A scale space multiresolution method for extraction of time series features

Leena Pasanen\textsuperscript{a}, Ilkka Launonen\textsuperscript{a} and Lasse Holmström\textsuperscript{a}

Received 8 October 2013; Accepted 19 October 2013

A scale space multiresolution feature extraction method is proposed for time series data. The method detects intervals where time series features differ from their surroundings, and it produces a multiresolution analysis of the series as a sum of scale-dependent components. These components are obtained from differences of smooths. The relevant sequence of smoothing levels is determined using derivatives of smooths with respect to the logarithm of the smoothing parameter. As time series are usually noisy, the method uses Bayesian inference to establish the credibility of the components. © The Authors. Stat published by John Wiley & Sons Ltd.

Keywords: Bayesian methods; simulation; smoothing; time series; visualization

1 Introduction

A typical aim of nonparametric regression is to estimate the signal from noisy data by smoothing, where the smoothness of the estimate is controlled by a smoothing parameter. The fundamental question then is, which level of smoothing uncovers the signal best, revealing such features as local minima and maxima. The aim of scale space analysis differs from this by viewing the signal on several smoothing scales, all of which are considered to contain valuable information about the data, instead of finding a single best fit.

Scale space analysis was made popular in computer vision literature by Lindeberg (1994). It was later introduced to statistics by Chaudhuri & Marron (1999, 2000) in the form of the significant zero crossings of the derivative (SiZer) method, which detects statistically significant increasing or decreasing intervals from nonparametric regression curves or density estimates on several scales. Later, a Bayesian SiZer (BSiZer) was proposed by Erästö & Holmström (2005). A different Bayesian approach was proposed by Godtliebsen & Øigård (2005). For reviews of statistical scale space methods and BSiZer, see Holmström (2010b, 2010a), respectively.

Let $\mu$, $y \in \mathbb{R}^n$ and $S_\lambda \in \mathbb{R}^{n \times n}$ be the underlying signal, the observed noisy series and a smoothing operator corresponding to a smoothing level $\lambda > 0$, respectively, and let $\{\lambda_i\}$ be an increasing sequence of smoothing levels. The SiZer and BSiZer analyses produce a map that consists of pixels $(i, j)$ where $i$ is the scale index and $j$ is the time index. The $i$th row of the map corresponds to the time derivative of the smooth $S_{\lambda_i} \mu$. The color of the pixel $(i, j)$ depends on the sign of the derivative. To determine the pixel color, SiZer applies frequentist statistical methods to the observed $y$, whereas BSiZer uses the posterior distribution $\rho(\mu | y)$.

*Department of Mathematical Sciences, University of Oulu, P.O. Box 3000, 90014 Oulu, Finland
*Email: leena.pasanen@oulu.fi

This is an open access article under the terms of the Creative Commons Attribution Non-Commercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.
Here, we take a different approach to time series feature extraction. Instead of detecting intervals where the derivative is credibly positive or negative, we detect intervals where the values of $\mu$ are higher or lower than its average in a local neighborhood and construct a multiresolution analysis of $\mu$ as a sum of scale-dependent components. This is accomplished by considering differences of smooths. In Holmström et al. (2011), such an approach was used for digital images. In practice, first a sample is drawn from the posterior distribution of $\mu$, and then, posterior samples of the scale-dependent components are obtained by applying a difference of smooths operator $S_{\lambda_i} - S_{\lambda_j}$, where $\lambda_i < \lambda_j$, to this sample. The credibly nonzero intervals of the components are then detected.

The smoothing levels used to compute the differences of smooths must be carefully chosen to correctly extract the components in the signal. For this, a **scale-derivative map** resembling SiZer and BSiZer maps is used (e.g., bottom panel of Figure 1). However, in contrast to SiZer and BSiZer, which consider the time derivative of the smoothed signal, each row $i$ of the scale-derivative map corresponds to the derivative of the smooth $S_{\lambda_i} \mu$ with respect to $\log \lambda$ computed at $\lambda_i$, where $\log$ denotes the natural logarithm. A column $j$ measures how the intensity of the smooth changes as a function of the scale at time $j$. The values $\lambda_i$ used in the final multiresolution analysis are the values that correspond to the rows of this map where the derivative is uniformly close to zero. Because $\mu$ is unknown, we use the posterior mean of $\mu$ to calculate the derivative of the smooth.

The difference of smooths is familiar also in image analysis literature, where it is used as an edge detection method. This is based on the fact that the edges of features are located at the zero crossings of the Laplace operator, which, by the heat equation, can be approximated by the less computationally intensive difference of Gaussian smooths (Marr & Hildreth, 1980). As we are dealing with one-dimensional data, differences of smooths can be used as an approximation of the second time derivative. SiZer and BSiZer detect local maxima and minima as zero crossings of the derivative, whereas here, in the spirit of edge detection, zero crossings of the second derivative correspond to intensity changes.

The article is organized as follows. Section 2.1. presents the idea of a scale space multiresolution analysis. The selection of smoothing levels and the scale-derivative map are discussed in Section 2.2.. Section 2.3. introduces credibility analysis for noisy data. Section 3 presents analyses of real and artificial time series, considers the rate of false-positives by analyzing samples from Gaussian noise and makes comparison with other SiZer-related methods. We conclude the article with a discussion in Section 4. A MATLAB implementation of the method is available online at http://cc.oulu.fi/~lpasanen/MRBSiZer1D/.

## 2 Bayesian multiresolution analysis of time series

### 2.1. A multiresolution decomposition

Consider a signal $\mu = [\mu_1, \ldots, \mu_n]^T$ observed at times $t_1 < t_2 < \cdots < t_n$. We try to decompose $\mu$ into scale-dependent components obtained from differences $S_{\lambda_i} \mu - S_{\lambda_i+1} \mu$ of smooths determined by an appropriate choice of the smoothing parameter sequence $\{\lambda_i\}$.

For smoothing, we use a discrete spline smoother

$$S_\lambda = (I + \lambda C^T C)^{-1},$$

where $C \in \mathbb{R}^{(n-2) \times n}$ is the second-order difference matrix, that is, $C_p = w$ with

$$w_j = \frac{\rho_{j+2} - \rho_{j+1}}{t_{j+2} - t_{j+1}} - \frac{\rho_{j+1} - \rho_{j}}{t_{j+1} - t_{j}}, \quad j = 1, \ldots, n - 2. \quad (2)$$

As $\lambda \to \infty$, $S_\lambda \mu$ converges to the linear regression line.
The smoothing operator $S_\lambda$ can be thought of as a low-pass filter. Hence, the difference of smooths $S_{\lambda_i} - S_{\lambda_j}$, with $\lambda_i < \lambda_j$, can be thought of as a bandpass filter that isolates features that are present at level $\lambda_i$ but not at $\lambda_j$ (Holmström et al., 2011). Let

$$0 = \lambda_1 < \lambda_2 < \cdots < \lambda_{L-1} < \lambda_L \leq \infty$$

be a sequence of smoothing levels and denote the mean of $\mu$ by $\bar{\mu}$. Because $S_{\lambda_1} \mu = \mu$, we can decompose $\mu$ into components as

$$\mu = \sum_{i=1}^{L-1} (S_{\lambda_i} - S_{\lambda_{i+1}}) \mu + S_{\lambda_L} \mu - \bar{\mu} 1 + \bar{\mu} 1 \equiv \sum_{i=1}^{L+1} z_i,$$

(3)

where $z_i = (S_{\lambda_i} - S_{\lambda_{i+1}}) \mu$ for $i = 1, \ldots, L - 1$, $z_L = S_L \mu - \bar{\mu} 1$ and $z_{L+1} = \bar{\mu} 1$. The component $z$ can be interpreted as the detail of $\mu$, which is smoothed out when smoothing is increased from $\lambda_i$ to $\lambda_{i+1}$. If $\lambda_L = \infty$, $z_i$ corresponds to the linear trend. It follows easily from the properties of the smoother (1) that $z_1, \ldots, z_{L-1}$ have no linear trend and $z_i = 0$, except for $z_{L+1}$ (cf. section 2.1 in Erästö & Holmström, 2005).

For exploratory purposes, such analysis can be performed directly on the actual observed noisy series $y$. In this case, the noise may be possible to filter out as the smallest-scale component. Figure 5 shows a scale-derivative map and scale-dependent components for raw data. However, in order to make inferences about the underlying structures of the series, the posterior distribution of $\mu$ is used instead. Given a sample from the distribution $p(\mu | y)$, we make inferences about the credibility of the scale-derivative map and the scale-dependent components. The inference is summarized in the form of maps and displays that use white, black and gray to indicate properties of the derivative and the signal components. The right panel of Figure 8 shows an example of inference for a scale-derivative map, and Figure 9 shows inference for scale-dependent components.

### 2.2. The scale-derivative map and the selection of smoothing levels

To achieve a good separation of the scale-dependent components, the sequence $\{\lambda_i\}$ in (3) needs to be chosen carefully. Intuitively, the idea is as follows. Consider a signal $\mu = \sum_{i=1}^{L-1} \mu_i$ that consists of a sum of components $\mu_i$ of different scales, ordered according to ascending magnitude of the scale, where $\mu_1$ is the smallest-scale component. We assume that $\mu_i$ has no linear trend for $i = 1, \ldots, L - 1$ and $\mu_L = 0$ except for the constant vector $\mu_{L+1}$. We should then find a sequence $\{\lambda_i\}$ for which $z_i = (S_{\lambda_i} - S_{\lambda_{i+1}}) \mu \approx \mu_i$, $i = 1, \ldots, L - 1$. This is possible because the smoother $S_{\lambda}$ acts as a low-pass filter. Therefore, for each $k = 1, \ldots, L$, there is a smoothing level $\lambda_k$ for which

$$S_{\lambda_k} \mu \approx \sum_{i=k}^{L+1} \mu_i,$$

if the scales of the components are sufficiently distinct. Thus, $(S_{\lambda_k} - S_{\lambda_{k+1}}) \mu \approx \mu_k$ for $k = 1, \ldots, L - 1$.

A principled method for choosing suitable smoothing levels $\lambda_i$ can be based on the derivative of the smooth $S_{\lambda} \mu$ with respect to $\log \lambda$. The logarithmic scale for $\lambda$ is preferable because the larger the smoothing levels, the larger the distance between successive values of $\lambda$ has to be to have a noticeable effect on the smooth. The derivative is

$$D_{\lambda} \mu = \lim_{\lambda' \to \lambda} \frac{S_{\lambda'} \mu - S_{\lambda} \mu}{\log \lambda' - \log \lambda} = -\lambda (1 + \lambda C' C)^{-1} C' (1 + \lambda C' C)^{-1} \mu.$$

We pick the values $\lambda_1, \ldots, \lambda_L$ as the local minima of the norm $\|D_{\lambda} \mu\|$ (e.g., Figure 2). This can be justified as follows. Consider a signal $\mu$ that consists of only one component with no linear trend and zero mean. Because of the low-pass nature of the smoother $S_{\lambda}$, the signal remains nearly unaltered for a sufficiently small $\lambda'$, that is, $S_{\lambda} \mu \approx \mu$ for $\lambda \leq \lambda'$. 

Stat 2013; 2: 273–291
For a sufficiently large smoothing level $\lambda''$, the signal will be close to zero for $\lambda \geq \lambda''$. Then, $\|D_\lambda \mu\|$ is close to zero for both $\lambda \leq \lambda'$ and $\lambda \geq \lambda''$. If $\mu$ is a sum of two zero-mean components $\mu_1$ and $\mu_2$ with sufficiently distinct scales, where $\mu_1$ has no linear trend, there is a smoothing level $\lambda'$ for which the smaller-scale component is mostly smoothed out but which doesn't affect the larger-scale component. Therefore, $\lambda'$ is a local minimum of $\|D_\lambda \mu\|$ and $\mu - S_{\lambda'} \mu \approx \mu_1$, $S_{\lambda'} \mu \approx \mu_2$. Similarly, assuming sufficient separation of successive scales, $\mu = \sum_{i=1}^{L-1} \mu_i$ can be decomposed using the $L - 1$ local minima of $\|D_\lambda \mu\|$.

The derivative $D_\lambda \mu$ itself also provides detailed information about the features of $\mu$ at different time points. The information is visualized using the scale-derivative map, where the color of the pixel $(i,j)$ corresponds to the $j$th element of the vector $D_\lambda \mu$. A suitable sequence of smoothing parameters can be searched also visually from the scale-derivative map by looking for rows where the derivative is uniformly close to zero between strongly colored areas. This may be necessary, for example, when a component has a variable scale with respect to time (cf. Section 3.3.). It is then broken into several subcomponents with a fixed scale over some time intervals while being nearly constant elsewhere. Close inspection of the scale-derivative map is also recommended when the data are highly nonequispaced.

We illustrate these ideas with the signal colored in black in the top panel of Figure 1. The signal $\mu$ consists of a sum of two sine wave series, $\mu_1$ and $\mu_2$, where the frequency of $\mu_1$ is higher than the frequency of $\mu_2$. Because of the choice of the time grid and phase, respectively, the series have zero mean and no linear trend. The signal should be decomposed into two components $z_1 = \mu - S_{\lambda'} \mu \approx \mu_1$ and $z_2 = S_{\lambda'} \mu \approx \mu_2$. We thus need to find a value of $\lambda$ for which the smooth $S_{\lambda} \mu$ is as close to the low-frequency sine $\mu_2$ as possible. The top panel of Figure 1 shows several smooths of $\mu$ where those coarser and smoother than $S_{\lambda'} \mu$ are colored blue and red, respectively. We have calculated the smooths and the derivative in a geometrically increasing grid $\lambda_i C 1_{\lambda_i \delta \lambda_i}^{10}$, starting from $10^{-1.5}$ and stopping at $2.6 \cdot 10^5$. The middle panel of Figure 1 displays all the smooths in a logarithmic scale where the $i$th row corresponds to

![Figure 1](image-url).

Figure 1. Analysis of a signal consisting of a sum of two sine waves. Top panel: the signal $\mu$ (black curve) and the smooths $S_{\lambda} \mu$ (red and blue curves). Middle panel: the rows correspond to smooths $S_{\lambda} \mu$. Bottom panel: corresponding map of the derivatives $D_{\lambda} \mu$, the scale-derivative map. Black lines indicate the minimum of $\|D_{\lambda} \mu\|$ with respect to $\log_{10} \lambda$ (cf. Figure 2).
the smooth $S_{\lambda, \mu}$ and the color of pixel $(i, j)$ indicates the value of $(S_{\lambda, \mu})_{ij}$. The bottom panel of Figure 1 shows the map of scale derivatives $D_{\lambda, \mu}$. The local minimum of $\|D_{\lambda, \mu}\|$ has been marked with a black line. The high-frequency and low-frequency sines are clearly separated into two oscillating bands of positive and negative values of the derivative.

In time points where the scale derivative is positive, the value of $S_{\lambda, \mu}$ increases with an increasing smoothing level, which means that $\mu$ is smaller there than its average in a local neighborhood, indicating a possible local minimum. Likewise, a negative derivative indicates a possible local maximum of $\mu$. The scale-derivative map therefore provides information about where the signal differs from its surroundings. Note however that for a signal with a single high peak, the derivative $D_{\lambda, \mu}$ is negative at the location of the peak and positive in its surrounding neighborhood. Therefore, positive and negative derivatives do not always correspond to local minima and maxima of the signal.

Figure 2 shows a plot of $\|D_{\lambda, \mu}\|$. The local minimum at $10^{1.78}$ is indicated by a black diamond and highlighted by a black line in the middle and bottom panels of Figure 1. This smoothing level separates the blue smooths from the red in the top panel. The extracted components $z_i$ and the original components $\mu_i$ are compared in Figure 3 and can be seen to be very close to one another, apart from the edges. The edge effects are caused by small local linear trends in the signal.

The scale-derivative map can also be interpreted through the heat equation. Consider a continuous signal $u(t)$ and its smooth $L(t, \lambda)$, obtained by Gaussian convolution. Then,

$$\frac{\partial L(t, \lambda)}{\partial \lambda} \propto \frac{\partial^2 L(t, \lambda)}{\partial t^2}.$$

Hence, the derivative with respect to scale corresponds to the second derivative with respect to time. Therefore, a positive derivative with respect to scale means local convexity of $u(t)$, and a negative derivative with respect to scale means local concavity. The zero crossings of the second derivative correspond to places where the rate of change of

Figure 2. $\|D_{\lambda, \mu}\|$ as a function of $\log_{10} \lambda$ for the example signal. The local minimum is indicated by a black diamond.

Figure 3. The original components of the example signal (cf. Figure 1) are colored black. The extracted components are in blue and red.
the signal achieves its local maximum, indicating an “edge” in the signal. Therefore, the zero crossings in the scale-derivative map indicate the edges of the signal, whereas in BSiZer and SiZer, the zero crossings correspond to local extrema. The spline smoother we use can be shown to approximate the Gaussian convolution, which makes this reasoning plausible. Note also that differentiating with respect to log λ only adds the positive multiplicative factor λ to the heat equation and therefore does not affect the preceding argument.

2.3. Bayesian credibility analysis of the components

Usually, we observe only a degraded noisy version \( y \) of the signal \( \mu \). We then assume the model

\[
y = \mu + \epsilon,
\]

where \( \epsilon \) is the Gaussian independently and identically distributed (i.i.d.) noise. The scale-derivative map and the decomposition (3) could be computed directly from \( y \), with the smallest-scale component \( z_1 \) representing the noise \( \epsilon \). However, as the noise cannot be completely separated from the other components \( z_i \), one should analyze which of the features in each \( z_i \) are real and not just artifacts of the noise.

Our approach is to analyze posterior distributions of the components \( z_i = (S_{\lambda_i} - S_{\lambda_{i+1}})\mu \). This is carried out by drawing a large sample from the posterior \( p(\mu | y) \) and then applying the transformation \( S_{\lambda_i} - S_{\lambda_{i+1}} \) to each sampled series. The sequence \( \{\lambda_i\} \) for multiresolution analysis is chosen from the local minima of \( \|\bar{E}(D_i\mu | y)\| \), where \( \bar{E}(D_i\mu | y) \) is the posterior mean of the derivative \( D_i\mu \). To analyze the scale-dependent multiresolution components \( z_i \), the posterior \( p(z_i | y) \) is then used to find the time points where \( z_i \) differs credibly from zero. These places could be detected simply by identifying for each \( z_i \) the time points where the marginal posterior probability for the event \( z_i \neq 0 \) exceeds some threshold value \( \alpha \). However, such pointwise inference is sensitive to false-positives. We therefore use simultaneous inference over all time points in \( z_i \) by applying the method of highest pointwise probabilities (HPW), first described by Erästö & Holmström (2005). The HPW is a greedy algorithm that selects time points in descending order with respect to their associated marginal posterior probabilities for as long as their joint probability is at least \( \alpha \). This analysis will be presented as bars of white, black and gray in the background of each plotted component \( z_i \), where the color tells whether \( z_i \) is credibly positive, negative or neither at the corresponding time points (e.g., Figure 9).

We also summarize the credibility of the scale-derivative map by constructing a map with each pixel colored either white, black or gray depending on whether the derivative \( D_i\mu | y \) is credibly positive, negative or neither (e.g., right panel of Figure 8). The credibility is determined with the HPW method simultaneously within each row.

3 Experiments

3.1. An artificial example

We first illustrate the proposed methods by an artificial example. Let \( \mu_{1j} = \sin(t_j), \mu_{2j} = \sin(3t_j) \) and \( \mu_{3j} = \sin(6t_j) \), where \( j = 0, \ldots, 1000 \), and define \( \mu_1 = \mu_1 - S_{\infty} \mu_1, \mu_2 = \mu_2 - S_{\infty} \mu_2 \) and \( \mu_3 = \mu_3 - S_{\infty} \mu_3 \) shown in the left panel of Figure 4. Any linear trend has been removed from the components \( \mu_1 \) and \( \mu_2 \), and all the components have zero mean. The preceding decomposition of \( \mu \) adheres thus strictly to the requirements of equation (3).

The “observed” noisy series \( y \), presented in the right panel of Figure 4, is obtained by adding standard Gaussian noise \( \epsilon \) to \( \mu, y = \mu + \epsilon = \mu_1 + \mu_2 + \mu_3 + \epsilon \). The scale-derivative map of the raw series \( y \) is presented in the left panel of Figure 5. This map shows the three sines as three separate alternating bands of blue and red, and the noise is seen as the high-frequency band below the highest-frequency sine. The black lines represent the local minima of \( \|D_3y\| \), whose graph is plotted in the left panel of Figure 6. These minima are then used to extract the components \( z_1, \ldots, z_4 \).
Multiresolution analysis of time series

Figure 4. The signal $\mu$ that consists of a sum of three sine waves (left panel) and the noisy series $y$ (right panel).

Figure 5. The left panel shows the scale-derivative map of $y$ presented in Figure 4. The black lines indicate the local minima of $\|D_1 y\|$ (cf. Figure 6). The right panel shows the components $z_1, \ldots, z_4$ obtained by using differences of smooths on $y$ with the smoothing levels marked by the black lines.

Figure 6. The norm $\|D_1 y\|$ (left panel) and $\|E(D_2 \mu | y)\|$ (right panel) as a function of $\log_{10} \lambda$ for the artificial data. The local minima are indicated by black diamonds.

shown in the right panel of Figure 5. When visualizing the components, we leave out here and in the rest of the examples the component $z_{L+1} = \mu_1$ (cf. (3)). The original components $\mu_i$ are shown in Figure 7 by the green lines, and they are compared with the extracted components shown by the black lines (almost the same as the blue line in the middle and bottom panels). The decomposition obtained directly from the noisy $y$ can be seen to be reasonably good. However, the left edges of the extracted components are distorted, which is due to the phase of the sines causing a local linear trend.
Figure 7. Two decompositions of the artificial signal $\mu$ are compared with its true components. Top panel: the original component $\mu_1$ (green) and its estimates $z_2$ obtained from $y$ (black) and from the posterior mean $E(\mu|y)$ (blue). Middle and bottom panels: the components $\mu_2$ and $\mu_3$ and their estimates as in the top panel.

Because the observed series $y$ contains noise, the extracted components $z_i, i > 1$, are also somewhat corrupted. Instead of finding the components $z_i$ from $y$, we therefore next perform the decomposition (3) on the posterior mean of $\mu$. Thus, consider the model

$$
y|\mu, \sigma^2 \sim N(\mu, \sigma^2 I),$

$$\sigma^2 \sim \text{Inv-}\chi^2(v_0, \sigma^2_0),$$

$$p(\mu|\kappa, \sigma^2) \propto \left(\frac{\kappa}{\sigma^2}\right)^{\frac{n+2}{2}} \exp\left(-\frac{\kappa}{2\sigma^2} ||C\mu||^2\right),$$

where $C$ is defined in (2). We thus assume that the noise $\epsilon$ is i.i.d. zero-mean Gaussian with a random variance $\sigma^2$.

The improper prior of $\mu$ has a smoothing effect, controlled by the prior smoothing parameter $\kappa$. This model has a multivariate-$t$ posterior distribution for $\mu$,

$$\mu|y \sim t_{v_0+n-2}(S_x y, \Sigma),$$

where

$$S_x y = (I + \kappa C^T C)^{-1} y,$$

$$\Sigma = \left(\frac{||y||^2 - y^T S_x y + v_0 \sigma^2_0}{v_0 + n - 2}\right) S_x.$$

Further information and extensions of this model can be found in the work of Erästö & Holmström (2005).

Care is needed in choosing the prior smoothing parameter $\kappa$ and the hyperparameters $\sigma^2_0$ and $v_0$. As in the work of Erästö & Holmström (2005), we could use a full Bayesian inference and also set $\kappa$ randomly. However, for fast
computations, we have used an empirical Bayes approach and estimated $\kappa$ and the prior mean of $\sigma^2$ from the data. The first local minimum in the left panel of Figure 6 at $\lambda = 10^{-1.42}$ corresponds to the value of $\lambda$, which smooths out most of the noise but leaves the small-scale features unsmoothed. Although this value could be used as $\kappa$, the corresponding posterior mean $S_\kappa y$ seems too rough in this case. We therefore used a maximum likelihood (ML) method to estimate both $\kappa$ and the prior mean of $\sigma^2$. This algorithm is proposed in Appendix A of Holmström & Pasanen (2012) for digital images and is easily adapted for time series. The values used for the parameters $\kappa$, $\sigma^2$ and $\nu_0$ were 0.15, $(8/10) \cdot 0.97^2$ and 10, respectively, which means that the prior mean of $\sigma^2$ is 0.97². The ML estimate. The small value of $\nu_0$ makes the prior of $\sigma^2$ vague. Six thousand realizations were sampled from the posterior distribution of $\mu$.

$\|E(D_\lambda \mu | y)\|$ is shown in the right panel of Figure 6 as a function of $\log_{10} \lambda$. The local minima, shown by the black diamonds, are located at $10^{0.46}$ and $10^{2.14}$. These are indicated by the two uppermost black lines in the scale-derivative map of the posterior mean, shown in the left panel of Figure 8. The map shown in the right panel of Figure 8 presents the credibility analysis of the scale-derivative map where the credibility threshold $\alpha = 0.95$. The features of the scale-derivative map are similar to those of the map in Figure 5, apart from the lower part containing the noise, which has been mostly smoothed out in the posterior mean. Besides the two local minima, we added one further smoothing level at $10^{-1.04}$, shown by the lowermost line in the scale-derivative map and its credibility analysis. Because $S_{10^{-1.04}} \mu \approx \mu$, this extra level does not smooth out any real features. However, it alleviates the detection of the true component $\mu_1$ in the credibility analysis, because the marginal posterior variance of $(S_{10^{-1.04}} - S_{10^{0.46}}) \mu$ is smaller than that of $(S_0 - S_{10^{0.46}}) \mu$. We chose the level $10^{-1.04}$ by eye at around the level where the gray area ends in the credibility map. The sequence of smoothing levels for multiresolution analysis is thus $\{0, 10^{-1.04}, 10^{0.46}, 10^{2.14}\}$, and the scale-dependent component $z_{i+1}$ corresponds to the true component $\mu_i$, $i = 1, 2, 3$.

The four components extracted with these smoothing levels are shown by the blue lines in the four lowermost panels of Figure 9. The uppermost panel compares the posterior mean with the signal $\mu$, colored blue and green, respectively. Judging from the component plots in the other panels, the first component $z_1$ differs only in a very short interval credibly from zero, whereas the peaks and valleys of the other components, representing the true components $\mu_1$, $\mu_2$ and $\mu_3$, are mostly credible.

**Figure 8.** The left panel shows the scale-derivative map of the posterior mean for the artificial data example. The two uppermost black lines indicate the local minima of $\|E(D_\lambda \mu | y)\|$ (cf. Figure 6). The bottom line is chosen by visual inspection. The right panel shows the credibility analysis of the derivative in the left panel with the corresponding lines in yellow. For each smoothing level $\lambda$, points where the derivative of the smooth with respect to $\log \lambda$ is credibly positive, negative or neither are colored white, black and gray, respectively.

Stat 2013; 2: 273–291
The true and extracted components are compared in Figure 7 where the components \( z_2, z_3 \) and \( z_4 \) are colored blue. The estimate of \( \mu \), obtained from the posterior mean of \( \mu \), is smoother than the one obtained from \( y \) because of additional smoothing. The estimates \( z_3 \) and \( z_4 \) are almost equal to the estimates obtained from \( y \). Overall, the true features are extracted reasonably well.

### 3.2. Mean sea level change in the Pacific

We analyze a real time series that considers sea level change in the Pacific ocean (CU Sea Level Research Group; cf. Nerem et al., 2010). This series, shown in Figure 10, covers the time span from the end of the year 1992 to the end of 2012, with the sea level change expressed in millimeters with respect to a reference level. The series contains 732 values, averaging three or four measurements per month. These data were analyzed also in the work of Holström & Launonen (2013), where Posterior Singular Spectrum Analysis was applied to detect credible oscillations.

The graph of \( \| D_x y \| \) and its four local minima is shown in the left panel of Figure 11. The scale-derivative map of the series \( y \) is presented in the upper panel of Figure 12, where the local minima of \( \| D_x y \| \) are indicated by the black lines. The noise seems to lie under the lowermost line.

Using model (4), we sampled 6000 realizations from the posterior distribution of the true signal \( \mu \) to find its credible features. We estimated the prior smoothing parameter \( \kappa \) as the first local minimum of \( \| D_x y \| \) and the prior mean of \( \sigma^2 \) as the variance of the residual series \( y - S_x y \). The hyperparameter values are then \( \nu_0 = 10, \sigma_0^2 = (8/10) \cdot 5.9^2 \) and \( \kappa = 10^{-1.27} \). We also tried the ML estimates \( \kappa = 0.14 \) and \( \sigma^2 = 6.4^2 \) (in place of \( 5.9^2 \)), but the posterior mean of \( \mu \)
Figure 10. The mean sea level change in the Pacific from the end of 1992 to the end of 2012, estimated from satellite altimeter readings. The data were obtained from the website of the CU Sea Level Research Group (2013).

Figure 11. The norm $\|D_\lambda y\|$ (left panel) and $\|E(D_\lambda \mu | y)\|$ (right panel) as a function of $\log_{10} \lambda$ for the sea level data. The local minima are indicated by black diamonds.

seemed oversmoothed. The red line in the upper panel of Figure 12 corresponds to the ML estimate of $\kappa$. It seems to overlap some cyclic features, confirming the oversmoothing.

The graph of $\|E(D_\lambda \mu | y)\|$ and its three local minima is shown in the right panel of Figure 11. The values of the local minima are $10^{1.45}$, $10^{3.27}$ and $10^{5.05}$. The scale-derivative map of the posterior mean is presented in the lower left panel of Figure 12, and its credibility analysis is shown in the lower right panel. As in the previous example, we added an extra smoothing level at $10^{-1.70}$ to detect the small-scale oscillatory features better. We also added another level at $\infty$ because the data contain a rising, approximately linear, trend. The smoothing levels for multiresolution analysis are then $\{0, 10^{-1.7}, 10^{1.45}, 10^{3.27}, 10^{5.05}, \infty\}$.

The components extracted from the posterior mean are presented in the bottom six panels of Figure 13, and the posterior mean is shown in the top panel. The scale-derivative map and the second component show a small-scale oscillatory feature clearly. This is expected, as the data are known to contain a seasonal cycle resulting from the heat expansion of the ocean during warmer months and contraction during cooler months. The third component contains six credibly positive intervals. A comparison with the posterior mean reveals that these six peaks correspond to high peaks in the posterior mean. The leftmost peak is detected because, on average, the values to the left of it are much lower. The black areas correspond to valleys between the high peaks. This component may correspond to the 4- to 6-year Pacific oscillation reported by Unal & Ghil (1995), who analyzed the data of multiple tide-gauge observation stations with multichannel Singular Spectrum Analysis and attributed this period to El Niño Southern Oscillation (ENSO). The fourth and fifth components capture deviation from the linear trend, and considering how similar they are.
The ISI's Journal for the Rapid Dissemination of Statistics Research

Figure 12. The scale-derivative map of \( y \) (upper panel), the scale-derivative map of the posterior mean \( \mathbb{E}(\mu | y) \) (lower left panel) and its credibility analysis (lower right panel) for the Pacific sea level data. See Figure 8 and the text for more information.

and how ambiguous the largest local minimum in Figure 11 is, one might also omit the smoothing level \( 10^{5.05} \) from the multiresolution decomposition altogether. The posterior mean has a slight S shape that causes the components to be negative in the beginning and in the end and positive in the middle. The last component detects the strong, rising linear trend.

3.3. Irregular data

The final example considers a case where the data contain no regular cyclic features. Two sets of Gaussian i.i.d. random vectors \( \mu'_1 \) and \( \mu'_2 \) of length 100 were first generated on an equidistant time grid \( 0.01, \ldots, 10 \). The vectors were smoothed with (1) using smoothing parameter values of 0.002 and 7, respectively, and interpolated to vectors of length 1000 on an equidistant grid from 0.01 to 10. The components \( \mu'_1 \) and \( \mu'_2 \) were multiplied by 10 and 50, respectively, \( \mu'_1 \) was detrended to obtain \( \mu_1 = \mu'_1 - S_\infty \mu'_1 \) and \( \mu'_2 \) had its mean removed, \( \mu_2 = \mu'_2 - \bar{\mu}'_2 \mathbf{1} \). The components \( \mu_1 \) and \( \mu_2 \) were then summed up to obtain the signal \( \mu \). Gaussian i.i.d. noise with a standard deviation of 3 was finally added to \( \mu \) to produce the “observed” signal

\[
y = \mu_1 + \mu_2 + \epsilon = \mu + \epsilon.
\]
Figure 13. Top panel: Posterior mean of the Pacific sea level data. Panels 2–7: The components $z_i$ of the posterior mean are colored blue. The zero level is highlighted by yellow. The background for the time intervals where $z_i$ is credibly positive, negative or neither is white, black and gray, respectively.

Figure 14. Left panel: The components $\mu_1$ and $\mu_2$ (red lines) of the artificial irregular data example, their sum $\mu$ (black line) and the “observed” data $y$ (blue dots). Right panel: The scale-derivative map of $y$. The black lines indicate the local minima of $\|D_\lambda y\|$. 
The components $\mu_1$ and $\mu_2$, their sum $\mu$, and the noisy series $y$ are shown in the left panel of Figure 14. The scale-derivative map of $y$ is displayed in the right panel. The local minima of $\|D_y\|$ are again indicated by the black lines. The lowermost segment of the scale-derivative map seems to capture the noise, and the second segment seems to capture the small-scale feature $\mu_1$. The large-scale component $\mu_2$ is divided further into two components where the smaller-scale component captures the features that are present in the right part of $\mu_2$ and the larger-scale component captures its overall U shape.

Figure 15. Top panel: Posterior mean of the irregular series (blue) and true signal $\mu$ (green). Panels 2–5: The components $z_i$ of the posterior mean are colored blue. The zero level is highlighted by yellow. The background for the time intervals where $z_i$ is credibly positive, negative or neither is white, black and gray, respectively.

Figure 16. Comparison of the true and estimated components for the irregular data example. The true components $\mu_1$ and $\mu_2$ are colored black. The estimated components $z_2$ and $z_3$ are colored blue and red, respectively.
We then used model (4) for credibility analysis with a sample size of 6000. As $\kappa$, we used $10^{-2.66}$, the smallest local minimum of $\|D_y, y\|$. The prior mean of $\sigma^2$ was estimated as the variance of the residual series $y - S_y, y$. The values of the hyperparameters $\kappa$, $\sigma_0^2$ and $\nu_0$ are then $10^{-2.66}$, $(8/10) \cdot 2.67^2$ and 10, respectively. The values of $\lambda$ for multiresolution analysis were obtained from the local minima of $\|E(D, \mu, y)\|$ with a small value $10^{-2.11}$ added to detect the small-scale features better, producing the sequence $\{0, 10^{-2.11}, 10^{1.21}, 10^{3.92}\}$. The four extracted components are presented in Figure 15. The second component seems to capture $\mu_1$ and $\mu_2$, having a variable scale along the time axis, is represented by the sum of the third and fourth components. We therefore take $Z_3 = z_3 + z_4$ and compare $Z_2$ with $\mu_1$ and $Z_3$ with $\mu_2$. This comparison is shown in Figure 16. The original signal components $\mu_1$ and $\mu_2$ are reasonably well extracted, although the component $Z_3$ is somewhat rougher than the original component $\mu_2$.

We also tried ML estimates for $\kappa$ and the prior mean of $\sigma^2$. The posterior mean was then smoother, but the extracted components were very similar. The choice of $\kappa$ and the prior mean of $\sigma^2$ does not seem very crucial for the extraction of the components.

### 3.4. False-positives

We compared the sensitivity of our method to false-positives with that of BSiZer by analyzing data that consist of standard Gaussian random vectors of length 100. The test was repeated 100 times, sampling 4000 vectors for each repetition from model (4) with $\nu_0 = 10$ and $\sigma_0^2 = (8/10)\sigma_{ML}^2$ and $\kappa = \lambda_{ML}$, where $\lambda_{ML}$ and $\sigma_{ML}^2$ are ML estimates. As a result, 74 of the credibility maps of the derivative were completely gray, and 59 of the BSiZer maps were completely gray. The means of the estimated values of $\lambda_{ML}$ and $\sigma_{ML}^2$ were $1.05 \cdot 10^4$ and $0.98^2$, respectively. The maps with false alarms typically had nongray areas only for large values of $\lambda$ but were otherwise blank.

### 3.5. Comparison with other SiZer-related techniques

We finally explore the previous examples with SiZer and BSiZer, comparing the results with the scale-derivative map credibility analysis. The left and right panels of Figure 17 present SiZer maps of the artificial and Pacific sea level series, respectively (cf. Sections 3.1. and 3.2.). The maps were created with the R software package SiZer (R Core Team, 2012; Sonderegger, 2012). Inference in the maps is simultaneous within rows, based on a significance level.
of 0.95, and they are constructed using the extreme value theory as presented by Hannig & Marron (2006). The SiZer maps are interpreted as follows. The pixels where the derivative with respect to time is significantly positive or negative are colored blue and red, respectively. The remaining pixels are colored purple, except for those colored gray where the data are too sparse for meaningful statistical inference.

In the left panel of Figure 17, the artificial series component $\mu_3$ is clearly found around $\lambda = 10^6$. The component $\mu_2$ is detected at around $\lambda = 10^{-0.5}$. However, the component $\mu_1$ does not seem to be detected. Overall, the separation of the original components is clearer in the scale-derivative map (Figure 8, right panel). In the right panel, SiZer does not

![Figure 18. BSiZer maps of the artificial series (left panel) and the Pacific sea level series (right panel). For each smoothing level $\lambda$, points where the discrete time derivative of the smooth is credibly positive, negative or neither are colored white, black and gray, respectively.](image)

![Figure 19. Modified BSiZer maps obtained using second-order discrete time derivatives for the artificial series (left panel) and the sea level series (right panel). For each smoothing level $\lambda$, points where the second-order discrete time derivative of the smooth is credibly positive, negative or neither are colored white, black and gray, respectively.](image)
seem to detect all the small-scale features. At higher scales, the high peaks and the S shape are also hard to detect. The separation of the cyclic components is also clearer in the scale-derivative map (Figure 12, lower right panel). BSiZer (Erästö & Holmström, 2005) was tried with the same posterior model and credibility used in Sections 3.1. and 3.2. The resulting maps shown in Figure 18 are similar to the SiZer maps, except that a few more small-scale features are detected.

We argued in Section 2.2. that the first derivative of a smooth with respect to $\lambda$ is approximately proportional to its second derivative with respect to time. We therefore also made comparisons with a modified BSiZer, where the first-order discrete time derivative is replaced with the second-order derivative. We used the same posterior model as before and drew the maps for the artificial and sea level series with the HPW method using a credibility threshold of 0.95. The results, presented in Figure 19, are rather similar to the scale-derivative map credibility analyses of Figures 8 and 12. However, in the left panel of Figure 19, the small-scale feature $\mu_1$ of the artificial series is not completely detected. In the right panel, the small-scale features of the sea level series do seem to be detected as well as in the scale-derivative map. The modified BSiZer maps and the credibility analyses of the scale-derivative maps differ in the upper parts where the smoothing level is large with respect to the length of the series, as boundary conditions start to influence the results. We also drew SiZer maps based on the second derivatives. They were relatively similar to the modified BSiZer maps, although fewer small-scale features were detected.

### 4 Discussion

We proposed a method for extracting time series features in different scales. The method constructs a scale-derivative map of the smooths of the series together with its credibility analysis. The maps help detect features of the data in different scales in the same spirit as SiZer-related methods. However, whereas SiZer considers time derivatives, we differentiate along the logarithmic scale axis. As a result, instead of finding where the series increases or decreases, we detect time points where the series is larger or smaller than its average in a local neighborhood. The local minima of the norm of the scale derivative are then selected as the sequence of smoothing levels that extract the scale-dependent components of the time series, producing a multiresolution analysis.

The scale-derivative map and the scale-dependent components can be computed directly from noisy observations, which can already provide valuable information about the data. However, to separate true features from artifacts of noise, Bayesian inference is used. We assumed an i.i.d. Gaussian noise, but the method is easily extended to different noise models or even to a case where the time grid is random (Erästö & Holmström, 2007), as the posterior model is independent of the rest of the method.

If the smooths were differentiated with respect to $\lambda$ instead of $\log \lambda$, the magnitude of the derivative would diminish, with larger scales making the upper parts of the maps nearly featureless. For the same reason, the maps of the second time derivatives would not be as meaningful if viewed directly. Hence, although the scale-derivative map credibility analyses and the modified BSiZer maps look similar, the advantage of differentiating with respect to $\log \lambda$ is that, instead of just indicating credibly positive or negative derivatives, the magnitude and sign of the derivative is more clearly displayed. For the same reason, the scale-dependent components cannot be extracted using the modified BSiZer.

Spline smoothing was employed, but the method could use other smoothers, too. The fact that the heat equation relates to Gaussian convolution motivated us to try local linear regression used in SiZer, as it resembles Gaussian convolution more than spline smoothing. We approximated the scale-derivative map with the difference quotient
\( -(S_\lambda - S_{\lambda^*}) \mu / \log \gamma \), whereas posterior sampling and credibility analysis were performed as before. The high-frequency sine of the artificial series was not detected completely. The results of the irregular series example were quite similar for local linear regression and smoothing splines, whereas in the sea level example, a few more small-scale spikes were detected with local linear regression. In the work of Marron & Zhang (2005), spline and local linear regression SiZers were compared, and according to the authors, the “smoothing spline method is effective in smoothing low noise but sparse data, while the local linear method performs better for high noise but dense data,” which is in line with our findings.

We demonstrated our method with two artificial examples and one real time series involving sea level change in the Pacific. The scale-derivative map separated the oscillatory features, and the scale-dependent components thus extracted were mostly credible. The last example was composed of an irregular artificial series that did not contain clear oscillatory features, but the proposed method seemed to perform well. The estimates of the prior smoothing parameter chosen as the first local minimum of \( \|D_\lambda y\| \) had smaller values than those calculated with the ML method. In the case of the artificial examples, ML seemed to produce posterior means that were closer to the true signal. For the sea level series, however, the prior smoothing parameter estimate from the scale-derivative map appeared to perform better than the ML estimate, which caused oversmoothing. Therefore, the scale-derivative map may also offer a means to evaluate the quality of the prior smoothing parameter estimate.

We compared the results of the first artificial example and the sea level example with SiZer maps and observed that especially the oscillatory features were more clearly separated in our scale-derivative maps. The scale-derivative maps of the examples seemed to detect features more sensitively than BSiZer, and the scale-derivative map may also be less prone to false-positives. The modified BSiZer based on second derivatives produced results that were rather similar to the scale-derivative maps. However, the credibility analysis of the scale-derivative maps seemed to color more areas credible, especially for the artificial series.

Acknowledgment

This study was funded by the Academy of Finland under grant no. 250862.

References

Chaudhuri, P & Marron, JS (1999), ‘SiZer for exploration of structures in curves’, Journal of the American Statistical Association, 94(447), 807–823.

Chaudhuri, P & Marron, JS (2000), ‘Scale space view of curve estimation’, The Annals of Statistics, 28(2), 408–428.

CU Sea Level Research Group (2013), Global and regional mean sea level time series. Available online at http://sealevel.colorado.edu/. Accessed in February 2013.

Erästö, P & Holmström, L (2005), ‘Bayesian multiscale smoothing for making inferences about features in scatter plots’, Journal of Computational and Graphical Statistics, 14(3), 569–589.

Erästö, P & Holmström, L (2007), ‘Bayesian analysis of features in a scatter plot with dependent observations and errors in predictors’, Journal of Statistical Computation and Simulation, 77(5), 421–431.

Godtliebsen, F & Øigård, T (2005), ‘A visual display device for significant features in complicated signals’, Computational Statistics & Data Analysis, 48(2), 317–343.
Hannig, J & Marron, JS (2006), ‘Advanced distribution theory for SiZer’, *Journal of the American Statistical Association*, **101**(474), 484–499.

Holmström, L (2010a), ‘BSiZer’, *Wiley Interdisciplinary Reviews: Computational Statistics*, **2**(5), 526–534. Available online at <http://dx.doi.org/10.1002/wics.115>.

Holmström, L (2010b), ‘Scale space methods’, *Wiley Interdisciplinary Reviews: Computational Statistics*, **2**(2), 150–159. Available online at <http://dx.doi.org/10.1002/wics.79>.

Holmström, L & Launonen, I (2013), ‘Posterior singular spectrum analysis’, *Statistical Analysis and Data Mining*, **6**, 387–402, DOI 10.1002/sam.11195.

Holmström, L & Pasanen, L (2012), ‘Bayesian scale space analysis of differences in images’, *Technometrics*, **54**(1), 16 –29, DOI 10.1080/00401706.2012.648862.

Holmström, L, Pasanen, L, Furrer, R & Sain, SR (2011), ‘Scale space multiresolution analysis of random signals’, *Computational Statistics & Data Analysis*, **55**(10), 2840 –2855. Available online at <http://dx.doi.org/10.1016/j.csda.2011.04.011>.

Lindeberg, T (1994), *Scale-space Theory in Computer Vision*, Kluwer Academic Publishers, Dordrecht, Netherlands.

Marr, D & Hildreth, E (1980), ‘Theory of edge detection’, *Proceedings of the Royal Society of London. Series B, Biological Sciences*, **207**(1167), 187–217, DOI 10.2307/35407.

Marron, JS & Zhang, J (2005), ‘SiZer for smoothing splines’, *Computational Statistics*, **20**, 481–502, DOI 10.1007/BF02741310.

Nerem, RS, Chambers, D, Choe, C & Mitchum, GT (2010), ‘Estimating mean sea level change from the TOPEX and Jason altimeter missions’, *Marine Geodesy*, **33**, 435.

R Core Team (2012), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Sonderegger, D (2012), *Sizer: Significant Zero Crossings*. R package version 0.1-4.

Unal, YS & Ghil, M (1995), ‘Interannual and interdecadal oscillation patterns in sea level’, *Climate Dynamics*, **11**, 255–278.