

SMEFiT: a flexible toolbox for global interpretations of particle physics data with effective field theories

Tommaso Giani, Giacomo Magni, and Juan Rojo

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

Nikhef Theory Group, Science Park 105, 1098 XG Amsterdam, The Netherlands

Abstract

The Standard Model Effective Field Theory (SMEFT) provides a robust framework to interpret experimental measurements in the context of new physics scenarios while minimising assumptions on the nature of the underlying UV-complete theory. We present the PYTHON open source SMEFiT framework, designed to carry out parameter inference in the SMEFT within a global analysis of particle physics data. SMEFiT is suitable for inference problems involving a large number of EFT degrees of freedom, without restrictions on their functional dependence in the fitted observables, can include UV-inspired restrictions in the parameter space, and implements arbitrary rotations between operator bases. Posterior distributions are determined from two complementary approaches, Nested Sampling and Monte Carlo optimisation. SMEFiT is released together with documentation, tutorials, and post-analysis reporting tools, and can be used to carry out state-of-the-art EFT fits of Higgs, top quark, and electroweak production data. To illustrate its functionalities, we reproduce the results of the recent ATLAS EFT interpretation of Higgs and electroweak data from Run II and demonstrate how equivalent results are obtained in two different operator bases.

Contents

1	Introduction	2
2	The SMEFIT framework	3
3	The ATLAS Higgs EFT analysis as case study	10
4	Summary and outlook	13

1 Introduction

Global interpretations of particle physics observables in the framework of the Standard Model Effective Field Theory (SMEFT) [1–3], see [4–13] for recent analyses, require the inference of several tens or eventually hundreds of independent Wilson coefficients from experimental data. For instance, the SMEFIT analysis [14] of Higgs, top, and diboson data from the LHC constrains 36 independent directions in the SMEFT parameter space, with 14 more coefficients fixed indirectly from electroweak precision observables. Exploring efficiently such a broad parameter space is only possible by means of the combination of a large number of measurements from different processes with state-of-the-art theoretical calculations. Realizing the ultimate goal of the SMEFT paradigm, a theory-assisted combination of measurements from the high-energy frontier down to the electroweak scale, flavour physics, and low-energy observables, demands a flexible and robust EFT fitting methodology amenable for inference problems involving up to hundreds of coefficients from a wide range of different physical observables, each of them characterised by their own statistical model.

In addition to being able to constrain such large parameter spaces, a SMEFT fitting framework suitable to the aforementioned goal ought to satisfy several other requirements. These include, but are not restricted to: carrying out fits both at the linear and at the quadratic level in the EFT expansion; accounting for all relevant sources of methodological, theoretical, and experimental uncertainties; exhibiting a modular structure enabling the seamlessly incorporation of new processes or improved theory calculations; accepting general likelihood functions beyond the multi-Gaussian approximation like those associated to unbinned or Poisson-distributed observables; providing statistical and visualization diagnosis tools to assist the interpretation of the results, from PCA and Fisher Information to basis rotation and reduction algorithms; and implementing theoretical constraints on the parameter space, such as those associated to the matching to UV-complete scenarios. Furthermore, the availability of such a fitting tool as open source would facilitate its adoption by interested parties both from the theory and the experimental communities.

Several fitting frameworks have been developed and deployed in the context of SMEFT interpretations of particle physics data, such as SMEFIT [4], FITMAKER [10], HEPFIT [15], EFTFITTER [16], and SFITTER [5] among others. SMEFIT was originally developed in the context of top quark studies [4], and then extended to Higgs and diboson measurements in [14] and to vector boson scattering in [17]. Whenever possible, these analyses strived to account for NLO QCD calculations in the EFT cross-sections such as those provided by the SMEFT@NLO package [18]. Additional studies based on SMEFIT are the implementation of the Bayesian reweighting method [19] and the LHC EFT WG report on experimental observables [20]. Advantages of SMEFIT as compared to related frameworks include independent and complementary statistical methods to carry out parameter inference (Monte Carlo optimisation and Nested Sampling), the lack of restrictions on the allowed functional dependence in c_i/Λ^2 for the fitted observables, and a competitive scaling of the running time with the number of fitted parameters.

The goal of this paper is to present and describe the release of SMEFIT as a PYTHON open source fitting framework, together with the corresponding datasets and theory calculations required to reproduce published analyses. SMEFIT is made available via its public GitHub repository

https://github.com/LHCfitNikhef/smefit_release

together with the documentation and user-friendly tutorials provided in

https://lhcfitsnikhef.github.io/smefit_release

which also includes a catalog of SMEFT analyses corresponding to different choices of the input dataset, theoretical settings, and statistical methodology. This online documentation is the main resource to consult in order to use SMEFIT, either to reproduce existing analyses or to extend them to new processes and

observables, and hence here we restrict ourselves to highlighting key representative results. Technical aspects of the framework, such as the format in which the data and the theory calculations are to be provided, are already described in the online documentation and therefore are not covered in this paper.

Here first we reproduce the results of the global SMEFT analysis of [14]. In doing so, we fix a number of small issues that were identified during the code rewriting process. Then we illustrate the capabilities of SMEFT by independently reproducing the ATLAS EFT interpretation of Higgs and electroweak data from the LHC, together with LEP measurements, presented in [21]. We demonstrate how when using the same inputs in terms of experimental data and EFT parametrisation one obtains the same bounds in the Wilson coefficients. Furthermore, we show how results are independent of the choice of fitting basis: equivalent results are obtained when using either the Warsaw basis or the rotated basis chosen in [21] to restrict the parameter space to directions with large variability.

This benchmark exercise displays the capabilities of SMEFT to contribute to the ongoing and future generation of SMEFT studies, where the careful comparison between the outcomes from different groups is instrumental to cross-check independent determinations. This program, partly carried out in the context of the LHC EFT WG activities, aims to bring the robustness of SMEFT analyses on par to that of SM calculations, simulations, and benchmark comparisons.

The outline of this paper is as follows. Sect. 2 describes the SMEFT framework and validates the code rewriting by comparing its outcome with that of the global SMEFT analysis of [14]. Sect. 3 illustrates the possible applications of SMEFT by reproducing the results of the ATLAS EFT fit of LHC and LEP data from [21] and demonstrating the fitting basis independence of the results. We conclude and outline possible future developments in Sect. 4.

2 The SMEFT framework

In this section we provide a concise overview of the main features and functionalities of the SMEFT framework, pointing the reader to the original publications [4, 14, 17, 19] as well as to the code online documentation for more details.

Installation. The SMEFT code can be installed by using the CONDA interface. An installation script is provided, allowing the user to create a CONDA environment compatible with the one which has been automatically tested, and where the SMEFT package can be installed and executed. CONDA lock files ensure that results are always produced using the correct version of the code dependencies, regardless of the machine where the environment is created, hence ensuring complete reproducibility of the results. In this same environment the code can be easily edited, allowing the users to contribute to the development of the open-source framework.

EFT cross-section parametrisation. In the presence of n_{op} dimension-six SMEFT operators, a general SM cross-section σ_{SM} will be modified as follows

$$\sigma_{\text{eft}}(\mathbf{c}/\Lambda^2) = \sigma_{\text{SM}} + \sum_{i=1}^{n_{\text{op}}} \tilde{\sigma}_{\text{eft},i} \frac{c_i}{\Lambda^2} + \sum_{i,j=1}^{n_{\text{op}}} \tilde{\sigma}_{\text{eft},ij} \frac{c_i c_j}{\Lambda^4}, \quad (1)$$

with $\tilde{\sigma}_{\text{eft},i}/\Lambda^2$ and $\tilde{\sigma}_{\text{eft},ij}/\Lambda^4$ indicating respectively the contributions to the cross-section arising from the interference with the SM amplitudes and from the square of the EFT ones, once the Wilson coefficients c_i are factored out. These cross-sections hence depend on n_{op} Wilson coefficients and on the cutoff scale Λ , with only the ratios $\mathbf{c}/\Lambda^2$ being accessible in a model-independent analysis. Eq. (1) can be generalised when other types of SMEFT operators, e.g. dimension-eight operators, are considered in the interpretation of the observable.

The terms σ_{SM} , $\tilde{\sigma}_{\text{eft},i}$ and $\tilde{\sigma}_{\text{eft},ij}$ in Eq. (1), are inputs to SMEFT and are provided by means of external calculations. In this respect, the fitting code is agnostic in the calculational settings used to produce them, provided they comply with the required format of the theory tables described below.

The user can also choose to adopt an alternative form for the theory predictions

$$\sigma_{\text{eft}}(\mathbf{c}/\Lambda^2) = \sigma_{\text{SM}} \left(1 + \sum_{i=1}^{n_{\text{op}}} \kappa_{\text{eft},i} \frac{c_i}{\Lambda^2} + \sum_{i,j=1}^{n_{\text{op}}} \kappa_{\text{eft},ij} \frac{c_i c_j}{\Lambda^4} \right), \quad (2)$$

with now the EFT contributions entering as K -factors multiplying the SM prediction. The multiplicative variant in Eq. (2) is equivalent to Eq. (1) only in cases where higher order QCD and electroweak corrections coincide in the SM and in the SMEFT. Eq. (2) benefits from reduced theory uncertainties on the EFT contribution, such as the one due to PDFs, which partially cancel when taking the K -factor ratios.

Experimental data and theory predictions. The theoretical predictions and experimental data for the processes entering an EFT interpretation are considered as external, user-provided inputs to SMEFT. As such, they are stored in the following separate GitHub repository

https://github.com/LHCfitNikhef/smefit_database

since in this way one separates code developments from changes in the external data and theory inputs. This repository should be cloned separately and then the local path specified in the runcard. Detailed instructions are given in the online documentation.

Currently, this database repository includes the tables for experimental data and theory predictions required to reproduce the global SMEFT analysis of [14], see also Fig. 2, as well as the ATLAS EFT interpretation of [21], to be discussed in Sect. 3. This database will be kept updated as additional processes and improved theory calculations entering the SMEFT global analyses are included. Users of the code can take the existing datasets as templates for the implementation of new processes for their own EFT fits.

Theory predictions are stored in JSON format files composed by a dictionary that contains, for each dataset, the central SM predictions, the LO and NLO linear and quadratic EFT cross-sections, and the theory covariance matrix. For the experimental data instead we adopt a YAML format which contains the number of data points, central values, statistical errors, correlated systematic errors, and the type of systematic error (additive or multiplicative), from which the covariance matrix of the measurements can be constructed. Alternatively, for datasets in which the breakdown of systematic errors is not provided, the user has to decompose the covariance matrix into a set of correlated systematic errors. Details regarding the format of data, uncertainties, and theory predictions are provided in the corresponding section of the code documentation

https://lhcfitsnikhef.github.io/smefit_release/data_theory/data.html

The SMEFT runcard which steers the code should list the experimental inputs that enter the fit and the corresponding theory calculations, including the path to the folders where these inputs are stored. In the code repository one can find examples of runcards that can be used to reproduce the two EFT interpretations mentioned above, together with the corresponding post-fit analysis reports.

Likelihood function. The goal of SMEFT is to determine confidence level intervals in the space of the Wilson coefficients given an input dataset $\mathcal{D}$ and the corresponding theory predictions $\mathcal{T}(\mathbf{c})$, with the latter given by Eq. (1) or generalizations thereof. The agreement between the dataset $\mathcal{D}$ and a theory hypothesis $\mathcal{T}(\mathbf{c})$ is quantified by the likelihood function $\mathcal{L}(\mathbf{c})$. The wide majority of measurements used in EFT interpretations are presented as multi-Gaussian distributions, for which the likelihood is given by

$$\mathcal{L}(\mathbf{c}) \propto \prod_{m,n=1}^{n_{\text{dat}}} \exp \left[-(\mathcal{T}_m(\mathbf{c}) - \mathcal{D}_m) (\text{cov}^{-1})_{mn} (\mathcal{T}_n(\mathbf{c}) - \mathcal{D}_n) \right]. \quad (3)$$

In such cases, the log-likelihood function $(-\log \mathcal{L})$ becomes either a quadratic or a quartic function of the Wilson coefficients $\mathbf{c}$, depending on whether the quadratic terms in the EFT parametrization of Eq. (1) are retained, according to the user specifications in the runcard.

Despite Eq. (3) being the only functional form for the likelihood which is currently implemented in SMEFT, the modular structure of the code can be easily extended to accomodate alternative likelihood functions. For instance, for unbinned observables [22] the likelihood would receive a contribution of the form

$$\mathcal{L}(\mathbf{c}) = \frac{\nu_{\text{tot}}(\mathbf{c})^{N_{\text{ev}}}}{N_{\text{ev}}!} e^{-\nu_{\text{tot}}(\mathbf{c})} \prod_{i=1}^{N_{\text{ev}}} \frac{1}{\sigma_{\text{fid}}(\mathbf{c})} \frac{d\sigma(\mathbf{x}, \mathbf{c})}{d\mathbf{x}}, \quad (4)$$

with N_{ev} being the number of events, ν_{tot} the expected event count from theory, and the event probability is determined by the cross-section differential in the event kinematics $\mathbf{x}$. Such unbinned likelihood could be implemented in SMEFT, allowing for a general EFT interpretation involving a combination of measurements each of which described by a different statistical model.

Nested Sampling. Once a statistical model $\mathcal{L}(\mathbf{c})$ for the input dataset $\mathcal{D}$ is defined in terms of the theory predictions $\mathcal{T}(\mathbf{c})$, SMEFIT proceeds to determine the most likely values of the Wilson coefficients $\mathbf{c}$ and the corresponding uncertainties by means of two different and complementary strategies. The first one is based on Nested Sampling (NS) via the MultiNest library [23, 24]. The installation of the latter is automatically carried out by the SMEFIT installation script.

The starting point is Bayes' theorem relating the posterior probability distribution of the parameters $\mathbf{c}$ given the observed data and the theory hypothesis, $P(\mathbf{c}|\mathcal{D}, \mathcal{T})$, to the likelihood function (conditional probability) and the prior distribution π ,

$$P(\mathbf{c}|\mathcal{D}, \mathcal{T}) = \frac{P(\mathcal{D}|\mathcal{T}, \mathbf{c}) P(\mathbf{c}|\mathcal{T})}{P(\mathcal{D}|\mathcal{T})} = \frac{\mathcal{L}(\mathbf{c}) \pi(\mathbf{c})}{\mathcal{Z}}, \quad (5)$$

with the Bayesian evidence $\mathcal{Z}$ ensuring the normalisation of the posterior distribution,

$$\mathcal{Z} = \int \mathcal{L}(\mathbf{c}) \pi(\mathbf{c}) d\mathbf{c}. \quad (6)$$

By means of Bayesian inference, NS maps the n_{op} -dimensional integral over the prior density into

$$X(\lambda) = \int_{\{\mathbf{c}: \mathcal{L}(\mathbf{c}) > \lambda\}} \pi(\mathbf{c}) d\mathbf{c}, \quad (7)$$

a one-dimensional function corresponding to the volume of the prior density $\pi(\mathbf{c})d\mathbf{c}$ associated to likelihood values greater than λ . Eq. (7) defines a transformation between the prior and posterior distributions sorted by the likelihood of each point in the EFT parameter space, is evaluated numerically, and results in n_{spl} samples $\{\mathbf{c}^{(k)}\}$ providing a representation of the posterior probability distribution from which one can evaluate confidence level intervals and related statistical estimators. The default SMEFIT analyses assumes a flat prior volume $\pi(\mathbf{c})$, although implementing alternative functional forms for the prior volume is an option available to the user.

One benefit of sampling methods such as NS is that they bypass limitations of numerical optimisation techniques such as local minima preventing reaching the absolute minimum, with a drawback being that the computational resources required in NS grow exponentially with the dimensionality n_{op} .

MCfit optimisation. The second strategy available in SMEFIT, denoted by MCfit, is based on the Monte Carlo replica method used e.g. by the NNPDF analyses of parton distributions [25, 26]. N_{rep} Gaussian replicas $\mathcal{D}_n^{(k)}$ of the experimental data $\mathcal{D}_n$, with $n = 1, \dots, n_{\text{dat}}$, are generated according to the covariance matrix of Eq. (3). Subsequently, the best-fit coefficients $\mathbf{c}^{(k)}$ for each data replica $\mathcal{D}_n^{(k)}$ are determined from the numerical minimisation of the log-likelihood function.

Several minimisers are available for this purpose in SMEFIT: the evolutionary CMA-ES algorithm [27] used in the fragmentation function fits of [28]; and two build-in minimizers provided by SCIPY [29, 30]. For each of them, the user can specify different settings controlling the efficiency and accuracy of the minimisation. Additional algorithms can be added by the user. We note that as opposed to the PDF fit case no cross-validation is required here, since overlearning is not possible for a discrete parameter space, where the best-fit value coincides with the absolute maximum of the likelihood.

The final result of MCfit is a sample of N_{rep} replicas $\{\mathbf{c}^{(k)}\}$ that provides a representation of the probability density in the space of SMEFT coefficients, and that can be processed in the same manner as its NS counterpart. While the posteriors obtained with MCfit and NS should be equivalent, in practice small residual differences can appear and traced back to numerical inefficiencies of the minimiser. In this respect, we recommend that in SMEFIT the NS method is adopted as baseline, with MCfit as an independent cross-check. As compared to NS, the computational performance of MCfit scales better with n_{op} with the duration of single-replica fits being the limiting factor.

Theoretical uncertainties. The covariance matrix that enters the Gaussian likelihood in Eq. (3) contains in general contributions of both experimental and theoretical origin. Assuming that these two sources of uncertainty are uncorrelated and that the latter can be approximated by a multi-Gaussian distribution [31, 32], the covariance matrix used in SMEFIT is defined by

$$\text{cov}_{nm} = \text{cov}_{nm}^{(\text{exp})} + \text{cov}_{nm}^{(\text{th})}, \quad n, m = 1, \dots, n_{\text{dat}}, \quad (8)$$

namely as the sum of the experimental and the theoretical covariance matrices. The latter should contain in principle all relevant sources of theory error such as PDF, missing higher orders (MHO), and MC integration uncertainties, with MHO being treated according to the formalism developed in [31,32]. In practice, theory errors should be specified in the theory tables for each measurement. Note that the code is agnostic with respect to the source of theory errors provided by the user, and in particular can account for the PDF and MHO uncertainties associated to the linear and quadratic EFT predictions whenever the theory covariance matrix provided in the SMEFT theory tables takes these into account. We note that in the current implementation, correlations between theory uncertainties corresponding to different datasets are neglected.

Constrained fits. Within a SMEFT interpretation of experimental data it is often necessary to impose relations between some of the fitted Wilson coefficients, rather than keeping all of them as free parameters. Such constraints in the SMEFT parameter space arise for instance as a consequence of the matching to specific UV-complete models, but also from the approximate implementation of electroweak precision observables via a restriction in the parameter space used in [14], as well as from simplified EFT interpretations with more restrictive flavour assumptions. One example of the latter, considered in [14] and proposed by the LHC Top working group in [33], is that of a top-philic scenario with new physics coupling preferentially to the top quark. This scenario is based on the assumption that new physics couples dominantly to the left-handed doublets and right-handed up-type quark singlet of the third generation as well as to gauge bosons, and as compared to the baseline settings in [14] this assumption introduces additional restrictions in the EFT parameter space.

In general, linear constraints can be implemented via the SMEFT runcard and lead to a speedup of the fitting procedure. The implementation of the same type of constraints a posteriori by means of the Bayesian reweighting method [34] was demonstrated in [19], showing that it leads to a large efficiency loss and hence is only reliable for moderate restrictions in the parameter space.

In several scenarios the matching procedure between the SMEFT and UV-complete models results in non-linear relations between the Wilson coefficients. The automated implementation of such non-linear constraints in SMEFT is work in progress and requires non-trivial modifications of the fit procedure. An upcoming publication focused on matching to UV-complete models will discuss this problem in more detail.

Basis selection and rotation. The baseline choice for the theory tables containing the linear and quadratic EFT predictions in SMEFT is that these are provided by the user in the Warsaw basis. In general, it might be more convenient to carry out the fit in a different basis, for instance one closer to the actual constraints imposed by the experimental data considered. The user can thus indicate how the chosen fitting basis is related to the Warsaw operators by means of a rotation matrix

$$\mathcal{O}_i^{(F)} = \sum_{j=1}^{n_{\text{op}}} R_{ij}^{(W \rightarrow F)} \mathcal{O}_j^{(W)}, \quad i = 1, \dots, n_{\text{op}}, \quad R^{(F \rightarrow W)} = \left(R^{(W \rightarrow F)} \right)^{-1}, \quad (9)$$

with $\mathcal{O}_i^{(W)}$ and $\mathcal{O}_j^{(F)}$ indicating the operators in the Warsaw and fitting bases respectively. Note that the number of operators can be different in the two bases considered, or more precisely, in the fit basis a number of operators can be set to zero, for instance when unconstrained by the data.

This rotation matrix can also be determined automatically from a principal component analysis (PCA) of the Fisher information matrix (defined below) which determines the directions with the highest variability. EFT directions with the lowest variability can be set to zero as a constrain in order to remove quasi-flat directions and thus increase the numerical stability of the fits. Results of an EFT interpretation should of course be basis independent, provided that the two bases are related by a rotation. We will exploit these functionalities of the SMEFT framework in Sect. 3 to reproduce the ATLAS EFT interpretation of [21] in two different bases and verify that results are identical.

Fisher information. A measure of the sensitivity of individual datasets to specific directions in the EFT parameter space is provided by the Fisher information matrix I_{ij} , defined as

$$I_{ij}(\mathbf{c}) = -\text{E} \left[\frac{\partial^2 \ln \mathcal{L}(\mathbf{c})}{\partial c_i \partial c_j} \right], \quad i, j = 1, \dots, n_{\text{op}}, \quad (10)$$

where $\text{E}[\]$ indicates the expectation value over the Wilson coefficients and $\mathcal{L}(\mathbf{c})$ is the likelihood function. The covariance matrix of the Wilson coefficients, $C_{ij}(\mathbf{c})$, is bounded by the Fisher information matrix,

$C_{ij} \geq (I^{-1})_{ij}$, the so-called Cramer-Rao bound, which illustrates how I_{ij} quantifies the constraining power of the dataset $\mathcal{D}$.

In the specific case of linear EFT calculations and a diagonal covariance matrix, the Fisher information matrix Eq. (10) simplifies to

$$I_{ij} = \sum_{m=1}^{n_{\text{dat}}} \frac{\sigma_{\text{eft},i,m} \sigma_{\text{eft},j,m}}{\delta_{\text{tot},m}^2}, \quad i, j = 1, \dots, n_{\text{op}} \quad (11)$$

with $\delta_{\text{tot},m}$ being the total uncertainty of the m -th data point, such that I_{ij} is independent of the fit results and can be evaluated a priori. Eq. (11) shows that at the linear EFT level the Fisher information is the average of the EFT corrections to the SM cross-section in the dataset $\mathcal{D}$ in units of the measurement uncertainty. We emphasize that in SMEFIT we always evaluate Eq. (10) in terms of the full covariance matrix and Eq. (11) is provided only for illustration purposes.

SMEFIT evaluates the Fisher information matrix Eq. (10) for the datasets and theory predictions specified in the runcard, and presents the results graphically to facilitate the interpretation of the results. The absolute normalisation of the Fisher matrix is arbitrary, since one can always rescale operator normalizations. Hence we normalise it such that it becomes independent of the choice of overall operator normalisation. As mentioned above, the user can apply the PCA to this Fisher information matrix to determine the directions (principal components) with highest variability, and eventually use them as fitting basis, rather than the original Warsaw basis, by applying a rotation of the form of Eq. (9).

Fig. 1 displays the PCA applied to the Fisher information matrix (in the linear EFT case) for the global dataset of [14]. For each principal component, we display the coefficients of the linear combination of fit basis operators and the corresponding singular value. Three flat directions, corresponding to three linear combinations of four-heavy-quark operators, have vanishing singular values indicating that cannot be constrained from the fit.

Fit report and visualization of results. The output of a SMEFIT analysis consists on the posterior probability distributions associated to the fit coefficients $\mathbf{c}$ as well as ancillary statistical estimators such as the fit quality per dataset. This output can be processed and visualized by means of a fit report, which is generated by specifying a separate runcard either for an individual fit or for pair-wise (or multiple) comparisons between fits.

By modifying this dedicated runcard the user can specify what to display in the report, with currently available options including comparisons between SM and best-fit EFT predictions for individual datasets, posterior distributions with associated correlation and confidence level bar plots, two-parameter contour plots, the log-likelihood distribution among replicas or samples, Fisher matrix by dataset, and the outcome of the PCA analysis among others. The online documentation contains the description of the report runcard and examples of fits reports obtained with specific runcards,

https://lhcfitsnikhef.github.io/smefit_release/report/running.html.

The SMEFIT report is produced both in .pdf and .html format to facilitate readability and visualization.

Code rewriting and validation. As compared to the version of the code used for the global SMEFT analysis of [14], the SMEFIT framework has been completely rewritten in preparation for its public release, streamlining its overall structure and enhancing its modular character and user-friendly interface. In this process, both the code itself and the data and theory tables have been repeatedly cross-checked using the baseline fit of [14] as benchmark. A number of small issues were identified and corrected in the theory tables, without affecting any of the main findings of the original study. We have verified that the only differences between [14] and the results shown in Fig. 2 obtained with the new code is related to bug fixes in the theory tables, and that if with the new code we use the same theory tables as in [14], identical posterior distributions are obtained both at the linear and the quadratic EFT levels.

To highlight the agreement between the results of [14] and the output of the new SMEFIT release code, Fig. 2 compares the posterior distributions in the SMEFT parameter space obtained in [14] with those based on the new version of the code and of the theory and data tables. Posterior distributions are evaluated with Nested Sampling for the global dataset, while EFT cross-sections account for both NLO QCD corrections and for quadratic $\mathcal{O}(\Lambda^{-4})$ effects. The 95% CL intervals obtained in both cases are very similar, with

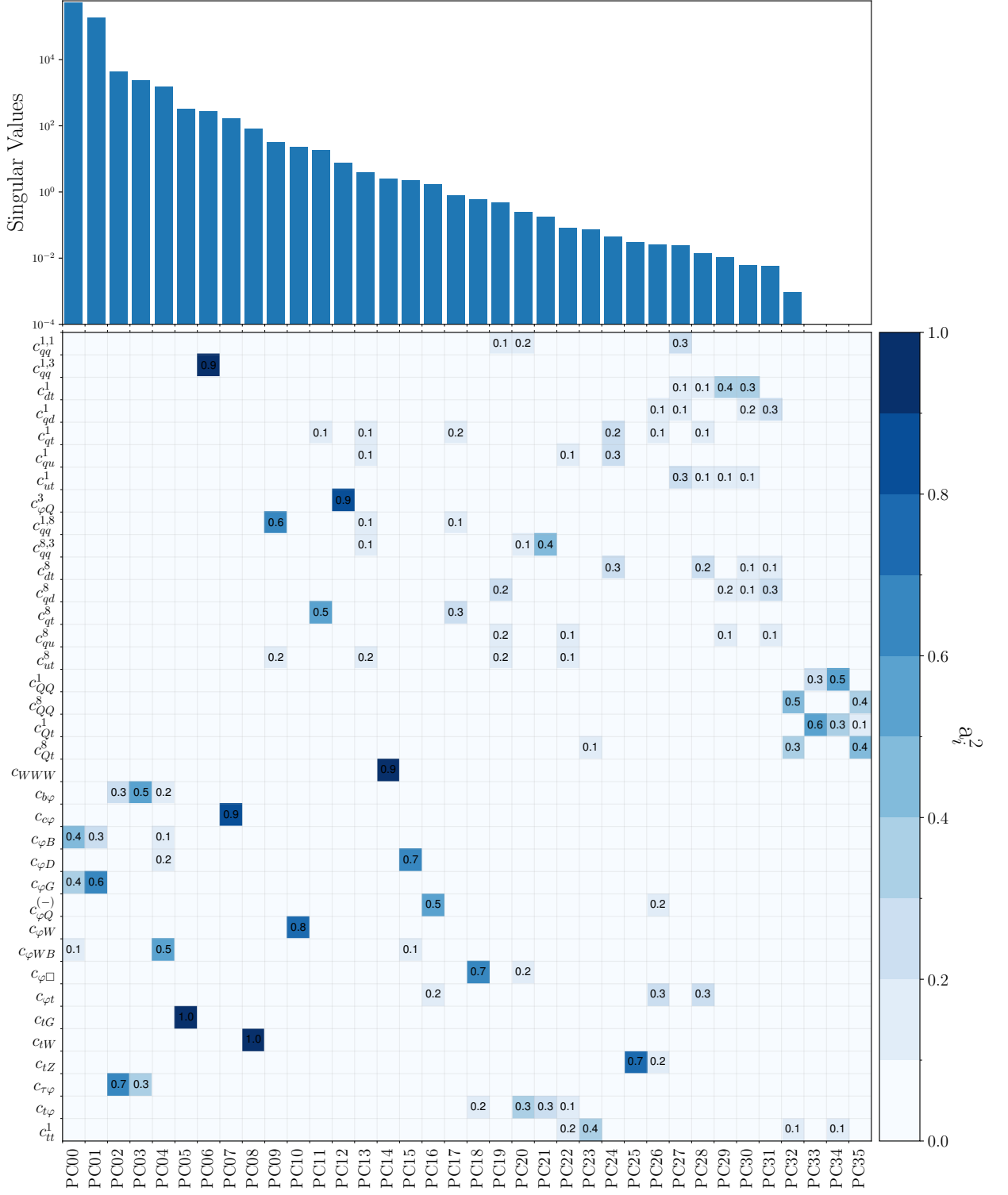


Figure 1. Results of the Principal Component Analysis applied to the Fisher information matrix (for linear EFT calculations) of the global dataset of [14]. For each principal component, we display the coefficients of the linear combination of fit basis operators and the corresponding singular value. Three flat directions, corresponding to three linear combinations of four-heavy-quark operators, have vanishing singular values indicating that cannot be constrained from the fit at the linear level.

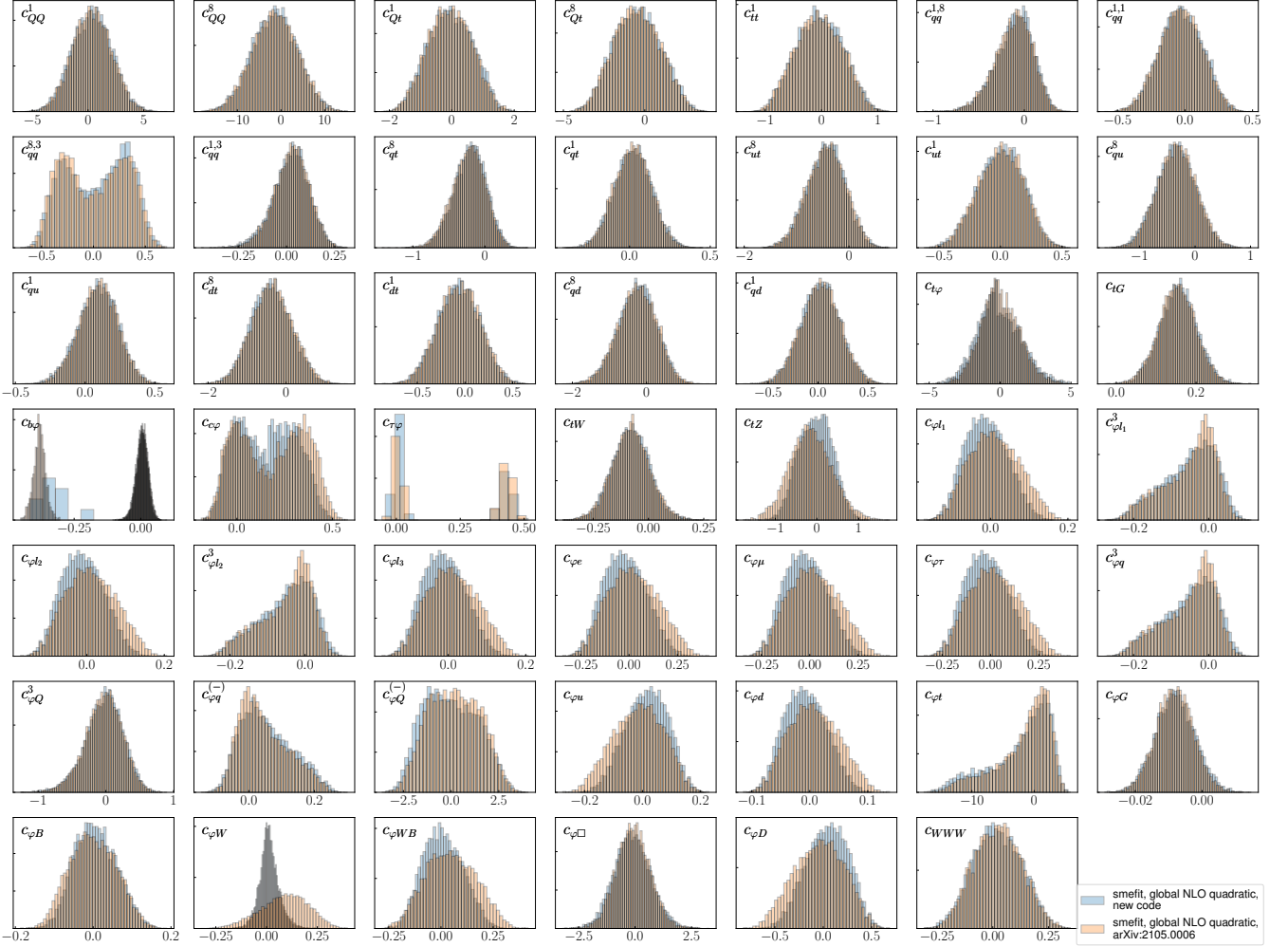


Figure 2. The posterior distributions in the SMEFT parameter space from the SMEFT global analysis of [14] compared to those obtained from the new code for the same theory settings and datasets, using in both cases NS and EFT cross-sections accounting for NLO QCD corrections and quadratic $\mathcal{O}(\Lambda^{-4})$ effects. The shown posteriors assume $\Lambda = 1$ TeV, and can be appropriately rescaled for other values of Λ .

possibly the exception of the $c_{\varphi W}$ bosonic operators where its uncertainty was somewhat overestimated in the original analysis. A similar or better level of agreement is found for the linear EFT fits and for the fits based on LO EFT cross-sections. Good agreement is also found for related fit estimators, such as the correlation matrix between Wilson coefficients.

As mentioned above, within the SMEFT framework one can choose between two alternative and complementary strategies to determine CL intervals on the Wilson coefficients entering the theory calculations, namely Nested Sampling and MCfit. Each method has its own advantages and disadvantages, for instance MCfit scales better with the number of fit parameters but may be affected by numerical inefficiencies of the minimisation procedure, specially for poorly constrained operators. Within the current framework, we recommend users to adopt NS as the default strategy and use MCfit as an independent cross-check.

Fig. 3 presents the comparison between linear EFT fits performed with NS and MCfit for the same data and theory settings as in global SMEFT fit of Fig. 2. The left panel compares the magnitude of the 95% CL intervals for c_i/Λ^2 for the $n_{\text{op}} = 49$ Wilson coefficients considered in the analysis, while the right panel displays the median and 68% CL and 95% CL (thick and thin respectively) intervals in each case. Results are grouped by operator family: from top to bottom we show the two-fermion, two-light-two-heavy four-fermion, the four-heavy-fermion, and the purely bosonic operators.

In the case of the linear EFT fits, results obtained with NC and MCfit are essentially identical at the level of the (Gaussian) posterior distributions, though a large number of replicas N_{rep} is required in MCfit to achieve sufficiently smooth shapes of the distributions, with residual differences moderate and confined to poorly constrained operators, such as the four-heavy top quark operators, that have a small contribution to the total χ^2 . Given that the NS and MCfit methods are orthogonal to each other, their agreement constitutes a non-trivial cross-check of the robustness of the global SMEFT analysis framework. The availability of such

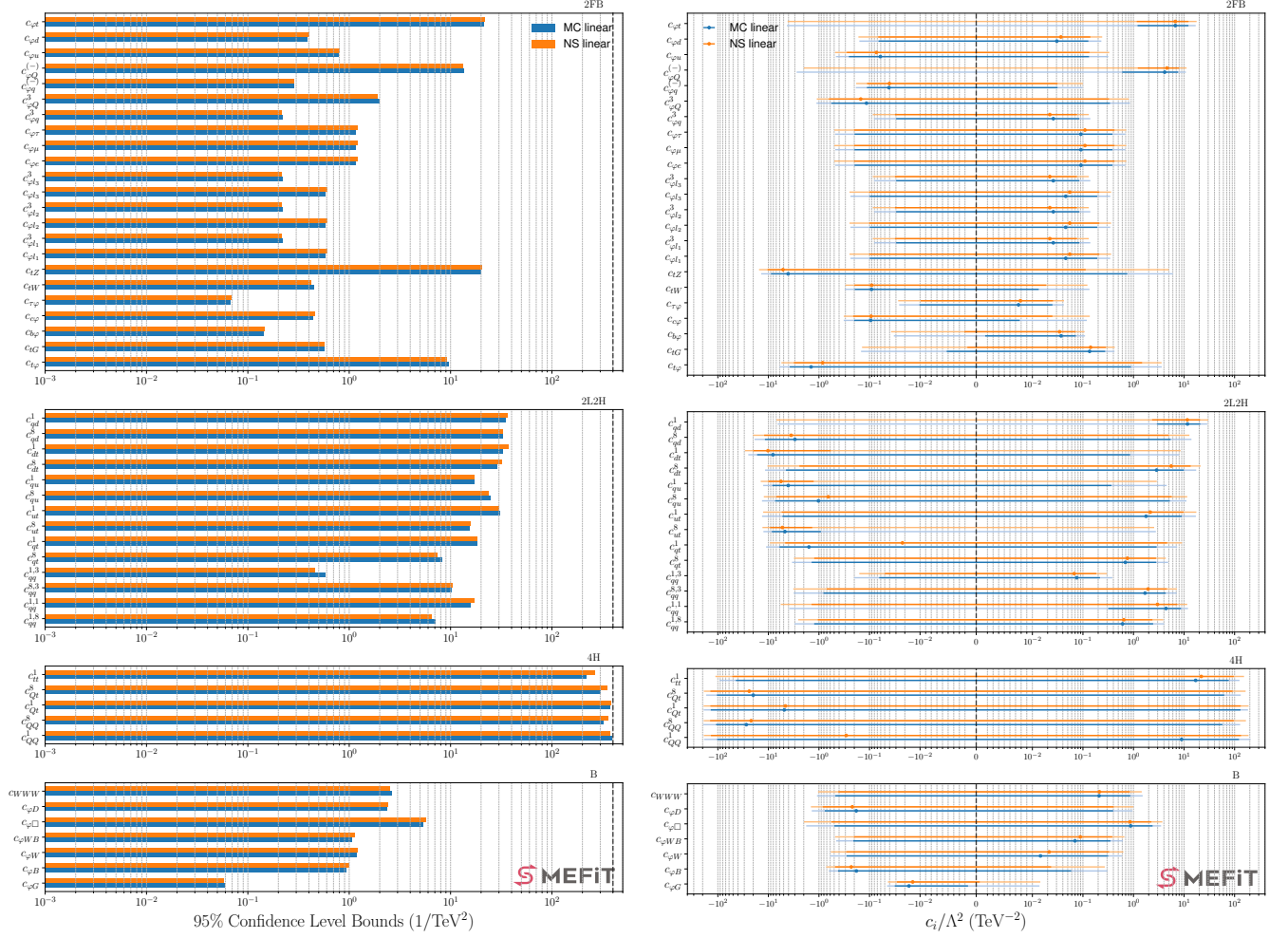


Figure 3. Results of global SMEFT fits with liner EFT corrections with the same inputs as in Fig. 2 obtained with either the NS or the MCfit methods. The left panel compares the magnitude of the 95% CL intervals for c_i/Λ^2 for the $n_{\text{op}} = 49$ Wilson coefficients considered in the analysis, while the right panel displays the median and 68% and 95% CL (thick and thin lines respectively) intervals in each case. Results are grouped by operator family: from top to bottom we show the two-fermion, two-light-two-heavy four-fermion, the four-heavy-fermion, and the purely bosonic operators.

functionality is specially useful to exclude that eventual deviations with respect to the SM baseline can be traced back to methodological limitations of the fitting framework.

On the other hand, when considering quadratic EFT fits, the agreement between NS and MCfit worsens for specific operators. This problem has been investigated in App. E of [35], where analytical calculations are performed for the posterior distributions in the case of single-parameter quadratic SMEFT fits, finding that MCfit results may not reproduce the correct Bayesian posteriors obtained from NS due to spurious solutions related to cancellations between the linear and quadratic EFT terms. This effect is most marked for observables where the quadratic EFT corrections dominate over the linear ones, and also whenever the SM cross-section overshoots sizably the central value of the experimental data. Hence [35] finds that NS and MCfit will only coincide at the quadratic level when all processes included in the fit are such that quadratic EFT corrections are subdominant as compared to the linear ones. However, for many of the observables considered in the global SMEFT fit, in particular those sensitive to the high energy tails accessible at the LHC, quadratic EFT effects are large and hence MCfit results may differ from the NS posteriors.

3 The ATLAS Higgs EFT analysis as case study

To illustrate the potential applications of the SMEFT framework, we independently reproduce the results of the ATLAS EFT interpretation presented in [21], which updates and supersedes previous ATLAS EFT studies [36, 37]. This analysis is based on the combination of ATLAS measurements of Higgs production, diboson production, and Z production in vector boson fusion with legacy electroweak precision observables

from LEP and SLC.

Our SMEFT-based reinterpretation is based on the same experimental data inputs, SM predictions, EFT cross-section parametrisations, and operator basis rotations as those used in [21]. This information is publicly available for the linear $\mathcal{O}(\Lambda^{-2})$ case: specifically, the central values, uncertainties and correlations of the experimental measurements have been extracted from Tables 4, 9, 10 and Fig. 18 of [21]; the definition of the rotated fit basis in terms of the Warsaw basis from Fig. 13; and for the linear EFT cross-sections a prescription based on the public numbers provided in [21, 38, 39] has been produced and used as input for the current analysis. These public inputs are also made available in the SMEFT repository, in order to facilitate the reproducibility of this benchmarking exercise.

The ATLAS EFT interpretation of [21] is based on Higgs boson production cross-sections and decay measurements carried out within the Simplified Template Cross-Section (STXS) framework [40] from the Run II dataset. It also contains selected electroweak Run II measurements, in particular diboson production in the WW , WZ , and 4ℓ final states as well as Z production in vector-boson-fusion, $pp \rightarrow Z(\rightarrow \ell^+\ell^-)jj$. Note that the diboson 4ℓ final state targets both on-shell ZZ as well as off-shell Higgs boson production. Table 1 summarizes the information associated to these ATLAS experimental inputs.

Process	Details	$\mathcal{L}$ [fb^{-1}]	Ref.
$pp \rightarrow h \rightarrow \gamma\gamma$	ggF, VBF, Wh , Zh , $t\bar{t}h$, th	139	[41]
$pp \rightarrow h \rightarrow ZZ^*$	ggF, VBF, Wh , Zh , $t\bar{t}h(4\ell)$	139	[42]
$pp \rightarrow h \rightarrow WW^*$	ggF, VBF	139	[43]
$pp \rightarrow h \rightarrow \tau\tau$	ggF, VBF, Wh , Zh , $t\bar{t}h(\tau_{\text{had}}\tau_{\text{had}})$	139	[44]
$pp \rightarrow h \rightarrow b\bar{b}$	Wh , Zh	139	[45–47]
	VBF	126	[48]
	$t\bar{t}h$	139	[49]
$pp \rightarrow e^\pm\nu\mu^\mp\nu$	$p_T^{\text{lead. lep.}} (m_{\ell\ell} > 55 \text{ GeV}, p_T^{\text{jet}} < 35 \text{ GeV})$	36	[50]
$pp \rightarrow \ell^\pm\nu\ell^+\ell^-$	$m_T^{WZ} (m_{\ell\ell} \in (81, 101) \text{ GeV})$	36	[51]
$pp \rightarrow \ell^+\ell^-\ell^+\ell^-$	$m_{ZZ} (m_{4\ell} > 180 \text{ GeV})$	139	[52]
$pp \rightarrow \ell^+\ell^-jj$	$\Delta\Phi_{jj} (m_{jj} > 1000 \text{ GeV}, m_{\ell\ell} \in (81, 101) \text{ GeV})$	139	[53]

Table 1. The ATLAS measurements included in the EFT interpretation of [21]. For each dataset we provide specific details of the measurement, the integrated luminosity $\mathcal{L}$, and the corresponding publication reference. For Higgs measurements we display which production modes are being targeted in the analysis. For electroweak measurements we indicate the differential distribution included in the fit and the main acceptance cuts. In this table $\ell = e, \mu$ denotes a first- or second-generation charged lepton.

The ATLAS data listed in Table 1 are complemented by the legacy LEP and SLC electroweak precision observables (EWPO) at the Z -pole from [54], required to constrain directions in the SMEFT parameter space not covered by LHC processes. Specifically, the analysis of [21] considers the inclusive cross-section into hadrons σ_{had}^0 , the ratio of partial decay widths R_ℓ^0 , R_q^0 , and the forward-backward asymmetries $A_{\text{fb}}^{0,\ell}$, $A_{\text{fb}}^{0,q}$ where q is measured separately for charm and bottom quarks and ℓ is the average over leptons. These EWPOs are defined as

$$\sigma_{\text{had}}^0 = \frac{12\pi}{m_Z^2} \frac{\Gamma_{ee}\Gamma_{\text{had}}}{\Gamma_Z^2}, \quad R_\ell^0 = \frac{\Gamma_{\text{had}}}{\Gamma_{\ell\ell}}, \quad R_q^0 = \frac{\Gamma_{qq}}{\Gamma_{\text{had}}}, \quad A_{\text{fb}}^0 = \frac{N_F - N_B}{N_F + N_B}, \quad (12)$$

where Γ_Z , Γ_{ee} , Γ_{had} , $\Gamma_{\ell\ell}$ and Γ_{qq} are the total and partial decay widths for the Z boson and $q = c, b$, and N_F (N_B) indicates the number of events in which the final-state fermion is produced in the forward (backward) direction. Further details about the EWPO implementation can be found in [21] and references therein.

Since the goal of this benchmarking exercise is to carry out an independent validation of the results of [21] using the same theory and data inputs but now with the SMEFT code, as indicated above we take the SM and linear EFT cross-sections from the ATLAS note and parse them into the SMEFT format adopting the

same flavour assumptions for the fitting basis, namely $U(2)_q \times U(2)_u \times U(2)_d \times U(3)_\ell \times U(3)_e$. CP conservation is assumed and Wilson coefficients are real-valued. SM Higgs cross-sections are taken from the LHC Higgs WG [55], while LHC electroweak processes are computed using SHERPA2.2.2 [56], HERWIG7.1.5 [57], and VBFNLO3.0.0 [58] at NLO, matched to SHERPA and PYTHIA8 [59] parton showers respectively. Linear EFT cross-sections are computed with MADGRAPH5_AMC@NLO [60] and SMEFTsim [61, 62], except for loop-induced processes in the SM such as $gg \rightarrow h$ and $gg \rightarrow Zh$ where SMEFT@NLO [18] is used for the calculation of 1-loop QCD effects. An analytical computation with NLO accuracy in QED [63] is used for $H \rightarrow \gamma\gamma$. SMEFT propagator effects impacting the mass and width of intermediate particles are computed using SMEFTsim. The MADGRAPH5_AMC@NLO+PYTHIA8 predictions are supplemented with bin-by-bin K -factors to account for higher-order QCD and electroweak corrections. Theory predictions for EWPOs in the SM and the SMEFT follow [64] adapted to the flavour assumptions of [21].

The SMEFT predictions in the Warsaw basis for the processes entering the analysis of [21] depend on $n_{\text{op}} = 62$ independent Wilson coefficients. However, at the linear level only a subset of directions can be constrained from the input measurements, with the other linearly independent combinations leading to flat directions in the likelihood function $\mathcal{L}(\mathbf{c})$. Numerical minimizers such as that used in [21] can only work with problems where there exists a point solution in the parameter space, unless all strictly flat directions are removed. For this reason, the ATLAS fit of [21] is carried out not in the Warsaw basis but in a rotated basis, corresponding to the directions with the highest variability as determined by a principal component analysis of the matrix

$$\mathcal{H}_{i,j} \equiv \sum_{m,n=1}^{n_{\text{dat}}} \sigma_{\text{eft},i,m} (\text{cov}^{-1})_{m,n} \sigma_{\text{eft},j,n}, \quad (13)$$

which can be identified with the Fisher information matrix, Eq. (10). The PCA defines the rotation matrix $R_{ij}^{(W \rightarrow A)}$ that implements this basis transformation

$$c_i^{(A)} = \sum_{j=1}^{n_{\text{op}}=62} R_{ij}^{(W \rightarrow A)} c_j^{(W)}, \quad R^{(A \rightarrow W)} = \left(R^{(W \rightarrow A)} \right)^{-1}, \quad (14)$$

where (A) indicates the ATLAS fit basis and (W) the Warsaw basis, see Sect. 5.2 and Fig. 8 of [21] for the explicit definitions. The ATLAS analysis is then performed in terms of the 28 PCA eigenvectors $\mathbf{c}^{(A)}$ with the highest variability, with the remaining 34 linear combinations set to zero. Below we demonstrate how the results of [21] are also reproduced when the fit is carried out directly in the original 62-dimensional Warsaw basis $\mathbf{c}^{(W)}$ rather than in the PCA-rotated basis.

The left panel of Fig. 4 compares the ATLAS EFT fit results from [21] with the corresponding results obtained with the SMEFT code when the same theory, data inputs, and fitting basis are adopted. The outcome of the ATLAS analysis is provided in [21] both for the full likelihood and for a simplified multi-Gaussian likelihood; here we consider the latter to ensure a consistent comparison with the SMEFT results. The dark and pale lines represent the 68% and 95% CL intervals respectively. Since the EFT calculations include only linear cross-sections, the resulting posteriors are by construction Gaussian. In both cases, the fits have been carried out in the 28-dimensional PCA-rotated basis $\mathbf{c}^{(A)}$ defined by Eq. (14). The SMEFT output corresponds to Nested Sampling, though equivalent results are obtained with MCFit.

Inspection of Fig. 4 confirms that good agreement is obtained both in terms of central values and of the uncertainties of the fitted Wilson coefficients. Furthermore, similar agreement is obtained for the correlations ρ_{ij} between EFT coefficients, displayed in Fig. 5 in the PCA-rotated basis, as can be verified by comparing with the results from [21]. The fact that the entries of correlation matrix displayed in Fig. 5 are typically small, with few exceptions, is a consequence of using a rotated fit basis which by construction reduces the correlations between fitted degrees of freedom.

As discussed in Sect. 2, within the SMEFT framework it is possible to rotate from the Warsaw basis to any user-defined operator basis. In addition, the user can choose to automatically rotate to a fitting basis defined by the principal components of the Fisher information via Eq. (9), indicating the threshold restricting the kept singular values. While being numerically less efficient, the presence of flat directions does not represent a bottleneck in SMEFT when the Nested Sampling strategy is adopted. One can therefore combine these two functionalities to repeat the EFT interpretation displayed in the left panel of Fig. 4 now in the original 62-dimensional Warsaw basis $\mathbf{c}^{(W)}$, rather than in the 28-dimensional PCA-rotated basis

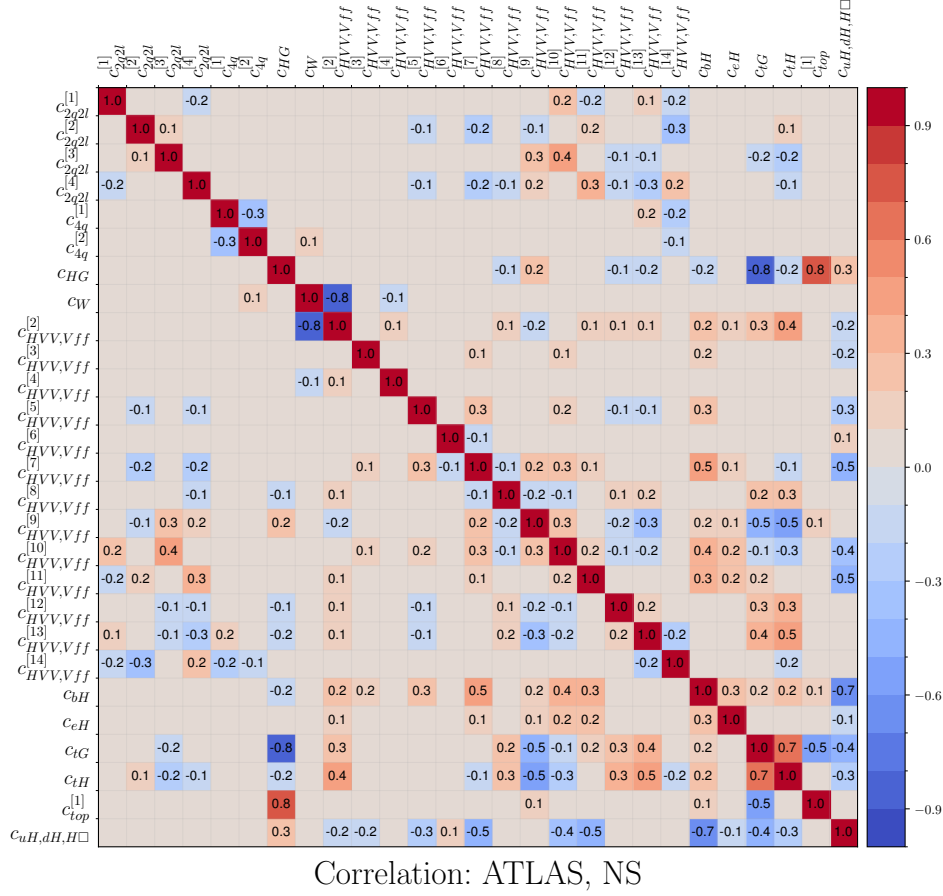


Figure 5. The correlation coefficients obtained in the SMEFT analysis reproducing [21] in the ATLAS fit basis. We do not display the numerical values of the correlation matrix entries with $|\rho_{ij}| < 0.1$.

applications. We have deliberately kept to a minimum the technical details, for instance concerning the format of the theory and data tables, and pointed the reader to the developing online documentation for further information. The features outlined in this paper represent only a snapshot of the full code capabilities, and in particular do not cover the extensive post-analysis visualization tools and statistical diagnosis methods provided with the code. While we have not discussed code performance, this is currently not a limiting factor for the analyses considered, with the global EFT fit (ATLAS EFT fit in Warsaw basis) in Fig. 2 (Fig. 4) taking around 2 (8) hours running on 24 cores.

This updated SMEFT framework will be the stepping stone making possible the realisation of a number of ongoing projects related to global interpretations of particle physics data in the SMEFT. To begin with, the implementation and (partial) automation of the matching between the SMEFT and UV-complete scenarios at the fit level, in a way that upon a choice of UV model, SMEFT returns the posterior distributions in the space of UV theory parameters such as heavy particle masses and couplings. This functionality is enabled by the SMEFT flexibility in imposing arbitrary restrictions between the Wilson coefficients, and will lead to the option of carrying out directly the fits in terms of the UV parameters, rather than in terms of the EFT coefficients. Second, to assess the impact in the global SMEFT fit of improved theory calculations such as the inclusion of renormalisation group running effects [65] and of electroweak corrections in high-energy observables. Third, to carry out projections quantifying the reach in the SMEFT parameter space [66, 67] of future lepton-lepton, lepton-hadron, and hadron-hadron colliders when these measurements are added on top of a state-of-the-art global fit. Fourth, extending the EFT determination to novel types of measurements beyond those based on a multi-Gaussian statistical model, such as the unbinned multivariate observables presented in [22]. Fifth, to validate related efforts such as the SIMUNET technique [68] developed to perform a simultaneous determination of the PDFs and EFT coefficients [69, 70], which should reduce to the SMEFT outcome in the fixed-PDF case for the same choice of theory and experimental inputs.

In addition to physics-motivated developments such as those outlined above, we also plan to further improve the statistical framework underlying SMEFT and expand the visualization and analysis post-

processing tools provided. One possible direction would be to implement new avenues to carry out parameter inference, such as the ML-assisted simulation-based inference method proposed in [71], as well as a broader range of optimisers for MCfit such as those studied in the benchmark comparison of [72]. It would also be advantageous to apply complementary methods to determine the more and less constrained directions in the parameter space. In particular, one could extend the linear PCA analysis to non-linear algorithms relevant to the case where the quadratic EFT corrections become sizable, such as with t-Distributed Stochastic Neighbor Embedding (t-SNE). Finally, one would like to run SMEFT in hardware accelerators such as multi graphics processing units (GPUs), leading to a further speed up of the code similar to that reported for PDF interpolations and event generators [73, 74].

The availability of this framework provides the SMEFT community with a new toolbox for all kinds of EFT interpretations of experimental data, with its modular structure facilitating the extension to other datasets and process types, updated theory calculations, and eventually its application to other EFTs such as the Higgs EFT. SMEFT will also streamline the comparisons and benchmarking between EFT determinations carried out by different groups, as the ATLAS analysis illustrates, and could be adopted by the experimental collaborations in order to cross-check the results obtained in their own frameworks.

Acknowledgments. We are grateful to Jaco ter Hoeve for testing the current version and for implementing the two-parameter contour feature in the fit report, and to Samuel van Beek, Jacob J. Ethier, Nathan. P. Hartland, Emanuele R. Nocera, and Emma Slade for their contributions to previous versions of the SMEFT framework. We thank Fabio Maltoni, Luca Mantani, Alejo Rossia, and Eleni Vryonidou for their inputs and suggestions concerning the SMEFT functionalities and the choice of formats for the data and theory tables. We are grateful to Rahul Balasubramanian, Lydia Brenner, Oliver Rieger, Wouter Verkerke, and Andrea Visibile for useful discussions about the ATLAS EFT Higgs analyses.

References

- [1] S. Weinberg, *Baryon and Lepton Nonconserving Processes*, *Phys. Rev. Lett.* **43** (1979) 1566–1570.
- [2] W. Buchmuller and D. Wyler, *Effective Lagrangian Analysis of New Interactions and Flavor Conservation*, *Nucl. Phys.* **B268** (1986) 621–653.
- [3] B. Grzadkowski, M. Iskrzynski, M. Misiak, and J. Rosiek, *Dimension-Six Terms in the Standard Model Lagrangian*, *JHEP* **10** (2010) 085, [[arXiv:1008.4884](#)].
- [4] N. P. Hartland, F. Maltoni, E. R. Nocera, J. Rojo, E. Slade, E. Vryonidou, and C. Zhang, *A Monte Carlo global analysis of the Standard Model Effective Field Theory: the top quark sector*, *JHEP* **04** (2019) 100, [[arXiv:1901.05965](#)].
- [5] I. Brivio, S. Bruggisser, F. Maltoni, R. Moutafis, T. Plehn, E. Vryonidou, S. Westhoff, and C. Zhang, *O new physics, where art thou? A global search in the top sector*, *JHEP* **02** (2020) 131, [[arXiv:1910.03606](#)].
- [6] A. Biekötter, T. Corbett, and T. Plehn, *The Gauge-Higgs Legacy of the LHC Run II*, *SciPost Phys.* **6** (2019) 064, [[arXiv:1812.07587](#)].
- [7] J. Ellis, C. W. Murphy, V. Sanz, and T. You, *Updated Global SMEFT Fit to Higgs, Diboson and Electroweak Data*, *JHEP* **06** (2018) 146, [[arXiv:1803.03252](#)].
- [8] E. da Silva Almeida, A. Alves, N. Rosa Agostinho, O. J. Éboli, and M. Gonzalez-Garcia, *Electroweak Sector Under Scrutiny: A Combined Analysis of LHC and Electroweak Precision Data*, *Phys. Rev. D* **99** (2019), no. 3 033001, [[arXiv:1812.01009](#)].
- [9] J. Aebischer, J. Kumar, P. Stangl, and D. M. Straub, *A Global Likelihood for Precision Constraints and Flavour Anomalies*, *Eur. Phys. J. C* **79** (2019), no. 6 509, [[arXiv:1810.07698](#)].
- [10] J. Ellis, M. Madigan, K. Mimasu, V. Sanz, and T. You, *Top, Higgs, Diboson and Electroweak Fit to the Standard Model Effective Field Theory*, *JHEP* **04** (2021) 279, [[arXiv:2012.02779](#)].

- [11] S. Bißmann, C. Grunwald, G. Hiller, and K. Kröninger, *Top and Beauty synergies in SMEFT-fits at present and future colliders*, *JHEP* **06** (2021) 010, [[arXiv:2012.10456](#)].
- [12] S. Bruggisser, R. Schäfer, D. van Dyk, and S. Westhoff, *The Flavor of UV Physics*, *JHEP* **05** (2021) 257, [[arXiv:2101.07273](#)].
- [13] S. Bruggisser, D. van Dyk, and S. Westhoff, *Resolving the Flavor Structure in the MFV-SMEFT*, [arXiv:2212.02532](#).
- [14] **SMEFT** Collaboration, J. J. Ethier, G. Magni, F. Maltoni, L. Mantani, E. R. Nocera, J. Rojo, E. Slade, E. Vryonidou, and C. Zhang, *Combined SMEFT interpretation of Higgs, diboson, and top quark data from the LHC*, *JHEP* **11** (2021) 089, [[arXiv:2105.00006](#)].
- [15] 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), no. 5 456, [[arXiv:1910.14012](#)].
- [16] N. Castro, J. Erdmann, C. Grunwald, K. Kröninger, and N.-A. Rosien, *EFTfitter—A tool for interpreting measurements in the context of effective field theories*, *Eur. Phys. J. C* **76** (2016), no. 8 432, [[arXiv:1605.05585](#)].
- [17] J. J. Ethier, R. Gomez-Ambrosio, G. Magni, and J. Rojo, *SMEFT analysis of vector boson scattering and diboson data from the LHC Run II*, *Eur. Phys. J. C* **81** (2021), no. 6 560, [[arXiv:2101.03180](#)].
- [18] C. Degrande, G. Durieux, F. Maltoni, K. Mimasu, E. Vryonidou, and C. Zhang, *Automated one-loop computations in the standard model effective field theory*, *Phys. Rev. D* **103** (2021), no. 9 096024, [[arXiv:2008.11743](#)].
- [19] S. van Beek, E. R. Nocera, J. Rojo, and E. Slade, *Constraining the SMEFT with Bayesian reweighting*, *SciPost Phys.* **7** (2019), no. 5 070, [[arXiv:1906.05296](#)].
- [20] N. Castro, K. Cranmer, A. V. Gritsan, J. Howarth, G. Magni, K. Mimasu, J. Rojo, J. Roskes, E. Vryonidou, and T. You, *LHC EFT WG Report: Experimental Measurements and Observables*, [arXiv:2211.08353](#).
- [21] **ATLAS** Collaboration, *Combined effective field theory interpretation of Higgs boson and weak boson production and decay with ATLAS data and electroweak precision observables*, tech. rep., CERN, Geneva, 2022. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2022-037>.
- [22] R. Gomez Ambrosio, J. ter Hoeve, M. Madigan, J. Rojo, and V. Sanz, *Unbinned multivariate observables for global SMEFT analyses from machine learning*, [arXiv:2211.02058](#).
- [23] F. Feroz, M. P. Hobson, E. Cameron, and A. N. Pettitt, *Importance Nested Sampling and the MultiNest Algorithm*, [arXiv:1306.2144](#).
- [24] F. Feroz and M. 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](#)].
- [25] **NNPDF** Collaboration, L. Del Debbio, S. Forte, J. I. Latorre, A. Piccione, and J. Rojo, *Unbiased determination of the proton structure function $F(2)^{**p}$ with faithful uncertainty estimation*, *JHEP* **03** (2005) 080, [[hep-ph/0501067](#)].
- [26] **The NNPDF** Collaboration, R. D. Ball et al., *A determination of parton distributions with faithful uncertainty estimation*, *Nucl. Phys.* **B809** (2009) 1–63, [[arXiv:0808.1231](#)].
- [27] N. Hansen and A. Ostermeier, *Completely derandomized self-adaptation in evolution strategies*, *Evolutionary Computation* **9** (2001), no. 2 159–195, [<https://doi.org/10.1162/106365601750190398>].
- [28] **NNPDF** Collaboration, V. Bertone, S. Carrazza, N. P. Hartland, E. R. Nocera, and J. Rojo, *A determination of the fragmentation functions of pions, kaons, and protons with faithful uncertainties*, *Eur. Phys. J. C* **77** (2017), no. 8 516, [[arXiv:1706.07049](#)].

- [29] R. H. Byrd, M. E. Hribar, and J. Nocedal, *An interior point algorithm for large-scale nonlinear programming*, *SIAM J. Optim.* **9** (1999) 877–900.
- [30] Y. Xiang, D. Y. Sun, W. Fan, and X. G. Gong, *Generalized simulated annealing algorithm and its application to the Thomson model*, *Physics Letters A* **233** (Feb., 1997) 216–220.
- [31] **NNPDF** Collaboration, R. Abdul Khalek et al., *A first determination of parton distributions with theoretical uncertainties*, *Eur. Phys. J. C* (2019) 79:838, [[arXiv:1905.04311](#)].
- [32] **NNPDF** Collaboration, R. Abdul Khalek et al., *Parton Distributions with Theory Uncertainties: General Formalism and First Phenomenological Studies*, *Eur. Phys. J. C* **79** (2019), no. 11 931, [[arXiv:1906.10698](#)].
- [33] D. Barducci et al., *Interpreting top-quark LHC measurements in the standard-model effective field theory*, [arXiv:1802.07237](#).
- [34] R. D. Ball, V. Bertone, F. Cerutti, L. Del Debbio, S. Forte, et al., *Reweighting and Unweighting of Parton Distributions and the LHC W lepton asymmetry data*, *Nucl.Phys.* **B855** (2012) 608–638, [[arXiv:1108.1758](#)].
- [35] Z. Kassabov, M. Madigan, L. Mantani, J. Moore, M. M. Alvarado, J. Rojo, and M. Ubiali, *The top quark legacy of the LHC Run II for PDF and SMEFT analyses*, [arXiv:2303.06159](#).
- [36] **ATLAS** Collaboration, *Combined measurements of Higgs boson production and decay using up to 139 fb^{-1} of proton-proton collision data at $\sqrt{s} = 13$ TeV collected with the ATLAS experiment*, tech. rep., CERN, Geneva, 2021. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2021-053>.
- [37] **ATLAS** Collaboration, *Combined effective field theory interpretation of differential cross-sections measurements of WW , WZ , 4ℓ , and Z -plus-two-jets production using ATLAS data*, tech. rep., CERN, Geneva, 2021. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2021-022>.
- [38] **ATLAS** Collaboration, *Combined measurements of Higgs boson production and decay using up to 139 fb^{-1} of proton-proton collision data at $\sqrt{s} = 13$ TeV collected with the ATLAS experiment*, .
- [39] **ATLAS** Collaboration, *Combined effective field theory interpretation of differential cross-sections measurements of WW , WZ , 4ℓ , and Z -plus-two-jets production using ATLAS data*, .
- [40] N. Berger et al., *Simplified Template Cross Sections - Stage 1.1*, [arXiv:1906.02754](#).
- [41] **ATLAS** Collaboration, *Measurement of the properties of Higgs boson production at $\sqrt{s}=13$ TeV in the $H \rightarrow \gamma\gamma$ channel using 139 fb^{-1} of pp collision data with the ATLAS experiment*, tech. rep., CERN, Geneva, 2020. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2020-026>.
- [42] **ATLAS** Collaboration, G. Aad et al., *Higgs boson production cross-section measurements and their EFT interpretation in the 4ℓ decay channel at $\sqrt{s}=13$ TeV with the ATLAS detector*, *Eur. Phys. J. C* **80** (2020), no. 10 957, [[arXiv:2004.03447](#)]. [Erratum: *Eur.Phys.J.C* 81, 29 (2021), Erratum: *Eur.Phys.J.C* 81, 398 (2021)].
- [43] **ATLAS** Collaboration, *Measurements of gluon fusion and vector-boson-fusion production of the Higgs boson in $H \rightarrow WW^* \rightarrow e\nu\mu\nu$ decays using pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, tech. rep., CERN, Geneva, 2021. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2021-014>.
- [44] **ATLAS** Collaboration, *Measurements of Higgs boson production cross-sections in the $H \rightarrow \tau^+\tau^-$ decay channel in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, tech. rep., CERN, Geneva, 2021. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2021-044>.

- [45] **ATLAS** Collaboration, G. Aad et al., *Measurements of WH and ZH production in the $H \rightarrow b\bar{b}$ decay channel in pp collisions at 13 TeV with the ATLAS detector*, *Eur. Phys. J. C* **81** (2021), no. 2 178, [[arXiv:2007.02873](#)].
- [46] **ATLAS** Collaboration, G. Aad et al., *Measurement of the associated production of a Higgs boson decaying into b -quarks with a vector boson at high transverse momentum in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *Phys. Lett. B* **816** (2021) 136204, [[arXiv:2008.02508](#)].
- [47] **ATLAS** Collaboration, *Combination of measurements of Higgs boson production in association with a W or Z boson in the $b\bar{b}$ decay channel with the ATLAS experiment at $\sqrt{s} = 13$ TeV*, tech. rep., CERN, Geneva, 2021. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2021-051>.
- [48] **ATLAS** Collaboration, G. Aad et al., *Measurements of Higgs bosons decaying to bottom quarks from vector boson fusion production with the ATLAS experiment at $\sqrt{s} = 13$ TeV*, *Eur. Phys. J. C* **81** (2021), no. 6 537, [[arXiv:2011.08280](#)].
- [49] **ATLAS** Collaboration, G. Aad et al., *Measurement of Higgs boson decay into b -quarks in associated production with a top-quark pair in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *JHEP* **06** (2022) 097, [[arXiv:2111.06712](#)].
- [50] **ATLAS** Collaboration, M. Aaboud et al., *Measurement of fiducial and differential W^+W^- production cross-sections at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *Eur. Phys. J. C* **79** (2019), no. 10 884, [[arXiv:1905.04242](#)].
- [51] **ATLAS** Collaboration, M. Aaboud et al., *Measurement of $W^\pm Z$ production cross sections and gauge boson polarisation in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *Eur. Phys. J. C* **79** (2019), no. 6 535, [[arXiv:1902.05759](#)].
- [52] **ATLAS** Collaboration, G. Aad et al., *Measurements of differential cross-sections in four-lepton events in 13 TeV proton-proton collisions with the ATLAS detector*, *JHEP* **07** (2021) 005, [[arXiv:2103.01918](#)].
- [53] **ATLAS** Collaboration, G. Aad et al., *Differential cross-section measurements for the electroweak production of dijets in association with a Z boson in proton-proton collisions at ATLAS*, *Eur. Phys. J. C* **81** (2021), no. 2 163, [[arXiv:2006.15458](#)].
- [54] **ALEPH, DELPHI, L3, OPAL, SLD, LEP Electroweak Working Group, SLD Electroweak Group, SLD Heavy Flavour Group** Collaboration, S. Schael et al., *Precision electroweak measurements on the Z resonance*, *Phys. Rept.* **427** (2006) 257–454, [[hep-ex/0509008](#)].
- [55] **LHC Higgs Cross Section Working Group** Collaboration, D. de Florian et al., *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, [arXiv:1610.07922](#).
- [56] **Sherpa** Collaboration, E. Bothmann et al., *Event Generation with Sherpa 2.2*, *SciPost Phys.* **7** (2019), no. 3 034, [[arXiv:1905.09127](#)].
- [57] M. Bähr et al., *Herwig++ Physics and Manual*, *Eur. Phys. J. C* **58** (2008) 639–707, [[arXiv:0803.0883](#)].
- [58] J. Baglio et al., *VBFNLO: A Parton Level Monte Carlo for Processes with Electroweak Bosons – Manual for Version 2.7.0*, [arXiv:1107.4038](#).
- [59] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, *An introduction to PYTHIA 8.2*, *Comput. Phys. Commun.* **191** (2015) 159–177, [[arXiv:1410.3012](#)].
- [60] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer, and T. Stelzer, *MadGraph 5 : Going Beyond*, *JHEP* **06** (2011) 128, [[arXiv:1106.0522](#)].
- [61] I. Brivio, Y. Jiang, and M. Trott, *The SMEFTsim package, theory and tools*, *JHEP* **12** (2017) 070, [[arXiv:1709.06492](#)].

- [62] I. Brivio, *SMEFTsim 3.0 — a practical guide*, *JHEP* **04** (2021) 073, [[arXiv:2012.11343](#)].
- [63] S. Dawson and P. P. Giardino, *Electroweak corrections to Higgs boson decays to $\gamma\gamma$ and W^+W^- in standard model EFT*, *Phys. Rev. D* **98** (2018), no. 9 095005, [[arXiv:1807.11504](#)].
- [64] T. Corbett, A. Helset, A. Martin, and M. Trott, *EWPD in the SMEFT to dimension eight*, *JHEP* **06** (2021) 076, [[arXiv:2102.02819](#)].
- [65] R. Aoude, F. Maltoni, O. Mattelaer, C. Severi, and E. Vryonidou, *Renormalisation group effects on SMEFT interpretations of LHC data*, [arXiv:2212.05067](#).
- [66] J. de Blas, Y. Du, C. Grojean, J. Gu, V. Miralles, M. E. Peskin, J. Tian, M. Vos, and E. Vryonidou, *Global SMEFT Fits at Future Colliders*, in *2022 Snowmass Summer Study*, 6, 2022. [arXiv:2206.08326](#).
- [67] J. de Blas et al., *Higgs Boson Studies at Future Particle Colliders*, *JHEP* **01** (2020) 139, [[arXiv:1905.03764](#)].
- [68] S. Iranipour and M. Ubiali, *A new generation of simultaneous fits to LHC data using deep learning*, *JHEP* **05** (2022) 032, [[arXiv:2201.07240](#)].
- [69] S. Carrazza, C. Degrande, S. Iranipour, J. Rojo, and M. Ubiali, *Can New Physics hide inside the proton?*, *Phys. Rev. Lett.* **123** (2019), no. 13 132001, [[arXiv:1905.05215](#)].
- [70] A. Greljo, S. Iranipour, Z. Kassabov, M. Madigan, J. Moore, J. Rojo, M. Ubiali, and C. Voisey, *Parton distributions in the SMEFT from high-energy Drell-Yan tails*, *JHEP* **07** (2021) 122, [[arXiv:2104.02723](#)].
- [71] B. K. Miller, A. Cole, G. Louppe, and C. Weniger, *Simulation-efficient marginal posterior estimation with swyft: stop wasting your precious time*, [arXiv:2011.13951](#).
- [72] **DarkMachines High Dimensional Sampling Group** Collaboration, C. Balázs et al., *A comparison of optimisation algorithms for high-dimensional particle and astrophysics applications*, *JHEP* **05** (2021) 108, [[arXiv:2101.04525](#)].
- [73] S. Carrazza, J. Cruz-Martinez, M. Rossi, and M. Zaro, *MadFlow: automating Monte Carlo simulation on GPU for particle physics processes*, *Eur. Phys. J. C* **81** (2021), no. 7 656, [[arXiv:2106.10279](#)].
- [74] S. Carrazza, J. M. Cruz-Martinez, and M. Rossi, *PDFFlow: Parton distribution functions on GPU*, *Comput. Phys. Commun.* **264** (2021) 107995, [[arXiv:2009.06635](#)].