

Machine-learning-inspired SMEFT simplified template cross sections: A case study in ZH production

Daniel Conde[✉], Miguel G. Folgado[✉], and Veronica Sanz[✉]

Instituto de Física Corpuscular (IFIC), 46980 Paterna, Spain, CSIC–Universitat de València, 46101 Valencia, Spain



(Received 1 April 2026; accepted 25 June 2026; published 24 July 2026)

The simplified template cross section (STXS) program has become the standard interface between Higgs measurements and global fits, but its fixed one-dimensional boundaries are not guaranteed to align with the phase-space directions to which the Standard Model effective field theory (SMEFT) is most sensitive. We propose a machine-learning-inspired extension of STXS in which supervised classifiers are used only at the design stage to identify simple, publishable phase-space boundaries. Using associated Higgs production, $pp \rightarrow ZH$, as a case study and a benchmark momentum-dependent bosonic SMEFT deformation, we show that the relevant signal-background separation is well captured by a linear boundary in the (p_T^Z, m_{ZH}) plane. We construct such boundaries with a linear support vector machine and with a deep-neural-network-assisted distillation procedure, and compare them directly with the standard STXS p_T^Z bins through a common single-region Asimov-significance analysis. In this proof-of-concept setup, the machine-learning-inspired regions systematically outperform the corresponding STXS regions, with the largest gains appearing in the boosted regime where SMEFT effects are concentrated. The final observable remains a simple linear cut, preserving the transparency and experimental portability that make STXS useful.

DOI: [10.1103/4zz1-dv63](https://doi.org/10.1103/4zz1-dv63)

I. INTRODUCTION

The Higgs sector remains one of the sharpest probes of physics beyond the Standard Model (SM). In the absence of direct evidence for new particles, the Standard Model effective field theory (SMEFT) provides a systematic language in which small departures from the SM can be organized through higher-dimensional operators suppressed by a heavy scale Λ . Global analyses of Higgs, electroweak, top, and diboson data now routinely use the SMEFT framework to translate precision measurements into constraints on Wilson coefficients; representative examples include Refs. [1,2]. Complementary fitting frameworks and EFT-translation tools commonly used in this context include fitmaker, SMEFTit, HEPfit, and SFitter [1,3–5].

A forewarning: Throughout this paper, process and system labels (ZH, m_{ZH}, VH) use uppercase H following experimental convention. Otherwise, like for the hVV vertex and equations in Sec. II A and Appendix A, H denotes the $SU(2)_L$ Higgs doublet and h the physical scalar after electroweak symmetry breaking. Among Higgs production modes, associated production with an electroweak vector

boson is especially attractive for SMEFT studies because several bosonic operators induce derivative hVV interactions whose effects grow with energy. This feature was already emphasized in early studies of boosted VH production [6–8] and was later quantified in SMEFT analyses of momentum-dependent Higgs-gauge interactions [9–13]. In practice, the relevant distortions populate a correlated high-energy region rather than a single kinematic tail: Events migrate simultaneously toward large vector-boson transverse momentum and large invariant mass of the VH system.

The experimental language used to report Higgs measurements, however, is the simplified template cross section (STXS) framework [14–17]. STXS is successful precisely because it is transparent, experimentally robust, and easy to reinterpret. But that robustness comes with a structural limitation: The regions are predefined and typically built from one-dimensional slices such as p_T^V . In a channel like ZH , where the SMEFT deformation is intrinsically correlated in (p_T^Z, m_{ZH}) , such bins need not be aligned with the most informative direction in phase space.

Machine learning (ML) offers an obvious route to improved discrimination [18]. Supervised classifiers can exploit multidimensional correlations and approximate likelihood-ratio information far beyond what is available in simple cut-based analyses [19–22]. Within this approach, however, we aim for portability over power. A nonlinear classifier output is not, by itself, a good public observable: Output like this depends on training details, architecture

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by SCOAP³.

choices, and internal calibrations, and is therefore poorly matched to the role played by STXS in the experimental-theory interface. The VH study of Ref. [23] illustrates this tension well, where we see that the classifier is powerful, but its output is not a measurement definition one would naturally publish as an STXS stage.

The central idea of this paper is to use machine learning only as a design tool. Rather than publishing an opaque discriminator, we let the classifier identify the physically relevant phase-space direction and then distill that information into a simple linear boundary in a low-dimensional plane. In the present case, the target plane is (p_T^Z, m_{ZH}) chosen because it captures the dominant high-energy deformation induced by momentum-dependent Higgs-gauge operators and because both variables are experimentally natural.

We develop this idea as a proof of concept in $pp \rightarrow ZH$, using a benchmark in which a single bosonic Wilson coefficient is switched on. We consider two constructions. The first is a purely linear support vector machine (SVM) trained directly on (p_T^Z, m_{ZH}) . The second uses a deep neural network (DNN) trained on a broader feature set and then distills its high-score region into a linear boundary in the same two-dimensional plane. In both cases, the final object is not the classifier score but a straight line, directly comparable to the official STXS boundaries.

Our statistical comparison is deliberately narrow and transparent. We compare each standard STXS region to an ML-inspired region within the same p_T^Z slice, and evaluate the resulting single-region sensitivity with the Asimov significance, including representative fractional background uncertainties. This paper therefore does not propose a full replacement for the STXS program, nor does it present a combined experimental analysis. It establishes a more modest point: Even when one insists on a simple publishable region definition, ML-guided boundaries can recover a meaningful fraction of the multidimensional information lost by fixed one-dimensional STXS cuts.

The rest of the paper is organized as follows. In Sec. II, we summarize the SMEFT motivation and the STXS baseline relevant for ZH production. Section III describes the event generation, feature construction, and statistical setup. In Sec. IV, we introduce the ML-inspired region-construction strategies based on a linear SVM and on DNN distillation. Section V presents the boundary comparisons and the significance results. We discuss the scope and limitations of the method in Sec. VI. Technical material is collected in the Appendixes.

II. PHYSICS TARGET AND STXS BASELINE

A. SMEFT origin of the boosted deformation

At dimension six, the SMEFT Lagrangian can be written as

$$\mathcal{L}_{\text{SMEFT}} = \mathcal{L}_{\text{SM}} + \sum_i \frac{c_i}{\Lambda^2} O_i. \quad (1)$$

For associated Higgs production, a particularly relevant class of deformations is generated by bosonic operators that modify the momentum structure of the hVV vertex. In the Warsaw basis, these include

$$O_{HW} = (H^\dagger H) W_{\mu\nu}^i W^{i\mu\nu}, \quad (2)$$

$$O_{HB} = (H^\dagger H) B_{\mu\nu} B^{\mu\nu}, \quad (3)$$

$$O_{HWB} = (H^\dagger \tau^i H) W_{\mu\nu}^i B^{\mu\nu}. \quad (4)$$

After electroweak symmetry breaking, these operators induce derivative interactions of schematic form $hV_{\mu\nu}V^{\mu\nu}$. Their contribution to the amplitude therefore grows with the characteristic event energy, which drives SMEFT effects into the boosted tails of associated production [6–13].

Appendix A gives the corresponding effective hZZ vertex. The three operators above generate the same Lorentz structure up to electroweak mixing factors. This does not prove that every optimized boundary is identical across operator space, but it does explain why the dominant deformation should generically track a correlated high-energy direction. In this paper, we work with a benchmark in which only c_{HW}/Λ^2 is nonzero; our claims are therefore restricted to a proof-of-concept demonstration in that setting.

These considerations motivate the use of p_T^Z and m_{ZH} as the primary variables for region design. The first tracks the boosted electroweak recoil and the second captures the hard scale of the full associated system. As we will see, most of the useful phase-space information can be organized in this two-dimensional plane.

B. Why STXS is the right baseline

The STXS framework was introduced to provide theory-friendly and experimentally stable Higgs measurements [14] and is now standard in ATLAS and CMS analyses [15,16]. Instead of reporting only inclusive rates, STXS partitions the production phase space into exclusive bins with limited extrapolation dependence. For ZH production the relevant stage-1/1.2 categories are organized in intervals of the vector-boson transverse momentum,

$$\begin{aligned} 75 \text{ GeV} < p_T^Z < 150 \text{ GeV}, \quad 150 \text{ GeV} < p_T^Z < 250 \text{ GeV}, \\ 250 \text{ GeV} < p_T^Z < 400 \text{ GeV}, \quad p_T^Z > 400 \text{ GeV}. \end{aligned} \quad (5)$$

These bins are routinely used in SMEFT reinterpretations and therefore form the natural baseline for the present study.

We see that in STXS the p_T^Z slicing is structurally one dimensional, and we suspect that in a channel where the relevant deformation occupies a correlated region of large p_T^Z and large m_{ZH} , a vertical cut in p_T^Z cannot be expected to be optimal. The aim of this work is therefore not to bypass

STXS but to ask whether one can preserve its experimental virtues while allowing the region boundaries themselves to be informed by the data morphology.

III. SIMULATION, OBSERVABLES, AND STATISTICAL SETUP

A. Event generation and benchmark definition

We consider

$$pp \rightarrow ZH \rightarrow \ell^+ \ell^- b \bar{b},$$

with $\ell = e, \mu$ at $\sqrt{s} = 13$ TeV. Events are generated with MadGraph5_aMC@NLO [24] using the SMEFTsim implementation of the Warsaw-basis dimension-six operators [25]. We use NNPDF3.1 [26] throughout.

To keep the proof-of-concept setup as controlled as possible, we compare an SM sample to a benchmark SMEFT sample with

$$\frac{c_{HW}}{\Lambda^2} = 1 \text{ TeV}^{-2}, \quad (6)$$

and all other Wilson coefficients set to zero.

In the statistical treatment below, we use the SM sample as the reference expectation B and the EFT deformation relative to it as the excess S . The significance quoted in this paper should therefore be read as a shape-sensitive benchmark discrimination metric inside a given region. This is in contrast to a fully realistic discovery significance against the complete set of experimental backgrounds. This distinction is essential to the interpretation of the results: The role of the metric is not to provide a full phenomenological sensitivity forecast, and the absolute numerical values of the significance metric should not be read as an experimental reach. The role of using this metric is to place the different region definitions on equal footing and ask which one captures more of the benchmark deformation. This tells us how much benchmark SMEFT-versus-SM separation is retained by a given region definition within the same counting setup.

B. Event selection and observable set

Events are required to contain two isolated, opposite-sign, same-flavor leptons reconstructing the Z boson and two b -tagged jets compatible with the Higgs decay. Standard acceptance and reconstruction cuts are applied to leptons and jets.

From each event, we construct a feature set containing the core kinematic information of the process:

- (i) transverse momenta $p_T^Z, p_T^H, p_T^{\ell_1}, p_T^{\ell_2}, p_T^{b_1}, p_T^{b_2}, p_T^{ZH}$
- (ii) transverse mass m_T^{ZH}
- (iii) invariant masses $m_{\ell\ell}, m_{bb}, m_{ZH}$
- (iv) angular variables $\eta_H, \Delta R_{\ell\ell}, \Delta R_{jH}$, and selected $\Delta\phi_{\ell_i b_j}$ combinations

Many of these observables are correlated. The purpose of the DNN is precisely to test whether those extra correlations materially change the final region design after projection into (p_T^Z, m_{ZH}) . A principal-component analysis is presented in Appendix C; it is consistent with the empirical observation that the dominant SMEFT-SM variation is strongly tied to an energylike direction in feature space.

C. Training samples and feature preprocessing

Each event is represented by a feature vector x and labeled as SM ($y = 0$) or SMEFT benchmark ($y = 1$). The full dataset contains 2×10^6 events. To estimate the stability of the pipeline, we repeat the full analysis over an ensemble of independent random train/test partitions generated from different random seeds. For each repetition, the data are split into training, validation, and test subsets in the ratio 70%:10%:20%. Whenever results in this paper are shown with distributions or quoted uncertainties, these correspond to the spread across that ensemble of repeated splits. The results shown in Sec. V are calculations of each model performed on respective unseen data from the corresponding test set.

Input features are standardized to zero mean and unit variance. Event weights are kept throughout so that the classifiers are trained with the correct physical class weighting rather than on an artificially balanced sample. The linear SVM is trained only on (p_T^Z, m_{ZH}) . The DNN is trained on the broader feature set of Sec. III B, but its output is never proposed as the final public observable, but instead projected down to the (p_T^Z, m_{ZH}) from its output via SVM.

To produce uncertainty estimations of both the boundary parameters and Asimov significance, we produced 20 different bootstrap samples of the original dataset, each with its own train-test-validation splits for which a boundary was found and evaluated in their respective test set. This means that for each methodology pipeline and background-uncertainty estimation, we obtained 20 boundaries from which we estimated both the boundary parameters slope intercept and Asimov-significance uncertainties, each differing from one another by the bootstrap-sample data used.

D. Single-region statistical metric

All comparisons in this paper are performed region by region. For any one region—an official STXS bin, an SVM-defined region, or a DNN-distilled region—we evaluate the sensitivity with the Asimov significance [27]. Without systematics,

$$Z_A = \sqrt{2 \left[(s+b) \ln \left(1 + \frac{s}{b} \right) - s \right]}. \quad (7)$$

Including a fractional background uncertainty $\epsilon = \sigma_B/B$, we use the profiled approximation

TABLE I. Linear SVM boundaries obtained in each official STXS p_T^Z slice. Uncertainties indicate the spread under the resampling procedure used in the analysis.

STXS slice	Slope a	Intercept b (GeV)
[75, 150]	-87 ± 5	9700 ± 500
[150, 250]	-59 ± 4	11700 ± 700
[250, 400]	-28.6 ± 0.6	9500 ± 200
[400, ∞]	-9.1 ± 0.1	6020 ± 50

$$Z_A^2 = 2 \left[(S+B) \ln \frac{(S+B)(B+\sigma_B^2)}{B^2 + (S+B)\sigma_B^2} - \frac{B^2}{\sigma_B^2} \ln \left(1 + \frac{S\sigma_B^2}{B(B+\sigma_B^2)} \right) \right], \quad (8)$$

with $\sigma_B = \epsilon B$. We study the representative values

$$\epsilon \in \{0.3, 0.5, 1.0\}. \quad (9)$$

The benchmark values are not chosen arbitrarily; they are intended to span the order of magnitude of current uncertainties encountered across VH measurements and STXS-like regions, especially once one moves to statistically limited or highly boosted corners of phase space. In this sense, ϵ should be interpreted as an effective analysis-level nuisance scale, not as a detailed decomposition of experimental systematics.

In the two-dimensional linear setting of this proof of concept, the cross-entropy loss and the Asimov-significance optimum are empirically close, so that a significance-aware loss function would not materially change the results when implemented in the DNN pipeline. The alignment between these objectives supports the use of the simpler, more transparent cross-entropy training pipeline instead of replacing the loss function with the Asimov significance, which would require introducing a soft surrogate for the threshold in training and careful implementation.

This choice is intentionally simple. We do not attempt a combined multibin likelihood or a full nuisance-parameter treatment. The paper is designed to answer a narrower question: Given a fixed p_T^Z slice, does an ML-informed region in (p_T^Z, m_{ZH}) capture more of the benchmark SMEFT sensitivity than the corresponding vertical STXS region?

IV. MACHINE-LEARNING-INSPIRED REGION CONSTRUCTION

A. Linear SVM in the (p_T^Z, m_{ZH}) plane

Our most direct construction uses a linear support vector machine trained on the two variables (p_T^Z, m_{ZH}) . We use the LinearSVC implementation in SCIKIT-LEARN with L_2 regularization, tolerance 10^{-4} , and a maximum of 10^4 iterations. Event weights proportional to the physical cross sections are included during training.

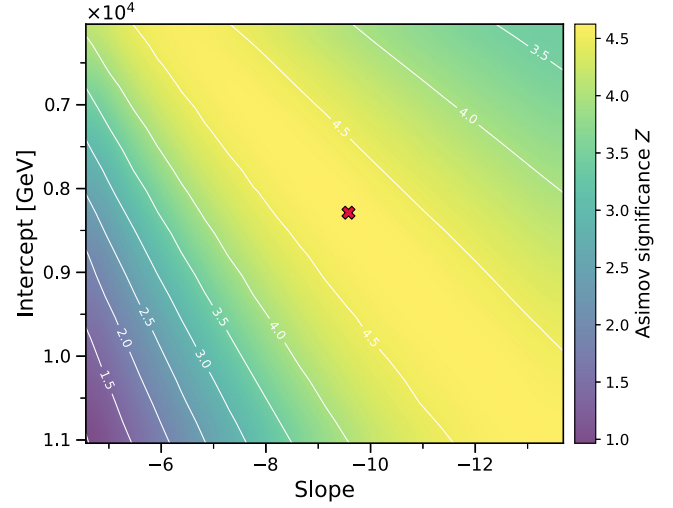


FIG. 1. Asimov significance in the slope-intercept plane for the scan around the SVM solution in the highest- p_T^Z slice, for a representative choice $\epsilon = 0.5$.

The classifier returns a linear separator in the standardized feature space, which we rescale and express in physical units as

$$f(p_T^Z) = ap_T^Z + b. \quad (10)$$

One side of the line $m_{ZH} > f$ is defined as the EFT-enriched region and the other $m_{ZH} < f$ as the SM-like region. To maintain a clean comparison with the official STXS setup, we construct one such boundary inside each of the four standard p_T^Z intervals of Eq. (5). The resulting parameters are listed in Table I.

The physical content of the result is already visible here: The learned boundary is tilted, selecting events that are simultaneously hard in p_T^Z and m_{ZH} , rather than merely large in one variable.

TABLE II. Significance-optimized SVM boundaries. The scan is performed separately for each STXS slice and each assumed fractional background uncertainty.

STXS slice	σ_B/B	Slope a	Intercept b (GeV)
[75, 150]	0.3	-78 ± 7	11600 ± 900
[75, 150]	0.5	-80 ± 10	12000 ± 2000
[75, 150]	1.0	-59 ± 10	11000 ± 1000
[150, 250]	0.3	-61 ± 6	15000 ± 1000
[150, 250]	0.5	-56 ± 6	14000 ± 1000
[150, 250]	1.0	-48 ± 7	13000 ± 2000
[250, 400]	0.3	-33 ± 3	12000 ± 1000
[250, 400]	0.5	-32 ± 4	12000 ± 1000
[250, 400]	1.0	-24 ± 2	10200 ± 800
[400, ∞]	0.3	-10.0 ± 0.9	7300 ± 600
[400, ∞]	0.5	-9.5 ± 0.3	8200 ± 200
[400, ∞]	1.0	-8.0 ± 0.8	9000 ± 700

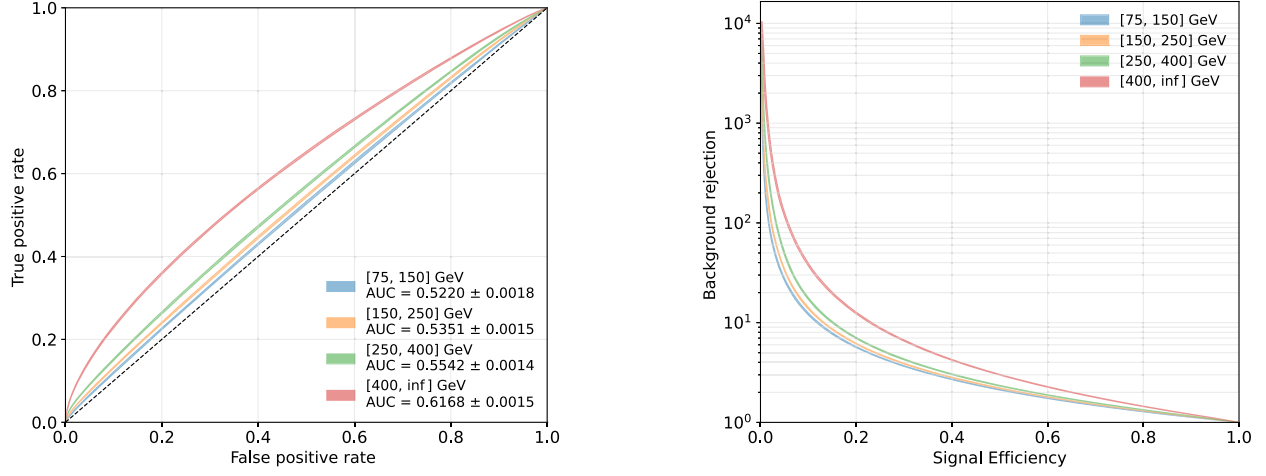


FIG. 2. Classifier performance inside the four official STXS p_T^Z slices. Left: receiver operating characteristic (ROC) curves for SM-benchmark discrimination with a shallow classifier trained separately in each slice. Right: background efficiency versus signal efficiency. The increasing separability toward large p_T^Z motivates the focus on the boosted region.

B. Significance-driven refinement of the linear boundary

The linear SVM is trained to maximize classification performance, not directly the significance metric of Eqs. (7) and (8). Since our aim is to optimize for the latter, we therefore perform a local scan in slope and intercept around the SVM solution and select the boundary that maximizes the Asimov significance on the training set for a fixed choice of ϵ .

Operationally, for each STXS slice and each value of ϵ we scan a family of lines around the SVM seed, compute the corresponding (s, b) yields, evaluate Z_A , and retain the maximizing point. Figure 1 shows the resulting significance landscape in the highest- p_T^Z slice for $\epsilon = 0.5$.

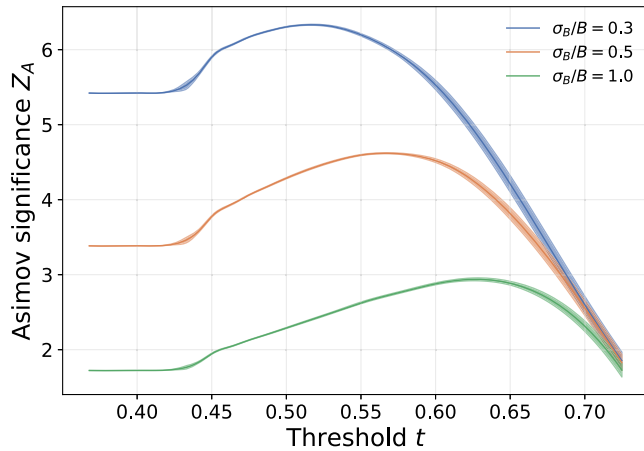


FIG. 3. Asimov significance as a function of the DNN-score threshold used to define the projected high-score region in the highest- p_T^Z slice.

TABLE III. Linear boundaries obtained by DNN-guided distillation into the (p_T^Z, m_{ZH}) plane.

STXS slice	σ_B/B	Slope a	Intercept b (GeV)
[75, 150]	0.3	-80 ± 40	12000 ± 5000
[75, 150]	0.5	-80 ± 40	12000 ± 5000
[75, 150]	1.0	-70 ± 40	12000 ± 5000
[150, 250]	0.3	-60 ± 20	14000 ± 4000
[150, 250]	0.5	-60 ± 20	14000 ± 4000
[150, 250]	1.0	-50 ± 20	13000 ± 3000
[250, 400]	0.3	-35 ± 4	13000 ± 1000
[250, 400]	0.5	-32 ± 4	12000 ± 1000
[250, 400]	1.0	-28 ± 4	12000 ± 1000
[400, ∞]	0.3	-12 ± 1	8600 ± 600
[400, ∞]	0.5	-11.2 ± 0.9	9700 ± 700
[400, ∞]	1.0	-10.6 ± 0.9	10800 ± 900

TABLE IV. DNN-distilled boundaries after the additional significance-driven slope-intercept scan.

STXS slice	σ_B/B	Slope a	Intercept b (GeV)
[75, 150]	0.3	-80 ± 30	12000 ± 5000
[75, 150]	0.5	-60 ± 40	11000 ± 6000
[75, 150]	1.0	-60 ± 30	11000 ± 5000
[150, 250]	0.3	-70 ± 10	16000 ± 3000
[150, 250]	0.5	-55 ± 6	14000 ± 2000
[150, 250]	1.0	-50 ± 20	13000 ± 4000
[250, 400]	0.3	-33 ± 5	12000 ± 2000
[250, 400]	0.5	-31 ± 5	12000 ± 2000
[250, 400]	1.0	-23 ± 5	10000 ± 2000
[400, ∞]	0.3	-10.8 ± 0.6	7800 ± 400
[400, ∞]	0.5	-9.7 ± 0.6	8400 ± 400
[400, ∞]	1.0	-9 ± 1	10000 ± 1000

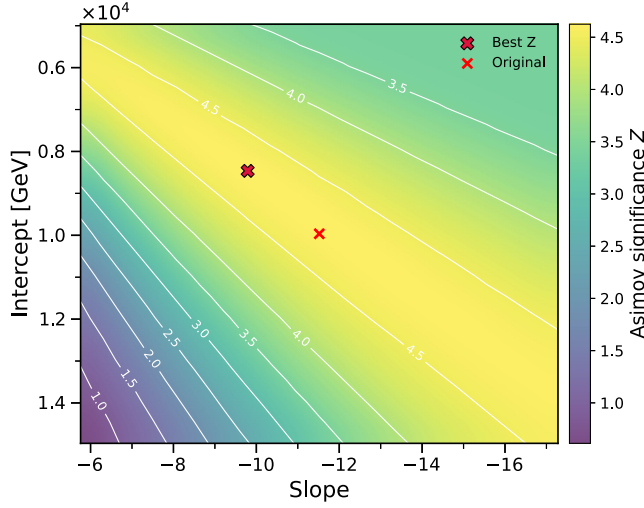


FIG. 4. Significance landscape for the postfit slope-intercept scan applied to the DNN-distilled boundary in the highest- p_T^Z slice, for $\epsilon = 0.5$.

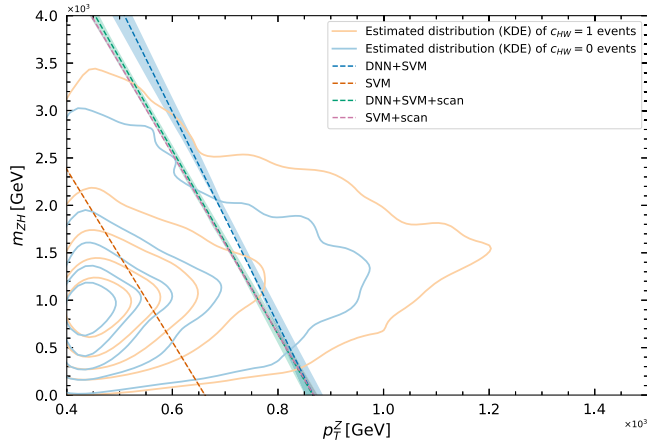


FIG. 5. Comparison of the official STXS boundary with the ML-inspired linear boundaries in the highest- p_T^Z slice shown for the representative background-uncertainty choices used in the analysis. The continuous lines represent the kernel density estimation of the distribution of the events in the (p_T^Z, m_{ZH}) space, classified by color for $c_{HW} = 0, 1$. The ML-guided straight dashed lines represent the final boundary obtained for each one of the proposed methods.

The optimized parameters are reported in Table II. Beyond the numerical values reported in the table, the

main point is the stability of the preferred orientation. Once significance is used as the objective instead of classification performance, the region remains aligned with the correlated high-energy tail.

C. DNN-guided distillation into a linear region

The second strategy asks whether a broader observable set materially changes the final low-dimensional region. We train a feedforward DNN on the full feature set of Sec. III B. The network uses three hidden layers of sizes (126, 84, 42) with ReLU activations, batch normalization after each hidden layer, Adam optimization with learning rate 10^{-4} , weighted binary cross-entropy loss, and early stopping monitored on the validation accuracy.

The classifier output

$$s_{\text{DNN}}(\mathbf{x}) \in [0, 1] \quad (11)$$

is used only as an internal ranking variable. It is not the final physics observable. Instead, for a threshold t , we select events with $s_{\text{DNN}}(\mathbf{x}) > t$, project that subset onto the (p_T^Z, m_{ZH}) plane, and fit a linear SVM to obtain a straight-line approximation of the high-score region. Scanning over t therefore produces a family of linear boundaries,

$$m_{ZH} = a_t p_T^Z + b_t, \quad (12)$$

from which we choose the one that maximizes the Asimov significance for a given ϵ .

This procedure has two advantages. First, it lets the DNN exploit correlations beyond the two-dimensional plane. Second, it forces the final output back into a form that could plausibly be published as an STXS-like region definition. Figure 2 shows the intrinsic classifier separation within each STXS slice, while Fig. 3 illustrates the threshold optimization for the highest- p_T^Z bin.

The distilled boundary parameters are shown in Table III. In practice, they are steeper than the STXS vertical cuts and comparable to the SVM-based solutions, indicating that the DNN confirms rather than overturns the picture suggested by the two-dimensional training.

For completeness, we also apply the same postfit slope-intercept scan used in the SVM case. The resulting parameters are listed in Table IV, and a representative scan is shown in Fig. 4. The numerical changes are modest,

TABLE V. Asimov significance for $\sigma_B/B = 0.3$. Each entry corresponds to a single-region comparison within a fixed STXS p_T^Z slice.

Method	Bin 1	Bin 2	Bin 3	Bin 4
STXS	2.244 ± 0.001	2.819 ± 0.001	3.656 ± 0.001	5.424 ± 0.001
SVM	2.393 ± 0.010	2.988 ± 0.009	3.911 ± 0.013	6.204 ± 0.016
DNN + SVM	2.529 ± 0.045	3.086 ± 0.038	3.983 ± 0.024	6.328 ± 0.023
SVM + scan	2.531 ± 0.051	3.084 ± 0.047	3.992 ± 0.027	6.324 ± 0.023
DNN + SVM + scan	2.515 ± 0.050	3.085 ± 0.038	3.991 ± 0.024	6.328 ± 0.023

TABLE VI. Asimov significance for $\sigma_B/B = 0.5$.

Method	Bin 1	Bin 2	Bin 3	Bin 4
STXS	1.347 ± 0.001	1.695 ± 0.001	2.216 ± 0.001	3.387 ± 0.001
SVM	1.438 ± 0.006	1.803 ± 0.006	2.401 ± 0.008	4.114 ± 0.011
DNN + SVM	1.514 ± 0.051	1.904 ± 0.027	2.519 ± 0.019	4.611 ± 0.025
SVM + scan	1.532 ± 0.058	1.907 ± 0.032	2.511 ± 0.019	4.631 ± 0.026
DNN + SVM + scan	1.494 ± 0.077	1.910 ± 0.033	2.518 ± 0.018	4.628 ± 0.026

TABLE VII. Asimov significance for $\sigma_B/B = 1.0$.

Method	Bin 1	Bin 2	Bin 3	Bin 4
STXS	0.674 ± 0.0003	0.849 ± 0.0004	1.113 ± 0.0003	1.724 ± 0.0003
SVM	0.720 ± 0.003	0.904 ± 0.003	1.212 ± 0.004	2.163 ± 0.006
DNN + SVM	0.806 ± 0.082	0.994 ± 0.049	1.330 ± 0.017	2.912 ± 0.024
SVM + scan	0.784 ± 0.072	0.973 ± 0.050	1.336 ± 0.024	2.952 ± 0.027
DNN + SVM + scan	0.772 ± 0.091	0.968 ± 0.042	1.328 ± 0.023	2.953 ± 0.026

which indicates that the threshold optimization already lands close to the significance optimum.

V. RESULTS

A. Comparison of the learned boundaries

Figure 5 compares, in the highest- p_T^Z slice, the official STXS region boundary with the linear SVM and DNN-distilled alternatives. Several qualitative features are robust across methods. First, both ML-inspired constructions tilt toward the correlated hard tail, rather than following a vertical cut. Second, the optimized lines move toward more selective regions as the assumed fractional uncertainty increases, reflecting the fact that significance favors a purer region when the background is less well controlled. Third, the difference between the DNN-distilled boundary and the SVM-based one is noticeable but not dramatic. Once projected into (p_T^Z, m_{ZH}) , the core geometric picture is already visible in the two-dimensional training.

This is the central geometric result of the paper. We explicitly set out to discover the most physically sensible correlated direction that can be encoded in a single straight line, instead of proposing an exotic nonlinear classifier.

B. Single-region significance comparison

Tables V–VII together with the box plots in Fig. 6 summarize the Asimov significance obtained with the official STXS region and with the ML-inspired alternatives inside each of the four p_T^Z slices. The quoted uncertainties come from the bootstrap procedure used in the numerical study. The three values of ϵ should be read as effective uncertainty levels motivated by the range currently encountered in VH analyses, from relatively controlled regions to more uncertainty-dominated boosted configurations.

The linear SVM improves over the official STXS region in every slice and for every uncertainty choice shown. The

gain is modest in the lowest- p_T^Z region and grows steadily toward the boosted regime, exactly where the benchmark SMEFT deformation is expected to be most concentrated.

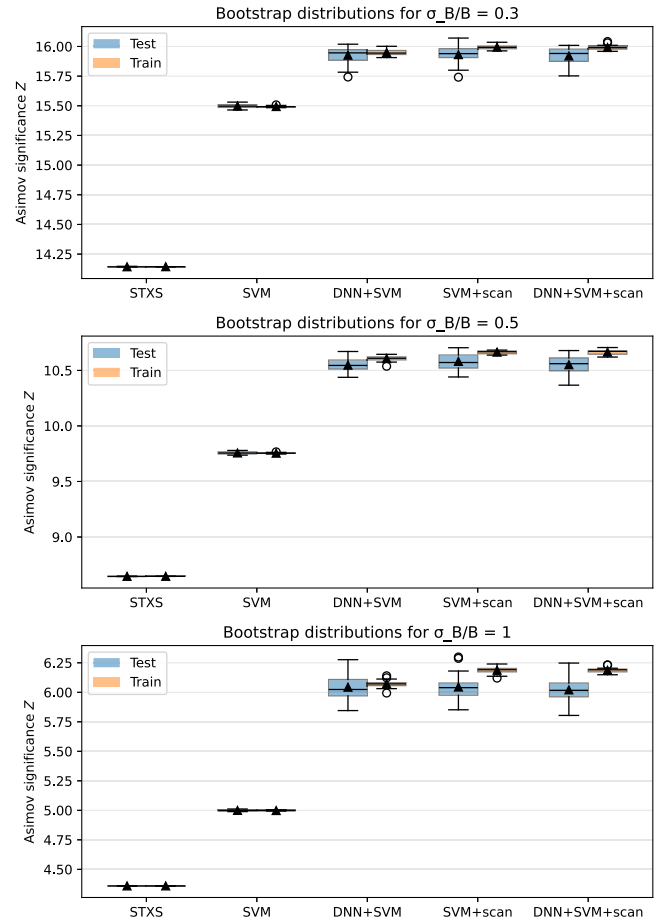


FIG. 6. Distribution of the final Asimov significance yields for each method summed over all bin regions. The distributions are obtained from the bootstrap procedure on both sample set and models.

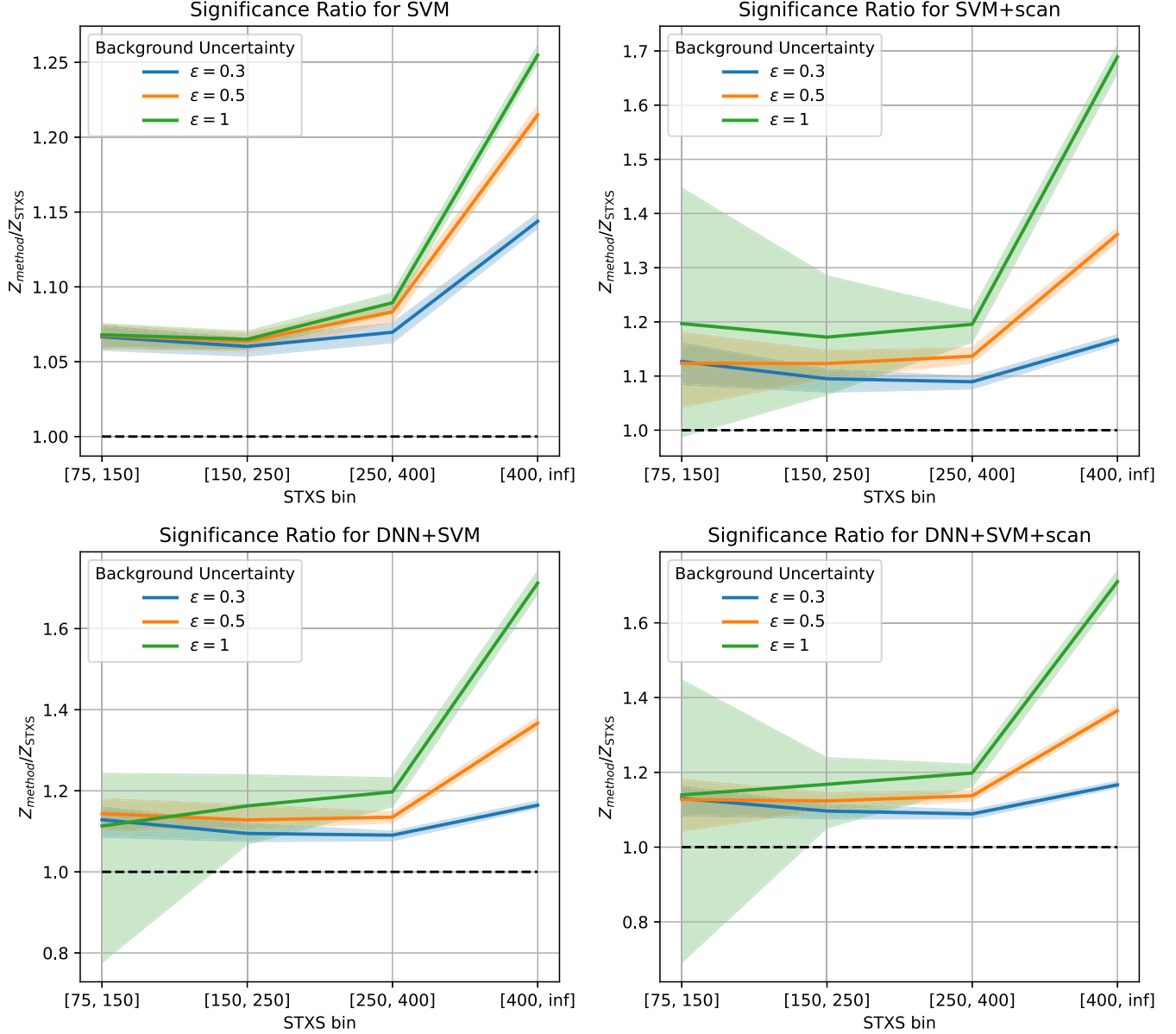


FIG. 7. Ratio of the Asimov significance obtained with each ML-inspired region to that of the corresponding official STXS region, for the different p_T^Z slices and uncertainty assumptions considered in the text.

The DNN-distilled boundary usually improves further, especially in the third and fourth slices. For the representative choice $\epsilon = 0.5$, the best ML-inspired boundary in the highest- p_T^Z slice reaches $Z_A \simeq 4.63$ compared with 3.39 for the corresponding STXS region, an increase of about 37%. For $\epsilon = 1.0$ the same comparison gives an increase of about 71%. The exact percentage is therefore uncertainty-model dependent, and it is more accurate to speak of a substantial gain in the boosted region than to quote a single universal number.

The postfit slope-intercept scan produces only marginal changes relative to the threshold-optimized DNN-assisted result. This is useful methodologically because it shows that the gain does not rely on aggressive fine-tuning of the

final line parameters. We observe evidence of overfit when comparing the box plots of the test set and the training set in Fig. 6. Even though it appears that, within uncertainty, the test results are compatible to the training results, we see that the mean of the box plots is higher and their spreads smaller, sometimes so much so that the lower bounds of the training results outperform three quartiles of the test results (as in the case of the “SVM + scan” and “DNN + SVM + scan” box plots for $\sigma_B/B = 1$ in the lower graph of Fig. 6). This is more prominent in the cases for higher background uncertainty, and, tellingly, for the procedures with postfit scans. We may conclude then that the threshold-optimized procedure (labeled “DNN + SVM”) might be more resistant to overfit when

compared to the grid search scans in parameter space; for the latter, the reason could be that the Asimov significance is an optimistically biased estimate of the true performance. The test-set results represented in the box plots and in the tables of this section should be taken as the more appropriate estimation of generalization performance, and the fact that they are compatible is evidence that the overfit is within the statistical noise of the procedure.

The larger bootstrap spread of the DNN-based methods in the lowest two slices suggests caution. The DNN still improves on the STXS baseline there, but the practical leverage is clearly concentrated in the high-energy bins. That is also the region where the proof of concept is most compelling physically.

The significance ratios relative to the STXS baseline are shown in Fig. 7. We recover the observations from before but relative to STXS: that significance yields increase in the high-energy regimes. However, the relative gain of significance in high background-uncertainty scenarios is much greater, despite the overall yield being low. A familiar observation emerges when comparing the significance-optimized methods—SVM + scan, DNN + SVM, and DNN + SVM + scan—to the simple SVM fit, for there is an increase in variance. This is markedly significant for the low bin and high background-uncertainty case, in which some bootstrap experiments showed no improvement over STXS. We see overall that the advantage of the ML-inspired regions is systematic and becomes strongest precisely where the fixed one-dimensional slicing is least adapted to the signal morphology.

The results above support three claims and only these three claims.

First, the standard STXS segmentation is not optimally aligned with the benchmark SMEFT deformation studied here. Second, a simple linear cut in (p_T^Z, m_{ZH}) already recovers a nontrivial fraction of the multidimensional information that the vertical STXS slicing leaves unused. The final low-dimensional region remains simple and interpretable.

What is not established here is a full replacement of STXS, a channel-independent universal boundary, or a complete experimental sensitivity projection. The paper should be read as a targeted proof of concept showing that publishable region definitions can be improved with machine-learning guidance without becoming opaque.

VI. CONCLUSIONS

We have presented a proof-of-concept extension of the STXS philosophy in which machine learning is used to design simplified regions rather than to replace them with opaque classifier outputs. Using $pp \rightarrow ZH$ as a benchmark and focusing on a momentum-dependent bosonic SMEFT deformation, we showed that the relevant SM-SMEFT separation is organized along a correlated direction in (p_T^Z, m_{ZH}) . The most attractive feature of the proposed

construction is the fact that the final object is a low-parameter cut that can be documented, reproduced, and interpreted in exactly the way STXS regions are interpreted now, but with the phase-space direction backed up by a robust ML procedure.

A linear SVM trained directly on that plane already defines a region that systematically outperforms the corresponding official STXS region inside each p_T^Z slice. A DNN trained on a broader observable also improves with respect to the STXS cuts, as much as a specialized significance optimization of the direct SVM output. The fact that these two latter approaches give similar results suggests that as long as the ML procedure takes significance optimization into account, the use of more kinetic observables in the DNN does not improve the significance yield much further. Seeing it from the other side, this result also shows that the gain does not rely on aggressive adjustment of the final line parameters, and so the DNN-assisted approach requires less fine-tuning.

We note that in the specific two-dimensional linear setting considered here, the scan over slope and intercept is itself a viable stand-alone procedure. The SVM and DNN steps contribute primarily as a principled initialization that identifies a good starting region in the parameter space and validates that a linear boundary in (p_T^Z, m_{ZH}) is close to the significance optimum. This reduces further scans in parameter space to a local refinement rather than a global search, which, additionally may be more susceptible to overfit than the proposed pipelines.

For the representative background-uncertainty choice $\sigma_B/B = 0.5$, the gain in the highest- p_T^Z slice is about 37%, and it is larger for more pessimistic uncertainty assumptions. The exact number is not universal, but the pattern is robust: The largest benefit appears precisely in the boosted regime where the benchmark SMEFT deformation is most concentrated. Likewise, we see that the regime where this happens is the higher- p_T^Z bins, consistent with the enhancement of the hZZ vertex of high-energy regions in ZH production.

For ZH production, this separation works because the underlying physics is simple: The benchmark SMEFT deformation enhances a correlated boosted tail, and a straight line in (p_T^Z, m_{ZH}) is already a good approximation to the relevant geometry. The DNN confirms this picture rather than replacing it with a qualitatively different one. That is an encouraging sign for experimental usability. The next step is to test how far this philosophy generalizes across operators, across channels, and within more realistic experimental likelihoods. Even at its present level, however, this study makes a concrete point that if STXS is meant to be the bridge between precision measurements and EFT inference, then allowing machine learning to optimize the shape of that bridge is a natural next step.

The authors also acknowledge the developers of MadGraph5_aMC@NLO, SMEFTsim, and open-source Python

libraries including SCIKIT-LEARN, PyTorch, and NumPy used throughout this study.

ACKNOWLEDGMENTS

We thank members of the LHC EFT and Higgs working groups for valuable discussions. This work is supported by the Spanish Grants No. PID2023-148162NB-C21, No. CNS-2022-135688, No. CEX2023-001292-S, and No. MMT24-IFIC-01, which comes from the European Union's Recovery and Resilience Facility–Next Generation, in the framework of the General Invitation of the Spanish Government's public business entity Red.es to participate in talent attraction and retention programs within Investment 4 of Component 19 of the Recovery, Transformation and Resilience Plan. Computational resources were provided by the ARTEMISA cluster at the Universitat de València. All authors contributed equally to this work.

DATA AVAILABILITY

The data that support the findings of this article are openly available [28].

APPENDIX A: EFFECTIVE hZZ COUPLING FROM DIMENSION-SIX OPERATORS

In this Appendix, we collect the momentum-dependent hZZ interaction induced by the bosonic SMEFT operators most relevant to associated Higgs production. In the Warsaw basis,

$$\begin{aligned}\mathcal{O}_{HW} &= (H^\dagger H) W_{\mu\nu}^i W^{i\mu\nu}, & \mathcal{O}_{HB} &= (H^\dagger H) B_{\mu\nu} B^{\mu\nu}, \\ \mathcal{O}_{HWB} &= (H^\dagger \tau^i H) W_{\mu\nu}^i B^{\mu\nu}.\end{aligned}\quad (\text{A1})$$

After electroweak symmetry breaking $H = (0, (v+h)/\sqrt{2})^T$, the terms linear in the Higgs field are

$$\mathcal{L} \supset \frac{vh}{\Lambda^2} \left[c_{HW} W_{\mu\nu}^i W^{i\mu\nu} + c_{HB} B_{\mu\nu} B^{\mu\nu} - \frac{1}{2} c_{HWB} W_{\mu\nu}^3 B^{\mu\nu} \right]. \quad (\text{A2})$$

Using

$$W_{\mu\nu}^3 = c_w Z_{\mu\nu} + s_w A_{\mu\nu}, \quad B_{\mu\nu} = c_w A_{\mu\nu} - s_w Z_{\mu\nu},$$

we obtain

$$\mathcal{L} \supset h \left[\kappa_{ZZ} Z_{\mu\nu} Z^{\mu\nu} + \kappa_{ZA} Z_{\mu\nu} A^{\mu\nu} + \kappa_{AA} A_{\mu\nu} A^{\mu\nu} \right], \quad (\text{A3})$$

with

$$\kappa_{ZZ} = \frac{v}{\Lambda^2} \left(c_{HW} c_w^2 + c_{HB} s_w^2 + \frac{1}{2} c_{HWB} s_w c_w \right), \quad (\text{A4})$$

$$\kappa_{ZA} = \frac{v}{\Lambda^2} \left[2s_w c_w (c_{HW} - c_{HB}) - \frac{1}{2} c_{HWB} (c_w^2 - s_w^2) \right], \quad (\text{A5})$$

$$\kappa_{AA} = \frac{v}{\Lambda^2} \left(c_{HW} s_w^2 + c_{HB} c_w^2 - \frac{1}{2} c_{HWB} s_w c_w \right). \quad (\text{A6})$$

The corresponding hZZ vertex, with all momenta incoming, is

$$V_{hZZ}^{\mu\nu}(k_1, k_2) = i2\kappa_{ZZ} [(k_1 \cdot k_2) g^{\mu\nu} - k_1^\nu k_2^\mu]. \quad (\text{A7})$$

The common Lorentz structure explains why this class of operators generically enhances correlated high-energy regions in ZH production.

APPENDIX B: DNN ARCHITECTURE AND TRAINING DETAILS

The DNN is used only as an intermediate representation from which a simple linear region can be distilled. It is therefore deliberately lightweight. The model is a fully connected multilayer perceptron with three hidden layers of sizes (126, 84, 42) and ReLU activations. The output layer consists of a single sigmoid neuron producing a score $s_{\text{DNN}}(x) \in [0, 1]$.

The hidden-layer widths were chosen from the scaling pattern $(9m_{\text{hl}}, 6m_{\text{hl}}, 3m_{\text{hl}})$, with the choice $m_{\text{hl}} = 14$ adopted in the final setup. Inputs are standardized to zero mean and unit variance. Training uses the Adam optimizer with learning rate 10^{-4} and weighted binary cross-entropy, where event weights encode the physical normalization of the SM and SMEFT samples. The network is trained for up to 200 epochs with batch size 1024, using early stopping on the validation loss with patience 3 and restoring the best weights.

After training, the DNN score is thresholded, the selected events are projected onto the (p_T^Z, m_{ZH}) plane, and a linear SVM is fitted to the selected subset. Scanning the threshold yields the family of DNN-guided linear boundaries discussed in Sec. IV C.

APPENDIX C: PRINCIPAL-COMPONENT ANALYSIS OF THE OBSERVABLE SET

To characterize the dominant directions of variation in the full observable set, we perform a principal-component analysis on the standardized features entering the DNN. The analysis is carried out separately in each official STXS p_T^Z slice. The purpose is to check the expressivity of each variable in the high-dimensional dataset.

Figure 8 shows that the leading principal component dominates the variance in every slice, with increasing weight in the boosted regime. Figure 9 shows that this leading component is driven primarily by invariant-mass- and transverse-momentum-related observables, while the next component carries more topological information. This

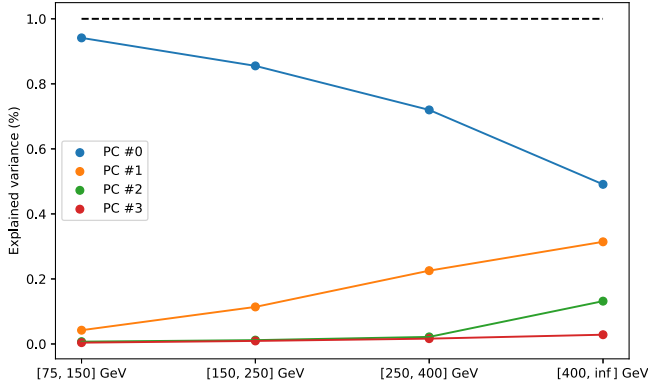


FIG. 8. Fraction of explained variance carried by the first four principal components in each official STXS p_T^Z slice. The leading component dominates throughout and becomes more pronounced in the boosted region.

is further motivation for a linear cut in (p_T^Z, m_{ZH}) which is expected to capture much of the useful structure.

APPENDIX D: GENERALIZATION TO 3D

The proof of concept presented in the main text is devoted to a single linear boundary in the two-dimensional (p_T^Z, m_{ZH}) plane. A natural question is whether the relative advantage of the ML-inspired pipelines over a direct parameter scan persist as the dimensionality of the search space grows. In this Appendix, we present a limited generalization to a three-dimensional boundary in order to address this question directly.

We extend the boundary parametrization of Eq. (10) to a hyperplane of the form

$$z = ax + by + c, \quad x \equiv p_T^Z, \quad y \equiv m_{ZH}, \quad z \equiv m_T^{ZH}, \quad (\text{D1})$$

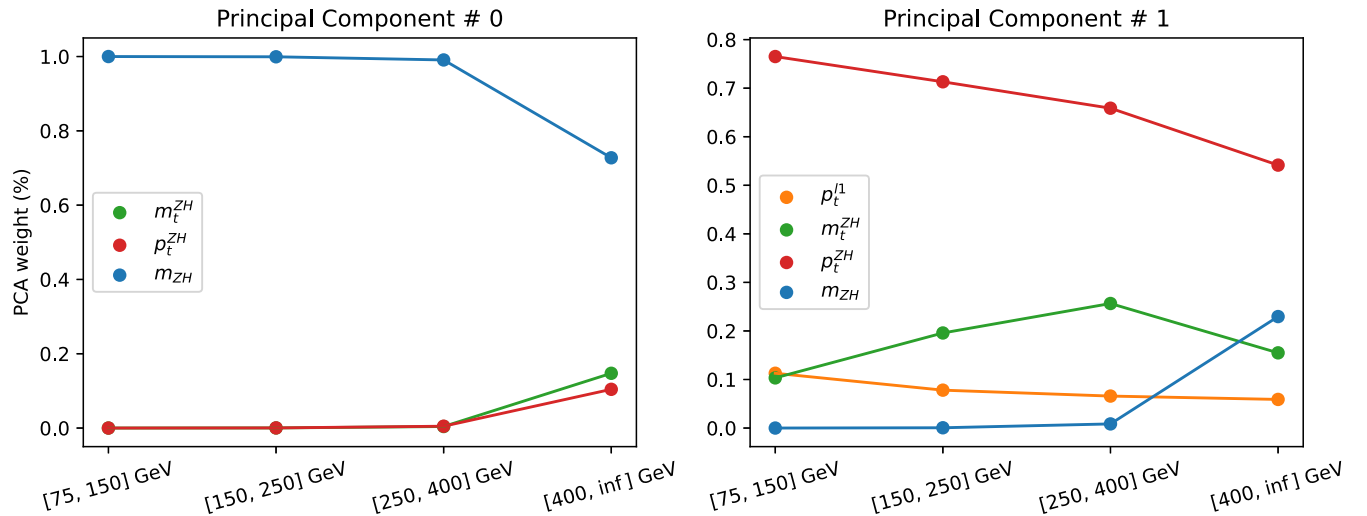


FIG. 9. Principal-component weights for the two leading components shown separately in each STXS slice. The first component is dominated by energylike observables supporting the use of (p_T^Z, m_{ZH}) as the plane in which to define the distilled linear boundary.

where the third variable m_T^{ZH} was selected on the basis of the principal-component analysis of Appendix D, which identifies m_T^{ZH} together with p_T^{ZH} as the directions carrying the most additional information beyond (p_T^Z, m_{ZH}) . We note that the specific choice of m_T^{ZH} may not be optimal for discriminating the benchmark c_{HW} deformation, which we return to when interpreting the results below.

We compare five constructions of the three-dimensional boundary, in addition to the official STXS baseline. We detail each of the methods used to obtain the boundaries, as a basis to later estimate the computational cost of each.

- (1) Scan: A global grid scan over the hyperplane parameters $a \in [-50, 50] \text{ GeV}^{-1}$, $b \in [-50, 50] \text{ GeV}^{-1}$, and $c \in [-500, 1000] \text{ GeV}$ evaluated on a $10 \times 10 \times 10$ grid ($10^3 = 1000$ evaluations of the Asimov significance) and optimized directly against this objective, with no machine-learning component.
- (2) Scan20: Identical to Scan, but with the grid resolution increased to $20 \times 20 \times 20$ points ($20^3 = 8000$ evaluations) over the same parameter ranges, included to test whether the failure of the global scan is a resolution artifact or a more fundamental limitation of the unguided search.
- (3) SVM: A linear support vector machine trained on the three input variables, optimizing classification accuracy (a single fit, with no scan over the Asimov significance), exactly as in the two-dimensional case of Sec. IV A.
- (4) SVM + scan: A refinement of the SVM solution via a local grid scan ($10 \times 10 \times 10$ grid, 1000 evaluations) over a window centered on the SVM seed $a, b \in [0, 1.5a_{\text{SVM}}]$ or $[1.5a_{\text{SVM}}, 0]$ depending on the sign of the seed value (and analogously for b), and $c \in [c_{\text{SVM}} - 500, c_{\text{SVM}} + 500] \text{ GeV}$ optimizing the Asimov significance.

TABLE VIII. Cost of each 3D boundary-construction method decomposed into the number of Asimov-significance evaluations on a fixed candidate boundary (each ≈ 0.05 s), the number of linear SVM fits (each ≈ 0.8 s), and the number of DNN fits (each ≈ 190 s).

Method	Significance evals.	SVM fits	DNN fits	Total cost
SVM	0	1	0	≈ 1 s
Scan	1,000	0	0	≈ 50 s
Scan20	8,000	0	0	≈ 400 s
SVM + scan	1,000	1	0	≈ 50 s
DNN + SVM	20	20	1	≈ 200 s

- (5) DNN + SVM: The same DNN-distillation procedure as in Sec. IV C, using the existing network trained on the full feature set of Sec. III B (no retraining required, since the network was not restricted to the two-dimensional plane), with 20 linear fits SVM taking the three variables and representative score thresholds (5% percentile spacing), and the optimization of Asimov significance scanning over the 20 SVM boundaries found.

As in the main text, all comparisons use the single-region Asimov significance of Eq. (8) for the three representative background-uncertainty values $\epsilon \in \{0.3, 0.5, 1.0\}$ evaluated on a held-out test partition, with five bootstrap repetitions used to estimate the spread of each result.

We report cost with respect to three elementary operations whose counts are fixed by the design of each procedure and are independent of the run-time environment: the number of Asimov-significance evaluations on a fixed candidate boundary, the number of linear SVM fits, and the number of DNN fits. From dedicated timing on the same hardware, a single significance evaluation takes approximately 0.05 s, and a single linear SVM fit takes approximately 0.8 s, and a single DNN built takes approximately 190 s. Table VIII reports the number of each operation required by every method, together with the resulting total cost.

We note that, in our setup, Scan and SVM + scan both require 1000 significance evaluations and are therefore comparable in raw computational cost (the single additional SVM fit in SVM + scan for this linear boundary in 3D is negligible against 1000 significance evaluations). With regard to DNN + SVM, this method is amortized by the previous 2D results: Given an already realized DNN, additional boundaries cost only 20 significance evaluations and 20 SVM fits totaling ≈ 2 s. This makes the DNN + SVM pipeline greatly generalizable because the DNN fit was done on all kinematic variables and can be

TABLE IX. Test-set Asimov significance (mean $\pm$ bootstrap standard deviation across five repetitions) summed over the four STXS p_T^Z slices, for each method and each representative background uncertainty ϵ .

Method	$\epsilon = 0.3$	$\epsilon = 0.5$	$\epsilon = 1.0$
STXS	14.143 ± 0.0004	8.645 ± 0.0003	4.359 ± 0.0001
Scan	9.170 ± 0.013	8.645 ± 0.0004	2.433 ± 0.002
Scan20	14.825 ± 0.008	9.482 ± 0.023	5.117 ± 0.029
SVM	15.394 ± 0.052	9.689 ± 0.033	4.964 ± 0.017
SVM + scan	15.871 ± 0.035	10.576 ± 0.050	5.972 ± 0.094
DNN + SVM	15.821 ± 0.036	10.553 ± 0.034	6.046 ± 0.012

used for any number of increments in the hyperplane parameter space.

Table IX reports the test-set Asimov significance for each method, summed across all four STXS p_T^Z slices, for the three representative values of ϵ , following the same bootstrap procedure used to produce the box plots of Fig. 6 in the main text.

The unguided global scan (Scan) performs poorly relative to the ML-seeded methods, and at $\epsilon = 1.0$ falls visibly below the STXS baseline (2.433 versus 4.359), despite requiring comparable computational cost. Increasing the grid resolution by a factor of 8 in evaluation count (Scan20, 8000 evaluations) recovers performance broadly comparable to the single-fit SVM, but does not reach the level of either SVM or DNN-seeded refinement. This is in contrast to the two-dimensional case discussed in the main text where a local scan around the SVM seed was already sufficient to closely approach the significance optimum. The three-dimensional result indicates that, as the boundary parameter space grows, an undirected search becomes substantially less reliable per unit of computational cost, supporting the use of machine learning to constrain the relevant region of parameter space before any refinement is attempted. These methods provide a disciplined and scalable route to identifying simple, publishable phase-space boundaries, in a setting where an unguided search is demonstrably less reliable, and less cost effective.

We additionally see that the overall gain of the 3D methods relative to the corresponding 2D results of the main text is modest. We attribute this primarily to the choice of m_T^{ZH} as the third variable, which may not be the most discriminating choice for the benchmark c_{HW} deformation considered here. A more thorough study of variable choice for the 3D extension is left to future work.

- [1] J. Ellis, M. Madigan, K. Mimasu, V. Sanz, and T. You, Top, Higgs, diboson and electroweak fit to the Standard Model effective field theory, *J. High Energy Phys.* **04** (2021) 279.
- [2] A. Falkowski and D. Straub, Flavourful SMEFT likelihood for Higgs and electroweak data, *J. High Energy Phys.* **04** (2020) 066.
- [3] 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**, 393 (2023).
- [4] J. de Blas *et al.*, Current and future constraints on Higgs couplings in the nonlinear Effective Theory, *J. High Energy Phys.* **07** (2018) 048.
- [5] J. Ellis and T. You, Updated global analysis of Higgs couplings, *J. High Energy Phys.* **06** (2013) 103.
- [6] J. Ellis, V. Sanz, and T. You, Prima Facie Evidence against Spin-Two Higgs Impostors, *Phys. Lett. B* **726**, 244 (2014).
- [7] J. Ellis, V. Sanz, and T. You, The effective Standard Model after LHC Run I, *J. High Energy Phys.* **03** (2015) 157.
- [8] J. Ellis, V. Sanz, and T. You, Complete Higgs sector constraints on dimension-6 operators, *J. High Energy Phys.* **07** (2014) 036.
- [9] A. Azatov, R. Contino, and J. Galloway, Model-independent bounds on a light Higgs, *J. High Energy Phys.* **04** (2012) 127.
- [10] A. Azatov and R. Contino, New prospects for Higgs compositeness in $h \rightarrow Z\gamma$, *Phys. Rev. D* **88**, 075019 (2013).
- [11] F. Bishara, R. Contino, and J. Rojo, Higgs pair production in vector-boson fusion at the LHC and beyond, *Eur. Phys. J. C* **77**, 481 (2017).
- [12] A. Biekötter, A. Knochel, R. Kogler, T. Plehn, and A. Voigt, The Global Higgs Picture at 27 TeV, *SciPost Phys.* **6**, 024 (2019).
- [13] I. Brivio, T. Corbett, and M. Trott, The Higgs width in the SMEFT, *J. High Energy Phys.* **10** (2019) 056.
- [14] D. de Florian *et al.* (LHC Higgs Cross Section Working Group), Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector, *CERN Yellow Rep. Monogr.* **2**, 1 (2017).
- [15] ATLAS Collaboration, Observation of $H \rightarrow b\bar{b}$ decays and VH production with the ATLAS detector, *Phys. Lett. B* **786**, 59 (2018).
- [16] CMS Collaboration, Measurements of Higgs boson production cross sections and couplings in the diphoton decay channel at $\sqrt{s} = 13$ TeV, *Phys. Rev. D* **104**, 092013 (2021).
- [17] C. Hays, V. Sanz Gonzalez, and G. Zemaityte, Constraining EFT parameters using simplified template cross sections, Technical Report LHCHXSWG-2019-004, CERN, Geneva, 2019.
- [18] J. Brehmer, S. Dawson, S. Homiller, F. Kling, and T. Plehn, Benchmarking simplified template cross sections in WH production, *J. High Energy Phys.* **11** (2019) 034.
- [19] J. Brehmer, K. Cranmer, G. Louppe, and J. Pavez, Constraining effective field theories with machine learning, *Phys. Rev. Lett.* **121**, 111801 (2018).
- [20] J. Brehmer, K. Cranmer, G. Louppe, and J. Pavez, A guide to constraining effective field theories with machine learning, *Phys. Rev. D* **98**, 052004 (2018).
- [21] K. Cranmer, J. Pavez, and G. Louppe, Approximating likelihood ratios with calibrated discriminative classifiers, [arXiv:1506.02169](https://arxiv.org/abs/1506.02169).
- [22] G. Kasieczka, B. Nachman, and D. Shih, The LHC Olympics 2020 a community challenge for anomaly detection in high energy physics, *Rep. Prog. Phys.* **84**, 124201 (2021).
- [23] F. F. Freitas, C. K. Khosa, and V. Sanz, Exploring the standard model EFT in VH production with machine learning, *Phys. Rev. D* **100**, 035040 (2019).
- [24] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations, *J. High Energy Phys.* **07** (2014) 079.
- [25] I. Brivio and M. Trott, Scheming in the SMEFT... and a reparameterization invariance!, *J. High Energy Phys.* **07** (2017) 148; **05** (2018) 136(A).
- [26] R. D. Ball *et al.* (NNPDF Collaboration), Parton distributions from high-precision collider data, *Eur. Phys. J. C* **77**, 663 (2017).
- [27] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, Asymptotic formulae for likelihood-based tests of new physics, *Eur. Phys. J. C* **71**, 1554 (2011).
- [28] M. G. Folgado, V. Sanz, and D. E. Conde Villatoro, Machine-learning-inspired SMEFT simplified template cross sections: A case study in ZH production, Zenodo, [10.5281/zenodo.19368109](https://zenodo.org/record/19368109) (2026).