QCD AND COLLIDER PHENO GROUP AT DAMTP, CAMBRIDGE

SMEFT, PDFS AND MORE

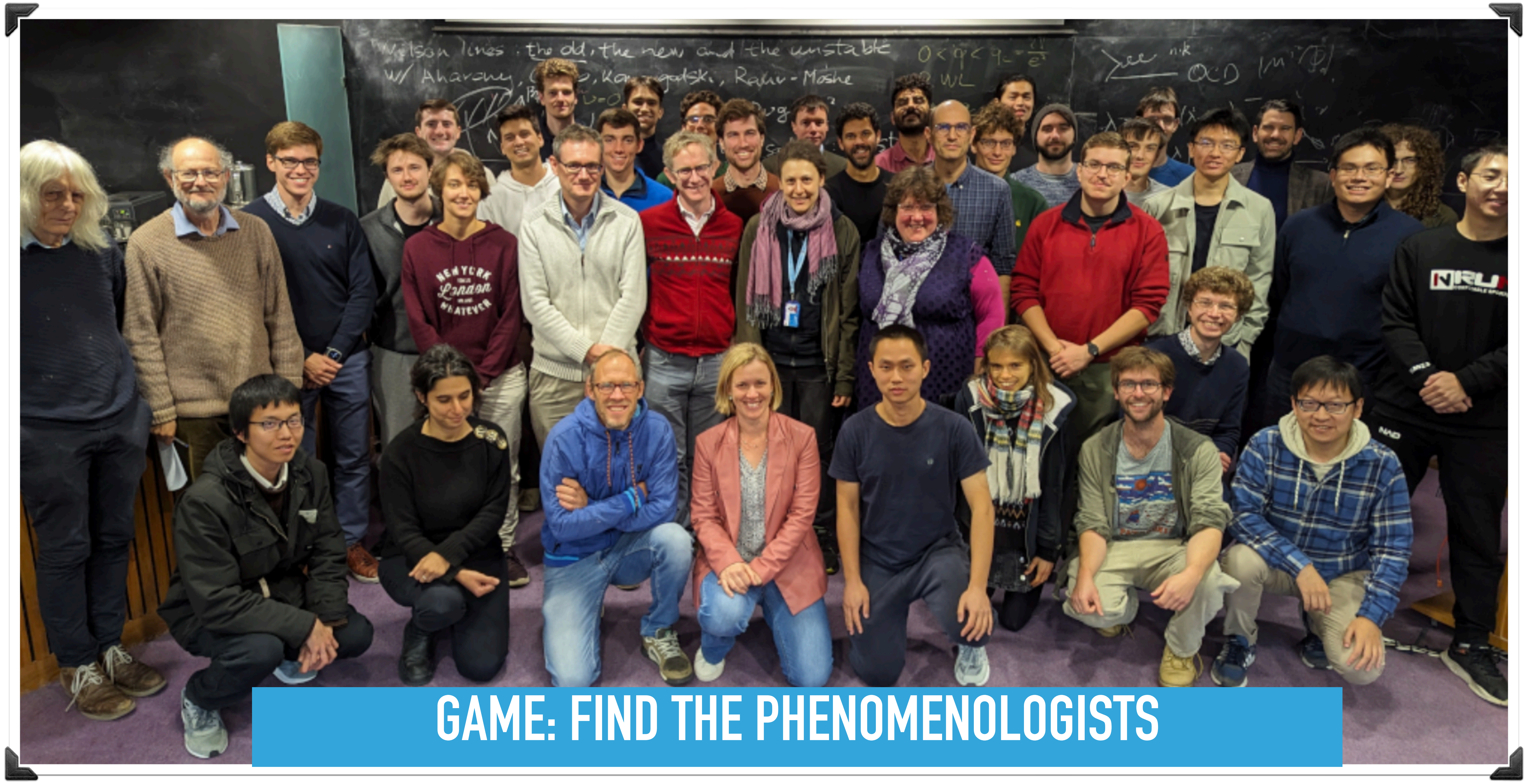
1st South-East UK QCD and collider phenomenology meeting

Oxford - 17th November 2023

PHENO AT DAMTP

- Ben Allanach + Hannah Banks + Nico Gubernari (BSM pheno, model building)

- Maria Ubiali + Luca Mantani + James Moore + Manuel Morales + Elie Hammou + Mark Costantini (QCD and SM/ BSM pheno)

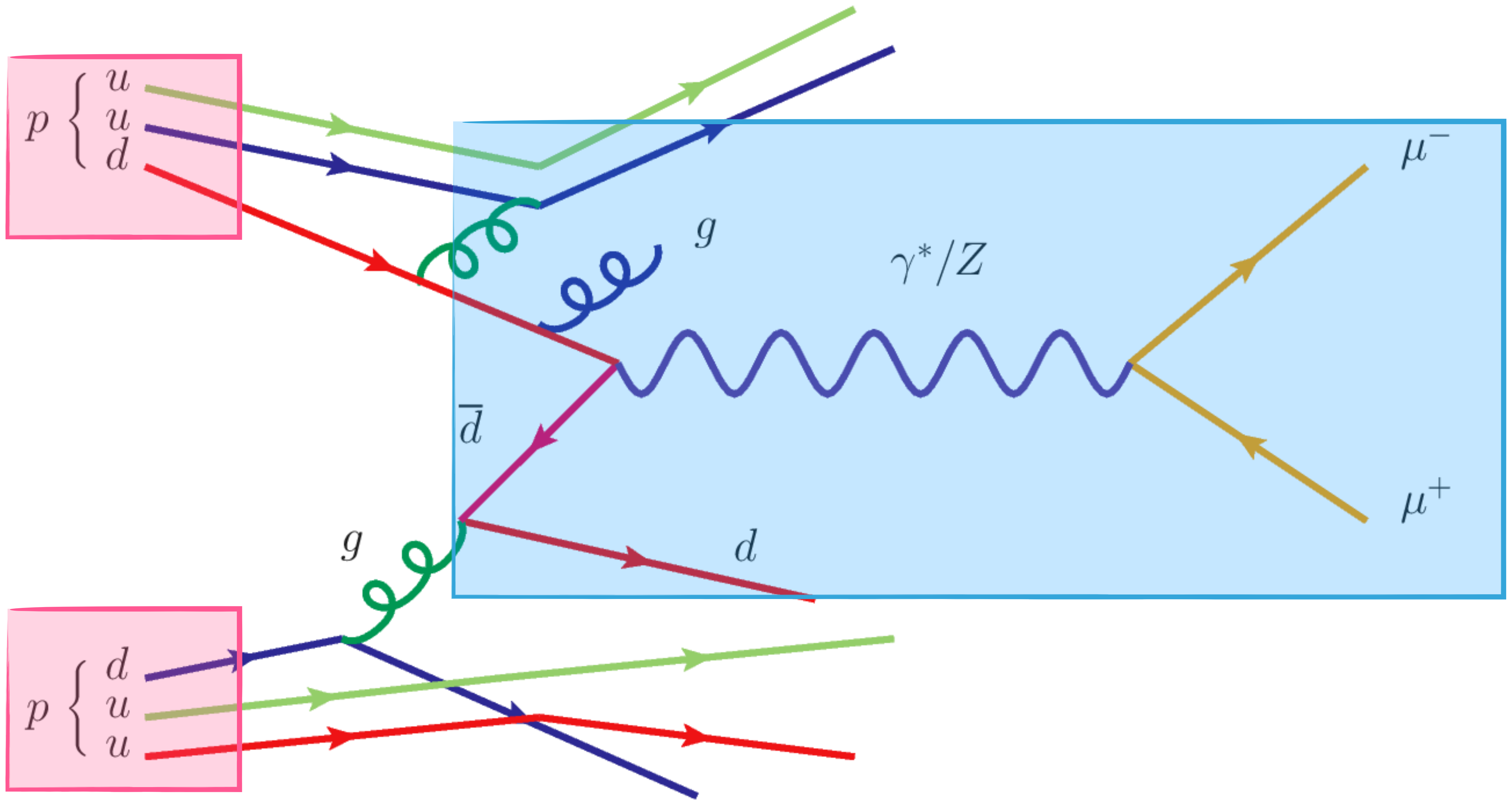


- My research interests: Parton Distribution Functions, Interplay between PDF fits and BSM signals, heavy quark fragmentations (in collaboration with Fabio Maltoni, Giovanni Ridolfi and Marco Zaro), collider signatures of weekly interacting particles (in collaboration with Fabian Esser, Maeve Madigan, Matthew McCullough, James Moore, Veronica Sanz), symbolic regression in HEP pheno (in collaboration with Daniel Conde, Manuel Morales, Veronica Sanz)

$$\sigma^{pp\to ab} = \sum_{i,j=-n_f}^{n_f} \int dz_1\, dz_2 \boxed{f_i(z_1,\mu_F)\, f_j(z_2,\mu_F)} \boxed{\hat\sigma^{ij\to ab}(z_1 z_2 S, \alpha_s(\mu_R), \mu_F)} + \mathcal{O}\left(\frac{\Lambda^n}{S^n}\right)$$

Collinear factorisation: separate long-distance universal information on proton structure in terms of quarks, antiquarks and gluons (partons) from from short-distance parton interaction (hard scattering)

Parton Distribution Functions



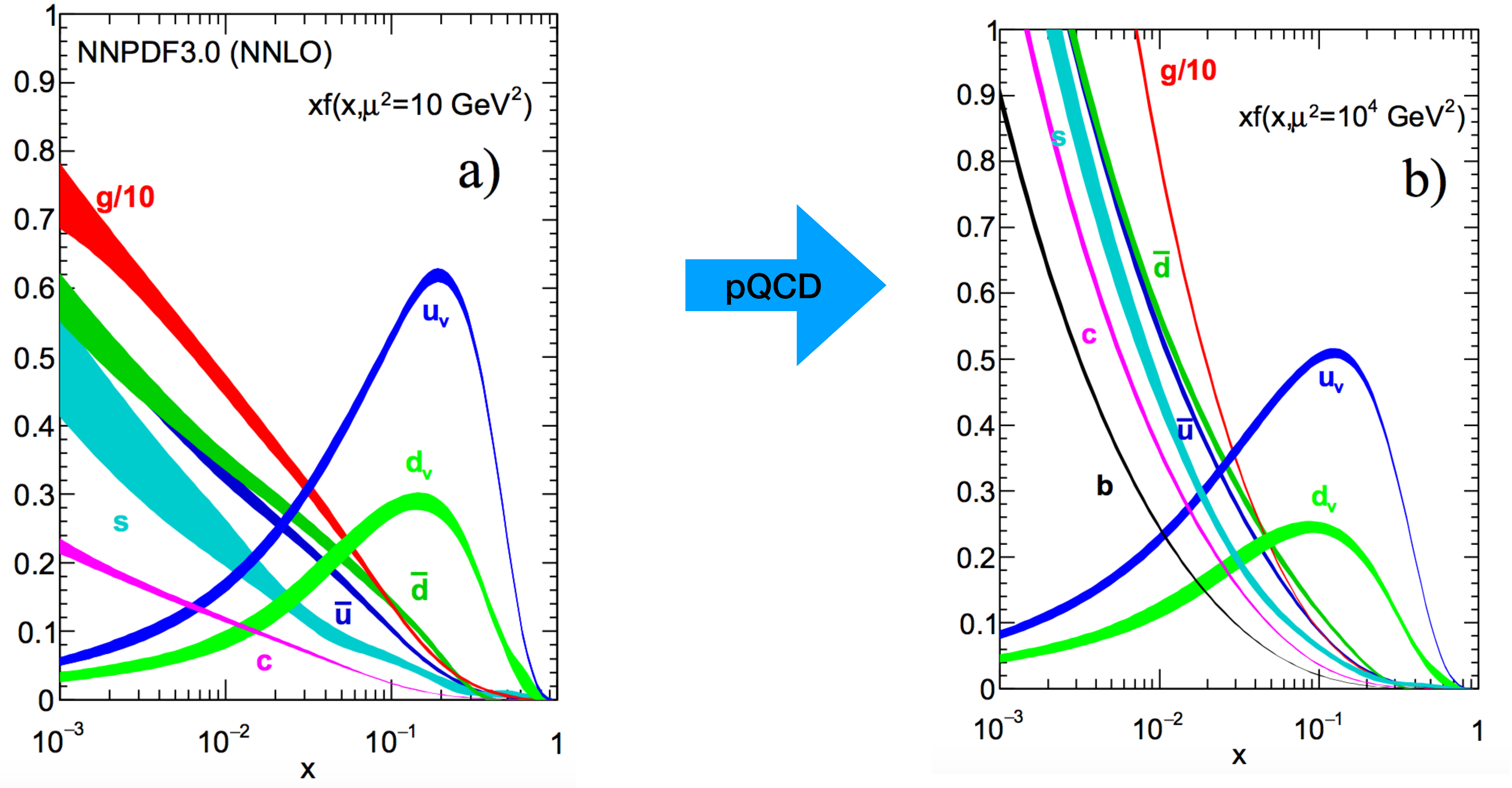
Hard Scattering: Perturbative QCD + EW

$$f_i(x, \mu)$$

Data

Perturbative QCD

Hadronic scale: global fit of PDFs



High scale: input to the LHC

PDG 2016

NNPDF: TOWARDS PRECISE AND ACCURATE PDFS

- Improving precision and accuracy of PDF determination (with the NNPDF collaboration)
  - **NNPDF40QED**: new QCD+QED determination of PDFs based on LuxQED
  - **NNPDF40MHOU**: first NNLO set of PDFs including missing higher order uncertainties based on the use of theory covariance matrix
  - **NNPDF40N3LO**: approximate N3LO PDF fit including incomplete higher order and missing higher order uncertainties
  - **NNPDF40pheno**: A comprehensive phenomenological study of the data-theory agreement with new experimental data (not included in NNPDF4.0)
  - Methodology studies based on **closure tests with inconsistent data**

EXTRACTING PARAMETERS FROM THE LHC DATA

$$\chi^2 = \frac{1}{N_{\rm dat}} \sum_{i=1}^{N_{\rm dat}} (T_i(\{\theta\},\{c\}) - D_i)\, {\rm cov}^{-1}_{ij}\, (T_j(\{\theta\},\{c\}) - D_j)$$

$$T_i(\{\theta\},\{c\}) = {\rm PDFs}(\{\theta\},\{c\}) \otimes \hat{\sigma}_i(\{c\})$$

$$\mathcal{L}_{\rm SMEFT} = \mathcal{L}_{\rm SM} + \sum_i^{N_{d6}} \frac{c_i}{\Lambda^2}\mathcal{O}_i^{(6)} + \sum_j^{N_{d8}} \frac{b_j}{\Lambda^4}\mathcal{O}_j^{(8)} + \dots$$

(B)SM parameters: $\alpha_s(M_z)$, M_w, θ_w ,**SMEFT WCs**……

Parameters determining PDFs at initial scale

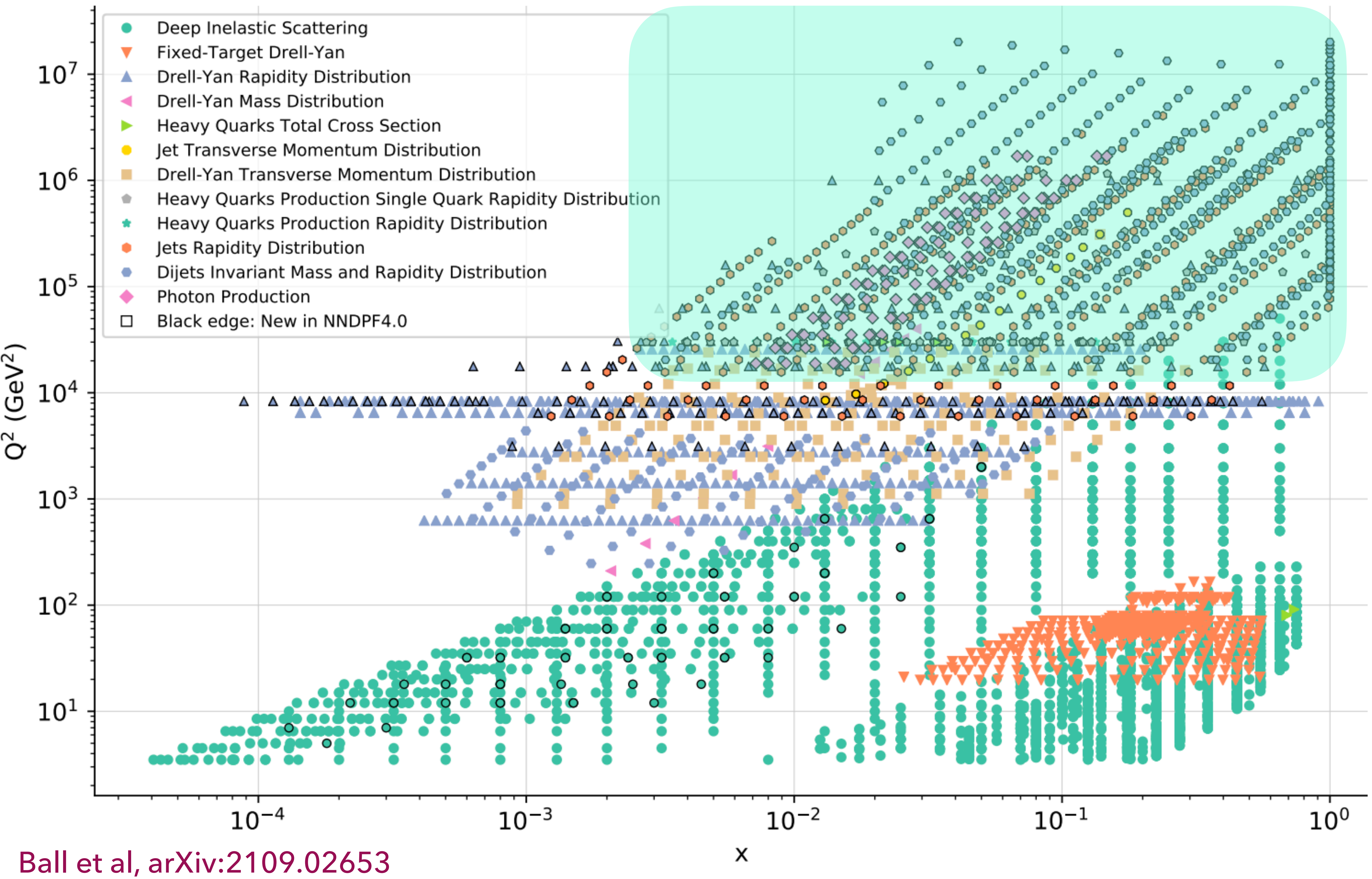
✓ In a PDF fit typically

$$T_i(\{\theta\}) = {\rm PDFs}(\{\theta\},\{c=0\}) \otimes \hat{\sigma}_i(\{c=0\})$$

✓ In a fit of SMEFT Wilson Coefficients

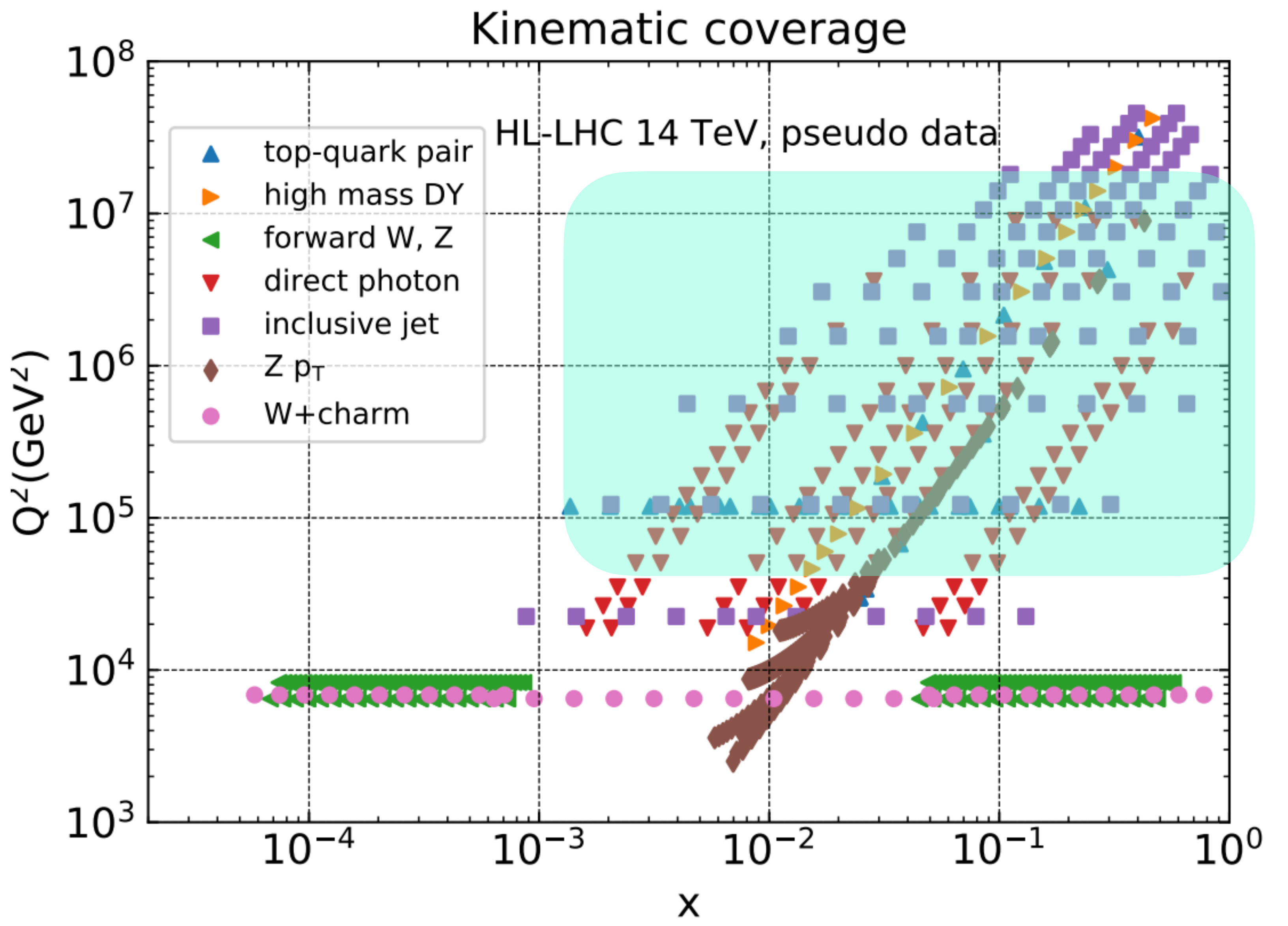
$$T_i(\{c\}) = {\rm PDFs}(\{\theta=\bar{\theta}\},\{c=0\}) \otimes \hat{\sigma}_i(\{c\})$$

EXTRACTING PARAMETERS FROM THE LHC DATA



Ball et al, arXiv:2109.02653

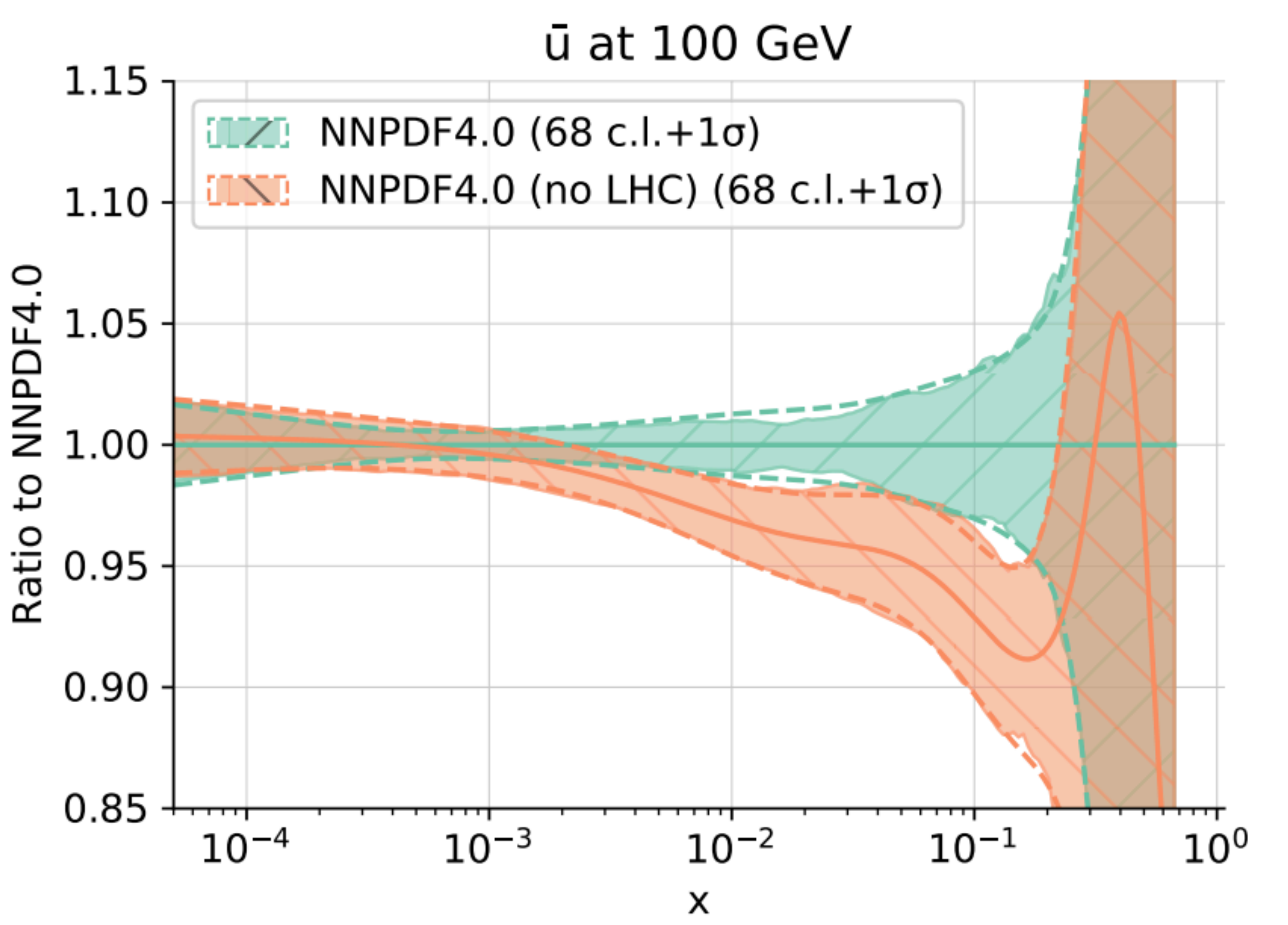
- ➡ Top pair production and single top data included in SMEFT analysis [Hartland et al 1901.05965] [Ellis et al 2012.02779]
- ➡ Dijets data in [Bordone et al 2103.10332] [Alioli et al 1706.03068]
- ➡ Drell-Yan data in [Farina et al 1609.08157, Torre et al 2008.12978]
- ➡ Jets and dijets [Alte et al 1711.07484]
- ➡ Overlap enhanced in HL-LHC projections [Abdul Khalek et al1810.03639]

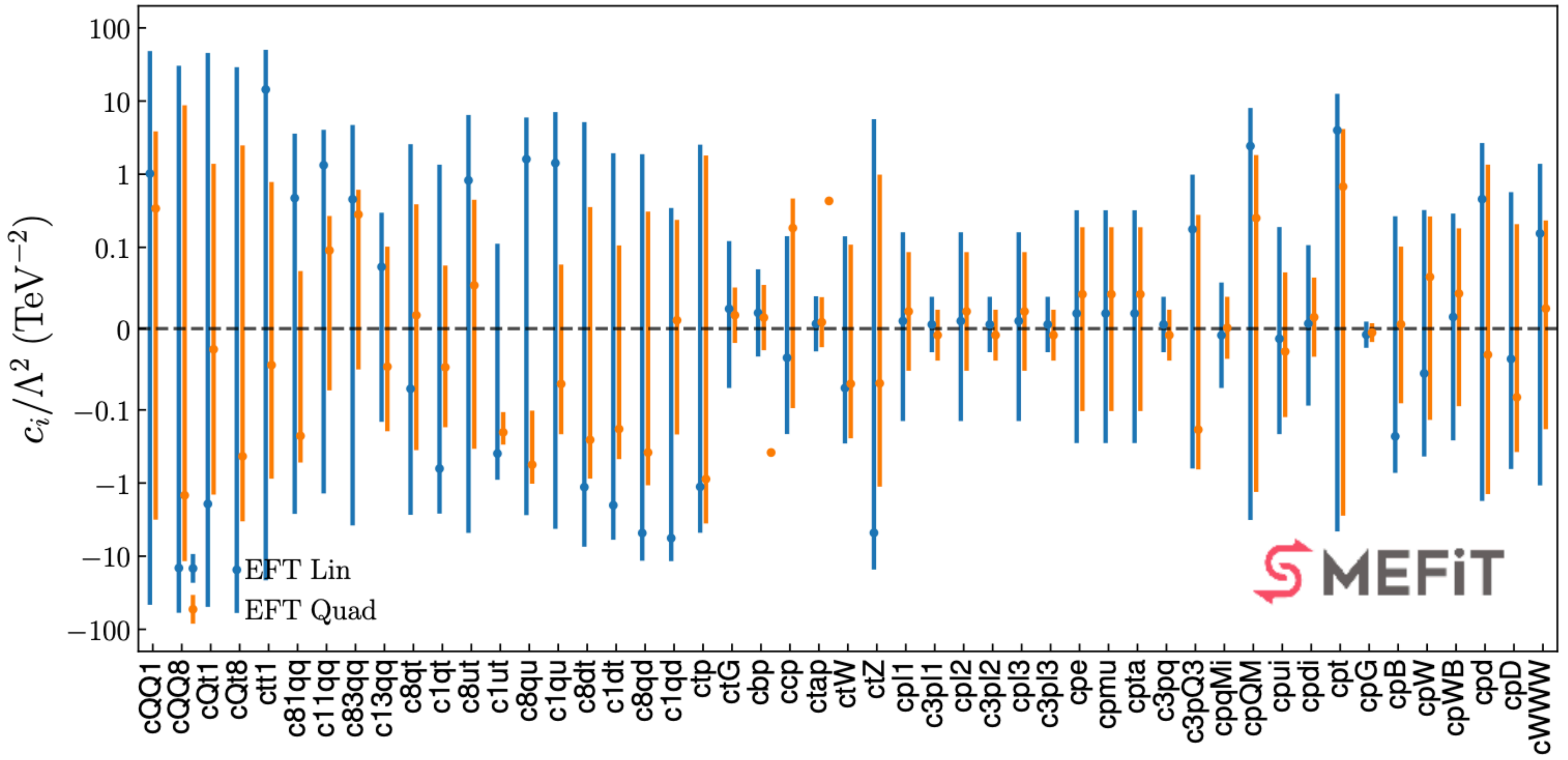


Abdul Khalek et al,1810.03639

PDF AND NEW PHYSICS INTERPLAY

- PDFs are low-scale quantities extracted from experimental data at all scales, without considering any potential high-scale contamination due to new physics.
- (SM)EFT fits are performed by assuming a priori that PDFs are SM-like.
- In principle low-scale physics is separable from high-scale physics, BUT the complexity of LHC environment might well intertwine them.





Ball et al, arXiv:2109.02653

Ethier et al, arXiv: 2105.00006

QUESTIONS

- 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?
- From the point of view of SMEFT fits:
  - ➡ Should I make sure I am using a clean set of PDFs in a SMEFT analysis? How to define it? Is it enough?
  - ➡ How would the bounds change if I was consistently using PDFs that include in the fit the same operators that I am fitting?

T

$$\boxed{d\sigma^{pp\to ab}} = \sum_{i,j} \boxed{f_i \otimes f_j \otimes d\hat{\sigma}^{ij\to ab}} + \ldots$$

Simultaneous fits can shed light on their interplay

$$T(\{\theta_k\},\{c_i\})$$

$$f(\{\theta_k\}) \qquad \mathcal{L}_{\rm SMEFT} = \mathcal{L}_{\rm SM} + \sum_i \frac{c_i}{\Lambda^2}\mathcal{O}_i^{(6)} + \cdots$$

THE PBSP GROUP

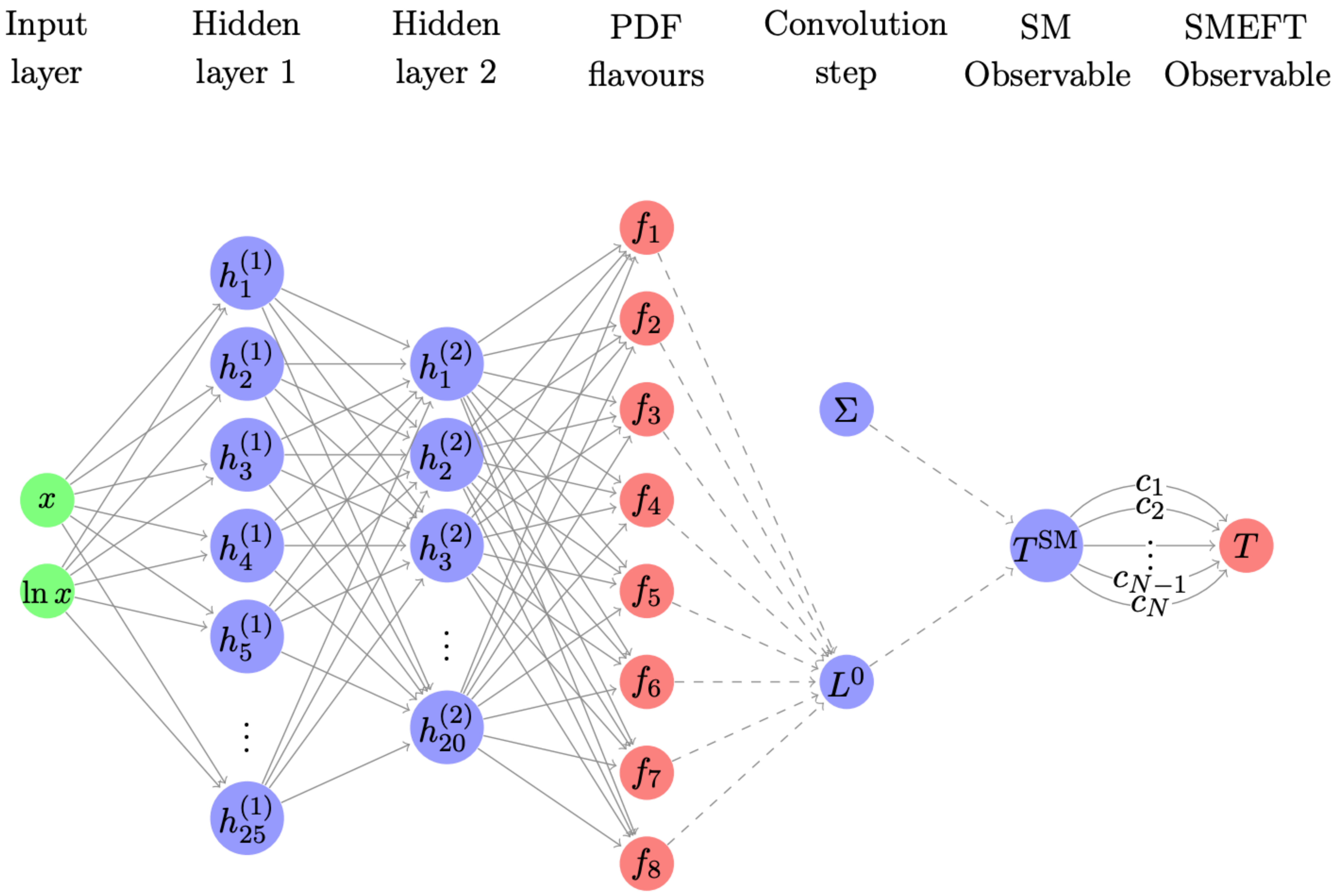
- PI: Maria Ubiali
- Postdocs: Luca Mantani, James Moore Zahari Kassabov (former), Maeve Madigan (former, now in Heidelberg)
- PhD students: Manuel Morales (III year), Elie Hammou (II year), Mark Costantini (II year), James Moore (now postdoc), Shayan Iranipour (former), Cameron Voisey (former)
- Visiting PhD students: Daniel Conde (PhD in Valencia) and Fabian Esser (visiting end of 2022 from Valencia)

PBSP

erc

SIMUNET: THE GENERAL IDEA [2201.07240]

- SimuNET yields a truly simultaneous fit, rather than a scan in benchmark point in WC space and it does not have limit in number of parameters that can be fitted alongside PDFs at the initial scale!



Linear dim-6 operator

$$\boxed{T(\hat{\theta}) = \Sigma(\{c_n\}) \cdot L^0(\theta) = T^{\rm SM}(\theta) \cdot \left(1 + \sum_{n=1}^{N} c_n R^{(n)}_{\rm SMEFT}\right)}$$

$$T^{\rm SM}(\theta) = \Sigma^{\rm SM} \cdot L^0(\theta)$$

Quadratic dim-6 operator

$$\boxed{T(\hat{\theta}) = T^{\rm SM}(\theta) \cdot \left(1 + \sum_{n=1}^{N} c_n R^{(n)}_{\rm SMEFT} + \sum_{1\leq n\leq m\leq N} c_{nm} R^{(n,m)}_{\rm SMEFT}\,,\right)}$$

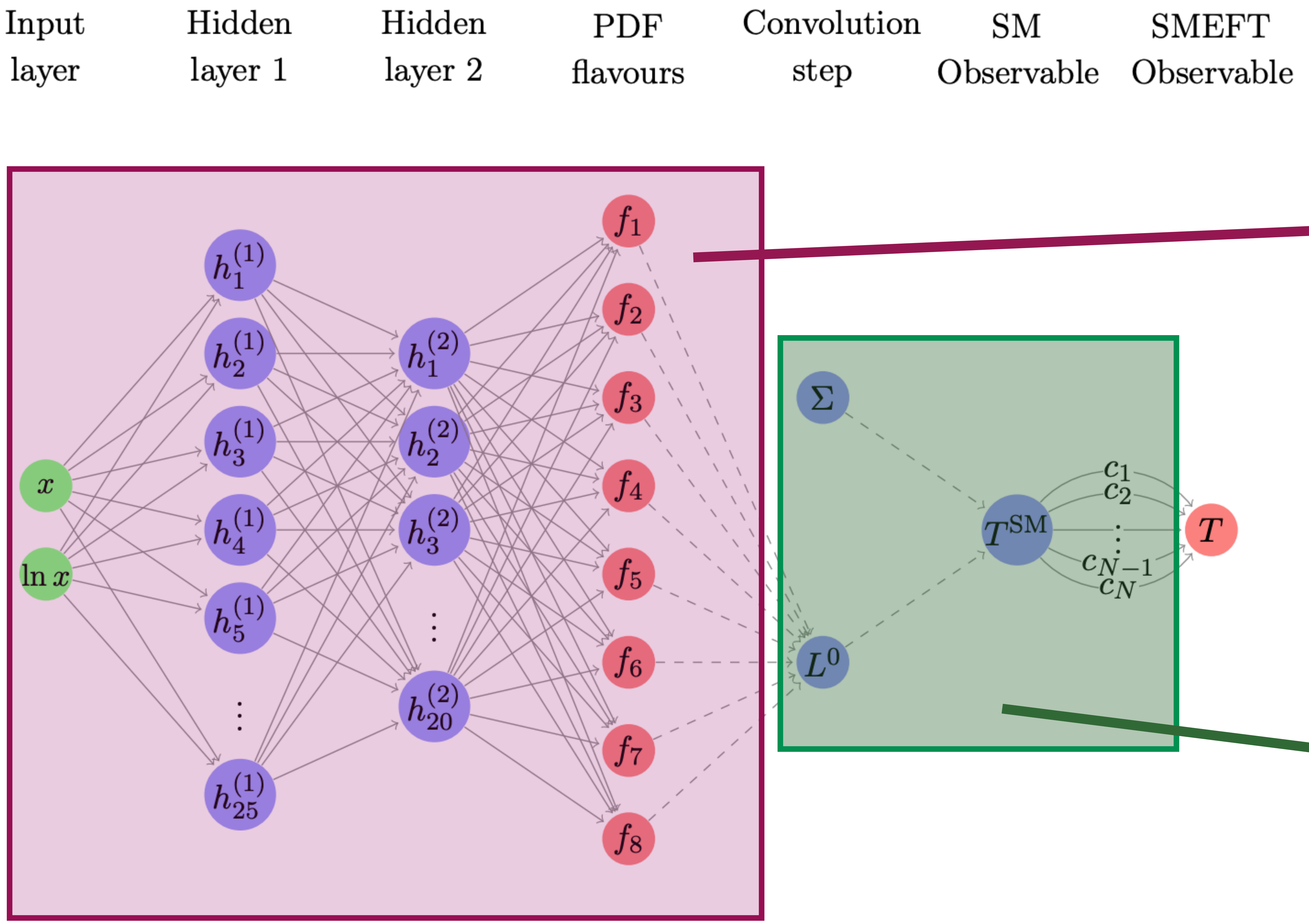
$$c_n c_m$$

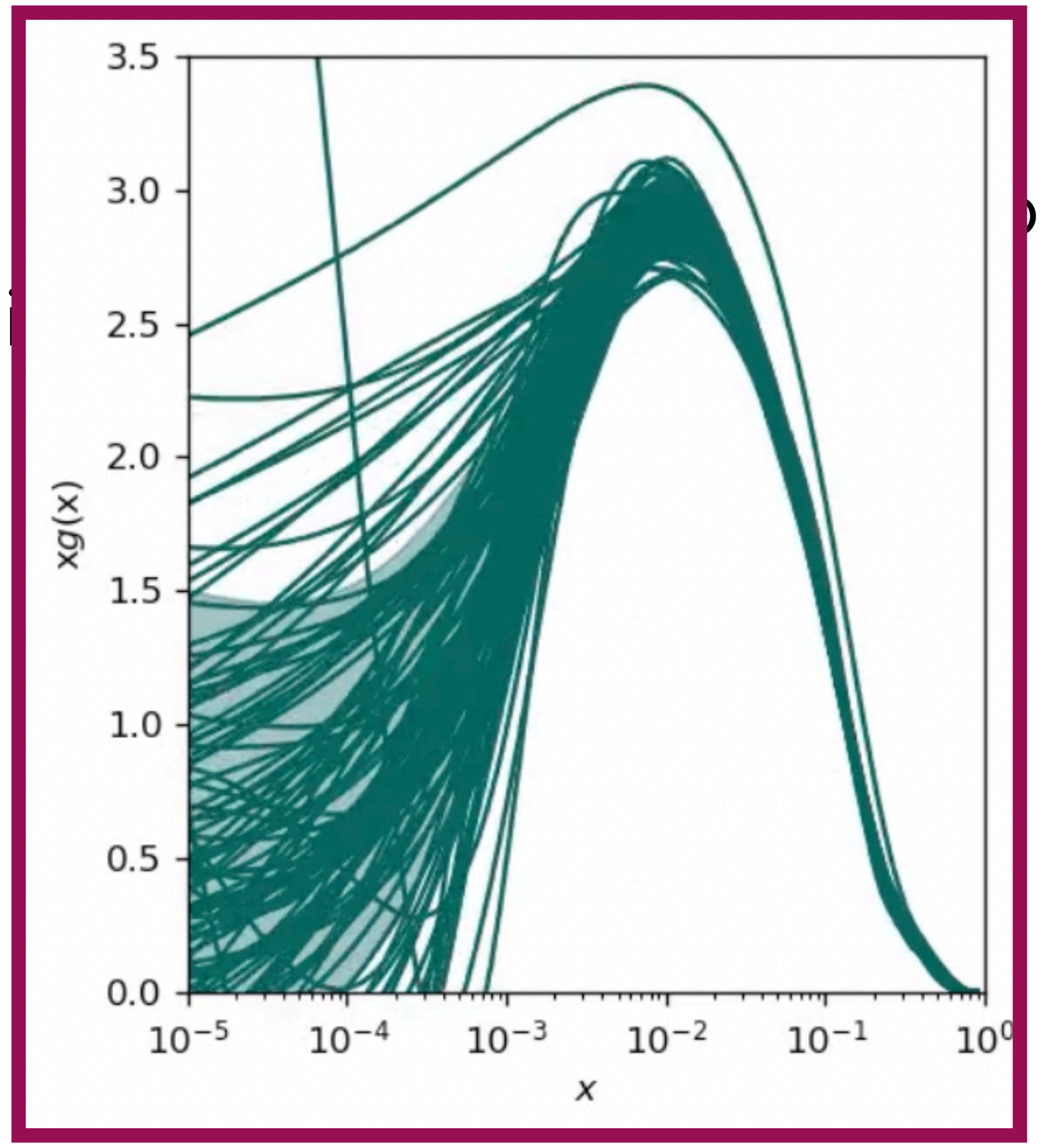
QUADRATIC POSSIBLE IN PRINCIPLE BUT NOT RELIABLE DUE TO MC UNCERTAINTY PROPAGATION

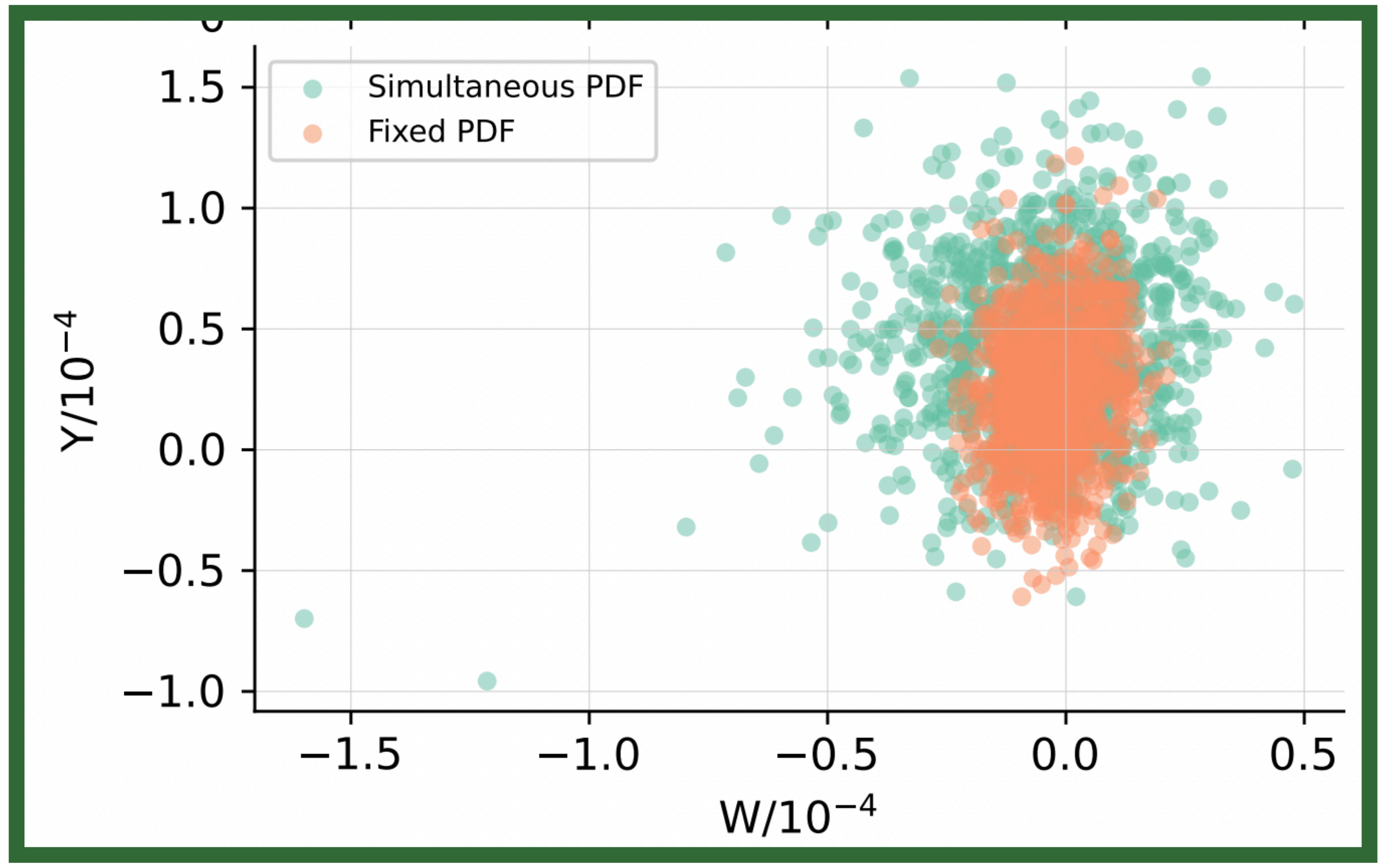
S. Iranipour, MU - arXiv: 2201.07240

SIMUNET: THE GENERAL IDEA

- SimuNET yields a truly simultaneous fit, rather than a scan in benchmark ... oes not have limit in number of parameters that can be fitted alongside PDFs at the ini



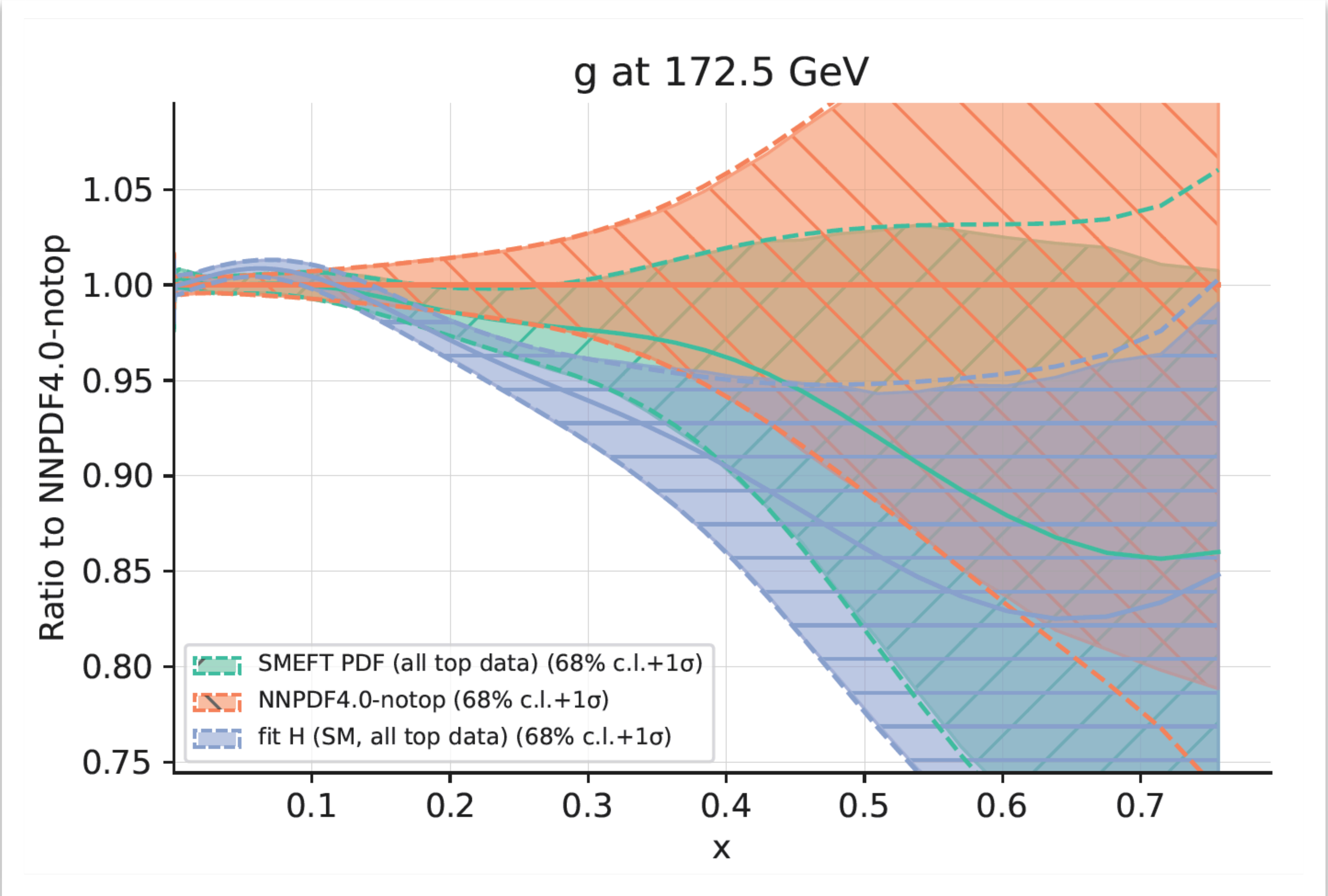


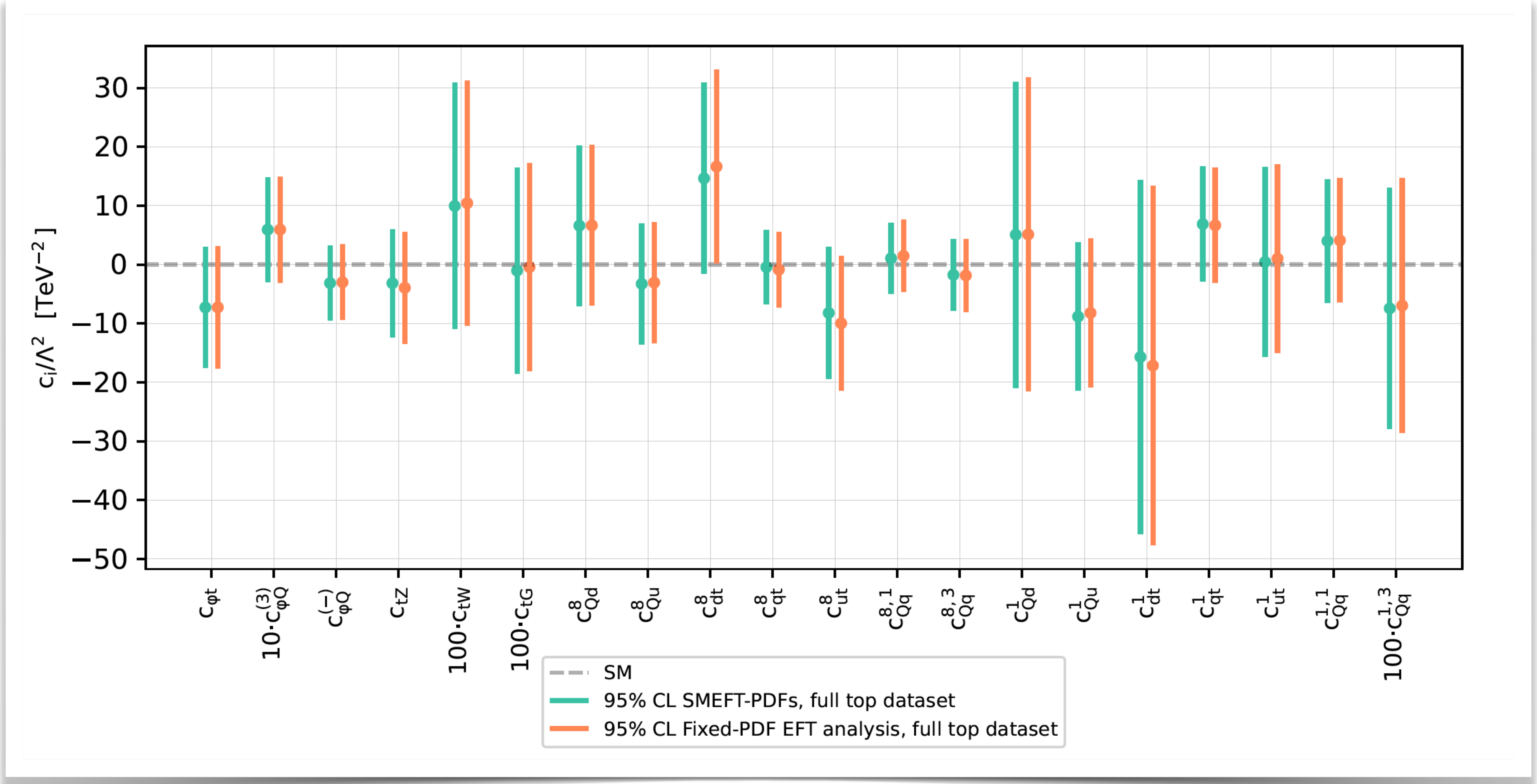


S. Iranipour, MU - arXiv: 2201.07240

THE TOP QUARK LEGACY OF THE LHC RUN II FOR PDF AND SMEFT ANALYSES [2303.06159]

We considered the most comprehensive and up-to-date top sector dataset



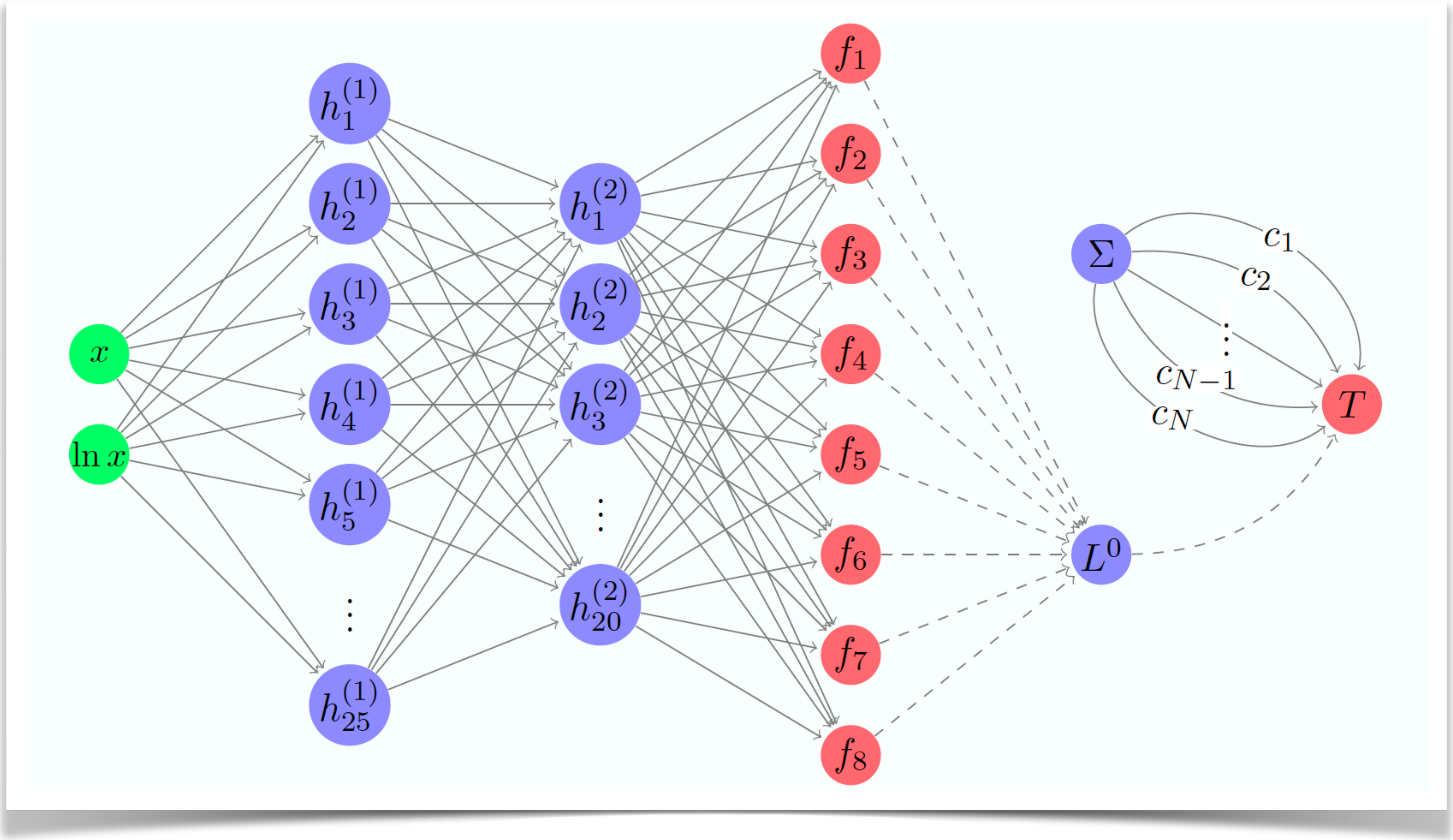


We performed PDF/SMEFT only fits, and simultaneous PDF-SMEFT fits to assess the interplay

Z. Kassabov, M. Madigan, L. Mantani, J. Moore, M. Morales, J. Rojo, M. Ubiali - arXiv: 2303.06159

SIMUNET PUBLIC RELEASE (COMING UP SOON)

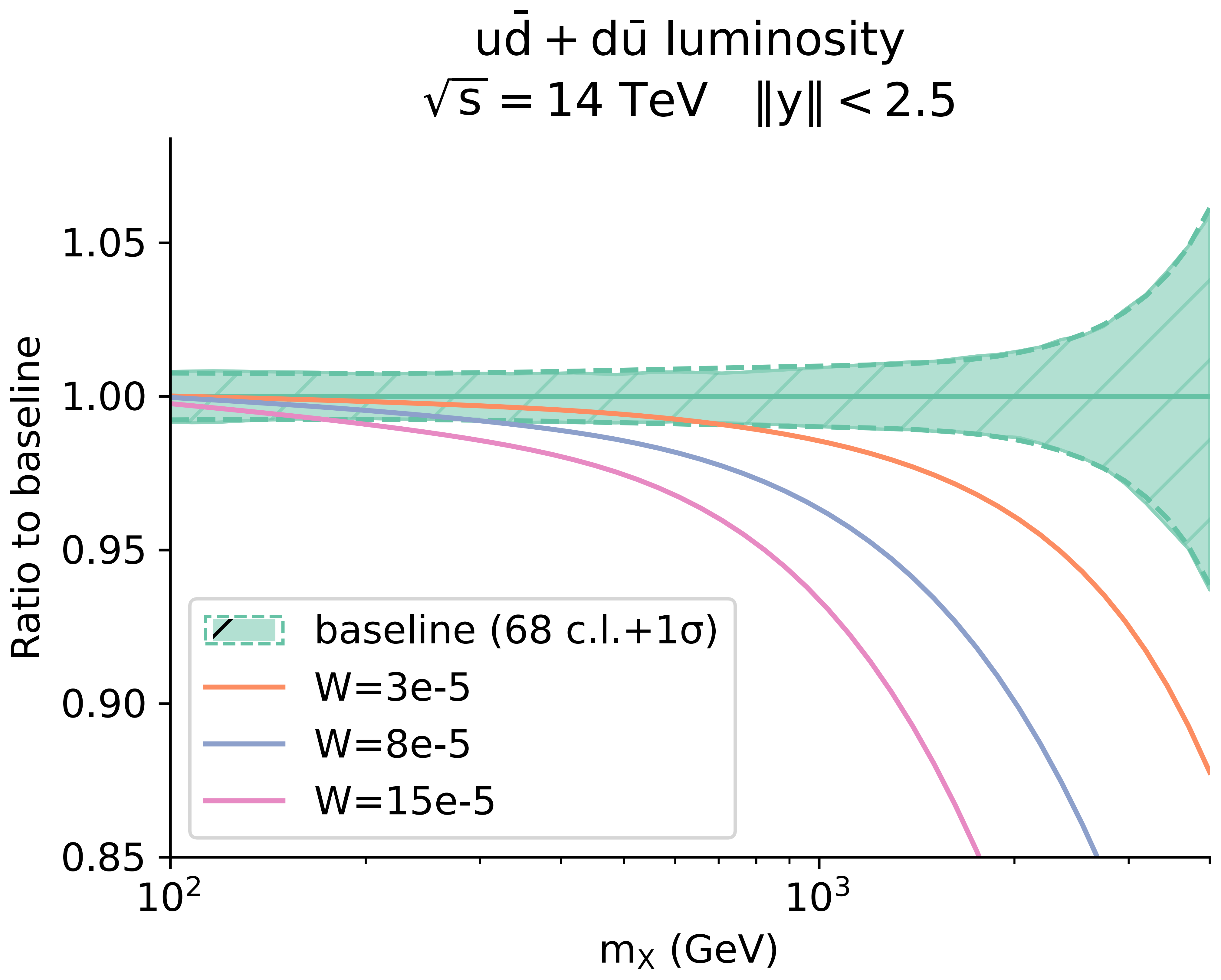
1. **Fully documented and open-source** methodology to perform simultaneous PDF-EFT fits
2. **New datasets** available (adding Higgs, EW, diboson, DY)
3. Extra features to **test new physics absorption** by the PDFs (next slides)



M. Costantini, E. Hammou, M. Madigan, L. Mantani, J. Moore, M. Morales, M. Ubiali - in preparation

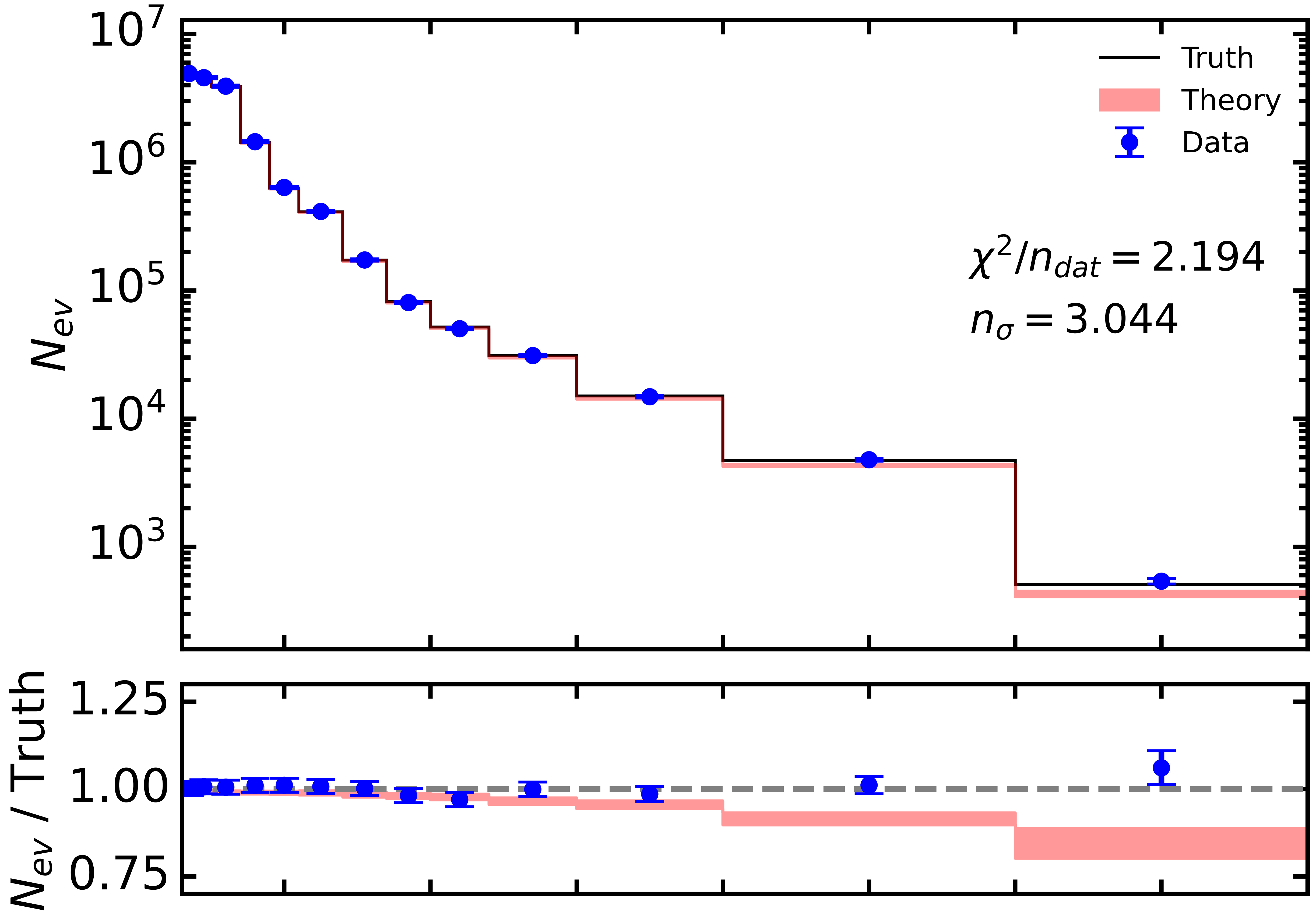
PDF "CONTAMINATION" FROM NEW PHYSICS [2307.10370]

PDF fitting in presence of W'



$$\sigma_{BSM} \otimes f_{true} \approx \sigma_{SM} \otimes f_{cont}$$

Fake deviations in other sectors



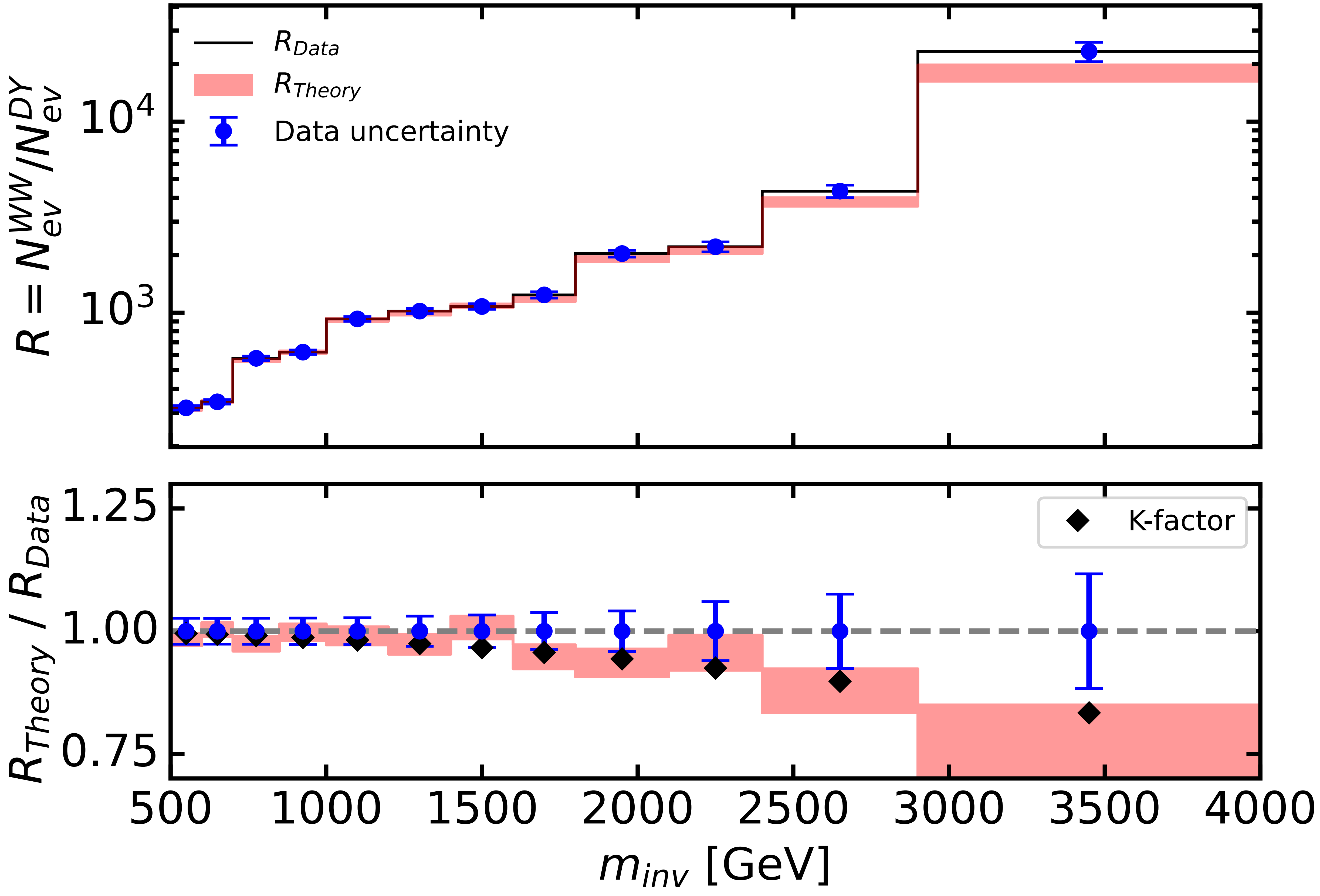
$$pp \to W^+W^-$$

E. Hammou, Z. Kassabov, M. Madigan, M. Mangano, L. Mantani, J. Moore, M. Morales, M. Ubiali - arXiv: 2307.10370

HOW TO DISENTANGLE SUCH EFFECTS?

Ratio of observables

WW / NC DY



Low-energy large-x data

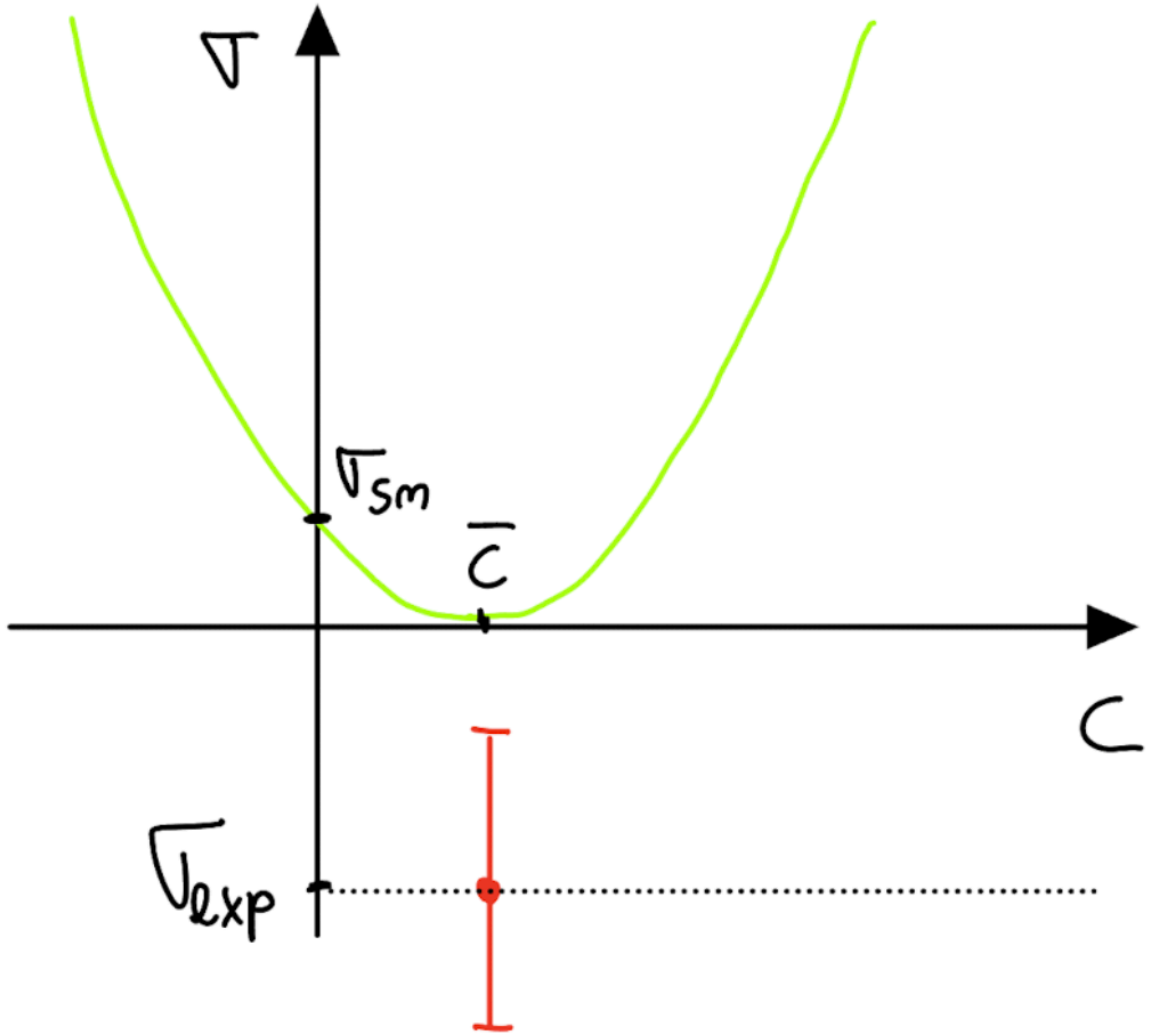
Data-Theory comparison

		Baseline	Contaminated
	Data points (ndata)	χ^2/ndata	χ^2/ndata
NuSea (2001)	15	1.350	1.823
NuSea (2003)	89	0.8017	0.9769
SeaQuest	6	0.4192	1.034
D0 detector	9	2.385	3.046
Total	119	0.9699	1.239

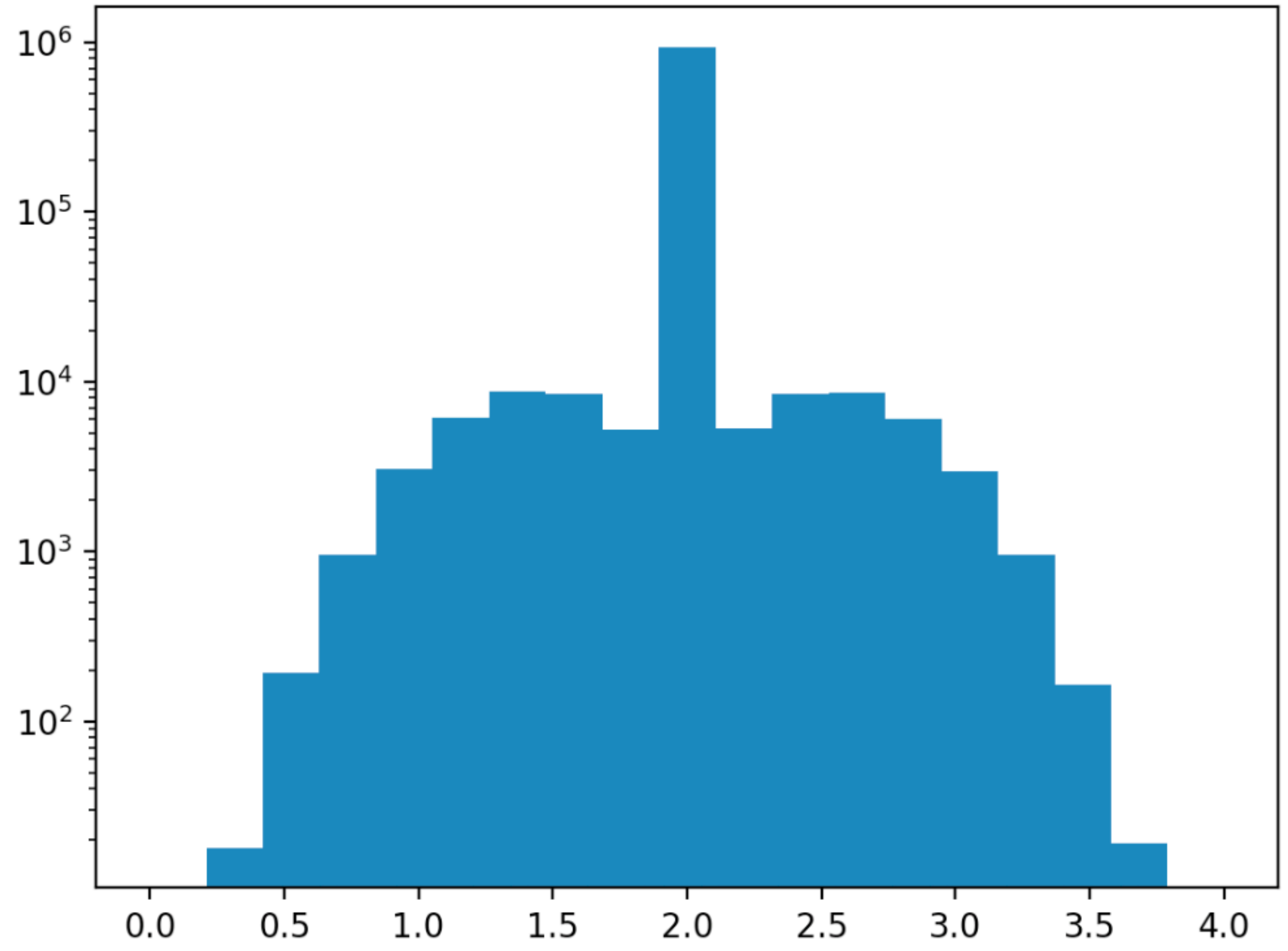
QUADRATIC FITS: A CHALLENGE

Let's consider a simple scenario: 1 operator, 1 datapoint

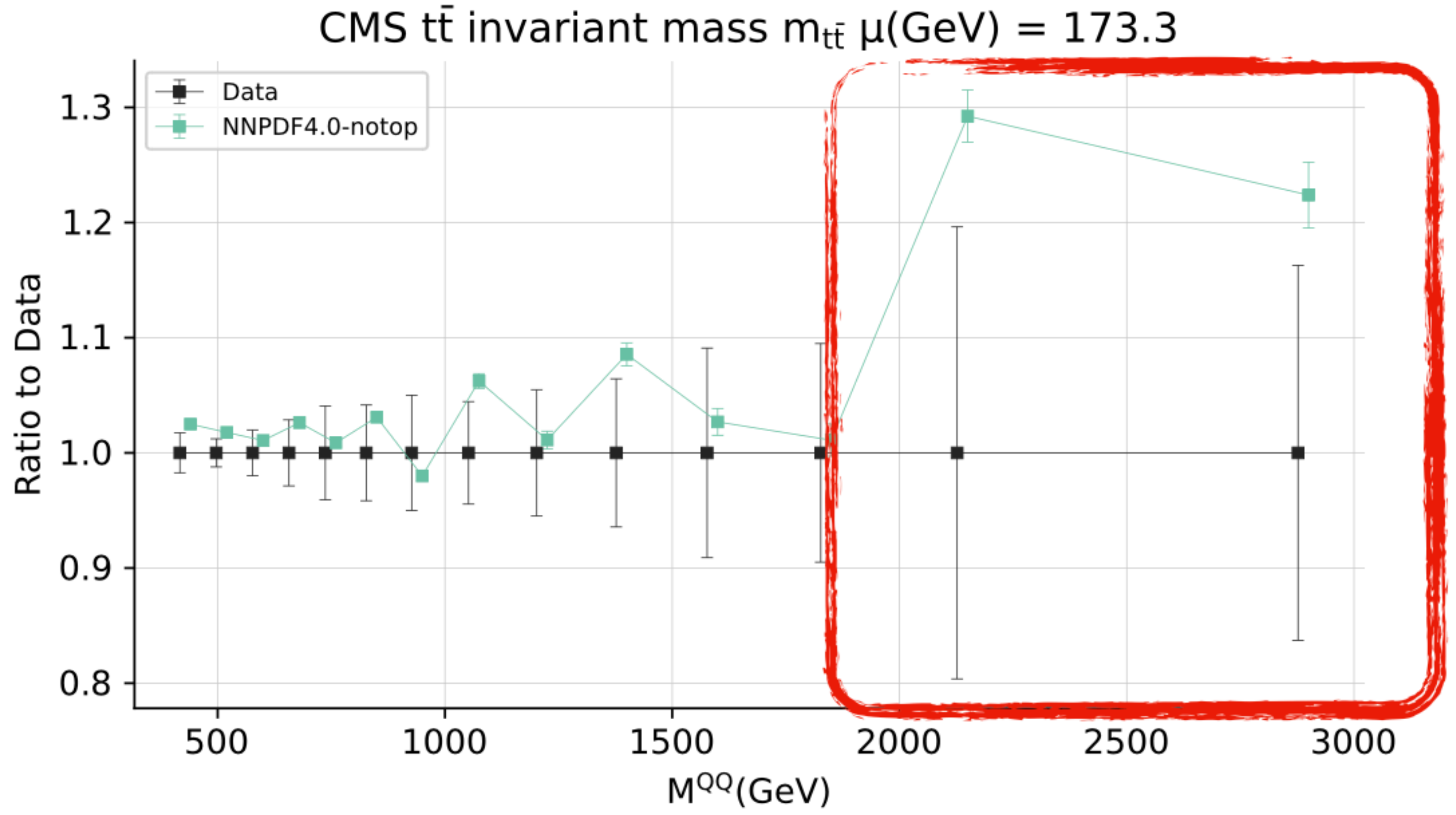
$$\chi^2 = \frac{(\sigma(c) - \sigma_{exp})^2}{\delta\sigma^2} \qquad \Delta\chi^2 = \chi^2 - \chi_{min} = 1 \longrightarrow [c_-, c_+]$$



Computed bounds completely wrong:
the spike dominates



EXAMPLE: CMS TTBAR INVARIANT MASS DATASET

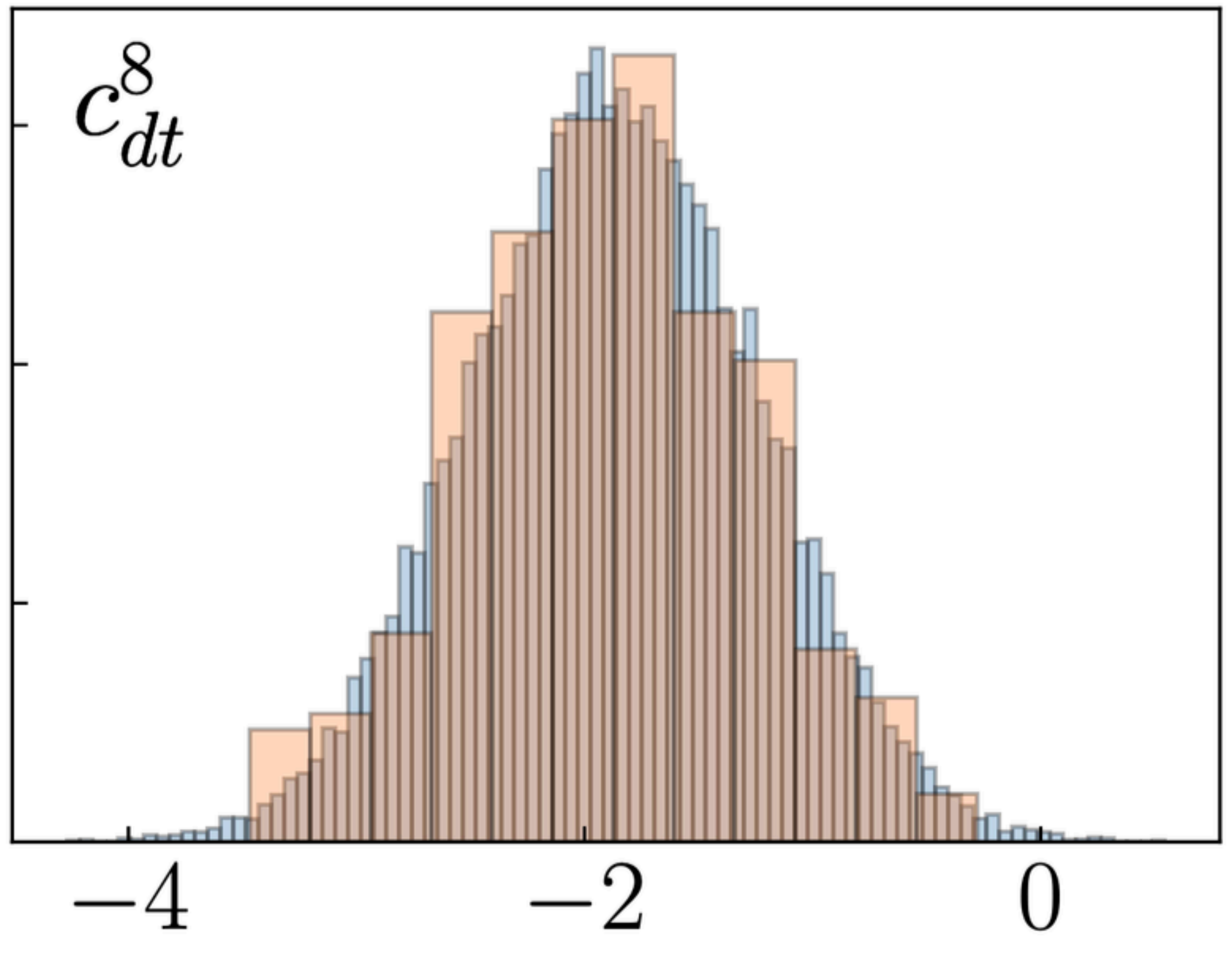


SM overshoots

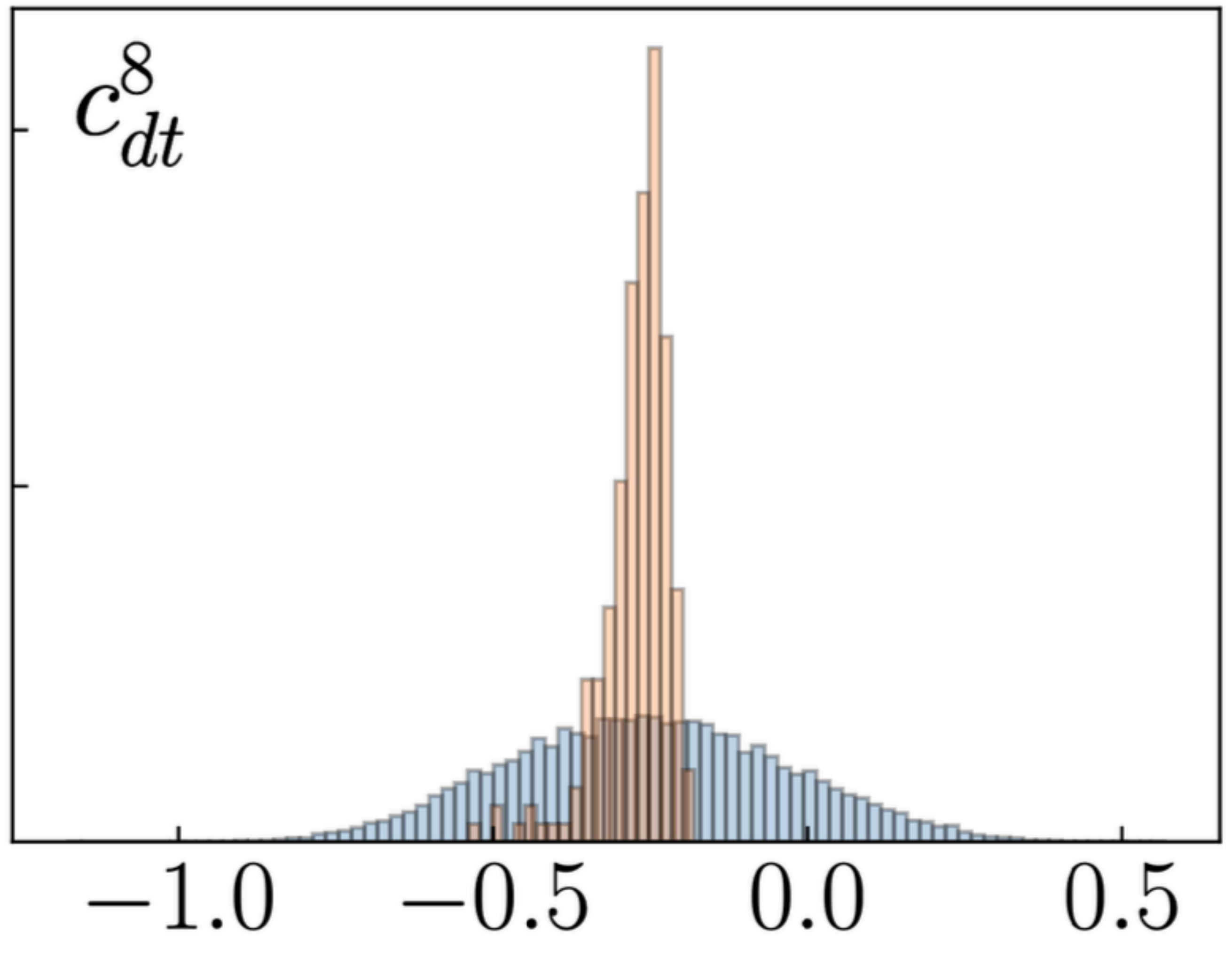
MC fit

NS fit

Linear fit



Quadratic fit



TOWARDS A NEW METHODOLOGY

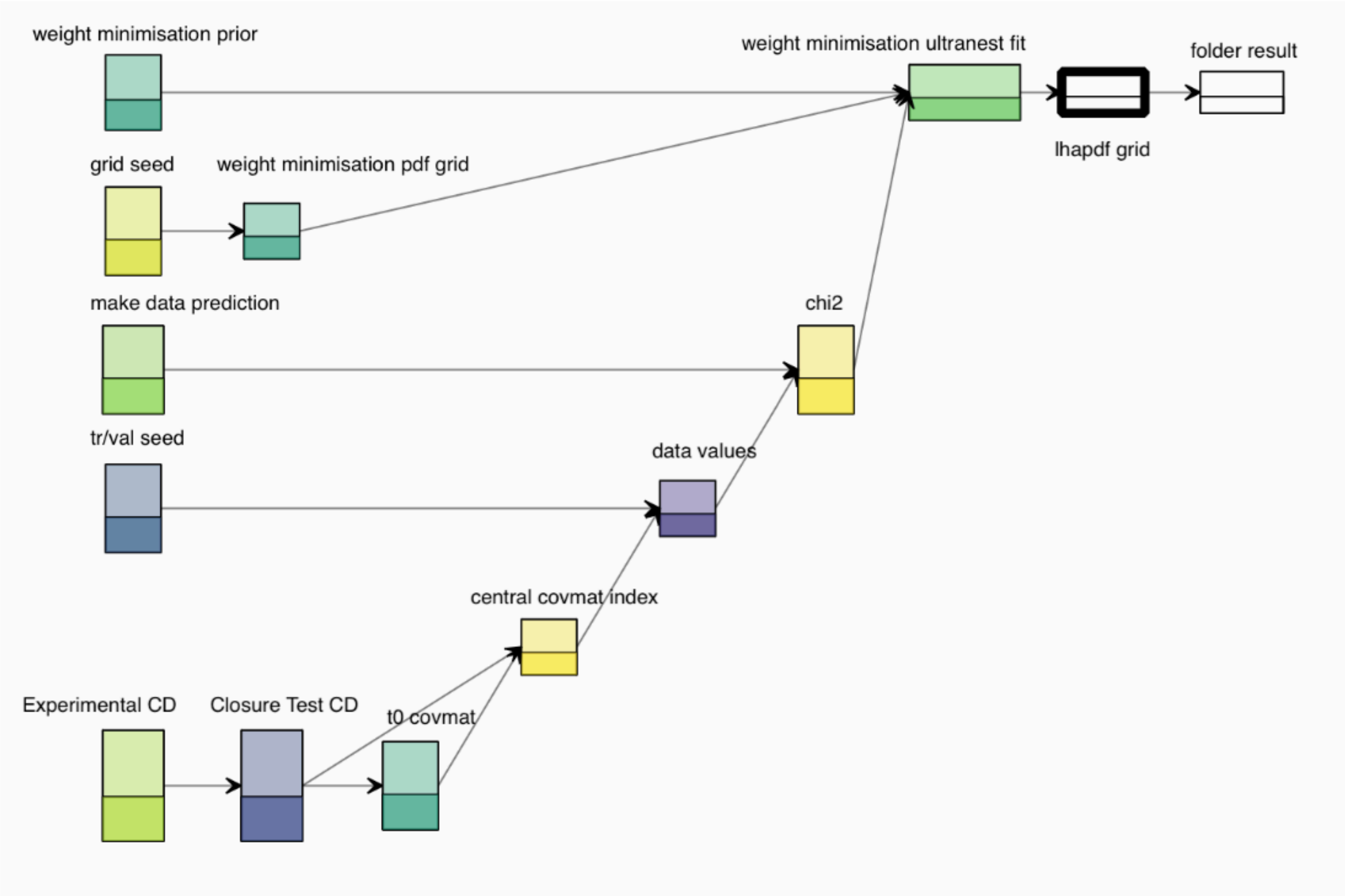
- In simultaneous PDF-SMEFT, a new PDF fitting framework is essential, one that does not rely on Monte Carlo replica error propagation.

- New Tool:

  -> perform both Monte Carlo as well as Bayesian (nested sampling) PDF fits

  -> independent on the PDF parameterisation that is being used

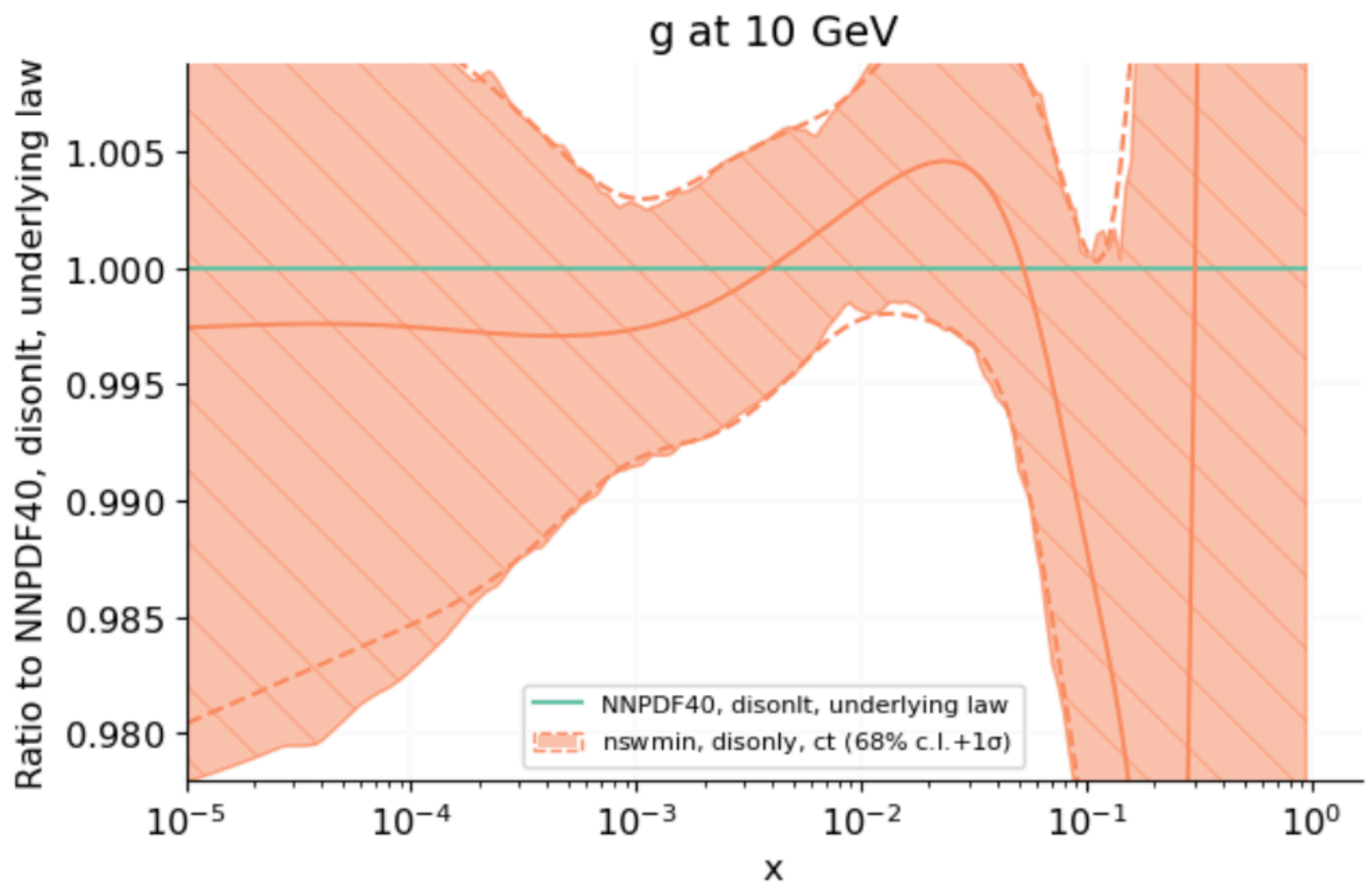
Diagrammatic Representation of Fitting Framework



WEIGHT MINIMISATION

Consider the following parameterisation:

$$f_{\mathrm{WM}}(x, Q^2) = \underbrace{f^{(j)}(x, Q^2)}_{j\text{-th replica}} + \sum_{i}^{N_{\mathrm{wm}}} \underbrace{w_i}_{\text{Weights}} \left(f^{(i)}(x, Q^2) - f^{(j)}(x, Q^2) \right)$$



-> Sum rules are automatically satisfied (provided that f satisfies them)

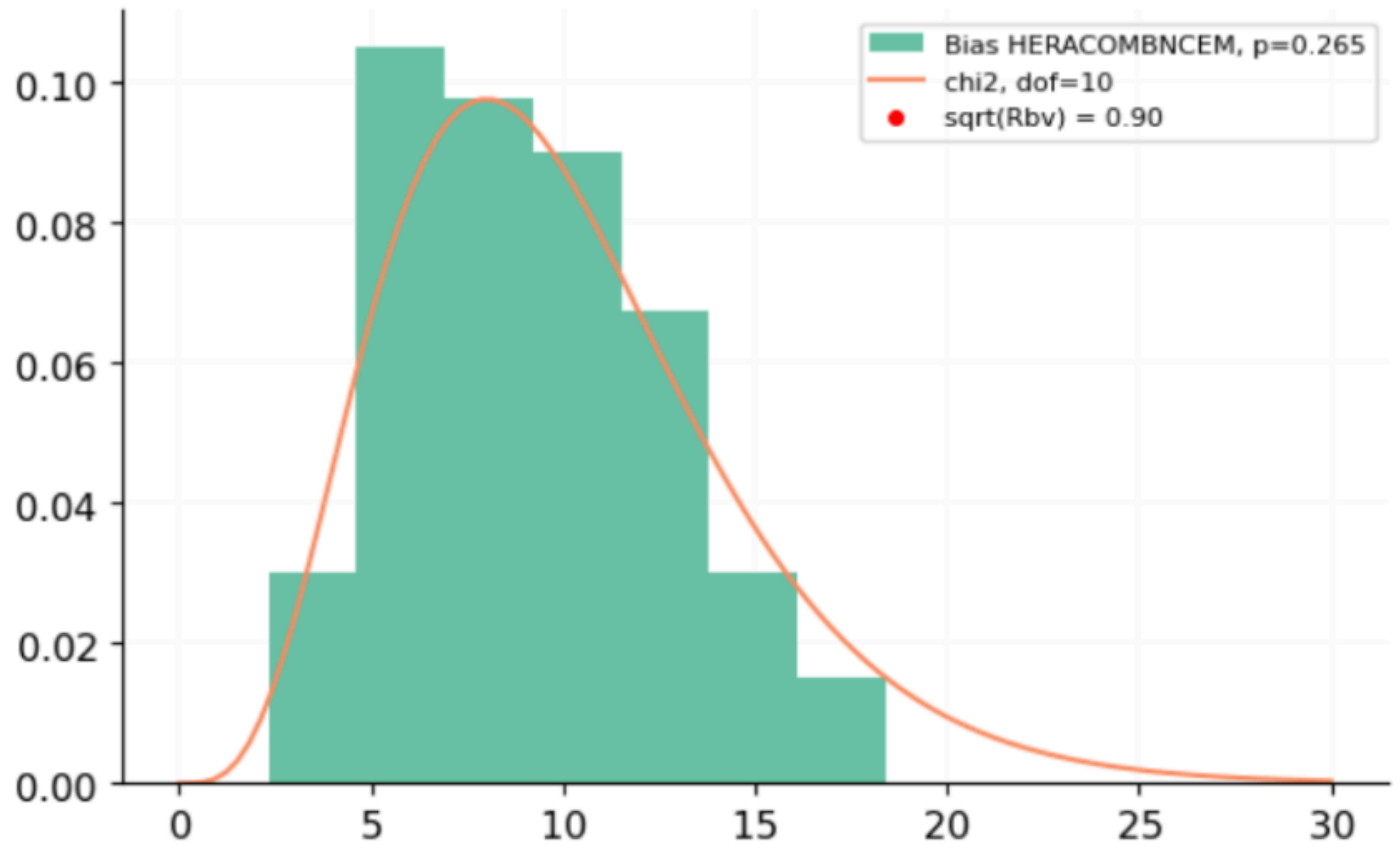
-> Easy to extend so as to include SMEFT coefficients dependence

MORE THAN PBSP

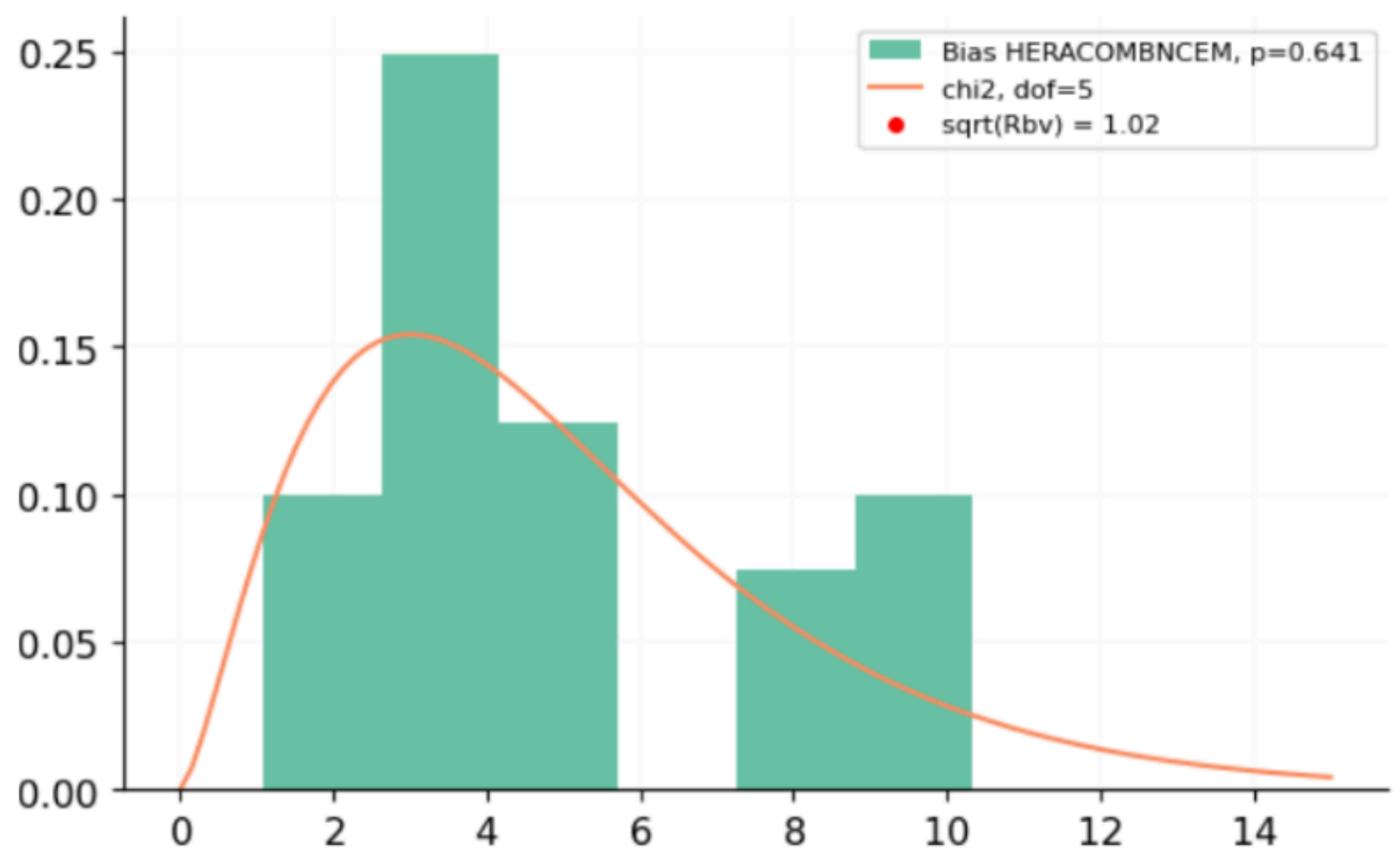
CLOSURE TEST WITH INCONSISTENT DATA

- Extend the (NNPDF) closure test framework so as to include data inconsistencies
- Data inconsistencies = underestimation of experimental uncertainties
- Study Bias distribution (χ^2 distributed) to assess how well the Neural Network (PDF model) is able to reabsorb the inconsistency

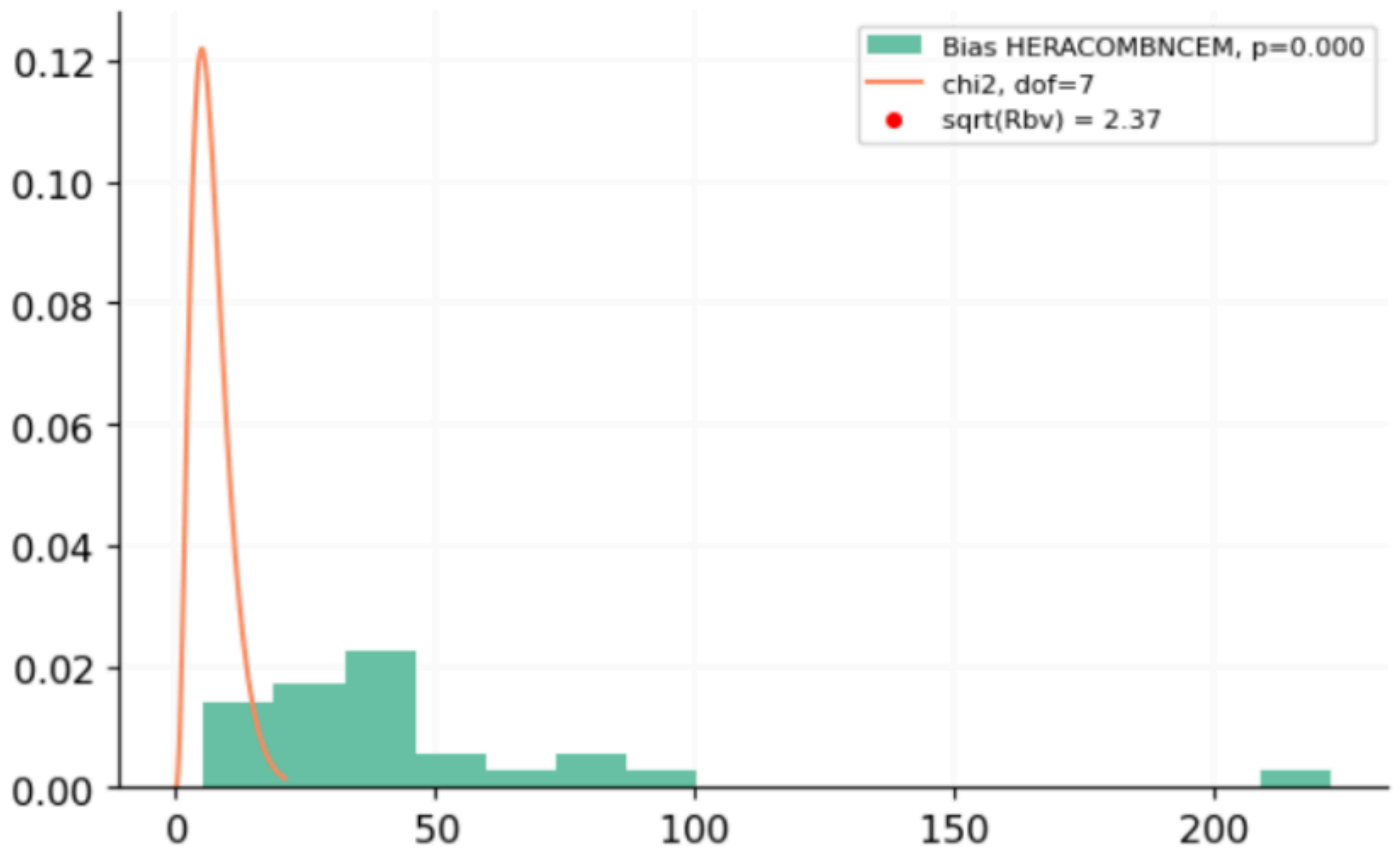
Consistent



Medium Inconsistency



Large Inconsistency



SMEFT EFFECTS IN PDF EVOLUTION

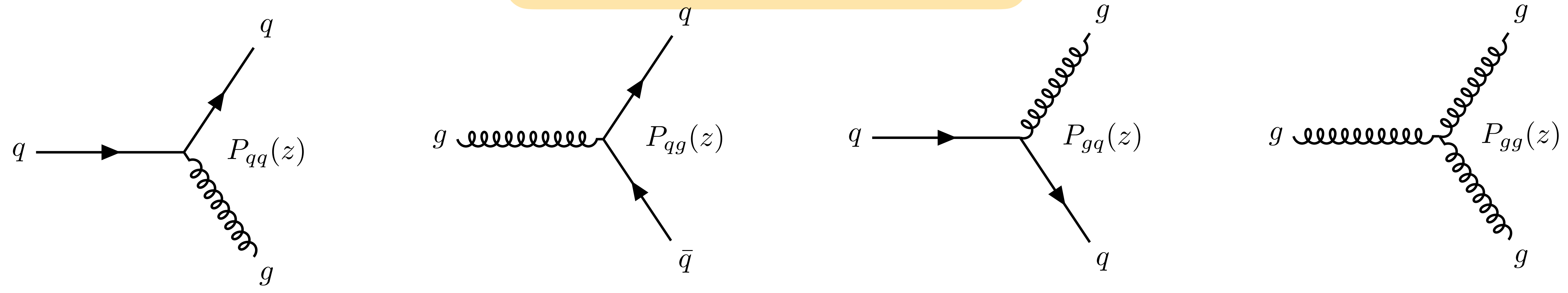
PDF evolution is crucial in global PDF fits, and it is described by the DGLAP evolution equations

$$\frac{\partial}{\partial \log(\mu^2)} \begin{pmatrix} q(x,\mu^2) \\ g(x,\mu^2) \end{pmatrix} = \frac{\alpha_S}{2\pi} \int_x^1 \frac{dz}{z} \begin{pmatrix} P_{qq}(z) & P_{qg}(z) \\ P_{gq}(z) & P_{gg}(z) \end{pmatrix} \begin{pmatrix} q(x/z,\mu^2) \\ g(x/z,\mu^2) \end{pmatrix}$$

Manu

We are interested in assessing explicitly if the SMEFT can affect the DGLAP equations in terms of

splitting functions

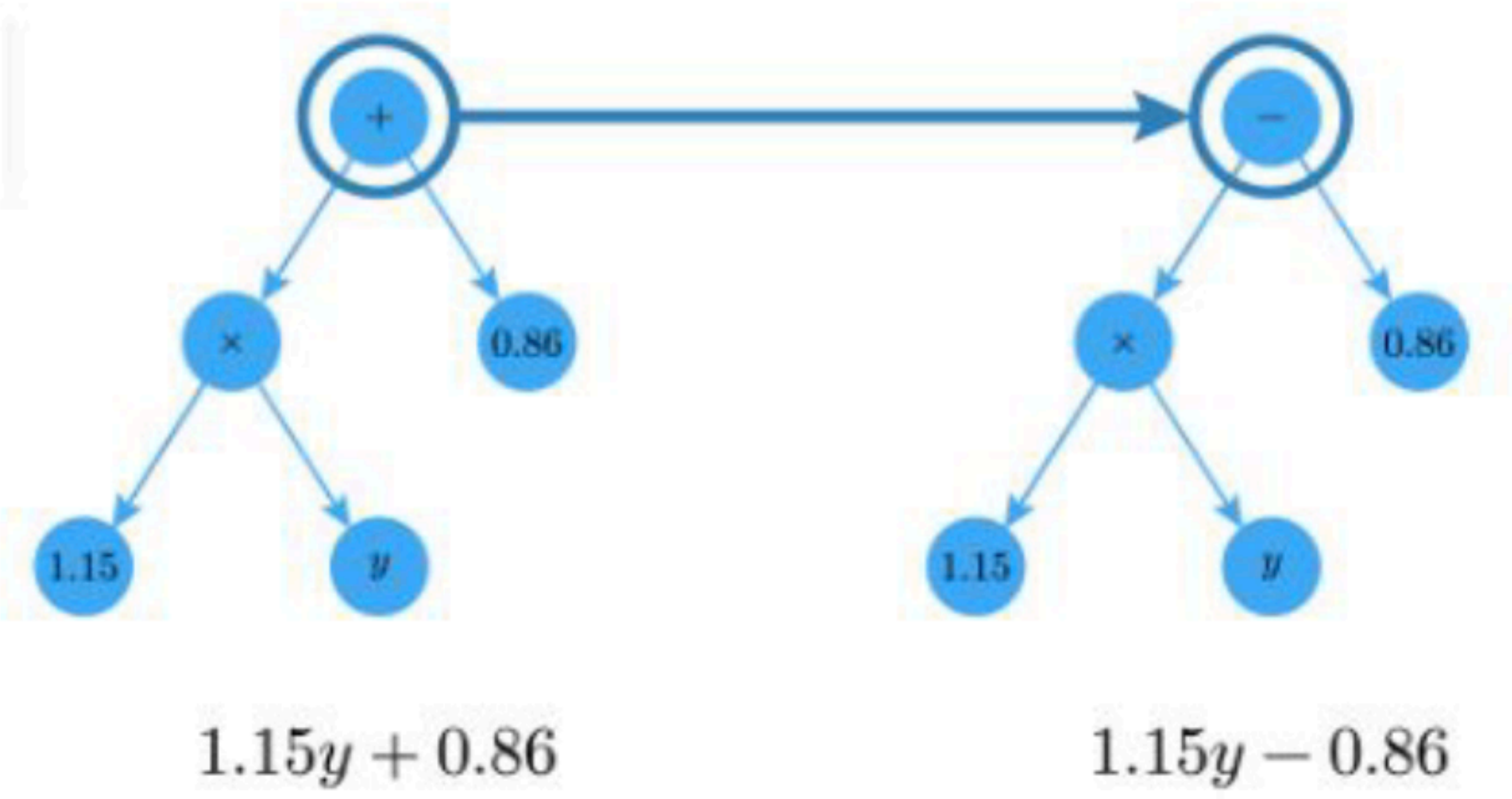


Shifts in the running of SM parameters

$$\alpha_S(\mu)$$

SYMBOLIC REGRESSION

- We want to understand the cause of the decision made by a Machine Learning model (interpretability).
- Math symbols is the language we've learned to understand patterns
- Symbolic regression automates our search for understandable patterns in the form of math symbols by making use of evolutionary algorithms.
  - PySR is an open-source library for practical symbolic regression. It uses a multi-population evolutionary algorithm



A mutation constitutes a possible step in a multi-population evolutionary algorithm in pySR

SYMBOLIC REGRESSION

- The differential cross-section describing the kinematics of the two Born-level leptons in W and Z production is given by

$$\frac{d^5\sigma}{dp_T\, d\eta\, dm\, d\cos\theta\, d\phi} = \frac{3}{16\pi}\frac{d^3\sigma}{dp_T\, d\eta\, dm}\left[(1+\cos^2\theta) + \sum_{i=0}^{7} P_i(\theta,\phi) A_i\right]$$

- We simulate, for example, the process $pp \to \mu^+\mu^-$ at MINNLOPS. The data obtained can be used to fit an empirical model for the cross section and angular coefficients A_i.
- In particular, we want a fully **interpretable** model for the coefficients in terms of a mathematical expression that depends on pT , |η| and m
  - We turn to **symbolic regression**: PySR is a open-access Machine Learning library which uses multi-population evolutionary algorithm to obtain mathematical expressions

QUANTUM TOMOGRAPHY @ LHC

$$\rho = \frac{\mathbb{1}_2 \otimes \mathbb{1}_2 + B_i^+ \sigma^i \otimes \mathbb{1}_2 + B_i^- \mathbb{1}_2 \otimes \sigma^i + C_{ij}\, \sigma^i \otimes \sigma^j}{4}.$$

$$\frac{1}{\sigma}\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega_+ \mathrm{d}\Omega_-} = \frac{1 + \mathbf{B}^+ \cdot \hat{\mathbf{q}}_+ - \mathbf{B}^- \cdot \hat{\mathbf{q}}_- - \hat{\mathbf{q}}_+ \cdot \mathbf{C} \cdot \hat{\mathbf{q}}_-}{(4\pi)^2}$$

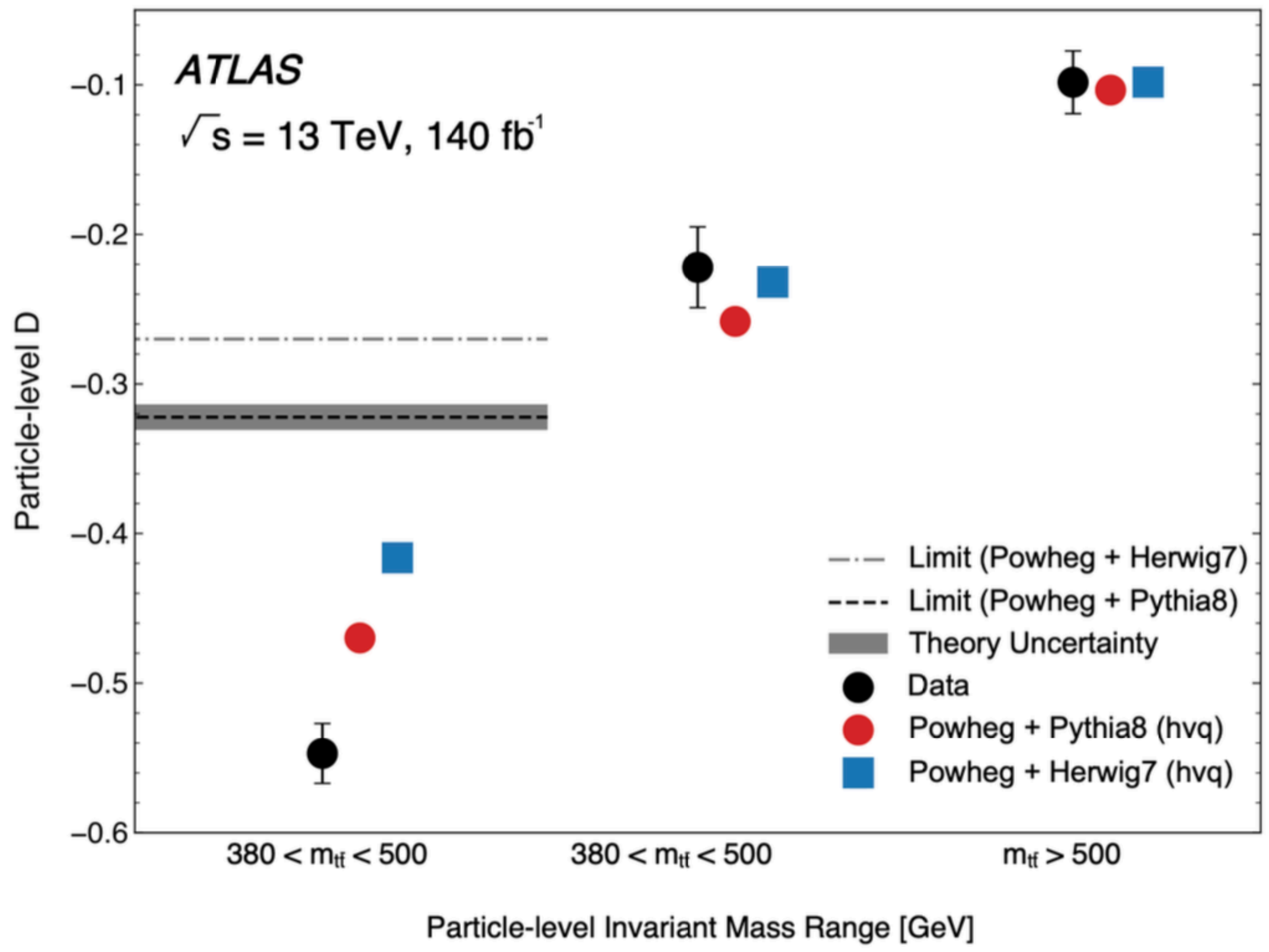
Direction of decay produced lepton

Interestingly, at threshold,
a specific angular distribution is
directly proportional to entanglement

$$C[\rho] = \max(-1 - 3D, 0)/2$$

$$\frac{1}{\sigma}\frac{d\sigma}{d\cos\varphi} = \frac{1}{2}(1 - D\cos\varphi)$$

$$D = \frac{\mathrm{tr}[\mathbf{C}]}{3}$$



[arXiv: 2311.07288]