

News from the CTEQ-TEA group

Marco Guzzi (Kennesaw State University, USA)
and **Pavel Nadolsky** (Southern Methodist University, USA)

With the CTEQ-TEA (Tung Et. Al.) working group

China: A. Ablat, S. Dulat, Y. Fu, T.-J. Hou, I. Sitiwaldi

Mexico: A. Courtoy

USA: M.G., P. N., T.J. Hobbs, J. Huston, H.-W. Lin, C. Schmidt, K. Xie, C.-P. Yuan

and other coauthors



2024-07-11



Guzzi & Nadolsky, LHC EW WG

• RESEARCH PROJECTS AND RESULTS •

<https://cteq-tea.gitlab.io/>

- CTEQ-TEA publications from INSPIRE
- LHAPDF grids for parton distributions
 - CT18 (N)NLO, CT18 QED, CT18 FC, ...
 - Subtracted heavy-quark PDFs in the S-ACOT-MPS scheme
- Public codes
 - ePump (Hessian updating for PDFs with tolerance > 1)
 - LHAexplorer (fast surveys of data using L2 sensitivities)
 - Fantômas (Bezier parametrizations)
 - mp4lhc/mcgen (MC PDFs, combination of PDFs)
 - ...

CT18up enhanced precision LHAPDF grids (2023)

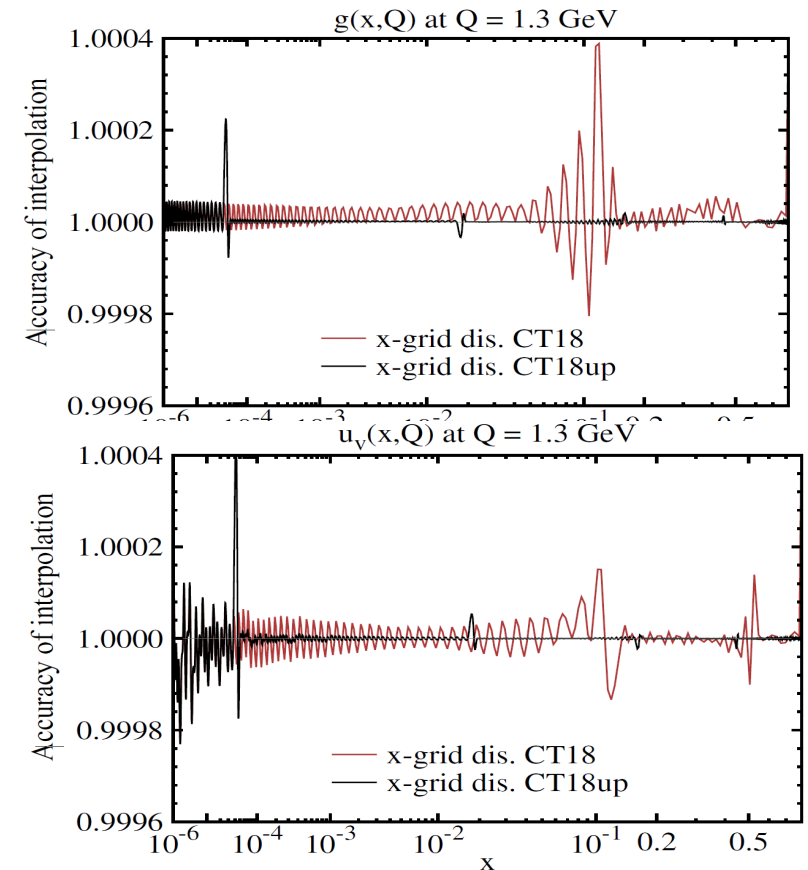
On <https://cteq-tea.gitlab.io/project/00pdfs/>

- CT18, A, X, Z NNLO PDFs (2019 edition) presented as LHAPDF grids with a 1.9x higher number of x and Q nodes
- Same PDFs as in the LHAPDF library, with even more precise interpolation at $10^{-4} \leq x \leq 1$
- Recommended for high-mass and precision calculations; 2019 grids ok in other cases

Numbers of x, Q nodes in LHAPDF grids

intervals in Q	CT18	CT18up
$[Q_0, m_c]$	2	4
$[m_c, m_b]$	8	11
$[m_b, m_t]$	14	18
$[m_t, Q_{\max}]$	13	16
Total	37	49

intervals in x	CT18	CT18up	intervals in x	CT18	CT18up
$[10^{-10}, 10^{-9}]$	1	1	$[0.1, 0.2]$	7	18
$[10^{-9}, 10^{-8}]$	11	11	$[0.2, 0.3]$	6	16
$[10^{-8}, 10^{-7}]$	12	12	$[0.3, 0.4]$	5	12
$[10^{-7}, 10^{-6}]$	11	11	$[0.4, 0.5]$	3	13
$[10^{-6}, 10^{-5}]$	12	12	$[0.5, 0.6]$	6	15
$[10^{-5}, 10^{-4}]$	11	15	$[0.6, 0.7]$	6	12
$[10^{-4}, 10^{-3}]$	12	23	$[0.7, 0.8]$	8	11
$[10^{-3}, 10^{-2}]$	11	23	$[0.8, 0.9]$	14	17
$[10^{-2}, 0.1]$	12	40	$[0.9, 1]$	15	38
Total	161	300			



Toward a new generation of CT202X PDFs

1. Multiple preliminary NNLO fits with LHC Run-2 (di)jet, vector boson, $t\bar{t}$ data
 - based on the selections of experiments recommended in [2305.10733](#), [2307.11153](#)
2. Work on implementation of N3LO contributions
3. Next-generation PDF uncertainty quantification: Bézier curves, META combination, ML stress-testing, multi-Gaussian approaches, ...
4. Physics applications
 - a. QCD+QED PDFs for a neutron (K. Xie et al., [2305.10497](#))
 - b. PDF dependence of forward-backward asymmetry (Y. Fu et al., [2307.07839](#))
 - c. An L2 sensitivity study using xFitter (L. Kotz, [2401.11350](#))
 - d. Fantômas Pion PDFs (L. Kotz et al., [arXiv:2311.08447](#))
 - e. AI/ML models for PDF generation (Kriesten and Hobbs, [arXiv:2312.02278](#), [2407.03411](#))

NNLO fits with new data at 8 and 13 TeV

Example

χ^2/N_{pt} for CT18+new data (CT18 in parentheses) NNLO fits; 68% CL

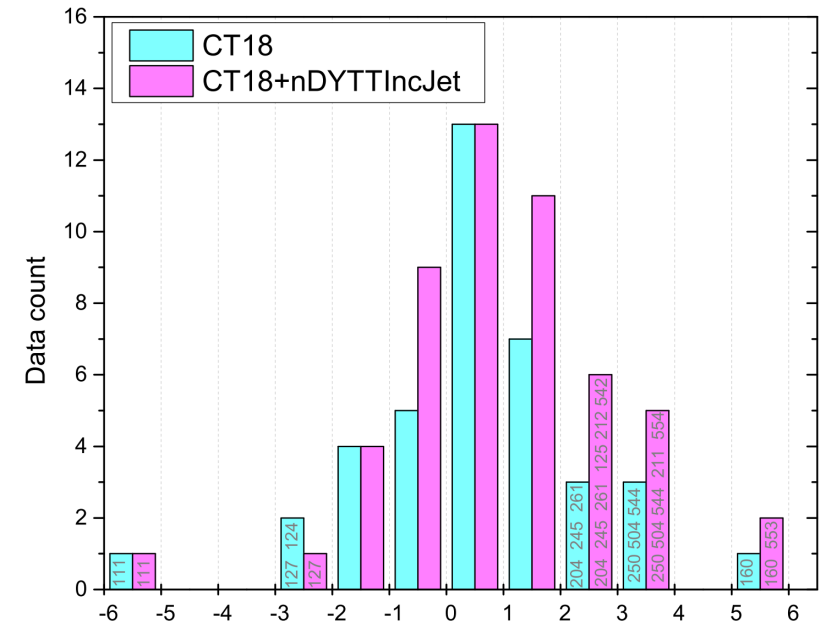
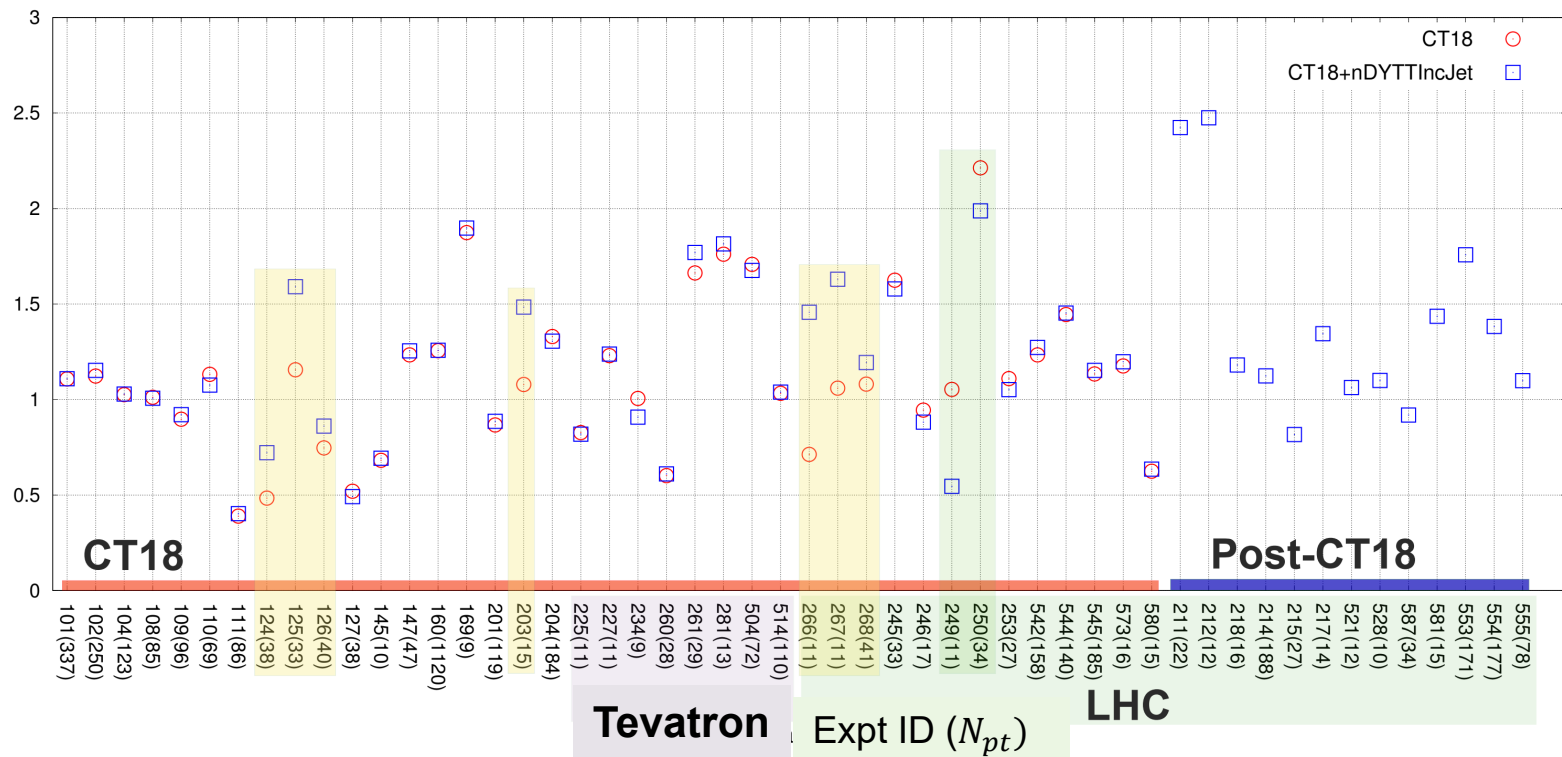
ID	Exp	N_{pt}	χ^2/N_{pt}		
Drell-Yan					
215	ATLAS 5.02 TeV W,Z	27	$0.82^{+0.55}_{-0.16}$	($1.15^{+1.22}_{-0.43}$)	} nDY
211	ATLAS 8 TeV W	22	$2.42^{+2.49}_{-1.51}$	($4.25^{+6.39}_{-3.34}$)	
214	ATLAS 8 TeV Z3D	188	$1.12^{+0.46}_{-0.02}$	($1.99^{+5.10}_{-1.85}$)	
212	CMS 13 TeV Z	12	$2.48^{+4.76}_{-0.88}$	($12.03^{+38.04}_{-21.84}$)	
217	LHCb 8 TeV W	14	$1.35^{+0.59}_{-0.61}$	($1.35^{+0.72}_{-0.64}$)	
218	LHCb 13 TeV Z	16	$1.18^{+1.42}_{-0.60}$	($1.49^{+1.74}_{-0.89}$)	
13 TeV $t\bar{t}$					
521	ATLAS all-hadronic $y_{t\bar{t}}$	12	$1.06^{+0.14}_{-0.09}$	($1.05^{+0.21}_{-0.10}$)	} nTT
528	CMS dilep $y_{t\bar{t}}$	10	$1.10^{+1.08}_{-0.68}$	($1.03^{+1.60}_{-0.74}$)	
587	ATLAS lep+Jet $m_{t\bar{t}} + y_{t\bar{t}} + y_{t\bar{t}}^B + H_T^{t\bar{t}}$	34	$0.92^{+0.32}_{-0.14}$	($0.94^{+0.59}_{-0.16}$)	
581	CMS lep+jet $m_{t\bar{t}}$	15	$1.44^{+1.18}_{-0.73}$	($1.37^{+1.86}_{-0.82}$)	
Inclusive Jet					
553	ATLAS 8 IncJet	171	$1.76^{+0.20}_{-0.12}$	($1.80^{+0.33}_{-0.16}$)	} nIncJet
554	ATLAS 13 IncJet	177	$1.38^{+0.13}_{-0.10}$	($1.39^{+0.20}_{-0.11}$)	
555	CMS 13 IncJet	78	$1.10^{+0.24}_{-0.17}$	($1.11^{+0.30}_{-0.16}$)	

Fits with 1 type of new data

A fit with all 3 types

A 3-data-type fit (CT18+nDYTTIncJet)

χ^2/N_{pt}



$$S_n \approx (\chi^2 - N_{pt}) / \sqrt{2N_{pt}}$$

The most precise new experiments tend to have an elevated χ^2/N_{pt} , in the same pattern as observed for CT18

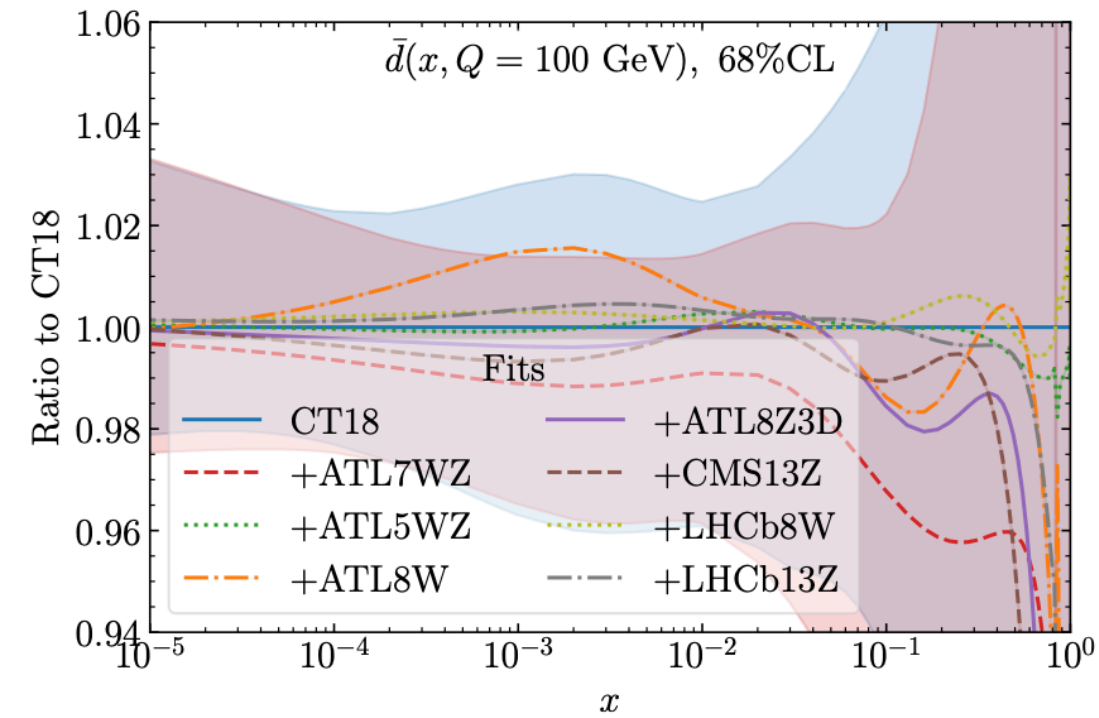
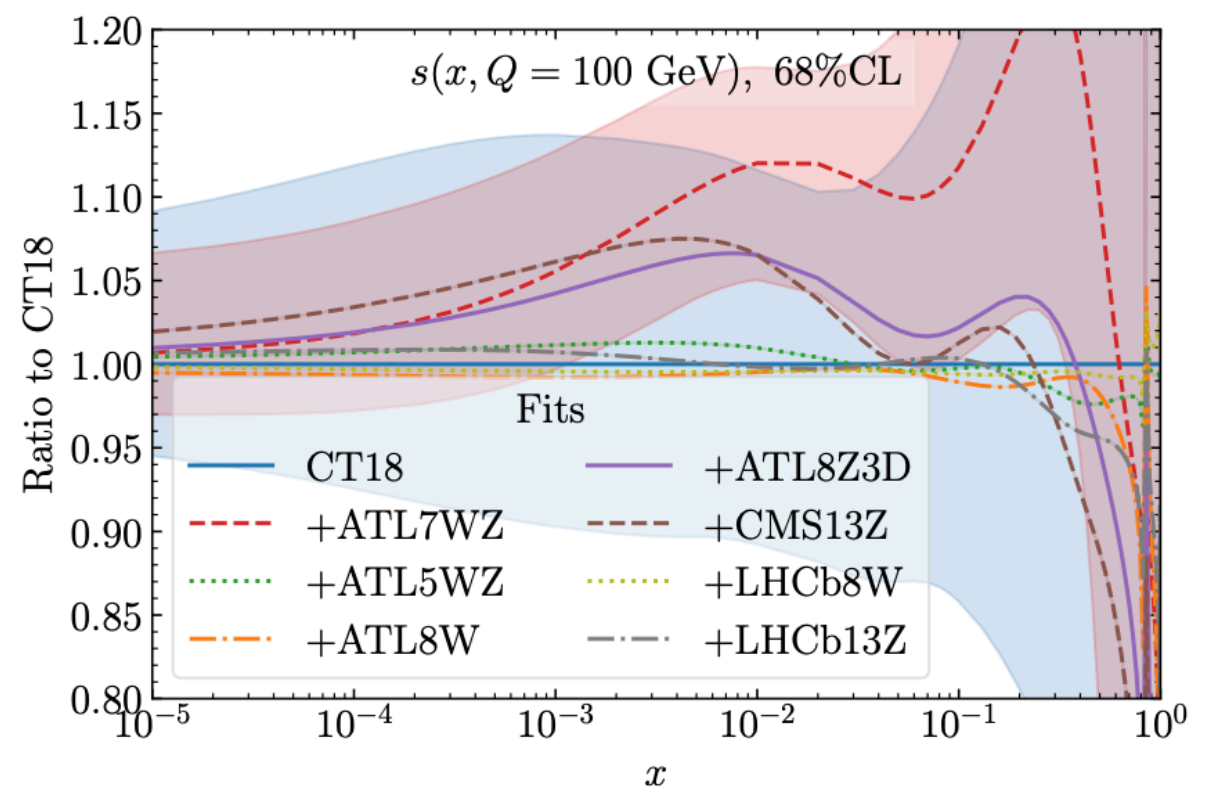
χ^2/N_{pt} increases for experiments 124 and 125 (NuTeV), 126 and 127 (CCFR) and 203 (E866 DY), 266 and 267 (CMS 7TeV Ach), 268 (ATLAS 7TeV W, Ach).

χ^2/N_{pt} decreases for experiments 249 (CMS 8 TeV Ach), 250 (LHCb 8 TeV W/Z)

Post-CT18 Drell-Yan data's impact

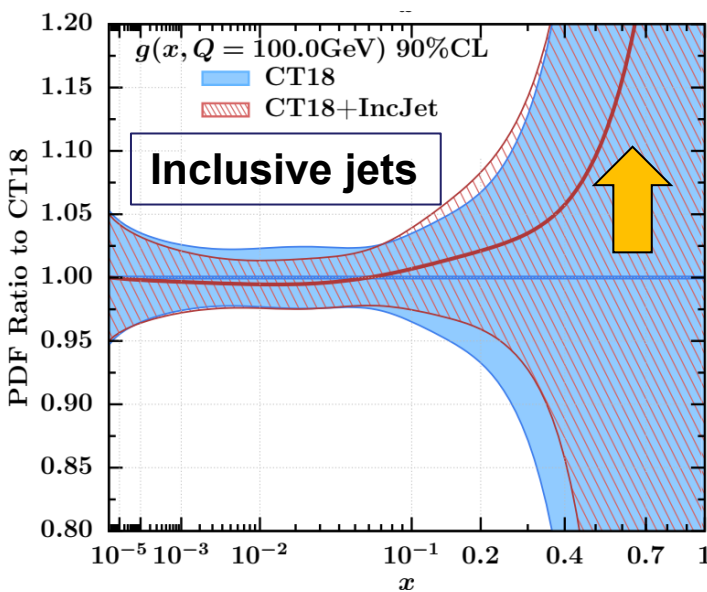
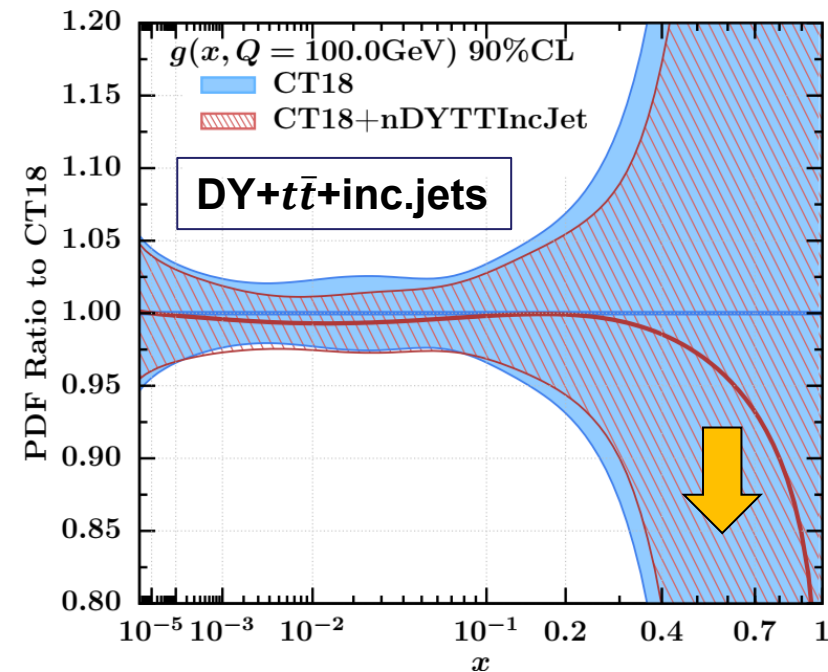
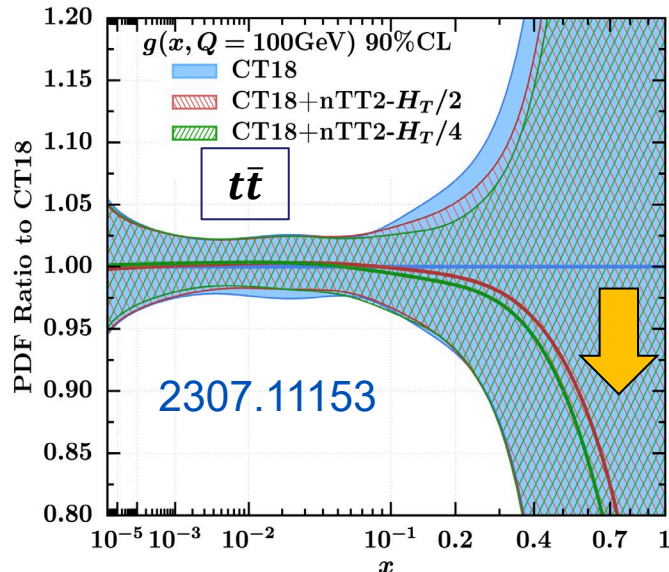
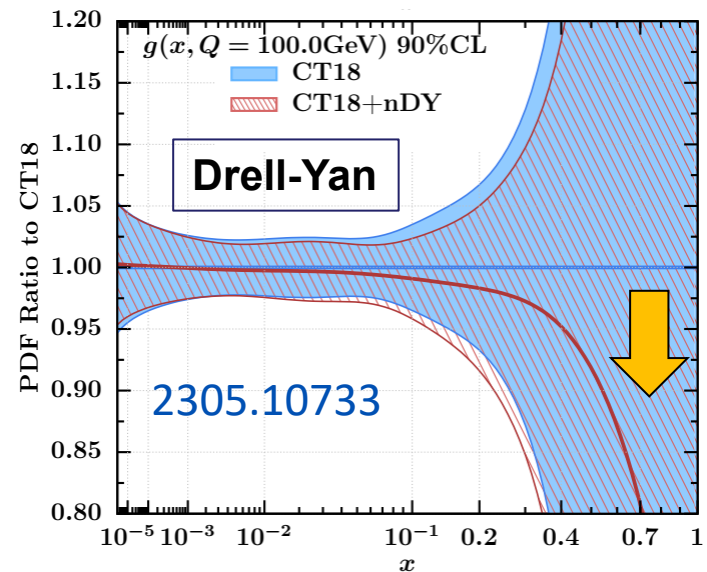
2305.10733 (PRD23')

ID	Experiment	N_{pt}			
			CT18	CT18A	CT18As
215	ATLAS 5.02 TeV W, Z	27	0.81	0.71	0.71
211	ATLAS 8 TeV W	22	2.45	2.63	2.51
214	ATLAS 8 TeV Z 3D [†]	188	1.12	1.14	1.18
212	CMS 13 TeV Z	12	2.38	2.03	2.71
216	LHCb 8 TeV W	14	1.34	1.36	1.43
213	LHCb 13 TeV Z	16	1.10	0.98	0.83
248	ATLAS 7 TeV W, Z	34	2.52	2.50	2.30
Total 3994/3953/3959 points			1.20	1.20	1.19



- Many new Drell-Yan (nDY) data came out after the release of CT18 PDFs.
- We found that most of the nDY data sets are consistent with the ATLAS 7 WZ precision data (16') and prefer enhanced strangeness at $x \sim 0.02$
- Only one exception: ATL8W has an opposite pull on d, \bar{d}
- CMS13Z and ATL8W have a similar χ^2/N_{pt} as ATL7WZ
- The more flexible strangeness parameterization in CT18As can relax the tension, but not completely resolve it.

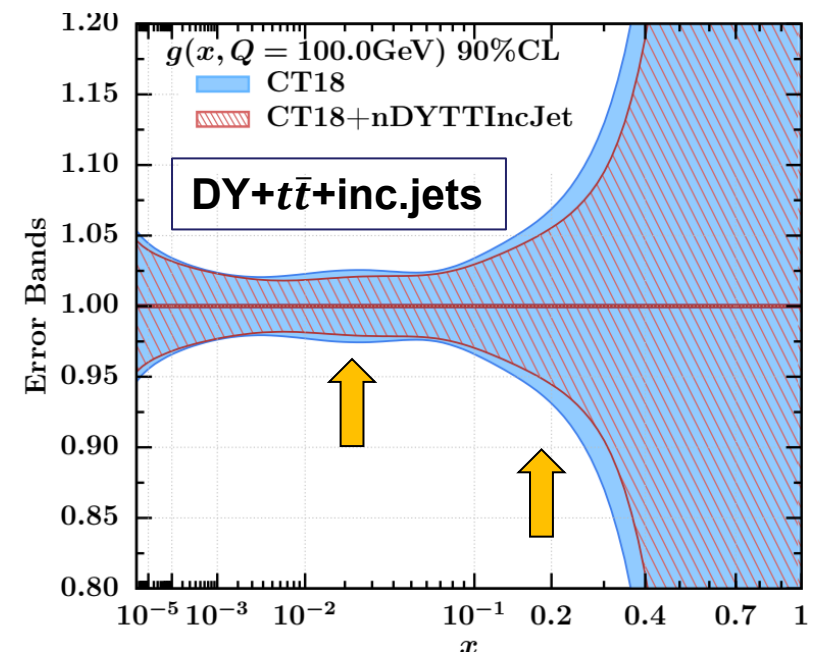
Pulls on the gluon PDF by the new data type



After including DY, $t\bar{t}$, and inc. jet data simultaneously, we get a softer gluon. Note that new DY and $t\bar{t}$ data favor a softer gluon, new inc. jet data prefer a harder gluon.

Mild changes in the gluon uncertainty

PRELIMINARY



Necessary components of an N3LO PDF analysis

Component		Availability
Splitting functions		Partial N3LO
Hard cross sections	• DIS, light flavors	Full N3LO
	• NC DIS, heavy flavors	Full N3LO (Blümlein et al.), not yet in fitting codes
	• Vector boson production	Full N3LO for some processes, fixed N3LO/NLO K-factor tables
	• CC DIS, jet, $t\bar{t}$ production	N2LO
	• $pp \rightarrow W + c$, $pp \rightarrow Z + b$, $pp \rightarrow b$	NLO (massive); NNLO (ZM)

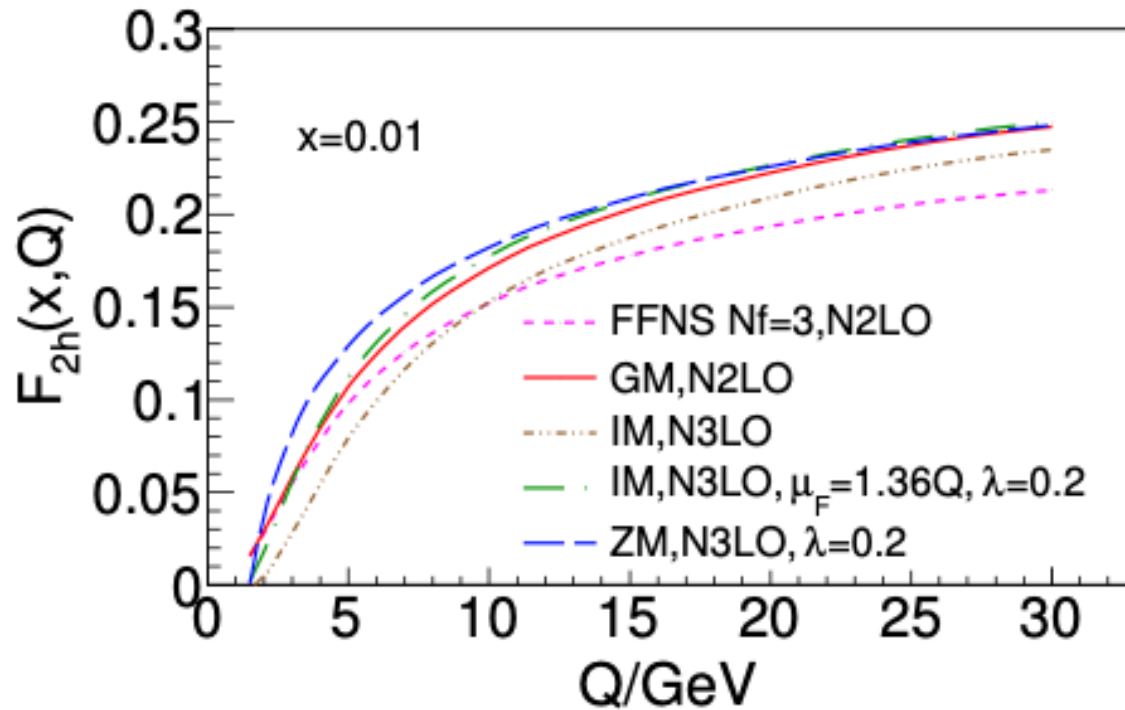
Looking forward to including all components **exactly and fully** to reduce the QCD scale uncertainty and guarantee the N3LO accuracy in the near future.

CTEQ-TEA and other groups include some N3LO contributions in their fitting codes: recent progress of MSHT and NNPDF in partial N3LO (aN3LO) fits

These partial N3LO calculations mostly agree with N2LO within their scale dependence

For $gg \rightarrow H^0$ production, the aN3LO-N2LO difference is comparable to other effects due to the remaining scale dependence, selection of experiments, treatment of systematic uncertainties

QCD cross sections @N3LO



- **DIS:** The CTEQ-TEA code implements complete flavor decompositions of DIS SFs at N3LO using approximate zero-mass Wilson coefficients with a rescaling variable (the **Intermediate-Mass VFN scheme**, cf. the figure)

Boting Wang's and Keping Xie's Theses, SMU

- **Working on the implementation of massive N3LO heavy-quark coefficients to obtain N3LO DIS cross sections in the SACOT-MPS General-Mass VFN scheme**

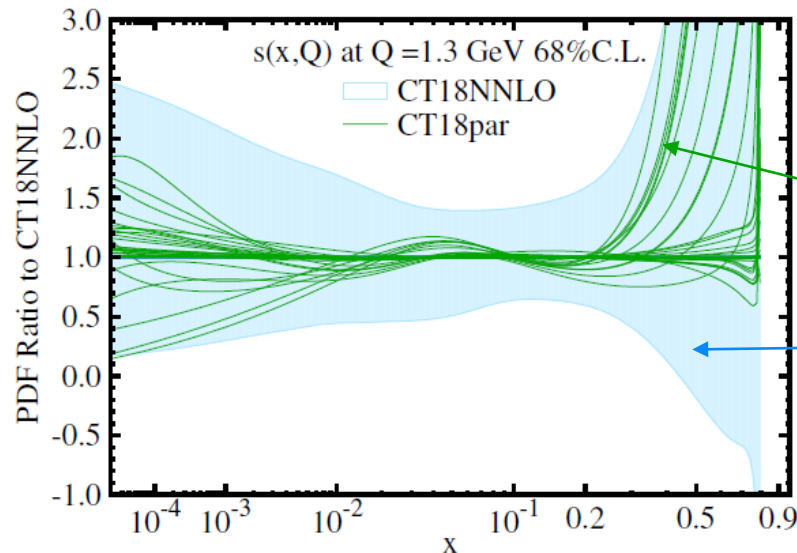
Factorization schemes	Mass dependence in the FC terms	Mass dependence of the FE and subtraction terms	Introduce heavy-quark PDFs at large Q
FFN	Exact	N/A	no
ZM	None	None	yes
IM	Approximate	Approximate	yes
GM	Exact	Approximate	yes

- **DGLAP evolution** is performed at N3LO with APFEL/APFEL++.
- **Drell-Yan:** Ongoing work to include N3LO DY effects using NNLO ApplFast + N3LO/N2LO K-factor tables

Taming PDF uncertainties in CT202X PDFs

Several efforts to refine PDF uncertainty quantification:

- understand conceptual underpinnings of the multivariate inverse problem. Much can be learned from non-HEP statistics applications
- suppress aleatory and perturbative uncertainties (e.g., from higher-order contributions)
- comprehensively estimate epistemic uncertainties (e.g., due to the PDF parametrization forms)



CT approach: “Bayesian exploration with Gaussian emulation”

preliminary PDFs for alternative parametrizations

final uncertainty with one parametrization

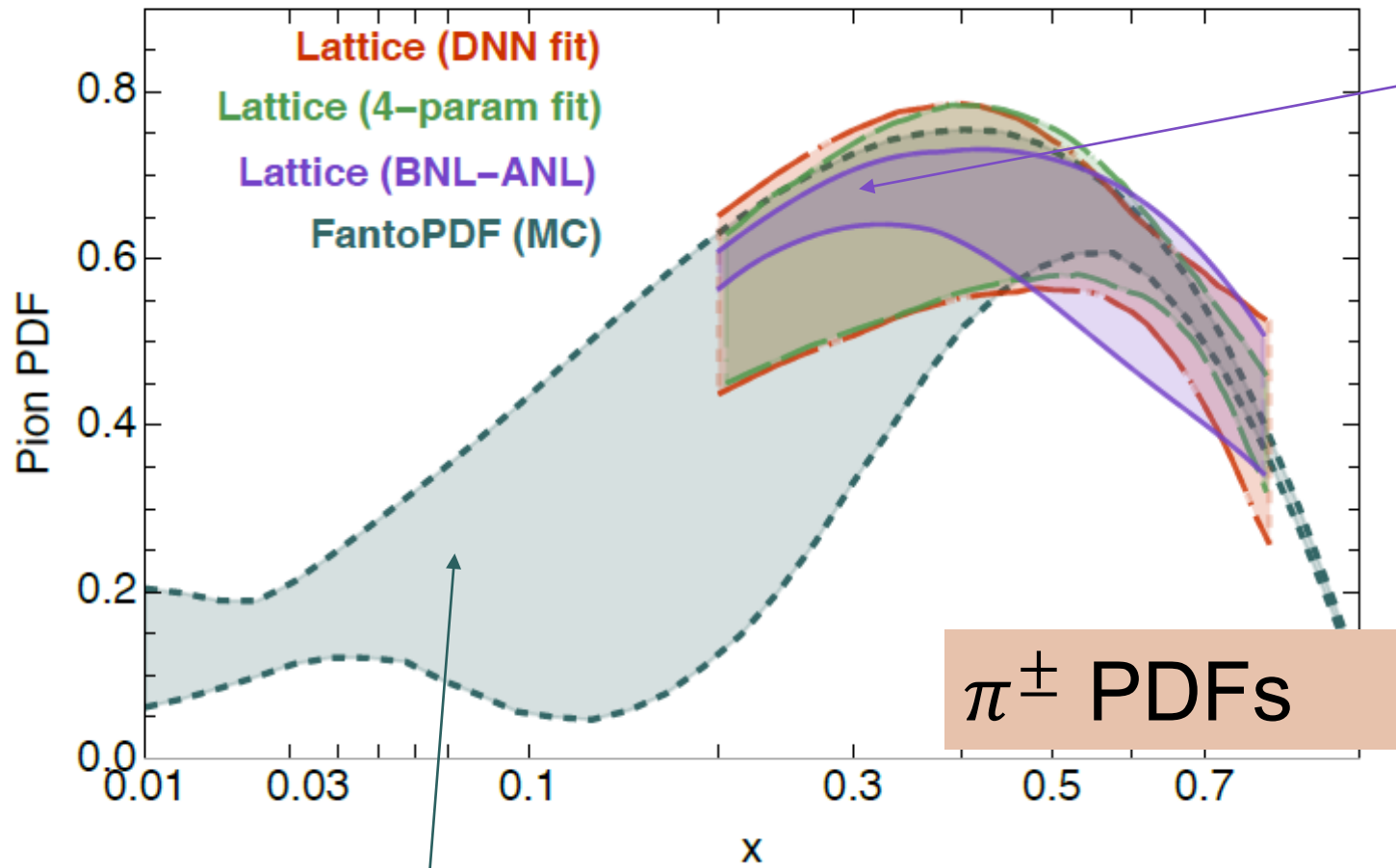
Preliminary fits explore experimental, theoretical, parametrization, methodological uncertainties

The final Hessian error set (50-60) approximates the total uncertainty due to the above factors.

Fantômas + mp4lhc 2.0: pion PDFs with advanced parametrization uncertainties

L. Kotz, A. Courtoy, M. Chavez, P. Nadolsky, F. Olness, and others, arXiv:2309.00152, arXiv:2311.08447

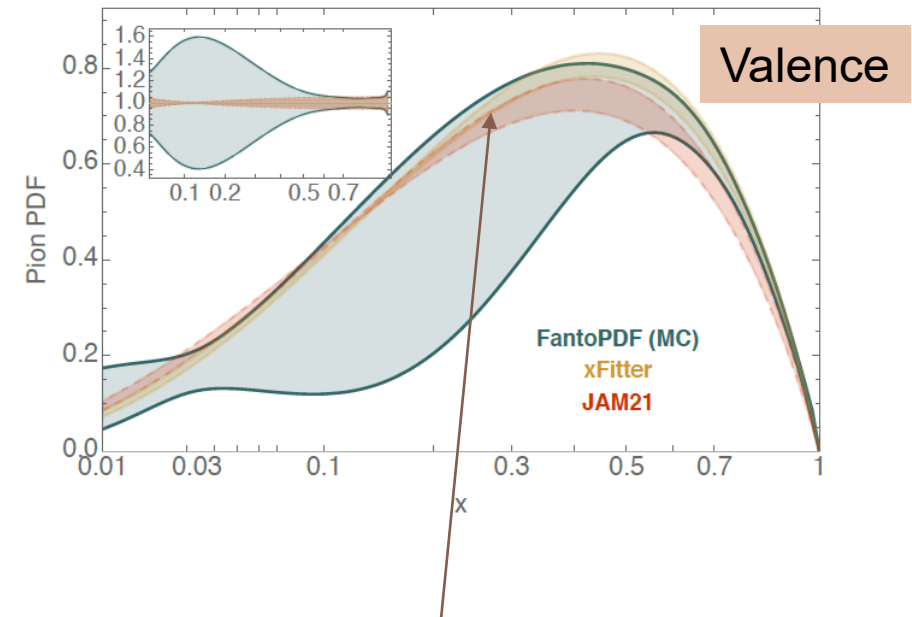
$xV(x, Q)$ at $Q=2. \text{ GeV}$, 68% c.l. (band)



Phenomenological analysis, including the parametrization dependence

are the lattice uncertainties fully estimated?

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



without parametrization dependence

Fantômas + mp4lhc 2.0: pion PDFs with advanced parametrization uncertainties

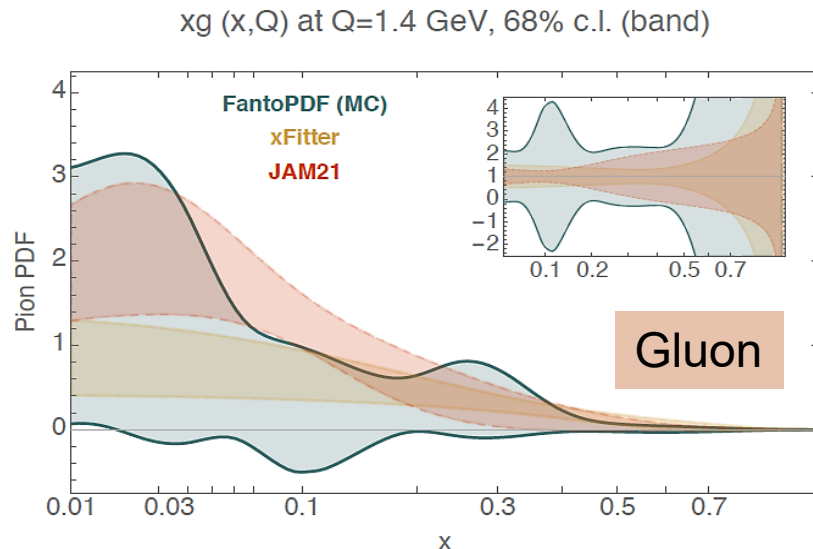
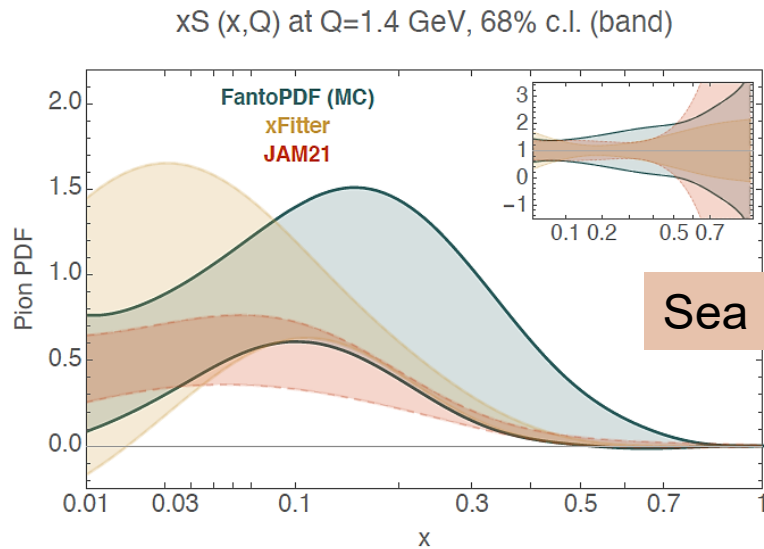
L. Kotz, A. Courtoy, M. Chavez, P. Nadolsky, F. Olness, and others, arXiv:2309.00152, arXiv:2311.08447

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

These polynomial curves are universal approximators.

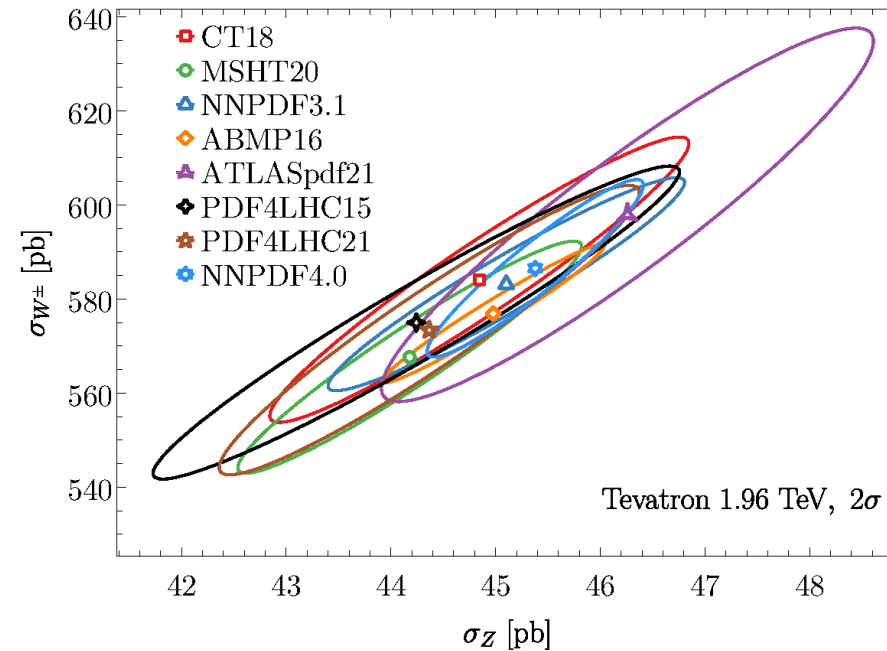
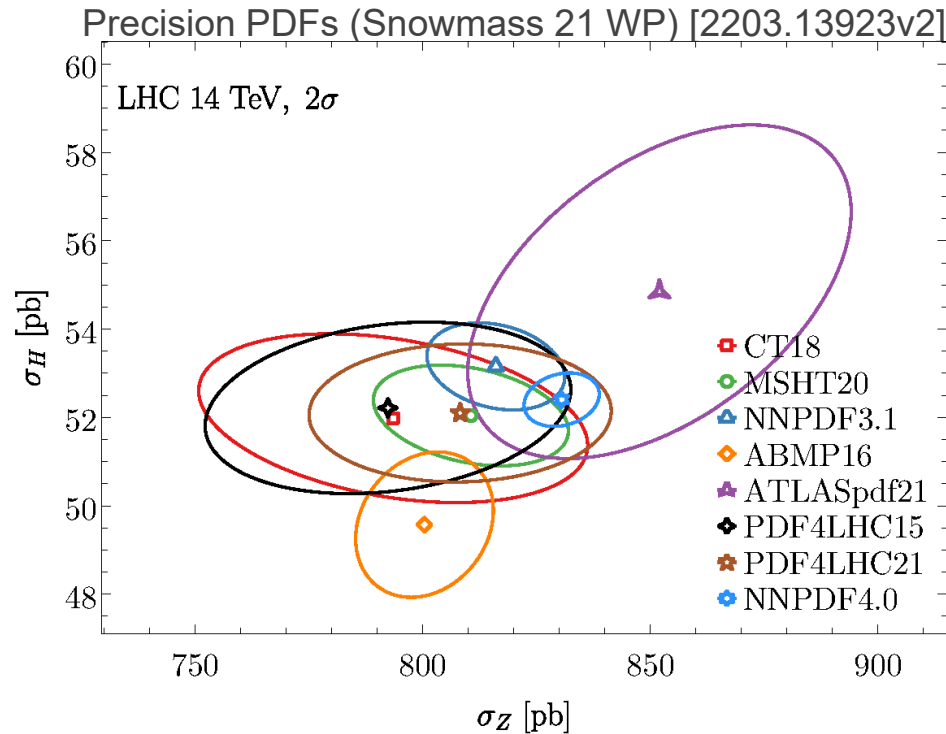
The Fantomas PDF error band is based on ~ 100 alternative parametrization forms with the same or better χ^2 as in the 2021 xFitter study [Novikov et al., arXiv:2002.02902]

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



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 α_s and M_W measurement. Important differences are traced to treatments of epistemic uncertainties.

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

Some thoughts on studies to understand differences between precision PDF sets

- General theme: **increasing precision** must be balanced by **replicability** of results.

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

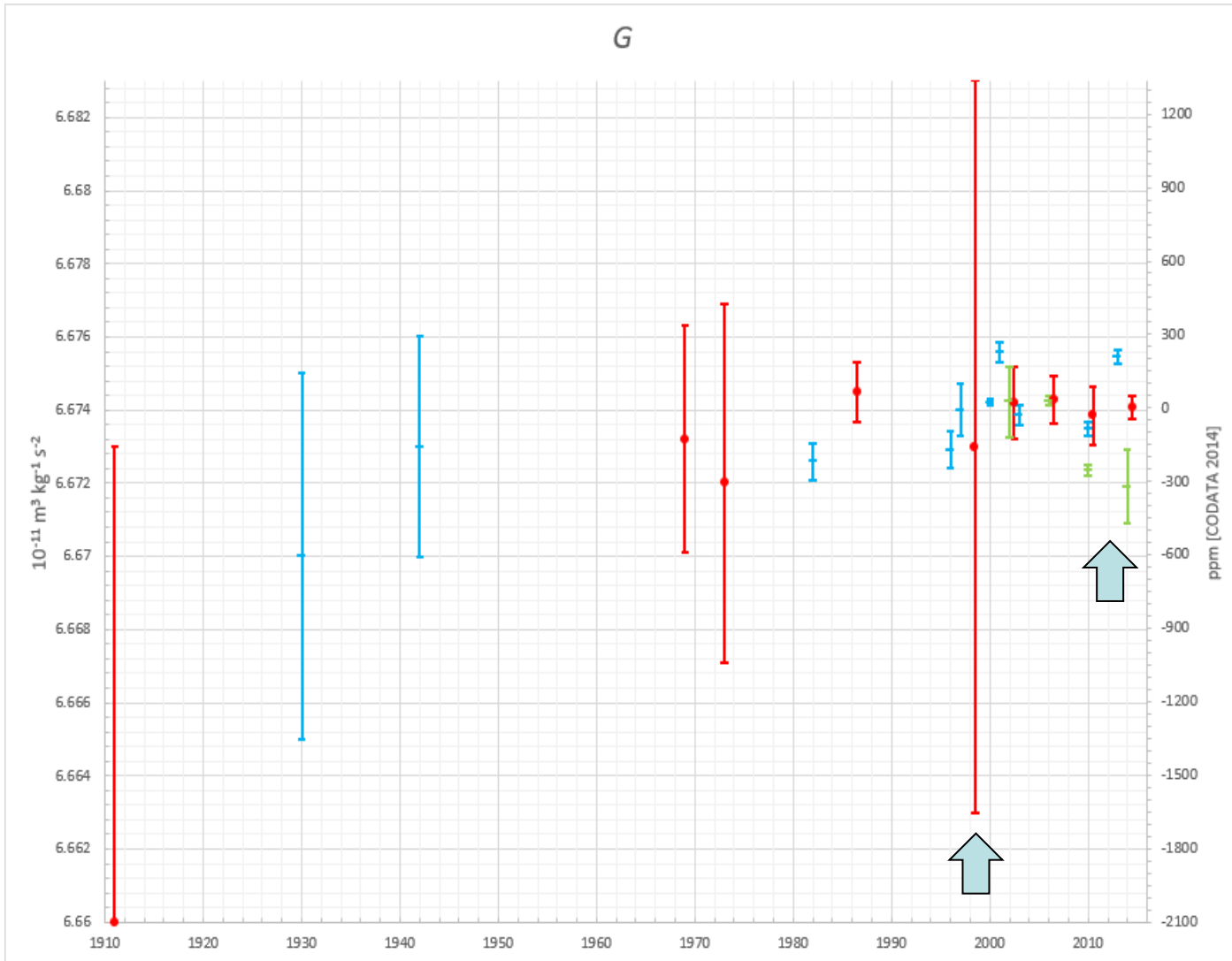
- Partial implementation of N3LO is an important step. N3LO contributions will be under control when fully implemented in key processes.

- Fundamental issues in propagating systematic uncertainties create a risk of negating N3LO improvements. Tensions between experiments and different χ^2 definitions have a large effect.

Can LHC collaborations publish streamlined (log-)likelihoods as standards for PDF fits?

- With $O(10 - 10^3)$ free parameters, including nuisance parameters, the $\Delta\chi^2 = 1$ criterion for 1σ PDF uncertainties with a fixed parametrization is almost certainly incomplete. Stop using it “as is”. There are strong mathematical reasons.
- PDF profiling is a fast estimate of the impact of a new data set, not a replacement for including the data into a global fit. It should use the same definition of χ^2 tolerance as in the global fit, or the uncertainties are misestimated. No standard profiling for PDF+ α_s dependence. The **ePump** program realizes a Hessian updating method that is consistent with the CT and MSHT tolerances.

World average for the gravitational constant

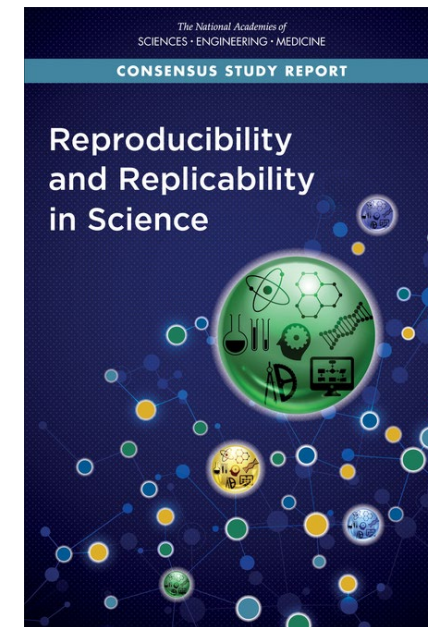


Timeline of measurements and recommended values for G since 1900: values recommended based on the NIST combination (red), individual torsion balance experiments (blue), other types of experiments (green).

The combination error bars are unstable after 1995

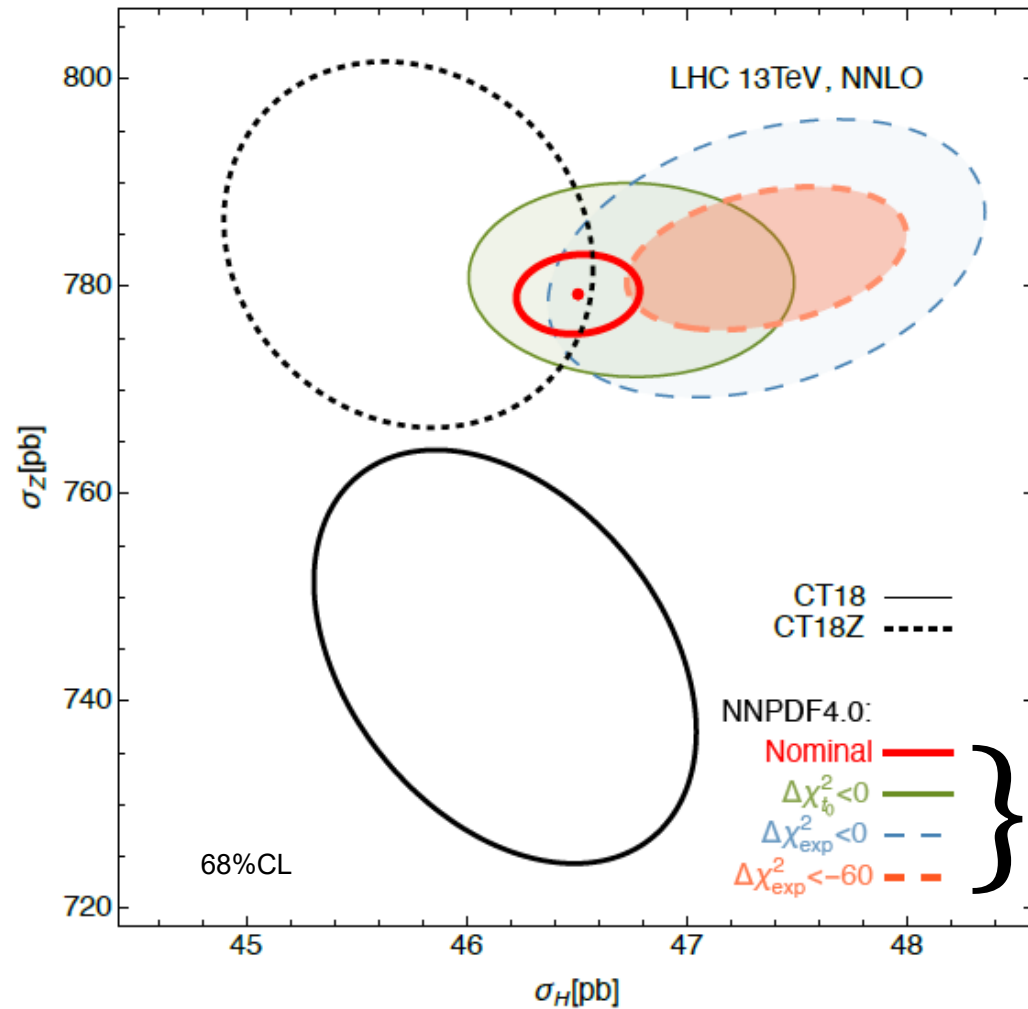
Some latest precise measurements conflict among themselves and with the post-2014 combination

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



Example: different χ^2 treatments produce discrepant uncertainty estimates

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



obtained with the same NNP4.0 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 does not cover them!

PDF wish list for systematic uncertainties

A proposal

Fundamental issues in propagating systematic uncertainties. Some possible remedies:

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 there 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. ...

Shown at the PDF4LHC meeting in Nov. 2023

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:

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

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

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

Also the next slide.

Augmented likelihood for PDFs with global tolerance

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

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

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

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

new experiment
priors on expt. systematics
and PDF params

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

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

5. **Inconsistent redefinitions:**

$$\chi^{2'}(\vec{\lambda}_{\text{exp}}, \vec{\lambda}_{\text{th}}) = \sum_{i=1}^{N_{pt}} \frac{[D_i + \sum_{\alpha} \beta_{i,\alpha}^{\text{exp}} \lambda_{\alpha,\text{exp}} - T_i - \sum_{\alpha} \beta_{i,\alpha}^{\text{th}} \lambda_{\alpha,\text{th}}]^2}{s_i^2} + \sum_{\alpha} \lambda_{\alpha,\text{exp}}^2 + \sum_{\alpha} \lambda_{\alpha,\text{th}}^2 \quad \text{and } P(D_i|T_i) \propto e^{-\chi^{2'}/2}$$

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

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

THANK YOU FOR YOUR ATTENTION!

Bézier curve

Bézier curves are convenient for interpolating discrete data

The interpolation through Bézier curves is unique if the polynomial degree = (# points - 1), there's a closed-form solution to the problem,

$$\mathcal{B}^{(n)}(x) = \sum_{l=0}^n c_l B_{n,l}(x)$$

with the Bernstein pol.

$$B_{n,l}(x) \equiv \binom{n}{l} x^l (1-x)^{n-l}.$$

The Bézier curve can be expressed as a product of matrices:

- \underline{T} is the vector of x^l
- $\underline{\underline{M}}$ is the matrix of binomial coefficients
- \underline{C} is the vector of Bézier coefficient, c_l , to be determined

$$\underline{\mathcal{B}} = \underline{T} \cdot \underline{\underline{M}} \cdot \underline{C}$$

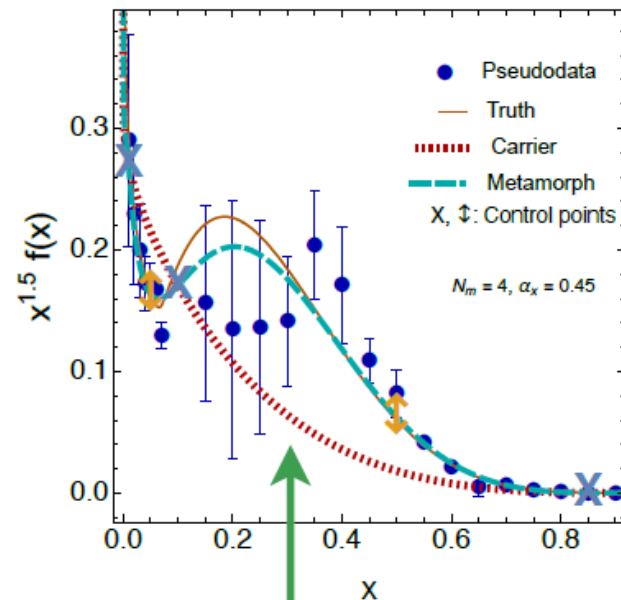
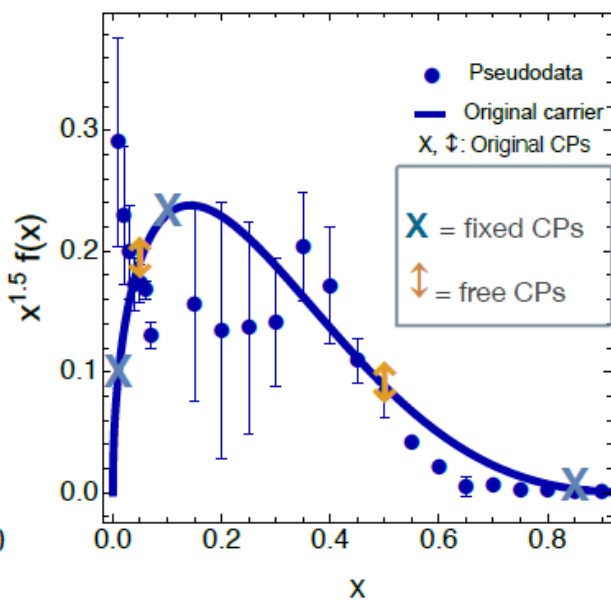
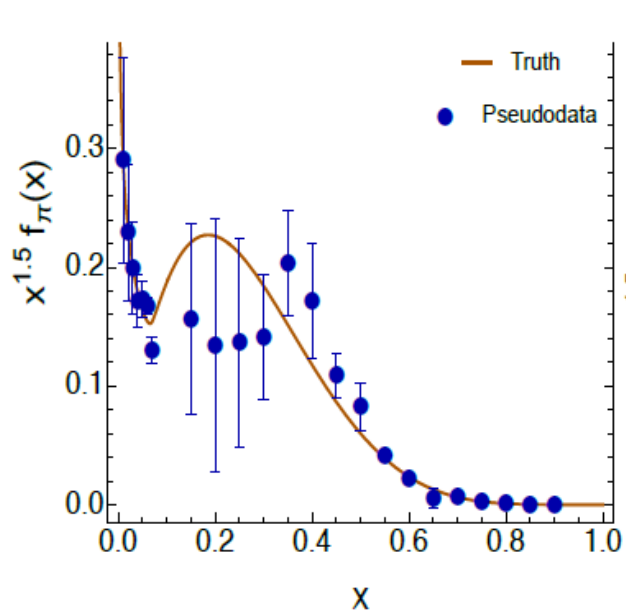
We can evaluate the Bézier curve at chosen **control points**, to get a vector of $\mathcal{B} \rightarrow \underline{P}$

$\underline{\underline{T}}$ is now a matrix of x^l expressed at the control points.

$$\underline{P} = \underline{\underline{T}} \cdot \underline{\underline{M}} \cdot \underline{C}$$

Slide by A. Courtoy

Bézier-curve methodology for global analyses — toy model



metamorph fit:

$$x q(x, Q_0^2) = A'_q x^{B_q} (1-x)^{C_q} \times \left(1 + \mathcal{B}^{(N_m)}(x^{\alpha_x}, Q_0^2; \underline{v}) \right)$$

with $N_m = \# \text{ CPs} - 1$ for a square-matrices system.

Shift of the control points ($\delta D_q, \dots$) replace free parameters

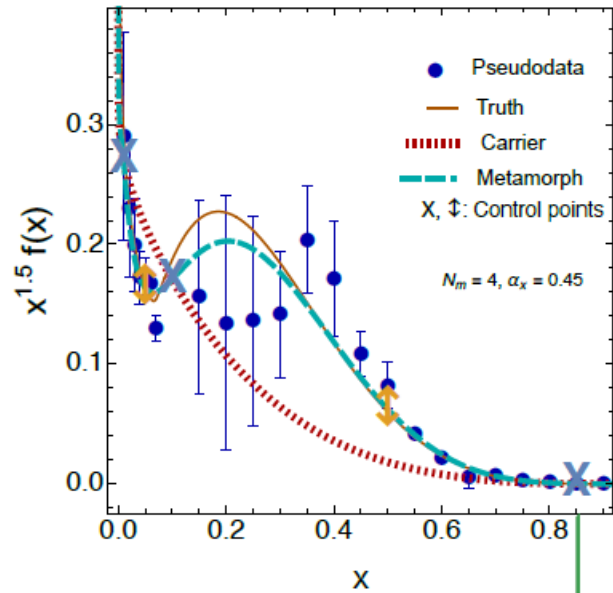
$N_m =$ degree of polynomial can vary

δB_q & δC_q allow the carrier to vary

α_x can vary

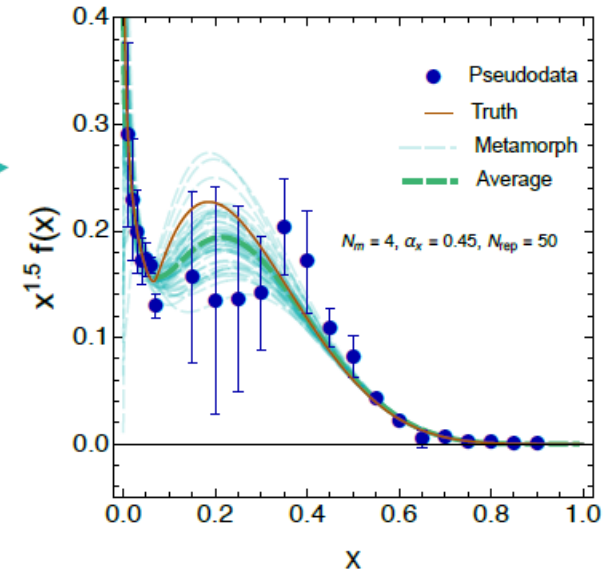
Slide by A. Courtoy

Bézier-curve methodology for global analyses — toy model

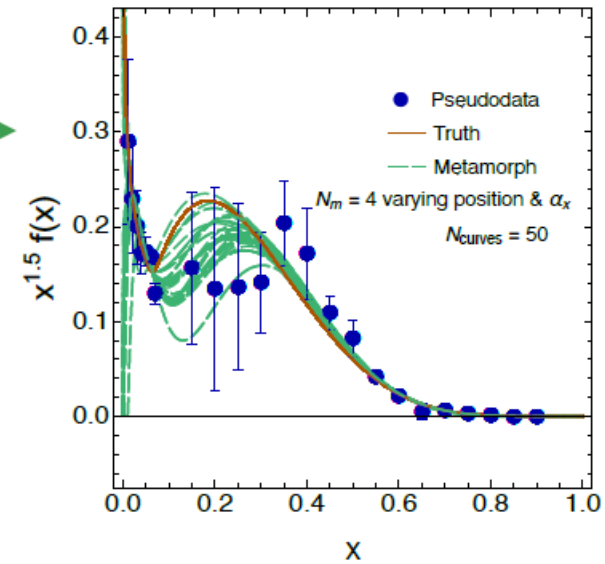


if bootstrapped

sampling on the distribution of data uncertainties



if sampled over metamorph settings
sampling over parametrizations



Both samplings can be done in the same analysis, they are not mutually exclusive.

Slide by A. Courtoy