Searching for periodic signals in kinematic distributions using continuous wavelet transforms

Hugues Beauchesne and Yevgeny Kats

Department of Physics, Ben-Gurion University,
Beer-Sheva 8410501, Israel

E-mail: beauches@post.bgu.ac.il, katsye@bgu.ac.il

ABSTRACT: Many models of physics beyond the Standard Model include towers of particles whose masses follow an approximately periodic pattern with little spacing between them. These resonances might be too weak to detect individually, but could be discovered as a group by looking for periodic signals in kinematic distributions. The continuous wavelet transform, which indicates how much a given frequency is present in a signal at a given time, is an ideal tool for this. In this paper, we present a series of methods through which continuous wavelet transforms can be used to discover periodic signals in kinematic distributions. Some of these methods are based on a simple test statistic, while others make use of machine learning techniques. Some of the methods are meant to be used with a particular model in mind, while others are model-independent. We find that continuous wavelet transforms can give bounds comparable to current searches and, in some cases, be sensitive to signals that would go undetected by standard experimental strategies.

Contents

1 Introduction

Experimental searches for physics beyond the Standard Model often look for peaks or dips in kinematic distributions. Although this is certainly well motivated, there exist potential signals that can take far more complicated forms and which have received very little attention.

One such possibility is periodic signals. These are a typical signature of models that include a large number of similar resonances with small mass splitting. Such models include the linear dilaton scenario [1–4], discrete and continuum clockwork models [3, 4], certain limits [5–7] of the Randall-Sundrum model [8] and more exotic warped extra dimensions [9]. An example of the two-photon invariant mass distribution is shown in figure 1 for the clockwork/linear dilaton (CW/LD) scenario [4].

Intuitively, one may think that taking the Fourier transform of a distribution like this would be an ideal strategy to exploit its periodic nature. There are however some complications with this in practice, as in most scenarios signals do not repeat themselves perfectly and indefinitely. For example, the repetitions might only occur over a finite interval. The position of that interval, which could potentially be used to discriminate signal from background, is essentially lost when passing to frequency space. Also, the frequency of the signal may not be constant. The Fourier transform would then not convey clearly which frequency is present at which point. All in all, certain characteristics of
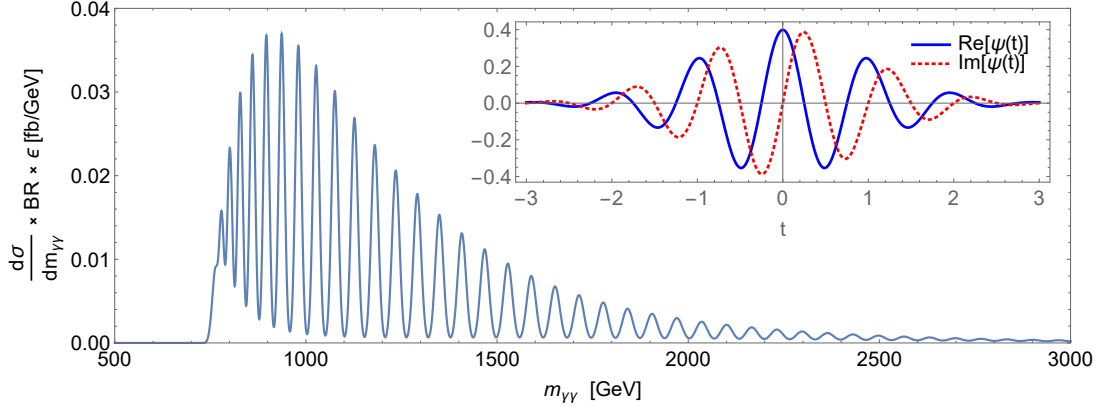
Figure 1: Example of a periodic signal in the diphoton spectrum from the clockwork/linear dilaton scenario. The parameters are set to $k = 750$ GeV and $M_5 = 3$ TeV. See appendix A for the technical details of the model and the modeling of the experimental resolution. The inset shows the Morlet wavelet (see eq. (2.4)).

realistic periodic signals are easier to see in the time domain, while others are more clear in the frequency domain. As such, neither the signal itself nor its Fourier transform is ideal to discover a signal whose periodicity changes with time.

Continuous Wavelet Transforms (CWT) address these issues by projecting a given signal over a basis of functions that are localized in both time and frequency space. An example of such a basis function is shown in the inset of figure 1. The output of the CWT is a scalogram, a two dimensional function which indicates how much a certain frequency is present at a given time. The CWT of the signal of figure 1, as well as a similar but weaker signal, is shown in figure 2 (with the input in figure 3). A signal that repeats itself with constant frequency would appear as a horizontal line, while one whose frequency changes with time as a line that moves up and/or down. A varying amplitude of the oscillations is also directly represented in the scalogram, as well as when the signal starts and ends. This makes the CWT very flexible when it comes to discovering generic periodic signals.

While signals of the kind shown in figure 1 can often be detected also directly by searches for resonances (e.g., [10–14]) or searches for continuum excesses at high masses (e.g., [10, 11, 15]), the coverage of such searches is not robust. The sensitivity of searches for single resonant peaks on top of a smooth background is likely reduced by the presence of neighboring partly-overlapping peaks, in a model- and search-procedure-dependent fashion. Furthermore, the expected presence of multiple peaks is not being exploited in an optimal manner by such searches. Searches for continuum excesses will have reduced sensitivity for scenarios in which the signal does not extend to high masses or it makes comparable positive and negative contributions to the spectrum due to quantum interference (see, e.g., [16–18]). In some cases, interference can even lead to pure deficits in the spectrum [18], challenging both the resonant and the continuum searches. It is also useful to note that the CWT analysis is almost insensitive to the systematic uncertainty on the normalization of the
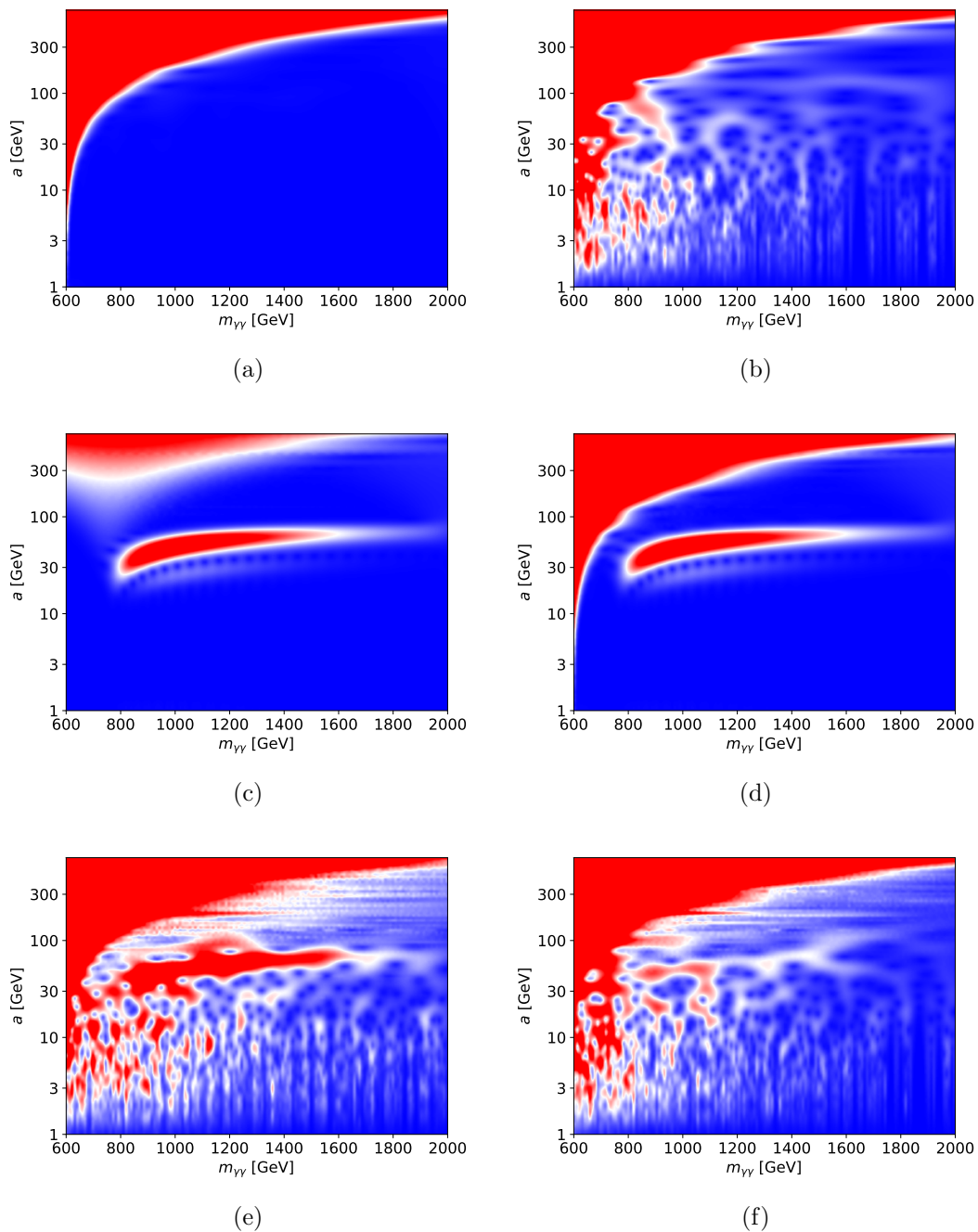
Figure 2: Examples of scalograms for the CW/LD with $k = 750$ GeV. (a) CWT of the smooth falling background from ref. [10] without statistical fluctuations. (b) CWT of the background with statistical fluctuations. (c) CWT of the signal alone, without statistical fluctuations, for $M_5 = 3$ TeV. (d) CWT of the same signal + background for $M_5 = 3$ TeV. (e) CWT of the signal + background with statistical fluctuations for $M_5 = 3$ TeV. (f) CWT of the signal + background with statistical fluctuations for $M_5 = 5$ TeV.
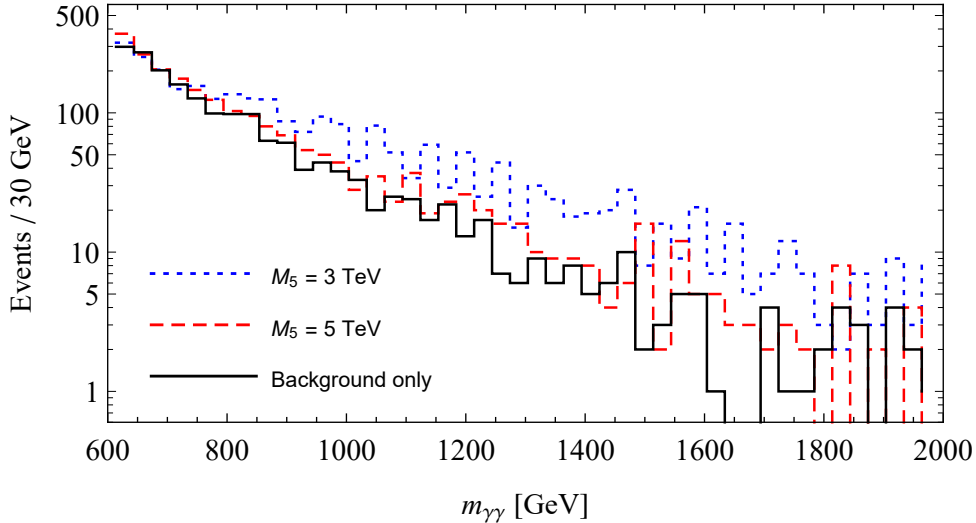
Figure 3: Mass spectra leading to the scalograms of figure 2. The black (solid), blue (dotted) and red (dashed) lines correspond to (b), (e) and (f), respectively.

background mass spectrum. The CWT approach is thus complementary to the existing search strategies and is therefore worth exploring.

Although the amount of work on CWT in the context of collider searches has been limited up to now (see ref. [19]), they have been used in many different fields of science. For example, they have been used in astronomy [20], biology [21], chaos theory [22], geophysics [23, 24], mechanical engineering [25] and signal processing [26]. Work on the statistical significance of signals in CWT includes refs. [27–35].

The goal of the current paper is to demonstrate that continuous wavelet transforms can be used to discover periodic signals in kinematic distributions, in particular in the context of new physics searches at colliders. To do this, we will present a series of methods through which CWT can be used to discover such signals. These will range from the use of a simple test statistic to more advanced machine learning techniques. We find that CWT can compete with current methods and in some cases be sensitive to signals that would otherwise go undetected by current analyses.

The paper is organized as follows. We begin by defining the continuous wavelet transforms more carefully. The set of methods are then introduced. Windowed Fourier transforms are then presented to serve as a comparison. The different methods are then compared in the context of the diphoton signal of the CW/LD benchmark, details about which are provided in the appendix. Some concluding remarks complete the paper.

2   Overview of continuous wavelet transforms

Assume $\psi(t)$ is a basis function localized in both time and frequency space. The continuous wavelet transform of a signal $f(t)$ at a scale $a > 0$ and translational parameter $b \in \mathbb{R}$ is

given by a projection over rescaled and shifted version of $\psi(t)$:

$$W(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} f(t)\, \psi^* \left(\frac{t-b}{a} \right) dt\,. \tag{2.1}$$

In practice, it is a measure of how much a certain frequency is present in the signal at a given time. The function $\psi(t)$ is known as the mother wavelet and its rescaled and shifted versions as daughter wavelets. The mother wavelet is required to satisfy two conditions:

$$\int_{-\infty}^{+\infty} |\psi(t)|^2 dt < \infty\,, \tag{2.2}$$

$$c_\psi \equiv 2\pi \int_{-\infty}^{+\infty} \frac{|\Psi(\omega)|^2}{|\omega|} d\omega < \infty\,, \tag{2.3}$$

where $\Psi(\omega)$ is the Fourier transform of $\psi(t)$. The first condition is simply that the mother wavelet has a finite norm and the second is known as the admissibility condition. The latter implies $\Psi(0) = 0$, which means in turn that an admissible wavelet must integrate to zero and as such that the CWT of a constant function is zero. Note that for practical uses the signal might be binned, in which case the integral is replaced by a sum over bins.

We will use the *Morlet wavelet* throughout this article. It consists of a localized wave packet and is given by:

$$\psi(t) \equiv \frac{1}{\sqrt{B\pi}}\, e^{-t^2/B} \left(e^{i2\pi Ct} - e^{-\pi^2 BC^2} \right), \tag{2.4}$$

where B and C are two constants that we will take as 2 and 1 respectively. With this choice, the wavelet transform of a signal will be maximum when its wavelength corresponds approximately to the scale a. The second term ensures that the admissibility condition is satisfied, though it can safely be ignored for our choice of parameters. The Morlet wavelet is shown in the inset of figure 1. Do note that there exist different conventions on the definition of the Morlet wavelet.[1]

In this article, we will be using the example of production of a set of resonances decaying to two photons. The invariant mass of the two photons $m_{\gamma\gamma}$ will play the role of t. When we discuss the distribution of $m_{\gamma\gamma}$ directly, we will say we are in mass space. When we deal with its wavelet transform, we will say we are in frequency space. The same ideas can be applied to dielectron, dimuon and other final states.

3 Search strategies with continuous wavelet transforms

We present in this section a series of methods through which continuous wavelet transforms can be used to discover periodic signals in kinematic distributions. These methods will be compared in section 4 by applying them to the CW/LD scenario [4] assuming the diphoton dataset of ref. [10] (37 fb^{-1} at 13 TeV). All illustrations are also taken from examples of that model.

---

[1]Our convention is chosen to conveniently match with that of the pywt Python package [36, 37].

(a)

(b)

(c)

(d)
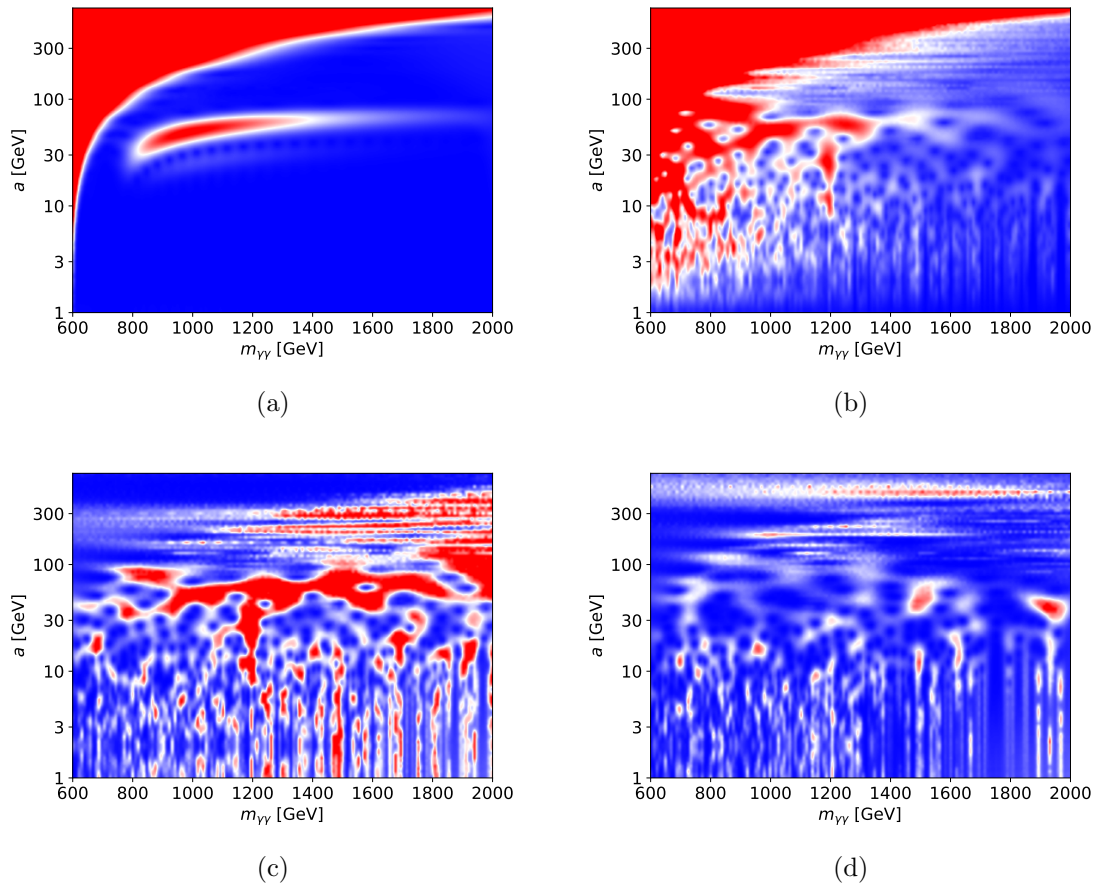
Figure 4: Example of (a) the CWT of the expected signal and background, (b) the same CWT with statistical fluctuations, (c) the corresponding $-\ln p$, and (d) $-\ln p$ of a typical background. The signal parameters used were $k = 750$ GeV and $M_5 = 4$ TeV.

3.1 Method 1: Model-specific search with a simple test statistic

The first strategy that we discuss is the use of a simple test statistic to detect a specific signal. A periodic signal is reflected in a scalogram as a series of ridges, each corresponding to a different harmonic. A given harmonic corresponds of course to a single scale for a given mass. Typically, the first harmonic is dominant and will have a large significance before the other harmonics are even visible. As such, we will concentrate on only the first one. Each point of a scalogram of the measured data can be assigned a local p-value by generating a large set of toy experiments with background only and determining which fraction of these have a larger norm of the wavelet coefficient at that point. Statistical fluctuations are simulated by finely binning the expected spectrum in mass space, fluctuating the number of events in each bin according to the Poisson distribution, and then applying the CWT. Of course, the bin size must be chosen to be smaller than the expected scale of the signal. A signal will appear as a valley of low p-value, i.e. an extended structure. This can be seen in figure 4. With these considerations in mind, a natural choice for a test statistic is:

$$t = -\sum_{i=i_{\min}}^{i_{\max}} \frac{1}{a_i} \ln p_i(a_i) \,, \qquad (3.1)$$

where the sum is over the mass bins of the scalogram, a_i is the scale of the first harmonic for bin i and $p_i(a_i)$ the local p-value for bin i at scale a_i. Roughly speaking, $i_{\min}$ is where the first harmonic starts to be discernible and $i_{\max}$ where it stops to be. In practice, it is best to perform an optimization of these parameters on a case by case basis. The division by the scale is performed to counteract the fact that fluctuations typically span a mass range of the order of their scale. This test statistic is also inspired by the Fisher method [38]. More advanced test statistics could in principle be used, but eq. (3.1) is easy to implement and will be shown to return good limits.

3.2 Method 2: Model-independent search with a simple test statistic

Arguably the greatest strength of continuous wavelet transforms is that they can reveal a periodic signal that was not predicted by any previously considered model. If an anomalous region is present in a scalogram, a question that would need to be answered is how significant it is. Previous work on the subject includes [28, 34].

First, bins of interest can be selected by asking that their local p-value be below a certain value. They are then grouped into continuous regions. The test statistic (3.1) is then applied to each region by taking the bin with the smallest p-value of each column. The largest test statistic is kept as a hyperstatistic. The statistical significance can then be obtained via a series of toy experiments.

3.3 Method 3: Neural network and a simple test statistic

Wavelet transforms map a periodic signal present over a background to an excess over an extended region in a scalogram. When the amount of statistics available is limited, such excesses can potentially be mistaken for simple statistical fluctuations of the background. As can be seen in figure 2, the scalogram of a given signal can take a very complicated form in practice. However, it is clear that a real signal will tend to present certain features that are typically absent from background fluctuations and vice-versa. These features can potentially be used to increase the statistical significance of a given signal. Trying to manually classify them is at the very least an extremely daunting task, but it is an obvious application of machine learning.

One possibility is to train a neural network to identify regions compatible with the signal searched for inside a scalogram and then calculate their significance using a test statistic. Note that after the mass spectrum is sampled with sufficient resolution to capture the details of the signal, the resulting scalogram can be sampled in a cruder fashion thanks to the extended nature of the excess. This helps making the size of the neural network's input manageable. The neural network is trained on an equal mix of pure backgrounds and examples with signals from random points in the parameter space of the model (with signal strength above some optimized threshold). In both cases, the network's input is the norm of the CWT of the signal + background divided by the expectation value of the norm of the CWT of the background only $\langle |W_b| \rangle$. When there is a signal, the output it is trained

Layer	Parameters
Input layer	63 mass bins $\times$ 56 scale bins
Convolutional layer 1	# filters = 32
	kernel size = (5, 5)
	Activation: Softplus
MaxPooling 1	Pooling size = (2, 2)
Convolutional layer 2	# filters = 64
	kernel size = (5, 5)
	Activation: Softplus
MaxPooling 2	Pooling size = (2, 2)
Convolutional layer 3	# filters = 128
	kernel size = (5, 5)
	Activation: Softplus
Dense 1	# of nodes = 5000
	Activation: Softplus
Output layer	# of nodes = 3528

(a)

| Setting	Choice
Optimizer	ADAM
Loss function	Mean squared error
# training experiments	5000
Validation split	0.2
Batch size	1000
# epochs	200
Callback	Smallest validation loss function

(b)

Table 1: (a) Structure of the convolutional neural network for the region finder. (b) Training parameters. All parameters not specified in these tables are left at their default KERAS values.

to return is $|W_{s,\exp}|/\langle|W_b|\rangle$, where $|W_{s,\exp}|$ is the norm of the wavelet transform of the expected signal. When there is no signal, the neural network is trained to return zero in every bin. The details of the neural network we used are given in table 1. When applied to a pseudo-experiment, the neural network will assign bins compatible with a signal-like excess a much larger value than those associated to the background. An example of this for a sample containing a signal is shown in figure 5, which shows that the neural network returns a very signal-like shape. For background-only samples, the neural network typically fails to return such well-defined regions. In principle, one could also train the neural network to return some other function, as long as its region of maximal value corresponds to where the excess should appear in the scalogram. All neural networks were implemented via the PYTHON deep learning library KERAS [39] with the TENSORFLOW backend [40] using the ADAM optimizer [41].

Having identified the region of interest, one can select for each mass column the scale with the largest neural network output and apply the test statistic of eq. (3.1) using the actual value of the wavelet coefficient for this bin. A minimal value of the neural network output is required for a bin to be counted in the test statistic. This way, the neural network's assessment of whether a signal-like excess is present at all, is taken into account. We scan the value of this threshold to maximize significance. The expected sensitivity limits are obtained by pseudo-experiments.
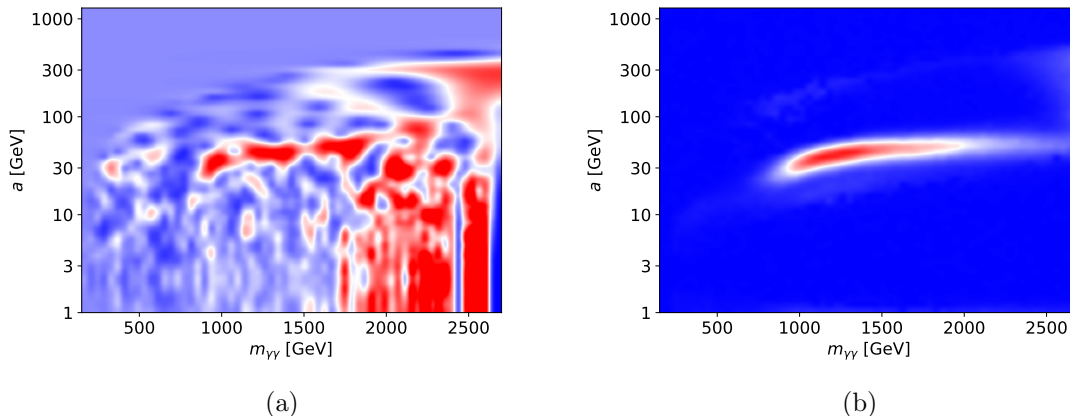
Figure 5: Example of the region finder algorithm input and output. (a) The input, which is the norm of the CWT divided by the expectation value of the background. (b) The output, which is trained to be $|W_{s,\exp}|/\langle|W_b|\rangle$. The parameters used were $k = 750$ GeV and $M_5 = 4$ TeV. The original signal is the same as in figure 4a.

3.4 Method 4: Classifier neural network

A more powerful way to use machine learning to discover a specific signal in a scalogram is with a classifier. First, toy experiments, with and without the signal (for a specific choice of model parameters), are generated and finely binned mass spectra are produced. The continuous wavelet transform of the mass spectrum is taken and rebinning is performed to make the size of the neural network's input manageable. The norm of each bin is then passed as an input to a convolutional neural network whose output is trained to be one when the signal is present and zero otherwise. The output of the neural network can then be used as a test statistic. The structure of the neural network that we used and additional parameters are provided in table 2.

3.5 Method 5: Autoencoder neural network

A more model-independent option to look for anomalies in scalograms is via autoencoders, similar to their application to jet images in refs. [42, 43]. The idea is to use an autoencoder network to compress a scalogram to a smaller set of parameters which are then used to reconstruct the original scalogram. The neural network is then trained on backgrounds only to reproduce the original scalogram as well as possible. After training, the neural network should be able to reproduce the original scalogram to good approximation if applied to a typical background sample, and fail if applied to a sample that contains a signal. One can then use the reconstruction loss function as a test statistic. The details of the neural network are presented in table 3 and are simply an adaption of the convolutional neural network of ref. [43]. The input of the neural network, which is also the output it is trained to return, is the negative log of the local p-value of each bin. Examples of inputs and outputs of the neural network are shown in figure 6. As can be seen, the autoencoder manages to reproduce approximately the major fluctuations of the background. At the

Layer	Parameters
Input layer	63 mass bins × 56 scale bins
Convolutional layer 1	# filters = 4
	kernel size = (3, 3)
	Activation: Elu
MaxPooling 1	Pooling size = (2, 2)
Convolutional layer 2	# filters = 8
	kernel size = (3, 3)
	Activation: Sigmoid
MaxPooling 2	Pooling size = (2, 2)
Convolutional layer 3	# filters = 16
	kernel size = (3, 3)
	Activation: Sigmoid
Dense 1	# of nodes = 200
	Activation: Sigmoid
Dense 2	# of nodes = 100
	Activation: Sigmoid
Output layer	# of nodes = 1

(a)

Setting	Choice
Optimizer	ADAM
Loss function	Binary cross entropy
# training experiments	4000
Validation split	0.2
Batch size	1000
# epochs	500
Callback	Smallest validation loss function

(b)

Table 2: (a) Structure of the convolutional neural network of the wavelet classifier. (b) Training parameters. All parameters not specified in these tables are left at their default KERAS values.

same time, as desired, it fails to reproduce the signal, which results in a much larger value for the reconstruction loss function.

3.6 Fourier analysis as a reference

Before moving on to comparing the methods, we discuss the use of Fourier transforms to discover periodic signals, which is the approach that was proposed in ref. [4] in the context of the CW/LD scenario. This will serve as a reference, though it is clear that CWT are far more general.

We define in general the power spectrum as:

$$P(T) = \left| \frac{1}{\sqrt{2\pi}} \int_{m_{\min}}^{m_{\max}} dm \frac{d\sigma}{dm} \frac{1}{L} \exp\left(i \frac{2\pi g(m)}{T} \right) \right|^2. \tag{3.2}$$

The quantity L is the parton luminosity given for CW/LD by $L(m^2) = \mathcal{L}_{gg}(m^2) + \frac{4}{3} \sum_q \mathcal{L}_{q\bar{q}}(m^2)$ (see eq. (A.4)). The mass spectrum is divided by this quantity to counteract the fast decrease that it causes in the signal and bring it closer to a regularly oscillating function [4]. For a general signal, the function $g(m)$ is defined such that the mass of the nth resonance is related to its index by $n = g(m_n)$. This is the quantity in terms of which the locations of the resonances are periodic and as such the quantity in terms of which

Layer	Parameters
Input layer	60 mass bins × 56 scale bins
Convolutional layer 1	# filters = 128
	kernel size = (3, 3)
	Activation: Elu
MaxPooling 1	Pooling size = (2, 2)
Convolutional layer 2	# filters = 128
	kernel size = (3, 3)
	Activation: Elu
MaxPooling 2	Pooling size = (2, 2)
Convolutional layer 3	# filters = 128
	kernel size = (3, 3)
	Activation: Elu
Dense 1	# of nodes = 40
	Activation: Elu
Dense 2 (*Encoded*)	# of nodes = 20
	Activation: Elu
Dense 3	# of nodes = 40
	Activation: Elu
Convolutional layer 4	# filters = 128
	kernel size = (3, 3)
	Activation: Elu
UpSampling 1	Upsampling factors = (2, 2)
Convolutional layer 5	# filters = 128
	kernel size = (3, 3)
	Activation: Elu
UpSampling 2	Upsampling factors = (2, 2)
Convolutional layer 6	# filters = 1
	kernel size = (3, 3)
	Activation: Elu
Output layer	60 mass bins × 56 scale bins

(a)

Setting	Choice
Optimizer	ADAM
Loss function	Mean squared error
# training experiments	5000
Validation split	0.2
Batch size	1000
Padding	Same
# epochs	100
Callback	Smallest validation loss function

(b)

Table 3: (a) Structure of the convolutional neural network for the autoencoder. (b) Training parameters. All parameters not specified in these tables are left at their default KERAS values.

the Fourier transform is best performed.[2] Simply doing a Fourier transform in terms of m would not lead to an optimal significance when $g(m)$ differs too much from m. This is because a signal that varies over a wide range of frequencies over its duration would lead to a very wide peak in Fourier space. For CW/LD, we take $g(m) = R\sqrt{m^2 - k^2}$ (see eq. (A.1)) [4]. Obviously, the power spectrum is expected to peak at $T = 1$. The value of the peak can then be used as a test statistic.

---

[2]As the masses m_n represent a discrete set and $g(m)$ must be a continuous function, simply requesting $g(m)$ to reproduce the correct masses does not fully define it and a smooth interpolation needs to be provided for the intermediate masses. In practice, there is usually an obvious definition for $g(m)$.
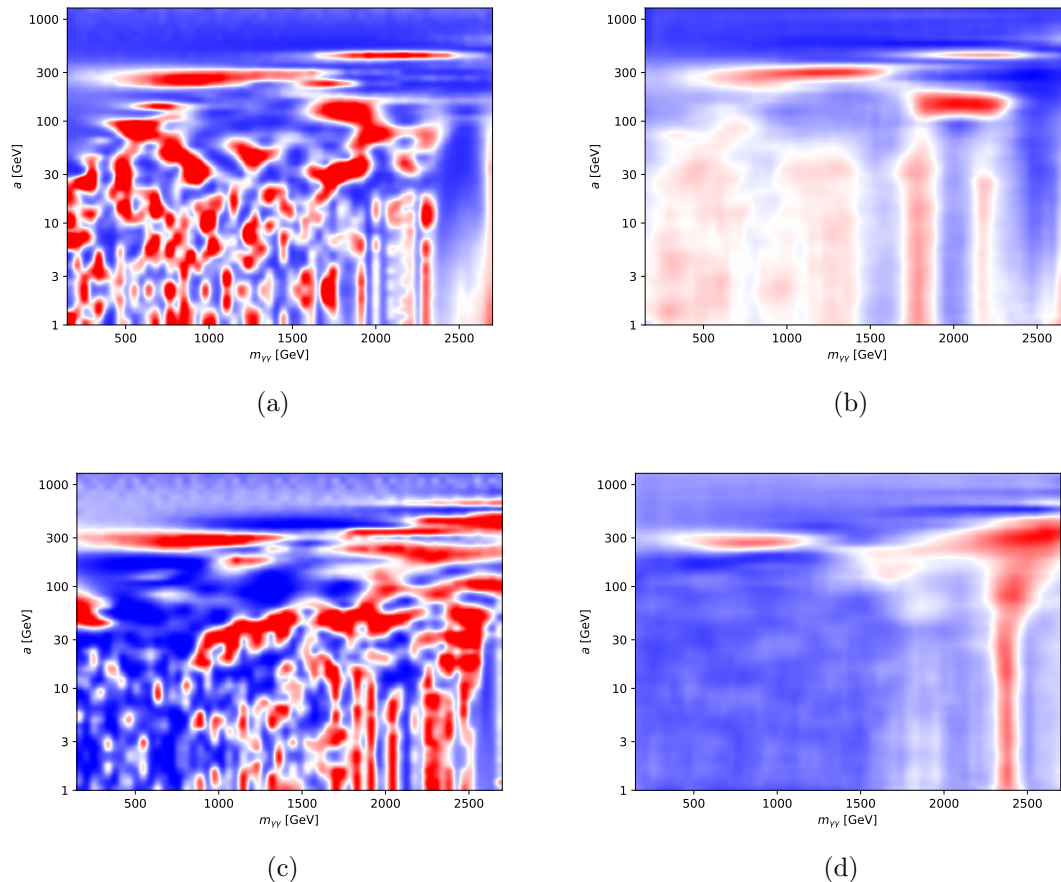
Figure 6: Examples of the input and output of the autoencoder. (a)→(b): Background-only sample and its reconstruction after encoding. (c)→(d): Sample containing a signal and its (mis)reconstruction after encoding. The parameters used for the signal were $M_5 = 4$ TeV and $k = 750$ GeV. The original signal is shown figure 4a.

4 Comparison between the different methods

To compare the different methods, we apply them to the CW/LD scenario with the $\gamma\gamma$ dataset of ref. [10]. The resulting reach in the parameter space of the model is shown in figure 7, where the contours correspond to a median expected significance of 2 sigma. As described in more detail in appendix A, the parameter k is approximately the mass at which the spectrum begins, while M_5 controls the cross section, which is approximately proportional to $1/M_5^3$. The structure of the resonance masses m_n, if described in terms of m_n/k, is essentially independent of k and M_5 in the range of parameters considered, and the asymptotic value of the mass splittings at high mass is given by $\Delta m \approx k/10$.

The blue (thin solid) curve corresponds to the test statistic method of section 3.1. The lower limit of the sum in eq. (3.1) was optimized to maximize the reach. A similar procedure was done for the upper limit, but we found that the results were virtually identical to those where the upper limit was taken as very large. In practice, we simply
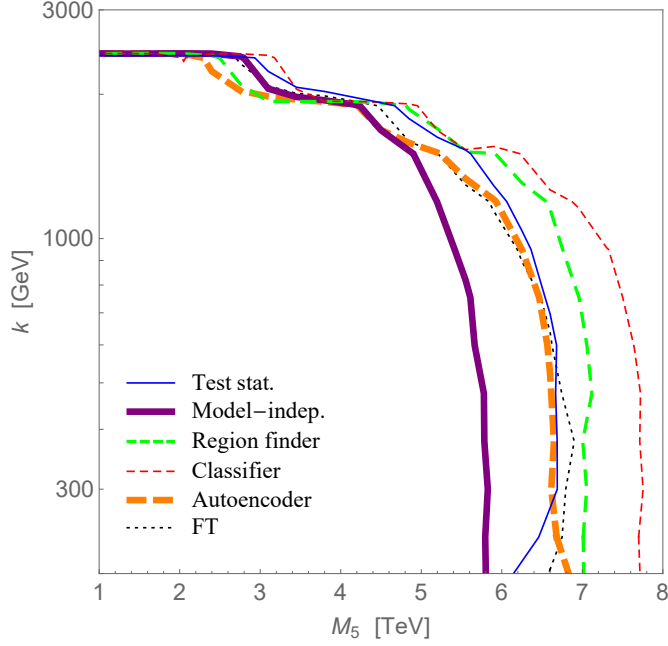
Figure 7: Comparison between the 2σ sensitivity of the various search methods, in the parameter space of the CW/LD model, based on the $\gamma\gamma$ dataset of ref. [10] (see appendix A for details). Thin lines correspond to methods designed for a specific model and specific parameters, medium lines to methods designed for a specific model but not a specific point and thick lines to searches for a general signal. Solid lines do not use machine learning and dashed ones do. The regions covered by the various methods are to the left of the corresponding curves.

took $\min(10k, 2.7\text{ TeV})$, where the latter is the upper limit of the experimental data. In terms of the reach in M_5, the sensitivity peaks at around $k \approx 600$ GeV. As k moves upward, the limit on M_5 decreases as the total number of gravitons produced decreases. As k moves downward, the part of the spectrum in which the splittings are resolvable experimentally also moves to lower masses, and the higher background at the lower masses lowers the sensitivity.

The purple (thick solid) curve corresponds to the model-independent search of section 3.2. Bins were considered as significant if their p-values were below 10%. This value represents a fair compromise. A much lower value would reduce the sensitivity to weak signals, while for much higher values the selected regions would have such an extent that their interpretation would become unclear. As this approach covers a vast space of possible models, the bounds are unsurprisingly weaker than with the previous approach.

The green (medium thickness, dashed) curve corresponds to the anomalous region finder of section 3.3. The neural network was trained with signals with k varying from 200 to 2700 GeV and M_5 varying from 1 to 6.5 TeV, where the latter number was obtained by optimization. The reach of this method is somewhat better than that of the first method

thanks to the neural network's ability to assess whether a given excess is signal-like.

The red (thin dashed) curve corresponds to the classifier method of section 3.4. It gives the strongest bounds. One should note, however, that in this case one still needs to account for the look-elsewhere effect.

The orange (thick dashed) curve corresponds to the autoencoder of section 3.5. While the bounds are somewhat weaker than those of some of the other methods, one should note that the look-elsewhere effect is already taken into account in this case.

The black (thin dotted) curve corresponds to the bounds from Fourier analysis as in section 3.6. As in ref. [4], the lower limit of the integral was taken to be $m_{\min} = k$, i.e. where the signal starts, and the upper limit $m_{\max}$ was optimized at each point in the parameter space to maximize the expected significance. The bounds are mostly similar to those obtained using the CWT and a test statistic without machine learning.

It is also interesting to ask how the reach of the CWT-based methods compares with that of more traditional search strategies. Conveniently, CMS have performed a search for the same CW/LD scenario, based on a similar diphoton dataset, looking for a continuum excess at high masses [11]. Their expected 95% CL exclusion limits on M_5 were around $M_5 \approx 10$ TeV, which is comparable to the 2σ reach of $M_5 \approx 8$ TeV that we obtain here. One should keep in mind that the performance of the two approaches may compare differently when applied to other final states, to larger datasets, etc. Even more importantly, while the CW/LD example that we considered here does not present any special challenge to continuum excess searches such as ref. [11], other scenarios with oscillating signals can be elusive to such searches. As we discussed in the Introduction, this can happen due to negative contributions from destructive interference or due to the spectrum not extending to sufficiently high masses. Single-resonance searches provide expected limits of $M_5 \approx 5$ TeV, as obtained in ref. [4] in the approximation that the sensitivity is dominated by the most prominent peak and that neighboring peaks do not hinder the search procedure. Obviously, higher sensitivity will be obtained by combining the contributions of multiple peaks, and the most natural way of implementing this in practice is via a Fourier or wavelet transform, as we do here.

5 Discussion

The LHC experiments have by now developed very comprehensive sets of analyses that provide good coverage of essentially all the simple final states, as well as many exotic ones. They have discovered the Higgs boson, and progress is constantly being made on covering more and more of the parameter space in which physics beyond the Standard Model may be found. While no signs of new physics are seen yet, the theoretical expectation that at least the solution to the electroweak-Planck hierarchy problem is likely to be within the energy reach of the LHC calls for continuing the searches. However, as the LHC experiments have by now matured, improvements in the reach of existing techniques will mostly be gradual. It can therefore be very useful, as an alternative to just waiting, to think about new ways of looking at the data.

In this paper we proposed that wavelet transforms offer such a new way. It is a very general method that can be applied to many different final states at the LHC to search for periodic signals in kinematic distributions in a rather model-independent way. We have also pointed out examples of theoretical models, including some that address the hierarchy problem, for which such searches could be relevant.

We have designed and simulated five different approaches for processing the scalograms produced by the wavelet transforms. Part of these approaches use machine learning techniques, which is another direction into which new physics searches at the LHC can expand, as has been also proposed recently in several other contexts (e.g., refs. [42–52]).

In our first approach, one assumes a specific new physics model and tests for the presence of its signature in the corresponding region of the scalogram using a simple test statistic. In the second, model-independent approach, the whole scalogram is being searched for extended regions of excess, and the most significant excess is assessed using the test statistic. In the third approach, the scalogram is being analyzed by a neutral network, which searches for regions of excess whose shape is consistent with excesses expected in a given class of models. The fourth approach tests for a specific signal using the classifier neural network, whose single output turns out to be a much more powerful test statistic than the more pedestrian test statistic of the first three methods. Finally, the fifth approach is a model-independent analysis in which an autoencoder neural network learns the background only, and identifies potential signals as deviations from a typical background.

We have exemplified the different methods and compared their sensitivities in the context of the diphoton invariant mass spectrum and the clockwork / linear dilaton model. We have also compared them with the Fourier transform method that was proposed in the context of the same model in ref. [4].

Our reach estimates should be viewed as conservative as there are various possible optimizations, either model-dependent or general, that we have left outside the scope of the current study. For example, we have not explored the possibility of using wavelets other than the Morlet wavelet. Also, we have done only very basic optimization of the architecture and the training parameters of the neural networks, so their performance is likely suboptimal. Nevertheless, all the methods resulted in sensitivities of up to 6–8 TeV in M_5, which is comparable to existing searches that would be sensitive to the same scenario [4, 11]. In addition, we would like to emphasize that similar analyses are interesting also in the dilepton [12, 15], dijet [13, 14] and other invariant mass spectra (and possibly other kinematic variables) and that the relative strengths and weaknesses of the different methods can vary depending on the final state, the new physics scenario, and the integrated luminosity.

Wavelet-space searches can be of special importance if the new physics signal does not extend to sufficiently high masses where the background is low; if the new physics makes both positive and negative contributions to the mass spectrum (due to quantum interference); and in situations in which the systematic uncertainty on the normalization of the mass spectrum is a limiting factor. In these kinds of cases, signals can go undetected by existing experimental strategies (given a finite integrated luminosity), but discovered in wavelet-space searches.

In conclusion, we hope to have convinced the reader that making wavelet-space searches part of the ATLAS and CMS toolkits is a possibility worth considering.

Note added: When this work was close to completion, ref. [53] appeared, which also proposed wavelet transforms as a way to search for new physics in kinematic distributions. Our approaches differ substantially due to the fact that ref. [53] uses the discrete wavelet transform based on the Haar wavelet, whereas our methods use the continuous wavelet transform.

Acknowledgments

A Benchmark model: clockwork / linear dilaton (CW/LD)

In the *linear dilaton* scenario, the Standard Model fields propagate on a brane in a space with one relatively large extra dimension. A particular scalar field, the dilaton, with a linear profile in the extra dimension, determines its warped geometry. The motivation for such a setup is its ability to explain the hierarchy between the electroweak and Planck scales. This scenario has first appeared in ref. [54], inspired by the seven-dimensional gravitational dual [55, 56] of Little String Theory [57, 58]. More recently, the same five-dimensional geometry has been rediscovered in [3] while exploring new applications for the *clockwork mechanism* [59–61], in the limit of a large number of sites. Many phenomenological aspects of this scenario have been studied in refs. [1, 2, 4, 62].

Most important for the collider phenomenology of the model are the Kaluza-Klein (KK) gravitons, whose masses (using the notation of ref. [4]) are given by

$$m_n^2 = k^2 + \frac{n^2}{R^2}, \qquad n = 1, 2, 3, \dots . \tag{A.1}$$

The model parameters k and R, which are related to the curvature and size of the extra dimension, are predicted to satisfy $kR \approx 10$ if this scenario is indeed responsible for the hierarchy. This implies a spectrum of narrowly-spaced resonances starting from mass $m_1 \simeq k$, with mass splittings that grow as a function of the mode number n before reaching an asymptotic value of $\Delta m \simeq 1/R \approx k/10$ for $n \gg kR$. Near the beginning of the spectrum, the relative splittings $\Delta m/m$ are around a few percent (almost independent of the value of k), while their decrease as $1/m$ at large n implies that at some point they fall below the experimental resolution. The intrinsic widths of the resonances are almost always negligible relative to the experimental resolution.

The KK graviton fields $h_n^{\mu\nu}$ couple to the Standard Model stress-energy tensor $T_{\mu\nu}$ as

$$\mathcal{L} \supset -\frac{1}{\Lambda_n} h_n^{\mu\nu} T_{\mu\nu}, \tag{A.2}$$

where

$$\Lambda_n^2 = M_5^3 \pi R \left(1 + \frac{k^2 R^2}{n^2} \right) . \tag{A.3}$$

Here M_5 is the five-dimensional reduced Planck mass, which is the fundamental scale of the theory. These couplings allow the KK gravitons to be produced from gg and $q\bar{q}$ in pp collisions with the cross sections

$$\sigma(pp \to G_n) = \frac{\pi}{48\Lambda_n^2} \left(3\,\mathcal{L}_{gg}(m_n^2) + 4 \sum_q \mathcal{L}_{q\bar{q}}(m_n^2) \right), \tag{A.4}$$

where

$$\mathcal{L}_{ij}(\hat{s}) = \frac{\hat{s}}{s} \int_{\hat{s}/s}^1 \frac{dx}{x} \, f_i(x) \, f_j \left(\frac{\hat{s}}{xs} \right) \tag{A.5}$$

are the parton luminosities, for which we take the LO MSTW2008 PDFs [63]. These couplings also allow the KK gravitons to decay to pairs of Standard Model particles, including $\gamma\gamma$. Heavy KK gravitons can also decay to pairs of lighter KK gravitons or KK scalars. We account for these decays in computing the $\gamma\gamma$ branching fraction, taking the case of rigid boundary conditions for the dilaton. However, since these decays start having an impact only for $m \gg k$, their effect is insignificant in the range of parameters we consider in this work. For additional details, see ref. [4]. The $\gamma\gamma$ branching fraction ends up being about 4%, almost independent of the model parameters or the KK graviton mass.

Since the parameters M_5, k and R must combine to give the known value of the four-dimensional reduced Planck mass, $M_P \equiv 1/\sqrt{8\pi G}$, as $M_P^2 = (e^{2\pi k R} - 1)M_5^3/k$, only two of the parameters are independent, and we choose them to be M_5 and k. The parameter k determines the beginning of the KK graviton spectrum, while M_5 fixes the cross section, which is approximately proportional to $1/M_5^3$.

We assume the experimental resolution in the diphoton invariant mass to be

$$\frac{\sigma(m_{\gamma\gamma})}{m_{\gamma\gamma}} = \sqrt{\frac{a^2}{m_{\gamma\gamma}(\text{GeV})} + \frac{c^2}{2}}, \tag{A.6}$$

with

$$a = 12\%, \qquad c = 1\%, \tag{A.7}$$

which is based on partial information from refs. [10, 64–66]. This resolution is used in figure 1 and in the rest of the paper. We take the background from the search for heavy resonances decaying to two photons of ref. [10] (37 fb^{-1} at 13 TeV), the "Spin-2 selection". A constant value of $\epsilon = 0.5$ is taken for the acceptance times efficiency.

References

[1] I. Antoniadis, A. Arvanitaki, S. Dimopoulos, and A. Giveon, "Phenomenology of TeV Little String Theory from Holography," *Phys. Rev. Lett.* **108** (2012) 081602, arXiv:1102.4043 [hep-ph].

[2] M. Baryakhtar, "Graviton Phenomenology of Linear Dilaton Geometries," *Phys. Rev.* **D85** (2012) 125019, arXiv:1202.6674 [hep-ph].

[3] G. F. Giudice and M. McCullough, "A Clockwork Theory," *JHEP* **02** (2017) 036, `arXiv:1610.07962 [hep-ph]`.

[4] G. F. Giudice, Y. Kats, M. McCullough, R. Torre, and A. Urbano, "Clockwork/linear dilaton: structure and phenomenology," *JHEP* **06** (2018) 009, `arXiv:1711.08437 [hep-ph]`.

[5] G. F. Giudice, T. Plehn, and A. Strumia, "Graviton collider effects in one and more large extra dimensions," *Nucl. Phys.* **B706** (2005) 455–483, `arXiv:hep-ph/0408320`.

[6] A. V. Kisselev, "RS model with a small curvature and two-photon production at the LHC," *JHEP* **09** (2008) 039, `arXiv:0804.3941 [hep-ph]`.

[7] R. Franceschini, P. P. Giardino, G. F. Giudice, P. Lodone, and A. Strumia, "LHC bounds on large extra dimensions," *JHEP* **05** (2011) 092, `arXiv:1101.4919 [hep-ph]`.

[8] L. Randall and R. Sundrum, "Large mass hierarchy from a small extra dimension," *Phys. Rev. Lett.* **83** (1999) 3370–3373, `arXiv:hep-ph/9905221`.

[9] P. Brax, S. Fichet, and P. Tanedo, "The Warped Dark Sector," *Phys. Lett.* **B798** (2019) 135012, `arXiv:1906.02199 [hep-ph]`.

[10] ATLAS Collaboration, M. Aaboud *et al.*, "Search for new phenomena in high-mass diphoton final states using 37 fb^{-1} of proton–proton collisions collected at $\sqrt{s} = 13$ TeV with the ATLAS detector," *Phys. Lett.* **B775** (2017) 105–125, `arXiv:1707.04147 [hep-ex]`.

[11] CMS Collaboration, A. M. Sirunyan *et al.*, "Search for physics beyond the standard model in high-mass diphoton events from proton-proton collisions at $\sqrt{s} = 13$ TeV," *Phys. Rev.* **D98** no. 9, (2018) 092001, `arXiv:1809.00327 [hep-ex]`.

[12] ATLAS Collaboration, G. Aad *et al.*, "Search for high-mass dilepton resonances using 139 fb^{-1} of pp collision data collected at $\sqrt{s} = 13$ TeV with the ATLAS detector," *Phys. Lett.* **B796** (2019) 68–87, `arXiv:1903.06248 [hep-ex]`.

[13] ATLAS Collaboration, "Search for New Phenomena in Dijet Events using 139 fb^{-1} of pp collisions at $\sqrt{s} = 13$ TeV collected with the ATLAS Detector," Tech. Rep. ATLAS-CONF-2019-007, CERN, Geneva, Mar, 2019. `https://cds.cern.ch/record/2668385`.

[14] CMS Collaboration, "Searches for dijet resonances in pp collisions at $\sqrt{s} = 13$ TeV using the 2016 and 2017 datasets," Tech. Rep. CMS-PAS-EXO-17-026, CERN, Geneva, 2018. `http://cds.cern.ch/record/2637847`.

[15] CMS Collaboration, A. M. Sirunyan *et al.*, "Search for contact interactions and large extra dimensions in the dilepton mass spectra from proton-proton collisions at $\sqrt{s} = 13$ TeV," *JHEP* **04** (2019) 114, `arXiv:1812.10443 [hep-ex]`.

[16] A. Djouadi, J. Ellis, and J. Quevillon, "Interference effects in the decays of spin-zero resonances into $\gamma\gamma$ and $t\bar{t}$," *JHEP* **07** (2016) 105, `arXiv:1605.00542 [hep-ph]`.

[17] S. P. Martin, "Signal-background interference for a singlet spin-0 digluon resonance at the LHC," *Phys. Rev.* **D94** no. 3, (2016) 035003, `arXiv:1606.03026 [hep-ph]`.

[18] N. Craig, S. Renner, and D. Sutherland, "Exploring Peaks and Valleys in the Diphoton Spectrum," *Phys. Rev.* **D94** no. 7, (2016) 075011, `arXiv:1607.06074 [hep-ph]`.

[19] M. Spalla, "Search for a jet-jet resonance associated with a W/Z, decaying leptonically, with the ATLAS detector: application of a multiresolution analysis," Master's thesis, Università Di Pisa, 2013. `https://etd.adm.unipi.it/theses/available/etd-10022013-194238/`.

[20] P. Lachowicz and B. Czerny, "Wavelet analysis of millisecond variability of Cygnus X-1 during its failed state transition," *MNRAS* **361** (2005) 645–658.

[21] I. Kiskin, D. Zilli, Y. Li, M. Sinka, K. Willis, and S. Roberts, "Bioacoustic detection with wavelet-conditioned convolutional neural networks," *Neural Comput. Applic.* (2018) . https://doi.org/10.1007/s00521-018-3626-7.

[22] R. Benitez, V. Bolós, and M. Ramírez, "A wavelet-based tool for studying non-periodicity," *CAMWA* **60** no. 3, (2010) 634–641.

[23] P. Kumar and E. Foufoula-Georgiou, "Wavelet Analysis for Geophysical Applications," *Rev. Geophys.* **35** (1997) 385–412.

[24] E. Flinchem and D. Jay, "An introduction to wavelet transform tidal analysis methods," *Estuarine, Coastal and Shelf Science* **51** no. 2, (2000) 177–200.

[25] H. Heidary, A. Refahi Oskouei, M. Hajikhani, B. Moosaloo, and M. Ahmadi Najafabadi, "Acoustic Emission Signal Analysis by Wavelet Method to Investigate Damage Mechanisms During Drilling of Composite Materials," in *ASME 2010 10th Biennial Conference on Engineering Systems Design and Analysis, ESDA2010*, vol. 1, pp. 409–416. 2010.

[26] I. W. Selesnick, R. G. Baraniuk, and N. C. Kingsbury, "The dual-tree complex wavelet transform," *IEEE Signal Proc. Mag.* **22** no. 6, (2005) 123–151.

[27] P. Lachowicz, "Wavelet analysis: a new significance test for signals dominated by intrinsic red-noise variability," `arXiv:0906.4176 [astro-ph.HE]`.

[28] J. A. Schulte, "Cumulative areawise testing in wavelet analysis and its application to geophysical time series," *Nonlin. Processes Geophys.* **23** no. 1, (2016) 45–57.

[29] J. A. Schulte, "Wavelet analysis for non-stationary, nonlinear time series," *Nonlin. Processes Geophys.* **23** no. 4, (2016) 257–267.

[30] D. Maraun, J. Kurths, and M. Holschneider, "Nonstationary Gaussian processes in wavelet domain: Synthesis, estimation, and significance testing," *Phys. Rev.* **E75** (2007) 016707.

[31] T. Rouyer, J.-M. Fromentin, N. C. Stenseth, and B. Cazelles, "Analysing multiple time series and extending significance testing in wavelet analysis," *MEPS* **359** (2008) 11–23.

[32] Z. Ge, "Significance tests for the wavelet power and the wavelet power spectrum," *Ann. Geophys.* **25** no. 11, (2007) 2259–2269.

[33] D. Maraun and J. Kurths, "Cross wavelet analysis: Significance testing and pitfalls," *Nonlin. Processes Geophys.* **11** (2004) 505–514.

[34] J. A. Schulte, C. Duffy, and R. G. Najjar, "Geometric and topological approaches to significance testing in wavelet analysis," *Nonlin. Processes Geophys.* **22** no. 2, (2015) 139–156.

[35] C. G. Torrence and G. Compo, "A Practical Guide to Wavelet Analysis," *Bulletin of the American Meteorological Society* **79** (1998) 61–78.

[36] G. R. Lee, R. Gommers, F. Wasilewski, K. Wohlfahrt, A. O'Leary, H. Nahrstaedt, and Contributors, "PyWavelets - Wavelet Transforms in Python." https://github.com/PyWavelets/pywt. 2019.

[37] G. R. Lee, R. Gommers, F. Wasilewski, K. Wohlfahrt, and A. O'Leary, "PyWavelets: A Python package for wavelet analysis," *JOSS* **4(36)** (2019) 1237.

[38] R. A. Fisher, *Statistical Methods for Research Workers, v.5.* Edinburgh: Oliver and Boyd, 1934.

[39] F. Chollet *et al.*, "Keras." `https://keras.io`, 2015.

[40] M. Abadi *et al.*, "TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems," `arXiv:1603.04467 [cs.DC]`.

[41] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," `arXiv:1412.6980 [cs.LG]`.

[42] T. Heimel, G. Kasieczka, T. Plehn, and J. M. Thompson, "QCD or What?," *SciPost Phys.* **6** no. 3, (2019) 030, `arXiv:1808.08979 [hep-ph]`.

[43] M. Farina, Y. Nakai, and D. Shih, "Searching for New Physics with Deep Autoencoders," `arXiv:1808.08992 [hep-ph]`.

[44] T. Cohen, M. Freytsis, and B. Ostdiek, "(Machine) Learning to Do More with Less," *JHEP* **02** (2018) 034, `arXiv:1706.09451 [hep-ph]`.

[45] 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 [hep-ph]`.

[46] J. H. Collins, K. Howe, and B. Nachman, "Anomaly Detection for Resonant New Physics with Machine Learning," *Phys. Rev. Lett.* **121** no. 24, (2018) 241803, `arXiv:1805.02664 [hep-ph]`.

[47] R. T. D'Agnolo and A. Wulzer, "Learning New Physics from a Machine," *Phys. Rev.* **D99** no. 1, (2019) 015014, `arXiv:1806.02350 [hep-ph]`.

[48] A. De Simone and T. Jacques, "Guiding New Physics Searches with Unsupervised Learning," *Eur. Phys. J.* **C79** no. 4, (2019) 289, `arXiv:1807.06038 [hep-ph]`.

[49] J. Hajer, Y.-Y. Li, T. Liu, and H. Wang, "Novelty Detection Meets Collider Physics," `arXiv:1807.10261 [hep-ph]`.

[50] O. Cerri, T. Q. Nguyen, M. Pierini, M. Spiropulu, and J.-R. Vlimant, "Variational Autoencoders for New Physics Mining at the Large Hadron Collider," *JHEP* **05** (2019) 036, `arXiv:1811.10276 [hep-ex]`.

[51] J. H. Collins, K. Howe, and B. Nachman, "Extending the search for new resonances with machine learning," *Phys. Rev.* **D99** no. 1, (2019) 014038, `arXiv:1902.02634 [hep-ph]`.

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

[53] B. G. Lillard, T. Plehn, A. Romero, and T. M. P. Tait, "Multi-scale Mining of Kinematic Distributions with Wavelets," `arXiv:1906.10890 [hep-ph]`.

[54] I. Antoniadis, S. Dimopoulos, and A. Giveon, "Little string theory at a TeV," *JHEP* **05** (2001) 055, `arXiv:hep-th/0103033`.

[55] O. Aharony, M. Berkooz, D. Kutasov, and N. Seiberg, "Linear dilatons, NS five-branes and holography," *JHEP* **10** (1998) 004, `arXiv:hep-th/9808149`.

[56] A. Giveon and D. Kutasov, "Little string theory in a double scaling limit," *JHEP* **10** (1999) 034, `arXiv:hep-th/9909110`.

[57] M. Berkooz, M. Rozali, and N. Seiberg, "Matrix description of M-theory on T^4 and T^5," *Phys. Lett.* **B408** (1997) 105–110, `arXiv:hep-th/9704089`.

[58] N. Seiberg, "New theories in six-dimensions and matrix description of M-theory on T^5 and T^5/Z_2," *Phys. Lett.* **B408** (1997) 98–104, `arXiv:hep-th/9705221`.

[59] K. Choi, H. Kim, and S. Yun, "Natural inflation with multiple sub-Planckian axions," *Phys. Rev.* **D90** (2014) 023545, `arXiv:1404.6209 [hep-th]`.

[60] K. Choi and S. H. Im, "Realizing the relaxion from multiple axions and its UV completion with high scale supersymmetry," *JHEP* **01** (2016) 149, `arXiv:1511.00132 [hep-ph]`.

[61] D. E. Kaplan and R. Rattazzi, "Large field excursions and approximate discrete symmetries from a clockwork axion," *Phys. Rev.* **D93** no. 8, (2016) 085007, `arXiv:1511.01827 [hep-ph]`.

[62] P. Cox and T. Gherghetta, "Radion Dynamics and Phenomenology in the Linear Dilaton Model," *JHEP* **05** (2012) 149, `arXiv:1203.5870 [hep-ph]`.

[63] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt, "Parton distributions for the LHC," *Eur. Phys. J.* **C63** (2009) 189–285, `arXiv:0901.0002 [hep-ph]`.

[64] "Search for resonances decaying to photon pairs in 3.2 fb^{-1} of pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector," Tech. Rep. ATLAS-CONF-2015-081, CERN, Geneva, Dec, 2015. `http://cds.cern.ch/record/2114853`.

[65] ATLAS Collaboration, "Search for scalar diphoton resonances with 15.4 fb^{-1} of data collected at $\sqrt{s}$=13 TeV in 2015 and 2016 with the ATLAS detector," Tech. Rep. ATLAS-CONF-2016-059, CERN, Geneva, Aug, 2016. `http://cds.cern.ch/record/2206154`.

[66] CMS Collaboration, V. Khachatryan *et al.*, "Search for high-mass diphoton resonances in proton–proton collisions at 13 TeV and combination with 8 TeV search," *Phys. Lett.* **B767** (2017) 147–170, `arXiv:1609.02507 [hep-ex]`.