



European Research Council

Established by the European Commission

MARIA UBIALI
UNIVERSITY OF CAMBRIDGE

UNPOLARISED COLLINEAR PROTON PDF_s

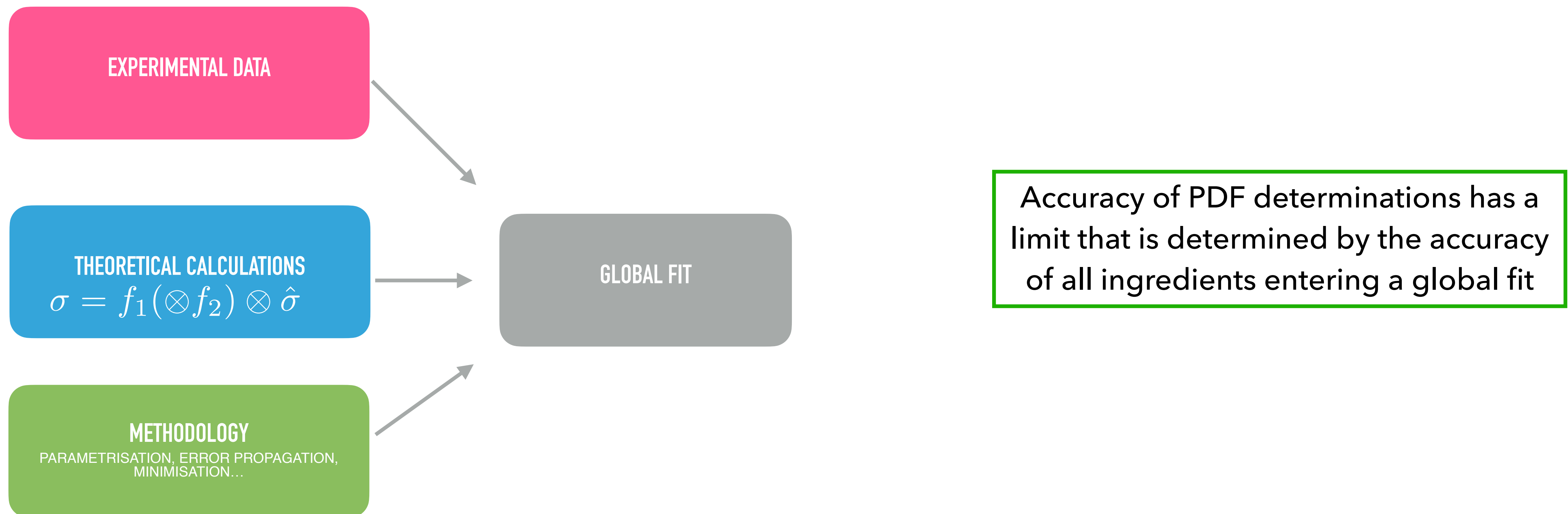
EIC-LHC CONNECTION

OUTLINE

- ➔ **Introduction:** progress and challenges in modern PDF global fits
- ➔ **EIC & LHC:** complementary experimental constraints
- ➔ **EIC & LHC:** preventing BSM bias at HL-LHC with EIC

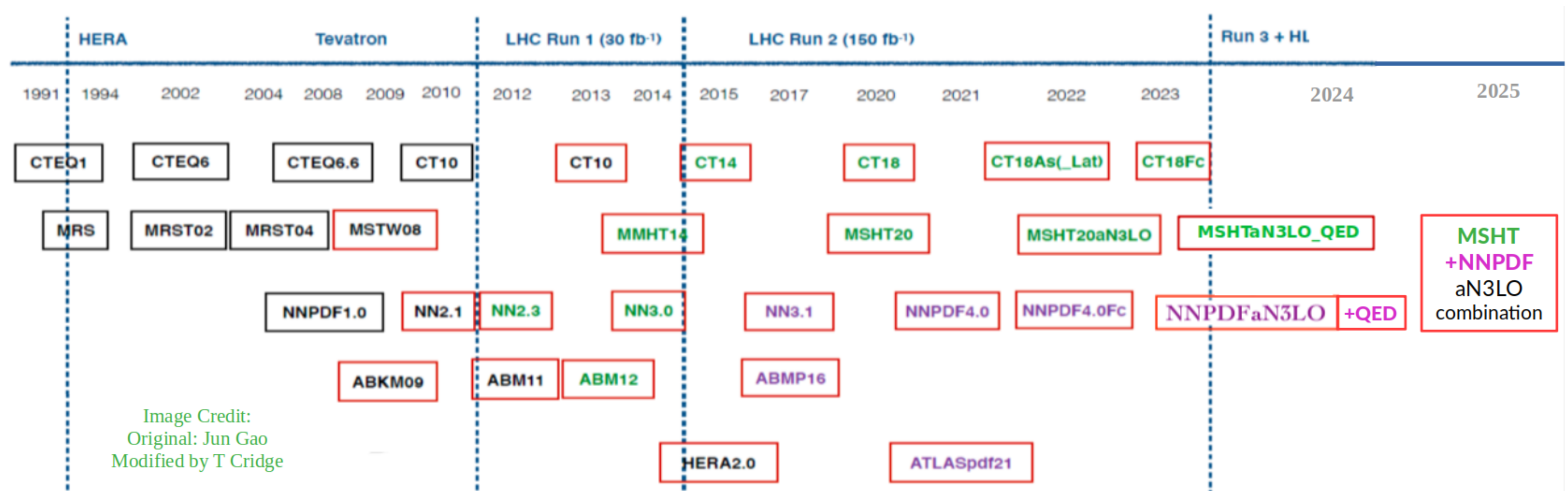
STATUS OF PDF DETERMINATION

- Huge progress from first PDF determinations to where we stand now
 - ➔ **Experimental data**: many more precise measurements constraining PDFs
 - ➔ **Theory**: heavy quark schemes, aN3LO PDFs, inclusion of MHOUs in PDF error bands, EW and QED corrections, initial photon...
 - ➔ **Methodology**: uncertainty propagation, parametrisation, optimal minimization



STATUS OF PDF DETERMINATION

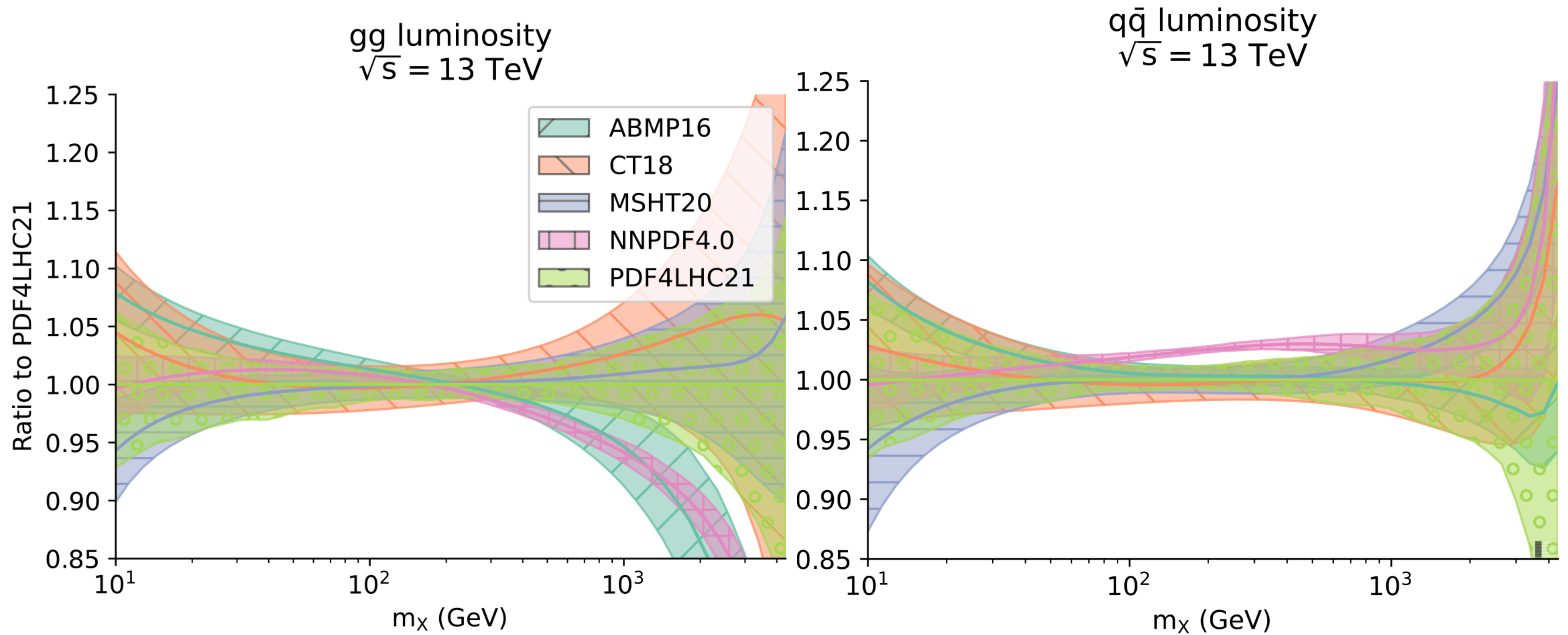
- Huge progress from first PDF determinations to where we stand now
 - ➔ **Experimental data**: many more precise measurements constraining PDFs
 - ➔ **Theory**: heavy quark schemes, aN3LO PDFs, inclusion of MHOUs in PDF error bands, EW and QED corrections, initial photon...
 - ➔ **Methodology**: uncertainty propagation, parametrisation, optimal minimization



STATUS OF PDF DETERMINATION

May 2025	NNPDF4.0	MSHT20	CT18	ABMP16	CJ22	JAM
Fixed Target DIS	✓	✓	✓	✓	✓	✓
HERA I+II	✓	✓	✓	✓	✓	✓
HERA jets	✓	✓	✗	✗	✗	✗
Fixed Target DY	✓	✓	✓	✓	✓	✓
SIDIS	✗	✗	✗	✗	✗	✓
Tevatron W,Z	✓	✓	✓	✓	✓	✓
Tevatron jets	✓	✓	✓	✗	✓	✓
LHC jets	✓	✓	✓	✗	✗	✗
LHC vector boson	✓	✓	✓	✓	✗	✓
LHC top	✓	✓	✓	✓	✗	✗
Stat. treatment	Monte Carlo	Hessian $\Delta\chi^2$ dynamical	Hessian $\Delta\chi^2$ dynamical	Hessian $\Delta\chi^2=1$	Hessian $\Delta\chi^2=1,10$	Monte Carlo
Parametrization	Neural Networks	Polynomial (Chebyshev)	Polynomial (Bernstein)	Polynomial	Polynomial	Polynomial (50 pars)
HQ scheme	FONLL	TR'	ACOT- χ	FFN (+BMST)	ACOT	ACOT
Order	NNLO/aN3LO	NNLO/aN3LO	NNLO	NNLO	NLO	NLO

STATUS OF PDF DETERMINATION

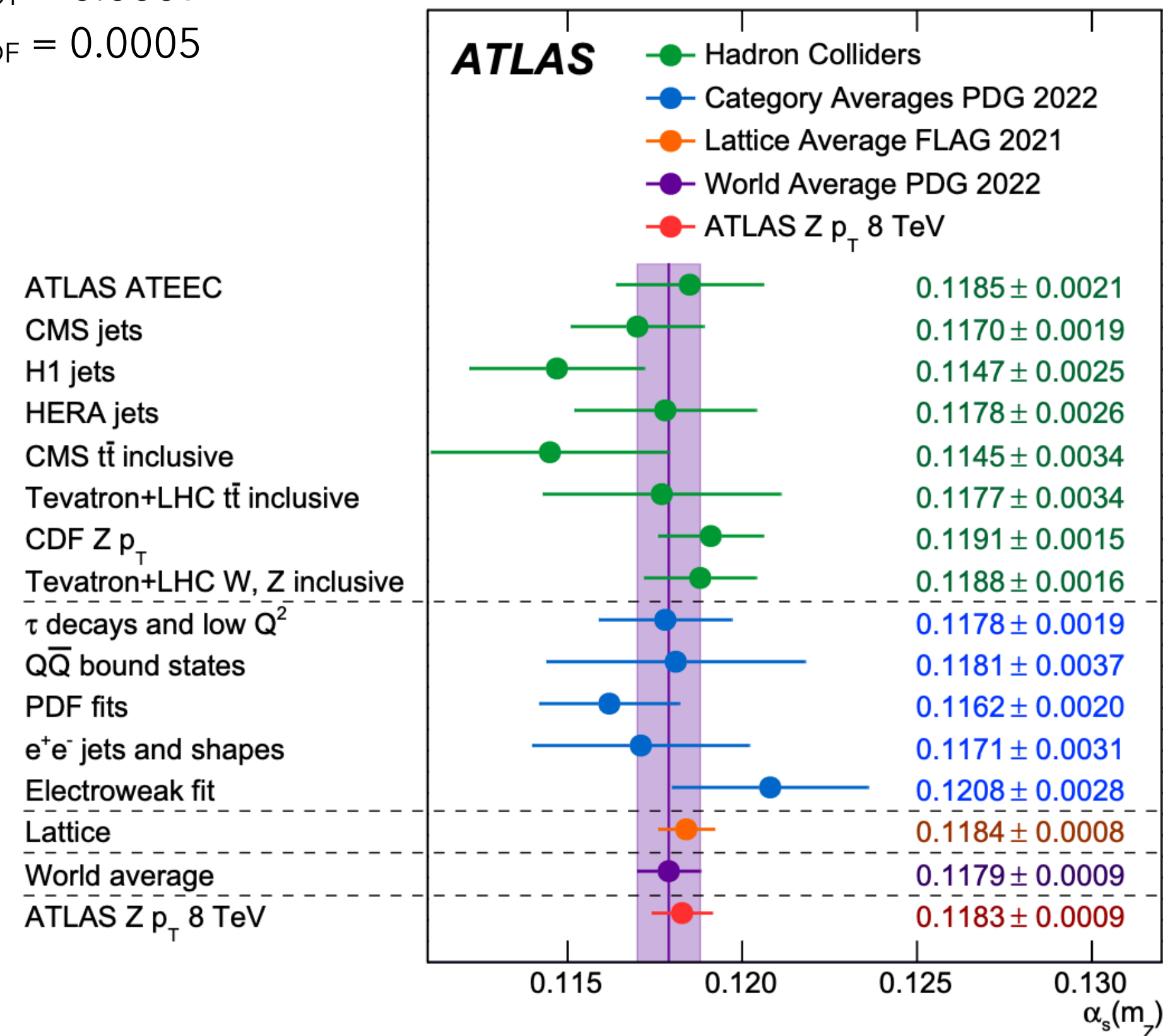


IMPLICATION ON DETERMINATION OF SM PARAMETERS

ATLAS α_s determination

$\Delta_{\text{TOT}} = 0.0009$

$\Delta_{\text{PDF}} = 0.0005$



ATLAS collaboration, arXiv: 2309.12986

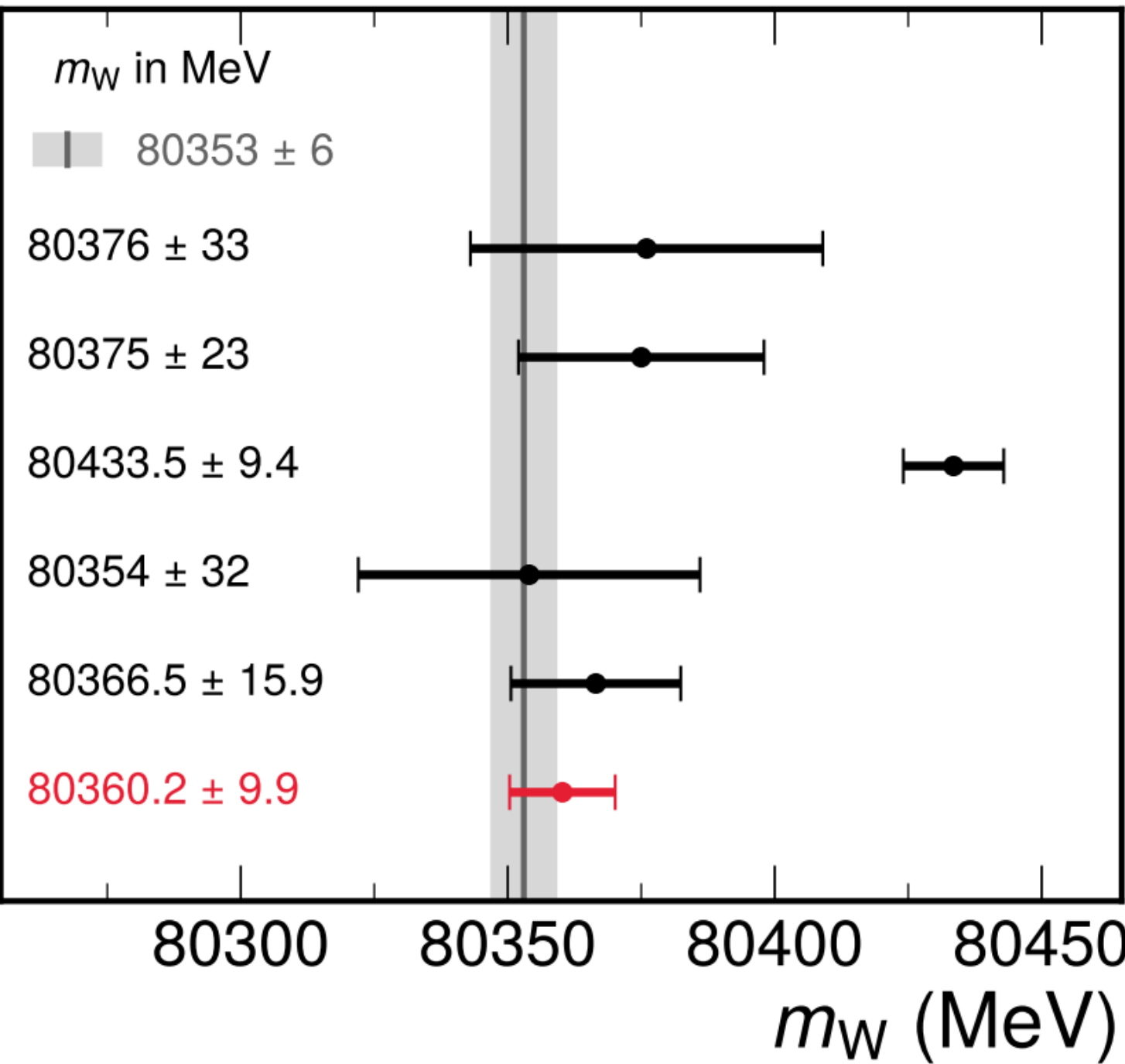
CMS M_W determination

$\Delta_{\text{TOT}} = 9.9 \text{ MeV}$

$\Delta_{\text{PDF}} = 4.4 \text{ MeV}$

Electroweak fit
PRD 110 (2024) 030001
LEP combination
Phys. Rep. 532 (2013) 119
D0
PRL 108 (2012) 151804
CDF
Science 376 (2022) 6589
LHCb
JHEP 01 (2022) 036
ATLAS
arXiv:2403.15085
CMS
This work

CMS



CMS collaboration, arXiv: 2412.13872

WAY FORWARD #1: STATISTICAL UNDERSTANDING OF METHODOLOGY

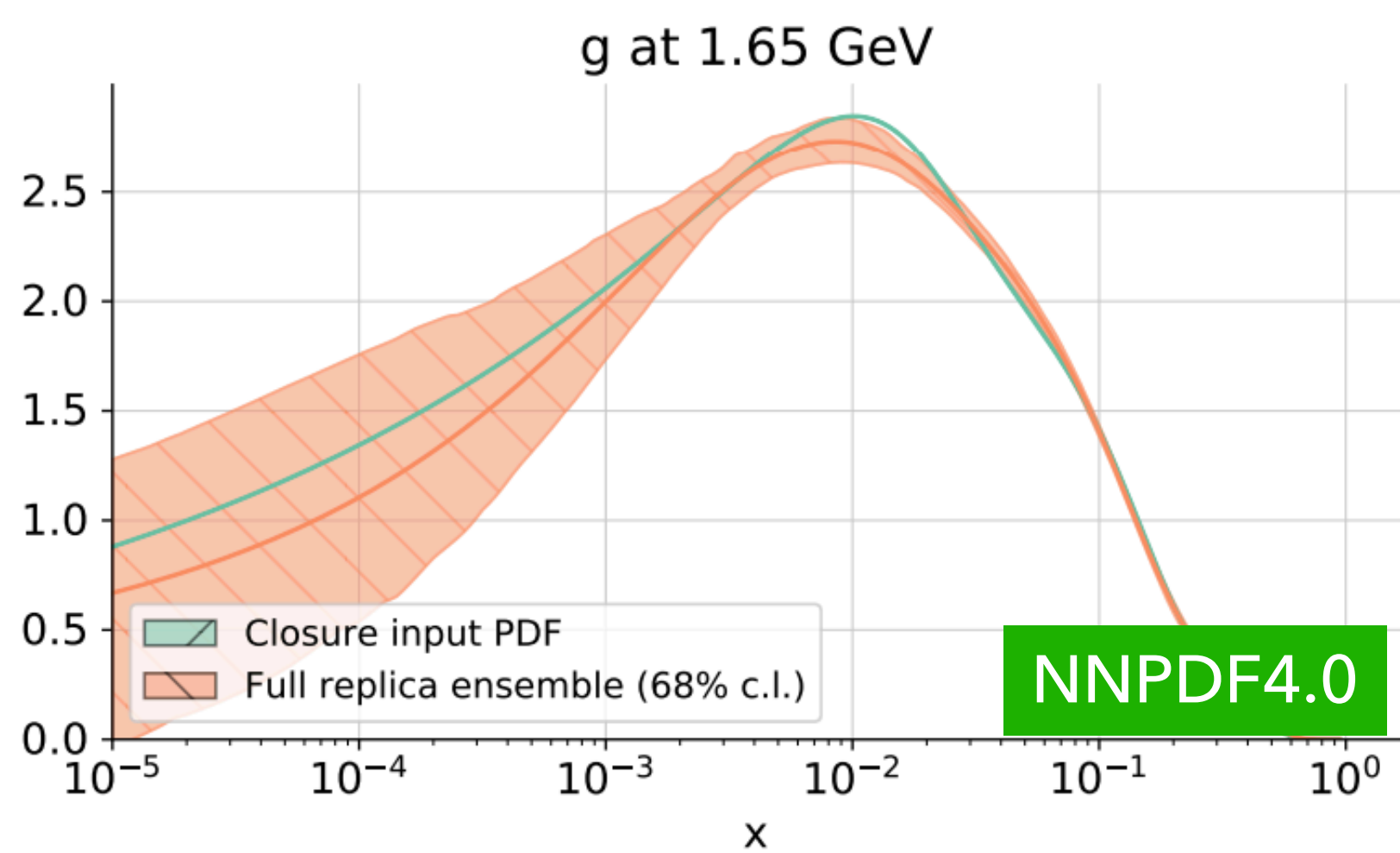
- Various studies of PDF parameterisation/methodology/uncertainty.
[Ablat et al, arXiv:2408.04020, Costantini et al arXiv:2503.17447, Hunt-Smith et al: arXiv: 2309.01454, etc...]
- **Closure tests** - validate PDF fit methodologies + assess faithfulness of PDF uncertainties.
- Can check the effect of experimental/theoretical inconsistencies and how PDF fits can absorb/deal with it
[Harland-Lang et al, arXiv:2407.07944, Costantini et al: arXiv:2503.17447]

$$y_0 = \mathcal{G}(w^0) + \eta$$

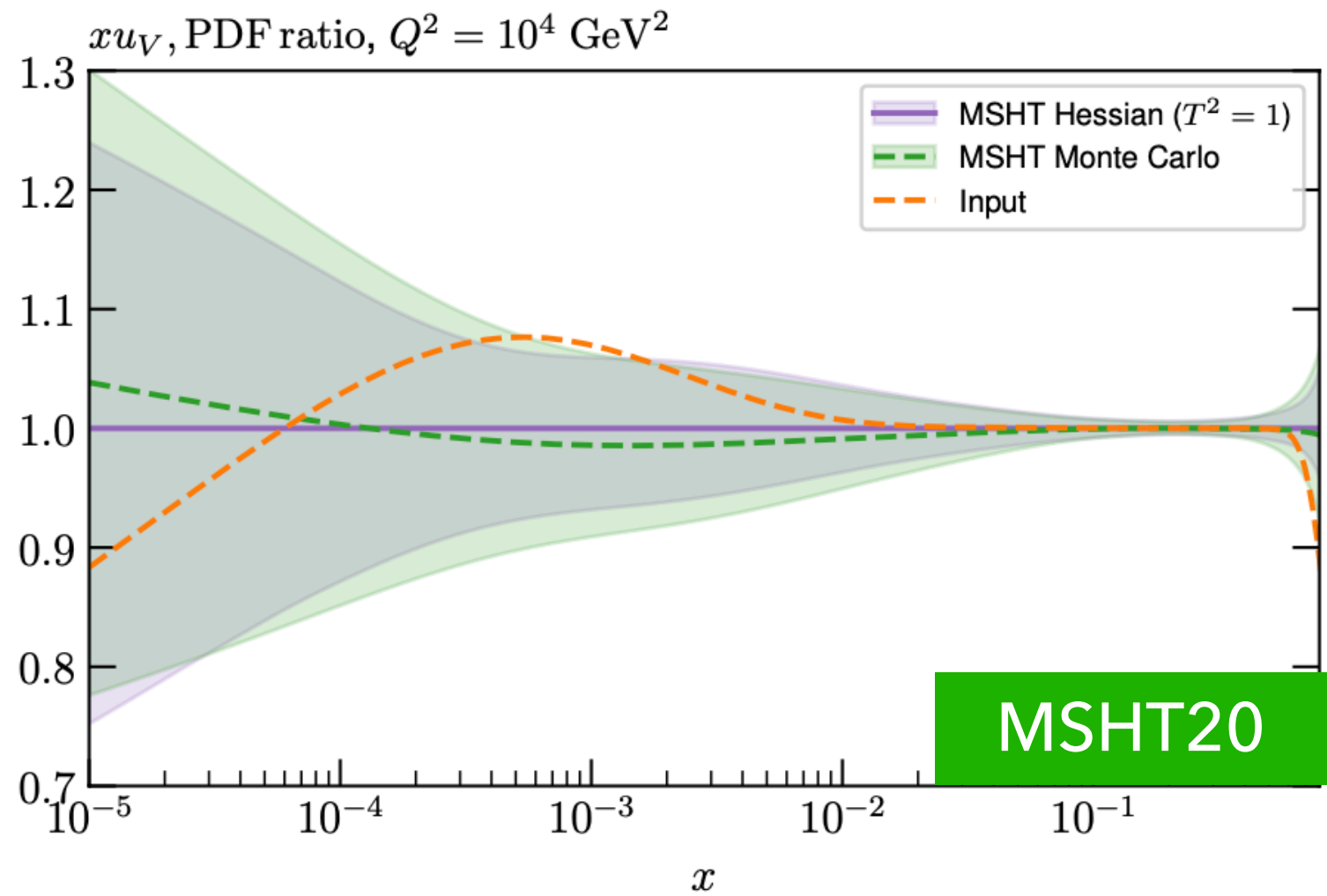
Experimental noise

Law of Nature

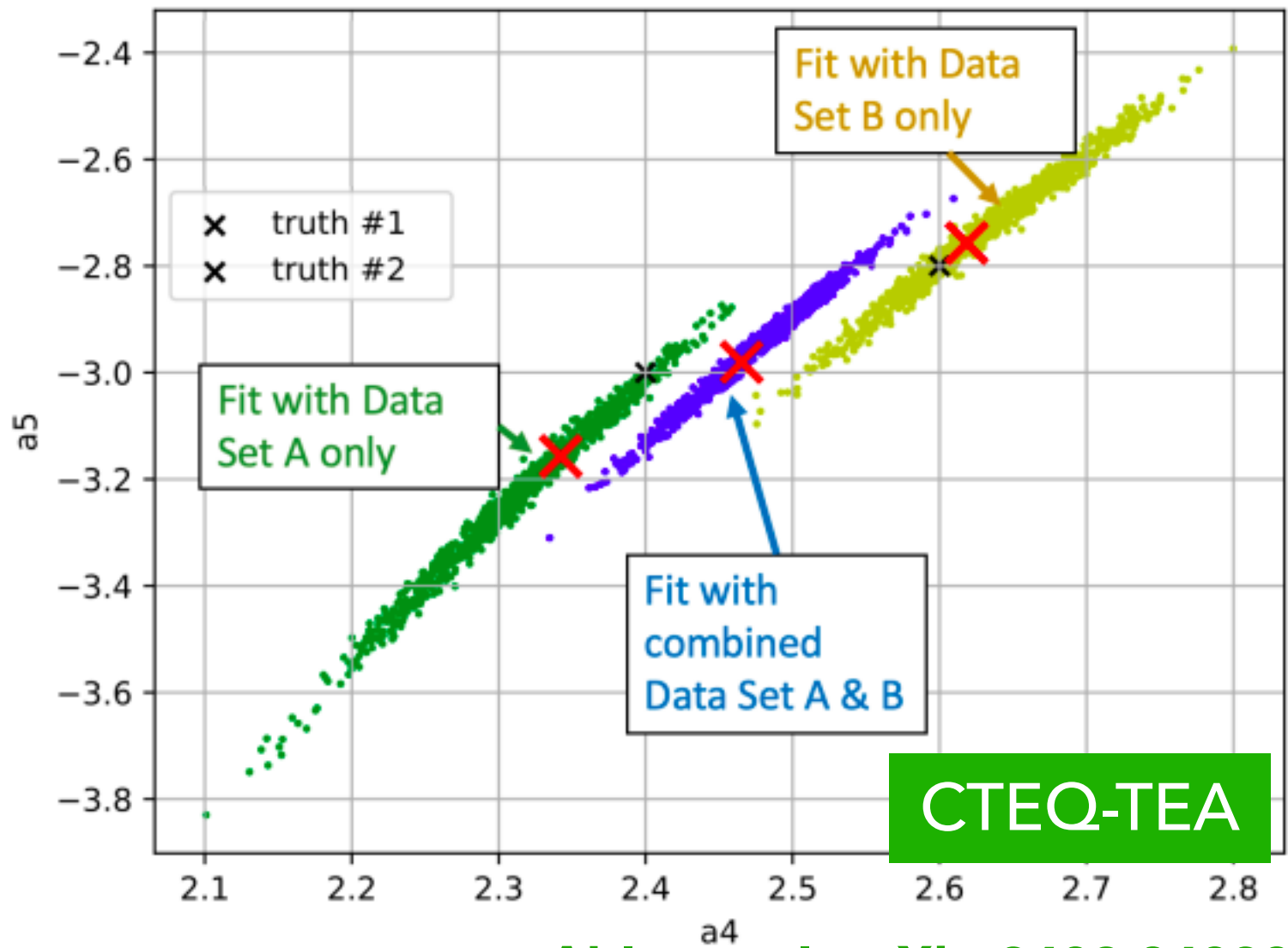
$\hat{\sigma}_{\text{NNLO}} \otimes (f \otimes f)$ “true” PDF



Del Debbio et al, [arXiv: 2111.05787]



Harland-Lang et al, arXiv:2407.07944



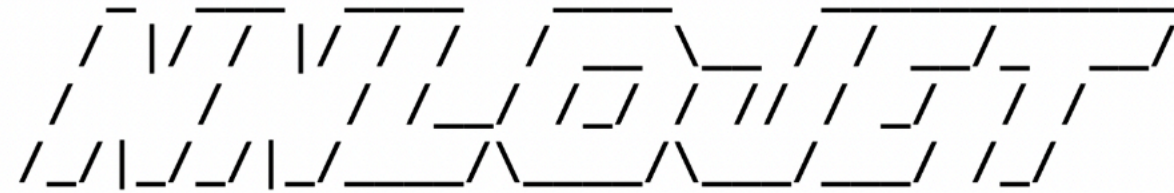
Ablat et al, arXiv:2408.04020

WAY FORWARD #2: CHECK PERFORMANCE AGAINST UNSEEN DATA

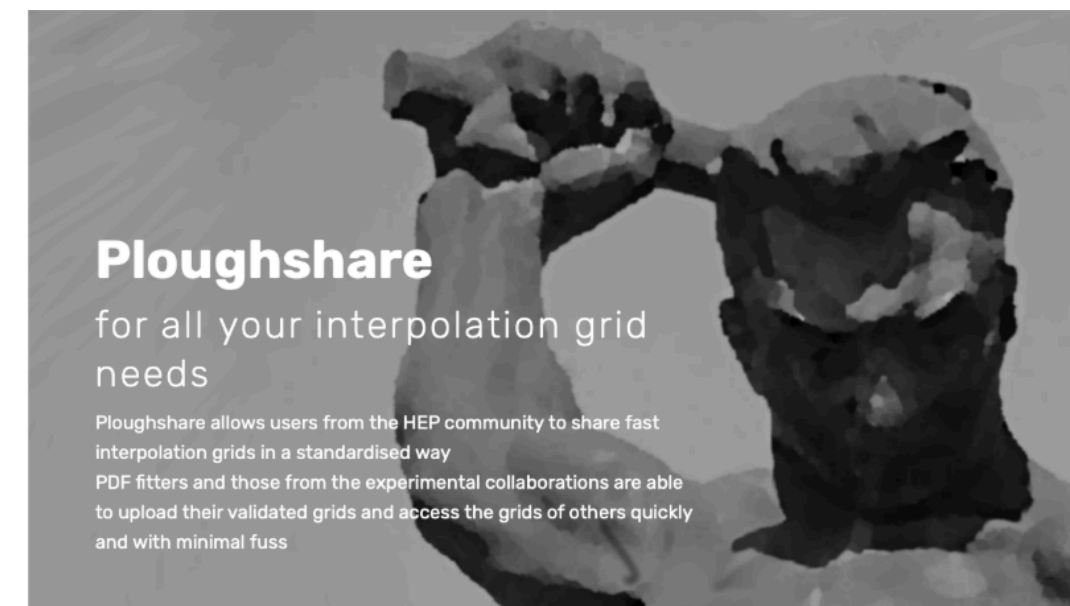
- To discriminate among PDF sets, test all global PDF sets against **new precise** data from LHC data & DIS HERA jets data. How well do various PDF sets describe data that are **not yet** included in the fit?
- **Generalisation test** is crucial and possible using exact NNLO predictions (no k-factor approximation) thanks to PineAPPL, NNLOJET, MATRIX and Ploughshare

PineAPPL

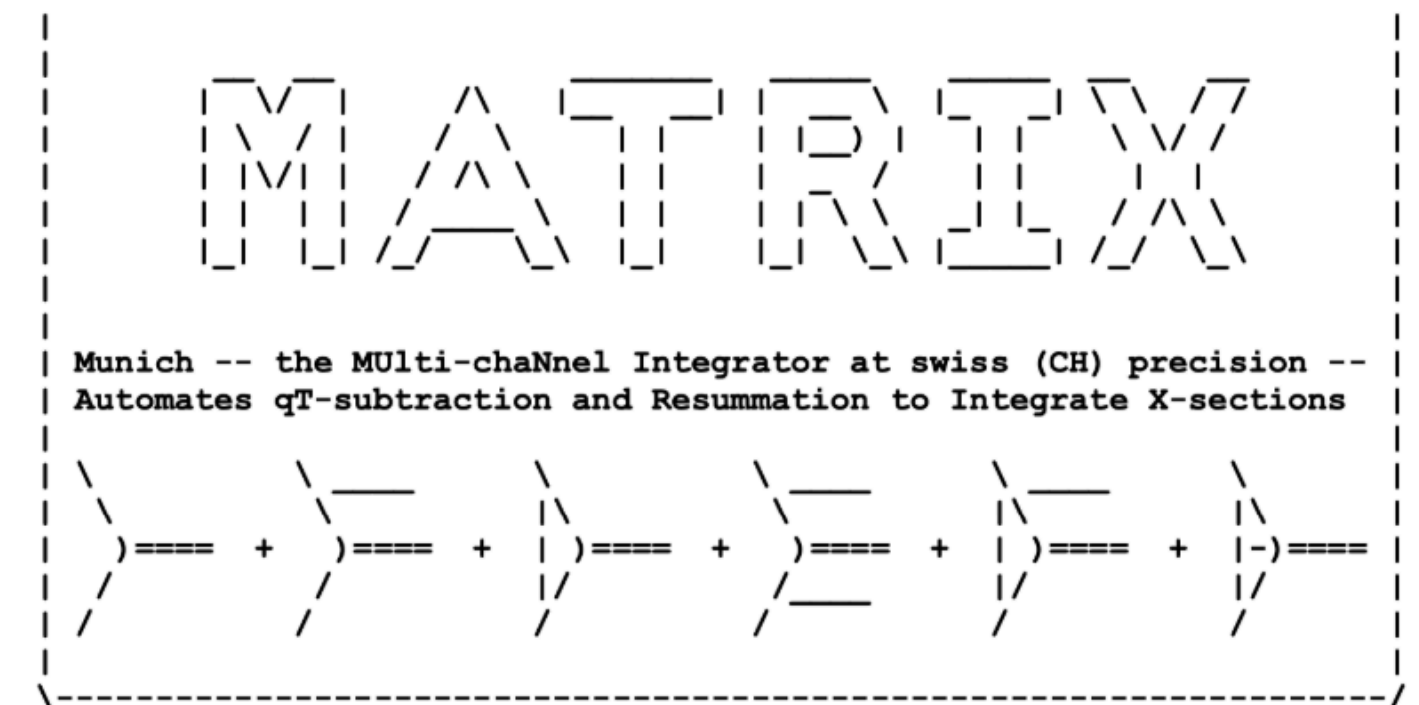
Carrazza et al: 2008.12789
Barontini et al: 2302.12124



Gehrmann-De Ridder et al: 1507.02850,
1605.04295, 2301.11827, Britzger et al:
2207.13735, Cruz-Martinez et al: 2501.13167



Ploughshare project, APPLgrid, FastNLO
Kluge et al: hep-ph/0609285
Carli et al: 0911.2985
Czakon et al: 1704.08551



Grazzini et al: 1711.06631

- Data-theory agreement with unseen data must be measured by considering **all sources of uncertainties**: experimental, missing higher order (MHOU), PDF uncertainties, α_s uncertainties by including all bin-by-bin correlations

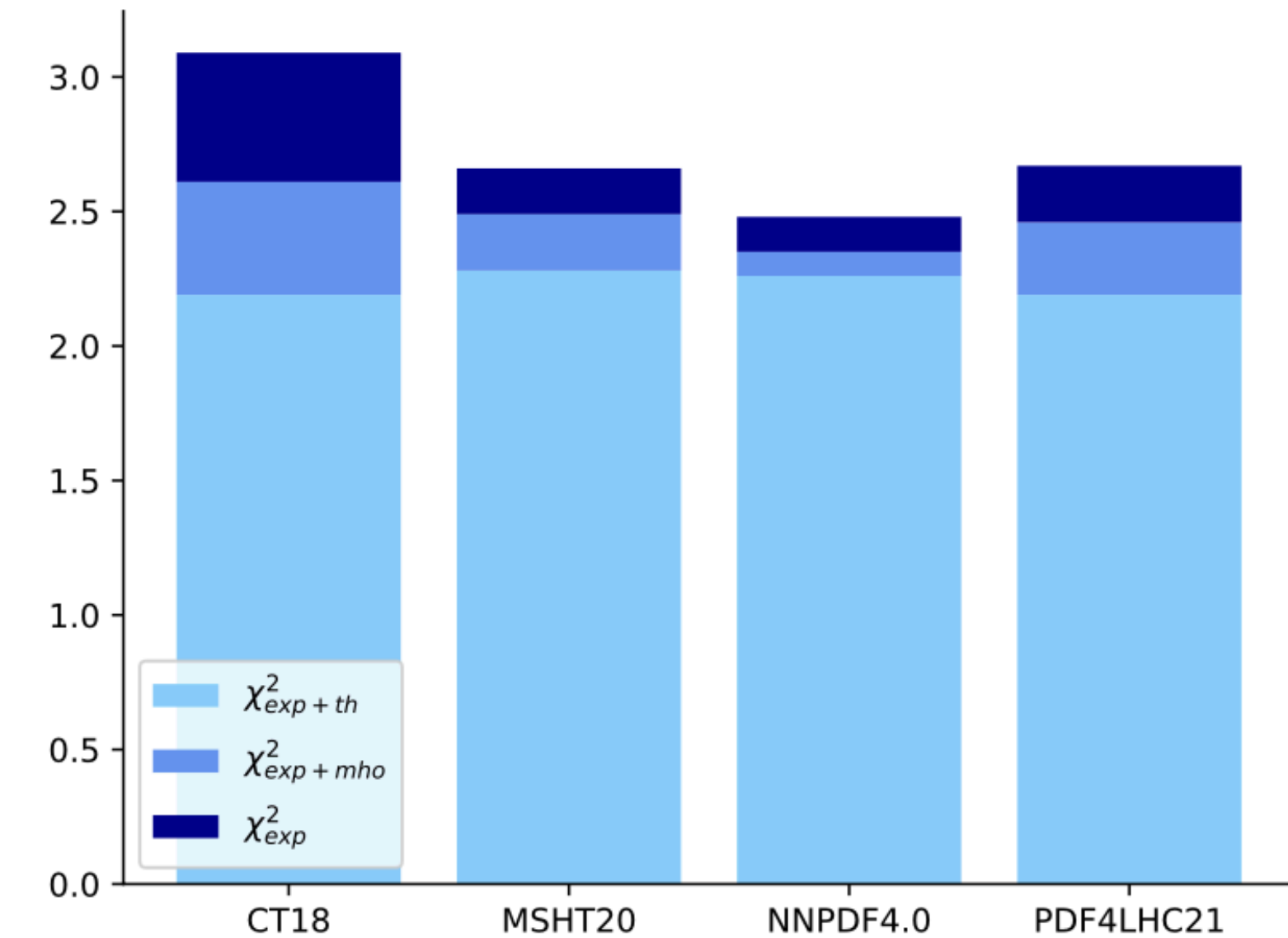
$$\chi^2 = \sum_{i,j=1}^{N_{\text{dat}}} \left(T_i^{(0)} - D_i \right) (\text{cov}^{-1})_{ij} \left(T_j^{(0)} - D_j \right)$$

$$\chi_{\text{exp+mho+pdf+as}}^2 \leftrightarrow \text{COV} = \text{COV}_{\text{exp}} + \text{COV}_{\text{mho}} + \text{COV}_{\text{pdf}} + \text{COV}_{\text{as}}$$

WAY FORWARD #2: CHECK PERFORMANCE AGAINST UNSEEN DATA

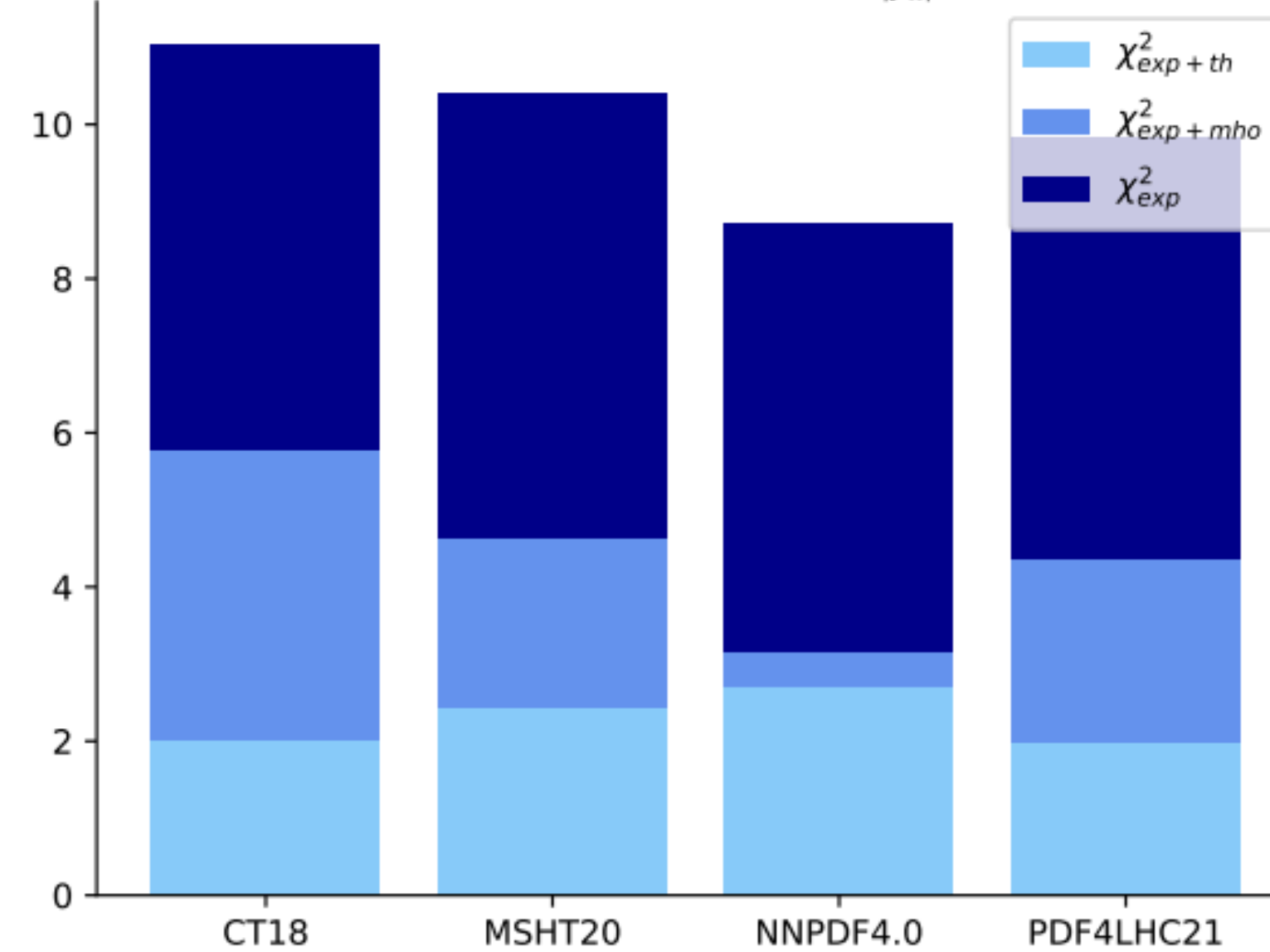
Chiefa et al: 2501.10359

LHCb $Z \rightarrow \mu\mu$



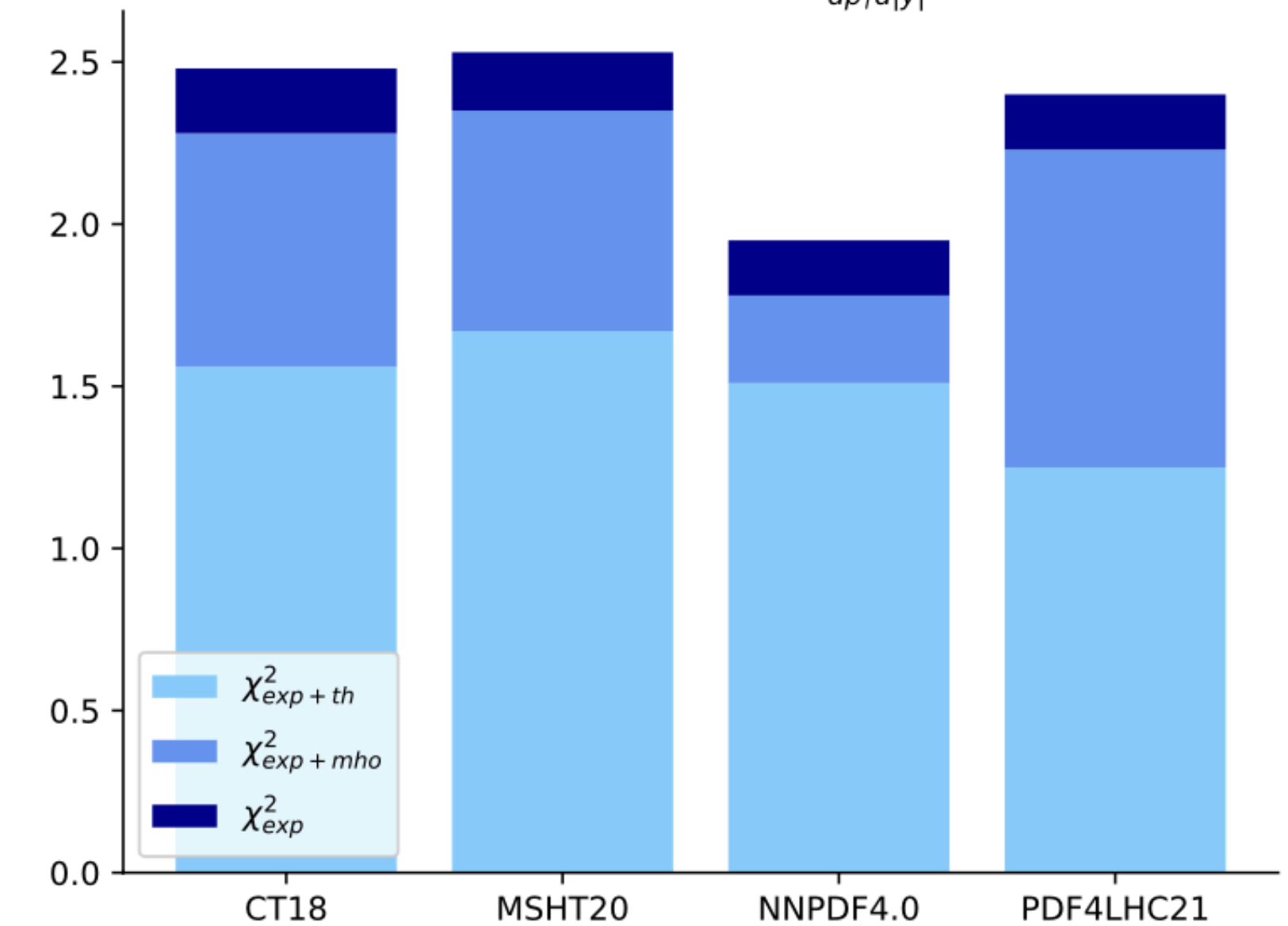
LHCb Z forward production at 13 TeV
LHCb collaboration [arXiv:2112.07458]

CMS 13 TeV $t\bar{t} \ell + j \frac{1}{\sigma} \frac{d\sigma}{d|y_{t\bar{t}}|}$



CMS $t\bar{t}$ production at 13 TeV, $\ell + j$
CMS collaboration [arXiv:2108.02803]

ATLAS Jet 13 TeV: $\frac{d^2\sigma}{dp_T d|y|}$



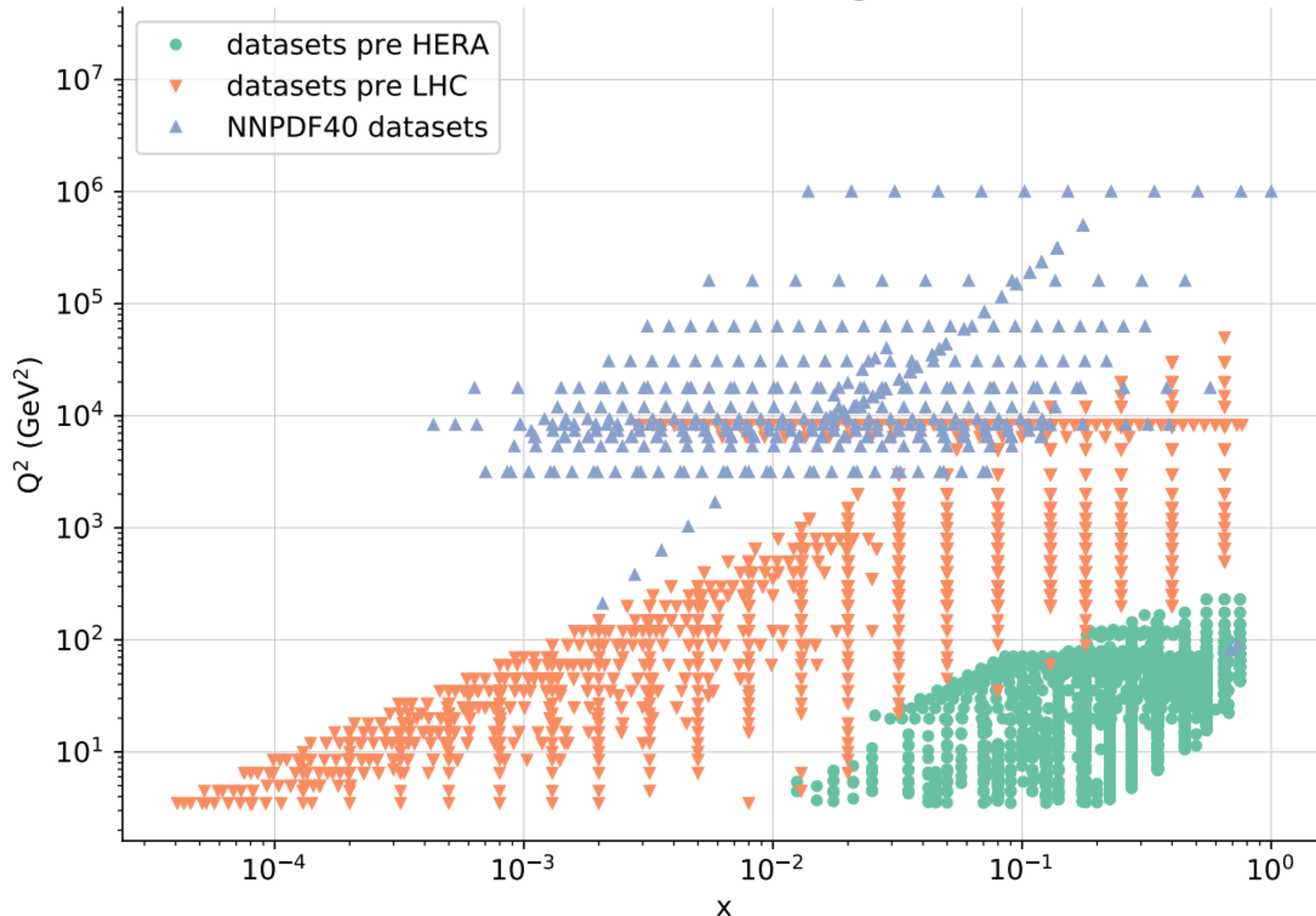
ATLAS incl. jet sec at 13 TeV, $R=0.6$, anti-kT
ATLAS collaboration [arXiv:1711.02692]

- Recently analysed 25 experimental distributions LHC@13 TeV and HERA jets measurements not yet included in any PDF analysis: inclusive jets, dijets, top differential distributions, W and Z production, Z pT, DIS jets [Chiefa et al, arXiv:2501.10359]
- Crucial to keep into account all sources of uncertainties and all bin-by-bin corrections.
- For some observables the inclusion of MHO uncertainties or PDF uncertainties changes the picture as compared to pure experimental χ^2
- LHC@13 TeV measurements **do not strongly discriminate among PDF sets so far**.
- NNPDF4.0, despite having smallest PDF uncertainties in data region yields good agreement with most data

EIC AND LHC CONNECTION #1: COMPLEMENTARY CONSTRAINTS

EXPERIMENTAL DATA

Kinematic coverage

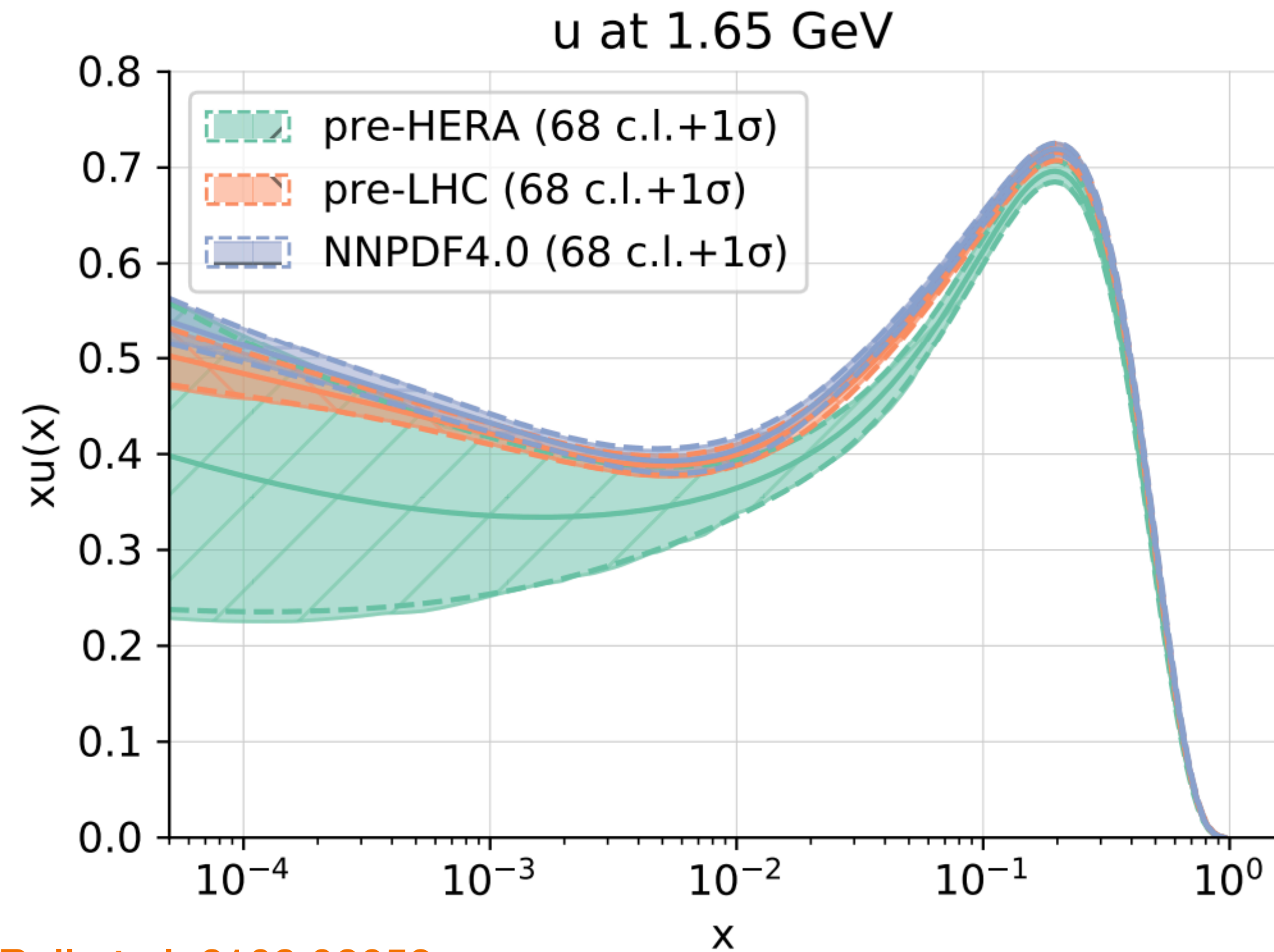


- Range of processes and colliders → complementary constraints on PDFs.
- Modern data show increasingly small statistical uncertainty and dominating correlated systematic uncertainties make global PDF fits challenging

[Kassabov, Nocera, Wilson 2207.00690](#)

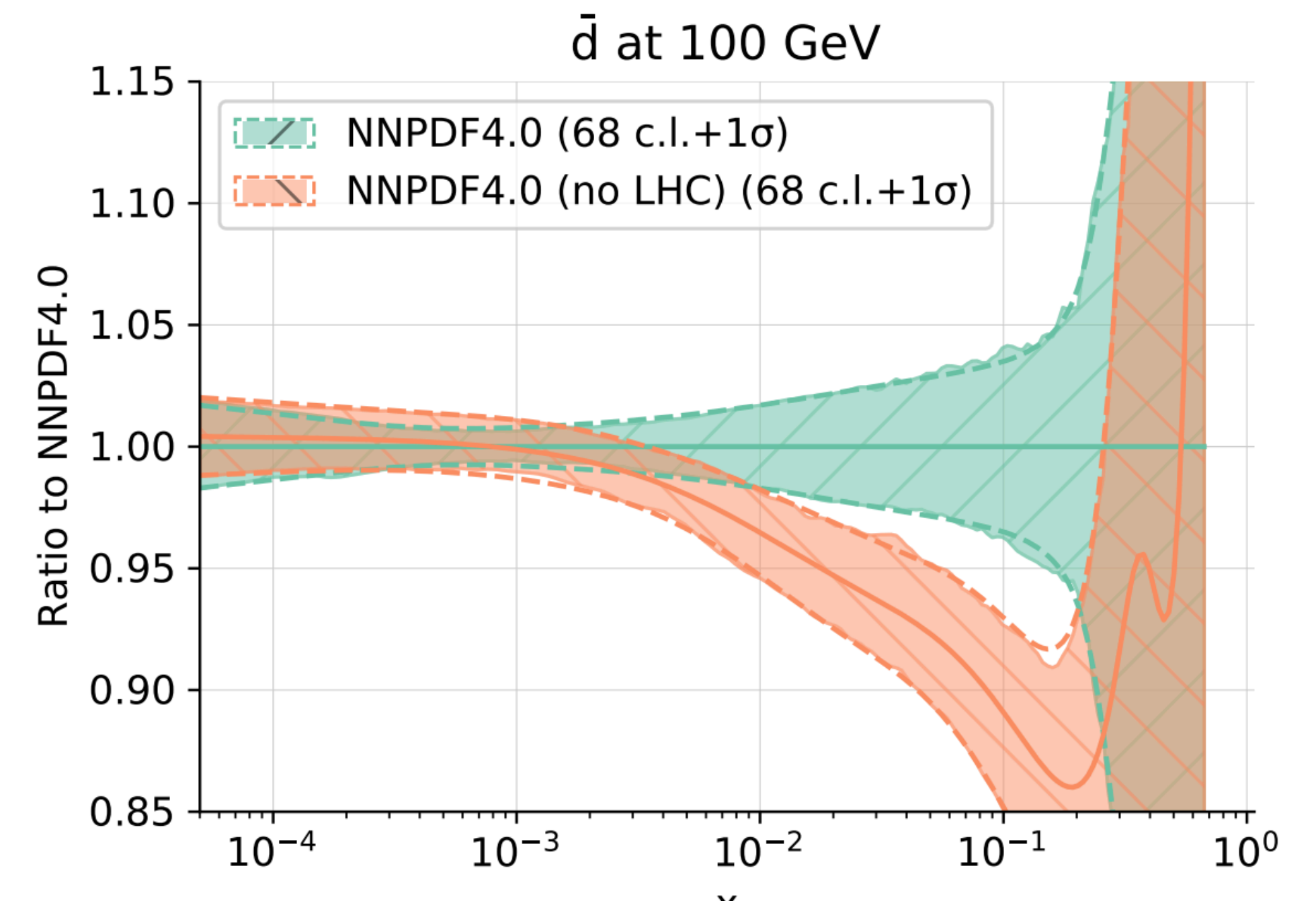
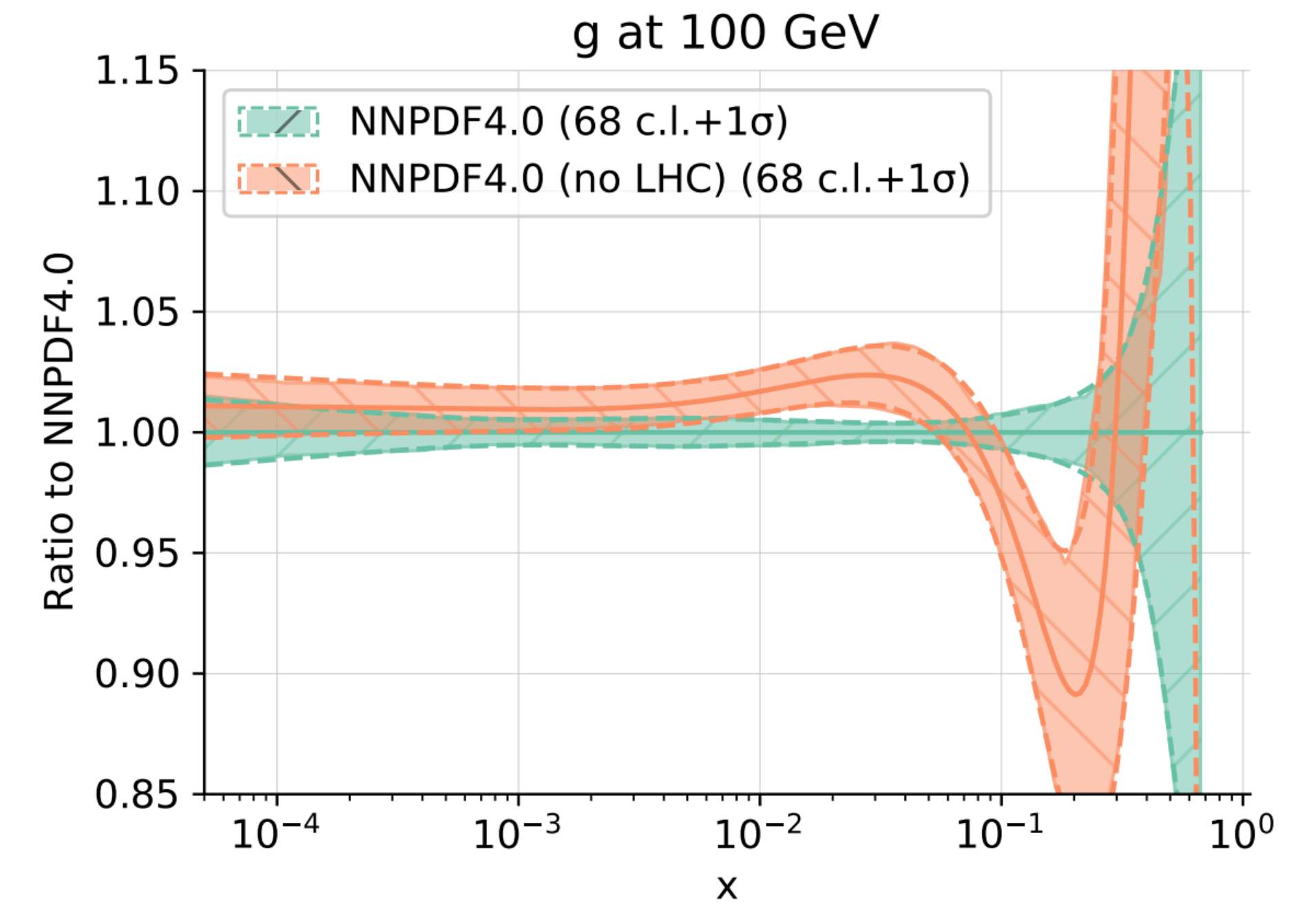
[Bailey, Harland-Lang 1909.10541](#)

EXPERIMENTAL DATA

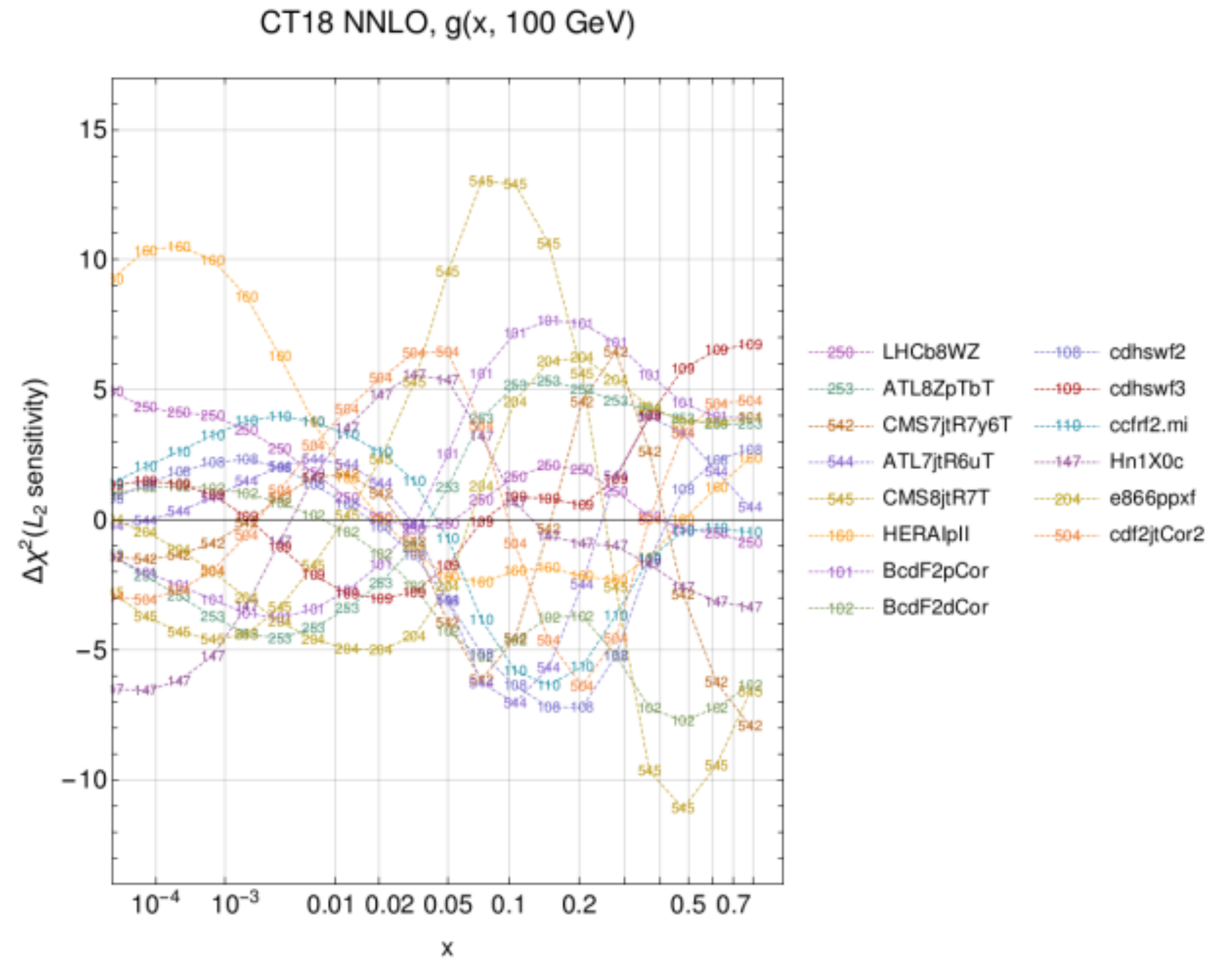
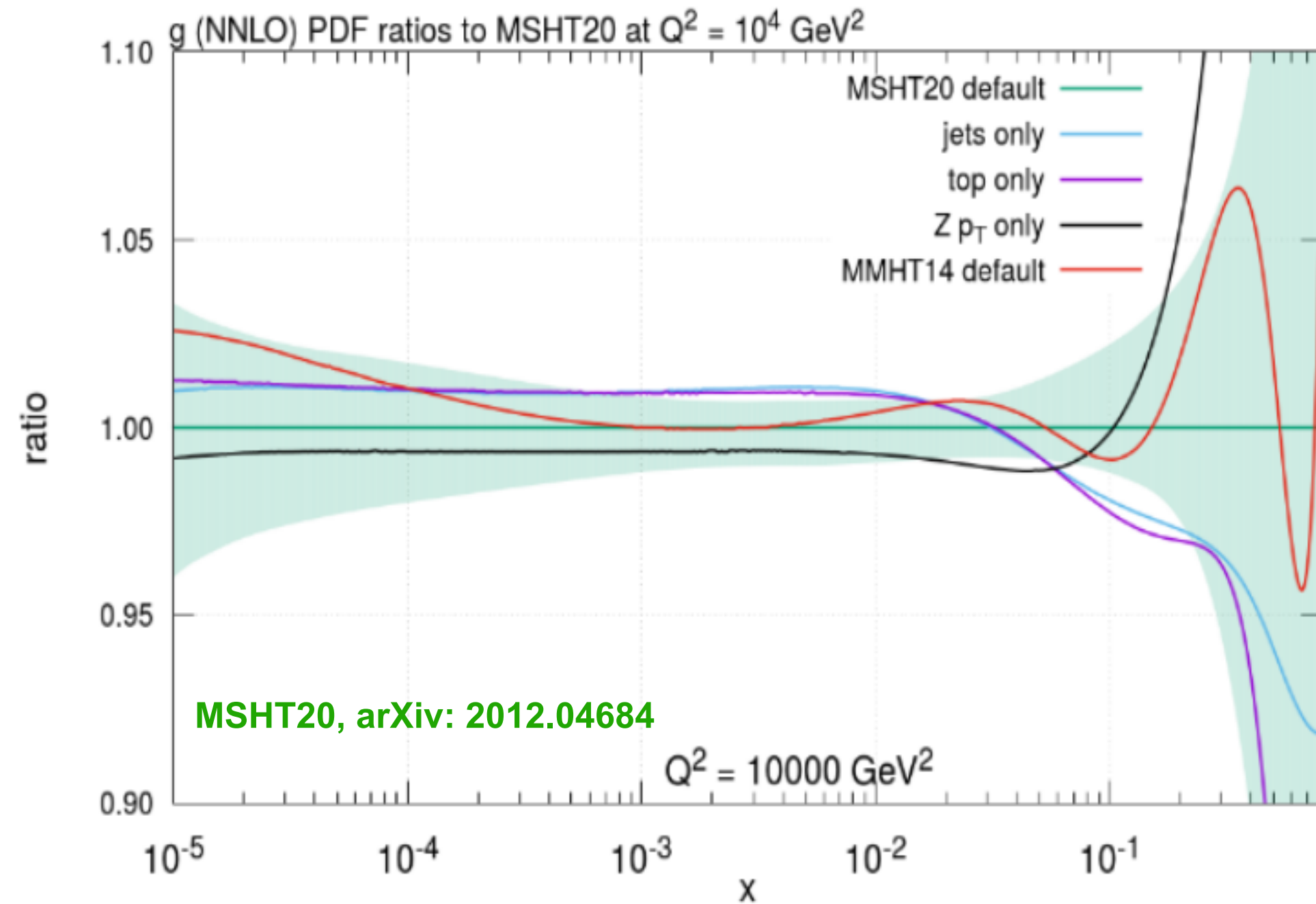


Ball et al, 2109.02653

- **HERA** - Valence quarks, some sea, low x gluon.
- **Fixed Target** (and others) - Flavour decomposition across intermediate to small x and gluon at small x
- **LHC** - Sea quarks, flavour decomposition, medium-large x gluon.
- Some **tension** with older fixed-target Drell-Yan and DIS data visible in the large- x region (especially gluon and anti-quarks)



EXPERIMENTAL DATA



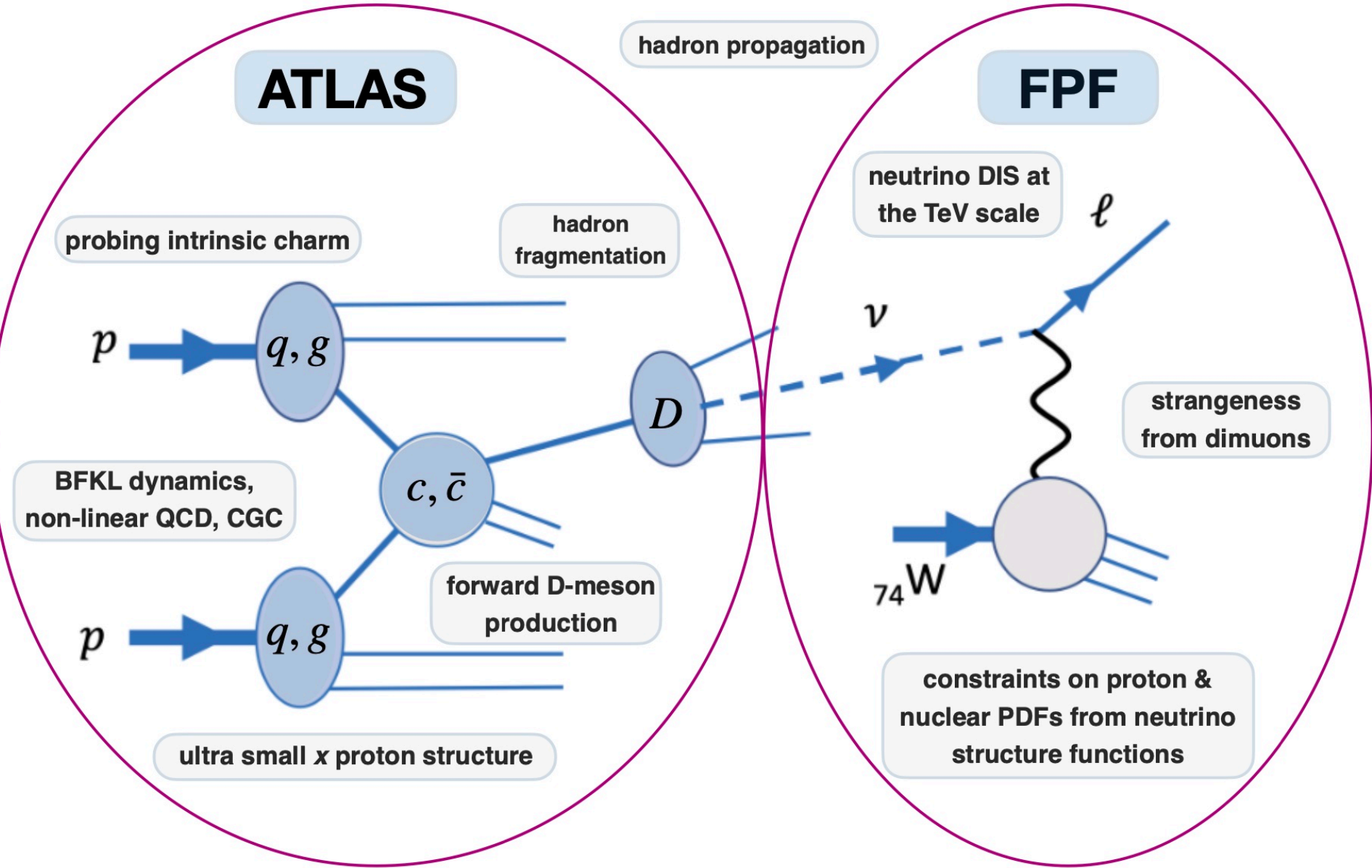
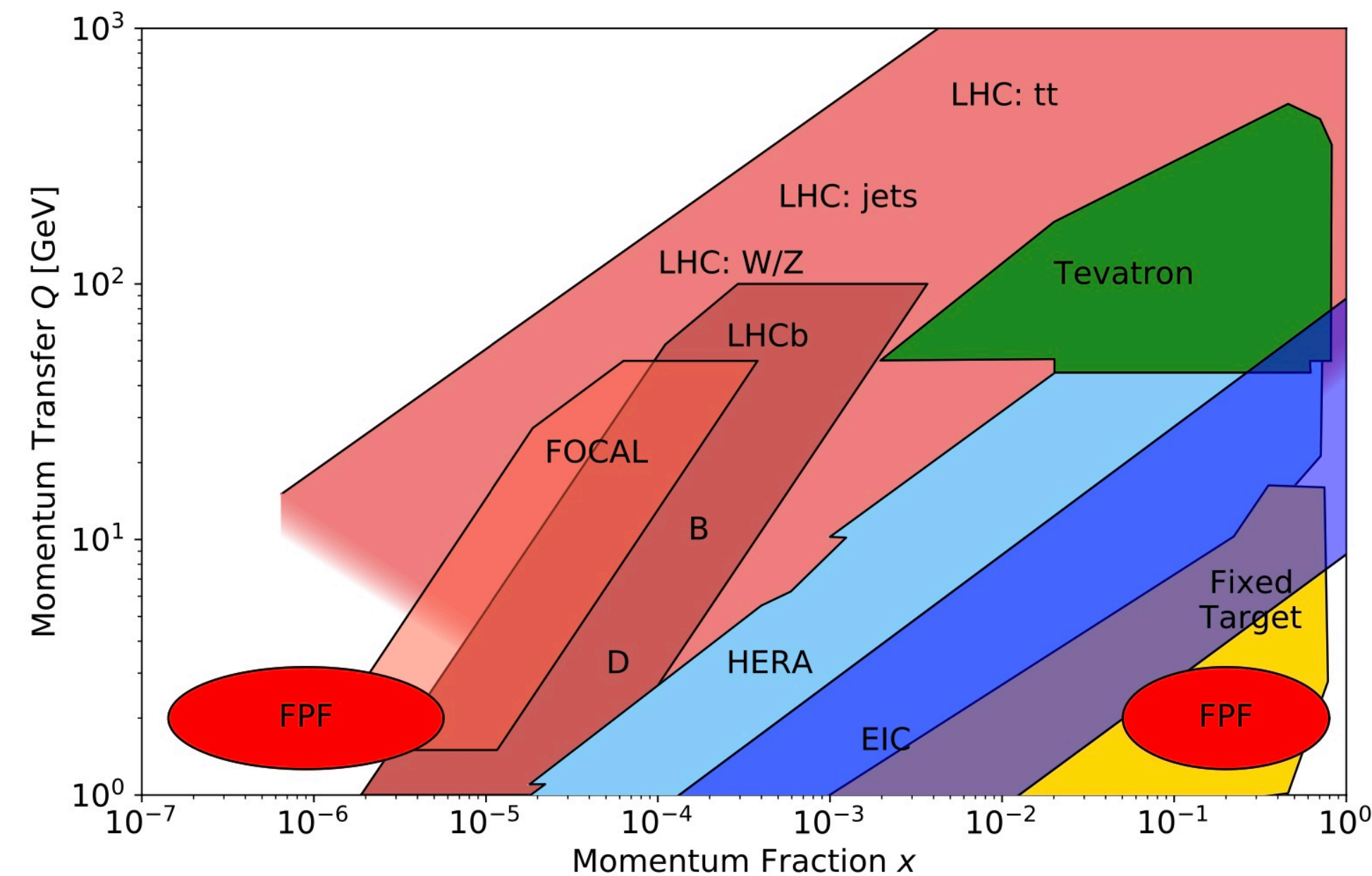
- **HERA** - Valence quarks, some sea, low x gluon.
- **Fixed Target** (and others) - Flavour decomposition across intermediate to small x and gluon at small x
- **LHC** - Sea quarks, flavour decomposition, medium-large x gluon.
- Some **tension** with older fixed-target Drell-Yan and DIS data visible in the large- x region (especially gluon and anti-quarks)

CT18, arXiv: 1912.10053 + joint CT-MSHT-exp study in 2306.03918

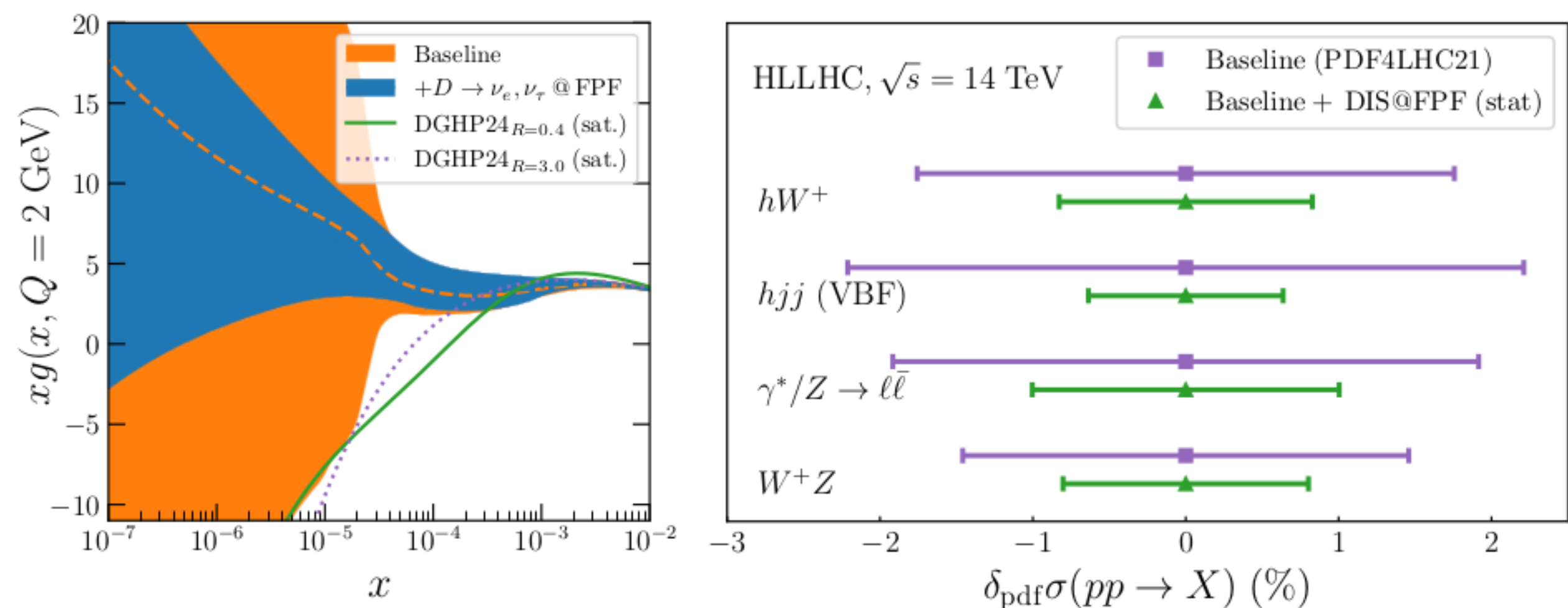
LOOKING AHEAD: FORWARD PHYSICS FACILITY

- Exploit HL-LHC produced neutrinos.
- Production $\rightarrow$ charm, gluon at high and low x.
- Interaction in ν DIS $\rightarrow$ strangeness.
- Significant uncertainty reductions and knock-on impacts on key LHC processes.
- Complementary with HL-LHC. Informs also FCC

FPF Snowmass arXiv:2203.05090



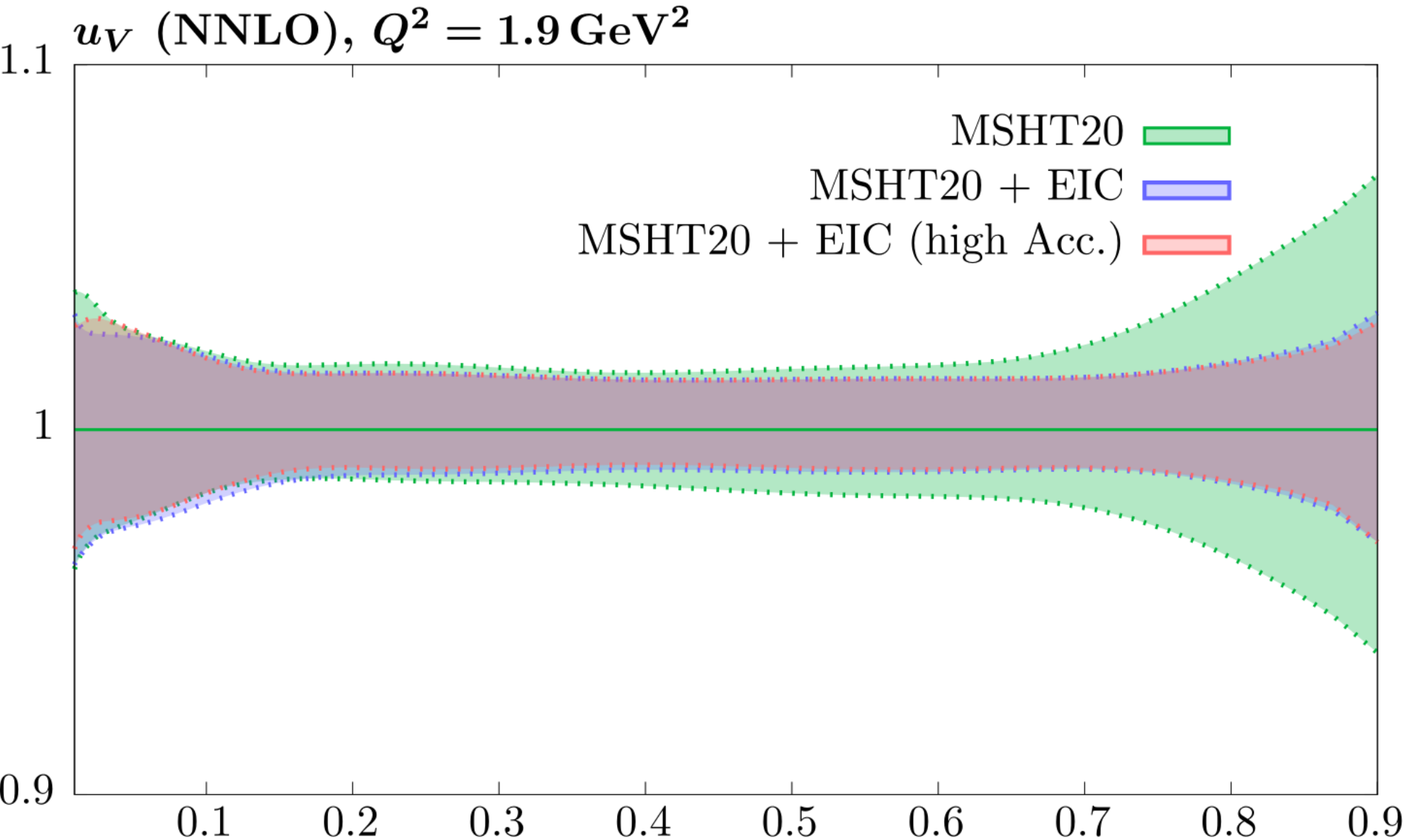
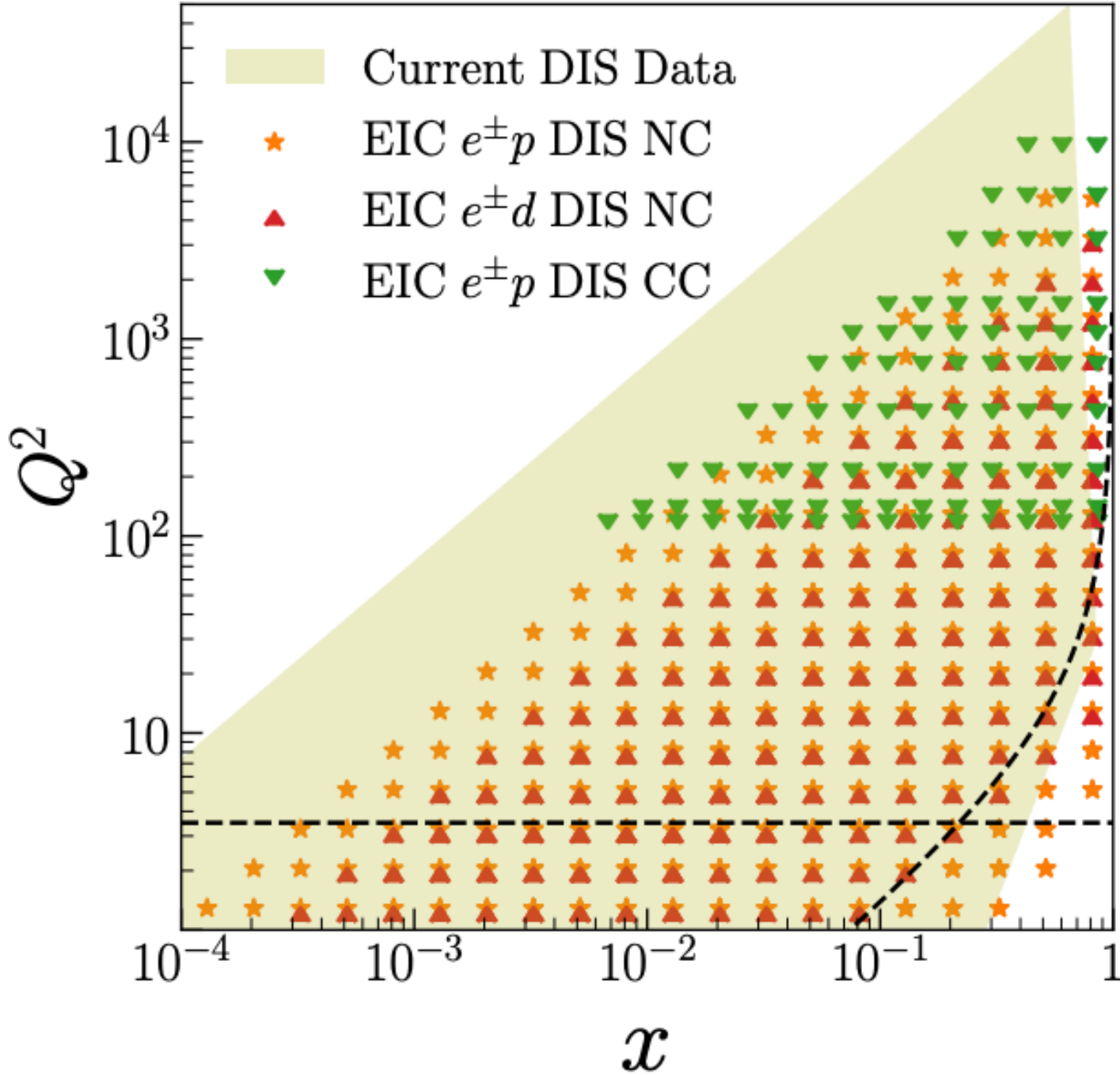
FPF Input to EPPSU arXiv:2503.19010



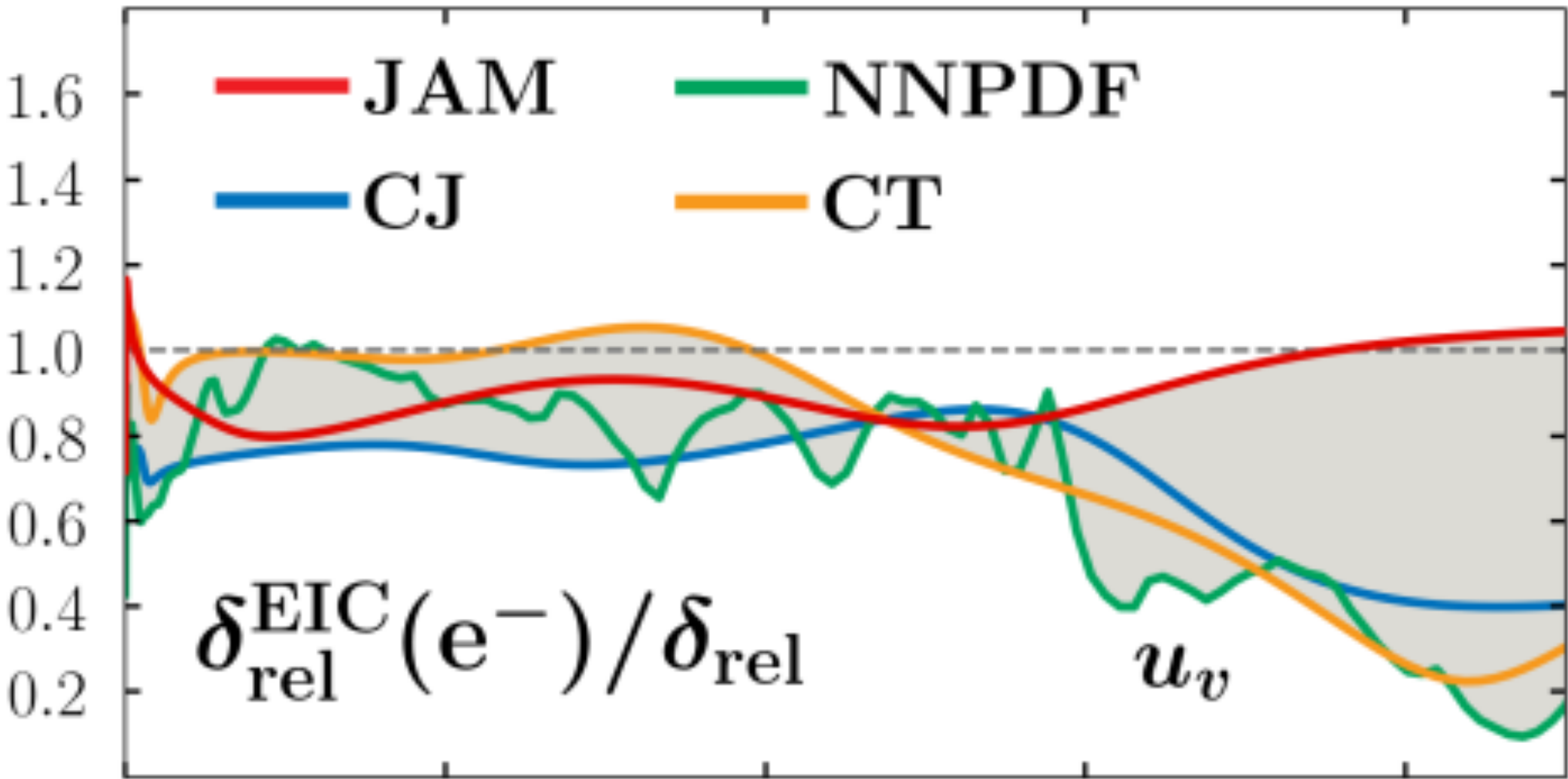
LOOKING AHEAD: ELECTRON ION COLLIDER

- EIC clean DIS environment - **precise, high x / moderate Q²**.
- Less sensitive to low Q²/W² corrections.
- Significant uncertainty reductions in **high x quarks**, smaller reduction in **gluon** uncertainty - BSM/Higgs.
- Further constraints from deuteron (or positron) data, SIDIS, FL etc. Also scope to constrain α_s .
- Complementary with HL-LHC.
- Potential for DIS only PDFs - simpler systematics.

Khalek et al, 2102.00018



Armesto et al arXiv:2309.11269

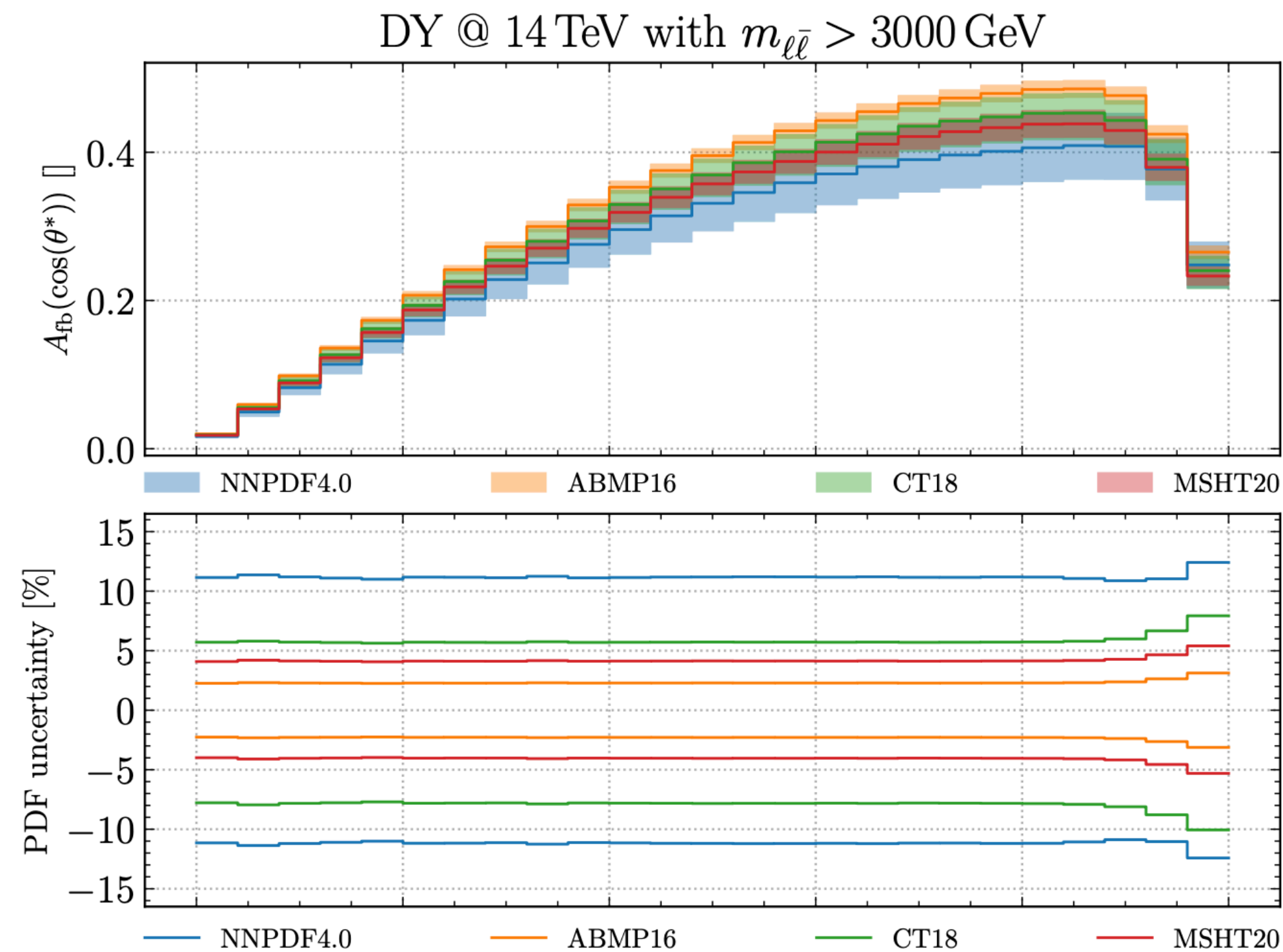


EIC Yellow Report: arXiv:2103.05419

See also EIC
Snowmass inputs
arXiv:2203.13199,
2203.13923

EIC AND LHC CONNECTION #2: PREVENT LARGE-X BSM CONTAMINATION

PDFS AND NEW PHYSICS SEARCHES



e. g. Heavy Z' search in DY Forward-Backward asymmetry

Ball et al arXiv: 2209.08115

$$x \approx \frac{M}{\sqrt{S}}$$

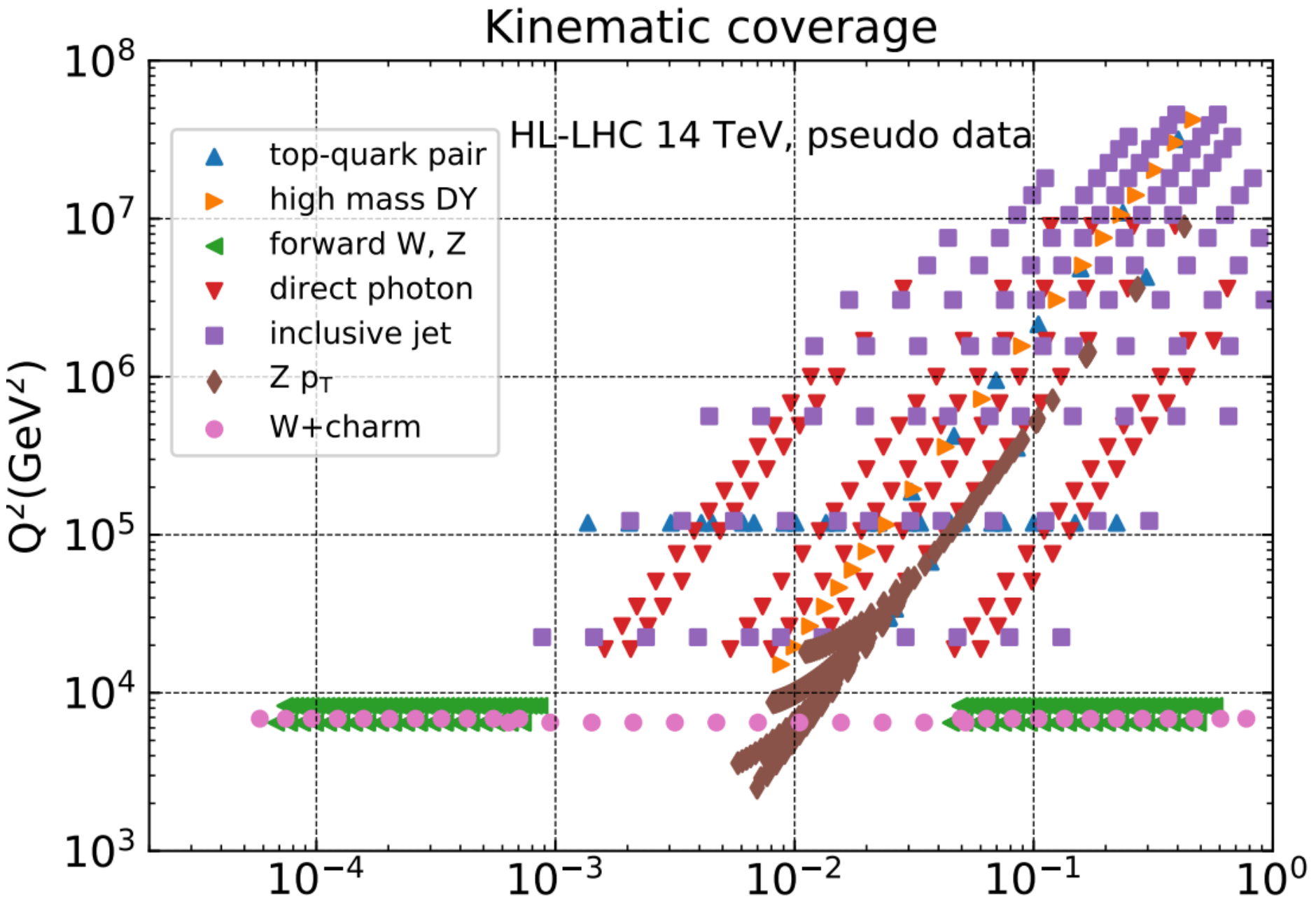
High-mass final states

$\Leftrightarrow$

Large-x PDFs

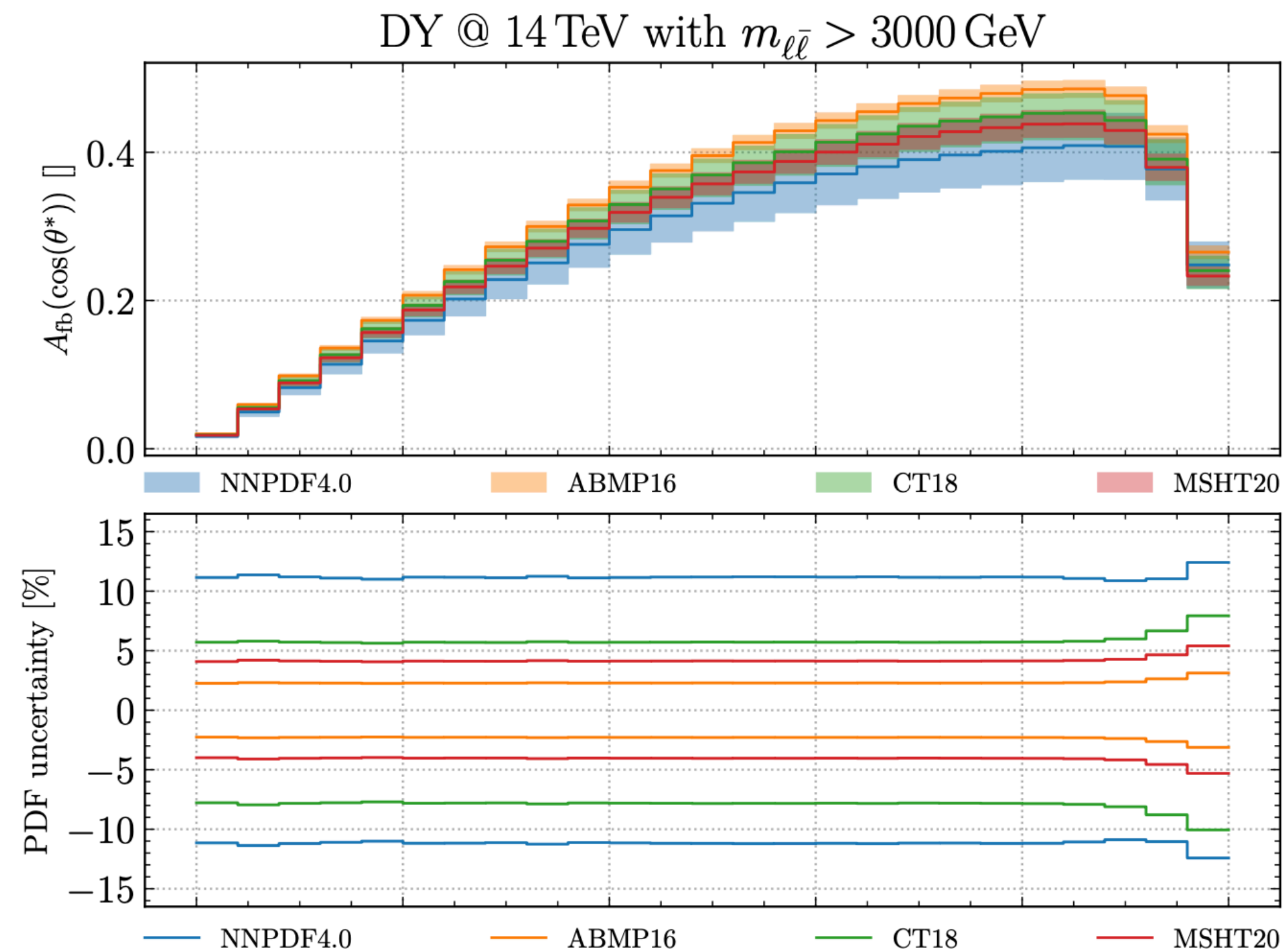
- Large-x PDF affected by large uncertainties
- Reducing PDF uncertainties in large-x region is essential to characterise any possible new physics and to interpret deviations in the tails of experimental distributions
- Looking forward: **HL-LHC (and FCC-hh)**

Large x $\leftrightarrow$ Large E and/or Large Y



x Khalek et al, 1810.03639

PDFS AND NEW PHYSICS SEARCHES

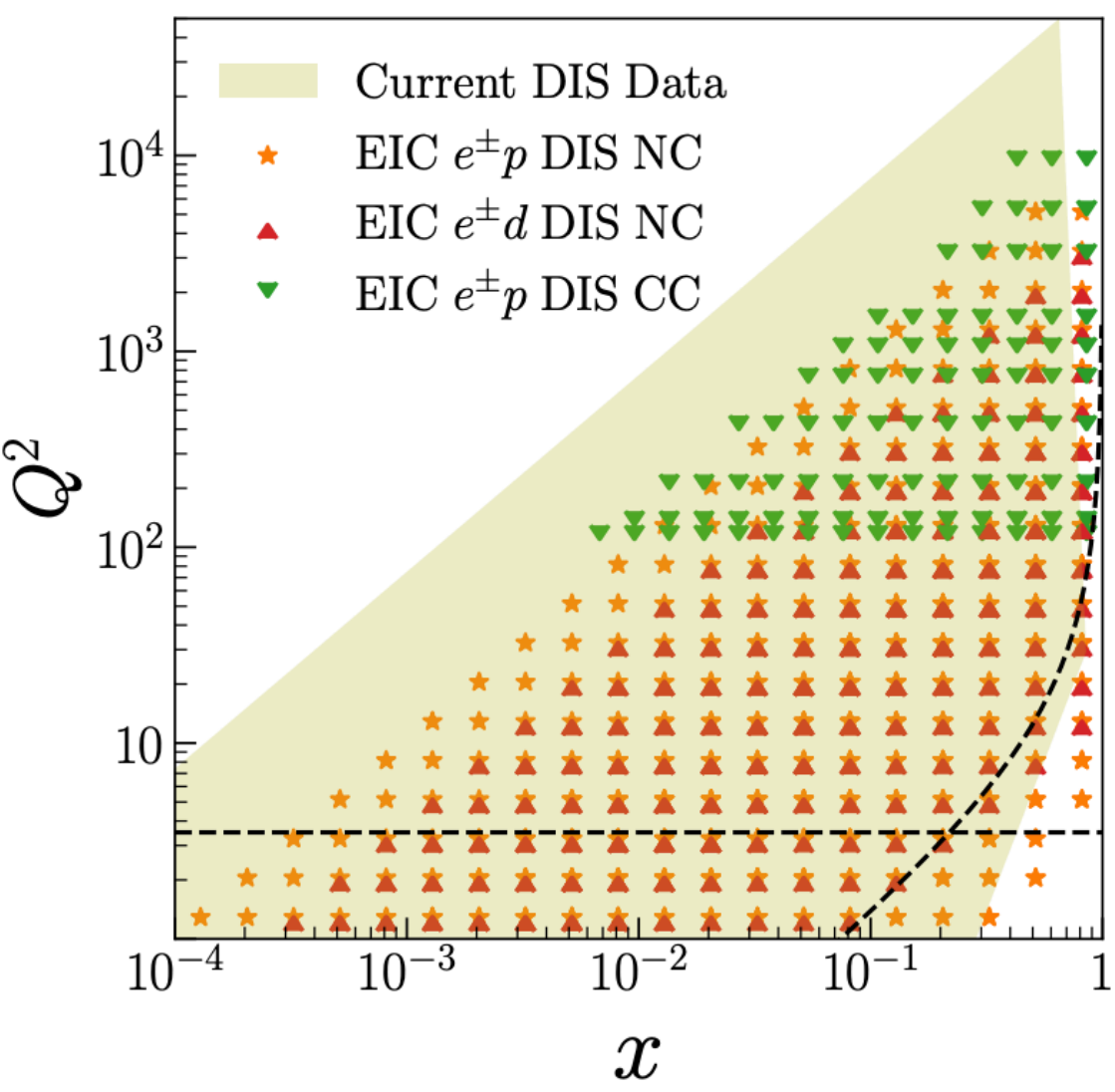


e. g. Heavy Z' search in DY Forward-Backward asymmetry

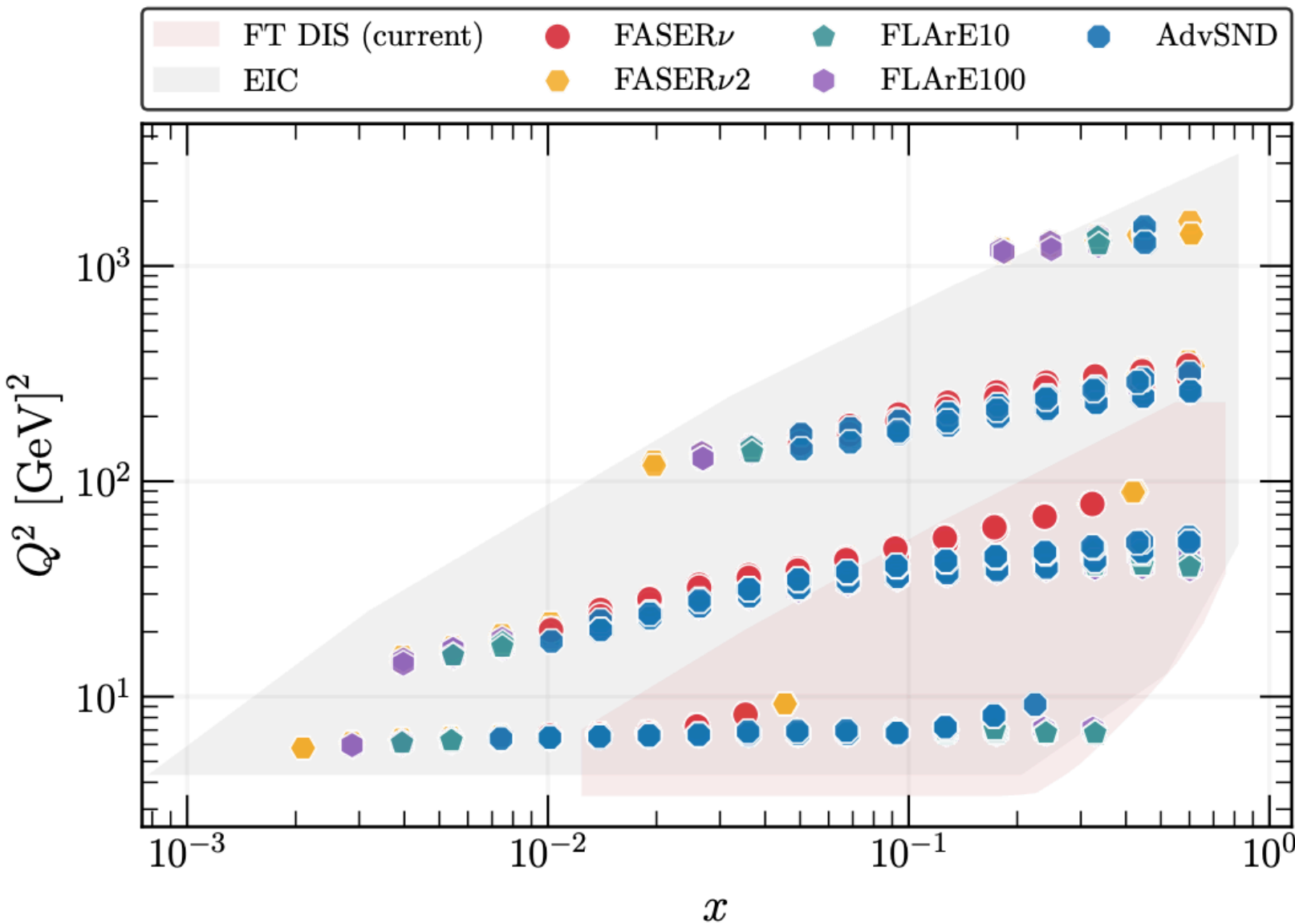
Ball et al arXiv: 2209.08115

- Large-x PDF affected by large uncertainties
- Reducing PDF uncertainties in large-x region is essential to characterise any possible new physics and to interpret deviations in the tails of experimental distributions
- Looking forward: **JLab, EIC, Forward Physics Facilities...**

Large x $\leftrightarrow$ Low/Medium E

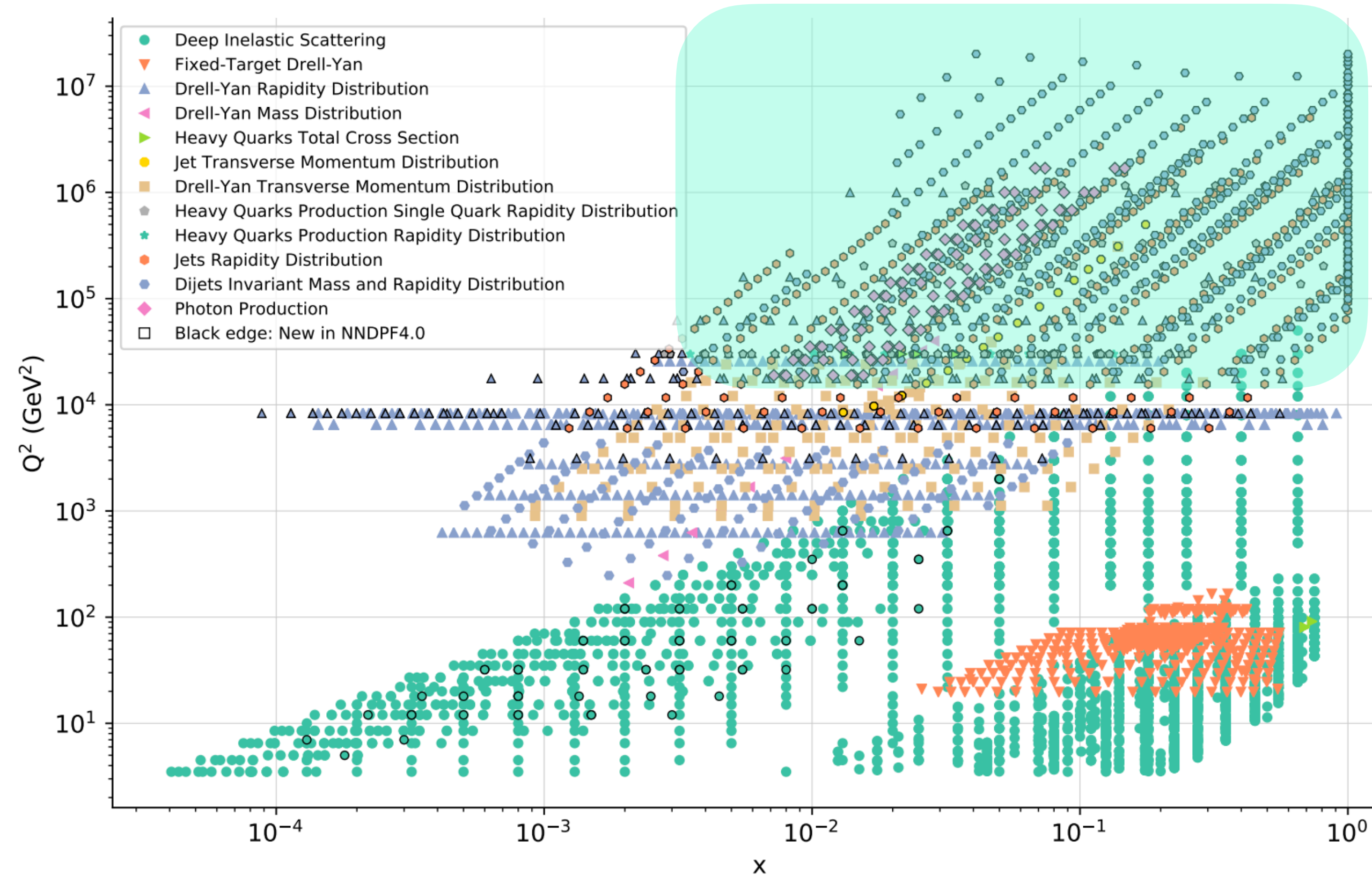


Khalek et al, 2102.00018

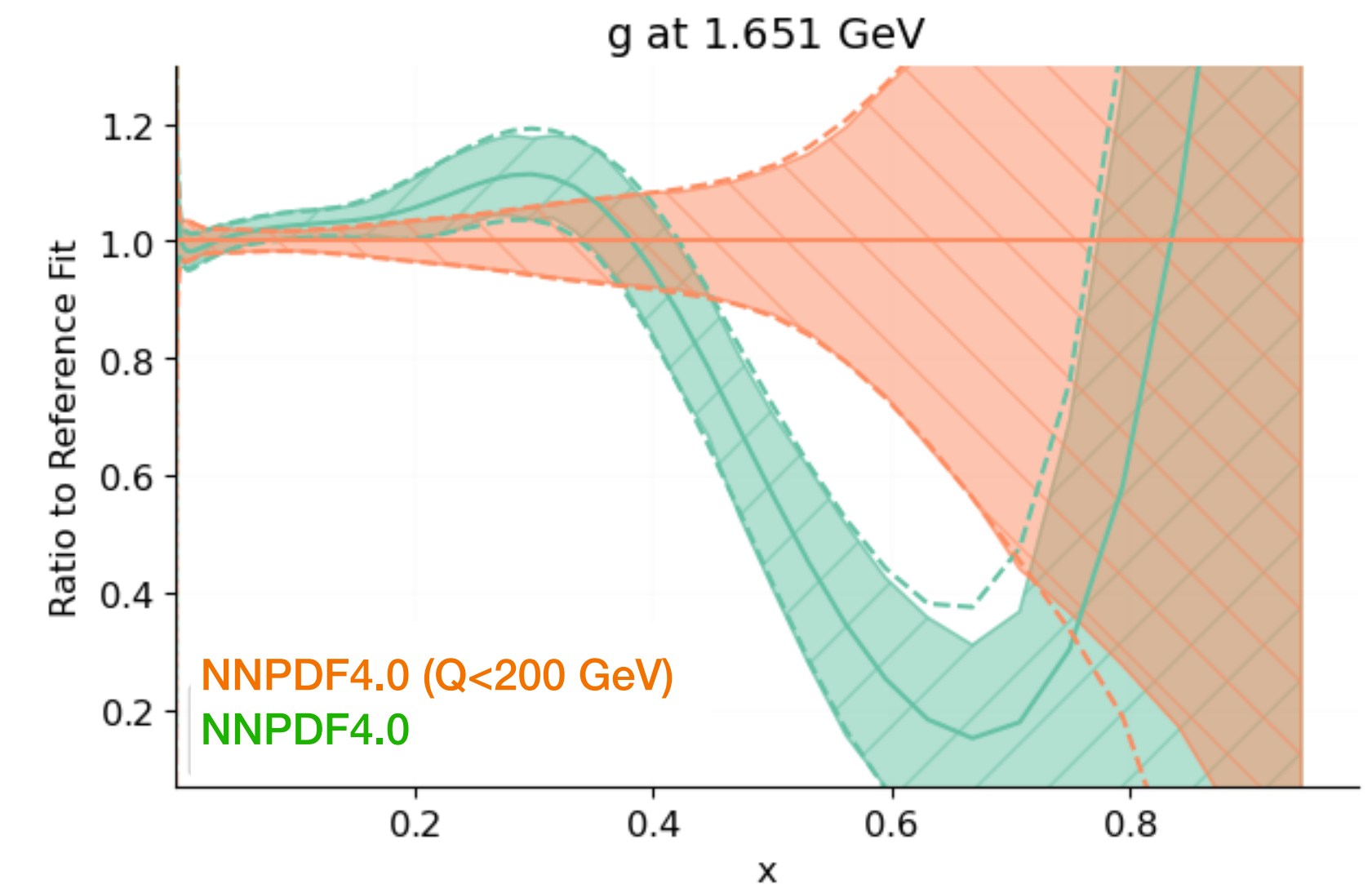
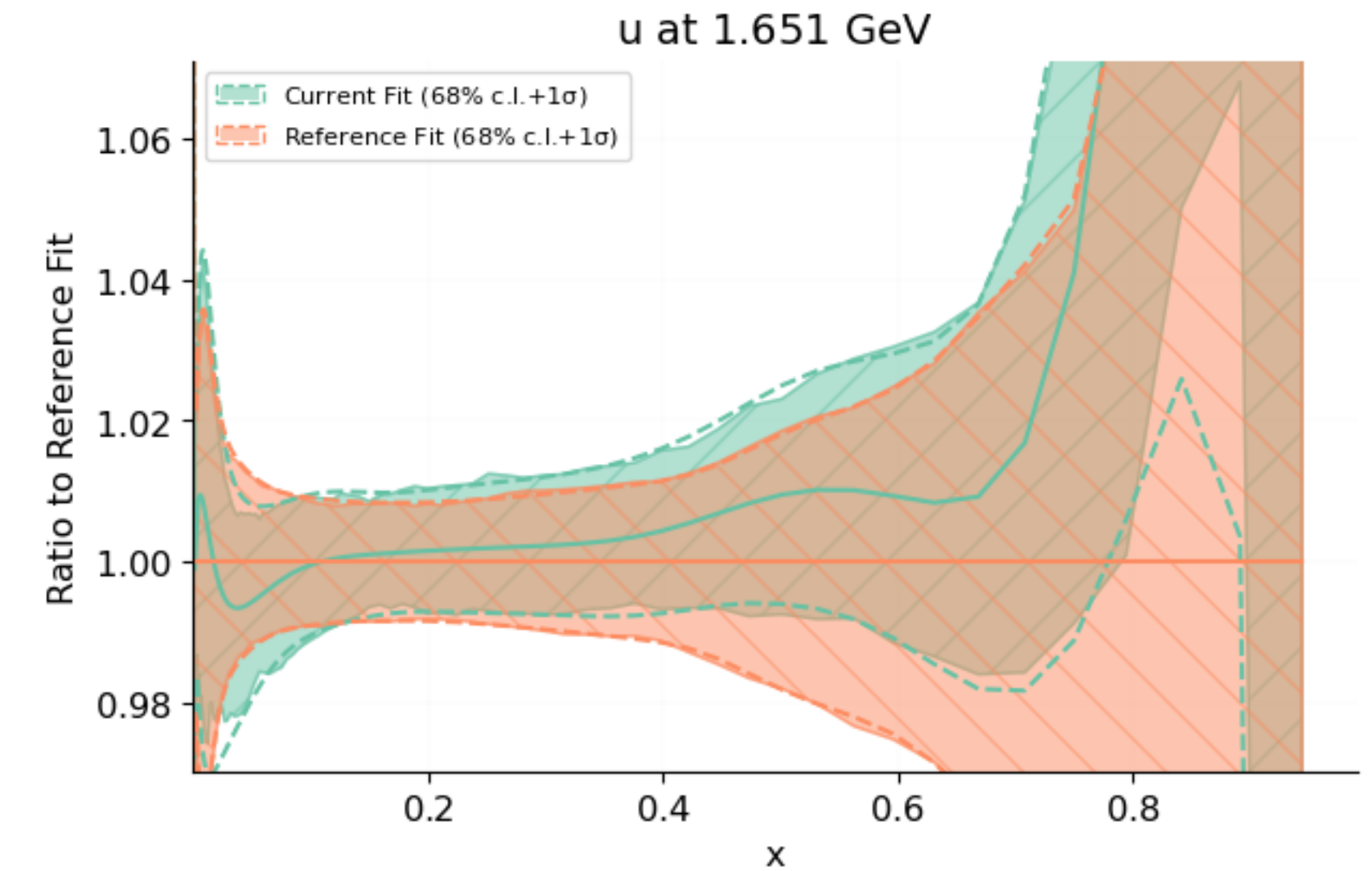


J Cruz-Martinez et al 2309.09581

HOW TO MAKE SURE THAT THE INPUT DATA ARE FREE OF POSSIBLE BSM EFFECTS?



- Naive solution: remove all data above $Q > Q_{\text{threshold}}$
- Cutting away all high energy data would result in dramatic uncertainty increase and loss of accuracy, as we would rely on old data and on PDF fitters' ability to control extrapolation region.



PDF AND BSM INTERPLAY

➡ From the point of view of PDF fits:

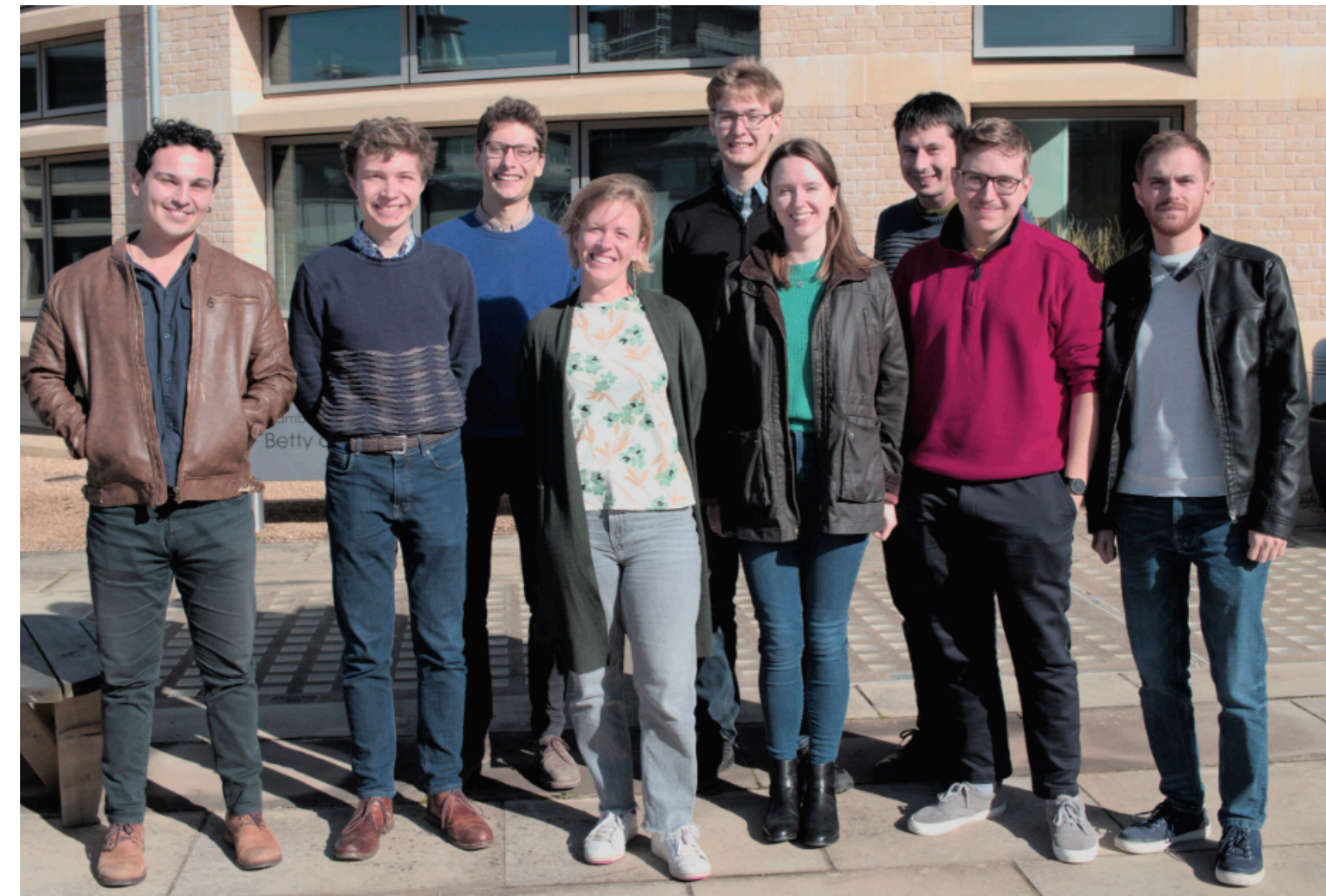
- How to make sure that new physics effects are not inadvertently fitted away in a PDF fit?
- Antiquarks ✓ [PBSP + Mangano: 2307.10370] Gluon ≈ [PBSP + Gomez-Ambrosio, Nocera, in progress]

➡ From the point of view of BSM indirect searches (or specifically in SMEFT fits):

- What PDFs should be used? Are conservative PDFs good enough? What about flavour? ≈ [PBSP, in progress]
- How would the bounds change if PDF fits included the same operators that are fitted in SMEFT fit?
 - ✓ [Top sector: PBSP + Rojo, 2303.06159]
 - ✓ [Drell-Yan sector: PBSP + Greljo : 2104.02723, Iranipour + MU: 2201.07240]
 - ≈ [Jet and Dijets: Greljo, Hammou, Merlotti, Smolkovic, MU, in progress]



MU + Zahari Kassabov, Maeve Madigan, Luca Mantani, James Moore, Manuel Morales Alvarado, Ella Cole, Elie Hammou, Mark Costantini, Francesco Merlotti



PDF AND BSM INTERPLAY

➔ From the point of view of PDF fits:

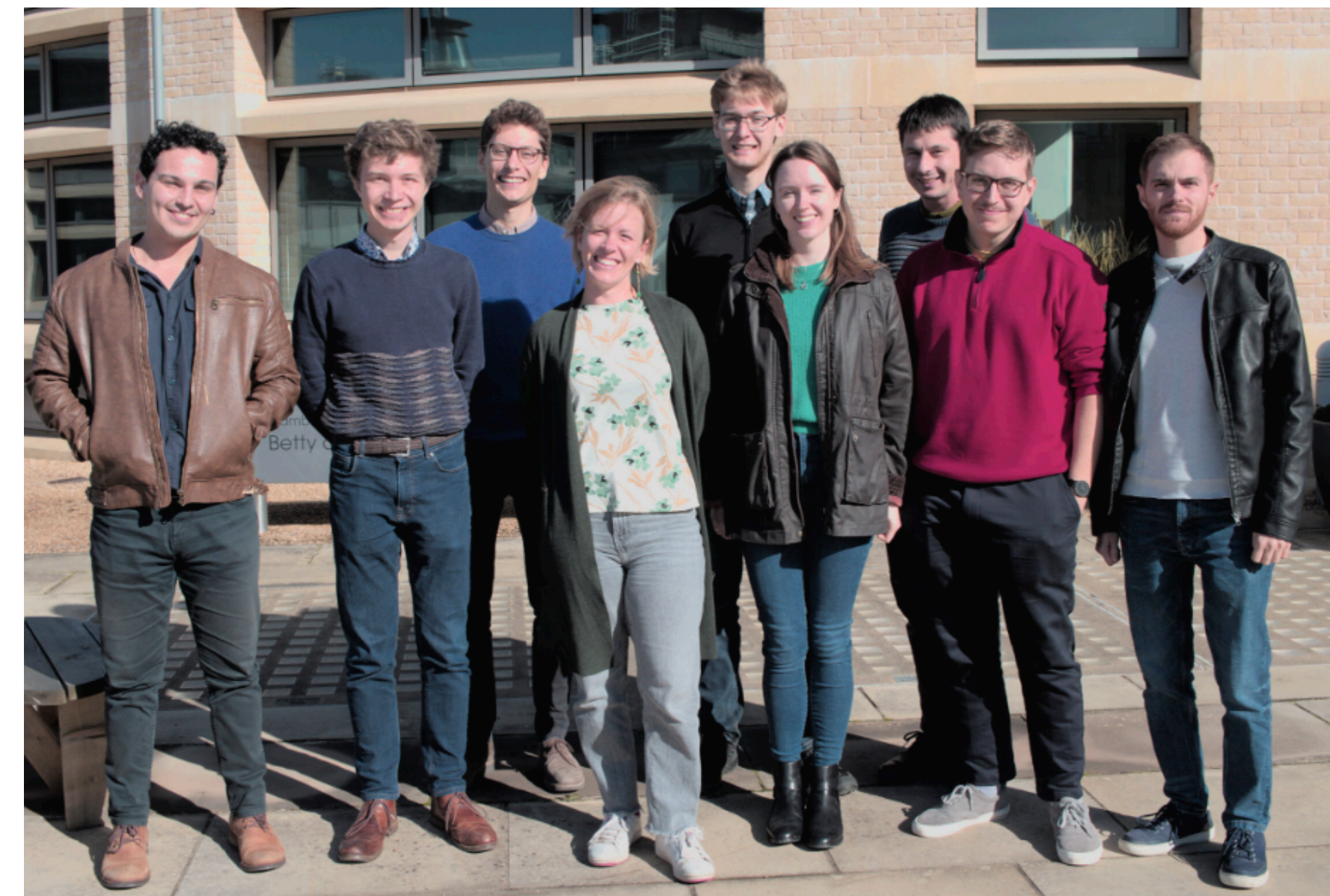
- How to make sure that new physics effects are not inadvertently fitted away in a PDF fit?
- Antiquarks ✓ [PBSP + Mangano: 2307.10370] Gluon ≈ [PBSP + Gomez-Ambrosio, Nocera, in progress]

➔ From the point of view of BSM indirect searches (or specifically in SMEFT fits):

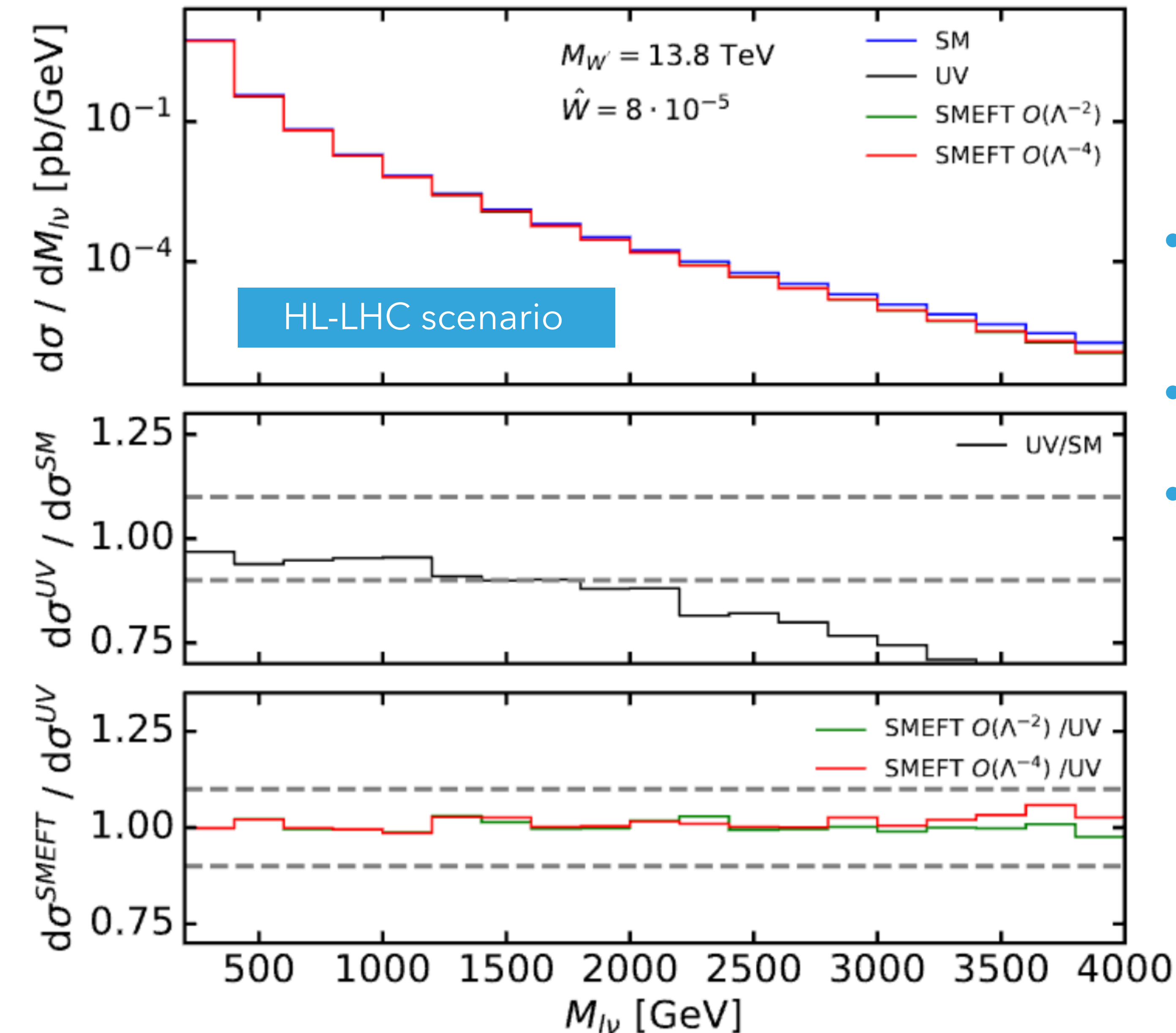
- What PDFs should be used? Are conservative PDFs good enough? What about flavour? ≈ [PBSP, in progress]
- How would the bounds change if PDF fits included the same operators that are fitted in SMEFT fit?
 - ✓ [Top sector: PBSP + Rojo, 2303.06159]
 - ✓ [Drell-Yan sector: PBSP + Greljo : 2104.02723, Iranipour + MU: 2201.07240]
 - ≈ [Jet and Dijets: Greljo, Hammou, Merlotti, Smolkovic, MU, in progress]



MU + Zahari Kassabov, Maeve Madigan, Luca Mantani, James Moore, Manuel Morales Alvarado, Ella Cole, Elie Hammou, Mark Costantini, Francesco Merlotti



CAN PDFS ABSORB NEW PHYSICS?



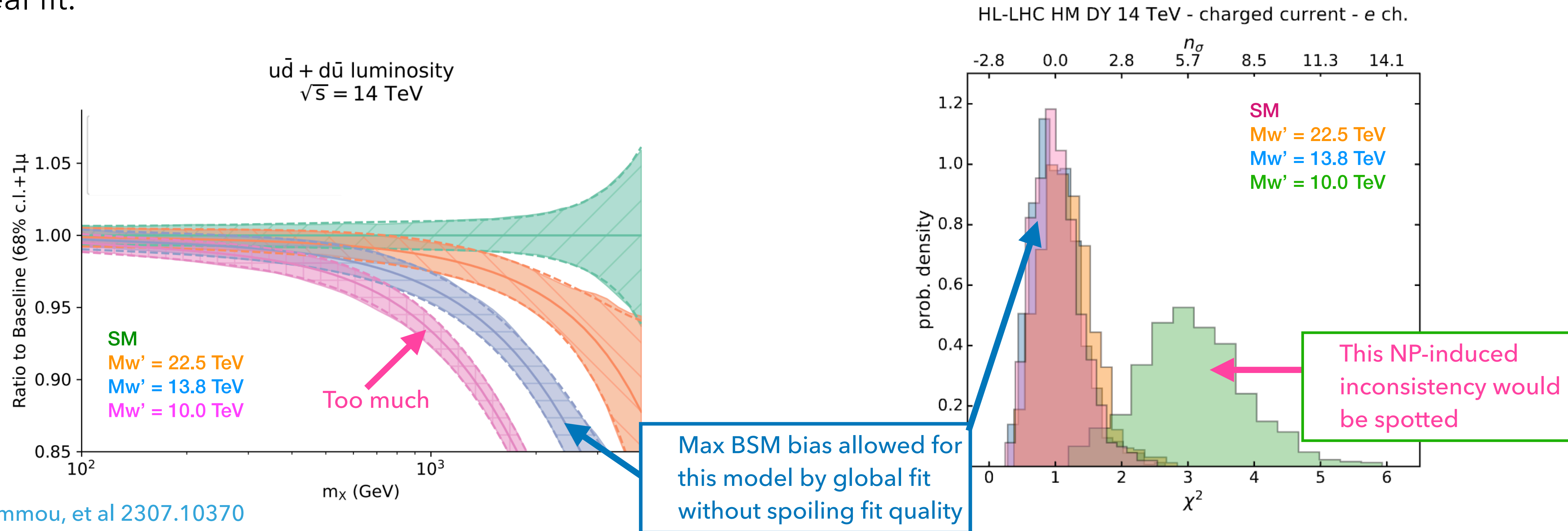
- Imagine to generate artificial data injecting a given underlying law:
"True" law of nature =
"True" PDFs + "True" BSM model
- Can inject directly UV model or SMEFT parametrisation in the region in which SMEFT approximation is good
- Fit PDFs **assuming SM**
- If fit quality of BSM (data) with SM (theory) does not deteriorate with respect to SM (data) with SM (theory) and PDFs deviate from SM PDFs then the BSM signal is absorbed by the PDFs!

$$\checkmark \quad \mathcal{L}_{SMEFT}^{W'} = \mathcal{L}_{SM} - \frac{g_{W'}^2}{2M_{W'}^2} J_L^{a,\mu} J_{L,\mu}^a \quad J_L^{a,\mu} = \sum_f \bar{f}_L T^a \gamma^\mu f_L$$

$$\times \quad \mathcal{L}_{SMEFT}^{Z'} = \mathcal{L}_{SM} - \frac{g_{Z'}^2}{2M_{Z'}^2} J_Y^\mu J_{Y,\mu} \quad J_Y^\mu = \sum_f Y_f \bar{f} \gamma^\mu f$$

CAN PDFS ABSORB NEW PHYSICS? YES!

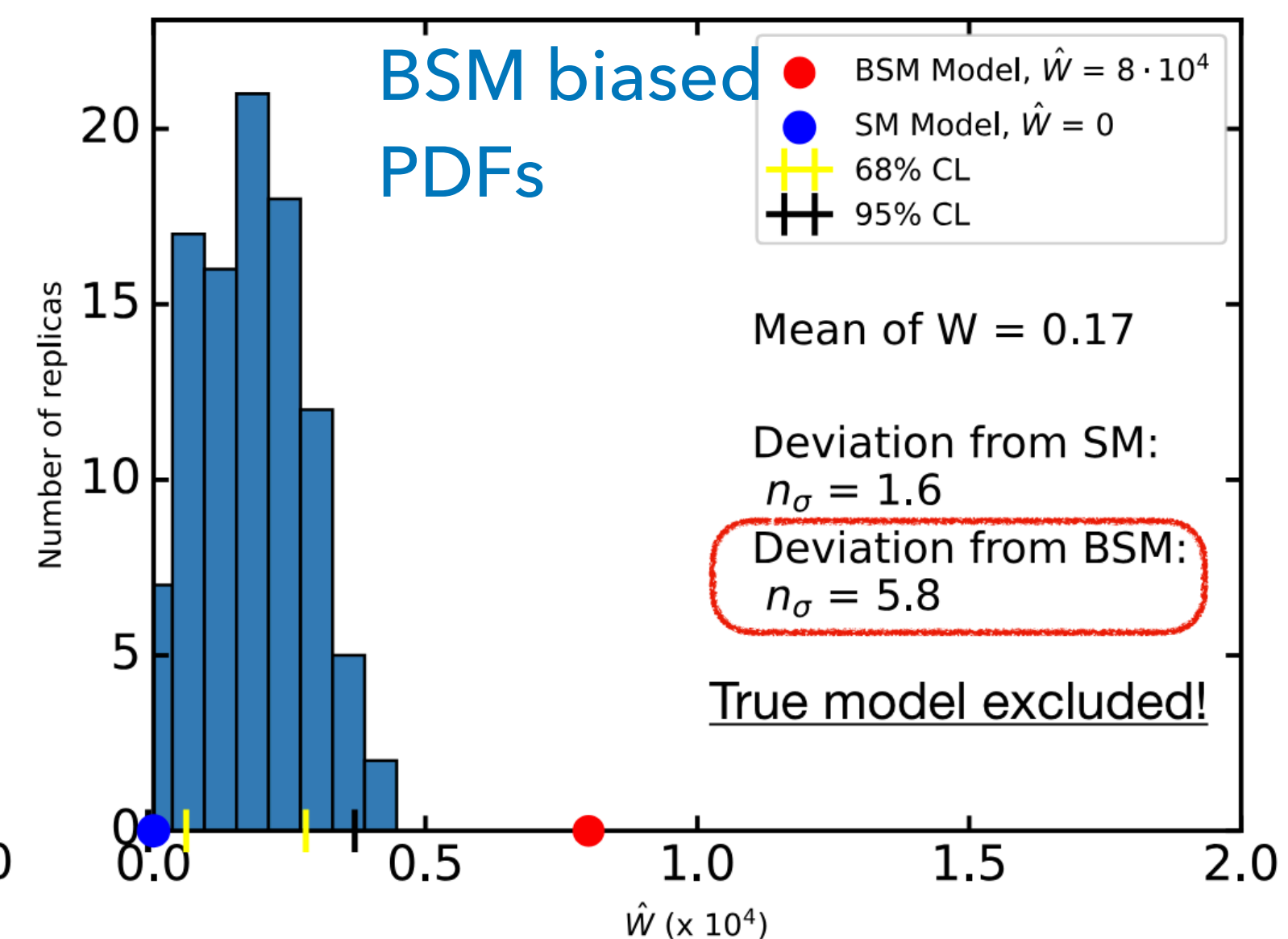
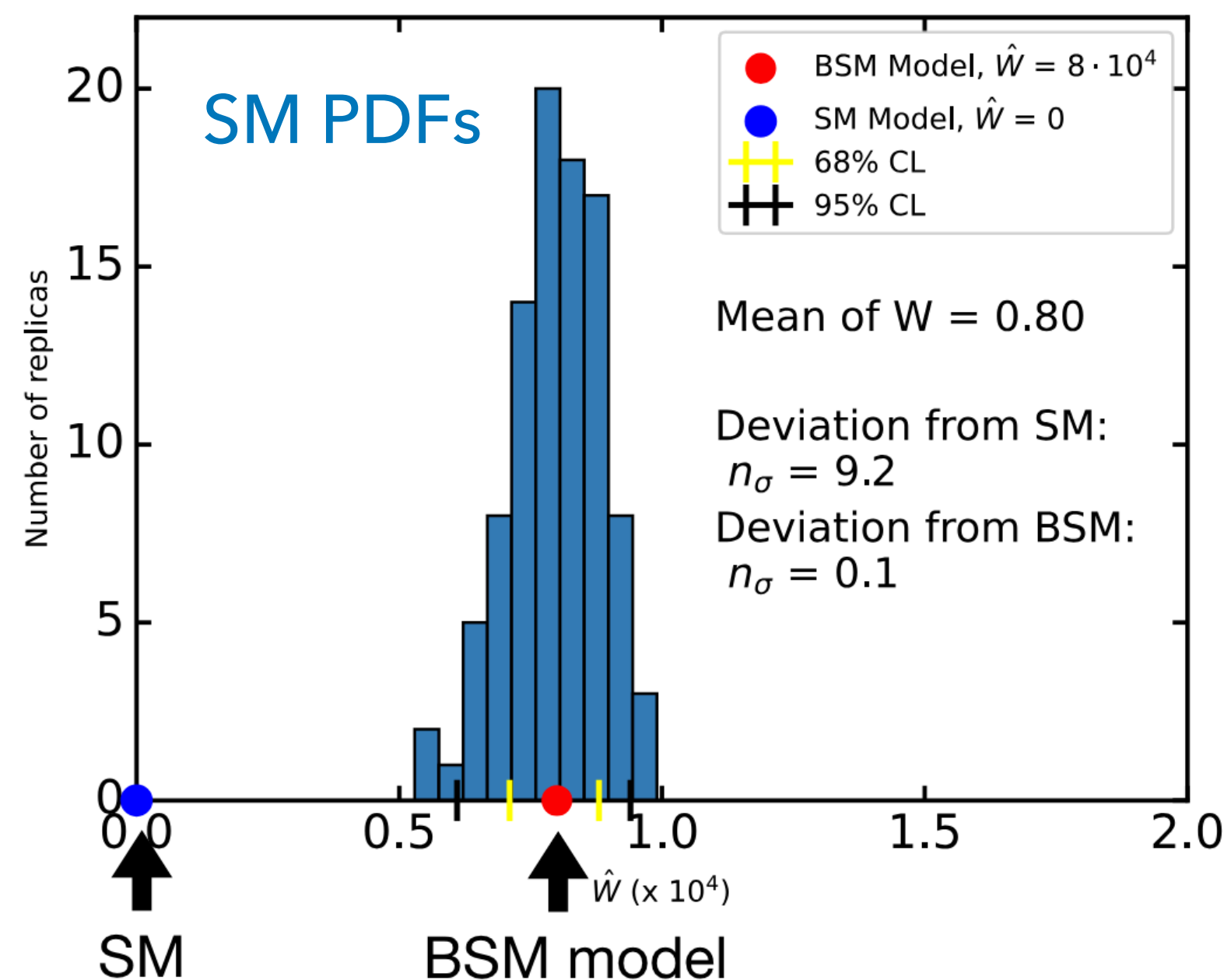
- The fit-quality of the global fit is unchanged even with signal from $M_{W'} = 13.8$ TeV injected in all data (mostly visible in HL-LHC NC and CC Drell-Yan data)
- Once we go beyond this point, the fit-quality deteriorates due to the HL-LHC neutral current and charged current Drell-Yan artificial data.
- Already for $M_{W'} = 13.8$ TeV the $q\bar{q}$ luminosity shifts far beyond the PDF uncertainties because anti-quark PDFs at large- x compensate or “fit away” the effect of New Physics and we would not know in a real fit.



PHENOMENOLOGICAL CONSEQUENCES

- The fit-quality of the global fit is unchanged even with signal from $M_{W'} = 13.8$ TeV injected in all data (mostly visible in HL-LHC NC and CC Drell-Yan data)
- Once we go beyond this point, the fit-quality deteriorates due to the HL-LHC neutral current and charged current Drell-Yan artificial data.
- Already for $M_{W'} = 13.8$ TeV the $qq \sim$ luminosity shifts far beyond the PDF uncertainties because anti-quark PDFs at large- x compensate or “fit away” the effect of New Physics and we would not know in a real fit.

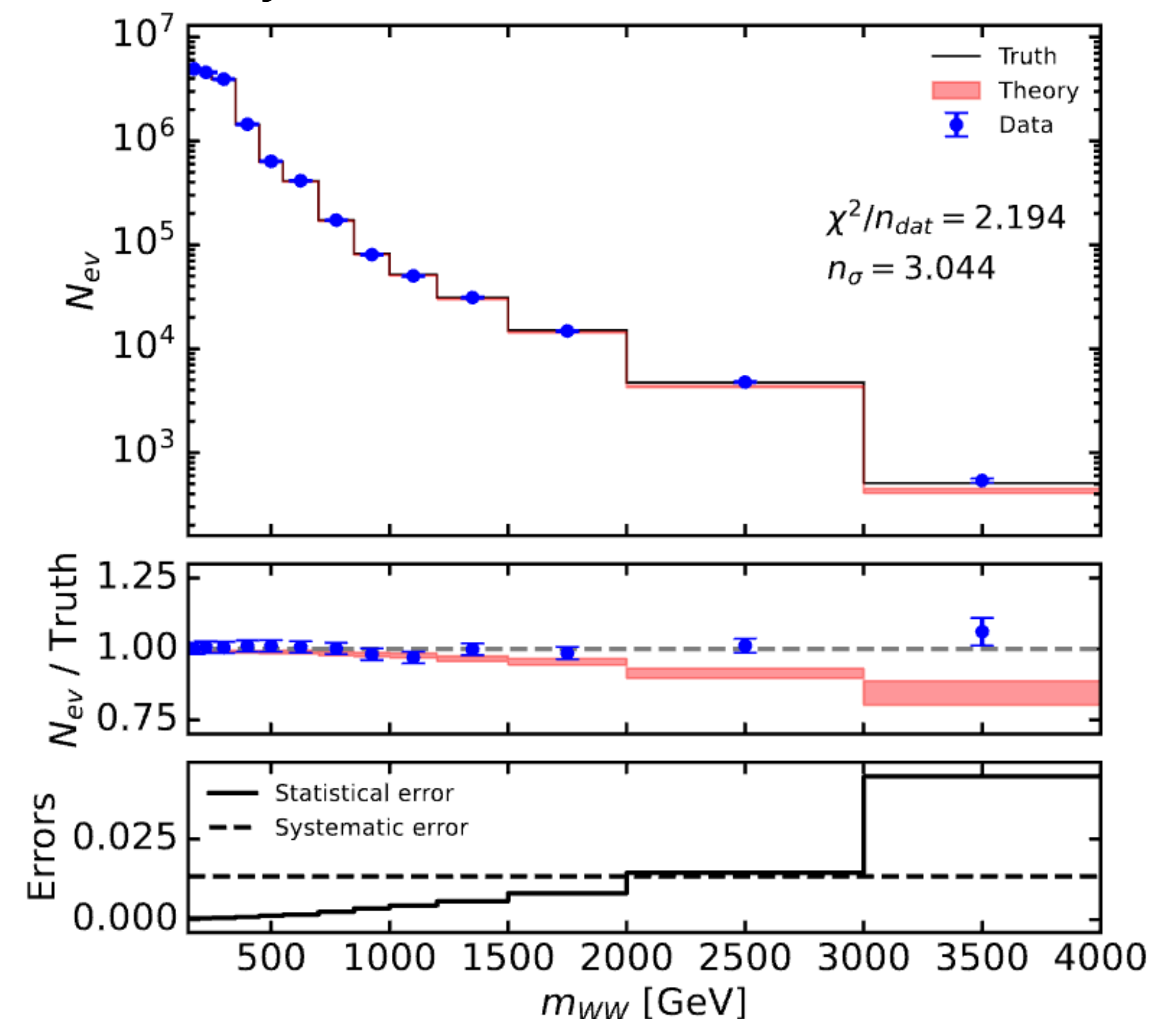
Consequence #1:
Would not see indirect effect of new physics as we would find SMEFT bounds compatible with the SM!



PHENOMENOLOGICAL CONSEQUENCES

- The fit-quality of the global fit is unchanged even with signal from $M_{W'} = 13.8$ TeV injected in all data (mostly visible in HL-LHC NC and CC Drell-Yan data)
- Once we go beyond this point, the fit-quality deteriorates due to the HL-LHC neutral current and charged current Drell-Yan artificial data.
- Already for $M_{W'} = 13.8$ TeV the $qq \sim$ luminosity shifts far beyond the PDF uncertainties because anti-quark PDFs at large- x compensate or “fit away” the effect of New Physics and we would not know in a real fit.

Consequence #2:
Would see New Physics effects where
there are none (for example in WW)



HOW TO AVOID NEW PHYSICS CONTAMINATION?

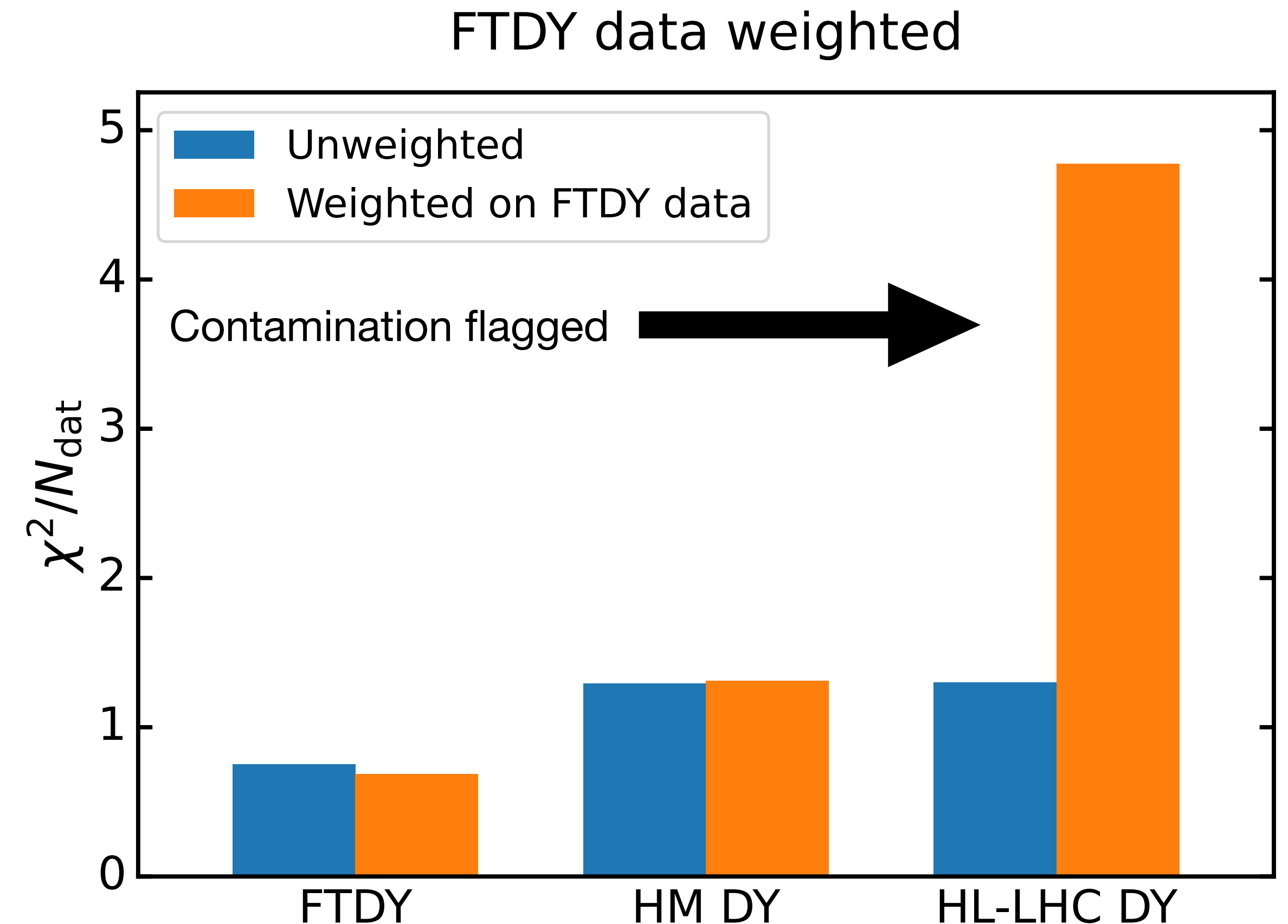
- LHCb on-shell at high rapidity data do not help as quark probed at large x , antiquark at small x
- If we fit giving more weight (artificially increasing statistics) of fixed-target DY data on proton/deuteron the BSM-induced inconsistency would be flagged
- Need more accurate low-energy/large- x constraining measurements to really disentangle such effects

Flat direction in the anti-quark at large- x :

- ➔ Accommodates PDFs to artificial data with BSM injected
- ➔ Allows for BSM bias in $qq \sim$ luminosity

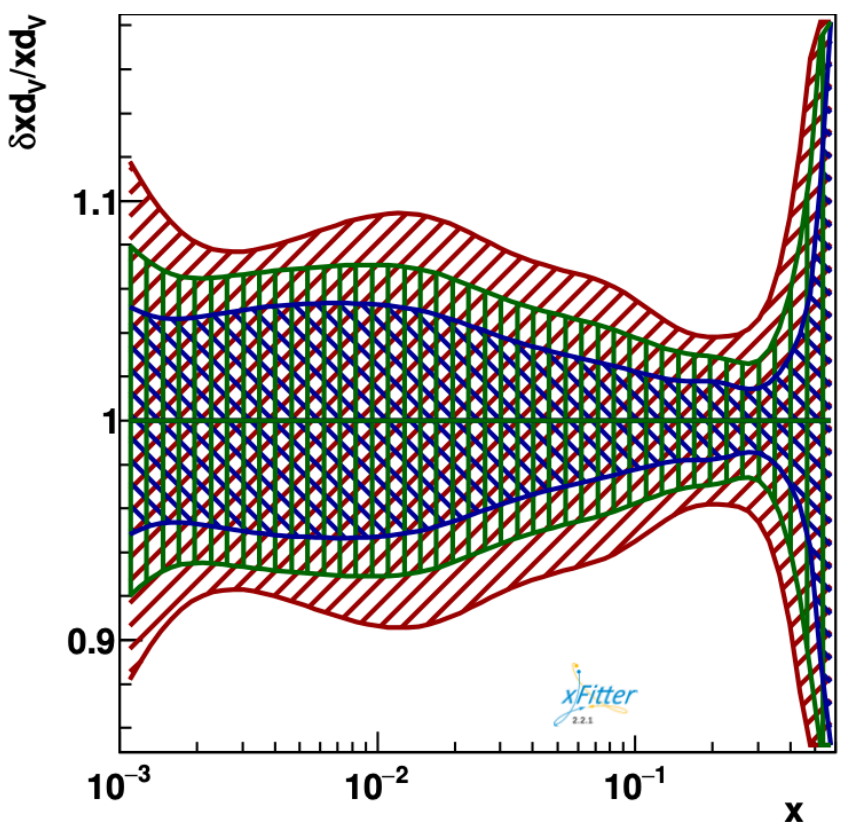
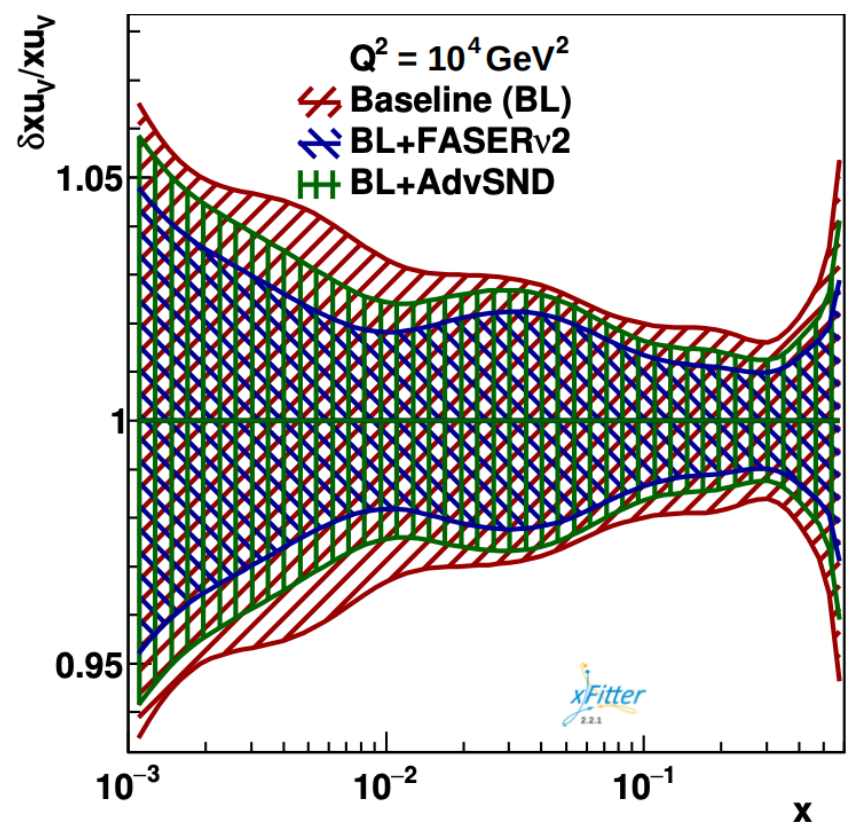
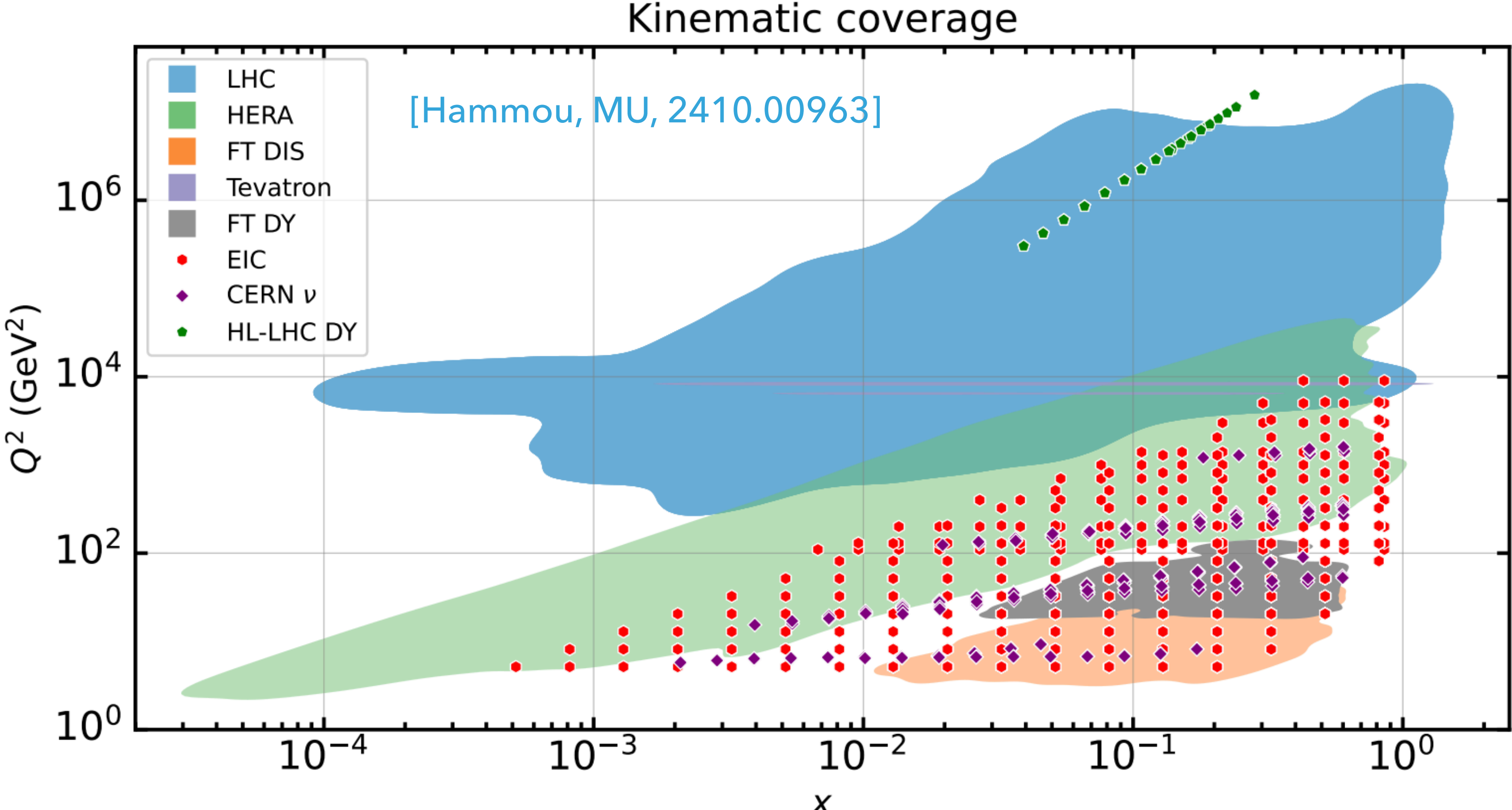
Including lower-energy large- x data:

- Constrain large- x region
- Safe from BSM contamination



PROJECTIONS FROM FPF

Observable	N_{dat}
FASERν	
$\tilde{\sigma}(\nu + N \rightarrow l^- + X)$	22
$\tilde{\sigma}(\bar{\nu} + N \rightarrow l^+ + X)$	16
FLArE	
$\tilde{\sigma}(\nu + N \rightarrow l^- + X)$	43
$\tilde{\sigma}(\bar{\nu} + N \rightarrow l^+ + X)$	39
$\tilde{\sigma}(\nu + N \rightarrow l^- + c/\bar{c} + X)$	31
$\tilde{\sigma}(\bar{\nu} + N \rightarrow l^+ + c/\bar{c} + X)$	19
FASERν2	
$\tilde{\sigma}(\nu + N \rightarrow l^- + X)$	44
$\tilde{\sigma}(\bar{\nu} + N \rightarrow l^+ + X)$	39
$\tilde{\sigma}(\nu + N \rightarrow l^- + c/\bar{c} + X)$	38
$\tilde{\sigma}(\bar{\nu} + N \rightarrow l^+ + c/\bar{c} + X)$	30
AdvSND	
$\tilde{\sigma}(\nu + N \rightarrow l^- + X)$	33
$\tilde{\sigma}(\bar{\nu} + N \rightarrow l^+ + X)$	29
$\tilde{\sigma}(\nu + N \rightarrow l^- + c/\bar{c} + X)$	17



- Neutrino DIS measurements at the LHC FPF operating concurrently with the HL-LHC would provide handles on PDFs of proton and nuclei in the medium-large-x region

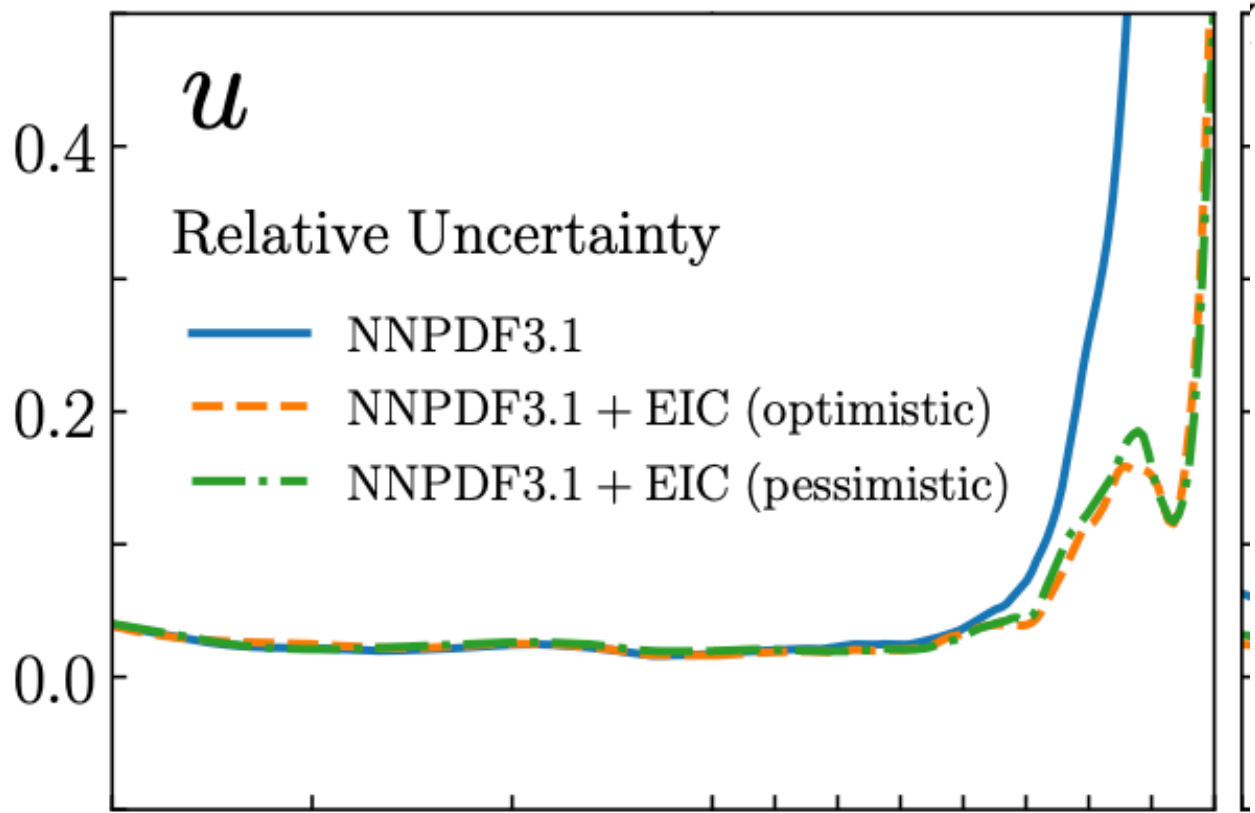
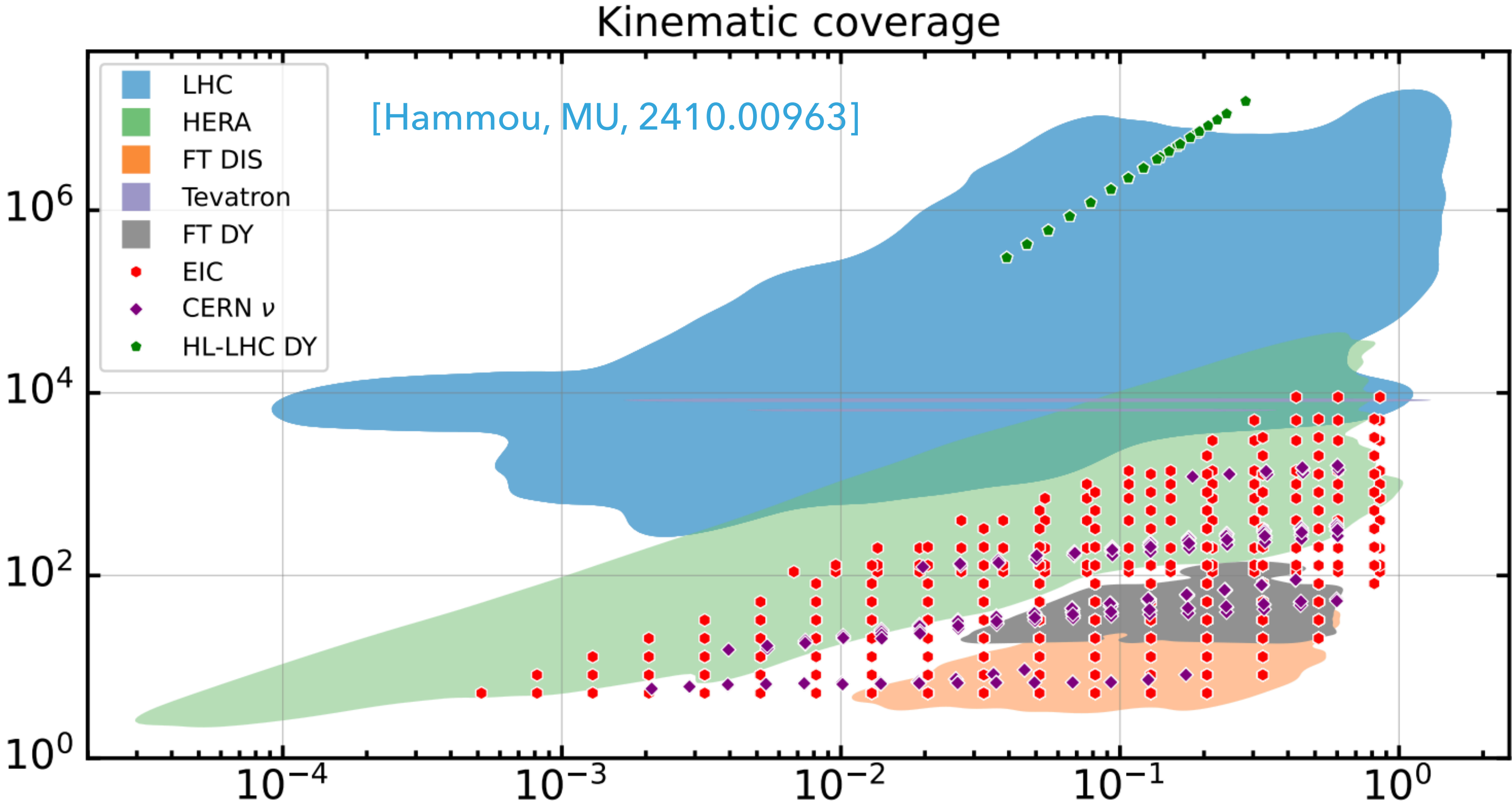
J Cruz-Martinez et al 2309.09581

PROJECTIONS FROM THE EIC

Observable	N_{dat}	$\sqrt{s}$ [GeV]	$\mathcal{L}$ [fb ⁻¹]
Charged Current $\tilde{\sigma}(e^- + p \rightarrow \nu + X)$ $\tilde{\sigma}(e^+ + p \rightarrow \bar{\nu} + X)$	89 (89)	140.7	100
	89 (89)	140.7	10
Neutral Current (proton) $\tilde{\sigma}(e^- + p \rightarrow e^- + X)$ $\tilde{\sigma}(e^+ + p \rightarrow e^+ + X)$	181 (131)	140.7	100
	181 (131)	63.2	100
	126 (91)	44.7	100
	87 (76)	28.6	100
	181 (131)	140.7	10
	181 (131)	63.2	10
	126 (91)	44.7	10
	87 (76)	28.6	10
Neutral Current (deuteron) $\tilde{\sigma}(e^- + d \rightarrow e^- + X)$ $\tilde{\sigma}(e^+ + d \rightarrow e^+ + X)$	116 (116)	89.0	10
	107 (107)	66.3	10
	76 (76)	28.6	10
	116 (116)	89.0	10
	107 (107)	66.3	10
	76 (76)	28.6	10

Optimistic
scenario

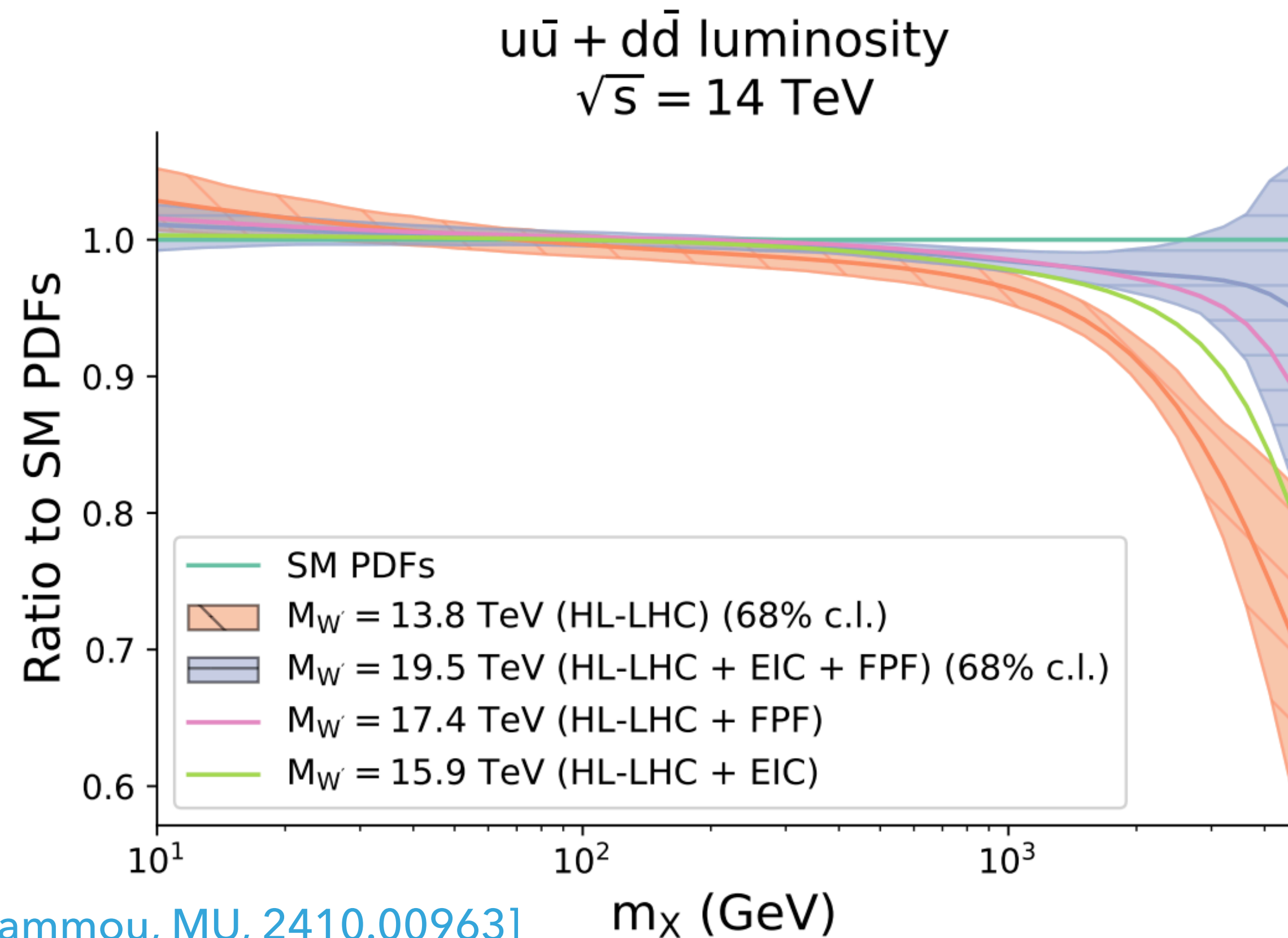
Pessimistic
scenario



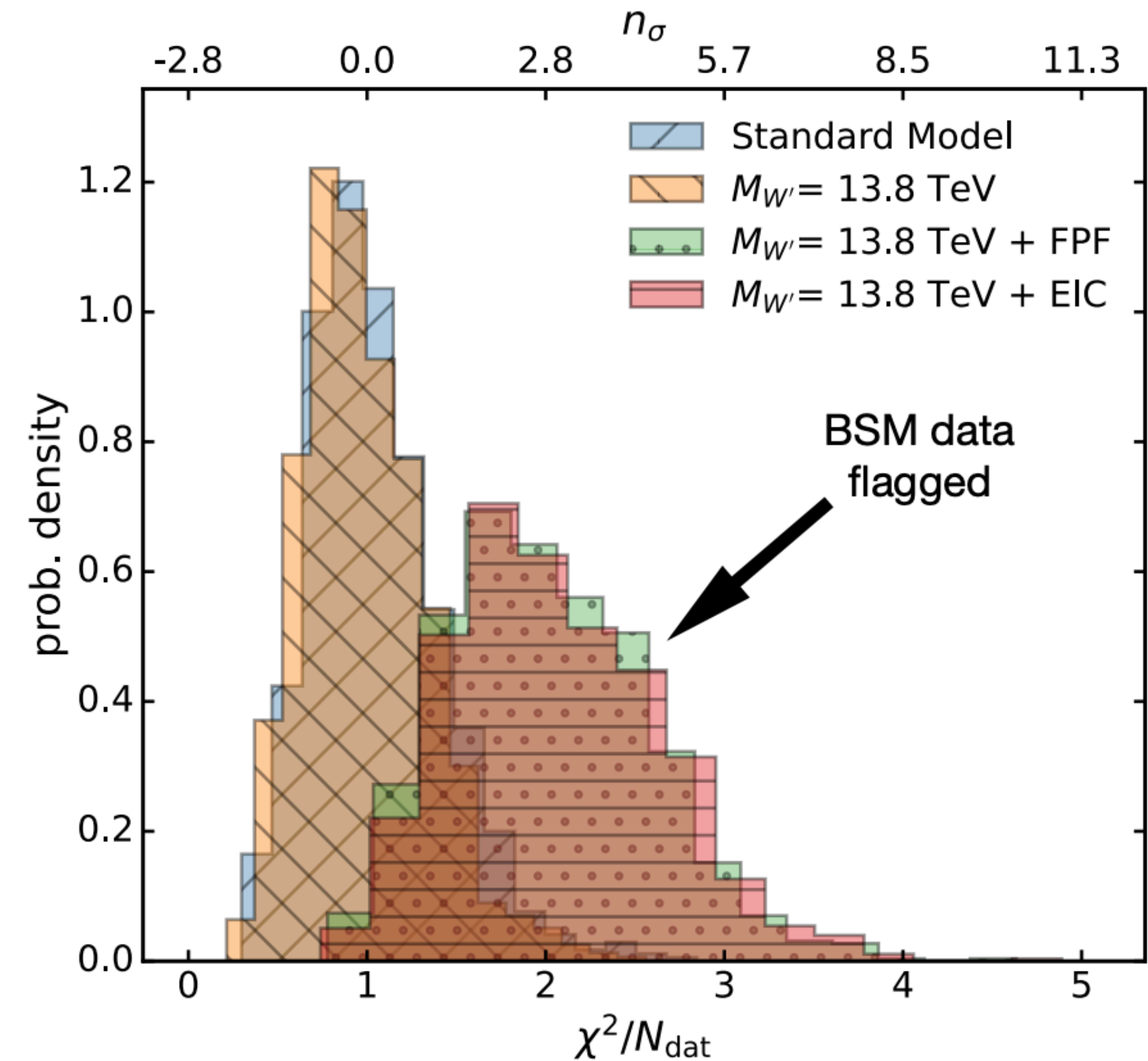
- NC and CC DIS measurements at the EIC would provide handles on up and down quarks at medium-large- x
- They would also constrain SMEFT 4-fermion operators

R. A. Khalek et al, 2102.00018
C. Bissolotti et al, 2307.09459

EIC AND FPF WOULD PREVENT BSM INDUCED BIAS



[Hammou, MU, 2410.00963]



- Including the EIC and FPF artificial data alongside the HL-LHC ones reduces the BSM-bias in PDF luminosity
- The HL-LHC data with $M_{W'} = 13.8$ TeV injected would be excluded from the fit as their theoretical description with the SM would be too poor and the data would be considered inconsistent
- EIC and FPF inclusion would push the threshold of BSM bias up to $M_{W'} \sim 20$ TeV and would make the effect on PDFs much smaller

EIC AND FPF WOULD PREVENT BSM INDUCED BIAS

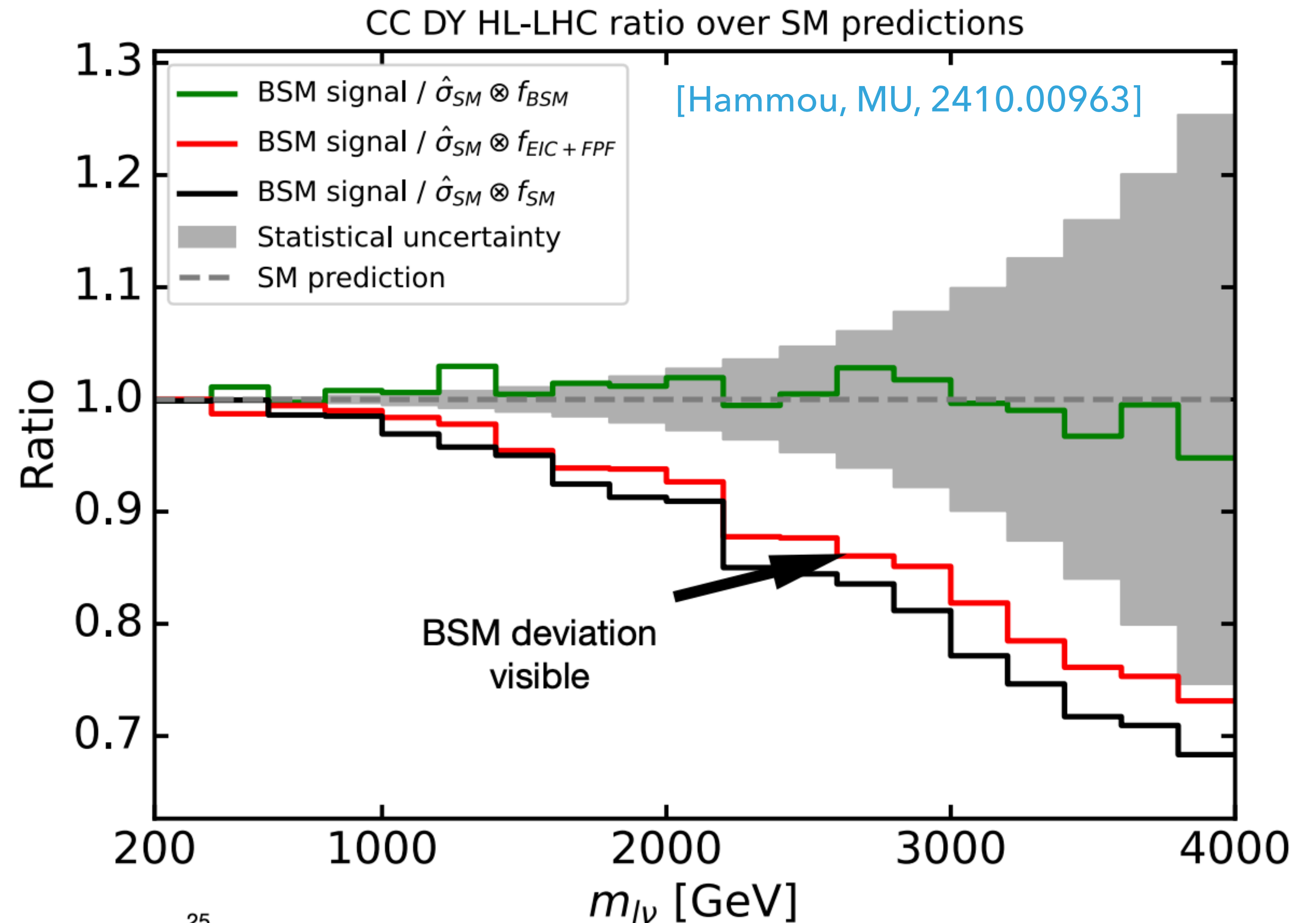
- ✓ Potential BSM effects, which might otherwise be absorbed into the PDFs, can be disentangled in high-energy measurements by incorporating EIC and FPF measurements in a global PDF analysis
- ✓ This minimises the risk of BSM-induced bias in PDF fits and allow for more consistent identification of BSM effects in high energy data, which can then be analysed separately.

$$\hat{\sigma}_{BSM} \otimes \mathcal{L}_{SM} \approx \hat{\sigma}_{SM} \otimes \mathcal{L}_{BSM}$$



$$\hat{\sigma}_{BSM} \otimes \mathcal{L}_{SM} \neq \hat{\sigma}_{SM} \otimes \mathcal{L}_{EIC+FPF}$$

$M_{W'} : 13.8 \text{ TeV}$



CONCLUSIONS AND OUTLOOK

- PDFs are fundamental to huge range of key SM/BSM processes + often limit sensitivity.
- Progress possible on several fronts:
 - **Data** - Substantial potential from complementarity of upcoming/proposed colliders.
Challenge - Understanding **systematics, correlations**, inconsistencies between experiments.
 - **Methodology** - Many developments on PDF methodology, parameterisation.
Challenge - Complexity and precision of global fits. Understanding origin/need for tolerance.
 - **Theory** - Moving towards **N3LO PDFs**, this + inclusion of **theoretical uncertainties** are crucial as precision demands increase.
Challenge - Understanding of theoretical uncertainties, higher order calculations.
 - **Standard Model/BSM Parameters** - PDF precision/accuracy improvements will drive advances in former.
Challenge - Understanding **assumptions of profiling** + impacts, need for **simultaneous fits**.
Separating out BSM from PDFs - aided by complementary data from multiple colliders.
 - **Lattice "data"** - Perhaps not competitive at the moment in polarised fits of PDFs, but already good enough to constrain the large-x region and pin down any possible BSM energy growing effect in the tails of the distributions that are included in PDF fits?

EIC & LHC 😊

EIC & LHC 😊