

New directions in precision PDF analysis: CT25 and beyond

Timothy J Hobbs

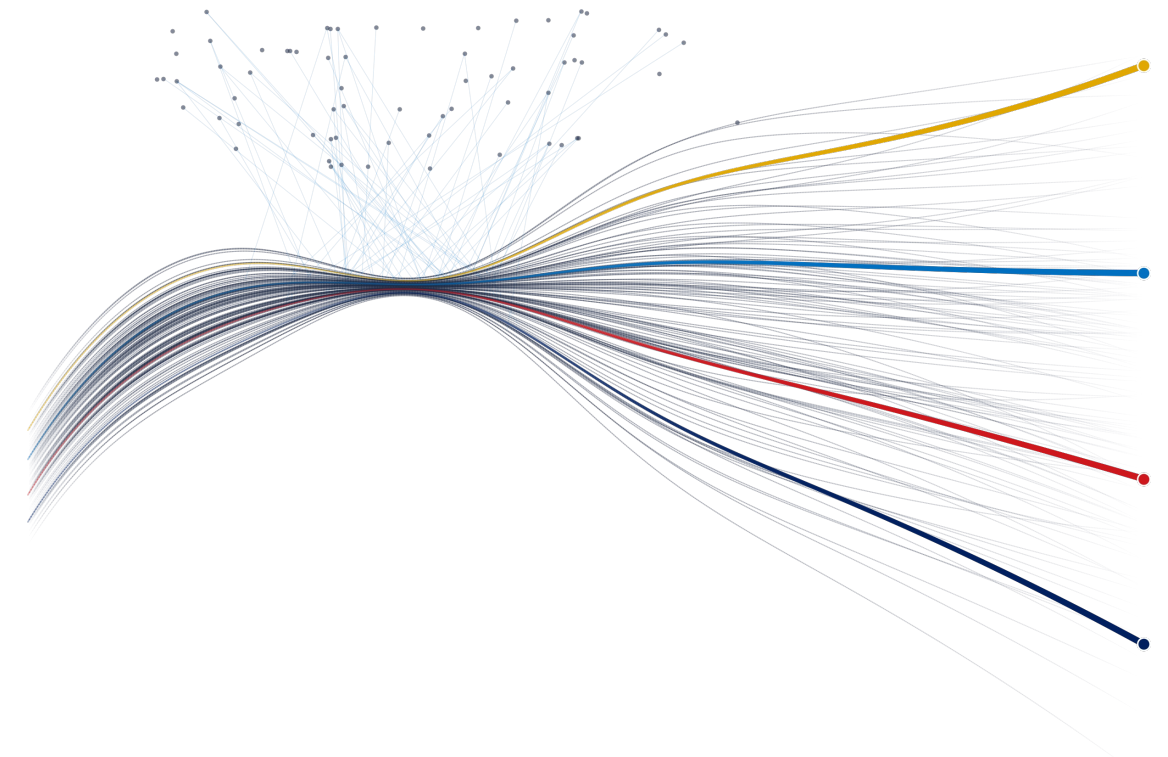
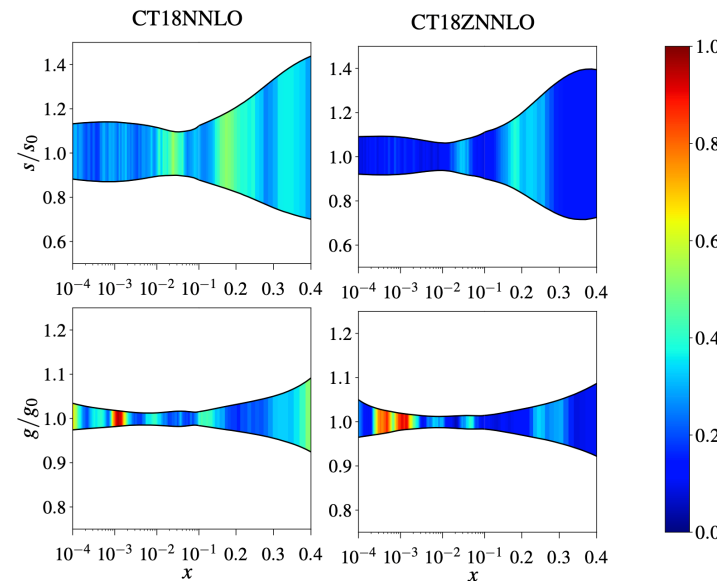
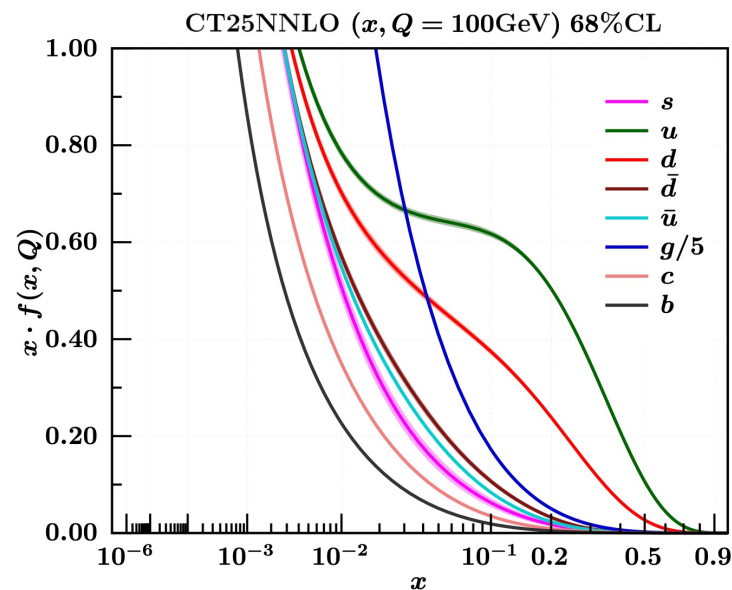
HEP Theory Group – Argonne National Laboratory, USA

QCD@LHC 2026

Oxford

28 Sept 2026

special thanks: CTEQ-TEA colleagues,
Max Ponce-Chavez, Brandon Kriesten, ...



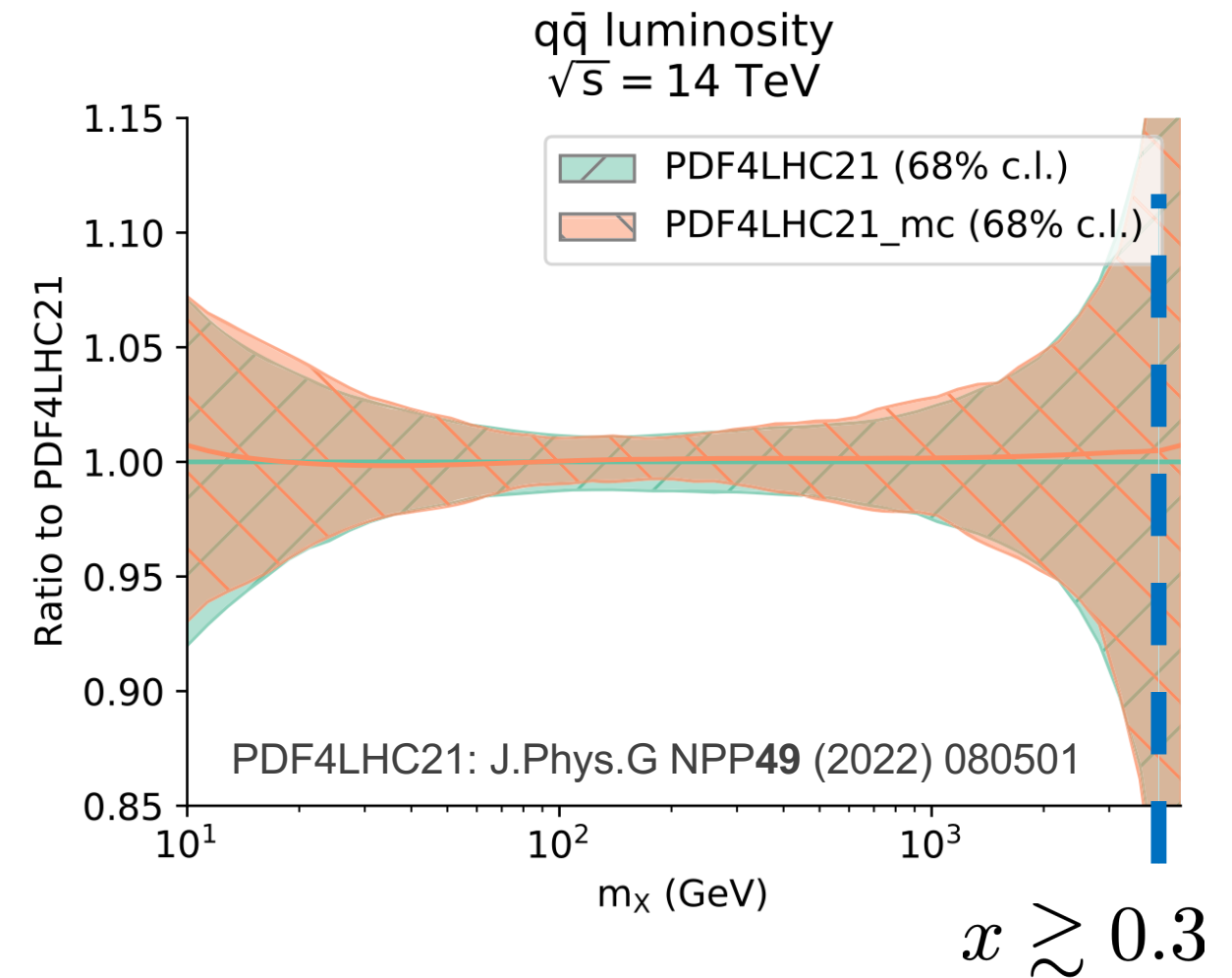
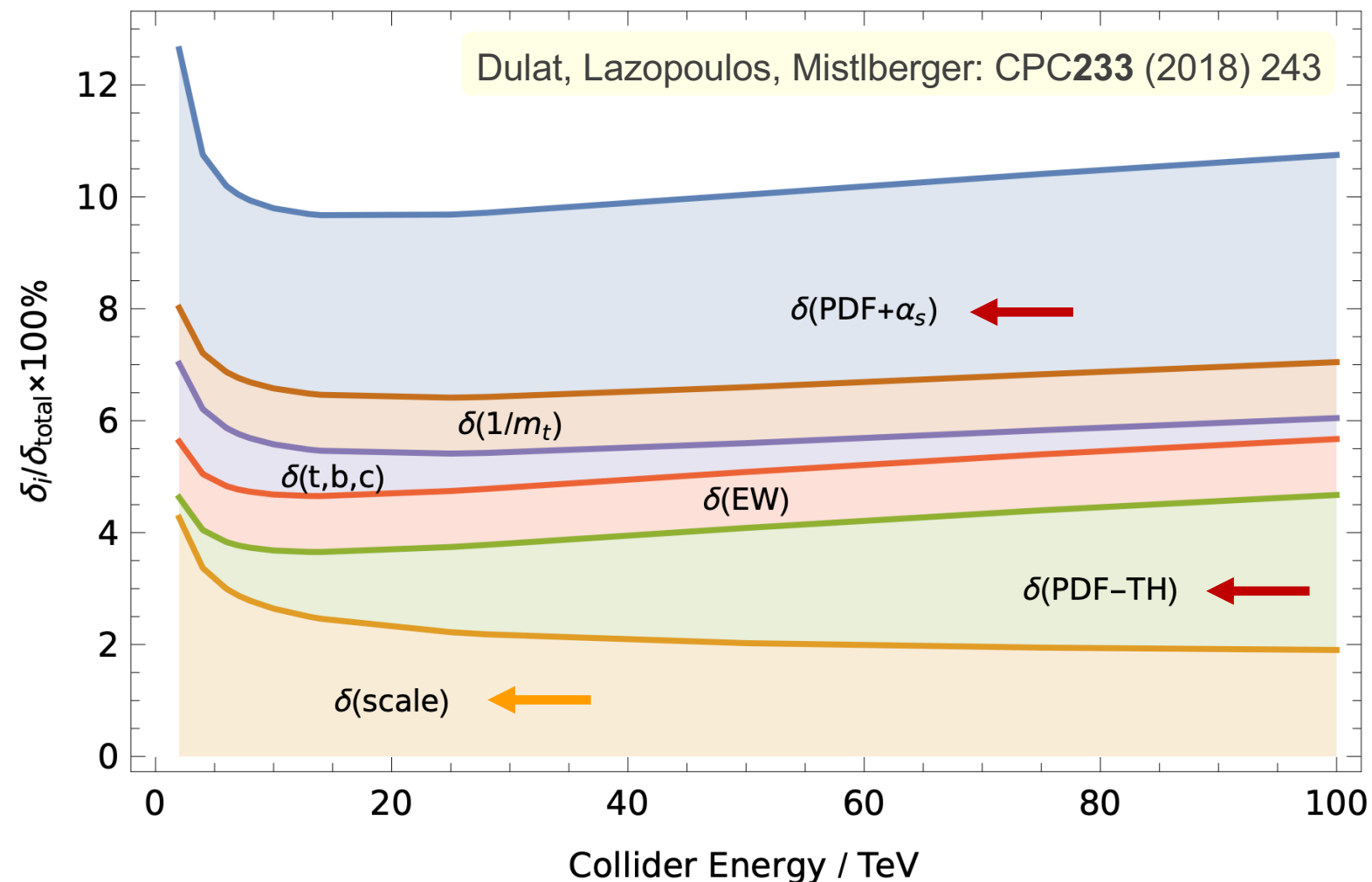
Argonne
NATIONAL LABORATORY

CTEQ

PDFs are a discovery and precision bottleneck at LHC

□ uncertainties in flagship processes [Higgs; Z' , W' , ..., production] receive substantial PDF contribution

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

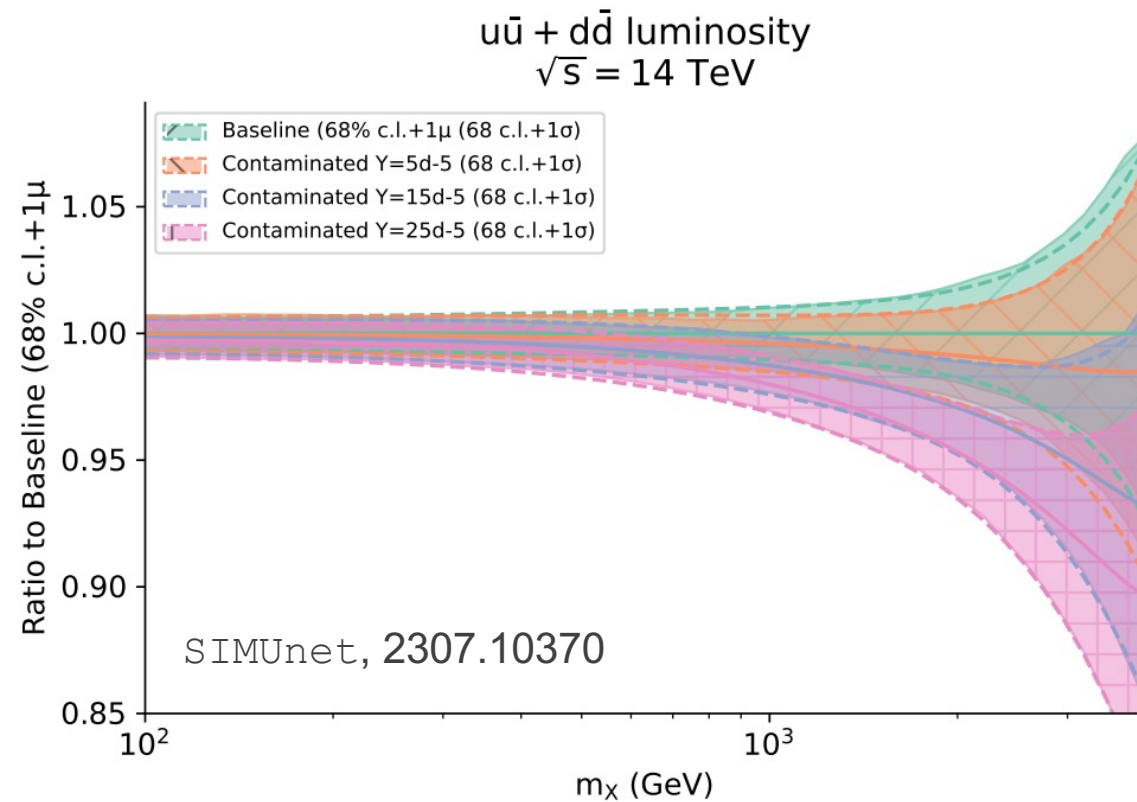


→ (high- x) PDF uncertainties limit searches for TeV-scale particles at the LHC

challenge of PDF-BSM (SMEFT) interdependence

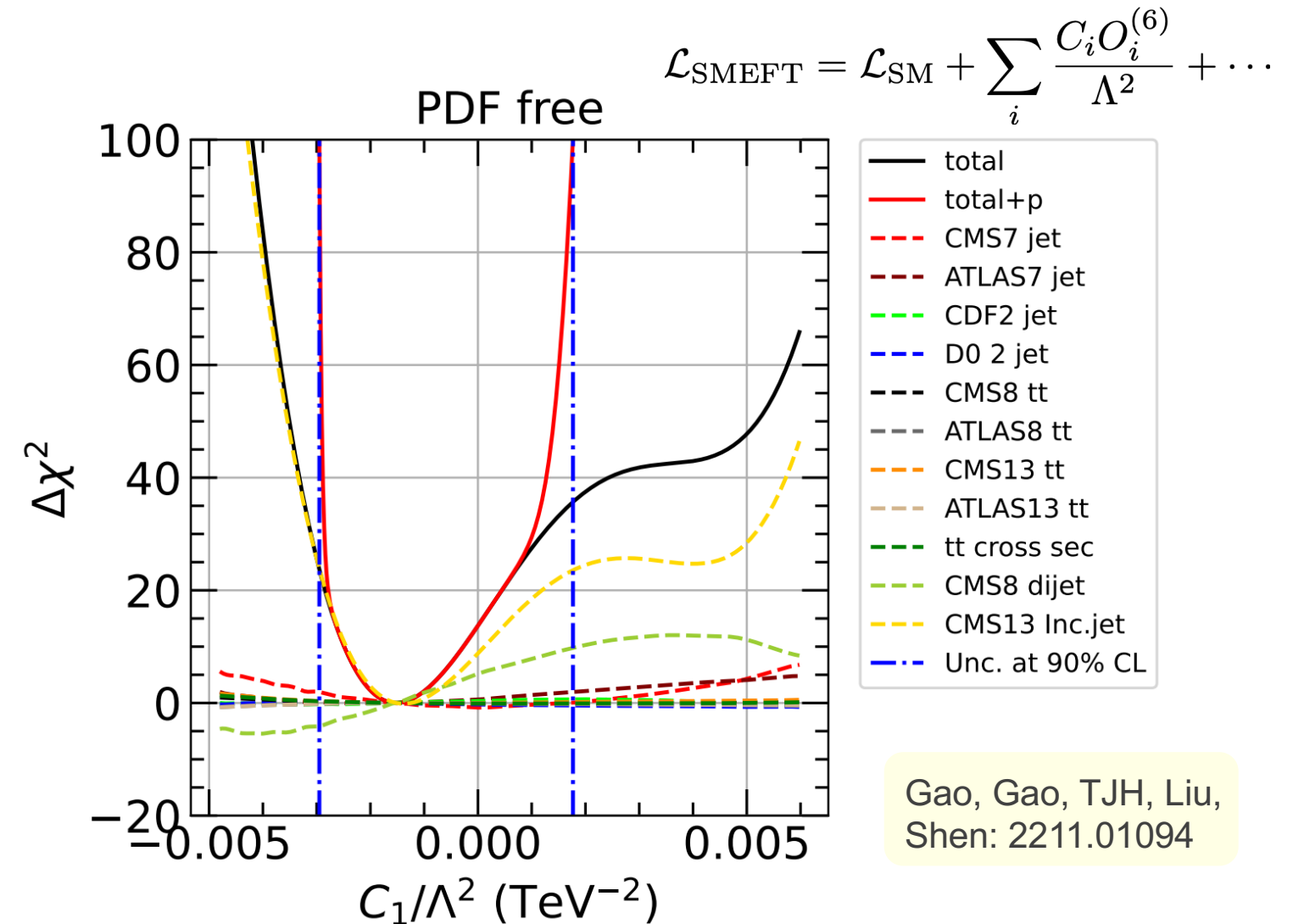


- BSM reach not only limited by PDF uncertainties: hadronic data (**esp., LHC**) used in New Physics searches also constrain PDF shape parameters!



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

[PDF uncertainties and BSM effects coupled: *e.g.*, high x]

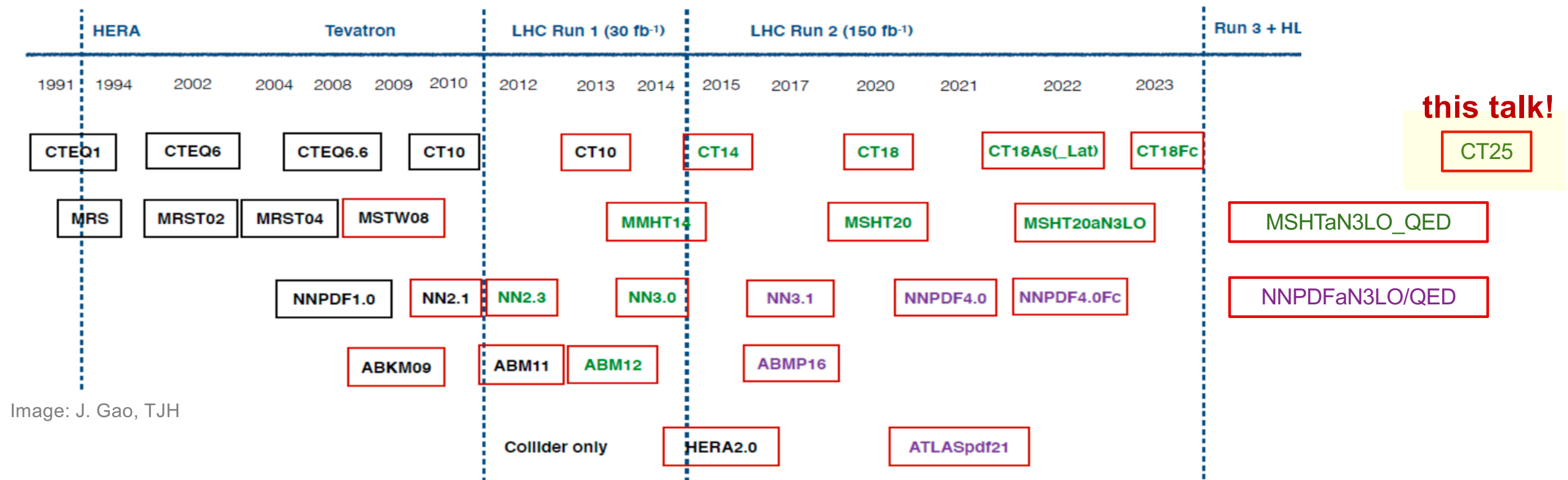


→ joint fits with PDFs (high- x gluon): **jets** accommodate wider contact interaction ranges (C_1/Λ^2) relative to fixed extractions

HEP PDF landscape has evolved steadily

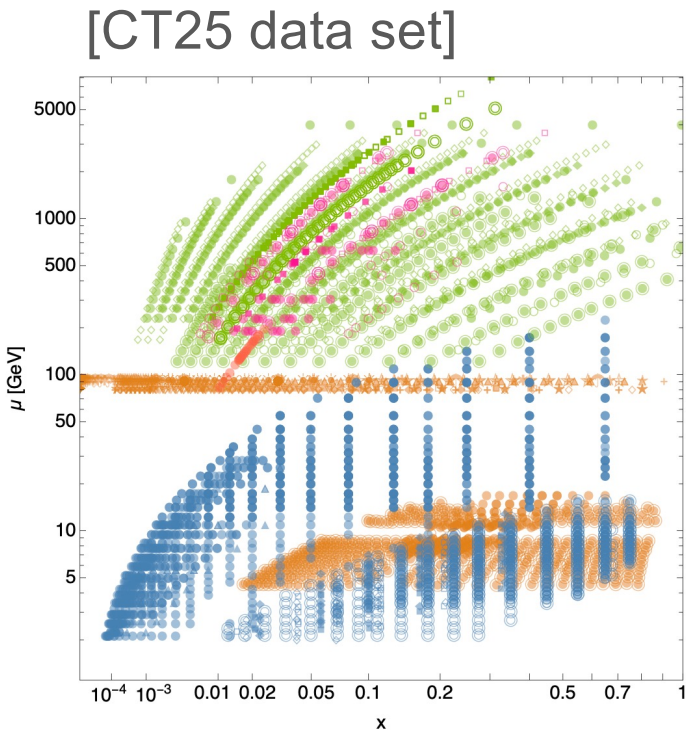
- PDFs as precision engines: global fits have evolved from ~order-of-magnitude explorations to few percent-level frameworks

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



require consistent accuracy for many processes, partonic interactions

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



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

Gao, Harland-Lang, Rojo: PR742 (2018) 1-121

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

recall: perturbative QCD ingredients entering PDFs



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

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

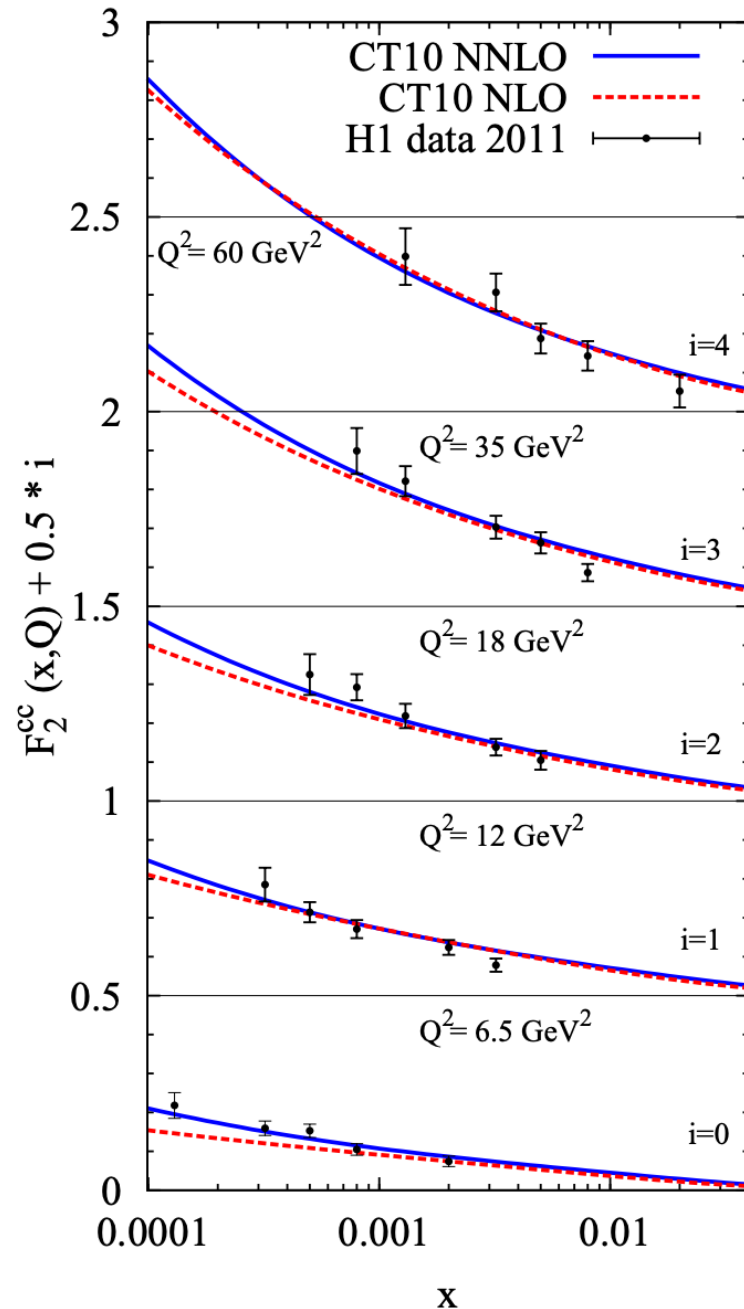
→ perturbative accuracy enters multiple sectors of the calculation:

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

[DGLAP splitting/kernels]
$$Q^2 \frac{\partial}{\partial Q^2} f_a(x, Q^2) = \sum_b P_{ab}(x, \alpha_s(Q^2)) \otimes f_b(x, Q^2)$$
$$P_{ab}(x, \alpha_s(Q^2)) = \sum_{n=0}^{\infty} \left(\frac{\alpha_s(Q^2)}{2\pi} \right)^{n+1} P_{ab}^{(n)}(x)$$

[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 are still emerging (*i.e.*, missing pieces remain outstanding)

- one scenario: *approximate* N3LO (aN3LO)

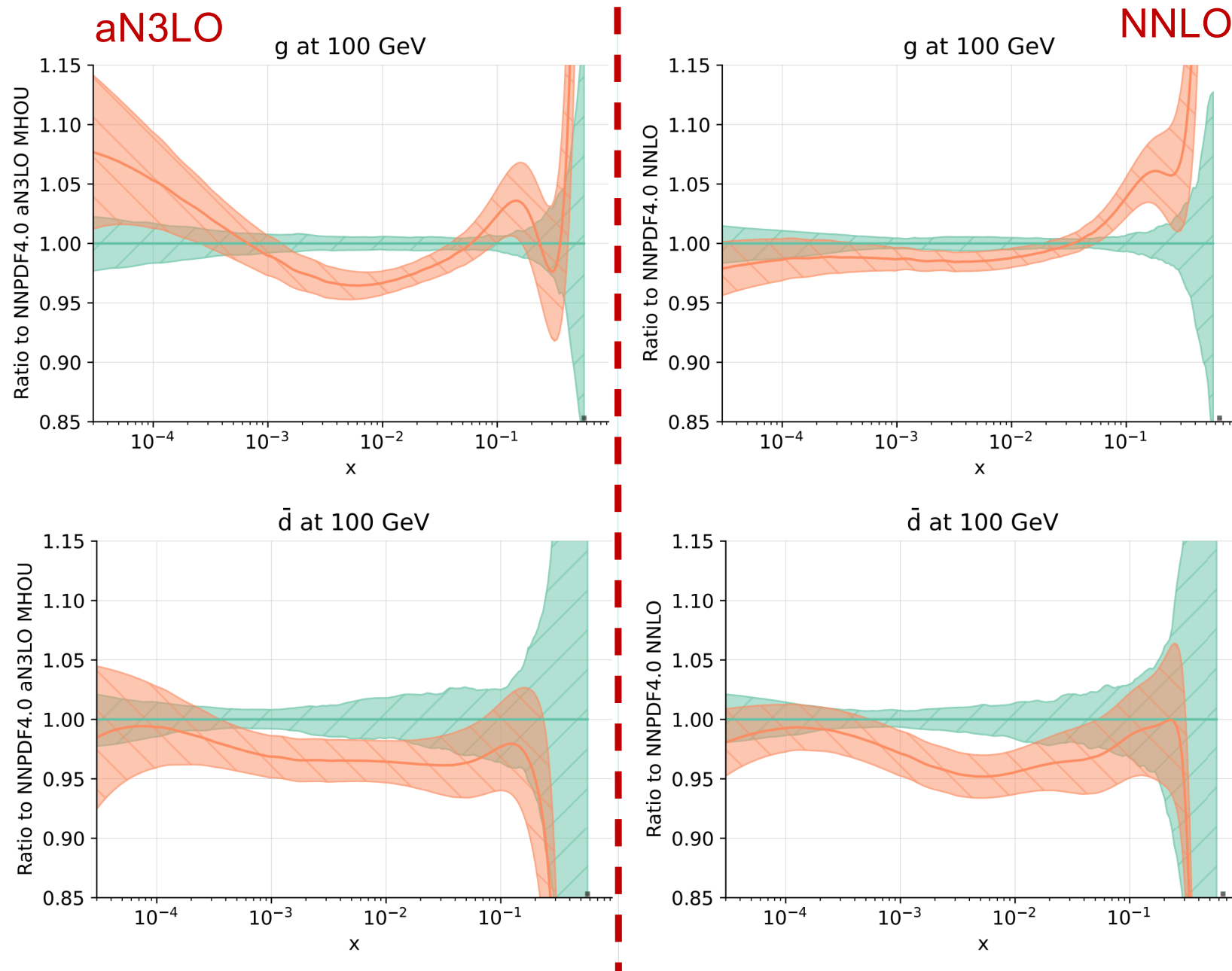
$$P_{ij}^{(3)}(x) = P_{ij}^{(3), \text{known}}(x) + \rho_{ij} P_{ij}^{(3), \text{approx}}(x)$$

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

- 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

[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 $pp \rightarrow B + X$, $\sigma_B^{N3LO} = \sigma_B^{NNLO} + \Delta^{N3LO} \sigma_B$, where

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

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

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

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

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

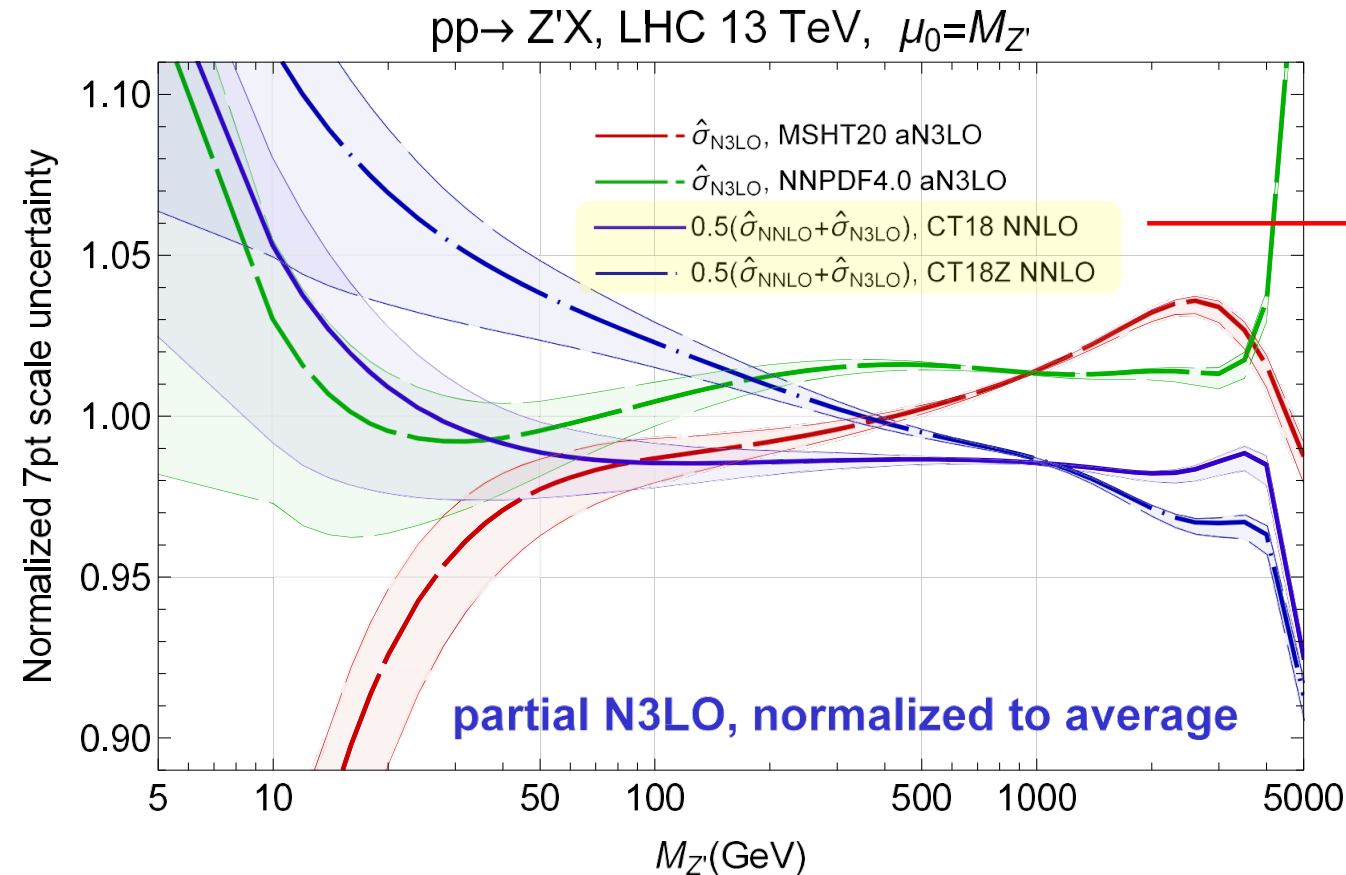
- prescription is then **epistemic estimator** on $\hat{\sigma}^{(3)} \in [0, \hat{\sigma}_{true}^{(3)}]$, where we take $\hat{\sigma}^{(3)} = \left(\frac{1}{2}\right) \cdot \hat{\sigma}_{true}^{(3)}$

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

total cross sections: testbed for pN3LO implementations

[TJH, Nadolsky, Ponce-Chavez: to appear.]

- 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, ...

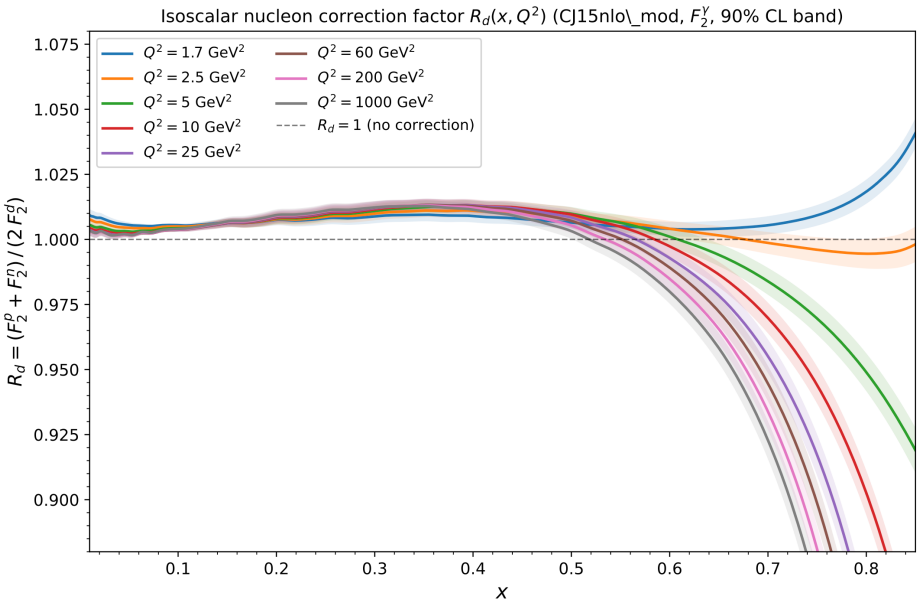
➔ these simply do not automatically decrease at NNLO+

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}	$\delta^{\text{NNLO-X}}$	$\delta^{\text{fixed d.c.}}$
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

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

(example) 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

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

DIS in CT25 at pN3LO

gg luminosity

CT25pN3LOprel: DIS is @pN3LO.

All other CT25 data sets are @NNLO.

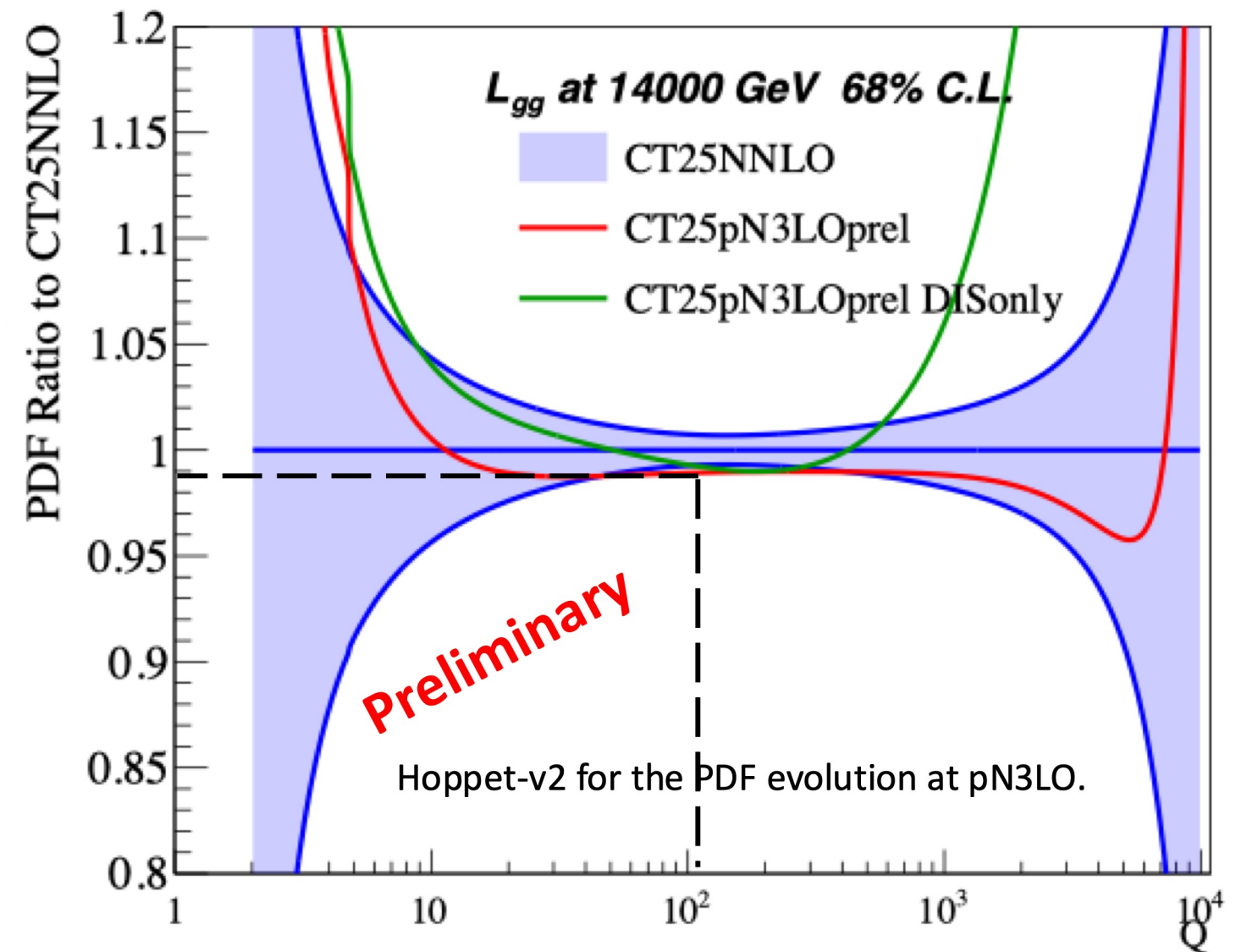
New LHC data sets in the CT25 NNLO baseline reduce gg luminosity by about 2.5% compared to CT18 NNLO, at Higgs mass.

Adding N3LO NC DIS in CT25N3LODISprel reduces $L_{gg}(M_H)$ by $\approx 1.2\%$, for a total reduction $\approx 4\%$

IHO/MHO = incomplete/missing higher-order uncertainties.

The pN3LOprel fit has IHO of order N3LO -- not small.

Increase in the gluon at small x is *predicted* by the DIS saturation scale in CT18X



NNLO, HERA DIS DATA; # of pts = 1120 ; $\chi^2 = 1401$ $\chi^2/N_{pt} = 1.251$

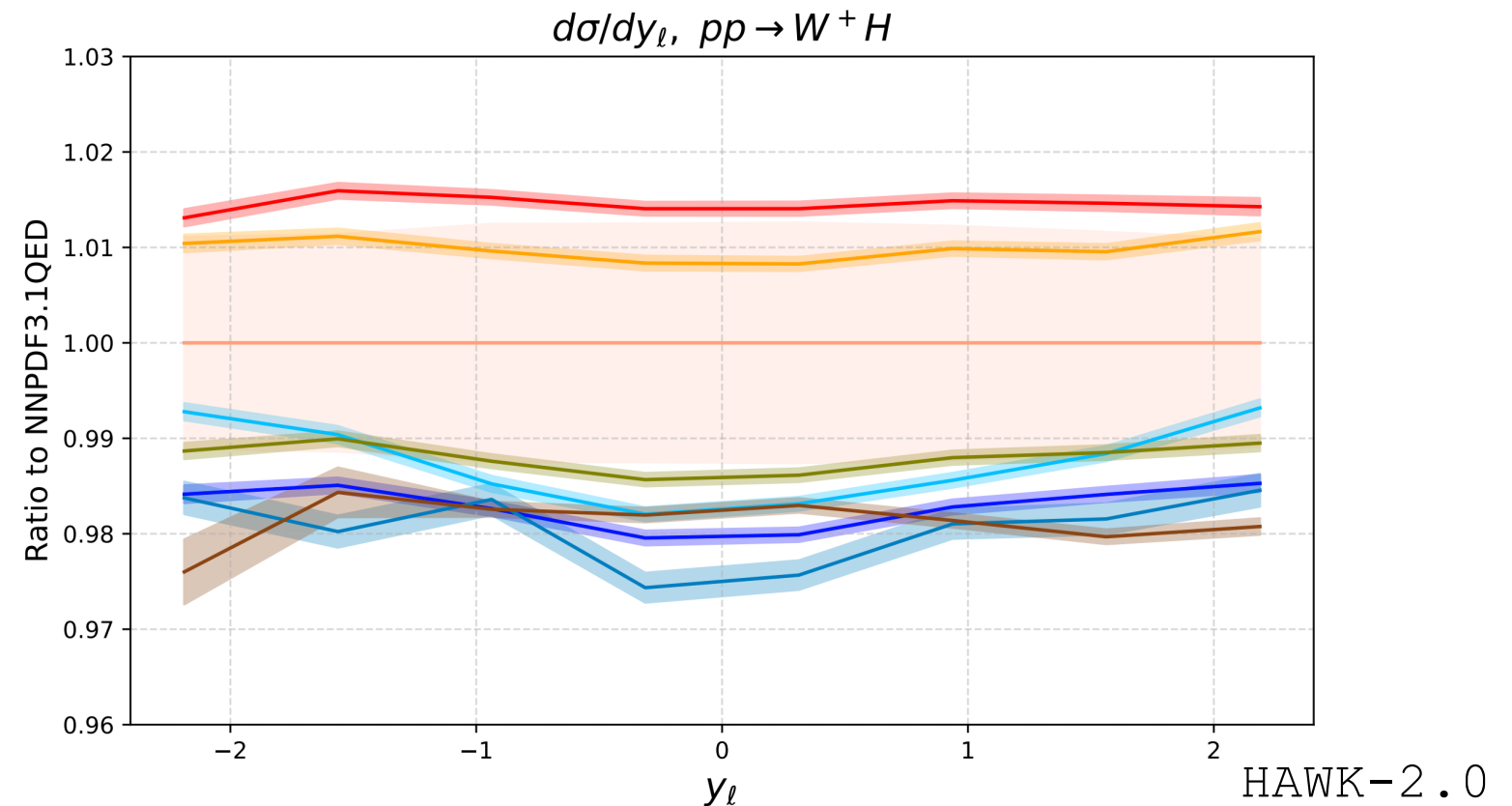
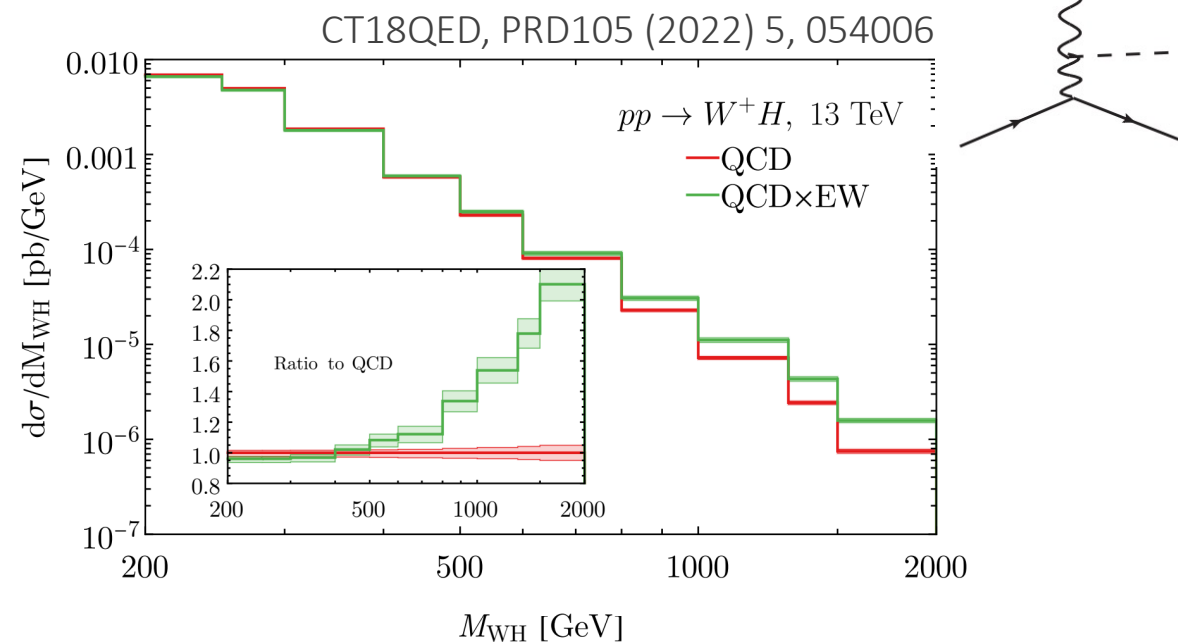
N3LO, HERA DIS DATA; # of pts = 1120 ; $\chi^2 = 1399$ $\chi^2/N_{pt} = 1.250$

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)

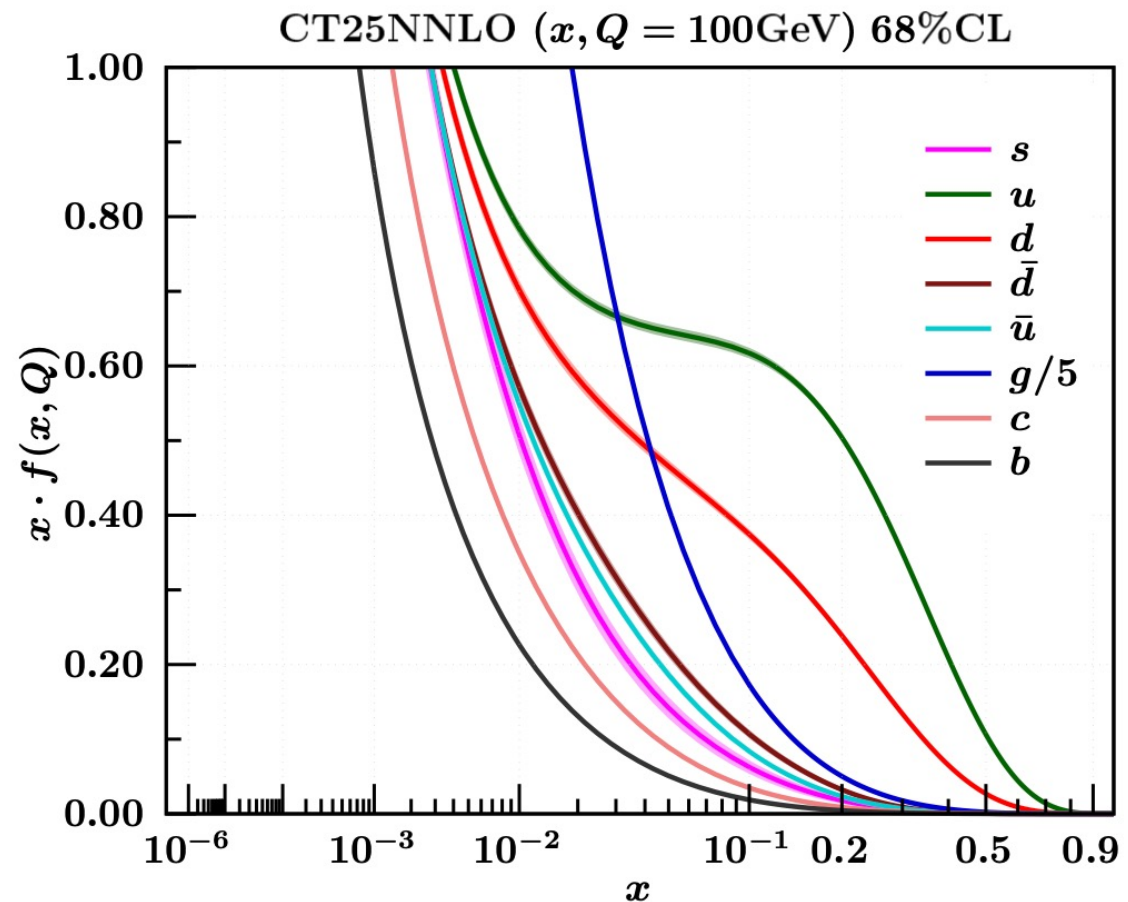
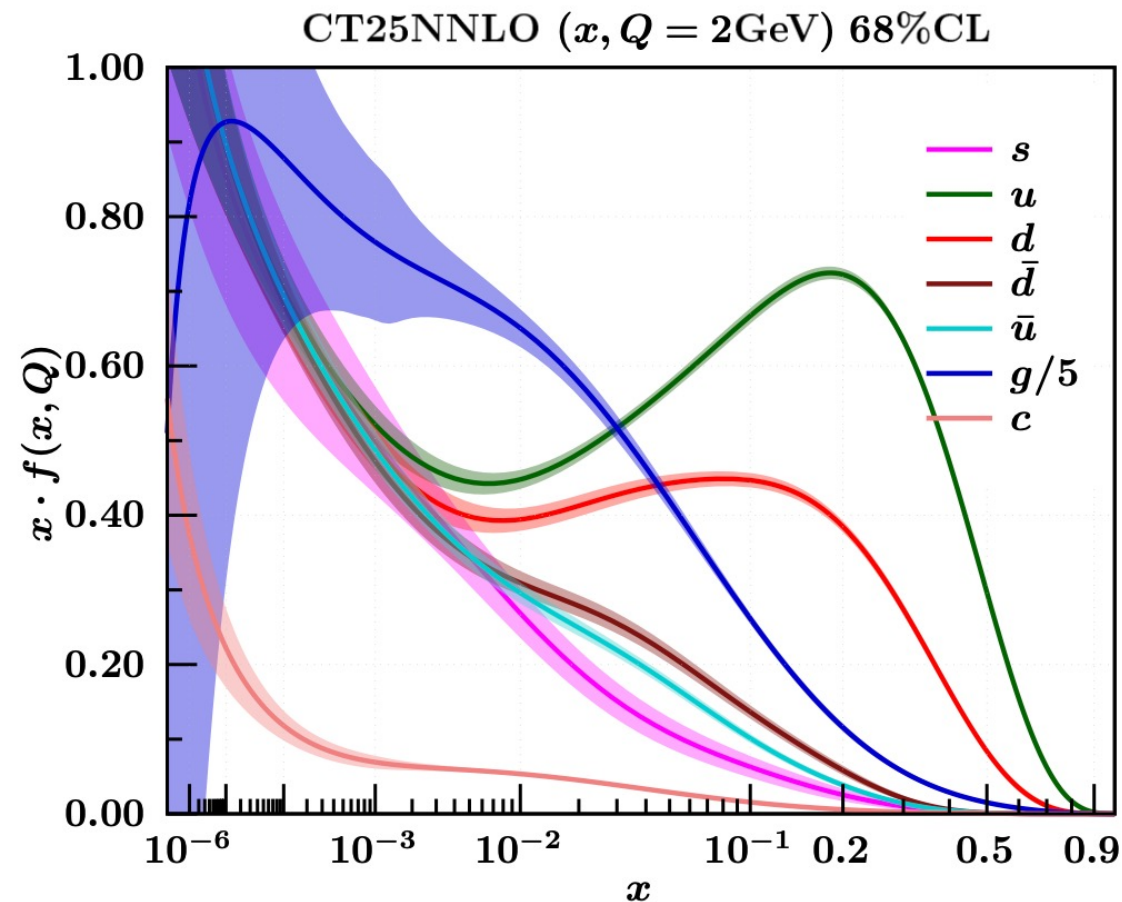


→ modern QED PDFs produce ~2-4% in VH distributions

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

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

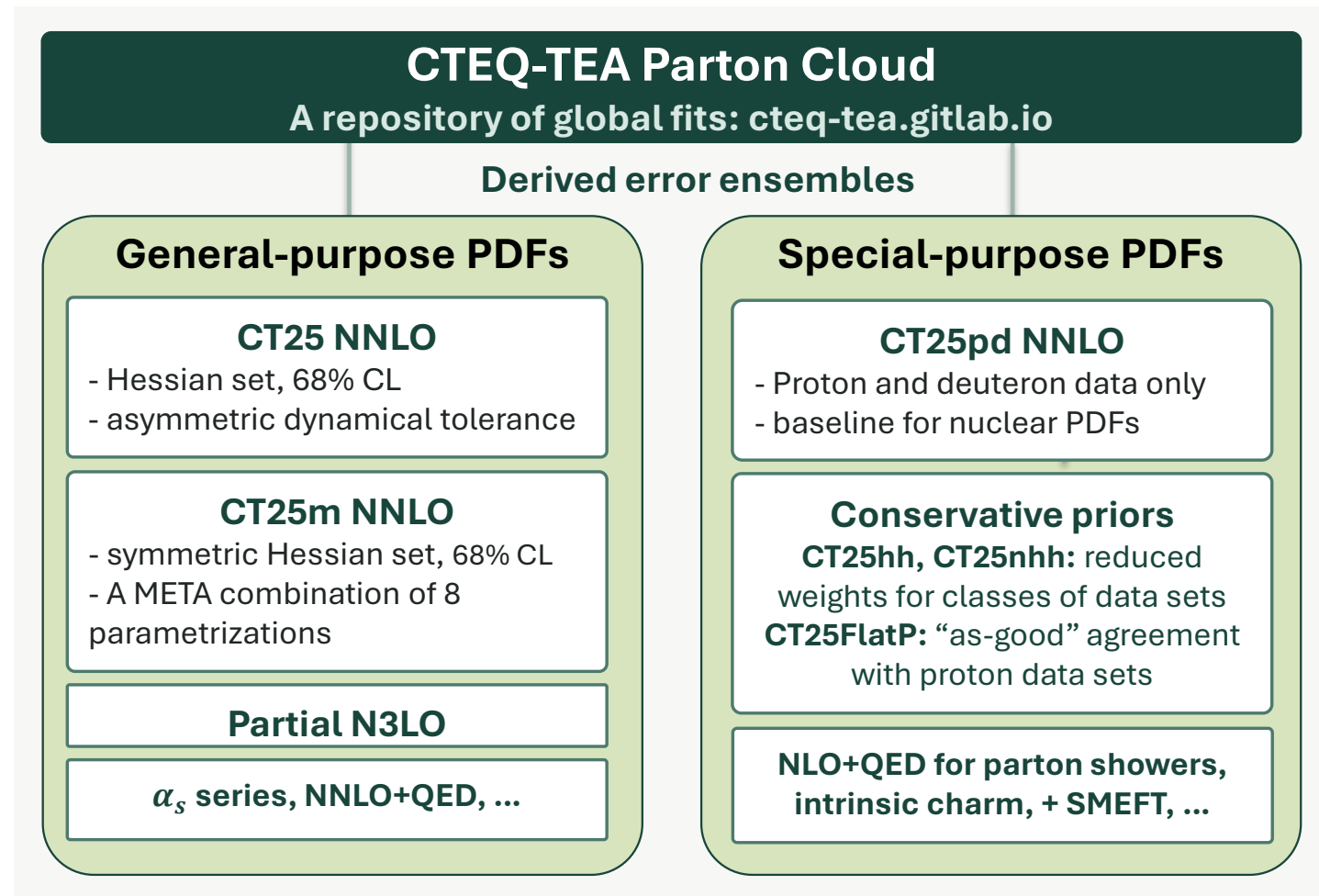
- CT25 global analysis PDF grids now public, [2512.19779](#) (includes corresponding series in α_s and n_F)
- forthcoming 'long paper': systematic exploration of fit variations; phenomenological implications



- 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); ...

CT25 is a Parton Cloud, not a single PDF set

- ❑ CT25 rests on a repository of **400+ (pN)NNLO global fits**: varied parametrizations, data baselines, methodology
 - public releases are derived error ensembles — one underlying cloud, each ensemble purpose-built for its use



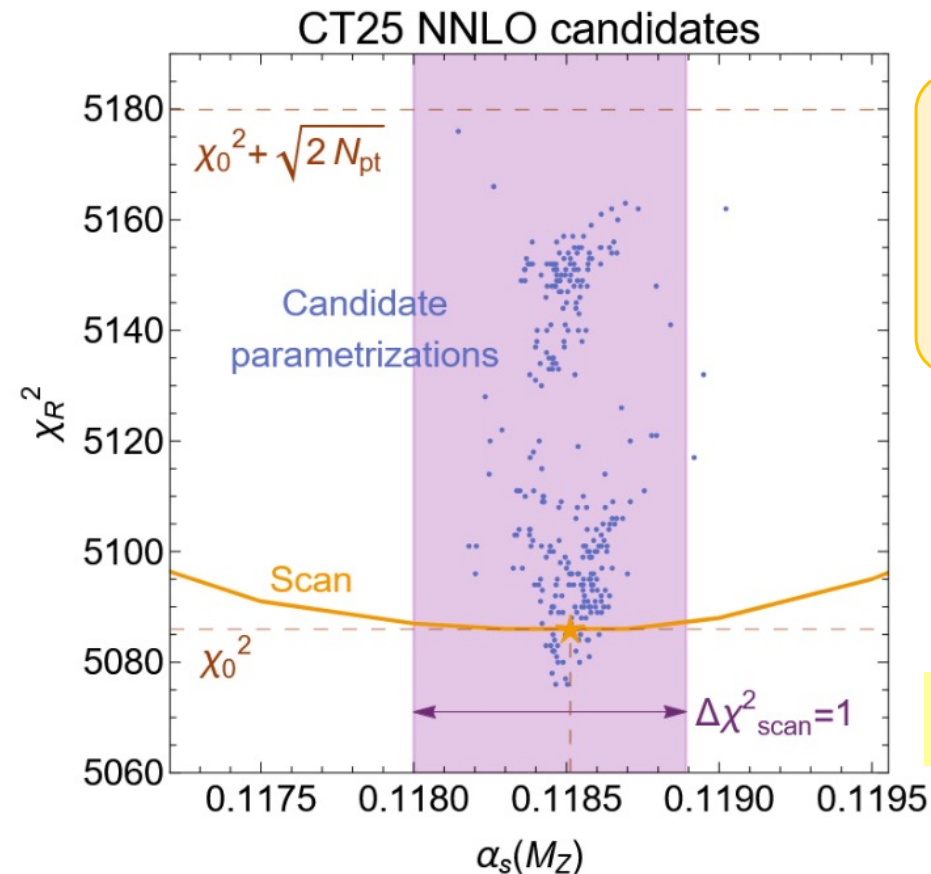
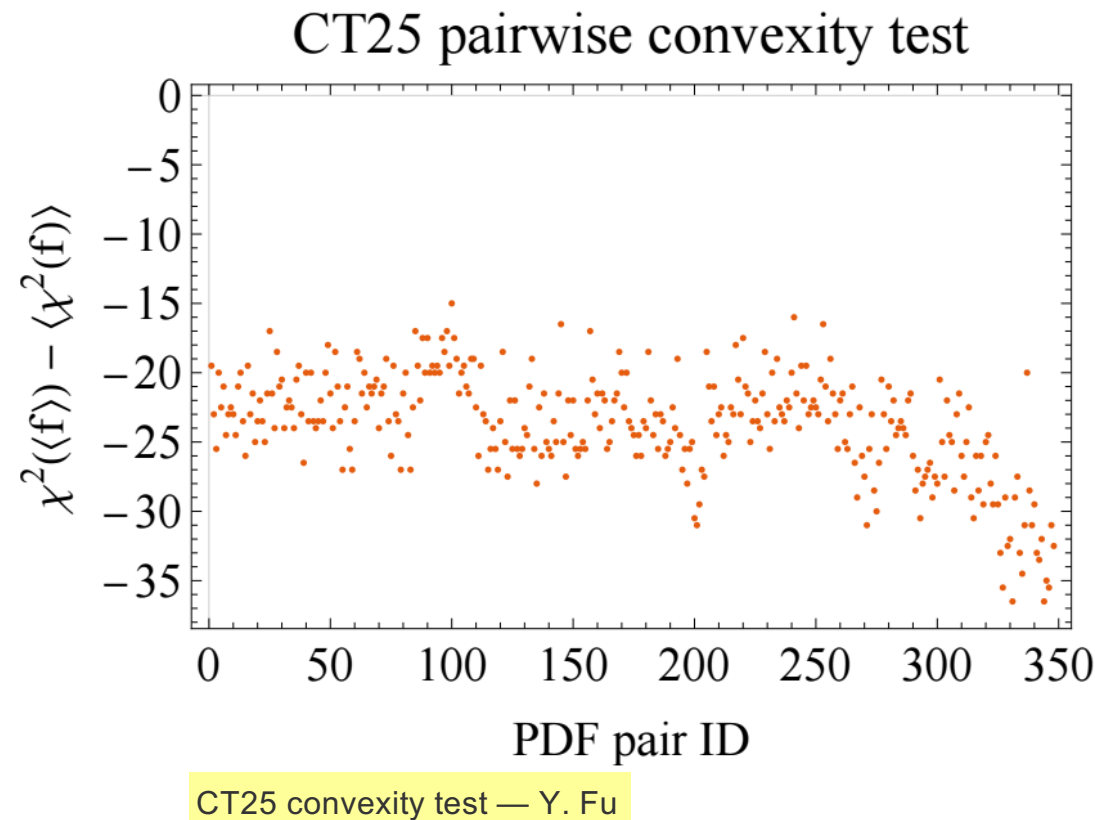
Parton Cloud preprint, this week

→ PDF set → sampled population of fits; the UQ question becomes which subset answers which question

why the Parton Cloud? convexity of χ^2 enables **scaled-up** fit generation ★

□ total χ^2 is convex across pairs of acceptable fits: $\chi^2(\langle f \rangle) - \langle \chi^2(f) \rangle < 0$ for all ~350 CT25 pairs (left)

→ strong Gaussianity of the likelihood → fits over many parametrizations can be generated and combined at scale



α_s scan over the population of fits, not a single curve

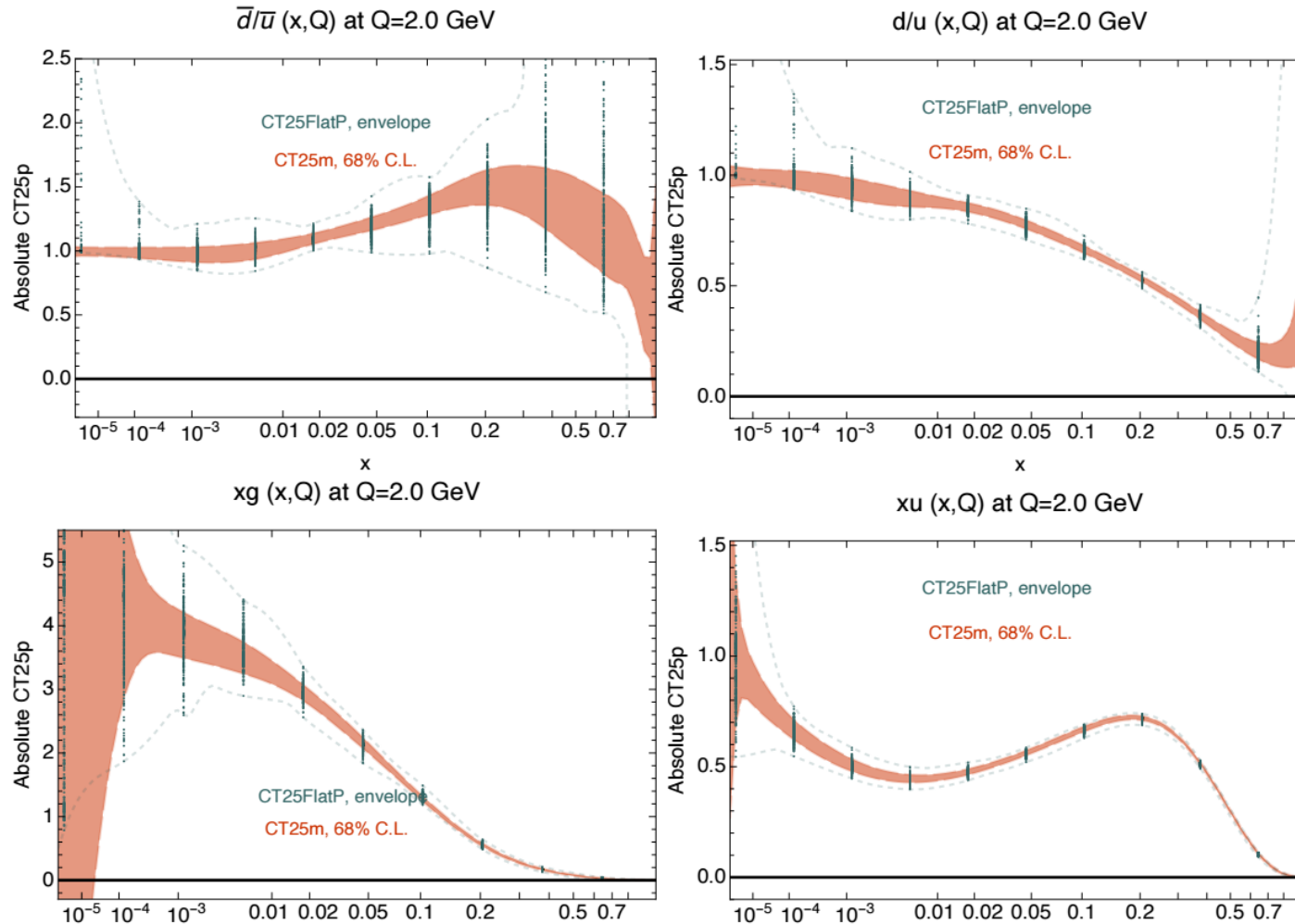
Ablat et al., 2512.23792
(287 alternative parametrizations)

□ at fixed α_s , χ^2 spread among **acceptable parametrizations** (right) $\gg \Delta\chi^2 = 1 \rightarrow$ parametrization is an epistemic uncertainty to be sampled, not fixed

→ basis for CT dynamical tolerance and for combining ensembles (CT25m); cf. parametrization UQ [Courtoy et al., 2205.10444]

Parton Cloud → pheno application: **purpose-built** error ensembles

□ different questions draw **different ensembles** from the Parton Cloud — uncertainty follows use case



CT25FlatP envelope vs. CT25m 68% CL

- CT25 / CT25m: general-purpose Hessian sets; CT25m META-combines 8 D-TOL ensembles over distinct parametrizations (PDF4LHC21-like)
- CT25FlatP: 350 proton-only fits with a flat prior, error = envelope → for LHC reweighting without double-counting; directly relevant to PDF-SMEFT entanglement (earlier)
- CT25pd: proton + deuteron data with CJ-CT nuclear correction → baseline for nuclear PDFs; deuteron effects rival N3LO (earlier)
- envelope vs. Hessian (left): FlatP envelope generally wider — aleatoric + epistemic (parametrization, data selection) pieces made explicit; e.g., $\bar{d}/\bar{u}$ at $x \gtrsim 0.1$

CT25 ensembles: a sampled model space rather than one tolerance — the UQ frontier revisited later (Fantômas, ML)

→ next: the LHC data that populate the Cloud

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

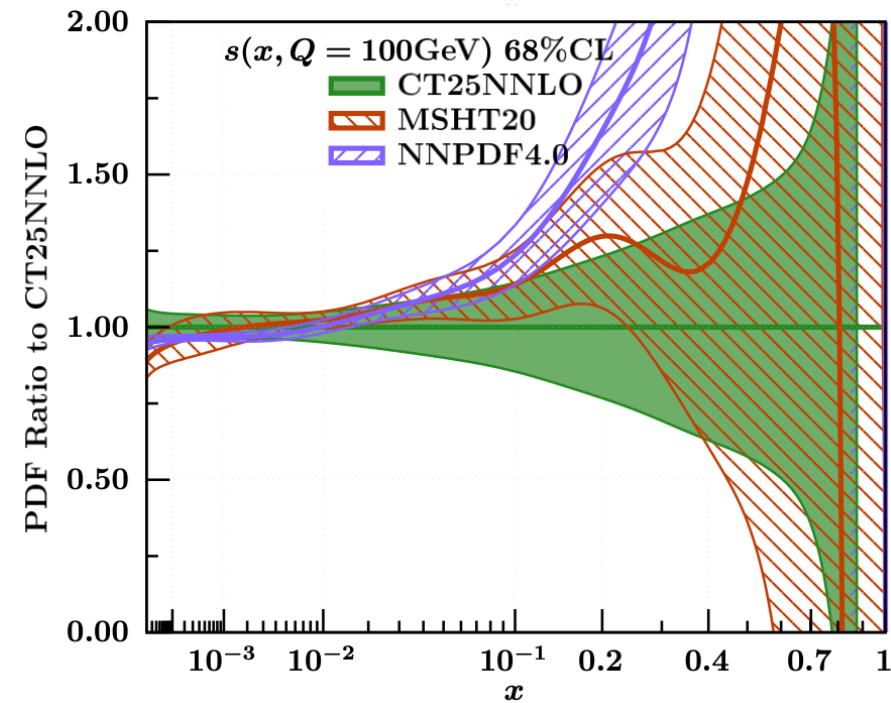
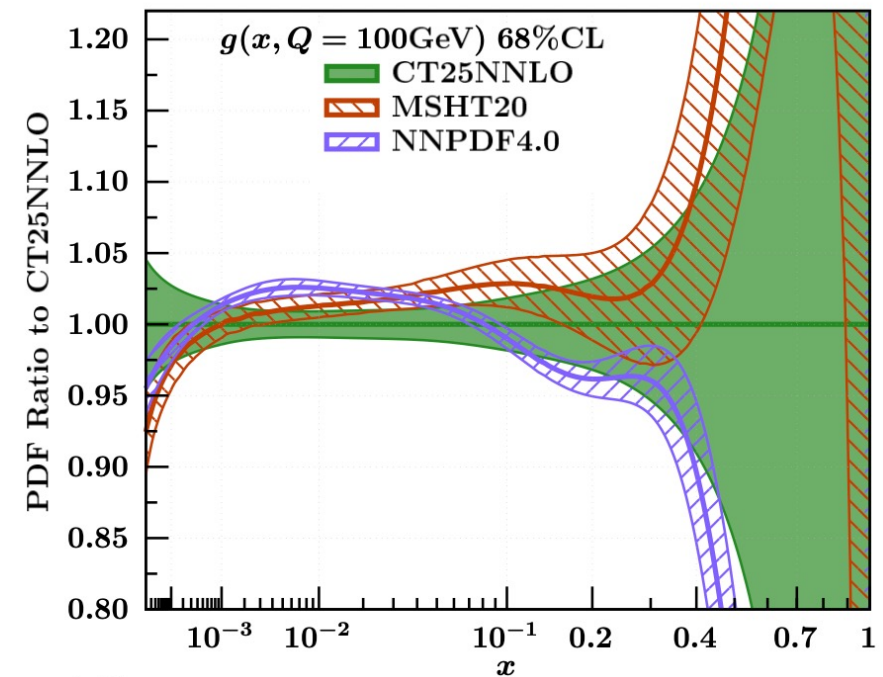
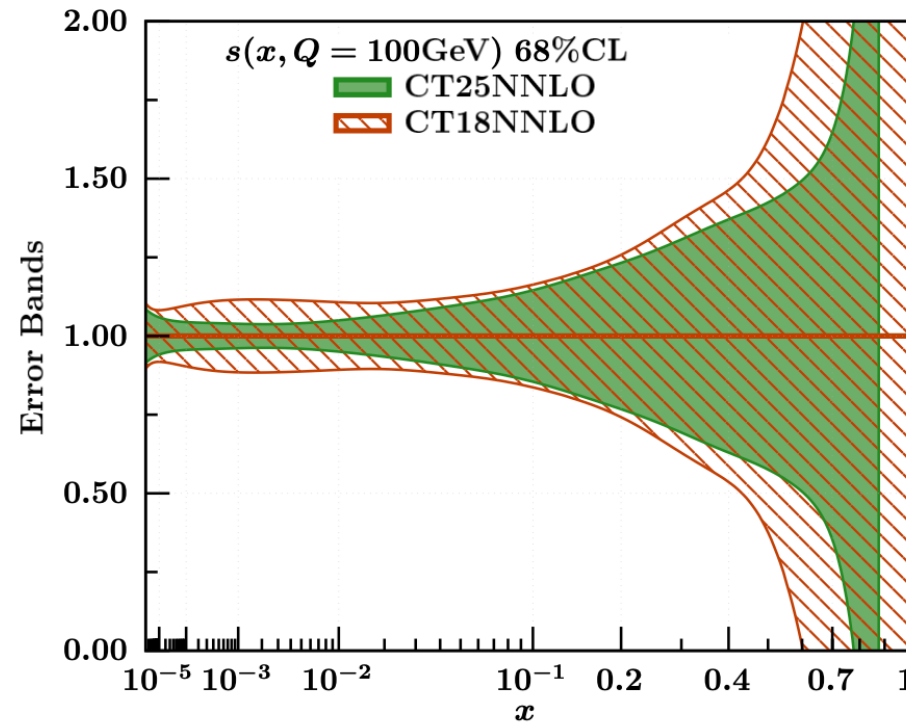
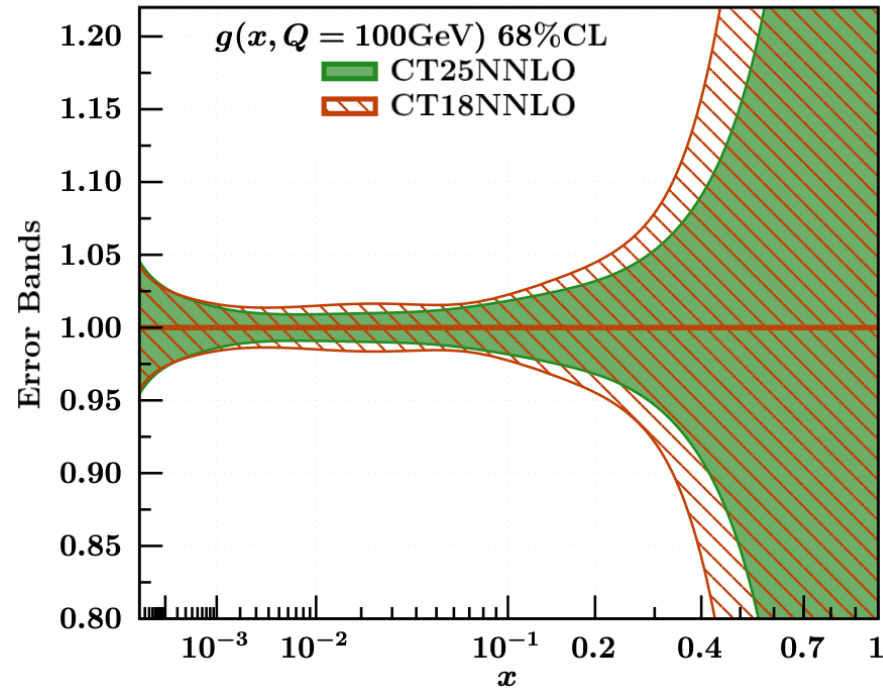
(new data span several LHC processes)

□ 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 : $p_T^{\ell,\nu_l} > 25 \text{ GeV}$, $ \eta_\ell < 2.5$, $m_T > 40 \text{ GeV}$; [11]	27	16.8	0.6	-1.5
211	ATLAS 8 TeV W : 20.2 fb^{-1} , $p_T^\mu > 25 \text{ GeV}$, $p_T^\nu > 25 \text{ GeV}$, $ \eta_\mu < 2.5$, $m_T > 40 \text{ GeV}$ [12]	22	44.6	2.0	2.8
214	ATLAS 8 TeV Z 3D: 20.2 fb^{-1} , $p_T^\ell > 20 \text{ GeV}$, $ \eta_\ell < 2.4$, $46 < m_{\ell\ell} < 200 \text{ GeV}$ [13]	188	217.7	1.2	1.5
212	CMS 13 TeV Z : 35.9 fb^{-1} , $\frac{d\sigma(Z)}{dy_Z}$, $p_T^\ell > 20 \text{ GeV}$, $ \eta_\ell < 2.4$, $ m_{\ell\ell} - M_Z < 15 \text{ GeV}$ [14]	12	24.2	2.0	2.1
217	LHCb 8 TeV W : 2.0 fb^{-1} , $W \rightarrow e\nu_e$, $p_T^e > 20 \text{ GeV}$, $2.00 < \eta < 4.25$ [15]	14	19.4	1.4	1.0
218	LHCb 13 TeV Z : 0.29 fb^{-1} $p_T^\ell > 20 \text{ GeV}$, $2.0 < \eta < 4.5$, $60 < m_{\ell\ell} < 120 \text{ GeV}$ [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 $\ell + j$ $y_{t\bar{t}} + m_{t\bar{t}} + y_{t\bar{t}}^B + H_T^{t\bar{t}}$: 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} , $\frac{d^2\sigma}{dp_T d y }$, $ y < 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} , $\frac{d^2\sigma}{dp_T d y }$, $ y < 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} , $\frac{d^2\sigma}{dp_T d y }$, $0 < y < 2.0$, $97 \leq p_T^{jet} \leq 3103 \text{ GeV}$, $R = 0.7$ [24]	78	88.6	1.1	0.9
556	CMS 7 TeV incl.jet: 5.0 fb^{-1} , $\frac{d^2\sigma}{dp_T d y }$, $0 < y < 2.5$, $114 \leq p_T \leq 1327 \text{ GeV}$, $R = 0.7$ [24]	118	87.3	0.7	-2.1
557	CMS 8 TeV incl.jet: 19.7 fb^{-1} , $\frac{d^2\sigma}{dp_T d y }$, $0 < y < 3.0$, $74 \leq p_T \leq 1784 \text{ GeV}$, $R = 0.7$ [24]	164	192.8	1.2	1.5

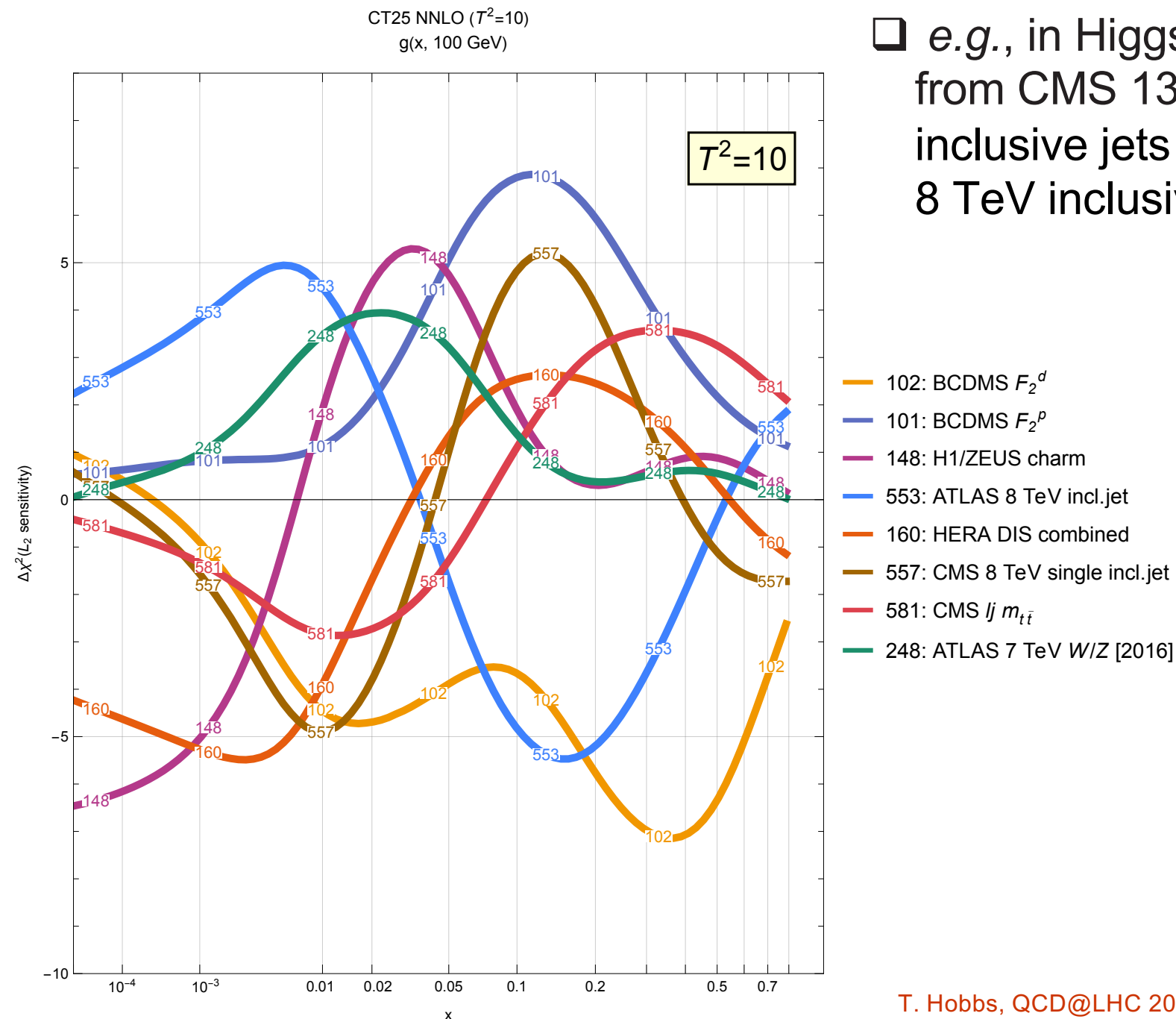
overview of CT25



- 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 CMS 8 TeV single-inclusive jets (557), but with a downward pull from ATLAS 8 TeV inclusive jet data (553)

- fixed-target DIS (e.g., BCDMS, 101/102) still have notable pulls at high x (> 0.1) as observed in CT18, as do 2016 ATLAS 7 TeV W/Z (248)

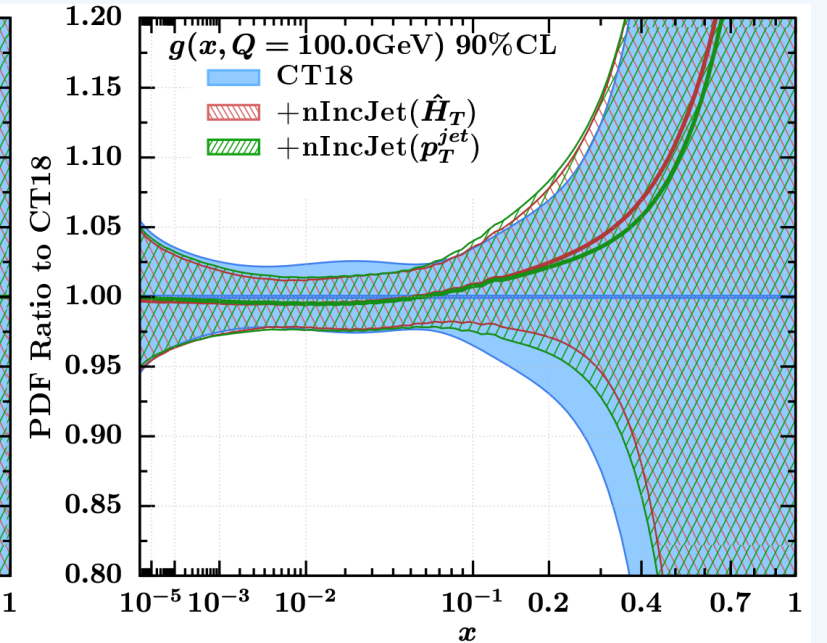
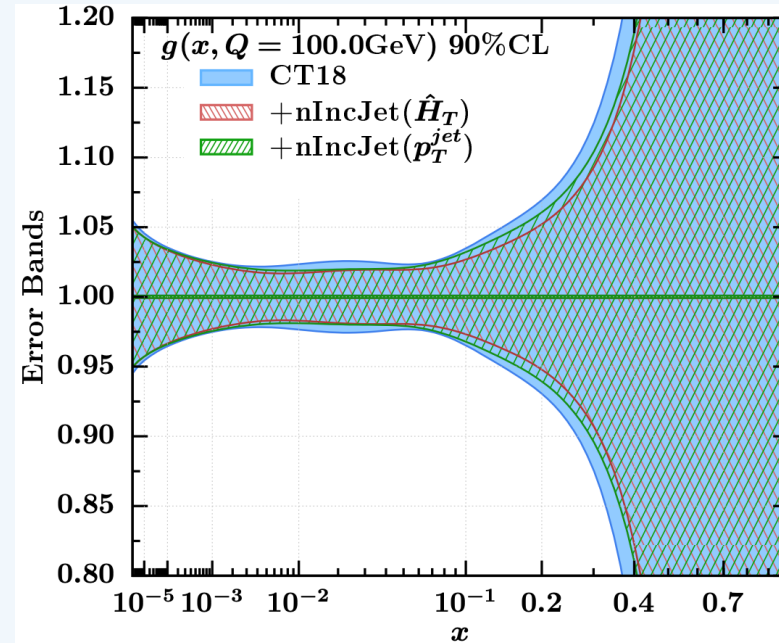
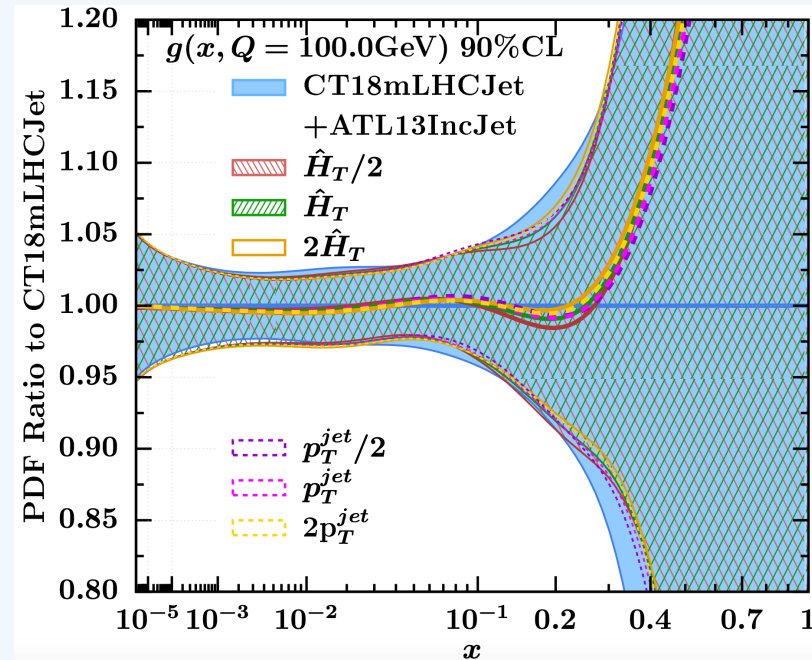
→ L_2 sensitivity method shown here originally introduced for PDF-Lattice synergy studies

[TJH, Wang, Nadolsky, Olness: 1904.00022]

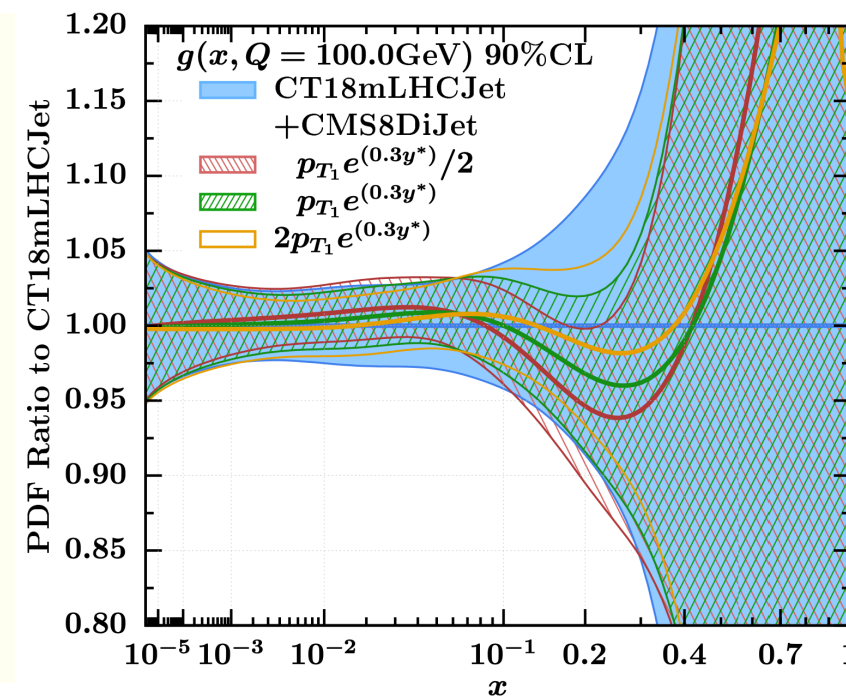
...complementary to XAIPDF and Shapley values methods.

inclusive jet vs. dijet data sets: scale dependence

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



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

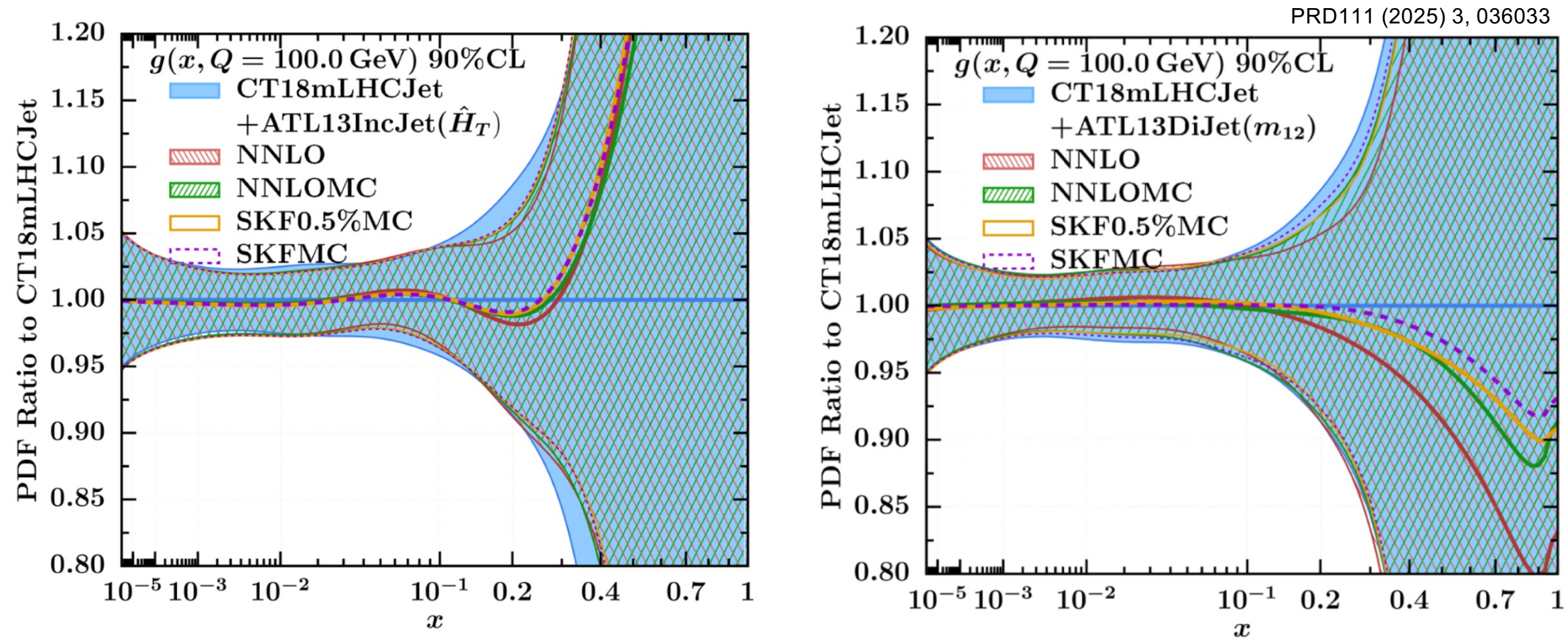


Ablat et al., PRD111 (2025) 3, 036033

- 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**

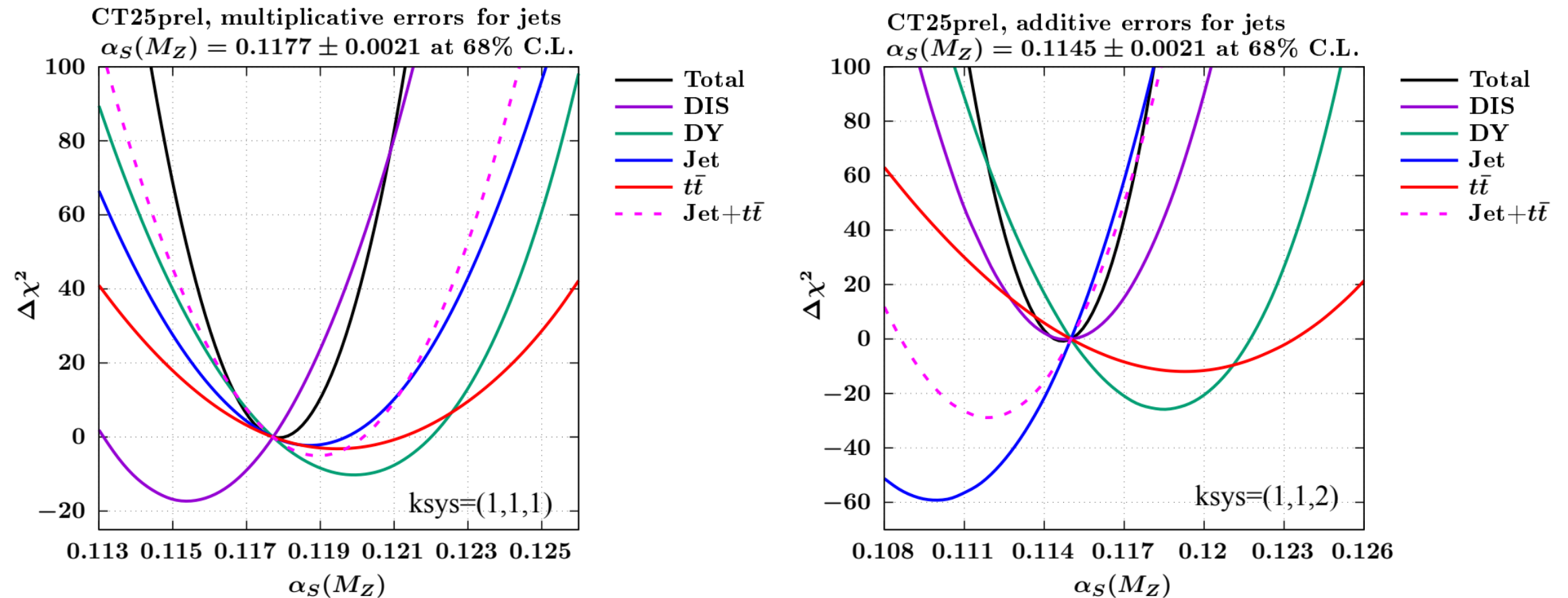
inclusive jet vs. dijet data sets: NNLO theory uncertainties

- 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]



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

- ❑ 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

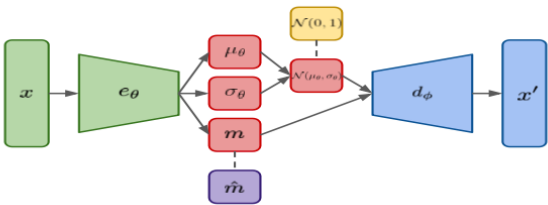


➤ treatment of jet systematics influential to pulls (de-correlation procedures; add./mult. prescriptions)

must explore alongside higher perturbative accuracy!

AI/ML approaches to PDF determination and interpretation

□ e.g., methods to refine PDFs via improvable (world) models [left]; connect theory to x behavior [right]



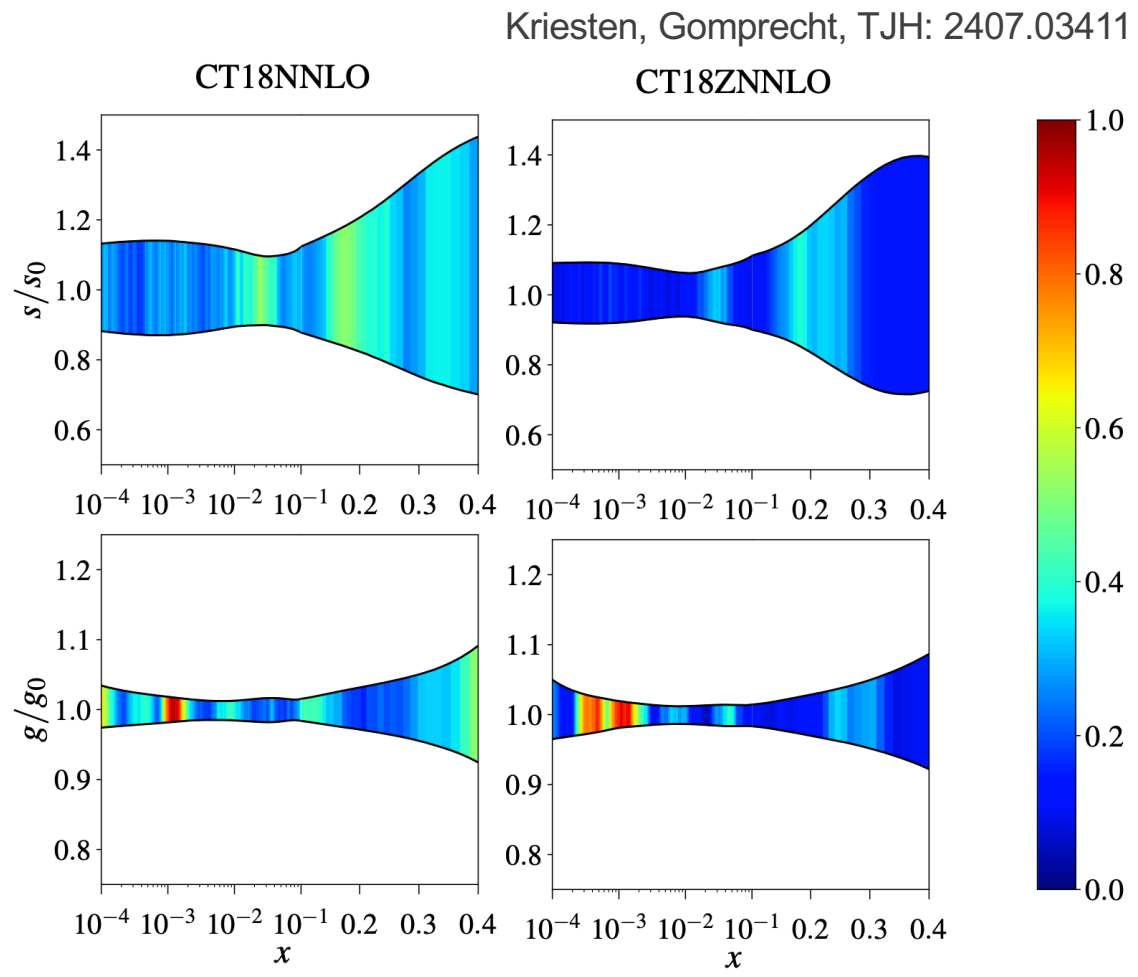
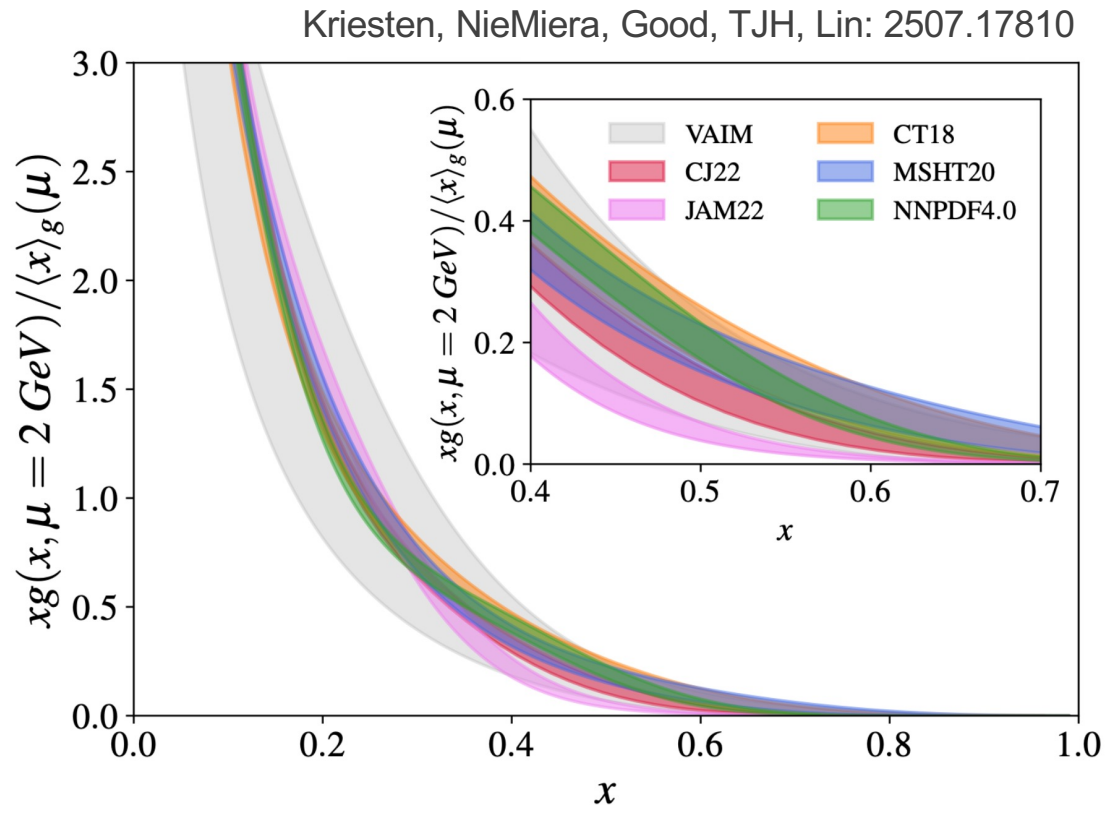
■ decoded gluon PDF from lattice QCD info (RpITDs) via **VAIM**; foundation model possibilities (MORPH)

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

[see talk, Thurs!]



Brandon Kriesten



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

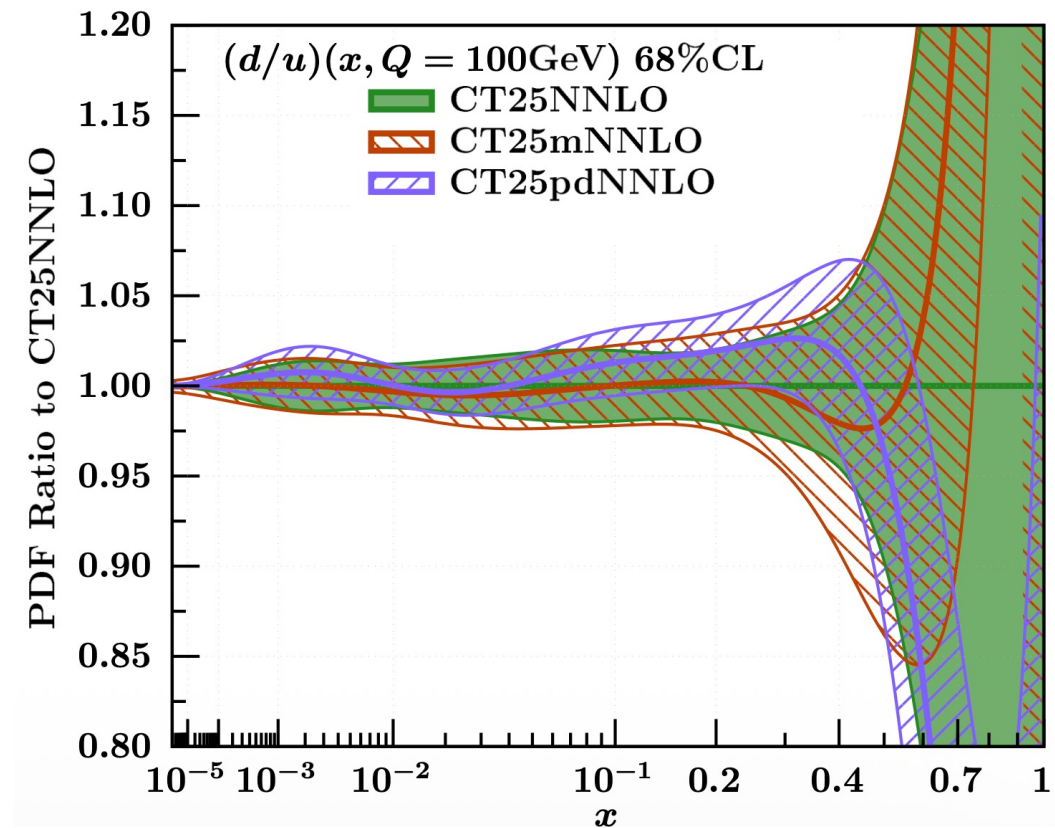
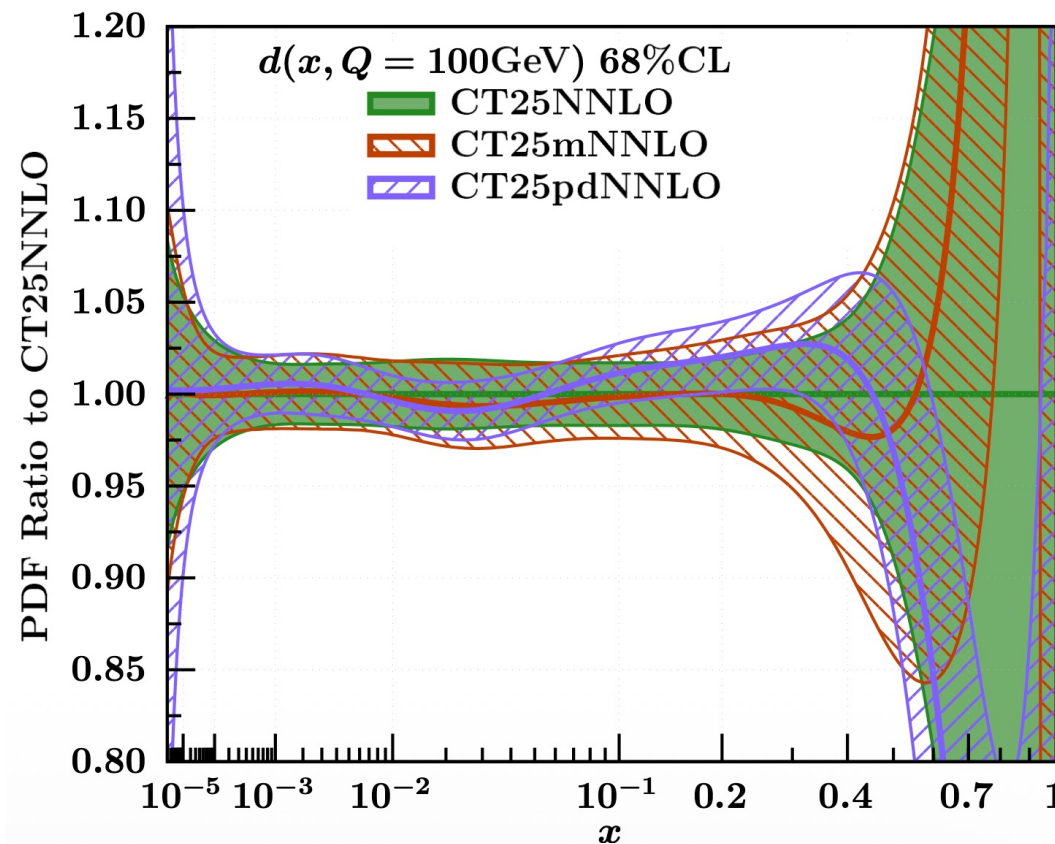
precision CT PDFs at the LHC: progress and outlook

- ❑ global PDF analyses: backbone of precision at the LHC; **progress hinges on consistent treatment of theory, data, and statistics**; **epistemic UQ** in CT Parton Cloud
- ❑ in HL-LHC era, sensitivity will require subtle interplay between PDFs and analysis choices
- ❑ CT25 fits: 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 is vital frontier with novel studies emerging (**new preprint this week**)
 - need further exploration of experimental cross talk (LHC EWK weighting studies; LHC-EIC; ...)
- ❑ AI/ML offers opportunities, challenges: novel PDF studies, reciprocal insights into statistical problems

Backup

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)

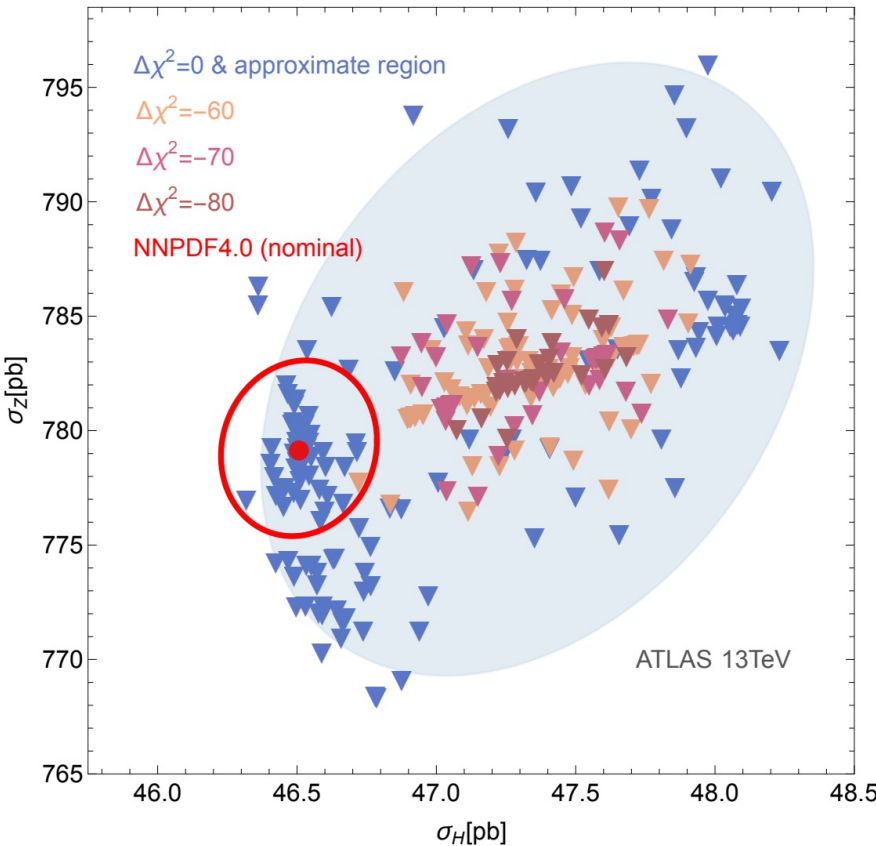
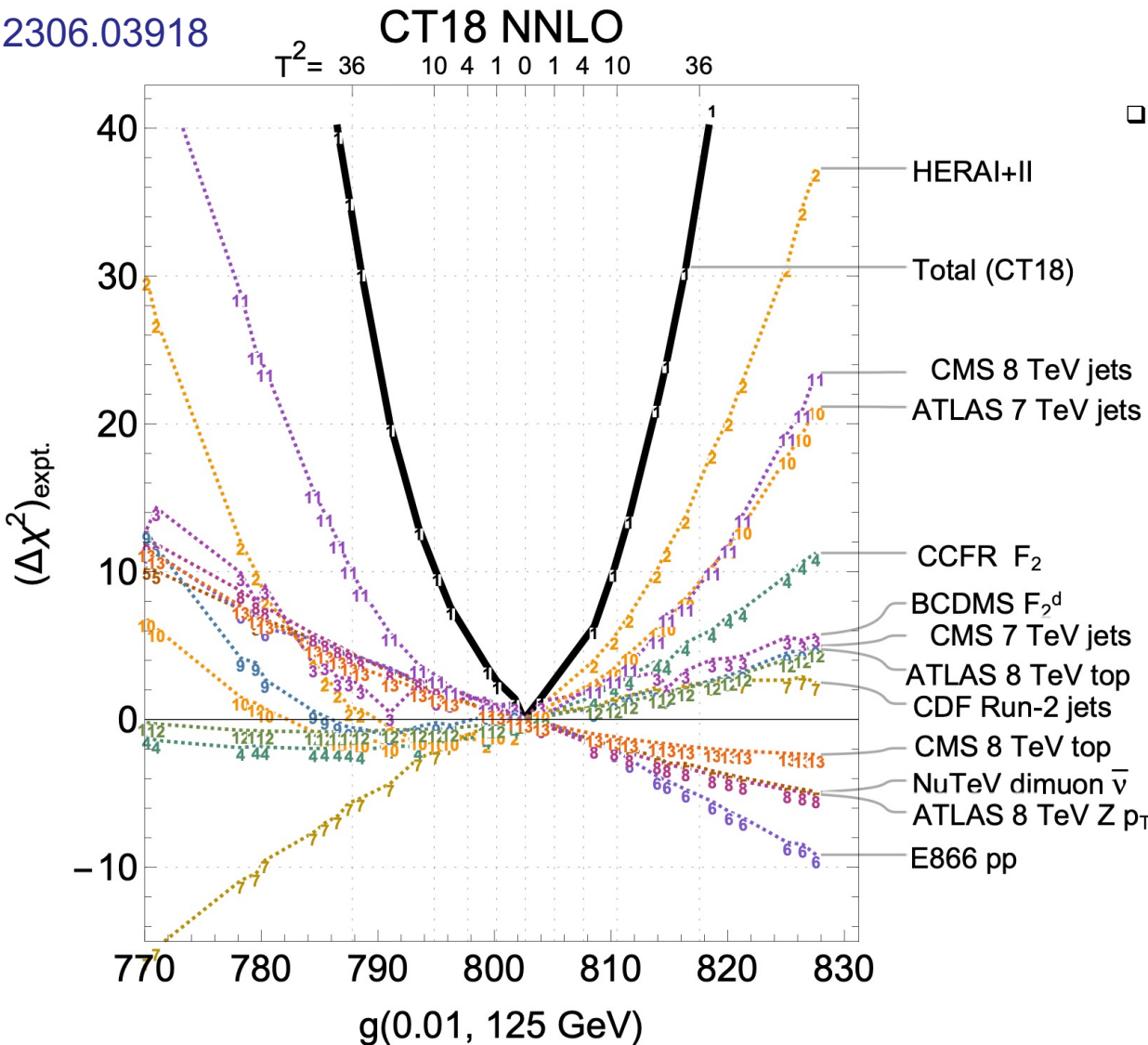


crucial frontier: reproducible, robust PDF uncertainties

- community-wide interest in quantifying PDF uncertainties

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

- need first-principles research in PDF uncertainties



Courtoy et al., 2205.10444

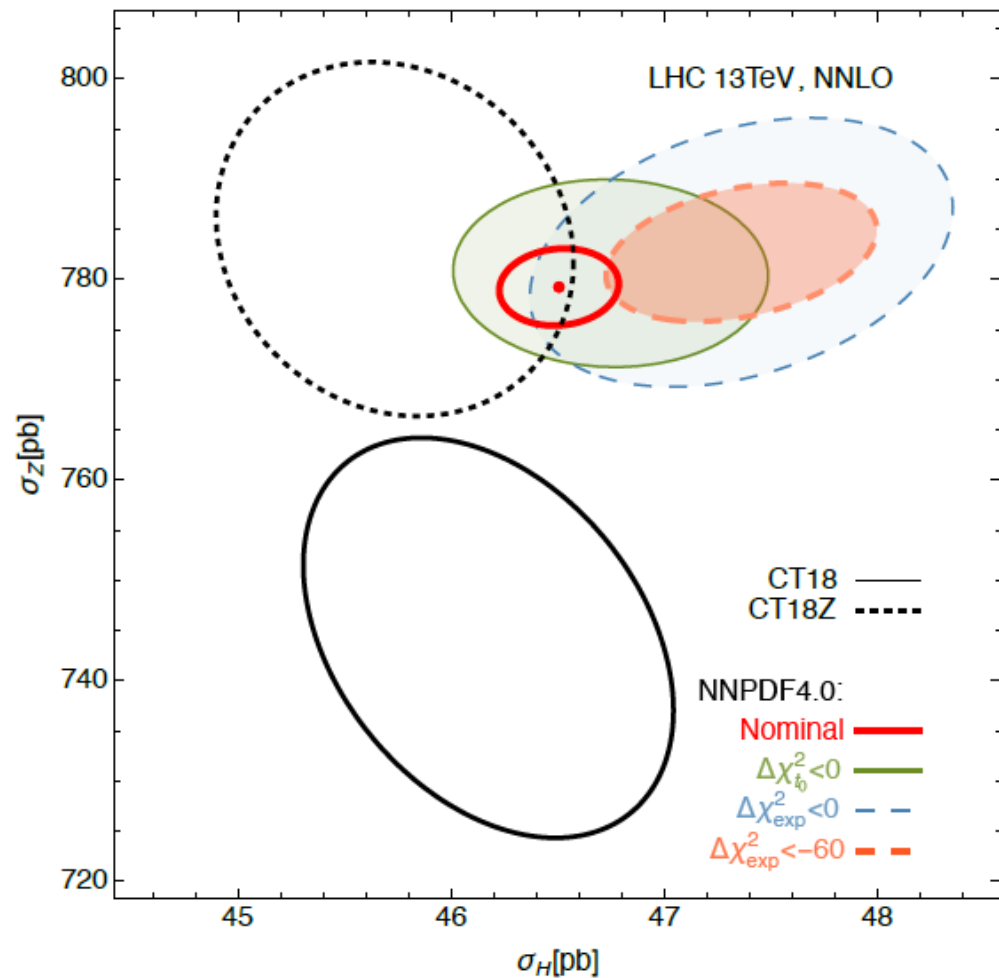
- ongoing investigation of relation(s) among resampling, Hessian, ML, ..., methods

uncertainty quantification (UQ) developments

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

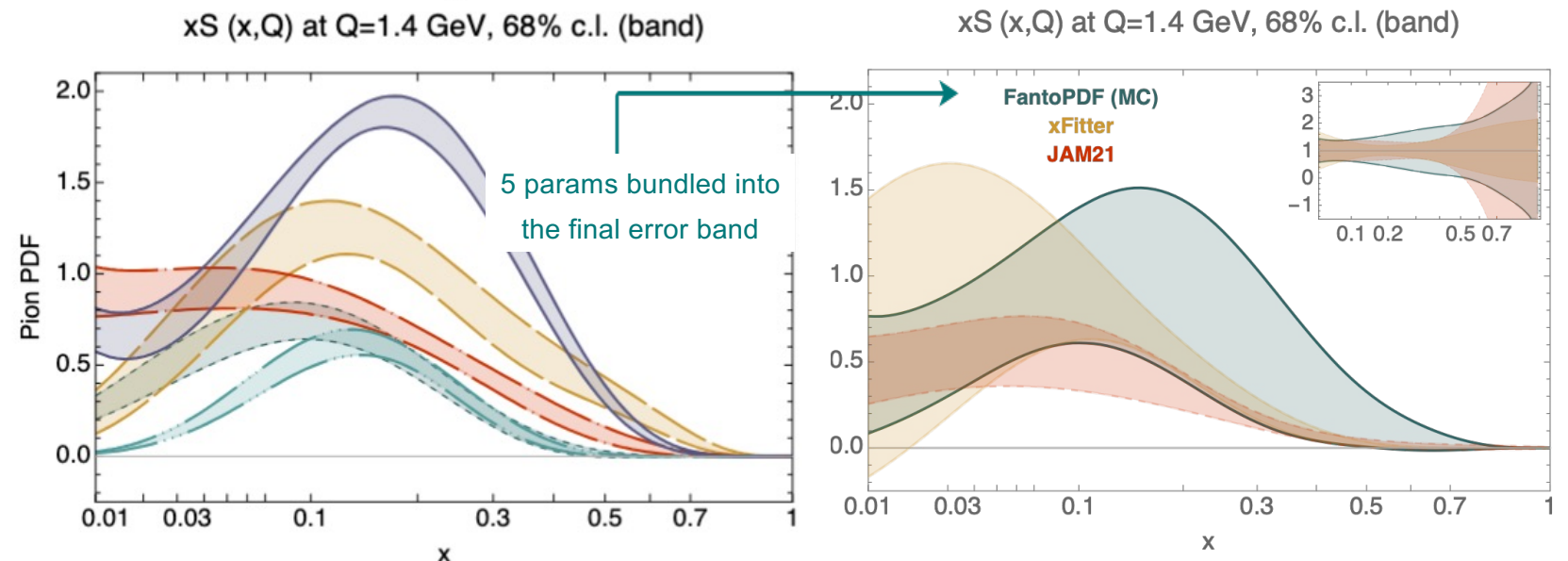
(essential for representative sampling over PDF models and fitting methodologies, robust pheno predictions)

[Courtoy et al., PRD107, [2205.10444](#)]



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

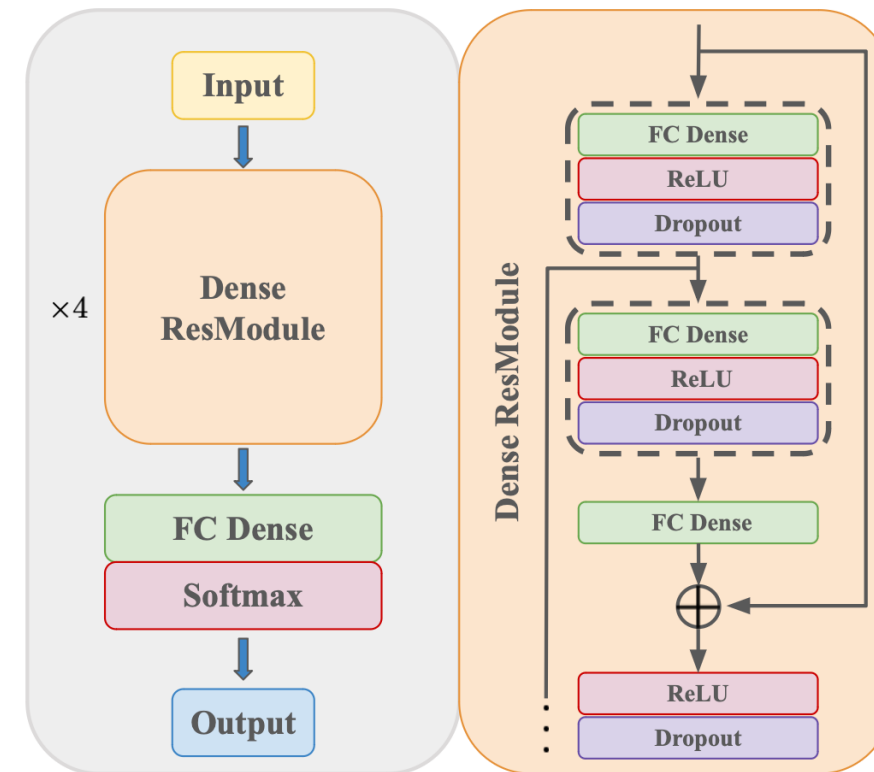
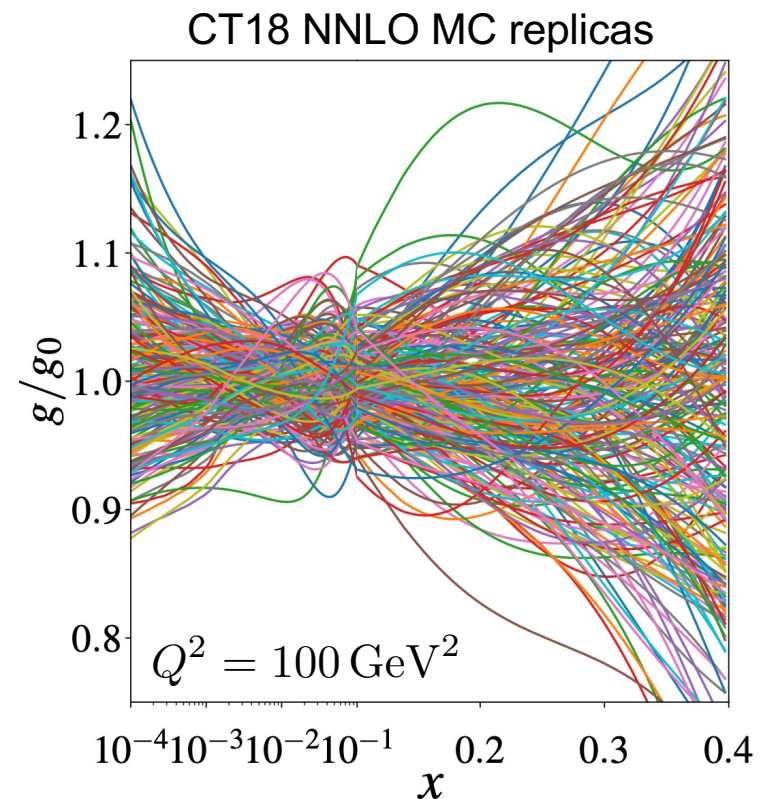
[Kotz et al., [2311.08447](#)]; Fantômas technical paper [[2507.22969](#)]



- complementary to parallel ML-based studies (later)

identify salient features: guided backpropagation (GBP)

Kriesten, Gomprecht, TJH: JHEP11 (2024) 007



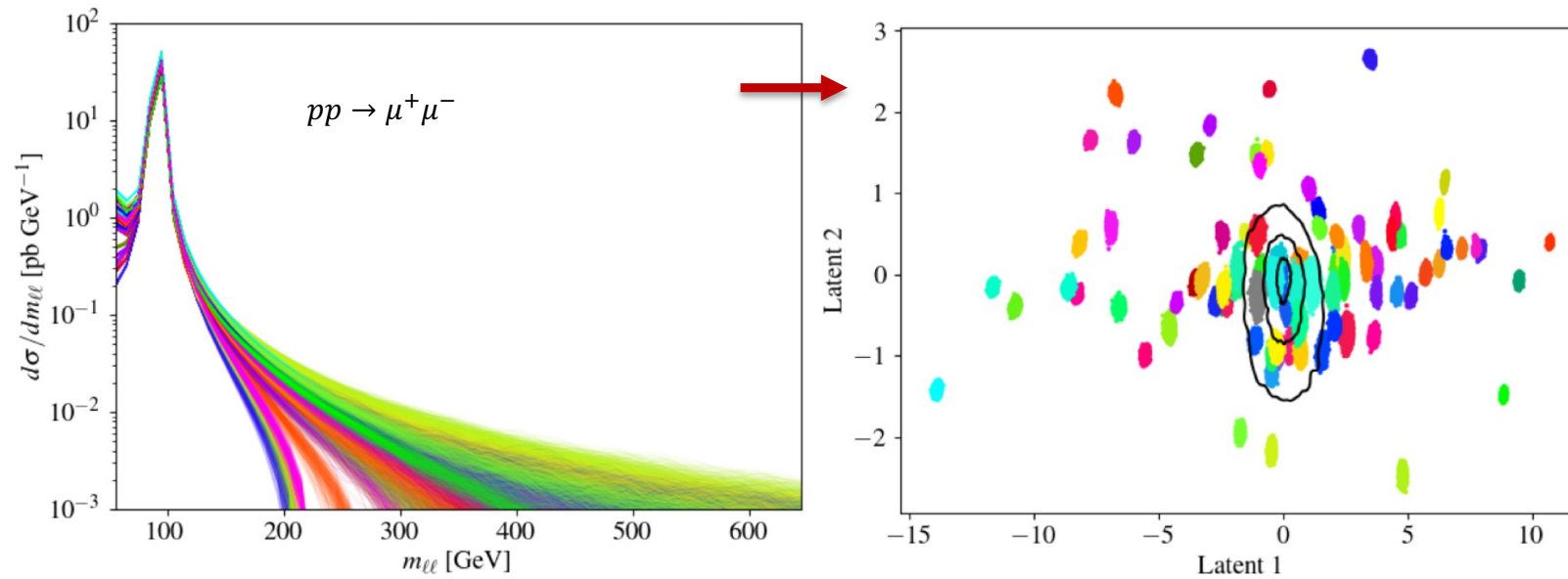
$$\frac{\partial f^{\text{out}}}{\partial f_i^\ell} = (f_i^\ell > 0) \cdot \left(\frac{\partial f^{\text{out}}}{\partial f_i^{\ell+1}} > 0 \right) \cdot \frac{\partial f^{\text{out}}}{\partial f_i^{\ell+1}}$$

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

[public code available]

foundation models for theory inferencing

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



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



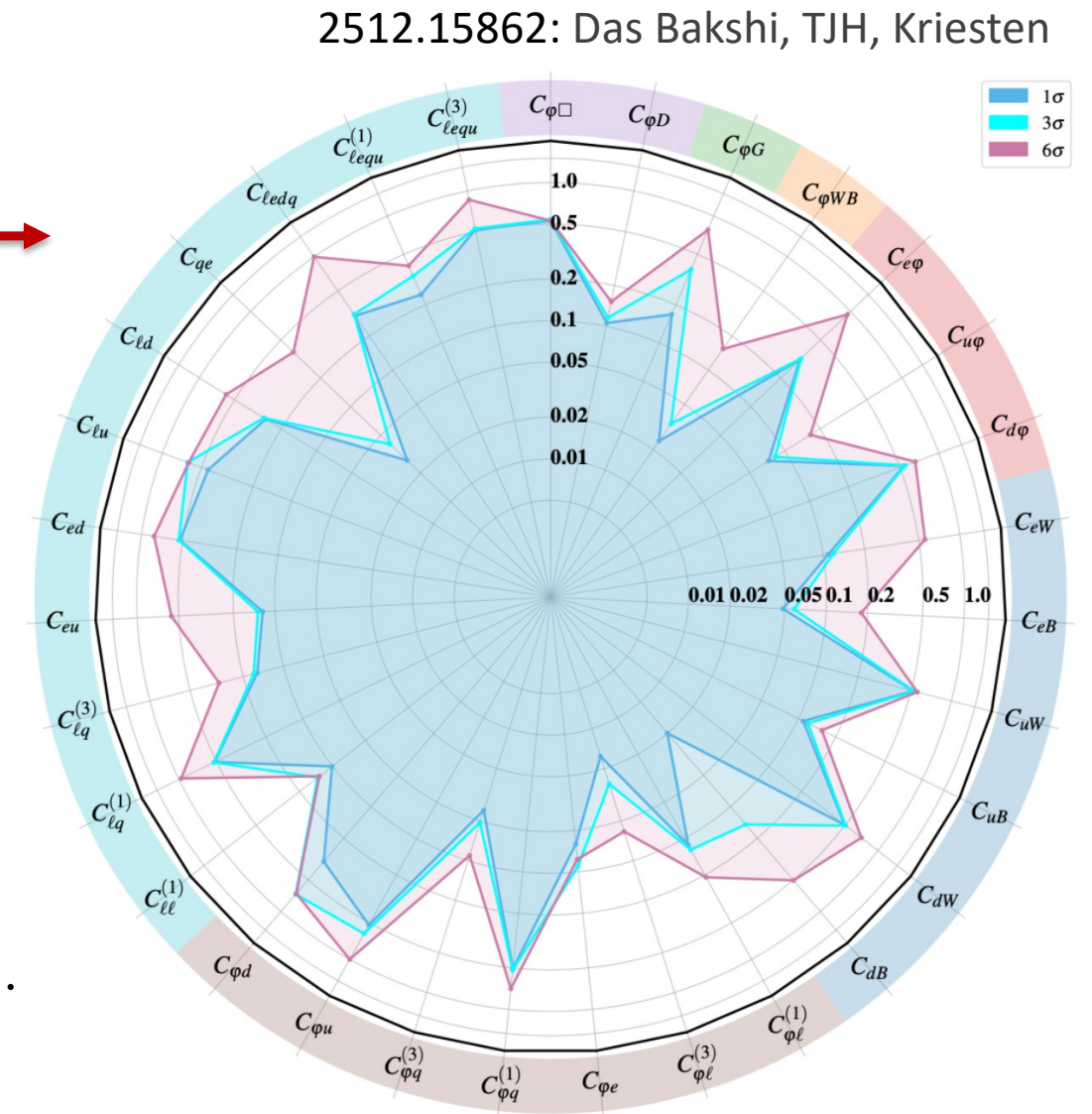
Supratim Das Bakshi



Brandon Kriesten

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

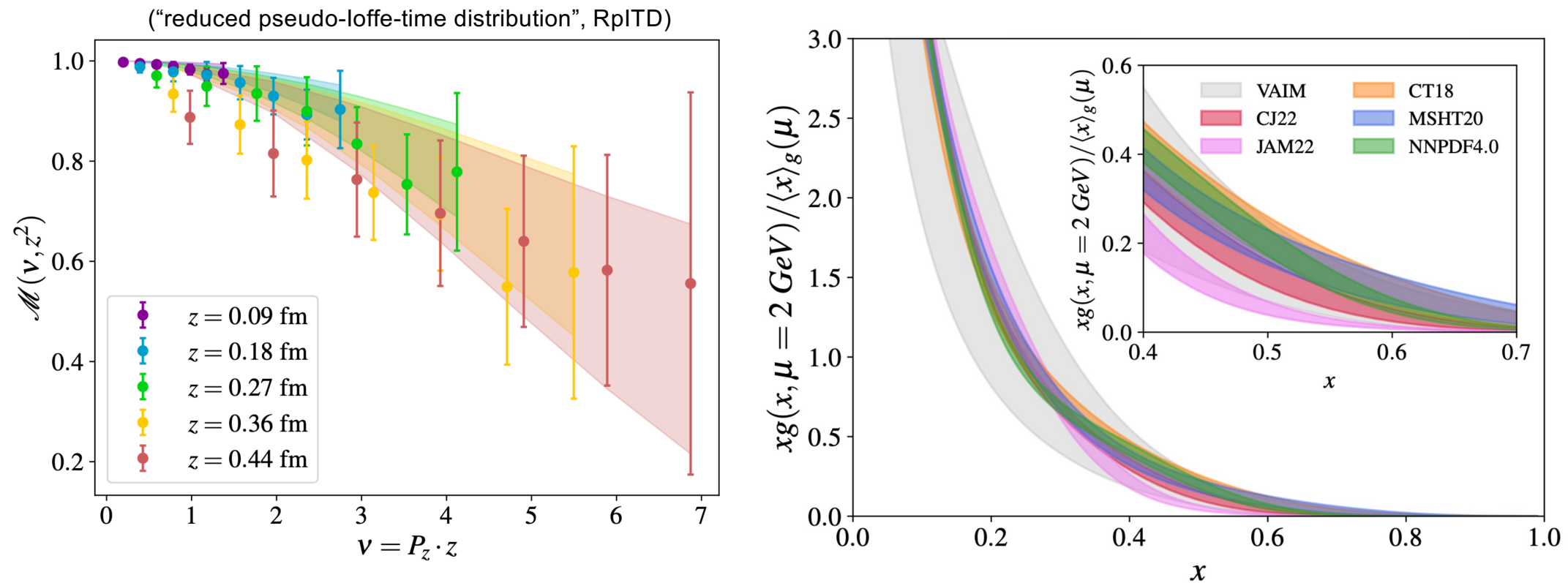
- template for inverse problems across HEP theory: amortize the forward model once, invert on demand



direct unfolding of gluon density

Kriesten, NieMiera, Good, TJH, Lin

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

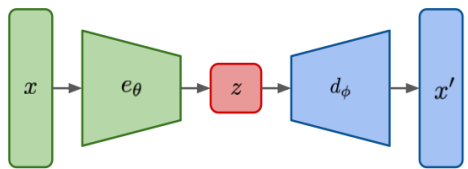
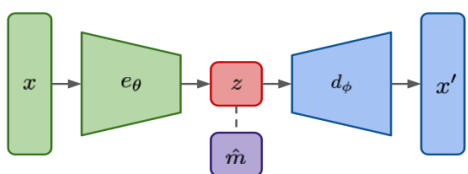
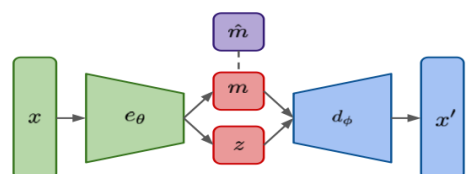
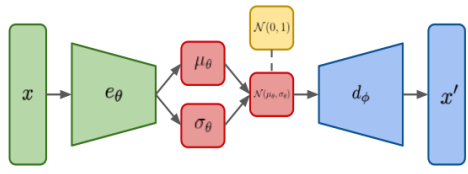
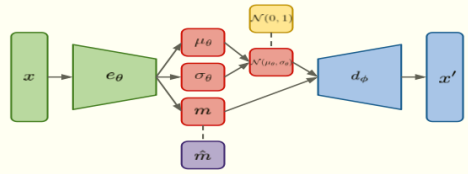


$$\mathcal{M}(\nu, z^2) = \int_0^1 dx \frac{xg(x, \mu^2)}{\langle x \rangle_g(\mu)} R_{gg}(x\nu, z^2 \mu^2) + \mathcal{O}(z^2 m^2, z^2 \Lambda_{\text{QCD}})$$

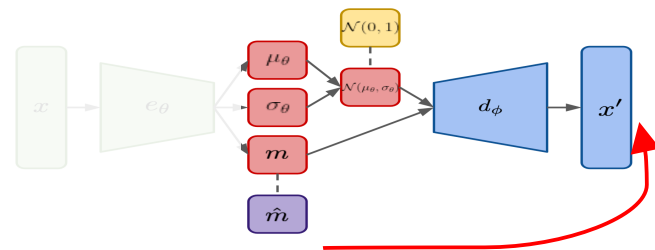
➤ good agreement with pheno PDFs! can be extended with lattice, empirical data

many alternative network models possible

- variations on encoder-decoder structure; different internal topologies

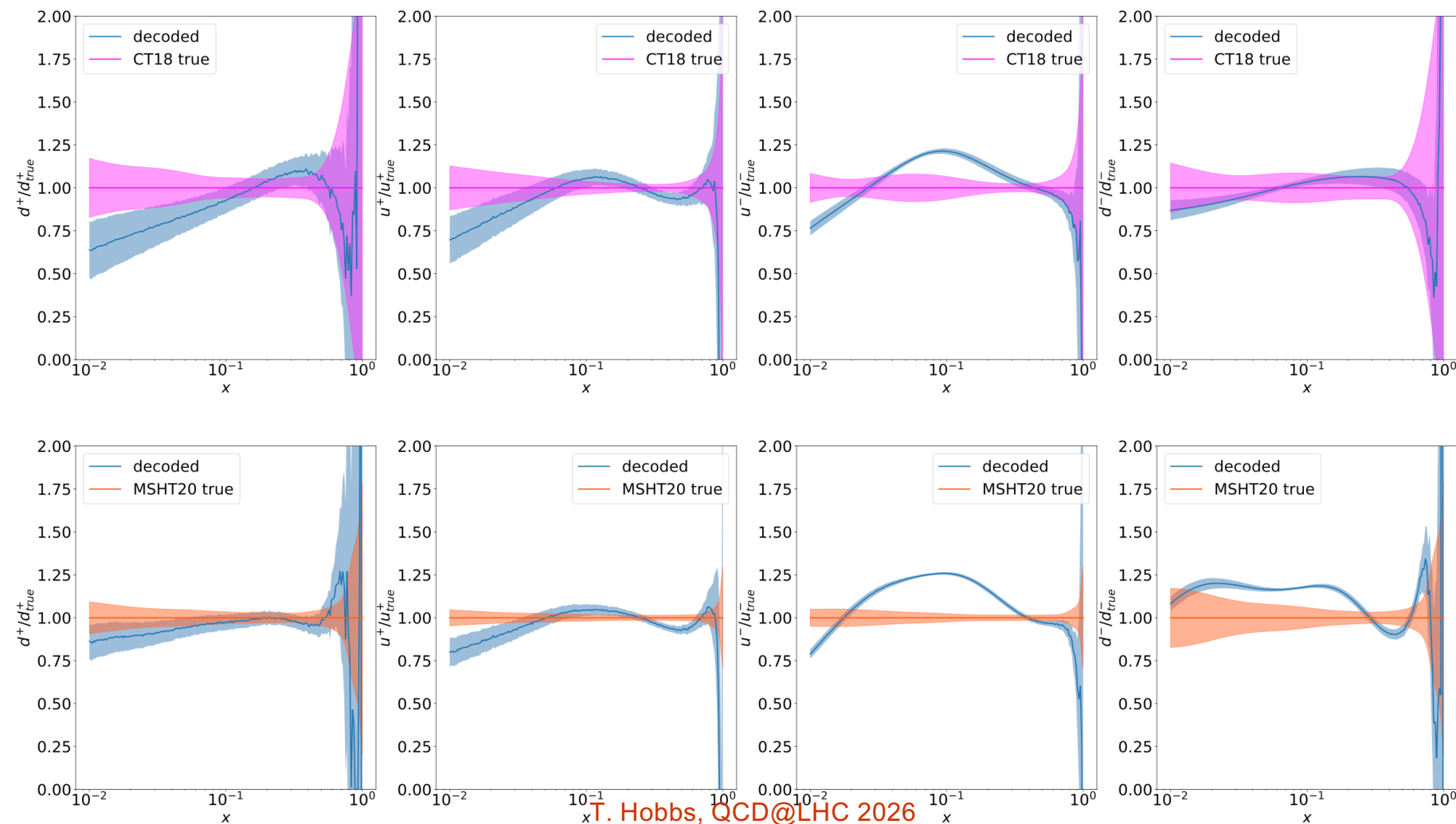
Name	Diagram	Loss	Recreates PDFs	Tractable Latent	Free Latent Dimension	Moment Constraint
AE		$\mathcal{L} = \ x - d_\phi(e_\theta(x))\ _2^2$	✓	✗	✓	✗
AE-CL		$\mathcal{L} = \ x - d_\phi(e_\theta(x))\ _2^2 + \ z - \hat{m}\ _2^2$	✓	✓	✓	✓
AE-WC		$\mathcal{L} = \ x - d_\phi(e_\theta(x))\ _2^2 + \ m - \hat{m}\ _2^2$	✓	✗	✓	✓
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] $\rightarrow$ [PDFs]

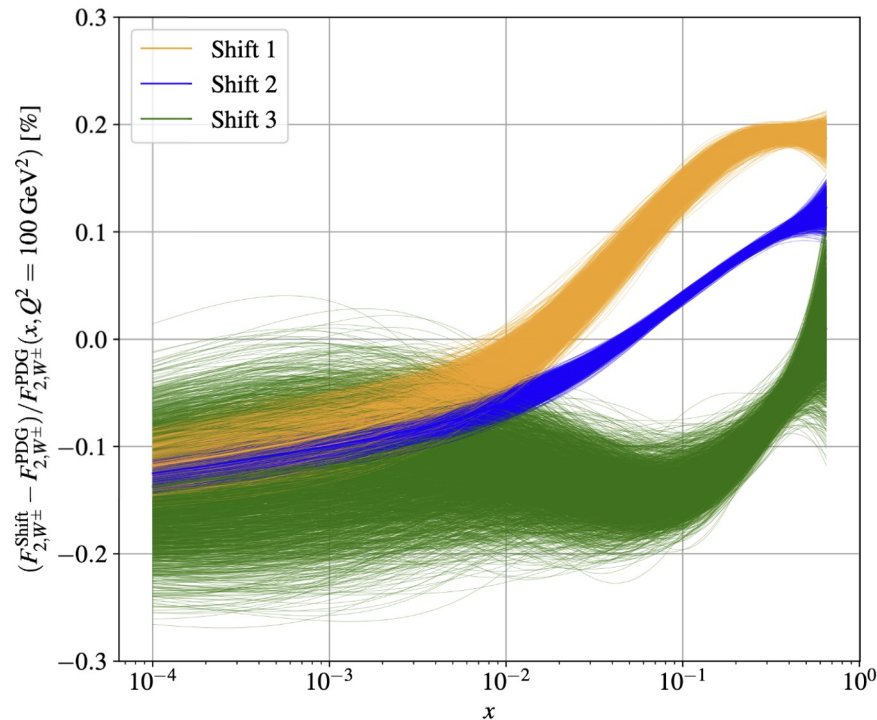
$\rightarrow$ predict pheno PDFs from their moments



(evidential) learning BSM theory differences alongside PDF variations

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

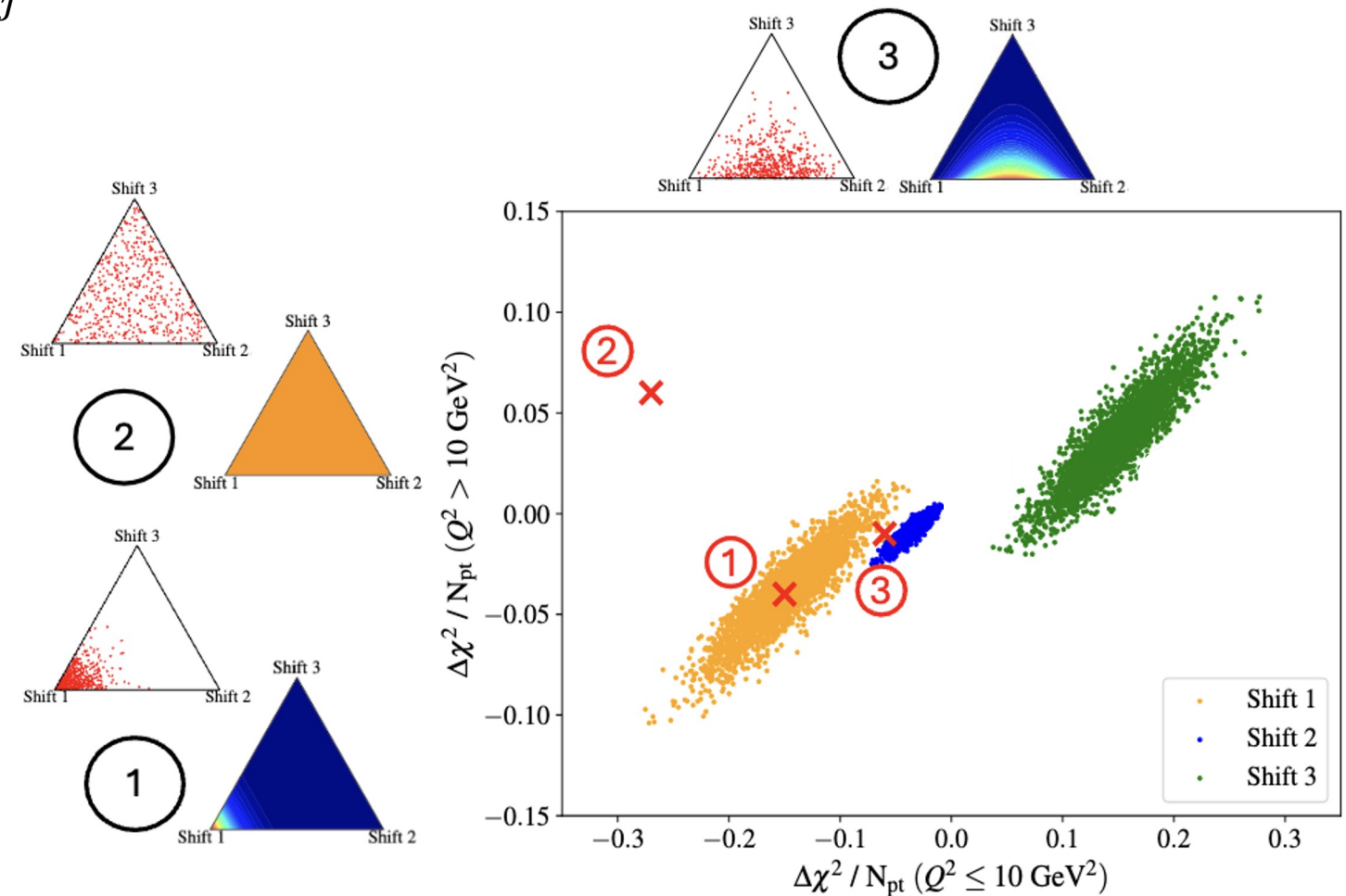
- non-standard neutrino interactions can shift high- x DIS structure functions (e.g., CDHSW predictions)



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

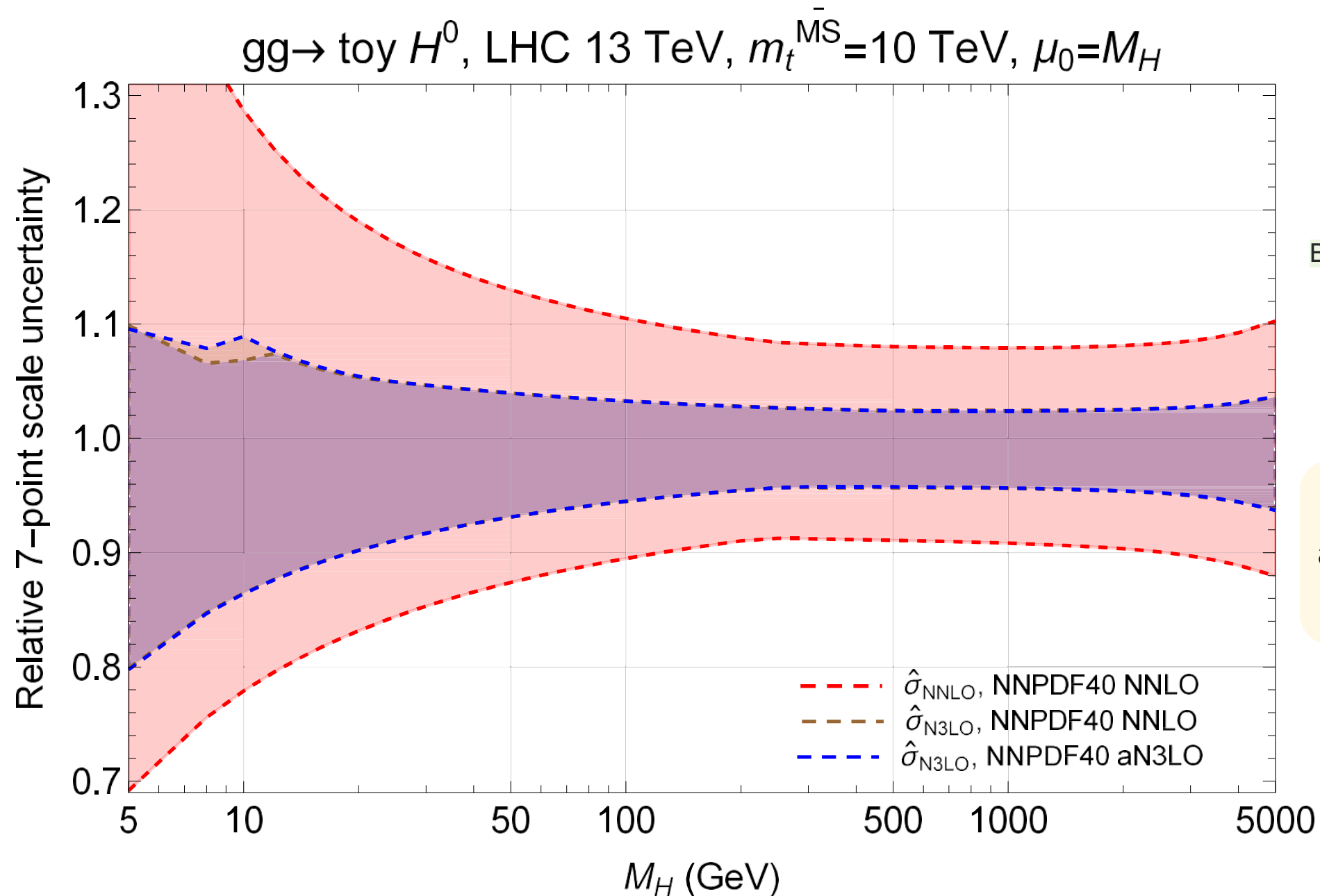
$$\mathcal{L}_{\text{WEFT}} \supset -\frac{2V_{ij}}{v^2} \left\{ \left[1 + \epsilon_L^{ij} \right]_{ab} (\bar{u}^i \gamma^\mu P_L d^j) (\bar{\ell}_a \gamma_\mu P_L \nu_b) + \text{h.c.} \right\}$$

Kriesten, TJH: 2412.16286



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

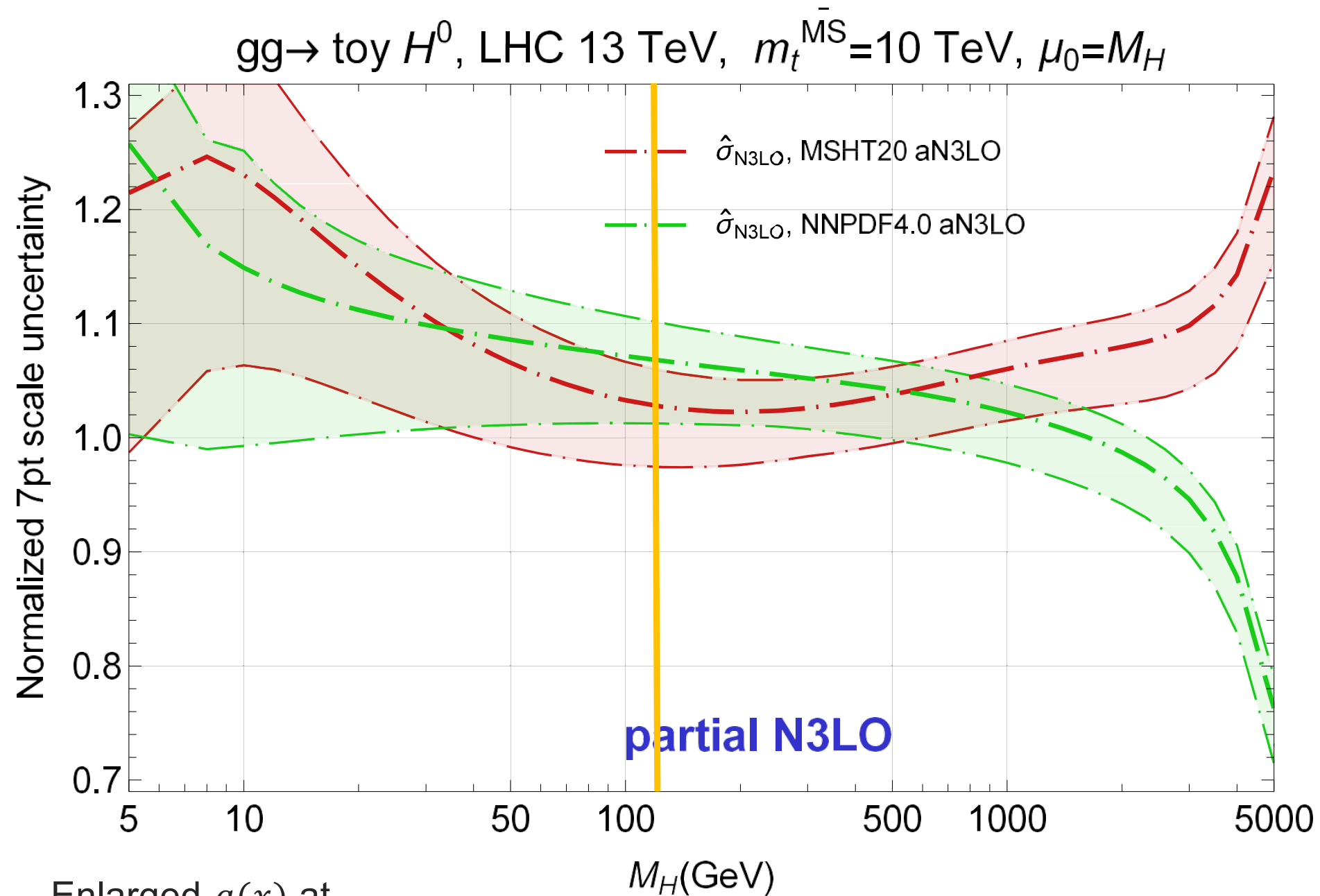
partial N3LO (pN3LO): total cross sections can serve as test bed



Baglio, Duhr, Mistlberger, Szafron (2209.06138)

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

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



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

