PROCEEDINGS OF SCIENCE

Determining SMEFT and PDF parameters simultaneously based on the CTEQ-TEA framework

Xiao-Min Shen,[a,b,*] **Jun Gao,**[a,b] **MeiSen Gao,**[a,b] **T. J. Hobbs**[c] **and DianYu Liu**[a,b]

[a] *INPAC, Shanghai Key Laboratory for Particle Physics and Cosmology, School of Physics and Astronomy, Shanghai Jiao-Tong University, Shanghai 200240, China*
[b] *Key Laboratory for Particle Astrophysics and Cosmology (MOE), Shanghai 200240, China*
[c] *High Energy Physics Division, Argonne National Laboratory, Lemont, IL 60439, USA*

E-mail: xmshen137@sjtu.edu.cn

The SM effective field theory (SMEFT) provides a model-independent and systematically improvable framework for new physics searches. In this talk, we outline our approach to simultaneously fitting SMEFT parameters and Probability Density Functions (PDFs) in an extension of the CT18 global analysis framework. To enhance the efficiency of our global fitting and Lagrange multiplier scans, we leverage machine-learning techniques. We focus on several representative operators relevant to top-quark pair production and jet production. Through this approach, we establish self-consistent limitations on the associated Wilson coefficients, and explore the correlations between these Wilson coefficients and the PDFs.

The European Physical Society Conference on High Energy Physics (EPS-HEP2023)
21-25 August 2023
Hamburg, Germany

*Speaker

PoS(EPS-HEP2023)291

1. Introduction

The Standard Model effective field theory (SMEFT) is a powerful framework for indirect beyond Standard Model (BSM) searches. Assuming lepton number conservation, the SMEFT Lagrangian may be written as

$$\mathcal{L}_{\text{SMEFT}} = \mathcal{L}_{\text{SM}} + \sum_i \frac{C_i}{\Lambda^2} O_i^{(6)} + \cdots , \tag{1}$$

where Λ is a mass scale usually chosen well above the electroweak scale, the Wilson coefficient C_i/Λ^2 quantifies the contribution of the dim-6 operators $O_i^{(6)}$ observed at a relatively low energy scale. Indirect BSM searches in the SMEFT framework have the merit of being model-independent and systematically improvable, and it will be sufficient to consider dim-6 operators only if the NP scale is large enough.

However, SMEFT-based BSM searches often involve the same data sets used in the studies of parton distribution functions (PDFs). So SMEFT analyses may be biased if the theoretical predictions are calculated with SM PDFs, which are extracted from the data assuming the absence of new physics. A more consistent way to determine the BSM parameters is to extract both SMEFT and PDF parameters simultaneously [1–8]. In Ref. [6], we perform such a joint SMEFT-PDF fit in the CT18 framework [9], and explore possible correlations between SMEFT and PDF parameters.

2. The framework

Our joint SMEFT-PDF fit uses all the data sets of the baseline CT18 fit, together with additional $t\bar{t}$ and jet data from the Tevatron and the LHC. All the $t\bar{t}$ production and jet production data sets used in our fit are summarized in Ref. [6].

As a demonstration study, we perform joint SMEFT-PDF fits with a selected set of dim-6 SMEFT operators which contribute to $t\bar{t}$ production and jet production at hadron colliders. For $t\bar{t}$ production, we consider the BSM contributions from

$$\begin{aligned}
O_{tu}^1 &= \sum_{i=1}^{2} \left(\bar{t}\gamma_\mu t\right)\left(\bar{u}_i\gamma^\mu u_i\right) , \\
O_{td}^1 &= \sum_{i=1}^{3} \left(\bar{t}\gamma^\mu t\right)\left(\bar{d}_i\gamma_\mu d_i\right) , \\
O_{tG} &= i g_s (\bar{q}_{L3}\tau^{\mu\nu}T^A t)\tilde{\varphi}G_{\mu\nu}^A + \text{h.c.} , \\
O_{tq}^8 &= \sum_{i=1}^{2} (\bar{q}_{Li}\gamma_\mu T^A q_{L,i})(\bar{t}\gamma^\mu T^A t) ,
\end{aligned} \tag{2}$$

where u_i, d_i are the right-handed quarks and q_{Li} is the left-handed quark doublet of the i-th generation, t is the right-handed top quark, φ is the Higgs doublet, $G_{\mu\nu}^A$ is the gluon field strength tensor, and g_s is the strong coupling. All the relevant Wilson coefficients are assumed to be real in our work. For jet production, we consider only BSM contribution from one of the quark contact interactions

$$O_1 = 2\pi \left(\sum_{i=1}^{3} \bar{q}_{L,i}\gamma_\mu q_{Li}\right)\left(\sum_{j=1}^{3} \bar{q}_{Lj}\gamma^\mu q_{Lj}\right) .$$

PoS(EPS-HEP2023)291

All the other DIS and Drell-Yan (DY) data sets used in our fit are not affected by these SMEFT operators at NLO in QCD.

We include full EFT contributions from these dim-6 operators at NLO in QCD, which are of the form

$$\frac{d\sigma}{d\hat{O}} - \frac{d\sigma_{\rm SM}}{d\hat{O}} = \sum_i \frac{d\tilde{\sigma}_i}{d\hat{O}} \frac{C_i}{\Lambda^2} + \sum_{i,j} \frac{d\tilde{\sigma}_{ij}}{d\hat{O}} \frac{C_i C_j}{\Lambda^4}, \tag{3}$$

for observable $\hat{O}$, where C_i, C_j are the Wilson coefficients of the corresponding SMEFT operators. EFT contributions to the top quark production observables are calculated at NLO in QCD using MadGraph5_aMC@NLO [10] together with the SMEFTatNLO model [11], while contributions to jet production from quark contact interactions are calculated at NLO in QCD by the CIJet framework [12, 13].

After collection of the data D and the corresponding theoretical predictions T, we may extract the best fit of both PDF parameters $\{\theta_{\rm PDF}\}$ and SMEFT parameters $\{C_i\}$ by minimizing the profiled log-likelihood function

$$\chi^2(\{\theta_{\rm PDF}\}, \{C_i\}) = \sum_{i,j=1}^{N_{\rm pt}} (T_i - D_i)[\mathrm{cov}^{-1}]_{ij}(T_j - D_j) \,, \tag{4}$$

where the covariance matrix includes both experimental and theoretical uncertainties. Uncertainties for all the fitted parameters are evaluated by Lagrange Multiplier scans [14, 15]. The LM scan method is generally computationally expensive when the number of parameters to be fitted is large. In order to efficiently scan the parameter space, we adopt and generalize the idea of Ref. [16], and model the profile of the log-likelihood functions by neural networks (NNs). An example of the NNs used in our work is shown in Fig. 1. Inputs of the NNs include the initial-scale PDFs, the top quark mass, the strong coupling constant, and Wilson coefficients of SMEFT. These NNs are trained so that they can predict the correct χ^2, and at the same time can be calculated much faster than the original log-likelihood functions.

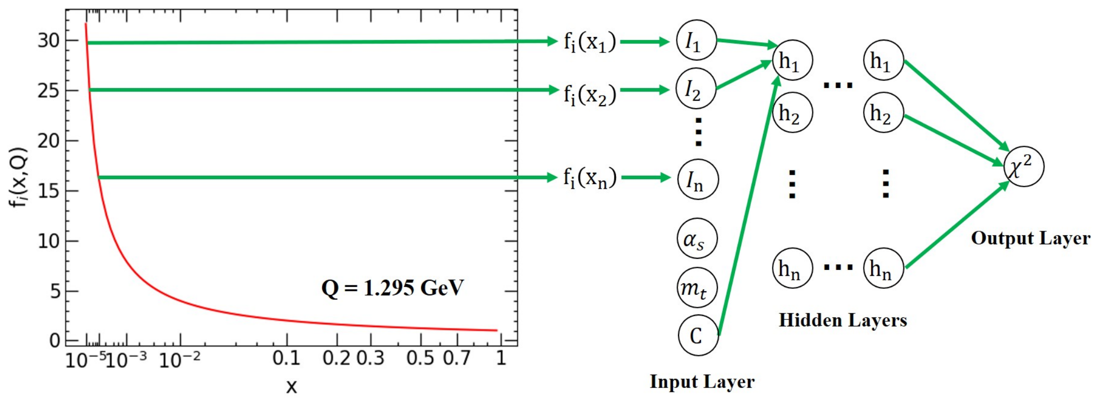

Figure 1: An example of the architecture of NNs adopted in this work.

3. Results of our joint SMEFT-PDF fits

As an example, we present here the results of our joint fits of the CT18 PDF parameters and Wilson coefficient of the quark contact interaction O_1. As mentioned above, the best-fit values of the SMEFT and PDF parameters are extracted by minimizing the χ^2 function, and the uncertainties of these parameters are predicted by LM scans. In Fig. 2, we show the results of LM scans over C_1/Λ^2, which is the Wilson coefficient of the quark contact interaction O_1. Our nominal results are shown in the left panel, where we treat PDF parameters and EFT coefficients on the same footing, and allow the CT18 PDF parameters to be changed when performing LM scans over C_1/Λ^2. The upper and lower bounds at 90% confidence level (CL) are given by the two vertical blue lines, which correspond to the CT18 tolerance criterion [9]. We find $C_1/\Lambda^2 = -0.0015^{+0.0033}_{-0.0014}$ TeV^{-2} at 90% CL, which is consistent with the SM.

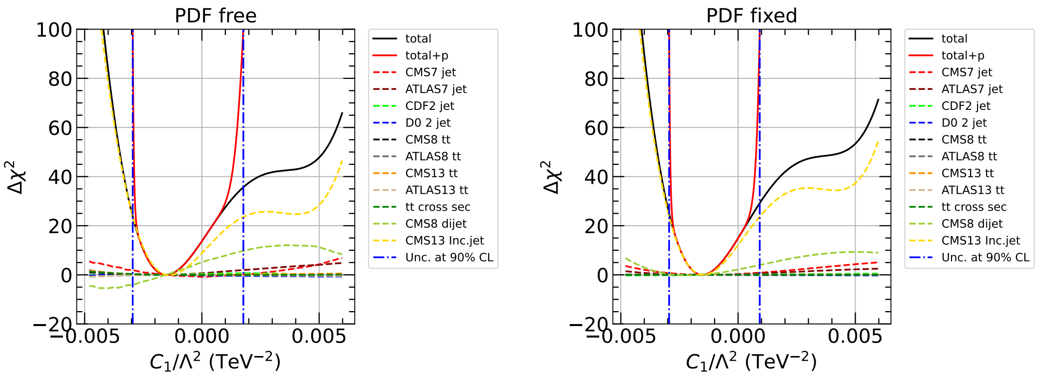

Figure 2: LM scans over C_1/Λ^2 in our nominal setup allowing the PDF parameters to freely float (left panel) or be fixed at the global minimum (right panel).

In order to study the correlation between PDF parameters and C_1, we perform similar LM scans, shown in the right panel of Fig. 2, where all the PDF parameters are fixed to their values at the global minimum of the SMEFT-PDF fit. The new LM scans give $C_1/\Lambda^2 = -0.0015^{+0.0024}_{-0.0014}$ TeV^{-2} at 90% CL, so the uncertainties are slightly underestimated as compared with our nominal result. This is a hint of small correlations between the PDFs and C_1. Results of our joint fits of PDF parameters and EFT parameters relevant for $t\bar{t}$ production can be found in Ref. [6], which give smaller correlations between PDFs and SMEFT.

The correlations between PDF and SMEFT parameters can also be studied by looking into the gluon PDF fitted with or without SMEFT contributions. In Fig. 3, we compare gluon PDFs at the initial scale $Q_0 = 1.295$ GeV determined by fitting with and without SMEFT contributions from O_1 and O_{tG}. As we can see, with the inclusion of SMEFT contributions, the gluon PDF is slightly changed in the large x region, but the variation is less than 5%. Also, the gluon PDF uncertainties are slightly enlarged with the inclusion of SMEFT operators.

Finally, we want to emphasize that the mild correlations between SMEFT and PDF may become larger at the HL-LHC for other future experiments. This possibility is also explored in Ref. [6].

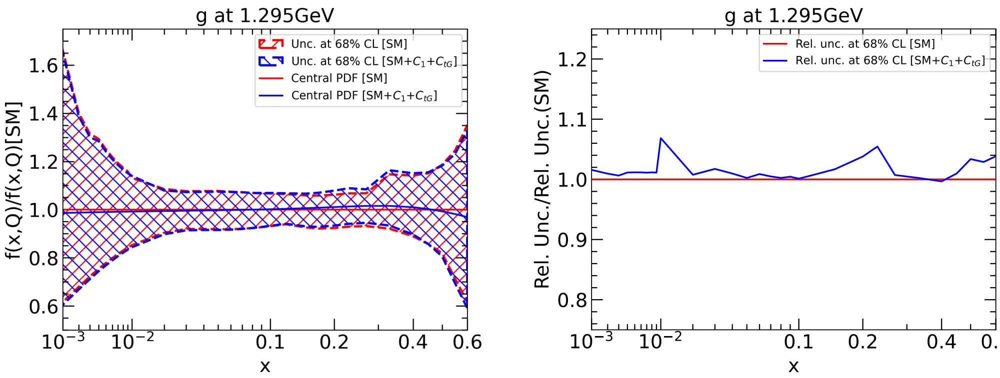

Figure 3: The central values and 68% CL uncertainties of the gluon PDF determined by fitting with and without SMEFT contributions from both O_1 and O_{tG} at $Q_0 = 1.295$ GeV. All the results are normalized to the central values of the SM gluon PDF. The ratios of relative uncertainties at 68% CL are shown in the right panel.

4. Summary

We present our methodology for performing simultaneous SMEFT-PDF fits based on an extension of the CT18 global analysis [6], where SMEFT corrections with full PDF dependence are calculated at NLO in QCD. In our work, global fits and Lagrange multiplier scans in the PDF+SMEFT parameter space are boosted with machine learning techniques, and our framework can be straightforwardly generalized to include more SMEFT parameters.

We show our results of self-consistent determination of the possible BSM effects in top-quark pair production and jet production at hadron colliders. We find mild correlations between SMEFT parameters and the PDFs, especially, the gluon PDF in the large momentum fraction region. We find these correlations may strengthen with growing precision and may be relevant for SMEFT fit on future colliders.

Acknowledgments

This work was sponsored by the National Natural Science Foundation of China under Grant No.12275173 and No.11835005. JG thanks the sponsorship from Yangyang Development Fund. The work of TJH at Argonne National Laboratory was supported by the U.S. Department of Energy, Office of Science, under Contract No. DE-AC02-06CH11357. XS is supported in part by the Helmholtz-OCPC Postdoctoral Exchange Program under Grant No. ZD2022004.

References

[1] **ZEUS** Collaboration, H. Abramowicz *et. al.*, *Limits on contact interactions and leptoquarks at HERA*, *Phys. Rev. D* **99** (2019), no. 9 092006 [1902.03048].

[2] S. Carrazza, C. Degrande, S. Iranipour, J. Rojo and M. Ubiali, *Can New Physics hide inside the proton?*, *Phys. Rev. Lett.* **123** (2019), no. 13 132001 [1905.05215].

[3] A. Greljo, S. Iranipour, Z. Kassabov, M. Madigan, J. Moore, J. Rojo, M. Ubiali and C. Voisey, *Parton distributions in the SMEFT from high-energy Drell-Yan tails*, *JHEP* **07** (2021) 122 [2104.02723].

[4] **CMS** Collaboration, A. Tumasyan *et. al.*, *Measurement and QCD analysis of double-differential inclusive jet cross sections in proton-proton collisions at $\sqrt{s}$ = 13 TeV*, *JHEP* **02** (2022) 142 [2111.10431].

[5] S. Iranipour and M. Ubiali, *A new generation of simultaneous fits to LHC data using deep learning*, *JHEP* **05** (2022) 032 [2201.07240].

[6] J. Gao, M. Gao, T. J. Hobbs, D. Liu and X. Shen, *Simultaneous CTEQ-TEA extraction of PDFs and SMEFT parameters from jet and $t\bar{t}$ data*, *JHEP* **05** (2023) 003 [2211.01094].

[7] Z. Kassabov, M. Madigan, L. Mantani, J. Moore, M. Morales Alvarado, J. Rojo and M. Ubiali, *The top quark legacy of the LHC Run II for PDF and SMEFT analyses*, *JHEP* **05** (2023) 205 [2303.06159].

[8] E. Hammou, Z. Kassabov, M. Madigan, M. L. Mangano, L. Mantani, J. Moore, M. M. Alvarado and M. Ubiali, *Hide and seek: how PDFs can conceal New Physics*, 2307.10370.

[9] T.-J. Hou *et. al.*, *New CTEQ global analysis of quantum chromodynamics with high-precision data from the LHC*, *Phys. Rev. D* **103** (2021), no. 1 014013 [1912.10053].

[10] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli and M. Zaro, *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, *JHEP* **07** (2014) 079 [1405.0301].

[11] C. Degrande, G. Durieux, F. Maltoni, K. Mimasu, E. Vryonidou and C. Zhang, *Automated one-loop computations in the standard model effective field theory*, *Phys. Rev. D* **103** (2021), no. 9 096024 [2008.11743].

[12] J. Gao, C. S. Li, J. Wang, H. X. Zhu and C. P. Yuan, *Next-to-leading QCD effect to the quark compositeness search at the LHC*, *Phys. Rev. Lett.* **106** (2011) 142001 [1101.4611].

[13] J. Gao, *CIJET: A program for computation of jet cross sections induced by quark contact interactions at hadron colliders*, *Comput. Phys. Commun.* **184** (2013) 2362–2366 [1301.7263].

[14] J. Pumplin, D. R. Stump and W. K. Tung, *Multivariate fitting and the error matrix in global analysis of data*, *Phys. Rev. D* **65** (2001) 014011 [hep-ph/0008191].

[15] D. Stump, J. Pumplin, R. Brock, D. Casey, J. Huston, J. Kalk, H. L. Lai and W. K. Tung, *Uncertainties of predictions from parton distribution functions. 1. The Lagrange multiplier method*, *Phys. Rev. D* **65** (2001) 014012 [hep-ph/0101051].

[16] D. Liu, C. Sun and J. Gao, *Machine learning of log-likelihood functions in global analysis of parton distributions*, 2201.06586.

PoS(EPS-HEP2023)291