The top quark legacy of the LHC Run II for PDF and SMEFT analyses

Zahari Kassabov,[a] **Maeve Madigan,**[a] **Luca Mantani,**[a] **James Moore,**[a]
Manuel Morales Alvarado,[a] **Juan Rojo**[b,c] **and Maria Ubiali**[a]

[a]*DAMTP, University of Cambridge,*
Wilberforce Road, Cambridge, CB3 0WA, U.K.

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

[c]*Nikhef Theory Group,*
Science Park 105, NL-1098 XG Amsterdam, The Netherlands

E-mail: zk261@cam.ac.uk, madigan@thphys.uni-heidelberg.de,
lm962@cam.ac.uk, jmm232@cam.ac.uk, mom28@cam.ac.uk, j.rojo@vu.nl,
M.Ubiali@damtp.cam.ac.uk

ABSTRACT: We assess the impact of top quark production at the LHC on global analyses of parton distributions (PDFs) and of Wilson coefficients in the SMEFT, both separately and in the framework of a joint interpretation. We consider the broadest top quark dataset to date containing all available measurements based on the full Run II luminosity. First, we determine the constraints that this dataset provides on the large-x gluon PDF and study its consistency with other gluon-sensitive measurements. Second, we carry out a SMEFT interpretation of the same dataset using state-of-the-art SM and EFT theory calculations, resulting in bounds on 25 Wilson coefficients modifying top quark interactions. Subsequently, we integrate the two analyses within the `SIMUnet` approach to realise a simultaneous determination of the SMEFT PDFs and the EFT coefficients and identify regions in the parameter space where their interplay is most phenomenologically relevant. We also demonstrate how to separate eventual BSM signals from QCD effects in the interpretation of top quark measurements at the LHC.

Contents

1 Introduction

The top quark is one of the most remarkable particles within the Standard Model (SM). Being the heaviest elementary particle known to date, with a mass around 185 times heavier than a proton, and the only fermion with an $\mathcal{O}(1)$ Yukawa coupling to the Higgs boson, the top quark has long been suspected to play a privileged role in potential new physics extensions beyond the Standard Model (BSM). For instance, radiative corrections involving top quarks are responsible for the so-called hierarchy problem of the SM, and the value of its mass m_t determines whether the vacuum state of our Universe is stable, metastable, or unstable [1–3]. For these reasons, since its discovery at the Tevatron in 1995 [4, 5] the properties of the top quark have been scrutinised with utmost attention and a large number of BSM searches involving top quarks as final states have been carried out. The focus on the top quark has further intensified since the start of operations at the LHC, which has realised an unprecedented top factory producing more than 200 million top quark pairs so far, for example.

In addition to this excellent potential for BSM studies, top quark production at hadron colliders also provides unique information on a variety of SM parameters such as the strong coupling constant $\alpha_s(m_t)$ [6, 7], the CKM matrix element V_{tb} [8], and the top quark mass m_t [9, 10], among several others. Furthermore, top quark production at the LHC constrains the parton distribution functions (PDFs) of the proton [11, 12], in particular the large-x gluon PDF from inclusive top quark pair production [13–15] and the quark PDF flavour separation from inclusive single top production [16, 17]. Indeed, fiducial and differential measurements of top quark pair production are part of the majority of recent PDF determinations. Reliably extracting SM parameters, including those parametrising the subnuclear structure of the proton in the PDFs, from LHC top quark production data has been made possible thanks to recent progress in higher order QCD and electroweak calculations of top quark production. Inclusive top quark pair production is now known at NNLO in the QCD expansion both for single- and double-differential distributions [18–22], eventually complemented with electroweak corrections [23], threshold resummation [24], and matching to parton showers [25]. NNLO QCD corrections are also known for single top quark production at the LHC, both in the t-channel [26, 27] and in the s-channel [28].

Even in BSM scenarios where new particles are sufficiently heavy such that direct production lies beyond the reach of the LHC, current and future measurements can still provide BSM sensitivity through low-energy signatures. These are typically revealed in the modification of SM particle properties, such as their interactions and coupling strengths. In this context, a powerful model-agnostic framework to parametrise, identify, and correlate the low-energy signatures of heavy BSM physics is the Standard Model Effective Field Theory (SMEFT). Several groups, both from the theory community and within the experimental collaborations, have presented interpretations of LHC top quark measurements in the SMEFT framework [29–42] to derive bounds on higher-dimensional EFT operators that distort the interactions of top quarks. Studies have also been performed in the framework of anomalous couplings [43, 44]. A key feature of these analyses is that the unprecedented energy reach of the LHC data increases the sensitivity to SMEFT operators via energy-growing effects entering the partonic cross-sections.

Therefore, in the LHC precision era, top quark measurements are being interpreted in (at least) two frameworks with rather different underlying assumptions. On the one hand, global PDF fits assume the SM and use top data to constrain the PDFs, producing Standard Model PDFs (denoted "SM-PDFs" in the following). On the other hand, SMEFT analyses assume that top data does not modify the SM predictions, and in particular that the proton PDFs are unchanged; we thus refer to SMEFT fits as "fixed-PDF" in the following. The two assumptions cannot be simultaneously correct, and hence one must answer two pressing questions concerning the interpretation of LHC top quark measurements. First, are SM-PDFs contaminated by BSM physics, encapsulated in the SMEFT framework, which are being reabsorbed into the fitted PDF boundary condition? Second, are the results of existing SMEFT interpretations dependent on the choice of PDFs entering the SM calculations, and is it consistent to use PDF sets that already include top quark data? It should be emphasized that for top quark production one cannot classify the data in two disjoint "SM-PDF" and "fixed-PDF" regions, since in both cases sensitivity arises from the high-energy regime.

These two questions can only be answered by means of the simultaneous determination of the PDFs and EFT coefficients from a common input dataset resulting in so-called "SMEFT-PDFs". A proof of concept of this strategy was presented for deep-inelastic scattering (DIS) data [45] and then extended to a joint analysis of DIS and Drell-Yan (DY) data [46] including projections for the HL-LHC; see also [47–49] for related work. The studies of [45, 46] were restricted to a small number of representative EFT operators, and extending them to the realistic case of processes sensitive to a large number of operators, such as top or jet production data, required the development of improved techniques. With this motivation, a new methodology dubbed `SIMUnet` was developed [50] making possible global SMEFT-PDF interpretations of LHC data suitable for processes depending on up to several tens of EFT operators. A key feature of `SIMUnet` is that it can be easily projected to both the SM-PDF case, in which it reduces to the NNPDF fitting methodology [51, 52], and to the fixed-PDF case, where it becomes equivalent to global EFT fitting tools such as SMEFiT [53].

The aim of this work is to extend the initial explorations of [45, 46] to a global determination of the SMEFT-PDFs from top quark production measurements. To this purpose, we consider the broadest top quark dataset used to date in either PDF or EFT interpretations, which in particular contains all available measurements from ATLAS and CMS based on the full Run II luminosity. By combining this wide dataset with the `SIMUnet` methodology, we derive bounds on 25 independent Wilson coefficients modifying top quark interactions, identify regions in the parameter space where the interplay between PDFs and SMEFT signatures is most phenomenologically relevant, and demonstrate how to separate eventual BSM signals from QCD effects in the interpretation of top quark measurements. As a non-trivial by-product, we also revisit the SM-PDF and fixed-PDF analyses by quantifying the information that our comprehensive top quark dataset provides. On the one hand, we assess the impact on the large-x gluon (SM-PDF), and on the other, we study the impact on the EFT coefficients (fixed-PDF), and compare our findings with related studies in the literature.

The structure of this paper is as follows. To begin with, in section 2 we describe the data inputs and the theory calculations (both in the SM and in the SMEFT) used in our study,

focusing on top quark sector measurements. The `SIMUnet` methodology deployed for the simultaneous extraction of PDFs and EFT coefficients, including its application to the fixed-PDF and SM-PDF analyses, is reviewed in section 3. Subsequently, in section 4 we present the results of the SM-PDF fits, and in particular we quantify the impact on the large-x gluon of recent high-statistics Run II measurements. In section 5 we consider the fixed-PDF analyses and present the most extensive SMEFT interpretation of top quark data from the LHC to date, including comparisons with previous results in the literature. The main results of this paper are presented in section 6, namely the simultaneous determinations of the PDFs and EFT coefficients and the comparison of these with both the fixed-PDF and SM-PDF cases. We summarise our results and outline some possible future developments in section 7.

Technical details of the analysis are collected in the appendices. Appendix A provides usage recommendations for interpretations of top quark measurements sensitive both to PDFs and SMEFT coefficients. Appendix B collects the theory settings for the SMEFT calculations, mainly concerning input schemes and operator definitions. Appendix C carries out a benchmark comparison between `SIMUnet` (in the fixed-PDF case) and the public SMEFiT code, demonstrating the agreement between the two frameworks at the linear level. The fit quality to the datasets considered in the analysis is presented in appendix D, and representative data-theory comparisons are given. Finally, in appendix E we discuss the difficulties in extending the simultaneous analysis of PDFs and EFT coefficients to the case where terms quadratic in the EFT Wilson coefficients are dominant.

2 Experimental data and theory calculations

We begin by describing the experimental data and theoretical predictions, both in the SM and in the SMEFT, used as input for the present analysis. We start in section 2.1 by describing the datasets that we consider, with emphasis on the top quark production measurements. Then in section 2.2 we use a modified version of the selection criteria defined in [51] to determine a maximally consistent dataset of top quark data to be used in the subsequent PDF and SMEFT interpretations. Finally, in section 2.3 we describe the calculation settings of the SM and SMEFT cross-sections for top quark processes, pointing the reader to the appendices for the technical details of their implementation.

2.1 Experimental data

With the exception of the top quark measurements, the dataset used in this work for fitting the PDFs both in the SM-PDF and SMEFT-PDF cases overlaps with that of the `NNPDF4.0` determination presentend in ref. [51]. In particular, the no-top variant of the `NNPDF4.0` dataset consists of 4535 data points corresponding to a wide variety of processes in deep-inelastic lepton-proton scattering [54–62] and in hadronic proton-proton collisions [63–101]; see [51] for more details.

Concerning the LHC top quark measurements considered in the present analysis, they partially overlap, but significantly extend, the top datasets included in global PDF fits such as `NNPDF4.0` [51] as well as in SMEFT analyses of the top quark sector [53, 102]. Here we discuss in turn the different types of measurements to be included: inclusive $t\bar{t}$ cross

sections and differential distributions; $t\bar{t}$ production asymmetries; the W-helicity fractions; associated top pair production with vector bosons and heavy quarks, including $\bar{t}tZ$, $\bar{t}tW$, $\bar{t}t\gamma$, $\bar{t}t\bar{t}t$, $\bar{t}t\bar{b}b$; $t-$ and $s-$channel single top production; and associated single top and vector boson production.

Choice of kinematic distribution. Many of these measurements, in particular those targeting top quark pair production, are available differentially in several kinematic variables, as well as either absolute distributions, or distributions normalised to the fiducial cross-section. We must decide which of the available kinematic distributions associated to a given measurement should be included in the fit, and whether it is more advantageous to consider absolute or normalised distributions.

Regarding the former, we note that correlations between kinematic distributions are in general not available, and only one distribution at a time can be included without double-counting (one exception is the ATLAS $t\bar{t}$ lepton+jet measurement at $\sqrt{s} = 8\,\mathrm{TeV}$ [103] where the full correlation matrix is provided). Therefore, wherever possible we include the top-pair invariant mass $m_{t\bar{t}}$ distributions with the rationale that these have enhanced sensitivity to SMEFT operators via energy-growing effects; they also provide direct information on the large-x PDFs. Otherwise, we consider the top or top-pair rapidity distributions, y_t and $y_{t\bar{t}}$ respectively, which also provide the sought-for information on the large-x PDFs; furthermore they benefit from moderate higher-order QCD and electroweak corrections [14].

Regarding the choice of absolute versus normalised distributions, we elect to use normalised distributions together with corresponding fiducial cross-sections throughout. Normalised distributions are typically more precise that their absolute counterparts, since experimental and theoretical errors partially cancel out when normalising. In addition, normalisation does not affect the PDF and EFT sensitivity of the measurement, provided the fiducial cross section measurements used for normalising are also accounted for. From the implementation point of view, since in a normalised measurement one bin is dependent on the others, we choose to exclude the bin with lowest $m_{t\bar{t}}$ value (the production threshold) to avoid losing sensitivity arising from the high-energy tails.

Inclusive $t\bar{t}$ production. A summary of the inclusive $t\bar{t}$ fiducial cross sections and differential distributions considered in this work is provided in table 1. We indicate in each case the centre of mass energy $\sqrt{s}$, the final-state channel, the observable(s) used in the fit, the luminosity, and the number of data points n_{dat}, together with the corresponding publication reference. In the last two columns, we indicate with a ✓ the datasets that are included for the first time here in a global PDF fit (specifically, those which are new with respect to `NNPDF4.0`) and in a SMEFT interpretation (specifically, in comparison with the global fits of [53, 102]). The sets marked with brackets have already been included in previous studies, but are implemented here in a different manner (e.g. by changing spectra or normalisation), as indicated in the table; more details are given in each paragraph of the section.

The ATLAS dataset comprises six total cross section measurements and five differential normalised cross section measurements. Concerning the latter, at 8 TeV we include three distributions from the dilepton and ℓ+jets channels. In the ℓ+jets channel, several kinematic distributions are available together with their correlations. Following the dataset

Exp.	$\sqrt{s}$ (TeV)	Channel	Observable	$\mathcal{L}$ (fb^{-1})	n$_{\text{dat}}$	Ref.	New (PDF fits)	New (SMEFT fits)
ATLAS	7	dilepton	$\sigma(t\bar{t})$	4.6	1	[104]		(✓)
	8	dilepton	$\sigma(t\bar{t})$	20.3	1	[104]		(✓)
			$1/\sigma d\sigma/dm_{t\bar{t}}$	20.2	5	[105]	$(y_{t\bar{t}} \to m_{t\bar{t}})$	(absolute → ratio)
		ℓ+jets	$\sigma(t\bar{t})$	20.2	1	[106]	✓	(✓)
			$1/\sigma d\sigma/d\lvert y_t\rvert$	20.3	4	[103]		$(m_{t\bar{t}}, p_t^T \to \lvert y_t\rvert, \lvert y_{t\bar{t}}\rvert)$
			$1/\sigma d\sigma/d\lvert y_{t\bar{t}}\rvert$	20.3	4	[103]		$(m_{t\bar{t}}, p_t^T \to \lvert y_t\rvert, \lvert y_{t\bar{t}}\rvert)$
	13	dilepton	$\sigma(t\bar{t})$	36.1	1	[107]	✓	✓
		hadronic	$\sigma(t\bar{t})$	36.1	1	[108]	✓	✓
			$1/\sigma d^2\sigma/d\lvert y_{t\bar{t}}\rvert dm_{t\bar{t}}$	36.1	10	[108]	✓	✓
		ℓ+jets	$\sigma(t\bar{t})$	139	1	[109]		(✓)
			$1/\sigma d\sigma/dm_{t\bar{t}}$	36	8	[110]	✓	(absolute → ratio)
CMS	5	combination	$\sigma(t\bar{t})$	0.027	1	[111]		✓
	7	combination	$\sigma(t\bar{t})$	5.0	1	[112]		✓
	8	combination	$\sigma(t\bar{t})$	19.7	1	[112]		✓
		dilepton	$1/\sigma d^2\sigma/dy_{t\bar{t}}dm_{t\bar{t}}$	19.7	16	[113]	$(m_{t\bar{t}}, y_t \to m_{t\bar{t}}, y_{t\bar{t}})$	
		ℓ+jets	$1/\sigma d\sigma/dy_{t\bar{t}}$	19.7	9	[114]		
	13	dilepton	$\sigma(t\bar{t})$	43	1	[115]		(✓)
			$1/\sigma d\sigma/dm_{t\bar{t}}$	35.9	5	[116]		(absolute → ratio)
		ℓ+jets	$\sigma(t\bar{t})$	137	1	[117]	✓	✓
			$1/\sigma d\sigma/dm_{t\bar{t}}$	137	14	[117]	✓	✓

Table 1. The inclusive cross-sections and differential distributions for top quark pair production from ATLAS and CMS that we consider in this analysis. For each dataset, we indicate the experiment, the centre of mass energy $\sqrt{s}$, the final-state channel, the observable(s) used in the fit, the integrated luminosity $\mathcal{L}$ in inverse femtobarns, and the number of data points n_{dat}, together with the corresponding publication reference. In the last two columns, we indicate with a ✓ the datasets that are included for the first time here in a global PDF fit and in a SMEFT interpretation, respectively. The sets marked with brackets have already been included in previous studies but here we account for their constraints in different manner (e.g. by changing spectra or normalisation), as indicated in the table and in the text description.

selection analysis carried out in [51], we select to fit the y_t and $y_{t\bar{t}}$ distributions as done in the NNPDF4.0 baseline. At 13 TeV, we include the normalised cross sections differential in $m_{t\bar{t}}$ from the ℓ+jets and hadronic channels, with both measurements being considered for the first time here in the context of a PDF analysis.

Moving to CMS, in the inclusive $t\bar{t}$ category we consider five total cross section and four normalised differential cross section measurements. At $\sqrt{s} = 8\,\text{TeV}$ we include differential distributions in the ℓ+jets and dilepton channels, the latter being doubly differential in $y_{t\bar{t}}$ and $m_{t\bar{t}}$. The double-differential 8 TeV measurement is part of NNPDF4.0 , but there the $(y_t, m_{t\bar{t}})$ distribution was fitted instead. At 13 TeV, we include the $m_{t\bar{t}}$ distributions in

Experiment	$\sqrt{s}$(TeV)	Channel	Observable	$\mathcal{L}$ (fb^{-1})	n_{dat}	Ref.	New (PDF fits)	New (SMEFT fits)
ATLAS	8	dilepton	A_C	20.3	1	[118]	✓	
	13	ℓ+jets	A_C	139	5	[119]	✓	✓
CMS	8	dilepton	A_C	19.5	3	[120]	✓	
	13	ℓ+jets	A_C	138	3	[121]	✓	
ATLAS & CMS combination	8	ℓ+jets	A_C	20	6	[122]	✓	

Table 2. Same as table 1 for the $t\bar{t}$ asymmetry datasets.

the dilepton and ℓ+jets channels. In the latter case we include the single $m_{t\bar{t}}$ distribution rather than the double-differential one in $(m_{t\bar{t}}, y_{t\bar{t}})$, which is also available, since we find that the latter cannot be reproduced by the NNLO SM predictions. We present a dedicated analysis of the double-differential distribution in section 5.3. As mentioned above, we will study the impact of our dataset selection choices by presenting variations of the baseline SM-PDF, fixed-PDF, and SMEFT-PDF analyses in the following sections.

$t\bar{t}$ asymmetry measurements. The $t\bar{t}$ production asymmetry at the LHC is defined as:

$$A_C = \frac{N(\Delta|y| > 0) - N(\Delta|y| < 0)}{N(\Delta|y| > 0) + N(\Delta|y| < 0)}, \tag{2.1}$$

with $N(P)$ being the number of events satisfying the kinematical condition P, and $\Delta|y| = |y_t| - |y_{\bar{t}}|$ is the difference between the absolute values of the top quark and anti-top quark rapidities. The asymmetry A_C can be measured either integrating over the fiducial phase space or differentially, for example binning in the invariant mass $m_{t\bar{t}}$. Measurements of A_C are particularly important in constraining certain SMEFT directions, in particular those associated to the two-light-two-heavy operators. However, they are unlikely to have an impact on PDF fitting due to their large experimental uncertainties; nevertheless, with the underlying motivation of a comprehensive SMEFT-PDF interpretation of top quark data, we consider here the A_C measurement as part of our baseline dataset, and hence study whether or not they also provide relevant PDF information. A summary of the asymmetry measurements included in this work is given in table 2.

W-helicity fractions. The W-helicity fractions F_0, F_L and F_R are PDF-independent observables sensitive to SMEFT corrections, and the dependence of the theory predictions with respect to the Wilson coefficients can be computed analytically. Since these W-helicity fractions are PDF-independent observables, to include them in the joint SMEFT-PDF analysis one has to extend the methodology presented in [50] to include in the fit datasets that either lack, or have negligible, PDF sensitivity and depend only on the EFT coefficients. We describe how this can be achieved within the `SIMUnet` framework in section 3.

In table 3 we list the LHC measurements of the W-helicity fractions considered in the current analysis. At $\sqrt{s} = 8\,\text{TeV}$ we include the combined ATLAS and CMS measurement from [123], while at 13 TeV we consider the ATLAS measurement of the W-helicities from [124], for the first time in a SMEFT fit.

Experiment	$\sqrt{s}$(TeV)	Observable	$\mathcal{L}$ (fb^{-1})	n_{dat}	Ref.	New (SMEFT fits)
ATLAS & CMS combination	8	F_0, F_L	20	2	[123]	
ATLAS	13	F_0, F_L	139	2	[124]	✓

Table 3. Same as table 1 for the W-helicity fraction measurements. These helicity fractions are PDF-independent and hence are only relevant in constraining the EFT coefficients.

Exp.	$\sqrt{s}$ (TeV)	Observable	$\mathcal{L}$ (fb^{-1})	n_{dat}	Ref.	New (PDF fits)	New (SMEFT fits)
ATLAS	8	$\sigma(t\bar{t}Z)$	20.3	1	[125]	✓	
		$\sigma(t\bar{t}W)$	20.3	1	[125]	✓	
	13	$\sigma(t\bar{t}Z)$	36.1	1	[126]	✓	
		$1/\sigma d\sigma(t\bar{t}Z)/dp_T^Z$	139	6	[127]	✓	✓
		$\sigma(t\bar{t}W)$	36.1	1	[126]	✓	
CMS	8	$\sigma(t\bar{t}Z)$	19.5	1	[128]	✓	
		$\sigma(t\bar{t}W)$	19.5	1	[128]	✓	
	13	$\sigma(t\bar{t}Z)$	35.9	1	[129]	✓	
		$1/\sigma d\sigma(t\bar{t}Z)/dp_T(Z)$	77.5	3	[130]	✓	(absolute → ratio)
		$\sigma(t\bar{t}W)$	35.9	1	[129]	✓	

Table 4. Same as table 1 for the measurements of top quark production in association with a vector boson.

Experiment	$\sqrt{s}$(TeV)	Observable	$\mathcal{L}$ (fb^{-1})	n_{dat}	Ref.	New (SMEFT fits)
ATLAS	8	$\sigma(t\bar{t}\gamma)$	20.2	1	[131]	
CMS	8	$\sigma(t\bar{t}\gamma)$	19.7	1	[132]	

Table 5. Same as table 1 for $t\bar{t}$ production in association with a photon. Theory predictions for these observables adopt a fixed PDF.

Associated top quark pair production. The next class of observables that we discuss is associated $t\bar{t}$ production with a Z- or a W-boson (table 4), a photon γ (table 5), or a heavy quark pair ($t\bar{t}b\bar{b}$ or $t\bar{t}t\bar{t}$, table 6). While measurements of $t\bar{t}V$ have been considered for SMEFT interpretations, we use them for the first time here in the context of a PDF determination. The rare processes $t\bar{t}\gamma$, $t\bar{t}b\bar{b}$, and $t\bar{t}t\bar{t}$ exhibit a very weak PDF sensitivity and hence in the present analysis their theory predictions are obtained using a fixed PDF, in the same manner as the W-helicity fractions in table 3.

Concerning the $t\bar{t}Z$ and $t\bar{t}W$ data, from both ATLAS and CMS we use four fiducial cross section measurements at 8 TeV and 13 TeV, and one distribution differential in p_T^Z at 13 TeV. These measurements are particularly interesting to probe SMEFT coefficients that modify the interactions between the top quark and the electroweak sector. For top-quark production associated with a photon, we include the fiducial cross-section measurements

Experiment	$\sqrt{s}$(TeV)	Channel	Observable	$\mathcal{L}$ (fb^{-1})	n_{dat}	Ref.	New (SMEFT fits)
ATLAS	13	multi-lepton	$\sigma_{\text{tot}}(t\bar{t}t\bar{t})$	139	1	[133]	
		single-lepton	$\sigma_{\text{tot}}(t\bar{t}t\bar{t})$	139	1	[134]	✓
		ℓ+jets	$\sigma_{\text{tot}}(t\bar{t}b\bar{b})$	36.1	1	[135]	
CMS	13	multi-lepton	$\sigma_{\text{tot}}(t\bar{t}t\bar{t})$	137	1	[136]	
		single-lepton	$\sigma_{\text{tot}}(t\bar{t}t\bar{t})$	35.8	1	[137]	
		all-jet	$\sigma_{\text{tot}}(t\bar{t}b\bar{b})$	35.9	1	[138]	
		dilepton	$\sigma_{\text{tot}}(t\bar{t}b\bar{b})$	35.9	1	[139]	
		ℓ+jets	$\sigma_{\text{tot}}(t\bar{t}b\bar{b})$	35.9	1	[139]	✓

Table 6. Same as table 1 for the measurements of $t\bar{t}$ production in association with a heavy quark pair. Theory predictions for these observables adopt a fixed PDF.

from ATLAS and CMS at 8 TeV; also available is a differential distribution at 13 TeV from ATLAS binned in the photon transverse momentum p_T^γ [140], but we exclude this from our analysis because of the difficulty in producing SMEFT predictions in the fiducial phase space (in the FitMaker analysis, its inclusion is only approximate, and in SMEFiT this distribution is neglected entirely). Finally, we include fiducial measurements of $t\bar{t}b\bar{b}$ and $t\bar{t}t\bar{t}$ production at 13 TeV considering the data with highest luminosity for each available final state.

Inclusive single-top pair production. The inclusive single-top production data considered here and summarised in table 7 comprises measurements of single-top production in the t-channel, which have previously been included in PDF fits [16, 51], as well as measurements of single-top production in the s-channel, which in the context of PDF studies have been implemented for the first time in this study. For t-channel production, we consider the ATLAS and CMS top and anti-top fiducial cross sections $\sqrt{s} = 7, 8$, and 13 TeV, as well as normalised y_t and $y_{\bar{t}}$ distributions at 7 and 8 TeV (ATLAS) and at 13 TeV (CMS). For s-channel production, no differential measurements are available and hence we consider fiducial cross-sections at 8 and 13 TeV from ATLAS and CMS.

Associated single top-quark production with weak bosons. Finally, table 8 lists the measurements of associated single-top production with vector bosons included in our analysis. We consider fiducial cross-sections for tW production at 8 and 13 TeV from ATLAS and CMS in the dilepton and single-lepton final states, as well as the tZj fiducial cross-section at 13 TeV from ATLAS and CMS in the dilepton final state. In addition, kinematical distributions in tZj production from CMS at 13 TeV are considered for the first time here in an EFT fit. For these differential distributions, the measurement is presented binned in either p_T^Z or p_T^t; here, we take the former as default for consistency with the corresponding $t\bar{t}Z$ analysis.

Exp.	$\sqrt{s}$ (TeV)	Channel	Observable	$\mathcal{L}$ (fb^{-1})	$\mathrm{n_{dat}}$	Ref.	New (PDF fits)	New (SMEFT fits)
ATLAS	7	t-channel	$\sigma_{\text{tot}}(t)$	4.59	1	[141]	($\checkmark$)	$\checkmark$
			$\sigma_{\text{tot}}(\bar{t})$	4.59	1	[141]	($\checkmark$)	$\checkmark$
			$1/\sigma d\sigma(tq)/dy_t$	4.59	3	[141]		$\checkmark$
			$1/\sigma d\sigma(\bar{t}q)/dy_{\bar{t}}$	4.59	3	[141]		$\checkmark$
	8	t-channel	$\sigma_{\text{tot}}(t)$	20.2	1	[142]	($\checkmark$)	$\checkmark$
			$\sigma_{\text{tot}}(\bar{t})$	20.2	1	[142]	($\checkmark$)	$\checkmark$
			$1/\sigma d\sigma(tq)/dy_t$	20.2	3	[142]		($\checkmark$)
			$1/\sigma d\sigma(\bar{t}q)/dy_{\bar{t}}$	20.2	3	[142]		($\checkmark$)
		s-channel	$\sigma_{\text{tot}}(t+\bar{t})$	20.3	1	[143]	$\checkmark$	
	13	t-channel	$\sigma_{\text{tot}}(t)$	3.2	1	[144]	($\checkmark$)	
			$\sigma_{\text{tot}}(\bar{t})$	3.2	1	[144]	($\checkmark$)	
		s-channel	$\sigma_{\text{tot}}(t+\bar{t})$	139	1	[145]	$\checkmark$	$\checkmark$
CMS	7	t-channel	$\sigma_{\text{tot}}(t)+\sigma_{\text{tot}}(\bar{t})$	1.17, 1.56	1	[146]		$\checkmark$
	8	t-channel	$\sigma_{\text{tot}}(t)$	19.7	1	[147]	($\checkmark$)	
			$\sigma_{\text{tot}}(\bar{t})$	19.7	1	[147]	($\checkmark$)	
		s-channel	$\sigma_{\text{tot}}(t+\bar{t})$	19.7	1	[148]	$\checkmark$	
	13	t-channel	$\sigma_{\text{tot}}(t)$	2.2	1	[149]	($\checkmark$)	
			$\sigma_{\text{tot}}(\bar{t})$	2.2	1	[149]	($\checkmark$)	
			$1/\sigma d\sigma/d|y^{(t)}|$	35.9	4	[150]	$\checkmark$	

Table 7. Same as table 1 for the inclusive single-top production datasets.

Experiment	$\sqrt{s}$(TeV)	Channel	Observable	$\mathcal{L}$ (fb^{-1})	$\mathrm{n_{dat}}$	Ref.	New (SMEFT fits)
ATLAS	8	dilepton	$\sigma_{\text{tot}}(tW)$	20.3	1	[151]	
		single-lepton	$\sigma_{\text{tot}}(tW)$	20.2	1	[152]	
	13	dilepton	$\sigma_{\text{tot}}(tW)$	3.2	1	[153]	
		dilepton	$\sigma_{\text{fid}}(tZj)$	139	1	[154]	
CMS	8	dilepton	$\sigma_{\text{tot}}(tW)$	12.2	1	[155]	
	13	dilepton	$\sigma_{\text{tot}}(tW)$	35.9	1	[156]	
		dilepton	$\sigma_{\text{fid}}(tZj)$	77.4	1	[157]	
		dilepton	$d\sigma_{\text{fid}}(tZj)/dp_T^t$	138	3	[158]	$\checkmark$
		single-lepton	$\sigma_{\text{tot}}(tW)$	36	1	[159]	$\checkmark$

Table 8. Same as table 1 for single-top production in association with an electroweak bosons.

2.2 Dataset selection

The top quark production measurements listed in tables 1–8 summarise all datasets that have been considered for the present analysis. In principle, however, some of these may need to be excluded from the baseline fit dataset to ensure that the baseline dataset is maximally consistent. Following the dataset selection procedure adopted in [51], here our baseline dataset is chosen to exclude datasets that may be either internally inconsistent or inconsistent with other measurements of the same process type. These inconsistencies can be of experimental origin, for instance due to unaccounted (or underestimated) systematic errors, or numerically unstable correlation models, as well as originating in theory, for example whenever a given process is affected by large missing higher-order perturbative uncertainties. Given that the ultimate goal of a global SMEFT analysis, such as the present one, is to unveil deviations from the SM, one should strive to deploy objective dataset selection criteria that exclude datasets affected by such inconsistencies, which are unrelated to BSM physics.

The first step is to run a global SM-PDF fit including all the datasets summarised in tables 1–8 (and additionally a fit with the data summarised therein, but with the CMS measurement of the differential $t\bar{t}$ cross-section at 13 TeV in the ℓ+jets channel replaced with the double-differential measurement) and monitor in each case the following two statistical estimators:

- The total χ^2 per data point and the number of standard deviations n_σ by which the value of the χ^2 per data point differs from the median of the χ^2 distribution for a perfectly consistent dataset,

$$n_\sigma \equiv \frac{|\chi^2 - 1|}{\sigma_{\chi^2}} = \frac{|\chi^2 - 1|}{\sqrt{2/n_{\rm dat}}}, \qquad (2.2)$$

where the χ^2 in this case (and in the rest of the paper unless specified) is the experimental χ^2 per data point, which is defined as

$$\chi^2 \equiv \chi^2_{\rm exp}/n_{\rm dat} = \frac{1}{n_{\rm dat}} \sum_{i,j=1}^{n_{\rm dat}} (D_i - T_i^0) \left({\rm cov}_{\rm exp}^{-1}\right)_{ij} (D_j - T_j^0), \qquad (2.3)$$

where T_i^0 are the theoretical predictions computed with the central PDF replica, which is the average over the PDF replicas, and the experimental covariance matrix is the one defined for example in eq. (3.1) of ref. [160].

Specifically, we single out for further examination datasets for which $n_\sigma \geq 3$ and $\chi^2 \geq 2$ per data point, where the poor description of the data is unlikely to be caused by a statistical fluctuation (note that these conditions relax those given in [51], which we hope gives the opportunity for the EFT to account for poor quality fits to data, rather than immediately attributing poor fits to inconsistencies). The question is then to ascertain whether this poor χ^2 can be explained by non-zero EFT coefficients (and in such case it should be retained for the fit) or if instead there one can find other explanations, such as the ones mentioned above, that justify removing it from the baseline dataset.

- The metric Z defined in ref. [161] which quantifies the stability of the χ^2 with respect to potential inaccuracies affecting the modelling of the experimental correlations. The calculation of Z relies exclusively on the experimental covariance matrix and is independent of the theory predictions. A large value of the stability metric Z corresponds to datasets with an unstable covariance matrix, in the sense that small changes in the values of the correlations between data points lead to large increases in the corresponding χ^2. Here we single out for further inspection datasets with $Z \geq 4$.

  As also described in [161], it is possible to regularise covariance matrices in a minimal manner to assess the impact of these numerical instabilities at the PDF or SMEFT fit level, and determine how they affect the resulting pre- and post-fit χ^2. To quantify whether datasets with large Z distort the fit results in a sizable manner, one can run fit variants applying this decorrelation procedure such that all datasets exhibit a value of the Z-metric below the threshold. We do not find it necessary to run such fits in this work.

In tables 9 and 10 we list the outcome of such a global SM-PDF fit, where entries that lie above the corresponding threshold values for χ^2, n_σ, or Z are highlighted in boldface. In the last column, we indicate whether the dataset is flagged. For the flagged datasets, we carry out the following tests to ascertain whether it should be retained in the fit:

- For datasets with $n_\sigma > 3$ and $Z > 4$, we run a fit variant in which the covariance matrix is regularised. If, upon regularisation of the covariance matrix, the PDFs are stable and both the χ^2 per data point and the $|n_\sigma|$ decrease to a value below the respective thresholds of 2.0 and 3.0, we retain the dataset, else we exclude it.

- For datasets with $\chi^2 > 2.0$ and $n_\sigma > 3$ we carry out a fit variant where this dataset is given a very high weight. If in this high-weight fit variant the χ^2 and n_σ estimators improve to the point that their values lie below the thresholds without deteriorating the description of any of the other datasets included the dataset is kept, then the specific measurement is not inconsistent, it just does not have enough weight compared to the other datasets. See ref. [51] for a detailed discussion on the size of the weight depending on the size of the dataset.

From the analysis of tables 9 and 10, one finds that only two datasets in the inclusive top quark pair production (lepton+jets final state) category are flagged as potentially problematic: the ATLAS $|y_{t\bar{t}}|$ distribution at 8 TeV and the CMS double-differential distributions in $m_{t\bar{t}}$ and $y_{t\bar{t}}$ at 13 TeV. The first of these was already discussed in the `NNPDF4.0` analysis [51]. It was observed that each of the four distributions measured by ATLAS and presented in ref. [103] behave somewhat differently upon being given large weight. The χ^2 of all distributions significantly improves when given large weight. However, while for the top transverse momentum and top pair invariant mass distributions this improvement is accompanied by a rather significant deterioration of the global fit quality, in the case of the top and top pair rapidity distributions the global fit quality is very similar and only the description of jets deteriorates moderately. The rapidity distributions thus remain largely compatible with the rest of the dataset, hence they are kept.

Experiment	$\sqrt{s}$ (TeV)	Observable, Channel	n_{dat}	$\chi^2_{\mathrm{exp}}/n_{\mathrm{dat}}$	n_σ	Z	flag
ATLAS	7	$\sigma^{\mathrm{tot}}_{t\bar t}$, dilepton	1	**4.63**	2.57	1.00	no
		$\sigma^{\mathrm{tot}}(t)$, t-channel	1	0.76	-0.17	1.00	no
		$\sigma^{\mathrm{tot}}(\bar t)$, t-channel	1	0.29	-0.50	1.00	no
		$1/\sigma\, d(tq)/dy_t$, t-channel	3	0.97	-0.04	1.28	no
		$1/\sigma\, d(\bar tq)/dy_{\bar t}$, t-channel	3	0.06	-1.15	1.39	no
	8	$\sigma^{\mathrm{tot}}_{t\bar t}$, dilepton	1	0.03	-0.69	1.00	no
		$1/\sigma\, d\sigma/dm_{t\bar t}$, dilepton	5	0.29	-1.12	1.61	no
		$\sigma^{\mathrm{tot}}_{t\bar t}$, ℓ+jets	1	0.28	-0.51	1.00	no
		$1/\sigma\, d\sigma/d|y_t|$, ℓ+jets	4	**2.86**	2.63	1.65	no
		$1/\sigma\, d\sigma/d|y_{t\bar t}|$, ℓ+jets	4	**3.37**	**3.35**	2.19	**yes (kept)**
		A_C, dilepton	1	0.67	-0.23	1.00	no
		$\sigma(t\bar tZ)$	1	0.23	-0.54	1.00	no
		$\sigma(t\bar tW)$	1	**2.44**	1.01	1.00	no
		$\sigma^{\mathrm{tot}}(t+\bar t)$, s-channel	1	0.21	-0.56	1.00	no
		$\sigma^{\mathrm{tot}}(tW)$, dilepton	1	0.54	-0.33	1.00	no
		$\sigma^{\mathrm{tot}}(tW)$, single-lepton	1	0.71	-0.21	1.00	no
	13	$\sigma^{\mathrm{tot}}_{t\bar t}$, dilepton	1	1.41	0.29	1.00	no
		$\sigma^{\mathrm{tot}}_{t\bar t}$, hadronic	1	0.23	-0.54	1.000	no
		$1/\sigma\, d^2\sigma/d|y_{t\bar t}|dm_{t\bar t}$, hadronic	10	1.95	2.12	2.33	no
		$\sigma^{\mathrm{tot}}_{t\bar t}$, ℓ+jets	1	0.50	-0.35	1.00	no
		$1/\sigma\, d\sigma/dm_{t\bar t}$, ℓ+jets	8	1.83	1.66	**7.61**	no
		A_C, ℓ+jets	5	0.99	-0.02	1.41	no
		$\sigma(t\bar tZ)$	1	0.75	-0.18	1.00	no
		$1/\sigma\, d\sigma(t\bar tZ)/dp_T(Z)$	5	1.93	1.47	2.27	no
		$\sigma(t\bar tW)$	1	1.43	0.30	1.00	no
		$\sigma^{\mathrm{tot}}(t)$, t-channel	1	0.72	-0.20	1.00	no
		$\sigma^{\mathrm{tot}}(\bar t)$, t-channel	1	0.39	-0.43	1.00	no
		$\sigma^{\mathrm{tot}}(t+\bar t)$, s-channel	1	0.70	-0.21	1.00	no
		$\sigma^{\mathrm{tot}}(tW)$, dilepton	1	1.15	0.36	1.00	no

Table 9. For the ATLAS measurements that we consider in this work, we list the outcome of a global SM-PDF fit with all measurements listed in tables 1–8 included. We display for each dataset the number of data points, the χ^2 per data point (eq. (2.3)), the number of standard deviations n_σ (eq. (2.2)), and the stability metric Z defined in [161]. The entries that lie above the corresponding threshold values are highlighted in boldface, In the last column, we indicate whether the dataset is flagged and is either kept or removed. See text for more details.

Experiment	$\sqrt{s}$ (TeV)	Observable	n_{dat}	$\chi^2_{\mathrm{exp}}/n_{\mathrm{dat}}$	n_σ	Z	flag
CMS	5	$\sigma^{\mathrm{tot}}_{t\bar{t}}$, combination	1	0.56	-0.31	1.00	no
	7	$\sigma^{\mathrm{tot}}_{t\bar{t}}$, combination	1	1.08	0.06	1.00	no
		$\sigma^{\mathrm{tot}}(t) + \sigma^{\mathrm{tot}}(\bar{t})$, t-channel	1	0.72	-0.20	1.00	no
	8	$\sigma^{\mathrm{tot}}_{t\bar{t}}$, combination	1	0.27	-0.52	1.00	no
		$1/\sigma d^2\sigma/dy_{t\bar{t}}dm_{t\bar{t}}$, dilepton	16	0.98	-0.06	2.33	no
		$1/\sigma d\sigma/dy_{t\bar{t}}$, ℓ+jets	9	1.15	0.31	1.63	no
		A_C, dilepton	3	0.05	-1.16	1.16	no
		$\sigma(t\bar{t}Z)$	1	0.47	-0.37	1.00	no
		$\sigma(t\bar{t}W)$	1	**2.27**	0.90	1.00	no
		$\sigma^{\mathrm{tot}}(t)$, t-channel	1	0.01	-0.70	1.00	no
		$\sigma^{\mathrm{tot}}(\bar{t})$, t-channel	1	0.09	-0.64	1.00	no
		$\sigma^{\mathrm{tot}}(t+\bar{t})$, s-channel	1	1.11	0.08	1.00	no
		$\sigma^{\mathrm{tot}}(tW)$, dilepton	1	0.38	-0.44	1.00	no
	13	$\sigma^{\mathrm{tot}}_{t\bar{t}}$, dilepton	1	0.06	-0.66	1.00	no
		$1/\sigma d\sigma/dm_{t\bar{t}}$, dilepton	5	**2.49**	2.36	1.61	no
		$\sigma^{\mathrm{tot}}_{t\bar{t}}$, ℓ+jets channel	1	0.22	-0.55	1.00	no
		$1/\sigma d\sigma/dm_{t\bar{t}}$, ℓ+jets	14	1.41	1.08	**4.57**	no
		$1/\sigma d\sigma/dm_{t\bar{t}}dy_t$, ℓ+jets	34	**6.43**	22.4	3.88	**yes (excl)**
		A_C, ℓ+jets	3	0.29	-0.87	1.00	no
		$\sigma(t\bar{t}Z)$	1	1.24	0.17	1.00	no
		$1/\sigma d\sigma(t\bar{t}Z)/dp_T(Z)$	3	0.59	-0.50	1.28	no
		$\sigma(t\bar{t}W)$	1	0.66	-0.24	1.00	no
		$\sigma^{\mathrm{tot}}(t)$, t-channel	1	0.88	-0.08	1.00	no
		$\sigma^{\mathrm{tot}}(\bar{t})$, t-channel	1	0.13	-0.62	1.00	no
		$1/\sigma d\sigma/d|y^{(t)}|$, t-channel	4	0.38	-0.88	1.70	no
		$\sigma^{\mathrm{tot}}(tW)$, dilepton	1	0.43	-0.40	1.00	no
		$\sigma^{\mathrm{tot}}(tW)$, single-lepton	1	**2.84**	1.30	1.00	no
ATLAS-CMS combination	8	A_C, ℓ+jets	6	0.602	-0.69	1.65	no

Table 10. Same as table 9 for the CMS and combined ATLAS-CMS datasets. *Note carefully:* the row corresponding to the CMS doubly-differential distribution at 13 TeV in the ℓ+jets channel comes from a separate fit, where the corresponding 1D distribution is replaced by this dataset.

Also shown in one row of table 10 is the fit-quality information for the CMS double-differential distribution at 13 TeV in the ℓ+jets channel, from a separate fit wherein the CMS single differential distribution at 13 TeV in the ℓ+jets channel is replaced by this dataset. We find that the 2D set is described very poorly, with a $\chi^2 = 6.43$, corresponding to a 22σ deviation from the median of the χ^2 distribution for a perfectly consistent dataset. To investigate this further, we performed a weighted fit; however, we find that the χ^2 improves only moderately (from $\chi^2 = 6.43$ to $\chi^2 = 4.56$) and moreover the χ^2-statistic of the other datasets deteriorates significantly (with total χ^2 jumping from 1.20 to 1.28). The test indicates that the double-differential distribution is both incompatible with the rest of the data and also internally inconsistent given the standard PDF fit. Hence we exclude this dataset from our baseline and include instead the single-differential distribution in $m_{t\bar{t}}$, which is presented in the same publication [117] and is perfectly described in the baseline fit. To check whether the incompatibility we observe in the double-differential distribution can be cured by the inclusion of SMEFT corrections, we will run a devoted analysis presented in section 5.3.

2.3 Theoretical predictions

In this section we describe the calculation settings adopted for the SM and SMEFT cross-sections used in the present analysis.

SM cross-sections. Theoretical predictions for SM cross-sections are evaluated at NNLO in perturbative QCD, whenever available, and at NLO otherwise. Predictions accurate to NLO QCD are obtained in terms of fast interpolation grids from MAD-GRAPH5_AMC@NLO [162, 163], interfaced to APPLGRID [164] or FASTNLO [165–167] together with AMCFAST [168] and APFELCOMB [169]. Wherever available, NNLO QCD corrections to matrix elements are implemented by multiplying the NLO predictions by bin-by-bin K-factors, see section 2.3 in [170]. The top mass is set to $m_t = 172.5$ GeV for all processes considered.

In the case of inclusive $t\bar{t}$ cross sections and charge asymmetries, a dynamical scale choice of $\mu_R = \mu_F = H_T/4$ is adopted, where H_T denotes the sum of the transverse masses of the top and anti-top, following the recommendations of ref. [18]. This scale choice ensures that the ratio of fixed order NNLO predictions to the NNLO+NNLL ones is minimised, allowing us to neglect theory uncertainties associated to missing higher orders beyond NNLO. To obtain the corresponding NNLO K-factors, we use the HIGHTEA public software [171], an event database for distributing and analysing the results of fixed order NNLO calculations for LHC processes. The NNLO PDF set used in the computation of these K-factors is either `NNPDF3.1` or `NNPDF4.0`, depending on whether a given dataset was already included in the `NNPDF4.0` global fit or not, respectively.

For associated $t\bar{t}$ and W or Z production, dedicated fast NLO grids have been generated. Factorisation and renormalisation scales are fixed to $\mu_F = \mu_R = m_t + \frac{1}{2}m_V$, where $m_V = m_W, m_Z$ is the mass of the associated weak boson, as appropriate. This scale choice follows the recommendation of ref. [172] and minimises the ratio of the NLO+NLL over the fixed-order NLO prediction. We supplement the predictions for the total cross section for

associated W and Z-production at $13\,\mathrm{TeV}$ with NLO+NNLL QCD K-factors taken from table 1 of [172]. On the other hand, the $t\bar{t}\gamma$, $t\bar{t}t\bar{t}$ and $t\bar{t}b\bar{b}$ data are implemented as PDF independent observables, and the corresponding theory predictions are taken directly from the relevant experimental papers in each case.

The evaluation of theoretical predictions for single top production follows [16]. Fast NLO interpolation grids are generated for both s- and t-channel single top-quark and top-antiquark datasets in the 5-flavour scheme, with fixed factorisation and renormalisation scales set to m_t. Furthermore, for the t-channel production we include the NNLO QCD corrections to both total and differential cross sections [26]. When the top decay is calculated, it is done in the narrow-width approximation, under which the QCD corrections to the top-(anti)quark production and the decay are factorisable and the full QCD corrections are approximated by the vertex corrections.

SMEFT cross-sections. SMEFT corrections to SM processes are computed both at LO and at NLO in QCD, and both at the linear and the quadratic level in the EFT expansion. Flavour assumptions follow the LHC TOP WG prescription of [29] which were also used in the recent SMEFiT analysis [53]. The flavour symmetry group is given by $\mathrm{U}(3)_l \times \mathrm{U}(3)_e \times \mathrm{U}(3)_d \times \mathrm{U}(2)_u \times \mathrm{U}(2)_q$, i.e. we single out operators that contain top quarks (right-handed t and SU(2) doublet Q). This also means that one works in a five-flavour scheme in which the only massive fermion in the theory is the top. As far as the electroweak input scheme is concerned, we work in the m_W-scheme, meaning that the 4 electroweak inputs are $\{m_W, G_F, m_h, m_Z\}$.

At dimension-six, SMEFT operators modify the SM Lagrangian as:

$$\mathcal{L}_{\mathrm{SMEFT}} = \mathcal{L}_{\mathrm{SM}} + \sum_{n=1}^{N} \frac{c_n}{\Lambda^2} \mathcal{O}_n \,, \tag{2.4}$$

where Λ is the UV-cutoff energy scale, $\{\mathcal{O}_n\}$ are dimension-six operators, and $\{c_n\}$ are Wilson coefficients. The 25 operators considered for this study are listed in table 14 in the Warsaw basis [173]. In this work we neglect renormalisation group effects on the Wilson coefficients [174]. For hadronic data, i.e. for proton-proton collisions, which are the only data affected by the SMEFT in this study, the linear effect of the n-th SMEFT operator on a theoretical prediction can be quantified by:

$$R_{\mathrm{SMEFT}}^{(n)} \equiv \left(\mathcal{L}_{ij}^{\mathrm{NNLO}} \otimes d\hat{\sigma}_{ij,\mathrm{SMEFT}}^{(n)} \right) / \left(\mathcal{L}_{ij}^{\mathrm{NNLO}} \otimes d\hat{\sigma}_{ij,\mathrm{SM}} \right) \,, \quad n = 1 \ldots, N \,, \tag{2.5}$$

where i, j are parton indices, $\mathcal{L}_{ij}^{\mathrm{NNLO}}$ is the NNLO partonic luminosity defined as

$$\mathcal{L}_{ij}(\tau, M_X) = \int_{\tau}^{1} \frac{dx}{x} \, f_i(x, M_X) f_j(\tau/x, M_X) \,, \quad \tau = M_X^2/s, \tag{2.6}$$

$d\hat{\sigma}_{ij,\mathrm{SM}}$ the bin-by-bin partonic SM cross section, and $d\hat{\sigma}_{ij,\mathrm{SMEFT}}^{(n)}$ the corresponding partonic cross section associated to the interference between $\mathcal{O}_n$ and the SM amplitude $\mathcal{A}_{\mathrm{SM}}$ when setting $c_n = 1$. This value of c_n is only used to initialize the potential contributions of the SMEFT operator; the effective values of the Wilson coefficient are found after the

fit is performed. Quadratic effects of the interference between the n-th and m-th SMEFT operators can be evaluated as

$$R_{\text{SMEFT}}^{(n,m)} \equiv \left(\mathcal{L}_{ij}^{\text{NNLO}} \otimes d\hat{\sigma}_{ij,\text{SMEFT}}^{(n,m)} \right) / \left(\mathcal{L}_{ij}^{\text{NNLO}} \otimes d\hat{\sigma}_{ij,\text{SM}} \right), \quad n, m = 1 \ldots, N, \qquad (2.7)$$

with the bin-by-bin partonic cross section $d\hat{\sigma}_{ij,\text{SMEFT}}^{(n,m)}$ 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 computation of the SMEFT contributions is performed numerically with the Feyn-Rules [175] model SMEFTatNLO [176], which allows one to include NLO QCD corrections to the observables. The obtained cross sections are then combined in so-called *BSM factors* by taking the ratio with the respective SM cross sections, in order to produce $R_{\text{SMEFT}}^{(n)}$ and $R_{\text{SMEFT}}^{(n,m)}$, respectively the linear and quadratic corrections. The current methodology relies on the assumption that bin-by-bin ratios between SM and SMEFT are basically unaffected by the choice of the PDF set. We have verified that they are rather stable for different available PDF sets and we ultimately computed them with the PDF set `NNPDF40_nnlo_as_01180`. If they were not stable, we would have iteratively re-computed them until reaching convergence. This is further discussed in section 6.

With these considerations, we can account for SMEFT effects in our theoretical predictions by mapping the SM prediction T^{SM} to

$$T = T^{\text{SM}} \times K(\{c_n\}), \qquad (2.8)$$

with

$$K(\{c_n\}) = 1 + \sum_{n=1}^{N} c_n R_{\text{SMEFT}}^{(n)} + \sum_{1 \leq n \leq m \leq N} c_{nm} R_{\text{SMEFT}}^{(n,m)}, \qquad (2.9)$$

with $c_{nm} = c_n c_m$. Eq. (2.8) is at the centre of the `SIMUnet` methodology, which we discuss in section 3.

3 Fitting methodology

In this work, the joint determination of the PDFs and the EFT coefficients is carried out using the `SIMUnet` methodology, first presented in [50], which is substantially extended in this work. The core idea of `SIMUnet` is to incorporate the Wilson coefficients into the optimisation problem that enters the PDF determination, by accounting explicitly for their dependence in the theoretical predictions used to fit the PDFs. Specifically, the neural network model used in the SM-PDF fits of `NNPDF4.0` is augmented with an additional layer, which encodes the dependence of the theory predictions entering the fit on the Wilson coefficients.

In this section, first we provide an overview of the `SIMUnet` methodology, highlighting the new features that have been implemented for the present study.

3.1 `SIMUnet` overview

The `SIMUnet` [50] methodology extends the `NNPDF4.0` framework [51, 52] to account for the EFT dependence (or, in principle, any parametric dependence) of the theory cross-sections entering the PDF determination. This is achieved by adding an extra layer to the

`NNPDF4.0` neural network to encapsulate the dependence of the theory predictions on the EFT coefficients, including the free parameters in the general optimisation procedure. This results in a simultaneous fit of the PDF as well as EFT coefficients to the input data. As in the NNPDF methodology, the error uncertainty estimation makes use of the Monte Carlo replica method, which yields an uncertainty estimate on both PDF and EFT parameters. We discuss the limitations of this method in appendix E.

The SM theoretical observables are encoded using interpolation grids, known as `FK`-tables [169, 177, 178], which encode the contribution of both the DGLAP evolution and the hard-scattering matrix elements and interface it with the initial-scale PDFs in a fast and efficient way.

The simultaneous fit is represented as a neural network using the `Tensorflow` [179] and `Keras` [180] libraries. The architecture is schematically represented in figure 1. Trainable weights are represented by solid arrows, and non-trainable weights by dashed arrows. Through a forward pass across the network, the inputs (x-Bjorken and its logarithm) proceed through hidden layers to output the eight fitted PDFs at the initial parametrisation scale Q_0. For each of the experimental observables entering the fit, these PDFs are then combined into a partonic luminosity $\mathcal{L}^{(0)}$ at Q_0, which is convolved with the precomputed `FK`-tables Σ to obtain the SM theoretical prediction $\mathcal{T}^{\text{SM}}$. Subsequently, the effects of the N EFT coefficients $\boldsymbol{c} = (c_1, \ldots, c_N)$, associated to the operator basis considered, are accounted for by means of an extra layer, resulting in the final prediction for the observable $\mathcal{T}$ entering the SMEFT-PDF fit. The `SIMUnet` code allows for both linear and quadratic dependence on the EFT coefficients. In linear EFT fits, the last layer consists of N trainable weights to account for each Wilson coefficient. In quadratic EFT fits, in addition to the N trainable weights, a set of $N(N + 1)/2$ non-trainable parameters, which are functions of the trainable weights, is included to account for all diagonal and non-diagonal contributions of EFT-EFT interference to the cross-sections. The results obtained with the quadratic functionality of `SIMUnet` are, however, not displayed in this work, for the reasons explained in appendix E. The PDF parameters $\boldsymbol{\theta}$ and the EFT coefficients $\boldsymbol{c}$ entering the evaluation of the SMEFT observable in figure 1 are then determined simultaneously from the minimisation of the fit figure of merit (also known as loss function).

The `SIMUnet` architecture can be minimally modified to deal with the fixed-PDF case, in which only the EFT coefficients are treated as free parameters in the optimisation process. This can be achieved by freezing the PDF-related weights in the network architecture to the values obtained in some previous fit, for example a SM-PDF determination based on `NNPDF4.0` . In this manner, `SIMUnet` can also be used to carry out traditional EFT fits where the PDF dependence of the theory predictions is neglected. Furthermore, for PDF-independent observables, computing an `FK`-table Σ is not required and the SM cross-section $\mathcal{T}^{\text{SM}}$ can be evaluated separately and stored to be used in the fit.

As illustrated in figure 1, within the `SIMUnet` framework a single neural network encapsulates both the PDF and the EFT dependence of physical observables, with the corresponding parameters being simultaneously constrained from the experimental data included
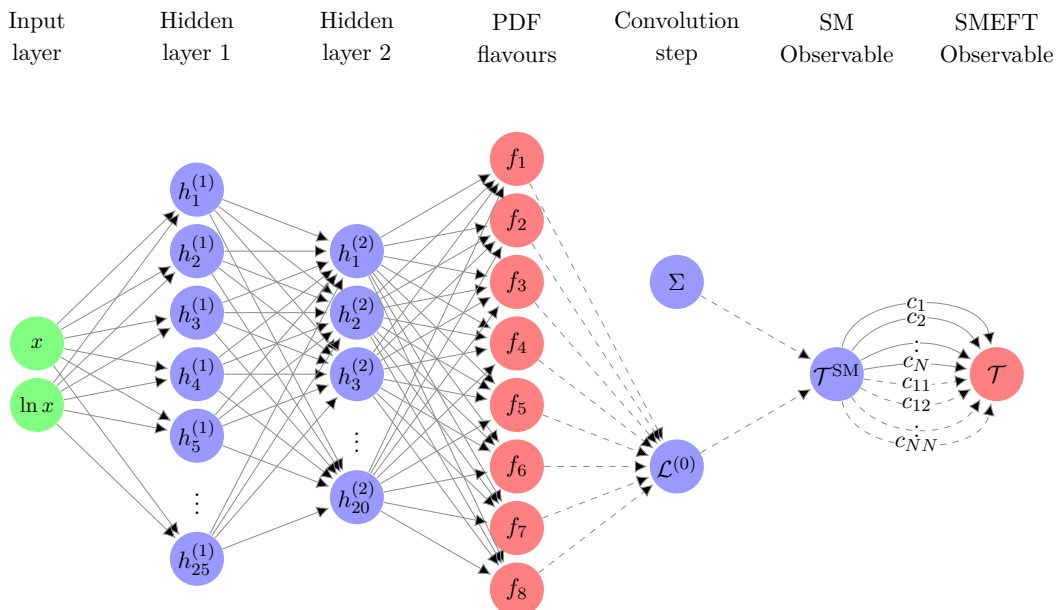
Figure 1. Schematic representation of the `SIMUnet` architecture for a general observable. Trainable weights are represented by solid arrows, and non-trainable weights by dashed arrows. Through a forward pass across the network, the inputs (x-Bjorken and its logarithm, in green) proceed through 2 hidden layers (in blue) to output the PDFs $f_1, \cdots, f_8$ (in red) at the initial parametrisation scale Q_0. For each of the experimental observables entering the fit, these PDFs are combined into a partonic luminosity $\mathcal{L}^{(0)}$ at Q_0, which is then convolved with precomputed `FK`-tables Σ to obtain the SM theoretical prediction $\mathcal{T}^{\mathrm{SM}}$. Subsequently, the effects of the EFT coefficients c_i are accounted for by means of an extra layer. In linear EFT fits this layer simplifies to just N trainable weights to account for each coefficient, and in quadratic EFT fits a set of $N(N+1)/2$ non-trainable weights has to be added to account for the EFT-EFT interference. The forward-pass of this layer results in the final prediction for the observable $\mathcal{T}$ entering the SMEFT-PDF fit. By setting the weights in the EFT layer to zero, one recovers the SM-PDF case. By freezing the PDF-related weights in the network architecture, one can carry out a fixed-PDF EFT determination or include in the joint SMEFT-PDF fit observables whose PDF dependence can be neglected.

in the fit. Specifically, we denote the prediction of the neural network as:

$$\mathcal{T} = \mathcal{T}(\hat{\boldsymbol{\theta}}) = \left(T_1(\hat{\boldsymbol{\theta}}), \ldots, T_n(\hat{\boldsymbol{\theta}}) \right) , \tag{3.1}$$

with $n = n_{\mathrm{dat}}$ and $\hat{\boldsymbol{\theta}} = (\boldsymbol{\theta}, \boldsymbol{c})$, where $\boldsymbol{\theta}$ and $\boldsymbol{c} = (c_1, \ldots, c_N)$ represent the weights associated to the PDF nodes of the network, and to the N Wilson coefficients from the operator basis, respectively. The uncertainty estimation uses the Monte Carlo replica method, where a large number N_{rep} of replicas $D^{(k)} = \left(D_1^{(k)}, \ldots, D_n^{(k)} \right)$ of the experimental measurements $D = (D_1, \ldots, D_n)$ are sampled from the distribution of experimental uncertainties with $k = 1, \ldots, N_{\mathrm{rep}}$. The optimal values for the fit parameters $\hat{\boldsymbol{\theta}}^{(k)}$ associated to each replica are obtained by means of a Stochastic Gradient Descent (SGD) algorithm that minimises

the corresponding figure of merit:

$$E_{\text{tot}}^{(k)}\left(\hat{\boldsymbol{\theta}}\right) = \frac{1}{n_{\text{dat}}} \sum_{i,j=1}^{n_{\text{dat}}} \left(D_i^{(k)} - T_i(\hat{\boldsymbol{\theta}})\right) \left(\text{cov}_{t0}^{-1}\right)_{ij} \left(D_j^{(k)} - T_j(\hat{\boldsymbol{\theta}})\right) , \qquad (3.2)$$

where the covariance matrix in eq. (3.2) is the t_0 covariance matrix, which is constructed from all sources of statistical and systematic uncertainties that are made available by the experiments with correlated multiplicative uncertainties treated via the 't0' prescription [181] in the fit to avoid fitting bias associated with multiplicative uncertainties.

Once eq. (3.2) is minimised for each replica, subject to the usual cross-validation stopping, one ends up with a sample of best-fit values for both the EFT coefficients and the PDF parameters:

$$\hat{\boldsymbol{\theta}}^{(k)} = \left(\boldsymbol{\theta}^{(k)}, \boldsymbol{c}^{(k)}\right) = \arg\min_{\hat{\boldsymbol{\theta}}} E_{\text{tot}}^{(k)}\left(\hat{\boldsymbol{\theta}}\right) , \qquad k = 1, \ldots, N_{\text{rep}} , \qquad (3.3)$$

from which one can evaluate statistical properties such as averages, variances, higher moments, or confidence level intervals. For example, the preferred value of the EFT coefficients could be evaluated over the mean over the replica sample,

$$c_\ell^* = \left\langle c_\ell^{(k)} \right\rangle_{\text{rep}} = \frac{1}{N_{\text{rep}}} \sum_{k=1}^{N_{\text{rep}}} c_\ell^{(k)} , \qquad (3.4)$$

though one could also define the preferred value as the median or mode of the distribution. Note that, in this methodology, the Monte Carlo error propagation automatically propagates the PDF uncertainty to the distribution of the best-fit values of the EFT coefficients. Hence the variance on the EFT coefficients reflects not only the experimental uncertainty of the data included in the fit, but also the functional uncertainty associated with the PDFs.

As we discuss below, the current implementation of the `SIMUnet` methodology also allows performing fixed-PDF fits, where only the Wilson coefficients are optimised. This is done by freezing the weights of the PDF part of the neural network during the minimisation of the loss function (3.2) from some other previous fit, $\boldsymbol{\theta}^{(k)} = \widetilde{\boldsymbol{\theta}}^{(k)}$, such that eq. (3.3) reduces to

$$\hat{\boldsymbol{\theta}}^{(k)} = \left(\widetilde{\boldsymbol{\theta}}^{(k)}, \boldsymbol{c}^{(k)}\right) = \arg\min_{\boldsymbol{c}} E_{\text{tot}}^{(k)}\left(\widetilde{\boldsymbol{\theta}}^{(k)}, \boldsymbol{c}\right) , \qquad k = 1, \ldots, N_{\text{rep}} . \qquad (3.5)$$

In this limit, `SIMUnet` reduces to a fixed-PDF EFT fit such as the MCfit variant of SME-FiT [182]. Likewise, by setting to zero the EFT coefficients,

$$\hat{\boldsymbol{\theta}}^{(k)} = \left(\boldsymbol{\theta}^{(k)}, \boldsymbol{c}^{(k)} = \boldsymbol{0}\right) = \arg\min_{\boldsymbol{\theta}} E_{\text{tot}}^{(k)}\left(\boldsymbol{\theta}^{(k)}, \boldsymbol{c}^{(k)} = \boldsymbol{0}\right) , \qquad k = 1, \ldots, N_{\text{rep}} , \qquad (3.6)$$

one recovers the same PDF weights $\boldsymbol{\theta}^{(k)}$ as in `NNPDF4.0`, or those of the SM-PDF fit being used as baseline in the analysis.

An important *caveat* here is that, while in the `SIMUnet` methodology the PDF uncertainty is propagated to the posterior distribution of the EFT coefficients via the Monte

Carlo replica method, in the MCfit variant of the SMEFiT methodology the fit of the EFT only considers the central PDF member (which in the `NNPDF4.0` case corresponds to the average of the PDF replicas) for all $N_{\rm rep}$ replicas, and the PDF uncertainty is propagated to the EFT coefficients by utilising an additional covariance matrix (both in the fit of the EFT coefficients and in the generation of the Monte Carlo replicas of the experimental data) that is added to t_0 covariance matrix. Namely,

$$\text{cov}_{\rm exp+th} = \text{cov}_{t^0} + \text{cov}_{\rm th}, \tag{3.7}$$

where $\text{cov}_{\rm th}$ includes the PDF contribution [33, 53], computed as

$$(\text{cov}_{\rm th})_{ij} = \langle T_i^{(k)} T_j^{(k)} \rangle_k - \langle T_i^{(k)} \rangle_k \langle T_j^{(k)} \rangle_k, \tag{3.8}$$

in which the average is taken over PDF replicas. The two ways of propagating PDF uncertainties to the distribution of the EFT coefficients are equivalent assuming that PDF uncertainties are Gaussian and uncorrelated.

`SIMUnet` adopts the same optimisation settings as those set in the `NNPDF4.0` analysis for the PDF-dependent part of the network. On the other hand it adjusts only those hyperparameters associated to the EFT-dependent layer. Within the joint SMEFT-PDF fit, several of the fit settings such as the prior ranges for the EFT parameters and the learning rates are improved in an iterative way until convergence is achieved. In doing so, we also iterate the t_0 covariance matrix and the preprocessing exponents as customary in the NNPDF procedure. In the fixed-PDF EFT fit, the user can decide both the ranges and the prior distributions to be used in the initial sampling of EFT coefficients as determined e.g. from a previous fit or from one-parameter scans.

3.2 New features

We now discuss some of the new features that have been implemented in `SIMUnet`, in comparison with [50], which are motivated by the needs of the SMEFT-PDF fits to LHC top quark data presented in this work. We consider in turn the following new features: the implementation of the quadratic contributions to the EFT cross-sections in the joint fits; fitting observables whose PDF dependence is negligible or non-existent; initialising the PDF weights of the neural network with the results of a previous fit; and finally, the improved initialisation of the EFT coefficients.

Quadratic EFT contributions. The version of `SIMUnet` used in [50] for the SMEFT-PDF fits of high-mass Drell-Yan data allowed the inclusion of quadratic contributions to the EFT cross-sections only under the approximation in which the cross-terms proportional to $c_i c_j$ with $i \neq j$ in eq. (2.9) were neglected. In the current implementation, `SIMUnet` can instead account for the full quadratic contributions to the EFT cross-sections, including the non-diagonal cross-terms. This feature can be especially important for the interpretation of top quark measurements at the LHC, given that for many observables quadratic corrections, including cross-terms relating different operators, can be sizeable specially in the high-energy region.

The implementation consists of explicitly accounting for the cross terms, as parameters which depend on the Wilson coefficients and can be differentiated as a function of them during the training procedure.

PDF-independent observables. In the original version of `SIMUnet`, only physical observables with explicit dependence on both the PDFs (via the `FK`-tables interface) and the EFT coefficients could be included in a simultaneous fit. We have now extended the `SIMUnet` framework to describe observables that are independent of the PDF parameters $\boldsymbol{\theta}$, namely the weights and thresholds of the network depicted in figure 1 that output the SM partonic luminosity $\mathcal{L}^{(0)}$. For these PDF-independent observables, the SM predictions T^{SM} are evaluated separately and stored in theory tables which can be used to evaluate the SMEFT cross-sections after applying the rescaling of eq. (2.9); hence, these observables only depend on the Wilson coefficients c_n.

In the current analysis the four-heavy cross-sections $\sigma_{\mathrm{tot}}(t\bar{t}b\bar{b})$ and $\sigma_{\mathrm{tot}}(t\bar{t}t\bar{t})$, the W-helicity measurements, and the associated top quark production cross-sections tZ, tW and $t\bar{t}\gamma$ are treated as PDF-independent observables, as for those cross-sections the PDF dependence can be neglected in comparison with other sources of theoretical and experimental uncertainty. The possibility to include PDF-independent observables makes `SIMUnet` a global SMEFT analysis tool on the same footing as SMEFiT [33, 182], FitMaker [102], HepFit [183], EFTfitter [184], and Sfitter [31] among others. This is demonstrated in appendix C, where it is shown that the results of a linear fixed-PDF SMEFT analysis performed with `SIMUnet` coincide with those obtained with SMEFiT [182] once the same experimental data and theory calculations are used. Moreover, the new feature will allow us to include in future analyses any non-hadronic observables, such as electroweak precision observables (EWPO) [185].

Fixed-PDF weight initialisation. Within the current `SIMUnet` implementation, one can also choose to initialise the PDF-dependent weights of the network in figure 1 using the results of a previous Monte Carlo fit of PDFs, for example an existing SM-PDF analysis obtained with the `NNPDF4.0` methodology. The weights of the latter are written to file and then read by `SIMUnet` for the network initialisation.

This feature has a two-fold application. First, instead of initialising at random the network weights in a simultaneous SMEFT-PDF fit, one can set them to the results of a previous SM-PDF fit, thus speeding up the convergence of the simultaneous fit, with the rationale that EFT corrections are expected to represent a perturbation of the SM predictions. Second, we can use this feature to compute EFT observables in the fixed-PDF case described above using the `FK`-table convolution with this previous PDF set as input, as opposed to having to rely on an independent calculation of the SM cross-section. Therefore, this PDF weight-initialisation feature helps realise `SIMUnet` both as a fixed-PDF EFT analysis framework, and to assess the stability of the joint SMEFT-PDF fits upon a different choice of initial state of the network in the minimisation.

Improved initialisation of the EFT coefficients. In the original implementation of `SIMUnet` it was only possible to initialise the EFT coefficients at specific values, selected

beforehand by the user. In this work, we have developed more flexible initialisation schemes for the Wilson coefficients, in the sense that they can now be sampled from a prior probability distributed defined by the user. Specifically, each Wilson coefficient c_i can be sampled from either a uniform $\mathcal{U}[a_i, b_i]$ or a normal $\mathcal{N}(\mu_i, \sigma_i)$ distribution. The ranges (a_i, b_i) of the uniform distribution $\mathcal{U}$ and the mean and standard deviation (μ_i, σ_i) of the Gaussian distribution $\mathcal{N}$ are now user-defined parameters, which can be assigned independently to each Wilson coefficients that enter the fit. This feature enhances the effectiveness and flexibility of the minimisation procedure by starting from the regions of the parameter space with the best sensitivity to the corresponding Wilson coefficient.

Another option related to the improved initialisation of EFT coefficients is the possibility to adjust the overall normalisation of each coefficient by means of a user-defined scale factor. The motivation to introduce such a coefficient-dependent scale factor is to end up with (rescaled) EFT coefficients entering the fit which all have a similar expected range of variation. This feature is advantageous, since the resulting gradients entering the SGD algorithm will all be of the same order, and hence use a unique learning rate which is appropriate for the fit at hand.

4 Impact of the top quark Run II dataset on the SM-PDFs

Here we present the results of a global SM-PDF determination which accounts for the constraints of the most extensive top quark dataset considered to date in such analyses and described in section 2. The fitting methodology adopted follows closely the settings of the `NNPDF4.0` study [51], see also [160] for a rebuttal of some critical arguments. This dataset includes not only the most up-to-date measurements of top quark pair production from Run II, but it also includes all available single top production cross-sections and distributions and for the first time new processes not considered in PDF studies before, such as the A_C asymmetry in $t\bar{t}$ production and the $t\bar{t}V$ and tV associated production (with $V = Z, W$).

We begin by summarising the methodological settings used for these fits in section 4.1. Then in section 4.2 we assess the impact of adding to a no-top baseline PDF fit various subsets of the top quark data considered in this study. In particular, we assess the impact of including updated Run II $t\bar{t}$ and single-top measurements in comparison with the subset used in the `NNPDF4.0` analysis, see the second-to-last column of tables 1–7. Furthermore, we quantify for the first time the impact of associated vector boson and single-top (tV) as well as associated vector boson and top-quark pair production ($t\bar{t}V$) data in a PDF fit. Finally in section 4.3 we combine these results and present a `NNPDF4.0` fit variant including all data described in section 2, which is compared to both the `NNPDF4.0-notop` baseline and to the original `NNPDF4.0` set.

4.1 Fit settings

An overview of the SM-PDF fits that are discussed in this section is presented in table 11. First of all, we produce a baseline fit, denoted by `NNPDF4.0-notop`, which is based on the same dataset as `NNPDF4.0` but with all top quark measurements excluded. Then we produce fit variants A to G, which quantify the impact of including in this baseline various

subsets of top data (indicated by a check mark in the table). Finally, fit variant H is the main result of this section, namely the fit in which the full set of top quark measurements described in section 2 is added to the no-top baseline.

In these fits, methodological settings such as network architecture, learning rates, and other hyperparameters are kept the same as in NNPDF4.0 , unless otherwise specified. One difference is the training fraction $f_{\rm tr}$ defining the training/validation split used for the cross-validation stopping criterion. In NNPDF4.0 , we used $f_{\rm tr} = 0.75$ for all datasets. Here instead we adopt $f_{\rm tr} = 0.75$ only for the no-top datasets and $f_{\rm tr} = 1.0$ instead for the top datasets. The rationale of this choice is to ensure that the fixed-PDF SMEFT analysis, where overfitting is not possible [53], exploits all the information contained in the top quark data considered in this study, and then for consistency to maintain the same settings in both the SM-PDF fits (this section) and in the joint SMEFT-PDF fits (to be discussed in section 6). Nevertheless, we have verified that the resulting SM-PDF fits are statistically equivalent to the fits obtained by setting the training fraction to be 0.75 for all data, including for the top quark observables.

Fits A–G in table 11 are composed by $N_{\rm rep} = 100$ Monte Carlo replicas after post-fit selection criteria, while the NNPDF4.0-notop baseline fit and fit H are instead composed by $N_{\rm rep} = 1000$ replicas. As customary, all fits presented in this section are iterated with respect to the t_0 PDF set and the pre-processing exponents.

4.2 Impact of individual top quark datasets

First we assess the impact of specific subsets of LHC top quark data when added to the NNPDF4.0-notop baseline, fits A–G in table 11. In the next section we discuss the outcome of fit H, which contains the full top quark dataset considered in this work.

Figure 2 displays the comparison between the gluon PDF at $Q = m_t = 172.5\,{\rm GeV}$ obtained in the NNPDF4.0 and NNPDF4.0-notop fits against fit D, which includes all top-quark pair (also the charge asymmetry A_C) and all single-top quark production data considered in this analysis. Results are normalised to the central value of the NNPDF4.0-notop fit, and in the right panel we show the corresponding PDF uncertainties, all normalised to the central value of the NNPDF4.0-notop baseline. From figure 2 one finds that the main impact of the additional LHC Run II $t\bar{t}$ and sinle-top data included in fit D as compared to that already present in NNPDF4.0 is a further depletion of the large-x gluon PDF as compared to the NNPDF4.0-notop baseline, together with a reduction of the PDF uncertainties in the same kinematic region. While fit D and NNPDF4.0 agree within uncertainties in the whole range of x, fit D and NNPDF4.0-notop agree only at the 2σ level in the region $x \approx [0.2, 0.4]$. These findings imply that the effect on the SM-PDFs of the new Run II top data is consistent with, and strengthens, that of the data already part of NNPDF4.0 , and suggests a possible tension between top quark data and other measurements in the global PDF sensitive to the large-x gluon, in particular inclusive jet and di-jet production. The reduction of the gluon PDF uncertainties from the new measurements can be as large as about 20% at $x \approx 0.4$. Differences are much reduced for the quark PDFs, and restricted to a 5% to 10% uncertainty reduction in the region around $x \sim 0.2$ with central values essentially unchanged.

Fit ID	Datasets included in fit					
	No-top baseline, section 2.1	Incl. $t\bar{t}$, table 1	Asymm., table 2	Assoc. $t\bar{t}$, table 4	Single-t, table 7	Assoc. single$-t$, table 8
NNPDF4.0-notop (Baseline)	✓					
A (inclusive $t\bar{t}$)	✓	✓				
B (inclusive $t\bar{t}$ and charge asymmetry)	✓	✓	✓			
C (single top)	✓				✓	
D (all $t\bar{t}$ and single top)	✓	✓	✓		✓	
E (associated $t\bar{t}$)	✓			✓		
F (associated single top)	✓					✓
G (all associated top)	✓			✓		✓
H (all top data)	✓	✓	✓	✓	✓	✓

Table 11. Overview of the SM-PDF fits discussed in this section. The baseline fit, NNPDF4.0-notop, is based on the same dataset as NNPDF4.0 with all top quark measurements excluded. The fit variants A to G consider the impact of including in this baseline various subsets of top data, while in fit H the full set of top quark measurements described in section 2 is added to the baseline.
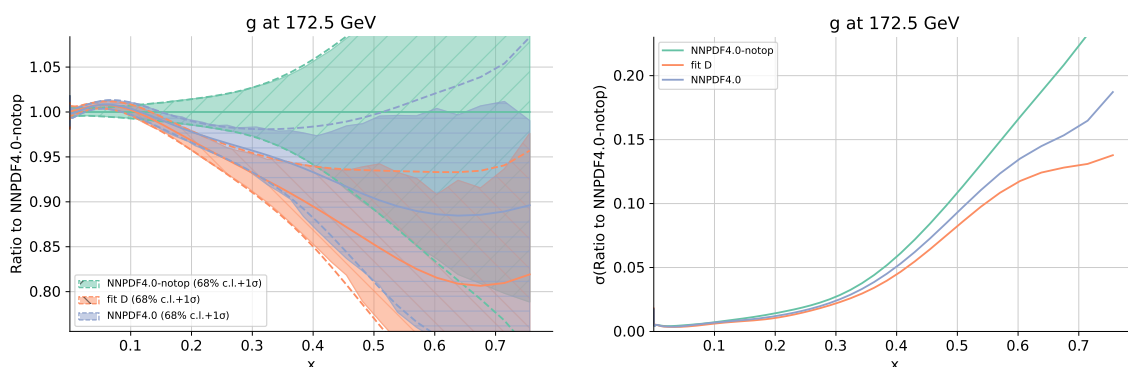

Figure 2. Left: comparison between the gluon PDF at $Q = m_t = 172.5\,\mathrm{GeV}$ obtained in the NNPDF4.0 and NNPDF4.0-notop fits against fit D in table 11, which includes all top-quark pair (also the charge asymmetry A_C) and single-top quark production data considered in this analysis. Results are normalised to the central value of the NNPDF4.0-notop set. Right: same comparison now for the PDF uncertainties (all normalised to the central value of the NNPDF4.0-notop set).
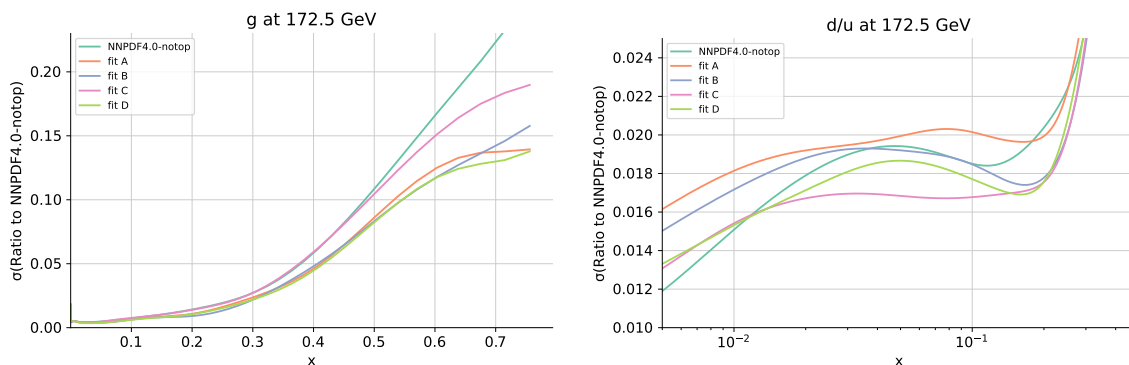
Figure 3. The ratio between the PDF 1σ uncertainty and the central value of the `NNPDF4.0-notop` baseline in the case of the gluon (left panel) and in the case of the d/u ratio (right panel). The uncertainty of the baseline fit at $Q = m_t = 172.5\,\text{GeV}$ is compared with the uncertainty associated with fits A, B, C, and D in table 11. These fit variants include the following top datasets: inclusive $t\bar{t}$ (A), inclusive $t\bar{t} + A_C$ (B), single top (C), and their sum (D).

To disentangle which of the processes considered dominates the observed effects on the gluon and the light quarks PDFs, figure 3 compares the relative PDF uncertainty on the gluon and on the d/u quark ratio in the `NNPDF4.0-notop` baseline fit at $Q = m_t = 172.5\,\text{GeV}$ with the results from fits A, B, C, and D. As indicated in table 11, these fit variants include the following top datasets: inclusive $t\bar{t}$ (A), inclusive $t\bar{t} + A_C$ (B), single top (C), and their sum (D). The inclusion of the top charge asymmetry A_C data does not have any impact on the PDFs; indeed fits A and B are statistically equivalent. This is not surprising, given that in eq. (2.1) the dependence on PDFs cancels out. Concerning the inclusion of single top data (fit C), it does not affect the gluon PDF but instead leads to a moderate reduction on the PDF uncertainties on the light quark PDFs in the intermediate-x region, $x \approx [0.01, 0.1]$, as shown in the right panel displaying the relative uncertainty reduction for the d/u ratio. This observation agrees with what was pointed out by a previous study [16], and the impact of LHC single-top measurements is more marked now as expected since the number of data points considered here is larger. We conclude that the inclusive $t\bar{t}$ measurements dominate the impact on the large-x gluon observed in figure 2, with single top data moderately helping to constrain the light quark PDFs in the intermediate-x region.

We now consider the effect of the inclusion of data that were not included before in any PDF fit, namely either $t\bar{t}$ or single-top production in association with a weak vector boson. Although current data exhibits large experimental and theoretical uncertainties, it is interesting to study whether they impact PDF fits at all, in view of their increased precision expected in future measurements; in particular, it is useful to know which parton flavours are most affected.

Figure 4 displays the same comparison as in figure 2 now for the `NNPDF4.0-notop` baseline and the variants E, F, and G from table 11, which include the $t\bar{t}V$ (E) and tV (F) data as well as their sum (G). The pull of $t\bar{t}V$ is very small, while the pull of the tV measurements is in general small, but consistent with those of the inclusive $t\bar{t}$ measurements, namely preferring a depletion of the large-x gluon. This result indicates
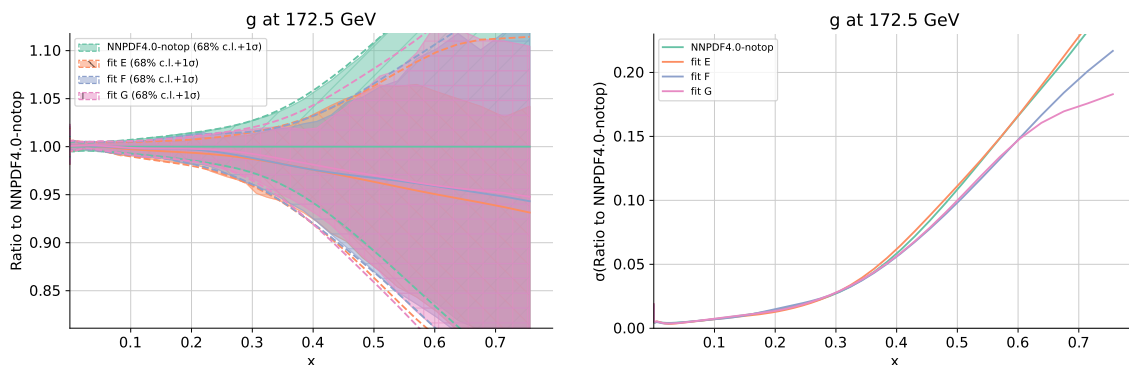
Figure 4. Same as figure 2 comparing the `NNPDF4.0-notop` baseline fit with variants E, F, and G from table 11. These variants include associated $t\bar{t}$ and vector boson production data (E), associated single top and vector boson production data (F) and their sum (G).
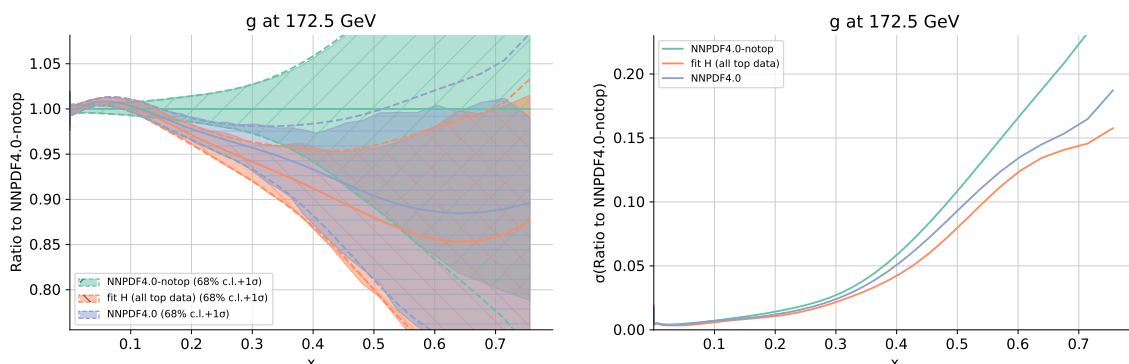

Figure 5. Same as figure 2 comparing NNPDF4.0 and NNPDF4.0 (no top) with fit variant H in table 11, which includes the full set of top quark measurements considered in this analysis.

that $t\bar{t}V$ and tV data may be helpful in constraining PDF once both future experimental data and theoretical predictions become more precise, although the corresponding inclusive measurements are still expected to provide the dominant constraints.

4.3 Combined effect of the full top quark dataset

The main result of this section is displayed in figure 5, which compares the `NNPDF4.0` and the `NNPDF4.0-notop` fits with variant H in table 11, namely with the fit where the full set of top quark measurements considered in this analysis has been added to the no-top baseline. As in the case of figure 2, we show the large-x gluon normalised to the central value of `NNPDF4.0-notop` and the associated 1σ PDF uncertainties (all normalised to the central value of the baseline). The results of fit H are similar to those of fit D, although slightly more marked. This is expected, since as shown above the associated production datasets $t\bar{t}V$ and tV carry little weight in the fit.

From figure 5 one observes how the gluon PDF of fit H deviates from the `NNPDF4.0-notop` baseline markedly in the data region $x \in [0.1, 0.5]$. The shift in the gluon PDF can be up to the 2σ level, and in particular the two PDF uncertainty bands
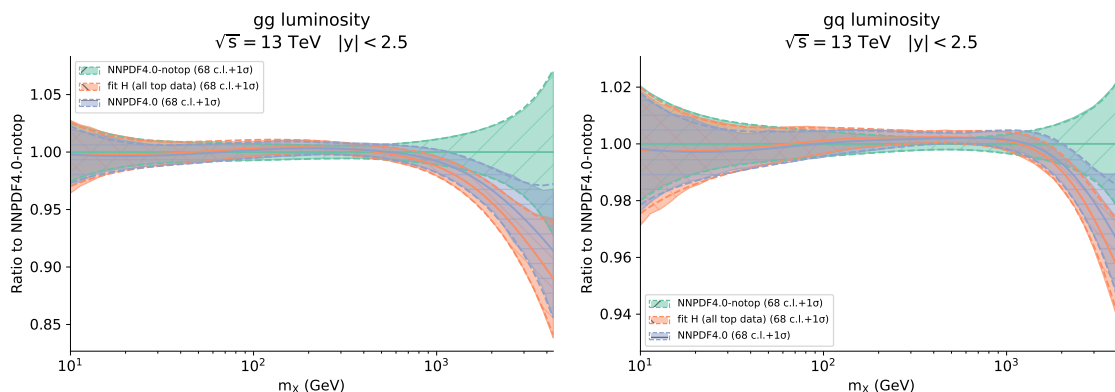
Figure 6. The gluon-gluon (left) and quark-gluon (right panel) partonic luminosities at $\sqrt{s} = 13\,\text{TeV}$ restricted to the central acceptance region with $|y| \leq 2.5$. We compare the `NNPDF4.0` and `NNPDF4.0-notop` fits with the predictions based on fit H, which includes the full top quark dataset considered here. Results are presented as the ratio to the `NNPDF4.0-notop` baseline fit.

do not overlap in the region $x \in [0.2, 0.35]$. As before, we observe that the inclusion of the latest Run II top quark measurements enhances the effect of the top data already included in `NNPDF4.0`, by further depleting the gluon in the large-x region and by reducing its uncertainty by a factor up to 25%. Hence, one finds again that the new Run II top quark production measurements lead to a strong pull on the large-x gluon, qualitatively consistent but stronger as compared with the pulls associated from the datasets already included in `NNPDF4.0`.

To assess the phenomenological impact of our analysis at the level of LHC processes, figure 6 compares the gluon-gluon and quark-gluon partonic luminosities at $\sqrt{s} = 13\,\text{TeV}$ (restricted to the central acceptance region with $|y| \leq 2.5$) between `NNPDF4.0`, `NNPDF4.0-notop`, and fit H including the full top quark dataset considered here and figure 6 compares their uncertainties. Results are presented as the ratio to the no-top baseline fit. The qq and $q\bar{q}$ luminosities of fit H are essentially identical to those of the no-top baseline, as expected given the negligible changes in the quark PDFs observed in fit H, and hence are not discussed further.

From figures 6–7 one observes that both for the quark-gluon and gluon-gluon luminosity the impact of the LHC top quark data is concentrated on the region above $m_X \gtrsim 1\,\text{TeV}$. As already reported for the case of the gluon PDF, also for the luminosities the net effect of the new LHC Run II top quark data is to further reduce the luminosities for invariant masses in the TeV range, with a qualitatively similar but stronger pull as compared to that obtained in `NNPDF4.0`. While `NNPDF4.0` and its no-top variant agree at the 1σ level in the full kinematical range considered, this is not true for fit H, whose error bands do not overlap with those of `NNPDF4.0-notop` for invariant masses m_X between 2 and 4 TeV. On the other hand, `NNPDF4.0` and fit H are fully consistent across the full m_X range, and hence we conclude that predictions for LHC observables based on `NNPDF4.0` will not be significantly affected by the inclusion of the latest LHC top quark data considered in this work.
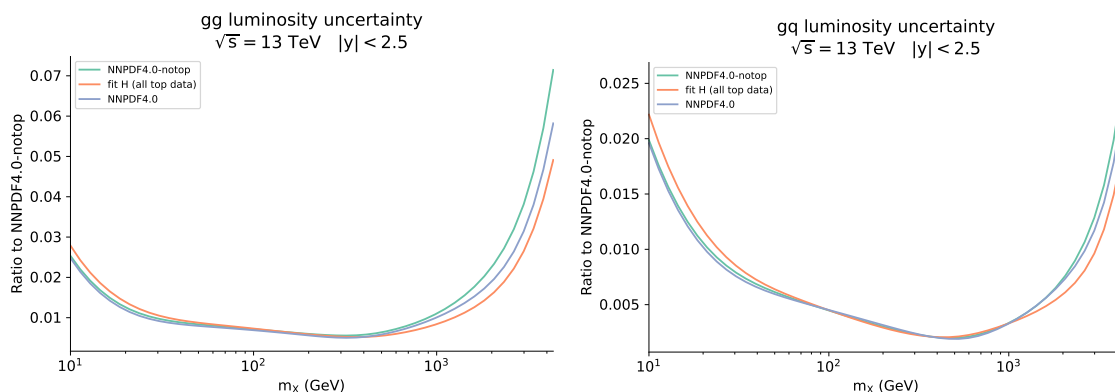
Figure 7. Same as figure 6, now for the relative luminosity uncertainties, all normalised to the `NNPDF4.0-notop` baseline fit.

Finally, concerning the fit quality of fit H is essentially stable, actually better relative to the `NNPDF4.0-notop` baseline. The experimental χ^2 per data point on their respective datasets is 1.156 for the no-top baseline, whilst for fit H is reduced to 1.144. A complete summary of the χ^2 information for all of the fits in this section is given in appendix D. It is interesting to observe that all new top data included in fit H are already described well by using the `NNPDF4.0` set, although clearly the χ^2 per data point improves (from 1.139 to 1.102) once all data are included in the fit. This confirms the overall consistency of the analysis.

5 Impact of the top quark Run II dataset on the SMEFT

We now quantify the impact of the LHC Run II legacy measurements, described in section 2, on the top quark sector of the SMEFT. As compared to previous investigations of SMEFT operators modifying top-quark interactions [29–36, 53, 102], the current analysis considers a wider dataset, in particular extended to various measurements based on the full LHC Run II luminosity. In the last column of tables 1–8 we indicated which of the datasets included here were considered for the first time within a SMEFT interpretation. Here we assess the constraints that the available LHC top quark measurements provide on the SMEFT parameter space, and in particular study the impact of the new measurements as compared to those used in [53, 102]. In this section we restrict ourselves to fixed-PDF EFT fits, where the input PDFs used in the calculation of the SM cross-sections are kept fixed. Subsequently, in section 6, we generalise to the case in which PDFs are extracted simultaneously together with the EFT coefficients.

The structure of this section is as follows. We begin in section 5.1 by describing the methodologies used to constrain the EFT parameter space both at linear and quadratic order in the EFT expansion. We also present results for the Fisher information matrix, which indicates which datasets provide constraints on which operators. In section 5.2, we proceed to give the results of the fixed-PDF EFT fits at both linear and quadratic order, highlighting the impact of the new Run II top quark data by comparison with

previous global SMEFT analyses. In section 5.3, we assess the impact of replacing the CMS 13 TeV differential measurement of $t\bar{t}$ in the ℓ+jets channel, binned with respect to invariant top quark pair mass, by the corresponding double-differential measurement binned with respect to both invariant top quark pair mass and top quark pair rapidity. In the dataset selection performed in Sect 2.2 we rejected the double-differential distribution due to its poor χ^2-statistic in the SM, which could not be improved by a weighted fit of PDFs; in the present section, it is interesting to see whether the SMEFT can help account for the poor fit of this dataset. Finally, in section 5.4 we evaluate the correlation between PDFs and EFT coefficients to identify the kinematic region and EFT operators which are potentially sensitive to the SMEFT-PDF interplay to be studied in section 6.

5.1 Fit settings

Throughout this section, we will allow only the SMEFT coefficients to vary in the fit, keeping the PDFs fixed to the SM-PDFs baseline obtained in the `NNPDF4.0-notop` fit discussed in section 4; with this choice, one removes the overlap between the datasets entering the PDF fit and the EFT coefficients determination. Our analysis is sensitive to the $N = 25$ Wilson coefficients defined in table 14, except at the linear level where the four-heavy coefficients c_{Qt}^8, c_{QQ}^1, c_{QQ}^8, c_{Qt}^1 and c_{tt}^1 (which are constrained only by $t\bar{t}t\bar{t}$ and $t\bar{t}b\bar{b}$ data) exhibit three flat directions [53]. In order to tackle this, we remove the five four-heavy coefficients from the linear fit[1] Hence, in our linear fit we have $N = 20$ independent coefficients constrained from the data, whereas in the quadratic fit, we fit all $N = 25$ independent coefficients.

The linear EFT fits presented in this section are performed with the `SIMUnet` methodology in the fixed-PDF option, as described in section 3; we explicitly verified that the `SIMUnet` methodology reproduces the posterior distributions provided by SMEFiT (using either the NS (Nested Sampling) or MCfit options) for a common choice of inputs, as explicitly demonstrated in appendix C. However, in the case of the quadratic EFT fits we are unable to use the `SIMUnet` methodology due to a failure of the Monte-Carlo sampling method utilised in the `SIMUnet` and SMEFiT codes; this is discussed in appendix E and a dedicated study of the problem will be the subject of future work. For this reason, quadratic EFT fits in this section are carried out with the public SMEFiT code using the NS mode [182]. To carry out these fits, the full dataset listed in tables 1–8, together with the corresponding SM and EFT theory calculations described in section 2.3, have been converted to the SMEFiT data format (this conversion was also already used for the benchmarking of appendix C).

Fisher information. The sensitivity to the EFT operators of the various processes entering the fit can be evaluated by means of the Fisher information, F_{ij}, which quantifies

---

[1]In principle one could instead rotate to the principal component analysis (PCA) basis and constrain the two non-flat directions in the four-heavy subspace, but even so, the obtained constraints remain much looser in comparison with those obtained in the quadratic EFT fit [53].

In our fits, we also keep the $t\bar{t}t\bar{t}$ and $t\bar{t}b\bar{b}$ datasets after removing the five four-heavy coefficients. We have verified that including or excluding theses sets has no significant impact whatsoever on the remaining coefficients.

the information carried by a given dataset on the EFT coefficients c_i [186]. In a linear EFT setting, the Fisher information is given by:

$$F_{ij} = L^{(i)T}(\text{cov}_{\text{exp}})^{-1}L^{(j)} \qquad (5.1)$$

where the k-th entry, $L_k^{(i)}$, of the vector $L^{(i)}$ is the linear contribution multiplying c_i in the SMEFT theory prediction for the k-th data point, and cov_{exp} is the experimental covariance matrix. In particular, the Fisher information is an $N \times N$ matrix, where N is the number of EFT coefficients, and it depends on the dataset. An important property of the Fisher information is that it is related to the covariance matrix C_{ij} of the maximum likelihood estimators by the *Cramer-Rao bound*:

$$C \geq (F^{-1}), \qquad (5.2)$$

where the notation used here implies that the matrix $C - F^{-1}$ is positive semi-definite, indicating that larger values of F_{ij} will translate to tighter bounds on the EFT coefficients.

Before displaying the results of the fixed-PDF SMEFT analysis in section 5.2, we use the Fisher information to assess the relative impact of each sector of top quark data on the EFT parameter space; this is done in the linear analysis, including $\mathcal{O}(1/\Lambda^2)$ SMEFT corrections. In the quadratic case, once $\mathcal{O}(1/\Lambda^4)$ SMEFT corrections are included, the dependence of F_{ij} on the Wilson coefficients makes interpretation more difficult. Writing $F_{ij}(D)$ for the Fisher information matrix evaluated on the dataset D, we define the relative constraining power of the dataset D via:

$$\text{relative constraining power of } D \text{ on operator } c_i = F_{ii}(D) \bigg/ \sum_{\text{sectors } D'} F_{ii}(D'). \qquad (5.3)$$

Since $F_{ii}(D)$ corresponds to the constraining power of the dataset D in a one-parameter fit of the Wilson coefficient c_i, this definition only quantifies how much a dataset impacts one-parameter fits of single Wilson coefficients in turn; however, this will give a general qualitative picture of some of the expected behaviour in the global fit too. We display the results of evaluating the relative constraining power of each top quark data sector on each of the parameters in figure 8, quoting the results in percent (%).

As expected, $t\bar{t}$ total cross sections constitute the dominant source of constraints on the coefficient c_{tG}. Each of the four-fermion operators receive important constraints from differential $t\bar{t}$ distributions and charge asymmetry measurements. Note that this impact is magnified when we go beyond individual fits, in which case measurements of charge asymmetries are helpful in breaking flat directions amongst the four-fermion operators [31, 102]. The coefficient $c_{Qq}^{1,3}$ is the exception as it is instead expected to be well-constrained by single top production. We note that the measurements of W helicities are helpful in constraining the coefficient c_{tW}, while $t\bar{t}Z$ measurements provide the dominant source of constraints on $c_{\varphi Q}^{(-)}$. We observe that the neutral top coupling c_{tZ} is entirely constrained by $t\bar{t}\gamma$, and that the effects of $t\bar{t}t\bar{t}$ and $t\bar{t}b\bar{b}$ are mostly restricted to the 4-heavy operators c_{Qt}^8, c_{QQ}^1, c_{QQ}^8, c_{Qt}^1 and c_{tt}^1.

5.2 Fixed-PDF EFT fit results

In this section, we present the results of the linear and quadratic fixed-PDF fits with the settings described in Sect 5.1.
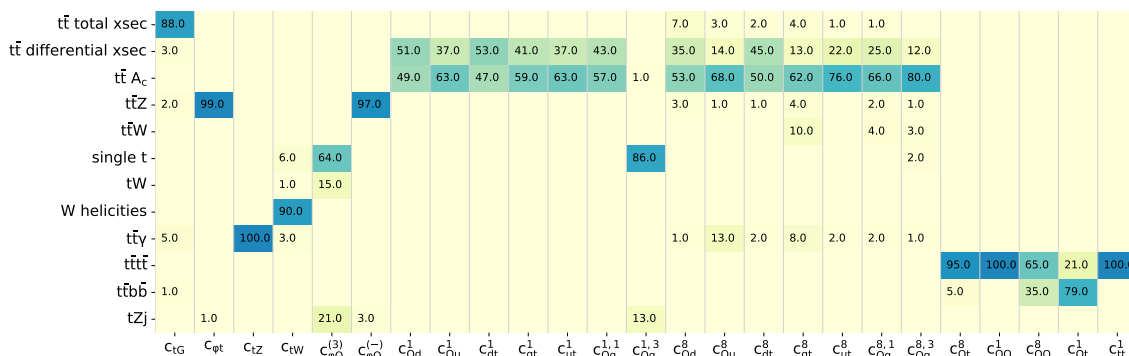
	c_{tG}	$c_{\varphi t}$	c_{tZ}	c_{tW}	$c_{\varphi Q}^{(3)}$	$c_{\varphi Q}^{(-)}$	c_{Qd}^1	c_{Qu}^1	c_{dt}^1	c_{qt}^1	c_{ut}^1	$c_{Qq}^{1,1}$	$c_{Qq}^{1,3}$	c_{Qd}^8	c_{Qu}^8	c_{dt}^8	c_{qt}^8	c_{ut}^8	$c_{Qq}^{8,1}$	$c_{Qq}^{8,3}$	c_{Qt}^8	c_{QQ}^1	c_{QQ}^8	c_{Qt}^1	c_{tt}^1
$t\bar t$ total xsec	88.0													7.0	3.0	2.0	4.0	1.0	1.0						
$t\bar t$ differential xsec	3.0						51.0	37.0	53.0	41.0	37.0	43.0		35.0	14.0	45.0	13.0	22.0	25.0	12.0					
$t\bar t$ A_c							49.0	63.0	47.0	59.0	63.0	57.0	1.0	53.0	68.0	50.0	62.0	76.0	66.0	80.0					
$t\bar t Z$	2.0	99.0				97.0								3.0	1.0	1.0	4.0		2.0	1.0					
$t\bar t W$																	10.0		4.0	3.0					
single t				6.0	64.0								86.0							2.0					
tW				1.0	15.0																				
W helicities				90.0																					
$t\bar t \gamma$	5.0		100.0	3.0										1.0	13.0	2.0	8.0	2.0	2.0	1.0					
$t\bar t t\bar t$																					95.0	100.0	65.0	21.0	100.0
$t\bar t b\bar b$	1.0																				5.0		35.0	79.0	
tZj		1.0			21.0	3.0							13.0												

Figure 8. Relative constraining power (in %) on each of the operators for each of the processes entering the fit, as defined in eq. (5.3).

Fit quality. We begin by discussing the fit quality of the global SMEFT determination, quantifying the change in the data-theory agreement relative to the SM in both the linear and quadratic SMEFT scenarios. Table 12 provides the values of the χ^2 per data point in the SM and in the case of the SMEFT at both linear and quadratic order in the EFT expansion for each of the processes entering the fit. Here, in order to ease the comparison of our results to those of SMEFiT and FitMaker, we quote the χ^2 per data point computed by using the covariance matrix defined in eq. (3.7), which includes both the experimental uncertainty and the PDF uncertainty. The corresponding values obtained by using the experimental χ^2 definition of eq. (2.3), along with a fine-grained fit quality description are given in appendix D.

We observe that in many sectors, the linear EFT fit improves the fit quality compared to the SM; notably, the $\chi^2_{\text{exp+th}}$ per data point for inclusive $t\bar t$ is vastly improved from 1.71 to 1.11. When quadratic corrections are also considered, the fit quality is usually poorer compared to the linear fit. For example, in inclusive $t\bar t$ the $\chi^2_{\text{exp+th}}$ per data point deteriorates from 1.11 to 1.69. This is not unexpected, however, since the flexibility of the quadratic fit is limited by the fact that for sufficiently large values of Wilson coefficients the EFT can only make positive corrections.[2]

It is also useful to calculate the goodness of fit, quantified by the χ^2 per degree of freedom, $\chi^2/n_{\text{dof}} = \chi^2/(n_{\text{dat}} - n_{\text{param}})$, which additionally accounts for the complexity of the models we are using in each fit. In our case, we find $\chi^2_{\text{exp+th}}/n_{\text{dof}} = 1.25$ in the SM and $\chi^2_{\text{exp+th}}/n_{\text{dof}} = 0.95$ and 1.33 in the linear and quadratic EFT scenarios respectively. We see that while the EFT at quadratic order does not provide a better fit than the SM, neglecting quadratic EFT corrections leads to a significant improvement in the overall fit quality.

Constraints on the EFT parameter space. Next, we present the constraints on the EFT parameter space. In figure 9, we display the 95% CL constraints on the 20 Wilson coefficients entering the linear fit. Two sets of constraints are shown; in green, we give the intervals obtained from a fit to the 175 data points introduced in section 2, whilst in

---

[2]This also has methodological implications. Large quadratic corrections can negatively impact the Monte-Carlo sampling method used by SIMUnet, as discussed in appendix E.

Process	n_{dat}	$\chi^2_{\mathrm{exp+th}}$ [SM]	$\chi^2_{\mathrm{exp+th}}$ [SMEFT $\mathcal{O}(\Lambda^{-2})$]	$\chi^2_{\mathrm{exp+th}}$ [SMEFT $\mathcal{O}(\Lambda^{-4})$]
$t\bar{t}$	86	1.71	1.11	1.69
$t\bar{t}$ AC	18	0.58	0.50	0.60
W helicities	4	0.71	0.45	0.47
$t\bar{t}Z$	12	1.19	1.17	0.94
$t\bar{t}W$	4	1.71	0.46	1.66
$t\bar{t}\gamma$	2	0.47	0.03	0.59
$t\bar{t}t\bar{t}$ & $t\bar{t}b\bar{b}$	8	1.32	1.06	0.49
single top	30	0.504	0.33	0.37
tW	6	1.00	0.82	0.82
tZ	5	0.45	0.30	0.31
Total	**175**	**1.24**	**0.84**	**1.14**

Table 12. The values of the χ^2 per data point for the fixed-PDF EFT fits presented in this section, both for individual groups of processes, and for the total dataset. Here the χ^2 is actually the $\chi^2_{\mathrm{exp+th}}$ defined by using the theory covariance matrix defined in eq. (3.7). In each case we indicate the number of data points, the $\chi^2_{\mathrm{exp+th}}$ obtained using the baseline SM calculations, and the results of both the linear and quadratic EFT fits.

orange, we give in the intervals obtained from a fit to the older top quark dataset used in the global analysis of ref. [53], obtained from a fit of 150 data points. This comparison allows us to quantify the information gained from the latest Run II datasets, relative to those available to previous analyses. The same comparison, this time at quadratic order in the EFT expansion, is shown in figure 10 (note that in this plot we display constraints on all 25 coefficients, including the 4-heavy coefficients c_{Qt}^8, c_{QQ}^1, c_{QQ}^8, c_{Qt}^1 and c_{tt}^1).

We first note that figures 9 and 10 both demonstrate good agreement between the fits using old and new datasets, and consistency between the new fit SMEFT bounds and the SM. At the linear level, the most noticeable improvement concerns c_{tG}; its 95% C.L. bounds decrease from $[-0.13, 0.41]$ to $[-0.18, 0.17]$, thanks to the increased amount of information in the input dataset, coming in particular from $t\bar{t}$ data. This results in both a tightening of the constraints by about 35% and a shift in the best-fit point. For many of the other coefficients, the bounds are either stable (e.g. c_{tZ}), or exhibit a shift in central value but no significant tightening (e.g. $c_{\varphi t}$, undergoes a shift of -14.3, but a decrease in the size of the constraints 95% C.L. by only 1%, and $c_{\varphi Q}^{(-)}$ undergoes a shift of -7.33, but its bounds only tighten by 2%). Finally, we note that some coefficients instead exhibit a broadening of constraints with the new dataset relative to the old dataset (for example, some of the four-fermion operators). The increase in the size of the constraints could point to some inconsistency within the new inclusive $t\bar{t}$ dataset; however, given that the bounds are very
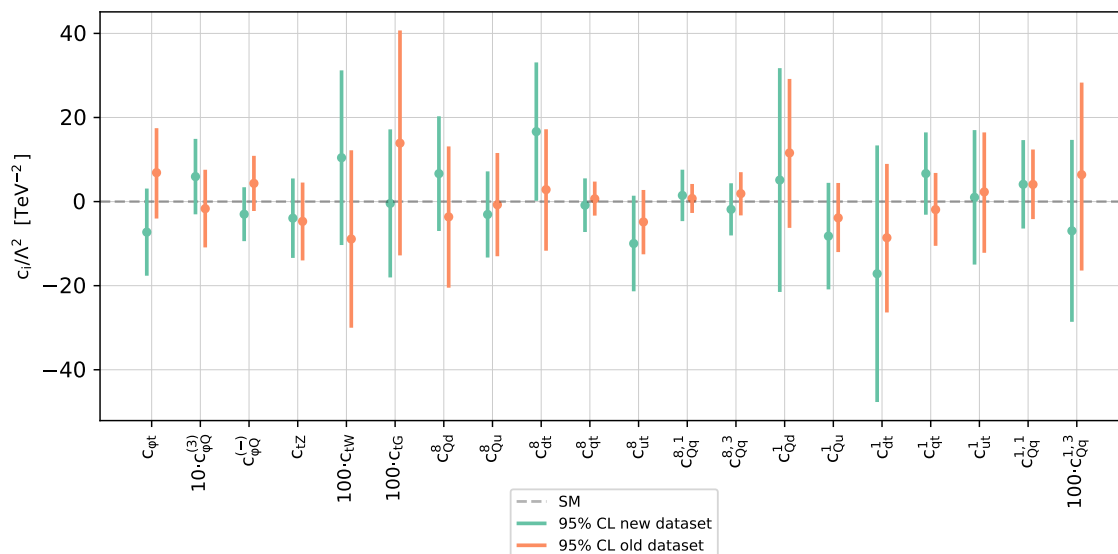
Figure 9. The 95% CL intervals on the EFT coefficients entering the linear fit, evaluated with `SIMUnet` on the dataset considered in this work, and evaluated with SMEFiT on the top quark dataset entering the analysis of [53] (note that at the linear EFT level, the results obtained with `SIMUnet` coincide with those provided by SMEFiT for the same dataset, as demonstrated in appendix C). Note also that the constraints on selected coefficients are rescaled by the factors shown, for display purposes.

large anyway, at the edges of the intervals we are likely to approach a region where the EFT is no longer valid in both cases, hence no definite conclusions may be drawn.

At the quadratic order in the EFT expansion, however, the impact of the latest Run II dataset is clear; we see a marked improvement in many of the SMEFT constraints. As shown in figure 10, the bounds on all 14 of the four-fermion operators become noticeably smaller as a result of the increase in precision in the $t\bar{t}$ sector. The constraint on c_{tZ} is improved by the inclusion of measurements of the $t\bar{t}\gamma$ total cross sections, resulting in a tightening of 24%. The addition of the p_T^γ spectrum [140] would yield an even stronger constraint, as seen in [102]. We will make use of this observable in future work when unfolded measurements are made available. Contrary to the linear fit, where we singled out $c_{\varphi t}$ and $c_{\varphi Q}^{(-)}$ as examples of coefficients which shift, but whose bounds are not improved, the constraints on $c_{\varphi t}$, $c_{\varphi Q}^{(-)}$ markedly tighten in the quadratic fit in the presence of new data; in particular, the size of the bounds on $c_{\varphi t}$, $c_{\varphi Q}^{(-)}$ decrease by 35% and 28%, respectively. On the other hand, despite the addition of new $t\bar{t}t\bar{t}$, $t\bar{t}b\bar{b}$ datasets, we find limited sensitivity to the five four-heavy coefficients c_{Qt}^8, c_{QQ}^1, c_{QQ}^8, c_{Qt}^1 and c_{tt}^1. In fact, with the new data we see a broadening of the bounds. As with the linear fit, this could point to either an inconsistency in the $t\bar{t}t\bar{t}$, $t\bar{t}b\bar{b}$ data, or simply to the ambiguity associated to the EFT validity in that particular region of the parameter space.

Correlations. Figure 11 shows the correlations the between Wilson coefficients evaluated in this analysis both at the linear and the quadratic order in the EFT expansion, shown on

Figure 10. The 95% CL intervals on the EFT coefficients entering the quadratic fit, evaluated with SMEFiT. We compare the results based on the full top quark dataset with the corresponding results obtained from the subset of top quark measurements entering the analysis of [53]. As in figure 9, the constraints on selected coefficients are rescaled by the factors shown, for display purposes.

the left and right panels respectively. In the linear fit, we first note a number of large correlations amongst the octet four-fermion operators which enter the $t\bar{t}$ production together. The singlet four-fermion operators are similarly correlated among themselves, although their correlations are comparatively suppressed. The coefficients $c_{\varphi t}$ and $c_{\varphi Q}^{(-)}$ exhibit a large positive correlation due to their entering into $t\bar{t}Z$ production together, while $c_{Qq}^{3,1}$ and $c_{\varphi Q}^{(3)}$ have positive correlations through their contribution to single top production. Further non-zero correlations are found, for example amongst the pairs $c_{Qq}^{8,3}$ & $c_{\varphi Q}^{(3)}$, c_{Qd}^{8} & c_{tZ} and c_{Qu}^{8} & c_{tZ}.

At quadratic order, however, we observe that many of these correlations are suppressed, as a result of the fact that the inclusion of quadratic corrections lifts many of the degeneracies in the fit. We observe that the pairs $c_{\varphi t}$, $c_{\varphi Q}^{(-)}$ and $c_{Qq}^{1,3}$, $c_{\varphi Q}^{(3)}$ remain correlated though. The 4-heavy operators are also included in this quadratic fit, and we find large anti-correlations between c_{QQ}^{1} and c_{QQ}^{8}, indicating that they are poorly distinguished in the $t\bar{t}t\bar{t}$ and $t\bar{t}b\bar{b}$ processes. Finally, note that we obtain subtle non-zero correlations between the octet and singlet four-fermion operators constructed from the same fields, for example between c_{Qd}^{8} and c_{Qd}^{1}. This is a result of the fact that $t\bar{t}$ measurements provide the dominant source of constraints on these coefficients and are very sensitive to quadratic corrections, and at this order the contribution from these operators differs only by a numerical factor.

5.3 Study of the CMS 1D vs 2D distribution

In the dataset selection discussed in section 2.2, the double-differential distribution measurements performed by CMS at $\sqrt{s} = 13\,\text{TeV}$ in the ℓ+jets channel [117] — binned with respect to the top quark pair invariant mass $m_{t\bar{t}}$ and the top quark pair rapidity $y_{t\bar{t}}$ — is found to be poorly described by the SM theory, with an experimental χ^2 that is 22σ away
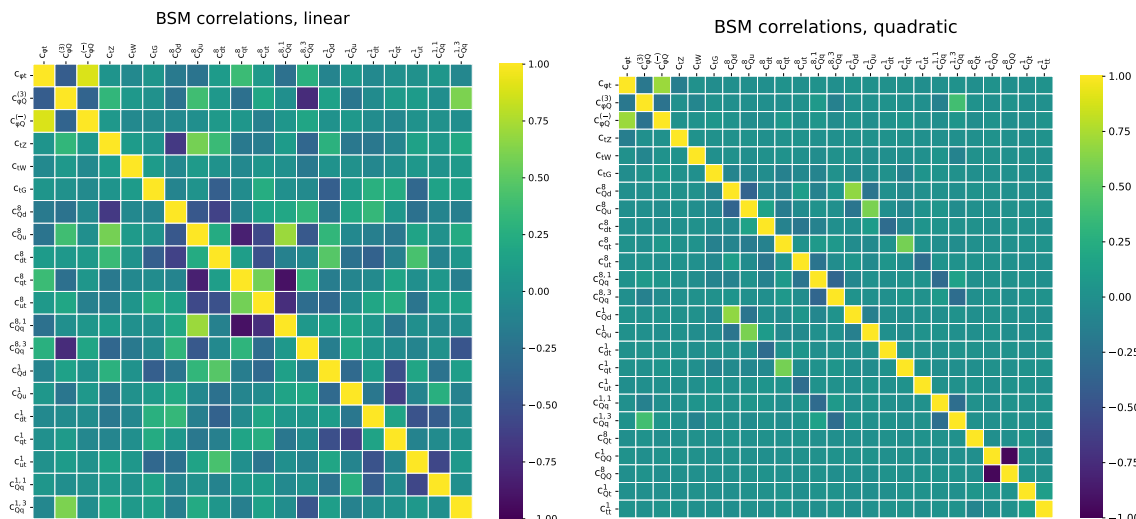
Figure 11. The values of the correlation coefficients evaluated between all pairs of Wilson coefficients entering the EFT fit at the linear (left) and quadratic (right) order. As explained in the next, the number of fitted DoFs is different in each case.

from the median of the χ^2 distribution of a perfectly consistent dataset made of $n_{\text{dat}} = 34$. Even by increasing the weight of this dataset in a weighted fit (see section 2.2 for a more detailed explanation), the χ^2_{exp} per data point improves only moderately to 4.56 and the χ^2-statistic of the other datasets deteriorates significantly, hinting to both an internal incompatibility of the CMS 2D distribution and to an incompatibility with the rest of the data. For this reason, the dataset was excluded from our analysis, and replaced by the single-differential distribution in $m_{t\bar{t}}$ — presented in the same publication [117]. The latter is well described by the SM theoretical predictions.

In this section we present a dedicated analysis to assess whether the SMEFT corrections can improve the theoretical description of this dataset. In particular, we compare a fixed-PDF fit including the 13 TeV CMS double-differential $(m_{t\bar{t}}, y_{t\bar{t}})$ distribution (CMS 2D) to the default one including the 13 TeV single-differential $m_{t\bar{t}}$ distribution (CMS 1D).

First of all, it is interesting to notice that the inclusion of quadratic SMEFT corrections in the fit does not improve much the quality of the fit of the CMS 2D distribution, with $\chi^2_{\text{exp+th}}/n_{\text{dat}}$ decreasing from 2.80 (in the SM) to $\chi^2_{\text{exp+th}}/n_{\text{dat}} = 2.57$ including SMEFT $\mathcal{O}(\Lambda^{-4})$ corrections. On the other hand, if SMEFT linear $\mathcal{O}(\Lambda^{-2})$ corrections are included, the fit quality of the CMS 2D distribution improves substantially with $\chi^2_{\text{exp+th}}/n_{\text{dat}} = 1.22$.

In order to assess what is the effect of the inclusion of the CMS 2D distribution on the other top sector data, in table 13 we compare the fit quality of the two (fixed-PDF) SMEFT fits, CMS 1D and CMS 2D, both for individual groups of processes, and for the total dataset. We observe that the fit quality deteriorates both in the SM and in the quadratic SMEFT fit once the CMS 2D distribution is fitted. The deterioration of the fit quality is mostly driven by a deterioration of the fit quality of the $t\bar{t}$ and $t\bar{t}W$ sectors. However at the level of linear SMEFT fit, the quality of the fit deteriorates only moderately and it is mostly driven by the slight deterioration in the inclusive $t\bar{t}$ sector and in the $t\bar{t}t\bar{t} \, \& \, t\bar{t}b\bar{b}$ one.

Process	n_{dat}	$\chi^2_{\text{exp+th}}$ [SM]		$\chi^2_{\text{exp+th}}$ [SMEFT $\mathcal{O}(\Lambda^{-2})$]		$\chi^2_{\text{exp+th}}$ [SMEFT $\mathcal{O}(\Lambda^{-4})$]	
		CMS 1D	CMS 2D	CMS 1D	CMS 2D	CMS 1D	CMS 2D
$t\bar{t}$	86	1.71	2.07	1.11	1.18	1.69	1.87
$t\bar{t}$ AC	18	0.58	0.58	0.50	0.47	0.60	0.60
W helicities	4	0.71	0.71	0.45	0.45	0.48	0.46
$t\bar{t}Z$	12	1.19	1.19	1.17	1.07	0.94	0.95
$t\bar{t}W$	4	1.71	1.71	0.46	0.46	1.66	1.82
$t\bar{t}\gamma$	2	0.47	0.47	0.03	0.03	0.58	0.18
$t\bar{t}t\bar{t}$ & $t\bar{t}b\bar{b}$	8	1.32	1.32	1.06	1.28	0.49	0.49
single top	30	0.50	0.50	0.33	0.34	0.37	0.35
tW	6	1.00	01.00	0.82	0.86	0.82	0.84
tZ	5	0.45	0.45	0.30	0.30	0.31	0.30
Total	**175**	**1.24**	**1.48**	**0.84**	**0.91**	**1.14**	**1.29**

Table 13. Same as table 12, now comparing the values of the $\chi^2_{\text{exp+th}}$ per data point for the fixed-PDF EFT fits presented in this section, the default one including the 13 TeV single-differential $m_{t\bar{t}}$ distribution (CMS 1D) and the one including the 13 TeV CMS double-differential $(m_{t\bar{t}}, y_{t\bar{t}})$ distribution (CMS 2D), both for individual groups of processes, and for the total dataset.

At the level of fit of the Wilson coefficients, the quadratic SMEFT fits yields similar 95% C.L. bounds on the EFT coefficients. This is somewhat expected, given that the fit quality of the CMS 2D data does not improve once quadratic SMEFT corrections are included, and the fit quality of the other datasets remains pretty much the same. On the other hand the bounds obtained from a SMEFT linear fit change more significantly depending on whether the CMS 1D or the CMS 2D distribution is used in the fit. In particular the central value of the c_{tG} 95% C.L. bounds is shifted upwards by 1σ, as well as the bounds on c_{Qd}^1 and c_{ut}^1, while the bounds on c_{Qu}^1 and c_{qt}^1 shift downwards, as expected from the EFT coefficient correlations displayed in figure 11.

As a result of our analysis, we observe that the bounds on the EFT in the linear case do depend on the input dataset quite significantly and careful consideration has to be made to the overall dataset compatibility and to the outcome of the fit once quadratic corrections are included. The current analysis in particular shows that the incompatibility of the 13 TeV CMS double differential distribution cannot be entirely attributed to new physics effects parametrised by the SMEFT expansion, rather there are internal experimental or theoretical incompatibilities that affect the results.

To conclude, it is interesting to observe that, if one tries to include the CMS double differential distribution dataset instead of the single differential dataset in a simultaneous fit of PDFs and SMEFT coefficients at the linear level, the improvement in the fit quality
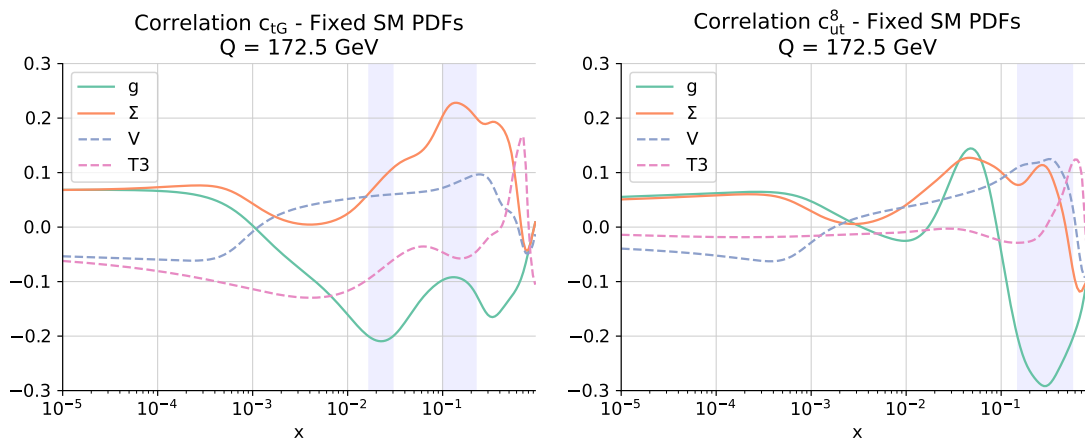
Figure 12. The 95% CL intervals on the EFT coefficients entering the liner fits, evaluated with `SIMUnet`. We compare the results based on the default full top quark dataset (CMS 1D) with the corresponding results obtained from the same set of top quark measurements in which the single-differential $m_{t\bar{t}}$ 13 TeV CMS distribution is substituted by the double differential $(m_{t\bar{t}}, y_{t\bar{t}})$ 13 TeV CMS distribution (CMS 2D).

that is obtained in a SMEFT linear fit is reverted. Indeed, the SMEFT PDF resulting from the simultaneous fit including the CMS double differential distribution (instead of the CMS single differential one) are basically unchanged with respect to the ones that will be presented in section 6, and the bounds on the SMEFT coefficients are quite similar. However, the fit quality of the CMS double differential distribution is rather poor, with $\chi^2_{\rm exp}/n_{\rm dat} = 5.69$ (only marginally decreasing from 6.43 in the PDF-only fit). This means that, in a simultaneous fit, the degrees of freedom of the PDFs tend to contrast the direction that the SMEFT coefficients would take in the absence of PDFs because the latter would require a change in the PDFs that is disfavoured by the other datasets in the global analysis. All results can be found on the public web page that includes the additional material that we do not show in the paper.[3]

5.4 Correlations between PDFs and EFT coefficients

We conclude this section by discussing the correlations observed between the PDFs and Wilson coefficients. The PDF-EFT correlation coefficient for a Wilson coefficient c and a PDF $f(x, Q)$ at a given x and Q^2 is defined as

$$\rho\left(c, f(x, Q^2)\right) = \frac{\left\langle c^{(k)} f^{(k)}(x, Q^2)\right\rangle_k - \left\langle c^{(k)}\right\rangle_k \left\langle f^{(k)}(x, Q^2)\right\rangle_k}{\sqrt{\left\langle (c^{(k)})^2\right\rangle_k - \left\langle c^{(k)}\right\rangle_k^2}\sqrt{\left\langle (f^{(k)}(x, Q^2))^2\right\rangle_k - \left\langle f^{(k)}(x, Q^2)\right\rangle_k^2}}, \qquad (5.4)$$

---

[3]https://www.pbsp.org.uk/topproject/.

Figure 13. The value of the correlation coefficient ρ between the PDFs and selected EFT coefficients as a function of x and $Q = 172.5\,\text{GeV}$. We show the results for the gluon, the total singlet Σ, total valence V, and non-singlet triplet T_3 PDFs. We provide results for representative EFT coefficients, namely c_{tG} and $c_{ut}^{(8)}$.

where $c^{(k))}$ is the best-fit value of the Wilson coefficient for the k-th replica and $f^{(k)}$ is the k-th PDF replica computed at a given x and Q, and $\langle \cdot \rangle_k$ represents the average over all replicas. We will compute the correlation between a SM PDF and the Wilson coefficients, both of which have been separately determined from the total dataset including all new top quark data. By doing so we hope to shed light on which Wilson coefficients, and which PDF flavours and kinematical regions, are strongly impacted by the top quark data and therefore exhibit a potential for interplay in a simultaneous EFT-PDF determination. The EFT corrections will be restricted to linear order in the EFT expansion.

Figure 13 displays a selection of the largest correlations. We observe that the gluon PDF in the large-x region is significantly correlated with the Wilson coefficients c_{tG}, $c_{ut}^{(8)}$. On the other hand, relatively large correlations are observed between c_{tG} and the total singlet Σ, while the total valence V, and non-singlet triplet T_3 PDFs show no relevant correlations with the selected coefficients. This is not surprising, given the impact of top quark pair production total cross sections and differential distributions in constraining these PDFs and Wilson coefficients. Whilst these correlations are computed from a determination of the SMEFT in which the PDFs are fixed to SM PDFs, the emergence of non-zero correlations provides an indication of the potential for interplay between the PDFs and the SMEFT coefficients; this interplay will be investigated in a simultaneous determination in the following section.

6 SMEFT-PDFs from top quark data

In this section we present the main results of this work, namely the simultaneous determination of the proton PDFs and the SMEFT Wilson coefficients from the LHC Run II top quark data described in section 2, following the `SIMUnet` methodology summarised in section 3. This determination of the SMEFT-PDFs from top quark data is carried out at the
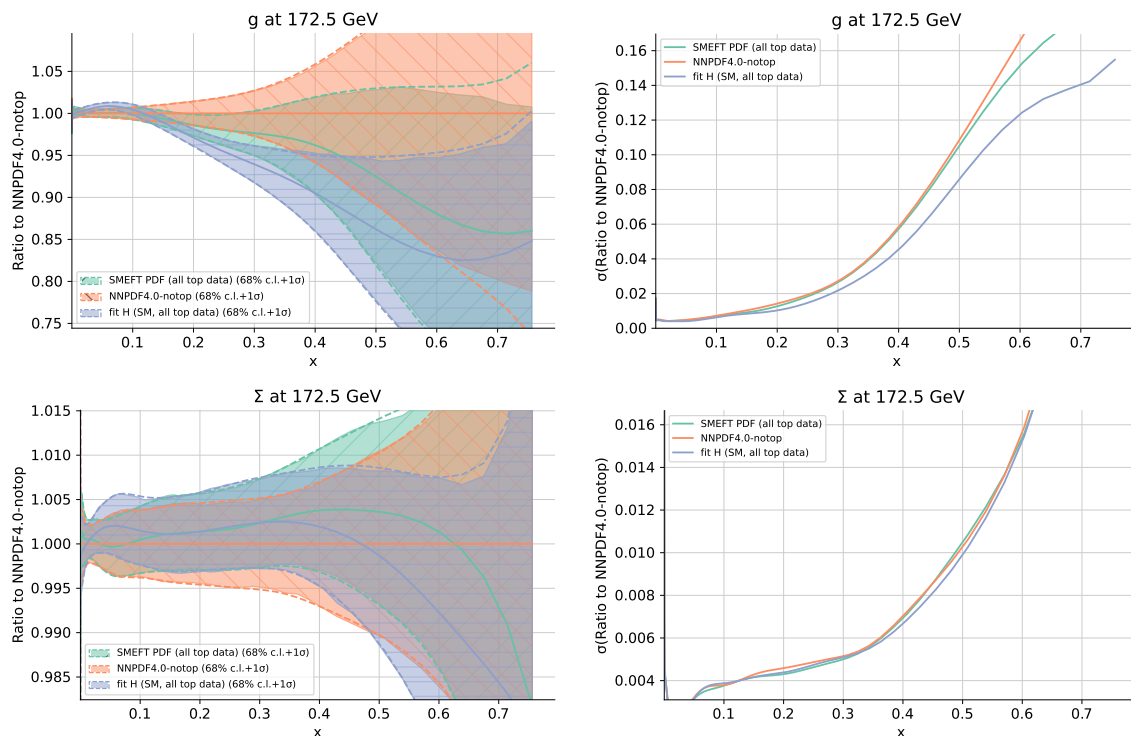
Figure 14. Left: a comparison between the gluon (upper panel) and quark singlet (lower panel) PDFs evaluated at $Q = m_t = 172.5\,\mathrm{GeV}$ in the large-x region. We display the `NNPDF4.0-notop` baseline, the SM-PDFs of fit H in table 11 which include the full top quark dataset, and their SMEFT-PDF counterparts. The results are normalised to the central value of the `NNPDF4.0-notop` fit. Right: the same comparison, but now for the relative 1σ PDF uncertainties.

linear, $\mathcal{O}(1/\Lambda^2)$, level in the EFT expansion. We do not perform simultaneous fits at the quadratic level due to shortcomings of the Monte Carlo replica method, on which `SIMUnet` is based; this is discussed in detail in appendix E, and will also be the subject of future work.

PDFs from a joint SMEFT-PDF fit. We begin by discussing the PDFs obtained through a joint fit of PDFs and Wilson coefficients from the complete LHC top quark dataset considered in this work. Simultaneously extracting the PDFs and the EFT coefficients from top quark data has a marked impact on the former, as compared to a SM-PDF baseline, but we shall see has much less impact on the latter, as compared to the results of the corresponding fixed-PDF EFT analyses. Figure 14 displays a comparison between the gluon and quark singlet PDFs, as well as of their relative 1σ PDF uncertainties, for the `NNPDF4.0-notop` baseline, the SM-PDFs of fit H in table 11 which include the full top quark dataset, and their SMEFT-PDF counterparts based on the same dataset. PDFs are compared in the large-x region for $Q = m_t = 172.5\,\mathrm{GeV}$. In the left panel they are normalised to the central value of the `NNPDF4.0-notop` fit.

While differences are negligible for the quark singlet PDF, both in terms of central values and uncertainties, they are more marked for the gluon PDF. Two main effects are observed therein. First, the central value of the SMEFT-PDF gluon moves upwards as

compared to the SM-PDF fit based on the same dataset, ending up halfway between the latter and the `NNPDF4.0-notop` fit. Second, uncertainties increase for the SMEFT-PDF determination as compared to the SM-PDFs extracted from the same data, becoming close to the uncertainties of the `NNPDF4.0-notop` fit except for $x \gtrsim 0.5$. In both cases, differences are restricted to the large-x region with $x \gtrsim 0.1$, where the impact of the dominant top quark pair production measurements is the largest.

The results of figure 14 for the gluon PDF therefore indicate that within a simultaneous extraction of the PDFs and the EFT coefficients, the impact of the top quark data on the PDFs is diluted, with the constraints it provides partially 'reabsorbed' by the Wilson coefficients. This said, there remains a pull of the top quark data as compared to the no-top baseline fit which is qualitatively consistent with the pull obtained in a SM-PDF determination based on the same dataset, albeit of reduced magnitude. Interestingly, as we show below, while the SMEFT-PDF gluon is significantly modified in the joint fit as compared to a SM-PDF reference, much smaller differences are observed at the level of the bounds on the EFT parameters themselves.

Figure 15 displays the same comparison as in figure 14 now for the case of the gluon-gluon and quark-gluon partonic luminosities at $\sqrt{s} = 13\,\text{TeV}$, as a function of the final-state invariant mass m_X. Consistently with the results obtained at the PDF level, one finds that the three luminosities are almost identical for $m_X \leq 1\,\text{TeV}$, and at higher invariant masses the SMEFT-PDF predictions are bracketed between the `NNPDF4.0-notop` fit from above and the SM-PDF which includes all top data (fit H in table 11) from below. Hence, the net effect of simultaneously fitting the PDFs and the EFT coefficients is a dilution of constraints provided by the top quark data on the former, which translates into larger PDF uncertainties (which end up being rather similar to those of `NNPDF4.0-notop`) and an increase in the large-m_X luminosity, e.g. of 5% in the gg case for $m_X \simeq 3\,\text{TeV}$, as compared to the SM-PDF luminosity.

This dilution arises because of an improved description of the top quark data included in the fit, especially in the high $m_{t\bar{t}}$ bins. In the SM-PDF case this can only be obtained by suppressing the large-x gluon, while in a SMEFT-PDF analysis this can also be achieved by shifting the EFT coefficients from their SM values. In other words, as compared to the `NNPDF4.0-notop` baseline, the gluon PDF experiences a suppression at large-x of up to 10% when fitting the top quark data, and this pull is reduced by approximately a factor two in the joint SMEFT-PDF determination due to the coherent effect of the linear EFT cross-sections.

It is worth mentioning that, since we include the SMEFT corrections by applying BSM factors computed bin-by-bin by taking ratios of the SMEFT contributions and the SM cross sections with a specific PDF set, a change of the PDFs can in principle translate into a change of these factors. Our current methodology relies therefore on performing this check a posteriori and in the case of modified BSM factors, the fit is reiterated with the new ones. However, in the present case, we find that the BSM factors are substantially unaffected and we do not reiterate the fit. The parton luminosities are indeed very similar to the ones of the PDF set `NNPDF40_nnlo_as_01180` used for the original calculation, with deviations of $2 - 3\%$ at most. At the current experimental sensitivity, changes of $\mathcal{O}(1\%)$ in the EFT corrections can be safely ignored as they do not impact in a significant way the fits.
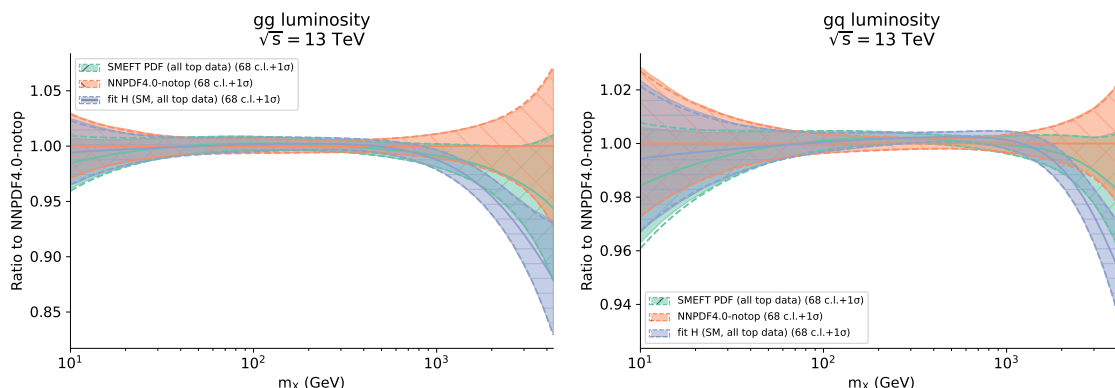
Figure 15. The gluon-gluon (left panel) and quark-gluon (right panel) partonic luminosities at $\sqrt{s} = 13\,\text{TeV}$ as a function of the final-state invariant mass m_X. We compare the `NNPDF4.0-notop` baseline fit with its SM-PDF counterpart including all top quark data considered (fit H in table 11) as well as with the SMEFT-PDF determination. Results are presented as the ratio to the central value of the `NNPDF4.0-notop` baseline.

EFT coefficients from a joint SMEFT-PDF fit. As opposed to the marked effect of the SMEFT-PDF interplay found for the large-x gluon, its impact is more moderate at the level of the bounds on the EFT coefficients, and is restricted to mild shifts in the central values and a slight broadening of the uncertainties. This is illustrated by figure 16, showing the posterior distributions for the Wilson coefficient c_{tG} associated to the chromomagnetic operator in the joint SMEFT-PDF determination, compared with the corresponding results from the fixed-PDF EFT analysis whose settings are described in section 5. The comparison is presented both for the fits which consider only top-quark pair production data and those based on the whole top quark dataset considered in this work. The leading effect of the chromomagnetic operator $\mathcal{O}_{tG}$ is to modify the total rates of $t\bar{t}$ production without altering the shape of the differential distributions, and hence it plays an important role in a simultaneous SMEFT-PDF determination based on top quark data.

For fits based only on inclusive $t\bar{t}$ data, as shown in the left panel of figure 16, the two posterior distributions are similar; the distribution based on the SMEFT-PDF analysis is slightly broader, approximately 10% so, as compared to the fixed-PDF EFT fit to the same measurements. This slight broadening is washed out in the fit to the full top quark dataset, as shown in the right panel of figure 16. In both cases, the determination of c_{tG} is consistent with the SM at a 95% CL, and the best-fit values of the coefficient are the same in the SMEFT-PDF and fixed-PDF EFT analyses. Hence, in the specific case of the chromomagnetic operator, the interplay between PDFs and EFT fits is rather moderate and restricted to a broadening of at most 10% in the 95% CL bounds. Similar comparisons have been carried out for other EFT coefficients as well as in the context of fits to a subset of the data and/or to a subset of the coefficients. We find that in general the impact of the SMEFT-PDF interplay translates to a broadening of the uncertainties in the EFT coefficients, which at most reaches 30%, and alongside which the best-fit values remain stable.[4]

---

[4]All results obtained with various subsets of the data and of the coefficients can be found at https://www.pbsp.org.uk/research/topproject.
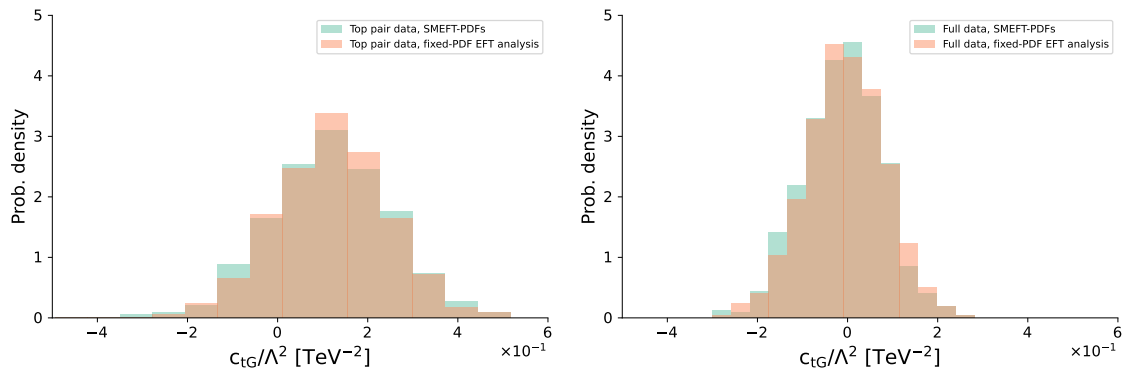
Figure 16. Posterior distributions for the Wilson coefficient c_{tG} associated to the chromomagnetic operator in the joint SMEFT-PDF determination, compared with the corresponding results from the fixed-PDF EFT analysis whose settings are described in section 5. We show results based on only top-quark pair production data (left) and in the whole top quark dataset considered in this work (right panel).

All in all, within the global fit based on the best available theory predictions, results for the EFT coefficients turn out to be very similar in the fixed-PDF EFT and SMEFT-PDF fits. This indicates that, provided a broad enough dataset and the most up-to-date theory calculations are used, the PDF dependence on the cross-sections entering an EFT interpretation of the LHC data is currently subdominant and can be neglected (this is not the case for the PDFs, see figure 15). Nevertheless, this statement applies only to the dataset currently available, and it is likely that the SMEFT-PDF interplay will become more significant in the future once HL-LHC measurements become available [187, 188], as demonstrated in the case of high-mass Drell-Yan [46].

The moderate impact of the SMEFT-PDF interplay on the Wilson coefficients for the full top quark dataset considered in this work is summarised in figure 17, which compares the 95% CL intervals of the 20 fitted Wilson coefficients relevant for the linear EFT fit obtained from the outcome of the joint SMEFT-PDF determination and the fixed-PDF EFT analysis. The latter is based on SM and EFT calculations performed with `NNPDF4.0-notop` as input; see also the description of the settings in section 5. The dashed horizontal line indicates the SM prediction, and some coefficients are multiplied by the indicated prefactor to facilitate the visualisation of the results. Figure 17 demonstrates that, other than slight broadenings and shifts in the central values, the results of the two analyses coincide.

Correlations. Figure 18 displays the correlation coefficients [177] between the SMEFT-PDFs and the Wilson coefficients evaluated at $Q = 172.5\,\text{GeV}$ as a function of x. Each panel displays the correlations of the coefficient c_k with the gluon and the total singlet Σ, total valence V, and non-singlet triplet T_3 PDFs, and we consider four representative EFT coefficients, namely c_{td}^8, c_{tu}^8, c_{tG}, and c_{tW}. The largest correlations within the EFT coefficients considered in this work are associated to the gluon PDF and four-fermion operators such as c_{td}^8 and c_{tu}^8 in the large-x region, peaking at $x \simeq 0.3$. Correlations for other values of x and for the quark PDFs are negligible for all operators entering the
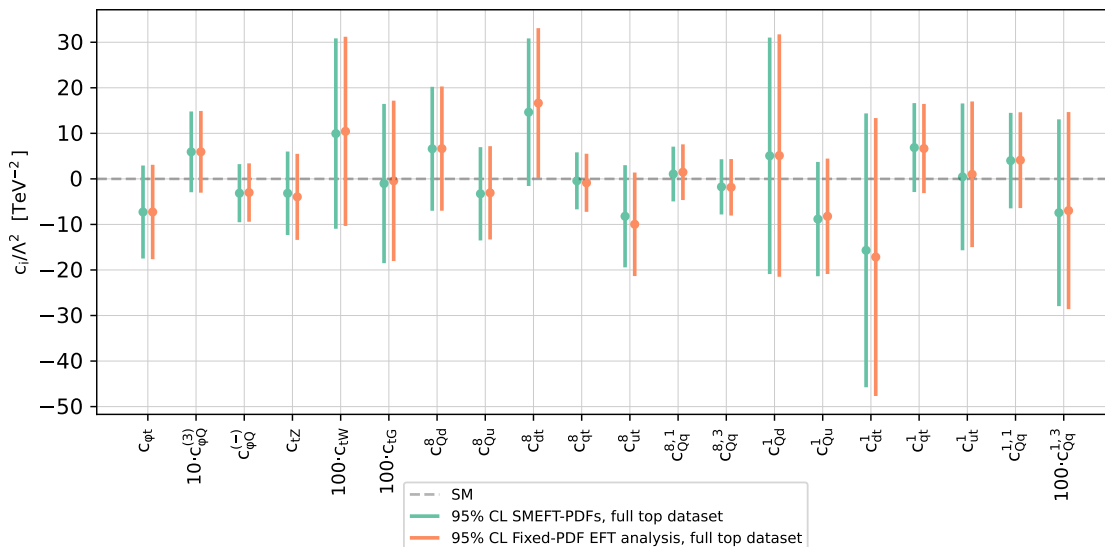
Figure 17. Comparison of the 95% CL intervals on the 20 Wilson coefficients considered in this work (in the linear EFT case) between the outcome of the joint SMEFT-PDF determination and that of the fixed-PDF EFT analysis. The latter is based on SM and EFT calculations performed with `NNPDF4.0-notop` as input, see also section 5. In both cases, results are based on the full top quark dataset being considered and EFT cross-sections are evaluated up to linear, $\mathcal{O}\left(\Lambda^{-2}\right)$, corrections. The dashed horizontal line indicates the SM prediction, $c_k = 0$. Note that some coefficients are multiplied by the indicated prefactor to facilitate the visualisation of the results.

analysis. We note that future data with an enhanced coverage of the high-$m_{t\bar{t}}$ region in top quark pair-production might alter this picture, given that for $m_{t\bar{t}} \gtrsim 3\,\mathrm{TeV}$ the $q\bar{q}$ luminosity starts to become more relevant and eventually dominates over the gg contribution.

Residuals. Finally, figure 19 displays a similar comparison as in figure 17 now at the level of the 68% CL fit residuals defined as

$$R_n = \frac{c_n^*}{\sigma_n},\tag{6.1}$$

where c_n^* and σ_n are the median value and the standard deviation of the Wilson coefficient c_n respectively, with $n = 1, \ldots, N$, where N is the number of operators. The outcome of the joint SMEFT-PDF determination is compared with that of a fixed-PDF EFT analysis where we use as input for the theory calculations the SMEFT-PDFs obtained in the joint fit, rather than the `NNPDF4.0-notop` set. That is, in both cases the information provided by the top quark data on the PDFs and Wilson coefficients is accounted for, but in one case the cross-correlations are neglected whereas they are accounted for in the other. The residuals are similar in the two cases; they are slightly bigger (in absolute value) in the fixed-PDF case in which the correlations between the SMEFT-PDFs and the EFT coefficients are ignored. This analysis further emphasises that, for the currently available top quark data, the cross-talk between PDFs and EFT degrees of freedom does not significantly modify the posterior distributions in the space spanned by the Wilson coefficients.
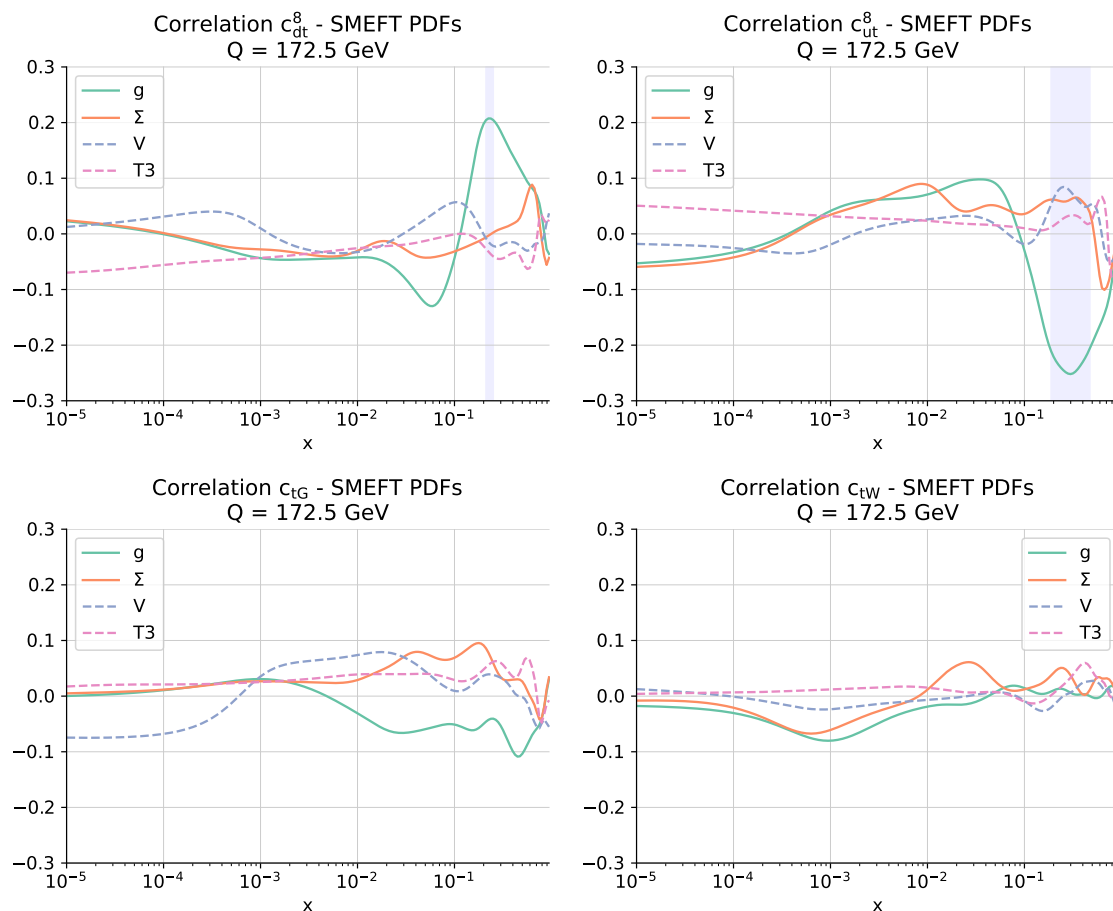
Figure 18. The correlation coefficient $\rho(f_i, c_k)$ between the SMEFT-PDFs f_i and the Wilson coefficients c_k evaluated at $Q = 172.5\,\mathrm{GeV}$ as a function of x. Each panel displays the correlations of the coefficient c_k with the gluon and the total singlet Σ, total valence V, and non-singlet triplet T_3 PDFs. We provide results for representative EFT coefficients, namely c_{td}^8, c_{tu}^8, c_{tG}, and c_{tW}. The largest correlations within the EFT coefficients considered in this work are associated to four-fermion operators such as c_{td}^8 and c_{tu}^8.

In summary, on the one hand we find that from the point of view of a PDF determination, SM-PDFs and SMEFT-PDFs extracted from top quark data differ by an amount comparable to their respective uncertainties in the case of the large-x gluon. On the other hand, at the level of Wilson coefficients the results are unchanged irrespective of the PDF set used as input for the theory calculations; that is, bounds based on `NNPDF4.0-notop` or the SMEFT-PDFs are almost the same. Hence, while EFT interpretations of top quark data can safely ignore the PDF dependence, at least for the settings adopted in this work, a global PDF fit could be significantly distorted if BSM physics were to be present in the large-energy top quark distributions. We use these findings in appendix A to provide recommendations for interpretations of LHC top quark data in the context of either PDF or EFT analyses.

Figure 19. The 68% CL residuals, eq. (6.1), for the same Wilson coefficients displayed in figure 17, comparing the outcome of the joint SMEFT-PDF determination with that of a fixed-PDF EFT analysis. In the latter, we use as input for the theory calculations the SMEFT-PDFs obtained in the joint fit rather than the `NNPDF4.0-notop` set used in section 5. The horizontal dashed lines indicate the $\pm 1\sigma$ regions.

7 Conclusions and outlook

In this work we have carried out a comprehensive theoretical interpretation of top quark data from the LHC, including all available measurements based on the full integrated luminosity of Run II, in terms of the proton PDFs and of Wilson coefficients in the SMEFT. We have integrated a global determination of PDFs with the fit of dimension-six Wilson coefficients modifying top quark interactions into a simultaneous determination of the SMEFT-PDFs and the EFT coefficients. The main outcome of our analysis is the assessment of the conditions under which the usual assumptions of SM-PDFs and fixed-PDF EFT analyses are valid, and establishing when the SMEFT-PDF interplay cannot be neglected without introducing a sizeable bias to the results. Furthermore, we have provided strategies to disentangle eventual BSM signals from QCD effects in the interpretation of these top quark measurements.

As a by-product of this determination of the SMEFT-PDFs, we have also presented state-of-the-art SM-PDF and fixed-PDF EFT interpretations of top quark data based on the broadest LHC experimental dataset to date, therefore fully exploiting the information contained in the legacy Run II top quark measurements. From the SM-PDF study, we have quantified the impact of the recent Run II top quark production data on the large-x gluon, and assessed their compatibility with the information provided by the datasets already included in previous global fits such as `NNPDF4.0` . From the fixed-PDF SMEFT analysis, we have benchmarked the `SIMUnet` performance reproducing the SMEFiT results, determined the improved constraints provided by the full-luminosity Run II data, and compared our findings with the results presented from other groups. All in all, we demonstrate the unpar-

alleled sensitivity of Run II top quark data both to probe the proton structure and to search for signatures of new physics arising as anomalous top quark properties and interactions.

Finally, we have uncovered the inaccuracy of the Monte Carlo error propagation in the case of a SMEFT fit in which the observables are dominated by quadratic SMEFT corrections, as we found a significant disagreement between quadratic SMEFT-only fits produced using the `SIMUnet` Monte-Carlo based methodology and the SMEFiT Nested Sampling option. Such a disagreement can be traced back to a deficiency in the Monte-Carlo sampling method used to propagate experimental error to the SMEFT coefficients, which has prevented us from applying the `SIMUnet` framework to joint SMEFT-PDF fits with quadratic EFT calculations. See appendix E for a more detailed discussion. The pitfall of the Monte Carlo sampling naturally raises a question on whether this affects not only the fit of the SMEFT coefficients, but also the fit of PDFs. Although the PDF fits are dominated by linear terms and although this question started to be investigated in ref. [189] using toy PDFs and no significant misrepresentation of PDF uncertainties was found (apart from a slight overestimate of PDF uncertainties by using Monte Carlo error propagation), we believe that this should be fully addressed beyond simple toy models. Such an investigation is being pursued [190] and will be presented in future works.

To coclude, the results presented in this work could be extended in a number of directions. One of the most pressing matters is the provision of a simultaneous determination of PDFs with EFT coefficients entering at *quadratic order*; the difficulties associated with using quadratics with our current `SIMUnet` methodology are described in appendix E. This will involve significant modification to the Monte-Carlo replica method used by `SIMUnet`, and will be the subject of future work.

On a different note, top quark measurements at the LHC have also been used to determine the strong coupling $\alpha_s(m_Z)$ as well as the top quark mass m_t, both independently and simultaneously with the PDFs [7]. It could hence be interesting to extend the `SIMUnet` framework to account for the determination of other SM parameters from top quark data in addition to the PDFs, as it is sketched in the original `SIMUnet` publication [50], and to study their interplay with eventual new physics effects. Additionally, it would be interesting to extend future versions of the `SIMUnet` methodology to account for potential renormalisation group effects on the Wilson coefficients [174, 191–193]. One may also want to assess the SMEFT-PDF interplay for other types of processes beyond those considered so far, and in particular study inclusive jet and di-jet production, which are instrumental for PDF fits in the same region constrained by the top data [194] and also provide information on a large number of poorly-known SMEFT operators. In the longer term, even processes that are never used for PDF fits, such as Higgs or gauge boson pair production, may reach a precision that makes them competitive, and in this case the only option to account for this information is by means of the simultaneous determination of PDFs and EFT coefficients.

Another possible development of the `SIMUnet` methodology would be to establish the interplay between PDFs and model parameters such as masses and couplings in specific UV-complete BSM scenarios by using the SMEFT as a matching bridge [195]. It would also be interesting to extend the joint SMEFT-PDF determination as implemented in `SIMUnet` to novel types of observables with enhanced or even optimal sensitivity to

either the PDFs of the EFT coefficients, such as the ML-assisted unbinned multivariate observables introduced in ref. [196]. Finally, following along the lines of the HL-LHC projections for high-mass W and Z production presented in [46], one can study the projected SMEFT-PDF synergies at future colliders, certainly at the HL-LHC but also at the Electron Ion Collider [197], sensitive to directions in the SMEFT parameter space not covered by LHC data, and at the Forward Physics Facility [198, 199], where proton and nuclear structure can be probed while also obtaining information on anomalous neutrino interactions parametrised by extensions of the SMEFT.

All results presented in this paper, as well as a broader selection of results, including fits performed with various theory settings, on subsets of datasets and/or operators are available at the webpage:

  https://www.pbsp.org.uk/research/topproject.

Readers are encouraged to contact the authors, should they need any specific SM-PDF or SMEFT-PDF fits that can be obtained with our methodology.

In order to facilitate that studies similar to those presented in this paper are carried out by the LHC experimental collaborations with their own data, we plan to release the open source `SIMUnet` framework together with documentation and user-friendly examples in a future publication, in a way that can be seamlessly integrated with the NNPDF fitting code [52]. The availability of `SIMUnet` as open source code would allow the LHC collaborations to study the SMEFT-PDF interplay in their own internal analyses, and in particular to modify the input datasets and EFT operator basis to match their own settings. In the meantime, we provide in appendix A usage recommendations concerning the joint interpretation of LHC measurements in terms of both PDFs and EFT coefficients based on the findings of this work.

As the LHC approaches its high-luminosity era, it becomes more and more important to develop novel analysis frameworks that make possible the full exploitation of the information contained in the legacy LHC measurements. The results presented in this work contribute to this program by demonstrating how to achieve a consistent simultaneous determination of the PDFs and EFT parameters from the same top quark dataset, bypassing the need for the assumptions required for the SM-PDF and fixed-PDF EFT interpretations.

Acknowledgments

A   Recommendations to quantify SMEFT-PDF interplay

The strategy presented here, based on carrying out a simultaneous determination of the SMEFT-PDFs and the Wilson coefficients, makes possible a quantitative assessment of the interplay between the PDF and SMEFT sensitivity when interpreting a given set of LHC measurements. As discussed in section 7, the SIMUnet framework will be made public only at a later stage. In the meantime, researchers interested in quantifying this interplay in their own analysis can use available open-source platforms like xFitter, as done by the CMS collaboration in their QCD and EFT analyses of double-differential inclusive jet cross sections at 13 TeV [49]. In several cases, however, a joint analysis may not be necessary, and one can estimate the possible role of the SMEFT-PDF interplay by means of a simplified approach. The discussion is presented here for the case of top quark measurements, but it can be straightforwardly extended to any other class of measurements.

- The most conservative strategy is to introduce a hard boundary between those processes used for PDF fits and those entering SMEFT interpretations. For instance, for a SMEFT analysis of LHC top quark measurements one can use PDF fits without top data, such as the NNPDF4.0-notop variant. We recall that with the NNPDF open-source code [52] one can produce fit variants with arbitrary input datasets.

- A somewhat less conservative assumption would be to introduce a hard boundary, not at the process level, but rather at the kinematic level. In the case of inclusive top quark pair production data, one can introduce a threshold on the top quark pair invariant mass $m_{t\bar{t}}$ such that data points below the threshold are used for the PDF fit and above it for the EFT interpretation, thus benefiting from the increased sensitivity to BSM effects in the high-energy tails of the LHC distributions. Again, using public PDF fitting codes one can produce fit variants with tailored kinematical cuts to separate the "PDF" region from the "EFT" region. One drawback of this approach is that there is never a clear-cut separation between these regions; depending on the EFT operators considered, the high-energy region may not be the dominant one.

- As demonstrated in this work, many measurements of relevance for EFT interpretations have limited sensitivity to PDFs and vice-versa. In such cases, the SMEFT-PDF interplay can be safely neglected. One can determine under which settings this condition is satisfied by adding the dataset under consideration either to a global PDF fit (using for example the NNPDF fitting code [52]) or to a global SMEFT fit (using the SMEFiT or FitMaker frameworks). If the results of either the PDF or EFT fits are unchanged, one can safely neglect the SMEFT-PDF interplay in this case.

- If instead the above analysis shows that a given dataset provides non-trivial information on both the PDFs and on the EFT Wilson coefficients, the next step is to compute the correlation between the EFT coefficients and the PDFs, as it is done in eq. (5.4). If the correlation is negligible across the whole $x-$range for all operators included in the analysis, one can safely ignore the interplay. On the other hand, if the correlations are visible but not dominant ($20\% \lesssim \rho \lesssim 40\%$) one can expect the SMEFT bounds to be stable upon PDF variations, although SMEFT PDFs might end up being visibly different with respect to the SM PDFs, as we observe in our current analysis.

- If significant correlations are observed (with $\rho \gtrsim 40\%$), the only options for a consistent theoretical interpretation are either introducing a hard boundary in the analysis (either at the process level or at the kinematic level) or to carry out the joint SMEFT-PDF interpretation of the full dataset using `SIMUnet` or other available tools.

Furthermore, it should be emphasised that, as opposed to the PDF case, the sensitivity to EFT coefficients of a given measurement depends in general on the choice of EFT operator considered. Hence, the above considerations assume a specific choice of operators and could be different if this choice is varied.

B EFT operator basis

Table 14 lists the dimension-six SMEFT operators relevant for this analysis, together with the corresponding degrees of freedom (DoF) entering the fit. In terms of the flavour assumptions, we follow the LHC top quark working group recommendations [29], which were also adopted in [53, 102]. The flavour symmetry group is given by $U(3)_l \times U(3)_e \times U(3)_d \times U(2)_u \times U(2)_q$, i.e. one singles out operators that contain top quarks (right-handed t and $SU(2)$ doublet Q). This means that we work in a 5-flavour scheme, where the only massive fermion in the theory is the top quark. The upper part in table 14 defines the relevant two-fermion operators modifying the interactions of the third-generation quarks. We also indicate the notation used for the associated Wilson coefficients; those in brackets are not degrees of freedom entering the fit, and instead the two additional DoF defined in the middle table are used. The bottom table defines the four-fermion DoF entering the fit, expressed in terms of the corresponding four-fermion Wilson coefficients associated to dimension-six SMEFT operators in the Warsaw basis.

EFT cross-sections at the linear and quadratic order in the EFT expansion are computed for the datasets considered in this analysis and for the DoF defined in table 14, both at LO and NLO in perturbative QCD. Furthermore we use the m_W-scheme, meaning that the four EW inputs are $\{m_W, G_F, m_h, m_Z\}$. In particular, the electric charge e becomes a dependent parameter and is shifted by the effects of higher-dimensional operators.

C Benchmarking with SMEFiT

Here we benchmark the performance of `SIMUnet` when operating as a fixed-PDF EFT fitter, by means of a tuned comparison with the SMEFiT framework when identical theory

Operator	Coefficient	Definition
$\mathcal{O}_{\varphi Q}^{(1)}$	$-\ (c_{\varphi Q}^{(1)})$	$i\big(\varphi^\dagger \overset{\leftrightarrow}{D}_\mu \varphi\big)\big(\bar{Q}\,\gamma^\mu Q\big)$
$\mathcal{O}_{\varphi Q}^{(3)}$	$c_{\varphi Q}^{(3)}$	$i\big(\varphi^\dagger \overset{\leftrightarrow}{D}_\mu \tau_I \varphi\big)\big(\bar{Q}\,\gamma^\mu \tau^I Q\big)$
$\mathcal{O}_{\varphi t}$	$c_{\varphi t}$	$i\big(\varphi^\dagger \overset{\leftrightarrow}{D}_\mu \varphi\big)\big(\bar{t}\,\gamma^\mu t\big)$
$\mathcal{O}_{tW}$	c_{tW}	$i\big(\bar{Q}\tau^{\mu\nu}\tau_I\, t\big)\tilde{\varphi}\, W_{\mu\nu}^I + \text{h.c.}$
$\mathcal{O}_{tB}$	$-\ (c_{tB})$	$i\big(\bar{Q}\tau^{\mu\nu}\, t\big)\tilde{\varphi}\, B_{\mu\nu} + \text{h.c.}$
$\mathcal{O}_{tG}$	c_{tG}	$i\big(\bar{Q}\tau^{\mu\nu}\, T_A\, t\big)\tilde{\varphi}\, G_{\mu\nu}^A + \text{h.c.}$

DoF	Definition
$c_{\varphi Q}^{(-)}$	$c_{\varphi Q}^{(1)} - c_{\varphi Q}^{(3)}$
c_{tZ}	$-\sin\theta_W c_{tB} + \cos\theta_W c_{tW}$

DoF	Definition (Warsaw basis)
c_{QQ}^1	$2c_{qq}^{1(3333)} - \frac{2}{3}c_{qq}^{3(3333)}$
c_{QQ}^8	$8c_{qq}^{3(3333)}$
c_{Qt}^1	$c_{qu}^{1(3333)}$
c_{Qt}^8	$c_{qu}^{8(3333)}$
c_{tt}^1	$c_{uu}^{(3333)}$
$c_{Qq}^{1,8}$	$c_{qq}^{1(i33i)} + 3c_{qq}^{3(i33i)}$
$c_{Qq}^{1,1}$	$c_{qq}^{1(ii33)} + \frac{1}{6}c_{qq}^{1(i33i)} + \frac{1}{2}c_{qq}^{3(i33i)}$
$c_{Qq}^{3,8}$	$c_{qq}^{1(i33i)} - c_{qq}^{3(i33i)}$
$c_{Qq}^{3,1}$	$c_{qq}^{3(ii33)} + \frac{1}{6}\big(c_{qq}^{1(i33i)} - c_{qq}^{3(i33i)}\big)$
c_{tq}^8	$c_{qu}^{8(ii33)}$
c_{tq}^1	$c_{qu}^{1(ii33)}$
c_{tu}^8	$2c_{uu}^{(i33i)}$
c_{tu}^1	$c_{uu}^{(ii33)} + \frac{1}{3}c_{uu}^{(i33i)}$
c_{Qu}^8	$c_{qu}^{8(33ii)}$
c_{Qu}^1	$c_{qu}^{1(33ii)}$
c_{td}^8	$c_{ud}^{8(33jj)}$
c_{td}^1	$c_{ud}^{1(33jj)}$
c_{Qd}^8	$c_{qd}^{8(33jj)}$
c_{Qd}^1	$c_{qd}^{1(33jj)}$

Table 14. Upper table: definition of the two-fermion dimension-six SMEFT operators relevant for this analysis. These operators modify the interactions of the third-generation quarks. We also indicate the notation for the associated Wilson coefficients; those in brackets are not degrees of freedom entering the fit. Middle table: the two additional degrees of freedom used in the fit involving two-fermion operators, defined in terms of the coefficients of the upper table. Bottom table: the four-fermion degrees of freedom considered here, expressed in terms of the corresponding four-fermion Wilson coefficients of dimension-six SMEFT operators in the Warsaw basis.
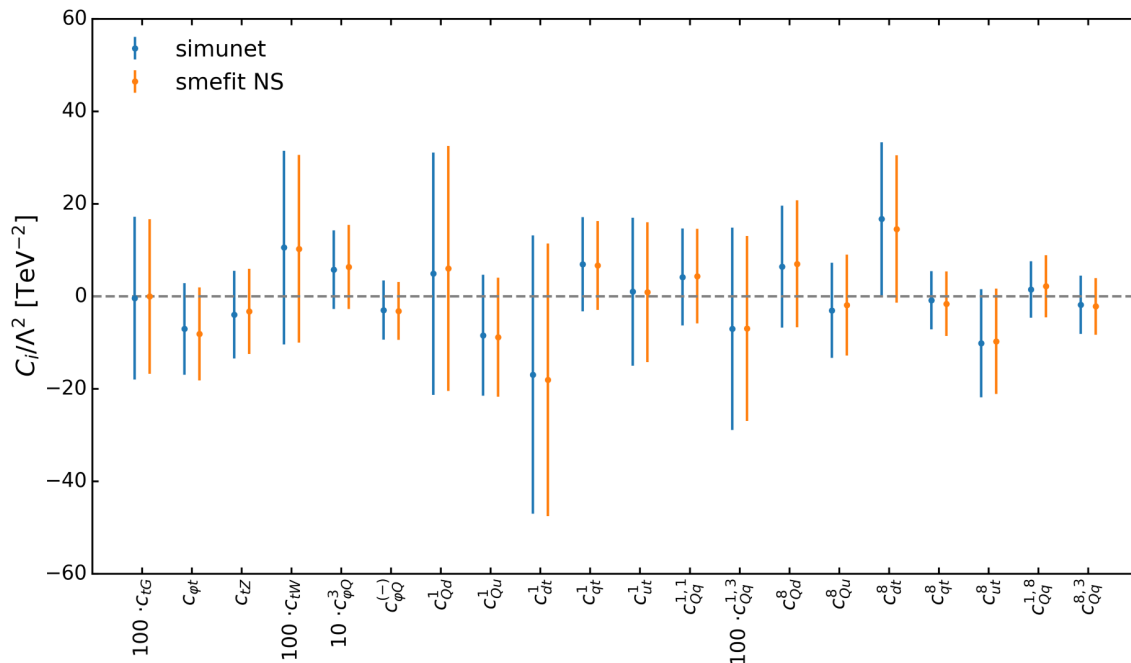
Figure 20. Comparison between the 95% CL intervals obtained with `SIMUnet` in the fixed-PDF linear EFT case (full top quark dataset) with the corresponding results based on the SMEFiT framework (with the Nested Sampling option). In both cases the same experimental and theoretical inputs are adopted, see text for more details. These `SIMUnet` results coincide with those displayed in figure 17 in the fixed-PDF EFT analysis.

and experimental inputs are used in both cases. SMEFiT currently provides two strategies to determine posterior distributions in the EFT parameter space. The first, known as MCfit, is based on the Monte Carlo replica method followed by parameter optimisation (as similarly used in `SIMUnet`). The second is based on Nested Sampling (NS), a purely Bayesian approach for parameter inference. At the linear EFT level, SMEFiT results based on MCfit and NS are identical [53]. At the quadratic level [182], however, the MCfit approach is affected by the potential pitfalls described in appendix E, which also prevent the use of `SIMUnet` for joint SMEFT-PDF fits in the presence of quadratic EFT corrections. We demonstrate now that, for a linear EFT determination and a common choice of inputs, the results obtained using the two frameworks display excellent agreement.

Figure 20 displays this comparison between the marginalised 95% CL intervals of `SIMUnet` and SMEFiT (operating in the Nested Sampling mode) for a linear EFT fit that takes as input in both cases the full dataset discussed in this work as well as the same SM and EFT theory calculations (and hence also the same operator basis). To ensure identical inputs, a parser has been written that automatically converts data and theory files from the `SIMUnet` to the SMEFiT standard formats. The `SIMUnet` results displayed here coincide with those displayed in figure 17 in the fixed-PDF EFT analysis. Satisfactory agreement between the two fitting frameworks is obtained, and similar benchmarks have been successfully performed for other variants of the fixed-PDF EFT fits presented in this work.

This benchmark study ensures that the optimisation algorithm adopted by `SIMUnet` is suitable both for PDF determinations (since it is based on the NNPDF4.0 settings) as well as for the determination of the EFT coefficients. The optimisation settings that provide the best performance of the SM-PDF and fixed-PDF EFT analyses are then combined for the simultaneous SMEFT-PDF extraction, where all weights of the `SIMUnet` network are allowed to be constrained by the data.

D Fit quality

Here we summarise the χ^2 (computed using the experimental definition, eq. 2.3) for the key PDF and EFT analyses presented in this work. Tables 15, 16 and 17 display the values of the χ^2 per data point for representative PDF and EFT fits, where for each dataset we also indicate the number of data points $n_{\rm dat}$. Specifically, we consider: (i) three SM-PDF fits: `NNPDF4.0-notop`, `NNPDF4.0` , and Fit H in table 11 (using our full top quark dataset); (ii) two fixed-PDF EFT fits, one based on `NNPDF4.0-notop` as input, and the other using Fit H as input; (iii) the outcome of the simultaneous SMEFT-PDF determination. We show sequentially the non-top datasets, the $t\bar{t}$ inclusive and associated production datasets, and the single-top inclusive and associated production datasets. We also provide the total χ^2 for separate groups of processes, including the full non-top and top datasets, and well as for their total sum.

In tables 16 and 17, entries in italics indicate datasets that do not enter the corresponding SM-PDF fit; for these datasets, we evaluate the associated χ^2 values *a posteriori* using the resulting PDFs. For instance, all top data is removed from the `NNPDF4.0-notop` fit, and for some top quark observables such as four-heavy-quark production, PDF dependence is neglected. Furthermore, we note that the χ^2 values for the `NNPDF4.0-notop` fit and for the fixed-PDF EFT fit based on `NNPDF4.0-notop` are identical in table 15, since the latter uses only top data as input. For the same reason, the entries in the columns for the SM-PDF Fit H and the fixed-PDF EFT fits based on Fit H are the same.

Several observations can be drawn from the results presented in tables 15, 16 and 17. First of all, we note that the description of the non-top data is improved in the simultaneous SMEFT-PDF analysis, with $\chi^2 = 1.137$ to be compared with the analogous SM-PDF analysis (Fit H) which leads to $\chi^2 = 1.146$. Given that we have $n_{\rm dat} = 4535$ no-top data points, this improvement corresponds to 40 units in the absolute χ^2. This improvement cannot traced back to a specific measurements or group of processes. In turn, this improvement is also reflected for the full dataset, whose χ^2 is the lowest (1.126 for $n_{\rm dat} = 4710$ points) in the SMEFT-PDF analysis. This lower χ^2 as compared to the SM-PDF baseline fits is not surprising, taking into account the fact that (as discussed in section 5) we now have more degrees of freedom in the fit, and in particular the inclusion of linear EFT corrections is instrumental in improving the χ^2 to the top quark data as compared to the SM-PDF fits. We also find that the fit quality to the top datasets is similar in the joint SMEFT-PDF analysis and in the fixed-PDF EFT fits with either `NNPDF4.0-notop` or Fit H as input PDFs, with $\chi^2 = 0.848, 0.868, 0.839$ respectively in each case for $n_{\rm dat} = 175$. The small differences between the three cases (less than five units in absolute χ^2) are consistent with

Dataset	$n_{\rm dat}$	$\chi^2/n_{\rm dat}$					
		SM-PDF fits			**Fixed-PDF EFT fits**		**SMEFT-PDFs**
		NNPDF4.0 (no top)	NNPDF4.0	Fit H	NNPDF4.0 (no top)	Fit H	Joint fit
SLAC	67	0.768	0.758	0.753	0.768	0.753	0.761
BCDMS	581	1.265	1.247	1.251	1.265	1.251	1.245
NMC	325	1.298	1.306	1.304	1.298	1.304	1.292
CHORUS	832	0.899	0.900	0.901	0.899	0.901	0.898
NuTeV	76	0.398	0.425	0.423	0.398	0.423	0.393
HERA	1208	1.204	1.203	1.204	1.204	1.204	1.200
Total DIS	**3089**	**1.122**	**1.119**	**1.119**	**1.122**	**1.119**	**1.114**
E886 $\sigma_{\rm DY}^d/\sigma_{\rm DY}^p$	15	0.527	0.524	0.536	0.527	0.536	0.544
E886 $\sigma_{\rm DY}^p$	89	1.566	1.559	1.615	1.566	1.615	1.602
E605 $\sigma_{\rm DY}^p$	85	0.456	0.456	0.460	0.456	0.460	0.464
E906 $\sigma_{\rm DY}^d/\sigma_{\rm DY}^p$	6	0.827	0.885	0.886	0.827	0.886	0.955
Total fixed-target DY	**195**	**0.981**	**0.979**	**1.008**	**0.981**	**1.008**	**1.006**
CDF $d\sigma_Z/dy_Z$	28	1.234	1.280	1.275	1.234	1.275	1.188
D0 $d\sigma_Z/dy_Z$	28	0.637	0.644	0.640	0.637	0.640	0.632
D0 $W \to \mu\nu$ asy.	9	1.958	1.929	1.905	1.958	1.905	1.641
ATLAS low-mass DY 7 TeV	6	0.875	0.879	0.883	0.875	0.883	0.888
ATLAS high-mass DY 7 TeV	5	1.694	1.691	1.661	1.694	1.661	1.629
ATLAS W, Z 7 TeV ($\mathcal{L} = 35$ pb^{-1})	30	1.000	0.988	1.000	1.000	1.000	0.991
ATLAS W, Z 7 TeV ($\mathcal{L} = 4.6$ fb^{-1})	61	1.689	1.686	1.685	1.685	1.685	1.599
ATLAS low-mass DY 2D 8 TeV	60	1.203	1.216	1.225	1.203	1.225	1.207
ATLAS high-mass DY 2D 8 TeV	48	1.123	1.110	1.118	1.123	1.118	1.147
ATLAS $\sigma_{W,Z}^{\rm tot}$ 13 TeV	3	0.727	0.778	0.706	0.727	0.706	0.471
CMS W electron asymmetry 7 TeV	11	0.836	0.833	0.836	0.836	0.836	0.892
CMS W muon asymmetry 7 TeV	11	1.728	1.714	1.733	1.728	1.733	1.699
CMS DY 2D 7 TeV	110	1.367	1.361	1.353	1.367	1.353	1.339
CMS W rapidity 8 TeV	22	1.375	1.361	1.365	1.375	1.365	1.162
LHCb $Z \to ee$ 7 TeV	9	1.630	1.648	1.636	1.630	1.636	1.669
LHCb $Z \to ee$ 8 TeV	17	1.272	1.325	1.334	1.272	1.334	1.266
LHCb $W, Z \to \mu$ 7 TeV	29	2.032	1.948	1.902	2.032	1.902	1.896
LHCb $W, Z \to \mu$ 8 TeV	30	1.494	1.426	1.417	1.494	1.417	1.386
LHCb $Z \to ee$ 13 TeV	15	1.701	1.726	1.717	1.701	1.717	1.640
LHCb $Z \to \mu\mu$ 13 TeV	16	0.970	0.993	0.978	0.970	0.978	0.928
Total inclusive gauge boson production	**548**	**1.344**	**1.339**	**1.335**	**1.344**	**1.335**	**1.297**
ATLAS $W^{\pm}$+jet 8 TeV	30	0.954	0.959	0.957	0.954	0.957	0.973
ATLAS Z p_T 8 TeV ($p_T, m_{\ell\ell}$)	44	0.911	0.905	0.904	0.911	0.904	0.910
ATLAS Z p_T 8 TeV (p_T, y_Z)	48	0.909	0.898	0.896	0.909	0.896	0.924
ATLAS incl. jets 8 TeV, $R = 0.6$	171	0.682	0.687	0.687	0.682	0.687	0.667
ATLAS dijets 7 TeV, $R = 0.6$	90	2.184	2.149	2.125	2.184	2.125	. 2.177
ATLAS isolated γ prod. 13 TeV	53	0.811	0.763	0.753	0.811	0.753	0.813
CMS Z p_T 8 TeV	28	1.447	1.401	1.407	1.447	1.407	1.473
CMS incl. jets 8 TeV	185	1.140	1.183	1.200	1.140	1.200	1.118
CMS dijets 7 TeV	54	1.788	1.810	1.809	1.788	1.809	1.800
Total jets, Z p_T and isolated photon	**703**	**1.154**	**1.157**	**1.158**	**1.154**	**1.158**	**1.148**
Total non-top data	**4535**	**1.148**	**1.145**	**1.146**	**1.148**	**1.146**	**1.137**

Table 15. The values of the χ^2 per data point (using the experimental definition of eq. (2.3)) for key PDF and EFT analyses presented in this work. For each dataset, we indicate the number of data points $n_{\rm dat}$ and the χ^2 values for *i)* three SM-PDF fits: `NNPDF4.0-notop`, `NNPDF4.0`, and Fit H in table 11 (full top quark dataset); *ii)* two fixed-PDF EFT fits: one based on `NNPDF4.0-notop` as input, and the other using Fit H as input; and finally *iii)* the outcome of the simultaneous SMEFT-PDF determination. Here we restrict ourselves to the non-top datasets considered in this analysis, the corresponding results for the top quark observables are provided in tables 16 and 17. We also provide the total χ^2 for separate groups of processes as well as for the full non-top dataset. The χ^2 values for the `NNPDF4.0-notop` fit and for the fixed-PDF EFT fit based on `NNPDF4.0-notop` are identical, since the latter uses only top data as input. For the same reason, the entries in the columns for the SM-PDF Fit H and the fixed-PDF EFT fits based on Fit H are the same.

Dataset	$n_{\rm dat}$	$\chi^2/n_{\rm dat}$					
		SM-PDF fits			**Fixed-PDF EFT fits**		**SMEFT-PDFs**
		NNPDF4.0 (no top)	NNPDF4.0	Fit H	NNPDF4.0 (no top)	Fit H	Joint fit
ATLAS $\sigma(t\bar t)$, dilepton, 7 TeV	1	*4.100*	4.540	4.599	1.733	1.824	1.974
ATLAS $\sigma(t\bar t)$, dilepton, 8 TeV	1	*0.005*	0.021	0.023	0.405	0.382	0.307
ATLAS $1/\sigma d\sigma/dm_{t\bar t}$, dilepton, 8 TeV	5	*0.262*	0.284	0.291	0.305	0.309	0.310
ATLAS $\sigma(t\bar t)$, ℓ+jets, 8 TeV	1	*0.218*	0.270	0.277	0.002	0.001	0.000
ATLAS $1/\sigma d\sigma/d|y_t|$, ℓ+jets, 8 TeV	4	*6.266*	3.219	2.827	1.178	1.021	1.144
ATLAS $1/\sigma d\sigma/d|y_{t\bar t}|$, ℓ+jets, 8 TeV	4	*7.978*	3.725	3.332	2.883	2.105	2.646
ATLAS $\sigma(t\bar t)$, dilepton, 13 TeV	1	*1.388*	*1.397*	1.374	0.004	0.004	0.002
ATLAS $\sigma(t\bar t)$, hadronic, 13 TeV	1	*0.231*	0.231	0.230	0.082	0.082	0.092
ATLAS $1/\sigma d^2\sigma/d|y_{t\bar t}|dm_{t\bar t}$, hadronic, 13 TeV	10	*2.276*	2.007	1.928	1.915	1.824	1.857
ATLAS $\sigma(t\bar t)$, ℓ+jets, 13 TeV	1	*0.489*	0.492	0.485	0.004	0.004	0.013
ATLAS $1/\sigma d\sigma/dm_{t\bar t}$, ℓ+jets, 13 TeV	8	*1.607*	1.788	1.835	2.087	2.072	2.090
CMS $\sigma(t\bar t)$, combined, 5 TeV	1	*0.442*	0.542	0.555	0.211	0.236	0.239
CMS $\sigma(t\bar t)$, combined, 7 TeV	1	*0.795*	1.032	1.065	0.004	0.010	0.027
CMS $\sigma(t\bar t)$, combined, 8 TeV	1	*0.170*	0.247	0.258	0.174	0.156	0.103
CMS $1/\sigma d^2\sigma/d|y_{t\bar t}|dm_{t\bar t}$, dilepton, 8 TeV	16	*1.094*	0.994	0.963	0.519	0.580	0.514
CMS $1/\sigma d\sigma/d|y_{t\bar t}|$, ℓ+jets, 8 TeV	9	*2.127*	1.245	1.131	0.928	0.995	0.935
CMS $\sigma(t\bar t)$, dilepton, 13 TeV	1	*0.064*	0.063	0.066	0.680	0.680	0.600
CMS $1/\sigma d\sigma/dm_{t\bar t}$, dilepton, 13 TeV	5	*2.760*	2.550	2.485	2.246	2.232	2.194
CMS $\sigma(t\bar t)$, ℓ+jets, 13 TeV	1	*0.230*	0.227	0.234	2.153	2.155	1.907
CMS $1/\sigma d\sigma/dm_{t\bar t}$, ℓ+jets, 13 TeV	14	*1.829*	1.481	1.393	0.962	0.911	0.912
ATLAS charge asymmetry, 8 TeV	1	*0.679*	0.678	0.675	0.567	0.587	0.571
ATLAS charge asymmetry, 13 TeV	5	*1.012*	0.997	0.989	0.872	0.786	0.836
CMS charge asymmetry, 8 TeV	3	*0.052*	0.052	0.053	0.072	0.069	0.071
CMS charge asymmetry, 13 TeV	3	*0.277*	0.283	0.286	0.443	0.510	0.451
ATLAS & CMS combined charge asy., 8 TeV	6	*0.603*	0.602	0.602	0.646	0.659	0.651
ATLAS W-hel., 13 TeV	2	*0.370*	*0.370*	*0.370*	0.037	0.037	0.038
ATLAS & CMS combined W-hel., 8 TeV	2	*1.046*	*1.046*	*1.046*	0.853	0.853	0.854
Total inclusive $t\bar t$	**108**	**1.700**	**1.328**	**1.271**	**1.032**	**0.994**	**1.002**
ATLAS $\sigma(t\bar tZ)$, 8 TeV	1	*0.264*	*0.232*	0.235	1.331	1.161	1.258
ATLAS $\sigma(t\bar tW)$, 8 TeV	1	*2.461*	*2.482*	2.430	0.751	0.708	0.722
ATLAS $\sigma(t\bar tZ)$, 13 TeV	1	*0.702*	*0.747*	0.742	0.001	0.000	0.000
ATLAS $1/\sigma d\sigma(t\bar tZ)/dp_T^Z$, 13 TeV	5	*1.961*	*1.940*	1.933	1.870	1.840	1.860
ATLAS $\sigma(t\bar tW)$, 13 TeV	1	*1.436*	*1.456*	1.417	0.000	0.002	0.000
ATLAS $\sigma(t\bar t\gamma)$, 8 TeV	1	*0.426*	*0.426*	*0.426*	0.037	0.009	0.034
ATLAS $\sigma(t\bar tt\bar t)$, multi-lepton, 13 TeV	1	*3.655*	*3.655*	*3.655*	4.289	3.784	4.145
ATLAS $\sigma(t\bar tt\bar t)$, single lepton, 13 TeV	1	*0.872*	*0.872*	*0.872*	0.997	0.898	0.969
ATLAS $\sigma(t\bar tb\bar b)$, ℓ+jets, 13 TeV	1	*2.062*	*2.062*	*2.062*	1.093	1.316	1.152
CMS $\sigma(t\bar tZ)$, 8 TeV	1	*0.432*	*0.470*	0.466	0.005	0.020	0.010
CMS $\sigma(t\bar tW)$, 8 TeV	1	*2.281*	*2.298*	2.255	0.808	0.770	0.783
CMS $\sigma(t\bar tZ)$, 13 TeV	1	*1.185*	*1.238*	1.233	0.085	0.113	0.110
CMS $1/\sigma d\sigma(t\bar tZ)/dp_T^Z$, 13 TeV	3	*0.628*	*0.598*	0.590	0.936	0.756	0.898
CMS $\sigma(t\bar tW)$, 13 TeV	1	*0.667*	*0.682*	0.652	0.297	0.329	0.307
CMS $\sigma(t\bar t\gamma)$, 8 TeV	1	*0.508*	*0.508*	*0.508*	0.022	0.045	0.024
CMS $\sigma(t\bar tt\bar t)$, multi-lepton, 13 TeV	1	*0.027*	*0.027*	*0.027*	0.124	0.041	0.096
CMS $\sigma(t\bar tt\bar t)$, single lepton, 13 TeV	1	*0.231*	*0.231*	*0.231*	0.262	0.238	0.255
CMS $\sigma(t\bar tb\bar b)$, all-jet, 13 TeV	1	*1.878*	*1.878*	*1.878*	1.334	1.466	1.369
CMS $\sigma(t\bar tb\bar b)$, dilepton, 13 TeV	1	*0.962*	*0.962*	*0.962*	0.312	0.447	0.347
CMS $\sigma(t\bar tb\bar b)$, ℓ+jets, 13 TeV	1	*0.900*	*0.900*	*0.900*	0.135	0.269	0.167
Total associated $t\bar t$	**26**	**1.255**	**1.255**	**1.246**	**0.925**	**0.888**	**0.913**

Table 16. Same as table 15 now for the inclusive and associated $t\bar t$ production datasets. Values in italics indicate datasets that do not enter the corresponding SM-PDF fit: for those processes, we evaluate a posteriori the physical observables and the associated χ^2 values using the resulting PDFs from the fit. For instance, top data is removed from the `NNPDF4.0-notop`, and for some top quark observables like $t\bar tt\bar t$ their PDF dependence is neglected. The last row indicates the total values adding together the non-top and top data contributions to the χ^2.

Dataset	n_{dat}	χ^2/n_{dat}					
		SM-PDF fits			**Fixed-PDF EFT fits**		**SMEFT-PDFs**
		NNPDF4.0 (no top)	NNPDF4.0	Fit H	NNPDF4.0 (no top)	Fit H	Joint fit
ATLAS t-channel $\sigma(t)$, 7 TeV	1	*0.785*	*0.756*	0.757	0.179	0.135	0.155
ATLAS t-channel $\sigma(\bar{t})$, 7 TeV	1	*0.304*	*0.296*	0.282	0.003	0.005	0.004
ATLAS t-channel $1/\sigma d\sigma(tq)/dy_t$, 7 TeV	3	*0.909*	0.959	0.971	0.885	0.946	0.900
ATLAS t-channel $1/\sigma d\sigma(\bar{t}q)/dy_{\bar{t}}$, 7 TeV	3	*0.061*	0.062	0.062	0.061	0.063	0.062
ATLAS t-channel $\sigma(t)$, 8 TeV	1	*0.795*	*0.739*	0.739	0.000	0.017	0.007
ATLAS t-channel $\sigma(\bar{t})$, 8 TeV	1	*2.314*	*2.277*	2.221	0.379	0.337	0.356
ATLAS t-channel $1/\sigma d\sigma(tq)/dy_t$, 8 TeV	3	*0.288*	*0.246*	0.241	0.290	0.245	0.283
ATLAS t-channel $1/\sigma d\sigma(\bar{t}q)/dy_{\bar{t}}$, 8 TeV	3	*0.196*	0.190	0.187	0.198	0.189	0.192
ATLAS s-channel $\sigma(t+\bar{t})$, 8 TeV	1	*0.211*	*0.203*	0.216	0.000	0.000	0.000
ATLAS t-channel $\sigma(t)$, 13 TeV	1	*0.738*	*0.722*	0.720	0.261	0.224	0.235
ATLAS t-channel $\sigma(\bar{t})$, 13 TeV	1	*0.400*	*0.393*	0.384	0.057	0.050	0.051
ATLAS s-channel $\sigma(t+\bar{t})$, 13 TeV	1	*0.700*	*0.688*	0.703	0.127	0.138	0.118
CMS t-channel $\sigma(t)+\sigma(\bar{t})$, 7 TeV	1	*0.765*	0.729	0.719	0.010	0.001	0.004
CMS t-channel $\sigma(t)$, 8 TeV	1	*0.010*	*0.006*	0.005	0.487	0.628	0.566
CMS t-channel $\sigma(\bar{t})$, 8 TeV	1	*0.079*	*0.084*	0.092	0.886	0.934	0.911
CMS s-channel $\sigma(t+\bar{t})$, 8 TeV	1	*1.108*	*1.112*	1.105	1.348	1.338	1.350
CMS t-channel $\sigma(t)$, 13 TeV	1	*0.904*	0.883	0.880	0.281	0.235	0.248
CMS t-channel $\sigma(\bar{t})$, 13 TeV	1	*0.136*	0.132	0.125	0.010	0.014	0.013
CMS t-channel $1/\sigma d\sigma/d|y^{(t)}|$, 13 TeV	4	*0.412*	0.387	0.383	0.405	0.378	0.396
Total single top	**30**	**0.509**	**0.498**	**0.495**	**0.332**	**0.330**	**0.331**
ATLAS $\sigma(tW)$, dilepton, 8 TeV	1	*0.510*	*0.529*	0.538	0.144	0.168	0.167
ATLAS $\sigma(tW)$, single-lepton, 8 TeV	1	*0.695*	*0.708*	0.713	0.428	0.449	0.448
ATLAS $\sigma(tW)$, dilepton, 13 TeV	1	*1.150*	*1.149*	1.148	0.840	0.847	0.849
ATLAS $\sigma_{\text{fid}}(tZj)$, dilepton, 13 TeV	1	*0.115*	*0.115*	*0.115*	0.871	0.611	0.795
CMS $\sigma(tW)$, dilepton, 8 TeV	1	*0.359*	*0.371*	0.376	0.134	0.150	0.149
CMS $\sigma(tW)$, dilepton, 13 TeV	1	*0.434*	*0.437*	0.438	1.612	1.573	1.563
CMS $\sigma_{\text{fid}}(tZj)$, dilepton, 13 TeV	1	*1.041*	*1.041*	*1.041*	0.279	0.427	0.316
CMS $d\sigma_{\text{fid}}(tZj)/dp_T^t$, dilepton, 13 TeV	3	*0.362*	*0.362*	*0.362*	0.130	0.156	0.134
CMS $\sigma(tW)$, single-lepton, 13 TeV	1	*2.837*	*2.833*	2.832	1.817	1.839	1.846
Total associated single top	**11**	**0.748**	**0.752**	**0.753**	**0.592**	**0.594**	**0.594**
Total top	**175**	**1.370**	**1.139**	**1.102**	**0.868**	**0.839**	**0.848**
Total dataset	**4710**	**1.156**	**1.145**	**1.144**	**1.138**	**1.135**	**1.126**

Table 17. Same as table 16 now for the inclusive and associated single-top production datasets. Values in italics indicate datasets that do not enter the corresponding fits: for those processes, we evaluate a posteriori the physical observables and the associated χ^2 values using the resulting PDFs from the fit. The last row indicates the total values adding together the non-top and top data contributions to the χ^2, and the next-to-last row the total top quark contribution.

the reported independence of the EFT interpretation of top quark data with respect to the choice of input PDF in the calculation.

Finally, we note the only class of top quark production process for which the χ^2 values are markedly different in the SM-PDF fits is inclusive top quark pair production. Indeed, the χ^2 values for the complete top quark dataset are $1.370, 1.139$, and 1.102 in the SM-PDF fits without top data, with NNPDF4.0 , and with the full top quark dataset (Fit H) respectively; this improvement arises mostly from the $t\bar{t}$ group, whose χ^2 values are $1.700, 1.328$, and 1.271 for the same three fits respectively. This observation is consistent with the findings of section 4, that the dominant PDF sensitivity in Fit H arises from inclusive $t\bar{t}$ production. We also note that the χ^2 for top quark data is similar between NNPDF4.0 and Fit H, again in agreement with the fact that the additional top quark measurements considered here as compared to those in NNPDF4.0 have a consistent pull on the PDFs, in particular on the large-x gluon.

E Pitfalls of the Monte-Carlo replica method for quadratic EFT fits

This analysis was originally prepared with the intention of performing a fully simultaneous SMEFT-PDF fit using NLO QCD theory, including *quadratic*, $\mathcal{O}\left(\Lambda^{-4}\right)$, contributions from the SMEFT. To this end, the capabilities of the `SIMUnet` framework were extended, as discussed in section 3.2, and all necessary SMEFT K-factors needed for the quadratic predictions were produced. However, whilst benchmarking our code, we noticed significant disagreement between quadratic SMEFT-only fits produced using the `SIMUnet` methodology and the SMEFiT Nested Sampling option;[5] on the other hand, we note perfect agreement between our `SIMUnet` quadratic SMEFT-only fits and the SMEFiT MCfit option.

This disagreement can be traced back to a deficiency in the Monte-Carlo sampling method used to propagate experimental error to the SMEFT coefficients, which currently prevents us from applying the `SIMUnet` framework to joint SMEFT-PDF fits with quadratic EFT calculations. In this appendix, we describe our current understanding of these limitations within the context of a toy model, and give a more realistic example from this work; further work on the topic is deferred to a future publication...

E.1 A toy model for quadratic EFT fits

In the following subsection, we consider a toy scenario involving a single data point d and only one Wilson coefficient c (we ignore all PDF-dependence). We suppose that our observed experimental data point d is a random variable drawn from a normal distribution centred on the underlying quadratic theory prediction, and with experimental variance σ^2:

$$d \sim N(t(c), \sigma^2). \tag{E.1}$$

We assume that the theory prediction is quadratic in the SMEFT Wilson coefficient c, taking the form:

$$t(c) = t^{\mathrm{SM}} + ct^{\mathrm{lin}} + c^2 t^{\mathrm{quad}}, \tag{E.2}$$

where we set $\Lambda = 1\,\mathrm{TeV}$ for convenience. Recall that $t^{\mathrm{quad}} > 0$, since it corresponds to a squared amplitude. Given the observed data d, we would like to construct interval estimates for the parameter c (usually *confidence intervals* in a frequentist setting or *credible intervals* in a Bayesian setting). Here, we shall describe the analytical construction of two interval estimates: first, using the Bayesian method, and second, using the Monte-Carlo replica method.

Bayesian method. In the Bayesian approach, c is treated as a random variable with its own distribution. By Bayes' theorem, we can write the probability distribution of c, given the observed data d, up to a proportionality constant (given by $1/\mathbb{P}(d)$, where $\mathbb{P}(d)$ is called the *Bayes' evidence*) as:

$$\mathbb{P}(c|d) \propto \mathbb{P}(d|c)\mathbb{P}(c), \tag{E.3}$$

---

[5]This disagreement is not present at the linear level; in this case, `SIMUnet` using the fixed-PDF option and SMEFiT using either the Nested Sampling or MCfit options perfectly coincide. See appendix C for more details.

where $\mathbb{P}(c|d)$ is called the *posterior distribution* of c, given the observed data d, and $\mathbb{P}(c)$ is called the *prior distribution* of c - this is our initial 'best guess' of the distribution of c before the observation of the data takes place. This distribution is often taken to be uniform in SMEFT fits; we shall assume this here.

Given a value of c, the distribution $\mathbb{P}(d|c)$ of the data d is assumed to be Gaussian, as specified in eq. (E.1). In particular, we can deduce that the posterior distribution of c obeys the following proportionality relation:[6]

$$\mathbb{P}(c|d) \propto \exp\left(-\frac{1}{2\sigma^2}(d-t(c))^2\right). \tag{E.4}$$

This posterior distribution can be used to place interval estimates on the parameter c. One way of doing this is to construct *highest density intervals*. These are computed as follows. For a $100\alpha\%$ credible interval, we determine the constant $p(\alpha)$ satisfying:

$$\int\limits_{\{c:\ \mathbb{P}(c|d)>p(\alpha)\}} \mathbb{P}(c|d)\ dc = \alpha. \tag{E.5}$$

An interval estimate for c is then given by $\{c : \mathbb{P}(c|d) > p(\alpha)\}$. In order to obtain such intervals then, we must construct the posterior $\mathbb{P}(c|d)$; efficient sampling from the posterior is guaranteed by methods such as Nested Sampling [200, 201].

The Monte-Carlo replica method. This method takes a different approach in order to produce a posterior distribution for the parameter c. Given the observed central data value d, one samples repeatedly from the normal distribution $N(d, \sigma^2)$ to generate a collection of *pseudodata replicas*, which we shall denote as $d^{(1)}, \ldots, d^{(N_{\text{rep}})}$, where N_{rep} is the total number of replicas. Given a pseudodata replica $d^{(i)}$, one obtains a corresponding best-fit value of the Wilson coefficient parameter $c^{(i)}$ by minimising the χ^2 of the theory to the pseudodata:

$$c^{(i)} = \arg\min_c \chi^2(c, d^{(i)}) = \arg\min_c \left(\frac{(d^{(i)} - t(c))^2}{\sigma^2}\right). \tag{E.6}$$

In this toy scenario, we can determine an analytical formula for $c^{(i)}$:

$$c^{(i)} = \begin{cases} -\dfrac{t^{\text{lin}}}{2t^{\text{quad}}}, & \text{if } d^{(i)} \leq \left(t^{\text{SM}} - (t^{\text{lin}})^2/4t^{\text{quad}}\right) ; \\ \dfrac{-t^{\text{lin}} \pm \sqrt{(t^{\text{lin}})^2 - 4t^{\text{quad}}(t^{\text{SM}} - d^{(i)})}}{2t^{\text{quad}}}, & \text{if } d^{(i)} \geq \left(t^{\text{SM}} - (t^{\text{lin}})^2/4t^{\text{quad}}\right) . \end{cases} \tag{E.7}$$

The first case arises when the χ^2 to the pseudodata $d^{(i)}$ has a single minimum, whilst the second case arises when the χ^2 has two minima. The two cases are split based on the value of

$$t_{\min} = t^{\text{SM}} - (t^{\text{lin}})^2/4t^{\text{quad}}, \tag{E.8}$$

which is the minimum value of the quadratic theory prediction $t(c) = t^{\text{SM}} + ct^{\text{lin}} + c^2 t^{\text{quad}}$. Note that for data replicas such that $d^{(i)} \leq t_{\min}$, the best-fit value $c^{(i)}$ becomes independent of $d^{(i)}$ and depends only on the ratio between linear and quadratic EFT cross-sections.

---

[6]Technically, truncated according to the end-points of the uniform prior.

Now, given that $d^{(i)}$ is a random variable drawn from the normal distribution $N(d, \sigma^2)$, one can infer the corresponding distribution of the random variable $c^{(i)}$, which is a function of the pseudodata $d^{(i)}$. For a real random variable X with associated probability density $P_X(x)$, a function $f : \mathbb{R} \to \mathbb{R}$ of the random variable has the distribution:

$$P_{f(X)}(y) = \int\limits_{-\infty}^{\infty} dx \; P_X(x) \delta(y - f(x)). \tag{E.9}$$

In our case, $c^{(i)}$ is a multi-valued function of $d^{(i)}$ given the two square roots, but the formula is easily generalised to this case. Recalling that the pseudodata replicas are generated according to a Gaussian distribution around the central measurement d with variance σ^2, we find that the probability density function for the Wilson coefficient replica $c^{(i)}$ is given (up to a proportionality constant) by:

$$P_{c^{(i)}}(c) \propto \int\limits_{-\infty}^{t_{\min}} dx \; \delta \left(c + \frac{t^{\text{lin}}}{2t^{\text{quad}}} \right) \exp \left(-\frac{1}{2\sigma^2}(x-d)^2 \right) \tag{E.10}$$

$$+ \int\limits_{t_{\min}}^{\infty} dx \; \delta \left(c - \left(\frac{-t^{\text{lin}} + \sqrt{(t^{\text{lin}})^2 - 4t^{\text{quad}}(t^{\text{SM}} - x)}}{2t^{\text{quad}}} \right) \right) \exp \left(-\frac{1}{2\sigma^2}(x-d)^2 \right)$$

$$+ \int\limits_{t_{\min}}^{\infty} dx \; \delta \left(c - \left(\frac{-t^{\text{lin}} - \sqrt{(t^{\text{lin}})^2 - 4t^{\text{quad}}(t^{\text{SM}} - x)}}{2t^{\text{quad}}} \right) \right) \exp \left(-\frac{1}{2\sigma^2}(x-d)^2 \right).$$

Simplifying the delta functions in the second and third integrals, we find:

$$P_{c^{(i)}}(c) \propto \delta \left(c + \frac{t^{\text{lin}}}{2t^{\text{quad}}} \right) \int\limits_{-\infty}^{t_{\min}} dx \; \exp \left(-\frac{1}{2\sigma^2}(x-d)^2 \right) + \frac{2}{|2ct^{\text{quad}} + t^{\text{lin}}|} \exp \left(-\frac{1}{2\sigma^2}(d - t(c))^2 \right). \tag{E.11}$$

This result is different from the posterior distribution obtained by the Bayesian method in eq. (E.4). Notable features of the posterior distribution $P_{c^{(i)}}(c)$ are: (i) the distribution has a Dirac-delta peak at $c = -t^{\text{lin}}/2t^{\text{quad}}$; (ii) elsewhere, the distribution is given by the Bayesian posterior distribution rescaled by a prefactor dependent on c. Therefore, the Monte Carlo replica optimisation method will not in general reproduce the Bayesian posterior.

However, one can note that in an appropriate limit, the Bayesian posterior *is* indeed recovered. In particular, suppose that the quadratic EFT cross-section is subdominant compared to the linear term, $t^{\text{lin}} \gg t^{\text{quad}}$; in this case, we have that $t_{\min} \to -\infty$ so that the first term in eq. (E.11) vanishes and the prefactor of the second term can be approximated with $2ct^{\text{quad}} + t^{\text{lin}} \approx t^{\text{lin}}$. Thus, the Bayesian posterior from eq. (E.4) is indeed recovered, and we see that the two methods are formally identical for a linear EFT analysis. Further, it is possible to show analytically that for multiple SMEFT parameters and multiple correlated data points, if only linear theory is used the two distributions agree exactly.

This calculation demonstrates that, for quadratic EFT fits, the Monte-Carlo replica method will not in general reproduce the Bayesian posteriors that one would obtain from, say, a nested sampling approach; agreement will only occur provided quadratic EFT corrections are sufficiently subdominant in comparison with the linear ones. For this reason, in this work we restrict the SMEFT-PDF fits based on `SIMUnet` (which rely on the use of the Monte-Carlo replica method) to linear EFT calculations; we defer the further investigation of the use of the Monte-Carlo replica method, and how it might be modified for use in `SIMUnet`, to future works.

E.2 Application to one-parameter fits

As demonstrated above, the Monte-Carlo replica method will lead to posterior distributions differing from their Bayesian counterparts whenever quadratic EFT corrections dominate over linear ones. Here, we show the numerical impact of these differences in a model case, namely the one-parameter fit of the coefficient c_{dt}^8 to the CMS 13 TeV $t\bar{t}$ invariant mass distribution measurement based on the ℓ+jets final-state [117]. Figure 21 compares the experimental data from this measurement with the corresponding SM theory calculations at NNLO using the NNPDF4.0 (no top) PDF set as input. We observe that the SM theory predictions overshoot the data, especially in the high $m_{t\bar{t}}$ regions, where energy-growing effects enhance the EFT corrections. Given that the pseudodata replicas $d^{(i)}$ are fluctuated around the central value d, the configuration where the SM overshoots the data potentially enhances the contribution of the upper solution in eq. (E.7) leading to the Dirac delta peak in the posterior eq. (E.11).

For the case of the c_{dt}^8 coefficient, the quadratic EFT corrections dominate over the linear ones and hence the net effect of a non-zero coefficient is typically an upwards shift of the theory prediction. Indeed, we have verified that for this coefficient the biggest negative correction one can obtain is of order $\sim 2\%$. For the last $m_{t\bar{t}}$ bin, the minimum of the theory cross section $t_{\min}$ in eq. (E.8) is obtained for a value $c_{dt}^8 \approx -0.2$, while for the second to last bin instead $t_{\min}$ is minimised by $c_{dt}^8 \approx -0.3$. The combination of these two features (a dominant quadratic EFT term, and a SM prediction overshooting the data) suggests that the Monte-Carlo replica method's posterior will be enhanced for $c_{dt}^8 \in (-0.3, -0.2)$ as compared to the Bayesian posterior.

In figure 22 we compare the posterior distributions for a one-parameter fit of the four-fermion coefficient c_{dt}^8 with the sole experimental input being the CMS $m_{t\bar{t}}$ distribution displayed in figure 21. Results are obtained with SMEFiT and we compare the outcome of Nested Sampling (in blue) with that of MCfit (in red) for linear and quadratic EFT fits. The agreement in the linear fit is lost for its quadratic counterpart, with the main difference being a sharp peak in the region $c_{dt}^8 \in (-0.3, -0.2)$ in which the contribution from the delta function in eq. E.11 is most marked.

The scenario displayed in figure 22 is chosen to display the maximum effect, based on a single coefficient with large quadratic EFT corrections, and a dataset where the SM overshoots the data in the $m_{t\bar{t}}$ region where EFT effects are the largest. Within a global fit, these differences are *partially* washed out (indeed the Bayesian and MCfit posterior distributions mostly agree well for the quadratic SMEFiT analysis, as shown in [182],
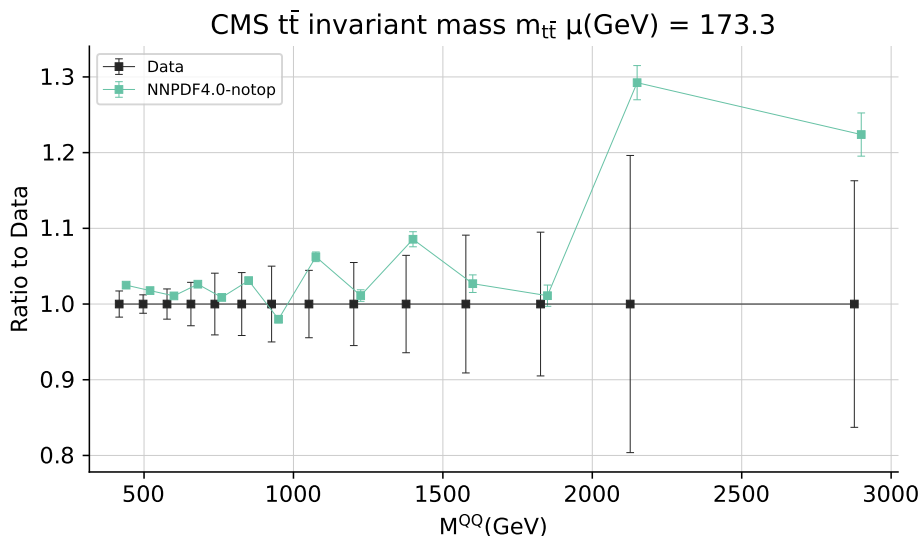
Figure 21. Comparison between the experimental data for the top-quark pair invariant mass $m_{t\bar{t}}$ distribution from the ℓ+jets CMS measurement at 13 TeV [117], with the corresponding SM theory calculations at NNLO using the NNPDF4.0 (no top) PDF set as input. For the latter, the error band indicates the PDF uncertainties, and for the former the diagonal entries of the experimental covariance matrix. Results are shown as ratios to the central value of the data. The SM theory predictions overshoot the data, especially in the high $m_{t\bar{t}}$ regions, where energy-growing effects enhance the EFT corrections.
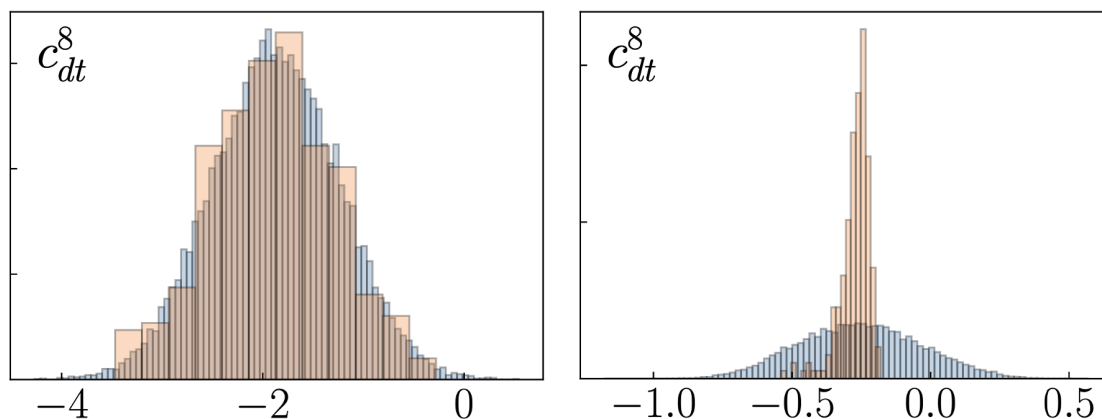
Figure 22. Posterior distributions for a one-parameter fit of the four-fermion coefficient c_{dt}^8 with the sole experimental input being the CMS $m_{t\bar{t}}$ distribution displayed in figure 21. Results are obtained with SMEFiT and we compare the outcome of Nested Sampling (in blue) with that of MCfit (in red) for linear (left panel) and quadratic (right panel) EFT fits.

for the majority of fitted coefficients). Nevertheless, at least in its current implementation, figure 22 highlights that the Monte-Carlo replica method is affected by pitfalls that prevent its straightforward application to global EFT interpretations of experimental data which include quadratic corrections.

Open Access. This article is distributed under the terms of the Creative Commons Attribution License ([CC-BY 4.0](#)), which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited. SCOAP^3 supports the goals of the International Year of Basic Sciences for Sustainable Development.

References

[1] G. Isidori, G. Ridolfi and A. Strumia, *On the metastability of the standard model vacuum*, *Nucl. Phys. B* **609** (2001) 387 [`hep-ph/0104016`] [INSPIRE].

[2] D. Buttazzo et al., *Investigating the near-criticality of the Higgs boson*, *JHEP* **12** (2013) 089 [`arXiv:1307.3536`] [INSPIRE].

[3] L. Di Luzio, G. Isidori and G. Ridolfi, *Stability of the electroweak ground state in the standard model and its extensions*, *Phys. Lett. B* **753** (2016) 150 [`arXiv:1509.05028`] [INSPIRE].

[4] CDF collaboration, *Observation of top quark production in $\bar{p}p$ collisions*, *Phys. Rev. Lett.* **74** (1995) 2626 [`hep-ex/9503002`] [INSPIRE].

[5] D0 collaboration, *Observation of the top quark*, *Phys. Rev. Lett.* **74** (1995) 2632 [`hep-ex/9503003`] [INSPIRE].

[6] CMS collaboration, *Determination of the top-quark pole mass and strong coupling constant from the $t\bar{t}$ production cross section in pp collisions at $\sqrt{s} = 7$ TeV*, *Phys. Lett. B* **728** (2014) 496 [*Erratum ibid.* **738** (2014) 526] [`arXiv:1307.1907`] [INSPIRE].

[7] A.M. Cooper-Sarkar et al., *Simultaneous extraction of α_s and m_t from LHC $t\bar{t}$ differential distributions*, `arXiv:2010.04171` [INSPIRE].

[8] CMS collaboration, *Measurement of CKM matrix elements in single top quark t-channel production in proton-proton collisions at $\sqrt{s} = 13$ TeV*, *Phys. Lett. B* **808** (2020) 135609 [`arXiv:2004.12181`] [INSPIRE].

[9] ATLAS collaboration, *Measurement of the top-quark mass using a leptonic invariant mass in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, `arXiv:2209.00583` [INSPIRE].

[10] CMS collaboration, *Measurement of the top quark mass using a profile likelihood approach with the lepton+jets final states in proton-proton collisions at $\sqrt{s} = 13$ TeV*, `arXiv:2302.01967` [INSPIRE].

[11] 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].

[12] S. Amoroso et al., *Snowmass* 2021 *whitepaper: proton structure at the precision frontier*, *Acta Phys. Polon. B* **53** (2022) A1 [`arXiv:2203.13923`] [INSPIRE].

[13] M. Czakon, M.L. Mangano, A. Mitov and J. Rojo, *Constraints on the gluon PDF from top quark pair production at hadron colliders*, *JHEP* **07** (2013) 167 [`arXiv:1303.7215`] [INSPIRE].

[14] M. Czakon et al., *Pinning down the large-x gluon with NNLO top-quark pair differential distributions*, *JHEP* **04** (2017) 044 [arXiv:1611.08609] [INSPIRE].

[15] M. Czakon et al., *An exploratory study of the impact of CMS double-differential top distributions on the gluon parton distribution function*, *J. Phys. G* **48** (2020) 015003 [arXiv:1912.08801] [INSPIRE].

[16] E.R. Nocera, M. Ubiali and C. Voisey, *Single top production in PDF fits*, *JHEP* **05** (2020) 067 [arXiv:1912.09543] [INSPIRE].

[17] J. Campbell, T. Neumann and Z. Sullivan, *Testing parton distribution functions with t-channel single-top-quark production*, *Phys. Rev. D* **104** (2021) 094042 [arXiv:2109.10448] [INSPIRE].

[18] M. Czakon, D. Heymes and A. Mitov, *Dynamical scales for multi-TeV top-pair production at the LHC*, *JHEP* **04** (2017) 071 [arXiv:1606.03350] [INSPIRE].

[19] M. Czakon, A. Mitov and R. Poncelet, *NNLO QCD corrections to leptonic observables in top-quark pair production and decay*, *JHEP* **05** (2021) 212 [arXiv:2008.11133] [INSPIRE].

[20] S. Catani et al., *Top-quark pair hadroproduction at next-to-next-to-leading order in QCD*, *Phys. Rev. D* **99** (2019) 051501 [arXiv:1901.04005] [INSPIRE].

[21] S. Catani et al., *Top-quark pair production at the LHC: fully differential QCD predictions at NNLO*, *JHEP* **07** (2019) 100 [arXiv:1906.06535] [INSPIRE].

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

[23] M. Czakon et al., *Top-pair production at the LHC through NNLO QCD and NLO EW*, *JHEP* **10** (2017) 186 [arXiv:1705.04105] [INSPIRE].

[24] M. Czakon et al., *Resummation for (boosted) top-quark pair production at NNLO+NNLL' in QCD*, *JHEP* **05** (2018) 149 [arXiv:1803.07623] [INSPIRE].

[25] J. Mazzitelli et al., *Top-pair production at the LHC with MINNLO_PS*, *JHEP* **04** (2022) 079 [arXiv:2112.12135] [INSPIRE].

[26] E.L. Berger, J. Gao and H.X. Zhu, *Differential distributions for t-channel single top-quark production and decay at next-to-next-to-leading order in QCD*, *JHEP* **11** (2017) 158 [arXiv:1708.09405] [INSPIRE].

[27] E.L. Berger, J. Gao, C.-P. Yuan and H.X. Zhu, *NNLO QCD corrections to t-channel single top-quark production and decay*, *Phys. Rev. D* **94** (2016) 071501 [arXiv:1606.08463] [INSPIRE].

[28] Z.L. Liu and J. Gao, *s-channel single top quark production and decay at next-to-next-to-leading-order in QCD*, *Phys. Rev. D* **98** (2018) 071501 [arXiv:1807.03835] [INSPIRE].

[29] D. Barducci et al., *Interpreting top-quark LHC measurements in the standard-model effective field theory*, arXiv:1802.07237 [INSPIRE].

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

[31] 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].

[32] S. Bißmann et al., *Constraining top-quark couplings combining top-quark and B decay observables*, *Eur. Phys. J. C* **80** (2020) 136 [arXiv:1909.13632] [INSPIRE].

[33] 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].

[34] G. Durieux et al., *The electro-weak couplings of the top and bottom quarks — global fit and future prospects*, *JHEP* **12** (2019) 98 [*Erratum ibid.* **01** (2021) 195] [arXiv:1907.10619] [INSPIRE].

[35] 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].

[36] B.R. Yates, *Using associated top quark production to probe for new physics within the framework of effective field theory*, in the proceedings of the 13th *international workshop on top quark physics*, (2021) [arXiv:2101.10828] [INSPIRE].

[37] R. Aoude, E. Madge, F. Maltoni and L. Mantani, *Quantum SMEFT tomography: top quark pair production at the LHC*, *Phys. Rev. D* **106** (2022) 055007 [arXiv:2203.05619] [INSPIRE].

[38] F. Maltoni, L. Mantani and K. Mimasu, *Top-quark electroweak interactions at high energy*, *JHEP* **10** (2019) 004 [arXiv:1904.05637] [INSPIRE].

[39] R. Aoude, H. El Faham, F. Maltoni and E. Vryonidou, *Complete SMEFT predictions for four top quark production at hadron colliders*, *JHEP* **10** (2022) 163 [arXiv:2208.04962] [INSPIRE].

[40] C. Degrande et al., *Single-top associated production with a Z or H boson at the LHC: the SMEFT interpretation*, *JHEP* **10** (2018) 005 [arXiv:1804.07773] [INSPIRE].

[41] H.E. Faham, F. Maltoni, K. Mimasu and M. Zaro, *Single top production in association with a WZ pair at the LHC in the SMEFT*, *JHEP* **01** (2022) 100 [arXiv:2111.03080] [INSPIRE].

[42] C. Severi and E. Vryonidou, *Quantum entanglement and top spin correlations in SMEFT at higher orders*, *JHEP* **01** (2023) 148 [arXiv:2210.09330] [INSPIRE].

[43] Q.-H. Cao, B. Yan, C.P. Yuan and Y. Zhang, *Probing $Zt\bar{t}$ couplings using Z boson polarization in ZZ production at hadron colliders*, *Phys. Rev. D* **102** (2020) 055010 [arXiv:2004.02031] [INSPIRE].

[44] Q.-H. Cao, B. Yan, J.-H. Yu and C. Zhang, *A general analysis of Wtb anomalous couplings*, *Chin. Phys. C* **41** (2017) 063101 [arXiv:1504.03785] [INSPIRE].

[45] S. Carrazza et al., *Can new physics hide inside the proton?*, *Phys. Rev. Lett.* **123** (2019) 132001 [arXiv:1905.05215] [INSPIRE].

[46] A. Greljo et al., *Parton distributions in the SMEFT from high-energy Drell-Yan tails*, *JHEP* **07** (2021) 122 [arXiv:2104.02723] [INSPIRE].

[47] D.Y. Liu, C.L. Sun and J. Gao, *Machine learning of log-likelihood functions in global analysis of parton distributions*, *JHEP* **08** (2022) 088 [arXiv:2201.06586] [INSPIRE].

[48] J. Gao et al., *Simultaneous CTEQ-TEA extraction of PDFs and SMEFT parameters from jet and $t\bar{t}$ data*, *JHEP* **05** (2023) 003 [arXiv:2211.01094] [INSPIRE].

[49] CMS collaboration, *Measurement and QCD analysis of double-differential inclusive jet cross sections in proton-proton collisions at $\sqrt{s} = 13$ TeV*, *JHEP* **02** (2022) 142 [*Addendum ibid.* **12** (2022) 035] [arXiv:2111.10431] [INSPIRE].

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

[51] NNPDF collaboration, *The path to proton structure at* 1% *accuracy*, *Eur. Phys. J. C* **82** (2022) 428 [arXiv:2109.02653] [InSPIRE].

[52] NNPDF collaboration, *An open-source machine learning framework for global analyses of parton distributions*, *Eur. Phys. J. C* **81** (2021) 958 [arXiv:2109.02671] [InSPIRE].

[53] SMEFiT collaboration, *Combined SMEFT interpretation of Higgs, diboson, and top quark data from the LHC*, *JHEP* **11** (2021) 089 [arXiv:2105.00006] [InSPIRE].

[54] New Muon collaboration, *Accurate measurement of F_2^d/F_2^p and $R_d - R_p$*, *Nucl. Phys. B* **487** (1997) 3 [hep-ex/9611022] [InSPIRE].

[55] New Muon collaboration, *Measurement of the proton and deuteron structure functions, F_2^p and F_2^d, and of the ratio σ_L/σ_T*, *Nucl. Phys. B* **483** (1997) 3 [hep-ph/9610231] [InSPIRE].

[56] L.W. Whitlow et al., *Precise measurements of the proton and deuteron structure functions from a global analysis of the SLAC deep inelastic electron scattering cross-sections*, *Phys. Lett. B* **282** (1992) 475 [InSPIRE].

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

[58] CHORUS collaboration, *Measurement of nucleon structure functions in neutrino scattering*, *Phys. Lett. B* **632** (2006) 65 [InSPIRE].

[59] NuTeV collaboration, *Precise measurement of dimuon production cross-sections in ν_μ Fe and $\bar{\nu}_\mu$ Fe deep inelastic scattering at the Tevatron*, *Phys. Rev. D* **64** (2001) 112006 [hep-ex/0102049] [InSPIRE].

[60] D.A. Mason, *Measurement of the strange-antistrange asymmetry at NLO in QCD from NuTeV dimuon data*, Ph.D. thesis, Oregon U., Eugene, OR, U.S.A. (2006) [InSPIRE].

[61] 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].

[62] H1 and ZEUS collaborations, *Combination and QCD analysis of charm and beauty production cross-section measurements in deep inelastic ep scattering at HERA*, *Eur. Phys. J. C* **78** (2018) 473 [arXiv:1804.01019] [InSPIRE].

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

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

[65] NuSea collaboration, *Improved measurement of the anti-d/anti-u asymmetry in the nucleon sea*, *Phys. Rev. D* **64** (2001) 052002 [hep-ex/0103030] [InSPIRE].

[66] 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].

[67] 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].

[68] 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].

[69] 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].

[70] CDF collaboration, *Measurement of the inclusive jet cross section using the* k_{T} *algorithmin* $p\bar{p}$ *collisions at* $\sqrt{s} = 1.96$ *TeV with the CDF II detector*, *Phys. Rev. D* **75** (2007) 092006 [*Erratum ibid.* **75** (2007) 119901] [hep-ex/0701051] [InSPIRE].

[71] 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].

[72] 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].

[73] 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].

[74] 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].

[75] 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].

[76] CMS collaboration, *Measurement of the muon charge asymmetry in inclusive* $pp \to W + X$ *production at* $\sqrt{s} = 7$ *TeV and an improved determination of light parton distribution functions*, *Phys. Rev. D* **90** (2014) 032004 [arXiv:1312.6283] [InSPIRE].

[77] 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].

[78] 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].

[79] 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].

[80] 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].

[81] LHCb collaboration, *Measurement of forward* $Z \to e^+e^-$ *production at* $\sqrt{s} = 8$ *TeV*, *JHEP* **05** (2015) 109 [arXiv:1503.00963] [InSPIRE].

[82] 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].

[83] 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].

[84] ATLAS collaboration, *Measurement of the Drell-Yan triple-differential cross section in pp collisions at $\sqrt{s} = 8$ TeV*, *JHEP* **12** (2017) 059 [arXiv:1710.05167] [INSPIRE].

[85] ATLAS collaboration, *Measurement of the cross-section and charge asymmetry of W bosons produced in proton-proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, *Eur. Phys. J. C* **79** (2019) 760 [arXiv:1904.05631] [INSPIRE].

[86] LHCв collaboration, *Measurement of forward $W \to e\nu$ production in pp collisions at $\sqrt{s} = 8$ TeV*, *JHEP* **10** (2016) 030 [arXiv:1608.01484] [INSPIRE].

[87] ATLAS collaboration, *Measurement of $W^{\pm}$ and Z-boson production cross sections in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *Phys. Lett. B* **759** (2016) 601 [arXiv:1603.09222] [INSPIRE].

[88] LHCв collaboration, *Measurement of the forward Z boson production cross-section in pp collisions at $\sqrt{s} = 13$ TeV*, *JHEP* **09** (2016) 136 [arXiv:1607.06495] [INSPIRE].

[89] 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].

[90] 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].

[91] 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].

[92] CMS collaboration, *Measurement of the differential cross sections for the associated production of a W boson and jets in proton-proton collisions at $\sqrt{s} = 13$ TeV*, *Phys. Rev. D* **96** (2017) 072005 [arXiv:1707.05979] [INSPIRE].

[93] ATLAS collaboration, *Measurement of inclusive jet and dijet production in pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector*, *Phys. Rev. D* **86** (2012) 014022 [arXiv:1112.6297] [INSPIRE].

[94] ATLAS collaboration, *Measurement of the inclusive jet cross section in pp collisions at $\sqrt{s} = 2.76$ TeV and comparison to the inclusive jet cross section at $\sqrt{s} = 7$ TeV using the ATLAS detector*, *Eur. Phys. J. C* **73** (2013) 2509 [arXiv:1304.4739] [INSPIRE].

[95] ATLAS collaboration, *Measurement of the inclusive jet cross-section in proton-proton collisions at $\sqrt{s} = 7$ TeV using $4.5\,fb^{-1}$ of data with the ATLAS detector*, *JHEP* **02** (2015) 153 [*Erratum ibid.* **09** (2015) 141] [arXiv:1410.8857] [INSPIRE].

[96] CMS collaboration, *Measurements of differential jet cross sections in proton-proton collisions at $\sqrt{s} = 7$ TeV with the CMS detector*, *Phys. Rev. D* **87** (2013) 112002 [*Erratum ibid.* **87** (2013) 119902] [arXiv:1212.6660] [INSPIRE].

[97] CMS collaboration, *Measurement of the inclusive jet cross section in pp collisions at $\sqrt{s} = 2.76$ TeV*, *Eur. Phys. J. C* **76** (2016) 265 [arXiv:1512.06212] [INSPIRE].

[98] ATLAS collaboration, *Measurement of the inclusive jet cross-sections in proton-proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, *JHEP* **09** (2017) 020 [arXiv:1706.03192] [INSPIRE].

[99] CMS collaboration, *Measurement and QCD analysis of double-differential inclusive jet cross sections in pp collisions at $\sqrt{s} = 8$ TeV and cross section ratios to 2.76 and 7 TeV*, *JHEP* **03** (2017) 156 [arXiv:1609.05331] [INSPIRE].

[100] ATLAS collaboration, *Measurement of the inclusive isolated prompt photon cross section in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, *JHEP* **08** (2016) 005 [arXiv:1605.03495] [INSPIRE].

[101] ATLAS collaboration, *Measurement of the cross section for inclusive isolated-photon production in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector*, *Phys. Lett. B* **770** (2017) 473 [arXiv:1701.06882] [INSPIRE].

[102] J. Ellis et al., *Top, Higgs, diboson and electroweak fit to the standard model effective field theory*, *JHEP* **04** (2021) 279 [arXiv:2012.02779] [INSPIRE].

[103] ATLAS collaboration, *Measurements of top-quark pair differential cross-sections in the lepton+jets channel in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector*, *Eur. Phys. J. C* **76** (2016) 538 [arXiv:1511.04716] [INSPIRE].

[104] ATLAS collaboration, *Measurement of the $t\bar{t}$ production cross-section using $e\mu$ events with b-tagged jets in pp collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS detector*, *Eur. Phys. J. C* **74** (2014) 3109 [*Addendum ibid.* **76** (2016) 642] [arXiv:1406.5375] [INSPIRE].

[105] ATLAS collaboration, *Measurement of top quark pair differential cross-sections in the dilepton channel in pp collisions at $\sqrt{s} = 7$ and 8 TeV with ATLAS*, *Phys. Rev. D* **94** (2016) 092003 [*Addendum ibid.* **101** (2020) 119901] [arXiv:1607.07281] [INSPIRE].

[106] ATLAS collaboration, *Measurement of the inclusive and fiducial $t\bar{t}$ production cross-sections in the lepton+jets channel in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, *Eur. Phys. J. C* **78** (2018) 487 [arXiv:1712.06857] [INSPIRE].

[107] ATLAS collaboration, *Measurement of the $t\bar{t}$ production cross-section and lepton differential distributions in $e\mu$ dilepton events from pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *Eur. Phys. J. C* **80** (2020) 528 [arXiv:1910.08819] [INSPIRE].

[108] ATLAS collaboration, *Measurements of top-quark pair single- and double-differential cross-sections in the all-hadronic channel in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector*, *JHEP* **01** (2021) 033 [arXiv:2006.09274] [INSPIRE].

[109] ATLAS collaboration, *Measurement of the $t\bar{t}$ production cross-section in the lepton+jets channel at $\sqrt{s} = 13$ TeV with the ATLAS experiment*, *Phys. Lett. B* **810** (2020) 135797 [arXiv:2006.13076] [INSPIRE].

[110] ATLAS collaboration, *Measurements of top-quark pair differential and double-differential cross-sections in the ℓ+jets channel with pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector*, *Eur. Phys. J. C* **79** (2019) 1028 [*Erratum ibid.* **80** (2020) 1092] [arXiv:1908.07305] [INSPIRE].

[111] CMS collaboration, *Measurement of the inclusive $t\bar{t}$ cross section in pp collisions at $\sqrt{s} = 5.02$ TeV using final states with at least one charged lepton*, *JHEP* **03** (2018) 115 [arXiv:1711.03143] [INSPIRE].

[112] S. Spannagel, *Top quark mass measurements with the CMS experiment at the LHC*, [*PoS* **DIS2016** (2016) 150](arXiv:1607.04972) [INSPIRE].

[113] CMS collaboration, *Measurement of double-differential cross sections for top quark pair production in pp collisions at $\sqrt{s} = 8$ TeV and impact on parton distribution functions*, [*Eur. Phys. J. C* **77** (2017) 459](arXiv:1703.01630) [INSPIRE].

[114] CMS collaboration, *Measurement of the differential cross section for top quark pair production in pp collisions at $\sqrt{s} = 8$ TeV*, [*Eur. Phys. J. C* **75** (2015) 542](arXiv:1505.04480) [INSPIRE].

[115] CMS collaboration, *Measurement of the top quark pair production cross section in proton-proton collisions at $\sqrt{s} = 13$ TeV*, [*Phys. Rev. Lett.* **116** (2016) 052002](arXiv:1510.05302) [INSPIRE].

[116] CMS collaboration, *Measurements of $t\bar{t}$ differential cross sections in proton-proton collisions at $\sqrt{s} = 13$ TeV using events containing two leptons*, [*JHEP* **02** (2019) 149](arXiv:1811.06625) [INSPIRE].

[117] CMS collaboration, *Measurement of differential $t\bar{t}$ production cross sections in the full kinematic range using lepton+jets events from proton-proton collisions at $\sqrt{s} = 13$ TeV*, [*Phys. Rev. D* **104** (2021) 092013](arXiv:2108.02803) [INSPIRE].

[118] ATLAS collaboration, *Measurements of the charge asymmetry in top-quark pair production in the dilepton final state at $\sqrt{s} = 8$ TeV with the ATLAS detector*, [*Phys. Rev. D* **94** (2016) 032006](arXiv:1604.05538) [INSPIRE].

[119] ATLAS collaboration, *Evidence for the charge asymmetry in $pp \to t\bar{t}$ production at $\sqrt{s} = 13$ TeV with the ATLAS detector*, arXiv:2208.12095 [INSPIRE].

[120] CMS collaboration, *Measurements of $t\bar{t}$ charge asymmetry using dilepton final states in pp collisions at $\sqrt{s} = 8$ TeV*, [*Phys. Lett. B* **760** (2016) 365](arXiv:1603.06221) [INSPIRE].

[121] CMS collaboration, *Measurement of the $t\bar{t}$ charge asymmetry in highly boosted events in the single-lepton channel at 13 TeV*, [CMS-PAS-TOP-21-014](), CERN, Geneva, Switzerland (2022) [INSPIRE].

[122] ATLAS and CMS collaborations, *Combination of inclusive and differential $t\bar{t}$ charge asymmetry measurements using ATLAS and CMS data at $\sqrt{s} = 7$ and 8 TeV*, [*JHEP* **04** (2018) 033](arXiv:1709.05327) [INSPIRE].

[123] CMS and ATLAS collaborations, *Combination of the W boson polarization measurements in top quark decays using ATLAS and CMS data at $\sqrt{s} = 8$ TeV*, [*JHEP* **08** (2020) 051](arXiv:2005.03799) [INSPIRE].

[124] ATLAS collaboration, *Measurement of the polarisation of W bosons produced in top-quark decays using dilepton events at $\sqrt{s} = 13$ TeV with the ATLAS experiment*, arXiv:2209.14903 [DOI:10.48550/arXiv.2209.14903].

[125] ATLAS collaboration, *Measurement of the $t\bar{t}W$ and $t\bar{t}Z$ production cross sections in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, [*JHEP* **11** (2015) 172](arXiv:1509.05276) [INSPIRE].

[126] ATLAS collaboration, *Measurement of the $t\bar{t}Z$ and $t\bar{t}W$ cross sections in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, [*Phys. Rev. D* **99** (2019) 072009](arXiv:1901.03584) [INSPIRE].

[127] ATLAS collaboration, *Measurements of the inclusive and differential production cross sections of a top-quark-antiquark pair in association with a Z boson at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *Eur. Phys. J. C* **81** (2021) 737 [arXiv:2103.12603] [INSPIRE].

[128] CMS collaboration, *Observation of top quark pairs produced in association with a vector boson in pp collisions at $\sqrt{s} = 8$ TeV*, *JHEP* **01** (2016) 096 [arXiv:1510.01131] [INSPIRE].

[129] CMS collaboration, *Measurement of the cross section for top quark pair production in association with a W or Z boson in proton-proton collisions at $\sqrt{s} = 13$ TeV*, *JHEP* **08** (2018) 011 [arXiv:1711.02547] [INSPIRE].

[130] CMS collaboration, *Measurement of top quark pair production in association with a Z boson in proton-proton collisions at $\sqrt{s} = 13$ TeV*, *JHEP* **03** (2020) 056 [arXiv:1907.11270] [INSPIRE].

[131] ATLAS collaboration, *Measurement of the $t\bar{t}\gamma$ production cross section in proton-proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, *JHEP* **11** (2017) 086 [arXiv:1706.03046] [INSPIRE].

[132] CMS collaboration, *Measurement of the semileptonic $t\bar{t} + \gamma$ production cross section in pp collisions at $\sqrt{s} = 8$ TeV*, *JHEP* **10** (2017) 006 [arXiv:1706.08128] [INSPIRE].

[133] ATLAS collaboration, *Evidence for $t\bar{t}t\bar{t}$ production in the multilepton final state in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *Eur. Phys. J. C* **80** (2020) 1085 [arXiv:2007.14858] [INSPIRE].

[134] ATLAS collaboration, *Measurement of the $t\bar{t}t\bar{t}$ production cross section in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *JHEP* **11** (2021) 118 [arXiv:2106.11683] [INSPIRE].

[135] ATLAS collaboration, *Measurements of inclusive and differential fiducial cross-sections of $t\bar{t}$ production with additional heavy-flavour jets in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *JHEP* **04** (2019) 046 [arXiv:1811.12113] [INSPIRE].

[136] CMS collaboration, *Search for production of four top quarks in final states with same-sign or multiple leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV*, *Eur. Phys. J. C* **80** (2020) 75 [arXiv:1908.06463] [INSPIRE].

[137] CMS collaboration, *Search for the production of four top quarks in the single-lepton and opposite-sign dilepton final states in proton-proton collisions at $\sqrt{s} = 13$ TeV*, *JHEP* **11** (2019) 082 [arXiv:1906.02805] [INSPIRE].

[138] CMS collaboration, *Measurement of the $t\bar{t}b\bar{b}$ production cross section in the all-jet final state in pp collisions at $\sqrt{s} = 13$ TeV*, *Phys. Lett. B* **803** (2020) 135285 [arXiv:1909.05306] [INSPIRE].

[139] CMS collaboration, *Measurement of the cross section for $t\bar{t}$ production with additional jets and b jets in pp collisions at $\sqrt{s} = 13$ TeV*, *JHEP* **07** (2020) 125 [arXiv:2003.06467] [INSPIRE].

[140] ATLAS collaboration, *Measurements of inclusive and differential cross-sections of combined $t\bar{t}\gamma$ and $tW\gamma$ production in the $e\mu$ channel at 13 TeV with the ATLAS detector*, *JHEP* **09** (2020) 049 [arXiv:2007.06946] [INSPIRE].

[141] ATLAS collaboration, *Comprehensive measurements of t-channel single top-quark production cross sections at $\sqrt{s} = 7$ TeV with the ATLAS detector*, *Phys. Rev. D* **90** (2014) 112006 [arXiv:1406.7844] [INSPIRE].

[142] ATLAS collaboration, *Fiducial, total and differential cross-section measurements of t-channel single top-quark production in pp collisions at 8 TeV using data collected by the ATLAS detector*, *Eur. Phys. J. C* **77** (2017) 531 [arXiv:1702.02859] [INSPIRE].

[143] ATLAS collaboration, *Evidence for single top-quark production in the s-channel in proton-proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector using the matrix element method*, *Phys. Lett. B* **756** (2016) 228 [arXiv:1511.05980] [INSPIRE].

[144] ATLAS collaboration, *Measurement of the inclusive cross-sections of single top-quark and top-antiquark t-channel production in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *JHEP* **04** (2017) 086 [arXiv:1609.03920] [INSPIRE].

[145] ATLAS collaboration, *Measurement of single top-quark production in the s-channel in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, arXiv:2209.08990 [INSPIRE].

[146] CMS collaboration, *Measurement of the single-top-quark t-channel cross section in pp collisions at $\sqrt{s} = 7$ TeV*, *JHEP* **12** (2012) 035 [arXiv:1209.4533] [INSPIRE].

[147] CMS collaboration, *Measurement of the t-channel single-top-quark production cross section and of the $| V_{tb} |$ CKM matrix element in pp collisions at $\sqrt{s} = 8$ TeV*, *JHEP* **06** (2014) 090 [arXiv:1403.7366] [INSPIRE].

[148] CMS collaboration, *Search for s channel single top quark production in pp collisions at $\sqrt{s} = 7$ and 8 TeV*, *JHEP* **09** (2016) 027 [arXiv:1603.02555] [INSPIRE].

[149] CMS collaboration, *Cross section measurement of t-channel single top quark production in pp collisions at $\sqrt{s} = 13$ TeV*, *Phys. Lett. B* **772** (2017) 752 [arXiv:1610.00678] [INSPIRE].

[150] CMS collaboration, *Measurement of differential cross sections and charge ratios for t-channel single top quark production in proton-proton collisions at $\sqrt{s} = 13$ TeV*, *Eur. Phys. J. C* **80** (2020) 370 [arXiv:1907.08330] [INSPIRE].

[151] ATLAS collaboration, *Measurement of the production cross-section of a single top quark in association with a W boson at 8 TeV with the ATLAS experiment*, *JHEP* **01** (2016) 064 [arXiv:1510.03752] [INSPIRE].

[152] ATLAS collaboration, *Measurement of single top-quark production in association with a W boson in the single-lepton channel at $\sqrt{s} = 8$ TeV with the ATLAS detector*, *Eur. Phys. J. C* **81** (2021) 720 [arXiv:2007.01554] [INSPIRE].

[153] ATLAS collaboration, *Measurement of the cross-section for producing a W boson in association with a single top quark in pp collisions at $\sqrt{s} = 13$ TeV with ATLAS*, *JHEP* **01** (2018) 063 [arXiv:1612.07231] [INSPIRE].

[154] ATLAS collaboration, *Observation of the associated production of a top quark and a Z boson in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *JHEP* **07** (2020) 124 [arXiv:2002.07546] [INSPIRE].

[155] CMS collaboration, *Observation of the associated production of a single top quark and a W boson in pp collisions at $\sqrt{s} = 8$ TeV*, *Phys. Rev. Lett.* **112** (2014) 231802 [arXiv:1401.2942] [INSPIRE].

[156] CMS collaboration, *Measurement of the production cross section for single top quarks in association with W bosons in proton-proton collisions at $\sqrt{s} = 13$ TeV*, *JHEP* **10** (2018) 117 [arXiv:1805.07399] [INSPIRE].

[157] CMS collaboration, *Observation of single top quark production in association with a Z boson in proton-proton collisions at $\sqrt{s} = 13$ TeV*, *Phys. Rev. Lett.* **122** (2019) 132003 [arXiv:1812.05900] [INSPIRE].

[158] CMS collaboration, *Inclusive and differential cross section measurements of single top quark production in association with a Z boson in proton-proton collisions at $\sqrt{s} = 13$ TeV*, *JHEP* **02** (2022) 107 [arXiv:2111.02860] [INSPIRE].

[159] CMS collaboration, *Observation of tW production in the single-lepton channel in pp collisions at $\sqrt{s} = 13$ TeV*, *JHEP* **11** (2021) 111 [arXiv:2109.01706] [INSPIRE].

[160] NNPDF collaboration, *Response to "Parton distributions need representative sampling"*, arXiv:2211.12961 [INSPIRE].

[161] Z. Kassabov, E.R. Nocera and M. Wilson, *Regularising experimental correlations in LHC data: theory and application to a global analysis of parton distributions*, *Eur. Phys. J. C* **82** (2022) 956 [arXiv:2207.00690] [INSPIRE].

[162] R. Frederix et al., *The automation of next-to-leading order electroweak calculations*, *JHEP* **07** (2018) 185 [*Erratum ibid.* **11** (2021) 085] [arXiv:1804.10017] [INSPIRE].

[163] 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].

[164] 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].

[165] T. Kluge, K. Rabbertz and M. Wobisch, *FastNLO: fast pQCD calculations for PDF fits*, in the proceedings of the 14th *international workshop on deep inelastic scattering*, (2006), p. 483 [DOI:10.1142/9789812706706_0110] [hep-ph/0609285] [INSPIRE].

[166] FASTNLO collaboration, *Theory-data comparisons for jet measurements in hadron-induced processes*, arXiv:1109.1310 [INSPIRE].

[167] FASTNLO collaboration, *New features in version 2 of the fastNLO project*, in the proceedings of the 20th *international workshop on deep-inelastic scattering and related subjects*, (2012), p. 217 [DOI:10.3204/DESY-PROC-2012-02/165] [arXiv:1208.3641] [INSPIRE].

[168] V. Bertone et al., *aMCfast: automation of fast NLO computations for PDF fits*, *JHEP* **08** (2014) 166 [arXiv:1406.7693] [INSPIRE].

[169] 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].

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

[171] *HighTEA (High-energy Theory Event Analyzer for collider processes) webpage*, https://www.precision.hep.phy.cam.ac.uk/hightea/.

[172] A. Kulesza et al., *Associated production of a top quark pair with a heavy electroweak gauge boson at NLO+NNLL accuracy*, *Eur. Phys. J. C* **79** (2019) 249 [arXiv:1812.08622] [INSPIRE].

[173] 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].

[174] R. Aoude et al., *Renormalisation group effects on SMEFT interpretations of LHC data*, arXiv:2212.05067 [INSPIRE].

[175] A. Alloul et al., *FeynRules 2.0 — a complete toolbox for tree-level phenomenology*, *Comput. Phys. Commun.* **185** (2014) 2250 [arXiv:1310.1921] [INSPIRE].

[176] C. Degrande et al., *Automated one-loop computations in the standard model effective field theory*, *Phys. Rev. D* **103** (2021) 096024 [arXiv:2008.11743] [INSPIRE].

[177] R.D. Ball et al., *A first unbiased global NLO determination of parton distributions and their uncertainties*, *Nucl. Phys. B* **838** (2010) 136 [arXiv:1002.4407] [INSPIRE].

[178] R.D. Ball et al., *Parton distributions with LHC data*, *Nucl. Phys. B* **867** (2013) 244 [arXiv:1207.1303] [INSPIRE].

[179] M. Abadi et al., *TensorFlow: large-scale machine learning on heterogeneous distributed systems webpage*, http://tensorflow.org/.

[180] D. Anderson, J.-R. Vlimant and M. Spiropulu, *An MPI-based python framework for distributed training with Keras*, in the proceedings of the *International conference for high performance computing, networking, storage and analysis*, (2017) [arXiv:1712.05878] [INSPIRE].

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

[182] T. Giani, G. Magni and J. Rojo, *SMEFiT: a flexible toolbox for global interpretations of particle physics data with effective field theories*, *Eur. Phys. J. C* **83** (2023) 393 [arXiv:2302.06660] [INSPIRE].

[183] J. De Blas et al., *HEPfit: a code for the combination of indirect and direct constraints on high energy physics models*, *Eur. Phys. J. C* **80** (2020) 456 [arXiv:1910.14012] [INSPIRE].

[184] N. Castro et al., *EFTfitter — a tool for interpreting measurements in the context of effective field theories*, *Eur. Phys. J. C* **76** (2016) 432 [arXiv:1605.05585] [INSPIRE].

[185] Z. Han and W. Skiba, *Effective theory analysis of precision electroweak data*, *Phys. Rev. D* **71** (2005) 075009 [hep-ph/0412166] [INSPIRE].

[186] J. Brehmer, K. Cranmer, F. Kling and T. Plehn, *Better Higgs boson measurements through information geometry*, *Phys. Rev. D* **95** (2017) 073002 [arXiv:1612.05261] [INSPIRE].

[187] R. Abdul Khalek et al., *Towards ultimate parton distributions at the high-luminosity LHC*, *Eur. Phys. J. C* **78** (2018) 962 [arXiv:1810.03639] [INSPIRE].

[188] 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].

[189] N.T. Hunt-Smith et al., *Determination of uncertainties in parton densities*, *Phys. Rev. D* **106** (2022) 036003 [arXiv:2206.10782] [INSPIRE].

[190] M.U. Madigan, L. Mantani and J. Moore, *Understanding the Monte Carlo replica method in presence of non-linearity*, in progress.

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

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

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

[194] 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].

[195] S. Dawson et al., *LHC EFT WG note: precision matching of microscopic physics to the Standard Model Effective Field Theory (SMEFT)*, arXiv:2212.02905 [InSPIRE].

[196] R. Gomez Ambrosio et al., *Unbinned multivariate observables for global SMEFT analyses from machine learning*, *JHEP* **03** (2023) 033 [arXiv:2211.02058] [InSPIRE].

[197] R. Boughezal et al., *Neutral-current electroweak physics and SMEFT studies at the EIC*, *Phys. Rev. D* **106** (2022) 016006 [arXiv:2204.07557] [InSPIRE].

[198] J.L. Feng et al., *The forward physics facility at the high-luminosity LHC*, *J. Phys. G* **50** (2023) 030501 [arXiv:2203.05090] [InSPIRE].

[199] L.A. Anchordoqui et al., *The forward physics facility: sites, experiments, and physics potential*, *Phys. Rept.* **968** (2022) 1 [arXiv:2109.10905] [InSPIRE].

[200] F. Feroz, M.P. Hobson, E. Cameron and A.N. Pettitt, *Importance nested sampling and the MultiNest algorithm*, *Open J. Astrophys.* **2** (2019) 10 [arXiv:1306.2144] [InSPIRE].

[201] F. Feroz and M.P. Hobson, *Multimodal nested sampling: an efficient and robust alternative to MCMC methods for astronomical data analysis*, *Mon. Not. Roy. Astron. Soc.* **384** (2008) 449 [arXiv:0704.3704] [InSPIRE].