delta X \Delta Y} =$$

$$\frac{1}{4\Delta X\,\Delta Y}\sum_{i=1}^{N}\left(X_i^{(+)} - X_i^{(-)}\right)\left(Y_i^{(+)} - Y_i^{(-)}\right)$$

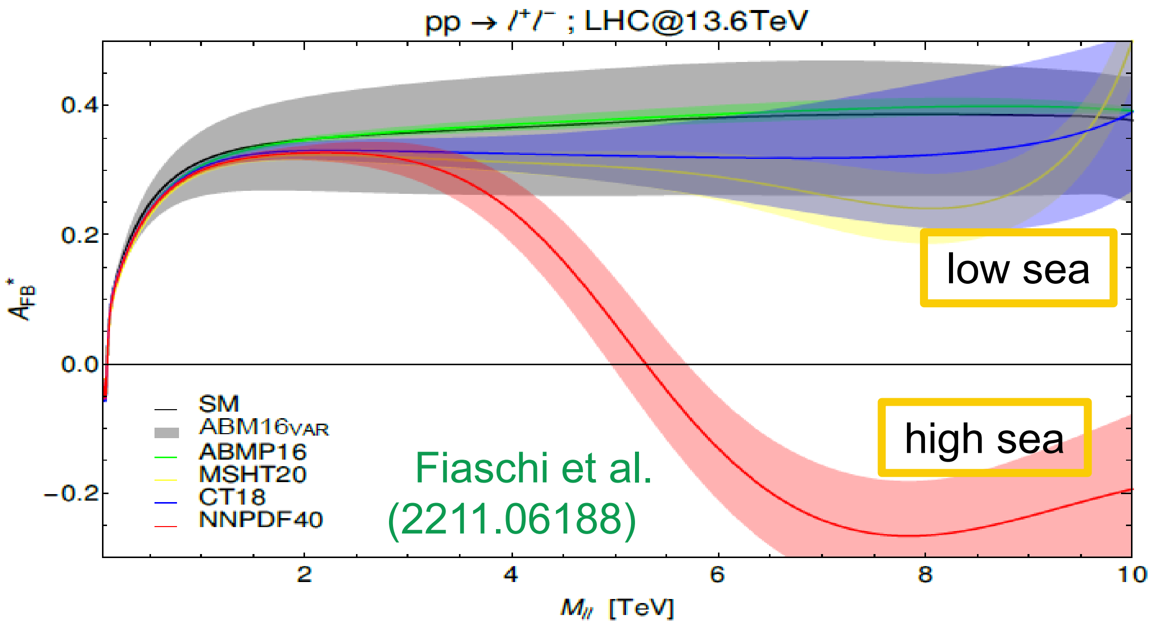


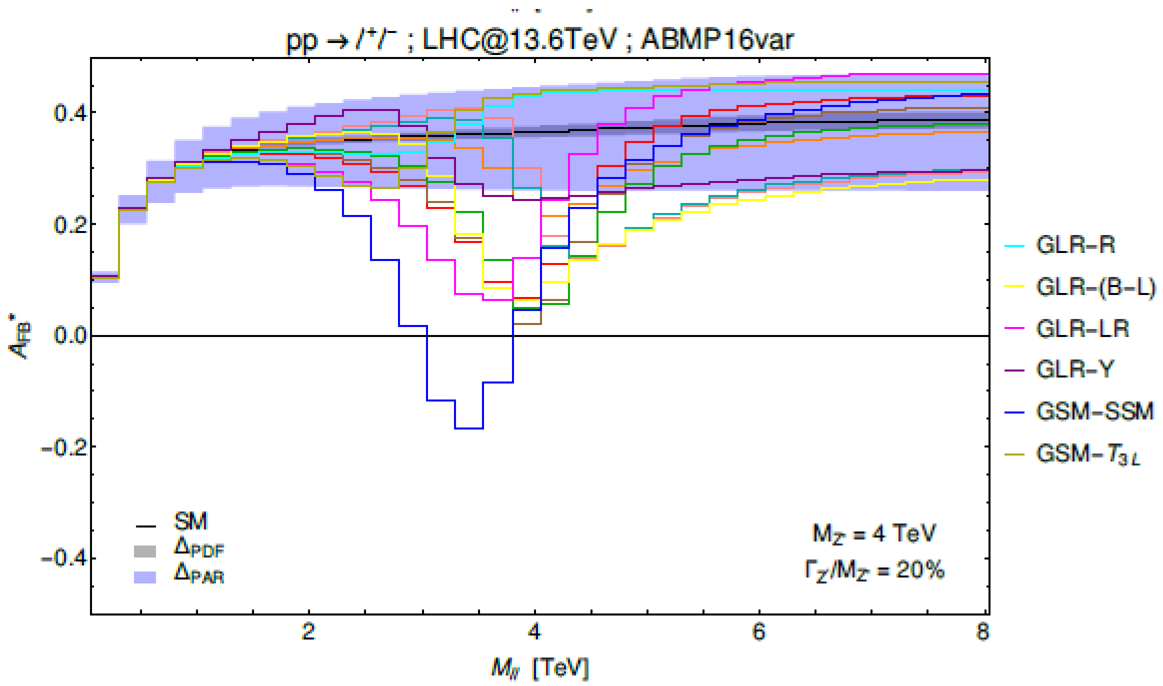
(b)
Orthonormal eigenvector basis

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



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



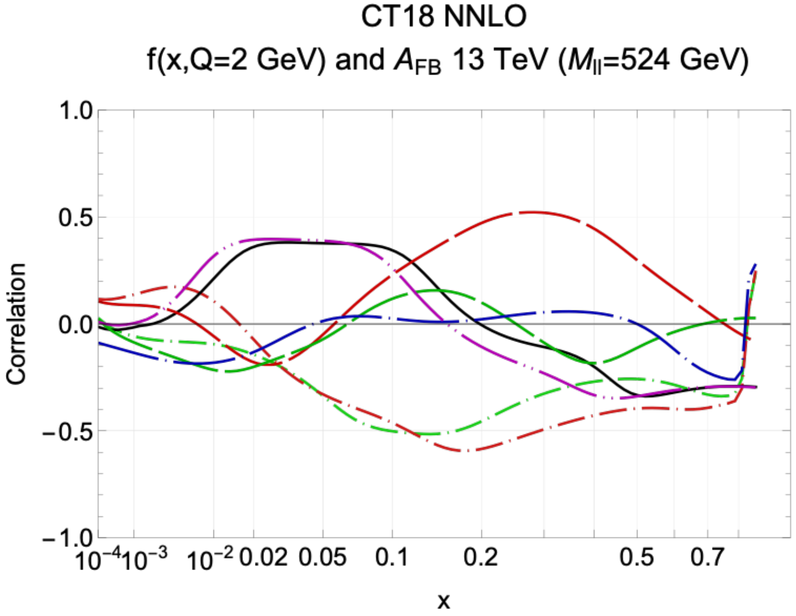


Relevant in searches for BSM resonances

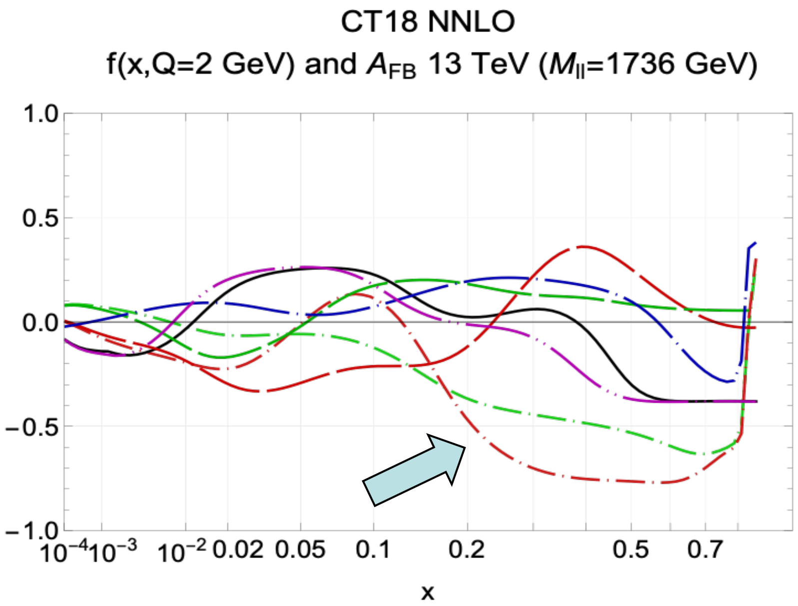
PDF correlations for A_{FB}

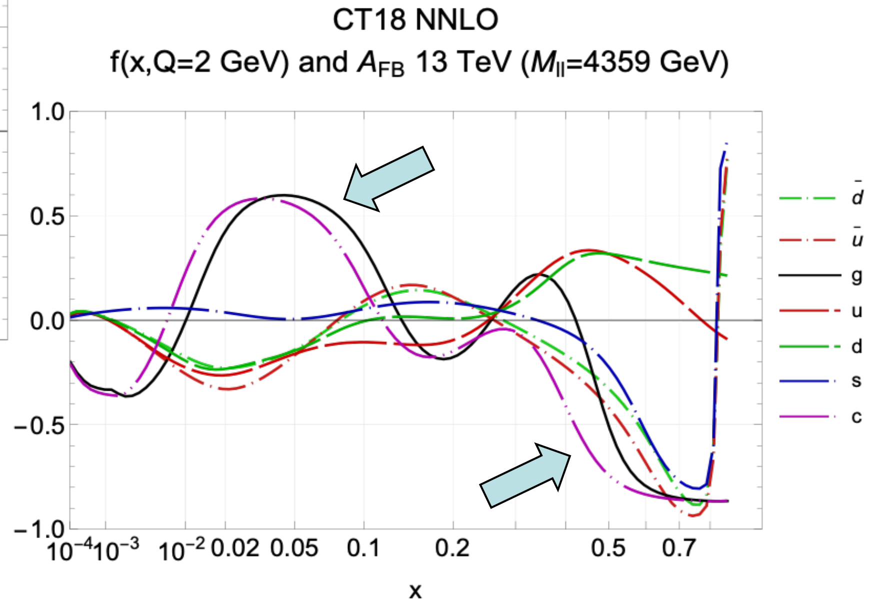
NEW:
Courtoy, Fu, Hou, Hobbs, PN, Yuan, 2023

Drell-Yan forward-backward asymmetry **can be** sensitive to light sea and gluon for increasing $M_{\ell\bar{\ell}}$.



Strong anti-correlation of AFB with $\overline{u}$ and $\overline{d}$ (and gluon) at large x for increasing M_{ll}.





PRELIMINARY

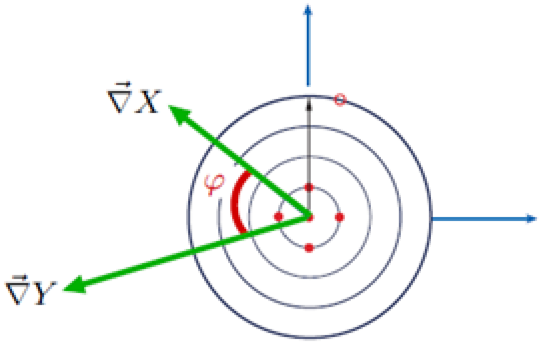
Growing correlation of AFB with gluon at $x < 0.2$.

see also [Ball et al, EPJC 2022 82]

L_2 sensitivity, definition

$S_{f,L_2}(E)$ for experiment E is the estimated $\Delta\chi^2_E$ for this experiment when a PDF $f_a(x_i, Q_i)$ increases by the +68% c.l. Hessian PDF uncertainty

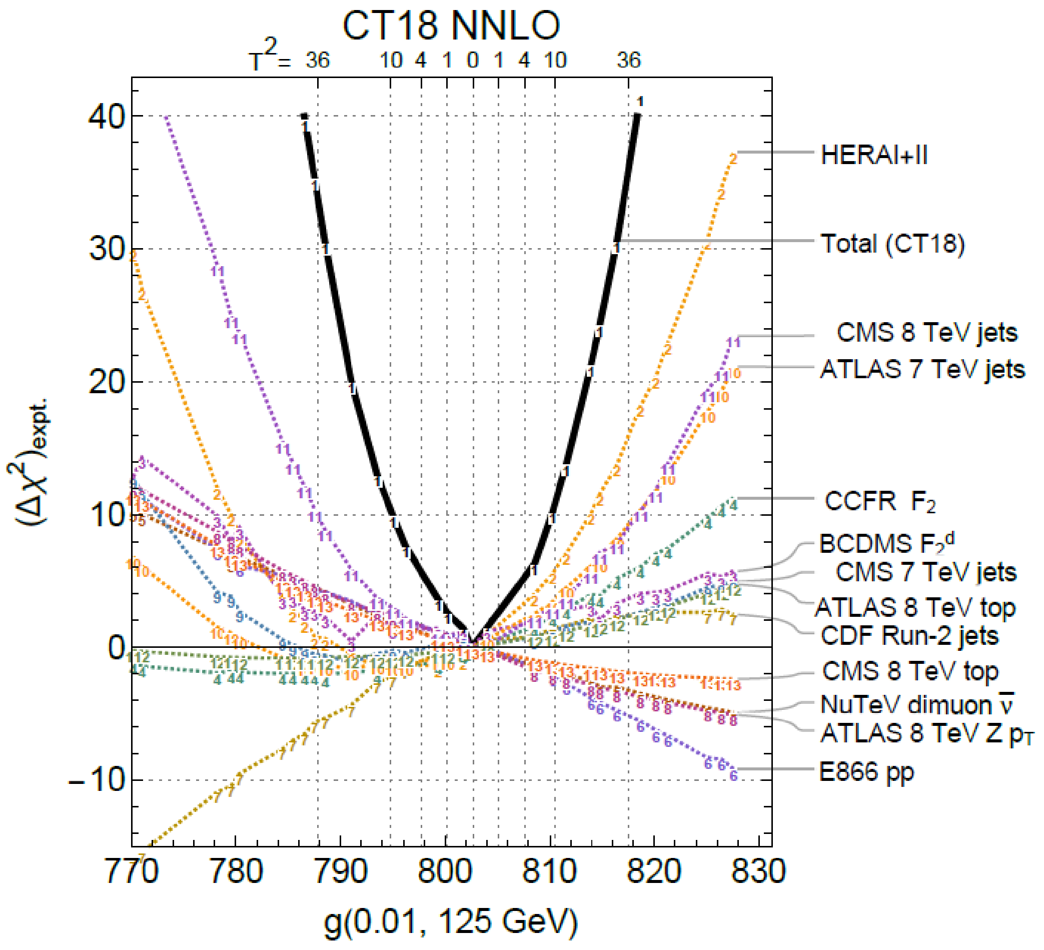
Take $X = f_a(x_i, Q_i)$ or $\sigma(f)$; $Y = \chi^2_E$ for experiment E.

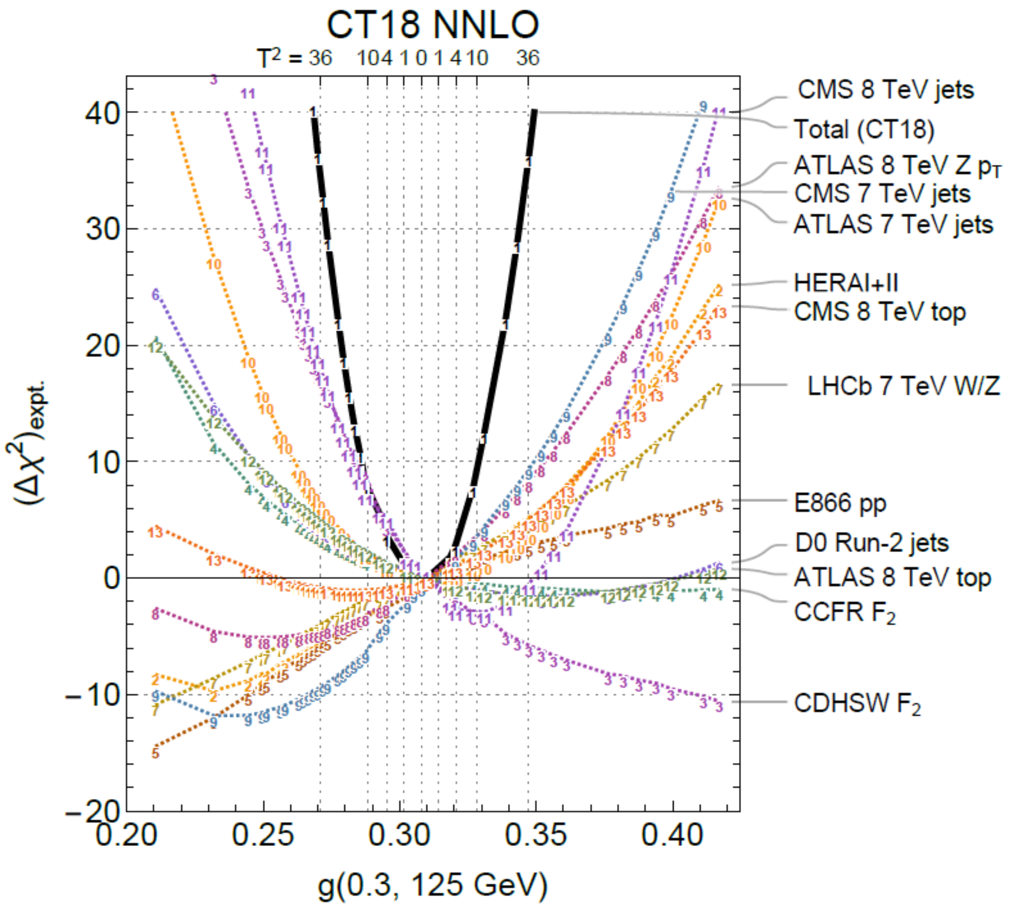


$$S_{f,L_2} \equiv \Delta Y(\vec{z}_{m,X}) = \vec{\nabla} Y \cdot \vec{z}_{m,X} = \vec{\nabla} Y \cdot \frac{\vec{\nabla} X}{|\vec{\nabla} X|} = \Delta Y \cos\varphi.$$

A fast version of the Lagrange Multiplier scan of χ^2_E along the direction of $f_a(x_i, Q_i)$!

Lagrange multiplier scans on $g(x, M_H)$ at $x = 0.01$ and 0.3





What is the L_2 sensitivity?

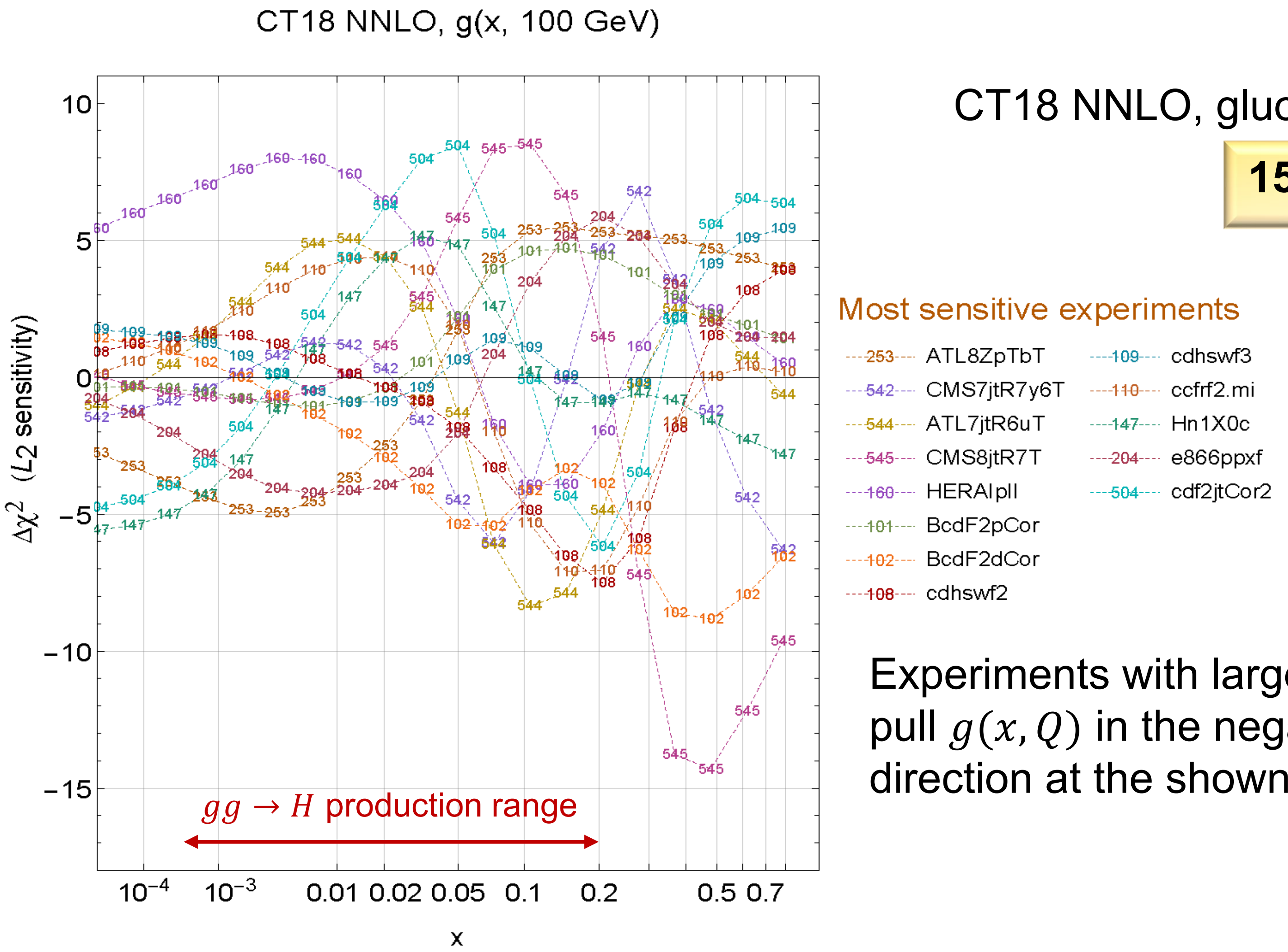
- The L_2 sensitivity is a way of vizualizing the pulls of fitted experiments on the best-fit PDF $f_a(x, Q)$, for a particular parton flavor x, as a function of x and Q
  – or, when plotted for a PDF luminosity, as a function of the final-state mass M_X
- The best-fit value for a particular $f_a(x, Q)$ is determined by the sum of these pulls

- Both the L_2 and LM methods explore the parametric dependence of the χ^2 function in the vicinity of the global minimum

- The L_2 sensitivity streamlines comparisons among independent PDF analyses using **published error PDFs**
- The L_2 sensitivity has been used internally by CT (in CT18), by the PDF4LHC21 benchmarking group (to determine which data sets should be in the reduced PDF fit used for benchmarking), and now by ATLASpdf, CT, and MSHT

Estimated χ^2 pulls from experiments

(L_2 sensitivity, T. J. Hobbs et al., arXiv:1904.00222)



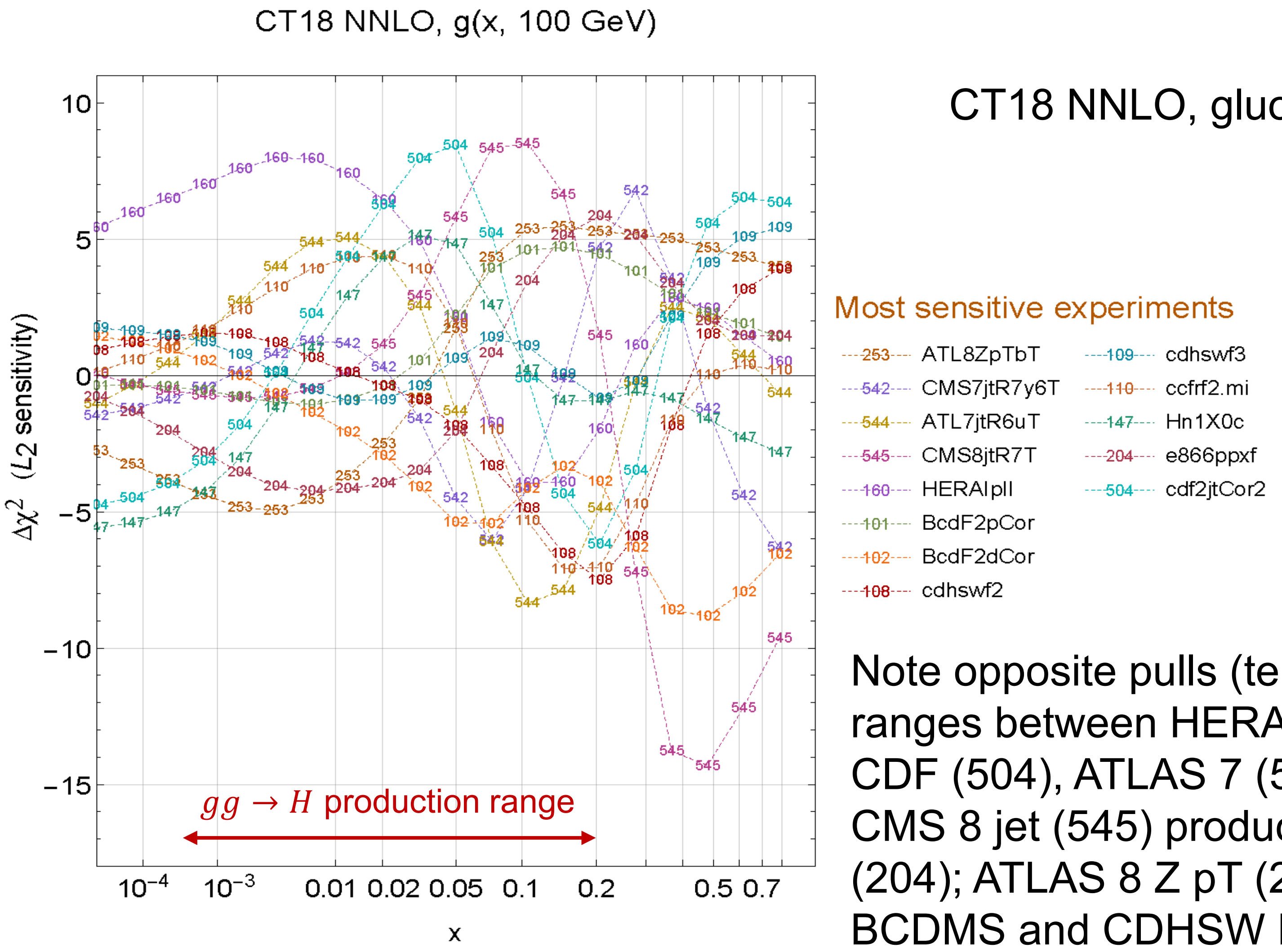
CT18 NNLO, gluon at Q=100 GeV

15 core-minutes

Experiments with large $\Delta\chi^2 > 0$ [$\Delta\chi^2 < 0$] pull $g(x,Q)$ in the negative [positive] direction at the shown x

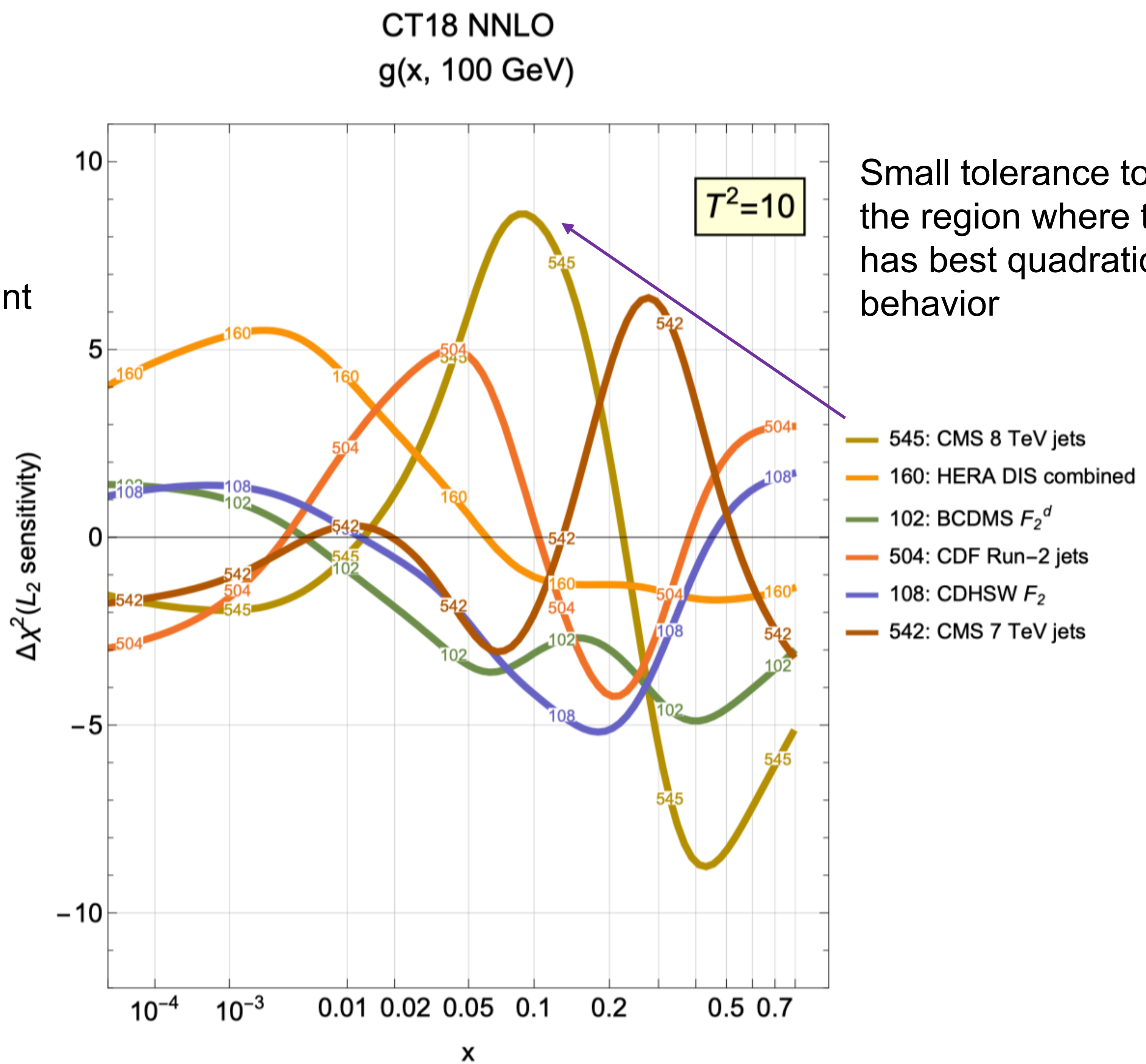
Estimated χ^2 pulls from experiments

(L_2 sensitivity, T. J. Hobbs et al., arXiv:1904.00222)



CT18 NNLO, gluon at Q=100 GeV

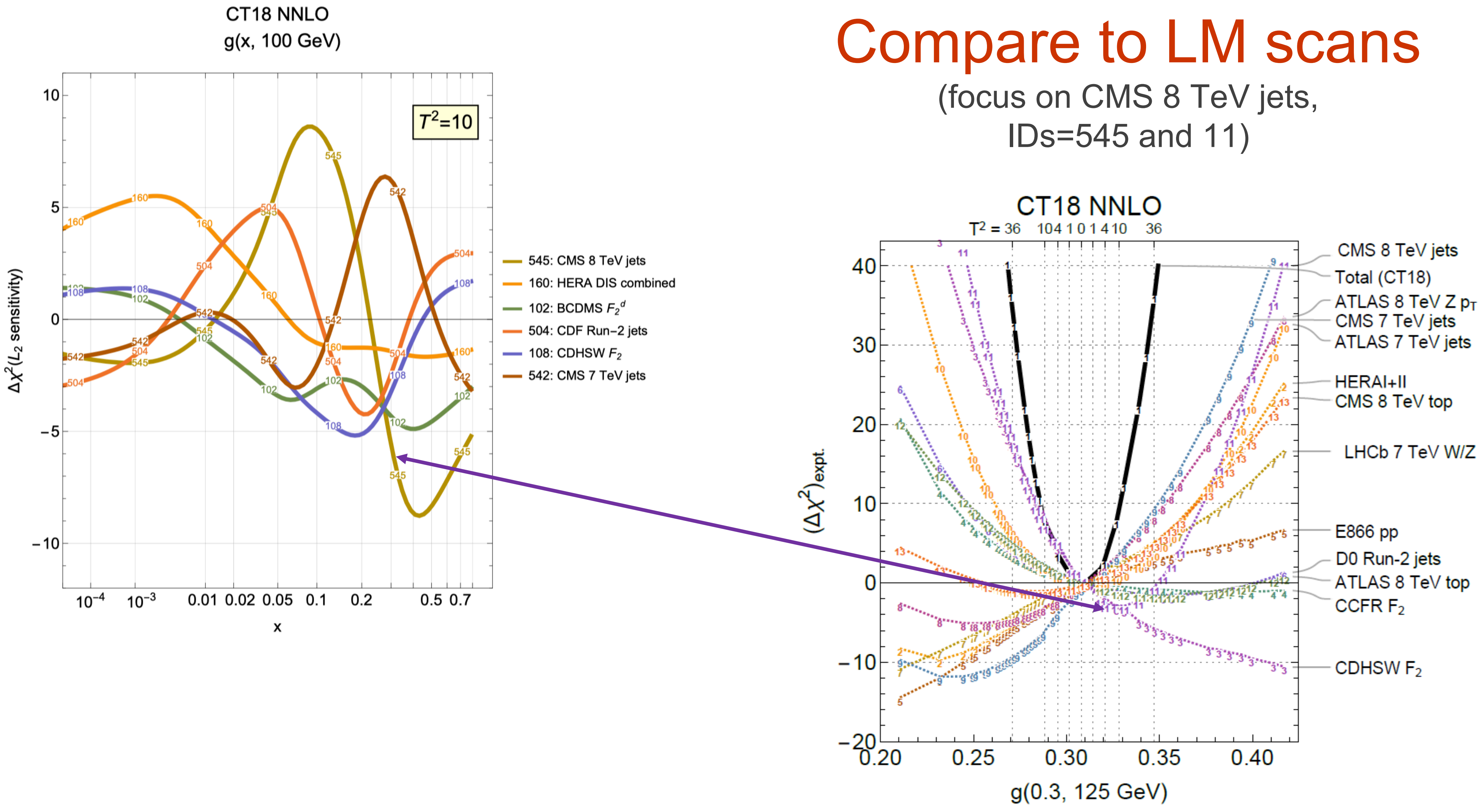
Note opposite pulls (tensions) in some x ranges between HERA I+II DIS (ID=160); CDF (504), ATLAS 7 (544), CMS 7 (542), CMS 8 jet (545) production; E866pp DY (204); ATLAS 8 Z pT (253) production; BCDMS and CDHSW DIS show only 6 most important experiments



Small tolerance to stay in the region where total χ^2 has best quadratic behavior

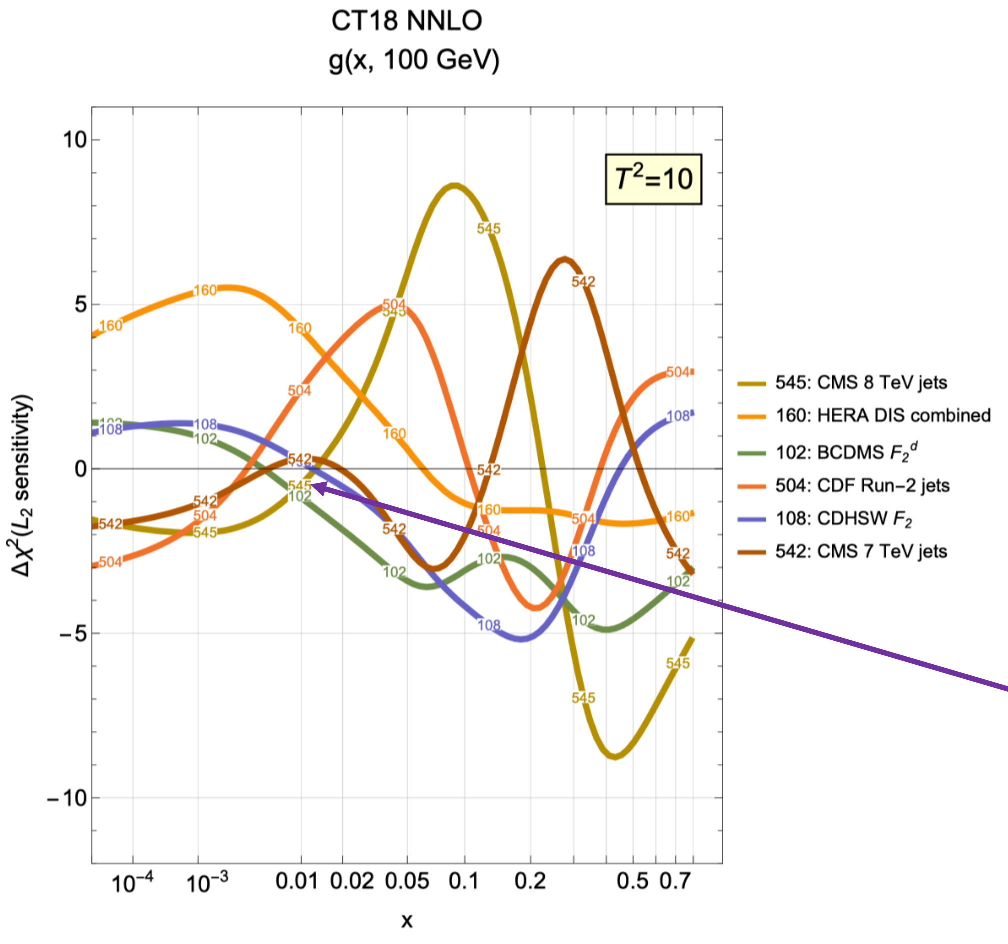
Compare to LM scans

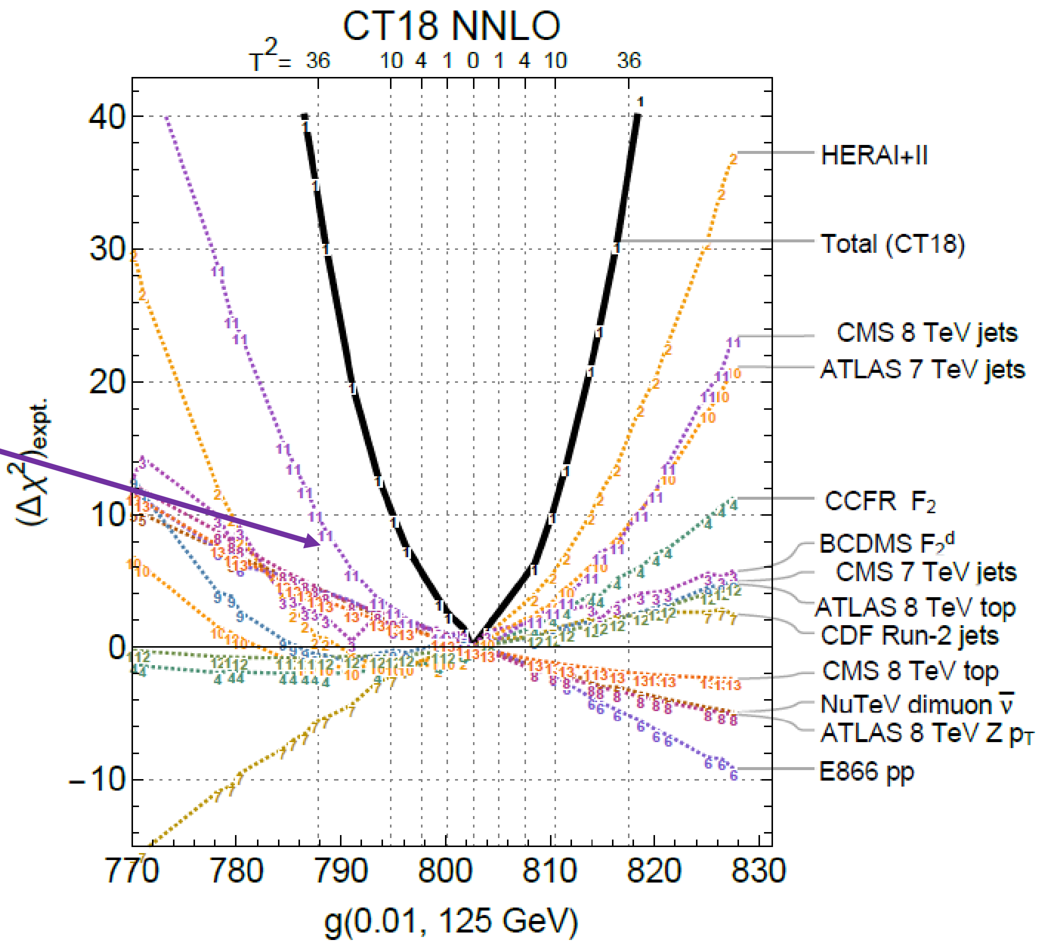
(focus on CMS 8 TeV jets,
IDs=545 and 11)



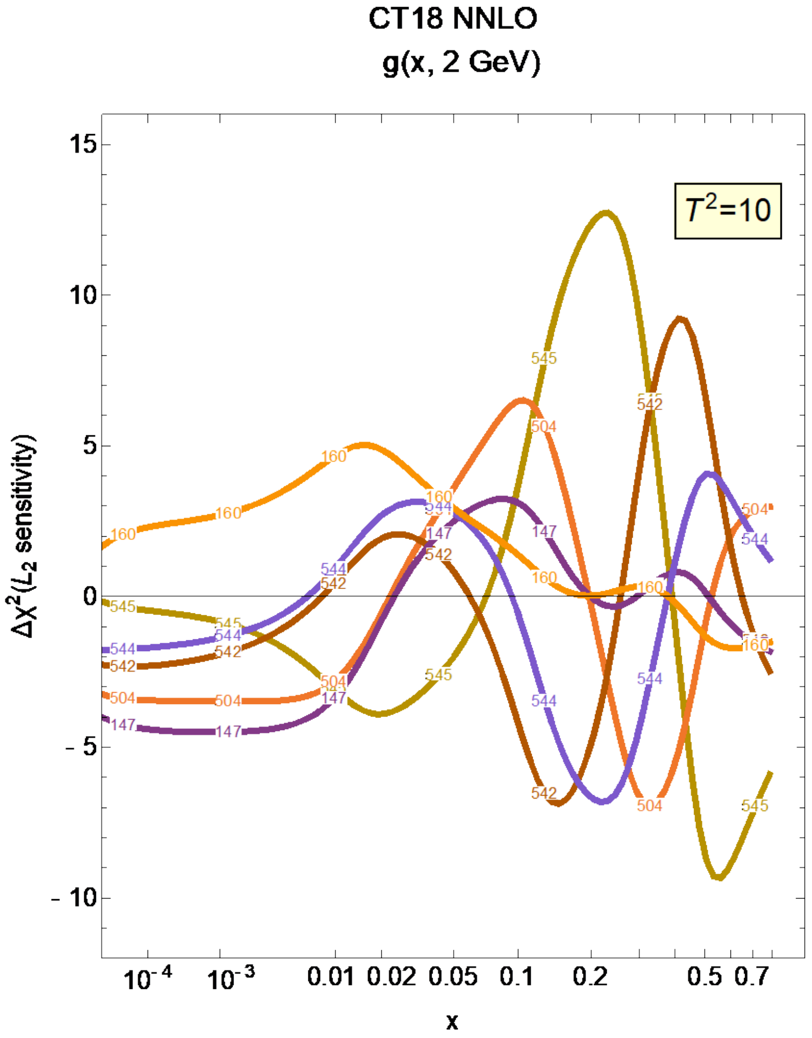
Compare to LM scans

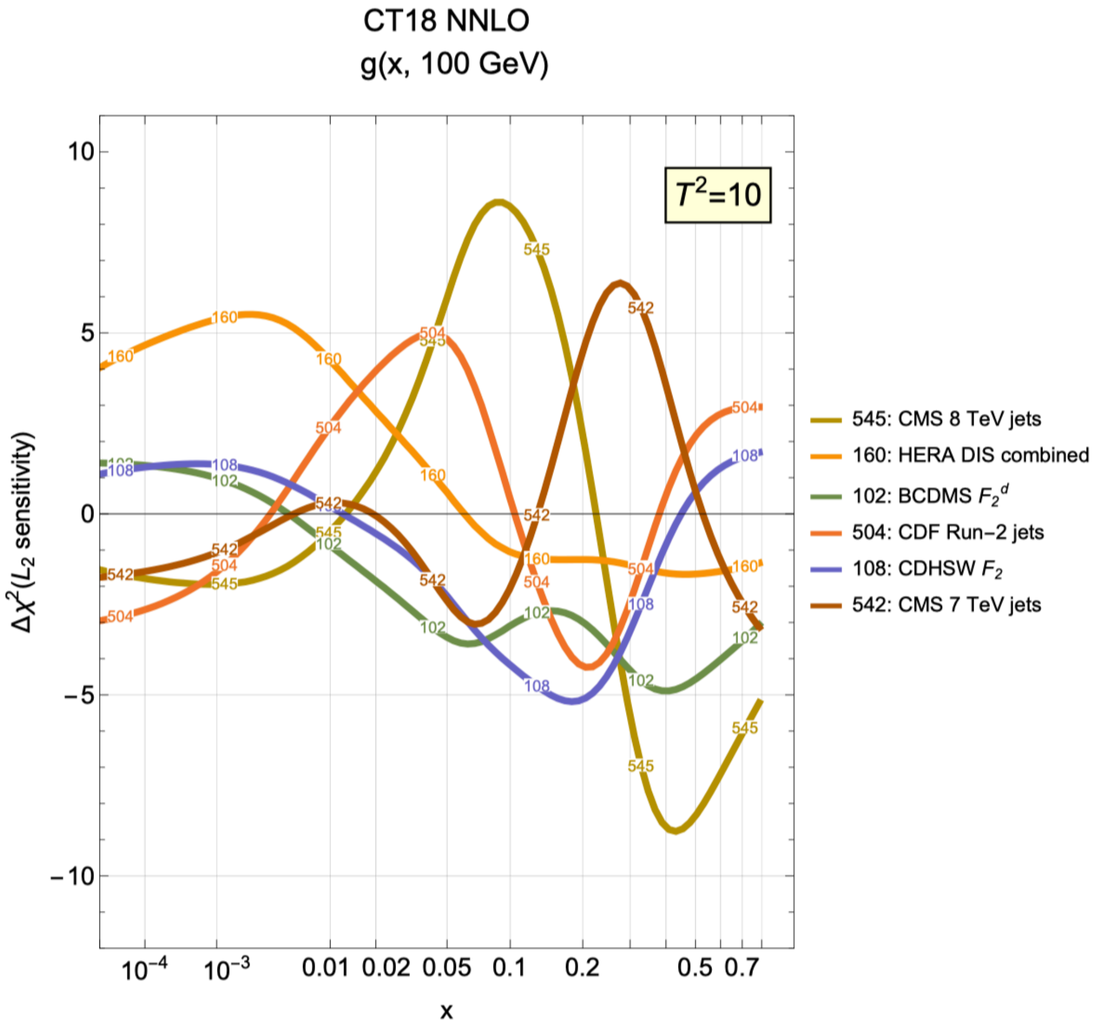
(focus on CMS 8 TeV jets, IDs=545 and 11)



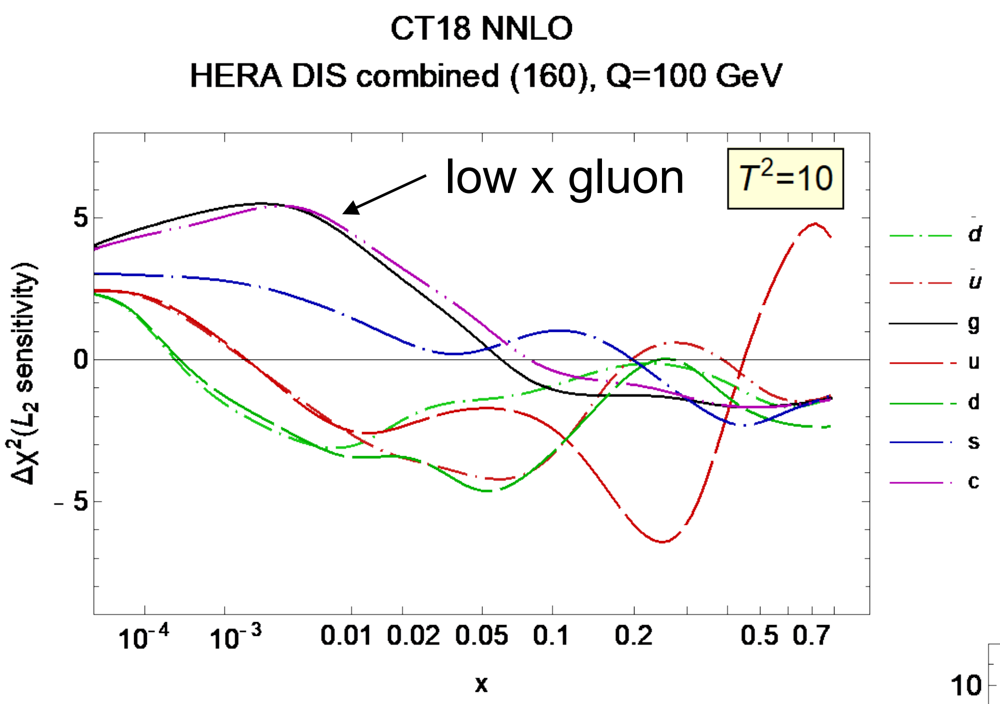


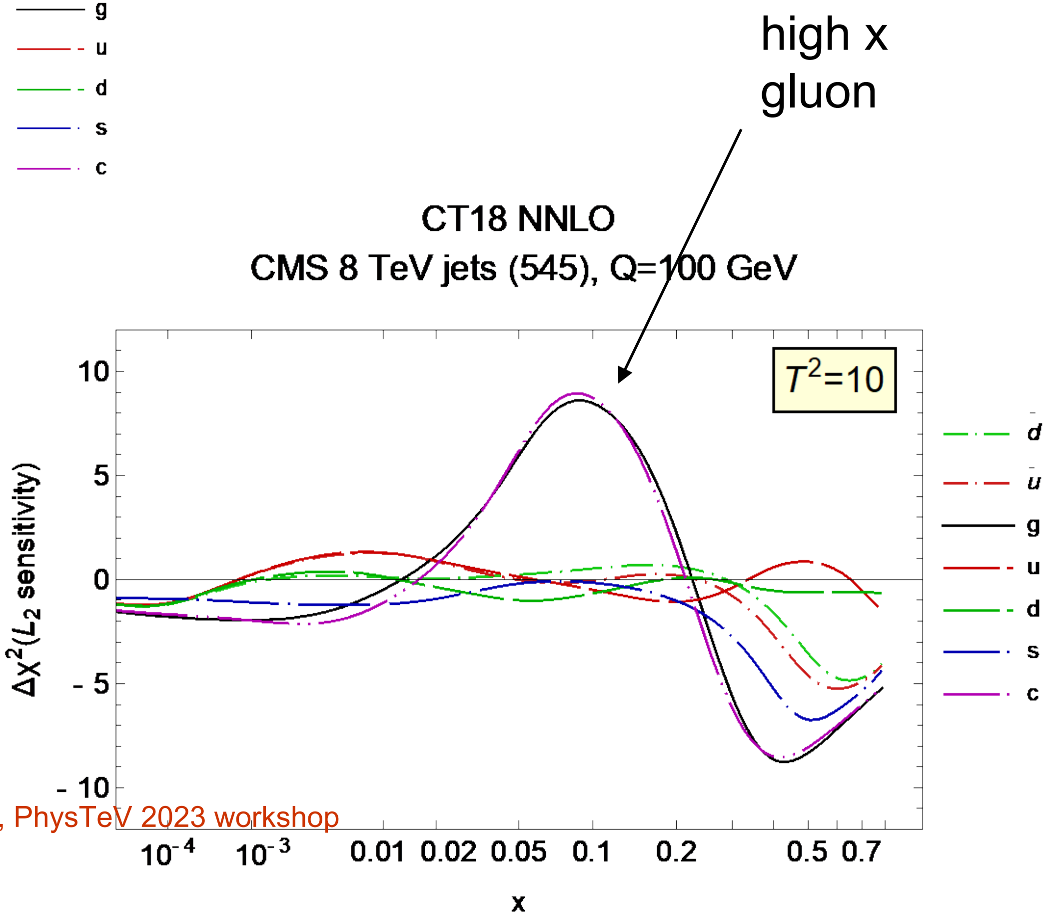
Can also look at L_2 for 2 GeV





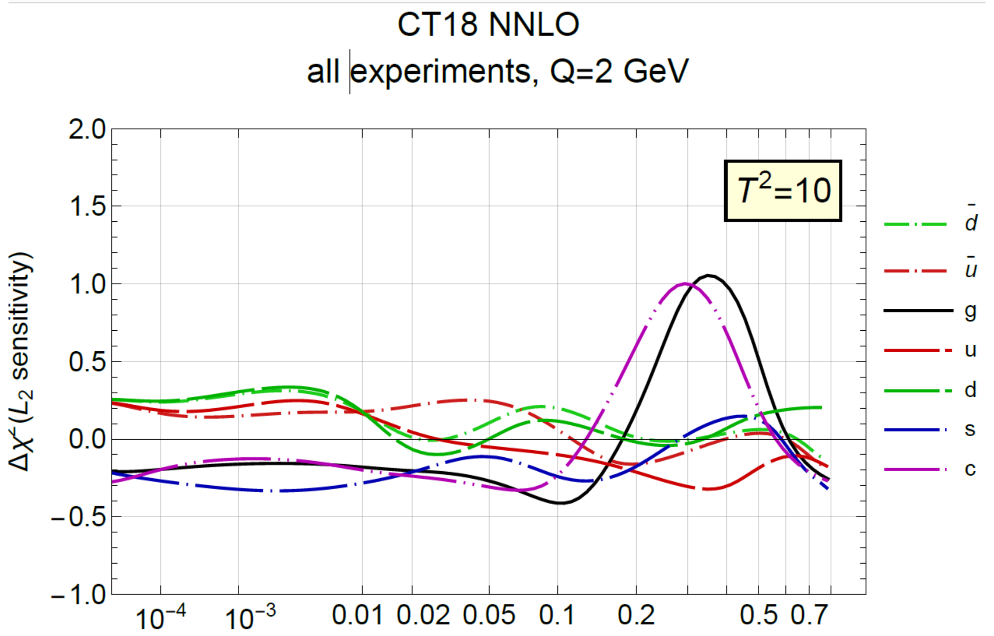
Examine the impact of each experiment on the different PDFs

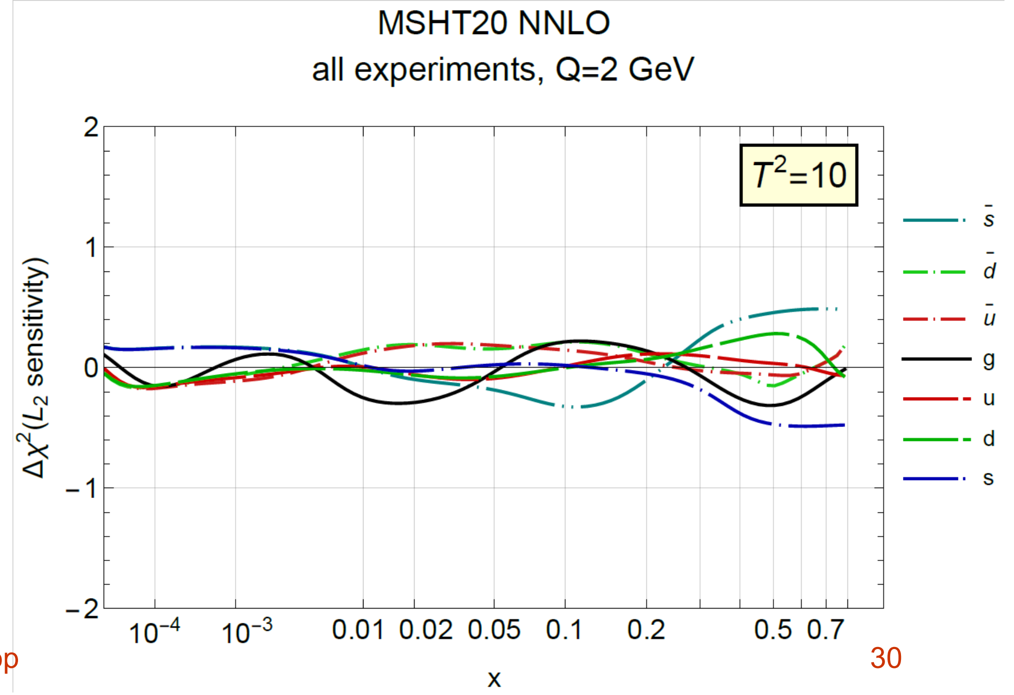




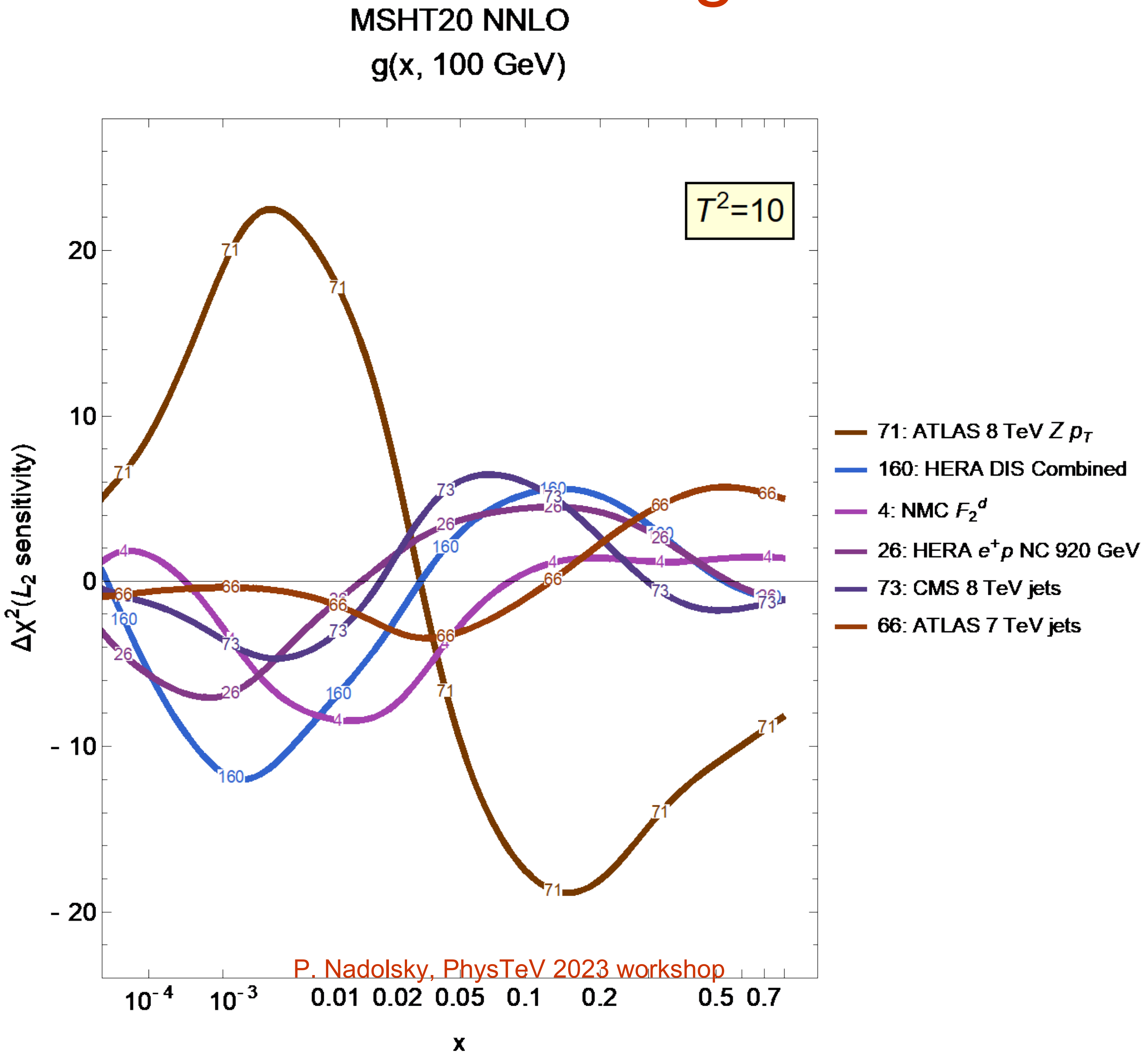
since first derivative of χ^2 vanishes at the global minimum, the sum of the L_2 sensitivities must be zero within uncertainties

$$\sum_E S_{f,L_2} \ll T^2 < \sum_E |S_{f,L_2}|$$



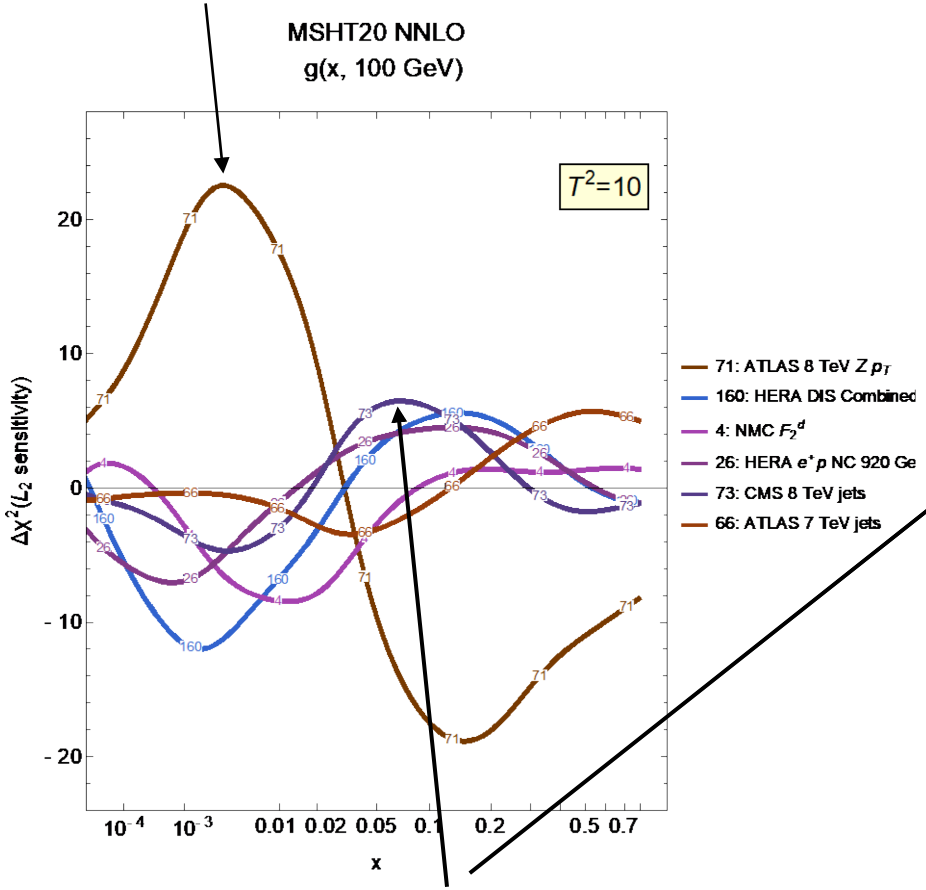


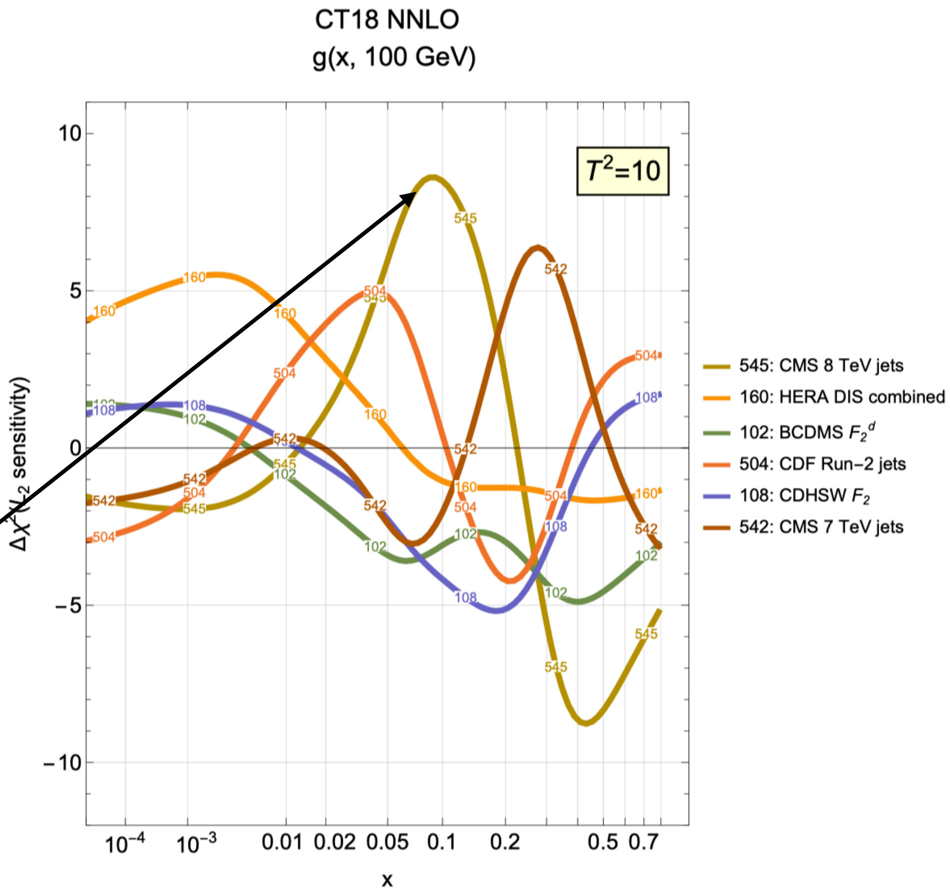
MSHT20 NNLO gluon



MSHT20 and CT18

Note importance of ATLAS Z p_T data (also, Z p_T data poorly fit at NNLO)



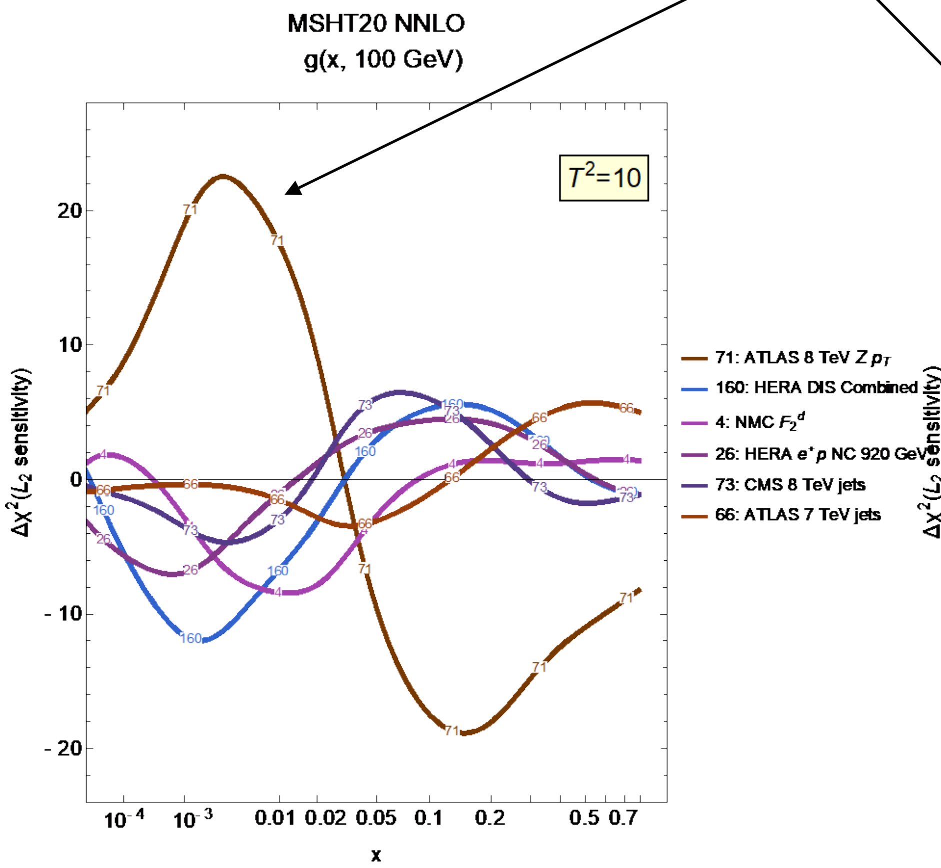


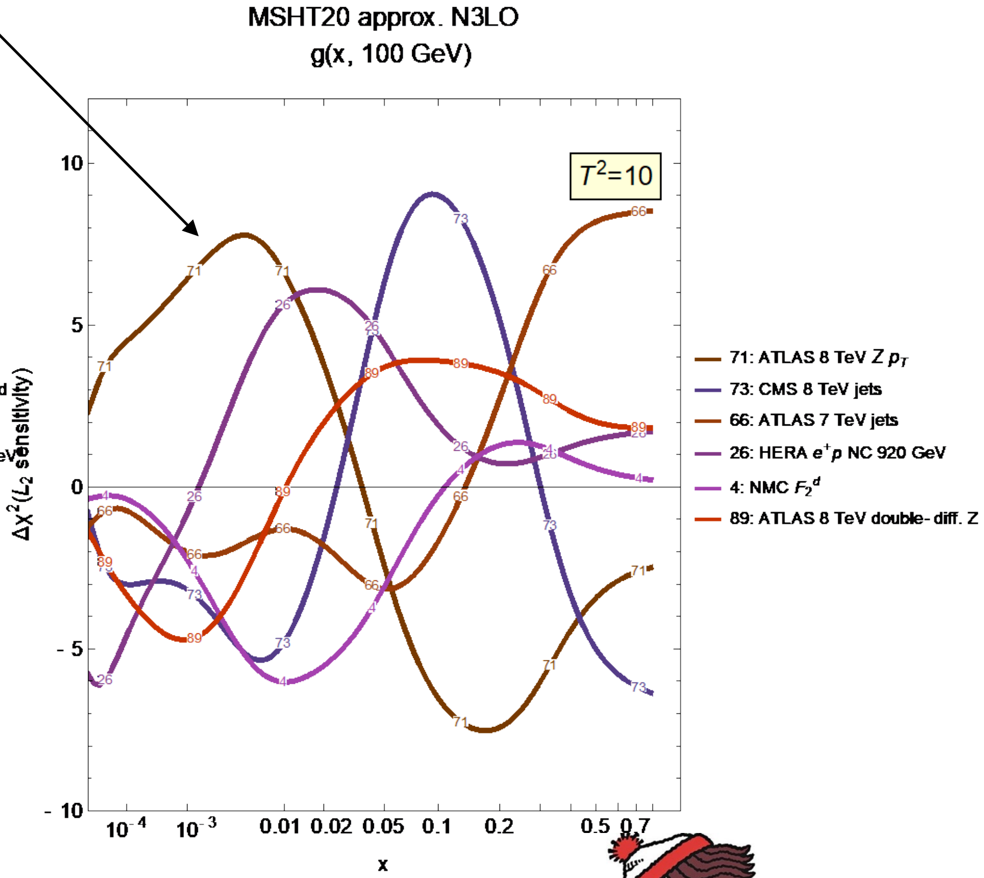
CMS 8 TeV jet data play a similar role as in CT18

ATLAS Z p_T not one of 6 most important experiments (fewer data points?)

MSHT20 NNLO and aN3LO

shape of L_2 sensitivity similar for two PDFs, but absolute value decreased by almost a factor of 3; significant change in low x gluon

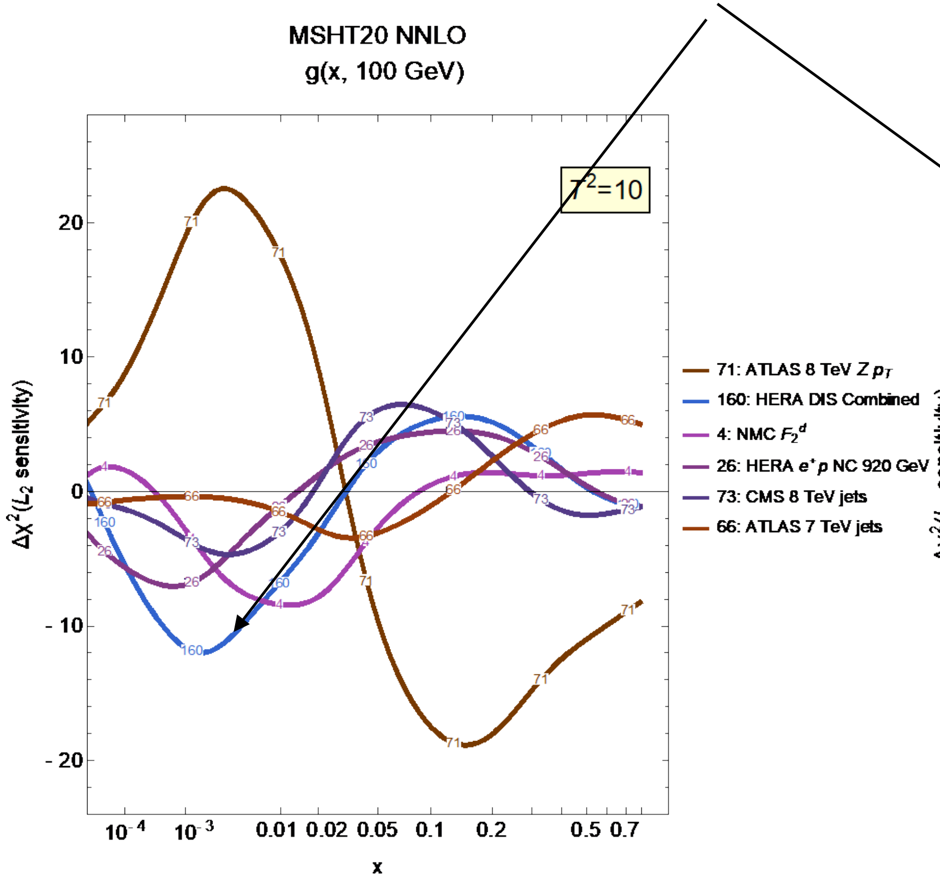


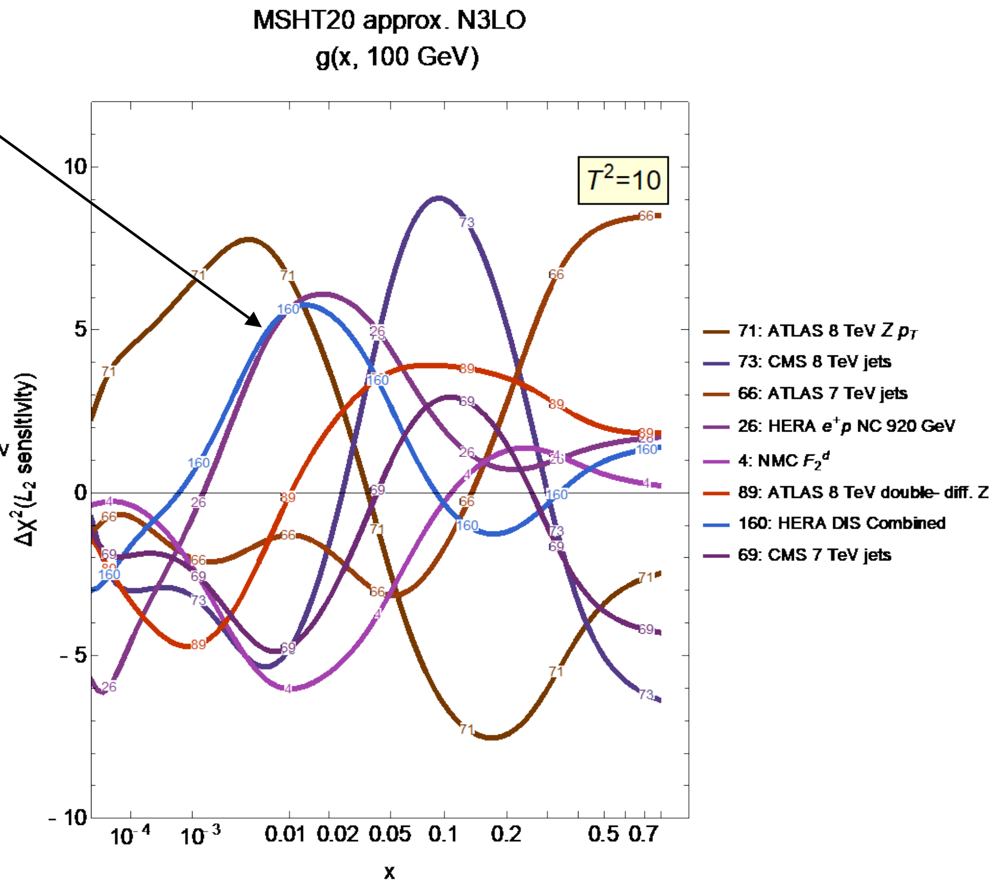


where's expt 160?

MSHT20 NNLO and aN3LO

at aN3LO, the two experiments now on same side; aN3LO needed for HERA





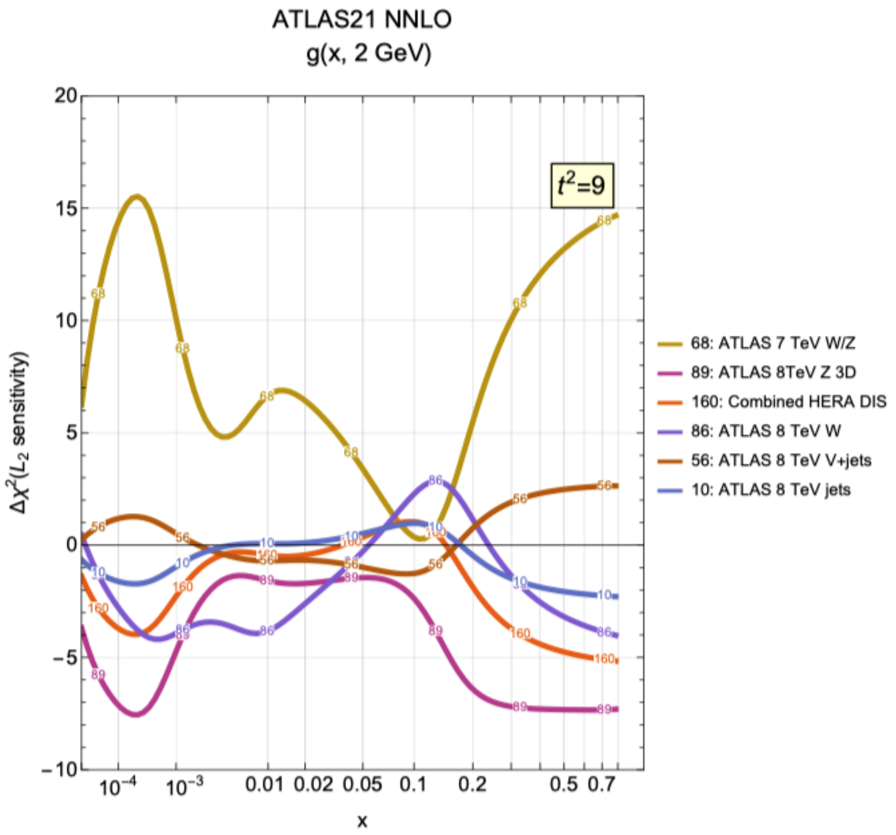
ATLASpdf21

ATLAS PDF fits are based on a more limited set of data, with HERA inclusive as the backbone

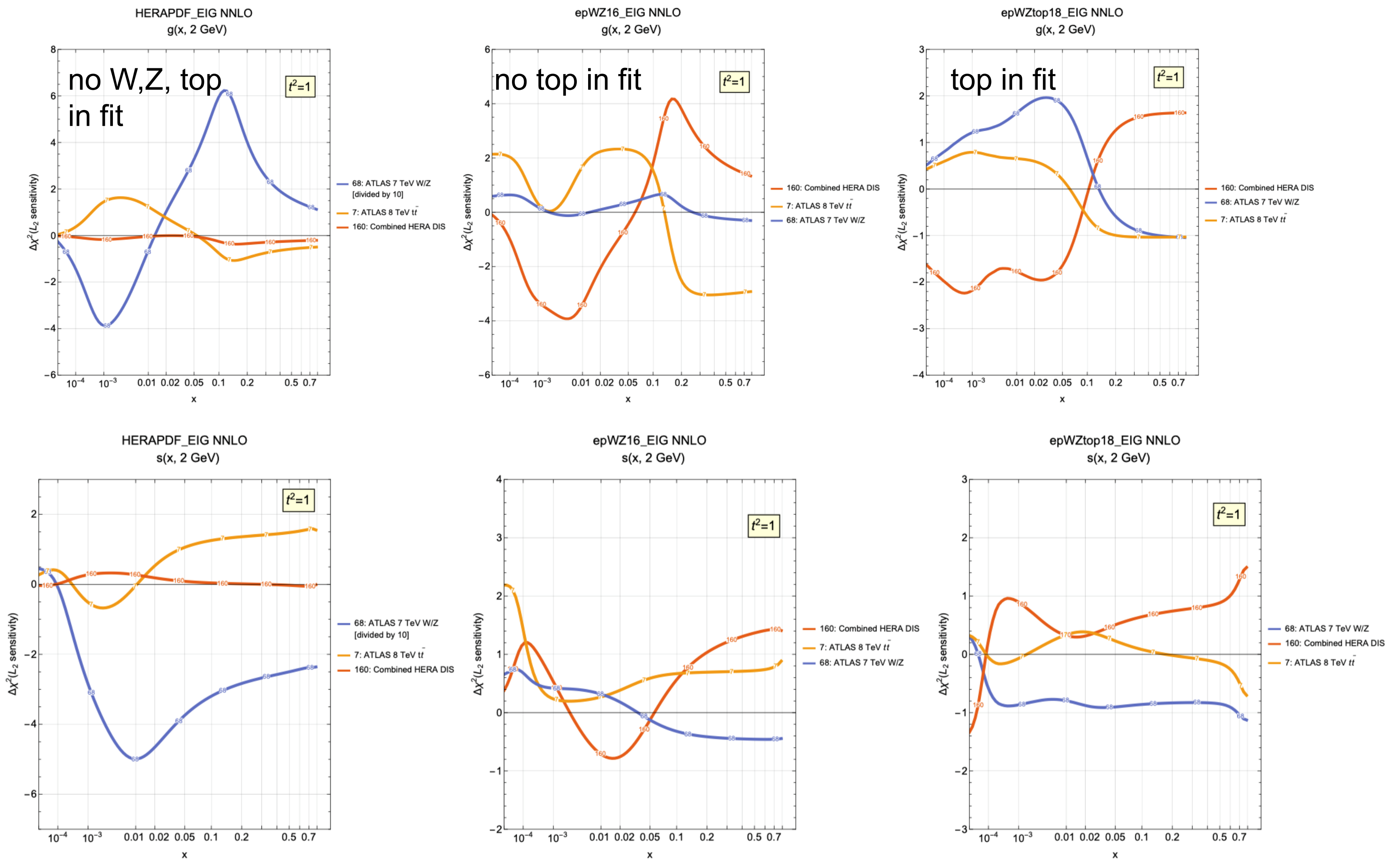
series of sequential PDF fits adding W/Z data,ttbar, W/Z+jets, inclusive jets and photon ratio data

full information on correlated systematic sources of uncertainty used (not available to to other PDF fits)

ID	Data set	$\sqrt{s}$ [TeV]	Luminosity [fb^{-1}]	Decay channel	Observables entering the fit
160	HERA inclusive DIS [26]	Varied	Varied		Reduced cross sections
68	Inclusive $W, Z/\gamma^*$ [27]	7	4.6	e, μ combined	η_ℓ (W), y_Z (Z)
89	Inclusive Z/γ^* [28]	8	20.2	e, μ combined	$\cos\theta^*$ in bins of $y_{\ell\ell}$, $m_{\ell\ell}$
86	Inclusive W [29]	8	20.2	μ	η_μ
56	$W^\pm$ + jets [30]	8	20.2	e	p_T^W
	Z + jets [31]	8	20.2	e	p_T^{jet} in bins of $\lvert y^{\text{jet}}\rvert$
7	$t\bar{t}$ [32, 33]	8	20.2	lepton + jets, dilepton	$m_{t\bar{t}}$, p_T^t, $y_{t\bar{t}}$
8	$t\bar{t}$ [34]	13	36	lepton + jets	$m_{t\bar{t}}$, p_T^t, y_t, $y_{t\bar{t}}^{\text{b}}$
9	Inclusive isolated γ [35]	8, 13	20.2, 3.2	-	E_T^γ in bins of η^γ
10	Inclusive jets [36–38]	7, 8, 13	4.5, 20.2, 3.2	-	p_T^{jet} in bins of $\lvert y^{\text{jet}}\rvert$



Impact of addition of W/Z, ttbar data to HERA inclusive



LHC $t\bar{t}$ production

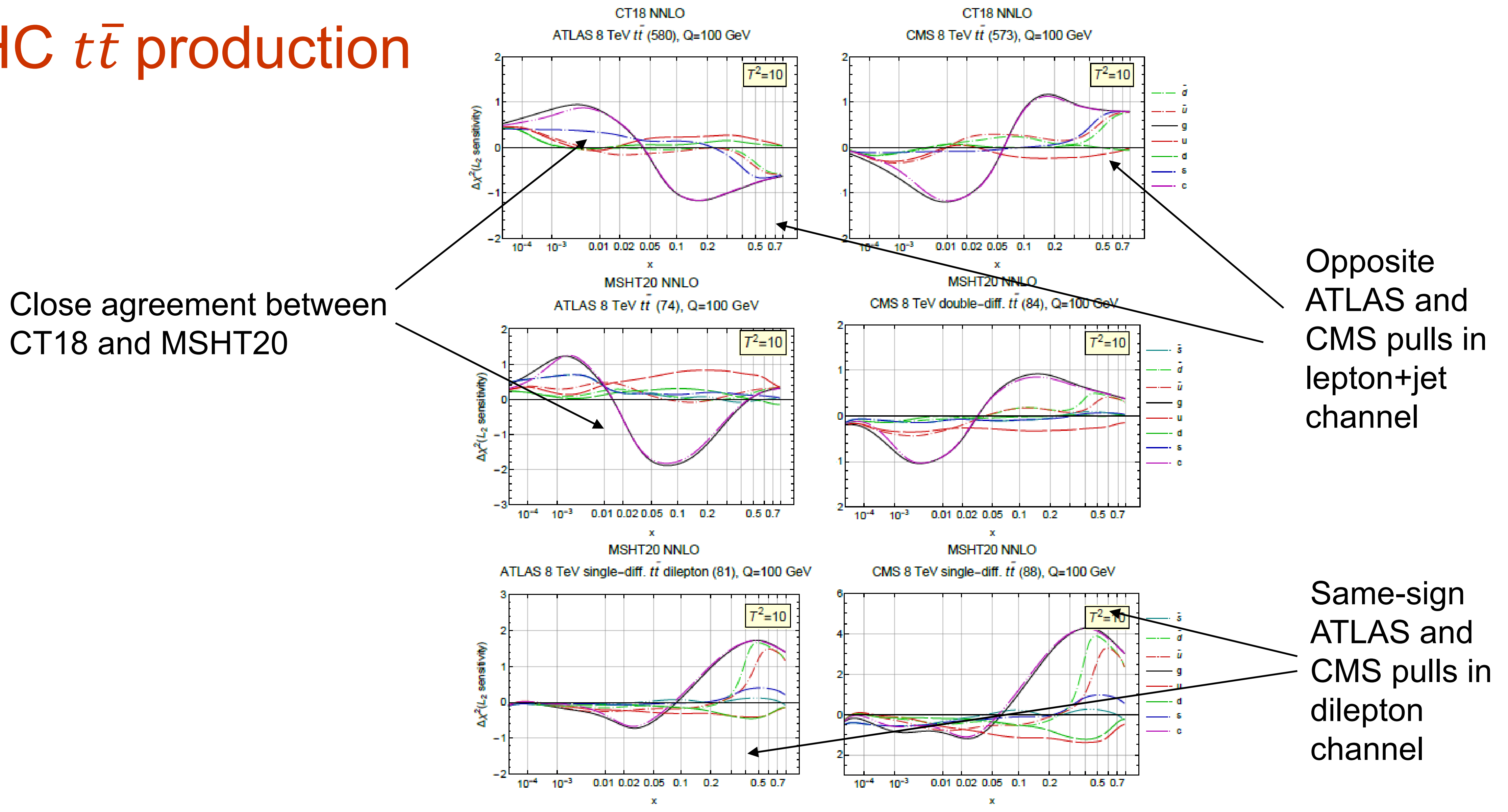


FIG. 22. Sensitivities for ATLAS 8 TeV (left column) and CMS 8 TeV (right column) $t\bar{t}$ data sets in the CT18 (upper row) and MSHT20 (middle and lower rows) analyses at $Q = 100$ GeV.

Many more such comparisons on
https://metapdf.hepforge.org/L2/

2. Implementations of systematic uncertainties

Explore using a hopscotch scan

for MC PDFs

arXiv:2205.10444 [PRD 107 (2023) 3, 034008]

by A. Courtoy, J. Huston, P. N., K. Xie, M. Yan, C.-P. Yuan

Goodness-of-fit functions in PDF analyses

Analysis	**χ^2 prescription to fit PDFs**	**χ^2 prescription to compare PDFs**	**Comments**
HERAPDF	HERA	HERA	
CT	Extended T +addl. prior	Extended T, Experimental	
MSHT'20	T	T	
NNPDF4.0	t_0 +addl. prior with fluctuated cross-sampled data	Experimental or t_0 with unfluctuated full data	t_0 prescription has pre- and post-NNPDF3.0 versions
…			
Hopscotch'2022	N/A	Experimental or t_0 [2022] with unfluctuated data	

Different prescriptions reflect modeling of additive and multiplicative systematic errors in covariance matrices. Neither prescription is complete because of the bias-variance dilemma. The χ^2 definition affects the PDF uncertainty.

Two forms of χ^2 in PDF fits

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

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

2. In terms of the covariance matrix

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

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

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.
Possible choices:

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

Systematic uncertainties and the bias-variance dilemma

$$\chi^2 = \sum_{i,j}^{N_{pt}} (T_i - D_i)(\text{cov}^{-1})_{ij}(T_j - D_j) \qquad (\text{cov})_{ij} = s_i^2 \delta_{ij} + \sum_{\alpha=1}^{N_\lambda} \beta_{i,\alpha}\beta_{j,\alpha}$$

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

D_i, T_i, s_i are the central data, theory, uncorrelated error

$\beta_{i,\alpha} \equiv \sigma_{i,\alpha}\hat{X}_i$ is the correlation matrix for N_λ nuisance parameters. Experiments publish $\sigma_{i,\alpha}$.

The "truth" normalizations $\hat{X}_i$ in the experiment are of order T_i or D_i. $\{\hat{\boldsymbol{X}}_i\}$ **are learned as a model** $\{\boldsymbol{X}_i\}$ **together with PDFs** $\boldsymbol{f}$ **and theory** $\{\boldsymbol{T}_i(\boldsymbol{f})\}$. For example, we can sample as $X_i = a_i D_i + b_i T_i$, with free $0 \leq a_i, b_i \lesssim 1$.

Mean variation δ_X^2 of the model from truth on an ensemble of replicas, for data point i:

$$\delta_X^2 \equiv \left\langle \left(X_i - \hat{X}_i\right)^2 \right\rangle = \underbrace{\left\langle \left(\hat{X}_i - \langle X_i \rangle\right)^2 \right\rangle}_{\text{model bias}} + \underbrace{\langle (X_i - \langle X_i \rangle)^2 \rangle}_{\text{variance}} = \underbrace{\left\langle \left(\hat{X}_i - \langle X_i \rangle\right)^2 \right\rangle}_{\text{model bias}} - \underbrace{\langle (D_i - \langle X_i \rangle)^2 \rangle}_{\text{data bias}} + \underbrace{\langle (D_i - X_i)^2 \rangle}_{\chi^2(D_i, T_i)}$$

Experimental definition, $X_i = D_i$: $\left\langle \left(X_i - \hat{X}_i\right)^2 \right\rangle = \left(\hat{X}_i - D_i\right)^2 \equiv \delta_D^2$

t_0 definition, $X_i = t_{0i}$: $\left\langle \left(X_i - \hat{X}_i\right)^2 \right\rangle = \left(\hat{X}_i - t_{0i}\right)^2 \equiv \delta_{t_0}^2$

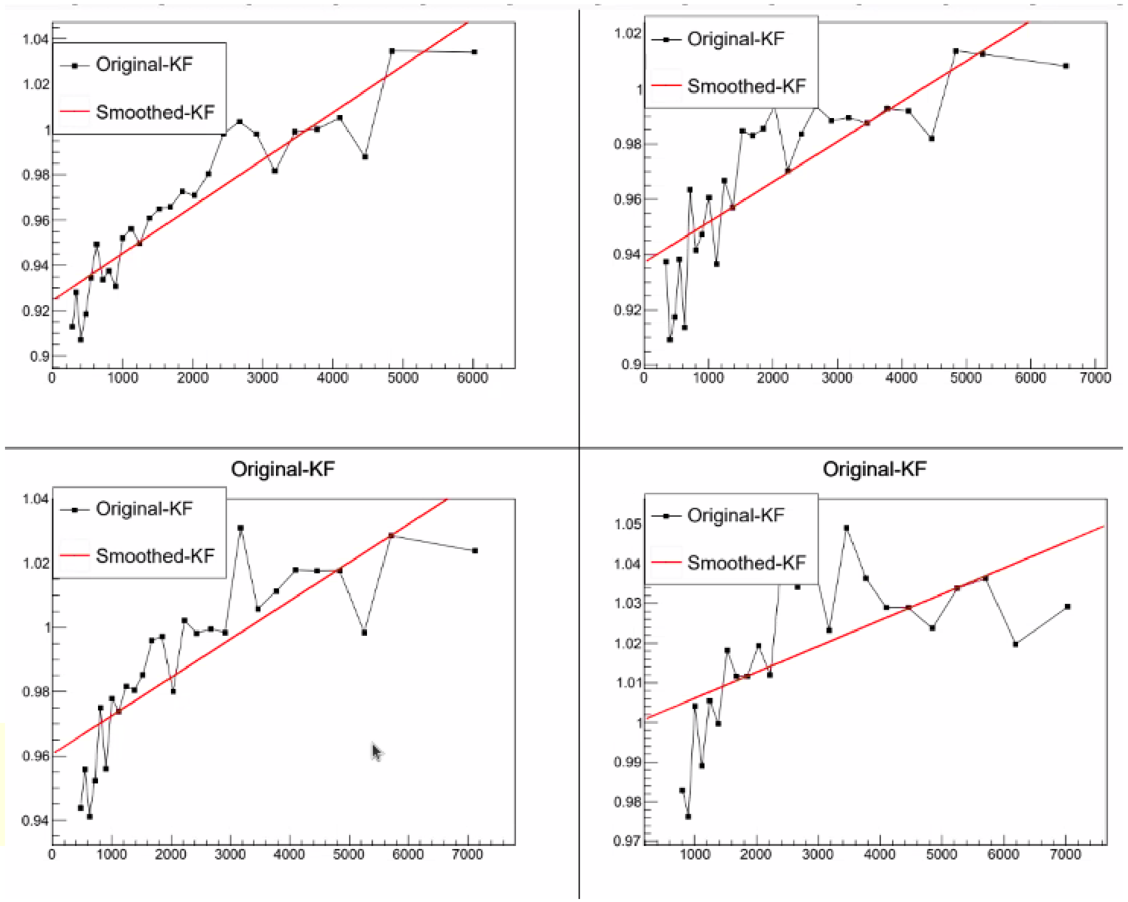
In general, not enough information to compare δ_D and δ_{t_0}

Smoothing of K-factors

An analogous **bias-variance tradeoff** arises during smoothing of MC integration errors for K-factor tables

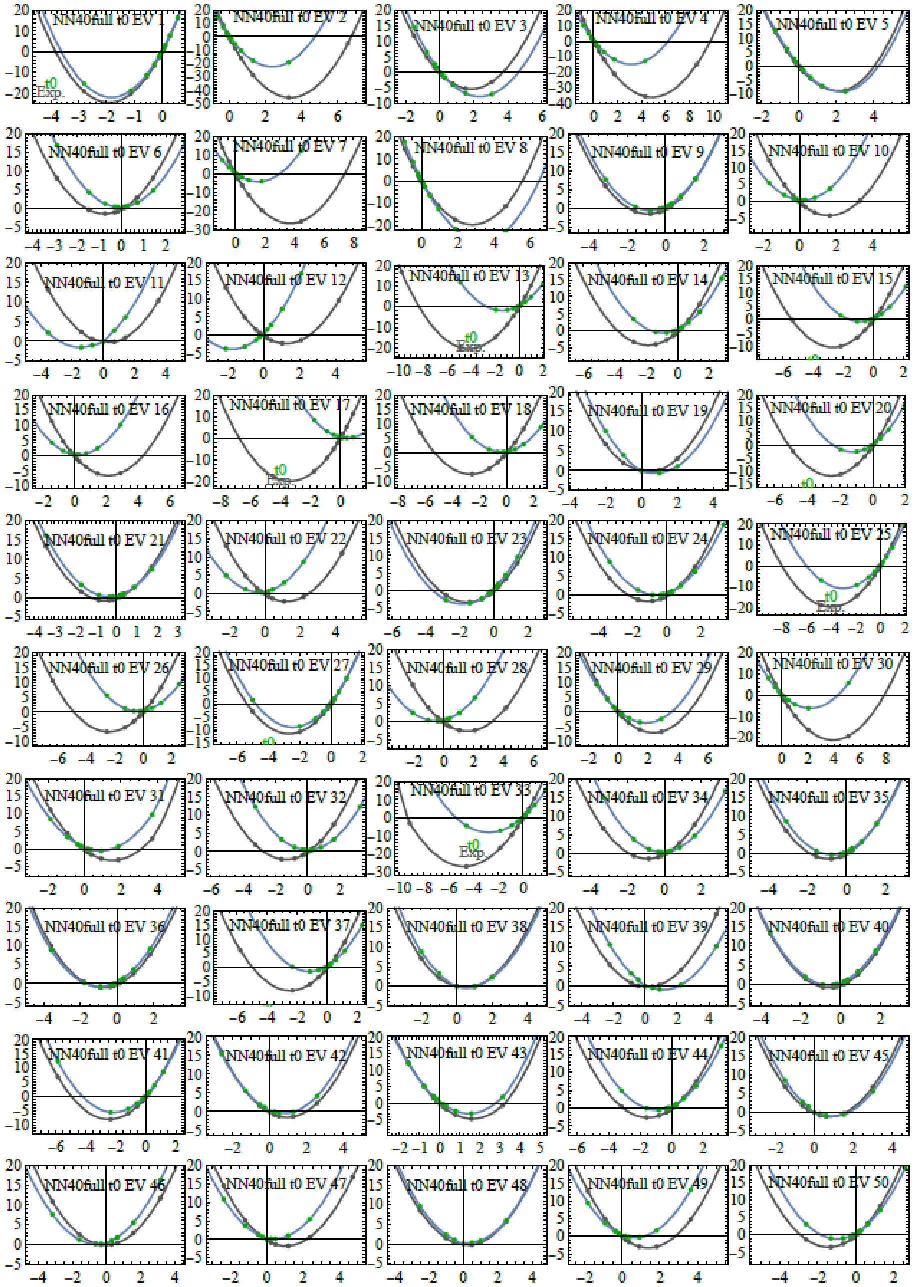
A smoother curve for theory reduces the χ^2 for the jet data, but the best-fit result retains some dependence on the fitted functional form

This dependence can be conservatively estimated by including an uncorrelated MC integration error (both the statistical error provided with the NNLO calculation and the epistemic error due to the functional form)

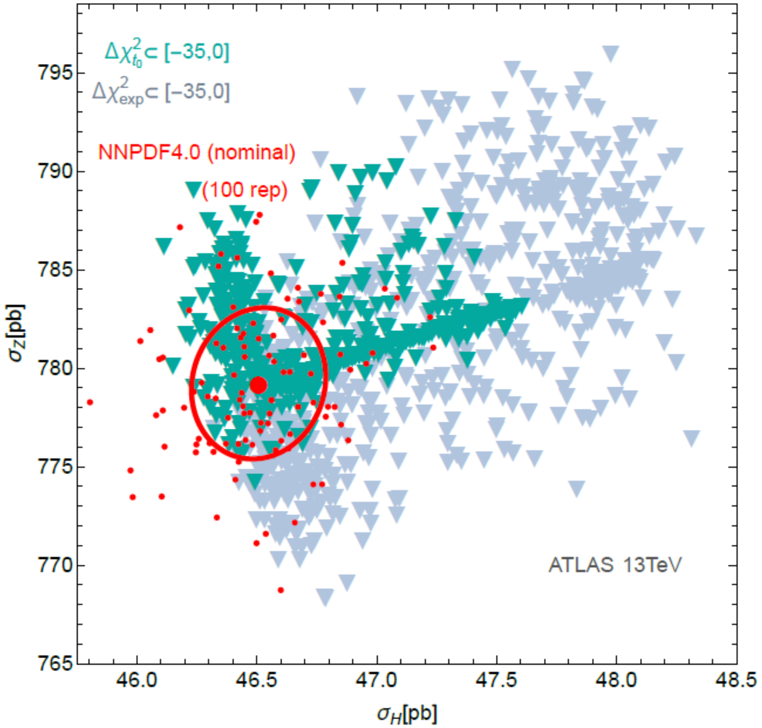


NNLO/NLO ratios for LHC 13 TeV jet production

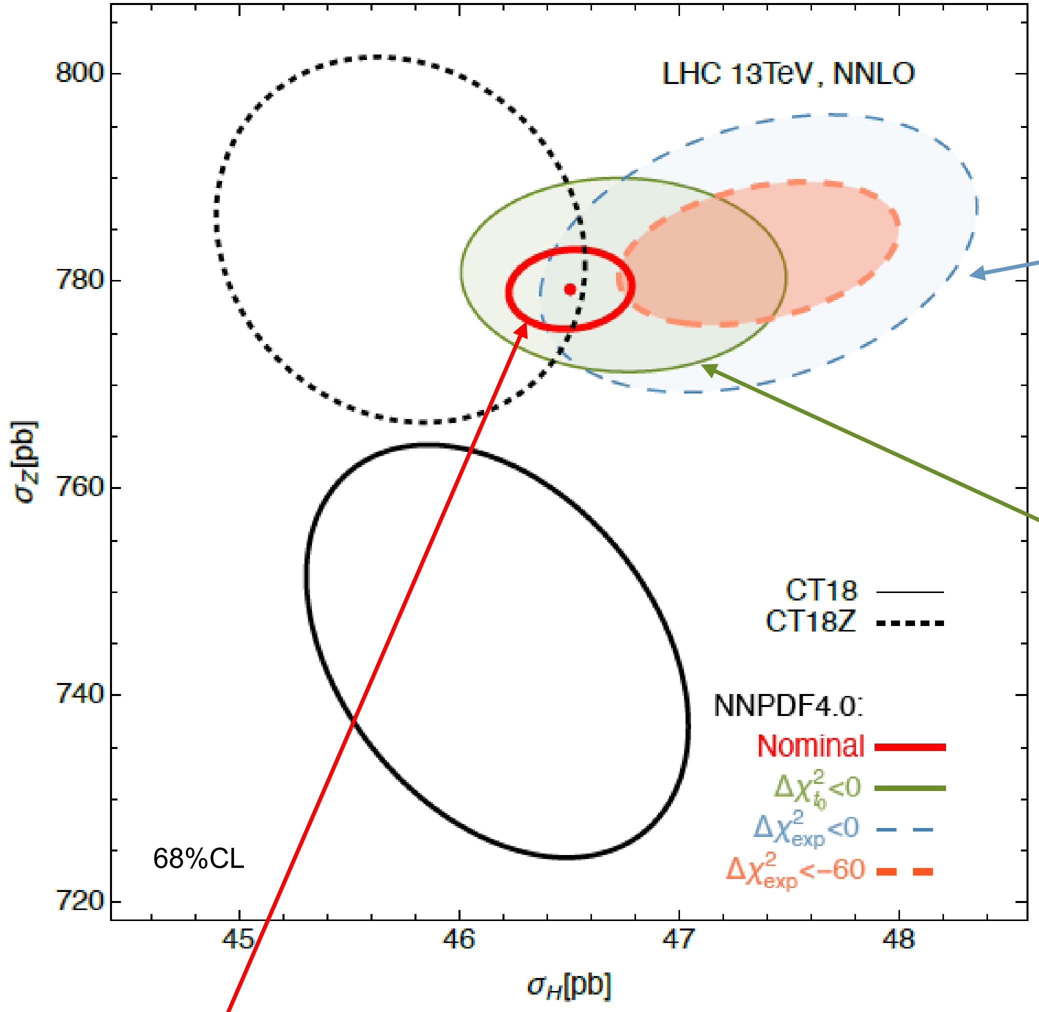
A **hopscotch study** of NN4.0 PDFs: look for signs of **epistemic** uncertainty



1. **Scan** the quasi-Gaussian χ^2 dependence along 50 Hessian EV directions
2. **Sample** with high density a few EV directions that drive the specific PDF uncertainty



Hopscotch scan+sampling of PDF parametrizations



Regions containing (very) good solutions according to the experimental form of χ^2 (is used in χ^2 summary tables of the NN4.0 article, is used in the NN4.0 public code when not doing the fits)

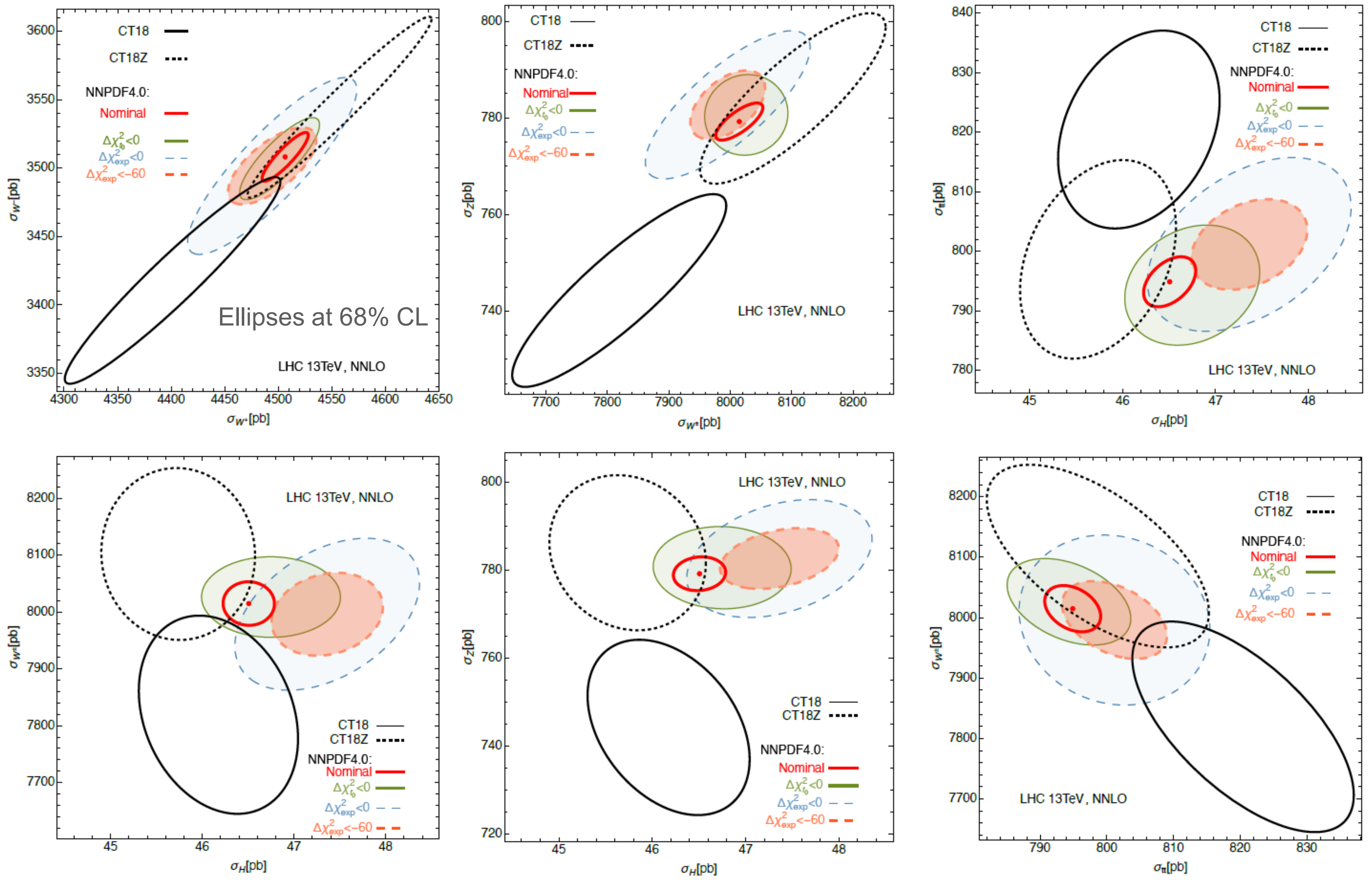
Region containing good solutions according to the NNPDF3.0 t_0 form of χ^2 (used to train NN4.0 replicas)

Nominal NN4.0 Hessian or MC 68%cl

These regions are approximate, at least as large as shown

The hopscotch scans: NNPDF4.0 vs CT18 uncertainties

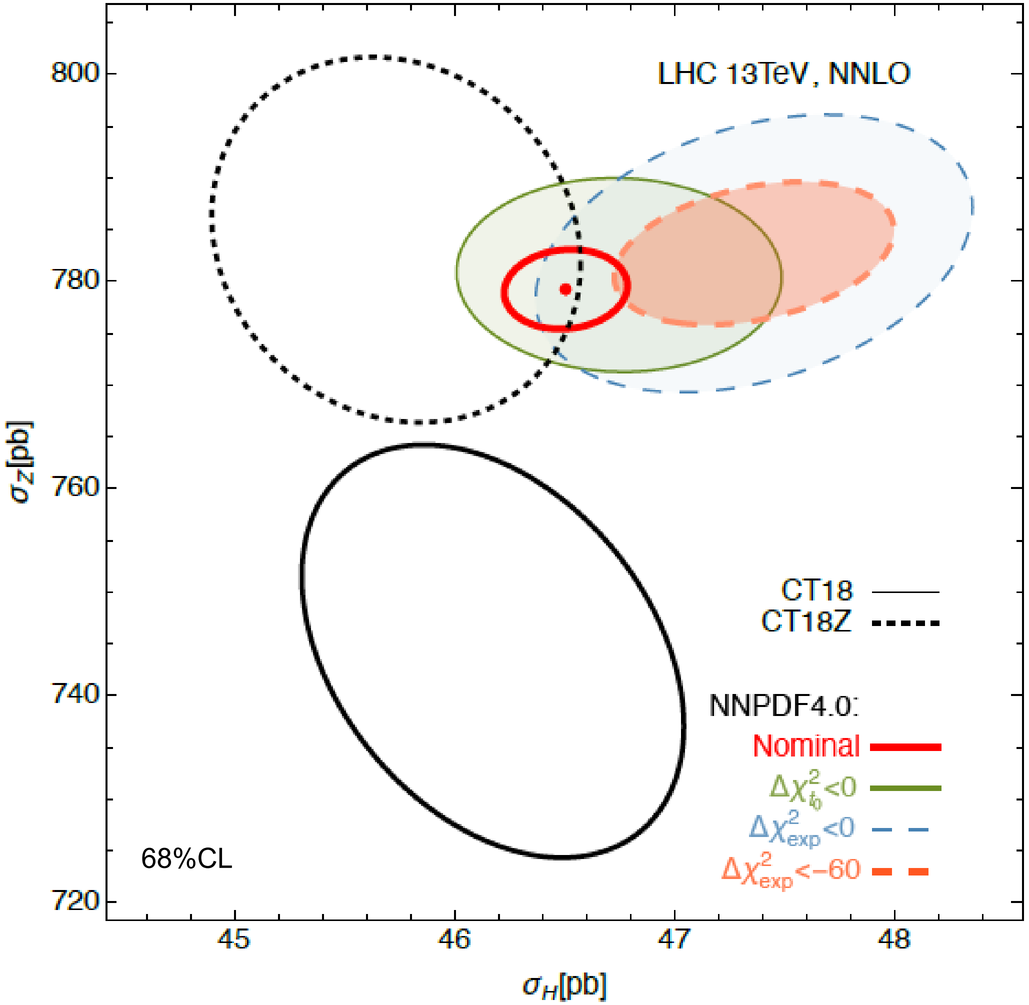
arXiv:2205.10444



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

Hopscotch scans realize the likelihood-ratio test



According to the LR test, the NN4.0 analysis discards PDFs in the green and blue regions based on the prior probabilities and differences in the likelihood definitions – both associated with prior terms

The allowed regions will change for the other acceptable χ^2 definitions, which exist in reflection of the bias-variance dilemma

PDF wish list for systematic uncertainties

A proposal

1. More complete representations for experimental likelihoods that do not need reverse engineering
2. Agreed-upon nomenclature for leading syst. sources
3. Is reducing dimensionality of published correlation matrices advisable? Is their a standard for it? E.g., fewer nuisance parameters; collect less relevant/certain nuisance parameters into one uncorrelated error; etc.
4. Mathematical consistency of covariance/correlation matrices (see Z. Kassabov et al.)
5. How do different implementations of syst. errors affect pulls on PDFs? L_2 sensitivities to nuisance parameters
6. …

Final remarks

Epistemic uncertainty (due to parametrization, methodology, parametrization/NN architecture, smoothness, data tensions, model for syst. errors, ...) is increasingly important in NNLO global fits as experimental and theoretical uncertainties decrease

Nominal PDF uncertainties in high-stake measurements (ATLAS W mass, Higgs cross sections…) thus should be tested for *control of tensions* and *robustness of sampling over acceptable methodologies.*

Smoothness of Hessian and NN PDFs is another such aspect associated with the prior that should be explored.

Such tests can be done outside of the PDF fits.

Tools for such studies exist using published PDFs and codes: L_2 sensitivities and hopscotch scans.

This is also necessary for combination of PDFs including data correlations
[LHC EW, Jet & Vector boson WGs, https://tinyurl.com/4wcnd8xn; https://tinyurl.com/2p8d8ba3; https://tinyurl.com/2p8tcn5b; Ball, Forte, Stegeman, arXiv:2110.08274].

The ambiguity in NNLO PDFs due to the χ^2 definition is significant. Must consider better formats to propagate experimental likelihoods into the PDF uncertainties. *[See also Cranmer, Prosper, et al., arXiv:2109.04981].*

Uncertainty quantification, a challenge for AI,
As we try to analyze PDFs and understand "why".
With machine learning methods we strive
To make sense of the data and derive.

But uncertainty presents a hurdle
As we try to make predictions and be certain.
It's a challenge that we must face
As we work to improve our models with grace.

Parton distributions, oh how they vex
As we try to understand their complex effects.
But still we persist, for we must know
The secrets that uncertainty has yet to show.

Early Microsoft Bing

Backup

Augmented likelihood for PDFs with global tolerance

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

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

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

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

new experiment

priors on expt. systematics and PDF params

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

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

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

consistent redefinition

5. Inconsistent redefinitions:

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

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

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

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

Why **augmented** likelihood?

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

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

new experiment

priors on expt. systematics and PDF params

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

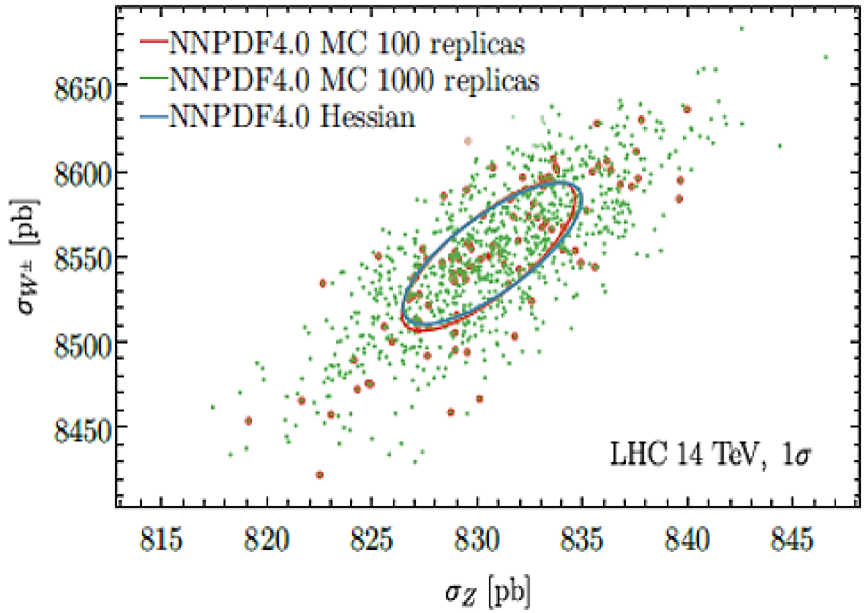
$$\chi^2 = \sum_{i,j}^{N_{pt}} (T_i - D_i)(\mathrm{cov}^{-1})_{ij}(T_j - D_j)$$

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

Computing uncertainty ΔX in the NNPDF analysis

1. By unweighted averaging of predictions for 100 (or 1000) MC replicas:

$$\langle X \rangle = \frac{1}{N_{rep}} \sum_{i=1}^{N_{rep}} X_i \,; \quad \Delta X^2 = \langle (X - \langle X \rangle)^2 \rangle$$



NNPDF calls it "**importance sampling**". The MC replicas are distributed according to the fluctuated data [Ball:2011gg] using the same training algorithm.

This estimates the **aleatory** uncertainty for a given methodology.

Replica 0 is the mean of 1000 MC replicas; has better unfluctuated χ^2 than MC replicas.

2. Using $N_{eig} = 50$ Hessian PDFs.

$$\Delta X^2 = \sum_{i=1}^{N_{eig}} (X_i - X_0)^2 \,.$$

NNPDF4.0 MC and Hessian uncertainties are in a good agreement.

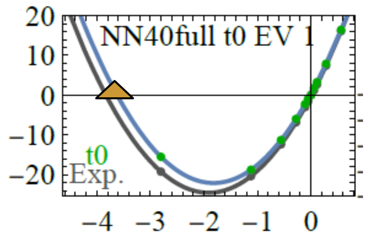
The hopscotch scan counterbalances the bias of the nominal replica ensemble

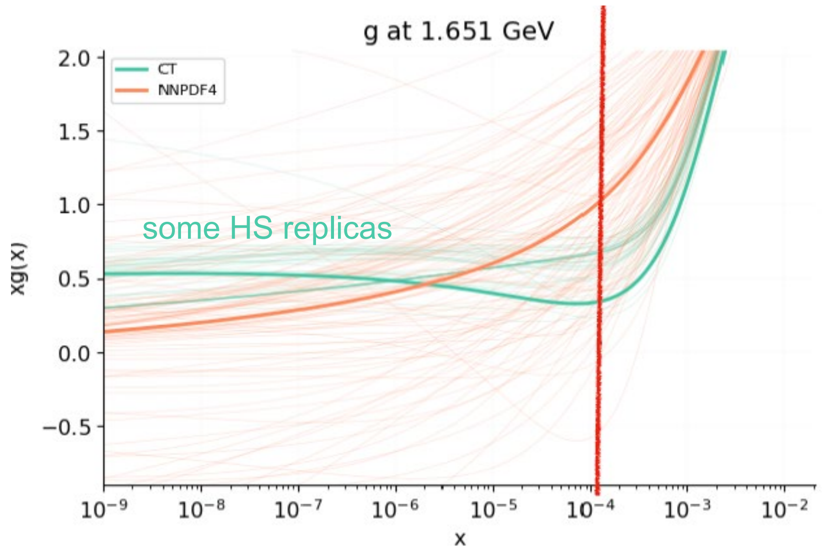
6.2 Creating a less biased sub-sample

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.

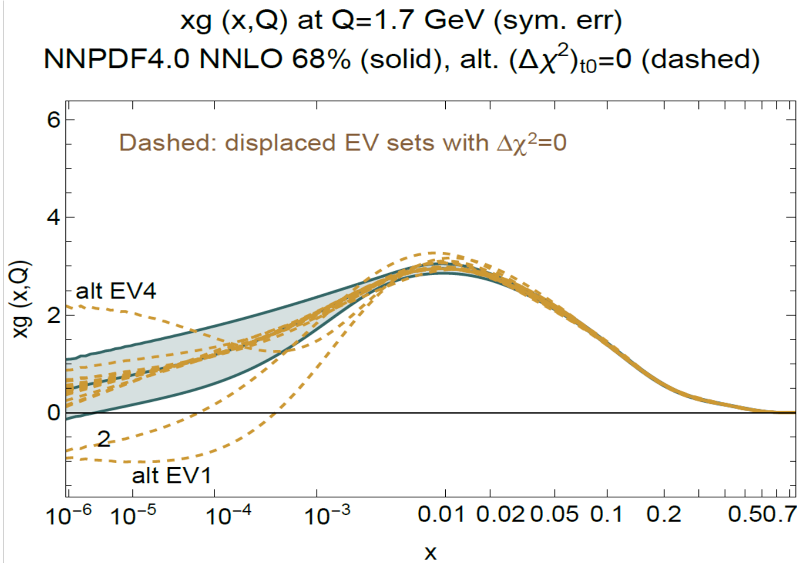
X.-L. Meng, Survey Methodology, Catalogue 12-001-X, vol. 48 (2022), #2

Why doesn't NNPDF4.0 find HS solutions?





NNPDF authors find that some HS replicas fail the initial-stage overfitting test
(M. Ubiali, HP2 2022 workshop, Durham, 2022-09-22)

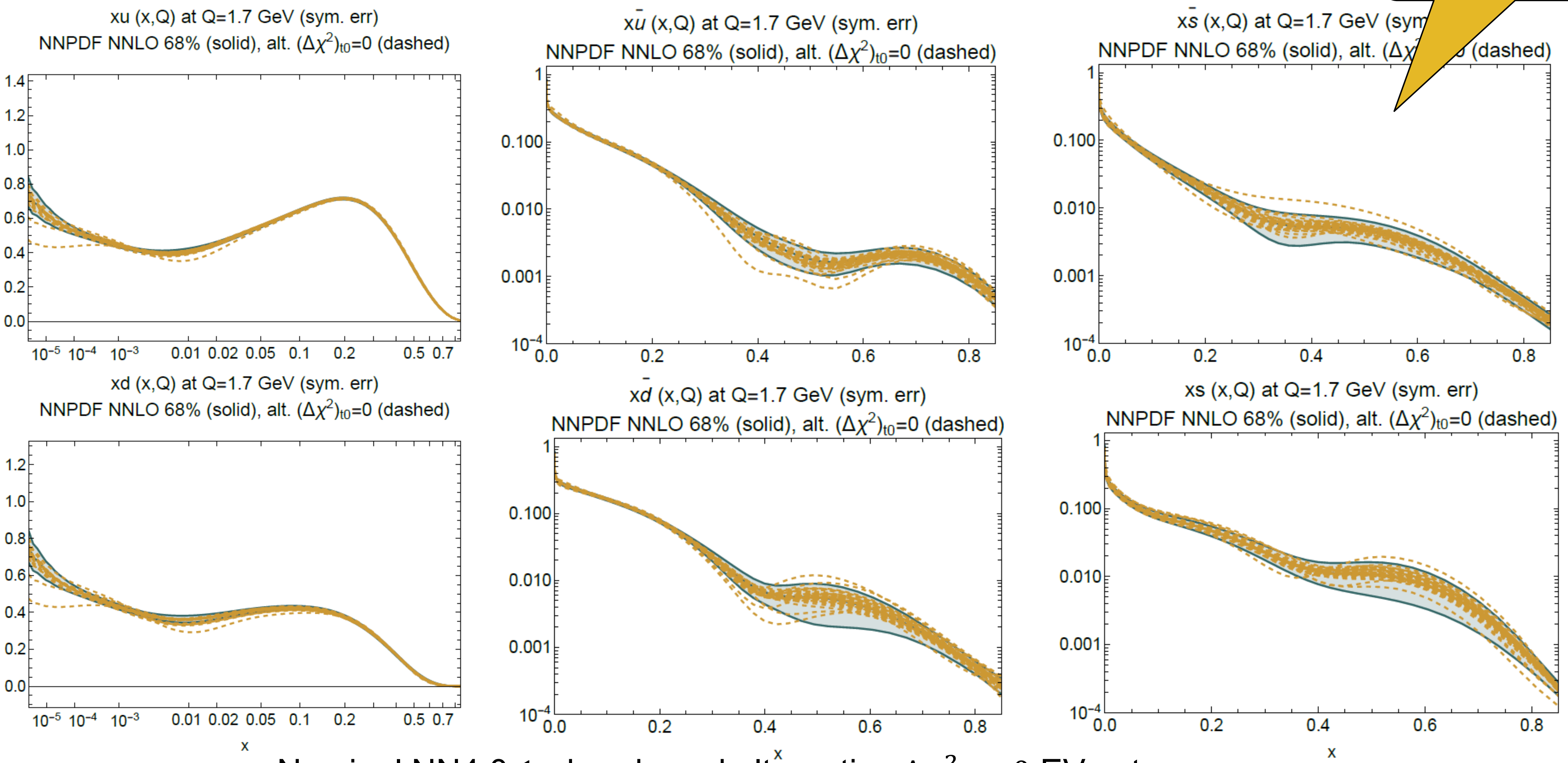


HS solutions have much lower χ^2 than NN MC replicas. HS PDFs are outside the 50-dim neighborhood of NN replica 0. We do not see evidence of "overfitting" according to CT18 criteria.

Hopscotch NN4.0 replicas

Error bands available at https://ct.hepforge.org/PDFs/2022hopscotch/

Smooth behavior of most replicas



Nominal NN4.0 1σ bands and alternative $\Delta\chi^2_{t_0} = 0$ EV sets

Search docs

Getting started

Fitting code: `n3fit`

Code for data: validphys

/ Chi square figures of merit | View page source

Chi square figures of merit

Within the NNPDF methodology various figures of merit are used, each of which can be used in different situations. To avoid confusion, it is important to understand the differences between the various figures of merit, and to understand which definition we are referring to in a given context. In particular, it is worth stressing that whenever a figure of merit is discussed, the t_0 method (discussed below) applies.

Note

From NNPDF2.0 onwards the t_0 formalism has been used to define the figure of merit used during the fitting of the PDFs.

Note

The t_0 method is **not** used by default in other validphys applications, and instead the default is to compute the experimental χ^2. To compute $\chi^2_{t_0}$, users need to specify

```
use_t0: True
t0pdfset: <Some LHAPDF set>
```

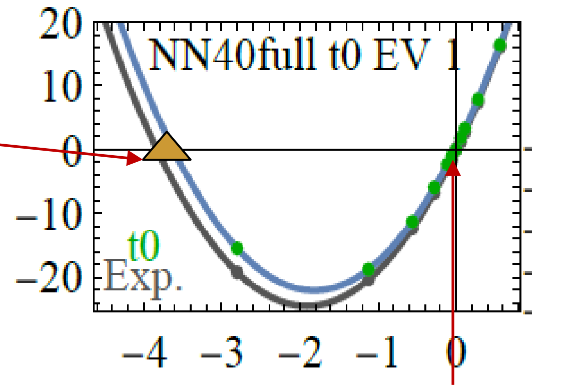
in the relevant namespace. This will instruct actions such as `validphys.results.dataset_chi2_table()` to compute the t_0 estimator.

https://docs.nnpdf.science/figuresofmerit/index.html, accessed on 2023-03-28

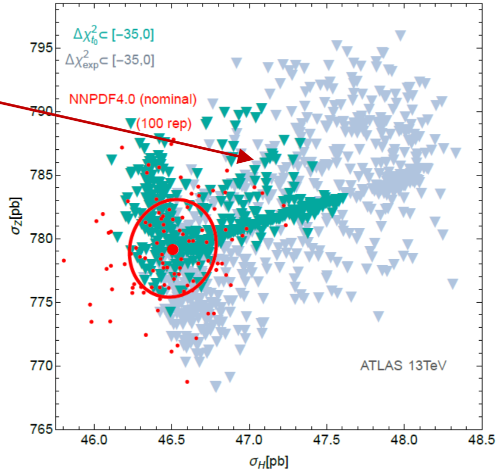
Hopscotch NN4.0 replicas

LHAPDF6 grids available at https://ct.hepforge.org/PDFs/2022hopscotch/

1. Alternative (second) EV sets with $\Delta\chi^2 = 0$, for 50 EV directions



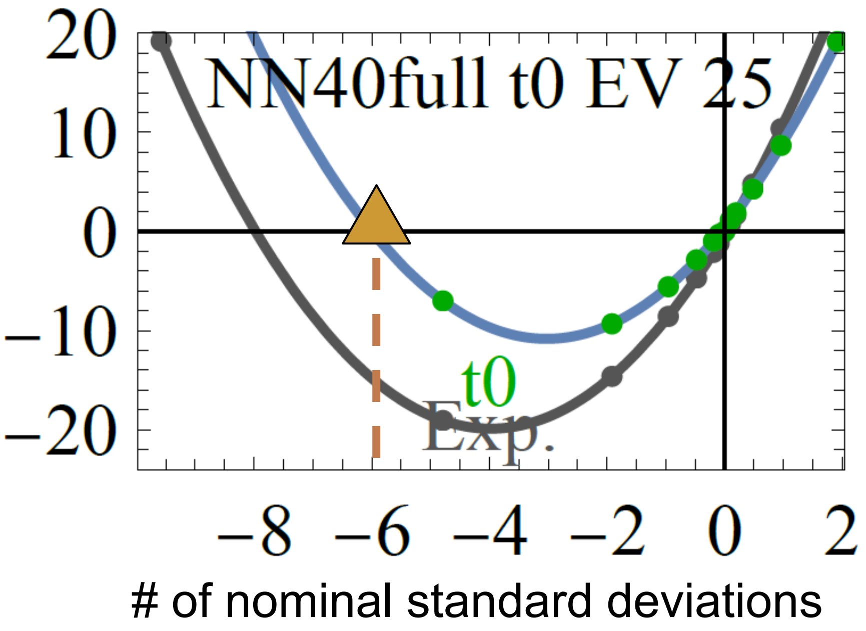
2. A total 2329 PDF sets from hopscotch scans on $\sigma_Z, \sigma_{W^+}, \sigma_{W^-}, \sigma_H, \sigma_{t\bar{t}}$ total inclusive cross sections at the LHC 13 TeV

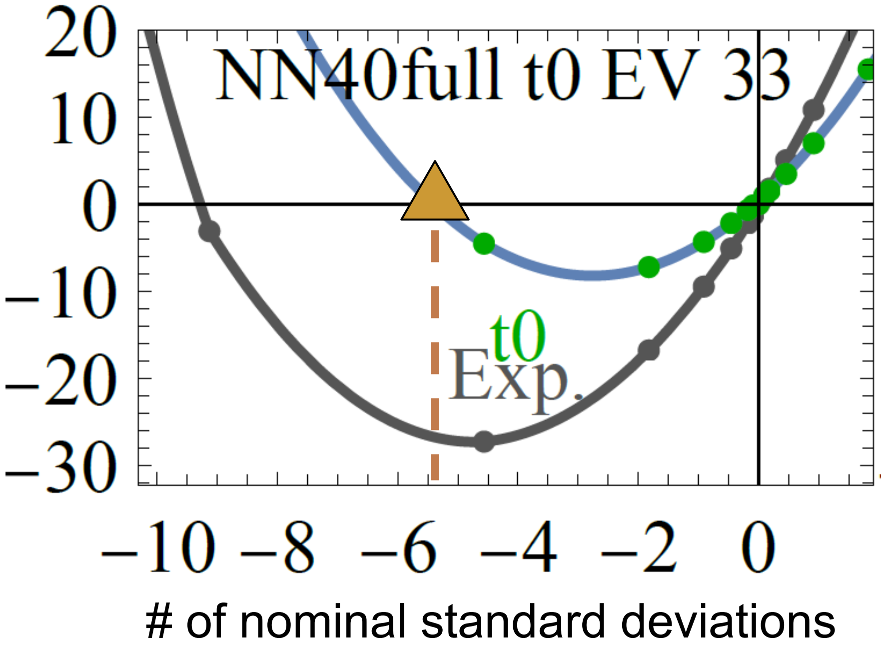


For $\chi^2_{t_0}$ and χ^2_{exp} definitions in the NNPDF4.0 code

Codes to generate LHAPDF grids for hopscotch replicas available by request.

Scans of the log-likelihood in EV directions 25 and 33





PRIOR PROBABILITY IN PDF FITS

✓ PDF fitting example of inverse problem: aim to find a posterior probability of ***f*** given the data ***D***.

✓ Parametrization of PDFs: finite-dimensional problem.

$$f(x) \approx \tilde{f}(x, \theta) \in \mathcal{F}$$

✓ The posterior probability for the parametrization depends on both the figure of merit that maximises the data likelihood given the parameters and on prior probability ***H.***

(M. Ubiali, HP2 2022 workshop, Durham, 2022-09-22)