Published for SISSA by Springer

Received: *January 11, 2024*
Accepted: *March 18, 2024*
Published: *April 11, 2024*

Non-resonant anomaly detection with background extrapolation

Kehang Bai,[a,b] **Radha Mastandrea**[b,c] **and Benjamin Nachman**[c,d]

[a]*Institute for Fundamental Science and Department of Physics, University of Oregon,
Eugene, OR 97403, U.S.A.*
[b]*Physics Division, Lawrence Berkeley National Laboratory,
Berkeley, CA 94720, U.S.A.*
[c]*Department of Physics, University of California,
Berkeley, CA 94720, U.S.A.*
[d]*Berkeley Institute for Data Science, University of California,
Berkeley, CA 94720, U.S.A.*

E-mail: kbai@uoregon.edu, rmastand@berkeley.edu, bpnachman@lbl.gov

Abstract: Complete anomaly detection strategies that are both signal sensitive and compatible with background estimation have largely focused on resonant signals. Non-resonant new physics scenarios are relatively under-explored and may arise from off-shell effects or final states with significant missing energy. In this paper, we extend a class of weakly supervised anomaly detection strategies developed for resonant physics to the non-resonant case. Machine learning models are trained to reweight, generate, or morph the background, extrapolated from a control region. A classifier is then trained in a signal region to distinguish the estimated background from the data. The new methods are demonstrated using a semi-visible jet signature as a benchmark signal model, and are shown to automatically identify the anomalous events without specifying the signal ahead of time.

Keywords: Models for Dark Matter, Specific BSM Phenomenology

ArXiv ePrint: 2311.12924

https://doi.org/10.1007/JHEP04(2024)059

Contents

1 Introduction

Despite the impressive predictability of the Standard Model (SM), it is a well-known fact that it does not account for all known phenomena and is plagued with a number of aesthetic issues. As such, the search for new, fundamental interactions is a central goal of particle physics across virtually every experiment. However, given that there is an uncountable number of new physics models, we simply do not have the bandwidth to test all possibilities, and it is certain that there are alternative hypotheses that we have not yet considered.

For this reason, a new paradigm has emerged to complement traditional, model-specific approaches. The new *anomaly detection* protocols seek to explore data with as little bias as possible in order to be broadly sensitive to many scenarios. Such protocols have been significantly advanced by modern machine learning (see [1–4] for a selection of machine learning reviews and anomaly detection challenges, and [5–102] for specific techniques). One powerful anomaly detection strategy ("weak supervision") is to isolate a region of phase space and compare data to a background-only reference. Signal model-agnostic evidence for new physics emerges when the data is not statistically consistent with the reference.

The core task of weakly supervised anomaly detection strategies is to construct the reference sample of background-only events. Existing strategies follow one of two approaches: (1) the reference sample is from (background-only) simulation and (2) the reference sample is learned from sidebands. The advantage of the simulation approach is that it puts few assumptions on the form of the signal. In particular, this approach can accommodate signals that are non-resonant, either due to the new physics being very massive or consisting of non-reconstructable (e.g. invisible) particles. The challenge with a simulation-based reference is that the simulation accuracy limits the anomaly detection sensitivity. In contrast, sideband methods posit that the signal, if it exists, is localized in at least one known dimension m. Focusing on a particular interval in m, a signal-sensitive region of phase space is then constructed using other features x, and regions in m away from the posited resonance are used to estimate the background $p_{\text{background}}(x|m)$. The benefit of sideband methods is that the reference is estimated directly from data, while the challenge is that not all signals are resonant.

We propose an approach to merge the advantages of the simulation-based and sideband-based cases to form a strategy that learns the reference from data without requiring a resonance. This method extends and modifies existing resonance approaches [19, 20, 34, 57, 85, 92, 103] to allow for extrapolation in m, as opposed to interpolation. Instead of the signal region being defined by an interval in a one-dimensional m, we now have a multidimensional m for which the signal region is defined by thresholds[1] $m_0 > c_0, m_1 > c_1, \ldots, m_n > c_n$ for $m, c \in \mathbb{R}^n$. The conditional reference $p_{\text{background}}(x|m)$ is learned (directly or indirectly) from the complement of the signal region and then extrapolated to the signal region. If the machine learning models are smooth and the background remains qualitatively the same in the signal region (e.g. the functional form of the background distribution, particularly the dependence on the context variables, is continuous across the CR-SR boundary), there is reason to believe that the extrapolation can be accurate.[2] A new feature of the non-resonant case is that we need to additionally estimate $p_{\text{background}}(m)$. This is also needed in the resonant case, but it is less important (since $p(m)$ is relatively constant in the signal region) and thus approximate or simplified methods have been considered so far. We propose to estimate $p_{\text{background}}(x|m)$ using a likelihood-ratio method [19] and then combine this estimate with three proposals for estimating $p_{\text{background}}(m)$ based on reweighting [19], generative models [57], and template morphing [103].

The growing anomaly detection for fundamental interactions literature includes a number of proposals that can accommodate non-resonant new physics. In the weakly supervised case, simulation-based background estimates do not require resonant new physics [6, 8, 67, 77]. Nearly all data-based weakly supervised methods include non-resonant x features [19, 20, 34, 57, 85, 92, 99, 103]. In special cases, symmetries can be used to define the reference directly in data without requiring m to be resonant [63, 64, 75, 79]. Extrapolation of generative models has been considered for background estimation, but combined with classical model-specific search strategies [105]. Unsupervised methods like those based on autoencoders (e.g. refs. [9, 10, 14] and many others) do not require resonances for achieving signal sensitivity in general. However, except in special cases, they often lack background estimation strategy [65], or any guarantees of optimality. Our approach is unique because it combines signal sensitivity with background estimation[3] and is asymptotically optimal [20] in the limit of large datasets and effective machine learning.

Non-resonant signals can arise from many new physics models. For example, models of a strongly interacting dark sector can produce jets which contain both stable and unstable dark

[1] A non-rectangular signal region is also possible, but makes the accounting more difficult.

[2] We leave additional studies to impose minimalism in the extrapolation to future work via e.g. monotonicity [104].

[3] The reference does not need to be used directly to estimate the background after making a cut on the anomaly score. One could use any classical non-resonant background estimation strategy like the ABCD method, possibly also with machine learning [106].

hadrons, predicting signatures of jets containing significant missing energy. Non-resonant signals are also characteristic of effective field theories, which can model new physics with masses beyond those accessible through on-shell production. ML-based anomaly detection can offer additional sensitivity to explore the vast signature space of non-resonant signals.

This paper is organized as follows. In section 2, we introduce the methodology we follow for our tail-based non-resonant anomaly detection procedure, providing details on the background extrapolation task. In section 3, we show a realistic physics example, considering jets emerging from a complex dark sector as our non-resonant signal. We conclude in section 4.

2 Methodology

2.1 Locating overdensities

The goal of a weakly-supervised anomaly search is to identify regions of phase space that are more represented by data (forming overdensities) than they are by the reference dataset. The search is performed in a signal region (SR) and the reference is estimated using information from a neighboring region.

In the resonant case, the SR is defined by an interval around a posited new particle mass $m \in [m_0 - \delta, m_0 + \delta]$. The neighboring region (called *sideband* or SB) is the region (or a subset of the region) outside of this interval. If the signal exists and is at mass m_0, we expect that most of the signal is in the SR with little contamination in the SB. The reference background template is estimated using information from the SB. Since we expect the phase space to be smoothly varying with m and since δ is expected to be small relative to regions over which large changes occur in $p_{\text{background}}(x|m)$, we expect the background to be well-estimated by interpolating from the SB (see further details in section 2.2). Given samples from the background dataset,[4] overdensities are identified with the Classification WithOut LAbels procedure (CWoLa) [5, 7, 107]. Events in data are all assigned the label "signal" and all events from the background template are labeled "background". An optimal classifier trained to distinguish these datasets will implicitly learn a function monotonically related to the signal-over-background likelihood ratio. Thresholding the classifier will then isolate an anomaly-like region of phase space.

In the non-resonant case, the setup is slightly different. Instead of having SB that surround the SR, the neighboring region is only on one side and is now called a *control region* or CR. However, we allow for m to be multidimensional — the more dimensions available to constrain $p_{\text{background}}(x|m)$, the better accuracy we might expect for reconstruction in the SR.[5] Resonances are in fact nearly a special case of this setup, since one could define e.g. $m' = 1/|m - m_0|$. As in the resonant case, we start by constructing background templates in the CR, and we expect most of the signal to be in the SR so that the CR can be used to estimate $p_{\text{background}}$ without bias. A schematic diagram of the non-resonant setup is presented in figure 1. The challenge now is to extrapolate into the SR instead of interpolating as in the resonant case. Given the background template, the actual anomaly detection proceeds

[4] There are variations on this where the density is used directly without samples [20], but the overall idea is the same.

[5] We validated this with simplified Gaussian samples.

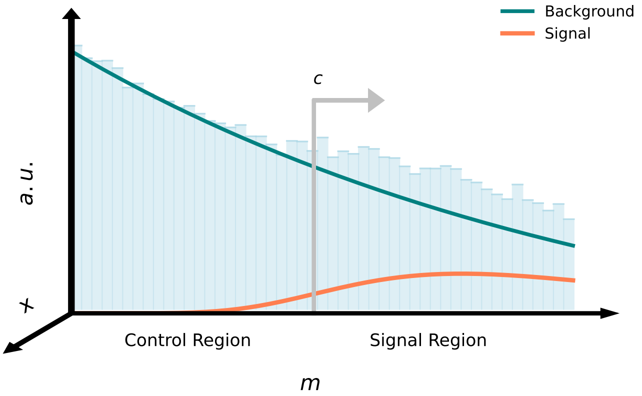


Figure 1. A schematic of the setup for non-resonant anomaly detection. The signal region (SR) is defined by a one-sided cut on a context variable $m > c$. While the figure shows a one-dimensional context variable, in practice the context can be multidimensional. A number of other features x are used for the CWoLa anomaly detection classifier. It is also possible to include m in the classifier.

using the same CWoLa procedure described above. The main aspect that distinguishes different approaches is how the background templates are generated.

2.2 Constructing background templates

To construct samples of background-only events in the SR, we take the following steps:

1. **Estimate $p_{\text{data}}(x|m)$ in the CR.** We first learn the distribution of the chosen features for background-only events, conditioned on the context variables, in the CR. It is assumed that the conditioning on the context will allow the learned distribution to be extrapolated into the SR. For the interpolation case (i.e resonant anomaly detection), the analogous first step would be to learn $p_{\text{data}}(x|m)$ in the SB. Since we assume that the signal is mostly absent from the CR, $p_{\text{data}} \approx p_{\text{background}}$.

2. **Estimate $p_{\text{background}}(m)$ in the SR.** In order to sample from $p_{\text{background}}(x|m)$ in the SR, we need to be able to generate SR-like context m. In the interpolation case, this might be done by performing a parametric fit to a histogram of m in the SB and interpolating to the SR. This could also work in the extrapolation case, but (1) the shape may be non-trivial in the tails of m and (2) this approach does not scale well to more than one dimension.

   To circumvent these challenges, we estimate $p_{\text{background}}(m)$ in the SR by reweighting context in the CR, $p_{\text{simulation}}(m)$, to match data and then extrapolating this function into the SR. A binary classifier parameterized in m [108, 109] is trained to distinguish simulated context from data context in the CR. The classifier output is then interpreted as a likelihood ratio $w(m)$ for estimating the background [19]. As in step 1, we assume that conditioning the network on the context will enable the learned weights to extrapolate into the SR.

3. **Sample $p_{\text{background}}(x|m)$ in the SR.** We first estimate $p_{\text{background}}(x|m)$ from the learned model in the CR, drawing context from $p_{\text{simulation}}(m)$, in the CR. The resulting

events are then weighted by $w(m)$ to arrive at our estimate of $p_{\text{background}}(x, m)$ in the SR, following eq. (2.1).

$$\begin{aligned} p_{\text{background}}(x, m) &= p_{\text{background}}(x|m) p_{\text{background}}(m) \\ &= p_{\text{background}}(x|m) p_{\text{simulation}}(m) \frac{p_{\text{background}}(m)}{p_{\text{simulation}}(m)} \\ &= p_{\text{background}}(x|m) p_{\text{simulation}}(m) w(m) \end{aligned} \tag{2.1}$$

We consider three approaches for approximating $p_{\text{data}}(x|m)$ in the CR, based on three techniques previously developed for the resonance case: Simulation-Assisted Likelihood-Free Anomaly Detection (SALAD) [19], Classifying Anomalies through Outer Density Estimation (CATHODE) [57], and Flow-Enhanced Transportation for Anomaly detection (FETA) [103]. Our proposed methods are:

1. **Reweight** (SALAD-inspired). The parameterized, binary classifier used to approximate $p_{\text{background}}(m)$ is extended to include also x so that the likelihood ratio in both x and m is estimated at the same time.

2. **Generate** (CATHODE-inspired). A generative neural network (normalizing flow [110]) is trained to model data in the CR, conditioned on the context m. Given values of m, one can sample features x from the normalizing flow.

3. **Morph** (FETA-inspired). A normalizing flow-based model is trained to map samples in the CR from MC to data, conditioned on the context m. Like SALAD, FETA starts from a simulation, but instead of reweighting, it morphs the features directly [111, 112].

Note that $w(m)$ is estimated with SALAD in all three cases, so our new approaches are SALAD-{SALAD, CATHODE, FETA} hybrid methods.

We evaluate the performance of the Reweight, Generate, and Morph-constructed background samples against two benchmarks: (1) a fully supervised classifier trained to discriminate pure background from pure signal, and (2) an "idealized" classifier trained to discriminate pure background from true background + signal. The idealized classifier represents the actual best performance to which our background estimation methods should asymptote.

2.3 Network architectures and training specifications

All networks are implemented in PyTorch [113] and optimized with Adam [114]. For the Generate and Morph methods, we construct normalizing flow networks with the nflows package [115]. All models are trained with a train-validation split of 2/3–1/3 and are evaluated at the epoch of lowest validation loss. For the Reweighting, Generate, and Morph methods, architectures and hyperparameters were optimized through manual tuning such that in the case of zero signal injection, a binary classifier could not distinguish reconstructed background in the CR from data in the CR. This was necessary in order to keep the SR blinded.

The Reweighting method is implemented with a binary classifier, consisting of a dense neural network with 3 layers of 100 nodes each. We use a batch size of 512, a learning rate of 10^{-3}, and train for up to 50 epochs with an early stopping patience of 5. The classifier that generates the context weights $w(m)$ is exactly the same as that for the Reweighting method. The Generate method is implemented with 4 layers of (Masked Affine Autoregressive Transform block of 128 hidden features, Reverse Permutation), where the transform block is built from a Masked Autoencoder for Distribution Estimation (MADE) architecture. Training is with a batch size of 216, a learning rate of 10^{-3}, and is carried out for up to 20 epochs with an early stopping patience of 5 epochs. For the Morph method, the base density flow uses the same architecture as the Generate method. The transfer flow is similar, but layers have only 64 hidden features, the batch size is reduced to 128, and the learning rate to 10^{-4}.

For the anomaly detection classifier networks, as well as the networks used for the closure tests (see section 3.2.1), we use a fully-connected architecture with 3 layers of 64 nodes each. We use a batch size of 512, a learning rate of 10^{-3}, and train for up to 50 epochs with an early stopping patience of 5 epochs. Networks are only trained on the non-context features. In order to ensure training stability, for a given set of generated samples, we crop the weights $w(m)$ by dropping all weights that are greater than 3 standard deviations above the mean. This has the effect of removing a very small number (usually on the order of a few tens, or $\sim 0.1\%$) of samples whose weights severely bias the resulting $p_{\text{background}}(x, m)$ distributions.

For the dataset, we consider a two-dimensional context m and a five-dimensional feature space x (to be explained in detail in section 3.1). The values of all seven variables are minmax-scaled to the range $(-2.5, 2.5)$ based on the minimum and maximum of the simulation (pure background) dataset before any network training is carried out. This was found to give more reliable flow training. When the number of signal events is small, the exact events that are injected into the "data" is important. Therefore, we rerun the procedures with 10 random signal injections and report the median and a 68-percentile spread. For each signal injection, we retrain the models with a different random initialization 20 times and average over these models. We found (as in previous studies) that this ensembling helps with sensitivity and stability.

3 Application with dark QCD jets

This section illustrates the use of the above background extrapolation method in the case of finding Beyond the Standard Model (BSM) physics at the LHC. The physics model we are interested in is jets that contain both stable and unstable hadrons from a strongly interacting dark sector, or dark QCD. This type of signal comes from e.g. Hidden Valley models [116–119]. In particular, we assume there exist some new massive quarks charged under dark QCD that are promptly produced and go through QCD-like showering and hadronization to form "dark hadrons". It is possible that a fraction of these dark hadrons decay back to the SM hadrons, and the rest remain invisible. This results in the type of dark QCD signature called semi-visible jets [120–125]. Semi-visible jets are challenging to distinguish from SM QCD jets, but the unique signature from certain dark hadron mass scales and missing energy may be used for identification.

We assume that the dark quarks are produced through a mediator such as a massive gauge boson Z'. When the mediator is produced on-shell, we can reconstruct the invariant mass of the mediator from the semi-visible jets. However, when the invisible fraction of the jet is large, there will be a large amount of missing energy, resulting in a poor invariant mass resolution so that the reconstructed signal is no longer resonant. The invisible fraction of

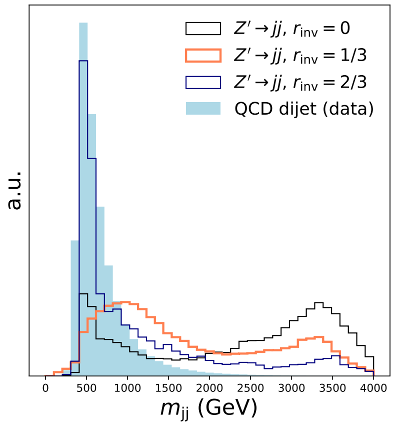


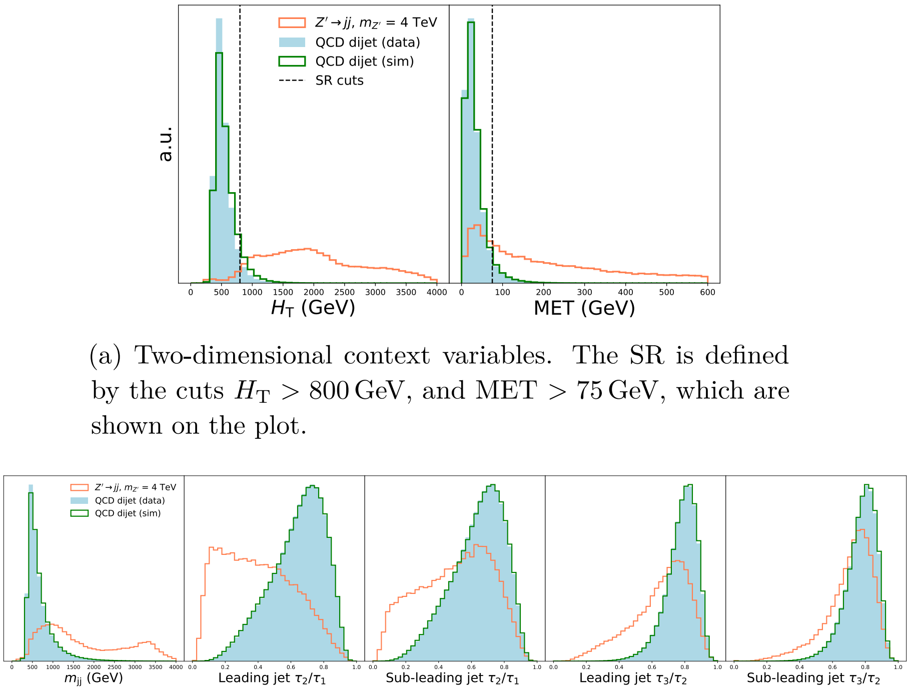
Figure 2. A histogram of m_{jj}, corresponding to the reconstructed mass of the 4 TeV Z' particle, for a selection of choices within the Hidden Valley model with parameter $r_{\rm inv}$. In this study, we set $r_{\rm inv} = 1/3$, at which point the Z' signal is no longer clearly resonant.

dark hadron decays is defined using the parameter, $r_{\rm inv}$, the ratio of the number of stable dark hadrons over number of total dark hadrons.

For this study, we choose a 4 TeV Z' decay into a pair of dark quarks. There are 3 flavours of dark quarks, all with the same mass of $m_{q_{\rm D}} = 250$ GeV. There are two types of dark hadrons, the dark pion and dark rho meson, whose masses are set to be the same as the confinement scale of the dark sector, $m_{\pi_{\rm D}} = m_{\rho_{\rm D}} = \Lambda_{\rm D} = 500$ GeV. Note that this choice of $\Lambda_{\rm D}$ results in a small number of dark mesons in the shower, and since the dark mesons are relatively heavy, the jet substructure comes from the dark meson decays and is less dependent on the hadronization process in the dark sector. A study of the effects from different mass parameters is presented in appendix A. The AD procedure is insensitive to the simulation modeling, and therefore is complementary to methods that minimizes the hadronization uncertainties [126, 127].

3.1 Dataset specifications

Signal events are generated using the Hidden Valley module from Pythia 8.310 [128]. Background ("data" and "simulation") events are generated using MadGraph5 aMC@NLO [129] and showered using Pythia 8.310. All events use Delphes [130] for detector simulation with the CMS card, and jets are clustered using the anti-k_t [131] algorithm with $R = 1.0$ using FastJet [132]. For the background samples, the data and simulation events differ in the choices of renormalization and factorization scales (1 vs. 2, respectively) as well as the tuning (14 vs. 25). We show results using a different background simulation in appendix B to validate the robustness of the anomaly detection (AD) methods against larger discrepancies in data and simulation. A minimum of 200 GeV transverse momentum is required for the leading jet in the event. In figure 2, we show the $m_{\rm jj}$ distribution for a few selections of $r_{\rm inv}$ values. To ensure the non-resonance of the signal, we set $r_{\rm inv} = 1/3$ in this study.



(a) Two-dimensional context variables. The SR is defined by the cuts $H_T > 800$ GeV, and MET > 75 GeV, which are shown on the plot.

(b) Five-dimensional feature variables.

Figure 3. Histograms of the observables used in the Hidden Valley non-resonant anomaly detection task, for background ("data" and "simulation") and signal events).

We use the scalar sum of jet transverse momenta (H_T) and missing transverse momentum (MET) as the context variables to define a two-dimensional SR.[6] For a five-dimensional feature space, we use the dijet invariant mass m_{jj}, and the two-pronged and three-pronged N-subjetiness [133, 134] for both the leading and subleading jets. Distributions of context and feature variables for signal and background is shown in figure 3. The SR is defined by the cuts $H_T > 800$ GeV, and MET > 75 GeV, which ensures the signal-over-background ratio is much lower in the CR than that of the SR.[7] We summarize the number of generated events in table 1.

3.2 Results

3.2.1 Closure tests

As a first test of extrapolation, we check that we are able to generate realistic background samples in both the CR and SR under zero-signal injection. In figure 4, we show the five feature distributions as constructed by the Ideal, Reweight, Generate, and Morph methods in relation to truth, the zero-signal data. Feature reconstruction is excellent in the CR,

[6] These were chosen since they are nearly independent. This is not strictly necessary for the method to work, but we found that it improves the extrapolation quality and the same features can be used for a classical matrix method / ABCD background estimate.

[7] In addition to the selection of features, the definition of CR and SR will introduce some model dependence. We expect that these regions will work well for a wide range of physics model parameters, especially when the Z' mass is higher.

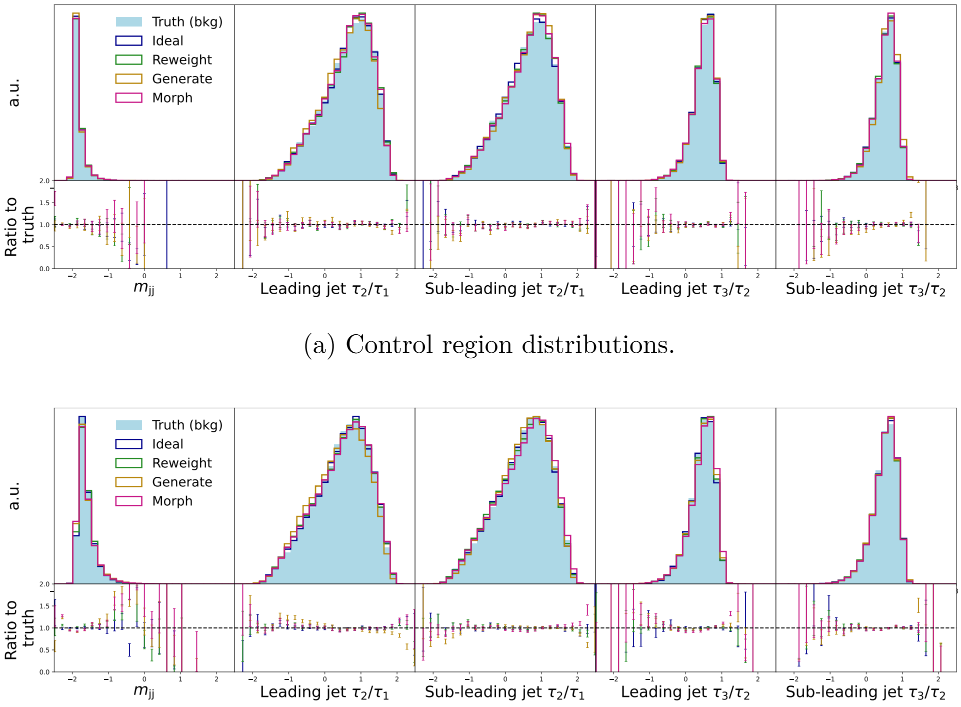
Purpose	Dataset	Number CR events	Number SR events
Training for generative models	Simulation (bkg.)	9,983k	126k
	Data (bkg.)	9,247k	72k
	Data (sig.)	14k	50k
Generated samples	Reweight	N/A	126k
	Generate		
	Morph		
Evaluation	Ideal AD samples	N/A	68k
	Fully supervised set	N/A	13k bkg., 30k sig.
	Test set	N/A	150k bkg., 20k sig.

Table 1. Breakdown of events generated for the Hidden Valley model. We found that large training sets in the CR were necessary for good performances of the Generate and Morph (the flow-based) methods, while the Reweight method worked well even with a smaller set. We did not explore oversampling in the SR, which may improve performance [57].

and the ratios of the marginals to the truth are consistent with unity across virtually all of the domains of the observables. Reconstruction is good overall in the SR, although the reconstructed distributions (particularly for the Morph method) deviate visibly. We calculated the one-dimensional Wasserstein distances as a quantitative measurement of the closure test, characterizing how well the observables are reconstructed in the signal region and control region by Reweight, Generate and Morph methods. The average Wasserstein distances of Ideal, Reweight, Generate, and Morph across five feature variables are 0.0048, 0.0073, 0.0237, 0.0116 in CR, and 0.0058, 0.0130, 0.0293, 0.0202 in SR, respectively. Note that in order to avoid premature unblinding, we withhold a set of 10k simulation and 10k data events in the CR from network training and use this set to evaluate CR closure.

In figure 5(a), we show a more rigorous test by training a binary classifier to discriminate between the constructed samples and background-only data. We show the spread of ROC AUCs (receiver operating characteristic area-under-curves) across 10 binary classifier runs; for a classifier trained on two identical datasets, the ROC AUC is expected to be 0.5 for a perfect classifier, in the limit of infinite statistics and training time. Again, we see that there is generally excellent reconstruction in the CR and slightly worse reconstruction in the SR. In figure 5(b), we show the classifier rejection (1/false positive rate) curves for the discrimination task in the SR. For all of the methods, the rejection found in the zero-signal case is compatible (to within errorbars) with that of a random classifier.

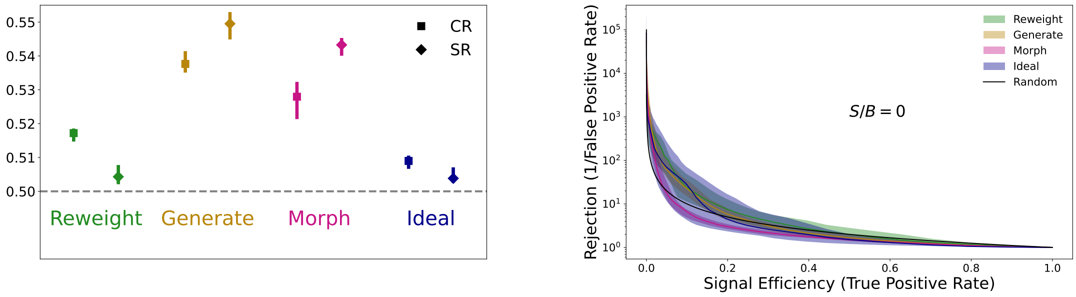
As stated in section 2.3, architectures and hyperparameters for the extrapolation networks were selected so as to make the spread of ROC AUCs in the CR consistent with random. There is a (seemingly unavoidable) systematic error associated with the extrapolation, such that Reweight, Morph, and Generate methods show a drop in similarity to true background in the SR. We believe that most of this drop can be attributed to the more general challenge of extrapolation for neural networks.



(a) Control region distributions.

(b) Signal region distributions.

Figure 4. Feature distributions as constructed by the Reweight, Generate, and Morph methods in the absence of signal. Errorbars in the ratio plots reflect Poisson statistics. Note that the values of the context and features have all been minmax-scaled to the range (-2.5, 2.5).



(a) ROC AUC spread in the CR and SR. The horizontal line is at 0.5, marking true random.

(b) Rejection in the SR.

Figure 5. Results for a classifier tasked with discriminating extrapolated background samples from true background. The “Ideal” corresponds to an idealized classifier, which has been trained to discriminate two statistically identical sets of background, and thus represents a realistic random classifier. Errorbars and uncertainty bands show a 68-percentile spread across 20 independent classifier runs.

3.2.2 Signal detection tests

In this section, we carry out an anomaly detection test for the dark QCD physics model. The test is done by generating a signal-background discriminator using a weekly-supervised binary classifier following the CWoLa procedure, as explained in section 2. The classifier is trained to discriminate the background reference against data that consists of SR background with a known amount of signal injection. For the simulation, we scan over seven signal injections of signal (S) over background (B) of $S/B \in \{0, 3.1, 6.5, 9.3, 12, 15, 18\} \times 10^{-3}$, corresponding to significances $S/\sqrt{B} \in \{0, 0.84, 1.75, 2.52, 3.32, 4.07, 4.99\}$, respectively.[8] In this section, we provide a number of metrics derived from this binary classifier.

We focus on metrics derived from the classifier SIC curves for the discrimination task in the SR. The SIC, defined as the true positive rate/$\sqrt{\text{false positive rate}}$, gauges the factor by which a signal's significance would improve by making a cut on the classifier score. We expect SIC $\gg 1$ for a well-performing classifier, while in the zero-signal case, the classifier would perform poorly.

In figure 6(a), we show the maximum SICs across all signal efficiencies, corresponding to the largest achievable significance improvements. In figure 6(b), we show SICs evaluated at a fixed rejection (given by 1/false positive rate) of 10^3. The results are encouraging, demonstrating comparable performance to the idealized anomaly detector for all three methods. The Reweight method enhances a signal significance from below $1(\sim 0.84)\sigma$ to the 5σ "discovery" threshold. The Generate and Morph methods perform less optimally with respect to the metric of SIC at a rejection of 10^3, but there is good evidence that the Generate and Morph methods could enhance a signal significance originally between 1–2σ to the discovery threshold. At zero signal injection, the higher SIC values for Reight is likely due to randomness, since there is nothing for the network to learn. What we see is that the mismodeling happens to be in a signal-like direction. It would be interesting for future studies to explore how these findings change for different signal and background models.[9]

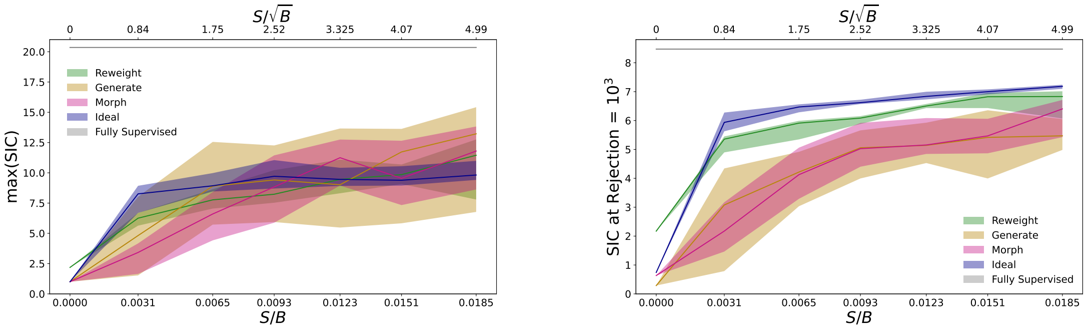
4 Conclusions

In this paper, we have extended background interpolation methods to the extrapolation case to enable weakly supervised, non-resonant anomaly detection (AD). Our strategy is based on the construction of a realistic set of background events through extrapolation from a background enriched control region (CR) into a signal region (SR). We proposed three extrapolation approaches based on reweighting, generation, and morphing which parallel the SALAD, CATHODE, and FETA methods, respectively. Studies in the resonant case have shown complementarity between these methods [94] which may be especially important in the extrapolation case.

The proposed methods are tested on a realistic BSM signal consisting of dark QCD jets. We found that in the zero-signal case, there was good closure between the reconstructed

[8] Note that with our chosen context region bounds, there is a small amount of signal contamination in the CR. For the largest signal injection, the contamination is $4.3\times10^{-3}\%$, from 384 events.

[9] See appendix A, where we show how the detection task changes when using a different set of signal parameters.



(a) Maximum of the Significance Improvement Characteristic.

(b) Significance Improvement Characteristic evaluated at a rejection of 10^3.

Figure 6. Metrics for an anomaly detection classifier tasked with discriminating extrapolated background samples from data in the SR. Uncertainty bands represent the 68-percentile spread over 10 different signal injections, all of which have been score-averaged over 20 independent classifier runs.

samples and the background data, showing that our background templates were well-modeled and unbiased. The agreement between predicted background and the true background in CR implies our methods are optimized for interpolation, and deviation in the SR might come from the intrinsic difficulty of extrapolation. Finally, we have also shown the performances of all the methods in a realistic AD task, demonstrating their abilities to elevate non-significant signal injection to the threshold of possible detection. While we only showed the performance for a single physics signal model, we stress that the protocol was optimized on only the background, and the sensitivity should extend to a broad class of new physics scenarios.

Our study, while encouraging, brings to light some of the many challenges associated with non-resonant AD. Future work is necessary in order to develop more robust non-resonant AD. Possible avenues may involve using networks or data preprocessing techniques that are especially well-suited for extrapolation. It is also worth testing the learning of $w(m)$ function given multiple different simulation sets that are qualitatively different from the data, in order to verify the effectiveness of the Reweight method.[10]

With background extrapolation methods, the range of possible signals that could be detected using an overdensity analysis would be greatly expanded. Not only are we able to explore a larger parameter phase space of known resonance signals that was previously not reconstructable, but we are also able to access non-resonant off-shell effects from heavy particles. It is likely that no one method will achieve the best sensitivity to all scenarios, so a set of techniques are required to achieve broad coverage.

Data and code availability. The physics datasets and parameter cards used for all simulations have been made available on Zenodo at https://zenodo.org/records/10154213. The analysis code is available at https://github.com/hep-lbdl/non-resonant-AD-extrapolation.

[10] See appendix B for an example of a modified simulation set.

Acknowledgments

We thank Timothy Cohen, Sascha Diefenbacher, Laura Jeanty, Simon Knapen, and Christiane Scherb for the useful discussions. In particular, we thank Christiane Scherb for the help on Hidden Valley simulations, and Laura Jeanty for providing detailed feedback on the manuscript. KB would like to thank Javier Montejo Berlingen for providing training and inspirations prior to this project. KB is supported by the U.S. Department of Energy, Office of Science, Office of High Energy Physics program under Award Number DE-SC0020244. KB's visit to LBNL for this collaboration is supported by the US ATLAS Center program. BN and RM are supported by the U.S. Department of Energy (DOE), Office of Science under contract DE-AC02-05CH11231 and Grant No. 63038 from the John Templeton Foundation. RM is additionally supported by Grant No. DGE 2146752 from the National Science Foundation Graduate Research Fellowship Program. This research used resources of the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231 using NERSC award HEP-ERCAP0021099.

A Comparing different signal parameters

In this section, we compare the performance of three proposed non-resonant AD methods (Reweight, Generate, and Morph) on different signal parameters of the dark QCD model. We choose the following two sets of parameters, both with $r_{\mathrm{inv}} = 1/3$ in addition to the 4 TeV Z' signal used in section 3:

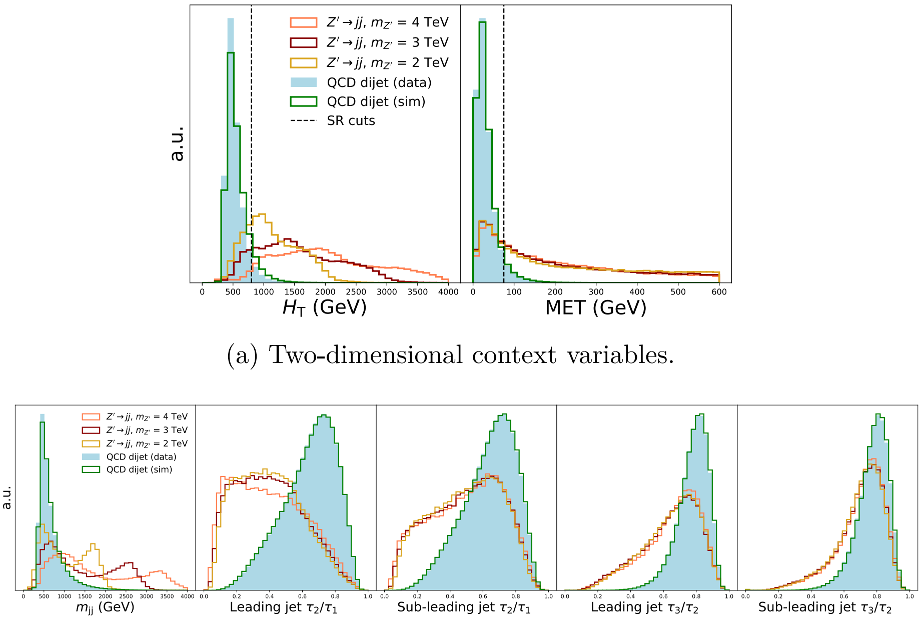
- $m_{Z'} = 2\,\mathrm{TeV}$, $m_{\pi_{\mathrm{D}}} = m_{\rho_{\mathrm{D}}} = \Lambda_{\mathrm{D}} = 200\,\mathrm{GeV}$, and $m_{q_{\mathrm{D}}} = 100\,\mathrm{GeV}$.
- $m_{Z'} = 3\,\mathrm{TeV}$, $m_{\pi_{\mathrm{D}}} = m_{\rho_{\mathrm{D}}} = \Lambda_{\mathrm{D}} = 300\,\mathrm{GeV}$, and $m_{q_{\mathrm{D}}} = 150\,\mathrm{GeV}$.

These two sets of parameters, denoted as the 2 TeV and 3 TeV signal, respectively, are chosen to have different m_{jj} and H_{T} distributions from the 4 TeV signal, with more signal located towards the bulk of the background. The choice of $m_{\pi_{\mathrm{D}}}$, $m_{\rho_{\mathrm{D}}}$, Λ_{D} and $m_{q_{\mathrm{D}}}$ values gives a similar two-pronged and three-pronged jet substructures as the 4 TeV signal in section 3.

Figure 7 shows the distributions of context and feature variables of three sets of signal parameters used in this study. We apply the same SR definition with $H_{\mathrm{T}} > 800\,\mathrm{GeV}$ and $\mathrm{MET} > 75\,\mathrm{GeV}$. Note that the distributions of the context variable H_T and the feature variables m_{jj} for the 2 TeV signal look much more similar to those of background, so we expect the discrimination task to be harder for lower signal masses.

In figures 8 and 9, we show the results for an anomaly detection classifier working on a range of signal detections for the 2 TeV and 3 TeV signals, respectively. (These plots are analogous to figure 6 for the 4 TeV signal.) As expected, the performances of all AD methods show low sensitivity to the 2 TeV signal, with Reweight achieving a similar significance improvement as the idealized classifier. Performance is better across the board for the 3 TeV signal, although the qualitative features are the same as for the 2 TeV signal.

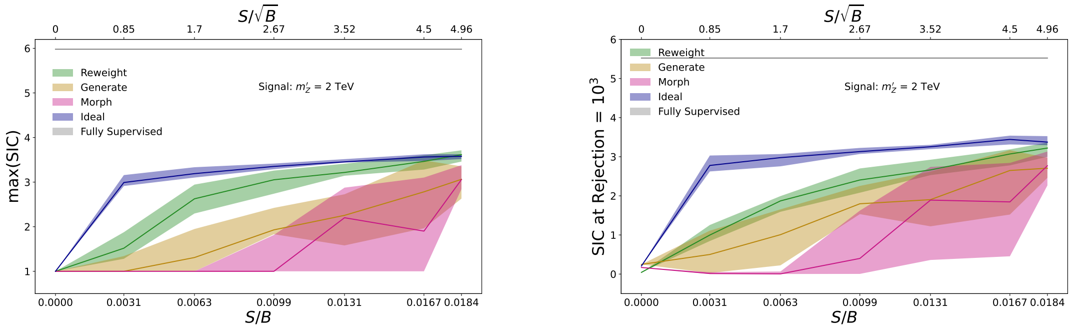
For both 2 TeV and 3 TeV signals, the gap in performance between the idealized classifier and the fully supervised classifier is much larger than it is for the 4 TeV signal. This difference might be ascribed to the fact that the lower-mass signals have a greater resemblance to the



(a) Two-dimensional context variables.

(b) Five-dimensional feature variables.

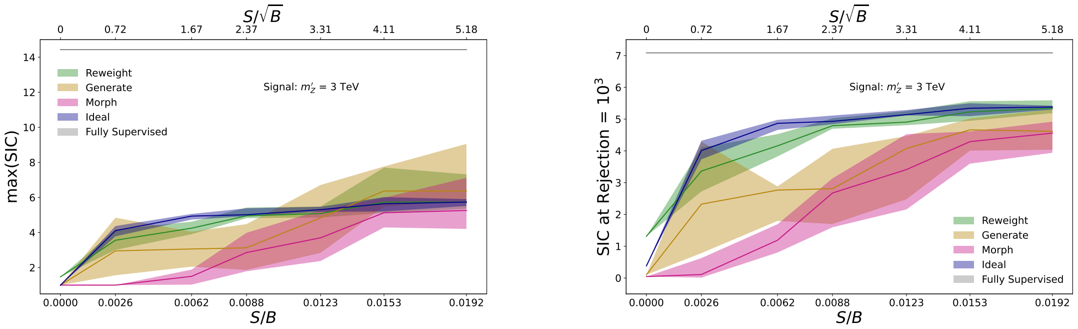
Figure 7. Histograms of the observables used in the Hidden Valley non-resonant anomaly detection task, for background (“data” and “simulation”) and signal events).



(a) Maximum of the Significance Improvement Characteristic.

(b) Significance Improvement Characteristic evaluated at a rejection of 10^3.

Figure 8. Metrics for an anomaly detection classifier tasked with discriminating extrapolated background samples from data in the SR. The background is QCD dijet events. The signal is a 2 TeV Z' producing two dark QCD jets with $2m_{q_D} = m_{\pi_D} = m_{\rho_D} = \Lambda_D = 200$ GeV. Uncertainty bands represent the 68-percentile spread over 10 different signal injections, all of which have been score-averaged over 20 independent classifier runs.



(a) Maximum of the Significance Improvement Characteristic.

(b) Significance Improvement Characteristic evaluated at a rejection of 10^3.

Figure 9. Metrics for an anomaly detection classifier tasked with discriminating extrapolated background samples from data in the SR. The background is QCD dijet events. The signal is a 3 TeV Z' producing two dark QCD jets with $2m_{q_D} = m_{\pi_D} = m_{\rho_D} = \Lambda_D = 300\,\mathrm{GeV}$. Uncertainty bands represent the 68-percentile spread over 10 different signal injections, all of which have been score-averaged over 20 independent classifier runs.

background process in variables such as H_T and m_{jj}. Additionally, the amount of signal contamination in the CR is higher for the lower mass Z' signals: for the 2 TeV Z', the highest signal injection has a CR contamination of 0.01% (1257 events), and for the 3 TeV Z', the highest contamination is 0.005% (551 events). A significant contamination could bias the extrapolated background distributions.

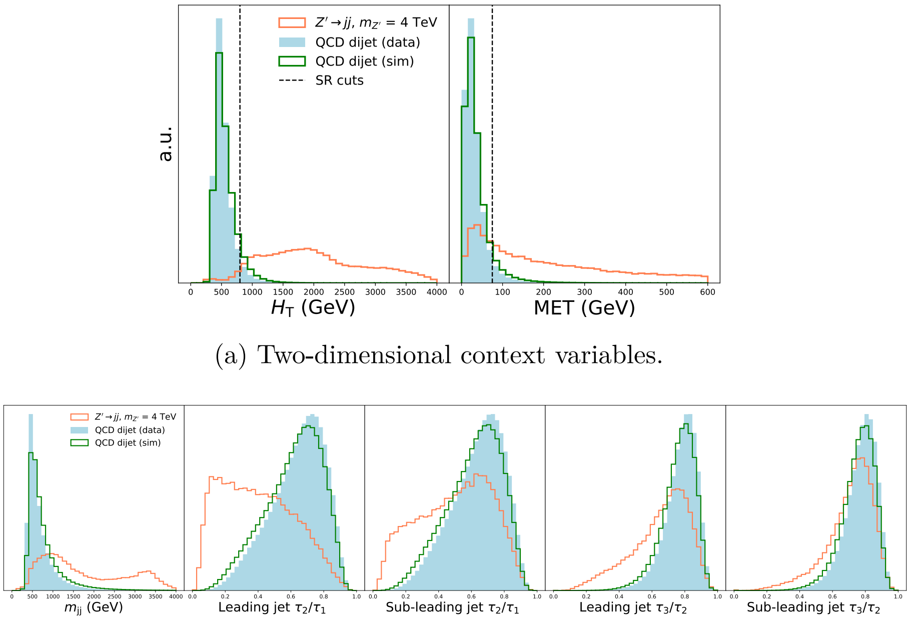
However, there is still sufficient evidence that the Reweight method could reliably enhance the signal significance from 1σ to the discovery threshold for all considered signal parameters. In contrast, the Generate and Morph methods appear to be more sensitive to the signal model, while still effective at picking up on moderate signal significances originating at $\sim 3\sigma$. It is worth exploring in which scenario the Reweight method becomes less effective than Generate and Morph in the future.

B Modifying the simulation

In the main text of this work, the distributions of simulated background are similar to the data distributions in most variables. This feature could decrease the difficulty of the reweighting procedure in all three methods. In this section, we investigate the effects of using a “morphed” simulation set that might better represent the expected discrepancy between simulation and actual collider data.

We introduce a hand-tuned modification to the simulation, both to the features and context, defined in eq. (B.1):

$$\begin{aligned} H_T:&\quad x \to x \\ \mathrm{MET}:&\quad x \to x\left(1+\frac{x}{500}\right) \\ m_{jj}:&\quad x \to x\left(1+\frac{x}{6000}\right) \\ \tau_i:&\quad x \to x^{1.1}. \end{aligned} \tag{B.1}$$



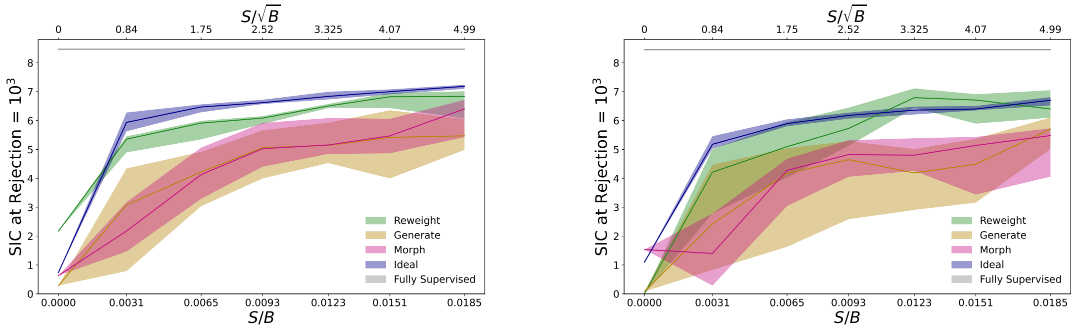
(a) Two-dimensional context variables.

(b) Five-dimensional feature variables.

Figure 10. Histograms of the observables used in the Hidden Valley non-resonant anomaly detection task, for background ("data" and "simulation") and signal events). The QCD dijet simulation has been modified to appear different from QCD dijet data.

Through this modification, the distributions of simulation are slightly morphed to be different from data, as shown in figure 10.

In figure 11, we compare the results for an anomaly detection classifier working on a range of signal detections for the modified simulation, using the same 4 TeV signal as was done in the main text. We make the comparison on the basis of the SIC at a rejection of 10^3; in figure 11(a), we show the results for the unmodified simulation, and in figure 11(b), we show the results from modified simulation. The performances are compatible, both quantitatively and qualitatively. In particular, the signal injections at which each method reaches the detection threshold are almost equal. The main difference is that the uncertainty bands are wider for the morphed simulation. This is logical given that when the simulation is more different from the "truth" background, the networks must learn a less trivial function to map from signal to background.



(a) Metric for unmodified simulation (identical to figure 6(b)).

(b) Metric for modified simulation.

Figure 11. SIC at a rejection of 10^3 for an anomaly detection classifier tasked with discriminating extrapolated background samples from data in the SR. Uncertainty bands represent the 68-percentile spread over 10 different signal injections, all of which have been score-averaged over 20 independent classifier runs.

References

[1] G. Kasieczka et al., *The LHC olympics* 2020 *a community challenge for anomaly detection in high energy physics*, *Rept. Prog. Phys.* **84** (2021) 124201 [arXiv:2101.08320] [INSPIRE].

[2] T. Aarrestad et al., *The dark machines anomaly score challenge: benchmark data and model independent event classification for the Large Hadron Collider*, *SciPost Phys.* **12** (2022) 043 [arXiv:2105.14027] [INSPIRE].

[3] G. Karagiorgi et al., *Machine learning in the search for new fundamental physics*, arXiv:2112.03769 [INSPIRE].

[4] M. Feickert and B. Nachman, *A living review of machine learning for particle physics*, arXiv:2102.02770 [INSPIRE].

[5] J.H. Collins, K. Howe and B. Nachman, *Anomaly detection for resonant new physics with machine learning*, *Phys. Rev. Lett.* **121** (2018) 241803 [arXiv:1805.02664] [INSPIRE].

[6] R.T. D'Agnolo and A. Wulzer, *Learning new physics from a machine*, *Phys. Rev. D* **99** (2019) 015014 [arXiv:1806.02350] [INSPIRE].

[7] J.H. Collins, K. Howe and B. Nachman, *Extending the search for new resonances with machine learning*, *Phys. Rev. D* **99** (2019) 014038 [arXiv:1902.02634] [INSPIRE].

[8] R.T. D'Agnolo et al., *Learning multivariate new physics*, *Eur. Phys. J. C* **81** (2021) 89 [arXiv:1912.12155] [INSPIRE].

[9] M. Farina, Y. Nakai and D. Shih, *Searching for new physics with deep autoencoders*, *Phys. Rev. D* **101** (2020) 075021 [arXiv:1808.08992] [INSPIRE].

[10] T. Heimel, G. Kasieczka, T. Plehn and J.M. Thompson, *QCD or what?*, *SciPost Phys.* **6** (2019) 030 [arXiv:1808.08979] [INSPIRE].

[11] T.S. Roy and A.H. Vijay, *A robust anomaly finder based on autoencoders*, arXiv:1903.02032 [INSPIRE].

[12] O. Cerri et al., *Variational autoencoders for new physics mining at the Large Hadron Collider*, *JHEP* **05** (2019) 036 [arXiv:1811.10276] [INSPIRE].

[13] A. Blance, M. Spannowsky and P. Waite, *Adversarially-trained autoencoders for robust unsupervised new physics searches*, *JHEP* **10** (2019) 047 [arXiv:1905.10384] [INSPIRE].

[14] J. Hajer, Y.-Y. Li, T. Liu and H. Wang, *Novelty detection meets collider physics*, *Phys. Rev. D* **101** (2020) 076015 [arXiv:1807.10261] [INSPIRE].

[15] A. De Simone and T. Jacques, *Guiding new physics searches with unsupervised learning*, *Eur. Phys. J. C* **79** (2019) 289 [arXiv:1807.06038] [INSPIRE].

[16] A. Mullin et al., *Does SUSY have friends? A new approach for LHC event analysis*, *JHEP* **02** (2021) 160 [arXiv:1912.10625] [INSPIRE].

[17] A. Casa and G. Menardi, *Nonparametric semisupervised classification for signal detection in high energy physics*, arXiv:1809.02977 [INSPIRE].

[18] B.M. Dillon, D.A. Faroughy and J.F. Kamenik, *Uncovering latent jet substructure*, *Phys. Rev. D* **100** (2019) 056002 [arXiv:1904.04200] [INSPIRE].

[19] A. Andreassen, B. Nachman and D. Shih, *Simulation assisted likelihood-free anomaly detection*, *Phys. Rev. D* **101** (2020) 095004 [arXiv:2001.05001] [INSPIRE].

[20] B. Nachman and D. Shih, *Anomaly detection with density estimation*, *Phys. Rev. D* **101** (2020) 075042 [arXiv:2001.04990] [INSPIRE].

[21] J.A. Aguilar-Saavedra, J.H. Collins and R.K. Mishra, *A generic anti-QCD jet tagger*, *JHEP* **11** (2017) 163 [arXiv:1709.01087] [INSPIRE].

[22] M. Romão Crispim, N.F. Castro, R. Pedro and T. Vale, *Transferability of deep learning models in searches for new physics at colliders*, *Phys. Rev. D* **101** (2020) 035042 [arXiv:1912.04220] [INSPIRE].

[23] M. Crispim Romão et al., *Use of a generalized energy Mover's distance in the search for rare phenomena at colliders*, *Eur. Phys. J. C* **81** (2021) 192 [arXiv:2004.09360] [INSPIRE].

[24] O. Knapp et al., *Adversarially learned anomaly detection on CMS open data: re-discovering the top quark*, *Eur. Phys. J. Plus* **136** (2021) 236 [arXiv:2005.01598] [INSPIRE].

[25] ATLAS collaboration, *Dijet resonance search with weak supervision using $\sqrt{s} = 13$ TeV pp collisions in the ATLAS detector*, *Phys. Rev. Lett.* **125** (2020) 131801 [arXiv:2005.02983] [INSPIRE].

[26] B.M. Dillon, D.A. Faroughy, J.F. Kamenik and M. Szewc, *Learning the latent structure of collider events*, *JHEP* **10** (2020) 206 [arXiv:2005.12319] [INSPIRE].

[27] M. Crispim Romão, N.F. Castro and R. Pedro, *Finding new physics without learning about it: anomaly detection as a tool for searches at colliders*, *Eur. Phys. J. C* **81** (2021) 27 [*Erratum ibid.* **81** (2021) 1020] [arXiv:2006.05432] [INSPIRE].

[28] O. Amram and C.M. Suarez, *Tag N' Train: a technique to train improved classifiers on unlabeled data*, *JHEP* **01** (2021) 153 [arXiv:2002.12376] [INSPIRE].

[29] T. Cheng et al., *Variational autoencoders for anomalous jet tagging*, *Phys. Rev. D* **107** (2023) 016002 [arXiv:2007.01850] [INSPIRE].

[30] C.K. Khosa and V. Sanz, *Anomaly awareness*, *SciPost Phys.* **15** (2023) 053 [arXiv:2007.14462] [INSPIRE].

[31] P. Thaprasop, K. Zhou, J. Steinheimer and C. Herold, *Unsupervised outlier detection in heavy-ion collisions*, *Phys. Scripta* **96** (2021) 064003 [arXiv:2007.15830] [INSPIRE].

[32] S. Alexander et al., *Decoding dark matter substructure without supervision*, arXiv:2008.12731 [INSPIRE].

[33] J.A. Aguilar-Saavedra, F.R. Joaquim and J.F. Seabra, *Mass Unspecific Supervised Tagging (MUST) for boosted jets*, *JHEP* **03** (2021) 012 [*Erratum ibid.* **04** (2021) 133] [arXiv:2008.12792] [INSPIRE].

[34] K. Benkendorfer, L.L. Pottier and B. Nachman, *Simulation-assisted decorrelation for resonant anomaly detection*, *Phys. Rev. D* **104** (2021) 035003 [arXiv:2009.02205] [INSPIRE].

[35] A.A. Pol et al., *Anomaly detection with conditional variational autoencoders*, in the proceedings of the *Eighteenth international conference on machine learning and applications*, (2020) [arXiv:2010.05531] [INSPIRE].

[36] V. Mikuni and F. Canelli, *Unsupervised clustering for collider physics*, *Phys. Rev. D* **103** (2021) 092007 [arXiv:2010.07106] [INSPIRE].

[37] M. van Beekveld et al., *Combining outlier analysis algorithms to identify new physics at the LHC*, *JHEP* **09** (2021) 024 [arXiv:2010.07940] [INSPIRE].

[38] S.E. Park et al., *Quasi anomalous knowledge: searching for new physics with embedded knowledge*, *JHEP* **06** (2020) 030 [arXiv:2011.03550] [INSPIRE].

[39] D.A. Faroughy, *Uncovering hidden new physics patterns in collider events using Bayesian probabilistic models*, *PoS* **ICHEP2020** (2021) 238 [arXiv:2012.08579] [INSPIRE].

[40] G. Stein, U. Seljak and B. Dai, *Unsupervised in-distribution anomaly detection of new physics through conditional density estimation*, in the proceedings of the 34th *conference on neural information processing systems*, (2020) [arXiv:2012.11638] [INSPIRE].

[41] P. Chakravarti, M. Kuusela, J. Lei and L. Wasserman, *Model-independent detection of new physics signals using interpretable semi-supervised classifier tests*, arXiv:2102.07679 [INSPIRE].

[42] J. Batson, C.G. Haaf, Y. Kahn and D.A. Roberts, *Topological obstructions to autoencoding*, *JHEP* **04** (2021) 280 [arXiv:2102.08380] [INSPIRE].

[43] A. Blance and M. Spannowsky, *Unsupervised event classification with graphs on classical and photonic quantum computers*, *JHEP* **08** (2020) 170 [arXiv:2103.03897] [INSPIRE].

[44] B. Bortolato, A. Smolkovič, B.M. Dillon and J.F. Kamenik, *Bump hunting in latent space*, *Phys. Rev. D* **105** (2022) 115009 [arXiv:2103.06595] [INSPIRE].

[45] J.H. Collins, P. Martín-Ramiro, B. Nachman and D. Shih, *Comparing weak- and unsupervised methods for resonant anomaly detection*, *Eur. Phys. J. C* **81** (2021) 617 [arXiv:2104.02092] [INSPIRE].

[46] B.M. Dillon, T. Plehn, C. Sauer and P. Sorrenson, *Better latent spaces for better autoencoders*, *SciPost Phys.* **11** (2021) 061 [arXiv:2104.08291] [INSPIRE].

[47] T. Finke et al., *Autoencoders for unsupervised anomaly detection in high energy physics*, *JHEP* **06** (2021) 161 [arXiv:2104.09051] [INSPIRE].

[48] D. Shih, M.R. Buckley, L. Necib and J. Tamanas, *via machinae: searching for stellar streams using unsupervised machine learning*, *Mon. Not. Roy. Astron. Soc.* **509** (2021) 5992 [arXiv:2104.12789] [INSPIRE].

[49] O. Atkinson et al., *Anomaly detection with convolutional graph neural networks*, *JHEP* **08** (2021) 080 [arXiv:2105.07988] [INSPIRE].

[50] A. Kahn et al., *Anomalous jet identification via sequence modeling*, 2021 *JINST* **16** P08012 [arXiv:2105.09274] [INSPIRE].

[51] T. Dorigo et al., *RanBox: anomaly detection in the copula space*, *JHEP* **01** (2023) 008 [arXiv:2106.05747] [INSPIRE].

[52] S. Caron, L. Hendriks and R. Verheyen, *Rare and different: anomaly scores from a combination of likelihood and out-of-distribution models to detect new physics at the LHC*, *SciPost Phys.* **12** (2022) 077 [arXiv:2106.10164] [INSPIRE].

[53] E. Govorkova et al., *LHC physics dataset for unsupervised new physics detection at* 40 *MHz*, *Sci. Data* **9** (2022) 118 [arXiv:2107.02157] [INSPIRE].

[54] G. Kasieczka, B. Nachman and D. Shih, *New methods and datasets for group anomaly detection from fundamental physics*, in the proceedings of the *Conference on knowledge discovery and data mining*, (2021) [arXiv:2107.02821] [INSPIRE].

[55] S. Volkovich, F. De Vito Halevy and S. Bressler, *A data-directed paradigm for BSM searches: the bump-hunting example*, *Eur. Phys. J. C* **82** (2022) 265 [arXiv:2107.11573] [INSPIRE].

[56] E. Govorkova et al., *Autoencoders on field-programmable gate arrays for real-time, unsupervised new physics detection at* 40 *MHz at the Large Hadron Collider*, *Nature Mach. Intell.* **4** (2022) 154 [arXiv:2108.03986] [INSPIRE].

[57] A. Hallin et al., *Classifying anomalies through outer density estimation*, *Phys. Rev. D* **106** (2022) 055006 [arXiv:2109.00546] [INSPIRE].

[58] B. Ostdiek, *Deep set auto encoders for anomaly detection in particle physics*, *SciPost Phys.* **12** (2022) 045 [arXiv:2109.01695] [INSPIRE].

[59] K. Fraser et al., *Challenges for unsupervised anomaly detection in particle physics*, *JHEP* **03** (2022) 066 [arXiv:2110.06948] [INSPIRE].

[60] P. Jawahar et al., *Improving variational autoencoders for new physics detection at the LHC with normalizing flows*, *Front. Big Data* **5** (2022) 803685 [arXiv:2110.08508] [INSPIRE].

[61] J. Herrero-Garcia, R. Patrick and A. Scaffidi, *A semi-supervised approach to dark matter searches in direct detection data with machine learning*, *JCAP* **02** (2022) 039 [arXiv:2110.12248] [INSPIRE].

[62] J.A. Aguilar-Saavedra, *Anomaly detection from mass unspecific jet tagging*, *Eur. Phys. J. C* **82** (2022) 130 [arXiv:2111.02647] [INSPIRE].

[63] R. Tombs and C.G. Lester, *A method to challenge symmetries in data with self-supervised learning*, 2022 *JINST* **17** P08024 [arXiv:2111.05442] [INSPIRE].

[64] C.G. Lester and R. Tombs, *Using unsupervised learning to detect broken symmetries, with relevance to searches for parity violation in nature. (Previously: "Stressed GANs snag desserts")*, arXiv:2111.00616 [INSPIRE].

[65] V. Mikuni, B. Nachman and D. Shih, *Online-compatible unsupervised nonresonant anomaly detection*, *Phys. Rev. D* **105** (2022) 055006 [arXiv:2111.06417] [INSPIRE].

[66] S. Chekanov and W. Hopkins, *Event-based anomaly detection for searches for new physics*, *Universe* **8** (2022) 494 [arXiv:2111.12119] [INSPIRE].

[67] R.T. d'Agnolo et al., *Learning new physics from an imperfect machine*, *Eur. Phys. J. C* **82** (2022) 275 [arXiv:2111.13633] [INSPIRE].

[68] F. Canelli et al., *Autoencoders for semivisible jet detection*, *JHEP* **02** (2022) 074 [arXiv:2112.02864] [INSPIRE].

[69] V.S. Ngairangbam, M. Spannowsky and M. Takeuchi, *Anomaly detection in high-energy physics using a quantum autoencoder*, *Phys. Rev. D* **105** (2022) 095004 [arXiv:2112.04958] [INSPIRE].

[70] L. Bradshaw, S. Chang and B. Ostdiek, *Creating simple, interpretable anomaly detectors for new physics in jet substructure*, *Phys. Rev. D* **106** (2022) 035014 [arXiv:2203.01343] [INSPIRE].

[71] J.A. Aguilar-Saavedra, *Taming modeling uncertainties with mass unspecific supervised tagging*, *Eur. Phys. J. C* **82** (2022) 270 [arXiv:2201.11143] [INSPIRE].

[72] T. Buss et al., *What's anomalous in LHC jets?*, *SciPost Phys.* **15** (2023) 168 [arXiv:2202.00686] [INSPIRE].

[73] S. Alvi, C.W. Bauer and B. Nachman, *Quantum anomaly detection for collider physics*, *JHEP* **02** (2023) 220 [arXiv:2206.08391] [INSPIRE].

[74] B.M. Dillon, R. Mastandrea and B. Nachman, *Self-supervised anomaly detection for new physics*, *Phys. Rev. D* **106** (2022) 056005 [arXiv:2205.10380] [INSPIRE].

[75] M. Birman et al., *Data-directed search for new physics based on symmetries of the SM*, *Eur. Phys. J. C* **82** (2022) 508 [arXiv:2203.07529] [INSPIRE].

[76] J.A. Raine, S. Klein, D. Sengupta and T. Golling, *CURTAINs for your sliding window: constructing unobserved regions by transforming adjacent intervals*, *Front. Big Data* **6** (2023) 899345 [arXiv:2203.09470] [INSPIRE].

[77] M. Letizia et al., *Learning new physics efficiently with nonparametric methods*, *Eur. Phys. J. C* **82** (2022) 879 [arXiv:2204.02317] [INSPIRE].

[78] C. Fanelli, J. Giroux and Z. Papandreou, *"Flux+Mutability": a conditional generative approach to one-class classification and anomaly detection*, *Mach. Learn. Sci. Tech.* **3** (2022) 045012 [arXiv:2204.08609] [INSPIRE].

[79] T. Finke, M. Krämer, M. Lipp and A. Mück, *Boosting mono-jet searches with model-agnostic machine learning*, *JHEP* **08** (2022) 015 [arXiv:2204.11889] [INSPIRE].

[80] R. Verheyen, *Event generation and density estimation with surjective normalizing flows*, *SciPost Phys.* **13** (2022) 047 [arXiv:2205.01697] [INSPIRE].

[81] B.M. Dillon et al., *A normalized autoencoder for LHC triggers*, *SciPost Phys. Core* **6** (2023) 074 [arXiv:2206.14225] [INSPIRE].

[82] S. Caron, R.R. de Austri and Z. Zhang, *Mixture-of-Theories training: can we find new physics and anomalies better by mixing physical theories?*, *JHEP* **03** (2023) 004 [arXiv:2207.07631] [INSPIRE].

[83] S.E. Park, P. Harris and B. Ostdiek, *Neural embedding: learning the embedding of the manifold of physics data*, *JHEP* **07** (2023) 108 [arXiv:2208.05484] [INSPIRE].

[84] J.F. Kamenik and M. Szewc, *Null hypothesis test for anomaly detection*, *Phys. Lett. B* **840** (2023) 137836 [arXiv:2210.02226] [INSPIRE].

[85] A. Hallin et al., *Resonant anomaly detection without background sculpting*, *Phys. Rev. D* **107** (2023) 114012 [arXiv:2210.14924] [INSPIRE].

[86] G. Kasieczka et al., *Anomaly detection under coordinate transformations*, *Phys. Rev. D* **107** (2023) 015009 [arXiv:2209.06225] [INSPIRE].

[87] J.Y. Araz and M. Spannowsky, *Quantum-probabilistic Hamiltonian learning for generative modeling and anomaly detection*, *Phys. Rev. A* **108** (2023) 062422 [arXiv:2211.03803] [INSPIRE].

[88] R. Mastandrea and B. Nachman, *Efficiently moving instead of reweighting collider events with machine learning*, in the proceedings of the 36th *conference on neural information processing systems: workshop on machine learning and the physical sciences*, (2022) [arXiv:2212.06155] [INSPIRE].

[89] J. Schuhmacher et al., *Unravelling physics beyond the standard model with classical and quantum anomaly detection*, *Mach. Learn. Sci. Tech.* **4** (2023) 045031 [arXiv:2301.10787] [INSPIRE].

[90] S. Roche et al., *Nanosecond anomaly detection with decision trees for high energy physics and real-time application to exotic Higgs decays*, arXiv:2304.03836 [INSPIRE].

[91] T. Golling et al., *The mass-ive issue: anomaly detection in jet physics*, in the proceedings of the 34th *conference on neural information processing systems*, (2023) [arXiv:2303.14134] [INSPIRE].

[92] D. Sengupta, S. Klein, J.A. Raine and T. Golling, *CURTAINs flows for flows: constructing unobserved regions with maximum likelihood estimation*, arXiv:2305.04646 [INSPIRE].

[93] V. Mikuni and B. Nachman, *High-dimensional and permutation invariant anomaly detection*, *SciPost Phys.* **16** (2024) 062 [arXiv:2306.03933] [INSPIRE].

[94] T. Golling et al., *The interplay of machine learning-based resonant anomaly detection methods*, *Eur. Phys. J. C* **84** (2024) 241 [arXiv:2307.11157] [INSPIRE].

[95] L. Vaslin, V. Barra and J. Donini, *GAN-AE: an anomaly detection algorithm for new physics search in LHC data*, *Eur. Phys. J. C* **83** (2023) 1008 [arXiv:2305.15179] [INSPIRE].

[96] ATLAS collaboration, *Anomaly detection search for new resonances decaying into a Higgs boson and a generic new particle X in hadronic final states using $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector*, *Phys. Rev. D* **108** (2023) 052009 [arXiv:2306.03637] [INSPIRE].

[97] S.V. Chekanov and R. Zhang, *Enhancing the hunt for new phenomena in dijet final states using anomaly detection filters at the High-Luminosity Large Hadron Collider*, *Eur. Phys. J. Plus* **139** (2024) 237 [arXiv:2308.02671] [INSPIRE].

[98] CMS ECAL collaboration, *Autoencoder-based anomaly detection system for online data quality monitoring of the CMS electromagnetic calorimeter*, arXiv:2309.10157 [INSPIRE].

[99] G. Bickendorf et al., *Combining resonant and tail-based anomaly detection*, arXiv:2309.12918 [INSPIRE].

[100] T. Finke et al., *Tree-based algorithms for weakly supervised anomaly detection*, *Phys. Rev. D* **109** (2024) 034033 [arXiv:2309.13111] [INSPIRE].

[101] E. Buhmann et al., *Full phase space resonant anomaly detection*, *Phys. Rev. D* **109** (2024) 055015 [arXiv:2310.06897] [INSPIRE].

[102] M. Freytsis, M. Perelstein and Y.C. San, *Anomaly detection in the presence of irrelevant features*, *JHEP* **02** (2024) 220 [arXiv:2310.13057] [INSPIRE].

[103] T. Golling, S. Klein, R. Mastandrea and B. Nachman, *Flow-enhanced transportation for anomaly detection*, *Phys. Rev. D* **107** (2023) 096025 [arXiv:2212.11285] [INSPIRE].

[104] O. Kitouni, N. Nolte and M. Williams, *Robust and provably monotonic networks*, *Mach. Learn. Sci. Tech.* **4** (2023) 035020 [arXiv:2112.00038] [INSPIRE].

[105] J. Lin, W. Bhimji and B. Nachman, *Machine learning templates for QCD factorization in the search for physics beyond the standard model*, *JHEP* **05** (2019) 181 [arXiv:1903.02556] [INSPIRE].

[106] G. Kasieczka, B. Nachman, M.D. Schwartz and D. Shih, *Automating the ABCD method with machine learning*, *Phys. Rev. D* **103** (2021) 035021 [arXiv:2007.14400] [INSPIRE].

[107] E.M. Metodiev, B. Nachman and J. Thaler, *Classification without labels: learning from mixed samples in high energy physics*, *JHEP* **10** (2017) 174 [arXiv:1708.02949] [INSPIRE].

[108] K. Cranmer, J. Pavez and G. Louppe, *Approximating likelihood ratios with calibrated discriminative classifiers*, arXiv:1506.02169 [INSPIRE].

[109] P. Baldi et al., *Parameterized neural networks for high-energy physics*, *Eur. Phys. J. C* **76** (2016) 235 [arXiv:1601.07913] [INSPIRE].

[110] D.J. Rezende and S. Mohamed, *Variational inference with normalizing flows*, in *Proceedings of the* 32$^{\text{nd}}$ *International Conference on Machine Learning* (2015), p. 1530 arXiv:1505.05770 [INSPIRE].

[111] T. Golling et al., *Morphing one dataset into another with maximum likelihood estimation*, *Phys. Rev. D* **108** (2023) 096018 [arXiv:2309.06472] [INSPIRE].

[112] S. Bright-Thonney, P. Harris, P. McCormack and S. Rothman, *Chained quantile morphing with normalizing flows*, arXiv:2309.15912 [INSPIRE].

[113] A. Paszke et al., *PyTorch: an imperative style, high-performance deep learning library*, arXiv:1912.01703 [INSPIRE].

[114] D.P. Kingma and J. Ba, *Adam: a method for stochastic optimization*, arXiv:1412.6980 [INSPIRE].

[115] C. Durkan, A. Bekasov, I. Murray and G. Papamakarios, *nflows: normalizing flows in PyTorch*, Zenodo, November 2020.

[116] M.J. Strassler and K.M. Zurek, *Echoes of a hidden valley at hadron colliders*, *Phys. Lett. B* **651** (2007) 374 [hep-ph/0604261] [INSPIRE].

[117] L. Carloni and T. Sjostrand, *Visible effects of invisible hidden valley radiation*, *JHEP* **09** (2010) 105 [arXiv:1006.2911] [INSPIRE].

[118] L. Carloni, J. Rathsman and T. Sjostrand, *Discerning secluded sector gauge structures*, *JHEP* **04** (2011) 091 [arXiv:1102.3795] [INSPIRE].

[119] S. Knapen, J. Shelton and D. Xu, *Perturbative benchmark models for a dark shower search program*, *Phys. Rev. D* **103** (2021) 115013 [arXiv:2103.01238] [INSPIRE].

[120] T. Cohen, M. Lisanti and H.K. Lou, *Semivisible jets: dark matter undercover at the LHC*, *Phys. Rev. Lett.* **115** (2015) 171804 [arXiv:1503.00009] [INSPIRE].

[121] T. Cohen, M. Lisanti, H.K. Lou and S. Mishra-Sharma, *LHC searches for dark sector showers*, *JHEP* **11** (2017) 196 [arXiv:1707.05326] [INSPIRE].

[122] H. Beauchesne, E. Bertuzzo, G. Grilli Di Cortona and Z. Tabrizi, *Collider phenomenology of hidden valley mediators of spin* 0 *or* 1/2 *with semivisible jets*, *JHEP* **08** (2018) 030 [arXiv:1712.07160] [INSPIRE].

[123] E. Bernreuther et al., *Casting a graph net to catch dark showers*, *SciPost Phys.* **10** (2021) 046 [arXiv:2006.08639] [INSPIRE].

[124] CMS collaboration, *Search for resonant production of strongly coupled dark matter in proton-proton collisions at* 13 *TeV*, *JHEP* **06** (2022) 156 [arXiv:2112.11125] [INSPIRE].

[125] ATLAS collaboration, *Search for non-resonant production of semi-visible jets using run* 2 *data in ATLAS*, *Phys. Lett. B* **848** (2024) 138324 [arXiv:2305.18037] [INSPIRE].

[126] T. Cohen, J. Doss and M. Freytsis, *Jet substructure from dark sector showers*, *JHEP* **09** (2020) 118 [arXiv:2004.00631] [INSPIRE].

[127] T. Cohen, J. Roloff and C. Scherb, *Dark sector showers in the Lund jet plane*, *Phys. Rev. D* **108** (2023) L031501 [arXiv:2301.07732] [INSPIRE].

[128] C. Bierlich et al., *A comprehensive guide to the physics and usage of PYTHIA* 8.3, arXiv:2203.11601.

[129] J. Alwall et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, *JHEP* **07** (2014) 079 [arXiv:1405.0301] [INSPIRE].

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

[131] M. Cacciari, G.P. Salam and G. Soyez, *The anti-k_t jet clustering algorithm*, *JHEP* **04** (2008) 063 [arXiv:0802.1189] [INSPIRE].

[132] M. Cacciari, G.P. Salam and G. Soyez, *FastJet user manual*, *Eur. Phys. J. C* **72** (2012) 1896 [arXiv:1111.6097] [INSPIRE].

[133] J. Thaler and K. Van Tilburg, *Maximizing boosted top identification by minimizing N-subjettiness*, *JHEP* **02** (2012) 093 [arXiv:1108.2701] [INSPIRE].

[134] J. Thaler and K. Van Tilburg, *Identifying boosted objects with N-subjettiness*, *JHEP* **03** (2011) 015 [arXiv:1011.2268] [INSPIRE].