Statistical Scale Space Methods

Lasse Holmström and Leena Pasanen[*]

Department of Mathematical Sciences
University of Oulu
Finland
e-mail: `lasse.holmstrom@oulu.fi`; `leena.pasanen@oulu.fi`

Abstract: The goal of statistical scale space analysis is to extract scale-dependent features from noisy data. The data could be for example an observed time series or digital image in which case features in either different temporal or spatial scales would be sought. Since the 1990s, a number of statistical approaches to scale space analysis have been developed, most of them using smoothing to capture scales in the data, but other interpretations of scale have also been proposed. We review the various statistical scale space methods proposed and mention some of their applications.

Keywords and phrases: smoothing, multi-scale, curve fitting, time series, image processing, applications.

1. Introduction

The goal of scale space analysis is to discover the salient features of an object of interest that appear in different scales. The object considered can be for example a time series or a digital image in which case features are sought correspondingly in different temporal or spatial scales. The scale space representation of a signal consists of a family of its smooths. Often smoothing is done by convolving the signal with a Gaussian kernel, that is, by computing a moving average with Gaussian weights. The variance of the kernel indexes the individual scale space family members. For an image, scale space representation is simply a family of blurs and for a curve the representation is a family of smooth curves; see Figure 1 and the upper panel of Figure 2 for examples. Each smooth is thought to provide information about the object of interest at a particular scale or resolution.

The papers of Witkin (Witkin, 1983, 1984) are usually regarded as the origin of scale space methodology although it was only relatively recently noticed that such a concept was also proposed in Japan already in the 1950's (Iijima, 1959; Weickert et al., 1999). In Witkin's scale-space filtering, a one dimensional signal is smoothed with a range of smoothing parameter values and the smooths sweep out a surface he refers to as a "scale-space image". The signal is then described qualitatively by tracking the zeros of the second derivative of the smooths across all scales. Witkin required that increasing smoothing level creates no new extrema, a property often referred to as the causality condition. The fact that Gaussian convolution is the only smoother that satisfies such a condition was first shown in Babaud et al. (1983) and Babaud et al. (1986). Koenderink (1984) pointed out that Gaussian scale-space representation can be obtained by solving a linear diffusion equation. This was the starting point of a whole field of research on linear diffusion in image processing (Sporring, 1997). Perona and Malik (1990) were the first to extend this approach for nonlinear
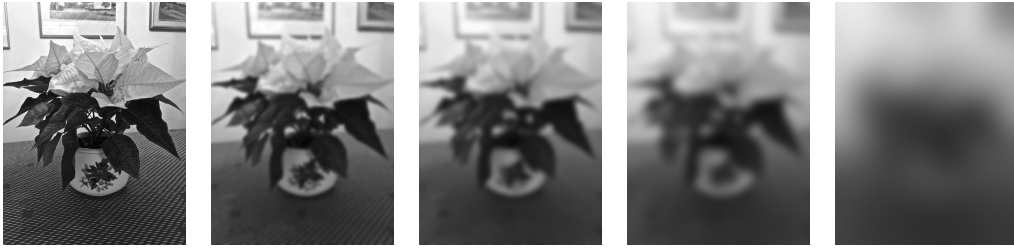
FIG 1. *Five images from the scale space representation of an image of a plant.*

diffusion processes. Since Witkin's work, various axiomatic foundations has also been developed for linear scale space theory (e.g. Chapter 6 in Sporring (1997)). Scale space methodology is nowadays well established in computer vision. Comprehensive introductions to the field can be found e.g. in Lindeberg (1994) and Sporring (1997).

The focus of this review is scale space as a tool in statistical analyses. Statistical scale space methodology began to emerge in the 1990s, first in mode detection for univariate and bivariate density estimation (Minnotte and Scott, 1993; Minnotte et al., 1998), gradually growing into a rich set of techniques quite distinct in character from the rest of scale space literature. The seminal work was the SiZer idea introduced in Chaudhuri and Marron (1999) and Chaudhuri and Marron (2000). As in Witkin's approach, a family of smooths was used to discover scale-dependent features of a curve. However, Chaudhuri and Marron allowed for the fact that the data at hand provides only an estimate of the underlying truth and their aim was to distinguish the actual structural features of the underlying curve from artifacts caused by noise. Statistical inference was therefore needed to assess which of the apparent features are structure and which are just sampling artifacts.

As in SiZer, the central goal in all statistics-based scale space analyses is to establish the significance or credibility of the scale-dependent features in the observed data (Holmström, 2010b). The inference results are summarized as easily interpretable "maps" that facilitate the application of these methods even by non-experts (e.g. Figure 2). Instead of the Gaussian convolution postulated in formal scale space theory, the statistics-based approaches often use other types of smoothers. The useful properties of a particular smoother is deemed to be more important than strict adherence to the scale space axioms. Outside statistical scale space literature such practice has also been motivated e.g. by computational efficiency (Wang and Lee, 1998).

There are, of course, other approaches one can use to analyze data in multiple scales, such as the various wavelet-based methodologies (e.g.Vidakovic (1999); Percival and Walden (2006)). We will however limit our scope only to techniques that use the level smoothing or some similar device to represent the scale in the object underlying the data. Even with our scope thus greatly narrowed, much territory needs to be covered as statistical scale space analysis has within a relatively short time blossomed into an impressively versatile ensemble of techniques with many interesting applications.

The rest of the article is organized as follows. Section 2 covers the original SiZer, its variants and some other related work. Univariate Bayesian variants of SiZer are described in Section 3 and various approaches to time series in Section 4. Extensions to higher dimensions and data on manifolds are considered in Section 5. Section 6 discusses various interpretations of scale, different uses of derivatives in feature detection and comparison of maps. Applications of scale

space methods are described in Section 7 and a short summary concludes the article in Section 8.

2. SiZer and its descendants

2.1. The original SiZer

We begin the more detailed discussion of statistical scale space methods by describing the SiZer approach of Chaudhuri and Marron. It has the closest connection to the formal scale space ideas when applied to density estimation. Thus, consider a sample $x_1, \ldots, x_n \sim p$ from a univariate probability density function p and its Gaussian kernel estimate,

$$\hat{p}_\lambda(x) = \frac{1}{n} \sum_{i=1}^{n} K_\lambda(x - x_i), \tag{1}$$

where the kernel K is the standard normal density function, $K_\lambda(x) = K(x/\lambda)/\lambda$, and $\lambda > 0$ is a smoothing parameter. Then the expectation

$$\mathbb{E}\hat{p}_\lambda(x) = \int_{-\infty}^{\infty} K_\lambda(x - y)p(y)dy = K_\lambda * p(x) \tag{2}$$

is a smooth of p obtained by Gaussian convolution. We will in the following also use the notations,

$$p_\lambda(x) = S_\lambda p(x) = K_\lambda * p(x),$$

where S_λ denotes the smoothing operator. Now $\{\hat{p}_\lambda | \lambda > 0\}$ can be viewed as the scale space representation of the data $x_1, \ldots, x_n$. By (2) this is an unbiased estimate of the scale space representation $\{p_\lambda | \lambda > 0\}$ of the true underlying density p. SiZer finds the statistically significant features in the smooths p_λ and summarizes the results in a map. Note here the difference between ordinary smoothing-based inference and the scale space approach. While in conventional inference the object of interest would be the unknown function p, scale space inference instead targets its smooths p_λ.

Figure 2 shows an example of SiZer analysis. The $n = 272$ observations x_i are the the durations in minutes of the eruption or the Old Faithful geyser in Yellowstone National Park, Wyoming, USA (Härdle, 1991; Azzalini and Bowman, 1990). The upper panel, the so-called family plot, displays part of the scale space of the data, that is, a family of kernel density estimates $\hat{p}_\lambda$ for a number of smoothing levels λ. The lower panel, the SiZer map, summarizes inference about the statistically significant features of the smooths p_λ. The inference is based on the sign of the derivative $p'_\lambda(x) = \mathbb{E}\hat{p}'_\lambda(x)$ – SiZer is an acronym for Significant ZERo crossings of derivatives. A pixel at $(x, \log_{10}(\lambda))$ is colored blue or red depending on whether the confidence interval of $p'_\lambda(x)$ (cf. (3)) is above or below 0. Purple means that the confidence interval of the derivative includes 0 and gray means that the data are too sparse for inference, indicated by an effective sample size smaller than 5 (cf. (4)). The black line marks the smoothing level that produces the black curve in the upper panel. In the largest scales the underlying function p is unimodal with maximum at around 3.5 minutes because the derivative is positive to its left and negative to its right. For a large range of scales there is clear evidence for two strong modes but none of the smaller modes turns out be significant.
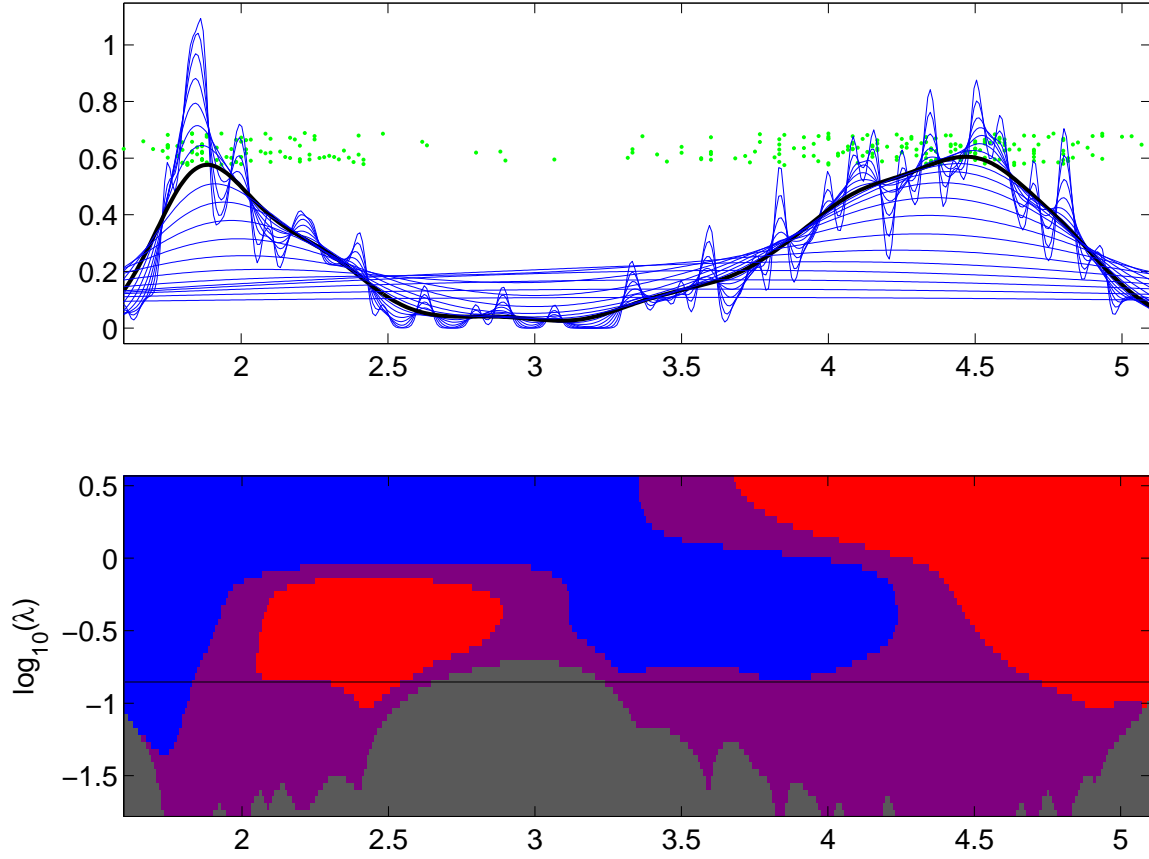
FIG 2. *SiZer analysis of the Old Faithful Data. Upper panel: the family plot of kernel density estimates of the data shown as jittered green dots. Lower panel: the SiZer map that summarizes the statistical significance of the sign of the derivative of the smooths of the density underlying the data, blue for positive slopes and red for negative slopes. Purple indicates a non-significant slope and gray data too sparse for inference. The smoothing level indicated by the black line corresponds to the smooth shown in black in the family plot. The level of significance used is* 0.05. *The analysis was produced by the software provided by J.S. Marron at* http: // www. unc. edu/ ~marron/ marron. html .

The confidence interval for $p'_\lambda(x)$ is of the form

$$\hat{p}'_\lambda(x) \pm q\widehat{\mathrm{SD}}(\hat{p}'_\lambda(x)), \tag{3}$$

with an appropriate quantile $q > 0$. The standard deviation in the confidence limits can be obtained by using the fact that the kernel estimator consists of sums of independent terms. Note how scale space analysis sidesteps the bias problem always present in nonparametric smoothing since the confidence interval is centered precisely on the quantity of interest, $\mathbb{E}\hat{p}'_\lambda(x) = p'_\lambda(x)$.

In Chaudhuri and Marron (1999) four alternatives to determine q were proposed, independent Gaussian quantile for each x, an approximate simultaneous quantile over all x based on approximately independent blocks of data, and two methods based on the bootstrap, one which is simultaneous over x and one which is simultaneous over both x and λ. For example, the approximate simultaneous method estimates the effective sample size (ESS) for each (x, λ) as

$$\mathrm{ESS}(x, \lambda) = \sum_{i=1}^{n} K_\lambda(x - x_i)/K_\lambda(0) \tag{4}$$

and defines the number of independent blocks as $m = n/\mathrm{avg}_x\mathrm{ESS}(x, \lambda)$. The associated quantile is obtained as

$$q = \Phi^{-1}\left(\frac{1 + (1 - \alpha)^{1/m}}{2}\right)$$

where $1-\alpha$ is the confidence level and Φ is the Gaussian distribution function. Nowadays the default inference method of SiZer software uses extreme value theory for Gaussian processes for improved multiple hypothesis testing required in simultaneous inference (Hannig and Marron, 2006). The map of Figure 2 in fact uses this method. Such inference based on advanced distribution theory is also used in many variants of SiZer developed in recent years. In a SiZer map, the range of smoothing values used should be selected so that all interesting features in the data are shown. This is discussed in Marron and Chung (2001) and Chaudhuri and Marron (1999).

The original SiZer considered also regression. The model is

$$y_i = \mu(x_i) + \varepsilon_i, \tag{5}$$

where $x_1 < \cdots < x_n$ are values of an explanatory variable, μ is the unknown regression function, and the errors ε_i are independent but possibly heteroscedastic. Local linear estimates of μ and μ' are employed, $\hat{\mu}_\lambda(x) = \hat{a}_\lambda$, $\hat{\mu}'_\lambda(x) = \hat{b}_\lambda$, where

$$(\hat{a}_\lambda, \hat{b}_\lambda) = \operatorname*{argmin}_{a,b} \sum_{i=1}^{n} [y_i - (a + b(x_i - x))]^2 K_\lambda(x - x_i), \tag{6}$$

and SiZer explores the significant features in the scale space representation $\{\mu_\lambda | \lambda > 0\}$, where

$$\mu_\lambda(x) = S_\lambda\mu(x) \equiv \mathbb{E}\hat{\mu}_\lambda(x).$$

In the regression setting the confidence interval for the derivative $\mu'_\lambda(x)$ also needs an estimate for the error variance in (5). Causality in scale space smoothing is now more complicated as for example in the case of local linear regression monotonicity of the number of extrema with respect to the scale does not hold anymore (Chaudhuri and Marron, 2000).
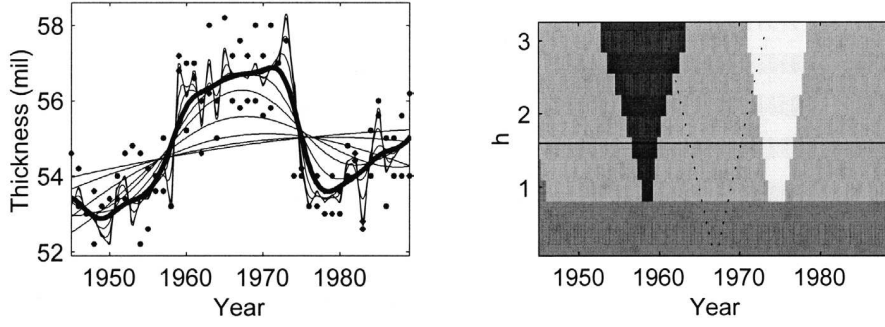
FIG 3. *Jump SiZer analysis of penny the thickness data from Kim and Marron (2006). Left panel: the family plot. Right panel: the SiZer map with black and white indicating positive and negative slope, respectively, and light and dark gray indicating non-significance and data sparseness, respectively. Funnels highlight two jumps. The smoothing parameter (here denoted by h) has a linear scale. The dotted lines show the effective size of the smoothing kernel.*

2.2. SiZer variants

A smoothing spline version of SiZer for regression was proposed in Marron and Zhang (2005). The authors argue that even though splines and local linear regression perform similarly, neither can replace the other. For instance, with few data points and low noise level, spline smoothing appears to be more effective than local linear regression whereas with more, noisier data, local linear regression generally seems to work better.

Hannig and Lee (2006) proposed a robust SiZer for regression. The estimator (6) is replaced by

$$(\hat{a}_{\lambda,c}, \hat{b}_{\lambda,c}) = \operatorname*{argmin}_{a,b} \sum_{i=1}^{n} \rho_c \left[\frac{y_i - (a + b(x_i - x))}{\hat{\sigma}} \right] K_\lambda(x - x_i), \tag{7}$$

where $\hat{\sigma}$ is a robust estimate of the error standard deviation and ρ_c is the Huber loss with cut-off $c > 0$,

$$\rho_c(x) = \begin{cases} x^2/2, & \text{if } |x| \le c, \\ |x|c - c^2/2, & \text{if } |x| > c. \end{cases}$$

Different values of c produce SiZer maps with different levels of robustness: the smaller the cut-off, the more robust the fit. Robust SiZer can also be used for outlier detection.

As the name suggests, the jump SiZer of Kim and Marron (2006) is designed to detect sudden jumps in a curve underlying noisy data. A jump produces a "funnel shape" in the SiZer map because the slopes are significant even for very small smoothing parameter values. The boundary of a jump funnel grows approximately linearly with the smoothing parameter and therefore visualization of inference results is more effective if the usual logarithmic scale $\log_{10} \lambda$ in a SiZer map is replaced by the linear scale λ. However, for the purposes of jump detection such a linear scale focuses the analysis too much on the larger scales and the authors therefore recommend that only the lower half of the usual SiZer smoothing parameter range is used when the results are displayed. Figure 3 shows an example of jump SiZer analysis. The Mutagram of Müller and Wai (2004) finds change points in a time series associated with a sudden jump or change in the slope. As in SiZer-like methods, multiple scales are considered and the results are summarized graphically.

Marron and de Uña-Álvarez (2004) extended SiZer for density and hazard rate estimation when the data are length biased, that is, when the probability of an observation appearing in the sample is proportional to its value. In addition to length biasedness the data can also be censored.

The quantile SiZer of Park et al. (2010b) targets the quantile structure of the data by replacing local linear regression by a local linear quantile smoother. The smooth corresponding to the quantile of order $0 < \tau < 1$ is defined by replacing (6) with

$$(\hat{a}_{\lambda,\tau}, \hat{b}_{\lambda,\tau}) = \operatorname*{argmin}_{a,b} \sum_{i=1}^{n} \rho_\tau[y_i - (a + b(x_i - x))]K_\lambda(x - x_i), \tag{8}$$

where ρ_τ is the so-called check loss function

$$\rho_\tau(x) = \begin{cases} x\tau, & \text{if } x \geq 0, \\ x(\tau - 1), & \text{if } x < 0. \end{cases}$$

Corresponding to a set of different quantiles, a collection of quantile SiZer maps is used to explore the structure underlying the data.

The causal SiZer (c-Sizer) of Skrøvseth et al. (2012c) is concerned with detecting change points in a live monitoring system such as disease surveillance. The authors consider two problems, changes in the distribution of event occurrence times and changes in measurements made at those times. The former involves density estimation and the latter regression. The analysis should be strictly causal, that is, a detected event should strictly precede the observed effect in time. However, due to its infinite support, use of the Gaussian kernel in estimation does not lead to this kind of causality since an estimate at some time x involves data that become available only after x. Causal SiZer is a variant of SiZer where the temporal causality of inference is enforced by employing a symmetric kernel with finite support lagged by one half of the support. Instead of single points, the method retrospectively produces reliable interval estimates of the change points.

In the case of discrete data, the least squares regression setting of the original SiZer may be inefficient and use of likelihood or quasi-likelihood based estimation can work better. Such a version of SiZer was proposed in Li and Marron (2005). In a local polynomial quasi-likelihood approach one uses a link function g and a quasi-likelihood Q and obtains a smooth of the regression function by replacing the square loss in the formula (6) by $Q(g^{-1}(a + b(x - x_i)), y_i)$, argmin by argmax, and then defining $\hat{\mu}_\lambda(x) = g^{-1}(\hat{a}_\lambda)$. Because the link function is usually monotone, inference and the construction of the SiZer map can be carried out for the transform $g \circ \mu$, instead of μ itself which simplifies the analysis. However, numerical maximization is required in the estimation and the authors develop an algorithm which manages to keep the computational burden close to the level of the original least squares SiZer. Park and Huh (2013b) improved on this work basing the SiZer inference on an extension of the Gaussian process approach of Hannig and Marron (2006) to discrete data. An asymptotic analysis of the method is also provided.

In a varying coefficient linear regression model the regression coefficients are allowed to be a nonparametric functions of a covariate,

$$y_i = \sum_{j=1}^{m} \beta_j(u_i)x_{ij} + \varepsilon_i,$$

where the u_i's and x_{ij}'s are values of covariates, conditional on which the error is zero mean with variance possibly depending on u_i. The coefficient functions β_j are modeled non-parametrically and Zhang and Mei (2012) described a version of SiZer to explore their significant features. Using local linear approximation one estimates $\hat{\beta}_{j,\lambda}(u) = \hat{a}_{j,\lambda}$, $\hat{\beta}'_{j,\lambda}(u) = \hat{b}_{j,\lambda}$, where

$$(\hat{\mathbf{a}}_\lambda, \hat{\mathbf{b}}_\lambda) = \operatorname*{argmin}_{\mathbf{a},\mathbf{b}} \sum_{i=1}^{n} \{y_i - \sum_{j=1}^{m} [a_j + b_j(u_i - u)]x_{ij}\}^2 K_\lambda(u_i - u), \qquad (9)$$

and $\mathbf{a} = [a_1, \ldots, a_m]^T$, $\mathbf{b} = [b_1, \ldots, b_m]^T$, and the minimizers are $\hat{\mathbf{a}}_\lambda = [\hat{a}_{1,\lambda}, \ldots, \hat{a}_{m,\lambda}]^T$, $\hat{\mathbf{b}}_\lambda = [\hat{b}_{1,\lambda}, \ldots, \hat{b}_{m,\lambda}]^T$. A separate SiZer map is constructed for each β_j. Zhang et al. (2013) proposed an robust alternative to this approach, replacing the square in (9) by absolute value. Confidence intervals are obtained with a modified version of the residual-based wild bootstrap.

Park and Huh (2013a) applied SiZer to the log-variance function in non-parametric regression. In a random design interpretation of (5), let $\varepsilon|x \sim \mathrm{N}(0, v(x))$ and consider the squared residual $r = (y - \mu(x))^2$. Then $r|x \sim v(x)\chi_1^2$ and the log-likelihood of r given $s(x) = \log(v(x))$ is

$$\ell(s(x), r) \equiv \log p(r|s(x)) = -(1/2)[\log(2\pi) + \log(r) + s(x) + r\exp(-s(x))].$$

In a local likelihood procedure one considers the data $r_i = (y_i - \mu(x_i))^2$ and replaces the square loss in (6) by $\ell(a + b(x_i - x), r_i)$ and argmin by argmax. The log-variance and its derivative are then estimated by $\hat{s}_\lambda(x) = \hat{a}_\lambda$, $\hat{s}'_\lambda(x) = \hat{b}_\lambda$. Here the unknown function μ is replaced by a pilot smooth and the pilot smoothing parameter becomes another dimension in scale space analysis in the same way as in Rondonotti et al. (2007) (cf. Section 4.1). Confidence intervals are estimated by applying the Gaussian process technique of Hannig and Marron (2006). Asymptotic analysis of the method is also provided.

2.3. Some other uses of the scale space

González-Manteiga et al. (2008) use SiZer for tests of the components of a non-parametric additive regression model. Consider a d-dimensional covariate vector $\mathbf{x} = [x_1, \ldots, x_d]^T$ and a model $y = \mu(\mathbf{x}) + \sigma(\mathbf{x})\varepsilon$, where ε is independent of $\mathbf{x}$, has zero mean and unit variance. An additive model assumes that the conditional mean is of the form

$$\mu(\mathbf{x}) = c + \sum_{j=1}^{d} \mu_j(x_j),$$

where where c is a constant and the component functions μ_j satisfy smoothness and identifiability conditions. Tests are proposed for additivity, component significance, the adequacy of parametric models for the components and for interactions. The idea is to use SiZer and its second derivate version, SiCon (cf. Section 6), to examine the residuals when a particular model is fitted. Significant features in the residuals indicate a poor fit. For interactions, one applies scale space analysis to differences of fits between models with and without interactions or, in case of logistic regression, differences between deviances. Additive models are considered also in the report Martínez-Miranda (2005) where a bootstrap-based local (adaptive) smoothing parameter selection strategy for component estimation is evaluated using a separate SiZer map for each μ_j. Using visual cues offered

by the SiZer map in local bandwidth selection was in fact contemplated also in the original work of Chaudhuri and Marron.

Several papers have proposed to investigate differences between regression functions using scale space analysis. Given two regression functions μ_1 and μ_2, Park and Kang (2008) make inferences about their difference by performing SiZer analysis on $\mu = \mu_2 - \mu_1$. However, instead of its derivative, the function μ itself is used and therefore the confidence interval (3) is replaced by $\hat{\mu}_\lambda(x) \pm q\widehat{\text{SD}}(\hat{\mu}_\lambda(x))$. To make comparisons between more than two regression functions the authors propose to analyze the distributions of the fitted residuals. Thus, consider m models,

$$y_{ji} = \mu_j(x_{ji}) + \varepsilon_{ji}, \;\; j = 1, \ldots, m, \; i = 1, \ldots n_j.$$

A pilot smoothing parameter λ_p is chosen and two sets of standardized residuals are computed, one based on the differences $y_{ij} - \hat{\mu}_{\lambda_p}(x_{ji})$ that correspond to the null hypothesis that all μ_j's are equal and another on the differences $y_{ij} - \hat{\mu}_{j,\lambda_p}(x_{ji})$ that corresponds to the alternative that some of the μ_j's may be different. The density functions of these two sets of residuals are then subjected to SiZer analysis in the same manner as in the case of the two regression functions μ_1 and μ_2 above and discovery of significant features indicates that the regression functions are not all equal. As before, the pilot smoothing parameter adds a new dimension to the scale space inference. This approach to the comparison of regression functions was extended to time series in Park et al. (2009) by employing SiZer for dependent data (cf. Section 4.1). However, for comparison of more than two time series, instead of considering residual densities, a pair of time series is assumed to be observed at the same time points x_i and SiZer analysis is applied directly to the difference of the residual time series $y_{ij} - \hat{\mu}_\lambda(x_i)$ and $y_{ij} - \hat{\mu}_{j,\lambda}(x_i)$. Recently, an ANOVA type approach for comparing multiple curves observed at arbitrary values of the dependent variable was developed in Park et al. (2014). The advantage of this method over the ones that rely on residuals is that, in the spirit of original SiZer, one obtains information also about the potential local scale dependent differences of the underlying curves themselves.

Differences between trends of two non-stationary time series was considered also in Godtliebsen et al. (2012) but, instead of analyzing the difference of means $\mu = \mu_2 - \mu_1$ itself, similarly to SiZer, inference was applied to the derivative $\mu'_\lambda = \mu'_{2,\lambda} - \mu'_{1,\lambda}$ for a range of scales. In addition to the original SiZer, the Bayesian BSiZer was also employed (cf. Section 3.1). In the context of image analysis, the Bayesian SiZer for images, iBSiZer, also considers differences of functions (Section 5.2).

An interesting application of scale space thinking is the multi-scale approach to supervised classification developed by A. Ghosh and his co-workers. In a two-class situation Ghosh et al. (2006b) propose to estimate the posterior probability in favor of class 1 by

$$\hat{p}_{\lambda_1,\lambda_2}(1|\mathbf{x}) = \frac{\pi_1 \hat{p}_{\lambda_1}(\mathbf{x}|1)}{\pi_1 \hat{p}_{\lambda_1}(\mathbf{x}|1) + \pi_2 \hat{p}_{\lambda_2}(\mathbf{x}|2)},$$

where π_1 and π_2 are the prior probabilities of the two classes and $\hat{p}_{\lambda_1}(\mathbf{x}|1)$ and $\hat{p}_{\lambda_2}(\mathbf{x}|2)$ are kernel estimates of the class densities. The posterior probability of an observation $\mathbf{x}$ or, alternatively, the "p-value" $P_{\lambda_1,\lambda_2}(\mathbf{x}) = \mathbb{P}\{\pi_1 \hat{p}_{\lambda_1}(\mathbf{x}|1) > \pi_2 \hat{p}_{\lambda_2}(\mathbf{x}|2)\}$ is then visualized with a two-dimensional plot using the smoothing parameters λ_1 and λ_2 as coordinates. The final classifier is based on a weighted average of class posterior probabilities corresponding to different smoothing parameter

combinations (λ_1, λ_2), where the weights take into account the estimated misclassification rate as well as the *p*-value of each observation. A related approach is applied to *k*-nearest neighbor classification in Ghosh et al. (2005, 2006a).

3. Univariate Bayesian methods

Two Bayesian approaches to univariate scale space regression have been proposed, BSiZer and Posterior Smoothing. Multi-scale analysis of features of the function μ in (5) is based on the posterior distribution $p(\mu'_\lambda(x)|\mathbf{y})$ of the slopes of the smooths $\mu_\lambda = S_\lambda\mu$, given the data $\mathbf{y} = [y_1, \ldots, y_n]^T$. The SiZer map is replaced by a feature credibility map computed from the posterior distribution of the slopes. One advantage of the Bayesian approach is that inference about the features of the underlying function can be done even in the smallest scales because data sparsity is not a problem. Of course, reliance on posterior inference assumes that the likelihood and the priors used are reasonable. The Bayesian methods are also often computationally more demanding than the frequentist approaches.

3.1. BSiZer

The BSiZer first introduced in Erästö and Holmström (2005) dealt only with a discrete parameter vector $\boldsymbol{\mu} = [\mu_1, \ldots, \mu_n]^T$ but it was subsequently extended in Erästö and Holmström (2007) to the case of a smooth regression function by using a spline model; see also Holmström (2010a) for a comprehensive review. Thus, assume that μ is a natural cubic spline, let $a \le x_1 < \cdots < x_n \le b$ be fixed values of the explanatory variable and let $\mu_i = \mu(x_i)$ be the value of the spline at the knot x_i. The spline μ is uniquely determined by its values at the knots. Denote by $\mathbf{S}_\lambda$ and $n \times n$ smoothing matrix and let $\boldsymbol{\mu}_\lambda = \mathbf{S}_\lambda\boldsymbol{\mu}$ be the smooth of $\boldsymbol{\mu}$. Instead of the posterior distribution $p(\mu'_\lambda|\mathbf{y})$ of the function μ'_λ one then analyzes the finite dimensional distribution $p(\mathbf{D}\boldsymbol{\mu}_\lambda|\mathbf{y})$, where $\mathbf{D}\boldsymbol{\mu}_\lambda = [\mu'_\lambda(\xi_1), \ldots, \mu'_\lambda(\xi_r)]^T$ is the vector of slopes of μ_λ at some fixed set of points $a < \xi_1 < \cdots < \xi_r < b$ and $\mathbf{D}$ is the matrix that computes the derivatives $\mu'_\lambda(\xi_j)$ from the vector $\boldsymbol{\mu}_\lambda$. Since the distribution $p(\mathbf{D}\boldsymbol{\mu}_\lambda|\mathbf{y})$ is obtained by applying the linear transformation $\mathbf{D}\mathbf{S}_\lambda$ to $p(\boldsymbol{\mu}|\mathbf{y})$ one only needs to deduce $p(\boldsymbol{\mu}|\mathbf{y})$.

Thus, let $\boldsymbol{\varepsilon} = [\varepsilon_1, \ldots, \varepsilon_n]^T$ be the vector of errors in (5) where in the simplest case one assumes that $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$. Then the likelihood $p(\mathbf{y}|\boldsymbol{\mu}, \sigma^2)$ is Gaussian with mean $\boldsymbol{\mu}$ and covariance $\sigma^2\mathbf{I}$. An inverse-χ^2 prior $p(\sigma^2)$ is assumed for the error variance and for $\boldsymbol{\mu}$ a smoothing prior is used,

$$p(\boldsymbol{\mu}|\lambda_0) \propto \lambda_0^{(n-2)/2} \exp\left[-(\lambda_0/2)\boldsymbol{\mu}^T\mathbf{K}\boldsymbol{\mu}\right], \tag{10}$$

where $\mathbf{K}$ is a positive semidefinitive matrix such that $\boldsymbol{\mu}^T\mathbf{K}\boldsymbol{\mu} = \int_a^b [\mu''(x)]^2 dx$. The prior penalizes μ for roughness and the level of penalty is controlled by the parameter $\lambda_0 > 0$ which can be assigned a separate hyper prior. The scale space smoothing matrix is defined as $\mathbf{S}_\lambda = (\mathbf{I} + \lambda\mathbf{K})^{-1}$.

The joint posterior $p(\boldsymbol{\mu}, \sigma, \lambda_0|\mathbf{y})$ is obtained from Bayes' theorem,

$$p(\boldsymbol{\mu}, \sigma, \lambda_0|\mathbf{y}) \propto p(\sigma^2)p(\boldsymbol{\mu}|\lambda_0)p(\lambda_0)p(\mathbf{y}|\boldsymbol{\mu}, \sigma^2).$$

The posterior credibility of the slopes $\mu'_\lambda(x)$ are inferred by first simulating a large sample from the marginal distribution $p(\boldsymbol{\mu}|\mathbf{y})$ and the transforming the sample by $\mathbf{D}\mathbf{S}_\lambda$. A credibility map

corresponding to the independent quantile SiZer map could then be constructed by choosing a threshold value $0.5 < \alpha < 1$ and coloring a map pixel (ξ_j, λ) blue or red according to whether $\mathbb{P}(\mu'_\lambda(\xi_j) > 0 \mid \mathbf{y}) \geq \alpha$ or $\mathbb{P}(\mu'_\lambda(\xi_j) < 0 \mid \mathbf{y}) \geq \alpha$ (and gray otherwise). However, the maps are in fact drawn based on the joint posterior probabilities over the locations ξ_j. An example of a BSiZer map is displayed in the lower panel of Figure 11. BSiZer analysis can also handle correlated errors as well as situations where the x_i's themselves contain errors (Erästö and Holmström, 2007).

3.2. Posterior Smoothing

A slightly different Bayesian interpretation of SiZer, Posterior Smoothing, was proposed in Godtliebsen and Øigård (2005). In the model (5), Gaussian independent errors are assumed and the function μ is thought of as a time-varying signal observed on an equispaced grid of times x_i. The model allows also a possible blurring by replacing $\boldsymbol{\mu}$ by $\mathbf{A}\boldsymbol{\mu}$ where $\mathbf{A}$ is a known matrix. A homoscedastic situation is considered and the error variances are assumed to be known or they are estimated from the data before Bayesian analysis is applied. A smoothing prior is again adopted,

$$p(\boldsymbol{\mu}|\lambda_0) \propto \exp\left[-\lambda_0 \sum_{i=1}^{n-1}(\mu_i - \mu_{i+1})^2\right], \tag{11}$$

but now λ_0 acts as the scale space smoothing parameter. The credibility map is computed from the approximate posterior distributions of the discrete derivatives and simultaneous inference is approximated with the independent blocks idea of the original SiZer.

Øigård et al. (2006) made several important improvements to the original Posterior Smoothing method. These include a better prior model for the signal, use of band-matrix algorithms for exact and computationally efficient inference, and a better way to summarize the results in a feature map. Assuming in (5) that $x_1 > 0$, an integrated Wiener process prior

$$\mu(x)|\{\lambda_0, \mu(0) = 0\} = \frac{1}{\sqrt{\lambda_0}} \int_0^x (x - h)dW(h) \tag{12}$$

is used where $W(h)$ is a standard Wiener process. This choice of prior allows one to evaluate the posterior distributions of the true derivatives $\mu'(x_i)$ instead of using their discrete approximations. Note that there is a well-known connection between the integrated Wiener process and the cubic splines that are used in the BSiZer prior (10) (Wahba, 1978). The key idea in improved Posterior Smoothing is to consider, instead of $\boldsymbol{\mu} = [\mu(x_1), \ldots, \mu(x_n)]^T$, the augmented random vector $\boldsymbol{\eta} = [\mu(t_1), \mu'(t_1), \ldots, \mu(t_n), \mu'(t_n)]^T$. It is viewed as a Gaussian Markov random field with an improper Gaussian prior distribution $p(\boldsymbol{\eta}|\lambda_0)$ which is of type (10). This, combined with a Gaussian likelihood produces a multivariate normal posterior $p(\boldsymbol{\eta}|\mathbf{y}, \lambda_0)$ whose mean and precision matrix (inverse covariance matrix) can be found exactly. The band structure of the precision matrix facilitates fast computations of the posterior mean and marginal variances. In the credibility map, an exact test of the form $|\mathbb{E}(\mu'(t_i)|\mathbf{y}, \lambda_0)| > q(\alpha/2)\text{SD}(\mu'(t_i)|\mathbf{y}, \lambda_0)$, can be applied where $q(\alpha/2)$ is the ordinary Gaussian $\alpha/2$ quantlile. A suitable range of smoothing levels for the credibility map is automatically determined using a concept of effective sample size (ESS) which is natural for the particular Bayesian model used.

BSiZer and Posterior Smoothing interpret SiZer somewhat differently. The idea of SiZer is to make inferences about the smooths of the underlying true function μ and BSiZer aims at the same

goal by first explicitly constructing its posterior distribution and then smoothing it. The separation of posterior modelling and scale space smoothing also facilitates changes to the observation model (5). Posterior Smoothing, on the other hand, implements the scale space idea by reconstructing the underlying μ using a range of alternative prior smoothing parameter values and making inferences about the features in these reconstructions.

4. Time series

4.1. Time domain analyses

Time series trend estimation can be interpreted as regression with a fixed equispaced design and correlated errors. Thus, in (5) $x_i = i$ and the ε_i's are dependent. In general, however, a non-constant mean of a non-stationary time series cannot be distinguished from a correlated error structure of a stationary time series on the basis of just one observed series. The time series SiZer of Rondonotti et al. (2007) assumes that the ε_i's are weakly stationary and then uses a pilot smooth of the data to estimate their autocovariance. The pilot smoothing parameter λ_p becomes a new scale space dimension and a correlation-adjusted SiZer map is produced for several alternative values of λ_p, each corresponding to a particular trade-off between an assumed trend and level of dependence in the data. Small λ_p postulates weakly correlated errors and a highly structured trend while a large λ_p allows the errors to explain most of the observed features in the data. In the spirit of a scale space, this SiZer for time series avoids commitment to a particular error model by considering a whole family of them simultaneously. However, in the final visualization of inference results only four values of the pilot smoothing parameter are considered that reflect a wide variety of trade-offs between the trend and error correlation. As a numerical measure of this trade-off one uses the ratio

$$\mathrm{IR}(\lambda_p) = \sum_{i=1}^{n} \hat{\varepsilon}_{\lambda_p,i}^2 / \left(\max_{\lambda_p} \sum_{i=1}^{n} \hat{\varepsilon}_{\lambda_p,i}^2 \right), \tag{13}$$

(Indicator of the Residual component), where $\hat{\varepsilon}_{\lambda_p,i}$'s are the residuals of the pilot smooth. The four pilot smooths correspond to values of λ_p for which the ratio is closest to 0%, 25%, 50%, and 75%. Figure 4 shows an example of such scale space analysis with time series SiZer.

Several improvements to time series SiZer were suggested in Park et al. (2009). The quantiles in the confidence intervals were estimated using the extreme value theory ideas of Hannig and Marron (2006) and the autocovariance function was estimated based on differenced time series. These changes result in a method that for time series with moderately correlated errors does not require pilot smoothing.

In a Bayesian approach, BSiZer also allows for a dependent Gaussian error structure and Erästö and Holmström (2007) discusses how it may be used to distinguish dependence in errors from an underlying trend.

Park et al. (2004) introduced the Dependent SiZer taking a goodness-of-fit approach to SiZer for time series where, instead of estimating the dependent error structure from the data and using it to adjust scale space inference accordingly, a specific time series model, such as a fractional Gaussian noise process is assumed and then one tests whether the model fits the data. Now the SiZer map flags deviations from the model while a good fit corresponds to feature non-significance.
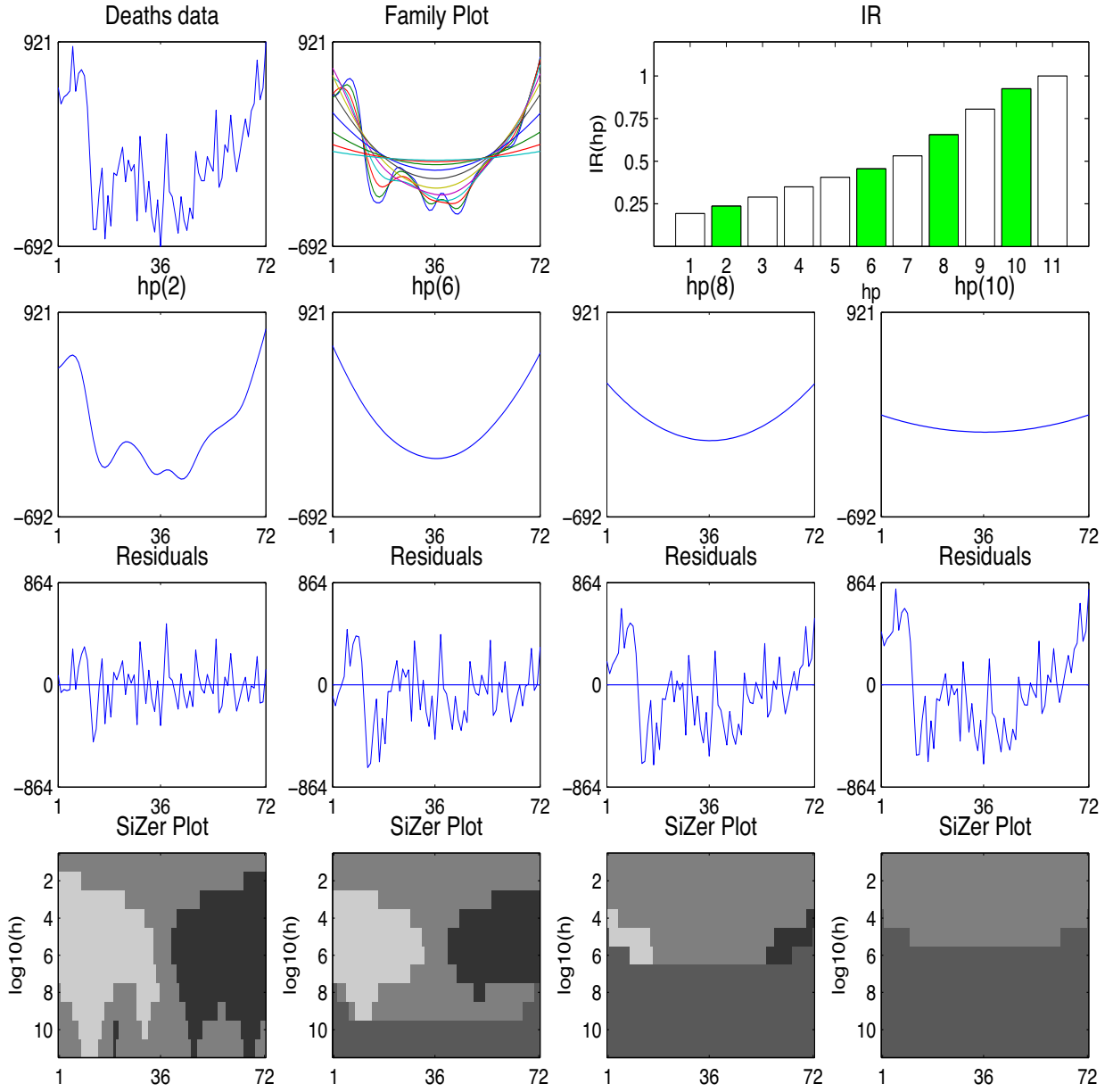
FIG 4. *Time series SiZer analysis of deaths data from Rondonotti et al. (2007). First row: the data, the family plot and the IR defined in (13); the smoothing parameter is denoted by h. Green highlights the IRs that correspond to the four pilot smoothing parameter values considered in the next rows. Second row: the smooths corresponding to the selected pilot smoothing parameters. Third row: the corresponding residuals. Fourth row: the SiZer maps where colors are interpreted as in Figure 3.*

The SiNos (Significant Non-stationarities) method of Olsen et al. (2007) detects deviations from stationarity of a time series in the form of a change in the mean, variance or 1-step autocorrelation. A pair of adjacent windows is slid along the time series and sample estimates of these statistics within the two windows are compared to test the null hypothesis of no departure from stationarity. The test statistics can be expressed as ratios of quadratic forms and Gaussian distribution-based saddle point approximation is used to determine their p-values. No smoothing is involved but the scale space idea is implemented by varying the length of the two windows used for testing. The global significance level on each scale is controlled by adopting the method of false discovery rate (FDR). In the wavelet SiZer and wavelet SiNos of Park et al. (2007) non-stationarities are detected by performing scale space analysis on sequences of squares (or other functions) of wavelet coefficients of a time series. Thus, for a fixed wavelet scale, SiZer (or SiNos) is applied to the time series of squared wavelet coefficients and the process is repeated for a number of wavelet scales. The analysis is facilitated by the fact that, for a fixed wavelet scale, the wavelet coefficients are only weakly correlated.

The MRAD (Multi-Resolution Anomaly Detection) procedure described in Zhang et al. (2008) and Zhang et al. (2014) detects outliers in time series data with long-range dependence, such as Internet traffic. The technique operates on multiple scales that correspond to window widths used to aggregate the data. An observed time series $\mathbf{y}_1 = [y_{11}, \ldots, y_{1n}]$ is aggregated by summing over a scale of dyadic window widths either using non-overlapping or overlapping sliding windows. At scale k, an aggregated series $\mathbf{y}_k = [y_{k1}, \ldots, y_{kn_k}]$ is produced, where $n_k = \lceil n/2^{k-1} \rceil$ and each y_{ki} is a sum over 2^{k-1} original observations normalized to have, under the null hypothesis of fractional Gaussian noise (no outliers), the same Gaussian marginal distribution that does not depend on k or i. In the case of non-overlapping windows,

$$y_{ki} = (2^{k-1})^{-H} \sum_{j=1}^{2^{k-1}} y_{1,(\lceil i/2^{k-1} \rceil - 1)2^{k-1}+j},$$

where H is the Hurst exponent of the process and similarly for overlapping windows. Such aggregation is similar to wavelet smoothing associated with Haar wavelets. The test to reject the null hypothesis at time i is of the form $\max_{k=1,\ldots,M} |y_{ki}| > C_\alpha^M$, where C_α^M depends on the estimated Hurst exponent, the chosen significance level α and M, the number of scales examined. In the MRAD map, the pixel color at a location (i, k) indicates the p-value of $|y_{ki}|$. Such multi scale testing can increase the power in outlier detection. Figure 5 shows an example of MRAD in action (Figure 3.19 of Zhang (2007)).

4.2. Frequency domain analyses

A local maximum in the spectral density of a stationary time series may indicate a periodic component in the data. However, a spectrum estimate such as the periodogram usually also shows features that are in fact just sampling artifacts. The idea in scale space spectral analysis is the same as in SiZer for density estimation: a feature, such as a local maximum is more certainly real if it appears in a substantial range of scales rather than just making a spurious appearance in a narrow range of smooths.
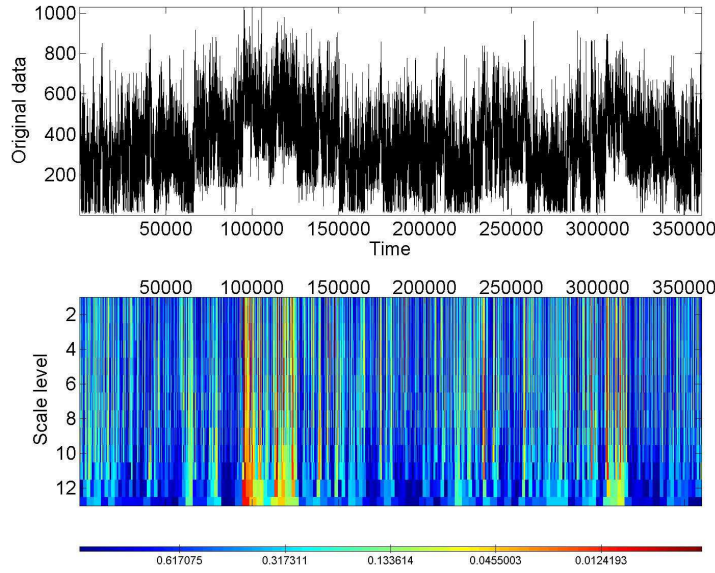
FIG 5. *MRAD analysis of an Internet traffic semi-experiment. Top: packet counts during an interval of one hour with an artificially injected port scan anomaly around 100000. The time unit is $10^{-2}s$. Bottom: MRAD map where the color of a pixel indicates p-value.*

Sørbye et al. (2009) propose a Bayesian scale space analysis of the log-spectral density. Consider an observed finite time series $z_1, \ldots, z_{2n}$ and let

$$P(\omega) = \frac{1}{2n} \left| \sum_{k=1}^{2n} z_t e^{-2\pi i \omega k} \right|^2, \qquad -1/2 \leq \omega \leq 1/2 \tag{14}$$

be the periodogram estimate of the spectral density. Let $\omega_j = j/(2n), j = 1, \ldots, n$ be the Fourier frequencies and denote by p the true spectral density of the process underlying the data. Then, asymptotically,

$$\frac{P(\omega_j)}{p(\omega_j)} \sim \begin{cases} \frac{1}{2}\chi_2^2, & \omega_j \neq 1/2, \\ \chi_1^2, & \omega_j = 1/2, \end{cases} \tag{15}$$

and the ratios $P(\omega_j)/p(\omega_j)$ are independent. Estimation of the log-spectral density $\mu = \log p$ is then viewed as a non-parametric regression problem, $y_j = \mu(\omega_j) + \varepsilon_j$, where $y_j = \log(P(\omega_j))$ and ε_j is distributed either as $\log(\chi_2^2)$ or $\log(\chi_1^2)$. Inference about the credible features of μ is based on the posterior density of its derivative which is found using an approach similar to the Posterior Smoothing method of Øigård et al. (2006) described in Section 3.2. However, since the noise ε_j is not Gaussian, a so-called simplified Laplace approximation method of Rue et al. is used for accurate and efficient computation of the credibility maps.

Olsen et al. (2008) use the spectral density to detect change points where the time series leaves one stationary state and enters another. What should be considered as change depends on the time scale considered which leads a scale space approach to the problem. A change at a time point is tested by computing the periodograms P_1 and P_2 in windows of length N to its left and right and then considering the ratios $P_2(\omega_j)/P_1(\omega_j)$, $j = 1, \ldots, N/2$, where $\omega_j = j/N$ is the jth Fourier frequency. The window length N has the role of the scale space smoothing parameter.
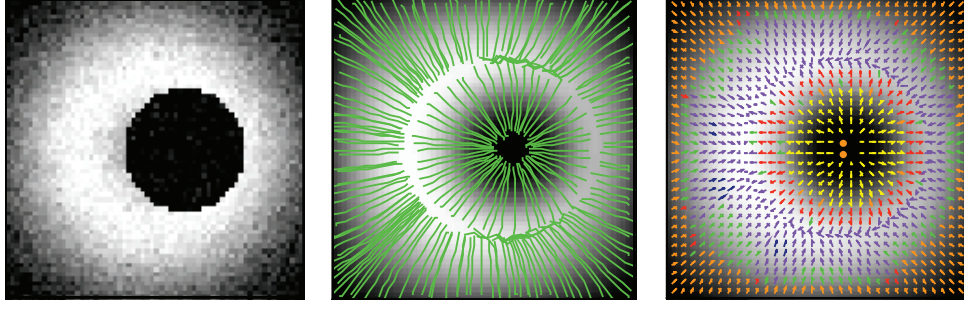
FIG 6. *Left: the noisy example image. Middle: streamlines computed from significant gradient directions. Right: arrows indicate significant gradient direction and different colors flag different types of significant local curvature as defined by the sign and significance of the eigenvalues $\kappa_{\lambda,1}$ and $\kappa_{\lambda,2}$: hole (yellow), long valley (orange), saddle point (red), long ridge (purple), and peak (dark blue). The example was obtained from* `http://www.unc.edu/~marron/marron.html`.

A Bayesian approach for finding credible features in discrete wavelet power spectra was described in Divine and Godtliebsen (2007). An i.i.d. Gaussian likelihood an a prior of the form (11) for the wavelet coefficients lead to a posterior that can be sampled exactly. Now the smoothing parameter in scale space analysis is the wavelet transform scale.

5. Higher dimensional settings and data on manifolds

5.1. S^3 *and its variants*

It is natural to consider the ideas developed for curves in a multivariate context. Godtliebsen et al. (2002) described a SiZer for two-dimensional densities. The features of a surface defined by a kernel density estimate of bi-variate data $\mathbf{x}_i \in \mathbb{R}^2$, $i = 1, \ldots, n$, were described in multiple scales using the gradient and the eigenvalues of the Hessian matrix. The methodology is referred to as S^3 for Significance in Scale Space and it was subsequently modified in Godtliebsen et al. (2004) for noisy digital images using a regression model. The inference and visualization methods are similar in these two 2D settings of SiZer and we describe next the latter in more detail.

Figure 6 shows an example of S^3 analysis where the underlying image is a volume of revolution of the Gaussian density with non-zero mean that has an off-centered cylinder removed. The noisy image is on the left and frames from feature analyses based on streamlines as well as significant gradients and curvatures are shown in the two other panels. These panels are in fact individual frames corresponding to a fixed scale of whole movies that runs through a range scales, visualizing features that appear in different image resolutions.

The model for a noisy image is

$$y_{ij} = \mu_{ij} + \varepsilon_{ij}, \ i = 1, \ldots, M, \ \ j = 1, \ldots, N, \tag{16}$$

where (i, j) is a pixel location in an $M \times N$ image. It is assumed that $\mu_{ij} = \mu(i, j)$, where μ is a smooth function of two continuous real variables and that the errors $\varepsilon_{i,j}$ are independent. Scale space smoothing is done using discrete Gaussian convolution with adjustment for boundary effects. Both the gradient (first order partial derivatives) and curvature (second order partial derivatives)

are needed for useful feature detection. Denote by $\mu_{\lambda,k}$ the kth partial derivative of the smooth μ_λ. The significance of the magnitude of the gradient is tested by

$$\frac{\hat{\mu}_{\lambda,1}(i,j)^2}{\hat{\sigma}_1^2} + \frac{\hat{\mu}_{\lambda,2}(i,j)^2}{\hat{\sigma}_2^2} > q_{\chi_2^2},$$

where the partials and their variances are estimated from a Gaussian kernel smooth of the observed noisy image and the χ^2-distribution is justified by the approximate bivariate normality of $[\hat{\mu}_{\lambda,1}(i,j), \hat{\mu}_{\lambda,2}(i,j)]^T$. However, no normal approximation is needed when the image noise is Gaussian. To explore curvature, denote by $\kappa_{\lambda,1} \leq \kappa_{\lambda,2}$ the two eigenvalues of the Hessian matrix of μ_λ. The null distribution of the test statistic $\max\{|\hat{\kappa}_{\lambda,1}(i,j)|, |\hat{\kappa}_{\lambda,2}(i,j)|\}$ is obtained by simulation using the estimated Hessian matrix. Holes, valleys, saddle points, ridges and peaks are flagged with different colors based on the significance and sign of the eigenvalues (Figure 6). Both for gradients and curvature, the idea of independent blocks of the original SiZer is applied to approximate simultaneous inference over the whole image.

The S^3 method was extended to images with spatially correlated noise in Vaughan et al. (2012). The covariance of the $\varepsilon_{i,j}$'s in (16) was assumed isotropic and S^3 was applied to analyze numerical climate model outputs. In Thon et al. (2012) a Bayesian version of S^3 was proposed. The method is called BMulA for Bayesian Multiscale Analysis and it can also be considered as a two-dimensional version of Posterior Smoothing (Godtliebsen and Øigård, 2005; Øigård et al., 2006) in that scale space smoothing is controlled by the image prior. The image model is again (16) but where the errors now are i.i.d. Gaussian with known variance. The image prior is an intrinsic Gaussian Markov random field and therefore the posterior is Gaussian which eliminates the need for simulation-based inference. However, the test statistic quantile needed for testing significance of curvature is still based on simulation. Computations are sped up by using toroidal image boundary conditions and Fourier methods. The method is applied to detecting hairs in dermatoscopic images of skin lesions.

Ganguli and Wand (2004) extended S^3 for geostatistical data. The model is

$$y_i = \mu(\mathbf{x}_i) + \varepsilon_i, \ i = 1\ldots,n, \tag{17}$$

where y_i is a scalar response at a geographical location $\mathbf{x}_i$, μ is a smooth function and the ε_i's are i.i.d. Gaussian. The data $\mathbf{x}_i$ are not equispaced and contain regions of data sparsity and therefore the setting is different from that of Godtliebsen et al. (2004). A mixed-model approach is taken where $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}$, $\boldsymbol{\mu} = [\mu(\mathbf{x}_1),\ldots,\mu(\mathbf{x}_n)]^T$, $\mathbf{X}$ and $\mathbf{Z}$ are a bi-variate polynomial and a low-rank thin plate spline design matrix, respectively, $\boldsymbol{\beta}$ is a fixed parameter vector and $\mathbf{u}$ is a zero-mean multivariate Gaussian random variable. The scale space smoothing parameter is $\lambda = \sigma_\varepsilon^2/\sigma_u^2$. A spline-based smoother offers a computationally efficient estimator in regions of data sparsity. Further, the mixed-model approach allows one to use distributional theory and software that already exists for such models. Simultaneous inference over the whole spatial grid can be done based on the asymptotic distribution theory of linear mixed models. As in Godtliebsen et al. (2004), the significant features in the spatial field are explored using gradient and curvature. In Ganguli and Wand (2007) the authors extend their approach to discrete responses using generalized additive models.

In Pedersen et al. (2008) the S^3 methodology was adapted for detection of differences between pairs of images to analyze the difference between numerical climate model output and remote sensing observations of global surface albedo climatology.

5.2. iBSiZer

In most of the above methods, finding significant features in two-dimensional data is based on inference about the first or second partial derivatives. In contrast, iBSiZer (Bayesian SiZer for images) and the related multiresolution BSiZer are descendants of BSiZer (Section 3.1) that consider the pixel intensities as such (Holmström and Pasanen, 2012; Holmström et al., 2011). In iBSiZer the model for an $M \times N$ image is again (16). Let $n = MN$ and denote by $\boldsymbol{\mu} = [\mu_1, \ldots, \mu_n]^T$ the vectorized underlying true image. The model can then be written as

$$\mathbf{y} = \boldsymbol{\mu} + \boldsymbol{\varepsilon}, \tag{18}$$

where, in the simplest case, $\boldsymbol{\varepsilon} \sim \mathrm{N}(\mathbf{0}, \sigma^2\mathbf{I})$ but other covariance structures such as isotropic or heteroscedastic noise can also be considered. Given a smoothing matrix $\mathbf{S}_\lambda$, iBSiZer makes inferences about the credible features of $\mathbf{S}_\lambda\boldsymbol{\mu}$ for a range of smoothing parameters λ. This is done by analyzing the the posterior distribution $p(\mathbf{S}_\lambda\boldsymbol{\mu}|\mathbf{y})$. The method is typically used for finding credible differences between two images $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$. When the associated errors $\boldsymbol{\varepsilon}_1$ and $\boldsymbol{\varepsilon}_2$ are independent, the difference image $\boldsymbol{\mu} = \boldsymbol{\mu}_2 - \boldsymbol{\mu}_1$ and its observed version $\mathbf{y} = \mathbf{y}_2 - \mathbf{y}_1$ can still be assumed to satisfy (18). Since smoothing amounts to local averaging of image pixel intensities, iBSiZer can be interpreted to provide information about the salient features of the truth $\boldsymbol{\mu}$ in different spatial scales or image resolutions. In the i.i.d. error model, inverse-χ^2 prior is assumed for σ^2 and for $\boldsymbol{\mu}$ one can use a smoothing Gaussian Markov random field prior

$$p(\boldsymbol{\mu}|\sigma^2, \lambda_0) \propto \left(\frac{\lambda_0}{\sigma^2}\right)^{(n-n_0)/2} \exp\left(-\frac{1}{2}\frac{\lambda_0}{\sigma^2}\boldsymbol{\mu}^T\mathbf{Q}\boldsymbol{\mu}\right),$$

where $\mathbf{Q}$ is a positive semidefinite matrix with null space dimension n_0. The scale space smoother is $\mathbf{S}_\lambda = (\mathbf{I} + \lambda\mathbf{Q})^{-1}$. Just as in BSiZer, $\lambda_0 > 0$ defines the level of smoothing when the signal is reconstructed from a noisy observation. Other, application-specific priors can also be used. Inference is based on posterior sampling and Fourier methods speed up computations.

Figure 7 shows iBSiZer analysis of a difference of two Landsat ETM+ satellite images taken over the same location in eastern Finland at two different times. The credibly positive and negative areas in the difference image are shown in blue and red, respectively. Gray is used for pixels that that do not differ credibly from 0. Inference is simultaneous over the images. See Holmström and Pasanen (2012) for more information on this example. In Pasanen and Holmström (2015) iBSiZer was extended for the analysis non-linear transformations of multidimensional satellite images such as the Normalized Difference Vegetation Index (NDVI) used in remote sensing.

In a sense, in iBSiZer analysis at smoothing level λ_i all the features corresponding to the scales $\lambda \geq \lambda_i$ are still present in the image. By considering differences of smooths, the multiresolution approach (Holmström et al., 2011) attempts to remedy this by isolating, for two levels $\lambda_i < \lambda_j$, those features that are present at level λ_i but not at λ_j. One way to think about the difference between conventional scale space analysis, such as iBSiZer, and the difference of smooths idea is that the former uses low pass filtering whereas the latter uses band pass filtering. Let $0 = \lambda_1 < \lambda_2 < \cdots < \lambda_{L-1} < \lambda_L = \infty$ be a set of increasing smoothing levels. We assume that $\lambda = 0$ corresponds to no smoothing so that $\mathbf{S}_0\boldsymbol{\mu} = \boldsymbol{\mu}$ and that $\mathbf{S}_\infty\boldsymbol{\mu}$ is the mean of $\boldsymbol{\mu}$. In BSiZer multiresolution analysis (MRBSiZer), inference is made about the credible features in the differences $\mathbf{z}_i = \mathbf{S}_{\lambda_i}\boldsymbol{\mu} - \mathbf{S}_{\lambda_{i+1}}\boldsymbol{\mu}$,

$$\lambda = 0 \qquad \lambda = 1 \qquad \lambda = 50 \qquad \lambda = 400 \qquad \lambda = 50000$$
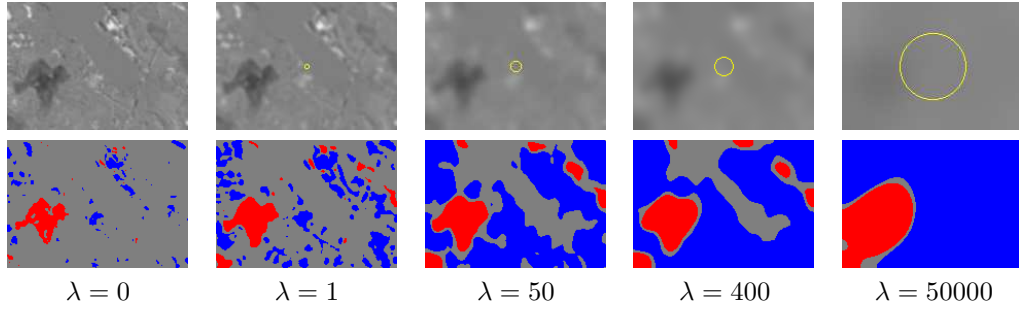
FIG 7. *First row: In the first panel is the posterior mean $\mathbb{E}(\boldsymbol{\mu}|\mathbf{y})$ of the difference of two Landsat ETM+ satellite images taken over eastern Finland on August 2, 1999 and May 29, 2002. The other four panels are posterior means $\mathbb{E}(\mathbf{S}_\lambda \boldsymbol{\mu}|\mathbf{y})$ corresponding to different levels of smoothing. The effective size of the smoother is indicated by the yellow circle. Second row: credibility maps that show the salient features in different spatial resolutions.*
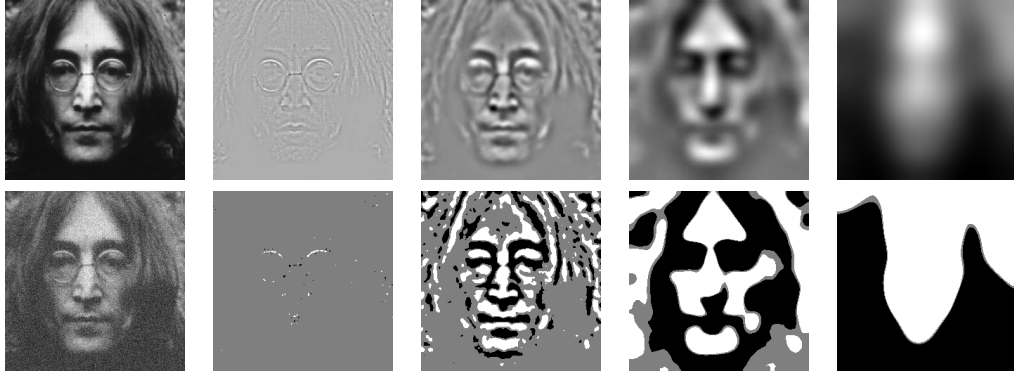

FIG 8. *iBSiZer scale space multiresolution analysis of an image of John Lennon. First row: the underlying true image and the mean components $\mathbb{E}(\mathbf{z}_i|\mathbf{y})$, $i = 1, \ldots, 4$. The overall mean image has been omitted. Second row: the noisy observed image and the credibility maps of the components. White and black indicate the credibly positive and negative pixels and gray pixels are not credibly different from zero.*

$i = 1, \ldots, L-1$ and $\mathbf{z}_L = \mathbf{S}_\infty \boldsymbol{\mu}$. The term multiresolution refers to the expansion $\boldsymbol{\mu} = \sum_{i=1}^{L} \mathbf{z}_i$ which resolves $\boldsymbol{\mu}$ into scale dependent components. Graphical and numerical tools are available for selecting an appropriate sequence of smoothing levels λ_i. Figure 8 shows an example an MRBSiZer multiresolution decomposition using $L = 5$ components. On the first row is the underlying true image and four image multiresolution components displayed as posterior means $\mathbb{E}(\mathbf{z}_i|\mathbf{y})$. The mean $\mathbf{z}_5$ has been left out. In the second row are the noisy observed image and credibility maps for the multiresolution components. Here white and black indicate the credibly positive and negative pixels and gray pixels are not credibly different from zero. A scale space technique based on differences of smooths was also proposed in Marvel et al. (2013). Like in Holmström et al. (2011), the method is used to analyze numerical climate model outputs in multiple scales by applying smoothing on a sphere.

5.3. Beyond 2D, non-linear manifolds

In principle, the inference part of many statistical scale space methods can be extended to multivariate random fields that depend on more than two covariates. This is for example clearly the

case for iBSiZer and gradient-based S^3 methods. However, visualization of inference results is an integral part of scale space data analysis and creating significance "maps" in contexts more complex than trivariate probability densities or random fields seems difficult. In a multivariate regression setting, one solution is to apply additive modelling and perform scale space analysis on each additive component separately (Miranda et al., 2002; Martínez-Miranda, 2005).

In Duong et al. (2008) scale space analysis of densities is considered in any dimension. Such curvature related features of surfaces as ridges and saddle points defined by second derivatives do not have clear counterparts in dimensions higher than two. The authors therefore focus on the significance of modes, that is, the local maxima of a multivariate density. Both the gradient and the Hessian matrix are used and while the approach resembles S^3, the hypothesis tests applied are not direct generalizations of Godtliebsen et al. (2002). Theoretical properties of the required kernel estimators are provided and visualization methods are developed for one, two and three dimensional data.

One could also try to approach higher dimensional problems through dimension reduction techniques such as principal component analysis or methods developed specifically for non-linear structures. Interestingly, scale space methods may be useful here, too. Wang and Marron (2008) proposed a scale space method for finding the effective dimensions in noisy data. The data are viewed within a spherical local neighborhood whose radius has the role of the scale space smoothing parameter. Noise can mask true underlying low-dimensional structure which becomes evident only after sufficiently large scales are considered. Given a set of noisy points sampled from a density defined on a submanifold of the euclidean space, the authors develop statistical tests to determine the effective manifold dimension in different scales. The tests are based on the local geometry of sample points and their theoretical properties are also considered.

Directional data that lie on a circle or on a sphere are common in various applications. Holmström et al. (2011), Vaughan et al. (2012) and Marvel et al. (2013) considered scale space analysis on a sphere (the globe). A version of SiZer for circular data, CircSiZer, was proposed in Oliveira et al. (2014). Both density and regression estimation were considered. Smoothing is done using the von Mises kernel and the quantiles as well the variance estimate in regression needed for significance tests are obtained using the bootstrap. Examples of CircSiZer maps are shown in Figure 9. Weather data from the atlantic coast of Galicia (NW Spain) are analyzed. In the left panel, density estimation is considered for wind directions and in the right panel regression of wind speed on wind direction is analyzed. The different colors are used in the same way as in the original SiZer (Section 2.1). The colors along a circle of a fixed radius indicate where the underlying function of interest is increasing or decreasing, with the radius corresponding to the level of smoothing applied. For more information on this example, see Oliveira et al. (2014). In Huckemann et al. (2014) it was pointed out that smoothing with the von Mises kernel does not satisfy scale space causality for all levels of smoothing and the authors proposed WSiZer (Wrapped SiZer) where the wrapped Gaussian kernel is used, instead.

6. Variations: scale, derivatives, maps

In formal scale space theory scale is defined by the variance of the Gaussian blurring kernel. In statistical scale space practice the concept of "scale" is often interpreted much more liberally simply as some particular component of a statistical model whose explorative variation helps to reveal
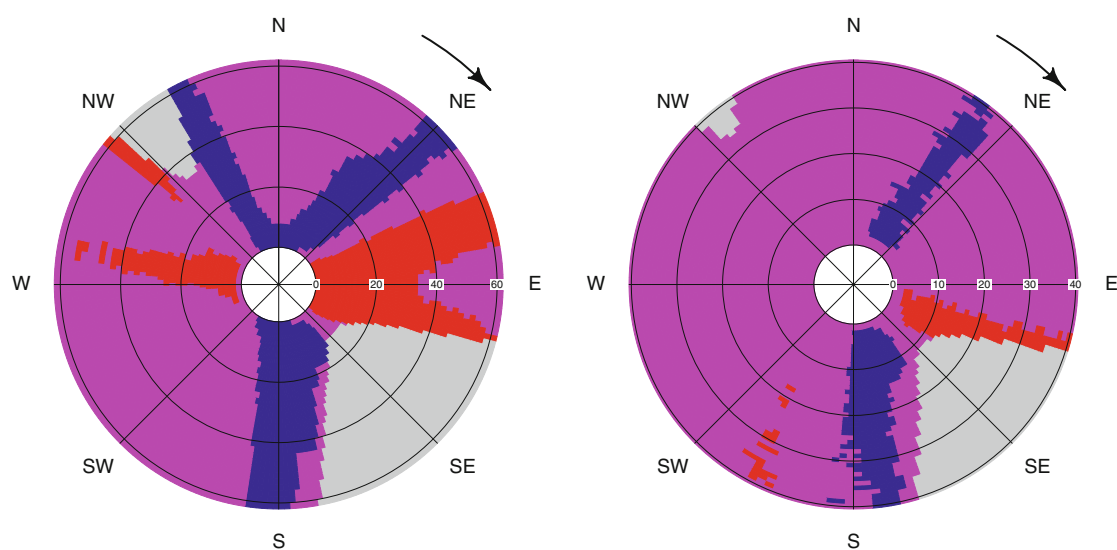
Fig. 9 CircSiZer map for kernel density estimator (*left*) for wind direction and CircSiZer map for circular–linear regression (*right*) for wind speed (m/s) with respect wind direction. For reading CircSiZer, take clockwise sense of rotation. Values of *ν* are indicated along the radius

FIG 9. *CircSizer analysis of weather data from the atlantic coast of Galicia (NW Spain). The original Figure 9 of Oliveira et al. (2014) is used with kind permission from Springer Science and Business Media. The colors along a circle of a fixed radius indicate where the underlying function of interest is increasing or decreasing with the radius (here referred to as ν) corresponding to the level of smoothing applied.*

the salient features of the phenomenon underlying the observed data. Temporal and spatial scales are the most typical choices but we have seen in the previous sections several examples of other interpretations of scales, too.

Regularization often has a smoothing effect on a regression estimate and can therefore be given a scale space interpretation. Boysen et al. (2007) studied the asymptotic behavior of piecewise constant least squares regression estimates (regressograms), when the number of jumps in the estimate is penalized. The penalty is a regularization parameter that the authors propose to use for scale space smoothing. Li and Xu (2015) propose a linear regression spline SiZer where the number of spline knots serves as the scale parameter. Fiducial inference is used and multiple testing adjustment is made to control the false discovery rate in the SiZer map. The LASSO is a linear regression method where the least squares sum is penalized by the absolute sum of the regression coefficients, thus enabling model fitting even when there are more explanatory variables than observations (Tibshirani, 1996). An estimate of the regression coefficient vector is obtained by minimizing a penalized least squares loss,

$$\boldsymbol{\beta}_\lambda = \underset{\boldsymbol{\beta}}{\mathrm{argmin}} \ \left\{ \|\boldsymbol{\beta}\mathbf{X} - \mathbf{y}\|_2^2 + \lambda\|\boldsymbol{\beta}\|_1 \right\},$$

where $\|\cdot\|_p$ is the L_p-norm. The regularization parameter λ can be viewed as a scale which motivated Pasanen et al. (2014) to propose a scale space view of the LASSO and the Bayesian LASSO. Hindberg (2012) introduced a test of multinormality based on weighted sums over consecutive dimensions along a data vector. Here the number of dimensions involved in the sum serves as a scale parameter. The technique can also be used to develop a test for the k-sample problem. Like the LASSO, both of Hindberg's methods can be applied in situations where the number of data is much smaller than the number of covariates.

Dümbgen and Walther (2008) made inferences about the slope of a univariate density (or a failure rate) by considering all intervals whose endpoints are observations. The length of such intervals can be thought of as a scale space parameter, although no smoothing is involved. Multiple hypotheses about density features across different scales can be tested simultaneously. For related theoretical work, see Dümbgen and Spokoiny (2001) and Rufibach and Walther (2010).

Yet another example of how to think of a scale is the thick pen transform of Fryzlewicz and Oh (2011) designed for the exploration of dependence structure of a time series. The technique is motivated by the scale space idea and the role of a scale is played by the thickness of the pen used to plot the time series.

In addition to the different definitions of "scale", there are various approaches to detecting interesting features in the data. The original SiZer (Chaudhuri and Marron, 1999) mostly concentrated on applying the first derivative although the possibility to use the second derivative or no derivative at all was also mentioned. The mathematical analysis of Chaudhuri and Marron (2000) in fact considered derivatives of arbitrary order. The second derivative version of SiZer has later been referred to as SiCon (Significance of Convexity). Chaudhuri and Marron (2002) compared SiZer and SiCon giving examples where the second derivative can detect structure better than the first derivative. One such case is the analysis of change points in a signal because a jump change corresponds to a local maximum of the first derivative, hence a local zero crossing of the second derivative. Often the use of the second derivative appears to facilitate the detection of interesting features of a curve that otherwise would be masked in the first derivative analysis by an overall

strong increase or decrease of the curve.

S^3 (Section 5.1) uses second derivatives to find features in two-dimensional data but iBSiZer and its multiresolution version (Section 5.2), the S^3 adaptation of Pedersen et al. (2008), as well the methods developed by C. Park and his co-workers for comparing curves (Section 2.3), are examples of approaches that use no differentiation at all. An alternative use of differentiation was proposed in Pasanen et al. (2013) where the derivative is computed with respect to the scale. Like in multiresolution BSiZer, the goal is to decompose an observed time series $\mathbf{y}$ into credible scale-dependent components using differences of smooths but the method is useful as an exploratory tool even without any associated inference. In this case one defines the so-called scale-derivative of $\mathbf{y}$ as $\mathbf{D}_\lambda\mathbf{y} = (d/d(\log\lambda))\mathbf{S}_\lambda\mathbf{y}$, where $\mathbf{S}_\lambda$ is a smoother. The local minima of the function $\lambda \mapsto \|\mathbf{D}_\lambda\mathbf{y}\|$ then define effective multiresolution smoothing levels. An example is shown in Figure 10, where the observed time series is a sum of three sine waves with wavelengths $\pi/3$, $2\pi/3$, 2π, and Gaussian noise. The usefulness of the scale-derivative in finding interesting features of a time series can be understood by considering the heat equation. Thinking of $\mathbf{S}_\lambda\mathbf{y}$ as a discrete-time estimate of a smooth μ_λ of an underlying function μ and assuming that the smoother approximates Gaussian convolution, the heat equation suggests that

$$\frac{\partial\mu_\lambda(x)}{\partial(\log\lambda)} = \lambda\frac{\partial\mu_\lambda(x)}{\partial\lambda} \propto \lambda\frac{\partial^2\mu_\lambda(x)}{\partial x^2}$$

holds approximately. Therefore, the zero-crossings of the scale-derivative are zero-crossings of the second time derivative of the underlying signal, places where its rate of change is largest. Compared to SiZer and BSiZer, the method consequently has the same benefits as SiCon.

Finally, given the great variety of available alternatives to statistical scale space analyses, including the many approaches outlined in the previous sections and the variations described above, how should one judge the performance of the different techniques? An interesting solution proposed by Hannig et al. (2013) is to compare the SiZer-type maps produced by different methods using distance metrics derived from ideas originally developed for digital images. Univariate regression is considered and one computes the distance of a map produced by a variant of SiZer to an "oracle" SiZer map based on known regression and variance functions. The smaller the distance, the better the performance of the SiZer-type method is deemed to be.
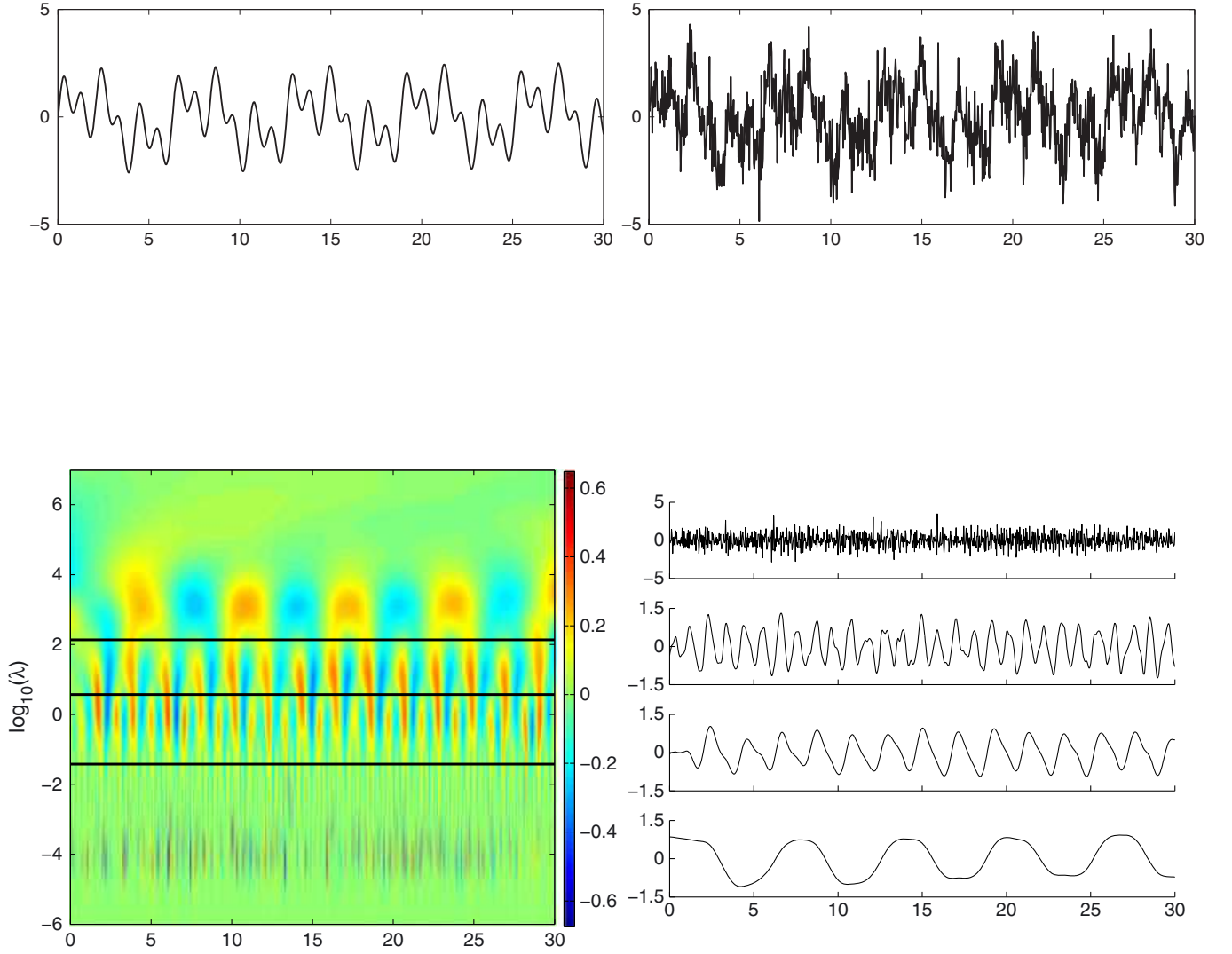
FIG 10. *Upper panels:* $\sin(x) + \sin(3x) + \sin(6x)$ *and its noisy version. Lower left: the scale-derivative map of the noisy signal with the minima of* $\lambda \mapsto \|\mathbf{D}_\lambda \mathbf{y}\|$ *indicated by the black horizontal lines. Lower right: the noise and three scale-dependent components extracted as differences of smooths where the smoothing levels used correspond to the black horizontal lines in the scale-derivative map. The mean of the signal is omitted. The figures originally appeared in Pasanen et al. (2013).*

Application	Methods	References
Internet traffic		
Detection of anomalies, non-stationarity	MRAD, SiZer, SiNos	Zhang et al. (2008, 2014); Park et al. (2007)
Traffic data modeling	(Dependent) SiZer	Park et al. (2004); Marron et al. (2004); Park et al. (2006); Hernandez-Campos et al. (2005); Park et al. (2005)
Medical technology and health care		
Medical imaging	(Dependent) SiZer, BMulA, S^3, MFD	Skrøvseth et al. (2010); Møllersen et al. (2010); Thon et al. (2012); Godtliebsen et al. (2004); Godtliebsen and Øigård (2005); Park et al. (2010a); Poline and Mazoyer (1994); Worsley et al. (1996)
Health monitoring, disease diagnosis	c-SiZer, SiZer, SiNos, SiCon	Skrøvseth et al. (2012b,a); Skrøvseth and Godtliebsen (2011); Mortensen et al. (2006); Rooper et al. (2014); Jacobs et al. (2012); Skrøvseth et al. (2012c); Skrøvseth et al. (2012)
Flow cytometry, genetic and cellular data	SiZer, SiCon, Bayesian LASSO, WSiZer	Zeng et al. (2002); Salganik et al. (2005); Duong et al. (2008); Barker et al. (2008); Vekemans et al. (2012); Bekaert et al. (2002); Zhao et al. (2004); Hannig and Marron (2006); Pasanen et al. (2014); Huckemann et al. (2014)
Earth sciences		
Climatology, meteorology	SiZer, BSiZer, S^3, SiCon, SiNos, Post. Smoothing, MRBSiZer, CircSiZer	Korhola et al. (2000); Holmström and Erästö (2002); Erästö and Holmström (2005, 2006, 2007); Weckström et al. (2006); Erästö et al. (2012); Pedersen et al. (2008); Rohling and Pälike (2005) Sørbye et al. (2009); Olsen et al. (2007, 2008); Holmström et al. (2011); Divine et al. (2009); Rydén (2010); Oliveira et al. (2014); Vaughan et al. (2012), Marvel et al. (2013)
Oceanograhy, glaciology	Wavelets, SiZer, SiCon, BSiZer	Divine and Godtliebsen (2007); Abram et al. (2013); Divine et al. (2005); Zamelczyk et al. (2014); Aagaard-Sørensen et al. (2014); Bolton et al. (2010); Miettinen et al. (2012); Justwan et al. (2008); McClymont et al. (2013); Wilson et al. (2011); Godtliebsen et al. (2012); Korsnes et al. (2002)
Geology and geophysics	Post. Smoothing, SiZer	Øigård et al. (2006); Chen et al. (2005); Rudge (2008); Condie and Aster (2010)
Ecology	SiZer, SiCon	Weis et al. (2014); Sonderegger et al. (2008)

TABLE 1
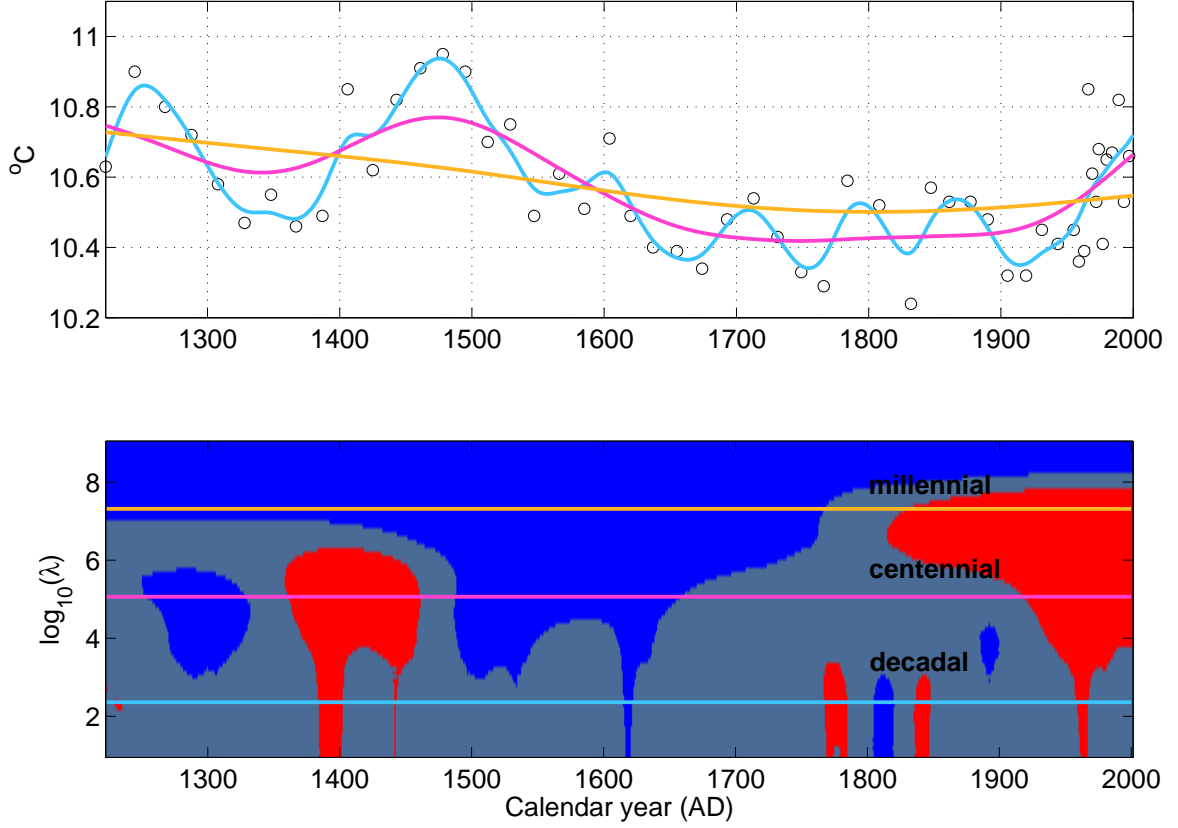*Applications of statistical scale space methods.*

FIG 11. *BSiZer analysis of a diatom-based temperature reconstruction in northern Fennoscandia for the last 800 years. The upper panel: the reconstructed mean July air temperatures shown as circles together with three smooths obtained as posterior means. The lower panel: BSiZer map with the three smoothing levels shown using the same colors as in the upper panel. As opposed to the original SiZer (e.g. Figure 2), here blue and red signify credibly negative and positive slope, respectively. The credibility threshold $\alpha = 0.8$ was used. Gray indicates a non-credible slope. For details see Weckström et al. (2006).*

7. Applications

The applications tackled with statistical scale space methods are too numerous to permit anything close to an exhaustive and detailed overview. Most of the papers reviewed in the preceding sections include analyses of real data, often comparing a proposed novel scale space technique with an earlier analysis of the same data. Beyond such performance comparisons and analyses of standard benchmark examples, there are a few of research fields that appear to be particularly amenable to a scale space approach and we have outlined them in Table 1. Interesting applications of scale space techniques can be found both to technological and scientific problems, sometimes in combination with more traditional statistical methods. For each application, Table 1 lists the scale space methods used together with some example references.

Analysis of Internet traffic data has been a popular topic. Park et al. (2004, 2007) and Zhang

et al. (2008, 2014) discussed in Section 4 all feature such data analyses and additional references are given in Table 1. Medical imaging applications have considered detection of melanoma as well as analysis of fMRI and PET images. The multifiltering detection (MFD) strategy of Poline and Mazoyer (1994) and associated work by Worsley et al. (1996) are included in the table since they work in the same spirit as the methods reviewed in the present article in that multiple smooths of a PET image are used for making inferences about interesting features. Scale space methods have also been used in monitoring type 1 and type 2 diabetes patients. Other monitoring and diagnostic applications include circulatory research, diagnosing of pleural effusion, analysis of the dependence of influenza risk on age and the incidence of whooping cough, as well as the use of accelerometer data for early detection of chronic obstructive pulmonary disease. Other applications to medical technology include analyses of flow cytometry, cellular and genetic data.

In basic scientific research, earth sciences have turned out to be a fertile ground for scale space methods. For example, climate science often deals with time series and data sets for which questions relating to temporal or spatial scales arise naturally. One particularly popular area has been the analysis of reconstructions of the paleoenvironment. Past temperatures going back hundreds or thousands of years can be inferred from lake or sea sediment cores and even longer time scales can be explored using ice cores. Figure 11 shows a lake sediment core based BSiZer analysis of the reconstructed mean July temperatures for the past 800 years in northern Fennoscandia. The large scale trend shows a decreasing temperature but in the centennial scale the Little Ice Age (about from A.D. 1550 to 1850) shows as a broad credible minimum. Many other scale space analyses of past and predicted climate variation are included in Table 1. Other work on earth sciences include applications to meteorology, oceanography, research on ice, earthquakes, and ecology.

Outside of the above main areas of applications, SiZer has been used also for instance in astronomy (Park et al., 2011), electrophysiology (Roca-Pardiñas et al., 2011), experimental psychology (Becker and Elliott, 2006), and economics (Pittau and Zelli, 2007; Marron and de Uña-Álvarez, 2004). Ganguli and Wand (2004) analyzed fishery data using S^3 for geostatistical data and Godtliebsen and Øigård (2005) used Posterior Smoothing in a pattern recognition application to detect bones in fish fillets. Finally, Holmström and Pasanen (2012) and Pasanen and Holmström (2015) proposed to use iBSiZer in satellite-based remote sensing.

For those interested in trying scale space methods in their own applications, software is available in the public domain. Matlab implementations include

- `http://www.unc.edu/~marron/marron_software.html`
- `http://www.stat.nus.edu.sg/~zhangjt/SubPages/SiZerSS/software.htm`
- `http://www.cs.helsinki.fi/u/kohonen/sizer/`

R packages are also available,

- `http://CRAN.R-project.org/package=SiZer`
- `http://CRAN.R-project.org/package=feature`
- `http://CRAN.R-project.org/package=NPCirc`

Matlab implementations of many of the Bayesian approaches discussed in this article can be obtained at

- `http://cc.oulu.fi/~lpasanen/`
- `http://mathstat.helsinki.fi/bsizer/`

- `http://www2.telemed.no/kevin/`

8. Summary

During the last 15 years a great variety of statistical scale space methods have been developed and they have found applications in many areas of research. The popularity of the scale space view can perhaps be explained by the highly intuitive and transparent way it deals with scales in the data, usually through smoothing while sometimes employing another natural idea of a scale. For non-statisticians, the often strikingly effective graphical summaries of scale space analyses are very appealing as they greatly facilitate the interpretation of the inferences about the salient features in the data. In the future, the toolbox of scale methods will undoubtedly keep expanding and we look forward to new applications both in the areas already well covered as well as in other fields of research where the concept of scale plays an important role in the analysis of data.

References

Aagaard-Sørensen, S., Husum, K., Hald, M., Marchitto, T., and Godtliebsen, F. (2014). Sub sea surface temperatures in the Polar North Atlantic during the Holocene: Planktic foraminiferal Mg/Ca temperature reconstructions. *The Holocene*, 24(1):93–103.

Abram, N. J., Mulvaney, R., Wolff, E. W., Triest, J., Kipfstuhl, S., Trusel, L. D., Vimeux, F., Fleet, L., and Arrowsmith, C. (2013). Acceleration of snow melt in an Antarctic Peninsula ice core during the twentieth century. *Nature Geoscience*, 6(5):404–411.

Azzalini, A. and Bowman, A. W. (1990). A Look at Some Data on the Old Faithful Geyser. *Applied Statistics*, 39(3):357–365.

Babaud, J., , Baudin, M., Witkin, A. P., and Duda (1983). Uniqueness of the Gaussian kernel for scale-space filtering. Fairchild TR 645, June 1983.

Babaud, J., Witkin, A. P., Baudin, M., and Duda, R. O. (1986). Uniqueness of the Gaussian Kernel for Scale-Space Filtering. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, PAMI-8(1):26–33.

Barker, M. S., Kane, N. C., Matvienko, M., Kozik, A., Michelmore, R. W., Knapp, S. J., and Rieseberg, L. H. (2008). Multiple paleopolyploidizations during the evolution of the Compositae reveal parallel patterns of duplicate gene retention after millions of years. *Molecular Biology and Evolution*, 25(11):2445–2455.

Becker, C. and Elliott, M. A. (2006). Flicker-induced color and form: Interdependencies and relation to stimulation frequency and phase. *Consciousness and cognition*, 15(1):175–196.

Bekaert, S., Koll, S., Thas, O., and Van Oostveldt, P. (2002). Comparing telomere length of sister chromatids in human lymphocytes using three-dimensional confocal microscopy. *Cytometry*, 48(1):34–44.

Bolton, C. T., Wilson, P. A., Bailey, I., Friedrich, O., Beer, C. J., Becker, J., Baranwal, S., and Schiebel, R. (2010). Millennial-scale climate variability in the subpolar North Atlantic Ocean during the late Pliocene. *Paleoceanography*, 25(4).

Boysen, L., Liebscher, V., Munk, A., and Wittich, O. (2007). *Scale space consistency of piecewise constant least squares estimators – another look at the regressogram*, volume 55 of *Lecture Notes–Monograph Series*, pages 65–84. Institute of Mathematical Statistics, Beachwood, Ohio, USA.

Chaudhuri, P. and Marron, J. (2002). Curvature vs. slope inference for features in nonparametric curve estimates. *Unpublished manuscript*.

Chaudhuri, P. and Marron, J. S. (1999). SiZer for exploration of structures in curves. *Journal of the American Statistical Association*, 94(447):807–823.

Chaudhuri, P. and Marron, J. S. (2000). Scale space view of curve estimation. *The Annals of Statistics*, 28(2):408–428.

Chen, S.-j., Jia, Q.-h., and Ma, L. (2005). SiZer for exploration of inhomogeneous structure in temporal distribution of earthquakes. *Acta Seismologica Sinica*, 18(5):572–581.

Condie, K. C. and Aster, R. C. (2010). Episodic zircon age spectra of orogenic granitoids: The supercontinent connection and continental growth. *Precambrian Research*, 180(3-4):227 – 236.

Divine, D., Isaksson, E., Kaczmarska, M., Godtliebsen, F., Oerter, H., Schlosser, E., Johnsen, S., Van Den Broeke, M., and Van De Wal, R. (2009). Tropical Pacific–high latitude south Atlantic teleconnections as seen in δ^{18}O variability in Antarctic coastal ice cores. *Journal of Geophysical Research: Atmospheres (1984–2012)*, 114(D11).

Divine, D. V. and Godtliebsen, F. (2007). Bayesian modeling and significant features exploration in wavelet power spectra. *Nonlinear Processes in Geophysics*, 14:79–88.

Divine, D. V., Korsnes, R., Makshtas, A. P., Godtliebsen, F., and Svendsen, H. (2005). Atmospheric-driven state transfer of shore-fast ice in the northeastern Kara Sea. *Journal of Geophysical Research: Oceans (1978–2012)*, 110(C9).

Dümbgen, L. and Spokoiny, V. G. (2001). Multiscale testing of qualitative hypotheses. *The Annals of Statistics*, 29(1):124–152.

Dümbgen, L. and Walther, G. (2008). Multiscale inference about a density. *The Annals of Statistics*, 36(4):1758–1785.

Duong, T., Cowling, A., Koch, I., and Wand, M. (2008). Feature significance for multivariate kernel density estimation. *Computational Statistics & Data Analysis*, 52(9):4225 – 4242.

Erästö, P. and 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. and Holmström, L. (2006). Selection of prior distributions and multiscale analysis in Bayesian temperature reconstructions based on fossil assemblages. *Journal of Paleolimnology*, 36(1):69–80.

Erästö, P. and 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.

Erästö, P., Holmström, L., Korhola, A., and Weckström, J. (2012). Finding a consensus on credible features among several paleoclimate reconstructions. *Annals of Applied Statistics*, 6(4):1377–1405. Available on-line at `http://dx.doi.org/10.1214/12-AOAS540`.

Fryzlewicz, P. and Oh, H.-S. (2011). Thick pen transformation for time series. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(4):499–529.

Ganguli, B. and Wand, M. (2004). Feature significance in geostatistics. *Journal of Computational and Graphical Statistics*, 13(4):954–973.

Ganguli, B. and Wand, M. (2007). Feature significance in generalized additive models. *Statistics and Computing*, 17(2):179–192.

Ghosh, A., Chaudhuri, P., and Murthy, C. (2005). On visualization and aggregation of nearest neighbor classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(10):1592–1602.

Ghosh, A., Chaudhuri, P., and Murthy, C. (2006a). Multi-scale classification using nearest neighbor density estimates. *IEEE Transactions on Systems, Man and Cybernetics, Part B*, 36(5):1139–1148.

Ghosh, A., Chaudhuri, P., and Sengupta, D. (2006b). Classification using kernel density estimates: Multiscale analysis and visualization. *Technometrics*, 48(1):120–132.

Godtliebsen, F., Holmström, L., Miettinen, A., Erästö, P., Divine, D. V., and Koc, N. (2012). Pairwise Scale-Space Comparison of Time Series with Application to Cli mate Research. *Journal of Geophysical Research*, 117, C03046. Available on-line at http://dx.doi.org/10.1029/2011JC007546.

Godtliebsen, F., Marron, J. S., and Chaudhuri, P. (2002). Significance in scale space for bivariate density estimation. *Journal of Computational and Graphical Statistics*, 11(1):1–21.

Godtliebsen, F., Marron, J. S., and Chaudhuri, P. (2004). Statistical significance of features in digital images. *Image and Vision Computing*, 22(13):1093–1104.

Godtliebsen, F. and Øigård, T. (2005). A visual display device for significant features in complicated signals. *Computational Statistics and Data Analysis*, 48(2):317–343.

González-Manteiga, W., Martínez-Miranda, M., and Raya-Miranda, R. (2008). SiZer map for inference with additive models. *Statistics and Computing*, 18(3):297–312.

Hannig, J. and Lee, T. (2006). Robust SiZer for exploration of regression structures and outlier detection. *Journal of Computational and Graphical Statistics*, 15(1):101–117.

Hannig, J., Lee, T., and Park, C. (2013). Metrics for SiZer map comparison. *Stat*, 2(1):49–60.

Hannig, J. and Marron, J. S. (2006). Advanced distribution theory for SiZer. *Journal of the American Statistical Association*, 101(474):484–499.

Härdle, W. (1991). *Smoothing Techniques: With Implementation in S*. Springer Series in Statistics. Springer.

Hernandez-Campos, F., Jeffay, K., Park, C., Marron, J., and Resnick, S. I. (2005). Extremal dependence: Internet

traffic applications. *Stochastic Models*, 21(1):1–35.

Hindberg, K. (2012). *Scale-Space Methodology Applied to Spectral Feature Detection, Multinormality Testing and the k-sample Problem, and Wavelet Variance Analysis*. PhD thesis, University of Tromsø.

Holmström, L. (2010a). BSiZer. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(5):526–534. Available on-line 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 on-line at http://dx.doi.org/10.1002/wics.79.

Holmström, L. and Erästö, P. (2002). Making inferences about past environmental change using smoothing in multiple time scales. *Computational Statistics & Data Analysis*, 41(2):289–309.

Holmström, L. and Pasanen, L. (2012). Bayesian scale space analysis of differences in images. *Technometrics*, 54(1):16–29. Available on-line at http://dx.doi.org/10.1080/00401706.2012.648862.

Holmström, L., Pasanen, L., Furrer, R., and Sain, S. R. (2011). Scale space multiresolution analysis of random signals. *Computational Statistics & Data Analysis*, 55(10):2840 – 2855. Available on-line at http://dx.doi.org/10.1016/j.csda.2011.04.011.

Huckemann, S. F., Kim, K.-R., Munk, A., Rehfeld, F., Sommerfeld, M., Weickert, J., and Wollnik, C. (2014). The circular SiZer, inferred persistence of shape parameters and application to stem cell stress fibre structures. Available as arXiv:1404.3300v1.

Iijima, T. (1959). Basic theory of pattern observation. *Papers of Technical Group on Automata and Automatic Control, IECE, Japan, Dec. 1959.* In Japanese.

Jacobs, J. H., Archer, B. N., Baker, M. G., Cowling, B. J., Heffernan, R. T., Mercer, G., Uez, O., Hanshaoworakul, W., Viboud, C., Schwartz, J., et al. (2012). Searching for sharp drops in the incidence of pandemic A/H1N1 influenza by single year of age. *PloS one*, 7(8):e42328.

Justwan, A., Koç, N., and Jennings, A. E. (2008). Evolution of the Irminger and East Icelandic Current systems through the Holocene, revealed by diatom-based sea surface temperature reconstructions. *Quaternary Science Reviews*, 27(15):1571–1582.

Kim, C. and Marron, J. S. (2006). SiZer for jump detection. *Nonparametric Statistics*, 18(1):13–20.

Koenderink, J. (1984). The structure of images. *Biological Cybernetics*, 50(5):363–370.

Korhola, A., Weckström, J., Holmström, L., and Erästö, P. (2000). A quantitative Holocene climatic record from diatoms in northern Fennoscandia. *Quaternary Research*, 54:284–294.

Korsnes, R., Pavlova, O., and Godtliebsen, F. (2002). Assessment of potential transport of pollutants into the barents sea via sea ice–an observational approach. *Marine pollution bulletin*, 44(9):861–869.

Li, N. and Xu, X. (2015). Spline Multiscale Smoothing to Control FDR for Exploring Features of Regression Curves. To appear in Journal of Computational and Graphical Statistics.

Li, R. and Marron, J. (2005). Local likelihood SiZer map. *Sankhyā: The Indian Journal of Statistics*, 67(3):476–498.

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

Marron, J. and Chung, S. S. (2001). Presentation of smoothers: the family approach. *Computational Statistics*, 16(1):195–207.

Marron, J. and de Uña-Álvarez, J. (2004). SiZer for length biased, censored density and hazard estimation. *Journal of Statistical Planning and Inference*, 121:149–161.

Marron, J. and Zhang, J. T. (2005). Sizer for smoothing splines. *Computational Statistics*, 20(3):481–502.

Marron, J. S., Hernández-Campos, F., and Smith, F. D. (2004). *A SiZer analysis of IP flow start times*, volume Volume 44 of *Lecture Notes–Monograph Series*, pages 87–105. Institute of Mathematical Statistics, Beachwood, OH.

Martínez-Miranda, D. (2005). *SiZer Map for Evaluation a Bootstrap Local Bandwidth Selector in Nonparametric Additive Models*. Reports in statistics and operations research. Universidade de Santiago de Compostela, Departamento de Estatística e Investigación Operativa.

Marvel, K., Ivanova, D., and Taylor, K. E. (2013). Scale space methods for climate model analysis. *Journal of Geophysical Research: Atmospheres*, 118(11):5082–5097.

McClymont, E. L., Sosdian, S. M., Rosell-Melé, A., and Rosenthal, Y. (2013). Pleistocene sea-surface temperature evolution: Early cooling, delayed glacial intensification, and implications for the mid-Pleistocene climate transition. *Earth-Science Reviews*, 123:173–193.

Miettinen, A., Divine, D., Koç, N., Godtliebsen, F., and Hall, I. R. (2012). Multicentennial Variability of the Sea Surface Temperature Gradient across the Subpolar North Atlantic over the Last 2.8 kyr. *Journal of Climate*,

25(12):4205–4219.

Minnotte, M., Marchette, D., and Wegman, E. (1998). The bumpy road to the mode forest. *Journal of Computational and graphical Statistics*, 7:239–251.

Minnotte, M. and Scott, D. (1993). The mode tree: a tool for visualization of nonparametric density estimates. *Journal of Computational and graphical Statistics*, 2:51–68.

Miranda, R., Miranda, M., and Carmona, A. (2002). Exploring the structure of regression surfaces by using SiZer map for additive models. In Härdle, W. and Rönz, B., editors, *Compstat*, pages 361–366. Physica-Verlag HD.

Møllersen, K., Kirchesch, H. M., Schopf, T. G., and Godtliebsen, F. (2010). Unsupervised segmentation for digital dermoscopic images. *Skin Research and Technology*, 16(4):401–407.

Mortensen, K. E., Godtliebsen, F., and Revhaug, A. (2006). Scale-space analysis of time series in circulatory research. *American Journal of Physiology-Heart and Circulatory Physiology*, 291(6):H3012–H3022.

Müller, H.-G. and Wai, N. (2004). Change trees and mutagrams for the visualization of local changes in sequence data. *Journal of Computational and Graphical Statistics*, 13(3):571–585.

Øigård, T., Rue, H., and Godtliebsen, F. (2006). Bayesian multiscale analysis for time series data. *Computational Statistics and Data Analysis*, 51(3):1719–1730.

Oliveira, M., Crujeiras, R., and Rodríguez-Casal, A. (2014). CircSiZer: an exploratory tool for circular data. *Environmental and Ecological Statistics*, 21(1):143–159.

Olsen, L., Chaudhuri, P., and Godtliebsen, F. (2008). Multiscale spectral analysis for detecting short and long range change points in time series. *Computational Statistics and Data Analysis*, 52:3310–3330.

Olsen, L., Sørbyea, S., and Godtliebsen, F. (2007). A scale-space approach for detecting non-stationarities in time series. *Scandinavian Journal of Statistics*, 35:119–138.

Park, C., Ahn, J., Hendry, M., and Jang, W. (2011). Analysis of long period variable stars with nonparametric tests for trend detection. *Journal of the American Statistical Association*, 106(495):832–845.

Park, C., Godtliebsen, F., Taqqu, M., Stoev, S., and Marron, J. (2007). Visualization and inference based on wavelet coefficients, SiZer and SiNos. *Computational Statistics and Data Analysis*, 51:5994–6012.

Park, C., Hannig, J., and Kang, K.-H. (2009). Improved SiZer for time series. *Statistica Sinica*, 19(4):1511.

Park, C., Hannig, J., and Kang, K.-H. (2014). Nonparametric comparison of multiple regression curves in scale-space. *Journal of Computational and Graphical Statistics*, 23(3):657–677.

Park, C., Hernández-Campos, F., Marron, J., and Smith, F. D. (2005). Long-range dependence in a changing internet traffic mix. *Computer Networks*, 48(3):401–422.

Park, C. and Huh, J. (2013a). Nonparametric estimation of a log-variance function in scale-space. *Journal of Statistical Planning and Inference*, 143(10):1766 – 1780.

Park, C. and Huh, J. (2013b). Statistical inference and visualization in scale-space using local likelihood. *Computational Statistics & Data Analysis*, 57(1):336–348.

Park, C. and Kang, K.-H. (2008). SiZer analysis for the comparison of regression curves. *Computational Statistics & Data Analysis*, 52(8):3954–3970.

Park, C., Lazar, N. A., Ahn, J., and Sornborger, A. (2010a). A multiscale analysis of the temporal characteristics of resting-state fMRI data. *Journal of Neuroscience Methods*, 193(2):334 – 342.

Park, C., Lee, T. C., and Hannig, J. (2010b). Multiscale exploratory analysis of regression quantiles using quantile SiZer. *Journal of Computational and Graphical Statistics*, 19(3):497–513.

Park, C., Marron, J., and Rondonotti, V. (2004). Dependent SiZer: goodness-of-fit tests for time series models. *Journal of Applied Statistics*, 31(8):999–1017.

Park, C., Shen, H., Marron, J., Hernández-Campos, F., and Veitch, D. (2006). Capturing the Elusive Poissonity in Web Traffic. In *Modeling, Analysis, and Simulation of Computer and Telecommunication Systems, 2006. MASCOTS 2006. 14th IEEE International Symposium on*, pages 189–196.

Pasanen, L. and Holmström, L. (2015). Bayesian scale space analysis of temporal changes in satellite images. *Journal of Applied Statistics*, 42(1):50–70.

Pasanen, L., Holmström, L., and Sillanpää, M. J. (2014). Bayesian LASSO, scale space and decision making in association genetics. To appear.

Pasanen, L., Launonen, I., and Holmström, L. (2013). A scale space multiresolution method for extraction of time series features. *Stat*, 2(1):273–291.

Pedersen, C., Godtliebsen, F., and Roesch, A. (2008). A scale-space approach for detecting significant differences between models and observations using global albedo distributions. *Journal of Geophysical Research*, 113:D10108

doi:10.1029/2007JD009340.

Percival, D. and Walden, A. (2006). *Wavelet Methods for Time Series Analysis*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

Perona, P. and Malik, J. (1990). Scale-space and edge detection using anisotropic diffusion. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 12(7):629–639.

Pittau, M. G. and Zelli, R. (2007). Exploring patterns of income polarization using SiZer. *Journal of Quantitative Economics*, 5(1):101.

Poline, J.-B. and Mazoyer, B. (1994). Analysis of individual brain activation maps using hierarchical description and multiscale detection. *Medical Imaging, IEEE Transactions on*, 13(4):702–710.

Roca-Pardiñas, J., Cadarso-Suárez, C., Pardo-Vazquez, J. L., Leboran, V., Molenberghs, G., Faes, C., and Acuna, C. (2011). Assessing neural activity related to decision-making through flexible odds ratio curves and their derivatives. *Statistics in medicine*, 30(14):1695–1711.

Rohling, E. and Pälike, H. (2005). Centennial-scale climate cooling with a sudden cold event around 8,200 years ago. *Nature*, 434:975–979.

Rondonotti, V., Marron, J., and Park, C. (2007). SiZer for time series analysis: A new approach to the analysis of trends. *Electronic Journal of Statistics*, 1:268–289.

Rooper, L. M., Ali, S. Z., and Olson, M. T. (2014). A minimum fluid volume of 75 ml is needed to ensure adequacy in a pleural effusion: A retrospective analysis of 2540 cases. *Cancer cytopathology*, 122(9):657–665.

Rudge, J. F. (2008). Finding peaks in geochemical distributions: A re-examination of the helium-continental crust correlation. *Earth and Planetary Science Letters*, 274(1):179–188.

Rufibach, K. and Walther, G. (2010). The block criterion for multiscale inference about a density, with applications to other multiscale problems. *Journal of Computational and Graphical Statistics*, 19(1):175–190.

Rydén, J. (2010). Exploring possibly increasing trend of hurricane activity by a SiZer approach. *Environmental and ecological statistics*, 17(1):125–132.

Salganik, M., Milford, E., Hardie, D., Shaw, S., and Wand, M. (2005). Classifying antibodies using flow cytometry data: class prediction and class discovery. *Biometrical journal*, 47(5):740–754.

Skrøvseth, S., Schopf, T., Thon, K., Zortea, M., Geilhufe, M., Mollersen, K., Kirchesch, H., and Godtliebsen, F. (2010). A computer aided diagnostic system for malignant melanomas. In *Applied Sciences in Biomedical and Communication Technologies (ISABEL), 2010 3rd International Symposium on*, pages 1–5.

Skrøvseth, S. O., Årsand, E., Godtliebsen, F., and Hartvigsen, G. (2012a). Mobile phone-based pattern recognition and data analysis for patients with type 1 diabetes. *Diabetes technology & therapeutics*, 14(12):1098–1104.

Skrøvseth, S. O., Arsand, E., Godtliebsen, F., and Joakimsen, R. M. (2012b). Model driven mobile care for patients with type 1 diabetes. *Studies in Health Technology and Informatics*, 180:1045–9.

Skrøvseth, S. O., Bellika, J. G., and Godtliebsen, F. (2012c). Causality in scale space as an approach to change detection. *PloS one*, 7(12):e52253.

Skrøvseth, S. O. and Godtliebsen, F. (2011). Scale space methods for analysis of type 2 diabetes patients' blood glucose values. *Computational and mathematical methods in medicine*, 2011.

Skrøvseth, S. O., Dias, A., Gorzelniak, L., Godtliebsen, F., and Horsch, A. (2012). Scale-space methods for live processing of sensor data. *Studies in Health Technology and Informatics*, 180:138–142.

Sonderegger, D. L., Wang, H., Clements, W. H., and Noon, B. R. (2008). Using SiZer to detect thresholds in ecological data. *Frontiers in Ecology and the Environment*, 7(4):190–195.

Sørbye, S., Hindberg, K., Olsen, L., and Rue, H. (2009). Bayesian multiscale feature detection of log-spectral densities. *Computational Statistics and Data Analysis*, 53(11):3746–3754.

Sporring, J. (1997). *Gaussian Scale-Space Theory*. Computational imaging and vision. Kluwer Academic Publishers, Dordrecht.

Thon, K., Rue, H., Skrøvseth, S. O., and Godtliebsen, F. (2012). Bayesian multiscale analysis of images modeled as Gaussian Markov random fields. *Computational Statistics & Data Analysis*, 56(1):49 – 61.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B*, 58(1):267–288.

Vaughan, A., Jun, M., and Park, C. (2012). Statistical inference and visualization in scale-space for spatially dependent images. *Journal of the Korean Statistical Society*, 41(1):115 – 135.

Vekemans, D., Proost, S., Vanneste, K., Coenen, H., Viaene, T., Ruelens, P., Maere, S., Van de Peer, Y., and Geuten, K. (2012). Gamma paleohexaploidy in the stem lineage of core eudicots: significance for MADS-box

gene and species diversification. *Molecular biology and evolution*, page mss183.

Vidakovic, B. (1999). *Statistical Modeling by Wavelets.* Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., New York.

Wahba, G. (1978). Improper priors, spline smoothing and the problem of guarding against model errors in regression. *Journal of Royal Statistical Society, Series B*, 40(3):364–372.

Wang, X. and Marron, J. S. (2008). A scale-based approach to finding effective dimensionality in manifold learning. *Electronic Journal of Statistics*, 2:127–148.

Wang, Y.-P. and Lee, S. L. (1998). Scale-space derived from B-splines. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20:1040–1055.

Weckström, J., Korhola, A., Erästö, P., and Holmström, L. (2006). Temperature Patterns over the Past Eight Centuries in Northern Fennoscandia Inferred from Sedimentary Diatoms. *Quaternary Research*, 66:78–86.

Weickert, J., Ishikawa, S., and Imiya, A. (1999). Linear scale-space has first been proposed in Japan. *Journal of Mathematical Imaging and Vision*, 10(3):237–252.

Weis, A. E., Wadgymar, S. M., Sekor, M., and Franks, S. J. (2014). The shape of selection: using alternative fitness functions to test predictions for selection on flowering time. *Evolutionary Ecology*, 28(5):885–904.

Wilson, L., Hald, M., and Godtliebsen, F. (2011). Foraminiferal faunal evidence of twentieth-century Barents Sea warming. *The Holocene*, 21(4):527–537.

Witkin, A. P. (1983). Scale-space filtering. In *8ht International Joint Conference of Artificial Intelligence*, pages 1019–1022.

Witkin, A. P. (1984). Scale-space filtering: A new approach to multi-scale description. In *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '84.*, volume 9, pages 150–153.

Worsley, K. J., Marrett, S., Neelin, P., and Evans, A. (1996). Searching scale space for activation in PET images. *Human brain mapping*, 4(1):74–90.

Zamelczyk, K., Rasmussen, T. L., Husum, K., Godtliebsen, F., and Hald, M. (2014). Surface water conditions and calcium carbonate preservation in the fram strait during marine isotope stage 2, 28.8ñ15.4 kyr. *Paleoceanography*, 29(1):1–12.

Zeng, Q., Wand, M., Young, A. J., Rawn, J., Milford, E. L., Mentzer, S. J., and Greenes, R. A. (2002). Matching of flow-cytometry histograms using information theory in feature space. In *Proceedings of the AMIA Symposium*, page 929. American Medical Informatics Association.

Zhang, H.-G. and Mei, C.-L. (2012). SiZer inference for varying coefficient models. *Communications in Statistics-Simulation and Computation*, 41(10):1944–1959.

Zhang, H.-G., Mei, C.-L., and Wang, H.-L. (2013). Robust SiZer approach for varying coefficient models. *Mathematical Problems in Engineering*, 2013.

Zhang, L. (2007). *Functional Singular Value Decomposition and Multi-Resolution Anomaly Detection.* PhD thesis, University of North Carolina at Chapel Hill.

Zhang, L., Zhu, Z., Jeffay, K., Marron, J., and Smith, F. (2008). Multi-Resolution Anomaly Detection for the Internet. In *INFOCOM Workshops 2008, IEEE*, pages 1–6.

Zhang, L., Zhu, Z., and Marron, J. S. (2014). Multiresolution anomaly detection method for fractional Gaussian noise. *Journal of Applied Statistics*, 41(4):769–784.

Zhao, X., Marron, J., and Wells, M. T. (2004). The functional data analysis view of longitudinal data. *Statistica Sinica*, 14(3):789–808.