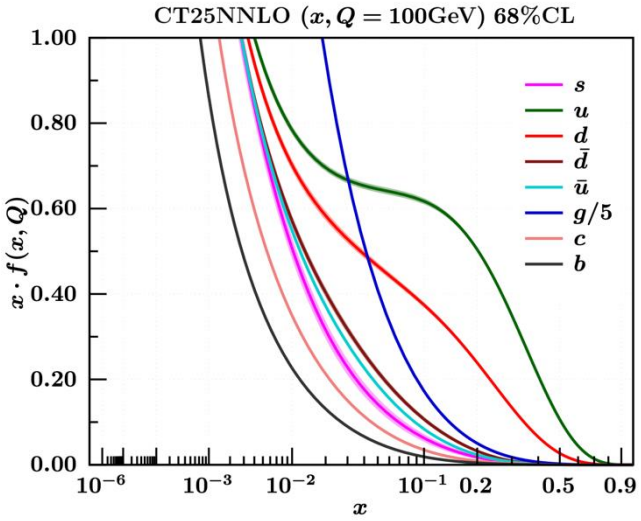
PDFs at the precision frontier: collider phenomenology & jet physics

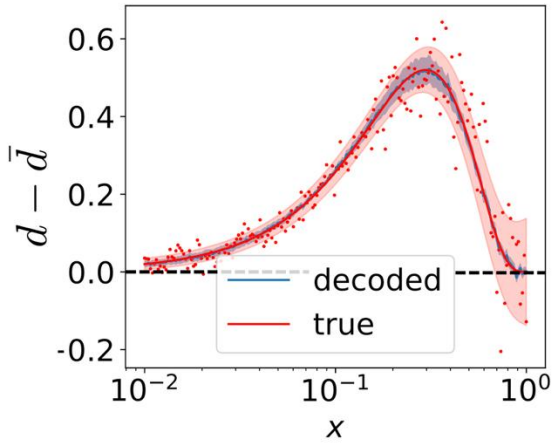
Timothy J Hobbs

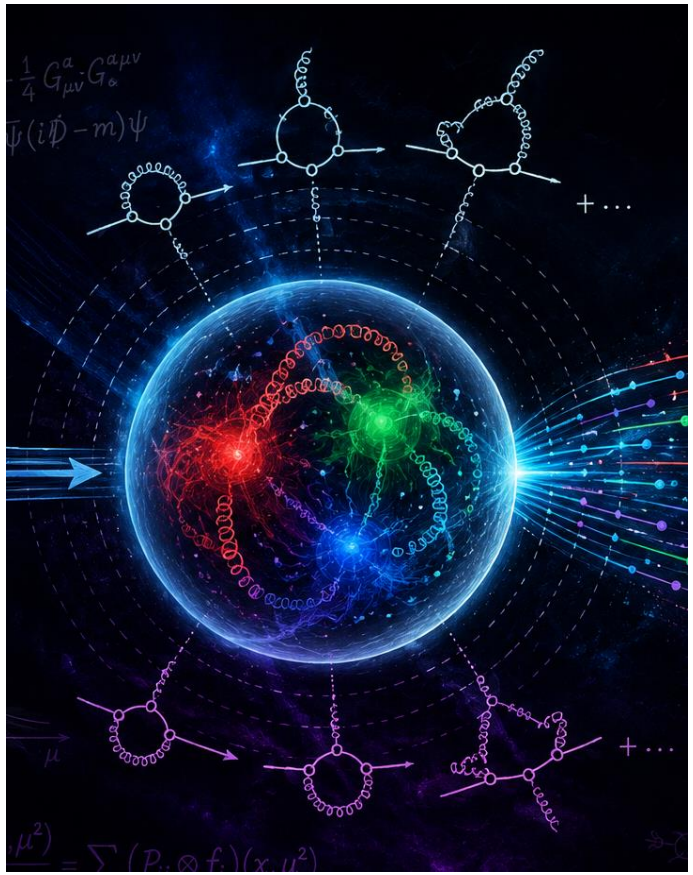
HEP Theory Group – Argonne National Laboratory, USA

LoopFest 2026
Brookhaven 28 May 2026

special thanks: CTEQ(-TEA) colleagues,
Max Ponce-Chavez, Brandon Kriesten, Supratim Das Bakshi







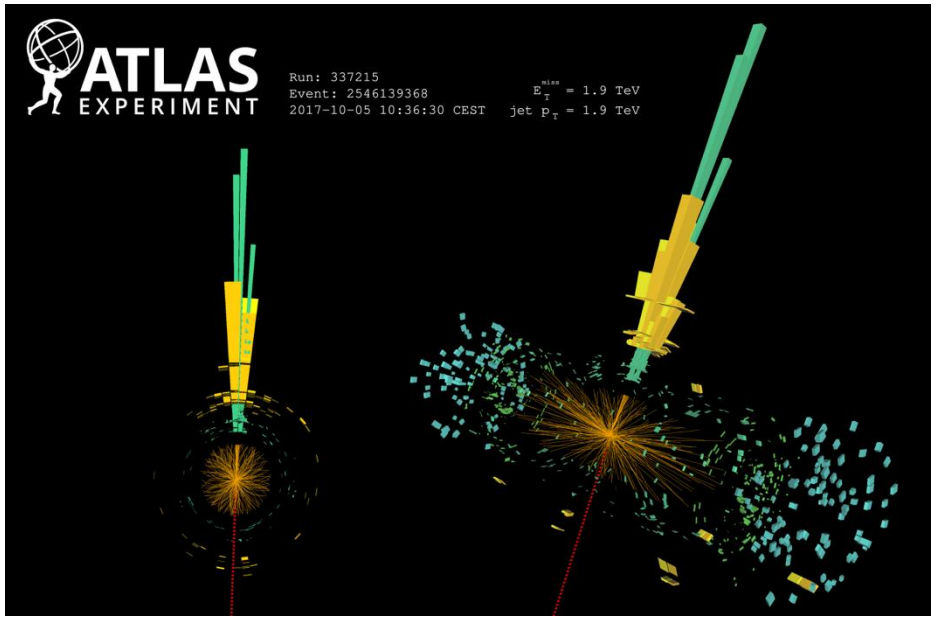
PDFs central to HEP pheno and perturbative theory efforts

- going toward HL-LHC, PDFs ‘not another systematic’ in jet or other pheno programs, but crux where perturbative theory meets Nature

- this (short) talk: several quick representative highlights

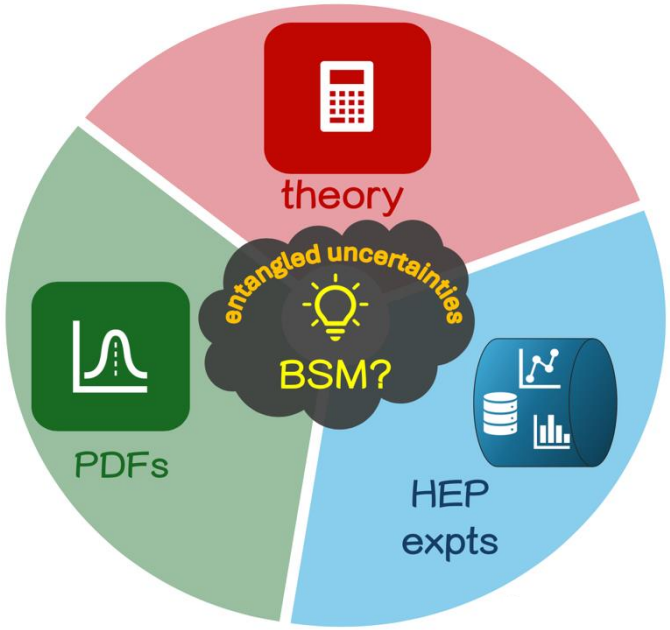
  → LHC context for PDFs

  → perturbative fundamentals; NNLO+ approximations



→ CT25 connections, especially to jet physics

→ AI for PDFs developments

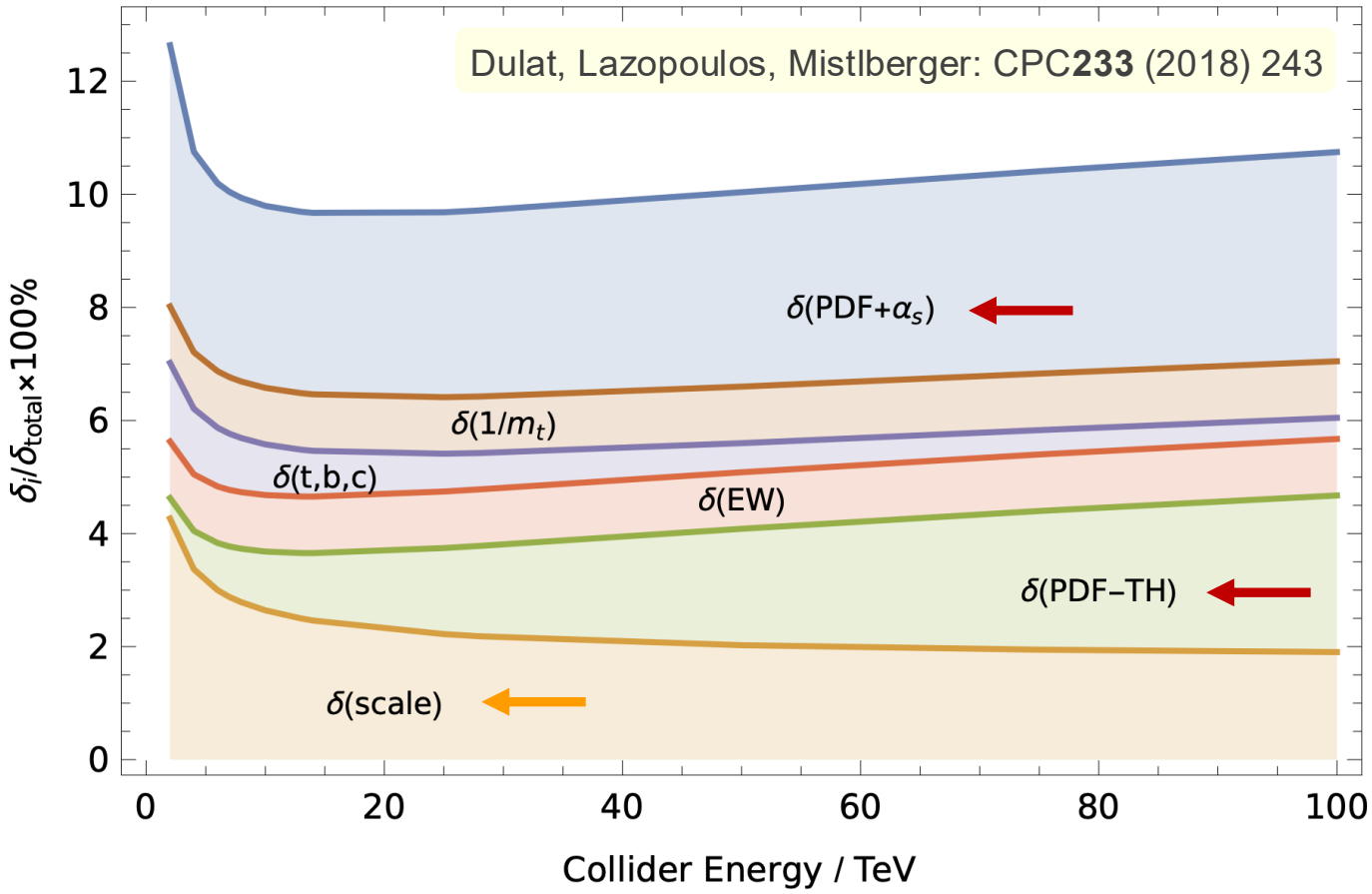


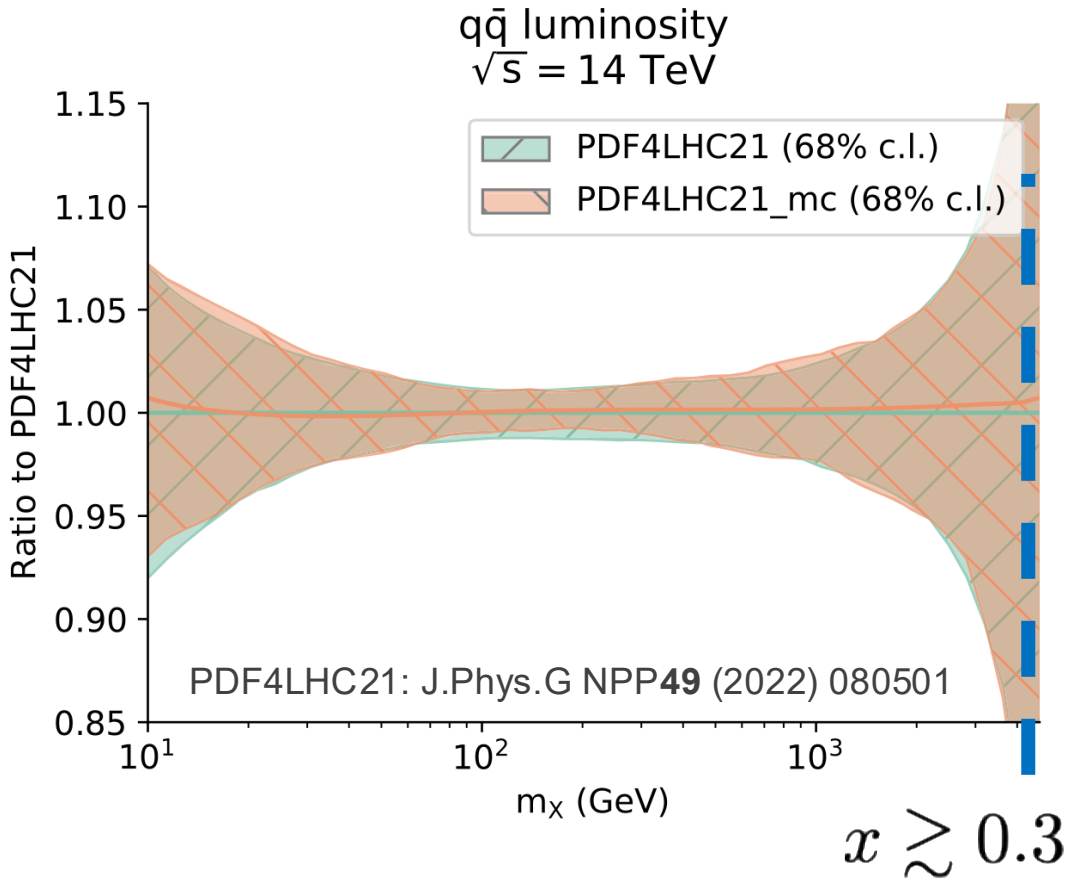
high perturbative accuracy; PDF precision; faithful uncertainties essential to theory interpretation of LHC, EIC, …, programs

PDFs as precision bottleneck at LHC

- Higgs; Z', W' production (as exemplar cases): uncertainties receive substantial PDF contribution

→ significant role from related effects (perturbative mismatch, scale, quark masses, …)





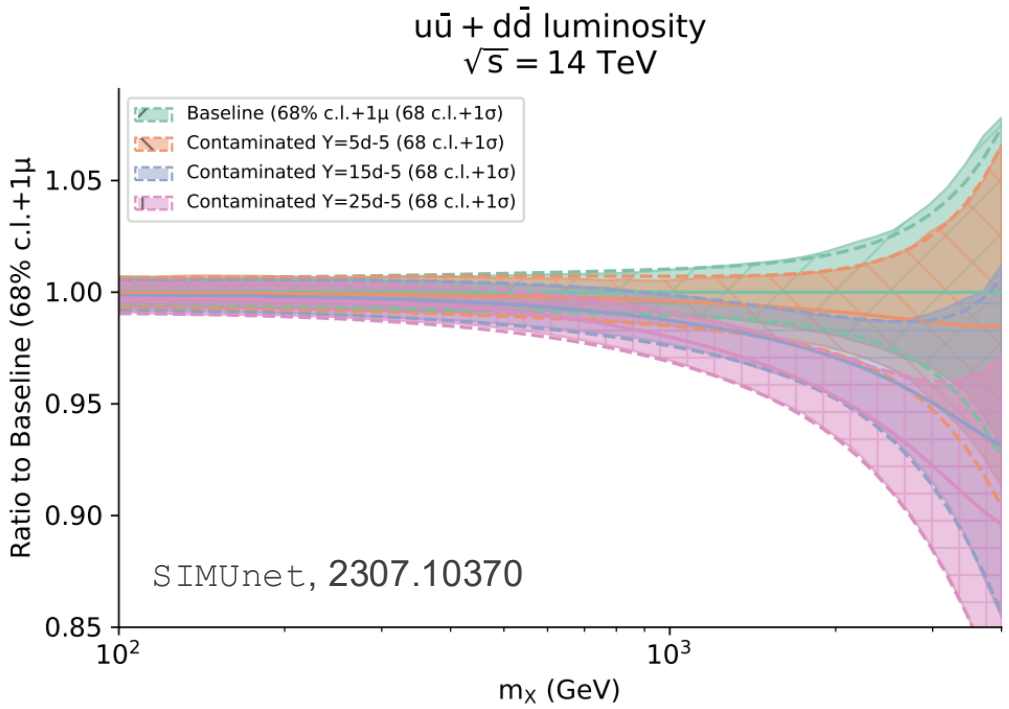
$x \gtrsim 0.3$

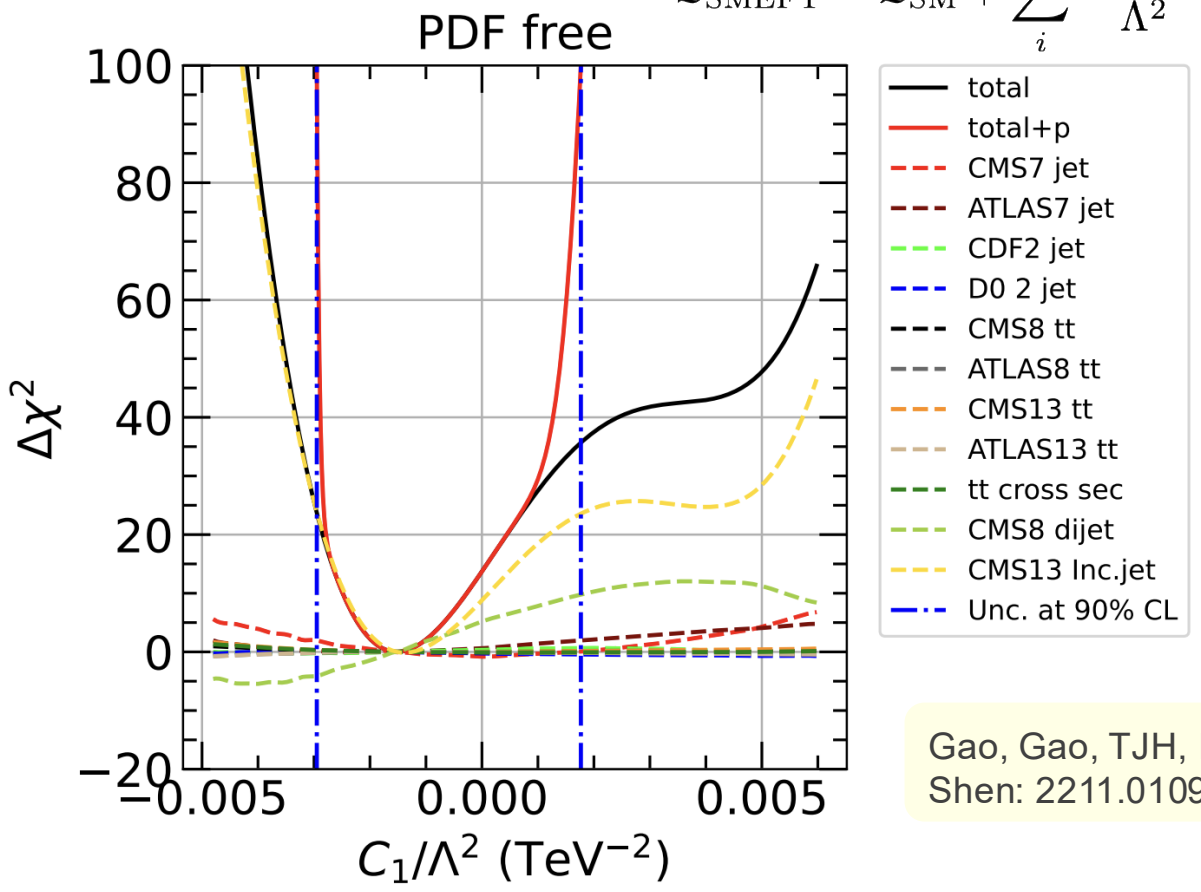
→ (high-x) PDF uncertainties limit searches for TeV-scale particles at the LHC

the PDF-BSM (SMEFT) entanglement problem

- BSM reach not only limited by PDF uncertainties: hadronic data used in New Physics searches also constrain PDF shape parameters!

$$\mathcal{L}_{\rm SMEFT} = \mathcal{L}_{\rm SM} + \sum_i \frac{C_i O_i^{(6)}}{\Lambda^2} + \cdots$$





Gao, Gao, TJH, Liu, Shen: 2211.01094

→ simultaneous analyses are only systematic approach to avoid spurious absorption of possible BSM signatures

→ joint fits alongside PDFs: **jet data** accommodate wider contact interaction ranges (C_1/Λ^2) relative to fixed-PDF extractions

[higher-loop order in multiple theories relevant…]

HEP PDF landscape has evolved steadily

- PDFs as precision engines: in subsequent decades, global fits have evolved from ~order-of-magnitude explorations to few percent-level frameworks

→ ecology of analysis approaches, (perturbative) theory choices; major progress but many **challenges remain**

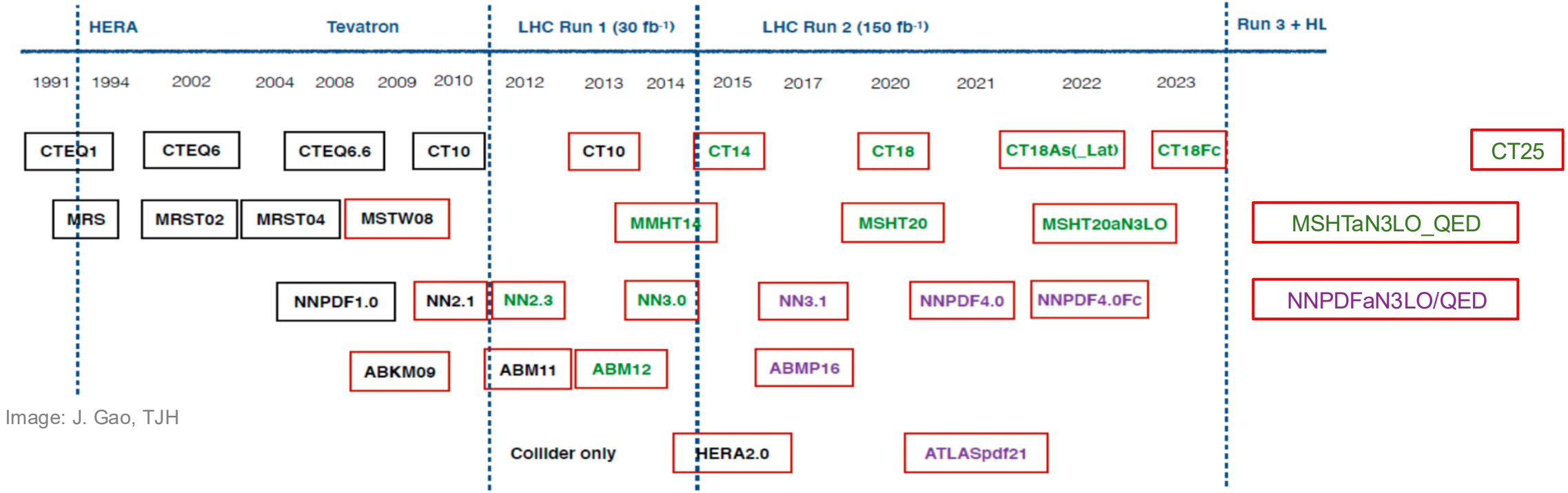


Image: J. Gao, TJH

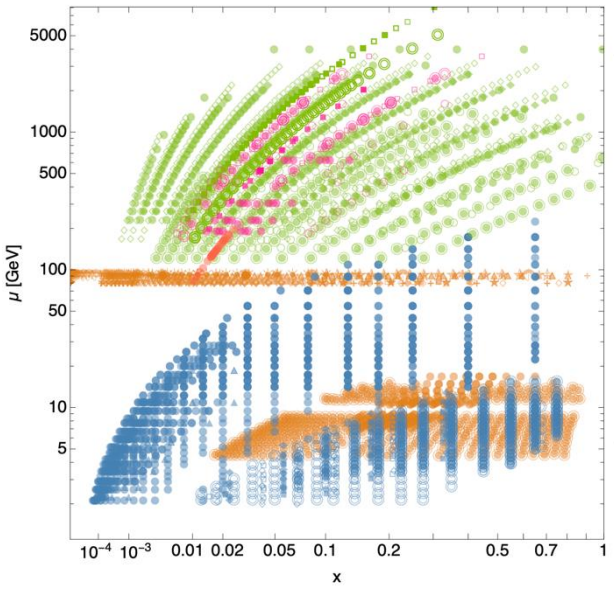
require consistent accuracy for many processes, partonic interactions

- Global fits stitch ~30 processes across every partonic channel and ~ten orders of magnitude in (x, Q^2)

	Process	Subprocess	Partons	x range
Fixed Target	$\ell^\pm\{p,n\} \to \ell^\pm + X$	$\gamma^* q \to q$	$q, \bar{q}, g$	$x \gtrsim 0.01$
	$\ell^\pm n/p \to \ell^\pm + X$	$\gamma^* d/u \to d/u$	d/u	$x \gtrsim 0.01$
	$pp \to \mu^+\mu^- + X$	$u\bar{u}, d\bar{d} \to \gamma^*$	$\bar{q}$	$0.015 \lesssim x \lesssim 0.35$
	$pn/pp \to \mu^+\mu^- + X$	$(u\bar{d})/(u\bar{u}) \to \gamma^*$	$\bar{d}/\bar{u}$	$0.015 \lesssim x \lesssim 0.35$
	$\nu(\bar{\nu})N \to \mu^-(\mu^+) + X$	$W^* q \to q'$	$q, \bar{q}$	$0.01 \lesssim x \lesssim 0.5$
	$\nu N \to \mu^-\mu^+ + X$	$W^* s \to c$	s	$0.01 \lesssim x \lesssim 0.2$
	$\bar{\nu} N \to \mu^+\mu^- + X$	$W^* \bar{s} \to \bar{c}$	$\bar{s}$	$0.01 \lesssim x \lesssim 0.2$
Collider DIS	$e^\pm p \to e^\pm + X$	$\gamma^* q \to q$	$g, q, \bar{q}$	$0.0001 \lesssim x \lesssim 0.1$
	$e^+ p \to \bar{\nu} + X$	$W^+\{d,s\} \to \{u,c\}$	d, s	$x \gtrsim 0.01$
	$e^\pm p \to e^\pm c\bar{c} + X$	$\gamma^* c \to c, \gamma^* g \to c\bar{c}$	c, g	$10^{-4} \lesssim x \lesssim 0.01$
	$e^\pm p \to e^\pm b\bar{b} + X$	$\gamma^* b \to b, \gamma^* g \to b\bar{b}$	b, g	$10^{-4} \lesssim x \lesssim 0.01$
	$e^\pm p \to \text{jet} + X$	$\gamma^* g \to q\bar{q}$	g	$0.01 \lesssim x \lesssim 0.1$
Tevatron	$p\bar{p} \to \text{jet} + X$	$gg, qg, qq \to 2j$	g, q	$0.01 \lesssim x \lesssim 0.5$
	$p\bar{p} \to (W^\pm \to \ell^\pm\nu) + X$	$ud \to W^+, \bar{u}\bar{d} \to W^-$	$u, d, \bar{u}, \bar{d}$	$x \gtrsim 0.05$
	$p\bar{p} \to (Z \to \ell^+\ell^-) + X$	$uu, dd \to Z$	u, d	$x \gtrsim 0.05$
	$p\bar{p} \to t\bar{t} + X$	$qq \to t\bar{t}$	q	$x \gtrsim 0.1$
LHC	$pp \to \text{jet} + X$	$gg, qg, q\bar{q} \to 2j$	g, q	$0.001 \lesssim x \lesssim 0.5$
	$pp \to (W^\pm \to \ell^\pm\nu) + X$	$u\bar{d} \to W^+, d\bar{u} \to W^-$	$u, d, \bar{u}, \bar{d}, g$	$x \gtrsim 10^{-3}$
	$pp \to (Z \to \ell^+\ell^-) + X$	$q\bar{q} \to Z$	$q, \bar{q}, g$	$x \gtrsim 10^{-3}$
	$pp \to (Z \to \ell^+\ell^-) + X, p_\perp$	$gq(\bar{q}) \to Zq(\bar{q})$	$g, q, \bar{q}$	$x \gtrsim 0.01$
	$pp \to (\gamma^* \to \ell^+\ell^-) + X$, Low mass	$q\bar{q} \to \gamma^*$	$q, \bar{q}, g$	$x \gtrsim 10^{-4}$
	$pp \to (\gamma^* \to \ell^+\ell^-) + X$, High mass	$q\bar{q} \to \gamma^*$	$\bar{q}$	$x \gtrsim 0.1$
	$pp \to W^+\bar{c}, W^-c$	$sg \to W^+c, \bar{s}g \to W^-\bar{c}$	$s, \bar{s}$	$x \sim 0.01$
	$pp \to t\bar{t} + X$	$gg \to t\bar{t}$	g	$x \gtrsim 0.01$
	$pp \to D, B + X$	$gg \to c\bar{c}, b\bar{b}$	g	$x \gtrsim 10^{-6}, 10^{-5}$
	$pp \to J/\psi, \Upsilon + pp$	$\gamma^*(gg) \to c\bar{c}, b\bar{b}$	g	$x \gtrsim 10^{-6}, 10^{-5}$
	$pp \to \gamma + X$	$gq(\bar{q}) \to \gamma q(\bar{q})$	g	$x \gtrsim 0.005$

Gao, Harland-Lang, Rojo: PR**742** (2018) 1-121

[CT25 data set]



consistency: matched accuracy across channels — **set by slowest $\hat{\sigma}$, not fastest**

perturbative QCD and PDFs

- QCD factorization couples perturbative accuracy to PDF dependence in hadronic cross sections

$$\frac{d^2\sigma^{\mathrm{DY}}}{dy\,dQ^2}\left(y, Q^2, \mu_R^2, \mu_F^2\right) = \sum_{a,b=q,\bar{q},g} \int_{\tau_1}^{1} dx_1\, f_a(x_1, \mu_F^2) \int_{\tau_2}^{1} dx_2\, f_b(x_2, \mu_F^2)\, \frac{d^2\hat{\sigma}_{ab}^{\mathrm{DY}}}{dy\,dQ^2}\left(x_1, x_2, y, Q^2, \mu_R^2, \mu_F^2\right)$$

→ perturbative accuracy enters multiple sectors of the calculation:

[partonic hard scatter]

$$\frac{d^2\hat{\sigma}_{ab}^{\mathrm{DY}}}{dy\,dQ^2}\left(x_1, x_2, y, Q^2, \mu_R^2, \mu_F^2\right) = \sum_{n=0}^{\infty} \left(\frac{\alpha_s(\mu_R^2)}{2\pi}\right)^n \frac{d^2\hat{\sigma}_{ab}^{(n)\mathrm{DY}}}{dy\,dQ^2}$$

[DGLAP splitting/kernels]

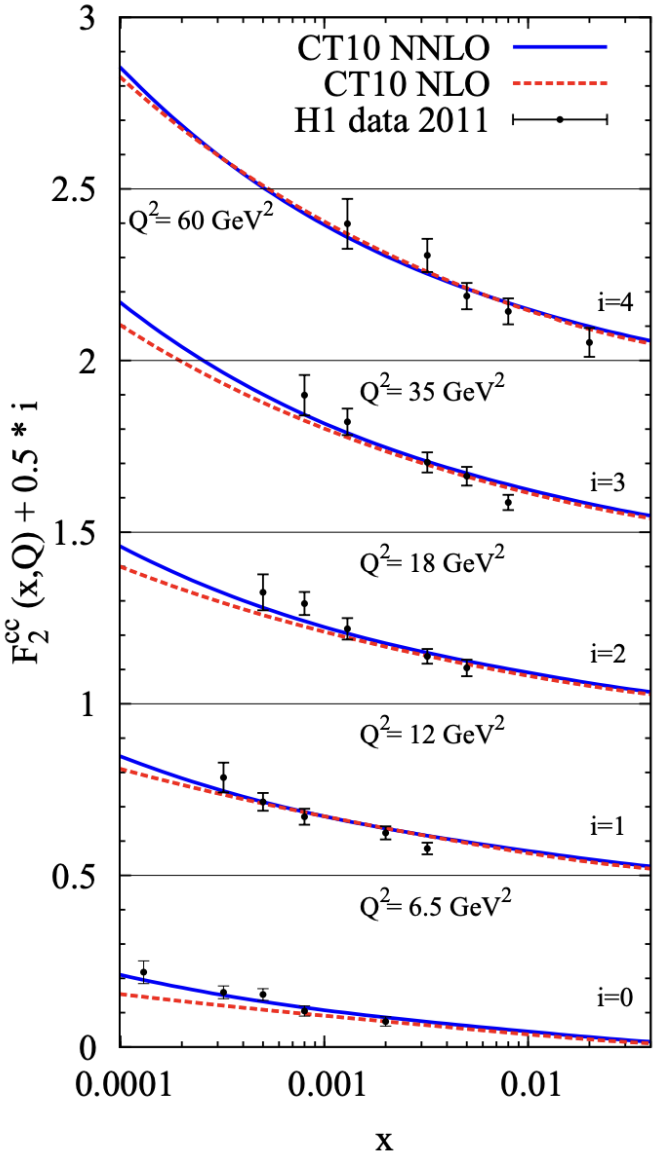
$$Q^2 \frac{\partial}{\partial Q^2} f_a(x, Q^2) = \sum_b P_{ab}\left(x, \alpha_s(Q^2)\right) \otimes f_b(x, Q^2)$$

$$P_{ab}\left(x, \alpha_s(Q^2)\right) = \sum_{n=0}^{\infty} \left(\frac{\alpha_s(Q^2)}{2\pi}\right)^{n+1} P_{ab}^{(n)}(x)$$

[α_s running]

$$\frac{d\ln\alpha_s(\mu_R^2)}{d\ln\mu_R^2} = -\sum_{n=0}^{\infty} \beta_n \left(\frac{\alpha_s(\mu_R^2)}{2\pi}\right)^{n+1}$$

the push toward higher perturbative accuracy



- NNLO corrections broadly extend descriptive reach of global fit
  - → (left) low-x rise of charm-tagged DIS cross section better modeled
  - → stronger control of scale uncertainties (see backup – GM VFNS)
  - → stabilize PDF extractions over wide Q^2 [shifts $\gtrsim$ PDF uncertainties]

- argues for extending to N3LO — BUT full N3LO ingredients not (yet) in hand

- one scenario: *approximate* N3LO (aN3LO)

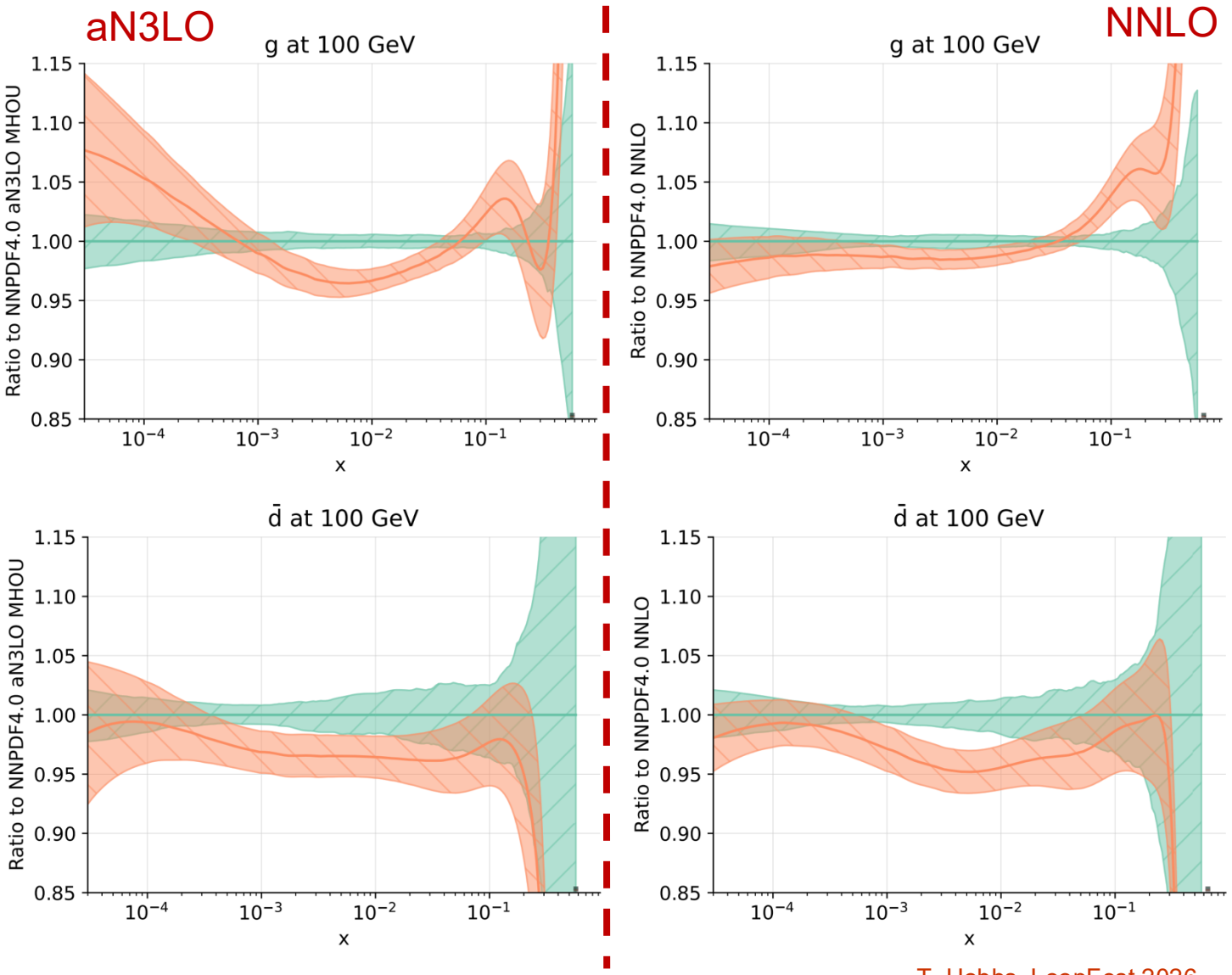
$$P_{ij}^{(3)}(x) = P_{ij}^{(3),\,\text{known}}(x) + \rho_{ij}\, P_{ij}^{(3),\,\text{approx}}(x)$$

$$\chi^2 = (D-T)^T \left(C_{\text{exp}} + C_{\text{th}}(\rho)\right)^{-1} (D-T)$$

  - → include known N3LO pieces; attempt UQ via nuisance parameters

recent practical aN3LO fit implementations

- growing set of N3LO α_s calculations increasingly available; recent efforts to incorporate into PDF fits



→ leads to analyses of mixed perturbative accuracy ('partial' or 'approximate' N3LO)

→ NNLO to aN3LO produces modest PDF-level shifts (more pronounced for gluon)

→ need to continue benchmarking and actively revisit as more N3LO theory ingredients emerge

[see talk, **M. Guzzi** for aN3LO details!]

[Cridge, Ball et al., 2411.05373]

benchmark estimator of missing N3LO corrections

- define generalizable prescription ('NNLO+') order-by-order for next-highest cross section; *e.g.*, for $pp \to B + X$, $\sigma_B^{N3LO} = \sigma_B^{NNLO} + \Delta^{N3LO}\sigma_B$, where

$$\sigma^{(i)} = \sum_{j=0}^{i} \hat{\sigma}^{(j)}$$

$$\Delta^{\mathrm{N3LO}}\sigma_B \equiv \sigma_B^{\mathrm{N3LO}} - \sigma_B^{\mathrm{NNLO}}$$

$$= \hat{\sigma}^{(3)} \otimes f^{\mathrm{N3LO}} + \left(\hat{\sigma}^{(0)} + \hat{\sigma}^{(1)} + \hat{\sigma}^{(2)}\right) \otimes \left[f^{\mathrm{N3LO}} - f^{\mathrm{NNLO}}\right]$$

$$f^{\mathrm{N3LO}} \overset{1\sigma}{\approx} f^{\mathrm{NNLO}} \implies \left(\sum_{i=0}^{2} \hat{\sigma}^{(i)}\right) \otimes \left[f^{\mathrm{N3LO}} - f^{\mathrm{NNLO}}\right] \approx 0$$

$$\Delta^{\mathrm{N3LO}}\sigma_B \approx \hat{\sigma}^{(3)} \otimes f^{\mathrm{NNLO}}$$

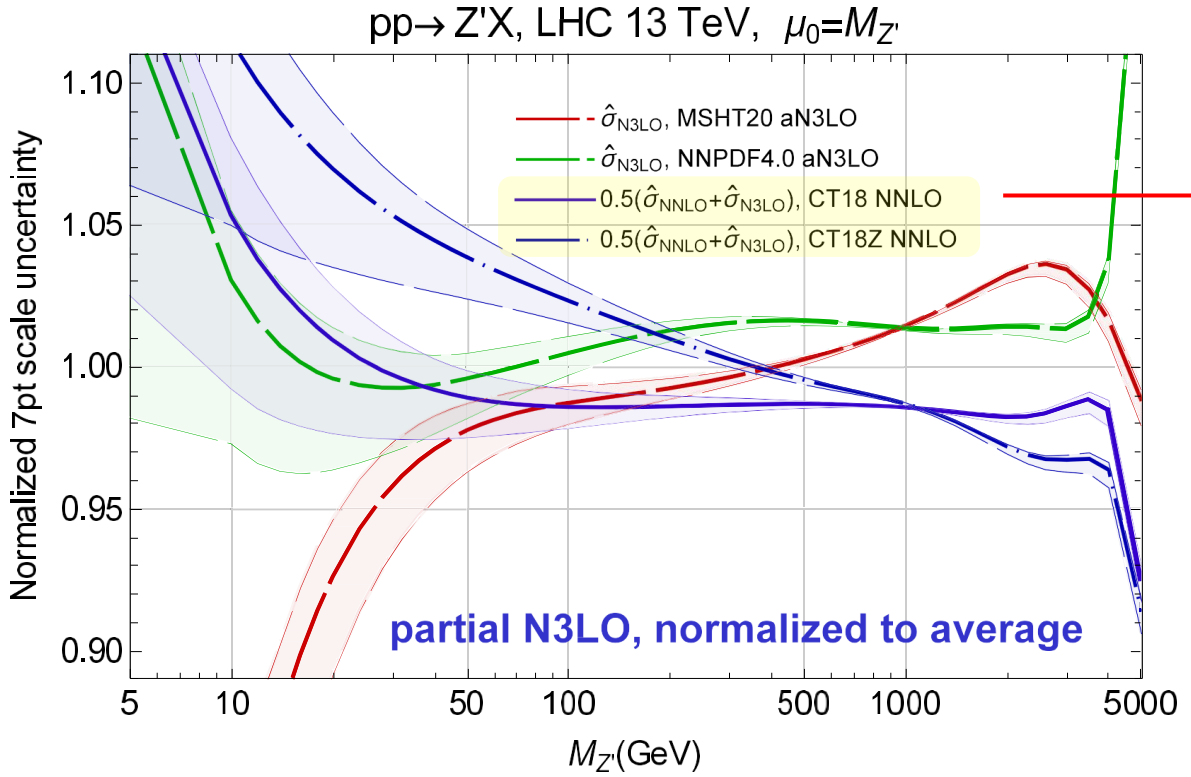
- prescription is then **epistemic estimator** on $\hat{\sigma}^{(3)} \in [0, \hat{\sigma}^{(3)}_{true}]$, where we take $\hat{\sigma}^{(3)} = \left(\frac{1}{2}\right) \cdot \hat{\sigma}^{(3)}_{true}$

→ implies simple benchmark approximator, testable when N3LO is known: $\sigma_B^{\mathrm{N3LO}} \approx \frac{1}{2}\left[\sigma^{(3)} + \sigma^{(2)}\right] \otimes f^{\mathrm{NNLO}}$

total cross sections: testbed for pN3LO implementations

[TJH, Nadolsky, Ponce-Chavez: appearing soon.]

❑ benchmark total cross sections [`n3loxs`]: unravel partial N3LO effect on precision, accuracy



→ test against pN3LO (NNLO+) prescription

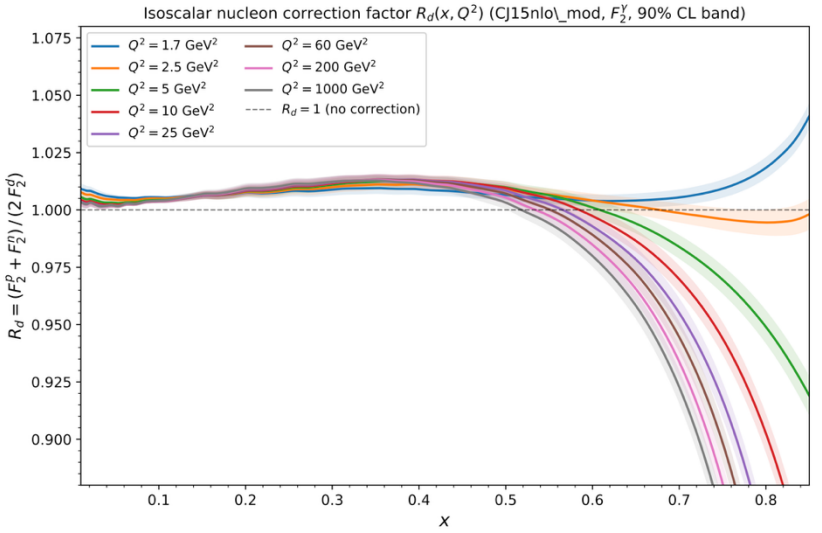
❑ mismatches among fits grow at small and large $M_{Z'}$: even as central PDFs are refitted at different orders

➢ other PDF uncertainties > (N3LO-NNLO) differences: *e.g.*, PDF priors, modeling of systematics, …

→ these simply do not automatically decrease at NNLO+

counter-example: deuteron corrections in NNLO PDF fits

- many nonperturbative corrections will not independently stabilize at N3LO
  - *d*-quark PDF substantially constrained by deuterium data (e.g., BCDMS; involves nuclear correction)



- fixed deuteron corrections can rival NNLO correction w.r.t. χ^2 shifts! [below]

Experiment	N_{pt}	no d.c. NLO	no d.c. NNLO	no d.c. NNLO-X	fixed d.c. NLO	δ^{NNLO}	$\delta^{\text{NNLO}-\text{X}}$	$\delta^{\text{fixed d.c.}}$
γ+jet	0	–	–	–	–	–	–	–
Jets Tevatron	182	225	236	233	225	11	8	0
DIS proton	1523	1818	1865	1839	1812	47	21	-6
DIS deuteron	373	401	406	405	387	5	4	-14
WZ Tevatron	101	113	120	110	129	7	-3	16
WZ LHC	185	267	259	185	347	-8	-82	80
Drell-Yan	318	454	364	432	348	-90	-22	-106
TOTAL	**3681**	**4289**	**4277**	**4208**	**4271**	**-12**	**-81**	**-18**

[see talk, **F. Olness** on nPDFs!]

Accardi, Jing, TJH, Nadolsky, 2102.01107

N3LO PDF analyses: essential pieces (recommendations)

❑ must include all PDF analysis theory ingredients **exactly and fully** to guarantee true N3LO accuracy

Theory Ingredients		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
	• $pp \to W + c$, $pp \to Z + b$, pp $\to b$	NLO (massive); NNLO (ZM)

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

full theory accuracy requires parallel exploration of these additional effects

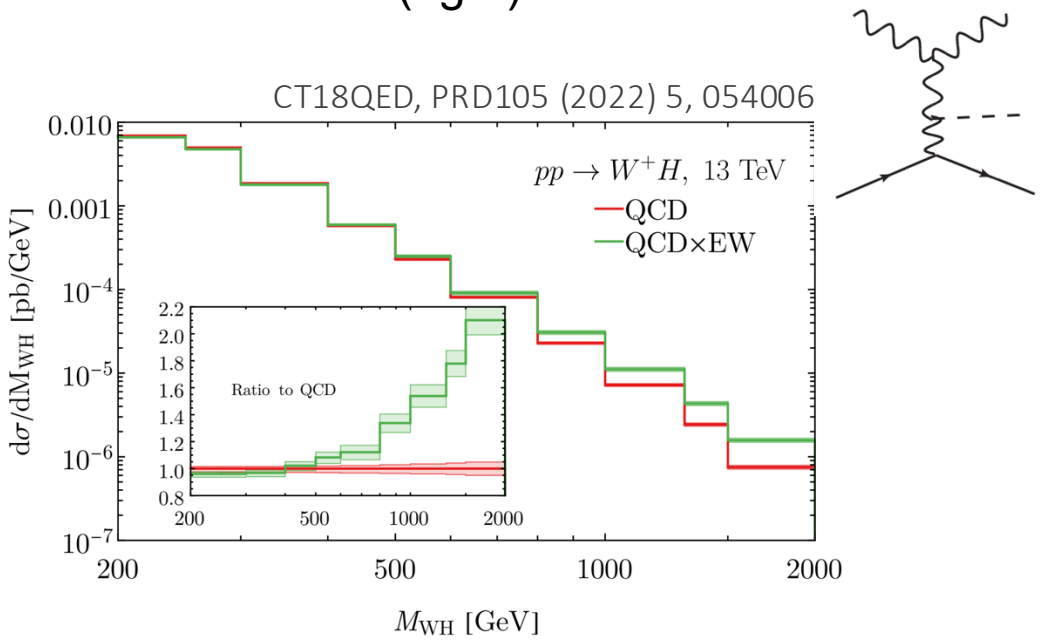
→ CTEQ-TEA, other groups already include some N3LO contributions in fitting codes; ongoing progress alongside MSHT, NNPDF on partial N3LO fits

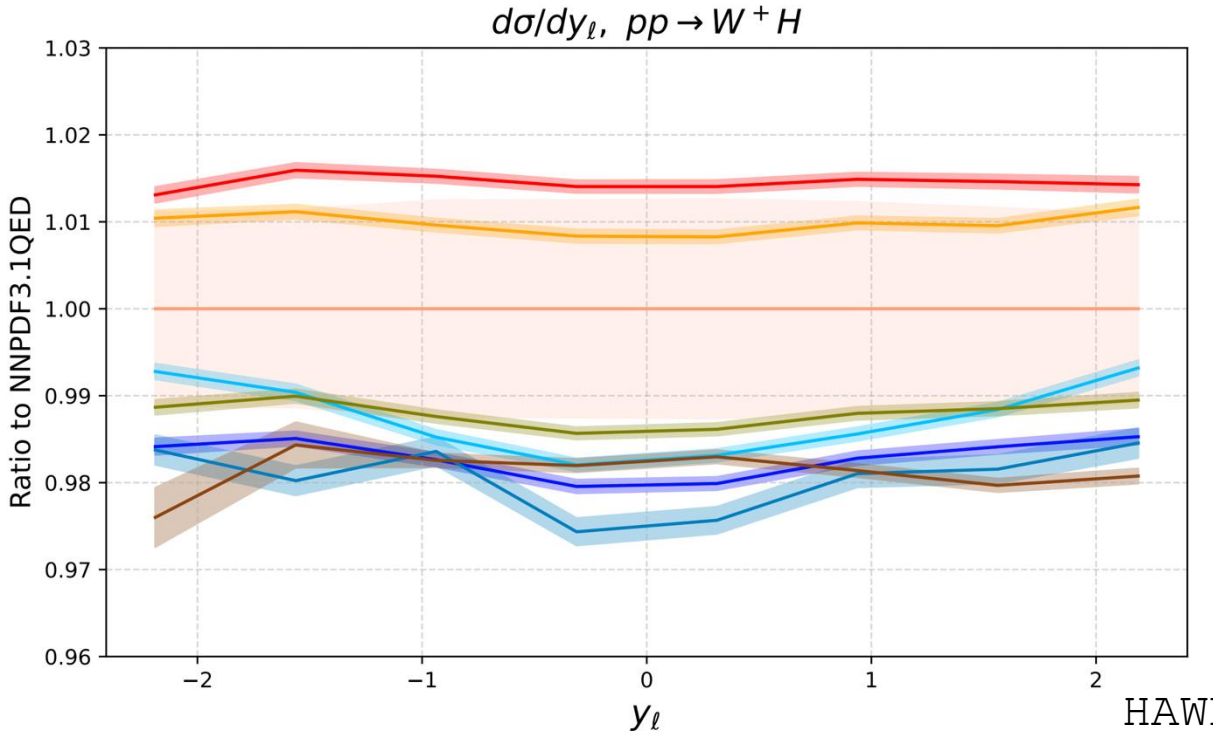
consistent theory accuracy: QCD → electroweak sector effects (photon PDF)

Cooper-Sarkar, Cridge, TJH et al: JHEP05 (2026) 228

❑ NNLO+ QCD accuracy implies need for complementary improvements in QED/EW corrections

➢ photon PDF necessary for precision pheno effects: *e.g.*, Higgs-strahlung rapidity distributions (right)



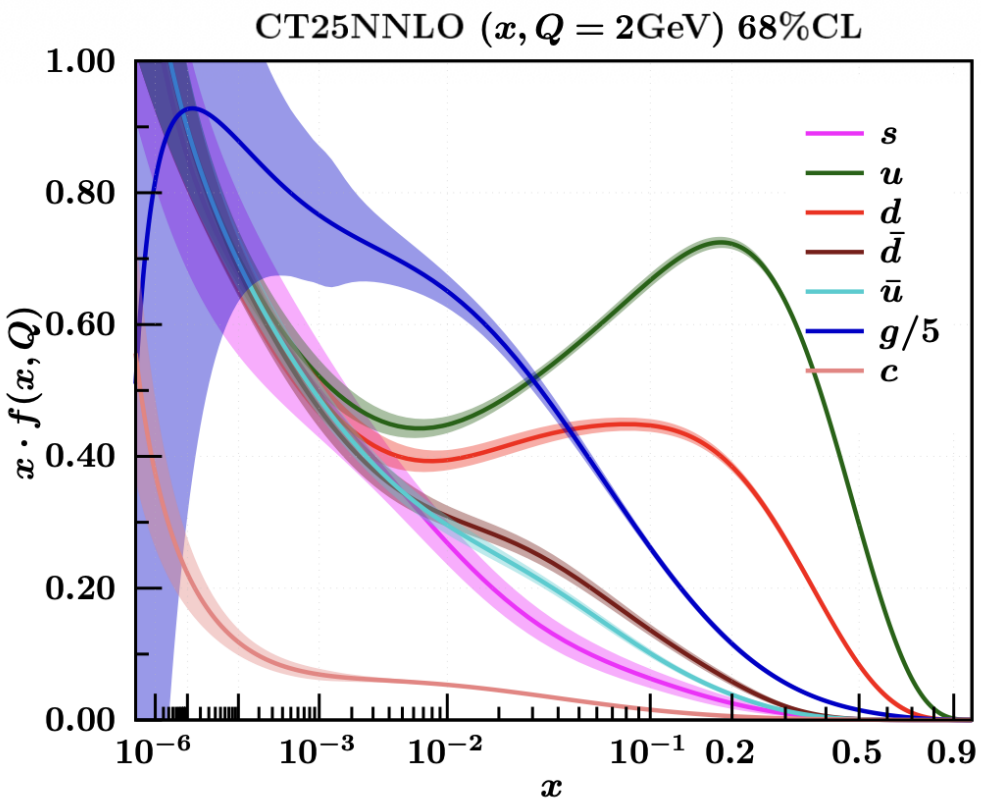


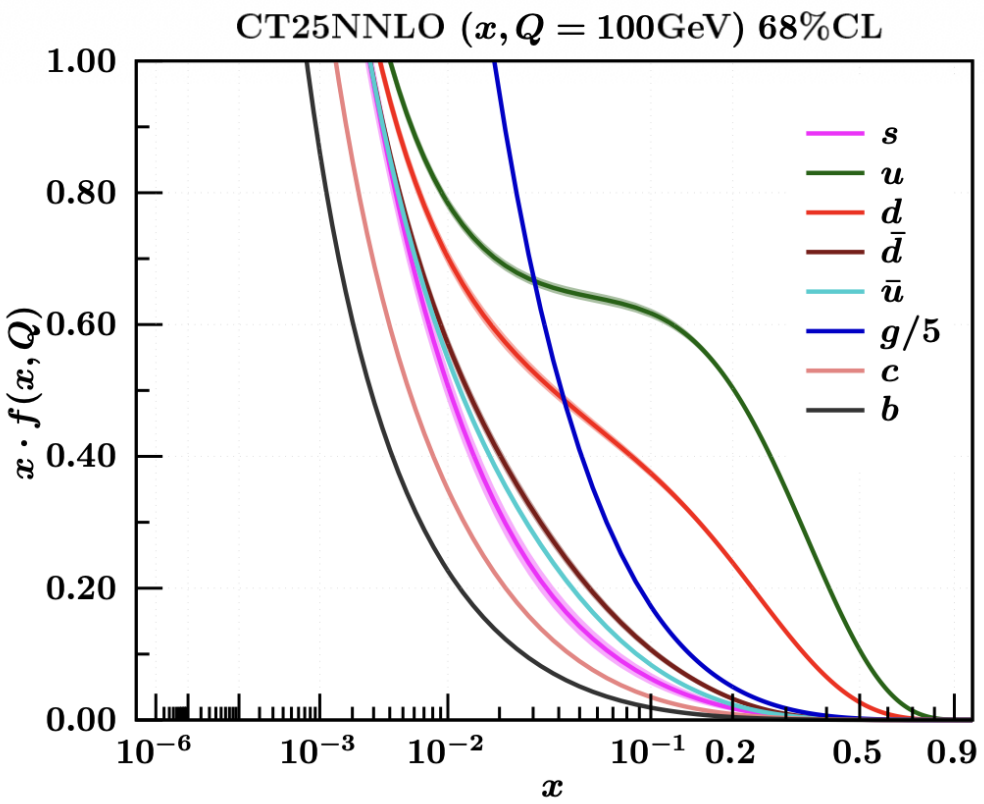
→ modern QED PDFs produce ~2-4% in VH distributions

(post LUXQED, QED PDF dependence of photon-initiated corrections has substantially stabilized [permille-level])

newest CTEQ-TEA global fit: **CT25** analysis

- CT25 global analysis PDF grids now public, 2512.19779 (includes corresponding series in α_s and n_F)
- forthcoming 'long paper': systematic exploration of fit variations; phenomenological implications





→ includes adoption of updated CT dynamical tolerance prescription for PDF uncertainties

→ analysis variants: CT25 (baseline); CT25FlatP (proton-only prior); CT25m (parametrization ensemble); CT25pd (deuteron-corrected data)

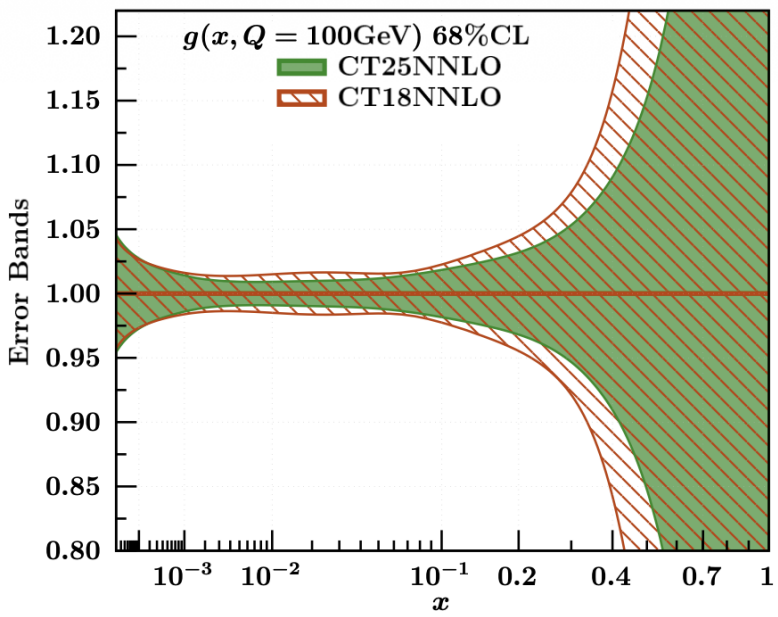
new CT25 data: LHC sets at 5, 7, 8, 13 TeV

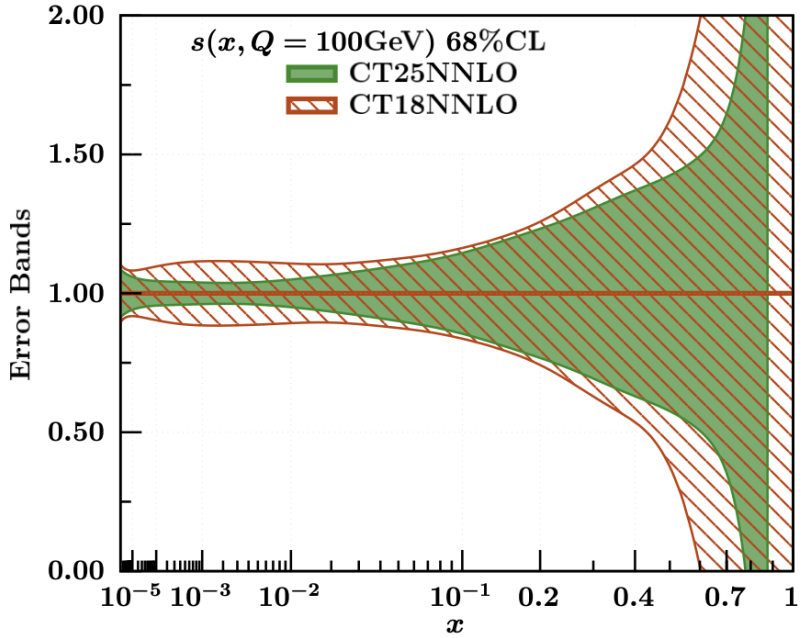
(new data span several LHC processes)

- substantial contingent [$N_{pt} > 700$] of (inclusive) jet-production data, ATLAS/CMS!
- (jet) and other new LHC data generally well-fitted at NNLO

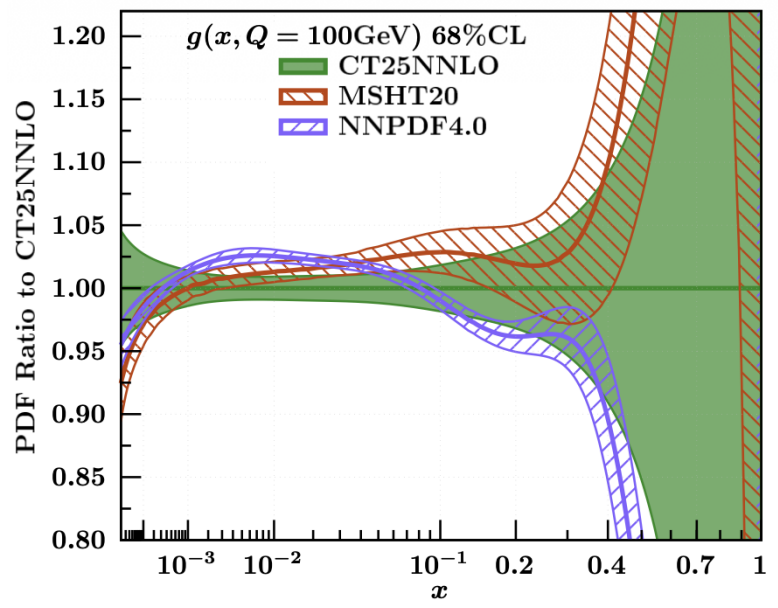
ID	Experimental data set		$N_{pt,E}$	χ^2_E	$\chi^2_E/N_{pt,E}$	S_E
215	ATLAS 5.02 TeV W/Z: 25 pb^{-1}, for W: $p_T^{\ell,\nu_l} > 25$ GeV, $\lvert\eta_\ell\rvert < 2.5$, $m_T > 40$ GeV;	[11]	27	16.8	0.6	-1.5
211	ATLAS 8 TeV W: 20.2 fb^{-1}, $p_T^\mu > 25$ GeV, $p_T^\nu > 25$ GeV, $\lvert\eta_\mu\rvert < 2.5$, $m_T > 40$ GeV	[12]	22	44.6	2.0	2.8
214	ATLAS 8 TeV Z 3D: 20.2 fb^{-1}, $p_T^\ell > 20$ GeV, $\lvert\eta_\ell\rvert < 2.4$, $46 < m_{\ell\ell} < 200$ GeV	[13]	188	217.7	1.2	1.5
212	CMS 13 TeV Z: 35.9 fb^{-1}, $\frac{d\sigma(Z)}{dy_Z}$, $p_T^\ell > 20$ GeV, $\lvert\eta_\ell\rvert < 2.4$, $\lvert m_{\ell\ell} - M_Z\rvert < 15$ GeV	[14]	12	24.2	2.0	2.1
217	LHCb 8 TeV W: 2.0 fb^{-1}, $W \to e\nu_e$, $p_T^e > 20$ GeV, $2.00 < \eta < 4.25$	[15]	14	19.4	1.4	1.0
218	LHCb 13 TeV Z: 0.29 fb^{-1} $p_T^\ell > 20$ GeV, $2.0 < \eta < 4.5$, $60 < m_{\ell\ell} < 120$ GeV	[16]	16	17.5	1.1	0.4
521	ATLAS 13 TeV all-hadronic $y_{t\bar{t}}$: 36.1 fb^{-1}, $t\bar{t}$ abs. cross sec. all-hadronic channel	[19]	12	13.3	1.1	0.4
587	ATLAS 13 TeV $\ell + j$ $y_{t\bar{t}} + m_{t\bar{t}} + y^B_{t\bar{t}} + H_T^{t\bar{t}}$: 36 fb^{-1}, $t\bar{t}$ abs. cross sections $\ell + j$ channel	[17]	34	36.2	1.1	0.3
528	CMS 13 TeV $\ell\ell$ $y_{t\bar{t}}$: 35.9 fb^{-1}, $t\bar{t}$ abs. $y_{t\bar{t}}$ cross sections dilepton channel	[20]	10	12.8	1.3	0.7
581	CMS 13 TeV $\ell + j$ $m_{t\bar{t}}$: 137 fb^{-1}, $t\bar{t}$ abs. cross sections $\ell + j$ channel	[21]	15	17.0	1.1	0.5
553	ATLAS 8 incl.jet: 20.3 fb^{-1}, $\frac{d^2\sigma}{dp_T d\lvert y\rvert}$, $\lvert y\rvert < 3.0$, $p_T^{jet} \in [70 : 2500]$, $R = 0.6$	[22]	171	273.4	1.6	4.7
554	ATLAS 13 incl.jet: 3.2 fb^{-1}, $\frac{d^2\sigma}{dp_T d\lvert y\rvert}$, $\lvert y\rvert < 3.0$, $p_T^{jet} \in [100 : 3937]$, $R = 0.4$	[40]	177	232.5	1.3	2.7
555	CMS 13 incl.jet: 36.5 fb^{-1}, $\frac{d^2\sigma}{dp_T d\lvert y\rvert}$, $0 < \lvert y\rvert < 2.0$, $97 \le p_T^{jet} \le 3103$ GeV, $R = 0.7$	[24]	78	88.6	1.1	0.9
556	CMS 7 TeV incl.jet: 5.0 fb^{-1}, $\frac{d^2\sigma}{dp_T d\lvert y\rvert}$, $0 < \lvert y\rvert < 2.5$, $114 \le p_T \le 1327$ GeV, $R = 0.7$	[24]	118	87.3	0.7	-2.1
557	CMS 8 TeV incl.jet: 19.7 fb^{-1}, $\frac{d^2\sigma}{dp_T d\lvert y\rvert}$, $0 < \lvert y\rvert < 3.0$, $74 \le p_T \le 1784$ GeV, $R = 0.7$	[24]	164	192.8	1.2	1.5

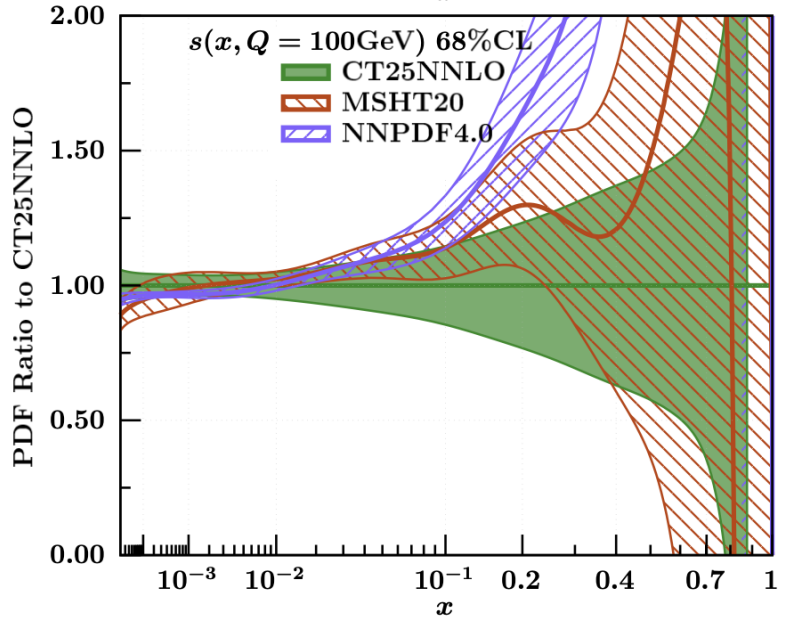
preview of CT25





- significant implications for gluon and strange PDFs; *e.g.*, $g(x)$ at high $x \gtrsim 0.1$ and $s(x)$ for $x \lesssim 0.01$

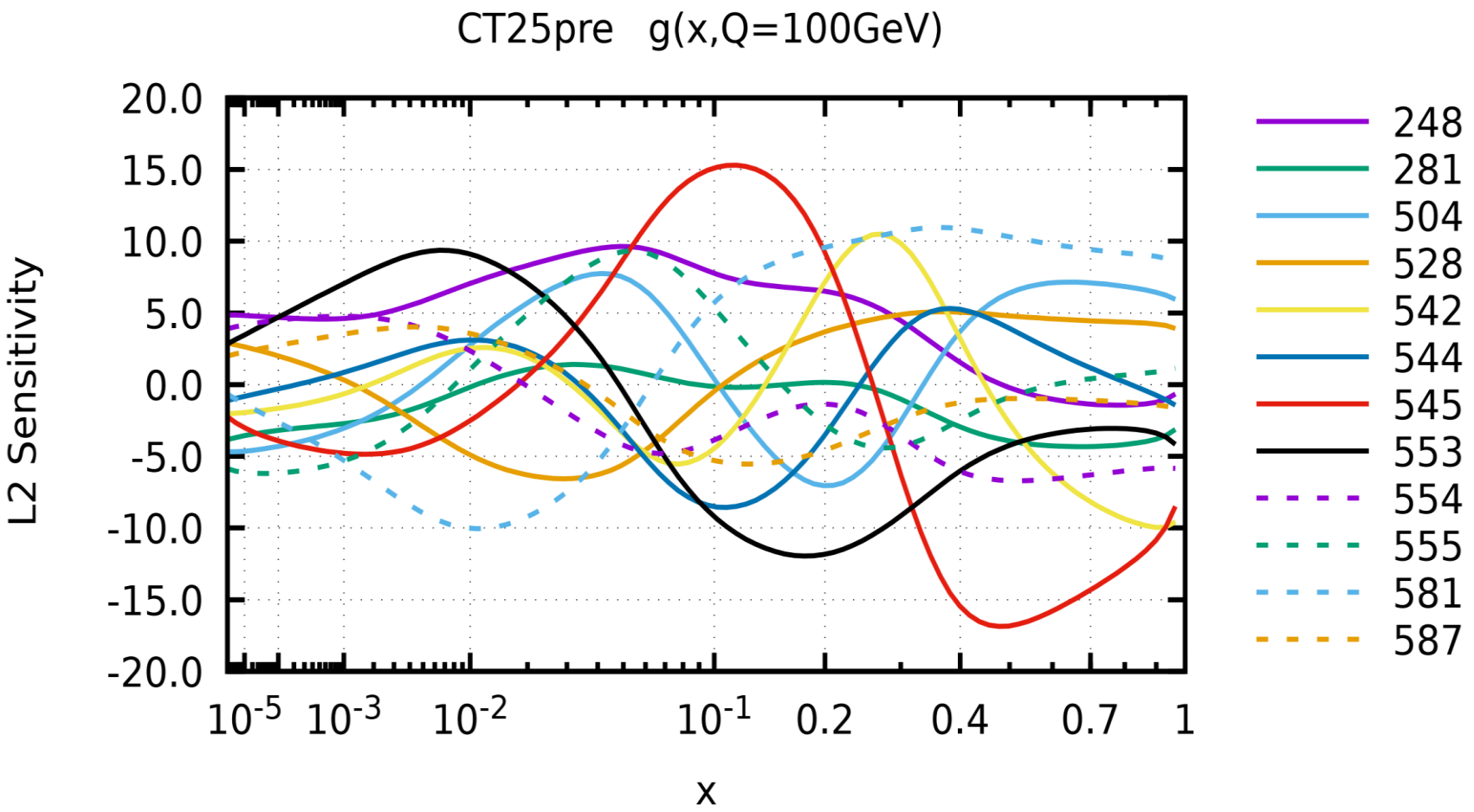




- general consistency among predictions, especially CT25 and MSHT
- shifts propagate to parton luminosities, *e.g.*, L_{gg} at large masses

L_2 sensitivities: significant PDF pulls for newly-fitted LHC sets

- *e.g.*, in Higgs-sensitive region ($x \sim 10^{-2}$), upward pulls from CMS 13 TeV $m_{t\bar{t}}$ (581) and $y_{t\bar{t}}$ (528), but with a downward pull from ATLAS 8 TeV inclusive jet data (553)



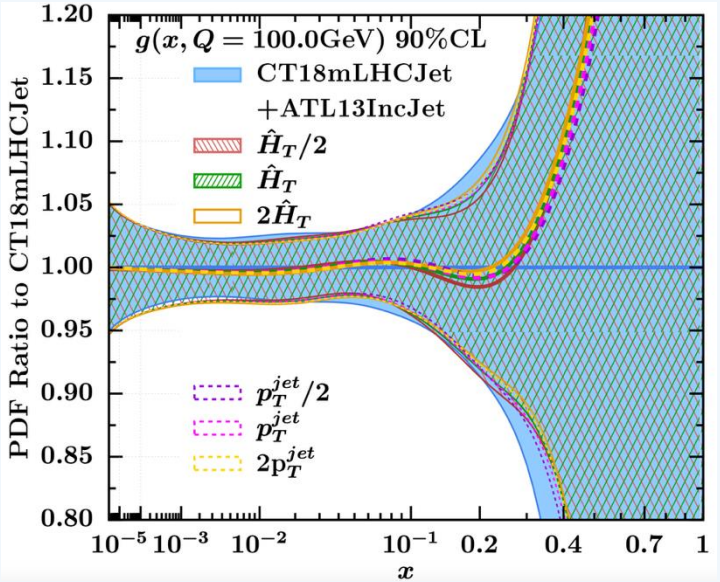
- inclusive jet data (*e.g.*, CMS 8 TeV, 545) still have notable pulls at high x (> 0.1) as observed in CT18

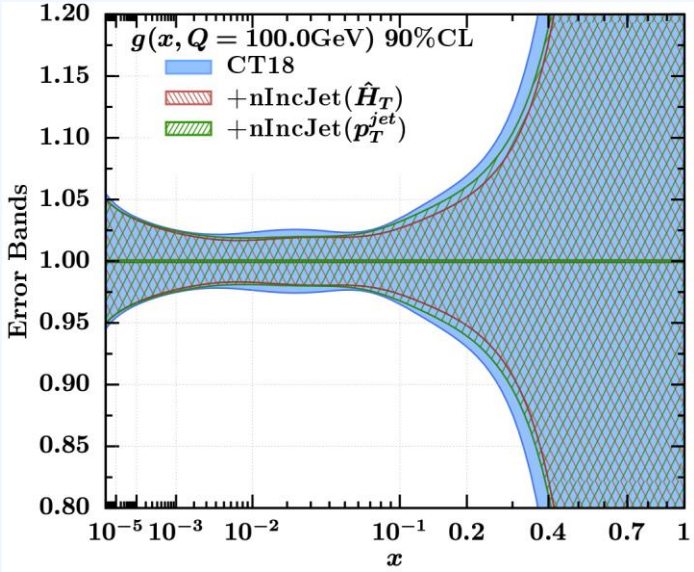
→ L_2 sensitivity method shown here originally introduced for PDF-Lattice synergy studies

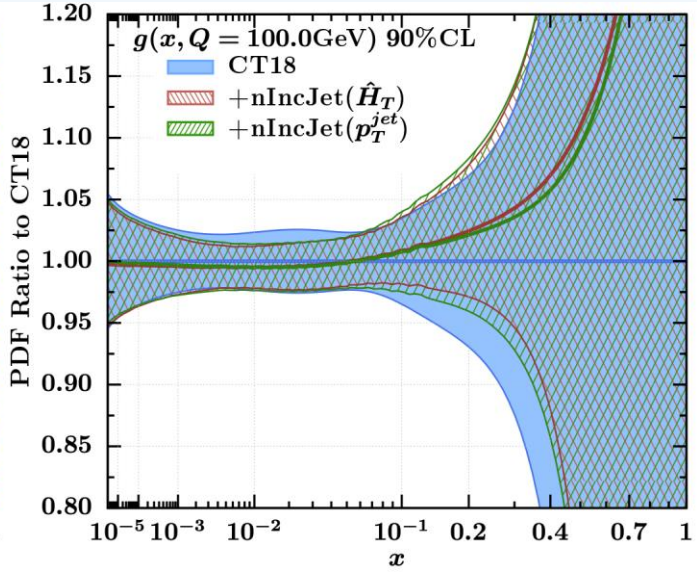
[TJH, Wang, Nadolsky, Olness: 1904.00022]

inclusive jet vs. dijet data sets: scale dependence

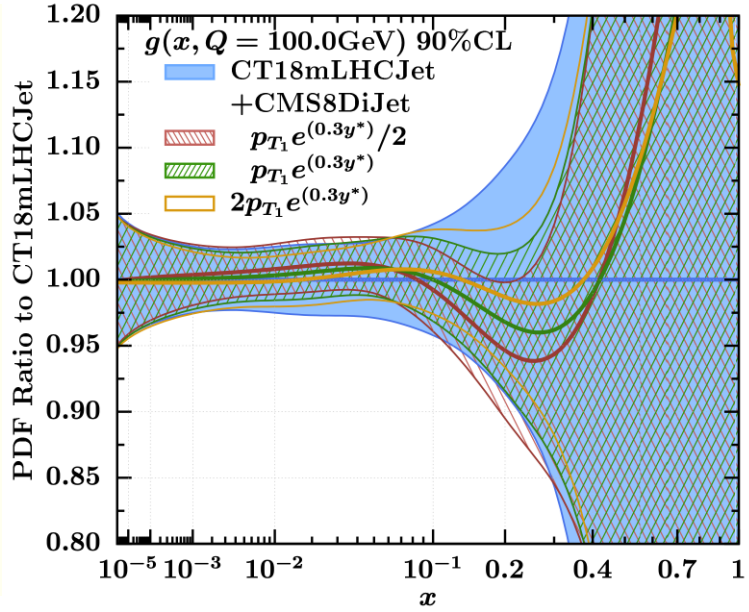
+ inclusive jets: small scale dependence, a harder $g(x, Q)$







+ dijets: <u>significant scale, dependence</u>, varied pulls on $g(x, Q)$



Ablat et al., PRD111 (2025) 3, 036033

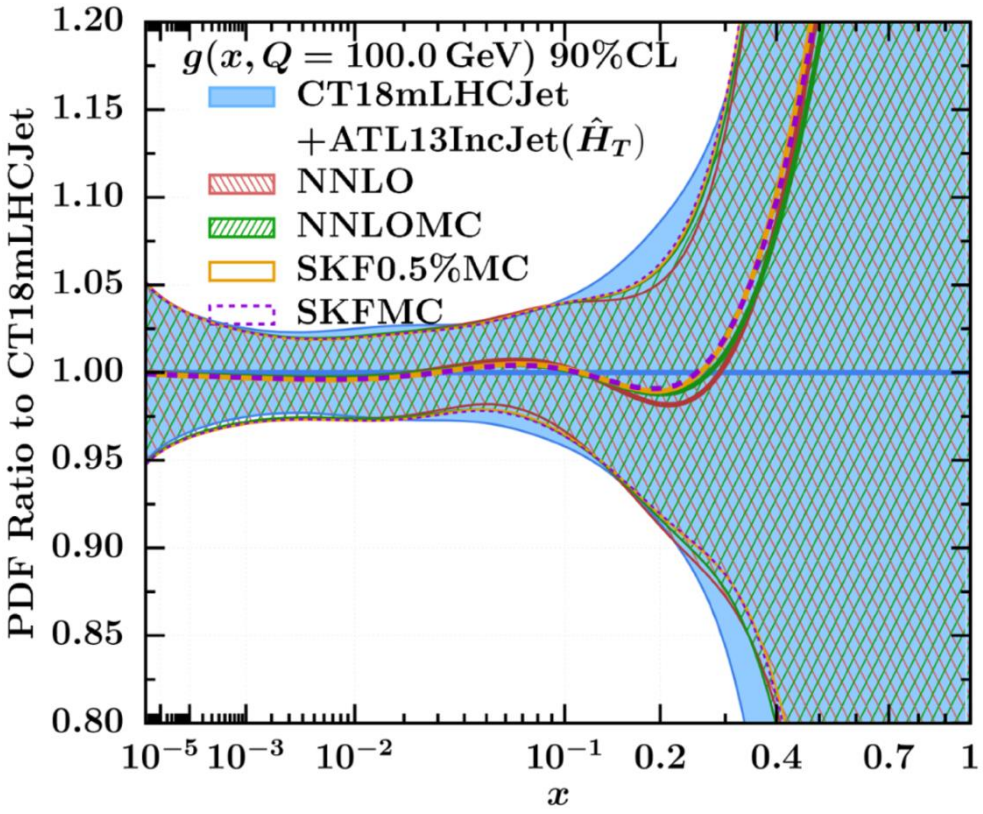
- Inclusive jet impact on $g(x, Q)$ is relatively independent of the scale choice. The final fit uses $\mu_{R,F} = p_T^j$, giving better χ^2
- **Dijet PDF impact substantially depends on scale choices**, especially for CMS8 TeV DiJet

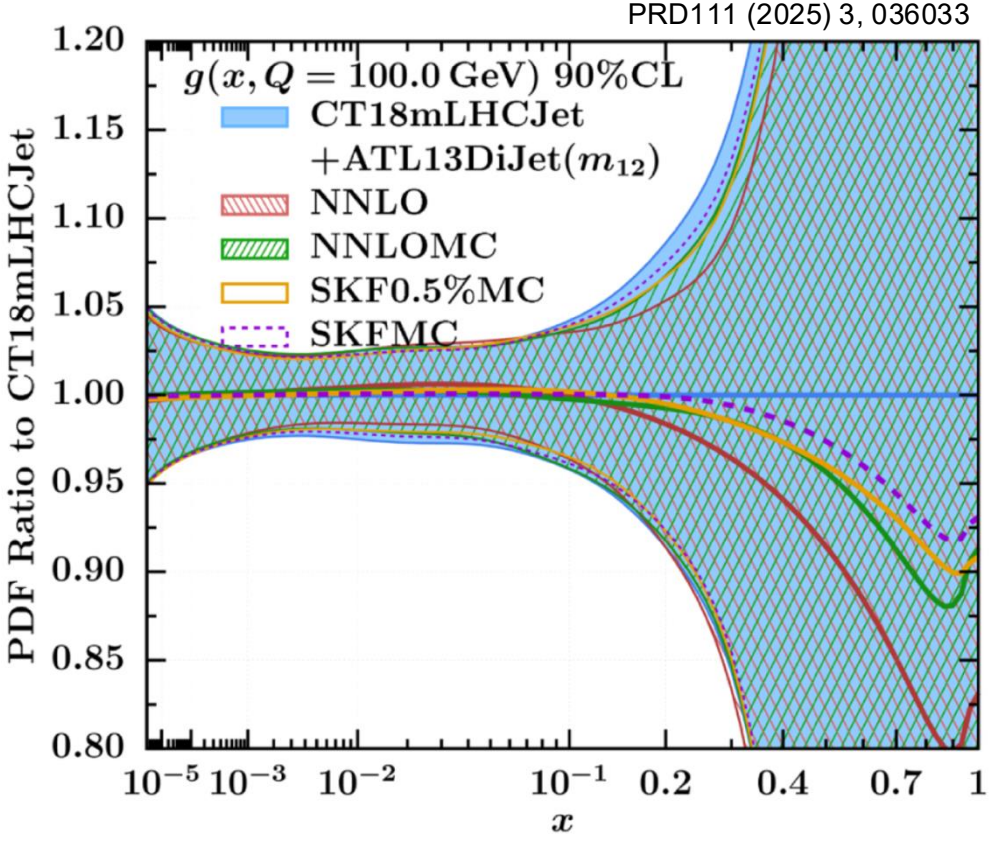
inclusive jet vs. dijet data sets: NNLO theory uncertainties

- relative stability of inclusive jet projection data applies also to theory choices in NNLO calculation

(use of K-factors; inclusion of 0.5% uncorr MC uncertainty; …) [NNLOjet]

PRD111 (2025) 3, 036033





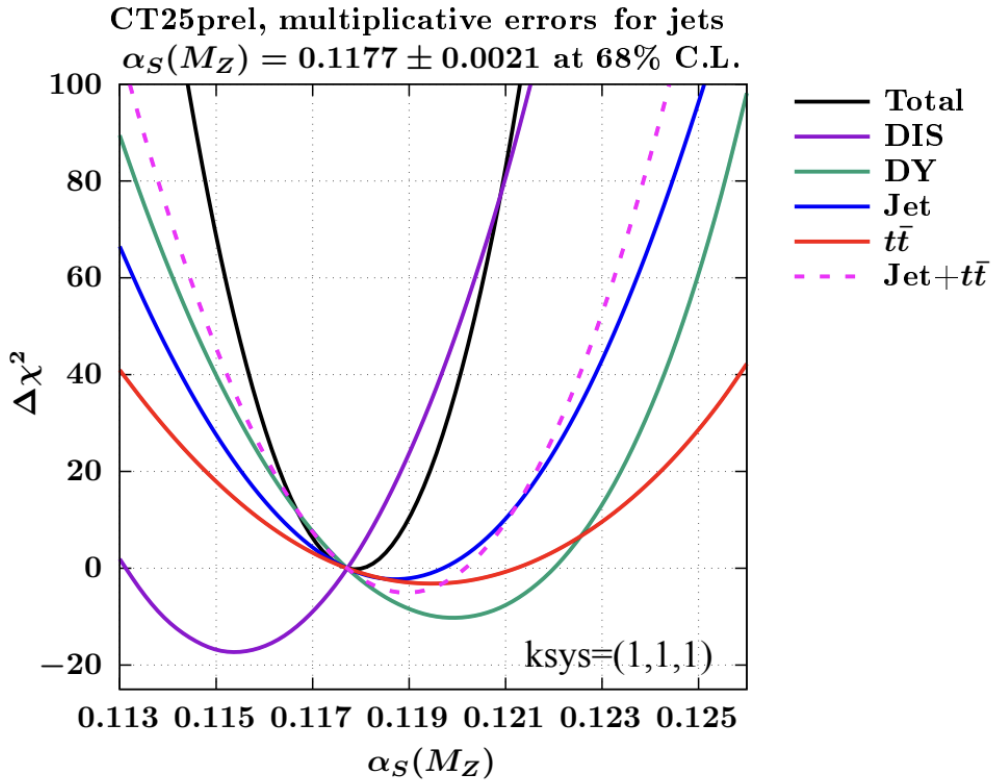
- Generet et al. (2604.25490): NNLL small-jet-radius resumm. can be consequential, including for incl. jets

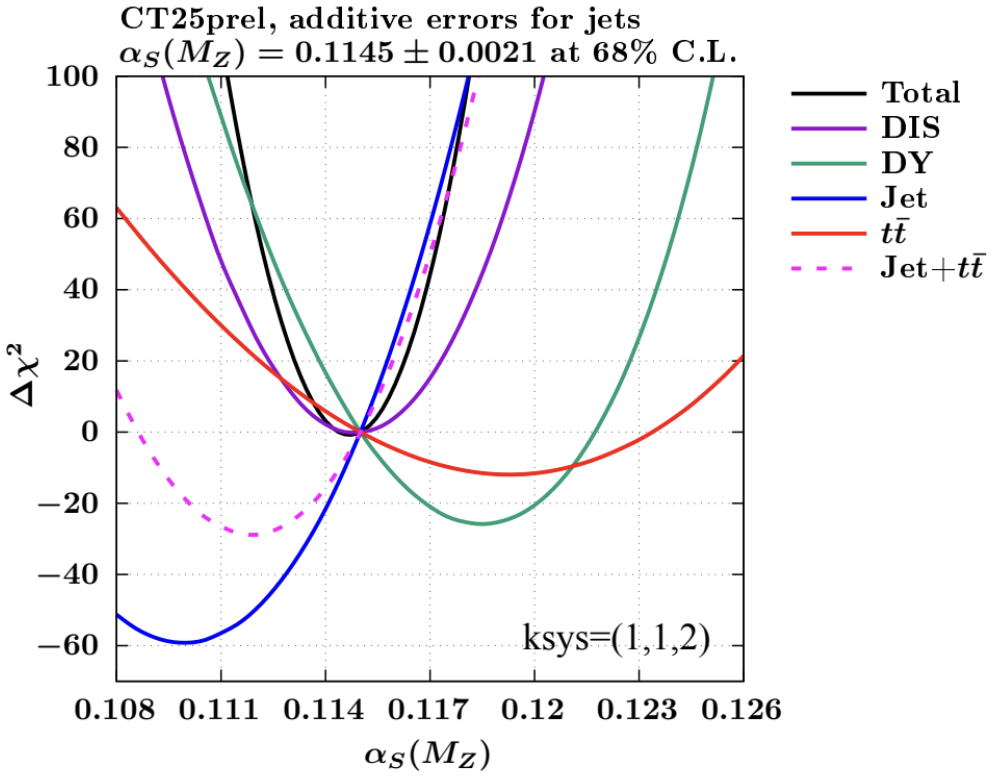
[see talks: **K. Lee, G. Stagnitto**!]

simultaneous fit(s) with α_s

2512.23792

- must extract alongside PDFs ('α_s series'): sequential refits at varying α_s
- significant upward pulls from inclusive jet data; $\Delta\chi^2$ for Jet+$t\bar{t}$ mainly driven by jet-production constraints



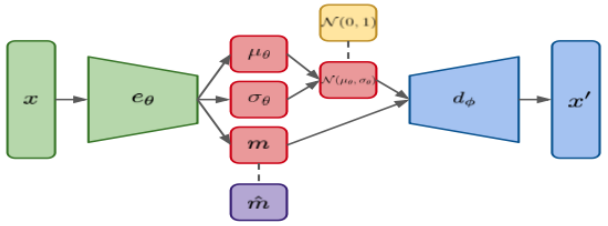


- treatment of jet systematics influential to pulls (de-correlation procedures; add./mult. prescriptions)

must explore alongside higher perturbative accuracy!

novel AI-based approaches to PDF extraction

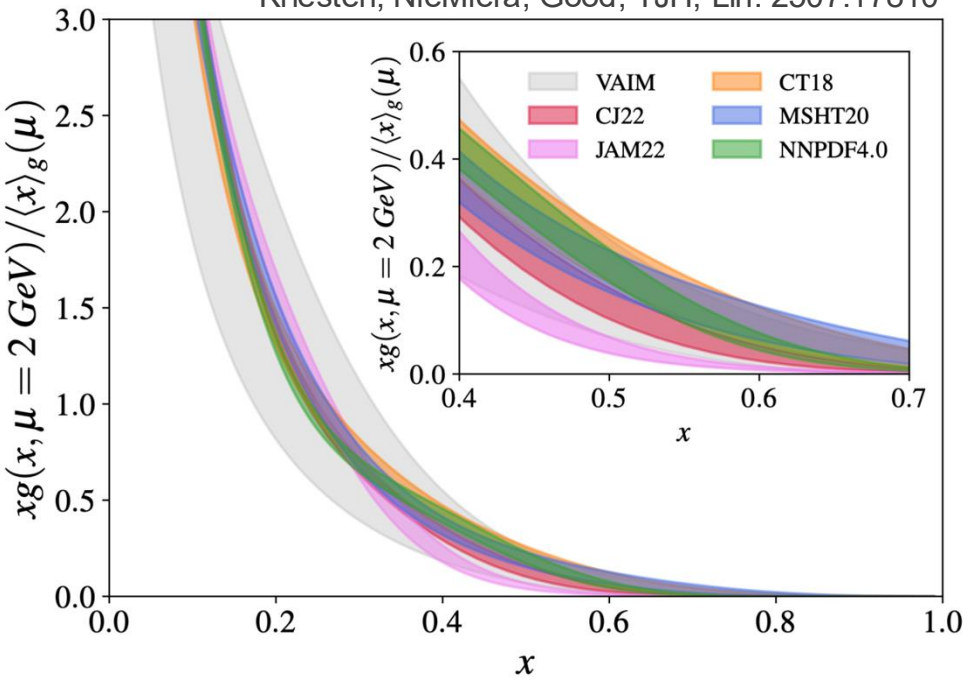
- AI/ML: emerging methods for HEP theory inverse problems



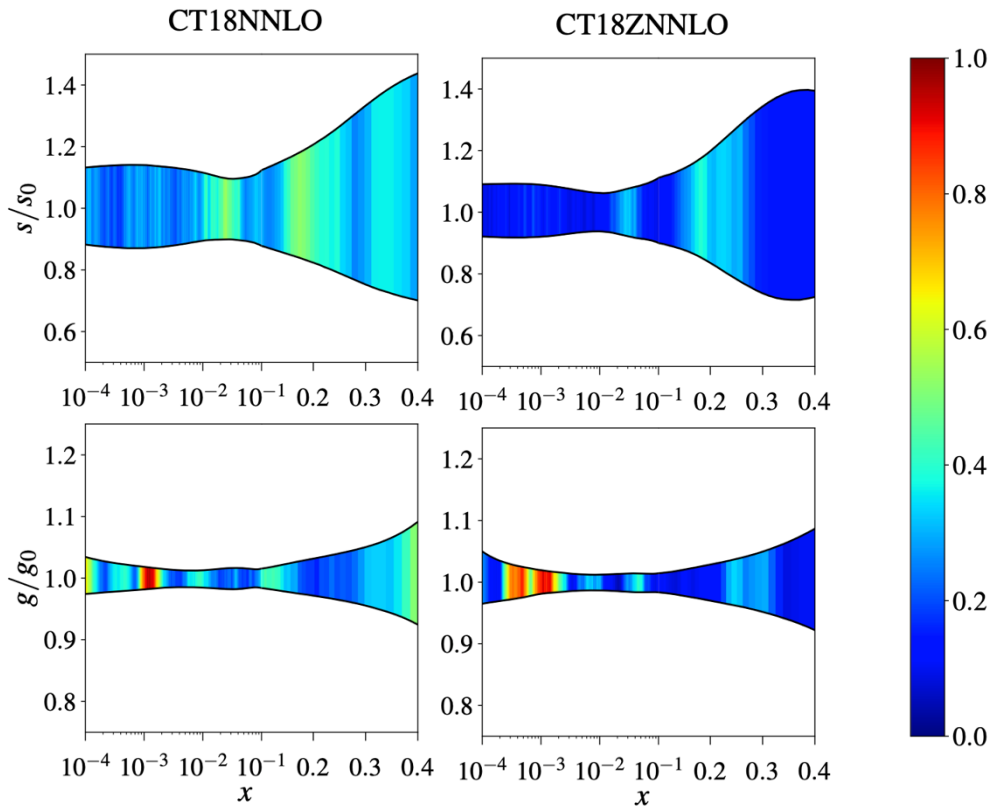
- *e.g.*, decode gluon PDF from lattice QCD info (RpITDs) through **VAIM** model [left]

$$\mathcal{M}(\nu, z^2) = \int_0^1 dx \, \frac{xg(x,\mu^2)}{\langle x \rangle_g(\mu)} R_{gg}(x\nu, z^2\mu^2) + \mathcal{O}(z^2 m^2, z^2 \Lambda_{\rm QCD})$$

Kriesten, NieMiera, Good, TJH, Lin: 2507.17810



Kriesten, Gomprecht, TJH: 2407.03411



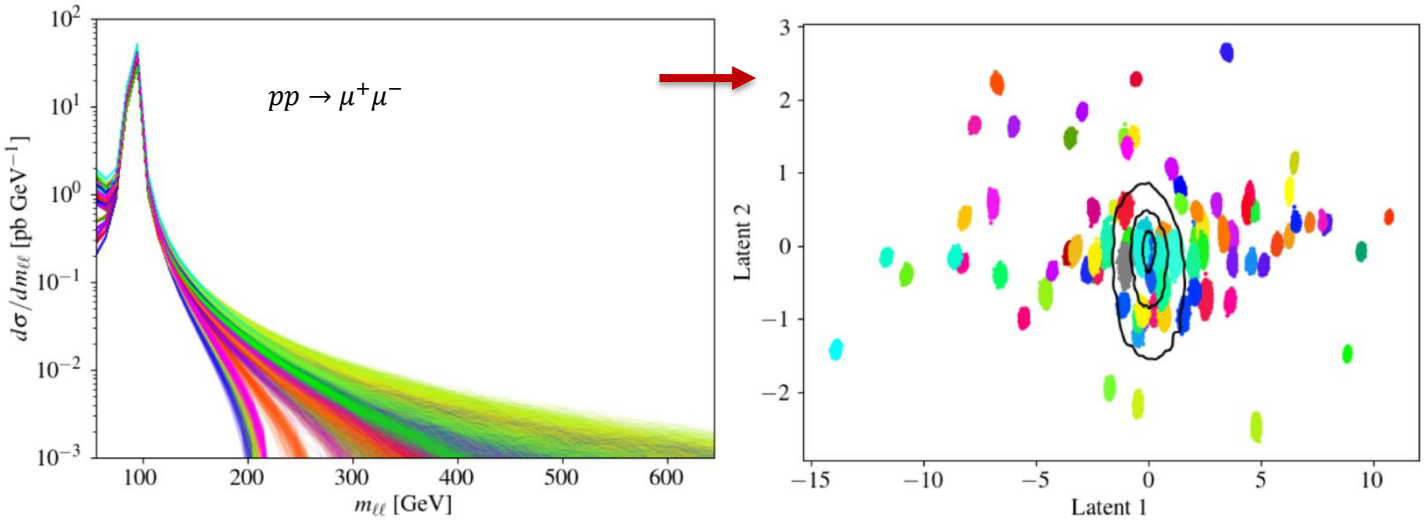
- **guided backpropagation**: train classifier to distinguish PDFs with varying theory settings [right]; extensible to fixed-order, aN3LO choices

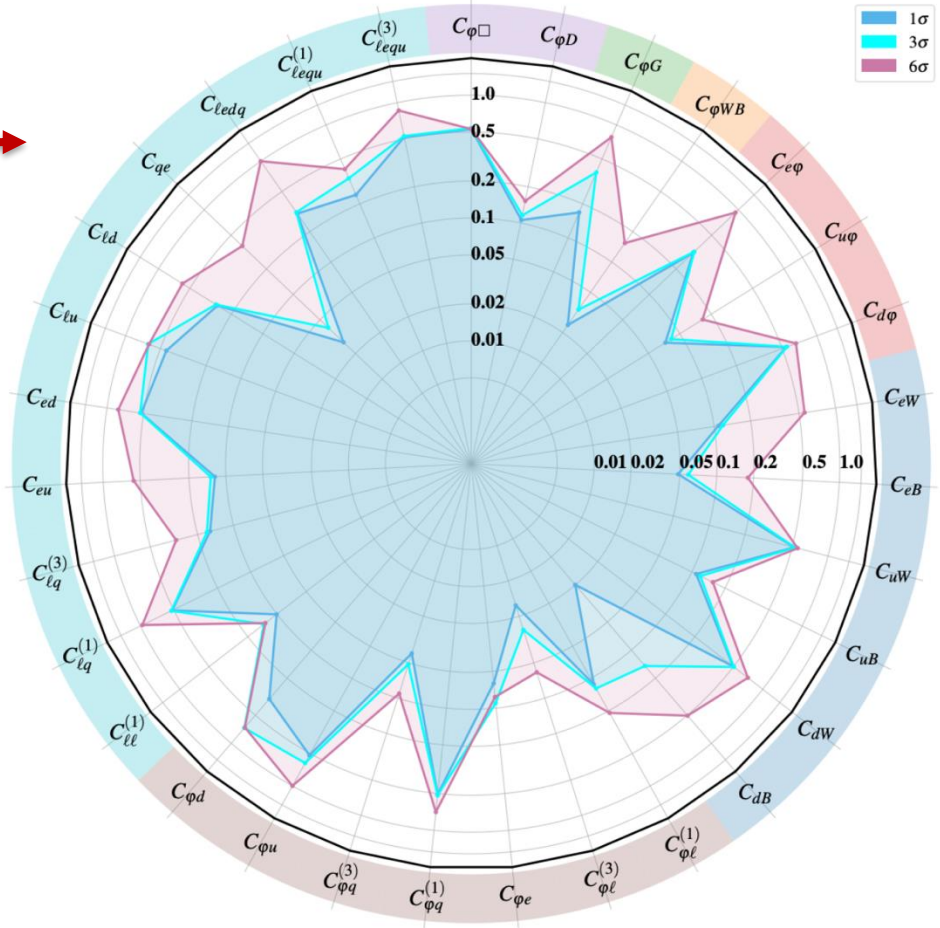
Brandon Kriesten

penultimate thought: foundation models for theory inferencing

- *idea*: pretrained foundation model (FM) aware of BSM sensitivities through simulated data

2512.15862: Das Bakshi, TJH, Kriesten





- FM acts as *fast analyzer*: *e.g.*, identifies sensitivity to specific Wilson coefficients (above, NC Drell-Yan with dim-6 sampling)

Supratim Das Bakshi

Brandon Kriesten

$$\mathcal{L}_{\text{SMEFT}} = \mathcal{L}_{\text{SM}} + \sum_i \frac{C_i O_i^{(6)}}{\Lambda^2} + \cdots$$

- template for inverse problems across HEP theory: amortize the forward model once, invert on demand

precision PDFs at the LHC: progress and outlook

- global PDF analyses: backbone of precision at the LHC; **progress hinges on consistent treatment of theory, data, and statistics**
- in EFT era, sensitivity depends on subtle interplay between PDFs and their BSM assumptions
- modern fits, including CT25: strong constraining power from 8, 13 TeV LHC sets; especially jets
  - ongoing progress to obtain, assemble higher-order corrections (QCD, EW, ...) on common footing
  - PDF UQ remains a vital frontier with new studies and approaches rapidly emerging
  - exciting cross-talk with planned experiments (EIC-LHC)
- AI/ML offers opportunities, challenges: novel PDF studies, reciprocal insights into big-data methods

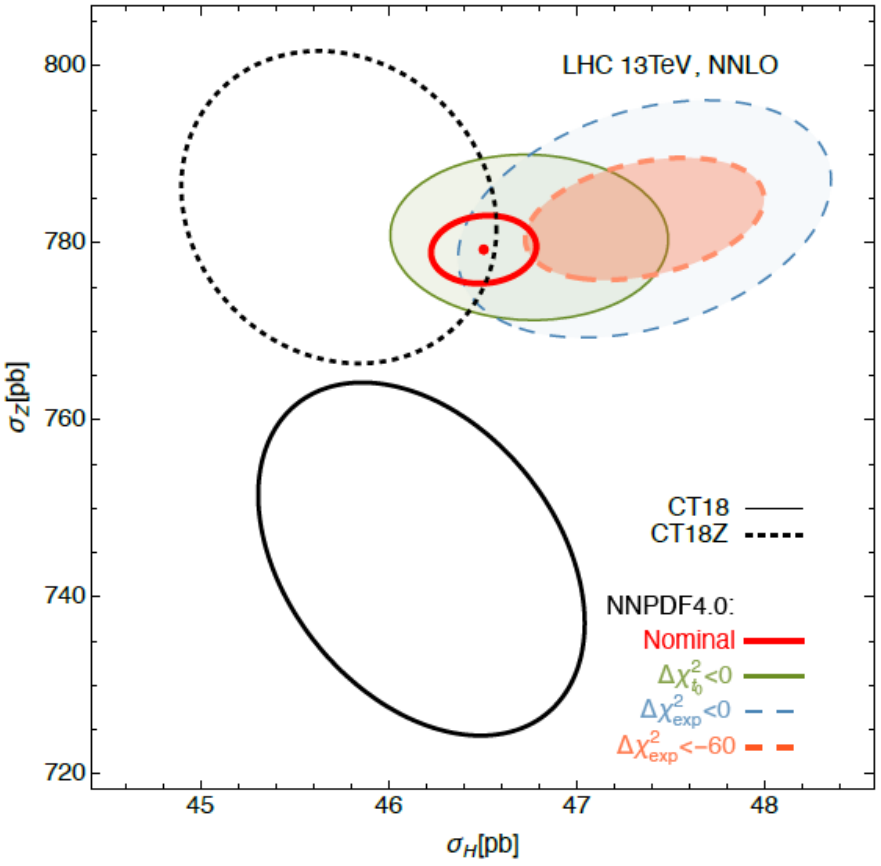
Backup

uncertainty quantification (UQ) developments

CT error bands quantify parametrization choice as a crucial source of epistemic uncertainty

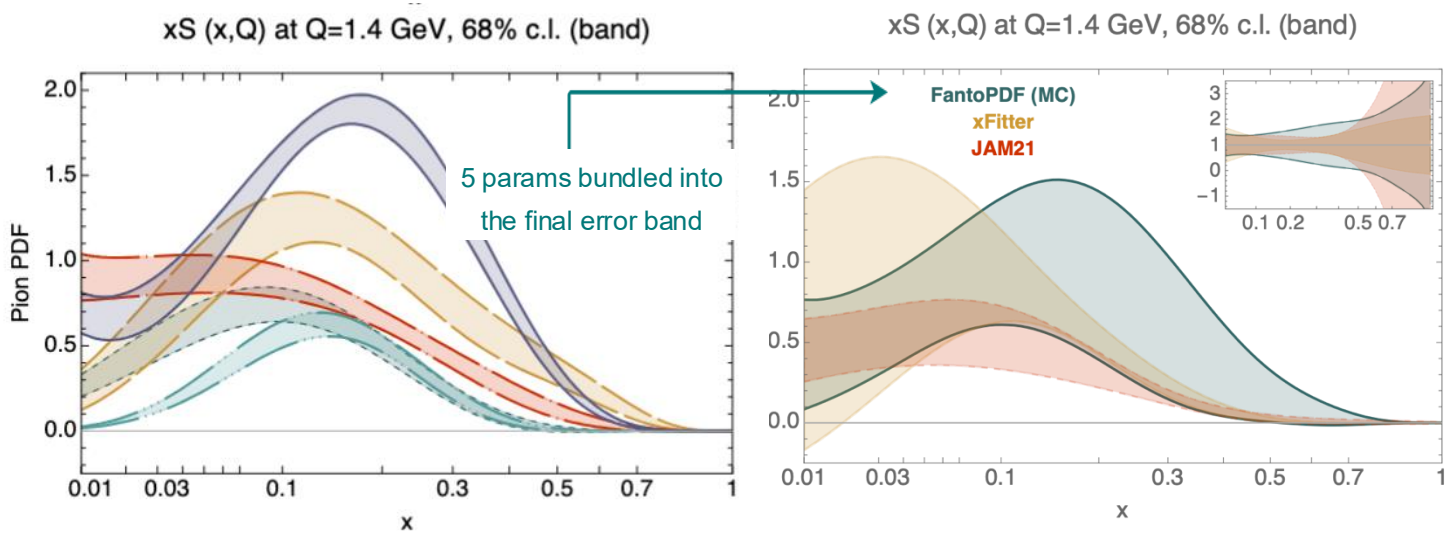
(essential for representative sampling over PDF models and fitting methodologies, robust pheno predictions)

[Courtoy et al., PRD107, 2205.10444]



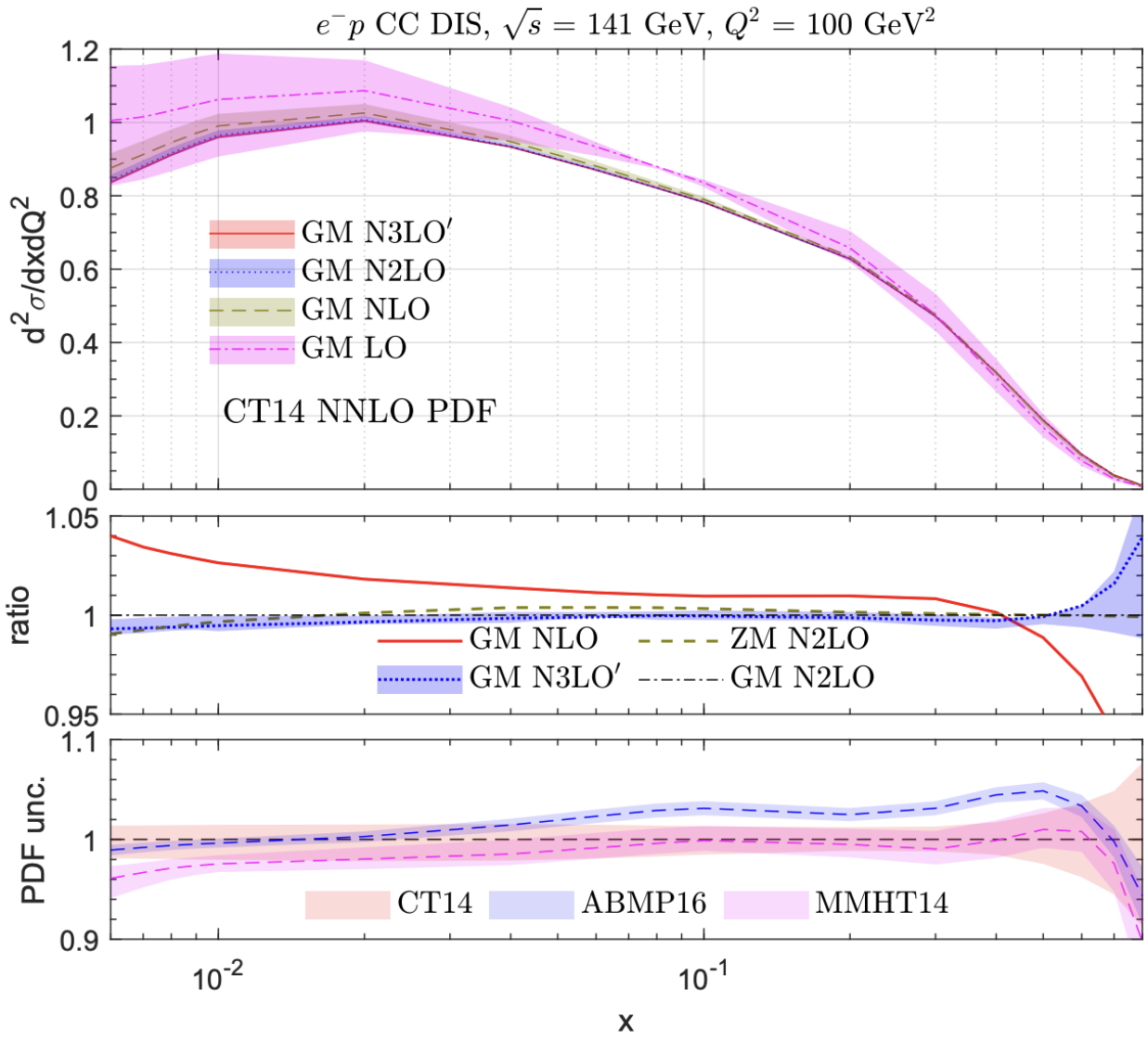
- studies of PDF parameterization dependence using Bézier curves: Fantômas4QCD — **testbed for exploring CT25 parametrizations**

[Kotz et al., 2311.08447]; Fantômas technical paper [2507.22969]



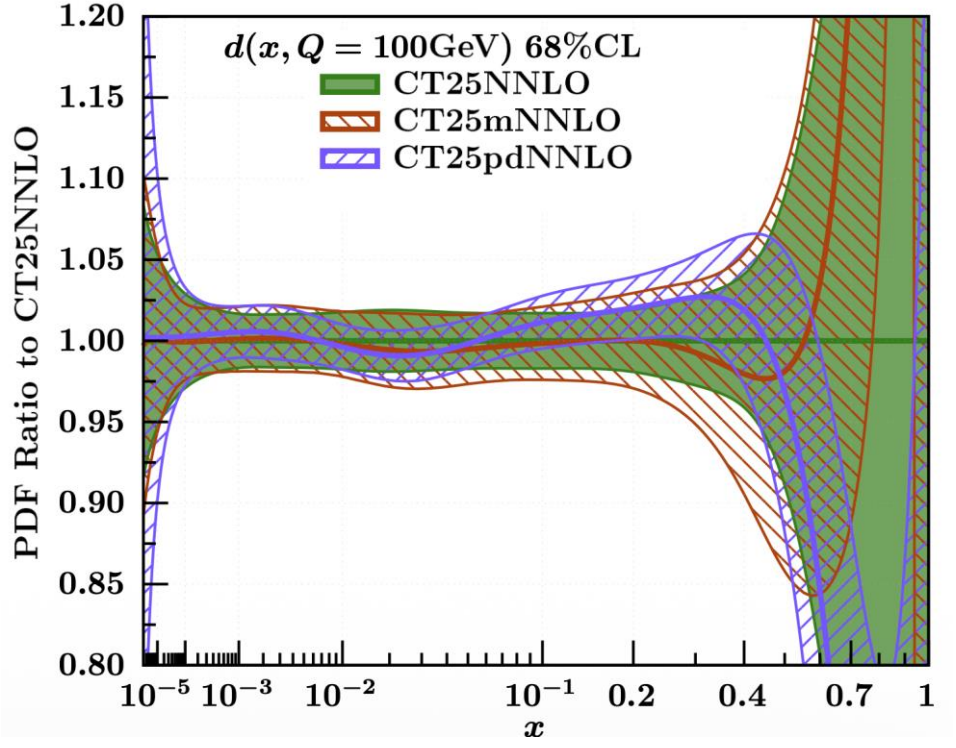
- complementary to parallel ML-based studies (later)

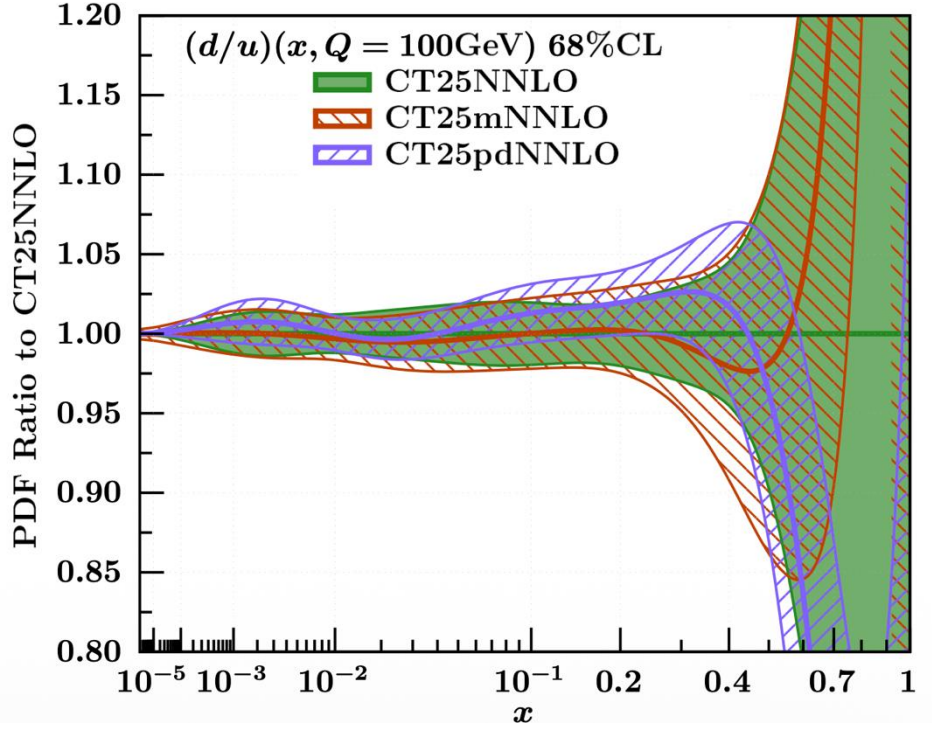
GM scheme up to N2LO (and ZM N3LO) DIS cross sections



CT25pd global fit results

- CT25pd deuteron corrections are significant *precision* effect
  - few-percent deviations in deuteron structure functions influence fitted PDFs
  - propagate to high-x shifts in d-PDF [below-left], d/u PDF ratio [below-right]
  - qualitative agreement with behavior in 2021 dedicated deuteron study (2102.01107)
- CT25pd PDFs consistent across deuteron correction models (*e.g.*, LT vs. HT)



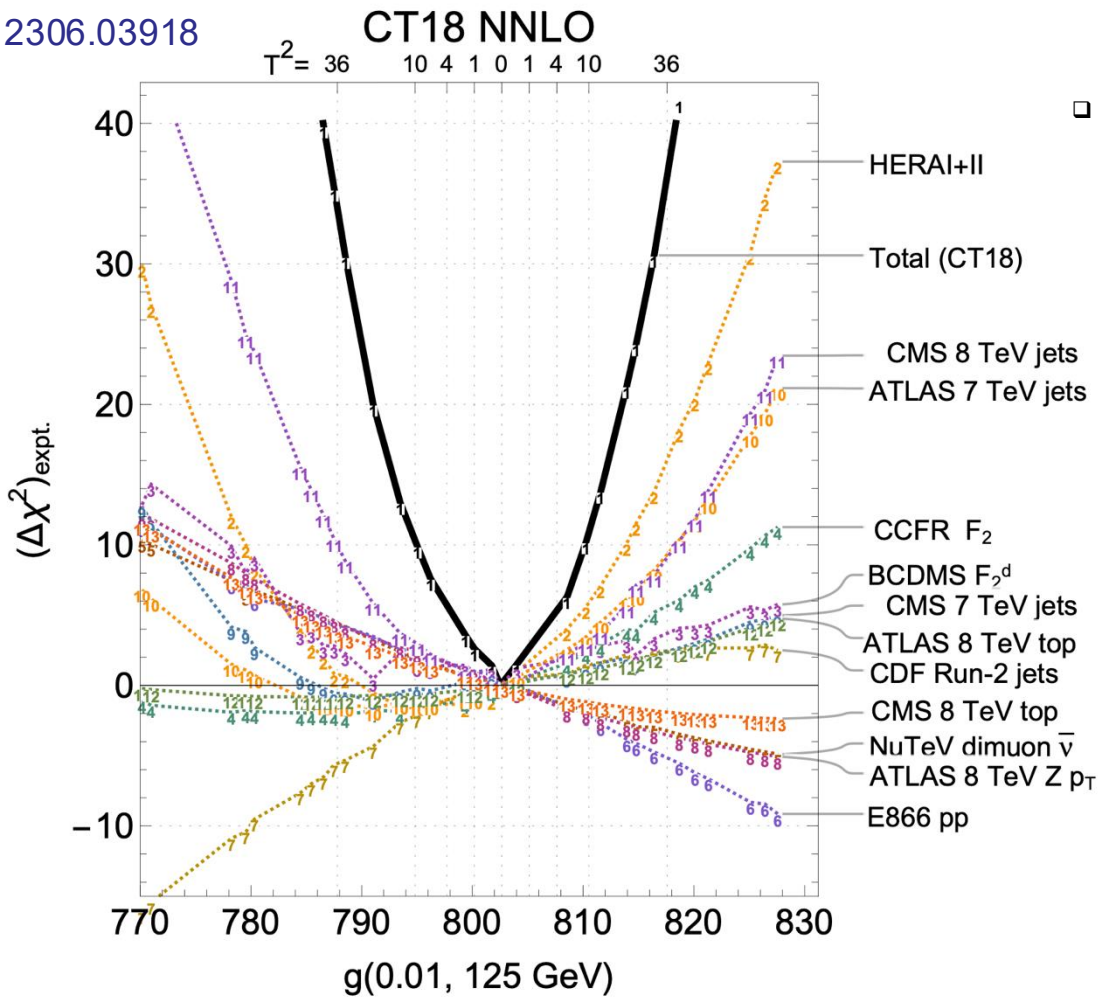


crucial frontier: reproducible, robust PDF uncertainties

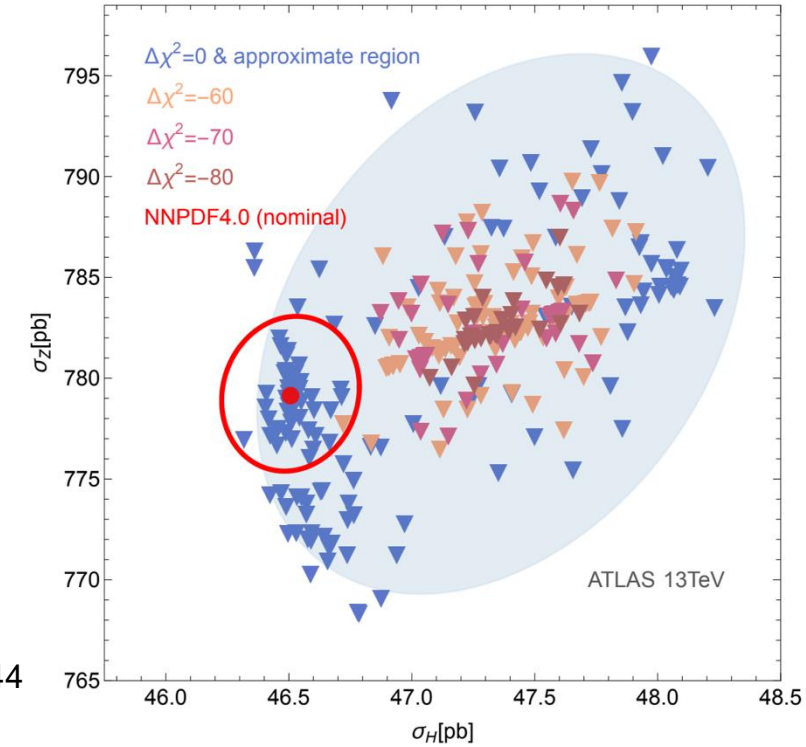
- community-wide interest in quantifying PDF uncertainties

→ quandaries in stat theory (*e.g.*, MC sampling challenges); exploration of large model spaces

2306.03918



- need first-principles research in PDF uncertainties



Courtoy et al., 2205.10444

- ongoing investigation of relation(s) among resampling, Hessian, ML, …, methods

Improved Hessian Method

Siqi Yang, et al,
arXiv: 2411.11645

- In standard Hessian analyses, applying the $\Delta\chi^2 = 1$ criterion will inevitably results in an underestimation of uncertainties arising from PDFs.

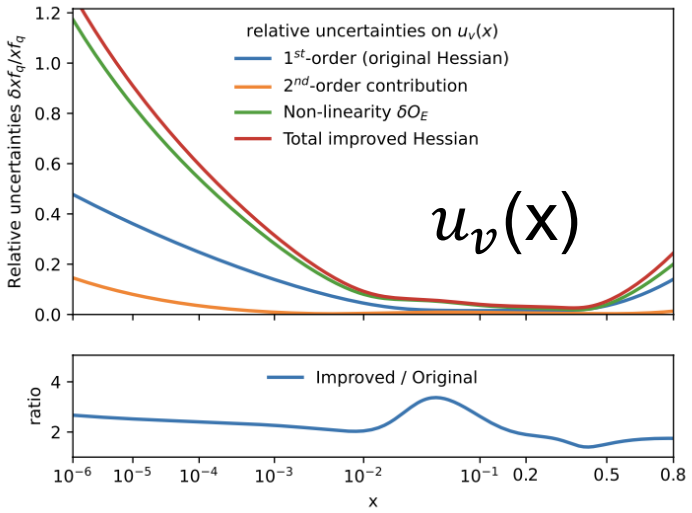
Standard Hessian Method
Assumption: physical observables depend linearly on PDF parameters

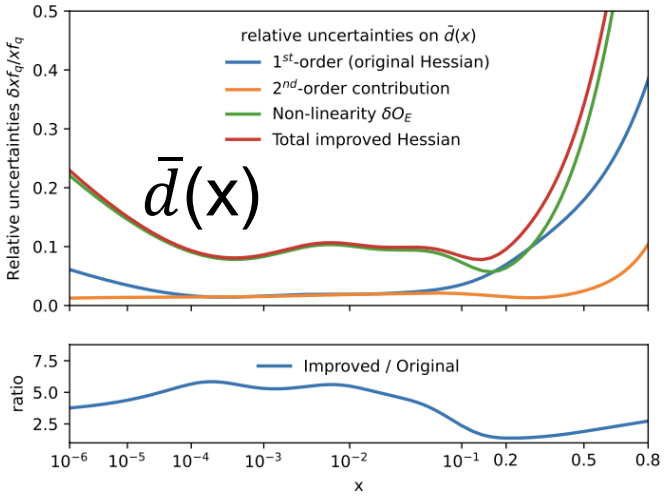
$$\mathcal{O}(a_i) \approx (O)(a_i^0) + \sum_i T_i(a_i - a_i^0) \approx \mathcal{O}(a_i^0) + \sum_i \frac{\partial \mathcal{O}}{\partial a_i} \mathrm{d}a_i$$

Improved Hessian Method introduced
1. Additional error sets derived from second-order derivatives.
2. A non-linearity uncertainty term, δO_E, based on deviations between linear approximations and full (nominal) fits.

$$\mathcal{O}(a_i) \approx \mathcal{O}(a_i^0) + \sum_i \frac{\partial \mathcal{O}}{\partial a_i} \mathrm{d}a_i + \sum_{ij} \frac{\partial \mathcal{O}}{\partial a_i \partial a_j} \mathrm{d}a_i \mathrm{d}a_j$$

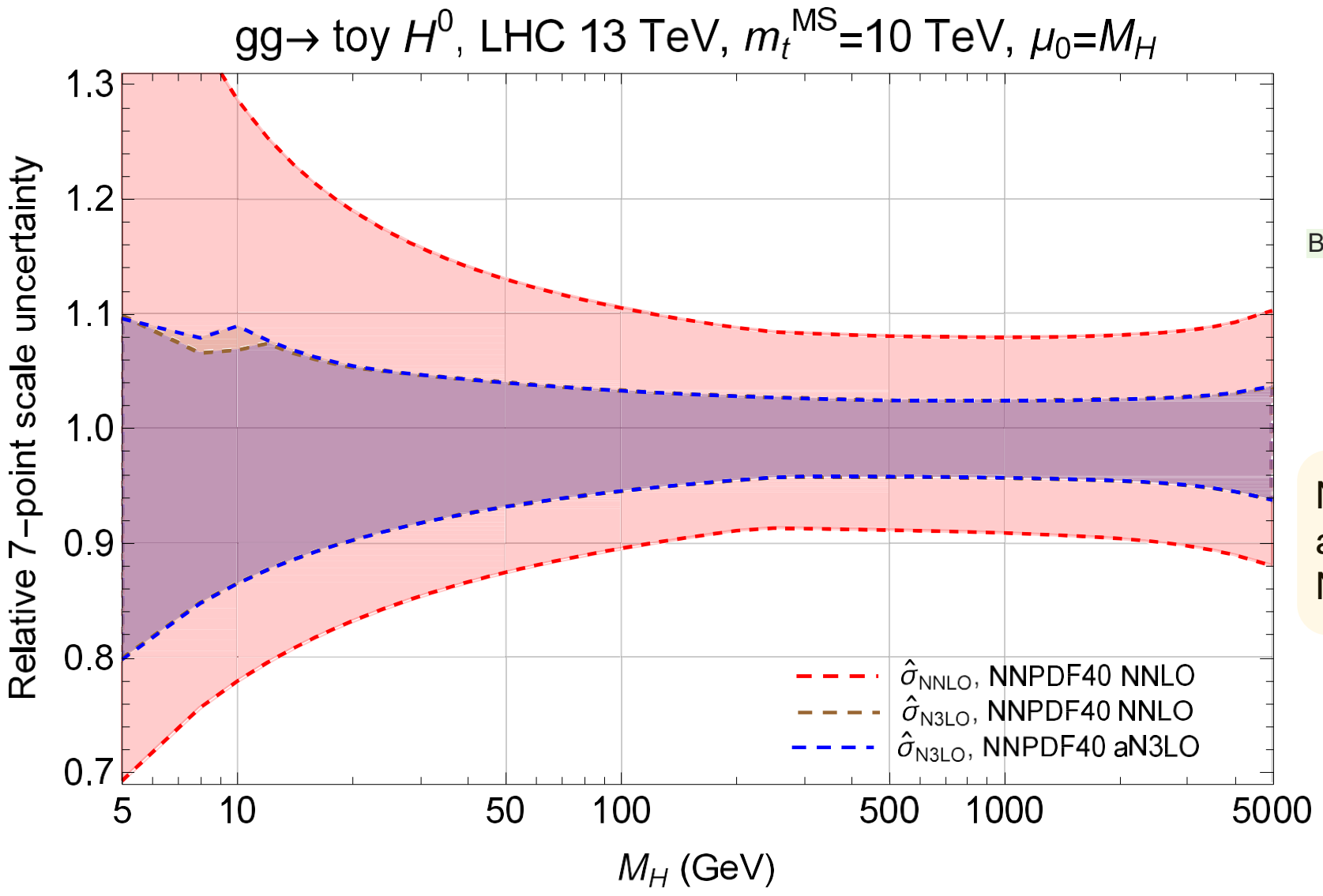
Relative uncertainties: Improved Hessian method tested in a pseudo global analysis





- The improved method shows that non-linear uncertainties can be significantly larger than the uncertainties estimated by the standard Hessian method.
- These effects are most pronounced in low or high x and for parton flavors (e.g., sea quarks) which are less constrained by data.
- Even with present precision data, non-linear effects remain sizable, suggesting that traditional uncertainty estimates. with $\Delta\chi^2 = 1$, may be overly optimistic.

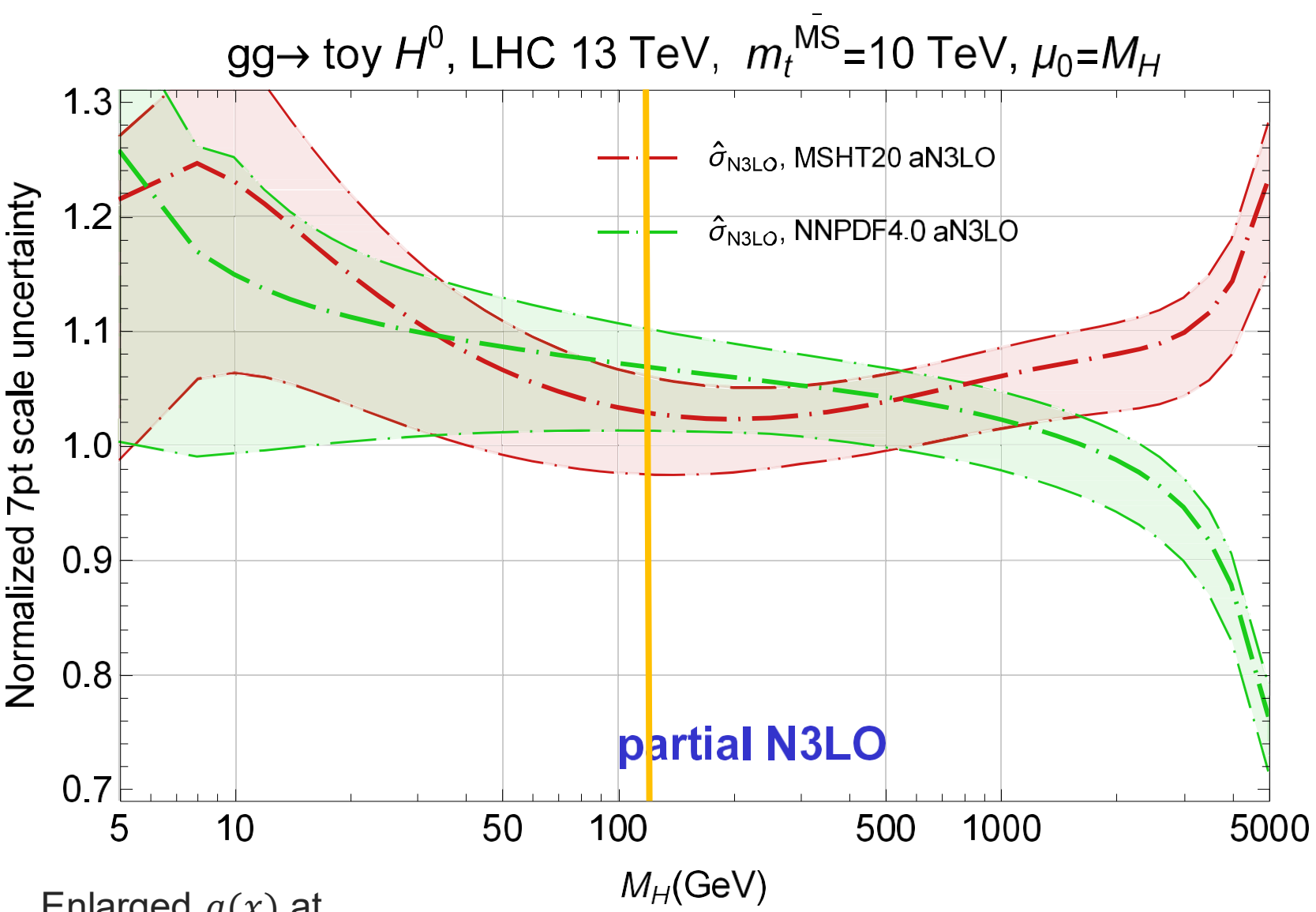
what about **partial N3LO** (pN3LO)? total cross sections can serve as test bed



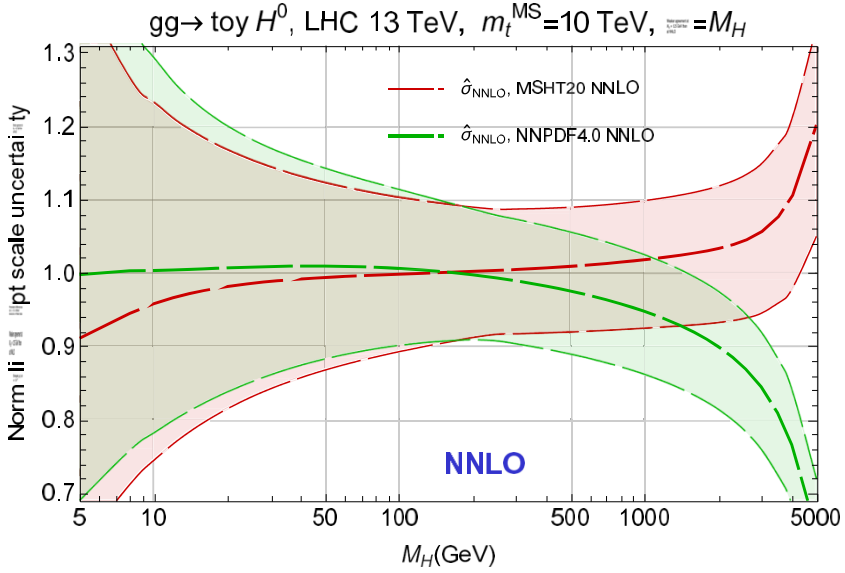
Baglio, Duhr, Mistlberger, Szafron (2209.06138)

N3LO scale uncertainty is about the same with either NNLO or aN3LO PDFs

At $M_H \approx 10$ GeV, more variability due to the $b\bar{b}$ mass threshold



cf.:



Enlarged $g(x)$ at $x < 10^{-3}$

Weaker agreement at $M_H = 125$ GeV than at NNLO

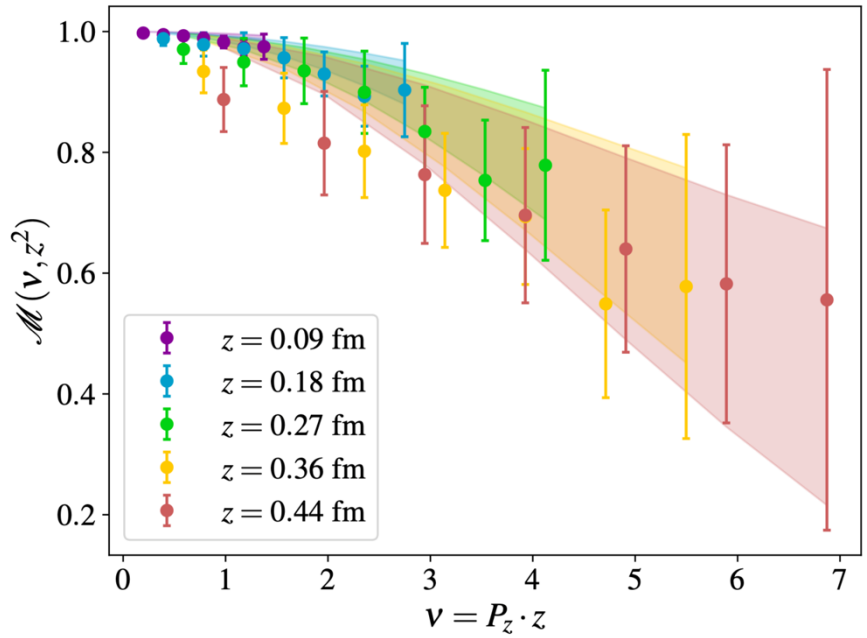
Persistent differences at $x > 0.1$ reflect tensions in fitted data; imprinted on separation at large $M_X \gtrsim 1$ TeV

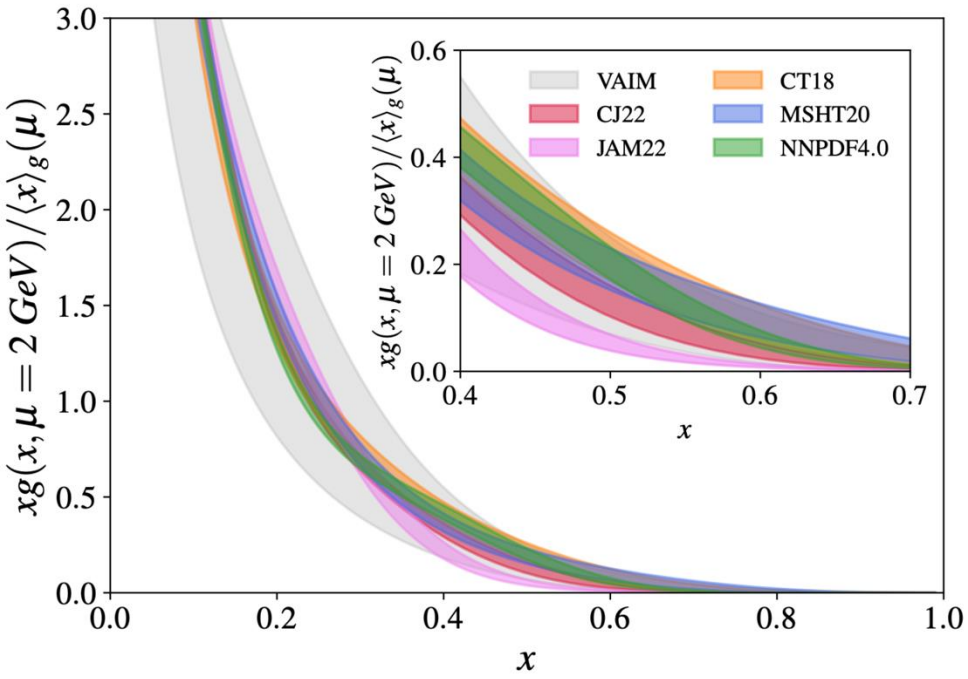
extension: direct unfolding of gluon density

Kriesten, NieMiera, Good, TJH, Lin

- generalizes – latent space of lattice M.E.s generatively predicts gluon PDF

("reduced pseudo-loffe-time distribution", RpITD)



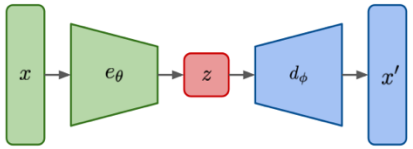
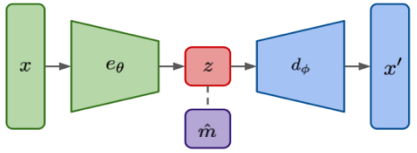
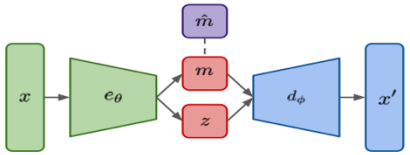
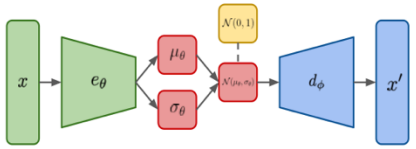
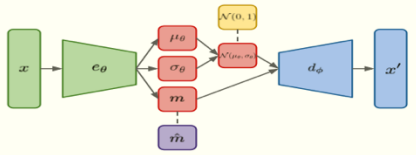


$$\mathcal{M}(\nu, z^2) = \int_0^1 dx \, \frac{xg(x, \mu^2)}{\langle x \rangle_g(\mu)} R_{gg}(x\nu, z^2\mu^2) + \mathcal{O}(z^2 m^2, z^2 \Lambda_{\mathrm{QCD}})$$

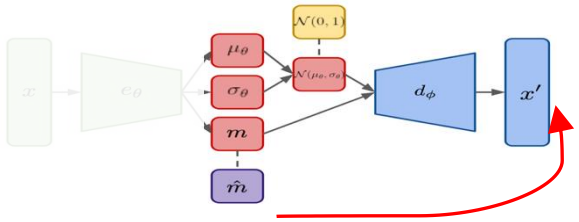
➢ *good agreement with pheno PDFs! can be extended with lattice, empirical data*

many alternative network models possible

- variations on encoder-decoder structure; different internal topologies

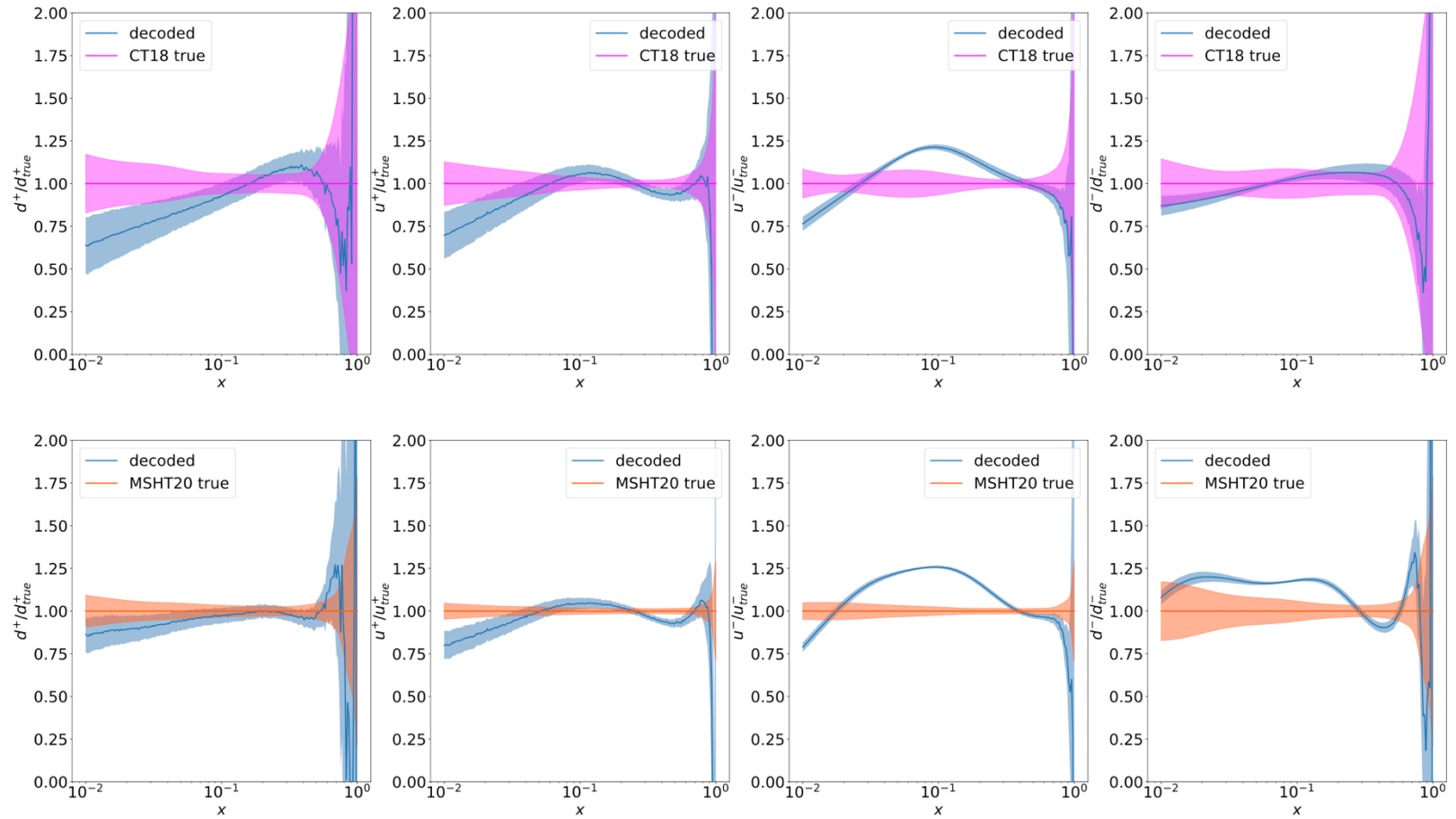
Name	Diagram	Loss	Recreates PDFs	Tractable Latent	Free Latent Dimension	Moment Constraint
AE		$\mathcal{L} = \|x - d_\phi(e_\theta(x))\|_2^2$	✓	✗	✓	✗
AE-CL		$\mathcal{L} = \|x - d_\phi(e_\theta(x))\|_2^2 + \|z - \hat{m}\|_2^2$	✓	✓	✓	✓
AE-WC		$\mathcal{L} = \|x - d_\phi(e_\theta(x))\|_2^2 + \|m - \hat{m}\|_2^2$	✓	✗	✓	✓
VAE		$\mathcal{L} = \|x - d_\phi(e_\theta(x))\|_2^2 + KL(\mathcal{N}(\mu_\theta, \sigma_\theta)\|\mathcal{N}(0,1))$	✓	✓	✓	✗
VAIM		$\mathcal{L} = \|x - d_\phi(e_\theta(x))\|_2^2 + \|m - \hat{m}\|_2^2 + KL(\mathcal{N}(\mu_\theta, \sigma_\theta)\|\mathcal{N}(0,1))$	✓	✓	✓	✓

compare ML model against pheno PDFs



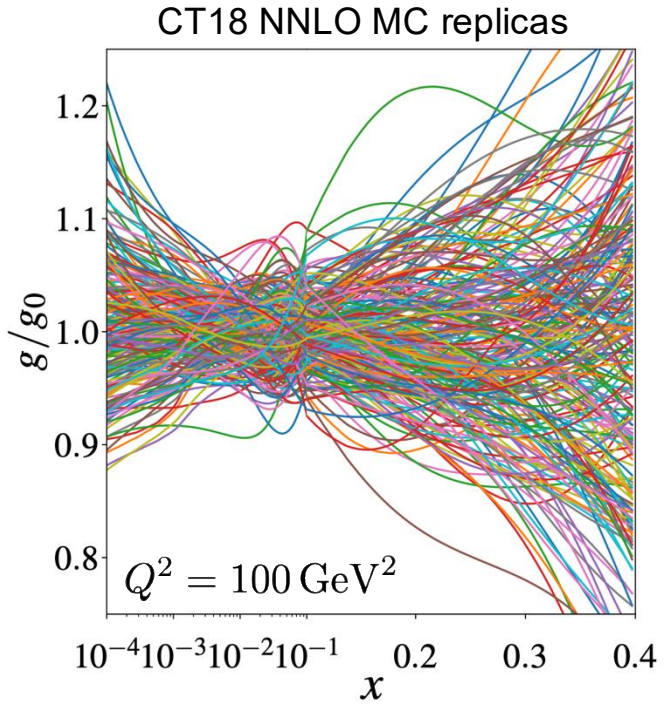
- given imposed tractability of the latent, amputate the VAIM: [moments] → [PDFs]

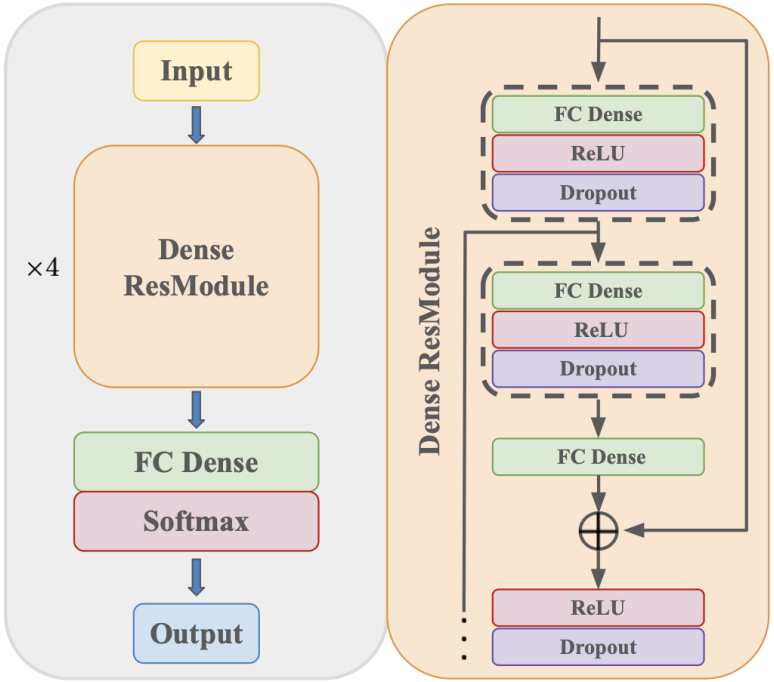
→ predict pheno PDFs from their moments



identify salient features: guided backpropagation (GBP)

Kriesten, TJH: JHEP11 (2024) 007





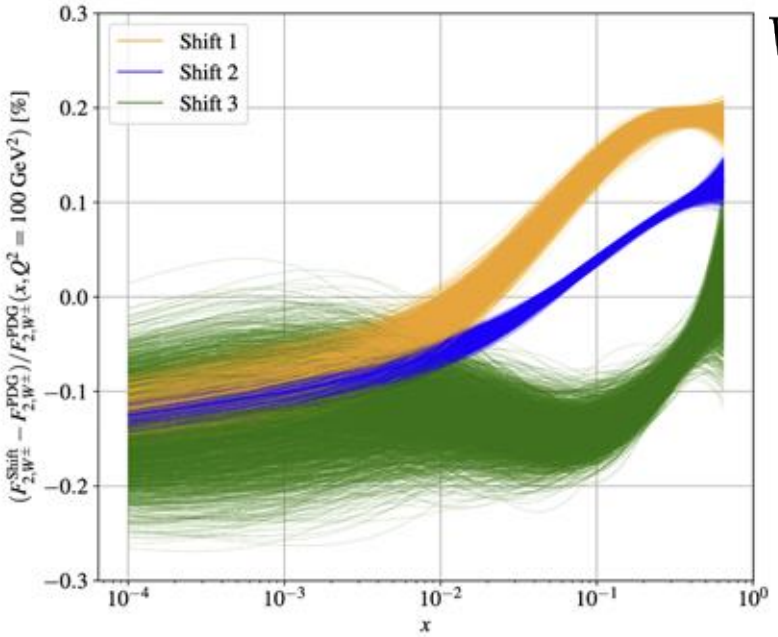
$$\frac{\partial f^{\text{out}}}{\partial f_i^\ell} = (f_i^\ell > 0) \cdot \left(\frac{\partial f^{\text{out}}}{\partial f_i^{\ell+1}} > 0 \right) \cdot \frac{\partial f^{\text{out}}}{\partial f_i^{\ell+1}}$$

train a ResNet-like model on MC PDF replicas;
backpropagate classification scores to PDF shapes

[public code available]

(evidential) learning BSM theory differences alongside PDF variations

$$F_2^{W^\pm}(x,Q^2) \equiv \frac{1}{2}\left(F_2^{W^+}(x,Q^2) + F_2^{W^-}(x,Q^2)\right)$$

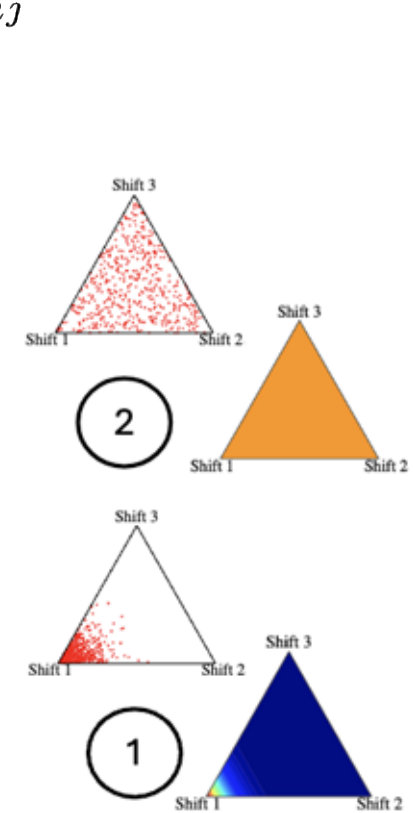


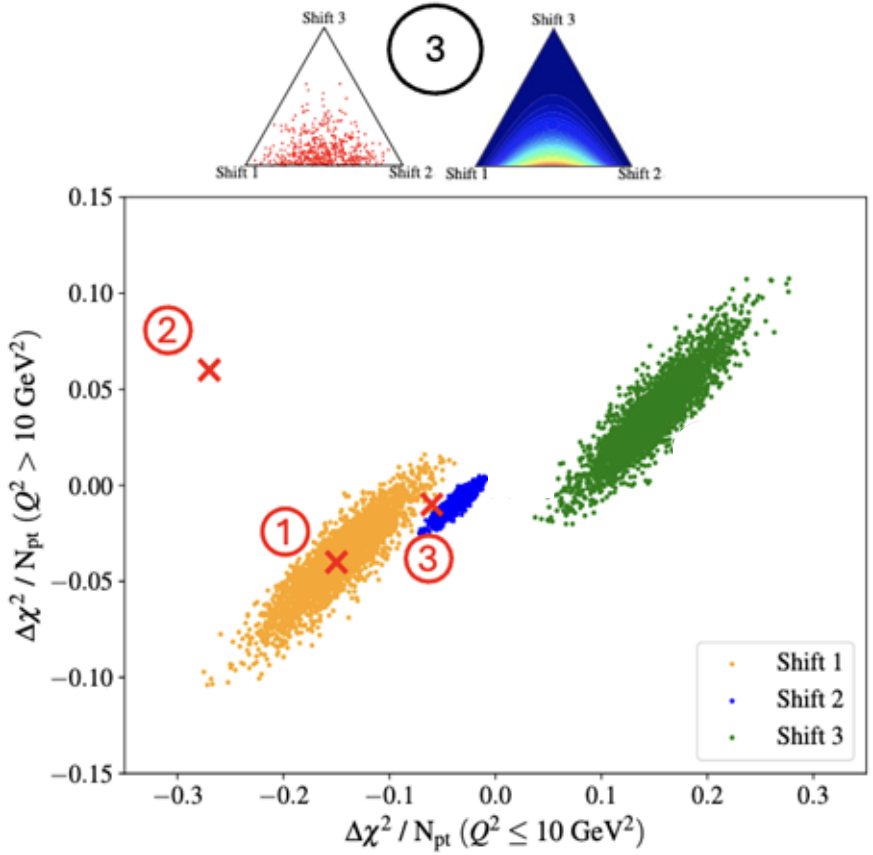
$$V_{ij} \to V'_{ij} = \Delta_{ij} \cdot V_{ij}$$

$$\mathcal{L}_{\text{WEFT}} \supset -\frac{2V_{ij}}{v^2}\left\{\left[1+\epsilon_L^{ij}\right]_{ab}\left(\bar{u}^i\gamma^\mu P_L d^j\right)\left(\bar{\ell}_a\gamma_\mu P_L \nu_b\right) + \text{h.c.}\right\}$$

Kriesten, TJH: 2412.16286

- non-standard neutrino interactions can shift high-x DIS structure functions (*e.g.*, CDHSW predictions)





- EDL techniques: distinguish BSM scenarios up to PDF uncertainties; quantify confidence in BSM separation