Unravelling new physics signals at the HL-LHC with EIC and FPF constraints

Elie Hammou[*] and Maria Ubiali[†]

DAMTP, University of Cambridge, Wilberforce Road, Cambridge, CB3 0WA, United Kingdom

(Received 22 October 2024; accepted 6 May 2025; published 23 May 2025)

Recent studies suggest that global fits of parton distribution functions (PDFs) might inadvertently "fit away" signs of new physics in the high-energy tails of the distributions measured at the high luminosity program of the LHC (HL-LHC). This could lead to spurious effects that might conceal key beyond the Standard Model (BSM) signatures and hinder the success of indirect searches for new physics. In this paper, we demonstrate that future deep-inelastic scattering (DIS) measurements at the Electron-Ion Collider (EIC), and at CERN via FASERν and SND@LHC at LHC Run III, and the future neutrino experiments to be hosted at the proposed Forward Physics Facility (FPF) at the HL-LHC, provide complementary constraints on large-x sea quarks. These constraints are crucial to mitigate the risk of missing key BSM signals, by enabling precise constraints on large-x PDFs through a "BSM-safe" integration of both high- and low-energy data, which is essential for a robust interpretation of the high-energy measurements.

DOI: 10.1103/PhysRevD.111.095028

Advancements in experimental analyses and theoretical predictions have ushered the LHC into a new era of precision physics. As a proton-proton collider, the accuracy of theoretical predictions heavily depends on the accurate knowledge of parton distribution functions (PDFs). PDF uncertainties represent a significant portion of the overall theoretical uncertainty at the LHC [1–3]. Historically, PDFs were primarily extracted from HERA, Tevatron and fixed-target (FT) experiments and then used as an input at the LHC. However, LHC data now play a pivotal role in global PDF analyses. For instance, in the recent NNPDF4.0 [4] determination, LHC data constituted nearly a third of the experimental input, providing unique constraints on large-x gluons, quarks, and antiquarks. The larger the mass and/or the rapidity of the final state, the larger is the kinematic region in x that is probed by the measurement, as at leading order the fraction of the proton's momentum carried by each parton is given by $x_{1,2} = Q/\sqrt{S}\exp(\pm\eta)$, where $\sqrt{S}$ is the center-of-mass energy of the experiment, Q is the energy scale of the process—roughly of the order of the mass of the final state object(s)—and η is its rapidity.

While LHC data significantly enhance PDF precision, they are also sensitive to beyond the Standard Model (BSM) dynamics. If BSM signals distort an experimental distribution included in a PDF fit—typically assumed to follow the SM—this can lead to inconsistencies. These inconsistencies may either exclude the dataset from the fit, ensuring consistent PDFs, or be missed by inconsistency tests and cause the PDFs to adapt and incorporate the BSM effects, resulting in *BSM-biased* PDFs. Such considerations cannot be considered purely speculative, as it was recently shown [5,6] that there exist BSM scenarios, in particular a scenario involving a new heavy $\mathrm{SU}(2)_L$ doublet W' with flavor universal coupling g', which would affect the tails of high energy invariant mass Drell-Yan (DY) distributions measured at the HL-LHC [7,8] and which can be completely absorbed, thus fitted away, by a flexible parametrization of the large-x quark and antiquark PDFs. As a consequence the BSM signals could be missed in indirect new physics searches as the *BSM-biased* PDFs might adapt in such a way that, when convoluted with the SM predictions, they mimic BSM effects. The BSM signal would be lost as it would appear compatible with the SM prediction; schematically,

$$\hat{\sigma}_{\mathrm{BSM}} \otimes f_{\mathrm{SM}} \approx \hat{\sigma}_{\mathrm{SM}} \otimes f_{\mathrm{BSM}}, \tag{1}$$

where $\hat{\sigma}_{\mathrm{BSM}}$ and $\hat{\sigma}_{\mathrm{SM}}$ correspond to the partonic cross sections computed according to a given BSM model and to the SM respectively, f_{SM} is the *true* SM PDF that parametrizes the proton's structure, which is obviously decoupled from any heavy BSM particles, while f_{BSM} is the *BSM-biased* PDF obtained from an inconsistent fit that has absorbed the BSM signals in the large-x PDFs.

[*]Contact author: eh651@cam.ac.uk
[†]Contact author: m.ubiali@damtp.cam.ac.uk

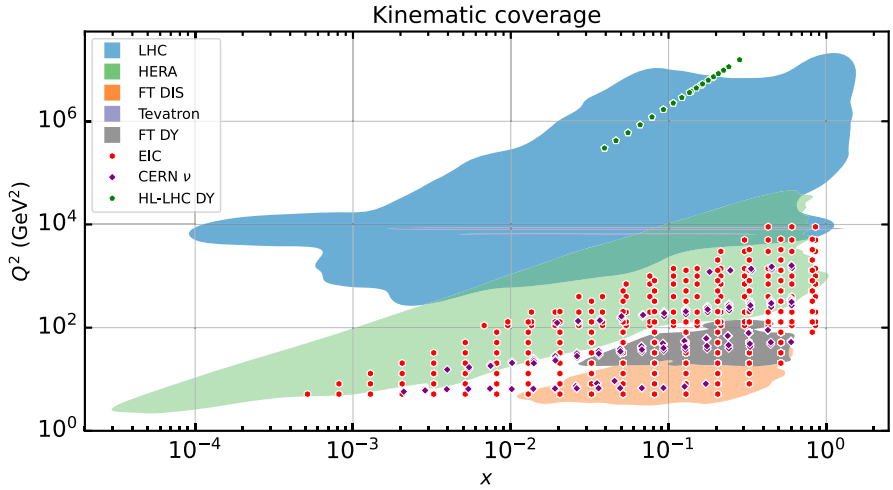

FIG. 1. Kinematic coverage of the data (shaded region) and projected data (scattered points) used in this study.

In this paper we show that deep-inelastic scattering (DIS) measurements from the Electron Ion Collider (EIC) at Brookhaven National Laboratory [9,10] along with neutrino DIS structure function measurements at CERN from FASERν [11–13] and SND [14,15] at Run III, and at the proposed Forward Physics Facility (FPF) experiments [16,17]—FASERν2, AdvSND, and FLArE—not only provide complementary sensitivity on PDF flavor decomposition [18–21] and on SMEFT Wilson coefficients [22,23], but that these experiments are essential for indirect BSM searches at high energy. The constraints on the large-x sea quarks and gluons that they provide are the key for a robust identification of BSM signal at the HL-LHC and at FCC-he and FCC-hh [24], for example in the Drell-Yan forward-backward asymmetry [25–27], thus proving once more the powerful synergy between diverse experimental programs.

In Fig. 1 we display the kinematic coverage of the data (shaded region) and projected data (scattered points) used in this study. The full list is provided in the Appendix. The green shaded area corresponds to the HERA measurements [28–30], the orange area to the fixed-target DIS data [31–35], the purple lines to the Tevatron measurements [36–41], the blue area corresponds to LHC data [42–71] and the gray one to fixed-target Drell-Yan (FTDY) measurements [72–74] included in the fit. The scattered points correspond to projected data for the EIC and the CERN neutrino experiments as well as for high-energy DY distributions at the HL-LHC. The main benefit of additional precise measurements covering the relatively low-energy region $Q \lesssim 100$ GeV on top of the high-energy region of $Q \approx 1$–3 TeV from the HL-LHC is that such observables, contrarily to the ones measured at the HL-LHC, are not susceptible to the effects of the BSM models identified in [5], and could therefore show a tension with the high-energy data, thus flagging a BSM-induced inconsistency in the global PDF fit.

In this work, as in Ref. [5], results are based on closure-test fits, which allow a systematic study of BSM-induced effects in PDF fits by isolating them from all other effects such as missing higher order uncertainties in theory predictions and experimental inconsistencies. Closure-test fits are done by generating the central values of the data included in the fit according to a known underlying law, given by a model and a perturbative order that determine the *true* partonic cross sections and a reference set of PDFs representing the *true* structure of the proton, and let them fluctuate according to experimental uncertainties. If a PDF fitting methodology is successful and the model used in the fit is the same as the model used in the data generation, then the fitted PDFs should reproduce the underlying law and the χ^2 per data point should be centered around one for all experiments. If however the data are generated according to a given BSM model but PDFs are fitted assuming the SM, then a closure test is successful only if the PDFs manage to *absorb* the BSM shifts, as schematically indicated in Eq. (1). More details about closure tests and related statistical estimators to estimate inconsistencies are given in the Appendix.

In Fig. 2 we compare the data-theory agreement for three subsets of the data included in two closure-test fits. In both fits, the data is generated by injecting the W' model discussed in [5,7,8] with $M_{W'} = 13.8$ TeV and $g' = 1$, and PDFs are fitted assuming the SM. Details of the model under consideration are summarized in the Appendix. The three experimental subsets correspond to fixed-target Drell-Yan (FTDY) data, the full LHC Run I and Run II high-mass DY data (HM DY) included in the NNPDF4.0 global fit [45,48,75–77], and HL-LHC high-mass DY projections (HL-LHC) that were considered in [5,6,8,78,79] to assess the potential impact of high-mass HL-LHC DY distributions on the PDFs. The height of the histogram bars represents the χ^2 per data point for each data category as obtained in two fits: a closure test performed on the

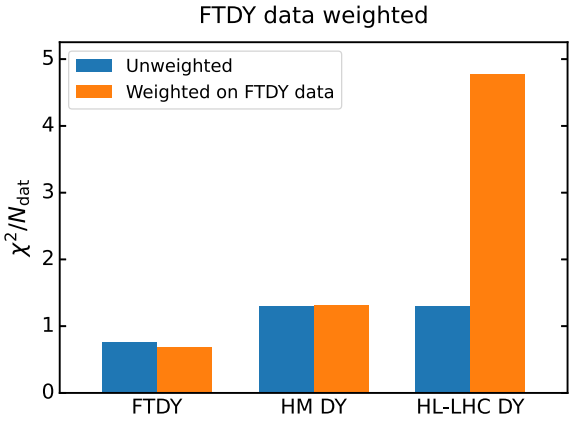

FIG. 2. Data-theory agreement for fixed-target DY data (FTDY), LHC Run I and Run II high-mass DY data (HM DY) and HL-LHC high-mass DY projections (HL-LHC) in two closure-test fits. In both fits the BSM model described in the main text is injected in the data and PDFs are fitted assuming the SM. The height of the blue histogram bars (unweighted) represents the χ^2 per data point obtained in a global NNPDF4.0-like closure-test fit augmented by the HL-LHC projections. The height of the orange bars (weighted on FTDY data) corresponds to the same quantity obtained in a similar closure test, in which more weight is given to the FTDY data.

global NNPDF4.0 dataset augmented by the HL-LHC projections (blue bars) and the same closure test in which more weight is given to the FTDY data (orange bars). In the case of the weighted fit the usual loss function

$$\chi^2 = \frac{1}{n_{\rm dat}} \sum_{i=1}^{n_{\rm dat}} n_{\rm dat}^{(i)} \chi_i^2, \tag{2}$$

is substituted by the weighted loss function

$$\chi_w^2 = \frac{1}{n_{\rm dat} - n_{\rm dat}^{(j)}} \sum_{i=1, i\neq j}^{n_{\rm exp}} n_{\rm dat}^{(i)} \chi_i^2 + w^{(j)} \chi_j^2, \tag{3}$$

in which the jth experiment, in this case the FTDY, is given a larger weight that is inversely proportional to the number of data points in the set

$$w^{(j)} = n_{\rm dat}/n_{\rm dat}^{(j)}. \tag{4}$$

The BSM model that we inject in the data affects mostly the HL-LHC DY projections and, although moderately, the Run I and Run II high-mass DY distributions. As we can see, the closure-test fit associated to the blue bars is successful, indicating that the PDF parametrization absorbs the effects of the BSM model injected in the data. In particular the fit quality of the HL-LHC data is extremely good, thus the BSM-induced inconsistency would not be spotted in the consistency checks across datasets that PDF fitters routinely run. The orange bars instead correspond to a fit in which the FTDY data was given more weight, thus features smaller uncertainties and has a stronger pull on the fit. In that case the χ^2 per data point for the FTDY data improves, while the one of the HL-LHC projections deteriorates to a level that an inconsistency would be spotted and these distributions would not be included in a global PDF analysis.

Typically, in a PDF fit, an experiment consisting of $n_{\rm dat}$ data points is considered inconsistent with the bulk of the other data included in the fit if the χ^2 per data point exceeds a critical threshold of $\chi^2_{\rm thr} = 1.5$, and if its deviation from 1 in terms of the standard deviation of a χ^2 distribution exceeds the critical threshold of 2, namely $n_\sigma = (\chi^2 - n_{\rm dat})/\sqrt{2n_{\rm dat}} > 2$. The thresholds for n_σ and the individual χ^2 in our current study are taken to be the same as in the NNPDF global analyses, and the choice of their specific values are discussed in detail in [4]. If a given dataset is flagged according to such criteria, an inconsistency with the other data entering the fit is outlined. Then, as in the NNPDF analyses, a weighted fit is performed, and if the inconsistency persists even upon weighting the flagged dataset or, if the other datasets are badly fitted upon weighting the flagged dataset, then the dataset is removed from the global fit.

To summarize, in the standard fit (blue bars) the FTDY data is not precise enough to flag the BSM-induced bias in the PDFs. However, when extra weight is given to it, the χ^2 of the HL-LHC increases dramatically, flagging the BSM-induced bias (orange bars). The reason behind this finding is that the low-energy FTDY data, thanks to a combination of proton and isoscalar targets constrain the $\bar{u}/\bar{d}$ ratio at large x, thus increasing the tension with the HL-LHC pull on the antiquarks.

While FTDY experiments are not currently planned by the HEP community, the EIC [18,19] has been approved by the United States Department of Energy at Brookhaven National Laboratory, and could record the first scattering events as early as 2030.

Our analysis starts by considering the constraining power of the EIC projections that were considered in [20]. The data points added to the PDF fits are listed in Table IV in the Appendix, and their kinematic coverage is displayed in Fig. 1. We consider measurements of charged-current (CC) and neutral-current (NC) DIS processes, using both electrons and positrons as projectiles and both protons and neutrons as targets. We generated the theory predictions using the `yadism` software [80] and the pineline toolchain [81–83]. Although the EIC will be able to use a greater variety of heavy ions as targets, we focus on the ones susceptible to constrain the proton PDF. We added these observables to the PDF fit dataset alongside the projected data from the HL-LHC that is susceptible to signs of the BSM model considered in this work and performed several fits for different W' masses by keeping its flavor-universal coupling $g' = 1$.

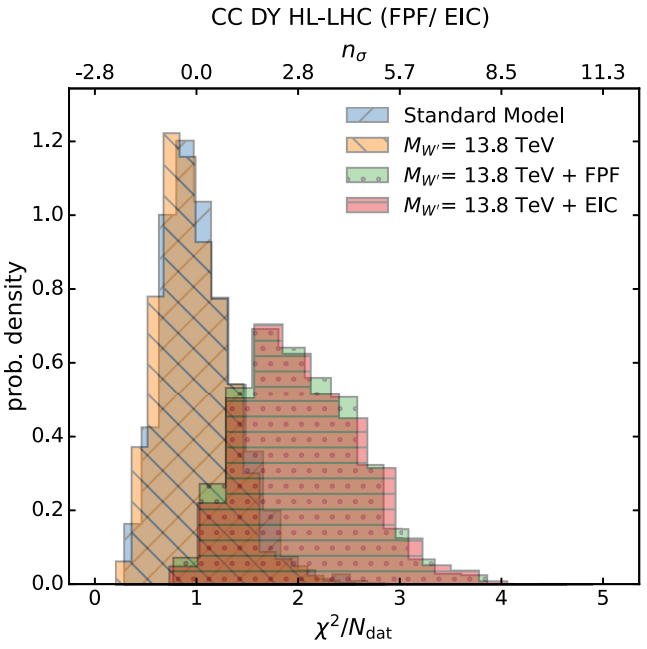

FIG. 3. Comparison of the distribution of the χ^2 per data point (bottom x-axis) and n_σ (top x-axis) for the high-mass HL-LHC DY CC projected data between a consistent fit in which both data and predictions are generated according to the SM (blue histogram), the fit biased by BSM including the HL-LHC projected data (orange histogram) and the same fit including both the HL-LHC and the EIC (red histogram) or the FPF (green histogram) projected data respectively.

The main observation is that the inclusion of EIC projections alongside the HL-LHC ones makes a significant difference in the ability of the large-x antiquark distributions to shift and fit away the BSM signal injected in the tails of the DY distributions. This can be observed in Fig. 3. There we display the $\chi^{2(k)}$ and $n_\sigma^{(k)}$ distributions across 1000 random instances (k) of the data generation in the closure-test fits for the datasets that are mostly affected by BSM corrections, in this case the HL-LHC CC DY projections. Looking at the blue and orange histograms, we observe that the global closure-test fit (in which data are generated according to a W' model with $M_{W'} = 13.8$ TeV and PDFs are fitted assuming the SM) displays the same good fit quality as the baseline (in which data are generated according to the SM and PDFs are fitted assuming the SM). As it was pointed out in Ref. [5] the value of $M_{W'} = 13.8$ TeV corresponds to the *maximal BSM threshold*, i.e., to the lowest $M_{W'}$ that can be absorbed by the PDFs without deteriorating the χ^2 of the fit, particularly of the BSM-affected data. The situation changes if the EIC projections as included. Indeed the red dotted histogram shows that the inclusion of this data would prevent PDFs from absorbing new physics, as the peak of the χ^2 and n_σ distributions shifts beyond the critical values of 1.5 and 2.0 respectively that would flag up the HL-LHC CC DY projections as inconsistent with the bulk of the datasets included in the analysis, hence it would exclude it from the global PDF fit.

Next we assess the effect of the CERN neutrino data. The recent observation of LHC neutrinos by the FASERν [11–13] and SND [14,15] far-forward experiments demonstrates that neutrino beams produced by pp collisions at the LHC can be deployed for physics studies [21]. Beyond the ongoing Run III, a dedicated suite of upgraded far-forward neutrino experiments would be hosted by the proposed Forward Physics Facility (FPF) [16,17] operating concurrently with the HL-LHC. We now investigate whether the projected measurements at the two facilities considered in [21] might provide a similar handle on the proton structure as the EIC structure function measurements, which would prevent BSM-induced bias in PDF fits. Here we only consider CC DIS processes, where the projectiles are neutrinos and antineutrinos, coming from the LHC beam, hitting isoscalar targets and the processes associated with charm production. As in [21] we ignore the SND Run III projections as the current statistics is too little to produce faithful projected data. The details of the datasets are presented in Table V in the Appendix and the kinematic coverage is displayed in Fig. 1. Looking at the green histogram in Fig. 3, we observe that, analogously to what happens in the case of the EIC, the inclusion of the FPF projections would shift the χ^2 distribution for the HL-LHC high mass DY data and would flag them up as inconsistent with the bulk of other data included in the fit. Note that the main constraint comes from the proposed upgraded far-forward neutrino experiments as FASERν has little impact on the PDFs.

The main finding is summarized in Fig. 4, in which we compare the data, featuring the BSM signal, i.e., the depletion in the tails associated with a new heavy W' with $M_{W'} = 13.8$ TeV and $g' = 1$, to the SM theoretical predictions obtained with a *BSM-biased* set of PDFs (green line) or with a set of PDFs obtained from a joint fit of the HL-LHC data and the FPF + EIC data, in which the CC HL-LHC data get flagged and excluded from the fit (red line), or finally with the *true* SM PDFs (black line). The inset shows the ratio between the BSM signal and the SM predictions for each of the PDF sets.

The green line closely matching the SM prediction demonstrates that the *BSM-biased* PDFs mimic the BSM signals present in the data and erase the deviation from the SM that should be observed, as in Eq. (1). As a result, using these PDFs in theoretical predictions would hide the BSM signal. This would not occur if the *true* PDFs (represented by the black line, which we would not know in reality) were used. However, the red line—derived from fitting the global dataset, including only the NC HL-LHC data compatible with EIC + FPF data—closely matches the black line and would allow us to detect the BSM signal.

To summarize, in this work we have identified future datasets that constrain the large-x region and can be included in PDF fits, enabling the safe integration of high-energy constraints from the HL-LHC. Our results

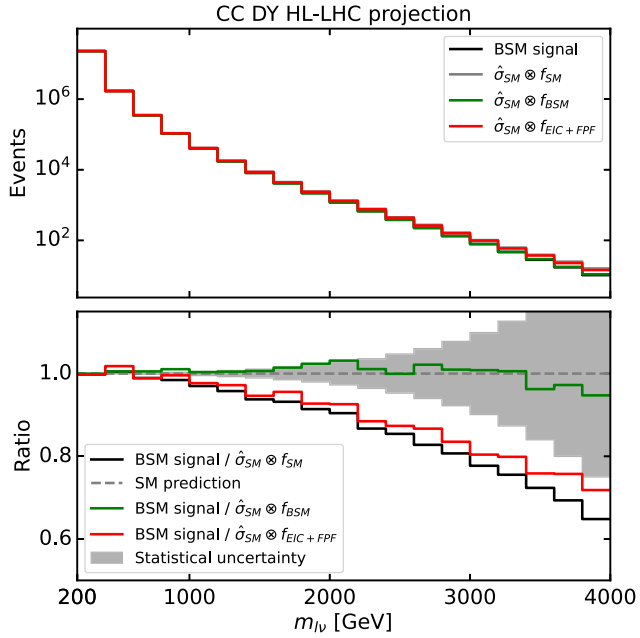

FIG. 4. Predictions for CC DY ($pp \to l^- \bar{\nu}_l$) inclusive cross sections differential in $M_{l\nu}$, with $l = (e, \mu)$. Factorization and renormalization scales are set to $\mu_F = \mu_R = M_T$, where M_T is the transverse mass of each bin. The BSM signal is compared to the SM predictions obtained with a *BSM-biased* set of PDFs (green line) or with a set of PDFs obtained from a joint fit of the HL-LHC data and the FPF + EIC data (red line), or finally with the *true* SM PDFs (black line). In the bottom inset the predictions obtained with the various PDF sets are compared to the SM prediction and its total uncertainty.

show that potential BSM effects, which might otherwise be absorbed into the PDFs, can be fully disentangled in high-energy measurements. Incorporating EIC and FPF measurements in a global PDF analysis will help minimize 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 analyzed separately.

Clearly the findings of Ref. [5], which this paper is based upon, are specific to the specific BSM models that were considered in the study. However the outcome of the study in Ref. [5] and of the follow-up study that we have presented here are far from trivial, as they show the first case of a realistic BSM model that can bias the outcome of a PDF global fit in the HL phase of the LHC, and how the complementary constraints from the future FPF and EIC facilities can prevent such an undesirable bias. We observe that in Ref. [5] another scenario was considered, corresponding to a new flavor-universal Z' boson with mass of the order of 10 TeV. Contrarily to the W' case, the effects of such a flavor-universal Z' could not be absorbed by the PDFs. The key difference between the two models is that the Z' only affects the NC Drell-Yan observables, while the W' model affects both the NC and the CC Drell-Yan observables, by thus allowing them to compensate the effects by exploiting the flat direction in the unconstrained antiquark sector.

Other very interesting directions can be explored in the future, following a similar methodology. For example, following the bottom-up SMEFT approach of Refs. [23,84] we may want to assess the impact of heavy new physics parametrized by specific combinations of SMEFT Wilson coefficients that affects not only the high tails of the invariant mass distributions, but also the somewhat lower energy DIS observables. Alternatively, one could adopt a top-down approach as in Ref. [85], in which several well-motivated BSM signals are injected into artificial data and recovered them by fitting specific distributions. Finally we could follow the approach of Ref. [86] which focuses on the renormalization group equation connection between SMEFT and UV parametrization in specific precision observables.

A further interesting study that would complement the findings of this paper would be addressing the question of whether the possible *BSM-induced* bias depends on the flexibility of the PDF parametrization. One could try to make the neural network structure used in NNPDF4.0 deeper or shallower, and check not only if an equally good fit quality can obtained without over-fitting, but also whether the BSM absorption by the PDFs becomes more or less visible. Alternatively one could perform such a test within a polynomial fit of PDFs, such as MSHT20 [87] or CT18 [88], by changing the number of parameters that affect the large-x behavior of PDFs and assess the effect of BSM-induced inconsistencies on the fit quality and on the resulting PDFs. For example, in Ref. [89] the authors use a closure test approach similar to the one that we describe in this work, and inject experimental inconsistencies in their input dataset to study their impact on the resulting PDFs. The methodology of [89] could be easily adapted to the injection of *BSM-induced* inconsistencies in the closure test data to assess whether the phenomenon we observe happens, and whether it is correlated with a particular choice of parametrization.

A further natural follow-up of our current study would be the extension to more complex BSM scenarios affecting both the large-x quarks and gluons. Both the code and the projected data used in this study are publicly available at the `SIMUnet` [6] repository.[1] Users are encouraged to test the robustness of our findings against different new physics scenarios and experimental projections.

ACKNOWLEDGMENTS

We are grateful to Juan Rojo for his precious suggestions concerning the presentation of our results. We thank Juan Rojo, Tanjona Rabemananjara and Juan Cruz-Martinez for providing us with the FPF projections and the

[1] https://github.com/HEP-PBSP/SIMUnet

corresponding theoretical predictions. We thank Emanuele Nocera for his assistance in identifying the EIC projected data and in generating theoretical predictions. We are most grateful for Felix Hekhorn for his precious help in using EKO and YADISM to generate interpolating grids for DIS observables. E. H. and M. U. are supported by the European Research Council under the European Union's Horizon 2020 research and innovation program (Grant agreement No. 950246), and partially by the STFC consolidated Grant No. ST/T000694/1 and No. ST/X000664/1.

APPENDIX: ESTIMATORS AND LIST OF DATA AND PSEUDODATA

In this Appendix, we provide additional information about the methodology and the definition of the statistical estimators used in our analysis. We also provide a list of the experimental data and pseudodata that we include, as well as details on the BSM model that we consider throughout the paper. Additionally, we show how our findings vary as a weaker BSM signal is injected in the data. We investigate what is the maximal BSM signal that may be absorbed in PDF fit even in the presence of EIC and FPF data and what the effect would be on indirect BSM searches.

Closure test. In order to systematically study possible BSM-induced bias in PDF fits, we work in a setting in which we assume we know the underlying law of nature. In our case the law of nature consists of the *true* PDFs, which are low-energy quantities that have nothing to do with new physics, and the *true* UV complete Lagrangian, which at low energy is well approximated by the SM Lagrangian but nonetheless comprises heavy new particles. We use these assumptions to generate the pseudodata in our analysis. We inject the effect of the new particles that we introduce in the Lagrangian in the artificially generated data, and their effect will be visible in some high-energy distributions depending on the underlying model. The methodology we use throughout this study is based on the NNPDF *closure-test* framework, first introduced in Ref. [90], and explained in more detail in Ref. [91]. Various statistical estimators are applied to check the quality of the fit (in broad terms, assessing its difference from the *true* PDFs), hence verifying the accuracy of the fitting methodology.

Detection of inconsistencies in global PDF fits. Once we produce a closure test fit, in which PDFs have been fitted assuming the SM while the pseudodata have been produced according to a given BSM model, we check whether it is possible to detect BSM-induced bias in the PDF fit, i.e., whether the datasets which *are* affected by the given BSM appear inconsistent with the bulk of the data in the global analysis and are poorly described by the resulting PDFs. To quantitatively address this point, we use the NNPDF dataset selection criteria, discussed in detail in [4]. We consider both the χ^2-statistic of the resulting fit to each dataset entering the fit, and also consider the number of standard deviations

$$n_\sigma = (\chi^2 - n_{\rm dat})/\sqrt{2n_{\rm dat}} \tag{A1}$$

of the χ^2-statistics from the expected χ^2 for each dataset. If $\chi^2/N_{\rm dat} > 1.5$ and $n_\sigma > 2$ for a particular dataset, the dataset would be flagged by the NNPDF selection criteria, indicating an inconsistency with the other data entering the fit.

Weighted fits. There are two possible outcomes from performing such a dataset selection analysis on a fit. In the first instance, the datasets affected by NP are flagged by the dataset selection criterion. If a dataset is flagged according to this condition, then a weighted fit is performed, i.e., a fit in which a dataset (j) is given a larger weight inversely proportional to the number of data points, according to Eqs. (2) and (4). If the data-theory agreement improves by setting it below thresholds and the data-theory agreement of the other datasets does not deteriorate in any statistically relevant way, then the dataset is kept, else the dataset is discarded, on the basis of the inconsistency with the remaining datasets.

Statistical distributions. Moreover, to check if the flag raised in the postfit analysis is statistically significant, we repeat the $\chi^{2(k)}$ and $n_\sigma^{(k)}$ computations across 1000 random instances (k) of the data generation in the closure-test fits (corresponding to independent runs of the Universe). We check it both globally and for each individual datasets, with a special emphasis on the datasets that are mostly affected by the BSM corrections associated to the given model. If the χ^2 and n_σ distributions are peaked above the critical thresholds of 1.5 and 2 respectively, then the weighted fit procedure is activated and the inconsistency is assessed according to the methodology outlined above.

BSM model. The model that we consider throughout this paper consists of the $SU(2)_L$ triplet field $W'^{a,\mu}$, where $a \in \{1,2,3\}$ denotes an $SU(2)_L$ index, with mass $M_{W'}$ and coupling coefficient by $g_{W'}$. The model is explained in details in Refs. [7,8] and is described by the Lagrangian

$$\begin{aligned}\mathcal{L}_{\rm UV}^{W'} = \mathcal{L}_{\rm SM} &- \frac{1}{4} W'^a_{\mu\nu} W'^{a,\mu\nu} + \frac{1}{2} M^2_{W'} W'^a_\mu W'^{a,\mu} \\ &- g_{W'} W'^{a,\mu} \sum_{f_L} \bar{f}_L T^a \gamma^\mu f_L \\ &- g_{W'} (W'^{a,\mu} \varphi^\dagger T^a i D_\mu \varphi + \text{H.c.}),\end{aligned} \tag{A2}$$

where the sum runs over the left-handed fermions: $f_L \in \{q, \ell\}$. The $SU(2)_L$ generators are given by $T^a = \frac{1}{2}\sigma^a$ where σ^a are the Pauli matrices. The kinetic term is given by $W'^a_{\mu\nu} = \partial_\mu W'^a_\nu - \partial_\nu W'^a_\mu - i g_{W'} [W'^a_\mu, W'^a_\nu]$. The covariant derivative is given by $D_\mu = \partial_\mu + \frac{1}{2} i g \sigma^a W^a_\mu + i g' Y_\varphi B_\mu$. The mixing with the SM gauge fields is neglected. The leading effect of this model on the dataset included in our analysis is to modify the Drell-Yan and deep inelastic

scattering observables, both charged current and neutral current. The tree-level matching of $\mathcal{L}_{\rm UV}^{W'}$ to the dimension-6 operators in the SMEFT is described in details in Ref. [8]. There it is shown that the impact at the LHC energy is dominated by the four-fermion interactions, which sum to

$$\mathcal{L}_{\rm SMEFT}^{W'} = \mathcal{L}_{\rm SM} - \frac{g_{W'}^2}{2M_{W'}^2} J_L^{a,\mu} J_{L,\mu}^a, \quad J_L^{a,\mu} = \sum_{f_L} \bar{f}_L T^a \gamma^\mu f_L, \tag{A3}$$

which is typically described by the $\hat{W}$ parameter

$$\mathcal{L}_{\rm SMEFT}^{W'} = \mathcal{L}_{\rm SM} - \frac{g^2 \hat{W}}{2m_W^2} J_L^{a,\mu} J_{L,\mu}^a, \qquad \hat{W} = \frac{g_{W'}^2}{g^2} \frac{m_W^2}{M_{W'}^2}. \tag{A4}$$

Using Fermi's constant, one can write the relation between the UV parameters and $\hat{W}$ as

$$\frac{g_{W'}^2}{M_{W'}^2} = 4\sqrt{2} G_F \hat{W}. \tag{A5}$$

By fixing $g_{W'} = 1$, each $M_{W'}$ can be associated to a value of $\hat{W}$.

Datasets included. We use all datasets included in the NNPDF4.0 global analysis [4]. The high-mass Drell-Yan measurements from ATLAS and CMS are impacted by the W' model we consider, we provide details on the data points in Table I. Several fixed-target Drell-Yan measurements from Fermilab with proton (p) and deuteron (d) targets probe the same processes at much lower energies. The effects of new physics are much more suppressed and they remain compatible with the SM predictions. More details is provided in Table II. On top of the existing measurements, we include the high-mass Drell-Yan projections that were introduced in [8], inspired by the HL-LHC projections studied in [78]. The invariant mass distribution projections are generated at 14 TeV assuming an integrated luminosity of 6 ab^{-1} (3 ab^{-1} collected by ATLAS and 3 ab^{-1} by CMS). Both in the case of NC and CC Drell-Yan cross sections, the MC data were generated using the MadGraph5 aMCatNLO NLO Monte Carlo event generator [92] with additional K-factors to include the NNLO QCD and NLO EW corrections. The MC data consist of four datasets (associated with NC/CC distributions with muons/electrons in the final state), each comprising 16 bins in the M_{ll} invariant mass distribution or transverse mass M_T distributions with both m_{ll} and M_T greater than 500 GeV, with the highest energy bins reaching $m_{ll} = 4$ TeV ($M_T = 3.5$ TeV) for NC (CC) data. The rationale behind the choice of number of bins and the width of each bin was outlined in Ref. [8], and stemmed from the requirement that the expected number of events per bin was big enough to ensure the applicability of Gaussian statistics. The choice of binning for the $M_{ll}(M_T)$ distribution at the HL-LHC is displayed in Fig. 5.1 of Ref. [8]. The details of the datasets included in the fit are presented in Table III.

As far as the EIC projections are concerned, we consider those that were built in Ref. [20], in particular the measurements of charged-current (CC) and neutral-current (NC) DIS processes, using both electrons and positrons as projectiles and both protons and neutrons as targets. Details about the projections used are presented in Table IV. We tested both the optimistic and pessimistic scenarios which vary in terms of how many data points can be collected, the former indicated outside and the latter indicated inside the brackets. We found that the choice of scenario does not impact our conclusions in any significant way.

TABLE I. Summary of the high-mass Drell-Yan data included in NNPDF4.0 [4].

Observable	$N_{\rm dat}$	$\sqrt{s}$ (TeV)
ATLAS		
$d\sigma^Z/dM_{\ell\ell}$	13	7
$d^2\sigma^Z/dM_{\ell\ell}d\lvert y_{\ell\ell}\rvert$	48	8
CMS		
$d\sigma^Z/dy_{\mu^+\mu^-}$	132	7
$d\sigma^Z/dy_{e^+e^-}$	132	7
$d\sigma^Z/dM_{\ell\ell}$	43	13

TABLE II. Summary of the fixed-target Drell-Yan data included in NNPDF4.0 [4].

Observable	$N_{\rm dat}$	$\sqrt{s}$ (GeV)
NuSea (E866)		
$d\sigma^{Z/\gamma^*}/dM_{ll}$ (d)	15	800
$d\sigma_{Z/\gamma^*}/dM_{ll}$ (p)	184	800
E605		
$d\sigma^{Z/\gamma^*}/dM_{ll}$	119	38.8
SeaQuest (E905)		
$d\sigma^{Z/\gamma^*}/dM_{ll}$ (d)	6	120

TABLE III. Summary of the HL-LHC projected data included in this work, taken from Ref. [8].

Observable	$N_{\rm dat}$	$\sqrt{s}$ (TeV)
Charged Current		
$d\sigma^{W^\pm}/dM_{T,e\nu}$	16	14
$d\sigma^{W^\pm}/dM_{T,\mu\nu}$	16	14
Neutral Current		
$d\sigma^Z/dM_{e^+e^-}$	12	14
$d\sigma^Z/dM_{\mu^+\mu^-}$	12	14

TABLE IV. Summary of the EIC projected data included in this work, taken from Ref. [20]. The number of data point for the optimistic and pessimistic scenarios are reported out and in the bracket respectively, with the pessimistic scenario being more conservative in terms of the number of data points.

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

For the FPF data, we consider the projected data built in Ref. [21]. The focus is on charged-current DIS processes, where the projectiles are neutrinos and anti-neutrinos, coming from the LHC beam, hitting isoscalar targets and the processes associated with charm production. For each of the far-forward LHC neutrino experiments the pseudorapidity coverage, target material, the acceptance for the charged lepton and hadronic final state, and the expected reconstruction performance are identified in [21] assuming that FASERν and SND acquire data for Run III with $\mathcal{L} = 150$ fb^{-1}. SND projections are too statistically insignificant to be included in the fit, due to the little statistics collected so far. As far as the FPF facility is concerned, FASERν2, AdvSND, and FLArE take data for the complete HL-LHC period, with $\mathcal{L} = 3$ ab^{-1}. The complete set of projected data that is included is summarized in Table V.

Maximal BSM threshold. To conclude, we consider how the maximal BSM threshold, i.e. the maximal BSM signal that can be absorbed by PDFs in a global fit, shifts as we include FPF and EIC projected data. Clearly, if we inject in the data the effect of a new heavy flavor-universal W', with mass larger than 13.8 TeV, the depletion in the high-energy tails measured at the HL-LHC will be less significant and thus the effect can still be fitted away in the PDFs. In this appendix we investigate by how much the EIC and CERN neutrino data would shift the maximal BSM threshold and whether the impact of the BSM bias—pushed at a higher threshold—would visibly impact the PDFs and thus possibly hamper the interpretation of indirect new physics searches. To answer this question, we made a scan over $M_{W'}$, by injecting increasingly large values of the W' mass, until the signal can be absorbed even in the presence of precise EIC and CERN neutrino data in a PDF global

TABLE V. Summary of the FPF projected data included in this work, taken from Ref. [21].

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

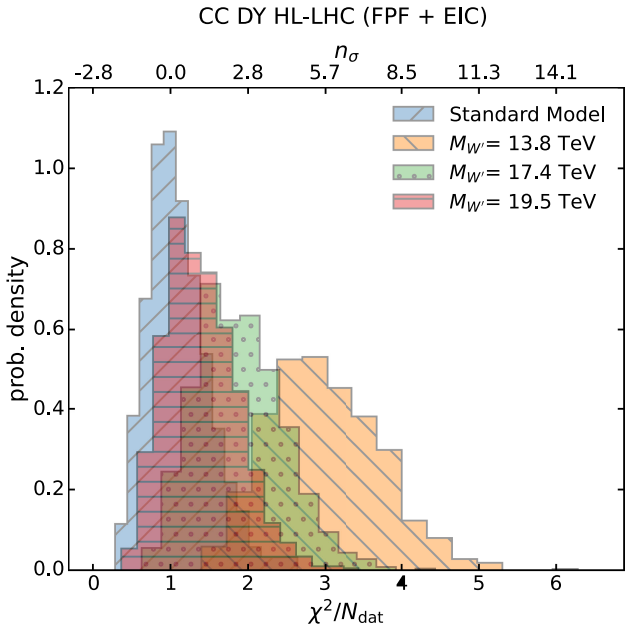

FIG. 5. Comparison of the distribution of the χ^2 per data point (bottom x-axis) and n_σ (top x-axis) for the high-mass HL-LHC DY CC projected data between the baseline fit (blue histogram) and fits with HL-LHC + EIC + FPF projections included by varying the strength of the BSM induced signal as $M_{W'}$ is increased from $M_{W'} = 13.8$ TeV (red histogram) to $M_{W'} = 17.4$ TeV (green histogram) to $M_{W'} = 19.5$ TeV (orange histogram).

TABLE VI. Impact of the inclusion of the EIC-only projected data (excluding the FPF projected data), both in optimistic (EIC opt.) and pessimistic (EIC pes.) scenarios provided in Ref. [20] on the fit quality, once the EIC projected data are included in a global fit alongside the BSM-contaminated HL Drell-Yan projected data, for two values of the heavy W', $M_{W'} = 13.8$ TeV and $M_{W'} = 15.9$ TeV. The difference between the optimistic and pessimistic scenarios mostly boils down to the number of data points, as it is detailed in Table IV.

$M_{W'}$ (TeV)	χ^2 (EIC opt.)	χ^2 (EIC pes.)
13.8	2.05	1.77
15.9	1.55	1.45

analysis. The summary results of the analysis are displayed in Fig. 5. There we see that, as we increase the value of $M_{W'}$ in our BSM model, thus decrease the size of the BSM corrections injected in the data, we have to get to $M_{W'} = 19.5$ TeV (orange histogram) for the signal to be absorbed by the PDFs.

In Table VI we show the χ^2 values of the HL-LHC charged current Drell-Yan projected data (who are contaminated by the injection of the BSM signal) once the EIC projected data are included in a global fit. We consider both the optimistic and pessimistic scenarios provided in Ref. [20]. For $M_{W'} = 13.8$ TeV, both cases the χ^2 values of the charged current Drell-Yan are above the threshold of 1.5, indicating that the fit is inconsistent. For a milder signal corresponding to $M_{W'} = 15.9$ TeV, the χ^2 values are very close to the threshold of 1.5 in both cases, indicating that the fit is consistent. The difference between the optimistic and pessimistic scenarios mostly boils down to the number of data points, as it is detailed in Table IV. These results correspond to a global fit in which we only included the EIC projections alongside the global datasets and the HL-LHC projections, and not the FPF ones. By assessing here the impact of the variation of the artificial data on our results, we observe small quantitative variations in the χ^2 value, as they tend to be slightly smaller in the pessimistic scenario compared to the optimistic one. This is completely expected as the pessimistic scenario has larger statistical uncertainties and thus do not constrain the PDF quite as tightly in the large-x region as the optimistic scenario does. However, there is no qualitative change. In the previous maximal mixing case (13.8 TeV), the pessimistic χ^2 is still sufficiently high to flag the fit as inconsistent. In the new maximal mixing case (15.9 TeV), both χ^2 are very close to 1.5, which is the threshold we chose for flagging the fit as inconsistent. Therefore, our results are stable under the variation of the EIC pseudodata. The small quantitative improvement in the optimistic case would suggest that if we will end up having even better statistics than the current projections indicate, we could hope to see a further improvement in our estimates.

Finally, we ask ourselves what is the effect of the new physics absorption now that the threshold has been pushed to a significantly larger value. Is this still going to spoil the interpretation of indirect BSM searches? To answer this question we assess the impact of the injection of our new physics model in the data on the fitted PDFs, by looking at the integrated luminosities for the parton pair i, j, which is defined as

$$\mathcal{L}_{ij}(m_X, \sqrt{s}) = \frac{1}{s}\int_{-y}^{y} d\tilde{y}\left[f_i\left(\frac{m_X}{\sqrt{s}}e^{\tilde{y}}, m_X\right) f_j\left(\frac{m_X}{\sqrt{s}}e^{-\tilde{y}}, m_X\right) + (i \leftrightarrow j)\right], \tag{A6}$$

where $f_i \equiv f_i(x, Q)$ is the PDF corresponding to the parton flavor i, and the integration limits are defined by

$$y = \ln\left(\frac{\sqrt{s}}{m_X}\right). \tag{A7}$$

In particular, we focus on the luminosities that are most constrained by the NC and CC Drell-Yan data respectively, namely

$$\mathcal{L}^{\mathrm{NC}}(m_X, \sqrt{s}) = \mathcal{L}_{u\bar{u}}(m_X, \sqrt{s}) + \mathcal{L}_{d\bar{d}}(m_X, \sqrt{s}), \tag{A8}$$

$$\mathcal{L}^{\mathrm{CC}}(m_X, \sqrt{s}) = \mathcal{L}_{u\bar{d}}(m_X, \sqrt{s}) + \mathcal{L}_{d\bar{u}}(m_X, \sqrt{s}). \tag{A9}$$

In Fig. 6 we show how the addition of the new projected data mitigates the BSM-induced PDF bias. We compare the NC and CC PDF luminosities defined in Eqs. (A8) and (A9) respectively as a function of m_X, which in the case of NC can be identified with the invariant mass of the produced leptons and in the case of CC can be identified with the transverse mass of the lepton-neutrino pair. We compare the luminosities obtained in a closure test in which the data have been generated by injecting a W' model with $M_{W'}$ set at different values and the data fitted assuming the SM. Each of the curves is compared to the *true* SM luminosities. The orange band represents the luminosities fitted on the whole NNPDF4.0 dataset plus the HL-LHC projected data displayed in Table III with a $M_{W'} = 13.8$ TeV signal injected in the data and the fit performed assuming the SM. Here the result is the same as the one presented in Ref. [5], namely the luminosities shift significantly as the PDFs have completely absorbed the BSM signal. On the other hand, if the EIC and the CERN forward neutrino projected data is included alongside the HL-LHC one, then the maximal BSM signal that can be absorbed corresponds to $M_{W'} = 19.5$ TeV. The corresponding PDF luminosities are displayed in blue. We observe a systematic reduction of the discrepancy with the true SM luminosities. In the case of the $u\bar{u} + d\bar{d}$ luminosity, the fit including EIC

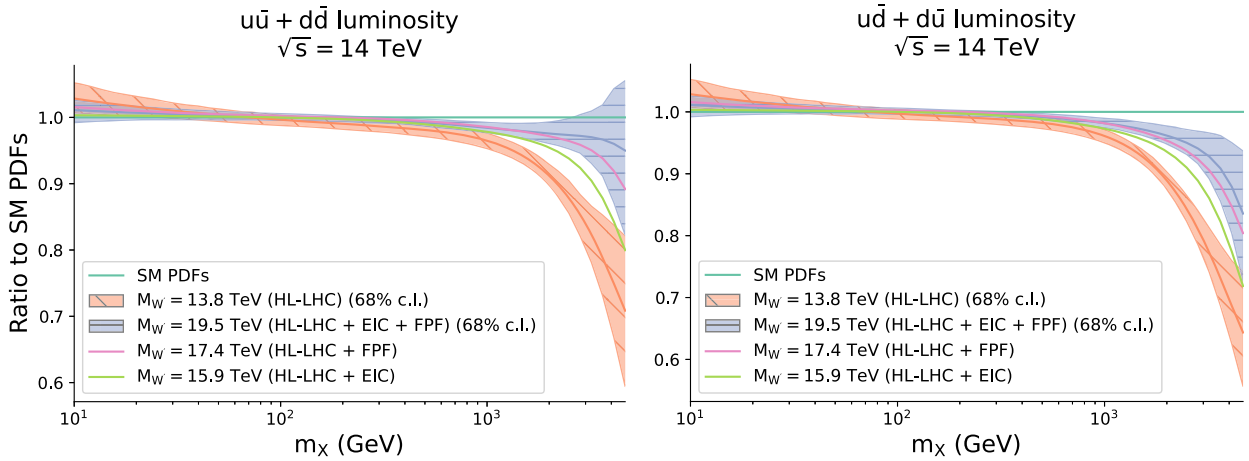

FIG. 6. Impact of the BSM-induced bias on the NC $u\bar{u} + d\bar{d}$ luminosity, Eq. (A8), (left) and CC $u\bar{d} + d\bar{u}$ luminosity, Eq. (A9), (right) for the different scenarios, all normalized to the *true* SM PDF luminosities. The orange bands correspond to the central value and PDF uncertainties of a fit including the global NNPDF4.0 datasets plus the HL-LHC projections considered here. The data have been generated according to a W' model with $M_{W'} = 13.8$ TeV and the fit is done assuming the SM. The blue bands correspond to the same fit in which also the EIC and the CERN forward neutrino projected data are included alongside the HL-LHC one, thus moving $M_{W'}$ to 19.5 TeV. The intermediate cases in which only EIC projections or only FPF projections are included are shown with only their central values in green and pink respectively, according to the maximal BSM signal that can be absorbed in each of these cases, namely $M_{W'} = 15.9$ TeV and $M_{W'} = 17.4$ TeV.

and CERN forward neutrino projections almost completely suppresses the discrepancy while for the $u\bar{d} + d\bar{u}$ a visible shift remains but its effect is greatly reduced. The intermediate case in which only EIC projections or only FPF projections are included are shown with only their central values in green and pink respectively, according to the maximal BSM signal that can be absorbed in each of these cases. To conclude, if the EIC and FPF projections are included in a global PDF analysis alongside the HL-LHC high-mass Drell-Yan distributions, the maximal BSM threshold, i.e., the maximal BSM signal that can be absorbed by PDFs in a global fit, shifts to larger masses or, equivalently, smaller deviations. As a result, while the effect can still be fitted away in the PDFs, the consequences are much milder and the BSM-biased PDFs are statistically undistinguishable from the *true* SM PDFs, thus effectively disentangling PDF effects from BSM effects.

[1] Julien Baglio, Claude Duhr, Bernhard Mistlberger, and Robert Szafron, Inclusive production cross sections at N^{3}LO, J. High Energy Phys. 12 (2022) 066.
[2] S. Amoroso *et al.*, Snowmass 2021 whitepaper: Proton structure at the precision frontier, Acta Phys. Pol. B **53**, 12 (2022).
[3] Maria Ubiali, Parton Distribution Functions and Their Impact on Precision of the Current Theory Calculations, arXiv:2404.08508.
[4] Richard D. Ball *et al.*, The path to proton structure at 1% accuracy, Eur. Phys. J. C **82**, 428 (2022).
[5] Elie Hammou, Zahari Kassabov, Maeve Madigan, Michelangelo L. Mangano, Luca Mantani, James Moore, Manuel Morales Alvarado, and Maria Ubiali, Hide and seek: How PDFs can conceal new physics, J. High Energy Phys. 11 (2023) 090.
[6] Mark N. Costantini, Elie Hammou, Zahari Kassabov, Maeve Madigan, Luca Mantani, Manuel Morales Alvarado, James M. Moore, and Maria Ubiali, SIMUnet: An open-source tool for simultaneous global fits of EFT Wilson coefficients and PDFs, Eur. Phys. J. C **84**, 805 (2024).
[7] Marco Farina, Giuliano Panico, Duccio Pappadopulo, Joshua T. Ruderman, Riccardo Torre, and Andrea Wulzer, Energy helps accuracy: Electroweak precision tests at hadron colliders, Phys. Lett. B **772**, 210 (2017).
[8] Admir Greljo, Shayan Iranipour, Zahari Kassabov, Maeve Madigan, James Moore, Juan Rojo, Maria Ubiali, and Cameron Voisey, Parton distributions in the SMEFT from high-energy Drell-Yan tails, J. High Energy Phys. 07 (2021) 122.
[9] R. Abdul Khalek *et al.*, Science requirements and detector concepts for the electron-ion collider: EIC yellow report, Nucl. Phys. **A1026**, 122447 (2022).
[10] R. Abdul Khalek *et al.*, Snowmass 2021 white paper: Electron ion collider for high energy physics, arXiv:2203.13199.

[11] Henso Abreu *et al.*, The FASER detector, J. Instrum. **19**, P05066 (2024).

[12] Henso Abreu *et al.*, First direct observation of collider neutrinos with FASER at the LHC, Phys. Rev. Lett. **131**, 031801 (2023).

[13] Roshan Mammen Abraham *et al.*, First Measurement of νe and $\nu\mu$ interaction cross sections at the LHC with FASER's emulsion detector, Phys. Rev. Lett. **133**, 021802 (2024).

[14] G. Acampora *et al.*, SND@LHC: The scattering and neutrino detector at the LHC, J. Instrum. **19**, P05067 (2024).

[15] R. Albanese *et al.*, Observation of collider muon neutrinos with the SND@LHC experiment, Phys. Rev. Lett. **131**, 031802 (2023).

[16] Luis A. Anchordoqui *et al.*, The forward physics facility: Sites, experiments, and physics potential, Phys. Rep., **968**, 1 (2022).

[17] Jonathan L. Feng *et al.*, The forward physics facility at the high-luminosity LHC, J. Phys. G **50**, 030501 (2023).

[18] A. Accardi *et al.*, Electron ion collider: The next QCD frontier: Understanding the glue that binds us all, Eur. Phys. J. A **52**, 268 (2016).

[19] E. C. Aschenauer, S. Fazio, J. H. Lee, H. Mantysaari, B. S. Page, B. Schenke, T. Ullrich, R. Venugopalan, and P. Zurita, The electron–ion collider: Assessing the energy dependence of key measurements, Rep. Prog. Phys. **82**, 024301 (2019).

[20] Rabah Abdul Khalek, Jacob J. Ethier, Emanuele R. Nocera, and Juan Rojo, Self-consistent determination of proton and nuclear PDFs at the electron ion collider, Phys. Rev. D **103**, 096005 (2021).

[21] Juan M. Cruz-Martinez, Max Fieg, Tommaso Giani, Peter Krack, Toni Mäkelä, Tanjona R. Rabemananjara, and Juan Rojo, The LHC as a neutrino-ion collider, Eur. Phys. J. C **84**, 369 (2024).

[22] Radja Boughezal, Frank Petriello, and Daniel Wiegand, Removing flat directions in standard model EFT fits: How polarized electron-ion collider data can complement the LHC, Phys. Rev. D **101**, 116002 (2020).

[23] Chiara Bissolotti, Radja Boughezal, and Kaan Simsek, SMEFT analysis with LHeC, FCC-eh, and EIC DIS pseudodata, arXiv:2307.09459.

[24] M. Aleksa *et al.*, Conceptual design of an experiment at the FCC-hh, a future 100 TeV hadron collider, 10.23731/CYRM-2022-002 (2022).

[25] Richard D. Ball, Alessandro Candido, Stefano Forte, Felix Hekhorn, Emanuele R. Nocera, Juan Rojo, and Christopher Schwan, Parton distributions and new physics searches: The Drell–Yan forward–backward asymmetry as a case study, Eur. Phys. J. C **82**, 1160 (2022).

[26] Andrii Anataichuk *et al.*, Exploring SMEFT couplings using the forward-backward asymmetry in neutral current Drell-Yan production at the LHC, Eur. Phys. J. C **84**, 1277 (2024).

[27] Juri Fiaschi, Francesco Giuli, Francesco Hautmann, and Stefano Moretti, Enhancing the Large Hadron Collider sensitivity to charged and neutral broad resonances of new gauge sectors, J. High Energy Phys. 02 (2022) 179.

[28] F. D. Aaron *et al.*, Combined measurement and QCD analysis of the inclusive $e^{\pm}p$ scattering cross sections at HERA, J. High Energy Phys. 01 (2010) 109.

[29] H. Abramowicz *et al.*, Combination and QCD analysis of charm production cross section measurements in deep-inelastic ep scattering at HERA, Eur. Phys. J. C **73**, 2311 (2013).

[30] H. Abramowicz *et al.*, Combination of measurements of inclusive deep inelastic $e^{\pm}p$ scattering cross sections and QCD analysis of HERA data, Eur. Phys. J. C **75**, 580 (2015).

[31] M. Arneodo *et al.*, Measurement of the proton and deuteron structure functions, F_2^p and F_2^d, and of the ratio σ_L/σ_T, Nucl. Phys. **B483**, 3 (1997).

[32] L. W. Whitlow, E. M. Riordan, S. Dasu, Stephen Rock, and A. Bodek, Precise measurements of the proton and deuteron structure functions from a global analysis of the SLAC deep inelastic electron scattering cross-sections, Phys. Lett. B **282**, 475 (1992).

[33] A. C. Benvenuti *et al.*, A high statistics measurement of the proton structure functions $F_2(x, Q^2)$ and R from deep inelastic muon scattering at high Q^2, Phys. Lett. B **223**, 485 (1989).

[34] G. Onengut *et al.*, Measurement of nucleon structure functions in neutrino scattering, Phys. Lett. B **632**, 65 (2006).

[35] M. Goncharov *et al.*, Precise measurement of dimuon production cross-sections in ν_μFe and $\bar{\nu}_\mu$Fe deep inelastic scattering at the Tevatron, Phys. Rev. D **64**, 112006 (2001).

[36] V. M. Abazov *et al.*, Measurement of the electron charge asymmetry in $p\bar{p} \to W + X \to e\nu + X$ events at $\sqrt{s} = 1.96$-TeV, Phys. Rev. Lett. **101**, 211801 (2008).

[37] T. Aaltonen *et al.*, Combined CDF and D0 upper limits on standard model higgs boson production with up to 8.2 fb^{-1} of data, arXiv:1107.5518.

[38] V. M. Abazov *et al.*, Measurement of the inclusive jet cross-section in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$-TeV, Phys. Rev. Lett. **101**, 062001 (2008).

[39] V. M. Abazov *et al.*, Measurement of the ratio of the $p\bar{p} \to W + c$-jet cross section to the inclusive $p\bar{p} \to W +$ jets cross section, Phys. Lett. B **666**, 23 (2008).

[40] Victor Mukhamedovich Abazov *et al.*, Measurement of the muon charge asymmetry in $p\bar{p} \to \mathrm{W} + \mathrm{X} \to \mu\nu + \mathrm{X}$ events at $\sqrt{s} = 1.96$ TeV, Phys. Rev. D **88**, 091102 (2013).

[41] Victor Mukhamedovich Abazov *et al.*, Measurement of the electron charge asymmetry in $\boldsymbol{p\bar{p} \to W + X \to e\nu + X}$ decays in $\boldsymbol{p\bar{p}}$ collisions at $\boldsymbol{\sqrt{s} = 1.96}$ TeV, Phys. Rev. D **91**, 032007 (2015); **91**, 079901(E) (2015).

[42] Georges Aad *et al.*, Measurement of the inclusive $W^{\pm}$ and Z/γ^* cross sections in the electron and muon decay channels in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector, Phys. Rev. D **85**, 072004 (2012).

[43] Morad Aaboud *et al.*, Precision measurement and interpretation of inclusive W^+, W^- and Z/γ^* production cross sections with the ATLAS detector, Eur. Phys. J. C **77**, 367 (2017).

[44] Georges Aad *et al.*, Measurement of the low-mass Drell-Yan differential cross section at $\sqrt{s} = 7$ TeV using the ATLAS detector, J. High Energy Phys. 06 (2014) 112.

[45] Georges Aad *et al.*, Measurement of the high-mass Drell–Yan differential cross-section in pp collisions at $\sqrt{s} =$ 7 TeV with the ATLAS detector, Phys. Lett. B **725**, 223 (2013).

[46] Serguei Chatrchyan *et al.*, Measurement of the electron charge asymmetry in inclusive W production in

pp collisions at $\sqrt{s} = 7$ TeV, Phys. Rev. Lett. **109**, 111806 (2012).
[47] Serguei Chatrchyan *et al.*, Measurement of the muon charge asymmetry in inclusive pp to WX production at $\sqrt{s} =$ 7 TeV and an improved determination of light parton distribution functions, Phys. Rev. D **90**, 032004 (2014).
[48] Sergei Chatrchyan *et al.*, Measurement of the differential and double-differential Drell-Yan cross sections in proton-proton collisions at $\sqrt{s} = 7$ TeV, J. High Energy Phys. 12 (2013) 030.
[49] Vardan Khachatryan *et al.*, Measurement of the differential cross section and charge asymmetry for inclusive pp $\to$ $W^{\pm} + X$ production at $\sqrt{s} = 8$ TeV, Eur. Phys. J. C **76**, 469 (2016).
[50] R Aaij *et al.*, Measurement of the cross-section for $Z \to e^+e^-$ production in pp collisions at $\sqrt{s} = 7$ TeV, J. High Energy Phys. 02 (2013) 106.
[51] Roel Aaij *et al.*, Measurement of the forward Z boson production cross-section in pp collisions at $\sqrt{s} = 7$ TeV, J. High Energy Phys. 08 (2015) 039.
[52] Roel Aaij *et al.*, Measurement of forward $Z \to e^+e^-$ production at $\sqrt{s} = 8$ TeV, J. High Energy Phys. 05 (2015) 109.
[53] Roel Aaij *et al.*, Measurement of forward W and Z boson production in pp collisions at $\sqrt{s} = 8$ TeV, J. High Energy Phys. 01 (2016) 155.
[54] Georges Aad *et al.*, Measurement of the double-differential high-mass Drell-Yan cross section in pp collisions at $\sqrt{s} =$ 8 TeV with the ATLAS detector, J. High Energy Phys. 08 (2016) 009.
[55] M. Aaboud *et al.*, Measurement of the Drell-Yan triple-differential cross section in pp collisions at $\sqrt{s} = 8$ TeV, J. High Energy Phys. 12 (2017) 059.
[56] Roel Aaij *et al.*, Measurement of forward $W \to e\nu$ production in pp collisions at $\sqrt{s} = 8$ TeV, J. High Energy Phys. 10 (2016) 030.
[57] Georges Aad *et al.*, Measurement of $W^{\pm}$ and Z-boson production cross sections in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector, Phys. Lett. B **759**, 601 (2016).
[58] Roel Aaij *et al.*, Measurement of the forward Z boson production cross-section in pp collisions at $\sqrt{s} = 13$ TeV, J. High Energy Phys. 09 (2016) 136.
[59] Georges Aad *et al.*, Measurement of the transverse momentum and ϕ^*_η distributions of Drell-Yan lepton pairs in proton-proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector, Eur. Phys. J. C **76**, 291 (2016).
[60] Vardan Khachatryan *et al.*, Measurement of the Z boson differential cross section in transverse momentum and rapidity in proton-proton collisions at 8 TeV, Phys. Lett. B **749**, 187 (2015).
[61] Morad Aaboud *et al.*, Measurement of differential cross sections and W^+/W^- cross-section ratios for W boson production in association with jets at $\sqrt{s} = 8$ TeV with the ATLAS detector, J. High Energy Phys. 05 (2018) 077; 10 (2020) 48.
[62] Albert M Sirunyan *et al.*, Measurement of the differential cross sections for the associated production of a W boson and jets in proton-proton collisions at $\sqrt{s} = 13$ TeV, Phys. Rev. D **96**, 072005 (2017).
[63] Georges Aad *et al.*, Measurement of inclusive jet and dijet production in pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector, Phys. Rev. D **86**, 014022 (2012).
[64] Georges Aad *et al.*, Measurement of the inclusive jet cross section in pp collisions at $\sqrt{s} = 2.76$ TeV and comparison to the inclusive jet cross section at $\sqrt{s} = 7$ TeV using the ATLAS detector, Eur. Phys. J. C **73**, 2509 (2013).
[65] Georges Aad *et al.*, Measurement of the inclusive jet cross-section in proton-proton collisions at $\sqrt{s} = 7$ TeV using 4.5 fb^{-1} of data with the ATLAS detector, J. High Energy Phys. 02 (2015) 153; 09 (2015) 141(E).
[66] Serguei Chatrchyan *et al.*, Measurements of differential jet cross sections in proton-proton collisions at $\sqrt{s} = 7$ TeV with the CMS detector, Phys. Rev. D **87**, 112002 (2013).
[67] Vardan Khachatryan *et al.*, Measurement of the inclusive jet cross section in pp collisions at $\sqrt{s} = 2.76$ TeV, Eur. Phys. J. C **76**, 265 (2016).
[68] Morad Aaboud *et al.*, Measurement of the inclusive jet cross-sections in proton-proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector, J. High Energy Phys. 09 (2017) 020.
[69] Vardan Khachatryan *et al.*, Measurement and QCD analysis of double-differential inclusive jet cross sections in pp collisions at $\sqrt{s} = 8$ TeV and cross section ratios to 2.76 and 7 TeV, J. High Energy Phys. 03 (2017) 156.
[70] Georges Aad *et al.*, Measurement of the inclusive isolated prompt photon cross section in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector, J. High Energy Phys. 08 (2016) 005.
[71] Morad Aaboud *et al.*, Measurement of the cross section for inclusive isolated-photon production in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector, Phys. Lett. B **770**, 473 (2017).
[72] M. R. Adams *et al.*, Proton and deuteron structure functions in muon scattering at 470-GeV, Phys. Rev. D **54**, 3006 (1996).
[73] R. S. Towell *et al.*, Improved measurement of the anti-d/anti-u asymmetry in the nucleon sea, Phys. Rev. D **64**, 052002 (2001).
[74] R. S. Towell *et al.*, Improved measurement of the anti-d/anti-u asymmetry in the nucleon sea, Phys. Rev. D **64**, 052002 (2001).
[75] Georges Aad *et al.*, Measurement of the double-differential high-mass Drell-Yan cross section in pp collisions at $\sqrt{s} =$ 8 TeV with the ATLAS detector, J. High Energy Phys. 08 (2016) 009.
[76] Vardan Khachatryan *et al.*, Measurements of differential and double-differential Drell-Yan cross sections in proton-proton collisions at 8 TeV, Eur. Phys. J. C **75**, 147 (2015).
[77] Albert M Sirunyan *et al.*, Measurement of the differential Drell-Yan cross section in proton-proton collisions at $\sqrt{s} = 13$ TeV, J. High Energy Phys. 12 (2019) 059.
[78] Rabah Abdul Khalek, Shaun Bailey, Jun Gao, Lucian Harland-Lang, and Juan Rojo, Towards ultimate parton distributions at the high-luminosity LHC, Eur. Phys. J. C **78**, 962 (2018).
[79] Shayan Iranipour and Maria Ubiali, A new generation of simultaneous fits to LHC data using deep learning, J. High Energy Phys. 05 (2022) 032.

[80] Alessandro Candido, Felix Hekhorn, Giacomo Magni, Tanjona R. Rabemananjara, and Roy Stegeman, Yadism: Yet another deep-inelastic scattering module, Eur. Phys. J. C **84**, 698 (2024).
[81] Alessandro Candido, Felix Hekhorn, and Giacomo Magni, EKO: Evolution kernel operators, Eur. Phys. J. C **82**, 976 (2022).
[82] S. Carrazza, E. R. Nocera, C. Schwan, and M. Zaro, PineAPPL: Combining EW and QCD corrections for fast evaluation of LHC processes, J. High Energy Phys. 12 (2020) 108.
[83] Andrea Barontini, Alessandro Candido, Juan M. Cruz-Martinez, Felix Hekhorn, and Christopher Schwan, Pineline: Industrialization of high-energy theory predictions, Comput. Phys. Commun. **297**, 109061 (2024).
[84] Chiara Bissolotti, Radja Boughezal, and Kaan Simsek, SMEFT probes in future precision DIS experiments, Phys. Rev. D **108**, 075007 (2023).
[85] Tyler Corbett, Top-down and bottom-up: Studying the SMEFT beyond leading order in $1/\Lambda^2$, SciPost Phys. Core **7**, 053 (2024).
[86] Admir Greljo, Ajdin Palavrić, and Aleks Smolkovič, Leading directions in the SMEFT: Renormalization effects, Phys. Rev. D **109**, 075033 (2024).
[87] S. Bailey, T. Cridge, L. A. Harland-Lang, A. D. Martin, and R. S. Thorne, Parton distributions from LHC, HERA, Tevatron and fixed target data: MSHT20 PDFs, Eur. Phys. J. C **81**, 341 (2021).
[88] Tie-Jiun Hou *et al.*, New CTEQ global analysis of quantum chromodynamics with high-precision data from the LHC, Phys. Rev. D **103**, 014013 (2021).
[89] L. A. Harland-Lang, T. Cridge, and R. S. Thorne, A Stress Test of Global PDF Fits: Closure Testing the MSHT PDFs and a First Direct Comparison to the Neural Net Approach, Eur. Phys. J. C **85**, 316 (2025).
[90] Richard D. Ball *et al.*, Parton distributions for the LHC Run II, J. High Energy Phys. 04 (2015) 040.
[91] Luigi Del Debbio, Tommaso Giani, and Michael Wilson, Bayesian approach to inverse problems: an application to NNPDF closure testing, Eur. Phys. J. C **82**, 330 (2022).
[92] 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, J. High Energy Phys. 07 (2014) 079.