Unbinning global LHC analyses

Henning Bahl[1], Tilman Plehn[1,2], and Nikita Schmal[1]

1 Institut für Theoretische Physik, Universität Heidelberg, Germany
2 Interdisciplinary Center for Scientific Computing (IWR), Universität Heidelberg, Germany

October 16, 2025

Abstract

Simulation-based inference has been shown to outperform traditional, histogram-based inference in numerous phenomenological and experimental studies. So far, these analyses have focused on individual high-profile processes. We study the combination of four different di-boson processes in terms of the Standard Model Effective Field Theory. Our results demonstrate how simulation-based inference also wins over traditional methods for global LHC analyses.

Contents

1 Introduction **2**

2 Methodology **2**
2.1 Learning the likelihood ratio 3
2.2 Backgrounds 4
2.3 Fractional smearing 6
2.4 Limit setting and empirical coverage 6

3 Processes, operators, and training setup **7**
3.1 SMEFT operators 7
3.2 Di-boson processes 8
3.3 Neural network setup and training 10

4 Results **11**
4.1 Single-process limits 11
4.2 Combined limits 16

5 Conclusions **18**

A Relevant SMEFT Feynman rules **20**

B Likelihood coverages **21**

C Single parameter results **25**

References **27**

---

1 Introduction

One of the biggest upcoming challenges in particle physics is to exploit the full potential of the current and future LHC datasets. Since the Higgs discovery, the LHC has been turned into the first precision-hadron collider in history, but extracting the full information from high-dimensional event data is becoming an increasingly problematic bottleneck. Traditional inference methods rely on, at most, few-dimensional histograms as summary statistics. This reduction of dimensionality and the binning result in a loss of information. Modern machine learning (ML) can overcome this bottleneck [1, 2]. In particular, they can be used to extract statistically optimal unbinned likelihood ratios [3].

Besides the ML-based matrix-element method [4, 5], simulation-based inference (SBI) [6–15] is a proven method for likelihood ratio estimation. The advantages of SBI for high-profile LHC processes have been demonstrated not only in a variety of phenomenological studies [10, 11, 13, 16–23], but recently also in the first experimental SBI analysis at the LHC [24, 25]. While SBI already improves the sensitivity to individual theory parameters, it has even greater potential if multiple theory parameters are to be inferred. Since SBI does not rely on lower-dimensional summary statistics, it is much more effective in disentangling different theory parameters than binned histograms.

A key aspect of modern LHC physics is the combination of many processes into a global analysis, to answer whether LHC data is explained by the Standard Model everywhere. The standard framework for such global analyses is the Standard Model Effective Field Theory (SMEFT) [26, 27]. Using SMEFT, numerous global LHC analyses have demonstrated that much stronger constraints on fundamental parameters can be obtained from the combination of measurements and processes [28–36]. These analyses are currently based on total rate or binned differential rate measurements. While it is methodologically clear how to integrate SBI into a global SMEFT analysis [19], it is not clear to what level the advantages of SBI persist for such global analyses.

We aim to quantify this effect for four different di-boson processes: WW, WZ, WH, and ZH production. These processes are highly relevant in global SMEFT analyses of the combined electroweak and Higgs sector [28, 37]: WW and WZ production constrain the resulting anomalous triple gauge couplings and purely electroweak operators, while WH and ZH are sensitive to Higgs–gauge interactions and add fermionic contact terms. Focusing on a subset of SMEFT operators, to which the di-boson processes are dominantly sensitive, we compare the sensitivity of the SBI and histogram methods in detail.

The paper is structured as follows. In Sec. 2, we describe the SBI methods used. Section 3 provides an overview of the considered processes and SMEFT operators. We present results in Sec. 4. Conclusions can be found in Sec. 5.

2 Methodology

We start by describing our methodology including learning the likelihood ratio using the derivative-learning approach, backgrounds, fractional smearing, and limit setting.

2.1 Learning the likelihood ratio

At parton level, the likelihood of the observables z_p given a parameter value θ is directly related to the differential cross section,

$$p(z_p|\theta) = \frac{1}{\sigma(\theta)} \frac{\mathrm{d}\sigma(z_p|\theta)}{\mathrm{d}z_p} \,. \tag{1}$$

The differential cross-section can be computed via the matrix element $\mathcal{M}$ and the parton distribution functions $f_{1,2}$,

$$d\sigma(z_p|\theta) = (2\pi)^4 \int dx_1 dx_2 d\Phi \frac{f_1(x_1,Q^2) f_2(x_2,Q^2)}{2x_1x_2s} |\mathcal{M}(z_p|\theta)|^2 \,. \tag{2}$$

Assuming that the theory parameters θ only affect the hard scattering element, the reconstruction-level likelihood is given by [6]

$$\begin{aligned} p(x|\theta) &= \int dz_d dz_s dz_p p(x|z_d) p(z_d|z_s) p(z_s|z_p) p(z_p|\theta) \\ &= \int dz_p p(x|z_p) p(z_p|\theta) \,. \end{aligned} \tag{3}$$

Here, $p(z_s|z_p)$ encodes the parton shower and hadronisation; $p(z_d|z_s)$, the detector response; and, $p(x|x_d)$, the reconstruction of the reconstruction-level observables x.

The integral in Eq.(3) is not tractable, implying that the $p(x|\theta)$ cannot be directly computed. We can, however, use neural networks to learn the likelihood ratio assuming the factorized form,

$$r(x|\theta,\theta_0) \equiv \frac{p(x|\theta)}{p(x|\theta_0)} \,. \tag{4}$$

This can be achieved using the loss function

$$\mathcal{L} = \left\langle \left[r(z_p|\theta,\theta_0) - r_\varphi(x|\theta,\theta_0) \right]^2 \right\rangle_{x,z_p \sim p(x|z_p)p(z_p|\theta);\theta\sim q(\theta)} \,, \tag{5}$$

where r_φ is the network output and $r(z_p|\theta,\theta_0)$ is the tractable parton-level likelihood ratio. The squared difference is averaged over combined parton-level and reconstruction-level Monte-Carlo samples with the theory parameters sampled from a prior $q(\theta)$.

Learning the likelihood ratio conditioned on the theory parameters is difficult and often numerically unstable. This can be avoided by exploiting the known dependencies of the likelihood ratio on the theory parameters. It is particularly useful for SMEFT, whose Lagrangian has the structure

$$\mathscr{L}_{\text{SMEFT}} = \mathscr{L}_{\text{SM}} + \sum_i \frac{c_i}{\Lambda^2} O_i \equiv \mathscr{L}_{\text{SM}} + \sum_i \theta_i O_i \,. \tag{6}$$

Expanding the squared matrix element to second order in the θ, the differential cross-section ratio

$$R(x|\theta,\theta_0) \equiv \frac{d\sigma(x|\theta)/dx}{d\sigma(x|\theta_0)/dx} = \frac{\sigma(\theta)p(x|\theta)}{\sigma(\theta_0)p(x|\theta_0)} \,, \tag{7}$$

can be expanded up to the same second order

$$\begin{aligned} R(x|\theta,\theta_0) &= 1 + (\theta-\theta_0)_i R_i(x) + (\theta-\theta_0)_i(\theta-\theta_0)_j R_{ij}(x) \\ R_i(x) &\equiv \frac{\partial}{\partial\theta_i} R(x|\theta,\theta_0)\Big|_{\theta=\theta_0} \\ R_{ij}(x) &\equiv \frac{\partial^2}{\partial\theta_i\partial\theta_j} R(x|\theta,\theta_0)\Big|_{\theta=\theta_0} . \end{aligned} \tag{8}$$

As in Eq.(5), we can learn these coefficients independently of each other using the parton-level ratios [10, 11]

$$\begin{aligned} R_i(z_p) &\equiv \frac{\partial}{\partial\theta_i}\frac{d\sigma(z_p|\theta)/dz_p}{d\sigma(z_p|\theta_0)/dz_p}\Big|_{\theta=\theta_0} = \frac{\partial_{\theta_i}|\mathcal{M}(z_p|\theta)|^2}{|\mathcal{M}(z_p|\theta_0)|^2}\Big|_{\theta_0} \\ R_{ij}(z_p) &\equiv \frac{\partial^2}{\partial\theta_i\partial\theta_j}\frac{d\sigma(z_p|\theta)/dz_p}{d\sigma(z_p|\theta_0)/dz_p}\Big|_{\theta=\theta_0} = \frac{\partial_{\theta_i}\partial_{\theta_j}|\mathcal{M}(z_p|\theta)|^2}{|\mathcal{M}(z_p|\theta)|^2}\Big|_{\theta_0} . \end{aligned} \tag{9}$$

The neural network estimators R_φ approximate the true ratios $R_i(x)$ and $R_{ij}(x)$,

$$\begin{aligned} R_{\varphi,i}(x) &\approx R_i(x) \\ R_{\varphi,ij}(x) &\approx R_{ij}(x) . \end{aligned} \tag{10}$$

The likelihood ratio is then obtained via

$$r(x|\theta,\theta_0) = \frac{\sigma(\theta_0)}{\sigma(\theta)}\frac{d\sigma(x|\theta)/dx}{d\sigma(x|\theta_0)/dx} = \frac{\sigma(\theta_0)}{\sigma(\theta)} R(x|\theta,\theta_0) . \tag{11}$$

We refer to this approach as derivative learning.

Alternatively, the likelihood ratio can be learned at specific benchmark points. The theory parameter dependence is then recovered by morphing the likelihood between these benchmark points [6,14]. In this work, we only use derivative learning as we previously found it to be more stable if the phase-space regions populated by the SM and BSM hypotheses have significant overlap [14].

2.2 Backgrounds

In the presence of background, we can further split up the squared matrix element into a signal, a background, and an interference component,

$$|\mathcal{M}(z_p|\theta)|^2 = |\mathcal{M}_{\text{sig}}(z_p|\theta)|^2 + 2\text{Re}\Big[\mathcal{M}_{\text{sig}}(z_p|\theta)\mathcal{M}^*_{\text{bkg}}(z_p)\Big] + |\mathcal{M}_{\text{bkg}}(z_p)|^2 . \tag{12}$$

For the processes considered in this work, the interference contribution is negligible — either due to different initial/final states or to the small Higgs width. In this case, the differential cross section can be written as the sum

$$d\sigma(x|\theta) = d\sigma_{\text{sig}}(x|\theta) + d\sigma_{\text{bkg}}(x) . \tag{13}$$

Correspondingly, the likelihood and likelihood ratio split into

$$p(x|\theta) = \frac{\sigma_{\text{sig}}(\theta)}{\sigma_{\text{sig}}(\theta)+\sigma_{\text{bkg}}} p_{\text{sig}}(x|\theta) + \frac{\sigma_{\text{bkg}}}{\sigma_{\text{sig}}(\theta)+\sigma_{\text{bkg}}} p_{\text{bkg}}(x)$$

$$r(x|\theta,\theta_0) = \frac{\sigma_{\text{sig}}(\theta_0)+\sigma_{\text{bkg}}}{\sigma_{\text{sig}}(\theta)+\sigma_{\text{bkg}}} \frac{1+\dfrac{\sigma_{\text{sig}}(\theta_0)}{\sigma_{\text{bkg}}}\omega(x|\theta_0) R(x|\theta,\theta_0)}{1+\dfrac{\sigma_{\text{sig}}(\theta_0)}{\sigma_{\text{bkg}}}\omega(x|\theta_0)}$$

$$\text{with} \qquad \omega(x|\theta_0) \equiv \frac{p_{\text{sig}}(x|\theta_0)}{p_{\text{bkg}}(x)} \,. \tag{14}$$

The likelihood ratio can be extracted through a signal–background classifier, whose classifier score D converges towards [21,38]

$$D_{\text{opt}}(x) = \frac{p_{\text{sig}}(x|\theta_0)}{p_{\text{sig}}(x|\theta_0)+p_{\text{bkg}}(x))} = \frac{\omega(x)}{1+\omega(x)} \qquad \Leftrightarrow \qquad \omega(x) = \frac{D_{\text{opt}}(x)}{1-D_{\text{opt}}(x)} \,. \tag{15}$$

The first derivative of the summed log-likelihood ratio is required to vanish at $\theta = \theta_0$. This can be derived analytically:

$$\begin{aligned}
&\frac{\partial}{\partial\theta_i}\left[\int dx\, p(x|\theta_0)\log r(x|\theta,\theta_0)\right]_{\theta_i=\theta_{0,i}} \\
&= \int \mathrm{d}x\, p(x|\theta_0) \frac{R_j(x)}{1+\frac{\sigma_{\text{bkg}}}{\sigma_{\text{sig}}(\theta_0)}\frac{1}{\omega(x|\theta_0)}} - \frac{\sigma_{\text{sig},j}(\theta_0)}{\sigma_{\text{sig}}(\theta_0)+\sigma_{\text{bkg}}} \\
&= \frac{\sigma_{\text{sig}}(\theta_0)}{\sigma_{\text{sig}}(\theta_0)+\sigma_{\text{bkg}}} \int \mathrm{d}x p_{\text{sig}}(x|\theta_0) \frac{\partial}{\partial\theta_j}\frac{\mathrm{d}\sigma_{\text{sig}}(x|\theta)}{\mathrm{d}\sigma_{\text{sig}}(x|\theta_0)} - \frac{\sigma_{\text{sig},j}(\theta_0)}{\sigma_{\text{sig}}(\theta_0)+\sigma_{\text{bkg}}} \\
&= \frac{1}{\sigma_{\text{sig}}(\theta_0)+\sigma_{\text{bkg}}} \frac{\partial}{\partial\theta_j} \int \mathrm{d}x \frac{\mathrm{d}\sigma_{\text{sig}}(x|\theta)}{\mathrm{d}x} - \frac{\sigma_{\text{sig},j}(\theta_0)}{\sigma_{\text{sig}}(\theta_0)+\sigma_{\text{bkg}}} \\
&= 0 \,,
\end{aligned} \tag{16}$$

where we introduced

$$\sigma_{\text{sig},j}(\theta_0) = \frac{\partial}{\partial\theta_j}\sigma_{\text{sig}}(\theta)\Big|_{\theta=\theta_0} \,. \tag{17}$$

If the classifier and the differential cross-section ratios are not learned perfectly or if the data samples are not large enough, the numerical cancellation between the first and second terms above might not be perfect. The likelihood will then not have its minimum at θ_0, even using a large dataset generated for $\theta = \theta_0$. To avoid this numerical issue, we set

$$\sigma_{\text{sig,j}} = \big(\sigma_{\text{sig}}(\theta_0)+\sigma_{\text{bkg}}\big) \left\langle \frac{R_{\varphi,j}(x_i)}{1+\frac{\sigma_{\text{bkg}}}{\sigma_{\text{sig}}(\theta_0)}\frac{1}{\omega(x_i|\theta_0)}} \right\rangle_{x\sim p(x|\theta_0)} \,, \tag{18}$$

when evaluating the kinematic part of the likelihood. This extends the procedure proposed in Ref. [14] for the background-free case.

2.3 Fractional smearing

Another numerical problem arises from outlier events associated with significantly larger derivatives R_i or R_{ij} than the bulk of the distribution. These events are passed through the parton shower and detector simulation only once, and the learned estimators can be biased. As proposed in Ref. [14], we use fractional smearing to avoid this issue. We pass each event with a large derivative through the parton shower and detector simulation n times and assign each resulting event a weight $1/n$, where n is chosen based on the size of the derivatives. These weights are then incorporated into the MSE loss. For the example of likelihood ratio regression for a fixed θ, this reads

$$\begin{aligned}\mathcal{L} &= \left\langle \left[r(z_p|\theta,\theta_0) - r_\varphi(x|\theta) \right]^2 \right\rangle_{x,z_p\sim p(x|z_p)p(z_p|\theta)} \\ &= \sum_i^N w_i \left[r(z_{p,i}|\theta,\theta_0) - r_\varphi(x_i|\theta) \right]^2 . \end{aligned} \tag{19}$$

2.4 Limit setting and empirical coverage

The full likelihood for a set of events $\{x\}$ is given by

$$p_{\text{full}}(\{x\}|\theta) = \text{Pois}(n|L\sigma(\theta)) \prod_i p(x_i|\theta) \qquad \text{with} \qquad \text{Pois}(k|\lambda) = e^{-\lambda}\frac{\lambda^k}{k!} . \tag{20}$$

Given the luminosity L, $\text{Pois}(n|L\sigma(\theta))$ is the total rate likelihood. The unbinned kinematic likelihood for each event is $p(x_i|\theta)$ [39]. The corresponding likelihood ratio reads

$$\log r_{\text{full}}(\{x\}|\theta,\theta_0) = \log \frac{\text{Pois}(n|L\sigma(\theta))}{\text{Pois}(n|L\sigma(\theta_0))} + \sum_i \log r(x_i|\theta,\theta_0) . \tag{21}$$

To derive expected limits, we replace the sum by

$$\sum_i \log r(x_i|\theta,\theta_0) \;\to\; \frac{N_{\text{exp}}}{N_{\text{MC}}} \left\langle \log r(x_i|\theta,\theta_0) \right\rangle_{x_i\in\{x\}_{\text{MC}}} , \tag{22}$$

to exploit the full Monte-Carlo sample of size N_{MC}. In the presence of background, we can split this into

$$\begin{aligned}\frac{N_{\text{exp}}}{N_{\text{MC}}} \left\langle \log r(x_i|\theta,\theta_0) \right\rangle_{x_i\in\{x\}_{\text{MC}}} \to{}& \frac{N_{\text{bkg,exp}}}{N_{\text{bkg,MC}}} \left\langle \log r(x_i|\theta,\theta_0) \right\rangle_{x_i\in\{x\}_{\text{MC,bkg}}} \\ &+ \frac{N_{\text{sig,exp}}}{N_{\text{sig,MC}}} \left\langle \log r(x_i|\theta,\theta_0) \right\rangle_{x_i\in\{x\}_{\text{MC,sig}}} \end{aligned} \tag{23}$$

using separate samples $\{x\}_{\text{bkg}}$ and $\{x\}_{\text{sig}}$ for background and signal, respectively. This is valid since background and signal samples are statistically independent. For multiple independent background processes, further splits are possible.

After deriving the full likelihood, the test statistic is given by

$$\begin{aligned} q(\theta) &= -2\log r_{\text{full}}(\{x\}|\theta,\hat{\theta}) \\ &= -2\left(\log r_{\text{full}}(\{x\}|\theta,\theta_0) - \log r_{\text{full}}(\{x\}|\hat{\theta},\theta_0)\right) , \end{aligned} \tag{24}$$

where $\hat{\theta}$ is the minimum of the likelihood. We approximate it by using

$$\hat{\theta} = \underset{\theta}{\text{argmax}} \log r_{\text{full}}(\{x\}|\theta, \theta_1); . \tag{25}$$

Based on Wilk's theorem, the distribution $p(q(\theta)|\theta)$ converges towards a chi-squared distribution for sufficiently many events. Based on this, we calculate the p-value for a parameter point θ via

$$p_\theta = \int_{q_{\text{obs}}(\theta)}^{\infty} dq\, p(q(\theta)|\theta) = 1 - F_{\chi^2}(q_{\text{obs}}(\theta)|k)\,. \tag{26}$$

It gives the confidence with which one can reject the parameter point θ, where $q_{\text{obs}}(\theta)$ is the observed value of $q(\theta)$ for the sample $\{x\}$. $F_{\chi^2}(y|k)$ is the cumulative chi-squared distribution function with k degrees of freedom. The γ confidence region is defined by all θ values for which $p_\theta < \gamma$.

As a cross-check for the learned likelihood ratio $r_\varphi(x|\theta, \theta_0)$, we take n samples for a given θ and evaluate for what fraction of samples the true θ lies within a given γ confidence region, defining the coverage

$$c_\gamma \equiv \Big\langle \mathbb{1}\left(p_{\theta_0}(\{x\}) > 1-\gamma\right) \Big\rangle_{\{x\}_{1,\dots,n}}\,. \tag{27}$$

Here, we use the indicator function $\mathbb{1}$ which is equal to one if the expression in the brackets is fulfilled and zero otherwise. If the fraction is higher than the nominal confidence level, $c_\gamma > \gamma$, our learned likelihood is conservative or underconfident. Inversely, if it is lower the learned likelihood is overconfident.

3 Processes, operators, and training setup

In our analysis we focus on four di-boson processes — WZ, WW, ZH, and WH production — and a selected subset of SMEFT operators.

3.1 SMEFT operators

Throughout our analysis, we neglect flavor- or $\mathcal{CP}$-violating operators. We, moreover, restrict ourselves to dimension-six operators. A list of all operators considered is given in Table 1. In this analysis we ignore the fact that they are also constrained by electroweak precision observables [30, 31, 36]. We list the relevant SMEFT Feynman rules in App. A.

Table 2 details which operators affect which process. While WH production is only affected by three of the considered operators, five operators contribute to ZH production. At first sight, this larger set of relevant operators significantly increases the complexity of inferring the likelihood ratio. In practice, however, the expansion performed in Eq.(8) allows for a straightforward inclusion of more operators.

operator	definition	operator	definition
$\mathcal{O}_{\Phi D}$	$(\Phi^\dagger D^\mu \Phi)^*(\Phi^\dagger D_\mu \Phi)$	$\mathcal{O}_{\Phi q}^{(3)}$	$\sum_{i=1,2}(\Phi^\dagger i\overleftrightarrow{D^a}_\mu \Phi)(\bar{q}_i \sigma^a \gamma^\mu q_i)$
$\mathcal{O}_{\Phi B}$	$\Phi^\dagger \Phi B_{\mu\nu} B^{\mu\nu}$	$\mathcal{O}_{\Phi W}$	$\Phi^\dagger \Phi W^a_{\mu\nu} W^{\mu\nu a}$
$\mathcal{O}_{\Phi WB}$	$\Phi^\dagger \sigma^a \Phi W^a_{\mu\nu} B^{\mu\nu}$	$\mathcal{O}_{WWW}$	$\epsilon^{abc} W^{\nu a}_\mu W^{\rho b}_\nu W^{\mu c}_\rho$

Table 1: Dimension-6 SMEFT operators considered in our analysis.

process	$\mathcal{O}_{\Phi D}$	$\mathcal{O}_{\Phi W}$	$\mathcal{O}_{\Phi B}$	$\mathcal{O}_{\Phi WB}$	$\mathcal{O}_{\Phi q}^{(3)}$	$\mathcal{O}_{WWW}$
WZ	✓			✓	✓	✓
WW	✓			✓	✓	✓
ZH	✓	✓	✓	✓	✓	
WH	✓	✓			✓	

Table 2: Operators contributing to the different di-boson processes.

3.2 Di-boson processes

In our analysis we focus on four di-boson production processes: $qq \to WZ$, $qq \to WW$, $qq \to WH$, and $qq \to ZH$ production. We neglect the subleading loop-induced $gg \to ZH$ contribution to ZH production. Regarding the boson decay channels, we consider the leptonic decay channels of the W and Z boson and the $H \to b\bar{b}$ decay channel. The resulting signatures are

$$\begin{aligned} qq' &\to W^{\pm}Z \to \ell^{\pm}\ell^{+}\ell^{-} + E_{T,\text{miss}} \\ q\bar{q} &\to W^{\pm}W^{\mp} \to \ell^{+}\ell^{-} + E_{T,\text{miss}} \\ q\bar{q} &\to ZH \to \ell^{+}\ell^{-} + b\bar{b} \\ qq' &\to W^{\pm}H \to \ell^{\pm} + b\bar{b} + E_{T,\text{miss}} \,. \end{aligned} \tag{28}$$

Pre-selection cuts and backgrounds

We impose a series of pre-selection cuts, inspired by experimental analyses [40–44]. These cuts are chosen to suppress backgrounds. All cuts are listed in Tab. 3, where the object definition cuts for leptons ℓ and jets j are universal. The employed high-level observables are

- $p_T^{\ell W}$: reconstructed transverse momentum of the lepton originating from the W boson decay,
- m_T^W: transverse mass of the W boson,
- $m_Z^{\ell\ell}$: invariant mass of the lepton pair originating from the Z boson decay,
- $m_{\ell\ell}$: di-lepton invariant mass,
- m_{bb}: invariant mass of the bottom pair.

For WZ production, the main relevant backgrounds originate from Z + jets, $Z\gamma$, $t\bar{t}$, and WW production with a fake lepton in the final state or — without a fake lepton — from ZZ, $t\bar{t}V$, or VVV production, where V is either a W or Z boson. In the signal region, the contribution of these backgrounds is about 20% [41]. Since we expect the histogram-based and SBI approaches to be affected in a similar manner, we do not consider the WZ backgrounds for our analysis.

processes	**pre-selection cuts**
jet/lepton	$p_T^{\ell} > 15$ GeV, $p_T^{j} > 20$ GeV, $\lvert\eta^i\rvert < 2.5$ with $i \in \{\ell, j\}$
WZ	$E_{T,\text{miss}} > 45$ GeV, $p_T^{\ell W} > 20$ GeV, $m_T^W > 30$ GeV, $81.2 < m_Z^{\ell\ell} < 101.2$ GeV, $N_\ell = 3$
WW	$E_{T,\text{miss}} > 45$ GeV, $m_{\ell\ell} > 15$ GeV, $\lvert m_{\ell\ell} - m_Z\rvert > 15$ GeV, $N_\ell = 2$, $N_j = 0$
WH and ZH	$p_T^b > 35$ GeV, $p_T^j < 30$ GeV, $80 < m_{bb} < 160$ GeV, $R_{bb}, R_{b\ell}, R_{bj}, R_{\ell j} > 0.4$
WH only	$E_{T,\text{miss}} > 25$ GeV, $N_\ell = 1$, $N_b = 2$
ZH only	$N_\ell = N_b = 2$

Table 3: Summary of analysis selection cuts.

Similarly, we also do not take the backgrounds for WW production into account. Here, the main background would be di-top production, which is, however, heavily suppressed by the applied jet veto. Further subleading backgrounds are Drell-Yan, $W+$jets, and WZ production. In the signal region, the signal contribution is dominant with the background constituting about 35% of the total yield [40].

For ZH production, the dominant background is $Zb\bar{b}$ production. In the pre-selection region, the $Zb\bar{b}$ contribution is larger than the ZH contribution. Therefore, we include this background in our analysis following the procedure outlined in Sec. 2.2. However, we neglect the SMEFT corrections to the background process, assuming that the trained signal–background classifier will favor the Higgs signal. The presence of additional Wilson coefficients like four-fermion operators could compensate for the impact of the considered operators. We do not include them since they do not affect the target di-boson processes. The cancellation for the $Zb\bar{b}$ process can, however, still be present in particular given that $Zb\bar{b}$ production is precisely measured with the measurements following the SM expectations.

For WH production, we include the three most relevant backgrounds, $t\bar{t}$, $t\bar{t}b\bar{b}$, and $Wb\bar{b}$. For the $Wb\bar{b}$ background, we again neglect the dependence on the considered Wilson coefficients based on the same considerations as for the $Zb\bar{b}$ background.

Histogram observables

For the histogram limit setting, we have to choose specific observables and binning. Our selection shown in Tab. 4 is based on existing experimental analysis for WW and WZ production [41,42], as well as the simplified template cross-section (STXS) stage 1.2 [45–48] for WH and ZH production. For the WW and WZ binning, $p_T^{\ell_1}$ is the p_T of the leading lepton and m_T^{WZ}, the transverse mass of the WZ system. The choice of observables and binnings can be improved to provide better sensitivity to our SMEFT operators [16]. We deliberately adopt the standard STXS binning, to get an indication of how much sensitivity can be gained by adopting an SBI approach in comparison to existing experimental analyses.

For the histogram limits, we derive the respective histograms for the parameter θ and θ_0 and use them as input for a log-likelihood test. In the presence of background, we weight the histogram entries by the classifier output. This gives better sensitivity than placing a selection cut based on the classifier score.

Event generation and dataset preparation

For event generation, we employ MadGraph5_aMC@NLO 3.5.0 [49]. Generation is done at leading order, employing the SMEFTatNLO [50] UFO model. We approximate the effect of next-to-leading order corrections by applying a flat K-factor [45, 51–55]. The boson decays are simulated with MadSpin [56]. For the parton shower, we use Pythia8 8.306 [57]; for the detector simulation, Delphes 3.5.0 [58]; and, for the jet algorithm, FastJet 3.3.4 [59]. We use the parton distribution function set `PDF4LHC15_nlo_30` [60] accessed via LHAPDF6 [61].

process	observable	binning
WW	$p_T^{\ell_1}$	[0, 40, 50, 60, 70, 80, 90, 120, 140, 160, 180, 200, 300, 500]
WZ	m_T^{WZ}	[0, 200, 400, 600, 800, 1000, 1500, 2500]
WH	p_T^W	[0, 75, 150, 250, ∞]
ZH	p_T^Z	[0, 75, 150, 250, ∞]

Table 4: Observables and bins used for the histogram limit setting.

All events are generated for a center-of-mass energy of 13.6 TeV. Using this setup, we generate $\sim 10^6$ events for each signal and background process after applying the preselection cuts. For the signal channels, we further generate separate fractionally smeared datasets of size $\sim 2{\cdot}10^6$.

Although the considered operators can modify the decays of the Z, W, and Higgs bosons, we neglect their effect, which is limited by the low momentum transfer in the decays. Most of the branching ratios are already known to agree with the SM prediction with high precision. The impact of large Wilson coefficients on these decays can, moreover, be compensated for by other Wilson coefficients, which we do not include in our analysis.

Finally, we neglect systematic uncertainties. These typically affect mostly the non-kinematic part of the likelihood in the presence of large backgrounds. The total rate contribution to the likelihood is the same for the SBI and histogram approaches. Given that SBI extracts more kinematic information, we expect systematics to degrade the histogram limits more than the SBI limits. We refer to Refs. [13,23] for a detailed treatment of systematic uncertainties in the SBI approach.

3.3 Neural network setup and training

Learning differential cross-section ratios

For learning the differential cross-section ratios R_i and R_{ij}, we use a multi-layer perceptron (MLP) with five hidden layers, each consisting of 128 nodes. As activation function, we use the cumulative distribution function of the Gaussian distribution (GELU). As inputs, we use the four-vectors of all final-state particles, complemented by a selection of high-level features like the VV-invariant mass. Training is done for each process separately using 80% of the fractionally smeared dataset for training and 20% for validation. After training for 100–200 epochs using a cosine annealing scheduler with an initial learning rate of 10^{-4}, we select the best model based on the validation loss.

By default, we regress all non-zero R_i and R_{ij} simultaneously, summing the corresponding losses. Only for the ZH and WW processes, we learn $R_{c_{\Phi D}}$ and $R_{c_{\Phi D}^2}$ separately, to improve the numerical stability. For the WH process, we also exploit that $c_{\Phi D}$ only rescales the SM amplitude, as can be seen via the Feynman rules listed in App. A. Consequently, $R_{c_{\Phi D}}$ and $R_{c_{\Phi D}^2}$ are constants, and $R_{c_{\Phi D} c_{\Phi q}^{(3)}} \sim R_{c_{\Phi q}^{(3)}}$ as well as $R_{c_{\Phi D} c_{\Phi W}} \sim R_{c_{\Phi W}}$.

As discussed in Sec. 2.4, we checked the coverage of the learned likelihoods finding good agreement with the nominal confidence level. More detailed results are collected in App. B.

Signal–background classifiers

For the signal–background classification, we use a simple MLP with five layers, each containing 256 nodes. The output of each layer is passed through a leaky ReLU activation function. As inputs, we use four-vector entries of the involved particles as well as a selection of high-level observables like the WH or ZH invariant mass. The networks are trained on balanced signal and background datasets, to which no fractional smearing is applied.

For WH production, the background sample is obtained by mixing the samples of the different background processes according to their relative contribution to the phase-space region defined by the pre-selection cuts. 70% of these datasets are used for training, while the remaining 30% are used for the actual limit setting. For training, we use a scheduler which reduces the learning rate if the training loss has reached a plateau with an initial learning rate of 10^{-3}, a decay factor of 0.1, and a patience of 10 epochs. We find very good performance after approximately 50 epochs. The achieved area-under-the-curve (AUC) values for

the receiver-operating characteristic (ROC) curve are 0.81 and 0.84 for WH and ZH production, respectively.

4 Results

In this Section, we present the confidence limits for the di-boson processes introduced in Sec. 3 using the methods of Sec. 2. We assume an integrated luminosity of $300\,\text{fb}^{-1}$, deriving expected limits based on the SM assumption. Throughout this section, we compare

1. limits based on the known parton-level likelihood ratio for reference;
2. limits based solely on the total rate as a lower sensitivity bound;
3. limits derived from reconstruction-level histograms; and
4. limits from the reconstruction-level SBI.

4.1 Single-process limits

We first discuss the limits for each individual process. We show two-dimensional limits, setting the Wilson coefficients not shown to zero, as well as one-dimensional profiled limits. The corresponding one-dimensional constraints, where we set all other Wilson coefficients to zero, are collected in App. C.

WZ production

We start with the WZ process for which we consider the effect of four SMEFT operators. The associated two-dimensional limits, setting the third and fourth Wilson coefficient to zero, are shown in Fig. 1. For each type of limit, two contours indicate the one-sigma and two-sigma regions.

The results agree with Ref. [14], modulo the additional $\mathcal{O}_{\Phi D}$ operator. The total rate limits shown as grey lines are least constraining, featuring flat directions for all 2D combinations. These flat directions are mostly lifted by kinematic information in the histogram approach (orange dashed). In the $c_{\Phi WB}$–$c_{\Phi D}$ plane, the histogram approach still struggles to disentangle the two operators. Using SBI tightens the limits in comparison to the histogram approach, with the largest improvements in the $c_{\Phi WB}$–$c_{\Phi D}$, $c_{\Phi WB}$–$c^{(3)}_{\Phi q}$, and $c^{(3)}_{\Phi q}$–$c_{\Phi D}$ planes. The SBI limits are, as expected, weaker than the parton-level limits (blue lines), which we show as a check of the SBI results.

Moreover, in Fig. 2 we show the single parameter constraints profiled over all Wilson coefficients not shown. We compare parton-level constraints (blue), SBI constraints (green), and histogram-based constraints (orange). We do not show the rate-only results, where the profiling wipes out all constraints. Again, the SBI result is, as expected, weaker than the parton-level constraint. It, however, consistently outperforms the histogram-based limits in particular for $c_{\Phi WB}$ and $c_{\Phi D}$. Looking at Fig. 1, we can trace this back to the partial degeneracy between $c_{\Phi WB}$ and $c_{\Phi D}$. The SBI approach breaks this degeneracy more effectively than the histogram-based approach, since it exploits the full dimensionality of the phase space and not only one lower-dimensional summary statistic.

WW production

Next, we discuss the WW process, for which we also consider four SMEFT operators. The resulting two-dimensional limits, setting the remaining Wilson coefficients to zero, are shown in

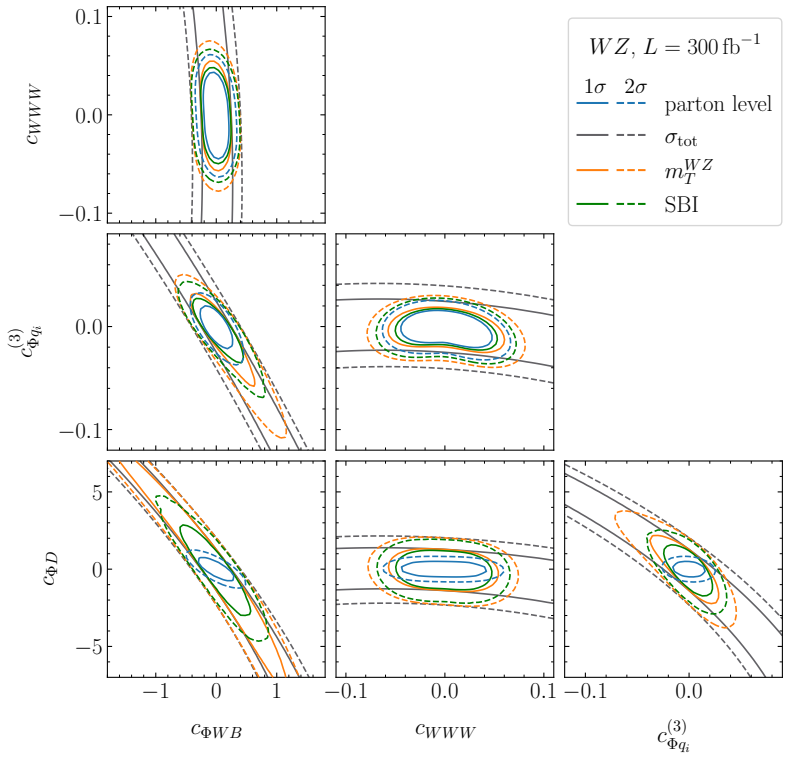

Figure 1: Expected two-dimensional constraints from ***WZ* production** setting the not-shown Wilson coefficient to zero.

Fig. 3. As for WZ, including kinematic information via histograms or SBI drastically improves the constraints, in particular in the c_{WWW} and $c_{\Phi WB}$ directions. Moreover, SBI outperforms the histogram approach with the largest improvements being present in the $c_{\Phi WB}$–$c_{\Phi D}$, the $c_{\Phi WB}$–c_{WWW}, and the $c_{\Phi q}^{(3)}$–$c_{\Phi D}$ planes. Here, SBI again profits from exploiting the full phase space and not only one particular low-dimensional summary statistic.

This is also reflected in the one-dimensional limits profiled over the full three-dimensional parameter space shown in Fig. 4. Using SBI leads to significantly stronger results for all four Wilson coefficients.

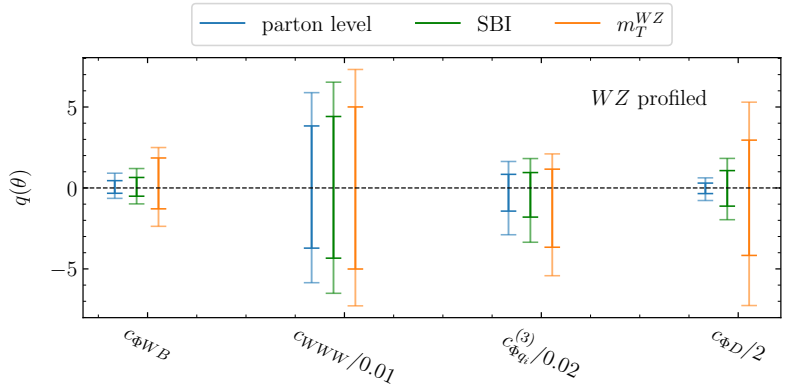

Figure 2: Expected profiled one-parameter confidence intervals for WZ production. The small horizontal lines indicate the one- and two-sigma confidence intervals.

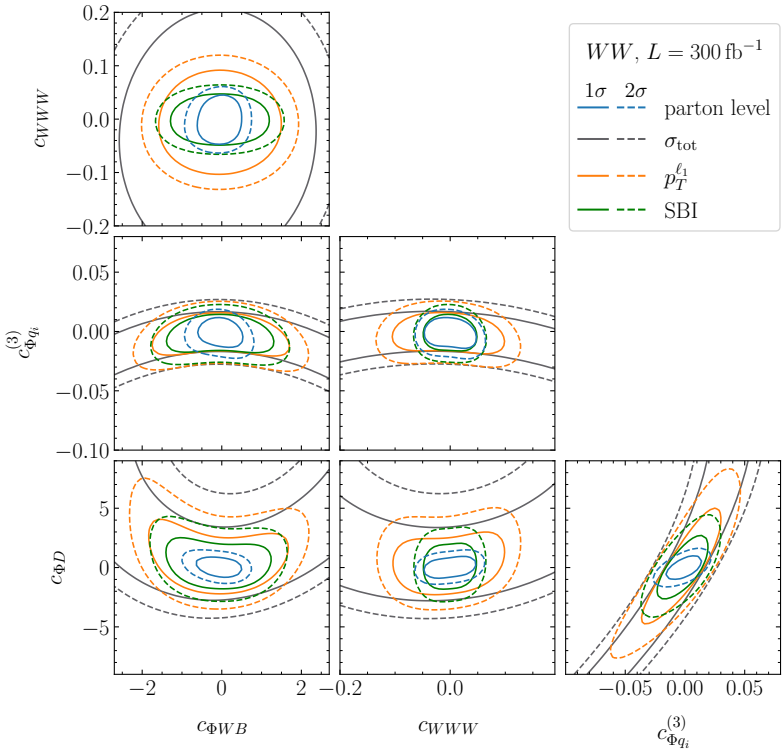

Figure 3: Expected two-dimensional constraints from WW **production** setting the not-varied Wilson coefficients to zero.

WH production

Now, we turn to the first Higgs-production process. WH production is only affected by three of our selected SMEFT operators. The corresponding two-dimensional limits, setting the third relevant Wilson coefficient to zero, are shown in Fig. 5. Here, we do not show parton-level bounds due to the presence of background and the corresponding need to construct the signal–background likelihood ratio. We again observe that the SBI constraints are stronger than the histogram limits. This is particularly evident for the $c_{\Phi D}$ and $c_{\Phi W}$ directions in which the p_T^W histogram is not able to extract significant information beyond the total rate.

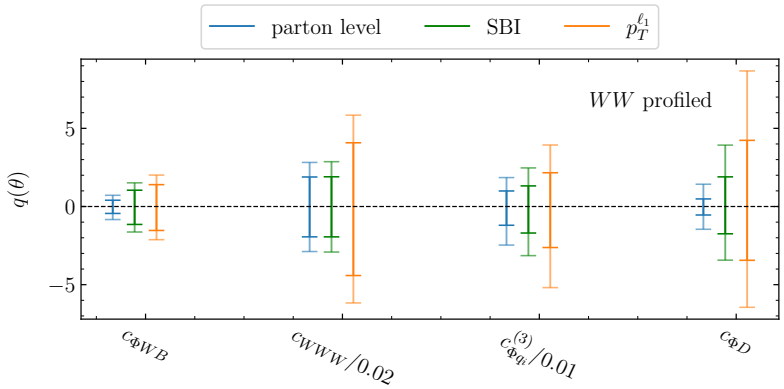

Figure 4: Expected profiled one-parameter confidence intervals for WW production. The small horizontal lines indicate the one- and two-sigma confidence intervals.

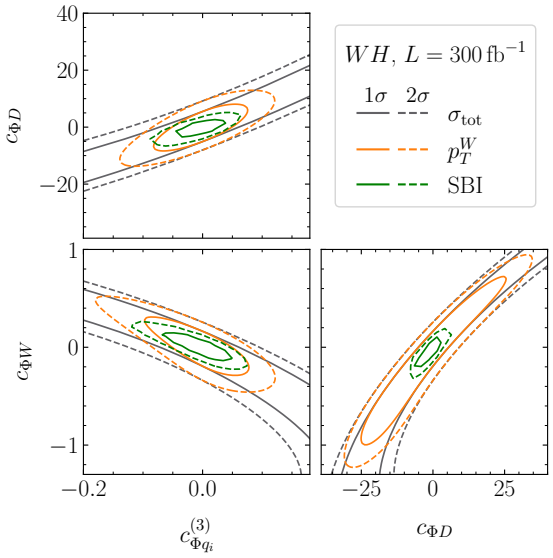

Figure 5: Expected two-dimensional constraints from ***WH* production** setting the not-varied Wilson coefficient to zero.

Looking at the one-dimensional limits when profiling over the other parameters, as shown in Fig. 6, we also clearly see the advantage of the SBI approach. The additional kinematic information extracted using the SBI approach significantly improves the limit on the three relevant Wilson coefficients. Our findings are in good agreement with Ref. [16]. In there, an alternative binning scheme is proposed to bring the histogram results closer to the SBI results.

ZH production

As the final process, we investigate ZH production, which is affected by five of the considered SMEFT operators. The 2D limits, setting the Wilson coefficients not shown to zero, are shown in Fig. 7.

First, the total rate constraints are relatively insensitive, making the kinematic information of the histogram and SBI approaches more relevant. This is a consequence of the large background contribution to the signal region. A harder cut on the signal–background classi-

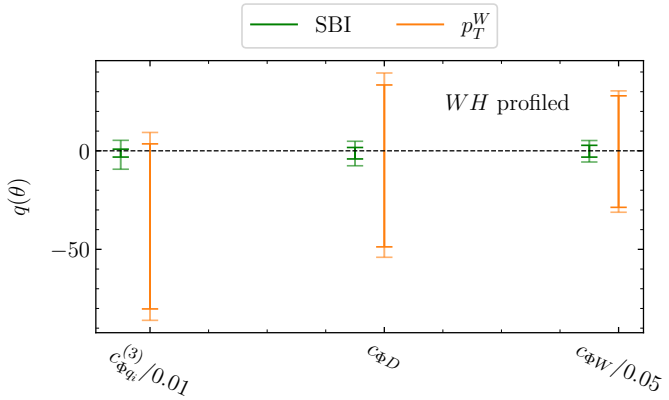

Figure 6: Expected profiled one-parameter confidence intervals for WH production, including backgrounds. The small horizontal lines indicate the one- and two-sigma confidence intervals.

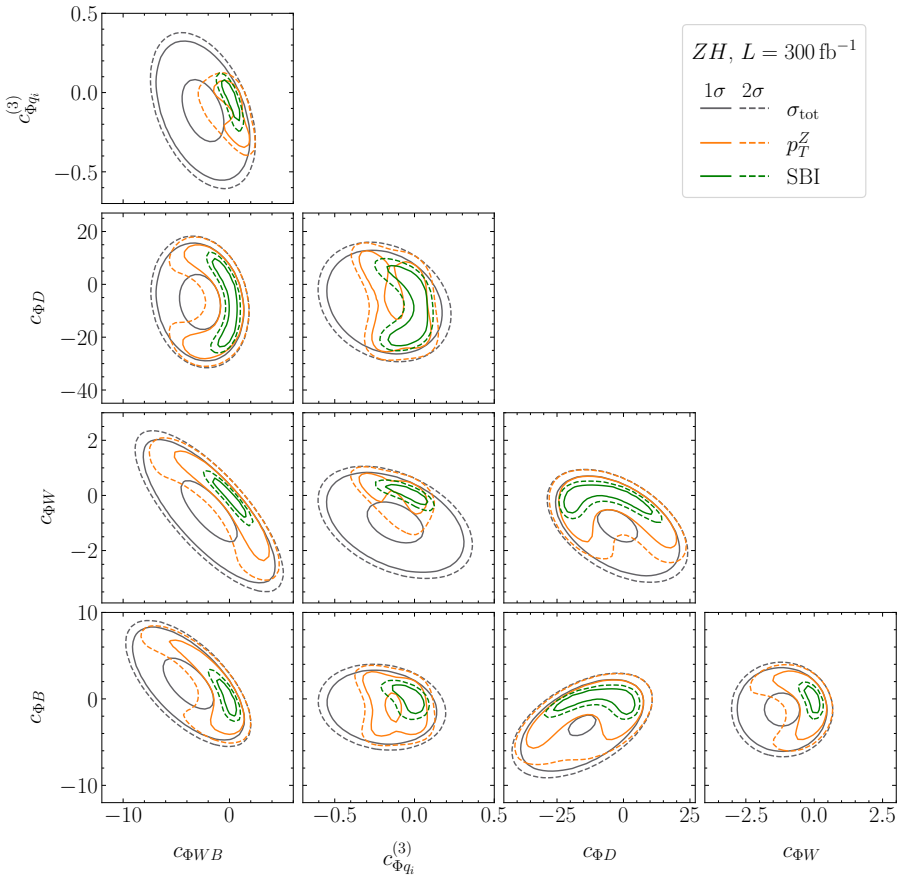

Figure 7: Expected two-dimensional constraints from ***ZH* production** setting the not-varied Wilson coefficient to zero. The SM point does not lie in the center of the parameter planes. For some of the shown constraints, in particular the total rates, the center of the parameter planes are expected to be excluded at the $1\,\sigma$ level resulting in the small central ellipsis.

fier score could enhance the rate constraints, but would reduce the kinematic constraints if combined with the kinematic SBI or histogram information.

For the histogram approach, the constraints get significantly tighter. They become even

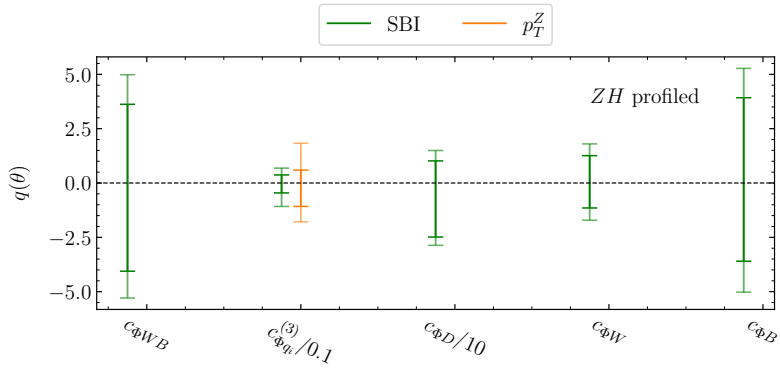

Figure 8: Expected profiled one-parameter confidence intervals for ZH production, including backgrounds. The small horizontal lines indicate the one- and two-sigma confidence intervals. If no confidence interval is shown, the corresponding parameter is not constrained.

tighter once the full kinematic information is extracted using the SBI approach. The effect is particularly strong for the $c_{\Phi W}$, $c_{\Phi B}$, $c_{\Phi WB}$, and $c_{\Phi D}$ directions and the two-dimensional combinations thereof. This improvement is due to the SBI approach being sensitive to the Z boson polarization, disentangling the left-handed and right-handed Z coupling to quarks, see App. A. In contrast, the p_T^Z histogram is not sensitive to the Z boson polarization.

In the upper half of the $c_{\Phi q}^{(3)}$–$c_{\Phi D}$ plane, we observe for $c_{\Phi D} \simeq 5$ and $c_{\Phi q}^{(3)} \simeq -0.2$ a small region in which the histogram-based approach is slightly stronger than the SBI limits. This is likely due to imperfect training of the $R_{c_{\Phi q}^{(3)} c_{\Phi D}}$ differential cross-section ratio.

The advantages of the SBI approach are even more visible in the one-dimensional profiled limits in Fig. 8. Since the p_T^Z histogram is not sensitive to the Z polarization, the $c_{\Phi WB}$, $c_{\Phi D}$, $c_{\Phi W}$, and $c_{\Phi B}$ Wilson coefficients form a degenerate set resulting in no one-dimensional profiled constraint for any of these coefficients. The SBI approach breaks this degeneracy and is able to constraint each of these coefficients even if profiling over the others. Only for $c_{\Phi q}^{(3)}$, the histogram-based approach yields a profiled one-dimensional limit, whose lower bound is weaker than the corresponding SBI constraint.

4.2 Combined limits

Until now, we have shown that SBI leads to significantly stronger constraints on the Wilson coefficients for each of the considered processes in comparison to the histogram approach. It remains to clarify whether this advantage persists when combining all four processes.

As mentioned above, we neglect systematic uncertainties. Moreover, the individual signal and background processes do not overlap within the regions defined by the pre-selection cuts introduced in Sec. 3.2. Therefore, we assume the different processes to be uncorrelated. The combined test statistic is then

$$q_{\text{comb}}(\theta) = q_{WW}(\theta) + q_{WZ}(\theta) + q_{WH}(\theta) + q_{ZH}(\theta)\,. \tag{29}$$

Based on this combined test statistic, we derive the combined confidence regions.

Before showing the combined limits, we investigate in Fig. 9 the contribution of each process. Here, we focus only on the SBI approach (solid lines) and the histogram approach (dashed lines). The different colors indicate the processes. Each process is important to constrain at least one of the directions. For example, the $c_{\Phi WB}$ direction is mostly constrained by WZ production; the $c_{\Phi W}$ direction, mostly by WH production; the $c_{\Phi B}$ direction, mostly by ZH production; and, the $c_{\Phi q}^{(3)}$ direction, mostly by WW production. This clearly shows the benefit of a combined analysis of all four di-boson processes. We also observe that SBI is more effective than the histogram approach in constraining most of the directions. This then translates into more stringent combined constraints.

We show the combined 2D limits, setting all other Wilson coefficients to zero, in Fig. 10. As before, the SBI approach consistently outperforms the histogram-based limits. In some parameter planes, e.g. the $c_{\Phi WB}$–$c_{\Phi q}^{(3)}$ plane, the differences are relatively small, because these directions are well constrained by the histograms for at least one of the considered processes. In other parameter planes, e.g. the c_{WWW}–$c_{\Phi B}$ plane, we find larger improvements with the SBI approach clearly outperforming the histogram limits. This is in particular true if the histogram approach is not able to extract information beyond the total rate, as seen e.g. in the $c_{\Phi W}$, $c_{\Phi B}$, and $c_{\Phi D}$ directions.

Finally, in Fig. 11 we show the one-dimensional limits profiling over all other Wilson coefficients using the combination of all four considered processes. For all six considered Wilson coefficients, we see significant improvements of the SBI limits over the histogram-based lim-

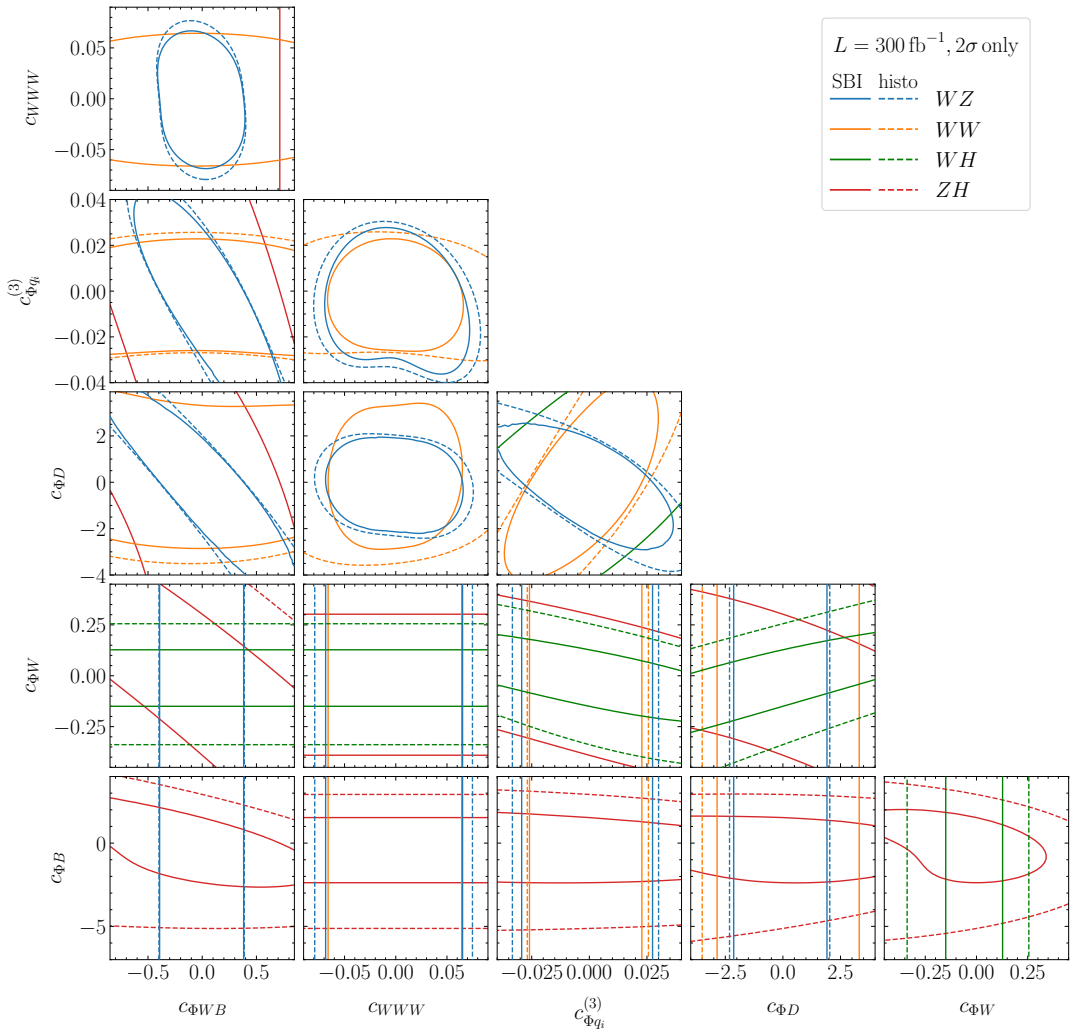

Figure 9: Expected two-dimensional 95% C.L. constraints setting the not-varied Wilson coefficient to zero. We compare the limits of the four single processes using either the histogram approach (dashed lines) or SBI (solid lines).

its. While the combination of the different processes avoids flat directions in the histogram limits, as observed for WH production in Fig. 6, combining different channels does not bring the sensitivity of the histogram approach to the same level as the SBI approach. Differences are particularly striking for $c_{\Phi D}$, $c_{\Phi W}$, $c_{\Phi WB}$, and $c_{\Phi B}$. As discussed for ZH production, these Wilson coefficients are strongly correlated if the analysis is not sensitive to the polarization of the Z boson. The SBI limits on these coefficients are a factor of ~ 2 stronger than the histogram-based limits. To reach the same sensitivity in the histogram approach, a factor ~ 4 more data would be needed if we assume a naive scaling of the limits with the inverse square root of the luminosity. But even for the c_{WWW} and $c^{(3)}_{\Phi q}$ Wilson coefficients, to which the chosen histogram observables are more directly sensitive, the SBI approach is $\sim 30\%$ more sensitive, corresponding to a factor of ~ 1.8 more data.

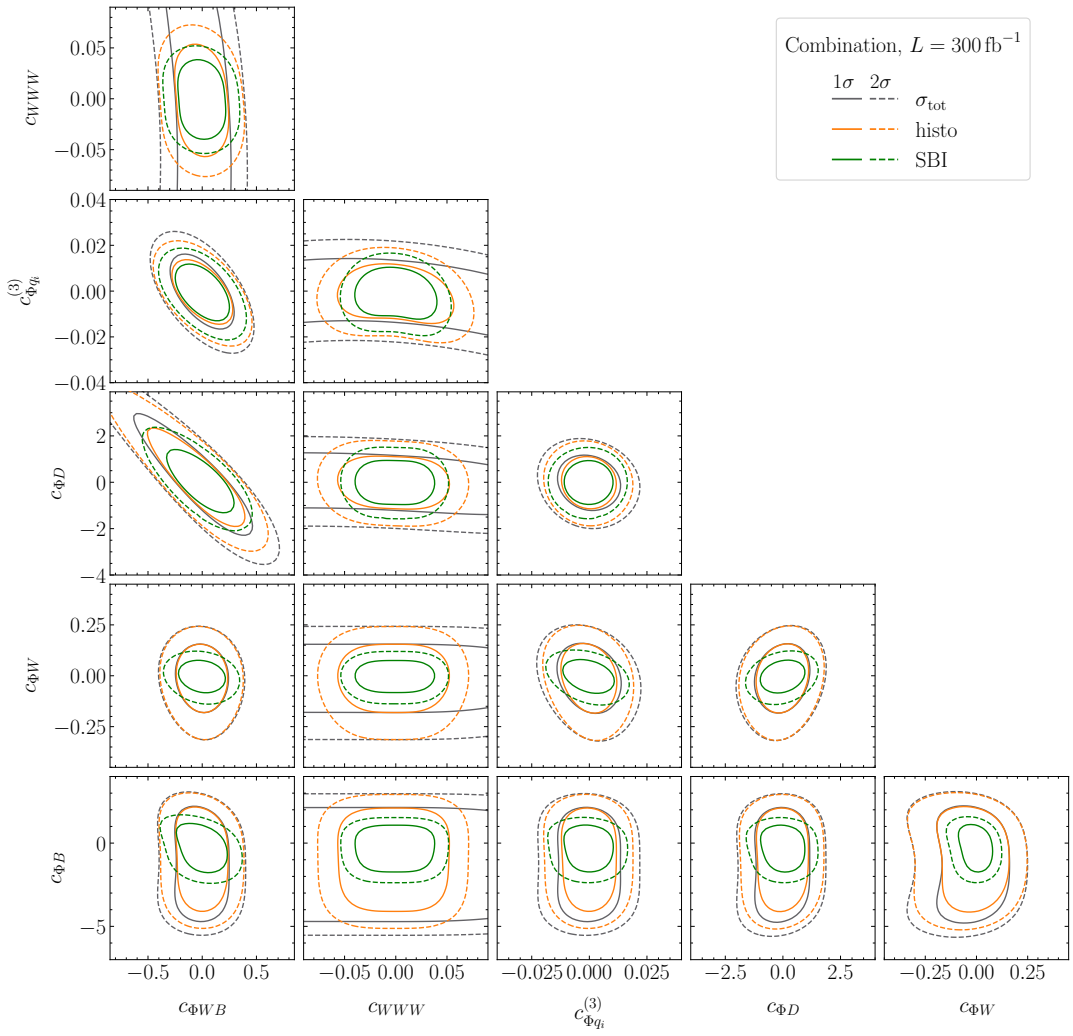

Figure 10: Expected two-dimensional constraints from the combined di-boson processes setting the not-varied Wilson coefficient to zero. We compare the parton-level limits with results based on SBI, histogram-based inference, and only the total rate information.

5 Conclusions

Simulation-based inference (SBI) is a key methodology if we want to unleash the full potential of the current and future LHC runs. In comparison to traditional approaches like histogramming, it does not rely on binned low-dimensional summary statistics. Exploiting the full high-dimensional event information not only leads to an improved sensitivity to single theory parameters but is also more effective in disentangling different parameters.

This is particularly evident in the SMEFT context. Often, a single process is affected by many different operators, each of which changes the kinematic distributions in a slightly different manner. We have shown in this paper that SBI clearly outperforms histogram-based methods in such scenarios, focusing on four different di-boson production channels, WW, WZ, WH, and ZH production, and six SMEFT operators. For each process, the SBI limits are significantly tighter than the histogram-based limits. In particular, SBI excels at lifting degeneracies between two or more Wilson coefficients, resulting in significantly stronger one-dimensional profiled limits.

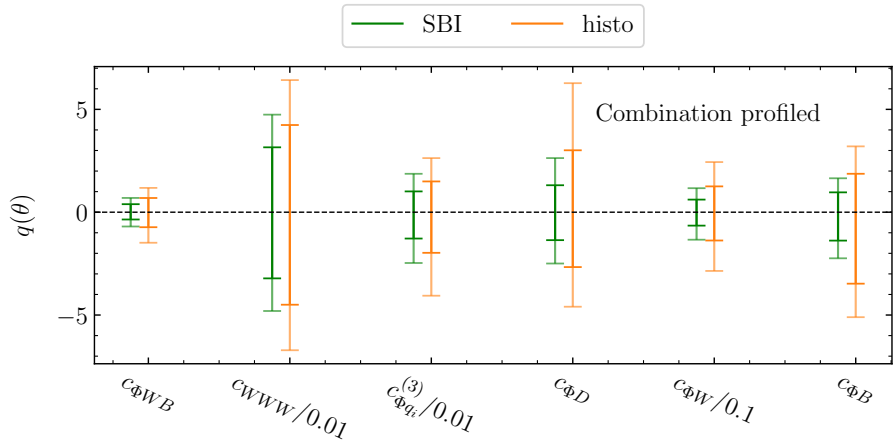

Figure 11: Expected profiled one-parameter confidence intervals combining all four di-boson processes, including backgrounds. The small horizontal lines indicate the one- and two-sigma confidence intervals.

Apart from showing that SBI outperforms the histogram approach for single-process limits, which has been shown before, we took a further step to demonstrate the potential of SBI. In particular, it has so far been unclear if the superior sensitivity of SBI persists when multiple processes are combined in a global analysis, as it is common practice for constraining SMEFT parameters. To answer this question, we have combined the limits of the four di-boson processes. In this global analysis context, the SBI shows persistent advantages over the histogram-based limits. It is significantly more sensitive, even for Wilson coefficients which strongly alter the distributions of the chosen histogram observables. Assuming a naive scaling of the constraints with the inverse square root of the luminosity, the SBI improvements over histogram-based inference corresponds to a factor two or more in luminosity.

Acknowledgments

This research is supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under grant 396021762–TRR 257: *Particle Physics Phenomenology after the Higgs Discovery*, and through Germany's Excellence Strategy EXC 2181/1 – 390900948 (the *Heidelberg STRUCTURES Excellence Cluster*). We would also like to thank the Baden-Württemberg Stiftung for financing through the program *Internationale Spitzenforschung*, project *Uncertainties – Teaching AI its Limits* (BWST_ISF2020-010). We acknowledge support by the state of Baden-Württemberg through bwHPC and the German Research Foundation (DFG) through grant no INST 39/963-1 FUGG (bwForCluster NEMO).

A Relevant SMEFT Feynman rules

Following Ref. [50, 62], we list here all relevant SMEFT Feynman rules neglecting $\mathcal{CP}$ and flavour violation:

$$
\begin{aligned}
c(u^{f_1}u^{f_2}Z_\mu) = {}& \frac{2ies_W}{3c_W}\delta_{f_1f_2}\gamma_\mu P_R + ie\left(\frac{s_W}{6c_W}-\frac{c_W}{2s_W}\right)\delta_{f_1f_2}\gamma_\mu P_L \\
& - \frac{ie}{2c_Ws_W}c_{\Phi q}^{(3)}\frac{v^2}{\Lambda^2}\delta_{f_1f_2}\gamma_\mu P_L - \frac{2}{3}iec_{\Phi WB}\frac{v^2}{\Lambda^2}\delta_{f_1f_2}\gamma_\mu P_L \\
& - \frac{1}{6}iec_{\Phi D}\frac{v^2}{\Lambda^2}\delta_{f_1f_2}\gamma_\mu\left[\left(\frac{1}{4c_Ws_W}+\frac{c_W}{s_W}\right)P_L + \left(\frac{1}{c_Ws_W}+\frac{c_W}{s_W}\right)P_R\right], \\
c(d^{f_1}d^{f_2}Z_\mu) = {}& -\frac{ies_W}{3c_W}\delta_{f_1f_2}\gamma_\mu P_R + ie\left(\frac{s_W}{6c_W}+\frac{c_W}{2s_W}\right)\delta_{f_1f_2}\gamma_\mu P_L \\
& + \frac{ie}{2c_Ws_W}c_{\Phi q}^{(3)}\frac{v^2}{\Lambda^2}\delta_{f_1f_2}\gamma_\mu P_L \\
& - \frac{1}{6}iec_{\Phi D}\frac{v^2}{\Lambda^2}\delta_{f_1f_2}\gamma_\mu\left[\left(\frac{1}{4c_Ws_W}+\frac{3c_W}{2s_W}\right)P_L + \left(\frac{1}{4c_Ws_W}+\frac{3c_W}{2s_W}\right)P_R\right], \\
c(u^{f_1}\bar{d}^{f_2}W_\mu^+) = {}& \frac{ie}{\sqrt{2}s_W}\delta_{f_1f_2}\gamma_\mu P_L + \frac{ie}{\sqrt{2}s_W}c_{\Phi q}^{(3)}\frac{v^2}{\Lambda^2}\delta_{f_1f_2}\gamma_\mu P_L\,, \\
c(u^{f_1}u^{f_2}Z_\mu h) = {}& -\frac{ie}{c_Ws_W}c_{\Phi q}^{(3)}\frac{v}{\Lambda^2}\delta_{f_1f_2}\gamma_\mu P_L\,, \\
c(d^{f_1}d^{f_2}Z_\mu h) = {}& \frac{ie}{c_Ws_W}c_{\Phi q}^{(3)}\frac{v}{\Lambda^2}\delta_{f_1f_2}\gamma_\mu P_L + \frac{ie}{c_Ws_W}c_{\Phi D}\frac{v}{\Lambda^2}\delta_{f_1f_2}\gamma_\mu P_R\,, \\
c(u^{f_1}d^{f_2}W_\mu^+h) = {}& \sqrt{2}\frac{ie}{s_W}c_{\Phi q}^{(3)}\frac{v}{\Lambda^2}\delta_{f_1f_2}\gamma_\mu P_L\,, \\
c(hW_\mu^+W_\nu^-) = {}& \frac{ie^2}{2s_W^2}vg_{\mu\nu}\left(1-\frac{1}{4}c_{\Phi D}\frac{v^2}{\Lambda^2}\right) \\
& + 4ic_{\Phi W}\frac{v}{\Lambda^2}(p_{2\mu}p_{3\nu}-p_2\cdot p_3 g_{\mu\nu})\,, \\
c(hZ_\mu^+Z_\nu^-) = {}& \frac{ie^2v}{2c_W^2s_W^2}g_{\mu\nu}\left(1+\frac{1}{4}c_{\Phi D}\frac{v^2}{\Lambda^2}\right) \\
& + 4i\left(c_W^2c_{\Phi W}+c_Ws_Wc_{\Phi WB}+s_W^2c_{\Phi B}\right)\frac{v}{\Lambda^2}\left(p_{2\mu}p_{3\nu}-p_2\cdot p_3 g_{\mu\nu}\right)\,, \\
c(h\gamma_\mu Z_\nu) = {}& 2i\left(2c_Ws_Wc_{\Phi W}+(s_W^2-c_W^2)c_{\Phi WB}+2c_Ws_Wc_{\Phi B}\right) \\
& \cdot\frac{v}{\Lambda^2}\left(p_{2\mu}p_{3\nu}-p_2\cdot p_3 g_{\mu\nu}\right)\,, \\
c(W_{\mu_1}^-W_{\mu_2}^+Z_{\mu_3}) = {}& -i\frac{ec_W}{s_W}\left(1+\frac{1}{4}\frac{v^2}{\Lambda^2}c_{\Phi D}\right)\left(g_{\mu_1\mu_2}(p_1-p_3)_{\mu_3}+g_{\mu_2\mu_3}(p_2-p_3)_{\mu_1}\right. \\
& \qquad\qquad \left. +g_{\mu_1\mu_3}(p_3-p_1)_{\mu_2}\right) \\
& - 6ic_W\frac{c_{WWW}}{\Lambda^2}\left(p_{3,\mu_1}p_{1,\mu_2}p_{2,\mu_3}-p_{2,\mu_1}p_{3,\mu_2}p_{1,\mu_3}\right. \\
& \qquad +g_{\mu_1\mu_2}(p_{1,\mu_3}p_2\cdot p_3-p_{2,\mu_3}p_1\cdot p_3) \\
& \qquad +g_{\mu_2\mu_3}(p_{2,\mu_1}p_1\cdot p_3-p_{3,\mu_1}p_1\cdot p_2) \\
& \qquad \left.+g_{\mu_1\mu_3}(p_{3,\mu_2}p_1\cdot p_2-p_{1,\mu_2}p_2\cdot p_3)\right) \\
& - iec_{\Phi WB}\frac{v^2}{\Lambda^2}\left(g_{\mu_1\mu_2}(p_1-p_2)_{\mu_3}-g_{\mu_1\mu_3}p_{1,\mu_2}+g_{\mu_2\mu_3}p_{2,\mu_1}\right)\,. \qquad (30)
\end{aligned}
$$

As electroweak input parameters, we use M_W, M_Z, and G_F. The parameters v, e, c_W, and s_W are then fixed via

$$v = \frac{1}{\sqrt{\sqrt{2}G_F}}\,, \quad e = \frac{2M_W s_W}{v}\,, \quad c_W = \frac{M_W}{M_Z}\,, \quad s_W = \sqrt{1-c_W^2}\,. \tag{31}$$

B Likelihood coverages

For computing the coverage of the learned likelihoods, we restrict us for computational efficiency to two-dimensional parameter scans. Since the likelihood is a second-order polynomial in the Wilson coefficients, two-dimensional parameter planes are sufficient to test all directions and correlations of the overall likelihood. Analogously, testing the coverage of the single-process likelihoods is also sufficient to test the coverage of the combined likelihood.

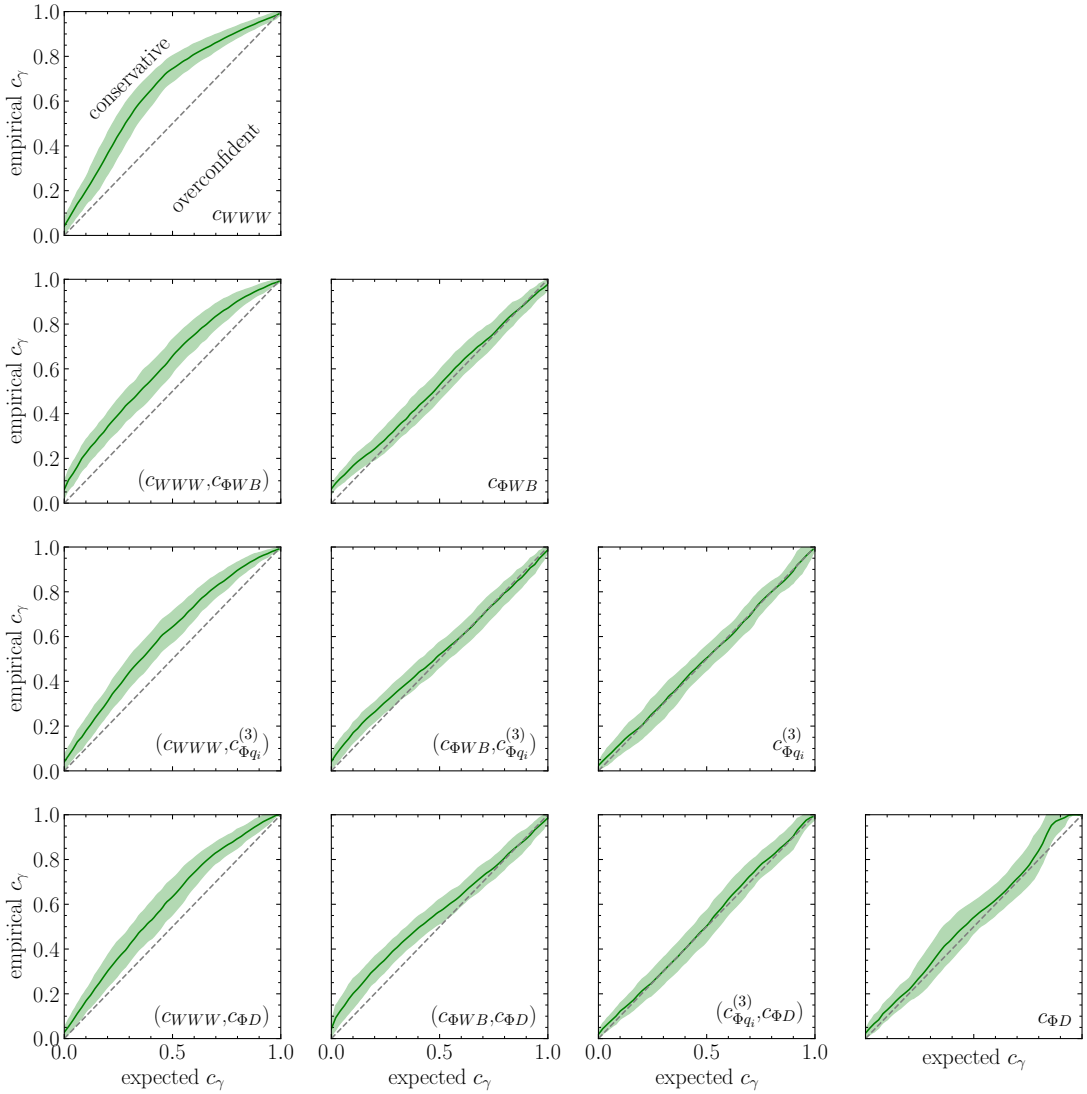

Figure 12: Two-dimensional coverage curves for all pairs of Wilson coefficients for the WZ dataset.

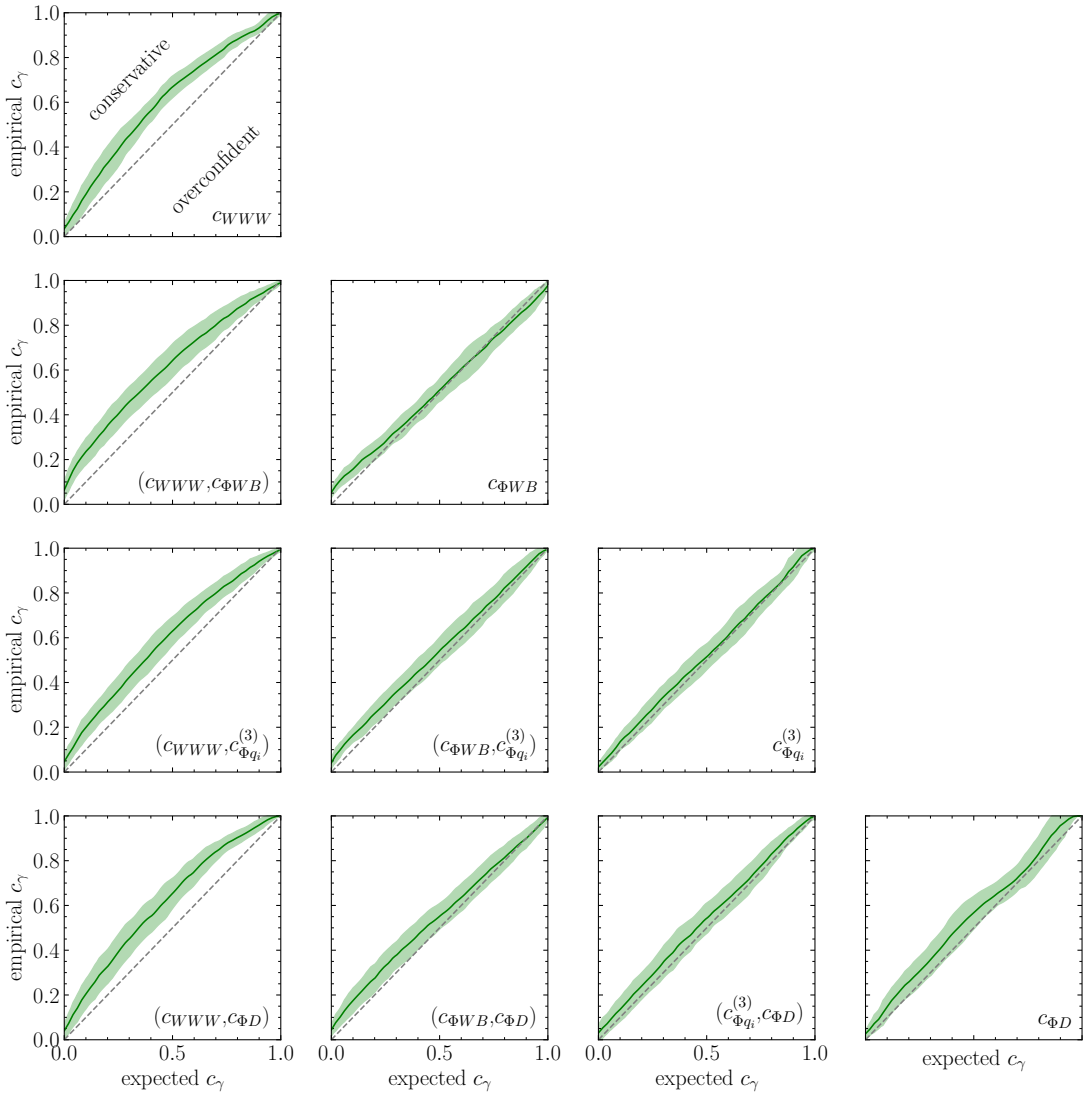

Figure 13: Two-dimensional coverage curves for all pairs of Wilson coefficients for the WW dataset.

The resulting coverage plots showing the emperical coverage against the expected coverage are shown in Figs. 12–15. For most of the parameter planes, the coverage curves lie, as expected, on the diagonal. Only for a few directions — e.g., c_{WWW} in WW production —, we observe that the learned likelihood is slighlty underconfident resulting in more conservative limits. In these case, an improved extraction of the likelihood would help to further tighten the expected constraints.

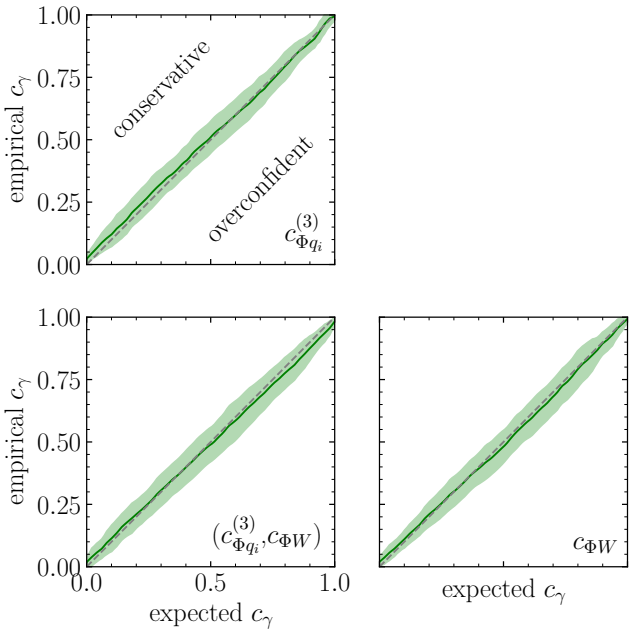

Figure 14: Two-dimensional coverage curves for all pairs of Wilson coefficients for the WH dataset. Here, we do not show parameter combinations involving $c_{\Phi D}$, since $c_{\Phi D}$ is only constrained by total rate information.

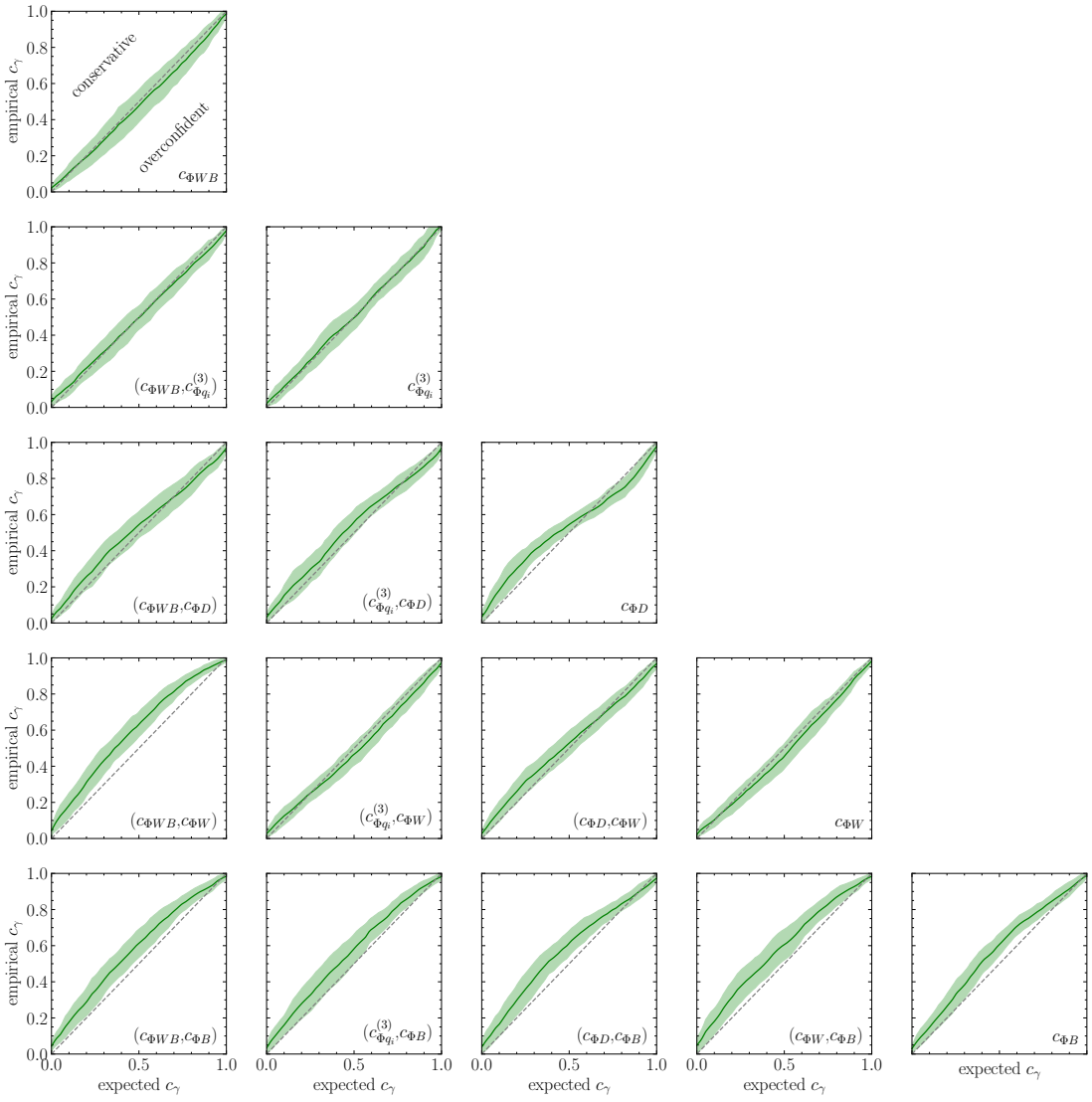

Figure 15: Two-dimensional coverage curves for all pairs of Wilson coefficients for the ZH dataset.

C Single parameter results

We show the one-dimensional limits for single Wilson coefficients setting all other Wilson coefficients to zero Fig. 16. Compared to the profiled limits, the difference between the SBI and histogram-based limits is smaller, since only a single Wilson coefficient needs to be constrained. SBI, however, still provides better or on-par sensitivity.

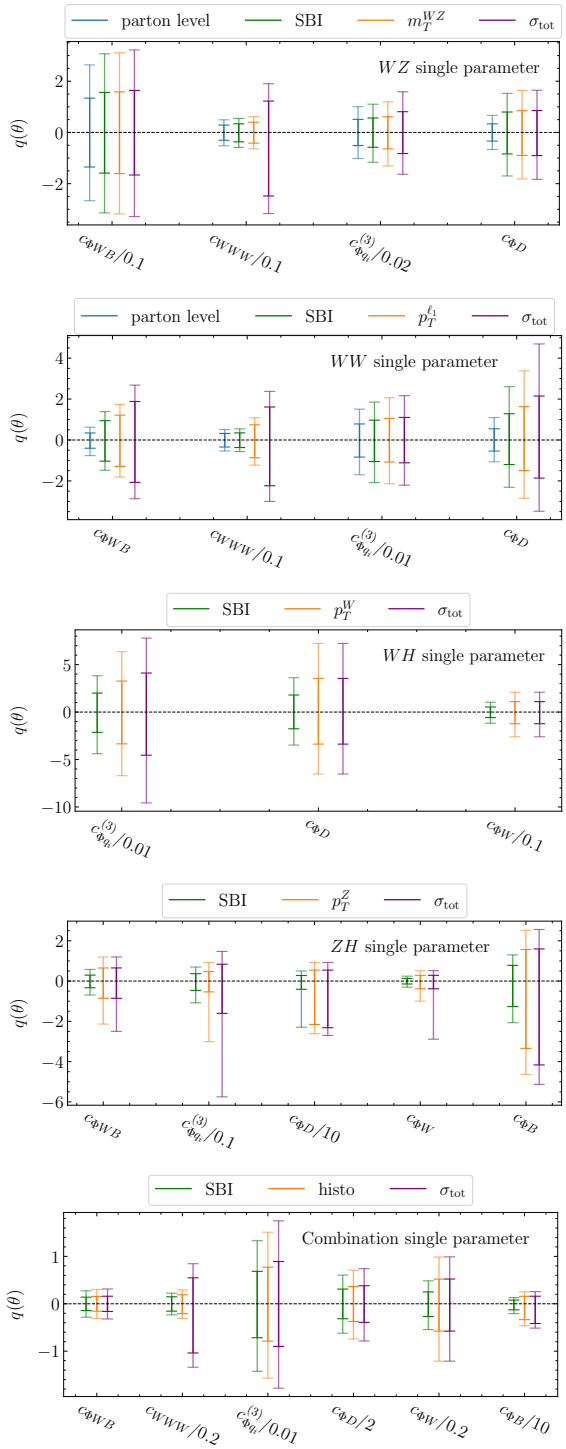

Figure 16: Projected confidence intervals for the individual processes and for their combination (lower). The small horizontal lines indicate the one- and two-sigma confidence intervals.

References

[1] S. Badger *et al.*, *Machine learning and LHC event generation*, SciPost Phys. **14** (2023) 4, 079, arXiv:2203.07460 [hep-ph].

[2] T. Plehn, A. Butter, B. Dillon, T. Heimel, C. Krause, and R. Winterhalder, *Modern Machine Learning for LHC Physicists*, arXiv:2211.01421 [hep-ph].

[3] K. Cranmer, J. Brehmer, and G. Louppe, *The frontier of simulation-based inference*, Proc. Nat. Acad. Sci. **117** (2020) 48, 30055, arXiv:1911.01429 [stat.ML].

[4] A. Butter, T. Heimel, T. Martini, S. Peitzsch, and T. Plehn, *Two invertible networks for the matrix element method*, SciPost Phys. **15** (2023) 3, 094, arXiv:2210.00019 [hep-ph].

[5] T. Heimel, N. Huetsch, R. Winterhalder, T. Plehn, and A. Butter, *Precision-Machine Learning for the Matrix Element Method*, arXiv:2310.07752 [hep-ph].

[6] J. Brehmer, K. Cranmer, G. Louppe, and J. Pavez, *A Guide to Constraining Effective Field Theories with Machine Learning*, Phys. Rev. D **98** (2018) 5, 052004, arXiv:1805.00020 [hep-ph].

[7] J. Brehmer, K. Cranmer, G. Louppe, and J. Pavez, *Constraining Effective Field Theories with Machine Learning*, Phys. Rev. Lett. **121** (2018) 11, 111801, arXiv:1805.00013 [hep-ph].

[8] J. Brehmer, G. Louppe, J. Pavez, and K. Cranmer, *Mining gold from implicit models to improve likelihood-free inference*, Proc. Nat. Acad. Sci. **117** (2020) 10, 5242, arXiv:1805.12244 [stat.ML].

[9] J. Brehmer, F. Kling, I. Espejo, and K. Cranmer, *MadMiner: Machine learning-based inference for particle physics*, Comput. Softw. Big Sci. **4** (2020) 1, 3, arXiv:1907.10621 [hep-ph].

[10] S. Chatterjee, N. Frohner, L. Lechner, R. Schöfbeck, and D. Schwarz, *Tree boosting for learning EFT parameters*, Comput. Phys. Commun. **277** (2022) 108385, arXiv:2107.10859 [hep-ph].

[11] S. Chatterjee, S. Rohshap, R. Schöfbeck, and D. Schwarz, *Learning the EFT likelihood with tree boosting*, arXiv:2205.12976 [hep-ph].

[12] K. Kong, K. T. Matchev, S. Mrenna, and P. Shyamsundar, *New Machine Learning Techniques for Simulation-Based Inference: InferoStatic Nets, Kernel Score Estimation, and Kernel Likelihood Ratio Estimation*, arXiv:2210.01680 [stat.ML].

[13] R. Schöfbeck, *Refinable modeling for unbinned SMEFT analyses*, Mach. Learn. Sci. Tech. **6** (2025) 1, 015007, arXiv:2406.19076 [hep-ph].

[14] H. Bahl, V. Bresó, G. De Crescenzo, and T. Plehn, *Advancing Tools for Simulation-Based Inference*, arXiv:2410.07315 [hep-ph].

[15] M. Silva, R. Barrué, I. Ochoa, and P. Conde Muíño, *Searching for HWW Anomalous Couplings with Simulation-Based Inference*, arXiv:2509.03307 [hep-ph].

[16] J. Brehmer, S. Dawson, S. Homiller, F. Kling, and T. Plehn, *Benchmarking simplified template cross sections in WH production*, JHEP **11** (2019) 034, arXiv:1908.06980 [hep-ph].

[17] R. K. Barman, D. Gonçalves, and F. Kling, *Machine learning the Higgs boson-top quark CP phase*, Phys. Rev. D **105** (2022) 3, 035023, arXiv:2110.07635 [hep-ph].

[18] H. Bahl and S. Brass, *Constraining $\mathcal{CP}$-violation in the Higgs-top-quark interaction using machine-learning-based inference*, JHEP **03** (2022) 017, arXiv:2110.10177 [hep-ph].

[19] R. Gomez Ambrosio, J. ter Hoeve, M. Madigan, J. Rojo, and V. Sanz, *Unbinned multivariate observables for global SMEFT analyses from machine learning*, JHEP **03** (2023) 033, arXiv:2211.02058 [hep-ph].

[20] R. Barrué, P. Conde-Muíño, V. Dao, and R. Santos, *Simulation-based inference in the search for CP violation in leptonic WH production*, JHEP **04** (2024) 014, arXiv:2308.02882 [hep-ph].

[21] R. Mastandrea, B. Nachman, and T. Plehn, *Constraining the Higgs potential with neural simulation-based inference for di-Higgs production*, Phys. Rev. D **110** (2024) 5, 056004, arXiv:2405.15847 [hep-ph].

[22] A. Ghosh, M. Griese, U. Haisch, and T. H. Park, *Neural simulation-based inference of the Higgs trilinear self-coupling via off-shell Higgs production*, arXiv:2507.02032 [hep-ph].

[23] L. Benato, C. Giordano, C. Krause, A. Li, R. Schöfbeck, D. Schwarz, M. Shooshtari, and D. Wang, *Unbinned inclusive cross-section measurements with machine-learned systematic uncertainties*, arXiv:2505.05544 [hep-ph].

[24] ATLAS, G. Aad *et al.*, *Measurement of off-shell Higgs boson production in the $H^* \to ZZ \to 4\ell$ decay channel using a neural simulation-based inference technique in 13 TeV pp collisions with the ATLAS detector*, Rept. Prog. Phys. **88** (2025) 5, 057803, arXiv:2412.01548 [hep-ex].

[25] ATLAS, G. Aad *et al.*, *An implementation of neural simulation-based inference for parameter estimation in ATLAS*, Rept. Prog. Phys. **88** (2025) 6, 067801, arXiv:2412.01600 [physics.data-an].

[26] C. Degrande, N. Greiner, W. Kilian, O. Mattelaer, H. Mebane, T. Stelzer, S. Willenbrock, and C. Zhang, *Effective Field Theory: A Modern Approach to Anomalous Couplings*, Annals Phys. **335** (2013) 21, arXiv:1205.4231 [hep-ph].

[27] I. Brivio and M. Trott, *The Standard Model as an Effective Field Theory*, Phys. Rept. **793** (2019) 1, arXiv:1706.08945 [hep-ph].

[28] A. Butter, O. J. P. Éboli, J. Gonzalez-Fraile, M. C. Gonzalez-Garcia, T. Plehn, and M. Rauch, *The Gauge-Higgs Legacy of the LHC Run I*, JHEP **07** (2016) 152, arXiv:1604.03105 [hep-ph].

[29] E. da Silva Almeida, A. Alves, N. Rosa Agostinho, O. J. P. Éboli, and M. C. Gonzalez-Garcia, *Electroweak Sector Under Scrutiny: A Combined Analysis of LHC and Electroweak Precision Data*, Phys. Rev. D **99** (2019) 3, 033001, arXiv:1812.01009 [hep-ph].

[30] A. Biekötter, T. Corbett, and T. Plehn, *The Gauge-Higgs Legacy of the LHC Run II*, SciPost Phys. **6** (2019) 6, 064, arXiv:1812.07587 [hep-ph].

[31] 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 [hep-ph].

[32] E. d. S. Almeida, A. Alves, O. J. P. Éboli, and M. C. Gonzalez-Garcia, *Electroweak legacy of the LHC run II*, Phys. Rev. D **105** (2022) 1, 013006, arXiv:2108.04828 [hep-ph].

[33] 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) 5, 393, arXiv:2302.06660 [hep-ph].

[34] N. Elmer, M. Madigan, T. Plehn, and N. Schmal, *Staying on Top of SMEFT-Likelihood Analyses*, SciPost Phys. **18** (2025) 108, arXiv:2312.12502 [hep-ph].

[35] T. Heimel, T. Plehn, and N. Schmal, *Profile Likelihoods on ML-Steroids*, arXiv:2411.00942 [hep-ph].

[36] J. de Blas, A. Goncalves, V. Miralles, L. Reina, L. Silvestrini, and M. Valli, *Constraining new physics effective interactions via a global fit of electroweak, Drell-Yan, Higgs, top, and flavour observables*, arXiv:2507.06191 [hep-ph].

[37] I. Brivio, S. Bruggisser, N. Elmer, E. Geoffray, M. Luchmann, and T. Plehn, *To profile or to marginalize - A SMEFT case study*, SciPost Phys. **16** (2024) 1, 035, arXiv:2208.08454 [hep-ph].

[38] K. Cranmer, J. Pavez, and G. Louppe, *Approximating Likelihood Ratios with Calibrated Discriminative Classifiers*, arXiv:1506.02169 [stat.AP].

[39] M. Arratia *et al.*, *Publishing unbinned differential cross section results*, JINST **17** (2022) 01, P01024, arXiv:2109.13243 [hep-ph].

[40] ATLAS, 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) 10, 884, arXiv:1905.04242 [hep-ex].

[41] ATLAS, 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) 6, 535, arXiv:1902.05759 [hep-ex].

[42] ATLAS, G. Aad *et al.*, *Measurements of W^+W^- production cross-sections in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, arXiv:2505.11310 [hep-ex].

[43] ATLAS, G. Aad *et al.*, *Measurements of Higgs boson production by gluon-gluon fusion and vector-boson fusion using $H \to WW^* \to e\nu\mu\nu$ decays in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Phys. Rev. D **108** (2023) 032005, arXiv:2207.00338 [hep-ex].

[44] CMS, A. M. Sirunyan *et al.*, *Search for a heavy vector resonance decaying to a Z boson and a Higgs boson in proton-proton collisions at $\sqrt{s} = 13$ TeV*, Eur. Phys. J. C **81** (2021) 8, 688, arXiv:2102.08198 [hep-ex].

[45] LHC Higgs Cross Section Working Group, D. de Florian *et al.*, *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, arXiv:1610.07922 [hep-ph].

[46] J. R. Andersen *et al.*, *Les Houches 2015: Physics at TeV Colliders Standard Model Working Group Report*, in *9th Les Houches Workshop on Physics at TeV Colliders*. 5, 2016. arXiv:1605.04692 [hep-ph].

[47] N. Berger *et al.*, *Simplified Template Cross Sections – Stage 1.1 and 1.2*, arXiv:1906.02754 [hep-ph].

[48] S. Amoroso *et al.*, *Les Houches 2019: Physics at TeV Colliders: Standard Model Working Group Report*, in *11th Les Houches Workshop on Physics at TeV Colliders: PhysTeV Les Houches*. 3, 2020. arXiv:2003.01700 [hep-ph].

[49] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer, and T. Stelzer, *MadGraph 5 : Going Beyond*, JHEP **06** (2011) 128, arXiv:1106.0522 [hep-ph].

[50] 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) 9, 096024, arXiv:2008.11743 [hep-ph].

[51] F. Febres Cordero, L. Reina, and D. Wackeroth, *W- and Z-boson production with a massive bottom-quark pair at the Large Hadron Collider*, Phys. Rev. D **80** (2009) 034015, arXiv:0906.1923 [hep-ph].

[52] M. Czakon and A. Mitov, *Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders*, Comput. Phys. Commun. **185** (2014) 2930, arXiv:1112.5675 [hep-ph].

[53] T. Gehrmann, M. Grazzini, S. Kallweit, P. Maierhöfer, A. von Manteuffel, S. Pozzorini, D. Rathlev, and L. Tancredi, *W^+W^- Production at Hadron Colliders in Next to Next to Leading Order QCD*, Phys. Rev. Lett. **113** (2014) 21, 212001, arXiv:1408.5243 [hep-ph].

[54] A. Denner and G. Pelliccioli, *NLO QCD predictions for doubly-polarized WZ production at the LHC*, Phys. Lett. B **814** (2021) 136107, arXiv:2010.07149 [hep-ph].

[55] L. Buonocore, S. Devoto, S. Kallweit, J. Mazzitelli, L. Rottoli, and C. Savoini, *Associated production of a W boson and massive bottom quarks at next-to-next-to-leading order in QCD*, Phys. Rev. D **107** (2023) 7, 074032, arXiv:2212.04954 [hep-ph].

[56] P. Artoisenet, R. Frederix, O. Mattelaer, and R. Rietkerk, *Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations*, JHEP **03** (2013) 015, arXiv:1212.3460 [hep-ph].

[57] 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, arXiv:1410.3012 [hep-ph].

[58] DELPHES 3, J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaître, A. Mertens, and M. Selvaggi, *DELPHES 3, A modular framework for fast simulation of a generic collider experiment*, JHEP **02** (2014) 057, arXiv:1307.6346 [hep-ex].

[59] M. Cacciari, G. P. Salam, and G. Soyez, *FastJet User Manual*, Eur. Phys. J. C **72** (2012) 1896, arXiv:1111.6097 [hep-ph].

[60] J. Butterworth *et al.*, *PDF4LHC recommendations for LHC Run II*, J. Phys. G **43** (2016) 023001, arXiv:1510.03865 [hep-ph].

[61] A. Buckley, J. Ferrando, S. Lloyd, K. Nordström, B. Page, M. Rüfenacht, M. Schönherr, and G. Watt, *LHAPDF6: parton density access in the LHC precision era*, Eur. Phys. J. C **75** (2015) 132, arXiv:1412.7420 [hep-ph].

[62] A. Dedes, W. Materkowska, M. Paraskevas, J. Rosiek, and K. Suxho, *Feynman rules for the Standard Model Effective Field Theory in R_ξ-gauges*, JHEP **06** (2017) 143, arXiv:1704.03888 [hep-ph].