Parton distributions in the SMEFT from high-energy Drell-Yan tails

Admir Greljo,[a,f] **Shayan Iranipour,**[b] **Zahari Kassabov,**[b] **Maeve Madigan,**[b]
James Moore,[b] **Juan Rojo,**[d,e] **Maria Ubiali**[b] **and Cameron Voisey**[c]

[a] *Albert Einstein Center for Fundamental Physics,*
 Institut für Theoretische Physik, Universität Bern,
 Sidlerstrasse 5, CH-3012 Bern, Switzerland

[b] *Department of Applied Mathematics and Theoretical Physics, University of Cambridge,*
 Wilberforce Road, CB3 0WA Cambridge, U.K.

[c] *Department of Physics, High Energy Physics group, University of Cambridge,*
 JJ Thomson Avenue, CB3 0HE Cambridge, U.K.

[d] *Department of Physics and Astronomy, Vrije Universiteit Amsterdam,*
 NL-1081 HV Amsterdam, The Netherlands

[e] *Theory Group, Nikhef,*
 Science Park 105, 1098 XG Amsterdam, The Netherlands

[f] *Theoretical Physics Department, CERN,*
 CH-1211 Geneva 23, Switzerland

 E-mail: admir.greljo@unibe.ch, s.iranipour@damtp.cam.ac.uk,
 zk261@cam.ac.uk, mum20@cam.ac.uk, jmm232@cam.ac.uk, j.rojo@vu.nl,
 M.Ubiali@damtp.cam.ac.uk, voisey@hep.phy.cam.ac.uk

Abstract: The high-energy tails of charged- and neutral-current Drell-Yan processes provide important constraints on the light quark and anti-quark parton distribution functions (PDFs) in the large-x region. At the same time, short-distance new physics effects such as those encoded by the Standard Model Effective Field Theory (SMEFT) would induce smooth distortions to the same high-energy Drell-Yan tails. In this work, we assess for the first time the interplay between PDFs and EFT effects for high-mass Drell-Yan processes at the LHC and quantify the impact that the consistent joint determination of PDFs and Wilson coefficients has on the bounds derived for the latter. We consider two well-motivated new physics scenarios: 1) electroweak oblique corrections $(\hat{W}, \hat{Y})$ and 2) four-fermion interactions potentially related to the LHCb anomalies in $R(K^{(*)})$. We account for available Drell-Yan data, both from unfolded cross sections and from searches, and carry out dedicated projections for the High-Luminosity LHC. Our main finding is that, while the interplay between PDFs and EFT effects remains moderate for the current dataset, it will become a significant challenge for EFT analyses at the HL-LHC.

Contents

1 Introduction

The Large Hadron Collider (LHC) at CERN is currently leading the exploration of the high-energy frontier. Thanks to a wealth of precision experimental measurements, combined with the staggering progress in precise theoretical calculations and simulations, the Standard Model (SM) of particle physics is being stress-tested at high energies to an unprecedented level. The hunt for deviations from the SM predictions, which could signal the

long-sought evidence for beyond-the-SM (BSM) dynamics, is being pursued at the LHC with two complementary strategies. On the one hand, by means of direct searches for new particles that might be light enough to be produced on shell. On the other hand, with indirect searches aiming to identify a consistent pattern of deviations in the properties of known particles, such as their coupling strengths, which would be induced by new particles and interactions whose characteristic energy scale lies above the direct reach of the LHC.

In the case of direct searches, the target is identifying an abrupt deviation from the SM, for example in terms of a resonance peak on top of a smoothly falling background, or with a systematic mismatch in the number of measured events as compared to the SM expectations for non-resonant searches. Indirect searches instead often aim at pinning down subtle distortions with respect to the SM predictions, such as a few-percent variation in the value of some production cross sections or decay rates, induced by yet-unknown particles and interactions. A powerful framework to identify, parametrise, and correlate in a model-independent manner the possible deviations from the SM predictions that arise in such indirect searches is provided by the effective field theory (EFT) framework, such as the Standard Model EFT (SMEFT) (e.g. [1–14]) or the Higgs EFT (HEFT) (e.g. [15–22]).

It is a common misconception that indirect BSM investigations based on EFTs are relevant only for low-energy phenomena, while instead direct searches dominate the sensitivity for high-p_T observables. On the contrary: accurate measurements of observables at high energies offer one of the most promising avenues towards a discovery of BSM physics at the LHC. While the collider has now (almost) reached the design energy, its integrated luminosity continues to grow steadily, thus greatly facilitating dedicated studies of the (currently statistics-limited) high-energy tails of distributions. With this motivation, several high-p_T LHC cross sections have been studied as indirect probes of new physics in the EFT framework, from the (di)lepton distributions in neutral- and charged-current Drell-Yan (DY) distributions [23–48] (the main subject of this work) to inclusive jets and dijets [49–52], top quark pair distributions [53–61], gauge boson pair production [61–67], and vector boson scattering [67–70], among several others.

Furthermore, many higher-dimensional EFT operators induce deviations from the SM which grow with the energy of the partonic collision, enhancing the BSM sensitivity of these high-energy tails. For instance, in the Drell-Yan process, the naïve scaling of the $\psi\psi \to \psi\psi$ scattering amplitude in an underlying EFT description leads to $\mathcal{A} \propto E^2/\Lambda^2$, where E is the energy of the process and Λ is the new physics scale. Thus, rather generically, a less precise measurement of a high-mass tail can compete with a low-energy precision measurement due to this energy enhancement, which can be traced back to the preservation of unitarity. This property leads to the somehow unintuitive result that the study of the high-energy lepton tails in the Drell-Yan process represents a competitive probe of BSM dynamics compared to electroweak precision tests and low-energy flavour physics test.

In order for the EFT interpretation of these high-p_T cross sections to convincingly uncover a BSM effect, it becomes crucial to ensure full control over the SM inputs and their uncertainties, such as those associated to the hard-scattering (partonic) cross sections or the parton distribution functions (PDFs) of the nucleon [71]. In this context, one of the key challenges hampering the applicability of the EFT programme to the LHC high-energy tails

is indeed related to the treatment of the PDFs. These are determined from experimental data assuming the validity of the SM, often using exactly the same processes that enter EFT analyses. This observation naturally prompts the question of how can one make sure that eventual BSM deviations arising in high-p_T tails are not being inadvertently reabsorbed into the PDFs.

A first take on this challenge was presented in a proof-of-concept study [72] where the simultaneous determination of PDFs and EFT coefficients from deep inelastic scattering (DIS) structure functions was demonstrated. There, it was found that the EFT corrections can indeed be partially reabsorbed into the PDFs but also that it is possible to robustly disentangle QCD and BSM effects by exploiting their different energy scaling. The main goal of the present work is to extend this approach to LHC processes, specifically with the joint determination of PDFs and EFT coefficients from DIS and Drell-Yan data. Drell-Yan processes in general, and high-mass measurements in particular, provide information on the light quark and anti-quark PDFs in a broad region of x representing an important ingredient in modern global PDF fits [73–76]. Furthermore, high-mass Drell-Yan data will be instrumental at the High-Luminosity LHC (HL-LHC) to pin down the large-x PDFs [77]. Considering that EFT signals can lead to significant deviations from the SM in these same high-energy DY tails, one would like to assess to what extent they can be reabsorbed into the PDFs and to define strategies to separate QCD from BSM effects.

In this work, in order to interpret the Drell-Yan data in the EFT framework we formulate simple, yet motivated, BSM benchmark scenarios, which are chosen to represent a wide class of UV-complete theories. In the first scenario [32], we consider the $\hat{W}$ and $\hat{Y}$ electroweak parameters generated in universal theories that modify the electroweak gauge boson propagators and lead to flavour-universal deviations which grow with the invariant mass. In the second benchmark [27], we consider a flavour-specific scenario motivated by the evidence of lepton flavour universality (LFU) violation in B-meson decays recently reported by the LHCb collaboration [78–81]. Our analysis accounts for available unfolded high-mass Drell-Yan cross section data, detector-level searches based on the full Run II luminosity, and dedicated HL-LHC projections. We do not consider other data beyond the DIS and DY processes, in order to ensure a theoretically consistent description of the parton distributions in the SMEFT. The price to pay for this reduced dataset is the information loss on some flavour combinations, specifically on the gluon PDF.

The structure of this paper is as follows. First of all, in section 2 we discuss the EFT benchmark scenarios that will be used in the fits. Then in section 3 we summarise the datasets used in our analysis and the corresponding theoretical calculations, both in the SM and in the SMEFT, and we discuss their impact on a PDF fit. We also describe the methodology we use to simultaneously fit PDFs and SMEFT coefficients, and how we deal with several sources of uncertainties in the fits. In section 4 we present the results for the simultaneous determination of the SMEFT coefficients and the PDFs from the available high-mass DY data from LHC Run I and Run II, in the two scenarios presented in section 2 and assess how they modify the interpretation of BSM searches based on the SM PDFs. In section 5 we present a summary of the constraints we find on the two scenarios we consider and we assess the outcome of a joint PDF and EFT analysis at the HL-LHC. Finally, we will outline our main conclusions and possible future developments in section 6.

More technical discussions are collected in the appendices, and include detailed comparisons of the SM PDF fits produced in this work with previous NNPDF global fits (appendix A), the quantitative assessment of the fit quality to the various input datasets (appendix B), a benchmarking study for the calculation of EFT cross sections (appendix C), and a study of the flavour dependence of the SMEFT PDFs (appendix D).

2 SMEFT benchmark scenarios

In this section we present the two SMEFT benchmark scenarios that will be used in this work to interpret the LHC Drell-Yan processes. The first scenario belongs to the class of electroweak precision tests and is sensitive to a broad range of UV-complete theories proposed in the literature. The second benchmark represents a consistency check of the existing hints of lepton universality violation in rare B-meson decays reported by the LHCb collaboration. Both scenarios highlight the interplay between the PDFs and the EFT dynamics, illustrating in particular how the former changes and how constraints to the latter are modified.

2.1 Benchmark I: oblique corrections $\hat{W}$ and $\hat{Y}$

The oblique corrections, as originally proposed in [82, 83], play a key role in testing theories beyond the Standard Model. They parametrise the self-energy $\Pi_V(q^2)$ of the electroweak gauge bosons W_μ^a and B_μ, where $V = W^3 W^3$, BB, $W^3 B$, and $W^+ W^-$. Truncating the momentum expansion at order q^4, while imposing proper normalisation and symmetry constraints, one concludes that there are only four oblique parameters which can be identified with dimension-six operators in the SMEFT. These are the well-known $\hat{S}$, $\hat{T}$, $\hat{W}$, and $\hat{Y}$ parameters [84]. The parameters $\hat{S}$ and $\hat{T}$ are well constrained from precision LEP measurements [84] and grow slowly with q^2, while $\hat{W}$ and $\hat{Y}$ scale faster implying that their effects will be enhanced for the high-energy dilepton tails at the LHC [32]. While $\hat{T} = \mathcal{O}(q^0)$ and $\hat{S} = \mathcal{O}(q^2)$, instead one has that $\hat{W}, \hat{Y} = \mathcal{O}(q^4)$. In the universal basis (see e.g. [12]), the $\hat{W}$ and $\hat{Y}$ parameters are the Wilson coefficients associated to the following two operators:

$$\mathcal{L}_{\text{SMEFT}} \supset -\frac{\hat{W}}{4m_W^2}(D_\rho W_{\mu\nu}^a)^2 - \frac{\hat{Y}}{4m_W^2}(\partial_\rho B_{\mu\nu})^2 , \tag{2.1}$$

where m_W indicates the W-boson mass, and D_ρ is the covariant derivative. The physical effects of the operators in eq. (2.1) on the Drell-Yan process arise from the difference in the propagators through the self-energy modifications, see eq. (1) of ref. [32].

Alternatively, using the equations of motion, the same operators can be rotated to a basis in which the modifications to the Drell-Yan cross sections are instead captured by four-fermion contact interactions,

$$\mathcal{L}_{\text{SMEFT}} \supset -\frac{g^2 \hat{W}}{2m_W^2} J_{L\mu}^a J_L^{a\mu} - \frac{g_Y^2 \hat{Y}}{2m_W^2} J_{Y\mu} J_Y^\mu , \tag{2.2}$$

where J_L and J_Y are $\text{SU}(2)_L$ and $\text{U}(1)_Y$ conserved fermionic currents,

$$J_L^{a\mu} = \frac{1}{2} \sum_{f=q,l} \bar{f} \sigma^a \gamma^\mu f , \qquad J_Y^\mu = \sum_{f=q,l,u,d,e} Y_f \bar{f} \gamma^\mu f . \tag{2.3}$$

Here q, l are the SM quark and lepton left-handed doublets, while u, d, e are the right-handed singlets. Also, g and g_Y are the corresponding electroweak gauge couplings, while σ^a are the Pauli matrices, and the hypercharge $Y_f = 1/6$, $-1/2$, $2/3$, $-1/3$, and -1 for q, l, u, d, e, respectively. Summation over flavour indices is assumed, which implies that in this scenario the fermionic currents respect the $U(3)^5$ global flavour symmetry.

Expanding eq. (2.2), one can relate the $\hat{W}$ and $\hat{Y}$ parameters to the coefficients of dimension-six operators in the Warsaw basis [5]. There, the operators relevant to the description of the Drell-Yan process are given by

$$
\begin{aligned}
\mathcal{O}_{ld} &= (\bar{l}\gamma_\mu l)(\bar{d}\gamma^\mu d)\,, & \mathcal{O}_{lu} &= (\bar{l}\gamma_\mu l)(\bar{u}\gamma^\mu u)\,, & \mathcal{O}_{lq}^{(1)} &= (\bar{l}\gamma^\mu l)(\bar{q}\gamma_\mu q)\,, \\
\mathcal{O}_{ed} &= (\bar{e}\gamma_\mu e)(\bar{d}\gamma^\mu d)\,, & \mathcal{O}_{eu} &= (\bar{e}\gamma_\mu e)(\bar{u}\gamma^\mu u)\,, & \mathcal{O}_{qe} &= (\bar{q}\gamma_\mu q)(\bar{e}\gamma^\mu e)\,, \\
\mathcal{O}_{lq}^{(3)} &= (\bar{l}\sigma^a \gamma^\mu l)(\bar{q}\sigma^a \gamma_\mu q)\,.
\end{aligned}
\tag{2.4}
$$

Again, the flavour indices are contracted within the brackets, for example $(\bar{l}\gamma_\mu l) \equiv (\bar{l}^1\gamma_\mu l^1 + \bar{l}^2\gamma_\mu l^2 + \bar{l}^3\gamma_\mu l^3)$. Taking into account this matching between the $\hat{W}$ and $\hat{Y}$ parameters and the corresponding Wilson coefficients in the Warsaw basis, we can express the SMEFT Lagrangian in this scenario, eq. (2.2), as follows

$$
\begin{aligned}
\mathcal{L}_{\text{SMEFT}} = \mathcal{L}_{\text{SM}} - \frac{g^2 \hat{W}}{4m_W^2}\mathcal{O}_{lq}^{(3)} - \frac{g_Y^2 \hat{Y}}{m_W^2}\Big(&Y_l Y_d\, \mathcal{O}_{ld} + Y_l Y_u\, \mathcal{O}_{lu} \\
&+ Y_l Y_q\, \mathcal{O}_{lq}^{(1)} + Y_e Y_d\, \mathcal{O}_{ed} + Y_e Y_u\, \mathcal{O}_{eu} + Y_e Y_q\, \mathcal{O}_{qe}\Big)\,.
\end{aligned}
\tag{2.5}
$$

The parametrisation in eq. (2.4) has been implemented using the `SMEFTsim` package [85] and cross-checked against the reweighting method used in ref. [27] (see also [37]), as will be discussed in section 3.2 and in appendix C.

The analysis in the ref. [32] reports the following 95% confidence level intervals on $\hat{W}$ assuming $\hat{Y} = 0$,

$$
\begin{aligned}
\hat{W} &\in [-3, 15] \times 10^{-4}\ (\text{ATLAS 8 TeV, } 20.3\,\text{fb}^{-1}\ [86])\,, \\
\hat{W} &\in [-5, 22] \times 10^{-4}\ (\text{CMS 8 TeV, } 19.7\,\text{fb}^{-1}\ [87])\,,
\end{aligned}
\tag{2.6}
$$

as well as, the 95% confidence level intervals for $\hat{Y}$ assuming $\hat{W} = 0$,

$$
\begin{aligned}
\hat{Y} &\in [-4, 24] \times 10^{-4}\ (\text{ATLAS 8 TeV, } 20.3\,\text{fb}^{-1}\ [86])\,, \\
\hat{Y} &\in [-7, 41] \times 10^{-4}\ (\text{CMS 8 TeV, } 19.7\,\text{fb}^{-1}\ [87])\,.
\end{aligned}
\tag{2.7}
$$

These bound have been computed by assuming SM PDFs. In our analysis, for this benchmark scenario, we see how the limits based on SM PDFs are modified once a consistent determination of the SMEFT PDFs is done, requiring a simultaneous fit of the PDFs together with the $\hat{W}$ and $\hat{Y}$ parameters from the high-mass Drell-Yan distributions.

2.2 Benchmark II: left-handed muon-philic lepton-quark interactions

Following ref. [27], here we consider gauge invariant four-fermion operators built from the SM quark and lepton $SU(2)_L$ doublets. In the Warsaw basis, eq. (2.4), these correspond

to the $\mathcal{O}_{lq}^{(3)}$ and $\mathcal{O}_{lq}^{(1)}$ operators. Expanding the $\mathrm{SU}(2)_L$ indices, we find that the SMEFT Lagrangian contains operators of the form

$$\mathcal{L}_{\mathrm{SMEFT}} \supset \frac{\mathbf{C}_{ij}^{U\mu}}{v^2}(\bar{u}_L^i \gamma_\mu u_L^j)(\bar{\mu}_L \gamma^\mu \mu_L) + \frac{\mathbf{C}_{ij}^{D\mu}}{v^2}(\bar{d}_L^i \gamma_\mu d_L^j)(\bar{\mu}_L \gamma^\mu \mu_L)\,, \tag{2.8}$$

where $v \approx 246\,\mathrm{GeV}$ is the Higgs vacuum expectation value and $\mathbf{C}_{ij}^{U\mu}$ and $\mathbf{C}_{ij}^{D\mu}$ represent matrices of Wilson coefficients. In eq. (2.8), $i, j = 1, 2, 3$ indicate quark flavour indices, and we have chosen to focus on those operators that couple the quark fields exclusively to the second lepton family.

The operators highlighted in eq. (2.8) have received a lot of attention recently in the context of the LHCb anomalies reported in rare B-meson decays [78–81]. The reason is that the CKM-like flavour structure relates the $b \to s\mu^+\mu^-$ decays to the neutral-current Drell-Yan process at the LHC $p\,p \to \mu^+\mu^-$ [27]. The explicit models which successfully describe the LHCb anomalies, based on the U(2) flavour symmetry and dominant dynamics with the third generation fermions [88–91], predict that the flavour channel dominating EFT effects in the Drell-Yan production is $b\,\bar{b} \to \mu^+\mu^-$ (see [92] for an explicit model example). The direct $b\bar{s}$ production channel is suppressed by V_{ts} and is therefore irrelevant. If the observed deviations in $R(K^{(*)})$ are due to new physics, in this class of models we generically expect $|\mathbf{C}_{33}^{D\mu}| \gtrsim 0.001$.

The ATLAS dimuon search reported in ref. [93] is recast in ref. [27] to set the limit on this scenario. In particular, the reported 95% confidence level interval is

$$\mathbf{C}_{33}^{D\mu} \in [-0.026, 0.021] \quad (\mathrm{ATLAS\ 13\,TeV},\ 36.1\,\mathrm{fb}^{-1}\ [93]). \tag{2.9}$$

For this second benchmark scenario we will assume that, out of the operators listed in eq. (2.8), only a single Wilson coefficient is allowed to be non-zero. Specifically, we allow $\mathbf{C}_{33}^{D\mu} \neq 0$, while setting to zero all the coefficients of the other four-fermion operators. This assumption implies that in this scenario the SMEFT Lagrangian is reduced to

$$\mathcal{L}_{\mathrm{SMEFT}} = \mathcal{L}_{\mathrm{SM}} + \frac{\mathbf{C}_{33}^{D\mu}}{v^2}(\bar{d}_L^3 \gamma_\mu d_L^3)(\bar{\mu}_L \gamma^\mu \mu_L)\,. \tag{2.10}$$

In contrast to the previous benchmark, now the electron channel is SM-like. This feature provides a useful handle to separate PDF and EFT effects in the Drell-Yan process, by using electron data to determine the former and muon data to constrain both. Another difference with respect to the first benchmark is that here the leading new physics effects arise at the dimension-six squared level, since the interference of the operator in eq. (2.10) with the SM is subleading [27]. (Dijet production is another prominent process relevant for flavour physics [94] that enters into PDF fits. A detail study of the interplay is left for future work.)

To summarise, in this second benchmark scenario there is a single non-zero Wilson coefficient, $\mathbf{C}_{33}^{D\mu}$. Therefore, the determination of the SMEFT PDFs from Drell-Yan data requires a simultaneous fit of the PDFs together with the $\mathbf{C}_{33}^{D\mu}$ parameter. One should also note that this operator enters in the description of the DIS neutral-current structure functions via $\mu\,b$ scattering, though this contribution is highly suppressed due to the smallness of the bottom PDFs and the low energy scale probed by DIS data.

3 Experimental data, theory predictions, and fit settings

In section 3.1 we present the LHC experimental data that will be used in the present analysis for the simultaneous determination of the PDFs and the EFT coefficients from high-mass Drell-Yan cross sections. We then describe in section 3.2 the corresponding theoretical calculations, both in the SM and in the two SMEFT benchmark scenarios described in section 2. In section 3.3 we discuss the settings of the baseline SM PDF fit and assess the specific impact of the Run I and Run II high-mass Drell-Yan data on PDFs. Finally, in section 3.4 we outline the fitting methodology adopted for the determination of the PDFs in the SMEFT, along with their simultaneous determination with the EFT Wilson coefficients.

3.1 Experimental data

The present analysis is based on the DIS and DY measurements which were part of the strangeness study of [95], which in turn was a variant of the NNPDF3.1 global PDF determination [74], extended with additional high-mass DY cross sections. The DIS structure functions include the same legacy HERA inclusive combination [96] used in the DIS-only joint fit of PDF and EFT effects of [72].

No other datasets beyond DIS and DY are considered. In particular, the inclusive jet and top quark production measurements used in [95] are excluded from the present analysis. The rationale behind this choice is the following. SMEFT at dimension-6 level introduces 2499 independent parameters, many of which contribute to the processes used to extract the parton distribution functions. The full PDF fit in the SMEFT (with the consistent power counting in the inverse powers of the new physics scale) is the ultimate future goal of this line of research. Before that, we are forced to make assumptions about the subset of operators and processes involved. The restricted choice of DIS and DY is motivated by the idea that other datasets, such as inclusive jet, could potentially receive corrections from other SMEFT operators, e.g. four-quark operators while being insensitive to the semi-leptonic operators. Including all datasets to effectively determine PDF, while considering one or two operators able to impact a subset of processes, would misrepresent the realistic case.

For the purposes of our study, the DY data can be classified into low-mass, on-shell, and high-mass datasets. Table 1 summarises the low-mass and on-shell datasets, where in each case we indicate the experiment, the centre-of-mass energy $\sqrt{s}$, the publication reference, the physical observable, and the number of data points. The only difference as compared to [95] is the removal of the $W \to e\nu$ asymmetry measurements from D0 [97], which were found to be inconsistent with the rest of the Drell-Yan data.

In table 2 we provide the same information as in table 1 but for the neutral-current high-mass Drell-Yan datasets. In table 2 we also indicate the final state, whether the distribution is 1D or 2D (thus differential only in the lepton invariant mass or differential in the lepton invariant mass and rapidity), the integrated luminosity $\mathcal{L}$, and the values of the dilepton invariant mass $m_{\ell\ell}$ for the most energetic bin. We note that while the ATLAS and CMS measurements at $\sqrt{s} = 7$ TeV [120, 121] were already part of the strangeness study of [95], the corresponding 8 TeV and 13 TeV measurements from [86, 87, 122] were not and are being considered for the first time in this analysis. For those datasets where

Exp.	$\sqrt{s}$ (TeV)	Ref.	Observable	n_{dat}
E886	0.8	[98]	$d\sigma^d_{\mathrm{DY}}/d\sigma^p_{\mathrm{DY}}$	15
E886	0.8	[99, 100]	$d\sigma^p_{\mathrm{DY}}/(dy\,dm_{\ell\ell})$	89
E605	0.04	[101]	$\sigma^p_{\mathrm{DY}}/(dx_F\,dm_{\ell\ell})$	85
CDF	1.96	[102]	$d\sigma_Z/dy_Z$	29
D0	1.96	[103]	$d\sigma_Z/dy_Z$	28
D0	1.96	[104]	$d\sigma_{W\to\mu\nu}/d\eta_\mu$ asy.	9
ATLAS	7	[105]	$d\sigma_W/d\eta_l, d\sigma_Z/dy_z$	30
ATLAS	7	[106]	$d\sigma_{Z\to e^+e^-}/dm_{e^+e^-}$	6
ATLAS	7	[107]	$d\sigma_W/d\eta_l, d\sigma_Z/dy_z$	61
ATLAS	7	[108]	$d\sigma_{W+c}/dy_c$	22
ATLAS	8	[109]	$d\sigma_Z/dp_T$	82
ATLAS	8	[110]	$d\sigma_{W+j}/dp_T$	32
CMS	7	[111]	$d\sigma_{W\to l\nu}/d\eta_\ell$ asy.	22
CMS	7	[112]	$d\sigma_{W+c}/dy_c$	5
CMS	7	[112]	$d\sigma_{W^++c}/d\sigma_{W^-+c}$	5
CMS	8	[113]	$d\sigma_Z/dp_T$	28
CMS	8	[114]	$d\sigma_{W\to\mu\nu}/d\eta_\mu$	22
CMS	13	[115]	$d\sigma_{W+c}/dy_c$	5
LHCb	7	[116]	$d\sigma_{Z\to\mu^+\mu^-}/dy_{\mu^+\mu^-}$	9
LHCb	7	[117]	$d\sigma_{W,Z}/d\eta$	29
LHCb	8	[118]	$d\sigma_{Z\to e^+e^-}/dy_{e^+e^-}$	17
LHCb	8	[119]	$d\sigma_{W,Z}/d\eta$	30
Total				**659**

Table 1. The low-mass and on-shell Drell-Yan datasets used in the present study. For each dataset we indicate the experiment, the centre-of-mass energy $\sqrt{s}$, the publication reference, the physical observable, and the number of data points.

data are available in terms of both Born and dressed leptons, the ATLAS 7 TeV analysis being an example thereof, we use the Born data so that it is not necessary to supplement our fixed-order predictions with final-state QED radiation corrections. The CMS 13 TeV data on the other hand are only provided in terms of dressed leptons. In total, there are either 270 or 313 data points in this high-mass category, depending on whether the 13 TeV CMS data are included in the combined channel or in the separate electron and muon channels.

From table 2 one can observe that, with the exception of the CMS 13 TeV data, only one specific leptonic final state is available to be used in the fit. For the CMS 13 TeV measurement instead, one can select between the combined channel or the individual electron and muon final states, which are statistically independent. The separate use of the electron and muon channels is potentially beneficial when considering BSM effects that are not lepton-flavour universal. For example, in benchmark scenario II described in section 2, the

Exp.	$\sqrt{s}$ (TeV)	Ref.	$\mathcal{L}$ (fb^{-1})	Channel	1D/2D	n_{dat}	$m_{\ell\ell}^{\max}$ (TeV)
ATLAS	7	[120]	4.9	e^-e^+	1D	13	[1.0, 1.5]
ATLAS (*)	8	[86]	20.3	$\ell^-\ell^+$	2D	46	[0.5, 1.5]
CMS	7	[121]	9.3	$\mu^-\mu^+$	2D	127	[0.2, 1.5]
CMS (*)	8	[87]	19.7	$\ell^-\ell^+$	1D	41	[1.5, 2.0]
CMS (*)	13	[122]	5.1	e^-e^+, $\mu^-\mu^+$ $\ell^-\ell^+$	1D	43, 43 43	[1.5, 3.0]
Total						**270 (313)**	

Table 2. Same as table 1 for the neutral-current high-mass Drell-Yan datasets considered in this work. We also indicate the final-state, whether the distribution is 1D (which are differential in the invariant mass, $m_{\ell\ell}$, of the final-state leptons) or 2D (which are differential in both the invariant mass of the leptons, $m_{\ell\ell}$, and in their rapidity, $y_{\ell\ell}$), and the values of $m_{\ell\ell}$ for the most energetic bin. Datasets indicated with (*) are used for the first time in this analysis in comparison with [95].

theoretical predictions for the DY electron data would be those of the SM while those of the muon data should include EFT corrections. On the other hand in the (flavour-universal) $\hat{W}$ and $\hat{Y}$ scenario, it is more convenient to include the data from the combined channel, which displays reduced systematic uncertainties.

3.2 Theoretical predictions

We now discuss the settings of the theoretical calculations, both in the SM and in the SMEFT. Appendix C contains further information regarding the computation and bench-marking of the SMEFT corrections for both the DIS structure functions, for which the effect in both scenarios is negligible, and the DY cross sections, for which the impact of SMEFT corrections is much more sizeable.

SM cross sections. The SM cross sections are computed at next-to-next-to-leading order (NNLO) in QCD and include next-to-leading order (NLO) EW corrections, the latter being especially significant in the high-mass region relevant for this study. In particular, the DIS reduced cross sections (combinations of structure functions) are evaluated at NNLO in the FONLL-C general-mass variable flavour number scheme [123] with APFEL [124] interfaced to APFELgrid [125]. The Drell-Yan differential distributions are computed using MCFM [126] and MadGraph5_aMC@NLO [127] interfaced to APPLgrid [128] and APFELgrid to generate fast NLO interpolation tables which are then supplemented by bin-by-bin K-factors to account for the NNLO QCD and NLO EW corrections. These K-factors are defined as

$$d\sigma_{pp} = \left(d\hat{\sigma}_{ij}\big|_{\mathrm{NLO\,QCD}} \otimes \mathcal{L}_{ij}^{\mathrm{NNLO}}\right) \times K_{\mathrm{QCD}} \times K_{\mathrm{EW}}, \tag{3.1}$$

where $\otimes$ represents the standard convolution product, $d\sigma_{pp}$ $(d\hat{\sigma}_{ij})$ is the short-hand notation for the bin-by-bin hadronic cross section (partonic cross section for partons i, j) differential in $m_{\ell\ell}$ (in case of neutral-current (NC) Drell-Yan) or m_T (in case of charged-current (CC)

Drell-Yan) and the partonic luminosities $\mathcal{L}_{ij}$ are defined as

$$\mathcal{L}_{ij}(\tau, m) = \int_\tau^1 \frac{dx}{x} \, f_i(x, m) f_j(\tau/x, m) \,, \tag{3.2}$$

where $m = m_{\ell\ell}$ in the NC case and $m = m_T$ in the CC case and are evaluated at NNLO. The QCD and EW K-factors are defined as

$$K_{\mathrm{QCD}} = \left(\mathcal{L}_{ij}^{\mathrm{NNLO}} \otimes d\widehat{\sigma}_{ij}\big|_{\mathrm{NNLO\,QCD}} \right) / \left(\mathcal{L}_{ij}^{\mathrm{NNLO}} \otimes d\widehat{\sigma}_{ij}\big|_{\mathrm{NLO\,QCD}} \right) \,, \tag{3.3}$$

$$K_{\mathrm{EW}} = \left(\mathcal{L}_{ij}^{\mathrm{NNLO}} \otimes d\widehat{\sigma}_{ij}\big|_{\mathrm{NLO\,QCD+EW}} \right) / \left(\mathcal{L}_{ij}^{\mathrm{NNLO}} \otimes d\widehat{\sigma}_{ij}\big|_{\mathrm{NLO\,QCD}} \right) \,, \tag{3.4}$$

The NNLO QCD K-factors have been computed using either `MATRIX` [129] or `FEWZ` [130] and cross-checked with the analytic computations of [131, 132]. The NLO EW K-factors have been evaluated with MADGRAPH5_AMC@NLO [127]. Eq. (3.4) accounts also for photon-initiated contributions (using the `NNPDF3.1QED` PDF set [133]) and final-state radiation effects, except when the latter has already been subtracted in the corresponding experimental analysis.

Figure 1 displays a comparison between the CMS Drell-Yan distributions at 13 TeV and the corresponding theoretical predictions as a function of the dilepton invariant mass $m_{\ell\ell}$, separately for the dielectron and dimuon final states. The theory calculations are presented at NLO QCD, NNLO QCD, and NNLO QCD combined with NLO EW corrections, in all cases with `NNPDF3.1QED_nnlo_as_0118` as input PDF set, to illustrate the effect of the K-factors of eq. (3.3) and (3.4). The CMS data are provided in terms of dressed leptons, and hence final state radiation (FSR) QED effects must be included in the electroweak corrections. Accounting for these effects is essential to improve the agreement between theory and data in the region below the Z-mass peak. NLO electroweak corrections are also important in the high-energy tail in $m_{\ell\ell}$, where they are driven by the interplay between (negative) virtual EW effects and (positive) photon-initiated contributions.

A quantitative assessment of the agreement between theoretical predictions and experimental data for the high-mass DY datasets listed in table 2 is presented in table 3, which collects the values of the χ^2 per data point evaluated using the full information on correlated systematics provided by the experimental covariance matrix

$$\chi^2 = \frac{1}{n_{\mathrm{dat}}} \sum_{i,j=1}^{n_{\mathrm{dat}}} (D_i - T_i) \, (\mathrm{cov}^{-1})_{ij} \, (D_j - T_j), \tag{3.5}$$

where T_i are the theoretical predictions, D_i the central value of the experimental data and where the multiplicative uncertainties in the experimental covariance matrix (cov_{ij}) are treated as explained in [134, 135]. One can observe how in general the NNLO QCD corrections are relatively small and that the NLO electroweak ones can be significant, especially for observables presented in terms of dressed leptons (such as the CMS 13 TeV ones) and are required to achieve a good description of the Drell-Yan data in the whole kinematical range available. Note that the input PDF sets used for these calculations include only a subset of these Drell-Yan measurements, in particular only the 7 TeV measurements, for which the data-theory agreement is comparable to the one observed in [74].
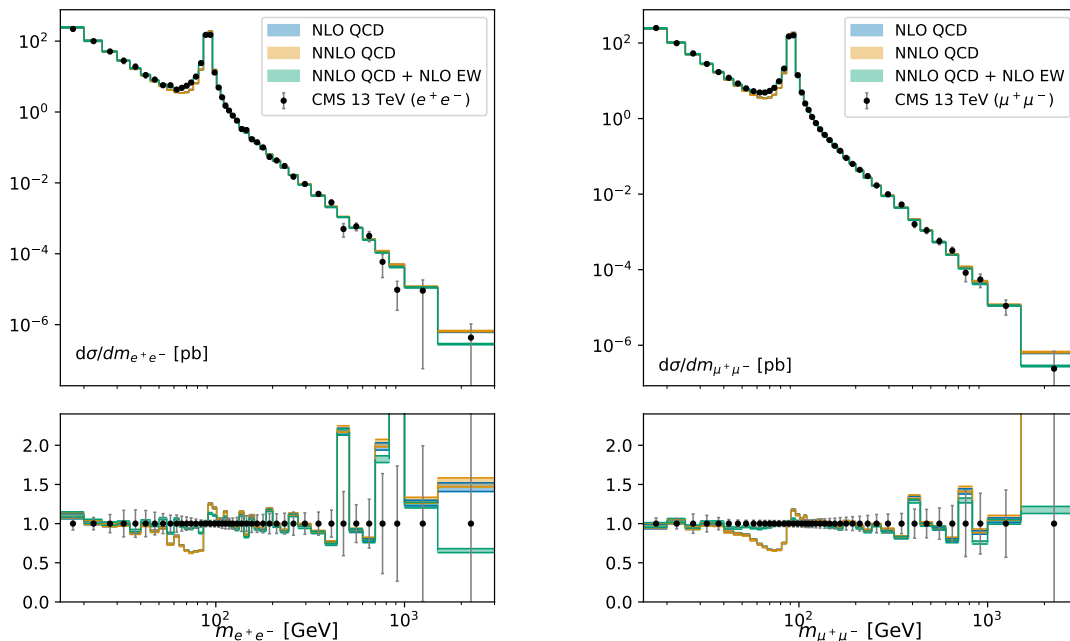
Figure 1. Comparison of the CMS Drell-Yan 13 TeV data with the corresponding theoretical calculations at different perturbative orders as a function of the dilepton invariant mass $m_{\ell\ell}$ in the dielectron (left) and dimuon (right panel) final states. The bottom panels display the ratio of the theory calculations to the central value of the experimental data. We display the sum in quadrature of the experimental uncertainties, and the error band in the theory predictions correspond to the one-sigma PDF uncertainties.

The data-theory agreement before including the 8 TeV and 13 TeV data in the PDF fit is generally good, once EW corrections are included, with the exception of the CMS 13 TeV data in the e^+e^- channel, for which the χ^2 per data point remains above 2. As can be observed in figure 1, the dielectron invariant mass distribution in this channel presents dips at about 500 GeV and 900 GeV which are not present in the $\mu^+\mu^-$ channel. These dips are the origin of this worse data-theory agreement, which is partially reduced once the dataset is included in the fit (see section 3.3). We have verified that excluding this dataset from the fit does not change the results of the analysis, and therefore decided to keep it. Further experimental analysis based on the full Runs II and III datasets will tell whether the dips in the distributions in the electron invariant mass will stay.

SMEFT cross sections. In this work, we augment the SM calculations of the high-Q^2 DIS reduced cross sections discussed in ref. [72] and the high-mass Drell-Yan cross sections listed in table 2 with the effects of dimension-six SMEFT operators following the two benchmark scenarios presented in section 2. These corrections are negligible for dilepton invariant masses of $m_{\ell\ell} \leq 200$ GeV and for DIS structure functions with $Q \leq (120)$ GeV,

Dataset	Final state	n_{dat}	χ^2/n_{dat}		
			NLO QCD	NNLO QCD	NNLO QCD + NLO EW
ATLAS 7 TeV	e^+e^-	13	1.45	1.77	1.73
ATLAS 8 TeV	$\ell^+\ell^-$	46	1.67	—	1.20
CMS 7 TeV	$\mu^+\mu^-$	127	3.40	1.27	1.54
CMS 8 TeV	$\ell^+\ell^-$	41	2.22	2.21	0.70
CMS 13 TeV	$\ell^+\ell^-$	43	18.7	19.7	1.91
CMS 13 TeV	e^+e^-	43	9.16	9.45	2.32
CMS 13 TeV	$\mu^+\mu^+$	43	15.7	15.8	0.81

Table 3. The values of the χ^2 per data point evaluated for the high-mass DY datasets listed in table 2, using theoretical predictions computed at different perturbative accuracy. The PDF sets used here are `NNPDF31_nlo_as_0118`, `NNPDF31_nnlo_as_0118` and `NNPDF31_nnlo_as_0118_luxqed` for the NLO QCD, NNLO QCD and NNLO QCD + NLO EW predictions respectively. For CMS 13 TeV, where different final states are available, we indicate the χ^2 values for each of them. For the ATLAS 8 TeV data, we only evaluated the combined NNLO QCD + NLO EW correction, and hence the pure NNLO QCD result is not given.

and hence there can safely adopt the SM calculations. In a similar manner as for higher-order QCD and EW corrections, we can define correction factors that encapsulate the linear and quadratic modifications induced by the dimension-six SMEFT operators. Adopting an operator normalisation such that

$$\mathcal{L}_{\mathrm{SMEFT}} = \mathcal{L}_{\mathrm{SM}} + \sum_{n=1}^{n_{\mathrm{op}}} \frac{c_n}{v^2}\,\mathcal{O}_n\,, \tag{3.6}$$

with n_{op} indicating the number of operators that contribute to a given benchmark scenario and c_n being the (dimensionless) Wilson coefficient associated to $\mathcal{O}_n$, the linear EFT corrections can be parametrised as

$$R_{\mathrm{SMEFT}}^{(n)} \equiv \left(\mathcal{L}_{ij}^{\mathrm{NNLO}} \otimes d\widehat{\sigma}_{ij,\mathrm{SMEFT}}^{(n)} \right) \Big/ \left(\mathcal{L}_{ij}^{\mathrm{NNLO}} \otimes d\widehat{\sigma}_{ij,\mathrm{SM}} \right)\,, \quad n = 1\ldots,n_{\mathrm{op}}\,, \tag{3.7}$$

with $\mathcal{L}_{ij}^{\mathrm{NNLO}}$ being the usual partonic luminosity evaluated at NNLO QCD, $d\widehat{\sigma}_{ij,\mathrm{SM}}$ the bin-by-bin partonic SM cross section, and $d\widehat{\sigma}_{ij,\mathrm{SMEFT}}^{(n)}$ the corresponding partonic cross section associated to the interference between $\mathcal{O}_n$ and the SM amplitude $\mathcal{A}_{\mathrm{SM}}$ when setting $c_n = 1$. Likewise, the ratio encapsulating the quadratic effects is defined as

$$R_{\mathrm{SMEFT}}^{(n,m)} \equiv \left(\mathcal{L}_{ij}^{\mathrm{NNLO}} \otimes d\widehat{\sigma}_{ij,\mathrm{SMEFT}}^{(n,m)} \right) \Big/ \left(\mathcal{L}_{ij}^{\mathrm{NNLO}} \otimes d\widehat{\sigma}_{ij,\mathrm{SM}} \right)\,, \quad n,m = 1\ldots,n_{\mathrm{op}}\,, \tag{3.8}$$

with the bin-by-bin partonic cross section $d\widehat{\sigma}_{ij,\mathrm{SMEFT}}^{(n,m)}$ now being evaluated from the squared amplitude $\mathcal{A}_n\mathcal{A}_m$ associated to the operators $\mathcal{O}_n$ and $\mathcal{O}_m$ when $c_n = c_m = 1$. The partonic

cross sections in these ratios are computed at LO. In terms of eqs. (3.7) and (3.8), we can define the EFT K-factors as

$$K_{\text{EFT}} = 1 + \sum_{n=1}^{n_{\text{op}}} c_n R_{\text{SMEFT}}^{(n)} + \sum_{n,m=1}^{n_{\text{op}}} c_n c_m R_{\text{SMEFT}}^{(n,m)} \,, \tag{3.9}$$

which allow us to express a general Drell-Yan or DIS cross sections accounting for the dimension-six operators in eq. (3.6) as

$$d\sigma_{\text{SMEFT}} = d\sigma_{\text{SM}} \times K_{\text{EFT}} \tag{3.10}$$

where the $d\sigma_{\text{SM}}$ is the state-of-the-art SM prediction including NNLO QCD and NLO EW corrections. In this approach, the SMEFT predictions inherit factorisable higher-order radiative correction [27, 37]. The SMEFT K-factors in eq. (3.9) are precomputed before the fit using a reference SM PDF set and then kept fixed. The effect of varying the input NNLO PDF in eqs. (3.7) and (3.8) is quantitatively assessed in appendix C and it is found to be at the permil level in Scenario I and slightly more significant but still at most at the percent level in Scenario II. As a result, this effect will be neglected in the following. Further details about the implementation and validation of these K-factors can be found in appendix C.

Figure 2 illustrates the size of the EFT corrections in benchmark scenario I from section 2 by comparing $(K_{\text{EFT}} - 1)$ with the relative experimental uncertainties for the ATLAS 7 TeV, CMS 8 TeV, and CMS 13 TeV Drell-Yan $m_{\ell\ell}$ distributions. We provide results for two representative points in the $(\hat{W}, \hat{Y})$ parameter space, namely $(\hat{W}, \hat{Y}) = (10^{-3}, 0)$ and $(0, 10^{-3})$. One can observe how for these values of $(\hat{W}, \hat{Y})$, and particularly for the ATLAS 8 TeV data, the SMEFT corrections to the Drell-Yan cross sections become comparable with the experimental uncertainties, increasing steadily with $m_{\ell\ell}$.

3.3 Baseline SM PDFs

We now describe the fit settings used to assemble the PDF set that represents the baseline SM PDFs in this work. In other words, the results that we present in this section correspond to PDFs extracted from the experimental data using the SM theoretical predictions, and then in the next section we will asses how these PDFs are modified once EFT corrections to the DIS and DY cross sections are accounted for in the fit of the PDFs.

The settings for this baseline SM PDF fit are the same as those used in the strangeness study of [95], itself a variant of NNPDF3.1 [74]. As described in section 3.1, in this work we consider only DIS and Drell-Yan datasets, with the latter augmented as compared to [95] with the new high-mass measurements indicated in table 2. A detailed comparison between this baseline SM PDF and the NNPDF3.1_str set from [95] is provided in appendix A, while table 10 in appendix B details the breakdown of the χ^2 for all the datasets that enter the fit. In general, the fit quality of the baseline SM PDF set is similar to that of the global fit of [95], although the description of the CMS 13 TeV invariant mass distribution in the combined electron and muon channels remains sub-optimal.

Figure 3 displays a comparison between this baseline SM PDF set, labelled "DIS+DY", with the same fit but without any datapoints from the high-mass DY datasets listed in
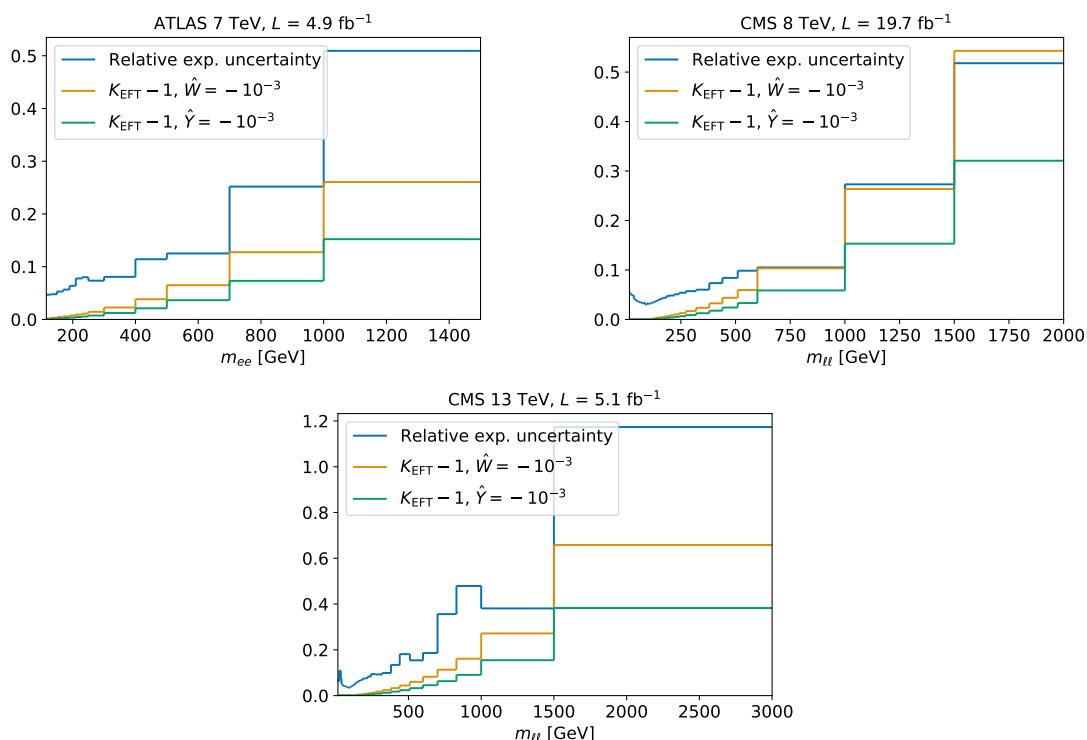
Figure 2. Comparison between the (relative) experimental uncertainties and the corresponding EFT corrections, $K_{\mathrm{EFT}}(\hat{W}, \hat{Y}) - 1$ in eq. (3.9), for the ATLAS 7 TeV, CMS 8 TeV, and CMS 13 TeV Drell-Yan $m_{\ell\ell}$ distributions, for two representative values of $\hat{W}$ and $\hat{Y}$.

table 2, labelled "DIS+DY (no HM)". We show results at $Q = 100\,\mathrm{GeV}$ both for the PDFs normalised to the central value of the baseline and for the relative PDF uncertainties. In the latter case we also display the PDF uncertainties from a corresponding DIS-only fit. The latter comparison shows that the DY cross sections significantly reduce the PDF uncertainties of the DIS-only fit. The addition of the high-mass DY data leads to a visible uncertainty reduction in the $0.005 \lesssim x \lesssim 0.3$ region as compared to the "DIS+DY(noHM)" reference as well an upwards shift of the up and down quarks and antiquark PDF.

We therefore find that the available high-mass DY data can have an appreciable impact on the light quark and antiquark PDFs, despite the fact that in terms of Run II data our analysis is restricted to a single low-luminosity high-mass DY dataset. Yet more stringent constraints on the PDFs are expected from the measurements based on the full Runs II and III datasets, as well as from those to be provided by the HL-LHC [77]. We study the anticipated impact of the HL-LHC measurements in section 5.

3.4 Methodology for the simultaneous PDF and EFT fits

Let us denote by $\mathbf{c} = (c_1, c_2, \ldots, c_{N_{\mathrm{op}}})$ the array containing the Wilson coefficients associated to the N_{op} dimension-six operators contributing to a given SMEFT scenario, where c_n are defined as in eq. (3.6). For each point $\mathbf{c}_i$ in the scan of the EFT parameter space, we evaluate the Drell-Yan and the DIS cross sections as described in section 3.2. Subse-
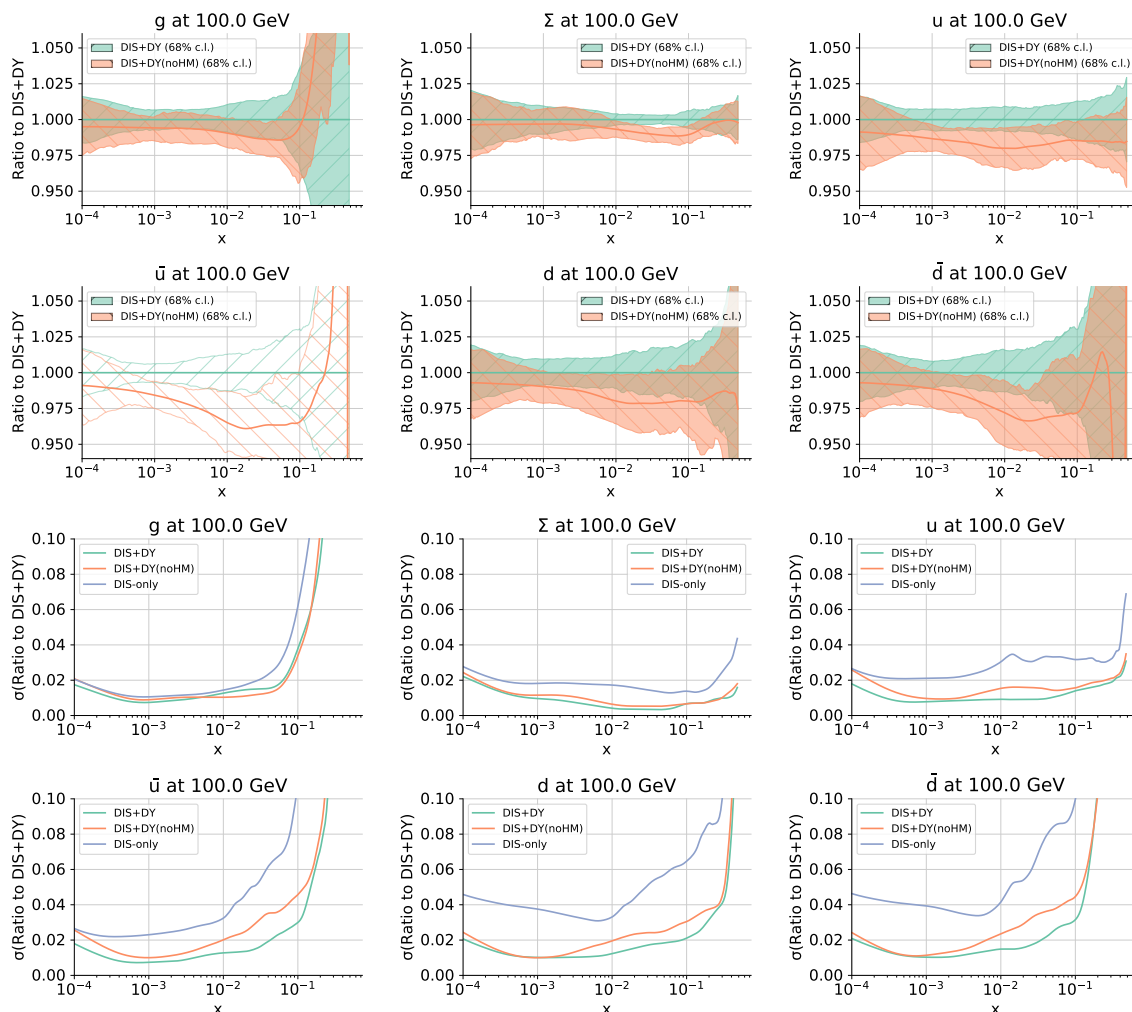
Figure 3. Comparison between the baseline SM PDF set of this work, labelled "DIS+DY", with the corresponding fit without high-mass DY data. We show results at $Q = 100\,\text{GeV}$ for PDFs normalised to the central value of the baseline (upper) and for the relative PDF uncertainties (lower panels). In the latter case, we also display the PDF uncertainties from the DIS-only fit.

quently, we determine the best-fit PDFs associated to $\mathbf{c}_i$ by means of the standard NNPDF methodology, which determines the minimum of the χ^2 in the space of the PDF parameters (subject to cross-validation, to avoid overlearning). We note that this χ^2, defined in eq. (3.5), keeps fully into account the experimental systematic correlations among all the measurements D_i included in the PDF analysis.

This procedure results in a sampling of the χ^2 values in the EFT parameter space, which we denote by $\chi^2_{\text{eftp}}(\mathbf{c}_i)$ (as in: EFT-PDFs). Alternatively, one could also evaluate the same DIS and DY cross sections using instead the baseline SM PDF set, ending up with χ^2 values which we denote by $\chi^2_{\text{smp}}(\mathbf{c}_i)$ (as in: SM-PDFs). The comparison between the resulting bounds on the EFT coefficients obtained from $\chi^2_{\text{eftp}}(\mathbf{c}_i)$ and from $\chi^2_{\text{smp}}(\mathbf{c}_i)$ quantifies the relevance of producing consistent joint determinations of PDFs and Wilson

coefficients when studying EFTs in high-energy tails. This strategy follows the one adopted in our proof-of-concept DIS-only study [72], now extended to LHC processes.

Close enough to a local minimum $\chi_0^2 = \chi^2\left(\mathbf{c}^{(0)}\right)$ associated with best-fit values $\mathbf{c}^{(0)}$, the χ^2 as a function of the EFT coefficients can be approximated by a quadratic form

$$\chi_i^2 \equiv \chi^2(\mathbf{c}_i) = \chi_0^2 + \sum_{n,m=1}^{N_{\mathrm{op}}} \left(c_{n,i} - c_n^{(0)}\right) H_{nm} \left(c_{m,i} - c_m^{(0)}\right), \tag{3.11}$$

with H_{nm} being the usual Hessian matrix in the EFT parameter space. Restricting the EFT calculations to their linear, $\mathcal{O}\left(\Lambda^{-2}\right)$, contributions, eq. (3.11) becomes exact in the case of $\chi_{\mathrm{smp}}^2(\mathbf{c}_i)$ (where cross sections are evaluated with SM PDFs). The reason is that in this case all dependence on the EFT coefficients is encoded in the partonic cross sections.

However, this is not true for $\chi_{\mathrm{eftp}}^2(\mathbf{c}_i)$, since now there will be a (non-linear) EFT back-reaction onto the PDFs and hence eq. (3.11) is only valid up to higher orders in the EFT expansion, even if the EFT cross sections themselves are evaluated in the linear approximation. Eq. (3.11) can thus be only considered a reasonable approximation in the case that the SMEFT PDFs are not too different from their SM counterparts.

Hence, if we work with linear EFT calculations, provided the sampling in the EFT parameter space is sufficiently broad and fine-grained, and that the EFT-induced distortion on the PDFs is moderate, we can extract the parameters χ_0^2 and $\mathbf{c}^{(0)}$ and the Hessian matrix H using least-squares regression from eq. (3.11), using χ_{smp}^2 for the SM PDFs and χ_{eftp}^2 for the SMEFT PDFs. The associated confidence level contours are determined by imposing

$$\Delta\chi^2(\mathbf{c}) \equiv \chi_i^2(\mathbf{c}) - \chi_0^2 = \sum_{n,m=1}^{N_{\mathrm{op}}} \left(c_n - c_n^{(0)}\right) H_{nm} \left(c_m - c_m^{(0)}\right) = \mathrm{constant}, \tag{3.12}$$

where this constant depends on the number of degrees of freedom. For linear EFT two-parameter fits, such as those for benchmark scenario I in the context of HL-LHC projections, imposing eq. (3.12) leads to elliptic contours in the $(\hat{W}, \hat{Y})$ plane.

In the case of fits to χ^2 profiles obtained from EFT calculations which include both linear, $\mathcal{O}\left(\Lambda^{-2}\right)$, and quadratic, $\mathcal{O}\left(\Lambda^{-4}\right)$, contributions, such as those arising in benchmark scenario II, rather than working in the Hessian approximation we instead carry out a (one-dimensional) quartic fit of the form

$$\chi^2(c) = \sum_{k=0}^{k_{\mathrm{max}}=4} a_k (c)^k \tag{3.13}$$

with the χ^2 values being χ_{smp}^2 (χ_{eftp}^2) for the SM (SMEFT) PDFs, and then determine confidence level intervals by imposing $\Delta\chi^2(c) = \chi_i^2(c) - \chi_0^2 = \mathrm{constant}$. We determine this constant numerically by finding the likelihood contour, $\mathcal{L} \propto \exp(-\chi^2/2)$, containing 95% of the total probability (for the 95% CL intervals).

To conclude this section, we give details on how we account for PDF uncertainties and the statistical uncertainty associated to the finite replica sample of the NNPDF Monte Carlo sets that we use here.

PDF uncertainty. In sections [4.1], [4.2] and [5.3] we will present bounds on the EFT parameters using the SM PDFs with and without the PDF uncertainties being accounted for. In order to estimate these, we follow the procedure detailed above to determine the confidence level intervals for the EFT parameters but now using the kth Monte Carlo replica of the PDF set, rather than the central replica $k = 0$ as done when PDF uncertainties are neglected. One ends up with N_{rep} values of the upper and lower bounds:

$$\left[\mathbf{c}_{\min}^{(k)}, \mathbf{c}_{\max}^{(k)}\right], \qquad k = 1, \ldots, N_{rep}, \tag{3.14}$$

and then the outermost bounds in the 68% envelope are considered to be the bounds on the EFT parameters $\mathbf{c}$, now including the 1σ-PDF uncertainty. This is very important to account for, given that in the case of the bounds determined using χ^2_{eftp}, the PDF uncertainty is already included by construction, given that the Wilson coefficients are determined from the global set of PDFs, exactly as in the case of the α_s determination from a global set of PDFs of [136, 137]. A more sophisticated way to extract parameters such as α_s of the Wilson coefficients from a global fit of PDFs, that includes the correlations between these parameters and the PDFs, is given by the correlated replica method proposed in the more recent α_s determination in [138]. The latter would allow better accounting of the correlations between Wilson coefficients and PDFs. However we do not use it here due to the fact that these correlations of the PDFs with the Wilson coefficients are much smaller than those with the strong coupling constant and due to its large computational cost. We endeavour to address this issue in future work.

Methodological uncertainty. In a simultaneous fit of PDFs and EFT coefficients, for each set of Wilson coefficients $\mathbf{c}_i$ one has a PDF fit composed of N_{rep} Monte Carlo replicas. The major methodological uncertainty is associated to finite-N_{rep} effects can be estimated by bootstrapping across the replicas, as explained in the $\alpha_s(m_Z)$ extraction of [138]. Specifically, for each value of $\mathbf{c}_i$ we perform N_{res} re-samples of all N_{rep} replicas with replacement, and compute the theory predictions:

$$\mathbf{T}_{i,lk}^{(res)}, \quad \begin{array}{l} l = 1, \ldots, N_{res}, \\ k = 1, \ldots, N_{rep}, \end{array} \tag{3.15}$$

such that there are N_{res} re-samples each composed of an N_{rep}-sized array of theory predictions. Since this re-sampling is done with replacement, it differs from the original sample in that it contains duplicates and missing values. The average theory prediction is then obtained for each of these bootstrapped sets:

$$\overline{\mathbf{T}}_{i,l} = \left\langle \mathbf{T}_{i,lk}^{(res)} \right\rangle_{rep}, \quad l = 1, \ldots, N_{res}. \tag{3.16}$$

These bootstrapped theory predictions $\overline{\mathbf{T}}_{i,l}$ are used to evaluate the χ^2 to data, with the finite-size uncertainty given by the standard deviation across each bootstrap re-sample:

$$\sigma_{\chi_i^2} = \text{std}\left(\chi_{i,l}^2\right)\Big|_{res}. \tag{3.17}$$

A value of $N_{res} \simeq 10^4$ re-samples is found to be sufficient to achieve stable results for the estimate of the finite-size uncertainties defined by eq. (3.17).

4 Results

In this section, we start by presenting results for the SMEFT PDFs extracted from DIS and Drell-Yan data in benchmark scenario I. We compare them with their SM counterparts at the level of partonic luminosities and assess how the bounds obtained on the $\hat{W}$ and $\hat{Y}$ parameters in this simultaneous SMEFT and PDF fit compare to those based on assuming SM PDFs. We then investigate the sensitivity of available high-mass Drell-Yan data to benchmark scenario II, where only the dimuon final state is modified by EFT effects. Finally, we quantify the impact that the consistent use of SMEFT PDFs has on the reinterpretation of high-mass dilepton BSM searches.

4.1 PDF and EFT interplay in current high-mass Drell-Yan data

By deploying the methodology described in section 3.4, we have extended the PDF analysis based on SM predictions presented in section 3.3 to account for the effects of non-zero EFT coefficients within benchmark scenario I defined in section 2.1. Here, we present results for one-dimensional fits where only one of the $\hat{W}$ or the $\hat{Y}$ parameter is allowed to be non-zero. The reason for this choice is that, in a fit including only high-mass neutral-current Drell-Yan processes, there exists a flat direction when $\hat{W}$ and $\hat{Y}$ are varied simultaneously, since both operators scale as q^4 and thus cannot both be constrained by a single 1D distribution. This degeneracy can only be lifted once high-mass charged-current DY data is included in the fit. As we demonstrate in section 5, thanks to the HL-LHC it will be possible to carry out a simultaneous fit of the PDFs and the two EFT parameters $(\hat{W}, \hat{Y})$.

Taking into account the existing bounds reported in section 2, as well as the sensitivity of available high-mass Drell-Yan data to the EFT coefficients illustrated by figure 2, here we have adopted the following sampling ranges for the $\hat{W}$ and $\hat{Y}$ parameters:

$$\left(\hat{W} \times 10^4\right) \in [-22, 14] \,, \qquad \left(\hat{Y} \times 10^4\right) \in [-20, 20] \,. \tag{4.1}$$

We used 21 sampling values of $\hat{Y}_i$ equally spaced in this interval, hence in steps of $\Delta \hat{Y} = 2 \times 10^{-4}$. In the case of $\hat{W}_i$ it was found convenient to instead use 15 points equally spaced between -14×10^{-4} and 14×10^{-4} in steps of $\Delta \hat{W} = 2 \times 10^{-4}$, and then to add two more values at $\hat{W}_i = -18 \times 10^{-4}$ and -22×10^{-4}.

Figure 4 displays the obtained values of $\Delta\chi^2$, eq. (3.12), as a function of $\hat{W}_i$ and $\hat{Y}_i$ in the case of the SMEFT PDFs. That is, using the values of $\chi^2_{\text{eftp}}(\mathbf{c}_i)$. These χ^2 values are evaluated as a sum over those datasets from table 1 and 2 that receive non-zero EFT corrections, namely the DIS datasets that have a reach in Q^2 above $(120)^2 \, \text{GeV}^2$ (namely HERA and NMC), and the ATLAS and CMS high-mass Drell-Yan measurements in table 2. The use of such a partial χ^2 rather than the global χ^2 is a necessary approximation due to the limitation of our current methodology. The statistical fluctuations of the global χ^2 are significantly larger than those of the partial χ^2 and can only be tamed by running a very large batch of replicas for each benchmark point in $\hat{W}$ and $\hat{Y}$ and by increasing the density of benchmark points in the region that is explored, as it was done for the scan of α_s in ref. [138]. However, while for the scan of α_s all processes contribute to the parabolic behaviour of the $\Delta\chi^2$, in this case the dominant contributions to χ^2_{eftp} come by far from
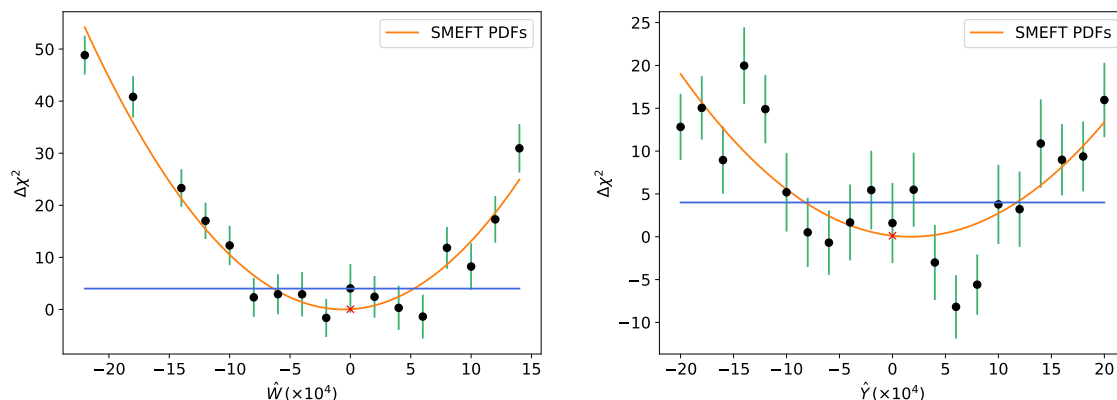
Figure 4. The values of $\Delta\chi^2$, eq. (3.12), obtained for the SMEFT PDFs (thus using the $\chi^2_{\text{eftp}}(\mathbf{c}_i)$ values) as a function of $\hat{W}_i$ (left) and $\hat{Y}_i$ (right panel) in the sampling ranges of eq. (4.1) together with the corresponding parabolic fits. The error bars indicate the finite-size uncertainties and the horizontal line corresponds to the $\Delta\chi^2 = 4$ condition defining the 95% CL intervals. The red cross indicates the SM expectation, $\hat{W} = \hat{Y} = 0$.

the SMEFT corrections to the hard cross section of these processes, and from the changes in the PDFs induced by non-zero Wilson coefficients. The latter changes in PDFs are confined to the large-x light quark and antiquark distributions, which affect the high-mass Drell-Yan data. The analysis of the χ^2_{eftp} computed on the subset of data captures the dominant effects, while minimising the level of statistical fluctuations.

A further approximation is given by the fact that only linear EFT effects are included in the calculation of the DIS and DY cross sections, while the (subleading) quadratic corrections are neglected in this scenario. The error bars in the $\Delta\chi^2_i$ points of figure 4 indicate the methodological finite-size uncertainties evaluated with the bootstrapping method described in section 3.4 and the horizontal line corresponds to the $\Delta\chi^2 = 4$ condition associated to a 95% CL interval. We also show in figure 4 the results of the associated parabolic fits,

$$\Delta\chi^2(\hat{W}) = \left(\hat{W} - \hat{W}^{(0)}\right)^2 / \left(\delta\hat{W}\right)^2 , \tag{4.2}$$

and likewise for $\Delta\chi^2(\hat{Y})$. From the results in figure 4, one observes that both the $\hat{W}$ and $\hat{Y}$ parameters agree with the SM expectation within uncertainties.

Figure 5 then compares the results of the parabolic fits based on the SMEFT PDFs as displayed in figure 4 with their counterparts obtained in the case of the SM PDFs. That is, in the latter case one carries out parabolic fits to the χ^2_{smp} values, as is customary in the literature for the EFT analyses. The insets highlight the region close to $\Delta\chi^2 \simeq 0$. For the $\hat{W}$ parameter, the consistent use of SMEFT PDFs leaves the best-fit value essentially unchanged but increases the coefficient uncertainty $\delta\hat{W}$, leading to a broader parabola. Similar observations can be derived for the $\hat{Y}$ parameter, though here one also finds a upwards shift in the best-fit values by $\Delta\hat{Y} \simeq 2 \times 10^{-4}$ in addition to a parabola broadening, when SMEFT PDFs are consistently used. We note that the SM PDF parabolas in figure 5 are evaluated using the central PDF replica and hence do not account for PDF uncertainties.
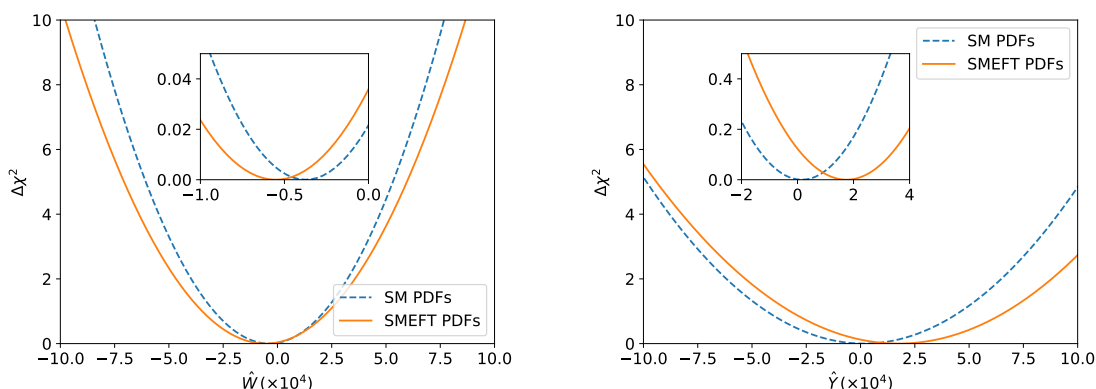
Figure 5. Comparison between the results of the parabolic fits to $\Delta\chi^2$, eq. (4.2), for the $\hat{W}$ (left) and $\hat{Y}$ (right panel) parameters for either the SMEFT PDFs (χ^2_{eftp}, already displayed in figure 4) or the SM PDFs (hence with χ^2_{smp}). The insets zoom on the region close to $\Delta\chi^2 \simeq 0$.

Table 4 summarises the 68% and 95% CL bounds on the $\hat{W}$ and $\hat{Y}$ parameters obtained from the corresponding parabolic $\Delta\chi^2$ fits using either the SM or the SMEFT PDFs shown in figure 5. The fourth and fifth column indicate the absolute shift in best-fit values and the percentage broadening of the fit parameter uncertainties when the SMEFT PDFs are consistently used instead of the SM PDFs (either without or with PDF uncertainties):

$$\text{best fit shift} \equiv \left(\hat{W}^{(0)}\Big|_{\text{SMEFT PDF}} - \hat{W}^{(0)}\Big|_{\text{SM PDF}} \right), \tag{4.3}$$

$$\text{broadening} \equiv \left(\delta\hat{W}^{(0)}\Big|_{\text{SMEFT PDF}} - \delta\hat{W}^{(0)}\Big|_{\text{SM PDF}} \right) \Big/ \delta\hat{W}^{(0)}\Big|_{\text{SM PDF}}, \tag{4.4}$$

and likewise for the $\hat{Y}$ parameter.

In the specific case of the SM PDF results, table 4 indicates the bounds obtained without (upper) and with (lower entry) PDF uncertainties accounted for; recall that the SMEFT PDF bounds already include PDF uncertainties by construction (see section 3.4). The methodological (finite-size) uncertainty is included according to the approach described in section 3.4 and it amounts to $4.7 \cdot 10^{-5}$ in the case of $\hat{W}$ and $1.0 \cdot 10^{-4}$ in the case of $\hat{Y}$, corresponding to 4% and 5% respectively of the 95% C.L. bounds for the $\hat{W}$ and $\hat{Y}$ coefficients.

By comparing the bounds obtained when PDF uncertainties are accounted for to those neglecting PDF uncertainty, one observes a systematic broadening of the bounds from both the lower and upper limits, as was also reported in [72].

When PDF uncertainties are neglected (accounted for) when using the SM PDFs to constrain the EFT parameters, the consistent use of the SMEFT PDFs leads to both a shift in the best-fit values of magnitude $\Delta\hat{W} = -2 \times 10^{-5}$ and $\Delta\hat{Y} = +1.6 \times 10^{-4}$ as well as to an increase (decrease) of the fit parameter uncertainties, with $\delta\hat{W}$ and $\delta\hat{Y}$ growing by 15% and 12% (decreasing by 11% and 13%) respectively. This result shows that, given available Drell-Yan data and once PDF uncertainties are accounted for, the bounds on the EFT parameters are actually *improved* once SMEFT PDFs are adopted.

	SM PDFs	SMEFT PDFs	best-fit shift	broadening
$\hat{W} \times 10^4$ (68% CL)	$[-3.0, 2.2]$	$[-3.5, 2.4]$	-0.2	$+13\%$
	$[-4.3, 3.8]$		-0.3	-27%
$\hat{W} \times 10^4$ (95% CL)	$[-5.5, 4.7]$	$[-6.4, 5.3]$	-0.2	$+15\%$
	$[-6.8, 6.3]$		-0.3	-11%
$\hat{Y} \times 10^4$ (68% CL)	$[-4.4, 4.7]$	$[-3.4, 6.9]$	$+1.6$	$+13\%$
	$[-6.7, 7.5]$		$+1.4$	-27%
$\hat{Y} \times 10^4$ (95% CL)	$[-8.8, 9.2]$	$[-8.3, 11.8]$	$+1.6$	$+12\%$
	$[-11.1, 12.0]$		$+1.3$	-13%

Table 4. The 68% CL and 95% CL bounds on the $\hat{W}$ and $\hat{Y}$ parameters obtained from the corresponding parabolic fits to the $\Delta\chi^2$ values calculated from either the SM or the SMEFT PDFs. For the SM PDF results, we indicate the bounds obtained without (upper) and with (lower entry) PDF uncertainties accounted for; the SMEFT PDF bounds already include PDF uncertainties by construction, while the methodological (finite-size) uncertainty is included according to the approached described in section 3.4. The fourth and fifth column indicate the absolute shift in best-fit values, eq. (4.3) and the percentage broadening of the EFT parameter uncertainties, eq. (4.4), when the SMEFT PDFs are consistently used instead of the SM PDFs.

All in all, the effect of the consistent treatment of the SMEFT PDFs in the interpretation of high-mass DY cross sections is moderate but not negligible, either loosening or tightening up the obtained bounds on the EFT parameters (depending on whether or not PDF uncertainties are accounted for to begin with) by up to 15% and, in the case of $\hat{Y}$ parameter, shifting its central value by one-third of the 68% CL parameter uncertainty. Such a relatively moderate effect can be partly understood from the limited availability of high-mass DY measurements for EFT interpretations, with a single dataset at 13 TeV, and even in this case, with it being restricted to a small fraction of the Run II luminosity. As we will demonstrate in section 5, the impact of SMEFT PDFs becomes much more significant once higher-statistics measurements of the NC and CC Drell-Yan tails become available at the HL-LHC, loosening the bounds on $\hat{W}$ and $\hat{Y}$ by up to a factor 5.

Comparing the limits on the $\hat{W}$ and $\hat{Y}$ parameters from table 4 with those of ref. [32] and reported in section 2, we observe that our bounds are more stringent. There are two main reasons that could explain this difference. On the one hand, on top of the ATLAS and CMS high-mass DY cross sections at 8 TeV, we also include the corresponding 7 and 13 TeV data that provide additional weight to the high invariant mass region of the spectrum in the fit. On the other hand, in our analysis we fit the whole invariant mass spectrum and do not cut away the low $m_{\ell\ell}$ region below 120 GeV, thus we do not ignore the correlations between the low and high ends of the spectrum which are important even if the former is not affected by SMEFT corrections.
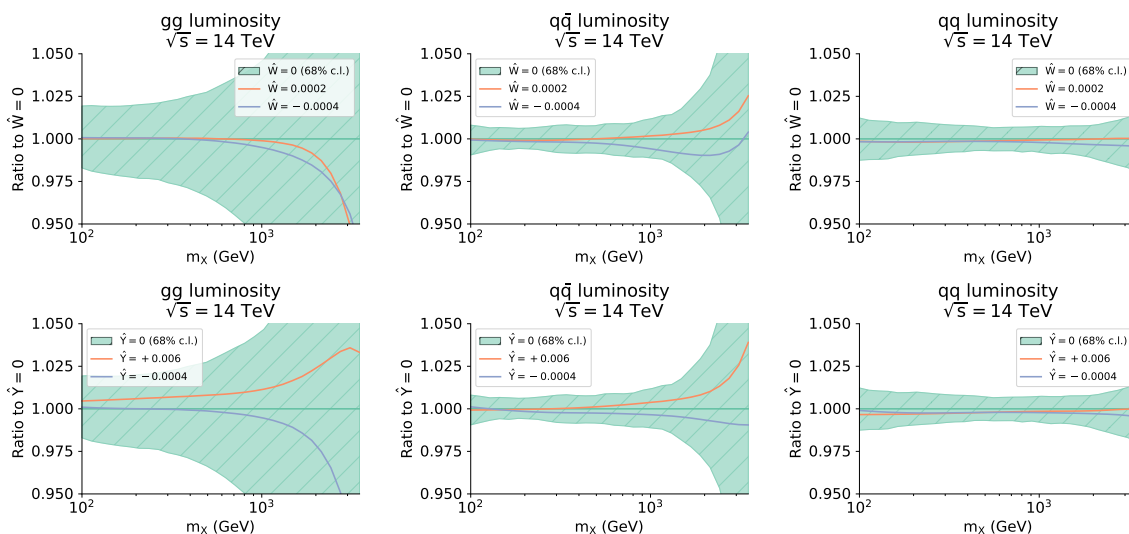
Figure 6. Comparison between the SM PDF luminosities with their SMEFT counterparts, displayed as ratios to the central value of the SM luminosities, for representative values of the $\hat{W}$ (upper) and $\hat{Y}$ (lower panel) parameters. The values of $\hat{W}$ and $\hat{Y}$ are chosen to be close to the upper and lower limits of the 68% CL intervals reported in table 4.

We now move to assess how the SMEFT PDFs relate to their SM counterparts, and determine the extent to which it is possible to reabsorb EFT effects into the PDFs. Figure 6 displays a comparison between the SM and the SMEFT PDF luminosities for representative values of the $\hat{W}$ (upper) and $\hat{Y}$ (lower panel) parameters. The values of $\hat{W}$ and $\hat{Y}$ are chosen to be close to the upper and lower limits of the 68% CL intervals reported in table 4. The error band in the SM PDFs corresponds to the 68% CL PDF uncertainty, while for the SMEFT PDFs only the central values are shown.

In all cases, one finds that the EFT-induced shifts on the luminosities are smaller than their standard deviation. The biggest differences, relative to uncertainties, are observed in the quark-antiquark luminosities for $m_X \gtrsim 500\,\text{GeV}$. This finding can be understood from the fact that the NC Drell-Yan cross section is proportional to the $u\bar{u}$ and $d\bar{d}$ combinations at leading order, but the up and down quark PDFs are already well constrained by lower-energy DIS measurements. Furthermore, we have verified that the size PDF uncertainties is unchanged in the SMEFT fits. The results of figure 6 are consistent with those of table 4 and demonstrate that, with current data, the interplay between EFT effects and PDFs in the high-mass Drell-Yan tails is appreciable but remains subdominant as compared to other sources of uncertainty.

One important question in this context concerns how one could disentangle the EFT-induced shifts in the PDF luminosities displayed in figure 6 (see also the corresponding PDF comparisons in figure 31) from other possible sources of deviations, such as internal inconsistencies in some datasets or missing higher orders in the SM calculations. An attractive strategy in this respect is based on exploiting the energy-growing effects associated to the higher-dimensional EFT operators, which translate into an enhanced sensitivity to the

$\hat{W}$ and $\hat{Y}$ parameters for large values of the dilepton invariant mass $m_{\ell\ell}$. To this purpose, it is useful to define the following ratio:

$$R_{\chi^2}\left(m_{\ell\ell}^{(\mathrm{max})}, \hat{W}, \hat{Y}\right) \equiv \frac{\chi^2\left(m_{\ell\ell}^{(\mathrm{max})}, \hat{W}, \hat{Y}\right)}{\chi^2\left(m_{\ell\ell}^{(\mathrm{max})} = 120 \text{ GeV}, \hat{W}, \hat{Y}\right)}, \qquad (4.5)$$

where $m_{\ell\ell}^{(\mathrm{max})}$ is the upper bound on the value of the dilepton invariant mass bins that enter the χ^2 calculation. In eq. (4.5), both the numerator and the denominator are evaluated using either χ^2_{smp} (for the SM PDFs) or χ^2_{eftp} (for the SMEFT PDFs), and the denominator corresponds to the χ^2 value (per data point) in the kinematic region for which EFT effects are negligible.[1]

The R_{χ^2} estimator defined in eq. (4.5) allows for the isolation of the contribution to the total χ^2 that arises from the high-$m_{\ell\ell}$ bins that dominate the overall sensitivity to the $\hat{W}$ and $\hat{Y}$ parameters. For small values of $m_{\ell\ell}^{(\mathrm{max})}$, say 200 GeV, one is cutting away all $m_{\ell\ell}$ bins with EFT sensitivity and hence one expects $R_{\chi^2} \simeq 1$. As $m_{\ell\ell}^{(\mathrm{max})}$ is increased, the χ^2 will include the contributions from the $m_{\ell\ell}$ bins more sensitive to EFT effects, and thus one expects to find a large deviation with respect to the $R_{\chi^2} \simeq 1$ reference value. Furthermore, EFT effects should induce an approximately monotonic growth of R_{χ^2} with $m_{\ell\ell}^{(\mathrm{max})}$, which would instead be absent from other possible sources of PDF distortion and thus represent a smoking gun for BSM physics in the high-energy DY tails.

These expectations are verified in figure 7, which displays the R_{χ^2} estimator (normalised to its SM value) as a function for $m_{\ell\ell}^{(\mathrm{max})}$ for representative values of the $\hat{W}$ and $\hat{Y}$ parameters both for the SM and the SMEFT PDFs, where the horizontal line indicates its reference SM value. Indeed we observe an approximately monotonic growth of R_{χ^2} arising from the energy-growing effects in the EFT. Due to the limited experimental information the binning in $m_{\ell\ell}$ is rather coarse, explaining the observed fluctuations. In the specific case of the $\hat{W}$ parameter, the SMEFT PDF curve lies slightly below the SM PDF one, highlighting how EFT effects are being partially (but not completely) reabsorbed into the PDFs.

4.2 EFT constraints on scenario II from current high-mass Drell-Yan data

In contrast to benchmark scenario I, which is flavour universal, the second SMEFT scenario to be explored in this work and described in section 2.2 contains a four-fermion interaction involving muons but not electrons, which therefore modifies the rates of the dilepton process $pp \to \mu^+\mu^-$ but not those of $pp \to e^+e^-$. This property implies that, without introducing further assumptions, the Wilson coefficient $\mathbf{C}_{33}^{D\mu}$ can be only constrained from DY measurements carried out in the dimuon (rather than in the dielectron or in the combined) final state. As indicated in table 2, only the CMS data at 7 TeV and 13 TeV include DY distributions in the dimuon final state.

Due to these restrictions in the input dataset, the interplay between PDFs and SMEFT effects is expected to be milder as compared to the results presented in section 4.1. For this

---

[1]Note that eq. (4.5) is computed *a posteriori* using existing fits, and that the kinematical cut in $m_{\ell\ell}^{(\mathrm{max})}$ is absent from the actual fits and it is only evaluated as a diagnosis tool.
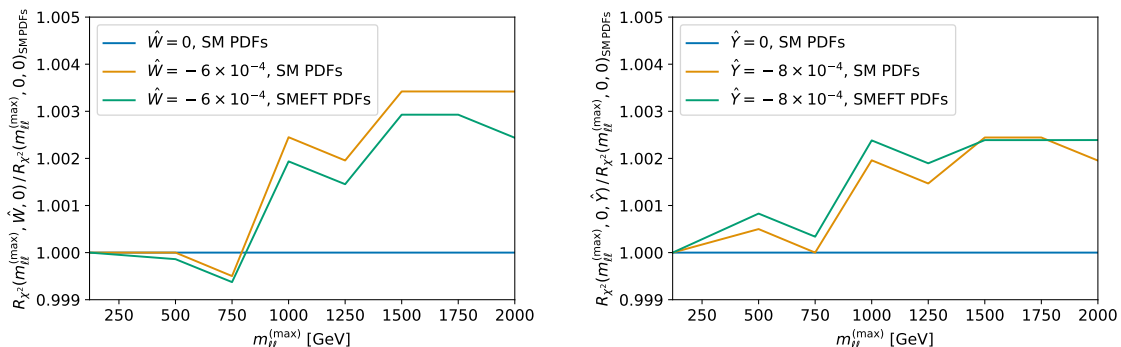
Figure 7. The R_{χ^2} estimator, eq. (4.5), normalised to its SM value, as a function for $m_{\ell\ell}^{(\mathrm{max})}$ for representative values of $\hat{W}$ (left) and $\hat{Y}$ (right panel). We display the results obtained both with SM and SMEFT PDFs, with the horizontal line indicating the reference SM value of R_{χ^2}.

reason, here we do not attempt to perform a joint determination of the PDFs and the $\mathbf{C}_{33}^{D\mu}$ coefficient, but rather restrict ourselves to quantifying the information that available DY data in the dimuon final state provide on this operator. We instead present a simultaneous determination including projections for the HL-LHC in section 5.

Figure 8 displays the results of three quartic fits to the $\chi^2\left(\mathbf{C}_{33}^{D\mu}\right)$ profile in benchmark scenario II, based on eq. (3.13), where here χ^2_{smp} includes only the contributions from the two available DY measurements in the dimuon final state. We present fits based on cross sections that account only for the linear, only for the quadratic, and for both the linear and quadratic terms in the EFT expansion. In all cases, these cross sections are computed using the baseline SM PDF set. The inset displays the outcome of the linear EFT fit with an enlarged x-axis range.

The results of figure 8 indicate that $\mathbf{C}_{33}^{D\mu}$ is essentially unconstrained at the linear EFT level, and only once quadratic corrections $\mathcal{O}(\Lambda^{-4})$ are accounted for is one able to obtain reasonable bounds on this coefficient. The reason for this behaviour is that for this operator the interference with the SM amplitude is suppressed, and hence the leading EFT effects arise at the quadratic level from the square of the EFT amplitude, thus being proportional to $\left(\mathbf{C}_{33}^{D\mu}\right)^2$ [27]. In the case of the polynomial fit to the χ^2 profile evaluated on the full quadratic EFT cross sections, we find the following 95% CL limits on this Wilson coefficient:

$$\left(\mathbf{C}_{33}^{D\mu} \times 10^2\right) \in [-1.2, 10.7] \,, \tag{4.6}$$

which can be compared with the bounds on the same operator obtained in [27] from recasting the ATLAS dilepton search data of [93], given by eq. (2.9). The fact that our bound in eq. (4.6) is around a factor three looser than in eq. (2.9) is explained because the dilepton search data from [93] benefits from an extended coverage in $m_{\ell\ell}$ as compared to the available unfolded DY cross sections. The same result, this time for the $\hat{W}$ and $\hat{Y}$ parameters, will be obtained in the next section where we assess the impact of the SMEFT PDFs in the EFT interpretation of the ATLAS dilepton search dataset.
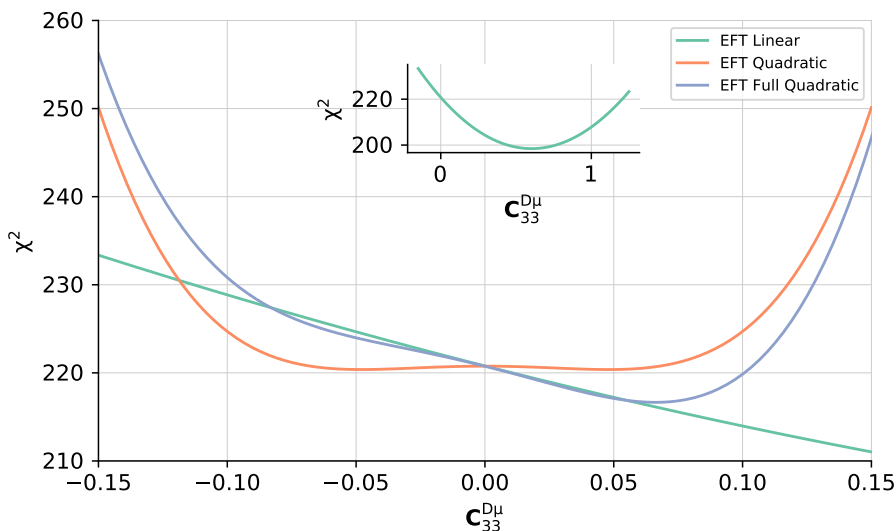
Figure 8. The results of polynomial fits to $\chi^2\left(\mathbf{C}_{33}^{D\mu}\right)$, eq. (4.2), in scenario II. This χ^2 includes only the contributions from the two DY measurements in the dimuon final state. We display results for fits based on cross sections that account only for the linear, only for the quadratic, and for both linear and quadratic terms in the EFT expansion, in all cases using the baseline SM PDF set. The inset displays the fit to the linear EFT values with an enlarged x-axis range.

4.3 On the EFT interpretation of high-mass dilepton searches

As mentioned above, a single high-mass DY cross section measurement is available at 13 TeV, and even in this case it is only based on a small subset of the Run II luminosity. As a consequence, the highest energy bin of this dataset is rather wide, $m_{\ell\ell} \in [1.5, 3.0]$ TeV. This implies limited sensitivity to deviations in the tails of DY distributions, for which using a large number of narrow bins is most beneficial to constrain heavy resonances, for instance. Here we would like to quantify the interplay between PDF and EFT effects at the level of a recent ATLAS 13 TeV search for Z' bosons in the dilepton channel [139] based on the complete Run II luminosity of $\mathcal{L} = 139\,\text{fb}^{-1}$. Since these are detector-level measurements, which cannot therefore be included in a PDF analysis, our aim is to use the SMEFT PDFs to investigate how the bounds on BSM physics are modified as compared to the standard approach based on computing theory predictions using SM PDFs.

The ATLAS data, displayed in figure 9, consist of event counts in 100 dilepton invariant mass, $m_{\ell\ell}$, bins in both the dimuon and dielectron channels in the range $m_{\ell\ell} \in (225, 6000)\,\text{GeV}$. We take this data from `HEPdata` [140] and denote the event count in the ith bin by n_i. The narrow binning and broad $m_{\ell\ell}$ coverage allowed ATLAS to constrain Z' masses to $M_{Z'} \gtrsim 4\,\text{TeV}$. This is a much higher reach than the DY cross section measurements used for the SMEFT PDF fits in the previous subsections, and should therefore provide stronger constraints on the EFT benchmark scenarios described in section 2. By including this search data in our study, we can investigate whether such strong constraints are sensitive to the EFT-induced modifications in the PDF luminosities highlighted in figure 6.
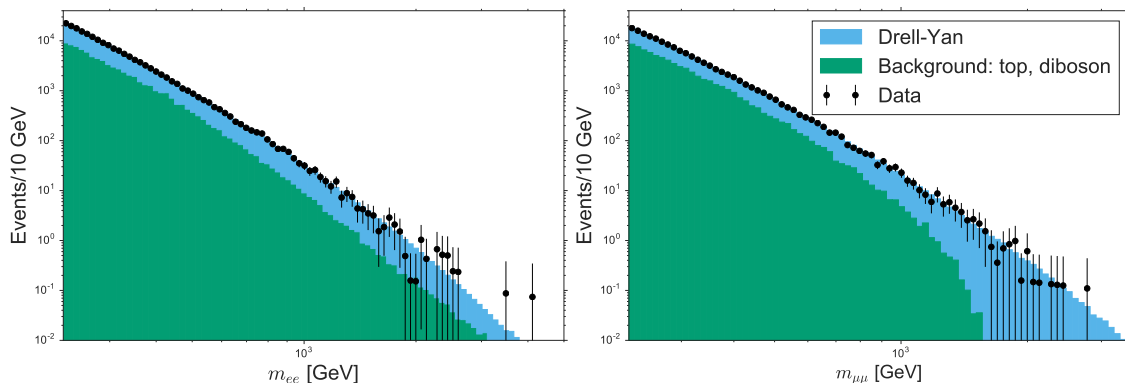
Figure 9. The data (number of events per 10 GeV bin) from the ATLAS Z' search from [139] in the di-electron (left) and di-muon (right) channels. We also display the theoretical predictions associated to the contributions from Drell-Yan and from the rest of the backgrounds, taken from the ATLAS publication.

In order to constrain the $\hat{W}$ and $\hat{Y}$ parameters in benchmark scenario I from the ATLAS dilepton search data, for each bin we compute a theory prediction $y_i = y_i(\hat{W}, \hat{Y})$ given by the sum of background b_i (top, diboson) and signal $s_i(\hat{W}, \hat{Y})$ (Drell-Yan) components. The ATLAS search provides an estimate of the total SM contribution (sum of top, diboson, and DY) without a breakdown into components. This estimate is provided as a continuous function of $m_{\ell\ell}$. We thus estimate our background (top and diboson) by subtracting our own DY simulation from the estimated total SM event counts found by evaluating this function at each bin centre. We compute the DY signal in each bin as

$$s_i(\hat{W}, \hat{Y}) = s_{i,\text{SM}} \times K(\hat{W}, \hat{Y}), \tag{4.7}$$

where $s_{i,\text{SM}}$ indicates the detector-level prediction for the ith bin of the $m_{\ell\ell}$ distribution evaluated at NLO QCD using MadGraph5_aMC@NLO [141], Pythia [142] and Delphes [143] and using the NNPDF31_nnlo_as_0118 set as PDF input set. In eq. (4.7), $K(\hat{W}, \hat{Y})$ is the K-factor calculated as the ratio of cross sections in each bin, accounting for the impact of non-zero EFT corrections $\hat{W}, \hat{Y} \neq 0$ both in the partonic cross section and in the PDFs

$$K_{\text{eftp}}(\hat{W}, \hat{Y}) \equiv \frac{\sum_q \int d\tau \mathcal{L}_{q\bar{q}}^{\text{SMEFT}}(\tau, \mu_F, \hat{W}, \hat{Y})\hat{\sigma}(\tau s_0, \hat{W}, \hat{Y})}{\sum_q \int d\tau \mathcal{L}_{q\bar{q}}^{\text{SM}}(\tau, \mu_F)\hat{\sigma}(\tau s_0, 0, 0)}, \tag{4.8}$$

where the integration in τ goes from $\tau_{\min}$ to $\tau_{\max}$ in each bin. For comparison, we will also present results where the K-factor is instead evaluated as usual in terms of the SM PDFs,

$$K_{\text{smp}}(\hat{W}, \hat{Y}) \equiv \frac{\sum_q \int d\tau \mathcal{L}_{q\bar{q}}^{\text{SM}}(\tau, \mu_F)\hat{\sigma}(\tau s_0, \hat{W}, \hat{Y})}{\sum_q \int d\tau \mathcal{L}_{q\bar{q}}^{\text{SM}}(\tau, \mu_F)\hat{\sigma}(\tau s_0, 0, 0)}. \tag{4.9}$$

The likelihood $\mathcal{L}$ is defined as the product of Poisson probabilities in each bin,

$$\mathcal{L}(n|\hat{W},\hat{Y}) = \prod_{i=1}^{m} \frac{y_i(\hat{W},\hat{Y})^{n_i}}{n_i!} e^{-y_i(\hat{W},\hat{Y})} , \qquad (4.10)$$

and the best-fit values of $\hat{W}, \hat{Y}$ are determined as the maximum-likelihood estimates. The test statistics in the individual fits of $\hat{W}$ and $\hat{Y}$ are the profile likelihood ratios defined as

$$\lambda(\hat{W}) = -2\ln\left(\frac{\mathcal{L}(n|\hat{W},0)}{\mathcal{L}(n|\hat{W},0)_{\max}}\right) \qquad (4.11)$$
$$\lambda(\hat{Y}) = -2\ln\left(\frac{\mathcal{L}(n|0,\hat{Y})}{\mathcal{L}(n|0,\hat{Y})_{\max}}\right)$$

and follow a χ^2 distribution with $n_{\mathrm{dof}} = 1$. The 1σ and 2σ bounds are found by solving the implicit equations $\lambda(\hat{W}) = 1, 4$ respectively.

PDF uncertainties can be included by taking the confidence level intervals on the bounds given by the NNPDF Monte Carlo replicas, as it is explained in section 3.4.

The results for the fits obtained by using SM PDFs in the K-factors, eq. (4.9), compared to those obtained by using SMEFT PDFs, eq. (4.8), are displayed in figure 10, with the corresponding bounds being provided in table 5. We find that inclusion of PDF uncertainties has a much smaller impact on the parabolic fit than in the previous analysis and thus only the parabola including PDF uncertainties is displayed. Secondly, we observe that the shift in the bounds that one has using SM versus SMEFT PDFs is not entirely negligible. This finding indicates that it is important to use consistent PDFs determined with the same settings as the theoretical predictions in the partonic cross section. We can similarly recast the ATLAS search data to constrain the scenario II Wilson coefficient $\mathbf{C}_{33}^{D\mu}$, finding the following constraints at 95% CL:

$$\left(\mathbf{C}_{33}^{D\mu} \times 10^2\right) \in [-1.6, 2.4] , \qquad (4.12)$$

in good agreement with the results reported in eq. (4.6).

4.4 Overview of current constraints

In order to summarise the results obtained in this section, figure 11 displays the 95% CL bounds derived on the EFT parameters $\hat{W}$ and $\hat{Y}$ (in scenario I) and on $\mathbf{C}_{33}^{D\mu}$ (in scenario II), both from the high-mass DY cross section measurements (table 4) and from the ATLAS Z' search data (table 5). These bounds are shown in the case of theoretical calculations evaluated either with SM PDFs or with SMEFT PDFs, and in the former case we indicate the results that account for PDF uncertainties (these are included by construction for the SMEFT PDFs). To compare with previous works, we also display the bounds derived in [32] for the $\hat{W}$ and $\hat{Y}$ parameters from the ATLAS 8 TeV data and in [27] for the $\mathbf{C}_{33}^{D\mu}$ coefficient from the same ATLAS Z' search data.

As discussed above, our main findings are that the consistent simultaneous determination of the PDFs together with the EFT parameters leads to a moderate increase in the
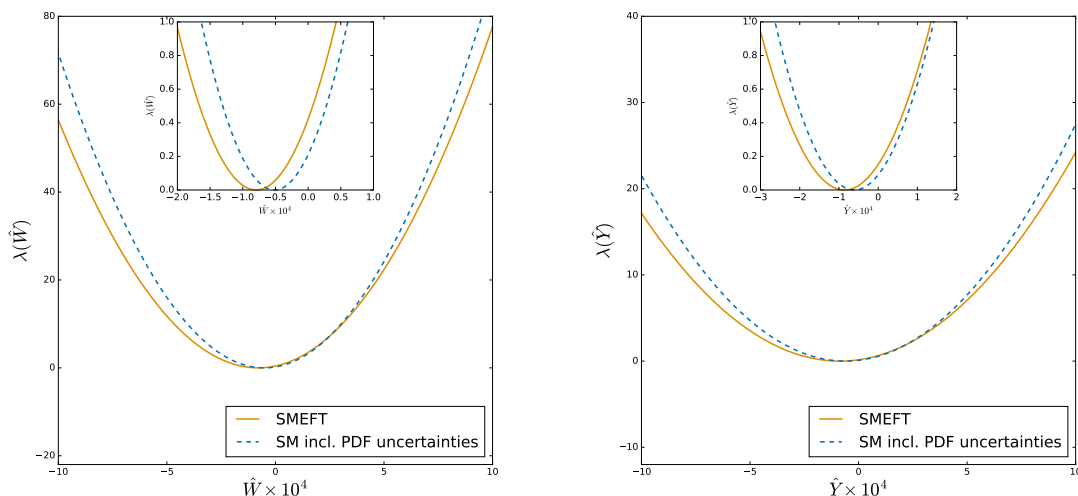
Figure 10. Comparison between the results of the parabolic fits to the ATLAS search data [139] for $\hat{W}$ (left) and $\hat{Y}$ (right panel) when using the SMEFT PDFs (χ^2_{eftp}) as compared to the SM PDF baseline (χ^2_{smp}). The insets zoom on the region close to $\lambda(\hat{W}), \lambda(\hat{Y}) \simeq 0$.

	SM PDFs	SMEFT PDFs	best-fit shift	broadening
$\hat{W} \times 10^4$ (68% CL)	$[-1.6, 0.6]$	$[-1.9, 0.5]$	-0.2	$+9\%$
	$[-1.6, 0.6]$		-0.2	$+9\%$
$\hat{W} \times 10^4$ (95% CL)	$[-2.7, 1.7]$	$[-3.1, 1.6]$	-0.25	$+7\%$
	$[-2.7, 1.7]$		-0.25	$+7\%$
$\hat{Y} \times 10^4$ (68% CL)	$[-2.6, 1.5]$	$[-3.1, 1.4]$	-0.3	$+10\%$
	$[-2.6, 1.6]$		-0.35	$+7\%$
$\hat{Y} \times 10^4$ (95% CL)	$[-4.7, 3.5]$	$[-5.3, 3.6]$	-0.25	$+9\%$
	$[-4.7, 3.6]$		-0.30	$+7\%$

Table 5. The 68% and 95% CL bounds on the $\hat{W}$ and $\hat{Y}$ parameters obtained from the corresponding parabolic fits to the ATLAS search data of [139] when using either the SMEFT PDFs or their SM counterparts. For the SM PDF results, we indicate the bounds obtained without (upper) and with (lower entry) PDF uncertainties accounted for; the SMEFT PDF bounds already include PDF uncertainties by construction. The fourth and fifth column indicate the absolute shift in best-fit values and the percentage broadening of the EFT parameter uncertainties when the SMEFT PDFs are consistently used instead of the SM PDFs.

Figure 11. Overview of the results obtained in this section concerning the EFT parameters $\hat{W}$ and $\hat{Y}$ (in scenario I) and $\mathbf{C}_{33}^{D\mu}$ (in scenario II). We compare the 95% CL bounds derived in [32] with those obtained in this work from the high-mass DY cross section measurements (table 4) and from the ATLAS Z' search data (table 5), in both cases displaying the results obtained with either the SM or the SMEFT PDFs. In the former case, we indicate the results that account for PDF uncertainties; these are included by construction for the SMEFT PDFs.

uncertainties (in this case, up to 10%) as well as to a small shift in their central values. As we demonstrate in the next section, the interplay between PDFs and EFT coefficients becomes much more marked in the case of the high-mass DY measurements that will become available at the HL-LHC.

5 Projections for the High-Luminosity LHC

The results presented in the previous section indicate that, given the available unfolded Drell-Yan measurements, the impact of a simultaneous determination of the PDFs together with the EFT parameters remains moderate. However, it is conceivable that this interplay between PDFs and BSM effects in the high-energy tails of Drell-Yan cross sections will become more significant once more data are accumulated. With this motivation, we revisit the analysis of section 4 now accounting for the impact of projected High-Luminosity LHC pseudo-data generated for the present study. We demonstrate that in the scenario under consideration, in which no other data apart from the high-mass Drell-Yan constrain the large-x quark and antiquark distributions, a consistent joint determination of PDFs is crucial for EFT studies at the HL-LHC. We will also discuss how the inclusion of further LHC data, which can constrain the large-x region without being affected by potential energy-growing new physics effects, can soften the interplay observed in this study and disentangle new-physics effects.

5.1 Generation of HL-LHC pseudo-data

Following the strategy adopted in [77] to estimate the ultimate PDF reach of the HL-LHC measurements (see also [144, 145]), here we generate HL-LHC pseudo-data for NC and CC high-mass Drell-Yan cross sections at $\sqrt{s} = 14$ TeV and for a total integrated luminosity of $\mathcal{L} = 6$ ab^{-1} (from the combination of ATLAS and CMS, which provide $\mathcal{L} = 3$ ab^{-1} each). For these projections, theoretical predictions are evaluated at NNLO in QCD including NLO EW corrections, as is explained in detail in section 3.2. The PDF set used as an input to generate the theoretical prediction is the DIS+DY baseline that was presented in section 3.3.

For the generation of the NC pseudo-data, we adopt as reference the CMS measurement at 13 TeV [122] based on $\mathcal{L} = 2.8$ fb^{-1}. The dilepton invariant mass distribution $m_{\ell\ell}$ is evaluated using the same selection and acceptance cuts of [122] but now with an extended binning in the $m_{\ell\ell}$ to account for the increase in luminosity. We assume equal cuts for electrons and muons and impose $|\eta_\ell| \leq 2.4$, $p_T^{\text{lead}} \geq 20$ GeV, and $p_T^{\text{sublead}} \geq 15$ GeV for the two leading charged leptons of the event. In the case of the CC pseudo-data, the lack of unfolded measurements of the m_T distribution at 13 TeV to be used as reference forces us to base our projections on the ATLAS search for W' bosons in the dilepton channel [146]. As in the case of the NC projections, theory predictions for the m_T distribution at high-mass are generated using the same selection and acceptance cuts as in [146] but now using an extended coverage in m_T.

Both in the case of NC and CC Drell-Yan cross sections, we restrict ourselves to events with either $m_{\ell\ell}$ or m_T greater than 500 GeV. Otherwise, the total experimental uncertainty would be limited by our modelling of the expected systematic errors and thus our projections could become unreliable. Furthermore, we require that the expected number of events per bin is bigger than 30 to ensure the applicability of Gaussian statistics. Taking into account these considerations, our choice of binning for the $m_{\ell\ell}$ (m_T) distribution at the HL-LHC is displayed in figure 12 (figure 13), with the highest energy bins reaching $m_{\ell\ell} \simeq 4$ TeV ($m_T \simeq 3.5$ TeV) for neutral-current (charged-current) scattering.

The percentage statistical and systematic uncertainties associated to the HL-LHC pseudo-data are displayed in the lower panels of figures 12 and 13 and have been estimated as follows. Let us denote by σ_i^{th} the theoretical prediction for the DY cross section, including all relevant selection cuts as well as the leptonic branching fractions. The expected number of events in this bin and the associated (relative) statistical uncertainty δ_i^{stat} are given by

$$N_i^{\text{th}} = \sigma_i^{\text{th}} \times \mathcal{L}, \quad \delta_i^{\text{stat}} \equiv \frac{(\delta N_i)_{\text{stat}}}{N_i^{\text{th}}} = \frac{1}{\sqrt{N_i^{\text{th}}}}. \tag{5.1}$$

Note that this bin-by-bin relative statistical uncertainty is the same both at the level of number of events and at the level of fiducial cross sections.

The HL-LHC systematic uncertainties are also estimated from the same reference measurements. If $\delta_{i,j}^{\text{sys}}$ denotes the jth relative systematic uncertainty associated to the ith bin of the reference measurement, and if this bin contains N_i^{th} events, then for our projections we assume that the same systematic error associated to a bin with a similar number of
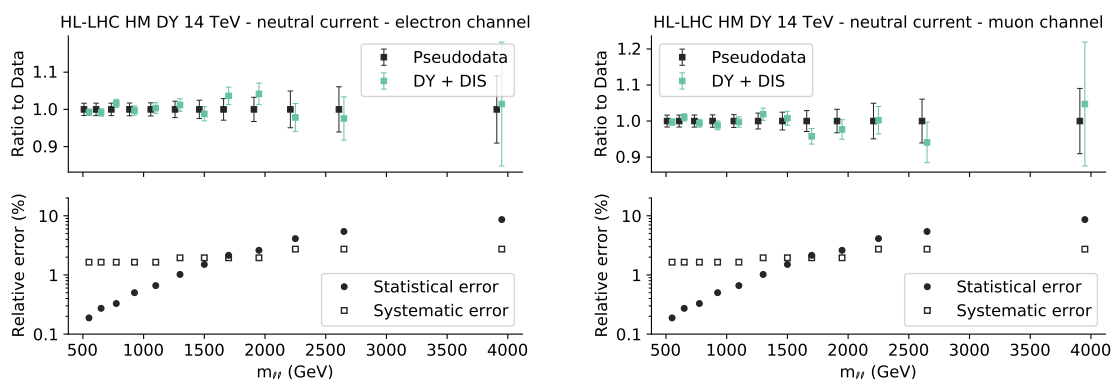
Figure 12. Top panels: comparison of the projected HL-LHC pseudo-data for high-mass neutral-current Drell-Yan in the dielectron (left) and dimuon (right) final states as a function of $m_{\ell\ell}$ with the corresponding theory predictions obtained from the SM PDF baseline. The theoretical predictions, generated according to eq. (5.3), are accompanied by their corresponding PDF uncertainties (green bars). Lower panels: the percentage statistical and systematic uncertainty in each $m_{\ell\ell}$ bin of the HL-LHC pseudo-data.

Figure 13. Same as figure 12 for charged-current Drell-Yan in bins of the transverse mass m_T.

expected events will be given by $f_{\mathrm{red},j}\delta_{i,j}^{\mathrm{sys}}$, where $f_{\mathrm{red},j}$ is the expected reduction in systematic errors foreseen at the HL-LHC.[2] This assumption is justified since most systematic errors improve with the sample size thanks to e.g. better calibration.

Adding in quadrature systematic uncertainties with the statistical error, the total relative uncertainty for the ith bin of our HL-LHC projections is

$$\delta_{\mathrm{tot},i}^{\mathrm{exp}} = \left(\left(\delta_i^{\mathrm{stat}}\right)^2 + \sum_{j=1}^{n_{\mathrm{sys}}} \left(f_{\mathrm{red},j}\delta_{i,j}^{\mathrm{sys}} \right)^2 \right)^{1/2}, \tag{5.2}$$

where n_{sys} indicates the number of systematic error sources. The central values for the HL-LHC pseudo-data is then generated by fluctuating the reference theory prediction by

---

[2]The binning of the CC reference measurement, the ATLAS W' search, is much finer than for our HL-LHC projections and hence we first match them by means of a weighted average.

the expected total experimental uncertainty, namely

$$\sigma_i^{\text{hllhc}} \equiv \sigma_i^{\text{th}} \left(1 + \lambda \delta_{\mathcal{L}}^{\text{exp}} + r_i \delta_{\text{tot},i}^{\text{exp}} \right), \qquad i = 1, \ldots, n_{\text{bin}}, \tag{5.3}$$

where λ, r_i are univariate Gaussian random numbers, $\delta_{\text{tot},i}^{\text{exp}}$ is the total (relative) experimental uncertainty corresponding to this specific bin (excluding the luminosity and normalisation uncertainties), and $\delta_{\mathcal{L}}^{\text{exp}}$ is the luminosity uncertainty, which is fully correlated amongst all the pseudo-data bins of the same experiment. We take this luminosity uncertainty to be $\delta_{\mathcal{L}}^{\text{exp}} = 1.5\%$ for both ATLAS and CMS, as done in ref. [77].

Here we adopt the baseline SM PDF set described in section 4, which is denoted as "DIS+DY", to evaluate the σ_i^{th} cross sections entering eq. (5.3). We have verified that, both at the pre- and post-fit levels, the fit quality to the HL-LHC pseudo-data satisfies $\chi^2/n_{\text{bin}} \simeq 1$ in the case of the SM PDFs as expected. Furthermore, we assume $f_{\text{red},j} = 0.2$ for all systematic sources, as done in the optimistic scenario of ref. [77]. We note that more conservative values for the reduction of systematic errors, such as $f_{\text{red},j} = 0.5$, are not expected to qualitatively modify our results. The reason is that, as indicated by the bottom panels of figures 12 and 13, for the highest energy bins (which dominate the EFT sensitivity), specifically above $m_{\ell\ell} \approx 1.7\,\text{TeV}$ and $m_T \approx 1.5\,\text{TeV}$, the measurement will be limited by statistical uncertainties.

5.2 Impact on PDF uncertainties

From figures 12 and 13, one can observe that the PDF uncertainties in the SM PDF baseline used to generate the pseudo-data are either comparable or larger than the corresponding projected experimental uncertainties at the HL-LHC. Specifically, for the highest $m_{\ell\ell}$ bin of the NC distribution the PDF errors are twice the experimental ones, while in the CC case the associated PDF errors become clearly larger than the experimental ones starting at $m_T \simeq 2\,\text{TeV}$. This comparison suggests that one should expect a significant uncertainty reduction once the HL-LHC pseudo-data is included in the PDF fit.

To validate this expectation, figure 14 displays the impact of the HL-LHC pseudo-data on the quark-antiquark luminosity $\mathcal{L}_{q\bar{q}}$ as a function of the final state invariant mass m_X at $\sqrt{s} = 14\,\text{TeV}$. We compare $\mathcal{L}_{q\bar{q}}$ for the SM PDF baseline fit (DIS+DY) with the same quantity from the corresponding fits including the HL-LHC pseudo-data, either only NC or also with CC cross sections. The right panel displays the associated relative PDF uncertainties. We find a significant reduction of the PDF uncertainties affecting the quark-antiquark luminosity (and hence the Drell-Yan cross sections) in the high mass ($m_X \gtrsim 1\,\text{TeV}$) region once the HL-LHC pseudo-data constraints are accounted for. For instance, at $m_X \gtrsim 2\,\text{TeV}$, PDF uncertainties on $\mathcal{L}_{q\bar{q}}$ decrease from $\simeq 5\%$ in the baseline down to $\simeq 2.5\%$ ($\simeq 1.5\%$) once the NC (NC+CC) HL-LHC pseudo-data is included in the fit. The effect of the inclusion of HL-LHC projections becomes more dramatic as m_X increases. On the other hand, other partonic luminosities such as the quark-quark and gluon-gluon ones are essentially unaffected by the HL-LHC constraints. In terms of fit quality, the only noticeable effect is a mild improvement in the χ^2 of the high-mass DY datasets listed in table 2.
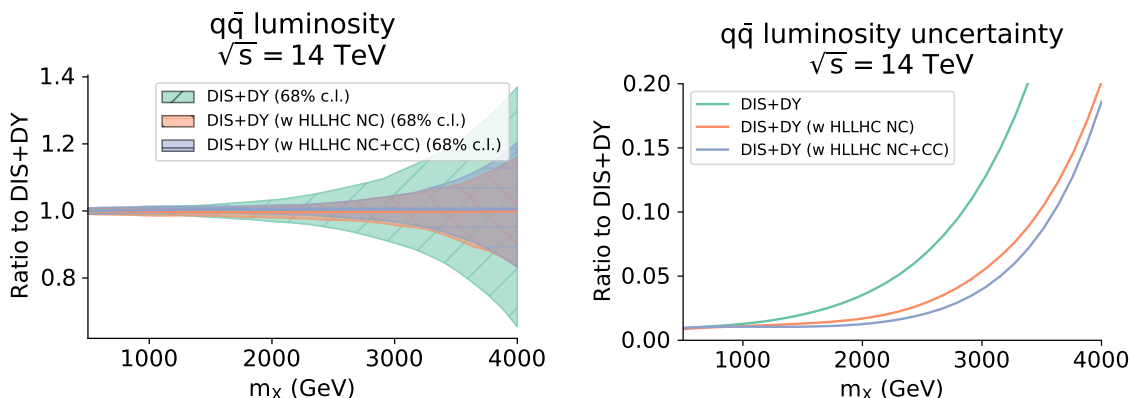
Figure 14. Impact of the HL-LHC pseudo-data on the quark-antiquark luminosity $\mathcal{L}_{q\bar{q}}$ of the SM PDF baseline fit as a function of m_X. Left: the luminosities $\mathcal{L}_{q\bar{q}}$ for the DIS+DY baseline and the corresponding fits including the HL-LHC pseudo-data, either only NC or also with CC cross sections, presented as a ratio to the central value of the former. Right: the relative PDF uncertainty in $\mathcal{L}_{q\bar{q}}$ (with the central value of the DIS+DY baseline as reference) for the same fits.

5.3 PDF and EFT interplay at the HL-LHC

The finding that the projected HL-LHC pseudo-data has a significant impact on the quark-antiquark PDF luminosity, summarised in figure 14, suggests that the interplay between PDFs and EFT effects in the high-energy DY tails should become enhanced as compared to the results reported in the previous section. With this motivation, we first of all repeat the joint determination of PDFs and the $\hat{W}, \hat{Y}$ coefficients from EFT scenario I presented in section 4.1 now accounting for the constraints of the HL-LHC pseudo-data. An important difference in this case is that the inclusion of CC data lifts the flat direction in the $(\hat{W}, \hat{Y})$ plane, making a full two-dimensional fit possible. Secondly, the availability of the HL-LHC pseudo-data allows us to assess the interplay between the PDFs and the EFT coefficient $\mathbf{C}_{33}^{D\mu}$ from benchmark scenario II, whose analysis in section 4.2 was restricted to fixed SM PDFs.

Scenario I. For the simultaneous determination of PDFs and the $\hat{W}, \hat{Y}$ coefficients accounting for the constraints provided by the HL-LHC pseudo-data, we use 35 sampling values of $(\hat{W}_i, \hat{Y}_i)$, 25 of which are equally spaced in either $\hat{W} \in (-1.6, 1.6) \times 10^{-5}$ or $\hat{Y} \in (-8, +8) \times 10^{-5}$ (hence in steps of $\Delta\hat{W} = 0.8 \times 10^{-6}$ and $\Delta\hat{Y} = 4 \times 10^{-6}$ respectively), and then 10 additional points along the diagonals. In order to assess the robustness of the results, we added 12 more sampling values, 8 further away from the origin and 4 more along the $\hat{W} = 0$ and $\hat{Y} = 0$ axes, and verified that the confidence level countours are stable upon their addition.

We find that the constraints on the $(\hat{W}, \hat{Y})$ parameters are completely dominated by the HL-LHC projections and that current data exhibit a much smaller pull, consistent with the findings of previous studies [32, 37]. Also, the χ^2_{eftp} contour is more stable and requires less replicas if only the HL-LHC projections are included in the computation of the χ^2.

	SM PDFs	SMEFT PDFs	best-fit shift	broadening
$\hat{W} \times 10^5$ (68% CL)	$[-0.7, 0.5]$	$[-4.5, 6.9]$	1.3	850%
	$[-1.0, 0.9]$		1.3	500%
$\hat{W} \times 10^5$ (95% CL)	$[-1.0, 0.8]$	$[-8.1, 10.6]$	1.4	940%
	$[-1.4, 1.2]$		1.4	620%
$\hat{Y} \times 10^5$ (68% CL)	$[-1.8, 3.2]$	$[-6.4, 8.0]$	0.1	190%
	$[-3.7, 4.7]$		0.3	70%
$\hat{Y} \times 10^5$ (95% CL)	$[-3.4, 4.7]$	$[-11.1, 12.6]$	0.1	190%
	$[-5.3, 6.3]$		0.3	110%

Table 6. Same as table 4 for the 68% CL and 95% CL marginalised bounds on the $\hat{W}$ and $\hat{Y}$ parameters obtained from the two-dimensional $(\hat{W},\hat{Y})$ fits that include the HL-LHC pseudo-data for NC and CC Drell-Yan distributions. As in table 4, for the SM PDFs we indicate the bounds obtained without (upper) and with (lower entry) PDF uncertainties accounted for.

The corresponding marginalised bounds on $\hat{W}$ and $\hat{Y}$ are reported in table 6 using the same format as in table 4.

From table 6, one can observe how including high-mass data at the LHC both in a fit of PDFs and in a fit of SMEFT coefficients and neglecting the interplay between them could result in a significant underestimate of the uncertainties associated to the EFT parameters. Indeed, the marginalised 95% CL bound on the $\hat{W}$ ($\hat{Y}$) parameter becomes looser once SMEFT PDFs are consistently used, with a broadening, defined in eq. (4.4), of 500% (110%), even once PDF uncertainties are fully accounted for. This effect would have been even more marked if PDF uncertainties had not been accounted for in EFT fits based on SM PDFs, where the same broadening factors would be 940% and 190% respectively.

A further important question is whether the bounds obtained with SM PDFs appearing on the left column of table 6 would become more comparable to those obtained from the simultaneous fit of PDFs and SMEFT coefficients, in case a conservative set of PDF was used in the analysis based on SM PDFs. To address this question, in table 7 we display the bounds that are obtained using a PDF set that does not include any of the high-mass Drell-Yan sets (neither the HL-LHC projections nor the current datasets listed in table. 2) and compare the bounds obtained using this set of PDFs to those obtained consistently using SMEFT PDFs. We observe that, once this set of conservative PDF is used as an input PDF set and the PDF uncertainty is included in the computation of the bounds, the latter increases as compared to the bounds in table 6. As a result, the size of the bounds obtained by keeping fixed SM PDFs is closer to the size obtained from the simultaneous fits, although still slightly underestimated. At the same time, the shift in the best-fit becomes more marked.

	SM cons. PDFs	SMEFT PDFs	best-fit shift	broadening
$\hat{W} \times 10^5$ (68% CL)	$[-1.0, 0.0]$	$[-4.5, 6.9]$	1.7	1000%
	$[-4.0, 2.8]$		1.8	70%
$\hat{W} \times 10^5$ (95% CL)	$[-1.4, 0.4]$	$[-8.1, 10.6]$	1.8	940%
	$[-4.3, 3.1]$		1.9	150%
$\hat{Y} \times 10^5$ (68% CL)	$[2.1, 7.0]$	$[-6.4, 8.0]$	-3.7	190%
	$[-3.4, 11.2]$		-3.6	-1%
$\hat{Y} \times 10^5$ (95% CL)	$[0.5, 8.5]$	$[-11.1, 12.6]$	-3.7	200%
	$[-5.0, 13.7]$		-3.6	30%

Table 7. Same as table 6 for the 68% and 95% CL marginalised bounds on the $\hat{W}$ and $\hat{Y}$ parameters obtained from the two-dimensional $(\hat{W}, \hat{Y})$ fits that include the HL-LHC pseudo-data for NC and CC Drell-Yan distributions. The input PDF set for the analysis done using fixed SM PDFs (corresponding to the results displayed in the column "SM cons. PDFs") is a conservative PDF set that does not include any of the high-mass distributions or the HL-LHC projections nor the Run I and Run II high-mass dataset listed in table 2. The limits obtained from the simultaneous fit of PDFs and Wilson coefficients (corresponding to the results displayed on the column "SMEFT PDFs") are the same as those in table 6.

Results are graphically displayed in figure 15, where the 95% confidence level contours in the $(\hat{W}, \hat{Y})$ plane obtained from the DIS+DY fits that include the high-mass Drell-Yan HL-LHC pseudo-data when using either SM PDFs, SM conservative PDFs or SMEFT PDFs are compared. All solid countours include PDF uncertainties, while the dashed contours that do not include PDF uncertainties are also indicated to visualise the impact of the inclusion of the PDF uncertainties.

To conclude, we should also emphasise that, while in this work we use pseudo-data and hence the best-fit values are by construction unchanged, this would not necessarily be the case in the analysis of real data, where improper treatment of PDFs could result in a spurious EFT 'signal', or even missing a signal which is indeed present in the data. A detailed study aimed at a precise definition of 'conservative' PDFs in a more general scenario is beyond the scope of this paper and will be the topic of future work; a thorough comparison of the consistent simultaneous approach, versus the use of conservative PDF sets, will be of particular interest in cases of EFT manifestations of new physics.

The increased role that the interplay between PDFs and EFT coefficients will play at the HL-LHC can also be illustrated by comparing the expected behaviour of the quark-antiquark luminosity, displayed in figure 16, for the SMEFT PDFs corresponding to representative values of the $\hat{W}$ and $\hat{Y}$ parameters of benchmark scenario I as compared to the SM PDFs. Note that the corresponding comparison for $\mathcal{L}_{q\bar{q}}$ in the fits to available Drell-Yan data was displayed in figure 6. Indeed, the central value of the quark-antiquark luminosity for SMEFT PDFs corresponding to values of $(\hat{W}_i, \hat{Y}_i)$ selected along the grid used to derive
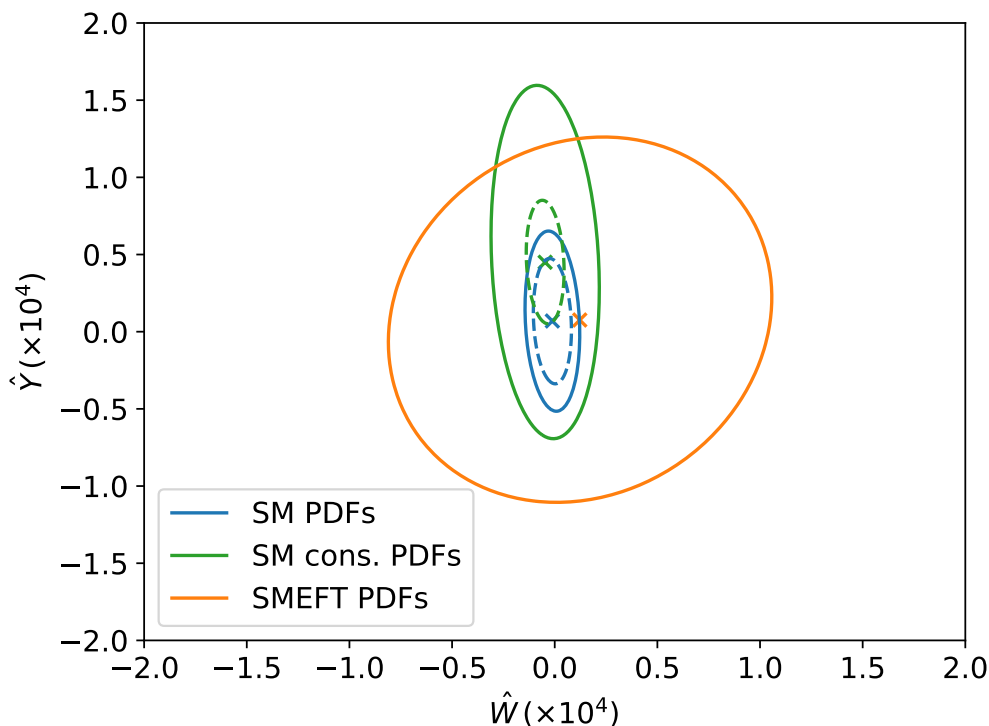
Figure 15. The 95% confidence level contours in the $(\hat{W},\hat{Y})$ plane obtained from the DIS+DY fits that include the high-mass Drell-Yan HL-LHC pseudo-data (both in the NC and CC channels) when using either SM PDFs (blue) or conservative SM PDFs (green). In both cases the ellipses are obtained by performing a parabolic fit to $\chi^2_{\rm smp}$ with fixed PDFs. PDF uncertainties are included in the solid lines and not included in the dashed lines. The results are compared to those obtained in a simultaneous fit, namely with SMEFT PDFs (orange). In this case, the parabolic fit is performed to $\chi^2_{\rm eftp}$ by varying simultaneously the Wilson Coefficients and the PDFs. The crosses indicate the best fits in the three cases discussed in the text.

figure 16 changes greatly, well outside the one-sigma error band of the SM PDFs, while the PDF uncertainties themselves are unchanged. This change in central value of the large-x PDFs partially reabsorbs the effects in the partonic cross section induced by the SMEFT operators and leads to better χ^2 values as compared to those obtained with the SM PDFs.

Even neglecting SMEFT PDF effects, we note that our marginalised bounds on the $\hat{W}$ and $\hat{Y}$ coefficients from HL-LHC pseudo-data using SM PDFs turn out to be more stringent than those reported in [37] by around a factor of 4 for $\hat{W}$ and a factor 2 for ($\hat{Y}$). This is due to a combination of factors. First of all we use the 13 TeV measurements as reference to produce the HL projections. Furthermore we assume a total integrated luminosity of $\mathcal{L} = 6\,{\rm fb}^{-1}$ (from the combination of ATLAS and CMS) rather than $3\,{\rm fb}^{-1}$ as well as a more optimistic scenario concerning the reduction of the experimental systematic uncertainties.

Figure 17 then displays the R_{χ^2} estimator, defined in eq. (4.5) and shown in figure 7 for the case of available LHC data, now evaluated from the fits including the HL-LHC pseudo-data. In this case, the $m^{(\rm max)}_{\ell\ell}$ cut applies to $m_{\ell\ell}$ for the neutral-current distributions and
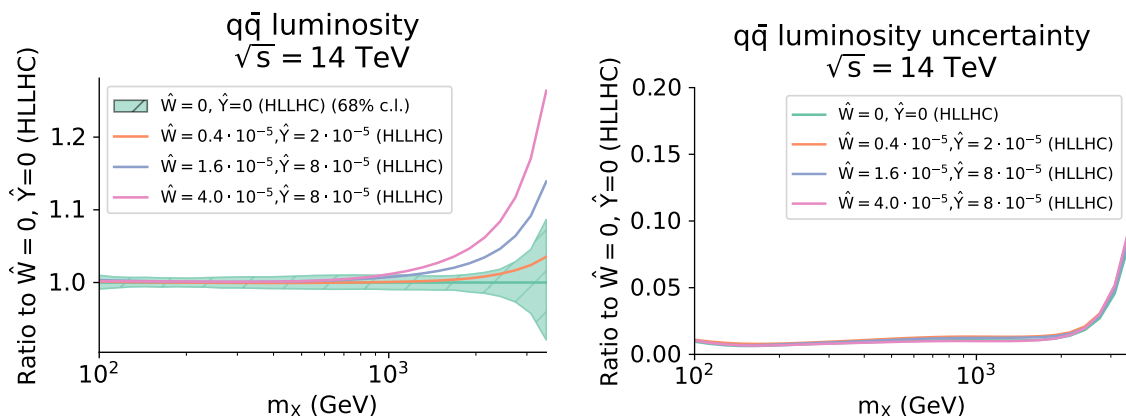
Figure 16. Same as figure 14, now comparing the quark-antiquark SM PDF luminosity in the fits including the HL-LHC pseudo-data with those obtained in the SMEFT PDF fits for representative values of the $\hat{W}$ and $\hat{Y}$ parameters. The corresponding comparison in the case of fits to available Drell-Yan data was shown in figure 6.
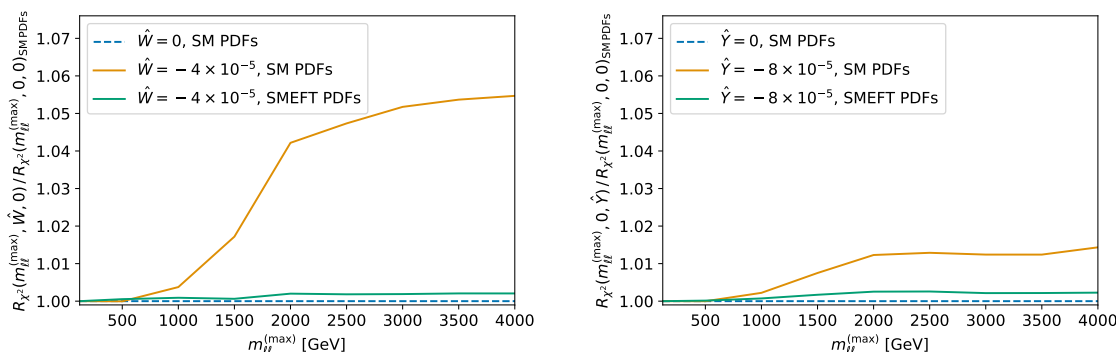

Figure 17. Same as figure 7 now for the fits including the HL-LHC pseudo-data. Note that in this case the $m_{\ell\ell}^{(\mathrm{max})}$ cut refers to the transverse mass m_T for the charged-current distributions.

to the transverse mass m_T for the charged-current ones. As in the case of figure 7, we observe an approximately monotonic growth of R_{χ^2} for the SM PDFs arising from the energy-growing EFT corrections that dominate the high-energy DY tails, an effect which is now rather larger thanks to the presence of the HL-LHC pseudo-data. The most striking difference as compared to figure 7 is that now the SMEFT PDF curve is much flatter, indicating that EFT effects are being almost totally reabsorbed into the PDFs.

The findings summarised by figure 17 demonstrate that, at the HL-LHC, EFT-induced deviations could be indeed inadvertently "fitted away" into a PDF redefinition, explaining the large broadenings reported in figure 16, and highlight the need to devise novel strategies to disentangle the effects of PDFs and EFT contributions from the high-energy tails of LHC cross-sections. Such strategies could exploit, for instance, the availability of measurements sensitive to large-x PDFs but not to high scales, such as forward electroweak gauge boson production by LHCb [77].
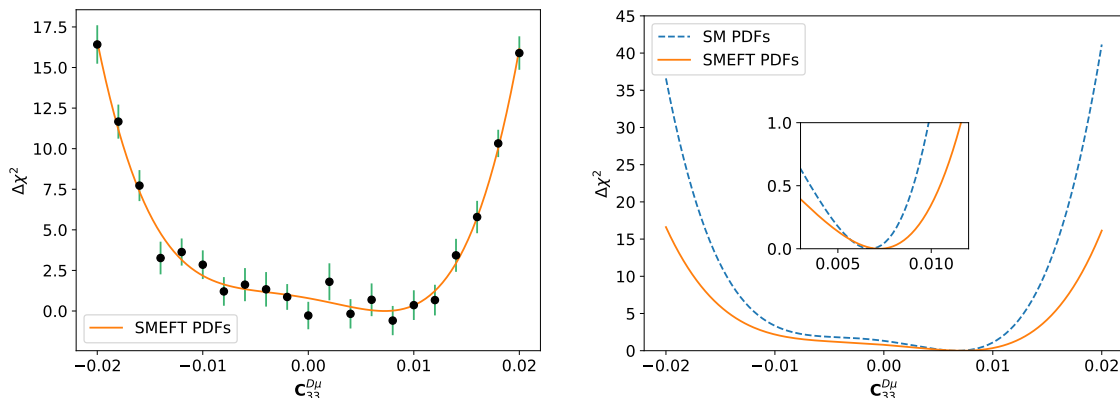
Figure 18. Left: the values of $\Delta\chi^2$ obtained for the SMEFT PDFs as a function of $\mathbf{C}_{33}^{D\mu}$ from the fits including the HL-LHC pseudo-data, together with the corresponding quartic polynomial fit. Right: comparison of the polynomial fit obtained with SMEFT PDFs and displayed in the left panel with its counterpart based on SM PDFs.

Scenario II. We now turn to present the corresponding results of the simultaneous fits of the PDFs and the EFT coefficients including the HL-LHC pseudo-data for the case of benchmark scenario II. As motivated in section 2.2, a non-zero value of the $\mathbf{C}_{33}^{D\mu}$ coefficient affects only the NC and CC muon final states, while the electron ones remain described by the SM calculations. This property implies that, in fits presented below, the EFT corrections modify only the shapes of the HL-LHC distributions in the muon channel, the right panels in figures 12 and 13, but not those of the electron pseudo-data.

Figure 18 displays the values of $\Delta\chi^2$ obtained for the SMEFT PDFs as a function of the EFT parameter $\mathbf{C}_{33}^{D\mu}$ from the joint fits that include the HL-LHC pseudo-data. The sampling is constituted by 21 points uniformly distributed in $\mathbf{C}_{33}^{D\mu} \in [-0.02, 0.02]$. As in figure 4, the error bars indicate the uncertainties associated to the finite number of Monte Carlo replicas used for each value of $\mathbf{C}_{33}^{D\mu}$. The profile in $\Delta\chi^2$ exhibits a double minimum structure (bimodal distribution), explained by the fact that in this scenario it is the quadratic rather than the linear terms in the EFT expansion that dominate. The corresponding quartic polynomial fit using eq. (4.2) can be seen to successfully reproduce the $\Delta\chi^2$ values obtained in this joint analysis. The right panel of figure 18 then compares the polynomial fit obtained with the SMEFT PDFs with the corresponding one when using instead fixed SM PDFs to determine the $\Delta\chi^2$ values, with the inset focusing on the region close to $\Delta\chi^2 \simeq 0$. The associated 68% and 95% CL bounds are then reported in table 8, where we note that since the 68% CL interval is disjoint we evaluate the shift and broadening only for the 95% CL bounds.

Inspection of figure 18 and table 8 indicates that, even at the HL-LHC, the interplay between PDFs and EFT coefficients remains moderate in this particular scenario. Indeed, in contrast with the marked effects in scenario I (figure 15), where the bounds on the $\hat{W}$ and $\hat{Y}$ worsened by up to an order of magnitude when the SMEFT PDFs were consistently used, in scenario II the obtained bounds on $\mathbf{C}_{33}^{D\mu}$ would only loosen by around 30%. The

	SM PDFs	SMEFT PDFs	best-fit shift	broadening
$\mathbf{C}_{33}^{D\mu} \times 10^2$ (68% CL)	$[-0.1, 1.1]$	$[-0.3, 1.2]$	0.06	25%
$\mathbf{C}_{33}^{D\mu} \times 10^2$ (95% CL)	$[-1.0, 1.2]$	$[-1.2, 1.4]$	0.06	18%

Table 8. Same as table 6, now for the $\mathbf{C}_{33}^{D\mu}$ parameter from EFT benchmark scenario II.
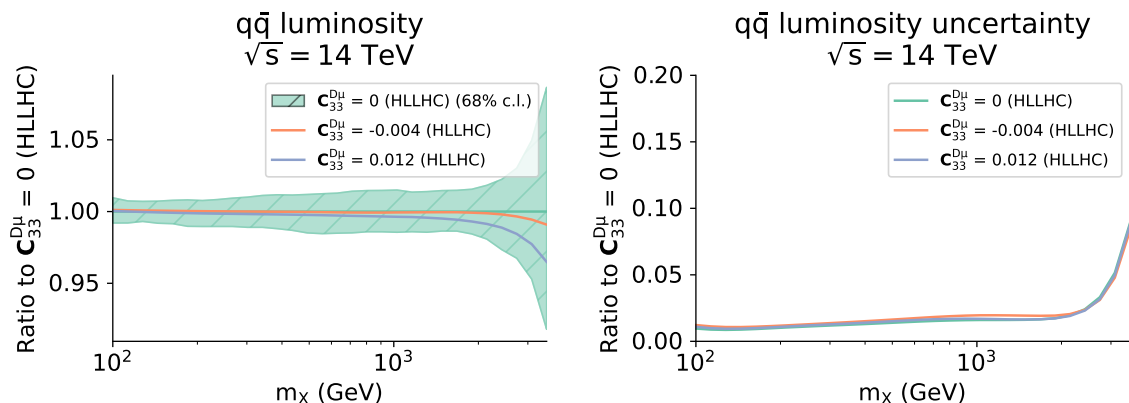

Figure 19. Same as figure 16 now in the case of the simultaneous fits of the PDFs and the $\mathbf{C}_{33}^{D\mu}$ EFT parameter taking into account HL-LHC pseudo-data.

origin of this rather different behaviour can be traced back to the fact that in scenario II the electron channel data do not receive EFT corrections, and hence all the information that they provide makes it possible to exclusively constrain the PDFs. The muon channel distributions then determine the allowed range for $\mathbf{C}_{33}^{D\mu}$, restricted by the well-constrained large-x quarks and antiquark PDFs from the electron data. This finding demonstrates how the availability of measurements in separate leptonic final states is of utmost importance to test BSM scenarios that account for violations of Lepton Flavour Universality.

In the same manner as in figure 16, figure 19 displays the comparison of the quark-antiquark luminosities at $\sqrt{s} = 14$ TeV in the fits with HL-LHC pseudo-data in the case of the SM PDFs and for the SMEFT PDFs for representative values of $\mathbf{C}_{33}^{D\mu}$. Specifically, we show $\mathbf{C}_{33}^{D\mu} = -0.004$ and 0.012, chosen to lie at the boundary of the 68% CL interval reported in table 8. The result that the two values lead to the same effect on $\mathcal{L}_{q\bar{q}}$ follows from the dominance of the quadratic EFT terms in this scenario. One finds that the shift in the central values of the quark-antiquark luminosity induced by a non-zero value of $\mathbf{C}_{33}^{D\mu}$ is well within PDF uncertainties. This is consistent with the result of figure 18 indicating that bounds on $\mathbf{C}_{33}^{D\mu}$ obtained with SM and with SMEFT PDFs are relatively similar in this scenario even after accounting for the HL-LHC constraints.

6 Conclusions and outlook

Indirect searches for new physics beyond the SM, such as those carried out in the SMEFT framework, often aim at pinning down subtle distortions with respect to the SM predictions,

such as a few-percent deviation in the value of a given production cross section or decay rate. Exploiting the full potential of current and future precision measurements at the LHC for these indirect BSM searches requires the development of novel data interpretation frameworks that are able to account for hitherto ignored effects that can no longer be neglected. A pressing example of this is the interplay between PDF and EFT effects in the high-energy tails of LHC distributions. Indeed, the very same datasets are being used both to determine the parton distributions (assuming SM cross sections) and, independently, to constrain EFT coefficients (assuming SM PDFs). Given that these LHC processes provide significant information for both PDF and EFT fits, it is of paramount importance to ascertain the extent for which eventual EFT signals can be reabsorbed into the PDFs, as well as how current bounds on the EFT coefficients are modified within a consistent simultaneous determination together with the PDFs.

In this work, building upon our previous DIS-only study [72], we have presented a first simultaneous determination of PDFs and EFT coefficients from high-energy LHC data, specifically from high-mass Drell-Yan cross sections. Our analysis has considered available unfolded measurements, detector-level searches based on the full Run II luminosity, and tailored HL-LHC projections. The EFT interpretation of the Drell-Yan data is formulated in terms of two benchmark scenarios, first a flavour universal one leading to modifications of the $\hat{W}$ and $\hat{Y}$ electroweak parameters [32], and second a flavour-specific scenario motivated by the recent evidence for lepton flavour universality violation in B-meson decays [27].

The main findings of this work are summarised in figure 20. We demonstrate how, for the analysis of all available unfolded Drell-Yan data, the consistent simultaneous extraction of the PDFs together with the EFT parameters leads to a modest increase in the uncertainties of the latter (up to 15%, in the case of the $\hat{W}$ and $\hat{Y}$ parameters), as well as to a shift in their central values by up to a third of a sigma. Furthermore, while our results indicate that for current data the interplay between PDF and EFT effects remains moderate, the impact of their cross-talk will become much larger at the HL-LHC: using SM rather than SMEFT PDFs would lead to artificially precise bounds, even mimicking new physics effects. This result indicates that including high-energy data in PDF fits should be done with care, as PDFs can actually absorbe the effects of new physics. On the other hand, we have seen that in this simple case, using a conservative set of PDFs that does not include any of the high-mass Drell-Yan data and accounting for the large contribution of the PDF uncertainty on the bounds inflates them and makes them of the same order of magnitude as those obtained in a simultaneous fit of PDFs and SMEFT coefficients.

At the same time, once real data at HL-LHC are considered, neglecting the PDF interplay and simply using conservative sets of PDFs might miss EFT manifestations of new physics or misinterpret them. One should also emphasise that estimators such as those shown in figure 5.6 only become available in the joint PDF+EFT fit, and cannot be defined in the "conservative PDF" approach. In particular, they provide information on the kinematic dependence of any possible deviation between the data and the SM predictions, and the extent to which this can be reabsorbed into the PDFs. Hence, they represent a powerful diagnosic tool to separate QCD effects from genuine BSM deviations. A complementary strategy to disentangle QCD effects from BSM effects would be to account for the con-
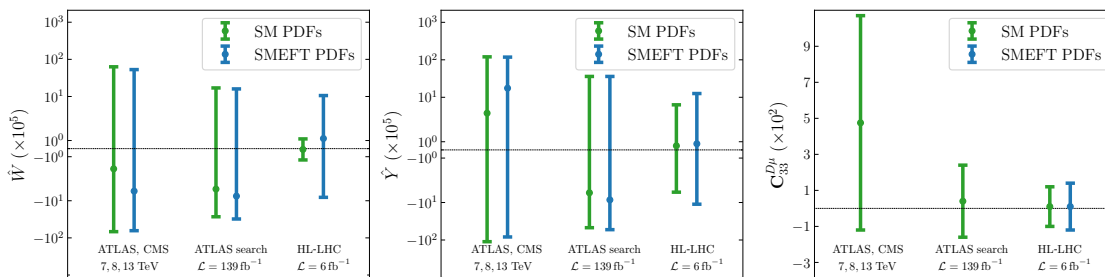
Figure 20. Overview of the 95% CL bounds obtained in this work the EFT parameters $\hat{W}$ (left), $\hat{Y}$ (middle), and $\mathbf{C}_{33}^{D\mu}$ (right panel) based on either the SM PDFs of the SMEFT PDFs. Both PDF and methodological uncertainties are accounted in the bounds, when available.

straints on the large-x PDFs arising from other processes for which EFT corrections can be neglected, such as forward W, Z production at LHCb. This way the uncertainties associated to the PDFs at large-x would be reduced and the indirect signal for new physics could be more easily disentangled. A detailed study aimed at a definition of conservative PDFs in a more general scenario is beyond the scope of this paper, and will be the topic of future work.

Furthermore, concerning the next steps in this program, it would be interesting to consider other more general EFT benchmark scenarios, as well as accounting for new 13 TeV measurements based on increased luminosity. One could also envisage complementing the EFT analysis of inclusive DY cross sections with related processes, such as dilepton production in association with extra jets, which provides sensitivity to different combinations of dimension-six operators.

In addition to increasing the dataset, one would need to address a major challenge of the current fitting methodology for the joint determination of PDFs and EFT coefficients, namely that for each point in the EFT parameter space one needs to carry out a full-fledged NNPDF fit, which requires intensive computing resources. To tame the instability of the χ^2, once all datasets are kept into account, one needs to run a very large number of replicas. Furthermore, in the simultaneous fit of EFT coefficients and PDFs, we currently ignore the correlation between the two sets of parameters. The kind of analysis presented in this work could then be streamlined and made more efficient by adapting the fitting methodology to exploit the relatively simple dependence of hadronic cross sections on the EFT coefficients, which is either linear at $\mathcal{O}(\Lambda^{-2})$ or quadratic at $\mathcal{O}(\Lambda^{-4})$. Such methodological developments would make it possible to simultaneously fit the PDFs with a large number of EFT coefficients, something that is currently infeasible but that is being developed.

Finally, beyond single gauge boson production, it would also be important to ascertain the interplay that arises between PDFs and EFT effects for the interpretation of gluon-dominated LHC processes, such as top-quark pair production and inclusive jet and dijet (or even multijet) production. The reason is that these two groups of processes have been shown to provide crucial information for, on the one hand, pinning down the gluon PDF over a broad range of x values [147, 148], and on the other hand, constraining a large number of EFT dimension-six operators [50, 55, 56, 149] which cannot be accessed by

other probes. Given that in modern global PDF analyses the gluon for $x \simeq 10^{-2}$ is almost entirely determined by these and related high-energy LHC processes, it is conceivable that the PDF and EFT interplay there could be more significant than for the inclusive DY processes studied in this work, even just accounting for the already available measurements.

Acknowledgments

A Detailed SM PDF comparisons

In this appendix we present detailed comparisons between different sets of SM PDFs to complement the discussions in section 3. To begin with, we compare the baseline SM PDF of this work, based on DIS+DY data, with the recent global `NNPDF3.1_str` fit obtained in the context of the proton strangeness study of [95]. Figure 21 is the counterpart of figure 3, now displaying the gluon, singlet, up, anti-up, down, and anti-down quark PDFs at $Q = 100 \, \text{GeV}$ both for the baseline SM PDF (labelled "DIS+DY") and for the global `NNPDF3.1_str` determination.

We observe an overall good compatibility between our DIS+DY baseline and the `NNPDF3.1_str` global fit, with PDFs in agreement at the one-sigma level in all cases except for the quark singlet Σ in the region $0.01 \lesssim x \lesssim 0.1$. As discussed in section 3.3, the new high-mass DY data included in this analysis as compared to [95] are responsible for the bulk of the differences observed in figure 21, both in terms of central values and uncertainties, for the quark and anti-quark PDFs. Specifically, the upwards shift in the central values of the quark and anti-quark PDFs in this x-region for the DIS+DY baseline as compared to the `NNPDF3.1_str` determination is consistent with the comparisons in figure 3 illustrating the
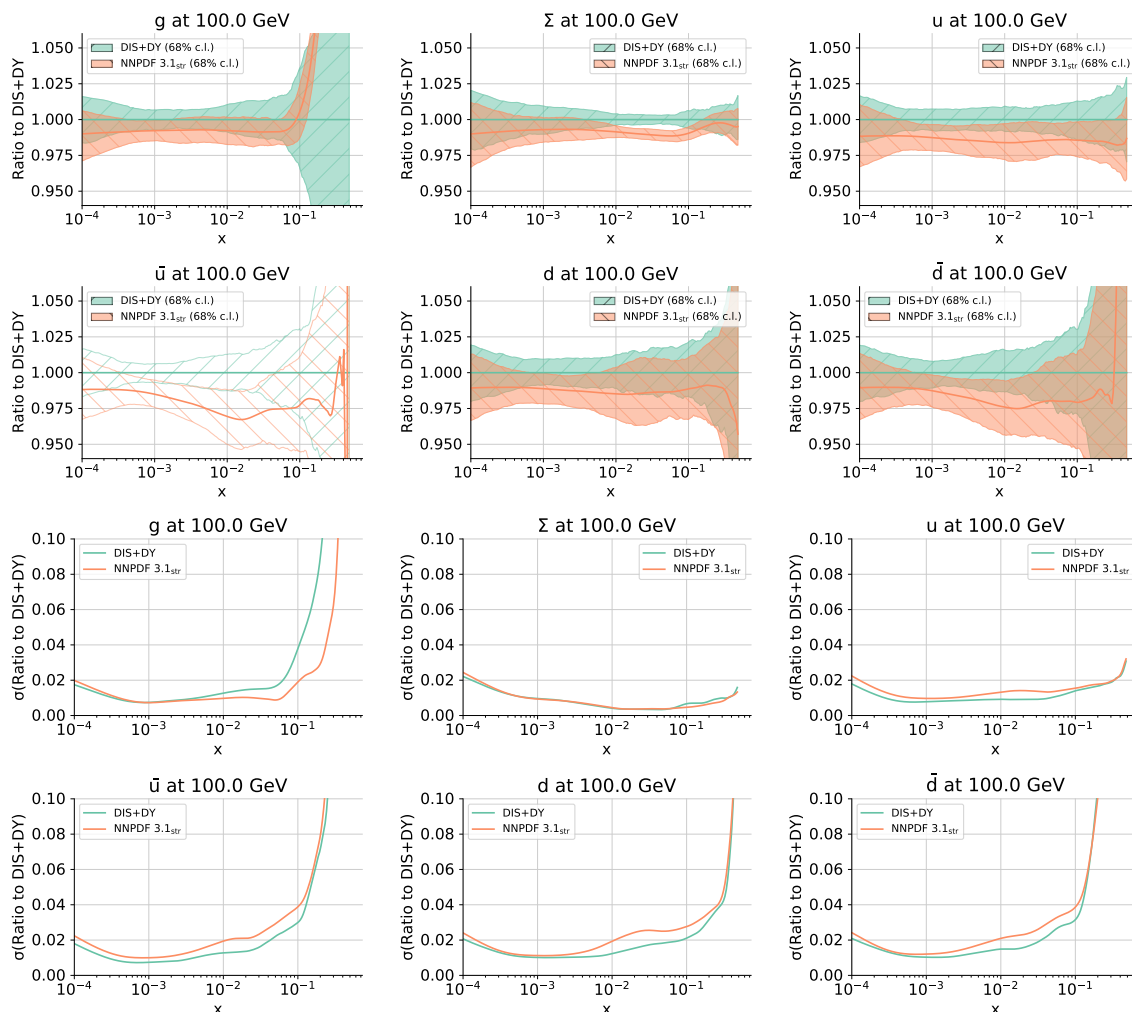
Figure 21. Same as figure 3 comparing the baseline SM PDF set used in this work (labelled "DIS+DY") with the global `NNPDF3.1_str` determination.

impact of the high-mass Drell-Yan data in the fit, and the same applies for the associated reduction of the quark and anti-quark PDF uncertainties.

As is well known, the PDF uncertainties on the gluon become rather enlarged in the DIS+DY baseline due to the lack of information from the top and jet cross sections. However, this does not impact the results of the present joint PDF and EFT interpretation, given that gluon-induced contributions to inclusive Drell-Yan processes enter only starting at NLO. Furthermore, we also find somewhat larger uncertainties in the strangeness of the DIS+DY baseline as compared to `NNPDF3.1_str` due to the missing constraints from the NOMAD neutrino dimuon cross sections. All in all, with the exception of gluon-initiated processes, we can conclude that the DIS+DY baseline to be used in this work is competitive with a full-fledged global PDF determination.

Next, we display in figure 22 the corresponding comparison between the baseline SM PDF set based on DIS and DY data (dubbed "DIS+DY") with the same fit but only in-
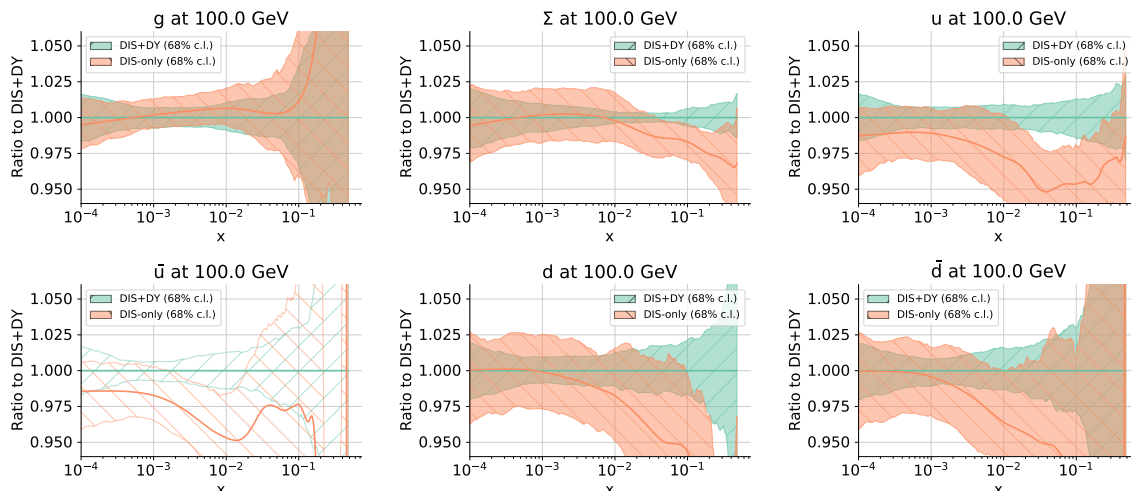
Figure 22. Same as the upper panels of figure 3 comparing now the baseline SM PDF set used in this work (labelled "DIS+DY") with the corresponding DIS-only fit. Note that the comparison between the PDF uncertainties in these two fits was already displayed in the lower panels of figure 3.

cluding DIS structure functions. Note that the comparison between the PDF uncertainties in these two fits was already displayed in the lower panels of figure 3. One can observe how in general there is excellent consistency between the two fits. Indeed, PDFs are in agreement at the one-sigma level except for very specific cases, such as the up quark PDF at $x \simeq 0.05$, but even there the differences are at most at the 1.5σ level. The very marked reduction of PDF errors is also appreciable in the DIS+DY fit as compared to the DIS-only fit, highlighting the importance of the DY data in the global PDF fit to constrain the light quark and antiquark PDFs in a broad range of x.

Finally, in figure 23 we compare the PDF luminosities in the DIS+DY baseline with those from the same fit excluding all the data of the high-mass DY datasets listed in table 2. The corresponding comparisons at the PDF level was shown in figure 3

We focus on the gluon-gluon, quark-antiquark, and quark-quark luminosities at $\sqrt{s} = 14\,\mathrm{TeV}$ as a function of the invariant mass m_X of the produced final state, and display both the luminosity ratio to the reference as well as the relative PDF uncertainties in each case.

Again, one finds that the high-mass DY measurements constrain the luminosities in the range $100\,\mathrm{GeV} \lesssim m_X \lesssim 2\,\mathrm{TeV}$, consistent with the kinematic coverage in $m_{\ell\ell}$ of the data used in the fit. Their main effects are a reduction of the $q\bar{q}$ uncertainty for m_X between $500\,\mathrm{GeV}$ and $2\,\mathrm{TeV}$ and an upwards (downwards) shift in the central values of the $q\bar{q}$ (gg) luminosities within this m_X region. Th uncertainty of $\mathcal{L}_{qq}$ is barely changed and its central value i shifted within 1σ PDF uncertainties once the high-mass Drell-Yan datasets of table. 2 are included in the fit. This comparison further highlights how the high-mass DY data provide useful information for constraining the PDF luminosities and in turn the high-p_T processes relevant for both direct and indirect BSM searches at the LHC.
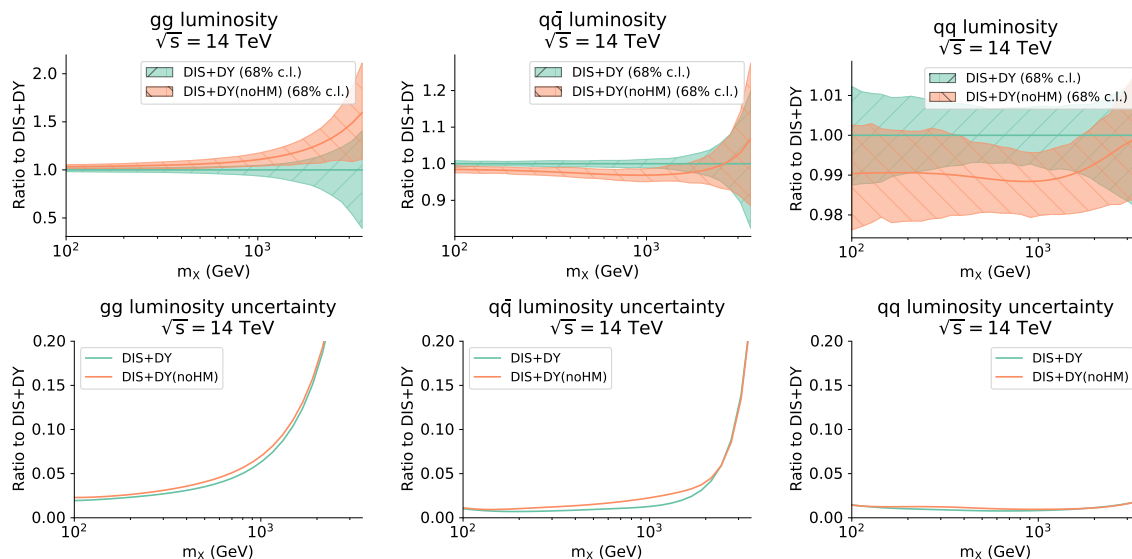
Figure 23. The gluon-gluon, quark-antiquark, and quark-quark luminosities at $\sqrt{s} = 14 \, \text{TeV}$ as a function of m_X for the DIS+DY baseline and for the same fit excluding the all the datapoints in the high-mass DY experimental sets listed in table 2. The top panels display the ratio of luminosities to the central value of the DIS+DY baseline, while the bottom panels compare the relative PDF uncertainties in each case.

Dataset	NNPDF3.1	This work
ATLAS 7 TeV	$m_{ee} \leq 200 \, \text{GeV}$	no cuts
ATLAS 8 TeV	not included	$y_{\ell\ell} < 1.68$ (for $m_{\ell\ell} \geq 1000 \, \text{GeV}$)
CMS 7 TeV	$m_{\mu\mu} \leq 200 \, \text{GeV}$	$y_{\mu\mu} \leq 2.2$ (for all $m_{\mu\mu}$)
	$y_{\mu\mu} \leq 2.2$	$y_{\mu\mu} < 1.5$ (for $m_{\mu\mu} \geq 850 \, \text{GeV}$)
CMS 8 TeV	not included	no cuts
CMS 13 TeV	not included	no cuts

Table 9. The kinematic cuts applied to the high-mass Drell-Yan datasets listed in table 2, compared to those used in NNPDF3.1.

B Fit quality for SM and SMEFT PDFs

In this appendix, we provide detailed information about the PDF fit quality as quantified by the figure of merit used in the fits: the χ^2 per data point, defined in eq. (3.5) and evaluated with the t_0 prescription described in ref. [134]. We will do this both for the SM PDFs based on different datasets and for the SMEFT PDFs from the fits with the baseline dataset and for different values of the EFT parameters $\hat{W}$ and $\hat{Y}$.

For completeness, we also provide here the χ^2 values obtained when using `NNPDF3.1_str` as the input PDF set, with all other settings such as the partonic matrix elements unchanged. Note that here the kinematical cuts are slightly different as compared to [95], the differences being summarised in table 9. The rationale behind having different

cuts is that in this work we include electroweak corrections to the high-mass DY cross sections, thus the $m_{\ell\ell} \leq 200\,\text{GeV}$ restriction applied in NNPDF3.1 is not necessary anymore. With the current cuts, essentially all high-mass DY data points can be included in the fits. The exception is a subset of points from the CMS 7 TeV and ATLAS 8 TeV datasets, where we restrict ourselves to the tree-level kinematic condition $|y_{\ell\ell}| \leq \ln(\sqrt{s}/m_{\ell\ell})$. The reason is that our calculation of the EFT corrections is based on tree-level SM cross sections which must satisfy this requirement. Furthermore, for the CMS 7 TeV dataset the last rapidity bin is excluded for all $m_{\mu\mu}$ bins, since it is found to deviate from the SM predictions by a large amount suggesting the need to account for threshold resummation effects [150]. Henceforth, here we evaluate the predictions based on `NNPDF3.1_str` for the same set of kinematical cuts as in this work.

Table 10 summarises the values of the χ^2 for the baseline SM PDF fit labelled "DIS+DY" compared to the most recent NNPDF global fit `NNPDF3.1_str`, as well as for the corresponding fits based on reduced datasets, namely the DIS-only fit and the fit excluding the high-mass DY data. The entries in italic indicate the datasets that do not enter the corresponding fit.

We observe that the quality of the description of the DIS data is similar across all fits considered. As far as hadronic data are concerned, we observe that the fit quality of the LHCb data slightly deteriorates when the high-mass Drell-Yan data are included. Also, the description of the CMS 13 TeV invariant mass distribution in the combined electron and muon channels is not optimal, even after including the data in the fit. However, the overall χ^2 is statistically equivalent to the most recent NNPDF3.1 set.

Then in table 11 we list again the χ^2 values in the SM PDF fit (same as the "DIS+DY" column of table 10) and compare them with those obtained with the SMEFT PDFs for the same representative values of $\hat{W}$ and $\hat{Y}$ parameters as used in section 4.2, see also the PDF-level comparisons in figure 31. Clearly the theoretical predictions are computed consistently, namely the partonic cross sections of the SM baseline "DIS+DY" are computed in the SM, while the partonic cross section of the other columns are augmented by the SMEFT contributions of the corresponding operators.

First of all, we observe that as expected the addition of the EFT corrections does not affect the description of the DIS structure functions. Differences are also small for the low-mass and on-shell DY data, and slightly larger for the HM measurements. For instance, the χ^2 to the high-mass Drell-Yan datasets is 1.471 for $\hat{W} = 0.0006$ to be compared with 1.478 for the SM PDFs. In any case, the differences at the level of χ^2 between the SM and SMEFT PDFs are reasonably small, consistent with the finding that the best-fit values of the $\hat{W}$ and $\hat{Y}$ parameters are close to the SM expectation.

C   Validation of the SMEFT K-factors

As described in section 3.2, in this work the effect of the dimension-six SMEFT operators considered in the two benchmark scenarios is accounted for at the level of cross sections via the K-factor approach, eq. (3.10). In this appendix, we provide further details about the calculation and validation of these EFT K-factors. Specifically, we compare the numerical

Dataset	n_{dat}	χ^2/n_{dat} (SM PDFs)			
		This work			Reference
		DIS-only	DIS+DY(noHM)	DIS+DY	NNPDF3.1_str
SLAC	67	1.032	0.807	0.780	0.772
BCDMS	581	1.150	1.222	1.230	1.229
NMC	325	1.320	1.347	1.378	1.346
CHORUS	832	1.058	1.188	1.228	1.191
NuTeV	76	0.796	0.642	0.684	0.703
HERA inclusive	1145	1.238	1.250	1.242	1.264
HERA charm	37	1.654	1.433	1.445	1.424
HERA bottom	29	1.304	1.328	1.326	1.343
Total DIS	**3092**	**1.172**	**1.217**	**1.230**	**1.225**
E886 $\sigma_{\mathrm{DY}}^d/\sigma_{\mathrm{DY}}^p$	15	*49.94*	0.484	0.484	0.509
E886 σ_{DY}^p	89	*1.306*	1.061	1.094	1.064
E605 σ_{DY}^p	85	*2.682*	0.972	0.982	1.006
CDF $d\sigma_Z/dy_Z$	29	*1.796*	1.443	1.460	1.459
D0 $d\sigma_Z/dy_Z$	28	*0.650*	0.595	0.602	0.594
D0 $W \to \mu\nu$ asy.	9	*6.729*	1.411	1.488	1.582
ATLAS W, Z 2010	30	*1.353*	0.817	0.866	0.846
ATLAS low-mass $Z \to ee$	6	*1.038*	0.985	0.949	0.995
ATLAS W, Z 2011	61	*6.077*	1.704	1.681	1.760
ATLAS $W + c$ rapidity	22	*0.497*	0.469	0.468	0.487
ATLAS $Z\, p_T$	92	*1.110*	0.989	0.942	1.029
ATLAS $p_{T,W}$ in W+jets	32	*2.074*	1.574	1.690	1.567
CMS W asy.	22	*5.362*	1.291	1.287	1.292
CMS σ_{W+c} 7 TeV	5	*0.555*	0.495	0.478	0.505
CMS $\sigma_{W^++c}/\sigma_{W^-+c}$ 7 TeV	5	*2.526*	1.826	1.687	1.710
CMS $Z\, p_T$	28	*1.289*	1.336	1.296	1.354
CMS $W \to \mu\nu$ rapidity	22	*5.022*	1.006	1.070	1.077
CMS $W + c$ rapidity 13 TeV	5	*0.638*	0.661	0.658	0.671
LHCb $Z \to \mu\mu$	9	*2.440*	1.630	1.652	1.676
LHCb $W, Z \to \mu$ 7 TeV	29	*13.62*	2.032	2.209	2.136
LHCb $Z \to ee$	17	*1.273*	1.118	1.124	1.114
LHCb $W, Z \to \mu$ 8 TeV	30	*8.835*	1.496	1.769	1.475
Total DY (excl. HM)	**670**	***4.185***	**1.166**	**1.191**	**1.193**
ATLAS DY high-mass 7 TeV	13	*2.261*	*2.014*	1.885	1.945
ATLAS DY high-mass 8 TeV	46	*1.393*	*1.227*	1.181	*1.215*
CMS DY high-mass 7 TeV	117	*1.603*	*1.617*	1.589	1.584
CMS DY high-mass 8 TeV	41	*0.796*	*0.891*	0.805	*0.838*
CMS DY high-mass 13 TeV	43	*1.837*	*1.981*	2.013	*1.952*
Total DY (HM-only)	**260**	***1.510***	***1.514***	**1.478**	***1.480***
Total	**4022**	**1.733**	**1.258**	**1.243**	**1.266**

Table 10. The values of the χ^2 per data point for the baseline SM PDF fit, labelled "DIS+DY", and for the corresponding fits based on reduced datasets. Here eq. (3.5) is evaluated using the t_0 prescription. We also include the results obtained using NNPDF3.1_str with the kinematic cuts used in this work and summarised in table 9. Values in italics indicate datasets that do not enter the corresponding fit.

Dataset	n_{dat}	χ^2/n_{dat}				
		SM	$\hat{W} =$ 0.0006	$\hat{W} =$ -0.0008	$\hat{Y} =$ 0.0012	$\hat{Y} =$ -0.0006
SLAC	67	0.780	0.806	0.798	0.763	0.833
BCDMS	581	1.230	1.223	1.224	1.228	1.227
NMC	325	1.378	1.349	1.346	1.377	1.364
CHORUS	832	1.228	1.233	1.223	1.230	1.237
NuTeV	76	0.684	0.698	0.651	0.706	0.655
HERA inclusive	1145	1.242	1.247	1.244	1.243	1.245
HERA charm	37	1.445	1.440	1.431	1.433	1.416
HERA bottom	29	1.326	1.324	1.314	1.330	1.323
Total DIS	**3092**	**1.230**	**1.229**	**1.224**	**1.230**	**1.231**
E886 $\sigma_{\text{DY}}^d/\sigma_{\text{DY}}^p$	15	0.484	0.454	0.475	0.513	0.458
E886 σ_{DY}^p	89	1.094	1.067	1.060	1.075	1.049
E605 σ_{DY}^p	85	0.982	1.025	1.013	1.005	1.032
CDF $d\sigma_Z/dy_Z$	29	1.460	1.446	1.452	1.459	1.463
D0 $d\sigma_Z/dy_Z$	28	0.602	0.604	0.600	0.602	0.603
D0 $W \to \mu\nu$ asy.	9	1.488	1.464	1.517	1.507	1.536
ATLAS W, Z 2010	30	0.866	0.875	0.869	0.868	0.863
ATLAS low-mass $Z \to ee$	6	0.949	0.960	0.964	0.949	0.952
ATLAS W, Z 2011 CC	46	1.837	1.921	1.899	1.893	1.864
ATLAS W, Z 2011 CF	15	1.254	1.264	1.269	1.260	1.253
ATLAS $W + c$ rapidity	22	0.468	0.470	0.445	0.478	0.460
ATLAS $Z\, p_T$	92	0.942	0.973	0.967	0.951	0.978
ATLAS $p_{T,W}$ in W+jets	32	1.690	1.675	1.737	1.640	1.697
CMS $W\, e$ asy.	11	0.838	0.829	0.827	0.856	0.825
CMS $W\, \mu$ asy.	11	1.736	1.747	1.746	1.760	1.735
CMS σ_{W+c} 7 TeV	5	0.478	0.483	0.487	0.471	0.501
CMS $\sigma_{W^++c}/\sigma_{W^-+c}$ 7 TeV	5	1.687	1.686	1.728	1.678	1.735
CMS $Z\, p_T$	28	1.296	1.280	1.276	1.293	1.273
CMS $W \to \mu\nu$ rapidity	22	1.070	1.054	1.088	1.031	1.048
CMS $W + c$ rapidity 13 TeV	5	0.658	0.671	0.680	0.654	0.695
LHCb $Z \to \mu\mu$	9	1.652	1.648	1.662	1.651	1.645
LHCb $W, Z \to \mu$	29	2.209	2.165	2.176	2.161	2.257
LHCb $Z \to ee$	17	1.124	1.139	1.129	1.130	1.143
LHCb $W, Z \to \mu$	30	1.769	1.773	1.748	1.741	1.823
Total DY (excl. HM)	**670**	**1.191**	**1.199**	**1.198**	**1.192**	**1.202**
ATLAS DY high-mass 7 TeV	13	1.885	2.090	1.643	2.127	1.781
ATLAS DY high-mass 8 TeV	46	1.181	1.230	1.243	1.247	1.181
CMS DY high-mass 7 TeV	117	1.589	1.585	1.553	1.598	1.577
CMS DY high-mass 8 TeV	41	0.805	0.793	0.833	0.820	0.787
CMS DY high-mass 13 TeV	43	2.013	1.876	2.203	1.881	2.064
Total DY (HM-only)	**260**	**1.478**	**1.471**	**1.497**	**1.486**	**1.473**
Total	**4022**	**1.243**	**1.244**	**1.242**	**1.245**	**1.247**

Table 11. Same as table 10, now comparing the χ^2 values (computed using the t_0 prescription, as above) of the SM PDFs with those of the SMEFT PDFs for different values of the $\hat{W}$ and $\hat{Y}$ parameters in benchmark scenario I, specifically those displayed in figures 6 and 31.

values for these K-factors, which have been obtained using `SMEFTsim` [85, 151] interfaced with MADGRAPH5_AMC@NLO, with the analytic calculation presented in [27].

DIS structure functions. SMEFT corrections to the neutral-current deep-inelastic structure functions F_2, F_3 in the benchmark scenario I of section 2 are obtained by means of a direct calculation in perturbation theory.[3] In order to determine these corrections, we rewrite eq. (2.5) as the linear combination of four-fermion operators of the form $\bar{q}_\lambda \gamma^\mu q_\lambda \bar{\ell}_{\lambda'} \gamma_\mu \ell_{\lambda'}$, where q_λ is a quark field of helicity λ (with $\lambda = +1$ for a right-handed field and $\lambda = -1$ for a left-handed field) and $\ell_{\lambda'}$ is a lepton field of helicity λ'. The relevant operators for the $\hat{Y}$ parameter are already of this form in eq. (2.5). For the $\hat{W}$ parameter, the associated operators can be expanded explicitly as:

$$\mathcal{L}_{\text{SMEFT}} \supset -\frac{g^2 \hat{W}}{4 m_W^2} \sum_{i=1}^{3} \left(\bar{e}_L^i \gamma^\mu e_L^i \bar{u}_L^i \gamma_\mu u_L^i - \bar{e}_L^i \gamma^\mu e_L^i \bar{d}_L^i \gamma_\mu d_L^i \right. \tag{C.1}$$
$$\left. - \bar{\nu}_L^i \gamma^\mu \nu_L^i \bar{u}_L^i \gamma_\mu u_L^i + \bar{\nu}_L^i \gamma^\mu \nu_L^i \bar{d}_L^i \gamma_\mu d_L^i \right),$$

where the index i runs over generations, and the flavour-changing contributions have been dropped since they only contribute to low energy CC structure functions.

The EFT corrections to the DIS structure functions induced by a specific four-fermion operator of the form

$$\mathcal{L}_{\text{SMEFT}} \supset \frac{c_{\lambda\lambda'}^{q\ell}}{\Lambda^2} \bar{q}_\lambda \gamma^\mu q_\lambda \bar{\ell}_{\lambda'} \gamma_\mu \ell_{\lambda'} \tag{C.2}$$

can be shown to be given by

$$\Delta F_2(x, Q^2) = \frac{c_{\lambda\lambda'}^{qe}}{\Lambda^2} \frac{Q^2}{2e^2} \left(e_q - K_Z \left(V^e - \lambda' A^e \right) \left(V^q - \lambda A^q \right) \right) \left(x f_q(x, Q^2) + x f_{\bar{q}}(x, Q^2) \right),$$
$$\Delta F_3(x, Q^2) = -\frac{c_{\lambda\lambda'}^{qe}}{\Lambda^2} \frac{Q^2}{2e^2} \left(\lambda\lambda' e_q - K_Z \left(\lambda' V^e - A^e \right) \left(\lambda V^q - A^q \right) \right) \left(f_q(x, Q^2) - f_{\bar{q}}(x, Q^2) \right),$$

where e is the positron charge, e_q is the charge on the quark q in units of the positron charge, θ_W is the Weinberg angle, and $K_Z = Q^2 / \sin^2(2\theta_W)(Q^2 + m_Z^2)$. The vector and axial couplings are given by $V^e = -\frac{1}{2} + 2\sin^2(\theta_W)$, $A^e = -\frac{1}{2}$, $V^q = I_3^q - 2\sin^2(\theta_W)e_q$ and $A^q = I_3^q$, where I_3^q is the third component of the quarks' weak isospin. These formulae are the natural generalisations of those derived in [72], where only right-handed four-fermion operators were considered. Taking combinations of these DIS structure-function corrections according to eq. (2.5) for the $\hat{Y}$ parameter and to eq. (C.1) for the $\hat{W}$ parameter yields the sought-for EFT corrections for DIS observables.

This calculation has been implemented in `APFEL` [124] following the strategy presented in [72]. FK tables are produced from `APFEL` [125] and then used to evaluate the DIS K-factors defined in eq. (3.10). Furthermore, we have used `APFEL` to include the higher-order QCD corrections in the SMEFT sector, so that in fact eq. (3.10) holds exactly for the DIS K-factors in our study.

---

[3]Scenario II is not relevant for DIS data, given that high-Q^2 structure functions only involve electrons.
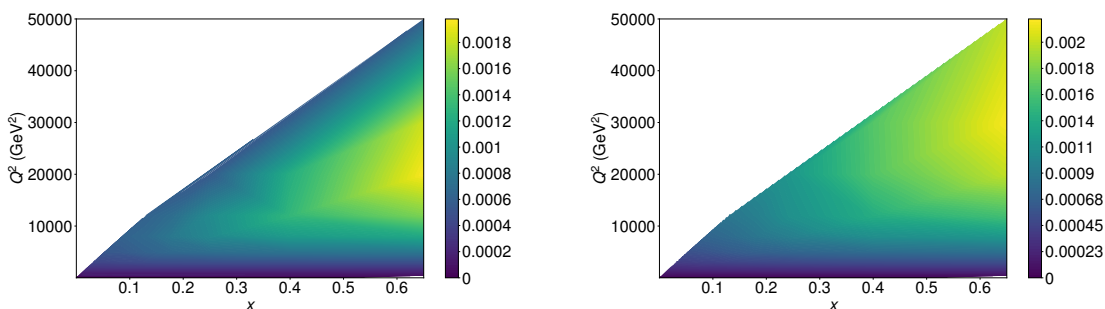
Figure 24. Contour maps indicating the value of the EFT correction, $K_{\mathrm{EFT}}(\hat{W}, \hat{Y}) - 1$ in eq. (3.9), for the DIS reduced cross sections as a function of x and Q^2 for two representative values of the EFT parameters: $\hat{W} = -10^{-4}$ (left panel) and $\hat{Y} = -10^{-4}$ (right panel).

Figure 24 displays contour maps indicating the EFT correction, $K_{\mathrm{EFT}}(\hat{W}, \hat{Y}) - 1$ in eq. (3.9), for the DIS reduced cross sections (which include both F_2 and xF_3) as a function of x and Q^2 for two representative values of the EFT parameters, $\hat{W} = -10^{-4}$ and $\hat{Y} = -10^{-4}$. These maps should be compared with figure 1 of [72], which considered different EFT scenarios. We find that the overall effect of non-zero $\hat{W}$ and $\hat{Y}$ parameters is rather small, well below the percent level even for the highest bins in Q^2 covered by the HERA data. This comparison highlights how, in this benchmark EFT scenario, the constraints on the $\hat{W}$ and $\hat{Y}$ parameters will be completely dominated by the high-mass Drell-Yan cross sections.

Drell-Yan distributions. As demonstrated in ref. [27], in the case of the dilepton $m_{\ell\ell}$ distribution in neutral-current Drell-Yan, one can derive the following analytic expression

$$\left(\frac{d\sigma}{dm_{\ell\ell}}\right)_{\mathrm{SMEFT}} = \left(\frac{d\sigma}{dm_{\ell\ell}}\right)_{\mathrm{SM}} \times \frac{\sum_{q,\ell} \mathcal{L}_{q\bar{q}}(m_{\ell\ell}^2/s, m_{\ell\ell})|F_{q\ell}(m_{\ell\ell}, \epsilon^{q\ell})|^2}{\sum_{q,\ell} \mathcal{L}_{q\bar{q}}(m_{\ell\ell}^2/s, m_{\ell\ell})|F_{q\ell}(m_{\ell\ell}, 0)|^2} , \qquad (C.3)$$

where $F_{ql}(m_{ll}, \epsilon^{q\ell})$ represents a form factor that depends on the values of the SMEFT coefficients

$$F_{q\ell}(m_{\ell\ell}, \epsilon^{q\ell}) = \delta^{ij} \frac{e^2 Q_q Q_\ell}{m_{\ell\ell}^2} + \delta^{ij} \frac{g_Z^q g_Z^\ell}{m_{\ell\ell}^2 - m_Z^2 + i m_Z \Gamma_Z} + \delta^{ij} \frac{\epsilon^{q\ell}}{v^2} . \qquad (C.4)$$

One can then match the Warsaw-basis parametrisation of eq. (2.4) to the contact terms $\epsilon_{ijkl}^{q\ell} \equiv \epsilon^{q\ell} \delta_{ij} \delta_{kl}$ in the above equation, where $i, j \ (k, l)$ are the quark (lepton) flavour indices. There are four combinations for quarks $q = u_L, u_R, d_L, d_R$ and two combinations for charged leptons $\ell = e_L, e_R$ relevant for the description of the neutral-current Drell-Yan process. Specifically, the matching is

$$\begin{aligned}
&\epsilon^{u_L e_L} = \hat{W} + \frac{1}{3} t_\theta^2 \hat{Y} , \quad \epsilon^{u_R e_L} = \frac{4}{3} t_\theta^2 \hat{Y} , \quad \epsilon^{d_L e_L} = -\hat{W} + \frac{1}{3} t_\theta^2 \hat{Y} , \quad \epsilon^{d_R e_L} = -\frac{2}{3} t_\theta^2 \hat{Y} , \\
&\epsilon^{u_L e_R} = \frac{2}{3} t_\theta^2 \hat{Y} , \qquad\quad \epsilon^{u_R e_R} = \frac{8}{3} t_\theta^2 \hat{Y} , \quad \epsilon^{d_L e_R} = \frac{2}{3} t_\theta^2 \hat{Y} , \qquad\quad \epsilon^{d_R e_R} = -\frac{4}{3} t_\theta^2 \hat{Y} ,
\end{aligned} \qquad (C.5)$$

where t_θ is the tangent of the Weinberg angle. We confirm these results using the modified form of the Feynman propagator for W^3 and B fields following directly from eq. (2.1).
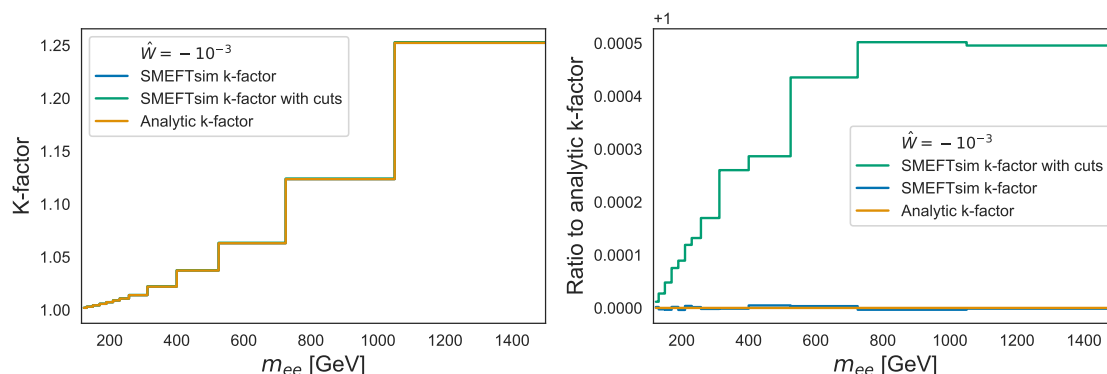
Figure 25. Left: comparison of the SMEFT K-factors, eq. (3.10), linearised in the EFT parameters, between the numerical and the analytical approaches for $\hat{W} = -10^{-3}$ and the kinematics of the ATLAS 7 TeV DY data. The label "cuts" indicates that we impose acceptance requirements of $p_T^\ell \geq 25$ GeV and $|\eta_\ell| \leq 2.5$ in the numerical (SMEFTsim) calculation. Right: relative difference in the K-factors shown in the left panel taking the analytical calculation as a reference.

In the unbroken phase, where $\hat{s} \gg m_W^2$, this simply amounts to the replacement $1/p^2 \to 1/p^2 - \hat{Y}/m_W^2$ for B and $1/p^2 \to 1/p^2 - \hat{W}/m_W^2$ for W^3 in the Standard Model calculation of $pp \to \ell^+\ell^-$. Note that there are two diagrams at the leading order with W^3 and B vector bosons propagating in the s-channel.

An analogous expression involving the transverse mass m_T can be derived in the CC case. These K-factors can be matched to the experimental data kinematics by integrating over the suitable ranges in $m_{\ell\ell}$ (m_T) and rapidity $y_{\ell\ell}$ (y_ℓ).

1D distributions. For the benchmarking between the analytical and numerical calculations, we use the relation between the $\hat{W}$ coefficient and the $c_{lq}^{(3)}$ coefficient in the Warsaw basis, $c_{lq}^{(3)} = -\hat{W}\Lambda^2/v^2$, where Λ is the SMEFT cut-off and v is the Higgs vev. We set $\Lambda = 1$ TeV and determine v using the (α_{EW}, m_Z, G_F) input scheme. We set m_Z and G_F using their PDG values: $m_Z = 91.1876$ GeV and $G_F = 1.1663787 \times 10^{-5}$ GeV^{-2}, and set $\alpha_{EW} = 1/127.951$. These values yield $v = 246.22$ GeV for the Higgs vev.

Figure 25 displays the comparison of the SMEFT K-factors, eq. (3.10), linearised in the EFT parameters, between the numerical and the analytical approaches for a representative value of $\hat{W} = -10^{-3}$ ($c_{lq}^{(3)} = 1.65 \times 10^{-2}$) and the kinematics of the ATLAS 7 TeV DY data. The label "cuts" indicates that we impose acceptance requirements of $p_T^\ell \geq 25$ GeV and $|\eta_\ell| \leq 2.5$ in the numerical (SMEFTsim) calculation on the final-state leptons; these cuts cannot be applied in the analytical calculation. The right panel shows the relative difference in these K-factors, with the analytical calculation as a reference.

From this comparison we observe, first of all, the perfect agreement between the analytical and numerical K-factors in the case of no acceptance cuts, and second, that the acceptance cuts on the leptonic variances leave the K-factor value essentially unchanged. Furthermore, we have verified that the same level of agreement in the calculation of the EFT K-factors between the numerical and analytical approaches is obtained in the case
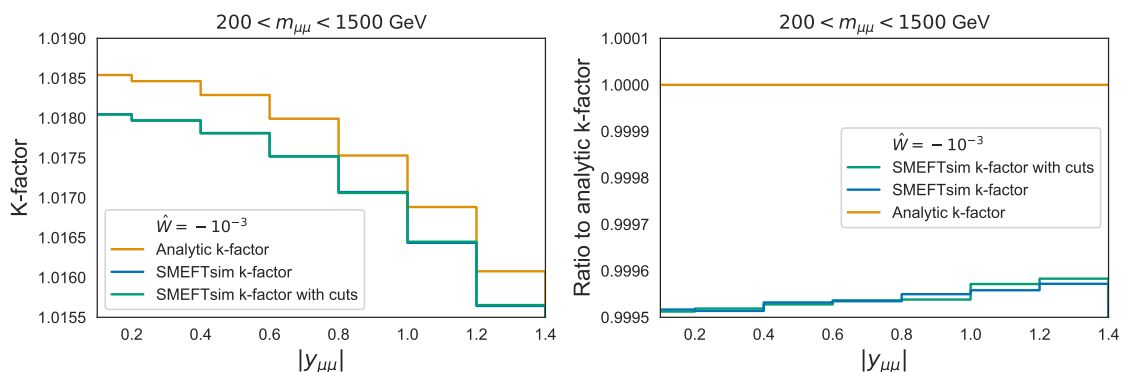
Figure 26. Same as figure 25 (also for $\hat{W} = 10^{-3}$) for the EFT K-factors for the double-differential Drell-Yan cross sections in $m_{\ell\ell}$ and $|y_{\ell\ell}|$, for the highest $m_{\ell\ell}$ bin of the CMS 7 TeV DY measurement. We only compute K-factors for $|y_{\ell\ell}|$ satisfying LO kinematics.

of the $\hat{Y}$ parameter, as well as once the quadratic EFT corrections are accounted for. In addition, the calculation of the EFT K-factors for the neutral current DY pseudo-data used in the HL-LHC projections of section 5 has been validated by ensuring that the analytical and `SMEFTsim` calculations are in perfect agreement.

As discussed in [27], far above the Z peak it is sufficient to include higher-order corrections to the SM cross section in eq. (C.3) to achieve good theoretical accuracy in the SMEFT, due to the (approximate) factorisation of higher-order QCD corrections in this region. We also note that renormalisation group evolution effects [6, 7] are not required in this calculation, since for the operators considered in our benchmark scenarios the corresponding anomalous dimensions are either Yukawa-suppressed or suppressed by NLO electroweak contributions.

2D distributions. In Drell-Yan datasets such as the CMS 7 TeV data of [121], the measurement is presented as a double-differential distribution in the dilepton invariant mass $m_{\ell\ell}$ (or equivalently $\tau = m_{\ell\ell}^2/s$) and rapidity $|y_{\ell\ell}|$. Also in this case, bin-by-bin EFT K-factors can be computed using the prescription of eq. (3.10). Figure 26 displays the same comparison as in figure 25 (also for $\hat{W} = -10^{-3}$) now for the EFT K-factors for the double-differential Drell-Yan cross sections in $m_{\ell\ell}$ and $|y_{\ell\ell}|$, for the highest $m_{\ell\ell}$ bin of the ATLAS 7 TeV DY measurement. Note that we only compute these K-factors for the $|y_{\ell\ell}|$ that satisfy the LO kinematics. As in the case of the 1D distributions, we find good agreement between the analytical and numerical calculations, with differences well below the absolute magnitude of the K-factors, and also that the impact of the acceptance cuts in the leptonic variables is negligible.

EFT K-factors for CC Drell-Yan at the HL-LHC. The left panel of figure 27 displays the EFT K-factor for high-mass charged-current DY production at the HL-LHC as a function of m_T, the transverse mass of the neutrino-lepton pair, for a parameter value of $\hat{W} = -10^{-3}$. We find rather large EFT corrections, with K-factors as large as $K_{\rm EFT} \simeq 5$ for the highest m_T bin. To validate this `SMEFTsim`-based calculation, we show a comparison
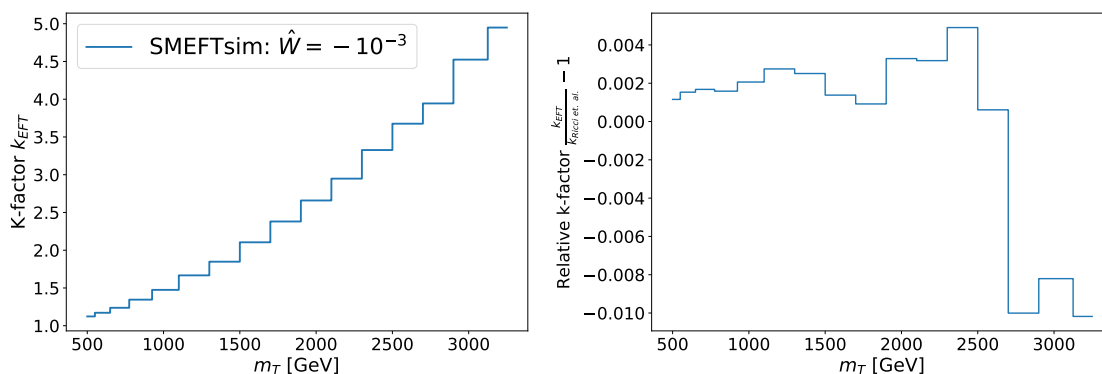
Figure 27. Left: the EFT K-factor for high-mass charged-current DY production at the HL-LHC as a function of m_T, the transverse mass of the neutrino-lepton pair, for $\hat{W} = -10^{-3}$. Right: the same quantity, now compared to the corresponding k-factors provided by the authors of [37].

between k_{EFT} and the linear k-factor $k_{\text{Ricci et. al}}$ provided by the authors of [37]. The quadratic k-factors are also shown in the left panel. In the right panel we plot the relative k-factor $\frac{k_{EFT}}{k_{\text{Ricci et.et}}} - 1$, finding good agreement between the two calculations. In addition, one can observe that the quadratic EFT corrections become larger than the linear ones for $m_T \gtrsim 3\,\text{TeV}$. Nevertheless, we point out that in our projections for the HL-LHC data from section 5 we restrict ourselves to the linear approximation in benchmark scenario I.

Effect of varying PDFs in the computation of the SMEFT K-factors. The SMEFT K-factors in eq. (3.9) are precomputed before the fit using a reference SM PDF set and then kept fixed. Here, we quantitatively assess the effect of varying the input NNLO PDF in eqs. (3.7) and (3.8).

We will first spell out the approximation made when using SMEFT K-factors, and then assess its impact. For definiteness, consider corrections from a non-zero $\hat{W}$ coefficient. Denote the partonic cross section by $\hat{\sigma}_{\text{SMEFT}} = \hat{\sigma}_{SM} + \hat{W}\hat{\sigma}_{\hat{W}}$ and let $\mathcal{L}_{\text{SMEFT}}$ denote the luminosity calculated with SMEFT PDFs. The total cross section for $\hat{W} \neq 0$ is given by

$$\sigma(\hat{W}) = (\hat{\sigma}_{\text{SM}} + \hat{W}\hat{\sigma}_{\hat{W}})_{ij} \otimes \mathcal{L}_{\text{SMEFT,ij}} \tag{C.6}$$
$$= \hat{\sigma}_{\text{SM,ij}} \otimes \mathcal{L}_{\text{SMEFT,ij}} + \hat{W}\hat{\sigma}_{\hat{W},ij} \otimes \mathcal{L}_{\text{SMEFT,ij}},$$

where we sum over the partons i, j. By using the SMEFT K-factor approach we approximate eq. (C.6) as

$$\sigma(\hat{W}) \approx \hat{\sigma}_{\text{SM,ij}} \otimes \mathcal{L}_{\text{SMEFT,ij}} \left(1 + \hat{W}\frac{\hat{\sigma}_{\hat{W},ij} \otimes \mathcal{L}_{\text{SM,ij}}}{\hat{\sigma}_{\text{SM,ij}} \otimes \mathcal{L}_{\text{SM,ij}}}\right) \tag{C.7}$$
$$= \hat{\sigma}_{\text{SM,ij}} \otimes \mathcal{L}_{\text{SMEFT,ij}} + W\hat{\sigma}_{\text{SM,ij}} \otimes \mathcal{L}_{\text{SMEFT,ij}} \left(\frac{\hat{\sigma}_{\hat{W},ij} \otimes \mathcal{L}_{\text{SM,ij}}}{\hat{\sigma}_{\text{SM,ij}} \otimes \mathcal{L}_{\text{SM,ij}}}\right).$$

Note that the first term of eqs. (C.6) and (C.7) are equal, thus we can express the approximation as

$$\frac{\hat{\sigma}_{\hat{W},ij} \otimes \mathcal{L}_{\text{SMEFT,ij}}}{\hat{\sigma}_{\text{SM,ij}} \otimes \mathcal{L}_{\text{SMEFT,ij}}} \approx \frac{\hat{\sigma}_{\hat{W},ij} \otimes \mathcal{L}_{\text{SM,ij}}}{\hat{\sigma}_{\text{SM,ij}} \otimes \mathcal{L}_{\text{SM,ij}}}, \tag{C.8}$$
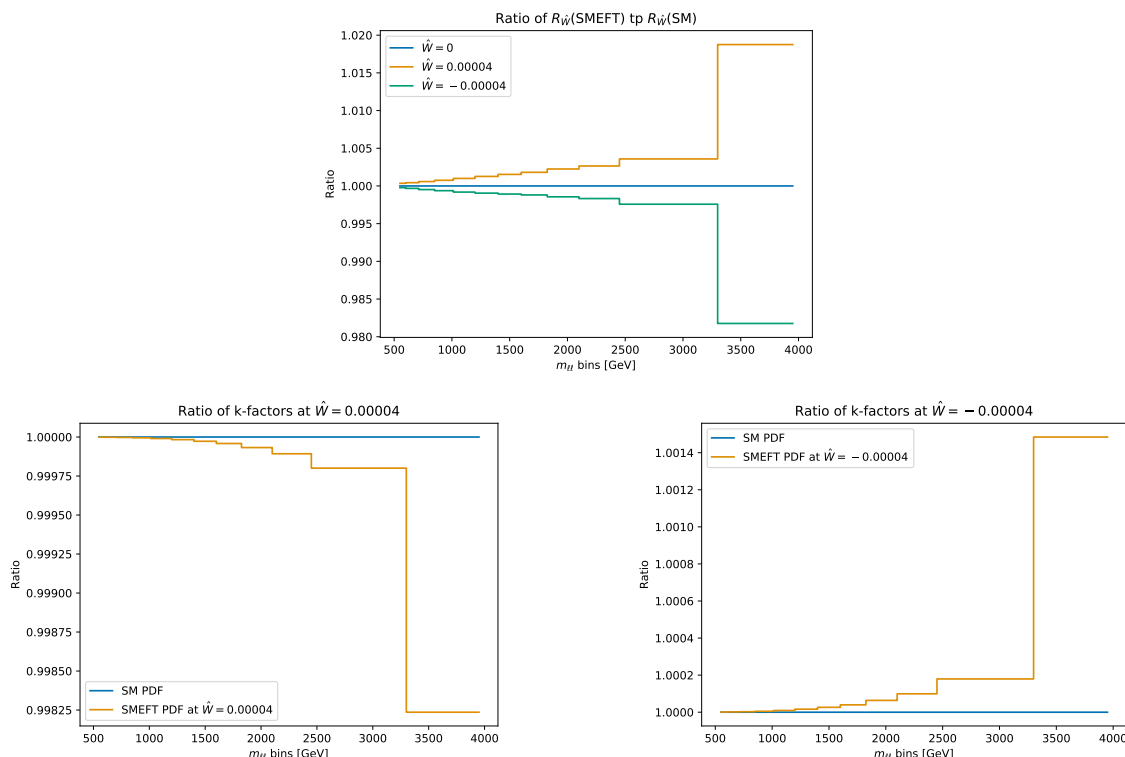
Figure 28. The impact of the SMEFT PDFs on R_W (above) and the SMEFT K-factor $K = 1 + \hat{W} R_W$ (below) calculated at the Scenario I benchmark points $\hat{W} = \pm 4 \cdot 10^{-5}$.

or, equivalently, as

$$R_{\hat{W}}(\text{SMEFT}) \approx R_{\hat{W}}(\text{SM}) \tag{C.9}$$

where R is defined in eqs. (3.7) and (3.8) and is computed either with SMEFT PDFs or fixed SM PDFs. In what follows we test whether $R(\text{SM})$ is a good approximation for $R(\text{SMEFT})$, taking each of the coefficients of Scenario I and II at a time.

We will first consider Scenario I and the $\hat{W}$ parameter. We use the PDFs including HL-LHC pseudodata and calculate $R_{\hat{W}}$ for the HL-LHC NC Drell-Yan bins outlined in section 5. In figure 28 we observe that, using fixed SM PDFs in the computation of $R_{\hat{W}}$ yields a 2% deviation in the highest invariant mass bin. In the same figure we assess the impact of these differences on the SMEFT K-factors themselves, $K_{\hat{W}} = 1 + \hat{W} R_{\hat{W}}$, calculated at each of the benchmark points $\hat{W} = \pm 4 \cdot 10^{-5}$ and we observe that the difference at the level of the observable is completely negligible, at the permil level.

In the case of the $\hat{Y}$ parameter we observe a rather larger deviation in the highest invariant mass bin, which reaches to 4%, as shown in figure 29. However, as in the case of $\hat{W}$, the impact of these discrepancies on the SMEFT K-factors, calculated at each of the benchmark points $\hat{Y} = \pm 1.2 \cdot 10^{-4}$, thus on the observable is below the percent level, which is still negligible compared to the experimental and theoretical uncertainties associated to the last bin if the invariant mass distribution.
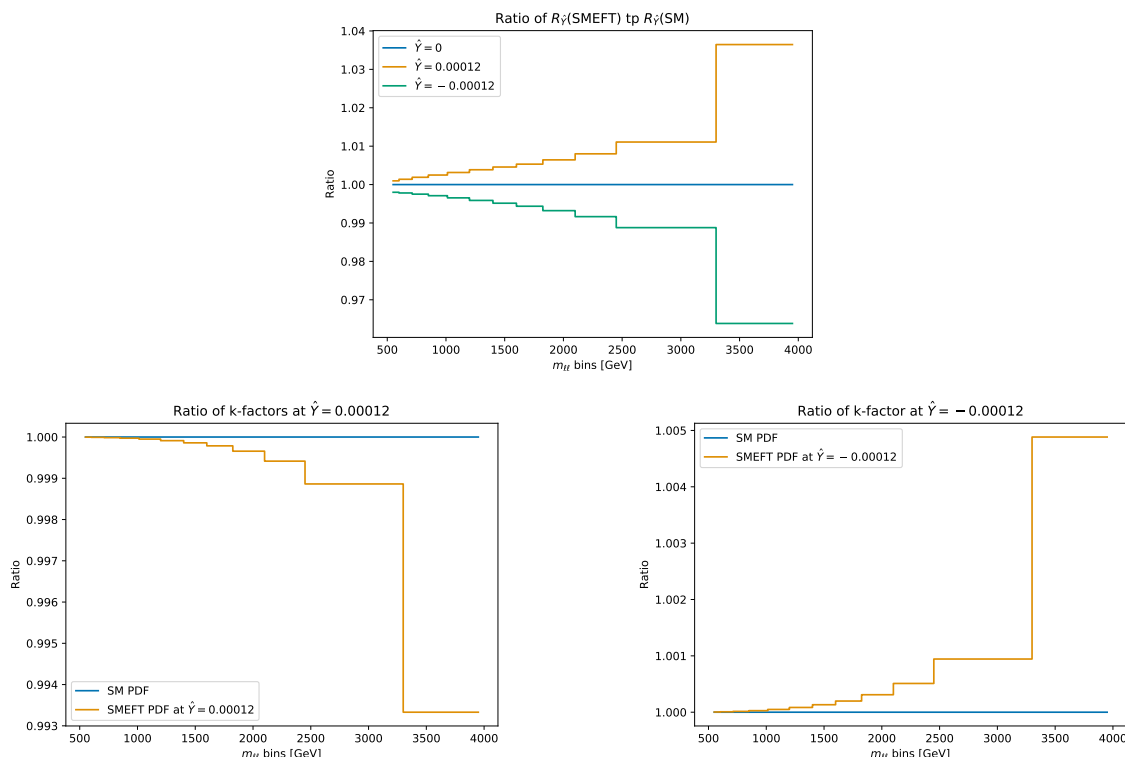
Figure 29. Same as figure 28, in the case of $\hat{Y}$ parameter.

Finally, we turn to Scenario II, in which we expect to observe the largest deviation between our approximation, based on using fixed SM PDFs in the computation of the SMEFT K-factors, and the full calculation. This is expected because we include both the linear and quadratic terms in the EFT expansion, and because of the flavour non-universal structure of Scenario II. Here we will denote the K-factor by $K = 1 + C_{33}^{D\mu} R^{\mathrm{lin}} + (C_{33}^{D\mu})^2 R^{\mathrm{quad}}$ and calculate the dependence of each R^{lin}, R^{quad} on the SMEFT PDFs. We observe a 10% deviation in the computation of R in the highest invariant mass bin, as shown in figure 30. However, the impact of such differences on the actual SMEFT K-factors, thus on the observables, calculated at each of the benchmark points $C_{33}^{D\mu} = \pm 0.014$, is still at the percent level, which is acceptable compared to the experimental and theoretical uncertainties.

D Flavour dependence of the SMEFT PDFs

In section 4, when discussing the results of the joint PDF and EFT fits to available Drell-Yan cross-section data, we presented the comparison between the SM and SMEFT PDFs in benchmark scenario I for different values of the $\hat{W}$ and $\hat{Y}$ parameters in terms of the partonic luminosities, figure 6. In this appendix we present the corresponding comparisons between SM and SMEFT PDFs at the level of individual PDF flavours. Figure 31 displays a comparison between the SM and the SMEFT PDFs at $Q = 100\,\mathrm{GeV}$ for representative values of the $\hat{W}$ (upper) and of $\hat{Y}$ (lower panels) parameters. The values of $\hat{W}$ and $\hat{Y}$ are

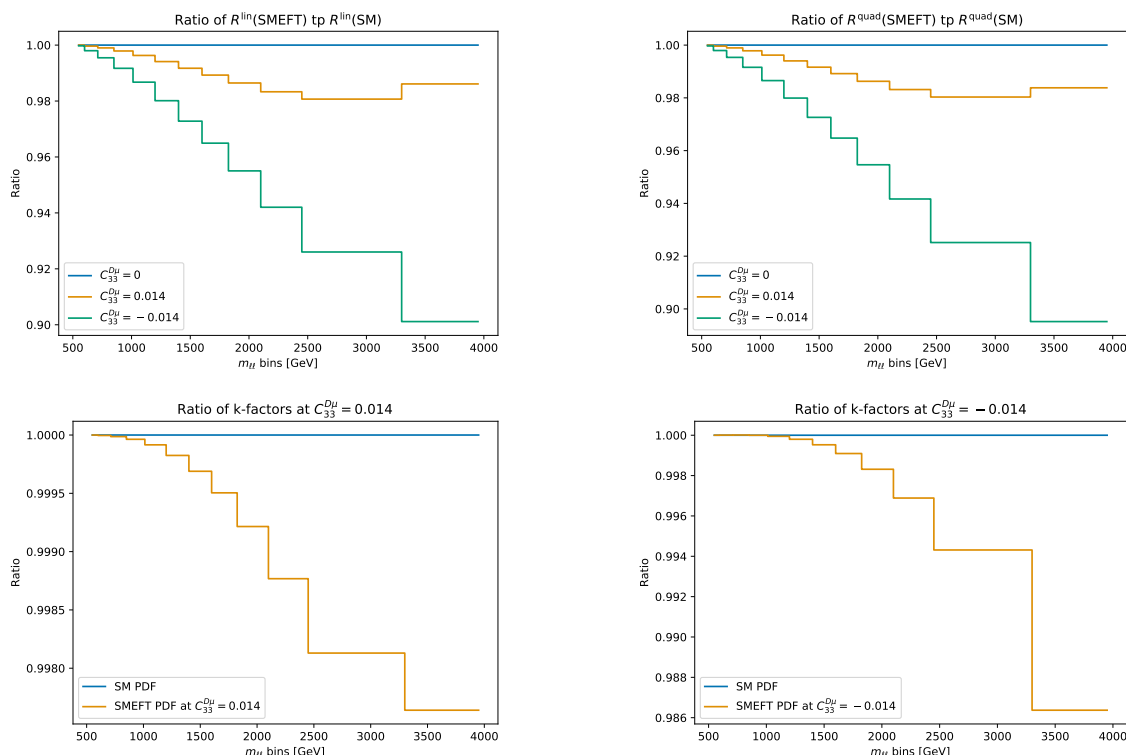**Figure 30**. The impact of the SMEFT PDFs on $R^{\rm lin}$, $R^{\rm quad}$ (above) and the SMEFT K-factor $K = 1 + C_{33}^{D\mu}R^{\rm lin} + (C_{33}^{D\mu})^2 R^{\rm quad}$ (below) calculated at the Scenario II benchmark points $C_{33}^{D\mu} = \pm 0.014$.

chosen to be close to the upper and lower limits of the 68% CL intervals reported in table 4. The error band in the SM PDFs corresponds to the 68% CL PDF uncertainty, while for the SMEFT PDFs only the central values are shown.

In all cases, one finds that the EFT-induced shifts on the PDFs are smaller than their uncertainties, though in some cases these shifts can represent up to one-third of a standard deviation. In particular, the up and down antiquarks in the region $x \gtrsim 10^{-2}$ are the PDF flavours most affected by the EFT effects. This finding can be understood from the fact that the NC Drell-Yan cross section is proportional to the $u\bar{u}$ and $d\bar{d}$ combinations at leading order, but the up and down quark PDF are already well constrained by lower-energy DIS measurements. Furthermore, we have verified that the PDF uncertainties themselves are unchanged in the SMEFT fits. The results of figure 31 are consistent with those of table 4 and demonstrate that, with current data, the interplay between EFT effects and PDFs in the high-mass Drell-Yan tails is appreciable but remains subdominant as compared to other sources of uncertainty.
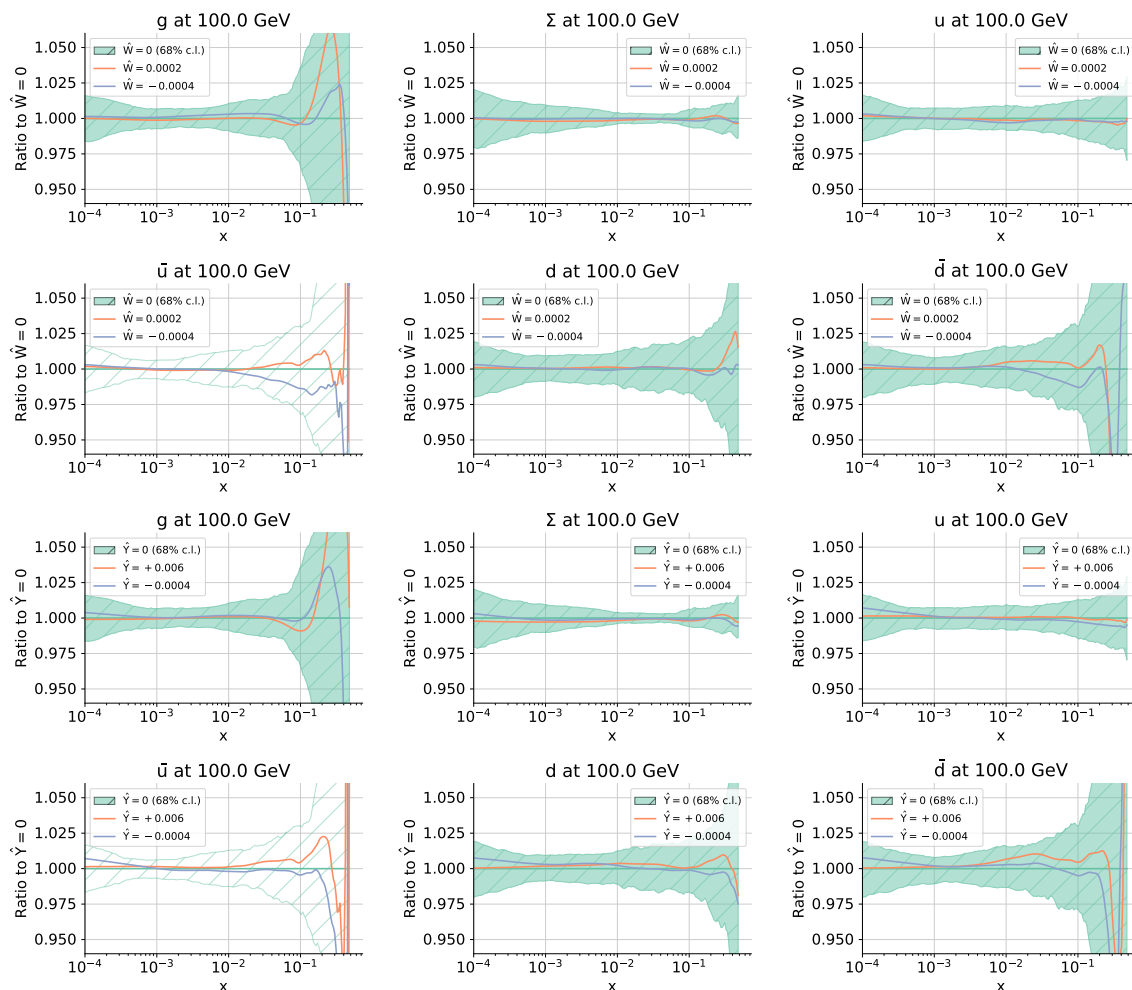
Figure 31. Comparison between the SM and the SMEFT PDFs at $Q = 100\,\text{GeV}$, displayed as ratios to the central value of the SM PDFs, for representative values of the $\hat{W}$ (upper) and of $\hat{Y}$ (lower panels) parameters. We show the gluon, the total quark singlet, the up quark and antiquark, and the down quark and antiquark PDFs. The values of $\hat{W}$ and $\hat{Y}$ are chosen to be close to the upper and lower limits of the 68% CL intervals reported in table 4.

References

[1] S. Weinberg, *Baryon and lepton nonconserving processes*, *[Phys. Rev. Lett.](#)* **43** (1979) 1566 [INSPIRE].

[2] W. Buchmüller and D. Wyler, *Effective Lagrangian analysis of new interactions and flavor conservation*, *[Nucl. Phys. B](#)* **268** (1986) 621 [INSPIRE].

[3] H. Georgi, *Effective field theory*, *[Ann. Rev. Nucl. Part. Sci.](#)* **43** (1993) 209 [INSPIRE].

[4] G.F. Giudice, C. Grojean, A. Pomarol and R. Rattazzi, *The strongly-interacting light Higgs*, *JHEP* **06** (2007) 045 [hep-ph/0703164] [INSPIRE].

[5] B. Grzadkowski, M. Iskrzynski, M. Misiak and J. Rosiek, *Dimension-six terms in the standard model Lagrangian*, *JHEP* **10** (2010) 085 [arXiv:1008.4884] [INSPIRE].

[6] R. Alonso, E.E. Jenkins, A.V. Manohar and M. Trott, *Renormalization group evolution of the standard model dimension six operators. Part III. Gauge coupling dependence and phenomenology*, *JHEP* **04** (2014) 159 [arXiv:1312.2014] [INSPIRE].

[7] E.E. Jenkins, A.V. Manohar and M. Trott, *Renormalization group evolution of the standard model dimension six operators. Part II. Yukawa dependence*, *JHEP* **01** (2014) 035 [arXiv:1310.4838] [INSPIRE].

[8] E.E. Jenkins, A.V. Manohar and M. Trott, *Renormalization group evolution of the standard model dimension six operators. Part I. Formalism and lambda dependence*, *JHEP* **10** (2013) 087 [arXiv:1308.2627] [INSPIRE].

[9] B. Henning, X. Lu and H. Murayama, *How to use the standard model effective field theory*, *JHEP* **01** (2016) 023 [arXiv:1412.1837] [INSPIRE].

[10] I. Brivio and M. Trott, *The standard model as an effective field theory*, *Phys. Rept.* **793** (2019) 1 [arXiv:1706.08945] [INSPIRE].

[11] J.D. Wells and Z. Zhang, *Effective theories of universal theories*, *JHEP* **01** (2016) 123 [arXiv:1510.08462] [INSPIRE].

[12] C. Englert, G.F. Giudice, A. Greljo and M. Mccullough, *The $\hat{H}$-parameter: an oblique Higgs view*, *JHEP* **09** (2019) 041 [arXiv:1903.07725] [INSPIRE].

[13] J. Fuentes-Martin, M. König, J. Pagès, A.E. Thomsen and F. Wilsch, *SuperTracer: a calculator of functional supertraces for one-loop EFT matching*, *JHEP* **04** (2021) 281 [arXiv:2012.08506] [INSPIRE].

[14] T. Cohen, X. Lu and Z. Zhang, *STrEAMlining EFT matching*, *SciPost Phys.* **10** (2021) 098 [arXiv:2012.07851] [INSPIRE].

[15] F. Feruglio, *The chiral approach to the electroweak interactions*, *Int. J. Mod. Phys. A* **8** (1993) 4937 [hep-ph/9301281] [INSPIRE].

[16] B. Grinstein and M. Trott, *A Higgs-Higgs bound state due to new physics at a TeV*, *Phys. Rev. D* **76** (2007) 073002 [arXiv:0704.1505] [INSPIRE].

[17] R. Contino, C. Grojean, M. Moretti, F. Piccinini and R. Rattazzi, *Strong double Higgs production at the LHC*, *JHEP* **05** (2010) 089 [arXiv:1002.1011] [INSPIRE].

[18] R. Alonso, M.B. Gavela, L. Merlo, S. Rigolin and J. Yepes, *The effective chiral Lagrangian for a light dynamical "Higgs particle"*, *Phys. Lett. B* **722** (2013) 330 [*Erratum ibid.* **726** (2013) 926] [arXiv:1212.3305] [INSPIRE].

[19] R. Alonso, I. Brivio, B. Gavela, L. Merlo and S. Rigolin, *Sigma decomposition*, *JHEP* **12** (2014) 034 [arXiv:1409.1589] [INSPIRE].

[20] G. Buchalla, O. Catà, A. Celis and C. Krause, *Fitting Higgs data with nonlinear effective theory*, *Eur. Phys. J. C* **76** (2016) 233 [arXiv:1511.00988] [INSPIRE].

[21] R. Alonso, E.E. Jenkins and A.V. Manohar, *A geometric formulation of Higgs effective field theory: measuring the curvature of scalar field space*, *Phys. Lett. B* **754** (2016) 335 [arXiv:1511.00724] [INSPIRE].

[22] T. Cohen, N. Craig, X. Lu and D. Sutherland, *Is SMEFT enough?*, *JHEP* **03** (2021) 237 [arXiv:2008.08597] [INSPIRE].

[23] V. Cirigliano, M. Gonzalez-Alonso and M.L. Graesser, *Non-standard charged current interactions: beta decays versus the LHC*, *JHEP* **02** (2013) 046 [arXiv:1210.4553] [INSPIRE].

[24] J. de Blas, M. Chala and J. Santiago, *Global constraints on lepton-quark contact interactions*, *Phys. Rev. D* **88** (2013) 095011 [arXiv:1307.5068] [INSPIRE].

[25] M. González-Alonso and J. Martin Camalich, *Global effective-field-theory analysis of new-physics effects in (semi)leptonic kaon decays*, *JHEP* **12** (2016) 052 [arXiv:1605.07114] [INSPIRE].

[26] D.A. Faroughy, A. Greljo and J.F. Kamenik, *Confronting lepton flavor universality violation in B decays with high-p_T tau lepton searches at LHC*, *Phys. Lett. B* **764** (2017) 126 [arXiv:1609.07138] [INSPIRE].

[27] A. Greljo and D. Marzocca, *High-p_T dilepton tails and flavor physics*, *Eur. Phys. J. C* **77** (2017) 548 [arXiv:1704.09015] [INSPIRE].

[28] V. Cirigliano, A. Falkowski, M. González-Alonso and A. Rodríguez-Sánchez, *Hadronic τ decays as new physics probes in the LHC era*, *Phys. Rev. Lett.* **122** (2019) 221801 [arXiv:1809.01161] [INSPIRE].

[29] A. Greljo, J. Martin Camalich and J.D. Ruiz-Álvarez, *Mono-τ signatures at the LHC constrain explanations of B-decay anomalies*, *Phys. Rev. Lett.* **122** (2019) 131803 [arXiv:1811.07920] [INSPIRE].

[30] S. Bansal, R.M. Capdevilla, A. Delgado, C. Kolda, A. Martin and N. Raj, *Hunting leptoquarks in monolepton searches*, *Phys. Rev. D* **98** (2018) 015037 [arXiv:1806.02370] [INSPIRE].

[31] A. Angelescu, D.A. Faroughy and O. Sumensari, *Lepton flavor violation and dilepton tails at the LHC*, *Eur. Phys. J. C* **80** (2020) 641 [arXiv:2002.05684] [INSPIRE].

[32] M. Farina, G. Panico, D. Pappadopulo, J.T. Ruderman, R. Torre and A. Wulzer, *Energy helps accuracy: electroweak precision tests at hadron colliders*, *Phys. Lett. B* **772** (2017) 210 [arXiv:1609.08157] [INSPIRE].

[33] S. Alioli, M. Farina, D. Pappadopulo and J.T. Ruderman, *Catching a new force by the tail*, *Phys. Rev. Lett.* **120** (2018) 101801 [arXiv:1712.02347] [INSPIRE].

[34] N. Raj, *Anticipating nonresonant new physics in dilepton angular spectra at the LHC*, *Phys. Rev. D* **95** (2017) 015011 [arXiv:1610.03795] [INSPIRE].

[35] M. Schmaltz and Y.-M. Zhong, *The leptoquark hunter's guide: large coupling*, *JHEP* **01** (2019) 132 [arXiv:1810.10017] [INSPIRE].

[36] G. Brooijmans et al., *Les Houches* 2019 *physics at TeV colliders: new physics working group report*, in 11th *Les Houches workshop on physics at TeV colliders: PhysTeV Les Houches*, (2020) [arXiv:2002.12220] [INSPIRE].

[37] R. Torre, L. Ricci and A. Wulzer, *On the W&Y interpretation of high-energy Drell-Yan measurements*, *JHEP* **02** (2021) 144 [arXiv:2008.12978] [INSPIRE].

[38] J. Fuentes-Martin, A. Greljo, J. Martin Camalich and J.D. Ruiz-Alvarez, *Charm physics confronts high-p_T lepton tails*, *JHEP* **11** (2020) 080 [arXiv:2003.12421] [INSPIRE].

[39] S. Alioli, V. Cirigliano, W. Dekens, J. de Vries and E. Mereghetti, *Right-handed charged currents in the era of the Large Hadron Collider*, *JHEP* **05** (2017) 086 [arXiv:1703.04751] [INSPIRE].

[40] S. Alioli, M. Farina, D. Pappadopulo and J.T. Ruderman, *Precision probes of QCD at high energies*, *JHEP* **07** (2017) 097 [arXiv:1706.03068] [INSPIRE].

[41] S. Alioli, W. Dekens, M. Girard and E. Mereghetti, *NLO QCD corrections to SM-EFT dilepton and electroweak Higgs boson production, matched to parton shower in POWHEG*, *JHEP* **08** (2018) 205 [arXiv:1804.07407] [INSPIRE].

[42] S. Alioli, R. Boughezal, E. Mereghetti and F. Petriello, *Novel angular dependence in Drell-Yan lepton production via dimension-8 operators*, *Phys. Lett. B* **809** (2020) 135703 [arXiv:2003.11615] [INSPIRE].

[43] G. Panico, L. Ricci and A. Wulzer, *High-energy EFT probes with fully differential Drell-Yan measurements*, arXiv:2103.10532 [INSPIRE].

[44] CMS collaboration, *Search for resonant and nonresonant new phenomena in high-mass dilepton final states at $\sqrt{s} = 13$ TeV*, arXiv:2103.02708 [INSPIRE].

[45] ATLAS collaboration, *Search for new phenomena in final states with two leptons and one or no b-tagged jets at $\sqrt{s} = 13$ TeV using the ATLAS detector*, arXiv:2105.13847 [INSPIRE].

[46] D. Marzocca, U. Min and M. Son, *Bottom-flavored mono-tau tails at the LHC*, *JHEP* **12** (2020) 035 [arXiv:2008.07541] [INSPIRE].

[47] Y. Afik, S. Bar-Shalom, J. Cohen and Y. Rozen, *Searching for new physics with $b\bar{b}\ell^+\ell^-$ contact interactions*, *Phys. Lett. B* **807** (2020) 135541 [arXiv:1912.00425] [INSPIRE].

[48] A. Alves, O.J.P.t. Eboli, G. Grilli Di Cortona and R.R. Moreira, *Indirect and monojet constraints on scalar leptoquarks*, *Phys. Rev. D* **99** (2019) 095005 [arXiv:1812.08632] [INSPIRE].

[49] F. Krauss, S. Kuttimalai and T. Plehn, *LHC multijet events as a probe for anomalous dimension-six gluon interactions*, *Phys. Rev. D* **95** (2017) 035024 [arXiv:1611.00767] [INSPIRE].

[50] S. Alte, M. König and W. Shepherd, *Consistent searches for SMEFT effects in non-resonant dijet events*, *JHEP* **01** (2018) 094 [arXiv:1711.07484] [INSPIRE].

[51] V. Hirschi, F. Maltoni, I. Tsinikos and E. Vryonidou, *Constraining anomalous gluon self-interactions at the LHC: a reappraisal*, *JHEP* **07** (2018) 093 [arXiv:1806.04696] [INSPIRE].

[52] R. Goldouzian and M.D. Hildreth, *LHC dijet angular distributions as a probe for the dimension-six triple gluon vertex*, *Phys. Lett. B* **811** (2020) 135889 [arXiv:2001.02736] [INSPIRE].

[53] A. Buckley et al., *Constraining top quark effective theory in the LHC run II era*, *JHEP* **04** (2016) 015 [arXiv:1512.03360] [INSPIRE].

[54] C. Englert, L. Moore, K. Nordström and M. Russell, *Giving top quark effective operators a boost*, *Phys. Lett. B* **763** (2016) 9 [arXiv:1607.04304] [INSPIRE].

[55] N.P. Hartland et al., *A Monte Carlo global analysis of the standard model effective field theory: the top quark sector*, *JHEP* **04** (2019) 100 [arXiv:1901.05965] [INSPIRE].

[56] I. Brivio et al., *O new physics, where art thou? A global search in the top sector*, *JHEP* **02** (2020) 131 [arXiv:1910.03606] [InSPIRE].

[57] G. Durieux, M. Perelló, M. Vos and C. Zhang, *Global and optimal probes for the top-quark effective field theory at future lepton colliders*, *JHEP* **10** (2018) 168 [arXiv:1807.02121] [InSPIRE].

[58] S. van Beek, E.R. Nocera, J. Rojo and E. Slade, *Constraining the SMEFT with Bayesian reweighting*, *SciPost Phys.* **7** (2019) 070 [arXiv:1906.05296] [InSPIRE].

[59] S. Bißmann, C. Grunwald, G. Hiller and K. Kröninger, *Top and Beauty synergies in SMEFT-fits at present and future colliders*, *JHEP* **06** (2021) 010 [arXiv:2012.10456] [InSPIRE].

[60] S. Bruggisser, R. Schäfer, D. van Dyk and S. Westhoff, *The flavor of UV physics*, *JHEP* **05** (2021) 257 [arXiv:2101.07273] [InSPIRE].

[61] J. Ellis, M. Madigan, K. Mimasu, V. Sanz and T. You, *Top, Higgs, diboson and electroweak fit to the standard model effective field theory*, *JHEP* **04** (2021) 279 [arXiv:2012.02779] [InSPIRE].

[62] A. Falkowski, M. Gonzalez-Alonso, A. Greljo and D. Marzocca, *Global constraints on anomalous triple gauge couplings in effective field theory approach*, *Phys. Rev. Lett.* **116** (2016) 011801 [arXiv:1508.00581] [InSPIRE].

[63] A. Falkowski, M. Gonzalez-Alonso, A. Greljo, D. Marzocca and M. Son, *Anomalous triple gauge couplings in the effective field theory approach at the LHC*, *JHEP* **02** (2017) 115 [arXiv:1609.06312] [InSPIRE].

[64] J. Baglio, S. Dawson and I.M. Lewis, *An NLO QCD effective field theory analysis of W^+W^- production at the LHC including fermionic operators*, *Phys. Rev. D* **96** (2017) 073003 [arXiv:1708.03332] [InSPIRE].

[65] J. Baglio, S. Dawson, S. Homiller, S.D. Lane and I.M. Lewis, *Validity of standard model EFT studies of VH and VV production at NLO*, *Phys. Rev. D* **101** (2020) 115004 [arXiv:2003.07862] [InSPIRE].

[66] G. Panico, F. Riva and A. Wulzer, *Diboson interference resurrection*, *Phys. Lett. B* **776** (2018) 473 [arXiv:1708.07823] [InSPIRE].

[67] J.J. Ethier, R. Gomez-Ambrosio, G. Magni and J. Rojo, *SMEFT analysis of vector boson scattering and diboson data from the LHC run II*, *Eur. Phys. J. C* **81** (2021) 560 [arXiv:2101.03180] [InSPIRE].

[68] C. Grojean, M. Montull and M. Riembau, *Diboson at the LHC vs LEP*, *JHEP* **03** (2019) 020 [arXiv:1810.05149] [InSPIRE].

[69] R. Gomez-Ambrosio, *Studies of dimension-six EFT effects in vector boson scattering*, *Eur. Phys. J. C* **79** (2019) 389 [arXiv:1809.04189] [InSPIRE].

[70] A. Dedes, P. Kozów and M. Szleper, *Standard model EFT effects in vector-boson scattering at the LHC*, *Phys. Rev. D* **104** (2021) 013003 [arXiv:2011.07367] [InSPIRE].

[71] J. Gao, L. Harland-Lang and J. Rojo, *The structure of the proton in the LHC precision era*, *Phys. Rept.* **742** (2018) 1 [arXiv:1709.04922] [InSPIRE].

[72] S. Carrazza, C. Degrande, S. Iranipour, J. Rojo and M. Ubiali, *Can new physics hide inside the proton?*, *Phys. Rev. Lett.* **123** (2019) 132001 [arXiv:1905.05215] [InSPIRE].

[73] NNPDF collaboration, *Parton distributions for the LHC run II*, *JHEP* **04** (2015) 040 [arXiv:1410.8849] [INSPIRE].

[74] NNPDF collaboration, *Parton distributions from high-precision collider data*, *Eur. Phys. J. C* **77** (2017) 663 [arXiv:1706.00428] [INSPIRE].

[75] T.-J. Hou et al., *New CTEQ global analysis of quantum chromodynamics with high-precision data from the LHC*, *Phys. Rev. D* **103** (2021) 014013 [arXiv:1912.10053] [INSPIRE].

[76] 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** (2021) 341 [arXiv:2012.04684] [INSPIRE].

[77] R. Abdul Khalek, S. Bailey, J. Gao, L. Harland-Lang and J. Rojo, *Towards ultimate parton distributions at the High-Luminosity LHC*, *Eur. Phys. J. C* **78** (2018) 962 [arXiv:1810.03639] [INSPIRE].

[78] LHCb collaboration, *Test of lepton universality using $B^+ \to K^+ \ell^+ \ell^-$ decays*, *Phys. Rev. Lett.* **113** (2014) 151601 [arXiv:1406.6482] [INSPIRE].

[79] LHCb collaboration, *Test of lepton universality with $B^0 \to K^{*0} \ell^+ \ell^-$ decays*, *JHEP* **08** (2017) 055 [arXiv:1705.05802] [INSPIRE].

[80] LHCb collaboration, *Search for lepton-universality violation in $B^+ \to K^+ \ell^+ \ell^-$ decays*, *Phys. Rev. Lett.* **122** (2019) 191801 [arXiv:1903.09252] [INSPIRE].

[81] LHCb collaboration, *Test of lepton universality in beauty-quark decays*, arXiv:2103.11769 [INSPIRE].

[82] M.E. Peskin and T. Takeuchi, *Estimation of oblique electroweak corrections*, *Phys. Rev. D* **46** (1992) 381 [INSPIRE].

[83] G. Altarelli, R. Barbieri and S. Jadach, *Toward a model independent analysis of electroweak data*, *Nucl. Phys. B* **369** (1992) 3 [*Erratum ibid.* **376** (1992) 444] [INSPIRE].

[84] R. Barbieri, A. Pomarol, R. Rattazzi and A. Strumia, *Electroweak symmetry breaking after LEP-1 and LEP-2*, *Nucl. Phys. B* **703** (2004) 127 [hep-ph/0405040] [INSPIRE].

[85] I. Brivio, Y. Jiang and M. Trott, *The SMEFTsim package, theory and tools*, *JHEP* **12** (2017) 070 [arXiv:1709.06492] [INSPIRE].

[86] ATLAS collaboration, *Measurement of the double-differential high-mass Drell-Yan cross section in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, *JHEP* **08** (2016) 009 [arXiv:1606.01736] [INSPIRE].

[87] CMS collaboration, *Measurements of differential and double-differential Drell-Yan cross sections in proton-proton collisions at 8 TeV*, *Eur. Phys. J. C* **75** (2015) 147 [arXiv:1412.1115] [INSPIRE].

[88] R. Barbieri, G. Isidori, J. Jones-Perez, P. Lodone and D.M. Straub, U(2) *and minimal flavour violation in supersymmetry*, *Eur. Phys. J. C* **71** (2011) 1725 [arXiv:1105.2296] [INSPIRE].

[89] A.L. Kagan, G. Perez, T. Volansky and J. Zupan, *General minimal flavor violation*, *Phys. Rev. D* **80** (2009) 076002 [arXiv:0903.1794] [INSPIRE].

[90] J. Fuentes-Martín, G. Isidori, J. Pagès and K. Yamamoto, *With or without* U(2)? *Probing non-standard flavor and helicity structures in semileptonic B decays*, *Phys. Lett. B* **800** (2020) 135080 [arXiv:1909.02519] [INSPIRE].

[91] D.A. Faroughy, G. Isidori, F. Wilsch and K. Yamamoto, *Flavour symmetries in the SMEFT*, *JHEP* **08** (2020) 166 [arXiv:2005.05366] [INSPIRE].

[92] A. Greljo, P. Stangl and A.E. Thomsen, *A model of muon anomalies*, arXiv:2103.13991 [INSPIRE].

[93] ATLAS collaboration, *Search for new high-mass phenomena in the dilepton final state using* 36.1 fb^{-1} *of proton-proton collision data at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, *JHEP* **10** (2017) 182 [arXiv:1707.02424] [INSPIRE].

[94] M. Bordone, A. Greljo and D. Marzocca, *Exploiting dijet resonance searches for flavor physics*, arXiv:2103.10332 [INSPIRE].

[95] F. Faura, S. Iranipour, E.R. Nocera, J. Rojo and M. Ubiali, *The strangest proton?*, *Eur. Phys. J. C* **80** (2020) 1168 [arXiv:2009.00014] [INSPIRE].

[96] H1 and ZEUS collaborations, *Combination of measurements of inclusive deep inelastic* $e^{\pm}p$ *scattering cross sections and QCD analysis of HERA data*, *Eur. Phys. J. C* **75** (2015) 580 [arXiv:1506.06042] [INSPIRE].

[97] D0 collaboration, *Measurement of the electron charge asymmetry in* $p\bar{p} \to W + X \to e\nu + X$ *decays in* $p\bar{p}$ *collisions at* $\sqrt{s} = 1.96$ *TeV*, *Phys. Rev. D* **91** (2015) 032007 [*Erratum ibid.* **91** (2015) 079901] [arXiv:1412.2862] [INSPIRE].

[98] NuSea collaboration, *Improved measurement of the* $\bar{d}/\bar{u}$ *asymmetry in the nucleon sea*, *Phys. Rev. D* **64** (2001) 052002 [hep-ex/0103030] [INSPIRE].

[99] NuSea collaboration, *Absolute Drell-Yan dimuon cross-sections in* 800 *GeV/c pp and pd collisions*, hep-ex/0302019 [INSPIRE].

[100] J.C. Webb, *Measurement of continuum dimuon production in* 800 *GeV/c proton nucleon collisions*, Ph.D. thesis, New Mexico State U., Las Cruces, NM, U.S.A. (2003) [hep-ex/0301031] [INSPIRE].

[101] G. Moreno et al., *Dimuon production in proton-copper collisions at* $\sqrt{s} = 38.8$ *GeV*, *Phys. Rev. D* **43** (1991) 2815 [INSPIRE].

[102] CDF collaboration, *Measurement of* $d\sigma/dy$ *of Drell-Yan* e^+e^- *pairs in the Z mass region from* $p\bar{p}$ *collisions at* $\sqrt{s} = 1.96$ *TeV*, *Phys. Lett. B* **692** (2010) 232 [arXiv:0908.3914] [INSPIRE].

[103] D0 collaboration, *Measurement of the shape of the boson rapidity distribution for* $p\bar{p} \to Z/\gamma^* \to e^+e^- + X$ *events produced at* $\sqrt{s} = 1.96$ *TeV*, *Phys. Rev. D* **76** (2007) 012003 [hep-ex/0702025] [INSPIRE].

[104] D0 collaboration, *Measurement of the muon charge asymmetry in* $p\bar{p} \to W + X \to \mu\nu + X$ *events at* $\sqrt{s} = 1.96$ *TeV*, *Phys. Rev. D* **88** (2013) 091102 [arXiv:1309.2591] [INSPIRE].

[105] ATLAS collaboration, *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** (2012) 072004 [arXiv:1109.5141] [INSPIRE].

[106] ATLAS collaboration, *Measurement of the low-mass Drell-Yan differential cross section at* $\sqrt{s} = 7$ *TeV using the ATLAS detector*, *JHEP* **06** (2014) 112 [arXiv:1404.1212] [INSPIRE].

[107] ATLAS collaboration, *Precision measurement and interpretation of inclusive* W^+, W^- *and* Z/γ^* *production cross sections with the ATLAS detector*, *Eur. Phys. J. C* **77** (2017) 367 [arXiv:1612.03016] [INSPIRE].

[108] ATLAS collaboration, *Measurement of the production of a W boson in association with a charm quark in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, *JHEP* **05** (2014) 068 [arXiv:1402.6263] [INSPIRE].

[109] ATLAS collaboration, *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** (2016) 291 [arXiv:1512.02192] [INSPIRE].

[110] ATLAS collaboration, *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*, *JHEP* **05** (2018) 077 [*Erratum ibid.* **10** (2020) 048] [arXiv:1711.03296] [INSPIRE].

[111] CMS collaboration, *Measurement of the electron charge asymmetry in inclusive W production in pp collisions at $\sqrt{s} = 7$ TeV*, *Phys. Rev. Lett.* **109** (2012) 111806 [arXiv:1206.2598] [INSPIRE].

[112] CMS collaboration, *Measurement of associated W + charm production in pp collisions at $\sqrt{s} = 7$ TeV*, *JHEP* **02** (2014) 013 [arXiv:1310.1138] [INSPIRE].

[113] CMS collaboration, *Measurement of the Z boson differential cross section in transverse momentum and rapidity in proton-proton collisions at 8 TeV*, *Phys. Lett. B* **749** (2015) 187 [arXiv:1504.03511] [INSPIRE].

[114] CMS collaboration, *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** (2016) 469 [arXiv:1603.01803] [INSPIRE].

[115] CMS collaboration, *Measurement of associated production of a W boson and a charm quark in proton-proton collisions at $\sqrt{s} = 13$ TeV*, *Eur. Phys. J. C* **79** (2019) 269 [arXiv:1811.10021] [INSPIRE].

[116] LHCb collaboration, *Inclusive W and Z production in the forward region at $\sqrt{s} = 7$ TeV*, *JHEP* **06** (2012) 058 [arXiv:1204.1620] [INSPIRE].

[117] LHCb collaboration, *Measurement of the forward Z boson production cross-section in pp collisions at $\sqrt{s} = 7$ TeV*, *JHEP* **08** (2015) 039 [arXiv:1505.07024] [INSPIRE].

[118] LHCb collaboration, *Measurement of the cross-section for $Z \to e^+e^-$ production in pp collisions at $\sqrt{s} = 7$ TeV*, *JHEP* **02** (2013) 106 [arXiv:1212.4620] [INSPIRE].

[119] LHCb collaboration, *Measurement of forward W and Z boson production in pp collisions at $\sqrt{s} = 8$ TeV*, *JHEP* **01** (2016) 155 [arXiv:1511.08039] [INSPIRE].

[120] ATLAS collaboration, *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** (2013) 223 [arXiv:1305.4192] [INSPIRE].

[121] CMS collaboration, *Measurement of the differential and double-differential Drell-Yan cross sections in proton-proton collisions at $\sqrt{s} = 7$ TeV*, *JHEP* **12** (2013) 030 [arXiv:1310.7291] [INSPIRE].

[122] CMS collaboration, *Measurement of the differential Drell-Yan cross section in proton-proton collisions at $\sqrt{s} = 13$ TeV*, *JHEP* **12** (2019) 059 [arXiv:1812.10529] [INSPIRE].

[123] S. Forte, E. Laenen, P. Nason and J. Rojo, *Heavy quarks in deep-inelastic scattering*, *Nucl. Phys. B* **834** (2010) 116 [arXiv:1001.2312] [INSPIRE].

[124] V. Bertone, S. Carrazza and J. Rojo, *APFEL: a PDF evolution library with QED corrections*, *Comput. Phys. Commun.* **185** (2014) 1647 [arXiv:1310.1394] [INSPIRE].

[125] V. Bertone, S. Carrazza and N.P. Hartland, *APFELgrid: a high performance tool for parton density determinations*, *Comput. Phys. Commun.* **212** (2017) 205 [arXiv:1605.02070] [INSPIRE].

[126] J. Campbell and T. Neumann, *Precision phenomenology with MCFM*, *JHEP* **12** (2019) 034 [arXiv:1909.09117] [INSPIRE].

[127] R. Frederix, S. Frixione, V. Hirschi, D. Pagani, H.S. Shao and M. Zaro, *The automation of next-to-leading order electroweak calculations*, *JHEP* **07** (2018) 185 [arXiv:1804.10017] [INSPIRE].

[128] T. Carli et al., *A posteriori inclusion of parton density functions in NLO QCD final-state calculations at hadron colliders: the APPLGRID project*, *Eur. Phys. J. C* **66** (2010) 503 [arXiv:0911.2985] [INSPIRE].

[129] M. Grazzini, S. Kallweit and M. Wiesemann, *Fully differential NNLO computations with MATRIX*, *Eur. Phys. J. C* **78** (2018) 537 [arXiv:1711.06631] [INSPIRE].

[130] Y. Li and F. Petriello, *Combining QCD and electroweak corrections to dilepton production in FEWZ*, *Phys. Rev. D* **86** (2012) 094034 [arXiv:1208.5967] [INSPIRE].

[131] C. Duhr, F. Dulat and B. Mistlberger, *Drell-Yan cross section to third order in the strong coupling constant*, *Phys. Rev. Lett.* **125** (2020) 172001 [arXiv:2001.07717] [INSPIRE].

[132] C. Duhr, F. Dulat and B. Mistlberger, *Charged current Drell-Yan production at N^3LO*, *JHEP* **11** (2020) 143 [arXiv:2007.13313] [INSPIRE].

[133] NNPDF collaboration, *Illuminating the photon content of the proton within a global PDF analysis*, *SciPost Phys.* **5** (2018) 008 [arXiv:1712.07053] [INSPIRE].

[134] NNPDF collaboration, *Fitting parton distribution data with multiplicative normalization uncertainties*, *JHEP* **05** (2010) 075 [arXiv:0912.2276] [INSPIRE].

[135] R.D. Ball et al., *Parton distribution benchmarking with LHC data*, *JHEP* **04** (2013) 125 [arXiv:1211.5142] [INSPIRE].

[136] S. Lionetti et al., *Precision determination of α_s using an unbiased global NLO parton set*, *Phys. Lett. B* **701** (2011) 346 [arXiv:1103.2369] [INSPIRE].

[137] R.D. Ball et al., *Precision NNLO determination of $\alpha_s(M_Z)$ using an unbiased global parton set*, *Phys. Lett. B* **707** (2012) 66 [arXiv:1110.2483] [INSPIRE].

[138] NNPDF collaboration, *Precision determination of the strong coupling constant within a global PDF analysis*, *Eur. Phys. J. C* **78** (2018) 408 [arXiv:1802.03398] [INSPIRE].

[139] ATLAS collaboration, *Search for high-mass dilepton resonances using $139\,fb^{-1}$ of pp collision data collected at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *Phys. Lett. B* **796** (2019) 68 [arXiv:1903.06248] [INSPIRE].

[140] E. Maguire, L. Heinrich and G. Watt, *HEPData: a repository for high energy physics data*, *J. Phys. Conf. Ser.* **898** (2017) 102006 [arXiv:1704.05473] [INSPIRE].

[141] J. Alwall et al., *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 [arXiv:1405.0301] [INSPIRE].

[142] T. Sjöstrand et al., *An introduction to PYTHIA* 8.2, *Comput. Phys. Commun.* **191** (2015) 159 [arXiv:1410.3012] [INSPIRE].

[143] DELPHES 3 collaboration, *DELPHES* 3, *a modular framework for fast simulation of a generic collider experiment*, *JHEP* **02** (2014) 057 [arXiv:1307.6346] [INSPIRE].

[144] P. Azzi et al., *Report from working group* 1: *standard model physics at the HL-LHC and HE-LHC*, *CERN Yellow Rep. Monogr.* **7** (2019) 1 [arXiv:1902.04070] [INSPIRE].

[145] M. Cepeda et al., *Report from working group* 2: *Higgs physics at the HL-LHC and HE-LHC*, *CERN Yellow Rep. Monogr.* **7** (2019) 221 [arXiv:1902.00134] [INSPIRE].

[146] ATLAS collaboration, *Search for a heavy charged boson in events with a charged lepton and missing transverse momentum from pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *Phys. Rev. D* **100** (2019) 052013 [arXiv:1906.05609] [INSPIRE].

[147] R. Abdul Khalek et al., *Phenomenology of NNLO jet production at the LHC and its impact on parton distributions*, *Eur. Phys. J. C* **80** (2020) 797 [arXiv:2005.11327] [INSPIRE].

[148] M. Czakon, N.P. Hartland, A. Mitov, E.R. Nocera and J. Rojo, *Pinning down the large-x gluon with NNLO top-quark pair differential distributions*, *JHEP* **04** (2017) 044 [arXiv:1611.08609] [INSPIRE].

[149] M. Farina, C. Mondino, D. Pappadopulo and J.T. Ruderman, *New physics from high energy tops*, *JHEP* **01** (2019) 231 [arXiv:1811.04084] [INSPIRE].

[150] M. Bonvini et al., *Parton distributions with threshold resummation*, *JHEP* **09** (2015) 191 [arXiv:1507.01006] [INSPIRE].

[151] I. Brivio, *SMEFTsim* 3.0 — *a practical guide*, *JHEP* **04** (2021) 073 [arXiv:2012.11343] [INSPIRE].