involve

A modified wavelet method
for identifying transient features in time signals
with applications to bean beetle maturation

David McMorris, Paul Pearson and Brian Yurk

msp

A modified wavelet method
for identifying transient features in time signals
with applications to bean beetle maturation

David McMorris, Paul Pearson and Brian Yurk

(Communicated by Kenneth S. Berenhaut)

We develop an averaging method, based on a modified Haar wavelet technique, for identifying when transient features occur in a time signal. We call this method the seaweed method and use it to identify different stages of bean beetle embryo maturation. We use randomized simulations to evaluate the seaweed method for accuracy and precision at different levels of noise. Our results support the efficacy of the seaweed method as a means for analyzing time-lapse photographs of bean beetle embryos and a wide variety of other time signals.

1. Introduction

Wavelets are commonly employed in signal processing for their usefulness in detecting sudden changes in a signal. In this paper we introduce a new method based a modification of a Haar wavelet transform. This method, which we call the "seaweed method," is designed to identify points signifying the beginning of a gradual change in the signal, indicating a change in the sign of the slope on a range of time scales. We call such points "transition points," as they often indicate important transitions in the underlying signal. We show this method is effective at identifying transitions in a noisy signal of pixel brightness values coming from time-lapse photography. We believe this method will prove capable of analyzing many kinds of noisy time signals, including audio signals, stock market data, and electrical signals from sensors. Spatial signals, such as population density of a given organism along a transect, may also benefit from analysis by the seaweed method.

The seaweed method was motivated by the need to identify biological markers in developing bean beetle (*Callosobruchus maculatus*) embryos, for the purpose of studying factors that influence insect development time. Ultimately, these studies should yield insights into how climate change will impact insect phenology and,
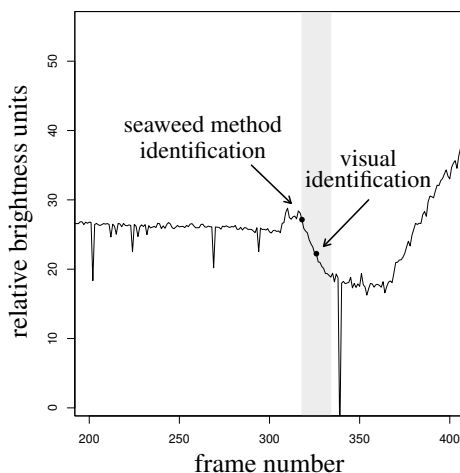
Figure 1. Graph of relative brightness of a particular bean beetle egg over time between frames 200 and 400. Frames are 20 minutes apart. The shaded region indicates the darkening of the embryo's head capsule, and arrows indicate the differences in identification of this marker by eye versus using the seaweed method. Also note that in this signal, the large spikes down are the spurious results of beetles crossing the calibration region on the bean.

hence, population dynamics (see, e.g., [Logan and Powell 2001; Logan et al. 1976; Yurk and Powell 2009]).

Bean beetles are agricultural pests native to Africa and Asia that infest legume crops both in fields and storehouses [Beck and Blumer 2011]. These insects lay their eggs on the surfaces of beans. Following the completion of embryonic development, the larvae burrow into the bean, where they feed and complete development before emerging as adults. In addition to being an important agricultural pest, bean beetles are studied as model organisms, as they are readily obtainable and easy to work with in a laboratory.

In order to better understand the timing of bean beetle embryonic development, we use biological markers to distinguish between different stages of maturation. The marker we use to signify the completion of embryonic development is the formation and darkening of the larval head capsule, which appears as a darkening in the eggs. Although this spot is visible to the naked eye, its formation is so gradual that visual identification of its initial darkening is impossible. Figure 14 shows a sequence of digital microscope images which demonstrates this darkening process.

Because of the difficulties posed by visually identifying the beginning of head-capsule darkening, we used digital time-lapse microscopy to record changes in color and brightness of the eggs at 20-minute intervals. Note that, unlike the images

in Figure 14, these images include many different eggs rather than a single egg. For each egg in the images, we focused on a small, fixed rectangular region of the images containing the egg. We used these images to investigate embryonic development time, and developed a method to pinpoint the precise time when maturation had finished. In each image, the brightness level of an egg, relative to a calibration region on the bean, can be extracted as the difference of two average grayscale values from different regions of the image. In the same way, we can extract a signal consisting of the relative brightness of a specific egg from a sequence of images. A portion of one such signal is displayed in Figure 1. Note that these signals are scaled and shifted for consistency, so the resulting quantities would be dimensionless. Based on our visual observations, we know that the embryo's head capsule darkens along the decline in the shaded region of Figure 1. Whereas a purely visual analysis of the eggs would identify the head-capsule darkening somewhere in the middle of this shaded region, the extraction of this signal and application of the seaweed method make it possible to accurately locate the beginning of this region. This allows for more precise measurement of embryonic development time. In general, the shape of the signal seen in Figure 1 is common to all the eggs we analyze.

The seaweed method is a graphical method based on a modification of the Haar wavelet algorithm. Herein we present the seaweed method, as well as statistical tests to validate it.

2. Background

2A. *Wavelets.* Discrete and continuous wavelet transforms have been used extensively over the last few decades to analyze time signals, as discussed in [Aboufadel and Schlicker 1999; Bénéteau and Van Fleet 2011; Mulcahy 1996]. They provide a way to express a time signal as a sum of component waves. Each component wave has a fixed frequency (or, equivalently, a fixed period) and is a sum of wavelets obtained from a "mother" wavelet by time shifting, time dilation, and amplitude scaling. These wavelet transforms measure changes in amplitude of all component waves over time and frequency simultaneously, thereby allowing signal features to be identified. Signal features include changes in amplitude over a range of time that identifies when an event occurred, changes in amplitude over a range of frequencies that identify the frequency signature of an event, or a combination of the two. This paper will focus on step detection (or change detection) in a time signal using a variation of the Haar wavelet transform.

The discrete wavelet transform separates a signal into a direct sum of wavelets by requiring that amplitude measurements for component waves occur over rectangles of equal area that tile the time-frequency plane but do not overlap, as shown in Figure 2. The width of the rectangles in the top row of Figure 2 is determined by the sampling rate of the signal.
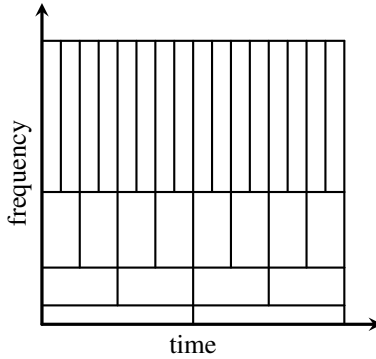
Figure 2. An example of the tiling of the time-frequency plane used for the discrete wavelet transform (DWT). The DWT expresses the signal as a sum of wavelets with one wavelet for each rectangle.

In contrast, the continuous wavelet transform performs more amplitude calculations by increasing the number of rectangular tiles in the time-frequency plane and allowing them to overlap. This overlap in the continuous wavelet transform allows for amplitude measurements in the time-frequency plane to vary continuously both in the time direction and in the frequency direction.

The Haar wavelet transform will be used as the basis for signal processing by wavelets in this paper. The prototype Haar wavelet and Haar scaling function are defined as follows. The Haar "mother" wavelet is the function $\psi : \mathbb{R} \to \mathbb{R}$ defined by

$$\psi(t) = \begin{cases} -1 & \text{if } 0 \leq t < \frac{1}{2}, \\ 1 & \text{if } \frac{1}{2} \leq t < 1, \\ 0 & \text{otherwise.} \end{cases} \tag{2-1}$$

Contrary to standard convention, we have chosen $\psi(t)$ to be an increasing, rather than decreasing, step function over $0 \leq t < 1$. This choice will make interpreting amplitude coefficients for Haar wavelets easier since positive coefficients will indicate an increasing function and negative coefficients a decreasing function. The Haar scaling function is the function $\phi : \mathbb{R} \to \mathbb{R}$ defined by

$$\phi(t) = \begin{cases} 1 & \text{if } 0 \leq t < 1, \\ 0 & \text{otherwise.} \end{cases} \tag{2-2}$$

2B. *Averaging and differencing.* In [Bénéteau and Van Fleet 2011; Mulcahy 1996], the discrete Haar wavelet transformation is described as a process of averaging and "differencing" as follows. Suppose a discrete, real-valued time signal is sampled at a constant rate, that Δt is the time between consecutive signal measurements, and that s_n denotes the n-th signal value ($n \geq 1$). The process of averaging and differencing starts by setting $\alpha_{0,n} = s_n$. Then, for each $j = 1, 2, 3, \ldots$ and each

	averages										differences							
$j \setminus n$	1	2	3	4	5	6	7	8		$j \setminus n$	1	2	3	4	5	6	7	8
0	8	20	32	36	32	20	8	4		0								
1		14		34		26		6		1		6		2		−6		−2
2				24				16		2				10				−10
3								20		3								−4

Table 1. Averages and differences arrays.

$n = 2^j,\ 2 \cdot 2^j,\ 3 \cdot 2^j, \ldots$, averages $\alpha_{j,n}$ and "differences" $\delta_{j,n}$ are defined by

$$\alpha_{j,n} = \frac{\alpha_{j-1,n-2^{j-1}} + \alpha_{j-1,n}}{2}, \quad \delta_{j,n} = \frac{\alpha_{j-1,n} - \alpha_{j-1,n-2^{j-1}}}{2}. \tag{2-3}$$

Each $\alpha_{j,n}$ is the average of 2^j signal samples, while each $\delta_{j,n}$ is half of the difference between two average values of the signal, where each average is of 2^{j-1} signal values. Thus, each $\delta_{j,n}$ measures changes in the average value of the signal. The averages $\alpha_{j,n}$ and differences $\delta_{j,n}$ are amplitude coefficients for time-shifted and time-dilated Haar scaling functions $\phi_{j,n}(t)$ and Haar wavelets $\psi_{j,n}(t)$, defined as

$$\phi_{j,n}(t) = \phi\left(\frac{t - (n - 2^j + 1)\Delta t}{2^j \Delta t}\right), \quad \psi_{j,n}(t) = \psi\left(\frac{t - (n - 2^j + 1)\Delta t}{2^j \Delta t}\right). \tag{2-4}$$

The time window over which $\phi_{j,n}(t)$ and $\psi_{j,n}(t)$ are nonzero has length $2^j \Delta t$.

Example 2.1. Suppose $s(t)$ is a time signal with amplitude measurements 8, 20, 32, 36, 32, 20, 8, 4 sampled uniformly over the time interval $0 \le t \le 8$, so that $\Delta t = 1$. By (2-3), the averages and differences for this signal are given by the values in Table 1.

The signal can be written as a sum of Haar scaling functions and Haar wavelets over disjoint time intervals of length 2 by using the row $j = 1$ averages and differences as coefficients:

$$s(t) = \sum_{n=2,4,6,8} \alpha_{1,n}\phi_{1,n}(t) + \sum_{n=2,4,6,8} \delta_{1,n}\psi_{1,n}(t). \tag{2-5}$$

Graphs of the signal, the Haar scaling functions, and the Haar wavelets from (2-5) are shown in the rows $j = 0, 1$ of Figure 3. Just as the signal was split into a sum of Haar scaling functions and Haar wavelets, the average values in (2-5) can be treated as a time signal and split into a sum of Haar scaling functions and Haar wavelets:

$$s(t) = \sum_{n=4,8} \alpha_{2,n}\phi_{2,n}(t) + \sum_{n=4,8} \delta_{2,n}\psi_{2,n}(t) + \sum_{n=2,4,6,8} \delta_{1,n}\psi_{1,n}(t). \tag{2-6}$$
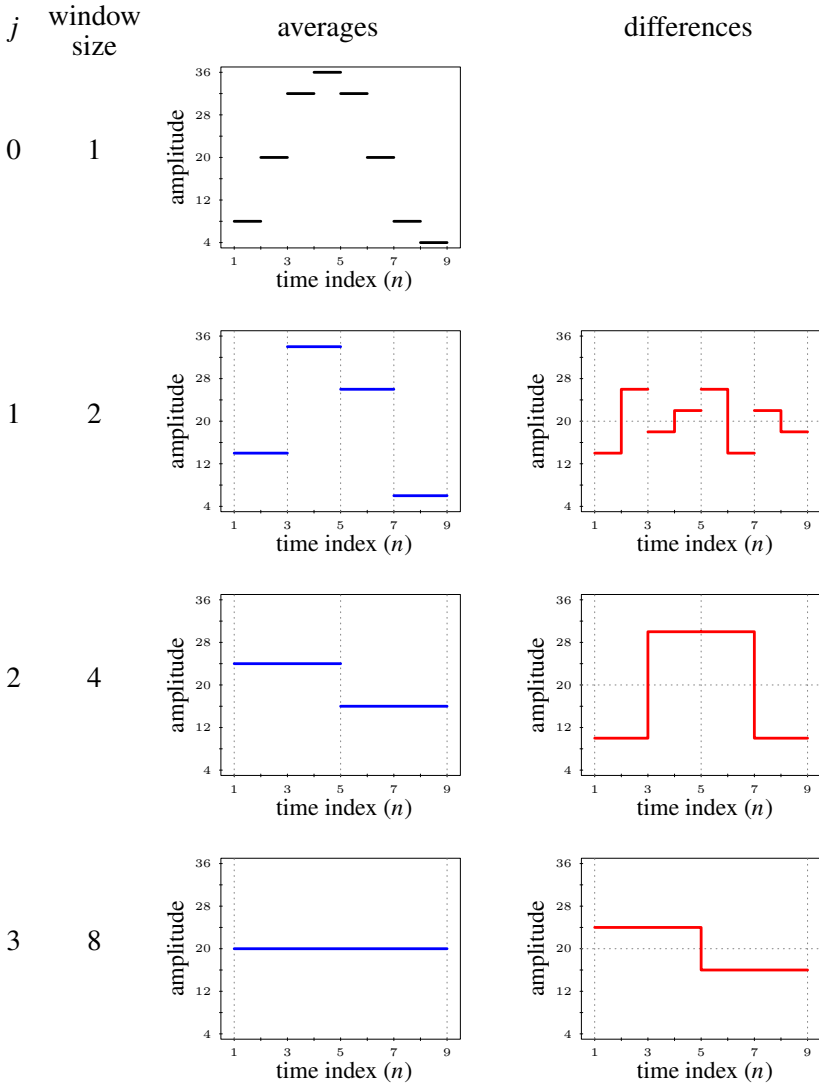
Figure 3. Plot of the signal (row $j = 0$), Haar scaling functions or average values of the signal (left graphs in rows $j = 1, 2, 3$), and Haar wavelets or component waves (right graphs in rows $j = 1, 2, 3$). The amplitudes of the Haar scaling functions and the Haar wavelets come from the coefficients in the averages array and differences array, respectively.

Equation (2-6) shows that the signal is a sum of Haar scaling functions and Haar wavelets over disjoint time intervals of length 4 plus Haar wavelets over disjoint time intervals of length 2, as shown in the rows $j = 1, 2$ of Figure 3. Similarly, the average
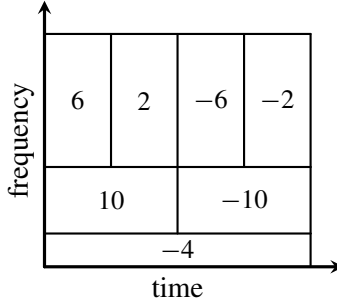
Figure 4. A plot of the Haar wavelet amplitude coefficients $\delta_{j,n}$ in the time-frequency plane can be used to identify features in the signal. For instance, the amplitudes change from positive values $(2, 10)$ to negative values $(-6, -10, -4)$ at time $t = 4$, indicating that the signal changes from increasing to decreasing across multiple frequencies at the same time. Note: the amplitude coefficients in the plot are horizontally centered in each time window (rectangle) to facilitate identifying the time at which a signal feature occurs.

values in (2-6) can be treated as a time signal and split into a sum of Haar scaling functions and Haar wavelets as in (2-7) and in the rows $j = 1, 2, 3$ of Figure 3:

$$s(t) = \sum_{n=8} \alpha_{3,n}\phi_{3,n}(t) + \sum_{n=8} \delta_{3,n}\psi_{3,n}(t) + \sum_{n=4,8} \delta_{2,n}\psi_{2,n}(t) + \sum_{n=2,4,6,8} \delta_{1,n}\psi_{1,n}(t).$$

(2-7)

Since each difference $\delta_{j,n}$ represents the amplitude of a Haar wavelet $\psi_{j,n}(t)$ that has frequency $1/(2^j \Delta t)$, the differences $\delta_{j,n}$ can be plotted in the time-frequency plane as a means to identify signal features, as shown in Figure 4.

2C. *Redefining the differences.* The Haar wavelet coefficients $\delta_{j,n}$ are defined recursively in (2-3), but they can be defined explicitly in terms of the signal values:

$$\delta_{j,n} = \frac{1}{2^j}\left(\sum_{i=0}^{2^{j-1}-1} s_{n-i} - \sum_{i=2^{j-1}}^{2^j-1} s_{n-i}\right).$$

(2-8)

By writing $1/2^j$ as $\frac{1}{2}(1/2^{j-1})$ and distributing the factor $1/2^{j-1}$ to each term in the sum in (2-8), we see that (2-8) succinctly states that $\delta_{j,n}$ is half of the difference between the average of the 2^{j-1} signal values between s_{n-2^j+1} and $s_{n-2^{j-1}}$ and the average of the 2^{j-1} signal values between $s_{n-2^{j-1}+1}$ and s_n. In other words, $\delta_{j,n}$ is one half of the difference between two average values of the signal over two adjacent time intervals. Each $\delta_{j,n}$ is the amplitude of a Haar wavelet $\psi_{j,n}(t)$ of wavelength $2^j \Delta t$ and frequency $1/(2^j \Delta t)$.

3. Methods

3A. *Modified algorithm.* We now modify the traditional Haar wavelet algorithm by adding more rows to the differences array thereby making it dense, rather than sparse. Adding more rows means having a higher resolution in the frequency domain, while making it dense means using a sliding time window. A sliding window can provide a better resolution in time and is accomplished by allowing n to be any integer, not just a multiple of a power of 2. In this regard, our modified algorithm provides a discrete approximation to the continuous Haar wavelet transform. We define $d_{j,n}$ to be the difference

$$d_{j,n} = \frac{1}{j}\left(\sum_{i=0}^{j-1} s_{n-i} - \sum_{i=j}^{2j-1} s_{n-i} \right). \tag{3-1}$$

In this definition, $2j$ is the size of the window being considered, and with each increase of n, this window is shifted to the right by one time step. Equation (3-1) states that each $d_{j,n}$ is the difference between the average of the j signal values between s_{n-j+1} and s_n and the average of the j signal values between s_{n-2j+1} and s_{n-j}.

There are several benefits of this modified algorithm. This algorithm examines differences in the average values of the signal, making it is easy to describe conceptually. Due to the addition of a sliding window, this method has a higher resolution for detecting signal features than the discrete Haar wavelet transform, as presented in Section 2B. One drawback is that this method is specific to the Haar wavelet and does not immediately generalize to other types of wavelets. Another drawback is that it is more computationally intensive than the discrete Haar wavelet transform, and may even be more computationally intensive than a continuous Haar wavelet transform that has been optimized for speed using the fast Fourier transform.

3B. *Related work.* The modified Haar wavelet algorithm in (3-1) is perhaps best thought of as a discrete approximation to the continuous Haar wavelet transform. It also has some similarities with wavelet frame theory [Christensen 2001; Daubechies et al. 2003; Teolis 1998] and overcomplete discrete wavelet transforms (OCDWT) [Auscher 1992; Bayram and Selesnick 2009; Selesnick 2011; Teolis 1998].

Using Haar wavelet frames, a signal can be represented as a linear combination of wavelets from a frame, which consists of a set of wavelets that span the function space (such as $L^1([a,b])$) but are not necessarily linearly independent, which means the coefficients in the linear combination are not necessarily uniquely determined. Using the modified Haar wavelet algorithm in (3-1), however, the signal is represented uniquely relative to a Haar wavelet basis *within each time window of length* $2j\Delta t$, and time windows are allowed to slide (and overlap) by

changing the value of n. A frame could be formed by taking the union of all wavelet bases over all time windows. However, in this study we choose to interpret the wavelet coefficients provided by (3-1) relative to individual time windows that slide through time, rather than relative to the union of all time windows.

The OCDWT and the modified Haar wavelet algorithm in (3-1) are both characterized by denser time-frequency lattices arising from increasing the resolution in the frequency domain, as can be seen in Figure 2 in [Bayram and Selesnick 2009]. The OCDWT has an inherent degree of noise robustness due to redundancy in its representation [Teolis 1998, p. 134], and the modified Haar wavelet algorithm also has this noise robustness due to redundancy. The modified Haar wavelet algorithm has a sliding time window that the OCDWT does not, which means that its time-frequency lattice is more dense than the OCDWT because its time windows slide and are allowed to overlap.

3C. *Identifying transition points.*   At this point, we employ the modified algorithm introduced in Section 3A and look for transition points in the signal by locating sign changes between $d_{j,n-1}$ and $d_{j,n}$. When the sign function (the function which takes a number and returns its sign as ± 1 or 0) of the product,

$$p_{j,n} = \text{sgn}(d_{j,n-1} \cdot d_{j,n}) \qquad (3\text{-}2)$$

is nonpositive, either a sign change has occurred or the signal is constant (which in practice is a rare occurrence). A sign change indicates that the corresponding Haar wavelets $d_{j,n-1}\psi_{j,n-1}(t)$ and $d_{j,n}\psi_{j,n}(t)$ changed from either constant or increasing to decreasing or that they changed from either constant or decreasing to increasing at time index n. This process of identifying transitions occurs across multiple time scales as the size of the sliding window increases.

When a transition point is identified, we plot a point at $(n, 2j)$. The type of point we plot depends on the kind of transition that occurs. If the transition signifies an increase in the signal, a closed circle is plotted. If the transition signifies a decrease, an open circle is used. Finally, if the transition signifies that the signal is constant, an open triangle is used. We illustrate this process with the following example.

Example 3.1. Consider the time signal $s(t)$ defined in Example 2.1. The new difference array and corresponding array of product signs for this signal are displayed in Table 2.

Based on the array of products in Table 2, it is clear that transition points have been identified at $n = 5, 6, 7$ using windows of size $2j = 2, 4, 6$ respectively. Figure 5 shows the signal as a piecewise-linear representation as well as the discrete representation with all identified transition points overlaid. The strand of points in Figure 5 identifies the region where the transition occurs in the signal. We call these strands of points "seaweed" due to their visual appearance for more complicated

	differences							
$j \setminus n$	1	2	3	4	5	6	7	8
1		12	12	4	-4	-12	-12	-4
2				20	8	-8	-20	-20
3					$\frac{28}{3}$	$-\frac{28}{3}$	$-\frac{68}{3}$
4								-8

	products							
$j \setminus n$	1	2	3	4	5	6	7	8
1			+	+	$-$	+	+	+
2					+	$-$	+	+
3							$-$	+
4								

Table 2. Array of differences between averages on each window $(2j = 2, 4, 6, 8)$ as they slide across the time domain, and corresponding product sign array. A negative sign indicates that a transition has occurred.

signals. The presence of multiple points in this strand reveals that this transition is apparent across multiple scales, and that it is seen with the highest resolution at the lowest point in the strand. For this reason, the transition point we are looking for is the lowest point, $(n, 2j) = (5, 2)$. In Section 3D we will introduce guidelines for identifying transition points in more general cases involving noisier signals.

Figure 6 depicts a more detailed look at the way in which the sign changes are identified and marked using the algorithm. The first graph in each row shows the last difference for a certain window that is positive; the second shows the first difference at that scale that is negative. As the window shifts from a positive difference to a negative difference, a point is plotted to indicate a transition has occurred. An
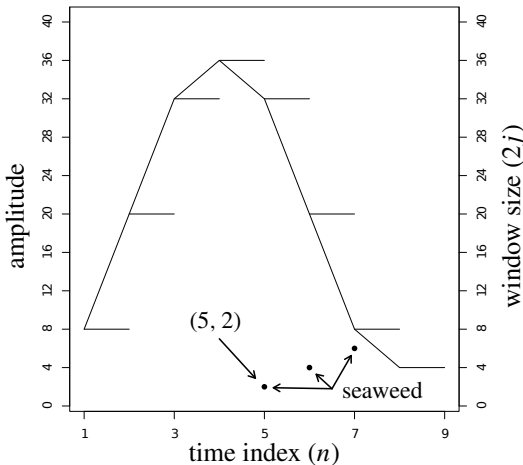
Figure 5. Plot of piecewise-linear representation of the signal used in Examples 2.1 and 3.1 as well as the discrete representation with transition points at all time scales marked.
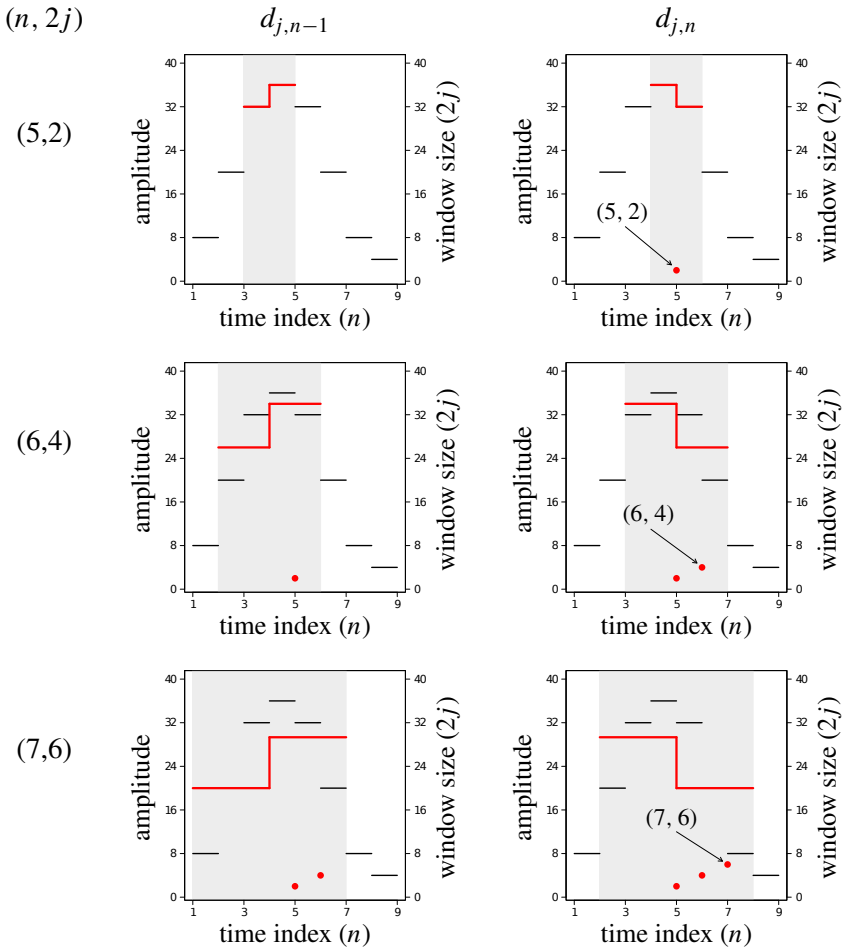
Figure 6. Plots of the signal used in Examples 2.1 and 3.1 depicting the differences in averages on three time scales and the transition points identified as a result of a sign change. The column $d_{j,n-1}$ shows the last positive difference in averages, and the column $d_{j,n}$ shows the first negative difference in averages.

implementation of the seaweed algorithm to the signal in Example 3.1 is included as seaweed_code.R in the online supplement, which can be obtained from this article's publication page.

3D. *Graphical interpretation.* To identify an important transition in the signal with the graphical output, we locate a long branch of seaweed and follow it down until it branches or becomes too hard to distinguish from other, generally shorter, seaweed branches. This happens when the seaweed is being strongly affected by
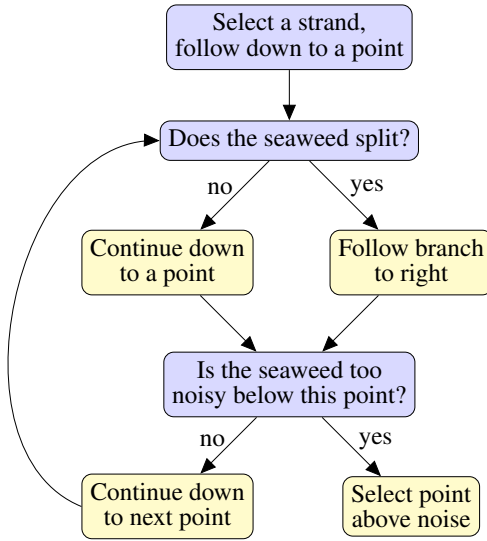
Figure 7. The decision process in the seaweed analysis.

small-scale high-frequency noise, and is no longer being strongly influenced by the global trend. In such a case we see that the seaweed itself becomes too noisy to be reliable at such a small scale. If the branch does split, we follow the general trend, typically on the right side. This is because the algorithm selects the right endpoints of a window, and so selecting the left branch may be departing from a particular strand of seaweed. The time index of the last distinguishable seaweed point on that branch is identified as the time index of a transition point. A decision tree representation of this process is seen in Figure 7.

Next, we examine three additional example signals with the seaweed method. For this we use a simulated base signal, seen in Figure 8, designed to emulate the signals observed in our bean beetle experiments, specifically the region of Figure 1 around frame 300. The equation for the base signal is given by

$$f(n) = \begin{cases} 0 & \text{if } 1 \leq n < \beta - 25, \\ 0.02n + 0.5 - 0.02\beta & \text{if } \beta - 25 \leq n < \beta - 10, \\ n + 10.3 - \beta & \text{if } \beta - 10 \leq n < \beta - 6, \\ 4.3 & \text{if } \beta - 6 \leq n < \beta, \\ -0.5n + 4.3 + 0.5\beta & \text{if } \beta \leq n < \beta + 10, \\ -0.25n + 1.8 + 0.25\beta & \text{if } \beta + 10 \leq n < \beta + 15, \\ -1.95 & \text{if } \beta + 15 \leq n < 70, \end{cases} \tag{3-3}$$

where $\beta + 1$ is the initial large-scale drop in amplitude, which we call the primary transition point, and n is the time index. In Example 3.2 we use the seaweed method to identify the primary transition point in the base signal.
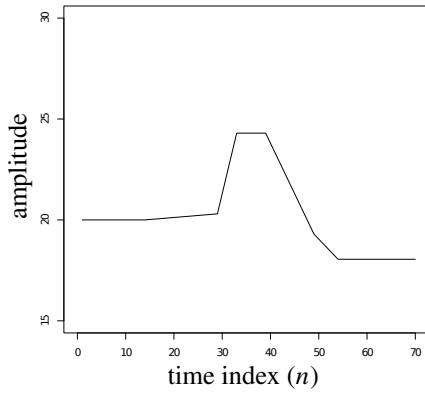
Figure 8. Base signal from (3-3) with primary transition point at $\beta + 1 = 40$.

Example 3.2. Consider the signal displayed in Figure 9. The primary transition point is located by following the long strand of solid-circle seaweed down. In following the right branch at the split, we end at the point $(n, 2j) = (40, 2)$, signifying that the transition occurs at $n = 40$. It is also interesting to note the stacks of triangle seaweed concentrated near the beginning and end of the signal. In each case, this behavior is the result of the constant amplitude of the signal, and the shape of these stacks is due to the changing size of the sliding window. Furthermore, in each case the stack is bordered by a strand of open-circle seaweed, which corresponds to the
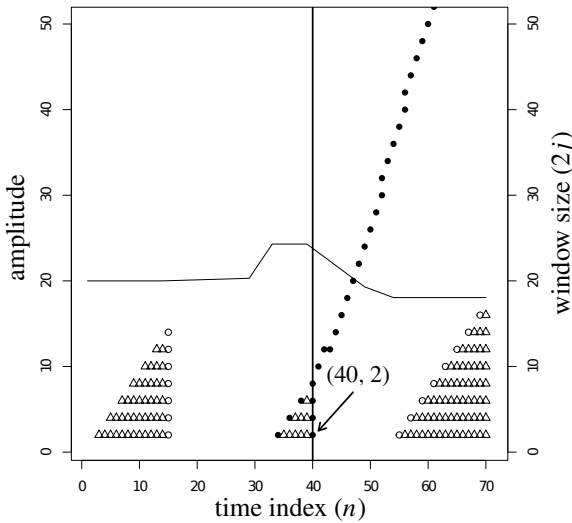
Figure 9. Base signal from (3-3) with primary transition point $\beta + 1 = 40$.
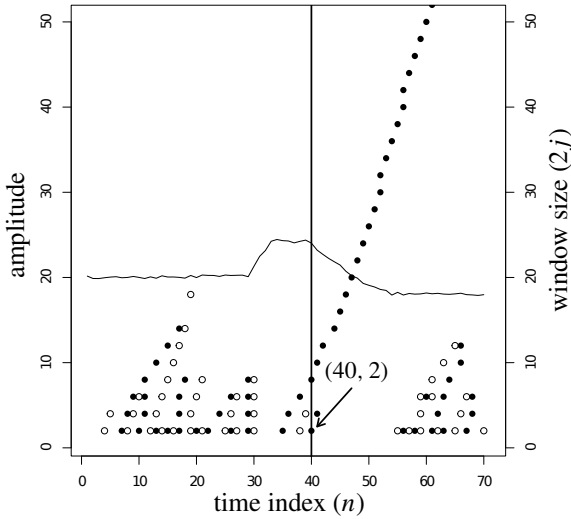
Figure 10. Example of a signal with low noise and limited seaweed branching. The primary transition point for this signal is $\beta + 1 = 40$, and the noise was sampled from a normal distribution with mean 0 and standard deviation 0.1.

first initial transition in the signal from constant to increasing, at $n = 15$, and the last transition in the signal from decreasing to constant, at $n = 55$.

We follow this example with another involving a slightly noisier signal with low levels of seaweed branching.

Example 3.3. Consider the signal displayed in Figure 10, which was obtained by adding noise to the base signal which was randomly sampled from a normal distribution with mean 0 and standard deviation 0.1. This example proceeds nearly identically to Example 3.2. We follow the largest strand of solid-circle seaweed down, following the right branch to the point $(n, 2j) = (40, 2)$. As before, this means that the primary transition point occurs at $n = 40$. Additionally, the window level corresponding to this point, $2j = 2$, indicates that the algorithm was able to recognize the transition point even at the highest possible resolution. Note that in this example, the presence of noise eliminates the constant behavior seen in Figure 9, and therefore we do not observe any open triangles being plotted in the output. Instead, we see branching strands of open- and closed-circle seaweed which appear on either side of localized perturbations due to noise.

Next, we consider an example with more noise and seaweed branching.

Example 3.4. Consider the signal displayed in Figure 11, which was obtained by adding noise to the base signal which was randomly sampled from a normal distribution with mean 0 and standard deviation 1.2. We begin again as in Example 3.2;
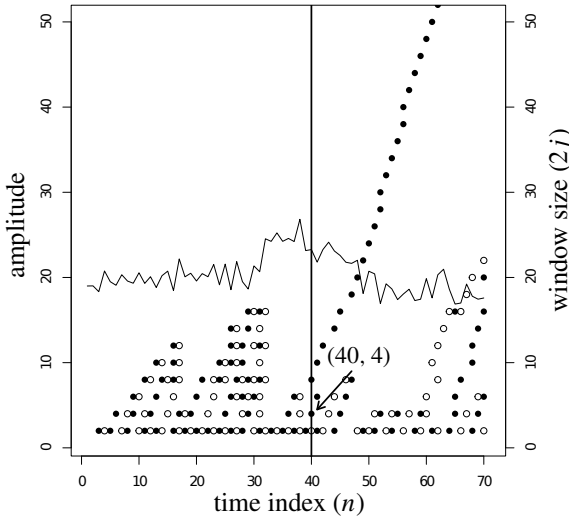
Figure 11. Example of a signal with a higher level of noise and seaweed branching. The primary transition point for this signal is $\beta + 1 = 40$, and the noise was sampled from a normal distribution with mean 0 and standard deviation 1.2.

however, in this case we can only follow the right strand to $(n, 2j) = (40, 4)$, as the seaweed below this point is focusing on high-frequency noise. Therefore, the primary transition occurs at $n = 40$.

3E. *Evaluating the seaweed method.* We designed experiments to test the accuracy and precision of the seaweed method, using the base signal shown in Figure 8 and defined in (3-3). The first experiment explores a broad range of noise levels with a small number of signals at each level. The second experiment explores a smaller range of noise levels with a larger number of signals at each level.

3E.1. *Experiment 1.* Noise was generated by randomly sampling from normal distributions with mean 0 and standard deviations 0, 0.2, 0.4, 0.6, 0.8, 1, 1.2, 1.4, 1.6, 1.8, 2. We injected this noise into base signals by adding it to base signal. The primary transition point, $\beta + 1$, had been randomly selected from the interval $26 \leq n \leq 56$. Examples of this sort of signal can be seen in Examples 3.2, 3.3, and 3.4.

For each of the eleven standard deviations, 30 signals were generated with randomly generated noise and a randomly positioned primary transition point. Each of the 330 signals was stored in a file in random order, and the true primary transition point and noise level for each were stored in a separate file. This second file was kept hidden while we used seaweed plots to identify the primary transition points in the synthetic signals. These primary transition points were then compared to those in the

hidden file. The script seaweed_test1.R used to generate the signals and guarantee the conditions required for a blind trial is included in the online supplement.

3E.2. *Experiment 2.* Noise with standard deviations of 0.1, 0.2, 0.3, 0.4, 0.5 was injected into the base signal. Additionally, in this experiment we generated 200 signals for each standard deviation. With the exception of these two changes, everything else proceeded as in the first experiment. The script seaweed_test2.R for this experiment is also included in the online supplement.

4. Results

4A. *Experiment 1.* The difference between the true primary transition points and the primary transition points identified by the seaweed method were computed for each signal, and the error distributions that resulted for each level of noise were examined. The standard deviations of errors for Experiment 1 are displayed in Table 3.

As can be seen from Table 3, the noisier the signal the greater the error in identifying primary transition points. Furthermore, as can be seen in Figure 12, this trend is approximately linear, with $r^2 = 0.728$.

4B. *Experiment 2.* As in Experiment 1, the primary transition point identification error and standard deviations for the resulting error distributions were calculated for Experiment 2. These standard deviations are displayed in Table 4.

Again, it is apparent that as we add more noise to the signal, we observe more error in the analysis. As can be seen in Figure 12, this trend is also approximately linear, with $r^2 = 0.825$. Histograms of each error distribution in this experiment are displayed in Figure 13. Each has a mode at 0 and exhibits a higher standard deviation of primary transition point identification error as the noise level increases.

4C. *A bean beetle example.* Recall the time signal for the relative brightness of a bean beetle egg shown in Figure 1. This time signal is based on time-lapse

noise s.d.	0	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2.0
error s.d. (time steps)	0	1.106	1.375	1.137	1.574	1.907	2.369	2.128	1.846	2.132	2.273

Table 3. Standard deviation of primary transition point identification error for each of the 11 levels of noise tested in Experiment 1.

noise s.d.	0.1	0.2	0.3	0.4	0.5
error s.d. (time steps)	0.615	0.808	0.845	0.990	0.941

Table 4. Standard deviation of primary transition point identification error for each of the 5 levels of noise tested in Experiment 2.
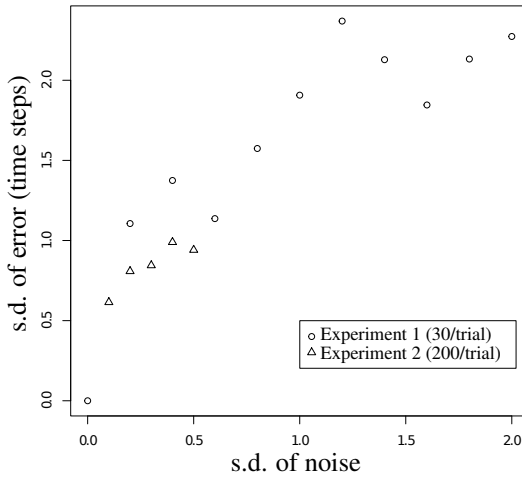
Figure 12. Scatterplot showing standard deviation of primary transition point identification error vs. standard deviation of noise for both experiments, showing their approximately linear relationship with $r^2 = 0.728$ for Experiment 1 and $r^2 = 0.825$ for Experiment 2.
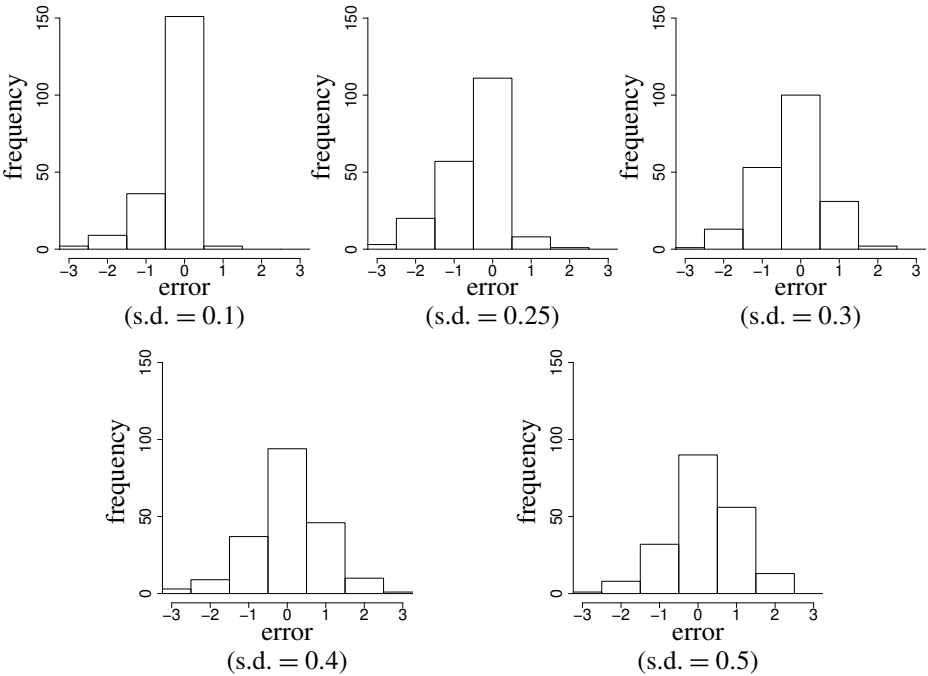
Figure 13. Histograms of error distributions in Experiment 2. All have a mode at 0 and exhibit a higher standard deviation of primary transition point identification error as the noise level increases.
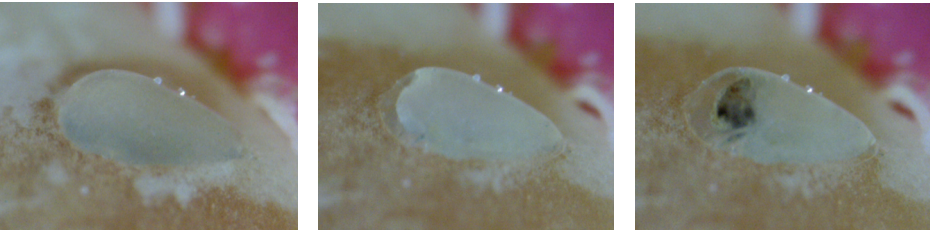
Figure 14. Digital microscope images of an egg of a developing bean beetle embryo. The image on the left was taken shortly after the egg was laid, the image in the middle was taken near the end of embryonic development, and the third image shows the embryo's darkened head capsule.

photographs of bean beetle eggs such as the ones shown in Figure 14. We used the seaweed method to analyze this signal, and determine the frame number associated with the beginning of head-capsule darkening. The signal with seaweed points is shown in Figure 15. In this case we used the seaweed method to select the point $(n, 2j) = (318, 4)$, meaning that the embryo's head capsule began to darken in frame 318. Furthermore, we know from experimental observations that this egg was laid in frame 93, which combined with the fact that the frames are 20 minutes apart, means that the embryo developed in approximately 3.125 days.
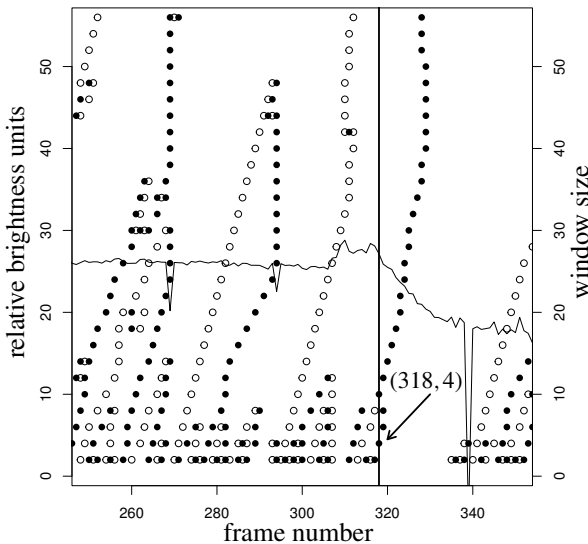
Figure 15. Graph of relative brightness of a particular bean beetle egg over time between frames 250 and 350 with seaweed overlaid. All of the frames are 20 minutes apart.
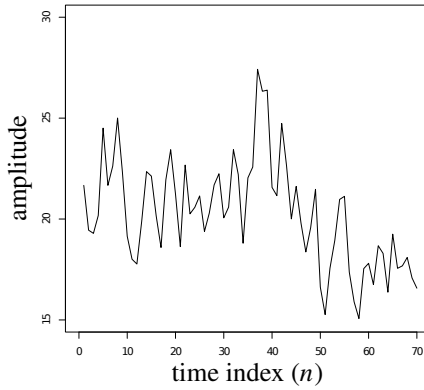
Figure 16. Base signal injected with noise sampled from a normal distribution with mean 0 and standard deviation 2, with primary transition point at $\beta + 1 = 40$.

5. Discussion

5A. *Experiment 1.* Despite the lower resolution with which we examined varying noise levels in Experiment 1, we were still able to glean a general idea of how the seaweed method performs at a wide range of noise. The correlation observed between noise level and primary transition point identification error seen in Table 3 and Figure 12 suggests that the seaweed method is most accurate at low levels of noise. However, even at higher levels of noise, such as noise with a standard deviation of 2, when it is difficult to visually recognize the presence of a signal beneath the noise, our method was still able to detect the primary transition point. An example of such a signal can be seen in Figure 16. For this level of noise, the resulting error distribution had a standard deviation of 2.273, meaning that most of the time the seaweed method is able to accurately identify the primary transition point within three time steps, even at such an extreme level of noise.

For lower levels of noise the standard deviations of the error distributions were closer to 1.5, meaning that at lower levels of noise this experiment suggested that the method is generally capable of accurately identifying the primary transition point within two time steps, or 40 minutes in our bean beetle experiments.

5B. *Experiment 2.* In Experiment 2 we chose to focus more closely on the range of noise with standard deviations around 0.2, the level of noise observed in the time signals from our bean beetle experiments. As can be seen from Table 4 and Figure 12, the approximately linear trend observed in Experiment 1 is also apparent at this scale, lending support to the conclusion that as more noise is introduced to the signal, more error is apparent in detection of primary transition points. Additionally, the increased sample size in this experiment resulted in a better representation of the

standard deviations of primary transition point identification error for this method, suggesting that the standard deviation of this error for low levels of noise is closer to one time step, or 20 minutes in our bean beetle experiments.

These conclusions are further reinforced through the histograms in Figure 13. All histograms exhibit a mode at 0 and standard deviations between 0 and 0.941, meaning that at the levels of noise examined in this experiment, the seaweed method is particularly accurate at identifying primary transition points. As the level of noise increases, the standard deviation of primary transition point identification error increases slightly, although the majority of the values fall between ± 1, meaning that our method is rarely more than one time step away in identifying the primary transition point for those levels of noise. Furthermore, none of the histograms include values more extreme than ± 3, meaning that our method is very rarely more than three time steps away from identifying the primary transition point.

We also observed a slight skew in the error distributions, in the form of a right skew at lower levels of noise and a left skew at higher levels of noise. The right skew may be the result of the asymmetry in the base signal, as seen in Figure 8, or potentially human subjectivity. At low levels of noise there may be a tendency to select the transition point without fully relying on the seaweed. This effect ought to be negligible, however, due to the manner in which the experiment was conducted. The left skew may be due to the fact that the algorithm selects the right endpoint of the window being considered and therefore may occasionally overshoot the transition point. Further experimentation, such as a replication of these experiments with a different base signal, would be needed to identify the source of the skew.

5C. *Conclusions.* The error distributions found in Experiments 1 and 2 have a mode at 0 and low standard deviations of primary transition point identification error, from which we can conclude that the seaweed method is both accurate and precise in identifying transition points. Although we have focused our discussion on transitions from positive to negative averages, our results should generalize to detecting other types of transition points. We have used the seaweed method to detect transitions from negative to positive averages to identify other biological markers in bean beetle development as well.

The conceptual transparency inherent to this method gives it the advantage of being approachable by a wide audience, including those without any substantial background in wavelets or signal processing. Its unique graphical representation of transition points also makes it easy to learn and apply. Although this may seem to imply a level of subjectivity, this can be avoided though careful adherence to the guidelines laid out in Section 3D.

This method seeks to identify transition points whose influence reaches across multiple time scales, localizing transitions of global importance. In addition, the

seaweed method provides a higher resolution of detecting these types of transitions when compared to the discrete wavelet transform.

The efficacy of this method also serves to support our analysis of bean beetle development times, as this method was essential in determining these times from signals like those discussed in Example 3.1. Furthermore, we believe that this method will prove capable of analyzing a much wider range of time signals, such as stock market data and audio signals, and may be used to analyze other types of signals (e.g., spatial) as well. One such spatial signal that could benefit from this type of analysis is population density of a given organism along a transect.

Acknowledgments

References

[Aboufadel and Schlicker 1999] E. Aboufadel and S. Schlicker, *Discovering wavelets*, Wiley, New York, NY, 1999. MR Zbl

[Auscher 1992] P. Auscher, "Wavelet bases for $L^2(\mathbb{R})$ with rational dilation factor", pp. 439–451 in *Wavelets and their applications*, edited by M. B. Ruskai et al., Jones and Bartlett, Boston, MA, 1992. MR Zbl

[Bayram and Selesnick 2009] İ. Bayram and I. W. Selesnick, "Overcomplete discrete wavelet transforms with rational dilation factors", *IEEE Trans. Signal Process.* **57**:1 (2009), 131–145. MR

[Beck and Blumer 2011] C. W. Beck and L. S. Blumer, "A handbook on bean beetles, *Callosobruchus maculatus*", 2011, available at http://www.beanbeetles.org/handbook/handbook.pdf.

[Bénéteau and Van Fleet 2011] C. Bénéteau and P. J. Van Fleet, "Discrete wavelet transformations and undergraduate education", *Notices Amer. Math. Soc.* **58**:5 (2011), 656–666. MR

[Christensen 2001] O. Christensen, "Frames, Riesz bases, and discrete Gabor/wavelet expansions", *Bull. Amer. Math. Soc.* (*N.S.*) **38**:3 (2001), 273–291. MR Zbl

[Daubechies et al. 2003] I. Daubechies, B. Han, A. Ron, and Z. Shen, "Framelets: MRA-based constructions of wavelet frames", *Appl. Comput. Harmon. Anal.* **14**:1 (2003), 1–46. MR Zbl

[Logan and Powell 2001] J. A. Logan and J. A. Powell, "Ghost forests, global warming and the mountain pine beetle", *Amer. Entomol.* **47**:3 (2001), 160–173.

[Logan et al. 1976] J. A. Logan, D. J. Wollkind, S. C. Hoyt, and L. K. Tanigoshi, "An analytic model for description of temperature-dependent phenomena in arthropods", *Environ. Entomol.* **5**:6 (1976), 1133–1140.

[Mulcahy 1996] C. Mulcahy, "Plotting and scheming with wavelets", *Math. Mag.* **69**:5 (1996), 323–343. MR Zbl

[Selesnick 2011] I. W. Selesnick, "Wavelet transform with tunable Q-factor", *IEEE Trans. Signal Process.* **59**:8 (2011), 3560–3575. MR

[Teolis 1998] A. Teolis, *Computational signal processing with wavelets*, Birkhäuser, Boston, MA, 1998. MR Zbl

[Yurk and Powell 2009] B. Yurk and J. A. Powell, "Modeling the evolution of insect phenology", *Bull. Math. Biol.* **71**:4 (2009), 952–979. MR Zbl

david.mcmorris@huskers.unl.edu

Department of Mathematics, University of Nebraska–Lincoln, 203 Avery Hall, PO BOX 880130, Lincoln, NE 68588, United States

pearsonp@hope.edu   *Department of Mathematics, Hope College, 27 Graves Place, Holland, MI 49422, United States*

yurk@hope.edu   *Department of Mathematics, Hope College, 27 Graves Place, Holland, MI 49423, United States*

involve

msp.org/involve

INVOLVE YOUR STUDENTS IN RESEARCH

Involve showcases and encourages high-quality mathematical research involving students from all academic levels. The editorial board consists of mathematical scientists committed to nurturing student participation in research. Bridging the gap between the extremes of purely undergraduate research journals and mainstream research journals, *Involve* provides a venue to mathematicians wishing to encourage the creative involvement of students.

involve